HAL
open science

Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite

Guillaume Deconde

To cite this version:

Guillaume Deconde. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite. Psychologie. Université Rennes 2; Université Européenne de Bretagne, 2009. Français. NNT : 2009REN20064 . tel-00585247

HAL Id: tel-00585247

https://theses.hal.science/tel-00585247v1

Submitted on 12 Apr 2011

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.

Sous le sceau de l'Université Européenne de Bretagne

UNIVERSITE RENNES 2

Ecole Doctorale - Sciences Humaines et Sociales

Centre de Recherches en Psychologie, Cognition et Communication -EA 1285-

ETUDE ITERATIVE DES LIENS ENTRE UTILISABILITE ET ACCEPTABILITE D'UN DISPOSITIF DE SAISIE ET DE RECONNAISSANCE DE L'ECRITURE MANUSCRITE

THESE
présentée en vue de l'obtention du grade de docteur en Psychologie Expérimentale,
soutenue publiquement par

Guillaume DECONDE

Directeurs de recherche
Eric JAMET (professeur des universités, Université de Haute Bretagne)
Alain LIEURY (professeur des universités, Université de Haute Bretagne)

Membres du jury
Annie PIOLAT, professeur des universités, Université de Provence (Rapporteur)
André TRICOT, professeur des universités, Université de Toulouse-le-Mirail (Rapporteur)
Eric ANQUETIL, maître de conférences, INSA Rennes
Eric JAMET, professeur des universités, Université Rennes 2 (Directeur de thèse)

Remerciements

Un immense merci à Eric Jamet pour sa patience. Me faire partager ses qualités d'expérimentateur et la justesse de ses analyses m'a permis de devenir un apprenti chercheur.

Un grand merci à Alain Lieury pour sa gentillesse et sa disponibilité. Ses connaissances encyclopédiques me montrent tout le chemin qu'il reste à parcourir.

Mes remerciements s'adressent aussi à Eric Anquetil et François Bouteruche de l'IRISA, qui m'ont fait bénéficier de leurs compétences en informatique en développant DIGIME, pierre angulaire de cette thèse.

Je remercie également Lucian Olar, Katharina Haaf et Hélène Hardy pour leur aide à l'élaboration et la passation des expériences 1 et 2.

Mille mercis vont à tous les membres du Laboratoire de Psychologie Expérimentale, qu'ils soient étudiants ou enseignants. Chacun d'entre eux m'a apporté son aide quand j'en avais besoin.

Pour finir, un très très gros merci à Mounette et Lou Pape pour leur soutien indéfectible.

Sommaire

REMERCIEMENTS **2**

SOMMAIRE **3**

INTRODUCTION **8**

1 UTILITE, UTILISABILITE ET ACCEPTABILITE DE L'ECRITURE MANUSCRITE SUR POCKET PC **11**

1.1. DE L'UTILITE DE LA PRISE DE NOTES SUR PDA A L'ACCEPTABILITE DE LA RECONNAISSANCE D'ECRITURE MANUSCRITE : LA NOTION D'UTILISABILITE **11**
1.2. HISTORIQUE DU CONCEPT D'UTILISABILITE **13**
1.3. UTILISABILITE OU FACILITE D'UTILISATION PERÇUE **14**
1.3.1. L'UTILISABILITE : UN ATTRIBUT DES PRODUITS ACCEPTABLES 14
1.3.2. LA FACILITE D'UTILISATION : UNE CROYANCE PRECURSEUR DE L'ACCEPTATION 32
1.4. VERS UNE UNIFICATION DES MODELES D'UTILISABILITE ET D'ACCEPTATION **47**
1.5. MESURER ET AMELIORER L'UTILISABILITE **49**
1.5.1. L'UTILISABILITE COMME PROCESSUS ET METHODE DE CONCEPTION 50
1.5.2. L'INSPECTION ERGONOMIQUE 51
1.5.3. LE TEST UTILISATEURS 52
1.5.4. LIMITES DES METHODES D'EVALUATION DE L'UTILISABILITE 58

2- ECRITURE ET MEMOIRE DE TRAVAIL (MDT) **61**

2.1. L'ECRITURE **61**
2.1.1. L'ECRITURE COMME OBJET D'ETUDE 61
2.1.2. REDACTION ET PRISE DE NOTES (PDN) 62
2.2. LA MEMOIRE DE TRAVAIL **66**
2.2.1. LE MODELE DE MEMOIRE DE TRAVAIL COMPONENTIEL DE BADDELEY (2000) 66
2.3. MEMOIRE DE TRAVAIL ET REDACTION DE TEXTES **74**
2.3.1. LE MODELE DE HAYES (1996) 74
2.3.2. LE MODELE DE KELLOGG (1996) 76
2.3.3. CRITIQUE DE L'INTEGRATION DE LA MEMOIRE DE TRAVAIL AUX MODELES DE PRODUCTION ECRITE 86
2.4. LA PRODUCTION GRAPHIQUE DE LETTRES **88**
2.4.1. DU MOT A LA LETTRE : LES REGISTRES MNESIQUES SOLLICITES PAR LES PROCESSUS ORTHOGRAPHIQUES 89
2.4.2. REALISATION GRAPHIQUE DES LETTRES 91
2.5. DE LA MEMOIRE DE TRAVAIL A LA PRODUCTION GRAPHIQUE DE LETTRES **94**

3- MEMOIRE, APPRENTISSAGE ET PERFORMANCE **95**

3.1. ENVIRONNEMENTS INFORMATISES POUR L'APPRENTISSAGE HUMAIN (EIAH) **95**
3.2. ARCHITECTURE COGNITIVE ET FONCTIONNEMENT MNESIQUE **96**
3.3. APPRENTISSAGE, MOTRICITE ET IMAGERIE MENTALE **98**
3.3.1. PASSER DU SAVOIR AU SAVOIR-FAIRE 100
3.3.2. FAVORISER L'APPRENTISSAGE PROCEDURAL ET MOTEUR 103

3.4. LA THEORIE DE LA CHARGE COGNITIVE (CLT, SWELLER, 1988)..... 110
3.4.1. LA CHARGE COGNITIVE 111
3.5. EXPERTISE ET ADAPTABILITE 112
3.5.1. L'EXPERTISE EN PSYCHOLOGIE COGNITIVE 113
3.5.2. ADAPTABILITE 116
3.6. MESURER LA CHARGE COGNITIVE..... 118
3.6.1. CHARGE MENTALE..... 119
3.6.2. EFFORT MENTAL ET PERFORMANCE 120
3.6.3. L'EFFICIENCE MENTALE 122
3.7. FACTEURS D'AUGMENTATION DE LA CHARGE INUTILE ET REMEDIATIONS..... 125
3.7.1. EFFETS DE PARTAGE DE L'ATTENTION ET DE CONTIGUÏTE SPATIALE..... 125
3.7.2. LE PRINCIPE DE PROXIMITE COMPATIBILITE (PROXIMITY COMPATIBILITY PRINCIPLE, PCP, WICKENS & CARSWELL, 1995) 126
3.8. CRITIQUE DE LA NOTION DE CHARGE..... 128

4- EXPERIENCE 1 : QUALITE ERGONOMIQUE DE L'INTERFACE DIGIME: EFFETS DE LA CONTIGUÏTE SPATIALE DU FEEDBACK ET DE LA CONSERVATION A L'ECRAN DES LETTRES SAISIES LORS D'UNE TACHE DE SAISIE DE TEXTE MANUSCRIT SUR PDA..... 130

4.1. INTRODUCTION..... 131
4.2. METHODE 136
4.2.1. PARTICIPANTS..... 136
4.2.2. MATERIEL 136
4.2.3. PROCEDURE EXPERIMENTALE 142
4.2.4. VARIABLES INDEPENDANTES..... 143
4.2.5. VARIABLES DEPENDANTES 146
4.2.6. HYPOTHESES..... 147
4.3. RESULTATS 147
4.3.1. EFFETS DES VI SUR LES PERFORMANCES..... 148
4.3.2. EFFETS DES VI SUR LES TESTS SUBJECTIFS..... 153
4.4. DISCUSSION..... 154

5- EXPERIENCE 2 : CREATION D'UN DIDACTICIEL POUR DIGIME : EFFETS DES ILLUSTRATIONS STATIQUES OU DYNAMIQUES SUR L'APPRENTISSAGE DU FONCTIONNEMENT DE L'INTERFACE ET SUR SON UTILISABILITE PERÇUE..... 160

5.1. INTRODUCTION..... 161
5.2. METHODE 165
5.2.1. PARTICIPANTS..... 165
5.2.2. MATERIEL 165
5.2.3. PROCEDURE EXPERIMENTALE 172
5.2.4. VARIABLES INDEPENDANTES..... 173
5.2.5. VARIABLES DEPENDANTES 173
5.2.6. HYPOTHESES..... 174
5.3. RESULTATS 174
5.3.1. EFFETS DES VI SUR LES PERFORMANCES..... 174
5.3.2. EFFETS DU TYPE DE DIDACTICIEL SUR L'UTILISABILITE PERÇUE 180
5.3.3. EFFETS DE LA PRESENCE D'UN DIDACTICIEL 181
5.4. DISCUSSION..... 182

6- EXPERIENCE 3 : TEST UTILISATEUR : COMPARAISON DE DIGIME AUX METHODES ALTERNATIVES DE SAISIE DE TEXTE SUR PDA 187

6.1. INTRODUCTION 188
6.2. METHODE 191
6.2.1. PARTICIPANTS 191
6.2.2. MATERIEL 191
6.2.3. PROCEDURE EXPERIMENTALE 199
6.2.4. VARIABLES INDEPENDANTES 199
6.2.5. VARIABLES DEPENDANTES 200
6.3. RESULTATS 200
6.3.1. EFFETS DU TYPE D'INTERFACE SUR LES TEMPS DE SAISIE A L'ENTRAINEMENT ET A LA PRISE DE NOTES 200
6.3.2. EFFETS DU TYPE D'INTERFACE SUR LES JUGEMENTS 202
6.4. DISCUSSION 205

7- EXPERIENCE 4 : EFFET DU NIVEAU D'EXPERTISE EN INFORMATIQUE SUR LES PERFORMANCES ET L'INTENTION D'UTILISATION DE DIGIME 209

7.1. INTRODUCTION 210
7.2. METHODE 213
7.2.1. PARTICIPANTS 213
7.2.2. MATERIEL 213
7.2.3. PROCEDURE EXPERIMENTALE 216
7.2.4. VARIABLES INDEPENDANTES 216
7.2.5. VARIABLES DEPENDANTES 216
7.2.6. HYPOTHESES 217
7.3. RESULTATS 217
7.3.1. EFFETS DU NIVEAU D'EXPERTISE EN INFORMATIQUE AU PRE-QUESTIONNAIRE 217
7.3.2. EFFETS DU NIVEAU D'EXPERTISE EN INFORMATIQUE AU TEST DE PERCEPTION 218
7.3.3. EFFETS DU NIVEAU D'EXPERTISE EN INFORMATIQUE SUR LES PERFORMANCES 220
7.3.4. EFFETS DU NIVEAU D'EXPERTISE EN INFORMATIQUE SUR LES JUGEMENTS 220
7.4. DISCUSSION 222

8- EXPERIENCE 5 : UTILISABILITE OBJECTIVE ET SUBJECTIVE : PREDIRE L'ACCEPTABILITE COMPAREE ENTRE DIGIME ET *NOTE* 226

8.1. INTRODUCTION 227
8.2. METHODE 232
8.2.1. PARTICIPANTS 232
8.2.2. MATERIEL 232
8.2.3. PROCEDURE EXPERIMENTALE 235
8.2.4. VARIABLES INDEPENDANTES 237
8.2.5. VARIABLES DEPENDANTES 237
8.2.6. HYPOTHESES 238
8.3. RESULTATS 240
8.3.1. EFFETS DES INTERFACES SUR LES VD (H1) 240
8.3.2. EFFETS DES TACHES CONCURRENTES SUR LA TACHE PRINCIPALE ET SUR LES JUGEMENTS (H2) 241
8.3.3. EFFETS DES MESURES DE PERFORMANCES SUR LA CHARGE DE TACHE MENTALE (H4, H5 ET H6) 241

8.3.4. EFFETS DES MESURES D'UTILISABILITE SUR L'ACCEPTABILITE COMPAREE (H6, H7, H8, H9 ET H10) 244
8.4. DISCUSSION 245

9- EXPERIENCE 5 BIS : UTILISABILITE OBJECTIVE ET SUBJECTIVE : PREDIRE L'ACCEPTABILITE COMPAREE ENTRE DIGIME ET LE CLAVIER VIRTUEL 250

9.1. INTRODUCTION 251
9.2. METHODE 253
9.2.1. PARTICIPANTS 253
9.2.2. MATERIEL 253
9.2.3. PROCEDURE EXPERIMENTALE 255
9.2.4. VARIABLES INDEPENDANTES 255
9.2.5. VARIABLES DEPENDANTES 256
9.2.6. HYPOTHESES 256
9.3. RESULTATS 258
9.3.1. ANALYSE FACTORIELLE 258
9.3.2. TEST DES HYPOTHESES 261
9.4. DISCUSSION 271

10- DISCUSSION GENERALE 275

11- BIBLIOGRAPHIE 285

12- INDEX DES FIGURES 311

13- INDEX DES TABLEAUX 313

14- INDEX DES FORMULES 313

15- ANNEXES 314

15.1. CARACTERISTIQUES LINGUISTIQUES DES MOTS UTILISES POUR L'EXPERIENCE 1 314
15.2. CARACTERISTIQUES LINGUISTIQUES DES MOTS UTILISES POUR L'EXPERIENCE 2 314
15.3. QUESTIONNAIRES 315
15.3.1. PRE-QUESTIONNAIRE 315
15.3.2. NASA-TLX 316
15.3.3. TEST D'UTILISABILITE DE NIELSEN 318
15.3.4. QUESTIONS ADDITIONNELLES 318
15.3.5. TEST SUS (SYSTEM USABILITY SCALE) 319
15.3.6. TEST TAM (TECHNOLOGY ACCEPTANCE MODEL) 320
15.4. DIDACTICIELS EXPERIENCE 1 322
15.4.1. CONDITION "CONSERVATION DE L'ENCRE AVEC FEEDBACK CONTIGU" 322
15.4.2. CONDITION "CONSERVATION DE L'ENCRE SANS FEEDBACK CONTIGU" 323
15.4.3. CONDITION "DISPARITION DE L'ENCRE AVEC FEEDBACK CONTIGU" 325
15.4.4. CONDITION "DISPARITION DE L'ENCRE SANS FEEDBACK CONTIGU" 326
15.5. DIDACTICIEL EXPERIENCE 2 328
15.5.1. CONDITION STATIQUE 328

15.6. ANALYSE DES EFFETS DU TYPE D'ILLUSTRATION SUR LES TEMPS DE SAISIE, LE NOMBRE D'ERREURS, LES DUREES ET OCCURRENCES DE CONSULTATION DES AIDES EN LIGNES EN FONCTION DU TYPE DE LETTRE 332
15.7. TEXTES UTILISES POUR LES EXPERIENCES 5 ET 5 BIS 337
15.7.1. LE CHAT 337
15.7.2. LE CHIEN 338

Introduction

. Un PDA (*Personal Digital Assistant*, littéralement, assistant numérique personnel) est un ordinateur de poche. Composé d'un processeur, de mémoires, d'un écran tactile avec stylet et de fonctionnalités réseau dans un boîtier de petite taille, son usage est prévu originellement dans un but d'organisation. Un assistant personnel fournit donc généralement des applications telles que gestionnaire de tâche, agenda, carnet d'adresse (coordonnées et messagerie électronique) ainsi qu'un client de messagerie rendant possible la consultation de son courrier électronique et la rédaction de nouveaux messages.

Ayant fait son apparition il y a maintenant une quinzaine d'années, la recherche concernant la saisie de texte sur les dispositifs mobiles de petite taille (tels que Pocket PC, PDA ou Smartphones) connaît aujourd'hui un essor important. MacKenzie et Soukoreff attribuent à cela plusieurs raisons (2002) :

- Premièrement, l'informatique mobile a engendré de nouveaux domaines d'applications (*Web* mobile, ubiquité, accès permanent à sa messagerie électronique, *etc.*).
- Deuxièmement, l'amélioration croissante des capacités de traitement des périphériques mobiles les rend toujours plus efficaces, et donc, toujours plus utiles.
- Troisièmement, un besoin fort des utilisateurs de méthodes de saisie de texte performantes et ergonomiques sur ces machines.
- Quatrièmement, et pour terminer, le succès incroyable de la messagerie instantanée sur téléphone, les très célèbres SMS (*Short Message Service*), dont la croissance exponentielle inspire nombre de spéculations quant aux développements à venir de la saisie de texte nomade[1].

S'il est vrai que le clavier ne connaît pour ainsi dire aucune concurrence en ce qui concerne les ordinateurs de bureau, il n'en va pas de même pour les ordinateurs de poche. Pourtant familier des utilisateurs de postes de travail, le choix du clavier comme méthode de

[1] A titre indicatif, 1881 milliards de SMS ont été envoyés au cours du second trimestre de l'année 2003 sur les trois réseaux français de téléphonie mobile existant à cette date. Il n'est pas possible d'accéder aux chiffres des opérateurs qui protègent leur stratégie marketing. Les chiffres d'organismes tels que l'ART (Autorité de Régulation de la Téléphonie) ou de l'INSEE (Institut National de la Statistique et des Etudes Economiques) sont accessibles mais peuvent témoigner d'une large variabilité (Martin, 2003). Quoi qu'il en soit, les chiffres sont astronomiques et ont vraisemblablement dû décupler depuis.

saisie de texte aurait pu également se généraliser à l'informatique nomade, si ce n'est la place qu'il occupe, la petitesse des touches que cela induit, et le choix souvent fait par les constructeurs de PDA de l'y remplacer par le tandem stylet-écran tactile (voir figure 1 ci-dessous). Par conséquent, un clavier, fût-il virtuel, ne figure qu'au rang d'option, parmi d'autres, pour écrire sur les terminaux portables. Les alternatives sont nombreuses, et nous ne les évoquerons pas toutes ici (voir Bouteruche, Deconde, Anquetil, & Jamet, 2005; Commaford, 2004; Curran, Woods, & O Riordan, 2006; Hinckeley, 2003; Jansen & Ayers, 2004; Mackenzie & Chang 1999; MacKenzie, Nonnecke, McQueen, Riddersma & Meltz, 1994; MacKenzie & Soukoreff, 2002; MacKenzie & Zhang, 1997; McQueen, MacKenzie, Nonnecke, Riddersma, & Meltz, 1994; Ortelli, 2003; Sears & Arora, 2002). Notre travail s'orientera sur l'une d'elles particulièrement, à savoir, la reconnaissance automatique d'écriture manuscrite produite grâce au stylet sur l'écran tactile du dispositif.

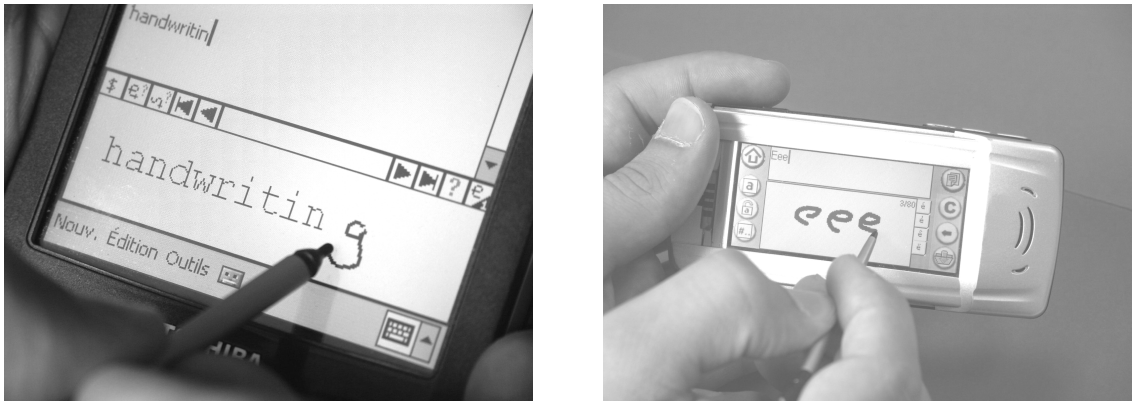


Figure 1: Dispositifs mobiles de petite taille (PDA à gauche et Smartphone à droite) équipés de stylets et d'écran tactile.

A ce titre, notre recherche vise à comprendre les processus cognitifs induits par l'utilisation d'un logiciel de saisie et de reconnaissance automatique de l'écriture manuscrite développé par l'IRISA (Institut de Recherche en Informatique et Systèmes Aléatoires) et désigné DIGIME (*DIGital Ink Micro-Editor*) (Anquetil & Bouchereau, 2002; Anquetil & Bouteruche, 2004; Bouteruche, Deconde, Anquetil, & Jamet, 2005). Le but est à terme d'en comprendre les mécanismes d'acceptabilité, c'est-à-dire ceux qui conduisent à faire en sorte qu'il soit choisi préférentiellement par les utilisateurs potentiels d'informatique nomade. Sur ce dernier point, il est à noter qu'il n'existe que peu d'études comparatives entre méthodes de saisie, si ce ne sont, par exemple, celle de Sears et Arora (2002) qui compare les logiciels *JOT*

et *Graffiti* en terme de vitesse et de satisfaction, ou celle de McKenzie et Zhang (1997) qui teste l'utilisabilité immédiate de *Graffiti*. Hélas, ce type d'études a un caractère résolument descriptif. Cela ne nous informe que peu sur les causes des différents jugements et performances engendrés par l'utilisation de l'une ou l'autre de ces méthodes de saisie. Plus encore, ces études ne permettent pas non plus de prédire dans quelle mesure ces méthodes de saisie engendreront chez l'utilisateur un comportement d'adoption.

1 Utilité, utilisabilité et acceptabilité de l'écriture manuscrite sur Pocket PC

1.1. De l'utilité de la prise de notes sur PDA à l'acceptabilité de la reconnaissance d'écriture manuscrite : la notion d'utilisabilité

Si les évolutions technologiques récentes font des PDA de véritables stations multimédia à l'instar de n'importe quel ordinateur, ce sont ses fonctions de base associées à sa portabilité qui donnent toute son utilité à ce dispositif en ce sens qu'il remplace le classique agenda papier. A cet égard, il joue un rôle de stockage externe d'informations disponibles hors du contexte de leur émission: on peut par conséquent l'assimiler, au moins en partie, à un dispositif électronique de prise de notes (Benton, Kiewra, Whitfill, & Dennison, 1993). Pour permettre la saisie de texte sur Pocket PC, plusieurs solutions existent. Qu'il s'agisse du clavier virtuel, reproduction sur l'écran tactile du « classique » clavier à touches, ou de tout autre dispositif, tels *Graffiti* ou *JOT*, misant respectivement sur la reconnaissance de formes ou de lettres manuscrites, ces dispositifs logiciels constituent une part importante de l'interface homme-machine (IHM) du PDA. Or, toutes ces méthodes de saisie ne sont pas équivalentes, aussi bien en termes qualitatifs qu'en termes de performance ou de satisfaction induite. Il est vrai que les interfaces de saisie et de reconnaissance de texte manuscrit offrent une solution pertinente en termes d'encombrement (disparition du clavier à touches et de la souris, remplacés par un dispositif de pointage et de saisie composé d'un stylet associé à un écran tactile). C'est de surcroît une solution qu'on imagine être acceptable du point de vue de l'utilisateur (Buisson, Sallagoïty, Athènes, & Mertz, 2003). En effet, la maîtrise de l'écriture manuscrite par tout un chacun peut raisonnablement laisser penser que ces méthodes de saisie présentent un caractère intuitif propre à en favoriser l'usage. Les données empiriques montrent cependant qu'il n'en est rien, confirmant une fois encore, s'il en était besoin, que le bon sens ne suffit pas à élaborer des interfaces ergonomiques de qualité (Baccino, Bellino, & Colombi, 2005; Bastien & Scapin, 2001; Kolski, 1997). En effet, les dispositifs de reconnaissance automatique de l'écriture ne sont pas (ou peu) utilisés par leurs possesseurs (Ortelli, 2003), alors qu'ils affichent des taux de reconnaissance des caractères de l'ordre de 97% (équivalent à un lecteur expert, et potentiellement acceptable pour un utilisateur) (MacKenzie & Chang, 1999). De fait, des dispositifs alternatifs de saisie comme les claviers virtuels sont majoritairement choisis par les utilisateurs de PDA pour y entrer du texte. On

peut donc en conclure que l'utilité même de la saisie de texte sur Pocket PC n'est pas à remettre en question puisque c'est une nécessité de le faire pour quiconque utilise ces ordinateurs de poche. La question est de comprendre pourquoi ces dispositifs de saisie qui accomplissent la même tâche, et donc, ont la même utilité théorique, vont être, selon les cas, acceptables – et donc utilisés par les consommateurs. Tricot et ses collaborateurs se sont à plusieurs reprises penchés sur les liens qui régissent utilité et acceptabilité. Afin d'éclairer cette problématique, ces auteurs évoquent, dans le sillage de plusieurs de leurs prédécesseurs (Bevan, Kirakowski, & Maissel, 1991; Dillon & Watson, 1996; Nielsen, 1993; Shackel, 1991), la notion d'utilisabilité. Ce terme est apparu au début des années 80. Auparavant, on parlait de « convivialité des interfaces » (*user friendly*). Cet emploi a disparu à cause de son manque de clarté et de sa connotation subjective (Bevan, Kirakowski, & Maissel, 1991). On peut définir l'utilisabilité comme étant, dans son sens le plus large, la facilité avec laquelle un individu peut utiliser un objet donné de sorte qu'il accomplisse un objectif fixé. Cependant, il est prématuré à ce stade de notre exposé de définir de façon péremptoire ce concept, tant il est vrai qu'il recouvre de multiples facettes. Ceci est le reflet des nombreux champs d'études et d'applications qui se le sont appropriés, pour en faire une dimension explicative centrale de la qualité d'utilisation des dispositifs informatiques. Toute la difficulté à appréhender la notion d'utilisabilité réside dans l'hétérogénéité des déterminismes (ergonomie cognitive, conception centrée utilisateur, psychologie ergonomique, ingénierie cognitive, sociologie…) qui la définissent par des moyens nombreux (études de terrain ou de laboratoire, tests utilisateurs, mesures objectives, questionnaires, enquêtes…) et à des fins qui le sont tout autant (mercantiles, démarche qualité, amélioration de conditions de travail, *etc.*). Cette considération n'a pas échappé à Tricot et Tricot (2000), qui relèvent que depuis la publication de *Usability Engineering* (Nielsen, 1993), les références portant sur l'utilisabilité des systèmes d'informations se comptent en milliers, faisant parfois d'elle, une notion « fourre-tout ». Notre propos ne sera pas de présenter de façon exhaustive l'utilisabilité en tant que telle (voir Keinonen, 1998, pour une revue de question). Nous nous attacherons dans les lignes qui suivent, à préciser ce que nous considérons être l'utilisabilité, bien que ce débat reste d'actualité (Brangier & Barcenilla, 2003).

1.2. Historique du concept d'utilisabilité

Nous l'évoquions précédemment, les nombreux apports théoriques et méthodologiques au concept d'utilisabilité lui confèrent de multiples acceptions (Bevan, Kirakowski, & Maissel, 1991; Brangier & Barcenilla, 2003).

Historiquement, l'ergonomie (du grec *ergon*, travail, et *nomos*, savoir, règles) propose dès le début du XXe siècle, d'étudier l'adaptation du travail à l'homme. La diversité des situations de travail ont imposé à cette discipline son caractère pluridisciplinaire. Le travail d'antan étant plutôt physique et répétitif, le champ d'intervention ergonomique relevait principalement des compétences des médecins et physiologistes. Grâce à l'automatisation de nombreuses procédures et aux progrès de la robotique, le travail est devenu davantage mental, impliquant principalement une activité humaine de contrôle et de supervision (Zhang & Luximon, 2005). Cette transformation du travail n'a pas été sans susciter de nouveaux problèmes aux opérateurs : surcharge mnésique, contraintes perceptives et attentionnelles, *etc.* (Baccino, Bellino, & Colombi, 2005). L'étude de l'utilisabilité s'inscrit donc dans la lignée des travaux en ergonomie ainsi que dans l'histoire de la diffusion de l'outil informatique au grand public (Shackel, 2000) : expurgée de la dimension « travail » de l'ergonomie (impliquant notamment des aspects tels que sécurité et santé), l'utilisabilité va s'attacher à rendre compte de l'adéquation d'une IHM avec les capacités perceptivo-motrices (Chua, Weeks, & Goodman, 2003), cognitives, et plus récemment, conatives (Fogg, 2003) et émotionnelles des individus (Brave & Clifford, 2003). En fait, cette idée d'utilisabilité s'applique à tout type d'IHM, que l'usage en soit professionnel ou pas, et incluant notamment, les Technologies de l'Information et de la Communication (TIC), dont l'utilisation par le grand public n'est désormais plus nouvelle. Cependant, si Tricot et Tricot (2000) constatent qu'utilisabilité est de plus en plus synonyme de « qualité générale d'un système d'information », c'est dû dans une très large mesure au caractère composite et hétérogène de ce concept. En effet, les différents champs disciplinaires qui ont à le manipuler en tant que variable explicative lui confèrent des définitions et des propriétés qui peuvent très largement varier d'un auteur à l'autre.

Différents modèles existent. Ils suivent principalement deux orientations. Suivant le type de mesures qu'ils préconisent (objectives ou subjectives), ils vont être, soit orientés sur l'interaction avec la machine, soit orientés sur les attitudes, croyances et comportements de

l'utilisateur. Les deux approches sont complémentaires en l'absence actuelle d'un modèle fiable les unifiant.

1.3. Utilisabilité ou facilité d'utilisation perçue

1.3.1. L'utilisabilité : un attribut des produits acceptables

Il existe plusieurs modélisations de l'utilisabilité qui la conçoivent comme un attribut mesurable des IHM. Toutes ne la considèrent pas comme relevant des mêmes facteurs ou mesures. Pourtant, l'examen de ces différents modèles fait indubitablement ressortir de nombreux points communs. Nous nous rangeons à l'avis de Folmer et Bosch (2003) qui estiment que ces approches, bien que différentes, renvoient en fait à la même idée. En témoigne ainsi l'utilisation généralisée d'attributs d'utilisabilité se recoupant largement d'un auteur à l'autre. Notre intérêt pour le concept d'utilisabilité se situe plutôt dans la possibilité de l'opérationnaliser dans une démarche d'évaluation. Les trois modélisations théoriques de l'utilisabilité que nous exposons dans leur ordre chronologique de publication n'envisagent l'utilisabilité qu'en relation avec l'acceptabilité et l'utilité de l'interface. Cela rejoint les propos visionnaires de Ken Eason qui déclarait que « le meilleur indicateur de l'utilisabilité est l'utilisation effective du produit » (1984, notre traduction). Tricot et ses collaborateurs en donnent les définitions suivantes (Tricot et al., 2003; Tricot & Tricot, 2000) :

- **Utilité** : C'est la concordance entre ce qu'un dispositif permet de faire et ce que veut faire son utilisateur.
- **Utilisabilité** : C'est la possibilité d'utilisation d'un dispositif, sa maniabilité.
- **Acceptabilité** : C'est l'intention d'usage suscitée par un dispositif.

Si ces définitions sont indubitablement consensuelles, c'est leur articulation qui diffère largement d'un modèle à l'autre.

1.3.1.1. L'utilisabilité selon Shackel (1991)

Précurseur en la matière (Shackel, 1959, cité par Shackel, 2000), il envisagea très tôt l'importance de l'utilisabilité des systèmes d'information, et son travail a fourni à d'autres

matière à réflexions et modifications (Dillon, 2001a; Folmer & Bosch, 2003). Considérant que la recherche en informatique envisageait « l'utilisateur final » sans se préoccuper de ses opinions ni de ses capacités, il est à l'origine de trois grandes idées :

- l'utilisabilité doit être considérée dans son contexte d'utilisation, de sorte qu'il soit possible d'évaluer l'acceptabilité de la machine, attribut ultime de la caractérisation de l'interaction.

- l'utilisabilité, concept faisant l'objet de définitions antérieures satisfaisantes, n'est pas mesurable en soi (concept multidimensionnel).

- l'utilisabilité est une dimension variable influencée par le contexte d'usage.

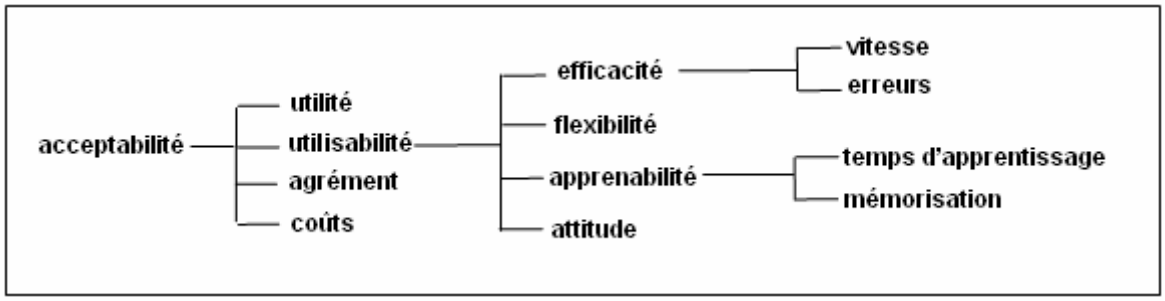


Figure 2: Modélisation des attributs de l'acceptabilité des systèmes selon Shackel (1991, notre traduction)

Il envisage l'acceptabilité d'une IHM comme étant l'alternative résultant de la mise en balance des facteurs d'utilisabilité, d'utilité, d'agrément, et ce, par rapport aux coûts. L'utilité renvoie à l'adéquation entre les besoins de l'utilisateur et les fonctionnalités de la machine, tandis que l'utilisabilité relève de la capacité à utiliser ces fonctions dans la pratique. L'agrément dépend, lui, d'un jugement affectif. De l'autre côté de la balance, on trouve les coûts tout à la fois économiques (achat et entretien du système), sociaux (ex : formation) et organisationnels (ex : perte d'expertise). Ce travail de Shackel a parfois fait l'objet d'une lecture approximative, et nous contestons le raccourci fait par certains qui substituent au concept d'acceptabilité (*acceptability*) celui d'acceptation (*acceptance*) (Folmer & Bosch, 2003; Keinonen, 1998). L'acceptabilité est un attribut des IHM au titre duquel elles peuvent engendrer l'acceptation, c'est-à-dire un comportement d'utilisation effectif.

Shackel note que l'évolution du marché de l'informatique allant vers des puissances de traitement de plus en plus élevées pour des tarifs toujours plus accessibles, l'utilisabilité tend

à devenir la dimension centrale de cette « équation d'acceptabilité ». Mais ceci s'inscrit dans un environnement particulier, impliquant l'outil, les tâches à accomplir et l'utilisateur, tout autant que les aides disponibles (entraînement, didacticiel, assistance). A cet égard, l'utilisabilité est autant le fait de la conception de la machine que de facteurs circonstanciels au cours desquels l'interaction aura lieu. Impliquant utilisateur, environnement, tâche et équipement, l'utilisabilité n'est pas une propriété constante transposable à toutes situations : elle dépend du contexte. L'utilisabilité, selon l'auteur, pourra autant être jugée sur des mesures subjectives de facilité d'utilisation que par des mesures objectives de performance et d'efficacité. L'évaluation pourra se fonder sur les critères suivants :

- **Efficacité**, en termes de vitesse de réalisation et d'erreurs produites,
- **Apprenabilité**, par la mise en relation de la performance avec la durée de pratique et la fréquence d'utilisation,
- **Flexibilité**, impliquant l'adaptation aux différents contextes possibles,
- **Attitudes**, en tant que niveaux humainement acceptables de fatigue, inconfort, frustration et efforts.

Par conséquent, l'utilisabilité est définie, pour un équipement, comme « la possibilité en termes de capacités humaines, d'être utilisé facilement, efficacement par un échantillon spécifique d'utilisateurs, dans un environnement donné, en fonction d'un entraînement particulier et des aides disponibles, et d'accomplir un certain nombre de tâches pour un certain nombre de scénarii » (Shackel, 1991, p.24, notre traduction). En résumé, c'est la possibilité pour un dispositif d'être utilisé facilement (évaluation subjective) et efficacement (évaluation objective).

1.3.1.2. L'utilisabilité selon Nielsen (1993)

Figure incontournable de la conception centrée-utilisateur, Jakob Nielsen[2] considère également l'utilisabilité comme un facteur déterminant l'acceptabilité d'un système. Cette dernière est la combinaison de son acceptabilité sociale et de son acceptabilité pratique. L'acceptabilité sociale peut renvoyer, par exemple, à des considérations éthiques, le jugement des pairs ou à des normes subjectives. Elle est de nature à contrebalancer l'impact de

[2] Le très célèbre *New York Times* l'a consacré « *Guru of Web page usability* »…

l'acceptabilité pratique, quand bien même cette dernière serait élevée. L'acceptabilité pratique se divise en plusieurs catégories, qui renvoient autant à des aspects matériels tels que le coût, la compatibilité avec les systèmes existant, la fiabilité, *etc.*, qu'à l'utilité pratique. Par utilité pratique, on entend la possibilité offerte par une IHM de mener à bien un but fixé. Nielsen l'envisage comme le fruit de l'utilité théorique du système (c'est-à-dire de fonctionnalités et de contenu) ainsi que de son utilisabilité (la façon dont les personnes vont pouvoir se servir de ses fonctionnalités).

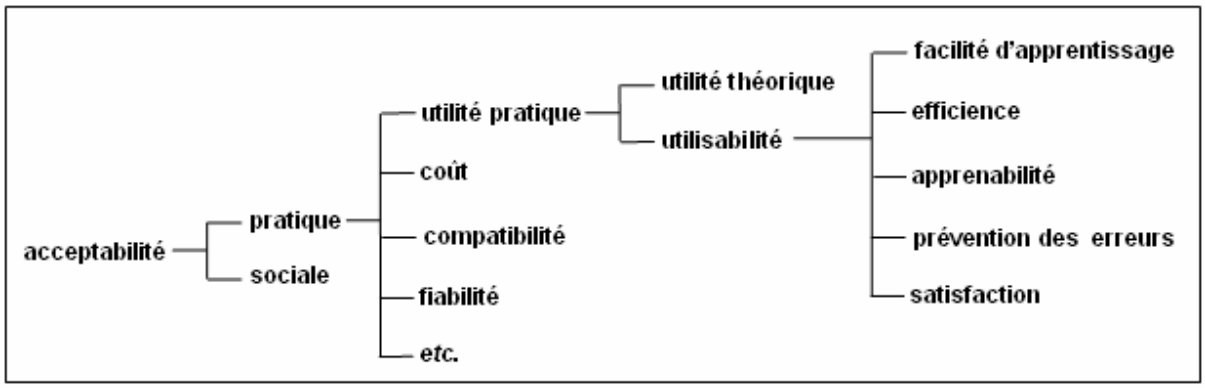


Figure 3: Modélisation des attributs de l'acceptabilité des systèmes selon Nielsen (1993, notre traduction)

Le modèle de Nielsen, ainsi représenté, serait donc à lire de la droite vers la gauche, comme cela a été remarqué (Tricot et al., 2003). L'utilisabilité y est à prendre en compte à tous les niveaux de l'interaction, aussi bien dans l'accomplissement d'une tâche donnée que pour des opérations d'installation ou de maintenance. Contrairement à Shackel (1991), Nielsen ne donne pas de définition précise de l'utilisabilité, arguant que l'on ne saurait la réduire à un concept unidimensionnel. Il l'associe à cinq dimensions dont il donne une définition, ce qui revient en fait à la définir opérationnellement (il rejoint Shackel sur ce point). Keinonen (1998) reprécise utilement certaines d'entre elles :

- **Apprenabilité** : C'est la possibilité pour l'utilisateur novice d'acquérir rapidement un niveau de performance raisonnable. Critère essentiel selon Nielsen, puisque tout dispositif nécessite un apprentissage préalable à l'obtention d'un niveau d'efficacité satisfaisant. On peut la mesurer par l'amélioration de la performance au fil du temps.
- **Prévention des erreurs** : Il s'agit du nombre d'erreurs que fait l'utilisateur. Dans la pratique, une distinction qualitative est introduite, suivant que les erreurs puissent

être facilement corrigées ou prévenues et que celles-ci ne soient pas trop importantes, par destruction ou perte du travail en cours.

- **Efficience** : C'est l'idée de rendement, une fois le maniement du système acquis. Autrement dit, la possibilité de faire beaucoup au moindre effort.
- **Mémorisation** : Capacité du produit à laisser une trace mnésique chez l'utilisateur, après une période d'inutilisation
- **Satisfaction** : C'est le plaisir que l'utilisation du dispositif suscite chez celui qui le manipule.

Ces cinq dimensions de l'utilisabilité ont fait l'objet d'une déclinaison sous forme de questionnaire en cinq items, et dont le but est de mesurer l'utilisabilité perçue par les utilisateurs. Les testeurs y répondent sur des échelles de type Lykert.

Pour terminer, Nielsen propose un descriptif pour caractériser les utilisateurs dont l'échantillonnage est essentiel.

1.3.1.3. L'utilisabilité selon la norme ISO 9241-11 (1998)

La norme ISO 9241 définit le standard international d'« exigences ergonomiques pour le travail de bureau avec terminaux à écrans de visualisation ». Si toutes les parties de cette norme n'ont pas fait l'objet d'une adhésion unanime de la part de la communauté de recherche sur les IHM, la partie 11 « lignes directrices relatives à l'utilisabilité », en dépit de son orientation sur les terminaux à écrans pour le travail de bureau, est applicable à de très nombreuses situations d'interaction homme-machine (Keinonen, 1998). Nous considérons donc qu'elle est à même d'éclairer certains problèmes d'utilisabilité relevant de l'usage des PDA, même s'ils n'ont pas vocation à ne servir qu'au bureau. La norme ISO 9241-11 s'assigne pour fonction de permettre la conception et l'évaluation de l'utilisabilité de produits, de sorte que leurs utilisateurs puissent accomplir les objectifs qu'ils se fixent, et ce, dans un contexte donné. L'utilisabilité y est définie comme « le degré auquel un produit peut être utilisé par des utilisateurs donnés pour accomplir des buts précis avec efficacité, efficience et satisfaction dans un contexte d'utilisation particulier » (notre traduction). Ce standard de recommandation est vraisemblablement l'outil de caractérisation et d'évaluation de l'utilisabilité le plus abouti : outre les définitions rigoureuses de chaque concept évoqué, la

norme ISO 9241-11 fournit également une méthodologie très précise d'évaluation, de la spécification des tâches, des utilisateurs, des buts, de l'environnement ou des équipements, jusqu'aux mesures à réaliser.

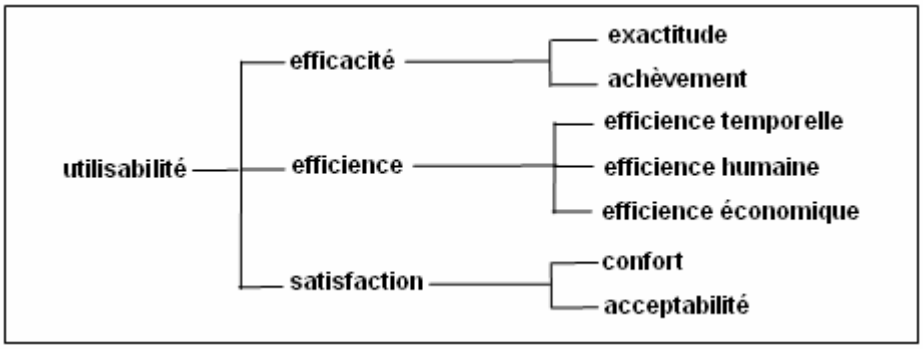


Figure 4: Dimensions de l'utilisabilité selon la norme ISO 9241-11 (1998, notre traduction)

Par ailleurs, toute assertion y est brillamment illustrée d'exemples concis et précis. Comme pour les modèles de Shackel (1991) et Nielsen (1993), l'utilisabilité ne s'y appréhende pas en soi, mais par la mesure de dimensions, ici au nombre de trois :

- **Efficacité** : C'est l'exactitude et le degré d'achèvement avec lesquels les utilisateurs vont accomplir les buts et sous-buts fixés. Elle mesure l'utilisabilité sous l'angle du produit de l'interaction. L'exactitude renvoie à la qualité du résultat (*e.g.* nombre d'erreurs), tandis que l'achèvement relève de la quantité produite en fonction des objectifs. Elle peut aussi être mesurée en pourcentage de buts accomplis ou pourcentage d'utilisateurs ayant accompli avec succès la tâche qui leur était assignée.
- **Efficience** : Il s'agit des ressources engagées en regard de l'exactitude et du degré d'achèvement avec lesquels les utilisateurs accomplissent leurs tâches. Ces ressources se traduisent aussi bien en tant qu'effort humain, temps passé, et coûts engendrés. L'idée d'effort humain requiert une attention particulière. Il peut être à la fois physique, et mental. Il est tout à fait envisageable d'en faire des mesures distinctes, subjectives et objectives (physiologiques, par exemple). Plutôt que d'approfondir cette voie, la norme préconise de s'intéresser à la charge de travail que différents outils, parmi lesquels le NASA-TLX, mesurent avec précision (Hart &

Staveland, 1988; Rubio, Diaz, Martin, & Puente, 2004). L'efficience y est à la fois temporelle, humaine, et économique. Chacune de ces mesures est le quotient d'une mesure d'efficacité divisée respectivement par le temps passé, la charge de travail ou les coûts additionnés de ce travail (formation, matériel, salaires…). Bevan et MacLeod (1994) proposent une relecture des spécifications de la norme pour la mesure de l'efficience en termes mathématiques. Ils n'apportent rien sur le plan théorique, mais formalisent les préconisations de la norme :

Efficience temporelle = efficacité ÷ temps d'accomplissement

Efficience humaine = efficacité ÷ effort

Efficience économique = efficacité ÷ coût total

Formule 1: Mesures d'efficiences selon la norme ISO 9241-11 interprétées par Bevan et MacLeod (1994)

Ils font de même avec la proposition d'unification de l'efficacité dans une seule mesure par le produit de l'achèvement et l'exactitude, ce qui augmente la portée opérationnelle des recommandations de la norme :

Efficacité (%) = 1/100 × (exactitude × achèvement)

Formule 2: Mesure d'efficacité selon la norme ISO 9241-11 interprétée par Bevan et MacLeod (1994)

- **Satisfaction** : C'est le confort et les attitudes relatives à l'usage. La norme 9241-11 reste assez évasive quant aux mesures à effectuer. Pour la satisfaction, elle envisage des échelles d'évaluation subjective associées à des techniques d'enquêtes et d'observations sur le terrain pour attester de l'usage effectif. Est également évoquée la possibilité de relever des commentaires pendant l'interaction, les postures et comportements, ou bien encore d'évaluer l'acceptabilité de la charge de travail induite.

La norme ISO 9241-11 ne précise pas quelle contribution chacune des dimensions (efficacité, efficience et satisfaction) va avoir sur l'utilisabilité globale, ceci pouvant dépendre à la fois de la mesure et du contexte. D'ailleurs, il n'y a, selon elle, pas d'utilisabilité

intrinsèque au produit, seulement une possibilité d'être utilisée pour un contexte donné. Elle attire aussi l'attention sur son incapacité à déterminer, pour un problème d'utilisabilité mis au jour, la cause de celui-ci. Ce n'est qu'au prix d'un examen minutieux de l'usage et le renvoi aux autres sections de la norme ISO 9241, notamment la 10, relative aux principes de dialogues, que l'origine de la complication émergera.

D'autre part, si l'utilisabilité peut être calculée de façon globale, il n'est pas exclu que toutes les dimensions d'utilisabilité du dispositif ne soient pas mises à profit dans le test. On peut y pallier, si les besoins de développement y incitent, en estimant l'utilisabilité sur une propriété particulière (annexe B). La liste fournie par la norme ne se veut pas exhaustive. On y trouve :

- Evaluation du besoin d'entraînement,
- Evaluation de la prise en main,
- Evaluation de la mémorisation (après une période de non utilisation ou à une fréquence donnée),
- Evaluation des besoins en aide,
- Evaluation de l'apprenabilité (contrairement à Nielsen (1993), il s'agit du nombre de fonctions retenues),
- Evaluation de la tolérance aux erreurs (pourcentage d'erreurs corrigées ou tolérées),
- Evaluation de la lisibilité.

Là encore, il sera possible de se référer à la section 10, et plus spécialement aux sections relatives à l'apprenabilité et à l'individualisation des interfaces.

1.3.1.4. Critique des modélisations de l'utilisabilité

Shackel (1991) et Nielsen (1993) considèrent l'utilisabilité comme un facteur de l'acceptabilité du dispositif par l'utilisateur (et consommateur, s'il on prend en compte la dimension « coût »). A l'inverse, la norme ISO 9241-11 développe une conception de l'utilisabilité centrée sur la qualité du travail produit par l'utilisateur, l'acceptabilité y contribuant par la satisfaction. Il existe encore une autre approche, centrée elle, sur le produit. En effet, dans son annexe D, la norme ISO 9241-11 précise ses relations avec les autres

standards internationaux. Parmi eux, la norme ISO 9126-1 (2000) relative à la qualité des logiciels et que d'aucuns font une référence en matière d'utilisabilité (Bevan, 2001; Bevan & Macleod, 1994; Folmer & Bosch, 2003). Effectivement, le terme utilisabilité est souvent employé comme étant la possibilité pour un produit d'être utilisé facilement. Selon la norme ISO 9126-1, l'utilisabilité a une contribution relative (parmi d'autres dimensions) à la qualité des logiciels dans leur conception et leur évaluation. L'utilisabilité y est définie comme « un ensemble d'attributs du logiciel, qui découle de l'effort imposé par son utilisation, ainsi que de l'évaluation personnelle d'une telle utilisation par un ensemble donné d'utilisateurs » (notre traduction). Par contre, elle définit la qualité d'utilisation (*quality in use*) comme étant « le degré auquel un produit peut être utilisé par des utilisateurs donnés pour accomplir des buts précis avec efficacité, productivité, sûreté et satisfaction dans un contexte d'utilisation particulier » (notre traduction). Cette définition est donc semblable en tous points à celle que donne la norme 9241-11 de l'utilisabilité, à cette exception près que l'idée de sûreté y a été ajoutée, et que la productivité a remplacé l'efficience. L'assimilation de la qualité générale d'un système à son utilisabilité a été repérée par Tricot et Tricot (2000) comme vecteur de confusion. Dès lors, comment appréhender cette nouvelle approche ? Si l'on examine la norme ISO 9126-1, on constate que la qualité d'utilisation est le fait des influences conjointes de l'utilisabilité, des fonctionnalités, de la fiabilité, de l'efficience, de la portabilité et de la facilité d'entretien du logiciel. Cela correspond en fait à l'idée d'acceptabilité pratique chez Nieslen (1993), qu'il évalue en fonction de l'utilité pratique (utilisabilité + utilité théorique ou fonctionnalités), de la fiabilité, la compatibilité, les coûts, etc.… On rejoint ainsi l'idée d'Eason (1984) selon laquelle « le meilleur indicateur de l'utilisabilité est l'utilisation effective du produit » dans une approche sociotechnique (Dillon & Watson, 1996). Cependant, cette assertion peut s'avérer fausse, et plusieurs exemples en témoignent. La boîte de vitesse automatique, en France, n'est que peu utilisée. Pourtant, les assureurs savent que la conduite de ce type de véhicules occasionne moins d'accidents et moins d'efforts que ceux équipés de boîtes manuelles. Autre exemple allant dans le même sens, la très large adoption des claviers AZERTY dont la disposition datant du XIXème siècle ne visait, pour les anciennes machines à écrire, que la séparation des bigrammes les plus fréquents afin d'éviter que ne se coincent les tiges supportant les fontes de touches avoisinantes. Leur organisation donc, si elle ne doit rien au hasard, n'est pas le fait d'un choix ergonomique. Il existe depuis des claviers dont l'ergonomie a été pensée pour faciliter la frappe de documents à une époque où se sont

perfectionnées les machines à écrire (ex : claviers Dvorak[3]). Or, ces claviers ne se sont que peu diffusés dans le monde de la bureautique et de l'informatique. Nous n'en connaissons pas les raisons. Simplement, nous notons que l'utilisabilité ne conduit pas nécessairement à l'acceptabilité. Ou bien alors, c'est que l'utilisabilité a une définition différente de celle que nous lui accordons : Bevan considère, lui, que l'utilisabilité est la qualité d'usage (1995). La citation d'Eason pourrait alors être remise au goût du jour en disant que le meilleur indicateur de la qualité d'un logiciel est son utilisation effective. Cette confusion entre utilisabilité et qualité des IHM pourrait ainsi trouver son explication dans le point de vue de Shackel (1991) selon lequel, par une diminution des coûts d'acquisition et d'entretien[4] du matériel informatique et une augmentation régulière des puissances de traitement[5], l'utilisabilité devient le facteur principal de l'acceptabilité. Il nous semble cependant que ces exemples soulignent d'autant mieux toute la puissance du modèle de Nielsen (1993) qui prend en considération à la fois l'acceptabilité pratique (qualité d'utilisation) et l'acceptabilité sociale. C'est ce qui justifie, selon nous, le succès mérité de ce dernier. Bevan (2006) ne s'y est d'ailleurs pas trompé quand il déclare que les approches des normes ISO 9126-1 et 9241-11 sont complémentaires : elles sont résumées, à notre sens, en une seule articulation dans le modèle de Nielsen.

Pour autant, il n'est pas exempt de manques, notamment en ce qui concerne la place accordée à l'utilité, à l'efficacité et à la description des erreurs. Keinonen (1998) a traité avec intelligence ces questions. S'agissant de l'utilité et de sa position dans les modèles de Nielsen ou de Shackel, que ce dernier subordonne à l'utilisabilité. Il fait remarquer que l'idée de départ repose sur le fait de mesurer l'utilisabilité d'une IHM par le biais de la qualité du travail qu'elle a permis de réaliser. Cela implique deux conditions : d'une part, que la

[3] Le clavier Dvorak a été conçu en 1936 par August Dvorak et William Dealey. L'idée est de mettre les lettres les plus utilisées sur la rangée principale du clavier, ce qui dépend de la langue considérée. On réduit ainsi le besoin de déplacer les doigts vers les autres rangées, ce qui augmente le confort et la vitesse de frappe, tout en diminuant le risque de développer des lésions des tissus mous comme les muscles, les tendons ou les ligaments.

[4] Bevan et MacLeod (1994) font remarquer que si l'idée d'efficience économique est juste sur le plan théorique, elle ne peut, dans la pratique, s'appliquer à l'utilisateur *lambda*. En effet, le prix d'achat du matériel fait souvent figure de variable couperet. L'idée d'efficience économique renvoie alors à la notion d'amortissement, plus applicable au monde du travail.

[5] La loi de Moore (1965) prédit que le nombre de transistors sur un microprocesseur double tous les deux ans. En conséquence, les machines électroniques sont devenues de moins en moins coûteuses et de plus en plus puissantes. Cette considération n'est pas anodine pour les PDA dont la petitesse a des répercussions sur la puissance de calcul (Bouteruche, Deconde, Anquetil, & Jamet, 2005), et dont l'amélioration future sera sans doute un facteur clé de l'acceptabilité des services qu'ils proposent.

communication entre l'humain et la machine soit suffisamment efficace, d'autre part, que les fonctionnalités de la machine le permettent. La première condition renvoie clairement à l'idée d'utilisabilité, notamment en termes de prévention d'erreurs dites « catastrophiques », qui ruineraient tous les efforts engagés (disparition complète du travail en cours). Pour la seconde, c'est plus compliqué : exclure l'utilité de l'utilisabilité n'a de sens que si le dispositif évalué a été conçu explicitement pour accomplir la tâche prescrite pour son évaluation. L'exemple fournit par Keinonen est très parlant : si l'on évalue l'édition de texte sur téléphone portable en excluant son utilité (à savoir, qu'il n'est pas fait pour ça), on constatera que l'édition de texte y est difficile, ou bien que le test n'est pas pertinent. Néanmoins, il peut arriver que l'on ait besoin de saisir du texte sur un téléphone portable. Dès lors, en incluant ceci en tant qu'utilité potentielle, on se rend compte de l'utilisabilité de la saisie de texte sur ce type de machine. C'est pourquoi Keinonen plaide pour l'introduction de l'idée de « flexibilité » (ou adaptabilité pour Bastien et Scapin (2001)) dans les modèles d'utilisabilité afin de définir de façon adéquate les tâches et les conditions de sa réalisation durant les tests. En ce qui concerne les erreurs, il souligne le manque de précision des modèles dans leur description : une erreur peut être grave (catastrophique) ou bénigne. Cela n'a évidemment pas la même valeur. En outre, il peut y avoir des effets pervers et antagonistes. Améliorer l'utilisabilité passe pour beaucoup par la réduction de la charge de travail et de l'effort humain. Il est à noter que si une trop forte charge de travail n'est évidemment pas souhaitable, il en va de même pour une charge trop faible, contrairement aux idées reçues (Delinières & Deschamps, 2000; Yerkes & Dodson, 1908). En effet, une faible charge de travail peut se traduire par un désengagement attentionnel, conduisant à des erreurs « d'inattention ». L'évaluateur doit en tenir compte. Autant de raisons qui font que si le modèle de Nielsen (1993) nous semble le plus intéressant sur le plan théorique, celui de la norme ISO 9241-11 est sans aucun doute le seul véritablement opérationnalisable.

Au final, qu'est-ce que l'utilisabilité ? Nous l'avons déjà précisé, il est difficile de trancher cette question. Une nouvelle définition ne serait qu'une définition de plus. Contrairement à Keinonen (1998) qui propose un modèle unifiant les trois précédents, notre choix se portera sur le modèle de Nielsen (1993). En effet, il nous semble que l'utilisabilité n'a d'intérêt que dans sa mise en relation avec l'utilité et l'acceptabilité puisqu'il n'est pas possible de faire des mesures absolues d'utilisabilité. A titre d'exemple, il est impossible de comparer l'utilisabilité d'une console de jeux et celle d'un panneau de contrôle aérien. C'est

toute l'importance de la prise en considération du contexte qui se joue ici, avec, à la clé, l'intention d'utilisation effective par l'utilisateur. Le modèle de Nielsen rend compte le mieux de cela en distinguant acceptabilité sociale et pratique. On peut assimiler cette dernière à la définition de la qualité logicielle de la norme ISO 9126-1. Si cette norme est assez généraliste, le modèle de Nieslen accorde beaucoup plus d'importance à l'utilisabilité, comme le fait Shackel. Il y prend en considération l'apprenabilité (facilité et durée), ce que ne fait pas directement la norme 9241-11. Néanmoins, et en fonction des besoins d'évaluation d'une interface, on pourra enrichir le modèle de Nielsen. Par exemple, on pourra tenir compte de mesures d'efficacité (comme le préconisent Shackel (1991) et la norme ISO 9241-11) si l'on veut tester en particulier la flexibilité d'un dispositif. De la même façon, on s'inspirera au besoin des mesures d'efficience proposées par Bevan et MacLeod (1994) pour compléter celle de Nielsen (1993).

Le modèle de Nielsen comporte d'autres lacunes. Si les trois modèles insistent sur l'importance de définir le contexte d'utilisation (environnement, tâche et utilisateurs), aucun n'inclue ces mesures. Elles pourraient pourtant jouer un rôle de variables modératrices (âge, volonté d'utilisation, expérience…) dans ces modèles. Depuis plusieurs années déjà, certains auteurs ont alerté sur la nécessaire prise en compte des différences intra et inter individuelles dans l'étude des IHM (Dillon & Watson, 1996). L'absence de ces dimensions doit être, dans la pratique, contrôlée par un échantillonnage rigoureux, que l'on sait être difficile. On peut penser que ces questions sont implicitement traitées par l'idée d'acceptabilité sociale, renvoyant à certaines normes subjectives propres à chacun. Cela renvoie aussi à l'idée d'accessibilité, c'est-à-dire, d'utilisabilité pour tous, et ce d'autant plus que nos sociétés évoluent vers le pluralisme culturel et la prise en considération de chacun (Marcus, 2003).

Par ailleurs, ces modèles ne mettent pas en relation les critères de qualité ergonomique avec les mesures d'utilisabilité correspondantes. Cela n'a pas échappé à Folmer et son équipe (Folmer, Van Gurp, & Bosch, 2003) qui proposent un cadre conceptuel pour faire le lien entre attributs d'utilisabilité (Nielsen, 1993; norme ISO 9126; Shackel, 1991) et qualité ergonomique (Bastien & Scapin, 2001; Nielsen, 1994). L'idée est très intéressante. Cependant, le modèle qu'ils présentent n'est que le fruit d'une méta-analyse de la littérature pour sélectionner les critères et attributs les plus représentatifs. Il lui manque encore une validation expérimentale.

Pour finir, si ces modèles préconisent de mesurer l'acceptabilité et la satisfaction, ils ne proposent pas beaucoup de solutions pour. En particulier, ils ne précisent pas dans quelle mesure l'intention d'utilisation (attitude) est prédictive de l'acceptation (comportement), ni en quoi la satisfaction va être liée aux mesures réalisées (Hornbæk, 2006). Nielsen et Levy (1994) relèvent cependant dans une méta-analyse qu'il existe une corrélation positive ($r = .44$) entre des mesures objectives de performance et des échelles de satisfaction. S'il y a bien un lien, il n'est pas pour autant très fort, et l'on ne saurait extrapoler une corrélation vers une relation de causalité sans autre précaution. D'autres études se montrent quant à elles beaucoup plus prudentes sur les relations entre performance et satisfaction, arguant que pour l'heure, on ne mesure pas encore la complexité des rapports entre les différentes mesures d'utilisabilité (efficacité, efficience et satisfaction). C'est pourquoi elles préconisent de tenir compte de chacune, sans présager de corrélations élevées entre elles (Frokjaer, Hertzum, & Hornbaek, 2000). A l'inverse, certains évoquent même une dissociation entre mesures subjectives de charge mentale et performance, pouvant conduire à des erreurs d'appréciation (et donc, d'acceptation) de certaines interfaces (Yeh & Wickens, 1984).

1.3.1.5. Evaluation de l'utilisabilité de la saisie de texte manuscrit sur PDA

C'est à Scott MacKenzie et ses collaborateurs que l'on doit la plus importante contribution à l'étude de l'utilisabilité des méthodes de saisie pour PDA et Smartphone*s*. Ils ont consacré plusieures de leurs études aux interfaces de saisie d'écriture manuscrite. Les tests auxquels ils procèdent sont très contrôlés et très axés sur la mesure précise des performances. Ils ont d'ailleurs développé un ensemble de calculs des taux d'erreurs pour évaluer la saisie de texte sur périphériques mobiles, qui hélas, sont à réserver à l'étude de la saisie sur clavier (Soukoreff & MacKenzie, 2004). S'ils prennent aussi en considération la satisfaction et l'acceptabilité, c'est de façon moins élaborée. Ces travaux sont plus représentatifs des mesures préconisées par les modèles de Shackle (1991), Nielsen (1993) et la norme ISO 9241-11, que de leur articulation entre elles. C'est dû à l'orientation ergonomique de ces travaux qui ne sont pas des travaux de psychologie.

MacKenzie et son équipe évaluent en 1997 l'utilisabilité immédiate de *Graffiti* sur un *Tablet* PC. Cette interface de saisie manuscrite reconnaît des lettres assez élaborées. Pour ce

faire ils soumettent 25 sujets novices au protocole suivant : dans une première phase, les participants étudient une minute un didacticiel de *Graffiti* et écrivent ensuite l'alphabet. Tout de suite après, on les laisse découvrir seuls *Graffiti* pendant cinq minutes, à la suite de quoi ils réécrivent l'alphabet (deuxième phase). La troisième phase se déroule une semaine après, et consiste à écrire l'alphabet sans aucun entraînement depuis la phase 2.

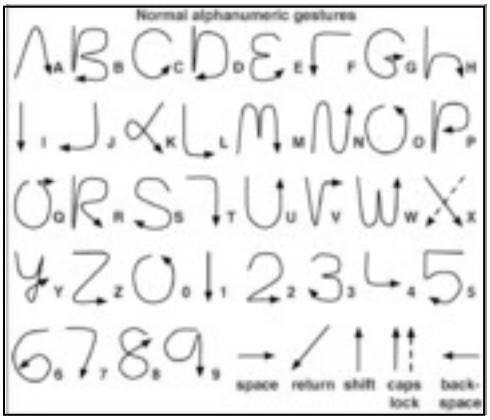


Figure 5: Alphabet *Graffiti*

Ils mesurent le taux de reconnaissance moyen des lettres (correction faite de leur fréquence d'apparition dans la langue anglaise). Ils constatent qu'à la phase 1, ce taux de reconnaissance est de 85,5%. Après un entraînement de cinq minutes, ce taux passe à 96,9%. Une semaine après, il monte à 97,2%. On se rend compte avec ces chiffres que l'apprenabilité de *Graffiti* est rapide par la pratique, après exposition à un didacticiel. Elle se maintient également dans la durée. Si les auteurs étudient avec attention les taux de reconnaissance lettre par lettre, pour connaître celles qui posent problème, ils ne s'intéressent pas à la satisfaction des participants ni à l'acceptabilité du dispositif. Cette étude n'est donc que descriptive de l'apprenabilité de *Graffiti* (MacKenzie & Zhang, 1997).

En 1999, MacKenzie et Chang proposent une expérience comparant deux interfaces de reconnaissance d'écriture : *Microsoft Pen for Windows* et *CIC's Handwriter 3.3*. Cette expérience se déroule sur une tablette graphique (*Wacom*) et non sur un PDA. Chaque participant recopie sur la tablette un ensemble de phrases présenté à l'écran selon quatre conditions :

- *Microsoft Pen for Windows*, lettres minuscules,

- *Microsoft Pen for Windows*, lettres minuscules et majuscules,
- *CIC's Handwriter 3.3*, lettres minuscules,
- *CIC's Handwriter 3.3*, lettres minuscules et majuscules.

En ce qui concerne l'efficacité, ils constatent que les deux méthodes de saisie ne diffèrent pas pour la vitesse d'écriture (nombre de mots par minute). Par contre, écrire en minuscules et en majuscules avec *Microsoft Pen for Windows* se traduit par un taux de reconnaissance des lettres plus bas que pour les autres conditions (précision). Ils évaluent l'apprenabilité, et montrent que si la vitesse de saisie s'améliore avec le temps, ce n'est pas le cas de la précision, dont le taux de reconnaissance est stable (puisque limité par le logiciel lui-même, selon eux). En fait, ils ne testent pas tant l'ergonomie de l'interface que le moteur de la reconnaissance lui-même, puisque l'interface est identique pour les deux logiciels.

Figure 6: Capture d'écran typique de l'expérience de MacKenzie et Chang (1999)

La mesure de la satisfaction donne lieu à une préférence accordée au *CIC's Handwriter 3.3.*, le taux de reconnaissance des lettres par celui-ci y étant pour beaucoup. L'intention d'utilisation relevée par les auteurs est fonction de l'utilité : les sujets préfèreraient les méthodes manuscrites au clavier pour remplir des tâches simples, comme par exemple remplir un formulaire. Pour des créations de texte plus élaborées, ils préfèreraient le clavier (matériel). Cette étude met donc en relation utilisabilité, utilité et acceptabilité, et c'est là qu'elle est la plus instructive. Par contre, elle ne compare pas ces méthodes d'écriture manuscrite avec d'autres alternatives (comme le clavier virtuel, par exemple). On peut faire ce même reproche à Sears et Arora (2002) quand ils comparent *JOT* à *Graffiti*, pour montrer que *JOT*, plus rapide, est également mieux apprécié.

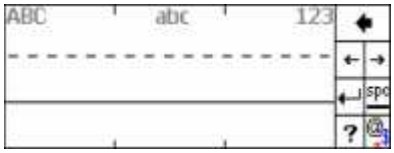


Figure 7: Interface *JOT*

Bien qu'ils prennent la précaution d'utiliser ces interfaces dans des tâches réalistes, ils n'expliquent pas le problème d'acceptabilité des systèmes de reconnaissance de l'écriture (McQueen, MacKenzie, Nonnecke, Riddersma, & Meltz, 1994).

C'est précisément ce que font les deux études suivantes. La première compare un logiciel de reconnaissance de l'écriture manuscrite, un clavier virtuel et deux autres méthodes dans une tâche de saisie de chiffres (MacKenzie, Nonnecke, McQueen, Riddersma, & Meltz, 1994). Chaque participant teste toutes les interfaces. Les auteurs montrent la supériorité du clavier virtuel, aussi bien pour le taux d'erreurs (1,2%) que la vitesse (30,4 chiffres par minute) sur la reconnaissance de l'écriture (10,4% et 18,5 chiffres par minute). La satisfaction des utilisateurs est évaluée sur une échelle en 5 points. Ils plébiscitent le clavier virtuel dont ils apprécient la performance, mais s'ils devaient choisir, préféreraient continuer à écrire « à la main » sur leur PDA, bien que cela soit moins rapide. Ce résultat est sans doute à remettre dans le contexte de nouveauté de ces interfaces d'écriture, dont MacKenzie et ses collaborateurs soulignent déjà qu'elles sont jugées « inutilisables » par la presse spécialisée.

Ils concluent sur la nécessité d'améliorer les taux d'identification de caractères des interfaces de reconnaissance de l'écriture, et proposent de compléter cette étude en étudiant la saisie de lettres (MacKenzie, Nonnecke, Riddersma, McQueen, & Meltz, 1994). Onze participants novices sont mis à contribution dans cette étude comparant *Microsoft Pen for Windows*, un clavier virtuel QWERTY et un clavier virtuel alphabétique.

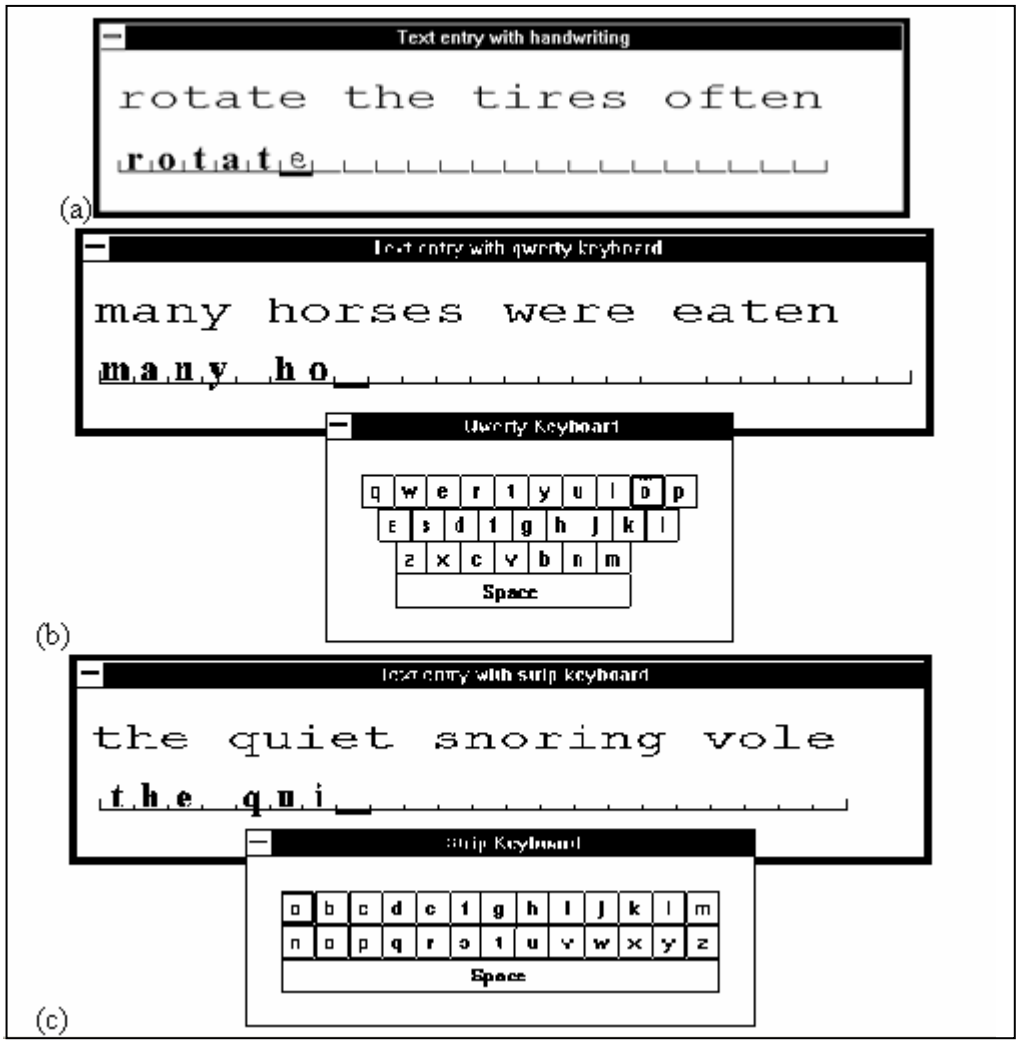


Figure 8: Les trois conditions expérimentales étaient (a) l'écriture manuscrite reconnue, (b) le clavier QWERTY, et (c) le clavier ABC (MacKenzie, Nonnecke, Riddersma, McQueen, & Meltz, 1994)

Ils doivent saisir plusieurs phrases présentées par la machine en utilisant les trois méthodes. En relevant les taux d'erreurs, les auteurs montrent que la reconnaissance d'écriture est bien moins efficace (8% d'erreurs) que les deux claviers (1%). Par contre, elle est un peu plus rapide que le clavier ABC. Au final, les utilisateurs apprécient autant le clavier QWERTY que la reconnaissance d'écriture, devant le clavier alphabétique. Pour les auteurs, la reconnaissance d'écriture est acceptable pour quelques personnes dont les performances sont bonnes.

En conclusion, on retiendra le caractère très contrôlé de ces études, assez loin de toute tentative de reproduction du contexte. Ces expériences se déroulent d'ailleurs sur des tablettes graphiques *Wacom* et non des PDA. La question de la satisfaction et de l'acceptabilité y est

traitée sommairement. Ce sont des approches très centrées sur la mesure de l'utilisabilité. Très descriptives, leur potentiel explicatif est limité. On peut en dire autant pour les travaux de Read (2007), très inspirés par ceux de MacKenzie, mais appliqués aux enfants. Il en ressort que les enfants de 7/8 ans composent des textes plus long avec une tablette graphique qu'avec le clavier matériel QWERTY, alors qu'ils sont novices dans l'emploi des deux dispositifs. Cela est peut-être une piste d'acceptabilité future par découverte précoce de la reconnaissance d'écriture, mais la satisfaction n'a pas été mesurée. Ces recherches ne répondent donc pas non plus à la question de l'acceptabilité.

Ce n'est pas le cas de l'étude de Költringer et Grechenig (2004). Ils comparent le clavier virtuel à *Graffiti* 2 et *Graffiti 2 + complétion de mot* dans une tâche de saisie de phrases (lettres, chiffres, URL, etc.).

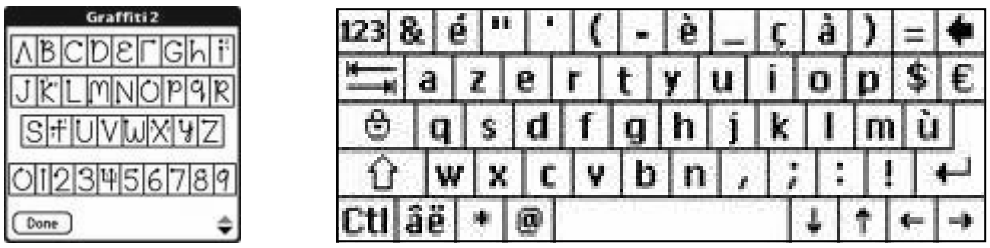


Figure 9: Alphabet *Graffiti* 2 et clavier virtuel

Ils relèvent que le clavier virtuel est la méthode la plus performante (vitesse et précision) par rapport aux deux autres, qui ne diffèrent pas entre elles. L'effet est d'autant plus marqué pour les commandes (entrée, échappement, tabulation) qui se font par geste ou signe sur *Graffiti*. Par contre, le test SUS (Brookes, 1996) ne témoigne pas de différences significatives entre les interfaces. Seul *Graffiti 2 + complétion de mots* engendre une charge de tâche (NASA-TLX) plus basse que pour les deux concurrents. Or, *Graffiti 2 + complétion de mots* recueille la plus grande satisfaction. On peut regretter que ces auteurs n'aient pas procédé à une mesure d'efficience humaine (telle que la recommande la norme ISO 9241-11, cf. formule 1). On aurait peut-être eu un indice supplémentaire si la différence d'efficience entre le clavier et *Graffiti 2 + complétion de mot* pouvait expliquer la différence d'acceptabilité.

Pour finir, évoquons la très ambitieuse étude de Curran et ses collaborateurs (Curran, Woods, & O Riordan, 2006) qui comparent plusieurs méthodes de saisie sur téléphones portables :

- Clavier de bureau QWERTY,
- Mini clavier QWERTY (organiseur de poche),
- Clavier de téléphone portable,
- Clavier de téléphone portable + complétion de mot T9,
- Clavier virtuel sur PDA,
- Saisie et reconnaissance d'écriture manuscrite sur PDA.

Pour conduire leur expérience, ils réunissent un échantillon de 24 participants qu'ils partagent en trois groupes selon leur niveau « d'expertise », le critère de répartition étant basé sur le nombre moyen de SMS qu'ils envoient par semaine. Dans chacun de ces groupes, les sujets sont répartis de façon équilibrée par tranche d'âge (jeunes ou âgés, la valeur seuil étant 30 ans) et genre. Les participants ont à écrire, après un court entraînement, 4 phrases assez « réalistes » (de type rendez-vous, adresse *e-mail*, pense-bête, et une phrase en abréviations courantes). Si cette manipulation est prometteuse, les résultats ne sont pas à la hauteur, et les différences de performances par âge et sexe assez ténues. En ce qui concerne les dispositifs de saisie, la reconnaissance d'écriture engendre les plus mauvaises performances. Cependant, elle est relativement bien jugée par les utilisateurs, juste derrière le clavier de bureau, le clavier de téléphone, et devant le clavier virtuel et le mini clavier.

1.3.2. La facilité d'utilisation : une croyance précurseur de l'acceptation

Ces manipulations, bien qu'instructives en ce qui concerne l'utilisabilité de la reconnaissance d'écriture sur PDA, ne donnent pas l'opportunité d'anticiper l'utilisation future de ce type de logiciel (faut-il encore que les participants soient de potentiels utilisateurs). D'autres modèles traitent cette question de façon très approfondie, au détriment de l'utilisabilité. La mesure de l'utilisabilité, si elle est considérée, n'est plus une VD (variable dépendante) mais une VI (variable indépendante) dont on testera alors l'influence sur l'acceptation (Dillon & Watson, 1996). On parle d'acceptation et non d'acceptabilité puisqu'il ne s'agit plus d'un attribut de l'interface, mais d'un comportement d'utilisation effective ou d'une intention d'usage. Ces modèles, qui se déclinent en outils (questionnaires, enquêtes), peuvent s'utiliser en laboratoire, mais leur élaboration est le fruit d'un travail de

terrain. Ils permettent de prédire, avec une certaine précision, si un dispositif va être adopté ou pas. Il y a donc un glissement de la question « traditionnelle » de l'utilisabilité vers les déterminants de l'acceptation ou intention d'usage (Dillon, 2001b). L'acceptation peut être définie comme étant l'intention d'usage avérée d'un système d'information pour réaliser une tâche pour laquelle il est conçu, et ce, par un groupe d'utilisateurs donné. Par conséquent, l'étude de l'acceptation se souciera moins de l'utilisation occasionnelle ou non-motivée d'un dispositif que des causes qui conduisent à l'adopter de façon voulue et raisonnée. L'idée qui sous-tend le développement des modèles d'acceptation est de prévenir, pendant la conception de systèmes d'informations, les risques de résistance ou de rejet d'une interface par ses utilisateurs potentiels. Il s'agit donc autant de prédire un comportement futur que de l'expliquer (Tornatzky & Klein, 1982).

1.3.2.1. Les premiers modèles

1.3.2.1.1. La Théorie de la Diffusion des Innovations (IDT, Rogers, 1962)

Le principal appui théorique à la notion d'acceptation, selon Dillon et Morris (1997), est sans doute la Théorie de la Diffusion des Innovations (*Innovation Diffusion Theory*, IDT) (1962, cité par Rogers, 1995). Bien que ne se préoccupant pas exclusivement des TIC, elle est à même de rendre compte du processus qui va conduire une invention en cours d'élaboration à sa plus large diffusion. L'approche de l'IDT est sociale, et non pas cognitive. Selon Rogers, cinq caractéristiques d'une innovation vont déterminer sa diffusion : l'avantage relatif (c'est-à-dire le degré d'amélioration fourni par l'innovation par rapports aux outils existants), la compatibilité (en regard des pratiques et des normes de l'utilisateur), la complexité (d'utilisation et d'apprentissage), la possibilité d'essai (c'est-à-dire de tester le dispositif avant de le choisir), et enfin l'observabilité (la visibilité de la qualité des résultats produits par le dispositif). Si ces critères, pris séparément, échouent à rendre compte de la diffusion future d'une interface, leur mise à contribution commune s'avère prédictive (Rogers, 1995). Parmi ces critères, une méta-analyse (Tornatzky & Klein, 1982) a mis en évidence que trois caractéristiques avaient la plus grande influence sur la diffusion : l'avantage relatif, la compatibilité et la simplicité d'usage. Dillon (2001b) remarque que ces critères trouvent écho dans des travaux traitant de la conception d'IHM (Nielsen, 1993), à cette différence près que

ces derniers envisagent l'utilisabilité comme étant le principal déterminant de l'acceptabilité. On retrouve ici, en partie, les déterminants de la qualité des logiciels de la norme ISO 9126-1, bien que Tornatsky et Klein alertent sur le manque de rigueur dans la définition de ces concepts, tout particulièrement celui d'avantage relatif.

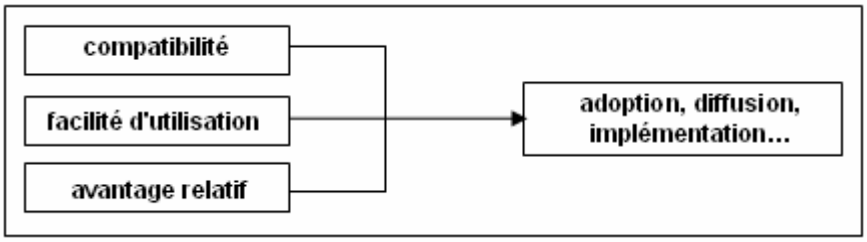


Figure 10: Déterminants principaux de la diffusion des innovations, selon Tornatzky et Klein (1982, notre traduction)

Si l'IDT détermine des facteurs de diffusion relatifs à la machine, elle propose également des critères centrés sur l'utilisateur. Rogers catégorise les individus en fonction de leur rapidité d'adoption selon six groupes : innovateurs, premiers adopteurs, première majorité, dernière majorité, derniers adopteurs, et retardataires. Ce découpage n'a pas de valeur théorique propre, puisque chaque catégorie représente un écart-type à partir de la moyenne d'une courbe représentant la diffusion de l'innovation en fonction du temps, et dont la distribution est gaussienne. A noter l'exception faite de la catégorie « retardataires » qui représente deux écarts-types, non différenciés, puisque moins influente que celle des innovateurs et primo-adopteurs dans le processus d'influence qui conduit à la diffusion. La construction de ces catégories a donné lieu à des recoupements en termes de personnalité ou de situation socio-économique, voire même, d'attitudes (Agarwal, Ahuja, Carter, & Gans, 1998). Cependant, cette approche socio-technique de la question de l'acceptation d'une nouvelle technologie ne fait qu'expliquer l'adoption une fois celle-ci entérinée. Elle ne permet pas de la prédire à un niveau individuel. C'est pourquoi d'autres champs de recherches se sont penchés sur ce problème d'un point de vue psychosocial et cognitif. Ces résultats issus de la recherche sur des modèles cognitivo-comportementaux ont inspiré à certains des développements de l'IDT dans le domaine des technologies de l'information, de sorte qu'à la lumière de ces modèles, les croyances et attitudes des individus innovateurs et « primo-adopteurs » soient mieux connues (Agarwal, Ahuja, Carter, & Gans, 1998).

1.3.2.1.2. La Théorie de l'Action Raisonnée (TRA, Fishbein & Acjek, 1975)

La question de l'acceptation, si elle n'est pas appréhendée en tant que telle par l'IDT, peut être envisagée sous l'angle de la Théorie de l'Action Raisonnée (*Theory of Reasoned Action, TRA*) (Fishbein & Ajzen, 1975). Elle n'a pourtant que peu été mise à profit pour l'explication de l'adoption de nouvelles technologies, n'ayant, à l'origine, pas été conçue pour cela. Ce modèle, issu du champ de la psychologie sociale, propose de prédire les comportements par la mesure des attitudes et de la norme subjective. En effet, elles influencent directement l'intention comportementale (variable intermédiaire entre comportement effectif et représentations) qui indique la volonté de mener à bien une action en fonction d'objectifs fixés. Si bien sûr, seule l'intention comportementale rend compte du comportement futur, l'avantage de cette modélisation est de connaître la contribution relative, entre attitudes et norme subjective, qui va expliquer le choix de l'individu. L'attitude, dans une acception large, consiste en l'opinion affective (c'est-à-dire positive ou négative) d'un individu sur un objet donné. Ici, elle évalue un comportement. L'attitude est elle-même le résultat de croyances quant aux conséquences du comportement envisagé, ainsi que de l'évaluation de ces dites conséquences (bonnes ou mauvaises). La norme subjective renvoie elle, au sentiment de la personne sur les attentes des individus de référence de son groupe d'appartenance (amis, famille ou collègues) (Channouf, Py, & Somat, 1996; Fishbein & Ajzen, 1975). Ces attitudes et croyances se fondent et évoluent sous l'influence de stimuli extérieurs. Ce modèle, bien qu'évoquant une action « raisonnée », admet la possibilité que l'intention comportementale et sa réalisation éventuelle se fassent de façon implicite.

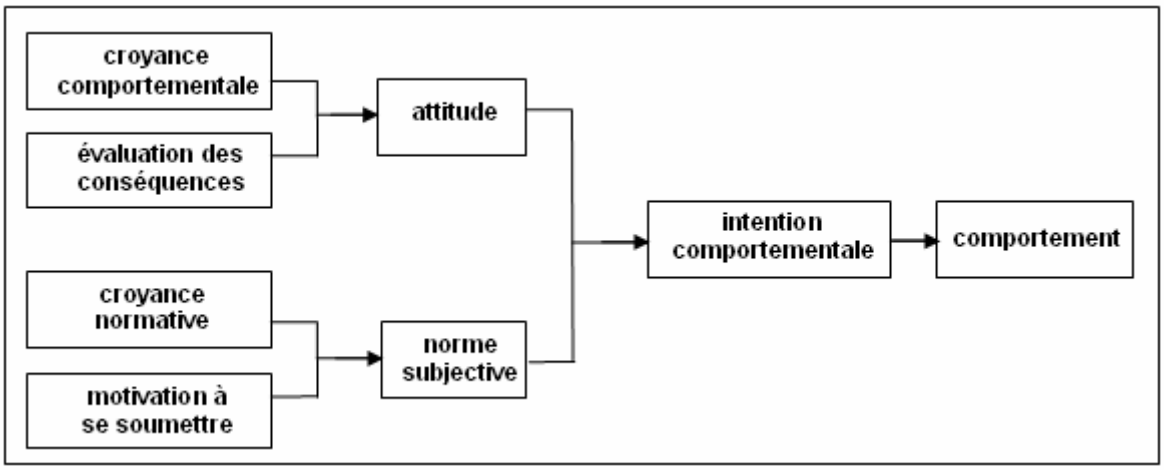


Figure 11: Théorie de l'action raisonnée (d'après Channouf, Py, et Somat, 1996)

Désormais ancienne, cette théorie fait toujours l'objet de recherches et d'améliorations. Parmi elles, la théorie du comportement planifié (*Theory of Planned Behaviour, TPB*, (Ajzen, 1985)). Elle ajoute à la TRA un nouvel antécédent à l'intention comportementale : le contrôle comportemental perçu. Il est défini comme la compétence ressentie par un individu de mener à bien un comportement. Ce concept est identique à celui d'auto-efficacité (Bandura, 1977), puisque les deux renvoient à la croyance d'un individu d'être capable de réaliser un comportement. Seule la façon de les mesurer diffère, puisque dans l'un ou l'autre cas, cette démarche poursuit des objectifs différents. Cela renvoie en fait à l'expérience antérieure de la personne quant à la réalisation du comportement considéré et des obstacles supposés.

Ni la TPB, la TRA ou l'IDT n'ont, à notre connaissance, été utilisées pour évaluer l'acceptation et la diffusion de logiciels de reconnaissance de l'écriture manuscrite. Par contre, la TPB (plus que la TRA) a été appliquée à l'acceptation de nouvelles technologies avec des résultats mitigés (Dillon & Watson, 1996). On retiendra cependant qu'une étude récente sur l'acceptation des PDA par 222 médecins hospitaliers a montré que le contrôle comportemental perçu peut se révéler prédictif de la facilité d'utilisation perçue ainsi que de l'intention d'usage (Yi, Jackson, Park, & Probst, 2006). Quoi qu'il en soit, dans d'autres champs de recherches, des méta-analyses témoignent du caractère prédictif indubitable de ces deux modèles (Hagger, Chatzisarantis, & Biddle, 2003).

1.3.2.2. Le Modèle d'Acceptation de la Technologie (TAM, Davis, 1989)

Conscient des difficultés à expliquer les mécanismes cognitifs d'acceptation de nouvelles technologies, Davis propose le modèle TAM (*Technology Acceptance Model*) dès 1986 (cité par Davis, Bagozzi, & Warshaw, 1989). Alors que la TRA est un modèle général applicable à de nombreux comportements et situations, le TAM se limite quant à lui au champ du management des systèmes d'informations (MIS, *Management of Information Systems*). Le but du TAM est de prédire l'acceptation d'un système d'information et de diagnostiquer les problèmes de conception de l'interface en amont de sa diffusion. S'il est permis de penser que la TRA est suffisante pour répondre à cette problématique, la perspective choisie par Davis et ses collaborateurs est beaucoup plus économique, aussi bien dans son articulation théorique que dans sa mise en œuvre dans le cadre d'une recherche. Concevoir un questionnaire s'appuyant sur la TRA exige un travail important de construction d'items basé sur une sélection de croyances des individus quant au comportement qu'ils pourraient projeter de réaliser. En effet, la TRA est un modèle général qui ne spécifie pas les croyances engagées dans la réalisation d'un comportement donné : il est à adapter à la situation testée. Davis va franchir cet obstacle en proposant que des croyances relatives à l'utilité perçue (*perceived usefulness*, PU) ainsi qu'à la facilité d'utilisation perçue (*perceived ease of use*, PEOU) suffisent à expliquer le comportement d'adoption d'un système d'information. Le choix de ces mesures n'est pas anodin puisqu'elles sont récurrentes dans un certain nombre de travaux sur la diffusion technologique. On pensera en particulier à la facilité d'utilisation et l'avantage relatif de l'IDT (assimilé par certains à l'utilité (Adams, Nelson, & Todd, 1992)) ou encore à l'auto-efficacité de Bandura (1977)). L'utilité perçue est définie par Davis comme étant « l'évaluation subjective par un individu, que l'utilisation potentielle d'un dispositif améliorera sa performance au travail» (notre traduction). La facilité d'utilisation perçue renvoie elle au « degré potentiel auquel un utilisateur espère qu'un système donné sera manipulable sans effort » (Davis, Bagozzi, & Warshaw, 1989, notre traduction). L'utilité et la facilité d'utilisation sont des perceptions fondées sur des croyances de l'utilisateur à propos du système. Ces deux dimensions ont fait l'objet d'une étude de construction d'échelles conduite par Davis visant à asseoir leur validité et leur fiabilité (1989). Une réplication, en 1992, a confirmé ces résultats (Adams, Nelson, & Todd, 1992), aussi bien en termes de fiabilité que de validité. De surcroît, trois études se sont assurées que l'ordre de présentation

des items n'était pas de nature à biaiser le recueil des résultats (Davis & Venkatesh, 1996). Initialement, le TAM envisageait que l'utilité et la facilité d'utilisation perçues avaient un impact sur les attitudes de l'utilisateur quant à l'usage du dispositif, définies comme étant les opinions, bonnes ou mauvaises, envers le système. L'intention comportementale y était fonction des influences combinées des attitudes et de l'utilité perçue. L'intention d'utilisation détermine, elle, le comportement effectif, dont plusieurs recherches montrent qu'elle en est le meilleur prédicteur (Adams, Nelson, & Todd, 1992; Davis, 1989; Davis, Bagozzi, & Warshaw, 1989). Les autres facteurs, dits « externes » n'exerceraient leur influence sur l'intention comportementale que par l'intermédiaire de l'utilité perçue et la facilité d'utilisation perçue. Ces variables externes renvoient aux caractéristiques de la tâche, aux propriétés de l'interface, aux caractéristiques de l'utilisateur (ex : style cognitif, personnalité, influence politique, *etc.*), ou encore au processus d'implémentation. Les antécédents de l'utilité et de la facilité d'utilisation perçues ont fait l'objet d'études visant à les identifier (Arning & Ziefle, 2007; Brosnan, 1999; Burton-Jones & Hubona, 2006; Hasan & Ahmed, 2007; Legris, Ingham, & Collerette, 2003; McFarland & Hamilton, 2006; Porter & Donthu, 2006; Saadé & Kira, 2007; Venkatesh & Davis, 1996).

Le modèle TAM est donc moins général que la TRA, mais plus adapté à la question de l'acceptabilité des technologies, et plus particulièrement encore, pour le monde du travail. Cette dernière considération n'est pas anodine, car bien que le TAM s'appuie sur les fondations théoriques de la TRA, sa mise à l'épreuve a imposé un certain nombre de modifications au modèle de départ. Notamment, la disparition de la norme subjective comme facteur explicatif de l'intention comportementale. A cela deux raisons : tout d'abord, la norme subjective est le facteur le moins bien connu de la TRA, et il n'est pas exclu qu'il influence les attitudes (par internalisation notamment), de sorte que l'évaluation de ses effets n'est pas aisée. Ensuite, Davis rapporte que l'usage de l'informatique par les professionnels est, sinon volontaire, plus dicté par les exigences de réalisation du travail demandé que par des sentiments ou croyances (Carlopio, 1988)… Si cela était sans doute vrai en 1989, la question reste entière aujourd'hui puisque l'informatique a largement pénétré la sphère privée. En tout état de cause, dans une étude comparant le TAM à la TRA, Davis *et al.* n'ont pas relevé d'influence significative de la norme subjective. Néanmoins, ils préconisent que ces mécanismes d'influence sociale fassent l'objet, dans d'ultérieures recherches,

d'approfondissements[6], car le logiciel de traitement de texte dont ils testent l'acceptabilité pour cette étude n'est pas adapté à mettre en évidence le rôle de la norme subjective. En effet, le choix d'un traitement de texte n'a que peu de visibilité sociale. A cet égard, l'utilisation d'un PDA est vraisemblablement plus connotée. En réalité, la norme subjective influence l'utilité perçue et l'intention d'utilisation d'un système quand son usage est imposé (Venkatesh & Davis, 2000). Ce constat a amené les auteurs à proposer le modèle TAM 2, adapté à cette situation particulière d'usage prescrit.

Autre différence importante, le TAM, après validation expérimentale, a montré une relation directe forte entre l'utilité perçue et l'intention d'utilisation, et ce, au détriment du pouvoir explicatif des attitudes, contrairement à la TRA. Celle-ci veut que les attitudes médiatisent complètement l'influence des croyances sur les intentions (dans le cas ou la norme subjective n'aurait aucune influence). Selon Davis, dans l'environnement de travail, l'intention d'utilisation des technologies est basée plus sur l'anticipation de la performance réalisable que sur le sentiment général de l'utilisateur envers le système. En d'autres termes, si un employé peut ne pas aimer un système, il pourra tout de même l'utiliser s'il pense qu'il peut améliorer ses performances. Ce manque d'influence des attitudes est d'autant plus prononcé que l'usage effectif est mesuré loin dans le temps (14 semaines après). Cela a conduit l'auteur à préconiser l'abandon de la mesure des attitudes peut-être prématurément s'il on en croit Yang et Youngjin (2004) qui montrent que si l'on fait une distinction entre attitudes cognitives et affectives, le potentiel explicatif des attitudes cognitives sur l'usage effectif devient plus de deux fois supérieur à celui de l'utilité perçue. D'autres également ont relevé l'influence des attitudes, sans pour autant faire cette distinction (Dillon & Morris, 1997), mais à plus courte échéance (2 semaines). Cette volonté de parcimonie par la suppression des attitudes que l'on retrouve chez Adams *et al.* (Adams, Nelson, & Todd, 1992) a pu conduire, semble-t-il, aux résultats parfois contradictoires et instables dont témoigne leur réplication (seules les influences sur l'usage de l'utilité et de la facilité d'utilisation perçues sont entrées dans leur modélisation structurale, et les auteurs préconisent également dans leur discussion de réexaminer le rôle médiateur des attitudes associé à celui de variables

[6] Cette problématique a par la suite fait l'objet de plusieurs travaux, conduits, entre autres, par Davis. Ils sont détaillés dans la section suivante et ont donné lieu à la création du modèle TAM 2 et du modèle UTAUT (*Unified Theory of Acceptance and Use of Technology*). Il est désormais admis que la norme subjective, de même que la pression sociale influencent conjointement l'intention d'usage, sous certaines conditions.

contextuelles)[7]. La dernière différence entre le TAM et la TRA est l'effet direct de la facilité d'utilisation perçue sur l'utilité perçue. Concrètement, quand deux systèmes proposent des fonctions identiques, un individu trouvera celui qui est le plus simple, le plus utile. Cette influence de la facilité d'utilisation sur l'utilité n'est pas réciproque. Dans leur travail de validation du TAM, Davis et ses collaborateurs ont trouvé une plus forte relation entre l'utilité perçue et l'intention d'utilisation qu'entre la facilité d'utilisation perçue et l'intention d'usage. La relation entre la facilité d'utilisation perçue et les intentions est largement médiatisée par l'utilité perçue.

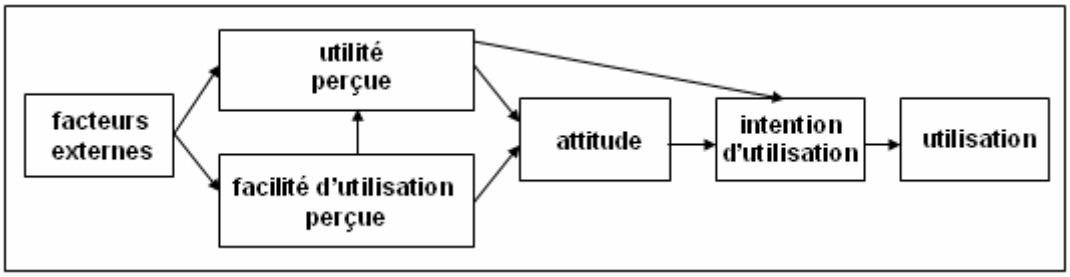


Figure 12: Modèle d'acceptation de la technologie (Davis, 1989)

En comparant le TAM à la TRA, Davis et son équipe montrent que le TAM est un meilleur prédicteur (basé sur la quantité de variance expliquée, R^2) de l'intention d'utiliser un traitement de texte, témoignant d'un $R^2 = .47$ immédiatement après la présentation du logiciel et d'un $R^2 = .51$ quatorze semaines après. Ces chiffres sont comparés respectivement aux .32 et .26 pour la TRA. Le modèle TAM a depuis été largement adopté par la communauté MIS. Plusieurs études font cas de sa robustesse, et des réplications ont été produites pour différentes technologies (principalement des logiciels ou service disponibles en ligne gratuitement). Citons pour l'exemple celle de Morris et Dillon (1997) qui teste l'adoption du navigateur *Netscape* auprès d'étudiants en informatique. Ils collectent des résultats qui concordent avec ceux obtenus par Davis *et al.*, à cette différence près qu'ils confirment l'influence des attitudes sur l'intention comportementale ($R^2 = .28$; contre .27 pour Davis *et al.* 1989) plutôt que d'évaluer directement l'impact de l'utilité et de la facilité d'utilisation perçues sur cette intention. Ils valident la puissance du TAM puisque qu'après une heure de découverte du logiciel, l'intention d'utilisation prédit l'utilisation deux semaines après (R^2 =

[7] La question de la parcimonie du TAM et des modèles plus récents qui ont vocation à l'améliorer est discutée à la section suivante.

