Apport de la Psychologie Cognitive

Minds are simply computers made of meat

Marvin Minsky

Apport de la Psychologie Cognitive

3.1. Introduction

L'intérêt de la psychologie cognitive, à la fois théorique et expérimentale, pour la conception des interfaces homme-machine est maintenant reconnu. Le travail séminal de Card, Moran et Newell [Card 1983] a ouvert la voie à de nombreuses collaborations fructueuses entre psychologues et informaticiens. La psychologie cognitive vise à comprendre et modéliser les mécanismes cognitifs mis en jeu dans les activités humaines. Les activités étudiées par la psychologie cognitive comprennent par exemple la planification et l'exécution d'un plan, l'apprentissage, la mémorisation, la perception de notre environnement.

Nous présentons dans ce chapitre le modèle cognitif Interacting Cognitive Subsystems (ICS) [Barnard 1985], qui offre l'intérêt de prendre en compte les aspects cognitifs de l'interaction d'un utilisateur avec des systèmes informatiques avancés comme les systèmes multimédias ou multimodaux[1]. ICS apporte des éléments précieux pour la conception d'interfaces de communication homme-homme médiatisée en offrant un support théorique pour l'analyse de la perception et la compréhension humaines de différentes sources d'information simultanées, comme par exemple son et image ou bien son et texte. Nous présentons ensuite l'expérimentation "Garden Movie" que nous avons menée en collaboration avec Jon May et Phil Barnard au MRC-Applied Psychology Unit à Cambridge (Grande-Bretagne). Cette expérience nous a permis d'étudier la collaboration entre deux utilisateurs disposant de moyens de communication audio/vidéo et s'est révélée riche d'enseignements pour la conception des systèmes de communication homme-homme médiatisée.

3.2. Sciences Cognitives et Psychologie Cognitive

Il nous semble opportun de rappeler les hypothèses traditionnelles sur lesquelles repose l'ensemble des sciences cognitives. Selon [Andler 1992], le paradigme classique ou "cognitivisme" propose les trois hypothèses fondamentales suivantes :

1. *Le complexe esprit/cerveau est susceptible d'une double description, matérielle ou physique au sens large (la physique intervenant en réalité par le biais des neurosciences), et informationnelle ou fonctionnelle ; ces deux niveaux sont largement indépendants, et le rapport qui s'établit entre eux est à l'image de celui*

[1] Pour une discussion détaillée des mots multimodal et multimédia, voir [Nigay 1994].

qui lie un ordinateur en tant que système physique à la description du même appareil en tant que système de traitement de l'information.

2. *Au niveau informationnel, le système cognitif de l'homme (…) est caractérisé par ses* états *internes ou mentaux et par les* processus *qui conduisent d'un état au suivant. Ces états sont* représentationnels *: ils sont dotés d'un contenu renvoyant à des entités externes (on dit aussi qu'ils sont* sémantiquement évaluables*).*

3. *Les états ou* représentations *internes sont des formules d'un langage interne (ou "mentalais") proche des langages formels de la logique. (…)*

Notons tout d'abord que ces trois postulats, fondamentaux pour les sciences cognitives, se veulent universels : ils sont applicables à tout être humain quelle que soit l'activité humaine considérée. Ils sont donc pertinents pour l'étude de l'interaction homme-machine.

On pourra regretter l'analogie faite dans le premier postulat avec l'ordinateur : en effet, l'ordinateur est un outil créé par l'homme et même s'il s'agit probablement de l'outil le plus complexe imaginé à ce jour, sa complexité et ses capacités sont dans l'ensemble bien inférieures à celle de notre cerveau. Cette analogie a cependant le mérite de clarifier les positions de l'approche "matérialiste" qui étudie les mécanismes biologiques et, à l'opposé, de l'approche qui considère le cerveau comme une "boîte noire" et l'étudie de façon externe.

L'approche cognitiviste est contestée en particulier par le mouvement connexionniste issu de l'Intelligence Artificielle. [Smolensky 1992] distingue le connexionnisme de l'approche reposant sur le "paradigme symbolique". Le paradigme symbolique affirme pouvoir formaliser les structures mentales sans connaître la relation que les structures neuronales sous-jacentes entretiennent avec les structures mentales. Selon Smolensky, les modèles reposant sur le paradigme symbolique ne permettent qu'une approximation grossière des comportements cognitifs. En revanche, l'approche connexionniste vise à décrire exactement le comportement cognitif directement à partir des structures neuronales, sans recourir aux structures mentales (qui incluent les notions de buts, connaissances, perceptions, actions, etc., sur lesquelles reposent largement les théories psychologiques appliquées jusqu'à présent à l'interaction homme-machine). Une approche intermédiaire que propose Smolensky repose sur le "paradigme subsymbolique" qui tente d'établir un pont entre les deux approches.

