



**HAL**  
open science

# Etude empirique de l'interaction multimodale mobile en situation naturelle

Guillaume Calvet-Inglada

► **To cite this version:**

Guillaume Calvet-Inglada. Etude empirique de l'interaction multimodale mobile en situation naturelle. Sciences de l'Homme et Société. Université Paul Sabatier - Toulouse III, 2008. Français. NNT : . tel-00443654

**HAL Id: tel-00443654**

**<https://theses.hal.science/tel-00443654>**

Submitted on 1 Jan 2010

**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.



# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse III - Paul Sabatier  
Discipline ou spécialité : Ergonomie Cognitive

---

Présentée et soutenue par *Guillaume CALVET-INGLADA*  
Le 24 novembre 2008

**Titre :** *Etude empirique de l'interaction multimodale mobile en situation naturelle*

---

### JURY

*Laurence Nigay, Rapporteur*  
*Françoise Détienne, Rapporteur*  
*Bernard Pavard, Directeur*  
*Christian Bastien, Examineur*  
*Julien Kahn, Examineur*  
*Pascal Salembier, Tuteur*

---

**Ecole doctorale :** *MITT*  
**Unité de recherche :** *IRIT (Institut de Recherche en Informatique de Toulouse)*  
**Directeur(s) de Thèse :** *Bernard Pavard*  
**Rapporteurs :**



## Remerciements

Vient enfin le temps des remerciements !!

Cette aventure n'aurait pas eu lieu sans l'exceptionnelle ouverture d'esprit de Bernard (Pavard), qui un beau jour m'a proposé, comme il sait si bien le faire, une fin de crédit et m'a ainsi fait mordre, pour longtemps je l'espère, à l'hameçon de l'ergonomie. C'est grâce à lui donc que j'ai pu entamer ce formidable cursus, et que j'ai pu avoir la chance d'évoluer dans une approche intellectuelle et théorique aussi originale que séduisante, une approche que j'aurai plus tard à assumer et à défendre à mon tour. Par cette pluridisciplinarité et cet éclectisme (pour le moins !), il a toujours su attirer ma curiosité vers des sujets semblant parfois éloignés du propos mais finalement devenant un jour ou l'autre pertinents. Merci encore Bernard.

Je tiens évidemment à remercier les membres du Jury, Laurence Nigay, Françoise Détienne, Christian Bastien, Julien Kahn et Pascal Salembier, d'avoir accepté d'examiner ce travail. Et je les remercie particulièrement d'avoir supporté mes nombreuses entorses faites au règlement, concernant notamment les délais serrés de lecture, ou encore pour les rapporteurs, l'examen de versions incomplètes.

Toutes mes pensées vont à l'ancienne équipe GRIC, aujourd'hui devenue un bout de IC3. Je commencerai évidemment par mon maître devant tous : Pascal Salembier, fidèle parmi les fidèles, qui, de par son exigence et sa discipline de fer m'a inculqué le plaisir de la souffrance au travail. Ce fut un vrai plaisir, à tous points de vue de travailler avec toi, et j'espère bien que nous n'en resterons pas là !

Parmi les anciens disciples, aujourd'hui tombés (comme moi) du côté obscur du Capital, il faut également citer les deux joyeux compères Moustapha Zouinar et Julien Kahn, qui m'ont fait découvrir les joies de la multimodalité, et de bien d'autres choses encore. Merci pour tout, en souvenir également de quelques soirées mémorables...

Toujours au sein du GRIC, et parmi les anciennes, une spéciale dédicace à Sandrine Darcy, dont je ne compte plus les services rendus, me rendant redevable pour toujours, voir au delà...

Un sourire amical à Julie Dugdale et Cécile. Et parmi les petits jeunes anciens du Gric, un gros salut à Nico, l'indispensable Italian Touch, célèbre et unique docteur distributeur de pages jaunes. Un scratch spécial pour AKKA Colin, le killer des dancefloors et des snowparks. Un grand bonjour également à Mehdi, qui doit être trop occupé à s'occuper de ses deux petits pour lire des remerciements de thèse.

Une pensée émue pour la maman du Gric, Chantal Morand, toujours là au bon moment, et toujours prête à recueillir et choyer ses nouveaux enfants, thésards et stagiaires, errant dans ce merveilleux monde de la recherche.

Et puis il y a les Autres... ceux des repas de midi à 11 h 31min et 43 sec, en particulier ce personnage étrange, Nadhem, toujours avec un lacet défait et une tâche quelque part, ayant toujours quelque chose à dire d'intéressant (quoique !!).

Un grand merci à Nicolas (collègue et néanmoins ami), pour les discussions enrichissantes qui ont contribué à certaines des conclusions de ce travail. Merci également à Ernest et Mister Olibou, ex-voisin et néanmoins ami, qui m'a prêté le portable qui couine me permettant de finaliser ce document.

Enfin, un grand merci à ma famille et à Cheval (en particulier la famille Ferrari) pour leur contribution morale, logistique, psychologique et folklorique évolutive à ce travail.

A ma sœur.



## Résumé :

L'ubiquité des dispositifs interactifs mobiles dans les sphères professionnelles, culturelles et domestiques constitue une occasion d'ouverture de nouveaux champs de recherche dans le domaine de l'étude des usages individuels et coopératifs des Nouvelles technologies de l'information et de la communication (NTICs). Les technologies mobiles amènent notamment à renouveler les points de vue sur des questions telles que, les problèmes d'utilisabilité spécifiques posés par la mobilité, les méthodes et les dispositifs de recueil de données, la question de l'identification des facteurs contextuels pertinents pour la conception de systèmes mobiles, la superposition des espaces d'interaction. De ce point de vue, nous posons que la mobilité constitue un nouveau paradigme d'interaction, générateur potentiel de variations contextuelles riches, pour lequel les styles d'interaction classiques basés sur la manipulation directe ne sont pas forcément les plus appropriés.

L'objectif de cette thèse est de documenter empiriquement la question de l'usage de dispositifs multimodaux en mobilité, de produire des éléments permettant de poser les bases d'un modèle de l'usage de dispositifs multimodaux en mobilité, et de fournir un ensemble de principes de recommandations pour la conception.

Dans un premier temps nous présentons une analyse critique de la littérature sur les questions relatives à la multimodalité, la mobilité et le Context Aware Computing (CAC). Nous définissons ensuite une méthode générale pour l'étude de l'usage de technologies multimodales en mobilité dans des situations écologiquement valides. Cette méthode, qui repose sur un dispositif d'enregistrement de données en situation et sur le recueil de verbalisations consécutives à l'activité assisté par les traces, est appliquée à une situation de jeu collaboratif en mobilité dans le contexte de la plateforme des usages LUTIN en situation naturelle (Cité des Sciences et de l'Industrie).

Dans un premier temps, nous présentons des résultats essentiellement quantitatifs qui permettent de décrire un ensemble d'usages des modalités, et d'examiner l'influence d'un certain nombre de facteurs contextuels (internes ou externes) sur les choix d'usage de ces modalités (contexte physique, social ou encore l'historique de l'interaction). Dans un second temps, une analyse plus fine de données qualitatives telles que les verbalisations des utilisateurs ou l'examen de chroniques d'activités, permet d'expliquer en partie ces choix d'usages mais également d'identifier de nouveaux facteurs potentiellement explicatifs.

Nous utilisons ensuite ces résultats pour proposer une ébauche de modélisation ainsi qu'un ensemble de recommandations pour la conception d'interfaces multimodales mobiles.

A Karen, et à nos trois filles



## Table des matières

<b>1. Introduction générale .....</b>	<b>14</b>
<b><u>Chapitre 1. Positionnement du problème et état de l'art .....</u></b>	<b><u>14</u></b>
<b>1. Introduction.....</b>	<b>14</b>
<b>2. Positionnement du problème .....</b>	<b>15</b>
2.1. La mobilité : un nouveau paradigme d'interaction ? Usages sédentaires, nomades et mobiles des systèmes interactifs .....	15
2.2. La mobilité comme facteur d'extension des variations contextuelles .....	15
2.3. Mobilité et gestion des ressources attentionnelles.....	16
2.4. La multimodalité et les systèmes sensibles au contexte comme solutions aux problèmes posés par l'interaction en mobilité.....	17
<b>3. La Mobilité : portabilité et usages des dispositifs d'interaction.....</b>	<b>18</b>
3.1. Introduction .....	18
3.2. Eléments de définition - Essai de taxonomie .....	18
3.2.1. Définitions.....	18
3.2.2. Catégories analytiques et taxonomies des situations mobiles .....	19
3.2.3. Classification "ontologique" .....	19
3.2.4. Classification "typologique" .....	21
3.2.5. Extension des cadres de description .....	22
3.2.6. Mobilité et activités collaboratives.....	22
3.3. Caractérisation des systèmes "mobiles" .....	24
<b>4. Relation Mobilité-Contexte.....</b>	<b>26</b>
4.1. Introduction .....	26
4.2. Eléments de définition.....	26
4.2.1. Le contexte en linguistique.....	26
4.2.2. Le contexte en psychologie.....	26
4.3. Prise en compte du contexte dans les systèmes.....	27
4.3.1. Systèmes de capture des informations contextuelles.....	27
4.3.2. Finalité de la capture d'informations contextuelles .....	29
4.3.3. Automatisation et Contrôle de l'application.....	29
4.3.4. Multimodalité & Context-aware computing.....	30
<b>5. Interaction multimodale : caractéristiques et usages des dispositifs .....</b>	<b>31</b>
5.1. Introduction .....	31
5.2. Multimodalité, neurosciences et physiologie .....	32
5.2.1. L'hypothèse centraliste invalidée ? .....	32
5.2.2. Le sens du mouvement : un sixième sens.....	33
5.2.3. Une perception active plutôt que passive.....	33
5.2.4. Les référentiels absolus et relatifs pour l'action .....	33
5.2.5. Modalité de contrôle.....	35
5.3. Multimodalité dans les IHM.....	36
5.3.1. Définitions.....	37
5.3.2. Espaces de conception.....	42
5.3.3. Etudes empiriques de la multimodalité en situation fixe et mobile .....	46
5.3.4. Usage des modalités .....	52
5.3.5. Facteurs influant l'usage de la multimodalité .....	57
<b>6. Problématique méthodologique .....</b>	<b>60</b>

6.1.	De l'étude de laboratoire (laboratory studies).....	60
6.2.	En passant par les expériences de terrain (field / quasi experiments).....	60
6.3.	A l'étude en contexte naturel (field studies).....	60
6.4.	Quels outils pour l'analyse des données ?.....	61
<b>7.</b>	<b><i>Synthèse et conclusion</i></b> .....	<b>62</b>
<b><u>Chapitre 2. Etude empirique : Méthode et cadre d'analyse</u></b> .....		<b>65</b>
<b>1.</b>	<b><i>Objectifs et principes Généraux</i></b> .....	<b>66</b>
<b>2.</b>	<b><i>Cadre d'analyse</i></b> .....	<b>67</b>
2.1.	Cadre d'analyse pour la mobilité.....	67
2.2.	Cadre d'analyse pour l'interaction multimodale.....	68
2.3.	Cadre général d'interaction : le modèle de Norman et Kirsh.....	68
2.3.1.	<i>Intérêts du modèle de Norman / Kirsh</i> .....	69
2.3.2.	<i>Critique du modèle de Norman / Kirsh</i> .....	70
2.3.3.	<i>Usages du modèle dans notre étude</i> .....	71
<b>3.</b>	<b><i>Méthode</i></b> .....	<b>72</b>
3.1.	Description de l'application.....	72
3.1.1.	<i>Architecture et fonctions de l'application</i> .....	72
3.1.2.	<i>Modalités d'interactions disponibles</i> .....	74
3.2.	Tâche et consigne.....	74
3.3.	Sujets et protocole expérimental.....	75
3.4.	Description du dispositif d'enregistrement et de recueil des données.....	76
3.4.1.	<i>Vue contextuelle</i> .....	76
3.4.2.	<i>Vue Subjective</i> .....	77
3.4.3.	<i>Enregistrement des fichiers Log (Logger)</i> .....	78
3.4.4.	<i>Sessions de Rejeu et d'autoconfrontations</i> .....	79
3.5.	Méthode et outil de codage des données.....	80
3.6.	Méthode de traitement des verbalisations.....	83
<b><u>Introduction des chapitres 3, 4, 5 et 6</u></b> .....		<b>84</b>
<b><u>Chapitre 3. Usages de la multimodalité</u></b> .....		<b>86</b>
<b>1.</b>	<b><i>Introduction</i></b> .....	<b>87</b>
<b>2.</b>	<b><i>Usage global des trois modalités équivalentes (stylet, vocal, g2d)</i></b> .....	<b>88</b>
2.1.	Usage global des modalités après décompte des commandes de correction (stylet, vocal, g2d).....	88
<b>3.</b>	<b><i>Usage des modalités par sujets</i></b> .....	<b>90</b>
3.1.	Nombre de commandes multimodales par sujets.....	90
3.2.	Répartition de l'usage des modalités par sujets.....	90
<b>4.</b>	<b><i>Evolution des usages par session</i></b> .....	<b>93</b>
4.1.	Evolution globale, tous sujets confondus.....	93
4.2.	Evolution intra-individuelle.....	93
<b>5.</b>	<b><i>Usage des commandes</i></b> .....	<b>95</b>
5.1.	Usage spécialisé de modalités par commandes au niveau global.....	95
5.2.	Evolution de l'usage global de modalité par commandes.....	96
5.3.	Spécialisation intra-individuelle et évolution des usages.....	97
<b>6.</b>	<b><i>Transitions et enchaînement des modalités</i></b> .....	<b>100</b>

6.1.	Enchaînement local des modalités .....	100
6.1.1.	<i>Changements de modalité pour les commandes multimodales et unimodales</i> .....	100
6.1.2.	<i>Changements de modalité en correction ou non</i> .....	102
6.2.	Enchaînement global des modalités .....	103
6.2.1.	<i>Sujet 1</i> .....	103
6.2.2.	<i>Sujet 2</i> .....	104
6.2.3.	<i>Sujet 3</i> .....	104
6.2.4.	<i>Sujet 4</i> .....	105
6.2.5.	<i>Sujet 5</i> .....	106
6.2.6.	<i>Sujet 6</i> .....	107
6.2.7.	<i>Synthèse des résultats sur les transitions au niveau global</i> .....	107
6.3.	Verbalisations concernant les séquences d'usage d'une modalité.....	107
6.3.1.	<i>Activation locale et globale de modalité (sujets 1, 2, 3, 6)</i> .....	108
6.3.2.	<i>Continuité et fluidité de l'interaction (sujets 1, 2, 3, 6)</i> .....	109
<b>7.</b>	<b><i>Les erreurs au niveau global (modalité vocale et tactile stylet)</i></b> .....	<b>110</b>
7.1.	Erreurs et commandes .....	110
7.2.	Les erreurs de la modalité vocale.....	112
7.3.	Les erreurs de la modalité tactile stylet .....	114
<b>8.</b>	<b><i>Verbalisations concernant les préférences individuelles</i></b> .....	<b>116</b>
8.1.	Test et appropriation de l'application et des modalités .....	116
8.1.1.	<i>Test de l'application et des modalités (S2, S4, S5)</i> .....	116
8.1.2.	<i>Moments de test (S1, S2)</i> .....	116
8.1.3.	<i>Aspect ludique, nouveauté (S1, S4, S5, S6)</i> .....	117
8.1.4.	<i>Contexte expérimental (S4)</i> .....	117
8.1.5.	<i>Test de la mise en œuvre des modalités (S1)</i> .....	117
8.1.6.	<i>Test de l'efficacité</i> .....	118
8.2.	Exigences de la tâche .....	119
8.2.1.	<i>Rapidité</i> .....	119
8.2.2.	<i>Criticité</i> .....	121
8.2.3.	<i>Effort de mise en œuvre</i> .....	122
8.2.4.	<i>Spécialisation (S1)</i> .....	126
<b>9.</b>	<b><i>Synthèse des résultats sur l'usage de la multimodalité</i></b> .....	<b>127</b>
9.1.	Usage des modalités au Niveau global .....	127
9.2.	Usage des modalités au Niveau individuel .....	127
9.3.	Evolution des usages au Niveau global .....	127
9.4.	Spécialisation et évolution des usages au niveau global.....	128
9.5.	Spécialisation et évolution des usages au niveau intra et inter-individuel .....	128
9.6.	Transitions de modalités.....	128
9.7.	Etudes des erreurs .....	129
9.8.	Préférences individuelles des modalités : .....	130
9.9.	Conclusion.....	131
<b><u>Chapitre 4. Facteurs contextuels influençant l'usage des modalités</u></b> .....		<b>132</b>
<b>1.</b>	<b><i>Introduction</i></b> .....	<b>133</b>
<b>2.</b>	<b><i>Influence du niveau sonore sur l'usage des modalités</i></b> .....	<b>134</b>
2.1.	Influence du niveau sonore sur les erreurs.....	134
2.2.	Influence globale du niveau sonore.....	134
2.3.	Influence individuelle du niveau sonore.....	135
2.4.	Perception subjective du niveau sonore.....	137

2.4.1.	<i>Gêne due au bruit (S3, S4, S6)</i> .....	137
2.4.2.	<i>Modification du mode opératoire (S6)</i> .....	139
2.4.3.	<i>Divers</i> .....	139
<b>3.</b>	<b><i>Influence de l'environnement social sur l'usage des modalités</i></b> .....	<b>141</b>
3.1.	Influence globale de l'environnement social .....	141
3.2.	Influence au niveau individuel de l'environnement social.....	142
3.3.	Perception subjective du contexte social .....	145
3.3.1.	<i>Expression de gêne liée au monde</i> .....	145
3.3.2.	<i>Expression d'absence de gêne liée à la présence de tiers</i> .....	148
3.3.3.	<i>Analyse et synthèse</i> .....	151
3.4.	Illustrations de commandes vocales en présence de tiers .....	151
<b>4.</b>	<b><i>Influence des niveaux d'activité sur l'usage des modalités</i></b> .....	<b>153</b>
4.1.	Influence de l'espace de travail (fenêtre) sur l'usage des modalités.....	153
4.1.1.	<i>Temps passé dans chaque espace de l'application</i> .....	153
4.1.2.	<i>Niveau global (tous sujets, toutes interactions)</i> .....	154
4.1.3.	<i>Niveau individuel</i> .....	154
4.2.	Influence de l'activité réelle sur l'usage des modalités .....	155
4.2.1.	<i>Durées de l'activité réelle</i> .....	156
4.2.2.	<i>Influence globale et individuelle de l'activité réelle sur l'usage des modalités</i> .....	156
4.2.3.	<i>Perception subjective de l'influence de l'activité</i> .....	158
<b>5.</b>	<b><i>Historique de l'interaction</i></b> .....	<b>160</b>
5.1.	Historique local (Intra-session) et inter-session : étude de cas.....	160
<b>6.</b>	<b><i>Synthèse des résultats sur les facteurs contextuels influençant l'usage des modalités</i></b> .....	<b>163</b>
6.1.	Contexte sonore .....	163
6.2.	Contexte Social.....	163
6.2.1.	<i>Expression de gêne liée au monde</i> .....	164
6.2.2.	<i>Expression d'absence de gêne liée à la présence de tiers</i> .....	164
6.3.	Activité en cours .....	164
6.3.1.	<i>Codage de l'activité via les espaces de travail de l'application (liste, édition, lecture)</i> .....	164
6.3.2.	<i>Influence de l'activité réelle sur l'usage des modalités</i> .....	165
6.4.	Historique de l'interaction .....	165
<b><u>Chapitre 5. Multimodalité, mobilité et ressources attentionnelles</u></b> .....		<b>166</b>
<b>1.</b>	<b><i>Introduction</i></b> .....	<b>167</b>
<b>2.</b>	<b><i>Observables des ressources attentionnelles</i></b> .....	<b>168</b>
<b>3.</b>	<b><i>Mobilité et interaction : Effets des modalités ?</i></b> .....	<b>169</b>
3.1.	Répartition temporelle et numérique globale des catégories de mobilité .....	169
3.2.	Mobilité et modalités.....	170
3.3.	Variations interindividuelles de mobilité.....	171
3.4.	Evolution de la mobilité au cours des sessions .....	173
3.5.	Mobilité et commandes .....	173
<b>4.</b>	<b><i>Direction du regard</i></b> .....	<b>176</b>
4.1.	Répartition temporelle de la direction du regard et interaction .....	176
4.2.	Répartition des directions du regard selon les modalités (Mode global) .....	176
4.3.	Direction du regard selon les sujets.....	177
4.4.	Positionnement du pda lors d'une commande vocale.....	179
4.5.	Répartition des directions du regard selon les modes Correction ou Premier .....	180

4.6.	Regard et mobilité.....	181
4.7.	Verbalisations concernant la catégorie « regard ».....	182
<b>5.</b>	<b><i>Mobilité et Activité : Effet de l'activité ?</i></b> .....	<b>188</b>
<b>6.</b>	<b><i>Transitions entre les états de mobilité</i></b> .....	<b>190</b>
6.1.	Changements de mobilité .....	190
6.2.	Changements de mobilité selon les modalités.....	191
<b>7.</b>	<b><i>Correction des erreurs</i></b> .....	<b>193</b>
7.1.	Influence des corrections sur la mobilité .....	193
7.2.	Influence des modalités de correction sur la mobilité .....	193
<b>8.</b>	<b><i>Perceptions des utilisateurs</i></b> .....	<b>195</b>
8.1.	Remarques générales liées à la mobilité et à l'application.....	195
8.2.	Remarques liées à la mobilité et aux modalités .....	198
<b>9.</b>	<b><i>Synthèse des résultats : Multimodalité, mobilité et ressources attentionnelles</i></b> .....	<b>201</b>
9.1.	Mobilité et interaction : Effets des modalités .....	201
9.2.	Regard et multimodalité .....	201
9.3.	Regard et mobilité (simultanéité).....	202
9.4.	Activité réelle et mobilité : Effet de l'activité.....	202
9.5.	Transitions de mobilité.....	203
9.6.	Résultats qualitatifs (verbalisations) sur la mobilité et les ressources attentionnelles .....	203
9.6.1.	<i>Remarques générales liées à la mobilité et à l'application</i> .....	203
9.6.2.	<i>Remarques liées à la mobilité et aux modalités</i> .....	204
<b><u>Chapitre 6. Saisie de médias</u></b> .....		<b>205</b>
<b>1.</b>	<b><i>Introduction</i></b> .....	<b>206</b>
<b>2.</b>	<b><i>Usages de la saisie de média par sujets</i></b> .....	<b>207</b>
<b>3.</b>	<b><i>Usage des médias en fonction du contexte</i></b> .....	<b>208</b>
<b>4.</b>	<b><i>Verbalisations concernant la saisie de médias</i></b> .....	<b>212</b>
<b>5.</b>	<b><i>Synthèse des résultats sur la saisie de médias</i></b> .....	<b>219</b>
<b><u>Chapitre 7. Synthèse, discussion et apports de la thèse</u></b> .....		<b>221</b>
<b>1.</b>	<b><i>Introduction</i></b> .....	<b>223</b>
<b>2.</b>	<b><i>Etude de cas et illustrations</i></b> .....	<b>224</b>
<b>3.</b>	<b><i>Synthèse des propriétés effectives et subjectives des modalités</i></b> .....	<b>229</b>
<b>4.</b>	<b><i>Proposition d'un modèle enrichi pour l'interaction mobile</i></b> .....	<b>231</b>
4.1.	Objectifs .....	231
4.2.	Intention et action .....	231
4.3.	Un modèle détaillé de l'action .....	231
4.3.1.	<i>Préparation (action utilisateur hors système)</i> .....	233
4.3.2.	<i>Entrée ou commande (action utilisateur sur le système)</i> .....	233
4.3.3.	<i>Modalité de contrôle (action utilisateur hors système) et types de repères</i> .....	233
4.3.4.	<i>Désengagement (action utilisateur hors système)</i> .....	234
4.3.5.	<i>Feedback de prise en compte de la commande (action système)</i> .....	234
4.3.6.	<i>Feedback lié à l'effet de la commande (action système)</i> .....	234

4.3.7.	<i>Organes et sens mis en œuvre</i> .....	235
4.3.8.	<i>Remarques sur la manipulation directe</i> .....	235
<b>5.</b>	<b><i>Proposition de modèle pour une mobilité finalisée</i></b> .....	<b>236</b>
5.1.	Objets de la tâche .....	238
5.2.	Finalité de la tâche et de la mobilité .....	238
5.3.	Objets de la mobilité .....	239
5.4.	Combinaisons possibles des objets de la tâche .....	239
<b>6.</b>	<b><i>Critères et recommandations pour la conception et l'évaluation d'interfaces mobiles multimodales</i></b> .....	<b>241</b>
6.1.	Compatibilité * .....	241
6.2.	Guidage .....	243
6.2.1.	<i>Lisibilité</i> * .....	243
6.2.2.	<i>Incitation</i> * .....	244
6.2.3.	<i>Groupement/Distinction entre items</i> .....	244
6.2.4.	<i>Feed-back immédiat</i> * .....	246
6.3.	Contrôle explicite .....	246
6.3.1.	<i>Actions explicites</i> * .....	246
6.3.2.	<i>Contrôle utilisateur</i> * .....	247
6.4.	Signifiante des codes, dénominations et comportements * .....	248
6.5.	Charge de travail .....	248
6.5.1.	<i>Charge physique</i> * .....	248
6.5.2.	<i>Brièveté</i> .....	249
6.5.3.	<i>Densité informationnelle</i> * .....	251
6.6.	Adaptabilité .....	251
6.6.1.	<i>Prise en compte de l'expérience de l'utilisateur</i> * .....	251
6.6.2.	<i>Flexibilité</i> * .....	252
6.7.	Homogénéité/Cohérence * .....	252
6.8.	Gestion des erreurs .....	253
6.8.1.	<i>Protection contre les erreurs</i> * .....	253
6.8.2.	<i>Qualité des messages d'erreur</i> * .....	253
6.8.3.	<i>Correction des erreurs</i> * .....	253
<b><u>Chapitre 8. Conclusion Générale et perspectives</u></b> .....		<b>255</b>
<b>1.</b>	<b><i>Introduction</i></b> .....	<b>256</b>
<b>2.</b>	<b><i>Rappel des résultats et apports de la thèse</i></b> .....	<b>257</b>
2.1.	Résultats empiriques : documentation des usages de dispositifs multimodaux en mobilité .....	257
2.2.	Résultats empiriques : apports méthodologiques .....	259
2.3.	Résultats théoriques .....	260
2.4.	Résultats technologiques .....	260
<b>3.</b>	<b><i>Perspectives</i></b> .....	<b>261</b>
3.1.	Conception d'applications sensibles au contexte (Contexte Aware Computing) .....	261
3.2.	Développement d'applications mobiles à ressources attentionnelles limitées .....	262
3.3.	Validation des critères et recommandations .....	262
3.4.	Perspectives méthodologiques .....	262
3.4.1.	<i>De l'utilisabilité à l'expérience utilisateur</i> .....	263
3.4.2.	<i>Etude de l'appropriation des applications dans le temps</i> .....	263
<b><u>Chapitre 9. Bibliographie et annexes</u></b> .....		<b>265</b>

**1. Bibliographie ..... 266**

**2. Annexes ..... 273**

2.1. Index des critères et recommandations pour la conception et l'évaluation des interfaces multimodales mobiles 273

2.2. Outils d'annotation de corpus multimodaux .....276

2.3. Index des figures .....277

2.4. Index des tableaux.....279





## 1. Introduction générale

La conjonction de rapides progrès technologiques, (notamment miniaturisation), et de la généralisation de l'accès ubiquitaire aux réseaux, rend possible l'évolution des pratiques d'interaction en mobilité qui permet aux utilisateurs de s'extirper des situations classiques d'interaction « sédentarisée ». Cette évolution est la résultante des progrès techniques, et de l'augmentation d'une demande liée à la nomadisation de nombreuses activités, qu'elles soient de loisirs, grand public ou professionnelles. La mobilité devient une composante à part entière d'un nombre grandissant d'activités, soit en tant que finalité (ou du moins pré requis) d'une tâche principale mobile, soit en tant que contexte d'une tâche principale indépendante du déplacement.

Dans un cas comme dans l'autre, le caractère mobile de l'interaction va provoquer des changements importants sur la nature même de l'interaction. En effet, comme l'indique (Kristoffersen, Steinar & Ljungberg, 1999), jusqu'à récemment, le principal paradigme d'interaction utilisé pour la conception d'interfaces, fixes ou mobiles, était le paradigme de la « manipulation directe » qui repose en grande partie sur une interaction très consommatrice de ressources attentionnelles visuelles.

Cette problématique a été soulevée par un certain nombre d'auteurs (Kristoffersen, Steinar & Ljungberg, 1999), (Roto *et al.*, 2004), qui soulignent plusieurs caractéristiques nouvellement identifiées de l'interaction en mobilité, notamment le fait que la réalisation de la tâche principale en situation d'activité mobile ne nécessite pas directement l'utilisation du dispositif informatique mais que celui-ci joue souvent un rôle facilitateur, remplit une fonction de médiation (communication),

Si miniaturisation et donc portabilité accrues permettent d'envisager dès maintenant des interactions réellement mobiles, c'est à dire ne se traduisant plus forcément par des interruptions des déplacements, et intégrant éventuellement des supports à la mobilité (guidage, support aux tâches mobiles etc.), il faut néanmoins également envisager la mobilité comme facteur de variations contextuelles accrues, qui peuvent rendre plus difficile ou moins fluide l'interaction, qui n'est pas forcément encore adaptée à la mobilité.

On peut en effet considérer que le glissement des situations d'interaction vers plus de mobilité a précédé un réel changement des styles d'interaction dont on souhaiterait qu'ils soient plus adaptés à la mobilité. Ce décalage entre extension de l'utilisation de dispositifs portables et conservation des modalités traditionnelles d'interaction est à l'origine des situations problématiques observables au quotidien (il suffit d'observer les difficultés que rencontre un utilisateur de téléphone portable en essayant de prendre des notes à la volée dans la rue). (Anhalt *et al.*, 2001) déclarent à ce sujet qu'un effort conséquent a été fourni sur le développement des technologies, mais que parallèlement à cela, un facteur limitant a été négligé à savoir **la gestion de l'attention de l'utilisateur**, la capacité d'une personne à se focaliser sur sa tâche principale.

On peut aller plus loin dans le constat en remarquant que l'introduction de la mobilité va potentiellement jusqu'à modifier la nature même des tâches que l'utilisateur va pouvoir effectuer (Nielsen, J., 1995; Pascoe, Ryan, & Morse, 2000).

Ceci peut-être rendu possible uniquement à condition que les styles d'interaction proposés puissent s'adapter à la tâche principale des utilisateurs. Dans ces différentes situations, les utilisateurs peuvent être soumis à des contraintes très fortes pouvant limiter le champ des interactions possibles, voire les rendre impossibles. Un constat que l'on peut en effet faire à propos de l'interaction en mobilité est qu'elle est par nature une activité multitâche qui provoque de fait une répartition des ressources physiques et cognitives plus ou moins volontairement contrôlée. En particulier, l'allocation de ressources attentionnelles devient critique lorsqu'il s'agit d'effectuer des interactions en mobilité « réelle », quand l'utilisateur doit au minimum gérer son propre déplacement et l'interaction avec l'artefact proprement dite.

Plusieurs pistes complémentaires de conception sont généralement envisagées (Anhalt *et al.*, 2001; Arroyo & Selker, 2000; Jaimes & Sebe, 2005; Pascoe, 2001), afin de pallier aux contraintes induites par la mobilité sur l'interaction. Il se dégage en particulier deux pistes intéressantes : l'interaction multimodale et

les systèmes sensibles au contexte (« context-aware computing » ou CAC), toutes deux pouvant potentiellement venir alléger la charge attentionnelle provoquée par l'interaction avec le dispositif.

En ce qui concerne le CAC, celui-ci, pourrait permettre un allègement de l'interaction via la diminution (qu'il s'agisse d'élimination ou d'automatisation) du nombre d'actions effectuées par l'utilisateur grâce aux routines déclenchées par la contextualisation de l'application.

Dans le cas de la multimodalité, la philosophie est différente bien que les objectifs poursuivis soient les mêmes : diminution de la charge cognitive et en particulier du poids des ressources attentionnelles mobilisées par l'interaction. La multimodalité offre en effet, la possibilité d'adapter les modalités d'entrées et de sorties aux ressources et contraintes contextuelles (Oviatt, S.L., 1999). Toujours dans cet esprit, (Oulasvirta, 2005) propose de porter une attention particulière à l'usage pertinent des ressources physiques et cognitives disponibles, par exemple en déchargeant des tâches sur des ressources non utilisées ou dans des modalités différentes, mais également en fournissant des feedbacks de modalités différentes selon les temps de délais de réponse du système.

Dans cette thèse nous procéderons dans un premier temps à un passage en revue des études menées sur les questions suivantes : activités instrumentées en mobilité, relation entre mobilité et contexte et enfin multimodalité. Même si par commodité chacune de ces questions est abordée de manière autonome, nous chercherons systématiquement à mettre en évidence la nature de leurs relations.

- Activités instrumentées en mobilité

Jusqu'à récemment la plupart des études sur le travail mobile se sont essentiellement focalisées sur des aspects technologiques (réseaux mobiles, RPC, systèmes de positionnement par satellite,...) ou sur des problématiques d'évaluation de l'utilisabilité de dispositifs portables tels que les PDAs. Cette double orientation renvoie à une posture de questionnement de la notion de mobilité uniquement fondée sur les conditions logistiques de la mobilité d'un dispositif technique qui doit rester fonctionnel quelque soit l'environnement dans lequel il est utilisé, et sur la compatibilité entre caractéristiques de portabilité et conditions d'utilisation efficace de l'outil (notamment ses caractéristiques physiques).

Par ailleurs, les possibilités ouvertes par l'instrumentation des environnements et l'informatique "pervasive" (bornes intégrées, capteurs,...) permet de reconsidérer la mobilité comme une modalité d'interaction à part entière (intentionnelle ou non, voir par exemple (Dix *et al.*, 2000)) grâce à de nouveaux modes d'interaction entre terminaux (fixes ou mobiles). Cette nouvelle approche nécessite de réfléchir à la définition de cadres de description et d'espaces de conception qui intègrent la notion de mobilité.

Afin de saisir l'ensemble des points pertinents spécifiques des activités mobiles, il convient d'étendre ce cadre initial de positionnement de la problématique en y intégrant notamment la nature des tâches à réaliser, les implications sociales de la mobilité (organisation collective de la mobilité) et l'articulation avec les aspects contextuels

- Contexte

Présente dans de nombreux champs disciplinaires, la notion de contexte connaît de multiples définitions plus ou moins précises. Bien qu'il n'existe pas une définition stable et partagée de cette notion, un point commun à une grande partie de ces définitions est l'idée que le contexte se réfère à des relations entre un élément particulier (individu, objet, processus, etc.) et un ensemble d'autres éléments qui sont intrinsèques ou extrinsèques à ce même élément. Nous nous attachons dans cette partie à : définir la notion de contexte au regard des conséquences de la mobilité sur l'interaction Homme-Machine ; recenser les modes et les techniques de capture et de traitement des informations contextuelles ; poser le problème de l'articulation entre contexte et multimodalité dans le cadre de l'étude.

- Interaction multimodale

Les progrès technologiques réalisés au cours de la dernière décennie ont ouvert de nouvelles perspectives dans le domaine de l'interaction homme-machine. L'une de ces perspectives explore la conception des interfaces ou systèmes multimodaux, c'est-à-dire des systèmes capables de permettre aux utilisateurs d'interagir via l'usage de différentes modalités communicationnelles (parole, gestes, etc.), mais aussi de traiter et de « comprendre » des actions ou expressions multimodales (combinaison de différentes modalités pour réaliser une même intention ou réalisation simultanée d'intentions différentes à

l'aide de modalités différentes). L'hypothèse qui sous-tend cette perspective est que de tels systèmes devraient rendre l'interaction homme-machine plus efficace, plus robuste et plus flexible (Oviatt, 1999). Par exemple, sur le plan de l'utilisabilité, on suppose que l'un des avantages de ce type d'interfaces est de permettre aux utilisateurs de choisir la modalité la plus efficace et la plus adaptée au contexte d'utilisation (caractéristiques de l'environnement physique, de l'utilisateur, de la tâche, etc.). Cet avantage est particulièrement intéressant lorsque le système est utilisé en situation de mobilité où certains aspects du contexte peuvent changer rapidement. Un autre avantage attendu de l'utilisation de ces interfaces est qu'ils facilitent la récupération d'erreurs en permettant par exemple aux utilisateurs de changer de modalité lorsque l'usage d'une modalité donnée génère des incompréhensions entre l'utilisateur et le système. Par ailleurs, on s'attend à ce que ces systèmes soient plus efficaces dans l'interprétation des actions de l'utilisateur grâce notamment à la « désambiguïsation mutuelle » (Oviatt (Oviatt, S.L., 1999), 1999) qui se réfère à la capacité du système à intégrer et confronter les informations provenant des différents modules d'interprétation des données issues des différentes sources d'entrée (modalités). L'intégration permet au système de compléter les informations provenant des différents modules, et de produire ainsi des interprétations plus fiables. La confrontation peut permettre au système de repérer des ambiguïtés ou des incohérences entre les traitements réalisés dans les différents modules.

Cette étude s'inscrit dans la continuité d'un premier travail de DEA mené sur l'utilisation d'un système portable multimodal en laboratoire (Calvet et al., 2001). Notre objectif est, dans le contexte de cette thèse, de documenter la question de l'usage de dispositifs d'interaction mobiles en environnement réel, de tester les effets potentiels de la mobilité sur l'usage d'un système interactif en situation réelle, et d'étudier empiriquement en quoi la versatilité des systèmes multimodaux, souvent présentée comme un de leurs avantages principaux (cf plus haut) peut aider l'utilisateur à gérer les contraintes induites par les changements de contexte inhérents à la mobilité (modification de l'environnement physique et social notamment).

D'un point de vue méthodologique, les artefacts mobiles posent également de nouveaux problèmes relatifs à la conception des dispositifs et des méthodes d'évaluation à mettre en œuvre pour l'étude empirique de leur usage en situation naturelle. On distingue ainsi par convention deux approches paradigmatiques : les études menées en laboratoire et les études de terrain. Les études en laboratoire se focalisent sur le contrôle de variables susceptibles d'influer sur l'interaction ; elles requièrent de recréer artificiellement une situation de mobilité (Beck, Christiansen, Kjeldskov, Kolbe, & Stage, 2003) qui se traduit généralement par un appauvrissement du contexte. Les études naturalistes insistent quant à elles sur le critère de validité écologique et s'efforcent de préserver la richesse contextuelle des activités mobiles. Récemment certains auteurs ont proposé sous les termes de « field experiments » (Pirhonen, Brewster, & Holquin, 2002) et de « quasi-experiments » (Roto et al., 2004) une voie médiane qui tente de concilier contrôle expérimental des variables et réalisme de la situation. Cette approche demeure essentiellement quantitative et s'inscrit dans la tradition orthodoxe de la psychologie expérimentale appliquée à l'interaction Homme-machine.

Après avoir présenté une rapide revue des études et des problématiques méthodologique, nous présenterons le type d'approche que nous avons retenue. En effet, le travail présenté ici se rattache à la catégorie des études naturalistes, mais il se donne des contraintes de rigueur quant aux phases de recueil et d'analyse des observables. Il s'efforce en outre d'articuler les données quantitatives (qui constituent des éléments de synthèse commodes et permettent de résumer les tendances générales des résultats) avec des données contribuant à donner du sens aux comportements observés (verbalisations des sujets).

Nous présenterons pour ce faire (Etude empirique : Méthode et cadre d'analyse, P65) un dispositif général d'étude et de recueil des activités mobiles en situation naturelle, mis en œuvre à l'occasion d'une étude empirique menée dans le contexte de la plateforme d'évaluation des usages LUTIN à la Cité des Sciences et de l'Industrie de Paris-La Villette.

Cette méthode, qui repose sur un dispositif d'enregistrement de données en situation naturelle et de recueil de verbalisations consécutives à l'activité assisté par les traces, est appliquée à une situation de jeu collaboratif en mobilité utilisant comme ressources les objets de l'environnement présents dans les expositions ouvertes au public.

## Introduction générale

Nous présenterons ensuite des résultats essentiellement quantitatifs qui permettent de décrire un ensemble d'usages des modalités (Chapitre 3, p86), et d'examiner l'influence (Chapitre 4, p132) d'un certain nombre de facteurs contextuels (internes ou externes) sur les choix d'usage de ces modalités (contexte physique, social ou encore l'historique de l'interaction). Le Chapitre 5 (p166) présente, quant à lui, un ensemble de résultats concernant les usages de la multimodalité en mobilité et de son influence sur les ressources attentionnelles.

Nous présentons également un ensemble de résultats concernant la saisie de média en mobilité (Chapitre 6, p205).

Lorsque les données le permettent, tout au long de cette première approche quantitative, nous procédons à une analyse plus fine de données qualitatives telles que les verbalisations des utilisateurs ou l'examen de chroniques d'activités, qui permet d'expliquer en partie les choix d'usages mais également d'identifier de nouveaux facteurs potentiellement explicatifs.

Nous présentons au Chapitre 7 (p221), une synthèse des résultats obtenus et discutons des apports de cette thèse. Nous utilisons ensuite ces résultats sous la forme d'une étude de cas détaillée, permettant de dégager plusieurs apports, constitués par une synthèse des propriétés des modalités et médias mis en œuvre ici, puis par la proposition d'une ébauche de modélisation concernant les finalités de la mobilité et de l'interaction en général.

Finalement, nous proposerons également un ensemble de recommandations pour la conception d'interfaces multimodales mobiles en reprenant la classification utilisée par (Bastien & Scapin, 1993) et mise à jour pour les environnements virtuels par (Bach, 2004).

Enfin, nous présenterons la conclusion et les perspectives de ce travail (Chapitre 8 Conclusion Générale et perspectives P255).

**Chapitre 1. Positionnement du problème et  
état de l'art**

## 1. Introduction

La première partie de ce chapitre (Positionnement du problème, p15) présente rapidement la problématique associée à l'interaction multimodale en mobilité en expliquant en quoi cette mobilité constitue un nouveau paradigme d'interaction présentant un certain nombre de caractéristiques liées, entre autre, à la variabilité des situations d'usages qu'elle peut provoquer. Cette variabilité présente la capacité de générer des variations de facteurs contextuels, ayant un impact potentiel sur la gestion des ressources attentionnelles des utilisateurs.

Cette première partie permet de mettre en lumière plusieurs thématiques que sont la mobilité, le contexte et la multimodalité, que nous nous proposons d'étudier dans les parties suivantes constituant l'état de l'art.

Nous commençons par étudier la notion de mobilité (La Mobilité : portabilité et usages des dispositifs d'interaction, p18) à travers ses différentes acceptions dans la littérature sur le sujet, tout en tentant de se dégager d'une approche technocentrée relativement répandue et en nous intéressant plus aux aspects liés aux usages en mobilité.

Dans une deuxième partie (Relation Mobilité-Contexte, p26), nous montrons à travers une revue d'études, la relation existant entre mobilité et contexte, tout en présentant les différentes définitions de cette dernière notion selon différentes disciplines.

Nous procédons également à une revue des systèmes sensibles au contexte (CAC) ainsi qu'à la finalité de cette capture du contexte (par exemple le déclenchement de comportements automatiques selon la situation).

Nous étudions ensuite la notion de multimodalité (Interaction multimodale : caractéristiques et usages des dispositifs, p31) sous des angles disciplinaires différents tels que les neurosciences et la psychologie. Nous en présentons ensuite son traitement dans la littérature de l'interaction homme-machine, à la fois en ce qui concerne ses définitions et l'étude de ses usages à travers une revue des applications mettant en œuvre la multimodalité.

Une dernière partie (Problématique méthodologique, 60), avant la conclusion, présente la problématique méthodologique qu'introduit l'étude des usages des applications multimodales mobiles en situation naturelle en dressant un rapide panorama des divers positionnements méthodologiques et théoriques et leurs conséquences sur les méthodes effectivement mises en œuvre par la communauté scientifique du domaine.

## 2. Positionnement du problème

### 2.1. *La mobilité : un nouveau paradigme d'interaction ? Usages sédentaires, nomades et mobiles des systèmes interactifs*

Le caractère mobile de l'interaction provoque des changements importants sur la nature même de l'interaction. En effet, comme l'indique (Kristoffersen, Steinar & Ljungberg, 1999), jusque récemment, le principal paradigme d'interaction utilisé pour la conception d'interfaces qu'elles soient fixes ou mobiles était le paradigme de "la manipulation directe"; qui fait appel à trois grands principes :

- « Une représentation continue des objets et des actions d'intérêt à l'aide de métaphores visuelles significatives »
- « Des actions physiques ou des pressions sur des boutons avec labels au lieu de la saisie d'une syntaxe complexe »
- « Des opérations incrémentales rapides et réversibles dont les effets sur les objets d'intérêts sont visibles immédiatement »

Ce paradigme repose donc en grande partie sur une interaction très consommatrice de ressources attentionnelles visuelles, puisqu'il requiert la plupart du temps la mise en place et le contrôle d'activités balistiques : identification d'une cible, visée pour lancer l'action correspondante et enfin vérification du bon déroulement de l'opération.

Autant ce paradigme peut convenir dans un contexte d'interaction dit « fixe », où les ressources visuelles sont principalement dédiées à la gestion de l'interaction, autant cela devient problématique dans un contexte de mobilité où va apparaître de façon critique le problème de partage de ressources, qu'elles soient perceptives, cognitives ou motrices.

Cette problématique a été soulevée par un certain nombre d'auteurs (Kristoffersen, Steinar & Ljungberg, 1999), (Roto et al., 2004), qui soulignent tous des caractéristiques nouvellement identifiées de l'interaction en mobilité :

- La tâche principale de l'activité mobile se déroule souvent à « l'extérieur » de l'ordinateur par opposition aux tâches prenant place à « l'intérieur » de l'ordinateur.
- Les mains des utilisateurs sont de fait souvent mobilisées pour la manipulation d'objets physiques et non pas dédiées exclusivement à l'interaction.
- Les utilisateurs sont souvent mobilisés par des tâches extérieures à l'interaction, qui sont très demandeuses en attention visuelle.
- Durant ces tâches les utilisateurs peuvent être extrêmement mobiles par opposition aux situations sédentaires.

De plus, ces contraintes liées au caractère mobile de l'interaction viennent s'ajouter aux contraintes liées à la miniaturisation des surfaces d'affichage et de contrôle des dispositifs, rançon de leur portabilité accrue. Les terminaux se voient en effet dotés systématiquement de petits écrans, et de manière générale d'une faible surface d'interaction que ce soit en sortie visuelle (écrans, LED) et sonore (hauts parleurs), ou en entrée (boutons logiciels et physiques). Toutes ces caractéristiques ont donc tendance à invalider le modèle dominant actuel de style d'interaction à savoir la manipulation directe, telle qu'elle est déclinée sur les supports traditionnels fixes.

### 2.2. *La mobilité comme facteur d'extension des variations contextuelles*

La mobilité vient également questionner les modes classiques d'interaction du fait de la multiplication des occasions de passages d'un contexte d'interaction à un autre. D'un paradigme où l'interaction avec un ordinateur était généralement motivée par la réalisation d'une tâche principale, voire unique, avec un certain nombre de paramètres contextuels peu fluctuants (par exemple pour le contexte social : présence ou non de collègues temporairement, travail en open space etc.), voire des paramètres figés (par exemple pour le contexte spatial : localisation fixe), et des ressources de travail relativement stables (par exemple pour les ressources matérielles et informationnelles : imprimantes disponibles, accès au réseau, accès à

des ressources physiques ou numériques telles que documentation ou bases de données, etc.), on passe à un contexte d'interaction mobile labile et fluctuant.

Les nouvelles possibilités de portabilité des dispositifs (PDA et autres Smartphones) permettent d'envisager un nouveau pas dans la direction de l'interaction « réellement mobile », c'est à dire une interaction ne venant plus forcément interrompre un déplacement, voire dans certains cas, une interaction facilitant la mobilité (guidage, support aux tâches mobiles,...). Cependant, il faut également envisager la mobilité comme facteur de variations contextuelles qui peuvent rendre plus difficile ou moins fluide l'interaction, qui n'est pas forcément encore adaptée à ce changement de paradigme. En effet, on peut constater que l'augmentation de la portabilité des dispositifs favorisant potentiellement la mobilité a eu lieu avant un réel changement du style d'interaction, provoquant des situations problématiques observables au quotidien (voir par exemple les difficultés rencontrées par un utilisateur de téléphone portable qui tente de prendre des notes à la volée dans la rue). (Anhalt et al., 2001) déclarent à ce sujet que si un effort conséquent a été fourni en ce qui concerne le développement des technologies, des facteurs limitant tels que la gestion de ses ressources attentionnelles par l'utilisateur, sa capacité à se focaliser sélectivement sur sa tâche principale, ont été négligés.

La mobilité vient donc potentiellement amplifier des phénomènes de changements de contextes, incluant une palette très variée de déterminants, allant du contexte social, à la localisation spatiale, aux facteurs physiques d'ambiance (sonore, lumineux), au contexte d'organisation proxémique (espace encombré, dégagé et), aux ressources disponibles (réseau, imprimante, accès à l'information). Il est important d'aller plus loin dans le constat en remarquant que l'introduction de la mobilité va jusqu'à modifier potentiellement la nature même de la tâche que l'utilisateur va pouvoir effectuer, ouvrant ainsi de nouveaux champs de développement des dispositifs d'interaction dans des contextes de tâches auparavant difficiles à instrumenter (Nielsen, J., 1995; Pascoe, Ryan, & Morse, 2000). Il est donc maintenant théoriquement possible d'envisager d'outiller numériquement toute tâche mobile, qu'elle relève de l'interaction sociale (travail collaboratif mobile), de la prise d'information sur le terrain (relevés ou comptages) ou encore de l'intervention sur des objets physiques (maintenance mobile par exemple), ceci à conditions que les styles d'interactions proposés puissent s'adapter à la tâche principale et à ses différents contextes de réalisation. Dans ces nouvelles situations d'usage, les utilisateurs peuvent être soumis à des contraintes très fortes pouvant limiter le champ des interactions possibles, voire les rendre impossibles.

### **2.3. Mobilité et gestion des ressources attentionnelles**

L'interaction en mobilité est par nature une activité multitâche provoquant de fait un partage des ressources physiques et cognitives plus ou moins volontairement contrôlé. En particulier, le partage de l'attention devient critique lorsqu'il s'agit d'effectuer des interactions en mobilité « réelle », où l'utilisateur doit au minimum gérer son propre déplacement, et l'interaction avec l'artefact proprement dite. Comme le souligne (Oulasvirta, 2005), ce genre de situations provoque une fragmentation de l'attention dont le focus passe constamment d'une tâche à l'autre, parfois sur des durées très faibles, pouvant tomber à moins de 4 secondes. Cette nature multitâche de l'activité s'accompagne donc d'interruptions permanentes de la tâche en cours, obligeant l'utilisateur à de lourds et permanents réajustements de chaque activité.

Les auteurs vont jusqu'à observer des stratégies mises en œuvre par les utilisateurs pour limiter les conséquences de ces interruptions. Celles-ci peuvent relever de simples « replis stratégiques » afin de prioriser des tâches plus importantes comme par exemple le ralentissement de la marche afin de libérer des ressources pour l'interaction ou le refus ou la mise en attente de tâches moins prioritaires. Des stratégies plus élaborées ont également été observées, telles que l'anticipation de situations connues, pour lesquelles les utilisateurs configurent par avance leurs ressources sachant qu'ils seront moins disponibles ultérieurement.

Il est donc important suite à ces constats, d'insister sur l'importance des choix de conception si l'on souhaite réellement soutenir les activités mobiles qui doivent par nature s'effectuer dans des contextes changeants, difficiles voire concurrents ou adverses, l'intérêt étant d'inverser la tendance observée actuellement en tentant de transformer d'éventuelles contraintes contextuelles en ressources potentielles de l'activité de l'utilisateur (incluant évidemment l'interaction).

#### **2.4. La multimodalité et les systèmes sensibles au contexte comme solutions aux problèmes posés par l'interaction en mobilité**

Plusieurs pistes de conception sont généralement envisagées (Anhalt et al., 2001; Arroyo & Selker, 2000; Jaimes & Sebe, 2005; Pascoe, 2001), afin de pallier les contraintes de la mobilité sur l'interaction telles que nous venons de les décrire.

Il se dégage en particulier deux pistes intéressantes que sont l'interaction multimodale et les systèmes sensibles au contexte (context-aware computing : CAC), toutes deux pouvant potentiellement venir alléger ou diminuer la charge attentionnelle provoquée par l'interaction rendue parfois d'autant plus complexe que les interfaces se réduisent en même temps que la taille des dispositifs.

En ce qui concerne le CAC, celui-ci, pourrait permettre un allègement de l'interaction via la diminution (qu'il s'agisse d'élimination ou d'automatisation) du nombre d'actions effectuées par l'utilisateur grâce aux automatismes déclenchés par la contextualisation de l'application. En effet, de nombreuses possibilités d'adaptation de l'application en fonction du contexte sont envisagées, telles que le filtrage d'information en sortie (par exemple l'affichage ou le non affichage d'informations en fonction du lieu ou de l'heure), l'enregistrement automatique de données par exemple les coordonnées géographiques pour le cas d'études de terrains d'écologues ou encore une reconfiguration de l'affichage de l'interface en fonction de la situation de mobilité (Pascoe, Ryan, & Morse, 2000).

La philosophie globale du CAC est donc de diminuer la « compétition » des ressources cognitives entre les tâches en en diminuant au maximum le nombre.

En ce qui concerne la Multimodalité, la philosophie est différente bien que poursuivant les mêmes objectifs de diminution de la charge cognitive et en particulier le poids des ressources attentionnelles liées à l'interaction. La multimodalité offre en effet, la possibilité d'adapter les modalités d'entrées et de sorties aux contraintes contextuelles (Oviatt, S.L., 1999).

Celle-ci permet à l'utilisateur de choisir la modalité mobilisant le moins les ressources physiques et attentionnelles facilitant ainsi un plus haut degré de mobilité et des interruptions moins intrusives ou moins coûteuses en terme de ressources cognitives ou physiques. (Arroyo & Selker, 2000) montrent par exemple les caractéristiques d'interruptions perturbatrices d'une tâche selon qu'elles sont dans une modalité ou une autre (lumière et chaleur) et leur impact sur la performance. Ils élargissent ainsi le champ des recherches sur l'interruption qui ont porté essentiellement sur « quand » et « comment » interrompre un utilisateur en introduisant la notion de modalité d'interruption.

Toujours dans cet esprit, (Oulasvirta, 2005) propose de porter une attention particulière sur l'usage pertinent des ressources physiques et cognitives disponibles, par exemple en déchargeant des tâches sur des ressources non utilisées ou dans des modalités différentes, mais également en fournissant des feedbacks de modalités différentes selon les temps de délais de réponse du système. Ces auteurs proposent de faciliter le monitoring des changements mineurs ou majeurs de l'application, ainsi que son contrôle et la gestion non sanctionnée des pauses et autres interruptions.

### 3. La Mobilité : portabilité et usages des dispositifs d'interaction

#### 3.1. Introduction

Jusqu'à récemment la plupart des études sur le travail mobile se sont essentiellement focalisées sur des aspects technologiques (réseaux mobiles, RPC, systèmes de positionnement par satellite,...) ou sur des problématiques d'évaluation de l'utilisabilité de dispositifs portables tels que les PDAs. Cette double orientation renvoie à une posture de questionnement de la notion de mobilité uniquement fondée sur les conditions logistiques de la mobilité d'un dispositif technique qui doit rester fonctionnel quelque soit l'environnement dans lequel il est utilisé, et sur la compatibilité entre caractéristiques de portabilité et conditions d'utilisation efficace de l'outil (notamment ses caractéristiques physiques).

Par ailleurs, les possibilités ouvertes par l'instrumentation des environnements et l'informatique "pervasive" (bornes intégrées, capteurs,...) permet de reconsidérer la mobilité en tant que modalité d'interaction à part entière (intentionnelle ou non) supportée par de nouveaux modes d'interaction entre terminaux (fixes ou mobiles). Cette nouvelle approche nécessite de réfléchir à la définition de cadres de description et d'espaces de conception qui intègrent la notion de mobilité.

Afin de saisir l'ensemble des points pertinents spécifiques des activités mobiles, il convient d'étendre ce cadre initial de positionnement de la problématique en y intégrant notamment la nature des tâches à réaliser, les implications sociales de la mobilité (organisation collective de la mobilité) et l'articulation avec les aspects contextuels. Ceci doit nous permettre d'élaborer un cadre descriptif pertinent pour l'analyse des situations mobiles qui aille au-delà des définitions de la mobilité généralement proposées, qui intègre de nouveaux modes de communication entre systèmes et qui en évaluent les conséquences sur la gestion multimodale de l'interaction.

#### 3.2. *Éléments de définition - Essai de taxonomie*

##### 3.2.1. Définitions

La difficulté à définir de façon satisfaisante la notion de mobilité a été soulignée à de nombreuses reprises (cf par exemple (Kristoffersen, S. & Ljunberg, 2000) : les définitions proposées sont soit trop précises et excluent de fait des situations, soit trop générales et elles ne permettent donc pas de pointer sur les caractéristiques importantes des activités mobiles.

Une première façon classique de définir la mobilité consiste à se référer à la notion d'*autonomie géographique de l'agent* ; le dispositif mobile est alors vu ici comme un moyen de s'affranchir de contraintes topologiques (Makimoto & Manner, 1997) mais également comme une source potentielle de nouvelles contraintes liées aux caractéristiques physiques du système mobile : zone d'affichage réduite, modalités d'entrée restreintes, capacités de traitement et de communication limitées (Ketola & Røykkee, 2001) (Kristoffersen, S. et al., 1998).

Cette orientation conduit généralement à se centrer sur la relation entre l'individu et le système mobile qu'il utilise (PDA, téléphone portable,...) du point de vue de l'*utilisabilité du système*. Le problème posé par l'intégration d'un contenu informationnel similaire sur des dispositifs d'affichage différents a ainsi donné lieu à un volume relativement important d'études (Nielsen & Sondergaard, 2000). Mais pour important qu'il soit cet aspect n'est pas spécifique des situations de mobilité : il concerne également l'utilisation de dispositifs de taille réduite dans des situations d'usage statique (Pascoe, Ryan, & Morse, 1998). Pour aller plus loin il convient donc de considérer que la conception de systèmes mobiles introduit des problématiques de nature radicalement nouvelle, qui remettent en question ou limitent la portée de principes implicitement mis en œuvre dans la conception de systèmes fixes (Dix & Beale, 1996) et qui nécessitent de considérer des aspects directement liés à la nature des tâches à réaliser et au contexte social de l'interaction.

Dans cet esprit, des extensions des définitions initiales (et quelque peu réductrices) ont été proposées afin de prendre en compte la nature des *interactions mises en œuvre par des agents mobiles* (Kakihara & Sørensen, 2001). L'accent mis sur les interactions plutôt que sur la composante "mobilité géographique" d'un individu s'est traduite par la mise au premier plan de considérations liées à la dimension *collaborative* des activités en mobilité et des conséquences sociales afférentes.

D'autres propositions de définition, tout en reprenant la composante "portabilité" du terminal, intègrent cette composante "interaction" en posant le problème de l'accès non-contraint à différents systèmes dans le cadre d'une activité mobile. Le principe selon lequel l'agent doit pouvoir interagir n'importe où, n'importe quand avec un panel étendu d'artefacts constitue à cet égard un élément important de la définition de la mobilité (Gallis, Kasbo, & Herstad, 2001) et ouvre vers les développements en cours dans le domaine de l'informatique ubiquitaire et pervasive.

### 3.2.2. Catégories analytiques et taxonomies des situations mobiles

L'orientation des taxonomies de la mobilité proposées dans la littérature est soit de type technologique (Dix & Beale, 1996) soit centrée sur les usages. L'accent est dans ce cas mis sur la caractérisation des propriétés de la mobilité en tant qu'élément de détermination d'une activité plutôt que sur les caractéristiques techniques du système mobile utilisé.

A l'intérieur de cette seconde classe on pourra par commodité distinguer entre :

- des tentatives de classification de type "ontologique" (ou intensionnel), dans lesquelles on s'efforce de caractériser la notion de mobilité en référence à un ensemble de propriétés intrinsèques ou en recourant à des catégories analytiques censées couvrir différentes dimensions constitutives de la notion étudiée ;
- des tentatives de classification de nature "typologique" (ou extensionnel), où la mobilité va être définie par la référence à des types principaux (ou catégories), qui vont permettre de classer des instances typiques de ces catégories.

### 3.2.3. Classification "ontologique"

Kakihara & Sørensen (Kakihara & Sørensen, 2001) proposent trois catégories analytiques qui renvoient à trois modes de mobilité : spatial, temporel, contextuel :

- Mobilité spatiale :

Afin de dépasser le niveau d'appréhension de la mobilité en termes d'indépendance géographique des agents (centré sur le déplacement et les moyens qui le rendent possibles), Kakihara & Sørensen suivant une analyse de Urry distinguent trois dimensions constitutives de la notion de mobilité qui vont définir de nouveaux patterns d'interaction entre humains (et pourrait-on ajouter entre humains et non-humains pour reprendre la terminologie de Latour) :

- *La mobilité des objets.* La mobilité des objets est souvent (implicitement) jugée superposable à la mobilité des agents qui les transportent ; aucune différenciation n'est effectuée entre les deux. Or pour les auteurs les objets suivent des parcours parfois beaucoup plus complexes que les agents humains (on pourra prendre comme exemple le dossier médical qui est utilisé en mobilité et en fixe, qui passe entre les mains de plusieurs agents, qui existe sous différentes formes matérielles et qui remplit des fonctions cognitives et sociales diverses ;(Luff & Heath, 1998).
- *La mobilité des symboles.* Cette seconde dimension renvoie à l'ensemble des supports de transmission (télévision, radio,...) qui permettent de diffuser simultanément des informations de nature différente (texte, images, sons) à un grand nombre d'agents distribués sur des surface étendues.
- *La mobilité de l'espace.* L'idée est ici que la création de nouveaux espaces d'interaction virtuels sur le Web a généré une réalité spatiale d'un nouveau type : la mobilité de l'espace. Les frontières géographiques se sont estompées, et l'organisation de l'espace est reconfigurable dynamiquement en fonction de l'émergence ponctuelle ou stabilisée de communautés d'intérêt.

- **Mobilité temporelle :**

L'utilisation des technologies de l'information et de la communication contribuent à modifier l'organisation temporelle des activités. Elles réduisent considérablement les temps de réponse dans les échanges entre agents et favorisent le développement de la polysynchronicité dans les activités : il devient possible (voire nécessaire) de gérer plusieurs tâches de façon simultanée. Ceci est d'autant plus vrai avec les systèmes mobiles qui permettent de gérer des activités et d'initier des interactions synchrones ou asynchrones quel que soit le moment de la journée.

- **Mobilité contextuelle**

Depuis les travaux de Suchman, la communauté HCI a intégré le rôle déterminant du contexte dans toute forme de situation de communication. De ce point de vue les descripteurs contextuels sont aussi importants que les descripteurs spatiaux ou temporels (qui peuvent d'ailleurs être considérés comme des éléments constitutifs du contexte). La mobilité introduit donc une possibilité d'extension des déterminants contextuels de l'activité ; en retour, et suivant en cela les travaux menés par Heath et ses collègues sur l'activité de visite de musées (vom Lehn, Heath, & Hindmarsh, 2001) on notera que le contrôle de son déplacement par l'acteur amène ce dernier à construire activement le contexte dans lequel il évolue.

De même que les dispositifs mobiles sont sensés relâcher les contraintes spatiales qui entravent les activités des agents, ils sont également supposés les libérer de certaines formes de contraintes contextuelles de nature essentiellement sociale. Ainsi certains aspects potentiellement limitatifs de la communication en face-à-face se trouvent-ils dissouts dans la désinhibition sociale relative inhérente aux situations de communication médiées.

Dans une veine plus technologique<sup>1</sup>, (Dix et al., 2000) proposent une taxonomie de la mobilité organisée en trois dimensions : le niveau de mobilité, la relation entre le système et son environnement (incluant d'autres dispositifs) et l'orientation (individuelle ou collective) du support fourni par le dispositif mobile (Tableau 1). Cette taxonomie constitue un élément d'un cadre conceptuel des systèmes mobiles qui repose sur la prise en compte des quatre composantes suivantes :

- *La localisation dans l'espace.* La localisation du système utilisé par un agent constitue une information susceptible de modifier le fonctionnement du système. Elle doit donc être prise en compte au cours du développement. La localisation pourra être envisagée, selon les besoins, sous l'angle d'un espace de coordonnées cartésiennes (position absolue) ou sous l'angle d'un espace de représentation topologique (position relative) dans lequel la localisation physique exacte du système n'est pas indispensable et pourra être avantageusement remplacée par une localisation relative (par rapport à la position d'autres objets ou capteurs par exemple). Cette distinction est considérée dans le contexte d'espaces réel et virtuel.

---

<sup>1</sup> La référence est ici le système et non pas l'utilisateur.

## Positionnement du problème et état de l'art

- *La mobilité dans l'espace.* Le système peut être qualifié de mobile en vertu de propriétés différentes : il peut être transporté par un utilisateur (PDA,...) ; il peut se mouvoir par lui-même (robot autonome) ; il peut-être inséré dans un autre objet mobile (ordinateur de bord). Cette insertion peut aller jusqu'à l'encapsulation et la distribution dans l'environnement (systèmes "pervasifs").
- *La nature des agents qui occupent cet espace.* L'espace (réel ou virtuel) dans lequel se meut un dispositif mobile est la plupart du temps peuplé par des systèmes qui peuvent être des agents humains, d'autres dispositifs mobiles, des objets,... avec lesquels le dispositif est susceptible d'interagir, et dont la seule présence peut modifier son comportement.
- *La construction de la "conscience" (awareness) de l'environnement.* Pour avoir un certain niveau de conscience de son environnement le dispositif mobile doit être capable de détecter, de capter et de traiter des variables de description de cet environnement. Ce peuvent être des variables qui décrivent des propriétés physiques de l'environnement et/ou des informations relatives à l'état de systèmes proches (agent humain ou autres dispositifs techniques). Se pose alors le problème de la forme du codage de ces données (cf paragraphe 4 Relation Mobilité-Contexte).

		Individuel	Groupe	Publique
Libre	Fixe	PC de bureau		
	Mobile	PDA	Tour guides	
	Autonome			Robot industriel
Embarqué	Fixe		Réfrigérateur actif	ATM
	Mobile	Vêtements augmentés	Ordinateur de voiture	Chariot intelligent
	Autonome		Pilote automatique	Monorail
"Pervasif"	Fixe		Pièce active	
	Mobile			Star Trek
	Autonome	Agent Web	HAL	Web crawler

**Tableau 1 : Taxonomie de différents niveaux de mobilité (Dix et al., 2000).**

Comme dans toute taxonomie le problème consiste à gérer de façon satisfaisante le nécessaire aller-retour entre catégories analytiques générées par la combinatoire des dimensions descriptives et réalité empirique ; certaines cases demeurent ainsi vides par manque de possibilité d'instanciation raisonnable : par exemple un système individuel, "pervasif" et fixe renverrait pour citer Dix & al. à une cellule active d'ermite difficilement envisageable. Certaines supra catégories, toujours d'après les auteurs, sont par nature problématiques (toute la catégorie "pervasive"-individuelle notamment).

### 3.2.4. Classification "typologique"

(Kristoffersen, S. & Ljunberg, 2000) ont proposé de conceptualiser la notion de mobilité en se référant à trois types principaux :

- La "*Promenade*" (*Wandering*) renvoie à une classe d'activité dans laquelle l'agent se déplace de façon fréquente dans un espace localisé (immeuble par exemple). On trouvera typiquement dans cette classe les tâches de surveillance et de support technique local.
- Le "*Voyage*" (*Travelling*) consiste en un déplacement d'un lieu à un autre en utilisant un moyen de locomotion.
- La "*Visite*" (*Visiting*) décrit l'ensemble des activités pour lesquelles l'agent va être amené à se rendre sur différents lieux où il va passer un certain temps afin de réaliser un ensemble de tâches (commercial rendant visite à des clients, médecin se déplaçant chez ses patients,...).

Mobilité / Technologie	Mobile	Portable	Fixe
"Promenade"	X		
"Voyage"	X	X	
"Visite"	X	X	X

**Tableau 2 : Types de mobilité et technologies associées (d'après (Kristoffersen, S. & Ljunberg, 2000)).**

A chaque type de situation correspond un ensemble de possibilités ou d'impossibilités de mise en œuvre d'un type de technologie (mobile, portable, fixe).

(Marchand, Salembier, & Zouinar, 2000) distinguent pour leur part trois catégories d'activité en situation de mobilité qui s'ordonnent sur un axe intégrant plusieurs descripteurs : définition a priori de l'espace de déplacement, niveau de mobilité de l'agent, niveau de mobilité des outils.

- *Les situations de nomadisme de niveau 1* : l'agent se déplace dans un espace prédéfini en utilisant des outils (notamment des supports à l'interaction) fixes. La situation prototypique est le déplacement d'un espace de travail à un autre (bureau) ; l'agent utilise les outils disponibles en accédant si besoin est à des ressources stockées sur d'autres postes de travail via un réseau.
- *Les situations de nomadisme de niveau 2* : l'agent se déplace d'un point à un autre dans un espace non prédéfini et emporte ses propres outils. Le cas classique est celui d'un agent utilisant un ordinateur portable ; il peut se connecter sur un serveur distant pour accéder à des ressources non disponibles sur son portable.
- *Les situations de mobilité "réelle"* : l'agent se déplace dans un espace non prédéfini et utilise en même temps des outils embarqués. Dans ce type de situation l'agent peut donc être amené à gérer simultanément trois types d'activité : la réalisation d'une tâche donnée, la gestion du déplacement, la coordination avec d'autres agents. Les outils utilisés doivent être susceptibles de fournir un support à ces trois activités.

### 3.2.5. Extension des cadres de description

Le propre des cadres de description évoqués ci-dessus est de fournir un ensemble d'axes de description et de classification des activités en regard du concept de mobilité (lequel est défini de façon plus ou moins fine notamment par rapport aux objets qu'il adresse). Le continuum sur lequel s'ordonne les situations part du niveau 0 de mobilité (activité fixe) à la mobilité "réelle" (réalisation d'une tâche requérant l'utilisation d'un outil alors que l'agent se déplace). Outre leur aspect pré-théorique (aucun des travaux cités ne repose sur une "théorie de la mobilité") ces cadres présentent un certain nombre de limitations et négligent souvent de faire apparaître explicitement une dimension importante de la gestion des activités en situation de mobilité, à savoir le niveau de contrôle requis et exercé par l'agent sur son déplacement. Ainsi dans le cas où la mobilité se caractérise par l'utilisation d'un moyen de locomotion, le niveau de contrôle nécessaire sera différent selon que l'agent est conducteur ou simple passager, selon le type de véhicule (train, voiture, avion,...) et également selon le contexte local du déplacement (conduite sur une autoroute déserte ou en ville par exemple).

### 3.2.6. Mobilité et activités collaboratives

Comme nous l'avons mentionné précédemment, les dispositifs mobiles ont parfois vocation à supporter des tâches collaboratives (Travail Coopératif Assisté par Ordinateur) et cela sous des formes variées selon les types de tâches supportées. En effet, la collaboration médiée par ordinateur peut par exemple être synchrone (téléphonie, chat, échange de documents etc.), asynchrone (messagerie, espace de partage de documents ou d'agenda etc.) ou encore revêtir des formes pouvant potentiellement modifier la nature même de la collaboration. Pour cette raison, il nous semble intéressant de dresser un bref descriptif de quelques études concernant ce champ, encore peu étudié dans un contexte de mobilité.

### 3.2.6.1. Positionnement du problème

Au cours des dernières années plusieurs critiques ont été formulées à l'encontre des travaux menés dans le champ du Travail Coopératif Assisté par Ordinateur (CSCW) pour avoir négligé ou purement et simplement ignoré la problématique du support aux activités collaboratives en mobilité (Bellotti & Bly, 1996) ; (Luff & Heath, 1998). Or la nature du travail en mobilité conduit à repenser les questions posées classiquement autour de la coopération et de la coordination dans le CSCW, du fait des modifications dans la nature de l'environnement de travail.

De la même façon, le caractère collaboratif ou coopératif irréductible des activités professionnelles mobiles a été souligné (Kristoffersen, S. & Ljunberg, 2000; Renevier, 2004) et a conduit à reconsidérer, dans le contexte de la mobilité, la validité d'un mode d'appréhension de type HCI uniquement centré sur l'interaction entre l'agent et son dispositif mobile (dans ce cas l'interaction est dite immersive car requérant toute l'attention de l'utilisateur).

Ainsi certains aspects sociaux du travail mobile qui vont au-delà des aspects individuels doivent être considérés : les conditions de création d'un "esprit de groupe", les mécanismes de collaboration mis en œuvre, la résolution des conflits, l'organisation sociotechnique d'un environnement de travail mobile.

### 3.2.6.2. Le traitement de la mobilité dans le cadre d'activités collaboratives

L'importance de la prise en compte de la mobilité dans la conception de collecticiels étant entendue, il reste néanmoins à préciser le champ des questions pertinentes à considérer.

A partir d'une étude menée chez un opérateur téléphonique scandinave<sup>2</sup> Wiberg & Grönlund (2000) ont identifié cinq domaines de recherche potentiels autour du CSCW mobile :

- *L'appartenance* aux groupes mobiles : comment devenir membre d'une communauté professionnelle et au-delà d'une communauté de pratique au sens de Wenger dans un environnement de travail mobile ? Comment initier et maintenir l'interaction dans un environnement mobile ? Comment réaliser l'intégration d'un nouveau membre dans une communauté mobile ? La technologie peut-elle être utile à cet égard ?
- La *coopération* et le partage de connaissance à distance : le travail mobile étant en effet par nature distal, décentralisé, individuel plutôt que stationnaire, centralisé et relatif à des tâches partagées, comment créer des espaces de partage et de confrontation des expériences et comment gérer le partage de connaissances implicites ?
- La coordination et le service clientèle : *comment* assurer la coordination entre agents mobiles ?

---

<sup>2</sup> Dans cette étude les auteurs ont étudié l'activité de techniciens chargés de la maintenance client (réparations de câbles, interférences sur les lignes,...). Les techniciens sont totalement mobiles : toute information nécessaire peut être accessible dans leur voiture et on peut les joindre de la même façon.

- Les conditions de création d'un environnement coopératif de *travail* : comment aller au-delà de la simple coordination d'activités individuelles ?
- Le maintien d'une communauté mobile : *comment* maintenir l'interaction sociale dans un environnement mobile solitaire ? La technologie peut-elle être utile de ce point de vue, notamment en maintenant l'"awareness" entre les membres de la communauté ? La mobilité affecte-t-elle les patterns de communication entre les individus ?

Ces cinq points, comme nous le verrons par la suite, ne couvrent pas l'ensemble des orientations de recherche possible autour de la coopération mobile ; ils ont néanmoins le mérite de permettre un décentrement des aspects purement technologiques tout en pointant sur la question de l'évaluation de l'apport possible d'outils aux processus de constitution des communautés de pratique, de coordination des activités et de partage des connaissances.

### 3.3. *Caractérisation des systèmes "mobiles"*

- Contraintes inhérentes à la portabilité des systèmes :

Comme nous l'avons vu, les situations de mobilité peuvent également introduire des contraintes inhérentes à la portabilité des systèmes, portant autant sur des aspects technologiques que sur les usages individuels ou collaboratifs.

On retrouvera donc dans cette partie l'ensemble des éléments identifiés par les études portant sur le couplage entre un utilisateur et un dispositif portable typique (PDA, téléphone portable, tablette,...) caractérisé principalement par une taille réduite d'affichage, des capacités généralement limitées de traitement et des possibilités de mise en réseau aléatoires. Néanmoins le contexte d'utilisation reste souvent statique et ne diffère fondamentalement pas des situations d'usage d'ordinateurs de type desktop.

Une des limitations classiques des systèmes portables est de ne présenter que des informations statiques (agenda, répertoire d'adresses,...) ; ceci est essentiellement dû à la difficulté de représenter des informations dynamiques sur un écran de taille réduite (Brewster & Murray, 2000). Cette limitation est paradoxale dans la mesure où un grand nombre d'activités requérant la mobilité des utilisateurs pourraient avantageusement bénéficier d'informations remises à jour de façon dynamique. Le caractère portable des dispositifs n'est donc pas exploité comme il le pourrait du point de vue des besoins inhérents aux activités mobiles.

On peut également s'interroger sur les éventuelles conséquences négatives de l'utilisation de dispositifs portables sur des dimensions non anticipées. Ainsi le format d'affichage limité et, paradoxalement, la taille réduite de ce type de système peuvent devenir des obstacles dans le contexte d'activités coopératives ; le style d'utilisation induit par les propriétés physiques de l'artefact est en effet avant tout centré sur l'individu et non sur le collectif. Des informations qui habituellement sont représentées publiquement et donc accessibles à l'ensemble des agents peuvent être condamnées à rester dans la sphère individuelle (Luff & Heath, 1998) et à ne plus supporter par exemple les phénomènes de "mutual awareness".

- Contraintes inhérentes à la mobilité :

Outre l'émergence de nouvelles situations d'interaction, l'utilisation de dispositifs mobiles peut également entraîner des modifications drastiques de certaines activités qui font déjà l'objet d'une instrumentation en situation fixe. Ainsi en est-il par exemple des activités de supervision de processus qui sont réalisées à distance dans des salles de contrôle confinées, où la relation avec l'objet s'effectue uniquement via des supports de représentation symbolique et nécessite donc régulièrement le recours ponctuel à des auxiliaires présents sur le terrain qui ne sont pas forcément contextuellement insérés dans les opérations de résolution en cours (Heath & Luff, 2000). L'utilisation de dispositifs mobiles constitue de ce point de vue une forme de réponse au problème posé par la réappropriation par les agents de la dimension spatiale de leur activité. Mais dans le même temps ce gain en contextualisation peut se traduire par l'apparition de nouvelles contraintes.

Ces contraintes inhérentes aux situations mobiles sont de nature diverse ; si on les considère du point de vue de l'activité mise en œuvre, elles peuvent être de nature individuelle et collective (cf plus haut).

Ainsi dans le cadre de travaux visant à l'instrumentation d'activité de recueil de données (archéologie, éthologie animale) Pascoe & al ont identifié quatre caractéristiques spécifiques des activités mobiles de terrain (fieldwork) (Pascoe et al., 2000):

## Positionnement du problème et état de l'art

- **Configuration dynamique de l'utilisateur** : ce dernier doit être en mesure de réaliser son activité quelles que soient les conditions environnementales, qu'il soit en situation stationnaire ou déambulatoire
- **Capacité d'attention limitée** : l'utilisateur doit être en mesure de pouvoir concentrer son attention sur la réalisation de sa tâche principale ; les ressources mobilisées par l'interaction avec le système doivent donc être limitées au minimum.
- **Interaction rapide** : en fonction des exigences de la situation, l'utilisateur doit pouvoir si nécessaire effectuer des traitements rapides avec le système (acquisition de volumes importants d'information par exemple).
- **Dépendance contextuelle** : en mobilité, les activités doivent souvent être indexées par rapport à des données du contexte (notamment la localisation géographique).

## 4. Relation Mobilité-Contexte

### 4.1. Introduction

Présente dans de nombreux champs disciplinaires, la notion de contexte connaît de multiples définitions plus ou moins précises. Bien qu'il n'existe pas une définition stable et partagée de cette notion, un point commun à une grande partie de ces définitions est l'idée que le contexte se réfère à des relations entre un élément particulier (individu, objet, processus, etc.) et un ensemble d'autres éléments qui sont intrinsèques ou extrinsèques à ce même élément. Ce chapitre contribue à Définir la notion de contexte au regard des conséquences de la mobilité sur l'interaction Homme-Machine ; Recenser les modes et les techniques de capture et de traitement des informations contextuelles ; Poser le problème de l'articulation entre contexte et multimodalité dans le cadre de l'étude.

### 4.2. Eléments de définition

#### 4.2.1. Le contexte en linguistique

En linguistique, le contexte est ainsi traditionnellement défini comme l'environnement linguistique immédiat d'un élément (un phonème dans un mot, un mot dans une phrase, etc.) ou comme l'ensemble des circonstances dans lesquelles a lieu une énonciation écrite ou orale et qui sont nécessaires à (approche normative) ou qui interviennent dans la compréhension de celle-ci. Ces circonstances peuvent être : l'environnement physique et social, l'identité des interlocuteurs, les événements qui ont précédé l'énonciation, leurs activités en cours, etc. Il convient de noter que certains auteurs préfèrent nommer ces circonstances « situation » et réservent ainsi le terme de contexte à l'environnement linguistique, que d'autres préfèrent nommer « cotexte ». Cette dichotomie a généré deux approches concernant le statut du contexte envisagé comme situation. Pour les auteurs concevant la langue comme système appréhendable indépendamment de toute situation, le contexte n'a qu'un effet secondaire et limité sur l'analyse sémantique des unités constituant les systèmes linguistiques. L'autre approche considère le contexte comme un constitutif fondamental de la langue. Une position forte de celle-ci consiste à considérer que le sens de tout acte linguistique dépend toujours des circonstances dans lesquelles il a été accompli, c'est la théorie de l'indexicalité généralisée (au sens où elle ne limite pas ce fait qu'aux indexicaux comme, par exemple, les pronoms de première et de deuxième personne dont la signification est implicite). Par ailleurs, certains auteurs définissent ce contexte comme une liste (potentiellement) identifiable d'éléments qui influencent les processus de production et de compréhension d'actes linguistiques. Une autre position qui est issue du courant ethnométhodologique envisage le contexte comme un « accomplissement » des interlocuteurs, c'est-à-dire un élément qui se construit progressivement au fil de l'interaction, et non identifiable *a priori* (la pertinence d'un élément contextuel dépend alors du point de vue que l'on prend). Produit et modifié à chaque moment par les acteurs qu'ils utilisent comme ressource d'intelligibilité ou compréhension mutuelle, il influence également en retour leurs comportements. À l'inverse de la première, cette approche met ainsi en avant le caractère fondamentalement dynamique et intersubjectif du contexte.

#### 4.2.2. Le contexte en psychologie

En psychologie, la notion de contexte est également définie de différentes façons. Pour certains auteurs, le contexte se réfère aux représentations ou de façon très large aux états mentaux qui jouent un rôle dans ou influence les processus cognitifs (perception, raisonnement, mémoire, compréhension, etc.). Dans le cadre des études de psychologie expérimentale, le contexte renvoie aux conditions dans lesquelles apparaît un stimulus cible, pour s'intéresser aux effets de celle-ci sur son traitement (par exemple, effets d'amorçage).

Une autre approche du contexte qui s'inspire de certaines des définitions présentées ci-dessus apparaît en ergonomie. Le contexte y est généralement défini comme la situation ou l'ensemble des conditions ou encore l'environnement dans lequel s'exerce une activité et qui déterminent ou influencent celle-ci (Hollnagel, 1993); (Leplat, 2001). Ces facteurs peuvent être : l'environnement physique (outils disponibles dans l'environnement, disposition spatiale des individus, etc.), l'environnement social et organisationnel,

etc. Certains auteurs font une distinction entre contexte « externe » et contexte « interne » (Leplat, 2001). Il s'agit d'une opposition entre la « représentation » que se construit un observateur-analyste des conditions de l'activité (c'est le contexte externe) et la représentation qu'en a l'acteur impliqué dans l'activité (c'est le contexte interne).

#### **4.3. Prise en compte du contexte dans les systèmes**

Dans le domaine récent du Context-Aware Computing (CAC), le contexte est généralement défini comme un ensemble d'informations qui portent sur de nombreux éléments ayant potentiellement un intérêt du point de vue la conception de nouvelles fonctionnalités/services informatiques et de l'interaction Homme-Machine. Par exemple, de nombreux auteurs considèrent comme contexte l'ensemble des informations suivantes : la localisation de l'utilisateur à un instant donné, ses caractéristiques sociales (identité, statut, fonction, groupe/communauté d'appartenance, etc.) et personnelles (intérêts, etc.), son état émotionnel et physiologique à ce même instant, les caractéristiques de son environnement physique (objets, espace géographique, lumière et bruit ambiants, et) et social (personnes se trouvant dans son entourage, leur identité, etc.), ou l'heure et ou la date. Ainsi, partant d'une revue de nombreuses définitions proposées dans la littérature du domaine, Dey, Abowd & Salber (Dey, Salber, & Abowd, 2001) envisagent le contexte comme toute information pouvant être utilisée pour caractériser à un instant donné des entités (personnes, lieux, objets) ayant ou pouvant avoir une incidence sur l'interaction entre un utilisateur et un système qui font eux mêmes partie du contexte ; l'information peut concerner le lieu de l'interaction, l'identité et l'état des (groupes) individus et des objets (informatiques ou non) de l'environnement. A partir de cette définition, les auteurs proposent une modélisation de ce qu'ils entendent par contexte. Le modèle se compose de trois types d'entités et de quatre catégories d'information contextuelle. Les entités sont : les lieux (bureaux, rues, chambres, ...), les individus ou groupes d'individus et les choses (objets physiques ou numériques). Les catégories regroupent l'identité, la localisation (position dans l'espace, relations spatiales entre les individus ou les objets, ...), états (activité des individus, leur état physiologique, température d'une pièce, niveau de bruit, états de fichiers informatiques et de ressources computationnelles, ...) et le temps (heure, date, ...). Ces différentes informations peuvent être obtenues directement (localisation géographique via un GPS par exemple) ou via des inférences qui en combinent d'autres (informations contextuelles).

Une autre définition qui recoupe celle de (Dey et al., 2001) est proposée par Moran & Dourish (Moran & Dourish, 2001). Selon ces auteurs, le contexte correspond aux caractéristiques physiques et sociales dans lesquelles se situent des outils informatiques/computationnels.

A côté de ces définitions qui sont plutôt statiques et cherchent une description précise des éléments qui forment le contexte, on trouve d'autres approches qui abordent le contexte du point de vue de ses fonctions. Ainsi, partant de la notion de contexte telle qu'elle est abordée en /linguistique/pragmatique/philosophie du langage, Winograd (Winograd, 2001) circonscrit celle-ci à son rôle dans la communication et plus précisément dans l'interprétation des actes communicatifs. Pour Winograd, un élément constitue un ou du contexte lorsqu'il joue un rôle dans l'interprétation de quelque chose.

« (...) something is context because of the way it is used in interpretation, not due to its inherent properties. (...) Features of the world become context through their use. »

Autrement dit, c'est l'usage que l'on fait de l'information qui lui procure ou non le statut de contexte : x (e.g. un événement ou une caractéristique) est un élément de contexte si l'interprétation de l'action d'une entité (un individu ou un agent artificiel) dépend de x. Partant de cette définition, Winograd propose de faire une distinction entre contexte et « setting ». Le setting se réfère aux caractéristiques de ce que Dey et al. considèrent comme contexte, c'est-à-dire les personnes, les lieux et les choses qui constituent à un moment donné l'environnement de l'utilisateur.

##### **4.3.1. Systèmes de capture des informations contextuelles**

Nous nous intéressons ici à l'acquisition du contexte pour l'élargissement et l'amélioration des applications sur artefact mobile. Deux modes d'acquisition sont envisagés. L'acquisition explicite nécessite l'intervention de l'utilisateur, autrement dit c'est ce dernier qui fournit l'information contextuelle, l'acquisition implicite se fait automatiquement, sur l'initiative du système. Outre ces deux modes

d'acquisition, on distingue deux approches dans les systèmes de capture du contexte. La première repose sur un environnement contrôlé augmenté par différents capteurs et objets communicants. Le système fournit ainsi à l'application mobile des informations contextuelles issues de cet environnement augmenté. La seconde approche repose sur des capteurs intégrés à l'artefact mobile. La première approche circonscrit le plus souvent l'usage de l'application à un espace géographique restreint. Cette contrainte d'usage a été soulevée par exemple dans le projet ParcTab (Want et al., 1995). A l'inverse les systèmes utilisant des capteurs directement intégrés à l'artefact mobile présentent a priori l'avantage de pouvoir être utilisés en mobilité totale. Cependant de tels systèmes proposés par exemple par Schmit dans (Schmidt, Albrecht, Beigl, & Gellersen 1999b) sont soumis aux mêmes contraintes de portabilité que celles évoquées au chapitre précédent.

Par ailleurs comme le montrent les définitions présentées au paragraphe précédent, le contexte ne se réduit pas à un certain nombre de données prélevées sur l'environnement à partir de capteurs (Rey, 2005).

On parle parfois de contexte de bas niveau ou de haut niveau (Chen & Kotz, 2000). Le contexte de bas niveau étant capturé directement par les capteurs tandis que le contexte de haut niveau (contexte social par exemple) dérive des données capturées ou de la fusion de contexte de bas niveau (Salber, Dey, & Abowd, 1999). Par conséquent il est possible d'assimiler la capture du contexte à un processus de reconnaissance du contexte qui est le terme choisi par Schmidt et al (Schmidt, Albrecht et al., 1999a). Un système de capture de contexte incorpore alors trois niveaux de traitement :

- le premier est l'acquisition de données contextuelles,
- le second est l'interprétation de ces données en faisant évoluer leurs représentations.
- le dernier est la fusion des informations (données interprétées) à différents niveaux d'interprétation.

### 4.3.1.1. Acquisition de données contextuelles

Dans les paragraphes suivants nous recensons quelques capteurs en fonction des informations contextuelles qu'ils permettent d'acquérir. Une autre classification serait possible en distinguant les **capteurs physiques** des **capteurs logiques** (Schmidt, Albrecht et al., 1999a). Les capteurs physiques sont des composants électroniques, mécaniques, optiques, etc. qui mesurent des paramètres physiques de l'environnement. Les capteurs logiques sont du type de l'horloge ou de la cellule GPS. Ces paragraphes reprennent pour l'essentiel l'état de l'art de (Bouchart et al., 1997) et (Chen & Kotz, 2000) que nous replaçons dans le cadre de cette thèse.

- Déterminer la *localisation*

En situation de mobilité, la localisation est une donnée qui change à chaque fois que l'utilisateur se déplace. La localisation est souvent utilisée dans le domaine du CAC (Pascoe et al., 2000), prise en compte du contexte dans les applications).

Il est facile de récolter des informations de localisation si l'utilisateur accepte et n'oublie pas de coopérer en fournissant sa localisation au système. Ce mode d'acquisition explicite peut être mis en œuvre par l'utilisation d'un lecteur de carte magnétique ou d'empreintes digitales à l'entrée et à la sortie d'une pièce, par exemple.

La localisation peut aussi être déterminée sans intervention explicite de l'utilisateur, par des technologies comme le GPS (Global Positionnement System) avec une précision de 10 à 20 mètres. La puissance du signal GPS est trop faible pour pénétrer la plupart des bâtiments et donc ce système ne marche pas pour la localisation à l'intérieur. On peut cependant se tourner vers d'autres solutions utilisant par exemple l'adressage IP (Internet Protocol) d'un réseau sans fil.

- Les capteurs *de mouvement*

La capture du mouvement n'est pas une idée neuve. Depuis des années elle est utilisée dans les systèmes de sécurité, le guidage des missiles, l'automobile, l'aviation. Elle repose sur des capteurs soit distribués dans l'environnement comme les capteurs infrarouges, les sonars, les radars, les caméras, soit directement intégrés aux dispositifs concernés comme les accéléromètres, inclinomètres et autres capteurs inertiels. Verplaetse propose dans (Verplaetse, 1996) une revue des technologies de capture du mouvement.

#### 4.3.2. Finalité de la capture d'informations contextuelles

Dans le domaine du CAC, différentes fonctions de l'acquisition de contexte ont été distinguées. Pour Dey et al. (Dey et al., 2001), l'un des objectifs principaux de l'acquisition du contexte est de déterminer par inférence ce que fait ou tente de faire l'utilisateur à un moment donné pour lui proposer des services et l'aider au mieux dans son activité. Ils distinguent trois types de fonction du recueil d'informations contextuelles :

- la présentation de l'information
- l'exécution de services
- et le stockage d'informations.

La première catégorie de fonctions décrit les applications qui utilisent les informations contextuelles pour présenter de l'information ou proposer des actions pertinentes à l'utilisateur (par exemple, présenter un choix d'imprimantes proches à un utilisateur - cf. (Schilit, Adams, & Want, 1994); montrer à l'utilisateur sa position géographique sur une carte - cf. (Abowd et al., 1997).

La deuxième, décrit les applications qui exécutent automatiquement une ou des fonctions (par exemple, délivrer un message ou une note lorsque l'utilisateur se retrouve à un endroit spécifié – cf. (Beigl, 2000), enregistrer automatiquement une réunion informelle – cf. (Brotherton, Abowd, & Truong, 1999).

Enfin, la troisième, décrit les applications qui permettent de récupérer ou d'ajouter de l'information contextuelle à d'autres informations (par exemple, le système de Brotherton et al. permet d'accéder sur le lieu de la réunion, d'en connaître la date et l'heure, les participants ; le système développé par Pascoe, Ryan & Morse, (Pascoe et al., 2000) permet aux zoologues de prendre des notes indexées spatialement et temporellement).

(Schmidt, A. , Van de Velde, & Kortuem, 2000) décrivent trois fonctions :

- l'adaptation au contexte d'usage des modalités d'entrée ou de sortie
- la réduction pour l'utilisateur du besoin de fournir de manière explicite (intentionnellement) des informations au système
- et la détermination des moments opportuns et des modes pertinents d'interruption de l'utilisateur.

#### 4.3.3. Automatisation et Contrôle de l'application

La problématique de context-aware computing, et d'éventuelles actions automatisées du système qu'elle pourrait provoquer, doit s'accompagner d'une réflexion sur l'allocation du contrôle de l'application, ainsi que de l'utilisabilité d'éventuels contrôles automatiques. Une illustration de ce concept a été présentée par (Mäntyjärvi & Seppänen, 2003). Les auteurs utilisent des senseurs continus de contexte afin d'enrichir une représentation floue du contexte qui peut permettre des actions automatiques du système du même ordre que celles présentées précédemment (présentation d'information adaptée au contexte, contrôle du volume sonore de sortie ou de l'intensité lumineuse de l'écran d'un PDA). Ils introduisent la notion de contexte floues, dans la mesure où, dans une situation de mobilité, celui-ci est rarement tranché et figé, mais plutôt dynamique et faisant intervenir un enchevêtrement de situations de natures variées, difficilement interprétables et dont les frontières ne sont généralement pas nettes. Par exemple, il est souvent impossible de définir clairement des zones d'entrée et de sortie d'un contexte, car cela ne reflèterait pas la complexité de la réalité de la situation.

Dans ce cas, des actions automatiques du système pourraient ne pas être appropriées si l'application s'adaptait aux moindres variations du contexte, et cela perturberait l'utilisateur dans la tâche qu'il est en train d'accomplir. Dans leur étude, les résultats montrent que les erreurs d'automatisation sont inacceptables du point de vue des utilisateurs et surtout qu'ils veulent garder le contrôle sur l'application. Ils présentent plusieurs recommandations intéressantes quant à d'éventuelles adaptations automatiques :

1. L'application doit demander confirmation à l'utilisateur avant d'effectuer un changement majeur dans l'application
2. l'application doit fournir une notification avant tout changement mineur de l'application
3. Après modification, l'utilisateur doit pouvoir revenir facilement à la configuration précédente

Ils relèvent également que certaines informations bénéficient en particulier de ce genre d'adaptation, par exemple la présentation ou non présentation d'informations d'horaires en fonction de l'heure actuelle.

Ils précisent également que l'application doit respecter les règles des comportements sociaux.

#### 4.3.4. Multimodalité & Context-aware computing

A notre connaissance, il existe peu d'études portant sur l'évaluation en contexte naturel d'applications multimodales sensibles au contexte. En effet, il existe une littérature relativement conséquente d'évaluations plus particulièrement centrées sur les aspects techniques, mais finalement la mise en œuvre de tels systèmes a encore peu été étudiée en contexte réel.

Le contexte est souvent envisagé du point de vue des contraintes qu'il peut imposer à l'interaction. Par exemple, pour leur système PHD (Portable Help Desk permet à partir de la carte d'un lieu, de localiser des collègues dans ce lieu, de voir automatiquement leurs coordonnées,...), Anhalt et al. (Anhalt et al., 2001) ont développé deux types d'interfaces : une interface sonore et une interface visuelle. L'interface sonore vise à permettre à l'utilisateur de pouvoir interagir avec le système sans avoir à regarder l'interface. Elle a été prévue pour les situations d'usage où la vision est allouée à une autre activité importante pour l'utilisateur (la marche par exemple). (Ciavarella & Paternò, 2003) utilisent le même procédé de guidage sonore dans un musée pour soulager la vision déjà sollicitée entre autres, par les tâches d'observation des œuvres et par l'orientation dans les salles. Dans cet exemple, des événements d'interaction tels que des bips sonores ou des affichages particuliers, sont déclenchés par la situation géographique des utilisateurs. En cela, ils illustrent bien les pistes complémentaires possibles de conception mettant en œuvre à la fois la multimodalité et la sensibilité au contexte.

## 5. Interaction multimodale : caractéristiques et usages des dispositifs

### 5.1. Introduction

Comme cela a été mentionné à plusieurs reprises dans les parties précédentes, les champs de recherche concernant la mobilité, le contexte et la multimodalité présentent de nombreux points de recoupements qui peuvent servir un projet commun. Si la mobilité constitue un nouveau paradigme d'interaction, nous avons pu voir qu'elle présentait potentiellement de nouvelles contraintes pouvant nuire à la qualité de l'interaction si ce nouveau paradigme ne s'accompagnait pas d'un nouveau style d'interaction ou d'une amélioration des styles existants. Entre autre, nous avons pu voir que la notion de « Context Aware Computing » (CAC) permettait d'envisager de soulager la tâche de l'utilisateur par divers mécanismes, avec pour conséquence attendue une diminution des sollicitations de l'attention, en particulier visuelle, libérant ainsi les possibilités d'évolutions dans le monde physique (mobilité, interaction physique ou visuelle avec les objets de l'environnement), mais aussi social (interactions avec les autres personnes). De la même manière que pour le CAC, l'interaction multimodale et multimédia est parfois également envisagée comme solution palliative aux contraintes exercées par la mobilité sur l'interaction. En effet, une interface multimodale, en laissant le choix de la modalité d'entrée à l'utilisateur pourrait permettre une interaction plus adaptée au contexte, et le cas échéant, favorisant la mobilité. Il en est de même pour le choix des modalités et médias de sorties qui pourraient soit être explicitement choisis par l'utilisateur, soit implicitement (automatiquement) via une reconnaissance et une interprétation adaptée du contexte (CAC).

En cela, la multimodalité (et le CAC) constituerait une première possibilité d'allègement du coût de la gestion des ressources attentionnelles particulièrement critique dans des situations de mobilité.

Une deuxième raison fréquemment évoquée pour l'introduction de la multimodalité dans les interactions homme-machine est sa capacité à faciliter la gestion des erreurs, à la fois en mode préventif comme en mode correctif.

Comme l'évoque (Oviatt, S.L. et al., 2000), un avantage de la multimodalité est qu'elle permet de mieux gérer les erreurs des artefacts et de l'interaction. En effet, tout en les prévenant, les interfaces multimodales permettent une correction plus aisée. Cela pour plusieurs raisons :

- **Prévention des erreurs** : L'utilisateur choisit la modalité qu'il juge la moins susceptible de provoquer une erreur (Oviatt, S. & Van Gent, 1996), ceci en fonction de son expérience et des essais ultérieurs qu'il a réalisés. On peut également envisager que le sujet choisira la modalité la plus appropriée en fonction du contexte, qu'il soit physique, social, historique etc...
- **Prévention des erreurs de reconnaissance** : Le langage de l'utilisateur est moins compliqué (langage opératif), ce qui réduit les erreurs de reconnaissance vocale (Oviatt, S. L., De Angeli, & kuhn, 1997).
- **Réparation d'erreur** : Les utilisateurs ont naturellement tendance à changer de modalité après une erreur (Oviatt, S.L., 2000), (Calvet et al., 2001). De même, les sujets ont tendance à modifier leurs expressions lexicales suite à une erreur, ce qui se confirme après des erreurs répétées.
- **Correction élégante des erreurs** : Les sujets ressentent moins de frustration qu'avec un système monomodal. En effet, les sujets ont tendance à préférer le changement de modalité pour corriger les erreurs, ils apprécient la flexibilité que cela apporte.
- **Désambiguïsation des entrées** : La fusion des différentes entrées qu'elles soient redondantes ou complémentaires permet de clarifier des ambiguïtés, ceci rendant le système plus robuste.
- **Coût cognitif de la correction** : Un taux élevé de verbalisations non-fluides a été associé avec un coût élevé de planification de phrases, de taux d'erreurs et de temps d'accomplissement de la tâche, cela peut refléter une charge cognitive élevée non désirable. Ce taux de verbalisations non fluides baisse lors de corrections d'erreur par changement de modalité, ce qui pourrait indiquer qu'il n'y a pas de surcharge cognitive liée au changement de modalité. Egalement (Cohen, P.R. & Oviatt, 1995) ont montré que le gain de temps escompté est, en réalité, perdu à corriger / rattraper les erreurs de reconnaissance de la parole. Qui plus est, l'étude (Mellor, Baber, & Tunley, 1996) précise qu'il faudrait atteindre un taux de reconnaissance de 94% pour égaler les performances obtenues par une entrée manuelle.

L'usage de la multimodalité laisse donc présager un progrès dans la qualité de l'interaction à la fois grâce à la souplesse d'usage qu'elle offre aux utilisateurs dans le choix des modalités les mieux adaptées à la tâche, aux ressources attentionnelles disponibles, et grâce à la robustesse de prévention et de correction des erreurs.

Cependant, le terme de multimodalité couramment utilisé peut prendre plusieurs acceptions selon les disciplines qui l'étudient, et il n'existe pas à notre connaissance de définitions communes à ces disciplines, de plus, à l'intérieure même d'un champ disciplinaire tel que l'étude des interactions homme machine, sa définition est également peu partagée.

Dans cette partie, nous nous attacherons donc à relever ce qu'entendent plusieurs champs disciplinaires (neurosciences, IHM) par multimodalité en essayant de dégager des points pouvant être intéressants pour la conception d'interfaces multimodales.

### **5.2. Multimodalité, neurosciences et physiologie**

Les neurosciences peuvent apporter un éclairage intéressant sur l'intégration sensorielle des sujets, que cela concerne une modalité individuelle, mais également, et c'est le sujet qui nous interroge ici, lorsque cela concerne plusieurs modalités. En effet, notre relation au monde sensoriel est largement de nature multimodale. Même dans les rares cas où un seul sens est sollicité, la perception du stimulus hérite tout de même des expériences vécues par tous les sens au cours de notre vie. Ceci sera explicité ci-dessous.

#### 5.2.1. L'hypothèse centraliste invalidée ?

En introduction de son article, Kinsbourne (Kinsbourne, 2003) évoque les questions persistantes concernant la perception multimodale :

Le cerveau est-il spécifiquement construit pour des fonctions multimodales ? Ce qui semblerait naturel puisque nous vivons dans un monde d'objets multimodaux. Il semblerait logique que le traitement d'informations multimodales ne soit pas un ajout occasionnel, mais plutôt inhérent à l'activité du cerveau depuis le départ.

Il ajoute qu'étonnement, les modèles traditionnels du cerveau impliquent l'inverse, c'est à dire qu'après l'identification des informations unimodales, une étape supplémentaire de fusion des modalités serait nécessaire pour les coordonner et les entraîner. L'information devrait donc converger vers ce que William James a appelé la « cellule pontificale » (« pontifical cell ») qui serait la seule structure du cerveau entièrement informée et donc qualifiée pour prendre les décisions (hypothèse de « l'homuncule »).

Les travaux de (Kinsbourne, 2003) permettent d'illustrer de manière simple qu'il n'existe pas à proprement parler une telle zone d'intégration des informations multimodales. Voici les commentaires de Kinsbourne sur une expérience menée par d'autres chercheurs (dont le sujet était par ailleurs tout autre que celui traité ici) :

Ces chercheurs ont déterminé quelles zones du cerveau étaient activées en réponse à un flash, puis à un clic et enfin à un flash et un clic en même temps.

Les stimuli unimodaux activent chacun les zones respectives du cerveau : le flash active la zone visuelle et le clic la zone acoustique.

Il est intéressant de noter que les deux stimuli simultanés activent leurs deux zones respectives sans toutefois activer une éventuelle nouvelle zone « cross-modale ». Il fait également remarquer qu'aucune zone présentant de telles caractéristiques n'a jamais été observée à travers l'ensemble des recherches en neurosciences.

Il semble donc qu'il n'y ait pas de point central vers lequel convergent toutes les informations monomodales pour être intégrées.

Il propose une version beaucoup plus distribuée de notre expérience au monde :

Notre expérience est toujours constituée de la configuration de tout le réseau neuronal. Il n'y a pas un endroit central où la conscience se produit, mais plutôt chaque partie du cerveau incorpore une conscience potentielle de ce qu'il fait.

En ce qui concerne notre sujet, ses conclusions sont relativement pertinentes :

Les domaines « d'awareness », de perception, d'intention, d'expérience, sont coordonnés de manière transparente (« seamlessly »), chaque sens par son canal. Cette unité n'est pas un collage mais émerge

d'une unité rudimentaire, qui est différenciée pour représenter les détails spécifiques du moment, de l'état « cerveau / cognitif / expérientiel » précédent qui sert alors de background.

Il remarque également que du point de vue visuel, l'identification et la localisation sont faites par deux voies parallèles, comme s'il s'agissait de deux modalités différentes. Ces propriétés pourraient être exploitées pour une présentation correcte de l'information et corroborent plusieurs postulats entre autres de la psychologie écologique (saillance par la localisation, le regroupement etc.).

### 5.2.2. Le sens du mouvement : un sixième sens

(Berthoz, 1997) explique que nous n'avons pas que cinq sens. En plus des capteurs de la vision, de l'audition, du toucher, du goût et de l'olfaction nous avons aussi des capteurs qui détectent le mouvement. Chacun de ces sens à lui seul ne peut pas mesurer le mouvement, c'est la coopération de tous ces sens qui constitue le sixième sens : le sens du mouvement. Le cerveau doit, à partir de ces sens, reconstruire une perception unique et cohérente des relations de notre corps et de l'espace. Il faut cesser de séparer les sens et insister sur le caractère multisensoriel de la perception. Un sens peut venir modifier la perception d'un autre sens.