.18 ; β = .46 et p < .001 ; chiffres comparables à ceux obtenus de Davis et son équipe : R^2 = .12 ; β = .35 et p < .001).

Il n'est bien évidemment pas envisageable de présenter ici toutes les études ayant mis à profit le TAM. C'est pourquoi quelques auteurs ont construit des méta-analyses de l'exploitation scientifique du TAM. Parmi eux, King et He (2006) ont synthétisé les résultats de 88 études appliquant le TAM à leurs problématiques respectives. Il en ressort que le TAM fait preuve d'une grande robustesse dans de multiples contextes, en expliquant jusqu'à plus de 40% de la variance d'intention d'usage, et ce, après une période d'utilisation relativement courte. En dépit de ses succès, le TAM n'est pas exempt de critiques. On retiendra notamment l'utilisation très importante d'échantillons issus de populations étudiantes, le choix des technologies testées (se limitant souvent à des applications bureautique en ligne), et pour finir, la mesure du comportement d'utilisation basée le plus souvent uniquement sur le témoignage auto-rapporté des sujets (Legris, Ingham, & Collerette, 2003, pour une autre méta-analyse).

1.3.2.3. Extensions du TAM

1.3.2.3.1. Les variables modératrices

Cependant, l'étude de King et He (2006) s'interroge sur l'importance de variables modératrices telles que le type d'utilisateur (tout venant, étudiant ou professionnel) et le type d'usage (bureautique / professionnel, général, ainsi qu'internet / e-commerce). Effectivement, si les corrélations entre études sont fortes, les auteurs constatent une forte variabilité entre études les invitant à supposer le rôle de variables modératrices pour justifier ces effets. Bien qu'il soit notable que le niveau de compétence préalable joue un rôle important dans le processus d'acceptation, cet aspect, souvent non pris en considération par les chercheurs pour leurs études, n'apparaît pas en tant que prédicteur dans leurs relevés de mesures, de sorte qu'il n'a pas été possible à King et He de l'inclure dans leur traitement. Il semble néanmoins que cette considération ne soit pas anodine si l'on s'attache au résultat suivant, mis en évidence par la méta-analyse : les influences respectives de l'utilité perçue et de la facilité d'utilisation perçue diffèrent significativement en fonction des populations dont sont issues les échantillons de test. S'il semble que l'on puisse assimiler les étudiants aux professionnels, ces derniers diffèrent d'une population tout-venant, pour laquelle la facilité d'utilisation revêt une

importance plus grande que ce que laissaient supposer les travaux antérieurs. En effet, il est souvent admis qu'après plusieurs semaines d'utilisation, ce *construct* n'a plus d'influence directe sur les attitudes mais se voit entièrement médiatisé par l'utilité perçue. Cette intuition de King et He (2006) a été confirmé plus récemment dans le cadre d'une adaptation du TAM circonscrite au *Web* (*Web Acceptance Model*, *WAM*) et dans laquelle il est expérimentalement mis en évidence le rôle modérateur des connaissances antérieures (Castañeda, Muñoz-Leiva, & Luque, 2007).

D'une manière générale, c'est dans toute leur diversité que les caractéristiques des individus peuvent modérer l'intention d'usage. Pour l'heure, aucune liste exhaustive de ces dernières n'a encore été dressée. C'est un champ d'étude et d'amélioration du TAM tout à fait récent. Les travaux qui s'y rapportent ne permettent pas encore d'apprécier avec le recul nécessaire l'importance de leur contribution future. Comme le soulignent Arning et Ziefle (2007), l'absence de réplication et les résultats parfois contradictoires de ces études interdisent toute conclusion prématurée. C'est pourquoi ils proposent une expérience visant à déterminer les rôles joués par l'âge, le sexe et la performance dans l'intention d'usage d'un PDA. Ils placent leurs sujets novices dans une situation d'utilisation d'un PDA, émulé sur un PC de bureau. L'épreuve consiste en des tâches de navigation dans un agenda électronique avec la souris. Les auteurs examinent plusieurs caractéristiques individuelles telles que l'âge, le sexe, l'expérience de l'informatique et la confiance technique subjective (*Subjective Technical Confidence*, STC, qui est la croyance d'un individu dans sa capacité à résoudre des problèmes techniques). Les résultats montrent que l'interaction du sexe et de l'âge influencent la confiance technique subjective, ainsi que l'utilité et la facilité d'utilisation perçues. A l'inverse, l'expérience en informatique n'a pas d'effet sur ces différences de jugement. Les jeunes hommes présentent les scores les plus élevés de confiance et de perception subjective. Concernant la performance (mesurée selon les critères d'efficacité et d'efficience préconisés par la norme ISO 9241-11), l'âge est un facteur fort de différences, les jeunes se montrant plus efficaces que leurs aînés. De façon inattendue, le sexe n'a aucun effet sur la performance ou les jugements, contrairement à ce dont témoignent d'autres recherches (Ong & Lai, 2006; Venkatesh, Morris, Davis, & Davis, 2003). Ces résultats se confirment à l'aune des analyses de régression. L'efficience est largement prédite par l'âge et la facilité d'utilisation (R^2 = .68), bien qu'un effet modérateur important de l'âge sois mis en évidence s'agissant du potentiel prédictif de la facilité perçue (R^2 = .22 pour les jeunes, contre R^2 = .41 pour les seniors).

L'efficacité est aussi largement expliquée par l'âge et la facilité perçue (R^2 = .71), à cette différence que l'âge perd son potentiel modérateur. La confiance technique subjective n'explique quant à elle qu'une part infime de la variance. Pour conclure sur ce versant de cette étude, la prise en considération de l'âge des participants met en lumière que les jeunes appuient leur perception de la facilité d'utilisation sur leur efficience, alors qu'en avançant en âge, les sujets se fondent principalement sur leur efficacité.

Plusieurs autres modérateurs ont été mis au jour par d'autres études. On notera par exemple, en plus de la population (étudiante, professionnelle ou tout-venant), le type de technologie utilisée (micro-ordinateur ou non, technologie familière ou inhabituelle) ou la culture (occidentale ou asiatique) (Schepers & Wetzels, 2007). Plus complète, une tentative d'unification de huit modèles prédictifs de l'acceptation des technologies a été menée afin de ne retenir que le meilleur de chacune de ces constructions théoriques. L'aboutissement de cet immense travail de synthèse a vu le jour sous le nom de *Unified Theory of Acceptance and Use of Technology* (Théorie Unifiée de l'Acceptation et de l'Utilisation des Technologies, UTAUT) (Venkatesh, Morris, Davis, & Davis, 2003). Cette étude a mis en évidence quatre modérateurs aux déterminants de l'intention d'utilisation et de l'usage effectif que sont le sexe, l'âge, l'expérience d'usage, et la volonté d'utilisation. Ce nouveau modèle, qui explique jusqu'à 70% de la variance de l'usage effectif repousse dans ses dernières limites notre capacité à prédire un comportement d'adoption d'une technologie. L'intention comportementale y est prédite par la performance prévue (*performance expectancy*), l'effort prévu (*effort expectancy*), l'influence sociale (*social influence*) et les conditions facilitantes (*facilitating conditions*). Ce modèle adjoint donc au TAM les dimensions sociales et environnementales qui lui faisaient originellement défaut (Legris, Ingham, & Collerette, 2003; Venkatesh & Davis, 2000).

1.3.2.3.2. Les facteurs externes

La plus importante limite du TAM est son incapacité à permettre de déterminer quels sont les facteurs externes sur lesquels le concepteur d'IHM va pouvoir intervenir pour améliorer leur acceptabilité, et de fait, leur acceptation. Ceci est vrai pour tous les modèles cognitivo-comportementaux développés précédemment. La description des mécanismes cognitifs et psycho-sociaux qui conduisent à l'adoption d'un dispositif ne permet pas d'envisager de se

fonder sur ces seuls résultats pour diriger efficacement une intervention sur le système dans le cadre d'un cycle de conception itératif. Quelques travaux toutefois se sont attachés à mettre en évidence les facteurs externes déterminant les prédicteurs de l'intention comportementale. Contrairement à l'intuition de Davis (1989) suivant Fishbein et Ajzen (1975), ces facteurs externes peuvent influencer directement l'intention d'usage (Burton-Jones & Hubona, 2006; McFarland & Hamilton, 2006). Une méta analyse fait l'inventaire d'une grande partie de ces facteurs externes étudiés (Legris, Ingham, & Collerette, 2003). Ces derniers relèvent le plus souvent de croyances, d'attitudes et de critères démographiques ou sociaux. On citera pêle-mêle anxiété ou techno-phobie (Brosnan, 1999; McFarland & Hamilton, 2006; Saadé & Kira, 2007), l'âge, le niveau d'éducation (Burton-Jones & Hubona, 2006; Porter & Donthu, 2006), l'auto-efficacité (Venkatesh & Davis, 1996) ou les revenus (Porter & Donthu, 2006). Hélas, seules quelques recherches ont tenté de mettre en relation l'utilisabilité avec l'intention d'utilisation (Arning & Ziefle, 2007; Hasan & Ahmed, 2007; Venkatesh & Davis, 1996).

S'agissant de l'étude de Hasan et Ahmed (2007), on ne peut pas à proprement parler d'utilisabilité. Cependant, les auteurs s'interrogent sur l'effet du style d'interface (lignes de commandes ou menus) sur la facilité d'utilisation et l'utilité perçues ainsi que sur l'intention d'utilisation. Par conséquent, s'ils ne procèdent pas à des mesures d'utilisabilité telles que nous les avons présentées, ils comparent deux interfaces dont l'utilisabilité est différente. Leur analyse révèle que le style d'interface influence à la fois la facilité d'utilisation perçue et l'utilité perçue (conjointement avec la facilité d'utilisation). Par contre, le style d'interface ne permet pas de prédire l'intention d'utilisation, dont la variance est assumée à 61% par la facilité d'utilisation et l'utilité perçues. Toutefois, une étude a pointé en tant que facteur externe d'acceptation d'un logiciel, l'expérience antérieure en informatique, et plus particulièrement, celle avec d'autres logiciels de même type que celui évalué. En effet, une interface peut être jugée différemment suivant qu'elle soit découverte par un utilisateur potentiel sans autre cadre de référence que le dispositif lui-même (désigné comme étant son utilisabilité intrinsèque), ou bien en comparaison d'une interface différente aux fonctionnalités identiques (auquel cas les utilisateurs jugeraient à partir de similarités de surface et de fonctions entre les deux systèmes). C'est ce qui a été mis en évidence en comparant différents styles d'interfaces sans pour autant en mesurer l'utilisabilité en tant que telle. Il apparaît notamment qu'une interface « difficile » sera jugée plus négativement après avoir testé une interface plus facile d'utilisation (Wiedenbeck & Davis, 1997).

Venkatesh et Davis (1996) soulignent que la notion d'interface est différente de celle d'utilisabilité, et que si l'idée d'interface permet de faire des comparaisons entre dispositifs en tant que variables factices, la comparaison de leur utilisabilité doit se faire par des mesures objectives. Pour ce faire, ils déterminent l'utilisabilité objective des systèmes qu'ils utilisent pour leurs expériences grâce au KLM (*Keystroke Level Model*, une adaptation du modèle GOMS, acronyme de *Goal, Operators, Methods and Selection rules*, qui est un modèle informatique d'utilisateur) ainsi qu'à des experts spécialistes de chaque interface (deux experts par interface). Cela leur permet de déterminer l'utilisabilité objective des interfaces en calculant le rapport entre le temps de réalisation des tâches par les experts et les temps mis par les sujets novices. Leur étude met clairement en évidence que l'utilisabilité objective peut prédire la facilité d'utilisation perçue (de même que le sentiment d'auto-efficacité) quand les participants ont eu l'opportunité de tester la machine. Cet effet de l'utilisabilité objective est modéré par la possibilité de tester effectivement le dispositif, car, si l'on se contente d'en faire une démonstration aux participants, cet effet s'efface au profit de la seule auto-efficacité. Bien que l'on puisse critiquer cette façon de mesurer l'utilisabilité, ce travail à l'immense mérite de poser la question de l'intention d'utilisation en relation avec l'utilisabilité du système. Les auteurs insistent en effet sur l'importance de pouvoir déterminer quelles caractéristiques de l'interface peuvent poser problème pour son acceptabilité. Malgré cela, leur choix de mesure de l'utilisabilité, bien qu'expliquant une part substantielle de la variance de la facilité d'utilisation perçue n'est pas de nature à fournir une explication quant à d'éventuels problèmes d'utilisabilité. On regrettera que cette voie de recherche n'ait pas donné lieu à d'autres travaux visant les mêmes objectifs, tout en tenant compte des développements ultérieurs en matière d'utilisabilité.

A ce titre, l'étude proposée par Arning et Ziefle (2007) tient compte des avancées réalisées en la matière. En mesurant l'utilisabilité par le biais de l'efficacité (nombre de tâches accomplies) et de l'efficience (temps de réalisation), ils mettent en évidence qu'elle prédit significativement la facilité d'utilisation perçue.

Ce type d'étude, encore très récent, est appelé à prendre de l'importance comme certains auteurs l'appellent de leurs vœux (Tricot & Tricot, 2000; Venkatesh & Davis, 1996; Zhang & Dillon, 2003).

1.3.2.4. Critique des modélisations de l'acceptation

Avec 70% de variance expliquée par le modèle UTAUT (Venkatesh, Morris, Davis, & Davis, 2003), dernière extension du modèle TAM (Davis, 1989; Davis, Bagozzi, & Warshaw, 1989), ce type de modèles arrive aux limites de notre capacité à prédire des comportements d'utilisation de technologie. Dans le champ MIS, les critiques de ces théories portent essentiellement sur les mesures de l'utilisation effective, qui, le plus souvent, sont auto-rapportées par les sujets et non vérifiées. Legris et ces collaborateurs évoquent à ce propos qu'une enquête réalisée sur des individus tout-venant témoignait que leur utilisation auto-rapportée du lavabo en sortant des toilettes était très largement surévaluée par rapport à ce que les caméras de surveillance permettaient de comptabiliser. Autre point de critique, l'utilisation quasi-exclusive d'applications bureautique ou e-commerce accessibles en ligne. L'acceptation de technologies matérielles a été largement délaissée (Legris, Ingham, & Collerette, 2003).

D'autre part, si, sur le plan cognitif, l'inspiration du TAM dans la TRA (Fishbein & Ajzen, 1975) lui donne un crédit certain, l'idée selon laquelle l'acceptation est déterminée par l'utilité (principalement) et la facilité d'utilisation (dans une moindre mesure), n'est pas partagée par tous. Dans le domaine du marketing notamment, comme le souligne Tricot et ses collaborateurs (Tricot et al., 2003), on fait valoir qu'une technologie « acceptable et utilisable […] est [prête] à être éventuellement utile » (Ketola & Röykkee, 2001). Peut-être que cette contradiction trouve tout simplement son explication dans la distinction nécessaire entre acceptation et acceptabilité, sur laquelle nous avons déjà insisté.

Si l'on sort du cadre MIS, la principale critique portera évidemment, de notre point de vue, sur les lacunes du TAM inhérentes à la prise en compte de l'utilisabilité en tant que déterminant de la facilité d'utilisation perçue. Malgré quelques tentatives prometteuses (Arning & Ziefle, 2007; Venkatesh & Davis, 1996), des travaux complémentaires sont indispensables. Pour finir, notons que les PDA, et, à plus forte raison, les dispositifs de saisie et de reconnaissance de l'écriture manuscrite n'ont que peu fait l'objet de tests d'acceptation. On peut, bien sûr, évoquer l'étude d'Arning et Ziefle (2007), mais le protocole de test passant par l'émulation d'un PDA sur PC entache considérablement la validité écologique du test.

1.4. Vers une unification des modèles d'utilisabilité et d'acceptation

Notre exposé montre que deux conceptions de l'acceptation / acceptabilité se dégagent : l'une s'intéresse aux déterminants sociaux et cognitifs de l'acceptation tandis que l'autre envisage l'acceptation comme la conséquence d'une démarche de conception de l'interface. Ces deux approches, connues sous les sigles de MIS (*Management of Information Systems*) et HCI (*Human-Computer Interaction*) font l'objet d'une discussion des champs d'intervention qu'elles ont en commun (Zhang & Dillon, 2003). Il ressort de cette analyse la nécessaire convergence des deux disciplines dont la capacité à prédire les usages par l'utilisabilité seule d'une part, et la capacité à envisager la part de l'utilisabilité dans l'acceptation d'autre part, touchent à leurs limites. De fait, si l'utilisabilité d'un système d'information est une condition nécessaire à son acceptation, elle n'est néanmoins pas suffisante (Dillon, 2001b). A l'inverse, prédire l'acceptation qu'un dispositif suscitera chez une population n'a d'intérêt pour la communauté HCI que si elle peut s'en servir pour déterminer à un stade précoce de son développement les modifications qui s'y imposent (Jamet & Février, 2008). Le rapprochement de ces deux perspectives de recherches n'a pas encore donné lieu à la création de modèles tout à fait satisfaisants. Il est vrai que la complexité des facteurs à prendre en compte (ergonomiques, cognitifs, psycho-sociaux et socio-techniques) rend la tâche particulièrement ardue. Toutefois, Dillon et Morris ont proposé un modèle unifiant ces deux perspectives : le modèle P3 pour *Power, Perception and Performance* (Potentiel, Perception et Performance, notre traduction) (1998, 1999). Ce modèle, qui n'est que théorique pour l'instant, propose de mettre en relation, pour un système donné, des mesures de fonctionnalités de la machine (indépendamment de la capacité des utilisateurs à les exploiter), des mesures attitudinales (telles que celles du TAM, par exemple) et des mesures de performances (mesures d'utilisabilité telles que développées précédemment). Ce modèle ambitionne de contourner le fait que l'on ne puisse pas considérer si la facilité d'utilisation perçue est l'utilisabilité, de la même façon que l'on ne peut pas assimiler l'utilité perçue au potentiel de la machine.

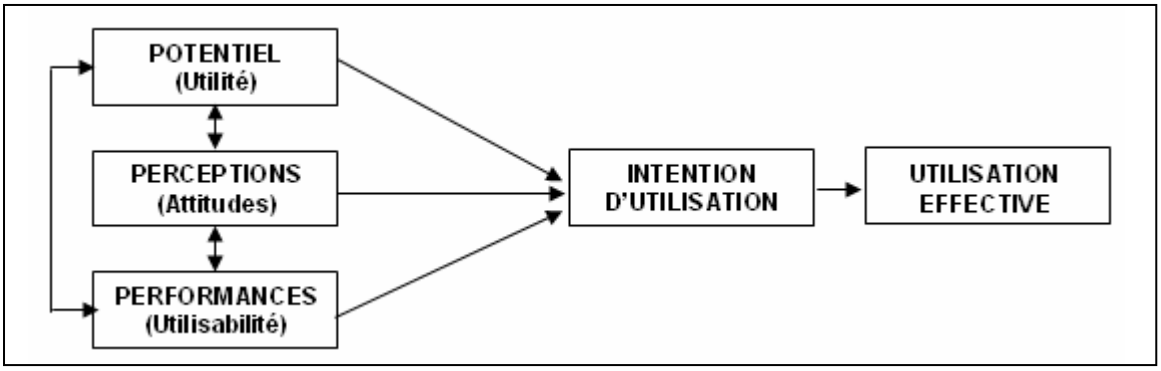


Figure 13: Le modèle P3 (d'après Dillon & Morris, 1998, notre traduction)

Le modèle P3 propose donc de prédire l'usage d'un dispositif à la fois par ce qu'il permet de faire objectivement (potentiel), par son utilisabilité (c'est-à-dire les performances qu'il autorise) et par les jugements que l'utilisateur porte sur son utilisation. En revanche, ses auteurs ne préconisent aucune méthode de mesure du potentiel des logiciels. Fiorello (1999) propose de s'appuyer sur la méthode des points de fonction (*Function Points Method*, Pressman, 1992, cité par Fiorello, 1999). Cette méthode, utilisée en ingénierie logicielle, permet d'estimer la taille et la complexité d'une application en cours d'élaboration. Fiorello signale que si cette méthode ne permet pas de déterminer le potentiel d'un logiciel, au moins en donne-t-elle une mesure qui lui est corrélée.

Dans les faits, lui-même n'utilise pas cette mesure pour son étude, et la définition objective du potentiel d'un dispositif reste éminemment problématique. Ce modèle prévoit que chacune de ces dimensions influence directement et/ou indirectement l'intention d'utilisation d'un dispositif technologique. Par ailleurs, ces dimensions peuvent s'influencer entre elles. Une tentative de validation de ce modèle a été proposée, avec un succès limité (Fiorello, 1999) dans le cadre d'un essai du logiciel *Dragon NaturallySpeaking* en vue de son implémentation au sein de l'armée de l'air états-unienne. Après deux séances (entraînement puis tâche de copie d'un texte) d'une heure et demie d'utilisation de ce logiciel de reconnaissance de discours (espacées de 48h maximum), les sujets remplissaient les questionnaires TAM (mesure de l'intention d'utilisation, de la facilité d'utilisation perçue, et de l'utilité perçue) (Davis, 1989; Davis, Bagozzi, & Warshaw, 1989), QUIS (Chin, Diehl, Norman, 1988, cités par Fiorello, 1999) et CSUQ (Lewis, 1995) (questionnaires d'utilisabilité perçue et de satisfaction). Par ailleurs, leurs performances étaient enregistrées (nombre d'erreurs, nombre de mots correctement saisis, durée de réalisation de la tâche). L'auteur met

en évidence que le TAM (avec la facilité d'utilisation et l'utilité perçues) est le meilleur prédicteur de l'intention d'utilisation. Il montre également que ni le nombre d'erreurs, ni le nombre de mots correctement saisis, ni le temps de réalisation de la tâche n'expliquent l'intention d'utilisation ou toute autre mesure subjective fournie par les questionnaires TAM, QUIS et CSUQ. On reprochera à cette étude de ne pas proposer, au chapitre des mesures d'utilisabilité, une mesure d'efficience (ne serait-ce que temporelle (Bevan & Macleod, 1994)) unifiant à la fois nombre de mots saisis, erreurs et temps de réalisation.

En dépit de cet échec à valider le modèle P3, ainsi que le manque de retentissement de cette théorie, aussi bien dans la communauté HCI que MIS, il apparaît que les relations entre utilité, utilisabilité et acceptabilité ne sont pas nécessairement orientées comme le laissent entendre les modèles présentés précédemment (Tricot et al., 2003). Exemples à l'appui, dans une réflexion menée dans le cadre des EIAH (Environnements Informatiques pour l'Apprentissage Humain), Tricot et ses collaborateurs montrent que l'acceptabilité peut, dans le cas des *pagers*, précéder l'utilité. Ils concluent que toutes relations entre utilité, utilisabilité et acceptabilité sont *a priori* possibles. Il suffit, pour s'en convaincre, d'examiner attentivement les différents modèles présentés. Le modèle de Nielsen (1993) fait de l'utilisabilité un préalable à l'utilité, ce qui n'est pas le cas chez Shackel (1991). Pour la norme ISO 9241-11 (1994), l'acceptabilité détermine en partie l'utilisabilité. L'idée n'est pas saugrenue si l'on considère qu'un système, avant d'être jugé utilisable, doit être utilisé. L'utilité est, en partie, incluse dans l'utilisabilité dans le modèle de Shackel (1991), comme le recommande Keinonen (1998). Pour finir, les modèles tels que le TAM placent l'utilité comme déterminant essentiel de l'acceptabilité (1989)… La complexité des ajustements entre la machine, les possibilités qu'elle offre, et l'utilisateur peuvent engendrer de multiples patterns d'acceptation. La recherche sur ces questions n'en est qu'à ses balbutiements.

1.5. Mesurer et améliorer l'utilisabilité

Bien que sur le plan théorique il subsiste des incertitudes quant à l'articulation de l'utilité et de l'acceptabilité avec l'utilisabilité, celle-ci bénéficie en revanche d'un large panel de méthodes permettant de la mesurer et, le cas échéant, de l'améliorer.

1.5.1. L'utilisabilité comme processus et méthode de conception

Au-delà même de la difficulté à trouver un consensus pour définir l'utilisabilité, il est possible d'appréhender ce concept de différentes façons. Pour certains auteurs, l'utilisabilité est avant toute chose une approche de la conception de produits (Bevan, 2006; Keinonen, 1998). Que l'on parle d'ingénierie ergonomique ou de conception centrée utilisateur, il est désormais courant dans le processus d'élaboration d'une IHM de lui faire bénéficier d'une prise en considération de l'utilisabilité finale du produit, en regard des attentes supposées des utilisateurs finaux et des contraintes de leur tâche (Stewart & Travis, 2003). Cela passe aussi bien par l'application d'heuristiques ou de recommandations ergonomiques que par des méthodes plus lourdes, souvent prises en charge par des départements spécialisés, telles que les tests utilisateurs ou la conception participative, mettant à contribution l'utilisateur cible (Keinonen, 1998). Bevan (2006) arrive à cette conclusion en envisageant les types de recommandations pour la conception des IHM et l'utilisabilité selon quatre catégories :

1. la capacité d'une organisation à utiliser et appliquer une conception centrée utilisateur (ex : Un projet de développement d'une interface a-t-il un département spécialisé dans la qualité ergonomique ou simplement la volonté de prendre en considération cette question).
2. le processus utilisé pour développer le produit (ex : évaluation heuristique, évaluation experte, test utilisateur, méthode expérimentale, questionnaires…).
3. L'interface utilisateur et l'interaction possible (ex : critères qui ont présidé à la conception du produit, c'est la qualité ergonomique intrinsèque du dispositif).
4. L'utilisation du produit, et ce qu'elle engendre en termes de performance et de satisfaction pour l'utilisateur (ex : efficacité, efficience, satisfaction… C'est-à-dire l'utilisabilité telle que nous l'entendons).

Bien que n'étant pas opposés à cette approche, nous pensons qu'elle entretient une certaine confusion quant à la définition de l'utilisabilité. L'inexactitude la plus fréquente consistant, selon nous, à assimiler la qualité ergonomique d'une interface à son utilisabilité. A cela trois raisons : tout d'abord, cette confusion est d'autant plus facile à faire que les méthodes d'évaluations de ces deux dimensions peuvent être identiques (cf.2). De plus, ce que nous considérons être l'utilisabilité (cf. 4) est une conséquence de la qualité de l'interface (cf.3). Bastien et Scapin (2001) font également la distinction entre utilisabilité et critères

ergonomiques. Ils associent cette distinction aux méthodes d'évaluation des interfaces requérant la participation directe des utilisateurs de celles qui s'en affranchissent.

1.5.2. L'inspection ergonomique

Bastien et Scapin (2001) proposent une taxonomie décrivant les différentes méthodes possibles pour évaluer les caractéristiques ergonomiques d'une interface sans la participation directe d'utilisateurs. Elles peuvent s'appuyer sur des évaluations expertes (Baccino, Bellino, & Colombi, 2005), ou être fondées sur des outils variés tels que les normes (notamment ISO 9241-10 relative au principes de dialogue), des recommandations (*guidelines*) pour PDA (Jansen & Ayers, 2004) ou encore des principes ergonomiques. L'utilisation de ces ressources va permettre de détecter, par inspection, de possibles erreurs de conception susceptibles d'entraver l'utilisation d'un dispositif.

Nous retiendrons les *critères ergonomiques* de Bastien et Scapin (2001) comme le plus important travail de synthèse et de validation de principes ergonomiques.

L'ambition présidant à la création de ces critères ergonomiques était de préciser les connaissances ergonomiques en les structurant et en les classant. Ils résument les principes fondamentaux indispensables à la conception centrée utilisateur selon 8 critères organisant 18 principes élémentaires. Par un processus de validation inter-juges, 800 recommandations collectées par les auteurs dans différents guides répertoriés sur la base *Ergonomics Abstracts* ont été condensés en un nombre restreints de critères. :

- **Guidage** : Le dialogue soutient et guide l'utilisateur dans l'apprentissage du système.
- **Charge de travail** : Il s'agit du coût engendré par l'utilisation du dispositif en termes de fatigue perceptive, mnésique, physique et de stress. Il convient de concevoir les interfaces de sorte de la réduire au maximum.
- **Contrôle explicite** : L'utilisateur peut initier et contrôler la direction et le rythme de l'interaction.
- **Adaptabilité** : L'interface peut être modifiée pour s'adapter aux besoins de la tâche, aux préférences ou aux compétences des utilisateurs.

- **Gestion des erreurs** : Le résultat attendu peut être obtenu avec ou sans action correctrice.
- **Homogénéité / Cohérence** : C'est l'adéquation contextuelle des éléments de dialogue (icônes, bouton, menu, *etc.*).
- **Signifiance des codes et dénomination** : Les éléments de dialogue sont sémantiquement liés aux fonctions et informations auxquelles ils sont attachés.
- **Compatibilité** : Concerne à la fois l'accessibilité, c'est-à-dire la concordance entre l'interface et les capacités perceptives, motrices et cognitives des utilisateurs possibles, ainsi que l'interopérabilité entre l'interface et son environnement.

L'originalité de leur démarche réside dans leur tentative de validation de ces critères par l'expérimentation, et ce, afin de s'assurer de leur qualité et de leur utilisabilité. Concrètement, ils ont mis en place un protocole en trois manipulations dont le but était de s'assurer que l'utilisation de leurs critères permettait d'obtenir l'identification du plus grand nombre possible de défauts de conception sur une interface donnée. A notre connaissance, ce type de démarche appliquée à la validation d'un guide de conception d'IHM reste rare dans la littérature et fait toute la différence avec d'autres travaux visant le même usage. L'évaluation de leurs critères ergonomiques a montré qu'ils permettaient de détecter plus de défauts de conception que la plupart des autres méthodes ergonomiques existantes.

1.5.3. Le test utilisateurs

Par test utilisateurs, nous entendons toute procédure d'évaluation d'une IHM impliquant des utilisateurs en interaction directe avec l'application (hors conception participative). Dans la pratique, cette interface se réduit à un prototype dont l'état d'avancement varie en fonction des objectifs visés par le test à un stade de développement donné. L'état d'avancement du prototype déterminera le choix du type de test utilisateur (Beaudoin-Lafon & Mackay, 2003; Snyder, 2003). Ces différents types tests, pour lesquels la contribution d'utilisateurs cible est capitale (d'où leur nom), se caractérisent par l'utilisation de prototypes différents (Hom, 1998, p.58; Kaufman, 2006). C'est pourquoi nous introduisons un distinguo entre test utilisateurs et test d'utilisabilité, assimilés l'un à l'autre. Nous n'envisageons le test d'utilisabilité (Dumas, 2003) que comme une méthode du test utilisateurs, au même titre que le test de perception. Ce

dernier permet plutôt d'apprécier la compréhension immédiate d'un utilisateur vis-à-vis de l'interface, sans manipulation. Ainsi, il est déjà possible d'estimer, en plus de la compréhension, une certaine forme de satisfaction dans l'esthétique de l'interface. Par ailleurs, les couleurs ont un rôle essentiel dans la construction de pictogrammes, icônes et autres balises de navigation qui perdraient de leur signification en leur absence (Brangier, Gronier, & Pino, 2001; Watzman, 2003, p.282)

D'autre part, il ne faut pas confondre test d'hypothèse et test utilisateurs, quand bien même leurs protocoles peuvent être en apparence identiques (Dumas, 2003). Appliqué à une démarche de recherche hypothético-déductive avec contrôle des variables, le protocole du test utilisateur va perdre de sa valeur diagnostique (scénario de tâche restreint, contrôle trop important) (Gray & Salzman, 1998). Il faut donc bien distinguer les deux conceptions.

1.5.3.1. Echantillonnage, tâche et contexte

La question de l'éventuelle supériorité des tests utilisateurs sur les méthodes d'inspection n'est pas encore tranchée et fait l'objet de recherches actuelles (Dumas, 2003). De toute évidence, c'est une technique fiable, dont la mise en œuvre est porteuse de résultats, quoique délicate. Ceci tient pour beaucoup au travail en amont qui doit être fait comme préalable au test utilisateurs. En premier lieu, l'échantillonnage : les participants doivent être de potentiels utilisateurs finaux, sans quoi le test perd de sa valeur. Enquêtes, sondages, observations de terrain (Nielsen, 1993) ou études de marché, que la sélection du panel des participants se fasse sur des critères d'intention d'achat du type de produit dont il est question ou sur des variables invoquées telles que le sexe ou l'âge, il est de toute évidence impossible de réunir l'échantillon idéal. Deuxièmement, l'élaboration du scénario de test et des tâches à faire accomplir aux testeurs va revêtir la plus grande importance. On peut être très directif, ce faisant, le participant utilise les fonctions que l'on veut. Cela veut-il dire que ce qu'on lui impose correspondra à l'utilisation future qu'il en fera ? Rien n'est moins sûr. A l'inverse, l'exploration libre peut amener le sujet à négliger certaines fonctions par absence de buts précis… Cela renvoie à la spécification du contexte : à quelles occasions, pourquoi, et dans quels buts l'utilisateur final fera-t-il usage de la machine ? C'est sans doute l'échec relatif à envisager tous ces aspects dans la préparation d'un test utilisateur qui justifie que la finesse d'une inspection experte puisse s'avérer également déterminante (Bevan, 1995; Bevan &

Macleod, 1994; Dumas, 2003). Dans la pratique on se contentera souvent d'un pré questionnaire pour s'assurer que les sujets sont bien novices (ou experts) dans l'utilisation du produit (Morahan-Martin & Schumacher, 2007). Des études ont montré que les attitudes envers les technologies ou l'informatique avaient des effets sur la charge de travail et les performances (Segall, Doolen, & Porter, 2005). Il est donc intéressant de prendre ces facteurs en considération.

1.5.3.2. Le test d'utilisabilité

Cette méthodologie du test utilisateurs, en dépit de ses failles (Dumas, 2003), présente néanmoins un souci d'objectivité important. Elle va permettre d'identifier concrètement les problèmes ainsi que l'utilisabilité qu'on pourra mesurer en calculant les performances des participants. Ce test peut se faire aussi bien en laboratoire que sur le terrain, avec les avantages et inconvénients que ces lieux d'études induisent (plus ou moins de contrôle pour plus ou moins de contexte) (Kaikkonen, Kallio, Kekäläinen, Kankainen, & Cankar, 2005). Cela donne lieu à différentes façon de procéder pour collecter les informations, suivant le type d'informations que l'on cherche, les problèmes que l'on veut détecter, et les caractéristiques du prototype à l'essai (voir Folmer & Bosch, 2003 pour un inventaire). Certaines sont très directives et intrusives (exploration guidée par le testeur, apprentissage par co-découverte, triple-tâche…), d'autres moins (vocalisation externe, double-tâche…), certaines pas du tout (observation et enregistrement à distance, mesures de performance, rétrospection différée…). Il est régulièrement avancé dans la littérature que pour réaliser un test d'utilisabilité, cinq participants suffisent. En fait, Nielsen a montré qu'avec cinq utilisateurs bien choisis, on peut détecter 80% des problèmes d'utilisabilité. Au-delà de cinq, le rapport coût / problèmes détectés devient moins intéressant. Cependant, si le test utilisateur vise à comparer entre elles plusieurs interfaces dans le cadre d'un test d'hypothèse, il faudra alors autant de sujets que ce que requiert l'analyse statistique (Gray & Salzman, 1998). Cette analyse ne se fera qu'en fonction des informations recueillies durant le test. Si certaines informations peuvent être collectées après l'épreuve, d'autres le seront au cours de celles-ci. L'intérêt est de pouvoir évaluer des processus dynamiques engagés par les individus pendant l'interaction (Jamet, 2006). L'enregistrement *a posteriori* de ceux-ci engendre nécessairement une perte de qualité (voir section suivante). La difficulté de l'enregistrement *on-line* réside dans la préservation du

contexte d'évaluation (en fonction des buts poursuivis) sans sacrifier à cela la finesse du recueil.

Parmi les mesures possibles, on retiendra tout d'abord l'enregistrement des mouvements oculaires. Cette méthode peut donner de très bons résultats en ce qui concerne les tâches impliquant un écran de bureau. Cependant, elle n'est pas encore au point pour l'étude de l'informatique mobile, et ce, moins pour des raisons de taille des PDA que de mobilité, car les oculomètres sont désormais très précis. A l'inverse, les lunettes-caméras, portables et discrètes, n'enregistrent que le champ de vision central (Zouinar, Relieu, Salembier, & Calvet, 2004).

Autres types de mesures *on-line* très informatives, les techniques de double ou triple tâches. Elles servent à quantifier objectivement la charge mentale, c'est-à-dire la quantité de ressources mentales investies par l'utilisateur dans la réalisation de sa tâche. Le principe de ces méthodes est fondé sur le postulat selon lequel un individu a un réservoir attentionnel et mnésique à capacité limitée (voir chap. 2 sur la mémoire de travail). Par conséquent, plus sa tâche dite principale (*i.e.* celle qu'il a à accomplir grâce au système testé) sera complexe, et moins il aura de ressources à allouer à une tâche secondaire consistant à appuyer rapidement sur un bouton à l'audition d'un signal sonore. C'est pourquoi un allongement de son temps de réaction sera interprété comme une augmentation de la charge mentale. Dans le cas d'une tâche triple, après chaque appui, le sujet se livre à une rétrospection dirigée, ce qui consiste à dire ce qu'il était en train de faire au moment du signal sonore. Des travaux sur l'écriture montrent qu'en dépit du caractère intrusif de cette méthode, la tâche principale n'est pas affectée (Piolat, Kellogg, & Farioli, 2001; Piolat & Olive, 2000; Piolat, Roussey, Olive, & Farioli, 1996). On ne sait pas si cela reste valable pour un test utilisateur. Un outil logiciel, SCRIPTKELL, a été conçu pour faciliter ce genre de protocole (Piolat, Olive, Roussey, Thunin, & Ziegler, 1999). Il a d'ailleurs fait ses preuves dans l'étude d'interfaces, permettant à la fois de catégoriser les actions réalisées, d'en déduire le temps moyen qui leur a été accordé, et leur coût en ressources attentionnelles (Bonnardel & Piolat, 2003).

La rétrospection dirigée n'est pas sans faire penser aux protocoles verbaux consistant à demander aux participants de verbaliser leurs pensées (Jamet, 2006). Si cette méthode est elle aussi intrusive (et par conséquent, susceptible de biaiser le contexte de réalisation), on la différera avantageusement en présentant *a posteriori* au sujet une capture d'écran dynamique

de son travail qui permettra également de procéder à différentes observations *on-line*, sans avoir à le gêner (pour relever des erreurs, *etc.*).

1.5.3.3. Les questionnaires subjectifs

Il existe de nombreux questionnaires subjectifs administrables à la fin d'un test (voir Folmer & Bosch, 2003 pour un inventaire). Ils permettent de mesurer des attitudes envers le dispositif et son utilisation. Plusieurs types de mesures sont possibles : satisfaction, intention d'usage, évaluation subjective de critères ergonomique, l'évaluation subjective des attributs d'utilisabilité ou charge de travail… Nous présenterons deux d'entre eux. Très souvent utilisés et cités, ils sont représentatifs de ce qui se fait en la matière. Créer un questionnaire est difficile et requiert des compétences particulières. C'est pourquoi Dumas (2003) recommande de n'utiliser que des questionnaires validés par des professionnels. C'est le cas de ceux-ci. Simples d'utilisation et d'analyse, ils sont efficaces et facilement disponibles. Par contre, ce ne sont que des outils de mesure. Ils ne sont pas sous-tendus par des modèles théoriques qui vont mettre en relation l'utilisabilité du produit avec son utilité et son acceptabilité.

Le test SUS (*System Usability Scale*, échelle d'utilisabilité des systèmes) (Brookes, 1996). Dix items fortement corrélés entre eux y sont présentés. On y répond sur des échelles ordinales en 5 ou 7 points (au choix). Ils ont fait l'objet d'une sélection sur un ensemble de départ de 50 items, à la suite de l'utilisation de deux logiciels ; l'un réputé facile, l'autre difficile. Les items sélectionnés concentraient sur l'une des valeurs de leur échelle au moins 50% des réponses. L'orientation des questions et des échelles est variée d'une question à l'autre, afin que les utilisateurs ne puissent pas répondre de façon « mécanique ». Les items ne sont pas regroupés en facteurs, mais ils recouvrent chacun différents aspects de l'utilisabilité du système, tels que le besoin d'aide, d'entraînement, la complexité. La validité est élevée. Le score au SUS est global. Les scores aux items seuls ne sont pas valables selon Brookes. Il propose un calcul pour unifier les scores de tous les items dans une mesure comprise entre 0 et 100.

Le NASA-TLX (*Task Load indeX*, Indice de charge de tâche) (Hart & Staveland, 1988). Il s'agit d'une échelle multidimensionnelle dont le but est de mesurer la charge de tâche subjective. Ses auteurs la définissent comme « le coût induit par la réalisation d'une tâche à un certain niveau de performance par un opérateur humain » (notre traduction). Cette

définition est orientée sur le point de vue de l'utilisateur. Tout est complètement subjectif. Ce questionnaire a fait l'objet de trois années de développement, et a été testé à plusieurs reprises, dans différents contextes. La charge de tâche globale qu'il mesure est le fruit d'un calcul impliquant six dimensions :

- **Complexité** : Mesure la difficulté à réaliser le travail.
- **Fatigue physique** : Mesure l'investissement physique nécessaire à l'accomplissement de la tâche.
- **Pression temporelle** : Mesure en quoi le temps imparti pour réaliser la tâche est suffisant ou pas.
- **Frustration** : Mesure le découragement, l'irritation, le stress…
- **Performance** : Mesure la satisfaction de l'utilisateur quant à l'accomplissement de la tâche.
- **Effort** : Mesure l'effort mental et physique demandé par l'exercice.

Les participants évaluent chacune des dimensions sur une échelle de 0 à 20. Des comparaisons par paire entre les dimensions permettent de leur attribuer un coefficient qui sert à calculer le score final, compris entre 0 et 20. C'est dans les 15 à 30 minutes qui suivent l'épreuve que le test doit être administré, sur les lieux mêmes de l'épreuve. Un délai trop important engendre des différences notables d'appréciation (Moroney, Biers, & Eggemier, 1995). Il existe d'autres échelles du même type, parmi lesquelles le SWAT (*Subjective Workload Assessment Technique*) (Reid, Nygren, Hancock, & Meshkati, 1998) ou le *Workload Profile* (Tsang & Velasquez, 1996). Ces trois questionnaires ont fait l'objet d'une comparaison (Rubio, Diaz, Martin, & Puente, 2004). Le *Workload Profile* apparaît comme l'outil le plus sensible et le plus valide. Mais la question de la validité de ces questionnaires reste problématique en l'absence d'une définition consensuelle de ce qu'est la charge de tâche (Tricot & Chanquoy, 1996). Elle l'est d'autant plus que ces outils supposent que la personne soit capable de s'auto-évaluer et attribuent aux items le sens que leur auteur leur a donné (Jamet, 2006). Quoiqu'il en soit, le NASA-TLX a été mis à contribution plusieurs fois dans l'étude de la saisie de texte sur PDA avec un certain succès (Költringer & Grechenig, 2004; Segall, Doolen, & Porter, 2005).

En dépit de leur qualité, on reprochera à ces questionnaires de ne pas être étayés par un modèle de l'utilisabilité qui la mette en relation avec l'utilité et l'acceptabilité du système.

1.5.4. Limites des méthodes d'évaluation de l'utilisabilité

Le principal atout des méthodes d'inspection ergonomique est de pouvoir être facilement intégrées au processus de conception puisque l'utilisateur futur n'y est pas impliqué. Néanmoins, ces méthodes ne permettent pas d'évaluer l'utilisabilité d'une interface (Bevan & Macleod, 1994). Au mieux permettent-elles d'évaluer la qualité ergonomique d'une interface, qualité qui est un préalable à l'utilisabilité de cette dernière, en détectant de façon précoce de potentiels problèmes d'utilisabilité. Si Keinonen (1998) et Bevan (2001) accordent à l'utilisabilité de multiples dimensions, nous considérons pour notre part qu'elle se limite à la qualité de l'interaction entre une interface donnée et son utilisateur en contexte (Bevan, 1995; Cockton, Lavery, & Woolrych, 2003). On appréciera sa valeur par un ensemble de mesures (objectives ou subjectives) de performances et de satisfaction qui pourront témoigner des propriétés d'un dispositif et de la qualité de son élaboration. A cet égard, la qualité ergonomique d'un système d'information est une condition nécessaire, mais pas suffisante, à son utilisabilité. Distinguer qualité ergonomique et utilisabilité trouve écho chez Folmer, Van Gurp, et Bosch (2003) qui se rangent également à cette idée en proposant un cadre conceptuel de relation entre utilisabilité et architecture logicielle. Ainsi, ils mettent en relation quelles recommandations vont avoir un impact sur tel attribut de l'utilisabilité (charge induite, gestion des erreurs, *etc*). De la même façon, Koohang (2004) introduit les principes de propriétés d'utilisabilité (du produit ou système) et d'attributs d'utilisabilité (qualité de l'interaction), quand d'autres parlent d'utilisabilité orientée-produit ou orientée-interaction (Rauterberg, 1996). Cette idée n'est donc pas neuve (pourtant déjà présente chez Nielsen, 1993, ce que conteste Bevan (1995)), mais elle clarifie en partie le flou qui entoure les acceptions de l'utilisabilité. Cependant, cette distinction est avant tout un problème de vocabulaire, puisque Folmer et Bosch (2003) évoquent eux-mêmes ces méthodes d'évaluation des interfaces en tant qu'« inspection d'utilisabilité » (*usability inspection*) comme le font Cockton, Lavery et Woolrych (2003) sans pour autant afficher la moindre ambiguïté sur le fond. Brangier et Barcenilla (2003), de leur côté, préfèrent éluder la question d'une définition stricte de l'utilisabilité, évoquant « un débat fécond et inachevé », pour se ranger au point de vue

suivant. Il consiste à considérer l'utilisabilité comme un attribut mesurable des IHM, relatif à l'utilisation effective du dispositif. Au-delà de la qualité de l'interface (identifiable par les méthodes citées précédemment), c'est dans ses relations aux concepts d'utilité et d'acceptabilité que l'utilisabilité va prendre tout son sens. Cela implique de mettre l'utilisateur en contact direct avec l'IHM à évaluer, quand bien même son développement ne serait pas achevé. C'est le but du test utilisateurs.

Il pose comme difficultés principales la prise en compte du contexte (Cockton, Lavery, & Woolrych, 2003) et le profil de l'utilisateur final (Stewart & Travis, 2003). C'est de surcroît une méthode lourde et coûteuse, et l'on peut s'interroger sur le bénéfice à tirer de cette méthode par rapport à l'inspection ergonomique. C'est un fait, le test utilisateurs n'est pas la panacée. La spécification et la reproduction du contexte d'usage restent assurément difficiles. L'usage en contexte induit des comportements d'utilisation particuliers qui, non reproduits en laboratoire (ex : scénario de test inapproprié ou incomplet), failliraient à faire émerger des problèmes d'utilisabilité (Bevan & Macleod, 1994; Cockton, Lavery, & Woolrych, 2003; Shackel, 1991). Attention cependant à ne pas assimiler conception participative et test utilisateurs (Dumas, 2003) : s'enquérir de l'avis d'un « groupe de mise au point » (*focus group*) ne relève pas du test d'utilisabilité. Seul un test mettant à contribution l'utilisateur final (ou pour le moins, l'idée que le testeur s'en fait par un échantillonnage adéquat) avec le dispositif permet de procéder à des mesures de la qualité de cette interaction et à des observations fiables (ex : observation distante) (Dumas, 2003; Vasalou, Ng, Wiemer-Hastings, & Oshlyansky, 2004). Ces mesures peuvent être objectives (qualitatives et quantitatives) ou subjectives (satisfaction, attitudes, ou perception d'un attribut de l'interface). L'adoption d'un modèle théorique particulier engendrera, de par son opérationnalisation, une sélection de mesures particulières. Leur choix n'est pas anodin. Plus que seulement témoigner d'une performance pour une tâche, elles peuvent rendre compte de mécanismes cognitifs sous-jacents. Ces processus, et les conditions dans lesquelles ils se réalisent permettent seuls de comprendre comment la qualité de l'interaction trouve son explication. Tout l'intérêt est de comprendre comment la configuration de l'interface va faciliter ou entraver l'exécution des processus cognitifs utiles à la réalisation de la tâche. PDA et Smartphones impliquant de la saisie de texte en mobilité, ils s'apparentent donc à des dispositifs électroniques de prise de notes. L'étude scientifique de la prise de note s'inscrit dans le cadre plus large de celui de la rédaction de texte. Elle est étudiée selon deux orientations majeures: l'une se focalise sur la

rédaction, tandis que l'autre traite de la graphomotricité. Cependant, chacune lie la production de texte à des registres mnésiques, et plus particulièrement, à la mémoire de travail (MdT). Les deux approches sont complémentaires pour pouvoir saisir les déterminants psychologiques de l'écriture sur PDA.

2- Ecriture et mémoire de travail (MdT)

2.1. L'écriture

2.1.1. L'écriture comme objet d'étude

Dans l'histoire de l'humanité, l'écriture n'a été inventée que récemment, aux alentours de la moitié du IVème millénaire avant notre ère. Ce constat de l'apparition tardive de l'écriture, autant sur les plans phylogénétique qu'ontogénétique atteste de la spécificité du langage écrit, par opposition au langage oral, et il y a fort à parier qu'il n'existe pas de programmation génétique du système nerveux central pour l'écriture. Une des illustrations en est la faible proportion de langues écrites par rapport aux milliers de langues orales existantes (Bonin, 2003). D'autre part, oralité et écriture font fondamentalement l'objet d'un apprentissage différent : implicite pour la parole (par immersion linguistique), explicite et spécialisé (notamment en milieu scolaire) pour l'écriture (Zesiger, 1995). Pour autant, un certain nombre de structures sont communes aux langages oral et écrit, de sorte que ces deux modes de production ne témoignent que d'une indépendance « relative », pour reprendre le terme de Bonin, Pacton et Fayol (2001). Cependant, quelles que soient les structures mnésiques impliquées dans l'écriture et le discours oral, et indépendamment des processus que ces activités sollicitent, l'écriture reste une activité plus coûteuse en termes de charge mentale que ne l'est l'expression orale. Dans une étude sur la production de discours ou de texte à partir d'une liste de mots, il s'est avéré que le mode de production écrit engendrait une plus grande difficulté à élaborer des idées à partir d'une liste de mots sémantiquement hétérogènes. Bien que ces résultats soient interprétés en termes de capacité d'activation de la MdT, ils témoignent d'une différence fondamentale entre parole et écriture : on ne saurait les différencier sur le seul mode de production (Bourdin & Fayol, 2002).

L'écriture assure deux fonctions principales : la première est de communication, et la seconde est d'être une « mémoire externe », comme le veut l'adage selon lequel « les écrits restent ». Ecrire sur un PDA, de quelque façon que ce soit, s'inscrit donc dans ce cadre. Cela témoigne aussi de la formidable évolution des méthodes de saisie, et des nouvelles orientations de recherche qu'imposent les nouvelles technologies (Favart & Olive, 2005).

Si l'écriture peut prendre différentes formes (logographique ou phonographique), et entretenir de multiples relations avec sa langue orale (pure, morpho-phonique, morpho-

syllabique ou morpho-consonantique), notre intérêt se portera exclusivement sur l'écriture manuscrite alphabétique française avec ses signes diacritiques.

Zesiger (1995) distingue trois niveaux d'analyse dans l'étude des productions écrites manuscrites :

- **La production de textes** : impliquant tout un ensemble de structures qui ne sont pas spécifiques au langage écrit (MdT, mémoire à long terme (MLT), mémoire épisodique, processus syntaxique, *etc.*), la rédaction de texte organise le récit en fonction de contraintes de but (que veut-on écrire, à qui, pourquoi, *etc.*).
- **La production de mots** : elle désigne les processus qui permettent l'accès au lexique orthographique et sa retranscription, ainsi que les processus de conversion des sons en lettres.
- **La production de lettres** : recouvrant les processus de plus bas niveau (jusqu'à périphérique, c'est-à-dire au niveau de l'organe effecteur), la production de lettres renvoie aux processus cognitifs qui permettent l'exécution motrice du geste inhérent à leur réalisation graphique.

2.1.2. Rédaction et prise de notes (PdN)

2.1.2.1. Considérations générales sur la prise de notes

Définir la PdN n'est pas simple. Piolat (2004b) reconnaît d'ailleurs qu'il est plus facile de l'envisager en tant que produit fini qu'en tant qu'activité. Il est vrai que les méthodes de PdN peuvent largement varier suivant les individus, les objectifs poursuivis, ou encore les situations. Très souvent, l'étude de la PdN se situe dans le contexte scolaire ou universitaire. A ce titre, elle y est souvent étudiée comme une activité induisant un apprentissage incident du contenu noté (Roussey & Piolat, 2003). Pourtant, les situations dans lesquelles la PdN s'avère utile dépassent très largement ce cadre (Hartley, 2002). La PdN est largement mise à contribution dans les activités professionnelles, voire même, dans le quotidien de tout un chacun (Piolat, 2006).

On peut retenir de la PdN qu'elle constitue une mémoire externe rendant un certain nombre d'informations disponibles en dehors de leur contexte d'émission (Benton, Kiewra,

Whitfill, & Dennison, 1993). Il ne faut cependant pas la réduire à cela : prendre des notes est une activité complexe, bien plus que ne peut l'être la simple copie, dont elle diffère à bien des égards (Piolat, 2004b). En effet, la PdN implique, au cours de la saisie, de la mémorisation, du jugement, de la sélection et de l'organisation d'informations souvent dispensées à un rythme sur lequel le noteur n'a pas de contrôle. Cela n'est pas anodin puisque la cadence d'émission du langage oral est beaucoup plus rapide que ne l'est celle de l'écriture. Par conséquent, les propos d'un conférencier ne peuvent pas être intégralement consignés dans une PdN. Conséquemment à cette pression temporelle, le noteur se doit de construire, à partir de l'analyse microstructurelle du message écrit ou auditif une représentation macrostructurelle qui en consigne seulement la signification globale (Piolat, Roussey, & Barbier, 2003). C'est donc une activité fortement contrainte, et ce d'autant plus que les situations matérielles de PdN sont souvent inconfortables (Piolat, 2004b).

Comme le résume Piolat (2004b; 2006), la PdN est « l'écriture de l'urgence ». Cela suppose que le scripteur soit capable à la fois de résumer (en sélectionnant les informations pertinentes) et de condenser (par l'utilisation de procédés abréviatifs) le message source en vue d'une exploitation ultérieure (Piolat, Roussey, & Barbier, 2003). Bien que cela soit rare, cette exploitation différée peut même être confiée à un tiers, ce qui impose en plus une lisibilité supérieure permettant le partage (Piolat, 2006). Par conséquent, noter c'est « comprendre et rédiger ». « Il s'agit pour le noteur de stocker (par écrit et/ou mentalement) des informations seulement entendues (ou lues), en gérant simultanément des processus de compréhension (accès au contenu et sélection des informations) et des processus rédactionnels (mise en forme de ce qui est transcrit à l'aide de procédés abréviatifs, de raccourcis syntaxiques, de paraphrases d'énoncés, et de mise en forme matérielle de ses notes) » (Piolat, 2004b).

Concrètement, le scripteur en situation de PdN va devoir développer un ensemble de stratégies qui, dans la mise en texte et la reformulation des informations qui lui sont fournies, vont lui permettre de compenser sa lenteur de saisie et de condenser le propos original. Différents « outils » de production écrite peuvent y concourir, notamment l'exploitation raisonnée de l'agencement spatial du texte ainsi que l'utilisation de procédés abréviatifs (ex : troncature des mots) et substitutifs (utilisation de symboles, sigles, *etc.*). Ces solutions, qui permettent d'augmenter artificiellement la cadence de saisie, requièrent, en contrepartie, un

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

effort cognitif important de la part du scripteur (Kobayashi, 2005; Peverly et al., 2007; Piolat, Olive, & Kellogg, 2005).

2.1.2.2. Prise de notes et saisie en mobilité sur PDA

La portabilité des PDA et leur rôle d'agenda font de ces outils des machines principalement dédiées à une certaine forme de PdN visant à organiser son quotidien (Piolat, 2006), voire même, à servir de pense-bête. Différentes applications servent ces objectifs. S'il est vrai que la saisie experte sur clavier est plus rapide que ne l'est l'écriture manuscrite (Curran, Woods, & O Riordan, 2006), cette considération n'est pas valable pour les Pockets PC. Ces derniers sont souvent dépourvus de claviers matériels, et leur émulation logicielle n'autorise pas les mêmes cadences de saisie. Par conséquent, des outils logiciels de reconnaissance de l'écriture manuscrite ont été développés, avec plus ou moins de succès (voir Bouteruche, Deconde, Anquetil, & Jamet, 2005; Commaford, 2004; Curran, Woods, & O Riordan, 2006; Hinckeley, 2003; Jansen & Ayers, 2004; Mackenzie & Chang 1999; MacKenzie, Nonnecke, McQueen, Riddersma & Meltz, 1994; MacKenzie & Soukoreff, 2002; MacKenzie & Zhang, 1997; McQueen, MacKenzie, Nonnecke, Riddersma, & Meltz, 1994; Ortelli, 2003; Sears & Arora, 2002). Ne visant pas explicitement la PdN, ces logiciels ne permettent pas directement l'insertion de signes sémio-graphiques (puces, flèches) ou de procédés substitutifs (idéo-, picto- et topogrammes tels que signes logiques, mathématiques ou lettres grecques), pourtant récurrents (Piolat, 2006). Si bien sûr, la PdN sur Pocket PC peut se faire de façon « naturelle », sur une feuille blanche électronique, la modernisation des outils de communication et les ressources offertes par Internet autorisent de nouvelles formes de PdN (copier/coller, *etc.*) qui sont déjà l'objet de recherches et de développement d'outils spécialement dédiés (Palaigeorgiou, Despotakis, Demetriadis, & Tsoukalas, 2006). On comprend donc l'intérêt de pouvoir prendre ses notes sur PDA par une méthode de saisie compatible avec des logiciels bureautiques, les réseaux sans fil connaissant une croissance exponentielle. Les enjeux présidant à ce type de recherches sont d'autant plus importants que prendre des notes en recherchant des informations sur le *Web* est une des activités cognitives les plus couteuses (plus que noter en écoutant, par exemple) (Piolat, 2004b; Piolat, Olive, & Kellogg, 2005).

2.1.2.3. Processus rédactionnels

La PdN est une activité rédactionnelle particulière, comme nous l'avons signalé précédemment. Par conséquent, son étude s'inscrit dans le cadre plus large de la rédaction de texte (Piolat, Roussey, & Barbier, 2003). Celle-ci fait l'objet d'un modèle de référence dont la notoriété et la longévité lui ont donné une large audience : le modèle de Hayes et Flower (1980). En utilisant la méthode d'analyse des protocoles verbaux (verbalisation à haute voix des activités mentales des scripteurs pendant la rédaction) associée à l'étude minutieuse des productions fournies par les participants, ces auteurs ont isolé trois « processus » engagés dans la rédaction de textes. Le terme même de processus peut être critiqué dans la mesure où demander à des personnes de caractériser et décrire leur activité cognitive suppose d'une part qu'ils en soient capables, et d'autre part qu'ils aient un accès conscient à ces mécanismes (Jamet, 2006; Zesiger, 1995). Sur le deuxième point, il est bien évidemment certain que tous les processus ne sont pas conscients. On préférera alors parler « d'étapes » de la rédaction, plutôt que de processus (Zesiger, 1995). Cette étude a retenu trois étapes majeures qu'elle met en relation avec l'environnement de la tâche (*i.e.* texte déjà écrit, et tâche assignée en termes de thème, de destinataire et de motivation) et la MLT (connaissances déclaratives et procédurales, plans de textes connus, *etc.*).

Tout d'abord, on distingue la *planification*, permettant au scripteur de définir les objectifs de sa production de texte, de mobiliser ses connaissances en mémoire, et d'y faire un premier tri en fonction de leur adéquation à la réalisation du but fixé. La réalisation de ce but passe par l'organisation hiérarchique de sous-buts qui vont servir de jalons à l'activité rédactionnelle. Viendrait ensuite la *traduction*, qui transforme les idées sélectionnées en matériel linguistique. L'unité d'étude en est la phrase, ou des parties de phrases (Hayes & Flower, 1986). Cette étape de la production est relativement automatisée chez le scripteur expert. Pourtant, dans le cadre de la PdN, l'utilisation massive des procédés abréviatifs pour gagner du temps oblige le noteur à rompre avec ses automatismes de conversion phono-orthographique pour s'en créer de nouveaux. Cela peut alourdir considérablement l'élaboration de la mise en texte (Piolat, 2006). Pour finir, la *révision* opérerait pour vérifier et corriger les deux précédentes étapes. Il est à noter que le déroulement de ces étapes n'est pas strictement linéaire, la révision pouvant intervenir à tout moment. Chacune de ces étapes a fait l'objet d'études spécialisées dans le but de mettre à jour les sous-processus qu'elles impliquent (Chanquoy & Alamargot, 2002). A cet égard, ce modèle s'est montré très heuristique. Cela n'a pourtant pas empêché un certain

nombre de critiques adressées à son encontre. Elles concernent principalement ses insuffisances dans la caractérisation des processus engagés dans les étapes de production, et particulièrement celui d'exécution motrice. C'est pourquoi le milieu des années 90 a vu émerger des modèles combinant à la fois processus rédactionnels et MdT. Ce faisant, l'idée que la capacité limitée de la mémoire à court terme pouvait expliquer d'éventuelles surcharges, déjà présente, s'est vu développée par une étude minutieuse du coût des processus associés à la MdT. Cette nouvelle orientation dans l'étude des processus rédactionnels a fait écho aux propositions de Kiewra qui la préconisait dans son modèle d'étude de la PdN dès 1988.

2.2. La mémoire de travail

2.2.1. Le modèle de mémoire de travail componentiel de Baddeley (2000)

Le concept de mémoire de travail trouve son origine dans celui de mémoire à court terme (MCT) (Gaonac'h & Larigauderie, 2000). La MCT désigne une structure mnésique à capacité limitée, dédiée au stockage temporaire des informations.

De longue date, des esprits éclairés ont remarqué que la conscience ne pouvait instantanément embrasser qu'une quantité limitée d'informations (Lieury, 1997; Miyake & Shah, 1999). Ce qui n'a été pendant des années qu'une intuition s'est vu confirmé en 1956 par Miller, en développant l'idée du « chiffre magique 7 ± 2 ». Dans cet article, synthèse de plusieurs travaux, Miller constate que l'empan de mémoire immédiate des adultes se situe autour de sept éléments qu'il appelle *chunks*. Cette quantité d'éléments que l'on peut simultanément appréhender en mémoire est relativement indépendante du type et de la taille du matériel mémorisé (lettres, chiffres, phrases, images). La réalité est en fait plus complexe, suivant le type de tâche (catégorisation immédiate, mémorisation immédiate ou perception immédiate) ou le type d'items utilisés (sons, images, mots, visages, goûts), comme le fait remarquer Baddeley (1994). Par ailleurs, Gaonac'h et Larigauderie (2000) signalent que le travail de Miller a pu faire l'objet d'une lecture approximative, dont la teneur a pu être négligée au profit du seul « chiffre magique 7 ». Des travaux postérieurs sur la question ont montré que différents facteurs pouvaient affecter la taille de l'empan de MCT (l'augmentant

ou le diminuant), de sorte que certains auteurs défendent aujourd'hui l'idée que le « chiffre magique » serait vraisemblablement plus proche de 4 ± 1 que de 7 (Cowan, 2001). D'autres vont même, par provocation, jusqu'à parler de « légende urbaine » autour de ce chiffre 7 (Jones, 2002). Quel que soit le volume de la MCT, la communauté scientifique s'accorde à dire qu'elle a une capacité limitée. Ce caractère restreint de la MCT s'exprime aussi dans la durée, puisque les éléments qui y sont contenus ont tendance à disparaître rapidement. Sans rappel, elles n'y persistent pas plus d'une vingtaine de secondes (Peterson & Peterson, 1959).

La MdT est « présentée comme un système hiérarchisé en plusieurs sous-systèmes articulés » (Gaonac'h & Larigauderie, 2000). Composée d'un administrateur central (*central executive*, AC) chargé du contrôle attentionnel, ce dernier supervise deux sous-systèmes dits esclaves, dévolus au stockage des informations en fonction de leur nature. L'existence de ces deux sous-systèmes a été postulée par Baddeley sur la base de données empiriques laissant supposer une telle partition. A la charge de la boucle phonologique (*articulatory loop*, BP), le stockage à court terme des informations verbales, et à celle du calepin visuo-spatial (*visuo-spatial scratch-pad*, CVS), celui des informations visuelles et spatiales. Cette organisation est largement admise (Martein, Kemps, & Vandierendonck, 1999). Cependant, l'auteur indique que, s'appuyant sur les résultats expérimentaux portés à sa connaissance, l'inventaire de ces sous-systèmes n'est pas clos (Baddeley, 1986). Par la suite, il ajoutera à son modèle un troisième système esclave qu'il nommera *buffer* épisodique ou mémoire tampon épisodique (*back-up store* ou *episodic buffer,* BE) (2000) (voir figure 14) .

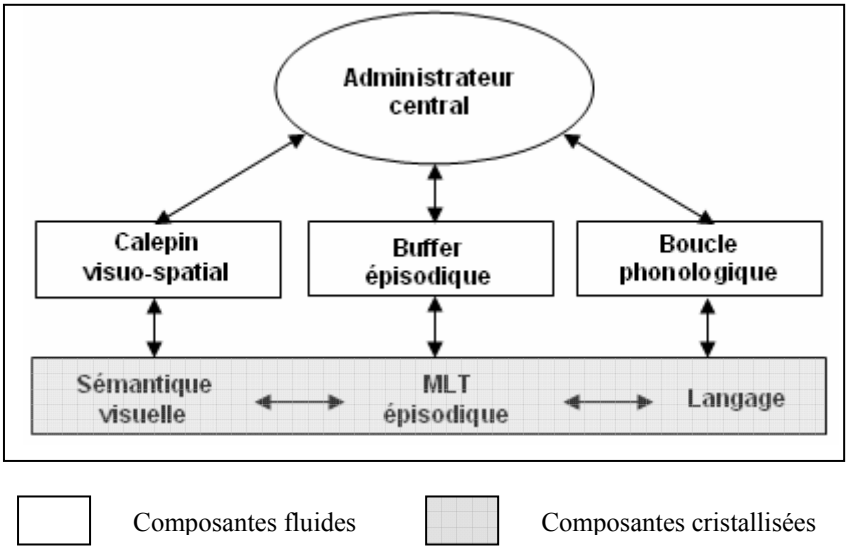


Figure 14: Modèle multipartite de la mémoire de travail (d'après Baddeley, 2000, notre traduction)

Baddeley (2000) définit la mémoire de travail comme « un système à capacité limitée organisant le stockage temporaire et la manipulation d'informations nécessaires à des tâches complexes telles que la compréhension, l'apprentissage et le raisonnement » (notre traduction). C'est donc bien une unité composite spécialisée dans le stockage à court terme et le traitement. Bien que la spécialisation des systèmes esclaves soit dépendante de la nature des informations à traiter, ce serait une erreur de ne l'envisager que comme une « porte d'entrée » (*gateway*) sensorielle, comme certaines publications ont pu le laisser entendre : il s'agit bien d'un espace de traitement cognitif. D'ailleurs, les données recueillies auprès de patients présentant des déficits de MdT n'indiquent pas de troubles perceptifs associés (Della Sala & Logie, 2002). La MdT se situe en fait, en tant qu'espace de stockage et de traitement temporaires, à l'interface entre les mémoires sensorielles et la mémoire à long terme. Siège de l'activité de pensée en tant que conscience orientée, les résultats des traitements qu'elle opère peuvent y être conservés (temporairement et en vue d'autres opérations), placés en MLT ou encore être implémentés dans un programme de commande motrice pour susciter un comportement moteur (Logie, Engelkamp, Dehn, & Rudkin, 2000).

2.2.1.1. La boucle phonologique (BP)

Plus étudié des systèmes esclaves, il en est le mieux connu, bénéficiant du caractère langagier que l'on attribuait au départ à la MCT (Baddeley, 1993). Un grand nombre de données expérimentales a été collecté sur elle. Par ailleurs, Baddeley la considère comme le plus simple des sous-systèmes (1993). Elle est impliquée dans le stockage temporaire des informations auditives et verbales. Les informations qui lui parviennent peuvent provenir de la mémoire sensorielle auditive ou échoïque (dont la durée n'excède pas quelques secondes) ou bien être issues d'un recodage graphophonologique dans le cas de la lecture. Lorsque la BP contribue à une tâche de stockage et de traitement (sous l'égide de l'administrateur central), on peut alors parler de MdT verbale ou phonologique (Baddeley, 1986). La BP s'organise autour de deux composantes :

- **Une composante passive de stockage des informations auditives (*phonological store*)** : Elle stocke à court terme les informations auditives. Les informations qui transitent par cette structure ont tendance à disparaître au bout de deux secondes (Baddeley, Chincotta, & Adlam, 2001).

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

- **Un système actif d'autorépétition et de recodage phonologique (*articulatory rehearsal*)** : Ce système agit en rafraîchissant les informations auditives présentes dans la composante de stockage par subvocalisation (Hardyk & Petrinovich, 1970, cités par Baddeley, 1986).

La BP est donc un sous-système à capacité limitée, autant pour le stockage passif (qui peut alors impliquer l'administrateur central en cas de surcharge) que pour l'autorépétition (la totalité des items à retenir devant être idéalement subvocalisable dans un délai de deux secondes, durée de maintien passif avant oubli).

S'agissant du rôle de la BP, on admet qu'elle est impliquée dans des activités complexes telles que l'apprentissage de la lecture, la compréhension du langage, l'apprentissage phonologique à long terme (acquisition d'une langue seconde par exemple) ou encore l'acquisition de vocabulaire (Baddeley, Gathercole, & Papagno, 1998). Des travaux récents ont montré qu'elle est aussi engagée dans le soutien de fonctions exécutives comme la commutation de l'attention (Baddeley, Chincotta, & Adlam, 2001). Bien évidemment, elle participe à la production écrite (Kellogg, 1996).

2.2.1.2. Le calepin visuo-spatial (CVS)

Défini comme étant le sous-système en charge « du maintien temporaire et de la manipulation des informations visuo-spatiales, il joue un rôle important dans l'orientation spatiale et dans la résolution de problèmes visuo-spatiaux » (Baddeley, 2002a, notre traduction). Il permet de combiner les informations visuelles et spatiales, ce qui recouvre un spectre plus large de modalités : motrices, tactiles ou haptiques (Baddeley, 2002a). Si les premiers travaux sur le CVS ont laissé penser que son fonctionnement était relativement indépendant de celui de l'administrateur central, des recherches plus récentes montrent qu'il n'en est rien (Logie & Marchetti, 1991; Rudkin, Pearson, & Logie, 2007). La tendance actuelle va désormais plutôt dans le sens d'une caractérisation fine des composantes du CVS (au même titre que pour la boucle phonologique), étant entendu que ce dernier traite et stocke des informations qui ne sont fondamentalement pas de même nature : elles peuvent être visuelles et/ou spatiales (Baeyens & Bruyer, 1999; Logie & Marchetti, 1991; Logie & Pearson, 1997; Rudkin, Pearson, & Logie, 2007). Les dernières études conduites sur le CVS y montrent une forte implication des mécanismes de contrôle exécutif, se caractérisant par des

corrélations élevées entre des mesures relevant respectivement du CVS et de l'AC (Darling, Della Sala, & Logie, 2007; Rudkin, Pearson, & Logie, 2007; Vecchi & Richardson, 2000).

Lorsque le CVS est mis à contribution dans une tâche de stockage et de traitement (sous l'égide de l'administrateur central), on peut alors parler de MdT visuo-spatiale (Baddeley, 1986). Tout comme la boucle phonologique, il s'organise autour de deux composantes (Darling, Della Sala, & Logie, 2007; Darling, Della Sala, Logie, & Cantagallo, 2006) :

- **Un système passif de stockage des informations visuo-spatiales (*visual cache*)** : Il permet d'encoder les patterns visuels de type forme ou couleur. Il traite aussi les tâches passives de stockage.

- **Un système d'autorépétition visuelle (*inner scribe*) :** Son rôle consiste à stocker des patterns spatiaux (séquences de mouvement), générer et transformer des images mentales, organiser des séquences de mouvements.

Bien que la connaissance de ce sous-système ait très largement bénéficié des recherches de Logie et de ses collaborateurs (Darling, Della Sala, & Logie, 2007; Logie & Marchetti, 1991; Rudkin, Pearson, & Logie, 2007), le CVS reste de toute évidence un système esclave dont la description n'est pas complète, et il conviendra d'en étudier de façon plus approfondie l'influence dans les processus rédactionnels (Olive & Piolat, 2005).

2.2.1.3. Le buffer épisodique (BE)

Un certain nombre de phénomènes expérimentaux observés n'ont pas trouvé d'explication dans le modèle de MdT tel que formulé en 1974 (Baddeley & Hitch) puis 1986 (Baddeley), le rendant l'objet de critiques. Elles ont notamment concerné les effets relativement faibles de la suppression articulatoire sur la présentation visuelle de chiffres. De la même façon, des effets de similarité visuelle sur des items verbaux ont donné à penser que les informations visuelles et phonologiques étaient vraisemblablement combinées. L'absence, au sein du modèle de MdT de 1986, d'une structure permettant à la fois « le rappel sériel, l'intégration visuelle et phonologique, et sans doute, d'autres types d'informations » était évidente (Baddeley, 2000, notre traduction). Ces limites ont pris d'autant plus d'importance que la « concurrence » des modèles d'activation se faisait forte pour expliquer « l'élasticité » des *chunks*. Par exemple, une phrase de cinq mots prend autant de place en mémoire qu'un seul mot hors contexte

(Baddeley, 2000). Faisant l'hypothèse qu'une partie de la MLT peut être activée dans un domaine d'expertise donné pour « augmenter » le volume de la MdT, ce type de modèles peut expliquer ces résultats (Ericsson & Kintsch, 1995). Bien que Baddeley ait pu dans un premier temps souscrire par défaut à cette analyse, il s'en est progressivement éloigné. Cette portion activable de MLT ne rendait pas compte, selon lui, d'un espace de travail dans lequel des opérations de transformation pouvaient être expliquées.

Pour combler cette lacune, Baddeley propose en 2000 d'ajouter un nouveau sous-système à son modèle de mémoire de travail : le *buffer* ou mémoire tampon épisodique. Conçu comme « un système de stockage temporaire à capacité limitée capable d'intégrer des informations issues de sources différentes », « il se place sous le contrôle de l'administrateur central, qui peut y récupérer des informations pour les rendre accessibles à la conscience, et, le cas échéant, les manipuler et les modifier » (notre traduction). Si Baddeley le compare au concept de mémoire épisodique développé par Tulving (1989), il insiste toutefois sur son caractère temporaire, sa capacité limitée, et son implication dans la recherche d'informations épisodiques, c'est-à-dire inhérentes au contexte d'encodage en MLT d'autres informations. A cet égard, il serait tout particulièrement impliqué dans la récupération des connaissances et dans la manipulation d'informations multimodales, puisque situé à l'interface des multiples sources d'informations.

Les implications de cette découverte sont potentiellement très importantes, car elle réinvente le rôle de la MdT dans la faculté de conscience. Cette approche de la mémoire, telle que s'en réclame Tulving, se veut plus large, plus complète, permettant la nuance entre le rappel et le souvenir, trop souvent considérés comme procédant d'un même mécanisme (Tulving, 1989). Par ailleurs, il permet de résoudre le doute exprimé par certains quant à la possibilité pour l'administrateur central de disposer d'un réservoir mnésique propre (Baddeley & Hitch, 1974). Le caractère récent de ce sous-système fait de lui le moins connu et le moins étudié de la MdT. C'est pourquoi on ne mesure pas encore l'étendue de sa contribution aux différentes activités mentales, et encore moins à l'écriture.

Pourtant, plusieurs auteurs (Chanquoy & Alamargot, 2002; Piolat, 2004a) font remarquer que la capacité de ce BE à manipuler des informations de natures diverses pourrait vraisemblablement lui permettre de jouer le rôle que Hayes attribue au registre sémantique de MdT, ajouté à son modèle de production écrite en 1996. Ce registre est dédié au maintien d'une information sémantique préverbale pendant l'élaboration syntaxique et lexicale du texte

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

dans la BP. Cette piste intéressante reste à étudier, comme, d'une manière générale, l'implication du BE dans les processus rédactionnels (Olive & Piolat, 2005).

2.2.1.4. L'administrateur central (AC)

« Dépotoir ou superviseur », l'AC a longtemps été négligé au point que Baddeley le présente en ces termes (1986). C'est pourtant l'attribut essentiel de la MdT, par opposition à la MCT, qui n'est qu'un espace de stockage temporaire et n'engage pas le traitement des informations. Le consensus voulait qu'il soit un « réservoir à capacité limitée de ressources pour les processus » (Baddeley, 2002a, notre traduction). L'AC a longtemps servi à recueillir (et occulter) des processus difficiles à étudier, comme l'intégration d'informations issues des systèmes esclave (BP et CVS) ou bien l'idée d'une MdT engagée dans la faculté de conscience (Baddeley, 1993). Cette impossibilité de rendre à la fois compte d'un éventuel espace de travail doublé d'un siège de la conscience bénéficie désormais de l'ajout du *buffer* épisodique dont nous venons de faire état (Baddeley, 2000). Par la suite, l'AC a fait l'objet de travaux permettant de clarifier sa structure, son fonctionnement, et ses attributions (Gaonac'h & Larigauderie, 2000).

On ne sait pas précisément si l'administrateur existe en tant que tel, auquel cas ses fonctions en seraient des émanations (approche cognitive), ou bien si il n'est qu'un « label » pour des fonctions indépendantes appelées fonctions exécutives (approche neuropsychologique). Les connaissances actuelles tendraient à faire pencher vers cette seconde hypothèse (Baddeley, 1996). L'ensemble de ces fonctions exécutives constitue un ensemble hétérogène de processus de haut niveau dont on localise le substrat organique au niveau des aires frontales et préfrontales, bien que d'autres soient également impliquées. C'est pourquoi Baddeley préconise de ne pas définir l'AC par son implémentation anatomique (1990, 1996). D'autres travaux confirment ce point de vue. Les épreuves de MdT peuvent activer de très larges aires corticales ou sous-corticales, qui s'étendent bien au-delà des aires frontales. La charge induite par ce type de tâches joue aussi un rôle dans la distribution organique de l'activation. D'autre part, des patients cérébro-lésés témoignent, après recouvrement de certaines fonctions exécutives déficitaires, d'activation d'aires adjacentes, voire controlatérales aux aires atteintes (Carpenter, Just, & Reichle, 2000).

L'AC a été, au départ, envisagé par Baddeley (1986) comme une instance de contrôle des activités mentales fonctionnant sur le mode inhibition/planification. Inhibition des processus inappropriés ou automatiques, et planification des actions en regard des objectifs fixés. Ces considérations ne sont pas anodines pour l'ergonome, comme le fait remarquer Baddeley dans un exemple (1993). Une machine facile à utiliser est une machine dont la compréhension du fonctionnement se fait sur un mode automatique, sans que l'on ait à fournir d'effort pour en deviner le principe. Il illustre cela avec la répartition des potentiomètres des plaques de cuisson qui ne présente jamais aucune analogie avec la position de celles-ci, ce qui les rend difficiles à utiliser.

Cette première approche de l'AC a permis de le développer au minimum, en l'absence d'études qui lui soient véritablement consacrées. Par la suite, des travaux se sont concentrés sur une description plus fine des processus soutenus par ce sous-système, que l'on considère être les fonctions exécutives en tant que telles. La production de texte relève explicitement des ces fonctions exécutives.

2.2.1.5. Critique du modèle de mémoire de travail de Baddeley

En dépit de son importante valeur heuristique, le modèle de MdT de Baddeley (2000) n'est pas exempt de critiques. Gaonac'h et Larigauderie (2000) proposent une lecture synthétique de ces dernières.

On relèvera tout d'abord que, pour faire face aux résultats inexpliqués, le modèle peut être facilement modifié par l'ajout d'un module adapté. L'histoire montre d'ailleurs qu'il l'a été, avec l'incorporation du BE au modèle de MdT (Baddeley, 2000). C'est afin d'éviter de subir ces mêmes écueils que les modèles d'activation opposent une vision unitaire, et non modulaire, de la cognition (Anderson, 1996; Cowan, 1988, Ericsson & Kintsch, 1995; Just & Carpenter, 1992).

Autre point critique, le contrôle exécutif. Qu'il s'agisse d'un administrateur central ou d'un ensemble de fonctions exécutives permettant la réalisation d'opérations intellectuelles (comme la compréhension (Just & Carpenter, 1992) ou l'écriture (Kellogg, 1996)), on peut presque parler « d'Arlésienne » jusqu'à une époque très récente. Le choix fait par Baddeley de remplacer cette Arlésienne par « l'Homoncule », sorte de conscience dans la conscience, n'a en rien clarifié ces lacunes (Miyake & Shah, 1999).

Il arrive également que certaines limites du modèle ne lui soient pas propres, et puissent concerner n'importe quel modèle de MdT. C'est le cas par exemple de la caractérisation de processus dits « automatiques », dont la définition peut se baser sur des critères variables d'un auteur à l'autre (Perruchet, 1988). Ceci est particulièrement problématique pour la caractérisation de certains processus d'acquisition de compétences pour lesquels la MdT est fortement mise à contribution, notamment en termes d'apprentissage moteur (Mounoud, 1986). Or, l'écriture est par essence une activité motrice.

2.3. Mémoire de travail et rédaction de textes

2.3.1. Le modèle de Hayes (1996)

L'ajout de la MdT de Baddeley (1986) au modèle de Hayes et Flower (1980) dans la révision qu'en propose Hayes en 1996 donne lieu à un modèle d'une rare exhaustivité (voir figure 15). Elaboré à partir de résultats expérimentaux, il tient compte de nombreux paramètres dépassant largement le cadre de la seule MdT (à laquelle il accorde une place centrale, ce dont rend compte la figure 15), puisqu'il intègre aussi les dimensions conatives et affectives de l'écriture.

Il ne nous revient pas dans ces lignes de discuter de la totalité de ces modifications. L'auteur lui-même admet que son modèle n'est qu'en partie validé au moment de sa publication. Cela reste valable à ce jour, à notre connaissance. En dépit de cela, l'importance qui est accordée à l'environnement physique de la tâche d'écriture y est bien étayée. On ne saurait négliger ce point lorsque l'on s'intéresse à l'écriture sur PDA, situation matérielle particulière s'il en est. Notons aussi que par l'ajout de la MdT au modèle, Hayes veut apporter un élément de réponse à la principale critique visant son précédent modèle (Hayes & Flower, 1980). Cela passe principalement par une plus grande considération des implications visuo-spatiales dans l'écriture, et dont le rôle est primordial pour la PdN (Piolat, 2006). L'incorporation du modèle de Baddeley y est complétée par un registre supplémentaire dédié au traitement de l'information sémantique. Le rôle de ce registre est de maintenir en mémoire une information sémantique préverbale pendant que la boucle phonologique est sollicitée pour apparier ces représentations à des représentations linguistiques (*i.e.* lexicales). Nous avons souligné que cette modification se justifie désormais par le rôle nouveau que Baddeley accorde au *buffer* épisodique (2000).

Si la MLT apparaît également dans cette révision, on remarquera surtout que les étapes de la rédaction de texte initialement prévues ont été fondues dans un corps plus large de processus non exclusivement dédié à l'écriture : la planification devient la réflexion, et la production de texte relève autant du langage écrit qu'oral (Piolat, 2004a).

Ce très ambitieux modèle, n'a, semble-t-il, pas eu la portée et l'audience du précédent. Comme le souligne à juste titre Piolat, un certain nombre de relations postulées par le modèle sont difficiles à étudier expérimentalement, particulièrement ce qui relève des affects et de la motivation, toujours très difficiles à étudier scientifiquement. Pourtant, il synthétise un ensemble cohérent de paramètres à considérer dans une analyse de la production écrite centrée sur le contenu produit plus que sur les processus (exception faite de la révision par la lecture qui est largement développée, mais qui procède au premier chef de la qualité de la production). Dans une revue de question récente, l'accent est mis sur l'interactivité de la production écrite entre, d'une part, les connaissances en MLT, et, d'autre part, l'environnement de la tâche. Si le rôle de la MdT dans la récupération des connaissances en MLT a fait intensivement l'objet de recherches, la situation de rédaction, qui joue un rôle important, reste souvent négligée. La double orientation des flèches (cf. figure 15) entre environnement et individu n'est pas anodine, et la production de texte implique autant des processus ascendants que descendants. Ceci plaide donc en faveur d'une étude de la MdT non plus orientée sur les mécanismes de production en tant que tels, mais sur sa dimension d'interface entre les connaissances (MLT) et les perceptions (environnement). A cet égard, les modèles d'activation, bien que relativement négligés jusqu'ici, sont tout indiqués dans le traitement de cette problématique, puisqu'ils peuvent rendre compte de l'orientation attentionnelle en fonction des connaissances antérieures, de même qu'ils peuvent expliquer l'activation de connaissances en mémoire à partir d'indices environnementaux, notamment la lecture du texte produit, mais aussi la lecture de sources documentaires extérieures (Alamargot, Chanquoy, & Chuy, 2005).

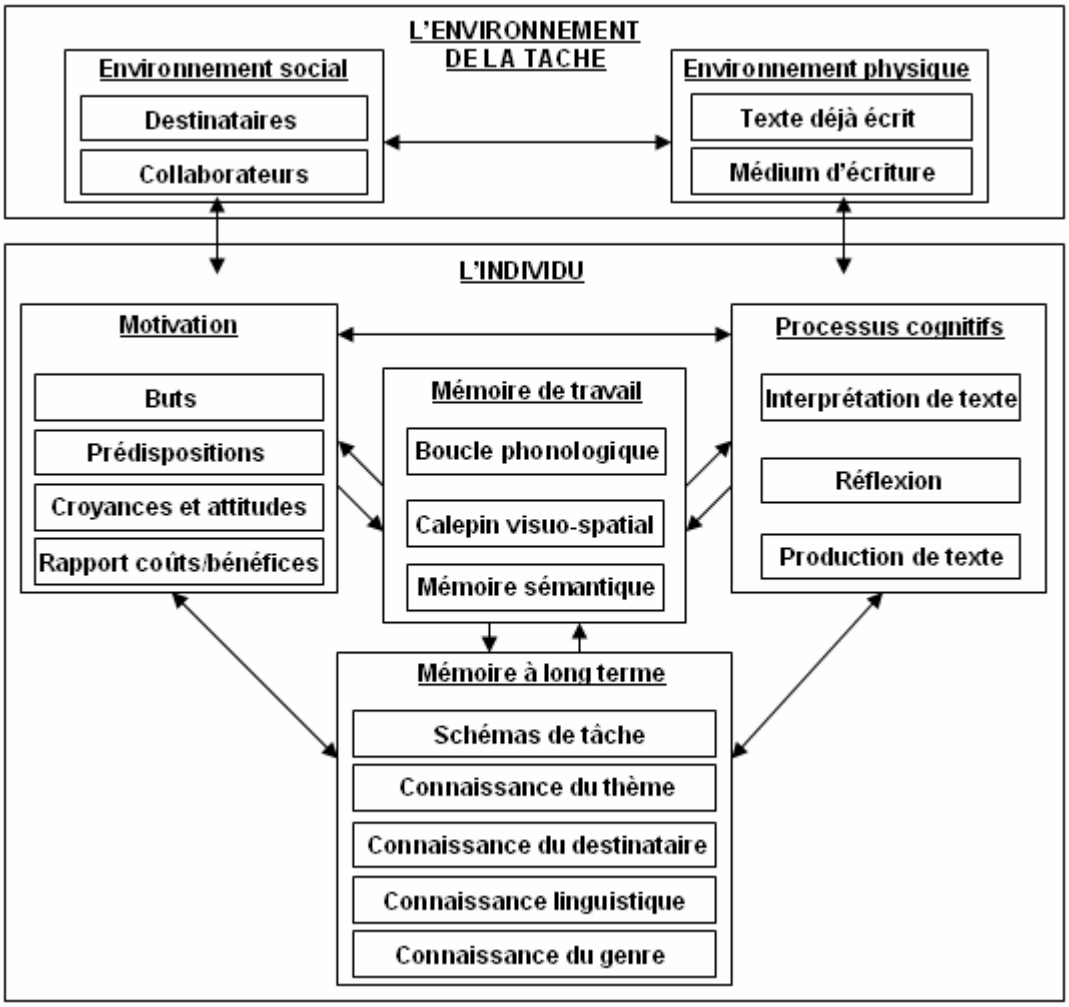


Figure 15: Modèle révisé des processus rédactionnels (Hayes, 1996, d'après Piolat, 2004b)

2.3.2. Le modèle de Kellogg (1996)

Beaucoup plus synthétique que celui de Hayes (1996), le modèle proposé par Kellogg la même année et dans le même ouvrage s'avère cependant plus heuristique et fécond (Piolat, 2004a). En concentrant ses efforts sur la mise en relation des processus rédactionnels avec les sous-systèmes de la MdT de Baddeley, Kellogg développe une théorie circonscrite à l'essentiel, mais en contrepartie très détaillée. Les étapes de la production écrite y sont largement inspirées du modèle de Hayes et Flower (1980), bien que faisant l'objet d'une révision. Elles en diffèrent dans leur appariement, qui se fonde sur leur dépendance à l'un ou l'autre sous-système de la MdT. D'autre part, il ajoute l'étape d'exécution, qui rend compte de la programmation motrice du geste d'écriture, ainsi que de sa mise en œuvre. C'est un

choix que nous jugeons particulièrement intéressant en regard de nos préoccupations en ce qui concerne la saisie de texte manuscrit sur PDA.

2.3.2.1. La procédure de Kellogg

La contribution de Kellogg à l'étude des processus rédactionnels, loin de n'être que théorique, est aussi méthodologique. Il appuie ses travaux sur une procédure de double ou de triple tâche, postulant que la MdT est, en plus d'un système de traitement à capacité limitée, un réservoir de ressources attentionnelles limitées (Olive, 2004; Piolat, Roussey, Olive, & Farioli, 1996). Le principe de la double tâche ne repose pas ici sur la saturation artificielle de l'un ou l'autre des sous-systèmes de MdT, mais sur la concurrence entre processus, et la répartition des ressources attentionnelles dont ils bénéficient. Cela se traduit par la mesure d'un temps de réaction (TR) à une stimulation auditive (bip sonore), devant entrainer une réaction rapide du sujet. Ce TR constitue donc un indice des ressources attentionnelles disponibles au moment du signal sonore : puisque les ressources attentionnelles sont en quantité limitée, le TR dépend des ressources non allouées à la réalisation de la tâche principale. Par conséquent, plus la tâche principale sera coûteuse attentionnellement et plus le TR sera long. C'est par comparaison avec le TR de base du sujet (c'est-à-dire le TR moyen au signal sonore en situation de tâche simple) que l'on peut ensuite déduire le coût des processus qui interviennent de façon concomitante avec la tâche double pendant le test. Il s'agit donc d'une mesure de charge mentale, c'est-à-dire de l'effort mental moyen au cours d'une tâche donnée (Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

Afin de déterminer quel est le processus mobilisé par le sujet au moment auquel le « bip » survient, Kellogg ajoute une troisième tâche qui suit immédiatement la seconde : la rétrospection dirigée. Cela consiste pour le sujet à expliciter (oralement ou par l'appui sur une touche labellisée convenue à l'avance) le processus ou l'étape de production auquel il se livrait à l'instant d'apparition du signal. Un entraînement préalable est bien évidemment nécessaire, ne serait-ce que pour apprendre aux participants à « sonder » leur cognition. Une telle capacité requiert certaines aptitudes métacognitives d'une part, et d'autre part, nous l'avons déjà signalé, que les processus étudiés puissent faire l'objet d'un accès conscient par les sujets. Dans la mesure où les processus évalués sont des « processus » de production écrite, c'est-à-dire une conjonction de plusieurs processus cognitifs, la catégorisation ne pose

pas de difficultés (sauf peut-être chez les sujets jeunes qui peuvent être plus indécis (Piolat, Roussey, Olive, & Farioli, 1996)). Un des principaux arguments en faveur de la rétrospection dirigée est la stabilité du pattern de mobilisation des processus que nous développerons plus loin (Kellogg, 2001a; Piolat, Roussey, Olive, & Farioli, 1996). Une étude a même montré que la rétrospection dirigée pouvait se porter sur plusieurs processus si le sujet estime qu'il en conduisait deux conjointement au moment de l'apparition du signal sonore (Pélissier & Piolat, 1999).

Il existe une autre méthode de triple tâche, basée sur l'analyse des protocoles verbaux en plus du TR à un bip sonore. Cette méthode, qui remplace la rétrospection dirigée par la verbalisation à haute voix des intentions du sujet au cours de la rédaction est plus adaptée à l'étude du déroulement et de l'enchaînement des processus. Par ailleurs, plus complexe à mettre en œuvre (notamment par l'enregistrement des cadences de saisie pour les mettre en correspondance avec les verbalisations), elle est aussi plus intrusive, et donc susceptible d'interférer sur la tâche principale (Piolat & Olive, 2000). L'intérêt de la rétrospection dirigée une fois la tâche seconde de détection terminée est de ne pas s'ajouter à la tâche principale. Elle ne fait que prolonger son interruption. Dans le cas de la procédure de Kellogg, la rétrospection se fait par appui sur des touches désignant chacune un processus ou une étape. En général, il s'agit des étapes de *planification*, *traduction* et *révision*, puisqu'il est admis que l'*exécution* est peu coûteuse. Une dernière touche sert à recueillir les incertitudes ou les pensées « parasites » qui peuvent légitimement, à un moment ou à un autre, dissiper le sujet durant sa composition.

Plusieurs études se sont concentrées sur les questions méthodologiques soulevées par l'utilisation de la procédure de Kellogg afin de déterminer si elle interfère significativement avec la tâche principale, et si oui, sous quelles conditions (Olive, Kellogg, & Piolat, 2002; Piolat, Kellogg, & Farioli, 2001; Piolat & Olive, 2000; Piolat, Roussey, Olive, & Farioli, 1996). Ces travaux ont montré que les paramètres adoptés par Kellogg quant au déroulement de la triple tâche (délai d'apparition des signaux sonores et entraînement à la rétrospection dirigée) sont appropriés. Bien que la rétrospection dirigée n'augmente pas les TR (à condition de ne pas entrainer les sujets à y accorder une trop grande importance), elle affecte cependant la complexité syntaxique des productions (Piolat, Roussey, Olive, & Farioli, 1996). Plus récemment, une étude a fait varier la consigne dans une épreuve de triple tâche, incitant les sujets à privilégier l'une ou l'autre des trois tâches. Elle n'a pas relevé, dans les TR, la

fréquence d'apparition des processus, ou la qualité de la rédaction, de différences significatives. Les auteurs concluent que la tâche principale de rédaction est tellement coûteuse attentionnellement qu'elle mobilise nécessairement le sujet avant tout autre chose. Ils répondent ainsi aux critiques portant sur la question de savoir sur quelle tâche le sujet oriente son attention (Piolat, Kellogg, & Farioli, 2001). La principale limite de ce protocole est de ne pas pouvoir distinguer dans quelle mesure deux processus peuvent être menés simultanément (Piolat, Roussey, Olive, & Farioli, 1996). Pourtant, il apparaît que dans des situations courantes de production écrite, la conduite concomitante de plusieurs processus soit possible (Olive & Kellogg, 2002; Olive & Piolat, 2002; Pélissier & Piolat, 1999). D'autre part, il est évident que les processus rédactionnels reposent eux-mêmes sur des processus cognitifs dont on ne peut pas prendre la mesure. Des labels très larges tels que *planification*, *traduction* ou *révision* deviennent alors « fourre-tout », et les TR qu'on leur prête ne représentent en fait qu'une moyenne des ressources allouées à des processus forts différents. Des travaux critiques conduits sur la description fine des processus rédactionnels le soulignent (Chanquoy & Alamargot, 2002; Favart & Olive, 2005). C'est pourquoi l'analyse des productions est déterminante pour permettre de comprendre quels aspects du discours écrit peuvent être altérés en fonction des exigences de la tâche (Kellogg, 2004; Kellogg, Olive, & Piolat, 2007).

Kellogg a utilisé dans certaines de ces études un système de clé vocale pour détecter au mieux le TR, à l'imprécision près de la réponse orale (réduisant la composante motrice au minimum). D'autres études ne mesurent le temps de réaction que par l'appui sur une touche grâce à la main non dominante (pour ne pas interférer avec celle qui écrit). Quelle que soit la méthode d'enregistrement des TR choisie, elle est toujours comparée au temps de réaction de base, préalablement enregistré. Un outil logiciel a été spécialement conçu pour cela : SCRIPTKELL (Piolat, Olive, Roussey, Thunin, & Ziegler, 1999).

2.3.2.2. Description des processus

En dépit d'une présentation discrétisée de l'enchaînement des étapes de la production écrite, Kellogg (1996) assure que les processus impliqués dans les différentes étapes peuvent se dérouler simultanément, si tant est que leur sollicitation conjointe ne dépasse pas les capacités limitées de la MdT.

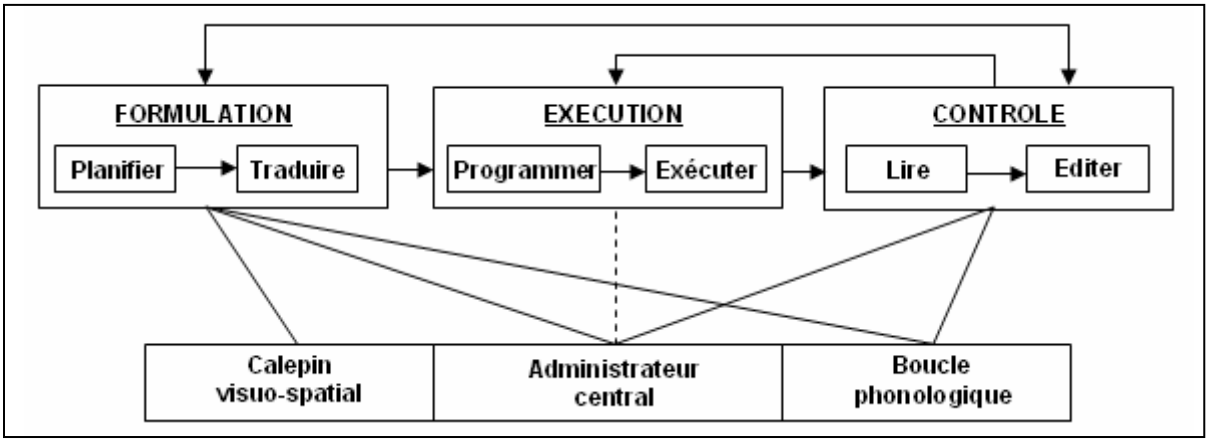


Figure 16: Processus rédactionnels et mémoire de travail (Kellogg, 1996)

Trois étapes principales sont proposées, reposant chacune elles-mêmes sur deux sous-processus. La *formulation* tout d'abord, qui s'appuie sur la *planification* et la *traduction*. Le processus de formulation relève des mécanismes qui permettent la transformation d'idées organisées en mots constituant des phrases. La *planification* organise les objectifs discursifs et les propositions qui peuvent, à ce stade, s'exprimer sous forme de représentations visuelles analogiques (plan, graphique, schéma ou organigramme), auquel cas la *traduction* n'en est pas complète. Cette étape se veut purement conceptuelle. Ce n'est qu'ensuite que la *traduction* va mettre en texte les idées, afin que le sens soit véhiculé par des mots grammaticalement organisés. Cela implique, à partir d'une information purement sémantique, de sélectionner les unités lexicales phonologiques et leur agencement syntaxique, qui, converties par une transcription des phonèmes en graphèmes, va permettre la production motrice du résultat.