Pour résumer l'opposition entre les écoles cognitiviste et connexionniste et pour clarifier notre position, nous serions tentés de reprendre la formule provocatrice de Marvin Minsky citée en exergue de ce chapitre en la modifiant : "L'esprit humain *peut être modélisé* comme un ordinateur".

Le modèle psychologique ICS que nous présentons ci-dessous repose, comme toutes les modélisations utilisées classiquement pour l'étude de l'interaction homme-machine, sur le paradigme symbolique et cognitiviste classique. Cependant, nous sommes conscients des possibles approximations que contient un tel modèle, approximations sur lesquelles insistent d'ailleurs ses auteurs. Malgré cette limitation, le modèle ICS permet de prédire et d'expliquer des comportements cognitifs et apporte une contribution déterminante au domaine de l'interaction homme-machine.

3.3. ICS, un modèle cognitif

Un des premiers modèles issu de la psychologie cognitive et qui a été appliqué à l'étude de l'interaction homme-machine est le célèbre modèle du processeur humain [Card 1983]. Ce modèle décompose le système cognitif humain en un ensemble de trois processeurs spécialisés (perceptif, moteur et cognitif) et de mémoires (de travail et à long terme). Comme le reconnaissent eux-mêmes ses auteurs, ce modèle est très simplifié, en particulier en ce qui concerne le système perceptif. Cette simplicité, suffisante pour des interfaces graphiques traditionnelles, est une limitation gênante pour l'étude des interfaces multimodales ou multimédia. En effet, en interagissant avec un système multimodal ou multimédia, l'utilisateur doit percevoir et interpréter des informations non seulement visuelles, mais aussi auditives, tactiles, etc. Le système perceptif monolithique tel qu'il est décrit dans le modèle du processeur humain ne permet pas d'étudier la façon dont, par exemple, nous intégrons des informations visuelles et auditives et parvenons à fusionner les deux sources d'information et à les interpréter comme un ensemble cohérent. Ce cas de figure correspond par exemple, outre l'utilisation de nos sens dans la vie quotidienne, au spectateur de cinéma, mais aussi à l'utilisateur d'un système de communication multimédia ou d'une interface multimodale.

Le modèle Interacting Cognitive Subsystems (ICS) [Barnard 1985] peut dans un premier temps être vu comme un affinement du modèle du processeur humain. Mais ces deux modèles ont des différences fondamentales : en particulier, ICS repose sur une architecture parallèle multi-processus, et non sur un processeur cognitif central ; il n'y a pas dans ICS de mémoire de travail centralisée et banalisée, mais un ensemble de mémoires locales ; d'autre part, ICS ne vise pas à expliquer précisément la nature de l'information traitée par le système cognitif humain, ni les mécanismes précis du

traitement de l'information, mais il tente de construire des modèles approchés des opérations cognitives en s'intéressant particulièrement à l'utilisation des ressources cognitives. Enfin, l'intérêt majeur d'ICS pour notre domaine d'étude est dû au fait qu'il s'intéresse plus aux phénomènes sensoriels qu'au raisonnement, à la différence de beaucoup de modèles cognitifs. Cet aspect d'ICS le rend plus adapté que le modèle du processeur humain à l'étude des interfaces multimodales et multimédia.

3.3.1. Présentation générale du modèle

Le modèle Interacting Cognitive Subsystems (ICS) structure le système de traitement de l'information humain en un ensemble de neuf sous-systèmes (figure 3.1).