### 5.2.3. Une perception active plutôt que passive

Pour Berthoz, Le cerveau est un SIMULATEUR d'action qui utilise la MEMOIRE pour PREDIRE les conséquences de l'action. Nous progressons vers la compréhension des mécanismes qui lient perception et action. Pour cela il faut réintégrer le corps sensible dans l'étude de la perception, de la pensée, de l'émotion.

Pour appuyer ses propos, Berthoz cite (Gibson, 67) : « Nous devons concevoir les sens externes d'une façon nouvelle, comme actifs plutôt que passifs, comme systèmes plutôt que comme canaux et comme en interaction plutôt que mutuellement exclusifs. Ce sont des systèmes perceptuels». Les « Affordances » sont les relations qui définissent les invariants utiles à une espèce particulière, les « faisabilité », mesures de la possibilité d'exécuter des actions compte tenu de son répertoire sensori-moteur.

Il cite également (Merleau-Ponty, 64) : Le visible et l'invisible : « Il faut nous habituer à penser que tout visible est taillé dans le tangible, tout être tactile promis, en quelque manière à la visibilité, et qu'il y a empiètement, enjambement, non seulement entre le touché et le touchant, mais aussi entre le tangible et le visible qui est incrusté en lui comme, inversement, lui-même n'est pas un néant de visibilité, n'est pas sans existence visuelle. »

La perception même non physique est donc ancrée dans un monde tangible, en évaluation permanente d'actions potentielles en étroite relation.

La perception est d'ailleurs différente dans le cas où le membre est touché passivement où si il y a un mouvement du sujet lui-même, car il y a un caractère actif, intentionnel, il y a prédiction et donc attente d'un effet. Il doit donc y avoir cohérence et unité de la perception.

Pour (Robertsson, Iliev, Palm, & Wide, 2007), la perception passive correspond au processus au cours duquel les sens reçoivent une information sans toutefois produire un feedback (ce qui dans le cas de la perception humaine ne se produit qu'à très bas niveau). Il ne peut y avoir de corrections de but par rapport à un changement de l'environnement. Le sens participe simplement à la représentation (fonction épistémique pure), sans qu'il y ait d'interaction immédiate.

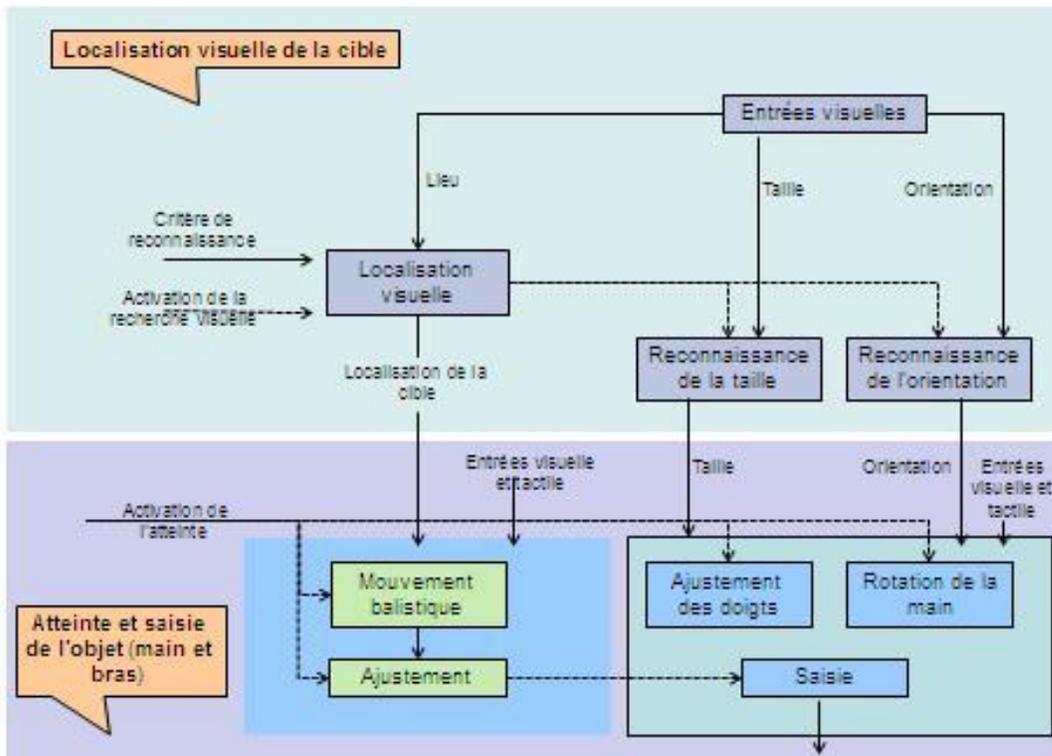
La perception active quant à elle, a la capacité de corriger son but en fonction de l'environnement. La différence entre voir et regarder illustre bien le concept de perception active et passive pour le sens de la vision. L'œil s'adapte pour percevoir un objet. Il ne faut pas confondre la perception active avec un sens actif, qui à la fois transmet et reçoit de l'information (par exemple la fonction épistémique et ergotique d'un geste).

### 5.2.4. Les référentiels absolus et relatifs pour l'action

Les référentiels permettent de comprendre les mécanismes et les repères de l'action en rapport à une tâche donnée.

Pour (Arbib, 1993): « La tâche de la perception n'est pas uniquement de reconstruire un modèle euclidien de l'environnement (référentiel absolu). La perception est orientée vers l'action, combinant des stimuli actuels et une connaissance mémorisée pour déterminer un déroulement de l'action approprié à l'action en cours ».

Ainsi (Jeannerod & Hochmann, 91) en s'inspirant du modèle proposé par Arbib, montre que la main se préforme à la dimension et à l'orientation de l'objet qu'elle va saisir au moment où le mouvement est amorcé, suggérant que le référentiel utilisé par le cerveau est relatif, incluant essentiellement les relations entre les doigts qui vont saisir l'objet.



**Figure 1 : Modèle d'Arbib montrant l'indépendance des segments composant une tâche de saisie. Il convient de noter les processus visuels parallèles traitant des différents aspects de l'objet devant être saisi et manipulé (localisation spatiale, taille, orientation). Ces processus sont corrélés aux contrôleurs du mouvement.**

Arbib ajoute : « Il n'y a pas d'espace absolu unique représenté dans un endroit du cerveau, seulement un couplage des espaces sensoriels et moteurs pour produire un mouvement et achever un but » (voir Figure 1).

Berthoz, pour illustrer ce phénomène de référence d'un mouvement dépendant de la tâche présente deux grands types de repères et en donne les exemples suivants :

- **Repère relatif** : par exemple, il convient de minimiser le mouvement relatif entre la main tenant un livre et les yeux quand on lit en marchant.
- **Repère absolu** : lorsque l'on marche avec un verre d'eau plein, le repère servant de référent est alors la verticale.

Berthoz précise alors : « mais il faut envisager que le cerveau construise une série de référentiels pour chaque phase d'un même mouvement. En effet, pour des tâches plus complexes mettant en œuvre des compositions de mouvements, les référentiels peuvent tout à fait être fort différents ».

La notion de référentiel relatif ou absolu pourrait être particulièrement utile pour l'étude de l'action, dont l'interaction en situation réelle de mobilité, où les tâches s'enchaînent les unes aux autres, mais peuvent également être effectuées en parallèle (en concurrence), mettant en œuvre une succession de référentiels dont la nature et la compatibilité doivent être prises en compte lors de la conception d'interface.

### 5.2.5. Modalité de contrôle

Avant de distinguer une première classification des sens, et des organes associés, de production et de perception de signes, il nous paraît utile d'introduire la notion de modalité de contrôle que peu d'auteurs décrivent en tant que telle.

Elle est parfois mentionnée au sujet, par exemple, du contrôle moteur lors de l'exécution d'un mouvement. Mais ce phénomène est souvent peu détaillé et souvent appliqué aux seuls mouvements des membres principaux (corps, bras et jambes) dans le but de leur bonne coordination dans l'exécution d'une action plus ou moins complexe.

Il nous semble intéressant d'appliquer ce concept à tous les organes de production de signes. En effet, autant il est évident qu'un contrôle proprioceptif intervient pour l'exécution d'un mouvement des membres principaux, autant il est peu mentionné pour le contrôle de l'usage des autres modalités de production de signes.

Pour illustrer ce propos, nous pouvons citer deux exemples :

- **Saisie d'un objet** : Pour l'exécution de la saisie d'un objet, interviennent plusieurs modalités de contrôle en fonction de la phase de l'action. Par exemple, du début du mouvement jusqu'à l'atteinte de l'objet, la **vision** est primordiale pour le contrôle de la visée. Le contrôle **proprioceptif** entre également en jeu, ainsi que le **toucher** au moment du contact. Il s'agit ici d'un exemple classique où les modalités de contrôle que sont la vision et la proprioception, coopèrent à des degrés différents selon les phases du mouvement et le degré de précision requis par la tâche.
- **Prononciation d'une phrase** : Dans un registre moins courant, lors de la prononciation d'une phrase, on « s'entend parler », on peut ainsi contrôler à la fois le contenu (sens de l'énoncé) et la forme (prosodie, volume) de la phrase prononcée, qui sont d'ailleurs indissociables pour leur bonne interprétation. (Nous pourrions pousser plus loin le raisonnement en ajoutant que la proprioception intervient également à ce niveau dans le contrôle des mouvements de la bouche et de la langue, ainsi que dans les vibrations ressenties, mais cela dépasse le cadre de notre étude).

A travers ces deux exemples, il apparaît donc que la modalité de contrôle telle que nous la définissons, consiste en la perception et au contrôle de ce que l'on est en train de faire. Elle est particulièrement sensible au choix de conception de la partie physique d'un dispositif, mais également à la partie logicielle.

En particulier, en ce qui nous concerne pour la suite de l'étude, et comme nous venons de le montrer dans les parties précédentes, il est primordial de noter que le contrôle n'est pas exclusivement moteur et qu'il peut être **inter modal**, **multimodal**, voire **intramodal**.

Pour préciser d'avantage, nous pouvons ajouter que :

- Le contrôle pourrait être **intermodal** lorsque la modalité de production du signe est différente de la modalité de contrôle (par exemple la parole et l'audition lors de la prononciation d'un énoncé).
- De même, le contrôle pourrait être **multimodal** lorsque plusieurs modalités coopèrent à la bonne exécution de la production d'action (exemple de la vision et de la proprioception lors de l'atteinte et de la saisie d'un objet).
- Et dans le cas assez particulier de la production d'un geste moteur « à l'aveugle », le contrôle pourrait être **intramodal** puisque faisant intervenir quasiment les mêmes organes (ou capteurs et effecteurs) dans une boucle perception-action couplée.

## **La notion de modalité de contrôle prendra toute son importance lorsque nous aborderons la problématique couplée de l'interaction et de la mobilité.**

De plus, au delà de la tâche d'identification de la nature qualitative de la modalité de contrôle (sonore, motrice, visuelle etc.) il serait intéressant, dans la mesure du possible de considérer le coût associé de la modalité de contrôle, en fonction du choix de la modalité d'entrée. Ceci pourrait être réalisé par exemple de manière relative à un autre choix de modalité pour en évaluer les impacts relatifs sur l'interaction. En effet, (Woodworth, 1899) explique :

« C'est la précision et la justesse d'un mouvement qui détermine l'intention (la conscience) et l'attention requises pour sa réalisation (ressources attentionnelles). Alors que peu de mouvements relèvent uniquement de la force brute peu coordonnée, dans la plupart des cas, il doit y avoir un degré considérable de contrôle et d'adaptation pour une finalité donnée ».

Il nous semble donc primordial d'évaluer la qualité et la quantité des modalités de contrôle mises en œuvre pour une interaction mobile et multimodale.

Il est intéressant de noter que le couplage de la notion de modalité de contrôle et de repères relatifs et absolus constitue un cadre qui peut s'avérer très pertinent pour l'analyse des situations d'interaction en mobilité. En effet, selon les choix techniques d'implémentation des modalités, les repères mis en œuvre pour l'interaction et les modalités de contrôle associées vont varier de manière corrélée.

En guise d'illustration de ces propos voici deux exemples :

- **Repère relatif** : Pour aller cliquer une icône sur un PDA, le sujet devra dans un premier temps placer et maintenir le PDA à une distance relative à la position des yeux permettant l'interaction. Pour cela, il devra mettre en œuvre au moins deux modalités de contrôles : proprioceptive et visuelle. Les repères mis en œuvre sont donc relatifs, et les modalités de contrôles requièrent dans un premier temps un faible degré de précision et de ressources attentionnelles. Puis pour la visée, et l'atteinte de l'icône, le repère relatif se précise puisqu'il s'agit maintenant de se focaliser sur la position de l'icône par rapport à la position du stylet, ce qui requiert un contrôle moteur et une attention visuelle soutenue.
- **Repère absolu** : Pour certaines interactions tangibles (l'artefact constitue en soi le moyen de contrôle de l'interaction), il s'agit par exemple d'incliner le PDA en avant ou en arrière du plan vertical pour faire scroller la page d'un message (Calvet, 2001). Dans ce cas, le repère est donc absolu et fait intervenir principalement des modalités de contrôle proprioceptives (position du PDA par rapport à la verticale) et peu visuelles. En effet, la vision en tant que modalité de contrôle est peu mobilisée, elle le sera peut-être davantage en tant que modalité d'interaction pour vérifier l'éventuelle effet de la commande (il s'agit donc ici d'une modalité de feedback et non pas d'une modalité de contrôle).

### **5.3. Multimodalité dans les IHM**

Les nombreuses études menées sur la composante multimodale de la communication Homme-Homme (voir par exemple (Kendon, 81, 82)) ont notamment conduit à mettre en relief leur interdépendance avec un processus de contextualisation. La transposition de cette problématique vers le domaine de l'interaction Homme-machine questionne les capacités des systèmes à élaborer une hypothèse sur le contexte qu'ils partagent avec l'utilisateur et sur les moyens dont ils disposent pour valider cette hypothèse. L'ensemble des informations que peut percevoir un système doit être mis en relation avec le sens qu'on peut leur donner d'un point de vue utilisateur. De fait, il est nécessaire de donner du sens à des congruences d'indices observables par la machine, de hiérarchiser la pertinence de ces indices lorsqu'ils sont réunis, et de créer des hypothèses à partir de ces indices. Le travail de modélisation en amont apparaît important. Dans cette veine, (Horchani, 2007) propose d'appréhender les stratégies de dialogue mises en œuvre dans l'interaction et ainsi d'améliorer la qualité des informations multimodales présentées en sortie par le système.

Une autre dimension clé des interactions homme-homme réside dans l'historique de la co-expérience d'interaction entre interactants. On peut à ce propos se poser la question des capacités de la machine à apprendre à reconnaître des patterns caractéristiques d'activité de l'utilisateur.

Dans cette partie, nous présentons la problématique de la multimodalité appliquée aux interactions homme-machine en présentant les définitions des termes et concepts principaux, quelques espaces de conceptions, des résultats et sur les domaines et les usages ainsi que les facteurs identifiés comme influençant les choix de modalités.

### 5.3.1. Définitions

On peut distinguer dans le domaine de l'interaction homme-machine différentes approches complémentaires (Nigay & Coutaz, 1996) :

- une approche centrée sur la technique,
- une approche centrée sur "l'humain"
- et un point de vue hybride qui concilie les deux points de vue.

Au sens large, une modalité est assimilée à une technique d'interaction, une manière d'interagir avec le système. Les ergonomes et les psychologues associent la modalité aux capacités perceptuelles du sujet humain mais aussi à ses facultés de compréhension.

Certains auteurs parlent de média pour traduire le niveau perceptuel et de modalité pour le niveau représentationnel. (Anastopoulou, Baber, & Sharples, 2001) mentionne les concepts communs et des différences relatifs aux systèmes multimédia et multimodaux. Les paragraphes ci-après visent à recenser les différentes définitions des notions de média et de modalité selon les points de vue.

#### 5.3.1.1. Média

- Une approche centrée sur la technique

Du point de vue technique, un média est un dispositif physique, un support d'information qui acquiert ou présente de l'information : ainsi, un clavier est un capteur (acquiert de l'information) ; un écran est un effecteur (présente de l'information). (Maybury & Lee, 1995) considère le média comme l'objet au travers duquel la communication se produit, que ce soit un support physique ou logique.

D'autres auteurs (Blattner, M. , Sumikawa, & Greenberg, 1989), (Frohlich, 1991) affinent cette définition en considérant un média comme toute forme logique (liée à la technique) véhiculant ou pouvant véhiculer de l'information. Cette distinction reflète une différence de niveaux d'abstraction.

- Une approche centrée sur "l'humain"

Le second point de vue centré sur l'humain (Bernsen, N. O., 1995) occulte la composante technique du média au profit de son lien avec les qualités perceptuelles et cognitives humaines. Ainsi, le média est considéré comme un couple  $\langle q, d \rangle$  où  $q$  représente un ensemble de qualités perceptives et  $d$ , les dispositifs sensoriels humain capables de percevoir  $q$ . Ce point de vue a donné lieu à la théorie des modalités que nous approfondirons plus en détails (5.3.1.2.1 Le modèle de Bernsen). Dans ce cadre, un média est associé aux qualités perceptuelles de l'être humain (le visuel pour la vision, le son pour l'audition, etc.).

Enfin un point du vue hybride tente de concilier le concept de média sous les angles de la perception et en tant que support matériel. Dans ce cadre, un média est un dispositif capable de véhiculer de l'information composée d'une "*certaine substance*" sous une "*certaine forme*" (Krus, 1995). Ainsi, si nous prenons comme information, une chanson, le son instancie la substance (matériau brut) tandis que la chanson proprement dite instancie la forme. Le média est ici fortement associé à la substance véhiculée mais son contenu sémantique échappe au média.

Finalement, la définition de média en tant que dispositif support de l'information est affinée suivant les finalités suivantes : matérielle (capteur/effecteur), technique (procédé) ou humaine. Nous considérons le

média comme un dispositif physique sur lequel plusieurs modalités peuvent coopérer. Ainsi, les modalités sonores verbales et non-verbales peuvent coopérer sur le média "haut-parleur".

### 5.3.1.2. Modalité

Pour (Maybury & Lee, 1995), la modalité est centrée sur l'agent de communication et réfère à l'environnement d'entrées/sorties par rapport aux processus perceptuels.

Selon (Bellik, 1995) —point de vue centré sur l'humain—, une modalité est définie par la structure des informations échangées telle qu'elle est perçue par l'utilisateur. Ainsi, la modalité est déterminée non pas, par la structure utilisée par la machine pour représenter l'information mais par la structure perçue par l'être humain. Sur cette définition, la présentation d'un texte sous forme de caractères ASCII ou sous forme de pixels constitue une seule modalité (Krus, 1995). A contrario, la présentation de données sous forme d'un tableau ou sous forme d'histogramme peut constituer deux modalités distinctes. (Bernsen, N. O., 1995) caractérise la modalité comme un système représentationnel de l'information (un texte, un tableau, ...). Enfin (Martin, J. -C., 1995) évoque la modalité comme un processus d'analyse opérant sur des ensembles de données d'entrées-sorties.

Au-delà des diverses tentatives définitoires la nécessité de construire des taxonomies des modalités, que ce soit en entrée ou en sortie, a été soulignée (Martin, J. -C., 1995). Dans cet esprit deux modèles ont plus particulièrement été proposés : le modèle de Bernsen (5.3.1.2.1, Le modèle de Bernsen) centré utilisateur et le modèle de Buxton (5.3.1.2.2, Le modèle de Buxton) centré système.

#### 5.3.1.2.1. Le modèle de Bernsen

(Bernsen, N. O., 1995) s'est intéressé au problème de la présentation optimale des informations échangées entre un utilisateur et son système<sup>3</sup>. L'objectif visé est d'aider le concepteur à identifier les modalités à utiliser afin de représenter au mieux les concepts du système. C'est ainsi que Bernsen définit une modalité par un couple <média, profil> qu'il appelle modalité pure :

- le média d'expression : prend une des trois valeurs <graphique, son, toucher> qui sont caractérisées par un ensemble de qualités de perception (visuelles, auditives ou tactiles). Le média prend ici le même sens que pour (Frohlich, 1991) mais se rapproche du mode de (Bellik, 1995).
- Le profil s'instancie parmi les huit propriétés suivantes :
- **Langagier/ Non Langagier** : une modalité est langagière quand elle fonctionne sur des relations entre des composants lexicaux, syntaxiques, pragmatiques pour produire du sens.

---

<sup>3</sup> "... of unimodal output modalities intended to serve the purpose of clear and efficient presentation of information."

- **Analogique/ Non Analogique** : une modalité est analogique quand elle est en relation avec la réalité. Par exemple, un son du monde réel (Gaver, 1988) est analogique en opposition avec les sons générés (Blattner, M. & Dannenberg, 1990), (Blattner, M. et al., 1989) dits non analogiques.
- **Arbitraire/Non arbitraire** : une modalité est arbitraire si elle fonctionne en dehors d'un système de convention. Par exemple, la parole est non arbitraire puisqu'elle repose sur un système sémantique acquis. Le langage gestuel utilisé jadis par les boursiers est également une modalité non arbitraire. La Langue des Signes est également non arbitraire.
- **Statique/Dynamique** : une modalité est statique lorsque la composante temporelle n'intervient pas dans la représentation de la modalité. Par exemple, une présentation d'un mot à l'écran est statique alors qu'un mot qui clignote à l'écran est dynamique.

Ainsi, dans la taxonomie de Bernsen, la modalité parole est définie comme le quintuplet **<son, langagier, non analogique, non arbitraire, dynamique>**.

Le média d'expression permet d'éliminer de nombreuses combinaisons de modalités pures. Par exemple, l'information présentée par le média d'expression sonore va induire la dynamique.

D'autre part, certains critères sont mutuellement exclusifs. Une modalité analogique (qui tire sa représentation de l'information présentée) ne peut pas être arbitraire.

En combinant les média d'expression (3) avec les valeurs possibles de la modalité (2 langagier/non langagier, 2 analogiques/non analogiques, 2 arbitraires/non arbitraires, 2 statiques/dynamiques), la taxonomie de Bernsen définit 48 modalités de base. Néanmoins, les relations inter-critères et les incohérences permettent d'en éliminer 20 (Figure 2) pour obtenir 28 modalités pures.

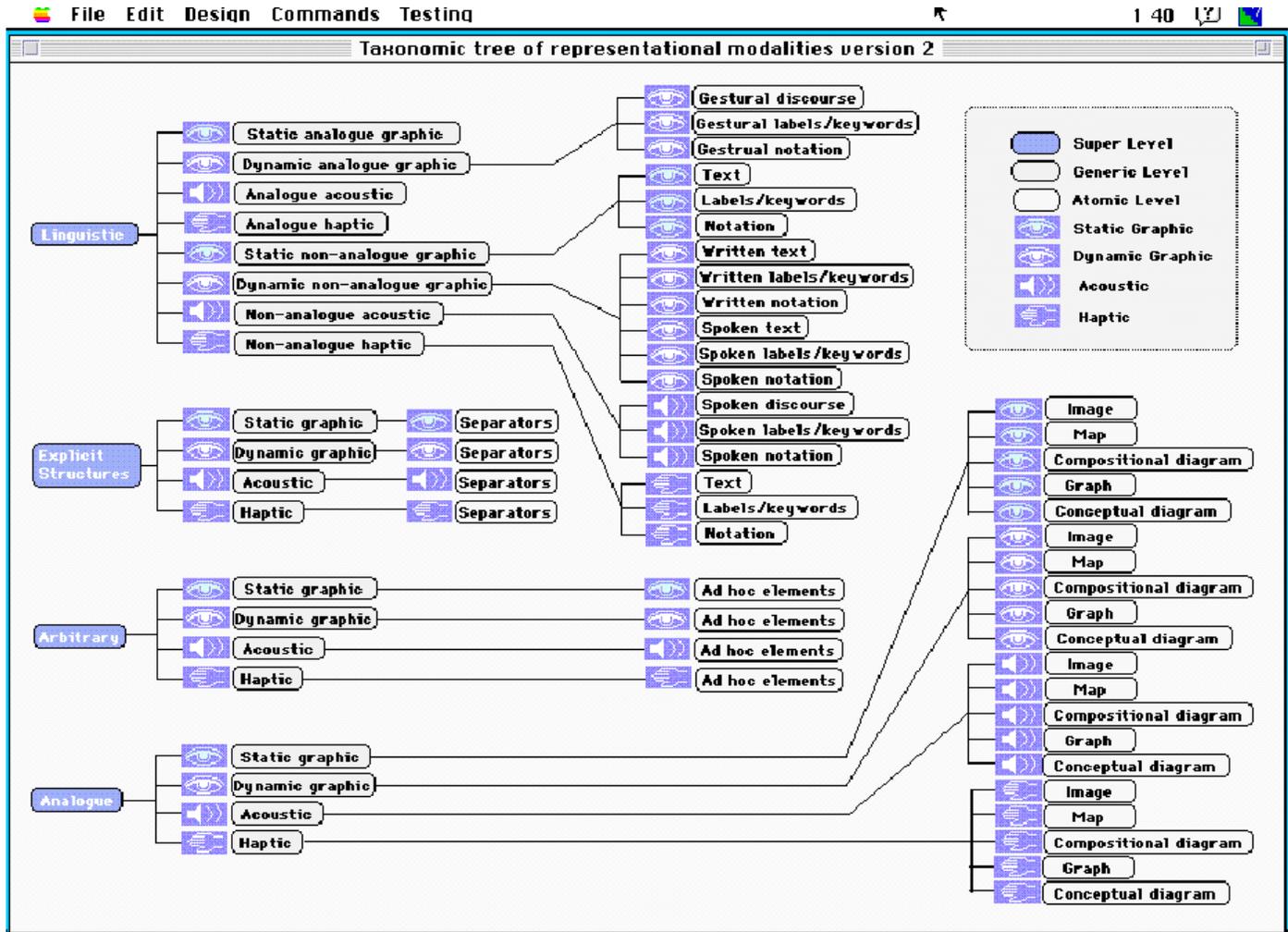


Figure 2 : Recensement des modalités selon (Berssen, N. O., 1995).

Les limitations identifiées de la taxonomie de Berssen concernent essentiellement:

## Positionnement du problème et état de l'art

- la difficulté d'attribuer les valeurs binaires de certains critères comme par exemple, pour le cas de la musique : langagier ou non langagier ? Ce critère est d'ailleurs discuté par (Bellik, 1995) qui considère que ses valeurs devraient être étendues à des valeurs continues (au sens mathématique). "*Certaines modalités telles que la musique sont difficilement positionnables par rapport au critère langagier/non langagier*". Nous rejoignons ici l'argumentaire de Bellik.
- et sur la non complétude des critères. En effet, le chant et la parole lue dans cette taxonomie se caractérisent de manière identique par <sonore, non analogique, non arbitraire, dynamique>. De même, comment différencier la parole chuchotée de la parole créée ?

### 5.3.1.2.2. Le modèle de Buxton

(Buxton, 1993), (Buxton, 1995) propose une catégorisation des modalités d'entrée, essentiellement haptique (Figure 3)<sup>4</sup>. Il adopte un point de vue centré ingénierie des périphériques.

		Number of Dimensions							
		1		2			3		
Property Sensed	Position	Rotary Pot	Sliding Pot	Tablet & Puck	Tablet & Stylus	Light Pen	Isotonic Joystick	3D Joystick	M
					Touch Tablet	Touch Screen			T
	Motion	Continuous Rotary Pot	Treadmill	Mouse			Sprung Joystick Trackball	3D Trackball	M
			Ferinstat				X/Y Pad		T
	Pressure	Torque Sensor					Isometric Joystick		T
		rotary	linear	puck	stylus finger hoiz.	stylus finger vertical	small fixed location	small fixed with twist	

Figure 3: Taxonomie des dispositifs d'entrée (emprunté à (Buxton, 1995)).

<sup>4</sup> Haptique signifie contact physique.

Les dispositifs de la Figure 3, sont regroupés selon deux niveaux de catégorisation. Le premier niveau (traits plein) correspond aux propriétés (position, mouvement et pression) et aux nombres de dimension (1D, 2D et 3D). Le second niveau (traits pointillés) différencie par ses lignes horizontales les dispositifs qui ont un intermédiaire mécanique (stylet par exemple) entre la main et le mécanisme de capture, de ceux qui sont sensibles au toucher. Les sous colonnes regroupent les dispositifs qui nécessitent des contrôles moteurs similaires pour leur manipulation.

### 5.3.2. Espaces de conception

L'objectif de ce type d'approche est de recenser les espaces de conception utiles à l'élaboration des interfaces multimodales. Les premiers espaces de conception (de (Foley & Van Dam, 1982), de Mackinlay (Mackinlay, Card, & Robertson, 1990) de (Buxton, 1993), etc.) dissocient l'interface d'entrée de l'interface de sortie. Cette limitation est forte pour la modélisation de la boucle « perception-action ».

Dans le cadre de ce document, nous nous limiterons principalement à la présentation des espaces de conception centrés sur la multimodalité.

#### 5.3.2.1. Espace de conception de Frohlich

(Frohlich, 1991) a proposé un espace de conception des interfaces incluant les entrées et les sorties. Il globalise les différents types de modalités et prend en compte les sens humains. Frohlich prend comme point de départ le modèle du processeur humain de (Card, Moran, & Newell, 1983). L'interaction, côté machine est le résultat d'un échange d'information entre deux interfaces : celle d'entrée et celle de sortie. Du côté humain, l'information suit le cycle "acquisition-traitement-action". Il en résulte deux espaces de conception : celui des interfaces d'entrée et celui des interfaces de sortie (Tableau 3).

Quatre éléments structurent cet espace : **le mode, le canal, le média et le style**.

- Le **mode** est défini comme "*des états dans lesquels, différentes actions de l'utilisateur peuvent avoir le même effet*". Deux modes sont distingués : le mode action et le mode langage. Dans le mode action, l'interface présente un monde dans lequel l'utilisateur peut agir directement. Dans l'autre mode, les objets de l'interface ne sont plus présentés explicitement et ne sont manipulables que par un intermédiaire. Le mode au sens de Frohlich est centré sur l'utilisateur : il permet de mettre en relief la propriété d'équivalence entre modalités (Tableau 3).
- Le **canal** est défini comme "*une interface qui opère une transformation d'énergie*". Cette définition permet de mettre en relation les capteurs humains et les effecteurs informatiques. A chaque canal correspond un dispositif et inversement au canal "audio" correspond les haut-parleurs.
- Le **média** est ici un système représentationnel qui permet l'échange d'informations. Comme nous pouvons le remarquer sur le Tableau 3, le média au sens de Frohlich est fixé par le couple **<mode, canal>**. Ainsi en sortie, le choix du média utilisé implique le choix du canal utilisé de l'interface.
- Enfin, les **styles** sont découplés en un style langage (déclinés en cinq catégories) produisant des "expressions" et un style action (décliné en trois catégories) produisant des "événements".

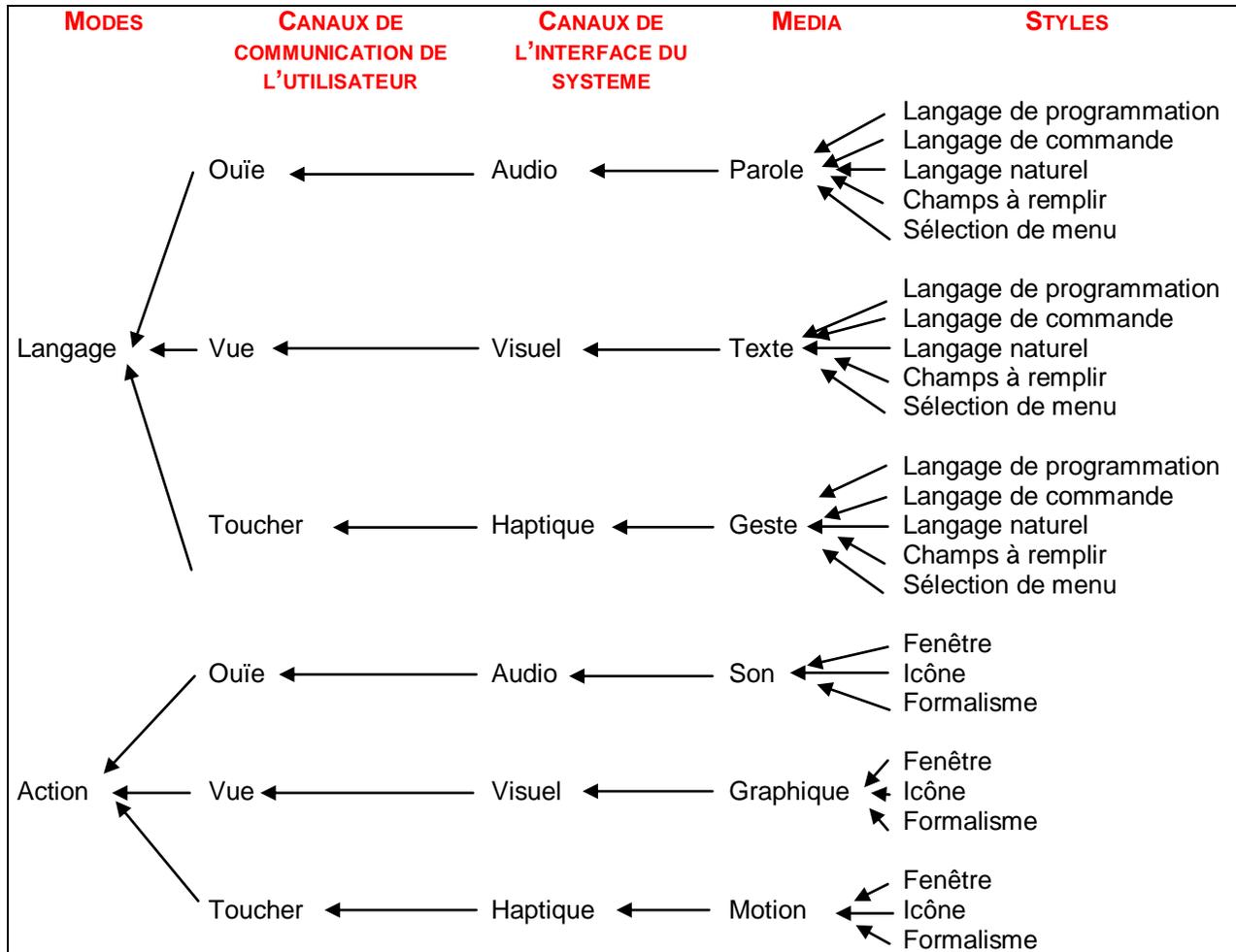
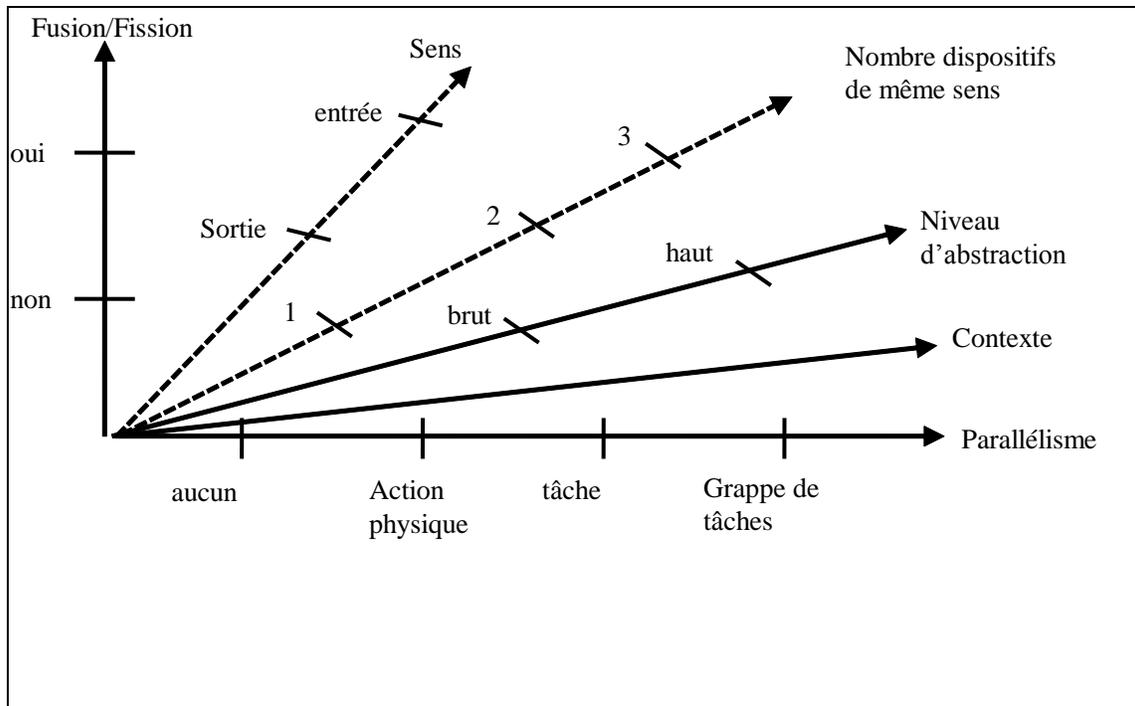


Tableau 3 : Espace de conception des interfaces en sortie de (Frohlich, 1991).

### 5.3.2.2. MSM (Multi Sensori Moteur)

(Nigay & Coutaz, 1993) a proposé un espace de travail de classification des systèmes multimodaux afin d'aider les concepteurs à identifier les implications et les contraintes logicielles induites par le développement d'un système interactif multimodal. Six axes <<n, s>, a, c, f, p> définissent cet espace de conception :

- n : le nombre de canaux du système ;
- s : le sens des canaux utilisés (entrée ou sortie) ;
- a : le niveau d'abstraction qui exprime le degré de transformations subies par les informations reçues ou émises par le système interactif ;
- c : le contexte d'interaction, défini ici comme un ensemble de paramètres d'état utilisés pour contrôler l'interprétation ou la présentation ;
- f : le paramètre de fusion ou de fission d'informations. Ce critère traduit les activités d'interprétation et de présentation ;
- p : l'utilisation en parallèle ou en séquence des médias de l'interface.



**Figure 4 : Le référentiel MSM : un espace de référence pour un système multi-sensori-moteur.**

Les axes n et s caractérisent les dispositifs d'entrée et de sortie. Les niveaux d'abstraction, le *contexte*, le *processus* (*fusion versus fission*), la granularité du parallélisme permettent de caractériser les capacités computationnelles des dispositifs d'entrée/sortie.

Pour (Nigay & Coutaz, 1993), les types de multimodalité sont définies sur cet espace de conception par le couple <présentation, utilisation des modalités>. Ce couple décrit ainsi quatre types de multimodalité CASE (Concurrence, Alternance, Synergie et Exclusivité) intégrant la notion de temporalité (Figure 5).

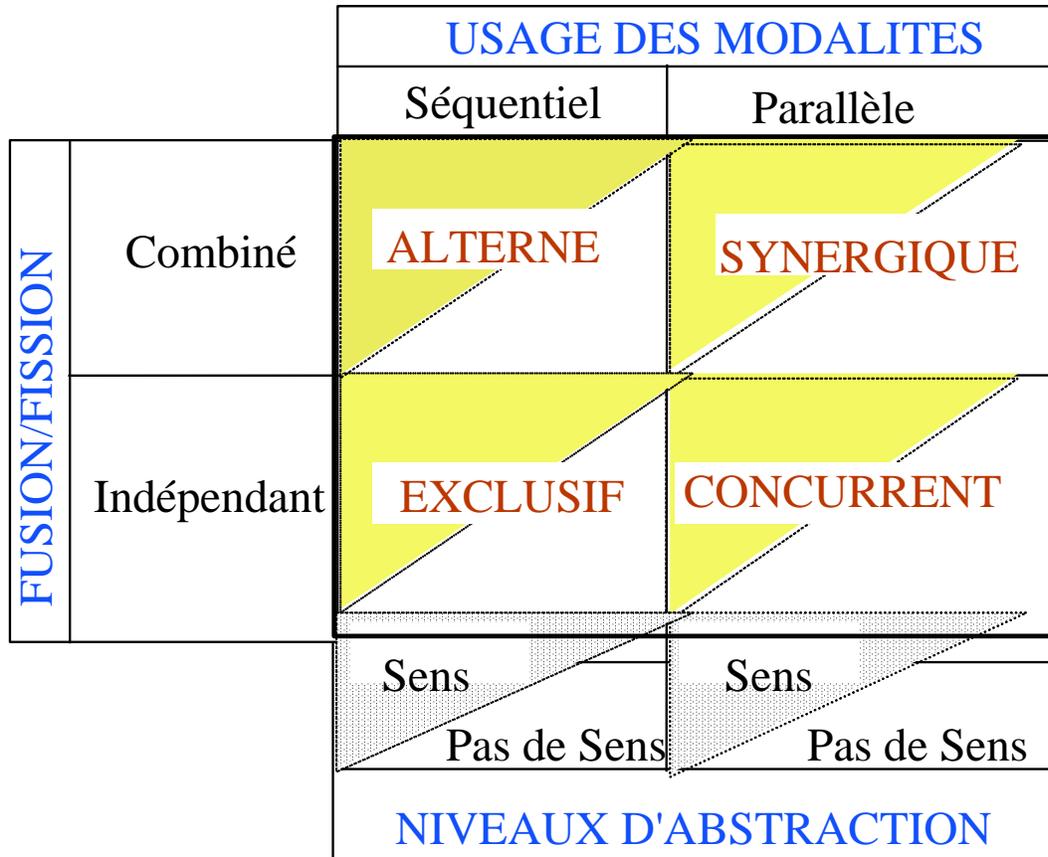


Figure 5 : Espace CASE (Concurrent, Alterné, Synergique, Exclusif).

Cet espace de conception permet ici de rendre explicite les utilisations des différentes modalités du système interactif.

À partir de cet espace, il est possible de spécifier les "types" de multimodalité supportés par le système. Une seconde interprétation de cet espace de conception a été proposée par (Bellik, 1995).

<b>Présentation des messages</b>	<b>Nombre de modalités par message</b>	<b>Utilisation des modalités</b>	<b>Type de multimodalité</b>
Séquentielle	Une	<i>Indépendant</i>	Exclusive
Séquentielle	Plusieurs	<i>Indépendant</i>	Alternée
Séquentielle	Plusieurs	<i>Combiné</i>	Synergique
Parallèle	Une	<i>Indépendant</i>	Parallèle exclusive
Parallèle	Une	<i>Combiné</i>	Parallèle concurrente
Parallèle	Plusieurs	<i>Indépendant</i>	Parallèle alternée
Parallèle	Plusieurs	<i>Combiné</i>	Parallèle synergique

Figure 6 : Les différents types de multimodalité selon (Bellik, 1995).

(Carbonell, Valot, Mignot, & Dauchy, 1994) a notamment analysé l'utilisation des modalités gestuelles et langagières lors de commandes de déplacement d'objets par rapport aux quatre types CASE. Les résultats sont mitigés quand à l'usage synergique des modes par rapport à ce que l'on pouvait en attendre a priori. Cette étude a permis de souligner que l'espace CASE décrit les interfaces multimodales plus en termes de propriétés du système qu'en termes de comportement de l'utilisateur.

### 5.3.2.3. L'espace « CARE » (Complémentarité, Assignation, Redondance et Équivalence) et ses dérivées

Les usages de la multimodalité par les sujets sont souvent regroupés sous le concept CARE — Complémentarité, Assignation, Redondance et Équivalence — (Coutaz & Nigay, 1994). Les auteurs ont formalisé les propriétés CARE en tant qu'éléments d'analyse pour l'utilité des interfaces multimodales d'un point de vue système et utilisateur. Cette classification a permis d'établir un lien entre les propriétés d'une situation (l'état), les buts à atteindre et les relations temporelles entre modalités susceptibles d'être utilisées. Les propriétés (au nombre de quatre) permettent de décrire les relations entre les médias et les modalités ainsi que les relations qui peuvent lier ces modalités.

#### Complémentarité

Plusieurs médias d'entrée ou de sortie sont complémentaires s'ils contribuent ensemble à produire un résultat pour une modalité donnée. De même, plusieurs modalités en sortie sont complémentaires si elles contribuent ensemble à présenter les messages multimodaux.

#### Assignation

Une modalité est assignée à un média de sortie ou d'entrée s'il n'existe aucun autre média pour produire le même résultat. De même, une modalité est assignée à un message si cette modalité est la seule à pouvoir le présenter.

#### Redondance

Les médias de sortie sont redondants lorsque l'utilisation d'un seul d'entre eux suffit à présenter l'information. Par exemple, l'affichage sur un écran et sur une plage tactile d'un texte sans mise en forme est redondant. Plusieurs modalités de sorties sont redondantes si elles expriment le même message de manière séparée. Elles sont redondantes en entrée si elles permettent de produire le même effet au niveau du noyau fonctionnel.

#### Équivalence

Les médias sont dits équivalents si ceux-ci peuvent produire la même expression d'une modalité. Les modalités sont équivalentes si elles présentent de la même façon les résultats d'un message.

Les propriétés CARE autorisent la modélisation de la coopération des modalités, mais qu'advient-il des caractéristiques de dynamicité et de temporalité des modalités ? En effet, ces propriétés ne caractérisent pas la notion de temporalité. (Zanella, M. L. & Caelen, 1996) a aussi proposé dans le cadre de la multimodalité en entrée, d'une part de compléter les propriétés CARE et d'autre part, d'y intégrer les types de multimodalité CASE. Ces propriétés ont essentiellement été étudiées pour des modalités d'entrées (Catinis & Caelen, 1995) et (Zanella, M.-L., 1997). D'autres auteurs ((Truillet, Ph., 1997)), (Truillet, P., 1999) ont transposé ces propriétés CARE en tenant compte des caractéristiques des modalités de substitution —sonores et tactiles— à la modalité visuelle et ont étendu la propriété d'équivalence sous les deux points de vue : informationnel et cognitif (Truillet, P., 1999), (Maurel et al., 2002).

#### 5.3.3. Etudes empiriques de la multimodalité en situation fixe et mobile

Bien que la littérature (en attestent le Tableau 4 et le Tableau 5 ci-après), (Benoit, Martin, Pelachaud, Schomaker, & Suhm, 1998) montre que de nombreuses applications peuvent bénéficier à termes de la multimodalité, il n'en demeure pas moins qu'il n'y a pas de consensus sur la catégorisation des applications multimodales. Ce recensement des applications multimodales est cependant nécessaire. Il permet d'identifier :

- les techniques d'interaction disponibles dans les laboratoires de recherche et en cours d'étude,
- les contextes applicatifs ;
- les dimensions de catégorisation des modalités.

Cette catégorisation est nécessaire par ailleurs pour compléter les dimensions des espaces de conception décrits en 5.3.2 qu'il est utile de considérer pour la conception des interfaces multimodales mobiles.

## Positionnement du problème et état de l'art

(Benoit et al., 1998) ont proposé deux taxonomies des applications multimodales. La première est centrée sur les techniques d'interaction selon l'axe entrée/sortie. Par exemple, pour les techniques d'interaction d'entrée, on trouve les systèmes de reconnaissance de la parole et du geste, les techniques de désignation, le suivi du regard, etc. La seconde est relative aux applications qui supportent des techniques d'interaction. (cartes multimodales, entrée de données, environnement virtuel, applications pour des personnes ayant des besoins spécifiques (handicapés, personnes âgées, etc.).

Compte tenu de notre objet, nous avons dans un premier temps identifié deux sous-ensembles d'applications : celles dites fixes et celles dites en situation de mobilité. Après avoir commenté les tableaux en termes de techniques d'interaction, une première discussion rapporte les effets de la variabilité inter et intra individuelle sur l'usage des modalités ainsi que de leur appropriation. Ensuite, il nous a paru important d'identifier une liste de dimensions ayant des effets sur le choix de l'usage de la multimodalité. Cette liste de paramètres n'est relatée dans aucun état de l'art du domaine à notre connaissance. Les facteurs suivants nous semblent toutefois d'intérêt scientifique et technologique pour l'étude de l'utilisabilité de la multimodalité dans le cadre de cette thèse. Parmi ces paramètres, nous avons retenu la liste suivante comme grille d'analyse et de discussion :

- Le choix d'usage de la multimodalité ;
- L'effet de la contrainte temporelle ;
- L'effet des champs perceptifs ;
- L'effet du contexte.

Enfin, nous discutons de l'apport de la multimodalité pour le traitement des erreurs dans l'interaction afin de lister les contraintes de la mobilité au regard de la conception d'interfaces multimodales.

Positionnement du problème et état de l'art

Groupe	Etude	Application	Dispositif	Modalités (E/S)	Combinaisons modalités (E/S)	Implémentation
Graphique Cartes	Oviatt, S. L., De Angeli, A., & kuhn, K (1997)	sélection immobilière sur une carte	E : stylet, micro S : écran (tablette), hp	E : tactile médié, vocale S : visuelle, auditive	E : Vocale et tactile (c ou r) S : Auditive-visuelle (c ou r)	Simulé WOZ
Graphique autres	Mignot, C., & Carbonell, N. (1996)	Agencement d'appartement	E : stylet, doigt, micro S : écran, hp	E : tactile médié ou non, vocale S : visuelle, auditive	E : Vocale et tactile (c ou r) S : ?	Simulé WOZ
Graphique autres	De Angeli, A., Wolff, F., Lopez, P., & Romary, R. (1999)	Déplacement de groupe d'objets dans des fichiers	E : stylet, micro S : écran, hp	E : tactile médié, vocale S : visuelle, auditive	E : Vocale et tactile (c ou r) S : Auditive-visuelle (c ou r)	Simulé WOZ
Graphique autres	Vergo, J. (1998).	Editeur de texte vocal	E : stylet, micro S : écran	E : tactile médié, vocale S : visuelle	E : Vocale et tactile (c ou r)	Simulé WOZ
Graphique autres	Isokoski, P. (2000)	Entrée de texte par eye track et sélection de cibles	E : eyetrack (casque Infra Rouge) S : écran, hp	E : visuelle (code de mvts des yeux) S : visuelle, auditive (feedback)	E : Pas de combinaison S : R visuelle et auditive	Implémenté
Graphique autres	Schapira, E., & Sharma, R. (2001)	Sélection de cibles	E : caméra, micro S : écran géant	E : gestuelle (sans contact avec l'écran), vocale S : visuelle	E : vocal valide la sélection gestuelle	Implémenté
Formulaire	Nigay, L. (1994) Notebook Matis	Réservation de billet d'avion, prise de notes	E : souris, clavier, micro S : écran	E : tactile (clavier), gestuelle (souris), vocale S : visuelle	E : gestuelle et vocale	Implémenté
Formulaire	Petrelli, D., De Angeli, A., Walter, G., & Guilia, C. (1997)	Remplissage de formulaire pour les étudiants	E : Clavier, souris, micro S : écran, hp	E : gestuelle (souris), vocale S : visuelle, auditive	E : gestuelle et vocale S : auditif et visuelle	Simulé WOZ
Formulaire	Quek, F. K. H. (1996)	Remplissage de formulaire (expérimental)	E : Caméra, clavier S : écran	E : gestuelle (sans contact avec l'écran), tactile (clavier) S : visuelle	Pas de combinaison	Implémenté
Réalité virtuelle	Duncan et al (1999)	Entraînement à la maintenance de Boeing en RV	E : capteurs magnétiques, micro S : écran, hp	E : gestuelle (sans contact avec l'écran), vocale S : visuelle, auditive	E : gestuelle et vocale (c et r) S : auditif et visuelle	Implémenté

**Tableau 4 : Recensement d'études faisant appel à plusieurs modalités en situation fixe.**

Positionnement du problème et état de l'art

Groupe	Etude	Application	Dispositif	Modalités (E/S)	Combinaisons modalités (E/S)	Implémentation
Graphique Cartes	Bernsen, N. O., & Dybkjær, L. (2001)	Assistance à la navigation automobile	Fixé sur tableau de bord E : bouton, micro-casque S : hp ; écran portable	E : vocale, tactile S : visuelle, auditive	S : Auditive-visuelle (R)	Simulé WOZ
Graphique Cartes	Cohen, P. R., Johnston, M., McGee, D., Oviatt, S. L., Clow, J., Smith, I. (1998)	placer des unités militaires sur une carte géo référencée	tablette wacom E : stylet, micro S : hp, tablette wacom	E : tactile médié, vocale S : visuelle, auditive	E : Vocale et tactile (c ou r) S : Auditive-visuelle (c ou r)	Implémenté
Graphique Cartes	Cheyser, A., & Julia, L. (1995)	carte Planification de voyage	PDA E : stylet, micro ou téléphone S : écran, hp	E : tactile médié (manipulation directe, dessin gestuel, écriture manuscrite), vocale S : visuelle, auditive	E : Vocale et tactile (c ou r) S : Auditive-visuelle (c ou r)	Implémenté
Graphique autres	Bers, J., Miller, S., & Makhoul, J. (1998),	Navigation web et commande de pièces	PDA E : stylet, micro S : écran,	E : tactile médié (écriture manuscrite, sélection), vocale S : visuelle,	E : Vocale et tactile (c ou r)	Implémenté
Formulaire	Holzman, T. (1999)	FMA. Formulaires médicaux vocaux sur le terrain	Wearable E : micro S : écouteurs	E : vocale S : auditive	Pas de combinaison	Implémenté
Formulaire	Holzman, T. (1999)	FMC. Formulaires médicaux graphiques sur le terrain	Tablette tactile E : micro, stylet, clavier virtuel S : Ecran	E : tactile médié (clavier virtuel), vocale S : visuelle	Pas de simultanéité ou de fusion (alternatif)	Implémenté
Formulaire	Pascoe, J., Ryan, N., & Morse, D. (2000)	Saisie de données pour les écologues	PDA E : pouce, stylet, boutons S : écran,	E : tactile médiée ou non S : visuelle	Pas de combinaison	Implémenté
Divers	Hinckley, K., Pierce, J., Sinclair, M., & Horvitz, E. (2000).	Utilisation classique d'un PDA, mémo vocal, écran en position portrait ou paysage, arrêt et	PDA E : Accéléromètre linéaire à deux axes, Capteur de toucher, Capteur infrarouge de proximité, stylet? S : écran	E : Embodied, tactile médiée S : visuelle	Pas de combinaison	Implémenté

Positionnement du problème et état de l'art

Groupe	Etude	Application	Dispositif	Modalités (E/S)	Combinaisons modalités (E/S)	Implémentation
		allumage automatique, etc				
Navigation	Harrison, A. B. L., Fishkin, K. P., Gujar, C., & Mochon, R. (1998)	Navigation dans un livre ou document et dans des listes séquentielles	PDA E : stylet, pads de pression (actifs et passifs) S : écran	E : tactile médiée ou non, gestuelle (code), embodied (code) S : visuelle	Pas de combinaison	Implémenté
Balladeur musique	Pirhonen, A., Brewster, S. A., & Holguin, C. (2002)	Commandes d'un balladeur musical	PDA E : doigt, écran tactile S : casque stéréo	E : tactile non médiée (code gestuel) S : auditive (earcons)	Pas de combinaison	Implémenté
Mel	Nim (2001)	Consultation de courrier électronique	PDA E : stylet, caméra, micro S : écran, hp	E : tactile, vocale, gestuelle (code), embodied (code) S : auditive, visuelle	Possibilité de redondance Complémentarité peu probable	Simulé WOZ

**Tableau 5 : Recensement d'études faisant appel à plusieurs modalités en situation mobile (ou portable).**

La synthèse a été scindée selon que les dispositifs et le support permettent ou non une certaine mobilité. En général, c'est le dispositif choisi qui est déterminant. Ainsi, les applications mobiles font toutes appel à des supports de type PDAs ou tablettes graphiques, en particulier lorsque l'utilisateur lui-même peut se déplacer (autodéplacement). Une étude (Bernsen, N. O. & Dybkjær, 2001) traite plutôt de la mobilité de type voyage où le dispositif est fixé au tableau de bord du véhicule.

La colonne indiquant les combinaisons de modalités ne sont pas toujours explicitées clairement dans les études : dans ce cas elle correspond à une interprétation de notre part.

### 5.3.3.1. Quelques domaines d'application

De nombreuses études traitent de l'usage des différentes modalités ainsi que de diverses possibilités de combinaison de ces modalités. Cependant, à la lecture du recensement condensé dans le Tableau 4 et le Tableau 5 qui sont loin de se vouloir exhaustif (le lecteur peut également consulter (Benoit et al., 1998)), nous constatons que les domaines d'applications choisis par les expérimentateurs se recoupent tous plus ou moins, dans des champs assez restreints. Il en est de même pour le choix des modalités mises en œuvre.

Globalement, il en ressort trois types d'applications : les applications graphiques spatiales, les applications de remplissage de formulaire, et les autres applications.

- Applications graphiques spatiales (souvent manipulation directe)

De nombreux auteurs s'accordent sur les avantages de la multimodalité pour les applications graphiques spatiales. Ce domaine constitue la majorité des études recensées dans le Tableau 4 et le Tableau 5. Ceci peut se justifier en partie du fait que la multimodalité est fréquemment et efficacement utilisée dans la communication homme-homme pour identifier des objets dans l'espace physique (Petrelli, De Angeli, Walter, & Guilia, 1997). L'acte de référence multimodal permet de compenser la complexité de l'espace physique perçu par la vision, qui ne peut être exprimé que partiellement via les expressions verbales.

Dans les applications graphiques spatiales, on trouve deux grands groupes, celui des cartes géographiques et celui des plans schématiques et autres représentations d'espaces.

- Cartes géographiques

Ce domaine est privilégié par de nombreuses études (Bernsen, N. O. & Dybkjær, 2001), (Cheyer & Julia, 1995; Oviatt, S. L. et al., 1997). etc. Il s'agit généralement des applications les moins contraintes et les plus complexes, du fait qu'elles sont souvent couplées avec de grandes bases de données donnant lieu à des requêtes croisées. L'utilisateur navigue dans un système profond et large (Card et al., 1983) où ses choix sont multiples et enchaînés. Qu'il s'agisse de requêtes complexes ou de créations d'objets, la modalité vocale confère un avantage indéniable car elle évite par exemple le passage par de longues listes déroulantes pour spécifier une requête (Cohen, P. R. et al., 1998), (voir également 5.3.4.1).

- Autres (plans d'aménagement, dessin de schémas)

(Cohen, P. R. et al., 1997), (Mignot & Carbonell, 1996), (De Angeli, Wolff, Lopez, & Romary, 1999), (Schapira & Sharma, 2001), (Vergo, 1998) (Brison, 1997). Ces applications sont diverses : de l'agencement d'appartement aux commandes de pièces détachées sur un catalogue web. Le style d'interaction est proche de celui des applications précédentes, mais reste néanmoins souvent moins complexe. Ceci est souvent dû à l'affichage simplifié de l'interface graphique, qui restreint les possibilités de requêtes à des contraintes données et d'un manque de représentation mentale qu'a l'utilisateur de l'application.

Dans les deux domaines d'applications présentés ci dessus, le système laisse régulièrement place à la manipulation directe (stylet, souris) qui est complétée/remplacée par un autre mode d'interaction reposant plus sur le principe du langage de commande ou du "faire-faire" (commandes vocales).

Ces applications sont souvent de nature assez complexe et donnent lieu à des interactions variées ainsi qu'à des combinaisons de commandes.

- Remplissage de formulaires

Ces applications (Nigay, 1994; Pascoe et al., 2000; Petrelli et al., 1997) même si elles reposent également sur des interfaces graphiques, se limitent souvent à des remplissages de champs. L'interaction consiste à sélectionner l'item et à satisfaire/remplir le champ de ce dernier.

En ce qui concerne les combinaisons de modalités, il semble qu'il s'agisse majoritairement de la combinaison de la modalité vocale avec une modalité de pointage gestuelle (souris) ou tactile (via le doigt, le stylet, etc.). On constate une évolution du champ d'application qui commence à intégrer la reconnaissance gestuelle, les techniques d'interaction "tangibles", etc.

- Autres applications

(Benoit et al., 1998) propose également le recensement d'applications autour de la réalité virtuelle, l'assistance aux personnes ayant des besoins spécifiques et le travail coopératif.

### 5.3.4. Usage des modalités

Dans la mesure où les modalités étudiées dans notre étude empirique sont parmi les plus utilisées, il convient en premier lieu de décrire leurs caractéristiques et leurs usages. Cette étude est regroupée sous trois grandes familles de modalité que sont les modalités vocales, gestuelles et tactiles présentées ci-après.

#### 5.3.4.1. Modalité vocale

Nous présentons une typologie des situations d'usage des applications interactives où l'utilisation de l'interaction vocale pourrait être avantageuse. Cette typologie concerne les différents artefacts de reconnaissance de la parole : **système de type commande, dictée vocale et parole continue**. Cette catégorisation réactualise et complète celle de (Oviatt, S.L. et al., 2000). Le lecteur pourra se reporter à (Privat, Vigouroux, & Truillet, 2002) pour une discussion de l'utilisabilité de l'interaction vocale en entrée. Nous discuterons deux contextes d'usage de l'interaction vocale : monomodale et multimodale.

##### 5.3.4.1.1. Usage de l'interaction vocale en contexte monomodale

Cette situation correspond à une situation d'usage où seule l'entrée orale est utilisable. Cette situation résulte entre autres de contraintes liées aux caractéristiques du système interactif, de l'environnement ou de l'utilisateur. Nous pouvons identifier les contraintes :

1. **Du système** : Le système ne permet pas d'employer une autre modalité en entrée, ou c'est la seule envisageable. C'est le cas des applications telles que les serveurs vocaux interactifs accessibles uniquement au travers du téléphone tel que ARISE (Pérennou, De Calmès, Lavelle, & Tronel, 1998) ou GALAXY (Goddeau et al., 1994) pour la navigation entre applications multithèmes. On vise ici des « Système de Reconnaissance Automatique de la Parole » (SRAP) de parole continue indépendant du locuteur.
2. **De l'environnement sur l'activité de l'utilisateur** : Cela signifie que l'opérateur a les mains et les yeux occupés pour une autre tâche. L'interaction vocale est alors un moyen pour entrer des données au système (cas des applications de maintenance en aéronautique, application militaire) (Simpson & Hetter, 1985), pour contrôler l'environnement (essentiellement habitacle de voiture et/ou domotique) ou pour énoncer des ordres dans une application comme par exemple dans le domaine de l'ATC (Air Traffic Control) (Truillet, P., 2002). Ces domaines d'application nécessitent des performances élevées de reconnaissance qui conditionnent partiellement les caractéristiques des SRAP : langage de commande ou mot-clé, petit vocabulaire (à faible confusion phonétique entre les mots), dépendance au locuteur, syntaxe a priori figée (bien que la tendance actuelle soit de s'exprimer en langage naturel ouvert pour l'application (Rosenfeld, Olsen, & Rudnicky, 2001).
3. **Des capacités de l'utilisateur** : Lorsque l'utilisateur est handicapé, les SRAP sont dépendants du service à rendre à la personne. Ils peuvent être utilisés :
  - par des personnes handicapées (Damper, 1984), comme modalité d'interaction *a priori* efficace (Christian, Kules, Schneiderman, & A., 2000) pour contrôler l'environnement domotique comme dans le projet HOME, pour aider à la mobilité en guidant les fauteuils roulants, ou pour la navigation dans des documents électroniques tels que IODE (Oriola,

Vigouroux, Pérennou, & Truillet, 1993), PipeBeach (<http://www.pipebeach.com>). Généralement, ces SRAP sont de type commande avec apprentissage par l'utilisateur ;

- pour faciliter la saisie d'écriture, couplé ou non à des systèmes de prédiction pour l'entrée de textes tels que VITIPI (Boissière, 2002), HandiAS (<http://www.li.univ-tours.fr/FRLI/BdChm/Projet2.htm>) et NEMO (<http://www.adaptive-technologies.com/Products/Nemo.html>). Les SRAP utilisés sont des systèmes du type dictée vocale, grand vocabulaire et dépendant du locuteur ;
- par des handicapés sourds voulant communiquer avec des "entendants" au travers du téléphone (Bernstein, 1988).

La modalité vocale en entrée (pseudo langage naturel ou langage naturel) est présente dans toutes les études recensées dans le Tableau 4 et le Tableau 5 sauf une (Pascoe et al., 2000) qui envisage de l'utiliser dans d'autres contextes. Les études ne précisent pas toujours s'il s'agit de « simples » commandes vocales (avec l'utilisation de mots clés faisant appel à un dictionnaire relativement restreint) dont les taux de reconnaissance sont proches des 100%, ou de langage naturel où l'utilisateur peut former des requêtes complexes parfois difficiles à interpréter et présentant donc des taux de succès nettement plus faibles. La quasi totalité fait également usage de la modalité tactile médiée ou non (stylet ou doigt).

D'après (Oviatt, S.L. et al., 2000), la modalité vocale offre la rapidité d'interaction, une large bande passante d'information et une relative facilité d'utilisation. Elle laisse également les yeux et les mains de l'utilisateur libre, ce qui est particulièrement intéressant dans les situations de terrain ou de mobilité. Les utilisateurs tendent à préférer le vocal pour des fonctions comme la description d'objets et d'événements, des groupes et des sous groupes d'objets, des objets hors de vue, des informations conjointes, des états passés et futurs, ainsi que pour passer des commandes pour des actions ou des actions itératives.

Par contre, elle soulève des questions d'ordre social et culturel telles que la confidentialité et la sécurité des données échangées, comme dans un service bancaire où il est inconcevable de donner son code secret oralement.

Elle souffre également des insuffisances des performances des artefacts de reconnaissance : la robustesse de l'interaction vocale, étroitement liée aux performances des SRAP et de la gestion des ruptures de communication, est un critère qui influe la qualité (continuité, coopération, convivialité) du dialogue (Bousquet & Vigouroux, 2001).

Une solution pourrait être l'étude du concept d'Interface Vocale Unifiée (USI (Shriver, Toth, Zhu, Rudnicky, & Rosenfeld, 2001) et (Rosenfeld et al., 2001)), qui repose sur une modélisation de l'interaction autour d'un noyau de fonctionnalités (aide, orientation, correction d'erreurs, etc.) communes à plusieurs applications, pour en faciliter l'appropriation par l'utilisateur.

Enfin, (Cohen, P.R. & Oviatt, 1995) a montré que le gain de temps escompté est, en réalité, perdu à corriger / rattraper les erreurs de reconnaissance. Qui plus est, l'étude de (Mellor et al., 1996) précise qu'il faudrait atteindre un taux de reconnaissance de 94% pour égaler les performances obtenues par une entrée manuelle.

L'interaction vocale peut néanmoins apporter beaucoup, comme moyen de communication palliatif ou couplé à d'autres modalités (apport de la multimodalité pour limiter les effets des erreurs de la reconnaissance, comme dans (Cohen, P. R. et al., 1998). Elle pourrait servir comme outils d'assistance aux personnes handicapées.

Dans le cadre de cette thèse, elle est plus particulièrement étudiée pour les aspects "mains libres" comme cela est décrit dans l'étude de (Bernsen, N. O. & Dybkjær, 2001) explorant l'interaction naturelle en voiture.

Les systèmes interactifs incluant la modalité vocale demeurent encore souvent confrontés au problème de l'activation et de la désactivation du canal audio déclenché par l'utilisateur pour faciliter la reconnaissance

d'énoncés en environnement difficile (contexte bruité, forte charge de travail). Certains activent la reconnaissance vocale quand le stylet touche l'écran (De Angeli et al., 1999), d'autres passent d'un mode commande à un mode dictée vocale à l'aide d'un interrupteur sur le micro (Vergo, 1998) ou plus généralement un bouton poussoir. Ces constatations ont permis d'identifier des critères de recommandations de conceptions d'interaction vocale (Truillet, P., 2002).

### 5.3.4.1.2. Usage de l'interaction vocale en contexte multimodal

Cet usage est également dépendant des contraintes précédemment citées. Cependant, ici, il s'agit également d'un choix de l'utilisateur vis à vis du système :

- L'interaction vocale peut être préférée à une autre modalité de communication, comme pour l'élimination des contraintes liées à l'utilisation de dispositifs technologiques par des personnes non habituées, telles que les personnes âgées. (Rudnicky, 1993) a montré, dans le cas de recherches dans de longues listes avec une seule modalité disponible en même temps, que les utilisateurs préféreraient utiliser la parole au clavier ou à la souris, même avec des performances moindres en terme d'efficacité de la recherche. On parle alors d'assignation au sens du modèle CARE décrit dans les espaces de conception décrits dans la partie 5.3.2.3 de ce document.
- En situation de multimodalité : Les travaux de Oviatt (Oviatt, S.L. et al., 2000) montrent que les performances et l'acceptabilité de certains systèmes pourraient être accrues par l'usage simultané de la parole avec d'autres modalités.

### 5.3.4.1.3. Discussion

La combinaison de la modalité vocale avec d'autres modalités correspond à des actions de communication pour exprimer les intentions de l'utilisateur plutôt que de faire des actions élémentaires (Mignot & Carbonell, 1996), (De Angeli et al., 1999). D'autre part, les verbalisations sont fortement influencées par le contexte visuel. Pour maximiser la pertinence de la communication, les éléments sémantiques sont distribués entre le langage, la vision et les gestes, qui sont complémentaires et fournissent différentes informations.

Les utilisateurs tendent à préférer les entrées vocales pour les informations descriptives.

Les études présentées ici privilégient particulièrement l'usage complémentaire de différentes modalités; en général, la modalité vocale est utilisée combinée à une autre modalité. Comme c'est le cas dans la communication homme-homme (par exemple le geste co-verbal dans la communication homme-homme), la multimodalité en interaction homme-machine permettrait un usage récurrent et plus naturel des déictiques qui se traduit dans ces études par le traditionnel "*put that there*" (Bolt, 1980). On retrouve alors dans l'usage de la complémentarité le couple (action/référent); on assigne généralement la modalité vocale à l'action à mener et la référence à une modalité complémentaire de type gestuelle, tactile, haptique ou autre.

L'utilisation de la modalité vocale implique une plus grande sollicitation de la mémoire à court terme et de la mémoire de travail. (Shneiderman, 2000) a montré que des sujets avaient plus de difficultés à mémoriser des symboles manipulés par des commandes vocales que par le clavier et la souris. Ceci peut donc être une limite à la conception de systèmes comportant une interaction vocale.

Il reste cependant nécessaire de mener des études sur l'utilité et l'utilisabilité de l'interaction vocale en situation de mobilité.

### 5.3.4.2. L'interaction gestuelle

Nous posons en introduction la question de la reconnaissance et de l'interprétation du geste naturel et présentons ensuite quatre classes d'artefacts d'interaction :

- Le geste 2D
- Le geste 3D
- Le suivi du regard
- Et le geste 3D de manipulation de l'artefact (nommée interaction tangible, ou « embodied interface » dans la suite de ce document).

### 5.3.4.2.1. Problématique

La question est ici de s'interroger sur l'adéquation, les limites et les performances des travaux de recherche relatifs à l'introduction du geste naturel dans sa dimension communicationnelle au sens de (Quek, 1996) pour l'interaction H-M.

La communication gestuelle en trois dimensions peut théoriquement être effectuée par toute partie du corps humain. Cependant, l'utilisation d'autres parties du corps que les mains pour communiquer de l'information de manière explicite a encore peu été étudiée. Cette partie s'attache donc particulièrement à l'étude du geste de la main et du bras qui est une des plus abouties à l'heure actuelle. Dans un deuxième temps, nous mentionnerons le mouvement de quelques autres organes qui peuvent être utilisés comme modalité de commande ou de pointage (regard, orientation du visage etc.).

L'observation de l'utilisateur par un système de vision entre dans ce contexte : elle n'impose pas à l'utilisateur un appareillage trop intrusif particulier ce qui facilite une interaction naturelle. Basée sur la reconnaissance de ses expressions, sur l'analyse de ses attitudes posturales et sur l'interprétation de ses gestes et de ses mouvements corporels, elle viserait à lui donner davantage l'impression d'être compris par l'ordinateur, ce qui donne à ce dernier un statut d'interlocuteur et place le dialogue à un niveau pertinent.

Quelques travaux (Baudel & Braffort, 1993), (Braffort, Baudel, & Teil, 1992), (Calvet et al., 2001), se sont intéressés à une définition du geste simple selon leur fonction. De ces travaux sur l'analyse de l'interaction gestuelle, nous pouvons extraire que les gestes émis par l'opérateur humain sont de différentes natures et peuvent jouer différents rôles dans la communication avec la machine que nous résumons ainsi :

- Geste instrumental (faire en montrant, manipuler des objets virtuels).
- Geste de désignation, de pointage.
- Configuration statique référençant une commande isolée.
- Geste de commande, désignant une action et ses paramètres (quoi et comment).
- Geste co-verbal accompagnant une communication parlée.
  - de manière inconsciente : par exemple certaines coordinations regard-mouvement des mains
  - de manière intentionnelle : dire et montrer comment.

Cependant l'usage de la communication gestuelle dans les interactions H-M par la vision demeure encore très limité. Ces restrictions d'usage résultent des capacités encore limitées des systèmes de vision à suivre en temps réel des objets articulés mobiles en environnement naturel ou encore à reconnaître des configurations présentant un grand nombre de degrés de liberté, que ces configurations soient statiques (configurations des mains) ou dynamiques (gestes).

D'autre part, il existe peu de niveaux intermédiaires, dans l'interaction gestuelle intentionnelle, entre les gestes de commandes (composant un code limité) et la langue des signes (constituant une langue complexe à apprendre).

De ce fait, les applications émergentes concernent surtout l'observation des déplacements et des expressions de l'utilisateur ou l'interprétation de gestes simples de désignation et de description.

Une hypothèse de recherche encore peu explorée est la prise en compte de la composante langagière du geste ce qui permettrait de dépasser le stade actuel des applications (Sensiva, <http://www.sensiva.com>), Charade ((Baudel, 1991; Baudel & Braffort, 1993)), DigiStrips (<http://www.tls.cena.fr/division/PII>) et de faire un saut qualitatif dans la communication gestuelle H-M. dans l'élaboration d'un langage (composant un code moins limité).

Il subsiste à ce sujet un grand nombre d'interrogations concernant l'utilisation de geste dans les interactions dont voici quelques exemples :

***Quels sont les indices à observer, dans le geste intentionnel ou dans les gestes co-verbaux ?***

- Est-on capable de les extraire et de les interpréter par un système d'interprétation ?
- A quelles conditions (complexité du lexique (nombre de codes), (ambiguïté), etc.) ?
- Peut-on les exploiter dans l'interaction et à quelles conditions ?

Certains auteurs explorent des modalités telles que les mouvements conscients et inconscients du corps. Les mouvements des mains sont particulièrement observés comme chez (Quek, 1996) qui propose deux types de reconnaissance gestuelle, l'une en 2D et l'autre en 3D.

Cependant, il faut noter qu'une approche constructiviste postule que la communication est un processus qui se construit par et en fonction du contexte qu'elle produit elle-même, mettant en œuvre des mécanismes complexes paraissant difficiles à transférer tels quels vers une interaction homme-machine.

### **Reconnaissance et suivi du geste en 2D et 3D**

#### ○ Geste 2D

L'auteur (Quek, 1996) implémente un suivi dynamique en deux dimensions horizontales de la main pour positionner le curseur à l'écran (FingerMouse), remplaçant ainsi la souris dans des interfaces de type WIMP (Windows, Icons, Menus, Pointing) et complétant l'utilisation d'un clavier. Le système reconnaît la configuration de la main lorsqu'elle est en position de pointage, et mémorise ainsi le référent lorsque le sujet presse une touche du clavier avec la main libre. L'objet est alors mis en surbrillance sur l'interface en temps réel. Ce type d'interaction répond à deux exigences : diminuer le temps de "homing" lié à l'utilisation d'une souris, et libérer de l'espace sur le plan de travail. Il semble que les sujets obtiennent de bons résultats après trois essais. Quek indique également que la vitesse d'interaction est relativement bonne comparée à celle d'une souris.

#### ○ Geste 3D

(Quek, 1996) s'intéresse également au geste communicatif explicite (geste sémiotique de Cadoz) et il se base sur la reconnaissance de gestes codés de type langage des signes, ceci selon trois critères : le coup du geste, la reconnaissance de la pose de la main et la dynamique du mouvement. Ils obtiennent ainsi un taux de reconnaissance de 94% des signes.

(Cassell & Stone, 1999), (Cassell, Nakano, Bickmore, Sidner, & Rich, 2001), quant à eux, s'intéressent au suivi des mouvements du corps en général, pour enrichir la connaissance du système sur le contexte de l'interaction. Il ne s'agit donc plus de commandes actives de l'utilisateur, mais de reconnaissance passive ou de prise d'information par le système afin d'améliorer la représentation par le système du contexte en cours.

#### ○ Geste 3D de manipulation de l'artefact (tangibile, embodied)

Dans des travaux précédents (Calvet, 2001), nous nous sommes intéressés à des types de modalités particulières dites « tangibles » ou « embodied », au sens où il s'agit d'une manipulation de l'artefact dans son ensemble qui constitue la commande. Dans notre étude, nous avons mis en œuvre un certain nombre de gestes sémantiquement signifiants afin d'interagir avec le dispositif, par exemple, une inclinaison du pda permettait de faire défiler du texte, le fait de le secouer effaçait le message sélectionné. Ce type de modalités expérimentales, introduites par (Rekimoto, 1996) et (Hinckley, Pierce, Sinclair, & Horvitz, 2000) sont aujourd'hui exploitées commercialement par des produits grand public tels que la Wii™, ou quelques téléphones portables. Un aspect intéressant de ces modalités, est que pour la plupart des cas d'implémentation, elles ne disposent plus forcément d'une interface d'entrée explicitement identifiée puisque c'est l'artefact lui-même qui la constitue. Lorsque les choix de conception sont pertinents, cela peut produire une interaction intuitive et naturelle et réduisant la courbe d'apprentissage. De plus, s'agissant d'une manipulation physique d'un objet, le lieu d'entrée de la modalité ne nécessite potentiellement que peu de ressources visuelles. Le degré de ressources visuelles nécessaire dépendra essentiellement de la modalité de feedback ou de sortie de l'application.

### **5.3.4.3. Modalité Tactile**

La modalité tactile met en œuvre le toucher de l'utilisateur. C'est une modalité portable. En plus d'être utilisé pour écrire des mots, le stylet sur table traçantes, sur PDA, etc. peut être utilisé pour tracer des symboles et des signes, des gestes, des graphiques simples, du travail artistique et les signatures (Oviatt, S.L. et al., 2000).

Elle peut être utilisée pour pointer, sélectionner des objets visibles, et comme un déclencheur de micro pour les entrées vocales. Dans un environnement public, elle permet une interaction plus privée et socialement acceptable et une alternative fiable à la modalité vocale dans les environnements bruyants.

Dans les domaines de type architecturaux, le stylet permet de faire des croquis. Sur une carte, il transmet également des informations spatiales sur des points précis, des lignes etc. Le stylet a un fort potentiel pour les applications graphiques, spécialement pour des tâches mobiles.

Les utilisateurs tendent à préférer les entrées au stylet pour les numéros, les symboles et les contenus graphiques (Oviatt, S.L. et al., 2000).

Il est à noter qu'on a développé des dispositifs permettant l'usage d'une modalité tactile avec retour de force.

Les exemples cités précédemment, dans leur variété, ont tous en commun l'usage d'un stylet ou du doigt mais au delà, il faut relever que les sémantiques utilisées sont de natures parfois très différentes et consistent donc en autant de modalités différentes, voire parfois pas forcément d'une modalité mais simplement d'une capture de média (par exemple pour le tracé d'un dessin non interprété).

### 5.3.5. Facteurs influant l'usage de la multimodalité

Après avoir caractérisé un ensemble de modalités, nous nous intéressons maintenant à leur mise en œuvre dans des applications, et surtout à leurs usages par les sujets. Ce passage en revue des usages de la multimodalité nous permettra ainsi de dégager des facteurs les influençant.

#### 5.3.5.1. Variations intra et inter-individuelles dans l'usage de la multimodalité

Plusieurs études relatent d'importantes variations intra et inter-individuelles dans l'usage de la multimodalité.

L'existence d'une forte variabilité interindividuelle amène à se poser la question des caractéristiques de l'utilisateur qui jouent un rôle sur le choix des modalités et sur la gestion de la multimodalité (Huls & Boss, 1995), (Mignot, 1995), (Calvet, 2001), (Calvet et al., 2001).

Plusieurs facteurs sont susceptibles de jouer un rôle dans la définition de l'expérience des utilisateurs. On peut citer l'ancienneté de la première expérience et la fréquence d'utilisation, mais aussi la nature des expériences qui peuvent permettre des transferts de procédures de l'utilisation d'un dispositif à un autre et la nature des connaissances déclaratives relatives au fonctionnement du dispositif.

Étant donné son importance potentielle, le niveau d'expérience des utilisateurs doit être considéré lors du processus de conception. Par exemple, certains résultats expérimentaux en matière de navigation sur le web (Derycke, Plénacoste, & Vanderdonck, 1998) invitent les concepteurs à avoir recours à des modalités différentes selon le niveau d'expérience des utilisateurs.

La première phase dans l'apprentissage d'un système multimodal débute selon (Mignot, 1995) et (Calvet, 2001), (Calvet et al., 2001) par une période de test exploratoire des possibilités du système au travers de plates-formes Magicien d'Oz. Durant cette phase, les différentes actions réalisées par l'utilisateur permettent à ce dernier d'identifier les caractéristiques de l'interface, d'appréhender les possibilités multimodales, la sémantique des commandes, et les particularités syntaxiques propres au langage de commandes. Chaque action lui permet de tester ce qui lui est possible de transférer d'une interaction homme-homme (interaction dite 'naturelle') à une interaction homme-machine multimodale (qui est forcément plus contraignante).

A ce niveau la représentation que l'utilisateur se construit des capacités du système peut jouer un rôle: elle va varier selon l'expérience antérieure qu'il a de dispositifs plus ou moins identiques. Il existe donc une grande variabilité dans les modes d'appréhension et d'utilisation de la multimodalité. (Martin, Jean-Claude, Julia, & Cheyer, 1998) notent que la variabilité interindividuelle dans l'utilisation de dispositifs multimodaux est très importante. Plusieurs études citées par ces auteurs semblent, par ailleurs, montrer que pour un même utilisateur la part d'utilisation multimodale progresse au cours des sessions. La variabilité intra-individuelle serait ainsi non négligeable. On peut penser que lorsqu'elle est observée en début de tâche, elle correspond principalement à la phase d'exploration de l'interface évoquée plus haut. Ainsi (Huls & Boss, 1995) soulignent-ils que les sujets changent de modalité d'interaction : la modalité d'interaction multimodale (dispositif de pointage plus langage avec utilisation de déictique) prend le pas, au fil de l'utilisation, sur la modalité de manipulation directe des objets.

On notera que ces résultats ont été obtenus avec des sujets peu familiers du dispositif testé. On peut penser que lorsque les modes d'utilisation du système par les sujets deviennent relativement stabilisés, les changements de modalités peuvent se révéler moins fréquents car éventuellement plus coûteux à mettre en œuvre. S'adapter à une modalité de navigation nouvelle, si adaptée puisse-t-elle être à la situation, n'est pas nécessairement moins coûteux pour un utilisateur qui a recours à des routines efficaces dans une autre modalité. Par exemple, un utilisateur habitué à consulter par la modalité vocale le répondeur téléphonique de son mobile pourrait préférer conserver cette modalité, même en environnement bruyant, plutôt que de naviguer en mode tactile qu'il n'utilise pas d'habitude.

(Petrelli et al., 1997), remarquent que l'usage de la multimodalité est très différent chez les experts (84 % des commandes), et chez les novices (30%). Les novices ont souvent recours à des référents linguistiques peu efficaces. Ils suggèrent de palier ce manque d'expertise par de l'aide contextuelle ou par des réponses directrices.

A partir des résultats d'une étude menée par (Mignot, 1995; Mignot & Carbonell, 1996), ces auteurs ont défini une typologie des styles d'usages de la multimodalité comportant quatre catégories :

- Les sujets qui utilisent préférentiellement une relation de type "manipulation" en début d'expérimentation (ils semblent "commander" le dispositif), puis évoluent vers une relation de type coopérative (ils « co-agissent » avec le dispositif, en utilisant par exemple préférentiellement des termes "nous" que des injonctions à l'infinitif) ;
- Les sujets dont l'usage de la multimodalité illustre la spécialisation des modes en fonction des tâches. Ces sujets utilisent la modalité gestuelle pour certaines tâches (déplacement d'objet) et les modalités gestuelle et orale pour d'autres tâches (introduction d'objet) ;
- Les sujets qui utilisent préférentiellement la modalité orale ;
- Les sujets qui n'ont pas utilisé les modalités de la même manière au cours des sessions, phénomène pour lequel il est difficile de trouver une interprétation satisfaisante.

(Calvet et al., 2001), (Calvet, 2001), viennent confirmer une partie des résultats précédents en ce qui concerne les variations inter et intra-individuelles ainsi que sur l'évolution de l'usage avec l'expertise. Les utilisateurs, après une phase de test exploratoire, tendent à stabiliser, voire à "s'autoassigner" l'usage d'une modalité pour une commande, ceci avec de fortes différences inter-individuelles appelées « spécialisation » par les auteurs. Mais cela, n'empêche pas la rupture de cette stabilité quand le contexte de l'interaction vient à changer. Les auteurs indiquent que l'usage des modalités repose sur une base interne de préférences propres aux utilisateurs ainsi que sur des facteurs plus externes liés au contexte de l'interaction (historique de l'activité, correction d'erreurs, changement de but, etc.).

### 5.3.5.2. Liberté de choix d'usage

La possibilité de choix d'usage d'une modalité ou d'une combinaison augmente le temps d'accomplissement de la tâche d'après une étude menée par (Huls & Boss, 1995). En effet, les auteurs implémentent un classement de fichiers avec la modalité écrite au clavier, la sélection graphique des fichiers, et imposent l'usage individuel d'une modalité, puis de l'autre et de leur combinaison. La tâche multimodale est celle qui prend le plus de temps. Cependant, il s'agit d'une application particulière mettant en œuvre des modalités peut-être inappropriées pour ce type de tâche (la longueur des noms de fichiers est variable et leur saisie au clavier peut être fastidieuse).

Ces résultats interpellent plutôt sur les préférences d'usage des utilisateurs, l'adaptabilité de l'interaction et l'assignation d'une modalité à une tâche. En effet, l'assignation de modalité ne semble pas recommandée, car elle occulte la flexibilité des interfaces multimodales, et empêchent les utilisateurs de choisir pertinemment, selon leurs préférences et selon le contexte d'interaction, la modalité la plus appropriée (Calvet et al., 2001). Cette étude a montré que de nombreux facteurs interviennent pour justifier un changement de modalité, et par conséquent, empêcher ces changements limite la flexibilité du système.

### 5.3.5.3. Contraintes temporelles

La contrainte temporelle semble intervenir sur le style de l'interaction. Une étude (Norman, 1991) a montré que les utilisateurs passent d'un style de manipulation directe au style communication (langage de commandes) sous pression temporelle. Ceci, comme mentionné au point précédent, indique que les utilisateurs ont une préférence individuelle d'interaction, même si elle n'est pas forcément perçue comme efficace, la pression temporelle pousse à passer à un style perçu comme efficace. L'efficacité ne serait donc pas forcément le facteur principal du choix de style d'interaction.

### 5.3.5.4. Effet des champs perceptifs

(Petrelli et al., 1997) (De Angeli et al., 1999) partent de l'idée que d'après la psychologie écologique les cycles perception-action sont médiés par les affordances, informations optiques qui véhiculent des propriétés fonctionnelles des objets (relation dérivée de la rencontre entre les propriétés des objets et le répertoire d'actions physiques disponibles à l'utilisateur). L'approche écologique tente d'étendre le concept d'affordance pour expliquer la variabilité des actions multimodales. Les gestes sont déterminés par la mutualité entre l'information fournie par l'objet auquel on se réfère et la palette de mouvements disponible du sujet. Elle attribue une grande importance à la perception visuelle dans la variabilité des actions en particulier dans l'agencement spatial.

La saillance (caractère cognitivement distinct ou/et effectivement saillant d'un aspect du champ psychologique) est un bon prédicteur de type d'action. Les groupes saillants appellent des accès de groupe (encercler un groupe d'icône), alors que les non saillants, appellent une sélection individuelle (cocher des icônes).

### 5.3.5.5. Contexte expérimental et naturel

La quasi totalité des études sont expérimentales et prennent peu en compte les éléments du contexte et la variabilité qui en découle. Cependant, certains auteurs mentionnent des aspects intéressants à ce sujet :

- (De Angeli et al., 1999) indiquent que le contexte spatial influence les modes d'interaction.
- (Calvet, 2001; Calvet et al., 2001) élaborent une première liste de facteurs contextuels intervenant dans les choix d'utilisation de telle ou telle modalité. Par exemple : le contexte de l'activité en cours va peser sur le choix de la modalité, selon l'intention du sujet et le type de message qu'il est en train de consulter.
- (Pascoe et al., 2000) attribuent une grande importance au contexte de l'interaction et leur étude est une des rares en situation "naturelle". Pour eux, le choix des modalités d'interaction dépendra de l'activité de l'utilisateur et du contexte dans lequel se déroule l'interaction.

Ici se pose à nouveau la question de la définition de ce que l'on entend par contexte, (voir la partie 4).

Après avoir abordé nos problématiques théoriques (Mobilité, contexte et multimodalité), nous présentons maintenant la problématique méthodologique.

## 6. Problématique méthodologique

Le développement des systèmes mobiles (PDA : Personal Data Assistants, téléphones portables, etc.) pose de nouveaux problèmes du point de vue méthodologique et technique pour étudier empiriquement leurs usages et leur utilisabilité dans des environnements naturels. La mobilité d'un utilisateur implique non seulement un changement d'environnement, mais aussi la gestion de son déplacement physique, ainsi que des changements rapides d'éléments du contexte qui rendent l'interaction avec le système plus complexe pour l'utilisateur et compliquent la tâche de recueil de données d'interaction pour l'analyste. Parmi les études empiriques relatives aux usages de systèmes mobiles, il y a une nette dichotomie entre deux types d'approches (les études en laboratoire « laboratory studies » et les études de terrain « field studies »), qui sont basées sur des présupposés épistémologiques différents (Kjeldskov, Skov, Als, & Hoegh, 2004).

### 6.1. De l'étude de laboratoire (laboratory studies)

La première approche consiste à étudier l'interaction dans un laboratoire. Il s'agit de se focaliser sur le contrôle des variables qui peuvent affecter l'interaction à travers la construction d'un environnement physique artificiel pour placer l'utilisateur dans une situation de mobilité. Par exemple, dans certaines études, on demande aux utilisateurs de marcher sur des tapis roulants ou de suivre un parcours prédéfini par un marquage au sol (Beck et al., 2003; Pirhonen et al., 2002). Le présupposé de ce type d'approche expérimentale est qu'il est possible de recréer dans un laboratoire une situation « naturelle » de mobilité d'un utilisateur. De ce point de vue, la mobilité est considérée principalement sous ses aspects physiques et d'awareness. Un tel réductionnisme pose la question traditionnelle de la validité écologique des données ainsi recueillies.

Pour un acteur, l'acte de déplacement implique des changements de contextes plus ou moins importants, non seulement d'un point de vue physique (luminosité, bruits, objets, etc.) mais aussi d'un point de vue socioculturel (lieux, gens, situations, etc.). Même s'ils n'ont pas un effet systématique sur l'activité, ces variations sont potentiellement pertinentes et peuvent avoir des effets difficiles à prévoir sur le cours d'action des utilisateurs et donc sur leur usage d'un dispositif. De plus, l'interaction homme machine génère elle-même son propre contexte de réalisation par une gestion dynamique de l'instant, orientée par un but donné (ex : rechercher une information, écrire un message, etc.).

### 6.2. En passant par les expériences de terrain (field / quasi experiments)

Certains auteurs ont identifié un compromis entre les deux approches, en utilisant la notion « d'expérimentations de terrain » (“field experiments”) (Goodman, Brewster, & Gray, 2004) ou de “quasi-expériences” (“quasi-experiments”) (Roto et al., 2004). Ils tentent de conserver un strict contrôle expérimental de facteurs tout en préservant le réalisme des usages. Leur approche est essentiellement quantitative et vise à expliquer les relations causales entre facteurs. Cette approche suit la tradition orthodoxe de la psychologie expérimentale appliquée à l'IHM (tester des hypothèses, contrôler des variables, utiliser des tests statistiques).

### 6.3. A l'étude en contexte naturel (field studies)

Afin de traiter les limitations des expérimentations en laboratoire quand il s'agit d'évaluer les artefacts mobiles, une troisième approche a été explorée, se focalisant sur les études « en milieu naturel » (écologique). Avec cette approche, le caractère naturel de la mobilité et donc son contexte est préservé autant que faire se peut, et ce qui est perdu en contrôle des variables est gagné en terme de validité écologique.

Cependant, un des challenges principaux de l'approche naturaliste est de définir des méthodes et des équipements qui permettent d'étudier empiriquement les usages et l'utilisabilité des technologies mobiles dans des situations réelles.

Des recherches récentes ont posé la question, de manière systématique, de la valeur ajoutée des études en situation naturelle dans l'identification des problèmes d'utilisabilité comparativement aux études de laboratoires (Kjeldskov et al., 2004).

Le travail présenté ici se rattache à la catégorie des études naturalistes, mais il se donne des contraintes de rigueur quant aux phases de recueil et d'analyse des observables. Il s'efforce en outre d'articuler les données quantitatives (qui constituent des éléments de synthèse commodes et permettent de résumer les tendances générales des résultats) avec des données contribuant à donner du sens aux comportements observés (verbalisations des sujets).

#### **6.4. Quels outils pour l'analyse des données ?**

L'étude de l'interaction en contexte naturel nécessite une récolte de données relativement exhaustive dont la quantité peut vite s'avérer conséquente. De plus, cette masse de données est forcément de nature hétérogène puisqu'elle fait appel à divers formats d'enregistrement de l'activité (vidéos subjectives et contextuelles, notes papier, fichiers informatiques (log) traçant les événements système, etc..

Sans un outil dédié à l'analyse de telles données, celle-ci peut très vite s'avérer fastidieuse, voire impossible à réaliser dans des délais raisonnables. En effet, le croisement de données de nature si variée est très complexe à réaliser sans un outil permettant la visualisation et l'indexation de ces divers flux.

(Dybkjær et al., 2001) ont réalisé un état de l'art relativement exhaustif des outils existants ainsi que des besoins pour l'annotation d'interactions naturelles et des données multimodales dans le cadre d'un programme nommé ISLE NIMM (Natural Interactivity and Multimodality). Ils mentionnent 12 outils ( voir en annexe 2.2 Outils d'annotation de corpus multimodaux) dont certains n'ont pas été réellement implémentés et sont au stade de l'étude ou des spécifications fonctionnelles. De manière générale, ces outils ont souvent été conçus pour des buts spécifiques tels que l'annotation de corpus de nature variée (dialogues parlés, expression faciale, gestes, posture du corps, ainsi que leur éventuelle relation) et des visées de standardisation différentes.

Comme nous le verrons par la suite, nos besoins spécifiques et les nombreuses limites que présentent ces outils, nous ont conduits vers l'adoption d'un logiciel dédié à l'analyse de l'activité nommé Actogram<sup>TM</sup> (Kerguelen, 1998).

## 7. Synthèse et conclusion

Que tirer de ce passage en revue de la littérature ? Il nous semble qu'il permet a minima de pointer sur quelques axes de recherche pertinents présentant un intérêt à la fois théorique, méthodologique et technologique.

- **L'utilisabilité des systèmes mobiles**

Un des axes les plus développés a trait à l'évaluation de l'utilisabilité des systèmes mobiles. La question posée ici concerne la possibilité de transfert des principes ergonomiques et l'applicabilité des méthodes d'évaluation développés dans le champ des systèmes interactifs « statiques » aux systèmes mobiles. Plusieurs études ont ainsi été menées pour questionner la position selon lequel la spécificité des situations de mobilité ne peut être recréée en laboratoire et que mener des études de terrain serait par conséquent indispensable à l'évaluation des systèmes mobiles. Les résultats obtenus sont souvent mitigés ; ainsi une étude de (Kjeldskov & Graham, 2003) semble montrer que le réalisme de la situation ne joue pas un rôle significatif sur l'identification quantitative de problèmes d'utilisabilité par les sujets ; mais dans le même temps la situation la plus réaliste permet de faire émerger des points non identifiés dans la situation de simulation moins réaliste. On retrouve des résultats similaires sur la complémentarité laboratoire-terrain dans d'autres études (Mersioli, 2005).

Une autre façon de considérer la question de l'utilisabilité des systèmes mobiles consiste à étudier l'effet de la mobilité sous l'angle des contraintes qu'elle génère sur l'activité des utilisateurs. Des études ont ainsi posé le problème en termes de conflit entre la composante motrice du déplacement et la gestion du dispositif interactif (Pirhonen et al., 2002). La situation type d'expérimentation va donc consister à provoquer artificiellement la déambulation d'un sujet (sur un tapis roulant par exemple) à qui l'on demande de réaliser une tâche sur un terminal portable (PDA par exemple). L'objectif est alors de relever les effets de la mobilité provoquée sur l'utilisation du dispositif. On peut également étendre ce schéma de base à une autre composante de la mobilité, à savoir l'orientation dans un espace naturel (Kjeldskov & Stage, 2004), ou artificiellement délimité en laboratoire (Barnard, Ji Soo, Jacko, & Sears, 2005). Dans ce cas le modèle prégnant de lecture de l'activité est celui de l'allocation concurrentielle de ressources cognitives entre l'environnement et le système (Jameson, 2002). Les recommandations ergonomiques pour la conception de systèmes mobiles vont donc viser à limiter les effets de cette « fragmentation » de l'activité (Oulasvirta, Tamminen, Roto, & Kuorelahti, 2005b) et à solliciter au minimum les ressources attentionnelles (Pascoe et al., 2000).

- **L'indexation contextuelle et l'aide au positionnement géographique**

Depuis quelques années, la communauté HCI a progressivement intégré le rôle déterminant du contexte dans toute forme de situation d'interaction. Dans le cadre des activités en mobilité cet aspect a par exemple été pris en compte dans la notion de dépendance contextuelle identifiée par (Pascoe et al., 2000) comme une des caractéristiques spécifiques des activités mobiles de terrain : en mobilité, les activités doivent souvent être indexées par rapport à des données de l'environnement qui sont d'ailleurs souvent limitées à la localisation topologique des utilisateurs. Cette réduction du contexte au positionnement relatif dans un espace instrumenté (ressources informationnelles, capteurs) ou non est également encore très présente dans le domaine de l'informatique sensible au contexte (*context-aware computing*).

Plusieurs études ont donc été menées au cours des dernières années, qui investiguent le rôle d'informations de type géolocalisation sur les activités mobiles individuelles mais également sur les processus de collaboration, notamment dans le domaine du jeu. Un des objectifs de ces travaux est, dans le champ des situations de mobilité, de questionner l'application possible des développements conceptuels réalisés dans le CSCW autour de la notion d'*awareness* (Nova, Girardin & Dillenbourg, 2005).

- **La continuité des espaces d'interaction (physiques et virtuels)**

Une application possible des technologies mobiles consiste à combiner espace physique et espace virtuel. Les utilisateurs peuvent alors se déplacer dans un lieu réel (un espace public urbain par exemple) mais également dans un espace virtuel qui n'existe que par le biais d'une projection numérique affichée sur le dispositif portable qui vient se superposer à l'espace physique. La correspondance entre les deux espaces de mobilité est soit réalisée par l'utilisateur (dans le cas d'un terminal affichant un ensemble de coordonnées virtuelles), soit effectuée directement dans un dispositif d'affichage unique (casque semi-transparent par exemple). Ces techniques ont souvent été utilisées dans le domaine des jeux collaboratifs (Thomas et al., 2002) et de la visite d'espaces culturels.

Les points d'intérêt théorique et conceptuel susceptibles de faire l'objet d'analyse sont nombreux dans ce type de situation ; on notera par exemple la manière dont les utilisateurs alignent dynamiquement les différentes écologies, notamment par le biais d'épisodes d'interaction (Licoppe & Inada, 2005) et la nature des stratégies mises en œuvre par les utilisateurs/joueurs pour tirer parti collectivement des informations contextuelles présentes dans des environnements mixtes (Flintham et al., 2003).

- **La multimodalité comme mode d'enrichissement de l'interaction et de gestion des erreurs**

La revue d'études indique dans un premier temps une certaine tendance des recherches à une focalisation essentiellement sur le type de multimodalité relevant de la « complémentarité ». En effet, ce type de multimodalité qui met en œuvre une fusion d'au moins deux modalités pour délivrer une commande unitaire (par exemple un geste de pointage et une action vocale à réaliser, le fameux « Put that there » de (Bolt, 1980)) est historiquement à l'origine de ce concept. Par la suite, plusieurs auteurs, et en particulier (Nigay & Coutaz, 1993) ont à travers leurs travaux, fait émerger une taxonomie de la multimodalité, faisant apparaître des propriétés de la multimodalité encore peu documentées. Nous pensons ici aux propriétés d'équivalence fonctionnelle pour l'aspect sémantique (la possibilité de réaliser une même commande à l'aide d'une modalité au choix) et à l'exclusivité pour l'aspect plus temporel (commandes séquentielles à modalité unique par commande). En effet, ce type de multimodalité, plus simple dans sa mise en œuvre technique (critère particulièrement pertinent pour les artefacts mobiles dont les ressources sont limitées) car ne faisant pas intervenir de fusion d'information a été, dans un premier temps, relativement peu étudiée par les scientifiques et à l'inverse le milieu industriel s'en est emparé massivement ces dernières années, tant dans des contextes statiques (Wii, etc.) que mobiles (Console Nintendo DS, Iphone et autres téléphones à commandes vocales), devançant les études sur leurs usages statiques ou mobiles. Une de nos études précédentes (Calvet, 2001) a tenté de dégager des premiers résultats d'observation des usages de la multimodalité équivalente, et d'en dégager une première liste de facteurs de choix d'usages, mais celle-ci était alors simulé par un magicien d'Oz, et se limitait à un contexte statique de laboratoire. Cette étude pointait, au delà de la flexibilité et de la possibilité d'adaptation au contexte des modalités choisies par les sujets, une autre propriété de la multimodalité équivalente : la robustesse aux erreurs. En effet, les sujets peuvent facilement changer de modalité afin de prévenir ou corriger une situation d'erreur de manière efficace.

- **Implications méthodologiques : pour une approche naturaliste de l'étude des usages des systèmes mobiles**

Certaines des études portant sur l'évaluation de l'utilisabilité des systèmes mobiles amènent à se poser des questions sur la finalité même de la démarche suivie. Est-il en effet pertinent de penser qu'il est dans tous les cas de figure utile et souhaitable de faire usage d'un système interactif tout en se déplaçant ? Et quel sens donner à des résultats obtenus dans des situations où l'on contraint un utilisateur à réaliser des tâches tout en marchant alors qu'il ne le ferait jamais de son propre chef ? Le problème est de ce point de vue tout à fait différent de ce que l'on peut rencontrer dans des situations de mobilité où l'utilisateur n'a pas d'autre choix que de gérer simultanément deux types d'action (conduite d'un véhicule et manipulation d'un système embarqué par exemple). Le type d'activité (professionnelle, ludique, domestique,...) dans lequel est engagé l'utilisateur constitue donc une dimension importante du contexte d'utilisation des dispositifs d'interaction dont il dispose.

Il apparaît clairement à la lecture des différents travaux publiés depuis quelques années sur le sujet que la mobilité introduit une possibilité d'extension des déterminants contextuels de l'activité en étendant le champ des situations possibles auxquelles sont dynamiquement confrontés les utilisateurs. Mais dans le

même temps on doit également considérer que le contrôle de son déplacement par l'utilisateur amène ce dernier à construire activement le contexte dans lequel il évolue, notamment par le biais des interactions sociales qu'il est susceptible de tisser (voir par exemple les travaux menés par Heath et ses collègues sur l'activité de visite de musées (vom Lehn et al., 2001)). La caractérisation de la mobilité comme expérience individuelle organisée par un projet collectif coopératif (et non plus comme simple déplacement dans un espace physique, (Urry, 2000) permet de dépasser une vision de la mobilité comme simple source de contrainte sur l'activité des utilisateurs.

Ces questions sont en partie à l'origine du débat entre partisans du recours à des situations de laboratoires qui permettent théoriquement de contrôler systématiquement les facteurs qui sont susceptibles d'influer sur l'utilisation de dispositifs mobiles, et promoteurs d'approches écologiques qui mettent l'accent sur la nécessité de tenir compte des modifications de contexte induites par la situation de mobilité. C'est notamment le cas par exemple des fluctuations de l'environnement social quand l'utilisateur se meut dans un espace public mouvant. Ce type de problématique a été posé par quelques auteurs (Jost, Haussler, Merdes, & Malaka, 2005) mais les méthodes utilisées (projection de l'utilisateur dans une situation imaginaire proposée par l'expérimentateur) sont peu convaincantes.

L'étude présentée dans le chapitre suivant participe d'un effort qui vise à documenter empiriquement ces questions.

Le cadre d'analyse présenté dans le chapitre suivant introduira notre positionnement suite à cette revue de la littérature.

## **Chapitre 2. Etude empirique : Méthode et cadre d'analyse**

## 1. Objectifs et principes Généraux

Notre objectif est de procéder à une étude empirique de l'interaction avec un dispositif multimodal et mobile, dans un environnement naturel, en se concentrant sur l'observation de pratiques réelles, donnant un sens aux actions des sujets. Le dispositif mobile testé est une application de messagerie implémentée sur un PDA qui peut être piloté en utilisant plusieurs modalités. Ce point est à noter dans la mesure où les interfaces multimodales sont souvent présentées comme une solution possible aux problèmes d'allocation des ressources perceptivo cognitives classiquement pointés dans les études portant sur l'utilisabilité des systèmes mobiles.

Un des objectifs de cette expérimentation était de questionner ces points de vue sur la base de données recueillies dans une situation « naturelle », où aucun mode opératoire ou contrainte particulière relatifs à l'usage du dispositif n'étaient imposés aux sujets.