L'*exécution* ensuite, va reposer sur la *programmation* motrice et son *exécution* proprement dite. Cette étape peut prendre différentes formes en fonction du type de réponse produite (orale, manuscrite ou dactylographiée). Dans le cas de l'écriture, la *programmation* nécessite de choisir la taille des lettres et leurs allographes. D'autres contraintes sont à gérer à ce niveau, telles que la force à appliquer ou la vitesse du mouvement. Le rétrocontrôle proprioceptif qui est renvoyé par l'exécution motrice est aussi pris en compte. Un courant d'étude de l'écriture se consacre exclusivement à ces aspects. Nous les présenterons à la section consacrée à la graphomotricité. De fait, le modèle de Kellogg étant orienté sur la rédaction de texte, le caractère automatique qu'il accorde à l'exécution motrice de l'écriture tend à lui faire négliger cette dimension de la production écrite.

Pour finir, le *contrôle* s'organise autour de la *lecture* du texte produit (macro et microstructure du texte), ainsi que sur l'*édition*, qui consiste à comparer le texte lu avec le résultat escompté. Le *contrôle* se fait à tous les niveaux (sémantique, lexical, syntaxique, *etc.*) et à n'importe quelle étape, comme l'indiquent les flèches. D'autre part, si le *contrôle* peut être sous la dépendance immédiate de la *formulation* (contrôle de l'adéquation aux objectifs visés), la *formulation* peut elle aussi faire l'objet d'un contrôle avant même sa *traduction* et son *exécution* (cf. double flèche). La complexité des interactions possibles entre ces mécanismes, qui reste encore mal connue, a fait l'objet d'une revue de question (Roussey & Piolat, 2005).

Le modèle de Kellogg ne spécifie pas comment ces processus rédactionnels s'enchaînent les uns aux autres, mais constate, au travers de différentes études, un pattern de mobilisation des processus relativement stable. D'une manière générale, les scripteurs traduisent plus souvent qu'ils ne planifient ou ne révisent. La *planification* est beaucoup utilisée en début de rédaction et moins à la fin. A l'inverse, la révision est plus sollicitée à la fin de la rédaction qu'au début (Kellogg, 2001a; Piolat, Roussey, Olive, & Farioli, 1996). Il est à noter que certains processus peuvent se dérouler simultanément. C'est ce qu'a montré une étude d'Olive et Piolat (2002) sur l'effet de la suppression du *feedback* visuel pendant l'écriture. Par un ingénieux protocole de composition écrite, suivi d'une tâche de copie du texte préalablement produit, les auteurs catégorisent les processus suivant qu'ils soient de bas (*exécution*) ou de haut niveau (*formulation* et *contrôle*). Pendant la composition, les TR enregistrés pendant l'écriture (*i.e.* l'activité motrice d'écriture) relèvent à la fois de la transcription et/ou de la f*ormulation*. Par contre, pendant les pauses, les TR ne concernent que la *formulation*. A l'inverse, pendant la copie, les TR relevés durant les phases d'écriture ne consistent qu'en l'*exécution*, alors que pour les pauses, aucun traitement n'est associé. Les résultats montrent que, si l'*exécution* est affectée par la suppression du *feedback*, les processus de haut niveau sont quant à eux préservés. Par contre, cette suppression engendre des difficultés chez les copieurs / rédacteurs à coordonner plusieurs processus simultanément. En effet, pendant les phases d'écriture en rédaction, avec *feedback*, les TR sont significativement plus élevés que sans *feedback*. Ce résultat donne à penser qu'en rédaction, les scripteurs coordonnent à la fois la *formulation* et la transcription, ce qu'ils peinent à faire en l'absence de *feedback*. Une autre étude corrobore ces résultats en s'appuyant sur des populations de scripteurs experts (écriture cursive adulte) et novices (écriture cursive d'enfants ou adultes écrivant en lettres script

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

majuscules). Quand le coût lié à la transcription augmente (chez les enfants ou dans le cas de l'écriture script pour les adultes), il devient difficile pour les rédacteurs de coordonner simultanément les processus de haut et de bas niveau (Olive & Kellogg, 2002). Sur le plan méthodologique, il semble qu'une modification de la triple tâche soit possible à moindre effort, en permettant aux sujets de désigner plus d'un processus à la suite des interruptions de la tâche seconde. C'est ce qu'ont démontré Pélissier et Piolat (1999). Il apparait que cette façon de procéder autorise une catégorisation plus fine des processus (la *révision* de Kellogg peut recouvrir de la révision – recherche d'erreurs - et de l'évaluation – adéquation aux objectifs) et met en évidence le déroulement simultané de ceux-ci. C'est le cas notamment de la *traduction* et de l'*exécution*, comme pour la révision et l'évaluation, dont les coûts additionnés semblent ne pas outrepasser les ressources cognitives disponibles.

2.3.2.3. Coûts des processus

Le modèle de Kellogg présente le double avantage d'identifier le coût des processus ainsi que le siège de leur élaboration (voir figure 17, ci-dessous).

	Sous-systèmes de la mémoire de travail :		
Processus :	Calepin visuo-spatial	Administrateur central	Boucle phonologique
Planification	√	√	
Traduction		√	√
Programmation		√	
Exécution			
Lecture		√	√
Edition		√	

Figure 17: Processus rédactionnels et ressources de MdT (Kellogg, 1996)

Selon Kellogg, la formulation requiert la plus importante contribution de la MdT, arguant que ses deux sous-processus sollicitent l'AC et les systèmes esclaves, chacun respectivement.

La planification, en activant des représentations analogiques, des images mentales, des schémas et la mémoire sémantique, met à contribution le CVS, dont la forte dépendance vis-à-vis de l'AC déclenche son implication conjointe. La traduction, qui se manifeste par le discours intériorisé (*inner speech*), fait appel à la BP pour produire les phrases. Les demandes au niveau de la BP sont telles qu'il est vraisemblable que l'AC y investisse aussi des ressources attentionnelles (maintien en mémoire, choix des mots et des constructions syntaxiques). A l'inverse, l'*exécution* ne s'appuie que sur l'AC, et dans une moindre mesure si la procédure motrice est automatisée (Bourdin & Fayol, 1994). Pourtant, même chez le scripteur adulte expert, la demande ne peut être considérée comme nulle, car la rédaction reste une activité plus coûteuse que la production orale (Bourdin & Fayol, 2002). Pour ce qui est du contrôle, la lecture/compréhension étant une tâche de MdT verbale (cf. Daneman & Carpenter, 1980), elle implique l'AC et la BP. L'*édition*, quant à elle, est une tâche complexe qui s'appuie sur l'AC uniquement.

D'une manière générale, la *planification* et la *révision* sont des processus particulièrement coûteux, plus que ne peut l'être, en moyenne, jouer aux échecs, y compris à haut niveau (Piolat, Olive, & Kellogg, 2005).

Le coût des processus a été étudié selon deux orientations. La première postule l'existence d'un réservoir de ressources attentionnelles à capacité limitée (Bourdin & Fayol, 1994; Kellogg, 2001a; McCutchen, 1996).

2.3.2.3.1. Approche capacitaire du coût des processus rédactionnels

L'idée d'une dépendance des processus à un réservoir unique de ressources était déjà présente chez Hayes et Flower (1980). Ils voyaient dans l'écriture une activité induisant une surcharge attentionnelle. Ce dernier point a été discuté par Piolat et Olive (2000) pour lesquels la capacité des scripteurs à préserver la qualité de leur production, même en situation de double tâche très contraignante (cadence élevée des interruptions), indiquerait qu'il reste à leur disposition des ressources inexploitées. Quoi qu'il en soit, les recherches conduites par Christensen et ses collaborateurs montrent que l'automatisation des processus orthographiques et graphomoteurs, liés à la transcription, libère des ressources de traitement. Celles-ci seraient, selon toute vraisemblance, affectées à la réalisation de processus de haut

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

niveau, comme en témoignent les différences de qualité rédactionnelles des écrits produits (Christensen, 2004; Jones & Christensen, 1999).

Cette notion de réservoir de ressources à capacité limitée a fourni à McCutchen (1996) le principal étayage à son modèle de production écrite, qu'elle adosse au modèle de MdT de Just et Carpenter (1992). Cette approche, dite capacitaire, repose sur l'idée que la MdT est une portion activée de MLT. En ce sens, la MdT n'est pas un espace de stockage à capacité limitée, mais c'est le réservoir d'activation de la MLT dont la capacité est limitée. Si ce modèle a été spécialement conçu pour rendre compte de la compréhension de texte, il n'exclue pas pour autant le traitement d'autres activités complexes telles que la production de texte (Miyake & Shah, 1999). Il explique les différences de compréhension entre individus en termes de quantité d'activation disponible, sur la base notamment, de l'empan de lecture (Daneman & Carpenter, 1980). Ce test consiste à faire lire à des sujets à haute voix des séries de phrases dont ils doivent retenir le dernier mot. Les sujets à fort empan retiennent typiquement quatre mots ou plus, les sujets moyens en retiennent trois, les sujets à faible empan n'en retiennent pas plus que deux. Cette approche apporte un éclairage nouveau sur l'idée de modularité de certains processus, expliquant qu'en cas de surcharge, leur déroulement puisse être encapsulé, et particulièrement les processus syntaxiques[8]. D'une manière générale, toute circonstance entrainant une diminution de la quantité d'activation disponible (*e.g.* maintien d'un ou deux mots en mémoire) se traduit par des erreurs ou l'allongement des latences de compréhension. A l'inverse, ce qui libère de l'activation permet d'améliorer les performances. La libération de ressources d'activation potentielle passe par l'automatisation de processus répétés et maîtrisés.

C'est donc sur cette conception de MdT que McCutchen va intégrer les processus rédactionnels empruntés à Hayes et Flower (1980). Cette formulation permet, dans le sillage du travail réalisé par Just et Carpenter (1992) sur la compréhension, d'expliquer le développement et les différences interindividuelles dans la production écrite. Plusieurs exemples attestent de l'opportunité de ce choix théorique. On retiendra le fait que chez les scripteurs adultes (experts), la *planification* ou la *révision* peuvent entraîner fréquemment

[8] Cela se traduit par une plus grande difficulté à comprendre, spécialement chez les individus à faible empan de lecture, les phrases à construction passive dans lesquelles il importe de distinguer sujet et complément d'agent, puisque le sujet ne réalise pas l'action. Just et Carpenter interprètent ce type de difficulté comme un déficit d'activation ne permettant pas « d'activer, de maintenir, et, plus important, de propager l'information dans de multiples systèmes linguistiques (*e.g.* lexical, syntaxique, discursif) » (McCutchen, 1996, notre traduction).

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

l'interruption du processus de traduction, ce qui donne à penser que ces étapes peuvent être trop « gourmandes » en ressources d'activation, de sorte qu'il n'est plus possible de les mener conjointement. Cela trouve confirmation dans l'étude des enfants apprentis scripteurs. Chez eux, la *traduction* n'étant pas du tout automatisée, elle mobilise toute leur attention. Elle induit même à elle seule des surcharges. Cela est attesté par l'articulation intelligible du contenu de la phrase en cours de rédaction pour faciliter son maintien en mémoire. Cela ne disparaît selon McCutchen, que vers l'âge de neuf ou dix ans, quand le processus commence à s'automatiser, libérant ainsi le volume d'activation nécessaire à une plus grande fluence[9] verbale écrite. Un lien de causalité entre fluence et empan de lecture a d'ailleurs été mis en évidence : quand l'empan augmente, particulièrement avec l'âge, la fluence s'améliore (McCutchen, Covill, Hoyne, & Mildes, 1994). D'autres arguments laissent également penser à un réservoir unique de ressources, indépendamment que l'on considère qu'il s'agisse de ressources d'activation ou de ressources attentionnelles, comme l'a montré Kellogg en comparant rédaction manuscrite (automatisée) et dactylographiée (contrôlée) (Kellogg, 2001a). En diminuant le coût du processus de planification, il met en évidence que les ressources libérées se distribuent sur les autres processus. De façon plus inattendue, l'allégement du processus d'exécution ne bénéficie qu'à la traduction et à la révision, sans que l'auteur n'y puisse fournir une explication satisfaisante, y compris en prenant en compte la modularité de la MdT selon Baddeley (1986, 1974 avec Hitch).

2.3.2.3.2. Approche componentielle du coût des processus

Cette modularité constitue le second axe d'étude du coût des processus, qui postule que les sous-systèmes de MdT ne sont pas sollicités de la même façon par les processus rédactionnels. Plusieurs études ont été conduites pour en témoigner, avec un succès limité, en regard de l'objectif initial qui était la validation du modèle de Kellogg. Les principales critiques portant sur ces recherches sont essentiellement méthodologiques quant aux choix des tâches interférentes (phonologiques et visuo-spatiales) ainsi que sur le contrôle de leur coût respectif. Par ailleurs, une relecture critique de ces investigations donne à penser qu' « il serait certainement préférable de réaliser des études plus fines et plus systématiques sur le fonctionnement de chacun des processus de production, en évaluant leur coût respectif dans le

[9] La fluence se mesure en nombre de mots produits pour une unité de temps donnée.

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

cadre […] des capacités limitées de maintien […] de la mémoire de travail » (Chanquoy & Alamargot, 2002). C'est pourquoi des travaux plus récents ont avantageusement mis en pratique cette recommandation, en abordant la question de la sollicitation des sous-systèmes de MdT par les processus rédactionnels de façon plus spécifique. Dans ce contexte de recherche, Kellogg (2004) a montré que la concurrence d'une épreuve interférente de rétention de six chiffres n'affectait que la longueur des phrases produites par les sujets, mais sans en diminuer la complexité grammaticale. Cette tâche interférente, mobilisant fortement la MdT verbale, n'a pesé que sur la longueur des phrases produites, contrairement aux tâches concurrentes de rétention visuelle, spatiale ou phonologique légère (trois chiffres à mémoriser) qui elles, n'ont pas eu d'effet. A l'inverse, la complexité grammaticale reste intacte, quelle que soit la condition. Ce type de résultat, interprété comme un indice de la modularité des traitements grammaticaux (cf. Just & Carpenter, 1992) est une illustration de l'implication certaine de la BP dans les processus lexicaux, même si elle n'est pas systématique (Bonin, 2005; Fayol & Gombert, 1999). Cette analyse a fait l'objet de développements très récents sous la plume de Kellogg et de ses collaborateurs, dans une série de trois expériences dont les résultats sont cohérents. En enregistrant les TR à une tâche de décision phonologique, visuelle, ou spatiale, ils font valoir l'implication indubitable de la MdT verbale, mais aussi celle de la MdT visuelle, suivant le type sémantique du matériel linguistique à traiter (mots concrets ou abstraits). Le CVS serait donc bien engagé dans la planification (pour le moins, de mots concrets, puisque les auteurs n'excluent pas l'idée d'une MdT sémantique, à rapprocher sans doute du BE de Baddeley (2000) ou du registre sémantique de Hayes (1996)). S'il est admis que les traitements linguistiques (phonologiques, grammaticaux, orthographiques) sont bien sous la dépendance de la MdT verbale, des incertitudes persistent quant à leur subordination à la MdT visuo-spatiale (Kellogg, Olive, & Piolat, 2007).

2.3.3. Critique de l'intégration de la mémoire de travail aux modèles de production écrite

Le modèle de Kellogg est indéniablement très riche, mais ses limites sont inhérentes à celles du modèle de MdT de Baddeley (Chanquoy & Alamargot, 2002). Si le modèle de Kellogg se veut ambitieux dans la description des représentations engagées dans l'activité

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

rédactionnelle (visuo-spatiales, phonologique, *etc.*), il peine à rendre compte de l'organisation des processus dans une approche plus dynamique de leur enchaînement. Les modèles d'activation (Anderson, 1996; Ericsson & Kintsch, 1995) seraient vraisemblablement plus adaptés au traitement de cette problématique, permettant aussi de déboucher sur la question cruciale de l'acquisition d'expertise rédactionnelle (Chanquoy & Alamargot, 2002). Cette critique de son modèle n'est pas restée étrangère à Kellogg, qui, à l'instar de McCutchen (2000), s'est penché sur la question de l'intégration du modèle MdT-LT à la production écrite (Kellogg, 2001b). On reprochera aussi au modèle de Kellogg de ne pas prendre en considération l'environnement de production écrite, et particulièrement le médium d'écriture. Cette question n'est en fait envisagée que de façon incidente par l'entremise du coût de l'exécution motrice qui peut, ou non, suivant la situation, être, ou pas, automatisée. C'est à notre sens l'un des intérêts majeurs du modèle de Hayes (1996), bien que cette prise en considération ne soit restée à ce stade de son élaboration qu'un vœu pieux.

C'est à cet effet qu'on regrettera que le modèle de Kellogg n'accorde que peu d'importance à la compréhension des mécanismes cognitifs impliqués dans la réalisation du processus d'exécution. Certes, il est vrai que ce modèle ambitionne d'expliquer les liens entre rédaction de texte et MdT. Cela concerne des niveaux de traitement assez élevés auxquels la description des processus orthographiques et graphomoteurs n'apporterait vraisemblablement qu'un faible éclairage. Pourtant, la question du coût de l'étape d'exécution n'est pas restée étrangère aux promoteurs de cette théorie (Kellogg, 2001a; Olive & Piolat, 2002) ainsi qu'à d'autres (Christensen, 2004; Jones & Christensen, 1999). En effet, toute situation de production écrite impliquant un contrôle attentionnel de l'exécution motrice est susceptible d'interférer avec les autres processus en diminuant les ressources nécessaires à leur accomplissement. Ceci n'est pas anodin pour qui s'intéresse à l'apprentissage d'une interface de saisie de texte manuscrit chez des utilisateurs novices, ce qui est notre cas. Par ailleurs, l'activité de contrôle de la production écrite s'applique à tous les niveaux de l'élaboration du contenu du texte : élaboration du propos, respect des règles orthographiques et syntaxiques, et, bien évidemment, dans le cadre de l'écriture manuscrite, lisibilité et forme des lettres. Ce dernier aspect est d'autant plus important que le texte produit doit être partagé ou lu par des tiers. Ceci est également vrai dans le cas d'une PdN impliquant une relecture ultérieure, et plus encore si les lettres doivent être reconnues par une machine. Ce type de contrôle visuel de l'écriture a fait l'objet d'études poussées, sans pour autant être lié explicitement à la MdT

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

(Serratrice & Habib, 1993; Zesiger, 1995). Comprendre ces mécanismes négligés par le modèle de Kellogg est donc un enjeu important de notre démarche, c'est pourquoi nous lui consacrons les lignes qui suivent.

2.4. La production graphique de lettres

La production langagière révèle une certaine hétérogénéité dans ses multiples modes de production : langage oral, écriture scripte ou cursive, ou bien dactylographie, pour n'en citer que quelques uns. Au niveau central du traitement de ces informations (par opposition au niveau périphérique qui renvoie au mode de production du langage, qu'il s'agisse de l'appareil bucco-pharyngé ou de l'appareil articulaire de la main), on n'observe pas de différence notable[10] : la *planification* (gestion des buts et de la stratégie de communication et du contenu sémantique) et la *traduction* sont relativement indépendantes du mode de production des informations langagières, pour peu que l'*exécution* ne soit pas trop gourmande en ressources attentionnelles (Bonin, Pacton, & Fayol, 2001; Bourdin & Fayol, 2002; Christensen, 2004; Jones & Christensen, 1999; Olive & Kellogg, 2002; Olive & Piolat, 2002). L'étude des processus induits par l'*exécution* (*programmation* et *exécution* motrices) sont l'objet d'un courant de recherche particulier, issu à la fois de la neuropsychologie et de la psychologie cognitive. Ces questions ne sont pas sans intérêt pour nous, car la méthode de saisie DIGIME impose à ses utilisateurs d'y saisir les mots lettre par lettre, séparément, en prenant le soin bien sûr, de choisir des allographes de lettres susceptibles d'être identifiés par la machine. Par conséquent, les enjeux recouverts par ces processus périphériques revêtent une importance capitale pour notre problématique.

[10] A cette nuance près que les intentions et circonstances conduisant à l'écriture ou au discours oral diffèrent largement. L'écriture est une situation d'énonciation monologique, c'est-à-dire privée, tandis que le discours oral implique une réactivité plus importante en regard des attentes de son interlocuteur. D'autre part, la syntaxe et le registre d'énonciation se différencient à l'oral et à l'écrit, sans pour autant que cette distinction soit parfaitement figée : une conférence orale tient plus de la rigueur de formulation propre à l'écrit, tandis que la messagerie électronique instantanée s'accommode volontiers des approximations de l'oral (Bonin, 2003).

2.4.1. Du mot à la lettre : les registres mnésiques sollicités par les processus orthographiques

L'individu en situation de production écrite va sélectionner un certain nombre d'unités lexicales pour composer ses phrases. Ces mots peuvent être issus d'une source orale (dictée), d'une source écrite (copie) ou encore de son propre registre sémantique (composition ou PdN à partir de l'élaboration macrostructurelle du texte à réaliser). A partir de cela, il va élaborer la microstructure de sa production en puisant soit dans son lexique graphémique ou orthographique (procédure d'adressage, pour les mots connus), soit en utilisant la voie phonologique qui consiste à convertir les phonèmes en graphèmes (procédure d'assemblage, pour les mots inconnus). Quelque soit la voie choisie, les informations aboutiront au tampon graphémique. Comme Zesiger le fait remarquer (1995), s'il existe plusieurs façons, suivant les auteurs, de définir le graphème, nous nous cantonnerons à évoquer celui-ci en tant que représentation abstraite d'une lettre. L'autre conception du graphème l'envisage comme étant la représentation abstraite écrite d'un phonème. Nous y ferons référence sous l'appellation grapho-syllabe.

2.4.1.1. Lexique orthographique

Cette structure mnésique à long terme stocke les mots sous forme de suite de lettres (ou graphèmes) indépendamment du type de sortie (casse et forme des lettres). C'est dans ce registre que le scripteur va puiser des mots qui lui sont suffisamment familiers pour s'y trouver. Contrairement à ce qui a été longtemps considéré à son sujet, l'unité de base du lexique orthographique n'est pas le mot, mais plutôt le morphème, c'est-à-dire l'unité minimale porteuse de sens (Badecker, Hillis, & Caramazza, 1990). L'accès à cette base de données orthographique se fait par une procédure dite d'adressage, car elle consiste en la récupération d'informations stockées en MLT. La constitution de cette mémoire se fait tout au long de la vie de l'individu, en se confrontant à de nouveaux mots dont l'orthographe finit par y être représentée (Zesiger, 1995).

2.4.1.2. La conversion phonème-graphème

Il arrive cependant parfois qu'un scripteur, en situation de dictée ou de PdN orale, soit amené à écrire un mot qui lui est inconnu et dont il ne connaît pas l'orthographe. Cette situation a largement été étudiée par l'utilisation de pseudo-mots, séquence de lettres plausible dans la langue, mais dénuée de sens. Dans ce cas, après analyse acoustique du mot, le scripteur va segmenter la représentation auditive qu'il s'en est fait. Fractionné en unités phonologiques, le mot peut être ainsi converti en graphèmes. Cette procédure, dite d'assemblage, consiste donc à associer, pour chaque phonème découpé, une séquence de lettres susceptible de porter le son auquel elle est associée. Cela se fait autant sur la base de fréquences d'occurrence dans la langue que de règles liées aux positions des lettres dans le mots et des lettres entre elles (Zesiger, 1995).

2.4.1.3. Le tampon graphémique

Quelle que soit la voie en mémoire que le mot « emprunte » (adressage ou assemblage), il aboutit au tampon graphémique. Cette structure de stockage maintient à court terme la séquence de graphèmes qui forme le mot à écrire avant que cette représentation soit traitée par des structures périphériques (Zesiger, 1995). Si l'on a précédemment pu croire que le tampon graphémique ne contenait que des informations relatives à « l'identité des représentations abstraites des lettres – autrement dit la spécification des graphèmes contenus dans le mot – et l'ordre dans lequel ces graphèmes sont agencés » (Zesiger, 1995), des travaux plus récents ont largement enrichi cette conception. En effet, des données issues de l'étude du patient L.B. (Caramazza & Miceli, 1990) ont permis d'asseoir l'idée que les représentations graphémiques ont une structure multidimensionnelle, contenant notamment des informations relatives aux grapho-syllabes, à l'alternance consonne-voyelle, et bien sûr, à l'identité des graphèmes en ordre. A ce dernier niveau s'ajoute un niveau supplémentaire qui code les consonnes géminées (doublées). Des résultats issus de la psychologie expérimentale ont confirmé cela en partie, notamment le bien-fondé de la syllabe comme unité de production en écriture manuscrite, ce qui se répercute au niveau moteur : les temps intermédiaires entre la rédaction de chaque lettre constitutive d'un mot sont plus longs entre deux syllabes qu'à l'intérieur de celles-ci (Kandel, Alvarez, & Vallée, 2006; Kandel, Valdois, & Orliaguet, 2003).

2.4.2. Réalisation graphique des lettres

2.4.2.1. Les niveaux postgraphémiques

En aval des processus orthographiques se déroulent les processus postgraphémiques. Ces processus se distinguent entre eux en fonction du mode de production des lettres choisi par le scripteur : épellation, dactylographie ou écriture manuscrite (Bonin, 2003; Zesiger, 1995). Cette différenciation se fonde, là encore, sur l'étude de patients déficitaires présentant des troubles spécifiques pour l'un ou l'autre de ces modes de sortie (Bonin, 2003). Dans le cas de l'écriture manuscrite, les niveaux postgraphémiques renvoient à deux aspects de l'agencement spatial du texte sur le support (Paillard, 1988; Zesiger, 1995):

- **Morphocinétique** : il s'agit des caractéristiques propres à la forme de la lettre dessinée par la trajectoire cursive de l'instrument d'écriture (plume, ou stylet sur PDA). On admet que l'élaboration des lettres est principalement sous la dépendance de processus de contrôle proactif, bien qu'il y ait quelques exceptions à cela. Ce contrôle proactif est à la fois visuel et proprioceptif, et permet plus d'éviter les erreurs que de les corriger (Serratrice & Habib, 1993).
- **Topocinétique** : il s'agit de l'agencement spatial du texte sur le support, c'est-à-dire le respect de la progression gauche→droite (pour l'alphabet latin) et de l'agencement général des phrases sur des lignes qui doivent se superposer parallèlement les unes aux autres jusqu'au bas de la feuille. Les implications relatives à cette question ne seront pas traitées ici, puisque la méthode de saisie DIGIME ne permet que de la saisie lettre par lettre. Par conséquent, l'organisation des mots sur l'écran est totalement prise en charge par la machine et n'est pas sous la dépendance de ses utilisateurs. Néanmoins, les aspects topocinétiques peuvent intervenir marginalement dans la formation de quelques lettres, notamment pour placer la barre du *t*, le point sur le *i* et bien sûr les signes diacritiques (accents, cédille) : il s'agit des exceptions dont nous faisions état précédemment pour lesquelles un contrôle rétroactif intervient pour former les lettres (Zesiger, 1995). L'implication de ces mécanismes topocinétiques dans l'élaboration de la forme même des lettres est sous la dépendance étroite du contrôle visuel. Par contre, la taille de l'écriture, qui fait l'objet d'ajustements contextuels pour s'adapter, par exemple, aux dimensions de la feuille, aux fins de lignes, *etc.,* est largement

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

automatisée par une programmation motrice adéquate, de la même façon que lorsque nous traversons une route, quelque soit la taille de notre pas, il « s'ajuste inconsciemment et à l'avance pour gravir le trottoir opposé » (Serratrice & Habib, 1993).

2.4.2.2. Niveau allographique

Les mots stockés dans le tampon graphémique sont ensuite traités au niveau allographique. Cela concerne les différentes formes graphiques que peut prendre un graphème donné, suivant la casse (majuscule ou minuscule), le type d'écriture (cursif ou script) et de même que toute variation de la forme canonique de la lettre propre à chaque scripteur. Ce codage serait de nature visuelle ou spatiale, selon les auteurs (Zesiger, 1995). Bien que des travaux de neuropsychologie cognitive laissent penser à des atteintes spécifiques de ce niveau, d'autres auteurs l'envisagent déjà comme une émanation du programme moteur qui va permettre sa réalisation par les muscles et articulations de la main (Bonin, 2003).

Au-delà de ce que le niveau allographique permet au scripteur de choisir un allographe sur la base d'informations contextuelles (ex : majuscule en début de phrase ou en initiale de noms propres) et des « préférences » propres à chacun dans le style d'écriture (cursif ou script), une recherche montre que beaucoup de scripteurs utilisent un style hybride (mélange de cursif et de script, et pouvant se distribuer sur un *continuum* favorisant plutôt l'un ou l'autre). Ces caractéristiques n'affectent pas de façon claire la vitesse de production ou la lisibilité, contrairement à ce que l'on a pu croire auparavant (Graham, Weintraub, & Berninger, 2001). En revanche, il est un poncif qui s'est vu confirmé par l'expérimentation : il existe bien une différence entre l'écriture des filles et des garçons. Cette différence est d'ailleurs repérée très tôt par les enfants. Pour autant, il est difficile de savoir précisément la part du biologique et du culturel dans ce constat, considérant qu'il serait très surprenant que les deux n'interviennent pas conjointement (Hartley, 1991).

2.4.2.3. Niveau moteur

La programmation motrice fait débat chez les spécialistes de la graphomotricité, car, loin de ne se limiter qu'à la production de lettres, elle englobe également toute réalisation

graphique, et indépendamment de l'organe effecteur (main, pied ou bouche, pour certaines situations expérimentales ou contextuelles particulières). Notre propos n'étant pas d'éclairer ce litige, nous nous cantonnerons à évoquer certaines connaissances sur le sujet susceptibles d'illustrer au mieux ses implications dans la production de lettres.

Le niveau moteur concerne à la fois des processus centraux (programmation motrice) et périphériques (implémentation de la commande au niveau de l'organe effecteur). Cette distinction est particulièrement importante pour justifier que la nature de la programmation motrice serait de nature spatiale, c'est-à-dire indépendante de l'organe effecteur et des contraintes matérielles de son élaboration (notamment en terme de taille) (Zesiger, 1995). En effet, la production de traces graphiques (lettre, ou toute autre forme) témoigne d'une certaine invariance. Spatiale, tout d'abord, puisque quel que soit l'organe effecteur ou la taille de la lettre, celle-ci conserve les caractéristiques de sa configuration, que l'on peut principalement définir par la hauteur, la largeur et les angles (courbure des traits) des lettres (Peeples & Retzlaff, 1991). Temporelle ensuite, car là encore, il existe une étonnante stabilité dans la durée proportionnelle du tracé des éléments constitutifs de la lettre par rapport à la durée totale de sa réalisation. Ces deux phénomènes, connus sous les termes d'homothéties spatiale et temporelle, entretiennent une certaine interdépendance entre eux et concourent largement à la lisibilité de l'écriture (Longstaff & Heath, 1997). Par conséquent, au niveau central, le programme moteur contiendrait essentiellement des informations relatives à la forme de la lettre (forme générale, taille relative et orientation des traits qui la composent) ainsi que l'ordre dans lequel la séquence de traits doit être réalisée.

Au niveau périphérique, le programme moteur serait paramétré par des informations sur la force à appliquer au geste (ce qui en détermine la taille) et sur les groupes musculaires à sélectionner pour sa mise en œuvre. Cette approche repose essentiellement sur un contrôle proactif de la production de traits. Néanmoins, le contrôle retroactif joue aussi un rôle dans l'exécution des lettres (Serratrice & Habib, 1993). Rappellons qu'Olive et Piolat (2002) ont montré une augmentation du coût de l'exécution graphomotrice en l'absence de feedback visuel. Concernant le feedback proprioceptif, s'il semble évident qu'il joue un rôle dans les caractéristiques morphocinétiques de la trace (en complément de la vision), la difficulté à opérationnaliser son expérimentation (exception faite des patients désafférentés) favorise sa méconnaissance. Malgré cela, il joue sans doute un rôle propre dans le contrôle de la prise du stylo par la main. Elle semble d'ailleurs faire preuve d'une assez grande adaptabilité aux

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

conditions matérielles de l'accomplissement de son mouvement. En particulier, le diamètre et la forme du crayon ne semblent pas affecter la qualité de la production selon une revue de question dont les auteurs soulignent tout de même le manque de travaux (Dobbie & Askov, 1995). Ceci a d'autant plus d'importance que, pour des questions d'encombrement, les stylets de PDA affichent des dimensions fortement réduites aussi bien en longueur qu'en diamètre.

2.5. De la mémoire de travail à la production graphique de lettres

Dans son ouvrage de référence, Zesiger (1995) évoque à plusieurs reprises l'existence de mémoires tampon mises à contribution dans la production graphique de lettres. Il rapproche ces registres mnésiques à court terme de la MdT sans apporter plus de précision. Lambert et Espéret (1997) franchissent le pas en proposant, exemples à l'appui, que l'acquisition et l'amélioration des aptitudes graphomotrices en écriture chez les enfants puissent être une émanation du phénomène de *chunking* en MdT. Ce processus, volontaire ou non, permet de dépasser les capacités limitées de la MdT en y associant plusieurs éléments, de sorte qu'ils n'y occupent plus que le volume d'un seul, comme on le fait, par exemple, en associant les chiffres des numéros de téléphone par deux ou trois, favorisant de cette manière leur mémorisation (Tricot & Chanquoy, 1996). Le même procédé serait donc applicable aux lettres (ou plus exactement à leur représentation en mémoire), dont la configuration résulterait de l'association de plusieurs traits les composant, mais ne sollicitant qu'un volume de ressources cognitives inférieur à ce que leur ensemble, pris séparément, représente. Cette conception se retrouve également chez les spécialistes de la motricité dans son acception la plus large (Mounoud, 1986). Ainsi considéré, l'apprentissage graphomoteur pourrait être assimilé à l'élaboration de schémas en mémoire (Cooper, 1998; Tricot, 1998, voir chapitre suivant). Cette approche est donc capacitaire. Il faudra donc se pencher sur, une fois encore, la compatibilité des modèles de MdT existant avec ce nouvel éclairage, et sur la nature même des représentations engagées afin de déterminer si ils sont « solubles » dans le modèle de Baddeley, et par conséquent, dans celui de Kellogg. Cette voie semble pouvoir bénéficier des travaux d'Alamargot et ses collaborateurs qui ont engagé une démarche ambitieuse d'élaboration d'un protocole combinant l'étude des mouvements oculaires du scripteur en parallèle d'une analyse dynamique de sa production écrite avec le dispositif « *eye and pen* » (Alamargot, Chesnet, Dansac, & Ros, 2006; Chesnet & Alamargot, 2005).

3- Mémoire, apprentissage et performance

3.1. Environnements Informatisés pour l'Apprentissage Humain (EIAH)

Shackel (1991) et Nielsen (1993) s'accordent à faire de l'apprenabilité d'un système une dimension « clé » de son utilisabilité. C'est tout l'intérêt de comprendre les mécanismes cognitifs impliqués par l'apprentissage de l'utilisation d'une IHM en conduisant l'utilisateur novice à acquérir un certain niveau d'expertise dans la manipulation de celle-ci. A cet égard, les dispositifs de saisie et de reconnaissance de l'écriture manuscrite ont ceci de particulier qu'ils nécessitent de ré-apprendre à écrire, de sorte que les lettres formées par l'utilisateur soient reconnues par la machine. Utiliser une IHM implique donc un apprentissage, ou, pour le moins, que l'utilisateur mette à profit certaines de ses connaissances préalables pour réaliser la tâche qu'il s'est assigné. Cet apprentissage se fait bien évidemment au cours de la pratique avec le dispositif, c'est-à-dire, pendant l'accomplissement de la tâche. On peut néanmoins favoriser l'apprentissage d'une interface en proposant des ressources qui, mises à disposition de l'utilisateur, vont en faciliter la découverte. La psychologie cognitive a largement traité ces questions dans le cadre des recherches menées sur les EIAH (Erhel & Jamet, 2006; Gyselinck, Jamet, & Dubois, 2008; Hidrio & Jamet, 2002, , 2008; Hidrio, Jamet, & Le Bohec, 2002; Jamet, 2002, , 2006; Jamet & Erhel, 2006; Jamet, Le Bohec, & Hidrio, 2003; Mayer, 2001; Sweller, 1999; Sweller, Van Merriënboer, & Paas, 1998). Bien que ceux-ci soient le plus souvent envisagés comme des supports éducatifs impliquant un apprentissage explicite, ils sont aussi bien souvent mis à contribution dans les IHM sous forme de tutoriel (manuel de l'utilisateur) ou de didacticiel (logiciel d'instruction). Dans ce dernier cas, le support électronique offre alors d'immenses possibilités de communication des informations par des canaux (ou modalités) aussi divers que variés : c'est l'apprentissage multimédia et son incroyable capacité à transmettre des connaissances difficilement verbalisables comme les actions ou les procédures (Najjar, 1996). La facilitation de ce type d'apprentissage est un aspect central de la qualité d'interaction avec la machine ainsi que nous l'avons souligné au cours du premier chapitre de notre exposé.

La racherche sur les EIAH vise à faciliter, non pas la tâche, car c'est impossible (sauf à la transformer), mais plutôt les conditions de sa réalisation. On facilite ainsi l'apprentissage d'une procédure en libérant une partie des ressources cognitives qui ne pouvait y être affectée, mobilisée par des traitements de moindre importance. On désigne sous le nom de charge « la

capacité de mémoire immédiate mobilisée par un sujet lors de la réalisation d'une tâche » (Tricot & Chanquoy, 1996). On peut donc envisager le concept de charge mentale comme une dimension de la réalisation d'une tâche au sens le plus général du terme (c'est-à-dire dépassant le cadre restreint de l'apprentissage). Ainsi considérée, la charge est la résultante de l'activation d'opérations mentales et d'actions. Les actions vont permettre de changer l'état d'avancement de la tâche à accomplir en passant d'un sous-but *n* à un sous-but *n + 1* (Tricot & Chanquoy, 1996). Ces actions peuvent être aussi bien physiques que mentales. On parlera alors d'opérations dans ce dernier cas. C'est au sein de l'architecture cognitive qui leur sert de support que ces opérations vont avoir lieu.

3.2. Architecture cognitive et fonctionnement mnésique

L'architecture cognitive s'articule autour de trois types de mémoires : les mémoires sensorielles, la mémoire de travail (MdT) et la mémoire à long terme (MLT) (Atkinson & Shiffrin, 1968, cités par Baddeley, 2002b). Chacune de ces mémoires présente des caractéristiques et des limitations qui lui est propre (Baddeley, 2002b; Cooper, 1998; Lieury, 1997).

Les mémoires sensorielles sont des registres mnésiques qui prennent en charge les informations issues des organes des sens. Elles ne permettent qu'un stockage éphémère de l'information, de quelques centaines de millisecondes pour la mémoire sensorielle visuelle à quelques secondes pour la mémoire sensorielle auditive (Lieury, 1997). Constamment rafraîchies par de nouvelles informations entrantes, ces mémoires ne durent que le temps du transfert de ce qui y est stocké vers des mémoires de plus haut niveau comme la MdT ou la MLT. Elles jouent donc le rôle de mémoires tampon.

En ce qui concerne la MdT, et en regard de la présentation détaillée que nous en avons faite au chapitre précédent, nous nous cantonnerons à rappeler que, plus qu'un espace de stockage passif, elle constitue un espace de traitement à capacité limitée. L'unité de stockage de la MdT est le *chunk*, dont la propriété d'être « élastique » et relativement indépendant des informations qu'il représente, permet d'excéder les capacités limitées de la MdT (Baddeley, Thompson, & Buchanan, 1975; Gaonac'h & Larigauderie, 2000; Jones, 2002). Ainsi, un *chunk* peut aussi bien être occupé par un seul chiffre que par une image mentale complexe, si

tant est que celle-ci est encodée sous forme d'une représentation combinée (Cooper, 1998). Cette particularité permet d'outrepasser les capacités de la MdT qui ne peut contenir simultanément que 7 ± 2 *chunks* (Miller, 1956), voir 4, quand le maintien à court terme des éléments est associé à du traitement. La MdT peut coordonner des informations issues aussi bien de la MLT (connaissances) que des registres sensoriels (stimuli) pour les transformer, les recombiner et ainsi, créer de nouvelles connaissances. C'est ainsi qu'en interaction avec l'environnement d'un individu, la MdT va permettre de mener à bien un certain nombre de tâches (actions, résolution de problèmes, *etc.*). Lorsque les informations traitées par la MdT sont encodées en MLT et intégrées au réseau de connaissances antérieures, on parle alors d'apprentissage. Bien que la MdT soit fortement impliquée dans l'apprentissage de connaissances déclaratives, son rôle est plus complexe dans les apprentissages procéduraux ou moteurs (Anderson, 1982, 1987; Maxwell, Masters, & Eves, 2003).

La MLT est un espace de stockage dont le volume ne connaît aucune limite puisqu'il est toujours possible d'y ajouter de nouvelles connaissances ou compétences, et ce, pour une durée indéterminée (Cooper, 1998; Lieury, 1997). Les connaissances en MLT peuvent être implicites ou explicites (Baddeley, 2002b), auquel cas elles sont stockées sous forme de schémas. Ceux-ci sont des représentations mentales qui enchâssent plusieurs sous-éléments informationnels, de telle sorte que leur récupération en MdT ne va occuper que le volume d'un seul élément (un *chunk*). Ces sous-éléments peuvent être eux-mêmes d'autres schémas de plus bas niveau. Ainsi, c'est toute la MLT qui est organisée hiérarchiquement dans une arborescence de schémas (Sweller, Van Merriënboer, & Paas, 1998). Au sein de la MdT, le *chunking* va permettre l'intégration des informations aux schémas préexistants pour en former de nouveaux (Shaffer, Doube, & Tuovinen, 2003). Le stockage de nouvelles connaissances est sous la dépendance de cette organisation préalable. Ces constructions mentales catégorisent les informations ainsi que les problèmes à résoudre afférents (Paas, Renkl, & Sweller, 2004). Le bénéfice tiré de la construction de ces schémas est double: d'une part, ils permettent une économie de ressources considérable lors de leur activation en MdT puisqu'ils peuvent contenir plus d'informations que ce que leur seul volume n'autorise (Kalyuga, Chandler, & Sweller, 1998). D'autre part, leur mise en œuvre répétée va là encore faire diminuer le coût de leur traitement en mémoire (Sweller, Van Merriënboer, & Paas, 1998). C'est le volet procédural de l'exploitation des compétences dont le coût en ressources s'amenuise par « automatisation » (bien que le concept même soit sujet à discussion selon

Perruchet, 1988). La pratique, l'expérience et les connaissances acquises pour un domaine donné vont contribuer aux différences interindividuelles dans l'élaboration de schémas de plus en plus riches et complexes. Par conséquent, apprendre peut être considéré comme l'acquisition de nouveaux schémas qui vont transformer durablement l'organisation des connaissances en mémoire (Cooper, 1998; Tricot, 1998).

3.3. Apprentissage, motricité et imagerie mentale

Au-delà des processus de haut niveau mis en œuvre dans la production de texte, et ce, quelque soit son support matériel, l'écriture impose, à son plus bas niveau, l'exécution motrice de mouvements. Bien que nous ayons précédemment développé les liens entre écriture et motricité, il convient également de s'intéresser aux processus et conditions qui vont permettre à un individu d'intégrer le fonctionnement d'une IHM. Cela recouvre un certain nombre de procédures (c'est-à-dire de règles d'utilisation de l'interface), mais aussi un ensemble de gestes moteurs (et plus précisément graphomoteurs, pour ce qui nous concerne), dont la maîtrise par l'utilisateur constitue un gage d'efficacité.

De même que la MdT est componentielle, la MLT témoigne aussi d'une certaine hétérogénéité. La première distinction dans l'organisation de la MLT tient au type de connaissances qu'elle stocke : explicites ou implicites (Baddeley, 2002b). La mémoire explicite relève des connaissances verbalisables, c'est-à-dire celles que l'on peut récupérer consciemment en mémoire pour les exprimer. A l'inverse, la mémoire implicite concerne des compétences, des savoir-faire, dont la récupération en mémoire ne peut se faire que sous la forme de la mise en œuvre d'une procédure. La littérature scientifique assimile le plus souvent explicite à déclaratif et implicite à procédural. Ceci peut être réducteur. Certains auteurs défendent qu'il est impossible d'identifier tout ce qui peut être appris implicitement, et de citer l'exemple suivant : « Comment déterminer […] si un individu anticipant le point d'arrivée d'une balle abstrait inconsciemment les règles de la balistique, ou a simplement mémorisé un certain nombre de lancés antérieurs à celui-ci ? » (Pacton & Perruchet, 2006). Autre exemple, celui de l'apprentissage du langage oral, souvent présenté comme prototypique de l'apprentissage implicite, et qui ne porte pas uniquement sur des connaissances procédurales. Comme nous le verrons par la suite, connaissances déclaratives

et procédurales peuvent être intimement liées (Anderson, 1982; 1987; Lee & Anderson, 1997; Taatgen & Lee, 2003).

Au sein même de cette première partition de la MLT, d'autres distinctions sont également pertinentes. Les connaissances procédurales reflètent autant le conditionnement classique, les connaissances non associatives (habituation et sensibilisation), que les habiletés et compétences, notamment motrices (Baddeley, 2002b). Ce sont ces dernières qui nous intéressent au premier chef, puisque écrire à la main implique, au plus bas niveau de sa réalisation, un comportement moteur. Par conséquent, apprendre à utiliser une interface de saisie et de reconnaissance de l'écriture manuscrite nécessite autant d'assimiler les procédures d'exploitation de l'interface (raccourcis, fonctions spéciales, *etc.*) que les allographes de caractères susceptibles d'être reconnus par la machine. Il s'agit donc d'apprentissages procéduraux. Ceux-ci peuvent se réaliser de façon incidente, par une utilisation répétée de l'interface, de la même façon que l'on apprend à faire du vélo, sans que l'on soit capable d'expliquer comment on finit par y parvenir. Il est aussi possible d'acquérir ce type de savoir à partir de connaissances déclaratives, qui, par définition, sont explicitables, et donc, à même d'être transmises et partagées. Certains auteurs alertent sur le fait qu'il est réducteur de penser que les savoirs déclaratifs ne deviennent procéduraux que par le simple mécanisme d'automatisation qui les rend implicites et inconscients (Griggs & Carot, 2002). Et de citer Anderson (1987, p.206) : « Notre système de savoirs déclaratifs a la capacité d'emmagasiner […] nos expériences dans n'importe quel domaine, y compris les enseignements, un modèle de comportement correct, succès et échecs de nos propres tentatives, *etc*. Une caractéristique de base du système déclaratif est qu'il ne requiert pas pour emmagasiner le savoir qu'on sache comment celui-ci sera utilisé. Cela signifie qu'on peut mettre aisément un savoir pertinent dans notre système, mais qu'un effort considérable doit être fait quand vient le moment de convertir ce savoir en comportement. » (traduit par Griggs & Carot, 2002).

Par conséquent, il convient, dans le cadre de notre problématique, de comprendre les mécanismes cognitifs qui permettent un apprentissage procédural (et plus particulièrement graphomoteur), que celui-ci soit implicite ou explicite.

3.3.1. Passer du savoir au savoir-faire

3.3.1.1. La compilation de connaissances (Anderson, 1982)

L'étude des relations entre les connaissances déclaratives et procédurales a considérablement tiré parti des travaux d'Anderson. En prenant appui sur le modèle d'acquisition de compétences développé par Fitts (1964, cité par Anderson, 1982), Anderson propose une théorie qui reprend également l'idée de trois stades successifs dans l'apprentissage de compétences (plus précisément, de compétences cognitives – *cognitive skills*).

Le premier niveau est celui des connaissances déclaratives (*declarative stage*). Des instructions y sont encodées en tant que connaissances factuelles sur la compétence à développer. Celles-ci peuvent donner lieu à interprétation, de telle façon qu'il sera possible de produire le comportement. A ce stade, une médiation verbale est fréquemment observée pour soulager la MdT d'une charge trop importante imposée par l'activation conjointe de toutes les connaissances nécessaires, parallèlement à l'élaboration de la procédure. Ceci explique que de nombreuses erreurs dans l'accomplissement de la tâche sont aussi fréquemment observées à ce stade, et imputables à une surcharge mentale (par exemple, disparition d'informations maintenues en MdT, pourtant essentielles à la poursuite de la tâche, *etc.*).

Le second niveau est celui de la compilation de connaissances (*knowledge compilation*). Compilation est à entendre dans son acception informatique, c'est-à-dire qu'il s'agit de convertir un « code source » en fichier exécutable par la machine. La métaphore informatique du fonctionnement cognitif prend une fois encore tout son sens : par une pratique répétée, les connaissances déclaratives vont être converties en connaissances procédurales directement utilisables, sans la médiation de procédures d'interprétation. C'est donc à ce niveau que les connaissances déclaratives sont converties en connaissances procédurales.

Le troisième et dernier niveau est celui dit procédural (*procedural stage*). Sa mise en place s'accompagne d'un ensemble d'ajustements apporté aux connaissances procédurales pour que leur réalisation soit toujours plus rapide et opportune.

Si la seule description de ces trois stades dans la « procéduralisation » des connaissances n'apporte rien à la formalisation proposée par Fitts, la contribution d'Anderson est de décrire les processus qui permettent de passer de l'un à l'autre. Il développe une approche des

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

connaissances procédurales en tant qu'ensemble intégré de règles de production conditionnelles de type « SI … ALORS … ». Chacune de ces règles est applicable à des connaissances déclaratives qui vont encadrer l'action (par exemple, SI *connaissance a* ALORS *action b*). La procéduralisation peut donc être appréhendée, non seulement comme une mise en application de connaissances déclaratives, mais surtout comme le résultat d'une analyse de type « moyen-fin » dont l'examen du but final va décomposer la procédure en plusieurs sous-buts. Cette structure d'élaboration inclut donc, en dépit de son caractère linéaire, des boucles de contrôle à tous les niveaux de résolution des sous-buts. Cela autorise d'éventuels retours en arrière le cas échéant.

La compilation va donc se constituer à partir de connaissances déclaratives qui, bien que présentant l'avantage de permettre une grande souplesse, sont cependant très coûteuses à interpréter, aussi bien en termes de temps que de ressources en MdT. Ceci s'explique par la nécessité de récupérer ces informations en MLT déclarative dont l'accès est lent. Compiler les connaissances consiste en fait à passer de l'interprétation des connaissances déclaratives, à des procédures (ou productions), qui appliquent directement le savoir au problème. Deux mécanismes président à la compilation. La composition tout d'abord, qui, en réduisant l'ensemble des règles de production en une seule et unique, permet un gain de vitesse dans l'accomplissement de la tâche. Il s'agit donc d'un processus d'encapsulation. Il est possible d'y voir une certaine analogie avec la construction de schémas en mémoire ou encore le *chunking* (Cooper, 1998; Tricot, 1998). La procéduralisation ensuite, qui va permettre à l'individu de s'affranchir de l'activation en MdT des connaissances déclaratives en les incluant dans la procédure elle-même. Ce mécanisme présente l'avantage de sa grande économie cognitive, mais en contrepartie, perd toute adaptabilité, ce qui n'est pas encore le cas avec la seule composition.

3.3.1.2. Critique de la compilation de connaissances

Ce modèle proposé par Anderson a depuis connu validations empiriques (Lee & Anderson, 1997) et développements (Taatgen & Lee, 2003), notamment grâce l'architecture cognitive ACT-R (*Adaptive Control of Thought-Rational*) (Anderson et al., 2004).

On pourra notamment s'interroger sur l'opportunité de considérer ce modèle comme pertinent pour rendre compte de l'acquisition de compétences motrices, et plus

particulièrement, graphomotrices. En effet, tel qu'Anderson l'expose (1982), rien ne permet de faire le rapprochement entre connaissances procédurales et motrices, à ceci près que dans les deux cas, il s'agit de connaissances considérées comme implicites (Thomas, 1997). Néanmoins, certains auteurs avancent que les habiletés motrices sont une émanation des habiletés cognitives. Dans une revue de question, Mounoud (1986) présente une série de travaux qui accréditent la thèse que « c'est par ce que nous expérimentons au travers de nos actions (ajustements empiriques) que nous élaborons de nouvelles connaissances, de même que c'est par ce que nous connaissons (représentations abstraites) que nos actions sont déterminées » (Mounoud, 1986, p.385). C'est pourquoi Thomas (1997) envisage la théorie d'Anderson (1982) sur la compilation de connaissances comme étant à même d'expliquer l'acquisition d'habiletés motrices. Le mouvement serait ainsi considéré comme la partie distale d'un processus mental plus profond qui le détermine en amont, comme le souligne Féry (2001).

En dépit de cela, une étude laisse entendre que ces connaissances procédurales motrices pourraient se former sans l'utilisation préalable de la MLT déclarative (Maxwell, Masters, & Eves, 2003). Un protocole de suppression de *feedback* dans une tâche d'apprentissage du *put*[11] au golf, montre qu'il entrave la formation de connaissances déclaratives. En situation de test, après cet apprentissage, les sujets n'ayant pas (ou moins) pu élaborer de représentations déclaratives, montrent de moins bonnes performances que ceux du groupe contrôle (apprentissage avec *feedback*). Mais à l'inverse, ils ne sont pas gênés par une tâche concurrente de MdT pendant la phase de test, ce qui laisse à penser que l'obtention de leur niveau de performance ne sollicite pas la MdT, et donc, pas de connaissances déclaratives. Ce type d'observation avait déjà été fait par Lieury (1997) dans une critique de la MdT de Baddeley, puisque il soulignait l'activation possible à court terme de la MLT par les techniques d'étude de la mémoire implicite. Cela confirme aussi que si la réalisation d'un mouvement implique une activité de contrôle, elle peut être sensible à toute autre activité concurrente, même si celle-ci est de nature différente. En fait, elle peut être sensible à toute augmentation de charge mentale, quels que soient les processus qui en sont à l'origine (Gaonac'h & Larigauderie, 2000).

Prenant acte de ces résultats, l'enjeu est de pouvoir fournir aux utilisateurs de PDA des connaissances nécessairement déclaratives (car seules pouvant être partagées), susceptibles de

[11] Le *put* au golf est le geste pratiqué pour pousser la balle dans le trou.

leur communiquer les informations utiles à la création en mémoire de procédures d'utilisation et de programmes moteurs relatifs à une certaine habileté, une certaine compétence dans l'exploitation de la machine.

3.3.2. Favoriser l'apprentissage procédural et moteur

La recherche sur les EIAH se penche depuis de longues années déjà sur les conditions de facilitation des apprentissages. Cette démarche a accordé une large part au rôle qu'y tiennent les illustrations. Ces travaux ont été conduits principalement dans le cadre de l'acquisition de connaissances déclaratives et procédurales (et spécialement pour la compréhension de documents à caractère scientifique ou technique), mais jamais motrices en tant que telles (pas à notre connaissances). Si motricité il y a, elle est abordée de façon incidente dans le cadre de l'étude scientifique de « l'action », qui renvoie plus à sa planification et à sa mémorisation (mémoire prospective) qu'à son exécution (Logie, Engelkamp, Dehn, & Rudkin, 2000). Pourtant, fournir des informations visuelles ou visuospatiales peut s'avérer crucial pour un apprentissage moteur puisque la motricité entretient d'étroits rapports avec l'imagerie mentale. C'est ce que rapporte Thomas (1997) dans une analyse du rôle conféré à l'apport de connaissances déclaratives lors d'un apprentissage moteur. Bien que ces dernières ne soient pas suffisantes à elles seules pour permettre un apprentissage moteur, et doivent s'accompagner nécessairement d'essais répétés par l'apprenant, leur utilité est aujourd'hui couramment admise (Chevalier, Hall, & Nadeau, 1990), notamment en début d'apprentissage, au stade déclaratif (Anderson, 1982; Thomas, 1997).

3.3.2.1. Image mentale et image motrice

Il existe trois types de relations entre imagerie mentale et motricité (Thomas, 1997). L'apport de l'image mentale à la réalisation d'un mouvement peut être de nature motivationnelle, psychophysiologique et cognitive. C'est ce dernier aspect qui retiendra toute notre attention, la problématique pouvant être posée en ces termes : comment améliorer l'acquisition d'un comportement moteur en favorisant la création en mémoire de représentations à partir desquelles il va pouvoir s'appuyer. Formulé autrement, il s'agit de comprendre les relations entre programmation motrice et imagerie mentale.

De fait, de nombreux résultats expérimentaux donnent à penser que la programmation motrice partage les mêmes propriétés que la représentation mentale de ce mouvement. Par exemple, penser le mouvement active les mêmes aires corticales et chaînes musculaires que le mouvement lui-même. Plus surprenant, imaginer une conduite motrice se traduit par des variations de l'activité du système végétatif (rythmes cardiaque et respiratoire) indexés sur l'intensité de l'activité représentée en mémoire (Féry, 2001). Par exemple, s'imaginer courir fait plus augmenter les pulsations cardiaques que s'imaginer marcher (sans pour autant faire augmenter la consommation d'oxygène, ce qui tend à limiter la portée de l'activation musculaire que nous évoquions dans les lignes précédentes). Pour finir, des épreuves de chronométrie mentale ont montré que se représenter un geste suit la même progression dans le temps que la réalisation du geste (Jeannerod, 1995).

Féry (2001), dans une revue de question, a traité de la récupération consciente des procédures motrices. Son analyse témoigne que verbaliser les conduites motrices n'est pas chose aisée. Cela ne veut cependant pas dire qu'il n'est pas possible d'y avoir accès. La difficulté ne réside en fait pas tant dans nos capacités à récupérer ce type d'informations, que dans celle à verbaliser et décrire nos mouvements. C'est pourquoi les images (statiques ou mieux encore, dynamiques), de par leur caractère analogique, sont plus à même de décrire nos procédures motrices que ne l'est le langage articulé.

Pour autant, Jeannerod (1995) préfère parler d'images motrices que d'images mentales. En effet, si l'introspection ne laisse aucun doute au sens commun quant à notre capacité à générer des images mentales (Fortis, 1994), les images motrices semblent pourtant ne pas leur être parfaitement assimilables. Pour étayer son propos, Jeannerod s'appuie sur la théorie du double codage, proposée par Paivio (1969). Ce modèle avance un certain nombre de résultats expérimentaux qui plaident en faveur de deux systèmes de traitement séparés : l'un verbal, l'autre imagé. En dépit de cette indépendance fonctionnelle, il est néanmoins possible de créer des « ponts » entre ces deux systèmes, de sorte qu'il est possible de décrire une image avec des mots, ou de dessiner un schéma à partir d'une description verbale. Or, il n'est justement pas possible (ou très difficile) de décrire une procédure motrice (Féry, 2001). C'est ce constat qui incite Jeannerod à parler d'image motrice plutôt que d'image mentale. Fondamentalement, images mentales et motrices sont des images, et c'est vraisemblablement ce caractère analogique qu'elles partagent. Néanmoins, l'image motrice contient en plus des informations de nature kinesthésiques, ainsi que portant sur la structure du but, ce qui lui confère sa

spécificité. Ces différences rendent compte en grande partie de la difficulté à verbaliser les comportements moteurs. Bien que ce rapprochement n'ait pas été fait dans la littérature, il nous semble que l'on puisse assimiler l'image motrice aux modèles mentaux (Johnson-Laird, 1983, 1993) ou de situation (Kintsch & Van Dijk, 1978) en ce sens qu'elle constitue une représentation analogique qui contient plus d'informations que sa seule traduction imagée, autorisant ainsi la production d'inférences.

Par conséquent, s'il n'est pas possible qu'un individu ait un accès conscient aux programmations motrices qui sous-tendent ses mouvements (puisque implicites), il a néanmoins la possibilité de s'en faire une représentation analogique : l'image motrice. Celle-ci va assurer deux fonctions principales :

- Tout d'abord, elle va participer à l'apprentissage de conduites motrices. En tant que représentation mentale, elle peut être manipulée, transformée, paramétrée, ajustée. Elle assure ainsi une fonction de simulation mentale qui permet à l'individu de ne pas seulement apprendre par essais et erreurs, mais également d'anticiper le résultat de modifications du programme moteur en amont de son exécution (Jeannerod & Frak, 1999). C'est justement, comme nous le soulignions, une forme d'inférence.
- Ensuite, elle va permettre le partage d'informations relatives à la séquence motrice avec des tiers. Bien évidemment, dans le cadre d'un partage de connaissances motrices avec autrui, la médiation ne se fait pas par le biais d'une illustration à proprement parler, mais plutôt de démonstrations de mouvements. Ainsi, voir quelqu'un exécuter un geste active les aires corticales qui sont associées à sa réalisation, de même que s'imaginer produire ce geste (Blake & Shiffar, 2007).

Ces mécanismes neuronaux sont très ancrés phylogénétiquement puisque la présentation d'objet à des singes[12] active le cortex prémoteur, aussi bien en présentation tactile que visuelle. Si ce résultat n'est pas surprenant dans le cas d'une stimulation tactile, l'activation des aires motrices sur la seule base d'une stimulation visuelle laisse entendre que les neurones du cortex prémoteur traduisent directement l'image visuelle en action motrice (Murata et al., 1997). En ce qui concerne l'humain, et plus particulièrement l'écriture, la simple présentation visuelle de lettres isolées à des sujets droitiers provoque l'activation d'une zone corticale prémotrice gauche, alors même qu'aucune réponse motrice n'est requise. Cette zone ne

[12] *Macaca Nemestrina*, macaque à queue de cochon

s'active pas lorsque des pseudo-lettres (caractères inconnus) sont présentées. En revanche, la même zone s'active fortement lorsque les sujets doivent écrire ces lettres. Cela suggère que chez l'adulte la représentation centrale des caractères intègre une composante motrice. Mais rien ne permet d'affirmer que l'activation de cette aire prémotrice joue un rôle dans la reconnaissance des lettres vues (Longcamp, Anton, Roth, & Velay, 2003). Pour faire la lumière sur cette question, Longcamp et son équipe ont comparé deux modes d'apprentissage de lettres inconnues (issues des alphabets Bengali et Guajarati), par écriture manuscrite et écriture au clavier. Les résultats montrent que les caractères appris à la main sont mieux reconnus en présentation miroir que ceux appris par dactylographie (Longcamp et al., 2008; Longcamp, Boucard, Gilhodes, & Velay, 2006).

Cela montre donc comment l'imagerie mentale peut tout autant permettre l'accès à la représentation du mouvement que participer à la création de celui-ci.

Dès lors se pose le problème de l'accès à ce type d'information en l'absence de démonstrateur (humain, ou avatar (Rickel & Johnson, 1999)). Si ce que Venkatesh et ses collaborateurs (2003) ont désigné comme étant des « conditions facilitantes » de l'acceptation des systèmes, parmi lesquelles l'assistance de tiers, les IHM ambitionnent le plus souvent de s'auto-suffire pour ce qui relève des aides et tutoriels. A défaut de pouvoir remplacer un démonstrateur humain, des solutions basées sur la production d'illustrations se sont faites jour, profitant de leur caractère justement analogique pour générer des représentations mentales qui le sont aussi (Denis & De Vega, 1993). Leur emploi se fonde sur les explications que nous venons de fournir, et la propension naturelle de chacun à apprendre par imitation (Blake & Shiffar, 2007; Thomas, 1997). Ces illustrations peuvent être de deux types suivant qu'elles soient statiques ou dynamiques.

3.3.2.2. Comparaison des illustrations statiques et dynamiques

L'étude des rapports qu'entretiennent entre elles visualisations internes (images mentales) et visualisations externes (illustrations) est relativement récente. Cela tient au fait que seules les dernières avancées dans les technologies informatiques ont permis de produire des illustrations véritablement efficaces (Jamet, Le Bohec, & Hidrio, 2003). Ce courant de recherche s'est principalement orienté vers la compréhension de processus ou de documents scientifiques et techniques. La question de l'étude de la motricité se cantonne, lorsqu'il s'agit

de sujets sains adultes, à la psychologie du sport. Le rôle que les illustrations peuvent jouer dans le cadre de tels apprentissages (pratique sportive) est plus spécifique, souvent basé sur des photographies ou des vidéos, qui renvoient alors à un type particulier d'illustrations, s'inscrivant cependant dans la distinction statique/dynamique que nous évoquions précédemment. Néanmoins, comme le souligne Thomas (1997), les recherches concernant l'apprentissage de procédures en général peuvent être décalquées sur les apprentissages moteurs qui en constituent un cas particulier.

Bien que circonscrits aux animations mentales de dispositifs techniques, les travaux de Hegarty et Sims sont instructifs quant aux relations entre imagerie mentale et illustrations. Si Jeannerod (1995; 1999, avec Frak) parle d'image motrice, la composante motrice prête bien évidemment un caractère dynamique à ces « images ». La connaissance des animations mentales ne peut donc que bénéficier à celle des images motrices, selon nous, puisque Thomas (1997) fait le rapprochement entre acquisition de connaissances motrices et procédurales en s'appuyant sur le modèle d'Anderson (1982). Hegarty et Sims proposent quatre types de relations possibles entre visualisations internes et externes.

Brièvement, elles définissent une visualisation externe comme un dispositif visuospatial dans lequel les informations sont véhiculées par l'arrangement spatial des éléments de la représentation (Hegarty & Sims, 1994; 1997). Une illustration, peut être aussi, pour peu qu'elle soit présentée sur un écran, dynamique. Elle peut représenter tous types de phénomènes, du plus concret (*e.g.* moteur) au plus abstrait (*e.g.* flux d'informations sur le *Web*). Les visualisations peuvent être tout à la fois internes à la cognition d'un individu (en tant qu'image mentale) ou alors être un artefact imprimé sur papier ou affiché sur un moniteur et potentiellement visible par quiconque. Cependant, précisent-elles encore, si nos capacités de visualisations internes n'ont pas changé significativement dans un passé récent, les nouvelles technologies ont permis d'augmenter considérablement nos possibilités de créations d'images externes. C'est ce qui explique selon elles que la recherche leur porte une attention toute particulière. Il y a par ailleurs une importante tradition de recherche sur l'imagerie mentale grâce à la psychologie cognitive. Pour autant, l'étude des animations mentales est plus orientée vers la nature même du système d'imagerie interne que du rôle qu'elles tiennent dans les activités de pensée et de raisonnement.

S'agissant des rapports qui lient les visualisations internes et externes, Hegarty et Sims proposent qu'une première possibilité soit que les représentations externes puissent se

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

substituer à la production de représentations internes. Tout un chacun peut produire une illustration d'un phénomène (un schéma par exemple), de sorte qu'une autre personne la voyant aura la même visualisation interne que celle de l'individu qui l'a créée. Si c'est le cas, une image peut servir de « béquille » pour des personnes qui ont de faibles aptitudes visuospatiales. Une deuxième possibilité est que les illustrations externes améliorent la qualité des visualisations internes par complément d'informations, qui s'ajouteraient alors à celles déjà fournies. Par exemple, une visualisation externe peut représenter un dispositif plus complexe que ce qui peut mentalement être visualisé du fait de la capacité limitée de la MdT. La troisième possibilité est que l'aptitude à produire des images mentales est un préalable au traitement et à la compréhension d'illustrations externes. Dans ce cas, le bénéfice éventuel de ces dernières serait subordonné aux possibilités de chacun en la matière, variant au gré des différences intra- et interindividuelles. Sur ce point, Mary Hegarty est rejointe par Gyselinck (1996). Pour finir, les auteurs évoquent aussi l'idée selon laquelle l'exposition aux visualisations externes peut améliorer les habiletés d'illustrations internes. Elles montrent que ces différentes possibilités ne sont pas mutuellement exclusives.

Par conséquent, cette approche fait sienne l'idée selon laquelle les mots sont parfois insuffisants ou inadaptés à la description de certains processus, notamment moteurs (Féry, 2001). C'est pourquoi, de par le caractère analogique des illustrations, proche de celui de certaines représentations, elles sont à même de faciliter leur élaboration mentale (Denis & De Vega, 1993; Glenberg & Langston, 1992). Ceci a été mis en évidence notamment dans le cadre de l'apprentissage d'une procédure de saisie de requète sur un moteur de recherche (Hidrio, Jamet, & Le Bohec, 2002; Jamet, Le Bohec, & Hidrio, 2003). Cependant, si l'on n'a pas à douter de l'apport des illustrations aux apprentissages procéduraux, et plus particulièrement aux apprentissages moteurs, il reste difficile de savoir, parmi les illustrations dynamiques ou statiques, celles qui sont les plus efficaces. C'est en tout cas ce que confirme une synthèse menée par Bétrancourt et Tversky (2000). En passant au crible douze études sur la comparaison entre illustrations statiques et dynamiques, elles rapportent que seules sept d'entre elles témoignent d'une supériorité des illustrations dynamiques sur les illustrations statiques. En conclusion, Bétrancourt et Tversky soulignent tout de même le manque de cadre théorique et de réflexion de fond sur les processus cognitifs mis à l'œuvre dans la compréhension des animations. De surcroît, un certain nombre de travaux ont pu manquer de rigueur sur le plan méthodologique. L'expérience de Jones et Scaife (1999) en est un

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

exemple : elle compare illustrations statiques sur papier et animations sur ordinateur. Si ces auteurs n'observent pas de différences entre les deux conditions, il est néanmoins permis de s'interroger sur la validité du résultat, les supports de présentation de l'information étant différents. Cette synthèse s'est vue complétée par Bétrancourt et ses collaborateurs qui ont montré que dans la plupart des expériences menées où les illustrations dynamiques se montraient plus efficaces que celles statiques, la quantité d'informations était différente d'une condition à l'autre (très souvent, l'animation donne des informations supplémentaires, ce qui peut justifier sa supériorité) (Bétrancourt, Bauer-Morrison, & Tversky, 2001). Une solution pour pallier à cela est l'utilisation d'illustrations statiques fléchées. Assez curieusement, il semble que la recherche en psychologie sur le rôle des illustrations ait délaissé celui tenu par l'ajout de flèches aux schémas explicatifs. A cela, nous invoquons une raison : il est à peine concevable de produire une illustration didactique digne de ce nom sans y avoir recours. De plus, elles peuvent tenir des fonctions différentes et ce, indépendamment de leur polymorphisme. En effet, elles peuvent tour à tour focaliser l'attention de l'apprenant sur une zone précise du document, témoigner d'un flux ou bien encore d'un déplacement, d'une pression… C'est de ce constat que découle la difficulté à formuler précisément le rôle qui incombe aux flèches dans une image explicative. Pourtant, parmi les rares travaux consacrés au rôle tenu par les flèches, certains se sont penchés sur la question en termes de comparaison de la quantité d'information véhiculée entre illustrations statiques fléchées et illustrations dynamique. Là encore, les résultats sont mitigés. S'il apparaît que la présence de flèches sur des illustrations statiques améliore leur compréhension par rapport à des illustrations non fléchées pour Gyselinck (1995) ou Mautone et Mayer (2001), d'autres auteurs concluent quant à eux qu'une illustation statique ou statique fléchée (les flêches représentant un mouvement) n'apporte aucun bénéfice à la compréhension du fonctionnement d'un moteur non illustré, contrairement aux animations (Hidrio & Jamet, 2002, 2008; Jamet, Le Bohec, & Hidrio, 2003).

Au final, l'efficacité des animations serait due à plusieurs effets, parmi lesquels le caractère dynamique de l'apprentissage (impliquant du mouvement ou l'apprentissage d'une procédure). Cette condition au bénéfice des animations par rapport aux simples illustrations est reportée par de nombreux auteurs (Park & Hopkins, 1993). Pour autant, cette condition, si elle est nécessaire, n'est pas forcément suffisante, comme en témoigne une étude sur l'apprentissage du montage d'un saxophone pour lequel la présence d'animations ne se révèle

pas plus efficace que celle d'illustrations statiques (Zacks & Tversky, 2003). En dépit de ces incertitudes, il ne faut pas pour autant nier les avantages à tirer des animations, ne serait-ce qu'en termes de motivation et d'intérêt suscité chez l'apprenant (Hidrio & Jamet, 2008).

Quoi qu'il en soit, on ne saurait réduire un apprentissage aux conditions matérielles de sa réalisation. C'est d'abord et avant tout un processus psychologique complexe. La théorie de la charge cognitive proposée par Sweller (1988) apporte d'importants éclairages sur la façon dont se construit l'apprentissage et permet justement, d'en faciliter l'élaboration du support matériel.

3.4. La théorie de la charge cognitive (CLT, Sweller, 1988)

La théorie de la charge cognitive (*Cognitive Load Theory, CLT*, 1988) relève autant des apprentissages éducatifs que de la résolution de problème. C'est pourquoi elle se révèle tout-à-fait pertinente dans le cadre de la manipulation d'une IHM si tant est que l'on considère une tâche à accomplir comme un problème à résoudre (Drommi, Ulferts, & Shoemaker, 2001). Dans le cadre de l'accomplissement de cette tâche peut se greffer, de façon incidente ou volontaire, un apprentissage du fonctionnement du dispositif, de sorte que l'opérateur acquiert progressivement une certaine maîtrise de son fonctionnement.

En consacrant ses recherches à l'apprentissage et à la résolution de problèmes depuis plus de trente ans, John Sweller a développé la CLT. Celle-ci a suscité une audience considérable et connu d'importants développements (Clark, Nguyen, & Sweller, 2006; Paas, Renkl, & Sweller, 2003; Sweller, 1988; Tricot, 1998 ; pour une présentation des travaux de Sweller). La CLT s'attache à expliquer et à améliorer la conception de matériels éducatifs. Elle a pour objectif de guider les concepteurs de supports pédagogiques, que ceux-ci soient, ou non, électroniques et/ou interactifs. « La théorie de la charge cognitive a été conçue pour fournir des recommandations visant à améliorer la présentation des informations de telle façon qu'elle facilite les processus d'apprentissage et optimise les performances intellectuelles » (Sweller, Van Merriënboer, & Paas, 1998 ; notre traduction). Cette définition n'exclut en rien la manipulation et l'apprentissage d'une interface. C'est pourquoi la CLT éclaire la théorie d'Anderson (1982) sur la compilation de connaissances, puisqu'elle postule elle aussi que l'accomplissement d'une tâche passe par une analyse de but de type moyen-fin. Cependant, elle aborde cela en tant que facteur de charge pesant sur la MdT. Anderson (1882) fait la

même observation au stade déclaratif. La CLT propose donc de réduire cette charge pour faciliter l'apprentissage de la procédure de résolution du problème (ou d'accomplissement de la tâche).

3.4.1. La charge cognitive

La charge cognitive trouve son origine dans une MdT (à capacité limitée) pouvant manipuler à la fois des informations issues de la perception et de la MLT (dont la capacité est elle, illimitée). Dans ce cadre, la charge cognitive renvoie aux coûts des processus mentaux pesant sur la MdT au cours de la réalisation d'une tâche donnée (Sweller, Van Merriënboer, & Paas, 1998). Par conséquent, tout traitement ou schéma consciemment activé en MdT et qui n'est pas automatisé (ou pré-attentionnel (Treisman & Kanwischer, 1998)) pèse en tant que charge sur cette dernière. Cette charge, dite « cognitive », résulte du fonctionnement normal de la MdT, mais ne devient problématique que lorsqu'elle excède les ressources disponibles de celle-ci (Paas, Renkl, & Sweller, 2004; Sweller, Van Merriënboer, & Paas, 1998).

La contribution fondatrice de Sweller à sa théorie est d'avoir mis en évidence qu'un individu confronté à un nouveau problème à résoudre va adopter une stratégie de résolution dite de « moyen-fin ». Ce type de stratégie, en nuisant à la création de schémas en MLT, s'avère contre-productif pour l'apprentissage de la procédure de résolution. Cela se traduit par le fait qu'un sujet novice, confronté à un problème nouveau, va consacrer une partie des ressources de sa MdT à maintenir en mémoire (1) le but final et les sous-buts qu'il s'est fixés, (2) l'état d'avancement de la réalisation de sa tâche, ainsi que (3) les opérateurs susceptibles de le conduire à la prochaine étape de résolution. Ce volume de MdT, ainsi occupé, n'est plus disponible pour créer et assimiler de nouveaux schémas qui justement, faciliteraient la résolution du problème. Ce constat va inspirer à Sweller et ses collaborateurs que la charge cognitive est le fait des contraintes cumulées de trois types de charges :

- **la charge intrinsèque**, relative à la difficulté même de la tâche. On ne peut la réduire, car elle dépend en fait du degré d'interactivité qu'il existe entre les éléments du problème à résoudre. De l'interactivité des éléments va en fait dépendre le nombre d'éléments à traiter simultanément en MdT (Chandler & Sweller, 1991; Clark, Nguyen, & Sweller, 2006) ;

- **la charge inutile** ou extrinsèque, qui renvoie à la façon dont l'environnement d'accomplissement de la tâche présente les informations. Dépendant du format de présentation des informations, la charge extrinsèque est induite par le traitement du support d'apprentissage ou de réalisation de la tâche (une IHM par exemple). Elle est par conséquent inutile et même nuisible à la compréhension et à l'apprentissage. La CLT propose justement des guides de conception d'interfaces visant à minimiser la charge extrinsèque (Chandler & Sweller, 1991; Clark, Nguyen, & Sweller, 2006) ;
- **la charge utile** ou *germane cognitive load*, correspondant aux ressources dévolues au traitement, à la construction et à l'automatisation des schémas. C'est justement pour favoriser l'allocation de ressources à ces processus qu'il faut réduire la charge inutile (Sweller, Van Merriënboer, & Paas, 1998).

L'objectif de la CLT est de faire en sorte que la quantité cumulée de ressources demandées par la réalisation de la tâche n'excède pas la quantité limitée de ressources en MdT dont dispose un individu (Cooper, 1998).

Il est à noter que le niveau d'expertise et de connaissances préalables des individus lors de l'interaction a un impact sur la charge. La complexité de la tâche va donc aussi dépendre des personnes, suivant qu'elles disposent ou pas de schémas élaborés et automatisés dans le domaine considéré. Par conséquent, on ne peut appréhender la difficulté d'un apprentissage sans se préocupper des connaissances antérieures des individus qui y sont soumis. C'est la question de l'expertise et de la possibilité de mettre à profit des connaissances antérieures pour en acquérir de nouvelles.

3.5. Expertise et adaptabilité

Pour Cooper (1990), l'apprentissage est le processus qui conduit un sujet novice à devenir expert dans un domaine donné. Or, la charge intrinsèque engagée par la réalisation d'un apprentissage dépend du niveau d'expertise du sujet. Il convient donc de clarifier la notion d'expertise. De surcroît, définir l'expert permet aussi, par exclusion, de définir le novice, ce qui est capital pour pouvoir envisager de tester l'apprenabilité d'une interface. Pourtant, personne n'est vierge de connaissances. Gaston Bachelard écrivait qu'on « connaît contre une

connaissance antérieure »[13]. Cette assertion reste plus que jamais d'actualité en regard des connaissances de la psychologie cognitive sur l'apprentissage, puisqu'elle pose implicitement la question de l'adaptabilité cognitive, c'est-à-dire la capacité à mobiliser des connaissances antérieures non spécifiques à la réalisation d'une tâche dans un contexte nouveau.

3.5.1. L'expertise en psychologie cognitive

Les premiers travaux sur l'expertise portaient exclusivement sur les joueurs d'échecs. Bien que le concept d'expertise soit désormais étudié dans des situations qui dépassent ce cadre restreint, ces premiers travaux ont eu le mérite de montrer que experts et novices ne se différenciaient pas sur les caractéristiques de leur MCT, mais sur leurs connaissances en MLT (Ericsson, 2005). Les travaux publiés depuis, en approfondissant la notion d'expertise, en ont aussi accentué le flou, en montrant parfois que les performances des « experts » pouvaient s'avérer moins bonnes ou équivalentes à celles de sujets non-experts ou novices (Ericsson, 2005; Farrington-Darby & Wilson, 2006). A l'inverse, d'autres recherches ont apporté sur l'expertise des éclairages novateurs et dont on ne mesure que depuis peu le potentiel heuristique, telles que la théorie de la mémoire de travail à long terme (MdT-LT, Ericsson & Kintsch, 1995).

Pour autant, l'étude de l'expertise en psychologie pâtit sans doute plus du manque de définition de ce qu'est l'expertise, que des processus qu'elle engendre et des mécanismes qui permettent d'y parvenir (Ericsson, Krampe, & Tesch-Romer, 1993; Ericsson & Lehmann, 1996; Van Merriënboer & Sweller, 2005). Cela est éminemment problématique pour nous qui évoquons autant l'expertise pour qualifier l'automatisation du processus d'exécution graphomotrice, que pour désigner la finalité d'un apprentissage procédural visant à assimiler le fonctionnement d'une IHM (MacKenzie & Soukoreff, 2002).

3.5.1.1. Définir l'expertise

Définir l'expertise n'est pas chose aisée. Selon Farrington-Darby et Wilson (2006), ce concept, à l'instar de l'utilisabilité, est employé par divers champs disciplinaires qui lui confèrent des définitions pouvant grandement diverger de l'un à l'autre. Ces auteurs, dans une

[13] *La formation de l'esprit scientifique*, 1938. Paris : Librairie Philosophique J. Vrin. (p.13)

revue de question, proposent un cadre théorique pour appréhender la notion d'expertise. En termes de performance tout d'abord, puisque l'expertise peut donner lieu à des performances sociales (notamment prise de décision et diagnostic en milieux professionnels et institutionnels), cognitives, ou physiques. Farrington-Darby et Wilson font remarquer que la distinction entre performance cognitive et physique n'est que peu pertinente, puisque ces deux dimensions sont réciproquement adossées l'une à l'autre (Mounoud, 1986). D'autre part, l'expertise « sociale » renvoie aussi à des aspects cognitifs (cognition distribuée, passant notamment par l'usage généralisé de systèmes socio-techniques requérant autant des aptitudes sociales que cognitives). C'est pourquoi notre approche de l'expertise n'exclura aucune de ces catégories. En ce qui concerne les disciplines qui manipulent la notion d'expertise, les auteurs extraient de leurs lectures l'intelligence artificielle, les sciences économiques et statistiques, et pour finir, les sciences cognitives. Là encore, admettre un tel partage est peu pertinent tant il est vrai que ces champs de recherche peuvent se recouvrir mutuellement.

Malgré cela, nous nous attacherons à aborder l'expertise sous l'angle de la psychologie cognitive. A cela deux raisons : tout d'abord, il est évident que la psychologie cognitive constitue le cadre théorique général de notre recherche. Ensuite, les auteurs rapprochent de ce domaine d'étude, et à raison, celui de l'ergonomie. C'est donc un choix théorique opportun, d'autant qu'il ressort de cette revue de question que la perspective adoptée en psychologie cognitive est résolument explicative, là où l'approche économique / statistique vise principalement l'augmentation du potentiel décisionnel d'un individu ou d'une institution.

Farrington-Darby et Wilson étendent cette distinction à celle existant entre micro- et macrocognition. Les aspects microcognitifs relèveraient de processus très concrets et circonscrits (fonctions exécutives, mémoire, construction de puzzles, *etc.*) tandis que la macrocognition s'inscrirait dans des processus plus larges et abstraits (planification, processus psychosociaux, attention, *etc.*). Ce partage entre déterminismes et processus cognitifs est également étendu au lieu de recherche : terrain pour les processus décisionnels abstraits et laboratoire pour les processus cognitifs concrets. Cette considération n'est pas sans intérêt pour nous, puisque si faire le choix d'accepter d'utiliser un dispositif de saisie de texte manuscrit est un processus décisionnel ancré dans le réel, la mesure de l'utilisabilité (au premier rang de laquelle, l'apprenabilité ou l'efficacité) est, elle, associée au laboratoire. On retrouve donc ici la ligne de partage qu'il existe entre les champs MIS et HCI (Zhang & Dillon, 2003).

Par conséquent, et en dépit de l'orientation cognitiviste du traitement que nous désirons accorder à l'expertise, il semble que notre problématique ne nous permette pas de faire l'économie d'une vision globale de ce sujet.

3.5.1.2. Définir l'expert

A défaut de pouvoir définir l'expertise, et comptant envisager cette notion dans son intégralité, il nous semble plus pertinent de tenter de définir l'expert. Bien que cela soit à l'avenant ambigu, un certain nombre de critères et méthodes ont été décrits pour cela. Farrington-Darby et Wilson en retiennent en cinq :

- l'expérience (nombre d'années de pratique dans la discipline considérée (Ericsson, Krampe, & Tesch-Romer, 1993)) ;
- la situation socioprofessionnelle (un professeur est considéré expert par rapport à ses élèves) ;
- la spécialisation (c'est la circonscription du champ de compétence d'un individu : plus il est réduit, plus on considère que celui-ci en est expert) ;
- la qualification (par diplôme ou autorisation, quelqu'un peut être considéré comme expert) ;
- le pré-test (l'utilisation d'un pré-questionnaire avant épreuve peut permettre de déterminer un ou plusieurs groupes d'individu dont le niveau d'expertise varie) ;
- la sélection par les pairs (elle renvoie en fait aux mêmes difficultés que la sélection par la situation ou la qualification, à ceci près qu'intervient un jugement extérieur).

Conscients des limites de ces méthodes de catégorisation des experts, les auteurs envisagent qu'il est possible de mesurer l'expertise sur la base de processus cognitifs. La psychologie est évidemment pour beaucoup dans le développement de ce type de méthodes. Cela n'est pas sans poser de problèmes, puisque la distinction entre processus micro- et macro-cognitifs reste plus que jamais cruciale. Or, il semble que les conditions d'évaluation de l'expertise revêtent une importance capitale, les études de terrain y étant plus favorables. Néanmoins, et concernant des habiletés psychomotrices, il apparaît que le laboratoire puisse

constituer un lieu d'étude approprié. Pour nous qui nous intéressons à l'écriture et sa production graphique, ce constat est de bon augure (Farrington-Darby & Wilson, 2006).

En effet, le scripteur adulte est souvent considéré comme scripteur « expert », dans le sens où l'exécution motrice de l'écriture est pour lui normalement automatisée. La question est donc de savoir comment ces connaissances acquises et fortement ancrées relevant du geste d'écriture peuvent être mises à contribution pour faciliter l'apprentissage d'une nouvelle écriture, potentiellement identifiable par un logiciel. De surcroit, notre propension à vivre dans un environnement de plus en plus informatisé invite à se demander si nos connaissances en informatique ne sont pas un gage de réussite à appréhender le fonctionnement de nouvelles applications. Ces interrogations relèvent du champ de l'adaptabilité cognitive, c'est-à-dire de notre capacité à mettre à profit des connaissances antérieures dans un apprentissage nouveau.

3.5.2. Adaptabilité

3.5.2.1. Transfert de connaissances

Le transfert de connaissances est une thématique de recherche assez ancienne en psychologie, puisqu'elle a fait l'objet d'études dès le début du siècle sous les plumes de Thorndike et Woodworth (1901). A l'époque, ils entrainaient des sujets à des tâches d'estimations de surface de formes prédécoupées (rectangles par exemple) et leur faisaient ensuite estimer les aires d'autres formes. Bien que ces auteurs n'aient pas réussi à démontrer de façon claire la possibilité de transmettre un apprentissage donné à une situation nouvelle, l'idée de transfert d'apprentissage était née. Cette notion concerne la recherche sur l'expertise, car, si les experts excellent dans leur domaine de prédilection, il arrive souvent qu'ils n'en tirent aucun avantage dans des situations nouvelles (Perkins & Salomon, 1994). Elle concerne aussi l'expertise dans son acception la moins « noble » : le scripteur « expert », celui qui a automatisé le processus d'exécution graphomotrice, est-il capable de mettre en application ses connaissances graphomotrices pour utiliser au mieux un logiciel de reconnaissance de l'écriture ? Elle a trait également aux novices, car jamais un nouvel apprentissage ne se fait sans transfert d'un apprentissage précédent (cf. Bachelard et d'une manière générale, l'élaboration de schémas à partir d'autres schémas (Cooper, 1998; Tricot, 1998)). C'est donc un point central en psychologie des apprentissages.

Contrairement à ce que l'intuition pourrait laisser penser, le transfert de connaissances peut être aussi bien positif (l'apprentissage antérieur améliore l'apprentissage suivant) que négatif (l'apprentissage suivant est parasité par l'apprentissage précédent). C'est même souvent ce dernier cas qui est le plus souvent observé (Perkins & Salomon, 1994).

Deux paramètres entrent en jeu dans le transfert de connaissances : la distance tout d'abord, c'est-à-dire la différence entre le contexte initial d'apprentissage et le domaine d'application de ces connaissances (par exemple, si l'on bénéficie tous du même apprentissage scolaire, on en use dans des domaines fort variés). Le mécanisme ensuite, suivant que le transfert se fasse par l'activation de routines bien établies (automatiques ou semi-automatiques, *low road*) ou par la médiation d'un effort conscient de planification pour puiser dans ses connaissances la solution à un problème nouveau (*high road*). Il est à noter que la notion de « distance de transfert » est tout-à-fait imprécise, et très difficile à mesurer objectivement. Il semble pourtant que les mécanismes *low road* s'exercent pour des transferts courts (forte similarité contextuelle) alors que les mécanismes *high road* permettraient des transferts de plus longues distances. Ceci est tout particulièrement marqué chez les experts qui peinent à exprimer leurs compétences dans des domaines éloignés du leur. Néanmoins, si fortes soient les connaissances « locales » d'un expert, elles ne laissent en rien présager de l'existence, chez le même individu, de connaissances plus générales, et donc, plus transférables.

3.5.2.2. Différences entre novices et experts :

En regard de ce dont nous avons fait état dans les lignes précédentes, il apparaît plus pertinent de décrire les experts par rapport aux novices sur la base de processus ou d'aptitudes mentales qui différeraient d'un groupe à l'autre. Pour la CLT, la principale différence entre experts et novices tient au nombre et au niveau d'automatisation des schémas (relatifs au domaine considéré) en MLT. Par conséquent, les schémas dont dispose l'expert vont lui permettre, en le renvoyant à des situations de problèmes déjà rencontrés et résolus, de s'orienter vers le choix de la procédure appropriée. Celle-ci serait mise en œuvre de façon automatique, ce qui assurerait la rapidité du traitement. Les différences entre novices et experts ne s'arrêtent cependant pas à cela. Si les capacités de MCT des experts ne diffèrent pas de celles des novices (pas plus que toute autre mesure psychométrique, telle que le

quotient intellectuel, par exemple), c'est par contre la possibilité pour les premiers d'utiliser une partie de leur MLT (relative au domaine de spécialité) comme une mémoire de travail qui change (Ericsson & Kintsch, 1995). Ainsi, l'encodage d'informations nouvelles et la récupération de connaissances anciennes y seraient aussi rapides (ou du même ordre) que pour la MdT. Cela peut expliquer ainsi une augmentation capacitaire du volume en MCT, dont les mesures d'empan classiques ne rendaient pas compte (Guida, Tardieu, & Nicolas, sous presse).

A l'inverse, les novices n'ont que peu de schémas, ce qui entrave leur identification des structures de problèmes autres que faciles, ou susceptibles de faire appel à des expériences antérieures de même type. Ces difficultés s'ajoutent au fait que, une fois la solution identifiée, sa mise en œuvre reste laborieuse (procédure non automatique et médiation verbale témoignant d'une surcharge de la MCT s'accompagnant d'erreurs (Anderson, 1982; Cooper, 1998)).

Dans une perspective évolutionniste Darwiniste, il serait même permis d'envisager ces schémas automatisés comme des structures de connaissances sélectionnées naturellement, car adaptées (Ericsson & Lehmann, 1996; Sweller, 2004). Cette approche va même jusqu'à les envisager comme une forme d'administration centrale fondée sur l'organisation des connaissances en MLT (Sweller, 2004). D'autres travaux accréditent cette thèse en soutenant l'idée que l'expertise de certains dans un domaine ne serait pas le fait d'un talent ou d'un don, mais uniquement la conséquence d'un travail délibéré, fourni et répété (Ericsson, Krampe, & Tesch-Romer, 1993; Ericsson & Lehmann, 1996). Au-delà des débats spécialisés, on mesure toutefois l'importance qu'il y a à faciliter l'élaboration de ces schémas, ce qui nécessite un volume de MdT substantiel. C'est tout l'enjeu porté par la diminution de la charge cognitive d'un individu en situation d'apprentissage. Cela implique cependant d'être capable de mesurer cette charge afin de pouvoir en apprécier objectivement le niveau.

3.6. Mesurer la charge cognitive

Il est possible d'évaluer la charge cognitive à partir de mesures des concepts de charge mentale, d'effort mental et de performance. C'est l'ensemble de ces indicateurs qui va permettre de donner une estimation de la charge cognitive (Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

3.6.1. Charge mentale

Selon Paas et son équipe, la charge mentale « est la dimension de la charge cognitive qui trouve son origine dans les caractéristiques de la tâche et des individus. […] La charge mentale peut être mesurée sur la base des connaissances préalables quant à la tâche à réaliser, et sur des caractéristiques individuelles. Ainsi, elle peut fournir une indication sur la demande en ressources cognitives et peut être considérée *a priori* comme une estimation de la charge cognitive » (Paas, Tuovinen, Tabbers, & Van Gerven, 2003, notre traduction).

Le concept de charge mentale est fort employé en psychologie, et particulièrement en psychologie ergonomique. Elle peut faire l'objet de définitions aussi variées que le sont les méthodes de mesure dont elle se prévaut. Cela explique qu'elle puisse prendre des noms différents. Qu'il s'agisse de charge mentale (Paas, Tuovinen, Tabbers, & Van Gerven, 2003), charge de travail (Bastien & Scapin, 2001), ou charge de tâche (Hart & Staveland, 1988), Tricot et Chanquoy (1996) n'hésitent pas à faire le rapprochement entre ces notions dans une analyse de leurs définitions, leurs mesures, ainsi que de leurs utilités heuristiques. Ils signalent notamment que dans son acception ergonomique, la charge de travail (ou charge de tâche) implique aussi une composante physiologique et physique, tant dans la prise en considération des caractéristiques de la tâche que des mesures de performance à celle-ci (ex : fatigue, stress, *etc.*). Le flou qui accompagne la notion de charge – quels que soient les qualificatifs qui lui sont adjoints - fait d'elle l'objet de critiques nombreuses et fondées (Jamet, 2006; Tricot & Chanquoy, 1996). En dépit des incertitudes quant à la définition de charge, c'est un concept largement employé, bien que difficile à opérationnaliser. Il existe en effet beaucoup de méthodes d'évaluation de la charge. En tant que charge mentale, elle peut être appréhendée expérimentalement par différentes méthodes, reposant principalement sur l'administration de questionnaires *post-test*, afin de ne pas interférer avec la réalisation de la tâche en l'interrompant. Il existe différents questionnaires tels que le NASA-TLX que nous avons déjà présentés au cours du premier chapitre en tant que méthode d'évaluation de l'utilisabilité des IHM (Hart & Staveland, 1988; Rubio, Diaz, Martin, & Puente, 2004). Ces outils sont reconnus pour être très fiables et sensibles. En ce qui concerne leur validité, il a été mis en évidence qu'ils donnent une indication fidèle de la charge cognitive investie dans une activité. Néanmoins, on peut avantageusement substituer à ces méthodes d'évaluation multidimensionnelles des méthodes unidimensionnelles, beaucoup plus économiques à mettre en œuvre, tout en conservant les qualités de sensibilité et de fiabilité évoquées précédemment.

En effet, une seule échelle d'évaluation de la charge mentale administrée à un sujet à l'issue d'une épreuve suffit à donner en une estimation juste (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Ces techniques s'appuient sur l'idée selon laquelle les individus sont capables de sonder par introspection leurs processus cognitifs et d'en estimer l'effort mental requis (Delinières & Deschamps, 2000). Ce postulat est bien évidemment critiquable (Jamet, 2006).

Afin de faire face à cette objection, d'autres méthodes de mesure de la charge mentale ont été mises au point. Parmi elles, les méthodes dites psychophysiogiques et les tâches doubles. On pourra cependant prétexter qu'elles mesurent plus l'effort mental que la charge mentale, qu'on assimilera alors à l'effort mental moyen (Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

3.6.2. Effort mental et performance

L'effort mental renvoie à la quantité de ressources cognitives engagées à un moment donné dans la réalisation d'une tâche. On le mesure pendant que les individus sont en train de la réaliser. Mesurer l'effort mental passe principalement par l'utilisation de méthodes objectives qui, idéalement, n'interfèrent peu ou pas avec l'activité du sujet. Les enregistrements psychophysiologiques vont mesurer l'effort mental en s'appuyant sur des indices physiologiques qui témoignent de modifications de l'activité psychologique d'un individu. Le rythme cardiaque, la dilatation pupillaire, l'activité cérébrale ou la conductance électrodermale peuvent être appréciées, avec, à la clé, plus ou moins de succès. Si l'enregistrement du rythme cardiaque n'a pas bonne presse (car invalide, insensible et intrusive), celui de la dilatation pupillaire s'avère lui beaucoup plus intéressant (Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

La réalisation de la tâche peut elle aussi donner lieu à des mesures de performances qui constituent aussi un aspect de la charge cognitive. Or, certaines techniques d'évaluation de l'effort mental sont basées sur des mesures de performances à la tâche. Elles sont de deux types. On pensera en premier lieu à des mesures telles que l'accomplissement (c'est-à-dire le pourcentage de réalisation de la tâche), le nombre d'erreurs ou le temps de réalisation (Bevan & Macleod, 1994). Ce sont des mesures portant sur la tâche principale. Ces mesures peuvent être complétées par des mesures de performances à une tâche seconde. Cette deuxième tâche entre en concurrence avec la tâche principale, puisque toutes deux dépendent d'un unique

réservoir de ressources cognitives (Baddeley, 1986; McCutchen, 1996; Miyake & Shah, 1999). Par conséquent, les ressources mentales doivent se partager entre les deux tâches, ce qui diminue la quantité de ressources disponibles pour chacune d'elles. Si la qualité du traitement d'une tâche dépend des ressources disponibles, alors la performance à la tâche seconde doit diminuer d'autant plus que les ressources allouées à sa réalisation diminuent également (attendu qu'elles sont attribuées à la tâche principale) (Brünken, Steinbacher, Plass, & Leutner, 2002). Cette procédure est analogue à celle de Kellogg et fait l'objet des mêmes critiques, notamment en ce qui concerne son caractère intrusif (Piolat, Kellogg, & Farioli, 2001; Piolat, Olive, Roussey, Thunin, & Ziegler, 1999; Piolat, Roussey, Olive, & Farioli, 1996). En effet, si la tâche principale requiert déjà toute les ressources du sujet, alors la tâche seconde ne peut que venir interférer avec la tâche principale. Ceci est d'autant plus vrai si la tâche seconde implique aussi du raisonnement ou de la décision (Cox-Fuenzalida & Angie, 2005). Cependant, des travaux méthodologiques et des recherches expérimentales ont montré, dans les cadres de la CLT et de la procédure de Kellogg, l'intérêt de ce type de mesure de l'effort mental et son innocuité sur la tâche principale (Bonnardel & Piolat, 2003; Brünken, Steinbacher, Plass, & Leutner, 2002; Piolat, Olive, Roussey, Thunin, & Ziegler, 1999).