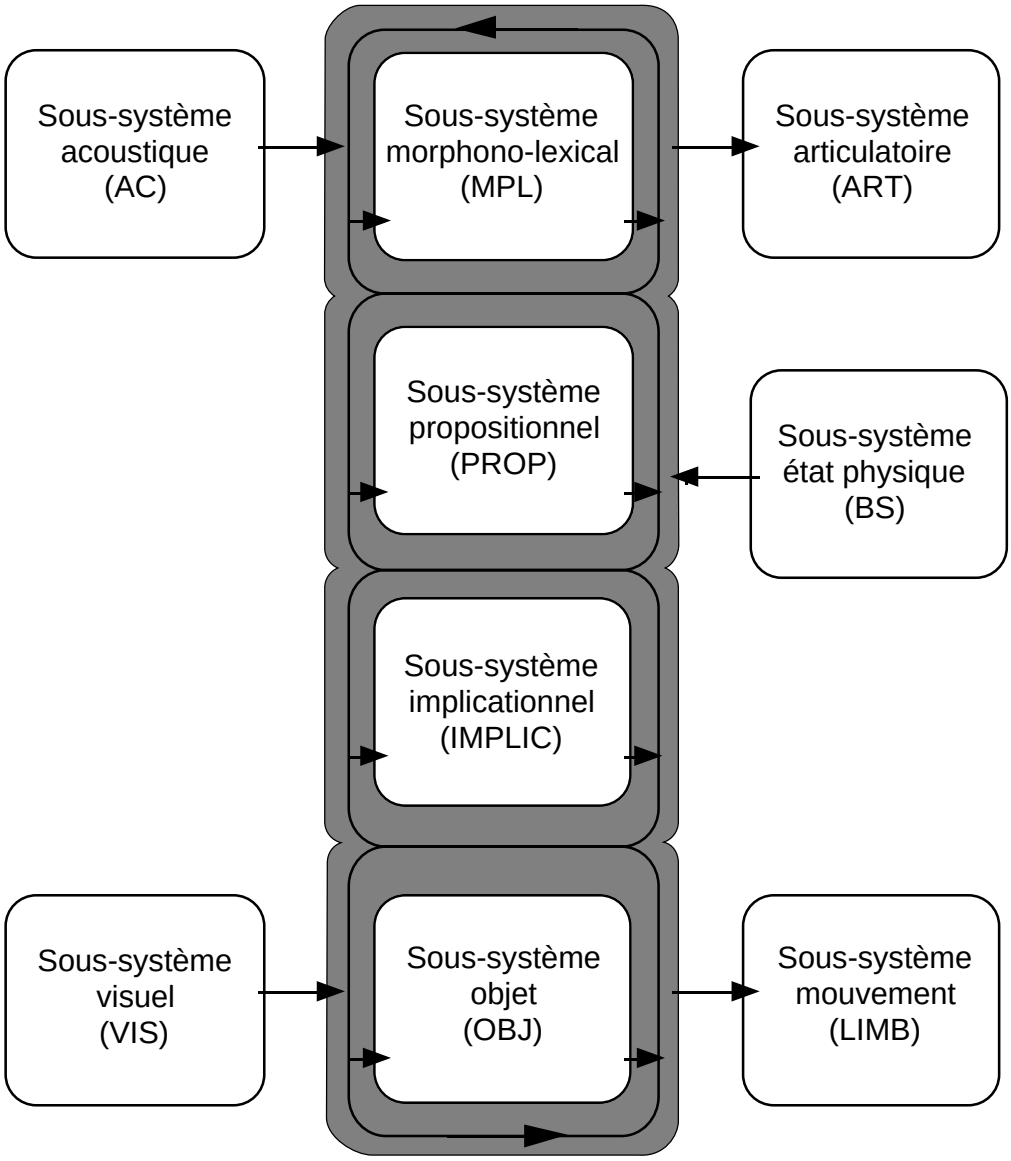

Figure 3.1. Architecture générale du modèle ICS. La partie grisée représente le réseau permettant aux sous-systèmes de communiquer.

Les sous-systèmes perceptifs (visuel, acoustique, état physique) acquièrent l'information en provenance de l'environnement ou du corps. Les sous-systèmes centraux (objet, morphono-lexical, propositionnel, implicationnel) réalisent des transformations de l'information acquise. Ces sous-systèmes sont au cœur des processus d'interprétation et de compréhension, ainsi que des mécanismes du raisonnement. Les sous-systèmes moteurs (articulatoire, mouvement) contrôlent les actions physiques. Tous ces sous-systèmes fonctionnent en parallèle.

Les sous-systèmes centraux sont reliés entre eux par un réseau grâce auquel ils peuvent communiquer. Les sous-systèmes perceptifs peuvent uniquement transférer des informations vers ce réseau et les sous-systèmes moteurs ne peuvent que recevoir des informations en provenance de ce réseau.

