

# Analyse du comportement multimodal de l'utilisateur humain dans une tâche de dessin

Lian Catinis, Jean Caelen

ICP-INPG, 46, Avenue F. Viallet  
38031 Grenoble Cedex

e-mail : catinis@icp.grenet.fr, jcaelen@icp.grenet.fr

## RESUME

Cet article décrit une expérience dans laquelle un groupe de sujets effectue une tâche de dessin avec un logiciel approprié, sous l'assistance d'un expert. L'objectif de l'expérience est d'étudier le comportement multimodal des sujets sous l'angle des propriétés CARE (Complémentarité, Assignation, Redondance, Équivalence) [14].

Par rapport à des résultats d'autres auteurs, [10] par exemple, l'usage de la multimodalité "synergique" semble plus important.

Nous avons observé une spécialisation des modes selon la tâche ainsi qu'une tendance à fiabiliser l'interaction par l'usage de la redondance. Ce facteur prend moins d'importance si l'action est complexe ou à faible risque : il semble que des considérations d'économie l'emportent alors pour limiter l'usage de la redondance.

Mais les véritables causes sont encore à cerner. A ce stade de notre investigation nous avons pu juger positivement de l'intérêt des propriétés CARE pour situer l'utilisateur par rapport à un contexte d'interaction.

## MOTS-CLÉS

interaction multimodale, ergonomie et comportement humain

## POSITION DU PROBLEME

L'usage de plusieurs modes et/ou modalités en communication homme-machine est conditionné par l'efficacité et l'utilisabilité des systèmes interactifs, et reste centré sur l'intérêt que peuvent apporter les interfaces multimodales. Il serait vain de concevoir de telles interfaces sans un espoir d'un usage effectif même si cet usage passe à terme par une phase d'apprentissage. La question est donc de prédire de manière aussi précise que possible, la portée d'un tel usage sans bien sûr recourir à des expérimentations en vraie grandeur qui nécessiteraient la réalisation de l'interface elle-même. Pour répondre à cette question, différentes expériences ont été réalisées [12], [21], [20], [22] en utilisant la technique dite *Magicien d'Oz* (*Moz*), technique mise au point dans d'autres secteurs d'utilisation (notamment par Richard, Morel, Falzon,

Amalberti, etc.). Carbonell et al. [10] ont étudié le comportement multimodal des usagers par rapport à une grille de critères, à savoir : exclusif, concurrent, alterné et synergique [7]. Les résultats semblent mitigés quant à l'usage effectif de la multimodalité — en particulier il ne semble pas évident que l'usage synergique des modes soit aussi fréquent qu'on pouvait le penser a priori. A la suite de tels résultats, on est en face de deux réactions possibles : (a) remettre en cause la technique *Moz* elle-même (ce dont de nombreux auteurs ne se privent pas de faire) et/ou (b) choisir une autre grille de critères pour poursuivre les expériences :

(a) remettre en cause une technique sans proposer d'alternative reste stérile. Il nous semble préférable de considérer les résultats *Moz* pour ce qu'ils sont et de les replacer dans un cadre comparatif plus vaste, en variant les approches et méthodes de mesure, les consignes et les contraintes des expériences,

(b) il faut remarquer que la grille {exclusif, concurrent, alterné et synergique} décrit un système multimodal et non le comportement de l'utilisateur. Depuis les travaux cités ci-dessus, différents chercheurs ont apporté des contributions nouvelles à une typologie des interactions multimodales du point de vue, de l'utilisateur : les propriétés CARE [14], [18], [23]. En reprenant les définitions à notre compte, nous résumons ces propriétés comme suit :

- C = Complémentarité, lorsque pour une action, l'utilisateur utilise une expression multimodale dans laquelle chaque mode est nécessaire (et contribue) à la compréhension de l'action,
- A = Assignation, lorsque pour une action, l'utilisateur choisit un mode récurrent particulier (ou un sous-ensemble de modes) pour s'exprimer (il se peut aussi que ce choix soit imposé par le concepteur du système),
- R = Redondance, lorsque pour une action, l'utilisateur utilise simultanément deux ou plusieurs modes à travers lesquels les informations sont redondantes,
- E = Équivalence, lorsque pour une action, l'utilisateur choisit indifféremment tel ou tel mode (ou un sous-ensemble de modes) pour s'exprimer.