Indépendamment du type de mesure de l'effort mental, les valeurs mesurées ne permettent pas de déterminer quelle est la part respective des différentes formes que peut prendre la charge cognitive (*i.e.* intrinsèque, inutile ou utile). Par contre, la moyenne des mesures d'effort mental donnera une estimation de la charge mentale. L'effort mental peut aussi être analysé au fil du temps. Les éventuelles fluctuations peuvent ainsi être mises en relation avec d'autres données enregistrées au même instant. Cela peut se faire *on-line* comme avec la méthode de triple tâche, auquel cas on s'intéresse à la tâche principale (par rétrospection dirigée) au moment où survient la tâche seconde (détection auditive) (Piolat, Kellogg, & Farioli, 2001; Piolat, Olive, Roussey, Thunin, & Ziegler, 1999; Piolat, Roussey, Olive, & Farioli, 1996). Cela peut aussi se faire de façon plus systématique, lorsque les mesures le permettent (dilatation pupillaire ou conductance électrodermale) (Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

3.6.3. L'efficience mentale

L'intérêt de ces mesures objectives de performances, aussi bien pour la tâche principale que pour la tâche seconde est de pouvoir donner lieu à des mesures d'efficience mentale. Paas et Van Merrienboër ont abondamment traité cette question, et leur contribution y est, à notre sens, essentielle. Mesurer la charge cognitive prend tout son sens lorsqu'on la met en relation avec les mesures de performance à la tâche réalisée. On comprend aisément que faire diminuer la charge inhérente à la réalisation d'une certaine tâche n'a d'intérêt que si cela ne réduit pas le niveau de performance. C'est ce rapport qui permet d'évaluer une interface : que permet-elle de faire, et pour quel effort consenti ? Cette nouvelle mesure de l'interface, dite, d'efficience mentale, apporte d'autres informations sur la charge cognitive que les mesures à partir desquelles elle est calculée, prises individuellement (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Paas et Van Merrienboër proposent un modèle mathématique pour cela (1993, voir formule 3). Sa particularité est de combiner les scores centrés et réduits (z) d'effort mental et de performance. Le score d'efficience mentale (E) y est déterminé par la distance orthogonale séparant un point (dont les coordonnées sont déterminées par les scores z de performance et d'effort mental) et la droite correspondant à $E = 0$. La formule est la suivante :

$$E = (z_{\text{Performance}} - z_{\text{Effort Mental}}) / \sqrt{2}$$

Formule 3: Mesure d'efficience mentale *E* (Paas & Van Merriënboer, 1993)

Il est plus facile de comprendre cette formalisation dans sa représentation graphique bidimensionnelle. Puisque les scores sont standardisés, l'ensemble des points déterminés par la mise en relation de l'effort mental et de la performance s'inscrit dans un carré, dont une diagonale représente l'efficience mentale $E = 0$ (comme le montre la figure 18, page suivante).

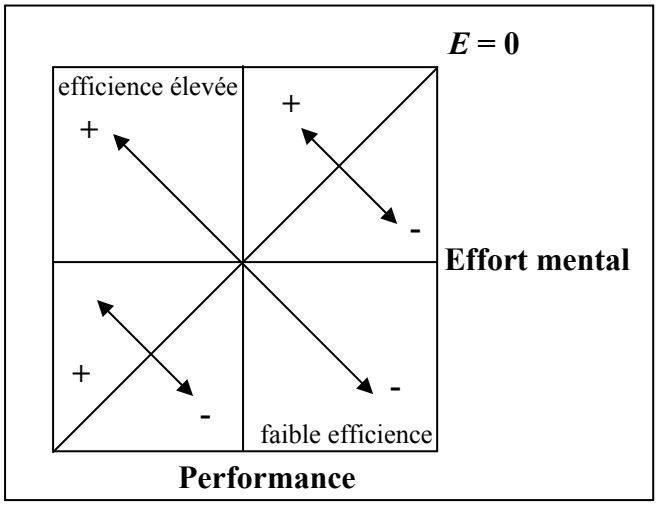


Figure 18: Représentation graphique de l'efficience mentale (d'après Paas, Tuovinen, Tabbers & Van Gerven, 2003)

Les scores *z* de performance et d'effort mental sont représentés sur un repère orthonormé. L'axe des ordonnés exprime la performance, tandis que celui des abscisses représente l'effort mental. L'efficience mentale est calculée par la formule *E* qui donne la distance orthogonale séparant un point donné de la diagonale $E = 0$. L'ensemble des points qui la composent renvoie en fait à un niveau de performance associé à un niveau d'effort mental pour lesquels on estime qu'il n'y a ni bénéfice, ni déficit. L'efficience y est considérée comme nulle. Si l'efficience est positive ($E > 0$), alors cela signifie que les performances sont bonnes en regard de l'effort mental investi. A l'inverse, une efficience faible se traduira par un score négatif ($E < 0$).

La formule de Paas et Van Merriënboer (1993) est dérivée de la formule générale de calcul de la distance orthogonale entre un point et un hyperplan. Dans un repère orthonormé, pour un point P de coordonnées (p_1 ; p_2 ; … ; p_n) et l'hyperplan H d'équation $a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0$, on a la distance d (P, H) = $|a_1p_1 + a_2x_2 + \ldots + a_nx_n| / \sqrt{a^2_1 + \ldots + a^2_n}$. Un hyperplan est un objet mathématique de dimension *n – 1* dans un espace de dimension *n*. Dans la figure 18, la droite $E = 0$ représente un hyperplan de dimension 1 (une droite) dans un espace de dimension 2 (un carré). Bien que la formule de calcul de la distance entre un point et un hyperplan préconise de ne tenir compte que des valeurs absolues de leurs coordonnées, Paas et Van Merriënboer ne les utilisent volontairement pas : la distance seule ne permettrait pas de savoir si l'efficience est positive ou négative, d'où l'intérêt de tenir compte des valeurs relatives. Ce calcul de distance, Tuovinen et Paas (2004) l'ont étendu à un espace de

dimension 3. Il leur a ainsi été possible de combiner trois types de mesures différentes et d'évaluer, pour chaque point de ce volume, quelle distance le sépare de l'hyperplan représenté par l'ensemble des solutions pour que son équation égale 0.

Cette méthode est extrapolable à n'importe quel nombre de dimensions, quand bien même ne pourrait-on plus en faire la représentation graphique. Bien que Tuovinen et Paas traitent chaque variable entrant dans l'équation à égalité, d'autres auteurs font remarquer qu'il est envisageable de leur faire bénéficier d'une pondération en fonction de leur importance supposée ou connue (Huang, Hong, & Eades, 2006). La formule générale de calcul devient alors celle présentée à la formule 4.

$$E = (\omega_1 Z_{\text{Score}1} - \omega_2 Z_{\text{Score}2} - \ldots - \omega_i Z_{\text{Score}\,i}) \,/\, \sqrt{(\omega^2{}_1 + \omega^2{}_2 + \ldots + \omega^2{}_i)}$$

Formule 4: Mesure d'efficience *E* (d'après Huang, Hong, & Eades, 2006)

Ces modifications apportées à la formule de départ impliquent que le repère dans lequel s'inscrivent les dimensions n'est plus orthonormé. Bien que chaque axe reste perpendiculaire à chacun des autres, la graduation, elle, variera en fonction du coefficient qui lui a été attribué. Par conséquent, le dénominateur de la formule ne sera plus la racine du nombre de dimensions[14], mais la racine de la somme des coefficients élevés au carré. Ainsi utilisée, cette formule ne mesure plus nécessairement une efficience, mais peut mesurer toute autre chose, pour peu que l'on en donne une définition en fonction des termes composant l'équation. C'est le cas par exemple, d'une étude qui met en relation performance et motivation (Darabi, 2004). Une autre utilisation possible de ce type de formule est, en mettant en relation différentes mesures, de déterminer une éventuelle dissociation entre elles. C'est ce qu'ont fait Yeh et Wickens (1984) pour mettre en évidence une dissociation entre performance et charge de tâche subjective.

[14] Le dénominateur n'est pas à proprement parler la racine du nombre de dimensions, mais la somme du carré des coefficients dans le cas où ils sont chacun égaux à 1, ce qui revient au nombre de dimensions.

3.7. Facteurs d'augmentation de la charge inutile et remédiations

3.7.1. Effets de partage de l'attention et de contiguïté spatiale

L'intérêt de mesurer la charge mentale est de pouvoir la faire diminuer afin de favoriser l'apprentissage. Ainsi que nous l'avons souligné plus haut, il n'est pas possible de réduire la charge intrinsèque propre à la réalisation d'une tâche. Il est cependant possible d'en réduire la charge inutile, c'est-à-dire celle liée au format de présentation des informations relatives à la réalisation de la tâche. Plusieurs facteurs d'augmentation de la charge extrinsèque ont été repérés au cours du développement de la CLT. Parmi eux, nous nous intéresserons à l'effet de partage de l'attention (*split-attention effect*). Il intervient lorsque les individus sont mis en situation de partager leur attention entre plusieurs sources informationnelles pour ensuite devoir les intégrer mentalement (Ayres & Sweller, 2005; Chandler & Sweller, 1992; Cooper, 1990). Les ressources mentales affectées à l'intégration de sources éloignées vont donc faire défaut à d'autres processus plus essentiels, c'est-à-dire ceux relevant de la charge utile (Sweller, Van Merriënboer, & Paas, 1998). Le corollaire de cet effet est celui de contiguïté spatiale (*spatial contiguity principle*). Il est en effet possible de faire diminuer la charge extrinsèque d'une tâche en intégrant, ou pour le moins, en rapprochant spatialement, des sources informationnelles disparates mutuellement référencées (Mayer, 2001). Ceci n'est valable que si les informations séparées ne sont pas intelligibles isolément. Si tel est le cas, on aura alors tout intérêt à les intégrer (c'est-à-dire les unifier en une seule et même source, approche proposée par Sweller et ses collaborateurs) ou tout simplement les rapprocher (approche préconisée par Mayer et ses collaborateur).

Cependant, et en dépit de la solidité de ces deux principes de conception d'interfaces, ceux-ci sont résolument orientés vers l'apprentissage de documents scientifiques illustrés : formation des orages, fonctionnement des freins et de pompes (Mayer, 2001), fonctionnement d'un circuit électrique (Chandler & Sweller, 1991; Cooper, 1990) ou encore biologie (Erhel & Jamet, 2006; Jamet, 2000; 2006) pour ne citer qu'une petite partie du corpus de travaux consacrés à cette thématique. Bien qu'il soit possible d'aborder la charge mentale dans le cadre très large de la réalisation de tâches (Tricot & Chanquoy, 1996), c'est dans l'étude et la conception de documents pédagogiques que la CLT s'est illustrée. Dès lors, il serait sans doute hâtif de vouloir appliquer à une interface de saisie et de reconnaissance d'écriture manuscrite le principe de « contiguïté spatiale » en tant que tel. Il existe toutefois une autre

conception théorique très proche de celle dont nous venons de faire la présentation. Elle ambitionne de diminuer la charge induite par les traitements qu'impose le format de présentation des informations dispensées par une interface. Il s'agit du principe de proximité compatibilité (*Proximity Compatibility Principle*, PCP), développé dans le domaine des interfaces de contrôle aérien (au sol ou embarquées) (Wickens & Carswell, 1995). Même si ce champ d'étude peut apparaître d'emblée éloigné du notre, c'est plutôt dans la notion de contrôle qu'il faut chercher la parenté. Or, nous l'avons vu au chapitre 2, le contrôle est une des étapes essentielles de la production écrite (Kellogg, 1996; Roussey & Piolat, 2005).

*3.7.2. Le Principe de Proximité Compatibilité (*Proximity Compatibility Principle*, PCP, Wickens & Carswell, 1995)*

Lors de la réalisation d'une tâche en interaction avec une interface, le système propose à son utilisateur un certain nombre d'informations, issues de différentes sources. Le postulat de départ du PCP est que souvent, plusieurs de ces informations doivent impérativement être traitées simultanément par l'opérateur. Par conséquent, le PCP propose des recommandations sur la façon dont les sources informationnelles doivent être organisées entre elles afin d'en faciliter la collecte et le traitement. Le PCP repose sur deux dimensions de proximité ou de similarité : la proximité perceptive (*perceptual proximity*) et la proximité de traitement (*processing proximity*). S'agissant de notre problématique, le terme de proximité sera préféré à celui de similarité, car plus proche sémantiquement de l'idée de « contiguïté spatiale ».

- **Proximité perceptive :** c'est le degré de similarité, en termes d'attributs perceptifs et de localisation, entre des canaux informatifs véhiculant des informations relatives à la tâche. Elle peut aussi bien désigner une ressemblance de forme, couleur, orientation, *etc.* qu'un rapprochement de position.

- **Proximité de traitement :** c'est le degré auquel deux ou plusieurs informations vont servir conjointement à la réalisation d'une partie de la tâche. Si plusieurs informations doivent être mentalement intégrées, alors elles présentent une forte proximité de traitement.

Le PCP prescrit que des informations présentant une forte proximité de traitement devront également être proches perceptivement. Les promoteurs du PCP préconisent un certain

nombre de « manipulations de similarité » qui vont permettre d'augmenter la proximité perceptive de sources informationnelles. Il est possible de faire varier le degré de proximité le long d'un *continuum* qui s'étend d'un pôle (aucune proximité) à l'autre (forte proximité). Le niveau de proximité perceptive doit bien sûr correspondre au niveau de proximité de traitement entre les informations que la réalisation de la tâche induit. Parmi les solutions proposées par Wickens et Carswell, nous retiendrons la proximité spatiale (rapprochement des sources) et l'intégration spatiale (fusion des sources) des informations qu'il est nécessaire d'associer mentalement. On fait donc aisément le rapprochement avec les travaux de Mayer et de Sweller, appliqués cependant à des interfaces de contrôle plutôt qu'à des documents pédagogiques.

Certains aspects des préceptes du PCP sont appliqués de longue date par la communauté des concepteurs d'IHM, suivant des recommandations aussi anciennes que la *Gestalt* elle-même (Tractinsky & Hassenzahl, 2005). La contribution propre du PCP est de faire le lien entre ses recommandations et le fonctionnement cognitif, sans pour autant être une théorie psychologique en tant que telle (Wickens & Carswell, 1995). C'est ainsi que les auteurs mentionnent le coût d'accès aux informations (*Information Access Cost*, IAC) et la MdT. Le rapprochement des sources dispensant des informations à traiter conjointement facilite leur comparaison et leur intégration mentale en diminuant l'effort pour les localiser et passer de l'une à l'autre. Cet effort relève autant des mouvements oculaires et faciaux que du déplacement du faisceau attentionnel. L'ensemble des coûts imputables à ces mouvements et opérations mentales est cumulé et désigné sous le sigle IAC. L'une des applications les plus connues et spectaculaires visant à diminuer l'IAC est la conception pour l'aviation de systèmes d'affichages dits « tête haute » (*head-up display*) qui affichent les données des instruments de bord sur un miroir transparent placé dans l'axe de la verrière. Cela évite au pilote de quitter des yeux son environnement tout en restant informé sur l'état de son appareil ou de tout autre paramètre utile (Wickens, 1996; Wickens & Carswell, 1995). Même s'il désigne spécifiquement le coût de recherche des informations, on peut rapprocher l'IAC de la charge extrinsèque de Sweller et son équipe. En effet, en pesant sur les ressources limitées de la MdT, l'intégration mentale d'informations provenant de sources disparates peut altérer la performance par surcharge.

3.8. Critique de la notion de charge

Une étude récente s'est attachée à démontrer expérimentalement le bien-fondé de l'application du PCP à une interface dédiée, pour l'expérience, à la résolution d'un problème mathématique. Si effectivement, la performance des participants a bénéficié d'une conception de l'interface fondée sur le PCP, l'administration *post-test* du NASA-TLX n'a pas permis de révéler de différences significatives entre les conditions. Cette expérience, tout en validant le bénéfice que les concepteurs d'IHM ont eu à s'inspirer du PCP pour élaborer une interface de qualité, montre également des limites du concept de charge (Rothrock et al., 2006). Tricot et Chanquoy (1996) se sont en effet interrogés sur la pertinence du concept de charge mentale en psychologie. Ils constatent que l'idée de charge mentale est parfois évoquée *a posteriori* pour justifier de résultats observés et dont l'explication trouverait son origine dans la charge mentale. C'est à cet égard qu'ils évoquent Molière et son Malade Imaginaire pour aborder la charge mentale. Leur examen de la notion de charge à travers différentes approches en psychologie leur fait conclure que le concept de charge est plus utile dans une approche « comparatiste-évaluative » telle que s'en réclame Sweller que dans une perspective psychométrique. L'étude de Rothrock et ses collaborateurs (2006) le confirme précisément : la manipulation de la conception d'une interface induit des niveaux de charge mentale différents dont les répercussions sur les performances sont directement observables. A l'inverse, la mesure de la charge mentale par le biais du NASA-TLX n'a pas permis de déceler de différence entre les interfaces. Il faut donc nuancer l'écart qui existe entre notre capacité à mesurer la charge mentale (y compris avec des outils éprouvés) et la possibilité de la mettre en évidence expérimentalement (visée comparatiste-évaluative).

Faut-il pour autant renoncer à mesurer la charge mentale ? Bien évidemment, non. Selon nous, la conception développée par Paas et ses collaborateurs est tout-à-fait judicieuse. En s'attachant à évaluer la charge cognitive par le biais des mesures de charge mentale, d'effort mental, de performance et d'efficience, ils proposent un ensemble d'indicateurs qui, à défaut de mesurer explicitement de la charge cognitive, constituent un « faisceau d'indices » pouvant utilement renseigner sur les niveaux de charge subis par les individus en situation d'interaction avec un dispositif.

Cependant, se donner les moyens de mesurer (ou d'évaluer) la charge mentale ne doit pas faire tomber dans l'écueil consistant à vouloir la faire diminuer à tout prix. La charge mentale est fondamentalement un phénomène cognitif normal, et il n'est ni possible, ni souhaitable de

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

la diminuer à l'extrême. On en prend la mesure si, comparant la charge mentale à la température corporelle, pour éviter qu'elle monte, on cherche à la réduire au maximum. S'il importe que la charge mentale n'outrepasse pas les capacités de la MdT, il faut garder en mémoire que la charge peut être « utile » (Keinonen, 1998; Paas, Renkl, & Sweller, 2004; Sweller, Van Merriënboer, & Paas, 1998). Delignières et Deschamps (2000) vont jusqu'à faire le rapprochement entre effort mental et activation physiologique du système nerveux central pour laquelle il est connu depuis fort longtemps qu'il existe un niveau optimal (ni trop faible, ni trop élevé) grâce auquel les processus cognitifs atteignent la plus grande efficacité (Yerkes & Dodson, 1908).

4- Expérience 1 : Qualité ergonomique de l'interface DIGIME: Effets de la contiguïté spatiale du feedback et de la conservation à l'écran des lettres saisies lors d'une tâche de saisie de texte manuscrit sur PDA

Résumé – On teste deux facteurs susceptibles d'améliorer l'ergonomie de l'interface DIGIME. La conservation de l'encre tout d'abord, qui consiste à maintenir à l'écran, ou pas, l'encre électronique laissée par le stylet de l'utilisateur sur l'écran du PDA. La contiguïté du feedback ensuite, qui consiste en l'affichage contigu ou éloigné de la zone de saisie, des lettres reconnues par le logiciel DIGIME. Les résultats montrent que la conservation de l'encre est une entrave à la bonne exploitation du dispositif, aussi bien en vitesse de saisie qu'en nombre d'erreurs, mais uniquement en début de tâche. A l'inverse, si le feedback contigu n'a pas d'effet sur les temps de saisie, il augmente le nombre d'erreurs produites. Ces effets n'ont que peu de répercussion sur les jugements des participants.

4.1. Introduction

Les interfaces stylo connaissent une forte expansion grâce au développement de l'informatique mobile. En premier lieu, les besoins inhérents aux situations de mobilité ont favorisé l'utilisation de ce type de méthodes d'entrée. Les machines telles que *Tablet* PC, PDA ou encore les téléphones mobiles de nouvelle génération (Smartphones) ont délaissé le tandem clavier-souris pour une interaction avec la machine via un stylet et un écran tactile (MacKenzie & Soukoreff, 2002). Le second facteur de développement de ce type d'interfaces tire parti du caractère *a priori* très intuitif de l'interaction stylo. C'est un mode de communication très utilisé par l'humain pour de nombreuses applications (prises de notes, rédaction, croquis, *etc.*) qui trouvent leurs pendants en informatique. Il semble donc intéressant de conserver ce mode d'interaction pour proposer des ergonomies intuitives et adaptées aux utilisateurs.

Nos partenaires de l'IRISA travaillent au développement de systèmes de reconnaissance d'écriture et d'IHM orientés stylo. Ceux-ci sont destinés principalement aux périphériques mobiles de petite taille (Smartphones et PDA). Ces systèmes présentent deux contraintes matérielles importantes à considérer : de faibles ressources physiques (mémoire et vitesse de processeur) et des écrans de petite taille. Le système de reconnaissance de caractères manuscrits isolés RESIFCar (Anquetil & Bouchereau, 2002), conçu par l'IRISA, a déjà été embarqué sur des Smartphones commercialisés en Europe (voir figure 19). Les contraintes matérielles ont dicté le choix d'une reconnaissance de caractères isolés, plutôt que celui de mots cursifs. La reconnaissance de mots est trop coûteuse en ressources pour obtenir des temps de réponse acceptables du point de vue de l'utilisateur. De tels systèmes de reconnaissance sont évalués en dehors de tout contexte applicatif, sur des bases de caractères manuscrits collectées préalablement (Bouteruche, Deconde, Anquetil, & Jamet, 2005). Le retour d'expérience associé au système RESIFCar montre que ses performances théoriques en termes de taux de reconnaissance, bien que proches des seuils de tolérance admis par l'utilisateur, ne permettent pas à elles seules d'assurer l'acceptation du système. Ce constat est en adéquation avec les résultats rapportés par des travaux antérieurs (MacKenzie & Chang, 1999; Ortelli, 2003).

Figure 19: Le logiciel de reconnaissance RESIFCar implémenté sur Smartphone

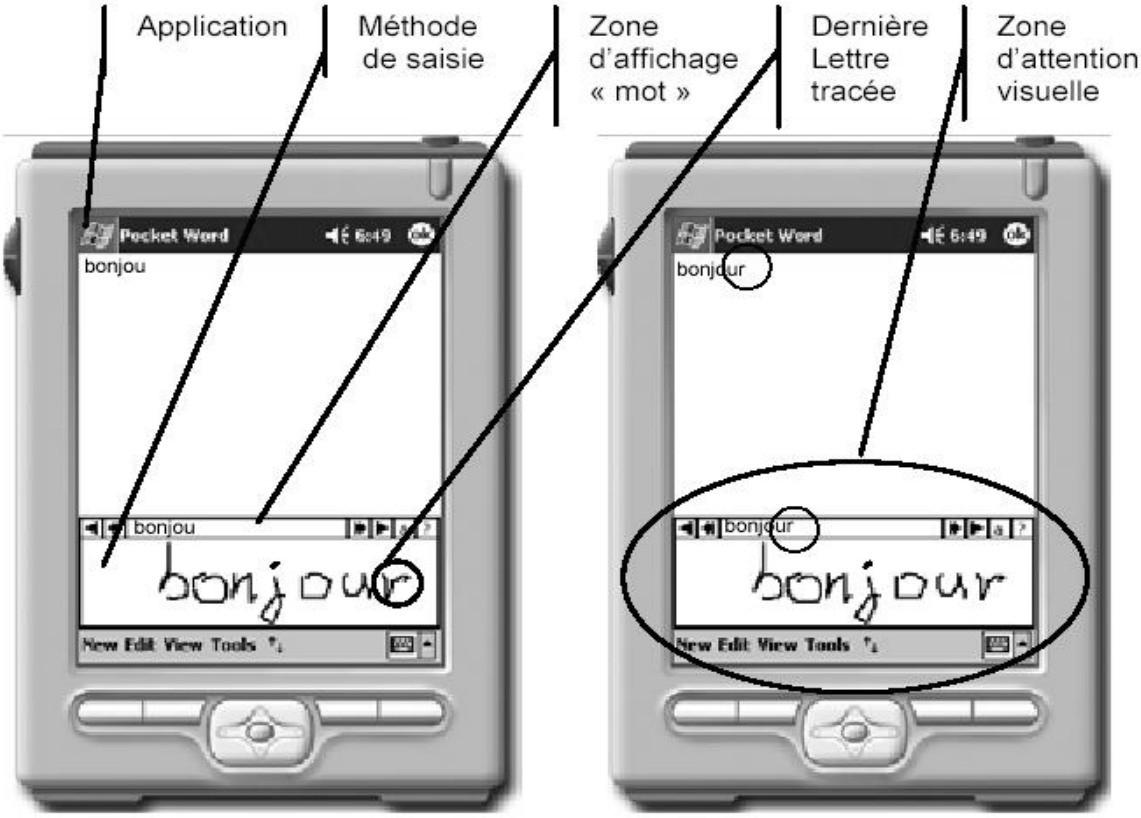


Figure 20: L'interface de saisie DIGIME implémentée sur pocket PC, sous l'application *Pocket Word*

Afin de pallier à ce déficit d'acceptabilité, il est possible d'améliorer l'ergonomie de l'interface de saisie (Baccino, Bellino, & Colombi, 2005; Bastien & Scapin, 2001; Kolski, 1997). On cherche donc à la rendre adaptée aux capacités cognitives des utilisateurs, pour

permettre une utilisation optimale du système de reconnaissance. On espère ainsi en favoriser l'acceptabilité (Arning & Ziefle, 2007; Hasan & Ahmed, 2007; ISO-9241-11, 1994; Keinonen, 1998; Nielsen, 1993; Shackel, 1991; Tricot & Tricot, 2000; Venkatesh & Davis, 1996). A partir de ce constat, nous proposons donc, en collaboration avec l'IRISA, de développer une interface de saisie de texte manuscrit sur périphériques mobiles de petite taille. Il s'agit de l'interface DIGIME. L'objectif poursuivi est d'améliorer ainsi l'acceptabilité du système de reconnaissance RESIFCar. Pour en faciliter la conception et l'évaluation, nous l'implémentons sur Pocket PC (voir figure 20). Malgré ce changement de support, l'interface conserve les propriétés et limitations qui étaient les siennes sur Smartphone, à savoir que la taille de la zone de saisie de texte reste inchangée, de même que le moteur de la reconnaissance RESIFCar.

Même s'il existe quelques travaux en ergonomie sur l'utilisabilité de ces systèmes d'identification de lettres écrites (Költringer & Grechenig, 2004; MacKenzie & Chang, 1999; MacKenzie, Nonnecke, McQueen, Riddersma, & Meltz, 1994; MacKenzie & Soukoreff, 2002; MacKenzie & Zhang, 1997; McQueen, MacKenzie, Nonnecke, Riddersma, & Meltz, 1994; Read, 2007), il n'en est pas moins vrai qu'à notre connaissance, ces recherches se cantonnent à des études descriptives, et n'apportent pas d'explications, notamment en termes de fonctionnement cognitif, sur les performances observées.

Pour développer DIGIME et apporter des améliorations significatives à l'interface de départ RESIFCar, nous choisissons de nous appuyer sur les *critères ergonomiques* de Bastien et Scapin (2001). Parmi l'ensemble des recommandations qu'ils formulent, nous retenons celle qui vise à faire diminuer la charge de travail. L'opportunité de ce choix s'inscrit dans le prolongement de la discussion de Tricot et Chanquoy (1996) sur la notion de charge mentale. Le concept de charge mentale présente l'intérêt de faire le « pont » entre les trois dimensions de notre problématique : ergonomie des systèmes, manipulation d'interface et production écrite. En fait, la notion de charge mentale n'a de sens qu'en regard des limitations capacitaires de la MdT (Baddeley, 1986; Cowan, 2001; Gaonac'h & Larigauderie, 2000; McCutchen, 1996; Miller, 1956; Tricot & Chanquoy, 1996). En tant qu'espace de traitement conscient à capacité limitée (Baddeley, 2000), elle est impliquée dans toute activité volontaire (Logie, Engelkamp, Dehn, & Rudkin, 2000). La manipulation d'une IHM s'inscrit dans ce cadre (Drommi, Ulferts, & Shoemaker, 2001). Par conséquent, un certain nombre de difficultés à en faire bon usage peut être imputable aux ressources limitées de la MdT, si tant

est que celles-ci ne sont pas suffisantes à la réalisation de la tâche par un utilisateur novice (Sweller, 1988; Tricot, 1998). En effet, l'utilisateur novice, contrairement à l'expert, n'a pas développé ni automatisé les schémas et procédures qui lui permettraient d'outrepasser les capacités limitées de sa MdT (Cooper, 1998; Ericsson, 2005; Tricot, 1998). Notre objectif est donc de trouver des solutions ergonomiques à même de minimiser l'allocation de ressources cognitives à la réalisation d'une tâche de saisie de texte sur l'interface DIGIME. Deux contributions théoriques peuvent nous y aider : la théorie de la charge cognitive (CLT, Sweller, 1988) associée au principe de compatibilité proximité (PCP, Wickens & Carswell; 1995) et le modèle de production écrite de Kellogg (1996).

S'agissant de la manipulation d'interfaces, nous avons vu qu'il était possible de l'assimiler à une forme de résolution de problème (Drommi, Ulferts, & Shoemaker, 2001). La théorie de la charge cognitive (CLT, Sweller, 1988) peut donc fournir un éclairage adapté à l'amélioration de l'interface DIGIME. Bien qu'il ne soit pas possible de réduire la charge mentale inhérente à la réalisation d'une tâche de saisie de texte sur PDA (charge intrinsèque, Chandler & Sweller, 1991; Clark, Nguyen, & Sweller, 2006), il est néanmoins possible de faire diminuer la charge induite par les conditions de sa réalisation (charge extrinsèque, Chandler & Sweller, 1991; Clark, Nguyen, & Sweller, 2006). La réduction de la charge inutile devrait, selon la CLT, procurer à l'utilisateur un gain de ressources cognitives qui pourront être utilement affectées à la construction en MLT de schémas relatifs à l'exploitation de l'interface de saisie (Cooper, 1998; Tricot, 1998). Ainsi, en favorisant les conditions d'apprentissage du fonctionnement de DIGIME, les performances des utilisateurs devraient s'améliorer. La CLT propose un certain nombre de recommandations pour la conception d'interfaces. Elles sont susceptibles de faire diminuer la charge induite par leur traitement (Sweller, Van Merriënboer, & Paas, 1998). Parmi elles, nous portons notre attention sur le principe de contiguïté spatiale (Chandler & Sweller, 1991; Cooper, 1990; Mayer, 2001). Lorsque deux sources informationnelles sont mutuellement référencées, il convient de les rapprocher physiquement afin d'éviter une surcharge cognitive due au partage de l'attention (Ayres & Sweller, 2005; Chandler & Sweller, 1992; Cooper, 1990). Or, écrire sur une interface de reconnaissance d'écriture manuscrite implique de la part de l'utilisateur qu'il contrôle que les lettres saisies sont correctement reconnues par la machine (en plus de l'activité de contrôle propre à toute situation de production écrite (Kellogg, 1996; Roussey & Piolat, 2005). Par conséquent, si la zone de saisie des lettres manuscrites est spatialement

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

éloignée de l'endroit où elles s'affichent une fois reconnues, cela peut entraîner une augmentation préjudiciable de la charge inutile. On peut envisager ce constat sous l'angle du principe de compatibilité proximité (PCP, Wickens et Carswell; 1995) et penser que s'il existe une forte proximité de traitement entre les lettres manuscrites produites par le scripteur et les lettres reconnues par la machine, alors cette proximité de traitement doit être accompagnée d'une proximité perceptive équivalente. Ceci doit normalement faire diminuer le coût d'accès à ces informations (IAC, *Information Access Cost*) interdépendantes.

En ce qui concerne la production écrite, nous ferons particulièrement cas de ce qui relève de l'exécution motrice de l'activité d'écriture. Si pour le modèle de production écrite de Kellogg (1996), l'exécution motrice de l'écriture est automatisée (pour le scripteur expert), et par conséquent, ne pèse pas sur la MdT, dans le cas de DIGIME, il en est sans doute autrement. Comme nous l'avons précisé plus haut, DIGIME est une interface de saisie et de reconnaissance de caractères isolés (écriture script). Ce type d'écriture est assez inhabituel pour la plupart des individus, même si nombre d'entre nous utilisons un style hybride, entre cursif et script (Graham, Weintraub, & Berninger, 2001). Par ailleurs, le support matériel de production étant lui aussi inhabituel (Dobbie & Askov, 1995; Serratrice & Habib, 1993), il y a fort à parier que pour les utilisateurs novices de DIGIME, l'exécution motrice de leur texte manuscrit ne sera pas automatique, mais contrôlée (Olive & Kellogg, 2002; Olive & Piolat, 2002). Or, l'étude d'Olive et Piolat (2002) a révélé que la suppression du feedback visuel en situation d'écriture était de nature à en augmenter le coût d'exécution. Par conséquent, faire en sorte que l'interface de saisie de texte permette à ses utilisateurs un contrôle rétroactif constituerait un facteur de diminution de la charge mentale (Serratrice & Habib, 1993). Dans ce contexte d'augmentation probable des ressources cognitives nécessaires à la réalisation du processus d'exécution, cet aspect peut s'avérer crucial pour l'utilisabilité du dispositif.

En faisant varier ces facteurs susceptibles de faire baisser la charge mentale engendrée par l'exploitation du dispositif, on cherche à en améliorer l'ergonomie, et, ainsi, l'utilisabilité. Parmi les outils dont nous disposons pour évaluer les répercussions de ces manipulations sur la charge, il est possible de faire usage du NASA-TLX (Hart & Staveland, 1988), dont l'utilisation dans le cadre de l'évaluation d'interfaces de reconnaissance de texte a déjà montré son utilité (Költringer & Grechenig, 2004; Segall, Doolen, & Porter, 2005). De même, l'intérêt que nous avons témoigné à la formalisation théorique proposée par Nielsen (1993) nous oriente aussi vers l'utilisation du *Nielsen's Attribute of Usability* (Test d'utilisabilité de

Nielsen) afin de guider les participants dans leur évaluation subjective et qualitative de l'interface. Nous nous appuierons aussi sur les recommandations de Nielsen (1993) pour dégager des critères propres à évaluer l'utilisabilité de l'interface DIGIME en fonction des conditions expérimentales. L'apprenabilité tout d'abord, qui évalue la facilité avec laquelle les utilisateurs se familiarisent avec le dispositif. Les erreurs commises, ensuite. L'efficacité également, par des mesures de performance. Pour finir, on pourra évaluer la satisfaction, notamment grâce au test subjectif d'utilisabilité de Nielsen.

4.2. Méthode

4.2.1. Participants

Quatre-vingt trois personnes ont participé à l'expérience. Il s'agissait, pour la plupart, d'étudiants en sciences humaines, d'un âge compris entre 18 et 25 ans. La moyenne d'âge des participants était de 20,9 ans. Quarante-huit de nos participants étaient des femmes, tandis que 35 étaient des hommes. La moitié d'entre eux provenaient de la deuxième année du cours de psychologie tandis que l'autre moitié a été recrutée sur le campus de Rennes 2. Aucun de nos participants n'était possesseur ou utilisateur de PDA, ce qui en faisait des utilisateurs novices dans le cadre de cette expérience. Les participants étaient tous volontaires et n'ont reçu aucune rémunération.

4.2.2. Matériel

4.2.2.1. Le logiciel de reconnaissance RESIFCar

L'un des aspects fondamentaux de la qualité d'interaction avec une interface de saisie d'écriture manuscrite est le moteur logiciel de la reconnaissance. Si son taux d'identification est trop faible, alors il s'avérera inutilisable, voir même, inutile. L'IRISA a développé le système de reconnaissance de caractères isolés RESIFCar. Cette application pouvait reconnaître des caractères isolés, ce qui incluait lettres, signes diacritiques, ponctuation, ainsi que des gestes d'édition (Anquetil & Bouchereau, 2002). Elle était le résultat de plusieurs années de recherches en la matière (Anquetil & Lorette, 1996; 1997a; 1997b). Cet identificateur traitait les formes écrites correspondant à la trajectoire du stylet sur l'écran entre

le moment où il y était posé et celui où il était relevé. La reconnaissance se faisait *on-line*, s'appuyant sur des informations propres à la cinématique du geste (vitesse, orientation et ordre de production des traits). Le logiciel a reçu un « apprentissage » à partir d'exemples de caractères produits par environ 200 personnes, de façon à ce que le dispositif soit confronté à de nombreuses et différentes écritures manuscrites. Si un minimum de 60 exemples est nécessaire pour construire la modélisation d'un caractère, RESIFCar a été conçu sur une base moyenne comprise entre 100 et 300 exemples pour chaque type de caractères, afin d'obtenir les meilleurs modèles possibles. La rigueur accordée à la modélisation informatique des caractères, associée à des algorithmes particuliers de décomposition des formes tracées, a assuré à RESIFCar robustesse et rapidité dans l'identification, et ce, indépendamment des limitations techniques de la machine (Bouteruche, Deconde, Anquetil, & Jamet, 2005).

4.2.2.2. L'interface DIGIME

DIGIME était l'IHM permettant d'exploiter le logiciel de reconnaissance de caractères RESIFCar. Si l'on se réfère au cadre conceptuel proposé par Ham et ses collaborateurs (2006), RESIFCar était l'interface utilisateur logique, DIGIME était l'interface utilisateur graphique, tandis que le Pocket PC était l'interface utilisateur physique. DIGIME a été développée pour PDA afin d'en faciliter l'évaluation, notamment le recueil des données, qui aurait été plus compliqué sur Smartphone. Nous présentons dans les lignes qui suivent les principes généraux de fonctionnement et d'utilisation de DIGIME. Il est à noter que certaines des caractéristiques dont nous faisons état dans ces lignes ont fait l'objet de modifications en fonction des modalités des VI que nous avons manipulées pour cette expérience. Nous nous cantonnons ici à présenter le fonctionnement de DIGIME de façon générale. Les changements apportés à l'interface dans le cadre du test de nos hypothèses sont précisés à la section consacrée à la présentation des VI.

4.2.2.2.1. Saisie de caractères manuscrits en contexte mot

L'application permettait d'écrire dans le contexte du mot. Le scripteur écrivait les mots en suivant la progression habituelle gauche→droite, chaque lettre étant détachée des autres. Bien que RESIFCar ne reconnaissait que l'écriture script, il pouvait néanmoins détecter différents

allographes d'une même lettre, même s'ils étaient ordinairement associés à l'écriture cursive, exception faite des lettres majuscules, non reconnues. L'encre électronique laissée par le geste de l'utilisateur restait affichée dans la zone de saisie, ce qui permettait de produire certains caractères en plusieurs traits, tels que les « i », « t », ou « x » (voir figure 21). La présence des lettres écrites assurait donc à l'utilisateur un feedback de sa progression dans la saisie : c'est le contexte mot. De la même façon, il était possible de transformer un caractère déjà saisi en réécrivant par-dessus. La présence des lettres précédemment tracées figurait parmi les facteurs que nous avons fait varier dans cette étude.

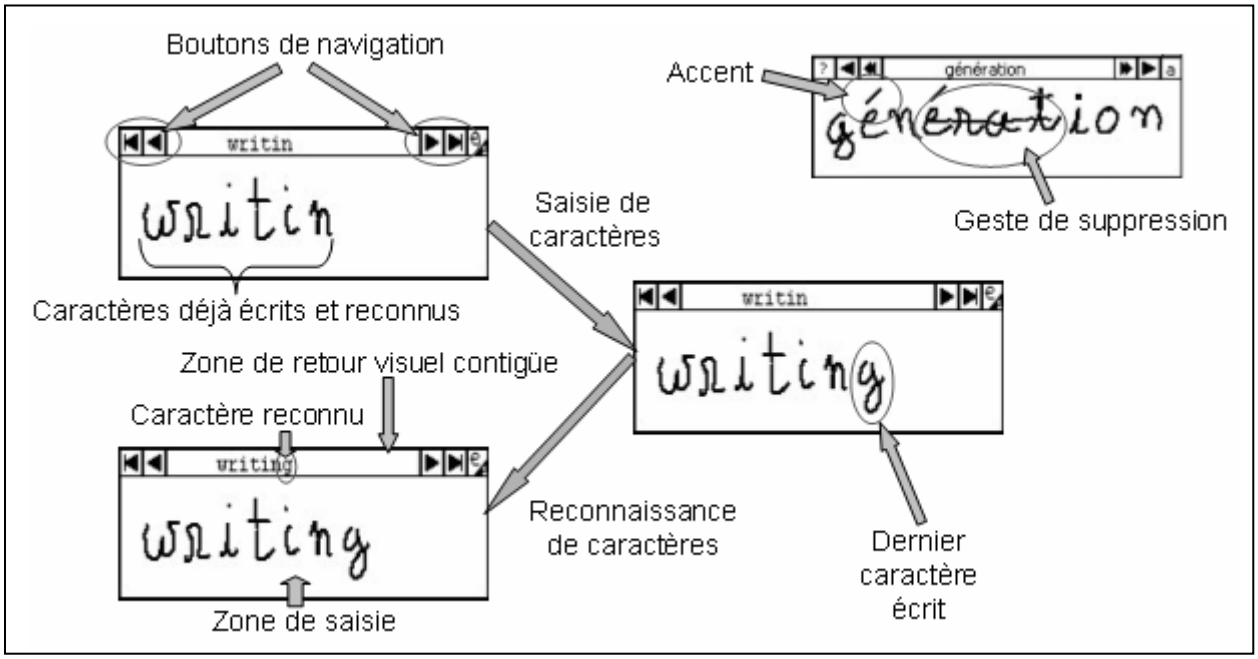


Figure 21 : Fonctionnement de l'interface DIGIME

4.2.2.2.2. Feedback visuel

Un aspect capital de l'utilisation d'une interface de reconnaissance de caractères est la possibilité de pouvoir contrôler si les lettres écrites sont correctement reconnues par la machine. Une zone de feedback contigüe a été intégrée au micro éditeur afin de limiter la distance entre zones de saisie et de contrôle (retour visuel, voir figure 20, « zone d'affichage mot »). La zone de retour visuel était donc située à proximité immédiate de la zone de saisie. Cela évitait à l'utilisateur d'avoir à déplacer son regard jusqu'à la zone d'application (traitement de texte *Pocket Word* en l'occurence). Chaque fois que l'utilisateur inscrivait un

nouveau caractère ou modifiait un caractère existant, les changements étaient immédiatement transmis à la zone de feedback. La présence contigüe du feedback visuel figurait également parmi les facteurs que nous avons fait varier dans cette étude.

4.2.2.2.3. Navigation dans le micro éditeur

Pour que l'utilisateur puisse disposer à sa convenance de l'espace de saisie, l'interface DIGIME proposait un système automatique de défilement des lettres saisies, afin de laisser visibles les dernières, sur la gauche de la zone de saisie. Ceci libérait de l'espace au centre, et n'entravait ainsi pas l'écriture des lettres suivantes. L'automaticité de cette procédure évitait à l'utilisateur de s'en charger lui-même par l'intermédiaire des flèches de défilement. Celles-ci lui permettaient néanmoins de naviguer dans le texte déjà produit. C'est un élément de contrôle explicite tel que préconisé par Bastien et Scapin (2001).

4.2.2.2.4. Signes diacritiques, ponctuation et gestes d'édition

Afin d'adhérer au principe de cohérence et d'homogénéité défendu par Bastien et Scapin (2001), l'interface DIGIME a tenté de minimiser au maximum l'utilisation de ressources type clavier virtuel, menus, et boutons. C'est pourquoi elle permettait de réaliser le plus naturellement, par tracés, accents, ponctuation, et gestes d'édition. Pour les signes diacritiques, le geste à réaliser était calqué sur celui fait lors d'une situation d'écriture manuscrite sur papier (y compris la localisation, au-dessus ou en-dessous des lettres). Pour les gestes d'édition, la suppression consistait à recouvrir les lettres à effacer d'un trait de la droite vers la gauche. Ce geste, s'il était réalisé sans recouvrir de caractère n'effaçait que le dernier saisi. Il pouvait également, suivant l'endroit de la zone de saisie où il était placé, n'effacer qu'un accent, une virgule ou une cédille.

4.2.2.3. Matériel informatique

Nous avons utilisé trois Pocket PC E400 Toshiba avec stylets (*Intel PXA 261* cadencés à 300 MHz, 31,22 Mo de mémoire vive, environnement *Windows CE*) ainsi que trois ordinateurs de bureau (*Intel Pentium III* cadencés à 1 GHz, 128 Mo de mémoire vive,

environnement *Windows XP*). Durant la passation de l'expérience, les PDA étaient reliés avec des ordinateurs distants pour le suivi en direct, par deux expérimentateurs, de l'activité des participants, ce afin de s'assurer du bon déroulement de la passation sans les gêner. La connexion entre les PDA et les PC était assurée par les logiciels *Microsoft ActiveSync* 3.7 et *Remote Display Control* 2.03.

4.2.2.4. Matériel linguistique

Une liste de quinze mots a été sélectionnée à partir de BRULEX, base de données lexicales regroupant de nombreux critères linguistiques pour chacun des 36000 mots qui la composent (Content, Mousty, & Radeau, 1990). Les mots de notre liste étaient répartis en trois groupes de cinq mots variant chacun selon leur nombre de lettres : quatre, sept ou dix lettres. Ces mots étaient choisis de manière à ce que ces groupes de mots soient équivalents en termes de classe grammaticale, de fréquence d'usage des formes, du nombre de variantes sémantiques, de valence d'imagerie, du nombre de variantes grammaticales, ainsi que de fréquence des bigrammes (voir en annexes, section 15.1)

4.2.2.5. Didacticiels

Un didacticiel a été créé et intégré à l'interface pour être présenté aux participants avant le début de l'épreuve. Il permettait de leur montrer la façon d'écrire quelques lettres, mais aussi comment procéder à quelques gestes d'édition. Le didacticiel se présentait sous quatre formes se voulant aussi similaires que possibles les unes aux autres, bien qu'adaptées à chacune des conditions. La durée de présentation du didacticiel était identique pour tous les participants car contrainte par le logiciel. Ce didacticiel était rudimentaire, et ne visait rien d'autre que la standardisation de la présentation du fonctionnement de l'interface. Il ne fournissait aux participants novices que les connaissances minimales nécessaires à l'accomplissement de la tâche. Il n'offrait pas la possibilité aux participants de s'exercer pendant sa présentation.

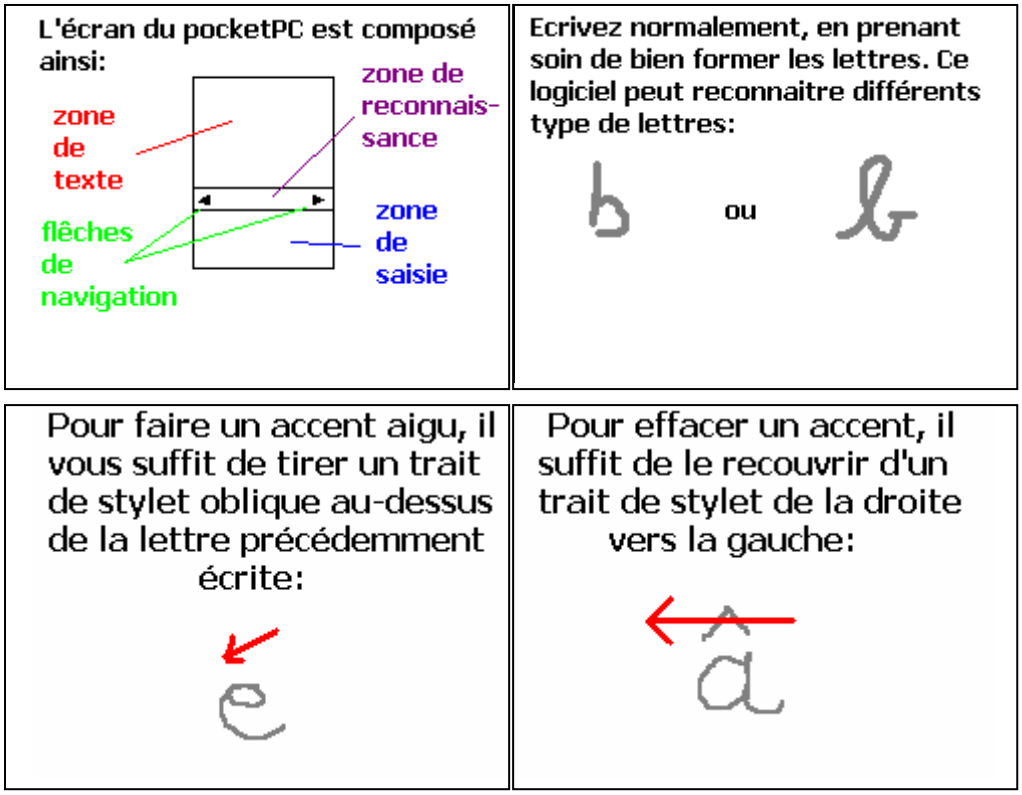


Figure 22: Exemples de planches proposées par les didacticiels

Même si DIGIME pouvait reconnaître plusieurs allographes pour chaque caractère, nous n'avons fait la présentation, dans le didacticiel que d'un allographe à chaque fois pour les quelques lettres qui ont servi d'exemple. Nous n'avons cependant pas manqué de signaler aux participants que le logiciel reconnaissait plusieurs formes de lettres. Les didacticiels sont consultables en annexes (section 15.4).

4.2.2.6. Questionnaires

En début d'expérience, un pré-questionnaire de cinq questions, portant sur leur éventuelle expérience en informatique et sur PDA, a été présenté aux participants. Les questions qui le composaient étaient du type : « Citez quelques marques de Pocket PC » ou « Donnez quelques exemples de systèmes d'exploitation employés sur les Pocket PC ». Nous n'avons écarté les participants que dans la mesure où ils faisaient une utilisation régulière du type de matériel que nous expérimentions, de sorte que leurs connaissances préalables ne biaisent pas les résultats (Farrington-Darby & Wilson, 2006; Morahan-Martin & Schumacher, 2007).

A la fin de la passation, les participants ont rempli le NASA-TLX. Ce test a été suivi du questionnaire d'utilisabilité de Nielsen (Nielsen, 1993). Les réponses à ce dernier questionnaire se faisaient sur une échelle de type Lykert à cinq niveaux, graduées de –2 (très négatif) à +2 (très positif). Les questions étaient les suivantes :

- Vous a-t-il semblé facile d'apprendre à utiliser le logiciel de reconnaissance ?
- Le logiciel de reconnaissance vous a-t-il semblé efficace ?
- Si vous deviez reproduire cet exercice dans une semaine seriez-vous capable d'atteindre le même niveau de performance ?
- Avez-vous eu du mal à gérer les erreurs, que ce soit celles du logiciel ou les vôtres ?
- Est-ce que le logiciel de reconnaissance vous a laissé une impression agréable ?

Nous avons complété ces questions avec trois items de notre cru, afin de mieux cerner les jugements des participants :

- Utiliseriez-vous ce système de reconnaissance ?
- Avez-vous changé d'écriture pour que le système reconnaisse votre écriture ?
- Avez-vous trouvé la situation expérimentale ennuyante ?

Tous ces questionnaires sont disponibles dans leur forme originale de présentation en annexes, à la section 15.3.

4.2.3. Procédure expérimentale

Dans un premier temps les participants répondaient au pré-questionnaire. Un didacticiel adapté à la condition expérimentale à laquelle ils étaient affectés leur était présenté sur le PDA. Ils étaient ensuite mis en situation de test dans l'une des quatre conditions expérimentales. Les participants avaient à reproduire successivement les quinze mots présentés aléatoirement sur l'écran du PDA, à l'aide du stylet, et à gérer les éventuelles erreurs de reconnaissance produites pour y parvenir. Les mots étaient présentés à l'écran tant que les participants n'avaient pas décidé de commencer la saisie, action débutant par un clic sur un bouton contenant le message « exercice» (voir figure 23). Le mot devait être écrit de façon non cursive, c'est-à-dire que les lettres devaient être séparées les unes des autres. L'écriture du mot demandé se terminait soit par la reconnaissance du mot par le logiciel, soit

par l'activation du mot « Suivant » au bout de trois minutes, sur lequel les participants devaient appuyer pour passer au mot suivant, en cas d'échec de la saisie correcte. Le temps qui séparait l'apparition d'un mot à écrire et sa saisie complète était enregistré par la machine, en millisecondes (sauf cas particulier d'une saisie excédant les trois minutes, auquel cas les participants étaient invités à passer au mot suivant). Les participants avaient la possibilité de corriger leurs erreurs, le procédé étant expliqué dans le didacticiel. La présentation des mots a été faite de façon aléatoire afin de contrebalancer les effets de rang. A la fin de l'expérience, nous avons soumis les participants au test NASA-TLX puis au test d'utilisabilité de Nielsen. La passation de l'expérience durait une trentaine de minutes.

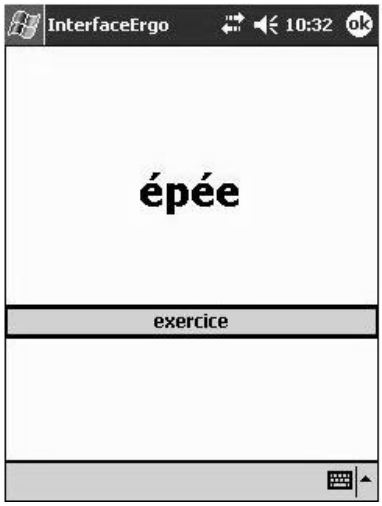


Figure 23: Présentation aléatoire d'un mot sur l'écran du PDA avant saisie par les participants

4.2.4. Variables indépendantes

Pour cette expérience, nous avons fait varier trois facteurs. Les deux premiers étaient intergroupes à deux modalités :

- **Contiguïté spatiale du feedback :** Il s'agissait de faire en sorte, suivant les modalités du facteur (présence ou absence du feedback contigu), que les lettres reconnues par la machine soient affichées aussi à côté de la zone de saisie ou bien seulement dans la zone traitement de texte située en haut de l'écran du PDA.
- **Conservation de la trace écrite :** Ce facteur a consisté, suivant ses modalités (disparition ou conservation de la trace écrite), à conserver ou non la trace écrite du

scripteur, c'est-à-dire l'encre électronique, laissée par le contact du stylet sur la zone de saisie.

Le croisement de ces deux VI a donné lieu à quatre déclinaisons de l'interface DIGIME que nous présentons à la figure 24, page suivante. Le troisième facteur était intragroupe, à trois modalités :

- **Position sérielle par groupe de cinq mots :** Afin d'observer un éventuel apprentissage se traduisant par une évolution de la performance au fil de l'épreuve, nous avons manipulé une troisième VI, à trois modalités. Elle a consisté dans le découpage des quinze mots à écrire en trois groupes de cinq mots (en fonction de leur ordre aléatoire de présentation, c'est-à-dire, un premier tiers de cinq mots, suivi du second tiers et du troisième). Ce découpage visait à nous permettre de traiter pour chaque « tiers » (début, milieu et fin d'épreuve), les niveaux de performances qui lui étaient associés.

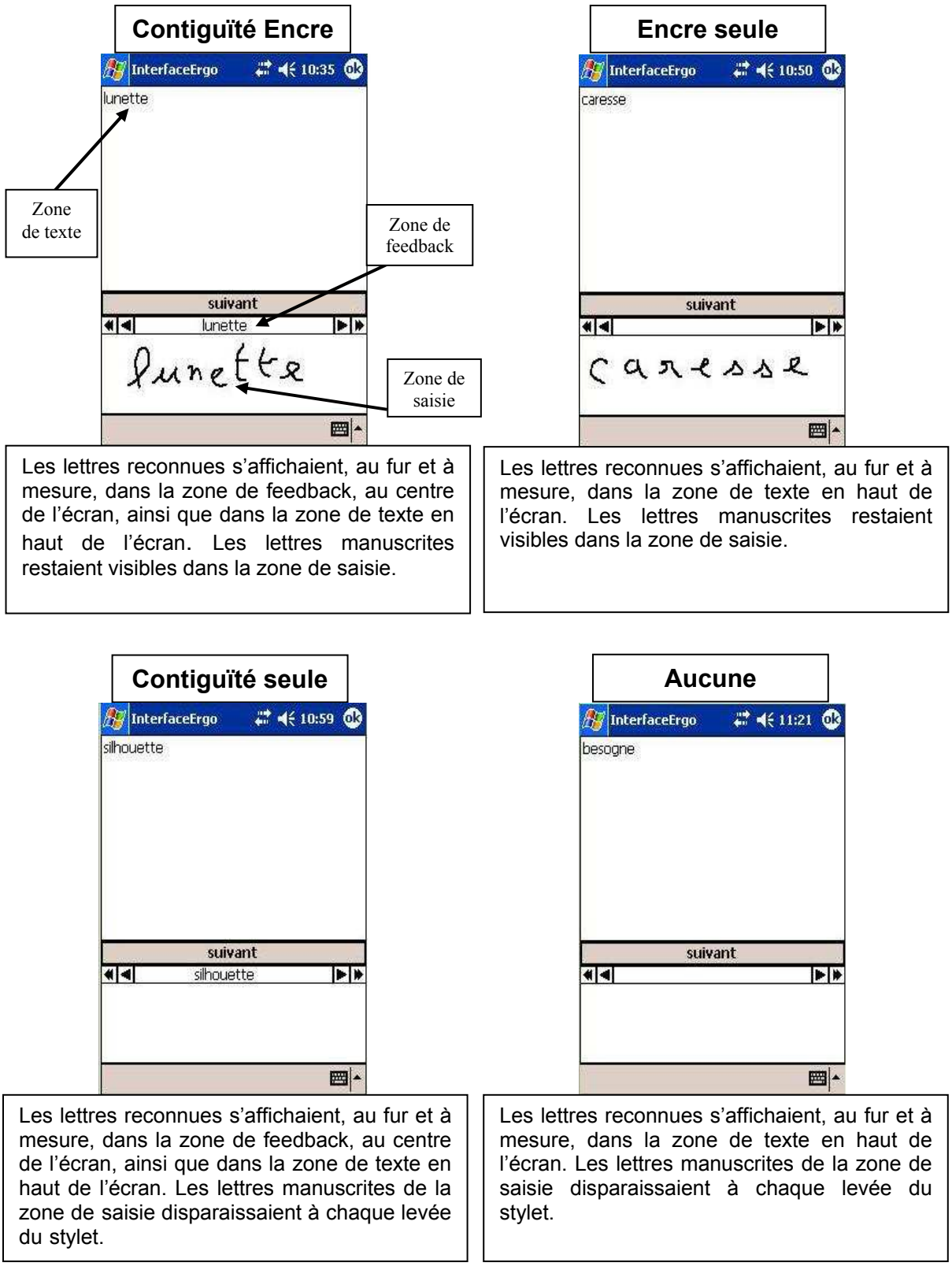


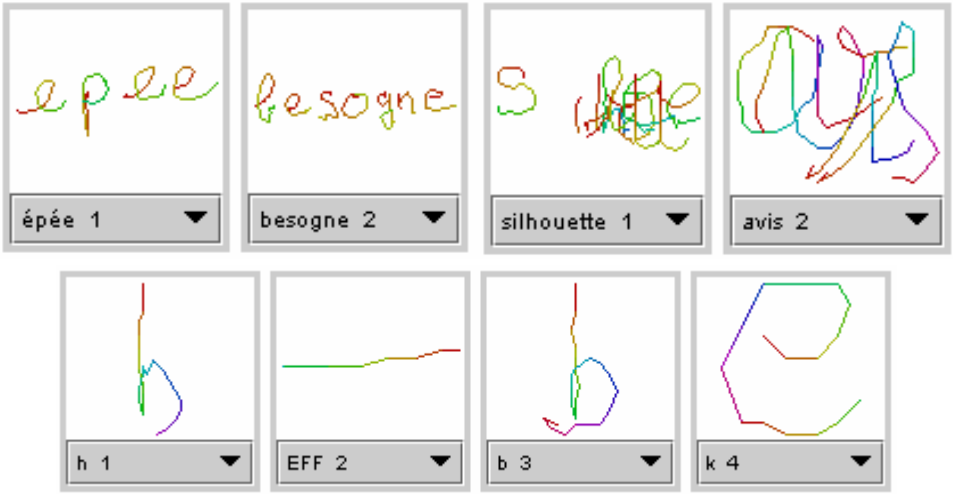
Figure 24: Présentation des quatre déclinaisons de l'interface DIGIME en fonction des modalités des VI intergroupes (contiguïté du feedback et conservation de l'encre)

4.2.5. Variables dépendantes

Nous avons mesuré les temps, en millisecondes, mis par les participants pour écrire chaque mot. Cela nous a permis, en prenant ces scores par groupe de cinq mots, d'établir les temps de saisie pour les cinq premiers, cinq suivants, et cinq derniers mots de l'épreuve, leur présentation aléatoire nous laissant supposer que chacun de ces groupes de mots était équivalent en nombre de lettres. Ces temps de saisie ont été enregistrés par le logiciel DIGIME, transformé pour l'occasion par nos partenaires de l'IRISA, et ce de façon totalement invisible pour les participants. Le chronomètre était lancé quand les participants activaient, par un clic sur le bouton, le mot à écrire, et arrêté par la saisie complète et correcte du mot. Ce délai, contraint par le programme, ne pouvait excéder trois minutes.

De la même façon, nous avons enregistré le nombre d'erreurs commises par les participants. Par erreurs, nous entendons tout geste ou lettre mal identifiée par la machine en fonction de ce qui était attendu dans la progression de l'écriture du mot. Ces erreurs pouvaient être autant le fait de la machine que de son utilisateur : il ne nous a pas été pas possible de le savoir objectivement. Notre analyse des erreurs s'est limitée à cela, considérant que cette mesure était représentative de la qualité de l'interaction avec l'interface. Nous n'avons pas tenu compte des erreurs de type fautes d'orthographe ou oublis, beaucoup plus difficiles à catégoriser et mesurer. Ce dernier type d'erreur était de toute façon, à terme, corrigé par l'utilisateur pour pouvoir accéder au mot suivant. L'accès à ces fautes s'est fait grâce à l'application Caractest, développé par l'IRISA, qui donnait accès *a posteriori*, aux traces écrites laissées par les scripteurs sur l'écran des pocket PC (voir figure 25, page suivante).

Pour finir, nous avons collecté les scores aux tests subjectifs de charge de tâche (NASA-TLX) ainsi qu'à chaque item du test d'utilisabilité de Nielsen. Nous avons complété ces questionnaires avec trois questions de notre invention, interrogeant les participants sur l'ennui ressenti pendant l'expérience, leur changement d'écriture, et leur intention d'utilisation de l'interface DIGIME à supposer qu'ils aient été possesseurs d'un PDA.



Note : En haut, traces écrites de mots, avec ou sans conservation de l'encre. Sans l'encre, les scripteurs écrivent toujours au même endroit, d'où la supperposition des lettres. En dessous, traces écrites de lettres, indépendamment des conditions de saisie. La couleur rouge indique le point de départ de la trace jusqu'au vert. Sous les traces, les mots ou lettres reconnus par la machine.

Figure 25: Captures d'écrans de l'application Caractest.

4.2.6. Hypothèses

- H1: La contiguïté spatiale du feedback aura un effet facilitateur sur la saisie des mots (en termes de vitesse, de nombre d'erreurs, de vitesse d'apprentissage et de score aux tests subjectifs) en évitant un partage attentionnel coûteux.

- H2: La conservation de l'encre à l'écran aura un effet facilitateur sur la saisie des mots (en termes de vitesse, de nombre d'erreurs, de vitesse d'apprentissage et de scores aux tests subjectifs) en soulageant la mémoire de travail d'un effort supplémentaire.

4.3. Résultats

Toutes les analyses statistiques de ce travail de recherche ont été conduites avec SPSS 11.0.

4.3.1. Effets des VI sur les performances

Nous avons procédé à une analyse de variance à mesures répétées, tenant compte à la fois des VI conservation de l'encre et contiguïté du feedback, ainsi que de la position des mots pris par groupe de cinq, au cours de l'épreuve. Nos données ne souscrivant pas à la condition de sphéricité, nous nous sommes appuyé sur les résultats de l'analyse multivariée de la variance qui elle, n'y est pas conditionnée (Howell, 1997, p.557). Puisque la variance des erreurs de nos VD était homogène, nous sommes passé outre le fait que toutes leurs distributions n'étaient pas gaussiennes (Howell, 1997, p.361). Nous avons conduit successivement deux ANOVA (modèle linéaire général, plan à mesures répétées) sur les deux types de VD temps de saisie et nombre d'erreurs puisque nous ne faisions pas, *a priori*, d'hypothèses sur le type de mesure de performance.

Afin de faciliter la lecture, nous ferons références aux moyennes par la lettre « *M* » et aux écart-types par le sigle « *SD* » (*Standard Deviation*, Ecart-type), aussi bien pour les temps moyens de saisie et le nombre moyen d'erreurs que pour les scores au NASA-TLX. Pour la même raison, bien que les temps de saisie aient été enregistrés au millième de seconde, nous n'en fournirons que des approximations tronquées à la seconde inférieure, niveau de précision suffisant pour rendre compte des effets observés.

VI inter	*Feedback*		*Encre*		
VD	Contigu	Eloigné	Persistante	Caduque	Total
Tps début (s)	315 (159)	333 (173)	376 (184)	277 (131)	323 (165)
Tps milieu (s)	173 (92)	159 (71)	183 (88)	152 (76)	167 (83)
Tps fin (s)	121 (67)	109 (56)	123 (71)	108 (53)	115 (62)
Tps total (s)	611 (254)	601 (262)	684 (286)	538 (206)	606 (256)
Nbr erreurs début	50,1 (31,1)	38,3 (22,2)	54,1 (31,7)	36,5 (21,1)	44,7 (27,9)
Nbr erreurs milieu	31,2 (24,2)	23,2 (13,1)	31,3 (23,5)	24,2 (16,5)	27,6 (20,2)
Nbr erreurs fin	21,6 (17,9)	14,3 (7,7)	19,7 (13,8)	17 (15,3)	18,3 (14,6)
Nbr erreurs total	103 (55,9)	75,9 (32,7)	105,2 (53,1)	77,9 (40,4)	90,7 (48,4)
NASA-TLX	9,03 (2,26)	9,42 (1,99)	9,60 (1,84)	8,85 (2,34)	9,2 (2,1)

Note : Moyenne (écart-type). Le score au NASA-TLX est compris entre 0, faible charge et 20, forte charge

Tableau 1: Principaux résultats de l'expérience, en fonction des VI manipulées

4.3.1.1. Effet des VI sur les temps de saisie

Les résultats ont montré une amélioration des temps de saisie au fil de l'épreuve (F(2, 77) = 117,001, *p* = .00). Cela témoigne d'un apprentissage par la manipulation de l'interface se traduisant par une plus grande rapidité au fil du temps. Du début d'épreuve (*M* = 323s, *SD* = 165), les temps de saisies se sont ensuite amenuisés (*M* = 167s, *SD* = 83), pour finir à *M* = 115s, *SD* = 62 en fin d'épreuve.

Ce résultat est à reconsidérer à la lumière de l'interaction entre le facteur conservation de l'encre et la position (F(2,77) = 3,952, *p* = .02). Les participants qui ont bénéficié de la disparition de l'encre ont montré de meilleurs temps de saisie en début d'épreuve (*M* = 277s, *SD* = 131) que ceux qui n'en bénéficiaient pas (*M* = 376s, *SD* = 184). Cependant, passée la saisie des cinq premiers mots, l'écart entre les deux conditions s'est resserré en milieu d'épreuve (*M* = 152s, *SD* = 76 pour la condition sans encre, contre *M* = 183s ; *SD* = 88 pour la condition encre conservée), pour finir dans un mouchoir de poche (*M* = 108s ; *SD* = 53 sans l'encre, contre *M* = 123s, *SD* = 71 avec l'encre).

De même, si la contiguïté du feedback n'a pas interagit avec le facteur position (F<1), l'interaction triple entre la position, la contiguïté et la conservation s'est avérée elle, être significative (F(2,77) = 3,904, *p* = .02). L'examen des mesures réalisées montre que la VI feedback a permis, quelque soit sa modalité, des temps de saisie équivalents pour chaque tiers d'épreuve. Or, ceux-ci n'ont différé de ceux dus au facteur encre que sur le premier groupe de mots (respectivement *M* = 333s, *SD* = 173 pour le feedback éloigné et *M* = 315s, *SD* = 159 pour le feedback contigu, contre *M* = 277s, *SD* = 131 sans l'encre et *M* = 376s, *SD* = 184 avec). A partir du second tiers, la différence s'est effacée (*M* = 159s, *SD* = 71 pour le feedback éloigné et *M* = 173s, *SD* = 173 pour le feedback contigu), cette tendance se confirmant en fin d'épreuve (*M* = 109s, *SD* = 56 pour le feedback éloigné et *M* = 121s, *SD* = 67 pour le feedback contigu, contre *M* = 108s ; *SD* = 53 sans l'encre, contre *M* = 123s, *SD* = 71 avec l'encre).

Ces résultats sont résumés page suivante à la figure 26.

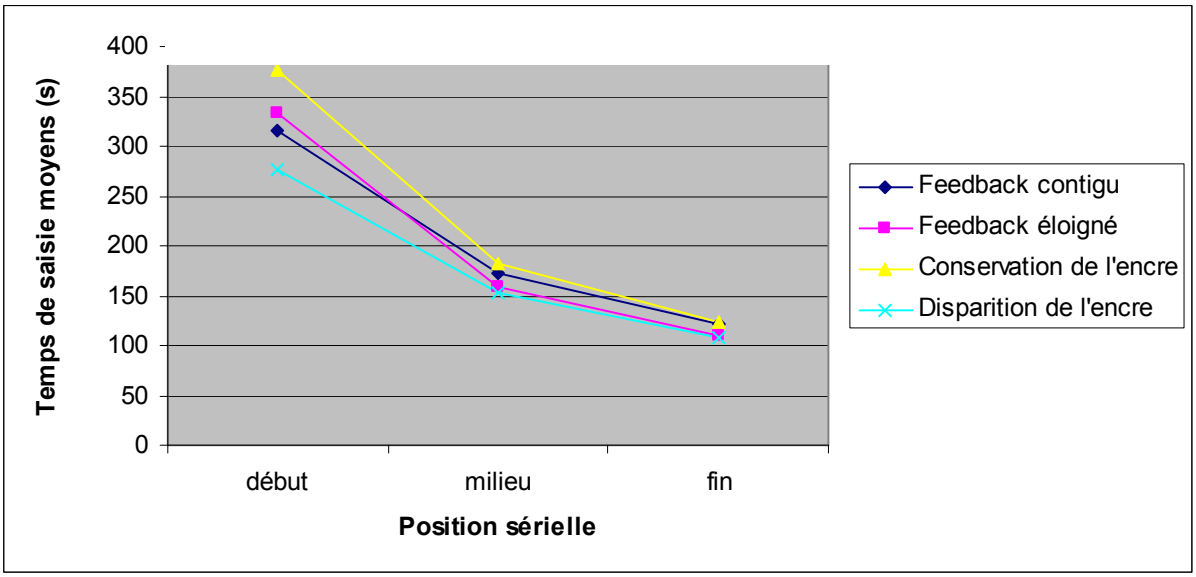


Figure 26: Effets des VI contiguïté du feedback et conservation de l'encre en fonction de la position sérielle sur les temps de saisie de groupes de cinq mots

Si l'on examine les effets de la contiguïté du feedback et de la conservation de l'encre sur les temps de saisie totaux (sur l'ensemble de l'épreuve), seul le facteur encre a montré un effet significatif ($F(1,78) = 5,943$, $p = .01$), ce dont témoigne la figure 27.

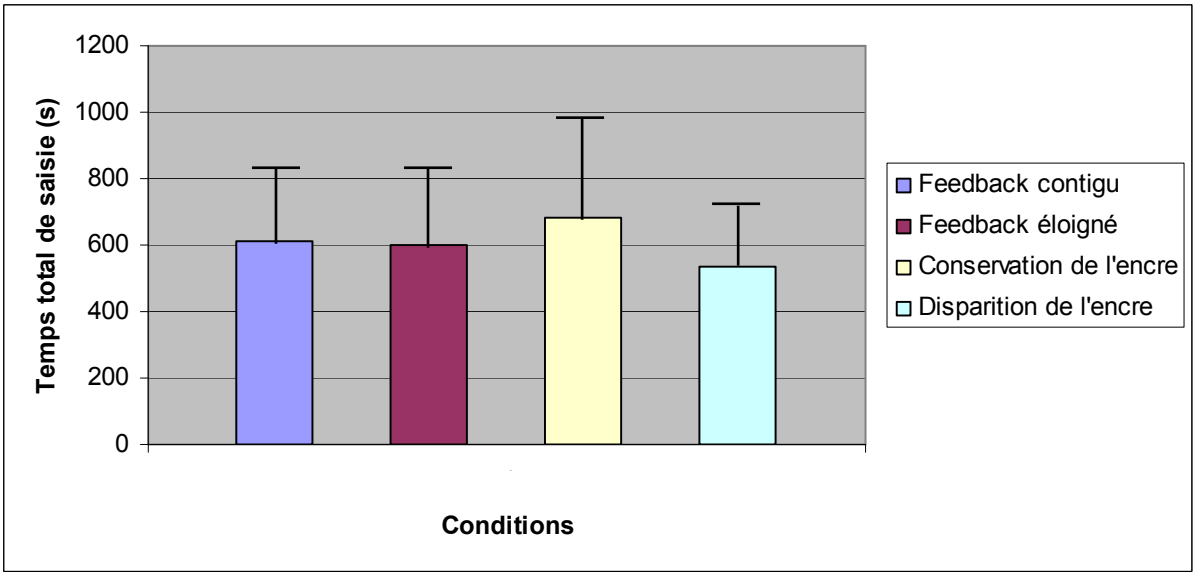


Figure 27: Effets des VI contiguïté du feedback et conservation de l'encre sur les temps de saisie des quinze mots

4.3.1.2. Nombre d'erreurs produites

Un certain nombre de données issues des fichiers d'analyse des erreurs n'était pas exploitable. Cela explique que le nombre de degrés de liberté soit différent entre l'analyse des temps de saisie et celle des erreurs, ci-après.

En ce qui concerne les erreurs, la position sérielle a exercé aussi un effet significatif ($F(2,72) = 47,438$, $p = .00$) qui est allé dans le sens d'une diminution du nombre d'erreurs au fil de l'épreuve. D'une moyenne de 44,7 erreurs en début d'épreuve (*SD* = 27,9), les scripteurs ont amélioré leur performance au fur et à mesure (*M* = 27,6, *SD* = 20,2 dès le second tiers de mots et *M* = 18,3, *SD* = 14,6 pour le dernier tiers).

Comme pour les temps de saisie, le nombre d'erreurs commises a subi les effets conjugués de la position sérielle et de la conservation de l'encre ($F(2,72) = 3,092$, $p = .05$). Si l'écart entre le nombre d'erreurs des conditions conservation (*M* = 54,1, *SD* = 31,7) et disparition (*M* = 36,5, *SD* = 21,1) de l'encre était assez important au début, il s'est amenuisé petit à petit (*M* = 24,2, *SD* = 16,5 pour la disparition, contre *M* = 31,39, *SD* = 23,5 pour la conservation de l'encre) pour disparaître en fin de test (respectivement *M* = 17,7, *SD* = 15,3 et *M* = 19,7, *SD* = 13,8 pour la disparition et la conservation de l'encre).

A l'inverse, ni l'interaction entre le facteur contiguïté du feedback et la position ($F<1$), ni l'influence conjointe des trois VI prises ensemble n'ont été significatives ($F(2,72) = 2,271$; $p = .11$). Pour ce qui est de l'absence d'effet entre contiguïté du feedback et déroulement de l'épreuve, un examen rapide de la figure 28 (page suivante) fait prendre la mesure du quasi parallélisme des courbes représentant le nombre d'erreurs associés à chaque modalité de la VI contiguïté. Il en va de même en ce qui concerne l'absence d'effet d'interaction entre les trois VI.

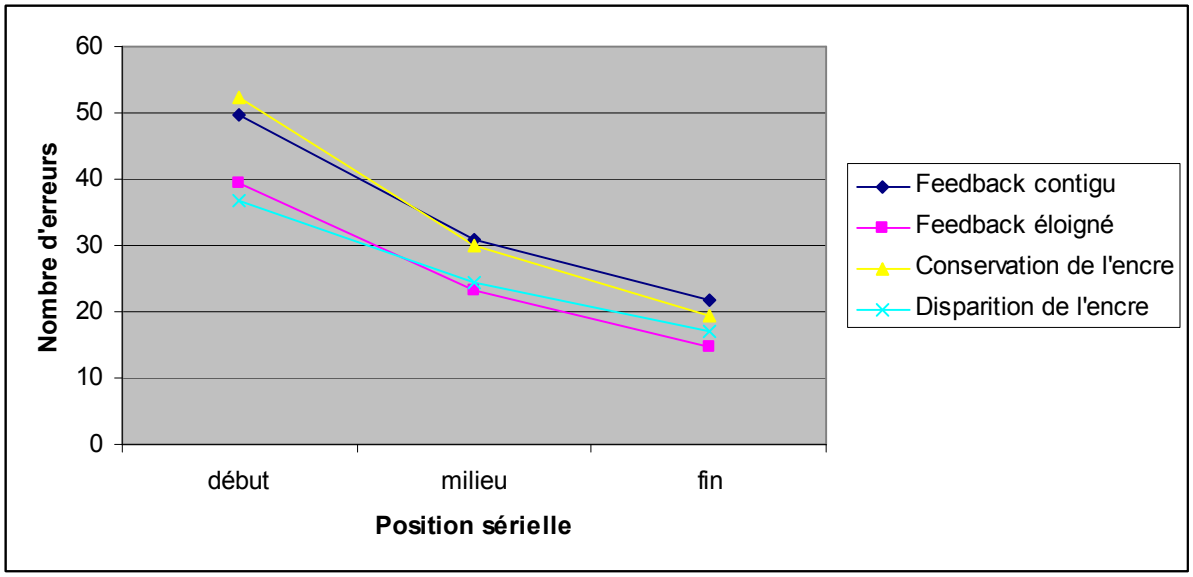


Figure 28: Effets des VI contiguïté du feedback et conservation de l'encre en fonction de la position sérielle sur le nombre d'erreurs commises sur chaque groupe de cinq mots

Pour finir, si l'on examine les effets de la contiguïté du feedback et de la conservation de l'encre sur le nombre total d'erreurs (sur l'ensemble de l'épreuve), on constate qu'à la fois la VI conservation de l'encre ($F(1,73) = 4,908$, $p = .03$) et la VI contiguïté du feedback ($F(1,73) = 5,481$, $p = .02$) ont témoigné d'un effet sensible sur le nombre d'erreurs enregistrées (voir figure 29, page suivante). C'est pourquoi il n'y a pas eu d'interaction entre les deux facteurs inter, contrairement à ce que nous avons mesuré pour les temps de saisie ($F(2,72) = 2,271$, $p = .11$).

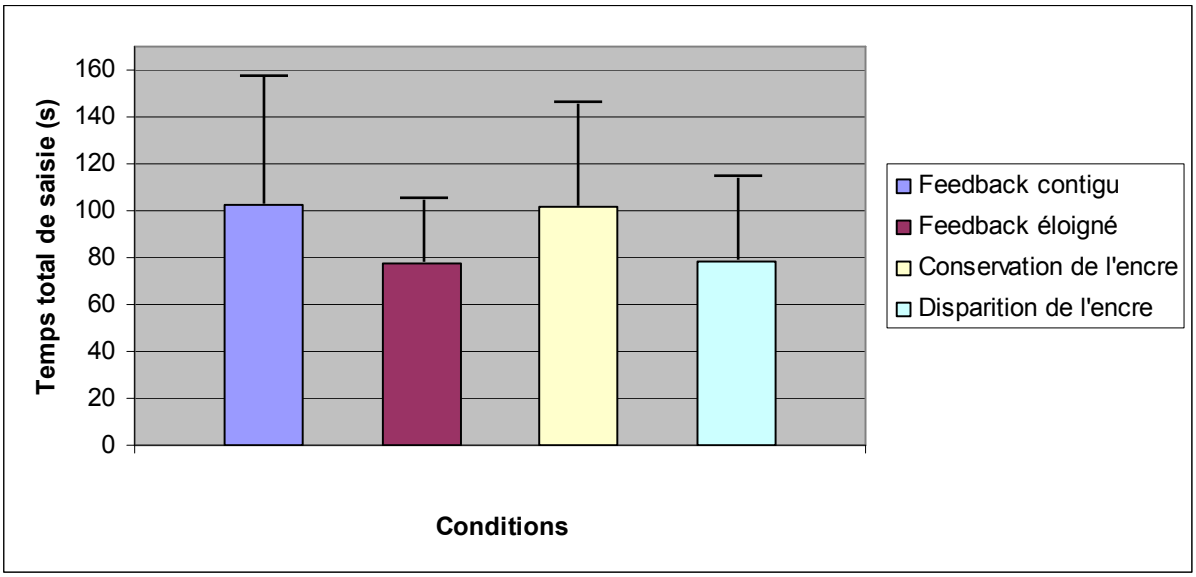


Figure 29: Effets des VI contiguïté du feedback et conservation de l'encre sur le nombre moyen d'erreurs commises pour la saisie des quinze mots

4.3.2. Effets des VI sur les tests subjectifs

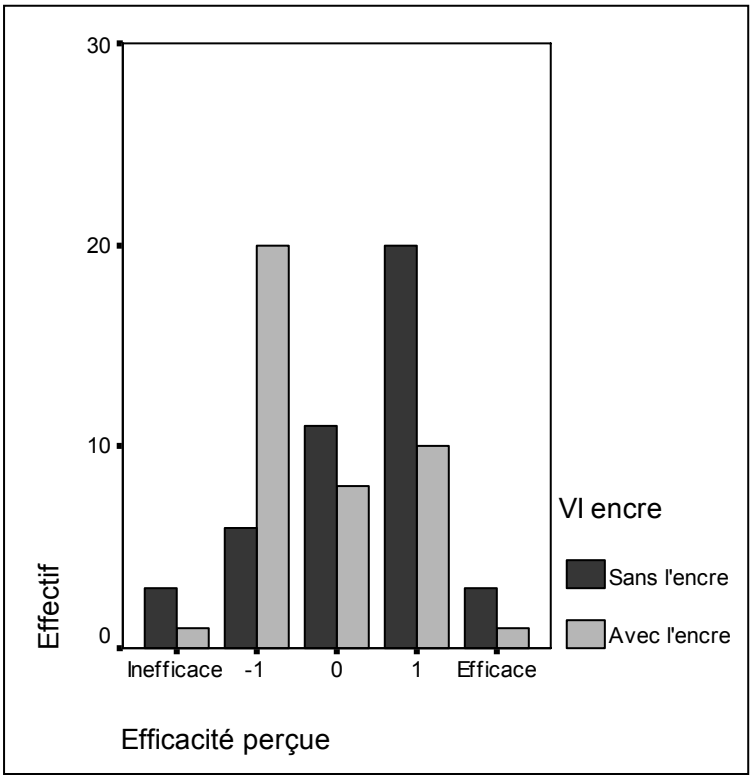
4.3.2.1. Test de charge de tâche NASA-TLX

Aucune des VI inter manipulées n'a eu d'effet sur la charge de tâche subjective mesurée par le NASA-TLX (ANOVA univariée, conditions d'applications respectées aussi bien pour la distribution que pour l'homogénéité des variances). Qu'il s'agisse du facteur conservation de l'encre ($F(1,78) = 3,001$, $p = .08$), du facteur contiguïté du feedback ($F(1,78) = 1,045$, $p = .31$) ou de leur interaction ($F<1$), les niveaux de charge rapportés par les participants sont assez similaires, et ce, quelles que soient les modalités des facteurs testés. Si le feedback contigu porte la charge de tâche subjective moyenne à 9,03 (*SD* = 2,26) contre 9,42 (*SD* = 1,99) lorsqu'il est éloigné, la conservation de l'encre montre une charge ressentie de 9,60 (*SD* = 1,84) contre 8,85 (*SD* = 2,34) lorsqu'elle disparaît.

4.3.2.2. Test d'utilisabilité de Nielsen

La conservation de l'encre a eu un effet significatif sur les jugements portés par les participants sur l'interface. L'efficacité perçue a été notamment jugée de façon plus positive

lorsque l'encre n'était pas conservée ($\chi^2(4,N = 83) = 13{,}254$, $p = .01$). La figure 30, ci-dessous, le montre.



Note : « Le logiciel de reconnaissance vous a-t-il semblé efficace ? »

Figure 30: Distribution des réponses à la question en fonction du facteur conservation de l'encre.

Aucun autre des items, qu'il s'agisse de ceux du test d'utilisabilité de Nielsen ou de nos questions supplémentaires ne s'est révélé sensible aux variables que nous manipulions.

4.4. Discussion

Pour résumer les résultats obtenus, nous retiendrons qu'effectivement, il y a bien eu un apprentissage de l'utilisation de l'interface DIGIME par la pratique. La diminution des temps de saisie ainsi que du nombre d'erreurs commises au fil de l'épreuve en atteste. Derrière ce premier constat, quelques nuances sont à considérer. Tout d'abord, les différences de performances dues à la manipulation des facteurs ne s'expriment qu'en début d'épreuve, notamment pour le facteur encre (temps de saisie et erreurs produites), qui interagit aussi avec le facteur contiguïté (ce dernier ne présentant pas d'effet simple sur les temps de saisie). Par

contre, la proximité du feedback a eu un effet néfaste sur les erreurs commises. Ces performances ne se sont que peu répercutées sur les jugements fournis par les participants, exception faite du sentiment d'efficacité qui a été affecté par la conservation de l'encre.

Par conséquent, force est de constater que le choix de nos hypothèses n'a pas été très heureux : les résultats mis en évidence par notre étude les invalident tout simplement.

Contrairement à nos attentes, la conservation de l'encre du scripteur a été une entrave aux performances, aussi bien pour la vitesse de saisie que pour les erreurs produites. Plusieurs explications permettent de comprendre cette différence avec les résultats obtenus par Olive et Piolat en 2002. Elles sont dues, selon nous, aux conditions de production écrite, qui varient largement entre les deux études. Bien que nous n'ayons pas à douter que la suppression du feedback visuel en écriture soit une source d'augmentation du coût de son exécution motrice, les conditions de réalisation de la tâche avec le logiciel DIGIME diffèrent suffisamment pour que d'autres entraves, imposées par la conservation de l'encre, soient de nature à en supplanter les effets. Les différences entre les types de tâches des deux études peuvent justifier ces résultats. L'explication de la gêne occasionnée par l'encre se situe dans le fait que l'écriture script n'est, d'une manière générale, pas automatique mais contrôlée (Olive & Kellogg, 2002; Olive & Piolat, 2002), même si rares sont ceux ayant une écriture purement cursive ou script (Graham, Weintraub, & Berninger, 2001). Par conséquent, l'exécution motrice de l'écriture étant chez le scripteur adulte automatique, la présence à l'écran de lettres a pu inciter les participants, par habitude, à lier entre elles plusieurs lettres. Cela n'est bien évidemment pas compatible avec les exigences du moteur d'identification RESIFCar. Cet attachement involontaire des lettres entre elles a pu se faire aussi à cause de la petite taille de l'écran. En dépit du défilement automatique vers la gauche des lettres déjà saisies, un problème de place a pu se poser pour écrire les nouveaux caractères alors que restaient affichés les précédents. Par ailleurs, l'utilisation par les scripteurs des flèches de défilement a peut-être été aussi un facteur d'augmentation des temps de saisie. De toute évidence, cela n'a pas pu nuire à ceux bénéficiant de la disparition de l'encre (puisqu'ils pouvaient écrire au centre de l'écran sans se préoccuper de ces aspects topocinétiques (Serratrice & Habib, 1993; Zesiger, 1995)). Il ne nous est pas possible de le vérifier. L'attachement de lettres entre elles ne fait en tout cas aucun doute, pour nous qui avons eu accès aux traces écrites des participants, bien que nous ne puissions pas, sur la seule base de ces informations, inférer objectivement qu'il s'agisse de cela. C'est pour cela qu'aucune analyse statistique n'a été

conduite sur une telle mesure en vue de le démontrer. C'est en tout cas une piste raisonnable pour expliquer l'augmentation du nombre d'erreurs produites par la conservation de l'encre sur le PDA, cet effet néfaste disparaissant au fur et à mesure de la pratique avec l'interface. L'entrave due à la conservation de l'encre n'est pas restée étrangère aux participants qui ont témoigné d'un moindre sentiment d'efficacité au test d'utilisabilité de Nielsen dans cette condition. A l'inverse, le NASA-TLX, n'a pas été sensible à cela. S'il on peut envisager que la difficulté ajoutée par la conservation de l'encre n'a pas été compensée des participants par un effort cognitif plus important (expliquant de fait, les moins bonnes performances), on peut y voir aussi un résultat explicable par la dissociation entre les mesures subjectives et les performances (Frokjaer, Hertzum, & Hornbaek, 2000; Yeh & Wickens, 1984). Néanmoins, l'approche comparatiste que nous adoptons, suivant les recommandations de Tricot et Chanquoy (1996), ne nous donne pas à penser que la charge mentale a effectivement été plus élevée pour la condition conservation de l'encre, puisque les performances d'apprentissage associées y ont été aussi bonnes que sans l'encre : bien que la prise en main de l'interface se soit soldée, dans cette condition, par des temps de saisie plus longs et des erreurs plus nombreuses, les performances étaient équivalentes en fin de test pour les deux conditions, avec et sans l'encre. Si l'obtention de ce niveau a été plus laborieuse en présence de l'encre, cela peut s'expliquer par le fait que la compréhension de l'interface était sans doute moins « intuitive » que sans l'encre (du fait de ces erreurs d'attachement témoignant de la rupture d'un automatisme bien ancré). Cette différence s'est néanmoins estompée relativement vite, pour arriver à des niveaux de performances comparables à ceux obtenus sans l'encre. Par conséquent, la conservation de l'encre n'entrave pas l'apprentissage, ce qui démontre qu'il reste aux participants suffisamment de ressources cognitives pour supporter l'apprentissage de l'interface. Malgré cela, le « handicap » des cinq premiers mots reste suffisamment fort pour que l'effet néfaste de l'encre s'exerce sur la totalité de l'épreuve et influence significativement les jugements des participants (efficacité perçue), qui porte lui aussi, sur l'épreuve entière. Ce décalage entre le score au NASA-TLX et performance n'est donc pas un résultat surprenant (Rothrock et al., 2006).

De la même façon, et contrairement à nos attentes, la contiguïté du feedback s'est avérée être une entrave à l'exploitation de l'interface DIGIME. Nos lectures nous laissaient pourtant penser que, les lettres reconnues par la machine partageant une forte proximité de traitement avec celles écrites par les participants, il convenait de leur faire partager aussi une forte

proximité perceptive, passant par un rapprochement physique (Wickens et Carswell; 1995). Ceci, s'apparentant au principe de contiguïté spatiale (Chandler & Sweller, 1991; Cooper, 1990; Mayer, 2001) de la CLT (Sweller, 1988), était supposé faire diminuer le coût d'accès aux informations, et donc, faire diminuer la charge mentale inutile (Chandler & Sweller, 1991; Clark, Nguyen, & Sweller, 2006). Ce gain de charge mentale devait, selon nous, libérer un certain nombre de ressources cognitives pour favoriser l'apprentissage du dispositif. Les résultats montrent qu'il n'en a rien été, bien au contraire. S'agissant des temps de saisie, la VI contiguïté a été tout simplement sans effet, puisque, comme nous le montre la figure 26, les courbes correspondant à chacune des deux modalités de ce facteur sont superposables de bout en bout, expliquant également l'absence d'effet sur l'ensemble de l'épreuve. Mais, plus gênant est l'effet néfaste d'un feedback contigu sur le nombre d'erreurs produites. Ceci est d'autant plus gênant que cet effet n'a pas diminué pas au fil de l'épreuve. Cela implique que, à la fois, la contiguïté du feedback n'a pas permis de libérer plus de ressources attentionnelles propres à améliorer l'apprentissage de DIGIME, mais qu'en outre, elle constitue un terreau favorable à la production d'erreurs sans que les participants aient pu y remédier véritablement (bien qu'il y ait une amélioration au fil du temps, mais pas par rapport à la condition sans feedback contigu). Le fait que la VI contiguïté n'ait eu aucun effet sur les temps de saisie peut donner lieu à deux interprétations : la première consiste à penser que, du fait de la petite taille de l'écran du PDA, le feedback contigu n'apportait rien de plus que le feedback éloigné, déjà suffisamment proche. La seconde consiste à penser que le feedback contigu était encore trop éloigné. Si l'on choisit cette seconde explication, on peut penser que la proximité de traitement entre les lettres écrites par les scripteurs et les lettres reconnues par la machine était si forte qu'il aurait fallu encore en augmenter la proximité perceptive jusqu'à une intégration complète (Chandler & Sweller, 1991; Cooper, 1990; Erhel & Jamet, 2006; Jamet, 2002; Jamet & Erhel, 2006; Mayer, 2001; Wickens, 1996; Wickens & Carswell, 1995). Cela n'explique cependant pas que la contiguïté du feedback ait été à l'origine d'un nombre plus important d'erreurs par rapport au feedback éloigné, et que la différence entre les deux conditions ne se soit pas réduite au fil de la tâche. Contrairement à la VI encre, la contiguïté du feedback ne modifiait pas la zone de saisie, elle n'était *a priori* pas de nature à augmenter le nombre d'erreurs. Néanmoins, si les erreurs ont été plus nombreuses avec le feedback proche, elles n'ont pas ralenti les participants dont les temps de saisie ont été équivalents à ceux qui n'avaient pour seul feedback que la zone application en haut de l'écran. D'autre part, ces erreurs induites par le feedback n'ont pas été mises en cause par les participants dans leurs

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

jugements. Sur la question de la charge induite par l'un ou l'autre des feedbacks (proche ou éloigné), rien ne nous autorise à penser à une différence. Les courbes d'apprentissage étant strictement parallèles, rien ne laisse penser que des ressources cognitives ont été libérées pour favoriser l'apprentissage. Il nous est donc très difficile de conclure sur ce résultat. Hormis le fait de l'invalidation évidente de notre hypothèse sur la contiguïté, nous pourrions trouver une explication dans la caractérisation des erreurs. En effet, nous avons choisi comme erreurs tout caractère, geste ou lettre, produit par les scripteurs mais ne donnant pas le résultat escompté au moment où il a été fait au cours de l'élaboration du mot. Cependant, les erreurs sont plus complexes que cela, sans qu'il ne nous soit possible, à partir de l'analyse de leurs traces écrites, de les catégoriser de façon plus fine. Bien que des tentatives de clarification sur la catégorisation des erreurs aient été proposées (Soukoreff & MacKenzie, 2004), elles étaient inapplicables à notre analyse. Une autre explication serait de penser que le feedback contigu, en étant redondant avec le feedback éloigné dans la zone traitement de texte, créait une surcharge attentionnelle. Elle devait d'ailleurs être d'autant plus importante que la conservation de l'encre constituait elle aussi un feedback supplémentaire qui s'ajoutait à ceux déjà évoqués. L'interaction triple entre l'encre, le feedback et la position plaide en faveur de ce type d'explication.

Pour conclure, si les résultats sont inattendus, ils n'en sont pas moins riches d'enseignements et nous ouvrent certaines pistes de recherche. A défaut de valider des règles de conception qu'il faut appliquer, cette manipulation aura plutôt eu le mérite de mettre en évidence ce qu'il ne faut pas faire. La conservation de l'encre tout d'abord, dont les effets néfastes, s'ils s'estompent rapidement, marquent les participants de façon négative et sensible. La contiguïté du feedback ensuite, dont l'influence sur le nombre d'erreurs est catastrophique, sans pour autant peser sur les jugements. C'est notamment le cas pour notre question concernant l'intention d'utilisation, qui constitue l'objectif de départ de notre démarche. Cette première manipulation s'intéressait en premier lieu à l'ergonomie de l'interface. C'est pourquoi nous n'avons pas traité la question de l'acceptabilité avec toute l'importance qu'elle mérite, eût égard aux outils méthodologiques existants (Ajzen, 1985; Davis, Bagozzi, & Warshaw, 1989; Fishbein & Ajzen, 1975; Venkatesh & Davis, 2000; Venkatesh, Morris, Davis, & Davis, 2003). Sur la question de la charge, et de sa source possible dans l'ergonomie de l'interface, nous aboutissons à une impasse : rien ne nous permet de conclure que les modifications apportées à l'interface, même dans les conditions pour lesquelles les

performances ont été les meilleures, sont à l'origine d'une diminution de la charge induite par sa manipulation. On retiendra donc de cette étude que la conservation de l'encre entrave la découverte de DIGIME et que le feedback n'était pas suffisamment intégré à la zone de saisie pour que les utilisateurs aient pu en bénéficier pleinement. En conséquence, deux axes de travail se dégagent : le premier relève des compétences de nos partenaires de l'IRISA, afin de faire en sorte que le feedback visuel soit totalement intégré à la zone de saisie de DIGIME. Le second, lui, nous revient. Les effets observés dus à la conservation de l'encre ne se sont exercé qu'en tout début d'épreuve. Or si ces effets sont de courte durée, leurs répercussions, à la fois sur la performance totale et sur les jugements, incitent à ne pas les négliger. Il convient donc de déterminer par quels moyens il serait possible de faciliter la prise en main de DIGIME. Ce problème d'apprenabilité est connu pour être crucial pour l'utilisabilité des interfaces (Nielsen, 1993; Shackel, 1991). C'est pourquoi nous choisissons d'orienter notre travail sur cette problématique.

5- Expérience 2 : Création d'un didacticiel pour DIGIME : Effets des illustrations statiques ou dynamiques sur l'apprentissage du fonctionnement de l'interface et sur son utilisabilité perçue.

Résumé – On teste l'effet du type d'illustration d'un didacticiel sur une nouvelle version de DIGIME, tenant compte des apports de la première expérience. On lui adjoint notamment un système d'aide en ligne. Trois conditions sont mises à l'épreuve : un didacticiel animé, un didacticiel statique et aucun didacticiel (condition contrôle). Si les résultats montrent que le type d'illustrations associé aux didacticiels n'a pas d'effet sur les performances, il ressort que les aides en ligne apportent aux utilisateurs une assistance non négligeable. La présence d'un didacticiel, à défaut d'améliorer les performances permet toutefois d'augmenter l'utilisabilité perçue du dispositif. Par rapport à la première étude, la prise en main de l'interface est globalement améliorée.

5.1. Introduction

L'expérience précédente a montré que, indépendamment des facteurs ergonomiques que nous avons manipulés pour tenter d'améliorer l'utilisabilité de l'interface DIGIME, les différences en termes d'apprentissage s'exercent en tout début d'épreuve, c'est-à-dire sur les cinq à dix premiers mots. C'est relativement court, pour autant que l'on puisse en juger, car la durée de la manipulation ne nous permettait pas de savoir comment aurait évolué la performance sur une période plus longue. Néanmoins, cette différence d'apprenabilité, due à la manipulation de la VI encre, si courte était-elle, s'est avérée suffisamment forte pour peser sur les jugements des participants et sur leurs performances à l'ensemble de l'épreuve. Or, le modèle de Nielsen (1993) met une certaine emphase sur l'apprenabilité des systèmes, c'est-à-dire leur capacité à permettre une découverte rapide et simple des principes de fonctionnement qui les régissent. En tant que dimension de l'utilisabilité d'une interface, améliorer son apprenabilité peut permettre d'en augmenter l'acceptabilité. Ce point de vue est couramment défendu, y compris dans d'autres modélisations de l'utilisabilité (ISO-9241-11, 1994; Keinonen, 1998; Shackel, 1991). Les modèles cognitivo-comportementaux les plus récents, tels que la Théorie Unifiée de l'Acceptation et de l'Utilisation des Technologies (UTAUT), en attestent par le caractère prédictif des conditions facilitantes sur l'acceptation des systèmes.