Les neuf sous-systèmes qui constituent ICS ont tous la même structure. Nous décrivons d'abord la structure d'un sous-système, puis nous détaillons le rôle de chacun des sous-systèmes et ses interactions possibles avec les autres sous-systèmes.

3.3.2. Description d'un sous-système d'ICS

Un sous-système du modèle ICS est constitué d'un ensemble d'entrées et de sorties, d'une capacité de traitement de l'information, et d'une mémoire locale (figure 3.2).

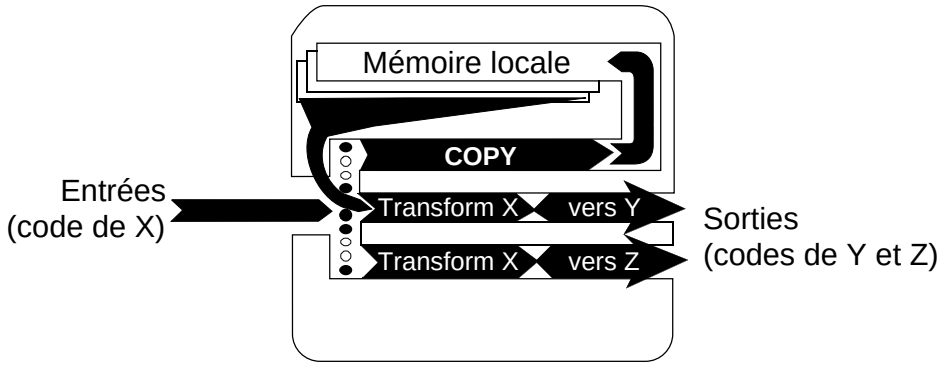

Figure 3.2. Un sous-système X. Le sous-système reçoit en entrée des représentations dans son code et fournit en sortie des représentations dans le code d'autres sous-systèmes (ici Y et Z).

Chaque sous-système dispose d'un code de représentation de l'information qui lui est propre : il ne peut traiter que des informations arrivant à ses entrées qui sont représentées dans son propre code. Il peut en revanche générer des représentations dans le code d'autres sous-systèmes. Par exemple, le sous-système objet acquiert les représentations en provenance du sous-système visuel (ces représentations sont donc exprimées dans le

code du sous-système objet) et les transforme par exemple en représentations exprimées dans le code du sous-système propositionnel.

La capacité de traitement de l'information d'un sous-système est représentée par un ensemble de processus qui opèrent sur les représentations manipulées par le sous-système. Ces processus sont de deux sortes : le processus de recopie locale (COPY) copie systématiquement toute représentation arrivant en entrée du sous-système vers la mémoire locale. Les processus de transformation (TRANSFORM) reçoivent en entrée soit les représentations présentes aux entrées du sous-système, soit des représentations provenant de la mémoire locale. Ces processus transforment les représentations et fournissent en sortie des représentations exprimées dans le code d'autres sous-systèmes. A un instant donné, un seul processus de transformation peut-être actif. Deux processus TRANSFORM d'un même sous-système ne peuvent pas fonctionner en parallèle. En revanche, COPY peut fonctionner en même temps qu'un processus TRANSFORM. Ce mode de fonctionnement montre que les capacités de traitement de l'information d'un sous-système sont limitées et peuvent donc être sujettes à surcharge.

La mémoire locale peut contenir quatre types de représentations :

- *Active Task Record* (ATR) : une représentation que le sous-système vient de traiter et qui est disponible pour une réutilisation immédiate,

- *Experiential Task Record* (ETR) : une représentation qui a été rencontrée dans le passé et qui peut être rappelée si le sous-système reçoit une représentation analogue,

- *Common Task Record* (CTR) : abstraction regroupant plusieurs représentations déjà rencontrées (ETR) et constituée de leur rappel simultané ; les similarités entre les représentations sont combinées et les différences gommées,

- *Entity Property Record* (EPR) : partie d'une représentation déjà rencontrée, élaborée à partir des deux catégories précédentes (CTR et ETR), et qui peut servir à compléter une représentation partielle qui se présente à l'entrée du sous-système.