Au-delà des conséquences physiques d'un déplacement sur l'activité, une des principales conséquences de la mobilité est d'induire des changements rapides de contexte, susceptibles d'avoir des effets directs sur les usages d'un dispositif.

Il est donc nécessaire de capturer les données concernant l'évolution du contexte, tant dans sa composante physique que sociale. Cependant, il est impossible d'identifier a priori tous les facteurs contextuels pertinents, en particulier lorsqu'il s'agit d'un environnement ouvert. Les données collectées doivent donc permettre à l'expérimentateur de pratiquer une analyse a posteriori, basée sur un enregistrement riche de l'activité et de son contexte.

Cette orientation méthodologique requiert un équipement portable et suffisamment autonome pour permettre à l'expérimentateur de suivre l'utilisateur dans des environnements différents. De plus, cet équipement se doit d'être le moins intrusif possible pour l'utilisateur, que ce soit physiquement, psychologiquement ou encore socialement.

Une solution possible est de suivre d'aussi près que possible l'utilisateur avec une caméra vidéo. Mais cette méthode présente de nombreux inconvénients. Premièrement elle peut provoquer un sentiment d'inconfort du fait qu'il n'est pas toujours socialement plaisant d'être suivi et filmé dans un espace public.

Deuxièmement, il est très difficile voire impossible d'enregistrer de manière simultanée les actions de l'utilisateur et l'interface du dispositif multimodal ; en effet, l'utilisateur peut rapidement passer d'une modalité d'entrée à une autre, alors qu'il est également nécessaire de garder une trace des modalités de sorties du système. L'expérimentateur devrait donc rapidement déplacer le champ de la caméra du sujet vers l'interface, en choisissant en temps réel l'information pertinente à capturer à un instant donné.

Troisièmement, il n'est pas possible de définir a priori les facteurs importants dans le contexte dynamique environnant qui peuvent expliquer les actions de l'utilisateur. L'expérimentateur doit donc choisir entre un focus sur la manipulation du dispositif tout en perdant les aspects contextuels, ou bien prendre un plan plus large, incluant par exemple l'interaction avec d'autres personnes (vom Lehn et al., 2001), mais en perdant cette fois les aspects précis de la séquence d'interaction.

Cet aspect en particulier, illustre, d'un point de vue technique, la différence entre deux traditions générales d'observation : d'un côté une approche s'inspirant de la sociologie et des ethno-sciences qui se focalisent sur les interactions entre les acteurs dans un même environnement physique, et d'un autre côté, l'approche IHM classique qui se focalise principalement sur l'interaction d'un sujet isolé avec un dispositif.

Dans un premier temps nous présenterons les différents cadres d'analyse que nous avons retenus, en accord avec la problématique méthodologique présentée précédemment, justifiant la méthodologie choisie que nous décrirons dans un second temps.

## 2. Cadre d'analyse

Etant donnée la diversité de thématiques abordées dans ce travail, il est impossible d'identifier un cadre d'analyse unique couvrant toutes les thématiques. Pour cette raison, nous avons choisi de mettre en application un cadre d'analyse composite faisant appel successivement aux champs thématiques traités dans ce travail. Il s'agit donc d'un cadre pour la mobilité, pour l'interaction multimodale et l'interaction en général.

### 2.1. Cadre d'analyse pour la mobilité

Le type d'application proposé et les tâches à réaliser par les sujets avaient pour vocation de provoquer et supporter des degrés de mobilités de natures et de degrés variables.

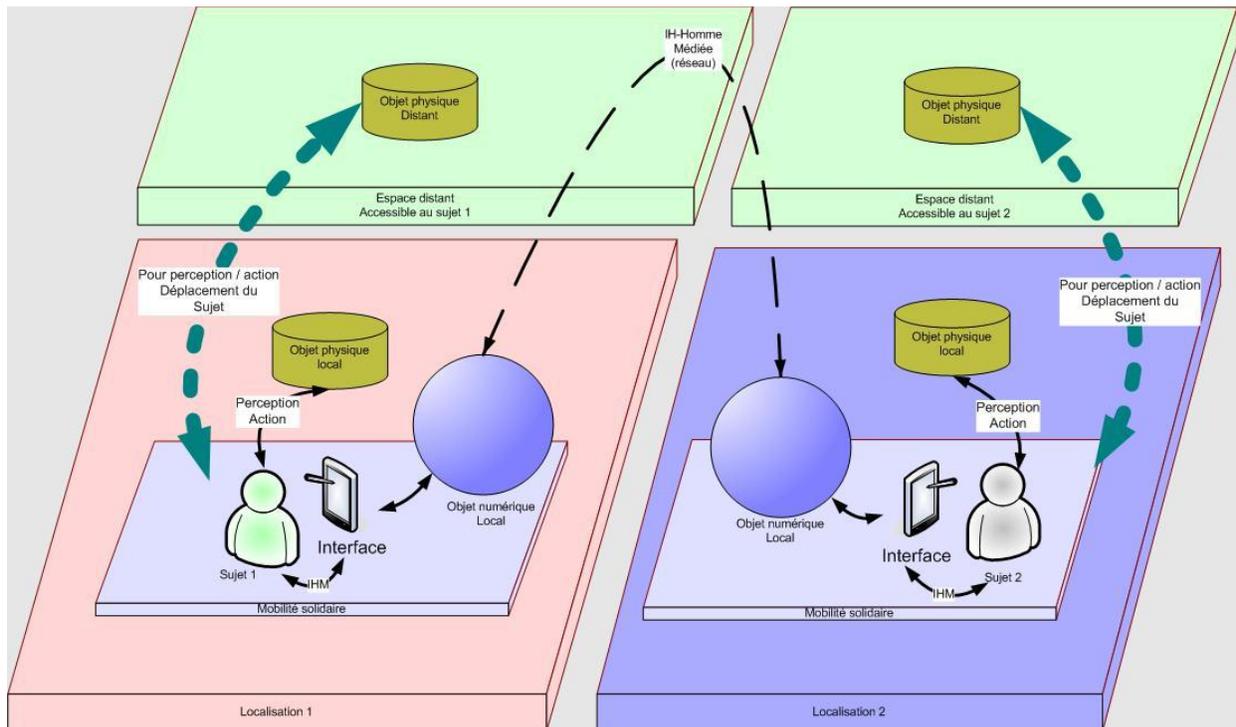


Figure 7 : Les différents types de mobilités mis en œuvre lors de nos expérimentations

En effet, comme l'a montrée la partie concernant l'étude de la mobilité, celle-ci peut concerner, au delà du déplacement d'un sujet, le déplacement d'informations, de dispositifs d'interaction, d'objets physiques etc. Dans notre cas, la mobilité que nous souhaitons provoquer prenait plusieurs formes (Figure 7) :

- **Mobilité des sujets :** Les sujets devaient se déplacer pour avoir accès aux **objets physiques distants** (objets impossibles à déplacer tels que des cartels de l'exposition, des sculptures etc.) afin éventuellement d'interagir avec ces objets physiques devenus locaux, ceci dans le but d'obtenir les réponses aux questions du quiz contenues dans leur PDA. Ceci constitue donc un cadre d'autodéplacement des sujets (gestion propre de leur déplacement).
- **Mobilité des dispositifs d'interaction :** Le dispositif d'interaction portable devait être **transporté** par le sujet. En effet, sans celui-ci, la tâche serait impossible à réaliser, car il constitue à la fois le média (l'interface) permettant toutes les interactions mais en plus il contient l'ensemble des **données numériques locales** qui sont également embarquées dans les déplacements (questions du quiz et réponses contenant les indices envoyés par leur partenaire) et qui constituent des ressources indispensables. Le sujet pouvant éventuellement interagir en même temps qu'il se déplace, situe le cadre dans celui de la mobilité « réelle ».

- **Coopération distante et mobilité de l'information** : Afin de pallier à la distance entre les deux sujets ne permettant pas l'échange d'information en **coprésence**, il leur était donc nécessaire de procéder à l'envoi de données numériques distantes devenant alors locales pour un usage également local. Cela nous place dans le cadre du travail collaboratif mobile médié par ordinateur (CSCW mobile).

## 2.2. Cadre d'analyse pour l'interaction multimodale

Contrairement à de nombreuses études, le type de multimodalité implémenté dans notre application ne traite pas des aspects complémentaires mettant en œuvre la fusion d'au moins deux modalités pour délivrer une commande en entrée du système, ou pour délivrer une information à l'utilisateur. En effet, dans le cadre d'une étude précédente (Calvet, 2001), nous avons montré que l'aspect complémentaire offert par une application n'est pas forcément exploité par les sujets si le domaine des tâches proposé ne s'y prête pas. En effet, dans cette étude, comme dans celle présentée ici, nous avons proposé une application de messagerie peu propice à l'usage de la multimodalité complémentaire (pas de manipulation directe en entrée, mais plutôt des commandes discrètes). Seuls quelques rares cas avaient pu être observés.

De plus, des contraintes techniques liées aux capacités de calcul des dispositifs techniques retenus et aux difficultés de développement d'applications mettant en œuvre la complémentarité, nous ont orientés vers d'autres types de multimodalités.

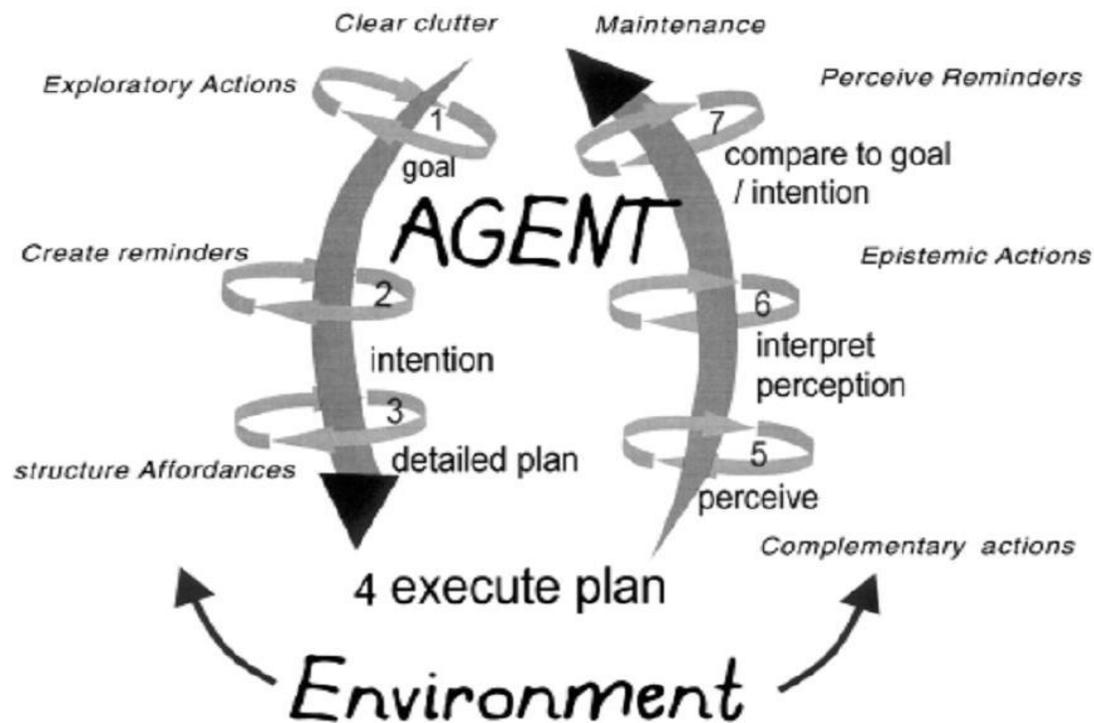
En effet, nous nous intéressons à des aspects identifiés par (Nigay & Coutaz, 1993) qui ont, à travers leurs travaux, fait émerger une taxonomie de la multimodalité, faisant apparaître des propriétés de la multimodalité encore peu documentées. Nous pensons ici aux propriétés **d'équivalence fonctionnelle** pour l'aspect sémantique (la possibilité de réaliser une même commande à l'aide d'une modalité au choix) et à **l'exclusivité** pour l'aspect plus temporel (commandes séquentielles à modalité unique par commande), propriétés que l'on retrouve dans les modèles CARE et CASE.

Afin de ne pas placer les sujets devant des contraintes techniques de choix d'usage d'une modalité, et du fait de la similarité de l'application avec celle de notre étude précédente (Calvet, 2001), un **principe d'équivalence fonctionnelle maximale** nous a guidé pour le développement des modalités d'interaction présentes dans l'application. L'objectif était de proposer aux sujets une palette de modalités équivalentes pour réaliser le plus de commandes possibles. Cependant, quelques fonctions n'ont pas pu être implémentées selon ce principe, telles que le scroll, la manipulation des médias audio et de manière générale l'insertion de médias. Ce fait constitue une des limites de notre étude.

## 2.3. Cadre général d'interaction : le modèle de Norman et Kirsh

Le modèle de Norman, enrichi par la contribution de (Kirsh, 1997) décrit de manière globale le déroulement d'une action du point de vue cognitif, en formulant l'hypothèse que les sujets forment des plans d'action correspondant à une intention relativement de haut niveau. L'exécution de ce plan est constituée d'une phase d'exécution, suivie d'une phase perceptuelle (épistémique) qui permet d'interpréter les résultats des actions et de les comparer au but poursuivi, et ainsi de reformuler une nouvelle intention.

Comme l'indique également (Stanton, 2003), l'action commence par le besoin de satisfaire un but particulier (ex : suivre un match à la radio), qui se traduit par la formulation d'intentions (allumer la radio) qui s'appuie sur la mémoire à long terme. Les intentions sont ensuite traduites en une séquence d'actions qui sont finalement exécutées. L'exécution de ces actions se fait grâce à leurs boucles de feedback, qui doivent être perçues, interprétées et comparées au but par le sujet (ex : la radio est bien allumée). Tout ceci se basant sur la mémoire à long terme. Une partie de cette activité se déroule à l'aide de la mémoire de travail consciente, alors qu'une autre partie peut être largement automatisée, sans ou avec très peu de contrôle conscient.



**Figure 8 : Modèle de Norman enrichi par Kirsh. Les flèches entre les ellipses représentent le couplage entre l'agent et l'environnement. A la différence du modèle du cycle de décision qui traite du but, de l'intention et des phases de planification comme des actions séquentielles, dans ce modèle, chaque étape impliquée dans la prise de décision est interactive. Il montre également que la nature de l'interactivité est diverse, allant d'une simple action à des actions préparatoires, exploratoires, de maintenance ou complémentaires.**

(Kirsh, 1997), en s'inspirant du courant de la cognition située et distribuée, complète le modèle de Norman en y ajoutant des éléments (Figure 8) qui resituent l'action et la cognition dans la matérialité de l'environnement, en insistant sur l'aspect opportuniste de l'action/cognition. Il explique comment les sujets vont organiser leur environnement (« structure affordances », « create reminders », etc.) pour faciliter la remise en contexte cognitif et ainsi rendre plus évidente, affordante, l'action suivante. Ce modèle présente plusieurs intérêts et inconvénients que nous allons décrire :

### 2.3.1. Intérêts du modèle de Norman / Kirsh

#### 2.3.1.1. Point de vue de la cognition située et distribuée

Tout en présentant des aspects contradictoires (voir ci dessous), ce modèle est l'un des rares à se baser sur la théorie de la cognition située et distribuée et dont la circularité indique un réajustement permanent de l'action en fonction des modifications de la situation suite à l'action du sujet.

#### 2.3.1.2. Prise en compte de la matérialité de la cognition

En décrivant de manière précise l'impact de l'environnement matériel sur la cognition, Kirsh est en accord avec le modèle de la cognition distribuée. Il va même au-delà en décrivant des actions volontaires des sujets pour faciliter le cours d'action en organisant de manière pertinente leur environnement (situation). Les humains exploitent les ressources matérielles mises à leur disposition comme support de médiation et de mémorisation de représentations, mais également comme support à « l'externalisation » lié au pouvoir de calcul des objets (tableaux, abaqués, cartes etc...).

### 2.3.1.3. Notion d'intention

La notion d'intention est intéressante car elle se situe clairement du côté du sujet et en cela, ce modèle est donc centré sur l'activité du point de vue d'un acteur (tout en prenant en compte son environnement), comme le fait l'ergonomie. Lorsqu'il s'agit d'étudier l'interaction homme-machine, la notion d'intention peut être pertinente car les actions de l'utilisateur sont souvent discrètes, relativement isolées et donc facilement identifiables. Dans ce cas, il est possible sans trop d'artifices d'associer une intention à une action, en particulier pour l'ensemble des commandes discrètes. Cette association, ce couple intention/action, permet dès lors d'identifier leur bonne adéquation et ainsi de distinguer des situations de distorsion, voire d'erreurs.

### 2.3.1.4. Séquentialité de l'action et de la cognition

L'aspect séquentiel du modèle de Kirsh présente des intérêts pratiques du point de vue de l'analyse de l'interaction, puisqu'il permet d'en identifier des phases, des séquences, qui se focalisent à chaque fois sur des aspects différents de l'action et de la cognition. Ces étapes sont au nombre de sept :

- formation d'un but
- formation d'une intention
- formation d'un plan détaillé
- exécution du plan
- perception des résultats de l'exécution de l'action
- interprétation des résultats de l'action
- comparaison avec le but

Et le cycle peut alors reprendre avec la formation éventuelle d'un nouveau but.

## 2.3.2. Critique du modèle de Norman / Kirsh

### 2.3.2.1. Focus et niveau d'intérêts

Le modèle de Norman est relativement de haut niveau et simplifié pour rendre compte de l'activité humaine, puisqu'il ne prend pas en compte de nombreux facteurs comme les émotions, la personnalité, le genre, les facteurs culturels et sociaux comme l'indique (Stanton, 2003).

De plus, en s'intéressant au niveau global de la cognition et de l'action, il en devient très peu descriptif à des niveaux plus bas de l'activité, et finalement, ne décrit pas le détail de l'interaction qui est souvent le focus de l'ergonomie des IHMs. En cela, son utilité est restreinte pour l'aide à la conception détaillée et de bas niveau qui nous concerne ici.

### 2.3.2.2. Problème du découpage séquentiel de l'activité

Tout en se réclamant du courant de la cognition distribuée et située, ce modèle procède à un découpage de l'activité en 7 phases qui semblent entrer en contradiction avec les principes théoriques de ce courant de pensée. En effet, soit ce modèle se situe à un niveau « méta » de la cognition (enchaînement de plusieurs séquences d'exécutions d'action), et en tant que telle, ne peut « parler » que d'elle, ce qui en soit constitue un intérêt (limité pour notre étude), soit il se place à un niveau plus bas de l'action/cognition (unité d'action), et dans ce cas, il devient insuffisamment précis pour aider à concevoir des systèmes.

De plus, au sein même de la communauté de la cognition située, la notion de découpage en séquences de l'action/cognition est largement remise en cause, en particulier la notion de plan. Ce modèle présente donc des ambiguïtés quant à son statut.

Si l'on regarde les quatre grandes séquences qui constituent ce modèle (Figure 8), les phases 1 à 3 consistent en une grande phase cognitive de formation d'un plan. Les phases 5 à 7 sont également des phases cognitives liées à la perception et à l'interprétation du résultat des actions (phase épistémique). Seule la phase 4 (exécution du plan) ne concerne pas explicitement la cognition mais l'action. Il y a donc ici une nette dichotomie entre les deux, ce qui semble peu compatible avec la notion de cognition située. En effet, l'action semble alors complètement déconnectée de la cognition, dans une boucle séquentielle faisant intervenir alternativement la cognition et l'action.

Comme nous l'avons mentionné plus haut, ce défaut du modèle peut être accepté dans le cas des commandes « discrètes » où l'action peut être facilement isolée de manière unitaire, et où chaque action

discrète appelle une réaction discrète corrélée. Il est alors acceptable et cohérent de séquentialiser les actions, même si la notion de plan reste toujours discutable.

Cependant, dans le cas de styles d'interaction de manipulation directe, l'unicité de l'action est difficile, voire impossible à identifier. Les utilisateurs sont plus souvent dans des phases d'exploration de l'interface par la manipulation et dans ce cas, n'ont pas toujours d'intention claire ni de plan de séquence pour obtenir un résultat. Le résultat souhaité n'est d'ailleurs pas forcément identifié au préalable, il s'agirait plus d'hypothèses de résultats souhaités que l'utilisateur confronte à la réalité des résultats obtenus (par exemple, lorsque l'on retouche une photo avec un outil pour gommer les yeux rouges, on a effectivement une intention (méta) qui est de gommer les yeux rouges, mais lorsque l'on ballade le crayon Photoshop, on explore le résultat au fur et à mesure que l'on agit).

### **2.3.2.3. Absence de description du comportement des systèmes**

Le modèle de l'action/cognition de Norman et Kirsh est clairement placé du côté du sujet agissant, cependant, le sujet agit en fonction de son environnement, du contexte et de la situation (Quéré, 1997). Ceci inclut la présence d'un système informatique qui peut avoir un comportement spécifique qui n'est absolument pas décrit dans ce modèle, et qui pourtant va influencer à son tour le comportement de l'utilisateur.

Cependant, selon (Stanton, 2003), il aide à expliquer différents aspects de l'activité humaine qui peuvent directement concerner l'ergonomie. En effet, le système peut se retrouver dans les différentes phases présentées dans ce modèle mais son niveau de précision et de description est faible. Un meilleur modèle de l'interaction devrait donc permettre de décrire d'une manière plus fine ses caractéristiques et son comportement.

### **2.3.3. Usages du modèle dans notre étude**

Ce modèle nous sera particulièrement utile lors de la phase d'analyse fine de séquences d'interactions supportées par des illustrations (2 Etude de cas et illustrations, p224). En effet, nous en reprendrons les principales composantes tout en identifiant certaines de ses limites pour lesquelles nous proposerons des alternatives. Au final, cela nous permettra de proposer un modèle détaillé de l'interaction, support à une analyse fine (voire micro) de l'interaction, pré requis nous semblant indispensable pour la conception d'interfaces multimodales mobiles.

## 3. Méthode

### 3.1. Description de l'application

Pour répondre aux besoins de l'expérimentation in situ, une application a entièrement été développée en collaboration avec l'équipe Diamant de l'IRIT. En effet, dans les conditions de mobilité envisagées, il était très difficile, voire impossible d'utiliser la technique du magicien d'Oz comme dans notre étude précédente (Calvet, 2001).

#### 3.1.1. Architecture et fonctions de l'application

Notre choix s'est porté sur une application de messagerie Wifi, multimédia, mobile et multimodale implémentée sur un PDA de type IPAQ TM de Compaq, à l'aide de laquelle les sujets peuvent consulter, éditer et recevoir des messages provenant de leur partenaire.

La Figure 9 illustre l'architecture fonctionnelle ainsi que les principales commandes disponibles dans notre application.

Celle-ci est composée de trois espaces principaux :

- **Liste des messages** : Espace contenant l'ensemble des messages et permettant l'accès aux messages archivés, envoyés et reçus. Depuis cet espace, les sujets ont éventuellement accès aux deux autres espaces.
- **Lecture d'un message** : Espace permettant de consulter le contenu d'un message et éventuellement d'y répondre ou de le transmettre. Les messages peuvent contenir différents types de média (son, image, texte dactylographié ou manuscrit).
- **Edition d'un message** : Espace permettant l'édition d'un message, l'insertion d'un média (texte, son, dessin), la sélection d'un destinataire et l'envoi du message.

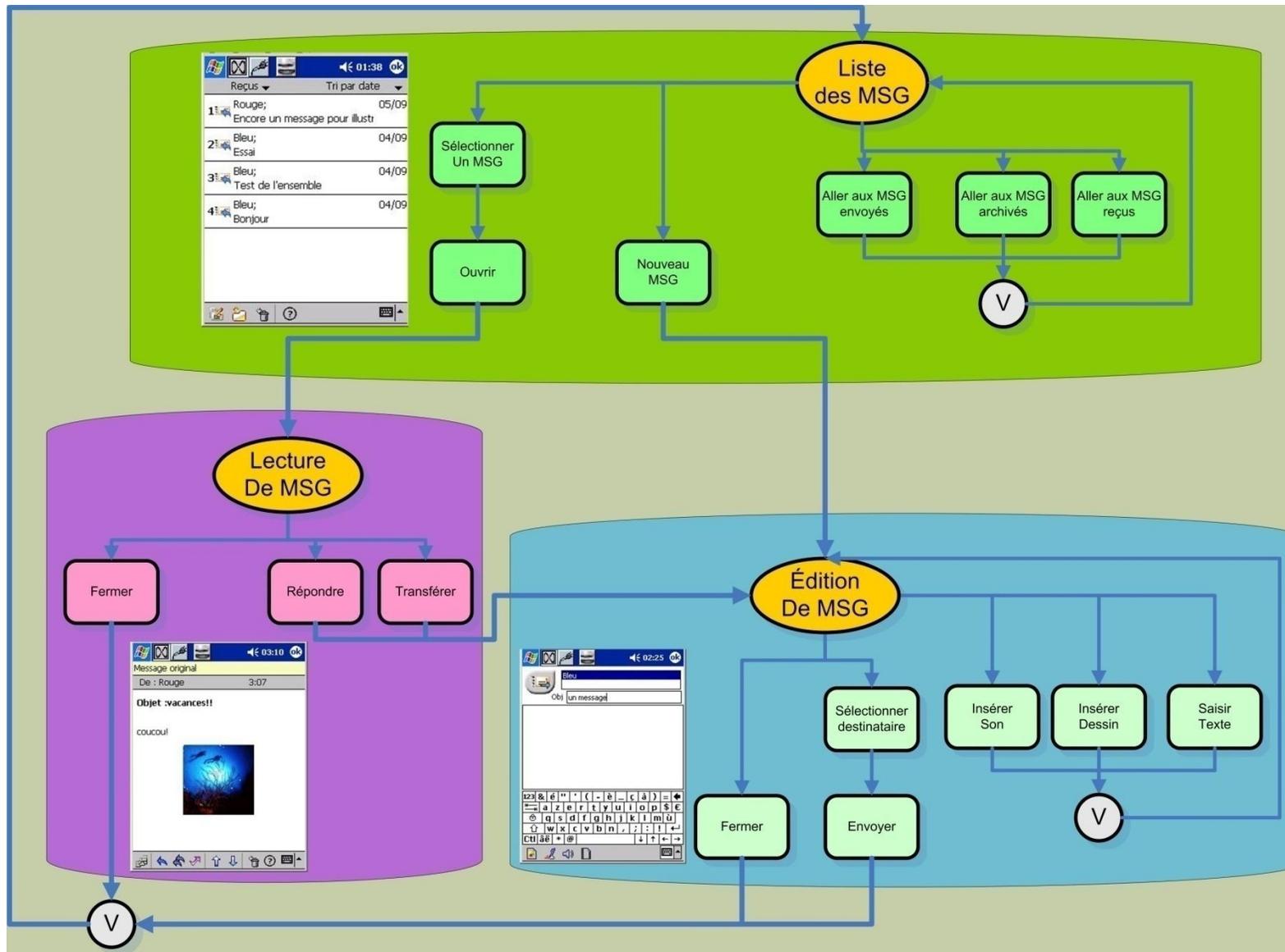


Figure 9 : Architecture fonctionnelle et principales commandes de l'application

En dehors de la liste ou du contenu propres des messages, le reste de l'application contenait deux barres de menus. La barre supérieure permettait la gestion des modes pour le pilotage des modalités d'entrées et de sortie (commandes vocales, Geste à deux dimensions (G2D), synthèse vocale) ainsi que le réglage du niveau sonore de l'application.

### 3.1.2. Modalités d'interactions disponibles

L'ensemble des modalités a été implémenté selon le principe d'équivalence fonctionnelle maximale (sauf pour le G2D) qui permet à l'utilisateur l'usage, pour toutes les commandes, de toutes les modalités.

En entrée, trois modalités d'interaction ont été implémentées :

- La modalité **tactile** « **stylet** » classique qui permet de pointer sur les boutons et icônes correspondants à des commandes et sur les zones de texte.
- La **modalité vocale** : l'application est capable de reconnaître une trentaine de mots clés. Il est possible de combiner certaines commandes vocales, par exemple « Ouverture du message 5 ».
- La **modalité geste 2D** : l'application est capable de reconnaître une demi-douzaine de gestes en deux dimensions, effectués à l'aide du stylet sur l'écran du PDA. Ces gestes sont prédéfinis et limités essentiellement à quelques commandes d'édition (copier, coller, supprimer, nouveau message,...).

En dehors des sorties graphiques visuelles classiques, l'application dispose, en sortie, de plusieurs médias et modalités :

- **Audio** : Pour chaque événement système, l'application émet un « clic » **sonore** indiquant le bon déroulement des séquences de commandes (feedback de prise en compte de la commande, feedback d'effet de la commande comme un changement d'état ou de fenêtre, etc.).
- **Modalité Synthèse vocale** : La messagerie dispose d'une **synthèse vocale** qui peut être activée et désactivée. Lorsque les commandes sont vocales, il y a systématiquement un retour vocal de réussite ou d'erreur (ex : « message envoyé à Rouge », « erreur de reconnaissance » etc.).
- **Média synthèse vocale** : La synthèse peut également, à la demande, être utilisées en tant que média pour lire les informations affichées dans l'écran courant, l'en-tête des messages ou leur contenu.

### 3.2. Tâche et consigne

La tâche consistait pour des couples de sujets à répondre collaborativement aux questions d'un jeu de type quiz. Les réponses devaient être trouvées dans des espaces distincts couvrant chacun approximativement 500 m<sup>2</sup> (Figure 10) (les sujets ne se rencontraient donc pas physiquement pendant le déroulement des sessions).



Figure 10 : Plan de l'espace d'expérimentation de la Cité de la Science et de l'industrie

Chacun des sujets avait cinq questions sur son PDA, et disposait de cinq indices correspondant aux questions de l'autre sujet. Il était impossible de répondre aux questions sans demander au partenaire les indices correspondants. Les indices étaient constitués de photos prises dans l'exposition et permettant d'orienter les sujets vers l'emplacement où se trouvaient les réponses dans l'environnement.

Chaque session se déroulait dans un nouvel espace avec des jeux de questions adaptées et renouvelées afin d'éviter la mémorisation des réponses d'un sujet à l'autre.

D'un point de vue expérimental l'objectif de la consigne et de la tâche était double :

- **Provoquer de l'interaction** : Favoriser les échanges entre les sujets tout en incitant de manière naturelle à interagir avec le PDA. Notre hypothèse consistait à déplacer la finalité de l'interaction qui était donc plus l'échange d'information entre les sujets (composante sociale) que l'interaction en elle-même.
- **Provoquer de la mobilité** : Provoquer un engagement réel finalisé dans l'exploration de l'environnement afin de trouver les réponses au quiz. Cette exploration pouvait se traduire par des déplacements du sujet ou simplement par une exploration visuelle.

Au final, notre objectif était donc de provoquer une obligation de partage des ressources attentionnelles entre l'interface et l'environnement.

L'aspect ludique, dynamique et social devait permettre de basculer le focus de la tâche primaire d'une interaction isolée de tout sens vers une activité complexe d'effectuation d'un quiz à l'aide d'un dispositif multimodal (l'interaction devenant de fait une tâche plus secondaire).

### 3.3. Sujets et protocole expérimental

Nous avons procédé à l'envoi du « manuel utilisateur » à tous les sujets au moins une semaine avant la passation du test. Celui-ci expliquait le contexte de l'expérimentation, décrivait brièvement le principe du quiz et le détail du fonctionnement de l'application ainsi que des modalités d'interaction.

L'accueil du couple de sujet et le remplissage des formalités se faisait dans le cadre de la Cité de la Science.

Nous avons alors procédé à une session de familiarisation avec le système où l'ensemble des fonctions et des modalités leur étaient montrées en parcourant le document utilisateur, puis les sujets pratiquaient quelques exercices dirigés puis libres.

Les 6 sujets ont effectué **trois** sessions d'interaction de 40 à 50 minutes, toutes espacées de plusieurs jours.

Date	Evènement	Lieu
J-15	Envoi du manuel utilisateur	Domicile des sujets
J	Accueil des sujets par paire Familiarisation et exercices	Cité de la science
>J+1	Rappel de l'application et de la consigne Première session Autoconfrontation libre et dirigée	Cité de la science Espaces 1 et 2 Laboratoire Lutin
>J+2	Rappel de l'application et de la consigne Deuxième session Autoconfrontation libre et dirigée	Cité de la science Espaces 2 et 3 Laboratoire Lutin
>J+3	Rappel de l'application et de la consigne Troisième session Autoconfrontation finale libre et dirigée	Cité de la science Espaces 3 et 1 Laboratoire Lutin

**Tableau 6 : Protocole expérimental**

Chaque session était suivie d'une autoconfrontation individuelle prenant les formes habituelles d'entretiens :

- **Libre** : Question type : comment ça c'est passé aujourd'hui ?
- **Semi-dirigée** : Question type : tu as utilisé la modalité vocale aujourd'hui ?
- **Dirigée** : La vidéo de la vue subjective de l'interaction a été utilisée comme support de remise en contexte du sujet dans son activité. De manière générale, l'expérimentateur faisait défiler l'entièreté de la vidéo de la session et s'arrêtait lorsque des événements intéressants se déroulaient ou lorsque le sujet prenait spontanément la parole pour décrire son activité. Lorsque la vue subjective ne suffisait pas au sujet pour une bonne remise en situation, nous avons parfois également recours à la vue contextuelle.

Les sujets ont été recrutés par une société spécialisée dans le domaine et rémunérés pour les besoins de l'expérimentation. Ils devaient simplement répondre à deux critères :

- être familiarisés avec l'usage du téléphone portable et/ou des PDAs
- ne pas être des professionnels d'un domaine proche de l'informatique

Nous n'avons effectué aucune étude concernant les catégories de sujets (genre, âge, expertise etc.) et d'éventuelles conséquences sur les usages, nous ne présentons pas les caractéristiques des sujets.

### 3.4. Description du dispositif d'enregistrement et de recueil des données

Nous avons conçu un dispositif de recueil de données qui se veut le moins intrusif possible, ceci tant au niveau physique, que du point de vue de sa discrétion. Ce dispositif doit permettre de recueillir des données sur les actions et les interactions de l'utilisateur avec le système, ainsi que le contexte d'utilisation. Le dispositif est composé de trois modules de recueil :

- **Vue « contextuelle »** : Un système de prise de vue large qui permet de situer l'utilisateur dans son contexte global d'usage (vue contextuelle observateur). Ceci est réalisé à l'aide d'une caméra DV classique portée par l'expérimentateur à une distance suffisamment grande pour être la moins intrusive possible tout en fournissant une vue générale du contexte de l'interaction.
- **Vue « subjective »** : Un système de prise de vue du champ visuel de l'utilisateur réalisé au moyen de lunettes-caméra (Figure 11). Cette vue capture une partie de l'interaction avec le système et fournit également des informations sur le contexte de cette interaction.
- **Fichiers « Log »** : Un système de capture automatique des actions de l'utilisateur sur le système et les sorties de celui-ci (capture log), facilitant ainsi un fastidieux codage manuel. Ce fichier log était enregistré via le serveur réseau sans fil gérant l'application de messagerie.

Ces données sont ensuite complétées par les verbalisations des sujets issues des entretiens d'autoconfrontation. Cette richesse des données permet une bonne finesse d'analyse de l'activité de l'utilisateur.

#### 3.4.1. Vue contextuelle

Afin d'identifier à posteriori les éléments contextuels qui rendent significatives les actions de l'utilisateur pour l'expérimentateur, il est crucial de capturer l'information contextuelle pertinente du point de vue de l'utilisateur. Ceci peut-être réalisé à l'aide de caméras judicieusement distribuées dans l'environnement (quand l'utilisateur se déplace dans un espace restreint et identifié), ou bien en utilisant une caméra portée par un expérimentateur.

Les caméras fixées dans l'environnement manquent de flexibilité, il n'y a qu'un point de vue sur la situation et il ne peut pas être modifié dynamiquement. D'un autre côté, un expérimentateur qui suit un utilisateur avec une caméra vidéo peut ajuster son cadrage en temps réel afin de capturer le point de vue le plus pertinent à chaque instant. En contrepartie, avec cette approche, les différents expérimentateurs ne se comporteront pas de manière uniforme face au choix de cadrage. Idéalement, il serait préférable que ce soit le même expérimentateur qui procède aux prises de vues tout le long de l'expérimentation.



**Figure 2. L'expérimentateur suit le sujet de manière non intrusive avec une caméra vidéo**

Nous avons choisi l'option d'un expérimentateur qui suit le sujet à distance pour avoir une vue large de la situation. Contrairement à d'autres approches, la présence de l'expérimentateur reste discrète et non intrusive du fait qu'il n'est pas nécessaire de capturer les interactions du sujet avec le système, ces informations étant fournies par la vue subjective et le fichier log décrits ci-après.

#### 3.4.2. Vue Subjective

La capture des informations contextuelles dans l'environnement n'est pas suffisante dans la mesure où elle ne permet pas l'accès au point de vue de l'utilisateur, notamment lorsqu'il s'intéresse aux artefacts ou s'il les manipule. De ce fait, nous avons utilisé une mini caméra portable et autonome, fixée sur des montures de lunettes (Figure 11) afin de fournir une vue la plus proche possible du champ de vision de l'utilisateur.



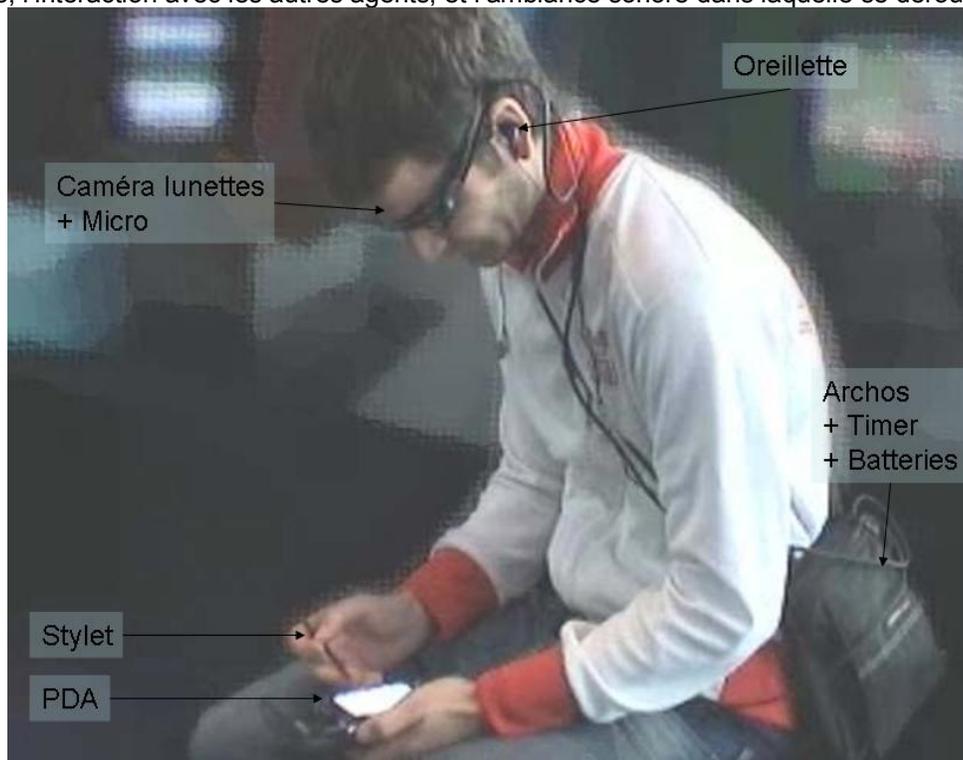
**Figure 11 : Les deux modèles de lunettes-caméra utilisés. A gauche, une caméra discrète et micro intégrés dans les branches, à droite une caméra et micro moins discrets montés sur rotule orientable**

Ceci permet de capturer des données concernant l'environnement immédiat dans le champ de vision de l'utilisateur, l'interaction avec le dispositif multimodal, et plus particulièrement, les manipulations de ce dispositif. Ces informations complètent les données « brutes » collectées dans le fichier log et facilitent leur analyse.



**Figure 3. Exemple de capture du point de vue subjectif de l'utilisateur**

La micro caméra est équipée d'un micro qui fournit les informations sonores de l'environnement, en particulier : les interactions vocales avec le module de reconnaissance vocale de l'application multimodale, l'interaction avec les autres agents, et l'ambiance sonore dans laquelle se déroule l'activité.



**Figure 12 : Dispositif mobile d'enregistrement**

Les données audio vidéo (incluant les sorties son de l'oreillette) sont directement enregistrées et compressées au format DivX sur un disque dur portable de type Archos™. Ce format permet de rendre disponibles immédiatement après les sessions les vidéos afin de procéder aux autoconfrontations.

L'utilisateur porte le dispositif et les batteries des camera-lunettes dans un sac à dos ou dans une sacoche selon les modèles utilisés (Figure 12). Il s'agit donc d'un dispositif d'observation mobile et compact, relativement autonome et peu intrusif.

#### 3.4.3. Enregistrement des fichiers Log (Logger)

La vue subjective n'est pas toujours suffisante pour capturer tout les événements se produisant sur l'interface du dispositif mobile. Ceci peut-être dû à des problèmes banals comme la présence de reflets sur l'écran, ou si l'utilisateur regarde quelque chose ou quelqu'un qui fait sortir le PDA du champ de la caméra.

De plus le codage manuel des données liées à l'interaction avec le dispositif à partir de fichiers vidéo peut s'avérer difficile et très long. Pour remédier à ces limitations, nous avons développé une application d'enregistrement des événements systèmes pertinents du point de vue de l'analyse des interactions. Les événements sont enregistrés sur un serveur connecté sur le réseau WIFI (802.11) (Merlin, 2004).

Journal d'évènements - C:\Documents and Settings\merlin\Mes documents\Démonstration1.tr						
Fichier Action Parametres						
Acteurs	Modalites	Fenetres	Messages	Messages systè...	Arguments à déba...	Temps (hh:mm:s...
Rouge	WIMP	EditeurMail	Selection du dest..			00:01:00 - 550
Rouge	WIMP	EditeurMail	Objet : saisie de B			00:01:04 - 273
Rouge	WIMP	EditeurMail	Objet : saisie de o			00:01:04 - 827
Rouge	WIMP	EditeurMail	Objet : saisie de n			00:01:05 - 262
Rouge	WIMP	EditeurMail	Objet : saisie de j			00:01:05 - 493
Rouge	WIMP	EditeurMail	Objet : saisie de o			00:01:05 - 758
Rouge	WIMP	EditeurMail	Objet : saisie de u			00:01:06 - 21
Rouge	WIMP	EditeurMail	Objet : saisie de r			00:01:06 - 524
Rouge	WIMP	EditeurHTML	selection de 1 à 1			00:01:07 - 168
Rouge	WIMP	EditeurHTML	saisie de E			00:01:10 - 573
Rouge	WIMP	EditeurHTML	selection de 2 à 2			00:01:10 - 580
Rouge	WIMP	EditeurHTML	saisie de s			00:01:10 - 982
Rouge	WIMP	EditeurHTML	selection de 3 à 3			00:01:10 - 993
Rouge	WIMP	EditeurHTML	saisie de s			00:01:11 - 175
Rouge	WIMP	EditeurHTML	selection de 4 à 4			00:01:11 - 180
Rouge	WIMP	EditeurHTML	saisie de a			00:01:12 - 197
Rouge	WIMP	EditeurHTML	selection de 5 à 5			00:01:12 - 203
Rouge	WIMP	EditeurHTML	saisie de i			00:01:12 - 931
Rouge	WIMP	EditeurHTML	selection de 6 à 6			00:01:12 - 937
Rouge	WIMP	EditeurHTML	saisie de ESPACE			00:01:13 - 447

**Figure 13 : Exemple de capture d'un fichier Log dans l'application de gestion des Logs. Dans l'exemple, l'utilisateur saisie au stylet le mot « bonjour »**

Une application d'édition et de gestion de ces fichiers Logs a également été développée (Figure 13). Celle-ci permet, l'affichage, le filtrage et l'export des données vers d'autres applications (Actogram™, Excel) permettant leur analyse.

Parmi les données retenues dans les fichiers logs, figurent le tag du temps, l'acteur de l'action (système ou sujet), la modalité employée, l'espace de l'application en cours, et la nature de l'événement (commande utilisateur, sortie système etc.).

#### 3.4.4. Sessions de Rejeu et d'autoconfrontations

L'idée générale de l'approche est de fournir à un sujet ou à un groupe de sujets un support constituant une trace d'activité (ex : écrits, schémas, annotations, données log, ou données audio vidéos) afin de recueillir des commentaires verbaux.

Les traces externes permettent également à l'analyste de recouper les commentaires verbaux avec les traces de l'activité.



**Figure 14 : Une session d'autoconfrontation**

Le terme d'autoconfrontation s'applique lorsqu'un sujet est « reconfronté » à une trace enregistrée de sa propre activité dans le but de le replacer dans le contexte de son expérience passée et ainsi d'obtenir des commentaires et explications sur son activité et en particulier sur des aspects non observables directement.

Dans notre cas, la vue subjective et contextuelle fournissent un support audiovisuel à la conduite des entretiens d'autoconfrontations avec les utilisateurs (Figure 14).

Dans le cadre de notre expérimentation, les fichiers vidéo étaient visualisés directement après les sessions de test sur une station PC équipée d'une version adaptée du logiciel Actogram<sup>TM</sup> (<http://www.actogram.net/>) qui permet l'annotation en temps réel de plusieurs fichiers audio vidéo (Kerguelen, 1998).

Les sessions d'autoconfrontations étaient elles-mêmes filmées afin d'en permettre une analyse ultérieure plus riche.

### **3.5. Méthode et outil de codage des données**

Les données audio-vidéo ont été codées manuellement à l'aide du logiciel Actogram<sup>TM</sup> qui permet la visualisation simultanée de plusieurs vidéos synchronisées (Figure 15). Ce logiciel possède une interface dédiée permettant le codage « à la volée » des informations selon des catégories contextuelles prédéfinies par l'expérimentateur afin de procéder à l'analyse.

Différentes taxonomies d'informations contextuelles ont été utilisées dans de récents travaux sur l'utilisabilité des dispositifs mobiles. La plupart d'entre elles sont essentiellement empiriques et ne se basent pas sur des concepts définis par la théorie.



**Figure 15 : L'outil d'analyse de données Actogram™**

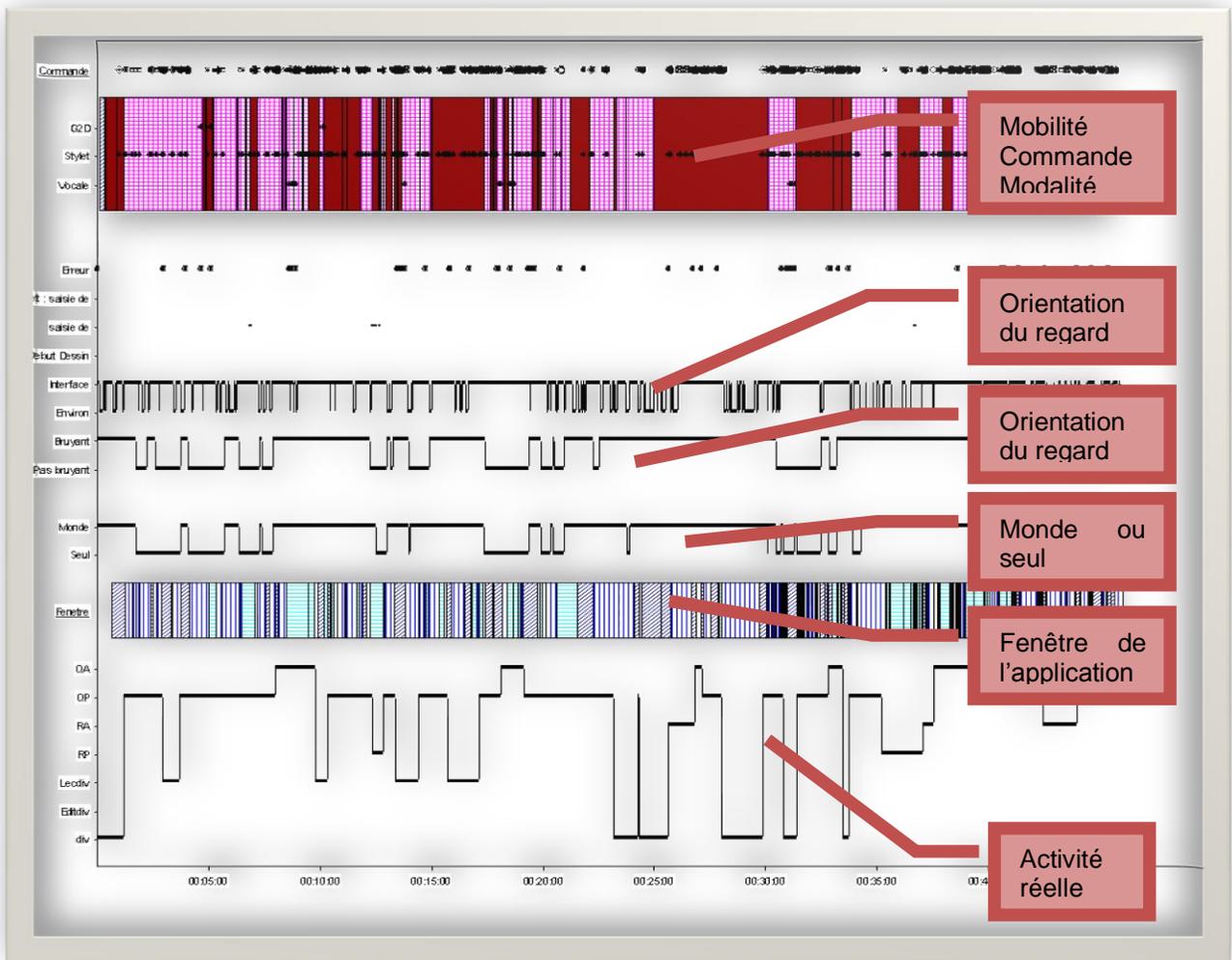
Nous avons codé les différents événements, incluant les actions des utilisateurs en trois catégories, afin de décrire les dimensions contextuelles des activités situées. Nous nous inspirons ici de la distinction opérée par (Quéré, 1997) entre « environnement », « contexte » et « situation ». Dans notre étude, « environnement » fait référence aux propriétés physiques de l'environnement (lumière, bruit, agencement physique de l'espace) qui pourraient jouer un rôle dans le choix et l'usage des modalités d'interaction disponibles. Ces propriétés sont identifiées et codées par l'expérimentateur au cours de visionnages répétés des vidéos (vues subjectives et contextuelles).

Le contexte fait référence à un élément qui donne du sens à une action : le type d'activité dans laquelle est engagé le sujet à un moment donné (ex : interaction avec le dispositif, acquisition d'information, gestion du déplacement, etc.), historique de la dynamique du cours d'action du sujet (ex : séquence d'erreurs, épisode d'interaction avec l'autre utilisateur, interaction avec d'autres acteurs, etc.). Ces informations contextuelles sont issues du croisement entre le visionnage des vidéos et des données tirées des fichiers logs.

La « situation » fait référence directement à l'expérience de l'utilisateur et à son engagement dans l'action à ce moment donné. Ce niveau d'analyse, qui nécessite une description de l'activité du point de vue de l'utilisateur, est basé sur les rapports d'autoconfrontation et l'étude des verbalisations qui en découlent.

L'ensemble des données ainsi récoltées et annotées permet la génération de fichiers d'observations complets qui permettent selon le filtrage effectué de réaliser des analyses quantitatives ou qualitatives à un niveau global (tout sujets) ou individuel.

Le logiciel permet d'obtenir des informations telles que les durées d'états, les nombres d'occurrences d'évènements, les simultanités (durées et occurrences).



**Figure 16 : Exemple d'un graphe d'activité fourni par Actogram™ pour un sujet et une session**

Au delà des données purement quantitatives, Actogram™ permet également de réaliser des chroniques d'activité donnant une vue graphique de certains évènements plus faciles ou impossibles à interpréter uniquement avec le support des chiffres (Figure 16).

Ce format permet entre autre, de repérer des proximités temporelles d'évènements de manière visuelle (groupement de modalités, séquences d'erreurs, etc.).

### 3.6. *Méthode de traitement des verbalisations*

Immédiatement après les sessions d'interactions, nous avons procédé aux autoconfrontations individuelles de chaque sujet à l'aide des traces vidéos.

Les verbalisations ont ensuite fait l'objet de retranscriptions systématiques et exhaustives puis ont été rassemblées dans un fichier unique Excel. Lors de ces retranscriptions, nous avons également procédé à un premier niveau de codage à l'aide d'indicateurs pertinents pour une analyse ultérieure des données.

Le premier niveau de codage consiste à tagger les thématiques abordées par des mots clés correspondants aux focus de l'étude. Il pouvait s'agir de remarques concernant chaque modalité, l'application en général, son appropriation, les stratégies choisies, les tests menés, etc. Voici quelques exemples de mots clés retenus :

**Tableau 7 : exemples de mots clés utilisés pour le codage des verbalisations**

regard	vocale	tactile	g2d	audio	clavier	manuscrit	synthèse	monde	mobilité	charge	test	bruit
--------	--------	---------	-----	-------	---------	-----------	----------	-------	----------	--------	------	-------

Ce codage a ensuite permis de regrouper thématiquement toutes les verbalisations afin de dégager des sujets émergents récurrents propres à chaque focus ainsi que des problématiques transverses exprimées par les sujets.

Ce traitement permet ainsi d'alimenter l'analyse qualitative de ce travail, en particulier en abordant l'aspect subjectif des usages de l'application multimodale. A chaque fois que des données qualitatives sont disponibles, elles sont associées à la suite des analyses quantitatives présentées pour venir en enrichir le contenu.

## **Introduction des chapitres 3, 4, 5**

### **et 6**

L'ensemble de ces trois chapitres présente les résultats de l'étude empirique effectuée dans le cadre du laboratoire des usages « LUTIN » à la Cité de la Science et de l'Industrie de la Villette. Trois couples de sujets ont eu à réaliser un quiz durant trois sessions à l'aide d'un dispositif multimodal, multimédia et mobile.

La présentation des résultats s'articule autour de trois parties principales en commençant par les résultats concernant les usages de la multimodalité en général (Chapitre 3, Usages de la multimodalité) indépendamment de son contexte d'usage. L'objectif est de dégager d'éventuelles caractéristiques intrinsèques à chaque modalité permettant en partie d'expliquer les préférences individuelles de chaque utilisateur.

La deuxième partie (Chapitre 4, Facteurs contextuels influençant l'usage des modalités), au contraire, réintroduit les usages de la multimodalité dans un contexte naturel afin d'identifier d'éventuels facteurs contextuels pouvant venir influencer les usages. Il sera en particulier question d'étudier les effets du contexte sonore et social, ainsi que ceux du niveau d'activité en cours et enfin ceux de l'historique de l'interaction.

La troisième partie (Chapitre 5, Multimodalité, mobilité et ressources attentionnelles) traite de l'articulation entre une situation de mobilité et de ses conséquences sur une interaction multimodale. Cette partie

## Introduction des chapitres 3, 4, 5 et 6

commence donc par évaluer les effets réciproques des choix de modalités d'interaction sur la mobilité, et inversement, étudie les effets de la situation de mobilité sur le choix d'une modalité d'interaction. Nous aborderons ensuite les problématiques de la gestion des ressources attentionnelles en situation mobile, puis celle de l'influence de l'activité en cours, et nous finirons par le sujet des erreurs et de leur traitement (correction) afin d'évaluer leur impact sur la mobilité.

Enfin en ce qui concerne les résultats, une dernière partie abordera le sujet de la saisie de média (Chapitre 6, Saisie de médias) à la fois pour en décrire les caractéristiques, les usages et leurs conséquences sur la gestion des ressources attentionnelles en situation de mobilité.

## **Chapitre 3. Usages de la multimodalité**

## 1. Introduction

Ce chapitre traite des premiers résultats concernant les usages de la multimodalité envisagés, dans un premier temps, indépendamment de la situation ou du contexte d'usage.

Les résultats présentés ici s'appuient en premier lieu sur des données quantitatives mais sont également enrichis de données qualitatives dans la mesure du possible lorsque des verbalisations sont disponibles.

La grille de lecture utilisée dans ce premier temps de l'analyse s'efforce de donner des éléments de réponse aux questions suivantes :

***Quel est le poids respectif de chaque modalité dans l'utilisation du système par les sujets ?***

***Y'a-t-il des différences interindividuelles marquées ?***

***Constate-t-on une évolution entre la première et la dernière session expérimentale ?***

***Observe-t-on un début de spécialisation (association récurrente d'une modalité et d'une commande) ?***

***Comment les modalités s'enchaînent-elles au niveau local (commandes successives) et global (au cours d'une session) ?***

***Quel usage est-il fait de la multimodalité en situation d'erreur ?***

***De manière générale comment les sujets perçoivent-ils les différentes modalités ?***

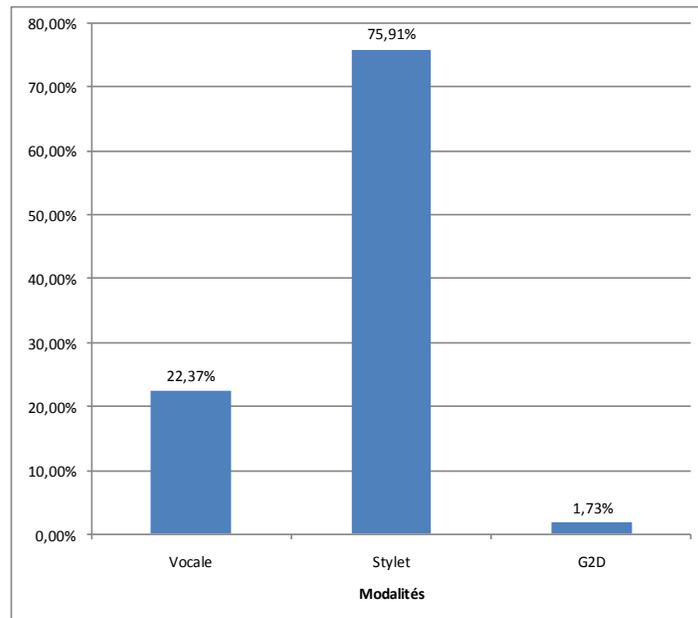
Ce chapitre est structuré en suivant cet enchaînement, et s'achève sur une synthèse rapide de l'ensemble des résultats obtenus dans cette partie.

## 2. Usage global des trois modalités équivalentes (stylet, vocal, g2d)

Cette première partie s'intéresse en particulier à la première question mentionnée ci-dessus :

### *Quel est le poids respectif de chaque modalité dans l'utilisation du système par les sujets ?*

La Figure 17 concerne les commandes totalement équivalentes pour la modalité vocale et tactile stylet, c'est à dire qui peuvent être effectuées indifféremment avec ces deux modalités. Cependant, le geste 2d n'a pu être implémenté que pour cinq commandes d'édition (nouveau message, couper, copier, coller, supprimer) et a rencontré de nombreuses contraintes techniques qui rendaient son utilisation relativement limitée. (L'usage du g2d sera étudié ultérieurement).



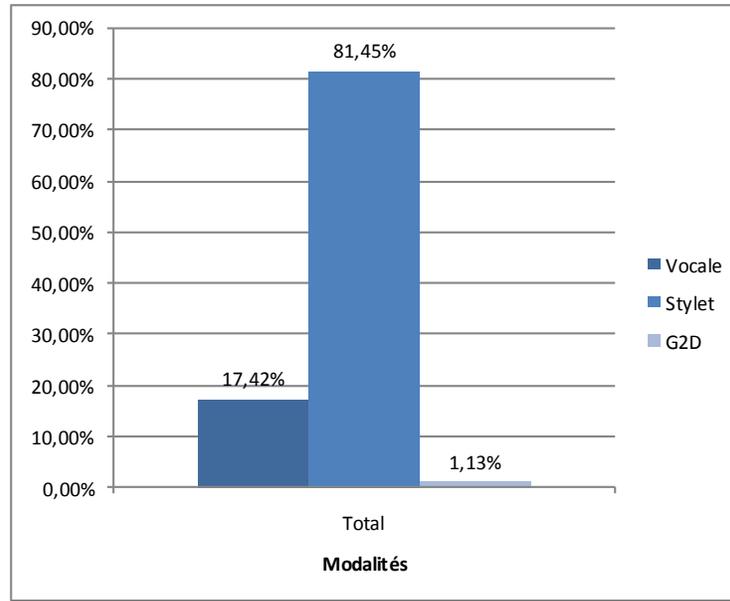
**Figure 17 Usage global des modalités toutes commandes, toutes sessions et tous sujets confondus (stylet, vocal, g2d)**

La Figure 17 indique un usage prédominant de la modalité tactile stylet (75,91%, n=3163 N=4167) ; la modalité vocale ne représente que 22,37% (n=932) des commandes émises par les sujets. On retrouve ici des résultats classiques (Oviatt, S.L., 1999) qui viennent contredire des positions « naïves » relatives à la primauté supposée du vocal dans les interfaces multimodales. La part du geste 2D est quant à elle quasi négligeable (1,73%, n=72).

### *2.1. Usage global des modalités après décompte des commandes de correction (stylet, vocal, g2d)*

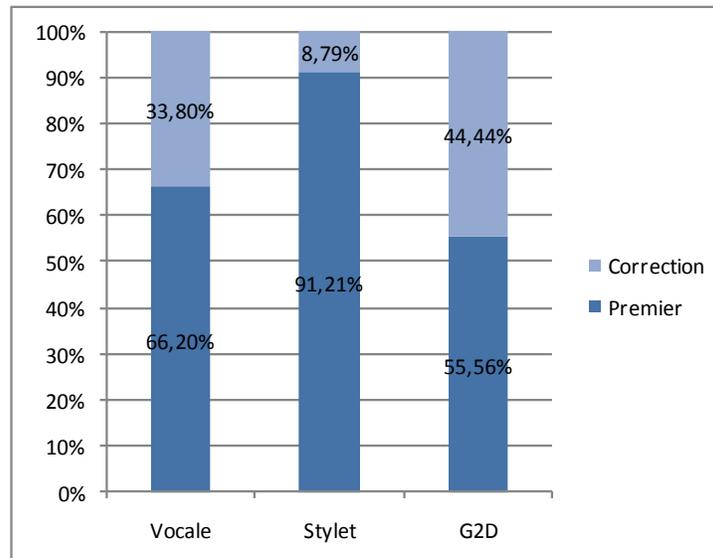
Une part non négligeable des commandes (N=625) correspond à des corrections d'erreurs consistant à la répétition de la même commande usant ou non de la même modalité. Nous distinguons donc ici les commandes dites « premières » constituant la première intention des sujets dans leur choix de modalité d'interaction, et les commandes dites de « correction ». Dans la Figure 18, ces dernières ont été soustraites du total.

## Usages de la multimodalité



**Figure 18 Usage global des modalités toutes commandes, toutes sessions et tous sujets confondus après décompte des commandes de correction**

Suite à ces corrections, les résultats indiquent une accentuation encore plus prononcée pour les commandes « premières » de la prédominance de la modalité tactile stylet (81,45%), et, parallèlement une diminution modérée de la modalité vocale (17,42%) et du Geste 2D (1,13%).



**Figure 19 Ratio des premières commandes et des commandes de correction**

La Figure 19 indique le ratio des premières commandes et des commandes de correction pour les trois modalités. Ainsi, 33,80% des commandes vocales correspondent à des répétitions de la même commande au cours de corrections. Le ratio est encore plus élevé en ce qui concerne le G2D dont 44,44% sont des corrections. Par contre, pour le stylet, on note des chiffres nettement moins élevés (8,79% de corrections).

Ces chiffres seront repris dans l'analyse détaillée des situations d'erreurs, mais ils indiquent déjà à un niveau global une faible robustesse de la modalité vocale et G2D dont les conséquences sur les choix d'usages des modalités sont à déterminer.

### 3. Usage des modalités par sujets

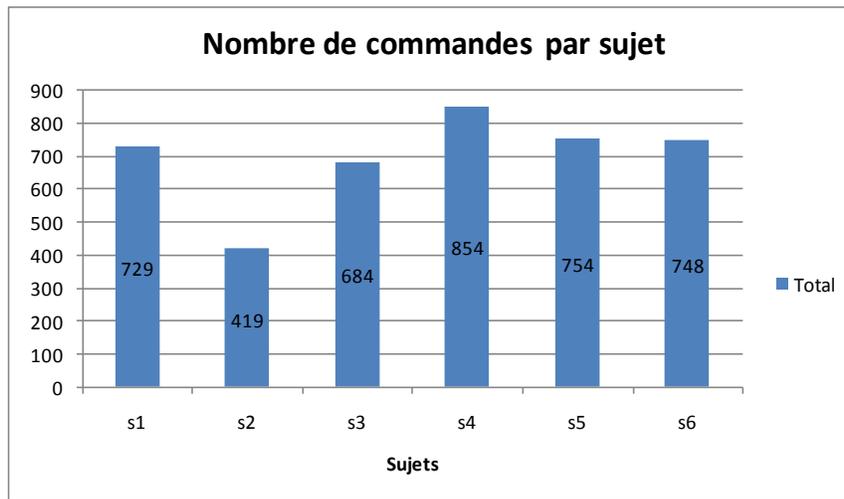
Dans cette partie, nous tenterons de répondre à la seconde question :

***Y'a-t-il des différences interindividuelles marquées dans l'usage des modalités ?***

Les résultats suivants ne tiennent plus compte des commandes concernant la modalité Geste 2D, celles-ci n'étant pas équivalentes pour l'ensemble des commandes étudiées dans la suite de l'analyse (N=4095).

**3.1. Nombre de commandes multimodales par sujets**

Dans un premier temps, nous allons regarder la quantité de commandes effectuées pour chaque sujet au cours des trois sessions.



**Figure 20 Nombre de commandes par sujet toutes sessions confondues**

La Figure 20 indique une assez bonne homogénéité dans le nombre de commandes multimodales effectuées par les sujets (la différence pour le sujet 2 étant due à l'absence de données pour une session entière). La moyenne du nombre de commandes multimodales au cours des trois sessions est proche de 700 occurrences. Ce résultat permet de faire des comparaisons entre sujets.

**3.2. Répartition de l'usage des modalités par sujets**

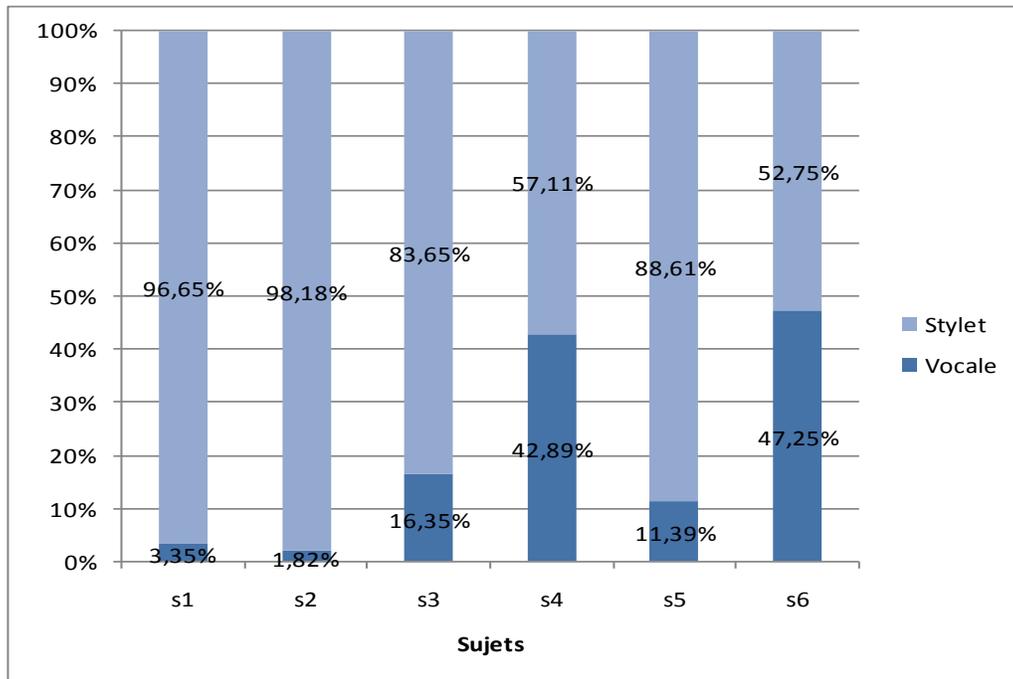
Les résultats précédents sur l'usage global des modalités indiquent une nette prédominance de la modalité tactile stylet sur le vocal.

***Peut-on extrapoler ces résultats sur chaque sujet ?***

***Ou bien observe-t-on des différences interindividuelles sur l'appropriation des modalités ?***

Pour répondre à ces interrogations, nous allons, dans cette partie, nous focaliser sur l'usage des modalités sujet par sujet.

## Usages de la multimodalité



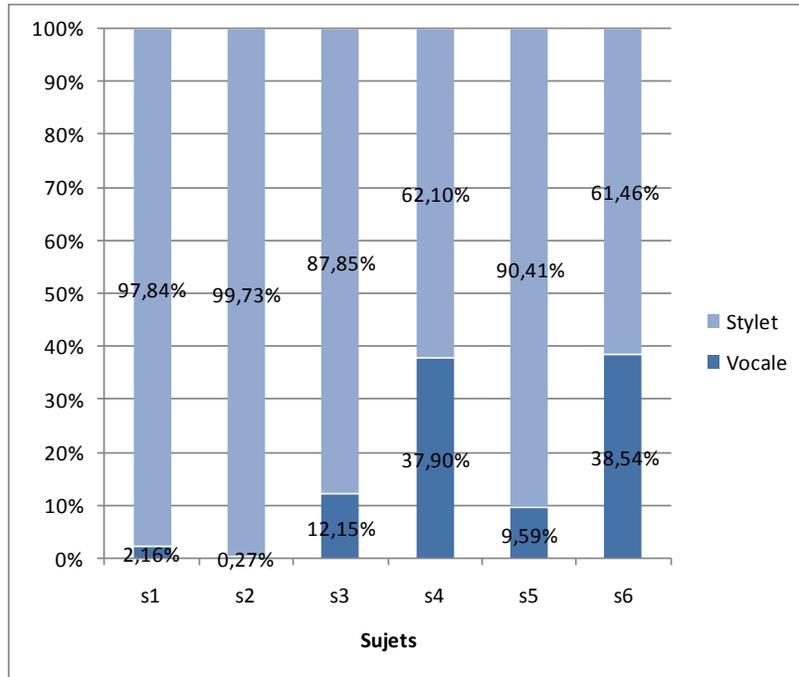
**Figure 21 Usage global des modalités par sujet, toutes commandes et toutes sessions confondues**

La Figure 21 indique une forte variabilité inter individuelle dans l'usage des modalités (de 3,35% de commandes vocales à 47,25%). On peut néanmoins distinguer trois groupes de sujets : Certains sujets (1 et 2) utilisent très peu la modalité vocale (en dessous des 4%). A l'inverse les sujets 4 et 6 l'ont beaucoup exploitée (47,25% pour le sujet 6 et 42,89% pour le sujet 4). Les sujets 3 et 5 occupent une position intermédiaire (avec respectivement 16,35% et 11,39% d'usage de la modalité vocale).

En ce qui concerne le geste 2D, seul le sujet 2 atteint 8,7% alors que son usage chez les autres sujets reste anecdotique.

La Figure 22 distingue les commandes « premières » et de corrections, permettant ainsi de différencier les premières « intentions » d'usage des modalités des sujets.

## Usages de la multimodalité



**Figure 22 Usage global des modalités par sujet, pour les commandes premières**

Là encore, si nous considérons uniquement les commandes « premières » (Figure 22), les résultats sont davantage marqués en faveur de la modalité tactile stylet, en particulier pour les sujets qui ont peu tendance à utiliser la modalité vocale (les sujets 1 et 2 passent respectivement de 3,35 à 2,1%, et de 1,82 à 0,27%). Ces valeurs indiquent que pour l'ensemble des sujets, une partie des effectifs des commandes vocales est due aux commandes de correction.

Dans la suite de l'étude, nous allons nous pencher sur les modes d'apprentissages et d'usage des modalités au cours des sessions et enfin, tenter de dégager d'éventuelles variations inter et intra-individuelles.

## 4. Evolution des usages par session

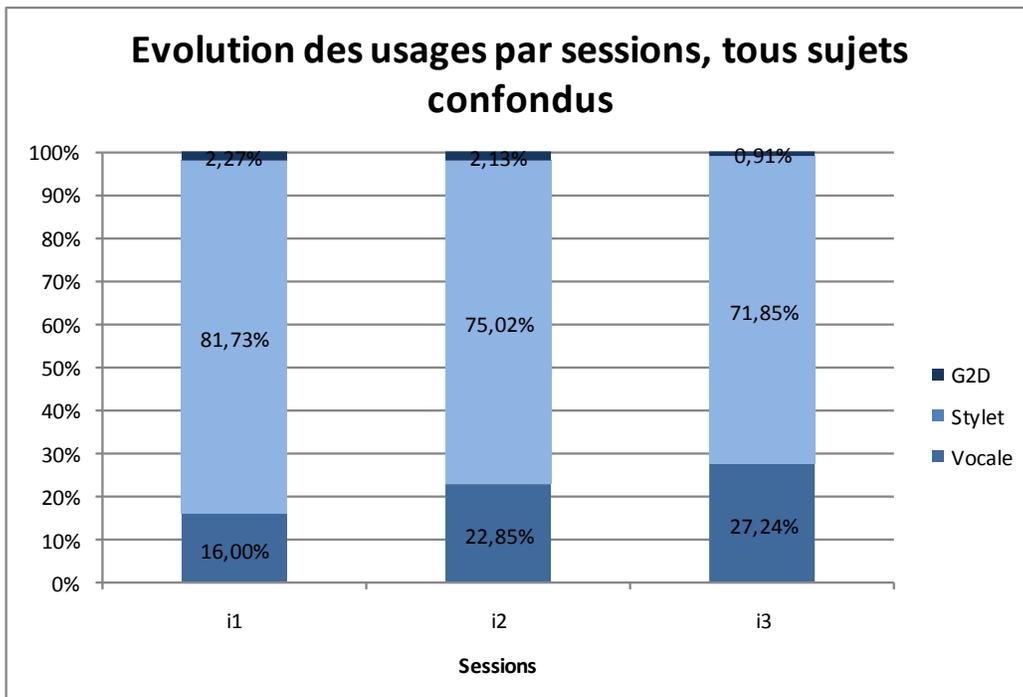
Après avoir traité des résultats globaux, c'est à dire ne tenant pas compte de l'évolution des usages au cours des sessions, nous pouvons aborder les questions suivantes :

**Constata-t-on une évolution des usages entre la première et la dernière session expérimentale ?**

**Constata-t-on des différences intra-individuelles dans l'évolution des usages pour chaque sujet, entre la première et la dernière session ?**

### 4.1. Evolution globale, tous sujets confondus

Pour répondre en partie à la première question, nous commençons par regarder à un niveau global (tous sujets confondus) l'évolution des usages entre la première et la troisième session d'interaction.



**Figure 23 Evolution des usages par sessions, tous sujets confondus**

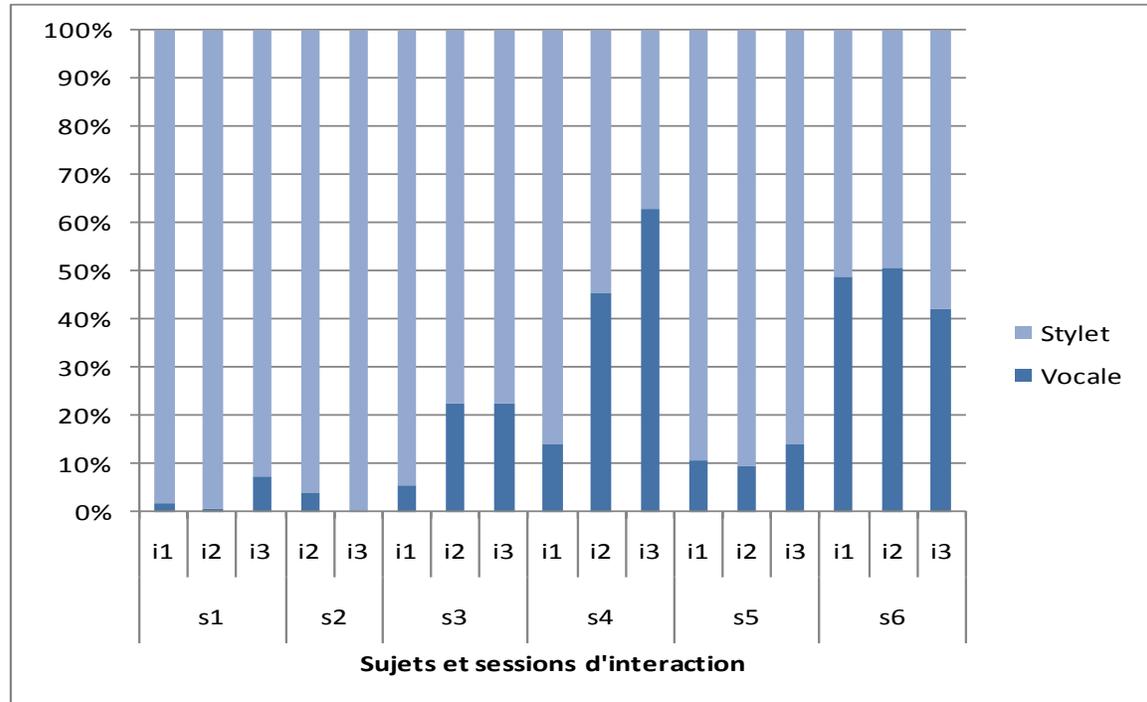
La Figure 23 montre une augmentation sensible de l'utilisation de la modalité vocale qui passe de 16 à 27,24% , ceci, au détriment de la modalité tactile, qui elle, passe de 81,73% à 71,85%. La modalité G2D passe de 2,27% à 0,01%.

A un niveau global, il y a donc une adoption de la modalité vocale plus importante que pour les autres modalités.

### 4.2. Evolution intra-individuelle

Au-delà des variations inter-individuelles globales, il est intéressant d'étudier l'évolution de l'appropriation des modalités pour chaque sujet, session après session.

## Usages de la multimodalité



**Figure 24 Evolution des usages par session et par sujet**

Les résultats obtenus ici (Figure 24), semblent différer des résultats précédents, car la répartition et l'évolution de l'usage des modalités entre la première et la dernière session sont loin d'être homogènes si l'on considère les sujets individuellement.

Alors que la Figure 23 indique une évolution globale croissante de la modalité vocale, nous observons ici, au contraire, de grandes différences inter et intra-individuelles. En effet, la Figure 24, permet de distinguer 2 catégories de sujets sur les trois sessions :

- Les sujets pour lesquels les choix de modalité sont quasiment acquis dès la première séance (sujets 1, 2, 5 et 6),
- Et les sujets pour lesquels on observe une évolution importante entre la première et la dernière session (sujets 3 et 4).

Il s'agit essentiellement d'un accroissement important de l'usage de la modalité vocale au détriment de la modalité tactile « stylet » (sauf pour le sujet 2 qui au contraire, va abandonner totalement la modalité vocale).

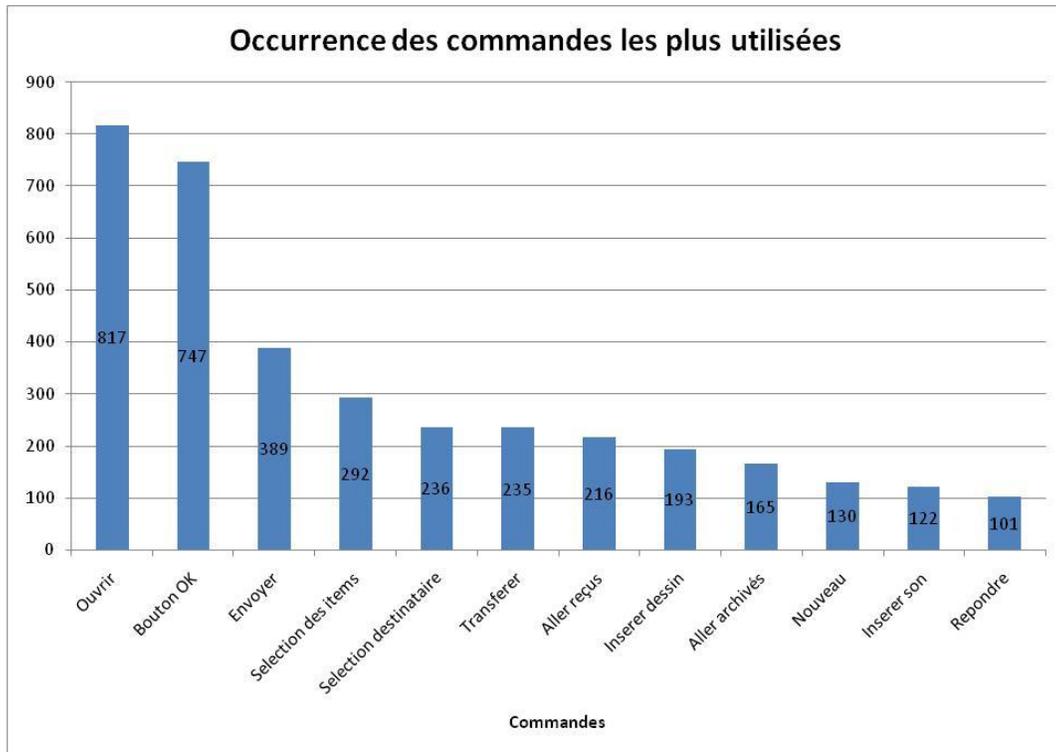
Si l'on considère par exemple le sujet 4, on note une inversion totale de la modalité dominante : l'usage de la modalité tactile « stylet » passant de 83,7% à la première interaction à 36,9% à la dernière session. Corrélativement, la modalité vocale passe de 13,6 à 63,1%.

Une étude précédente (Calvet, 2001) montre une phase de « test informel » de l'usage des modalités et de l'application qui se traduit par une répartition relativement homogène des modalités dans la première session. Vient ensuite une stabilisation des usages des modalités après environ trois sessions d'apprentissage (sur un total de 7 sessions), les dernières sessions ne faisant que confirmer les préférences individuelles.

Dans cette étude, le faible nombre de sessions (3) ne permet pas d'observer une stabilisation marquée des usages et il est donc impossible d'indiquer si les usages de la session 3 correspondent à une adoption de modalité relativement stabilisée par les sujets.

## 5. Usage des commandes

Avant d'étudier l'apparition d'éventuelles spécialisations sur certaines commandes, nous commençons par en étudier leur occurrence. Cette partie comptabilise donc les commandes (équivalentes) les plus utilisées par les sujets (au delà de 80 occurrences), ce qui permet à un niveau quantitatif de rendre compte de l'activité générale des sujets et de les classer par ordre d'occurrence.



**Figure 25 Les commandes les plus utilisées au niveau global**

Les commandes les plus largement utilisées sont « ouvrir » et « fermer » (bouton ok) avec respectivement 817 et 747 occurrences (N= 3643). Il s'agit de commandes de consultation des messages et il est logique qu'elles soient du même ordre de grandeur puisqu'elles sont des passages obligés de la tâche. Vient ensuite la commande « envoyer » qui atteint quasiment 400 occurrences (389), elle est également une commande sans alternative pour envoyer un message édité. La suite des commandes oscille entre 292 occurrences (« sélection du message » (292)) et 101 (« répondre »).

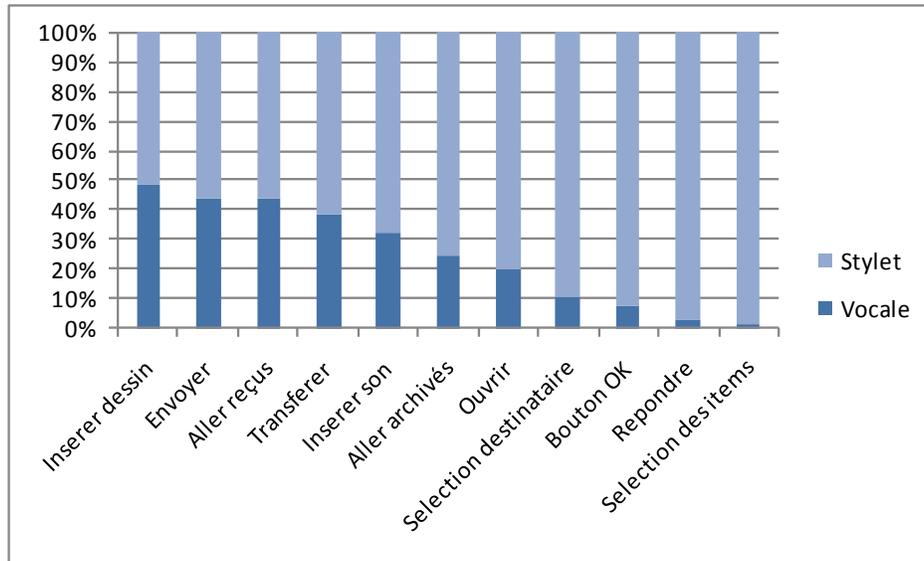
### 5.1. Usage spécialisé de modalités par commandes au niveau global

Nous avons vu qu'il y avait de fortes variations inter et intra-individuelles, et il convient maintenant de déterminer si celles-ci se traduisent par des spécialisations éventuellement communes à tous les sujets (niveau global).

#### **Observe-t-on un début de spécialisation (association récurrente d'une modalité et d'une commande) ?**

Nous considérons qu'il y a spécialisation d'une modalité lorsqu'elle est utilisée à 80% pour une commande donnée. Par soucis de clarté et pour simplifier la lecture des tableaux, nous ne conserverons dans les analyses suivantes que les commandes les plus utilisées (au delà de 80 occurrences).

## Usages de la multimodalité



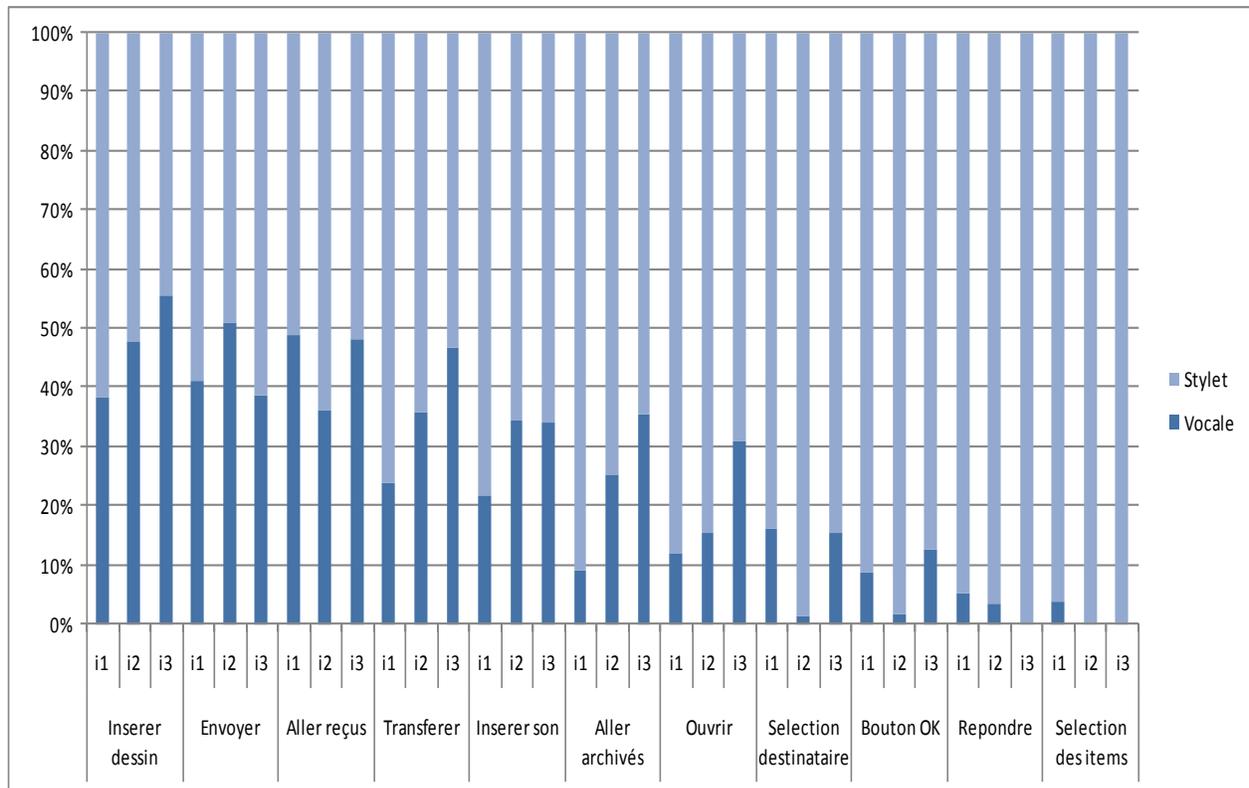
**Figure 26 Usage global des modalités par commandes tous sujets et sessions confondus**

Les résultats mettent en évidence qu'un nombre important de commandes sont effectuées uniquement avec la modalité stylet (Figure 26 ). En effet, on peut considérer qu'à un niveau global, il y a une spécialisation de la modalité tactile stylet pour les commandes « ouvrir », « sélectionner un destinataire », « Bouton Ok » (fermer), « répondre » et enfin « sélectionner un item ». Pour l'ensemble des sujets on n'observe pas de spécialisation intégrale de la modalité vocale pour une commande donnée. Il existe, cependant, une certaine tendance à la préférence vocale pour quelques commandes telles que « insérer dessin », « envoyer », « aller à reçus », « transférer », etc. qui avoisinent tout de même les 40%.

### **5.2. Evolution de l'usage global de modalité par commandes**

Ces tendances aux préférences globales doivent être examinées de plus près afin de déterminer si elles apparaissent au cours du temps ou si elles sont acquises dès la première session.

***Observe-t-on, au niveau global, une évolution des spécialisations au cours des sessions ?***



**Figure 27 Evolution de l'usage global des modalités par commandes tous sujets confondus**

Les résultats n'indiquent aucune tendance globale et forte à la spécialisation session après session. Cependant, nous pouvons tout de même observer une augmentation de l'usage de la modalité vocale pour un certain nombre de commandes telles que « insérer dessin », « transférer », « insérer son », « aller archivés », « ouvrir » indiquant une adoption de cette modalité pour ces commandes, mais cela n'atteignant jamais la spécialisation. Il s'agira donc d'usages préférentiels pouvant être majoritaires dans un cas (« Insérer dessin ») à la dernière session.

Il est également intéressant de noter que certaines commandes ont de forts taux de commandes vocales dès la première session. Celles-ci ont tendance à rester stables pour les commandes « envoyer », « aller reçus », ou à augmenter pour la commande « insérer dessin ».

En revanche, la commande « répondre » qui dès la première session est spécialisée dans la modalité tactile stylect voit se phénomène s'accroître jusqu'à la quasi disparition de la modalité vocale.

Il est nécessaire d'aller voir au niveau individuel comment se répercutent ces tendances et s'il ne s'agit pas simplement d'un effet de moyenne dû à des comportements isolés.

### 5.3. Spécialisation intra-individuelle et évolution des usages

Nous avons vu dans la partie 3 que les sujets ont un usage différent (varié, propre) des modalités proposées. Si l'on introduit la notion de commandes, de nouvelles questions émergent alors :

**Les différences inter-individuelles d'usage des modalités se répercutent-elles sur toutes les commandes ?**

**Observe-t-on, au niveau individuel, un début de spécialisation (association récurrente d'une modalité et d'une commande) ?**

## Usages de la multimodalité

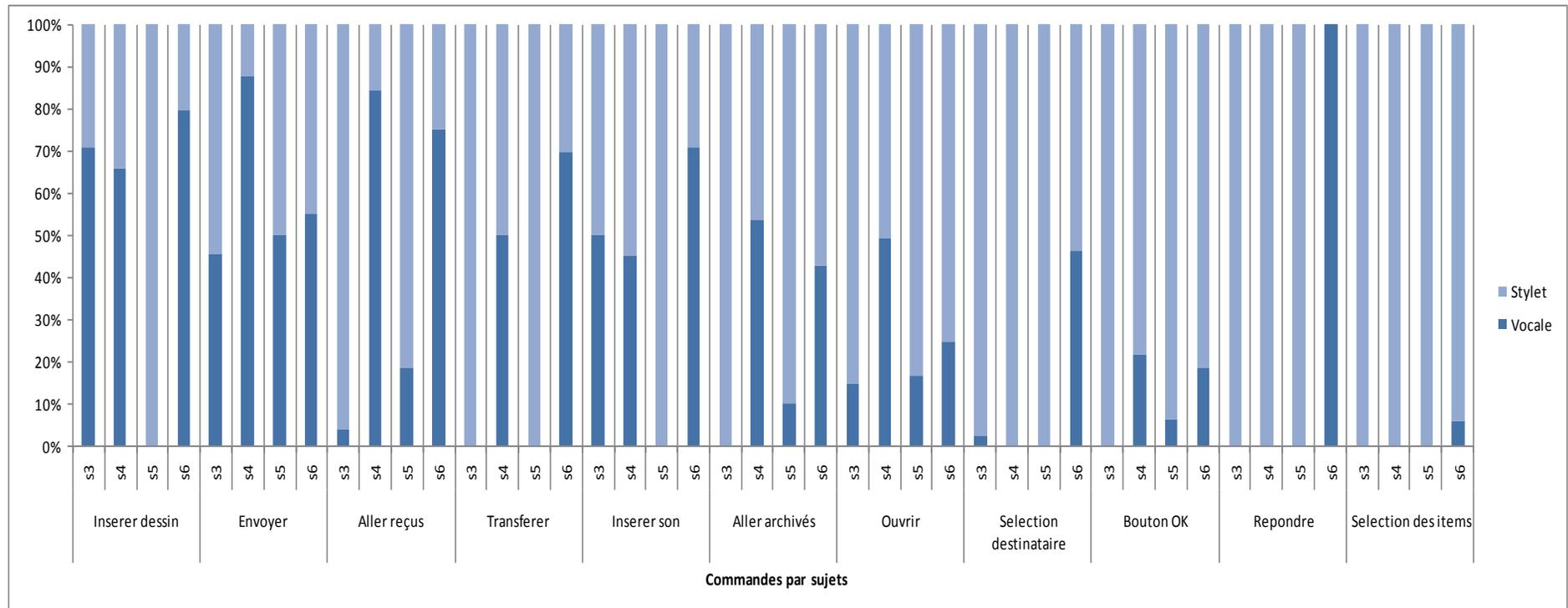


Figure 28 Différences inter-individuelles dans l'usage des modalités par commandes toutes sessions confondues

Si l'on considère les résultats sujet par sujet on constate que le phénomène de spécialisation est davantage marqué et qu'il y a de grandes différences à la fois inter et intra individuelles dans l'adoption des modalités et dans certaines spécialisations qui en découlent.

Ainsi, pour répondre à la première interrogation, nous voyons que la spécialisation observée au niveau global, d'une modalité pour une commande donnée, ne se répercute pas au niveau individuel. En effet, il peut y avoir de grandes disparités d'usages de modalités pour une commande, par exemple, pour la commande « transférer », les sujets 3 et 5 n'ont absolument pas eu recours à la modalité vocale alors que le sujet 6 atteint des taux proches de 70% pour cette commande.

En ce qui concerne la deuxième interrogation, nous constatons de nombreuses tendances à la spécialisation pour l'une ou l'autre des modalités, cependant, il n'y a pas de spécialisation totale pour la modalité vocale (usage exclusif d'une modalité pour une commande) sauf pour le sujet 6 et la commande « répondre ».

En revanche, il existe un certain nombre de spécialisations pour la modalité tactile stylet, et ceci, parfois de manière partagée par plusieurs sujets (« Sélection des items », « Répondre », « Sélection destinataire »). Pour ces commandes, l'analyse au niveau global reste donc pertinente.

Ce résultat quantitatif n'a pas pu trouver d'explication au niveau qualitatif via les entretiens d'autoconfrontations ou via une analyse fine des propriétés des commandes. Le fait que les sujets ne se spécialisent pas sur les mêmes commandes et les mêmes modalités semble plutôt indiquer des facteurs de choix non liés aux caractéristiques intrinsèques des commandes pour lesquelles une modalité serait éventuellement plus appropriée, et donc l'explication pourrait plutôt tendre vers des préférences individuelles (internes) non partagées entre les sujets.

## 6. Transitions et enchaînement des modalités

A travers cette partie, nous souhaitons aborder deux types de questionnements liés aux schémas d'usage des modalités :

***L'usage d'une modalité se fait-il en séquences d'une même modalité ou bien y a t il alternance des modalités ?***

***L'usage d'une modalité se fait-il de manière regroupée dans le temps ?***

Nous nous intéressons donc ici à l'enchaînement des modalités entre elles, et ceci selon deux points de vue :

- **L'enchaînement local (immédiat) des modalités** : nous analysons les enchaînements immédiatement successifs des modalités entre-elles afin de dégager d'éventuels cycles d'usage d'une modalité. Il s'agit de vérifier si l'on observe un phénomène local, à très court terme, d'activation de modalité.
- **L'enchaînement global des modalités** : à l'aide de chroniques d'usages des modalités, nous recherchons d'éventuels regroupements temporels (grappes) de modalités, sans pour autant que l'usage d'une même modalité soit immédiatement successif. Il s'agit ici, de vérifier si l'on observe un phénomène plus global d'activation de modalité par phases de la session d'interaction.

### 6.1. Enchaînement local des modalités

Cette partie renvoie au questionnement sur d'éventuels schémas d'usage des modalités en termes de routines (heuristiques), et plus particulièrement ici, en termes d'enchaînement de séquence d'usage d'une même modalité (économie cognitive). Les questions sous-jacentes sont donc les suivantes :

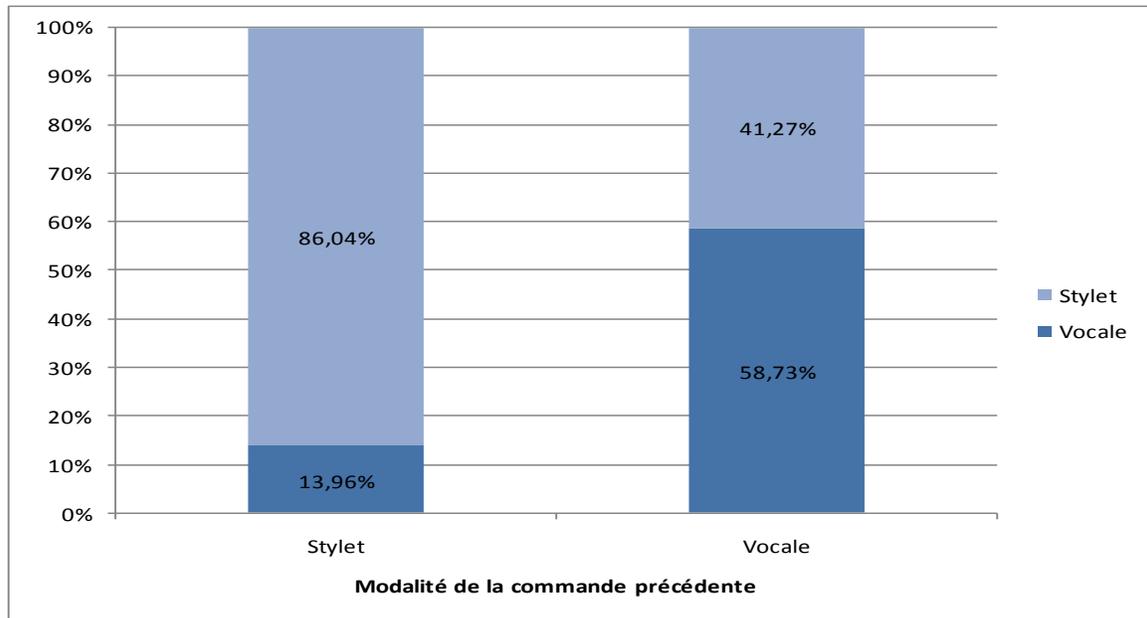
***Les sujets ont-ils tendance à faire usage de la même modalité de manière séquentielle ?***

***Si oui, quelles en sont les raisons (économie cognitive, routines) ?***

#### 6.1.1. Changements de modalité pour les commandes multimodales et unimodales

Dans les résultats présentés précédemment, seules les commandes « équivalentes » (qui peuvent être faites avec toutes les modalités) étaient prises en compte puisqu'il s'agissait de laisser le choix à l'utilisateur. Ici, nous nous intéressons à l'enchaînement immédiat des modalités et il nous faut donc réintroduire dans les effectifs toutes les autres commandes non équivalentes qui auraient pu s'intercaler entre les commandes équivalentes (les défilements, les insertions de média, etc.).

## Usages de la multimodalité

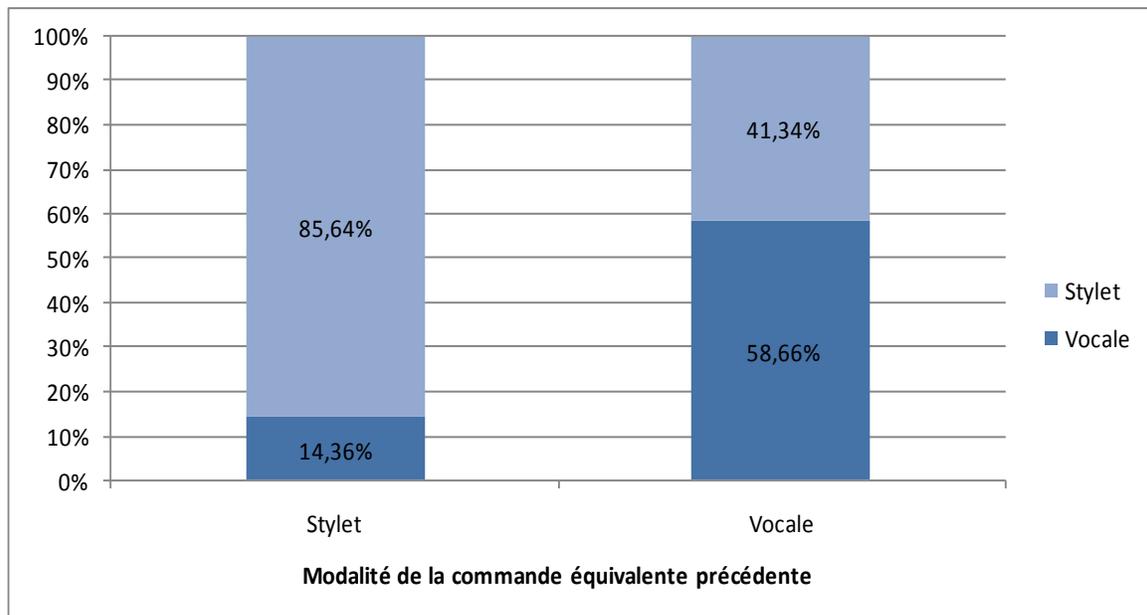


**Figure 29 Transition des modalités pour toutes les commandes (incluant les commandes non équivalentes)**

La Figure 29 montre à un niveau global, un effet fort de la modalité précédente sur le choix de la modalité suivante. Les commandes qui ont été effectuées au stylet, sont suivies à plus de 86% de commandes de la même modalité et donc seulement de 13,96% de commandes vocales.

De même, les commandes vocales sont, elles aussi, suivies préférentiellement de commandes de la même modalité mais ceci dans une moindre mesure que pour les commandes au stylet (58,73%).

La figure suivante, à la différence de la figure précédente, inclut uniquement les commandes équivalentes (m) qui peuvent être exécutées indifféremment à l'aide de la modalité vocale ou tactile.



**Figure 30 Transition entre la modalité précédente et la modalité suivante**

Les résultats diffèrent peu des résultats précédents, en effet, la Figure 30 montre à un niveau global que la modalité tactile stylet se fait suite massivement (plus de 85%). La modalité vocale quant à elle, est suivie à plus de 58% par elle-même. Ceci indique que concernant les deux modalités, il y a une tendance globale à l'enchaînement des mêmes modalités. (Dans une moindre mesure pour le vocal).

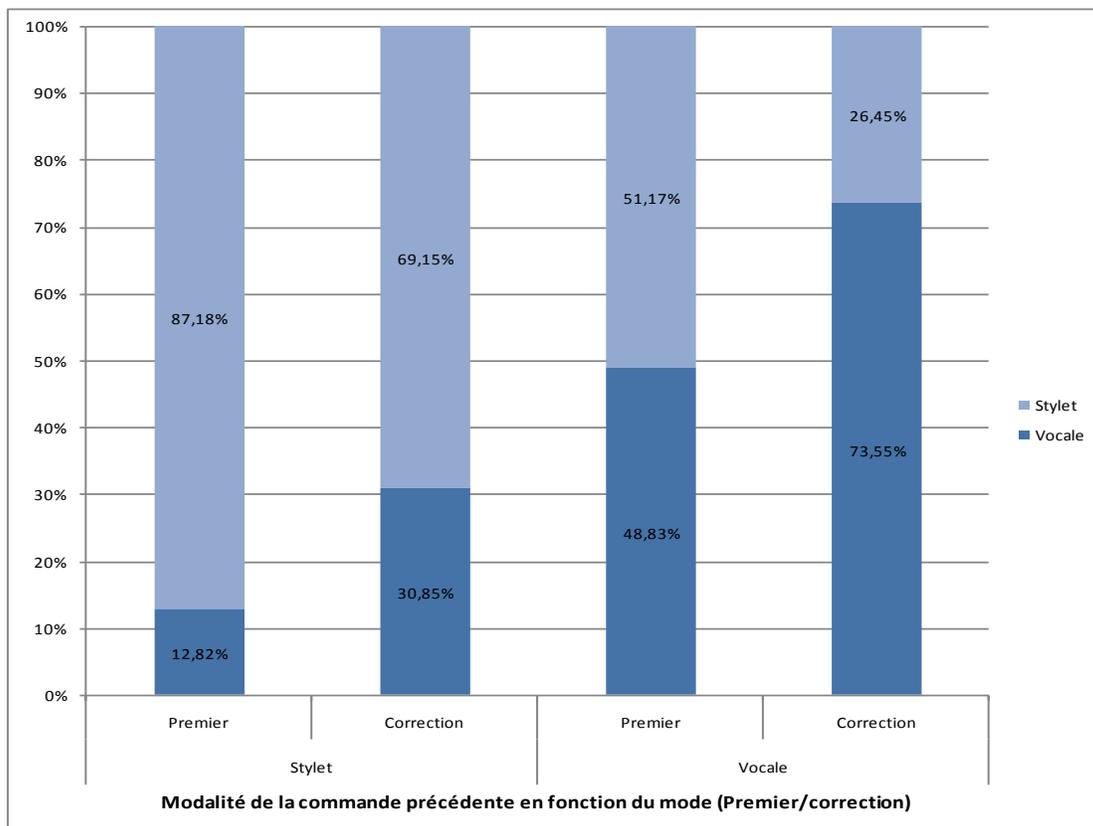
#### 6.1.2. Changements de modalité en correction ou non

Nous regardons ici l'influence du mode (première ou correction) de la commande sur le choix de la modalité afin de répondre aux questions suivantes :

***Y a t'il des usages différents des modalités en situation d'erreur ?***

***La commande première et les commandes de correction ont-elles tendance à être les mêmes ?***

Cela nous permettra également de savoir si les sujets ont tendances ou non à changer de modalité pour effectuer des corrections d'erreur.