La première version de DIGIME (*modulo* l'influence des facteurs que nous avons manipulés) n'était accompagnée que d'un didacticiel minimal, dont l'unique fonction était d'apporter aux participants les connaissances suffisantes à la réalisation de la tâche. Il est cependant possible de proposer un didacticiel dont la conception reposerait sur des recommandations issues de la recherche sur les EIAH. Il s'agit pour nous de déterminer comment offrir aux utilisateurs de DIGIME le support à même de leur permettre de développer le plus rapidement possible les compétences indispensables à son utilisation. C'est le *guidage*, selon Bastien et Scapin (2001). La difficulté réside dans le fait de décrire aux utilisateurs des connaissances de natures procédurales et graphomotrices. En effet, la manipulation de DIGIME implique à la fois que l'utilisateur intègre un ensemble de règles d'usage (écriture script minuscule, connaissances des fonctions inhérentes à chaque zones ou boutons, *etc.*), mais également qu'il acquiert des connaissances motrices relatives à la production d'allographes qui seront reconnus par le logiciel. Ce partage de connaissances entre les concepteurs du dispositif et l'utilisateur final ne peut se faire que par la médiation

d'informations, qui, si elles véhiculent des connaissances procédurales, seront traitées en tant que connaissances déclaratives. Anderson (1982) a largement documenté cette question, et il ne fait aucun doute que l'acquisition de savoir-faire est possible grâce à la compilation de connaissances (Anderson, 1987; Lee & Anderson, 1997; Taatgen & Lee, 2003). Cela est aussi bien valable pour des procédures, que pour des actions à forte composante motrice, comme le souligne Thomas (1997). De fait, les aptitudes motrices sont issues de connaissances procédurales que l'on applique à son propre corps (Féry, 2001; Mounoud, 1986). Néanmoins, le soutien apporté par ces informations n'est pas suffisant à l'apprentissage, qui, s'il peut se faire sans médiation de la MdT (Maxwell, Masters, & Eves, 2003), doit nécessairement être accompagné par la pratique (Chevalier, Hall, & Nadeau, 1990), notamment en début d'apprentissage (Thomas, 1997).

S'agissant du meilleur format pour transmettre ces connaissances procédurales, il semble que les images soient le mieux adaptées, comme en témoignent un certain nombre de résultats expérimentaux. Ceci vaut autant pour les procédures au sens large, que pour les compétences motrices, ou plus précisément, graphomotrices. En effet, en l'absence de démonstrateur pour apprendre à l'utilisateur novice à utiliser une interface, la production d'une image simulant le comportement à accomplir constitue un bon palliatif. A cela, plusieurs raisons : en premier lieu, le fait que les procédures soient difficiles à verbaliser oblige à les faire partager de façon analogique (Denis & De Vega, 1993; Féry, 2001; Glenberg & Langston, 1992; Paivio, 1969). Si l'on ne dispose pas de modèle humain, la solution la plus simple est l'utilisation d'images (Féry, 2001). Ensuite, il existe de fortes compatibilités de traitement entre l'observation d'un geste et sa réalisation puisque la seule perception d'un mouvement provoque l'activation d'aires corticales associées à sa réalisation (Blake & Shiffar, 2007). Cela va même plus loin puisque la présentation du résultat d'un geste (tel qu'une lettre) active aussi les aires motrices qui sont affectées normalement à sa production (Longcamp, Anton, Roth, & Velay, 2003; Longcamp et al., 2008; Longcamp, Boucard, Gilhodes, & Velay, 2006). On comprend donc mieux comment une image va pouvoir participer à l'élaboration de ce que Jeannerod (1995) nomme les images motrices, qui procèdent autant du stockage du comportement moteur que de sa paramétrisation (Jeannerod & Frak, 1999). Ces images motrices sont, selon nous, à rapprocher du niveau de traitement morphocinétique de la production de lettres (Bonin, 2005; Paillard, 1988; Serratrice & Habib, 1993; Zesiger, 1995). A la suite de la procédure d'adressage (Zesiger, 1995), la séquence à écrire est stockée dans le tampon graphémique

(Caramazza & Miceli, 1990; Kandel, Alvarez, & Vallée, 2006; Kandel, Valdois, & Orliaguet, 2003; Zesiger, 1995) avant de subir un traitement postgraphémique. C'est à ce niveau que se jouent les aspects morphocinétiques. S'ils sont automatisés chez le scripteur expert, la première manipulation nous a montré que les scripteurs utilisant DIGIME n'étaient effectivement plus en situation d'expertise. Par conséquent, ces traitements sont pour eux contrôlés, et, compte tenu de leur nature spatiale, visuelle et motrice, vraisemblablement sous la dépendance du CVS (Baddeley, 2002a), et donc, de l'AC (Darling, Della Sala, & Logie, 2007; Rudkin, Pearson, & Logie, 2007; Vecchi & Richardson, 2000). De façon beaucoup plus marginale, les aspects topocinétiques sont aussi à l'œuvre pour placer les barres des « t », les points sur les « i » et « j » et ainsi que les signes diacritiques (Zesiger, 1995). Tout ceci va concourir à la détermination de la forme de la lettre, et donc, d'une certaine forme de sélection d'allographe à un niveau moteur (Bonin, 2003).

Au-delà de ces questions de programmation périphérique, l'assimilation de procédures peut renvoyer à l'application de règles plus « abstraites » que ne le sont les mouvements. L'imagerie mentale joue là-encore un rôle prépondérant (Denis & De Vega, 1993; Glenberg & Langston, 1992). De par sa capacité à produire des simulations mentales qui permettent à tout un chacun, pour peu qu'ils en aient les capacités, de se représenter le déroulement d'une séquence, l'imagerie mentale est au cœur de l'apprentissage de procédures (Hegarty, & Sims, 1994; 1997). Ces animations mentales peuvent parfois être difficiles à produire, si les individus ont de mauvaises aptitudes visuospatiales (Gyselinck, 1996) ou si l'effort relatif à sa construction en MdT dépasse les ressources de cette dernière. L'utilisation d'images peut alors se montrer un support adapté à ce type d'apprentissage.

Si le rôle bénéfique des images n'est pas à démontrer dans le cadre d'un apprentissage procédural ou moteur, ne serait-ce que du fait des dispositions de chacun à apprendre par imitation (Blake & Shiffar, 2007; Thomas, 1997), il reste à savoir quel type d'illustrations utiliser. Elles peuvent être statiques (images) ou dynamiques (animations). Bien que la littérature sur le sujet soit abondante, la question n'est pas encore tranchée quant à la supériorité des unes ou des autres (Bétrancourt, Bauer-Morrison, & Tversky, 2001; Bétrancourt & Tversky, 2000). Le consensus actuel irait plutôt, non pas vers le choix définitif de l'un ou l'autre type comme étant supérieur, mais plutôt vers la caractérisation des conditions qui vont révéler leur potentiel. A ce titre, les animations semblent toutes indiquées lorsque l'on cherche à faire la démonstration d'un phénomène dynamique ou décrivant une

procédure (Hidrio & Jamet, 2002; 2008; Hidrio, Jamet, & Le Bohec, 2002; Park & Hopkins, 1993). Notre problématique se situe bien dans ce cadre, mais ce constat ne s'appuie par sur des travaux relevant à la fois d'un apprentissage procédural et graphomoteur. Néanmoins, dans l'étude de Zacks et Tversky (2003) sur l'apprentissage du montage d'un saxophone (tâche procédurale à forte composante motrice), l'utilisation d'illustrations statiques et dynamiques n'a pas montré de différence significative entre les deux formats de présentation des informations imagées. Au demeurant, quand la supériorité des animations a été « avérée », un examen plus minutieux a souvent révélé que les illustrations dynamiques véhiculaient plus d'informations que les simples images (Bétrancourt, Bauer-Morrison, & Tversky, 2001). On peut y remédier par l'utilisation de flèches qui vont permettre d'illustrer le mouvement de façon statique (Gyselinck, 1995; Mautone & Mayer, 2001). Cela n'élude cependant pas complètement la question de la quantité d'informations qui reste différente d'une condition à l'autre, et par conséquent, peut ne pas avoir les mêmes effets (Hidrio & Jamet, 2002; Jamet, Le Bohec, & Hidrio, 2003).

En faisant varier les conditions d'apprentissage de l'interface DIGIME, on souhaite, dans le cadre des recommandations issues de la CLT (Sweller, Van Merriënboer, & Paas, 1998), libérer un maximum de ressources cognitives qui pourront être allouées utilement à l'apprentissage des procédures et des gestes (Anderson, 1982). La façon la plus simple pour le mesurer est d'évaluer comment le type d'illustration va affecter les performances des participants, aussi bien en vitesse de saisie que d'erreurs produites. On peut également traiter cette question différemment, en termes de transfert d'apprentissage (Perkins & Salomon, 1994). Par expérience, nous savons que nous ne prenons jamais connaissance des manuels et autres tutoriels attachés aux machines que nous utilisons… Ou alors très peu. Il est donc intéressant de savoir dans quelle mesure un didacticiel ne fournissant que des connaissances parcellaires (démonstration de quelques lettres parmi l'alphabet) va permettre à ses lecteurs d'en inférer la méthode de saisie des autres lettres, dont il n'aura pas étudié spécifiquement la production. Il s'agit d'une situation de transfert d'apprentissage de courte distance, *a priori* relativement facile. Si tel est le cas, alors les performances de saisie devraient être meilleures, ce qui influera également sur la satisfaction des utilisateurs.

5.2. Méthode

5.2.1. Participants

Quarante-neuf personnes ont participé à l'expérience. Il s'agissait d'individus tout venants, néanmoins largement composé d'étudiants, d'un âge compris entre 17 et 56 ans. La moyenne d'âge des participants était de 29,8 ans. Vingt-neuf d'entre eux étaient des femmes, tandis que 21 étaient des hommes. Aucun de nos participants n'était possesseur ou utilisateur de PDA, ce qui en faisait des utilisateurs novices dans le cadre de cette expérience. Les participants étaient tous volontaires et n'ont reçu aucune rémunération.

5.2.2. Matériel

5.2.2.1. Nouvelle interface DIGIME

Les résultats obtenus à l'issue de notre première expérience ont inspiré à nos partenaires de l'IRISA un certain nombre de modifications à l'interface DIGIME (figure 31). Cette nouvelle interface bénéficiait, en plus des améliorations ergonomiques dont nous faisons la présentation dans les lignes qui suivent, d'une version améliorée de RESIFCar, augmentant légèrement son taux d'identification des caractères (toujours calculé en dehors de tout contexte applicatif sur des bases de caractères manuscrits collectés préalablement (Bouteruche, Deconde, Anquetil, & Jamet, 2005)).

Afin d'intégrer les résultats de la première expérience, DIGIME ne laissait plus à l'écran la trace écrite du scripteur, qui disparaissait dès que le caractère était reconnu. La trace écrite du geste était remplacée dans la zone d'écriture par la lettre reconnue par la machine. La trace écrite ne s'affichait que la durée du geste graphomoteur, ce qui autorisait un contrôle rétroactif minimal (Serratrice & Habib, 1993). Cela a permis d'intégrer complètement le feedback à la zone de saisie et de tenter de répondre à l'un des problèmes soulevés par la discussion des résultats précédents, à savoir que le feedback contigu était peut-être encore trop éloigné des lettres de l'utilisateur, et redondant avec les informations fournies à la fois par l'encre du scripteur et la zone traitement de texte. C'est donc, suivant les recommandations du PCP (Wickens & Carswell, 1995) et de la CLT (Chandler & Sweller, 1991) une intégration complète des informations mutuellement référencées. Ainsi, une lettre saisie et reconnue s'affichait dans la zone d'écriture sous forme de caractère dactylographié

police « `Courrier New` ». Ce choix n'était pas anodin. Il était en effet possible pour l'utilisateur d'intervenir directement sur les lettres reconnues dans la zone de saisie (transformation, remplacement ou insertion de lettres, ainsi qu'ajout d'accents). Cette police présente comme particularité que toutes les lettres qui la composent on la même largeur (voir la barre horizontale du « `i` » par exemple), ce qui permettait au scripteur d'intervenir sur n'importe quel caractère avec la même facilité. La police « `Courrier New` » a permis aussi, du fait de la largeur des caractères, d'apporter une innovation à DIGIME : l'insertion de lettres. Cette procédure permettait, en tirant un trait de bas en haut entre deux lettres déjà reconnues, d'y insérer un espace. L'utilisateur pouvait ensuite écrire directement une lettre à l'intérieur de cet espace, qui se voyait transformé en fonction du caractère reconnu.

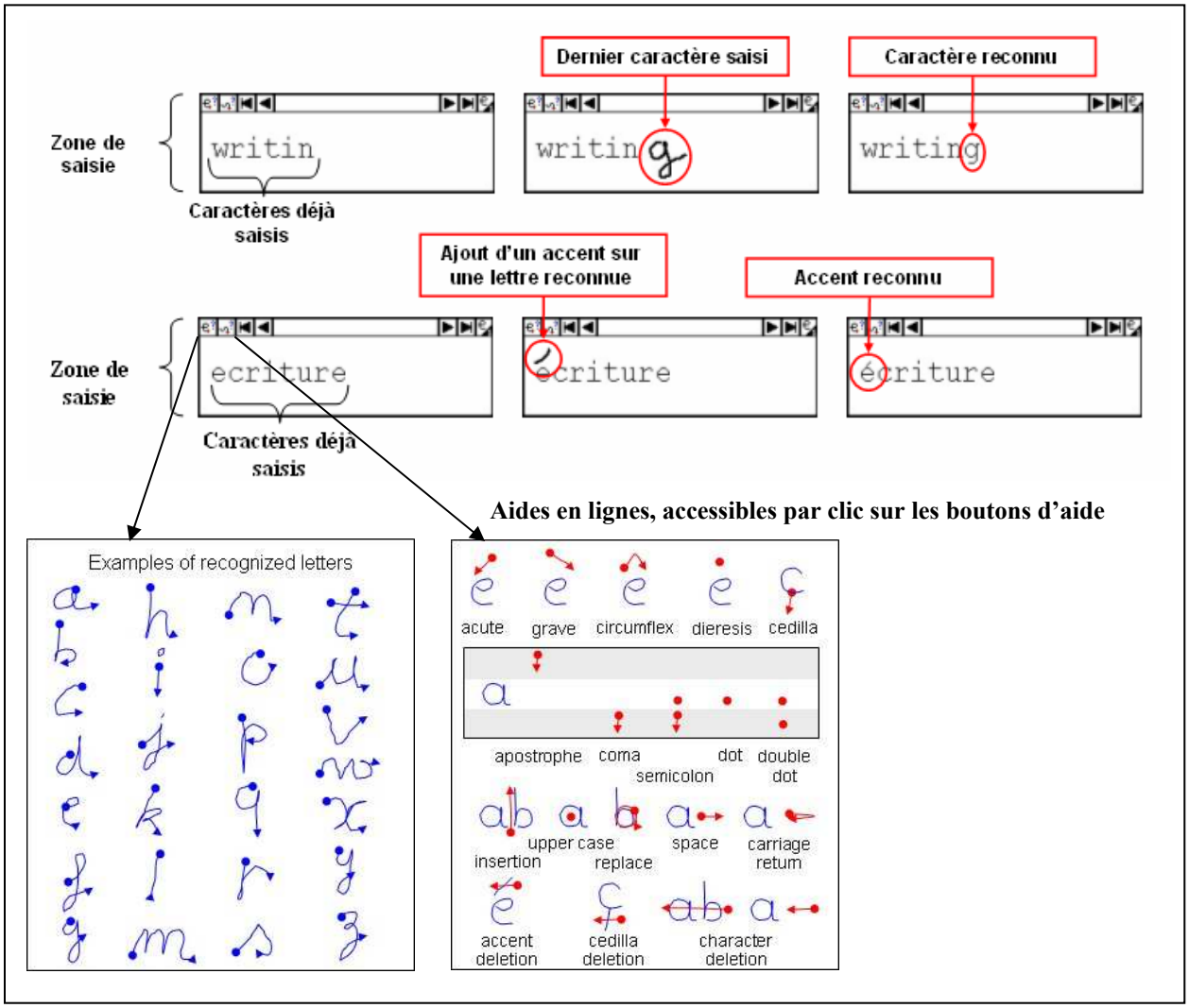


Figure 31: Fonctionnement de la nouvelle interface DIGIME (version anglaise)

Dernière modification, l'ajout de deux boutons, situés en haut à gauche de la zone de saisie, et sur lesquels un clic permettait d'accéder à deux aides en ligne (c'est-à-dire accessibles depuis l'interface sans ouverture d'une autre application). Ces aides consistaient en deux planches qui, une fois ouvertes, couvraient la totalité de la zone application (traitement de texte par exemple). Elles présentaient respectivement, des formes de lettres susceptibles d'être identifiées par le logiciel, ainsi que les gestes d'édition, signes diacritiques et la ponctuation. En présence des ces aides, l'accès à la zone de saisie était toujours possible, permettant ainsi à l'utilisateur de produire les caractères ou gestes qui lui posaient problème avec les modèles sous les yeux. Les mouvements y étaient représentés par un point (point de départ du geste) suivi d'une flèche désignant la direction geste à accomplir). La fermeture des aides en ligne se faisait par l'appui sur un bouton « ok » situé en haut à droite de l'écran avec le stylet. Ce type de pictogramme et la fonction qui y est liée est largement connu (Watzman, 2003). Le fonctionnement général de l'interface n'a pas changé par rapport à DIGIME puisque le moteur de la reconnaissance était RESIFCar : écriture script en contexte mot. A noter que si la figure 31 présente la version anglicisée de DIGIME, notre expérience a été conduite avec la version française.

5.2.2.2. Matériel informatique

Nous avons utilisé deux Pocket PC E400 Toshiba avec stylets (*Intel PXA 261* cadencé à 300 MHz, 31,22 Mo de mémoire vive, environnement *Windows CE*) ainsi que deux ordinateurs portables (*Intel Pentium III* cadencé à 1 GHz, 256 Mo de mémoire vive, environnement *Windows XP*). Durant la passation de l'expérience, les PDA étaient reliés avec des ordinateurs distants pour le suivi en direct, par deux expérimentateurs, de l'activité des participants, ce afin de s'en assurer du bon déroulement sans les gêner. La connexion entre les PDA et les PC distants était assurée par les logiciels *Microsoft ActiveSync* 3.7 et *Remote Display Control* 2.03.

5.2.2.3. Diaporama

Nous avons conçu un diaporama à l'aide de *Microsoft Powerpoint* pour faire une présentation rapide et illustrée aux participants de ce qu'était un PDA et ce à quoi il pouvait

servir. Le but de la chose était de les informer sur les fonctions générales de ce type de machines qui leur étaient souvent inconnues.

5.2.2.4. Matériel linguistique

Pour réaliser cette expérience, nous avons déterminé, sur la base de critères variés, un ensemble de tâches de saisie de plus en plus complexes.

L'analyse des erreurs produites pour chaque lettre par tous les participants lors de la première expérience nous a permis de déterminer quelles lettres étaient plus facilement reconnaissables par le logiciel. Ces taux d'erreur de reconnaissance (imputables aux utilisateurs ou à la machine sans que l'on puisse en déterminer la part respective) ont été calculés, à dessein, sans tenir compte de la fréquence d'apparition des lettres au cours de l'expérience 1. A titre d'exemple, plus de 2000 erreurs ont touché la lettre « e », mais celle-ci est apparue à 22 reprises. Le « h » n'est apparu qu'une seule fois, mais a donné lieu à plusieurs centaines d'erreurs. Il était sans doute au moins aussi difficile à faire reconnaître au logiciel que le « e ». Cependant, en regard de sa fréquence d'apparition, il était nettement moins problématique que ce dernier. C'est pourquoi nous l'avons considéré comme une lettre facile à écrire, ou, plus précisément, comme une lettre qui ne posait pas (beaucoup) de problème, par rapport à sa fréquence d'apparition dans la langue française. Nous avons donc choisi, de façon arbitraire, à partir de la figure 32, « a, b, h, m, o, et p » comme lettres faciles, et « e, i, n, r, t, et u » comme lettres difficiles. Cette méthode d'évaluation est contestable, mais le matériel linguistique de la première expérience présentait dix-huit lettres sur les vingt-six que contient l'alphabet, et leurs fréquences d'apparition dans notre liste concordent relativement bien avec les fréquences d'apparition des lettres dans la langue française (New & Pallier, 2001). Ces caractères ont été complétés par un ensemble de signes diacritiques : « ç, é, è, et ë ». Nous n'avons pas pu juger de leur difficulté, puisque seul le « é » était présent dans les mots de l'expérience 1.

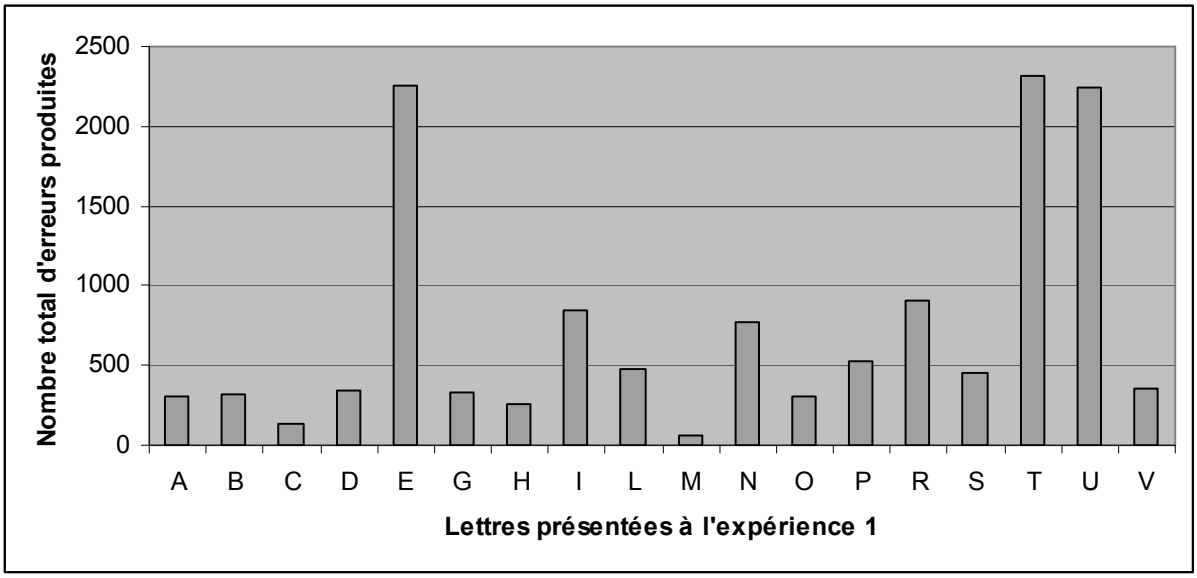


Figure 32: Nombre total d'erreurs par lettres produites pour l'ensemble des participants de l'expérience 1

Ensuite, nous avons choisi dans BRULEX (Content, Mousty, & Radeau, 1990) cinq mots qui présentaient entre eux de fortes similitudes linguistiques : égard, hôtel, maman, odeur et vertu (voir en annexes, section 15.2). Là encore, le choix était arbitraire.

Pour finir, en nous appuyant sur les travaux de MacKenzie et Soukoreff (2003), nous avons élaboré trois phrases que les participants avaient à écrire. Suivant les recommandations des auteurs, nous avons fait en sorte qu'elles soient « courtes, faciles à retenir, et représentatives du niveau d'énonciation visé » (MacKenzie & Soukoreff, 2003, notre traduction). Ainsi, nous avons proposé « c'est le printemps », « l'expérience est intéressante », et « rendez-vous demain ».

5.2.2.5. Didacticiels

Nous avons conçu un didacticiel qui présentait de façon linéaire, les différentes fonctions et méthodes propres à DIGIME. Ce didacticiel était décliné sous deux formes : l'une avec illustrations statiques (voir en annexes, section 15.5), l'autre avec illustrations dynamiques. Cette différence n'apparaissait que pour la présentation des lettres, signes diacritiques et gestes d'édition. Toutes les autres planches étaient identiques, que la condition soit statique ou dynamique. Dans la condition statique, le mouvement était représenté par l'ajout de flèches à côté du caractère (en noir) dont on faisait la démonstration de l'écriture, simulant ainsi la

progression du geste dans le temps (voir figure 33). Dans la condition dynamique, le mouvement était représenté lui-même par la coloration progressive du caractère noir en rouge, s'achevant sur une image fixe représentant l'image statique (noire) correspondant, avec les flèches (rouge). Ainsi, la quantité d'informations devait être identique d'une condition à l'autre (Bétrancourt, Bauer-Morrison, & Tversky, 2001).

Le didacticiel n'offrait pas la possibilité aux participants de s'exercer eux-mêmes pendant sa présentation. Pour chaque exemple de caractère, il ne présentait qu'un allographe, mais informait les participants que pour une même lettre, différentes formes pouvaient être reconnues. Le déroulement du didacticiel était sous le contrôle total des participants : ils pouvaient passer d'une planche à l'autre quand bon leur semblait, et s'attarder sur une planche s'ils le souhaitaient. Dans le cas du didacticiel animé, l'animation se répétait tant que les participants ne sollicitaient pas l'affichage d'une autre planche.

Le didacticiel, indépendamment de sa version (statique ou dynamique), ne présentait en guise d'exemples que les lettres « a, i, m, n, p, et t », ainsi que les signes diacritiques « é, ê, et ë ».

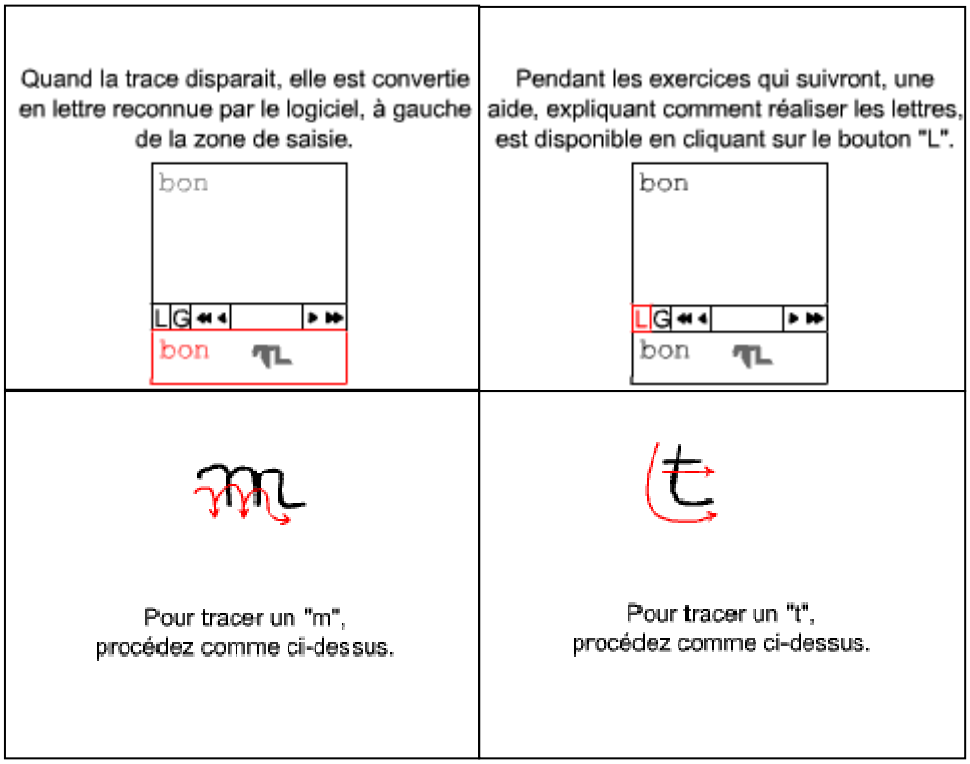


Figure 33: Exemples de planches proposées par le didacticiel statique

5.2.2.6. Questionnaires

En début de passation, nous avons administré aux participants un pré-questionnaire. Il consistait en plusieurs questions dont nous pensions qu'elles étaient à même de « trahir » des utilisateurs experts (Farrington-Darby & Wilson, 2006; Morahan-Martin & Schumacher, 2007). A titre d'exemple, « Que signifie PDA ? » ou « Qu'est-ce que *Transcriber* ? » sont des questions qui y figuraient.

En fin de test, nous avons choisi de faire remplir aux participants le test SUS (*System Usability Scale*, voir en annexes, section 15.3.5) (Brookes, 1996). Nous avons estimé ce test être le plus apte à révéler la satisfaction des utilisateurs au travers de l'utilisabilité de DIGIME. Le test SUS ambitionne justement d'évaluer l'utilisabilité en contexte. Or, cette expérience vise à évaluer l'utilisabilité en fonction du contexte didacticiel, plus qu'intrinsèquement, comme cela était le cas pour la première expérience. Les dix items le composant étaient les suivants :

1. Je pense que j'aimerai utiliser ce système de saisie manuscrite fréquemment.
2. J'ai trouvé ce système de saisie manuscrite inutilement complexe.
3. J'ai pensé que ce système de saisie manuscrite a été facile à utiliser.
4. Dans une utilisation future, je pense que j'aurais besoin de l'aide d'une personne expérimentée pour pouvoir utiliser le système.
5. J'ai trouvé que les différentes fonctions ont été bien intégrées au système de saisie manuscrite.
6. J'ai trouvé qu'il y avait trop d'incohérences dans le système de saisie manuscrite.
7. J'imagine que la plupart des gens apprendront rapidement à utiliser ce système de saisie.
8. J'ai trouvé ce système très désagréable à utiliser.
9. Je me suis senti(e) très confiant(e) pendant l'utilisation de ce système.
10. Il a fallu que j'apprenne beaucoup de choses avant de pouvoir utiliser ce système.

Les questions ont été traduites et adaptées par nos soins. Les réponses des participants étaient consignées sur des échelles type Lykert en cinq points, s'étendant de « pas du tout d'accord » à « tout-à-fait d'accord » (de 0 à 4, pour le calcul du score global). Brookes (1996)

insiste sur l'importance du choix d'un nombre de graduation impair pour permettre aux personnes interrogées de se « réfugier » sur la position centrale en cas de refus de prendre position. Le score global au test SUS, se calcule en additionnant, pour les items 1, 3, 5, 7, et 9 le score à la question moins un, et pour les items 2, 4, 6, 8, et 10, cinq moins le score à la question. Le résultat obtenu, multiplié par 2,5 donne un résultat compris entre 0 (inutilisable) et 100 (très utilisable).

5.2.3. Procédure expérimentale

L'accueil des participants était suivi par la complétion du pré-questionnaire, avant que ne leur soit présenté le diaporama explicatif. A la suite de quoi ils ont été soumis, suivant la condition expérimentale, à la présentation de l'un ou l'autre des didacticiels, soit à aucun didacticiel (condition contrôle). Dans ce dernier cas, ils commençaient immédiatement la tâche de saisie. Ils étaient néanmoins avertis de la possibilité d'utiliser les aides en ligne. Pour les participants qui bénéficiaient de la présentation d'un didacticiel, son déroulement était entièrement sous leur contrôle. Ce contrôle était linéaire (bouton suivant). Ils n'avaient pas accès à un menu hypertexte leur permettant de choisir quelle partie du didacticiel ils allaient consulter. Par conséquent, tout le didacticiel défilait sous leurs yeux, même très rapidement s'ils le désiraient. L'épreuve de saisie s'est déroulée de la façon suivante, par blocs : d'abord les lettres, puis les signes diacritiques, venaient ensuite les mots, et pour finir, les phrases. Cet ordre de présentation par bloc était important puisque nous pensions que les effets des didacticiels ne se feraient sentir qu'en début d'épreuve, avant que les participants aient eu accès aux aides en ligne et à l'entraînement par la pratique (Chevalier, Hall, & Nadeau, 1990). A l'intérieur de chaque bloc, l'ordre de présentation était randomisé, afin d'éviter au maximum des effets d'ordre. L'apparition des lettres, accents, mots ou phrases suivants ne pouvait se faire qu'à l'issue de la réussite complète de l'exercice en cours. La présentation des items à écrire s'est faite de la même façon que pour l'expérience 1 (voir figure 23). A partir du moment où un item était présenté au participant, c'est l'appui sur la touche « exercice » qui permettait la saisie du mot, jusqu'à ce que la séquence de lettres à écrire ait été complètée, activant un bouton « suivant ». Quand l'épreuve était terminée, les participants répondaient au questionnaire SUS. La passation durait une trentaine de minutes.

5.2.4. Variables indépendantes

Un facteur intergroupe était manipulé dans cette expérience.

- **Didacticiel :** Ce facteur à trois modalités consistait à faire varier le type d'illustrations, suivant qu'elles étaient statiques ou dynamiques. Une dernière modalité associée à cette VI était l'absence de didacticiel (condition contrôle).

Deux facteurs intragroupes étaient aussi introduits dans le plan expérimental.

- **Difficulté des lettres à être reconnues :** Cette VI consistait, sur la base des erreurs produites au cours de la première expérience, à choisir des lettres dont on savait qu'elles posaient des problèmes d'identification par DIGIME. Elle avait deux modalités, facile et difficile. Les lettres faciles étaient, pour mémoire, « a, b, h, m, o, et p », et les lettres difficiles « e, i, n, r, t, et u » (indépendamment des autres blocs de l'épreuve comme les mots ou les phrases).
- **Apprentissage :** Cette VI à deux modalités faisait varier le fait que parmi l'ensemble de lettres que les participants ont eu à écrire, certaines ont été explicitement vues au cours du didacticiel, alors que d'autres non. Ce facteur ne s'appliquait évidemment pas à la condition contrôle pour laquelle l'ensemble des lettres était inconnu. Les lettres présentées par le didacticiel étaient, pour mémoire, « a, i, m, n, p, et t », tandis que « b, e, h, o, r, et u » ne l'étaient pas (indépendamment des autres blocs de l'épreuve comme les mots ou les phrases).

5.2.5. Variables dépendantes

Pour cette expérience, l'interface DIGIME a été modifiée par ses concepteurs de l'IRISA. Elle enregistrait ainsi de façon invisible pour les scripteurs les temps de saisie de chaque lettre, mot, ou signe diacritique. Le chronomètre se déclenchait dès le clic sur le bouton exercice pour n'être arrêté que par la réussite à l'exercice. Ceci n'a cependant pas été possible pour les phrases, du fait de difficultés techniques. Par contre, l'application enregistrait le nombre de gestes de correction produits par les participants. L'intérêt était pour nous d'avoir un bon indicateur du nombre d'erreurs sans avoir à passer par le traitement fastidieux des traces écrites avec Caractest. Nous avons également eu accès aux durées de consultation des

aides en ligne, associées au nombre d'occurrences de consultation, ainsi qu'aux gestes d'insertion, ceci pour chaque type de lettres, accents, mots et phrases.

Grâce à la procédure de randomisation par bloc, nous avons traité les temps de saisie par type de lettres, en fonction du croisement des VI intra : lettres faciles apprises (a, m, et p), lettres faciles non apprises (b, h, et o), lettres difficiles apprises (i, n, et t) et lettres difficiles non apprises (e, r, et u). Les signes diacritiques étaient randomisés de la même façon, mais n'ont pas été traités selon les VI apprentissage et facilité qui ne s'appliquaient qu'aux lettres.

Nous avons enregistré aussi le temps total de consultation du didacticiel (sans qu'il ne nous ait été possible de savoir combien de temps les participants ont passé sur chaque planche). Cette mesure ne nous a servi qu'à vérifier que les participants consultaient véritablement le didacticiel.

La dernière VD était le score global au test SUS, calculé selon la méthode que nous avons développée plus haut.

5.2.6. Hypothèses

- H1: La consultation préalable du didacticiel dynamique engendre de meilleures performances (temps de saisie plus courts, moins de corrections et de consultation des aides en lignes) et une plus grande utilisabilité perçue, que le didacticiel statique, lui-même plus efficace que l'absence de didacticiel.

- H2: Cette supériorité du dynamique sur le statique est plus forte sur les lettres difficiles que sur les lettres faciles. De même, cet effet est plus marqué sur les lettres non apprises que sur les lettres apprises.

5.3. Résultats

5.3.1. Effets des VI sur les performances

Constatant que les distributions de nos VD n'étaient pas normales, associé au fait que la condition d'homogénéité des variances n'était pas remplie pour un grand nombre d'entre

elles, nous avons procédé à des analyses statistiques non paramétriques. Ceci est gênant puisque nous avons formulé une hypothèse d'interaction entre la VI difficulté d'une part, et la VI apprentissage d'autre part, avec le type d'illustration du didacticiel. C'est pourquoi, pour tester ces hypothèses, nous avons pris quand même le risque de violer les conditions d'application de l'ANOVA (modèle linéaire général, plan à mesures répétées). Nous avons limité cette transgression aux VD concernant le bloc « lettres », seul concerné par le traitement de nos VI intra. Nous nous sommes appuyés sur les résultats des tests multivariés puisque nos données ne respectaient pas la condition de sphéricité.

Par la suite, nous avons utilisé le test de Kruskall Wallis pour analyser toutes les autres VD.

Les performances moyennes sont indiquées par la lettre *M*, suivi de l'écart-type *SD*. Ces résultats sont exprimés en secondes pour les durées, sans unités pour les nombre d'occurrences.

5.3.1.1. Analyse des effets du type d'illustration sur les temps de saisie, le nombre d'erreurs, les durées de consultation des aides en lignes en fonction du type de lettre

Le détail des effets sur les performances moyennes des participants par condition (avec didacticiel statique ou dynamique) est disponible en annexe (section 15.6). Nous ne reproduisons pas ici ces résultats pour deux raisons : ils sont fort nombreux (nous avons mesuré beaucoup de VD) et aucun n'a témoigné d'un effet significatif du facteur didacticiel suivant que les lettres étaient faciles, ou pas, et apprises, ou non. Le facteur didacticiel renvoie ici aux modalités statique et dynamique uniquement, pour cette analyse, puisque la distinction entre lettres apprises et non apprises n'était pas valable pour la condition contrôle qui n'a pas vu le didacticiel.

Bien qu'aucune interaction entre type de didacticiel (statique/dynamique) et type de lettres (faciles/difficiles ; apprises/non apprises) n'ait été révélée par l'analyse, des effets propres aux types de lettres sont apparu. En particulier, on notera que la facilité a un effet sur le temps de consultation de l'aide pour les gestes d'édition ($F(1,29) = 4,240$, $p = .04$) se traduisant par de

plus longues consultation quand les lettres sont difficiles. A l'inverse, lorsque les lettres n'ont pas été apprises, elles ont occasionné plus de corrections (F(1,29) = 4,574, *p* = .04).

5.3.1.2. Analyse des effets du type de didacticiel (statique, dynamique, aucun)

Ci-dessous sont résumés les principaux résultats de l'expérience.

Type de tâche	VD	Aucun	Statique	Dynamique
Lettres faciles	Tps de saisie (s)	138,1 (53,4)	146,1 (56,6)	77,0 (58,5)
	Nbr d'effacements	4,7 (1,3)	4,6 (1,4)	5,9 (1,5)
	Tps aide gestes (s)	20,5 (4,2)	0,1 (4,4)	2,8 (4,6)
	Tps aide lettres (s)	15,1 (3,4)	5,0 (3,6)	8,1 (3,8)
Lettres difficiles	Tps de saisie (s)	120,0 (21,5)	132,8 (22,8)	78,7 (23,5)
	Nbr d'effacements	8,5 (2,4)	8,2 (2,5)	4,9 (2,6)
	Tps aide gestes (s)	9,1 (4,1)	13,0 (4,4)	3,3 (4,5)
	Tps aide lettres (s)	11,0 (3,0)	8,1 (3,2)	3,0 (3,3)
Total lettres	Tps de saisie (s)	258,1 (62,2)	278,9 (65,9)	155,8 (68,1)
	Nbr d'effacements	13,2 (3,1)	12,9 (3,3)	10,8 (3,4)
	Tps aide gestes (s)	29,7 (5,3)	13,2 (5,6)	6,1 (5,8)
	Tps aide lettres (s)	26,2 (5,8)	13,1 (6,1)	11,1 (6,3)
Signes diacritiques	Tps de saisie (s)	105,6 (89,2)	103,1 (75,2)	82,0 (46,4)
	Nbr d'effacements	9,9 (10,4)	10,1(9,4)	6,3 (5,1)
	Tps aide gestes (s)	2,7 (3,5)	3,1 (3,7)	6,7 (7,6)
	Tps aide lettres (s)	14,2 (16,1)	14,3 (19,4)	9,5 (7,6)
Mots	Tps de saisie (s)	248,1 (60,1)	167,9 (63,8)	141,5 (65,8)
	Nbr d'effacements	18,3 (3,1)	21,0 (3,3)	13,0 (3,4)
	Tps aide gestes (s)	4,6 (1,6)	1,5 (1,7)	1,3 (1,8)
	Tps aide lettres (s)	3,4 (2,0)	3,7 (2,1)	3,6 (2,2)
Phrases	Nbr d'effacements	38,5 (5,3)	40,4 (5,6)	31,7 (5,8)
	Tps aide gestes (s)	17,7 (3,7)	14,5 (3,9)	13,8 (4,1)
	Tps aide lettres (s)	1,9 (1,0)	2,4 (1,1)	2,8 (1,1)
	Nbr insertions	0 (0)	0,1 (.12)	1,6 (.67)

Tableau 2: Principaux résultats de l'expérience, en fonction de la VI didacticiel (Moyenne, et, entre parenthèses, écart-type)

Pour conduire cette analyse, nous avons créé de nouvelles variables en additionnant les scores de performance (temps de saisie, nombre d'erreurs, temps et occurrences de consultation des aides en lignes) obtenus pour les lettres faciles d'une part, et les lettres difficiles d'autre part, indépendamment qu'elles aient été apprises ou pas, puisque la

condition contrôle, privée de didacticiel, n'a pas pu y être sensible. Nous avons calculé également, par addition, et pour ces mêmes VD, les scores pour l'ensemble des lettres.

Les résultats du test de Kruskall Wallis sur nos données ont montré des différences significatives entre les conditions, suivant que les participants n'aient bénéficié d'aucun didacticiel, d'un didacticiel statique ou d'un didacticiel dynamique. Cet effet du type de didacticiel n'est apparu que sur les temps de consultation de l'aide pour les gestes d'édition au cours de la saisie des lettres faciles (χ^2 (2,N = 49) = 12,896, p = .00). Les participants qui n'ont pas eu accès au didacticiel ont consulté cette aide beaucoup plus (M = 21,7s, SD = 27) que ceux qui ont vu le didacticiel (respectivement M = 0,1s, SD = 0,7 en statique et M = 2,8s, SD = 9,2 en dynamique). De façon tout-à-fait logique, ce résultat a été confirmé par le nombre d'occurrences de consultation de cette aide (χ^2 (2,N = 49) = 11,637, p = .00). Cet effet ne s'est pas manifesté sur les autres VD telles que temps de saisie (χ^2 (2,N = 49) = 3,067, p = .21), nombre de correction (χ^2 (2,N = 49) = .413, p = .81), ou temps de consultation de l'aide pour les lettres (χ^2 (2,N = 49) = 3,674, p = .15).

A l'inverse, les groupes ne se sont pas différenciés sur la saisie des lettres difficiles (temps de saisie (χ^2 (2,N = 49) = 1,785, p = .41), nombre de correction (χ^2 (2,N = 49) = .779, p = .67), temps de consultation des aides pour les lettres (χ^2 (2,N = 49) = 3,083, p = .21) et les gestes (χ^2 (2,N = 49) = 1,804, p = .40)). Pourtant, l'effet précédemment évoqué (temps de consultation de l'aide pour les geste d'édition) s'est confirmé sur la totalité des lettres isolées à écrire (χ^2 (2,N = 49) = 14,272, p = .00 ; avec respectivement M = 8,7s, SD = 16,5 sans didacticiel, M = 13,0s, SD = 24,1 en statique et M = 3,3s, SD = 7,0 pour le dynamique). Aucune autre des mesures réalisées sur l'ensemble des lettres seules n'a été affectée par le facteur didacticiel (temps de saisie (χ^2 (2,N = 49) = 4,244, p = .12), nombre de correction (χ^2 (2,N = 49) = .530, p = .76), et temps de consultation des aides pour les lettres (χ^2 (2,N = 49) = 3,152, p = .20))

Les figures 34 et 35 donnent un aperçu de ces résultats, page suivante.

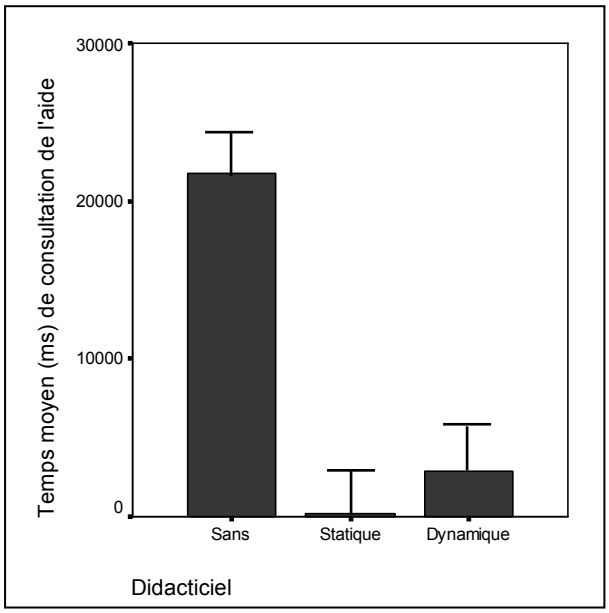


Figure 34:Durées moyennes de consultation de l'aide gestes d'édition pendant la saisie des lettres faciles en fonction du didacticiel

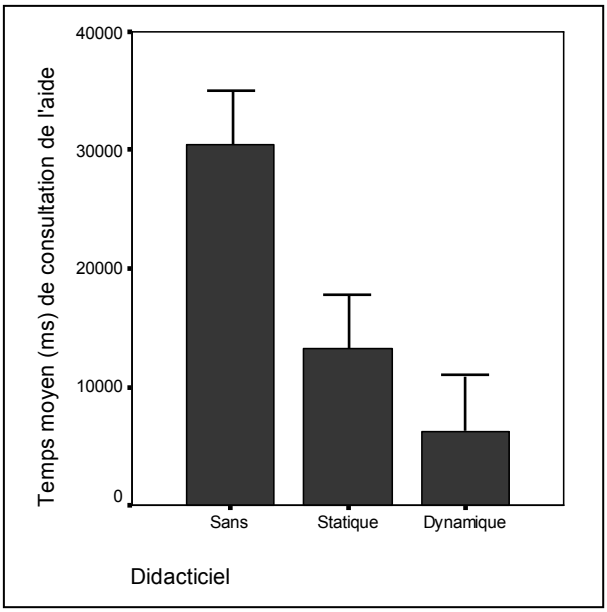


Figure 35: Durées moyennes de consultation de l'aide gestes d'édition pendant la saisie de toutes les lettres en fonction du didacticiel

Les signes diacritiques n'ont montré aucune sensibilité au type de didacticiel, et sur aucune des VD mesurées (temps de saisie (χ^2 (2,N = 49) = .157, p = .92), nombre de correction (χ^2 (2,N = 49) = 1,191, p = .55), temps de consultation des aides pour les lettres (χ^2 (2,N = 49) = 3,477, p = .17) et les gestes (χ^2 (2,N = 49) = .232, p = .89)). Cependant, la tâche de saisie des différents accents et de la cédille n'était pas pure : elle impliquait la saisie préalable de la lettre support, en l'occurrence, « e » ou « c ». Or, le « e » figurait parmi nos lettres difficiles, et, comme nous venons de le voir, les lettres difficiles n'ont pas été affectées par la VI didacticiel. A cela s'ajoute que le « e » était répété trois fois pour les accents « ë », « ê » et « è ».

L'absence d'effet a également été valable pour les mots (temps de saisie (χ^2 (2,N = 49) = .853, p = .41), nombre de correction (χ^2 (2,N = 49) = 2,318, p = .31), temps de consultation des aides pour les lettres (χ^2 (2,N = 49) = .506, p = .77) et les gestes (χ^2 (2,N = 49) = .808, p = .66)) ainsi que les phrases (nombre de correction (χ^2 (2,N = 49) = .483, p = .78), temps de consultation des aides pour les lettres (χ^2 (2,N = 49) = .905, p = .95) et les gestes (χ^2 (2,N = 49) = 1,787, p = .40)). A noter cependant que pour la saisie des phrases, le nombre de procédures d'insertion a varié significativement entre les conditions ((χ^2 (2,N = 49) = 9,626, p = .00)), puisque sans didacticiel, les participants n'ont jamais utilisé cette procédure, alors qu'en statique (M = 0,1, SD = 00) et en dynamique (M = 1,6, SD = 2,6), cela a été le cas. Même si cet effet n'a eu qu'une faible amplitude, il est cependant très intéressant (voir figure 36).

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

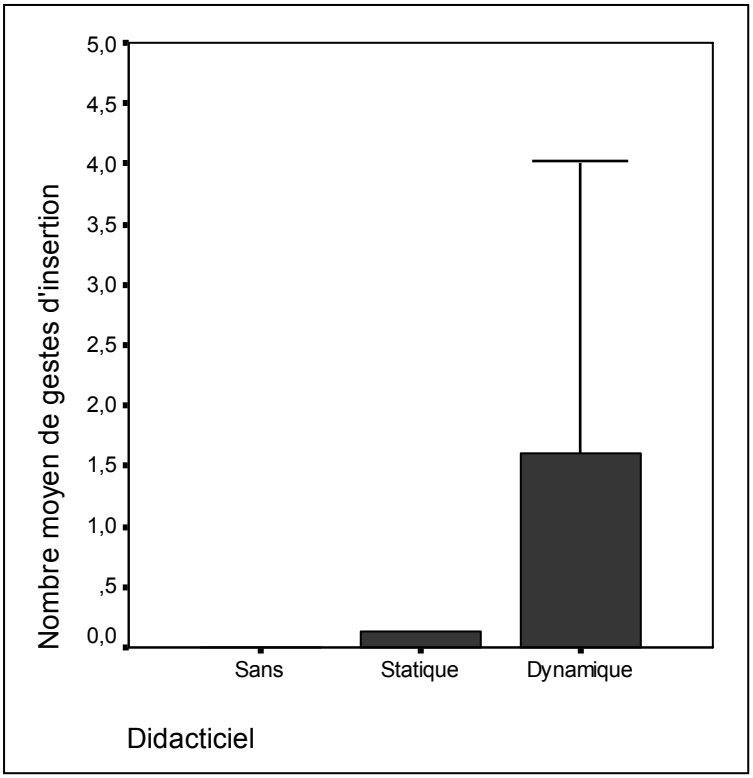


Figure 36: Nombre moyen de geste d'insertion en fonction du didacticiel pendant la saisie des phrases en fonction du didacticiel

S'agissant des insertions, il est à noter que les participants n'en avaient jamais fait usage avant la saisie de phrases, même pour les mots. Cette VD ne rentrait d'ailleurs dans notre analyse statistique qu'à partir des phrases, puisque, n'ayant pas été mise à contribution avant, nous n'avons pas jugé utile d'encombrer nos fichiers de résultats de colonnes vides.

5.3.2. Effets du type de didacticiel sur l'utilisabilité perçue

Le didacticiel n'a eu qu'un effet marginal sur l'utilisabilité perçue mesurée par le test SUS (χ^2 (2,N = 49) = 5,418, p = .06). Sans didacticiel, le score au test SUS est de M = 32,8, SD = 11,9 contre M = 39,3, SD = 12,7 en statique et M = 41,8, SD = 14,2 pour le dynamique.

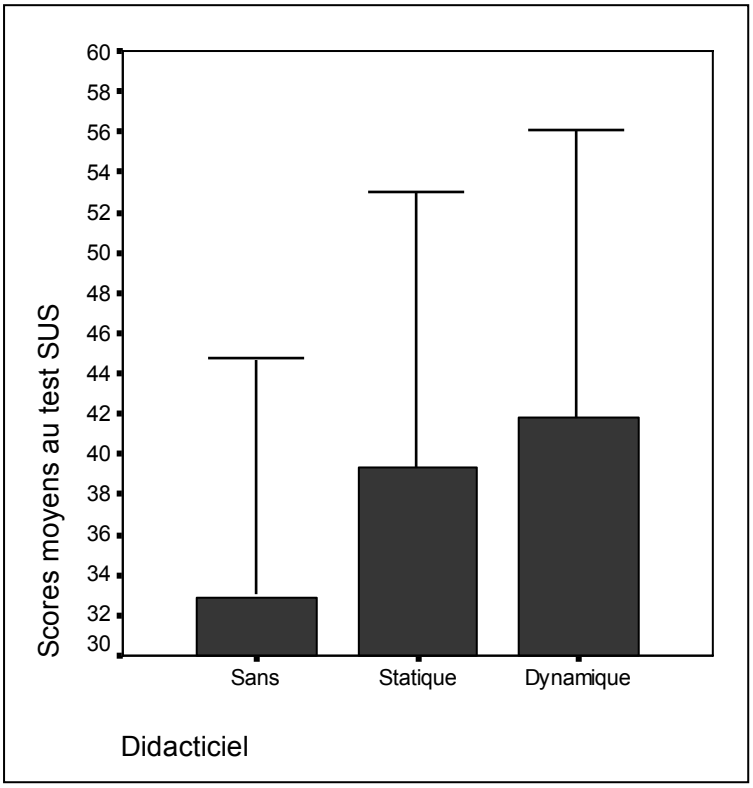


Figure 37: Utilisabilité perçue de l'interface DIGIME en fonction du didacticiel

5.3.3. Effets de la présence d'un didacticiel

Les résultats au test SUS en fonction du type didacticiel nous laissent penser que la présence d'un didacticiel avait un intérêt. Si, bien sûr, montrer que la présence d'un didacticiel (qu'il soit statique ou dynamique) autorise de meilleures performances qu'en l'absence de didacticiel est trivial, il nous semble cependant intéressant de démontrer que sa présence est accueillie positivement, en termes d'utilisabilité perçue de l'interface, par les participants. Rappelons que le déroulement du didacticiel, quand il était montré aux participants, ne dépendait que d'eux (s'il l'on excepte que le contrôle de l'activité des participants sur un ordinateur distant, ce dont ils étaient informés, a dû les inciter sans nul doute à essayer de faire les choses telles qu'ils pensaient que l'expérimentateur les voulait).

Pour tester cela, nous avons codé une nouvelle VI didacticiel, intergroupe, à deux modalités : avec ou sans le didacticiel (indépendamment de son caractère statique ou dynamique). Les résultats au test de Mann-Whitney ont montré que l'utilisabilité perçue de

l'interface DIGIME a été sensiblement affectée par la présence d'un didacticiel (U=188,00 ; Z=-2,135, *p* =.00), puisque avec un didacticiel, le score moyen rapporté au test SUS est de *M* = 40,5, *SD* = 13,3 (contre *M* = 32,8, *SD* = 11,9 sans didacticiel).

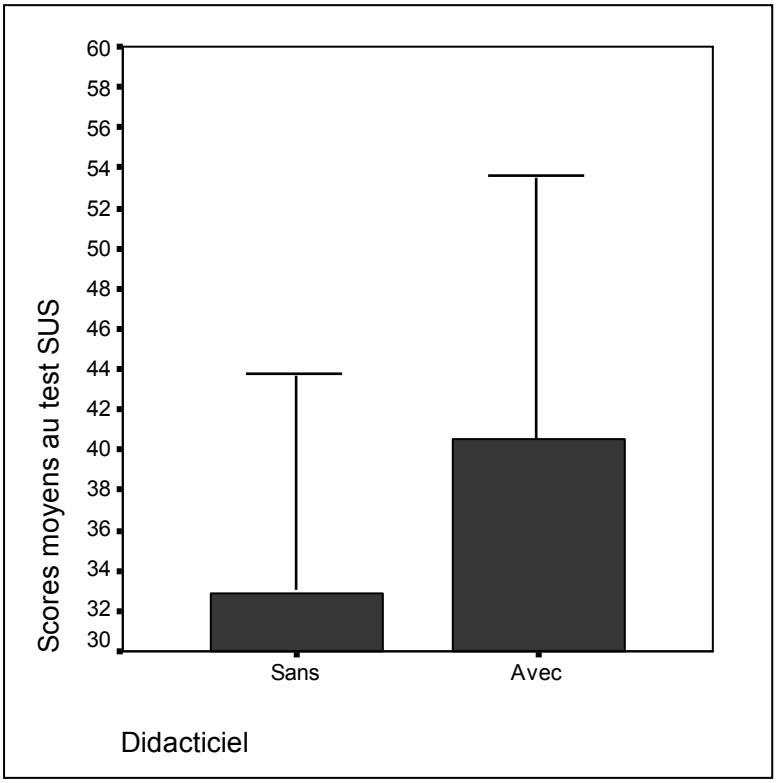


Figure 38: Utilisabilité perçue de l'interface DIGIME en fonction de la présence ou de l'absence de didacticiel

5.4. Discussion

Nos hypothèses ne sont que très partiellement validées. Nous avons échoué à mettre en évidence que le didacticiel dynamique permettait un gain de performances (aussi bien pour les temps de saisie que ceux de consultation des aides, ainsi qu'en termes d'erreurs), et ce, plus particulièrement pour les lettres difficiles et les lettres non apprises. Nous ne pouvons donc que conclure qu'il n'y a pas eu d'effet du didacticiel sur le transfert de connaissance de l'apprentissage de certaines lettres à d'autres, non apprises. Des travaux antérieurs ont laissé entendre que ce transfert ne se produisait que dans certaines conditions de similarité entre les deux contextes d'apprentissage, et il est possible qu'apprendre à un utilisateur à faire une reconnaître une lettre à DIGIME n'améliore pas la reconnaissance d'une autre (non apprise),

car trop différente (Perkins & Salomon, 1994; Thorndike & Woodworth, 1901). Cette équivalence des illustrations dynamiques avec les illustrations statiques est cependant bien connue, et n'a rien de surprenant en regard des études antérieures sur la comparaison statique/dynamique (Bétrancourt, Bauer-Morrison, & Tversky, 2001; Bétrancourt & Tversky, 2000; Zacks & Tversky, 2003). Cela peut s'expliquer d'autant mieux par le soin que nous avons accordé à contrôler la quantité d'information entre les deux formats de présentation par l'ajout de flèches (Gyselinck, 1995; Hidrio & Jamet, 2002, 2008; Jamet, Le Bohec, & Hidrio, 2003; Mautone & Mayer, 2001). De surcroît, le fait d'être passé outre les conditions d'application de l'ANOVA n'a pu que nuire à la puissance du test, sur un échantillon déjà de petite taille. Mais plus encore, la consultation libre des aides en ligne n'a pu qu'annihiler les effets du type d'illustration du didacticiel, ces effets devant être, de toute évidence relativement faibles.

Par contre, les résultats montrent très clairement l'intérêt de l'intégration d'un didacticiel (qu'il soit statique, ou dynamique) à DIGIME. Le temps que les participants de la condition contrôle ont passé à consulter l'aide en ligne pour s'informer sur les gestes d'édition pendant la saisie de les lettres faciles le montre. De ce constat, on peut déduire un certain nombre d'enseignements :

Pour commencer, les durées de consultation de l'aide en ligne pendant la saisie des lettres isolées montrent que ce qui n'a pas été appris pendant la consultation d'un didacticiel (qu'il soit statique ou dynamique) doit l'être à un moment donné. L'aide en ligne pour les gestes leur évite sans doute des tâtonnements inutiles. Ce temps de consultation, en tout état de cause, reste relativement réduit par rapport au temps de saisie total des lettres puisque l'on ne relève aucun effet du didacticiel sur les temps d'écriture (que ce soit sur les lettres faciles ou difficiles, ou même sur les temps de saisie de tout autre item). D'autre part, l'entraînement progressif, par la pratique en atténue encore plus les effets (Chevalier, Hall, & Nadeau, 1990; Thomas, 1997).

L'effet de la présence d'un didacticiel ne s'est vérifié que sur les temps de consultation de l'aide pour les gestes d'édition pendant l'écriture des lettres. On peut l'expliquer en arguant que les gestes d'édition étant spécifiques à l'interface DIGIME, ils étaient donc inconnus des utilisateurs qui n'avaient pas vu un didacticiel. La consultation de cette aide leur a permis de prendre connaissance du fonctionnement général de l'interface et d'élaborer des schémas en MLT déclarative, en attendant que la pratique les transforme, par compilation de

connaissances, en automatismes procéduraux (Anderson, 1982 ; Anderson, 1987; Lee & Anderson, 1997; Taatgen & Lee, 2003). A défaut de pouvoir vérifier si cette dernière affirmation est exacte, tout donne à penser que c'est le cas, puisque les effets du didacticiel ne s'exercent plus sur les VD enregistrées pendant la saisie des mots ou des phrases. L'entraînement a joué là encore, sans surprise, son rôle primordial (Chevalier, Hall, & Nadeau, 1990; Thomas, 1997).

Rien n'a montré un effet de la présence ou de l'absence de didacticiel sur la qualité de la production écrite, en termes de « lisibilité » pour la machine (temps ou erreurs). On ne peut donc pas juger en quoi il a favorisé la sélection d'allographes adaptés et la programmation motrice qui les déterminait (Caramazza & Miceli, 1990; Kandel, Alvarez, & Vallée, 2006; Kandel, Valdois, & Orliaguet, 2003; Serratrice & Habib, 1993; Zesiger, 1995). Ce résultat n'est cependant pas surprenant. Ce type d'apprentissage est relativement long, et la durée de l'épreuve sans doute insuffisante. Il apparaît quand même, que, si la présentation visuelle d'une lettre manuscrite active les aires corticales qui permettent sa réalisation (Longcamp, Anton, Roth, & Velay, 2003; Longcamp et al., 2008; Longcamp, Boucard, Gilhodes, & Velay, 2006), les connaissances antérieures et automatismes des participants sur la façon d'écrire, ainsi que l'accès indifférencié à l'aide pour les lettres expliquent sans nul doute des performances équivalentes. L'importance de l'entraînement pour ce genre d'apprentissage s'affirme là aussi de toute évidence, et la faiblesse des effets issus de nos didacticiels (s'ils existent), ne permet pas les révéler (Chevalier, Hall, & Nadeau, 1990; Thomas, 1997). On en prend la mesure en constatant les faibles durées de consultations de l'aide pour les lettres en fin de test, pour la saisie des phrases, et ce, toutes conditions confondues. Si nos choix théoriques sur cette question ne sont pas des plus opportuns par rapport à notre problématique (Caramazza & Miceli, 1990; Kandel, Alvarez, & Vallée, 2006; Kandel, Valdois, & Orliaguet, 2003; Serratrice & Habib, 1993; Zesiger, 1995), nous soulignerons que l'apprentissage, chez le scripteur adulte, d'allographes visant à être reconnus par un dispositif électronique est assez mal documenté…

Un résultat intéressant tient à la confirmation du bien-fondé du partage des lettres que nous avons choisi de faire écrire aux participants, suivant qu'elles soient faciles ou difficiles. Contrairement à nos attentes, la réalisation des lettres difficiles n'a pas profité de la présence du didacticiel. Cela signifie qu'en situation de difficulté, même les participants ayant visionné le didacticiel (statique ou dynamique) consultent autant les aides (pour les lettres et gestes

d'édition) que les participants n'y ayant pas eu accès. Cela révèle donc bien toute l'importance qu'il y a à mettre à la disposition des utilisateurs un système d'aide rapide et simple, immédiatement accessible. Ces aides, si simples soient-elles dans leur conception, sont capables de rivaliser avec le didacticiel (statique ou dynamique), et sont vraisemblablement plus utile que lui. En effet, même en ayant été exposés à un didacticiel, les participants les ont utilisées autant que ceux qui ne l'avaient pas vu avant l'épreuve (exception faite des lettres faciles, pour lesquelles l'apport du didacticiel a suffit à la réussite de la tâche).

En fait, l'intérêt de la présence d'un didacticiel se situe, en dépit de ses faibles effets sur les performances, sur la satisfaction et l'utilisabilité perçue. Même si nos résultats montrent que de simples aides en lignes suffisent à assurer la majeure partie des besoins des utilisateurs en termes d'apport de connaissances, la consultation préalable d'un didacticiel a significativement influencé la perception subjective de l'interface DIGIME. Ce résultat n'est pas anodin si nous le mettons en relation avec le fait que seuls les participants en ayant bénéficié ont utilisé la procédure d'insertion de caractères. Cette procédure, bien que présentée dans l'aide pour les gestes, fait cependant appel à un niveau de stratégie dans la pratique de l'interface assez élaboré, puisqu'elle limite l'effacement de caractères déjà reconnus et exacts, pour pouvoir placer, là où il le faut, une lettre eronnée dans une phrase. Le fait que cette procédure ait été utilisée seulement pour les phrases peut vouloir dire que son usage n'a d'intérêt qu'à partir d'une certaine quantité de lettres dont on juge que l'effacement intempestif serait préjudiciable à la performance. Cependant, si tel était le cas, l'effet se serait fait sentir aussi sur le nombre de gestes de corrections (que l'on a assimilé au nombre d'erreurs) pendant l'écriture des phrases. Or, ce n'est pas le cas. Cela témoignerait plutôt que la saisie d'une séquence de lettres assez longue (telle qu'une phrase, par opposition aux mots de cinq lettres utilisés pour cette épreuve) peut engendrer, peut-être par surcharge mentale, des oublis dans le contrôle des erreurs produites, erreurs qui ne seraient détectées qu'à l'occasion de contrôles plus espacés dans le temps. La satisfaction apportée par la résolution rapide de ces fautes détectées à l'occasion de contrôle « inopinés » joue peut-être un rôle. Quoi qu'il en soit, nous n'avons aucun moyen d'en avoir la certitude.

En conclusion, on retiendra tout de même, que, comme le montre la figure 38, même avec un didacticiel, l'utilisabilité perçue dépasse à peine les 40%. Certes, si les participants n'avaient pas la possibilité de comparer diverses interfaces entre elles, l'utilité du diaporama

était justement de pouvoir les affranchir de ces lacunes, afin qu'ils puissent affiner leur jugement. 40%, c'est peu, de toute évidence. Néanmoins, le peu de gestes d'effacement, y compris en début d'épreuve (une moyenne d'à peu près 12 corrections pour la saisie de 12 lettres, contre 50 erreurs pour 35 lettres en début d'expérience 1) montrent que les modifications que nos collègues de l'IRISA ont apporté à DIGIME, associées à la présence d'aides en ligne, ont éliminé le principal handicap de la première version de DIGIME, à savoir, son caractère non intuitif. Il nous faudra déterminer si ces améliorations sont suffisantes pour véritablement améliorer l'acceptabilité de DIGIME.

6- Expérience 3 : Test utilisateur : Comparaison de DIGIME aux méthodes alternatives de saisie de texte sur PDA

Résumé – On compare DIGIME aux méthodes de saisie alternatives les plus courantes sur PDA (clavier virtuel, *Note* et *JOT*). Il apparaît que DIGIME reste en-deçà de ses concurrents, et ce, en dépit des améliorations qui lui ont été apportées. Elle ne rivalise pas avec *JOT*, qui mise également sur la reconnaissance de caractères. Une difficulté dans la réalisation de gestes d'édition en est vraisemblablement la cause. Malgré cela, elle n'est pas moins acceptable que les autres méthodes de saisie, exception faite du clavier virtuel. Bien que l'échantillonnage de cette étude soit criticable, de même que les tâches proposées, il semble que les participants rejettent l'écriture manuscrite sur PDA au profit du clavier virtuel, confirmant les résultats issus de la littérature.

6.1. Introduction

Nos précédentes études se sont focalisées sur l'utilisabilité de DIGIME. Nous avons travaillé, d'une part, sur l'ergonomie de l'interface, et d'autre part sur son apprenabilité. Nous souhaitons recentrer notre travail, toujours dans une perspective de psychologie ergonomique, sur l'acceptabilité de DIGIME. Nous l'avons déjà évoqué, RESIFCar a été embarqué sur des Smartphones commercialisés en Europe. Or, les échos de sa diffusion commerciale ont montré un problème d'acceptabilité, à savoir que les utilisateurs des téléphones proposant RESIFCar (Anquetil & Bouchereau, 2002) n'en ont pas fait usage. La question de l'utilité de ce logiciel de reconnaissance de lettres manuscrites n'est pas à mettre en doute puisqu'il est nécessaire de pouvoir entrer du texte sur ce type de machine. Cela sous-entend que les possesseurs de ces téléphones lui ont préféré d'autres méthodes de saisie. Il en existe beaucoup, et elles proposent des principes d'interaction très variés. Parmi elles, le clavier virtuel suscite une large adhésion par rapport aux logiciels de reconnaissance d'écriture manuscrite (Bouteruche, Deconde, Anquetil, & Jamet, 2005; Commaford, 2004; Curran, Woods, & O Riordan, 2006; Hinckeley, 2003; Jansen & Ayers, 2004; Mackenzie & Chang 1999; MacKenzie, Nonnecke, McQueen, Riddersma & Meltz, 1994; MacKenzie & Soukoreff, 2002; MacKenzie & Zhang, 1997; McQueen, MacKenzie, Nonnecke, Riddersma, & Meltz, 1994; Ortelli, 2003; Sears & Arora, 2002).

Nous désirons conduire une étude visant à comparer DIGIME aux principales autres méthodes de saisie de texte sur PDA. Il nous sera ainsi possible d'évaluer en quoi les transformations apportées à DIGIME ont permis d'améliorer son acceptabilité par rapport aux solutions concurrentes (Bevan, 2006; Keinonen, 1998; Stewart & Travis, 2003). Pour ce faire, nous souhaitons utiliser une méthode proche du test d'utilisabilité (Dumas, 2003), bien qu'il s'agisse, de par la comparaison d'interface entre elles, d'un test d'hypothèse. Bien sûr, l'inspiration méthodologique de cette étude pourra donner lieu à des observations qualitatives issues de l'inspection ergonomique (Bastien & Scapin, 2001), mais elle se veut avant tout un moyen d'évaluer objectivement DIGME par rapport aux autres méthodes de saisie (Gray & Salzman, 1998).

Faire un test d'utilisabilité n'est possible que si le prototype que l'on veut tester présente un niveau de finition suffisamment abouti pour que la tâche puisse être aussi réaliste que possible. En ce qui nous concerne, nos collaborateurs de l'IRISA ont développé et mis à notre disposition DIGIME de telle façon que l'interface propose toutes les fonctionnalités

nécessaires à un usage en conditions réelles. L'interface graphique y est parfaitement soignée, comme lors de nos précédentes manipulations. C'est donc un prototype complet, d'autant qu'il est installé sur des PDA eux aussi éprouvés (Ham et al., 2006; Nielsen, 1993; Vanderdonckt & Coyette, 2006).

La méthode du test d'utilisabilité impose à ses concepteurs un travail de sélection des tâches et des utilisateurs, dont l'objectif est de donner au test la plus grande validité écologique en fonction du contexte d'usage supposé de l'interface (Bevan, 1995; Bevan & Macleod, 1994; Dumas, 2003; Nielsen, 1993). Dans le cadre de cette étude, cela sous-entend de réunir un panel de personnes susceptibles de faire l'achat d'un PDA ou d'un Smartphone, sans pour autant en posséder un, puisqu'il nous importe qu'elles soient novices dans leur manipulation. C'est bien évidemment difficilement envisageable pour nous, dont les échantillons de participants sont issus d'une population principalement jeune, étudiante, et, disons-le, désargentée. Or, ces dispositifs restent relativement coûteux, et, quand bien même tendent-ils à se généraliser (par la *Nintendo DS*, notamment), leur usage reste encore confidentiel. Ceci n'est pas crucial cependant, puisque l'objectif initial de cette étude reste la comparaison ergonomique de différentes méthodes de saisie sur PDA. Dès lors, si les échantillons composant les groupes des conditions expérimentales sont issus de la même population, le problème est moindre. Par contre, il est important que le choix des tâches à réaliser soit guidé par le souci du réalisme. Comme le fait remarquer Keinonen (1998), si l'on exclue l'utilité de l'utilisabilité, cela sous-entend que la tâche à accomplir rentre dans les attributions du dispositif. Par conséquent, il faut choisir avec soin ce que devront faire les participants au test. Cela a d'autant plus d'importance que, en dépit du lieu de test, limité au laboratoire en ce qui nous concerne, les contraintes que les tâches imposent aux utilisateurs doivent être les plus proches possible de ce pour quoi le dispositif est conçu (Folmer & Bosch, 2003; Kaikkonen, Kallio, Kekäläinen, Kankainen, & Cankar, 2005).

Dans le cas des PDA et autres Smartphone*s*, leur utilité réside principalement dans leurs fonctions de communication (téléphonie, navigation sur le *Web*, messagerie électronique, SMS). Ils peuvent également, en tant que mémoires portables, servir à stocker toute information dont l'utilisateur estime qu'il aura un besoin ultérieur (Benton, Kiewra, Whitfill, & Dennison, 1993). Dans les deux cas, ces fonctions nécessitent de pouvoir saisir du texte. Plus précisément, elles nécessitent de prendre des notes, ou, pour le moins, de produire des écrits dont la forme finale s'apparente à de la prise de notes. Ceci passe notamment par

l'utilisation massive de procédés abréviatifs (Piolat, 2004b). Il y a donc fort à parier que la production de texte sur PDA et Smartphones implique principalement de la saisie de rendez-vous, de numéros de téléphones, d'adresses, d'URL, *etc*. en plus des SMS (Curran, Woods, & O Riordan, 2006; Költringer & Grechenig, 2004), dont la saisie partage de nombreux points communs avec la PdN. Comme elle, leur rédaction fait usage d'abréviations liées à des contraintes temporelles (Hartley, 2002; Piolat, 2006). Cependant, si les logiciels de saisie de texte sur PDA vont servir à des activités rédactionnelles assimilables à de la PdN, ils ne sont pas fondamentalement fait pour ça et n'incluent pas la possibilité d'insérer des signes sémiographiques ou substitutifs pourtant courants dans la PdN (Piolat, 2006). Ces procédés, bien que coûteux en ressources mentales, puisque très contrôlés, sont toutefois ceux qui vont permettre au scripteur de gagner du temps (Piolat, 2004b; 2006). On peut donc estimer objectivement que le meilleur logiciel de saisie de texte sur PDA est celui qui permet d'écrire le plus de choses en un minimum de temps, indépendamment des efforts qu'il requièrt, qui de toute façon figurent parmi les plus élevés (Kobayashi, 2005; Peverly et al., 2007; Piolat, Olive, et Kellogg, 2005). Cette dernière affirmation relève cependant du point de vue de la communauté HCI. L'autre point de vue, celui de la communauté MIS mérite aussi toute notre attention (Zhang & Dillon, 2003).

Bien que nous ayons toutes les raisons de penser que l'amélioration de l'acceptabilité de DIGIME soit adossée à celle de son utilisabilité (Arning & Ziefle, 2007; Hasan & Ahmed, 2007; ISO-9241-11, 1994; Keinonen, 1998; Nielsen, 1993; Rogers, 1995; Shackel, 1991; Tornatzky & Klein, 1982; Tricot & Tricot, 2000; Venkatesh & Davis, 1996), seul compte au final l'avis des utilisateurs sur la question. Plusieurs outils existent pour le mesurer de façon fiable (Ajzen, 1985; Davis, Bagozzi, & Warshaw, 1989; Dillon & Morris, 1997; Fishbein & Ajzen, 1975; Venkatesh & Davis, 2000; Venkatesh, Morris, Davis, & Davis, 2003). Parmi ces différents questionnaires qui permettent d'évaluer et prédire l'adoption future d'une technologie, le modèle TAM est sans conteste le plus adapté à notre problématique (Davis, 1989; Davis, Bagozzi, & Warshaw, 1989). Beaucoup plus spécifique que la TRA (Fishbein & Ajzen, 1975) ou le TPB (Ajzen, 1985), il est reconnu pour donner de meilleurs résultats dans le cadre du test des nouvelles technologies (Adams, Nelson, & Todd, 1992; Davis, Bagozzi, & Warshaw, 1989; Dillon & Watson, 1996) et est, de par sa construction, plus simple à mettre en œuvre (Davis, Bagozzi, & Warshaw, 1989). Sa construction, en termes de validité, fiabilité et fidélité a fait l'objet d'études, et ne fait aucun doute (Adams, Nelson, & Todd, 1992; Davis,

1989; Davis & Venkatesh, 1996). Même si plusieurs améliorations ont été apportées au TAM depuis sa création, augmentant considérablement sont potentiel prédictif (Venkatesh & Davis, 2000; Venkatesh, Morris, Davis, & Davis, 2003), nous choisissons néanmoins d'utiliser ce questionnaire dans sa première version, comme l'ont fait Dillon et Morris (1997). Nous faisons ce choix au motif que, dans le contexte d'un test de laboratoire dont les déterminants principaux relèvent de l'utilisabilité des interfaces, le TAM d'origine est suffisant et adapté avec ses mesures d'utilité et de facilités d'utilisation perçues, associées aux attitudes et à l'intention d'usage. Les autres modèles tels que le TAM 2 (Venkatesh & Davis, 2000) ou UTAUT (Venkatesh, Morris, Davis & Davis, 2003), s'ils sont plus puissants, accordent plus d'importance aux déterminants sociaux et psychosociaux de l'intention d'usage. Or, ces questions ne nous concernent pas vraiment, c'est pourquoi nous préférons l'économie qu'offre un questionnaire relativement court, par rapport à d'autres, beaucoup plus longs. Pour autant, et comme l'ont fait Dillon et Morris (1997), nous prenons la précaution de mesurer la norme subjective.

6.2. Méthode

6.2.1. Participants

Principalement étudiants à Rennes 2, 47 individus tout-venant ont participé à cette expérience. Parmi eux 10 garçons et 37 filles, dont la moyenne d'âge était de 22,2 ans. Tous ces participants étaient novices dans l'utilisation des PDA. Leur participation à l'expérience était volontaire et n'a donné lieu à aucune rémunération ou récompense.

6.2.2. Matériel

6.2.2.1. Matériel informatique

Nous avons utilisé deux Pocket PC E400 Toshiba avec stylets (*Intel PXA 261* cadencé à 300 MHz, 31,22 Mo de mémoire vive, environnement *Windows Mobile*) ainsi que quatre ordinateurs portables (Respectivement Acer Aspire *Intel Pentium III* cadencé à 1 GHz, 256 Mo de mémoire vive ; Dell Inspiron 6000 *Intel Mobile* cadencé à 1,73 GHz, 502 Mo de mémoire vive ; ainsi que deux Toshiba Satellite 2800 *Intel Pentium III* cadencé à 1GHz, 256

Mo de mémoire vive ; tous étant sous environnement *Windows XP*). Durant la passation de l'expérience, les PDA étaient reliés chacun avec un ordinateur distant pour le suivi en direct, par un expérimentateur, de l'activité des participants. Le but était de s'assurer du bon déroulement de la passation sans les gêner. La connexion entre les PDA et les PC était assurée par les logiciels *Microsoft ActiveSync* 3.7 et *Remote Display Control* 2.03. Deux autres ordinateurs servaient à la présentation du commentaire oral à prendre en note par les participants. Ils permettaient aussi l'enregistrement des durées de consultation des séquences de celui-ci. C'est pourquoi nous avons utilisé des casques audio de qualité professionnelle.

6.2.2.2. Méthodes de saisie

La vocation de cette expérience était de comparer DIGIME à quelques unes des alternatives existant sur le marché des PDA. Nous avons choisi de comparer notre interface au clavier virtuel, à *Note*, ainsi qu'à un logiciel de reconnaissance de caractères (fondé sur la technologie *JOT*).

6.2.2.2.1. DIGIME

Les deux premières expériences ont permis d'apporter beaucoup d'améliorations à DIGIME. Malgré cela, ses principes de fonctionnement étaient identiques puisque le moteur de la reconnaissance était le même (hormis la légère amélioration du taux d'identification évoquée pour la deuxième expérience).

La principale modification a consisté en l'ajout d'un didacticiel complet, accessible d'un clic sur le point d'interrogation bleu en haut à droite de la zone de saisie. Il lançait une application supplémentaire et donnait accès à un menu hypertexte. Ce menu proposait une vidéo de démonstration de quelques minutes. Elle présentait rapidement les différentes fonctions de DIGIME, exemples à l'appui. L'utilisateur avait tout contrôle sur cette animation : avance et retour rapides, pause, lecture. Le reste du menu proposait une présentation thématique de la saisie des lettres, accents, ponctuation, gestes d'édition, chiffres et caractères spéciaux (un lien par thème). La navigation dans le didacticiel se faisait par trois boutons « précédent », « continuer » et « menu », pour revenir à l'accueil. Il suffisait aux utilisateurs de cliquer sur un pictogramme « ok » pour fermer le didacticiel et retourner à

l'application DIGIME. Ce didacticiel était complété par deux aides en lignes, l'une pour les lettres, l'autre pour les gestes d'édition et la ponctuation. Elles étaient identiques à celles que nous avons utilisé pour l'expérience 2 et étaient accessibles directement via l'interface DIGIME.

Aux deux boutons situés sur la gauche de la zone de saisie, et permettant d'ouvrir les aides en ligne, les développeurs de l'IRISA ont ajouté un troisième bouton sur lequel un clic ouvrait un petit clavier virtuel permettant, par simple contact du stylet sur les symboles qu'ils représentaient, d'insérer des caractères spéciaux non reconnus par le logiciel (ex : ? ! / €). De la même façon, un quatrième bouton (situé à l'extrême droite de DIGIME) permettait à l'utilisateur de passer du mode de saisie de lettres à celui de saisie de chiffres. Ces modifications ont permis de faire de DIGIME une application interactive complète que l'on a pu tester en conditions réelles, sans restriction. Pour envoyer le texte écrit dans la zone application en haut de l'écran, les sujets devaient faire le geste « entrée », correspondant à un trait horizontal décrivant un aller-retour (comme le retour chariot d'imprimerie).

Figure 39: Interface DIGIME complète (à gauche), avec menu hypertexte du didacticiel (au centre) et exemple de planche du didacticiel (à droite).

6.2.2.2.2. Clavier virtuel

Le clavier virtuel que nous avons utilisé est le pendant logiciel des claviers AZERTY que nous trouvons sur la plupart des ordinateurs, qu'ils soient ou non portables. La saisie s'y faisait par contact du stylet sur le caractère que les participants voulaient insérer. Son

fonctionnement a été entièrement calqué sur celui des claviers matériels, notamment pour l'accès aux majuscules et aux caractères de la « rangée du haut » (touche Maj.). Il permettait également d'ouvrir un pavé numérique logiciel si le besoin s'en faisait sentir. Aucun didacticiel ne lui était associé.

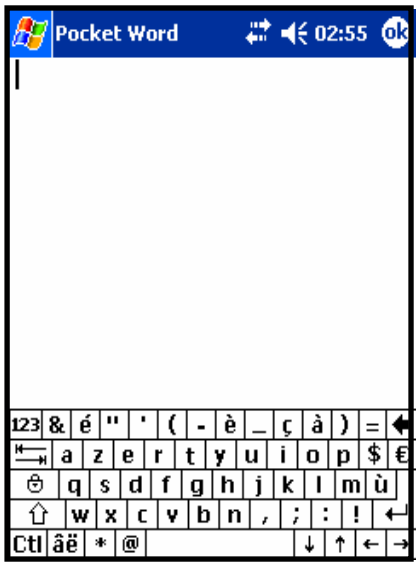


Figure 40: Interface de saisie clavier virtuel

6.2.2.2.3. Note

L'application *Note* pour Pocket PC permettait d'écrire sur l'écran tactile comme sur une feuille de papier. Elle n'impliquait aucune reconnaissance. Par conséquent, un utilisateur pouvait s'en servir pour faire un dessin, un schéma, ou écrire des signes sémio-graphiques. La « feuille » électronique pouvait s'étirer autant que le souhaitait le scripteur (dans les limites bien sûr, de la mémoire vive disponible, mais non contraignante par rapport aux besoins de l'expérience). Cette feuille comportait des lignes. Les participants étaient libres d'écrire dessus ou non.

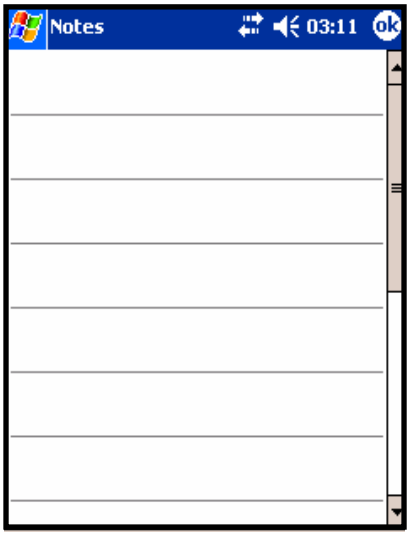


Figure 41: Interface d'écriture *Note*

6.2.2.2.4. Reconnaissance de lettres (JOT)

Cette interface de saisie et de reconnaissance de lettres manuscrites a été conçue, à l'instar de DIGIME, pour transformer, après identification, des lettres manuscrites en caractères d'imprimerie. On y écrit des lettres minuscules, détachées les unes des autres. Comme pour DIGIME, ce logiciel pouvait reconnaître plusieurs allographes pour une même lettre. L'endroit où l'on y formait les caractères avait son importance : à gauche, pour les majuscules (bien qu'elles devaient y être écrites en minuscule), au centre pour les minuscules, à droite pour les chiffres. La ponctuation pouvait être saisie n'importe où dans la zone d'écriture. Il fallait écrire sur la ligne horizontale continue, comme le montre l'exemple pour la lettre « o ». En-dessous de cette ligne, on faisait dépasser les queues des « p », ou les boucles des « j » par exemple. Au-dessus de la ligne pointillée, on faisait dépasser les boucles des « b » ou « l », et on plaçait les accents. Les lettres reconnues s'affichaient immédiatement dans la zone traitement de texte, en haut de l'écran. Les corrections étaient faites grâce à la flèche noire épaisse orientée vers la gauche. Les deux petites flèches situées en-dessous avaient pour fonction de déplacer le curseur parmi les lettres déjà reconnues. Il s'agit de boutons sur lesquels il fallait appuyer, le cas échéant. La touche « entrée » (flèche coudée) et la touche « espace », fonctionnaient de la même manière. *JOT* proposait, comme DIGIME, un didacticiel chapitré par un menu. Il était composé d'animations faisant la démonstration de la façon dont saisir les caractères. Un dernier bouton, en bas à droite, activait un clavier virtuel pour l'insertion de symboles et autres caractères spéciaux (comme DIGIME).

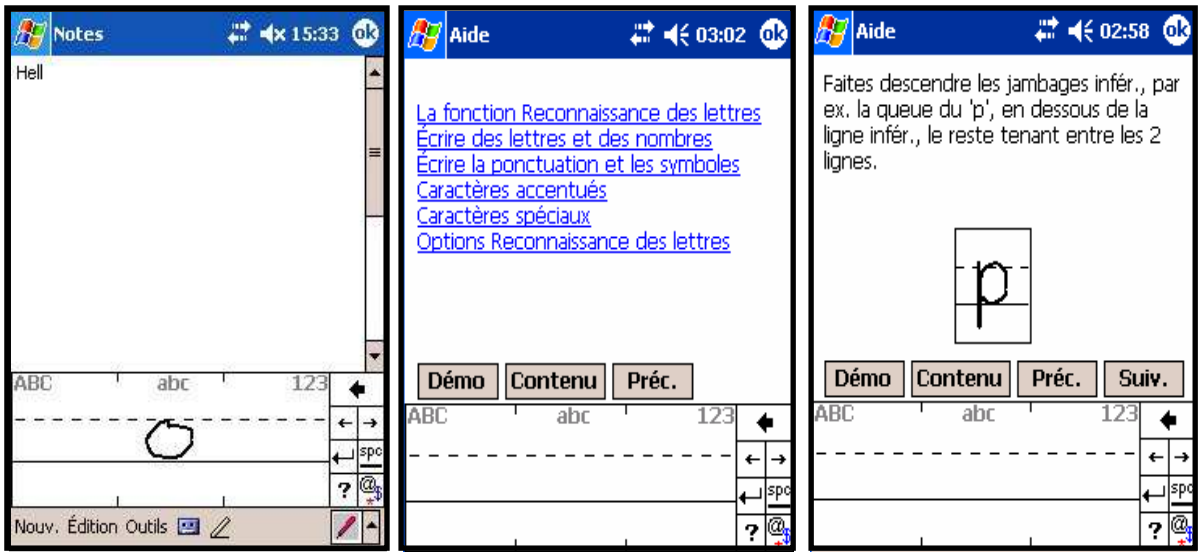


Figure 42: Interface de saisie et de reconnaissance de caractères manuscrits *JOT* (à gauche), avec menu hypertexte du didacticiel (au centre) et exemple de planche du didacticiel (à droite).

6.2.2.3. Matériel linguistique

Sans une phase préalable d'entraînement et de découverte, il était impensable de faire faire à des sujets novices dans la manipulation d'un PDA, une tâche se voulant réaliste. Pour s'assurer que les participants avaient compris les principes de fonctionnement essentiels de l'interface à laquelle ils étaient confrontés, ils devaient reproduire à l'identique cette planche (figure 43) sur l'écran de leur PDA. Cela incluait autant l'orthographe que l'ordre de présentation des caractères, ainsi que les retours à la ligne.

0123456789
aAbBcCdDeEfFgGhHiIjJkKlLmMnNoOpPq
QrRsStTuUvVwWxXyYzZ
é è ë î
Zéro, un, deux, trois, quatre, cinq, six, sept, huit, neuf.

Figure 43: Planche de caractères à reproduire en phase d'entraînement (taille réelle)

Une fois accompli cet entraînement, l'épreuve proprement dite consistait à prendre en note le texte suivant :

Marie est disponible : - lundi 14 octobre à 8h,

- jeudi 17 octobre à 14h,
- vendredi 18 octobre à 8h ;

Benoît est disponible : - mardi 15 octobre à 9h,
- mercredi 16 octobre à 14h ;

Noël est disponible toute la semaine ;

François est disponible : - mardi 15 octobre à 9h,
- jeudi 17 octobre à 14h ;

Lucie est disponible : - mardi 15 octobre à 9h,
- jeudi 17 octobre à 14h ;

Cédric est disponible : - jeudi 17 octobre à 14h,
- vendredi 18 octobre à 11h ;

Marlène est disponible : - jeudi 17 octobre à 16h,
- vendredi 18 octobre à 8h,
- vendredi 18 octobre à 14h.

Guy est disponible : - mardi 15 octobre à 9h,
- jeudi 17 octobre à 14h.

Ce texte a été enregistré oralement par l'expérimentateur. Il a ensuite été découpé en séquences. Une séquence est constituée d'un prénom suivi de « est disponible » ou bien d'une date complète. Chacune de ces séquences a été enregistrée dans une application Powerpoint interactive de telle façon qu'en appuyant sur la touche « espace » du clavier, la lecture de la séquence se lançait en déclenchant un chronomètre (précision à la seconde). En appuyant à nouveau sur « espace », les participants lançaient la séquence suivante. Cela enregistrait la durée écoulée depuis le précédent appui tout en réinitialisant le chronomètre, et ainsi de suite, jusqu'à ce que les participants aient entendu « l'expérience est terminée ». Bien qu'ayant utilisé Powerpoint de façon détournée pour les besoins de notre expérience, aucune image

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

n'était affichée à l'écran des ordinateurs diffusant le commentaire audio. L'écran était blanc. Le choix des prénoms était principalement justifié par la présence de tous les signes diacritiques possibles de la langue française.

6.2.2.4. Questionnaires

Comme pour chacune de nos précédentes expériences, nous avons mis à contribution le même pré-questionnaire afin de s'assurer de l'inexpérience des sujets en matière de PDA.

Pour mesurer de la façon la plus fiable possible l'acceptabilité de ces interfaces de saisie, nous avons utilisé le TAM (Davis, 1989; Davis, Bagozzi, & Warshaw, 1989). Nous avons choisi de traduire la version du questionnaire utilisée par Morris et Dillon en 1997. Les auteurs fournissent les données suivantes, issues de la littérature (Davis, Bagozzi, & Warshaw, 1989), sur la fiabilité des items qu'ils utilisent pour chaque dimension :

- l'utilité perçue (quatre items, ex : Utiliser ce logiciel de saisie de texte pourrait améliorer mes performances au quotidien) ($\alpha = .92$);
- la facilité d'utilisation perçue (sept items, ex : Je trouve facile de faire au logiciel ce que je veux qu'il fasse) ($\alpha = .90$);
- attitude quant à l'usage (quatre items, ex : Utiliser ce logiciel de saisie de texte est une …….. idée) ($\alpha = .85$);
- intention d'usage (deux items, ex : J'ai l'intention d'utiliser ce logiciel de saisie à l'avenir) ($\alpha = .91$);
- norme subjective (deux items, ex : Les gens qui sont importants pour moi pensent que je devrais utiliser ce logiciel de saisie) (fiabilité non rapportée).

Les réponses aux questions se font sur des échelles de type Lykert à sept graduations, qui s'étendaient d'un pôle très favorable (ou tout-à-fait d'accord, ou très positif) à un pôle très défavorable (ou pas du tout d'accord, ou très négatif).

Nous avons complété ce questionnaire par le test d'utilisabilité de Nielsen (1993). Pour ce test, les sujets répondaient sur des échelles graduées de -2 (très négatif) à +2 (très positif). L'intérêt de ces dernières questions était de cerner au mieux les différences qualitatives ressenties par les participants quant aux possibilités de saisie des différentes interfaces.

6.2.3. Procédure expérimentale

Les participants commençaient par remplir le pré-questionnaire. Ils étaient ensuite affectés aléatoirement à l'une ou l'autre des interfaces de saisie testées. L'épreuve commençait par la phase d'entraînement. Les participants avaient devant eux une petite carte plastifiée comme le montre la figure 43. La consigne était de reproduire le contenu de cette carte à l'identique sur l'écran du PDA grâce à la méthode de saisie qui leur était affectée. Aucune indication sur le fonctionnement de l'interface ne leur était donnée. Par contre, ils ont été autorisés à rechercher des informations sur le fonctionnement de l'interface comme bon leur semblait. Ils pouvaient donc naviguer librement dans le contenu du Pocket PC et tout essayer pour réaliser la tâche dans les meilleurs délais (selon la consigne). Une fois cette première tâche terminée (saisie complète de la planche 43 en respectant la disposition originale), débutait alors l'épreuve de prise de notes. La consigne enjoignait les participants à prendre en notes les disponibilités de plusieurs personnes fictives en vue de fixer une date de réunion à laquelle un maximum d'entre elles pourrait venir. Ils pouvaient prendre leurs notes comme ils le désiraient, sans aucune autre restriction que d'orthographier correctement les noms propres (avec majuscule en initiale) et d'être capables de se relire pour pouvoir organiser ce rendez-vous. Ceci a fait l'objet d'une vérification en fin d'épreuve. En dehors de ces deux impératifs, ils pouvaient employer les abréviations qu'ils voulaient, laisser des fautes d'orthographe qu'ils ne souhaitaient pas corriger, *etc.* Là encore, ils pouvaient, s'ils en ressentaient le besoin, chercher de l'aide dans les ressources qu'offraient le PDA et le logiciel qu'ils utilisaient. L'accent a été mis, dans la consigne, sur la rapidité de leur prise de notes. Le commentaire oral était distillé par de courtes séquences audio lorsqu'ils appuyaient sur la touche « espace » de l'ordinateur. Des casques audio reliés aux ordinateurs fournissaient le son. A l'issue de cela, ils ont répondu aux questions du TAM et du test d'utilisabilité de Nielsen. L'expérience a duré entre 30 et 90 minutes, en fonction des conditions expérimentales.

6.2.4. Variables indépendantes

Nous avons manipulé une VI intergroupe à quatre modalités. Il s'agit du type d'interface suivant que les participants réalisaient l'expérience avec DIGIME, *Note*, le clavier virtuel ou *JOT*.

6.2.5. Variables dépendantes

Nous avons enregistré les temps de saisie du texte d'entraînement, ainsi que les temps de prise de notes des disponibilités. Ces données ont été complétées par les réponses aux différents items composant les questionnaires (TAM et test de Nielsen).

6.3. Résultats

6.3.1. Effets du type d'interface sur les temps de saisie à l'entraînement et à la prise de notes

Les données recueillies sur les temps de saisies aux deux tâches (entraînement et prise de notes) ne souscrivaient pas aux conditions de distribution et d'homogénéité des variances, nous les avons donc traitées avec des tests non paramétriques : Kruskal-Wallis pour les comparaisons d'ensemble et Mann-Whitney pour les comparaisons deux à deux. Nous avons pris la précaution, pour les comparaisons multiples, de corriger $\alpha_{\text{par comparaison}}$ de façon à ce que $\alpha = 1 - (1 - \alpha_{\text{par comparaison}})^{\text{nombre de comparaisons}}$, dans le but de réduire la probabilité de faire une erreur de type I (Onwuegbuzie & Daniel, 2003). Puisque nous nous sommes limités à comparer DIGME aux autres interfaces, $\alpha_{\text{par comparaison}} = .02$.

Les résultats sont résumés par la figure 44 (page suivante). Sans surprise, l'interface a eu un effet aussi bien pour les temps de réalisation de l'entraînement (χ^2 (3, N = 43) = 37,425, p = .00) que pour ceux de la prise de notes (χ^2 (3, N = 43) = 36,770, p = .00).

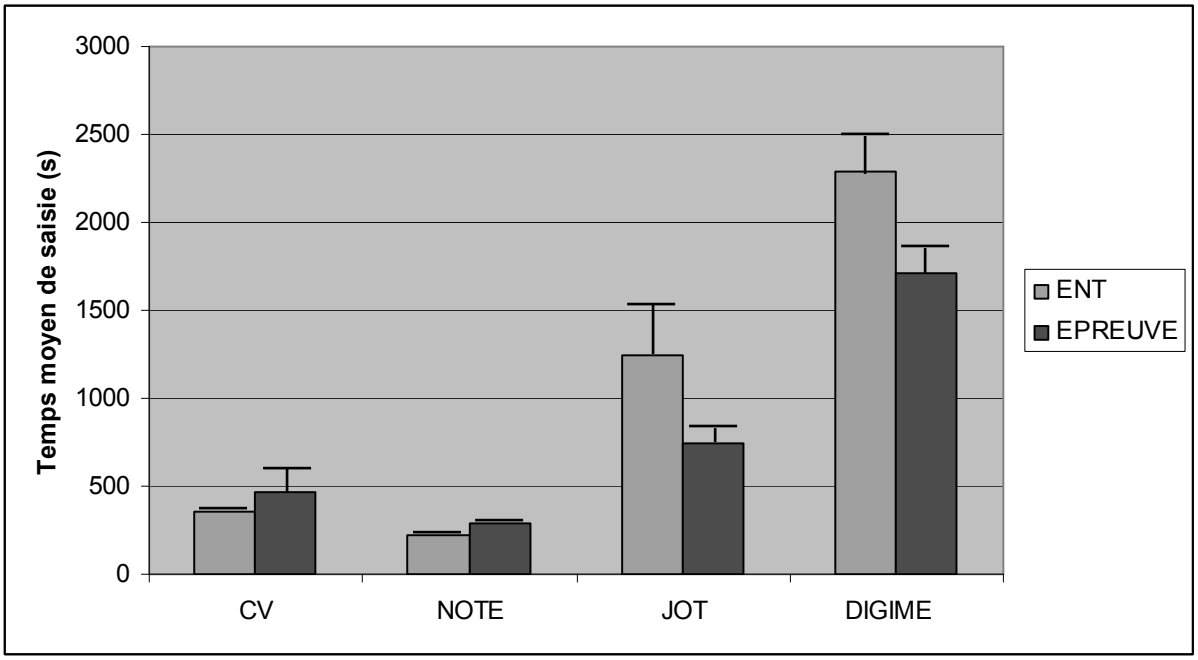


Figure 44:Temps moyens de saisie (s) aux tâches d'entraînement et de prises de notes

Cependant, il était plus intéressant pour nous de comparer DIGIME par rapport aux autres méthodes de saisie, et principalement à *JOT*, qui fonctionnait sur le même principe. La comparaison entre DIGIME et le clavier virtuel était également légitime du fait qu'il représentait l'alternative la plus couramment choisie par rapport aux logiciels de reconnaissance d'écriture manuscrite. Pour finir, comparer DIGIME et *Note* a permis d'évaluer la différence entre deux situations d'écriture manuscrite sur PDA dont, l'une, n'impliquant pas de reconnaissance, pouvait être considérée comme une situation d'écriture manuscrite « idéale ».