La mémoire locale d'un sous-système est à rapprocher de la mémoire de travail décrite dans le modèle du processeur humain : le nombre de représentations qu'elle peut contenir est réduit, et la mémoire locale est alimentée par les entrées du sous-système (sous forme d'ATRs) et par la mémoire à long terme (ETRs, CTRs, et EPRs). Le modèle du processeur humain fait aussi intervenir des mémoires très fugitives liées au système

perceptif (*auditory image store* et *visual image store*) qui transfèrent immédiatement leur contenu à la mémoire de travail. Dans ICS, ces mémoires fugitives sont considérées comme faisant partie intégrante de la mémoire de travail (ce sont des ATRs des sous-systèmes acoustique et visuel) et la mémoire de travail est ici distribuée dans l'ensemble des sous-systèmes. La distinction entre les différents types de représentation manipulées par la mémoire locale d'un sous-système d'ICS, et en particulier les CTRs et EPRs permet de mettre en évidence la nature associative de la mémoire humaine. ICS montre aussi que les informations stockées en mémoire à long terme sont représentées dans le code du sous-système concerné et ne peuvent donc être rappelées que par ce sous-système. Enfin les EPRs montrent qu'une représentation partielle présentée à un sous-système peut être complétée grâce à l'expérience accumulée par ce sous-système.

Les neuf sous-systèmes du modèle ICS fonctionnent tous comme décrit ci-dessus. Mais chaque sous-système est spécialisé dans le traitement d'une partie de l'activité cognitive. Nous décrivons ci-dessous le rôle joué par chacun d'entre eux.

3.3.3. Les sous-systèmes perceptifs

Les sous-systèmes perceptifs sont les sous-systèmes visuel (VIS), acoustique (AC) et état physique (BS pour *body state*). Chacun de ces sous-systèmes reçoit de l'information du monde physique par nos récepteurs sensoriels et transforme ces représentations de l'information à l'intention des sous-systèmes centraux.

Plus précisément, le sous-système visuel transforme les informations vues (par exemple, couleur, contour, luminosité) en représentations à l'intention du sous-système objet (OBJ). Ce processus de transformation, que l'on note VIS $\Rightarrow$ OBJ, fournit en sortie du sous-système VIS des informations exprimées dans le code de OBJ (par exemple, formes des objets vus).

Le sous-système acoustique opère sur les informations auditives (par exemple, hauteur des sons, rythme, timbre). Si les sons entendus sont par exemple des paroles, le sous-système acoustique va fournir en sortie des représentations pour le sous-système morphono-lexical (MPL) par la transformation AC $\Rightarrow$ MPL.

De façon similaire, le sous-système BS va acquérir des informations en provenance du corps (par exemple, retour d'information tactile, température, texture) et transformer ces informations à l'intention des sous-systèmes centraux.

En règle générale, chacun des sous-systèmes perceptifs a un partenaire privilégié parmi les sous-systèmes centraux, comme OBJ pour VIS ou MPL pour AC. Les transformations correspondantes abstraient l'information du niveau perceptif vers un niveau intermédiaire qui permettra ensuite l'interprétation des informations perçues.

3.3.4. Les sous-systèmes centraux

Quatre sous-systèmes forment les sous-systèmes centraux : les sous-systèmes propositionnel (PROP) et implicationnel (IMPLIC) qui sont parfois appelés *moteur central de la cognition* et les sous-systèmes objet (OBJ) et morphono-lexical (MPL) dont nous avons vu au paragraphe précédent qu'un de leurs rôles est de fournir une représentation abstraite des perceptions.

Le sous-système objet (OBJ) peut être vu comme ayant la charge de "l'imagerie mentale". C'est là par exemple que se forment les images mentales qui ne sont pas directement perçues mais qui sont générées par le système cognitif. Ce sous-système va en général, à partir des représentations fournies par VIS à ses entrées, élaborer des propositions, c'est-à-dire des relations sémantiques entre entités, à l'intention du sous-système PROP. Cette transformation correspond à l'identification d'un objet et à l'établissement de relations avec les objets qui l'entourent (par exemple, l'objet vu est un stylo et il est posé sur une table). Un cas particulier notable est celui de la perception du texte écrit : dans ce cas, OBJ va élaborer une représentation à l'intention du sous-système MPL qui est en charge du langage.