**Figure 31 Transition entre la modalité précédente et la modalité suivante en commande première ou correction**

La Figure 31 indique (comparativement à la figure précédente (Figure 30) un renforcement de la modalité vocale lorsqu'il s'agit de corriger une erreur (mode correction) et ceci que la modalité précédente aie été tactile ou vocale.

En effet, pour les commandes précédentes passées au stylet, les valeurs passent de 12,82% pour les commandes premières à 30,85% pour les corrections.

Pour les commandes précédentes passées avec la modalité vocale, les valeurs passent de 48,83% pour les commandes premières à 73,55% pour les corrections.

Donc, que les commandes précédentes aient été effectuées en vocal ou stilet, le mode correction augmente fortement la tendance d'une correction en modalité vocale.

Et donc à l'inverse, on observe une forte baisse de la modalité vocale pour les commandes premières. En effet, alors que les chiffres globaux de la Figure 30 avoisinaient les 14% de commandes vocales pour les commandes premières au stilet, les résultats ici sont légèrement inférieurs pour les commandes premières (12,82%).

De même pour les commandes premières vocales qui étaient au-dessus des 58% dans la Figure 30, les résultats indiquent ici une baisse jusqu'à 48,83%.

Ceci indique que les commandes premières, qui correspondent à la première intention de modalité des sujets, ont tendance à avoir des taux de modalité vocale encore plus faibles qu'au niveau global, et à l'inverse, les sujets se trouvant dans une situation d'erreur ont tendance à essayer la modalité vocale pour corriger la situation.

### 6.2. Enchaînement global des modalités

Les résultats obtenus dans la partie précédente traitaient de la nature des enchaînements locaux (de commandes consécutives) des modalités entre elles, il s'agit maintenant de s'intéresser à l'aspect temporel d'un éventuel regroupement des modalités au cours de la session et ainsi de vérifier si l'on observe des « grappes » d'usage de la même modalité.

Pour cela, il est impératif de raffiner les analyses à un niveau individuel et session par session. Pour chaque sujet et chaque session, nous présentons un graphe constituant une chronique temporelle des usages des deux modalités retenues (tactile stilet et vocale) afin de repérer visuellement les groupements de modalité.

Dans les figures suivantes (Figure 32 à Figure 37), chaque symbole représente une commande (triangle rouge en haut pour la modalité vocale, losange bleu en bas pour le stilet) avec en abscisses la ligne de temps.

#### 6.2.1. Sujet 1

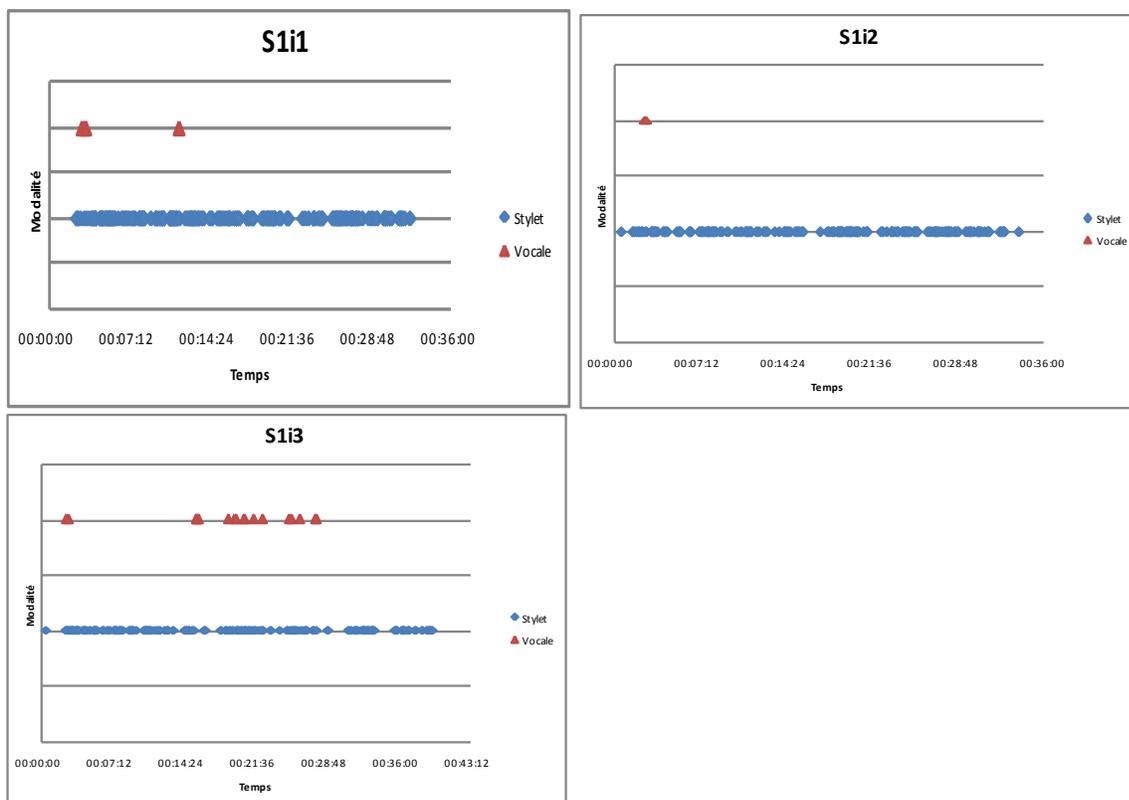
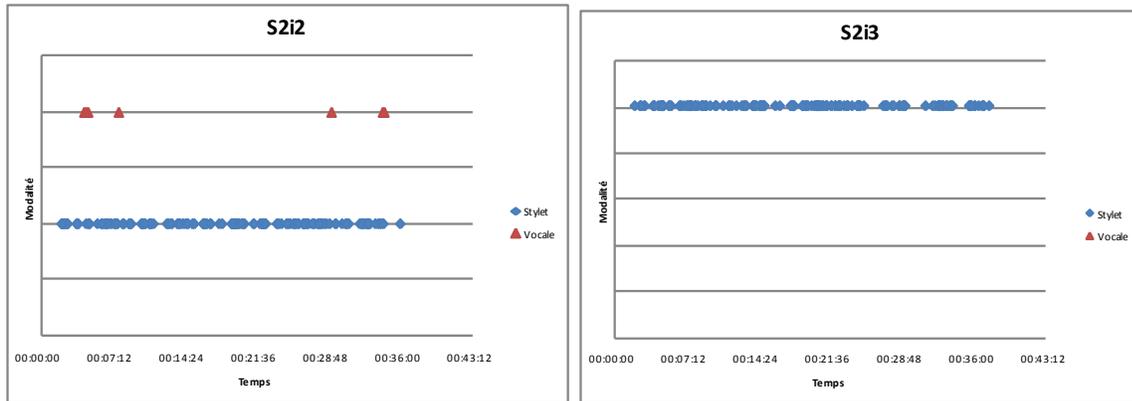


Figure 32 : Enchaînement temporel global des modalités pour le sujet 1 et pour les trois sessions

Pour le sujet 1, on observe lors de la première et deuxième session quelques tentatives groupées d'usage de la modalité vocale en début d'expérimentation, puis un abandon définitif de cette modalité. La session 3 est identique sauf qu'une situation particulière a obligé le sujet à avoir recours à la modalité vocale (en effet un bug de l'application rendait le tactile stylet inopérant).

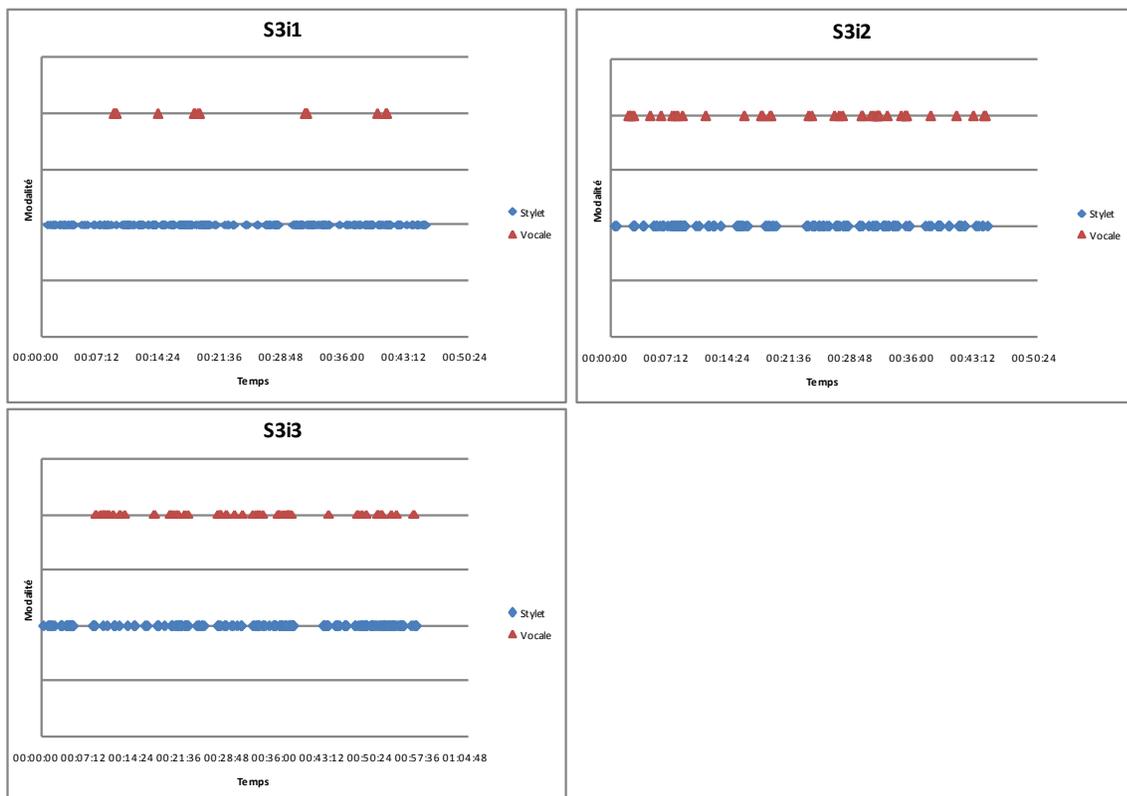
6.2.2. Sujet 2



**Figure 33 : Enchaînement temporel global des modalités pour le sujet 2 et pour les deux sessions disponibles**

Le sujet 2, lui aussi procède à quelques tentatives groupées d'usage de la modalité vocale, en début et fin de la session 2, puis l'abandonne totalement à la session 3.

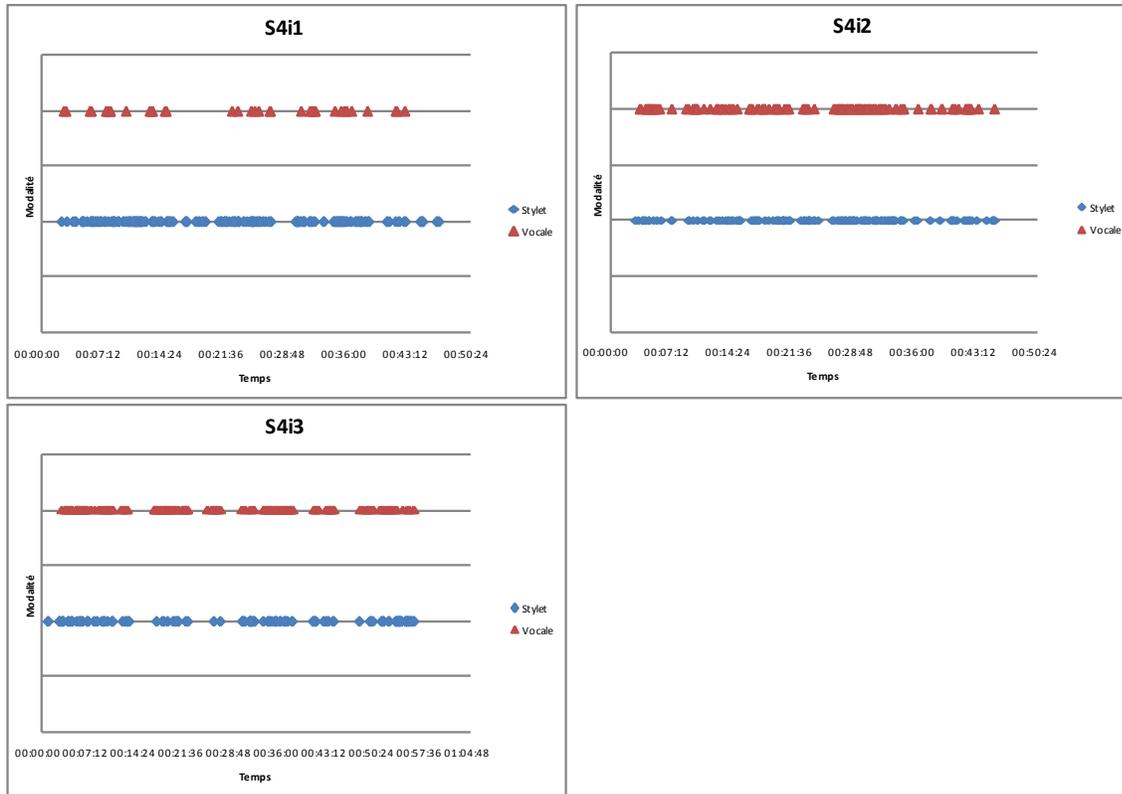
6.2.3. Sujet 3



**Figure 34 : Enchaînement temporel global des modalités pour le sujet 3 et pour les trois sessions**

Lors de la session 1, le sujet 3 procède à quelques tentatives groupées d'usages de la modalité vocale réparties dans l'ensemble de la session. Ces tentatives constituent dans un premier temps des îlots de séquences de quelques commandes relativement isolés les uns des autres, puis dans les sessions 2 et 3, ceux-ci deviennent plus fréquents et se répartissent alors tout au long de la session, en fonction des phases d'interaction.

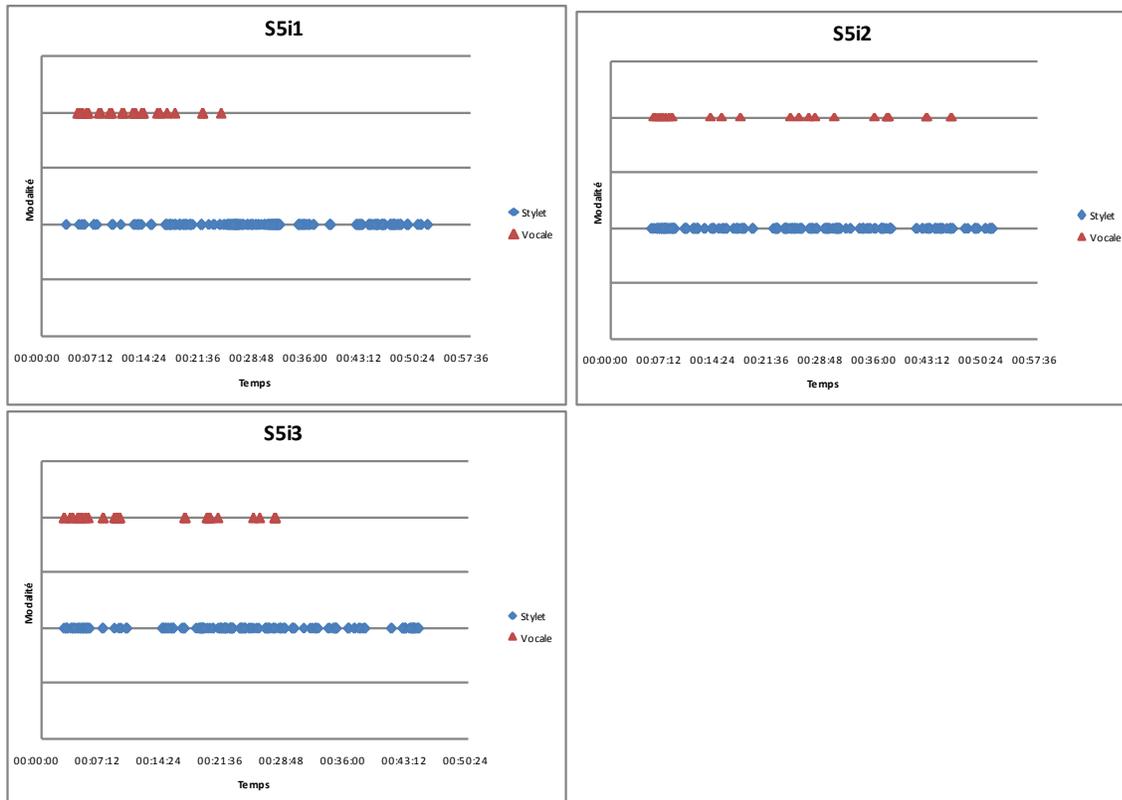
6.2.4. Sujet 4



**Figure 35 : Enchaînement temporel global des modalités pour le sujet 4 et pour les trois sessions**

Il en va de même pour le sujet 4, qui dès la première session a déjà plus tendance à utiliser la modalité vocale tout au long de la session mais avec encore quelques îlots identifiables et espacés dans le temps. Les sessions suivantes montrent un usage régulier et constant de cette modalité qui devient majoritaire en venant remplacer progressivement l'usage de la modalité tactile, et dont l'usage vient former à son tour des grappes plus facilement identifiables et isolées.

6.2.5. Sujet 5



**Figure 36 : Enchaînement temporel global des modalités pour le sujet 5 et pour les trois sessions**

Le sujet 5, montre en début de session 1, une forte propension à l'usage majoritaire de la modalité vocale, puis un abandon total jusqu'à la fin de la session. Puis, son usage forme des grappes (îlots) répartis tout au long de la deuxième et troisième session, avec cependant à nouveau un abandon de la modalité vocale aux deux tiers de la session 3.

6.2.6. Sujet 6

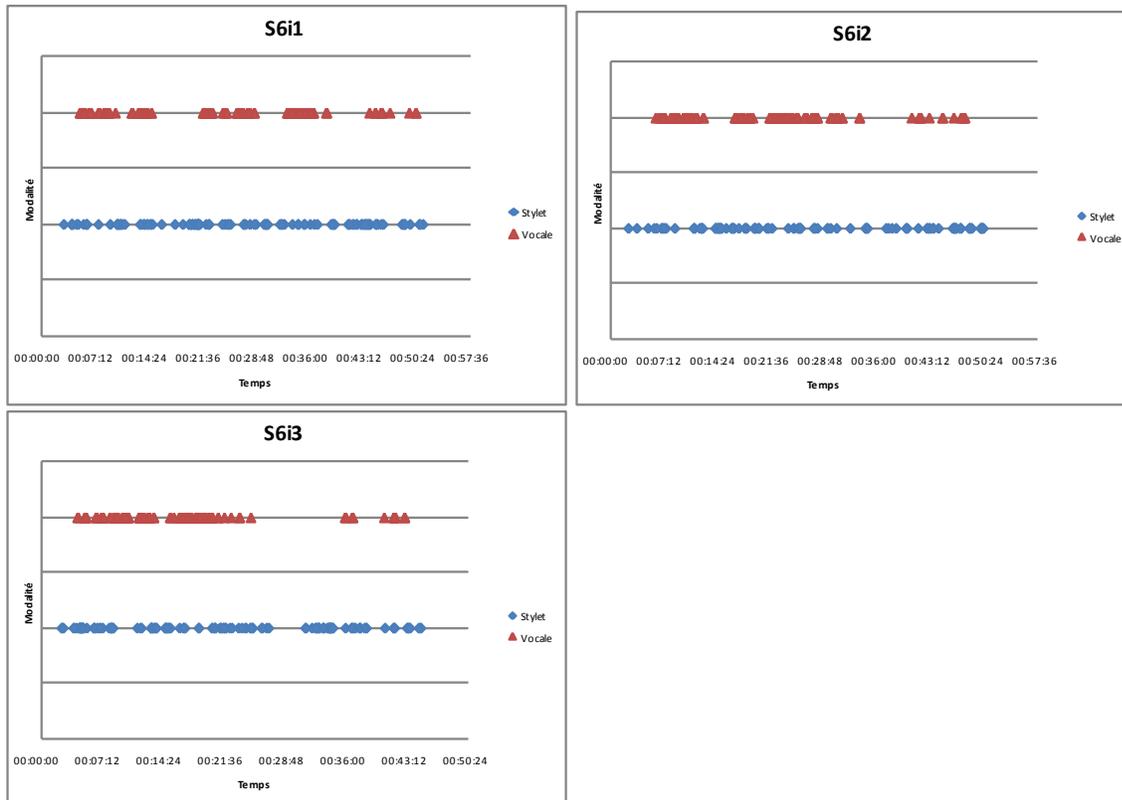


Figure 37 : Enchaînement temporel global des modalités pour le sujet 6 et pour les trois sessions

Le sujet 6 présente également des « îlots » assez longs d’usage de la modalité vocale se répartissant tout le long des sessions. Mais il présente également des interruptions d’usage de cette modalité, pouvant être assez longues dans le cas de la session 3.

6.2.7. Synthèse des résultats sur les transitions au niveau global

De manière générale, et en cohérence avec les résultats sur les enchaînements locaux de modalité, nous observons une tendance nette au regroupement des usages d’une modalité en « îlots » parfois isolés, en particulier pour les premières sessions.

En ce qui concerne les sujets utilisant très peu la modalité vocale (sujets 1 et 2), ils en font systématiquement usage de manière regroupée en îlots épars de quelques commandes successives. Ces îlots s’observent essentiellement en début de session et dans certains cas particuliers en cours de session pour les cas de correction d’erreur (voir pour une analyse plus détaillée : Correction des erreurs).

Pour les sujets qui vont adopter la modalité vocale, les premières sessions d’interaction sont également ponctuées par des îlots plus réguliers d’usage de la modalité vocale, puis leur répartition devient plus homogène et régulière au cours des sessions suivantes. Pour certains sujets (sujet 6), nous observons tout de même une tendance plus forte à l’usage en séquence de la modalité vocale.

Les explications de ces résultats pourront être explorées en partie à l’aide des verbalisations (voir la partie ci-après et Verbalisations concernant les préférences individuelles) et d’une analyse fine de chaque cas.

6.3. Verbalisations concernant les séquences d’usage d’une modalité

Les verbalisations recueillies concernant l’enchaînement des modalités permettent d’étayer quelques pistes expliquant l’adoption ou le rejet de telle ou telle modalité. Plusieurs thématiques ont été abordées par les sujets et sont regroupées sous les deux thèmes suivants : Activation locale et globale de modalités, continuité et fluidité de l’interaction.

6.3.1. Activation locale et globale de modalité (sujets 1, 2, 3, 6)

La notion d'activation d'une modalité a souvent été mentionnée par les sujets pour justifier des enchaînements d'une même modalité en séquence.

*S1 : i2 : Je préfère vraiment rester, si je commence avec le stylet, je fais tout avec le stylet. Et si je commence à faire euh, même les commandes vocales, j'sais pas, il faudrait que je puisse tout faire en vocal*

Dans cet exemple, le sujet 1 indique que lorsqu'il a commencé à utiliser une modalité il a tendance à la conserver pour les commandes suivantes. Dans le cas de ce sujet, du fait de la non-équivalence totale des commandes, l'impossibilité de pouvoir faire la totalité des commandes à l'aide de la modalité vocale a contribué à son rejet.

*Expé : là tu choisis d'envoyer au stylet*

*S3i2 : Là il me dit de sélectionner le destinataire, vu que j'étais déjà dans le truc (modalité tactile), j'ai sélectionné (tactile) et en même temps j'ai envoyé*

*S6 i2 : Ouais, là j'envoie au stylet au lieu de repasser au mode vocal, je finis tout au stylet quoi*

Dans ces deux exemples, les sujets illustrent à nouveau l'activation d'une modalité par une commande précédente de la même modalité. Pour ces deux sujets, leur tendance est de sélectionner le destinataire et d'envoyer le message avec la modalité vocale, et cependant, dans ces deux cas, ils ont utilisé la modalité tactile du fait de la proximité de la commande précédente tactile. Pour les deux sujets, cette suite de commandes est perçue comme unitaire.

Mais au-delà de l'activation locale d'une modalité, nous relevons également des verbalisations traitant de phénomènes à plus long terme.

*Expé : tu fais envoyer en vocal alors que tu faisais du tactile avant ?*

*S3 i3 : Ben ouais, j'étais dans le move du vocal, donc je me suis dit bon allez. Tu sais généralement, quand tu es dans une action, ça se fait automatiquement*

*C'est machinalement, j'y aurai pas pensé (à faire du tactile), je suis comme un ordinateur, même si c'était devant mes yeux, non*

Ici le sujet fait référence à sa tendance globale à utiliser la modalité vocale et en particulier pour cette commande. Il est intéressant de noter que ce choix rompt pourtant une séquence précédente d'usage de la modalité tactile. Le choix d'une modalité peut donc obéir à des règles supérieures à la simple activation locale d'une modalité par son usage pour la commande précédente, et relève donc d'heuristiques à plus long terme, au cours d'une session voire entre sessions.

Le sujet note également le caractère automatique (non conscient) du choix de la modalité.

*S1 : i3 : Je voulais y aller (faire une commande vocale), finalement je lance avec le stylet*

*expé : qu'est ce qui a fait que tu as changé d'avis?*

*Je sais pas, l'utilisation du stylet avant je pense. Je voulais essayer puis finalement je reste dans le stylet.*

Dans cet autre exemple, le sujet témoigne d'une première intention de faire une commande vocale, qu'il utilise peu en général, mais le fait qu'il ait utilisé une autre modalité avant l'incite à conserver ce mode. L'heuristique dominante d'usage de la modalité tactile a donc tendance à lui faire conserver ce mode d'interaction.

6.3.2. Continuité et fluidité de l'interaction (sujets 1, 2, 3, 6)

Comme cela a été en partie mentionné précédemment, la possibilité de continuité de l'interaction dans une même modalité semble être un argument fort pour son éléction. Mais il semble qu'à l'inverse, l'impossibilité d'interagir avec une unique modalité a été pour au moins deux sujets, une des causes importantes de son rejet.

*S1i2 : Il faudrait que je puisse parler, si je pouvais tout faire, justement avoir mon PDA qui, peut-être limite dans la poche et tout faire à partir de la poche, ce serait bien mais y avait trop de coupures en fait dans le sens où faut cliquer, bien gérer la distance, ya aussi le fait que, quand je voulais envoyer des messages vocaux, on peut pas les envoyer qu'en vocal, puisque ça coupait, il fallait reprendre le stylet, donc ya vraiment beaucoup de coupures dans le processus et au final, j'sais pas le vocal, j'ai vraiment des réticences*

Cet exemple illustre bien le besoin ressenti par les sujets d'avoir accès à une équivalence totale entre les modalités pour effectuer toutes les commandes de l'application. En effet, la limitation du système obligeant parfois à utiliser la commande tactile est perçue comme une rupture dans la continuité d'un processus unifié d'interaction. Les arguments avancés par les sujets sont essentiellement liés aux différences des styles d'interaction entre les deux modalités. En effet, les contraintes et coût physiques intrinsèques de chacune sont de nature très différente, et obligent donc les utilisateurs à ajuster l'interaction pour chaque changement de modalité (pour le vocal : « où faut cliquer, bien gérer la distance », et pour le tactile : « **il fallait reprendre le stylet** »).

Ces ajustements sont ressentis comme des « coupures » peu appréciées.

*S2i2 : Comme j'étais dans le truc, j'étais préparé, ça allait vite, c'était beaucoup plus fluide, pour des raisons de rapidité justement, comme j'étais dans un mouvement, ça casse le rythme en fait d'utiliser la commande vocale, avec le stylet c'est beaucoup plus rapide*

*Expé : Comment ça, ça casse le rythme ?*

*S2i2 : En fait, changer de mode d'interaction, quand ya certains modes d'interaction dont on ne peut pas se passer, par exemple on est toujours obligé d'utiliser le stylet quoiqu'il en soit, moi je dis autant se cantonner au stylet parce que quand on est dans un mouvement comme ça et que de toute façon on peut pas faire exclusivement un seul mode d'interaction, où là dans ce cas le rythme serait pas cassé, le fait de changer de mode d'interaction, ça a pas d'intérêt*

Le sujet 2 exprime ici le même sentiment que celui mentionné dans l'exemple précédent, et il ajoute également en plus du critère de continuité, des critères de fluidité et de rapidité. Il se dit « préparé » à utiliser une modalité déjà activée, ce qui rend son accès plus facile et surtout plus rapide, et procure un sentiment de fluidité, de rythme et donc également d'efficacité.

*S2i2 : Mieux vaut rester sur le stylet parce qu'on est sur la lancée et on perd du temps à changer, après si ya avait intégralement la commande vocale, peut-être que j'aurai utilisé la commande vocale, mais le fait de devoir changer alors que j'étais dans une lancée, ça a fait que j'ai pas utilisé la commande vocale, ou sauf à certains moments quand le stylet marchait pas très bien*

Le sujet 2 confirme et justifie son rejet de la modalité vocale en partie en raison de la rupture du rythme d'interaction, cependant, il relève que celle-ci a pu lui être utile lors de situations dégradées d'usage de sa modalité par défaut (voir la partie 7 Les erreurs au niveau global (modalité vocale et tactile stylet)).

## 7. Les erreurs au niveau global (modalité vocale et tactile stylet)

L'étude générale du nombre d'erreurs indique la robustesse effective de l'application liée aux erreurs du système et à celles des utilisateurs. Elle donne en cela une vision objective de la robustesse du système qui permettra sa comparaison avec la perception qu'en ont les sujets. En effet, comme nous le verrons dans les verbalisations ce critère est important pour l'adoption ou le rejet de modalités.

Dans un premier temps, les erreurs ont été divisées en deux catégories :

- Erreurs de la modalité tactile stylet : Concerne les erreurs non liées à la reconnaissance vocale
- Erreurs de la modalité vocale

Modalité	Erreur	Succès	Total
Vocale	12,14%	10,62%	22,76%
Stylet	5,18%	72,06%	77,24%
<b>Total</b>	<b>17,31%</b>	<b>82,69%</b>	<b>100,00%</b>

**Tableau 8 Répartition des erreurs et succès pour la totalité des commandes**

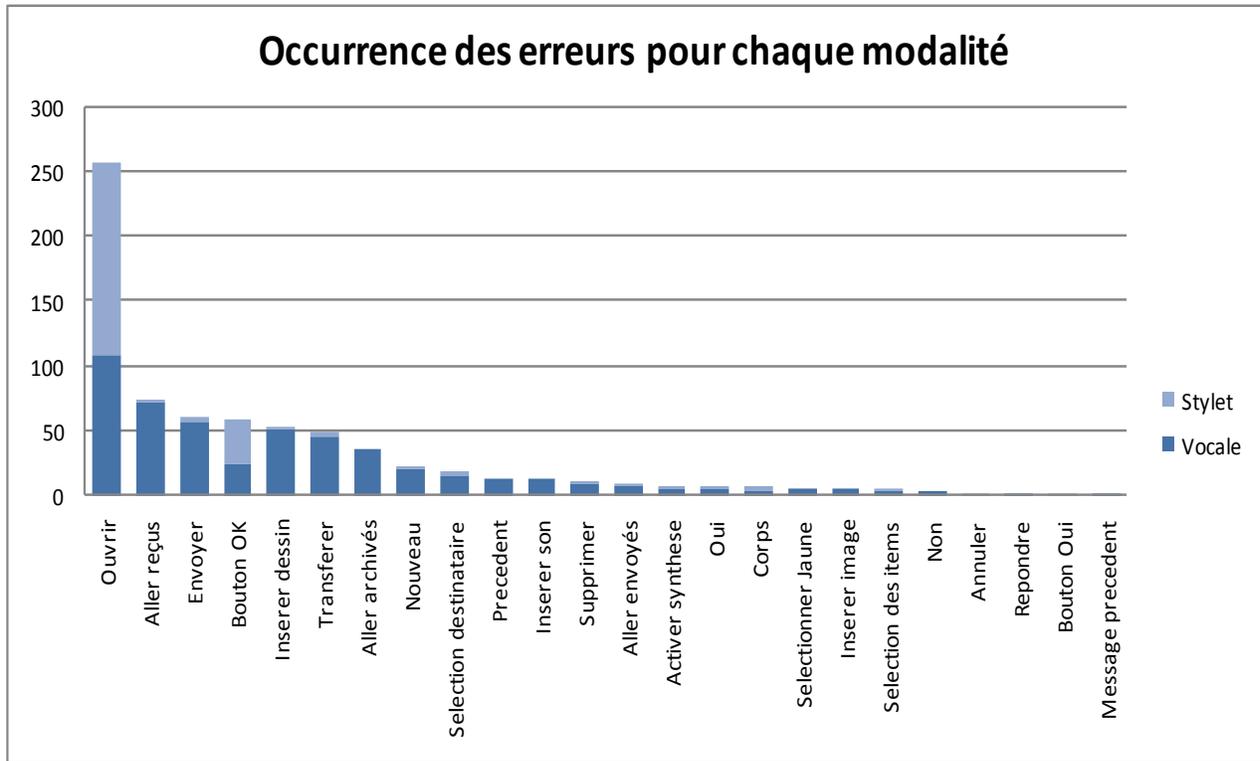
Le Tableau 8 indique que 17,31% des commandes (vocales ou tactile stylet) sont suivies d'une erreur. La répartition entre les deux modalités est la suivante : 12,14% d'erreurs pour les commandes vocales et 5,18% pour les commandes tactile stylet.

Il y a donc un taux d'erreur trois fois plus élevé pour la modalité vocale par rapport au stylet qui présente donc une meilleure robustesse.

L'analyse détaillée des erreurs pour chaque modalité est effectuée dans le paragraphe suivant.

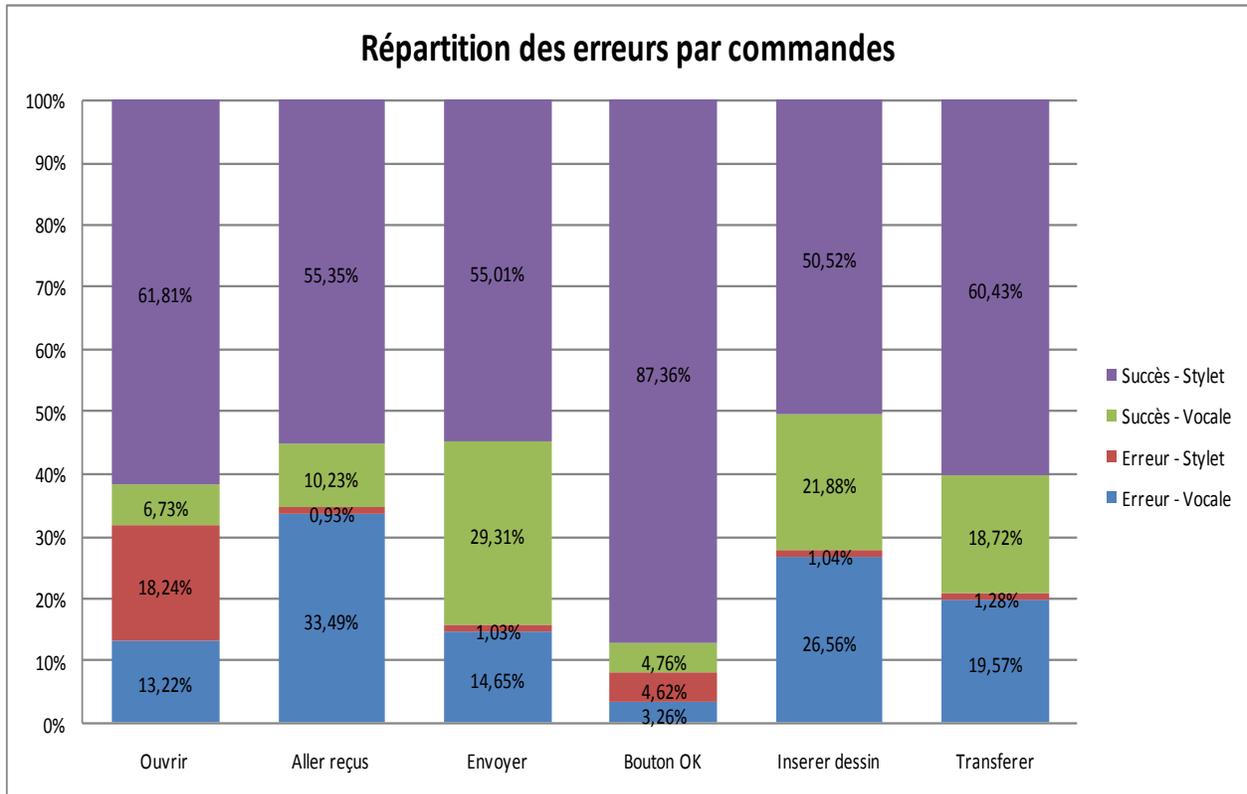
### 7.1. Erreurs et commandes

La Figure 38 hiérarchise le nombre d'erreurs pour toutes les commandes. Cette classification est faite pour nous orienter vers d'éventuelles situations d'activité qui provoqueraient plus d'erreur.



**Figure 38 Répartition des erreurs selon les commandes pour la modalité vocale et stylet**

En nombre absolu, la Figure 38 montre une grande variation du nombre d'erreur selon les commandes. La commande « ouvrir » présente le plus grand effectif d'erreur (plus de 250), suivi de loin par les commandes « Aller reçus », « Envoyer », « Bouton OK » (= fermer), « Insérer dessin » et enfin « transférer » dont les effectifs avoisinent 50 occurrences. Les autres commandes présentent des effectifs plus faibles et pour la suite de l'analyse des erreurs nous ne conserverons que les commandes pour lesquelles il y a le plus d'erreur. Nous conservons donc les six premières commandes pour les analyser plus finement.



**Figure 39 Répartition des erreurs pour les 6 commandes les plus utilisées pour la modalité vocale et stylet**

La Figure 39 relativise le nombre d'erreur par rapport aux succès, elle indique un taux d'erreur proche des 30% pour les commandes « Ouvrir », « Aller reçus » et « insérer dessin », et des taux entre 9 et 20% pour les commandes « Bouton OK » (=fermer), « Envoyer » et « transférer ».

Il faut également noter une répartition inégale des erreurs selon les modalités. En effet, seules les commandes « Ouvrir » et « Bouton OK » (=fermer) sont sujettes à des taux d'erreurs similaires avec les deux modalités vocale et tactile stylet. Les autres commandes quant à elles, sont quasiment exclusivement liées à la modalité vocale.

C'est donc la commande « ouvrir » qui rencontre le plus forts taux d'erreurs pour la modalité tactile et dans une moindre mesure, la commande « fermer ». Pour la modalité vocale, les commandes les moins robustes sont « aller reçus », « insérer dessin », « transférer » et enfin « envoyer » qui rencontrent des taux d'erreur très élevés.

Ces chiffres ne mentionnent pas les origines des erreurs et dans la suite de l'analyse, nous nous intéressons aux erreurs détaillées et spécifiques de chaque modalité.

### 7.2. Les erreurs de la modalité vocale

Les erreurs de la modalité vocale ont été classées en deux catégories :

- **Erreur Reco** : Erreurs liées au **système** de reconnaissance vocale proprement dit, alors que l'utilisateur a donné la bonne commande.
- **Impossible Reco** : Erreurs dues aux utilisateurs (mauvais libellé de commande, commande impossible etc.).

La modalité vocale, par comparaison au stylet, fait preuve d'une robustesse nettement plus faible. Le Tableau 9 indique un total de 53,33% d'erreur sur la totalité des commandes vocales.

Modalité	Impossible Reco	Erreur Reco	Succès	Total Erreur
Vocale	10,62%	42,70%	46,67%	53,33%

**Tableau 9 Répartition des types d'erreurs pour la totalité des erreurs liées à la modalité vocale**

Ces résultats indiquent une majorité d'interaction problématique concernant la modalité vocale, qui explique en partie la mauvaise appropriation de cette modalité par certains sujets, ce qui est confirmé par plusieurs verbalisations recueillies lors des autoconfrontations. (ex : « j'ai essayé plusieurs fois la reco, mais comme ça ne marchait pas, j'ai laissé tomber »). (voir 8 Verbalisations concernant les préférences individuelles)

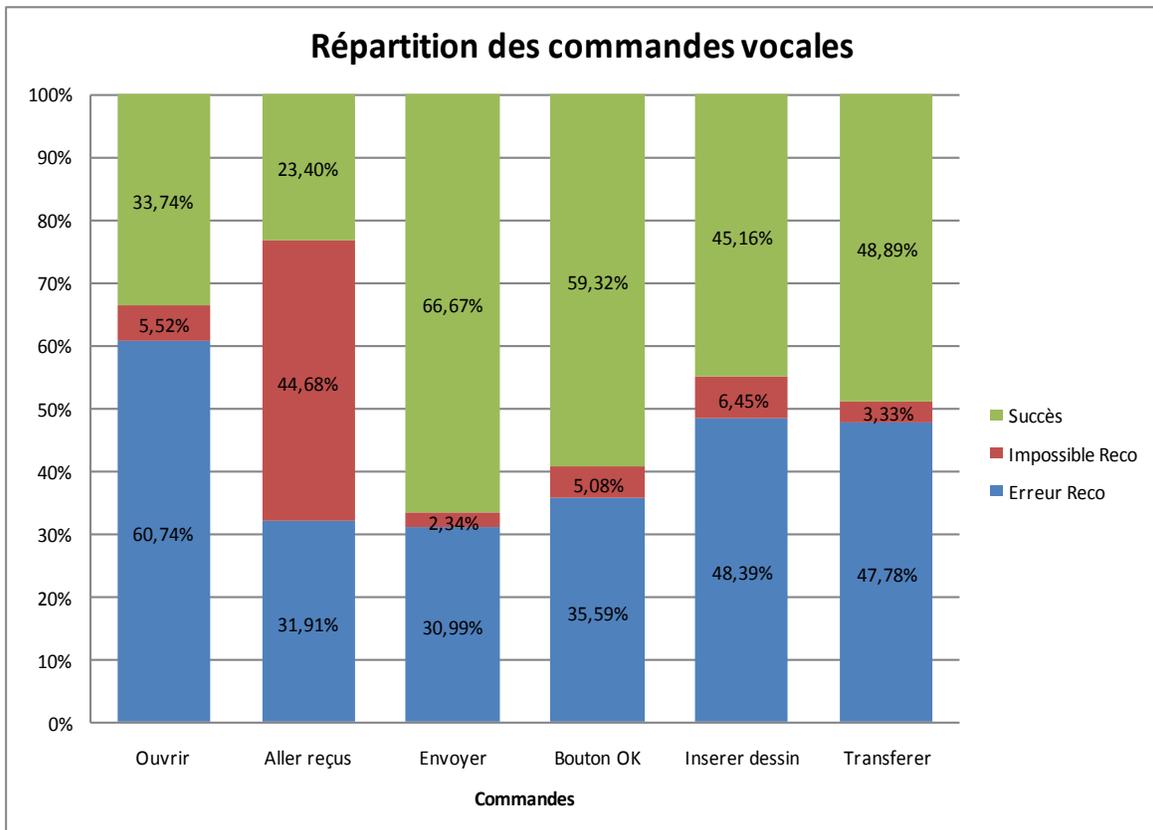
Le tableau suivant détaille ces résultats en ce focalisant uniquement sur la totalité des erreurs de la modalité vocale.

Modalité	Impossible Reco	Erreur Reco
Vocale	19,92%	80,08%

**Tableau 10 Répartition des types d'erreurs pour la totalité des erreurs liées à la modalité vocale**

Le Tableau 10 indique que parmi les erreurs de la modalité vocale, 80,08% sont des erreurs de reconnaissances liées au système lui-même alors que 19,92% des erreurs sont liées à des erreurs d'utilisation de la reconnaissance vocale par les sujets.

Une analyse des erreurs des sujets sera menée par la suite, mais il est important de noter que ce fort taux relève certainement d'un problème d'utilisabilité de l'application car dans la plupart des cas, les sujets pensaient effectivement effectuer une commande valide alors que ce n'était pas le cas (par exemple demander d'aller aux messages reçus, alors qu'ils sont dans la fenêtre d'édition).



**Figure 40 Répartition des erreurs selon les commandes pour la modalité vocale**

D'ores et déjà, la Figure 40 indique de fortes variations du taux d'erreur selon les commandes (de 33% à 77%), ainsi qu'une faible participation des erreurs de l'utilisateur sauf pour la commande « aller reçus » qui comptabilise 44,88% de commandes erronées. Cette concentration d'erreurs sur une seule commande confirme donc un problème d'utilisabilité de l'application. Les « erreurs utilisateur » liées à cette commande sont certainement liées pour la plupart au fait que l'utilisateur pense pouvoir aller à l'espace reçus depuis tous les autres espaces, les verbalisations ont montré qu'elle est perçue comme une sorte de raccourci accessible depuis toutes les fenêtres de l'application.

Loin derrière (6,45%), la deuxième commande avec un fort taux d'erreur est « insérer dessin ». Dans ce cas, il a également été identifié un problème d'utilisabilité lié à la confusion entre deux possibilités d'insertion de média. Il s'agissait d'une confusion entre l'insertion d'un dessin et d'une image qui a provoqué la plupart de ces erreurs.

Il est à noter que dans tous ces cas, nous n'avons pu prévoir de feedbacks spécifiques permettant d'informer l'utilisateur de l'impossibilité d'effectuer ces commandes et ainsi de corriger la situation, ce qui a provoqué beaucoup d'incompréhension pendant les premières sessions d'interaction, ainsi qu'une certaine frustration et perte de confiance dans les capacités du système à comprendre les intentions de l'utilisateur.

### 7.3. Les erreurs de la modalité tactile stylet

Trois catégories d'erreurs stylet ont été distinguées :

- Erreur uti : erreur de l'utilisateur, mauvaise exécution de la commande, commande inappropriée, (clique à côté, pas assez vite)
- Erreur sys : le système ne prend pas en compte la commande de l'utilisateur
- Erreur : Erreurs pour lesquelles la cause n'a pas pu être déterminée

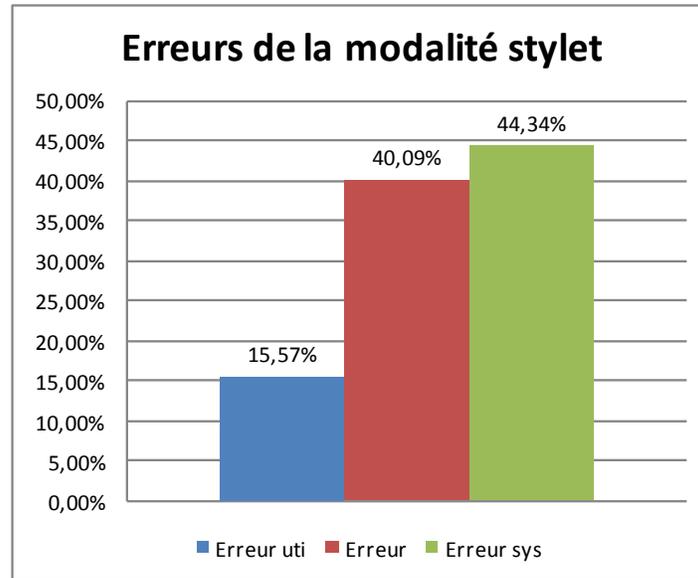


Figure 41 Répartition des types d'erreurs pour la modalité tactile stylet

La Figure 41 indique un grand nombre (40,09%) d'erreurs dont la cause n'a pu être identifiée. Comme pour la modalité vocale, la source principale d'erreurs provient du système (44,34%) et dans une moindre mesure de l'utilisateur (15,57%).

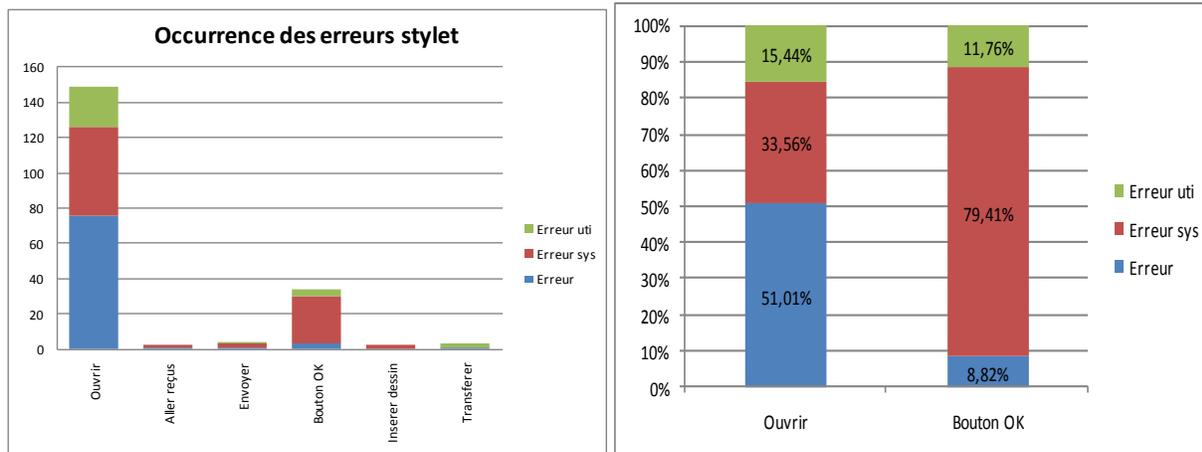


Figure 42 Répartition des erreurs selon les commandes pour la modalité stylet

Comme pour la modalité vocale, en terme d'occurrences, la principale commande source d'erreurs est « ouvrir » (n=149) et loin derrière « bouton ok » (=fermer) (n=34). Ce résultat montre également la concentration des erreurs de la modalité tactile stylet sur une commande quasi unique et ceci autant pour les erreurs systèmes que pour les erreurs utilisateurs. L'analyse détaillée de ces erreurs après dépouillement des vidéos a montré un problème à la fois d'utilisabilité et du système lié à l'implémentation du double clic au stylet qui est requis pour cette commande. Les problèmes liés au système (lenteur d'ouverture des messages et absence de feedback d'attente) ont rajouté à la confusion des utilisateurs sur la validité de leur double clic, expliquant ainsi un grand nombre de leurs erreurs en tentatives de corrections parfois. Les erreurs dont la cause n'est pas claire, relève d'ailleurs la plupart du temps de ces cas.

## 8. Verbalisations concernant les préférences individuelles

Dans cette partie, faisant suite à une analyse quantitative, nous abordons la question de l'usage des modalités à la fois avec l'appui des données quantitatives vues précédemment au niveau individuel, global et session par session, mais nous introduisons également une approche plus qualitative en intégrant une analyse des verbalisations des sujets.

Ces verbalisations sont présentées de manière thématique afin de dégager des problématiques récurrentes sur les causes d'adoption et de rejet de certaines modalités. Certaines de ces thématiques, sont reprises plus en détail dans leurs parties respectives, en particulier en ce qui concerne les liens entre modalités et mobilité (8 Perceptions des utilisateurs). Nous nous intéressons donc plus particulièrement aux propriétés intrinsèques des modalités, ainsi qu'à leur adoption par les sujets, selon leurs préférences, leur expérience, etc.. Ceux-ci vont se forger au fil du temps des préférences d'usages qui vont modeler leur style d'interaction propre. Ce phénomène apparaît dans les verbalisations sous deux grandes thématiques : une phase de test et d'appropriation de l'application, et la vérification de l'adéquation de l'application aux exigences de la tâche.

### 8.1. Test et appropriation de l'application et des modalités

Dans une étude précédente (Calvet, 2001) nous avons abordé la problématique des premiers usages d'une application en identifiant une phase de test informel et de familiarisation avec l'application qui se traduisait par des usages des modalités non stabilisés dans un premier temps.

Cette question apparaît également ici puisque nous pouvons noter une évolution des usages entre les sessions, ainsi que des verbalisations durant les autoconfrontations qui abordent cette problématique sous divers aspects que nous avons regroupés par thèmes. En effet, s'agissant d'une application et d'une consigne relativement complexes, dans un environnement également riche, les sujets ont du appréhender un certain nombre d'informations nouvelles dès le début des expérimentations, provoquant ainsi une charge de travail relativement élevée et une disponibilité cognitive variable qui a également joué sur l'adoption de l'application.

#### 8.1.1. Test de l'application et des modalités (S2, S4, S5)

Il a plusieurs fois été mentionné explicitement par les sujets qu'ils désiraient tester chaque possibilité qu'offrait l'application, que ce soit en termes de fonctionnalités ou de modalités d'interactions. Cette volonté en soi, justifie l'usage, au moins dans un premier temps, de modalités nouvelles telles que la reconnaissance vocale ou le Geste 2D.

*S2 : C'est vrai qu'en plus je voulais essayer cette application là (reconnaissance vocale), mais j'avais pas mal de soucis*

Dans cet exemple, le sujet manifeste sa volonté d'essayer la modalité vocale, mais témoigne également d'une expérience négative qui va contribuer à renforcer le rejet de cette modalité.

*S5 : Ben non, en fait ouais, ben j'ai voulu changer, pour pouvoir juger toutes les fonctions  
Expé : t'avais pas peur que ça fonctionne pas ?  
Non, j'ai testé (le vocal) et ça marchait donc*

Dans ce deuxième exemple, il s'agit d'une exploration relativement systématique et volontaire de toutes les modalités et fonctions afin de se construire une représentation globale des possibilités et des caractéristiques de l'application. Dans ce cas, l'exploration se solde par une expérience positive de l'usage de la modalité vocale favorisant son adoption.

#### 8.1.2. Moments de test (S1, S2)

Pour certains, ce test informel se déroule souvent dans les interstices de moindre charge de travail, laissés lorsque certaines tâches ont pu être effectuées de manière prioritaire, comme par exemple envoyer un indice à leur partenaire de Quiz.

*S1 : Le vocal quand j'ai essayé, c'était vraiment pour essayer, ça marchait jamais donc, dès que j'avais un peu de temps j'essayais de le faire*

*S2 : Moi-même, quand je cherchais mes réponses, je devais jongler entre beaucoup de choses et c'est vrai que j'ai pas eu le temps de penser à utiliser le mode d'interaction vocale*

Ces deux exemples montrent bien (pour les sujets 1 et 2) un temps spécifique de test de modalités non privilégiées. Pour eux, la tâche principale est prioritaire (« je devais chercher mes réponses ») et la modalité vocale n'est pas leur modalité privilégiée, par défaut, puisqu'ils considèrent qu'elle demande un effort supplémentaire et un temps dédié à son usage.

#### 8.1.3. Aspect ludique, nouveauté (S1, S4, S5, S6)

Le fait que l'application propose des modalités et des fonctionnalités innovantes a provoqué un intérêt et une curiosité poussant également à l'exploration des modalités et des fonctions, par son aspect ludique et nouveau.

*S4 : Euh, non, là c'est venu tout seul, c'est par rapport à la manip qu'on avait faite le 1er soir et puis vu que c'est le truc le plus nouveau pour moi sur un appareil comme ça, autant l'utiliser (la modalité vocale) quand il est là*

Cet aspect a joué également au niveau social chez le sujet 6 qui a trouvé que l'usage d'une application innovante pouvait attirer l'attention positivement sur lui.

#### 8.1.4. Contexte expérimental (S4)

Les sujets ont peu mentionné comme explication de choix d'usage des modalités le fait qu'il s'agisse d'un contexte expérimental, relativement contraint.

*Expé : t'avais l'impression que c'était un test?*

*S4 : Plus ouais, parce que c'est vrai que depuis le début je fais que avec le stylet*

Cependant, et en liaison avec les points mentionnés précédemment, certains sujets (sujet 4 par exemple), se sont sentis « obligés » de tester la modalité vocale afin « d'honorer » leur mandat de sujets d'expérimentations, même si la consigne et les échanges étaient on ne peut plus clairs à ce sujet. Les sujets avaient la totale liberté du choix d'usage des modalités et n'avaient aucune obligation de tester exhaustivement l'application.

#### 8.1.5. Test de la mise en œuvre des modalités (S1)

Au-delà même des qualités intrinsèques de chaque modalité, l'implémentation de mise en œuvre choisie par les concepteurs va jouer un rôle important pour l'adoption des modalités.

*S1 : « messages reçus ». Là j'ai essayé, j'y arrivais pas, pourtant j'avais bien mis, j'avais mis le micro, j'avais appuyé. J'ai dit " message reçus", c'était peut-être "aller à message reçus", je crois que j'ai essayé d'autres choses. Ah j'ai réessayé pas mal de fois, celle là elle a marché je crois, pasque je voulais essayer quand même, voir si ça pouvait être potentiellement bien, mais j'arrivais vraiment pas à m'en servir*

Pour la modalité vocale, cela peut concerner à la fois le test de la mise en œuvre physique de l'interface : mode vocal activé (« j'avais mis le micro »), enclenchement de la reconnaissance (« j'avais appuyé »), mais également cela peut concerner la mise en œuvre sémantique (« j'ai dit « messages reçus », c'était peut-être « aller à messages reçus », je crois que j'ai essayé d'autres choses »).

Pendant les premiers temps d'interactions, les sujets réalisent donc des tests de mise en œuvre de chaque modalité et fonctionnalité afin de renforcer leur apprentissage et leur expertise, tout en commençant à se forger des préférences d'usages.

### 8.1.6. Test de l'efficacité

L'efficacité est un critère important que certains sujets (en particulier les sujets 1 et 2) déclarent avoir testé en début de session expérimentale (voir également la partie 6 Transitions et enchaînement des modalités). Cette notion est apparue sous trois formes lors des entretiens : la rapidité d'interaction et son opposée la lenteur, et enfin la notion de raccourci associée aux modalités vocale et G2D.

- Rapidité VS. Lenteur : contrôle de l'application (S1, S2, S3)

La notion de rapidité est évidemment subjective (il s'agit en fait de la perception de rapidité), mais elle entre en compte de manière importante dans l'adoption ou le rejet de telle ou telle modalité. Certains des sujets aiment que l'application soit la plus réactive possible et ne tolèrent que très peu de latence, entre l'intention de commande, la commande et la réponse du système.

*S1 : c'était un peu pour tester le vocal, je voulais voir si c'était rapide et tout, c'est pareil, j'ai réessayé encore une fois. C'est pas un mode qui m'a plu du tout.*

Dans l'exemple ci-dessus, le sujet 1 indique vouloir tester la rapidité de la modalité vocale, et ceci à plusieurs reprises.

*S3 : Comme ça, j'ai voulu essayer (le vocal), voir si ça allait plus vite (que les autres modalités)*

Dans cet autre exemple, le sujet 3 déclare aussi vouloir tester la rapidité de la modalité vocale, mais ici en comparaison avec les autres modalités en particulier la modalité tactile stylet considérée comme la référente. D'après les verbalisations il convient donc de considérer deux types de tests de rapidité :

**Rapidité absolue : rapidité intrinsèque de chaque modalité, indépendamment des autres**

**Rapidité relative : comparaison relative de la rapidité de chaque modalité entre elles**

Dans l'exemple suivant, le sujet 1 explique en partie son abandon de la modalité vocale (« C'est pas un mode qui m'a plu du tout ») par sa lenteur et la lourdeur du processus à mettre en œuvre.

*S1 : Le processus je le trouve vraiment trop long, même si elle m'avait dit, oui, msg envoyé. Oui, le fait qu'elle répète tout ce qu'on dit, elle répète non, je crois? Elle valide au moins la commande, ben le fait qu'il y ait des fois une réponse, je trouve ça vraiment trop long. On attend de savoir..ah, c'est bon, non c'est pas bon. Ya trop de temps d'attente, entre le moment où on parle et le moment où les choses arrivent*

En effet, pour les sujets 1 et 3, qui ont très peu utilisé la modalité vocale, la lenteur et le nombre d'événements associés à sa mise en œuvre, sont des critères quasiment rédhibitoires pour une adoption. Le sujet 1 rajoute également une appréciation très négative concernant le feedback en synthèse vocale pour la validation d'une commande et se plaint du manque d'efficacité globale de la modalité vocale. L'aspect non déterministe et incertain du résultat d'une commande vocale est perçu comme une perte de contrôle sur l'interaction et nuit à sa fluidité.

- Association du vocal et G2D à des raccourcis (S1)

La notion de raccourcis a également été mentionnée à plusieurs reprises, celle-ci inclut plusieurs notions :

- encore une fois la rapidité de mise en œuvre, qui peut résider dans le fait de rassembler plusieurs commandes dans une seule,
- mais aussi la possibilité d'effectuer une commande actuellement non visible sur l'interface et donc non accessible directement à l'aide de la modalité tactile stylet.

*S1 : En fait je me demandais quel raccourci je voulais garder je pense, parce que pour moi, je les voyais comme des raccourcis en fait ces deux trucs (vocal et g2D)...c'était juste pour faire des choses plus rapides. Par exemple, est-ce que je vais utiliser pour aller plus vite le mode vocal ou le mode 2d, et en fait aucun des deux ne me servait*

Ici, le sujet donne l'impression d'avoir une modalité par défaut (la modalité tactile stylet) et manifeste son attente d'avoir des modalités annexes qui ponctuellement serviraient de raccourcis plus rapides. Or, l'application n'offre pas ces possibilités puisque le principe de stricte équivalence a guidé sa conception (toutes les modalités permettent l'accès aux mêmes commandes). Cette attente, comme nous l'avons vu dans la partie concernant les erreurs utilisateurs de la modalité vocale, a généré un assez grand nombre d'erreurs utilisateurs, en particulier pour la commande « aller reçus » qui permet d'afficher la fenêtre contenant la liste des messages reçus et que de nombreux sujets ont cherché à atteindre depuis des fenêtres ne le permettant pas.

## 8.2. Exigences de la tâche

Les critères d'évaluation de l'application et de ses modalités sont relatifs aux tâches que les sujets doivent exécuter. Dans ce cadre, plusieurs déterminants ont été mentionnés lors des entretiens. Ceux-ci ont été regroupés dans 4 thématiques :

- **Rapidité** : Comme expliqué dans la partie précédente (test de l'application), la rapidité est un critère déterminant pour l'adoption d'une modalité. Mais ici, nous abordons ce critère sous la forme de l'adéquation de la rapidité de la modalité proposée aux exigences de la tâche demandée.
- **Criticité** : Le critère de criticité est mentionné par certains auteurs (Karsenty, 2006), mais étant donné le caractère ludique et expérimental de notre étude, ce critère ne semble pas avoir été pris en compte par les sujets et nous ne disposons pas de verbalisations à ce propos.
- **Effort de mise en œuvre** : L'effort de mise en œuvre d'une modalité s'inscrit ici non seulement dans les qualités intrinsèques des modalités mais aussi dans leur insertion dans une séquence d'activité mettant éventuellement en œuvre d'autres modalités ou actions physiques et cognitives.
- **Spécialisation** : La spécialisation pourrait être appréhendée ici comme une autoassignation récurrente par les sujets de l'usage d'une modalité spécifique pour une commande donnée.

### 8.2.1. Rapidité

L'exigence de rapidité prend son importance au regard de plusieurs caractéristiques de l'activité de l'utilisateur et elle se décline avec des sous critères dont :

- la pression temporelle et sociale,
- la charge de travail,
- la comparaison entre modalités
- et la priorité entre les tâches.

Nous développons ici chacun de ces thèmes séparément, sachant que bien évidemment, ils peuvent être considérés conjointement par les sujets.

- Pression temporelle et sociale (S1, S2, S6)

Le contexte de l'expérimentation étant un Quiz collaboratif, certains des sujets (S1 et S2) ont ressentis une sorte de pression temporelle, d'urgence, afin de répondre entre autre aux attentes de leur collègue à qui ils doivent transmettre des indices pour leur permettre de trouver leurs réponses. Entre ces deux sujets, s'est souvent instauré un esprit de « challenge » afin de compléter au plus vite, avec un léger esprit de compétition, l'ensemble du Quiz. Cette thématique est également abordée dans l'analyse de l'activité réelle des sujets (voir 4.2 Influence de l'activité réelle sur l'usage des modalités)

*S2 : Parce qu'en même temps, y avait une forme de pression, comme j'avais dit aussi la dernière fois qui faisait qu'on devait quand même être assez rapide dans l'envoi des réponses, peut-être dans une situation un petit peu plus posée, on pourrait prendre le temps d'utiliser la commande vocale mais vraiment, là, comme y a un rythme assez rapide, on envoie des messages, ya des impératifs, on doit répondre assez rapidement à l'autre, Thibaud, j'allais pas le faire poireauter trop longtemps donc*

Ceci a pour conséquence une certaine exigence d'efficacité et donc de rapidité qui laisse peu de place soit au test d'une nouvelle modalité, soit à une modalité déjà testée mais considérée comme plus lente.

- Charge de travail (S6)

Il apparaît également souvent une notion corrélée à la pression temporelle qui est la charge de travail, mobilisant plus ou moins les ressources cognitives.

*S6 : Je peux utiliser les trucs (modalité vocale, g2d) quand je contrôle exactement ce que je vais faire, où je vais, mais quand je sais pas exactement, je suis un peu paumé autant dans l'application que dans mon raisonnement, enfin dans ma démarche intellectuelle, ... il faut que j'utilise le stylet quoi*

Dans cet exemple, le sujet 6 exprime sa difficulté à utiliser des modalités inhabituelles, moins maîtrisées, lorsqu'il est dans une situation de forte charge de travail cognitif, ce qui est le cas, en l'occurrence, lorsqu'il est dans le doute quant à la marche à suivre pour remplir sa tâche. Dans ce cas, il préfère utiliser une modalité mieux maîtrisée, plus familière (tactile stylet) comme une sorte de modalité par défaut, garante de certitude quant à son usage et qui soulage ou au moins n'augmente pas, sa charge de travail. Le sujet mentionne également la notion de **contrôle** qui semble faire défaut à la modalité vocale et pour laquelle les sujets semblent lui préférer la modalité tactile qui est souvent considérée comme une modalité « sûre » et source de contrôle par défaut.

- Comparaison vocal vs tactile (S3, S6)

Comme nous l'avons mentionné dans la partie test de l'application, les modalités acquièrent au fil des usages et des tests de l'application, des propriétés perçues par les sujets qui viennent déterminer leurs préférences individuelles. Comme nous l'avons également mentionné, ces préférences sont acquises en partie par comparaison des caractéristiques intrinsèques de chaque modalité, parmi l'ensemble des modalités proposées. Cela permet aux sujets d'établir une sorte de classement personnel en matière d'efficacité et donc de rapidité.

*S3 : A la fin ça allait, j'arrivais à manipuler le stylet, ça allait plus vite, au lieu de "insérer dessin" (modalité vocale) que ça ouvrait pas, je me suis dit, vas y, faut que j'y aille rapidement donc insérer avec le stylet, ça allait deux fois plus vite*

On voit ici que le sujet commence à acquérir une certaine maîtrise de l'usage des modalités et il peut ainsi choisir en fonction des exigences locales de la tâche une modalité appropriée. En l'occurrence, ici, le sujet qui a l'habitude d'insérer un dessin avec la modalité vocale, va modifier son comportement pour pallier une erreur du système alors que la pression temporelle se fait sentir.

- Priorité de tâche (S3, S6)

L'interaction en mobilité, implique souvent la conduite de plusieurs tâches en parallèle, ne serait-ce que la gestion de son propre déplacement, mais également dans le contexte de notre étude, l'accomplissement du Quiz. Le choix des modalités va donc s'insérer dans ce canevas de contraintes liées aux diverses activités en cours. Ces activités sont plus ou moins hiérarchisées et acquièrent des priorités variables selon les contextes. Selon cette priorisation, les tâches obtiennent des statuts particuliers liés à un certain nombre de critères dont :

- ✓ la possibilité d'interruption de la tâche,
- ✓ la possibilité et/ou facilité de reprise de cette tâche,
- ✓ le caractère obligatoire ou urgent de la tâche (pression temporelle et/ou sociale, gestion d'un déplacement difficile)
- ✓ L'inscription de la tâche dans une séquence unitaire à compléter

Cette priorisation hérite, comme nous l'avons également vu, des caractères intrinsèques des tâches mais aussi de leur perception par les sujets et il en résulte donc que la priorisation de l'un ne correspondra pas forcément à la priorisation de l'autre.

Dans les exemples que nous présentons ci-dessous, les sujets présentent des positions complètement opposées sur leur perception de l'efficacité de la modalité vocale, et donc de leur usage en fonction de la priorité de tâche à réaliser.

- La modalité vocale vue comme plus efficace :

Le sujet 3 perçoit la modalité vocale d'une manière particulière puisque de son point de vue, elle est plus efficace que la modalité tactile stylet, et il va donc la mettre en œuvre dans les situations qu'il considère appropriées :

*S3 : Oui, mais c'est différent, ya des moments où j'ai besoin de regarder à côté parce que je cherche un indice, t'en as d'autres, non. C'est quand j'ai plus de temps, dans ce cas là, je préfère prendre mon temps (et utiliser le stylet), quand j'en ai pas, je préfère que ce soit fait automatiquement (en vocal)*

*S3 : De temps en temps, quand t'es pressé, t'as pas le temps, c'est excellent (le vocal), mais quand t'as tout le temps, je préfère à la limite le faire moi-même (en tactile)*

Pour le sujet 3, la modalité vocale est perçue comme plus efficace, rapide, et comme une sorte de mode automatique qui « travaille à sa place », et donc qui libère des ressources pour faire autre chose. Il s'agit en quelque sorte d'une délégation de tâche. Il choisira donc la modalité vocale lorsqu'il doit effectuer une tâche qu'il considère comme prioritaire, en particulier pour les tâches non centrées sur l'interaction, comme ici, par exemple où il cherche un indice dans l'environnement et ne veut pas interrompre son activité.

En revanche, la modalité tactile stylet est réservée pour les moments où il a du temps, et des ressources attentionnelles à consacrer à l'interaction.

De plus, cette modalité est considérée comme un mode direct d'action sur l'interface (« je préfère le faire moi-même ») par opposition à la modalité vocale « où c'est la machine qui travaille ».

- La modalité vocale vue comme moins efficace :

Au contraire, pour le sujet 6 (et la plupart des autres sujets), la modalité tactile stylet est considérée comme la plus rapide et donc privilégiée dans les moments où la pression temporelle augmente :

*S6 : Le stylet c'est plus rapide que la reco vocale, t'es toujours obligé d'attendre qu'elle finisse de dire le truc, ça marche pas forcément du premier coup. C'est plus agréable parce que t'es plus libre, mai ça va un peu plus lentement, et là, Je repasse en vocal comme j'ai rattrapé un peu de retard*

Dans cet exemple, le sujet rattrape son retard par une série de commandes à l'aide du stylet, qui n'est pas sa modalité préférentielle, mais qui ici répond mieux aux exigences de la tâche.

Il est également intéressant de noter ici que le sujet 6 (ainsi que la plupart des autres sujets) reconnaît la moindre efficacité de la modalité vocale en termes purement objectifs (rapidité, mise en œuvre, propension aux erreurs, etc.), mais il la reconnaît comme une modalité privilégiée du point de vue subjectif (« agréable parce que tu es plus libre »). Il accepte donc un compromis d'efficacité objective inférieure pour un confort subjectif supérieur. Cette notion de confort subjectif sera davantage développée dans la partie concernant les ressources attentionnelles.

### 8.2.2. Criticité

La notion de criticité n'a pas été mentionnée dans les entretiens que nous avons menés, mais comme le suggère (Karsenty, 2006), ce critère pourrait être déterminant dans des applications présentant des caractères critiques (irréversibilité, conséquences importantes des commandes, contraintes temporelles fortes) qui laisseraient moins de place aux préférences subjectives. Dans notre application, même si de nombreuses commandes revêtaient effectivement un caractère irréversible, les conséquences en étaient peu critiques et de fait, ce déterminant a peu joué dans les choix d'usages des modalités ce qui explique certainement que ce critère n'a jamais été mentionné dans les entretiens d'autoconfrontation.

### 8.2.3. Effort de mise en œuvre

L'usage de chaque modalité représente un effort, un coût de mise en œuvre propre tant cognitif que physique, qui est évalué, comme nous l'avons mentionné précédemment, au cours de l'apprentissage de l'application. Cet effort peut être évalué de manière absolue (effort de mise en œuvre d'une modalité isolée, valeur intrinsèque) ou relative aux caractéristiques des autres modalités, ponctuellement ou encore dans son insertion dans une séquence d'usage, cas qui constituera le premier point abordé dans la suite de l'exposé. Viendront ensuite les points suivants : aspect « raccourci » de certaines commandes, libération de l'attention visuelle, sémantique des commandes vocales. Tous ces points sont apparus comme déterminants dans les choix d'usage des modalités.

- Séquence d'usage (S1, S2, S3, S6) (voir également la section « transition »)

L'étude des déterminants des usages d'une modalité ne peut pas être abordée de manière isolée. En effet, une action est déterminée par les actions passées et à son tour, celle-ci vient déterminer les actions futures.

La mise en œuvre d'une modalité (qui a été abordée dans la partie « test de l'application ») active des mécanismes tant cognitifs que physiques qui peuvent avoir tendance à s'auto entretenir. Comme nous allons le montrer dans les deux points suivants, ces mécanismes peuvent ainsi générer des séquences d'usages et ceci pour plusieurs raisons :

- Pré activation de modalité (S1)

Lorsqu'un sujet utilise une modalité, il y a une certaine tendance à enchaîner la commande suivant immédiatement avec la même modalité.

*S1 : Je voulais y aller (au vocal), finalement je lance avec le stylet*

*Expé : qu'est ce qui a fait que tu as changé d'avis?*

*S1 : Je sais pas, l'utilisation du stylet avant je pense. Je voulais essayer puis finalement je reste dans le stylet.*

...

*Pour moi, le vocal, tout va en même temps, c'est vraiment un ensemble, si elle parle, il faut que moi je parle aussi*

Dans les deux extraits ci-dessus, plusieurs points intéressants sont à relever.

Le sujet 1 explique que son intention première était de faire une commande vocale, mais que finalement, il fait une commande tactile en raison du fait qu'il venait d'utiliser cette modalité juste avant. Cela indique une pré activation, une sorte de facilitation, certainement par la remise en mémoire court terme de la mise en œuvre d'une modalité par elle-même du point de vue cognitif. Mais également, du point de vue de la mise en œuvre physique, les membres sollicités pour passer une commande sont déjà mobilisés et prêts à agir du fait de leur activation pour la commande précédente.

Dans la deuxième partie de la verbalisation, un autre point est abordé, celui de la perception d'ensemble d'un style d'interaction. Pour le sujet 1, si l'application parle, lui aussi doit parler, comme s'il fallait respecter l'homogénéité de l'interaction par des entrées et sorties faisant appel aux mêmes modalités. Il s'agit donc également d'une activation de modalité de commande par une modalité de sortie du système. Cependant, cette thématique semble avoir été peu abordée dans l'ensemble.

- Enchaînement de modalité (tous les sujets)

Pour plusieurs sujets, cet enchaînement de modalité généré par sa pré activation, constitue un critère important mentionné comme raison de non-usage d'une modalité.

*S1 : La reconnaissance vocale, euh, on l'avait travaillée avant mais euh, ça marchait bien, j'ai vu que ça marchait bien, mais je sais pas c'est pas naturel quoi. (...) enfin je suis resté sur l'idée de pas coupler voix et stylet. Je préfère vraiment rester, si je commence avec le stylet, je fais tout avec le stylet. Et si je commence à faire euh, même les commandes vocales, j'sais pas, il faudrait que je puisse tout faire en vocal, je trouve que ça reste trop long, trop compliqué, ya la personne qui redit derrière*

Dans cet exemple, le sujet 1 explique, en cohérence avec les verbalisations du point précédent, qu'il n'aime pas avoir recours à plusieurs modalités d'interactions pour une même séquence de commandes. Il y a une volonté de non-couplage de styles d'interactions différents qui entre autre est justifiée par l'activation cognitive et physique d'une modalité au détriment d'une autre. La cohabitation de deux modalités mettant en œuvre des mécanismes différents n'est pas souhaitée.

Pour obtenir une meilleure « fluidité » de l'interaction (S1, S2) préfèrent une certaine unité d'usage de modalité.

Au-delà de l'activation d'une modalité, le fait que l'application exige parfois le recours obligatoire à la modalité tactile stylet pour certaines actions (malheureusement, l'application n'a pas pu être développée totalement selon le principe de l'équivalence), provoque des interruptions de séquences de modalité unique qui sont peu appréciées et qui défavorisent la modalité vocale.

*S1 : (expé : t'as pas eu envie de te le faire lire le msg ?) Non, j'y ai même pas pensé à ça, j'aurai pu le faire ouais. C'est vrai que j'ai pas du tout pensé à la reconnaissance.. à la synthèse. Mais j'aime pas sa voix. Pour moi, le vocal, tout va en même temps, c'est vraiment un ensemble, si elle parle, il faut que moi je parle aussi. Il faudrait que je puisse parler, si je pouvais tout faire, justement avoir mon pda qui, peut-être limite dans la poche et tout faire à partir de la poche, ce serait bien mais y avait trop de coupures en fait dans le sens où faut cliquer, bien gérer la distance, ya aussi le fait que, quand je voulais envoyer des messages vocaux, on peut pas les envoyer qu'en vocal, puisque ça coupait, il fallait reprendre le stylet, donc ya vraiment beaucoup de coupures dans le processus et au final, j'sais pas le vocal, j'ai vraiment des réticences, même elle qui me parle parce que je trouve ça trop long*

Dans cet exemple, le sujet regrette de ne pouvoir effectuer toutes les commandes uniquement à l'aide d'une modalité unique. Par exemple, l'édition d'un message audio ne pouvait être pilotée qu'avec la modalité tactile ce qui est regrettable puisque sollicitant l'attention visuelle, alors que les modalités et médias concernés sont tous audio (« limite avoir le PDA dans la poche »). Avoir recours à une modalité visuelle pour utiliser un média ou une modalité non visuelle est donc perçu négativement.

La remarque est également valable du point de vue de la mobilisation des membres (mains et bras en particulier) qui sont peu sollicités pour une commande vocale et fortement pour une commande tactile.

- Aspect raccourci (rapide) (S1, S2, S5)

Comme nous l'avons mentionné dans la partie mentionnant le test de l'application, certaines commandes vocales peuvent être perçues comme des sortes de raccourcis facilitant leur mise en œuvre et augmentant donc leur rapidité.

*Expé : qu'est-ce qui te plaît dans cette commande ?*

*S5 : En fait, non, c'est que je sais que si par exemple "ouvrir", "fermer", des fois, faut toujours répéter, répéter, donc avec le stylet ça va vite, mais là donc "envoyer à rouge", c'est la rapidité, le côté pratique en fait*

Ici, l'exemple de la commande « envoyer à rouge » peut avoir dans certains cas la particularité de rassembler deux commandes en une seule. En effet, dans les cas où le destinataire n'est pas défini à l'avance (nouveau message, transférer), (à l'inverse de répondre qui sélectionne automatiquement le bon destinataire), afin de procéder à l'envoi d'un message, il convient de d'abord en sélectionner le destinataire. La commande vocale « envoyer à rouge » peut donc être une combinaison de deux commandes qui raccourci l'interaction du point de vue du nombre d'actions à effectuer et donc de la rapidité d'exécution.

En contrepartie, sa complexité pourrait augmenter sa propension aux erreurs.

- Libération de l'attention et de la mobilisation des membres (S6)

Cette question est traitée en détail dans la partie (Chapitre 5 Multimodalité, mobilité et ressources attentionnelles) mais nous la mentionnons tout de même ici car il s'agit d'un déterminant important de la préférence d'usage de la modalité vocale.

*S6 : Le stylet c'est plus rapide que la reco vocale, t'es toujours obligé d'attendre qu'elle finisse de dire le truc, ça marche pas forcément du premier coup. C'est plus agréable parce que t'es plus libre, mai ça va un peu plus lentement*

Le sujet 6 énumère un certain nombre d'inconvénients de la modalité vocale (attente de la confirmation par la synthèse, risque d'erreur, lenteur) qui sont pourtant souvent outrepassés par l'avantage important qu'est la libération de l'attention visuelle et la démobilitation de la main et du bras tenant le stylet.

- Sémantique des commandes vocales (Tous)

En ce qui concerne la sémantique des commandes vocales, nous disposons de nombreuses verbalisations des sujets abordant plusieurs points caractérisant cette modalité et qui expliquent en partie les raisons d'adoption ou de rejet de cette modalité.

- Difficultés à mémoriser le vocabulaire (S1, S2, S3, S5, S6 doute à cause des erreurs)

Quasiment tous les sujets ont à un moment ou à un autre exprimé des remarques quant à la difficulté ou tout au moins à l'ambiguïté qui réside dans la mémorisation du vocabulaire de la reconnaissance vocale.

*S1 : J'ai dit "messages reçus", c'était peut-être "aller à messages reçus", je crois que j'ai essayé d'autres choses*

Le sujet 1 manifeste ici un certain doute, une ambiguïté dans la sémantique appropriée pour éviter des erreurs de reconnaissance. Ce doute peut constituer une charge cognitive supplémentaire ou au moins **un sentiment d'incertitude et d'inconfort** suffisant pour ne pas chercher à le rencontrer en évitant d'utiliser la modalité vocale.

Cela est d'autant plus vrai lorsqu'il y a beaucoup d'erreurs qui font douter de la bonne procédure (des « bonnes clés de commandes », « tournures de phrases bien précises », « nombreuses commandes à mémoriser »).

*S3 : Non, c'est peut-être que j'avais pas le mot en tête, c'est vrai, quand j'ai pas le mot, j'ai pas envie de lui dire 36000 et qu'il retrouve pas donc quand je vois que j'ai pas le mot, je zappe (fais une commande tactile stylet)*

Le sujet 3, ici, ne se souvient plus du libellé de la commande (« j'avais pas le mot en tête »), plutôt que d'essayer de manière plus ou moins hasardeuse un autre libellé et pour éviter de se retrouver en situation d'erreur, il préfère directement avoir recours à une autre modalité moins ambiguë (tactile stylet).

*S5 : Les inconvénients ? Des problèmes de mémorisation des commandes vocales. Au départ oui, parce qu'on connaît pas trop mais après, justement, quand on est dans le bain, on s'y fait un peu, non, on trouve ça bien. Au contraire, plus le temps passe et plus on trouve que ya le côté utile*

Le sujet 5 reconnaît avoir eu des difficultés de mémorisation des commandes vocales, mais uniquement au début, qui se sont estompées au fur et à mesure pour laisser place au contraire à un sentiment de facilité.

*Expé : et en terme de mémorisation de ce qu'il ya a dire ?*

*S6 : Ça va, à l'usage, on en a pas beaucoup utilisé si tu regardes. Je pense qu'à l'usage ça vient. Ya des trucs qui sont super intuitifs, d'autres un peu moins, mais c'est un apprentissage, je pense qu'il faut l'apprendre*

Comme le sujet 5, le sujet 6 reconnaît la possibilité d'une difficulté de mémorisation en début d'expérimentation mais relativise par la relative simplicité de l'application et la facilité d'apprentissage de sa manipulation.

Il mentionne également « l'intuitivité » de certains des libellés qui ne nécessitent que peu d'efforts de mémorisation (« répondre », « envoyer », etc.). De plus, une partie des boutons de l'interface possède

des labels qui correspondent aux clés de commandes de la modalité vocale. Encore une fois, la désincarnation de l'interface liée à l'usage de la modalité vocale pose donc des problèmes de mémorisation des actions possibles et surtout de la sémantique appropriée qui y est associée.

- Articulation et prononciation (S1, S6)

Le fait d'interagir vocalement avec une machine provoque des modifications de l'articulation et de la prononciation des sujets qui tentent de s'adapter aux capacités de reconnaissance du système.

*S1 : A chaque fois je parlais comme un robot, parce que je croyais qu'il fallait parler comme un robot si je voulais être compris. Mais je pense que même si je parle normalement, ça comprend bien.*

Comme ici, le sujet 1 a le sentiment (déplaisant) de devoir parler comme un robot pour être compris, il ajoute également qu'il pense que ce n'est pas forcément nécessaire mais qu'il le fait quand même. Ceci dénote une relation particulière à la machine.

*S6 : J'ai essayé de parler un peu plus vite, "sélectionner jaune" apparemment ça a marché, j'en avais marre de parler nanana, (articulé)*

Ici aussi, le sujet 6 explique qu'il a tendance à accentuer son articulation pour être sûr d'être compris, au prix d'un sentiment d'inconfort.

- Modification du phrasé en cas d'erreur (S1, S3, S4, S6)

L'ambiguïté du langage de commande avec des commandes équivalentes peut aussi être source d'une certaine souplesse à la fois pour la correction d'erreur de manière plus élégante mais aussi pour permettre aux sujets de choisir le mot clé qui leur convient le mieux.

*S4 : Quand ça marchait pas (le vocal), je modifiais le phrasé. Souvent ça marche. "messages reçus" ça passe aussi, je sais pas si ya une règle spécifique, donc j'essaye plusieurs phrases*

Ici, le sujet 4 « modifie son phrasé » pour corriger une erreur de la reconnaissance vocale. Il indique également qu'il n'est pas certain de la ou des formulations prescrites et qu'il tente donc plusieurs hypothèses, en prenant le risque de se retrouver à nouveau en situation d'erreur.

- Souplesse de la reconnaissance (S6)

En relation avec les trois points précédents qui abordent plutôt des aspects négatifs de la complexité du langage de commande, le sujet 6 indique que d'un autre côté cette possibilité de choix parfois ambiguë ou complexe procure également une certaine souplesse, une liberté de choix.

*Expé: et pour le vocal en terme de nombre de commandes ?*

*S6 : C'est pareil, à l'utilisation, je pense que ça va, y en a pas tant que ça et puis ya quand même de la marge. Si tu dis un mot en trop, si elle reconnaît quelques mots clés, ça marche*

Il insiste également sur le fait que le système peut faire preuve d'une relative et confortable souplesse dans la reconnaissance partielle ou incomplète d'un certain nombre de termes, suffisants pour la compréhension de la commande.

- Style de langage d'interaction (S2)

Comme mentionné plus haut, s'adresser verbalement à une machine n'est pas neutre, d'autant que le système a des capacités de reconnaissances limitées qui réduisent la souplesse d'usage.

*S2 : Si je pouvais lui parler plus directement, (message), des choses plus brèves, ça fait pas neuneu mais l'impression d'être bête comme si.. que c'est un robot en fait. C'est vrai que parler à un robot, c'est pas un truc que j'assimile bien. En plus c'est vrai c'est une femme qui répond*

Ici le sujet 2 explique qu'il est mal à l'aise du fait de devoir s'adresser « à un robot » qui a un mode d'expression très limité l'obligeant à utiliser un style « simpliste » (« neuneu »). Cette question sera également développée lorsque nous aborderons l'influence du contexte social sur le choix des modalités, en effet, dans ce cas, le sujet déclare que s'exprimer ainsi devant des gens le dérangeait.

- Longueur des commandes (S3)

La longueur des commandes a aussi été mentionnée comme critère de sélection de modalité.

*S3 : Je crois que quand ça plante, ça me décourage, je pars. Mais je suis plus à l'aise pour les petits trucs avec le vocal*

Par exemple ici, le sujet 3 déclare préférer effectuer des commandes courtes avec la modalité vocale (« envoyer », « ouvrir », etc.), et implicitement effectuer les commandes aux libellés plus longs (« sélectionner jaune (destinataire jaune) », « aller à messages reçus ») avec le stylet.

Il est intéressant de noter ici que le choix de modalité est moins dicté par une préférence globale pour la modalité vocale que simplement par la longueur du langage de commande.

Les données quantitatives dénotent quelques contradictions avec ces propos car la commande la plus vocale de ce sujet (« insérer un dessin ») peut être considérée comme complexe et longue. Il faut donc considérer ces propos comme une perception subjective qui ne se traduit pas forcément en actes concrets ou bien que d'autres critères plus importants viennent supplanter celui de la longueur des commandes vocales.

Cette problématique est à rapprocher de la notion de spécialisation à suivre.

#### 8.2.4. Spécialisation (S1)

Lorsque le sujet d'une éventuelle spécialisation de modalité pour certaines commandes a été abordé, il a été difficile d'obtenir des verbalisations conséquentes. Bien que la question ait été systématiquement abordée lors des entretiens, il ne semble pas que les sujets aient vraiment conscience de certaines de leurs spécialisations ou du moins, cela ne ressort pas dans les verbalisations.

Seuls deux sujets (S1 et S3) ont plus ou moins explicité quelques commentaires :

*S1 : (expé : quel type de commandes tu as fait) Commandes vocales ? Plutôt pour me diriger, des « aller à messages reçus », « aller à messages archivés », des choses comme ça, plus, c'était quasiment que pour ça que j'ai utilisé les commandes vocales.*

Le sujet 1, qui a très peu utilisé les commandes vocales, explique qu'il y a eu recours surtout pour changer d'espace dans l'application (« pour me diriger »), perception qui correspond aux données quantitatives puisque les seules commandes vocales qu'il a utilisées étaient : « aller à messages reçus », « aller à messages archivés ».

Quant à d'éventuelles explications sur ces préférences, il a été difficile d'en obtenir. Seul le sujet 1 a mentionné éventuellement une certaine longueur des mots clés comme facteur de spécialisation (voir la partie précédente sur la longueur des mots clés).

*S4i2 : J'ai pas encore le réflexe de faire tout à la voix, transférer, fermer*

*Expé : Tu souhaiterais le faire?*

*Peut-être à force, mais comme aujourd'hui, la voix ça marchait pas, j'ai plus le réflexe d'utiliser le stylet*

## 9. Synthèse des résultats sur l'usage de la multimodalité

### 9.1. Usage des modalités au Niveau global

Au niveau global (tous sujets et toutes sessions confondues), et pour l'ensemble des commandes, la modalité tactile stylet est la plus utilisée avec un ratio de 75%, alors que le recours à la modalité vocale atteint environ 22%. La part du G2D est, quant à elle, quasiment négligeable (1,73%), en partie pour des raisons techniques et de choix d'implémentation (seules quelques commandes d'éditations étaient disponibles).

Ces résultats sont accentués si l'on tient compte uniquement des commandes dites « premières », correspondant aux intentions premières d'usage des modalités par les sujets. Dans ce cas, les commandes de correction, quelque soit la modalité utilisée, ne sont pas comptabilisées. Ceci indique qu'une part des changements de modalité est à imputer à des situations de correction d'erreurs à l'aide d'une modalité non souhaitée en première intention.

De manière générale, nous avons constaté des ratios d'erreurs assez élevés pour l'ensemble des modalités avec pour la modalité vocale, un taux d'erreur de 33%, de 44% pour le G2D et enfin de 8% pour la modalité tactile stylet. Cette faible robustesse, en particulier de la modalité vocale, constitue une première piste d'influence sur le choix d'une modalité d'usage et surtout de non usage d'une modalité.

En termes purement quantitatifs, nous avons observé une très bonne homogénéité dans le nombre de commandes effectuées par sujets au niveau global (toutes sessions confondues), ce qui nous a permis de comparer les résultats entre sujets.

### 9.2. Usage des modalités au Niveau individuel

Une analyse des usages au niveau individuelle permet de relativiser les résultats obtenus au niveau global et fait apparaître de très importantes variabilités inter individuelles dans l'usage des modalités. En effet, les valeurs observées vont de 3,35% pour le sujet ayant le moins recours aux commandes vocales à 47,25% pour celui qui l'a le plus utilisée. Les 6 sujets se répartissent en trois groupes relativement distincts constitués de ceux qui utilisent très peu la modalité vocale, ceux qui y ont recours moyennement et finalement ceux pour qui son usage est majoritaire.

De la même manière que précédemment, ces résultats sont accentués si l'on ne considère que les commandes premières. Dans ce cas, pour les sujets utilisant très peu la modalité vocale, les valeurs d'usage tendent vers 0 indiquant à nouveau un recours contraint à la modalité vocale pour la récupération d'erreurs liées à l'usage de la modalité tactile stylet.

### 9.3. Evolution des usages au Niveau global

A un niveau global (tous sujets confondus) nous avons constaté une augmentation de l'usage de la modalité vocale de 16 à 27,24% indiquant une augmentation de son adoption.

Cependant, cette adoption n'est pas répartie de manière homogène pour tous les sujets et nous avons observé au contraire, de grandes différences inter et intra-individuelles concernant l'évolution des usages des modalités session après session. Deux types de tendances sont apparus :

- ✓ une catégorie de sujets dont les usages sont stables dès les premières sessions,
- ✓ et une catégorie montrant une évolution des usages (en particulier une appropriation de la modalité vocale).

Cette adoption est constituée, en début de session, d'une phase de test informel des propriétés de l'application en général et des modalités en particulier. Celle-ci permet la consolidation des heuristiques mises en place tendant vers une certaine stabilisation des usages. Cependant, le faible nombre de sessions (3 sessions par sujet) ne permet pas d'indiquer de réelles stabilisations des usages.

#### **9.4. Spécialisation et évolution des usages au niveau global**

En ce qui concerne l'usage des commandes, les plus largement utilisées dans notre application sont « ouvrir » et « fermer », puis envoyer.

Nous avons observé de fortes spécialisations pour la modalité tactile au niveau global pour un certain nombre de commandes (« ouvrir », « sélectionner un destinataire », « Bouton Ok » (fermer), « répondre » et enfin « sélectionner un item »). En revanche aucune spécialisation pour la modalité vocale au niveau global n'a été observée, mais de nombreuses tendances à la préférence vocales pour un certain nombre de commandes sont apparues (« insérer dessin », « envoyer », « aller à reçus », « transférer »).

Les résultats n'indiquent aucune tendance globale et forte à la spécialisation session après session. Cependant, nous pouvons tout de même observer une augmentation de l'usage de la modalité vocale pour un certain nombre de commandes telles que « insérer dessin », « transférer », « insérer son », « aller archivés », « ouvrir » indiquant une adoption de cette modalité pour ces commandes, mais cela n'atteignant jamais la spécialisation. Il s'agira donc d'usages préférentiels pouvant être majoritaires dans un cas (« Insérer dessin ») à la dernière session.

Il est également intéressant de noter que certaines commandes ont de forts taux de commandes vocales dès la première session. Celles-ci ont tendance à rester stables pour les commandes « envoyer », « aller reçus », ou à augmenter pour la commande « insérer dessin ».

En revanche, la commande « répondre » qui dès la première session est spécialisée dans la modalité tactile stylet voit se phénomène s'accroître jusqu'à la quasi disparition de la modalité vocale.

#### **9.5. Spécialisation et évolution des usages au niveau intra et inter-individuel**

Au niveau intra-individuel le phénomène de spécialisation est davantage marqué et il y a de grandes différences à la fois inter et intra individuelles dans l'adoption des modalités et dans certaines spécialisations qui en découlent.

Il n'y a donc pas de spécialisation commune, mais de nombreuses tendances à la spécialisation individuelles pour l'une ou l'autre des modalités, mais sans aller jusqu'à la spécialisation totale (non usages d'une autre modalité).

En revanche, il existe un certain nombre de spécialisations pour la modalité tactile stylet, et ceci, parfois de manière partagée par plusieurs sujets (« Sélection des items », « Répondre », « Sélection destinataire »). Pour ces commandes, l'analyse au niveau global reste donc pertinente.

Ce résultat quantitatif n'a pas pu trouver d'explication au niveau qualitatif via les entretiens d'autoconfrontations ou via une analyse fine des propriétés des commandes. Le fait que les sujets ne se spécialisent pas sur les mêmes commandes et les mêmes modalités semble plutôt indiquer des facteurs de choix non liés aux caractéristiques intrinsèques des commandes pour lesquelles une modalité serait éventuellement plus appropriée, et donc l'explication pourrait plutôt tendre vers des préférences individuelles (internes) non partagées entre les sujets.

#### **9.6. Transitions de modalités**

- Transitions au niveau local :

Au niveau local (enchaînement immédiat) et global (tous sujets et toutes sessions), il y a un effet fort de la modalité précédente sur le choix de la modalité suivante, en particulier pour la modalité tactile (à plus de 86%) et dans une moindre mesure pour la modalité vocale (58,73%).

Nous avons observé un renforcement de la modalité vocale lorsqu'il s'agit de corriger une erreur (mode correction) et ceci que la modalité précédente aie été tactile ou vocale.

Ceci indique que les commandes premières, qui correspondent à la première intention de modalité des sujets, ont tendance à avoir des taux de modalité vocale encore plus faibles qu'au niveau global, et à l'inverse, les sujets se trouvant dans une situation d'erreur ont tendance à essayer la modalité vocale pour corriger la situation.

- Transitions au niveau global :

De manière générale, et en cohérence avec les résultats sur les enchaînements locaux de modalité, nous observons une tendance nette au regroupement des usages d'une modalité en « îlots » parfois isolés, en particulier pour les premières sessions.

En ce qui concerne les sujets utilisant très peu la modalité vocale (sujets 1 et 2), ils en font systématiquement usage de manière regroupée en îlots épars de quelques commandes successives. Ces îlots s'observent essentiellement en début de session et dans certains cas particuliers en cours de session pour les cas de correction d'erreur.

Pour les sujets qui vont adopter la modalité vocale, les premières sessions d'interaction sont également ponctuées par des îlots plus réguliers d'usage de la modalité vocale, puis leur répartition devient plus homogène et régulière au cours des sessions suivantes. Pour certains sujets (sujet 6), nous observons tout de même une tendance plus forte à l'usage en séquence de la modalité vocale.

- Verbalisations concernant les transitions :

Les verbalisations font apparaître deux thématiques principales qui sont :

- ✓ **Activation locale et globale des modalités** : La notion d'activation d'une modalité par la modalité précédente a souvent été mentionnée par les sujets pour justifier des enchaînements d'une même modalité en séquence. Mais au-delà de l'activation locale d'une modalité, nous relevons également des verbalisations traitant de phénomènes à plus long terme. Le choix d'une modalité peut donc obéir à des règles supérieures à la simple activation locale d'une modalité par son usage pour la commande précédente, et relève donc d'heuristiques à plus long terme, au cours d'une session voire entre sessions.
- ✓ **Continuité et fluidité de l'interaction** : Comme cela a été en partie mentionné précédemment, la possibilité de continuité de l'interaction dans une même modalité semble être un argument fort pour son éléction. Mais il semble qu'à l'inverse, l'impossibilité d'interagir avec une unique modalité a été pour au moins deux sujets, une des causes importantes de son rejet. Ici le principe d'équivalence fonctionnelle maximale prend toute son importance car s'il n'est pas mis en œuvre, cela oblige le sujet à changer de modalité de manière incessante et contre son gré. Les arguments avancés par les sujets sont essentiellement liés aux différences des styles d'interaction entre les deux modalités. En effet, les contraintes et coût physiques intrinsèques de chacune sont de nature très différente, et obligent donc les utilisateurs à ajuster l'interaction pour chaque changement de modalité.

### 9.7. Etudes des erreurs

Sur la totalité des sessions, 17,31% des commandes (vocales ou tactile stilet) sont suivies d'une erreur. La répartition entre les deux modalités est la suivante :

- ✓ 12,14% d'erreurs pour les commandes vocales
- ✓ 5,18% pour les commandes tactile stilet.

Il y a donc un taux d'erreur trois fois plus élevé pour la modalité vocale par rapport au stilet qui présente une meilleure robustesse.

Il y a de grandes variations du nombre d'erreur selon les commandes (taux d'erreur proche des 30% pour les commandes « Ouvrir », « Aller reçus » et « insérer dessin », et des taux entre 9 et 20% pour les commandes « Bouton OK » (=fermer), « Envoyer » et « transférer »).

Il faut également noter une répartition inégale des erreurs selon les modalités. En effet, seules les commandes « Ouvrir » et « Bouton OK » (=fermer) sont sujettes à des taux d'erreurs similaires avec les deux modalités vocale et tactile stilet. Les erreurs des autres commandes quant à elles, sont quasiment exclusivement liées à la modalité vocale.

Nous avons observé un total de 53,33% d'erreur sur la totalité des commandes vocales (parmi les erreurs de la modalité vocale, 80,08% sont des erreurs de reconnaissances liées au système lui-même alors que 19,92% des erreurs sont liées à des erreurs d'utilisation de la reconnaissance vocale par les sujets).

Comme pour la modalité vocale, la source principale d'erreurs Tactile provient du système (44,34%) et dans une moindre mesure de l'utilisateur (15,57%).

Ce résultat montre également la concentration des erreurs de la modalité tactile stilet sur une commande quasi unique et ceci autant pour les erreurs systèmes que pour les erreurs utilisateurs (il s'agit de la commande ouvrir qui requiert un double clic au stilet qui est problématique).

### 9.8. Préférences individuelles des modalités :

Les préférences individuelles et l'adoption par les sujets des modalités sont liées en premier lieu aux propriétés intrinsèques des modalités qui vont déterminer leur expérience. Ceux-ci vont se forger au fil du temps des préférences d'usages qui vont modeler leur style d'interaction propre. Ce phénomène apparaît dans les verbalisations sous deux grandes thématiques : une phase de test et d'appropriation de l'application, et la vérification de l'adéquation de l'application aux exigences de la tâche.

- test et appropriation de l'application

Cette étude vient confirmer une étude précédente (Calvet, 2001) dans laquelle nous avons abordé la problématique des premiers usages d'une application en identifiant une phase de test informel et de familiarisation avec l'application qui se traduisait par des usages des modalités non stabilisés dans un premier temps. L'ensemble des facteurs suivants a été identifié comme justifiant du choix d'usage d'une modalité :

- Test de l'application et des modalités
- Moments de test
- Aspect ludique, nouveauté
- Contexte expérimental
- Test de la mise en œuvre des modalités
- Test de l'efficacité
  - Rapidité VS. Lenteur : contrôle de l'application
  - Association du vocal et G2D à des raccourcis

- Exigences de la tâche

Les critères d'évaluation de l'application et de ses modalités sont relatifs aux tâches que les sujets doivent exécuter. Dans ce cadre, plusieurs déterminants ont été mentionnés lors des entretiens. Ceux-ci ont été regroupés dans 4 thématiques :

- **Rapidité** : Comme expliqué dans la partie précédente (test de l'application), la rapidité est un critère déterminant pour l'adoption d'une modalité. Mais ici, nous abordons ce critère sous la forme de l'adéquation de la rapidité de la modalité proposée aux exigences de la tâche demandée. Les quatre facteurs suivants ont été mentionnés :
  - la pression temporelle et sociale,
  - la charge de travail,
  - la comparaison entre modalités
  - et la priorité entre les tâches (possibilité d'interruption, possibilité et/ou facilité de reprise de cette tâche, caractère obligatoire ou urgent de la tâche, inscription de la tâche dans une séquence unitaire à compléter).
- **Criticité** : Le critère de criticité est mentionné par certains auteurs (Karsenty, 2006), mais étant donné le caractère ludique et expérimental de notre étude, ce critère ne semble pas avoir été pris en compte par les sujets et nous ne disposons pas de verbalisations à ce propos.
- **Effort de mise en œuvre** : L'effort de mise en œuvre d'une modalité s'inscrit ici non seulement dans les qualités intrinsèques des modalités mais aussi dans leur insertion dans une séquence d'activité mettant éventuellement en œuvre d'autres modalités ou actions physiques et cognitives.
  - Séquence d'usage et activation de modalité (idem que pour les transitions)
  - Libération de l'attention et de la mobilisation des membres (traité en détail dans la partie mobilité)
  - Sémantique des commandes vocales (Difficultés à mémoriser le vocabulaire, Articulation et prononciation, Modification du phrasé en cas d'erreur, Souplesse de la reconnaissance, Style de langage d'interaction, Longueur des commandes)
- **Spécialisation** : La spécialisation pourrait être appréhendée ici comme une autoassignation récurrente par les sujets de l'usage d'une modalité spécifique pour une commande donnée (nous considérons la présence d'une spécialisation au delà de 80% d'usage de cette modalité). Il ne semble pas que les sujets aient vraiment conscience de

certaines de leurs spécialisations et malgré un questionnement systématique, les raisons des choix d'usage ne ressortent pas des verbalisations.

### **9.9. Conclusion**

Pour conclure sur l'usage des modalités, il semble donc que les sujets ont des tendances à des usages préférentiels variés et non systématiquement partagés au niveau inter-individuel, ce qui tend vers une hypothèse de facteurs internes fortement subjectifs pour les choix d'usage de modalités. En effet, les propriétés intrinsèques des modalités strictement objectives, étudiées ici, dans le cadre d'une stricte équivalence fonctionnelle, semblent être perçues de manière très variable selon les sujets expliquant les fortes différences de préférences allant parfois jusqu'à la spécialisation. Comme nous le verrons par la suite, un certain nombre de facteurs contextuels vont également entrer en jeu pour expliquer les choix d'usage des modalités.

**Chapitre 4. Facteurs contextuels influençant  
l'usage des modalités**

## 1. Introduction

Comme nous venons de le voir dans le chapitre précédent, le choix d'usage d'une modalité repose en partie sur des préférences individuelles qui correspondent aux modalités « par défaut » que l'utilisateur aura tendance à utiliser (en fonction également du contexte).

Cependant, de nombreuses publications émettent l'hypothèse que le contexte d'usage d'une application (la situation) va également venir jouer un rôle important dans le choix des modalités.

Dans cette partie, nous identifions quelques éléments constitutifs du contexte et essayons de déterminer leur influence sur l'usage des modalités. Nous avons retenu quatre types d'observables liés au contexte :

- Le niveau sonore
- L'environnement social
- Le type d'activité en cours
- L'historique de l'interaction

Tout en suivant cet ordre de questionnement nous procéderons pour chaque thématique à une présentation quantitative des résultats, puis nous irons interroger systématiquement la perception qu'ont eu les sujets quant à l'influence de ces facteurs sur leur usage des modalités.