Le test de Mann-Whitney a confirmé que les performances entre *JOT* et DIGIME étaient différentes pour les deux tâches (U=10,00 ; Z= -3,304, *p* =.00 pour l'entraînement et U=.00 ; Z=-3,848, *p* =.00 pour la prise de notes). Idem pour la comparaison entre DIGIME et le clavier virtuel (U=.00 ; Z=-4,033, *p* =.00 pour l'entraînement et U=.00 ; Z=-3,909, *p* =.00 pour la PdN) ainsi que *Note* (U=.00 ; Z= -3,582, *p* =.00 pour l'entraînement et U=.00 ; Z= -3,677, *p* =.00 pour la PdN), qui se sont démarqués de DIGIME pour les deux phases de l'épreuve. A noter également que la comparaison entre le clavier virtuel et *Note* a montré des temps de saisie statistiquement différents (U=13,00 ; Z= -3,038, *p* =.00 pour l'entraînement et U=9,00 ; Z= -3,306, *p* =.00 pour la PdN).

Fait important, *JOT* et DIGIME sont les deux seules interfaces pour lesquelles l'entraînement a duré plus longtemps que la prise de notes.

6.3.2. Effets du type d'interface sur les jugements

6.3.2.1. Test TAM

L'utilisation du TAM dans une le cadre d'une expérience doit normalement donner lieu à une analyse factorielle confirmatoire et une analyse de fiabilité. Ces précautions permettent de s'assurer de la validité et de la fiabilité du questionnaire dans un contexte nouveau (Onwuegbuzie & Daniel, 2003). La petitesse de notre échantillon n'était pas compatible avec ce type d'analyse. Cependant, depuis près de 20 ans, le TAM a montré une robustesse qui n'a pas été démentie (King & He, 2006). Afin de simplifier notre analyse et éviter de multiplier les tests, nous avons choisi de calculer la médiane des réponses aux questions relevant de chacune des dimensions identifiées dans le TAM. Ce sont ces nouvelles variables que nous avons traitées. Ces résultats sont consignés dans le tableau 3. Les valeurs médianes sont données suivies de l'intervalle interquartile entre parenthèses.

Dimensions	Clavier virtuel	*Note*	*JOT*	DIGIME
Facilité d'utilisation perçue	2,0 (.50)	2,0 (1,0)	3,0 (1,0)	5,0 (1,0)
Utilité perçue	3,0 (1,75)	4,5 (3,0)	3,25 (4,0)	6,0 (2,75)
Attitude	3,0 (1,5)	4,0 (1,75)	3,5 (3,00)	4,75 (1,38)
Intention d'utilisation	4,0 (3,0)	6,0 (2,0)	6,0 (3,25)	6,5 (1,0)
Norme subjective	4,0 (1,0)	4,0 (1,0)	4,0 (2,50)	5,2 (2,25)

Note : Médiane (intervalle interquartile). Les échelles de réponses sont orientées entre un pôle très positif à 1 et un pôle très négatif à 7.

Tableau 3: Effets du type d'interface sur les dimensions du TAM

Nous avons traité ces résultats avec des tests non paramétriques (Kruskal-Wallis et Mann-Whitney), et ce, indépendamment de l'examen des distributions des données. De par leur nature même, elles ne peuvent être représentées que par leurs médianes et non leurs moyennes. Ce type de tests, en plus d'être adaptés au traitement des médianes, s'y révèlent aussi particulièrement sensible, ce qui justifie notre choix (Howell, 1997, p.726).

Il y a bien eu un effet du type d'interface sur la facilité d'utilisation perçue (χ^2 (3, N = 43) = 21,783, p = .00), de même que pour l'utilité perçue (χ^2 (3, N = 43) = 11,379, p = .00), l'attitude vis-à-vis du logiciel (χ^2 (3, N = 43) = 16,467, p = .01) et l'intention d'utilisation (χ^2 (3, N = 43) = 11,017, p = .01). Par contre, la norme subjective n'a pas été affectée par le type d'interface (χ^2 (3, N = 43) = 3,029, p = .38).

Si l'on examine plus attentivement les différences entre DIGIME et ses concurrents, on se rend compte qu'il n'a suscité des opinions différentes de *JOT* que pour la facilité d'utilisation perçue (5,0 (1,0) pour DIGIME contre 2,0 (.50) pour *JOT*, U=14,00, Z=-3,485, p =.00). A l'inverse, les deux systèmes ont été jugés de la même façon pour l'utilité (6,0 (2,75) pour DIGIME contre 3,0 (4,0) pour *JOT*, U=40,00, Z=-1,902, p =.06). Aucune différence n'a été révélée pour les attitudes (4,75 (1,38) pour DIGIME contre 3,5 (3,13) pour *JOT*, U=53,00, Z=-1,109, p =.29) ou l'intention d'usage (6,5 (1,0) pour DIGIME contre 6,0 (3,25) pour JOT, U=48,00, Z=-1,444, p =.17).

En ce qui concerne *Note*, nous avons constaté des différences avec DIGIME. Sur la facilité perçue, bien sûr (5,0 (1,0) pour DIGIME contre 2,0 (1,0) pour *Note*, U=9,5 ; Z=-3,237, p =.00), mais aussi pour les attitudes (4,75 (1,38) pour DIGIME contre 4,0 (1,75) pour *Note*, U=12,00 ; Z=-3,096, p =.00). Par contre, les analyses n'ont révélé aucune différence pour l'utilité (6,0 (2,75) pour DIGIME contre 4,5 (3,0) pour *Note*, U=32,5, Z=-1,559, p =.12) et l'intention d'utilisation (U=26,00 ; Z=-1,908, p =.06)).

Pour finir, penchons-nous sur le clavier virtuel, dont l'utilisation par nos participants a suscité chez eux des avis très contrastés par rapport à ceux qu'ont porté les utilisateurs de DIGIME. En effet, le clavier virtuel est jugé plus facile (5,0 (1,0) pour DIGIME contre 2,0 (.50) pour le clavier virtuel, U=10,5 ; Z=-3,829, p =.00), plus utile (6,0 (2,75) pour DIGIME contre 3,0 (1,75) pour le clavier virtuel, U=15,5 ; Z=-3,437, p =.00), de telle sorte que cela se répercute sur les attitudes (4,75 (1,38) pour DIGIME contre 3,0 (1,5) pour le clavier virtuel, U=7,5 ; Z=-3,881, p =.00) et, bien évidemment, sur l'intention d'usage (6,5 (1,0) pour DIGIME contre 4,0 (3,0) pour le clavier virtuel, U=21 ; Z=-3,141, p =.00).

6.3.2.2. Test d'utilisabilité de Nielsen

Les réponses médianes à ces items sont consignées dans le tableau 4. Le test de Kruskal-Wallis nous a appris que les réponses des participants se sont différenciées en fonction de l'interface de saisie uniquement sur certaines de nos VD. Cela s'est exprimé notamment sur la facilité d'apprentissage perçue (χ^2 (3, N = 43) = 14,945, p = .00), l'efficacité perçue (χ^2 (3, N = 43) = 13,200, p = .00), et la gestion des erreurs perçue (χ^2 (3, N = 43) = 9,625, p = .02). A l'inverse, les sujets ont répondu de la même façon quant à l'apprenabilité perçue (c'est-à-dire leur capacité à obtenir le même niveau de performance ultérieurement) (χ^2 (3, N = 43) = 5,485, p = .14). Pour finir, la satisfaction n'a pas été un critère de distinction entre les groupes (χ^2 (3, N = 43) = 5,542, p = .13).

	Clavier virtuel	*Note*	*JOT*	DIGIME
Facilité d'apprentissage perçue	5,0 (.50)	4,0 (1,0)	3,0 (2,5)	3,5 (3,0)
Efficacité perçue	4,0 (.50)	3,0 (1,5)	4,0 (2,25)	2,0 (1,0)
Apprenabilité perçue	5,0 (0)	5,0 (0)	5,0 (.25)	4,5 (1,0)
Gestion des erreurs	2,0 (2,5)	4,0 (3,0)	3,0 (1,5)	4,0 (0)
Satisfaction	4,0 (1,0)	3,0 (1,5)	3,0 (1,75)	2,5 (2,75)

Note : Médiane (intervalle interquartile). Les échelles de réponses sont orientées entre un pôle très négatif ou difficile à 0 et un pôle très positif ou facile à 5.

Tableau 4: Effets du type d'interface sur les items du test d'utilisabilité de Nielsen

L'examen particulier des différences entre DIGIME et les autres interfaces a précisé cela. Rappelons que pour ces comparaisons, la valeur critique de α = .02.

JOT, tout d'abord, ne s'est différencié de DIGIME ni sur la facilité d'apprentissage perçue (3,5 (3,0) pour DIGIME contre 3,0 (2,5) pour *JOT*, U=63,00, Z= -.532, p =.63), ni sur l'efficacité perçue (2,0 (1,0) pour DIGIME contre 4,0 (2,25) pour *JOT*, U=46,00, Z= -1,461, p =.19), alors que cela a été le cas pour la gestion des erreurs perçue, mieux jugée pour DIGIME (4,0 (.00) pour DIGIME contre 3,0 (1,5) pour *JOT*, U=27,00, Z= -2,844, p =.00).

La comparaison de *Note* à DIGIME n'a elle révélé aucune différence significative au test d'utilisabilité de Nielsen.

Par contre, les avis sont contrastés suivant qu'ils proviennent de participants ayant réalisé l'expérience avec le clavier virtuel ou avec DIGIME. C'est notamment le cas pour le sentiment d'efficacité (2,0 (1,0) pour DIGIME contre 4,0 (.50) pour le clavier virtuel, U=16,50, Z= -3,564, p =.00) et la gestion des erreurs perçue (2,0 (1,0) pour DIGIME contre 4,0 (.50) pour le clavier virtuel, U=31,00, Z= -2,672, p =.01). De même, les jugements sur la facilité d'apprentissage perçue ont varié entre ces deux interfaces (3,5 (3,0) pour DIGIME contre 5,0 (.50) pour le clavier virtuel, U=35,00, Z= -2,593, p =.01).

Un fait intéressant est que, en dépit des différences qui ont été mesurées par le test d'utilisabilité de Nielsen, les jugements qui y sont rapportés sont relativement bons.

6.4. Discussion

Le but de cette recherche était de comparer DIGIME à des logiciels assurant des fonctions équivalentes sur PDA. Nous voulions vérifier si les améliorations que nous avons apportées à DIGIME étaient suffisamment utiles pour en augmenter l'acceptabilité par rapport aux solutions concurrentes. Tout donne à penser le contraire. De toutes les interfaces utilisées dans cette expérience, DIGIME a été celle permettant les moins bonnes performances. En situation de découverte après un entraînement, les participants ont mis, en moyenne plus de 1500s (25 mn) pour prendre en notes le commentaire. Il a fallu moitié moins de temps aux utilisateurs de *JOT* pour réaliser la même tâche. L'observation des scripteurs en situation avec ces deux interfaces a montré que les utilisateurs de DIGIME avaient eu beaucoup de difficulté à assimiler le geste d'édition pour « entrée ». Avec *JOT*, ce problème ne se pose pas puisque cette fonction était assurée par un bouton sur lequel appuyer avec le stylet. La difficulté à réaliser ce geste avec DIGIME ne tenait pas tant, selon nous, à la réalisation du geste (assez facile en soi), qu'à sa compréhension à partir du didacticiel et des aides en lignes. Cette entrave a, de toute évidence, largement handicapé les utilisateurs de DIGIME. Ces problèmes de saisie des gestes d'édition ne sont pas inconnus, puisque Költringer et Grechenig (2004) avaient déjà fait cette observation. Cependant, même avec *JOT*, on mesure toute la marge de progression qu'il reste à franchir pour pouvoir faire rivaliser les logiciels de reconnaissance d'écriture manuscrite avec le clavier virtuel. Cela ne fait que confirmer nombre d'études dont

nous avons fait la présentation dans notre introduction théorique (Commaford, 2004; Curran, Woods, & O Riordan, 2006; Hinckeley, 2003; Jansen & Ayers, 2004; MacKenzie & Chang, 1999; MacKenzie, Nonnecke, McQueen, Riddersma, & Meltz, 1994; MacKenzie & Soukoreff, 2002; MacKenzie & Zhang, 1997; McQueen, MacKenzie, Nonnecke, Riddersma, & Meltz, 1994; Ortelli, 2003; Sears & Arora, 2002). Pourtant, écrire avec le clavier virtuel revient à ne taper qu'avec un doigt : le stylet. L'écart est d'autant plus important si l'on met en perspective *JOT* et DIGIME face à *Note*, qui constituait l'étalon d'une situation idéale de saisie de texte manuscrit sur PDA, puisqu'elle n'impliquait pas de reconnaissance. Il est permis de s'interroger sur les causes de cette différence. Malgré tout le soin apporté à l'élaboration d'interfaces de reconnaissance d'écriture manuscrite, et les méthodes « d'apprentissage » dont elles bénéficient, les taux de reconnaissance théoriques qu'elles affichent sont loin de ce que l'expérimentation révèle (Anquetil & Bouchereau, 2002; Anquetil & Bouteruche, 2004; Anquetil & Lorette, 1996, 1997a, 1997b; Bouteruche, Deconde, Anquetil, & Jamet, 2005).

Le fait que l'entraînement ait duré plus longtemps avec *JOT* et DIGIME montre qu'il y a bien eu un apprentissage. Cette phase de découverte n'existe manifestement pas avec le clavier et *Note*. Or, la facilité d'apprentissage est une dimension importante de l'utilisabilité. Cette différence d'apprenabilité entre les interfaces peut à elle seule constituer un motif, sinon de rejet, au moins de diminution de l'acceptation. Il est évident que si l'utilisabilité est un facteur d'acceptabilité (Arning & Ziefle, 2007; Hasan & Ahmed, 2007; ISO-9241-11, 1994; Keinonen, 1998; Nielsen, 1993; Rogers, 1995; Shackel, 1991; Tornatzky & Klein, 1982; Tricot & Tricot, 2000; Venkatesh & Davis, 1996), ce que Nielsen (1993) appelle « l'utilité pratique » y contribue plus encore. Or, manifestement, les interfaces de reconnaissance de l'écriture manuscrite n'ont pas une utilité pratique en adéquation avec les attentes des utilisateurs. *JOT* et DIGIME font bien la même chose que le clavier virtuel, c'est l'utilité théorique. Cependant, l'utilité pratique varie largement entre ces méthodes de saisie. Cela influence certainement les jugements des utilisateurs.

Les jugements exprimés par les participants sur l'interface qu'ils ont eu à manipuler peuvent être assez contradictoires. C'est notamment le cas pour les scores au test d'utilisabilité de Nielsen, pour lequel DIGIME a été jugé de la même façon que *Note*. Sans vouloir éluder ces résultats, rappellons que Keinonen (1998) fait remarquer que comparer entre eux des dispositifs assurant des fonctions différentes n'a que peu de sens. Selon les

interprétations, il est possible de considérer que *Note* et DIGIME ont, ou pas, les mêmes fonctions. Dans les deux cas, ces interfaces servent à écrire. *Note* est cependant plus adapté à la PdN, là où DIGIME est lui, compatible avec d'autres logiciels (*e.g.* traitement de texte ou tout autre champ). Leur utilité peut donc varier, en fonction des situations et des interprétations. Par contre, l'examen des différences engendrées par la pratique de DIGIME *vs. JOT* et le clavier virtuel ne pose pas le même problème, puisque chacune de ces interfaces assure précisément la même fonction. Des différences très nettes d'utilisabilité perçues ressortent de la comparaison entre DIGIME et le clavier virtuel. C'est moins le cas avec *JOT*, en dépit d'une très nette différence de performance entre les deux. Cela peut être interprété comme un refus global de la reconnaissance d'écriture manuscrite sur PDA. Néanmoins, les sujets de l'expérience n'ont pas eu l'opportunité de comparer les interfaces entre elles. Ils ont donné leur avis « à l'aveugle », sans cadre de référence puisqu'ils ne sont pas utilisateurs de PDA. C'est une lacune importante. L'examen attentif des scores produits au test de Nielsen montre d'ailleurs des valeurs médianes relativement bonnes, quelle que soit l'interface. Il est donc difficile de conclure à partir de ces données.

Les mêmes remarques peuvent être faites à propos du TAM. Il montre que l'intention d'utilisation de chaque interface, même hypothétique, reste relativement faible. C'est aussi vrai pour *Note*, contrairement au clavier virtuel. Faut-il y voir un certain rejet de l'écriture manuscrite sur écran tactile? Rien ne nous permet de l'affirmer à partir des données, mais les témoignages libres des sujets ont pu nous le laisser entendre. Cela semble d'autant plus plausible que le clavier virtuel est l'interface la plus acceptable. DIGIME ne diffère de *JOT* que sur la facilité d'utilisation perçue, pas sur l'utilité, ni les attitudes ou l'intention d'usage. Comparé à *Note*, DIGIME a été perçu moins facile et moins attrayant (attitudes). Cela n'a cependant pas suffit à faire de *Note* un dispositif significativement plus acceptable que DIGIME. Par contre, le clavier virtuel a été perçu comme plus facile, plus utile, plus agréable, et par conséquent, plus acceptable. De même, le test d'utilisabilité de Nielsen n'a montré que des différences assez ténues entre *JOT* et DIGIME, qui ne se sont différenciées que pour la gestion des erreurs, et à l'avantage de DIGIME. Ce constat est encore plus vrai pour la comparaison DIGIME / Note. Seul le clavier virtuel s'est véritablement distingué de DIGIME en matière d'utilisabilité perçue. Bien sûr, cette analyse ne vaut que pour l'échantillon sur lequel nous avons conduit cette expérience, qui, soulignons-le, n'aspirait sans doute pas à faire l'acquisition d'un PDA à court terme.

Si notre explication d'un rejet possible de l'écriture manuscrite sur PDA est exacte, cela pourrait signifier que les Pocket PC ne sont pas des outils de prise de notes pour les participants. Certes, la saisie d'informations écrites sur PDA est utile et nécessaire, mais peut-être plus pour des séquences de mots plus courtes que ce que l'épreuve imposait (ex : numéro de téléphone, rendez-vous dans un agenda électronique, URL, SMS). Les PDA et Smartphones, bien qu'utiles, puisque utilisés par un nombre croissant d'individus, ne sont peut-être pas dévolus à des tâches de PdN. Une étude de MacKenzie et Chang (1999) avait également pointé que le clavier était préféré pour la saisie de textes plutôt longs, alors que les logiciels de reconnaissances l'étaient pour des textes courts (type formulaire). Il est pourtant difficile de comparer cette recherche à la notre : elle s'est faite sur tablette graphique et ordinateur de bureau. Il ne s'agissait donc pas de PdN sur PDA, et les claviers étaient matériels. Quoi qu'il en soit, le clavier virtuel montre, une fois encore, qu'il est parfaitement adapté aux attentes des utilisateurs pour entrer du texte sur un PDA. Il l'est même plus que Note.

7- Expérience 4 : Effet du niveau d'expertise en informatique sur les performances et l'intention d'utilisation de DIGIME

Résumé – Nous testons l'utilisabilité et l'acceptabilité de DIGIME en faisant varier la population cible : experts ou novices en informatique. Les résultats nous apprennent que les experts, en dépit de connaissances antérieures en informatique réellement supérieures à celles des novices, ne disposent pas de plus de connaissances préalables sur l'interface. Ils n'ont pas non plus une plus grande envie de tester l'interface sur sa seule présentation visuelle, mais pensent avoir plus de facilité à le faire. Les experts obtiennent de meilleures performances que les novices. Cela se répercute sur les jugements portés en termes d'acceptabilité et d'utilisabilité perçue. Néanmoins, DIGME reste assez mal considéré par les participants, quels qu'ils soient. Cela peut s'expliquer, en partie, par le fait que les défauts constatés précédemment n'ont pas encore été corrigés. Il ressort de cette étude qu'une convergence entre les champs MIS et HCI est nécessaire à l'amélioration du traitement des questions d'utilisabilité et d'acceptabilité.

7.1. Introduction

Nous avons évoqué toute l'importance de l'échantillonnage pour la conduite d'un test d'utilisabilité. Le problème soulevé par le choix d'une population cible, pour ce type d'expérience est crucial (Nielsen, 1993). En effet, le choix des participants va déterminer, pour une bonne part, la qualité des mesures d'interaction relevées au cours de ce type de test. Les développeurs de logiciels de saisie et de reconnaissance de l'écriture manuscrite visent à les populariser auprès d'un public large, sans restriction. L'échantillon que nous avons réuni pour l'expérience 3 était issu d'une seule et unique population, par conséquent, l'affectation aléatoire des sujets n'a pas pu biaiser les performances d'un groupe à l'autre. Cependant, cet échantillon n'était ni représentatif de la population dans son ensemble, ni « technophile » ou susceptible de faire l'achat d'un PDA. Cette expérience a, selon nous, témoigné d'un certain rejet de l'écriture manuscrite sur PDA pour nos participants. Cette interprétation, prise en dehors de tout contexte, n'a que peu de sens : les jugements produits, notamment pour l'intention d'usage, ont subi la double influence de l'absence de comparaison des interfaces entre elles, ainsi que le fait que la population choisie n'était pas à proprement parler « technophile » ou désireuse d'acquérir un PDA. Or, si un échantillon n'est composé que d'individus qui n'envisagent même pas l'utilité d'un PDA, la question de l'intention d'utilisation, même posée au conditionnel, est problématique. Or, l'acceptation d'un système n'est pas que le fait d'un choix individuel. L'IDT (1995) a traité cette question et a montré que la diffusion d'une technologie passait par un processus d'influence sociale déclenché par les « innovateurs » et « primo-adopteurs ». Ces personnes sont en général des individus intéressés par l'innovation, capables de prise de risque face à la nouveauté. Cela signifie qu'une large partie des individus ne se risque à l'adoption d'une technologie qu'après qu'un certain nombre de primo-adopteurs l'aient fait eux-mêmes. Par conséquent, le traitement de la question de l'acceptabilité gagnerait sans doute à être envisagé chez ces innovateurs dont on sait justement qu'ils peuvent initier la diffusion d'un dispositif nouveau (Rogers, 1995). Par ailleurs, il est connu que si la sensibilité des utilisateurs par rapport à l'innovation a une influence sur leurs jugements et comportement d'adoption, elle en a aussi sur leurs performances d'utilisation (Morahan-Martin & Schumacher, 2007; Segall, Doolen, & Porter, 2005). Dès lors, il est possible que les logiciels de reconnaissance de l'écriture manuscrite sur PDA ne soient pas bien acceptés car peu entrés dans nos mœurs. Ainsi, les individus pourraient, par ignorance, préjuger de leur qualité avant même l'usage, ce qui, incidemment,

nuirait à leur utilisation effective en termes d'efficacité (et conséquemment, de jugements). Ce cercle vicieux ne serait brisé que par les processus sociaux d'influence qu'a expliqués l'IDT. D'où l'intérêt de se pencher précisément sur l'acceptabilité du système avec une population dont on suppose *a priori* qu'elle n'a pas de réticences face à l'informatique et aux TIC.

L'acceptation d'une technologie ne repose pas seulement sur des mécanismes sociaux. La question a aussi été traitée au niveau individuel et cognitivo-comportemental. Quelques travaux laissent effectivement entendre que les caractéristiques psychologiques ou démographiques des individus peuvent influer ou modérer leur intention comportementale d'utilisation d'un dispositif. Cet aspect de l'acceptation des technologies a interrogé King et He (2006) dont la méta-analyse, par manque de données, n'a pas pu conclure sur les effets des compétences antérieures des participants à des « tests d'acceptation ». Il leur semble pourtant que cela joue un rôle négligé. Il est vrai que beaucoup d'études sur le TAM ou ses variantes (*e.g.* WAM de Castañeda, Muñoz-Leiva, & Luque, 2007 ou UTAUT de Venkatesh, Morris, Davis, & Davis, 2003) s'appuient sur une population d'étudiants *a priori* sensible à l'usage de nouvelles technologies, et dont ils pourraient avoir besoin (applications bureautiques en lignes sur des campus universitaires). Dès lors que la population est tout-venant, on observe des modifications dans les opinions par rapport à d'autres populations (étudiantes ou professionnelles). Quoi qu'il en soit, ce champ de recherche sur l'acceptation des technologies est fort mal connu car très récent (Arning & Ziefle, 2007). Pourtant, il existe déjà un certain nombre de résultats expérimentaux qui ont fait la démonstration de l'impact des caractéristiques démographiques (âge, sexe) et cognitives (auto-efficacité, techno-phobie) sur les intentions d'usage (Burton-Jones & Hubona, 2006; Legris, Ingham, & Collerette, 2003; Ong & Lai, 2006; Porter & Donthu, 2006; Saadé & Kira, 2007; Schepers & Wetzels, 2007; Venkatesh & Davis, 1996; Venkatesh, Morris, Davis, & Davis, 2003). C'est pourquoi l'étude que nous proposons souhaite traiter cette question.

La conception d'interface stylo se veut intuitive pour l'utilisateur novice. Que l'on soit ou non spécialiste de l'informatique, tout un chacun sait écrire. Cela ne relève *a priori* pas de compétences particulières. L'interface DIGIME exploite principalement les capacités de chacun à l'écriture, d'une façon qui se veut assez naturelle (Bouteruche, Deconde, Anquetil, & Jamet, 2005). Elle ne repose pas sur une architecture dont la couche d'abstraction (arborescence) est très profonde (Vanderdonckt & Coyette, 2006). Il serait cependant

maladroit d'en juger sur notre seule interprétation. Il est possible que cette interface, conçue dans un souci d'adéquation à des normes et principes (Bastien & Scapin, 2001; Jansen & Ayers, 2004) couramment admis, fasse appel à des connaissances particulières à la pratique de l'informatique. Ainsi, utiliser DIGIME ne constituerait pas seulement un transfert de compétences en écriture, mais également un transfert de connaissances plus générales en informatique. Il est possible de le tester en évaluant les performances et jugements engendrés par l'utilisation de DIGIME chez des experts en informatique, comparés à des novices (les deux populations étant, bien évidemment, novices dans l'utilisation de DIGIME) (Morahan-Martin & Schumacher, 2007). Les travaux sur l'expertise montrent que les experts ont acquis et automatisé un certain nombre de schémas qui leur permettent de catégoriser plus rapidement les problèmes afin de leur choisir une procédure de résolution adaptée (Anderson, 1982; Cooper, 1990, 1998; Tricot, 1998). Il arrive cependant que les experts, dans leur domaine de prédilection, ne se montrent pas plus doués que les novices (Ericsson, 2005; Farrington-Darby & Wilson, 2006). Bien que la catégorisation des experts par rapport aux novices puisse être problématique, plusieurs critères sont reconnus pour en être de bons indicateurs (Farrington-Darby & Wilson, 2006) : l'expérience, la spécialisation ou la qualification sont des qualités qu'il est facile de vérifier. La difficulté tient plus au choix des indicateurs qui permettront de savoir en quoi l'expertise informatique améliore la qualité de l'interaction avec le système. En effet, si l'on peut considérer intuitivement que les informaticiens n'écrivent pas mieux que les autres, l'approche de l'expertise développée par certains donne à penser qu'elle joue un rôle d'adaptation à l'environnement (Ericsson & Lehmann, 1996; Sweller, 2004). Or, notre environnement étant de plus en plus informatisé, ces experts pourraient faire figure d'individus particulièrement adaptés à ce type de situation. Ils seraient capables de transférer utilement leurs connaissances générales sur ce thème à une situation nouvelle de telle façon qu'elle soit mieux résolue que ne l'aurait fait un novice (Perkins & Salomon, 1994). Si tel est le cas, des experts devraient, au cours d'un test de perception, identifier facilement les fonctions attachées aux boutons et zones de l'interface (Beaudoin-Lafon & Mackay, 2003; Snyder, 2003; Watzman, 2003). On pourrait ainsi considérer cela comme une émanation de leur capacité à définir le problème à résoudre que pose la manipulation de l'interface. Cet avantage se répercuterait alors sur les performances et sur les opinions engendrées par la manipulation de l'interface. Cet effet serait même d'autant plus important que l'expertise en informatique donne, avant même la découverte de l'interface, un sentiment de confiance plus important, constituant un terreau favorable à une

prise en main réussie de l'interface (sentiment d'auto-efficacité, Bandura, 1977). C'est d'ailleurs pour tenir compte de cela qu'Ajzen (1985) a modifié la TRA pour créer le TPB en y ajoutant le contrôle comportemental perçu.

7.2. Méthode

7.2.1. Participants

Pour conduire cette expérience, nous avons réuni deux échantillons issus de deux populations différentes. Un premier échantillon de 26 personnes constituait le groupe de novices. Agés en moyenne de 19,6 ans, il était composé de 20 filles pour 6 garçons parmi les étudiants en licence 1 et 3 de psychologie à l'université Rennes 2. Le second échantillon, constitué de 12 personnes, a été recruté au sein de la troisième année d'école d'ingénieur en informatique de l'INSA (Institut National des Sciences Appliquées). Il était composé de 9 garçons et 3 filles d'une moyenne d'âge de 20,0 ans. Aucun des participants n'était utilisateur de PDA, de même qu'ils étaient tous novices dans l'utilisation de DIGIME. Leur participation à l'expérience était volontaire et n'a donné lieu à aucune rémunération ou récompense.

7.2.2. Matériel

7.2.2.1. Matériel informatique

Nous avons utilisé deux Pocket PC E400 Toshiba avec stylets (*Intel PXA 261* cadencé à 300 MHz, 31,22 Mo de mémoire vive, environnement *Windows Mobile*) ainsi que deux ordinateurs portables (Dell Inspiron 6000 *Intel Mobile* cadencé à 1,73 GHz, 502 Mo de mémoire vive; et un Toshiba Satellite 2800 *Intel Pentium III* cadencé à 1GHz, 256 Mo de mémoire vive; sous environnement *Windows XP*). Durant la passation de l'expérience, les PDA étaient reliés avec deux ordinateurs distants pour le suivi en direct, par l'expérimentateur, de l'activité des participants, ce afin de s'assurer du bon déroulement de la passation sans les gêner. La connexion entre les PDA et les PC distants était assurée par les logiciels *Microsoft ActiveSync* 3.7 et *Remote Display Control* 2.03.

7.2.2.2. Matériel linguistique, questionnaires et interfaces

Nous avons choisi de réutiliser la planche de la figure 43 comme texte à écrire par les participants. Cet ensemble de lettres, de chiffres et de mots n'était pas réaliste, mais présentait l'avantage de contenir tout l'alphabet et les dix chiffres. La durée de l'expérience précédente avec DIGIME ne nous permettait pas d'envisager de reprendre son protocole à l'identique.

Les questionnaires utilisés pour cette expérience sont les mêmes que pour le test utilisateur (pré-questionnaire compris). Nous y avons cependant ajouté le NASA-TLX puisque Segall, Doolen et Porter (2005) assurent que les attitudes envers les technologies pouvaient modifier la charge de tâche.

S'agissant du pré-questionnaire, il faut remarquer que si pour les manipulations précédentes, il ne visait qu'à s'assurer de l'équivalence des différents groupes entre eux, ainsi que de l'inexpérience des sujets avec les PDA, il nous a permis ici de vérifier l'expertise réelle des informaticiens. Les items le composant étaient les suivants :

- Selon quelle périodicité utilisez-vous l'informatique ? (Jamais, mensuelle, hebdomadaire ou quotidienne).
- Avez-vous un PDA ? (oui, non)
- Selon quelle périodicité utilisez-vous un PDA ? (Jamais, mensuelle, hebdomadaire ou quotidienne).
- Citez quelques marques de Pocket PC.
- Donnez quelques exemples de systèmes d'exploitation employés sur les Pocket PC.
- Citez deux applications de reconnaissance de l'écriture manuscrite.

L'interface DIGIME utilisée est la même que pour l'expérience 3. Le délai séparant ces deux expériences étant trop court pour faire en sorte que, dans le cadre du cycle de conception itératif de DIGIME, l'interface ait pu bénéficier des suggestions d'améliorations que nous avons soumises à ses concepteurs (notamment pour la saisie du geste d'édition « entrée »).

7.2.2.3. Test de perception

Afin de tester notre hypothèse, nous avons conçu un prototype papier de DIGIME servant de support à un test de perception. Nous avons imprimé sur une carte plastifiée de format A4 une capture d'écran en couleur de DIGIME. Cette image était légendée, mais la légende était vide. A partir de cela, la consigne pour les participants était de déterminer à quoi pouvait servir les zones désignées par la légende (voir figure 45).

Ce test de perception a été augmenté de deux questions préalables à l'utilisation effective de l'interface :

- La capture d'écran du pda vous donne-t-elle envie d'utiliser ce logiciel de saisie et de reconnaissance de l'écriture manuscrite ?
- La capture d'écran du pda vous donne-t-elle l'impression que ce dispositif est facile à utiliser?

Les réponses à ces questions se faisaient sur deux échelles types Lykert graduées de -2 (pas du tout) à +2 (tout à fait).

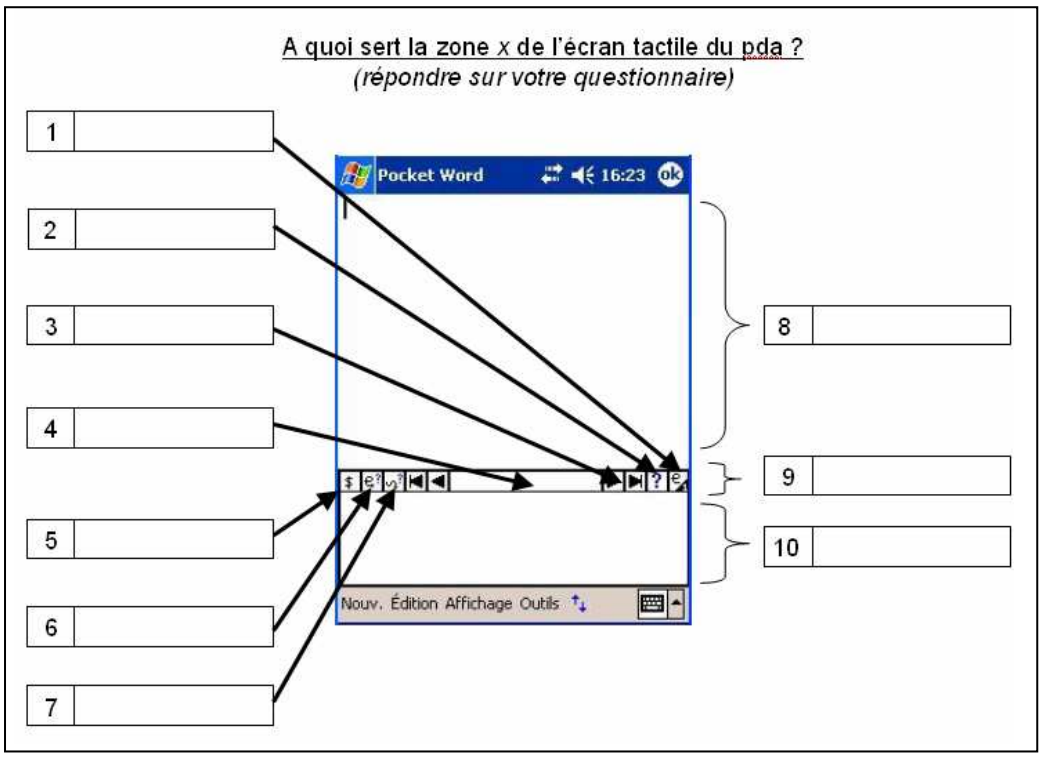


Figure 45: Test de perception

7.2.3. Procédure expérimentale

Immédiatement après avoir répondu au pré-questionnaire, les participants devaient, à partir du test de perception, déterminer quelle fonction était selon eux affectée à chaque zone de l'interface DIGIME. A la suite de cela, il leur fallait reproduire sur l'écran du PDA les caractères tels que présentés à la figure 43. Ils devaient respecter scrupuleusement, comme pour la phase d'entraînement de l'expérience 3, l'organisation du texte de la carte. Les sujets n'ont reçu aucune indication sur le fonctionnement de l'interface DIGIME. Ils avaient par contre l'autorisation de chercher par eux-mêmes dans l'interface toute information susceptible de leur permettre de mener cette tâche à bien dans les meilleurs délais. La durée de réalisation de cette épreuve était limitée, pour des raisons pratiques, à un délai de 30 minutes maximum. Une fois la tâche terminée ou le délai épuisé, les participants répondaient aux questionnaires TAM, NASA-TLX et test d'utilisabilité de Nielsen. L'expérience durait 45 minutes.

7.2.4. Variables indépendantes

Une seule VI était ici manipulée. Elle était invoquée, intergroupe, à deux modalités. Il s'agit du niveau de connaissance en informatique (expert ou novice). Dans la pratique, les étudiants de Rennes 2 en psychologie étaient considérés comme novices, tandis que les étudiants en troisième année d'école d'ingénieur en informatique à l'INSA étaient considérés comme experts.

7.2.5. Variables dépendantes

Nous avons enregistré les temps d'écriture (seconde), et, le cas échéant, le nombre de caractères manquant (ou erronés). Cette dernière mesure n'a eut lieu que dans le cas où la durée de saisie pour arriver au résultat escompté excédait les 30 minutes autorisées. A cela s'ajoute les scores obtenus à chacun des trois tests TAM, NASA-TLX ainsi qu'au test d'utilisabilité. Pour finir, nous avons mesuré le nombre de bonnes réponses fournies au test de perception, associées aux deux questions qui le complètent sur la facilité d'utilisation perçue avant usage et sur l'envie d'utilisation.

7.2.6. Hypothèses

- H1: Les scores au pré-questionnaire des experts en informatique seront meilleurs que ceux des sujets novices.

- H2: Les experts en informatique auront un meilleur score au test de perception et témoigneront d'une plus grande envie d'utilisation de DIGIME, jugeant l'interface plus facile à utiliser avant usage.

- H3: Les experts en informatique auront de meilleures performances (temps de saisie) à l'épreuve, ce qui se traduira également par de meilleurs jugements en termes d'utilisabilité perçue, d'intention d'utilisation et de charge de tâche.

7.3. Résultats

7.3.1. Effets du niveau d'expertise en informatique au pré-questionnaire

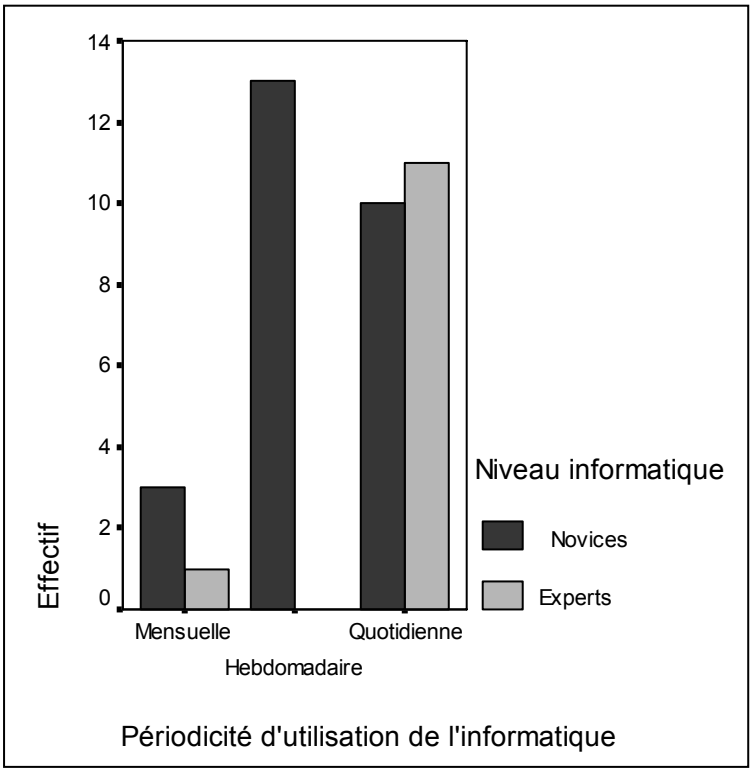
Experts ou novices en informatique, aucun des participants n'était possesseur ou utilisateur de PDA (voir tableau 5).

	Novices	Experts
Possession PDA	non	non
Périodicité d'usage PDA	jamais	jamais
Scores questions ouvertes	.12 (.43)	.75 (.86)

Note : Score moyen (écart-type)

Tableau 5: Effet du niveau en informatique sur les scores au pré-questionnaire

Par contre, les élèves ingénieurs utilisaient significativement plus l'informatique que les étudiants en psychologie (χ^2 (2, N = 38) = 10,282, p = .00, voir figure 46, page suivante).



Note : « Selon quelle périodicité utilisez-vous l'informatique ? »

Figure 46: Effet du niveau en informatique sur la fréquence d'utilisation de l'informatique

Pour finir, les experts ont été meilleurs que les novices au pré-questionnaire (U=90,00, Z= -2,914, *p* =.00). En effet, ils ont obtenu au questionnaire un score moyen de .75 (*SD* = .86) alors que les novices n'y ont comptabilisé que .12 (*SD* = .43).

7.3.2. Effets du niveau d'expertise en informatique au test de perception

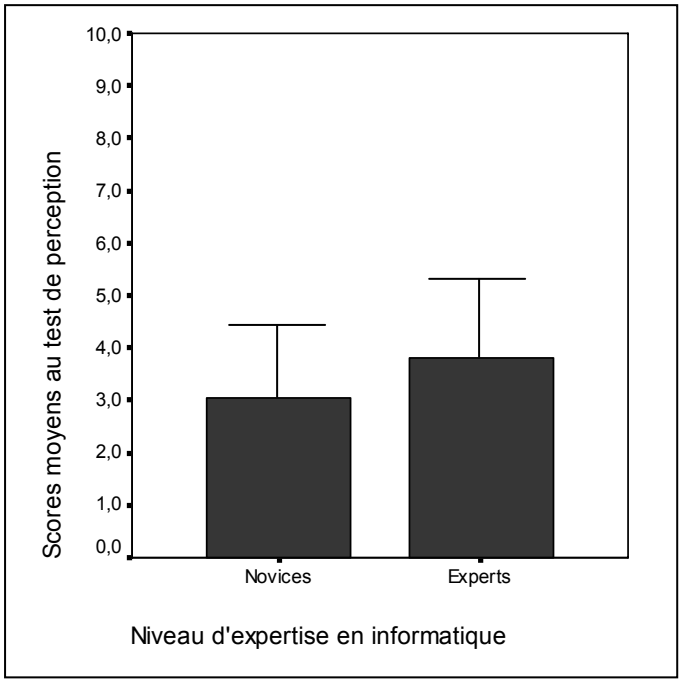
Les réponses aux questions portant sur l'envie d'utilisation de DIGIME et sa facilité d'utilisation perçue d'après le seul examen du prototype papier se sont faites sur des échelles ordinales. Nous les avons donc traitées avec le test de Mann-Whitney, adapté à ce type d'échelles (Howell, 1997, p.726). Nos analyses n'ont pas montré de différences entre les deux groupes, suivant qu'ils étaient experts ou novices en informatique, sur l'envie préalable d'utilisation de DIGIME (U=110,50, Z= -1,606, *p* =.15). Par contre, les experts ont jugé *a priori* DIGIME plus facile d'utilisation à partir du prototype (U=90,00, Z= -2,173, *p* =.03). Les scores à ces questions sont résumés dans le tableau 6.

	Novices	Experts
Envie d'utilisation préalable	1,0 (1,0)	1,0 (0,75)
Facilité d'utilisation perçue préalable	0 (2,0)	1,5 (1,75)

Note : Médiane (intervalle interquartile). Les échelles de réponses sont orientées entre un pôle très négatif ou difficile à 0 et un pôle très positif ou facile à 5.

Tableau 6: Effets du niveau en informatique sur les scores aux questions du test de perception

Pour les scores au test de perception (identification des zones de l'interface), les conditions de normalité et d'homogénéité des variances ayant été réunies, nous leur avons appliqué une analyse de variance (ANOVA à un facteur). Aucune différence n'apparaît d'un groupe à l'autre (F(1,37) = 2,339, p = .13, voir figure 47 ci-dessous), puisque les novices ont déterminé en moyenne la fonction de M = 3,0 (SD = 1,3) zones contre M = 3,7 (SD = 1,6) pour les experts.



Note : Score compris entre 0 et 10

Figure 47: Effet du niveau en informatique sur les scores au test de perception

7.3.3. Effets du niveau d'expertise en informatique sur les performances

La durée de l'épreuve était, par commodité, limitée à 30 minutes. Cela n'a pas été sans affecter la distribution des données, d'autant plus que le nombre de caractères manquant ou erronés n'a été calculé que dans le cas où les participants n'avaient pu mener la tâche à bien. Or, seuls les experts ont tous réussi à finir la tâche dans le temps imparti. Par conséquent, ces données ont été traitées avec un test non paramétrique. Le test de Mann-Whitney a montré qu'aussi bien les temps de saisie (U=69,00, Z= -2,924, *p* =.00) que le nombre de caractères manquants (U=42,00, Z= -3,829, *p* =.00) diffèrent entre les groupes, comme en témoigne le tableau 7.

	Novices	Experts
Temps de saisie (s)	1719 (180)	1567 (173)
Nombre d'erreurs et caractères manquants	57,45 (31,60)	0 (0)

Note : Moyenne (écart-type)

Tableau 7: Effets du niveau en informatique sur les performances de saisie à l'épreuve

7.3.4. Effets du niveau d'expertise en informatique sur les jugements

7.3.4.1. Test TAM

Les scores médians aux dimensions du test TAM sont résumés dans le tableau 8. Comme pour l'expérience précédente, le faible effectif que nous avons réuni ne permettait pas de procéder à une analyse factorielle confirmatoire suivi d'une analyse de fiabilité. C'est pourquoi, confiants dans la robustesse du TAM, nous avons fait le choix de traiter les dimensions du test en regroupant les items qui leurs étaient subordonnés afin d'en en calculer la médiane.

L'intérêt que nous avons porté à ces mesures résidant principalement dans l'expression de leur valeur médiane, nous les avons traitées une fois encore avec le test de Mann-Whitney.

Aucune différence significative n'est apparue entre les groupes pour la norme subjective (U=146,50, Z= -.307, *p* =.76) et les attitudes (U=97,50, Z= -1,877, *p* =.06). Par contre, les experts ont jugé DIGIME plus facile (U=82,50, Z= -2,387, *p* =.01), plus utile (U=62,50, Z= -3,004, *p* =.00), et aussi plus acceptable (U=84,00, Z= -2,297, *p* =.02).

	Novices	Experts
Facilité d'utilisation perçue	4,0 (3,0)	3,0 (2,25)
Utilité perçue	6,0 (1,0)	5,0 (1,88)
Attitude	3,5 (2,5)	3,75 (.88)
Intention d'utilisation	6,0 (2,5)	5,75 (3,13)
Norme subjective	5,0 (2,0)	4,75 (1,87)

Note : Médiane (intervalle interquartile). Les échelles de réponses sont orientées entre un pôle très positif à 1 et un pôle très négatif à 7.

Tableau 8: Effets du type d'interface sur les dimensions du TAM

7.3.4.2. Test d'utilisabilité de Nielsen

Pour les mêmes raisons que citées précédemment, le test de Mann-Whitney a été retenu pour analyser ces données dont le tableau 9 fourni un récapitulatif. Un certain nombre d'items au test de Nielsen ont été sensibles au niveau en informatique des sujets. Pour commencer, l'apprentissage de DIGIME a été perçu moins facile par les novices en informatique (U=72,50, Z= -2,599, p =.01). Idem pour le sentiment d'efficacité, moins affirmé chez les non-informaticiens (U=87,00, Z= -2,274, p =.03), ainsi que pour l'apprenabilité de l'interface (U=94,50, Z= -2,299, p =.05). Par contre, les deux groupes ont appréhendé de la même façon la gestion des erreurs (U=111,50, Z= -1,522, p =.16). Au final, la satisfaction était identique, indépendamment que les participants étaient rompus, ou non, à l'informatique (U=105,00, Z= -1,650, p =.11).

	Novices	Experts
Facilité d'apprentissage perçue	2,0 (3,0)	3,5 (2,5)
Efficacité perçue	3,0 (1,0)	3,5 (1,0)
Apprenabilité perçue	4,0 (2,0)	5,0 (0)
Gestion des erreurs	4,0 (1,0)	4,0 (1,75)
Satisfaction	3,0 (3,0)	3,0 (2,0)

Note : Médiane, (intervalle interquartile). Les échelles de réponses sont orientées entre un pôle très négatif ou difficile à 0 et un pôle très positif ou facile à 5.

Tableau 9: Effets du niveau en informatique sur les items du test d'utilisabilité de Nielsen

7.3.4.3. Test de charge de tâche NASA-TLX

Puisqu'ils souscrivaient à ses conditions d'application, nous avons analysé les résultats au NASA-TLX par l'ANOVA à un facteur. Elle a montré qu'il n'y avait pas de différence significative entre les deux groupes M = 13,82 (SD = 3,71) pour les novices contre 14,37 (3,70) pour les experts ($F(1,37) < 1$).

7.4. Discussion

La méthode d'échantillonnage que nous avons choisie en s'appuyant sur des critères d'expérience, de spécialisation et de qualification, suivant les recommandations de Farrington-Darby et Wilson (2006) a porté ses fruits. Bien que novices dans l'utilisation de PDA, les informaticiens ont produit de meilleures réponses au pré-questionnaire que les novices. Si ce résultat n'était pas surprenant, il fallait néanmoins prendre d'importantes précautions pour opérationnaliser la notion d'expertise. Nous savons qu'elle est sujette à un certain flou quant à sa définition, et la prudence s'imposait (Ericsson, 2005; Farrington-Darby & Wilson, 2006).

Nous avions émis l'hypothèse que les experts en informatique obtiendraient de meilleurs scores au test de perception, ce qui se traduirait par une envie plus forte d'utiliser l'interface DIGIME, car jugée plus facile. Elle n'est que très partiellement validée. L'inspection du prototype papier par les participants, qu'ils soient ou pas experts en informatique, ne conduit à aucune différence dans la l'attribution supposée d'une fonction à une zone de l'interface (Beaudoin-Lafon & Mackay, 2003; Snyder, 2003; Watzman, 2003). De la même façon, l'envie d'utiliser DIGIME suscitée par la présentation d'une capture d'écran a été égale parmi les deux groupes. En revanche, la facilité d'utilisation perçue préalable de DIGIME a été plus forte chez les informaticiens que chez les autres. Ces indicateurs ont donc montré que mis en situation de découverte de l'interface DIGIME, experts et novices avaient autant de connaissances sur l'interface. L'expertise en informatique n'a pas apporté de savoirs particuliers, qui, par transfert de connaissances, aurait donné à ceux en bénéficiant une « maîtrise innée » des fonctions du dispositif (Perkins & Salomon, 1994). Par conséquent, si nous n'avons pas à douter de l'expertise de notre panel d'informaticiens, leur niveau de connaissances en informatique n'a présenté *a priori* que peu d'avantage pour la découverte de l'interface DIGIME. Hormis la facilité préalable ressentie, aucune des VD du test de

perception n'a permis d'expliquer la supériorité des experts sur les novices dans la réalisation de la tâche. La perception de DIGIME est-elle une activité trop éloignée de la leur pour permettre un transfert de connaissance (Perkins & Salomon, 1994) ou s'agit-il d'un autre cas d'absence de supériorité d'experts sur des novices (Ericsson, 2005; Farrington-Darby & Wilson, 2006) ? Rien ne nous permet de le dire. Pourtant, si l'on reconnaît leur expertise comme telle, se traduisant notamment par l'élaboration de schémas en MLT (Anderson, 1982; Cooper, 1990, 1998; Tricot, 1998), les performances dont ils ont fait preuve à la tâche de saisie montrent qu'ils n'usurpent pas ce statut par rapport aux novices. De même, leur sentiment de confiance plus important quant à la facilité d'utilisation préalable de DIGIME laisse penser à une réelle capacité à s'adaptater à cette situation nouvelle (Ericsson & Lehmann, 1996; Sweller, 2004). Sans connaissances déclaratives autres que celles des novices sur l'interface DIGIME avant son utilisation, ils se sentent d'emblée plus confiants (Ajzen, 1985; Bandura, 1977). Plus important encore, ils sont aussi, par la suite, plus performants. Dans un environnement quotidien de plus en plus informatisé, les experts informaticiens feraient figure d'individus adaptés à ces situations nouvelles. Leurs connaissances antérieures, sans être spécifiquement applicables à un contexte nouveau, leur permettraient pourtant d'obtenir des niveaux de performances avec lesquelles les novices en informatique ne rivalisent pas. C'est une interprétation possible du constat selon lequel les individus sensibles aux technologies font preuve d'une plus grande efficacité dans leur manipulation (Morahan-Martin & Schumacher, 2007; Segall, Doolen, & Porter, 2005). Ainsi, et quand bien même les mesures du test de perception ne l'ont pas mis en évidence, les experts disposeraient effectivement de schémas en mémoire leur permettant une catégorisation et une résolution plus rapide des problèmes.

L'examen des temps de saisie et des caractères manquant montrent indubitablement une supériorité des experts en informatique, comme le prévoyait notre seconde hypothèse. Ces derniers ont d'ailleurs tous fini la tâche dans le temps imparti, ce qui n'a pas été le cas des novices. Cette supériorité permet de mieux comprendre le rôle des facteurs modérateurs dans l'intention d'utilisation d'un système. En effet, les études d'acceptation négligent le plus souvent les mesures de performances pour juger de l'acceptabilité d'un dispositif, puisqu'elles s'appuient sur des mesures d'usage auto-rapportées, qui ne donnent pas lieu à des mesures d'utilisabilité (Arning & Ziefle, 2007; Burton-Jones & Hubona, 2006; Legris, Ingham, & Collerette, 2003; Porter & Donthu, 2006; Saadé & Kira, 2007; Schepers & Wetzels, 2007;

Venkatesh & Davis, 1996). Cette expérience montre justement que la qualité de l'interaction avec la machine s'accompagne de jugements différenciés sur celle-ci. Elle est jugée à la fois plus utile, plus facile et plus acceptable par les élèves informaticiens. Bien que les groupes ne se soient pas différenciés sur la charge de tâche subjective, certaines dimensions de l'utilisabilité perçue comme la facilité d'apprentissage, l'efficacité et l'apprenabilité ont été jugées plus faibles par les novices. D'autre part, notre échantillon d'experts présentait un certain nombre de caractéristiques, par rapport aux novices, qui sont connues pour être des modérateurs de l'intention d'usage. Les informaticiens étaient majoritairement des garçons, dont la confiance en la facilité d'usage préalable de DIGIME était plus grande que pour les novices, principalement composés de filles. Ce type de données démographiques et d'attitudes sont des modérateurs avérés de l'intention d'usage d'un dispositif (Ajzen, 1985; Arning & Ziefle, 2007; Venkatesh & Davis, 1996). Il n'en reste pas moins vrai que leurs performances suffisent à comprendre les différences de jugements entre novices et experts. La question de savoir si le sexe et la confiance peuvent expliquer les performances, ou bien si plus simplement les informaticiens sont majoritairement des garçons sûrs d'eux ne sera pas résolue ici. Par ailleurs, ces variables invoquées ont à plusieurs reprises donné lieu à des résultats contradictoires (Arning & Ziefle, 2007; Ong & Lai, 2006; Venkatesh, Morris, Davis, & Davis, 2003). A défaut d'éclairer cette question, cette étude a le mérite de montrer qu'il semble bien vain de poser la question des modérateurs de l'intention d'usage sans tenir compte des performances qu'ils engendrent : des variables confondues en biaisent sans doute l'analyse. Il y a donc bien un vrai besoin d'unifier les approches HCI et MIS pour cerner de façon complète les questions d'acceptation des technologies (Zhang & Dillon, 2003).

Quoi qu'il en soit, cette étude a effectivement montré que les individus experts, ayant des attitudes positives vis-à-vis des technologies, étaient aussi plus efficaces dans leur utilisation (Morahan-Martin & Schumacher, 2007; Segall, Doolen, & Porter, 2005), sans que l'on sache comment cela se répercute sur les jugements qu'ils leur portent. Si effectivement, l'échantillonnage tient une grande importance dans le succès d'une étude d'utilisabilité (Nielsen, 1993), cette expérience souligne tout l'intérêt qu'il existe à tester une nouvelle technologie sur des personnes sensibles à l'innovation. En effet, l'IDT (Rogers, 1995) prévoit que la diffusion d'une technologie s'explique, en plus de caractéristiques propres au dispositif (avantage relatif, compatibilité, ou facilité d'utilisation) par des processus d'influence sociale. Or, ceux-ci sont initiés par les individus innovateurs, avant de se diffuser, le cas échéant, dans

le reste de la population. Les experts en informatique partagent vraisemblablement des caractéristiques avec les primo-adopteurs. Nous n'en avons pas fait la vérification, mais, de façon incidente, on peut légitimement penser que de futurs développeurs ne sont pas hostiles à l'innovation technologique. Par conséquent, ils pouvaient constituer un panel de référence sur lequel tester l'innovation avant qu'elle ne s'intègre et pénètre toutes les couches de la population.

Dans le cas de DIGIME précisément, cette étude permet de faire la part des choses entre utilisabilité et caractéristiques des utilisateurs. Les résultats sont alors assez mitigés : même si les jugements des experts sont meilleurs que ceux des novices sur des aspects tels que facilité d'utilisation perçue, utilité perçue et intention d'utilisation, cela reste à des niveaux assez bas. Une fois encore, DIGIME ne se montre pas très acceptable. L'utilité perçue, assez basse, en explique sans doute une large part. Il est vrai aussi que les défauts de conception de DIGIME que l'expérience 3 avait révélés n'étaient pas encore corrigés pour la présente étude. Les niveaux d'utilisabilité perçue ont pourtant été assez élevés chez les experts. Ceci est à pondérer par le fait que les participants n'ont pas pu comparer plusieurs interfaces entre elles, comme pour les expériences précédentes.

La question de l'utilité semble plus que jamais primordiale en fonction du choix de l'échantillon. Or, aucun des participants novices ou experts n'était utilisateur de PDA, condition *sine qua non* de leur inexpérience en la matière. En l'absence d'utilisateurs réguliers et convaincus de PDA pour le tester, améliorer l'interaction avec DIGIME se heurtera toujours au problème de l'utilité…

8- Expérience 5 : Utilisabilité objective et subjective : prédire l'acceptabilité comparée entre DIGIME et *Note*

Résumé – En permettant aux participants de comparer DIGIME et *Note* au cours d'une tâche de copie, nous espérons combler l'une des principales lacunes des expériences précédentes. On souhaite ainsi faire glisser la question de l'intention d'utilisation vers celle de l'acceptabilité des interfaces. Cette recherche vise à valider en partie de le modèle P3 (Dillon & Morris, 1998, 1999) en testant l'effet de mesures objectives (effort mental, efficacité et efficience) et subjectives (charge de tâche mentale) sur l'acceptabilité comparée de DIGIME et de *Note*. Pour ce faire, nous mettons à profit la procédure de Kellogg afin de mesurer objectivement l'effort mental. Une précaution méthodologique par groupe contrôle nous assure que la triple tâche ne parasite pas les performances des participants, ni même les jugements qu'ils portent sur les interfaces. Les résultats montrent que l'acceptabilité comparée des interfaces n'est prédite que par la charge de tâche mentale que l'on assimile à une mesure d'utilisabilité subjective. Aucune des mesures objectives d'utilisabilité n'y parvient. La charge de tâche médiatise leurs effets par l'intermédiaire de l'efficience mentale qui n'explique qu'une faible part de la variance du score au NASA-TLX. Nous en concluons que les utilisateurs d'une interface ne s'appuient que peu et sur des représentations déformées de l'utilisabilité, pour juger de l'acceptabilité.

8.1. Introduction

Les précédentes études nous ont permis de dégager un certain nombre d'enseignements quant à l'acceptabilité de DIGIME. Il apparaît que les améliorations apportées à l'ergonomie de l'interface, qui conditionnent son utilisabilité, en interaction avec l'utilisateur et ses connaissances, n'aient pas suffit à son acceptabilité. Des modèles bien établis, tels que celui de Nielsen (1993) ou la norme ISO 9126-1 (2000) nous le confirment. D'autre part, il nous semble que les méthodes que nous avons utilisées se sont heurtées à plusieurs obstacles dans l'amélioration itérative de DIGIME : jamais toutes les conditions nécessaires à l'émergence de résultats nets n'ont été réunies. En effet, aux problèmes d'échantillonnage s'est ajoutée l'impossibilité, pour les participants, de comparer les interfaces entre elles (Fiorello, 1999; Wiedenbeck & Davis, 1997). De surcroît, la dernière expérience a montré qu'il était nécessaire de prendre en compte les performances pour comprendre l'acceptation.

Les modèles qui prédisent l'acceptation des technologies identifient principalement des déterminants sociaux ou psychosociaux (croyances, attitudes, normes subjectives) à l'intention d'utilisation (Ajzen, 1985; Davis, Bagozzi, & Warshaw, 1989; Dillon & Morris, 1997; Fishbein & Ajzen, 1975; King & He, 2006; Legris, Ingham, & Collerette, 2003; Rogers, 1995; Tornatzky & Klein, 1982; Venkatesh & Davis, 2000; Venkatesh, Morris, Davis, & Davis, 2003). Parmi ces déterminants, la facilité d'utilisation est une dimension clé de la compréhension de l'acceptabilité des systèmes, particulièrement au cours des premiers jours, sinon des premiers instants, lors de la prise en main d'un dispositif. Passé ce délai, l'utilité perçue devient le seul déterminant de l'intention d'utilisation, quelle explique largement, et ce, dès la prise en main (Davis, 1989; Davis, Bagozzi, & Warshaw, 1989; Morris & Dillon, 1997). Pour l'instant, rien n'autorise à penser que la facilité d'utilisation perçue est assimilable à l'utilisabilité, fût-elle perçue (Dillon & Morris, 1997; Frokjaer, Hertzum, & Hornbaek, 2000; Nielsen & Levy, 1994; Yeh & Wickens, 1984). Puisque la facilité d'utilisation perçue se mesure en tant que telle, elle n'inclue aucune mesure d'utilisabilité. Franchir ce pas de lier des mesures d'utilisabilité à l'intention d'utilisation pourrait permettre de mieux comprendre dans quelle mesure l'utilisabilité influence l'intention d'utilisation d'une part, et d'autre part d'être capable d'identifier des causes possibles de problèmes d'acceptabilité pendant la phase de conception des logiciels (Arning & Ziefle, 2007; Dillon & Morris, 1997; Fiorello, 1999; Hasan & Ahmed, 2007; Jamet & Février, 2008; Zhang & Dillon, 2003). Pour l'instant seules quelques tentatives de ce type ont

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

vu le jour (Arning & Ziefle, 2007; Fiorello, 1999; Hasan & Ahmed, 2007; Venkatesh & Davis, 1996). Une seule cependant a proposé de tester un modèle théorique unifiant les champs MIS et HCI : il s'agit de l'étude de Fiorello (1999) s'appuyant sur le modèle P3 (Dillon & Morris, 1998, 1999). Les autres études se sont limitées, dans le cadre du développement du TAM (Davis, 1989; Davis, Bagozzi, & Warshaw, 1989) à chercher dans des mesures d'utilisabilité d'éventuels antécédents aux dimensions de ce modèle (Arning & Ziefle, 2007; Hasan & Ahmed, 2007; Venkatesh & Davis, 1996). Le modèle P3 (Dillon & Morris, 1998, 1999) fait l'hypothèse que l'acceptation est explicable autant par des mesures subjectives (attitudinales) que des mesures objectives de performance ou d'utilité. Si ce modèle n'a pas reçu les validations empriques suffisantes à assurer sa diffusion (Fiorello, 1999), il n'en est pas moins vrai qu'il mérite que l'on s'y penche. Il est couramment admis que l'utilisabilité est un facteur d'acceptabilité (Arning & Ziefle, 2007; Eason, 1984; Hasan & Ahmed, 2007; ISO-9241-11, 1994; Jamet & Février, 2008; , sous presse; Keinonen, 1998; Nielsen, 1993; Rogers, 1995; Shackel, 1991; Tornatzky & Klein, 1982; Tricot & Tricot, 2000; Venkatesh & Davis, 1996; Zhang & Dillon, 2003). La question est de savoir quelles mesures d'utilisabilité, objectives et subjectives, sont susceptibles de prédire l'acceptation. Il est évident que les mesures préconisées par la norme ISO 9241-11 sont les plus facilement opérationnalisables, car centrées sur le produit. Les propositions de Nieslen (1993) et Shackel (1991) pour mesurer l'utilisabilité sont elles centrées sur la qualité de l'interaction, et un certain nombre de mesures sont assez difficiles à mettre en œuvre, notamment pour la gestion des erreurs. Les mesures de la norme ISO 9241-11 ont été utilisées avec succès dans plusieurs études (Arning & Ziefle, 2007; Költringer & Grechenig, 2004). Nous proposons les mesures suivantes :

- Achèvement : nombre de caractères écrits (espaces compris).
- Erreurs : nombre d'erreurs restant sur le texte produit.
- Effort mental moyen : temps de réaction moyen à une tâche seconde de détection auditive (procédure de Kellogg (1996).
- Utilisabilité subjective : mesurée par le NASA-TLX (Hart & Staveland, 1988), comme le suggère la norme ISO 9241-11. Certains considèrent le NASA-TLX comme une mesure d'utilisabilité subjective (Keinonen, 1998).

A partir de ces VD, il est possible de constituer d'autres mesures, composites, sur lesquelles elles vont se fonder. Pour ce faire, nous proposons de faire usage des formulations mathématiques qu'ont introduites récemment certains auteurs (Darabi, 2004; Huang, Hong, & Eades, 2006; Paas, Tuovinen, Tabbers, & Van Gerven, 2003; Paas & Van Merriënboer, 1993; Tuovinen & Paas, 2004).

On pourra notamment proposer une mesure d'utilisabilité s'appuyant sur les dimensions recommandées par la norme ISO 9241-11. En procédant de la sorte, nous définissons à l'avance la « composition » de l'utilisabilité :

- Efficacité (performance), $P = ACHEVEMENT - ERREURS$. C'est une mesure de quantité de travail sans erreurs.
- Efficience mentale, $E = (Z_{\text{Performance}} - Z_{\text{Effort}}) / \sqrt{2}$.
- Utilisabilité, $U = (Z_{\text{Performance}} + Z_{\text{Efficience}} + Z_{\text{Satisfaction}}) / \sqrt{3}$. Nous obtenons ainsi une mesure unidimensionnelle de l'utilisabilité qui prend en compte toutes les dimensions de son appréciation selon la norme ISO 9241-11. La mesure de satisfaction sera celle issue du NASA-TLX. Sachant que la norme ISO 9241-11 est très évasive sur la définition de la satisfaction et de son opérationnalisation, nous contenterons de cette unique mesure dont on a tout lieu de penser qu'elle constitue une mesure fiable et valide d'utilisabilité subjective. Elle prend en considération plusieurs aspects de l'interaction et du résultat obtenu, et sa notoriété n'est plus à faire. Par ailleurs, la norme ISO 9241-11 estime qu'une mesure de charge de tâche mentale peut constituer une mesure de satisfaction.

La notion de charge mentale est très largement utilisée en psychologie, et ce, dans des champs de recherche aussi divers que l'étude des EIAH (Jamet, 2006; Tricot & Chanquoy, 1996) ou celle des IHM (Bonnardel & Piolat, 2003; Tricot & Chanquoy, 1996). Ce concept de charge mentale pose deux problèmes principaux que sont sa définition et sa mesure, comme en témoignent les différents noms qui lui sont attribués, sans que l'on sache précisément s'ils renvoient ou non à la même idée : charge mentale (Paas, Tuovinen, Tabbers, & Van Gerven, 2003), charge de travail (Bastien & Scapin, 2001), ou charge de tâche (Hart & Staveland, 1988). Tricot et Chanquoy (1996) ont déjà repéré ce problème. Leur analyse de la question les incite à conclure que ce concept a toute sa place dans le cadre de la recherche sur les apprentissages telle que l'envisage Sweller et ses collaborateurs. Certains d'entre eux ont

apporté un soin particulier à définir le concept de la charge mentale (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). S'ils font la distinction entre charge cognitive, charge mentale et effort mental, les travaux en ergonomie sur la charge mentale de travail sont moins précis. Il est néanmoins possible de mesurer cette charge par le biais du NASA-TLX (Hart & Staveland, 1988), ainsi que par la procédure de double tâche (Bonnardel & Piolat, 2003; Piolat, Olive, Roussey, Thunin, & Ziegler, 1999). Cela permet d'en avoir deux mesures, l'une, subjective, l'autre, objective. Grâce au logiciel SCRIPTKELL (Piolat, Olive, Roussey, Thunin, & Ziegler, 1999), on peut déterminer la charge mentale en calculant l'effort mental moyen (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Cette méthode est cependant assez intrusive (Jamet, 2006). C'est pourquoi plusieurs recherches se sont attachées, avec succès, à démontrer sa relative innocuité sur la tâche principale, exception faite de la complexite syntaxique à laquelle nous ne portons pas d'intérêt (Brünken, Steinbacher, Plaas, & Leutner, 2002; Olive, Kellogg, & Piolat, 2002; Piolat, Kellogg, & Farioli, 2001; Piolat & Olive, 2000; Piolat, Roussey, Olive, & Farioli, 1996).

En ce qui concerne la charge mentale, on sait que chez le scripteur adulte, l'exécution motrice de l'écriture manuscrite cursive est un processus largement automatisé (Bourdin & Fayol, 1994; 2002). Modifier ce processus habituel entraîne une charge supplémentaire de la part du scripteur en termes d'effort mental, c'est-à-dire de ressources de MdT allouées à la réalisation de la programmation motrice du nouveau geste. On l'observe lorsque des individus adoptent l'écriture script (Olive & Kellogg, 2002). Ces résultats ont été mis en évidence grâce à la méthode de double tâche. Plus les ressources allouées à la tâche principale sont importantes, plus le temps de réaction à la tâche seconde augmente (Kellogg, 2001b, 2001b, 2004; Kellogg, Olive, & Piolat, 2007; Olive, 2004; Olive & Kellogg, 2002; Olive, Kellogg, & Piolat, 2002; Olive & Piolat, 2002, , 2005; Piolat, Kellogg, & Farioli, 2001; Piolat & Olive, 2000; Piolat, Olive, & Kellogg, 2005; Piolat, Olive, Roussey, Thunin, & Ziegler, 1999; Piolat, Roussey, & Barbier, 2003; Piolat, Roussey, Olive, & Farioli, 1996). DIGIME est un logiciel de reconnaissance de l'écriture manuscrite script. Quand bien même le logiciel reconnaît un certain nombre d'allographe pour chaque lettre (y compris des allographes de type cursif), la saisie se fait nécessairement en détachant chaque lettre (Anquetil & Bouteruche, 2004; Bouteruche, Deconde, Anquetil, & Jamet, 2005). On peut donc supposer que la saisie de texte sur DIGIME en écriture script fait augmenter l'effort mental nécessaire à sa réalisation. D'autre part, il est possible que le fait d'écrire avec un stylet sur un écran de PDA soit,

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

associé aux problèmes de reconnaissance des lettres, suffisant pour faire augmenter la charge mentale de façon importante (Dobbie & Askov, 1995; Favart & Olive, 2005). Nous pensons que l'augmentation de la charge mentale induite par l'exécution motrice du geste d'écriture avec DIGIME peut expliquer à elle seule la désaffection que cette interface suscite chez des utilisateurs novices. Afin de tester cette hypothèse, nous proposons de faire comparer aux sujets de l'expérience l'interface de saisie DIGIME à l'interface *Note*, qui ne consiste qu'en une feuille blanche électronique n'impliquant pas de reconnaissance (cf. expérience 3). Ainsi, on a l'opportunité de tester DIGIME et la charge que son utilisation fait peser sur le processus d'exécution motrice de l'écriture manuscrite par rapport à une situation d'écriture plus naturelle, si ce n'est le fait qu'elle se déroule sur PDA. En n'impliquant aucune reconnaissance, on fait de *Note* l'étalon d'une situation idéale de saisie d'écriture manuscrite sur PDA. Cette méthodologie permet aux sujets d'exprimer leur opinion sur DIGIME en comparaison de cette situation de référence, et réciproquement. Fiorello (1999) a utilisé une telle méthode et nous pensons cette démarche intéressante. Notre problématique ne tient pas compte ici des autres dimensions connues déterminant l'acceptation. On néglige volontairement l'utilité (Adams, Nelson, & Todd, 1992; Davis, 1989; Dillon, 2001b; Dillon & Morris, 1997; Tricot et al., 2003; Tricot & Tricot, 2000). L'intérêt de cet oubli volontaire est de permettre, en n'incluant pas l'utilité dans l'utilisabilité, de comparer DIGIME à *Note*. Ces interfaces n'ont pas la même utilité, même si leurs champs d'utilité possible peuvent se recouvrir en partie (Keinonen, 1998). Par conséquent, ce n'est pas tant l'intention d'utilisation de DIGME qui va nous intéresser que la différence d'intention d'utilisation qu'elle suscite par rapport à *Note*. C'est en cela que la méthode de Fiorello (1999) est novatrice. Les études d'acceptation testent le plus souvent, pour une interface donnée, un ensemble de prédicteurs de l'intention d'usage. La question même de l'acceptation en valeur absolue (c'est-à-dire de savoir si oui ou non le logiciel est accepté, n'y est jamais traitée). L'approche comparative que nous proposons permet de faire glisser la question de l'acceptation (comportement ou intention comportementale) vers celle de l'acceptabilité (attribut du logiciel par rapport à un autre, et qui, le cas échéant, peut conduire à son acceptation). Nous espérons de cette façon, contourner le problème que pose notre échantillonnage. C'est pourquoi les mesures que nous évoquions précédemment sont à considérer, non pas en tant que telles, mais en comparaison d'une interface à l'autre (pour peu que l'on contrôle l'ordre de présentation des interfaces (Wiedenbeck & Davis, 1997)).

Nous proposons de tester le potentiel prédictif du NASA-TLX (Hart & Staveland, 1988). Nous faisons l'hypothèse qu'il sera un prédicteur important de l'acceptabilité. De la même façon, nous faisons l'hypothèse, comme l'envisage le modèle P3, que les mesures d'utilisabilité réelle (c'est-à-dire d'efficacité et d'efficience) vont prédire l'acceptabilité de DIGIME. Nous faisons également l'hypothèse que les mesures subjectives d'utilisabilité (ici, le NASA-TLX) médiatisent une partie de l'effet prédictif des mesures de performances (notamment effort, efficacité et efficience). De fait, s'il importe de tenir compte de l'efficience, car la preformance vaut aussi par l'effort qui l'a engendrée (Paas, Tuovinen, Tabbers, & Van Gerven, 2003), rien n'indique que la première dépend du second. Savoir si l'effort mental peut expliquer l'efficacité figure donc parmi les questions que nous souhaitons traiter. Pour finir, nous pensons également que l'utilisabilité telle que nous comptons la calculer sera aussi prédictive de l'acceptabilité comparée de DIGIME.

8.2. Méthode

8.2.1. Participants

Soixante-douze personnes de 18 à 26 ans (20,5 ans en moyenne) ont participé à l'expérience. Il s'agissait d'étudiants en licence de psychologie à l'Université Rennes 2. Leur participation était parfaitement volontaire et n'a fait l'objet d'aucune rétribution de quelque nature que ce soit. Vingt personnes ont été affectée au groupe contrôle et 52 au groupe expérimental. Ce déséquilibre était volontaire et visait à maximiser la fiabilité des tests relatifs aux effets intragroupes du groupe expérimental. Ces participants étaient novices dans l'utilisation des PDA.

8.2.2. Matériel

8.2.2.1. Matériel informatique

Le matériel expérimental était composé de deux pocket PC Toshiba e400 sur lesquels était installé DIGIME, développé par l'IRISA, ainsi que *Note*. Nous avons aussi utilisé deux ordinateurs portables : l'un Acer Aspire 1600, processeur Intel Pentium 4 1,5 GHz, 256 Mo de RAM, écran 15 pouces ; l'autre Dell Inspiron 6000, processeur Intel Mobile 1,73 GHz, 512

Mo de RAM, écran 15 pouces sur lesquels étaient installé le logiciel SCRIPTKELL (Piolat & Olive, 2000; Piolat, Olive, Roussey, Thunin, & Ziegler, 1999). Ces machines étaient reliées chacune à un casque audio de façon à ce que les sujets puissent entendre les signaux sonores de la tâche seconde sans perturber l'environnement sonore.

8.2.2.2. Interfaces DIGIME et *Note*

L'interface DIGIME a subi une dernière transformation pour cette expérience. Les observations réalisées sur l'interface au cours des expériences précédentes nous ont inspiré des améliorations. Inspiré, car ces modifications ont été proposées, non pas spécifiquement à partir des mesures que nous avions réalisées, mais sur la base d'une inspection experte, c'est-à-dire sur les observations que nous avons faites des participants en train d'interagir avec DIGIME (Bevan, 1995; Bevan & Macleod, 1994; Dumas, 2003). C'est pourquoi des lignes ont été ajoutées à la zone de saisie, afin que les participants placent au mieux leurs caractères et gestes d'édition pour une plus grande efficacité. De même, ayant constaté la difficulté des utilisateurs de DIGIME à réaliser le geste d'édition « entrée » (retour chariot), nos partenaires de l'IRISA lui ont substitué un bouton représenté par un flèche coudée (comme sur les claviers matériels ou virtuels). Une autre flèche, dirigée vers la droite assurait également, de même que le bouton « entrée », l'envoi du texte saisie dans la zone application sans le faire suivre d'un retour à la ligne. Cela permettait de traiter plus facilement des corrections dans le traitement de texte sans avoir à subir le retour à la ligne.

L'interface *Note* est celle utilisée à l'expérience 3.

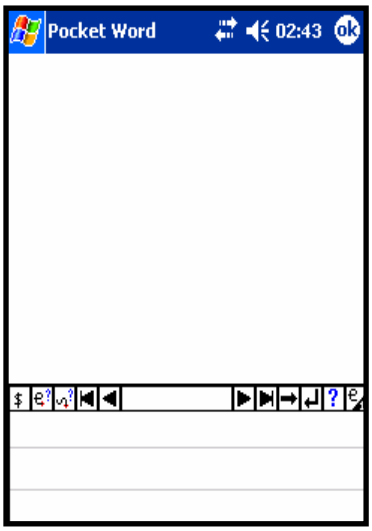


Figure 48: Dernière version de l'interface DIGIME

8.2.2.3. Questionnaires

Nous avons utilisé deux questionnaires : le premier était un pré questionnaire et avait pour fonction de déterminer le niveau de connaissances préalables des sujets en informatique ainsi que dans l'utilisation des PDA. Le second, post test, était le NASA-TLX. La charge de tâche mentale y est calculée de 0 (faible charge) à 20 (forte charge). Une question d'intention d'utilisation a été ajoutée également. Les réponses à cette question étaient collectées sur une échelle type Lykert à 5 points, graduée de -2 (aucune intention d'utilisation) à +2 (forte intention d'utilisation).

Ces questionnaires ont été dupliqués afin que les sujets donnent leur avis sur DIGIME et sur *Note*. L'ordre de présentation de chaque duplicata de ces questionnaires n'était pas contrebalancé et les sujets les remplissaient en fonction de l'ordre de présentation des interfaces durant la passation de l'expérience (qui lui était contrebalancé). Un dernier item évaluait quelle interface de saisie était préférée par les participants, justification à l'appui.

8.2.2.4. Textes

Nous avons utilisé deux textes différents. Issus de l'encyclopédie libre et en ligne Wikipédia, ils traitaient respectivement pour l'un des chiens, et pour l'autre des chats. Si ce choix peut sembler surprenant, il a été guidé par le souci de ne pas désintéresser les participants de la tâche. La saisie avec *Note* étant très rapide, il était difficilement

envisageable de créer du texte à copier pour une durée de dix minutes uniquement à partir de rendez-vous, d'URL ou de numéros de téléphone. Nous avons donc choisi des textes thématiques très généraux, assez longs pour permettre dix minutes de copie ininterrompues avec *Note* (voir en annexe, section 15.7).

8.2.3. Procédure expérimentale

Les particpants remplissaient le pré questionnaire. A la suite de cette première étape, ils se pliaient à la tâche d'enregistrement du niveau de base de leurs temps de réaction. Pour ce faire, après une brève explication les invitant à la plus grande rapidité, ils plaçaient le casque audio sur leurs oreilles et attendaient la présentation aléatoire de vingt-cinq signaux sonores (5 à 15 secondes de temps intermédiaire entre chaque signal, pour un intervalle moyen de 10 secondes). Ils devaient, dès qu'ils entendaient le signal, appuyer sur la touche « espace » du clavier de l'ordinateur portable à côté duquel ils se trouvaient. Les ordinateurs étaient placés de telle façon que le clavier était à portée immédiate de leur main non dominante, elle-même placée à côté du PDA sur lequel les participants devraient copier les textes proposés. De ces vingt-cinq mesures de temps de réaction, seules les vingt dernières ont été conservées pour déterminer le niveau de base. Ces paramètres sont ceux utilisés par Olive et Piolat (2002).

Le déroulement de la tâche était ensuite exposé par l'expérimentateur, notamment en ce qui concernait la rétrospection dirigée. Cinq actions étaient catégorisées : écrire ou corriger sur le PDA, lire le texte à copier, relire le texte saisi, chercher de l'aide et autre (pensées parasite). Néanmoins, les sujets pouvaient évoquer assez librement leur activité au moment de la survenue du signal sonore. Cette retrospection ne visait pas à déterminer la charge induite par tel ou tel processus comme c'est le cas dans les travaux de Kellogg et ses collaborateurs (Olive, Kellogg, & Piolat, 2002; Piolat, Kellogg, & Farioli, 2001; Piolat & Olive, 2000; Piolat, Roussey, Olive, & Farioli, 1996). Le protocole que nous avons mis en place, inspiré par celui d'Olive et Piolat (2002), nous assurait que la tâche n'impliquait que de l'exécution motrice. La rétrospection dirigée nous a permis de retirer les TR associés à la lecture du texte à copier ou toute autre activité qui ne procédait pas de la copie (lecture du texte, distraction, et consultation des aides en lignes). La rétrospection dirigée se faisait à l'oral, et était prise en note par l'expérimentateur après chaque appui. Cela ne concernait bien sûr que les sujets de la

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

condition expérimentale dont la tâche de copie était interrompue par la triple tâche. Pour le groupe contrôle, seul le niveau de base a été enregistré.

L'expérimentateur présentait ensuite aux participants, selon une procédure standardisée, comment utiliser l'interface de saisie qu'ils allaient avoir à utiliser en premier (DIGIME ou *Note*, une procédure de contrebalancement faisant en sorte qu'autant de sujets commencent l'expérience avec l'une ou l'autre des interfaces, de même que pour les textes à copier). Pour ce faire, l'expérimentateur, à l'aide d'un PDA faisait une démonstration d'utilisation sous les yeux des participants. Ainsi, ils prennaient connaissance des fonctions de base nécessaires à l'accomplissement de l'épreuve. Nous avons fait le choix de cet apprentissage guidé des interfaces pour des questions de temps, et pour que les sujets ne passent pas l'essentiel des dix minutes de copie à consulter les aides. Commençait alors l'épreuve (tâche de copie). Les participants étaient informés qu'il fallait copier le texte à l'identique, en respectant l'orthographe ainsi que la présentation générale (ex : retours à la ligne). La consigne les enjoignait à faire le plus et le mieux possible. Cette épreuve était entrecoupée par la présentation aléatoire de signaux sonores toutes les 15 à 45 secondes (paramètres Olive et Piolat (2002)). Entre 19 et 21 « bips » ont été émis au cours de l'épreuve. Un délai maximum de 3 secondes était laissé aux participants pour appuyer sur la touche « espace ». Passé ce délai, la réaction était considérée comme oubliée, et le temps de réaction n'était pas enregistré.

A la fin de ces 10 minutes de copie, l'expérimentateur laissait trois minutes de répit aux sujets avant de reprendre le déroulement de l'expérience, calqué sur la procédure évoquée ci-dessus, mais avec l'autre logiciel.

Pour terminer, les participants remplissaient les deux questionnaires post test, relatifs à chacune des méthodes de saisie qu'ils avaient évaluées. Il était explicitement dit aux sujets que si ces questions étaient données en toute fin de passation, c'était pour leur permettre de comparer les interfaces entre elles. Nous espérions ainsi limiter, grâce au contrebalencement, tout effet d'interaction entre le type d'interface (plus ou moins facile) et son odre de présentation (Wiedenbeck & Davis, 1997).

L'expérience durait une heure.

8.2.4. Variables indépendantes

Nous avons manipulé une première VI intergroupe à deux modalités (condition contrôle et condition expérimentale). Il s'agissait de l'utilisation de la méthode de triple tâche grâce au logiciel SCRIPTKELL. L'utilité de la condition contrôle était de s'assurer que l'ajout de deux tâches à la tâche principale n'affectait pas significativement cette dernière.

La seconde VI est intragroupe à deux modalités. Il s'agit de l'interface : DIGIME ou *Note*.

8.2.5. Variables dépendantes

Ci-dessous, la liste des VD enregistrées au cours de l'expérience :

- Achèvement : nombre de caractères écrits (espaces compris).
- Erreurs : nombre d'erreurs restant sur le texte produit.
- Effort mental : temps de réaction moyen à une tâche seconde de détection auditive. Grâce à la rétrospection dirigée, cette charge n'a concerné que l'exécution motrice de la copie (écriture et correction). L'effort mental est calculé de la façon suivante : $EM = TR_{Moyen} - TR_{de\ base}$. Le TR_{Moyen} n'a aucun sens en dehors de la comparaison à son niveau de base, c'est-à-dire sans tâche concurrente, puisqu'il est inhérent à la capacité limitée de la MdT (Baddeley, 1986; McCutchen, 1996; Miyake & Shah, 1999).
- Charge de tâche mentale : mesurée par le NASA-TLX (Hart & Staveland, 1988), elle pouvait également, selon la nomre ISO 9241-11, constituer une mesure de satisfaction. Nous avons choisi la traiter comme telle en tant que mesure d'utilisabilité perçue.
- Intention d'utilisation.

A partir de ces VD, il a été possible de créer d'autres mesures en utilisant notamment la formule de Huang, Hong et Eades (2006) :

- Efficacité/Performance, $P = ACHEVEMENT - ERREURS$. Cette mesure d'efficacité était également une mesure d'efficience temporelle, puisque la tâche de copie se déroulait sur une durée prédéfinie (10 mn). Il ne nous est pas possible de produire

une mesure d'efficacité telle que proposée par Bevan et Macleod (1994, pour mémoire Efficacité (%) = 1/100 × (exactitude × achèvement)). En effet, cette formule préconise de calculer le produit de l'achèvement (en pourcentage du travail achevé) par l'exactitude (en pourcentage d'erreurs). Or, puisque la tâche de copie confiée aux sujets se déroulait sur une durée finie, il n'était pas possible d'en calculer le pourcentage d'achèvement, d'où l'impossibilité de calculer l'efficacité de cette manière.

- Efficience mentale, $E = (Z_{\text{Performance}} - Z_{\text{Effort}}) / \sqrt{2}$.
- Productivité/Rendement R = Efficacité / Effort. Il s'agit d'une mesure d'efficience calculée sur les scores bruts (non standardisés) afin de permettre des comparaiasons entre interfaces, ce que ne permettent pas les scores Z dont la moyenne est par définition 0.
- Utilisabilité, $U = (Z_{\text{Performance}} + Z_{\text{Efficience}} + Z_{\text{Satisfaction}}) / \sqrt{3}$.

Toutes ces mesures ont été triplées, puisqu'elles étaient valables pour DIGIME, *Note*, ainsi que la différence entre les deux Δ = DIGIME − *Note*. Ces dernières mesures, basées sur la comparaison entre DIGIME et *Note* ont eu pour nous le plus grand intérêt puisque ce protocole expérimental avait pour principal atout de permettre aux participants de comparer les interfaces entre elles, ce qui faisait défaut à nos précédentes manipulations.

8.2.6. Hypothèses

Une hypothèse portait sur l'effort mental inhérent à l'utilisation de l'une ou l'autre des interfaces.

- H1: L'effort mental et la charge de tâche mentale engendrés par la copie sur DIGIME seront plus important que sur *Note*. Réciproquement, l'efficacité et l'efficience (productivité/rendement) seront plus faibles avec DIGIME.

Une seconde hypothèse portait sur l'innocuité de la triple tâche sur la tâche principale. C'est la raison d'être de notre groupe contrôle.

- H2: La triple tâche aura un effet néfaste sur les performances à la tâche principale de copie et sur les jugements ressentis.

La série d'hypothèses suivantes portaient sur des effets intragroupes, et relevaient du modèle page suivante, très inspiré du modèle P3 (Dillon & Morris, 1998, 1999), à ceci près qu'il ne tenait pas compte de l'utilité (réelle ou perçue) :

- H3: L'effort mental influence positivement l'efficacité.
- H4: L'effort mental et l'efficacité influencent positivement la charge de tâche mentale.
- H5: L'efficience influence positivement la charge de tâche mentale.
- H6: La charge de tâche mentale influence positivement l'acceptabilité comparée.
- H7: L'utilisabilité (*U*) influence positivement l'acceptabilité comparée.
- H8: L'effort mental influence positivement l'acceptabilité comparée.
- H9: L'efficience influence positivement l'acceptabilité comparée.
- H11: L'efficacité influence positivement l'acceptabilité comparée.

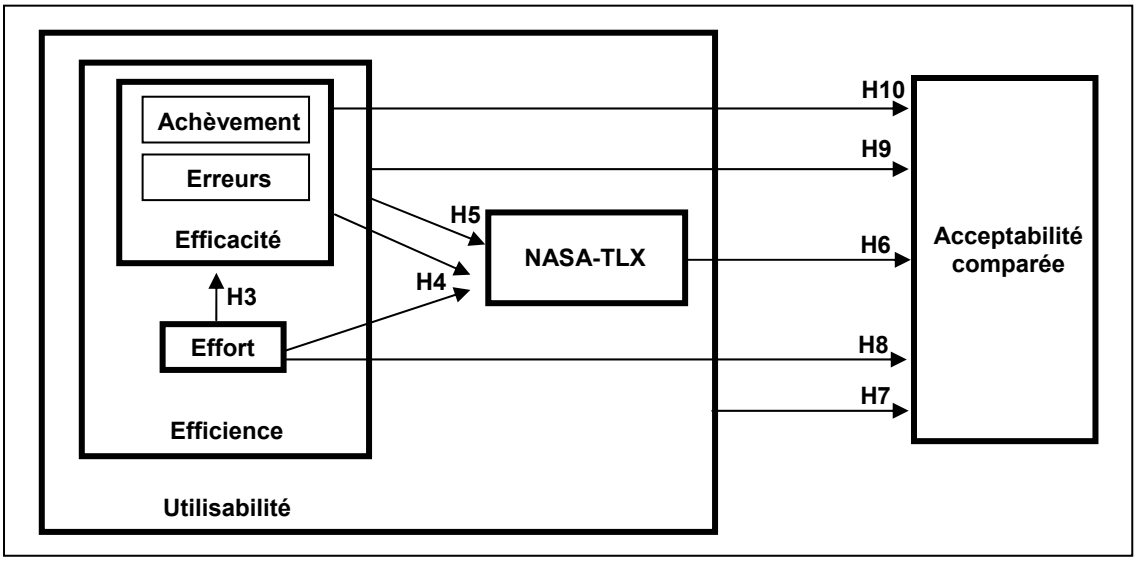


Figure 49: Modèle de prédiction de l'intention d'utilisation par des mesures objectives et subjectives d'utilisabilité

8.3. Résultats

8.3.1. Effets des interfaces sur les VD (H1)

En ce qui concerne le logiciel préféré, 11 participants sur les 72 (6 sur les 52 du groupe expérimental) ont rapporté préférer DIGME à *Note*. Ce choix a toujours été justifié par l'utilité que les participants attribuaient à DIGIME pour remplir des champs ou communiquer des informations à un logiciel (saisie proprement dite plutôt qu'écriture). Les sujets qui ont dit préférer *Note* l'ont choisi pour des raisons de simplicité d'utilisation. Les résultats montrent effectivement des différences notables d'une interface à l'autre. Les principaux résultats sont présentés dans le tableau 10, ci-dessous.

	DIGIME	*Note*
Effort mental (s)	608 (189)	451 (163)
Efficacité (caractères saisis)	90,5 (43,5)	758,8 (140,8)
Charge de tâche (de 0 à 20)	13,5 (3,2)	8,8 (3,1)
Productivité	.16 (.10)	2,03 (1,17)
Intention d'utilisation*	2,0 (3,0)	4,0 (1,7)

Note : Moyenne (écart-type) sauf * Médianne (intervalle interquartile)

Tableau 10: Principales mesures d'utilisabilité en fonction des interfaces

Les comparaisons intragroupes sur ces VD ont témoigné de différences significatives. Les distributions étant normales, nous leur appliquons le *t* de student pour échantillons appariés (sauf pour l'intention d'utilisation, mesurée sur une échelle ordinale, et par conséquent traitée avec le test de Wilcoxon, ainsi que pour la productivité, ne souscrivant pas aux conditions d'application d'un test paramétrique). Comme nos hypothèses l'avaient envisagé, les mesures réalisées sur les participants diffèrent en fonction de l'interface d'écriture, et ce, autant pour l'effort mental ($t(48) = 5{,}52$, $p = .00$), la charge de tâche mentale ($t(51) = 8{,}53$, $p = .00$), la productivité ($Z = -5{,}90$, $p = .00$), et l'intention d'utilisation ($Z= -4{,}047$, $p =.00$). Les chiffres de l'efficacité parlent d'eux-mêmes.

8.3.2. Effets des tâches concurrentes sur la tâche principale et sur les jugements (H2)

Un grand nombre de VD ne présentant pas une distribution normale, nous les avons analysées avec le test non paramétrique de Mann-Whitney. Pour limiter le nombre d'analyses, nous avons uniquement traitées les mesures Δ puisqu'elles résumaient à la fois celles de DIGIME et celles de *Note*. Par ailleurs, il s'agissait des plus importantes pour cette étude. Les scores obtenus par groupe pour les différentes VD sont consignés dans le tableau 11.

Mesures Δ	Groupe Contrôle	Groupe Expérimental
Intention d'utilisation	-.50 (3,50)	-1,0 (3,00)
NASA-TLX	3,80 (3,96)	4,50 (3,79)
Efficacité (Nbre caractères justes)	659 (127)	668 (148)
Temps de réaction de base (ms)	607 (103)	578 (109)

Note : Partie supérieure médiane (intervalle interquartile). Partie inférieure moyenne (écart-type).

Tableau 11: Scores aux VD pour chaque groupe

	U	*Z*	*p*
Intention d'utilisation	403,5	-1,27	.20
NASA-TLX	383,5	-1,43	.15
Efficacité	345,5	-.93	.35
Temps de réaction de base	477	-.541	.58

Tableau 12: Résultats du test de Mann-Whitney appliqué à la comparaison entre groupe contrôle et groupe expérimental

Les résultats ont montré que le groupe contrôle et le groupe expérimental ne présentaient aucune différence significative, comme le montre le tableau 12, ci-dessus.

8.3.3. Effets des mesures de performances sur la charge de tâche mentale (H4, H5 et H6)

Les tableaux 13, 14 et 15, ci-après, présentent les corrélations entre les différentes VI. Toutes les échelles de mesures ont été orientées suivant la même direction. Par conséquent, lorsque les corrélations sont négatives, il existe bien une relation inverse entre les variables.

Mesures Δ	NASA-TLX	Efficacité	Effort	Efficience	Utilisabilité
NASA-TLX	_				
Efficacité	.31*	_			
Effort mental	.12	-.11	_		
Efficience	.33*	.59**	.67**	_	
Utilisabilité	.69**	.77**	.35*	.82**	_

Note : rho de Spearman, corrélations bivariées, ** $p \leq .01$, * $p \leq .05$

Tableau 13: Corrélations entre les VI comparées de l'analyse

Mesures DIGIME	NASA-TLX	Efficacité	Effort	Efficience	Utilisabilité
NASA-TLX	_				
Efficacité	.36**	_			
Effort mental	.09	.11	_		
Efficience	.32*	.76**	.68**	_	
Utilisabilité	.64**	.87**	.39**	.86**	_

Note : rho de Spearman, corrélations bivariées, ** $p \leq .01$, * $p \leq .05$, N=52

Tableau 14: Corrélations entre les VI de l'analyse pour DIGIME

Mesures *Note*	NASA-TLX	Efficacité	Effort	Efficience	Utilisabilité
NASA-TLX	_				
Efficacité	.33*	_			
Effort mental	-.17	.10	_		
Efficience	.17	.67**	.76**	_	
Utilisabilité	.62**	.84**	.31*	.78**	_

Note : rho de Spearman, corrélations bivariées, ** $p \leq .01$, * $p \leq .05$, N=52

Tableau 15: Corrélations entre les VI de l'analyse pour *Note*

L'examen de ces tableaux a révèlé certaines corrélations significatives. Il n'a pas été surprenant de constater que l'utilisabilité était corrélée à l'efficience, l'efficacité, l'effort et le NASA-TLX puisqu'elle a été calculée à partir de ces mesures. De même, les corrélations fortes entre efficicence d'une part, et efficacité et effort mental d'autre part, peuvent être expliquées par les mêmes raisons.

Plus intéressant est l'absence de corrélation entre effort mental et NASA-TLX, alors qu'à l'inverse, l'efficacité, et, dans une moindre mesure, l'efficience sont liées à la charge de tâche mentale.

Nous procédons à plusieurs régressions linéaires, simples et multiples (introduction simultanée), en vue de tester nos hypothèses. La distribution des résidus a, à plusieurs reprises, violé les conditions de normalité et d'homoscédasticité de la variance nécessaires à la conduite d'une analyse de régression. Nous avons cependant pris le risque de poursuivre notre analyse sans en tenir compte, misant sur la robustesse de la régression linéaire (Howell, 1997, p.587). Dans le cas de l'introduction simultanée de plusieurs VI, nous avons vérifié les statistiques de colinéarité qui ont donné lieu à des valeurs de tolérence et de facteur d'inflation de la variance acceptables.

Nous avons testé si l'effort mental moyen associé à la copie d'un texte permettait d'en prédire l'efficacité (H3). L'effet n'est pas significatif, comme le montre le tableau 16.

		Δ				DIGIME				*Note*			
VD	VI	R^2	β	*B*	*ES B*	R^2	β	*B*	*ES B*	R^2	β	*B*	*ES B*
Efficacité		-.01				.00				.00			
Effort mental			-.06	-.05	.12		.16	.03	.03		.19	.18	.13

Note: ** $p \leq .01$, * $p \leq .05$, R^2 est ajusté, *ES B* désigne l'erreur standard de *B, N=52*.

Tableau 16: Effet de l'effort mental sur l'efficacité

Nous avons ensuite testé si l'effort mental et l'efficacité permettaient de prédire la charge de tâche mentale (H4). La méthode d'introduction est simultanée. On constate, comme l'indique le tableau 17 (page suivante), que si l'efficacité a eu un potentiel prédicitif sur le score du NASA-TLX pour DIGIME ($sr^2 = .17$), ce seul effet n'a pas résisté à la comparaison avec *Note*, pour lequel il n'est pas apparu. Pour finir, nous vérifions si l'efficience a bien eu un effet sur la charge de tâche mentale (H5). Effectivement, il est apparu que l'efficience dans l'utilisation de DIGIME rend compte d'une part substantielle de la variance de la charge de tâche, de sorte que, si cet effet ne s'est pas observé avec *Note*, il se répercute sur la comparaison entre les deux interfaces.