C'est donc avec ces points de vue (a) et (b) que nous proposons une nouvelle expérimentation, à savoir

évaluer la multimodalité sous l'angle des propriétés CARE et en nous plaçant en situation de communication humaine contrainte par un support informatique dans le cadre d'une tâche donnée : il s'agit pour un usager de réaliser un dessin avec l'aide d'un assistant, tous les deux se trouvant côte-à-côte et face à un écran d'ordinateur (il n'y a plus d'ambiguïté pour le sujet sur la présence ou non d'un compère). Le choix de ce type de tâche et de situation sont justifiés par le souci de continuer et d'approfondir une démarche antérieure [21] mais surtout de progresser de façon maîtrisée pour obtenir des résultats ultérieurement comparables.

### DESCRIPTION ET CONDITIONS DE L'EXPERIENCE

L'objectif de la tâche proposée aux sujets est de dessiner quatre figures (fig. 1) dans un environnement de dessin "Draw" sur PC à partir d'éléments géométriques simples (triangles, carrés, cercles, lignes verticales et horizontales) (fig. 2) rangés dans une palette placée sous la zone de travail, vide au départ.

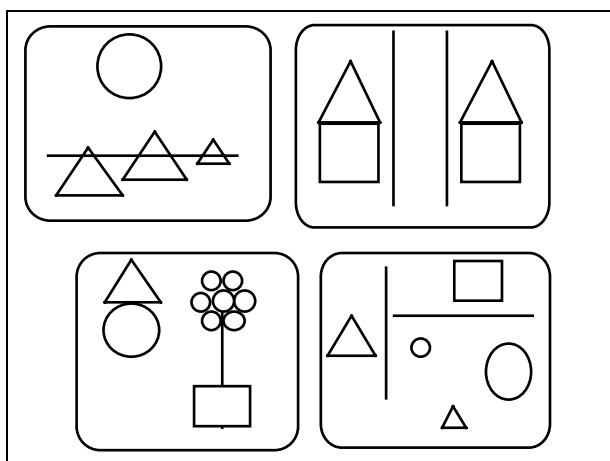


Figure 1 : les 4 figures à dessiner

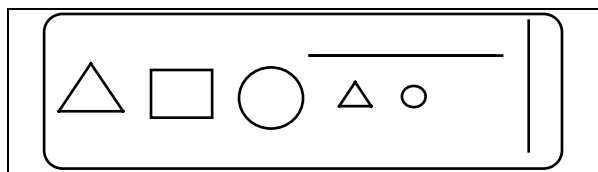


Figure 2 : la palette des objets disponibles

Les scènes figuratives sont de complexité variable vis-à-vis de la symétrie géométrique, de la symbolique de la scène, de la répétitivité des éléments et de la forme (gestalt). Nous avons volontairement limité le nombre de figures pour épargner la fatigue des sujets tout en essayant de faire varier au maximum ces paramètres de complexité d'une figure à la suivante (sans atteindre toutefois une combinatoire complète).

Afin de faire émerger des comportements multimodaux récurrents nous avons imposé aux sujets un ordonnancement des actions : nous comptons ainsi régulariser la charge de planification des actions pour tous les sujets et réduire son influence sur le choix des modes. Sans prétendre avoir effacé toute influence, cela nous a permis tout au moins de situer les sujets dans un même cadre normatif de tâche. Au cours de la session, le sujet joue le rôle d'instructeur vis-à-vis de son partenaire qui a un rôle d'assistant. Cet assistant est toujours le même pour tous les sujets ; il a subi un entraînement précis de façon à se comporter d'une manière aussi reproductible que possible. Nous avons encore contraint l'expérience en imposant au sujet un vocabulaire de commande limité et une syntaxe restreinte non ambiguë afin que l'assistant n'ait aucune difficulté ni marge d'interprétation d'une part et que d'autre part les sujets soient tous placés dans les mêmes conditions langagières ne favorisant pas *a priori* l'usage de la multimodalité. Cette nécessité s'imposait par ailleurs pour ne pas privilégier un mode par rapport à un autre c'est-à-dire pour annuler l'effet de préférence d'un mode d'expression — éventuellement plus riche — au détriment d'un autre : c'est pourquoi nous avons appauvri le mode langagier (notons au passage que ceci correspond mieux également aux performances actuelles des systèmes en reconnaissance automatique de la parole). Finalement, le sujet peut :

- énoncer des ordres multimodaux (parole et geste de désignation avec le doigt sur l'écran) ou monomodaux (parole seule) à l'adresse de l'assistant qui exécute ces ordres,
- dessiner lui-même la figure (ou un élément de la figure) avec les possibilités (souris+clavier) que lui offre le logiciel Draw.