Nous achevons ensuite ce chapitre par une rapide synthèse des résultats.

## 2. Influence du niveau sonore sur l'usage des modalités

Il est souvent mentionné dans les études que le niveau sonore d'une situation pourrait influencer dans le sens du non-usage de modalités ou de médias sonores telles que l'enregistrement de messages vocaux (audio) ou vocales. Cependant, aucune étude en situation réelle, à notre connaissance, n'a démontré cette hypothèse. De plus, la dimension subjective de la perception du niveau sonore n'est pas non plus prise en compte.

Dans un premier temps, nous distinguons ici deux niveaux d'ambiance sonore qui sont relativement simplistes : « bruyant » et « non bruyant » afin de dégager des tendances du point de vue quantitatif sur l'effet de ces ambiances, que ce soit aussi bien à un niveau global qu'individuel.

Dans un second temps, nous irons interroger la perception qu'ont eu les sujets quant au niveau sonore et à son influence sur les usages des modalités.

Les questions abordées ici et auxquelles nous tenterons de répondre sont les suivantes :

***Le niveau sonore « objectif » influence-t-il le choix des modalités ?***

***Le niveau sonore « perçu » influence-t-il le choix des modalités ?***

***Y a-t-il des différences inter-individuelles dans ces influences ?***

### 2.1. Influence du niveau sonore sur les erreurs

Nous commencerons d'abord par une rapide étude de l'influence du niveau sonore sur la quantité d'erreurs de la modalité vocale, en effet, le contexte sonore est souvent incriminé lorsque les taux d'erreur de reconnaissance sont élevés.

Dans le tableau suivant nous ne retenons donc que les erreurs de reconnaissance liées au système.

Modalité	Contexte Sonore	Erreur Reco	Succès	Total
Vocale	Bruyant	48,94%	51,06%	100,00%
	Pas bruyant	46,81%	53,19%	100,00%
Total		47,78%	52,22%	100,00%

**Tableau 11 : Taux d'erreur de la modalité vocale en milieu bruyant ou non**

Les résultats du Tableau 11 indiquent que le niveau sonore ne semble pas avoir ici d'influence sur le taux d'erreur de la modalité vocale, en effet les taux d'erreur dans les deux types de contexte sont très proches (46,81% en milieu non bruyant et 48,94% en milieu bruyant).

Du point de vue des sujets, il conviendra également de vérifier si le niveau sonore et la perception d'un risque d'erreur, est un critère déterminant ou non pour le choix ou le rejet de la modalité vocale.

### 2.2. Influence globale du niveau sonore

Le contexte de l'expérimentation étant celui de la Villette, de nombreuses installations muséales sont susceptibles de générer de fortes variations dans les ambiances sonores, indépendamment même des déplacements des sujets. En effet, de nombreuses installations sont soit vidéo ou audio et tournent en boucles générant ainsi dans un endroit donné des variations très importantes.

Il est donc utile dans un premier temps de se demander si les sujets interagissent plus ou moins selon le contexte sonore indépendamment de la question des usages des différentes modalités.

	Total	%(t)	% Nb de commandes
Bruyant	07:45:27	60	53,13
Pas bruyant	05:06:21	39,5	46,87

**Tableau 12 Niveau sonore global**

Le Tableau 12 indique que 60 % du temps global d'expérimentation s'est déroulé dans un environnement bruyant, voir très bruyant.

Quant au nombre de commandes effectuées dans ces deux conditions, la différence est moindre puisque l'on a 53,13% seulement de commandes dans un environnement bruyant et 46,87% dans un environnement non bruyant.

Il semble donc que les sujets ont moins interagi dans les contextes bruyants sans que pour l'instant nous n'ayons d'explications à ce résultat.

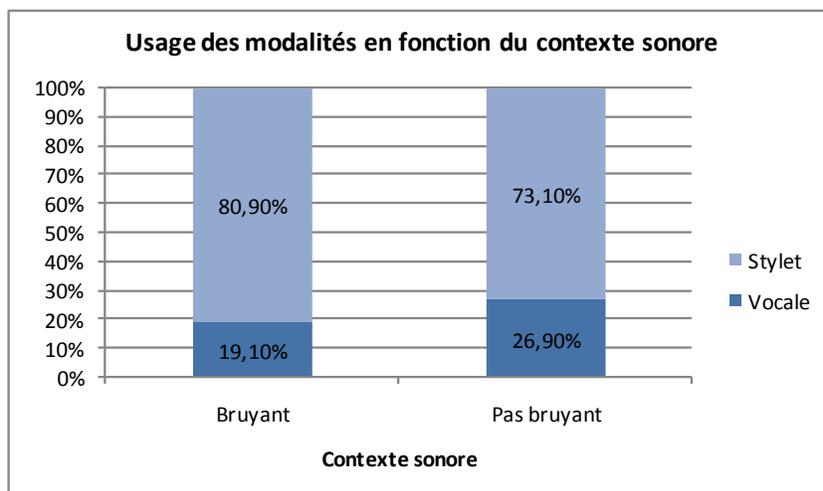
Ces premiers résultats soulèvent cependant un certain nombre d'interrogations telles que :

**Les sujets s'arrangent-ils pour faire moins d'interactions en milieu bruyant ?**  
**Question correspondant à l'hypothèse adaptative : assimilation, activité dirigée par les ressources.**

**Les sujets essaient-ils de se placer dans des contextes moins bruyants pour interagir ?**  
**Question correspondant à l'hypothèse adaptative : accommodation, activité dirigée par les exigences, maîtrise de la situation.**

Nous tenterons de répondre en partie à ces interrogations ultérieurement à travers une analyse qualitative via l'étude des verbalisations.

Mais d'abord, regardons l'effet du niveau sonore sur les choix d'usage de modalité vocale ou tactile.



**Figure 43 Effet de l'environnement sonore sur l'usage des modalités pour l'ensemble des sujets**

Au niveau global, le contexte sonore semble avoir un effet sur le choix d'usage des modalités. La modalité tactile stylet qui est utilisée à 80,90% dans des conditions bruyantes, passe à 73,10% dans des situations non bruyantes, et inversement pour la modalité vocale, qui elle, passe de 19,10% à 26,90%.

Le contexte bruyant semble donc ne pas favoriser l'usage des modalités vocales.

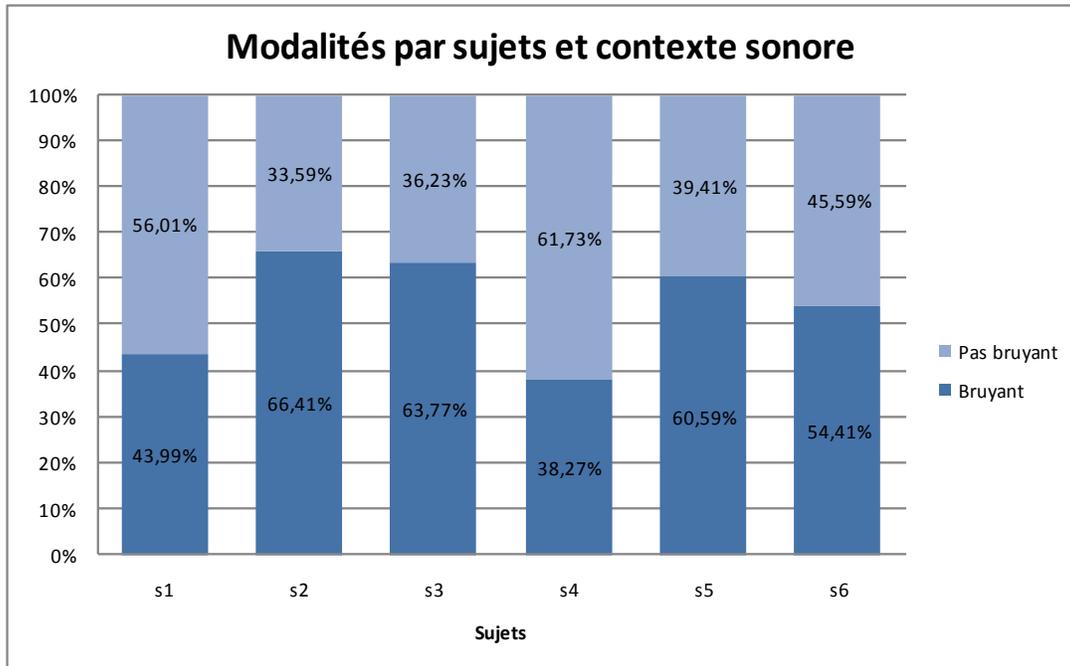
### 2.3. Influence individuelle du niveau sonore

Après avoir vu l'effet du contexte sonore à un niveau global, il convient maintenant d'examiner cette question au niveau individuel en répondant à l'interrogation suivante :

**Cette influence du niveau sonore au niveau global se retrouve-t-elle au niveau individuel ?**

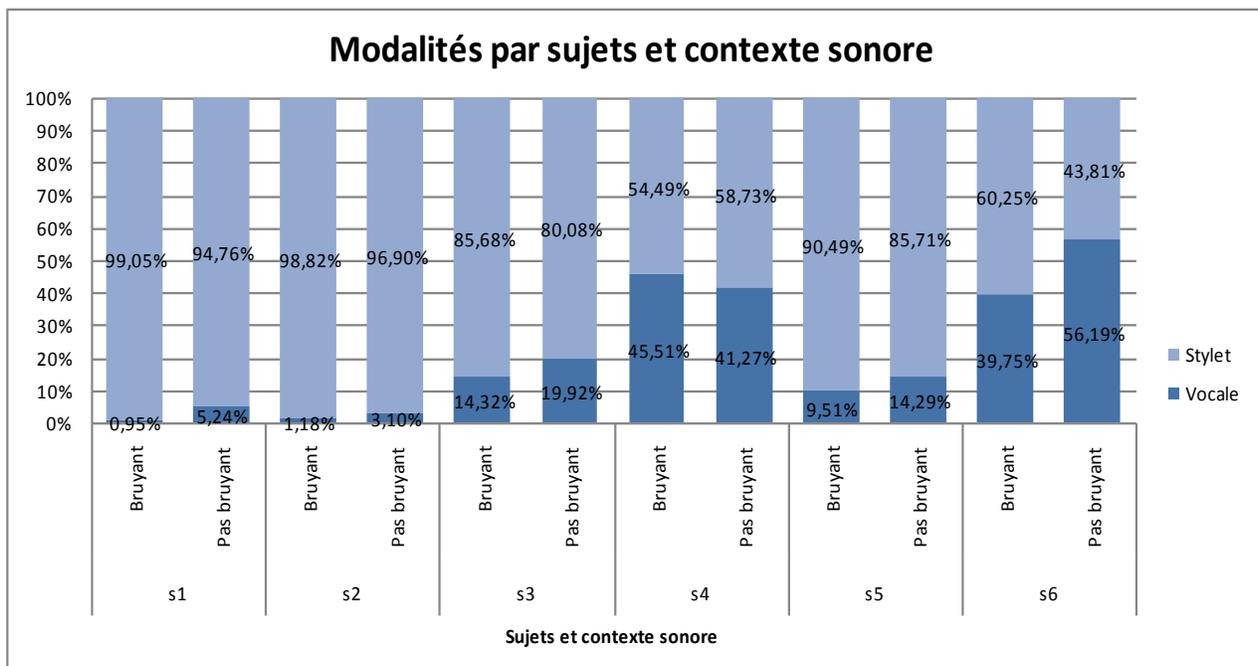
Afin de rendre compte de l'ambiance sonore dans laquelle a évolué chaque sujet, nous commençons par étudier le ratio du nombre de commandes selon le contexte sonore par sujet.

## Facteurs contextuels influençant l'usage des modalités



**Figure 44 : Ratio du nombre de commandes selon le contexte sonore par sujet**

Au niveau individuel, le nombre de commandes effectuées dans un contexte bruyant ou non diffère (Figure 44). En effet, on passe de 40% de commandes émises en milieu bruyant pour le sujet 4, à 65% pour le sujet 2. Seuls les sujets 1 et 4 sont en dessous des 50%, alors que les autres sujets sont plutôt autour des 60%. Il s'agit donc globalement d'ambiances sonores plutôt bruyantes (ce qui est cohérent avec les durées des niveaux sonores).



**Figure 45 : Influence du contexte sonore au niveau individuel**

En accord avec les résultats précédents, la Figure 45 et le Tableau 13, montrent une certaine tendance des sujets à émettre moins de commandes vocales que de commandes tactiles en milieu bruyant. Cet

écart n'est pas indicatif pour les sujets 1 et 2 car les effectifs de commandes vocales sont très faibles (24 et 7 respectivement) mais reste tout de même cohérent avec le niveau global. Pour les sujets 3 et 5 qui ont beaucoup utilisé les commandes vocales, l'écart est moyen (respectivement 5,60% et 4,77%) mais reste également cohérent avec les résultats globaux.

Commandes Vocales	Bruyant	Pas bruyant	Différence
S1	0,95%	5,24%	4,28%
S2	1,18%	3,10%	1,92%
S3	14,32%	19,92%	5,60%
S4	45,51%	41,27%	-4,24%
S5	9,51%	14,29%	4,77%
S6	39,75%	56,19%	16,45%

**Tableau 13 : Différences, par sujets, du nombre de commandes vocales selon le contexte sonore**

L'écart le plus important concerne le sujet 6, qui a le plus utilisé les commandes vocales et pour qui l'influence du contexte sonore semble prépondérant (16,45% d'écart entre le milieu bruyant et non bruyant).

Quant au sujet 4, les valeurs sont inversées par rapport aux autres sujets, ce qui peut s'expliquer entre autre par le fait qu'il a un nombre très important de commandes effectuées dans un environnement non bruyant (plus de 60%).

Ces résultats doivent être mis en perspective avec la perception qu'ont les sujets de cet environnement.

#### 2.4. Perception subjective du niveau sonore

Lors des entretiens d'autoconfrontations, le sujet de la perception du niveau sonore a systématiquement été abordé. Le plus grand nombre de verbalisations concerne la relation entre le contexte sonore et l'éventuel usage ou non-usage de la modalité vocale. Nous avons rassemblé ces verbalisations sous trois thèmes :

- Les gênes perçues dues au bruit
- Les modifications de modes opératoires liés au bruit
- Un thème divers

##### 2.4.1. Gêne due au bruit (S3, S4, S6)

De manière générale, le bruit ambiant peut-être perçu comme un élément perturbateur, mais sa perception semble relativement inégale et fortement contextualisée selon les sujets, la tâche en cours etc.

- Crainte d'erreurs potentielles dues au Bruit (S3, S4, S6)

En premier lieu, le bruit peut être perçu par certains sujets comme une source potentielle d'erreur du système de reconnaissance vocale même si les résultats quantitatifs sont venus infirmer cette crainte.

*S3 : (Suite à plusieurs erreurs vocales) Non, j'étais plus vénère qu'autre chose, mais non, je me suis dit que c'était soit par rapport au bruit qu'il y avait autour, qu'il fallait peut-être que je parle plus fort*

Dans cet exemple, suite à des erreurs successives de la modalité vocale, le sujet 3 essaye de diagnostiquer les causes d'erreur et envisage le niveau sonore comme raison principale. Il modifie le volume sonore de sa commande vocale pour pallier le fort bruit, sans changer de modalité.

Ici, il s'agit d'un diagnostique et d'une prise de conscience (constat) **a posteriori**, lorsque l'erreur s'est produite.

- Changement de modalité en anticipation ou suite à des erreurs avec bruit (S4)

Dans certains cas, proche du cas précédemment évoqué, les sujets vont aller jusqu'à changer de modalité pour pallier des erreurs de reconnaissance vocale qu'ils ont perçues comme liées au niveau sonore élevé.

- Changement de modalité due au bruit (S3, S4, S6)

Plusieurs sujets (3, 4 et 6) reconnaissent avoir anticipé et pris en compte le niveau sonore de l'environnement pour effectuer un choix de modalité de commande. Dans ce cas, il s'agit d'une prise en compte du contexte sonore **a priori**, afin d'anticiper un risque d'erreur.

*S6 : (expé : là, tu insères un son en stylet?) S6 : Je sais pas pourquoi, enfin si ça fait du bruit quand même dans ce truc et qu'inconsciemment je me dis, on va peut-être pas l'entendre, donc je le fais au stylet*

Le sujet 6 utilise beaucoup la modalité vocale, en particulier pour la commande vocale concernée ici (« insérer un son »), mais dans ce cas, il préfère utiliser le stylet en raison du bruit environnant. Il indique également que ce choix n'est pas forcément fait de manière consciente. Cette déclaration est tout à fait conforme aux résultats quantitatifs puis qu'il est le sujet qui a été le plus sensible au niveau sonore pour le choix de la modalité vocale (plus de 16% de différence de commandes vocales entre le milieu bruyant et non bruyant).

*S3 : Mais, si je suis où ya pas de bruit, la reco, c'est vrai ce serait très bien*

Pour le sujet 3, le bruit semble également être perçu comme un déterminant du choix de modalité, mais dans une moindre mesure selon les résultats quantitatifs (5,60% de différence).

*S4 : Pour l'application elle-même, pour changer de page comme ça, c'est vrai que je me pose pas la question, (mais c'est vrai que pour les réponses notamment, quand j'enregistre ma voix, je vais faire un peu gaffe au niveau sonore, parce que c'est un peu le bordel)*

Le sujet 4, au contraire, ne semble pas être sensible au niveau sonore pour choisir sa modalité d'interaction (« je me pose pas la question »), ce qui correspond aux résultats quantitatifs puisqu'il semble avoir fait plus de commandes vocales dans un environnement bruyant. Cependant, il indique percevoir le bruit mais ne pas être dérangé pour effectuer des commandes vocales, mais il introduit une nuance en distinguant bien la commande du dialogue homme machine qui n'est destinée qu'au système et l'enregistrement d'un message vocal (message audio), qui lui sera écouté par son partenaire et dont la qualité sera évaluée. Cette remarque sera étudiée dans la partie traitant de l'influence du contexte social et semble indiquer un filtrage ou une orientation de la perception du bruit en fonction des exigences de la tâche.

- Stratégie d'anticipation (S6)

Une des questions posées précédemment était de savoir si les sujets avaient ou non tendance à établir des stratégies d'anticipation afin d'éviter des situations d'erreur liées au bruit ou si tout simplement ils évitaient d'interagir en situation bruyante.

Nous avons dans l'exemple suivant un début de réponse pour un sujet.

*Expé: donc tu avais préparé en vocal ton message*

*S6 : Je savais qu'il y avait du bruit, si ça se trouve, ça aurait très bien marché. Je sais pas, en y allant, je me suis dit, hop "nouveau msg", "insérer un son"*

Le sujet 6, dans un souci d'anticipation de changement de niveau sonore et dans la crainte d'une erreur potentielle de reconnaissance, préfère préparer en avance son message en conservant son mode opératoire habituel (commandes vocales) avant de rejoindre le milieu bruyant. Il s'agit bien là d'une stratégie d'anticipation qui témoigne d'un certain degré de perception du niveau sonore actuel et à venir, et d'une stratégie d'accommodation de l'interaction à son contexte.

## Facteurs contextuels influençant l'usage des modalités

- Stratégie de retrait du milieu bruyant pour effectuer une commande vocale (S4)

Dans le même ordre d'idée que le point précédent, nous avons pu observer un sujet qui a attendu d'avoir effectué un retrait du milieu bruyant avant de passer une commande vocale mais aucune verbalisation n'est disponible à ce sujet.

Dans ce cas, il s'agirait plus d'un exemple d'assimilation, avec une activité plutôt dirigée par les ressources.

### 2.4.2. Modification du mode opératoire (S6)

Que ce soit en situation nominale ou en situation d'erreur, les sujets peuvent procéder à des modifications de leur mode opératoire habituel en raison du niveau sonore élevé.

- Augmentation du volume de la parole (S6)

*S6 : (suite à une erreur en milieu bruyant, le sujet répète la commande en parlant plus fort) Je me suis pas énervé mais y avait du bruit autour donc je pensais ... Je me dis peut-être qu'elle va moins comprendre mais tu verras, à d'autres moments ya du bruit, je parle quand même*

Ici, le sujet va répéter sa commande en parlant plus fort en espérant être mieux compris de l'application. Il s'agit ici d'une modification du mode opératoire **a posteriori**, suite à la constatation d'une erreur, mais il a pu également être observé (pas de verbalisation disponible) des augmentations du volume de la voix pour effectuer des commandes vocales en milieu bruyant (**a priori**).

- Rapprochement du pda près de la bouche (tous sujets)

Une autre possibilité de modification du mode opératoire est le rapprochement du PDA (et donc du micro) près de la bouche pour tenter de limiter les risques d'erreur en milieu bruyant. Il s'agit d'une sorte de mode préventif, souvent réflexe qui a souvent été observé.

*S6 : (rapproche le pda de la bouche pour une commande vocale) Je sais ya pas besoin. Après au fur et à mesure de l'expé, je le rapproche de moins en moins. Peut-être un peu plus cette fois comme ya vachement de bruits autour, j'ai tendance à le rapprocher*

Le sujet 6 indique qu'il est conscient que cette manœuvre n'est pas forcément nécessaire, et qu'il va la faire de moins en moins au fur et à mesure qu'il constate le bon fonctionnement de la reconnaissance en milieu non bruyant. Cependant, il pense que dans ce cas, le milieu bruyant a pu l'inciter à rapprocher davantage le PDA de la bouche.

### 2.4.3. Divers

- Appropriation de l'application (S6)

Le degré d'expérience d'usage de l'application peut lui aussi venir modifier la perception et l'influence du bruit sur le mode opératoire associé à la modalité vocale.

*S6 : Je pense que c'est dû au bruit. Mais aussi, j'm'en sers pas tout les jours, comme c'est un truc ou tu parles, t'as tendance à rapprocher mais de moins en moins au fur et à mesure. Mais comme ya du bruit, je me dis toujours est ce que ça va marcher*

Pour faire suite aux constatations précédentes, il faut noter également des modifications de modes opératoires liés à l'appropriation de l'application (« t'as tendance à rapprocher mais de moins en moins au fur et à mesure »). Cependant, le déterminant bruit reste apparemment une préoccupation constante poussant parfois au retour à l'ancien mode opératoire (rapprochement du PDA de la bouche).

- Perception a posteriori du niveau sonore (durant les autoconfrontations)

## Facteurs contextuels influençant l'usage des modalités

De manière générale, nous avons pu constater que les sujets n'étaient pas forcément très sensibles au niveau sonore ambiant. Plusieurs verbalisations attestent du fait que la constatation de bruit était faite a posteriori lors du visionnage des vidéos. Le contexte sonore de La Cité de La Science étant globalement plutôt bruyant, les sujets semblent s'être accoutumés au point d'une certaine désensibilisation, voire indifférence pour certains.

- Focalisation sur l'interaction et perte relative du contexte sonore

Un autre phénomène vient renforcer et expliquer partiellement le point précédent. Il s'agit de la focalisation des sujets sur des tâches bien précises et parfois requérant une forte mobilisation attentionnelle qui provoque un effet tunnel où le sujet semble isolé de son environnement.

S6 (Doit envoyer un message à son partenaire et effectuer la tâche, d'où focalisation)

### 3. Influence de l'environnement social sur l'usage des modalités

Le comportement social attendu dans un contexte comme la Villette est d'une nature particulière puisqu'il s'agit d'un lieu public relativement fréquenté par un public très varié. Il s'agit d'un environnement focalisé sur la science dans lequel les technologies sont omniprésentes. L'atmosphère générale est plutôt bruyante et l'ambiance n'est pas particulièrement orientée vers une contemplation calme des expositions, mais plutôt à une participation active des visiteurs, ce qui pourrait différencier quelque peu ce lieu d'un musée classique où le silence et le calme sont souvent de rigueur.

Bien évidemment, il convient de rappeler que les sujets sont soumis à un contexte expérimental biaisant leur comportement social naturel.

Les questions abordées ici seront les suivantes :

***L'environnement social « objectif » influence-t-il le choix des modalités ?***

***L'environnement social « perçu » influence-t-il le choix des modalités ?***

***Y a-t-il des différences inter-individuelles dans les influences du contexte social ?***

L'environnement social est pris ici dans son acception la plus basique et ne prend absolument pas en compte l'engagement des sujets dans une interaction sociale quelconque. De plus, le contexte de visite d'exposition de la Villette génère majoritairement un certain type d'environnement social, plus proche de la coprésence que d'un véritable engagement social. Ici, donc, nous ne différencierons que la présence ou non de tiers dans l'environnement du sujet. L'expérimentateur lorsqu'il n'interagit pas avec le sujet sera codé comme laissant le sujet seul. Nous ne dissocions pas non plus le type d'interactions que l'on pourrait trouver avec les visiteurs, de fait, il y a très peu d'interactions explicites entre les sujets et d'éventuels visiteurs.

#### 3.1. Influence globale de l'environnement social

Nous retrouvons donc uniquement deux codages de l'environnement social :

- « **seul** » : Le sujet ne peut percevoir aucune présence dans son environnement en dehors de celle de l'expérimentateur.
- « **monde** » : Le sujet peut percevoir une présence visuellement ou de manière sonore (voix, bruits etc.).

La notion de distance n'est pas prise en compte dans le codage.

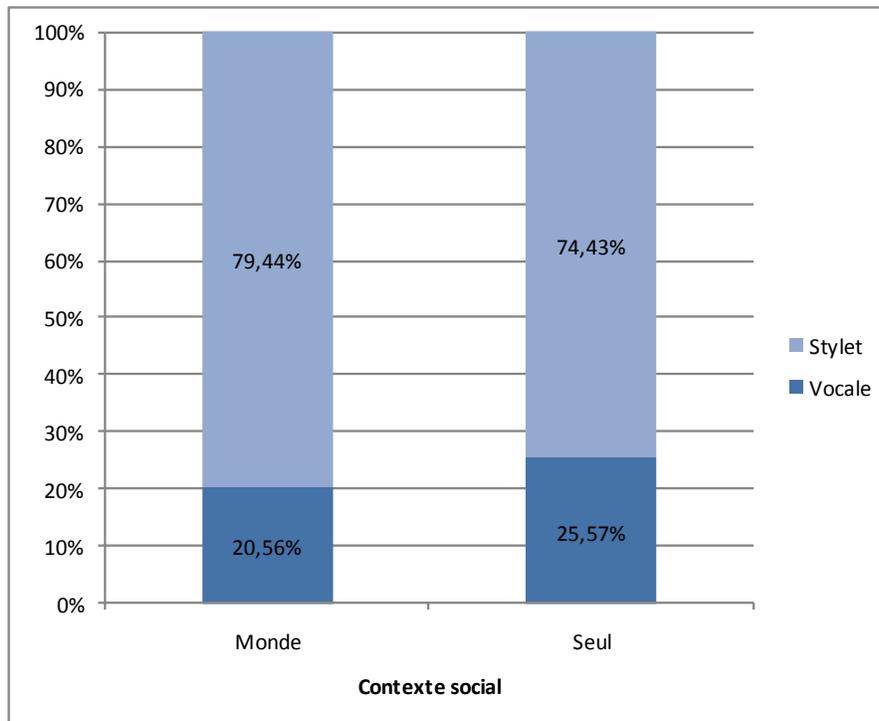
Comme nous l'avons fait pour le contexte sonore, nous commençons par étudier l'ambiance sociale générale dans lesquels les sujets ont évolués, ainsi que le nombre d'interactions correspondant à chaque contexte.

	% de la durée des sessions	% du nombre de commandes
Monde	64,1 %	58,22 %
Seul	35,2 %	41,78 %

**Tableau 14 : Environnement social des expérimentations : durée et nombre de commandes**

Sur la totalité de la durée des sessions, 64,1% du temps correspond à des situations où les sujets sont en présence de tiers (Tableau 14), durant lesquelles 58,22% des commandes ont été émises. Ces ratios sont assez similaires et permettent ainsi la comparaison du nombre de commandes vocales et tactile stylet selon les différents contextes sociaux.

## Facteurs contextuels influençant l'usage des modalités



**Figure 46 : Effet du contexte social sur l'usage des modalités pour l'ensemble des sujets**

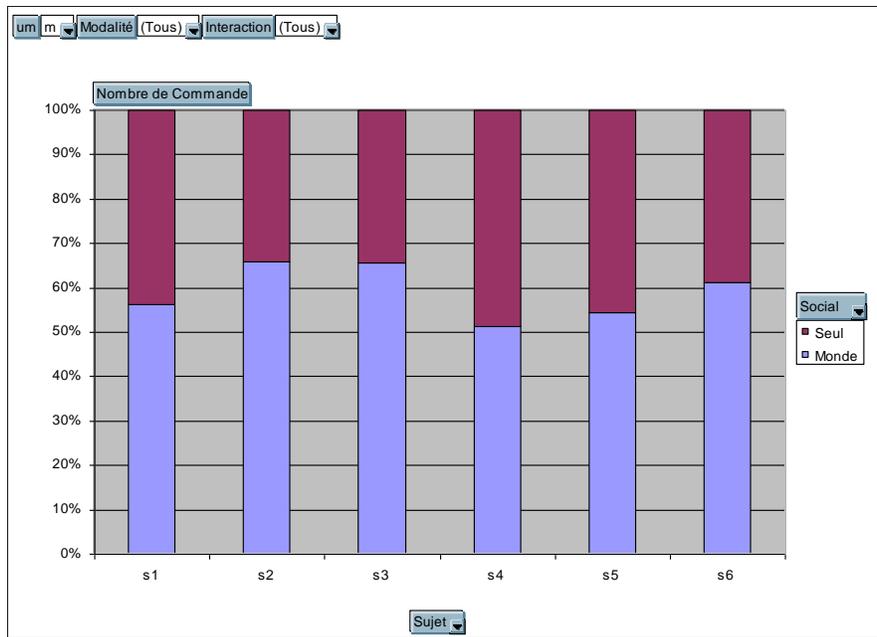
Au vu de la Figure 46, l'effet du contexte social, au niveau global, semble moins important que celui du contexte sonore. L'écart d'usage de la modalité vocale en situation « monde » ou « seul » est à peine de 5% (de 20,56% à 25,57%) et ne permet donc pas de tirer de conclusions particulières en dehors de l'absence d'un effet massif au niveau global du contexte social.

### **3.2. Influence au niveau individuel de l'environnement social**

Comme pour les autres données observées, il convient de regarder les résultats au niveau individuel, d'autant plus que la perception de la présence d'autrui peut être très subjective comme nous le verrons par la suite.

Pour avoir une idée du contexte social d'interaction de chaque sujet, nous commençons par comparer le nombre de commande émise selon les deux valeurs de contexte retenues (seul, monde).

## Facteurs contextuels influençant l'usage des modalités

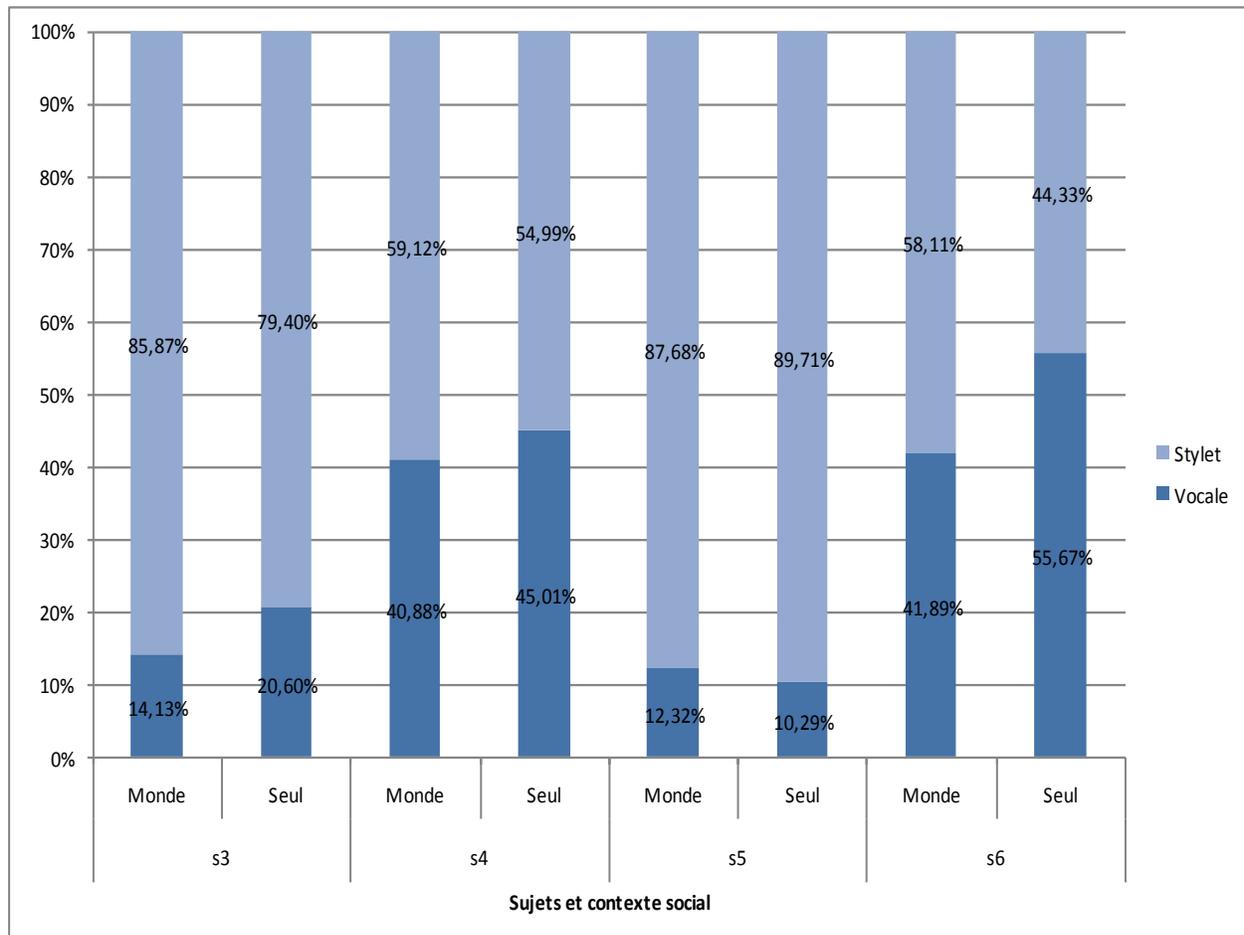


**Figure 47 : Ratio du nombre de commandes selon le contexte social par sujet**

Au niveau individuel, il y a globalement peu d'écart entre les sujets pour le nombre de commandes selon l'environnement social. Toujours au delà des 50%, les chiffres vont de 51 % à 65 % de commandes en environnement « monde »

Intéressons-nous maintenant à la répartition des modalités au niveau individuel selon les contextes sociaux.

## Facteurs contextuels influençant l'usage des modalités



**Figure 48 : Ratio ( ?) du nombre de commandes selon le contexte social par sujet**

Les valeurs ne sont pas significatives pour les sujets 1 et 2 qui n'ont quasiment pas utilisé de commandes vocales et ne sont donc pas présentées ici.

Les résultats au niveau individuel sont relativement conformes aux résultats du niveau global. L'effet du contexte social sur l'usage des modalités semble faible, en particulier pour les sujets 4 et 5 (Tableau 15). En revanche, les écarts de nombres de commandes vocales en contexte social « monde » ou « seul » (Tableau 15) sont plus importants pour les sujets 3 et 6 (6,48% et 13,78% respectivement).

Commandes Vocales	Monde	Seul	Différence
s3	14,13%	20,60%	6,48%
s4	40,88%	45,01%	4,13%
s5	12,32%	10,29%	-2,02%
s6	41,89%	55,67%	13,78%

**Tableau 15 : Différences, par sujets, du nombre de commandes vocales selon le contexte social**

Il est donc possible de conclure à d'éventuelles différences inter individuelles concernant l'effet du contexte social sur le choix des modalités d'interaction.

Dans la suite de l'exposé, il conviendra d'essayer de trouver des explications à ces différences d'usages à l'aide des verbalisations de chaque sujet concernant cette question.

### 3.3. Perception subjective du contexte social

La perception subjective du contexte social pourrait en partie expliquer les variations interindividuelles sur les raisons d'usage ou de non usage d'une modalité, et en particulier en ce qui concerne la modalité vocale pour laquelle les commentaires sont les plus nombreux à ce sujet. En effet, les verbalisations concernant l'interaction tactile stylet en présence de monde sont très peu nombreuses, et semblent indiquer une banalisation de ce type d'usages.

Verbalisations concernant l'interaction vocale et le contexte social :

Ces verbalisations sont regroupées dans deux parties, l'une rassemblant les expressions de gêne liée à l'interaction vocale en présence de tiers, l'autre à son absence.

#### 3.3.1. Expression de gêne liée au monde

Les sujets 1, 2 et 6 sont ceux qui ont manifesté le plus de sensibilité au contexte social quant à l'usage de la reconnaissance vocale et ceci pour des raisons diverses explicitées ci-dessous de manière thématisée. D'une manière générale, pour les sujets 1 et 2, la timidité sous des formes variées est un argument qui est souvent avancé comme cause de non usage de la modalité vocale.

Mais il peut y avoir également des gênes qui ne sont pas de l'ordre de la timidité, par exemple la crainte de déranger ou encore la crainte d'erreurs potentielles de la reconnaissance vocale liées au bruit produit par des tiers.

- Inhibition (Timidité) (S1, S2)

Les sujets 1 et 2 déclarent explicitement leur timidité, et la gêne que leur procure le fait de devoir faire des commandes vocales devant un tiers. Cette gêne se décline en plusieurs types :

- Difficultés à parler (seul) à une machine devant des gens (S1, S2)

*S1 : Le vocal, c'est vraiment trop laborieux par rapport au stylet. Le stylet, je suis plus habitué, et là le vocal, j'sais pas quand ya des gens et tout je trouve que... c'est bizarre devant les gens de parler*

*(expé : ça te dérange ?). Non ça me dérange pas, mais je trouve que c'est pas un contact naturel, ya vraiment rien de naturel, je suis comme ça (se penche ?) à moitié sur le truc, j'sais jamais si ça va marcher, j'sais pas, j'aimais pas trop le.., j'aurai eu un kit main libre justement comme on le disait, j'aurai peut-être eu moins de réticence à le faire, limite j'aurai pu le ranger dans ma poche*

Parmi les raisons que cite le sujet 1, se trouve le sentiment de malaise de devoir parler à une machine devant des gens, qui dans son cas a été un critère quasi rédhibitoire quant à l'usage de la modalité vocale puisque celle-ci représente à peine 3,35% de l'ensemble des commandes effectuées.

*S2 : ben à part pour la commande vocale que j'ai pas... pas pour la commande vocale, mais pour enregistrer un message vocal, mais ça aurait été la même chose si j'avais eu à utiliser la commande vocale, le fait qu'il y ait des gens, je trouve que c'est un handicap. Après, ya des gens qui sont moins timides, mais moi, personnellement, ça me gêne. Je pense qu'il ya beaucoup de gens qui sont dans mon cas*

Il en est de même pour le sujet 2 qui lui aussi déclare être timide et ne pas pouvoir effectuer de commandes vocales devant un tiers. Cela se traduit également par un très faible taux de commandes vocales (1,82%).

*S2 : Pour moi plutôt, je suis un peu timide, vous allez voir, j'attends qu'ils partent et après je commence*

Cela peut aller jusqu'à une certaine inhibition en présence d'autrui et l'attente de se retrouver seul pour enfin effectuer une commande vocale (message vocal en fait).

## Facteurs contextuels influençant l'usage des modalités

- Crainte d'attirer l'attention (S1, S2)

Pour les sujets timides (S1 et S2), la crainte d'attirer l'attention semble bien présente, et ceux ci vont donc éviter de se retrouver au centre de l'attention en effectuant des commandes vocales.

*S2 : expé : tu pensais que ça va gêner les gens ou que ça va te gêner toi? Les deux moi je pense, les gens pas tellement, mais c'est vrai, c'est gênant quand tout le monde se retourne comme ça*

Il indique également qu'il ne pense pas déranger, mais que c'est lui qui est mal à l'aise.

*S1 : Y avait pas mal de gens qui me regardaient. Ils avaient l'air surpris*

- Les heuristiques dépassent parfois la gêne

Certains sujets sont conscients ou sensible à la présence de tiers sans qu'il n'y ait vraiment de timidité. Mais cette présence semble tout de même avoir un effet sur le choix des modalités d'interaction.

*S3 : (expé : a failli envoyer au stylet (enfants autour) et le fait en vocal) Ca doit être l'habitude de dire "envoyer à rouge"*

Ici, le sujet 3 a eu un temps d'hésitation à effectuer une commande vocale du fait de la présence d'enfants et esquisse une ébauche de commande au stylet, mais au dernier moment il finit par faire une commande vocale. Il attribue ce choix à l'habitude d'effectuer cette commande (« envoyer à rouge ») avec la modalité vocale, ce qui est effectivement le cas puisque dans les deux dernières sessions, ce sujet atteint 60% de modalité vocale pour cette commande.

- Gestualité associée au vocal non naturel (S1)

La gestualité associée à la mise en œuvre de la modalité vocale peut être perçue comme une attitude gênante vis à vis des autres.

*S1 : , mais je trouve que c'est pas un contact naturel, ya vraiment rien de naturel, je suis comme ça (se penche ?)*

Ici, le sujet mime le geste de rapprocher le PDA de la bouche tout en se penchant de manière exagérée dessus, comme pour insister sur le malaise perçu.

Cette question fera l'objet d'une courte partie afin de montrer qu'il y a également de la variabilité inter et intra individuelle dans la gestualité accompagnant une commande vocale, ainsi donc que dans la perception de gêne qu'elle peut procurer.

- Kit main libre plus acceptable (discrétion) (S1)

Pour pallier ce désagrément et augmenter la discrétion, le sujet 1, toujours, évoque l'idée d'un kit main libre plus discret au moins du point de vue de la gestualité de la mise en œuvre de la commande vocale, mais pas du point de vue de la discrétion sonore ou du mode d'expression qui sont perçus comme les plus gênants.

*S1 : ... j'aurais eu un kit main libre justement comme on le disait, j'aurais peut-être eu moins de réticence à le faire, limite j'aurais pu le ranger dans ma poche*

- Peur de se tromper (S1)

*S1 : j'sais jamais si ça va marcher, j'sais pas, j'aimais pas trop*

L'incertitude (justifiée vu les faibles taux de reconnaissance) du succès de la commande vocale, font craindre au sujet 1 de devoir répéter plusieurs fois les mêmes commandes et ainsi d'augmenter la gêne perçue.

- Mode d'expression gênant, Vocabulaire simpliste (S1)

Le style d'interaction et la sémantique choisis pour les commandes vocales semblent avoir un effet sur certains sujets particulièrement « timides ».

*S1 : je peux pas parler avec mon pda devant les gens (expé : plus discret) Ouais, je sais que c'est pas euh pas un truc que j'aimerais faire. Et je trouve que justement j'ai l'impression, pas passer pour bête mais c'est vrai que le rapport qu'on a avec le pda devant les gens c'est ... (« aller à »).. ça fait assez, neuneu je trouve, enfin ça me plait pas trop*

Le sujet 1 explique que le rapport qui est institué par le type de vocabulaire pour communiquer avec le système ne lui convient pas du tout, en particulier devant des tiers. Apparemment, il semble ressentir un sentiment proche du ridicule (« ça fait assez neuneu je trouve »).

Il est intéressant de noter que les exemples de commandes qu'il cite « aller à reçus » ou « aller à archivés », sont effectivement des commandes complexes, composées et dont la construction n'est pas des plus correcte. Il s'agit d'ailleurs quasiment des seules commandes de ce type dans l'application, même s'il en existe d'autres complexes. Etonnement, et en accord avec les résultats quantitatifs, il s'agit quasiment des seules commandes vocales testées par ce sujet.

*S1 : Si je pouvais lui parler plus directement, « message », des choses plus brèves, ça fait pas neuneu mais l'impression d'être bête comme si.. que c'est un robot en fait. C'est vrai que parler à un robot, c'est pas un truc que j'assimile bien*

Comme formulée dans sa requête ci-dessus, il aurait préféré avoir à sa disposition un style plus direct, avec des commandes courtes et simples (« message »), qui l'auraient peut-être mis plus à l'aise. Il est aussi curieux de constater qu'il avait également à sa disposition des commandes du type souhaité (« fermer », « répondre », « transférer » etc.) qu'il n'a pas du tout essayées. Ceci peut s'expliquer en partie par une verbalisation déjà citée précédemment concernant la perception par ce sujet de la modalité vocale comme une sorte de raccourci :

*S1 : En fait je me demandais quel raccourci je voulais garder je pense, parce que pour moi, je les voyais comme des raccourcis en fait ces deux trucs (vocal et g2D)*

Plus précisément, il voit ces commandes comme des manières d'aller d'un espace à un autre, de naviguer, et moins comme des actions que le système devrait effectuer. Ici, donc, contrairement à d'autres sujets, il préfère faire les commandes « lui-même », en mode direct (au stylet), plutôt que de « déléguer » l'action au système via une commande vocale.

- Crainte de gêner les autres (ou à l'inverse, constat de non gêne (S6))

Le sujet de la crainte de gêner les autres en utilisant la modalité vocale a souvent été abordé, et le constat indique que les sujets semblent peu avoir le sentiment de déranger.

*S2 : expé : tu pensais que ça allait gêner les gens ou que ça va te gêner toi? Les deux, moi je pense, les gens pas tellement, mais c'est vrai, c'est gênant quand tout le monde se retourne comme ça*

Dans cet exemple, le sujet 2 n'aborde pas la question spontanément, après une hésitation il explique qu'il ne pense pas tellement gêner les gens autour de lui, mais il rappelle que c'est plutôt le regard des autres qui lui pose problème.

Cette question sera abordée dans la partie suivante concernant le constat de « non gêne ».

- Crainte d'erreur potentielle due au bruit associé (S3, S4, S5, S6)

Pour la plupart des sujets, la présence de tiers, au-delà d'une éventuelle gêne sociale ressentie, a été perçue comme une cause potentielle d'erreur de reconnaissance vocale liée à l'éventuel bruit produit par les personnes.

*S5 : Ben ça pouvait gêner par rapport aux gens ou au bruit aussi, je me suis dit à un moment, est-ce que ça va pas écouter ce qui joue, ça va peut-être dérégler le truc, mais non, apparemment, ça a pas joué énormément*

Ici, le sujet 5 exprime sa crainte première (« écouter ce qui joue, ça va peut-être dérégler le truc »), et explique ensuite qu'il a pu vérifier qu'apparemment le bruit n'a pas été une cause importante d'erreurs (voir les résultats quantitatifs).

Ce sujet est également traité dans la partie concernant l'influence du bruit sur les usages des modalités et il en ressort, comme ici, que les sujets, après une certaine appréhension d'erreur, testent l'application et vérifient son bon fonctionnement dans différents contextes sonores.

- Besoin d'intimité (S6)

Le besoin d'intimité, qui pourrait sembler primordial pour une application incluant des commandes vocales, a été très peu abordé lors des verbalisations (s'explique par le contexte expérimental ?).

*Expé : si y avait pas l'oreillette?*

*S6 : Je pense que ce serait pas pareil quand même, déjà je lui ferai pas lire les messages et puis même, tu sais, tu reçois tout les trucs de Linda tout ça, non. Ouais, non, il faut qu'il y ait l'oreillette, après moi je m'en fous de parler parce que tout le monde parle autour. En plus le son il est pas terrible, et puis tout le monde a pas besoin de profiter de ce que je suis en train de faire, mails, machin etc.*

Cet exemple rassemble un certain nombre de points importants associés au besoin d'intimité et de discrétion lié à une interaction vocale en entrée comme en sortie du système.

Le sujet 6 ne place pas au même niveau de besoin d'intimité, les sorties du système et ses commandes (ou messages vocaux). Il insiste bien sur la nécessité d'avoir une oreillette pour les sorties afin de conserver les informations reçues dans une sphère personnelle, alors qu'au contraire, cela ne le dérange pas de dicter des commandes (ou des messages vocaux) devant des tiers. Il accepte donc de « partager » ce qu'il produit mais pas ce qu'il reçoit (est ce parce qu'on ne sait pas ce que l'autre a fait, et que l'on contrôle ce que nous même nous faisons ?).

### 3.3.2. Expression d'absence de gêne liée à la présence de tiers

Un certain nombre de sujets (S3, S4, S5, S6) ont quant à eux, manifesté une relative absence de gêne due aux personnes environnantes, durant l'interaction avec le système. Plusieurs types d'arguments ont été avancés afin d'exprimer ce sentiment.

- Cadre spécifique

Le fait que l'expérimentation se soit déroulée dans le cadre de la Cité de la Science et de l'Industrie à la Villette, lieu public à forte fréquentation, public bénéficiant d'une certaine présomption de technophilie, a généré un contexte d'interaction spécifique, venant se rajouter au contexte propre du cadre expérimental.

- Attente de comportement social spécifique au lieu (S1, S3, S5, S6)

Les sujets (S1, S3, S5, S6) ont tous fait remarquer qu'il s'agissait d'un cadre spécifique où les attentes de comportement étaient particulières.

*S1 : (expé : t'as eu l'impression que ça modifiait ce que tu faisais ?) Non, pas du tout, je pense pas. Enfin, je sais pas, c'est pas les gens qui viennent ici qui... je m'en foutais un peu quoi, et ça*

## Facteurs contextuels influençant l'usage des modalités

*reste discret quoi, je trouve que tout reste discret, j'avais pas l'impression de passer pour un martien*

Par exemple, ici, le sujet 1 indique que le public venant dans un tel lieu peut s'attendre à voir des technologies innovantes et a donc un « seuil d'étonnement » un peu plus élevé que dans d'autres circonstances. Cela semble diminuer un peu la contrainte de discrétion que ce sujet souhaite (voir la suite sur le dispositif expérimental).

*S3 : (expé : et là, tu passes une commande vocale avec du bruit autour) Ca m'a pas dérangé, avec le cadre et tout, les gens ils sont pas trop étonnés donc ça gêne en rien*

Il en est de même pour le sujet 3 qui explique que le fait de passer une commande vocale est d'autant moins gênant dans un tel cadre, que le public a des attentes de comportements différents de ceux attendus dans un autre contexte.

*S5 : la gêne ? ah... je pense que ça gêne pas, de toutes façons, tout le monde parle, c'est vrai que si on était que, dans un autre espace, ou on pouvait rien dire, il faut pas parler comme dans une bibliothèque par exemple, là, je pense que j'aurai utilisé une autre modalité, mais là, non, ça m'a pas gêné,*

*S3 : Par exemple, si je suis dans la foule j'utiliserai plus le stylet pour faire discret, par rapport au bruit autour de moi*

Ici, les sujets 5 et 3 déclarent adapter leur comportement au contexte local de l'interaction, et en particulier au comportement des autres personnes présentes (« ça gêne pas (de parler), tout le monde parle »).

Ils projettent également leur éventuel comportement dans un autre contexte, comme dans une bibliothèque par exemple où l'attitude normale attendue est silencieuse et donc où une commande vocale est impossible.

Par ailleurs, le sujet 3 donne l'exemple d'une foule, dans laquelle il utiliserait préférentiellement la modalité tactile par soucis de discrétion et pour ne pas faire de bruit et ainsi attirer l'attention.

*S5 : (fait ouvrir en vocal avec des gens autour) Ca m'a pas gêné, je fais mes trucs et ils font leurs trucs en fait*

Le sujet 5 rajoute également qu'il ne tient pas compte de l'activité et du comportement des personnes qui l'entourent.

- Nature du cadre social : situation d'expérimentation

Nous avons pu récolter quelques remarques concernant le cadre et le dispositif expérimental et son éventuelle influence sur le choix des modalités d'interaction, une partie spécifique sera dédiée à ce sujet, mais ici nous abordons en particulier les aspects liés aux contraintes sociales.

- Désinhibition liée à la situation (S6)

Le fait qu'il s'agisse d'une expérimentation a certainement eu une influence sur le comportement des sujets et il est impossible d'en évaluer l'importance, cependant, plusieurs sujets mentionnent ce fait.

*S6 : je pense qu'il ya peut-être un peu de ça, je suis au centre des attentions, les gens ils me regardent, je sais qu'ils vont rien me dire, je suis filmé, j'ai les lunettes caméra etc, je pense qu'il y a peut-être un peu de ça*

Comme nous le verrons également par la suite, le sujet 6 a plusieurs fois déclaré être moins inhibé par la présence d'autres personnes dans l'environnement du fait qu'il s'agit d'une expérimentation dont il est au cœur et cela a eu des influences positives sur son usage de la modalité vocale.

- Dispositif technique discret

Le sujet 1 fait remarquer que le dispositif technique reste relativement discret et n'attire pas tellement l'attention. Pour les autres sujets également, le dispositif technique n'a pas semblé constituer une contrainte très importante.

*S1 : et ça reste discret quoi, je trouve que tout reste discret, j'avais pas l'impression de passer pour un martien*

- Infirmation de crainte de parler devant des gens

Plusieurs sujets ont mentionné leur absence de gêne de communiquer avec un système devant les autres visiteurs.

*S4 : aux visiteurs ? Je m'en foutais. Et là, en plus aujourd'hui c'était la seule journée ou y avait vraiment du monde, les autres jours, c'était beaucoup plus calme, et même aujourd'hui où y avait du monde, des gamins et compagnie, non, pas grand monde m'a regardé*

*S6 : Ca m'a pas gêné que les gens soient à côté, faudrait qu'il y ait plus de monde pour voir, si y avait une foule, je sais pas trop comment ça se passerait*

Le sujet 4 justifie en partie son absence de gêne par l'absence de réaction d'étonnement des personnes autour de lui, mettant en avant une certaine normalité de son comportement.

Le sujet 6, quant à lui exprime le même sentiment en précisant que la quantité de personne l'entourant pourrait tout de même avoir des conséquences différentes qu'il ne connaît pas à l'avance.

- Approche utilitariste (Recherche de performance du vocal sans soucis des gens autour)

Ici encore certains sujets confirment leur absence de gêne due aux autres visiteurs par le fait qu'ils ont une tâche à réaliser et que les circonstances d'usages ont finalement peu d'importance.

*Expé : tu crois que l'environnement t'as influencé?*

*S6 : Non, t'es concentré sur ce que t'as à faire moi je savais que je devais envoyer des msg à Linda, il fallait que je prenne des informations, mais ça m'a absolument pas perturbé, bon, c'était pas une foule dense, je me faisais pas bousculer, je sais pas, ya plein de conditions d'utilisation*

*Expé: si c'était hors expérimentale?*

*S6 : Si j'étais là pour ça, si j'avais des infos à donner à quelqu'un, absolument aucun problème, même je me la pêterai tu sais, je suis comme ça, avec mon pda, c'est James Bond, tu sens que t'es pas pareil que les autres gens, t'as ton pda, tu parles un peu au truc*

Dans ces exemples, le sujet 6 témoigne de la focalisation sur la tâche et de l'obligation de devoir l'effectuer plaçant de manière volontaire le contexte d'usage au second plan.

- Focalisation sur l'interaction et perte relative du contexte social (S1 ; S3, S5)

Il s'agit ici d'une problématique proche de la précédente mais ici le contexte social est occulté de manière involontaire en raison d'une focalisation excessive sur la tâche (effet tunnel).

*S3 : Mais le pire, c'est que je fais pas attention au monde qu'il y a autour de moi, quand je suis dans le pda, je suis vraiment dans le pda. Par exemple, là, les personnes qui sont à côté, je les avais pas remarquées*

Le sujet 3 déclare a posteriori qu'il n'avait pas perçu la présence de tiers et implicitement indique qu'ils n'avaient donc que peu d'influence sur le choix d'usage des modalités.

- Valorisation personnelle (S6)

Pour le sujet 6, l'usage d'un dispositif expérimental élaboré a été source de valorisation personnelle et a poussé à l'utilisation de la modalité vocale en particulier de par son caractère innovant.

*Expé: si c'était hors expérimentale?*

*S6 : Si j'étais là pour ça, si j'avais des infos à donner à quelqu'un, absolument aucun problème, même je me la pêterai tu sais, je suis comme ça, avec mon pda, c'est James Bond, tu sens que t'es pas pareil que les autres gens, t'as ton pda, tu parles un peu au truc*

*Rapidos mais la classe, non, je pense que je voulais un peu impressionner les gens, genre le mec qui parle à son pda, il est fou, Peut-être inconsciemment ya de ça, mais plusieurs fois, je me suis dit je vais parler, (car) ya des gens à côté*

Le sujet 6 indique ainsi sa volonté de se mettre en valeur vis à vis des autres visiteurs.

### 3.3.3. Analyse et synthèse

Le Tableau 16 rassemble et synthétise l'ensemble des données quantitatives et qualitatives concernant les usages selon le contexte social.

**Tableau 16 : Comparaison des données quantitatives et déclaratives par sujet, concernant le nombre de commandes vocales selon le contexte social**

Sujets	Différence commandes vocales Seul / Monde	Déclarations	Analyse
S1	Pas de vocal	Gêne, vocabulaire	Cohérence entre les données et le discours
S2	Pas de vocal	Gêne, timidité	Cohérence entre les données et le discours
S3	6,48%	Cadre spécifique, conscience de la présence d'autrui	Cohérence entre les données et le discours, Heuristiques fortes qui incitent au vocal
S4	4,13%	Pas de gêne si discret	Cohérence entre les données et le discours, tendance à parler doucement, s'orienter à l'opposé des gens pour les commandes vocales
S5	-2,02%	Cadre spécifique, indifférence	Cohérence entre les données et le discours
S6	13,78%	Cadre spécifique, valorisation	Incohérence entre les données et le discours, Conscience de la présence d'autrui exacerbée (volontarise la commande vocale ?)

Ce tableau permet de montrer la variabilité des résultats quantitatifs quant aux écarts d'usages de la modalité vocale, mais également il montre la variabilité du niveau subjectif de la perception du contexte social.

### 3.4. Illustrations de commandes vocales en présence de tiers

Le Tableau 17 illustre le passage de commandes vocales en présence de tiers.

## Facteurs contextuels influençant l'usage des modalités

Il s'agit de l'enchaînement de deux commandes (« Sélectionner un destinataire » et « envoyer ») qui sont représentées de manières séquentielle dans le tableau. Une illustration des codages associés est également présentée.

**Tableau 17 : Illustrations de commandes vocales en présence de tiers**

S6i3 (11'57''02) Série de Commandes vocales Monde Regard vers l'environnement		
Phase	Illustration	Description
Entrée 1		<p>Sujet : « sélectionner (destinataire) jaune »</p> <p>Monde Bruyant</p> <p>Engagement Physique : Vision environ Pouce gauche appui Main droite libre Cognitif : Attente FB (PEC et ES)</p>
Entrée 2		<p>Sujet : « envoyer »</p> <p>Monde Bruyant</p> <p>Engagement Physique : Vision environ Pouce gauche appui Main droite libre Cognitif : Attente FB (PEC et ES)</p>

Dans cet exemple, le sujet est en présence de deux personnes (relativement éloignées), et cela ne l'empêche pas d'effectuer une série de commandes vocales.

## 4. Influence des niveaux d'activité sur l'usage des modalités

Dans cette partie, nous allons tenter d'identifier une éventuelle influence de la tâche en cours et de son espace de travail associé sur les choix d'usages des modalités. En effet, il se peut qu'un certain nombre de caractéristiques d'une tâche (ressources associées, exigences de la tâche, besoin de mobilité etc.) puissent ne pas être neutre sur les choix de modalité par les sujets.

Dans un premier temps, nous avons pu grâce au fichier Log, retracer automatiquement un premier niveau d'activité correspondant à l'espace de travail dans lequel se trouve l'utilisateur.

Dans un deuxième temps, un codage manuel correspondant à l'activité réelle des utilisateurs a été effectué, permettant de mieux rendre compte des implications de la tâche en cours.

La question globale à laquelle nous cherchons donc à répondre dans cette partie est donc la suivante :

### *L'activité en cours peut-elle influencer sur le choix d'usage des modalités ?*

#### *4.1. Influence de l'espace de travail (fenêtre) sur l'usage des modalités*

L'application de messagerie proposée aux sujets est tout à fait classique dans son architecture. Elle propose une division en trois fenêtres principales qui ont été enregistrées automatiquement grâce aux fichiers Logs que l'application génère :

- **Liste des messages** : Cette fenêtre permet la consultation de la liste des messages indiquant l'expéditeur ou le récepteur, l'objet, la date d'envoi et éventuellement une notification de présence d'une pièce jointe. Elle comporte également trois sous espaces :
  - Liste des messages reçus
  - Liste des messages envoyés
  - Liste des messages archivés (cette espace contient les questions du Quiz auxquelles les sujets doivent répondre)
- **Lecture des messages** : Cette fenêtre permet la consultation des messages ainsi que des éventuelles pièces jointes (image, fichier sonore, ou texte manuscrit)
- **Edition des messages** : Cette fenêtre permet la création de messages ainsi que l'adjonction d'éventuelles pièces jointes (image, fichier sonore, ou texte manuscrit)

Ces trois fenêtres correspondent donc à des types de tâches et d'activités bien différentes et cette classification permettra donc un premier niveau d'analyse.

##### 4.1.1. Temps passé dans chaque espace de l'application

Le temps passé dans chaque espace de travail permet de rendre compte du niveau d'activité global pour l'ensemble des sujets

Fenêtre de l'application	Durée Totale	%(t)
Liste des messages	03:30:30	27,1
Edition d'un message	04:06:12	31,7
Lecture d'un message	04:50:33	37,5

**Tableau 18 : Temps passé dans chaque espace de l'application**

Les résultats montrent une répartition relativement homogène du temps passé dans chaque fenêtre de l'application avec le temps minimum (27,1% du temps) pour la liste des messages. Cet espace supporte une activité plutôt limitée et transitoire qui est la sélection pour l'ouverture ou la suppression d'un message.

## Facteurs contextuels influençant l'usage des modalités

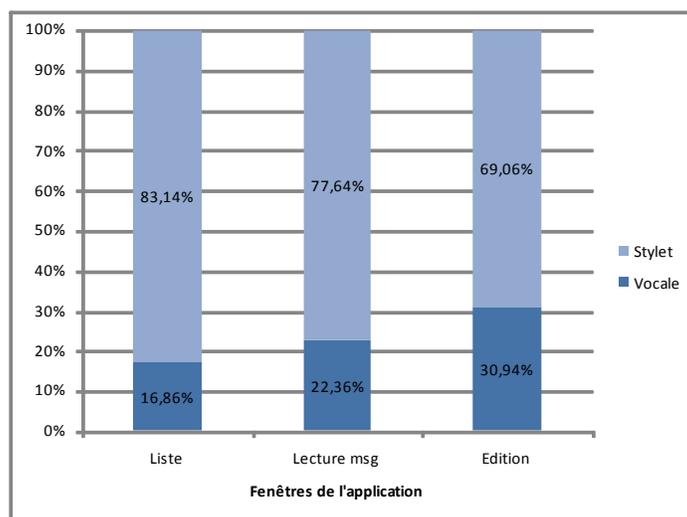
Viens ensuite la fenêtre édition de message (avec 31,7% du temps) où les sujets pouvaient soit demander ou envoyer des indices à leur collègue, ou encore répondre à une question résolue du Quiz.

Finalement vient la fenêtre de lecture de message (avec 37,5% du temps) qui effectivement supporte une grande partie de l'activité des sujets consistant à la recherche d'une réponse dans leur environnement à une question du Quiz, et ceci à l'aide d'un indice visuel contenu dans un message envoyé par leur collègue.

Ce raffinement d'analyse de l'activité sera davantage poussé dans la partie étudiant l'activité réelle des sujets, qui elle, a été codée manuellement.

### 4.1.2. Niveau global (tous sujets, toutes interactions)

La Figure 49 indique à un niveau global, une répartition très différente de l'usage des modalités selon les espaces de l'application. En effet, la fenêtre Edition est celle qui contient le ratio de commandes vocales le plus élevé (30,94%), alors que les fenêtres Lecture et Liste des messages ont des ratios respectifs de 22,36 et 16,86 %.



**Figure 49 Influence de l'espace de travail (fenêtre) sur l'usage des modalités**

Il y a donc une forte influence de l'espace de travail, et donc à travers cela de l'activité, sur l'usage des modalités, mais les résultats globaux présentés ici ne permettent pas d'en identifier les causes ou d'en donner une explication précise.

### 4.1.3. Niveau individuel

Les données disponibles permettent de raffiner au niveau individuel afin de vérifier si l'on peut extrapoler les résultats globaux.

## Facteurs contextuels influençant l'usage des modalités

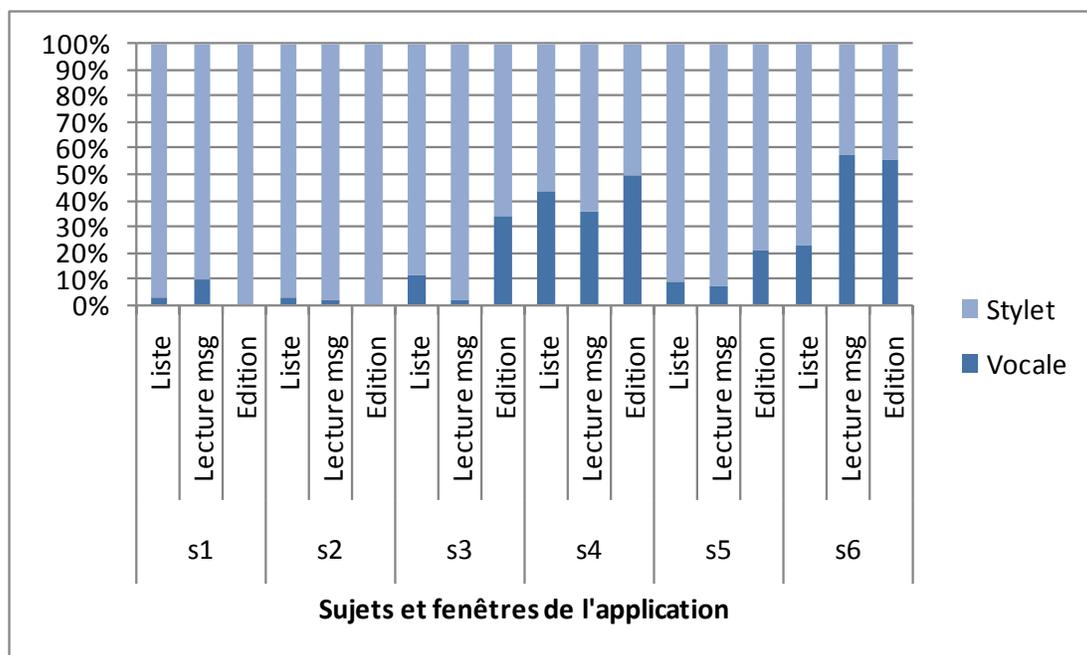


Figure 50 : Influence par sujet de l'espace de travail (fenêtre) sur l'usage des modalités

Les sujets 1 et 2 qui ont un usage très faible de la modalité vocale sont traités séparément : Il est intéressant de noter que les quelques commandes vocales qu'ils ont émises, l'ont été uniquement depuis les fenêtres de Liste des messages et Lecture des messages (en particulier pour le sujet 1 qui a émis quasiment 10% de commandes vocales dans cet espace).

En ce qui concerne le reste des sujets, les résultats au niveau individuel sont assez similaires à ceux du niveau global pour la fenêtre Edition qui est bien celle depuis laquelle ont été émises le plus de commandes vocales, avec une petite exception pour le sujet 6 dont le ratio de commandes vocales est quasiment identique dans la fenêtre Lecture. Ce dernier résultat vient certainement expliquer le poids de la fenêtre Lecture au niveau global, car pour tous les autres sujets, la fenêtre Liste est la deuxième en termes d'importance du ratio de commandes vocales.

Là encore, il est impossible d'en expliquer les raisons sans aller dans le détail de l'analyse à l'aide des verbalisations et des chroniques d'activité. Cette analyse sera faite dans la partie suivante qui s'intéresse à l'activité réelle des sujets qui est plus représentative de la perception des sujets.

### 4.2. Influence de l'activité réelle sur l'usage des modalités

Afin d'étudier l'influence de l'activité réelle, du point de vue subjectif, nous avons établi un certain nombre d'observables permettant de la décrire simplement. Il en résulte 7 catégories descriptives qui se déclinent ainsi :

Acronyme	Nom complet	Description de l'activité associée
QA	Questions de l'autre	Lire les questions de l'autre et chercher à lui envoyer les indices correspondants. Cela inclut également toute action collaborative orientée vers la résolution du Quiz
QP	Questions Perso	Lire ses questions, demander les indices, et chercher la réponse dans l'environnement
RA	Réponse de l'autre	Cela consiste simplement à lire la réponse que l'autre a envoyée concernant ses questions

<b>RP</b>	Réponse Perso	Lorsque la réponse a été trouvée dans l'environnement, écrire un message à l'autre contenant la réponse
<b>Lecdiv</b>	Lecture Divers	Lecture de messages non directement orientés vers la résolution du Quiz
<b>Editdiv</b>	Edition Divers	Edition de messages non directement orientés vers la résolution du Quiz
<b>div</b>	Divers	Autres activités non décrites ci-dessus, incluant les situations de résolution de problèmes techniques ou autres

**Tableau 19 : Description des catégories descriptives de l'activité réelle**

Comme nous pouvons le voir, ces catégories sont d'un niveau assez élevé et peuvent regrouper des actions de natures assez différentes telles que l'édition, la lecture. De manière transversale, nous retrouvons également des critères de nature sociale puisque ceux-ci sont parfois orientés vers la satisfaction d'une exigence liée à la collaboration entre les sujets (en particulier QA et RA ). Comme nous l'avons vu dans la partie (3 Influence de l'environnement social sur l'usage des modalités) ce genre d'activité peut parfois provoquer une certaine pression temporelle et sociale afin de remplir les objectifs de collaboration.

#### 4.2.1. Durées de l'activité réelle

Le Tableau 20 présente les durées globales de chaque activité et permet d'en évaluer leur répartition relative.

	Total	%(t)
<b>QP</b>	04:22:22	33,8
<b>RP</b>	02:12:03	17
<b>QA</b>	02:12:01	17
<b>div</b>	01:30:32	11,7
<b>Lecdiv</b>	00:58:11	7,5
<b>Editdiv</b>	00:49:36	6,4
<b>RA</b>	00:32:13	4,2

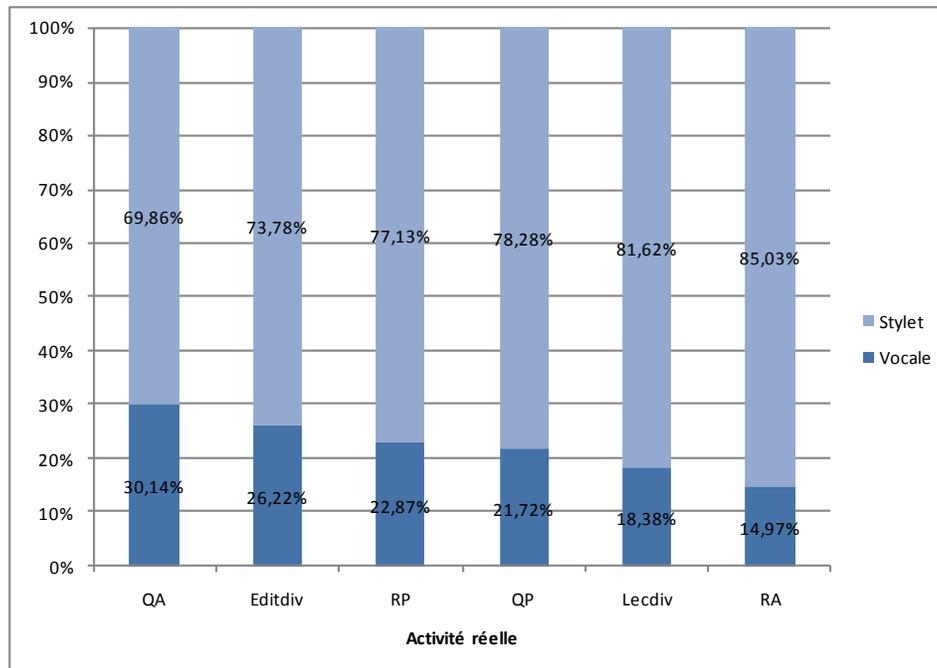
**Tableau 20 : Durée globale de l'activité réelle**

Du point de vue temporel, c'est l'activité QP (Question Perso avec 33,8% du temps) qui, loin devant les autres, a occupé le plus les sujets. Puis viennent les activités de Réponse Perso et Question de l'Autre avec 17% du temps des sessions. Globalement les sujets ont donc occupé majoritairement leur temps avec des activités orientées par leurs questions et réponses personnelles (en tout 50,8% de leur temps), avec une pression sociale faible a priori. Les activités orientées vers la collaboration avec leur collègue occupent 21,2% au total.

#### 4.2.2. Influence globale et individuelle de l'activité réelle sur l'usage des modalités

La Figure 51 croise les valeurs d'activité avec l'usage des modalités, afin d'en évaluer une éventuelle influence.

## Facteurs contextuels influençant l'usage des modalités



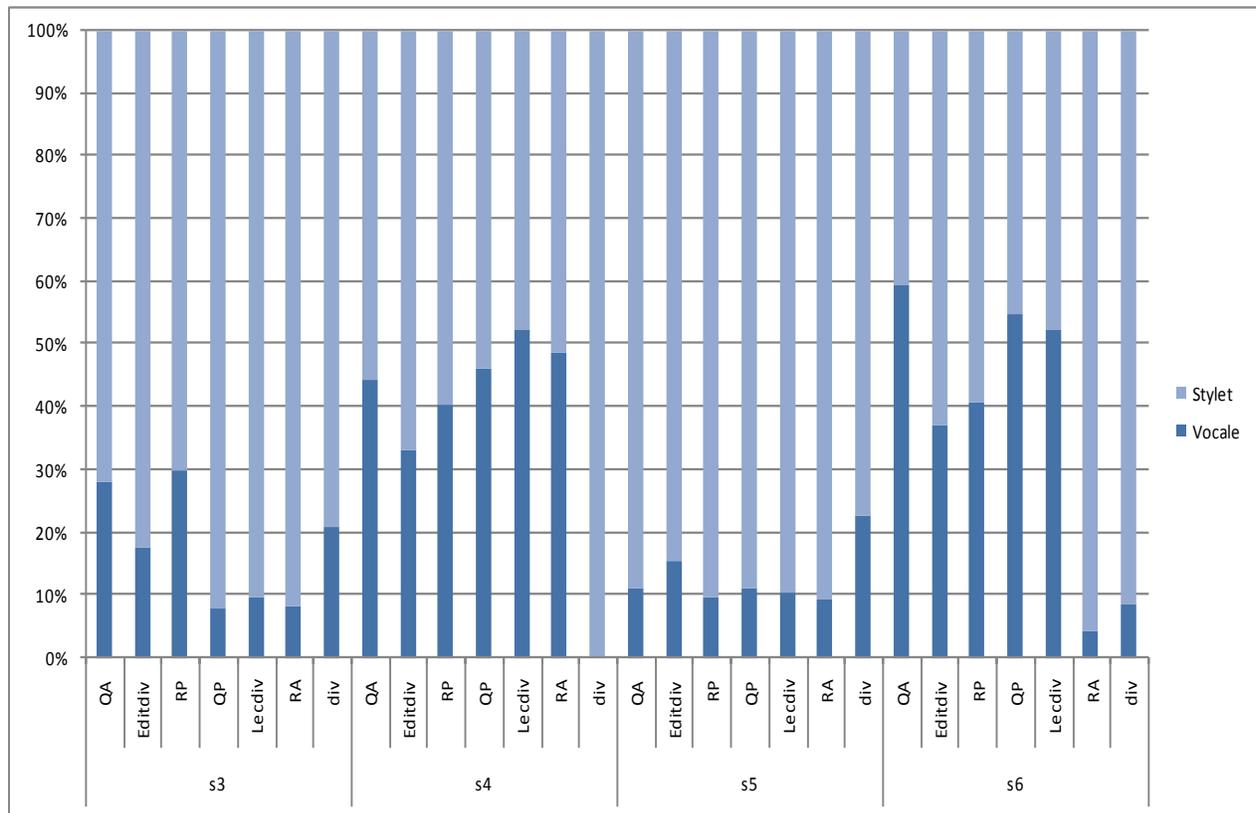
**Figure 51 : Influence de l'activité réelle sur l'usage des modalités**

L'activité à un niveau global semble avoir une influence sur le choix des modalités avec des différences importantes entre QA, l'activité qui connaît le plus haut ratio de commandes vocales (30,14%) et les autres types d'activité (au minimum 9,18%).

Cependant, la Figure 52 qui fait apparaître les résultats au niveau individuel semble montrer qu'il s'agit d'un effet de moyenne, en particulier dû au sujet 6 et dans une moindre mesure au sujet 4 dont les ratios de commandes vocales pour QA sont particulièrement élevés (pratiquement 60% pour S6, et 45% pour S4).

Les sujets 1 et 2 ne sont pas représentés en raison du faible effectif de commandes vocales. Quant à l'ensemble des autres sujets, il est impossible de tirer des résultats généraux et communs, en effet, on note d'assez fortes variations inter individuelles quant à l'usage de la modalité vocale en fonction du contexte d'activité.

## Facteurs contextuels influençant l'usage des modalités



**Figure 52 : Influence de l'activité réelle sur l'usage des modalités au niveau individuel**

Les niveaux de codage de l'activité choisis ne semblent donc pas suffisamment fins à eux seuls pour tenter d'expliquer les usages de modalités, cependant, nous verrons plus loin, que couplé à l'étude de la mobilité, ils peuvent être utiles à une certaine discrimination de critères influents.

### 4.2.3. Perception subjective de l'influence de l'activité

*S1 i2 : Stratégie globale : j'ai regardé toutes mes questions en fait, y'en avait quelques-unes, je savais où il fallait voir parce que je savais ce que c'était, donc j'ai commencé par lui demander celle ou je savais pas trop, j'ai fait tout une par une.. dans l'ensemble les questions j'ai bien vu que c'était genre..il devait m'envoyer une photo d'où j'allais trouver le truc*

*S1 i2 : j'essayais d'alterner ses réponses, mes questions et chercher*

*S1 i2 : Là j'arrivais un peu mieux à me déplacer dans le sens où je connaissais les temps, donc je cliquais en avançant, j'avais peut-être mieux la mesure des choses, je savais combien de temps il mettait, je savais que je pouvais regarder vite fait, jeter un coup d'œil en même temps marcher. Un peu mieux ouais, mais bon pareil, je pouvais pas non plus trop me balader, c'était plus essayer de changer de salle, le temps que tout arrive, le temps qu'il se mette en route.*

*S2 i2 : Parce que justement y avait l'impératif de répondre aux questions, personnellement, j'ai pris ça, pas comme un challenge mais, mais je me suis dit, ben allez, on va essayer de répondre aux questions le plus vite possible, moi même je me suis fait pas violence, je me suis mis un peu moi-même la pression, pour les besoins de l'expérience  
Je voulais le battre en fait Thibaud, c'est pour ça que je vais vite*

- *Pression temporelle liée au caractère social de l'activité*

La catégorie QA (Question de l'Autre : recherche des indices de l'autre sur son propre PDA pour les lui envoyer) est celle pour laquelle la pression temporelle peut être la plus forte car le partenaire dépend du traitement de cette activité pour à son tour chercher ses réponses.

*Expé : tu passes au stylet*

*S4 i1 : Il a pas fonctionné tout de suite, donc, comme je disais, comme j'ai passé beaucoup de temps à essayer de copier coller et que le vocal là, marchait pas, ben j'ai voulu aller le plus vite possible, pour ne pas faire perdre de temps non plus à Jezabelle, donc il fallait aller vite pour les réponses*

Le sujet 4 mentionne ici une conjonction de deux facteurs conduisant au passage à la modalité tactile stylet. Il s'agit de la récupération d'une situation d'erreur couplée à une activité orientée socialement et plaçant une certaine pression temporelle.

*S6 i2 : (est assis) Là, je passe un petit moment à tout répondre, à lui envoyer correctement, et là, je fais tout au stylet pour aller plus vite*

Dans cet autre exemple, le sujet 6 déclare également prendre en compte un facteur social justifiant le choix de la modalité tactile perçue comme plus rapide. Il est intéressant de noter que dans ce cas, le sujet accorde un temps particulier à cet activité en y dédiant la totalité de ses ressources attentionnelles.

## 5. Historique de l'interaction

Nous avons vu que l'usage des modalités variait en fonction du nombre de sessions d'interaction. Dans cette partie, nous nous intéresserons donc au mode d'appropriation de l'usage de l'artefact construit par l'expérience de l'utilisateur dans l'interaction et appréhendé ici sous la forme d'un historique des interactions.

### 5.1. Historique local (Intra-session) et inter-session : étude de cas

Les interactions successives, qu'elles soient constituées de succès ou d'erreurs viennent participer à la construction de l'appropriation du système par l'utilisateur. Cette appropriation peut relever aussi bien de l'expérience immédiate (intra-session) que de l'expérience des sessions précédentes (expérience inter-sessions), en particulier lorsqu'elles sont rapprochées comme dans notre cas.

Afin de mener une telle analyse, nous avons construit des chroniques d'activité sujet par sujet et session par session qui nous permettent d'avoir une « photographie » des événements constituant l'historique de l'interaction. Ces chroniques couplées avec les verbalisations des sujets permettent de plus d'y introduire un meilleur degré de subjectivité et donc de se rapprocher de l'expérience telle que vécue par les sujets.

Pour illustrer cette analyse, nous ferons une étude de cas en nous concentrant sur le sujet 1 et les difficultés qu'il a rencontré avec la modalité vocale allant en partie jusqu'à son abandon.

Ces chroniques ont été filtrées afin de présenter de manière synthétique l'usage des différentes modalités (lignes du haut) indexé temporellement ainsi que les situations d'erreurs rencontrées (ligne du bas).

Le sujet 1 a très peu utilisé la modalité vocale, ce qu'il a en partie justifié par la gêne que cela représente de parler à une machine en public (voir partie 3 Influence de l'environnement social sur l'usage des modalités). Cependant, comme nous allons le voir, d'autres facteurs liés à l'historique de l'interaction viennent également contribuer à l'abandon de cette modalité.

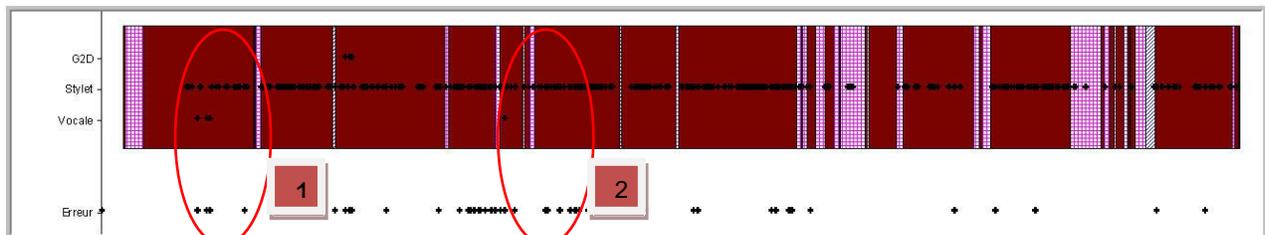


Figure 53 : Chroniques d'activité du sujet 1 session 1 (focus sur les erreurs et leur récupération)

Les chroniques indiquent deux séquences d'interactions vocales (cercles rouges 1 et 2) toutes deux suivies d'erreurs, pour lesquelles nous disposons de verbalisations :

*S1i1 : Séquence 1 : C'était un peu pour tester le vocal, je voulais voir si c'était rapide et tout, c'est pareil, j'ai réessayé encore une fois après (séquence 2). C'est pas un mode qui m'a plu du tout.*

Le sujet explique qu'il souhaitait dans un premier temps tester la modalité vocale en faisant la commande « aller aux messages reçus », mais il expérimente une première erreur (séquence 1), liée au fait que cette commande est impossible depuis l'espace où il se trouve. Cependant, le système ne lui donne pas l'information pour comprendre son erreur et la corriger. Il réitère donc deux fois de suite cette erreur en pensant pouvoir la corriger.

Puis en milieu de session il retente la même commande (séquence 2) qui est à nouveau suivie cette fois d'une erreur de reconnaissance liée au système :

*S1i1 : Séquence 2 : J'ai jamais réussi à le faire en fait, j'sais pas pourquoi. Du coup, j'ai laissé tomber, j'ai essayé deux trois phrases, je me suis dit que ça allait plus vite de cliquer*

Suite à ce deuxième test infructueux et à la lourdeur des corrections, il déclare abandonner cette modalité qui lui semble beaucoup moins efficace et rapide que la modalité tactile stylet.

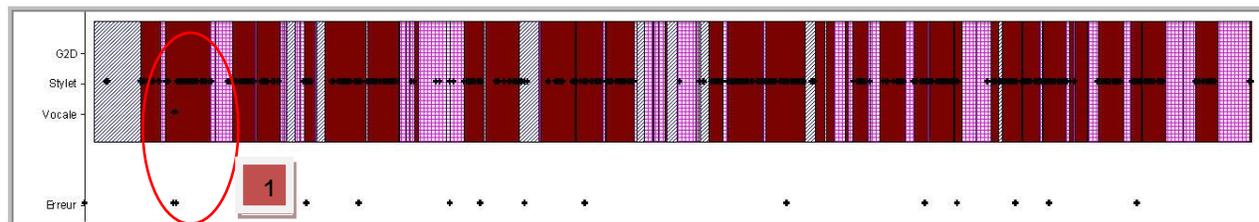


Figure 54 : Chroniques d'activité du sujet 1 session 2 (focus sur les erreurs et leur récupération)

Dès le début de la deuxième session, le sujet tente de nouveau une commande vocale (« nouveau message ») qui vient également se solder par une erreur, cette fois encore liée au système de reconnaissance. Il tente de corriger l'erreur avec la même modalité mais sans succès et passe finalement au stylet.

*S1i2 : Séquence 1 : J'ai fait « Nouveau msg » et ça a pas marché, j'ai arrêté après. La reconnaissance vocale, j'avais envie d'essayer, mais à chaque fois, j'ai beau essayer, je trouve ça trop long, trop compliqué, trop lourd à chaque fois. Dès que j'essaye une fois... Quand je suis pressé, quand ya des msg qui arrivent, des choses, je peux pas entendre, msg qui arrive plus machin enregistré, j'trouve que ça reste vraiment trop lourd, ya beaucoup de choses à faire, appuyer, bien penser à mettre le petit clic sur le micro, parler, enfin mine de rien il faut quand même faire attention à la distance, pas trop loin ni trop près. Je trouve que c'est trop long en fait vraiment trop long.*

Suite à cette première tentative infructueuse, qui vient se cumuler avec les expériences négatives des sessions précédentes, le sujet déclare arrêter les tentatives d'usage de la modalité vocale (ce qui se vérifie dans la chronique d'activité). Aux situations d'erreurs qu'il a rencontrés, viennent se rajouter des arguments liés à la lourdeur de mise en œuvre perçue par le sujet.

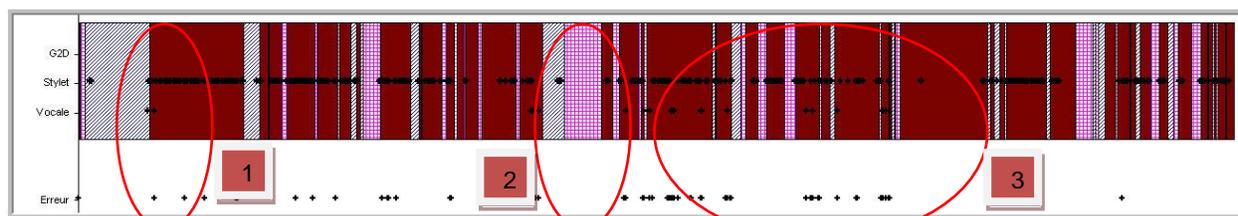


Figure 55 : Chroniques d'activité du sujet 1 session 3 (focus sur les erreurs et leur récupération)

A nouveau, au début de la session 3, le sujet tente une commande vocale (« aller reçus ») qui se solde par un échec corrigé par la modalité tactile et l'abandon de la modalité vocale.

*s1i3 : Séquence 1 : A chaque fois je pense que ça va pas marcher et que ça va prendre deux fois plus de temps de toutes façons. Même si c'est pas le cas, ça plante assez souvent, je me dis que ça va pas marcher donc...*

Ce cumul d'expériences négatives à travers les trois sessions le conduit donc à ne plus considérer cette modalité comme effective. Cependant, la chronique montre deux nouvelles séquences de commandes vocales (séquences 2 et 3) venant contredire les préférences du sujet. La séquence 3 est étonnamment longue et fait intervenir un nombre de commandes vocales totalement inhabituelles pour ce sujet. Si l'on regarde dans le détail de l'interaction, on peut voir qu'il s'agit d'une séquence où le système a tendance à bugger, même pour la modalité tactile.

*s1i3 : Séquence 3 : Mon PDA, il a eu quelques petits bugs, je crois, j'arrivais pas vraiment à aller dans les fenêtres où je voulais aller, fallait que je passe par les commandes vocales et après que*

## Facteurs contextuels influençant l'usage des modalités

*je ferme, parce que je sais pas pourquoi quand je disais (aller à msg reçus), il me mettait je sais pas où, autre part quoi et je pouvais pas fermer les msg. (Fait donc une commande vocale car l'application bug). J'avais l'impression que ça marchait, que ça le débuggait en fait.*

Dans cet exemple, le sujet témoigne de son usage « contraint » de la modalité vocale afin de sortir d'une situation de bug de la modalité tactile.

Lors de la session de débriefing à l'issue des trois sessions de test, le sujet donne son impression générale sur la modalité vocale.

*s1i3 : Je suis pas très mode vocal, même là, ça marchait mieux, j'ai essayé de m'en servir un peu plus, même si ça buggait, elle reconnaissait tout le temps ce que je disais. J'ai essayé plus les modalités vocales cette fois, ça marchait assez bien, pas tout le temps mais ça allait quoi*

Le sujet confirme donc que la modalité vocale n'est pas sa modalité de prédilection, même si en fin de dernière session, il a connu des expériences plus positives où celle-ci lui a permis de sortir de situation d'erreur. Il déclare finalement qu'elle marche assez bien.

Le choix d'usage d'une modalité au dépend d'une autre peut donc reposer sur des événements plus ou moins ponctuels (événements intra-session) ou des expériences antérieures plus diffuses (événements inter-sessions), qu'elles soient considérées comme positives ou négatives, et qui viennent renforcer les représentations existantes du sujet à propos des propriétés de telle ou telle modalité et expliquer en partie les choix d'usages des modalités.

## 6. Synthèse des résultats sur les facteurs contextuels influençant l'usage des modalités

### 6.1. Contexte sonore

Les résultats quantitatifs au niveau global indiquent que le niveau sonore ne semble pas avoir d'influence sur le taux d'erreur de la modalité vocale.

Le contexte sonore bruyant, voir très bruyant représente 60 % du temps global. Quant au nombre de commandes effectuées dans ces deux conditions, la différence est moindre puisque l'on a 53,13% seulement de commandes dans un environnement bruyant et 46,87% dans un environnement non bruyant.

Au niveau global, le contexte sonore semble avoir un effet sur le choix d'usage des modalités. La modalité tactile stylet qui est utilisée à 80,90% dans des conditions bruyantes, passe à 73,10% dans des situations non bruyantes, et inversement pour la modalité vocale, qui elle, passe de 19,10% à 26,90%.

A un niveau global, le contexte bruyant semble donc ne pas favoriser l'usage des modalités vocales. Cependant ce résultat est à relativiser au niveau individuel qui montre une certaine tendance des sujets à émettre moins de commandes vocales que de commandes tactiles en milieu bruyant mais avec des variabilités inter-individuelles assez prononcées.

Au niveau des résultats qualitatifs, nous avons rassemblé les verbalisations sur la perception subjective du niveau sonore et de son influence sous trois thèmes :

- Les gênes perçues dues au bruit
  - Crainte d'erreurs potentielles dues au Bruit
  - Changement de modalité en anticipation ou suite à des erreurs avec bruit (Changement de modalité due au bruit, Stratégie d'anticipation, Stratégie de retrait du milieu bruyant pour effectuer une commande vocale)
- Les modifications de modes opératoires liées au bruit
  - Augmentation du volume de la parole
  - Rapprochement du pda près de la bouche
- Un thème divers
  - Appropriation de l'application : Le degré d'expérience d'usage de l'application peut lui aussi venir modifier la perception et l'influence du bruit sur le mode opératoire associé à la modalité vocale.
  - Non perception ou perception a posteriori du bruit : De manière générale, nous avons pu constater que les sujets n'étaient pas forcément très sensibles au niveau sonore ambiant
  - Focalisation sur l'interaction et perte relative du contexte sonore

### 6.2. Contexte Social

Sur la totalité de la durée des sessions, 64,1% du temps correspond à des situations où les sujets sont en présence de tiers, durant lesquelles 58,22% des commandes ont été émises.

Les résultats quantitatifs indiquent un effet du contexte social, au niveau global, moins important que celui du contexte sonore.

Au niveau individuel, nous avons observé de fortes variations inter-individuelles.

L'effet du contexte social sur l'usage des modalités semble également faible pour les sujets 4 et 5.

En revanche, les écarts de nombres de commandes vocales en contexte social « monde » ou « seul » sont plus importants pour les sujets 3 et 6 (6,48% et 13,78% respectivement).

Au niveau qualitatif, les verbalisations concernant le contexte social traitent plus particulièrement de la modalité vocale :

Les verbalisations concernant l'interaction tactile stylet en présence de monde sont très peu nombreuses, et semblent indiquer une banalisation de ce type d'usages.

### 6.2.1. Expression de gêne liée au monde

Les sujets 1, 2 et 6 sont ceux qui ont manifesté le plus de sensibilité au contexte social quant à l'usage de la reconnaissance vocale et ceci pour des raisons diverses explicitées ci-dessous de manière thématique. D'une manière générale, pour les sujets 1 et 2, la timidité sous des formes variées est un argument qui est souvent avancé comme cause de non usage de la modalité vocale.

Mais il peut y avoir également des gênes qui ne sont pas de l'ordre de la timidité, par exemple la crainte de déranger ou encore la crainte d'erreurs potentielles de la reconnaissance vocale liées au bruit produit par des tiers.

- **Inhibition (Timidité)** : Difficultés à parler (seul) à une machine devant des gens, Crainte d'attirer l'attention
- **Heuristiques fortes** : Les heuristiques dépassent parfois la gêne
- **Gestualité** : Gestualité associée au vocal non naturel
- **Discrétion** : Un Kit main libre serait plus acceptable
- Peur de se tromper
- Mode d'expression gênant, Vocabulaire simpliste,
- Crainte de gêner les autres,
- Crainte d'erreur potentielle due au bruit associé,
- Besoin d'intimité

### 6.2.2. Expression d'absence de gêne liée à la présence de tiers

Un certain nombre de sujets (S3, S4, S5, S6) ont quant à eux, manifesté une relative absence de gêne due aux personnes environnantes, durant l'interaction avec le système. Plusieurs types d'arguments ont été avancés afin d'exprimer ce sentiment.

- **Cadre spécifique** (expérimentation et Lieu public) :

Le fait que l'expérimentation se soit déroulée dans le cadre de la Cité de la Science et de l'Industrie à la Villette, lieu public à forte fréquentation, public bénéficiant d'une certaine présomption de technophilie, a généré un contexte d'interaction spécifique, venant se rajouter au contexte propre du cadre expérimental.

- Attente de comportement social spécifique au lieu
- Nature du cadre social : situation d'expérimentation
- Désinhibition liée à la situation
- **Dispositif technique discret**
- **Absence de crainte** de parler devant des gens
- **Approche utilitariste** (Recherche de performance du vocal sans soucis des gens autour) : certaines verbalisations témoignent d'une focalisation sur la tâche et de l'obligation de devoir l'effectuer plaçant de manière volontaire le contexte d'usage au second plan.
- **Focalisation sur l'interaction et perte relative du contexte social** : Il s'agit ici d'une problématique proche de la précédente mais ici le contexte social est occulté de manière involontaire en raison d'une focalisation excessive sur la tâche (effet tunnel).
- **Valorisation personnelle** : l'usage d'un dispositif expérimental élaboré a été source de valorisation personnelle pour certains sujets et a poussé à l'utilisation de la modalité vocale en particulier de par son caractère innovant.

## 6.3. *Activité en cours*

L'activité des utilisateurs a été abordée selon deux points de vue : un point de vue plus déterministe qui code automatiquement la fenêtre en cours dans l'application et un codage manuel a posteriori indiquant l'activité réelle de l'utilisateur.

### 6.3.1. Codage de l'activité via les espaces de travail de l'application (liste, édition, lecture)

Les résultats quantitatifs montrent une répartition relativement homogène du temps passé dans chaque fenêtre de l'application avec le temps minimum (27,1% du temps) pour la fenêtre présentant la liste des messages. Ils indiquent également, à un niveau global, une répartition très différente de l'usage des

modalités selon les espaces de l'application. En effet, la fenêtre « Edition » est celle qui contient le ratio de commandes vocales le plus élevé (30,94%), alors que les fenêtres « Lecture » et « Liste des messages » ont des ratios respectifs de 22,36 et 16,86 %. Il semble donc qu'il y ait, au niveau global, une forte influence de l'espace de travail, et donc à travers cela de l'activité, sur l'usage des modalités.

Les résultats au niveau individuel sont assez similaires à ceux du niveau global pour la fenêtre Edition qui est bien celle depuis laquelle ont été émises le plus de commandes vocales. De plus, un tel niveau de codage ne permet pas de rendre compte finement de l'activité des utilisateurs et son influence sur le choix des modalités et il est impossible d'en expliquer les raisons sans aller dans le détail de l'analyse à l'aide des verbalisations.

### 6.3.2. Influence de l'activité réelle sur l'usage des modalités

Globalement les sujets ont occupé majoritairement leur temps avec des activités orientées par leurs questions et réponses personnelles (catégorie QP et RP) (en tout 50,8% de leur temps), avec une pression sociale faible a priori. Les activités orientées vers la collaboration avec leur collègue occupent 21,2% au total.

Bien que l'activité à un niveau global semble avoir une influence sur le choix des modalités, les résultats au niveau individuel ne montrent aucune tendance commune et plutôt de fortes variations inter-individuelles quant à l'usage de la modalité vocale en fonction du contexte d'activité réelle. Le codage de l'activité réelle choisi, bien que plus fin que le précédent, ne permet donc pas de rendre compte des déterminants des choix de modalités liés à l'activité des sujets

### 6.4. *Historique de l'interaction*

Les interactions successives, qu'elles soient constituées de succès ou d'erreurs viennent participer à la construction de l'appropriation du système par l'utilisateur. Cette appropriation peut relever aussi bien de l'expérience immédiate (intra-session) que de l'expérience des sessions précédentes (expérience inter-sessions), en particulier lorsqu'elles sont rapprochées comme dans notre cas.

Nous avons illustré cette thématique par des chroniques d'activité graphiques croisées avec les verbalisations correspondantes, permettant une analyse fine des situations d'activité.

Dans notre exemple, suite à plusieurs tentatives infructueuses dans une session, qui viennent se cumuler avec les expériences négatives des sessions précédentes, le sujet déclare arrêter les tentatives d'usage de la modalité vocale (ce qui se vérifie dans la chronique d'activité). Aux situations d'erreurs qu'il a rencontrés, viennent se rajouter des arguments liés à la lourdeur de mise en œuvre perçue par le sujet.

Ce cumul d'expériences négatives à travers les trois sessions le conduit donc à ne plus considérer cette modalité comme effective. Cependant, la chronique montre deux nouvelles séquences de commandes vocales (séquences 2 et 3) venant contredire les préférences du sujet. Si l'on regarde dans le détail de l'interaction, on peut voir qu'il s'agit en fait d'une séquence où le système a tendance à causer de nombreuses erreurs (même pour la modalité tactile) et le sujet témoigne dans les verbalisations de son usage « contraint » de la modalité vocale afin de sortir d'une situation de bug de la modalité tactile.

Le choix d'usage d'une modalité au dépend d'une autre peut donc reposer sur des événements plus ou moins ponctuels (événements intra-session) ou des expériences antérieures plus diffuses (événements inter-sessions), qu'elles soient considérées comme positives ou négatives, et qui viennent renforcer les représentations existantes du sujet à propos des propriétés de telle ou telle modalité et expliquer en partie les choix d'usages des modalités.

**Chapitre 5. Multimodalité, mobilité et  
ressources attentionnelles**

## 1. Introduction

Après avoir étudié les usages de la multimodalité indépendamment de la situation d'usage (Chapitre 3, p86), nous avons tenté de les « resituer » en les rattachant au contexte au sens large (Chapitre 4, p132).

Dans cette troisième partie de résultats, nous abordons la problématique globale des usages situés dans un contexte mobile, en particulier par le biais de la question du partage des ressources attentionnelles.

Il est communément admis (Oulasvirta et al., 2005b; Oviatt, S.L., 2000) que l'usage des modalités a des conséquences diverses sur les ressources attentionnelles, qui sont partagées par les activités de l'utilisateur. Par exemple, la modalité tactile est généralement reconnue comme mobilisant le plus l'attention visuelle comparée à la modalité vocale, cependant, très peu d'études traitent de ce sujet dans des conditions de mobilité réelle (tout au moins, en dehors d'un laboratoire).

Dans cette partie, nous chercherons donc à étudier, en nous basant sur des observables que nous décrivons ci-dessous, les incidences de l'usage de telle ou telle modalité sur les ressources attentionnelles.

La présentation des résultats suivra l'ordre suivant :

- Effets des modalités sur l'interaction mobile
- Effet de la direction du regard sur la mobilité
- Effet de l'activité sur la mobilité
- Transitions entre les états de mobilité selon les modalités
- Effet des corrections d'erreurs et des modalités sur la mobilité

Pour chaque partie, après une présentation quantitative des résultats, nous présenterons une analyse qualitative issue des verbalisations des sujets.

Le chapitre se clôt avec une synthèse des résultats.

## 2. Observables des ressources attentionnelles

Loin de chercher à quantifier la charge attentionnelle (subjective, perçue, ressentie), nous nous attachons plutôt à une analyse de l'activité à l'aide de critères observables qui nous renseignent sur les possibilités (potentialités) que permettent l'usage d'une modalité en terme d'activité « consommatrices » (« concurrentes ») en ressources attentionnelles.

En effet, les sujets ont diverses tâches à réaliser en parallèle, toutes consommatrices en ressources attentionnelles et physiques à des degrés et à des moments divers. La consigne et la tâche ont été élaborées pour conduire à ce partage des ressources attentionnelles selon plusieurs focus :

- **Interaction avec le dispositif** : les sujets doivent parcourir les questions du quizz, demander à leur binôme des indices concernant la question, transmettre leurs réponses, mais la réciproque étant vraie, ils doivent également répondre aux questions de leur binôme. On trouve aussi un certain nombre de messages ayant plus trait à la régulation sociale (entre les participants).
- **Exploration de l'environnement** : les réponses aux questions se trouvant réparties dans l'exposition, les sujets sont amenés à explorer l'environnement à la fois visuellement, mais également par le biais de leur propre déplacement (mobilité).

Transversalement à ces deux focus, nous avons retenu et croisé essentiellement trois types d'observables qui nous semblaient suffisamment indicatifs du degré de mobilisation de ressources attentionnelles.

- **Mobilité** : La mobilité est une composante nécessaire à la réalisation de la tâche demandée, tout au long de la session, les utilisateurs doivent gérer leur déplacement afin d'explorer physiquement leur environnement.
- **L'attention visuelle** : Il en est de même pour l'attention visuelle, qui elle, se répartit essentiellement entre la gestion visuelle du déplacement pour l'exploration de l'environnement et l'interaction visuelle avec le système.
- **L'activité en cours** : Les deux observables précédents sont fortement influencés par l'activité en cours de l'utilisateur, en effet, selon la phase d'activité dans laquelle se trouve l'utilisateur, le besoin d'exploration physique ou visuelle n'est pas le même. Par exemple un sujet qui recherche une question sur son pda, sera plus focalisé sur l'interaction que s'il recherche la réponse à une question dans l'espace environnant.

Dans cette partie, nous croiserons donc l'analyse des observables « Mobilité », « Direction du regard » avec « l'Activité en cours » afin de dégager d'éventuelles relations ou impact sur le choix d'usage des modalités.

L'analyse quantitative sera suivie par une analyse qualitative des verbalisations des sujets concernant ces thématiques, ainsi que des thématiques plus transversales qui n'ont pas pu être codées systématiquement (charge cognitive, appropriation de l'application etc.)

### 3. Mobilité et interaction : Effets des modalités ?

A travers cette partie, nous chercherons à répondre à une question générale :

***L'usage des modalités joue-t-il un rôle sur les possibilités de mobilité ?***

Et nous tenterons également d'aborder la question réciproque d'une influence de la mobilité sur l'usage des modalités.

Pour cela, lors du codage de la mobilité, nous avons distingué trois catégories la caractérisant :

- « **déplacement** » : déplacement rapide d'un point à un autre généralement avec une destination facilement identifiable
- « **micromobilité** » : déplacement lent, déambulation, piétinement
- « **statique** » : aucun déplacement

#### 3.1. Répartition temporelle et numérique globale des catégories de mobilité

Afin de donner une idée globale du degré de mobilité de l'ensemble des sujets au cours des trois sessions, nous commençons par établir un total des temps liés à la mobilité. Le Tableau 21 présente la répartition temporelle des trois catégories de mobilité tous sujets confondus au cours des trois sessions.

	Total	%(t)
statique	08:07:12	62,8
micromobilité	03:51:23	29,8
déplacement	00:51:04	6,6

**Tableau 21 Répartition temporelle des catégories de mobilité**

Les résultats montrent qu'à un niveau global, 62,8% de la durée des sessions se déroulent en situation statique et 29,8% en micromobilité. Le déplacement ne représente que 6,6% de la durée des sessions. Nous rappelons ici qu'aucune consigne particulière n'a été donnée concernant l'obligation d'interaction en mobilité, de même, en ce qui concerne d'éventuelles contraintes temporelles. Les sujets étaient donc relativement libres en ce qui concerne leur gestion de la session du point de vue temporel et de la mobilité. Par contre, nous rappelons également que la tâche impliquait de fait des niveaux de mobilité différents afin de chercher les réponses aux questions du Quiz réparties dans l'environnement.