Le sous-système morphono-lexical (MPL) est lié principalement au langage. On peut décrire les représentations qu'il manipule comme "ce que nous entendons dans notre tête". Ce sous-système a la connaissance des mots et des formes lexicales. Les représentations fournies à ses entrées par AC pour le langage parlé ou par OBJ (après transformation VIS $\Rightarrow$ OBJ) pour le langage écrit sont transformées en propositions pour le sous-système PROP. Cette transformation extrait le sens des formes langagières entendues ou lues. C'est aussi MPL qui permet la génération de formes langagières. Pour cette génération, le sous-système PROP va transformer une proposition en une représentation pour MPL. Cette représentation va ensuite être transformée par MPL à destination du sous-système articulatoire qui va l'exprimer sous forme de langage parlé ou écrit.

Le sous-système propositionnel (PROP) manipule des représentations correspondant à des propositions. Ces propositions peuvent être représentées sous forme de prédicats logiques. Les entrées de PROP sont principalement alimentées par les sous-systèmes

OBJ, MPL et IMPLIC. Les représentations circulant dans PROP correspondent au contenu sémantique des informations perçues par les sous-systèmes perceptifs mais aussi des propositions engendrées par IMPLIC.

Le sous-système PROP a pour partenaire privilégié le sous-système IMPLIC. IMPLIC reçoit en entrée des représentations qui proviennent principalement de PROP et génère de nouvelles représentations à l'intention de PROP. Pour simplifier et très schématiquement, on peut assimiler le couple PROP/IMPLIC à un système expert où PROP serait la base de connaissances et où IMPLIC jouerait le rôle du moteur d'inférence. Le cycle PROP $\Rightarrow$ IMPLIC et IMPLIC $\Rightarrow$ PROP modélise le raisonnement tel que l'entend le sens courant.

Le sous-système PROP génère aussi des représentations dans le code des sous-systèmes OBJ et MPL qui vont nous permettre d'agir sur l'environnement par l'intermédiaire des sous-systèmes moteurs.

3.3.5. Les sous-systèmes moteurs

Les deux sous-systèmes moteurs contrôlent nos actions sur le monde physique. Le sous-système articulatoire (ART) est spécialisé dans l'émission du langage qu'il soit articulé, mais aussi écrit ou tapé au clavier. Le sous-système mouvement (LIMB) se charge de tous les autres mouvements du corps. Notons qu'en règle générale, le sous-système LIMB reçoit des représentations du sous-système objet, traduisant le fait que nous préparons visuellement un geste avant de l'effectuer. Toutefois dans le cas particulier des actes réflexes (par exemple, on retire sa main rapidement lorsque l'on touche un objet brûlant), le sous-système état physique communique directement avec le sous-système mouvement. C'est l'un des cas où deux sous-systèmes peuvent communiquer sans mettre en jeu un sous-système central (un autre cas survient dans la même circonstance : on criera "Aïe!", ce qui fera aussi intervenir une communication directe BS $\Rightarrow$ ART).

3.4. Application à la communication homme-homme médiatisée

Notre présentation d'ICS met en évidence deux intérêts majeurs du modèle :

- ICS est un modèle qualitatif ; contrairement au modèle du processeur humain qui s'intéresse à la performance de l'utilisateur et aux aspects quantitatifs de la cognition (temps de réaction, limitations de capacité de la mémoire à court terme, etc.), ICS met en avant les aspects qualitatifs de la cognition en termes d'utilisation de ressources. Les différents sous-systèmes peuvent être considérés

comme un ensemble de ressources cognitives et ICS peut indiquer et justifier qu'une ressource risque d'être surchargée dans une circonstance donnée.

- ICS affine le système perceptif en trois sous-systèmes : VIS, AC et BS. De plus il met en évidence différents niveaux d'abstraction et modélise l'interprétation et la compréhension des informations perçues comme une suite de transformations, chaque transformation élevant le niveau d'abstraction. Pour la compréhension d'une information visuelle non textuelle par exemple, un parcours typique des représentations dans l'architecture ICS sera VIS $\Rightarrow$ OBJ puis OBJ $\Rightarrow$ PROP, suivi éventuellement d'un ou plusieurs cycles PROP $\Rightarrow$ IMPLIC / IMPLIC $\Rightarrow$ PROP.