Ceci définit les trois modes suivants :

- (pdd) parole et désignation du doigt sur l'écran,
- (p) parole seule,
- (s) souris+clavier (sans parole).

Remarquons que la modalité gestuelle dans le mode (pdd) ne passe pas par un capteur. Nous nous attendons donc à ce que ce mode soit plus spontané et plus "naturel" qu'un geste médiatisé par la souris. Par défaut, comme il est moins précis, il engendre un coût de traitement inférentiel de la cible plus grand de la part de l'interprète et donc peut-être une demande implicite de précision langagière plus forte de sa part vis-à-vis du sujet (pour le moment l'assistant humain mais plus tard la machine). Ces deux faits font que peut-être ce mode mixte (pdd) sera plus propre à supporter la multimodalité que d'autres modes mixtes étudiés jusque là, fondés sur la souris.

Pour rendre la situation plus engageante et plus réaliste pour le sujet nous lui avons demandé d'exécuter le dessin en un minimum de temps possible (mais cette contrainte a peut-être engendré un biais vis-

à-vis de l'assistant que certains pouvaient considérer comme un examinateur).

L'assistant avait pour rôle d'interpréter les ordres du sujet et de les exécuter. Il avait pour consigne de ne pas exécuter les ordres qui contenaient des mots ne faisant pas partie du vocabulaire ou des énoncés ayant une syntaxe trop complexe (propositions relatives, reprises, etc. étaient interdites). Le sujet et l'assistant avaient sous les yeux en permanence la liste des mots autorisés afin de ne pas pénaliser l'usage éventuel de la langue par rapport aux autres modes par un effort mémoriel supérieur. La sémantique devait se réduire impérativement au schéma *Action - Objet - Lieu* sous peine d'être rejetée par l'assistant, avec :

**Action** = [dessiner] [déplacer] [effacer] [stop]

**Objet** = [ligne verticale] [ligne horizontale] [grand triangle] [petit triangle] [grand cercle] [petit cercle] [carré] [*anaphoriques* = il, elle, le, la, les, etc.] [*déictiques* = ce, cette, ça, celui-ci, celui-là, etc.]

**Lieu** = [PositionRelative(objet)] [Position] [PositionAutoRelative]

**PositionRelative(objet)** = [au-dessus-de(objet)] [au-dessous-de(objet)] [à-gauche-de(objet)] [à-droite-de(objet)]

**Position** = [ici] [là] [au-centre] [en-haut] [en-bas] [à-droite] [à-gauche]

**PositionAutoRelative** = [plus haut] [plus bas] [plus à gauche] [plus à droite]

Malgré des contraintes aussi strictes, nous n'étions pas encore totalement assurés que l'assistant (pourtant entraîné) les respecte scrupuleusement et sans fatigue. Pour atténuer ce biais possible nous avons réduit la durée des expériences au minimum et conservé le même assistant pour toutes les expériences (ainsi tous les sujets étaient en face des mêmes conditions interlocutoires).

## DEROULEMENT DE L'EXPERIENCE

Les sujets n'étaient pas tous familiers des logiciels de dessin. Sur les 26 sujets testés 3 se trouvaient pour la première fois devant un écran et 5 étaient habitués à utiliser l'informatique. Les autres étaient de niveau variable (occasionnels). Les sujets novices pouvaient se familiariser avec le logiciel avant l'expérience pour se décontracter et éviter des blocages psychologiques (en effet il pouvait s'instaurer involontairement des relations de compétence hiérarchique avec l'assistant). Les domaines d'activité des sujets étaient très larges : des programmeurs, des ingénieurs de plusieurs spécialités, des médecins, des économistes, des secrétaires, des étudiants, des économistes.

Les expériences étaient filmées et analysées manuellement : on a relevé pour chaque action les propriétés CARE des sujets. Les actions retenues par l'analyse sont données dans l'arbre (fig. 3), c'est-à-dire essentiellement les actions à résultat immédiat.

Les analyses s'intéressaient aussi aux relations temporelles dans le cas des commandes avec déictiques [là] [ici] [ça] accompagnées de gestes de désignation du doigt sur l'écran. Ces relations temporelles ont été notées ( $\boxed{k} > D$ ,  $\boxed{k} = D$ ,  $\boxed{k} < D$ ,  $\boxed{k} \ll D$ ) pour indiquer respectivement que (a) le geste est consécutif au mot prononcé, (b) le geste est sensiblement simultané avec le mot prononcé, (c) le geste précède le mot prononcé, (d) le geste précède la séquence parlée qui contient la deixis et le doigt reste posé sur l'écran ensuite pendant l'énoncé verbal.