Mobilité	déplacement	micromobilité	statique	Total
Nb de Commandes	94	1012	2989	4095
% de Commandes	2,30%	24,71%	72,99%	100,00%

**Tableau 22 Répartition du nombre de commandes selon les catégories de mobilité**

Le Tableau 22 indique clairement la répartition des interactions en fonction du degré de mobilité de l'utilisateur (N= 4095). En effet, plus des deux tiers des interactions se déroulent en situation statique (n= 2989 ) et moins d'un tiers en micromobilité (n= 1012). Moins d'une centaine (n=94) de commandes sont effectuées en situation de déplacement.

Du point de vue du nombre de commandes, les ratios sont cohérents avec le tableau précédent et la répartition est encore plus marquée en faveur de la catégorie statique. Globalement, les commandes se répartissent à 73% en statique et 24,71% en micromobilité. Seules 2,3% des commandes sont effectuées en mode déplacement.

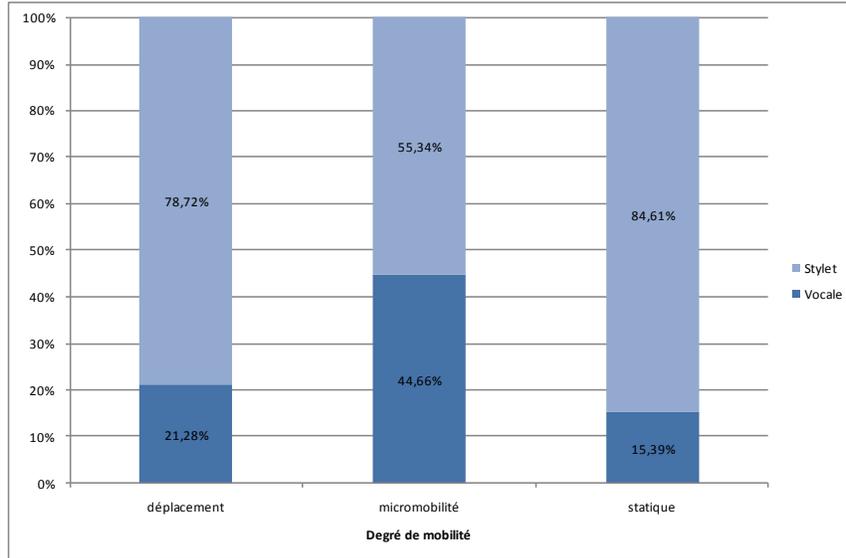
Ces résultats, ne prenant pas encore en compte les modalités d'interaction, soulèvent les deux questions suivantes :

***Lorsque les sujets en ont la possibilité, préfèrent-ils s'arrêter pour interagir avec le système ?***  
ou bien,

**L'interaction avec le dispositif les contraint-elle à s'arrêter ?**

**3.2. Mobilité et modalités**

Les résultats précédents indiquent donc une forte proportion d'interactions statiques (73%) et il convient donc maintenant d'évaluer la répartition des états de mobilité selon les modalités afin de détecter une éventuelle influence de celles-ci sur les possibilités de déplacements.



**Figure 56 Usage des modalités selon les catégories de mobilité**

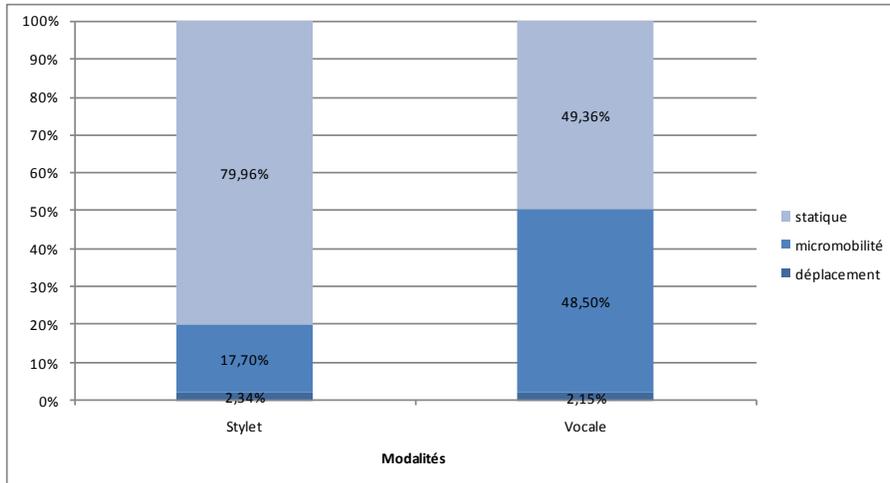
Les données disponibles (Figure 56) indiquent comme on pouvait s'y attendre que la modalité qui requière l'utilisation du stylet (Tactile Stylet) et conséquemment un niveau de contrôle visuel précis de l'action est effectuée à plus de 84% en statique alors que la modalité vocale est utilisée quasiment à 45% en micromobilité.

La modalité tactile stylet permet tout de même un bon degré de micromobilité puisqu'elle atteint plus de 55%, cependant, comme nous le verrons par la suite, la qualité de la mobilité n'est pas la même selon les modalités, notamment en ce qui concerne la mobilisation du regard.

Il est étonnant de remarquer que le stylet est largement majoritaire (proche de 80%) lors des interactions en déplacement, mais les effectifs de cette catégorie sont relativement bas (n=94) et certainement peu indicatifs.

Dans la figure suivante, nous avons mis la mobilité en tant que variable et les résultats sont encore plus nets.

## Multimodalité, mobilité et ressources attentionnelles



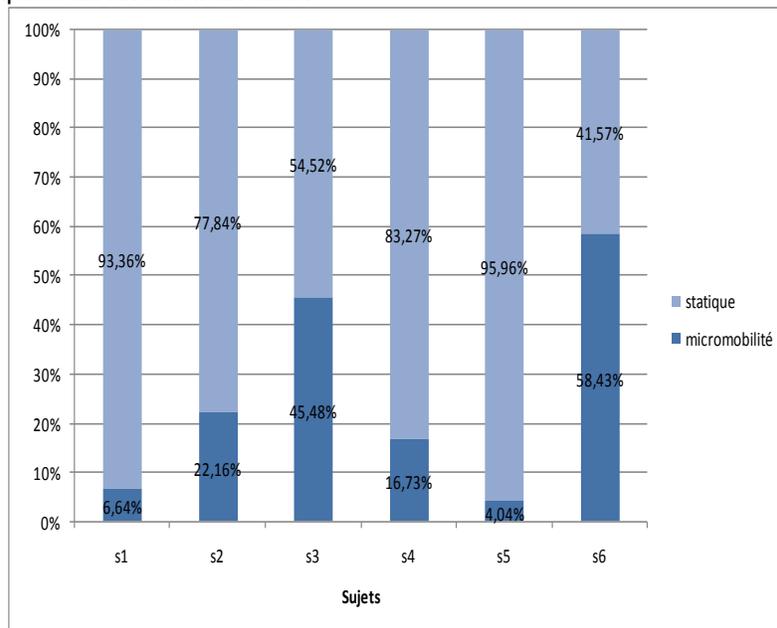
**Figure 57 : Mobilité selon les usages des modalités**

Présentés ainsi les résultats indiquent que pratiquement 80% des commandes tactiles stylet sont effectuées en statique alors qu'au contraire, on passe à un peu plus de 49% de commandes statiques pour la modalité vocale, ce qui fait un total d'un peu plus de 50% de commandes effectuées en mobilité. Il est donc clair que la modalité vocale favorise, ou du **moins permet un meilleur degré de mobilité**. Le croisement des résultats des deux figures précédentes semblent donc plutôt plaider pour l'hypothèse des modalités agissant soit comme contrainte, soit comme potentiel de la mobilité.

### 3.3. Variations interindividuelles de mobilité

Il est intéressant de regarder si tous les sujets ont des comportements identiques face à la possibilité d'interagir en mobilité « réelle ».

Les effectifs globaux de l'observable « déplacement » étant très faibles, nous n'en tiendrons plus compte dans la suite de l'exposé au niveau individuel.



**Figure 58 Répartition des catégories de mobilité par sujets**

La Figure 58 indique une forte variabilité interindividuelle pour le comportement lié à la mobilité des six sujets. Pour tous les sujets sauf le sujet 6 et le sujet 3, la grande majorité des commandes a été faite en situation statique (les résultats vont de 4% à 22% environ d'interactions mobile pour ces sujets). Ces

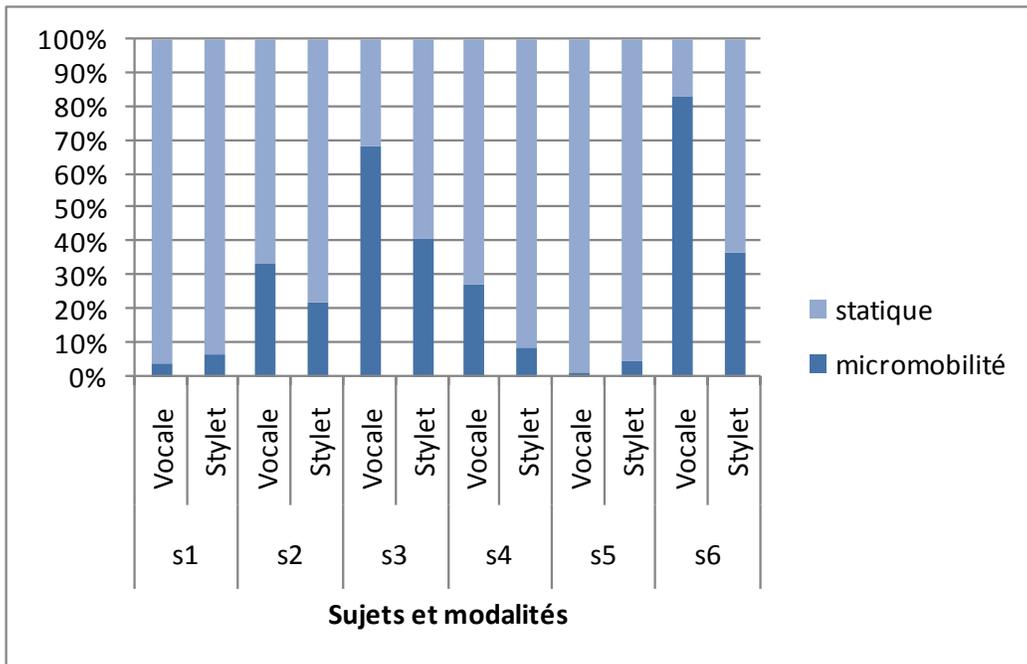
différences très importantes montrent bien que les sujets n'appréhendent absolument pas de la même manière l'interaction en mobilité et en particulier la possibilité d'interagir en mobilité « réelle », résultats qui répondent de manière satisfaisante à la question posée en début de partie :

**Lorsque les sujets en ont la possibilité, préfèrent-ils s'arrêter pour interagir avec le système ?**

Il semble donc que la réponse est de nature complexe puisque cela dépendra des sujets, et que l'on observe de fortes disparités dans les comportements. La réponse est donc relativement massivement oui pour les sujets 1, 2, 4 et 5. Elle est non pour les sujets 3 et 6.

La figure suivante réintroduit les modalités dans l'analyse afin d'établir un éventuel lien entre elles et les degrés de mobilité et répondre à une autre question posée précédemment :

**L'usage des modalités joue-t-il un rôle sur les possibilités de mobilité ?**



**Figure 59 Répartition des catégories de mobilité par sujets**

Les résultats de la Figure 59 sont également très marqués et permettent de répondre à la question du poids des modalités sur les possibilités d'interaction en mobilité « réelle ». Mis à part les sujets 1 et 5 pour qui les effectifs de modalité vocale sont très faibles et les résultats peu indicatifs, la figure montre un niveau de mobilité systématiquement plus élevé pour tous les sujets quand l'interaction est vocale.

Sujets	Vocale	Stylet	Différence
S1	4,35%	6,72%	-2,37%
S2	33,33%	21,97%	11,37%
S3	68,22%	41,11%	27,11%
S4	27,50%	8,60%	18,90%
S5	1,19%	4,40%	-3,21%
S6	83,13%	36,84%	46,29%

**Tableau 23 : synthèse des différences de micromobilité selon la modalité (par sujet)**

Comme l'indique le Tableau 23, les différences sont systématiquement supérieures à 11% et peuvent aller jusqu'à plus de 46% pour le sujet 6.

Ceci veut donc dire que même pour des sujets qui font peu d'interactions en mobilité « réelle », la modalité vocale est celle qui la favorise le plus. De plus, comme nous l'avons déjà mentionné, la qualité de la mobilité est très différente selon la modalité si on l'analyse finement.

### 3.4. Evolution de la mobilité au cours des sessions

La question posée dans cette partie est de savoir si au cours des sessions et de la familiarisation avec l'application, les sujets s'approprient l'usage des modalités et si cela a une incidence sur la mobilité.

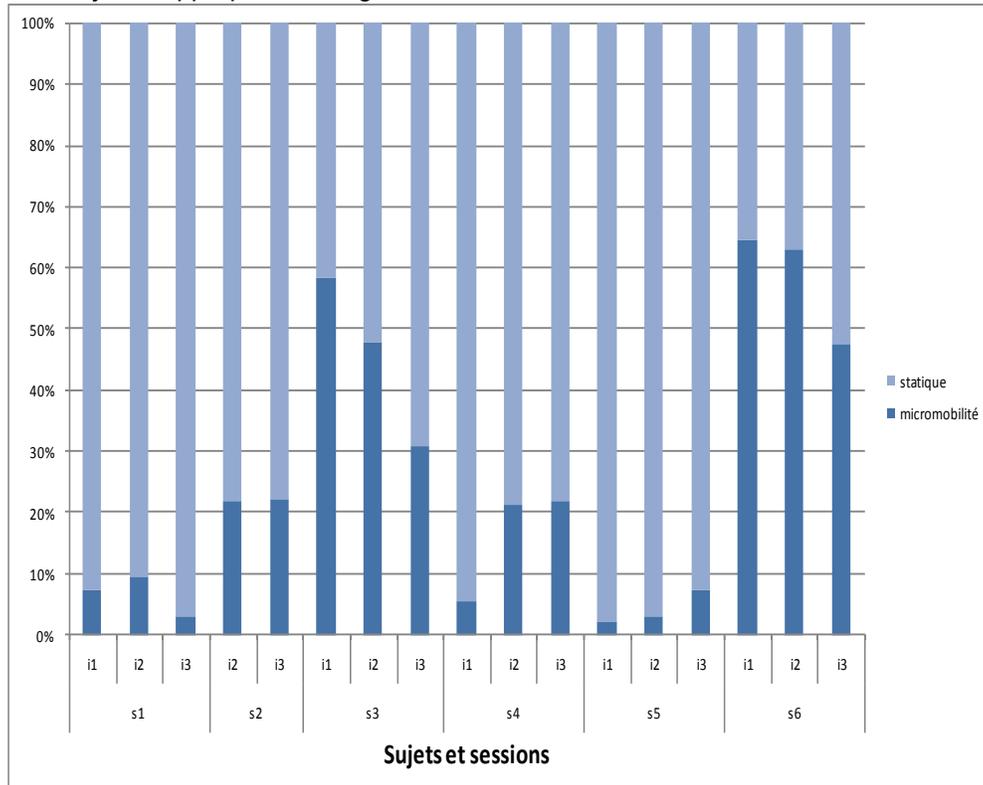


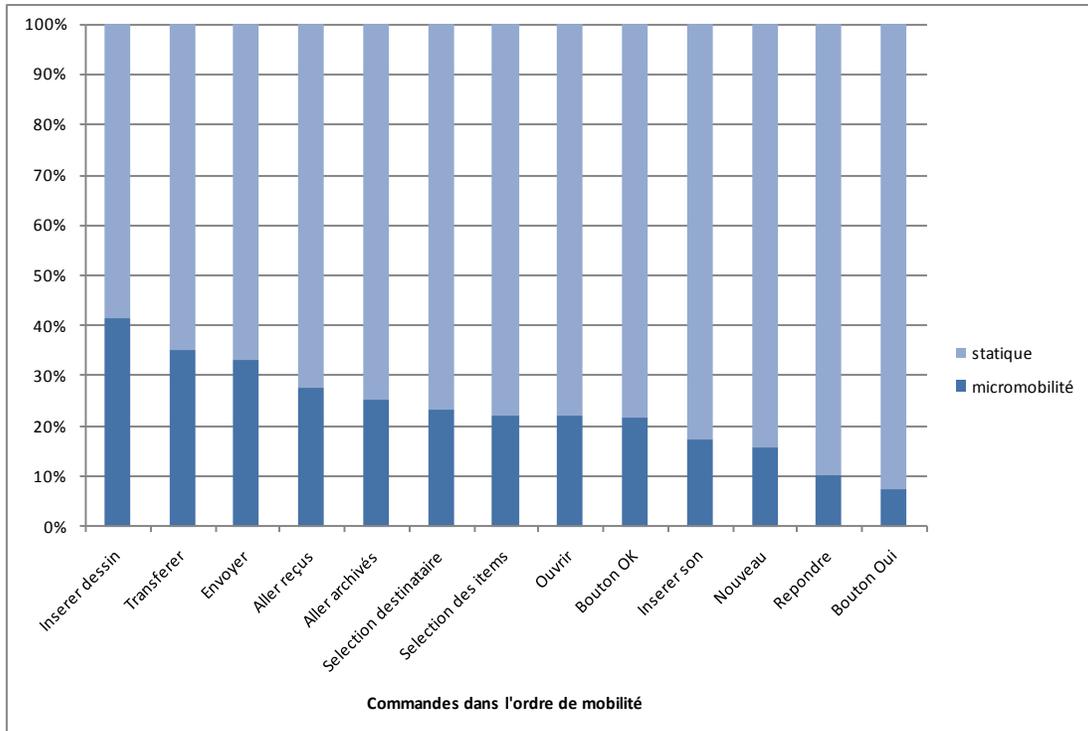
Figure 60 Evolution de la mobilité au cours des sessions

Là encore les résultats sont mitigés et seuls les sujets 3 et 6 semblent montrer une progression importante et nette mais contrairement à ce que l'on aurait pu croire, elle va dans le sens d'une diminution de l'interaction en mobilité « réelle » alors que ces sujets ont des taux d'interactions vocales en mobilité très élevés. Aucune explication précise ne peut être apportée en l'état des données, cependant il serait intéressant d'analyser la possibilité d'une réorganisation de leur activité pour limiter les interactions mobiles avec une augmentation de leur expertise. Il est possible également que leur expertise dans la tâche de résolution du Quiz leur ait permis de moins « tâtonner » et ainsi de diminuer le besoin d'interaction en mobilité réelle grâce à une meilleure répartition et séparation des tâches (d'interaction et de mobilité).

### 3.5. Mobilité et commandes

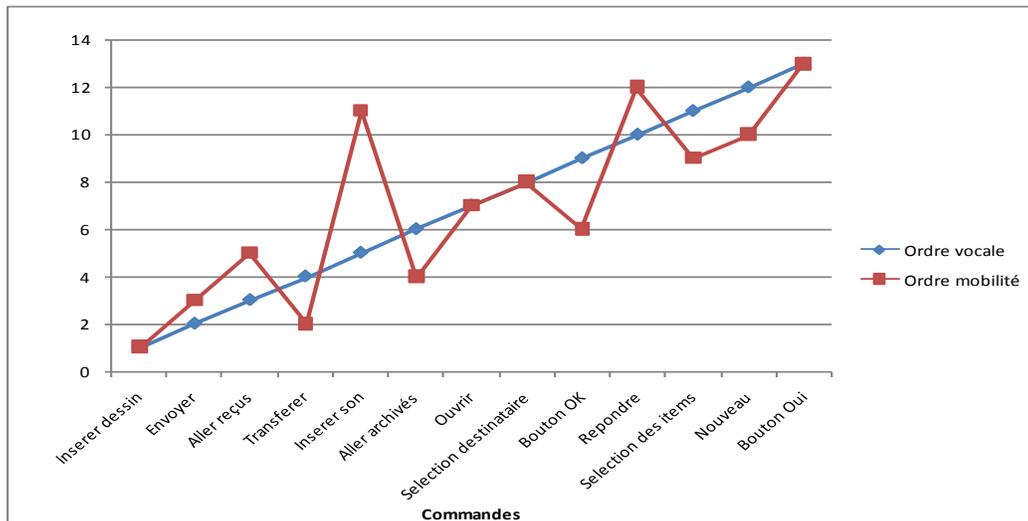
Nous raffinons ici l'analyse afin de valider ou d'invalider les résultats vus précédemment au niveau d'une commande individuelle.

## Multimodalité, mobilité et ressources attentionnelles



**Figure 61 Usage des commandes selon les catégories de mobilité**

Nous commençons d'abord par classer les commandes dans leur ordre de mobilité, ce qui permet de distinguer également de grandes disparités dans les résultats. C'est la commande « insérer un dessin » qui est la plus « mobile » avec plus de 40% d'effectifs en micromobilité, puis vient « transférer » etc. Afin de voir une éventuelle corrélation entre leur degré de mobilité et les modalités dans lesquelles elles ont été effectuées, nous les avons ordonnées en fonction de ces deux critères (l'ordre 1 étant le plus mobile et le plus vocal).



**Figure 62 : Comparaison des ordres de mobilité et de modalité vocale**

Les résultats au niveau d'une commande, là encore, sont cohérents avec les résultats précédents puisqu'il semble y avoir une corrélation forte entre l'ordre des commandes les plus vocales et celles les plus mobiles.

Les résultats présentés ici ne permettent pas de connaître les raisons expliquant ces ordres de mobilité et de modalité vocale. Même à travers les verbalisations, il nous a été impossible d'obtenir plus

d'informations à ce sujet et seules quelques inférences pourront être faites dans l'analyse des causes de choix de modalité vocale par commande.

## 4. Direction du regard

En revanche, il est possible d'établir un lien entre le degré de mobilité et la direction du regard, également en fonction des usages de modalités. Pour cela, la direction du regard lors d'une commande a été codée. Etant donné les variantes et les subtilités possibles qu'offre la direction du regard, et face à la difficulté d'en rendre compte finement, nous avons volontairement choisi des critères de codages simplistes et non ambigus. En effet, ces observables ne comprennent que deux catégories :

- **Interface** : L'utilisateur a le regard dirigé vers l'interface
- **Environ** : L'utilisateur a le regard dirigé vers l'environnement.

A l'aide de ces deux catégories et de leur croisement avec d'autres observables liés aux modalités, nous souhaitons apporter des débuts de réponses aux questions suivantes concernant l'orientation du regard pendant l'interaction :

***Comment s'oriente le regard pendant une interaction ?***

Et

***Observe-t-on des différences d'orientation du regard selon les modalités ?***

Dans un second temps, ces résultats seront croisés avec ceux de la mobilité.

### 4.1. Répartition temporelle de la direction du regard et interaction

Pour commencer, nous avons totalisé les temps d'orientation du regard pour tous les sujets et toutes les sessions.

Regard	Total	%(t)
Interface	08:59:04	69,5
Environ	03:48:44	29,5

**Tableau 24 Répartition globale de la direction du regard**

Le Tableau 24 montre une répartition d'environ un tiers du temps des sessions avec le regard orienté vers l'environnement et deux tiers orientés vers l'interface. Indépendamment de toute interaction, l'orientation visuelle pendant les sessions est donc relativement focalisée sur le contenu de l'application et la tâche demandée aux sujets semble donc être principalement orientée vers le contenu numérique de la messagerie.

Si maintenant nous couplons l'observation des interactions et les directions du regard, nous obtenons le tableau suivant :

Regard	Nb de commandes	Pourcentage
Environ	137	3,35%
Interface	3958	96,65%
Total	4095	100,00%

**Tableau 25 Répartition de la direction du regard**

Sur un total de 4095 commandes multimodales, seules 137 (3,35%) ont été effectuées le regard dirigé vers l'environnement et 3958 (96,65%) le regard dirigé vers l'interface.

L'interaction est donc globalement très fortement mobilisatrice de l'attention visuelle.

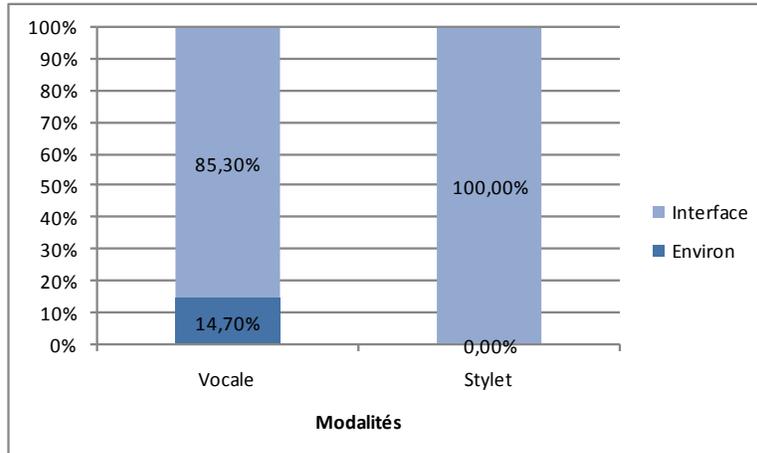
### 4.2. Répartition des directions du regard selon les modalités (Mode global)

Nous cherchons ici à voir dans quelle mesure les différentes modalités mobilisent l'attention visuelle.

Regard	Vocale	Stylet	Total
Environ	137	0	137
Interface	795	3163	3958
Total	932	3163	4095

**Tableau 26 Répartition de la direction du regard selon la modalité utilisée**

Sur un effectif de 4095 commandes, seules 137 ont été effectuées avec le regard orienté vers l'environnement.

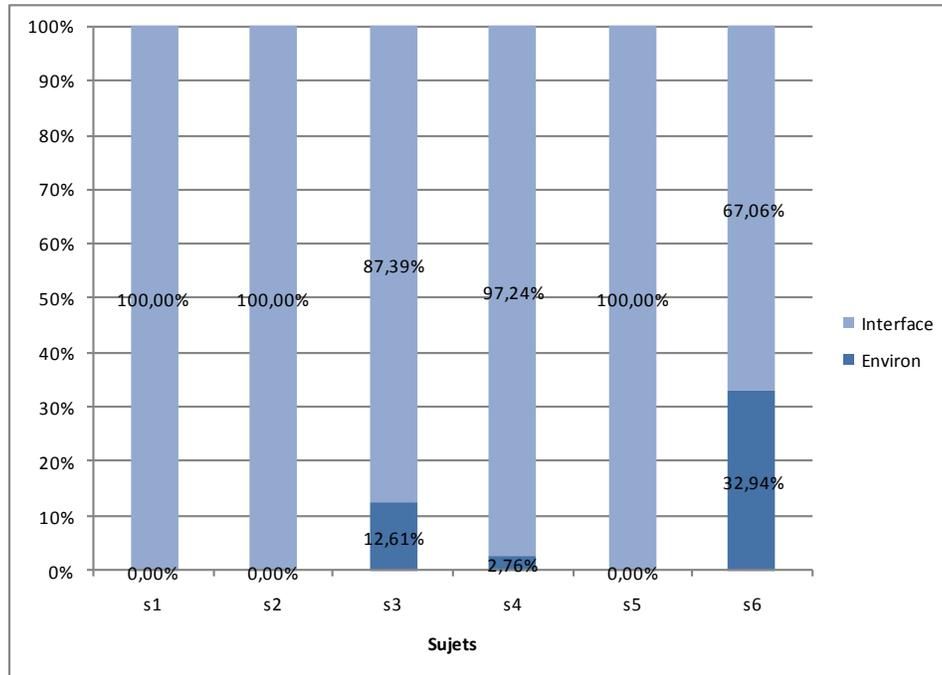


**Figure 63 Mobilisation de l'attention visuelle en fonction des modalités**

La Figure 63 indique que 100% des commandes tactiles ont nécessité une attention visuelle dirigée vers l'interface. La modalité vocale a permis 14,7% de commandes avec le regard orienté vers l'environnement. Ceci montre que la modalité vocale, dans une faible mesure, permet la libération de l'attention visuelle qui est intrinsèquement impossible avec la modalité tactile et que par ailleurs les sujets n'utilisent pas systématiquement la potentialité de libération du regard de la modalité vocale.

#### **4.3. Direction du regard selon les sujets**

Les sujets ayant effectué des commandes vocales (en particulier les sujets 5, 3, 4 et 6) peuvent donc potentiellement libérer leur attention visuelle durant la commande. Conformément à cette hypothèse, la Figure 64 montre une répartition non homogène des directions des regards des sujets et en particulier une orientation du regard vers l'environ plus marquée pour certains sujets qui ont tendance à utiliser la modalité vocale.



**Figure 64** Direction du regard selon les sujets

Cependant, on note que parmi ces 4 sujets (5, 3, 4 et 6), seuls les sujets 6, 3 et 4 dans une moindre mesure ont effectivement émis des commandes vocales avec le regard orienté vers l'environnement, ceci dans des proportions allant de 2,76% pour le sujet 4 à 32% pour le sujet 6. Le sujet 5 quant à lui, n'a pas mis à profit cette libération potentielle du regard.

Pour les sujets qui utilisent la modalité vocale, nous pouvons nous poser la question d'une éventuelle évolution de la direction du regard vers l'environnement au fur et à mesure des sessions et de leur appropriation de la modalité vocale.

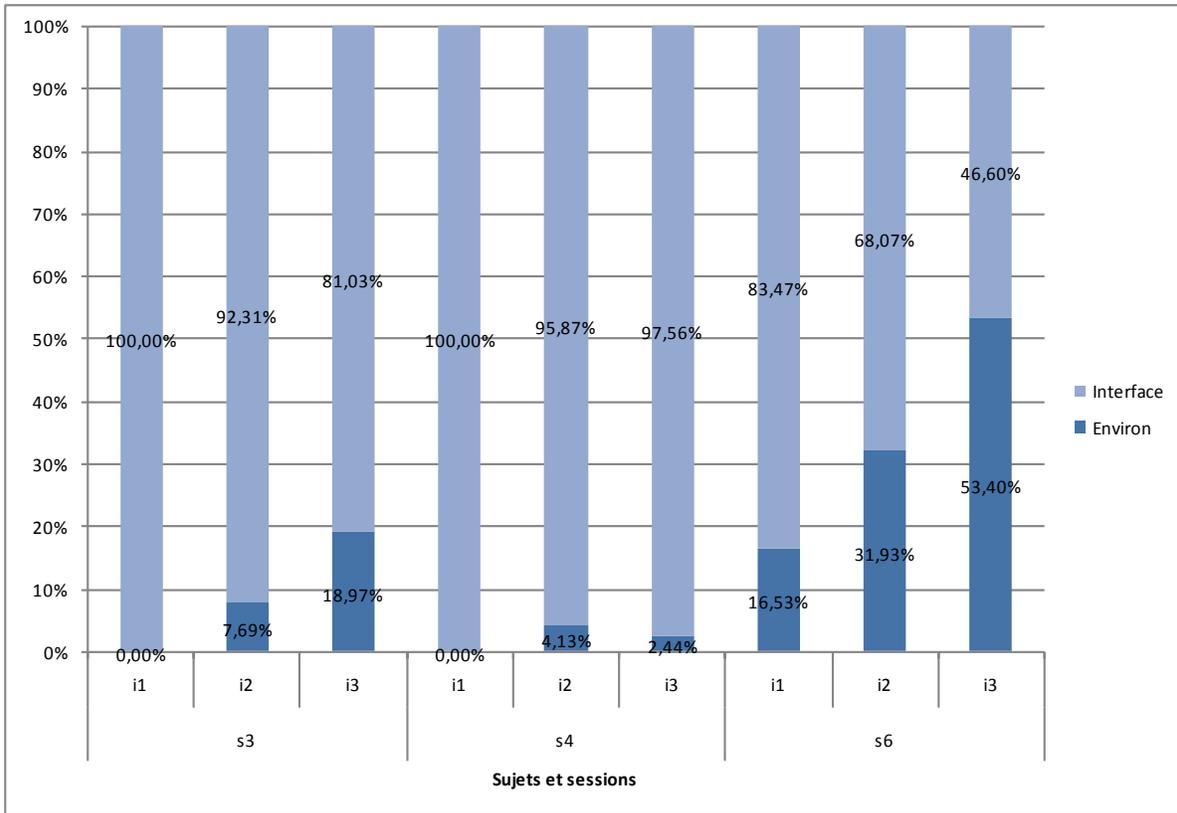


Figure 65 Evolution de la direction du regard selon les sujets

Il est intéressant de noter que pour les sujets 3 et 6, il y a une forte évolution de l'orientation du regard au cours des sessions (de 0% à 18,97% pour le sujet 3 et de 16,53% à 58,4% pour le sujet 6). Ces résultats semblent indiquer une bonne appropriation de l'application et en particulier de la modalité vocale permettant une amélioration de la libération de l'attention visuelle au fur et à mesure des sessions. Pour le sujet 4 les ratios de commandes orientés vers l'environnement sont très faibles, nuls pour la première session bien en dessous de 5% pour la deuxième et troisième.

#### 4.4. Positionnement du pda lors d'une commande vocale

Le passage d'une commande vocale peut entraîner différents comportements des utilisateurs. Ils peuvent soit approcher le pda du visage à la manière d'un microphone, soit laisser le pda en position basse pour passer la commande (la consigne donnée précisait qu'il n'était absolument pas nécessaire de porter le pda à la bouche pour passer une commande vocale).

Regard	Position micro	Position basse	Total
Interface	79,62%	20,38%	100,00%

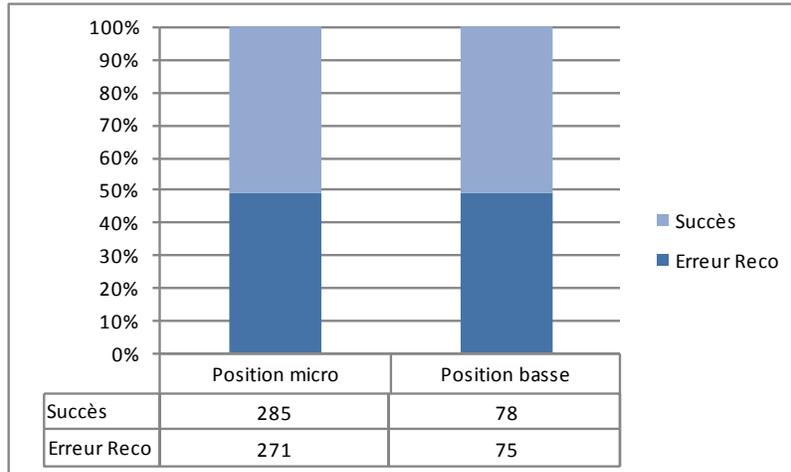
Tableau 27 Répartition de la position du pda lors des commandes vocales

Le Tableau 27 indique une forte tendance des utilisateurs à porter le pda en position micro pour effectuer une commande vocale (79,62%). Or, ces cas n'ont pas été comptabilisés dans la catégorie « environ » lors du codage de la direction du regard, car il s'agit souvent d'un mouvement relativement rapide d'aller retour qui ne relève pas réellement d'une intention de regarder l'environnement (lorsque c'est le cas, ces événements ont d'ailleurs été codés en tant que « environ »).

Cependant lorsque la commande est passée, par un certain opportunisme des sujets, le regard est généralement libéré pour une exploration visuelle rapide de l'environnement mais qui revient rapidement à l'interface. Comme nous le verrons à travers l'analyse des verbalisations, ce relâchement de la contrainte

visuelle a été fortement apprécié par les utilisateurs comme un avantage intrinsèque à la modalité vocale, même si cela ne constitue pas purement une interaction non visuelle au sens du code « environ ». En effet, dans ce cas, seule le « coup », c'est à dire la commande est non visuelle, mais toute la préparation et le contrôle post commande est à nouveau visuel.

Nous étudions ici rapidement une éventuelle influence du positionnement du pda sur les erreurs de reconnaissance vocale qui pourrait expliquer cette propension majoritaire des sujets à amener le PDA en position micro.



**Figure 66 Positionnement du pda et erreurs de reconnaissance vocale**

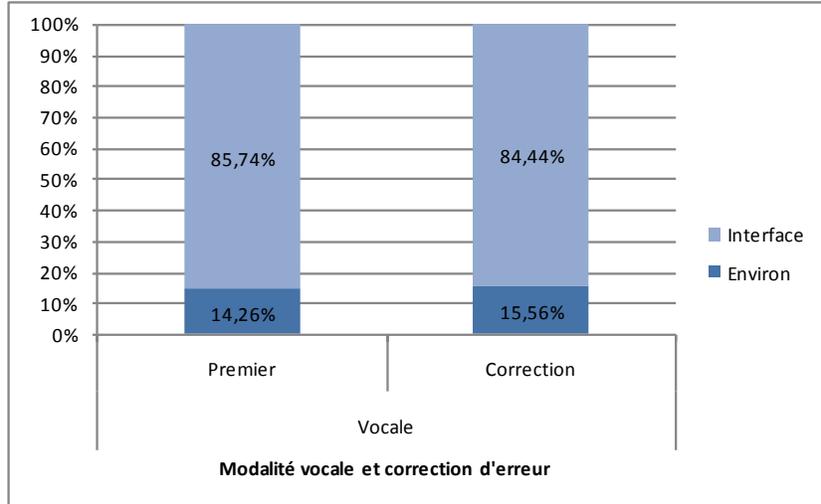
La Figure 66 indique que la position du pda au moment de la commande n'a aucune influence sur le taux d'erreur de reconnaissance vocale et ne peut donc être retenu comme critère objectif et explicatif de ce comportement (il se peut qu'au niveau subjectif, la perception des sujets aille tout de même dans ce sens, mais nous ne disposons pas de verbalisations en témoignant).

**4.5. Répartition des directions du regard selon les modes Correction ou Premier**

Nous cherchons dans cette partie à étudier la possibilité d'une influence de la modalité de correction d'une erreur sur l'orientation du regard et ainsi à répondre à la question suivante :

***La correction d'erreur mobilise-t-elle plus l'attention visuelle ?***

Nous regardons ici, la différence de l'orientation du regard en fonction du mode premier ou correction uniquement pour la modalité vocale qui est la seule à avoir des occurrences « environ ».

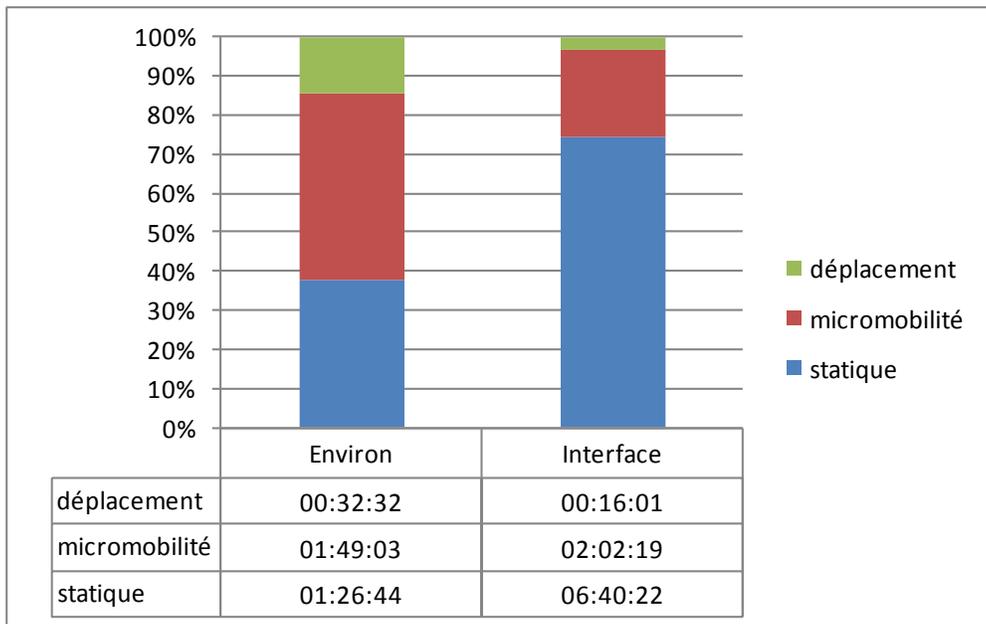


**Figure 67 Mobilisation de l’attention visuelle pour la modalité vocale en mode correction et premier**

La Figure 67 ne semble pas indiquer de différence significative entre les deux modes. Il semble même que les corrections en mode vocal (qui sont la plupart du temps des répétitions de la même commande) libèrent un peu plus l’attention visuelle. Ceci est cohérent avec les résultats sur les corrections et la mobilité. Il semble donc que la modalité vocale se prête à une correction moins coûteuse au niveau des ressources attentionnelles contrairement aux résultats répandus indiquant un coût cognitif élevé des corrections. Comme nous le verrons par la suite également dans la partie mobilité, lorsque les corrections de commandes vocales sont de simples répétitions de la même commande, son coût sur l’interaction est très faible.

**4.6. Regard et mobilité**

Pour étudier l’influence de l’orientation du regard sur le degré de mobilité, nous commençons par regarder en terme de temps, les durées d’états de mobilité et d’orientation du regard au niveau global.



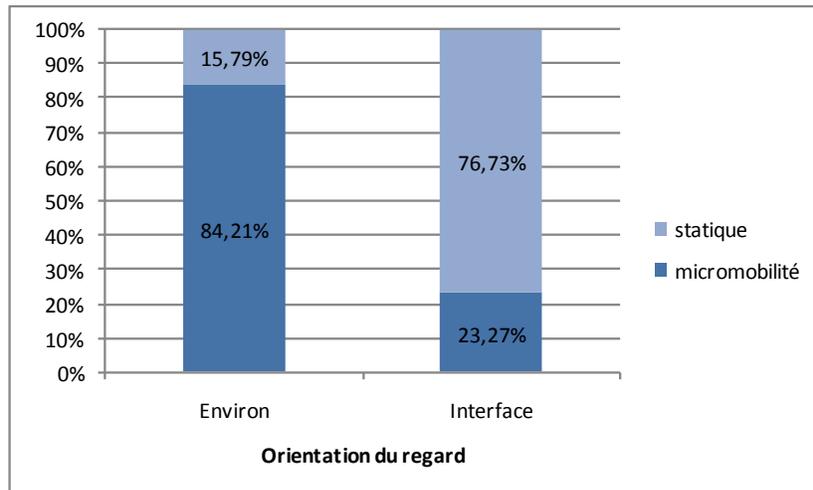
**Figure 68 : Durée des catégories de mobilité en fonction de l’orientation de l’attention visuelle tous sujets et toutes sessions**

La Figure 68 montre un ratio de catégorie « statique » de plus de 70% du temps où le regard est orienté vers l'interface et reste tout de même à un peu moins de 40% quand il est orienté vers l'environnement. En ce qui concerne la micromobilité, le ratio est proche de 20% pour « interface » et atteint quasiment 45% pour « environ ». La catégorie « déplacement » atteint plus de 15% pour « environ », mais il s'agit ici de cas particuliers dont nous ne tiendrons plus compte par la suite.

Deux résultats à retenir pour la suite de l'exposé :

- Lorsque les sujets regardent l'interface, ils sont largement statiques, cependant, ils parviennent tout de même à être en micromobilité 20% du temps.
- Lorsque les sujets regardent l'environnement, ils restent statiques tout de même 40% du temps et mobiles autour de 45%.

Pour la figure suivante, nous réintroduisons les interactions selon leurs catégories de mobilité et d'orientation du regard.



**Figure 69 : Catégories de mobilité en fonction de l'orientation de l'attention visuelle tous sujets et toutes sessions**

Comme nous l'avons vu dans la partie mobilité qui traitait de la répartition des interactions en fonction du degré de mobilité de l'utilisateur, plus des deux tiers des interactions se déroulaient en situation statique et moins d'un tiers en micromobilité. En croisant avec la direction du regard, les résultats sont encore plus marqués comme le montre la Figure 69 où plus des 84% des interactions orientées vers l'environnement ont été faites en mobilité. Ce taux tombe à 23% lorsque le regard est orienté vers l'interface. Il est donc assez clair que l'orientation du regard vers l'interface « ancre » les sujets en situation statique.

#### 4.7. Verbalisations concernant la catégorie « regard »

Nous avons obtenu un assez grand nombre de verbalisations concernant la perception de l'impact des usages des modalités sur l'attention visuelle, et conséquemment sur la possibilité de continuer une tâche principale ou parallèle. Un grand nombre de thématiques ont pu être abordées que nous avons thématiquées autant que possible.

- Partage de l'attention et attention floue (focus ailleurs que sur le lieu de l'interaction) (S1)

Le sujet 1 a fait très peu de commandes vocales, et il a donc quasiment exclusivement utilisé le stylet pour émettre ses commandes. La tâche de résolution du Quiz étant très mobile, il doit partager son attention entre l'interaction et la gestion de la tâche primaire ; il explique comment il réussit à relâcher un peu l'attention visuelle.

*S1 : Parfois je regarde pas vraiment l'interface, ?? mes yeux et je clique dans la direction et en fait je clique à côté quoi*

Le sujet explique qu'il essaie de déplacer son focus attentionnel, pour élargir sa vision périphérique quitte à perdre en précision sur le point central de l'interaction et ainsi à provoquer des erreurs ou des loupés de la cible qu'il cherchait à atteindre.

Cet exemple illustre bien la compétition qui existe quant à l'allocation des ressources attentionnelles et le choix de compromis de leur allocation auquel les sujets doivent procéder, au risque d'imprécision et de baisse de performance dans l'une ou l'autre des tâches.

- Effet des modalités sur l'attention visuelle (S3, S4, S6)
  - Effet de la modalité vocale (S3, S6) :

*S3 : Une fois que ça marche avec le vocal, du coup, je reste là quoi, c'est beaucoup plus pratique, parce que ce qui est bien c'est que quand je dis "envoyer", ça l'envoie et en même temps j'ai ma question tout de suite, donc j'ai l'impression qu'il ya deux actions qui sont faites simultanément, en même temps, je peux continuer à regarder ma question et puis, ça me dit dans les oreilles*

*En fait le fait de l'utiliser (la modalité vocale), ça permet en même temps de relâcher mon attention sur l'écran et de chercher autour de moi*

Cet exemple illustre bien chez certains sujets, l'impression de relâchement de l'attention visuelle que procure la modalité vocale qui permet de continuer une autre tâche (« je peux continuer à regarder ma question », « chercher autour de moi »).

- Interaction non visuelle (S6) :

Cet autre exemple illustre la capacité de certains sujets à effectuer des interactions avec de grandes phases non visuelles.

*Expé: tu arrives à faire presque tout sans regarder?*

*S6 : Ah ouais franchement, si tu sais ce que tu as à faire, jusqu'à ce que tu rentres le message ou même j'aurai pu aller jusqu'à répondre. Elle me lit le message, la voix, je peux faire insérer un son et répondre, j'ai même pas regardé le pda jusqu'à ce que j'ai enregistré le son quoi. Tu peux presque faire tout sans regarder*

Le sujet 6 est celui qui a le plus utilisé la modalité vocale et en particulier qui a fait un assez grand nombre de commandes non visuelles. Il explique que la connaissance à l'avance du comportement de l'application, ainsi que la sortie sonore et la synthèse vocale lui permettraient d'envisager une interaction absolument non visuelle si l'application n'était pas limitée techniquement (il était impossible d'insérer des messages audio à l'aide de la modalité vocale).

- Modalité tactile stylet (S6) :

A l'inverse de l'exemple précédent, le sujet 6 par exemple commente l'effet de l'usage de la modalité tactile sur l'attention visuelle en comparaison de la modalité vocale.

*S6 : C'est vrai que t'es pas tout le temps (le regard) dessus (avec la modalité vocale), quand t'écris oui. Tandis qu'avec le stylet t'es obligé de regarder le pda*

Son expérience montre que relativement à la modalité vocale, l'usage du stylet est beaucoup plus coûteux en ressources visuelles (« avec le stylet t'es obligé de regarder le pda »).

- Facteurs influençant la charge attentionnelle :

- Appropriation de l'application (S4, S6) :

Dans la continuité des remarques du sujet 6, le sujet 4 indique que l'appropriation de l'application joue un grand rôle dans la capacité à libérer l'attention visuelle de la modalité vocale.

*S4 : La reco, oui, parce que pendant ce temps là, on peut regarder ailleurs quand même, ça c'est clair, même si comme au début, l'appareil on le connaissait pas bien donc on a quand même les yeux rivés dessus*

En effet, le fait d'utiliser la modalité vocale ne permet d'alléger la charge cognitive que si l'on connaît exactement les actions à effectuer, car au-delà d'une commande individuelle vocale, il y a toute la gestion de l'interaction et des enchaînements d'actions à effectuer.

- Enchaînement d'activités visuelles (qui poussent à garder le regard dans la position courante) :

Plusieurs verbalisations montrent la présence de cycles de tâches requérant plus ou moins d'attention visuelle et incitant donc le sujet à ne pas modifier l'état courant de l'orientation du regard. Au moins deux exemples ont été cités : pour le premier, il s'agit de tâches d'interaction très visuelles focalisées sur le PDA, telles que la recherche et la sélection de dessins ou la lecture de messages par exemples. Mais il en est de même pour des tâches plutôt situées dans l'environnement (S3) :

*S3 : En fait le fait de l'utiliser (la modalité vocale), ça permet en même temps de relâcher mon attention sur l'écran et de chercher autour de moi*

*Expé : Le relâchement visuel, ça c'est vrai pour les autres commandes aussi (fermer etc)*

*S3 : Oui, mais c'est différent, ya des moments où j'ai besoin de regarder à côté parce que je cherche un indice, t'en as d'autres, non. C'est quand j'ai plus de temps, dans ce cas là, je préfère prendre mon temps, quand j'en ai pas, je préfère que ce soit fait automatiquement (en vocal)*

Ici, le sujet a fait une commande vocale alors que les commandes précédentes étaient plutôt tactiles et il justifie ce fait en précisant que la tâche qu'il était en train d'effectuer était demandeuse en attention visuelle pour explorer l'environnement afin de chercher un indice. Il semble donc qu'ici ce soit l'exigence de mobilité et d'exploration visuelle de la tâche qui aient contribué au choix d'usage de la modalité vocale.

- Effet du design sur l'attention visuelle

Les sujets ont fait quelques retours concernant l'impact de certains choix de design sur l'utilisabilité de l'interface, que l'interaction soit tactile ou vocale.

- Taille et proximité des icônes (S1) :

Plusieurs constatations concernant la taille des icônes et leur proximité ont été mentionnées par les sujets :

*S1 : C'est vrai que les icônes je trouve qu'elles sont pas très ... elles sont petites, on peut pas aller dessus un peu les yeux levés, il faut quand même bien cliquer précisément dessus sinon il veut pas.*

*Ça oblige à bien regarder. Autant pour l'enregistrement (grosses icônes) à la limite, je suis sûr que je pourrais pas trop regarder et quasiment mettre dessus.*

*Là c'est vrai que les petites icônes des fichiers son et dessin elles sont justes à côté les unes des autres*

Le sujet fait ici remarquer l'impact de la petite taille des icônes et de la difficulté que représente la visée sur les ressources visuelles (« **Ça oblige à bien regarder** »). Il fait également remarquer que les icônes plus grosses au contraire permettent quasiment une interaction non visuelle.

La deuxième partie de la remarque concerne la proximité des icônes entre-elles venant s'ajouter à leur petite taille et rendant la visée et l'atteinte encore plus coûteuse. Il faut également remarquer que ces commandes particulières sont parmi celles qui ont les plus forts taux d'interactions vocales.

- Emplacement des icônes (S6) :

L'emplacement des icônes semble avoir aussi joué en faveur d'une interaction vocale plutôt que tactile :

*S6 : (fais une commande vocale) C'est vrai, je suis pas beaucoup descendu sur les icônes en bas, tout ce qui était insertion de trucs, je l'ai fait quand même plus à la voix que... (au stylet) je trouve que les icônes en bas, tu les vois pas tellement, ils sautent pas aux yeux quoi, tu les vois si tu regardes, mais ton regard, il se dirige plus vers le haut du pda que vers le bas, la partie en bas, tu la vois moins*

Là encore, le fait que les icônes (les mêmes petites icônes que mentionnées dans l'exemple précédent) soient en position basse de l'interface ne semble pas favoriser leur accès avec la modalité tactile stylet, et le cas échéant, cette interaction sera encore plus couteuse du fait de la moins bonne accessibilité visuelle perçue.

- Effet des sorties sonores et visuelles sur l'attention visuelle

Suite aux remarques concernant le design en entrée, nous abordons ici quelques points liés aux effets des sorties sonores et visuelles sur l'attention dont l'usage de la synthèse vocale (en lecture de médias et feedbacks), la présence de feedbacks sonores et visuels.

- Usage et effet de la Synthèse vocale sur l'attention visuelle (S2, S3, S4, S6) :

Dans notre application, la synthèse vocale a deux fonctions :

- ✓ Une fonction de lecture de médias dont le contenu est textuel (nom de l'expéditeur, objet des messages, contenu des messages)
- ✓ Une fonction de production de feedbacks d'effectuation d'une commande (par exemple des confirmations d'actions effectuées du type « message envoyé »), qui sont des régulateurs de l'interaction

Dans ce paragraphe c'est la première acception qui est commentée.

De nombreux sujets ont mentionné que la synthèse vocale était un atout pour libérer l'attention visuelle avec cependant parfois des critiques concernant son utilité pas toujours avérée.

*Expé : tu écoutais la synthèse vocale?*

*S4 : J'écoutais et ça permettait de regarder ailleurs que le truc (pda) et comme juste avant j'avais visité la petite bougie et en fait j'avais vu la réponse à la question qu'elle était en train de me lire, j'étais un peu aidé*

Le sujet 4, comme la plupart des autres sujets (2, 3, 6) indique que la synthèse vocale lui « permet de regarder ailleurs » que l'interface. Dans ce cas, le sujet doit se déplacer pour aller chercher la réponse précise à une question qu'il est en train de se faire lire, ce qui lui permet de se déplacer en même temps.

*S4 : Nan, je l'ai pas utilisée (la synthèse vocale), non. Non, parce que ça sert à rien, enfin je veux dire, si j'ai les yeux ailleurs, oui, je vais l'utiliser, mais là, j'ai les yeux rivés au pda, donc pour lire deux lignes ça sert à rien. A la limite si c'est long d'accord, mais là c'est pour lire deux trois lignes, ça sert vraiment à rien.*

Dans la suite de la verbalisation, il ajoute cependant une nuance sur les conditions de l'efficacité de la synthèse vocale et notamment le peu d'intérêt de cette fonctionnalité lorsque la quantité de texte à lire est faible et qu'une lecture visuelle est plus rapide qu'une écoute parfois un peu fastidieuse. Il précise tout de même en ajoutant que si son attention était déjà dirigée ailleurs que sur l'interface alors que la synthèse démarre, cela pourrait être profitable.

*S4 : Non, mais j'avoue, là que quand elle parle en même temps, c'est assez hallucinant n'empêche, elle lit ce qu'il y a écrit, ça permet d'avoir les yeux rivés sur autre chose que sur le système*

Il fait également remarquer (ainsi que les autres sujets) la forte impression que lui a faite la synthèse vocale quant à sa capacité à lire correctement un message complexe à la volée, lui permettant ainsi une interaction moins visuelle, voire non visuelle.

- Feedback audio (sonore et synthèse) (S3, S4, S6) :

Les feedbacks audio correspondent à la seconde acception des sorties sonores présentées plus haut. Leur fonction de phatique ou régulateur d'écoute est illustrée par le sujet 3 :

*Expé : En passant la commande vocale, tu regardes ailleurs, tu n'attendais pas le retour ?*

*S3 : Non puisque c'est audio, donc forcément, j'allais entendre un son quoi, du coup je m'attardais pas dessus.*

*Expé : tu t'es pas sentie obligée de regarder ?*

*S3 : Non, j'avais confiance, ça allait me dire quelque chose dans tous les cas*

Ici le sujet intègre très bien dans sa représentation du comportement du système, la présence de feedbacks audio, à la fois sous forme de clics audio (petits clics sonores lorsque le système reçoit une commande) servant de feedbacks de prise en compte de la commande mais aussi de feedbacks en synthèse vocale pour la confirmation de l'exécution de la commande (« message supprimé » par exemple).

Cette propriété de l'application lui permet de relâcher son attention visuelle puisque la régulation et le contrôle de l'interaction seront effectués de manière sonore (« tu t'es pas sentie obligée de regarder ? Non, j'avais confiance, ça allait me dire quelque chose dans tous les cas »).

- Besoin de confirmation visuelle (S6) :

Le sujet 6 confirme ici la possibilité d'effectuer des interactions non visuelles en entrées comme en sortie, cependant, il apporte une nuance concernant le besoin de confirmation de prise en compte de la commande :

*S6 : Je vérifie juste un petit coup d'œil, alors qu'à l'usage y aurait pas besoin je pense, on doit pouvoir s'affranchir des vérifications comme ça, mais tu vérifies toujours si la commande a été prise en compte. Mais je pense que si je l'utilise deux mois, trois mois, au bout d'un moment, je regarde plus le pda en fait*

Dans sa verbalisation, le sujet 6 déclare avoir encore besoin d'une confirmation ou vérification visuelle (à l'inverse du sujet précédent qui déclarait avoir confiance dans la capacité du système à l'informer de son état). Cependant, il ajoute qu'avec un temps d'appropriation supérieur, cette vérification visuelle ne serait plus nécessaire. Il faut également remarquer que ce sujet est celui qui, de loin, a fait le plus de commandes totalement non visuelles à l'aide de la commande vocale.

- Simplicité de l'application (S6) :

Dans ce même esprit, et en continuité avec la remarque précédente, le sujet relève que la complexité de l'application est un facteur à prendre en compte pour rendre possible une interaction non visuelle :

*S6 : Tu peux presque tout gérer sans regarder le pda, surtout parce que l'application elle est pas très compliquée*

En effet, la relative simplicité de l'application a permis aux sujets de s'en construire une représentation et ainsi de mémoriser et prévoir ses comportements. Les heuristiques apprises lors des interactions visuelles ont ainsi permis leur reproduction soit avec d'autres modalités, soit de manière non visuelle,

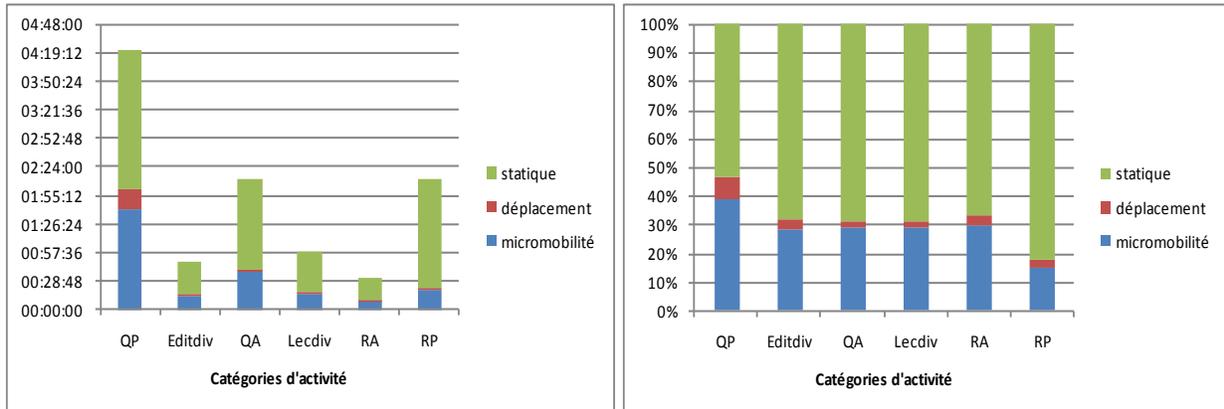
entre autre grâce au principe d'équivalence totale des modalités. L'apprentissage du fonctionnement est donc suffisamment simple pour mémoriser les étapes d'interaction sans qu'elles soient médiatisées par un support cognitif visuel.

De plus, au delà des remarques émises par les sujets, il faut remarquer que le type de l'application et les tâches qui y sont associées se prêtent à une interaction éventuellement moins visuelle. En effet, la totalité des commandes sont des commandes discrètes facilement effectuables avec l'une ou l'autre des modalités l'équivalence fonctionnelle (à l'inverse de commandes continues de type manipulation directe).

## 5. Mobilité et Activité : Effet de l'activité ?

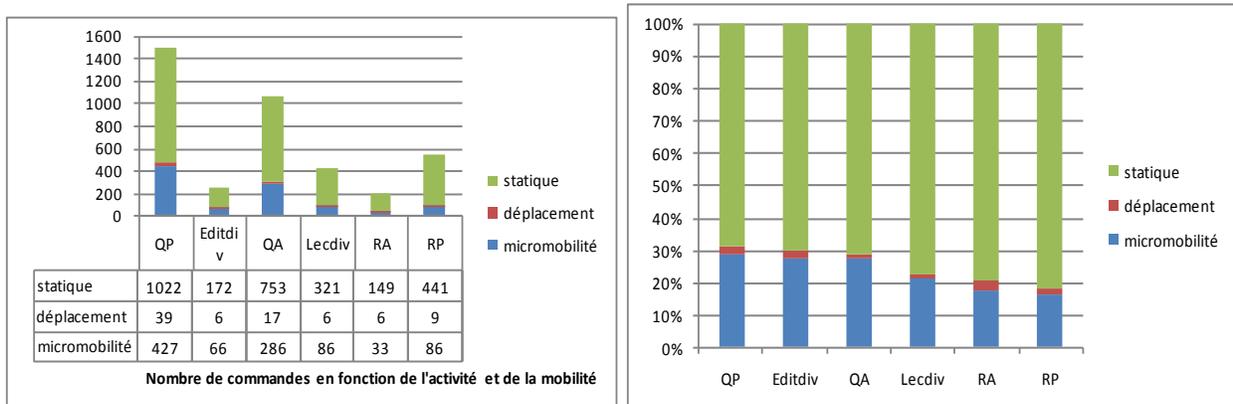
Nous croisons ici les catégories d'activité réelle et la mobilité afin d'identifier un éventuel lien entre les deux.

Nous observerons dans un premier temps les durées puis les interactions.



**Figure 70 Simultanéité de l'activité et de la mobilité**

Comme nous l'avons vu dans la partie « Activité » précédente (partie Influence de l'activité réelle sur l'usage des modalités), c'est le traitement des questions personnelles et la recherche de la réponse (catégorie QP) qui occupent la plus grande part de l'activité des sujets, viennent ensuite la réponse aux questions personnelles (RP) et le traitement des questions de l'autre (QA) qui durent environ la moitié de QP. Toutes ces activités ont une part minimum de mobilité supérieure à environ 18% du temps (minimum pour RP) de chaque catégorie. Toutes les autres catégories d'activité ont un ratio de mobilité supérieur à 30% du temps avec un pic à plus de 45% pour QP. Les activités sont donc toutes majoritairement statiques, avec cependant une base mobile relativement importante.



**Figure 71 Nombres de commandes mobiles selon l'activité**

Si l'on s'intéresse à l'ordre du nombre de commandes par activité (Figure 71), nous pouvons constater que celui-ci correspond globalement à l'ordre des durées de chaque activité (avec dans l'ordre décroissant : QP, QA, RP, Lecdiv etc.).

Et si l'on compare avec les résultats du tableau précédent, nous pouvons voir que l'ordre de la mobilité est à peu près le même mais avec des ratios de commandes mobiles plus faibles (avec des maximums autour de 30% de commandes mobiles pour QP, Editdiv, QA). Dans le contexte des activités de notre

expérimentation, il y a donc une certaine cohérence dans le degré de mobilité entre la durée passée en mobilité et le nombre de commandes passées pour un type d'activité.

## 6. Transitions entre les états de mobilité

L'étude des transitions d'états de mobilité permet de donner un aperçu de la nature des enchaînements successifs de mobilité en fonction des interactions. Pour ce faire, nous avons relevé systématiquement l'état de mobilité précédent une commande afin d'en évaluer l'impact sur l'état suivant de mobilité.

### 6.1. Changements de mobilité

Dans cette première partie, nous commençons à regarder à un niveau global, l'enchaînement des états successifs de mobilité sans tenir compte des modalités mises en œuvre.

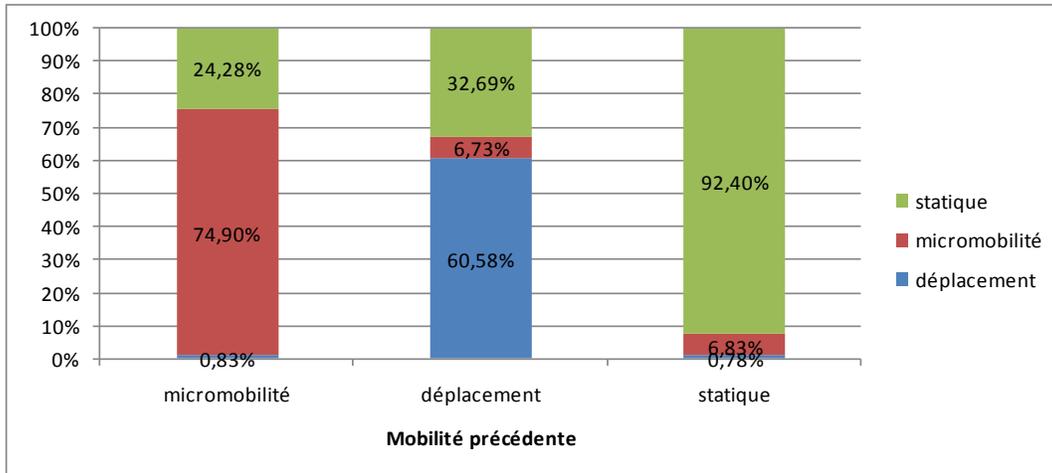


Figure 72 : Transition de mobilité entre une commande et la suivante

D'une manière générale, les états de mobilité montrent une grande continuité d'une commande à l'autre. En effet, les états de mobilité pour une commande et la suivante sont massivement identiques pour « statique » (92,40%) et très élevés également pour « micromobilité » (74,90%) avec un minimum pour « déplacement » de 60,58%.

Il semble donc, au vu de ces résultats, que le besoin d'interaction ne modifie que dans des proportions faibles les états de mobilités et que les sujets s'adaptent autant que possible au contexte de mobilité en cours.

Cependant, si nous considérons les résultats dans le détail, même si globalement, la continuité est forte, nous pouvons tout de même noter des différences non négligeables de continuité entre les états de mobilité, avec 32% d'écart entre « statique » et « déplacement » par exemple.

L'état « statique » est celui qui a donc le plus tendance à se renouveler en cas de commandes successives. Autrement dit, lorsque les sujets ont fait une commande à l'arrêt, les commandes suivantes étaient également massivement faites à l'arrêt.

Ce phénomène est tout de même moins marqué pour les autres états de mobilité précédentes tels que « micromobilité » ou « déplacement ». Lorsque les sujets ont fait des commandes en mode « déplacement », celui-ci n'a pu être conservé que dans 60% des cas, ce qui est tout de même moins important que dans le cas du mode « statique ». Plus de 32% des commandes effectuées en « déplacement » sont suivies d'une commande statique, ce qui indique tout de même une forte propension à l'arrêt de la mobilité lié à une interaction.

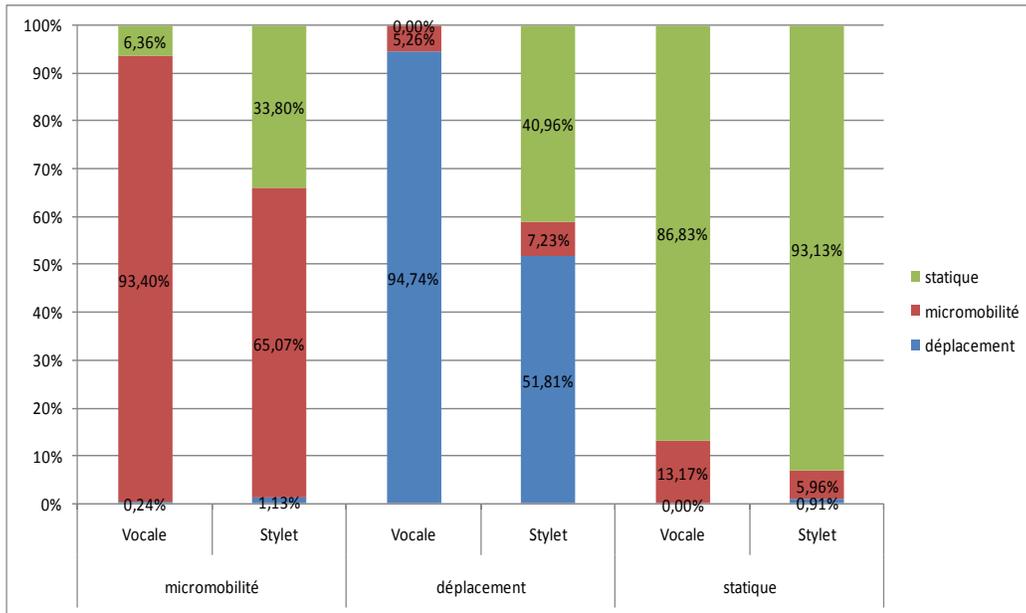
Pour le mode « micromobilité », les résultats sont intermédiaires avec un taux de renouvellement de la mobilité de presque 75% et un taux d'arrêt autour de 24% indiquant qu'il n'y a donc pas systématiquement arrêt de la mobilité pour interagir. Le mode « micromobilité » présente des caractéristiques intermédiaires aux deux autres modes avec une interruptibilité ou une continuité plus souple ou du moins plus utilisée par les sujets. Ce mode semble en effet représenter un compromis d'allocation possible des ressources attentionnelles et cognitives à la fois sur la gestion du déplacement et

l'interaction. Cette attention flottante permet des transitions plus souples entre les modes de mobilité car les deux tâches sont toujours en cours en parallèle.

En revanche, l'interaction statique, qu'elle soit due à une tâche d'interaction très demandeuse en ressource visuelle ou à une activité de nature statique extérieure au PDA (réponse à une question du Quiz), ne fait pas intervenir de gestion du déplacement, et permet donc une plus grande mobilisation des ressources attentionnelles sur l'interaction, qui ne favorise pas particulièrement une reprise fluide de la mobilité. Comme l'ont souligné à plusieurs reprises les sujets, l'interaction provoque souvent un effet de capture des ressources cognitives et attentionnelles qui restreint la capacité à effectuer une autre tâche en parallèle.

**6.2. Changements de mobilité selon les modalités**

Nous réintroduisons maintenant les modalités dans l'étude des transitions afin de déterminer leur éventuel impact sur les états de mobilité.



**Figure 73 : Enchaînement des états de mobilité selon les modalités**

Au regard des résultats précédents, la Figure 73 montre des différences parfois importantes sur les enchaînements d'états de mobilité selon la modalité mise en œuvre pour la commande suivante. De manière générale, lorsque la commande suivante est vocale, les taux de mobilité sont plus importants que lorsque la commande est tactile avec des écarts parfois très importants allant de plus 6% jusqu'à plus de 40% ( Tableau 28 ).

**Tableau 28 : Différences pour l'état « statique » selon les modalités**

Mobilité précédente	Modalité suivante		Différence
	Vocale	Stylet	
<b>micromobilité</b>	33,80%	6,36%	27,44%
<b>déplacement</b>	40,96%	0,00%	40,96%
<b>statique</b>	93,13%	86,83%	6,31%

Lorsque le mode précédent est « statique », la modalité qui suit a une incidence faible sur l'état suivant avec, comme nous l'avons vu au paragraphe précédent, une tendance massive à conserver un état « statique » quelque soit la modalité d'interaction suivante (valeurs au delà de 86% pour les deux modalités). La différence entre les deux modalités est tout de même supérieure à 6% avec un ratio de mobilité supérieur pour la modalité vocale favorise un peu plus un démarrage de mobilité.

Pour le mode de mobilité précédent « déplacement », en revanche, la différence est fortement marquée selon la modalité suivante avec une différence d'état « statique » de plus de 40%. Il faut noter également qu'avec la modalité vocale, il n'y a aucune interruption du mode « déplacement ». Les résultats concernant ce mode, sont à relativiser en raison du très faible effectif des données.

Pour le mode de mobilité précédent « micromobilité », les résultats indiquent également de fortes différences (27,44% pour le mode « statique ») selon la modalité utilisée à la suite. La modalité vocale permet donc une bien meilleure continuité de micromobilité dans l'interaction et à l'opposé, l'usage de la modalité « stylet » provoque d'avantage d'interruption de mobilité.

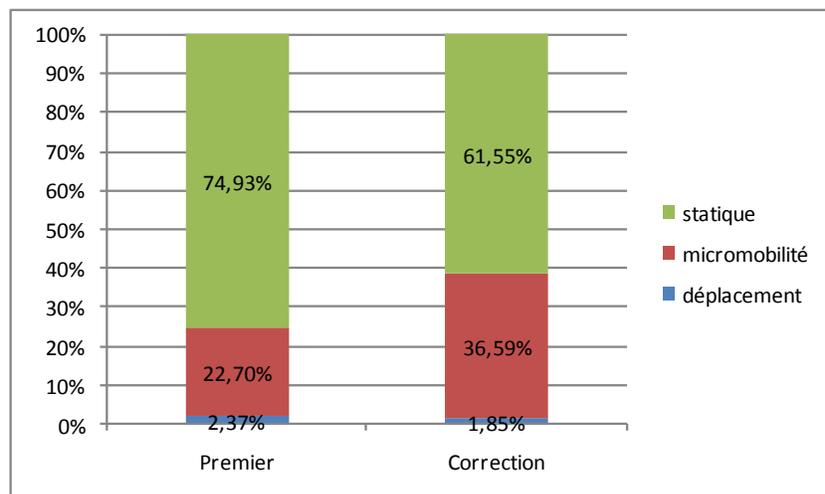
## 7. Correction des erreurs

Pour la suite de l'analyse, nous dissocions les commandes dites « premières » des commandes de correction. Les commandes premières correspondent à la première tentative de commande du sujet qui éventuellement, donne lieu à une erreur (de tout type) et qui est généralement suivie d'une commande de correction.

### 7.1. Influence des corrections sur la mobilité

**Quelle est l'influence des corrections d'erreurs sur le niveau de mobilité ?**

Nombre de Commande	de Mobilité			Total
	déplacement	micromobilité	statique	
Corr/Prem				
Premier	2,37%	22,70%	74,93%	100,00%
Correction	1,85%	36,59%	61,55%	100,00%
Total	2,30%	24,71%	72,99%	100,00%



**Figure 74 Répartition des catégories de mobilité selon le type de commande (premier et correction)**

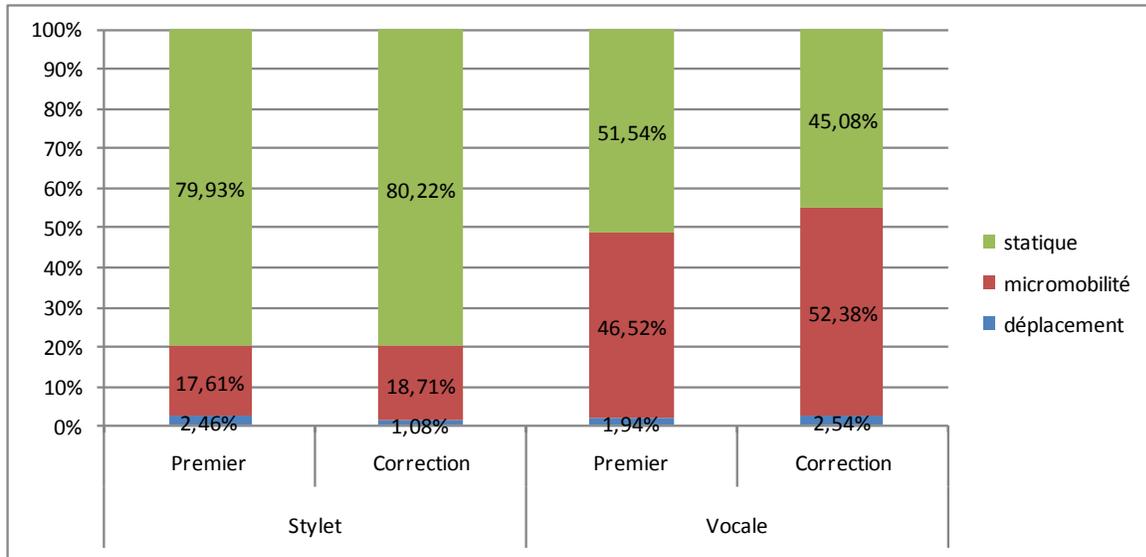
Au niveau global (toutes modalités confondues), la correction des erreurs semble avoir une influence sur le degré de mobilité des sujets. Contrairement à ce que l'on pourrait attendre, les corrections d'erreurs semblent permettre un plus haut niveau de mobilité que les commandes premières puisque les commandes statiques sont à quasi 75% pour les commandes premières, alors qu'elles baissent à 61,55% pour les corrections.

**1. Le mode correction semble permettre plus de mobilité que le mode premier (toutes modalités confondues)**

### 7.2. Influence des modalités de correction sur la mobilité

Nous cherchons ici à approfondir le constat précédent afin d'expliquer cette augmentation de la mobilité en mode correction. Dans un premier temps, nous nous posons la question suivante :

**La modalité de correction influence-t-elle le niveau de mobilité ?**



**Figure 75 Influence des modalités de correction sur la mobilité**

La Figure 75 permet d'observer plusieurs résultats intéressants :

2. **Comme nous l'avons déjà observé, la modalité tactile stylet présente un faible degré de mobilité (en dessous de 19%) alors que la modalité vocale atteint jusqu'à 52% de commandes en mobilité.**

Pour la modalité tactile stylet, le mode premier ou correction ne présente pas de différence quant au degré de mobilité (17,61% contre 18,71%).

Au contraire, pour la modalité vocale, le mode correction est effectué davantage en mobilité que le mode premier (de 46,5 à 52,38% pour la micromobilité en mode correction).

Au niveau global, les corrections se font à 78,4% en micromobilité et à 21,76% en statique.

Les corrections en modalité vocale se font à 57,87% en micromobilité, ce qui semble jouer en faveur de la continuité de l'activité de déplacement. Les corrections en modalité stylet, quant à elles, ne permettent que 20,37% de micromobilité.

## 8. Perceptions des utilisateurs

Les verbalisations et leur analyse que nous présentons ici concernent la perception que les utilisateurs ont eue de la possibilité d'interagir en mobilité de manière générale, en rapport avec l'application puis

### 8.1. Remarques générales liées à la mobilité et à l'application

Comme nous l'avons déjà mentionné plusieurs fois, la tâche que les utilisateurs avaient à réaliser a été explicitement construite pour provoquer un partage des ressources attentionnelles sur plusieurs focus en rapport avec des objets de la tâche différents.

Une partie de cette tâche consistait à explorer l'environnement en se déplaçant afin de trouver les réponses aux questions du Quiz, une autre consistait à interagir avec le PDA pour lire ses questions, y répondre et collaborer avec son partenaire. Il y a donc de fait un partage, voire une concurrence de l'allocation des ressources physiques et cognitives liées à la nature multitâche de cette activité.

- Objectifs divergents et caractère multitâche de l'activité (voire concurrents)
  - Exploration de l'exposition et interaction (S1, S6) :

*S1 : Ouais, là j'essayais d'avancer et en même temps de lire mes questions pour regarder, j'y arrivais pas quoi, dès que je commençais à cliquer sur le palm, c'était fini, je plongeais dedans, je pouvais pas avancer, je voyais pas à côté en fait*

Dans cet exemple, le sujet 1 illustre bien la nature multitâche de la consigne et de l'activité qui y est associée. Il décrit les difficultés qu'il rencontre afin de réaliser ses deux objectifs parallèles qui demandent des ressources similaires en simultané (ici se déplacer et lire les questions sur le PDA sollicitent toutes deux fortement la vision).

Dès qu'il commence à interagir (au stylet), sa vision est focalisée sur l'interaction (« je plongeais dedans ») et ne lui permet plus de se déplacer. Il est à noter que ce sujet est un de ceux qui a le moins eu recours à la modalité vocale.

*S6 : Mais après je voulais utiliser la reco parce que c'est bien, tu parles et tu te ballades. J'avais envie de regarder les trucs, de voir l'expo*

Dans cet autre exemple, le sujet 6 qui a beaucoup utilisé la modalité vocale, justifie ce choix en partie à cause de son désir de pouvoir visiter l'exposition en même temps qu'il interagit (ce que lui permet la modalité vocale).

- Partage de l'attention et Gestion de l'activité multitâche (S3) :
  - Compromis d'allocation de ressources cognitives et visuelles :

A l'inverse, le sujet 3 fait remarquer qu'il réussit quand même à partager son attention visuelle entre le déplacement et l'interaction à condition que celles-ci ne soient pas trop demandeuses :

*S3 : J'ai pas beaucoup regardé quand même; avec la vision que t'as, tu vois quand même quand t'avances, tu vois si ya un obstacle ou pas. J'étais quand même sur le pda mais j'étais pas vraiment rivé dessus, je regardais quand même à côté*

Il réserve une part de la vision périphérique pour une gestion du déplacement relativement grossière (« tu vois si ya un obstacle »). On note donc un compromis où il réduit les ressources allouées à l'interaction (« j'étais pas vraiment rivé dessus ») afin d'assurer un minimum de contrôle de son déplacement.

- Attention « flottante », « floue » ou focalisée sur l'interface selon l'activité (attention épisodique) :

Lorsque l'interaction est majoritairement tactile, les sujets doivent mettre en œuvre des stratégies pour réussir à se déplacer en même temps lorsque c'est possible :

*S1 : En fait ce que je faisais, je me déplaçais en même temps je cliquais quoi, je me déplaçais juste en même temps je cliquais, je regardais pas vraiment. Marcher, ouvrir les trucs, j'y arrivais quoi en fait, mais après bon, pour les lire vraiment, enfin les consulter, j'ai eu l'impression qu'il fallait pas tracer quoi, qu'il fallait pas aller trop vite, je cliquais, j'attendais un petit peu en marchant, je faisais les choses tranquillement*

Le sujet 1 explique qu'il régule l'attention qu'il porte à l'interaction en fonction de l'activité précise qu'il est en train d'effectuer car selon lui, les différentes tâches qu'il peut avoir à effectuer ne sollicitent pas toutes les mêmes ressources. Il précise par exemple qu'il peut très bien réussir à marcher tout en ouvrant un message en ayant une attention « floue » sur l'interface (regard non concentré pendant l'interaction), mais que par contre, il lui est impossible de se déplacer s'il doit lire le contenu détaillé d'un message.

*Expé : ça te dérange pas de marcher en interagissant ?*

*S3 : Ah non, j'ai l'habitude, j'arrive à faire des trucs en marchant, J'envoie des sms ou des mms en poussant avec la poussette ! Obligée quoi, avec une main, ça va. C'est un peu galère mais ça va*

Le sujet 3 fait ici remarquer qu'il est habitué à partager ses ressources cognitives et physiques au quotidien où il effectue déjà régulièrement des activités en parallèle (« J'envoie des sms ou des mms en poussant avec la poussette ! »), même si cela n'est pas facile.

- Baisse de la performance (cognitive) due au partage de tâches (S2,

Le sujet 2 et de manière générale tous les sujets, notent une baisse de leurs performances pour interagir lorsqu'ils se déplacent :

*S2 : C'est vrai que si on bouge en même temps, on est moins efficace sur le pda*

La concurrence ou compétition d'allocation des ressources va donc se jouer à plusieurs niveaux dépendants parfois les uns des autres. Les exemples vus précédemment illustraient aussi bien les sollicitations cognitives que physiques (mobilisation du regard et des membres en général) qui peuvent varier selon les activités en cours.

- Orientation de la direction du regard, tête basse (S1,

Dans cet esprit, le sujet 1 détaille les raisons qui l'empêchent de pouvoir interagir en mobilité réelle en dissociant bien la gestion du déplacement du point de vue moteur, et du point de vue des ressources attentionnelles :

*S1 : Dès que je me sers de l'ordinateur, c'est vrai que je voyais plus du tout ce qui se passait, je pouvais pas avancer. Enfin, si je pouvais avancer en même temps, j'sais pas avec quelque chose, des lunettes infrarouge ou un écran, je le ferais, c'est pas le mouvement qui m'embête, c'est vraiment le fait de rien voir quand je marche.*

Il précise bien que ce qui l'empêche de se déplacer est la focalisation visuelle en tête basse sur l'interface qui ne lui permet pas d'appréhender l'environnement dans lequel il veut se déplacer. Il ajoute également que la gestion du déplacement au sens moteur ne lui pose pas particulièrement de problème.

Selon lui, c'est donc plus la compétition d'allocation de la ressource visuelle qui constitue éventuellement un frein à la mobilité, le déplacement étant peu sollicitant du point de vue strictement moteur. Il va jusqu'à

donner des exemples d'outils qui pourraient lui permettre de voir à la fois l'interface et l'environnement (« des lunettes ou un écran ») et ainsi d'interagir en mobilité réelle.

- Focalisation sur l'interaction :

Les sujets ont souvent fait mention de la réduction du champ de vision accompagnant l'interaction, allant parfois jusqu'à une sensation d'isolement sensoriel et cognitif.

*S3 : Je suis resté dans le truc (pda), j'avancais, et j'étais plus là, j'étais carrément dans le pda*

Ce phénomène a également été observé quant à la perception de l'environnement sonore et social, de nombreux sujets mentionnent une sorte d'effet tunnel, dans lequel le focus sur l'interaction est tel qu'ils se ferment cognitivement à l'environnement.

- Orientation difficile (S2)

Une des conséquences de la focalisation sur l'interface est la difficulté à s'orienter pour rechercher des indices dans l'exposition :

*S2 : Surtout comme j'avais les yeux rivés sur le pda, j'avais du mal à m'orienter*

Comme nous l'avons vu précédemment, la simultanéité d'une interaction et d'un déplacement est éventuellement possible si la gestion du déplacement mobilise des ressources cognitives de bas niveaux du type réaction à l'environnement plus de l'ordre du tactique. En revanche, lorsqu'il s'agit de tâches plus stratégiques d'orientation ou de planification de trajet, mobilisant des ressources cognitives de haut niveau (traitement symbolique d'informations), la simultanéité d'une interaction devient difficile voire impossible.

- Stratégies permettant les activités en parallèle

En raison des difficultés rencontrées par les sujets pour gérer en parallèle le partage des ressources attentionnelles, de nombreuses stratégies sont mises en place au fil du temps et de l'appropriation de l'application.

- Appropriation de l'application, gestion du temps de réaction du système etc. :

*S1 : Là j'arrivais un peu mieux à me déplacer dans le sens où je connaissais les temps, donc je cliquais en avançant, j'avais peut-être mieux la mesure des choses, je savais combien de temps il mettait, je savais que je pouvais regarder vite fait, jeter un coup d'œil et en même temps marcher. Un peu mieux ouais, mais bon pareil, je pouvais pas non plus trop me balader, c'était plus essayer de changer de salle, le temps que tout arrive, le temps qu'il se mette en route.*

De nombreux sujets sont abordés dans cet extrait de verbalisation. Le sujet rappelle qu'il est difficile de se déplacer de manière précise en interagissant, mais qu'avec la connaissance plus précise de l'application, il arrive à prédire son comportement et ainsi à libérer son attention visuelle pendant de courtes périodes, mais suffisamment pour réussir à se déplacer en réaction à l'environnement (déplacement plutôt tactique et opportuniste). Il y a donc une alternance plus ou moins gérée et prévisible de phases d'interaction et de gestion du déplacement.

*S2 : Ben je pense que comme j'étais beaucoup mieux habitué maintenant à manipuler le pda, ça m'arrivait de marcher en écrivant, alors qu'avant même avec Thibaud ça doit être pareil, on restait assis enfin on restait posés, là je marchais un peu plus quand j'écrivais*

- Dissociation des mouvements des yeux et de la tête :

*S2 : Non, non, j'arrive bien à marcher comme ça en fait. Mais là, on voit pas mais mes yeux ils bougent, je lève pas la tête, même si je lève la tête, je lève le pda en même temps*

Certains sujets, afin de ne pas perdre le focus du PDA et de le garder en permanence dans le champ de vision, évitent de trop bouger la tête et ne procèdent qu'à des balayages oculaires rapides pour gérer leur déplacement.

Comme nous l'avons remarqué précédemment, la facilité de déplacement va en partie dépendre de l'environnement :

*S1 : ouais, c'est les escaliers, j'ai pas trop de chances de...( tomber) si le chemin est facile à suivre, je peux faire un peu des deux, pas besoin de regarder devant moi*

Lorsque le parcours est facile à évaluer ou très directif, le sujet 1 déclare que cela lui permet de réaliser les deux tâches d'interaction et de déplacement en parallèle, car la vision est peu sollicitée.

- Repérage de l'espace et anticipation (S3)

Parmi les stratégies observées lors des expérimentations pour gérer le déplacement, certaines ont fait l'objet de verbalisations des sujets :

*Expé : tu arrives à avancer et à interagir avec le pda en même temps?*

*S3 : Oui (rires). Si tu veux au départ j'avais déjà vu la distance que j'allais faire donc, j'ai mémorisé et après j'avançais facilement quoi.*

*J'ai du faire un balayage rapide pour voir que j'allais avancer tout droit, tourner, donc après c'était plus facile*

Le sujet réalise un « balayage rapide » de l'environnement afin de repérer et mémoriser son passage, d'en évaluer approximativement les distances et les points clés et ainsi anticiper l'allègement nécessaire de ses ressources visuelles pour une future interaction. Les sujets anticipant l'impossibilité de traitement parallèle de la planification d'un déplacement et de l'interaction, décalent temporellement ces deux tâches et n'auront plus qu'à gérer les ajustements tactiques en situation.

- Perception des possibilités de mobilité de l'application :

La plupart des sujets ont précisé qu'ils percevaient le potentiel mobile de l'application même si certains déclarent ne pas l'avoir exploité.

*S1 : Je vois que j'étais vachement figé en fait, je pense que je bougeais pas assez dans l'expo, j'aurai pu et le pda le permettait justement. J'ai pas cherché à bouger plus quoi. Je sais pas pourquoi, j'ai eu besoin de rester figé, alors que techniquement, sûrement, j'aurai pu plus bouger et facilement, en écoutant peut-être la voix, des choses comme ça. Mais euh, je sais pas, c'est pas que j'y arrivais pas, j'ai pas eu du tout, je me sentais pas de bouger, je restais figé.*

Ce sujet insiste en remarquant que malgré un dispositif technique permettant des interactions en mobilité réelle, en l'occurrence ici, la portabilité du PDA et l'usage de la synthèse vocale, celui-ci n'a pas réussi à en tirer profit, et le regrette. Une des explications peut être que ce sujet a très peu utilisé la modalité vocale et la synthèse qui y est associée. Cet usage statique semble avoir été perçu comme un frein à son efficacité dans le bon déroulement de la tâche d'exploration de l'environnement.

Ce résultat permet de noter que la portabilité du dispositif est une condition nécessaire mais pas forcément suffisante pour permettre la mobilité si par ailleurs, le style d'interaction proposé aux utilisateurs les ancre dans une position figée.

## **8.2. Remarques liées à la mobilité et aux modalités**

Dans cette partie, nous nous focalisons sur certains arguments mentionnés par les sujets concernant la possibilité d'interagir en mobilité en relation aux modalités d'interactions utilisées.

- Délégation de tâche à la machine via la modalité vocale :

Plusieurs sujets mentionnent une relation particulière à la machine lors de l'usage de la modalité vocale :

*S3 : "(« corps » en vocal) Je commençais à m'habituer de plus en plus au côté vocal, donc c'est vrai que ça me dérangeait pas plus. Je cherchais en même temps mes réponses donc là je préfère le faire automatiquement*

*S3 : Je devais lire et en même temps chercher à répondre*

*Expé : donc, là le vocal t'as libéré un peu?*

*S3 : Oui, ben généralement si, parcequ'au moins, le vocal il fait tout*

Dans ces deux exemples, le sujet 3 mentionne le fait qu'il avait des objectifs concurrents à réaliser en parallèle (lire les questions, chercher les réponses dans l'environnement) et ressent donc une forte charge cognitive. Pour ce sujet, un moyen de libérer des ressources à la fois cognitives, attentionnelles et motrices consiste à déléguer une partie de la tâche au système en utilisant des commandes vocales. Il perçoit ce mode d'interaction comme un mode automatique qui fait faire l'action au système (« je préfère le faire automatiquement »), car la mise en œuvre physique de la commande vocale est beaucoup moins demandeuse que la modalité tactile qui mobilise le regard et la deuxième main pour aller cliquer sur les boutons.

*S3 : Mais, si je suis où ya pas de bruit, la reco, c'est vrai ce serait très bien, si je suis en train de marcher, c'est très bien, si j'ai pas le temps, ouvrir, fermer, aller à msg, tu demandes, ça te fais tout, te dit tout,*

Dans ce troisième extrait, le sujet mentionne trois exemples de raisons contextuelles le poussant à passer en mode « automatique » :

- ✓ contexte sonore (« pas de bruit »)
- ✓ besoin mobilité (« en train de marcher »)
- ✓ pression temporelle, charge de travail (« si j'ai pas le temps »)

Le sujet précise à nouveau que ce mode d'interaction est perçu comme une délégation de tâche où le système « fait tout », que ce soit en entrée (vocale) ou en sortie (synthèse), comme si la perception du coût de la commande vocale était négligeable (au moins comparativement au coût d'une commande tactile).

- Sensation de liberté liée à la modalité vocale :

De manière générale, les sujets qui ont utilisé la modalité vocale ont souvent fait mention d'une sorte de sentiment de liberté lié à son usage :

*S6 : Oui, si, quand ça passait (la reco vocale), c'est toujours pareil, c'est vachement bien, tu regardes pas le truc, tu fais les manips sans regarder le... t'es vachement libre, tu peux regarder l'expo, j'ai quand même pu toucher le fœtus, tu verras, avec la caméra, je regarde d'autres trucs etc..quoi*

Ici encore les arguments de libération du regard et des membres participent à renforcer le sentiment d'un faible coût d'interaction rendant possible le parallélisme de tâche et procurant donc une sensation de liberté.

- Apport de la synthèse vocale :

Comme mentionné à plusieurs reprises, la synthèse vocale participe de ce sentiment de légèreté du coût d'interaction :

*S4 : Et puis la synthèse vocale, vu qu'on fait plusieurs tâches en même temps, envoyer, recevoir, enregistrer des sons, la synthèse vocale c'est vrai qu'elle est un peu chiant des fois, mais c'est utile dans le sens où quand on se ballade d'un endroit à un autre pendant quelques dizaines de mètres, on fait rien sur le truc, le fait qu'elle parle ça permet de faire autre chose, c'est clair, au lieu d'avoir les yeux rivés sur l'écran*

Après avoir noté quelques aspects négatifs de la synthèse vocale (en partie liés à des mises en routes intempestives et non commandées), le sujet rappelle que lors d'un déplacement où l'interaction n'est pas possible, la synthèse vocale permet tout de même d'être actif, via la libération du regard, pendant un temps de déplacement considéré comme perdu. Cette modalité de sortie permet de profiter des interstices durant lesquels l'interaction n'était pas possible auparavant.

*Expé: tu as eu le sentiment d'être focalisé sur le pda?*

*S6 : Ah non, on le verra, mais pendant qu'elle me parle, vu que j'ai l'oreillette, je peux regarder autre chose, écouter d'une oreille, c'est ça qui est bien aussi, vraiment, c'est comme ça que j'ai pu savoir où étaient les trucs, j'ai pu regarder l'expo et en même temps avancer le quiz*

Ici le sujet confirme l'exemple précédent en précisant l'intérêt de l'oreillette permettant une activité multitâche en parallèle.

Dans ces deux exemples mentionnés par les sujets, la sortie audio de la synthèse concerne plutôt la lecture de médias textuels que les feedbacks de commandes.

## 9. Synthèse des résultats : Multimodalité, mobilité et ressources attentionnelles

### 9.1. Mobilité et interaction : Effets des modalités

Les résultats montrent qu'à un niveau global, 62,8% de la durée des sessions se déroulent en situation statique et 29,8% en micromobilité et que globalement, les commandes se répartissent à 73% en statique et 24,71% en micromobilité.

Les données disponibles indiquent comme on pouvait s'y attendre que la modalité qui requière l'utilisation du stylet (Tactile Stylet) et conséquemment un niveau de contrôle visuel précis de l'action représente plus de 84% des commandes statiques alors que la modalité vocale est utilisée représente 45% des commandes en micromobilité. La modalité tactile stylet permet tout de même un bon degré de micromobilité puisqu'elle atteint plus de 55% des commandes en micromobilité.

Il est donc clair que la modalité vocale favorise, ou du **moins permet un meilleur degré de mobilité**.

Les résultats indiquent également une forte variabilité inter-individuelle six sujets pour le comportement lié à la mobilité. Pour tous les sujets sauf le sujet 6 et le sujet 3, la grande majorité des commandes a été faite en situation statique (les résultats vont de 4% à 22% environ d'interactions mobile pour ces sujets). Ces différences très importantes montrent bien que les sujets n'appréhendent absolument pas de la même manière l'interaction en mobilité et en particulier la possibilité d'interagir en mobilité « réelle ». Cependant, il faut noter que même pour des sujets qui font peu d'interactions en mobilité « réelle », la modalité vocale est celle qui la favorise le plus.

En ce qui concerne l'évolution de la mobilité au cours des sessions, là encore les résultats sont mitigés et seuls les sujets 3 et 6 semblent montrer une progression importante et nette mais contrairement à ce que l'on aurait pu croire, elle va dans le sens d'une diminution de l'interaction en mobilité « réelle » alors que ces sujets ont des taux d'interactions vocales en mobilité très élevés, de plus le faible nombre de sessions ne permet pas de statuer sur l'apparition de stabilisation des usages.

Les résultats au niveau d'une commande sont cohérents avec les résultats précédents puisqu'il semble y avoir un lien fort entre l'ordre des commandes les plus vocales et celles les plus mobiles.

### 9.2. Regard et multimodalité

Les résultats indiquent une répartition d'environ un tiers du temps des sessions avec le regard orienté vers l'environnement et deux tiers orientés vers l'interface. Sur un total de 4095 commandes multimodales, seules 137 (3,35%) ont été effectuées le regard dirigé vers l'environnement et 3958 (96,65%) le regard dirigé vers l'interface.

Parmi les effectifs de la modalité vocale, les résultats montrent que 14,7% des commandes ont été faites avec le regard orienté vers l'environnement. Ceci montre que la modalité vocale, dans une faible mesure, permet la libération de l'attention visuelle qui est intrinsèquement impossible avec la modalité tactile et que par ailleurs les sujets n'utilisent pas systématiquement la potentialité de libération du regard de la modalité vocale.

On note également une forte variabilité inter individuelle. Il est intéressant de noter que pour les sujets 3 et 6, il y a une forte évolution de l'orientation du regard au cours des sessions (de 0% à 18,97% pour le sujet 3 et de 16,53% à 58,4% pour le sujet 6). Ces résultats semblent indiquer une bonne appropriation de l'application et en particulier de la modalité vocale permettant une amélioration de la libération de l'attention visuelle au fur et à mesure des sessions.

Nous avons également noté une forte tendance des utilisateurs à porter le PDA en position micro pour effectuer une commande vocale (79,62% des commandes vocales). De plus, cette position (micro ou basse) du PDA au moment de la commande n'a aucune influence sur le taux d'erreur de reconnaissance vocale.

Les modes Correction ou Premier ne présentent pas de différence significative en terme de mobilité. Il semble même que les corrections en mode vocal (qui sont la plupart du temps des répétitions de la même commande) libèrent un peu plus l'attention visuelle. Il semble donc que la modalité vocale se prête à une correction moins coûteuse au niveau des ressources attentionnelles contrairement aux résultats répandus indiquant un coût cognitif élevé des corrections.

### 9.3. *Regard et mobilité (simultanéité)*

Deux résultats à retenir concernant la durée des orientations du regard :

- Lorsque les sujets regardent l'interface, ils sont largement statiques, cependant, ils parviennent tout de même à être en micromobilité 20% du temps.
- Lorsque les sujets regardent l'environnement, ils restent statiques tout de même 40% du temps et mobiles autour de 45%.

Il faut également retenir que plus de 84% des interactions orientées vers l'environnement ont été faites en mobilité et que ce taux tombe à 23% lorsque le regard est orienté vers l'interface. Il est donc assez clair que l'orientation du regard vers l'interface « ancre » les sujets en situation statique.

Au niveau qualitatif, nous avons obtenu un assez grand nombre de verbalisations concernant la perception de l'impact des usages des modalités sur l'attention visuelle, et conséquemment sur la possibilité de continuer une tâche principale ou parallèle. Un grand nombre de sujets ont pu être abordés que nous avons thématiques autant que possible.

- **Partage de l'attention et attention floue** (focus ailleurs que sur le lieu de l'interaction) : Il existe une compétition quant à l'allocation des ressources attentionnelles et les sujets doivent procéder à un choix de compromis de leur allocation, au risque d'imprécision et de baisse de performance dans l'une ou l'autre des tâches.
- **Effet des modalités sur l'attention visuelle**
  - **Effet de la modalité vocale** : Impression de relâchement de l'attention visuelle que procure la modalité vocale qui permet de continuer une autre tâche.
  - **Interaction non visuelle** : La connaissance à l'avance du comportement de l'application, ainsi que la sortie sonore et la synthèse vocale permettraient d'envisager une interaction absolument non visuelle si l'application n'était pas limitée techniquement (si l'application avait été développée selon le principe d'équivalence fonctionnelle maximale).
  - **Effet de la modalité tactile stylet** : Relativement à la modalité vocale, l'usage du stylet est beaucoup plus coûteux en ressources visuelles.
- **Facteurs influençant la charge attentionnelle** :
  - Appropriation de l'application
  - Enchaînement d'activités visuelles : Celles-ci poussent à garder le regard dans la position courante.
- **Effet du design sur l'attention visuelle**
  - Taille et proximité des icônes
  - Emplacement des icônes
- **Effet des sorties sonores et visuelles sur l'attention visuelle**
  - Usage et effet de la Synthèse vocale sur l'attention visuelle
  - Feedback audio (sonore et synthèse)
  - Persistance du besoin de confirmation visuelle
  - Simplicité de l'application permettant une interaction peu visuelle

### 9.4. *Activité réelle et mobilité : Effet de l'activité*

Les activités (réelles) sont toutes majoritairement statiques, avec cependant une base mobile relativement importante.

Dans le contexte des activités de notre expérimentation, il y a une certaine cohérence dans le degré de mobilité entre la durée passée en mobilité et le nombre de commandes passées pour un type d'activité. De plus, c'est bien le type d'activités qui détermine le degré de mobilité (par exemple la réponse aux

questions personnelles (catégorie QP) qui demande une exploration de l'environnement est l'activité la plus mobile).

### 9.5. *Transitions de mobilité*

D'une manière générale, les états de mobilité montrent une grande continuité d'une commande à l'autre. Il semble donc, au vu de ces résultats, que le besoin d'interaction ne modifie que dans des proportions faibles les états de mobilités et que les sujets s'adaptent autant que possible au contexte de mobilité en cours.

L'état « statique » est celui qui a donc le plus tendance à se renouveler en cas de commandes successives. Autrement dit, lorsque les sujets ont fait une commande à l'arrêt, les commandes suivantes étaient également massivement faites à l'arrêt. Ce phénomène est tout de même moins marqué pour les autres états de mobilité précédentes tels que « micromobilité » ou « déplacement ».

Les résultats indiquent une forte propension à l'arrêt de la mobilité lié à une interaction.

Le mode « micromobilité » présente des caractéristiques intermédiaires aux deux autres modes avec une interruptibilité ou une continuité plus souple ou du moins plus utilisée par les sujets. Ce mode semble en effet représenter un compromis d'allocation possible des ressources attentionnelles et cognitives à la fois sur la gestion du déplacement et l'interaction. Cette attention flottante permet des transitions plus souples entre les modes de mobilité car les deux tâches sont toujours en cours en parallèle.

Le mode correction semble permettre plus de mobilité que le mode premier (toutes modalités confondues). De plus, les corrections des erreurs en modalité vocale se font à 57,87% en micromobilité, ce qui semble jouer en faveur de la continuité de l'activité de déplacement. Les corrections en modalité stylet, quant à elles, ne permettent que 20,37% de micromobilité.

### 9.6. *Résultats qualitatifs (verbalisations) sur la mobilité et les ressources attentionnelles*

#### 9.6.1. Remarques générales liées à la mobilité et à l'application

- **Objectifs divergents et caractère multitâche de l'activité** (voire concurrents)
  - Exploration de l'exposition et interaction : Nombreuses verbalisations concernant la nature multitâche de la consigne et de l'activité qui y est associée et décrivant les difficultés rencontrées afin de réaliser ces deux objectifs parallèles qui demandent des ressources similaires en simultané.
- **Partage de l'attention et Gestion de l'activité multitâche**
  - **Compromis d'allocation de ressources cognitives et visuelles :**

Certains sujets, à l'inverse, font remarquer qu'ils parviennent quand même à partager leur attention visuelle entre le déplacement et l'interaction à condition que celles-ci ne soient pas trop demandeuses

- **Attention « flottante », « floue » ou focalisée sur l'interface selon l'activité** (attention épisodique) : Lorsque l'interaction est majoritairement tactile, les sujets doivent mettre en œuvre des stratégies pour réussir à se déplacer en même temps lorsque c'est possible
- **Baisse de la performance (cognitive) due au partage de tâches** : La concurrence ou compétition d'allocation des ressources va donc se jouer à plusieurs niveaux dépendants parfois les uns des autres. Plusieurs exemples illustrent aussi bien les sollicitations cognitives que physiques (mobilisation du regard et des membres en général) qui peuvent varier selon les activités en cours.
- **Orientation de la direction du regard, tête basse** : Plusieurs sujets ont évoqué les raisons qui les empêchent de pouvoir interagir en mobilité réelle en dissociant bien la gestion du déplacement du point de vue moteur, et du point de vue des ressources attentionnelles. Il s'agit principalement de la compétition d'allocation de la ressource visuelle qui constitue éventuellement un frein à la mobilité, le déplacement étant peu sollicitant du point de vue strictement moteur.
- **Focalisation sur l'interaction et réduction du champ de vision périphérique**
- **Orientation difficile** : la simultanéité d'une interaction et d'un déplacement est éventuellement possible si la gestion du déplacement mobilise des ressources cognitives de bas niveaux du type réaction à l'environnement plus de l'ordre du tactique. En

revanche, lorsqu'il s'agit de tâches plus stratégiques d'orientation ou de planification de trajet, mobilisant des ressources cognitives de haut niveau (traitement symbolique d'informations), la simultanéité d'une interaction devient difficile voire impossible.