Ces deux aspects d'ICS en font un bon candidat en tant que théorie psychologique adaptée aux nouveaux médias et combinaisons de médias utilisés aujourd'hui dans les interfaces homme-machine, et en particulier dans les interfaces des systèmes de communication homme-homme médiatisée. Nous nous sommes particulièrement intéressés à l'aspect perceptif de l'interaction dans le cadre de ces systèmes. Nous verrons qu'ICS apporte des éléments utiles pour l'étude de l'intégration et de la combinaison des nouveaux médias. ICS permet en particulier de raisonner sur la surcharge cognitive qui peut provenir de la présentation simultanée à l'utilisateur de plusieurs sources d'information. Le modèle permet aussi de déterminer les conditions que les différentes sources d'information doivent respecter afin d'être interprétées comme un tout cohérent. Ces points sont maintenant illustrés au paragraphe suivant.

3.4.1. Informations combinant plusieurs médias : l'évidence expérimentale

Dans de multiples situations de la vie courante, nous sommes amenés à combiner des informations provenant de différents sens. Parmi les possibilités, la combinaison d'informations visuelles et auditives est privilégiée. Lorsque nous discutons, lorsque nous assistons à la projection d'un film, nous construisons la signification de notre conversation ou du film à partir d'informations perçues à la fois par la vue et l'ouïe. Lorsque des informations visuelles et auditives sont présentées simultanément, elles doivent cependant respecter certains critères pour faire sens.

Un exemple familier est celui d'un film dont la bande-son est désynchronisée : si l'écart entre l'image et la bande-son est trop important, le film devient incompréhensible. Dans certains cas, une incohérence entre les informations auditives et visuelles peut même conduire à une interprétation erronée. L'effet McGurk [McGurk 1976] est une bonne

illustration de ce phénomène : on présente à un sujet une bande vidéo montrant un orateur. L'orateur prononce la syllabe "ga" mais la bande-son a été réenregistrée avec la syllabe "ba". Lors des tests, 98% des sujets adultes rapportent entendre la syllabe "da".

Ces exemples montrent que lorsque des informations visuelles et auditives sont présentées à l'aide de moyens techniques, elles doivent satisfaire certaines propriétés afin d'être correctement perçues et interprétées. Ce point est particulièrement crucial pour la conception d'outils de communication audio/vidéo. En effet, les techniques disponibles à l'heure actuelle ne permettent pas en général de transmettre une communication audio/vidéo avec une qualité équivalente à celle du cinéma ou de la télévision. Il est alors nécessaire de rechercher des compromis. En nous donnant un fondement théorique pour analyser les mécanismes perceptifs, ICS nous permet de déduire des principes pour guider ces compromis. Nous avons appliqué les propriétés issues de ces principes dans la conception du système de communication UserLink décrit au chapitre 8 dans la partie plus technique de ce mémoire.

3.4.2. Critères de combinaison d'informations dans ICS

ICS distingue plusieurs types de combinaisons d'informations perçues et donne des critères pour que ces combinaisons soient possibles. La combinaison d'informations distinctes est principalement effectuée par les sous-systèmes centraux (MPL, IMPLIC, PROP ou OBJ). Les informations susceptibles d'être combinées peuvent provenir soit des sous-systèmes perceptifs, soit des sous-systèmes centraux. Différentes sources d'informations perçues simultanément peuvent être combinées à différents niveaux cognitifs : au niveau capteur, dans les sous-systèmes perceptifs, dans les sous-systèmes centraux.

3.4.2.1. Combinaison au niveau capteur

Le tout premier niveau de combinaison est celui des capteurs physiques : les limitations intrinsèques de nos sens ou certaines de leurs particularités nous permettent une perception globale d'informations discrètes. Par exemple, l'image sur un écran est perçue comme un tout alors qu'elle est composée d'un ensemble de points. La persistance rétinienne nous fait voir l'image cinématographique comme un mouvement continu alors qu'elle est constituée de 24 images par seconde. Notons que ces limitations de nos capteurs sensoriels sont déjà exploitées par certains systèmes de compression. La norme JPEG de compression des images photographiques dégrade volontairement l'image originale en éliminant des détails que notre sens de la vue n'est pas capable de discerner. De façon similaire, la norme MPEG audio pour la compression du son tient compte de particularités physiologiques de l'audition pour éliminer certains détails. En d'autres

termes, le principe psychologique de limitation de nos capacités sensorielles permet de s'affranchir de la propriété d'intégrité (définie au chapitre 4) pour certains types d'informations.