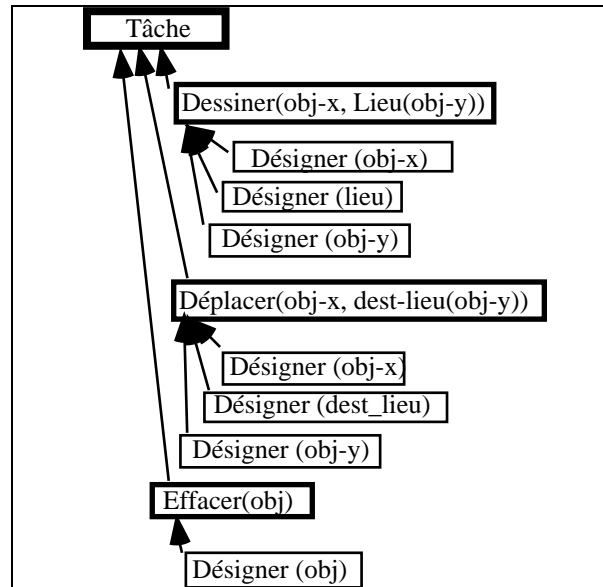


Figure 3 : Dans les cadres gras, les actions multimodales retenues dans l'analyse dans l'ordre de leur complexité sémantique décroissante (avec leur décomposition en actions plus élémentaires).

Enfin, toutes les observations intéressantes étaient notées, ainsi que les commentaires des sujets et leurs réactions en cas d'erreur d'exécution de l'assistant ou de leurs propres erreurs.

## RESULTATS TIRES DE L'EXPERIENCE

### L'équivalence des modes

L'équivalence des modes désigne le fait d'obtenir un même résultat en utilisant différentes modalités [17]. On peut distinguer deux notions : l'équivalence de résultat et l'équivalence fonctionnelle.

Équivalence de résultat : on peut constater que les modes proposés dans cette expérience (pdd) (p) et (s) sont équivalents vis-à-vis du résultat final (obtenir une figure). En effet chaque mode est suffisant à lui seul pour effectuer la tâche à accomplir. L'analyse des statistiques montre ce résultat : sur les 104 figures

dessinées par les 26 sujets, 40% ont été dessinées de bout en bout avec le mode (pdd), 6% avec le mode (p), 16% avec le mode (s), le reste des figures ayant été dessinées en variant les modes au cours de la séance. Il y a équivalence des modes mais pas de préférence de la part des sujets.

**Équivalence fonctionnelle :** L'équivalence fonctionnelle étant basée sur la précision et l'aisance d'usage de chaque mode [17], nous concluons de notre expérience qu'il n'y a pas d'équivalence fonctionnelle entre les modes. En effet nous avons considéré pour le paramètre *précision* les actions Désigner(objet) et Désigner(lieu) — qui seules nécessitent une certaine précision d'exécution — et nous avons relevé le nombre de fois où l'un et/ou l'autre mode ont été utilisés. Les résultats sont les suivants :

toutes catégories de sujets		
(pdd)	(p)	(s)
66%	14%	20%

Tableau 1 : usage des modes

catégorie expert		
(pdd)	(p)	(s)
81%	9%	10%

Tableau 2 : usage des modes

catégorie occasionnel		
(pdd)	(p)	(s)
53%	17%	30%

Tableau 3 : usage des modes

catégorie novice		
(pdd)	(p)	(s)
78%	22%	0%

Tableau 4 : usage des modes

On constate, à partir de ces 4 tableaux (tableau 1 à tableau 4), qu'il y a en moyenne et dans chaque catégorie un usage prédominant du mode pdd. L'usage des deux autres modes (p et s) dépend de la catégorie d'usager, les plus réticents à la souris étant les novices, et les plus favorables les occasionnels.

Le tableau 5 affine quelque peu ces résultats : il présente les fréquences d'utilisation du geste de désignation du doigt et de la désignation verbale des lieux et des objets pour les trois catégories de sujets.

expert			
désignation-lieu		désignation-objet	
doigt	parole	doigt	parole
90%	10%	22%	78%
occasionnel			
désignation-lieu		désignation-objet	
doigt	parole	doigt	parole
87%	13%	23%	77%
novice			
désignation-lieu		désignation-objet	
doigt	parole	doigt	parole
88%	12%	22%	78%

Tableau 5 : usage de la modalité "doigt" ou "parole" pour désigner un lieu ou un objet selon le type d'usager

Ces résultats (remarquablement stables) montrent que le lieu est désigné préférentiellement par un geste, tandis qu'un objet est plutôt désigné verbalement par son nom ou par un de ses traits pertinents (de couleur, de position, etc., par exemple "déplace le rouge à gauche"), ceci indépendamment de la catégorie des sujets.