- **Mise en œuvre de stratégies permettant les activités en parallèle :**
  - **Appropriation de l'application**, gestion du temps de réaction du système etc
  - **Dissociation des mouvements** des yeux et de la tête
  - **Repérage de l'espace et anticipation**
  - **Perception des possibilités de mobilité de l'application**

#### 9.6.2. Remarques liées à la mobilité et aux modalités

- **Délégation de tâche à la machine via la modalité vocale :** Pour certains sujets, un moyen de libérer des ressources à la fois cognitives, attentionnelles et motrices consiste à déléguer une partie de la tâche au système en utilisant des commandes vocales. Ils perçoivent alors ce mode d'interaction comme un mode automatique qui fait faire l'action au système, car la mise en œuvre physique de la commande vocale est beaucoup moins demandeuse que la modalité tactile qui mobilise le regard et la deuxième main pour aller cliquer sur les boutons.
- **Sensation de liberté liée à la modalité vocale**
- **Apport de la synthèse vocale comme facilitant la mobilité**

## Chapitre 6. Saisie de médias

## 1. Introduction

Bien que la saisie de média diffère par de nombreux points de l'interaction multimodale (entre autre par le fait que lors d'une interaction, le sujet s'adresse à la machine alors qu'une saisie de média est destinée à une personne), celles-ci ont tout de même plusieurs points communs qu'il est intéressant d'étudier.

En particulier nous nous pencherons sur l'influence du contexte sur le choix de saisie de média, comme nous l'avons fait lors de l'étude de l'usage des modalités.

Pour rappel, en terme de saisie de médias à insérer dans le corps des messages, les sujets avaient plusieurs possibilités relativement équivalentes au niveau fonctionnel (mais évidemment pas au niveau de la mise en œuvre) :

- ***Saisie de texte au clavier virtuel à l'aide du stylet***
- ***Saisie de texte manuscrit sous forme de dessin***
- ***Saisie d'un message audio dicté***

La saisie de texte manuscrit ne sera pas traitée car elle a été très peu utilisée (ceci pour des causes techniques), en revanche, les deux autres possibilités de saisie de médias ont été utilisées de manières très variées selon les sujets et les contextes.

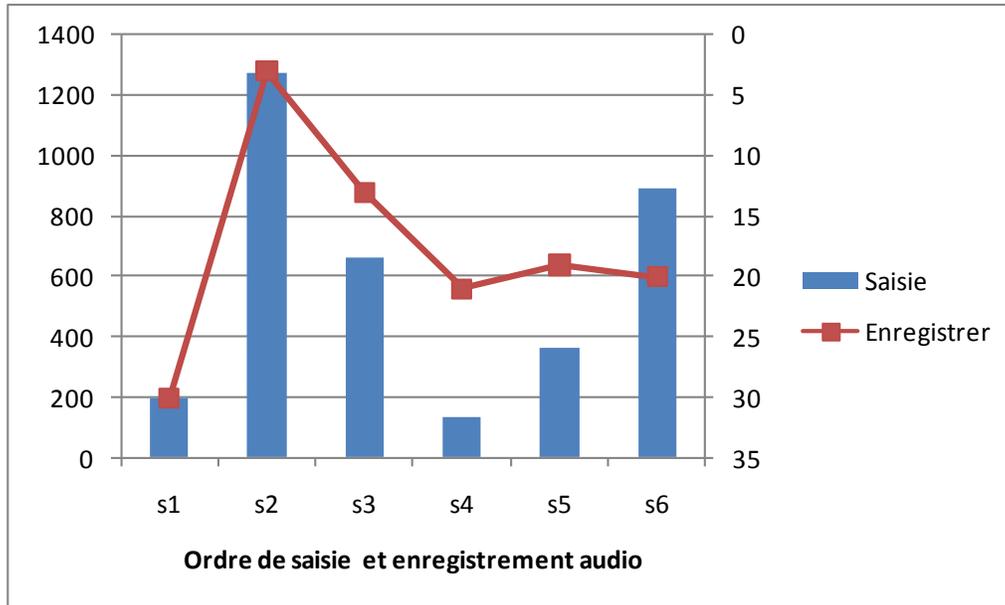
Les résultats sont présentés sous la forme suivante :

- Usages de la saisie de média par sujet
- Usage des médias en fonction du contexte
- Verbalisations concernant la saisie de médias

Une synthèse des résultats est présentée pour clore ce chapitre.

## 2. Usages de la saisie de média par sujets

Nous comparons ici les occurrences respectives de la saisie de texte au clavier et des enregistrements vocaux pour chaque sujet.



**Figure 76 Comparaison de saisie de média (message vocal VS clavier virtuel)**

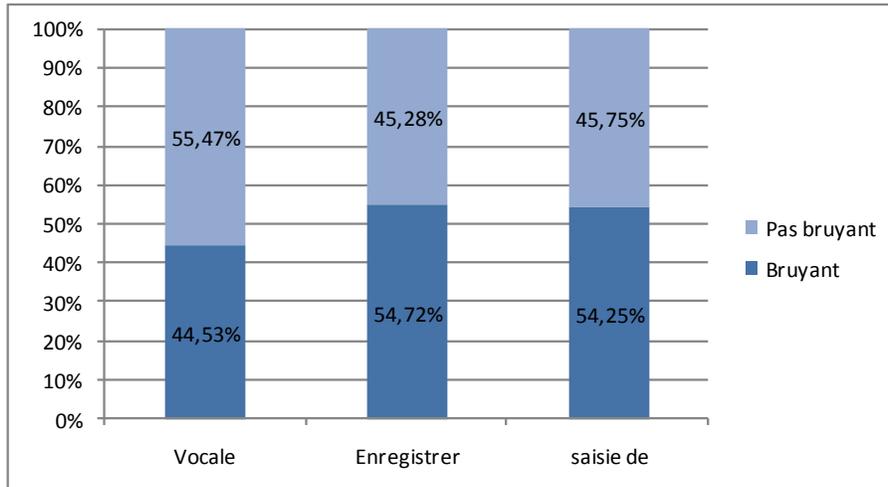
L'usage de saisie de médias présente de grandes variabilités au niveau interindividuel qu'il s'agisse de l'enregistrement de messages vocaux ou de la saisie de texte au clavier virtuel. Pour la saisie de texte les valeurs peuvent varier de moins de 200 caractères saisis pour l'ensemble des sessions (sujet 4) à plus de 1200 pour le sujet 2.

Il en est de même pour la saisie des messages vocaux pour lesquelles les occurrences varient de 3 pour le sujet 2 à 30 pour le sujet 1 pour l'ensemble des 3 sessions.

Si l'on met en relation ces deux modes de saisie de média (Figure 76) les résultats montrent une forte corrélation, avec une forte opposition entre les deux modes. De manière générale, plus les sujets utilisent un mode de saisie de média et moins ils utilisent l'autre (pour le sujet 6 les résultats sont moins marqués mais restent cohérents). Cela témoigne d'une sorte d'exclusivité (ou spécialisation) de média de saisie pour les sujets dont les préférences sont les plus marquées et pour lesquels nous disposons de verbalisations expliquant en partie leurs choix (voir partie 4 Verbalisations concernant la saisie de médias).

### 3. Usage des médias en fonction du contexte

De la même manière que l'usage des modalités, le choix de la saisie de médias pourrait être influencé par des facteurs contextuels comme le niveau sonore ou la présence de tiers, en particulier pour l'enregistrement de messages vocaux qui partagent de nombreuses propriétés avec la modalité vocale.

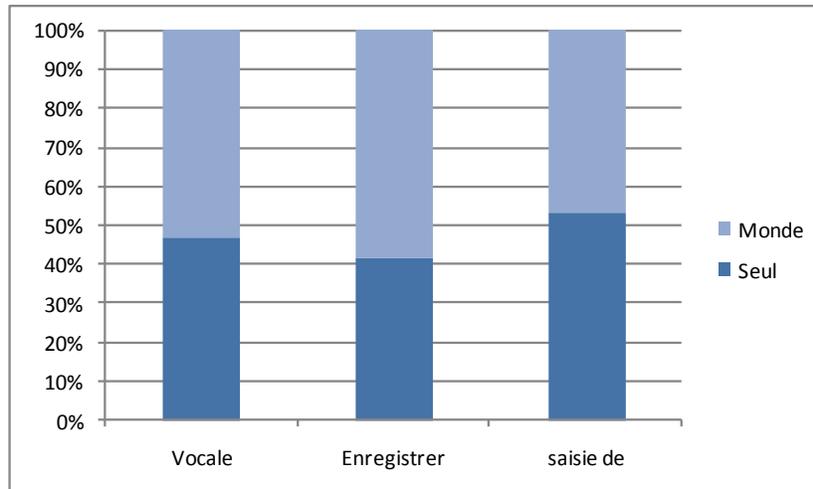


**Figure 77 : Saisie de médias selon le contexte sonore (niveau global)**

En ce qui concerne le contexte sonore et la saisie de médias, la Figure 77 ne montre aucune différence au niveau global entre les ratios de l'enregistrement audio et de la saisie de texte au stylet. Le contexte sonore ne semble donc pas avoir d'influence sur le choix d'un mode de saisie de média. Nous avons également placé la modalité vocale dans la figure afin de comparer ses taux avec l'enregistrement de messages audio. Les résultats montrent une différence de 10% entre les deux avec apparemment une plus grande tolérance au niveau sonore pour l'enregistrement de messages vocaux de la part des utilisateurs. A ce sujet, nous disposons de nombreuses verbalisations venant illustrer et plutôt contredire ce propos (voir partie 4 Verbalisations concernant la saisie de médias).

Nous examinons maintenant de la même manière l'hypothèse d'une influence du contexte social sur le choix de saisie des médias.

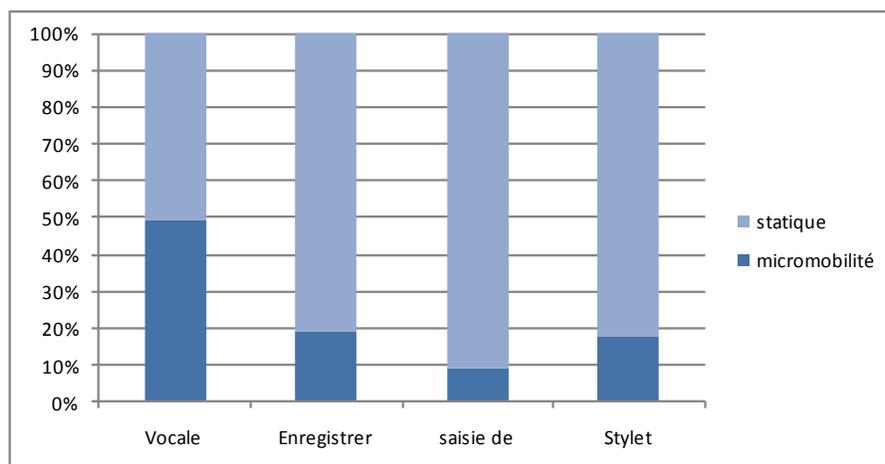
## Saisie de médias



**Figure 78 : Saisie de médias selon le contexte social (niveau global)**

La Figure 78 montre une différence de ratio d'environ 11% entre la saisie au clavier et les messages vocaux, avec étonnamment, un ratio supérieur pour la saisie de texte lorsque les sujets sont seuls. La commande « enregistrer » est majoritairement effectuée en contexte « seul », mais cela n'est pas massif avec un ratio de 40%. Le contexte social ne semble donc pas non plus avoir un effet très important sur le choix d'usage d'un média de saisie.

Chaque mode de saisie de média ayant ses caractéristiques propres quant à la mobilisation des ressources attentionnelles et cognitives, il est intéressant d'analyser leurs effets sur les états de mobilité. Pour cela, et à titre de comparaison, nous réintroduisons les valeurs de mobilité pour les modalités (vocale et stylet).



**Figure 79 : Saisie de médias et mobilité associée (niveau global)**

La Figure 79 indique de s variations dans les degrés de mobilité selon les médias mais également entre médias et modalités. Les différents modes de saisie montrent tout deux des degrés de mobilité très faibles avec plus de 90% de saisie de texte en « statique » et également plus de 80% pour « enregistrer ». La comparaison entre les deux modes montre donc tout de même une différence d'environ 10% de « mobilité » en plus pour « enregistrer ». Le mode de saisie de média oral, tout en restant largement statique, est donc plus mobile que la saisie de texte au stylet. Cependant, il faut noter que la commande de lancement d'enregistrement vocal est une commande uniquement tactile (avec une icône relativement

grande) et donc à tendance plus statique. En revanche, une fois l'enregistrement lancé, nous avons observé des forts niveaux de mobilité mais qui ne sont pas documentés quantitativement.

La comparaison de mobilité entre saisie de média et modalité révèle également des résultats intéressants. En effet, la saisie de messages vocaux, au regard de la modalité vocale (différence de 30%) qui permet un assez bon niveau de commandes mobiles (quasiment 50%), montre un taux de mobilité plutôt faible (moins de 20%) alors qu'elles partagent des caractéristiques proches. Nous verrons par la suite les explications de ces importantes différences.

Dans une moindre mesure le même phénomène est observé pour la comparaison de la saisie de texte au clavier et l'interaction tactile stylet avec une différence inférieure à 10% en faveur d'une meilleure mobilité pour la modalité. Plusieurs hypothèses pourraient en partie expliquer ce résultat, dont le fait que la saisie de texte relève d'une tâche relativement différente d'une interaction au stylet pour au moins un certain nombre de raisons :

- **Séquence de saisie vs commande plus isolée** : la saisie d'un texte relève quasiment toujours d'une séquence de saisie de plusieurs caractères et la mobilisation du regard est donc de plus longue durée favorisant peu la mobilité.
- **Atteinte de la cible** : La visée pour une commande individuelle est très différente de celle d'un caractère sur un clavier virtuel. L'icône de la commande est souvent plus grosse et bien distinguée par rapport à une touche de clavier noyée parmi les autres et dont la recherche est plus couteuse. La tâche d'identification de la cible, même avec une bonne mémorisation des emplacements est tout de même plus compliquée pour la saisie de texte et mobilise donc également la vision de manière plus intense.

Après avoir observé que le contexte sonore ou social a une influence limitée sur le choix des médias, et que ceux-ci montrent des degrés de mobilité différents, nous réintroduisons maintenant le niveau d'activité afin d'en vérifier son impact.

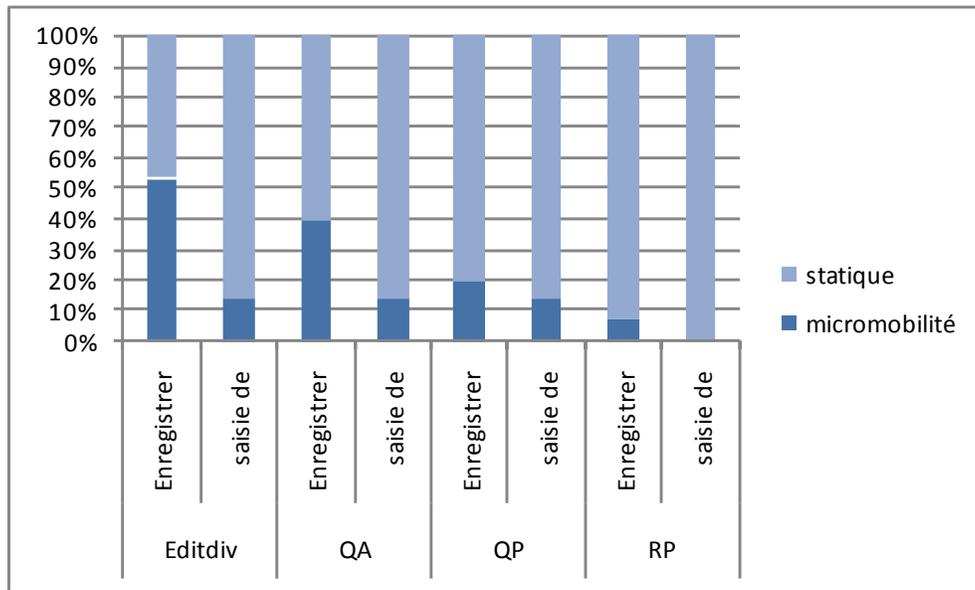


Figure 80 : Degré de mobilité selon l'activité et la saisie de média

Pour le mode de saisie « enregistrer », on observe de fortes variations de mobilité selon le niveau d'activité en cours avec des différences pouvant atteindre 45% (entre l'activité Editdiv et RP). L'activité, comme c'est le cas pour les modalités, semble bien déterminer le niveau de mobilité du mode de saisie audio. Dans le cas des deux activités cités (Editdiv et RP), la demande en mobilité est en effet bien différente puisque pour RP (Réponse Perso), il s'agit de dicter (ou recopier pour le cas du stylet) une réponse obtenue dans l'exposition et requérant donc un état statique relativement important. A l'inverse, l'activité Editdiv (Edition divers) consiste à dicter un message libre improvisé avec un contenu servant

plutôt à des régulations de l'activité collaborative non directement liées à une question précise du Quiz. Dans ce cas, l'attention visuelle n'a aucune contrainte d'orientation et permet une certaine mobilité. Dans les deux autres cas d'activité (QA et QP), l'origine du message audio est contenue dans le PDA puisqu'il s'agit de demander ou d'envoyer des informations relatives aux questions personnelles (QP) ou à celles de l'autre (QA), ce qui demande donc un peu plus d'attention visuelle.

Pour le cas de la saisie de texte au stylet, qui est massivement statique (autour de 90% pour tous les types d'activités sauf RP où le ratio est de 100%), les taux de mobilité ne varient pas selon les activités, sauf pour RP qui est absolument « statique ». Ceci peut s'expliquer, comme nous l'avons mentionné ci-dessus, par le fait que RP, dans le cas de la saisie de texte au clavier, consiste à recopier ou retranscrire une information écrite ou audio contenue dans l'exposition et requiert donc une attention soutenue difficilement compatible avec un déplacement. Au delà des propriétés intrinsèques de la saisie de texte au stylet, qui permettent peu de saisie mobile, l'activité peut donc éventuellement contraindre davantage le niveau de mobilité.

## 4. Verbalisations concernant la saisie de médias

- **Caractéristiques du média audio**

- Robustesse au contexte sonore, Qualité satisfaisante :

De la même manière que pour l'usage des modalités, le contexte sonore peut avoir une influence sur le choix de la saisie de média, mais ici, l'argumentation est tout à fait différente puisqu'il s'agit d'un enregistrement destiné à être écouté et interprété par un tiers et dont l'exigence de qualité sonore du rendu est un critère important :

*Expé : Tu te rends compte quand même que l'environnement sonore... ?*

*S1 : Ouais, j'ai pas réécouté mon message d'ailleurs, mais en général, j'ai trouvé que les messages ils avaient toujours une bonne qualité. Je pense que là il aurait compris ce que je disais. Ya un bon rendu donc, c'est pour ça que j'utilise justement ça parce que dans l'ensemble je trouve que c'est une modalité (message audio) qui a bien marché. Lui ce que je recevais ça marchait, lui ce que je lui envoyais, je pense que ça marchait aussi*

Dans cet exemple, le sujet manifeste sa conscience du niveau sonore de l'environnement et sa prise en compte pour le choix du média audio, mais son expérience positive attestant de la « bonne qualité » des messages lui fait relativiser ce critère qui devient mineur. Il exprime également sa satisfaction globale avec ce mode de saisie de média que ce soit en réception (« lui ce que je recevais, ça marchait ») comme en production (« lui ce que je lui envoyais, je pense que ça marchait aussi ») et justifie ainsi de son usage quasi exclusif de ce mode de saisie (sujet ayant le maximum d'occurrences de saisie audio).

- Présence de tiers :

Il est également intéressant que ce sujet, à l'inverse, est un de ceux qui a eu le moins recours à la modalité vocale. Pour rappel, nous reproduisons ici une de ces verbalisations concernant l'usage de la modalité vocale en présence de tiers :

*S1 : je peux pas parler avec mon pda devant les gens*

*Expé : plus discret*

*S1 : Ouais, je sais que c'est pas euh pas un truc que j'aimerais faire. Et je trouve que justement j'ai l'impression, pas passer pour bête mais c'est vrai que le rapport qu'on a avec le pda devant les gens c'est ... (« aller à »).. ça fait assez, neuneu je trouve, enfin ça me plait pas trop*

Concernant la modalité vocale, le sujet avait justifié de son non usage principalement par la gêne qu'il peut ressentir quant à la présence de tiers, témoins de son interaction vocale qui le rendent mal à l'aise. Il précisait également que cela était en partie du au style de relation et au vocabulaire utilisé pour l'interaction. Effectivement, cet argument semble être renforcé par le fait qu'à l'inverse il ne semble pas gêné d'enregistrer des messages vocaux devant des tiers. Pour la plupart des sujets, même s'il y a souvent une certaine confusion dans les verbalisations et le vocabulaire utilisé pour parler du média vocal et de la modalité vocale, il y tout de même une perception et une dissociation bien nette entre les deux. Dans l'exemple mentionné par le sujet 1, dicter un long message audio en langage naturel devant des tiers semble beaucoup moins l'incommoder que de commander son interface avec des mots clés « simplistes » (propos par ailleurs en parfaite cohérence avec ses résultats quantitatifs).

A l'inverse, d'autres sujets manifestent une gêne identique à celle ressentie pour les commandes vocales, pour dicter un message audio devant des tiers :

*S2 : Et puis même, alors en fait, là, je voulais enregistrer un message vocal, parce que j'ai vu que le menu il est long, je pouvais pas le réécrire, j'allais pas avoir le temps. Donc, j'ai attendu que les gamins y partent, parce que c'est vrai que c'est un peu gênant*

*Expé : pour eux ou pour toi?*

*S2 : Pour moi plutôt, je suis un peu timide, vous allez voir, j'attends qu'ils partent et après je commence*

Le sujet 2 précise ici que son intention première était de faire un message vocal en raison de la grande quantité d'informations à saisir et de la pression temporelle, mais la présence de tiers l'oblige à modifier ses intentions et à adopter une stratégie d'attente. Il ajoute à nouveau que sa timidité est un facteur important sur le choix du mode de saisie de média.

*S2 : C'est pas le bruit qui m'a empêché de faire le message audio, c'est par rapport au regard des gens*

Dans cet autre exemple, le sujet hiérarchise les facteurs conduisant à son choix de média de saisie. Il appuie bien le fait que pour lui, c'est la présence de tiers qui est prépondérante par rapport au niveau sonore.

*S2 : Alors là c'est intéressant. Pour la définition de l'ADN, c'était compliqué à écrire. A ce moment là, j'ai voulu utiliser l'enregistrement vocal, mais y avait quelqu'un à côté. Donc je l'ai pas fait. J'ai voulu le faire aussi pour une autre réponse, la machine*

*Expé : tu pensais que ça allait gêner les gens ou te gêner toi?*

*S2 : Les deux moi je pense, les gens pas tellement, mais c'est vrai, c'est gênant quand tout le monde se retourne comme ça. Y avait juste une personne à côté de moi, à gauche, c'est pour ça, je mets un peu de temps parce que je me suis décidé enfin à écrire*

De même, dans ce deuxième exemple, le sujet illustre l'obligation qu'il a ressentie de modifier ses intentions premières en raison de la présence d'une personne, malgré une réponse particulièrement complexe et longue à saisir. Il commence par une modification légère de son mode opératoire en patientant, mais la personne ne partant pas, il prend finalement la décision de changer de mode de saisie et de taper la réponse. Cet exemple illustre bien le poids du contexte social sur le choix du média de saisie sur ce sujet qui est particulièrement timide, alors que l'exigence de la tâche pousse de manière forte à choisir un autre média de saisie.

- Adaptation au contexte sonore et Modification du mode opératoire :

L'exigence de qualité mentionnée précédemment peut aller jusqu'à provoquer des modifications du mode opératoire des sujets en cas de bruit :

*S6 : Quand je lis les réponses, je le rapproche moins et je vois que ça marche. T'as tendance à rapprocher mais de moins en moins au fur et à mesure. Mais tu verras, ya une réponse ou deux que j'ai enregistrées, on entendait des trucs derrière quoi*

Dans cet exemple, le sujet explique que son expérience vient corriger petit à petit son mode opératoire original. En effet, au début des expérimentations, la plupart des sujets avaient tendance à rapprocher le PDA de la bouche pour enregistrer des messages vocaux ou passer des commandes vocales. Ici le sujet précise qu'il dissocie bien son comportement pour ces deux types d'actions, et qu'il a tendance, avec le temps à ne plus tant le rapprocher de la bouche vu les bons résultats du point de vue qualitatif.

*S6 : Là où je l'ai monté aussi (le pda), c'est dans le truc avec toutes les télé, y avait vachement de bruit, je voulais répondre à la question. On entend pas mal, j'ai été assez surpris, je pensais qu'on entendrait un truc dégueu*

Dans cet autre exemple le sujet se rappelle l'influence du bruit sur son mode opératoire en prévention d'une dégradation de la qualité du message audio, mais encore une fois, il se rend compte du bon rendu.

- **Meilleure efficacité des messages vocaux même avec du bruit (S4, S6) :**

Le facteur bruit peut avoir des conséquences sur le mode opératoire des sujets mais n'est pas le critère principal déterminant le choix du média :

*Expé : Et là, l'environnement était assez bruyant? Ça a changé quelque chose ?*

*S4 : Non, par rapport au choix (du média) de la réponse? Non, parce que la réponse était super longue*

En effet, dans cet exemple, l'exigence de la tâche, nécessitant la saisie d'une trop longue réponse à saisir au stylet semble l'emporter sur la contrainte que peut représenter le niveau sonore.

*S6 : Oui mais je me suis rendu compte que le micro il est assez directionnel, et puis là, c'était long...*

Et encore une fois, la présence de bruit est compensée par la qualité de l'enregistrement et la longueur de la réponse à saisir.

- **Perception du niveau sonore (S1, S2, S3, S4)**

Les témoignages concernant la perception subjective du niveau sonore sont très nombreux et illustrent la variabilité de la sensibilité au bruit en fonction des sujets mais également en fonction de la tâche et du contexte.

- Sensibilité au niveau sonore (S4) :

La sensibilité et la prise en compte du contexte sonore sont variables en grande partie selon la tâche que le sujet souhaite accomplir :

*S4 : Pour (piloter) l'application elle-même, pour changer de page comme ça, c'est vrai que je me pose pas la question, mais c'est vrai que pour les réponses notamment, quand j'enregistre ma voix, je vais faire un peu gaffe au niveau sonore, parce que c'est un peu le bordel*

Ici, le sujet témoigne de sa prise en compte fine et finalisée du contexte sonore en fonction du type d'interaction qu'il souhaite réaliser avec le système. Il dissocie très clairement l'interaction (les commandes dirigées vers le système) de la saisie d'un message vocale (destinée à son partenaire de Quiz). En effet, l'exigence de qualité dans les deux cas, n'obéit absolument pas aux mêmes contraintes :

- ✓ Éventuellement des critères de robustesse aux erreurs de la reconnaissance vocale en contexte bruyant qui sont peu pris en compte par l'ensemble des sujets (« pour piloter l'application, je me pose pas la question »). Ici les enjeux ne concernent que la relation fluide du sujet avec le système.
- ✓ Des critères d'exigence de qualité liés au fait que le message est destiné à un tiers qui jugera de la qualité de l'enregistrement et qui surtout devra être capable de le comprendre. Il y a donc une certaine attente sociale dans l'usage des médias en général du fait de leur adresse à un tiers.

- Insensibilité au niveau sonore (S2) :

D'autres sujets évoquent une moindre sensibilité au niveau sonore dans le choix d'enregistrer un message audio ou non :

*S2 : C'est pas le bruit qui m'a empêché de faire le message audio, c'est par rapport au regard des gens*

Dans ce cas, le sujet explique que le bruit n'est pas le critère retenu pour ne pas effectuer un message audio mais plutôt la présence de personnes à proximité, comme il l'avait déjà exprimé dans le cadre de

l'usage de la modalité vocale. Pour ce sujet, la distinction interaction et saisie de média n'est donc pas abordée de la même manière que dans l'exemple précédent, et le critère « environnement social » va impacter l'usage des deux types de tâches. Dans ses déclarations précédentes, ce sujet avait fait part de sa grande sensibilité concernant le contexte social et de l'impossibilité d'effectuer des commandes vocales en présence de tiers.

- Non perception du bruit (S3, S4)

Au-delà même de la prise en compte ou non du niveau sonore, plusieurs sujets ont fait part de leur habitude au contexte sonore allant jusqu'à sa non-perception :

*Expé : et le fait que ce soit bruyant?*

*S3 : Non, tu sais en fait tu es habitué, tu vois plus ou moins pas le monde autour de toi*

*Expé : tu t'en rendais compte de ce brouhaha ?*

*S4 : Non, pas autant non.*

*Expé : ça t'as pas dérangé?*

*S4 : Non, vu que je mettais la bouche près du micro, mais là, c'est vrai qu'il y a un sacré bruit derrière*

Dans ces deux exemples les sujets témoignent d'une perte d'objectivité quant à la perception du niveau sonore du fait de leur immersion dans un univers bruyant et de la focalisation sur la tâche qu'ils sont en train de réaliser. Seul le visionnage a posteriori de la vidéo de la situation leur fait prendre conscience du niveau de bruit important.

- Retrait du milieu bruyant pour effectuer un message vocal (S4)

Nous avons également pu observer des stratégies de retrait du milieu bruyant (ou avec des personnes présentes) pour effectuer un message vocal mais nous ne disposons pas de verbalisations pour illustrer ce propos.

- Appropriation du média audio :

Les appréciations concernant le média audio sont nombreuses et généralement positives. Dans l'exemple suivant, le sujet explique comment il l'a appréhendé :

*S1 : J'sais pas si c'est en rapport avec le premier (message audio que j'ai reçu), mais j'avais peut-être pas pensé dès le début à faire ça. Le fait d'en recevoir un, je me suis peut-être dit qu'il fallait faire ça, enfin, moi j'ai bien aimé le principe de recevoir des messages audio comme ça. C'était le premier message que j'ai eu, c'était audio et je sais pas, j'ai trouvé ça bien. Celui-là il était marrant, le son rendait bien, c'est pour ça que j'ai aussi fait pas mal de message audio. Je trouvais que c'était sympa*

Il explique qu'il n'aurait peut-être pas pensé de lui même à utiliser ce média pour faire des messages, mais que le fait d'en avoir reçu lui en a donné une impression très positive et l'a conduit à l'utiliser de nombreuses fois par la suite.

*S1 : Tout ce qui était logiciel du son, je trouvais ça assez simple à employer, c'était vraiment, allez, deux boutons on envoie et c'est fini, c'est vrai qu'à chaque fois, j'ai rien trouvé de plus rapide, parce que dès que c'est un peu long, euh*

La simplicité d'emploi de ce média ainsi que sa rapidité en font un média privilégié pour ce sujet mais également pour la plupart des autres sujets.

*S1 : Là, j'écrivais en même temps le texte pour lui envoyer les deux, au début j'essayais juste pour voir, pour être sûr qu'il comprenne bien, je sais pas je me disais que peut-être ça allait pas marcher ou le son*

Dans un premier temps, et avant de mieux appréhender ce média, le sujet préfère doubler la saisie avec le texte correspondant pour éviter tout problème de compréhension, pratique qui sera vite abandonnée suite à la vérification du bon fonctionnement de ce mode de saisie.

*S1 : Inconsciemment... je crois que je préfère écrire, même écrire les choses, que les enregistrer vocalement. J'ai peur que vocalement ça parte ou ça se perde, je sais pas pourquoi, j'aurai plus tendance à écrire pour vraiment être, j'ai l'impression que je vais être mieux compris si j'écris les mots texto, plus que vocalement.*

Dans cet autre exemple le sujet pointe le caractère labile du mode audio qui selon lui, manque de matérialité, ce qui pourrait provoquer des pertes ou des incompréhensions et justifie ainsi sa préférence pour la saisie de texte.

*S1 : Dans l'ensemble j'ai trouvé que ça marchait bien, j'ai pas eu de pb, j'ai envoyé mes questions, j'ai eu mes réponses, au stylet c'était rapide. J'écrivais quand je voulais pas, pas trop m'embêter, j'écrivais des petites questions, quand c'était des petites questions je les tapais au stylet, sinon, c'était plus au mode audio*

Il précise ici son mode opératoire selon l'exigence de la tâche et ses préférences, qui sont partagées par la plupart des sujets. La saisie de texte au stylet est réservée à la saisie de petits textes (dont il considère tout de même la mise en œuvre plus simple que la saisie audio), en revanche, pour les textes compliqués ou longs, il utilise le mode audio.

- Mise en œuvre des médias :

La mise en œuvre et l'implémentation de chaque média (et des modalités également) a joué sur leur appropriation et leurs usages :

*S1 : Ça oblige à bien regarder (pour la commande « insérer un son »). Autant pour l'enregistrement à la limite, je suis sûr que je pourrais pas trop regarder et quasiment mettre dessus. Là c'est vrai que les petites icônes des fichiers son et dessin elles sont justes à côté*  
*S1 : Là j'ai mis un son, parce que vraiment, c'était facile. T'as juste à parler. La manipulation elle est simple parce que justement, c'est ce que je disais c'est des gros écrans des gros boutons, on clique et envoyer c'est parti.*

Le sujet distingue ici le contraste des icônes des médias et les modalités en ce qui concerne leur taille et leur disposition. En effet, les icônes d'accès à la saisie de médias (son et dessin) sont relativement petites, proches les unes des autres et placées en bas à gauche de l'interface, ce qui les rendait peu accessibles, comme l'on fait remarquer plusieurs sujets. L'étape d'accès à la saisie de médias (autre que la saisie de texte qui se fait directement dans le corps du message) est donc coûteuse en attention visuelle qui contraste par exemple avec une interface dédiée au messages vocaux dont les icônes sont bien dimensionnés et facile à mettre en œuvre (symbologie claire, grosses icônes, position mise en évidence etc). Le gain lié au relâchement de l'attention visuelle par l'enregistrement vocal est donc partiellement occulté par la phase d'appel de la saisie de média.

*S2 : Je crois que j'ai du envoyer deux trois messages vocaux mais pas plus, quand vraiment les réponses elles étaient trop longues par exemple quand je devais noter le menu type dans une cabine spatiale, c'était vraiment vachement long le menu donc, j'allais pas tout réécrire et parce qu'avec le stylet y avait une assez grande fluidité, je pouvais aller très vite. Ça m'embêtait d'aller dans enregistrer nouveau message, enregistrer, insérer réécouter, je trouvais que c'était une perte de temps alors que le stylet marchait beaucoup mieux*

Dans cet autre exemple, le sujet justifie sa préférence de saisie au stylet par des critères de facilité de mise en œuvre du stylet comparée au nombre d'étapes de la saisie d'un message audio. Cependant il place également le critère de la longueur du contenu comme déterminant.

*S5 : Je sais que j'aurai plus utilisé le vocal (msg audio), si j'avais su y accéder, mais je savais pas y accéder*

Ici encore, pour le sujet 5, la mise en œuvre du message vocal a constitué un frein à son usage jusqu'à ce qu'il redécouvre son mode de fonctionnement (problème d'icônes petites et difficiles à identifier).

*S6 : Le stylet, les inconvénients, c'est quand tu écris, c'est long d'écrire un message avec le stylet, ya que ça comme inconvénient, tout le reste c'est des avantages, c'est rapide, en plus les boutons y sont à côté, comme l'écran d'un pda c'est assez petit, ça va assez vite. C'est juste un mouvement de poignet en fait. Tu bouges même pas le bras, donc ça va super vite*

Le sujet 6 fait le bilan de sa perception du mode stylet pour la saisie de texte en y incluant en partie celle de la modalité tactile. Mise à part pour la saisie de textes longs, il plébiscite ce mode en particulier en raison de la petite taille de l'écran qui rend les mouvements de sélection très minimes.

- Exigence de la tâche incertaine :

Comme nous l'avons vu, un des critères les plus importants pour le choix du média est la taille du contenu à saisir, cependant, il n'est pas toujours évident de l'évaluer à l'avance surtout s'il s'agit d'un message dont le contenu est improvisé :

*S1 : Mais ça me dérange pas de taper avec le clavier, j'ai un palm, ça m'embête pas quoi. C'est plus long, quand je suis lancé, tant pis. Mais bon, je me suis dit après que j'aurai peut-être mieux fait de passer par un message audio encore.*

*S1 : (Fait un msg audio) J'avais tapé un texte trop long avant, je voulais plus taper de texte, donc je suis repassé par les commandes vocales (message audio). En plus dans ma tête, je savais pas trop quoi lui dire, donc j'ai fait ça en direct*

*Expé : ça permet d'improviser ?*

*S1 : Ouais voilà, je pense que, on pense directement sur le truc.*

Cet exemple illustre bien l'incertitude à laquelle doit faire face le sujet quant au choix du média de saisie approprié lorsqu'il s'agit d'improviser le contenu. Suite à une expérience de saisie au stylet trop longue et laborieuse, le sujet va privilégier le mode audio qui est plus permissif à ce niveau.

*Expé : Qu'est ce qui te plait pas dans le message audio?*

*S3 : Parce que je parle trop, j'ai l'impression qu'il faut que je dise beaucoup de choses, alors que là au moins l'écrit je mets les mots qu'il faut, c'est bref et puis voilà. L'audio, j'aurai tendance à déborder de la question en fait. Je vais trop dire pour pas grand chose, donc ça sert à rien, je sais, je suis une femme..(Rires)*

De la même manière, le sujet 3 explique qu'il se contraint à utiliser le mode écrit par crainte de dériver dans le propos, du fait de la grande souplesse qu'offre le mode audio. Il va donc privilégier le mode écrit quand les réponses le permettent (réponses courtes) et faire des messages audio uniquement pour les réponses longues.

- Personnalisation ou neutralité du message :

Les modes, audio ou vocal, ont des propriétés différentes quant à la possibilité de richesse et de subtilité du contenu qu'ils peuvent potentiellement véhiculer.

*S3 : Je voulais lui parler directement, je savais pas trop ce que j'allais dire, donc je vais lui envoyer un son, c'est plus personnel, ça n'a rien à voir avec les réponses à taper et tout*

Pour le sujet 3 qui improvise également le message qu'il va envoyer, l'audio semble être un moyen beaucoup plus riche permettant de personnaliser le contenu (prosodie, intonation etc.) comparativement à la rigueur que propose la saisie au clavier.

*S6 : Je voulais pas dire par la voix "t'as pas répondu à toutes les questions, parce qu'après ça fait un téléphone, un pda c'est pas un téléphone. Elle m'a envoyé une phrase sonore, une ou deux phrases, comme c'était pas très long, je lui ai répondu par écrit. Et puis c'était pour lui dire qu'elle avait pas répondu à toute la question, je voulais pas que ça fasse genre*

*Expé : tu croyais que le ton en vocal ..*

*S6 : Aurait pu prêter à confusion genre "il est gonflé, il se prend pour qui"*

Le sujet 6, ici joue de la neutralité du mode écrit pour faire passer un message « sensible », délicat à faire passer avec un message audio et qui aurait pu être mal interprété si le ton avait été trop ambigu.

- Contrôle de l'enregistrement :

Lors du lancement de l'enregistrement, le vumètre indique le niveau de l'enregistrement et constitue un feedback qui semble important pour les sujets.

*Expé : tu ne réécoutes pas le msg?*

*S4 : Non je savais pas que c'était possible, si, si on l'avait fait. Mais quand je vois la courbe qui se met à bouger quand je parle, pour moi c'est bon*

En effet, le feedback d'enregistrement en cours tel que l'enfoncement du bouton d'enregistrement ne permet qu'un contrôle en début d'enregistrement audio, mais ne garanti pas le bon fonctionnement sur une durée plus longue. L'aspect dynamique et temps réel du vumètre constitue en cela une solution appréciable qui permet de se passer d'une réécoute du message enregistré.

- Ressources attentionnelles :

Comme pour la modalité vocale, nous disposons de nombreuses verbalisations au sujet de la faible mobilisation des ressources attentionnelles liée à l'usage de l'enregistrement audio.

*Expé : tu fais ton message (audio) et tu continues à avancer?*

*S5 : Ben ouais, ça gêne pas, au contraire, c'est pratique à utiliser, c'est pour ça qu'on marche en même temps*

*S6 : Quand je parle déjà, je regarde pas mon pda, je regarde devant, même si je le rapproche pas, je regarde autour toujours, je regarde mon environnement en fait, le vocal (en fait le média vocal) c'est ce qu'il y a de plus confortable*

Pour les deux sujets, le mode audio est celui qui est le plus pratique et confortable à utiliser, en particulier en raison de son faible coût en terme de ressources visuelles, qui permet de réaliser d'autres tâches en parallèle.

## 5. Synthèse des résultats sur la saisie de médias

L'usage de saisie de médias présente de grandes variabilités au niveau interindividuel qu'il s'agisse de l'enregistrement de messages vocaux ou de la saisie de texte au clavier virtuel.

De manière générale, plus les sujets utilisent un mode de saisie de média et moins ils utilisent l'autre. Cela témoigne d'une sorte d'exclusivité (ou spécialisation) de média de saisie pour les sujets dont les préférences sont les plus marquées.

Au niveau quantitatif global, le contexte sonore ne semble pas avoir d'influence sur le choix d'un mode de saisie de média.

Le contexte social ne semble pas non plus avoir un effet très important sur le choix d'usage d'un média de saisie.

Le mode de saisie de média oral, tout en restant largement statique, est plus mobile que la saisie de texte au stylet. Cependant, il faut noter que la commande de lancement d'enregistrement vocal est une commande uniquement tactile (même si elle dispose d'une icône relativement grande) et donc à tendance plus statique. En revanche, une fois l'enregistrement lancé, nous avons observé des forts niveaux de mobilité mais qui ne sont pas documentés quantitativement.

L'activité, comme c'est le cas pour les modalités, semble bien déterminer le niveau de mobilité du mode de saisie audio.

Pour le cas de la saisie de texte au stylet, qui est massivement statique (autour de 90% pour tous les types d'activités sauf RP (Réponse Perso) où le ratio est de 100%), les taux de mobilité ne varient pas selon les activités.

Verbalisations sur la saisie de médias :

- **Caractéristiques du média audio**
  - Robustesse au contexte sonore, Qualité satisfaisante
  - Sensibilité à la présence de tiers
  - Stratégies possibles d'adaptation au contexte sonore et Modification du mode opératoire
- **Meilleure efficacité des messages vocaux même avec du bruit**
- **Perception du niveau sonore :**
  - **Sensibilité au niveau sonore :** La sensibilité au contexte pour les messages audio, s'exprimant parfois via l'exigence de qualité, n'obéit absolument pas aux mêmes contraintes que pour la reconnaissance vocale :
    - Les critères de robustesse aux erreurs de la reconnaissance vocale en contexte bruyant sont peu pris en compte par l'ensemble des sujets (« pour piloter l'application, je me pose pas la question »). Ici les enjeux ne concernent que la relation fluide du sujet avec le système.
    - Les critères d'exigence de qualité liés au fait que le message est destiné à un tiers qui jugera de la qualité de l'enregistrement et qui surtout devra être capable de le comprendre. Il y a donc une certaine attente sociale dans l'usage des médias en général du fait de leur adresse à un tiers.
  - **Insensibilité au niveau sonore :** parfois, perception du bruit mais sans effet sur le choix des modalités.
  - **Non perception du bruit :** Au-delà même de la prise en compte ou non du niveau sonore, plusieurs sujets ont fait part de leur habitude au contexte sonore allant jusqu'à sa non-perception
  - **Stratégies :** Retrait du milieu bruyant pour effectuer un message vocal. Nous avons également pu observer des stratégies de retrait du milieu bruyant (ou avec des personnes présentes) pour effectuer un message vocal.
- **Appropriation ou rejet du média audio :**
  - **Usage « contaminant » :** Certains sujets n'auraient peut-être pas utilisé spontanément ce média pour faire des messages, mais le fait d'en avoir reçu leur en

ayant donné une impression très positive a pu les conduire à l'utiliser de nombreuses fois par la suite.

- **Mise en œuvre des médias** : La simplicité d'emploi de ce média ainsi que sa rapidité en font un média privilégié pour la plupart des sujets.
  - **Caractère labile du mode audio** : le manque de matérialité, de rémanence pourrait provoquer des pertes ou des incompréhensions et justifie pour certains la préférence pour la saisie de texte.
  - **Mode opératoire** : Choix du média selon l'exigence de la tâche et les préférences, qui sont partagées par la plupart des sujets. La saisie de texte au stylet est réservée à la saisie de petits textes (dont la mise en œuvre est parfois considérée comme plus simple que la saisie audio), en revanche, pour les textes compliqués ou longs, le mode audio est privilégié.
- **Exigence de la tâche incertaine** : Un des critères les plus importants pour le choix du média est la taille du contenu à saisir, cependant, il n'est pas toujours évident de l'évaluer à l'avance surtout s'il s'agit d'un message dont le contenu est improvisé : il existe en effet, une incertitude à laquelle doit faire face le sujet quant au choix du média de saisie approprié lorsqu'il s'agit d'improviser le contenu.
    - Le mode audio est plus permissif à ce niveau.
    - Certains sujets se sont contraints à utiliser le mode écrit par crainte de dériver dans le propos, du fait de la grande souplesse qu'offre le mode audio. Pour ceux-là, le mode écrit est donc privilégié quand les réponses le permettent (réponses courtes) et les messages audio ne sont utilisés que pour les réponses longues.
  - **Personnalisation ou neutralité du message** : Les modes, audio ou vocal, ont des propriétés différentes quant à la possibilité de richesse, de personnalisation et de subtilité du contenu qu'ils peuvent potentiellement véhiculer. Selon le degré de neutralité ou de personnalisation, un mode ou l'autre peut-être privilégié.
  - **Contrôle de l'enregistrement** : Lors du lancement de l'enregistrement, le vumètre indique le niveau de l'enregistrement et constitue un feedback qui semble important pour les sujets et constitue en cela une solution appréciable qui permet de se passer d'une réécoute du message enregistré.
  - **Ressources attentionnelles** : Pour plusieurs sujets, le mode audio est celui qui est le plus pratique et confortable à utiliser, en particulier en raison de son faible coût en terme de ressources visuelles, qui permet de réaliser d'autres tâches en parallèle.

**Chapitre 7. Synthèse, discussion et apports  
de la thèse**



## 1. Introduction

Dans cette partie, nous nous proposons à travers différents modes, de présenter une discussions des résultats obtenus au cours de ce travail, qu'ils soient quantitatifs ou qualitatifs, afin d'en extraire un ensemble d'apports.

Nous commençons par présenter une étude de cas, illustrant une séquence d'actions liées au passage d'une commande vocale. Nous proposons une analyse très détaillée des évènements se produisant tant du point de vue du système que de l'utilisateur, ceci à l'aide de la grille de lecture que constitue le cadre d'analyse de l'interaction retenu proposé par (Kirsh, 1997) et introduit dans la partie (2.3 Cadre général d'interaction : le modèle de Norman et Kirsh, p68).

Cette analyse fine croisée avec une partie des résultats quantitatifs et qualitatifs obtenus à travers l'analyse des verbalisations nous a permis de dégager une synthèse des propriétés effectives et subjectives des propriétés des modalités et des médias mis en œuvre lors de notre étude. Celle-ci est présentée sous la forme d'un tableau listant les différents médias et modalités étudiés au regard de leurs propriétés.

Dans la continuité de ces résultats, nous présentons ensuite une proposition de modèle détaillé permettant une analyse fine de l'interaction pour la conception ou l'évaluation d'interface multimodales mobiles.

De même, suite à l'analyse des observations sur la mobilité que nous avons réalisée, nous dégageons une proposition de cadre conceptuel pour une mobilité finalisée, permettant une identification des finalités de la mobilité et de l'interaction, afin de supporter une conception d'interface mobile appropriée.

Enfin, nous terminons ce chapitre par une première liste de critères et recommandations pour la conception et l'évaluation d'interfaces multimodales mobiles.

## 2. Etude de cas et illustrations

Grâce aux études de cas et aux illustrations, il est possible d'analyser et de décrire finement les mécanismes mis en jeu lors d'une interaction avec un dispositif portable. Nous avons distingué d'une part les mécanismes d'ordre physique (mobilisation des membres, du regard etc.) et d'autre part les mécanismes d'ordre cognitif (activité cognitive, attente de feedback, etc.).

Nous avons également tenté de distinguer les grandes phases de l'interaction, en distinguant les événements liés au système de ceux liés aux actions de l'utilisateur, ceci en nous basant autant que possible, sur le cadre d'analyse général retenu pour notre étude (modèle de Norman et Kirsh).

Cette analyse est réalisée à l'aide d'un tableau comprenant une première cellule contenant le titre de la séquence, puis d'une cellule comprenant un synopsis court. Le tableau est ensuite divisé en trois colonnes : la première intitulée « Phase », nomme la phase de la séquence étudiée, la seconde, nommée « illustration » contient une image illustrant la phase correspondante, enfin, la troisième colonne contient une description textuelle des mécanismes physiques et cognitifs.

Nous avons illustré ici un exemple de commande vocale pendant laquelle le sujet oriente le regard vers l'environnement car il s'agit d'un cas riche à décrire.

La comparaison de ces séquences permet de dégager des invariants propres aux interactions multimodales, ainsi que de dégager des propriétés des différentes modalités (vocale et tactile) qui sont synthétisées au paragraphe 3 (Synthèse des propriétés effectives et subjectives des modalités p229).

Le Tableau 29 illustre le passage d'une commande vocale avec le regard temporairement orienté vers l'interface puis l'environnement. Une analyse fine des phases de l'interaction permet de repérer des points intéressants pour la régulation d'une interaction de moins en moins visuelle.

S6i3 (11'45'') Commande vocale avec regard vers l'environnement		
<b>Synopsis</b> : Le sujet arrive devant un stand dans lequel est la réponse à une de ses questions. Il doit donc lire le cartel. De plus, il est en train de transférer un message contenant l'indice (dessin) à son partenaire (activité QA).		
Phase	Illustration	Description
0) Pas d'interaction		<b>Sujet</b> : fin d'interaction précédente <b>Engagement</b> : <ul style="list-style-type: none"> <li>• Physique : <ul style="list-style-type: none"> <li>○ Vision environ</li> <li>○ Pouce gauche libre</li> <li>○ Main droite libre</li> </ul> </li> <li>• Cognitif : <ul style="list-style-type: none"> <li>○ Non orienté vers l'interaction</li> <li>○ Exploration</li> </ul> </li> </ul>
1) Préparation / Engagement		<b>Sujet</b> : se prépare à passer une commande vocale <b>Engagement</b> : <ul style="list-style-type: none"> <li>• Physique : <ul style="list-style-type: none"> <li>○ Vision interface</li> <li>○ Pouce gauche appui</li> <li>○ Main droite libre</li> </ul> </li> <li>• Cognitif : <ul style="list-style-type: none"> <li>○ Mémo sémantique</li> <li>○ Vérif mode voc engagé</li> </ul> </li> </ul>
2) Entrée		<b>Sujet</b> : « insérer un dessin » <b>Engagement</b> <ul style="list-style-type: none"> <li>• Physique : <ul style="list-style-type: none"> <li>○ Vision environ</li> <li>○ Pouce gauche appui</li> <li>○ Main droite libre</li> </ul> </li> <li>• Cognitif : <ul style="list-style-type: none"> <li>○ Attente Feedback</li> <li>○ Prise en Compte et FB Effet Commande</li> </ul> </li> </ul>

<p>3) FB PEC Feedback de prise en compte de la commande</p>		<p><b>Système</b> : « clic »</p> <p>Désengagement</p> <ul style="list-style-type: none"> <li>• Physique :             <ul style="list-style-type: none"> <li>○ Vision environ (exploration)</li> <li>○ Pouce gauche relâché</li> <li>○ Mains relâchées</li> </ul> </li> <li>• Cognitif :             <ul style="list-style-type: none"> <li>○ Attente FB ES</li> </ul> </li> </ul>
<p>4) Réalisation et FB EC Feedback d'effet de la commande</p>		<p><b>Système</b> : Synthèse « Insertion d'un dessin » ; fenêtre insertion dessin</p> <p>Préparation</p> <ul style="list-style-type: none"> <li>• Physique :             <ul style="list-style-type: none"> <li>○ Vision interface</li> <li>○ Main gauche position moy</li> </ul> </li> <li>• Cognitif :             <ul style="list-style-type: none"> <li>○ Vérif effet commande</li> <li>○ Préparation commande suivante</li> </ul> </li> </ul>

**Tableau 29 : Illustration d'une commande vocale avec regard temporairement dirigé vers l'environnement**

Analyse du Tableau 29 :

**Synopsis** : Le sujet arrive devant un stand dans lequel se trouve la réponse à une de ses questions. Il doit donc lire le cartel. De plus, il est en train de transférer un message contenant un indice (dessin) à son partenaire.

Il est donc typiquement dans une activité à double tâche concurrente puisque non associées.

**Phase 0 : Fin d'interaction précédente (ou pas d'interaction) (phase Sujet)**

Dans cette phase, le **Sujet** n'est pas dans une interaction ou bien comme ici, il est en cours de finalisation de l'interaction précédente. Il s'agit d'un faible degré d'engagement physique et cognitif orienté vers l'interaction :

**Au niveau Physique :**

- La vision est libérée et orientée vers l'environnement (exploration visuelle à la recherche des réponses aux questions du Quiz)
- Pouce gauche libre (non activation du micro)
- Main droite libre en position basse (stylet en main)
- Main gauche tient le PDA en position très basse (hors champ)
- Au niveau Cognitif :
  - Le sujet est peu orienté vers l'interaction, cependant, il a l'intention d'insérer un indice
  - Exploration et recherche de la réponse qui est finalement repérée sur un stand

**Phase 1 : Préparation /Engagement (phase Sujet)**

Il s'agit ici d'une phase où le **Sujet** se prépare à passer une commande vocale pour insérer un dessin. En parallèle, il se rapproche du stand (aquarium) où est censée être la réponse à sa question.

L'engagement physique et cognitif augmente et commence à s'orienter vers l'interaction mais avec un besoin parallèle de ressources pour l'exploration de l'environnement :

- Engagement Physique :
  - Le regard se porte vers l'interface (et rend plus difficile l'exploration de l'environnement)
  - Le Pouce gauche appui sur le bouton de commande du micro afin de déclencher la commande vocale
  - La main droite reste libre
- Au niveau Cognitif :
  - Le sujet doit se remémorer la commande désirée et sa sémantique
  - Il peut éventuellement vérifier l'activation du mode vocal

Pour cette phase, le regard sur l'interface n'est pas forcément nécessaire.

### **Phase 2 : Entrée (phase Sujet)**

Il s'agit de l'effectuation de la commande proprement dite. Le **Sujet** dit « insérer un dessin » et dans ce cas, place le PDA en position micro, ce qui lui libère l'attention visuelle.

A ce moment l'engagement physique et cognitif peut déjà commencer à se relâcher :

- Engagement Physique :
  - La vision sur l'interface n'est plus nécessaire pour cette phase et peut à nouveau se porter sur l'environnement libérant ainsi des ressources pour l'exploration
  - Le pouce gauche appui sur le bouton de déclenchement du micro
  - La main droite est libre
- Cognitif :
  - Une fois la commande effectuée, la seule tâche cognitive est de vérifier le bon déroulement et la régulation de l'interaction, qui consiste dans ce design, en l'attente d'un feedback de prise en compte de la commande (FB PEC) et d'un feedback d'effet de la commande (FB EC)

Pour cette phase, le regard sur l'interface n'est pas forcément nécessaire étant donné que les feedbacks sont majoritairement audio dans un premier temps.

### **Phase 3 : Feedback de prise en compte de la commande (FB PEC) (phase Système)**

Il s'agit ici d'une phase « Système » de régulation de l'interaction où celui-ci produit un « clic », feedback indiquant la prise en compte d'une commande qui permet au sujet un certain relâchement et désengagement en terme d'interaction :

- Physique :
  - La vision peut rester orientée vers « environ » et éventuellement permettre une exploration visuelle rapide de l'environnement en attendant de devoir à nouveau interagir visuellement. En effet, le FB PEC est implémenté en redondance fonctionnelle via une modalité sonore (« clic ») et une modalité visuelle (enfoncement d'un éventuel bouton)
  - Pouce gauche relâché car la commande vocale est terminée
  - Les deux mains peuvent être relâchées car aucune action physique n'est requise pour cette phase
- Cognitif :
  - Une fois que le FB PEC a été validé et interprété comme une confirmation du bon déroulement de l'interaction, le sujet doit attendre le prochain feedback (FB ES). En effet, il ne pourra reprendre l'interaction qu'une fois le système à nouveau disponible et l'étape

commandée précédemment effectuée. Cette étape est souvent considérée comme un temps mort par les sujets dont il faut éventuellement profiter pour continuer une autre tâche, selon les durées de celle-ci.

A ce stade de l'interaction, le sujet ne peut pas encore savoir s'il y a eu une erreur système ou non et il ne peut donc qu'attendre la prochaine disponibilité du système.

#### **Phase 4 : Feedback d'effet de la commande (FB EC) (phase Système)**

Cette phase correspond à l'effectuation par le **Système** de la commande demandée, ainsi que de la production d'un Feedback d'effet de la commande (FB EC) indiquant que le système est prêt pour une éventuelle nouvelle commande, dans le cas d'un feedback positif ou prêt pour une correction en cas d'erreur.

Le FB EC prend ici deux formes fonctionnellement redondantes mais usant de deux modalités différentes :

##### ***Feedback en synthèse vocale déclarant « Insertion d'un dessin »***

- Affichage de la fenêtre d'insertion de dessins

Les sujets ont donc le choix quant à la modalité de monitoring de la survenue du FB PEC, qui doit alerter le sujet de la possibilité d'interagir. Selon les commandes, le délai entre le FB PEC et le FB EC peut être important et ce temps peut-être mis à profit pour réaliser une autre tâche, en l'occurrence ici la recherche de la réponse sur un cartel de l'exposition. Dans le cas où la commande constituerait la fin ou l'unique commande d'une séquence, le sujet pourrait alors se désengager totalement de l'interaction, mais ici, la commande sera suivie d'une autre (sélection d'un dessin à insérer) et le sujet se réengage donc dans l'interaction en fonction de la modalité qu'il souhaite employer. Dans ce cas, on se retrouve dans la situation présentée en phase 0.

#### **Préparation :**

- Physique :
  - Le sujet peut vouloir avoir une certaine redondance (en plus du FB EC vocal) dans la confirmation de l'effectuation de la commande et valider visuellement la fin du cycle d'interaction
  - En fonction de la modalité suivante, le regard doit à nouveau s'orienter ou non vers l'interface
  - La main gauche se replace en position moyenne
  - La main droite se positionne en fonction de la modalité suivante
- Cognitif :
  - Le sujet dans cette phase vérifie la validation de la commande et ses effets
  - Le cas échéant, il doit préparer la commande suivante et choisir sa modalité

Cette illustration permet de faire ressortir un ensemble de caractéristiques des séquences d'interaction multimodales en mobilité qui sont rarement documentées dans les études sur les usages. Celles-ci seront reprises par la suite (Proposition d'un modèle enrichi pour l'interaction mobile, p231) dans une proposition d'enrichissement du modèle d'interaction de (Kirsh, 1997) afin de les réintégrer dans les démarches de modélisation. En effet, des notions comme la modalité de contrôle, l'engagement et désengagement, ainsi que la prise en compte des séquences détaillées de l'enchaînement des actions, semblent tout à fait pertinentes à prendre en compte tant dans une démarche de conception que d'évaluation.

Par ailleurs, ces notions seront également intégrées dans notre proposition de critères et recommandation pour la conception d'interfaces multimodales mobiles (p241).

### 3. Synthèse des propriétés effectives et subjectives des modalités

Le Tableau 30 rassemble et synthétise l'ensemble des propriétés effectives et subjectives (telles que perçues par les sujets et obtenues via les autoconfrontations) des modalités et médias mis en œuvre dans cette étude.

**Tableau 30 : Tableau récapitulatif des propriétés des modalités et médias**

<b>Modalité Média</b>	<b>Propriétés</b>
<b>Vocale entrée</b>	Labile Non visuelle Modalité de contrôle non visuelle Mise en œuvre de repères absolus ou relatifs peu précis (selon l'implémentation) Peu sensible au design (taille etc.) des icônes et de l'interface Favorise la mobilité Libère les membres (main, bras) Nécessité de Mémorisation des commandes Sensible à la sémantique choisie  Mémorisation des séquences d'actions nécessaire Sensible au contexte social Manifeste au niveau sonore Discrète au niveau visuel Sensible au contexte sonore Sensible aux erreurs (performance moyenne) Correction des erreurs facile et peu visuelle Oblige à gérer les modes d'interaction Obligation de « push to talk »
<b>Modalité Vocale sortie (synthèse vocale pour l'interaction)</b>	Labile Non visuelle Sensible à la sémantique choisie Sensible à la vitesse de lecture Sensible au contexte social Sensible au contexte sonore Favorise la mobilité
<b>Audio sortie</b>	Labile Non visuelle Sensible à la sémantique choisie (arbitraire ou analogique) Sensible au contexte social Sensible au contexte sonore Favorise la mobilité
<b>Tactile entrée (digital et stylet)</b>	Plus ou moins visuel Modalité de contrôle plus ou moins visuelle Mise en œuvre de repères relatifs allant de très précis à peu précis Mobilise en général les deux mains N'oblige pas à gérer les modes d'interaction Pas d'obligation de « push to clic » mais obligation de gestion de

	veille Manifeste au niveau visuel Discrète au niveau sonore Peu sensible au contexte social Ne favorise pas la mobilité
<b>Tactile digital</b>	Idem précédent et : Pas de gestion de stylet Faible possibilité de précision de la visée Modalité de contrôle moins visuelle
<b>Tactile stylet</b>	Idem précédent et : Gestion d'un stylet obligatoire Précision possible de la visée Modalité de contrôle très visuelle Mobilise en général les deux mains  Possibilité d'intégrer de petits objets interactifs (clavier virtuel etc.) Très visuel Peu mobile
<b>Média vocal sortie (lecture de contenu par synthèse vocale)</b>	Labile Non visuel Sensible au contexte social Sensible au contexte sonore Sensible à la vitesse de lecture Favorise la mobilité Nécessité de pilotage du média si taille importante
<b>Média visuel sortie</b>	Rémanent Visuel Manifeste au niveau visuel Discret au niveau sonore Peu sensible au contexte social Ne favorise pas la mobilité Nécessité de navigation dans le média si taille importante

Un certain nombre de ces propriétés, est fortement lié au caractère subjectif de leur perception par les sujets, et il convient donc de les relativiser selon les sujets, les types d'implémentation etc. Ce tableau constitue en cela, une première tentative de lister un ensemble de caractéristiques à prendre éventuellement en compte pour la conception d'applications mettant en œuvre la multimodalité en mobilité.

## 4. Proposition d'un modèle enrichi pour l'interaction mobile

### 4.1. Objectifs

Dans le cadre de notre étude, nous nous intéressons particulièrement au partage des ressources attentionnelles et physiques de l'utilisateur qui doit répartir son activité au moins entre trois types de tâches potentiellement concurrentes : la gestion de sa mobilité, l'interaction avec le système, l'exploration de son environnement. Ces trois tâches peuvent largement s'entremêler et font intervenir aussi bien des actions physiques que cognitives. Selon l'activité du sujet, l'une des tâches peut prendre le dessus sur les autres, et il est quasiment impossible de les réaliser toutes en même temps. La bonne gestion des ressources devient donc primordiale pour réaliser cet enchaînement de tâches complexes.

Afin de concevoir des systèmes destinés à être utilisés dans de telles conditions de mobilité, il est utile de disposer d'un modèle de l'interaction homme-machine qui prenne en compte les critères déterminants l'usage des ressources attentionnelles que ce soit du côté du sujet, mais également concernant le comportement du système informatique.

### 4.2. Intention et action

Le modèle que nous proposons s'intéresse en particulier à la séquence 4 du modèle de Norman, dans lequel la séquence d'exécution de l'action n'est pas du tout décrite en détail.

Une des difficultés rencontrées lorsque l'on s'intéresse au détail, voire au micro-détail d'une interaction, est de faire la part de la cognition consciente de celle des automatismes à la fois cognitifs et moteurs, où la notion d'intention devient floue et difficilement identifiable. Celle-ci peut parfois être recueillie via les verbalisations (synchrone ou asynchrone) des sujets, inférée par l'expérimentateur via son insertion dans le cours d'action, ou simplement impossible à formuler (soit du fait de son inexistence ou d'un degré d'automatisme la rendant inconsciente).

Dans ce contexte, la notion d'intention ne peut être facilement utilisable et identifiable que si l'on s'intéresse uniquement aux intentions d'un niveau d'abstraction relativement élevé, dont le sujet a conscience et qu'il est capable de verbaliser. C'est le cas de la plupart des activités liées à l'interaction avec le système qui sont de l'ordre des commandes discrètes. Elles se traduisent généralement lors des autoconfrontations par des verbalisations du style : « là, je voulais créer un nouveau message » ou encore : « là, j'attendais que le message se ferme ».

### 4.3. Un modèle détaillé de l'action

Dans le cas des commandes discrètes, il est donc possible de découper les phases de l'interaction en séquences successives d'actions à la fois du sujet et du système. Cet enchaînement d'événements forme une boucle pouvant rappeler en partie le cycle de Norman.

Dans notre proposition de modèle, nous conservons l'idée globale d'un cycle composé d'une intention suivie par une action et l'évaluation des effets de cette action en vue d'ajuster une nouvelle intention. Cependant, nous souhaitons détailler davantage les agencements possibles de certains événements, qu'ils concernent aussi bien l'action ou la perception, en s'attachant particulièrement aux ressources attentionnelles mobilisées pour l'interaction.

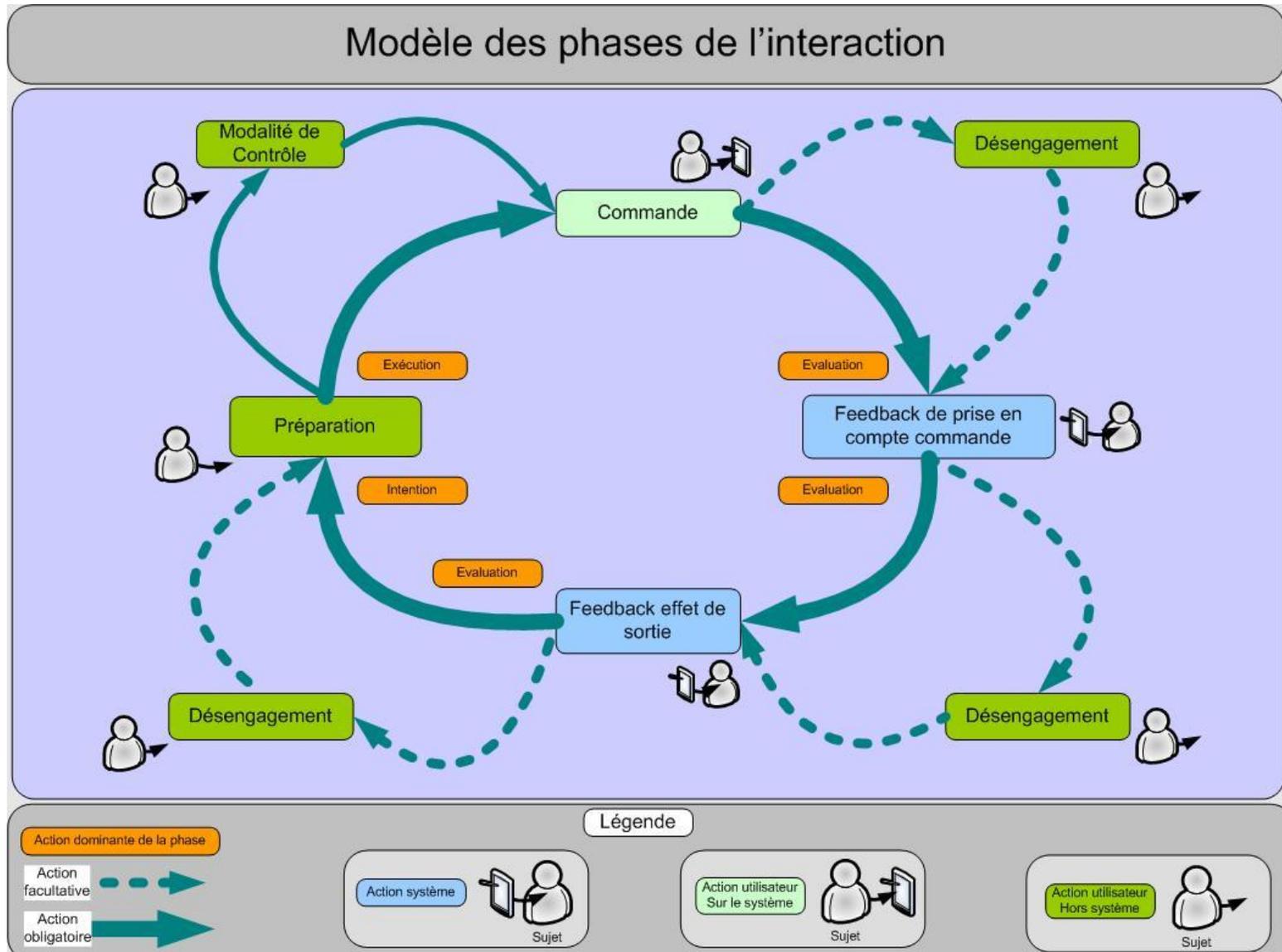


Figure 81 : Proposition d'un modèle détaillé de l'interaction

Proposition d'un modèle enrichi pour l'interaction mobile

Les séquences que nous proposons sont chronologiques et doivent être envisagées dans le cadre d'interactions en situation de mobilité provoquant une certaine concurrence entre les ressources attentionnelles (partagées à la fois entre la gestion du déplacement, de l'activité principale et de l'interaction avec le dispositif (Calvet & Salembier, 2006; Calvet, Salembier, Kahn, & Zouinar, 2005)).

La séquence proposée est constituée d'actions qui ont des propriétés propres en terme d'intensité de mobilisation des ressources attentionnelles. Les aspects cognitifs sont également indiqués en filigrane, (souvent sous la forme de cycles de Norman de bas niveau, proche du niveau « habiletés » ou « skills » selon Rasmussen). Une partie des événements est liée aux actions de l'utilisateur et une autre à celles du système.

Voici leur enchaînement : Préparation, Entrée, Contrôle, Désengagement, Feedback de prise en compte de la commande, Feedback lié à l'effet de la commande.

Nous décrivons par la suite chacune de ces phases.

#### 4.3.1. Préparation (action utilisateur hors système)

Au moment précédent l'exécution d'une commande (liée à une intention donnée), le sujet se prépositionne en effectuant une séquence motrice des membres concernés par la modalité d'interaction choisie, anticipant ainsi la séquence d'actions suivantes (phénomène de homing sur la souris par exemple, ou rapprochement du PDA vers la bouche et placement du pouce sur le bouton en vue d'une interaction vocale).

#### 4.3.2. Entrée ou commande (action utilisateur sur le système)

Une fois la préparation de la commande effectuée, le sujet peut interagir proprement dit avec le système en effectuant l'action souhaitée. A chaque modalité vont être associées des modalités de contrôle mettant en œuvre des ressources attentionnelles différentes selon les modes d'implémentation que nous décrivons ci-après.

#### 4.3.3. Modalité de contrôle (action utilisateur hors système) et types de repères

Les phases 1 et 2 (préparation et action) sont toutes deux des phases d'action, fortement orientées par une intention et un objectif précis. Il s'agit du point culminant de l'interaction qui requiert une attention plus soutenue. Toutes deux font intervenir les modalités de contrôle associées aux modalités d'entrées et de sorties, selon l'implémentation du système choisie.

Pour rappel, la notion de modalité de contrôle se situe en amont de la réponse logicielle de l'artefact à l'interaction dans la boucle action- (contrôle) - perception. Il s'agit de bien distinguer la modalité d'émission (ou de production de la commande), de la modalité de contrôle qui permet au sujet, dans un premier temps, de "contrôler" l'acte moteur de commande (contrôler la visée d'un bouton avec le stylet, ou s'écouter et se sentir prononcer une commande vocale par exemple) (voir Chapitre 1, 5.2.5 Modalité de contrôle).

La modalité de contrôle peut être associée dans un deuxième temps à la réponse de l'artefact à l'interaction qui se retrouve sous deux formes :

- la **réponse physique** (non logicielle) du dispositif au contact du membre effecteur de la commande (sensation de contact direct du dispositif avec le stylet pour la commande tactile, ou sensation de toucher du pouce au contact d'un bouton physique par exemple)
- le **retour système** implémenté (logiciel) proprement dit (voir la suite sur les feedbacks).

Les modalités de contrôle mises en œuvre dépendront à la fois fortement des modalités d'entrées choisies par l'utilisateur mais également du design physique et logiciel du dispositif choisis par les concepteurs.

Les modalités de contrôle mises en œuvre vont également dépendre très fortement de la nature et des enchaînements de repères associés à chaque modalité d'entrée selon leur implémentation. En effet, comme nous l'avons déjà mentionné, le détail d'implémentation de chaque phase d'usage d'une modalité est déterminant en ce qui concerne la mobilisation des ressources attentionnelles selon que les repères sont absolus ou relatifs. Par exemple, la présence d'un bouton physique d'activation de la modalité vocale (Push to talk) doit être soigneusement conçu si l'on ne veut pas rendre la phase d'appui et d'enclenchement excessivement visuelle. En effet, si le design physique du dispositif ne permet pas l'appui à l'aveugle, par exemple grâce à une distinction par la localisation et la texture (format) du bouton,

le sujet devra regarder le dispositif pour localiser le bouton et grèvera ainsi une partie des propriétés non visuelles de l'interaction vocale. Dans ce cas, cela obligerait à l'usage d'un repère relatif précis plus coûteux en ressources attentionnelles (usage de la vision pour viser le bouton, ressources tactilo-proprio-kinesthésiques pour discriminer l'appui sur le bon bouton etc.), alors que la phase de commande de la modalité vocale met plutôt en œuvre des repères relatifs très peu précis (pour de nombreux sujets, le regard est porté vers l'environnement et le PDA est soit vaguement rapproché de la bouche, ou laissé en position basse neutre).

Les notions de modalité de contrôle et de repères absolus ou relatifs sont essentielles pour la conception de dispositifs mobiles et sont pourtant très peu présentes dans la littérature du domaine.

#### 4.3.4. Désengagement (action utilisateur hors système)

A tout moment, l'utilisateur peut se désengager de l'interaction, en raison éventuellement de sollicitations importantes d'autres activités (gestion du déplacement, exploration de l'environnement, etc.), mais également du fait d'un choix de conception de l'interaction qui l'avertira en cas de nouvel événement système attendu ou inattendu.

Le désengagement peut revêtir plusieurs formes selon les modalités d'entrées utilisées et selon la situation. Le désengagement peut simplement consister en un simple détournement bref du regard hors de l'interface, jusqu'à un repositionnement neutre des membres et du dispositif.

#### 4.3.5. Feedback de prise en compte de la commande (action système)

Si l'exécution de la commande n'est pas immédiate (ce qui est souvent le cas pour des dispositifs mobiles dont les ressources sont très sollicitées), il peut y avoir un décalage temporel entre l'action (la commande) et la perception (sortie du système) qui peut elle-même être scindée en deux phases, selon les cas. Le feedback proprement dit qui ici est considéré comme la réponse (logicielle) du système confirmant la prise en compte de la commande (ex : bouton qui s'enfonce, bip, etc.). Ce feedback permet à l'utilisateur de vérifier que son ordre a bien été pris en compte et que son exécution est en cours, le cas échéant, cela peut lui permettre de se désengager momentanément de l'interaction.

#### 4.3.6. Feedback lié à l'effet de la commande (action système)

Si la commande prend également du temps à être exécutée, un deuxième feedback indiquant l'exécution effective de la commande, peut être implémenté. Celui-ci permet, comme le feedback de prise en compte de la commande, de libérer l'attention du sujet et de l'avertir lorsque le système est à nouveau disponible pour un nouveau cycle d'interaction. Ce feedback peut prendre des formes diverses comme un simple signal visuel, audio ou vocal pour indiquer l'envoi effectif d'un message qui a pris du temps comme par exemple :

- un pop-up « message envoyé »
- un bip sémantiquement signifiant
- une synthèse vocale disant « message envoyé »
- ou encore une vibration du dispositif sémantiquement signifiant

Si la commande est effectuée immédiatement, le feedback d'effet de la commande peut être temporellement confondu avec le feedback de prise en compte puisque l'effet devient de fait un feedback implicite de prise en compte.

Pour l'ensemble des feedbacks présentés ici, la modalité choisie sera déterminante sur le coût en ressources attentionnelles. Il nous semble ici que le principe d'équivalence fonctionnelle maximale en sortie doit être également pris en compte afin de fournir à l'utilisateur la possibilité d'accès au feedback quelque soit la configuration en cours de ses ressources attentionnelles. Bien entendu, l'utilisateur doit pouvoir tout de même garder le contrôle sur les modes de sorties selon ses préférences ou selon le contexte local, en désactivant par exemple la synthèse vocale ou l'ensemble des sorties audio.

Dans le même ordre d'idée, et dans le cas où les applications présentent un certain nombre d'actions automatisées, comme c'est le cas des applications de Context Aware Computing (CAC), par exemple le

rafraichissement d'informations en fonction de la localisation ou de l'heure (Oulasvirta & Saariluoma, 2006), ou encore une présentation différente de ces informations en fonction du type d'activité (Pascoe et al., 2000), il conviendrait de fournir des feedbacks de changement d'état ou de rafraichissement de l'application dans des modalités appropriées.

De même pour des éléments ponctuels ou discrets tels que les avertissements d'arrivée de message, ou d'entrée dans une zone de disponibilité du réseau par exemple, il conviendra de fournir des feedbacks appropriés.

Ce type d'évènements ne sont pas représentés dans notre modèle, entre autre car notre application n'était pas sensible au contexte.

#### 4.3.7. Organes et sens mis en œuvre

Pour chacune des phases décrites dans ce cycle, il est possible de distinguer les organes et les sens mis en œuvre selon les modalités d'entrées disponibles et le design physique et logiciel du dispositif. Dans le cadre d'une interaction mobile, cette potentialité prend toute son importance dans la mesure où l'interaction avec le dispositif va devoir s'insérer dans le reste de l'activité de l'utilisateur dont entre autre la gestion de son déplacement.

Il convient donc dans la conception d'un dispositif multimodal mobile d'identifier en détail chaque phase d'interaction potentiellement mise en œuvre et les organes, modalités d'interaction (entrées et sorties), modalités de contrôle, repères (absolus et relatifs, précis ou non) associés. Il s'agit donc d'une micro analyse de l'interaction qui nous semble indispensable

#### 4.3.8. Remarques sur la manipulation directe

Lorsque la réponse du système à une commande est immédiate, la boucle perception-action est raccourcie, il y a unification des deux phases (manipulation directe par exemple), la phase épistémique (perception) est confondue avec la phase commande (action) et les phases décrites ici peuvent alors devenir moins pertinentes. Cependant, et pour rappel, notre application ne fait pas intervenir le style d'interaction de manipulation directe mais plutôt des commandes discrètes pour le tactile et du langage de commande pour le vocal.

Il conviendra donc de reconduire une analyse de ce type pour des interfaces de type manipulation directe car l'ensemble des résultats décrits ici ne s'appliquent pas forcément directement.

Cependant nous pouvons d'ores et déjà arguer, comme nous l'avons fait en introduction, qu'un tel style d'interaction reposant essentiellement sur les ressources visuelles n'est pas forcément des plus pertinent pour des interactions réellement mobiles. De plus le concept d'équivalence fonctionnelle maximale devrait alors être étudié dans le détail pour être valide.

## 5. Proposition de modèle pour une mobilité finalisée

Pour démarrer notre synthèse concernant la mobilité, nous proposons de nous recentrer sur la problématique de la mobilité du point de vue ergonomique en s'intéressant donc à l'activité d'un sujet ayant une tâche à accomplir à l'aide d'un certain nombre de ressources. En effet, cette approche permet de réintroduire la notion de finalité de l'activité qui peut fournir un cadre intéressant pour décrire différents types de mobilité selon les objets, les finalités visées.

La Figure 82 est une proposition de cadre d'analyse d'un point de vue centré sur l'utilisateur et son activité. Elle constitue une synthèse des observations que nous avons menées en situation de mobilité en contexte naturelle, avec des aspects théoriques abordés dans notre état de l'art.

Avant de s'intéresser à l'aspect mobile de son activité, il est utile de commencer par dresser un bilan des objets de l'interaction du sujet avec son environnement qui est constitué de plusieurs types d'entités produisant à son tour plusieurs types de finalités. Nous abordons la description de ce modèle sous ses deux aspects principaux : l'objet de la tâche puis la finalité de la tâche et de la mobilité.

Notre cadre se concentre ici sur des types d'objets exclusivement physiques ou numériques et ne traite donc pas du champ, en plein essor, qu'est l'étude des systèmes mixtes, comprenant des possibilités de réalité augmentée et/ou de virtualité augmentée (Coutrix & Nigay, 2006).

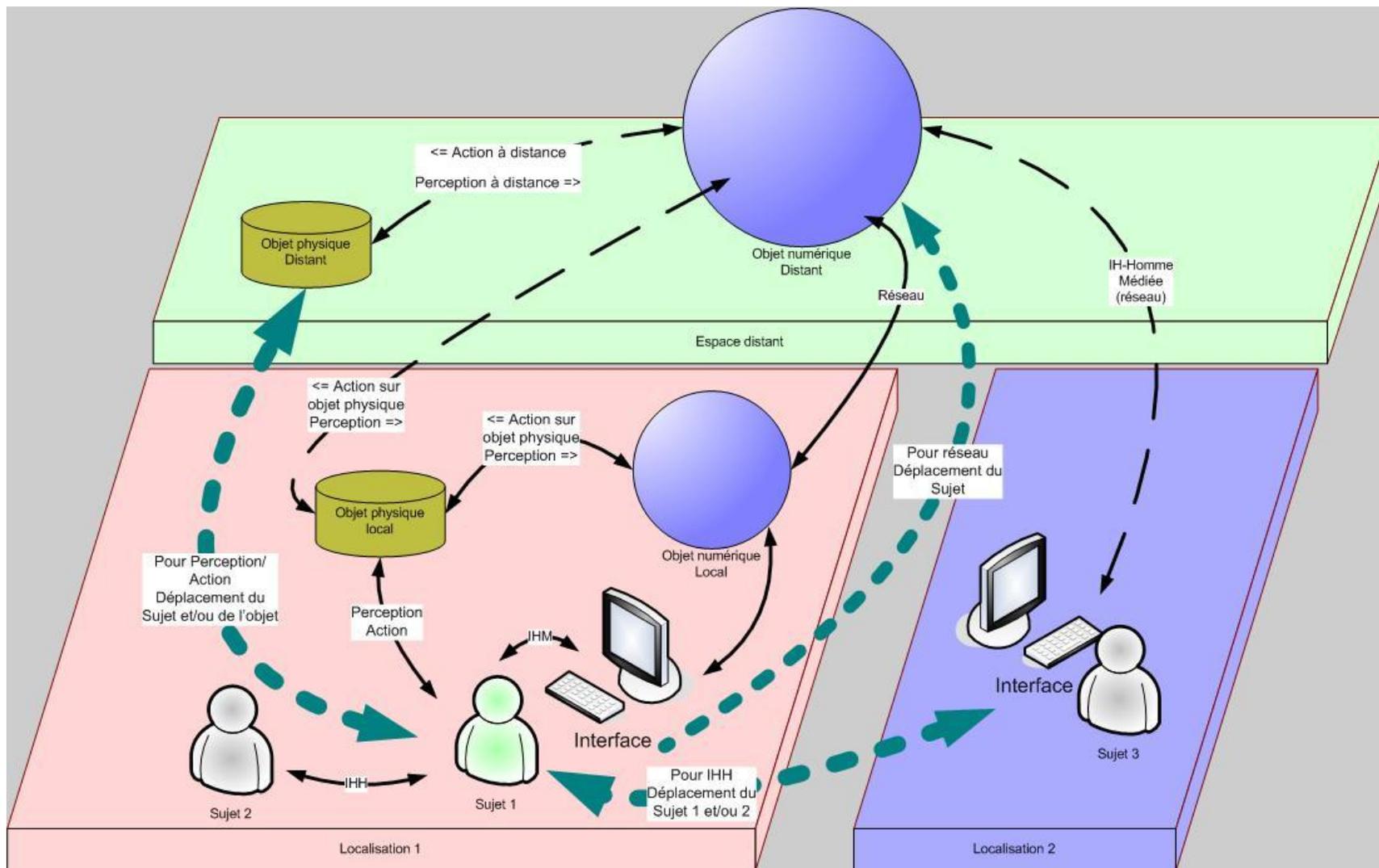


Figure 82 : Cadre descriptif général des finalités de l'interaction et de la mobilité

### 5.1. Objets de la tâche

Un sujet ayant une tâche finalisée à réaliser dispose donc de plusieurs types de ressources exploitables : les objets virtuels, les objets physiques, d'autres sujets pouvant être à sa portée. Bien que le courant de la cognition distribuée (Hutchins, 1995) nous invite à toujours considérer ces diverses ressources en interaction les unes avec les autres, par souci de clarté, nous les présenterons de manière individuelle dans un premier temps :

- **Interaction Homme-Homme (objet social de la tâche)** : L'objet de l'activité d'un sujet peut-être constituée en partie ou entièrement par une interaction avec un tiers. L'interaction peut d'ailleurs être une finalité complète en soi, c'est-à-dire que l'objet de la tâche du sujet réside purement dans un échange avec autrui (par exemple demander à un collègue un avis d'expert), mais le tiers peut également constituer une ressource partielle pour la complétude d'une tâche comportant un objet autre. Dans un cas comme dans l'autre, la condition de satisfaction minimum à atteindre dans un premier temps est la **coprésence** des sujets.  
Dans le schéma présenté ici, il s'agit du cas des sujets 1 et 2 qui sont localisés dans un même espace, en situation de coprésence, partageant un même espace contextuel leur permettant d'interagir à la fois verbalement, gestuellement, etc. et en partageant éventuellement une même interface ou objet physique local.
- **Interaction Homme-Machine (objet numérique de la tâche)** : L'objet de l'activité du sujet peut également être constitué de l'interaction avec un système donnant accès à la création, au stockage et à la modification, etc. d'objets numériques via une interface d'entrée et de sortie. En effet, une part grandissante de l'activité humaine s'effectue uniquement dans le monde numérique avec parfois très peu d'incidence dans le monde matériel. Cependant, à l'inverse, une activité essentiellement numérique peut avoir pour finalité la création d'un objet physique, par exemple la retouche d'une image numérique en vue de son impression sur papier photo. Sur notre schéma, il s'agit de la flèche libellée « IHM » reliant les sujets à leur interface.
- **Interaction Homme-Objet physique (objet physique de la tâche)** : Il s'agit de la relation la plus naturelle de l'homme à son environnement matériel. L'homme perçoit le monde matériel, agit sur la matière, tout en en retirant de l'information en retour. A ce sujet (Cadoz, 1992) décrit précisément la relation de l'homme au monde matériel, en particulier en ce qui concerne les différents types de gestes. Le geste ergotique transmet de l'énergie à la matière, tout en la transformant, le geste épistémique permet lui de percevoir les états physiques de la matière et d'en prendre connaissance. Le geste sémiotique intervient au niveau informationnel pour transmettre des signes (incluant l'interaction Homme-Machine). La relation de l'homme à son environnement peut aussi se passer du contact physique en se focalisant par exemple sur la vision. Dans le cas de l'étude de (Pascoe et al., 1998), des écologues procèdent à un comptage à distance du nombre d'animaux rencontrés sur un parcours bien précis. Sur notre schéma, il s'agit de la flèche libellée « Perception/ Action » reliant le sujet à un objet physique local.

Il est bien évident que ce découpage des objets de la tâche en trois catégories est extrêmement réducteur, et que la réalité de l'activité est bien souvent constituée par un mélange combinant l'ensemble des objets présentés ci-dessus. Cependant, il reste tout de même utile en terme descriptif pour illustrer un grand nombre de finalités de la mobilité.

### 5.2. Finalité de la tâche et de la mobilité

Pour un grand nombre d'activités professionnelles mais également grand public, il est possible d'identifier une/des finalités expliquant le besoin de mobilité. Il est d'ailleurs également possible de distinguer au moins deux types de mobilité, selon qu'elles sont incluses ou non dans l'objet de la tâche.

- **Tâche et mobilité ayant un but identique (mobilité « équifinale »)** : Par exemple, des écologues pour compter les animaux doivent obligatoirement se mouvoir sur une ligne droite bien définie (transec), la mobilité est donc ici un prérequis, et constitue une partie de la tâche. Il y a

équifinalité entre la mobilité et la tâche (mobilité équifinale). Dans notre étude, nous avons procédé au codage de l'activité et de la mobilité. Le code « Réponse Perso » (RP), lorsqu'il est associé au déplacement du sujet correspond à ce type de mobilité équifinale. En effet, le sujet se déplace afin de chercher la réponse à une question du Quiz tout en interagissant avec son PDA contenant des indices permettant de trouver cette réponse.

Dans ces cas de figure, la conception de l'interface doit donc prendre en compte l'aspect mobile inhérent à la nature de la tâche.

- **Tâche et mobilité ayant un but différent (mobilité « alterfinale »)** : Par opposition, la mobilité en cours peut ne pas être constitutive ou corrélée à la tâche en cours en parallèle. Il s'agit finalement de la majorité des situations d'interactions courantes, en particulier pour le grand public. Par exemple, répondre à un mel professionnel avec son téléphone portable tout en se rendant au cinéma. Le fait de se déplacer pour aller au cinéma n'a aucun rapport avec le mel professionnel, il y a « alterfinalité » entre la mobilité et la tâche. Dans notre étude, nous avons également de nombreux cas de mobilité et d'interaction alterfinales. Par exemple, lorsque le sujet se déplace tout en cherchant ses propres réponses, et en même temps répond à une question de son compère n'ayant rien à voir avec sa propre question et n'ayant donc aucun lien avec son déplacement.

Dans cet autre cas de figure, la conception de l'interface doit prendre en compte l'éventualité d'une interaction en mobilité sans que la tâche soit forcément effectuée en mobilité. On peut donc imaginer par exemple pour une même tâche, une interface pour une interaction sans mobilité (disons avec des boutons plus petits), et une autre au contraire où tous les objets sont magnifiés pour faciliter leur sélection en mobilité.

### 5.3. Objets de la mobilité

Nous avons donc vu à travers notre schéma, que l'on peut distinguer trois grands objets de tâches : l'interaction avec un objet physique, un objet virtuel et avec un/d' autres sujets, chacune de ces instances pouvant se situer à des degrés divers de mobilité. Cette interaction nécessite la possibilité d'accès à l'une ou l'autre de ces sources par une proximité géographique des sujets et de l'objet de leur tâche. Si cette condition n'est pas respectée (coupure d'une flèche relationnelle locale de notre schéma), il est nécessaire de trouver une procédure alternative si la tâche doit être tout de même effectuée.

Dans ce cas, les objets de la tâche deviennent distants et le sujet, pour interagir avec eux, doit alors procéder à son déplacement et/ou à celui de son objet ou bien disposer de ressources supplémentaires :

- **L'interface** : Elle devient ici indispensable pour interagir avec les objets numériques mais également avec les objets physiques et les sujets distants. Comme nous avons pu l'étudier dans cet état de l'art, certaines interfaces nouvelles peuvent être transportées et peuvent même permettre d'être utilisées à des degrés divers de mobilité (du nomadisme de niveau 1 à la mobilité réelle).
- **L'objet de la tâche** : L'interface en elle-même n'est utile que si elle est associée à un objet sur lequel interagir. Celle-ci est donc à l'heure actuelle quasiment toujours accompagnée d'un objet numérique local (application locale avec mémoire associée), mais l'augmentation de la connectivité permanente, permet aujourd'hui d'envisager parfois les interfaces uniquement comme des terminaux permettant d'interagir sur des objets virtuels, voire physiques, distants. L'interaction avec un ou des tiers est donc également médiée dans ce cas.

Pour résumer simplement, lorsque l'objet de la tâche n'est pas accessible, soit c'est l'information concernant cet objet qui se déplace, soit c'est le sujet, voire les deux.

### 5.4. Combinaisons possibles des objets de la tâche

Comme nous le mentionnions précédemment, nous avons procédé à un découpage simplifié des objets de la tâche dans un but de clarification ne correspondant que peu à la réalité complexe de nombreuses activités faisant intervenir des combinaisons multiples des situations que nous venons de décrire. Nous décrivons ici quelques cas à titre d'exemple.

Cas du CSCW (Computer Supported Colaborative Work) :

Dans le cas d'activités complexes généralement liées au CSCW, nous pouvons aisément envisager toutes les possibilités d'interaction présentées dans notre modèle (Figure 82). S'agissant de tâches complexes, de haut niveau, parfois étalées sur de longues périodes et faisant intervenir un nombre variable d'acteurs, la nécessité de recours au travail à distance devient quasiment impossible à éviter. La notion de mobilité ici n'est pas forcément vue comme une notion dynamique de déplacement relatif à un objet mais plutôt comme une impossibilité (ou un coût trop élevé) à son accès direct (local).

Cas des activités à mobilité variable :

Dans l'étude de (Pascoe et al., 1998), ceux-ci décrivent différentes phases de l'activité des écologues qui doivent préparer au laboratoire des expéditions sur le terrain, conduire ces expéditions, puis à nouveau regagner le laboratoire pour analyser les données récoltées. A chaque étape de leur activité, on peut observer des degrés de mobilité très différents (de l'interaction en mobilité réelle avec un PDA sur le terrain, à l'interaction classique de bureau fixe avec un ordinateur statique). L'interface de l'application est ici conçue spécifiquement en relation avec la nature mobile de la tâche et en particulier le degré et le lieu du focus de l'attention requise.

## 6. Critères et recommandations pour la conception et l'évaluation d'interfaces mobiles multimodales

A notre connaissance, il n'existe aucune liste synthétique de critères d'évaluation ou de classification de recommandations ergonomiques dédiées aux interfaces multimodales et mobiles. En effet, comparé au nombre conséquent de recommandations ergonomiques pour le web ou les applications d'ordinateurs de bureaux (GUI), seules quelques recommandations éparses peuvent être identifiées dans la littérature concernant ces sujets. Ce faible nombre de recommandations dédiées aux interactions multimodales mobiles obère, pour l'instant, toute tentative de construction et a fortiori de validation de critères ergonomiques dédiés à ce type d'interactions. Notre objectif ici n'est donc pas de proposer une liste de critères exhaustive et consolidée construite en tant qu'objet de recherche, mais plutôt d'établir une première liste de recommandations issue des résultats de nos expérimentations (et dans certains cas adaptés de la littérature) et dont la validité pourrait être éprouvée dans un second temps. Cette première contribution au domaine s'inscrit donc dans une approche globale et incrémentale de recherches qui aspirent à améliorer l'utilisabilité globale des systèmes interactifs multimodaux mobiles.

Afin de bénéficier des connaissances ergonomiques établies, sur la base d'études portant sur d'autres systèmes interactifs (GUI, Web, Environnements 3D), et dans un souci de capitalisation, de gestion des connaissances du domaine de l'ergonomie des systèmes interactifs, nous avons choisi de classer nos recommandations à l'aide des Critères Ergonomiques (CE) de (Bach & Scapin, 2005). En effet, cette version des CE, qui est le résultat d'une adaptation des CE de (Bastien & Scapin, 1993) aux Environnements Virtuels, semble bien adaptée, à la problématique des interfaces multimodales mobiles, car elle prend en compte à la fois la spatialisation (prise en compte de la 3D) et la multimodalité des interactions.

Nos recommandations seront présentées suivant le formalisme suivant :

- intitulé du CE, sa définition (entre guillemets) telle que proposée par Bach & Scapin (2005),
- description de nos recommandations associées classées par le CE considéré,
- le cas échéant l'objet ou la modalité sur laquelle s'applique la recommandation (sans précision cela signifie que la recommandation s'applique de façon générale à l'ensemble du système),
- la justification de la recommandation,
- quand cela est indiqué, un exemple d'instanciation de cette recommandation.

Les critères élémentaires (les critères qui s'appliquent) sont dotés d'un astérisque.

### 6.1. **Compatibilité \***

Définition :

« Le critère Compatibilité se réfère à l'accord pouvant exister entre les caractéristiques des utilisateurs (mémoire, perceptions, anatomie, habitudes, compétences, âge, attentes, etc.) et des tâches, d'une part, et l'organisation des sorties, des entrées et du dialogue d'une application donnée, d'autre part.

De plus, la Compatibilité concerne également le degré de similitude entre divers environnements ou applications. »

*Recom. 1 :* Prendre en compte toutes les phases potentielles du cycle d'interaction proposé (comprenant : Préparation, Entrée, Contrôle, Désengagement, Feedback de prise en compte de la commande, Feedback lié à l'effet de la commande) pour la conception d'une interface mobile multimodale, en particulier pour le choix des modalités à mettre en œuvre.

Justification :

En fonction de la tâche et de l'implémentation de l'application, une lacune de prise en compte de l'une de ces phases peut générer des écarts entre les attentes de l'utilisateur et l'implémentation effective du

système. Ces écarts peuvent entraîner des risques d'erreurs, une charge cognitive importante, une faible confiance en la fiabilité du système ou encore une insatisfaction globale chez les utilisateurs.

*Recom. 2 : Veiller à la continuité et à la cohérence des organes, sens, et modalités mis en œuvre dans toutes les phases du cycle en entrée comme en sortie du système.*

Justification :

Pour une interaction en général et une interaction mobile en particulier, les modalités (dont la modalité de contrôle) mises en œuvre dans une seule de ces phases va influencer l'ensemble des ressources attentionnelles nécessaires pour l'ensemble des phases (un cycle d'interaction). Il est donc critique de veiller à la cohérence des modalités mises en œuvre en particulier pour permettre une interaction non visuelle.

Contexte mobile :

*Recom. 3 : Dans le cas d'une application à mobilité et tâche « équifinale », l'interface doit être spécifiquement conçue pour le contexte de mobilité « réelle » requis par la tâche primaire.*

Justification :

« Equifinale » signifie que la tâche primaire, extérieure à l'interaction, requiert une interaction mobile « réelle » (interaction et déplacement simultanés) pour être satisfaite. Il convient de faciliter ou au moins de ne pas perturber la tâche primaire extérieure par une interaction inadaptée. Il convient également d'identifier les différents types de mobilité requis pour les différentes phases de la tâche afin d'adapter au mieux l'interface (mobilité supportée par ordinateur, Computer Supported Mobility).

Exemple :

Un employé de maintenance relève à la volée l'état des rails de chemin de fer en marchant le long de la voie.

Un écologue compte les animaux qu'il rencontre le long d'une ligne de déplacement (transect).

*Recom. 4 : Dans le cas d'une application à mobilité et tâche « alterfinale », l'interface doit au moins être à « attention minimale ».*

Justification :

« Alterfinale » signifie que les tâches d'interaction, et celles extérieures à l'interaction, ont des finalités différentes et indépendantes et sont donc concurrentes en terme de ressources attentionnelles. Dans ce cas, la tâche d'interaction doit être compatible avec la gestion du déplacement.

Exemple :

Le choix de la modalité de la commande pour accepter un appel téléphonique ne doit pas interrompre le déplacement de l'utilisateur.

*Recom. 5 : Identifier les repères absolus/ relatifs et les modalités de contrôle associées pour toutes les phases potentielles du cycle d'interaction proposé (comprenant : Préparation, Entrée, Contrôle, Désengagement, Feedback de prise en compte de la commande, Feedback lié à l'effet de la commande)*

Justification :

Toute action (dont l'interaction) met en œuvre des repères absolus (par exemple la gravité ou l'horizontalité) et ou relatifs (par exemple la distance entre les yeux et le dispositif) qui vont dimensionner la qualité et la quantité des ressources attentionnelles mobilisées par les utilisateurs. Leur identification permet de les prendre en compte dans la conception de l'application.

Exemple :

L'appui à l'aveugle sur un bouton physique de PDA à l'aide du pouce implique d'abord son positionnement (repère relatif entre la position du pouce et celle du bouton, modalité de contrôle tactilo-proprio-kinesthésique (TPK)), l'appui sur le bouton (repère relatif avec modalité de contrôle TPK), éventuellement un feedback de prise en compte de la commande (par exemple audio, repère absolu) etc.

*Recom. 6 : Proposer des modalités (y compris les modalités de contrôle) adaptées aux repères relatifs/absolus mis en œuvre pour toutes les phases potentielles du cycle d'interaction proposé (comprenant : Préparation, Entrée, Contrôle, Désengagement, Feedback de prise en compte de la commande, Feedback lié à l'effet de la commande)*

Justification :

Les repères absolus et ou relatifs dimensionnent la qualité et la quantité des ressources attentionnelles mobilisées par les utilisateurs. L'usage de modalités d'entrée ou de sortie adaptées peut diminuer l'impact des ressources attentionnelles mises en œuvre pour l'interaction.

Exemple :

Proposer un feedback audio sémantiquement codé de prise en compte d'une commande permet à l'utilisateur de ne pas regarder l'interface pendant le temps que le système réalise effectivement l'action demandée et fournisse à ce moment un feedback audio de réalisation de la commande.

*Recom. 7 : Il convient de contextualiser la présentation d'informations afin de faciliter son interprétation par l'utilisateur*

Justification :

Le filtrage d'information en fonction du contexte et de la tâche en cours permet d'éviter à l'utilisateur un tri manuel des données pertinentes et non pertinentes et diminue donc la charge attentionnelle de l'utilisateur.

Exemple :

Si l'utilisateur recherche les horaires des films en se déplaçant à proximité d'un cinéma, d'abord lui présenter les horaires des films proches de l'heure actuelle (contextualisation temporelle) et dans le cinéma où il se trouve (contextualisation géographique).

## **6.2. Guidage**

« Le Guidage est l'ensemble des moyens mis en œuvre, avec les diverses modalités disponibles, pour conseiller, orienter, informer, et conduire l'utilisateur lors de ses interactions avec l'interface. »

### **6.2.1. Lisibilité \***

Définition :

« Le critère Lisibilité concerne les caractéristiques multimodales de présentation des informations dans une interface pouvant entraver ou faciliter la lecture de ces informations (luminance, contraste, dimension des objets, espacement entre les objets, fréquence sonore, intensité, timbre, etc.)

Par convention, le critère Lisibilité ne concerne ni le feedback ni les messages d'erreurs. »

*Recom. 8 : Pour une interaction mobile en particulier, les objets interactifs ou non doivent être facilement perceptibles et localisables*

Justification :

Toute difficulté de lecture des éléments de l'interface augmente les ressources attentionnelles nécessaires et peut donc nuire voir empêcher le bon déroulement des tâches parallèles (gestion de la mobilité et autres tâches).

De plus, les ressources allouées par les sujets pour l'interaction peuvent être réduites et l'interface doit permettre des interactions à attention minimum.

*Recom. 9 : L'interface en générale et les objets interactifs en particulier doivent être lisibles dans un large spectre de conditions environnementales.*

Justification :

L'interaction en situation de mobilité peut provoquer de nombreuses variations de conditions d'éclairage (environnement sombre ou reflets par exemple) qui doivent limiter le moins possible les possibilités d'interaction.

#### 6.2.2. Incitation \*

*Recom. 10 : Proposer des sauts de curseur correspondants aux séquences d'actions à réaliser*

Justification :

De manière générale, un guidage de tâche facilite l'interaction, oriente l'utilisateur sur l'action suivante à réaliser diminuant ainsi les ressources cognitives et attentionnelles nécessaires.

Exemple :

Dans une fenêtre d'édition de message, lorsque l'utilisateur a terminé de saisir l'adresse du récepteur ou d'un groupe de récepteurs, un saut de curseur dans le champ « objet » lui permet de savoir qu'il peut entrer un texte à ce niveau.

*Recom. 11 : Si l'application propose une gestion par mode d'interaction, le mode actif doit être clairement identifié ainsi que les autres modes disponibles*

Justification :

L'affichage clair du mode actif en cours permet à l'utilisateur de savoir facilement avec quelle modalité il peut actuellement interagir, ce qui permet d'éviter des erreurs de mode (vouloir interagir à l'aide d'une modalité qui n'est pas activée). Par ailleurs cela facilite et réduit la courbe d'apprentissage du système multimodal.

*Recom. 12 : Le moyen de passage d'un mode à l'autre doit être clairement identifié*

Justification :

L'interface doit indiquer à l'utilisateur comment il peut passer d'un mode à l'autre. Dans le cas contraire et devant la difficulté de changer de mode, l'utilisateur pourrait être contraint d'utiliser une modalité non choisie ce qui diminuerait sa satisfaction envers le système.

#### 6.2.3. Groupement/Distinction entre items

Définition :

« Le critère Groupement/Distinction entre Items concerne l'organisation visuelle, auditive, proprioceptive, etc., des items d'information les uns par rapport aux autres. Ce critère prend en compte la topologie (localisation) et certaines caractéristiques multimodales et multimédia afin d'illustrer les relations entre les divers items présentés, leur appartenance ou non-appartenance à une même classe ou encore dans le but de montrer la distinction entre différentes classes d'items. Ce critère concerne aussi l'organisation des items à l'intérieur d'une même classe. »

##### 6.2.3.1. Groupement/Distinction par la localisation \*

*Recom. 13 : De manière générale, il est préférable de regrouper les objets, interactifs ou non, ayant des caractéristiques similaires au même endroit*

Justification :

Grâce à la mémorisation spatiale liée au regroupement des fonctionnalités, l'utilisateur peut retrouver plus facilement les éléments nécessaires. Dans le contexte d'une interaction mobile, cela permet à l'utilisateur de (pré)-orienter son action et d'identifier plus facilement les objets utiles à sa tâche.