		Δ				DIGIME				*Note*			
VD	VI	R^2	β	*B*	*ES B*	R^2	β	*B*	*ES B*	R^2	β	*B*	*ES B*
NASA-TLX		.05				.15				.03			
Effort mental			.17	.00	.00		.08	.00	.00		-.12	.00	.00
	Efficacité		.27	.00	.00		.42	.03	.01*		.25	.00	.00
		.06				.11				-.01			
	Efficience		.31	1,30	.60*		.36	1,15	.41*		.10	.31	.42

Note: ** $p \leq .01$, * $p \leq .05$, R^2 est ajusté, *ES B* désigne l'erreur standard de *B, N=52*.

Tableau 17: Effet de l'effort mental et de l'efficacité d'une part, de l'efficience d'autre part, sur la charge de tâche

8.3.4. Effets des mesures d'utilisabilité sur l'acceptabilité comparée (H6, H7, H8, H9 et H10)

Nous avons testé successivement les effets de nos différentes mesures d'utilisabilité sur l'acceptabilité des interfaces. Il s'agissait de l'effort mental, de l'efficacité, de l'efficience, de la charge de tâche mentale ainsi que de l'utilisabilité. Ces VI présentant des corrélations élevées (ou pour le moins, non négligeables), elles ont été introduites séparément. Les résultats obtenus sont consignés dans le tableau 18, page suivante.

La charge de tâche mentale a expliqué l'acceptabilité comparée (H6). Cet effet que l'on a également observé sur les mesures relevées avec *Note* n'est pas apparu avec DIGIME. Le NASA-TLX explique 22% de la variance de l'acceptabilité comparée, ce qui n'est pas négligeable.

La mesure d'utilisabilité que nous avons proposée n'a prédit, pour *Note* uniquement, que 10% de la variance d'intention d'utilisation, sans que cela ne se soit répercuté sur l'acceptabilité comparée (H7).

Aucune autre des VI que nous avons introduites dans la régression n'a expliqué l'acceptabilité des interfaces (H8, H9 et H10).

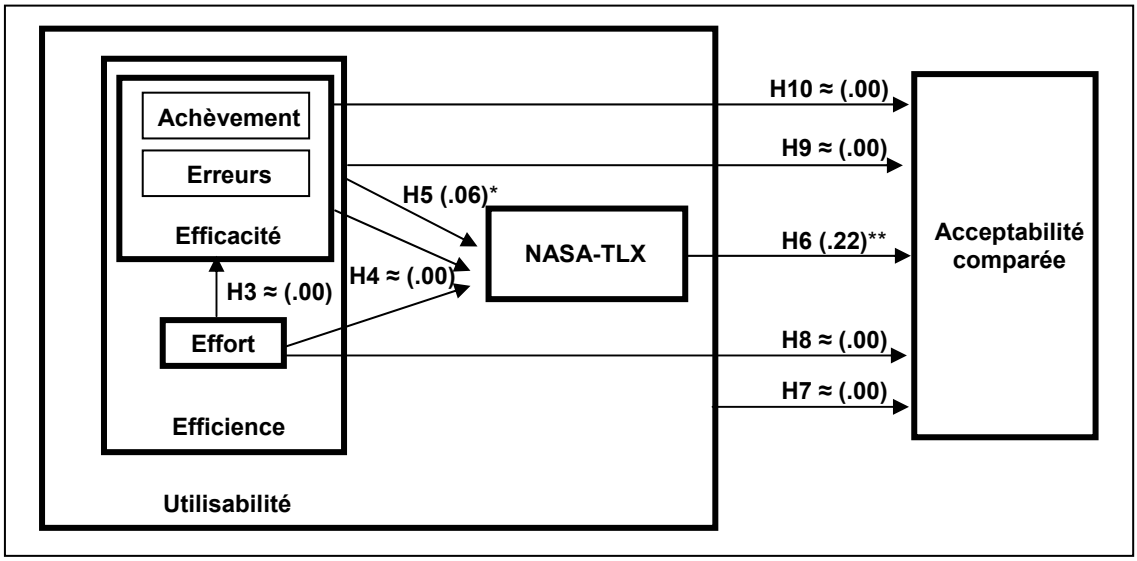
		Δ				DIGIME				*Note*		
VD	VI	R^2	β	*B*	*ES B*	R^2	β	*B*	*ES B*	R^2	β	*B*	*ES B*
Acceptabilité		.22				.01				.30			
	NASA-TLX		.48	.21	.05**		.17	.06	.05		.56	.25	.05**
		.04				.02				.10			
	Utilisabilité		.25	.36	.19		.21	.18	.12		.34	.37	.14*
		-.02				-.02				-.01			
	Effort mental		-.01	.00	.00		.00	.00	.00		-.07	.00	.00
		.02				.03				-.01			
	Efficience		.20	.35	.26		.22	.26	.16		.66	.09	.19
		.00				.03				.00			
	Efficacité		.14	.22	.22		.24	.00	.00		.14	.00	.00

Note: ** $p \leq .01$, * $p \leq .05$, R^2 est ajusté, *ES B* désigne l'erreur standard de *B, N=52*.

Tableau 18: Effet de la charge de tâche, de l'utilisabilité, de l'effort mental, de l'efficience et de l'efficacité sur l'acceptabilité

8.4. Discussion

Afin de faire un bref rappel des nombreuses hypothèses que nous avons soumises à l'expérimentation, la figure 50 résume graphiquement les principaux résultats obtenus.



Note: (R^2 ajusté) ** $p \leq .01$, * $p \leq .05$

Figure 50: Résumé graphique des hypothèses testées et des résultats obtenus sur les mesures Δ

Contrairement à nos attentes, l'effort mental qu'ont investi les participants dans la tâche n'a pas permis d'expliquer leur efficacité (H3). Il est en effet possible que les performances produites par les sujets dépendent d'autres processus cognitifs et que l'effort mental n'y compte que pour portion congrue.

Pourtant, l'effort investit dans l'utilisation de DIGIME a été significativement plus élevé que celui nécessité par *Note* (H1). D'autre part, la mise en relation de l'effort mental investit et de l'efficacité obtenue a permis d'expliquer une petite part de variance de la charge de tâche (pour DIGIME et en comparaison avec *Note*, H5). Il y a donc bien intérêt à prendre en compte cette mesure d'efficience (ISO-9241-11, 1998; Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Elle contient de fait d'autres informations que ce que les mesures sur lesquelles elle est calculée recèlent (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Preuve en est que l'effort mental associé à l'efficacité n'ont pas permis de rendre compte de la variance de la charge de tâche (H4).

Les résultats quant à notre capacité de prédire l'acceptabilité à partir de mesure d'utilisabilité sont assez mitigés puisque seule la charge de tâche mentale y est parvenue (H6). Cet effet, qui ne s'est observé que pour *Note* et la comparaison avec DIGIME, représente tout de même un R^2 de .22, ce qui est comparable aux valeurs obtenues dans le cadre du TAM lorsque la facilité d'utilisation perçue est testée sur les attitudes ou que ces dernières sont testées sur l'intention d'utilisation (Adams, Nelson, & Todd, 1992; Davis, Bagozzi, & Warshaw, 1989; Morris & Dillon, 1997). Ce résultat n'est donc pas surprenant si l'on accepte l'idée que la charge de tâche mentale, constituant une mesure de satisfaction ou d'utilisabilité perçue (ISO-9241-11, 1994; Keinonen, 1998) représente incidemment une mesure attitudinale (Channouf, Py, & Somat, 1996; Davis, 1989; Fishbein & Ajzen, 1975).

L'utilisabilité n'a pas été en mesure de prédire l'acceptabilité (exception faite de l'acceptabilité de *Note*, H7). De toute évidence, quand bien même l'a-t-elle pu, cela n'a été que grâce à la contribution de la satisfaction dans son calcul global puisque aucune autre des mesures d'utilisabilité objective (effort, efficacité et efficience) n'y sont parvenues (H8, H9 et H10). Il apparaît donc que si des facteurs externes peuvent influer directement sur l'intention d'usage (Burton-Jones & Hubona, 2006; McFarland & Hamilton, 2006), ces mesures de performance y ont échoué.

Plusieurs enseignements sont à tirer de ces résultats. Sur la validation du modèle P3 (Dillon & Morris, 1998, 1999) tout d'abord. Il semble que l'utilisabilité ne puisse pas à elle

seule expliquer l'intention d'utilisation. Quand bien même le peut-elle, c'est dans de si faibles proportions que cela ne présente que peu d'intérêt. En effet, la contribution d'une mesure d'utilisabilité subjective explique à elle seule un cinquième de l'intention comportementale. En l'incluant dans une mesure globale d'utilisabilité contenant également des mesures objectives de performances, on ne fait que diminuer son potentiel prédictif. Par conséquent, les mesures attitudinales sont confirmées dans leur rôle connu de meilleurs prédicteurs des intentions comportementales (Channouf, Py, & Somat, 1996; Davis, 1989; Fishbein & Ajzen, 1975). Par contre, il existe bien un lien de causalité entre les jugements portés sur l'interface et l'efficience. A ce titre, notre étude confirme les permiers travaux sur la question (Arning & Ziefle, 2007; Hasan & Ahmed, 2007; Venkatesh & Davis, 1996). Bien que cet effet existe, il est néanmoins très faible ($R^2 = .06$). Sur les plans cognitif et décisionnel, dans notre démarche d'unification des champs et HCI et MIS (Zhang & Dillon, 2003), de nouvelles perspectives se font jour.

Il est évident que les mesures objectives ne peuvent pas en elles-même expliquer un comportement ou son intention. C'est par le filtre cognitif qu'elles sont traitées et interprétées pour être transformées en représentations. Par conséquent, si elles ne permettent pas d'expliquer l'intention d'usage, c'est que les utilisateurs ne s'appuient pas sur elles pour formuler leurs jugements, ou bien c'est qu'ils s'en font une représentation assez éloignée de ce qu'elles sont vraiment. Une dernière possibilité est que les représentations qui sous-tendent l'acceptation intègrent, en plus des représentations de la performance, d'autres critères. La contribution de la performance serait ainsi complètement diluée. La capacité de l'efficience à expliquer une petite part de la variance de la charge de tâche nous incite plutôt à envisager les deux dernières explications. Si effectivement, comme le suggèrent les résultats, l'utilisabilité perçue (ici, par le biais du NASA-TLX) peut se substituer à la facilité d'utilisation perçue du TAM, on sait que l'utilité perçue en médiatise une large part. Par conséquent, si la charge de tâche médiatise elle-même la performance, cette dernière est nécessairement transformée au cours de la construction de ces représentations mentales. Les représentations des performances subiraient donc des modifications dans la « chaine causale décisionnelle » interne à un individu. Elles subiraient aussi le phénomène de dilution que nous évoquions précédemment. Les différents items du NASA-TLX font que cette mesure subjective prédictive de l'acceptation s'appuie autant sur des représentations de la performance et de l'effort mental que de la fatigue physique ou de la frustration. Yeh et Wickens (1984) ont déjà

pointé le fait que la dissociation entre mesures objectives de performances et mesures subjectives puisse conduire à des erreurs d'appréciation dans l'évaluation d'IHM par des utilisateurs. Bien qu'il n'y ait aucun doute sur le fait que l'intention d'utilisation d'un dispositif s'appuie sur des processus sociaux et psychosociaux (Ajzen, 1985; Davis, Bagozzi, & Warshaw, 1989; Dillon & Morris, 1997; Fishbein & Ajzen, 1975; King & He, 2006; Legris, Ingham, & Collerette, 2003; Rogers, 1995; Tornatzky & Klein, 1982; Venkatesh & Davis, 2000; Venkatesh, Morris, Davis, & Davis, 2003), la qualité de l'interaction entre l'humain et la machine est aussi déterminante (Shackel, 1991; Nielsen, 1993; ISO-9241-11, 1994). On peut donc s'interroger sur la relative difficulté des études que nous avons évoquées (et auxquelles cette expérience s'associe) à le mettre en évidence (Arning & Ziefle, 2007; Fiorello, 1999; Hasan & Ahmed, 2007; Venkatesh & Davis, 1996). Cela tient, selon nous, à l'opérationnalisation de l'utilisabilité. Dès notre introduction théorique, en discutant les différents modèles qui la formalisent (Shackel, 1991; Nielsen, 1993; ISO-9241-11, 1994; ISO-9126-1), nous avons fait l'éloge du modèle de Nielsen en regard des autres. Cependant, nous avons alerté sur la difficulté à opérationnaliser certains des critères qu'il recommande de prendre en compte. Parmi eux, la gestion des erreurs. C'est une lacune importante de nos expériences, nous l'avons déjà souligné. Cela tient principalement à la difficulté de définir les erreurs, les catégoriser et les mesurer. Il est admis que les erreurs doivent être catégorisées en fonction du préjudice qu'elles engendrent (Nielsen, 1993; Keinonen, 1998). Cependant, si personne ne doute qu'une erreur est une action réalisée n'ayant pas donné le résultat escompté ou une action inapropriée, on peut également l'envisager comme une action moins bonne qu'une autre. Ces aspects ont été traités dans le cadre de la saisie de texte sur téléphone mobile par Soukoreff et MacKenzie (2002), mais deviennent infiniment plus complexes pour les interfaces stylo du fait des possibilités d'interaction qu'elles offrent. Souvenons-nous que, dans l'expérience 2, le didacticiel dynamique avait permis au groupe en bénéficiant d'utiliser significativement plus des stratégies de corrections évoluées (insertion de caractères). Dans ce contexte, ne pas utiliser l'insertion n'est pas une erreur en soit, mais elle peut constituer, sinon une perte de temps mesurable, une perte de confort pour l'utilisateur. Cela renvoie à des considérations sur les stratégies d'exploitation de l'interface qui laissent entendre qu'en dépit de nos précautions (adaptation du protocole d'Olive et Piolat, 2002), l'effort mental mesuré implique peut-être plus que la seule exécution motrice de l'écriture sur DIGME. Ceci nous amène donc à penser que le caractère opérationnalisable de la norme ISO 9241-11 lui font peut-être négliger des aspects centraux de l'interaction, sans que l'on ait la possibilité, à

l'heure actuelle, de faire mieux ou différemment. La conception d'Eason (1984) selon laquelle l'utilisation effective d'un logiciel est la meilleure preuve de son utilisabilité n'est donc pas soutenable en l'état. Cela signifie qu'un test utilisateur ne nous permet pas, dans l'état actuel de nos connaissances, d'anticiper l'acceptation future du dispositif évalué sur la base des performances relevées. Le rapprochement des communautés HCI et MIS, s'il est nécessaire, n'est donc pas encore entériné en un même modèle. Pour terminer sur ces considérations, notre volonté de négliger l'utilité s'est avéré être un très mauvais choix. Les avis recueillis auprès des participants pour justifier le choix de leur interface préférée ont très clairement indiqué des motivations en termes d'utilité et de facilité d'utilisation. Or, *Note* et DIGIME n'ont pas précisément la même utilité. Cela constitue vraisemblablement un biais dans nos interprétations par rapport à la méthode que nous avons choisie. D'autant plus qu'il est apparu à plusieurs reprises un effet modérateur du type d'interface sur les relations entre variables. Il faudra donc y remédier par des choix méthodologiques plus adaptés. Une comparaison entre DIGIME et le clavier virtuel pourrait à cet égard nous éclairer. Cet écueuil est d'autant plus génant qu'au terme de cette expérience, il ne nous est pas encore possible d'affirmer que l'utilisabilité (même perçue) est substituable à la facilité d'utilisation du TAM.

Sur le plan méthodologique, cette étude confirme l'intérêt de la procédure de Kellogg pour évaluer des IHM (Bonnardel & Piolat, 2003). C'est d'autant plus important que l'efficience a été la seule mesure objective à avoir un effet sur la satisfaction. Or c'est grâce à SCRIPTKELL (Piolat, Olive, Roussey, Thunin, & Ziegler, 1999) que nous l'avons mesurée. Il conviendra, à l'avenir de se pencher sur le fait que notre mesure d'efficience ait exercé un effet sur la charge de tâche alors que l'efficacité n'y a pas réussit. Rappelons que dans leur étude, Arning et Ziefle (2007) ont mis en évidence que l'efficacité (nombre de tâches accomplies) et l'efficience (durée de réalisation) influençaient la facilité d'utilisation perçue. Pourtant, notre mesure d'efficacité constituait aussi une mesure d'efficience temporelle puisqu'elle était mesurée pour une durée constante.

9- Expérience 5 bis : Utilisabilité objective et subjective : prédire l'acceptabilité comparée entre DIGIME et le clavier virtuel

---

Résumé – Nous réutilisons le protocole de l'expérience précédente. Les participants y comparent cette fois DIGIME au clavier virtuel, dans un souci de contrôler l'utilité théorique des interfaces. Cette recherche vise aussi à valider en partie de le modèle P3 (Dillon & Morris, 1998, 1999) en testant l'effet de mesures d'utilisabilité objectives (efficience) et subjectives (Test de Nielsen) sur l'acceptabilité mesurée par le TAM. Nous réutilisons la procédure de Kellogg afin de mesurer objectivement l'effort mental, et ainsi, l'efficience. Les résultats montrent que la facilité d'utilisation perçue est le meilleur prédicteur de l'acceptabilité comparée des interfaces, à utilité égale. Le test de Nielsen révèle que la facilité d'apprentissage rend compte de la facilité d'utilisation perçue. Ce résultat est interprété comme accréditant l'idée que, couramment diffusé, le clavier (ici virtuel) n'a pas à faire l'objet d'un apprentissage, contrairement à DIGIME. La découverte d'une interface nouvelle entrainerait donc une diminution mesurable de la performance. Pour autant, les résultats montrent que l'efficience n'explique aucun des jugements formulés.

9.1. Introduction

Plusieurs limites sont apparues à l'issue de l'expérience 5. En premier lieu, le fait de ne pas avoir tenu compte de l'utilité perçue des logiciels. Trois raisons justifient ce constat. Tout d'abord, il est largement admis que l'utilité perçue est le plus fort déterminant de l'intention d'utilisation dans le TAM (Adams, Nelson, & Todd, 1992; Arning & Ziefle, 2007; Davis, Bagozzi, & Warshaw, 1989; Davis & Venkatesh, 1996; Dillon & Morris, 1997; Morris & Dillon, 1997). C'est cependant à dessein que nous avions négligé sa contribution. Nous souhaitions évaluer comment l'exécution motrice d'écriture sur DIGIME pouvait être affectée par ce mode de saisie inhabituel et évaluer ses répercussions sur la charge mentale (c'est-à-dire l'effort mental moyen). C'est ce qui a justifié le choix d'une comparaison avec *Note*. D'autre part, l'utilité perçue médiatise une partie de l'influence de la facilité d'utilisation perçue. Cette seconde raison conditionne la dernière : si nous nous intéressons à l'utilisabilité en tant que substitut possible à la facilité perçue, nous ne pouvons pas faire l'économie d'étudier ses relations avec l'utilité (Adams, Nelson, & Todd, 1992; Arning & Ziefle, 2007; Davis, Bagozzi, & Warshaw, 1989; Davis & Venkatesh, 1996; Dillon & Morris, 1997; Morris & Dillon, 1997). Par conséquent, tester DIGIME en comparaison d'une interface présentant une utilité théorique identique est essentiel. Le clavier virtuel, de part son écrasante supériorité sur le marché concurrentiel de la saisie de texte sur PDA constitue sans aucun doute un point de comparaison idéal par rapport à notre interface. Cependant, bien que DIGIME et le clavier virtuel aient la même utilité théorique, ils n'ont sans doute pas la même utilité pratique (Nielsen, 1993), notamment par ce mécanisme d'influence de l'utilité perçue par la facilité d'utilisation perçue. Cette utilité théorique identique est un atout particulier pour une nouvelle tentative de validation du modèle P3 (Dillon & Morris, 1998; 1999) : en maintenant constant le « potentiel » (*Power*) des interfaces, on se dispense d'avoir à le mesurer objectivement, ce qui est difficile (Fiorello, 1999). Cette méthode permet donc de comparer des logiciels sur une base que l'on sait être égale.

C'est pourquoi nous proposons de réutiliser le protocole de l'expérience 5 presque à l'identique. La précaution méthodologique que nous avons prise pour nous assurer que la procédure de Kellogg (incluant la rétrospection dirigée) ne parasitait pas plus la réalisation de la tâche principale que les jugements produits, il ne nous semble pas nécessaire de la reproduire dans cette nouvelle étude. De nombreux travaux méthodologiques ont attestés, avant notre contribution, du bien-fondé de la procédure de Kellogg en tant que méthode peu

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

intrusive d'évaluation de la charge mentale (Olive, Kellogg, & Piolat, 2002; Piolat, Kellogg, & Farioli, 2001; Piolat & Olive, 2000; Piolat, Roussey, Olive, & Farioli, 1996). De surcroît, ayant supposé qu'il était possible que l'effort mental ne soit pas que la conséquence de la seule exécution motrice de l'écriture, mais aussi le fait de stratégies d'exploitation de l'interface, une seule double tâche sans rétrospection dirigée nous semble suffisante et appropriée, puisque nous n'avons pas abordé la question des stratégies d'utilisation des IHM. N'ayant pas d'hypothèses solides sur ce point, nous ne le traiterons pas, ou plus précisément, pas de cette façon.

En effet notre problématique s'articule donc autour de la mise en relation de mesures subjectives (TAM) et objectives (efficacité, efficience, et effort mental), tel que le recommande le modèle P3 (Dillon & Morris, 1998; 1999) pour prédire l'acceptation (ou l'acceptabilité suivant la nuance que nous avons introduite). S'il est difficile de catégoriser objectivement certains aspects de la qualité d'interaction entre un humain et une machine (*i.e.* erreurs, stratégies d'exploitation), il est possible de s'enquérir de l'avis des utilisateurs sur la question. A cet égard, le test d'utilisabilité subjective de Nielsen (1993) est tout indiqué. La question est de savoir s'il peut se substituer aux items de facilité d'utilisation perçue du TAM pour permettre aux concepteurs d'IHM, tout en conservant le potentiel prédictif de ce questionnaire, de mieux cerner les causes d'acceptation ou de rejet d'un système à partir de son utilisabilité perçue. L'utilisabilité peut aussi être évaluée objectivement. Qu'il s'agisse de la littérature (Arning & Ziefle, 2007; Hasan & Ahmed, 2007; Venkatesh & Davis, 1996) ou de l'expérience précédente, l'efficience fait figure de mesure privilégiée pour représenter l'utilisabilité et prédire l'intention d'utilisation (ou pour le moins, ses déterminants). Il semble assez intuitif que l'utilisabilité d'un dispositif se traduise en termes de productivité (ISO/IEC 9126-1) par la mise en relation de l'effort fourni et du résultat obtenu. L'efficience permet aussi, à l'instar de la mesure d'utilisabilité que nous avons proposée pour l'expérience précédente, d'unifier en une seule mesure plusieurs dimensions de la performance. Cela permet d'apporter une information nouvelle que ne contenaient pas initialement chacune des mesures la composant, prises séparément (Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

Lorsque l'on calcule l'efficience à partir de l'efficacité et de l'effort mental moyen, la mesure est parfaitement objective et ne fait appel à aucun jugement de la part de l'utilisateur. L'expérience 5 nous a appris que le niveau d'effort mental associé à l'utilisation des interfaces ne permettait pas d'anticiper les autres mesures que nous avons collectées :

efficacité, charge de tâche ou intention d'utilisation. Nous avons interprété ce résultat comme le fait que les utilisateurs ne s'appuyaient pas sur ce critère pour étayer leurs jugements. Ce peut être parce qu'ils n'ont pas accès à ce type d'informations internes à leur cognition, ou bien parce qu'ils s'en font une représentation erronée ou « diluée » parmi d'autres.

Nous proposons donc tester un modèle inspiré du modèle P3 (Dillon & Morris, 1998; 1999). Il ambitionne de répondre à la question suivante : peut-on considérer que l'utilisabilité (objective ou subjective) est assimilable à la facilité d'utilisation perçue ? Nous pensons que c'est le cas. Cette démarche vise deux objectifs. Sur le plan théorique, il s'agit de faire le point sur la notion d'utilisabilité qui, composite et justiciable de plusieurs définitions, est souvent assimilée à une forme de facilité d'utilisation (au sens commun du terme, sans faire référence au TAM). Sur le plan décisionnel, elle permettra de mieux connaître l'impact des mesures objectives et subjectives d'utilisabilité sur l'acceptabilité. On espère ainsi faire progresser le rapprochement des champs MIS et HCI (Zhang & Dillon, 2003). L'objectif de notre démarche est de s'inscrire dans ce courant récent qui ambitionne de déterminer comment le concepteur d'IHM peut anticiper l'acceptation du dispositif qu'il élabore à partir de mesures d'utilisabilité.

9.2. Méthode

9.2.1. Participants

Soixante étudiants en licence 1, 2, et 3 à l'université de Brest ont participé à cette expérience. Les résultats de trois d'entre eux se sont avérés complètement inexploitables du fait de problèmes informatiques. Restaient donc 56 personnes, parmi lesquelles 5 garçons et 51 filles d'un âge moyen de 21,4 ans. Leur participation était volontaire et non rémunérée. Tous étaient novices dans l'utilisation des PDA.

9.2.2. Matériel

9.2.2.1. Matériel informatique

Le matériel expérimental était identique à celui de l'expérience 5.

9.2.2.2. Interfaces DIGIME et clavier virtuel

L'interface DIGIME était la même que pour l'expérience 5. Le clavier virtuel était celui utilisé à l'expérience 3.

9.2.2.3. Questionnaires

Nous avons utilisé trois questionnaires et une question supplémentaire : le premier était un pré questionnaire. Identiques à ceux des autres expériences, il servait à détecter une éventuelle expertise des sujets dans la manipulation de PDA.

Le second, post test, était le test d'utilisabilité perçue de Nielsen (réponse sur échelles type Lykert à 5 points, graduées de -2 (aucune intention d'utilisation) à +2 (forte intention d'utilisation). Pour mémoire, il était composé des items suivants :

- Vous a-t-il semblé facile d'apprendre à utiliser le logiciel de reconnaissance ?
- Le logiciel de reconnaissance vous a-t-il semblé efficace ?
- Si vous deviez reproduire cet exercice dans une semaine seriez-vous capable d'atteindre le même niveau de performance ?
- Avez-vous eu du mal à gérer les erreurs, que ce soit celles du logiciel ou les vôtres ?
- Est-ce que le logiciel de reconnaissance vous a laissé une impression agréable ?

Le dernier questionnaire était le TAM. Nous avons utilisé la même version que pour les expériences 3 et 4, traduite par nos soins d'après une étude de Morris & Dillon (1997). Cette fois encore, les échelles de réponses contenaient sept points d'ancrage compris entre un pôle très positif et un pôle très négatif.

Ces questions étaient posées explicitement dans l'hypothèse selon laquelle les participants possederaient un PDA ou projetteraient d'en faire l'acquisition (conditionnel).

Comme précédemment, ces questionnaires ont été dupliqués afin de que les sujets donnent leur avis sur DIGIME et sur le clavier virtuel (selon les mêmes modalités que pour l'expérience 5, c'est-à-dire pas de contrebalancement de l'ordre de présentation des questionnaires). Un dernier item évaluait quelle interface de saisie était préférée par les participants, justification à l'appui.

9.2.2.4. Textes

Les textes utilisés sont les mêmes que pour la manipulation précédente.

9.2.3. Procédure expérimentale

Les sujets remplissaient le pré questionnaire et enregistraient le niveau de base de leurs temps de réaction.

Le déroulement de la tâche était ensuite exposé. Contrairement à l'expérience précédente, nous avons opté pour une double tâche, sans rétrospection dirigée, afin de prendre en compte la totalité de l'effort mental affecté à la réalisation de la tâche.

La découverte des interfaces était toujours guidée par l'expérimentateur selon une procédure standardisée et éprouvée afin d'éviter des erreurs dans la démonstration.

Les deux tâches de copie (clavier virtuel + DIGIME avec l'un ou l'autre des textes, contrebalancement complet) duraient également dix minutes chacune. Elles ont été entrecoupées par les signaux sonores de la double tâche. Les sujets devaient y réagir en appuyant sur la touche « espace » du clavier situé à portée immédiate de leur main non dominante (paramètres Olive et Piolat (2002) identiques à ceux de l'expérience précédente). La consigne insistait sur la qualité et la rapidité de la copie à réaliser.

L'expérience s'achevait par l'administration des questionnaires pour permettre aux participants de comparer les interfaces entre elles. L'expérience durait une heure.

9.2.4. Variables indépendantes

Confiants dans le caractère non intrusif de la procédure de Kellogg (Olive, Kellogg, & Piolat, 2002; Piolat, Kellogg, & Farioli, 2001; Piolat & Olive, 2000; Piolat, Roussey, Olive, & Farioli, 1996), nous ne réitérons pas la précaution méthodologique de l'expérience précédente, ce d'autant plus qu'elle se limite ici à une simple double tâche. Par conséquent, la seule VI manipulée est intragroupe à deux modalités. Il s'agit du type d'interface : DIGIME ou clavier virtuel.

9.2.5. Variables dépendantes

Ci-dessous, la liste des VD enregistrées au cours de l'expérience :

- Achèvement : nombre de caractères écrits (espaces compris).
- Erreurs : nombre d'erreurs restant sur le texte produit.
- Effort mental : temps de réaction moyen à une tâche seconde de détection auditive. L'effort mental était calculé de la façon suivante : $EM = TR_{Moyen} - TR_{de\ base}$.
- Utilisabilité perçue : mesurée par les items du test d'utilisabilité de Nielsen (1993).
- Scores aux dimensions du TAM. Une analyse factorielle nous a permis d'en valider les dimensions. Par la suite, nous avons travaillé sur les scores factoriels attachés à chaque dimension extraites de l'analyse.

A partir de ces VD, nous avons créé les mesures suivantes:

- Efficacité/Performance, $P = ACHEVEMENT - ERREURS$.
- Efficience mentale, $E = (Z_{\text{Performance}} - Z_{\text{Effort}}) / \sqrt{2}$.
- Productivité/Rendement, R = Efficacité / Effort.

Nous n'avons pas souhaité retravailler sur la mesure d'utilisabilité de la dernière expérience. Les résultats ayant montré qu'elle ne permettait de prédire l'acceptabilité que parce qu'une mesure de satisfaction y était incorporée, elle ne présentait aucun intérêt pour cette manipulation.

Toutes ces mesures ont été triplées. Elles s'appliquaient donc à DIGIME, au clavier virtuel (CV), et plus encore, à la différence entre les deux Δ = DIGIME − CV. Rappelons que nous avons utilisé ce protocole de comparaison des interfaces pour contourner les problèmes liés à l'échantillonnage de nos expériences.

9.2.6. Hypothèses

Cette hypothèse portait sur les effets de la VI intragroupe que nous avons manipulée, à savoir, le type d'interface :

- H1: Le clavier virtuel permet de meilleurs performances et jugements que DIGIME.

D'autres hypothèses portaient sur les effets intragroupes suivants, tels que présentés par la figure 51, ci-dessous. Elles se fondent sur les relations connues entre dimensions du TAM (utilité et facilité d'utilisation perçues), à ceci près que les attitudes y ont été retirées. Il s'agit d'un TAM légèrement modifié pour y inclure des mesures d'utilisabilité subjective (test de Nielsen) et objective (efficience) afin de s'approcher au mieux du modèle P3.

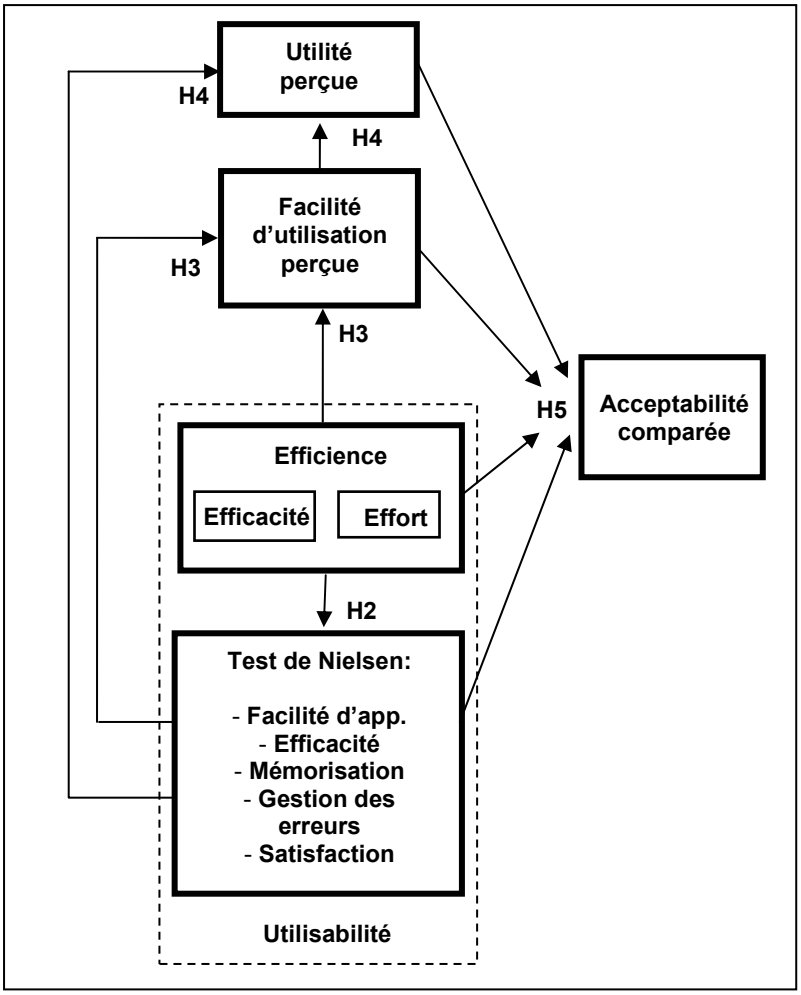


Figure 51: Modèle de prédiction de l'intention d'utilisation par des mesures objectives et subjectives d'utilisabilité

- H2: L'efficience influence positivement les scores au test de Nielsen.
- H3: L'efficience et les scores au test de Nielsen influencent positivement la facilité d'utilisation perçue.

- H4: Le test de Nielsen et la facilité d'utilisation perçue influencent positivement l'utilité perçue.
- H5: L'efficience, le test de Nielsen, la facilité d'utilisation et l'utilité perçues influencent positivement l'acceptabilité comparée.

9.3. Résultats

9.3.1. Analyse factorielle

Pour vérifier la structure factorielle des items du TAM, nous avons procédé à une analyse factorielle en choisissant la méthode des moindres carrés non pondérés puisque nos données s'y prêtaient (données ordinales, distribution non gaussienne) avec rotation OBLIMIN (Durand, 2005). La règle veut qu'il y ait minimum 5 sujets par items. Notre échantillon était trop petit pour. Cependant, tous les items du TAM ne nous intéressaient pas. Nous n'avions pas d'hypothèse sur la norme subjective, c'est pourquoi nous n'avons pas introduit dans l'analyse les items s'y rapportant. De même, nous n'avons pas tenu compte des attitudes dont le rôle est parfois discuté (Adams, Nelson, & Todd, 1992; Davis, Bagozzi, & Warshaw, 1989; Yang & Youngjin, 2004). Au final, nous avons introduit dix items relatifs à l'utilité perçue, à la facilité d'utilisation perçue et à l'intention d'utilisation.

Nous avons successivement conduit l'analyse sur les mesures Δ (auxquelles nous portons le plus grand intérêt puisque révélatrice de ce que nous avons appelé « l'acceptabilité comparée ». Nous avons cependant fait de même pour chaque logiciel, en dehors de toute comparaison (hormis le fait que la consigne associée aux questionnaires enjoignait les participants à comparer les interfaces entre elles).

9.3.1.1. Analyse factorielle sur les mesures Δ

Cette analyse n'a porté que sur les mesures Δ puisque c'était elles qui rendaient compte de la possibilité pour les participants de comparer les interfaces entre elles. Or, les particpants n'étant sans doute pas de potentiels acquéreurs de PDA, il était plus intéressant de se pencher sur ce que nous avons appelé « l'acceptabilité comparée » que sur l'intention d'utilisation au sens strict.

L'analyse a révélé un indice Kaiser-Meyer-Olkin (KMO) de .65. Il est admis que c'est une valeur limite[15] en termes de solution factorielle. Par ailleurs, la qualité de la représentation initiale s'est montrée bonne puisque les valeurs associées à nos items ont dépassé le seuil de rejet de l'analyse à .20.

Un rejet d'item a été justifé par un cas dit « Heywood » qui se produit quand un item présente une saturation factorielle supérieure à 1 (ce qui est théoriquement impossible). Ce type de résultat, dû à une trop forte corrélation entre cet item avec d'autres, impliquait son retrait. Il concernait l'item de facilité d'utilisation perçue du TAM « L'utilisation de ce logiciel est entièrement sous mon contrôle ».

Aucun item n'a présenté un caractère de complexité, c'est-à-dire une saturation supérieure à .30 pour plus d'un facteur.

La solution factorielle révélée par l'analyse s'est montrée juste sur le plan théorique. L'analyse a dégagé trois facteurs. Nous les avons désignés comme étant respectivement l'utilité perçue, la facilité d'utilisation perçue, et l'intention d'utilisation. La matrice des types partiellement reproduite ci-dessous fournit les saturations les plus importantes des items aux facteurs (>.50) associées au résultat de l'analyse de fiabilité (Alpha de Crombach) :

Facteur 1 : Facilité d'utilisation perçue (valeur propre initiale : 3,73 ; α = .83)	
Apprendre à utiliser ce logiciel serait facile pour moi	.65
Je trouverai facile d'utiliser ce logiciel de saisie de texte	.53
Je serai capable d'utiliser ce logiciel de saisie de texte	.69
J'ai la capacité à utiliser ce logiciel de saise de texte	.52
Facteur 2 : Utilité perçue (valeur propre initiale : 1,74 ; α = .82)	
Utiliser ce logiciel pourrait améliorer mes performances au quotidien	-.59
Utiliser ce logiciel pourrait me permettre d'améliorer mon efficacité	-.91
Utiliser ce logiciel pourrait me permettre d'améliorer ma productivité	-.83
Facteur 3 : Intention d'utilisation (valeur propre initiale : 1,12 ; α = .82)	
J'ai l'intention d'utiliser ce logiciel à l'avenir	-.73
J'ai l'intention d'utiliser ce logiciel de saisie fréquemment à l'avenir	-.91

[15] Une valeur de KMO de moins de .40 est inacceptable, .50 est misérable, .60 est médiocre, .70 est moyenne, .80 est méritoire, et .90 est merveilleuse (Norusis, M. J., 1993, SPSS professional statistics cité par Durand, 2005)

9.3.1.2. Analyse factorielle sur les mesures de DIGIME

L'analyse montre un indice KMO de .74. La qualité de la représentation initiale y est bonne (largement supérieure à .20). Ici encore, la solution factorielle révélée par l'analyse s'est montrée juste sur le plan théorique. L'analyse a extrait trois facteurs que nous avons désignés comme étant respectivement l'utilité perçue, la facilité d'utilisation perçue, et l'intention d'utilisation. La matrice des types avec analyse de fiabilité est reproduite ci-dessous :

Facteur 1 : Utilité perçue (valeur propre initiale : 5,91 ; α = .89)	
Utiliser ce logiciel pourrait améliorer mes performances au quotidien	.60
Utiliser ce logiciel pourrait me permettre d'améliorer mon efficacité	.94
Utiliser ce logiciel pourrait me permettre d'améliorer ma productivité	.95
Je trouverais ce logiciel de saisie utile au quotidien	.64
Facteur 2 : Facilité d'utilisation perçue (valeur propre initiale : 2,12 ; α = .81)	
Apprendre à utiliser ce logiciel serait facile pour moi	.53
Il me serait facile de devenir compétent dans l'utilisation de ce logiciel de saisie	.67
Je serais capable d'utiliser ce logiciel de saisie de texte	.87
J'ai la capacité à utiliser ce logiciel de saise de texte	.74
Facteur 3 : Intention d'utilisation (valeur propre initiale : 1,19 ; α = .91)	
J'ai l'intention d'utiliser ce logiciel à l'avenir	-.76
J'ai l'intention d'utiliser ce logiciel de saisie fréquemment à l'avenir	-.60

9.3.1.3. Analyse factorielle sur les mesures du clavier virtuel

Pour cette analyse, l'indice KMO s'élève à .71. De même, elle témoigne d'une excellente qualité de représentation. Contrairement aux deux précédentes analyses, quatre facteurs ont été extraits. Si l'intention d'utilisation et l'utilité perçue ne posent pas de problème de labellisation, la facilité d'utilisation perçue s'étend sur les deux derniers facteurs. Dans le souci d'adhérer au mieux aux données de la littérature, nous ne retiendrons que le premier de ces facteurs. En effet, la théorie ne fait pas état de nuances particulières au sein même du concept de facilité d'utilisation perçue. Par ailleurs, la valeur propre initiale du quatrième facteur est tout juste supérieure à 1 (1,09), critère couramment retenu pour déterminer le seuil

de rejet des facteurs. Au final, la matrice partielle des types avec analyse de fiabilité est la suivante :

Facteur 1 : Intention d'utilisation (valeur propre initiale : 4,65 ; α = .86)	
J'ai l'intention d'utiliser ce logiciel à l'avenir	.87
J'ai l'intention d'utiliser ce logiciel de saisie fréquemment à l'avenir	.88
Facteur 2 : Utilité perçue (valeur propre initiale : 2,49 ; α = .87)	
Utiliser ce logiciel pourrait améliorer mes performances au quotidien	.79
Utiliser ce logiciel pourrait me permettre d'améliorer mon efficacité	.84
Utiliser ce logiciel pourrait me permettre d'améliorer ma productivité	.78
Facteur 3 : Facilité d'utilisation perçue (valeur propre initiale : 1,52 ; α = .81)	
Apprendre à utiliser ce logiciel serait facile pour moi	.91
Il me serait facile de devenir compétent dans l'utilisation de ce logiciel de saisie	.70
Je trouverais facile d'utiliser ce logiciel de saisie de texte	.59
Je serais capable d'utiliser ce logiciel de saisie de texte	.53

Nous avons donc pu constater que la théorie se vérifiait dans cette analyse et que les facteurs qui en ont été extraits présentaient des propriétés psychométriques satisfaisantes malgré les quelques modifications que nous avons fait subir aux items originaux, ne serait-ce que du fait de la traduction.

9.3.2. Test des hypothèses

9.3.2.1. Effets des interfaces sur les VD (H1)

En ce qui concerne le logiciel préféré, tous les participants ont préféré le clavier virtuel à DIGIME. Ce choix a toujours été justifié par la facilité d'utilisation du clavier virtuel ne nécessitant pas, selon eux, d'apprentissage. Les résultats montrent effectivement des différences notables d'une interface à l'autre, alors même que l'utilité pratique est identique. Les principaux résultats sont présentés dans le tableau 19, page suivante.

Les comparaisons intragroupes sur ces VD ont témoigné de différences significatives. Les distributions n'étant pas normales, nous leur appliquons le test de Wilcoxon. Il en va de même pour l'intention d'utilisation et les items du test de Nielsen, mesurés sur des échelles

ordinales. Conformément à nos attentes, les deux interfaces se sont différenciées aussi bien sur l'effort mental (Z = -5,02, *p* = .00) et l'efficacité (Z = -6,39, *p* = .00) que sur la productivité (Z = -6,39, *p* = .00). A l'inverse, le nombre d'erreurs restantes est la seule mesure de performance qui n'a pas varié statistiquement d'une interface à l'autre (Z = -1,08, *p* = .27). S'agissant des mesures subjectives, les jugements des participants ont été nettement contrastés suivant le logiciel de saisie : facilité d'apprentissage (Z = -6,28, *p* = .00), reproductibilité (Z = -6,53, *p* = .00), efficacité perçue (Z = -4,07, *p* = .00), gestion des erreurs (Z = -5,76, *p* = .00) et satisfaction (Z = -6,40, *p* = .00). D'une manière générale, les jugements ont été meilleurs pour le clavier virtuel. De même, l'utilité perçue (Z = -5,75, *p* = .00) et la facilité perçue (Z = -6,37, *p* = .00) ont montré une nette préférence pour le clavier. Par conséquent et sans surprise, les intentions d'utilisation ont penché en faveur du clavier virtuel par rapport à DIGIME (Z = -5,99, *p* = .00).

	DIGIME	Clavier virtuel
Effort mental (s)	538 (215)	417 (200)
Efficacité (caractères corrects saisis)	93 (39)	590 (125)
Erreurs restantes	6,8 (9,1)	11,6 (38,3)
Productivité	.20 (.13)	1,89 (1,47)
Facilité d'apprentissage*	2,0 (2,0)	5,0 (1,0)
Efficacité (perçue)*	2,0 (1,0)	5,0 (1,0)
Reproductibilité*	4,0 (2,0)	5,0 (1,0)
Gestion des erreurs*	0,5 (1,0)	3,0 (2,25)
Satisfaction*	2,0 (2,0)	5,0 (1,0)
Facilité d'utilisation perçue	5,0 (2,0)	2,0 (.00)
Utilité perçue	6,0 (2,5)	3,0 (1,5)
Intention d'utilisation**	6,5 (1,0)	4,0 (2,5)

Note : Moyenne (écart-type) ; sauf * et ** Médianne (intervalle interquartile). * Les échelles de réponses sont orientées entre un pôle très négatif ou difficile à 0 et un pôle très positif ou facile à 5. ** Les échelles de réponses sont orientées entre un pôle très positif à 1 et un pôle très négatif à 7.

Tableau 19: Principales mesures d'utilisabilité en fonction des interfaces

9.3.2.2. Test des effets intragroupes (H2, H3, H4 et H5)

Nous procédons à plusieurs régressions linéaires, simples et multiples hiérarchiques (introduction simultanée), en vue de tester nos hypothèses. Même si la distribution des résidus a, à plusieurs reprises, violé les conditions de normalité et d'homoscédasticité de la variance nécessaires à la conduite d'une analyse de régression, nous avons cependant compté sur la robustesse de la régression (Howell, 1997, p.587). Dans le cas de l'introduction simultanée de plusieurs VI, nous avons vérifié les statistiques de colinéarité qui ont donné lieu à des valeurs de tolérence et de facteur d'inflation de la variance acceptables. Les tableaux 20, 21 et 22 (ci-dessous et page suivante) fournissent les principales mesures de corrélations entre les variables mesurées. Les différentes échelles ont été orientées dans la même direction afin que le signe des corrélations reflète véritablement le sens des relations entre les variables (sauf pour les scores factoriels des dimensions du TAM).

	Dimensions du TAM			Test d'utilisabilité perçue de Nielsen					Perf.
	Facilité perçue	Utilité perçue	Inten. d'util.	Facilité d'app.	Effica.	Reprod.	Gestion erreurs	Satisfac	Effici.
Facilité perçue	–								
Utilité perçue	-.40**	–							
Intention d'utilisation	.61**	-.26	–						
Facilité d'app.	.43*	-.12	.19	–					
Efficacité	.27*	-.21	.29*	.17	–				
Reproductibilité	.07	-.01	.22	.13	.05	–			
Gestion des erreurs	.27*	.09	.19	.18	.02	.03	–		
Satisfaction	.36**	-.09	.24	.41**	.35**	.15	.37**	–	
Efficience	.08	.22	.06	.01	-.19	-.05	.07	.10	–

Note : rho de Spearman, corrélations bivariées, ** $p \leq .01$, * $p \leq .05$, N=56

Tableau 20: Corrélations entre mesures Δ de l'expérience

	Dimensions du TAM			Test d'utilisabilité perçue de Nielsen					Perf.
	Facilité perçue	Utilité perçue	Inten. d'util.	Facilité d'app.	Effica.	Reprod.	Gestion erreurs	Satisfac	Effici.
Facilité perçue	–								
Utilité perçue	.26*	–							
Intention d'utilisation	.42**	.28*	–						
Facilité d'app.	.44**	.15	.30*	–					
Efficacité	.39**	.45**	-.58**	.27*	–				
Reproductibilité	.05	-.15	.08	.20	.00	–			
Gestion des erreurs	.29*	.04	.40**	.40**	.04	.15	–		
Satisfaction	.42**	.26*	.64**	.45**	.53**	.12	.29	–	
Efficience	.23	-.15	-.20	.01	.10	.05	.23	-.02	–

Note : rho de Spearman, corrélations bivariées, ** $p \leq .01$, * $p \leq .05$, N=56

Tableau 21: Corrélations entre mesures de l'expérience pour DIGIME

	Dimensions du TAM			Test d'utilisabilité perçue de Nielsen					Perf.
	Facilité perçue	Utilité perçue	Inten. d'util.	Facilité d'app.	Effica.	Reprod.	Gestion erreurs	Satisfac	Effici.
Facilité perçue	–								
Utilité perçue	-.09	–							
Intention d'utilisation	-.21	.22	–						
Facilité d'app.	.49**	-.04	-.07	–					
Efficacité	.40**	.01	-.28*	.51**	–				
Reproductibilité	.32*	-.09	.02	.18	.30*	–			
Gestion des erreurs	.39**	.19	-.10	.50**	.24	.00	–		
Satisfaction	.31*	.01	-.09	.59**	.37*	.15	.17	–	
Efficience	.26	.05	.08	.16	.23	.19	.14	.34**	–

Note : rho de Spearman, corrélations bivariées, ** $p \leq .01$, * $p \leq .05$, N=56

Tableau 22: Corrélations entre mesures de l'expérience pour le clavier virtuel

9.3.2.3. Effets de la mesure d'efficience sur les scores au test de Nielsen (H2)

Le tableau 23 donne les résultats de l'analyse pour H2. Nous avons procédé à cinq analyses séparées pour chacune des VD. Nous avons constaté que l'efficience n'a expliqué aucun des scores aux items du test d'utilisabilité perçue, exception faite de la satisfaction pour le clavier virtuel. Cet effet, relativement mince (R^2 = .05), ne se répercute pas sur la comparaison entre les deux logiciels.

Test de Nielsen		Δ				DIGIME				CV			
VD	VI	R^2	β	*B*	*ES B*	R^2	β	*B*	*ES B*	R^2	β	*B*	*ES B*
Facilité d'apprentissage		-.01				-.01				.02			
	Efficience		.09	.08	.16		.03	.04	.16		.19	.15	.10
Efficacité		.01				-.01				.02			
	Efficience		-.16	-.19	-.11		.08	.06	.10		.20	.09	.06
Reproductibilité		.00				-.01				-.01			
	Efficience		-.13	-.10	.17		.07	.08	.15		.09	.06	.09
Gestion des erreurs		.01				.02				-.01			
	Efficience		.10	.07	.17		.21	.24	.15		.09	.09	.14
Satisfaction		.00				.00				.05			
	Efficience		.10	.09	.14		-.10	-.09	.13		.26	.22	.11*

Note: ** $p \leq .01$, * $p \leq .05$, R^2 est ajusté, *ES B* désigne l'erreur standard de *B, N=56*.

Tableau 23: Effet de l'efficience sur les réponses aux items du test de Nieslen

9.3.2.4. Effets de l'efficience et des scores au test de Nielsen sur la facilité d'utilisation perçue (H3)

Pour cette analyse, nous avons introduit simultanément dans la régression multiple l'efficience ainsi que les scores aux items du test de Nielsen : facilité d'apprentissage, efficacité, reproductibilité, gestion des erreurs et satisfaction. Les résultats ont montré que seule la facilité d'apprentissage expliquait la facilité d'utilisation perçue (sr^2 = .11) pour la comparaison entre les interfaces. Cet effet s'est retrouvé pour les mesures de DIGIME prises individuellement (sr^2 = .10). Par contre, pour le clavier virtuel, seule la gestion perçue des erreurs a semblé pouvoir expliquer la facilité d'utilisation perçue (sr^2 = .08), sans que cela ne se traduise sur les mesures Δ.

		Δ				DIGIME				CV		
VD	VI	*R²*	*β*	*B*	*ES B*	*R²*	*β*	*B*	*ES B*	*R²*	*β*	*B*	*ES B*
Facilité d'utilisation		.15				.20				.35			
Facilité d'apprentissage			.25	.04	.10**		.33	.25	.10**		.20	.21	.14
Efficacité			.17	.18	.12		.16	.19	.16		.21	.35	.20
Reproductibilité			.05	.08	.08		-.05	-.04	.10		.19	.21	.12
Gestion des erreurs			.11	.18	.08		.06	-.05	.11		.25	.19	.09*
Satisfaction			.11	.18	.08		.16	.15	.14		.07	.06	.12
Efficience			.00	.00	.11		.15	.14	.11		.17	.14	.09

Note: ** p ≤ .01, * p ≤ .05, R² est ajusté, ES B désigne l'erreur standard de B, N=56.

Tableau 24: Effets du test de Nielsen et de l'efficience sur la facilité d'utilisation perçue

9.3.2.5. Effets du test de Nielsen et de la facilité d'utilisation perçue sur l'utilité perçue (H4)

Nous avons commencé par étudier comment le test de Nielsen et la facilité d'utilisation perçue pouvait prédire l'utilité perçue pour les mesures Δ. Etant donné que la facilité d'apprentissage (l'item 1 du test de Nielsen) a contribué à expliquer la facilité d'utilisation perçue, nous l'avons introduit dans un premier bloc de la régression hiérarchique afin que son influence ne soit pas mesestimée. Le second bloc contient les autres items du test de Nieslen ainsi que le score de facilité d'utilisation perçue. La régression nous a appris que seule la facilité d'utilisation perçue permettait de prédire l'utilité perçue (sr^2 = .13).

Bloc 1			Δ		
VD	VI	*R²*	*β*	*B*	*ES B*
Utilité perçue		-.01			
Facilité d'apprentissage			-.05	.03	.10

Bloc 2			Δ		
VD	VI	*R²*	*β*	*B*	*ES B*
Utilité perçue		.12			
Facilité d'apprentissage			.05	.07	.10
Efficacité			-.17	-.17	.13
Reproductibilité			.01	.02	.08
Gestion des erreurs			.14	.18	.09
Satisfaction			.11	.20	.12
Facilité d'utilisation perçue			-.45	-.42	.15**

Note: ** *p* ≤ .01, * *p* ≤ .05, *R²* est ajusté, *ES B* désigne l'erreur standard de *B, N=56*.

Tableau 25: Effets du test de Nielsen et de la facilité d'utilisation perçue sur l'utilité perçue comparée entre interfaces

Nous avons procédé de la même façon avec les mesures pour DIGIME, en contrôlant l'influence de la facilité d'apprentissage du test de Nielsen. Nous avons pu constater que c'est le seul sentiment d'efficacité qui a rendu compte de la variance de l'utilité perçue (sr^2 = .18) (voir tableau 26).

Pour finir, nous avons évalué comment les variables pour le clavier virtuel ont pu expliquer la variance de l'utilité perçue (tableau 27, page suivante). La procédure de contrôle de l'effet de la gestion perçue des erreurs a fait apparaître qu'au sein des autres variables, elle est seule à pouvoir rendre compte de l'utilité perçue, à hauteur de sr^2 = .07, sans toutefois permettre que le modèle explicatif pris dans sa totalité se montre statistiquement significatif.

Bloc 1		DIGIME			
VD	VI	R^2	β	*B*	*ES B*
Utilité perçue		.00			
Facilité d'apprentissage			.13	.10	.10

Bloc 2		DIGIME			
VD	VI	R^2	β	*B*	*ES B*
Utilité perçue		.22			
	Facilité d'apprentissage		.00	.00	.10
	Efficacité		.46	.57	.16**
	Reproductibilité		-.18	-.15	.10
	Gestion des erreurs		.00	.00	.10
	Satisfaction		.04	.04	.13
	Facilité d'utilisation perçue		.10	.10	.14

Note: ** $p \leq .01$, * $p \leq .05$, R^2 est ajusté, *ES B* désigne l'erreur standard de *B, N=56*.

Tableau 26: Effets du test de Nielsen et de la facilité d'utilisation perçue sur l'utilité perçue de DIGIME

Bloc 1			CV		
VD	VI	R^2	β	*B*	*ES B*
Utilité perçue		.03			
Gestion des erreurs			.23	.19	.19

Bloc 2			CV		
VD	VI	R^2	β	*B*	*ES B*
Utilité perçue		-.02			
Gestion des erreurs			.31	.25	.12*
Facilité d'apprentissage			-.13	-.14	.19
Efficacité			.05	.09	.28
Reproductibilité			-.01	-.02	.17
Satisfaction			.05	.05	.16
Facilité d'utilisation perçue			-.14	-.14	.17

Note: ** $p \leq .01$, * $p \leq .05$, R^2 est ajusté, *ES B* désigne l'erreur standard de *B, N=56*.

Tableau 27: Effets du test de Nielsen et de la facilité d'utilisation perçue sur l'utilité perçue du clavier virtuel

9.3.2.6. Effets du test de Nielsen, de la facilité d'utilisation perçue, de l'utilité perçue et de l'efficience sur l'intention d'utilisation (H5)

Nous avons une fois encore procédé à des régressions hiérarchiques.

S'agissant des mesures comparées, puisque la facilité d'utilisation perçue a contribué à expliquer l'utilité perçue, de même que la facilité d'apprentissage avec la facilité d'utilisation, nous avons introduit dans un premier bloc la facilité d'apprentissage, et dans le second, la facilité d'utilisation perçue.

Le tableau 28, résumant les résultats de la régression hiérarchique page suivante, montre que seule la facilité d'utilisation perçue a permis d'expliquer l'acceptabilité comparée (sr^2 = .19).

Bloc 1			Δ		
VD	VI	R^2	β	*B*	*ES B*
Intention d'utilisation		.00			
Facilité d'apprentissage			-.12	.16	.10

Bloc 2			Δ		
VD	VI	R^2	β	*B*	*ES B*
Intention d'utilisation		.32			
Facilité d'apprentissage			.07	.10	.09
Facilité d'utilisation perçue			-.66	-.63	.13**

Bloc 3			Δ		
VD	VI	R^2	β	*B*	*ES B*
Intention d'utilisation		.31			
Facilité d'apprentissage			.09	.12	.09
Facilité d'utilisation perçue			-.57	-.55	.15**
Utilité perçue			.02	.02	.12
Efficacité			-.17	-.17	.12
Reproductibilité			-.11	-.16	.08
Gestion des erreurs			-.04	-.07	-.08
Satisfaction			.00	.00	.10
Efficience			.03	.04	.10

Note: ** $p \leq .01$, * $p \leq .05$, R^2 est ajusté, *ES B* désigne l'erreur standard de *B, N=56*.

Tableau 28: Effets du test de Nielsen, de la facilité d'utilisation perçue, de l'utilité perçue et de l'efficience sur l'acceptabilité comparée

Nous avons procédé de la même façon pour les mesures propres à DIGIME et au clavier virtuel.

En ce qui concerne DIGIME, le tableau 29 (page suivante) montre que l'effet du sentiment d'efficacité (mesuré par un item du test de Nielsen) sur l'intention d'utilisation a perdu son pouvoir explicatif (sr^2 = .12) lorsqu'il a été introduit avec les autres variables. Parmi elles, la gestion perçue des erreurs (sr^2 = .10) et la satisfaction (sr^2 = .19) rendent compte d'une part non négligeable de la variance de l'intention d'utilisation de DIGIME.

Pour le clavier virtuel, seule l'utilité perçue peut prédire l'intention d'utilisation (sr^2 = .12).

Bloc 1			DIGIME		
VD	VI	R^2	β	*B*	*ES B*
Intention d'utilisation		.03			
Facilité d'apprentissage			-.22	-.16	.10

Bloc 2			DIGIME		
VD	VI	R^2	β	*B*	*ES B*
Intention d'utilisation		.30			
Facilité d'apprentissage			-.09	-.06	.08
	Efficacité		-.54	-.65	.13**

Bloc 3			DIGIME		
VD	VI	R^2	β	*B*	*ES B*
Intention d'utilisation		.52			
	Facilité d'apprentissage		.12	.09	.08
	Efficacité		-.33	-.39	.15**
	Reproductibilité		.00	.01	.15
	Gestion des erreurs		-.25	-.20	.08*
	Satisfaction		-.39	-.36	.10**
	Efficience		-.09	-.08	.09
	Facilité d'utilisation perçue		-.08	-.08	.11
	Utilité perçue		-.143	-.13	.11

Note: ** $p \leq .01$, * $p \leq .05$, R^2 est ajusté, *ES B* désigne l'erreur standard de *B, N=56*.

Tableau 29: Effets du test de Nielsen, de la facilité d'utilisation perçue, de l'utilité perçue et de l'efficience sur l'intention d'utilisation de DIGIME

Bloc 1			CV		
VD	VI	R^2	β	*B*	*ES B*
Intention d'utilisation		.00			
	Gestion des erreurs		.13	.11	.11

Bloc 2			CV		
VD	VI	R^2	β	*B*	*ES B*
Intention d'utilisation		.16			
	Gestion des erreurs		.13	.52	.12
	Facilité d'apprentissage		-.24	-.27	.18
	Efficacité		.26	.48	.26
	Reproductibilité		-.13	-.16	.16
	Satisfaction		.12	.12	.15
	Efficience		-.08	-.07	.11
	Facilité d'utilisation perçue		.22	.24	.18
	Utilité perçue		-.35	-.35	.13**

Note: ** $p \leq .01$, * $p \leq .05$, R^2 est ajusté, *ES B* désigne l'erreur standard de *B, N=56*.

Tableau 30: Effets du test de Nielsen, de la facilité d'utilisation perçue, de l'utilité perçue et de l'efficience sur l'intention d'utilisation du clavier virtuel

9.4. Discussion

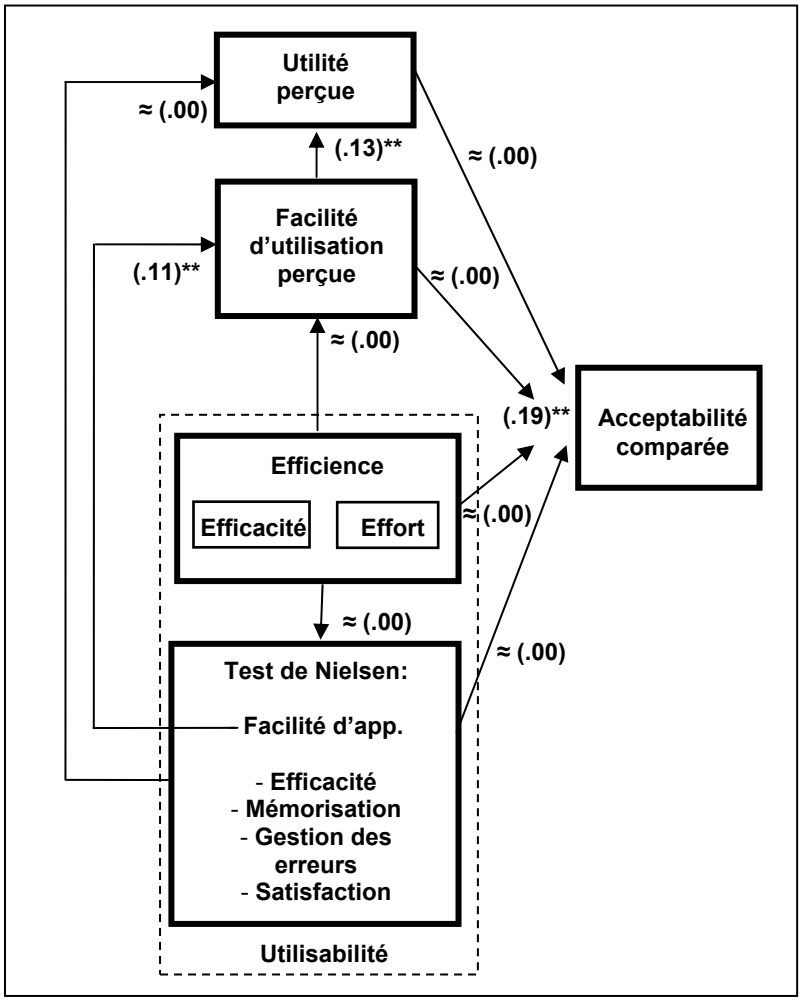
La grande volatilité des patterns explicatifs révélés par les analyses de régression, suivant le logiciel, incite à la prudence. La mise en relation des scores par interface et des effets intragroupes dans les régressions peuvent sembler surprenants. A utilité théorique égale, le clavier virtuel a été jugé plus utile que DIGIME. Pourtant, la facilité d'utilisation perçue n'en rend compte sur aucune mesure alors qu'elle était aussi plus grande pour le clavier. Toutefois, un rôle modérateur des interfaces est suspecté dans les modèles d'acceptation (Schepers & Wetzels, 2007). Des problèmes de colinéarité entre VI ont pu aussi jouer malgré nos précautions. Ainsi, la méthode d'introduction simultanée dans les régressions a peut-être facilité des chevauchements de variance entre mesures et masquer certains effets. Cette remarque est valable pour toute méthode d'introduction (ascendante, descendante ou pas à pas, d'autant plus criticables selon nous (Onwuegbuzie & Daniel, 2003)). Par contre, la comparaison des deux interfaces donne des résultats assez faciles à interpréter. S'agit-il d'un heureux hasard ? Il nous est difficile de le savoir. La question mérite cependant d'être posée, ce d'autant plus que la littérature témoigne des résultats les plus contradictoires quand il s'agit de mettre en relation des mesures objectives et subjectives (Frokjaer, Hertzum, & Hornbaek, 2000; Nielsen & Levy, 1994; Tricot & Chanquoy, 1996; Yeh & Wickens, 1984). Nous proposons donc une interprétation des résultats s'appuyant sur les mesures comparées entre logiciels, puisque notre protocole était avant tout conçu pour permettre cette comparaison. De surcroît, la comparaison entre interface visait à donner aux participants du « contexte » à leurs jugements, notamment en leur permettant d'apprécier entre elles des méthodes de saisie concurrentes.

Le premier résultat de cette expérience est la validité du test TAM que nous avons utilisé dans ce contexte. Une fois encore, il a fait preuve de la robustesse des dimensions sur lesquelles il s'articule, et ce, indépendamment du contexte d'utilisation d'une technologie matérielle, auquelle il n'a été jusque-là que peu consacré.

Deuxième résultat important, l'efficience n'a pas permis de prédire les mesures subjectives que nous avons collectées (particulièrement celles de facilité et d'efficacité perçues).

Troisième résultat, le test d'utilisabilité subjective peut prédire la facilité d'utilisation perçue. En l'occurrence, pour cette recherche, par le biais de la facilité d'apprentissage.

Dernier point, la facilité d'utilisation perçue explique seule l'acceptabilité comparée, à hauteur de sr^2 =. 19. Ces résultats sont résumés à la figure 52.



Note: (sr^2) ** $p \leq .01$, * $p \leq .05$

Figure 52: Résumé graphique des hypothèses et des principaux résultats obtenus sur les mesures Δ

Alors que nous pensions, autant sur la base de nos lectures (Arning & Ziefle, 2007; Hasan & Ahmed, 2007; Venkatesh & Davis, 1996) qu'à partir des résultats de l'expérience précédente, que l'efficience serait un prédicteur des jugements des participants, cela s'est avéré faux. Ce résultat, inattendu, nous est assez difficile à expliquer. Manifestement, cette variable n'a de toute évidence pas été prise en considération par les sujets pour se forger des opinions sur leur interaction avec les logiciels. Il est aussi permis de penser que nous n'avons pas fait un bon choix en mesurant sans distinction tout l'effort mental, indépendamment qu'il relève de la copie de texte ou d'autres processus.

Le fait que l'utilité perçue n'ait pas eu d'effet sur l'acceptabilité comparée est somme toute assez logique : de fait, les deux interfaces proposaient le même service. Par conséquent, toute l'intention d'utilisation s'appuie sur la facilité d'utilisation perçue. Si celle-ci ne peut pas être expliquée par l'efficience, c'est qu'elle se fonde sur d'autres critères dont l'efficience n'a pas rendu compte.

Ces critères peuvent prendre leur origine dans les jugements portés sur la qualité de l'interaction. C'est ce que montre le test de Nielsen qui a prédit une portion significative de la variance de la facilité d'utilisation perçue. Il est impossible de décider, à partir de ces seuls résultats, si c'est la facilité d'apprentissage qui prédit, en général, la facilité perçue, ou bien si, en fonction des circonstances d'interaction (c'est-à-dire types de logiciel, de tâches et d'environnements), d'autres items du test de Nielsen peuvent s'avérer eux aussi prédictifs (ce qui est le cas par interface). Nous penchons néanmoins pour la deuxième hypothèse puisque tous les sujets de l'expérience ont rapporté préférer le clavier virtuel au motif que sa pratique est plus simple et plus habituelle. Si tel est le cas, on mesure le bénéfice que les concepteurs d'interfaces auraient à généraliser l'utilisation de questionnaires d'utilisabilité subjective. En prenant en compte l'avis des utilisateurs sur ces dimensions, ils pourraient anticiper la facilité d'utilisation perçue tout en concentrant leurs efforts spécifiquement sur les aspects de l'interaction qui doivent être améliorés. Pour autant, l'utilisabilité perçue n'est pas substituable à la facilité d'utilisation perçue. Les corrélations entre ces mesures, si elles peuvent être significatives, restent assez faibles, et l'utilisabilité perçue n'a pas pu expliquer la variance de l'utilité perçue, contrairement à la facilité d'utilisation. Il en est de même pour l'intention d'utilisation, expliquée par la seule facilité d'utilisation perçue. Par conséquent, le modèle P3 (Dillon & Morris, 1998; 1999) qui postule *a priori* toutes relations possibles entre Potentiel, Perceptions et Performances fait sans doute erreur. Les outils de managements des systèmes d'informations (MIS) tels que le TAM ont une efficacité inégalée à prédire l'intention d'utilisation. Les mesures d'utilisabilité subjective sont insuffisantes dans l'état actuel de leur opérationnalisation, pour prédire l'acceptation. Sans doute le peuvent-elles seules, comme le NASA-TLX dans l'expérience précédente, mais, y échouent mises en concurrence avec des outils spécialement développés pour cela. Néanmoins, en amont de la décision d'utilisation, il est possible d'évaluer des critères qui vont permettre d'anticiper une part de la variance de cette intention comportementale. Il est très probable que l'utilisabilité n'explique pas directement l'acceptation. Par contre, elle y contribue sans nul doute par la

facilité d'utilisation perçue. La question qui reste en suspend est de savoir sur quels critères mesurables objectivement les utilisateurs s'appuient-ils pour juger de la qualité de l'interaction qu'ils ont eu avec le dispositif. L'efficience, qui nous apparaissait en être un (Paas, Tuovinen, Tabbers, & Van Gerven, 2003), ne l'est pas ici. Ce problème est crucial : au plan cognitif, cela permettrait de comprendre sur quels critères les utilisateurs évaluent leurs performances. Au plan décisionnel, cela permettrait de prévoir l'acceptation à partir de mesures de performance. Notre étude hélas, ne permet pas d'y répondre.

10- Discussion générale

Notre objectif de départ était double : comprendre pourquoi les interfaces de saisie et de reconnaissance de l'écriture manuscrite souffraient d'un problème d'acceptabilité et tenter d'y remédier dans le cadre de la conception de DIGIME. Si nous pensons avoir compris pourquoi ce type d'interface de saisie est massivement rejeté, nous n'avons pas été en mesure d'y apporter des solutions efficaces.

Notre travail s'est tout d'abord orienté sur l'ergonomie de l'interface DIGIME (expériences 1 et 2). Plusieurs travaux du début des années 90 ont développé la notion d'utilisabilité, attribut des interfaces qui témoigne de la facilité avec laquelle ont peut accomplir des objectifs que l'on se fixe (ISO-9241-11, 1994; Nielsen, 1993; Shackel, 1991, Tricot & al., 2003). Ces modèles conditionnent l'acceptabilité des systèmes à l'utilisabilité, bien qu'elle n'en soit pas le seul déterminant. L'utilisabilité trouve son origine dans certaines caractéristiques de la machine que l'on désigne comme étant son ergonomie. Cependant, l'ergonomie ne détermine pas à elle seule l'utilisabilité. C'est dans l'interaction entre une machine et un utilisateur, en contexte, que l'utilisabilité va prendre son origine. La psychologie ergonomique dispose de modèles éprouvés qui permettent d'anticiper, améliorer et comprendre les processus de conception qui font qu'une ergonomie est adaptée ou pas aux utilisateurs.

La première étude proposait de tester le principe de contiguïté spatiale appliqué à la saisie de texte sur PDA. Ecrire, quel qu'en soit le support, implique une activité de contrôle passant notamment par la lecture régulière du texte produit (Kellogg, 1996). Cette activité de contrôle existe également dans toute réalisation de tâche et de résolution de problème. Utiliser une IHM implique donc de résoudre un certain nombre de problèmes dont l'utilisateur doit s'assurer que les solutions qu'il propose sont en adéquation avec les buts qu'il poursuit (Drommi, Ulferts, & Shoemaker, 2001; Sweller, 1991; Tricot, 1998). Par conséquent, si les sources informationnelles que l'uilisateur doit intégrer pour mener à bien sa tâche sont éloignées, il en résulte un partage de l'attention préjudiciable à la performace par augmentation de la charge cognitive (Chandler & Sweller, 1992; Wickens & Carswell, 1995). Pour éviter cela, nous avons proposé de rapprocher spatialement la zone dans laquelle l'utilisateur écrit sur DIGIME, de celle où s'affichent les lettres reconnues. Ecrire impose par ailleurs, même sur papier, un contrôle minimal, particulièrement visuel. Or, on sait que la suppression du feedback visuel fait augmenter la charge d'exécution motrice, c'est-à-dire la

quantité d'effort mental qui est attribuée à la réalisation du geste moteur. Cela nous a amené à penser que ne pas afficher l'encre du scipteur quand il écrivait allait se révéler pénalisant. Les résultats ont montré que ces choix théoriques étaient inadéquats. La présence de l'encre s'est avérée être une entrave du fait du fonctionnement de DIGIME. Ce logiciel ne reconnaît que les caractères isolés, et la présence de lettres manuscrites à l'écran a incité les utilisateurs à lier les lettres entre elles, empéchant le bon fonctionnement du dispositif. En effet, l'écriture script est assez inhabituelle, et oblige à rompre des automatismes bien ancrés (Graham, Weintraub, & Berninger, 2001). Par conséquent, le bénéfice que les utilisateurs pouvaient retirer de la conservation à l'écran de leur trace écrite a complètement disparu. Ce d'autant plus qu'il y avait au minimum un feedback, même éloigné, les informant de leur progression dans la tâche. Ce feedback, lorsqu'il était contigu à la zone de saisie, et contrairement à nos attentes, a nuit à la qualité de l'interaction avec la machine. Bien que ce résultat nous ait surpris, nous l'avons interprété comme étant un problème de surcharge informationnelle. La contiguïté du feedback, telle que nous l'avons appliquée, ne se contentait pas de le rapprocher de la zone de saisie, mais également de l'ajouter au feedback éloigné qui restait affiché lui aussi. Cet effet était d'ailleurs d'autant plus important que l'encre conservée constituait également un feedback. Une autre piste explicative consistait à penser que le feedback proche, bien que contigu, était insuffisamment intégré à la zone de saisie. Ces constats ont donné lieu à des améliorations de DIGIME allant dans ce sens. Cette première étude a permis, quoi qu'il en soit, de constater que les effets néfastes observés ne duraient pas plus que le début de la prise en main du logiciel.

Cela nous a incités à travailler sur un didacticiel pour DIGIME (expérience 2). Il n'existe pas de recherches, à notre connaissance, sur l'apprentissage graphomoteur chez le scripteur adulte. Or, apprendre à utiliser DIGIME, c'est apprendre une autre façon d'écrire. Cette thématique étant assez mal connue, nous l'avons abordée sous l'angle plus général des apprentissages moteur. Ceux-ci constituent un cas particuliers des apprentissages procéduraux (Anderson, 1982; Thomas, 1997). L'apprentissage moteur est sous-tendu par des représentations mentales que Jeannerod (1995) appelle les images motrices. Celles-ci partagent le caractère analogique des illustrations. Nous avons fait l'hypothèse que des illustrations dynamique seraient sans doute plus efficaces que des images statiques pour transmettre cette information procédurale (Bétrancourt, Bauer-Morrison, & Tversky, 2001; Zacks & Tversky, 2003). Cette hypothèse, couramment traitée en psychologie ergonomique,

Deconde, Guillaume. Etude itérative des liens entre utilisabilité et acceptabilité d'un dispositif de saisie et de reconnaissance de l'écriture manuscrite - 2009

trouve écho dans les données issues de la neuropsychologie selon laquelle la présentation visuelle d'une lettre active le cortex moteur qui est normalement affecté à sa production (Longcamp & al., 2008). On espérait ainsi, à l'aide d'animations, favoriser ce type d'apprentissage. Les résultats de l'expérience n'ont pas montré de différences entre les illustrations statiques et dynamiques pour l'apprentissage de DIGIME. Si ces résultats sont récurrents dans ce type d'études (Bétrancourt, Bauer-Morrison, & Tversky, 2001; Zacks & Tversky, 2003), il faut noter aussi que l'interface DIGIME avait été enrichi d'aides en lignes qui permettaient à tout moment aux utilisateurs d'avoir à leur disposition une information illustrée statiquement. Cela n'a pu que diminuer les effets de l'exposition courte à un didacticiel dynamique en début d'épreuve (voire même de compenser l'absence de didacticiel). Toutefois, le fait que les participants aient abondamment consulté ces aides a fait preuve de leur utilité. Pour finir, un test d'utilisabilité perçue (SUS, Brookes, 1996) a permis de conclure au bien-fondé de la présence d'un didacticiel (particulièrement dynamique, mais aussi statique) : les sujets qui en avaient bénéficié ont utilisé des statégies de correction plus évoluées que ceux qui n'en avaient pas profité.

Ce résultat a donc donné lieu à l'installation d'un didacticiel sur DIGIME. L'interface étant dès lors complète en termes de prototypage (Vanderdonckt & Coyette, 2006), nous l'avons comparé à d'autres solutions de saisie de texte sur PDA (expérience 3). A cette occasion, nous avons constaté que les transformations apportées à DIGIME ne lui permettaient pas encore de rivaliser avec des interfaces du même type que lui (*JOT*), et encore moins avec le clavier virtuel, dont l'écrasante supériorité des performances qu'il autorise justifie pleinement son utilisation généralisée. De ce test utilisateurs, il est ressorti ce que nous avons interprété comme un rejet non pas de DIGIME, mais de toute forme d'écriture manuscrite sur PDA, y compris celle n'impliquant pas de reconnaissance.

Le fait que les participants aient ainsi rejetté l'idée d'écrire à la main sur un écran de Pocket PC nous a alertés sur la composition de nos échantillons. En effet, issus d'une population étudiante ne voyant pas l'utilité de ce type de machine, ils n'étaient pas à même de rendre compte de son acceptabilité, quand bien même l'utilisabilité y était améliorée. Plusieurs recherches, aussi bien dans les communautés HCI (Segall, Doolen, & Porter, 2005) que MIS (Agarwal, Ahuja, Carter, & Gans, 1998; King & He, 2006; Legris, Ingham, & Collerette, 2003) montrent que certaines populations (ou pour le moins des populations présentant certaines caractéristiques) acceptent plus facilement l'innovation et les

technologies, de sorte que cela puisse se répercuter même sur les performances d'utilisation (Segall, Doolen, & Porter, 2005). Partant du principe que nous n'avions pas évalué DIGIME sur des échantillons adéquats, nous avons décidé de tester si le niveau en informatique des utilisateurs pouvaient influer sur leurs performances d'usage, et donc, sur leurs jugements quant à la qualité de l'interface (expérience 4). C'est pourquoi nous avons comparé l'utilisation de DIGIME chez des étudiants en troisième année d'école d'ingénieur en informatique et des étudiants en psychologie. Selon nous, les experts en informatiques devaient, de par leurs connaissances générales sur le sujet (Cooper, 1990; Ericsson, 2005; Farrignton-Darby & Wilson, 2006)) être capables de les transférer utilement à cette situation d'interaction inconnue (Perkins & Salomon, 1994), justifiant ainsi de meilleures performances. Les résultats ont montré que si effectivement, les experts disposaient de connaissances générales en informatique plus importante que les novices, elles ne leur étaient d'aucun secours pour la prise en main de DIGIME, à ceci près qu'ils s'estimaient plus confiants. De fait, les experts se sont montrés plus performants avec DIGIME, et ils l'ont jugé plus acceptable, sans que l'on n'ait pu attribuer cela à leur expertise, leur confiance, ou leur sexe, qui constituaient autant de modérateurs possibles de l'intention d'usage identifiés par la littérature. Les études conduites sur le rôle des modérateurs dans l'acceptation des technologies négligent toujours les performances puisqu'elles ne s'appuient que sur de l'usage auto-rapporté (Legris, Ingham, & Collerette, 2003). Or, cette étude a montré que des modérateurs de l'intention d'utilisation pouvaient être accompagnés de niveaux de performances différents qui pouvaient eux aussi expliquer l'acceptation. Au moins autant que des caractéristiques démographiques ou attitudinales. D'où l'idée que l'étude de l'acceptation est indissociable de celle de l'interaction (utilisabilité et performance).

Ces deux thématiques, que l'on a désigné comme étant les champs réservés des communautés MIS et HCI ont été traitées jusqu'à présent de façon complètement indépendantes, sauf rares exceptions (Arning & Ziefle, 2007; Fiorello, 1999; Hasan & Ahmed, 2007; Venkatesh & Davis, 1996). Pourtant, plusieurs auteurs encouragent cette voie (Jamet & Février, 2008; Zhang & Dillon, 2003). Conscients de cette nécessité également, nous avons pris le parti de réorienter résolument nos recherches dans cette voie plutôt que continuer à travailler sur l'ergonomie de DIGIME. L'expérience 3 avait montré que le chemin à parcourir pour que DIGIME puisse rivalier avec le clavier virtuel était assez important. D'autre part, il nous semblait que la démarche hypothètico-déductive appliquée à un cycle de

conception itératif n'était pas nécessairement la méthode la plus adaptée ni la plus rapide par rapport à une inspection experte. En effet, l'évaluation de la qualité de l'interaction par des mesures d'utilisabilité ne renseigne pas forcément sur la cause ergonomique des difficultés rencontrées (Folmer, Van Gurp, & Bosch, 2003). Nous avons donc préféré comprendre les mécanismes du rejet de DIGIME, plus qu'en identifier les causes ergonomiques en vue d'y remédier.

C'est pourquoi nous avons initié une tentative de rapprochement entre les champs d'étude HCI et MIS (expérience 5 et 5 bis). Nous avons donc essayé de mettre en relation des mesures objectives de performances avec les jugements produits par les utilisateurs. Cette perspective, encore récente, n'est pas étayée par un modèle permettant de comprendre sur quels indices de performances ou de qualité de l'interaction les utilisateurs d'un logiciel vont s'appuyer pour se forger des opinions à son propos. D'autre part, conscients des problèmes d'échantillonnage que posait la mise à contribution répétée d'étudiants pour ce type d'étude, nous avons pris le parti de faire comparer aux participants plusieurs interfaces entre elles, afin que leurs jugements ne soient pas donnés « à l'aveugle ». En effet, pour beaucoup des sujets de nos expériences, ces manipulations ont été l'occasion de découvrir ce qu'était un PDA… En nous appropriant le protocole expérimental de Fiorello (1999), nous avons tablé sur le fait que les participants nous fourniraient non plus une intention d'utilisation à proprement parler (intention qui de toute façon était hypothétique, donc, sans grande valeur diagnostique), mais plutôt une valeur d'acceptabilité comparée, c'est-à-dire leur interface préférée en comparaison d'une autre. Cette méthode, à défaut de régler les problèmes d'échantillonnage que nous évoquions précédemment nous a permis, nous l'espérons, de le contourner. L'inspiration chez Fiorello avait également pour objectif de s'approcher du modèle P3 (Dillon & Morris, 1998; 1999). Ce modèle uniquement théorique propose d'expliquer l'acceptation des systèmes sur la base de leur utilité (Potentiel), des jugements (Perceptions) et de l'utilisabilité (Performance). Il postule *a priori* toutes relations possibles entre ces trois dimensions mais n'a jamais été validé. Nous considérons ce modèle comme la première pierre du rapprochement théorique dans lequel nous avons voulu nous inscrire entre ingéniérie décisionnelle et conception centrée utilisateur. Dans cette optique, nous avons fait l'hypothèse que l'écriture manuscrite étant largement automatisée chez les sujets adultes, c'est l'augmentation de la charge mentale induite par l'exécution motrice inhabituelle de l'écriture sur DIGIME qui pouvait en expliquer le rejet (expérience 5). En conséquence, les sujets devaient comparer DIGIME à *Note*, logiciel

d'écriture sur PDA ne consistant qu'en une feuille électronique, et n'impliquant pas de reconnaissance. A partir de cette mesure de charge mentale (procédure de Kellogg) associée au processus d'exécution motrice lors d'une tâche de copie (Olive, Kellogg, & Piolat, 2002; Olive & Piolat, 2002; Piolat, Kellogg, & Farioli, 2001; Piolat & Olive, 2000; Piolat, Roussey, Olive, & Farioli, 1996), nous avons pu composer plusieurs mesures d'utilisabilité : efficience notamment, en y incluant l'efficacité (Paas & Van Merriënboer, 1993) et utilisabilité en unifiant, selon les recommandations de la norme ISO 9241-11, efficacité, efficience et satisfaction (Darabi, 2004; Huang, Hong, & Eades, 2006; Paas, Tuovinen, Tabbers, & Van Gerven, 2003; Paas & Van Merriënboer, 1993; Tuovinen & Paas, 2004). Notre intérêt se portant sur la comparasion entre les interfaces, nous ne nous sommes pas intéressés aux mesures propres à chacune mais à là la différence entre les deux (mesures Δ). Nous avons mis en évidence que si le seul effort mental était insuffisant à prédire l'efficacité, l'acceptabilité comparée ou l'utilisabilité subjective, l'efficience s'est montrée prédictive de cette dernière. De même, la charge de tâche mentale (utilisabilité subjective) a expliqué 22% de la variance de l'acceptabilité comparée. De cette étude, nous avons tiré trois enseignements. Tout d'abord, la mesure d'utilisabilité que nous avons proposée n'a aucun intérêt prédictif pour l'intention d'utilisation. Ensuite, les mesures d'utilisabilité objective (efficacité, efficience, charge mentale) n'ont eu qu'un faible potentiel prédictif car les utilisateurs ne se sont que peu appuyés sur elles pour formuler leurs jugements. Une autre explication est qu'ils les ont inclues dans des représentations plus larges, intégrant d'autres dimensions. Pour finir, nous avons discuté les lacunes de notre étude quant à notre capacité à opérationnaliser l'utilisabilité, notamment pour ce qui est de la gestion des erreurs et des stratégies d'usage.

Cette interprétation des résultats nous a inspiré une dernière expérience (5 bis). Etant donné que les logiciels comparés dans l'expérience 5 n'avaient pas la même utilité, et que nous avions volontairement négligé ce paramètre, nous avons voulu réutiliser le protocole de la manipulation précédente afin de faire comparer DIGIME au clavier virtuel. Si ces deux interfaces de saisie ont la même utilité théorique, elles diffèrent par leur utilité pratique (Nielsen, 1993). Le modèle P3 tient compte de l'utilité objective des systèmes sans pour autant proposer de solutions pour la mesurer. En comparant deux dispositifs ayant la même utilité théorique, nous avons pensé pouvoir éluder la question. L'expérience précédente avait montré que les sujets avaient appuyé leur intention d'utilisation sur l'utilité perçue des logiciels. Nous ne pouvions donc plus faire l'économie de prendre ce facteur en considération.

En ce qui concerne le problème que nous avions soulevé, relatif à l'opérationnalisation de l'utilisabilité, nous avons proposé de mettre à profit le test d'utilisabilité de Nielsen. Bien qu'il ne consiste pas en des mesures objectives d'utilisabilité, nous avons pensé qu'il pouvait permettre aux utilisateurs de mieux catégoriser leur perception de l'interaction. L'idée sous-jacente étant de permettre aux concepteurs d'interfaces de pouvoir, à l'issue d'un test utilisateur, améliorer l'interface sur la base des jugements fournis par les participants, tout en pouvant anticiper l'augmentation que cela peut engendrer sur l'intention d'utilisation. Les résultats ont montré que le test de Nielsen a pu expliquer la facilité d'utilisation perçue du TAM à partir de la facilité d'apprentissage. Par contre, le test de Nielsen n'a pas permis de prévoir l'acceptabilité comparée, uniquement adossée à la facilité d'utilisation perçue (l'utilité théorique étant identique pour les deux interfaces). Cette facilité d'utilisation perçue a également eu une influence sur l'utilité perçue comme le prévoit la littérature (Adams, Neslon, & Todd, 1992; Davis, Bagozzi, & Warshaw, 1989; Morris & Dillon, 1997). A l'inverse, le test de Nielsen n'a pas eu d'effet sur l'utilité. Si la mesure d'efficience n'a pas révélé le potentiel prédictif que nous lui prétions par hypothèse, cela confirme d'autant plus la faiblesse de son implication dans l'élaboration des jugements des participants. Ceci reste cependant à considérer avec prudence. Des études antérieures ont montré que l'efficience était un antécédant de la facilité d'utilisation perçue (Arning & Ziefle, 2007; Venkatesh & Davis, 1996). Les particularités de ce protocole de comparaison aussi bien que le dispositif ont pu jouer un rôle modérateur (Wiedenbeck & Davis, 1997). Encore une fois, cette orientation scientifique est assez récente et l'on manque certainement de recul pour en apprécier toute la complexité. Cette étude aura néanmoins montré qu'il était possible et de façon simple de faire le lien entre utilisabilité et intention d'utilisation par le biais de mesures d'utilisabilité subjectives. En l'occurrence, c'est la facilité d'apprentissage qui a fait le pont entre la facilité d'utilisation perçue et la qualité de l'interaction.

Bien que nous n'ayons pas mesuré d'indicateur témoignant de la facilité de l'apprentissage comparée entre DIGIME et le clavier virtuel, il ne fait aucun doute que devoir réapprendre à écrire, est, pour des scripteurs experts, pour le moins frustrant. La facilité d'apprentissage est une dimension de l'utilisabilité aussi bien chez Shackel (1991) que chez Nielsen (1993). On a dans ce résultat la réponse à la question : pourquoi les logiciels de saisie et de reconnaissance de l'écriture manuscrite ne sont-ils pas acceptés ? Tout simplement parce qu'ils ne sont pas utilisables au sens le plus scientifique du terme. Souvenons nous que

l'utilisabilité n'existe pas intrinsèquement, elle s'appréhende en contexte. Lorsque les utilisateurs de PDA ou de Smartphones doivent saisir du texte dans leur machine, ils choisissent, à utilité égale, le logiciel de saisie le plus utilisable, c'est-à-dire celui qui a la plus grande utilité pratique. Les claviers AZERTY sont désormais entrés dans toutes les maisons, écoles, bureaux et lieux publics. Sans être dactylographe, tout le monde connaît et utilise les claviers matériels. Par conséquent, l'analogie que développent les claviers virtuels avec les claviers matériels les dispensent d'avoir à être appris. Ce n'est pas le cas des logiciels de saisie manuscrite dont l'apprentissage est difficile puisqu'il oblige à rompre des automatismes très anciens et très solides. Incidemment, un autre motif de rejet tient selon nous à la déception des utilisateurs d'interfaces de saisie manuscrite. DIGIME a été conçu, nous l'avons signalé, à partir de l'idée qu'il pouvait offrir une méthode de saisie intuitive (Bouteruche, Deconde, Anquetil, & Jamet, 2005). Les utilisateurs d'interface stylo attendent sans doute ce caractère intuitif, sans le trouver. Les difficultés d'apprentissage le montrent. Il ne nous appartient pas de discuter la conception même de ces logiciels. Cependant, les taux de reconnaissance théorique qu'ils affichent sont très éloignés de ce que les 6 expériences que nous avons menées permettent d'inférer.

A-t-on néanmoins amélioré l'utilsabilité de DIGIME ? Vraisemblement, du fait des améliorations ergonomiques que nos partenaires de l'IRISA lui ont apportées, et qui ont fait ponctuellement l'objet de tests, à l'intérieur des expériences 1 et 2. Il est clair qu'entre la première version de DIGIME (expérience 1, avec feedback contigu et conservation de l'encre) et la seconde (présence d'aides en ligne, feedback intégré sans conservation de l'encre, didacticiel), les performances ont été accrues. Un examen rapide des durées de saisie en fonction du nombre de lettres pour chaque expérience (indépendamment du contenu) aurait tendance à montrer que notre contribution est petite : les vitesses de saisie moyennes sont de l'ordre de 10 lettres par minute (expériences 1, 2, 5 et 5 bis). En fait, l'amélioration ne se situe pas tant sur les temps de saisie que sur le nombre d'erreurs commises, compensé par la consultation des aides en lignes. Les variations des conditions d'une étude à l'autre ne permettent pas de statuer précisément sur la question, mais c'est de toute évidence trop faible pour rendre DIGIME plus acceptable. Il en va de même si l'on examine l'utilisabilité subjective (satisfaction) et l'intention d'utilisation, identiques, pour les expériences 3 et 5 bis. Le prototype a pourtant bénéficié d'une importante évolution entre ces deux manipulations (suppression et remplacement du geste d'édition « entrée » par un bouton). Ceci peut

cependant s'expliquer par le fait que l'utilité reste le plus important déterminant de l'intention d'usage. Or, nous n'avons pas travaillé sur l'utilité même de DIGIME.

Peut-on alors, espérer améliorer l'utilisabilité de DIGIME et son acceptabilité ? Bien sûr. Le temps plaide en notre faveur par l'augmentation exponentielle des puissances de calcul de l'informatique nomade. D'autre part, toutes les solutions ergonomiques n'ont pas été testées sur DIGIME. Nous nous sommes principalement penchés sur celles qui relèvent du psychologue ergonome, c'est-à-dire celles qui permettent de favoriser les traitements cognitifs et faire diminuer la charge mentale. D'autres solutions existent, du côté de la machine notamment. On pensera entre autres aux systèmes qui apprennent l'écriture de leurs propriétaires ou les lexiques embarqués qui proposent des mots à partir des premières lettres tapées ou écrites (écriture T9 sur téléphones mobiles par exemple). MacKenzie et Chang (1999) ont notamment pointé la machine comme facteur limitant de la performance.

Ceci dit, nous pensons que l'acceptation de ses logiciels passera aussi par des processus de diffusion de l'innovation (Rogers, 1995). Deux exemples à ce propos. Nous avons été assez surpris de constater au fil des expériences qui se sont étendues de 2004 à 2008 l'évolution des mentalités sur les logiciels de saisie manuscrite. Pour ainsi dire complétement inconnus en 2004, la commercialisation de la *Nintendo DS* a fait pénétrer dans la sphère familliale les interfaces stylo dont est dotée cette console de jeu portable. Elle dispose d'un système de reconnaissance de l'écriture manuscrite. Si l'on admet l'idée que les enfants en sont les utilisateurs priviliégiés, l'acceptation de ces logiciels passera aussi par l'acculturation. L'expérience de Read (2007) donne à penser que l'exposition précoce des enfants apprentis scripteurs sera un facteur clé de la diffusion et de l'acceptation de ces dispositifs. Son étude a révélé que des enfants de 7 à 8 ans produisaient des textes plus longs sur tablette graphique que sur claviers matériel à niveau de compétence égale pour les deux méthodes de saisie.

En conclusion, nous mettrons en exergue les contributions scientifiques de ce travail. Celles-ci portent plus sur la compréhension du rôle de l'utilisabilité à l'acceptabilité, que précisément sur les remédiations susceptibles de favoriser l'acceptabilité de logiciels de reconnaissance de l'écriture manuscrite (particulièrement DIGIME).

Sur la mesure de l'utilisabilité tout d'abord. Bien que cette contribution soit modeste, et non déterminante pour la prédiction de l'acceptation, la mesure d'utilisabilité que nous avons proposée en nous inspirant des travaux de Huang, Hong, et Eades (2006, eux-même s'inspirant de Paas, Tuovinen, Tabbers, et Van Gerven (2003)) apporte une solution

mathématique à la quantification de l'utilisabilité dont tous les spécialistes s'accordent à dire qu'elle est multidimensionnelle. Ce caractère composite peut rendre son évaluation difficile. C'est particulièrement vrai au cours d'un cycle de conception itératif, si les mesures d'utilisabilité varient dans des directions différentes. Dès lors, comment faire la synthèse de ces variations pour déterminer objectivement comment l'utilisabilité a évolué ? La formule que nous avons proposée, en déterminant la distance orthogonale entre un point et un hyperplan de l'espace dans lequel il se situe permet de l'apprécier de façon fiable.

Sur les aides en lignes, dont les consultations répétées attestent de l'utilité. S'il est difficile de conclure à une supériorité d'un type de didacticiel (statique ou dynamique), leur présence a de toute évidence été accueillie positivement par les utilisateurs.

Pour finir, cette étude s'est inscrite dans une voie nouvelle de recherche, amené à prendre de l'importance de par la nécessité qui l'impose. En cherchant à comprendre les processus qui conduisent à l'acceptation à partir de mesures objectives et subjectives d'utilisabilité, nous avons dégagé des pistes de travail futures. En termes de représentations de l'utilisateur d'une part et, d'autre part sur des choix méthodologiques qui peuvent aider l'ergonome et le développeur d'IHM à déterminer les causes de failles dans la qualité d'interaction des utilsateurs avec les systèmes évalués. Cela avec l'assurance que les corrections apportées se traduiront de fait par un gain d'acceptabilité. Cette voie de travail et d'investigation se justifie d'autant plus qu'à une heure où la communauté MIS concentre ces efforts à déterminer la part des facteurs modérateurs dans l'acceptation, nous avons mis en évidence que cet aspect était indissociable de la performance.

.

11- Bibliographie

Adams, D. A., Nelson, R. R., & Todd, P. A. (1992). Perceived usefulness, ease of use and usage of information technology: A replication. *MIS Quarterly, 16*(2), 982-1003.

Agarwal, R., Ahuja, M., Carter, P. E., & Gans, M. (1998, December). *Early and late adopters of IT innovations: Extensions to innovation diffusion theory.* Paper presented at the Diffusion Interest Group of Information Technology, DIGIT'98, Helsinki, Finland.

Ajzen, I. (1985). From intentions to actions: A theory of planned behaviour. In J. Kuhl & J. Beckman (Eds.), *Action-Control: From Cognition to Behaviour*. Heidelberg, Germany: Springer.

Alamargot, D., Chanquoy, L., & Chuy, M. (2005). L'élaboration du contenu du texte: De la mémoire à long terme à l'environnement de la tâche. *Psychologie Française, 50*, 287-304.

Alamargot, D., Chesnet, D., Dansac, C., & Ros, C. (2006). Eye and Pen: A new device to study reading during writing. *Behaviour Research Methods, Instruments and Computers, 38*(2), 287-299.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-403.

Anderson, J. R. (1987). Skill Acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94*(2), 192-210.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51*, 355-365.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036-1060.

Anquetil, E., & Bouchereau, H. (2002, August). *Integration of an on-line handwriting recognition system in a smart phone device.* Paper presented at the 16th IAPR International Conference on Pattern Recognition, Québec.

Anquetil, E., & Bouteruche, F. (2004, June). *Conception d'un micro éditeur d'encre électronique et embarquement d'un système de reconnaissance d'écriture manuscrite sur téléphone mobile.* Paper presented at the Conférence Francophone: Mobilité et Ubiquité '04, Nice, France.