Le paramètre *aisance d'usage* est plus difficile à mesurer par observation directe, il aurait fallu installer un dispositif de mesure du temps d'exécution des actions dans le logiciel et surtout disposer d'une interface réellement multimodale. Nous en sommes donc restés à une estimation subjective obtenue à la suite des commentaires donnés par les sujets eux-mêmes : 22/26 des sujets ont préféré le mode (pdd).

Les résultats sont donc clairs dans notre expérience : le mode (pdd) (parole et désignation du doigt sans contrainte instrumentale) multimodal par essence, qui semblait *a priori* le plus "naturel" emporte effectivement l'adhésion des sujets qui l'utilisent dans la majorité des cas.

### L'assignation des modes

Pour estimer le degré d'adéquation d'un mode à un type d'action (assignation) nous avons calculé pour chaque action et selon son degré de réalisation (succès ou échec), le pourcentage relatif des modes utilisés. Avec plus de détails par types d'action nous avons :

(a) pour les cas de réussite de l'action

"dessiner"		
(s)	(p)	(pdd)
18,6%	6,6%	74,8%
"déplacer"		
(s)	(p)	(pdd)
5,8%	31,4%	62,8%
"effacer"		
(s)	(p)	(pdd)
22%	27%	51%

(b) pour les cas d'échec de l'action (les valeurs indiquées sont les nombres d'actions échouées)

qui sont en elles-mêmes peu nombreuses par rapport au nombre total d'actions tentées)

"dessiner"		
(s)	(p)	(pdd)
1	1	15
"déplacer"		
(s)	(p)	(pdd)
0	2	3
"effacer"		
(s)	(p)	(pdd)
0	2	0

Tableau 6 : usage des modes selon le type d'action

Remarquons tout d'abord que l'action "déplacer" est assez souvent effectuée en préambule de l'action "dessiner" c'est-à-dire que le sujet commence à placer un objet sur l'écran à partir de la palette, puis utilise l'action "dessiner" pour positionner plus finement l'objet. Pour ce faire il utilise plus volontiers la souris qui permet une désignation plus précise du lieu. Cela explique pourquoi la souris est davantage sollicitée dans "dessiner" que dans "déplacer" (de plus pour une désignation vague la parole est plus économique). Une assignation se dessine donc entre l'usage de la parole (p) et l'usage de la souris (s). Pour l'action "effacer", parole et souris s'équilibrent : des raisons de meilleure "sécurité" n'ont pas fait préférer (s) contre (p). Remarquons que le mode (pdd) reste en tête des usages et ceci malgré les échecs plus nombreux (mais encore rares donc probablement supportables) — les échecs qui ont été répertoriés ici ne sont dus qu'à des erreurs du sujet (et non celles qui proviennent d'une mauvaise interprétation de l'assistant).

### La complémentarité

La complémentarité se définit pour une commande multimodale par rapport aux informations véhiculées par chaque modalité : lorsque ces informations se complètent pour donner un sens à l'énoncé multimodal, on dit que les modalités sont complémentaires — cela ne présage pas de leur usage synchrone (et donc synergique) ou asynchrone (alterné). Dans notre expérience cela ne concerne que le mode (pdd) : la complémentarité se mesure dans une tâche en comptant le nombre d'actions qui, utilisent des modalités complémentaires sur le total des commandes multimodales. Les résultats donnent pour la tâche complète de dessin et tous sujets confondus :

"dessiner"	"déplacer"	"effacer"	moyenne pour toutes les actions
90.2%	59.3%	58%	66.6%

Tableau 7 : usage de la complémentarité selon le type d'action

Ces résultats montrent que la complémentarité est davantage mise à profit par les usagers pour les opérations sémantiquement complexes (voir fig. 3). En effet le recours à la complémentarité dans un énoncé peut être interprété comme une recherche d'économie du côté de la production : il est évident qu'un énoncé dans lequel chaque élément est nécessaire et suffisant est plus "économique" qu'un énoncé redondant qui aurait toutes choses égales par ailleurs, les mêmes effets. C'est la raison pour laquelle une commande déjà coûteuse à formuler au plan sémantique aura des chances d'être moins redondante qu'une commande plus simple. Il est évident que ceci peut être contredit par le critère de "fiabilité" qui va à l'encontre de l'économie de production et qui favorise l'économie de résultat. Nous discutons ce point maintenant.