*Recom. 14 : Il est préférable de regrouper les éléments de gestion des modes (activation et désactivation d'une modalité) au même endroit*

Justification :

Dans une application multimodale, la gestion des modes peut être proposée et, le cas échéant, celle-ci sera utilisée régulièrement. Il convient donc de veiller à sa facilité d'usage entre autre via un regroupement des commandes et sorties la concernant.

*Recom. 15 : Il est préférable de séparer/ regrouper les moyens de contrôle des modes selon qu'ils concernent les modalités d'entrées ou de sortie*

Justification :

La gestion des modalités d'entrées et de sorties est parfois dissociée du point de vue des sujets et leur usage est donc distinct. Une séparation des commandes de modes d'entrées et de sorties permet une meilleure identification de celles-ci.

#### **6.2.3.2. Groupement/Distinction par le format \***

*Recom. 16 : De manière générale, il est préférable d'attribuer des formats proches aux objets, interactifs ou non, ayant des caractéristiques similaires*

Justification :

Une meilleure saillance des commandes permet à l'utilisateur de retrouver facilement les éléments nécessaires et dans le contexte d'une interaction mobile, cela permet à l'utilisateur de consacrer moins d'attention à l'identification des objets de l'interface.

*Recom. 17 : Il est nécessaire de distinguer par le format les moyens de contrôle des modes des autres commandes disponibles dans l'interface*

Justification :

Dans une application multimodale, la gestion des modes peut être proposée et, le cas échéant, celle-ci sera utilisée régulièrement. Il convient donc de veiller à sa facilité d'usage entre autre via une distinction par le format des commandes et sorties la concernant.

*Recom. 18 : Il est nécessaire de distinguer par le format les moyens de contrôle des modes selon qu'ils concernent les modalités d'entrées ou de sortie*

Justification :

La gestion des modalités d'entrées et de sorties est parfois dissociée du point de vue des sujets et leur usage est donc distinct. Une distinction des commandes de modes d'entrées et de sorties permet une meilleure identification de celles-ci.

*Recom. 19 : Les commandes non disponibles doivent être clairement distinctes des commandes disponibles*

Justification :

Le format d'une commande doit renseigner l'utilisateur sur son état, notamment son état actif ou inactif.

Exemple : L'activation des différentes commandes ou fonctions doivent refléter l'état d'avancement d'une séquence de tâche afin de faciliter le guidage de l'utilisateur dans cette séquence. Par exemple, le bouton

« envoyer le message » doit être actif une fois que le nom d'un destinataire est renseigné dans le champ correspondant.

### 6.2.3.3. Groupement/Distinction par le comportement \*

Pas de recommandations identifiées pour ce critère.

### 6.2.4. Feed-back immédiat \*

Gestion et contrôle de l'interaction :

*Recom. 20 :* Lorsque les temps de réponse du système sont longs, prévoir au moins deux types de feedbacks : un premier feedback immédiat de prise en compte de la commande, puis un second feedback d'effet de la commande

Justification :

En cas d'interaction mobile ou à tâches concurrentes, le premier feedback de prise en compte de la commande permet à l'utilisateur un désengagement momentané de l'interaction pour réaliser une autre tâche jusqu'au feedback suivant (feedback d'effet de la commande) indiquant la disponibilité du système.

*Recom. 21 :* Fournir des feedbacks dans les modalités adaptées, en particulier lorsque les temps de réponse du système sont longs.

Justification :

La présence d'un feedback à chaque étape de l'interaction peut permettre le désengagement du focus sur l'interface et l'accomplissement d'une autre tâche. En particulier lorsque les délais sont longs, un feedback non visuel (audio, synthèse vocale, vibration) peut permettre d'attirer l'attention de l'utilisateur sur le nouvel état du système sans interrompre la tâche en cours.

*Recom. 22 :* Les feedbacks doivent être redondants au moins sous une forme visuelle et non visuelle

Justification :

Selon la modalité d'entrée utilisée, ou le délai de réponse du système, l'utilisateur peut avoir quitté des yeux l'interface. Un feedback visuel peut donc ne pas être suffisant et il convient de fournir également un feedback non visuel (audio, synthèse, vibration etc.).

*Recom. 23 :* Pour les applications nécessitant des rafraîchissements réguliers, les rendre facilement repérables et adapter la modalité de sortie, (Oulasvirta, 2005)

Justification :

Les actions sur l'initiative du système, si elles provoquent des changements d'état, doivent fournir un feedback afin d'alerter l'utilisateur sur la nature du changement.

## 6.3. Contrôle explicite

### 6.3.1. Actions explicites \*

Synthèse vocale :

*Recom. 24 :* N'utiliser un retour système en synthèse vocale que si celle-ci a été activée par l'utilisateur.

Justification :

Il convient en effet de ne pas associer systématiquement un retour système à l'aide de la synthèse vocale suite à une commande vocale de l'utilisateur.

Plusieurs sujets ont exprimé leur désapprobation d'un retour systématique en synthèse vocale qui alourdit ou ralenti le processus d'interaction.

Celle-ci ne doit être utilisée que si elle a été explicitement activée. En cas de désactivation de la synthèse vocale, une alternative possible est l'usage d'un feedback audio sémantiquement signifiant (dans le cas où l'audio n'est pas désactivé).

### 6.3.2. Contrôle utilisateur \*

*Recom. 25 : L'application doit supporter et surtout ne pas sanctionner l'organisation et la réorganisation temporelle des tâches.*

Justification :

L'interaction mobile est par nature imprévisible et les utilisateurs doivent faire face à de nombreux aléas : interruptions de tâche, changement de tâche, mise en pause de tâche, reprise de tâche, etc.

Par exemple, dans les systèmes complexes à applications multiples, le passage d'une application à une autre ne devrait pas avoir d'incidence sur l'état de la fonction précédente, ce qui permet à l'utilisateur de reprendre une tâche dans l'état et à l'endroit où il l'a laissée.

*Recom. 26 : La mise en veille ne doit pas provoquer ou avoir des conséquences de changement d'état de l'application.*

Justification :

L'utilisateur doit pouvoir reprendre la tâche à l'endroit où il en était avant l'interruption, et la mise en veille ne doit donc pas remettre l'application dans une position par défaut de manière automatique.

Synthèse vocale :

*Recom. 27 : Fournir un moyen simple d'activation et de désactivation de la synthèse vocale lorsque celle-ci est disponible dans l'application.*

Justification :

Selon le contexte (social, sonore etc.) et/ou les préférences de l'utilisateur, la synthèse vocale peut ne pas être appropriée, voire intempestive. Les sujets doivent donc pouvoir l'activer et la désactiver facilement.

*Recom. 28 : Fournir un moyen de pilotage (au minimum les fonctions pause, play) de la synthèse vocale lorsque celle-ci est disponible dans l'application.*

Justification :

En particulier dans un contexte mobile, la synthèse vocale peut devoir être interrompue par une autre tâche. Dans ce cas, l'utilisateur doit pouvoir effectuer une pause de lecture et la reprendre quand il le souhaite.

*Recom. 29 : Fournir la possibilité de faire varier la vitesse de lecture de la synthèse vocale*

Justification :

Certains sujets ont justifié l'abandon de l'usage de la synthèse vocale après un certain temps d'expérience, par la lenteur de la lecture, en particulier pour des messages courts très rapidement lus visuellement.

*Recom. 30 : Fournir la possibilité de supprimer des champs lus par la synthèse vocale*

Justification :

Certains utilisateurs ont apprécié la lecture d'une partie des champs contenus dans un message comme l'expéditeur et l'objet, mais par exemple trouvent la lecture de la totalité du contenu trop laborieuse ou inutile.

Média audio :

*Recom. 31 : Fournir un moyen simple de pilotage des médias audio lorsque ceux-ci sont disponibles dans l'application.*

Justification :

Lorsque l'application fait usage de medias audio, ceux-ci doivent pouvoir être pilotés facilement. Il faut également pouvoir effectuer une pause et une reprise de lecture aisée.

Sortie Audio :

*Recom. 32 : Fournir un moyen centralisé d'activation et de désactivation des sorties audio (incluant les sons et la synthèse vocale).*

Justification :

Selon le contexte (social, sonore etc.) et/ou les préférences de l'utilisateur, les sorties audio et vocales peuvent ne pas être appropriées, voire intempestives. Les sujets doivent donc pouvoir les activer et les désactiver facilement.

#### **6.4. Signifiante des codes, dénominations et comportements \***

Pas de recommandations identifiées pour ce critère.

#### **6.5. Charge de travail**

Définition :

« Le critère Charge de Travail concerne l'ensemble des éléments de l'interface qui ont un rôle dans la réduction de la charge perceptive, mnésique ou physique des utilisateurs et dans l'augmentation de l'efficacité du dialogue.

Trois sous-critères participent au critère Charge de Travail : Brièveté (qui inclut les critères Concision et Actions Minimales), Charge Physique et Densité Informationnelle. »

Justification(s) :

« Plus la charge de travail est élevée, plus grands sont les risques d'erreurs et d'incidents. De même, moins l'utilisateur sera distrait par des informations non pertinentes, plus il pourra effectuer sa tâche efficacement. Moins l'utilisateur sera gêné par les dispositifs, plus il y aura de chance pour qu'il puisse interagir correctement avec l'application. Par ailleurs, plus les actions requises seront courtes, plus rapides seront les interactions. »

##### **6.5.1. Charge physique \***

*Recom. 33 : La partie physique du dispositif doit être adaptée à une interaction à une main. La saisie du dispositif et l'appui sur les boutons physiques doivent être aisés*

Justification :

D'une manière générale, la saisie du dispositif et l'action sur les boutons ne doivent pas constituer une charge ou une préoccupation trop importante, la tenue doit non seulement inspirer confiance, mais en plus doit être assurée pour des situations d'action mouvementées.

*Recom. 34 : L'appui sur les boutons physiques doit être facilité par une conception mettant en œuvre des modalités de contrôles non visuelles*

Justification :

Les boutons physiques doivent pouvoir être facilement identifiés afin de faire intervenir le moins possible la vision. Leur texture, forme et localisation, par exemple, doivent permettre de les discriminer facilement autant que possible par les seuls sens tactiles et proprioceptifs.

#### 6.5.2. Brièveté

Définition :

« Le critère Brièveté concerne la charge de travail au niveau perceptif, mnésique et physique à la fois pour les éléments individuels d'entrée ou de sortie et les séquences d'entrées (i.e., les suites d'actions nécessaires à l'atteinte d'un but, à l'accomplissement d'une tâche). Il s'agit ici de limiter autant que possible le travail de lecture, d'entrée et les étapes par lesquelles doivent passer les utilisateurs. Deux sous-critères participent au critère Brièveté : Concision et Actions Minimales. »

Justification(s) :

« Les capacités de la mémoire à court terme sont limitées. Par conséquent, plus courtes sont les entrées, plus limités sont les risques d'erreurs. Aussi, plus les actions nécessaires à l'atteinte d'un but sont nombreuses et compliquées, plus la charge de travail augmente et par conséquent plus les risques d'erreurs sont élevés. »

##### 6.5.2.1. Actions minimales \*

Définition :

« Le critère Actions Minimales concerne la charge de travail quant aux actions nécessaires à l'atteinte d'un but, à l'accomplissement d'une tâche. Il s'agit ici de limiter autant que possible les étapes par lesquelles doivent passer les utilisateurs. »

Justification(s) :

« Plus les actions nécessaires à l'atteinte d'un but sont nombreuses et compliquées, plus la charge de travail augmente et par conséquent plus les risques d'erreurs sont élevés. »

Reconnaissance vocale :

*Recom. 35 : Éviter une gestion de déclenchement des modalités par mode d'interaction lorsque le nombre et la nature des modalités disponibles dans l'application le permettent.*

Justification :

Il est préférable de ne pas imposer aux utilisateurs, lorsque c'est possible, la gestion de modes d'interaction qui complexifie l'interaction et constitue un préalable de vérification d'activation du mode adéquat avant toute interaction.

En effet, les utilisateurs ont souvent exprimé des difficultés quant à la gestion de mode qui représentent une incertitude quant à la possibilité d'interagir avec une modalité directement sans avoir à vérifier l'activation du mode correspondant.

De plus, la gestion par mode est susceptible de provoquer des erreurs supplémentaires.

*Recom. 36 : Lorsque la gestion par mode est inévitable, celle-ci doit être la plus simple et son usage doit être le moins visuel possible en entrée comme en sortie. Elle doit également être accessible via plusieurs modalités si possible.*

Justification :

La vérification du mode en cours (sortie du système), le changement de mode si nécessaire (entrée du système) et la vérification du changement de mode sont des tâches secondaires qui alourdissent le

processus d'interaction et peuvent mobiliser des ressources attentionnelles de manière non appropriée (en particulier pour les modalités peu ou pas visuelles).

*Recom. 37 : Automatiser ou éliminer les tâches au maximum, en particulier dans un contexte de mobilité*

Justification :

Moins l'utilisateur a d'actions à effectuer et moins ses ressources cognitives et attentionnelles sont sollicitées.

Exemples :

Optimiser l'automatisation des sauts de curseurs correspondants aux séquences d'actions (automatisation des actions) (Oulasvirta, 2005)

Lorsque la tâche consiste à saisir des valeurs de géolocalisation associées à une donnée quelconque, automatiser la capture de la position par le système (Pascoe et al., 2000)

Mise en veille :

*Recom. 38 : La mise en veille ne doit pas conduire l'utilisateur à faire des actions supplémentaires.*

Justification :

La mise en veille du dispositif en réinitialisant l'application peut obliger l'utilisateur à recommencer des étapes qu'il avait déjà réalisées auparavant.

*Recom. 39 : La mise en veille dans la mesure du possible (et au moins dans un premier temps après une commande utilisateur) ne devrait pas être opaque afin de ne pas masquer l'état des applications.*

Justification :

Dans un contexte d'activité multitâche, l'utilisateur ne peut pas toujours effectuer la totalité des interactions d'une seule traite. Dans ces cas, la mise en veille totalement opaque nuit à la continuité de l'interaction et oblige l'utilisateur à effectuer une remise en marche de l'application.

#### **6.5.2.2. Concision \***

Définition :

« Le critère Concision concerne la charge de travail au niveau perceptif et mnésique pour ce qui est des éléments individuels d'entrée ou de sortie. Par convention, la Concision ne concerne pas le feedback ni les messages d'erreurs. »

Justification(s) :

« Les capacités de la mémoire à court terme sont limitées. Par conséquent, plus courtes sont les entrées, plus limités sont les risques d'erreurs. »

*Recom. 40 : Raccourcir les unités d'interaction en petits épisodes afin de libérer les ressources attentionnelles*

Justification:

En découpant les séquences d'interaction en petites unités, ou sous forme d'épisodes, les utilisateurs peuvent fractionner leur attention ce qui facilite l'accomplissement d'éventuelles tâches parallèles (Oulasvirta, 2005).

*Recom. 41 : La sémantique des commandes vocales doit être simple et utiliser dans la mesure du possible des mots ou combinaisons de mots courts*

Justification :

Les commandes vocales complexes (ou les combinaisons de mots) sont difficiles à mémoriser, de plus elles se prêtent à des variations de vocabulaire qui peuvent provoquer des erreurs de reconnaissance par le système.

### 6.5.3. Densité informationnelle \*

Définition :

« Le critère Densité Informationnelle concerne la charge de travail du point de vue perceptif et mnésique, pour des ensembles d'éléments et non pour des items. »

Justification(s) :

« Dans la plupart des tâches, la performance des utilisateurs est influencée négativement quand la charge informationnelle est trop élevée ou trop faible. La probabilité d'erreur augmente. Il faut donc supprimer les éléments sans lien avec le contenu de la tâche en cours.

Il faut aussi éviter d'imposer à l'utilisateur la mémorisation de longues et nombreuses informations ou procédures (la mémoire à court terme est limitée), ou toute activité nécessitant de sa part la mise en œuvre d'activités cognitives complexes lorsque la tâche ne le requiert pas. »

*Recom. 42 : Le moyen de pilotage de la synthèse vocale doit mobiliser le moins possible les ressources visuelles.*

Justification :

La synthèse vocale étant entre autre un moyen de libérer l'attention visuelle, son pilotage doit également être à attention minimale.

*Recom. 43 : Le moyen de pilotage des médias audio (du moins non visuels) doit mobiliser le moins possible les ressources visuelles.*

Justification :

Les médias audio peuvent permettre à l'utilisateur une interaction non visuelle. Les moyens de pilotage doivent également être le moins visuels possible, en particulier dans le cas où le contenu des médias concerne une tâche extérieure à l'interaction.

*Recom. 44 : Le moyen d'activation et de désactivation des sorties audio (du moins non visuelles) doit mobiliser le moins possible les ressources visuelles.*

Justification :

L'audio étant par nature non visuel, il est préférable que son moyen de contrôle le soit également.

## 6.6. Adaptabilité

### 6.6.1. Prise en compte de l'expérience de l'utilisateur \*

*Recom. 45 : Fournir la possibilité de diminuer le nombre de feedbacks ou d'incitations et la qualité de ceux-ci pour chaque étape de l'interaction en fonction de l'expérience des utilisateurs*

Justification :

Après un temps d'apprentissage, la présence de trop nombreux feedbacks ou une incitation trop présente peut ralentir le processus d'interaction et il convient de fournir aux utilisateurs la possibilité de les supprimer. Par contre, après une période importante de non-utilisation du système il conviendra de pouvoir réinitialiser les feedbacks, l'incitation et le niveau de guidage initial de l'application afin de faciliter un ré-apprentissage du système par l'utilisateur en s'appuyant sur un mécanisme de rappel des différentes fonctions et de leur usage.

### 6.6.2. Flexibilité \*

*Recom. 46 : Autant que possible, fournir la possibilité d'un choix de modalités d'interaction pour l'ensemble des commandes disponibles (principe d'équivalence fonctionnelle maximale)*

Justification :

Les utilisateurs ont des préférences personnelles d'usage des modalités et il convient de leur laisser le choix de la modalité d'interaction. De plus des changements de contexte peuvent les conduire à avoir recours temporairement ou ponctuellement à une autre modalité.

*Recom. 47 : Fournir aux utilisateurs un moyen facile pour changer de modalité d'interaction (gestion de mode)*

Justification :

Certains utilisateurs spécialisent l'usage d'une modalité pour certaines commandes et ont donc recours à plusieurs modalités dans une séquence d'interaction. D'autres ont des préférences fixes mais des événements peuvent les pousser ponctuellement à en changer. Le passage d'une modalité à l'autre doit donc être aisé.

*Recom. 48 : Fournir un ou plusieurs moyens de pilotage non visuels des médias non-visuels (synthèse vocale et messages audio etc.)*

Justification :

Les médias non visuels permettent de libérer le regard et leur pilotage doit conserver cette propriété en proposant une interface adaptée au contexte. Il peut s'agir par exemple de boutons soft plus gros que l'on active au doigt pour piloter un contenu audio sans avoir à sortir le stylet. Cela peut également être des boutons hard dédiés contextuellement au pilotage des médias.

### 6.7. Homogénéité/Cohérence \*

Définition :

« Le critère Homogénéité/Cohérence se réfère à la façon avec laquelle les choix de conception de l'interface (dispositifs, modalités, codes, dénominations, formats, procédures, comportements, etc.) sont conservés pour des contextes identiques, et sont différents pour des contextes différents. »

*Recom. 49 : La sémantique des textes et labels de l'interface doit être identique à la sémantique utilisées pour toutes les modalités (en particulier pour les commandes vocales).*

Justification :

Selon le contexte (social, sonore etc.) et/ou les préférences de l'utilisateur, la synthèse vocale peut ne pas être appropriée, voire intempestive. Les sujets doivent donc pouvoir l'activer et la désactiver facilement.

*Recom. 50 : Dans la mesure du possible, veiller à la possibilité d'effectuer une séquence de commandes avec une seule modalité*

Justification :

Les utilisateurs ont tendance à vouloir conserver la modalité en cours d'usage et il convient de ne pas les obliger à en changer dans le cas où une commande n'est disponible que via une seule autre modalité (absence d'équivalence fonctionnelle).

*Recom. 51 : De manière générale, l'interface et en particulier les commandes et objets interactifs doivent être disposés de manière stable et prévisible*

Justification :

Durant des interactions mobiles, les sujets peuvent anticiper ou préparer une action sur une commande, par exemple en amorçant le positionnement du moyen de contrôle de manière non visuelle afin d'être prêt à effectuer la commande visuelle en peu de temps lorsque le sujet dispose d'un créneau de faible charge de travail. Si les moyens de contrôle de l'application varient trop, la tâche d'identification des contrôles risque d'alourdir le processus d'interaction.

*Recom. 52 : Les comportements et les fonctions associées aux boutons contextuels, hard ou soft doivent être stables, clairement identifiés et prévisibles*

Justification :

De part leur définition, les boutons contextuels ont vocation à changer de fonction selon le contexte, pour limiter le besoin de ressources attentionnelles liées à leur usage, il convient de porter une attention particulière à leur conception.

## **6.8. Gestion des erreurs**

### 6.8.1. Protection contre les erreurs \*

*Recom. 53 : Les commandes non utiles à la réalisation d'une tâche doivent être inactives afin de prévenir leur utilisation*

Justification :

La correction d'erreur a un impact parfois lourd sur la charge attentionnelle. La prévention d'erreurs permet une interaction plus fluide et ne nécessitant pas de ressources attentionnelles supplémentaires.

### 6.8.2. Qualité des messages d'erreur \*

*Recom. 54 : Fournir des messages d'erreur mobilisant les mêmes sens et modalités que la modalité d'entrée utilisée pour la commande (seulement si ces modes ont été activés par l'utilisateur).*

Justification :

Nous avons observé que la correction d'erreur suite à l'usage d'une modalité non visuelle est souvent suivie d'une correction non visuelle également afin d'assurer la continuité de la tâche en cours. Les messages d'erreurs doivent donc favoriser cette continuité en reprenant au moins la même modalité que celle causant l'erreur.

*Recom. 55 : Les messages d'erreur non visuels doivent indiquer la nature de l'erreur afin d'en faciliter la compréhension et la correction sans solliciter la vision*

Justification :

Cette recommandation, couplée à la précédente, permet d'assurer un mode non visuel de correction des erreurs.

Par exemple un message d'erreur de reconnaissance doit indiquer la nature de l'erreur de manière vocale : « erreur de reconnaissance », « commande impossible » etc.

### 6.8.3. Correction des erreurs \*

*Recom. 56 : Toujours permettre, à minima, la correction d'une erreur à l'aide de la même modalité que la commande ayant provoqué l'erreur*

Justification :

Nous avons observé que la correction d'erreur suite à l'usage d'une modalité non visuelle est souvent suivie d'une correction non visuelle également afin d'assurer la continuité de la tâche en cours. La

correction des d'erreurs doit donc favoriser cette continuité en permettant au moins l'usage de la même modalité que celle causant l'erreur.

*Recom. 57 : Permettre à l'utilisateur de choisir la modalité de correction d'une erreur*

Justification :

Les utilisateurs changent parfois de modalité pour corriger une erreur et il convient donc de ne pas les obliger à n'utiliser qu'une seule modalité.

**Chapitre 8. Conclusion Générale et perspectives**

## 1. Introduction

L'étude réalisée dans le cadre de cette thèse s'inscrit dans la continuité d'un premier travail de DEA mené sur l'utilisation d'un système portable multimodal limité à un contexte de laboratoire (Calvet et al., 2001).

Notre objectif était de documenter la question de l'usage de dispositifs d'interaction mobiles en environnement réel, de tester les effets potentiels de la mobilité sur l'usage d'un système interactif en situation, et d'étudier empiriquement en quoi la versatilité des systèmes multimodaux, souvent présentée comme un de leurs avantages principaux, peut aider l'utilisateur à gérer les contraintes induites par les changements de contexte inhérents à la mobilité (modification de l'environnement physique et social notamment). Nous voulions également vérifier empiriquement l'intérêt de cette versatilité, également présentée comme un moyen utile de pallier aux erreurs d'interaction (soit liées au système, ou à l'utilisateur) tant en prévention qu'en correction, en permettant aux utilisateurs de choisir la modalité d'interaction la plus adaptée à la situation.

Cette perspective d'étude nous a orienté vers deux problématiques principales. La première, d'ordre théorique consistait à étudier les usages de dispositifs multimodaux mobiles, afin de tenter de produire un ensemble de connaissances potentiellement utiles pour la conception.

Concernant les aspects théoriques de l'interaction multimodale en mobilité, nous avons posé que celle-ci constituait un nouveau paradigme d'interaction et qu'elle ne pouvait-être rendue possible qu'à la condition que les styles d'interaction proposés puissent s'adapter à la tâche principale des utilisateurs. En effet, en situation de mobilité, les utilisateurs peuvent être soumis à des contraintes très fortes pouvant limiter le champ des interactions possibles, voire les rendre impossibles. Un constat que l'on peut en effet faire à propos de l'interaction en mobilité est qu'elle est par nature une activité multitâches qui provoque de fait une allocation des ressources physiques, perceptives et cognitives plus ou moins volontairement contrôlée (Oulasvirta, 2005). En particulier, l'allocation de ressources attentionnelles devient critique lorsqu'il s'agit d'effectuer des interactions en mobilité « réelle », quand l'utilisateur doit au minimum gérer son propre déplacement et l'interaction avec l'artefact proprement dit. Nous avons pointé deux pistes d'études complémentaires liées à la question de la régulation de la compétition pour les ressources attentionnelles générée par la mobilité. Il s'agit de la multimodalité et des applications sensibles au contexte (Context Aware Computing ou CAC) qui poursuivent en partie un but commun de réduction des ressources attentionnelles nécessaires à l'interaction mais avec des moyens différents : la réduction ou l'automatisation de tâches pour le CAC, la possibilité d'adaptation à la situation pour l'utilisateur pour la multimodalité.

La thématization de cette problématique nécessite donc de considérer conjointement des questions relatives à l'articulation entre multimodalité, mobilité, contexte et attention. Le traitement théorique et technologique de cette articulation constitue un champ propice au développement de nouvelles études.

La deuxième problématique, d'ordre méthodologique, visait à réfléchir aux techniques de recueil de ces données d'usages. En effet, les situations de mobilité introduisent des questionnements complexes sur les dispositifs à mettre en œuvre pour observer ces nouvelles situations d'usage, à la fois d'un point de vue quantitatif, mais également et surtout selon nous, d'un point de vue plus qualitatif afin d'en dégager les significations pour les acteurs et les déterminants de leur activité.

Dans cet esprit, nous avons donc présenté un dispositif général d'étude et de recueil des activités mobiles en situation naturelle, mis en œuvre à l'occasion d'une étude empirique menée dans le contexte de la plateforme d'évaluation des usages LUTIN à la Cité des Sciences et de l'Industrie de Paris-La Villette. Cette méthode, qui reposait sur un dispositif d'enregistrement de données en situation naturelle et de recueil de verbalisations consécutives à l'activité assisté par les traces, a été appliquée à une situation de jeu collaboratif en mobilité utilisant comme ressources les objets de l'environnement présents dans les expositions ouvertes au public.

## 2. Rappel des résultats et apports de la thèse

### 2.1. Résultats empiriques : documentation des usages de dispositifs multimodaux en mobilité

- Usages, préférences d'usages et déterminants contextuels

D'un point de vue quantitatif, nous avons pointé l'intérêt que ces données représentent afin d'appréhender des tendances générales d'usage de la multimodalité et le poids d'un certain nombre de déterminants contextuels. En effet, nous avons pu montrer de manière chiffrée des tendances d'usage ou de non usage de modalité avec une répartition globalement favorable à la modalité tactile et un usage moins important de la modalité vocale, avec des influences relativement peu marquées sur les choix d'usage, de déterminants tels que le contexte sonore ou social qui ne permettent pas de tirer de conclusions générales sur leur importance (toujours d'un point de vue quantitatif).

Le raffinement des analyses quantitatives au niveau individuel, nous a également permis de pointer le risque qu'il peut y avoir à vouloir exploiter des résultats globaux et d'en tirer des conclusions hâtives en raison des risques de masquage de variabilité ou d'artefacts liés au moyennage des données. En effet, en accord avec les résultats d'études d'autres chercheurs (Mignot, 1995), ou de ceux que nous avons obtenus en situation de laboratoire (Calvet, 2001), nous avons observé des variabilités inter-individuelles très importantes dans les choix d'usage des modalités, variations évidemment impossibles à identifier si l'on s'en tient à un niveau global d'analyse. De plus, à ces variations inter individuelles viennent se superposer des variations intra-individuelles plus ou moins marquées difficiles à expliquer si l'on s'en tient à un niveau quantitatif de restitution et d'analyse des données.

Nous avons ainsi pu observer chez de nombreux sujets des phénomènes marqués de préférences individuelles d'association pour une modalité, ou pour l'association plus ou moins récurrente d'une modalité à une commande (lorsque les taux de préférence vont au delà de 80%, nous l'avons nommée « spécialisation », qui est une sorte d'autoassignation au sens de (Nigay & Coutaz, 1995)), mais sans que cela ne soit forcément partagé par l'ensemble des sujets. Ce résultat quantitatif n'a pas pu trouver d'explication au niveau qualitatif via les entretiens d'autoconfrontation ou via une analyse fine des propriétés des commandes. Le fait que les sujets ne se spécialisent pas sur les mêmes commandes et les mêmes modalités semble plutôt indiquer des facteurs de choix non liés aux caractéristiques intrinsèques des commandes pour lesquelles une modalité serait éventuellement plus appropriée, et donc l'explication pourrait plutôt tendre vers des préférences individuelles (idiosyncrasies) propres à chaque sujet.

Les données globales, même si elle permettent un éclairage intéressant et une forme de restitution synthétique des résultats d'observation, ont donc un faible caractère explicatif. Elles ont surtout été exploitées en tant qu'éléments de guidage vers des questionnements plus qualitatifs.

Ainsi, à travers l'analyse des verbalisations, nous avons pu faire apparaître de nombreux déterminants expliquant en partie les préférences individuelles et l'adoption par les sujets des modalités. Cette analyse montre notamment que ces préférences sont liées en premier lieu aux propriétés intrinsèques des modalités qui vont en partie déterminer l'expérience individuelle d'usage des sujets, mais surtout, elles indiquent le caractère fortement subjectif de la perception de ces propriétés (par exemple un sujet peut percevoir la modalité vocale comme particulièrement efficace et un autre sujet à l'inverse, peut préférer la modalité tactile pour la même raison). Les sujets vont donc se forger au fil du temps des préférences d'usages (incluant à la fois des facteurs internes et externes) qui vont modeler leur style d'interaction propre expliquant ainsi les variations inter-individuelles et les spécialisations observées.

Les verbalisations font également ressortir l'importance du poids de facteurs externes - que la littérature IHM pourrait qualifier classiquement de « contexte » (Dey et al., 2001), et que nous qualifierons plutôt de situation, au sens de (Quéré, 1997) afin de restituer les dimensions « engagement de l'acteur » et historique de l'activité - pour expliquer les choix d'usage des modalités. De manière générale, nous avons pu observer une certaine stabilité des choix d'usage de modalités en fonctions de préférences

individuelles qui constituent des heuristiques privilégiées, des séquences d'actions intégrées mettant en œuvre des mécanismes subjectivement peu coûteux pour les situations d'usage courant. En revanche, lors de variations –souvent externes (mais parfois internes comme par exemple la fatigue)- de la situation, les sujets justifient des changements de modalités, ou des déviations de leurs heuristiques au prix d'efforts cognitifs ou physiques supplémentaires. Nous avons également montré que les déterminants de la situation justifiant les changements d'heuristiques sont très nombreux, de nature complexe, fortement situés et subjectifs (préférences, nature de la tâche, contexte social, contexte de mobilité, historique etc.). Ils relèvent donc plus d'un paradigme de la complexité, non déterministe ou la situation se construit, au point de faire « émerger » naturellement une modalité d'interaction privilégiée, expliquant également en partie les variations d'usages intra-individuels. Nous pourrions aller jusqu'à affirmer que la situation, construite et perçue en partie par les sujets va « afforder », au niveau subjectif, une modalité privilégiée d'interaction<sup>5</sup>.

- Usages de la multimodalité en mobilité et conséquences sur les ressources attentionnelles

En ce qui concerne la deuxième piste d'allègement des ressources attentionnelles en situation de mobilité, nous avons montré, tant d'un point de vue quantitatif que qualitatif, que la multimodalité, permettait bien, par la possibilité du choix de la modalité d'interaction, un ajustement approprié à la situation en particulier en terme d'allocation des ressources attentionnelles. Là encore, le choix d'une modalité appropriée à la situation de mobilité ne relève pas d'un processus déterministe partagé par l'ensemble des sujets, mais d'un choix complexe, guidé en premier lieu par les préférences individuelles. En effet, certains sujets, reconnaissant pourtant la supériorité de la modalité vocale en tant que mode facilitant un meilleur degré de mobilité, préfèrent tout de même continuer à avoir recours à une interaction privilégiée au stylet, parfois au prix de l'arrêt de la mobilité ou de la mise en œuvre de stratégie coûteuses au niveau physique ou cognitif (stratégie d'anticipation, interaction et déplacement avec le regard orienté vers le bas, ralentissement etc.).

Pour les sujets ayant plus recours à la modalité vocale, nous avons mis en évidence un véritable effet d'allègement des ressources attentionnelles (regard moins orienté vers l'interface) favorisant une mobilité plus fluide et moins interrompue, y compris pour les situations de corrections d'erreurs (allant jusqu'à des corrections faites en vocal plus mobiles que les commandes précédentes).

A l'inverse, nous avons montré que l'usage de la modalité tactile stylet, fortement consommatrice en ressources visuelles, oblige les sujets à se focaliser sur l'interaction au détriment de leur expérience avec l'environnement et donc également au détriment de la mobilité.

---

<sup>5</sup> Il convient néanmoins d'utiliser ici le terme d'affordance avec prudence, suivant en cela les avertissements de Norman lui-même...

### **2.2. Résultats empiriques : apports méthodologiques**

D'un point de vue méthodologique, notre questionnement s'est articulé autour de trois problématiques principales correspondant à notre approche théorique prenant en compte l'aspect complexe, situé et construit des situations d'interaction exigeant de fait une étude des usages en situation naturelle. Il s'agissait donc pour cela de :

- Proposer aux sujets une expérimentation en contexte naturel
- Construire un dispositif mobile de recueil des données et autoconfrontations
- Concevoir un dispositif de codage et d'analyse de données variées et complexes

La construction et le déroulement de notre expérimentation a permis de mettre à l'épreuve nos concepts et développements selon ses trois axes.

- Expérimentation en contexte naturel

Notre volonté de situer, autant que possible, l'expérimentation en contexte naturel s'est traduite par l'élaboration d'un protocole expérimental, d'une tâche à réaliser et de consignes appropriées qui se sont avérés relativement efficaces au regard de nos objectifs.

Le choix d'une tâche de réalisation d'un jeu collaboratif a permis une certaine immersion et implication des sujets de part son caractère ludique et social mettant en œuvre des échanges informels entre les sujets.

De même, l'objectif de provocation de phases alternées de tâches d'interaction et d'exploration visuelle de l'environnement en mobilité a été atteint grâce à l'emploi d'un jeu (de type quiz) mettant en œuvre à la fois des ressources situées dans l'environnement réel des utilisateurs et des ressources collaboratives médiées par l'application multimodale.

Le parti pris de non imposition de consignes strictes, laissant le choix des modalités et des méthodes de réalisations de la tâche aux sujets, a permis d'observer des usages plus naturels et une adoption d'heuristiques moins biaisées.

- Dispositif mobile de recueil des données et autoconfrontations

La mise au point d'un dispositif mobile de recueil de donnée s'est avéré être un challenge technique compliqué mais qui s'est soldé par des résultats extrêmement riches et pertinents impossibles à récolter à l'aide de méthodes classiques utilisées en situations fixes. En effet, l'usage de caméra lunettes fournissant une vue dite subjective, couplée à une vue externe plus classique et à une récolte de fichiers logs, a permis de répondre aux contraintes liées à notre approche non déterministe et située de l'interaction. Cette capture large du contexte, et l'enregistrement local des données en temps réel sur un disque de stockage multimédia nous a également permis de procéder à des autoconfrontations « à chaud » des sujets, sur la base des vidéos, en particulier la vue subjective qui aide le sujet à se replacer dans le contexte de son activité grâce à une vision centrée sur l'interaction et avec moins de biais liés à une analyse reconstruite a posteriori via des vues plus larges et moins précises (les vues larges dites « contextuelles » mettent en scène le sujet plutôt d'un point de vue de tierce personne, et l'analyse peut s'orienter différemment). Le même dispositif nous a également permis, de manière commode, d'enregistrer en vidéo les autoconfrontations.

- Dispositif de codage et d'analyse de données variées et complexes

En ce qui concerne le codage des différentes sources de données (vue subjective, contextuelle et fichiers logs), l'utilisation du logiciel Actogram s'est avérée indispensable, entre autre grâce aux modifications apportées qui permettent de visionner et d'annoter plusieurs vues simultanément, mais également grâce à ses ajouts de capacités de croisement, d'import et d'export de données sous des formats variés.

Cet outil a également facilité l'articulation des données quantitatives et qualitatives, dépassant les limites des données globales permettant un éclairage intéressant, mais ayant un faible caractère explicatif, grâce à un codage fin des interactions lié à l'enrichissement que peuvent fournir entre autres les verbalisations.

### **2.3. Résultats théoriques**

Nous avons formulé une première proposition de modèle détaillé de l'interaction, mettant en œuvre une analyse détaillée, voir micro, de l'activité de l'utilisateur. En effet, en situation de mobilité, et en particulier dans le cas de l'autodéplacement, l'échelle pertinente peut souvent relever de détails ou de délais très réduits faisant la différence dans la complexité de l'interaction et dans la compétition du partage des ressources attentionnelles.

Nous avons introduit, en ce sens les notions de modalité de contrôle et de repères relatifs et absolus qui nous semblent essentielles à prendre en compte pour la phase ultérieure de conception d'interfaces multimodales mobiles.

### **2.4. Résultats technologiques**

En nous basant à la fois sur l'état de l'art, les résultats de notre étude et les modèles que nous avons produits, nous avons proposé un ensemble de critères et de recommandations pour la conception et l'évaluation d'interfaces à attention minimale.

Nous avons également montré l'intérêt de privilégier le principe d'équivalence fonctionnelle maximale pour l'implémentation d'applications multimodales, qui permet aux sujets d'effectuer des séquences complètes de commandes à l'aide d'une ou plusieurs modalités indifféremment et ainsi de libérer potentiellement le regard via une interaction non visuelle.

### 3. Perspectives

#### 3.1. Conception d'applications sensibles au contexte (*Contexte Aware Computing*)

La piste « applications sensibles au contexte » qui permet d'alléger le coût des ressources attentionnelles lié aux interactions en mobilité nécessite d'approfondir nos connaissances et nos approches théoriques de conception de telles applications.

Il nous semble notamment nécessaire de porter une attention particulière aux relations qui se tissent entre contextualisation et incidence sur l'identification des choix de modalités d'interaction Homme-homme ou homme-machine. Il s'agit de suivre les conséquences d'une première rupture avec une modélisation de la communication en termes d'échange d'information : le « contexte » n'est pas une ressource stable et périphérique, une entité préexistante, que l'analyste ou le sujet met en relation avec des contenus - qu'il s'agisse de constater l'altération de ces contenus à son contact (le « contexte » comme source de bruit, comme contrainte), ou les informations qu'il permet de leur adjoindre (le « contexte » comme source de désambiguïsation, comme ressource). Le « contexte », puisqu'il s'agit dans un premier temps de conserver ce vocable quelque peu encombrant (Quéré, 1997), n'est pas ce qui s'adjoint au « texte » ou à la communication, mais ce qui fait le « texte » et ce qui est fait par le « texte ».

Contre un usage quelque peu positiviste de la notion de « contexte », l'ethnographie de la communication et l'analyse conversationnelle, entre autres, se sont employées à décrire la nature dynamique et située de la recherche et de la construction de « contextes » par des acteurs en interaction. Il est intéressant de noter le déplacement opéré par ces travaux qui préfèrent la notion de « contextualisation » à celle de « contexte ». Le « contexte » n'est plus alors une boîte noire, que l'analyste recompose en identifiant les traits ou les caractéristiques qu'il juge pertinents. Le « contexte » devient un objet de recherche ordinaire ; la notion de « contextualisation » fait alors référence à cet ensemble d'heuristiques dont les tenants et les aboutissants sont étroitement liés au caractère social de nos interactions.

En effet, ce parti pris constructiviste se voit confirmé par l'ensemble des résultats que nous avons obtenus dans ce travail, tant d'un point de vue quantitatif que qualitatif.

Cette dimension a été également soulignée par (Oulasvirta, Tamminen, & Höök, 2005a). Pour ces auteurs le caractère non déterministe des déterminants du contexte ou de la situation, est un enjeu scientifique débattu de longue date opposant les approches positivistes (ou réalistes) et les approches constructivistes. La reconnaissance du caractère complexe des notions de situation et de contexte rend donc théoriquement caduques toutes tentatives de reconnaissance, d'interprétation du contexte, et d'automatisation d'actions si celui-ci ne s'est pas construit en accord avec la perception qu'en ont les sujets.

De ce point de vue, il serait intéressant de s'inspirer de certaines notions développées par l'ethnolinguiste John Gumperz (Gumperz, 1989) qui a développé un vocabulaire d'enquêteur et d'historien pour rendre compte de nos heuristiques ordinaires. Il parle ainsi d'indices de contextualisation (*contextualization clues*), indices dont la découverte permet de générer des inférences et de produire une ou plusieurs interprétations. Nous pourrions ainsi définir le « contexte » comme l'ensemble des éléments pertinents permettant d'élaborer une ou plusieurs interprétations sur la situation en cours. Dans le cas d'applications sensibles au contexte, la notion d'indices de contextualisation pourrait s'avérer utile dans un processus interactif de construction du contexte mutuellement partagé entre le sujet et la machine. Comme le suggèrent également (Oulasvirta et al., 2005a), la notion de contexte est souvent vue par les réalistes comme ce que (Quéré, 1997) qualifie « d'environnement » qui, de manière simplifiée, relève de tout élément qui est « détectable » et interprétable de manière relativement simple comme l'heure, la position, le contexte spatial relatif, etc. Les comportements automatiques issus de ces interprétations sont souvent perçus comme problématiques car rarement opportuns, souvent intrusifs, et venant potentiellement perturber les tâches en cours. Le courant constructiviste optera plutôt pour une approche de gestion et de

construction des ressources (dont les données détectables), facilité et préparée par une reconnaissance de la situation construite en accord avec l'utilisateur.

Une autre approche issue des sciences humaines pourrait être transposée aux interactions hommes-machines sensibles à la situation pour la construction d'un contexte partagé. En effet, le vocabulaire de la scène et du jeu utilisé à des fins analytiques par Erving Goffman (Goffman, 1973, 1987) permet de décrire l'engagement et le risque propre à la production des inférences et des interprétations sur le contexte en cours. Pour lui, nous mettons en jeu au cours de nos interactions une interprétation de la situation en cours, avec ses erreurs, ses régulations etc. Plus une action est référée à des inférences, plus son auteur prend le risque de manquer son but, de susciter de l'incompréhension ou encore d'aboutir à un malentendu. A contrario, la possibilité de faire référence à un contexte partagé par nos actions célèbre l'existence d'un lien social et culturel, ce que l'on appelle ordinairement, selon le degré de partage, la connivence ou encore la complicité (ou « mutual awareness » dans le champs du CSCW).

La perspective du développement d'applications sensibles au contexte et à la situation semble donc prometteuse pour alléger la tâche d'utilisateurs aussi bien en situation fixe qu'en situation de mobilité, mais le chemin à parcourir en suivant une approche complexe semble encore à défricher et relèvera d'une conception obligatoirement centrée sur l'utilisateur et sa relation à la situation et au contexte.

### **3.2. Développement d'applications mobiles à ressources attentionnelles limitées**

La piste de développement d'applications multimodales favorisant l'allègement des ressources attentionnelles et donc la mobilité étant avérée, il conviendrait de tester les propositions de conceptions avancées dans les modèles que nous avons produits ainsi que les principes avancés.

Il s'agirait dans un premier temps de proposer des applications mettant en œuvre le concept d'équivalence fonctionnelle maximale permettant des séquences ininterrompues d'usage d'une même modalité, sans contraintes de choix pour les utilisateurs et permettant éventuellement des séquences complètes d'interaction non visuelle.

Suivant cette même approche, il serait prometteur d'axer les recherches vers le développement d'applications mettant en œuvre un choix élargi de modalités, et en particulier des modalités non visuelles telles que la modalité tangible que nous avons étudiée en situation de laboratoire (Calvet, 2001) et qui s'était avérée très prometteuse en terme d'adoption par les sujets.

Nous avons également pointé le paradoxe d'usage pour les applications mobiles du style d'interaction dominant que constitue le paradigme de la manipulation directe, fortement basée sur une interaction visuelle. Il conviendrait donc soit d'étudier des adaptations du style de manipulation directe en mobilité, soit d'en étudier des alternatives spécifiquement conçues pour la mobilité.

### **3.3. Validation des critères et recommandations**

Les critères et recommandations que nous avons proposés ne constituaient pas l'objet de cette recherche et en cela, ils ne constituent qu'une base de propositions. Ces critères seront donc à mettre à l'épreuve d'une conception et d'une évaluation scientifique pour être validés et enrichis.

### **3.4. Perspectives méthodologiques**

La volonté de tenir dans le même mouvement des objectifs contradictoires (contrôle expérimental strict et préservation du réalisme des situations visées) conduit à construire des dispositifs pour le moins surprenant. Ainsi les tentatives faites pour reproduire certaines propriétés des situations de mobilité en laboratoire donnent-elles naissance à des situations totalement artificielles (Barnard et al., 2005) (Pirhonen et al., 2002).

Outre ce premier type de reproche, cette démarche s'expose également à une objection qui a trait à l'appauvrissement de ce qui est évalué, et qui entraîne une limitation de la portée des résultats obtenus. Deux éléments (au moins) sont ici en cause :

- le caractère limité du type de données recueillies
- et la restriction temporelle de l'évaluation au moment circonscrit par le temps de la passation expérimentale

### 3.4.1. De l'utilisabilité à l'expérience utilisateur

La première limitation peut être traitée en recourant à des modalités de recueil qui permettent d'obtenir des données empiriques plus riches : les données comportementales synthétiques quantitatives (telles que temps de réalisation, nombre d'erreurs,...) prises comme indices de performance doivent être ainsi complétées par des données de verbalisation permettant d'avoir un accès à la signification que donnent les sujets à leur activité. Cette approche permet de documenter rigoureusement et systématiquement des dimensions qui échappent habituellement aux méthodes classiques basées uniquement sur des approches en « troisième personne ». Elle permet de dépasser le paradoxe de certaines évaluations qui semblent démontrer qu'une innovation n'apporte rien sur des critères de performance quantitatifs mais qui dans le même temps apporte une « satisfaction » plus importante aux utilisateurs<sup>6</sup> (reste bien évidemment à argumenter concrètement ce que recouvre la notion de « satisfaction » trop souvent utilisée de manière commode mais floue dans les échelles du même nom).

Cet élargissement des dimensions d'analyse habituellement utilisées dans le champ des études sur l'utilisabilité des dispositifs techniques s'est récemment traduit par l'apparition du label « new-usability » (Thomas, 2002) et par l'émergence d'un champ d'études centré sur « l'expérience de l'utilisateur » qui intègre des composantes cognitives, émotionnelles et sociales (Hassenzahl & Tractinsky, 2006).

### 3.4.2. Etude de l'appropriation des applications dans le temps

La seconde limitation nécessite une reconsidération drastique de la nature de l'évaluation même. En effet malgré la présence de phases organisées de familiarisation préalables aux sessions expérimentales d'évaluation qui ont pour objectif de dépasser (illusoirement sans doute) les effets produits par la nouveauté du dispositif et de ses modes opératoires associés, les tests d'utilisabilité ne fournissent pas le recul temporel nécessaire à l'appropriation du système par les agents et ne constituent donc que des instantanés qui ne peuvent rendre compte des évolutions possibles dans la mise en œuvre des artefacts en situation réelle d'utilisation.

Cette problématique de l'appropriation est au cœur des travaux sur les usages qui se sont développés avec le déploiement récent des technologies de l'information et de la communication dans des contextes variés et qui favorisent l'interpénétration des types d'activité (professionnelle et privée). L'étude approfondie de l'appropriation à plus long terme implique la mise en œuvre de dispositifs de recueils de données innovants qui doivent s'appliquer aussi bien à des données de nature quantitatives que

---

<sup>6</sup> Voir le cas bien connu de la visio-conférence par exemple.

## Conclusion Générale et perspectives

qualitatifs à travers par exemple, l'usage de sondes telles que celles utilisées par (Demunieux & Losquin, 2005), de fichiers logs enregistrés sur le réseau, ainsi que l'usage de logbooks permettant aux utilisateurs de rendre compte de la nature subjective de leur expérience.

## Chapitre 9. Bibliographie et annexes

## 1. Bibliographie

- Abowd, G. D., Atkeson, C. G., Hong, J., Long, S., Kooper, R., & Pinkerton, M. (1997). Cyberguide: A mobile context-aware tour guide. *ACM Wireless Networks*, 5(3), 421-433.
- Anastopoulou, S., Baber, C., & Sharples, M. (2001, September 26th-27th). *Multimedia and multimodal systems : commonalities and differences*. Paper presented at the 5th Human Centred Technology Postgraduate Workshop, University of Sussex.
- Anhalt, J., Smailagic, A., P.Siewiorek, D., Gemperle, F., Salber, D., Weber, S., et al. (2001). Toward Context-Aware Computing: Experiences and Lessons. *IEEE Intelligent systems*, 16(3), 38-46.
- Arbib, M. A. (1993). Schemas for high-level vision: the problem of instantiation. In *Computational neuroscience* (pp. 340-351): MIT Press.
- Arroyo, E., & Selker, T. (2000). Self-adaptive multimodal interruption interfaces.
- Bach, C. (2004). *Elaboration et validation de Critères Ergonomiques pour les Interactions Homme-Environnements Virtuels*. Université de Metz.
- Bach, C., & Scapin, D. L. (2005). *Critères ergonomiques pour les Interactions Homme Environnement Virtuel: définitions, justifications et exemples* (Rapport de recherche No. 5531). Rocquencourt: INRIAo. Document Number)
- Barnard, L., Ji Soo, Y., Jacko, J. A., & Sears, A. (2005). An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices. *International Journal of Human-Computer Studies*, 62(4), 487-520.
- Bastien, J. M. C., & Scapin, D. L. (1993). *Ergonomic criteria for the evaluation of human-computer interfaces* (Technical report N° 156). Rocquencourt, France: Institut National de Recherche en Informatique et en Automatiqueo. Document Number)
- Baudel, T. (1991). *Spécificité de l'interaction gestuelle dans un environnement multimodal*. Paper presented at the IHM'91, LRI Paris sud.
- Baudel, T., & Braffort, A. (1993). *Reconnaissance de gestes de la main dans un environnement multimodal*. Paper presented at the L'interface des mondes réels et virtuels, Montpellier.
- Beck, E., Christiansen, M., Kjeldskov, J., Kolbe, N., & Stage, J. (2003). *Experimental Evaluation of Techniques for Usability Testing of Mobile Systems in a Laboratory Setting*. Paper presented at the OzCHI 2003, Brisbane, Australia.
- Beigl, M. (2000). MemoClip: A location based remembrance appliance. *Personal Technologies*, 4(4).
- Bellik, Y. (1995). *Interfaces multimodales : concepts, modèles et architectures*. Unpublished Thèse de Doctorat en informatique de Paris XI, université Paris XI, Paris.
- Bellotti, V., & Bly, S. (1996). Walking Away from the Desktop Computer: Distributed Collaboration and Mobility in a Product Design Team. In K. Ehrlich & C. Schmandt (Eds.), *Proceeding of the sixth Conference on Computer Supported Cooperative Work* (pp. 209-218). Cambridge, MA: ACM Press.
- Benoit, C., Martin, J. C., Pelachaud, C., Schomaker, L., & Suhm, B. (1998). Audio-Visual and Multimodal Speech Systems. In D. Gibbon (Ed.), (Handbook of Standards and Resources for Spoken Language Systems ed., Vol. Supplement, to appear).
- Bernsen, N. O. (1995). *A taxonomy of input modalities* (Amodeus Esprit Basic Research Action No. WP22)o. Document Number)
- Bernsen, N. O., & Dybkjær, L. (2001, December). *Exploring natural interaction in the car*. Paper presented at the CLASS Workshop on Natural Interactivity and Intelligent Interactive Information Representation, Verona, Italy.
- Bernstein, J. (1988). Applications of speech recognition technology in rehabilitation. In J. E. Harkins & B. M. Virvan (Eds.), *Speech to text: Today and tomorrow* (Vol. n°2).
- Berthoz, A. (1997). *Le sens du mouvement*. Paris: Odile Jacob.
- Blattner, M., & Dannenberg, R. (1990). CHI'90 Workshop on multimedia and multimodal interface design. *SIGCHI Bulletin, ACM*, 22(2), 54-58.
- Blattner, M., Sumikawa, D., & Greenberg, R. (1989). *Earcons and Icons: Their Structure and Common Design Principles*. Paper presented at the Human-Computer Interaction.
- Boissière, P. (2002, 13-14 JUIN). (à paraître). *Vers un modèle d'aide à l'évaluation de systèmes d'assistance à l'écriture : application à VITIPI*. Paper presented at the Handicap 2002, Paris.

- Bolt, R. A. (1980). Put that there: Voice and gesture at the graphics interface. *ACM Computer Graphics*, 14(3), 262-270.
- Bouchart, I., Camus, G., Coello, Y., Courbois, Y., Edel, M., Kaczmarek, R., et al. (1997). *AccessIt: La performance spatio-temporelle de la personne handicapée dans la Communication Homme-Machine et l'Accès à l'Information* (No. 1): Institut Régional de Recherche sur le Handicapo. Document Number)
- Bousquet, C., & Vigouroux, N. (2001, 29-31 October). *Context Use to Improve the Speech Understanding Processing*. Paper presented at the International Workshop Speech & Computer - SPECOM-2001, Moscow.
- Braffort, A., Baudel, T., & Teil, D. (1992). *Utilisation des gestes de la main pour l'interaction Homme-Machine*. Paper presented at the IHM'92.
- Brewster, S., & Murray, R. (2000). Presenting dynamic information on mobile computers. *Personal Technologies*, 4(2), 209-212.
- Brison, E. (1997). *Stratégies de compréhension dans l'interaction multimodale*. Unpublished Thèse d'informatique, Université Paul Sabatier, Toulouse.
- Brotherton, J., Abowd, G. D., & Truong, K. (1999). *Supporting capture and access interfaces for informal and opportunistic meetings (GIT-GVU-99-06)*. Atlanta: GA: Georgia Institute of Technology, GVU Center.o. Document Number)
- Buxton, W. (1993). HCI and the inadequacies of Direct Manipulation systems. *SIGCHI Bulletin, ACM*, 25(1), 21-22.
- Buxton, W. (1995). Touch, gesture & marking. Chapter 7. In R. M. Baecker, J. Grudin, W. Buxton & S. Greenberg (Eds.), *Readings in human computer interaction : toward the year 2000*. San Francisco: Morgan Kaufmann.
- Cadoz, C. (1992, 03 mai). *Le geste canal de communication homme-machine*. Paper presented at the Journées de l'AFIA.
- Calvet, G. (2001). *Etude empirique de l'usage de la multimodalité dans l'interaction homme-machine* (Mémoire de DEA d'Ergonomie Cognitive). Paris: CNAMo. Document Number)
- Calvet, G., Kahn, J., Zouinar, M., Salembier, P., Briois, J. C., Nigay, L., et al. (2001). *Etude empirique de l'usage de la multimodalité avec un ordinateur de poche*. Paper presented at the People and Computers XV - Interaction without frontiers, Joint proceedings of HCI 2001 and IHM 2001, Lille.
- Calvet, G., & Salembier, P. (2006, 26-28 Juin). *"Cris et chuchotements"*. *Interactions mobiles en contexte réel*. Paper presented at the CITE'06, Nantes.
- Calvet, G., Salembier, P., Kahn, J., & Zouinar, M. (2005, 27-30 Septembre). *Etude empirique de l'interaction multimodale en mobilité : approche méthodologique et premiers résultats*. *Actes de la conférence*. Paper presented at the IHM'05, Toulouse.
- Carbonell, N., Valot, C., Mignot, C., & Dauchy, P. (1994). *Étude empirique : usage du geste et de la parole en situation de communication homme-machine*. Paper presented at the ERGO'IA, Biarritz.
- Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*.
- Cassell, J., Nakano, Y., Bickmore, T., Sidner, C., & Rich, C. (2001). *Non-verbal Cues for Discourse Structure*. Paper presented at the Association for Computational Linguistics Joint EACL - ACL Conference.
- Cassell, J., & Stone, M. (1999). *Living Hand to Mouth: Psychological Theories about Speech and Gesture in Interactive Dialogue Systems*. Paper presented at the AAAI 1999 Fall Symposium on Narrative Intelligence.
- Catinis, L., & Caelen, J. (1995, 13 octobre 1995). *Analyse du comportement multimodal de l'utilisateur humain dans un tâche de dessin*. Paper presented at the IHM'95, Toulouse.
- Chen, G., & Kotz, D. (2000). *A Survey of Context-Aware Mobile Computing Research* (No. TR2000-381): Dept. of Computer Science, Dartmouth Collegeo. Document Number)
- Cheyner, A., & Julia, L. (1995). Multimodal Maps : an agent based approach. In *Multimodal Human-Computer Communication* (pp. 111-121).
- Christian, K., Kules, B., Schneiderman, B., & A., Y. (2000). *A Comparison of Voice Controlled and Mouse Controlled Web Browsing*. Paper presented at the ACM ASSETS 2000 Conference, New York.
- Ciavarella, & Paternò, F. (2003, September 2003). *Design criteria for location-aware, indoor, PDA applications*. Paper presented at the 5th International Symposium, Mobile HCI 2003, Udine.

- Cohen, P. R., Johnston, M., McGee, D., Oviatt, S. L., Clow, J., & Smith, I. (1998). *The efficiency of multimodal interaction : A case study*. Paper presented at the International Conference on Spoken Language.
- Cohen, P. R., Johnston, M., McGee, D., Oviatt, S. L., Pittman, J. A., Smith, L., et al. (1997). *QuickSet: Multimodal interactions for distributed applications*. Paper presented at the Fifth ACM international Multimedia Conference.
- Cohen, P. R., & Oviatt, S. L. (1995). *The Role of Voice Input for Human-Machine Communication*. Paper presented at the the National Academy of Sciences.
- Coutaz, J., & Nigay, L. (1994, 8-9 décembre). *Les propriétés "CARE" dans les interfaces multimodales*. Paper presented at the IHM'94, Lille.
- Coutrix, C., & Nigay, L. (2006). *Mixed reality: A model of mixed interaction*. Paper presented at the AVI'06.
- Damper, R. I. (1984). Voice-Input aids for physically handicapped. *International Journal of Man-Machine Studies*, 541-553.
- De Angeli, A., Wolff, F., Lopez, P., & Romary, R. (1999, August 9-20). *Relevance and perceptual constraints in multimodal referring actions*. Paper presented at the Workshop on Deixis, Demonstration and Deictic Belief, Eleventh European Summer School in Logic, Language and Information" (ESSLLI-99), Utrecht, The Netherlands.
- Demunieux, R., & Losquin, P. (2005). *Collectionner les usages réels des clients de téléphonie (un outil embarqué)*.
- Derycke, A., Plénacoste, P., & Vanderdonckt, J. (1998, 2-4 septembre). *Vers des critères ergonomiques de conception/évaluation pour les sites Web*. Paper presented at the Contribution to the Workshop "Interactions sur le WEB", 10 ièmes Journées francophones sur L'Interaction Homme-Machine IHM'98, Nantes.
- Dey, A. K., Salber, D., & Abowd, G. D. (2001). A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *IJHCS*, 16.
- Dix, A., & Beale, R. (1996). *Remote cooperation : CSCW issues for mobile and teleworkers*. London: Springer-Verlag.
- Dix, A., Rodden, T., Davies, N., Trevor, J., Friday, A., & Palfreyman, K. (2000). Exploiting space and location as a design framework for interactive mobile systems. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(3), 285-321.
- Dybkjær, L., Berman, S., Kipp, M., Wegener Olsen, M., Pirrelli, V., Reithinger, N., et al. (2001). *Survey of existing tools, standards and user needs for annotation of natural interaction and multimodal Data (Report): ISLE Natural Interactivity and Multimodality Working Group*. Document Number
- Foley, J. D., & Van Dam, A. (Eds.). (1982). *Fundamentals of Interactive Computer Graphics* (Addison-Wesley ed.). Reading (Massachusetts).
- Frohlich, D. (1991). *The Design Space of Interfaces, Multimedia Systems*. Paper presented at the Interaction and Applications, 1st Eurographics Workshop, Stockholm, Suède.
- Gallis, H., Kasbo, J. P., & Herstad, J. (2001). *The multidevice paradigm in Knowmobile - Does one size fit all ?* Paper presented at the IRIS 24 Conference, Ulvik in Hardanger, Norway.
- Gaver, W. (1988). *Everyday Listening and Auditory Icons*. Unpublished Doctoral dissertation, University of California.
- Goddeau, D., Brill, E., Glass, J., Pao, C., Philips, M., Polifroni, J., et al. (1994). *GALAXY: a Human-Language Interface to On-line Travel Information*. Paper presented at the ICSLP 94, Yokohama, Japon.
- Goffman, E. (1973). *La mise en scène de la vie quotidienne. 1. La présentation de soi*. Paris: Les Editions de Minuit.
- Goffman, E. (1987). *Façons de parler*. Paris: Les Editions de Minuit.
- Goodman, J., Brewster, S., & Gray, P. (2004). *Using field experiments to evaluate mobile guides*.
- Gumperz, J. (1989). *Engager la conversation. Introduction à la sociolinguistique interactionnelle*. Paris: Les Editions de Minuit.
- Hassenzahl, M., & Tractinsky, N. (2006). User experience – a research agenda. *Behaviour & Information Technology*, 25(2), 91-97.
- Heath, C., & Luff, P. (2000). *Technology in action*. Cambridge: Cambridge University Press.
- Hinckley, K., Pierce, J., Sinclair, M., & Horvitz, E. (2000). *Sensing Techniques for Mobile Interaction*. Paper presented at the ACM UIST.
- Hollnagel, E. (1993). *Human reliability analysis. Context and Control*.

- Horchani, M. (2007). *Vers une communication Humain-Machine naturelle: stratégie de dialogue et de présentation multimodales*. Unpublished Informatique, Université Joseph Fourier, Grenoble.
- Huls, C., & Boss, E. (1995). Studies into Full Integration of Language and Action. *Multimodal Human-Computer Communication*, 312-325.
- Hutchins, E. (1995). *Cognition in the wild*. Bradford: Brooks-MIT Press.
- Jaimés, A., & Sebe, N. (2005, October 21). *Multimodal Human Computer Interaction: A Survey*. Paper presented at the IEEE International Workshop on Human Computer Interaction in conjunction with ICCV 2005, Beijing, China.
- Jameson, A. (2002, 17.19 June). *Usability Issues and Methods for Mobile Multimodal Systems*. Paper presented at the ISCA Tutorial and Research Workshop on Multi-Modal Dialogue in Mobile Environments, Kloster Irsee, Germany.
- Jeannerod, M., & Hochmann, J. (91). *Esprit, où es-tu? Psychanalyse et neurosciences*. Paris: Odile Jacob.
- Jost, M., Haussler, J., Merdes, M., & Malaka, R. (2005). Multimodal interaction for pedestrians: an evaluation study. In *Proceedings of the 10th international conference on Intelligent user interfaces* (pp. 59-66). San Diego, California, USA: ACM Press.
- Kakihara, M., & Sørensen, C. (2001). *Mobility Reconsidered: Topological Aspects of Interaction*. Paper presented at the IRIS 24, Ulvik in Hardanger, Norway.
- Karsenty, L. (2006). *Les déterminants du choix d'une modalité d'interaction avec une interface multimodale*. Paper presented at the ErgoIA, Biarritz.
- Kendon, A. (81). Geography of gesture: some observations on its history. *Semiotic Inquiry*, 2(1), 25-62.
- Kendon, A. (82). The study of gesture. *Semiotica*, 37, 129-163.
- Kerguelen, A. (1998). *Quels outils concevoir pour aider au relevé d'observation sur le terrain ?* Paper presented at the XXXIIIe Congrès de la SELF, "Temps et Travail", Paris.
- Ketola, P., & Røykkee, M. (2001). *The Three Facets of Usability In Mobile Handsets*. Paper presented at the CHI Workshop: Mobile Communications: Understanding Users, Adoption and Design.
- Kinsbourne, M. (2003, 14 novembre 2003). The multimodal mind: how the senses combine in the brain *Virtual Symposia*, from <http://www.semioticon.com/virtuals/multimodality.htm>
- Kirsh, D. (1997). Interactivity and multimedia interfaces. *Instructional Science*, 25(2), 79-96.
- Kjeldskov, J., & Graham, C. (2003). *A review of mobile HCI research methods*. Paper presented at the International Mobile HCI 2003 conference, Udine, Italy.
- Kjeldskov, J., Skov, M., Als, B., & Hoegh, R. (2004). Is it worth the hassle ? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In *Proceedings of the 6th International Mobile HCI 2004 conference, Glasgow, Scotland* (pp. 61-73): Lecture Notes in Computer Science, Berlin, Springer-Verlag.
- Kristoffersen, S., Herstad, J., Ljunberg, F., Løbersli, F., Sandbakken, J. R., & Thoresen, K. (1998). *Developping mobile CSCW scenarios*. Paper presented at the IRIS 21. Department of Computer Science, Aalborg University, 1998.
- Kristoffersen, S., & Ljunberg, F. (2000). *Mobile use of IT*. Paper presented at the IRIS 22 Conference, 7-10 August, Keuruu, Finland.
- Kristoffersen, S., & Ljunberg, F. (1999). "Making Place" to Make IT Work: Empirical Explorations of HCI for Mobile CSCW.
- Krus, M. (1995). *Présentation Multimodale d'Information*. Unpublished DEA d'Informatique, Université Paris XI.
- Leplat, J. (2001). La gestion des communications par le contexte. *Pistes*, 3(1).
- Luff, P., & Heath, C. (1998, November 14-18). *Mobility in collaboration*. Paper presented at the CSCW'98, Seattle.
- Mackinlay, J., Card, S. K., & Robertson, G. (1990). A semantic analysis of the design space of input devices. *Human computer interaction*, 5, 145-190.
- Makimoto, T., & Manner, D. (1997). *Digital Nomad*. Chichester: John Wiley & Sons.
- Mäntyjärvi, J., & Seppänen, T. (2003). Adapting applications in handheld devices using fuzzy context information. *Interacting with Computers*(15 ), 521-538.
- Marchand, T., Salembier, P., & Zouinar, M. (2000). *Activités coopératives distribuées en situation de mobilité* (No. Rapport de convention GRIC-IRIT/FT-R&D, Projet MAGIC). Issy-les-Moulineaux: FT-R&Do. Document Number)

- Martin, J.-C. (1995). *Coopération entre modalités et liage par synchronie dans les interfaces multimodales*. Ecole Nationale Supérieure des Télécommunications, Paris.
- Martin, J.-C., Julia, L., & Cheyer, A. (1998). *A theoretical Framework for Multimodal User Studies*. Paper presented at the cmc, Tilburg, The Netherlands.
- Maurel, F., Luc, C., Vigouroux, N., Mojahid, M., Virbel, J., & Nespoulous, J.-I. (2002, 29-30 Janvier 2002). *Transposition à l'oral des structures énumératives à partir de leurs paramètres formels*. Paper presented at the ISLsp 2002, Inscription Spatiale du Langage : structures et processus, IRIT.
- Maybury, M., & Lee, J. (1995). The Structure of Multimodal Dialogue II. In M. Taylor, F. Néel & D. G. Bouwhuis (Eds.), *Multimedia and Multimodal Interaction Structure*. Amsterdam: version électronique.
- Mellor, B. A., Baber, C., & Tunley, C. (1996). *Evaluating Automatic Speech Recognition as a Component of a Multi-input Human-computer Interface*. Paper presented at the International Conference on Spoken Language Processing.
- Merlin, B. (2004). *Plateforme générique d'observation des interactions* (Rapport Technique IRIT-DIAMANT). Toulouse: IRITo. Document Number
- Mignot, C. (1995). *Usage de la parole et du geste dans les interfaces multimodales - Etude expérimentale et modélisation*. Université Henri Poincaré, Nancy.
- Mignot, C., & Carbonell, N. (1996). Commande orale et gestuelle : étude empirique. *Technique et Science Informatiques*(15(10)), pp.1399-1428.
- Moran, T. P., & Dourish, P. (2001). Introduction to the special issue on Context-Aware Computing. *IJHCS*, 16.
- Nielsen, C., & Sondergaard, A. (2000). Designing for mobility : providing integration and overview on large and small screens. In L. Svensson, U. Snis, C. Sorensen, H. Fägerlind, T. Lindroth, M. Magnusson & C. Östlund (Eds.), *Proceedings of IRIS 23*. University of Trollhättan, Uddevalla.
- Nigay, L. (1994). *Conception et modélisation logicielles des systèmes interactifs : application aux interfaces multimodales*. Université Joseph Fourier - Grenoble 1, Grenoble.
- Nigay, L., & Coutaz, J. (1993). *A Design Space for Multimodal System: Concurrent Processing and Data Fusion*. Paper presented at the INTERCHI'93, Amsterdam.
- Nigay, L., & Coutaz, J. (1995). *A generic platform for addressing the multimodal challenge*. Paper presented at the CHI' 95.
- Nigay, L., & Coutaz, J. (1996). Espaces conceptuels pour l'interaction multimédia et multimodale. *Techniques et sciences informatiques*, 15(9), 1195-1225.
- Norman, D. (1991). Cognitive artifacts. In J. M. Carroll (Ed.), *Designing Interaction: Psychology at the Human-Computer Interface* (pp. 17-38). Cambridge: Cambridge University Press.
- Oriola, B., Vigouroux, N., Pérennou, G., & Truillet, P. (1993). IODE : une interface multimodale de consultation de documents électroniques pour déficients visuels (pp. 90-91). Forum des Recherches en Informatique - Ecole Polytechnique.
- Oulasvirta, A. (2005). The Fragmentation of Attention in Mobile Interaction, and What to Do with It. *Interactions, novembre + décembre*, 16-18.
- Oulasvirta, A., & Saariluoma, P. (2006). Surviving task interruptions: Investigating the implications of long-term working memory theory. *International Journal of Human-Computer Studies*, 64(10), 941-961.
- Oulasvirta, A., Tamminen, S., & Höök, K. (2005a). *Comparing two approaches to context: Realism and Constructivism*. Paper presented at the Critical Computing 2005, New York.
- Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. (2005b). Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 919-928). Portland, Oregon, USA: ACM Press.
- Oviatt, S., & Van Gent, R. (1996). *Error Resolution During Multimodal Human-Computer Interaction*. Paper presented at the International Conference on Spoken Language Processing, University of Delaware & A. I. duPont Instit.
- Oviatt, S. L. (1999, November,). *Ten myths of multimodal interactions*. Paper presented at the Communications of the ACM.
- Oviatt, S. L. (2000). *Multimodal System Processing in Mobile Environments*. Paper presented at the Thirteenth Annual ACM Symposium on User Interface Software Technology (UIST'2000), New York.

- Oviatt, S. L., Cohen, P. R., Wu, L., Vergo, J., Duncan, L., Suhm, B., et al. (2000). Designing the User Interface for Multimodal Speech and Pen-based Gesture Applications: State-of-the-Art Systems and Future Research Directions. *Human Computer Interaction [Reprinted in Human-Computer Interaction in the New Millennium]*, pp. 421-456.
- Oviatt, S. L., De Angeli, A., & Kuhn, K. (1997). *Integration and synchronization of input modes during multimodal human-computer interaction*. Paper presented at the CHI'97.
- Pascoe, J. (2001). *Context-Aware Software*. University of KENT, Canterbury.
- Pascoe, J., Ryan, N., & Morse, D. (1998). Human-computer-giraffe interaction : HCI in the field. In *Proceedings of the first workshop on Human Computer Interaction with mobile devices: GIST Technical report G98-1*.
- Pascoe, J., Ryan, N., & Morse, D. (2000). Using While Moving: HCI Issues in Fieldwork Environments. *ACM Transactions on Computer-Human Interaction*, 7(3), 417-437.
- Pérennou, G., De Calmès, M., Lavelle, A., & Tronel, R. (1998). *Un système de dialogue oral spontané pour l'accès téléphonique aux informations d'horaires de train* (Problème de robustesse, Systèmes complexes, systèmes intelligents et interfaces). Document Number)
- Petrelli, D., De Angeli, A., Walter, G., & Guilia, C. (1997, July 7-12). *Referring in multimodal systems: The importance of user expertise and system features*. Paper presented at the Referring phenomena in a multimedia context and their computational treatment. Un atelier en liaison avec le 35th Annual Meeting of the Association for Computational Linguistics - 8th Conference of the European Chapter of the Association for Computational Linguistics (ACL-EACL'97 Joint Conference ).
- Pirhonen, A., Brewster, S. A., & Holguin, C. (2002). *Gestural and Audio Metaphors as a Means of Control for Mobile Devices*. Paper presented at the ACM CHI2002, Minneapolis, MN.
- Privat, R., Vigouroux, N., & Truillet, P. (2002). Utilisabilité de l'interaction vocale pour l'accès aux systèmes interactifs par les personnes âgées. *Publication HANDICAP*.
- Quek, F. K. H. (1996). Unencumbered Gestural Interaction. *IEEE multimedia*, 3(4), 36-47.
- Quéré, L. (1997). La situation toujours négligée. *Réseaux*(n°85), 163-192.
- Rekimoto, J. (1996). *Tilting Operations for Small Screen Interfaces*. Paper presented at the User Interface and Software Technology (UIST'96).
- Renevier, P. (2004). *Systèmes mixtes collaboratifs sur supports mobiles : conception et réalisation*. Unpublished Informatique, Université Joseph Fourier, Grenoble.
- Rey, G. (2005). *Contexte en interaction homme-machine : le contexteur*. Unpublished Informatique, Université Joseph Fourier, Grenoble.
- Robertsson, L., Iliev, B., Palm, R., & Wide, P. (2007). Perception modeling for human-like artificial sensor systems. *International Journal of Human-Computer Studies*, 65(5), 446-459.
- Rosenfeld, R., Olsen, D., & Rudnicky, A. (2001). Universal Speech Interfaces. *Interactions*, VIII(6), 34-44.
- Roto, V., Oulasvirta, A., Haikarainen, T., Kuorelahti, J., Lehmuskallio, H., & Nyyssönen, T. (2004). *Examining mobile phone use in the wild with quasi-experimentation* (HIIT Technical Report No. 2004-1). Helsinki: HIITo. Document Number)
- Rudnicky, A. I. (1993). *Mode preference in a simple data-retrieval task, in ARPA*. Paper presented at the Human Language Technology Workshop, Princeton, New Jersey.
- Salber, D., Dey, A. K., & Abowd, G. D. (1999). *The Context Toolkit: Aiding the Development of Context-Enabled Applications*. Paper presented at the CHI'99, Pittsburgh.
- Schapiro, E., & Sharma, R. (2001, November). *Experimental Evaluation of Vision and Speech based Multimodal Interfaces*. Paper presented at the Workshop on Perceptive User Interfaces, Orlando, FL.
- Schilit, B., Adams, N., & Want, R. (1994). *Context-aware computing applications*. Paper presented at the Proceedings of the 1st International Workshop on Mobile Computing Systems and Applications., Los Alamitos, CA.
- Schmidt, A., Aidoo, K. A., Takaluoma, A., Tuomela, U., Laerhoven, K. V., & Velde, W. V. d. (1999a, September 27-29,). *Advanced Interaction in Context*. Paper presented at the International Symposium on Handheld and Ubiquitous Computing 1999, (HUC 99). Karlsruhe (Germany).
- Schmidt, A., Beigl, M., & Gellersen, H.-W. (1999b). There is more to context than location. *Computers and Graphics*, 23(6), 893-901.
- Schmidt, A., Van de Velde, W., & Kortuem, G. (2000). Situated Interaction in Ubiquitous Computing, *Proceedings of CHI 2000 workshop on Situated Interaction in Ubiquitous Computing*.
- Shneiderman, B. (2000). The limits of Speech Recognition. *Communications of the ACM*, 43(9), 63-65.

- Shriver, S., Toth, A., Zhu, X., Rudnicky, A., & Rosenfeld, R. (2001). *A Unified Design for Human-Machine Voice Interaction*. Paper presented at the CHI.
- Stanton, N. A. (2003). Human Error Identification in Human Computer Interaction. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: fundamentals, evolving technologies, and emerging applications*: Lawrence Erlbaum Associates.
- Sympson, J. B., & Hetter, R. D. (1985). *Controlling item-exposure rates in computerized adaptive testing*. Paper presented at the 27th Annual Meeting of the Military Testing Association, San Diego, CA: Navy Personnel Research and Development Center.
- Thomas, P. (2002). Introduction to The New Usability. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 9(2), 69-73.
- Truillet, P. (1997, 10-12 Septembre). *Étude de l'interaction multimodale dans une tâche de lecture de documents par des déficients visuels*. Paper presented at the IHM'97, Futuroscope.
- Truillet, P. (1999). *Modélisation de coopérations intermodales : application à l'interaction non-visuelle*. Unpublished Thèse d'université, Université Paul Sabatier, Toulouse.
- Truillet, P. (2002). Projet Vocal InteractiOn in air Traffic cOntrol. from <http://www.tls.cena.fr/~truillet>
- Vergo, J. (1998). *A statistical approach to multimodal natural language interaction*. Paper presented at the AAAI'98 Workshop on Representations for Multimodal Human-Computer Interaction.
- Verplaetse, C. (1996). Inertial proprioceptive devices: Self-motion-sensing toys and tools. *IBM SYSTEMS JOURNAL*, 35(3&4), 639-650.
- vom Lehn, D., Heath, C., & Hindmarsh, J. (2001). Exhibiting interaction : conduct and collaboration in museums and galleries. *Symbolic Interaction*, 24(2), 189-216.
- Want, R., Schilit, B. N., Adams, N. I., Gold, R., Petersen, K., Goldberg, D., et al. (1995, Dec). An overview of the ParcTab Ubiquitous Computing experiment. *IEEE Personal Communications*, 28-43.
- Winograd, T. (2001). Architectures for context. *IJHCS*, 16.
- Woodworth, R. S. (1899). Accuracy of voluntary movement. *Psychological Review*, 3(13), 1-114.
- Zanello, M.-L. (1997). *Contribution à la connaissance sur l'utilisation et la gestion de l'interface multimodale*. Thèse d'Université, Grenoble.
- Zanello, M. L., & Caelen, J. (1996, octobre). *Une approche centrée tâche de la multimodalité*. Paper presented at the IHM'96, Grenoble.

## 2. Annexes

### 2.1. *Index des critères et recommandations pour la conception et l'évaluation des interfaces multimodales mobiles*

Recom. 1 :	Prendre en compte toutes les phases potentielles du cycle d'interaction proposé (comprenant : Préparation, Entrée, Contrôle, Désengagement, Feedback de prise en compte de la commande, Feedback lié à l'effet de la commande) pour la conception d'une interface mobile multimodale, en particulier pour le choix des modalités à mettre en œuvre. 241
Recom. 2 :	Veiller à la continuité et à la cohérence des organes, sens, et modalités mis en œuvre dans toutes les phases du cycle en entrée comme en sortie du système. .... 242
Recom. 3 :	Dans le cas d'une application à mobilité et tâche « équifinale », l'interface doit être spécifiquement conçue pour le contexte de mobilité « réelle » requis par la tâche primaire. 242
Recom. 4 :	Dans le cas d'une application à mobilité et tâche « alterfinale », l'interface doit au moins être à « attention minimale »..... 242
Recom. 5 :	Identifier les repères absolus/ relatifs et les modalités de contrôle associées pour toutes les phases potentielles du cycle d'interaction proposé (comprenant : Préparation, Entrée, Contrôle, Désengagement, Feedback de prise en compte de la commande, Feedback lié à l'effet de la commande) ..... 242
Recom. 6 :	Proposer des modalités (y compris les modalités de contrôle) adaptées aux repères relatifs/absolus mis en œuvre pour toutes les phases potentielles du cycle d'interaction proposé (comprenant : Préparation, Entrée, Contrôle, Désengagement, Feedback de prise en compte de la commande, Feedback lié à l'effet de la commande)..... 243
Recom. 7 :	Il convient de contextualiser la présentation d'informations afin de faciliter son interprétation par l'utilisateur..... 243
Recom. 8 :	Pour une interaction mobile en particulier, les objets interactifs ou non doivent être facilement perceptibles et localisables..... 243
Recom. 9 :	L'interface en générale et les objets interactifs en particulier doivent être lisibles dans un large spectre de conditions environnementales. .... 244
Recom. 10 :	Proposer des sauts de curseur correspondants aux séquences d'actions à réaliser ..... 244
Recom. 11 :	Si l'application propose une gestion par mode d'interaction, le mode actif doit être clairement identifié ainsi que les autres modes disponibles..... 244
Recom. 12 :	Le moyen de passage d'un mode à l'autre doit être clairement identifié ..... 244
Recom. 13 :	De manière générale, il est préférable de regrouper les objets, interactifs ou non, ayant des caractéristiques similaires au même endroit..... 244

## Bibliographie et annexes

Recom. 14 :	Il est préférable de regrouper les éléments de gestion des modes (activation et désactivation d'une modalité) au même endroit .....	245
Recom. 15 :	Il est préférable de séparer/ regrouper les moyens de contrôle des modes selon qu'ils concernent les modalités d'entrées ou de sortie .....	245
Recom. 16 :	De manière générale, il est préférable d'attribuer des formats proches aux objets, interactifs ou non, ayant des caractéristiques similaires.....	245
Recom. 17 :	Il est nécessaire de distinguer par le format les moyens de contrôle des modes des autres commandes disponibles dans l'interface.....	245
Recom. 18 :	Il est nécessaire de distinguer par le format les moyens de contrôle des modes selon qu'ils concernent les modalités d'entrées ou de sortie .....	245
Recom. 19 :	Les commandes non disponibles doivent être clairement distinctes des commandes disponibles	245
Recom. 20 :	Lorsque les temps de réponse du système sont longs, prévoir au moins deux types de feedbacks : un premier feedback immédiat de prise en compte de la commande, puis un second feedback d'effet de la commande.....	246
Recom. 21 :	Fournir des feedbacks dans les modalités adaptées, en particulier lorsque les temps de réponse du système sont longs. ....	246
Recom. 22 :	Les feedbacks doivent être redondants au moins sous une forme visuelle et non visuelle.	246
Recom. 23 :	Pour les applications nécessitant des rafraîchissements réguliers, les rendre facilement repérables et adapter la modalité de sortie, (Oulasvirta, 2005) .....	246
Recom. 24 :	N'utiliser un retour système en synthèse vocale que si celle-ci a été activée par l'utilisateur.	246
Recom. 25 :	L'application doit supporter et surtout ne pas sanctionner l'organisation et la réorganisation temporelle des tâches. ....	247
Recom. 26 :	La mise en veille ne doit pas provoquer ou avoir des conséquences de changement d'état de l'application. ....	247
Recom. 27 :	Fournir un moyen simple d'activation et de désactivation de la synthèse vocale lorsque celle-ci est disponible dans l'application.....	247
Recom. 28 :	Fournir un moyen de pilotage (au minimum les fonctions pause, play) de la synthèse vocale lorsque celle-ci est disponible dans l'application.....	247
Recom. 29 :	Fournir la possibilité de faire varier la vitesse de lecture de la synthèse vocale .....	247
Recom. 30 :	Fournir la possibilité de supprimer des champs lus par la synthèse vocale .....	247
Recom. 31 :	Fournir un moyen simple de pilotage des médias audio lorsque ceux-ci sont disponibles dans l'application.....	248
Recom. 32 :	Fournir un moyen centralisé d'activation et de désactivation des sorties audio (incluant les sons et la synthèse vocale). ....	248

## Bibliographie et annexes

- Recom. 33 : La partie physique du dispositif doit être adaptée à une interaction à une main. La saisie du dispositif et l'appui sur les boutons physiques doivent être aisés..... 248
- Recom. 34 : L'appui sur les boutons physiques doit être facilité par une conception mettant en œuvre des modalités de contrôles non visuelles..... 249
- Recom. 35 : Eviter une gestion de déclenchement des modalités par mode d'interaction lorsque le nombre et la nature des modalités disponibles dans l'application le permettent. .... 249
- Recom. 36 : Lorsque la gestion par mode est inévitable, celle-ci doit être la plus simple et son usage doit être le moins visuel possible en entrée comme en sortie. Elle doit également être accessible via plusieurs modalités si possible. .... 249
- Recom. 37 : Automatiser ou éliminer les tâches au maximum, en particulier dans un contexte de mobilité 250
- Recom. 38 : La mise en veille ne doit pas conduire l'utilisateur à faire des actions supplémentaires... 250
- Recom. 39 : La mise en veille dans la mesure du possible (et au moins dans un premier temps après une commande utilisateur) ne devrait pas être opaque afin de ne pas masquer l'état des applications. 250
- Recom. 40 : Raccourcir les unités d'interaction en petits épisodes afin de libérer les ressources attentionnelles250
- Recom. 41 : La sémantique des commandes vocales doit être simple et utiliser dans la mesure du possible des mots ou combinaisons de mots courts..... 250
- Recom. 42 : Le moyen de pilotage de la synthèse vocale doit mobiliser le moins possible les ressources visuelles. .... 251
- Recom. 43 : Le moyen de pilotage des médias audio (du moins non visuels) doit mobiliser le moins possible les ressources visuelles. .... 251
- Recom. 44 : Le moyen d'activation et de désactivation des sorties audio (du moins non visuelles) doit mobiliser le moins possible les ressources visuelles. .... 251
- Recom. 45 : Fournir la possibilité de diminuer le nombre de feedbacks ou d'incitations et la qualité de ceux-ci pour chaque étape de l'interaction en fonction de l'expérience des utilisateurs.... 251
- Recom. 46 : Autant que possible, fournir la possibilité d'un choix de modalités d'interaction pour l'ensemble des commandes disponibles (principe d'équivalence fonctionnelle maximale)252
- Recom. 47 : Fournir aux utilisateurs un moyen facile pour changer de modalité d'interaction (gestion de mode) 252
- Recom. 48 : Fournir un ou plusieurs moyens de pilotage non visuels des médias non-visuels (synthèse vocale et messages audio etc.)..... 252
- Recom. 49 : La sémantique des textes et labels de l'interface doit être identique à la sémantique utilisées pour toutes les modalités (en particulier pour les commandes vocales). .... 252
- Recom. 50 : Dans la mesure du possible, veiller à la possibilité d'effectuer une séquence de commandes avec une seule modalité..... 252

Recom. 51 : De manière générale, l’interface et en particulier les commandes et objets interactifs doivent être disposés de manière stable et prévisible ..... 252

Recom. 52 : Les comportements et les fonctions associées aux boutons contextuels, hard ou soft doivent être stables, clairement identifiés et prévisibles ..... 253

Recom. 53 : Les commandes non utiles à la réalisation d’une tâche doivent être inactives afin de prévenir leur utilisation..... 253

Recom. 54 : Fournir des messages d’erreur mobilisant les mêmes sens et modalités que la modalité d’entrée utilisée pour la commande (seulement si ces modes ont été activés par l’utilisateur). 253

Recom. 55 : Les messages d’erreur non visuels doivent indiquer la nature de l’erreur afin d’en faciliter la compréhension et la correction sans solliciter la vision..... 253

Recom. 56 : Toujours permettre, à minima, la correction d’une erreur à l’aide de la même modalité que la commande ayant provoqué l’erreur ..... 253

Recom. 57 : Permettre à l’utilisateur de choisir la modalité de correction d’une erreur ..... 254

**2.2. Outils d’annotation de corpus multimodaux**

Outil	Objets d’intérêt	Fonctionnalités principales
<b>Anvil</b>	Langage et geste	Application java pour l’annotation de fichiers vidéo. Il permet d’annoter des comportements non-verbaux (gestes) et des concepts relatifs au discours sur plusieurs niveaux, ainsi que leur visualisation
<b>ATLAS</b>	Pas d’outil disponible. Support à l’annotation de tout signal linguistique	Outil cherchant à généraliser la formation et l’usage de corpus langagier
<b>CLAN</b>	Texte, discours et geste	Programme spécifiquement conçu pour l’analyse de données transcrites au format “Child Language Data Exchange System » (CHILDES). Les fichiers retranscrits peuvent être liés à des fichiers audio ou vidéo. Il permet de faire des analyses automatiques de corpus larges de retranscriptions (fréquence, comptage, cooccurrence, analyse d’interaction, etc.)
<b>CSLU</b>	Toolkit Speech, TTS, expression faciale	Suite d’outils regroupant RAD (Rapid Application Developer), BaldiSync (pour l’animation faciale), SpeechView, OGLsable (un outil d’annotation), des outils de reconnaissance de la parole et un environnement de programmation permettant l’animation d’un agent multimodal (TTS) (CSLUsh).
<b>MATE Workbench</b>	Discours et textes	Annotation de corpus de dialogues et extraction d’information des corpus annotés.
<b>MPI tools</b>	Discours et gestes	Tout les deux sont des outils d’annotation et

<b>(EUDICO and CAVA)</b>		d'analyse de fichiers audiovisuels en cours de développement. Ils permettent l'indexation de fichiers audio et vidéo.
<b>MultiTool</b>	Discours et gestes	Outils Java dédié à l'établissement et à l'usage de corpus multimodaux de langage parlé (audio et vidéo).
<b>The Observer</b>	Geste et expression faciale	Système professionnel pour la capture, l'analyse, la gestion et la présentation de données d'observations pouvant être utilisé pour l'enregistrement de l'activité, des postures, mouvements, expressions faciales, interactions sociales etc. sous une forme séquentielle d'informations indexées temporellement.
<b>SignStream</b>	Discours, gestes et expression faciale	Outil pour la navigation et l'annotation de données linguistiques capturées sur vidéo à l'aide de couches descriptives.
<b>SmartKom</b>	Discours, gestes, biométrie	N/A Le projet se base entre autre, sur l'étude, l'annotation et la production d'événements multimodaux.
<b>SyncWriter</b>	Discours et gestes	Outil de transcription et d'annotation d'événements synchrones comme un discours et des données issues de vidéos.
<b>TalkBank</b>	Discours (Transcriber) Texte, discours et gestes (CLAN, cf. ci-dessus)	Outil de retranscription à partir de fichiers audio.

**Tableau 31 : Outils d'annotation de corpus multimodaux (tiré de (Dybkjær et al., 2001))**

### 2.3. Index des figures

Figure 1 : Modèle d'Arbib montrant l'indépendance des segments composant une tâche de saisie. Il convient de noter les processus visuels parallèles traitant des différents aspects de l'objet devant être saisi et manipulé (localisation spatiale, taille, orientation). Ces processus sont corrélés aux contrôleurs du mouvement. ....	34
Figure 2 : Recensement des modalités selon (Bernsen, N. O., 1995).....	40
Figure 3: Taxonomie des dispositifs d'entrée (emprunté à (Buxton, 1995)).....	41
Figure 4 : Le référentiel MSM : un espace de référence pour un système multi-sensori-moteur. ....	44
Figure 5 : Espace CASE (Concurrent, Alterné, Synergique, Exclusif). ....	45
Figure 6 : Les différents types de multimodalité selon (Bellik, 1995). ....	45
Figure 7 : Les différents types de mobilités mis en œuvre lors de nos expérimentations. ....	67
Figure 8 : Modèle de Norman enrichi par Kirsh. Les flèches entre les ellipses représentent le couplage entre l'agent et l'environnement. A la différence du modèle du cycle de décision qui traite du but, de l'intention et des phases de planification comme des actions séquentielles, dans ce modèle, chaque étape impliquée dans la prise de décision est interactive. Il montre également que la nature de l'interactivité est diverse, allant d'une simple action à des actions préparatoires, exploratoires, de maintenance ou complémentaires. ....	69
Figure 9 : Architecture fonctionnelle et principales commandes de l'application.....	73
Figure 10 : Plan de l'espace d'expérimentation de la Cité de la Science et de l'industrie.....	74
Figure 11 : Les deux modèles de lunettes-caméra utilisés. A gauche, une caméra discrète et micro intégrés dans les branches, à droite une caméra et micro moins discrets montés sur rotule orientable.....	77
Figure 12 : Dispositif mobile d'enregistrement. ....	78
Figure 13 : Exemple de capture d'un fichier Log dans l'application de gestion des Logs. Dans l'exemple, l'utilisateur saisie au stylet le mot « bonjour » .....	79
Figure 14 : Une session d'autoconfrontation.....	80
Figure 15 : L'outil d'analyse de données Actogram <sup>TM</sup> .....	81

## Bibliographie et annexes

Figure 16 : Exemple d'un graphe d'activité fourni par Actogram <sup>TM</sup> pour un sujet et une session .....	82
Figure 17 Usage global des modalités toutes commandes, toutes sessions et tous sujets confondus (stylet, vocal, g2d) .....	88
Figure 18 Usage global des modalités toutes commandes, toutes sessions et tous sujets confondus après décompte des commandes de correction.....	89
Figure 19 Ratio des premières commandes et des commandes de correction .....	89
Figure 20 Nombre de commandes par sujet toutes sessions confondues .....	90
Figure 21 Usage global des modalités par sujet, toutes commandes et toutes sessions confondues .....	91
Figure 22 Usage global des modalités par sujet, pour les commandes premières .....	92
Figure 23 Evolution des usages par sessions, tous sujets confondus .....	93
Figure 24 Evolution des usages par session et par sujet .....	94
Figure 25 Les commandes les plus utilisées au niveau global .....	95
Figure 26 Usage global des modalités par commandes tous sujets et sessions confondus .....	96
Figure 27 Evolution de l'usage global des modalités par commandes tous sujets confondus .....	97
Figure 28 Différences inter-individuelles dans l'usage des modalités par commandes toutes sessions confondues ..	98
Figure 29 Transition des modalités pour toutes les commandes (incluant les commandes non équivalentes) .....	101
Figure 30 Transition entre la modalité précédente et la modalité suivante .....	101
Figure 31 Transition entre la modalité précédente et la modalité suivante en commande première ou correction ..	102
Figure 32 : Enchaînement temporel global des modalités pour le sujet 1 et pour les trois sessions .....	103
Figure 33 : Enchaînement temporel global des modalités pour le sujet 2 et pour les deux sessions disponibles .....	104
Figure 34 : Enchaînement temporel global des modalités pour le sujet 3 et pour les trois sessions .....	105
Figure 35 : Enchaînement temporel global des modalités pour le sujet 4 et pour les trois sessions .....	105
Figure 36 : Enchaînement temporel global des modalités pour le sujet 5 et pour les trois sessions .....	106
Figure 37 : Enchaînement temporel global des modalités pour le sujet 6 et pour les trois sessions .....	107
Figure 38 Répartition des erreurs selon les commandes pour la modalité vocale et stylet .....	111
Figure 39 Répartition des erreurs pour les 6 commandes les plus utilisées pour la modalité vocale et stylet .....	112
Figure 40 Répartition des erreurs selon les commandes pour la modalité vocale .....	114
Figure 41 Répartition des types d'erreurs pour la modalité tactile stylet.....	115
Figure 42 Répartition des erreurs selon les commandes pour la modalité stylet.....	115
Figure 43 Effet de l'environnement sonore sur l'usage des modalités pour l'ensemble des sujets .....	135
Figure 44 : Ratio du nombre de commandes selon le contexte sonore par sujet .....	136
Figure 45 : Influence du contexte sonore au niveau individuel .....	136
Figure 46 : Effet du contexte social sur l'usage des modalités pour l'ensemble des sujets.....	142
Figure 47 : Ratio du nombre de commandes selon le contexte social par sujet .....	143
Figure 48 : Ratio ( ?) du nombre de commandes selon le contexte social par sujet .....	144
Figure 49 Influence de l'espace de travail (fenêtre) sur l'usage des modalités.....	154
Figure 50 : Influence par sujet de l'espace de travail (fenêtre) sur l'usage des modalités .....	155
Figure 51 : Influence de l'activité réelle sur l'usage des modalités.....	157
Figure 52 : Influence de l'activité réelle sur l'usage des modalités au niveau individuel.....	158
Figure 53 : Chroniques d'activité du sujet 1 session 1 (focus sur les erreurs et leur récupération) .....	160
Figure 54 : Chroniques d'activité du sujet 1 session 2 (focus sur les erreurs et leur récupération) .....	161
Figure 55 : Chroniques d'activité du sujet 1 session 3 (focus sur les erreurs et leur récupération) .....	161
Figure 56 Usage des modalités selon les catégories de mobilité.....	170
Figure 57 : Mobilité selon les usages des modalités .....	171
Figure 58 Répartition des catégories de mobilité par sujets .....	171
Figure 59 Répartition des catégories de mobilité par sujets .....	172
Figure 60 Evolution de la mobilité au cours des sessions.....	173
Figure 61 Usage des commandes selon les catégories de mobilité.....	174
Figure 62 : Comparaison des ordres de mobilité et de modalité vocale .....	174
Figure 63 Mobilisation de l'attention visuelle en fonction des modalités.....	177
Figure 64 Direction du regard selon les sujets.....	178
Figure 65 Evolution de la direction du regard selon les sujets .....	179
Figure 66 Positionnement du pda et erreurs de reconnaissance vocale .....	180
Figure 67 Mobilisation de l'attention visuelle pour la modalité vocale en mode correction et premier.....	181
Figure 68 : Durée des catégories de mobilité en fonction de l'orientation de l'attention visuelle tous sujets et toutes sessions.....	181

Figure 69 : Catégories de mobilité en fonction de l'orientation de l'attention visuelle tous sujets et toutes sessions	182
Figure 70 Simultanéité de l'activité et de la mobilité	188
Figure 71 Nombres de commandes mobiles selon l'activité	188
Figure 72 : Transition de mobilité entre une commande et la suivante	190
Figure 73 : Enchaînement des états de mobilité selon les modalités	191
Figure 74 Répartition des catégories de mobilité selon le type de commande (premier et correction)	193
Figure 75 Influence des modalités de correction sur la mobilité	194
Figure 76 Comparaison de saisie de média (message vocal VS clavier virtuel)	207
Figure 77 : Saisie de médias selon le contexte sonore (niveau global)	208
Figure 78 : Saisie de médias selon le contexte social (niveau global)	209
Figure 79 : Saisie de médias et mobilité associée (niveau global)	209
Figure 80 : Degré de mobilité selon l'activité et la saisie de média	210
Figure 81 : Proposition d'un modèle détaillé de l'interaction	232
Figure 82 : Cadre descriptif général des finalités de l'interaction et de la mobilité	237

#### 2.4. Index des tableaux

Tableau 1 : Taxonomie de différents niveaux de mobilité (Dix et al., 2000).	21
Tableau 2 : Types de mobilité et technologies associées (d'après (Kristoffersen, S. & Ljunberg, 2000).	22
Tableau 3 : Espace de conception des interfaces en sortie de (Frohlich, 1991).	43
Tableau 4 : Recensement d'études faisant appel à plusieurs modalités en situation fixe.	48
Tableau 5 : Recensement d'études faisant appel à plusieurs modalités en situation mobile (ou portable).	50
Tableau 6 : Protocole expérimental	75
Tableau 7 : exemples de mots clés utilisés pour le codage des verbalisations.	83
Tableau 8 Répartition des erreurs et succès pour la totalité des commandes	110
Tableau 9 Répartition des types d'erreurs pour la totalité des erreurs liées à la modalité vocale.	113
Tableau 10 Répartition des types d'erreurs pour la totalité des erreurs liées à la modalité vocale	113
Tableau 11 : Taux d'erreur de la modalité vocale en milieu bruyant ou non	134
Tableau 12 Niveau sonore global	134
Tableau 13 : Différences, par sujets, du nombre de commandes vocales selon le contexte sonore.	137
Tableau 14 : Environnement social des expérimentations : durée et nombre de commandes	141
Tableau 15 : Différences, par sujets, du nombre de commandes vocales selon le contexte social	144
Tableau 16 : Comparaison des données quantitatives et déclaratives par sujet, concernant le nombre de commandes vocales selon le contexte social	151
Tableau 17 : Illustrations de commandes vocales en présence de tiers	152
Tableau 18 : Temps passé dans chaque espace de l'application	153
Tableau 19 : Description des catégories descriptives de l'activité réelle.	156
Tableau 20 : Durée globale de l'activité réelle	156
Tableau 21 Répartition temporelle des catégories de mobilité	169
Tableau 22 Répartition du nombre de commandes selon les catégories de mobilité	169
Tableau 23 : synthèse des différences de micromobilité selon la modalité (par sujet).	172
Tableau 24 Répartition globale de la direction du regard	176
Tableau 25 Répartition de la direction du regard	176
Tableau 26 Répartition de la direction du regard selon la modalité utilisée	177
Tableau 27 Répartition de la position du pda lors des commandes vocales	179
Tableau 28 : Différences pour l'état « statique » selon les modalités	191
Tableau 29 : Illustration d'une commande vocale avec regard temporairement dirigé vers l'environnement	226
Tableau 30 : Tableau récapitulatif des propriétés des modalités et médias.	229
Tableau 31 : Outils d'annotation de corpus multimodaux (tiré de (Dybkjær et al., 2001)).	277



**TITLE :**

Empirical study of mobile multimodal interaction in natural situation

**ABSTRACT:**

Ubiquity of mobile interactive devices either in professional, cultural or domestic domain constitutes a new research perspective to both individual and cooperative usage of new communication technologies. Mobile technologies rise, among others, new questioning about mobile usability specific issues, adapted methodologies and field data collection, identification of relevant contextual factors for mobile systems design, superimposition of interaction spaces. We believe that mobility is a new interaction paradigm carrying potential contextual variations, preventing traditional direct manipulation from being the most appropriate interaction style without serious adaptation because it currently requires a high level of visual attention .

In this PhD, we first present a review of current knowledge concerning Context Aware Computing (CAC) as well as mobile and multimodal interaction. We then present an empirical study of a mobile multimodal application in the context of a situated collaborative quiz-answering task.

Based on quantitative and qualitative results, our research first describes global and individual usage of the available modalities (tactile and vocal). In a second time, we study the influence of some identified internal and external contextual factors on the modalities selection,. For instance, we examine the impact of factors such as noise context, social context or history of the interaction on the modalities usage. A third part of our results highlights the relation between users choice of a modality depending on the attentional resources required by the mobile interaction.

All along the quantitative results, an extensive analysis of qualitative data such as user's verbalisation during self confrontation sessions, or activity chronicles examination, allowed us to argue and explain some modality choices and helped us identify some new factors taken into account by the users.

Finally, we use these results to propose some models for mobile multimodal applications, as well as some criteria and recommendations for their design.

**KEYWORDS:**

Multimodality, Mobility, Cognitive ergonomics, Context, Human-Machine interaction

AUTEUR :  
GUILLAUME CALVET-INGLADA

TITRE :  
*Etude empirique de l'interaction multimodale mobile en situation naturelle*

DIRECTEUR DE THESE :  
BERNARD PAVARD

LIEU ET DATE DE SOUTENANCE :  
TOULOUSE, 24 novembre 2008

RESUME :

L'ubiquité des dispositifs interactifs mobiles dans les sphères professionnelles, culturelles et domestiques constitue une occasion d'ouverture de nouveaux champs de recherche dans le domaine de l'étude des usages individuels et coopératifs des Nouvelles technologies de l'information et de la communication (NTICs). Les technologies mobiles amènent notamment à renouveler les points de vue sur des questions telles que, les problèmes d'utilisabilité spécifiques posés par la mobilité, les méthodes et les dispositifs de recueil de données, la question de l'identification des facteurs contextuels pertinents pour la conception de systèmes mobiles, la superposition des espaces d'interaction. De ce point de vue, nous posons que la mobilité constitue un nouveau paradigme d'interaction, générateur potentiel de variations contextuelles riches, pour lequel les styles d'interaction classiques basés sur la manipulation directe ne sont pas forcément les plus appropriés.

L'objectif de cette thèse est de documenter empiriquement la question de l'usage de dispositifs multimodaux en mobilité, de produire des éléments permettant de poser les bases d'un modèle de l'usage de dispositifs multimodaux en mobilité, et de fournir un ensemble de principes de recommandations pour la conception.

Dans un premier temps nous présentons une analyse critique de la littérature sur les questions relatives à la multimodalité, la mobilité et le Context Aware Computing (CAC). Nous définissons ensuite une méthode générale pour l'étude de l'usage de technologies multimodales en mobilité dans des situations écologiquement valides. Cette méthode, qui repose sur un dispositif d'enregistrement de données en situation et sur le recueil de verbalisations consécutives à l'activité assisté par les traces, est appliquée à une situation de jeu collaboratif en mobilité dans le contexte de la plateforme des usages LUTIN en situation naturelle (Cité des Sciences et de l'Industrie).

Dans un premier temps, nous présentons des résultats essentiellement quantitatifs qui permettent de décrire un ensemble d'usages des modalités, et d'examiner l'influence d'un certain nombre de facteurs contextuels (internes ou externes) sur les choix d'usage de ces modalités (contexte physique, social ou encore l'historique de l'interaction). Nous examinons également la relation existant entre le choix d'une modalité et son impact sur les ressources attentionnelles lors d'une interaction en mobilité. Tout au long de la présentation de ces résultats, une analyse plus fine de données qualitatives telles que les verbalisations des utilisateurs ou l'examen de chroniques d'activités, permet d'expliquer en partie ces choix d'usages mais également d'identifier de nouveaux facteurs potentiellement explicatifs.

Nous utilisons ensuite ces résultats pour proposer une ébauche de modélisation ainsi qu'un ensemble de recommandations pour la conception d'interfaces multimodales mobiles.

MOTS CLES :  
Multimodalité, mobilité, ergonomie cognitive, contexte, interaction homme-machine

DISCIPLINE : Ergonomie Cognitive

LABORATOIRE :  
Institut de Recherche en Informatique de Toulouse (IRIT)  
Université Paul Sabatier  
118 route de Narbonne  
F31062 Toulouse