Anquetil, E., & Lorette, G. (1996, July). *Automatic generation of fuzzy classification systems based on explicit fuzzy rules deduced from possibilistic clustering: Application to on-line handwritten character recognition.* Paper presented at the information Processing and Management of Uncertainly in Knowledge-Based Systems, Granada, Spain.

Anquetil, E., & Lorette, G. (1997a). On-line handwriting character recognition system based on hierarchical qualitative fuzzy modeling. In A. C. Downton & S. Impevodo (Eds.), *Progress in Handwriting Recognition* (pp. 109-116): World Scientific.

Anquetil, E., & Lorette, G. (1997b, August). *Perceptual model of handwriting drawing, application to handwriting segmentation problem.* Paper presented at the International Conference on Document Analysis and Recognition, Ulm, Germany.

Arning, K., & Ziefle, M. (2007). Understanding age differences in PDA acceptance and performance. *Computers in Human Behaviour, 23*, 2904-2927.

Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 135-146). New York: Cambridge University Press.

Baccino, T., Bellino, C., & Colombi, T. (2005). *Mesure de l'utilisabilité des interfaces*. Paris: Hermes Science - Lavoisier.

Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.

Baddeley, A. D. (1993). *La mémoire humaine, théorie et pratique*. Grenoble: Presses Universitaires de Grenoble.

Baddeley, A. D. (1994). The magical number seven: Still magic after all these years? *Psychological Review, 101*(2), 353-356.

Baddeley, A. D. (1996, November). *The fractionation of working memory.* Paper presented at the Memory: Recording Experience in Cells and Circuits, Irvine, CA.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A. D. (2002a). Is working memory still working? *European Psychologist, 7*(2), 85-97.

Baddeley, A. D. (2002b). The psychology of memory. In A. D. Baddeley, M. D. Kopelman & B. A. Wilson (Eds.), *Handbook of Memory disorders* (2nd ed., pp. 3-15). Hove, UK: Psychology Press.

Baddeley, A. D., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology, 130*(4), 641-657.

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158-173.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York: Academic Press.

Badecker, W., Hillis, A., & Caramazza, A. (1990). Lexical morphology and its role in the writing process: Evidence from a case of acquired dysgraphia. *Cognition, 35*, 205-243.

Baeyens, C., & Bruyer, R. (1999). Dissociating visual and spatial components of the visuospatial slave system of working memory. *Cahiers de Psychologie Cognitive / Current Psychology of Cognition, 18*(3), 307-320.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.

Bastien, J. M. C., & Scapin, D. L. (2001). Evaluation des systèmes d'information et critères ergonomiques. In C. Kolski (Ed.), *Systèmes d'information et interactions homme-machine. Environnements évolués et évaluation des IHM. Interaction homme-machine pour les SI.* (Vol. 2, pp. 53-79). Paris: Hermes.

Beaudoin-Lafon, M., & Mackay, W. (2003). Prototyping tools and techniques. In J. A. Jacko & A. Sears (Eds.), *The Human Computer Interaction Handbook* (pp. 1118-1138). London: Lawrence Erlbaum Associates.

Benton, S. L., Kiewra, K. A., Whitfill, J. M., & Dennison, R. (1993). Encoding and external-storage effects on writing processes. *Journal of Educational Psychology, 85*(2), 267-280.

Bétrancourt, M., Bauer-Morrison, J., & Tversky, B. (2001). Les animations sont-elles vraiment plus efficaces? *Revue d'intelligence artificielle, 14*(1-2), 149-166.

Bétrancourt, M., & Tversky, B. (2000). Effect of computer animation on user's performance: A review. *Le Travail Humain, 63*(4), 311-329.

Bevan, N. (1995, July). *Usability is quality of use.* Paper presented at the 6th International Conference on Human Computer Interaction, Yokohama.

Bevan, N. (2001). International Standards for HCI and Usability. *International Journal of Human-Computer Studies, 55*(4), 533-552.

Bevan, N., Kirakowski, J., & Maissel, J. (1991, September). *What is usability?* Paper presented at the 4th International Conference on HCI, Stuttgart.

Bevan, N., & Macleod, M. (1994). Usability measurement in context. *Behaviour and Information Technology, 13*, 132-145.

Blake, R., & Shiffar, M. (2007). Perception of Human Motion. *Annual Review of Psychology, 58*, 47-73.

Bonin, P. (2003). *Production verbale de mots. Approche cognitive*. Bruxelles: DeBoeck Université.

Bonin, P. (2005). Comment accède-t-on à un mot en production verbale écrite? *Psychologie Française, 50*(3), 323-338.

Bonin, P., Pacton, S., & Fayol, M. (2001). La production verbale écrite: Evidences en faveur d'une (relative) autonomie de l'écrit. *Psychologie Française, 46*(1), 77-88.

Bonnardel, N., & Piolat, A. (2003). Design activities: How to analyse cognitive effort associated to cognitive treatments? *The International Journal of Cognitive Technology, 8*(1), 6-15.

Bourdin, B., & Fayol, M. (1994). Is written language production more difficult than oral language production? A working memory approach. *International Journal of Psychology, 29*(5), 591-620.

Bourdin, B., & Fayol, M. (2002). Even in adults, written production is still more costly than oral production. *International Journal of Psychology, 37*(4), 219-227.

Bouteruche, F., Deconde, G., Anquetil, E., & Jamet, E. (2005, September). *Design and evaluation of handwriting input interfaces for small-size mobile devices.* Paper

presented at the 1st Workshop on Improving and Assessing Pen-Based Input Techniques, Edinburgh, Scotland.

Brangier, E., & Barcenilla, J. (2003). *Concevoir un produit facile à utiliser*. Paris: Editions d'Organisation.

Brangier, E., Gronier, G., & Pino, P. (2001). La conception d'icônes permettant la communication entre de grands handicapés moteurs aphasiques et leur entourage: éléments de communication palliative. *Revue d'Interaction Homme-machine, 2*(2), 31-54.

Brave, S., & Clifford, N. (2003). Emotion in Human-Computer Interaction. In J. A. Jacko & A. Sears (Eds.), *The Human Computer Interaction Handbook* (pp. 1118-1138). London: Lawrence Erlbaum Associates.

Brookes, J. (1996). A Quick and Dirty Usability Scale. In P. W. Jordan, B. Thomas, W. B.A. & I. L. McClelland (Eds.), *Usability Evaluation in industry*. London: Taylor & Francis.

Brosnan, M. J. (1999). Modeling technophobia: A case of word processing. *Computers in Human Behaviour, 15*, 105-121.

Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology, 49*(2), 109-119.

Buisson, M., Sallagoïty, I., Athènes, S., & Mertz, C. (2003). *De l'analyse du mouvement à la conception des interfaces: Application à l'écriture et au geste.* Paper presented at the 15ème Conférence Francophone sur l'Interaction Homme-Machine, IHM'03, Université de Caen, France.

Burton-Jones, A., & Hubona, G. S. (2006). The mediation of external variables in the technology acceptance model. *Information & Management, 43*(6), 706-717.

Caramazza, A., & Miceli, G. (1990). The structure of graphemic representations. *Cognition, 37*, 243-297.

Carlopio, J. (1988). A history of social psychological reactions to new technology. *Journal of Occupational Psychology, 61*, 67-77.

Carpenter, P. A., Just, M. A., & Reichle, E. D. (2000). Working memory and executive fonction: Evidence from neuroimaging. *Current Opinion in Neurobiology, 10*, 195-199.

Castañeda, J. A., Muñoz-Leiva, F., & Luque, T. (2007). Web Acceptance Model (WAM): Moderating affects of user experience. *Information & Management, 44*, 384-396.

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*, 233-246.

Channouf, A., Py, J., & Somat, A. (1996). Prédire des comportements à partir des attitudes: Nouvelles perspectives. In J.-C. Deschamps & J.-L. Beauvois (Eds.), *Des attitudes aux attributions* (pp. 55-65). Grenobe: Presses universitaires de Grenobles.

Chanquoy, L., & Alamargot, D. (2002). Mémoire de travail et rédaction de textes: Evolution des modèles et bilan des premiers travaux. *L'Année Psychologique, 102*, 363-398.

Chesnet, D., & Alamargot, D. (2005). Analyse en temps réel des activités oculaires et grapho-motrices du scripteur. Intérêt du dispositif "Eye and pen". *L'Année Psychologique, 105*(3), 477-520.

Chevalier, N., Hall, C., & Nadeau, C.-H. (1990). Imagerie et répétition mentale du mouvement: Perspectives de recherche. *STAPS, 11*(22), 7-9.

Christensen, C. A. (2004). Relationship between orthographic-motor integration and computer use for the production of creative and well-structured written text. *British Journal of Educational Psychology, 74*, 551-564.

Chua, R., Weeks, D. J., & Goodman, D. (2003). Perceptual-motor interaction: Some implications for human-computer interaction. In J. A. Jacko & A. Sears (Eds.), *The Human Computer Interaction Handbook* (pp. 1118-1138). London: Lawrence Erlbaum Associates.

Clark, R., Nguyen, F., & Sweller, J. (2006). An introduction to efficiency in learning. In *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. San Francisco, USA: Pfeiffer.

Cockton, G., Lavery, D., & Woolrych, A. (2003). Inspection-based evaluations. In J. A. Jacko & A. Sears (Eds.), *The Human Computer Interaction Handbook* (pp. 1118-1138). London: Lawrence Erlbaum Associates.

Commaford, P. M. (2004). An investigation of text thoughput speeds associated with pocket PC input method editors. *International Journal of Human Computer Interaction, 17*(3), 293-308.

Content, A., Mousty, P., & Radeau, M. (1990). BRULEX. Une base de données lexicales informatisée pour le français écrit et parlé. *Année Psychologique, 90*(4), 551-566.

Cooper, G. (1990). Cognitive load theory as an aid for instructional design. *Australian Journal of Educational Technology, 6*(2), 108-113.

Cooper, G. (1998). Research into cognitive load theory and instructional design at UNSW. Retrieved Mai, 2004, sur internet à l'url: http://wwwarts.unsw.edu.au/education/CLT.html

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences, 24*(87-185).

Cox-Fuenzalida, L.-E., & Angie, A. D. (2005). The effects of workload history on dual task performance. *Current Psychology: Developmental, Learning, Personality, Social, 24*(3), 171-179.

Curran, K., Woods, D., & O Riordan, B. (2006). Investigating text input methods for mobile phones. *Telematics and Informatics, 23*, 1-21.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Darabi, A. (2004, october). *Measuring motivation in instruction and training: A cognitive load perspective.* Paper presented at the 46th International Military Testing Association Annual Conference, Brussels.

Darling, S., Della Sala, S., & Logie, R. H. (2007). Behavioural evidence for separating components of visuo-spatial working memory. *Cognitive Processing, 8*, 175-181.

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*, 176-180.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoritical models. *Management Science, 35*(8), 982-1003.

Davis, F. D., & Venkatesh, V. (1996). A critical assessment of potential measurement biases in the technology acceptance model: Three experiments. *International Journal of Human Computer Studies, 45*, 19-45.

Delinières, D., & Deschamps, T. (2000). L'effort mental. In D. Delinières (Ed.), *L'effort* (pp. 25-40). Paris: Edition Revue EPS.

Della Sala, S., & Logie, R. H. (2002). Working memory. In *Encyclopedia of Human Brain* (Vol. 4, pp. 819-830).

Denis, M., & De Vega, M. (1993). Modèles mentaux et imagerie mentale. In M. F. Ehrlich (Ed.), *Les modèles mentaux. Approche cognitive des représentations*. Paris: Masson.

Dillon, A. (2001a). The evaluation of software usability. In W. Karwowski (Ed.), *Encyclopedia of Human Factors and Ergonomics*. London: Taylor and Francis.

Dillon, A. (2001b). User acceptance of information technology. In W. Karwowski (Ed.), *Encyclopedia of Human Factors and Ergonomics*. London: Taylor and Francis.

Dillon, A., & Morris, M. (1997). User acceptance of new information technology: Theories and models. *Annual Review of Information Science and Technology, 31*, 3-32.

Dillon, A., & Watson, C. (1996). User analysis in HCI: The historical lesson from individual differences research. *International Journal of Human Computer Studies, 45*(6), 619-637.

Dobbie, L., & Askov, E. N. (1995). Progress of handwriting research in the 1980's and futur prospects. *The Journal of Educational Research, 88*(6), 339-351.

Drommi, A., Ulferts, G. W., & Shoemaker, D. (2001, November). *Interface design: A focus on cognitive science.* Paper presented at the 21st Annual Information systems Education Conference, ISECON 2001, Cincinnati, USA.

Dumas, J. S. (2003). User-based evaluations. In J. A. Jacko & A. Sears (Eds.), *The Human Computer Interaction Handbook* (pp. 1118-1138). London: Lawrence Erlbaum Associates.

Eason, K. D. (1984). Towards the experimental study of usability. *Behaviour and Information Technology, 3*(2), 133-143.

Erhel, S., & Jamet, E. (2006). Using pop-up windows to improve multimedia learning. *Journal of Computer Assisted Learning, 22*, 137-147.

Ericsson, K. A. (2005). Recent advances in expertise research: A commentary on the contributions to the special issue. *Applied Cognitive Psychology, 19*, 233-241.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211-245.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of delibarate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task. *Annual review of Psychology, 47*, 273-305.

Farrington-Darby, T., & Wilson, J. R. (2006). The nature of expertise: A review. *Applied Ergonomics, 37*, 17-32.

Favart, M., & Olive, T. (2005). Modèles et méthodes d'études de la production écrite. *Psychologie Française, 50*, 273-285.

Fayol, M., & Gombert, J. E. (1999). L'apprentissage de la lecture et de l'écriture. In J. A. Rondal & E. Espéret (Eds.), *Manuel de Psychologie de l'Enfant* (pp. 565-594). Bruxelles: Mardaga.

Féry, Y.-A. (2001). Que savons-nous de nos mouvements? *STAPS, 55*, 7-22.

Fiorello, A. P. (1999). *Power, Performance, and Perception (P3): Integrating usability metrics and technology acceptance determinants to validate a new model for predicting system usage.* Air Force Institute of Technology Air University, Wright-Patterson AFB.

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention and Behaviour: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.

Fogg, B. J. (2003). Motivating, influencing, and persuading users. In J. A. Jacko & A. Sears (Eds.), *The Human Computer Interaction Handbook* (pp. 1118-1138). London: Lawrence Erlbaum Associates.

Folmer, E., & Bosch, J. (2003). Architecting for usability: A survey. *Journal of systems and software, 70*(1), 61-78.

Folmer, E., Van Gurp, J., & Bosch, J. (2003). A framework for capturing the relashionship betwenn usability and software architecture. *Software Process: Improvement and Practice, 8*(2), 67-87.

Fortis, J.-M. (1994). Image mentale et représentation propositionnelle. *Intellectica*(19), 253-305.

Frokjaer, E., Hertzum, M., & Hornbaek, K. (2000, April). *Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated?* Paper presented at the ACM CHI 2000 Conference on Human Factors in Computing Systems, The Hague, The Netherlands.

Gaonac'h, D., & Larigauderie, P. (2000). *Mémoire et fonctionnement cognitif*. Paris: Armand Colin.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language, 31*, 129-151.

Graham, S., Weintraub, N., & Berninger, V. W. (2001). The relationship between handwriting style and speed and legibility. *Journal of Educational Research, 91*(5), 290-296.

Gray, W. D., & Salzman, M. C. (1998). Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction, 13*(3), 203-261.

Griggs, P., & Carot, R. (2002). La dimension cognitive dans l'apprentissage des langues étrangères. *Revue Française de Linguistique Appliquée, 2*(72), 25-38.

Guida, A., Tardieu, H., & Nicolas, S. (sous presse). Mémoire de travail à long terme : Quelle est l'utilité de ce concept ? Emergence, concurrence et bilan de la théorie d'Ericsson et Kintsch (1995). *L'Année Psychologique*.

Gyselinck, V. (1995). *Les modèles mentaux dans la compréhension de textes: Le rôle des illustrations.* Université René Descartes, Paris V.

Gyselinck, V. (1996). Illustrations et modèles mentaux dans la compréhension de textes. *L'Année Psychologique, 96*(3), 195-226.

Gyselinck, V., Jamet, E., & Dubois, V. (2008). The role of working memory component in multimedia comprehension. *Applied Cognitive Psychology, 22*, 353-374.

Hagger, M. S., Chatzisarantis, N. L. D., & Biddle, S. J. H. (2003). A meta-analytic review of the theories of reasoned action and planned behaviour in physical activity: An examination of predictive validity and the contribution of additional variables. *Journal of Sport and Exercise Psychology, 24*(1), 3-32.

Ham, D.-H., Heo, J., Fossick, P., Wong, W., Park, S., Song, C., et al. (2006, November). *Conceptual framework and models for identifying and organizing usability impact factors of mobile phones.* Paper presented at the Australian Computer-Human Interaction Conference, OZCHI 2006, Sydney, Australia.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 239-250). Amsterdam: North Holland Press.

Hartley, J. (1991). Sex differences in handwriting: A comment on Spear. *British Educational Research Journal, 17*(2), 141-146.

Hartley, J. (2002). Notetaking in non-academic settings: A review. *Applied Cognitive Psychology, 16*, 559-574.

Hasan, B., & Ahmed, M. U. (2007). Effects of interface style on user perceptions and behavioral intention to use computer systems. *Computers in Human Behaviour, 23*, 3025-3037.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 1-27). Mahwah (NJ): Lawrence Erlbaum.

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing*. Hillsdale, NJ: Lawrence Erlbaum.

Hayes, J. R., & Flower, L. S. (1986). Writing research and the writer. *Amercian Psychologist, 41*(10), 1106-1113.

Hidrio, C., & Jamet, E. (2002, August). *Comprehension of a spoken text: Effects of different types of illustrations.* Paper presented at the EARLI-SIG Conference on Multimedia Comprehension, Poitiers, France.

Hidrio, C., & Jamet, E. (2008). Learning from a multimedia explanation: A comparison of static pictures and animation. In J.-F. Rouet, R. Lowe & W. Schnotz (Eds.), *Understanding Multimedia Documents* (pp. 103-118). New York: Springer.

Hidrio, C., Jamet, E., & Le Bohec, O. (2002, September). *Using dynamical illustrations and oral explanations in web sites: An application to a search engine's syntax.* Paper presented at the 11th European Conference on Cognitive Ergonomics (ECCE11) on Cognition, Culture and Design, Catania, Italy.

Hinckeley, K. (2003). Input technologies and techniques. In J. A. Jacko & A. Sears (Eds.), *The Human Computer Interaction Handbook* (pp. 151-168). London: Lawrence Erlbaum Associates.

Hom, J. (1998). The Usability Methods Toolbox Handbook. 2005, sur internet à l'url: http://jthom.best.vwh.net/usability/usable.htm

Hornbæk, K. (2006). Current practice in measuring usability : Challenges to usability studies and research. *International Journal of Human-Computer Studies,* 64, 79-102.

Howell, D. C. (1997). *Méthodes statitstiques en sciences humaines*. Paris: De Boeck.

Huang, W., Hong, S.-H., & Eades, P. (2006). *Predicting graph reading performance: A cognitive approach*. Sidney, Australia: IMAGEN Program, National ICT Australia School of Information Technologies, University of Sydney.

ISO-9241-11. (1998). *Ergonomics requirements for office work with visual display terminals. Part 11: Guidance on usability*. Geneva: ISO.

ISO/IEC-9126-1. (2000). *Software engineering - Product quality - Part 1: Quality model*. Geneva: ISO.

Jamet, E. (2002). L'apport des nouvelles technologies de l'information et de la communication dans la conception des documents techniques. *Psychologie Française, 47*(1), 33-40.

Jamet, E. (2006). Une présentation des principales méthodes d'évaluation des EIAH en psychologie cognitive. *Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation (STICEF), 13*, 129-146.

Jamet, E., & Erhel, S. (2006). Les effets de l'intégration spatiale de fenêtres ponctuelles sur la compréhension de documents pédagogiques illustrés. *Psychologie Française, 51*(1), 73-86.

Jamet, E., & Février, F. (2008). Utilisabilité, utilité et acceptabilité des nouvelles technologies dans l'entreprise: Une approche de psychologie ergonomique. In M. Benedetto-Meyer & R. Chevallet (Eds.), *Analyser les usages des systèmes d'information et des TIC : quelles démarches, quelles méthodes ?* Paris: Editions de l'Agence Nationale pour l'Amélioration des Conditions de Travail (ANACT).

Jamet, E., Le Bohec, O., & Hidrio, C. (2003). Comment présenter l'information dans les documents numériques éducatifs? Une approche de psychologie cognitive. *Document numérique, 7*(1-2), 1-14.

Jansen, W., & Ayers, R. (2004). *Guidelines on PDA forensics* (No. Special Publication 800-72). Gaithersburg: National Institute of Standards and Technology, Technology Administration, U.S. Department of Commerce.

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia, 33*(11), 1419-1432.

Jeannerod, M., & Frak, V. (1999). Mental imaging of motor activity in humans. *Current Opinion in Neurobiology, 9*, 735-739.

Jones, D., & Christensen, C. A. (1999). Relationship between automaticity in handwriting and students' ability to generate written text. *Journal of Educational Psychology, 91*(1), 44-49.

Jones, D. M. (2002, october). *The 7±2 urban legend.* Paper presented at the Motor Industry Software Reliability Association Conference, MISRA Conference Proceedings.

Jones, S., & Scaife, M. (1999, June). *Diagram representation: A comparison of animated and static format.* Paper presented at the ED MEDIA 99: World conference on educational multimedia, hypermedia and telecommunications Seattle, USA.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122-149.

Kaikkonen, A., Kallio, T., Kekäläinen, A., Kankainen, A., & Cankar, M. (2005). Usability testing of mobile applications: A comparison between laboratory and field. *Journal of Usability Studies, 1*(1), 4-46.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*(1), 1-17.

Kandel, S., Alvarez, C. J., & Vallée, N. (2006). Syllabes as processing units in handwriting production. *Journal of Experimental Psychology: Human perception and performance, 32*(1), 18-31.

Kandel, S., Valdois, S., & Orliaguet, J.-P. (2003). Etude de la production écrite en copie: Une approche visuo-orthographique et graphomotrice. *Le Langage et l'Homme, 38*(2), 5-24.

Kaufman, J. (2006, November). Card sorting: An inexpensive and practical usability technique. *Intercom,* pp. 17-19.

Keinonen, T. (1998). *One-dimensional usability: Influence of usability on consumers' product preference.* Helsinki: University of Art and Design Helsinki.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 57-71). Mahwah, NJ: Lawrence Erlbaum.

Kellogg, R. T. (2001a). Competition for working memory among writing processes. *American Journal of Psychology, 114*(2), 175-191.

Kellogg, R. T. (2001b). Long-term working memory in text production. *Memory and Cognition, 29*(1), 43-52.

Kellogg, R. T. (2004). Working memory components in written sentence generation. *American Journal of Psychology, 117*(3), 341-361.

Kellogg, R. T., Olive, T., & Piolat, A. (2007). Verbal, visual, and spatial working memory in written language production. *Acta Psychologica, 124*, 382-397.

Ketola, P., & Röykkee, M. (2001, April). *The three facets of usability in mobile handsets Nokia.* Paper presented at the CHI 2001, Seattle.

King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & Management, 43*, 740-755.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.

Kobayashi, K. (2005). What limits the encoding effect of note-taking? A meta-analytic examination. *Contemporary Educational Psychology, 30*, 242-262.

Kolski, C. (1997). *Interfaces homme-machine. Application aux systèmes industriels complexes.* Paris: Hermes.

Költringer, T., & Grechenig, T. (2004, April). *Comparing the immediate usability of Graffity 2 and virtual keyboard.* Paper presented at the CHI '04 extended abstracts on human factors in computing systems, Vienna, Austria.

Koohang, A. (2004). Expanding the concept of usability. *Informing Science Journal, 7*, 129-141.

Lee, F. J., & Anderson, J. R. (1997, August). *Learning to act: Acquisiton and optimization of procedural skill.* Paper presented at the 19th Annual Conference of the Cognitive Science Society, Stanford.

Legris, P., Ingham, J., & Collerette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. *Information & Management, 40*, 191-204.

Lewis, J. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation. *International Journal of Human-Computer Interaction, 7*(1), 57-78.

Lieury, A. (1997). *Manuel de psychologie générale*. Paris: Dunod.

Logie, R. H., Engelkamp, J., Dehn, D., & Rudkin, S. J. (2000). Actions, mental actions, and working memory. In M. Denis, R. H. Logie, C. Cornoldi & J. Engelkamp (Eds.), *Imagery, Language and Visuo-spatial thinking*. Hove, UK: Psychology Press.

Logie, R. H., & Marchetti, C. (1991). Visuo-spatial working memory: Visual, spatial or central executive? In R. H. Logie & M. Denis (Eds.), *Mental images in human cognition* (pp. 105-115). Amsterdam: Elsevier Sciences Publishers.

Logie, R. H., & Pearson, D. G. (1997). The inner eye and the inner scribe of visuo-spatial working memory: Evidence for developmental fractionation. *European Journal of cognitive psychology, 9*(3), 241-257.

Longcamp, M., Anton, J.-L., Roth, M., & Velay, J.-L. (2003). Visual presentation of single letters activates a premotor area involved in writing. *NeuroImage, 19*, 1492-1500.

Longcamp, M., Boucard, C., Gilhodes, J.-C., Anton, J.-L., Nazarian, B., & Velay, J.-L. (2008). Learning through hand- or type-writing influences visual recognition of new graphic shapes: Behavioral and functional imaging evidence. *Journal of Cognitive Neuroscience, 20*(5), 802-815.

Longcamp, M., Boucard, C., Gilhodes, J.-C., & Velay, J.-L. (2006). Remembering the orientation of newly learned characters depends on the associated writing knowledge: A comparison between handwriting and typing. *Human Movement Science, 25*, 646-656.

Longstaff, M. G., & Heath, R. A. (1997). Space-time invariance in adult handwriting. *Acta Psychologica, 97*, 201-214.

MacKenzie, I. S., & Chang, L. (1999). A performance comparison of two handwriting recognizers. *Interacting with Computers, 11*, 283-297.

MacKenzie, I. S., Nonnecke, B., McQueen, C., Riddersma, S., & Meltz, M. (1994, October). *A comparison of three methods of character entry on pen-based computers.* Paper presented at the Human Factors and Ergonomics Society, 38th annual Meeting, Santa Monica, CA.

MacKenzie, I. S., Nonnecke, B., Riddersma, S., McQueen, C., & Meltz, M. (1994). Alphanumeric entry on pen-based computers. *International Journal of Human-Computer Studies, 41*, 775-792.

MacKenzie, I. S., & Soukoreff, W. R. (2002). Text entry for mobile computing: Models and methods, theory and practice. *Human-Computer Interaction, 17*, 147-198.

MacKenzie, I. S., & Soukoreff, W. R. (2003). Phrase sets for evaluating text entry techniques. In *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems - CHI 2003* (pp. 754-755). New York: ACM.

MacKenzie, I. S., & Zhang, S. (1997, May). *The immediate usability of Graffiti.* Paper presented at the Proceedings of Graphics Interface '97, Toronto.

Marcus, A. (2003). Global and intercultural user-interface design. In J. A. Jacko & A. Sears (Eds.), *The Human Computer Interaction Handbook* (pp. 442-463). London: Lawrence Erlbaum Associates.

Martein, R., Kemps, E., & Vandierendonck, A. (1999). The role of working memory in a double span task. *Psychologica Belgica, 39*(1), 15-27.

Martin, C. (2003). *Téléphone portable chez les jeunes adolescents et leurs parents: Quelle légitimation des usages?* Paper presented at the 2ème Workshop de Marsouin, Brest.

Mautone, P. D., & Mayer, R. E. (2001). Signaling as cognitive guide in multimedia learning. *Journal of Educational Psychology, 93*(2), 377-389.

Maxwell, J. P., Masters, R. S. W., & Eves, F. F. (2003). The role of working memory in motor learning and performance. *Consciousness and Cognition, 12*(3), 276-402.

Mayer, R. E. (2001). *Multimedia Learning*. Cambridge, UK: Cambridge University Press.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review, 8*(3), 299-325.

McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational psychologist, 35*(1), 13-23.

McCutchen, D., Covill, A., Hoyne, S. H., & Mildes, K. (1994). Individual differences in writing: Implications of translating fluency. *Journal of Educational Psychology, 86*(2), 256-266.

McFarland, D. J., & Hamilton, D. (2006). Adding contextual specificity to the technology acceptance model. *Computers in Human Behaviour, 22*, 427-447.

McQueen, C., MacKenzie, I. S., Nonnecke, B., Riddersma, S., & Meltz, M. (1994, May). *A comparison of four methods of numeric entry on pen-based computers.* Paper presented at the Graphics Interface '94, Toronto, canada.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Miyake, A., & Shah, P. (1999). Models of working memory: An introduction. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanism of active maintenance and executive control* (pp. 1-27). New York: Cambridge University Press.

Morahan-Martin, J., & Schumacher, P. (2007). Attitudinal and experiential predictors of technological expertise. *Computers in Human Behaviour, 23*, 2230-2239.

Moroney, W. F., Biers, D. W., & Eggemier, F. T. (1995). Some measurement and methodological considerations in the application of subjective workload mesurement techniques. *The International Journal of Aviation Psychology, 5*(1), 87-106.

Morris, M. G., & Dillon, A. (1997). The influence of user perceptions on software utilisation: Application and evaluation of a theoretical model of technology acceptance. *IEEE Software, 14*(4), 58-69.

Mounoud, P. (1986). Action and cognition. Cognitive and motor skills in a developmental perspective. In M. G. Wade & H. T. A. Whiting (Eds.), *Motor development in children* (pp. 373-390). Dordrecht: M. Nijhoff.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., & Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *Journal of Neurophysiology 78*(4), 2226-2230.

Najjar, L. J. (1996). Multimedia information and learning. *Journal of Educational Multimedia and Hypermedia, 5*, 129-150.

New, B., & Pallier, C. (2001). Fréquence des lettres. Retrieved 12/2005, 2005, from http://www.lexique.org/listes/liste_lettres.php

Nielsen, J. (1993). *Usability engineering*. Boston: Academic Press.

Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability Inspection Methods*. New York, NY.

Nielsen, J., & Levy, J. (1994, April). *Measuring usability: Preference vs. performance.* Paper presented at the Communications of the ACM.

Olive, T. (2004). Working memory in writing: Empirical evidences from the dual-task technique. *European Psychologist, 9*(1), 32-42.

Olive, T., & Kellogg, R. T. (2002). Concurrent activation of high- and low-level production processes in written composition. *Memory and Cognition, 30*, 594-600.

Olive, T., Kellogg, R. T., & Piolat, A. (2002). Studying text production with the triple task technique. In T. Olive & C. M. Levy (Eds.), *Contemporary tools and techniques for studying writing* (pp. 31-58). Dordrecht: Kluwer Academic Publisher.

Olive, T., & Piolat, A. (2002). Suppressing visual feedback in written composition: Effects on processing demands and coordination of the writing processes. *International Journal of Psychology, 37*(4), 209-218.

Olive, T., & Piolat, A. (2005). Le rôle de la mémoire de travail dans la production écrite de textes. *Psychologie Française, 50*, 373-390.

Ong, C.-S., & Lai, J.-Y. (2006). Gender differences in perceptions and relationships among dominants of e-learning acceptance. *Computers in Human Behaviour, 22*, 816-829.

Onwuegbuzie, A. J., & Daniel, L. G. (2003). Typology of analytical and interpretational errors in quantitative and qualitative educational research [Electronic Version]. *Issues in Education*, 6. Retrieved Juillet 2006 from http://cie.ed.asu.edu/volume6/number2/.

Ortelli, R. (2003). Styles d'interaction dans les pocket PC: Analyses et comparaisons. Université de Genève, Faculté de Psychologie et des Sciences de l'Education, TECFA (TeChnologies des Formations et de l'Apprentissages).

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational psychologist, 38*(1), 1-4.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional science, 32*, 1-8.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist, 38*(1), 63-71.

Paas, F., & Van Merriënboer, J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*, 737-743.

Pacton, S., & Perruchet, P. (2006). L'apprentissage implicte: du labo à l'école. In P. Dessus & E. Gentaz (Eds.), *Comprendre les apprentissages* (Vol. 2, pp. 59-73). Paris: Dunod.

Paillard, J. (1988). Les bases nerveuses du contrôle visuo-manuel de l'écriture. In C. Sirat, J. Irigoin & E. Poulle (Eds.), *L'écriture, l'oeil, le cerveau et la main* (Bibliologia ed., Vol. 10, pp. 23-58). Paris: Brepols-Turnhout.

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review, 76*, 241-263.

Palaigeorgiou, G. E., Despotakis, T. D., Demetriadis, S., & Tsoukalas, I. A. (2006). Synergies and barriers with electronic verbatim notes (eVerNotes): Note taking and report writing with eVerNotes. *Journal of Computer Assisted Learning, 22*, 74-85.

Park, O., & Hopkins, R. (1993). Instructional conditions for using dynamic visual displays: A review. *Instructional science, 21*, 427-449.

Peeples, E. D., & Retzlaff, P. D. (1991). A component analysis of handwriting. *The Journal of General Psychology, 118*(4), 369-374.

Pélissier, A., & Piolat, A. (1999). *Are writing processes activated sequentially or in parallel? The way to appropriety use Kellogg's three task method.* Paper presented at the The 1998 European Writing Conference: Writing and Learning to Write at the dawn of the 21st Century, Poitiers: MSHS.

Perkins, D. N., & Salomon, G. (1994). Transfer of learning. In T. Husen & T. N. Postelwhite (Eds.), *International Handbook of Educational Research* (2nd ed., Vol. 11, pp. 6452-6457). Oxford: Pergamon Press.

Perruchet, P. (1988). Une évaluation critique du concept d'automaticité. In P. Perruchet (Ed.), *Les automatismes cognitifs* (pp. 27-54). Liège: Mardaga.

Peterson, L., & Peterson, M. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193-198.

Peverly, S. T., Ramaswamy, V., Brown, C., Sumowski, J., Alidoost, M., & Garner, J. (2007). What predicts skill in lecture note taking? *Journal of Educational Psychology, 99*(1), 167-180.

Piolat, A. (2004a). Approche cognitive de l'activité rédactionnelle et de son acquisition. Le rôle de la mémoire de travail. *LINX (Linguistique Institut Nanterre Paris X), 51*, 55-74.

Piolat, A. (2004b). La prise de notes: Ecriture de l'urgence. In A. Piolat (Ed.), *Ecriture: Approches en sciences cognitives* (pp. 205-229). Aix-en-Provence: Publications de l'université de Provence.

Piolat, A. (2006). *La prise de notes* (2ème ed.). Paris: Presses Universitaires de France.

Piolat, A., Kellogg, R. T., & Farioli, F. (2001). The triple task technique for studying writing processes: On which task is attention focused? *Current Psychology Letters: Behaviour, Brain and Cognition, 4*, 67-83.

Piolat, A., & Olive, T. (2000). Comment étudier le coût et le déroulement de la rédaction de textes? La méthode de la triple tâche: Un bilan méthodologique. *L'Année Psychologique, 100*, 465-502.

Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology, 19*, 291-312.

Piolat, A., Olive, T., Roussey, J. Y., Thunin, O., & Ziegler, J. C. (1999). SCRIPTKELL: A tool for measuring cognitive effort and time processing in writing and other complex cognitive activities. *Behaviour Reaserch Methods, Instruments & Computers, 31*(1), 113-121.

Piolat, A., Roussey, J. Y., & Barbier, M.-L. (2003). Mesure de l'effort cognitif: Pourquoi est-il opportun de comparer la prise de notes à la rédaction, l'apprentissage et la lecture de divers documents? *Arob@se, 1-2*, 118-140.

Piolat, A., Roussey, J. Y., Olive, T., & Farioli, F. (1996). Charge mentale et mobilistation des processus rédactionnels: Examen de la procédure de Kellogg. *Psychologie Française, 41*(4), 339-354.

Porter, C. E., & Donthu, N. (2006). Using the technology acceptance model to explain how attitudes determine internet usage: The role of perceived access barriers and demographics. *Journal of Business Research, 59*, 999-1007.

Rauterberg, M. (1996, November). *How to measure and to quantify usability attributes of man-machine interfaces.* Paper presented at the 5th IEEE International Workshop on Robot and Human Communication - ROMAN'96, Tsukuba, Japan.

Read, J. C. (2007). A study of usability of handwriting recognition for text entry by children. *Interacting with Computers, 19*, 57-69.

Reid, G. B., Nygren, T. E., Hancock, P. A., & Meshkati, N. (1998). The Subjective Workload Assessment Technique: A scaling procedure for measuring mental workload. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 185-218). North-Holland: Elsevier Science Publishers.

Rickel, J., & Johnson, W. L. (1999). Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence, 13*, 343-382.

Rogers, E. M. (1995). *Diffusions of Innovations* (4th ed.). New York: Free Press.

Rothrock, L., Barron, K., Simpson, T. W., Frecker, M., Ligetti, C., & Barton, R. R. (2006). Applying the proximity compatibility and the control-display compatibility principles to engineering design interface *Human Factors and Ergonomics in Manufacturing, 16*(1), 61-81.

Roussey, J. Y., & Piolat, A. (2003). Prendre des notes et apprendre. Effet du mode d'accès à l'information et de la méthode de prise de notes. *Arob@se, 1-2*, 47-68.

Roussey, J. Y., & Piolat, A. (2005). La révision du texte: Une activité de contrôle et de réflexion. *Psychologie Française, 50*, 351-372.

Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and Workload Profile methods. *Applied Psychology, 53*(1), 61-86.

Rudkin, S. J., Pearson, D. G., & Logie, R. H. (2007). Executive processes in visual and spatial working memory tasks. *The Quarterly Journal of Experimental Psychology, 60*(1), 79-100.

Saadé, R. G., & Kira, D. (2007). Mediating the impact of technology usage on perceived ease of use by anxiety. *Computers and Education, 49*(4), 1189-1204.

Schepers, J., & Wetzels, M. (2007). A meta-analysis of the technology acceptance model: Investigating subjective norm and moderation effects. *Information & Management, 44*, 90-103.

Sears, A., & Arora, R. (2002). Data entry for mobile devices: An empirical comparison of novice performance with Jot and Graffiti. *Interacting with Computers, 14*(5), 413-433.

Segall, N., Doolen, T. L., & Porter, J. D. (2005). A usability comparison of PDA-based quizzes and paper-and pencil quizzes. *Computers and Education, 45*, 417-432.

Serratrice, G., & Habib, M. (1993). *L'écriture et le cerveau, mécanismes neuro-physiologiques*. Paris: Masson.

Shackel, B. (1991). Usability - context, framework, design and evaluation. In B. Shackel & B. Richardson (Eds.), *Human Factors for Informatics Usability* (pp. 21-38). Cambridge: Cambridge University Press.

Shackel, B. (2000). People and computers - some recent highlights. *Applied Ergonomics, 31*, 595-608.

Shaffer, D., Doube, W., & Tuovinen, J. E. (2003, April). *Applying cognitive load theory to computer science education.* Paper presented at the Psychology of Programming Interest Group, PPIG 2003, Keele, UK.

Snyder, C. (2003). *Paper Prototyping*. San Francisco: Morgan Kaufmann Publishers.

Soukoreff, W. R., & MacKenzie, I. S. (2004, April). *Recent developments in text-entry error rate measurement.* Paper presented at the ACM Conference on Human factors in Comuting Systems, CHI 2004, New York.

Stewart, T., & Travis, D. (2003). Guidelines, standards, and style guides. In J. A. Jacko & A. Sears (Eds.), *The Human Computer Interaction Handbook* (pp. 1118-1138). London: Lawrence Erlbaum Associates.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(1), 257-285.

Sweller, J. (1999). *Instructional Design in Technical Areas*. Melbourne: ACER.

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional science, 32*, 9-31.

Sweller, J., Van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

Taatgen, N. A., & Lee, F. J. (2003). Production Compilation: A simple mechanism to model Complex Skill Acquisition. *Human Factors, 45*(1), 61-76.

Thomas, R. (1997). *L'apprentissage moteur*. Paris: Presses Universitaires de France.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 8*, 247-261.

Tornatzky, L. G., & Klein, K. J. (1982). Innovation characteristics and innovation adoption-implementation: A meta-analysis of findings. *IEEE Transactions on Engineering Management, 29*, 28-45.

Tractinsky, N., & Hassenzahl, M. (2005). Arguing for aesthetics in human-computer interaction. *I-Com, 3*, 66-68.

Treisman, A. M., & Kanwischer, N. G. (1998). Perceiving visually presented objects: recognition, awareness, and modularity. *Current Opinion in Neurobiology, 8*(21), 218-226.

Tricot, A. (1998). Charge cognitive et apprentissage. Une présentation des travaux de John Sweller. *Revue de Psychologie de l'Education, 3*, 37-64.

Tricot, A., & Chanquoy, L. (1996). La charge mentale, "vertu dormitive" ou concept opérationnel? Introduction. *Psychologie Française, 41*(4), 313-318.

Tricot, A., Plégat-Soutjis, F., Camps, J.-F., Amiel, A., Lutz, G., & Morcillo, A. (2003). Utilité, utilisabilité, acceptabilité: Interpréter les relations entre trois dimensions de l'évaluation des EIAH. In C. Desmoulins, P. Marquet & D. Bouhineau (Eds.), *Environnements informatiques pour l'apprentissage humain* (pp. 391-402). Paris: ATIEF/ INRP.

Tricot, A., & Tricot, M. (2000). *Un cadre formel pour interpréter les liens entre utilisabilité et utilité des systèmes d'information (et généralisation à l'évaluation d'objets finalisés).* Paper presented at the Ergo-IHM, Biarritz, France.

Tsang, P. S., & Velasquez, V. L. (1996). Diagnosticity and multidimensionnal subjective workload ratings. *Ergonomics, 39*(3), 358-381.

Tulving, E. (1989). Memory: Performance, knowledge, and experience. *European Journal of cognitive psychology, 1*(1), 3-26.

Tuovinen, J. E., & Paas, F. (2004). Exploring multidimensional approaches to the efficiency of instructional conditions. *Instructional science, 32*, 133-152.

Van Merriënboer, J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational psychologist Review, 17*(2), 147-177.

Vanderdonckt, J., & Coyette, A. (2006, Octobre). *Vers un prototypage des interfaces graphiques incluant vraiment l'utilisateur final.* Paper presented at the Ergo IA 2006, Bidart / biarritz France.

Vasalou, A., Ng, B., Wiemer-Hastings, P., & Oshlyansky, L. (2004, june). *Human-moderated remote user testing: Protocols and applications.* Paper presented at the 8th ERCIM Workshop, User Interfaces for All, Wien, Austria.

Vecchi, T., & Richardson, T. E. (2000). Active processing in visuo-spatial working memory. *Cahiers de Psychologie Cognitive / Current Psychology of Cognition, 19*(1), 3-32.

Venkatesh, V., & Davis, F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision Science, 27*, 451-481.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science, 46*(2), 186-204.

Venkatesh, V., Morris, M., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425-478.

Watzman, S. (2003). Visual design principles for usable interfaces. In J. A. Jacko & A. Sears (Eds.), *The Human Computer Interaction Handbook* (pp. 263-285). London: Lawrence Erlbaum Associates.

Wickens, C. D. (1996). *Attention and information access effort in head-up displays*. Savoy, Illinois, USA: Aviation Research Laboratory, Institute of Aviation, NASA Ames Research Center.

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.

Wiedenbeck, S., & Davis, S. (1997). The influence of interaction style and experience on user perceptions of software packages. *International Journal of Human-Computer Studies*, *46*(5), 563-588.

Yang, H.-D., & Youngjin, Y. (2004). It's all about attitude: Revisiting the technology acceptance model. *Decision Support System, 38*, 19-31.

Yeh, Y.-Y., & Wickens, C. D. (1984). *An investigation of the dissociation between subjective measures of mental workload and performance*. Urbana-Champaign: University of Illinois.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*, 459-482.

Yi, M. Y., Jackson, J. D., Park, J. S., & Probst, J. C. (2006). Understanding information technology acceptance by individual professionals: Toward an integrative view. *Information & Management, 43*, 350-363.

Zacks, J. M., & Tversky, B. (2003). Structuring information interfaces for procedural learning. *Journal of Experimental Psychology: Applied, 9*, 88-100.

Zesiger, P. (1995). *Ecrire*. Paris: Presses Universitaires de France.

Zhang, P., & Dillon, A. (2003). HCI and MIS: Shared concerns. *Interational Journal of Human-Computer Studies, 59*, 397-402.

Zhang, Y., & Luximon, A. (2005). Subjective mental workload measures. *Ergonomia IJE&HF, 27*(3), 199-206.

Zouinar, M., Relieu, M., Salembier, P., & Calvet, G. (2004, Juin). *Observation et capture de données sur l'interaction en mobilité.* Paper presented at the Mobilité & Ubiquité '04, Nice.

12- Index des figures

Figure 1: Dispositifs mobiles de petite taille (PDA à gauche et Smartphone à droite) équipés de stylets et d'écran tactile. 9
Figure 2: Modélisation des attributs de l'acceptabilité des systèmes selon Shackel (1991, notre traduction) 15
Figure 3: Modélisation des attributs de l'acceptabilité des systèmes selon Nielsen (1993, notre traduction) 17
Figure 4: Dimensions de l'utilisabilité selon la norme ISO 9241-11 (1998, notre traduction) 19
Figure 5: Alphabet Graffiti 27
Figure 6: Capture d'écran typique de l'expérience de MacKenzie et Chang (1999) 28
Figure 7: Interface JOT 29
Figure 8: Les trois conditions expérimentales étaient (a) l'écriture manuscrite reconnue, (b) le clavier QWERTY, et (c) le clavier ABC (MacKenzie, Nonnecke, Riddersma, McQueen, & Meltz, 1994) 30
Figure 9: Alphabet Graffiti 2 et clavier virtuel 31
Figure 10: Déterminants principaux de la diffusion des innovations, selon Tornatzky et Klein (1982, notre traduction) 34
Figure 11: Théorie de l'action raisonnée (d'après Channouf, Py, et Somat, 1996) 36
Figure 12: Modèle d'acceptation de la technologie (Davis, 1989) 40
Figure 13: Le modèle P3 (d'après Dillon & Morris, 1998, notre traduction) 48
Figure 14: Modèle multipartite de la mémoire de travail (d'après Baddeley, 2000, notre traduction) 67
Figure 15: Modèle révisé des processus rédactionnels (Hayes, 1996, d'après Piolat, 2004b) 76
Figure 16: Processus rédactionnels et mémoire de travail (Kellogg, 1996) 80
Figure 17: Processus rédactionnels et ressources de MdT (Kellogg, 1996) 82
Figure 18: Représentation graphique de l'efficience mentale (d'après Paas, Tuovinen, Tabbers & Van Gerven, 2003) 123
Figure 19: Le logiciel de reconnaissance RESIFCar implémenté sur Smartphone 132
Figure 20: L'interface de saisie DIGIME implémentée sur pocket PC, sous l'application Pocket Word 132
Figure 21 : Fonctionnement de l'interface DIGIME 138
Figure 22: Exemples de planches proposées par les didacticiels 141
Figure 23: Présentation aléatoire d'un mot sur l'écran du PDA avant saisie par les participants 143
Figure 24: Présentation des quatre déclinaisons de l'interface DIGIME en fonction des modalités des VI intergroupes (contiguïté du feedback et conservation de l'encre) 145
Figure 25: Captures d'écrans de l'application Caractest. 147
Figure 26: Effets des VI contiguïté du feedback et conservation de l'encre en fonction de la position sérielle sur les temps de saisie de groupes de cinq mots 150
Figure 27: Effets des VI contiguïté du feedback et conservation de l'encre sur les temps de saisie des quinze mots 150
Figure 28: Effets des VI contiguïté du feedback et conservation de l'encre en fonction de la position sérielle sur le nombre d'erreurs commises sur chaque groupe de cinq mots 152
Figure 29: Effets des VI contiguïté du feedback et conservation de l'encre sur le nombre moyen d'erreurs commises pour la saisie des quinze mots 153
Figure 30: Distribution des réponses à la question en fonction du facteur conservation de l'encre. 154
Figure 31: Fonctionnement de la nouvelle interface DIGIME (version anglaise) 166
Figure 32: Nombre total d'erreurs par lettres produites pour l'ensemble des participants de l'expérience 1 169
Figure 33: Exemples de planches proposées par le didacticiel statique 170
Figure 34:Durées moyennes de consultation de l'aide gestes d'édition pendant la saisie des lettres faciles en fonction du didacticiel 178
Figure 35: Durées moyennes de consultation de l'aide gestes d'édition pendant la saisie de toutes les lettres en fonction du didacticiel 178
Figure 36: Nombre moyen de geste d'insertion en fonction du didacticiel pendant la saisie des phrases en fonction du didacticiel 180
Figure 37: Utilisabilité perçue de l'interface DIGIME en fonction du didacticiel 181
Figure 38: Utilisabilité perçue de l'interface DIGIME en fonction de la présence ou de l'absence de didacticiel 182
Figure 39: Interface DIGIME complète (à gauche), avec menu hypertexte du didacticiel (au centre) et exemple de planche du didacticiel (à droite). 193
Figure 40: Interface de saisie clavier virtuel 194
Figure 41: Interface d'écriture Note 195

Figure 42: Interface de saisie et de reconnaissance de caractères manuscrits JOT (à gauche), avec menu hypertexte du didacticiel (au centre) et exemple de planche du didacticiel (à droite). 196
Figure 43: Planche de caractères à reproduire en phase d'entraînement (taille réelle) 196
Figure 44:Temps moyens de saisie (s) aux tâches d'entraînement et de prises de notes 201
Figure 45: Test de perception 215
Figure 46: Effet du niveau en informatique sur la fréquence d'utilisation de l'informatique 218
Figure 47: Effet du niveau en informatique sur les scores au test de perception 219
Figure 48: Dernière version de l'interface DIGIME 234
Figure 49: Modèle de prédiction de l'intention d'utilisation par des mesures objectives et subjectives d'utilisabilité 239
Figure 50: Résumé graphique des hypothèses testées et des résultats obtenus sur les mesures Δ 245
Figure 51: Modèle de prédiction de l'intention d'utilisation par des mesures objectives et subjectives d'utilisabilité 257
Figure 52: Résumé graphique des hypothèses et des principaux résultats obtenus sur les mesures Δ 272

13- Index des tableaux

Tableau 1: Principaux résultats de l'expérience, en fonction des VI manipulées 148
Tableau 2: Principaux résultats de l'expérience, en fonction de la VI didacticiel (Moyenne, et, entre parenthèses, écart-type) 176
Tableau 3: Effets du type d'interface sur les dimensions du TAM 202
Tableau 4: Effets du type d'interface sur les items du test d'utilisabilité de Nielsen 204
Tableau 5: Effet du niveau en informatique sur les scores au pré-questionnaire 217
Tableau 6: Effets du niveau en informatique sur les scores aux questions du test de perception 219
Tableau 7: Effets du niveau en informatique sur les performances de saisie à l'épreuve 220
Tableau 8: Effets du type d'interface sur les dimensions du TAM 221
Tableau 9: Effets du niveau en informatique sur les items du test d'utilisabilité de Nielsen 221
Tableau 10: Principales mesures d'utilisabilité en fonction des interfaces 240
Tableau 11: Scores aux VD pour chaque groupe 241
Tableau 12: Résultats du test de Mann-Whitney appliqué à la comparaison entre groupe contrôle et groupe expérimental 241
Tableau 13: Corrélations entre les VI comparées de l'analyse 242
Tableau 14: Corrélations entre les VI de l'analyse pour DIGIME 242
Tableau 15: Corrélations entre les VI de l'analyse pour Note 242
Tableau 16: Effet de l'effort mental sur l'efficacité 243
Tableau 17: Effet de l'effort mental et de l'efficacité d'une part, de l'efficience d'autre part, sur la charge de tâche 244
Tableau 18: Effet de la charge de tâche, de l'utilisabilité, de l'effort mental, de l'efficience et de l'efficacité sur l'acceptabilité 245
Tableau 19: Principales mesures d'utilisabilité en fonction des interfaces 262
Tableau 20: Corrélations entre mesures Δ de l'expérience 263
Tableau 21: Corrélations entre mesures de l'expérience pour DIGIME 264
Tableau 22: Corrélations entre mesures de l'expérience pour le clavier virtuel 264
Tableau 23: Effet de l'efficience sur les réponses aux items du test de Nieslen 265
Tableau 24: Effets du test de Nielsen et de l'efficience sur la facilité d'utilisation perçue 266
Tableau 25: Effets du test de Nielsen et de la facilité d'utilisation perçue sur l'utilité perçue comparée entre interfaces 266
Tableau 26: Effets du test de Nielsen et de la facilité d'utilisation perçue sur l'utilité perçue de DIGIME 267
Tableau 27: Effets du test de Nielsen et de la facilité d'utilisation perçue sur l'utilité perçue du clavier virtuel 268
Tableau 28: Effets du test de Nielsen, de la facilité d'utilisation perçue, de l'utilité perçue et de l'efficience sur l'acceptabilité comparée 269
Tableau 29: Effets du test de Nielsen, de la facilité d'utilisation perçue, de l'utilité perçue et de l'efficience sur l'intention d'utilisation de DIGIME 270
Tableau 30: Effets du test de Nielsen, de la facilité d'utilisation perçue, de l'utilité perçue et de l'efficience sur l'intention d'utilisation du clavier virtuel 270
Tableau 31: Caractéristiques linguistiques des mots 314
Tableau 32: Caractéristiques linguistiques des mots 314

14- Index des formules

Formule 1: Mesures d'efficiences selon la norme ISO 9241-11 interprétées par Bevan et MacLeod (1994) 20
Formule 2: Mesure d'efficacité selon la norme ISO 9241-11 interprétée par Bevan et MacLeod (1994) 20
Formule 3: Mesure d'efficience mentale E (Paas & Van Merriënboer, 1993) 122
Formule 4: Mesure d'efficience E (d'après Huang, Hong, & Eades, 2006) 124

15- Annexes

15.1. Caractéristiques linguistiques des mots utilisés pour l'expérience 1

Mots	Classe gram-maticale	Fré-quence d'usage des formes	Nombre de variantes sémanti-ques	Valence d' imagerie	Nombre de lettres	Nombre de syllabes	Nombre de variantes gram-maticales	Fré-quence de digram-mes
aveu	nom	3582	0	0	4	2	0	2,98
avis	nom	6598	0	2	4	2	0	2,96
épée	nom	3105	0	0	4	2	0	2,25
obus	nom	2692	0	0	4	2	0	2,33
pavé	nom	2254	0	5	4	2	0	2,55
Moyenne		**3646,2**	**0**	**1,4**	**4**	**2**	**0**	**2,61**
écart-type		**1721,8**	**0**	**2,1**	**0**	**0**	**0**	**0,34**
besogne	nom	3190	0	0	7	2	0	2,57
caresse	nom	2616	0	0	7	2	0	2,98
lunette	nom	2450	0	5	7	2	0	2,72
période	nom	6836	1	2	7	2	0	2,68
stupeur	nom	2327	0	0	7	2	0	2,59
Moyenne		**3483,8**	**0,2**	**1,4**	**7**	**2**	**0**	**2,70**
écart-type		**1902,8**	**0,4**	**2,1**	**0**	**0**	**0**	**0,16**
brouillard	nom	3301	1	5	10	2	0	2,84
changement	nom	7257	0	2	10	2	0	2,97
groupement	nom	2054	0	0	10	2	0	2,98
lieutenant	nom	288	0	0	10	2	0	3,06
silhouette	nom	2658	0	0	10	2	0	2,69
Moyenne		**3631,6**	**0,2**	**1,4**	**10**	**2**	**0**	**2,90**
écart-type		**2076,2**	**0,4**	**2,1**	**0**	**0**	**0**	**0,14**

Tableau 31: Caractéristiques linguistiques des mots

15.2. Caractéristiques linguistiques des mots utilisés pour l'expérience 2

Mots	Classe gram-maticale	Fré-quence d'usage des formes	Nombre de variantes sémanti-ques	Valence d' imagerie	Nombre de lettres	Nombre de syllabes	Nombre de variantes gram-maticales	Fré-quence de digram-mes
égard	nom	12103	0	0	5	2	0	2,38
hôtel	nom	12954	0	0	5	2	0	2,19
maman	nom	12273	0	0	5	2	0	2,62
odeur	nom	12214	0	0	5	2	0	2,6
vertu	nom	14694	0	0	5	2	0	2,7

Tableau 32: Caractéristiques linguistiques des mots

15.3. Questionnaires

15.3.1. Pré-questionnaire

Veuillez répondre le plus rigoureusement à ce questionnaire. Les informations recueillies sont strictement confidentielles et ne serviront qu'à la conduite de cette expérience.

Age :

Sexe :

Etudes suivies :

Latéralité (gaucher ou droitier) :

Numéro de participant (fourni par l'expérimentateur) :

Problème de vue (si votre problème de vue est corrigé par le port de lunettes au cours de l'expérience, notez non) :

Pour les questions qui suivent, entourez la réponse qui vous semble le mieux vous représenter :

1) Vous avez une utilisation de l'informatique

   Inexistante Mensuelle Hebdomadaire Quotidienne

2) Avez-vous un PDA (Personnal Digital Assistant) ?

   OUI NON

3) Vous avez une utilisation de PDA

   Inexistante Mensuelle Hebdomadaire Quotidienne

Pour les questions qui suivent, répondez lisiblement par écrit. Si vous ne savez pas, notez NSP :

4) Citez quelques marques de Pocket PC.

5) Donnez quelques exemples de systèmes d'exploitation employés sur les Pocket PC.

6) Que signifie PDA ?

7) Qu'est-ce que Transcriber ?

15.3.2. NASA-TLX

Selon vous laquelle de ces deux affirmations vous a posé le plus de difficulté au cours de la tâche ?
Choisissez ce qui vous a paru le plus difficile au cours de la tâche parmi la succession de deux affirmations suivantes :

1) a) Le travail demandé est complexe. b) La tâche était physiquement contraignante.	2) c) Le temps était réduit pour le travail imparti. d) La performance a été difficile à obtenir.
3) a) Le travail demandé est complexe. e) La performance n'est pas satisfaisante.	4) b) La tâche était physiquement contraignante. c) Le temps était réduit pour le travail imparti.
5) a) Le travail demandé est complexe. d) La performance a été difficile à obtenir.	6) e) La performance n'est pas satisfaisante. f) L'épreuve n'est pas agréable.
7) d) La performance a été difficile à obtenir. e) La performance n'est pas satisfaisante.	8) a) Le travail demandé est complexe. f) L'épreuve n'est pas agréable.
9) b) La tâche était physiquement contraignante. d) La performance a été difficile à obtenir.	10) a) Le travail demandé est complexe. c) Le temps était réduit pour le travail imparti.
11) b) La tâche était physiquement contraignante. e) La performance n'est pas satisfaisante.	12) c) Le temps était réduit pour le travail imparti. f) L'épreuve n'est pas agréable.
13) d) La performance a été difficile à obtenir. f) L'épreuve n'est pas agréable.	14) c) Le temps était réduit pour le travail imparti. e) La performance n'est pas satisfaisante.
15) b) La tâche était physiquement contraignante. f) L'épreuve n'est pas agréable.	

Veuillez répondre le plus sérieusement à ce questionnaire en cochant dans les échelles parmi les 20 positions possible :

1) Avez-vous trouvé le travail demandé simple ou complexe ?

simple complexe

2) Avez-vous trouvé la tâche physiquement contraignante ?

Pas du tout contraignant Très contraignant

3) Avez-vous ressenti une pression temporelle lors de la réalisation de la tâche ?

Aucune pression Forte pression

4) Vous a-t-il semblé difficile, mentalement et physiquement, d'obtenir votre niveau de performance ?

Facile difficile

5) Etes-vous satisfait de votre performance dans l'accomplissement de la tâche ?

Pas du tout satisfait Très satisfait

6) Vous êtes vous senti découragé, irrité, stressé ou au contraire motivé, content, satisfait durant l'accomplissement de la tâche

Découragé Content

15.3.3. Test d'utilisabilité de Nielsen

Entourez le chiffre correspondant à votre opinion :

1) Vous a-t-il semblé facile d'apprendre à utiliser le logiciel de reconnaissance ?

-2	-1	0	1	2
Pas du tout				Tout à fait

2) Le logiciel de reconnaissance vous a-t-il semblé efficace ?

-2	-1	0	1	2
Pas du tout				Tout à fait

3) Si vous deviez reproduire cet exercice dans une semaine seriez-vous capable d'atteindre le même niveau de performance ?

-2	-1	0	1	2
Pas du tout				Tout à fait

4) Avez-vous eu du mal à gérer les erreurs, que ce soit celles du logiciel ou les vôtres ?

-2	-1	0	1	2
Pas du tout				Tout à fait

5) Est-ce que le logiciel de reconnaissance vous a laissé une impression agréable ?

-2	-1	0	1	2
Pas du tout				Tout à fait

15.3.4. Questions additionnelles

6) Utiliseriez-vous ce système de reconnaissance ?

-2	-1	0	1	2
Pas du tout				Tout à fait

7) Avez-vous changé d'écriture pour que le système reconnaisse votre écriture ?

-2	-1	0	1	2
Pas du tout				Tout à fait

8) Avez-vous trouvé la situation expérimentale ennuyante ?

-2	-1	0	1	2
Pas du tout				Tout à fait

15.3.5. Test SUS (System Usability Scale)

SVP, veuillez répondre sérieusement à ce questionnaire. Vos réponses demeureront confidentielles et anonymes.

	Pas du tout d'accord				**Tout à fait d'accord**
1. Je pense que j'aimerais utiliser ce système de saisie manuscrite fréquemment.	1	2	3	4	5
2. J'ai trouvé ce système de saisie manuscrite inutilement complexe.	1	2	3	4	5
3. J'ai pensé que ce système de saisie manuscrite a été facile à utiliser.	1	2	3	4	5
4. Dans une utilisation future, je pense que j'aurais besoin de l'aide d'une personne expérimentée pour pouvoir utiliser le système.	1	2	3	4	5
5. J'ai trouvé que les différentes fonctions ont été bien intégrées au système de saisie manuscrite.	1	2	3	4	5
6. J'ai trouvé qu'il y avait trop d'incohérences dans le système de saisie manuscrite.	1	2	3	4	5
7. J'imagine que la plupart des gens apprendront rapidement à utiliser ce système de saisie.	1	2	3	4	5
8. J'ai trouvé ce système très désagréable à utiliser.	1	2	3	4	5
9. Je me suis senti(e) très confiant(e) pendant l'utilisation de ce système.	1	2	3	4	5
10. Il a fallu que j'apprenne beaucoup de choses avant de pouvoir utiliser ce système.	1	2	3	4	5

15.3.6. Test TAM (Technology Acceptance Model)

Choisissez l'item correspondant à votre sentiment

1) Apprendre à utiliser le dispositif de saisie de texte serait facile pour moi.

1	2	3	4	5	6	7
extrêmement probable	assez probable	peu probable	ni l'un ni l'autre	peu improbable	assez improbable	extrêmement improbable

2) Utiliser ce dispositif de saisie de texte pourrait améliorer mes performances au quotidien.

1	2	3	4	5	6	7
extrêmement probable	assez probable	peu probable	ni l'un ni l'autre	peu improbable	assez improbable	extrêmement improbable

3) Je trouve facile de faire faire au dispositif de saisie de texte ce que je veux qu'il fasse

1	2	3	4	5	6	7
extrêmement probable	assez probable	peu probable	ni l'un ni l'autre	peu improbable	assez improbable	extrêmement improbable

4) Utiliser ce dispositif de saisie de texte est une …….. idée.

1	2	3	4	5	6	7
très bonne	bonne	assez bonne	ni bonne ni mauvaise	assez mauvaise	mauvaise	très mauvaise

5) Il me serait facile de devenir compétent dans l'utilisation de ce dispositif de saisie de texte.

1	2	3	4	5	6	7
extrêmement probable	assez probable	peu probable	ni l'un ni l'autre	peu improbable	assez improbable	extrêmement improbable

6) Utiliser ce dispositif de saisie pourrait améliorer mon efficacité au quotidien.

1	2	3	4	5	6	7
extrêmement probable	assez probable	peu probable	ni l'un ni l'autre	peu improbable	assez improbable	extrêmement improbable

7) Utiliser ce dispositif de saisie de texte est une idée …….. .

1	2	3	4	5	6	7
très stupide	stupide	assez stupide	ni stupide ni sage	assez sage	sage	très sage

8) Je trouverais facile d'utiliser ce dispositif de saisie de texte.

1	2	3	4	5	6	7
extrêmement probable	assez probable	peu probable	ni l'un ni l'autre	peu improbable	assez improbable	extrêmement improbable

9) Utiliser ce dispositif de saisie pourrait améliorer ma productivité au quotidien.

1	2	3	4	5	6	7
extrêmement probable	assez probable	peu probable	ni l'un ni l'autre	peu improbable	assez improbable	extrêmement improbable

.

10) Je ……. l'idée d'utiliser ce dispositif de saisie de texte.

1	2	3	4	5	6	7
aime fortement	aime	aime peu	sans importance	déteste peu	déteste	déteste fortement

11) Je trouverai ce dispositif de saisie utile au quotidien.

1	2	3	4	5	6	7
extrêmement probable	assez probable	peu probable	ni l'un ni l'autre	peu improbable	assez improbable	extrêmement improbable

12) L'utilisation de ce dispositif de saisie est entièrement sous mon contrôle.

1	2	3	4	5	6	7
fortement d'accord	d'accord	peu d'accord	ni l'un ni l'autre	léger désaccord	désaccord	fort désaccord

13) Utiliser ce dispositif de saisie de texte serait ……..

1	2	3	4	5	6	7
extrêmement plaisant	assez plaisant	Peu plaisant	ni l'un ni l'autre	peu déplaisant	assez déplaisant	extrêmement Déplaisant

14) Les gens qui influencent ma conduite penseraient que je devrais utiliser ce dispositif de saisie de texte.

1	2	3	4	5	6	7
fortement d'accord	d'accord	peu d'accord	ni l'un ni l'autre	léger désaccord	désaccord	fort désaccord

15) Je serais capable d'utiliser ce dispositif de saisie de texte.

1	2	3	4	5	6	7
fortement d'accord	d'accord	peu d'accord	ni l'un ni l'autre	léger désaccord	désaccord	fort désaccord

16) J'ai l'intention d'utiliser ce dispositif de saisie de texte à l'avenir.

1	2	3	4	5	6	7
fortement d'accord	d'accord	peu d'accord	ni l'un ni l'autre	léger désaccord	désaccord	fort désaccord

17) J'ai les compétences, les connaissances, et la capacité à utiliser ce dispositif de saisie de texte.

1	2	3	4	5	6	7
fortement d'accord	d'accord	peu d'accord	ni l'un ni l'autre	léger désaccord	désaccord	fort désaccord

18) Les gens qui sont importants pour moi penseraient que je devrais utiliser ce dispositif de saisie de texte.

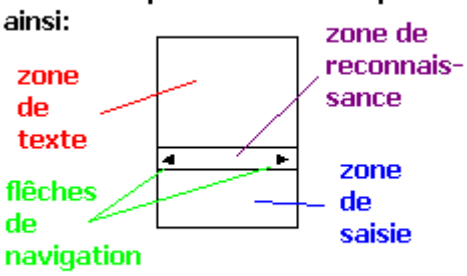
1	2	3	4	5	6	7
fortement d'accord	d'accord	peu d'accord	ni l'un ni l'autre	léger désaccord	désaccord	fort désaccord

19) J'ai l'intention d'utiliser ce dispositif de saisie de texte fréquemment à l'avenir.

1	2	3	4	5	6	7
fortement d'accord	d'accord	peu d'accord	ni l'un ni l'autre	léger désaccord	désaccord	fort désaccord

15.4. Didacticiels experience 1

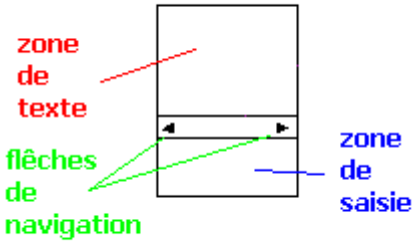
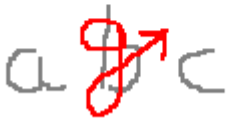
15.4.1. Condition “conservation de l’encre avec feedback contigu”

Bonjour, merci de participer à cette expérience. Il vous suffit de vous laisser guider par les instructions à l'écran et de cliquer sur les boutons avec le stylet. Nous allons vous expliquer comment utiliser ce pocketPC. Appuyez sur "suivant".	Pour commencer, nous allons vous expliquer comment fonctionne ce logiciel de reconnaissance de l'écriture. Soyez très attentifs, les explications vont défiler les unes à la suite des autres. Appuyez sur le bouton suivant.
Nous allons vous présenter des mots que vous allez devoir écrire avec le stylet sur l'écran du pocketPC. Faites très attention à l'orthographe des mots. La zone de saisie se trouve en bas de l'écran. Vous devez écrire chaque lettre séparément, non-attachées.	Ecrivez normalement, en prenant soin de bien former les lettres. Ce logiciel peut reconnaitre différents type de lettres: ou
L'écran du pocketPC est composé ainsi: 	Les lettres reconnues s'affichent au fur et à mesure dans la zone de texte en haut de l'écran, ainsi que dans la zone de reconnaissance, au centre de l'écran. Vous pouvez utiliser les flêches de navigation pour vous déplacer dans le mot manuscrit en cours de rédaction.
Pour corriger une erreur, il suffit de recouvrir la ou les lettres désirées d'un trait de stylet de la droite vers la gauche:	Pour faire un accent aigu, il vous suffit de tirer un trait de stylet oblique au-dessus de la lettre précédemment écrite:

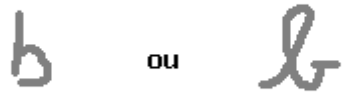
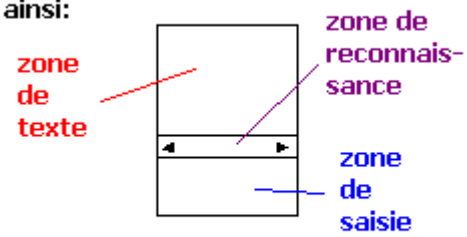
Pour effacer un accent, il suffit de le recouvrir d'un trait de stylet de la droite vers la gauche:	Vous pouvez compléter un caractère incomplet où qu'il soit dans le mot, cependant, il faut que le second trait de stylet soit en contact avec le premier:
Pour modifier une lettre en milieu de mot, il vous suffit d'écrire la nouvelle lettre sur celle que vous désirez changer:	L'expérience va commencer. Vous devrez lire attentivement chacun des mots qui apparaîtra et appuyer sur "exercice"pour pouvoir les écrire. Lorsque vous avez fini d'écrire un mot, appuyez sur "suivant". Le bouton "suivant" ne s'active que si le mot est correctement orthographié.

15.4.2. Condition "conservation de l'encre sans feedback contigu"

Bonjour, merci de participer à cette expérience. Il vous suffit de vous laisser guider par les instructions à l'écran et de cliquer sur les boutons avec le stylet. Nous allons vous expliquer comment utiliser ce pocketPC. Appuyez sur "suivant".	Pour commencer, nous allons vous expliquer comment fonctionne ce logiciel de reconnaissance de l'écriture. Soyez très attentifs, les explications vont défiler les unes à la suite des autres. Appuyez sur le bouton suivant.
Nous allons vous présenter des mots que vous allez devoir écrire avec le stylet sur l'écran du pocketPC. Faites très attention à l'orthographe des mots. La zone de saisie se trouve en bas de l'écran. Vous devez écrire chaque lettre séparément, non-attachées.	Ecrivez normalement, en prenant soin de bien former les lettres. Ce logiciel peut reconnaitre différents type de lettres: ou

L'écran du pocketPC est composé ainsi:  	Les lettres reconnues s'affichent au fur et à mesure dans la zone de texte en haut de l'écran. Vous pouvez utiliser les flèches de navigation pour vous déplacer dans le mot manuscrit en cours de rédaction.
Pour corriger une erreur, il suffit de recouvrir la ou les lettres désirées d'un trait de stylet de la droite vers la gauche: 	Pour faire un accent aigu, il vous suffit de tirer un trait de stylet oblique au-dessus de la lettre précédemment écrite:
Pour effacer un accent, il suffit de le recouvrir d'un trait de stylet de la droite vers la gauche: 	Vous pouvez compléter un caractère incomplet où qu'il soit dans le mot, cependant, il faut que le second trait de stylet soit en contact avec le premier:
Pour modifier une lettre ne milieu de mot, il vous suffit d'écrire la nouvelle lettre sur celle que vous désirez changer: 	L'expérience va commencer. Vous devrez lire attentivement chacun des mots qui apparaîtra et appuyer sur "exercice"pour pouvoir les écrire. Lorsque vous avez fini d'écrire un mot, appuyez sur "suivant". Le bouton "suivant" ne s'active que si le mot est correctement orthographié.

15.4.3. Condition “disparition de l’encre avec feedback contigu”

Bonjour, merci de participer à cette expérience. Il vous suffit de vous laisser guider par les instructions à l'écran et de cliquer sur les boutons avec le stylet. Nous allons vous expliquer comment utiliser ce pocketPC. Appuyez sur "suivant".	Pour commencer, nous allons vous expliquer comment fonctionne ce logiciel de reconnaissance de l'écriture. Soyez très attentifs, les explications vont défiler les unes à la suite des autres. Appuyez sur le bouton suivant.
Nous allons vous présenter des mots que vous allez devoir écrire avec le stylet sur l'écran du pocketPC. Faites très attention à l'orthographe des mots. La zone de saisie se trouve en bas de l'écran. Vous devez écrire chaque lettre séparément, non-attachées.	Ecrivez normalement, en prenant soin de bien former les lettres. Ce logiciel peut reconnaitre différents type de lettres:  ou
L'écran du pocketPC est composé ainsi: 	Les lettres reconnues s'affichent au fur et à mesure dans la zone de texte en haut de l'écran, ainsi que dans la zone de reconnaissance, au centre de l'écran. Chacunes de vos lettres manuscrites disparaissent de la zone de saisie dès que le logiciel les identifie.
Pour corriger une erreur, il suffit de faire un trait de stylet de la droite vers la gauche, ce qui efface la dernière lettre:	Pour faire un accent aigu, il vous suffit de tirer un trait de stylet oblique dans le tiers supérieur de la zone d'écriture après la disparition de la lettre précédemment écrite:

Pour effacer un accent, il suffit de tirer un trait de stylet de la droite vers la gauche dans le tiers supérieur de la zone d'écriture:	L'expérience va commencer. Vous devrez lire attentivement chacun des mots qui apparaîtra et appuyer sur "exercice"pour pouvoir les écrire. Lorsque vous avez fini d'écrire un mot, appuyez sur "suivant". Le bouton "suivant" ne s'active que si le mot est correctement orthographié.
Pour modifier une lettre en milieu de mot, vous devez effacer successivement les lettres qui suivent celle que vous voulez changer et réécrire ensuite la fin du mot:	L'expérience va commencer. Vous devrez lire attentivement chacun des mots qui apparaîtra et appuyer sur "exercice"pour pouvoir les écrire. Lorsque vous avez fini d'écrire un mot, appuyez sur "suivant". Le bouton "suivant" ne s'active que si le mot est correctement orthographié.

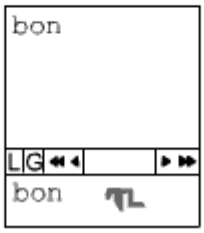
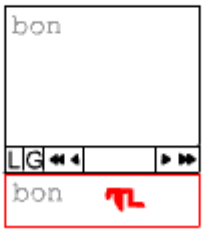
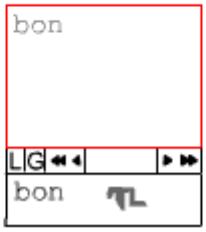
15.4.4. Condition "disparition de l'encre sans feedback contigu"

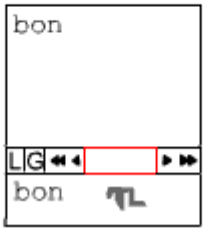
Bonjour, merci de participer à cette expérience. Il vous suffit de vous laisser guider par les instructions à l'écran et de cliquer sur les boutons avec le stylet. Nous allons vous expliquer comment utiliser ce pocketPC. Appuyez sur "suivant".	Pour commencer, nous allons vous expliquer comment fonctionne ce logiciel de reconnaissance de l'écriture. Soyez très attentifs, les explications vont défiler les unes à la suite des autres. Appuyez sur le bouton suivant.
Nous allons vous présenter des mots que vous allez devoir écrire avec le stylet sur l'écran du pocketPC. Faites très attention à l'orthographe des mots. La zone de saisie se trouve en bas de l'écran. Vous devez écrire chaque lettre séparément, non-attachées.	Ecrivez normalement, en prenant soin de bien former les lettres. Ce logiciel peut reconnaitre différents type de lettres: ou

L'écran du pocketPC est composé ainsi: zone de texte zone de saisie	Les lettres reconnues s'affichent au fur et à mesure dans la zone de texte en haut de l'écran. Chacunes de vos lettres manuscrites disparaissent de la zone de saisie dès que le logiciel les identifie.
Pour corriger une erreur, il suffit de faire un trait de stylet de la droite vers la gauche, ce qui efface la dernière lettre:	Pour faire un accent aigu, il vous suffit de tirer un trait de stylet oblique dans le tiers supérieur de la zone d'écriture après la disparition de la lettre précédemment écrite:
Pour effacer un accent, il suffit de tirer un trait de stylet de la droite vers la gauche dans le tiers supérieur de la zone d'écriture:	Les caractères doivent être écrits d'un seul trait, c'est-à-dire sans relever le stylet, sauf "i", "j", et "t".
Pour modifier une lettre en milieu de mot, vous devez effacer successivement les lettres qui suivent celle que vous voulez changer et réécrire ensuite la fin du mot:	L'expérience va commencer. Vous devrez lire attentivement chacun des mots qui apparaîtra et appuyer sur "exercice"pour pouvoir les écrire. Lorsque vous avez fini d'écrire un mot, appuyez sur "suivant". Le bouton "suivant" ne s'active que si le mot est correctement orthographié.

15.5. Didacticiel experience 2

15.5.1. Condition statique

Merci de participer à cette expérience. Celle-ci se déroule en deux parties. Nous allons tout d'abord vous expliquer comment fonctionne ce logiciel de reconnaissance de l'écriture. Ensuite, nous testerons ce que vous avez retenu de ces consignes aux cours de petits exercices... Cliquez sur "suivant" avec le stylet.	Les explications vous seront présentées successivement. A la fin de chacunes d'entre elles, cliquez sur "suivant" pour passer à l'explication suivante. A la fin du didacticiel, il vous sera proposé de le revoir (si vous le souhaitez), ou de commencer directement les exercices. Cliquez sur "suivant" pour commencer le didacticiel.
L'écran du pocket PC sur lequel vous allez écrire se présente de la façon suivante: 	En rouge, en bas de l'écran, se trouve la zone de saisie. C'est là que vous devez écrire:
L'écran conserve une trace de votre écriture quelques secondes seulement, le temps, par exemple, d'ajouter un accent sur la lettre. 	Quand la trace disparait, elle est convertie en lettre reconnue par le logiciel, à gauche de la zone de saisie.
En cas de besoin, vous pouvez faire des corrections sur ces lettres reconnues, comme nous le verrons plus tard.	En rouge, la zone traitement de texte, où votre texte s'inscrit. C'est là que s'affiche aussi le didacticiel que vous consultez. 

Vous n'aurez pas à utiliser cette zone pour écrire. Vous ne pouvez qu'y lire les lettres reconnues. 	Ces flêches, en rouge, vous permettent de faire défiler les lettres reconnues lorsque leur longueur éxcède la largeur de l'écran.
Dans la zone rouge, vous trouverez les boutons sur lesquels vous devrez cliquer avec le stylet. 	Pendant les exercices qui suivront, une aide, expliquant comment réaliser les lettres, est disponible en cliquant sur le bouton "L".
De même, une aide pour les gestes d'édition (espace, effacement, entrée...) est disponible en cliquant sur le bouton "G". 	Ces deux aides en ligne sont à votre disposition durant les exercices. N'hésitez pas à vous en servir.
Nous allons maintenant vous montrer comment réaliser quelques lettres. Cliquez sur "suivant".	Ce logiciel de reconnaissance des lettres manuscrites ne reconnait que l'écriture script, c'est-à-dire qu'il faut faire des lettres détachées les unes des autres.
Ce logiciel de reconnaissance des lettres manuscrites peut reconnaître différentes formes pour une même lettre, par exemple: 	Néanmoins, si l'une de vos lettres n'était pas reconnue à plusieurs reprises, ne vous obstinez pas trop et essayez une autre forme possible parmi celles que vous connaissez. Ecrivez à vitesse normale, de façon naturelle.

Pour tracer la lettre "a ", procédez comme ci-dessus.	Pour tracer un "m", procédez comme ci-dessus.
Pour tracer un "i ", procédez comme ci-dessus.	Pour tracer un "n ", procédez comme ci-dessus.
Pour tracer un "p", procédez comme ci-dessus.	Pour tracer un "t", procédez comme ci-dessus.
Voyons maintenant comment réaliser quelques accents. vous pouvez faire les accents au-dessus de la trace écrite de la lettre pendant quelques secondes, avant qu'elle ne disparaisse. Mais vous pouvez tout aussi bien les faire au-dessus des lettres déjà reconnues.	Pour faire un accent aigu, tracez un trait oblique au-dessus de la lettre à accentuer. Cela fonctionne aussi bien dans la zone de saisie que sur les lettres reconnues.
Pour faire un accent circonflexe, tracez un trait oblique comme ci-dessua. Cela fonctionne aussi bien dans la zone de saisie que sur les lettres reconnues.	Pour faire un accent trema, faites un point au-dessus de la lettre.

Maintenant, voyons comment réaliser quelques gestes d 'édition tels que effacement, espace ou entrée...	Pour effacer un accent que vous venez d'écrire, il suffit de tirer un trait de la droite vers la gauche dans la partie supérieure de la zone de saisie.
Pour effacer un accent sur les lettres déjà reconnues, faites un trait de la droite vers la gauche sur l'accent.	Pour faire "entrée", tracez le geste comme ci-dessus.
Pour effacer le dernier caractère (lettre ou espace) reconnu, tirez un trait de la droite vers la gauche.	Pour effacer un ou plusieurs caractères (lettre ou espace), recouvrez d'un trait, de la droite vers la gauche, les caractères reconnus que vous voulez faire disparaitre.
Pour créer un espace, tirez un trait de la gauche vers la droite dans la zone de saisie.	Pour créer un espace entre les lettres reconnues, tirez un trait de haut en bas entre les deux caractères à séparer. L'espace apparait sous la forme d'un carré blanc.
Vous pouvez compléter un caractère, même si vous avez levé le stylet, comme ci-dessus. Le second trait doit être en contact avec le premier.	Le didacticiel est terminé. Nous allons vous soumettre à une courte série d'exercices afin de tester ce que vous avez retenu du didacticiel. Il s'agit d'exercices pratiques durant lesquels vous devrez écrire des mots imposés. Si vous le souhaitez, vous pouvez revoir le didacticiel (en cliquant sur "revoir"). Sinon, cliquez sur "exercices" pour commencer les exercices.

15.6. Analyse des effets du type d'illustration sur les temps de saisie, le nombre d'erreurs, les durées et occurrences de consultation des aides en lignes en fonction du type de lettre

Facteurs intragroupes introduits dans l'analyse à mesures répétées:

Niveau de facilité	Apprentissage	Variable dépendante
Facile	Appris	lettres faciles apprises
	Non appris	lettres faciles non apprises
Difficile	Appris	lettres difficiles apprises
	Non appris	lettres difficiles non apprises

Facteur intergroupe introduit dans l'analyse à mesures répétées:
Type d'illustration (statique ou dynamique)

Variables dépendantes:
- Temps de saisie
- Nombre de corrections
- Temps de consultation aide lettres
- Temps de consultation aide gestes d'édition

VD : Temps de saisie

Statistiques descriptives

	condition	Moyenne	Ecart-type	N
Lettres faciles apprises	Statique	397,2	33,1	16
	Dynamique	450,3	53,0	15
	Total	422,9	43,2	31
Lettres faciles non apprises	Statique	106,4	335,2	16
	Dynamique	32,0	32,2	15
	Total	70,4	241,0	31
Lettres difficiles apprises	Statique	56,7	65,2	16
	Dynamique	31,4	25,8	15
	Total	44,4	51,0	31
Lettres difficiles non apprises	Statique	76,0	86,1	16
	Dynamique	47,3	56,1	15
	Total	62,1	73,4	31

Tests multivariés

Effet	Valeur	F	ddl de l'hypothèse	Erreur ddl	Signification
Facilité	,001	,016	1,000	29,000	,902
Facilité * Condition	,001	,026	1,000	29,000	,872
Apprentissage	,037	1,104	1,000	29,000	,302
Apprentissage * Condition	,032	,965	1,000	29,000	,334
Facilité * Apprentissage	,001	,042	1,000	29,000	,838
Facilité * Apprentissage * Condition	,025	,729	1,000	29,000	,400

Tests des effets inter-sujets

Source	Somme des carrés de type III	ddl	Moyenne des carrés	F	Signification
Intercept	365932526424,862	1	365932526424,862	16,283	,000
CONDITION	29355025181,410	1	29355025181,410	1,306	,262
Erreur	651726524880,009	29	22473328444,138		

VD : Nombre de corrections

Statistiques descriptives

	condition	Moyenne	Ecart-type	N
Lettres faciles apprises	Statique	3,50	3,42	16
	Dynamique	4,20	6,29	15
	Total	3,84	4,94	31
Lettres faciles non apprises	Statique	1,19	1,27	16
	Dynamique	1,73	1,83	15
	Total	1,45	1,56	31
Lettres difficiles apprises	Statique	4,31	4,72	16
	Dynamique	2,60	2,72	15
	Total	3,48	3,92	31
Lettres difficiles non apprises	Statique	3,94	5,05	16
	Dynamique	2,33	2,84	15
	Total	3,16	4,14	31

Tests multivariés

Effet	Valeur	F	ddl de l'hypothèse	Erreur ddl	Signification
Facilité	,034	1,035	1,000	29,000	,31
Facilité * Condition	,102	3,282	1,000	29,000	,08
Apprentissage	,136	4,574	1,000	29,000	,04
Apprentissage * Condition	,000	,000	1,000	29,000	,98
Facilité * Apprentissage	,116	3,808	1,000	29,000	,06
Facilité * Apprentissage * Condition	,001	,015	1,000	29,000	,90

Tests des effets inter-sujets

Source	Somme des carrés de type III	ddl	Moyenne des carrés	F	Signification
Intercept	1096,719	1	1096,719	41,404	,000
CONDITION	8,300	1	8,300	,313	,580
Erreur	768,168	29	26,489		

VD : Temps de consultation aide lettres

Statistiques descriptives

	condition	Moyenne	Ecart-type	N
Lettres faciles apprises	Statique	2,3	4,5	16
	Dynamique	4,8	9,5	15
	Total	3,5	7,3	31
Lettres faciles non apprises	Statique	2,6	7,5	16
	Dynamique	3,3	8,7	15
	Total	2,9	8,0	31
Lettres difficiles apprises	Statique	4,6	10,8	16
	Dynamique	1,1	3,0	15
	Total	2,9	8,1	31
Lettres difficiles non apprises	Statique	3,4	6,2	16
	Dynamique	1,8	4,6	15
	Total	2,7	5,5	31

Tests multivariés

Effet	Valeur	F	ddl de l'hypothèse	Erreur ddl	Signification
Facilité	,007	,198	1,000	29,000	,66
Facilité * Condition	,105	3,414	1,000	29,000	,07
Apprentissage	,003	,092	1,000	29,000	,76
Apprentissage * Condition	,000	,001	1,000	29,000	,97
Facilité * Apprentissage	,001	,021	1,000	29,000	,88
Facilité * Apprentissage * Condition	,015	,433	1,000	29,000	,51

Tests des effets inter-sujets

Source	Somme des carrés de type III	ddl	Moyenne des carrés	F	Signification
Intercept	1147256165,693	1	1147256165,693	17,966	,000
CONDITION	7916619,564	1	7916619,564	,124	,727
Erreur	1851851525,484	29	63856949,155		

VD : Temps de consultation aide gestes d'édition

Statistiques descriptives

	condition	Moyenne	Ecart-type	N
Lettres faciles apprises	Statique	,00	,00	16
	Dynamique	,89	3,4	15
	Total	,43	2,4	31
Lettres faciles non apprises	Statique	,19	,77	16
	Dynamique	1,9	5,8	15
	Total	1,0	4,1	31
Lettres difficiles apprises	Statique	5,0	17,9	16
	Dynamique	,00	,00	15
	Total	2,6	12,9	31
Lettres difficiles non apprises	Statique	8,0	18,6	16
	Dynamique	3,3	7,0	15
	Total	5,7	14,2	31

Tests multivariés

Effet	Valeur	F	ddl de l'hypothèse	Erreur ddl	Signification
Facilité	,128	4,240	1,000	29,000	,04
Facilité * Condition	,112	3,661	1,000	29,000	,06
Apprentissage	,034	1,031	1,000	29,000	,31
Apprentissage * Condition	,001	,025	1,000	29,000	,87
Facilité * Apprentissage	,016	,482	1,000	29,000	,49
Facilité * Apprentissage * Condition	,000	,005	1,000	29,000	,94

Tests des effets inter-sujets

Source	Somme des carrés de type III	ddl	Moyenne des carrés	F	Signification
Intercept	732351799,164	1	732351799,164	7,194	,012
CONDITION	98417116,358	1	98417116,358	,967	,334
Erreur	2952017465,384	29	101793705,703		

15.7. Textes utilisés pour les expériences 5 et 5 bis

15.7.1. Le chat

Le chat est un mammifère carnivore de la famille des félidés. Le mot chat vient du bas-latin cattus (chat sauvage). Le chat domestique Felis silvestris catus est particulièrement proche du chat sauvage européen Felis silvestris silvestris et du chat sauvage africain (chat ganté) Felis silvestris libyca, qui forment certainement tous trois une unique espèce : Felis silvestris (l'appellation Felis catus n'est plus valide).

Squelette et muscles

Comme tous les carnivores, la dernière prémolaire supérieure et la première molaire inférieure forment les carnassières. Celles-ci permettent au chat de déchirer sa nourriture, grâce à des muscles puissants fixés aux parois latérales de son crâne, et de l'avaler sans la mâcher.

Griffe avec le nerf visible

Le squelette est composé de 250 os. Les vertèbres du cou sont courtes, et la colonne vertébrale est très souple. Les vertèbres caudales prolongent la colonne, mais il est à noter que leur nombre est variable en fonction des races. La queue joue un rôle dans l'équilibre. Les pattes antérieures se terminent par cinq doigts pourvus de griffes rétractiles constituées de kératine, les pattes postérieures, plus longues que les pattes antérieures, se terminent par quatre doigts également pourvus de griffes rétractiles.

Les muscles du dos sont très souples et ceux des pattes postérieures sont puissants. Ces spécificités confèrent à l'animal une grande souplesse et une "détente" ample, lors des sauts par exemple.

Du fait d'une alimentation plus diversifiée et moins riche en protéines, les intestins du chat domestique sont plus longs que chez ses ancêtres sauvages. Avec une diminution de sa taille, c'est l'adaptation la plus notable à son nouveau mode de vie. Un chat pèse entre 2,5 et 9 kg.

Pelage

Le pelage du chat est composé de poils longs (jarre) et portant les marques de la robe (taches par exemple). En dessous se trouvent les poils plus courts (bourre), puis le duvet. Cette organisation permet une bonne isolation du corps.

Les types de pelages sont nombreux, car très variables en fonction des races. Il existe des poils longs, courts, frisés, et même crépus. Le sphynx, parfois appelé chat-nu, est une race presque dépourvue de poils : un très léger duvet recouvre le corps, ainsi que la queue.

De même, la couleur de la fourrure du chat peut prendre de nombreuses teintes et marques. Certains individus présentent de larges taches, d'autres des rayures ou des mouchetures, d'autres encore un pelage uni.

Cette variété est accentuée par les croisements obtenus en élevage, afin d'avoir des teintes inédites.

Dans certains pays, la fourrure du chat fait l'objet, comme celle du chien, d'une demande importante dans les industries de la mode. De nombreuses associations de protection des animaux condamnent cet usage des chats.

15.7.2. Le chien

Le chien (Canis lupus familiaris) est un mammifère domestique de la famille des canidés, proche du loup. Le chien n'est plus, pour le moment, considéré comme une espèce à part entière mais comme une sous-espèce de Canis lupus. Chien et loup sont ainsi de même espèce. Les anciennes appellations scientifiques du chien, Canis canis ou encore Canis familiaris, ne sont donc plus d'actualité.

C'est aussi une appellation pour plusieurs autres espèces de canidés de type Atelocynus et Speothos, voire de rongeurs du genre Cynomys (chien de prairie).

La femelle du chien s'appelle la chienne et un jeune chien est appelé un chiot. Le chien glapit, jappe ou aboie.

Il existe de nombreuses races de chiens. Environ les trois quarts de celles reconnues sont très anciennes et issues de la sélection naturelle sur la morphologie, combinée à une sélection artificielle sur le comportement (exemples : Husky sibérien, Berger de Brie). Les dernier quart est issu d'une sélection artificielle récente (exemples : Berger allemand, Golden Retriever) ; ces dernières sont souvent les plus populaires car modelées à la convenance de l'homme.

La domestication débute entre 16000 et 15000 avant J.-C. dans des groupes de chasseurs. En comparaison, le cheval sera domestiqué par des groupes nomades entre 4000 et 3000 avant J.-C.

Le chien aurait été simplement apprivoisé parmi d'autres animaux, tels les chacals ou les rongeurs. Mais c'est le seul maintenu en dépendance, car il aurait montré le plus d'aptitudes à une socialisation primitive. Le chien a pour ancêtre le loup mais des expériences, en cours depuis une cinquantaine d'années avec des croisements sélectifs de renards semblent donner des résultats similaires à ceux observés chez le chien (comportement particulièrement social, pédomorphisme, tempérament enfantin...).

L'étude des chiens et des races de chiens est appelée cynologie.
La Fédération Cynologique Internationale reconnaît 335 races. C'est elle qui définit les "standards", c'est-à-dire l'ensemble des caractéristiques définissant une race.

Anatomie

Le squelette du chien compte environ 300 os, le nombre étant variable d'une race à l'autre.

Malgré sa domestication et la dépendance à l'homme qui en découle, le chien a gardé sa musculature athlétique qui en fait un animal sportif et actif. Il possède un thorax large et descendu, et des pattes qui ne reposent au sol que par leur troisième phalange. Les membres antérieurs comportent 4 doigts, les postérieurs généralement 5, le pouce pouvant manquer chez certaines races (s'il existe, il ne touche pas le sol). Les doigts se terminent par des griffes et sont soutenus par des coussinets plantaires. Le chien est donc un digitigrade.

La tête du chien comporte une mâchoire puissante. La denture définitive, constituée de 42 dents, est en place vers 6 mois.

Chez le chien, taille et poids sont très variables d'une race à l'autre : dans les extrêmes, le poids du chihuahua peut être de 900 g et celui du Mastiff peut atteindre 140 kg.

L'espérance de vie de cet animal est en moyenne de 11 ans, mais peut aller de 8 à 20 ans.

Son sens de l'orientation est beaucoup plus précis que celui de l'homme. De même, son sens de l'équilibre serait légèrement plus aiguisé.

ETUDE ITERATIVE DES LIENS ENTRE UTILISABILITE ET ACCEPTABILITE D'UN DISPOSITIF DE SAISIE ET DE RECONNAISSANCE DE L'ECRITURE MANUSCRITE

Résumé Les interfaces de saisie et de reconnaissance automatique d'écriture manuscrite sur Pocket PC, bien que témoignant de performances théoriques élevées (McKenzie, 2002), ne sont pourtant pas jugés utilisables, et donc, acceptables, par leurs utilisateurs qui leur préfèrent le clavier virtuel (Ortelli, 2003, Költringer & Grechenig, 2004). Notre travail vise à comprendre ce problème d'acceptabilité, et s'articule autour de deux axes. Le premier consiste à améliorer la machine (ergonomie) tandis que le second consiste à comprendre l'utilisateur (efficience mentale et attitudes). Nous proposons tout d'abord une série de deux expériences et un test utilisateur dont le but est d'améliorer l'ergonomie d'une interface de saisie et de reconnaissance d'écriture manuscrite (DIGIME, développée en partenariat avec l'IRISA, Institut de Recherche en Informatique et Systèmes aléatoires), et ce afin d'en améliorer l'utilisabilité. Pour ce faire, nous mettons à contribution les modèles de la charge cognitive (Sweller, 1991), du format de présentation des illustrations (Mayer, 2001), ainsi que des critères ergonomiques (Bastien & Scapin, 2001). Les améliorations obtenues au niveau de l'ergonomie (expé 1) et du didacticiel (expé 2) ne s'avèrent pas suffisantes pour concurrencer les méthodes de saisies alternatives (expé 3). Par conséquent, dans un second temps, nous mettons en place trois autres expériences (dont un test de perception) qui tentent de mettre en relation processus rédactionnels, charge mentale (Kellogg, 1996) et acceptabilité (Davis, 1989). Les résultats montrent que s'il est vain de prendre en compte les avis des participants sans leur performance (expé 4), l'efficience mentale de l'écriture ne permet pas d'envisager de prévoir l'acceptabilité de DIGIME, de même que tout autre jugement porté sur cette interface (expé 5 et 5 bis). Ce cycle itératif de conception nous permet néanmoins de conclure que le défaut d'utilisabilité des logiciels de reconnaissance d'écriture est dû à un problème d'apprenabilité. L'acceptabilité s'en ressent, principalement du fait de la concurrence des claviers virtuels dont le fonctionnement, calqué sur les claviers matériels, est connu du plus grand nombre.

Mots clés: utilisabilité, acceptabilité, ergonomie, écriture manuscrite, pocket PC, interaction homme-machine, management des systèmes d'informations.

ITERATIVE STUDY OF THE RELATIONS BETWEEN USABILITY AND ACCEPTABILITY OF AN HANDWRITING INPUT AND RECOGNITION METHOD

Abstract Handwriting input methods on Pocket PC, although testifying high theoretical performances (McKenzie, 2002), are considered neither usable nor acceptable by users. Thus, they prefer virtual keyboard (Ortelli, 2003, Költringer & Grechenig, 2004). Our work aims to understand this lack of acceptability from an ergonomic and cognitive psychology framework. It pursues two goals: firstly, improving the software (ergonomics), and secondly understanding the user (mental efficiency and attitudes). First of all, we conducted a serie of two experiments and a usability testing, whom goals were to improve ergonomics (therefore ils utilisisability) of a handwriting input method (DIGIME, developed in partnership with IRISA, Institut de Recherche en Informatique et Systèmes aléatoires). The cognitive load theory (Sweller, 1991), the illustrations design principles (Mayer, 2001) and the ergonomic criteria (Bastien & Scapin, 2001) were used in these researches. The improvements obtained on ergonomics (study 1) and on the teachware (study 2) were not sufficient to compete with alternative input methods (study 3). Therefore, three more experiments were set up (including a paper prototyping) in order to connect editorial processes, mental effort (Kellogg, 1996) and acceptance (Davis, 1989). Results showed it seemed useless to take participant's attitude into account without considering performance (study 4). Mental efficiency of writing did not explain DIGIME's acceptability, or any other judgement related to this interface (studies 5 and 5 bis). This iterative cycle of design enables us to conclude that the lack of handwriting input methods usability is due to a learning problem. Acceptability decreases, mainly because of virtual keyboards competition, copied on hardware keyboards of whom functioning is well-known by the greater number.

Keywords: Usability, acceptance, ergonomics, handwriting, pocket PC, human-computer interaction, management of information systems.

Laboratoire de Psychologie Expérimentale – CRPCC
Place du Recteur Henri Le Moal
35043 Rennes CEDEX