### La redondance

La redondance multimodale à laquelle on s'intéresse ici (il peut en effet y avoir aussi une redondance monomodale) se définit en opposition avec la complémentarité : elle renvoie aux informations redondantes véhiculées par plusieurs modalités qui ont les mêmes résultats dans l'interprétation de la commande. Du point de vue de l'énonciateur, la redondance ne peut représenter une économie de production, mais du point de vue du destinataire cela peut représenter une économie d'interprétation : elle lui sert souvent à désambigüiser la commande ou à l'éclaircir. En retour, la redondance est un moyen pour l'énonciateur de fiabiliser sa commande. Donc, complémentarité et redondance s'équilibrent sur l'axe économie-fiabilité du point de vue de l'énonciateur.

Le tableau ci-après indique pour les actions "dessiner", "déplacer" et "effacer" les redondances pour l'expression d'un lieu ou la référence d'un objet. On retrouve bien le fait que l'opération "à risque" (effacer) devant être la plus fiable est aussi celle qui est la plus redondante. Avec ce même type de raisonnement on peut avancer que l'objet de l'action "dessiner" est l'objet de plus d'attention de la part du sujet que le lieu.

"dessiner"		"déplacer"		"effacer"
désignation		désignation		désignation
objet	lieu	objet	lieu	objet
7.4%	2.4%	19.8%	20.9%	42%

Tableau 8 : usage de la redondance selon le type d'action et les éléments sémantiques à l'intérieur des énoncés de ces actions

On a remarqué par ailleurs deux tendances chez les sujets qui ont passé une deuxième fois l'expérience : (1) voulant être plus précis dans la désignation des lieux (certainement pour augmenter la fiabilité de leurs commandes et donc l'efficacité générale de leur action) ils ont augmenté le taux de redondance des modalités (par exemple en disant "ici à droite du cercle" et en

montrant la position correspondante). (2) En même temps, s'étant rendu compte que la désignation des objets n'était pas très ambiguë dans cette application (les objets sont facilement discriminables et nommables), ils ont maintenu le même taux de complémentarité pour la dénomination des objets<sup>1</sup>.

### LES RELATIONS TEMPORELLES DANS LES CAS DE DEIXIS

Un second objectif de l'expérience était de mesurer les relations temporelles entre la prononciation de mots déictiques (ici, là, ça) et la désignation gestuelle des objets et des lieux correspondants. Les analyses ont montré que dans 67.3% des cas de deixis l'utilisateur a désigné avec le doigt en même temps — cette notion de synchronie se rapporte à la perception d'un observateur humain — qu'il prononçait le déictique (la synchronie est notée :  $\boxed{k} = D$ ) : nous appelons ce cas (a) synchronie vraie synergique. Deux autres cas de synergies ont été observés :

- (b)  $D < \boxed{k}$  : asynchronie anticipative (ou fausse synchronie) de la parole sur le geste,
- (c)  $\boxed{k} < D$  : asynchronie anticipative du geste sur la parole,

ainsi que deux cas d'asynchronie alternée :

- (d)  $\boxed{k} \ll D$  : le geste précède le temps de parole (et a fortiori la deixis contenue dans la séquence de parole),
- (e)  $D \ll \boxed{k}$  : le geste survient nettement après le temps de parole.

$\boxed{k} \ll D$	$\boxed{k} < D$	$\boxed{k} = D$	$D < \boxed{k}$ et $D \ll \boxed{k}$
3.2%	28.8%	67.3%	0.7%

Tableau 9 : taux de synchronie et d'asynchronie dans les actions multimodales (voir significations de D et  $\boxed{k}$  dans le texte).

On notera le faible nombre de cas où la parole anticipe sur le geste, cela est peut-être dû à l'usage du doigt qui "traîne" sur l'écran entre deux actions. Dans

<sup>1</sup>Mais ne voulant pas étudier spécialement les phénomènes liés à l'apprentissage, nous n'avons pas poussé cette expérience avec de nombreux sujets, il faut donc prendre ces résultats avec précaution et les considérer comme des tendances purement qualitatives.

les commandes de type "mets ça là" les désignations gestuelles ont toujours respecté l'ordre des déictiques c'est-à-dire d'abord pour satisfaire "ça" et ensuite pour satisfaire "là".

### OBSERVATIONS GENERALES

#### Ellipses

Les formes d'expression comportaient naturellement des ellipses, anaphores et déictiques mais aussi des ellipses déictiques (omission de ça par exemple dans "déplace" + geste).

#### Réparation des erreurs

L'assistant a fait quelques erreurs dans l'exécution des commandes. Les erreurs ont porté sur la taille ou la position d'objets, ou sur la référence de l'objet, ou plus rarement sur la commande. Le sujet n'a pas toujours réparé ces erreurs, il a même parfois pensé que cette erreur venait de lui. Quand il a réparé les erreurs il a réitéré sa commande (seule stratégie qui lui était permise dans cette expérience) plus comme s'il s'agissait d'une mauvaise audition que d'une mauvaise compréhension de la part de l'assistant.

**1-Réparation monomodale :** Les améliorations de l'expression verbale n'ont pas été sensibles, en particulier les cas d'utilisation d'ellipses pour réparer les situations d'échec ont été de 67% ce qui montre bien que le sujet ne tente pas d'améliorer la précision de son expression même dans une situation difficile, à moins qu'il ne veuille au contraire attirer l'attention que sur l'élément incriminé, comme dans l'exemple suivant :

commande : "dessine un carré là"  
 exécution : un grand cercle est dessiné  
 réparation : "un carré ..... un carré ..."

**2-Réparation multimodale :** Le tableau 10 donne quelques éléments d'appréciation sur la manière dont les sujets réparent les erreurs en s'aidant de la multimodalité, soit en réitérant la commande sur un même mode soit en changeant de mode. Le mode pdd étant dans tous les cas renforcé, il semble qu'il apparaisse plus sécurisant pour le sujet.

mode correction	mode ayant provoqué l'erreur	
	p	pdd
s	0%	0%
p	12%	6%
pdd	88%	94%

Tableau 10 : fréquence d'utilisation des modes dans la phase de réparation

#### Stratégies de placement des objets

On peut tirer quelques observations qualitatives sur le déroulement de la tâche :

- certains usagers procédaient par approximation successive pour placer les objets. Les cas d'ambiguïté de déictiques devaient alors être résolus par l'assistant de façon contextuelle.

- quelques usagers désignent les lieux ou les objets par entourage.

- pour déplacer un objet, certains usagers font redessiner un nouvel objet identique au premier par l'assistant puis lui font effacer l'ancien.

- quelques cas d'attente importante entre l'énoncé de la commande et la désignation gestuelle se sont produits : des usagers attendaient le "feed-back" de la machine avant de préciser le lieu de destination de l'objet.

### CONCLUSION ET PROLONGEMENTS

Les résultats de l'expérience donnent une bonne idée sur le comportement multimodal de l'utilisateur — dans la situation de communication humaine contrainte dans laquelle nous l'avons placé et pour une tâche de dessin assisté.

Par rapport à des résultats d'autres auteurs ([10] par exemple), l'usage de la multimodalité semble plus important : cela est peut-être dû au fait qu'ici le geste n'est pas médiatisé par un capteur et que la communication se situe dans un contexte humain. Nos résultats s'apparentent à ceux de Bellik [2] concernant l'évaluation de l'interface LIMSIDraw réellement implémentée : cela permet d'avancer que des simulations peuvent être mises en place valablement pour prédire des résultats cohérents et que des expériences telles que celles-ci peuvent être poursuivies.

En ce qui concerne les propriétés CARE, nous avons observé une assignation naturelle des modes selon la tâche ainsi qu'une tendance à fiabiliser l'interaction par l'usage de la redondance (également dans [1]). Ce facteur prend moins d'importance si l'action est complexe ou à faible risque : il semble que des considérations d'économie l'emportent alors pour limiter l'usage de la redondance. En fait la multimodalité apparaît plus comme une multimodalité actionnelle que comme une multimodalité informationnelle. Nous pensons donc que d'autres critères que les critères CARE doivent être pris en compte pour juger de la multimodalité car ils sont trop liés à la notion d'information (de manière similaire à la théorie de Bernsen [3]) et pas assez à celle d'action ou d'interaction (par exemple pour juger de la rémanence d'un processus actionnel ou de l'efficacité d'une action).

Les raisons d'utilisation de tel ou tel mode sont encore à affiner par de nouvelles expériences. Le travail sera poursuivi ; mais déjà, à ce stade de notre investigation nous avons pu juger positivement de l'intérêt des propriétés CARE pour situer l'utilisateur par rapport à un contexte d'interaction dans une visée informationnelle.

### REMERCIEMENTS

Cette étude a été menée en étroite collaboration avec A. Bisseret, qu'il en soit chaleureusement remercié ainsi que la société "Systems-International" qui a offert son soutien et son aide à la réalisation des expériences.

### BIBLIOGRAPHIE

1. Bell D., Johnson P., General models of multimedia interaction. ERCIM Workshop Report 94-W003, Inria, Nancy, p. 25-39, 1994.

2. Bellik Y., Interfaces multimodales : concepts, modèles et architectures. Thèse de l'Université d'Orsay, LIMSI, 1995.

3. Bernsen N.O., Modality theory : supporting multimodal interface design. ERCIM Workshop Report 94-W003, Inria, Nancy, p. 13-23, 1994.

4. Bretan Y., Karlgren J., Synergy effects in natural language-based multimodal interfaces. ERCIM Workshop Report 94-W003, Inria, Nancy, p. 43-58, 1994.

5. Bisson P., Nogier J-F., Interaction homme-machine multimodale : le système MELODIA dans "ERGO-IA92" Biarritz, p. 69-90, 1992.

6. Bourguet M.L., Conception et réalisation d'une interface de dialogue personne-machine multimodale, thèse de doctorat, ICP/INPG, Grenoble 1992.

7. Caelen J., Coutaz J., Interaction homme-machine multimodale: quelques problèmes. IHM'91, Troisième journées sur l'ingénierie des interfaces homme-machine. Dourdan 11-13 December 1991.

8. Caelen J., Compte rendu du workshop IHMM organisé par le GDR-PRC "Communication Homme-Machine", Dourdan 13-14 April, p. 213-228, 1992.

9. Caelen J., Multimodal Human-Computer Interface, First AI-SHAM International Conference on Information Technology, Damascus - Syria , May 1994.

10. Carbonell N., Valot Cl., Mignot Ch., Dauchy P., Etude empirique : usage du geste et de la parole en situation de communication homme-machine. Actes du congrès ERGO'IA, p. 128-139, 1994.

11. Catinis L., Caelen J., Multimodal Man-Machine Interaction in Administration and Public Services, First AI-SHAM International Conference on Information Technology, Damascus - Syria , May 1994.

12. Chapelier L., Fay-Varnier Ch., Roussalany A., Saint-Dizier V., Recueil et analyse d'un corpus d'interactions multimodales homme-machine. Actes du congrès ERGO'IA, p. 96-107, 1994.

13. Coutaz J. & Gourdol A., Communication homme-machine multimodale: perspectives pour la recherche. Deuxième Journées Nationales GRECO PRC CHM, Toulouse, p. 17-28, January 1991.

14. Coutaz J. et Nigay L., Les propriétés "CARE" dans les interfaces multimodales, IHM'94 Sixième journées sur l'ingénierie des interfaces Homme-Machine. Lille, p. 7-14, 8-9 décembre 1994.

15. Faure C., Julia L. , Interaction homme-machine par la parole et le geste pour l'édition de documents :

TAPAGE. Actes interface des mondes réels et virtuels. Montpellier, p. 71-180, 1994.

16. Faure C., Arnold M., L'interaction homme-machine du point de vue des principes d'économie. Actes de cinquièmes journées sur l'ingénierie des interface Homme-Machine IHM'93, Lyon, p. 3-8, 19-20 Octobre 1994.

17. IHM'93, Synthèse de l'atelier "Interfaces Multimodales", sous-groupe : "Formes de multimodalité en situation d'interaction utilisateur-machine".

18. Martin J-C, Béroùle D., Types et buts de coopération entre modalités, IHM'93, 5ème journées sur l'ingénierie des interfaces homme-machine, Lyon, p. 17-22, 1993.

19. Martin J-C, Etude fondée sur des Types et buts de coopération entre modalités. Troisièmes journées

internationales "L'interface des mondes réels et virtuels". Montpellier 1994.

20. Mignot Ch., Valot Cl., Carbonell N., An experimental study of future "natural" multimodal human-computer interaction. Proc. InterCHI'93, Amsterdam, p. 67-68, 1993.

21. Ozkan, N., Caelen J., Designation of graphical objects in human-computer interaction. Proc. of WWDU'92 congresss. Berlin, 1992.

22. Valot Cl., Interface Multimodale Projet Grenoblois, Rapport d'activité du pôle Interfaces Homme-Machine Multimodales, GDR-PRC CHM, 1993.

23. N. Vigouroux, E. Brisson, Interprétation des événements dans l'interaction multimodale. Actes IHM'94, p. 23-28, 1994.