3.4.2.2. Combinaison dans les sous-systèmes perceptifs

La combinaison d'informations perçues sur un même canal sensoriel se produit également au niveau des sous-systèmes perceptifs. Sur un même canal, nous pouvons combiner différentes informations à condition qu'elles soient liées par une structure commune. Barnard [Barnard 1992] cite l'exemple d'un vol d'oiseaux : on voit l'ensemble des oiseaux comme un tout, et non chaque oiseau individuellement. D'autres théories psychologiques, en particulier la Gestalt théorie peuvent ici apporter une aide. Cette théorie définit en effet un certain nombre de principes (telles la similarité, la symétrie, la proximité, etc.) qui ont une influence sur la façon dont nous percevons des informations visuelles (figure 3.3).

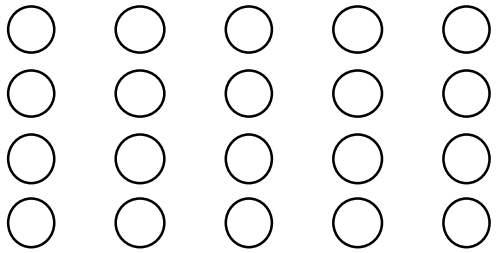

Figure 3.3. Application de la Gestalt théorie. La disposition de ces cercles nous conduit à les voir comme "cinq colonnes de cercles" plutôt que comme "quatre lignes de cercles". La proximité des cercles fait apparaître une propriété émergente de la figure : une disposition en colonnes.

La combinaison au niveau perceptif (c'est-à-dire dans les sous-systèmes perceptifs et sur un seul canal sensoriel) peut déjà avoir des conséquences importantes sur les niveaux supérieurs. En effet, le modèle ICS prédit que la combinaison d'informations provenant de différents canaux sensoriels dans les niveaux supérieurs ne peut avoir lieu que si les flots d'information issus des sous-systèmes perceptifs sont cohérents. Un contre-exemple est donné par les chaînes de télévision dites "mosaïque" qui affichent simultanément plusieurs chaînes sur un même écran (par exemple 16 pour la mosaïque des chaînes du câble à Grenoble). Si le son correspondant à la mosaïque est celui d'une des chaînes affichées, il est très difficile, voire impossible d'identifier à quelle image correspond le son entendu. L'attention du spectateur passant d'une image à l'autre, le flot fourni par le sous-système VIS n'est pas cohérent avec AC et par la suite, la combinaison entre les informations visuelles et auditives ne peut s'effectuer dans les sous-systèmes centraux.

3.4.2.3. Combinaison dans les sous-systèmes centraux

A un niveau cognitif supérieur, dans les sous-systèmes centraux de l'architecture ICS, la combinaison d'informations par un sous-système peut avoir lieu si les informations se présentant aux entrées du sous-système sont :

- exprimées dans le code du sous-système : cette contrainte est nécessaire pour que le sous-système puisse traiter les informations,

- cohérentes à l'intérieur d'un même flot d'informations, comme expliqué ci-dessus,

- cohérentes entre deux flots d'informations : si par exemple, la bande-son d'un film ne correspond pas à l'image, la combinaison est impossible ou donne lieu à une interprétation erronée. C'est le cas de l'effet McGurk exposé plus haut,

- simultanées : les flots d'informations doivent être présents simultanément aux entrées du sous-système, avec une certaine marge de tolérance qui peut être calculée expérimentalement et qui dépend des informations considérées. L'exemple évoqué plus haut de la désynchronisation de la bande-son d'un film est un cas typique où les informations ne sont pas simultanées et gênent, voire empêchent la combinaison.

Ces conditions de combinaison peuvent être vues comme des principes issus de la psychologie. Nous en déduisons des propriétés des médias au chapitre 4, comme la synchronisation et la régularité des flots. La combinaison d'informations peut se produire dans chacun des sous-systèmes centraux de l'architecture ICS sous réserve de respecter les critères énoncés ci-dessus. La combinaison d'informations visuelles et auditives par exemple peut avoir lieu dans le sous-système MPL (via OBJ pour les informations visuelles). Dans les autres sous-systèmes centraux ont lieu des combinaisons d'information à plus haut niveau, après abstraction des informations sensorielles par les sous-systèmes PROP et IMPLIC.

PROP est le lieu de la combinaison d'informations sensorielles (via OBJ et MPL) et sémantiques (via IMPLIC). Nous décrivons plus loin l'expérimentation Garden Movie qui fait intervenir une combinaison au niveau du sous-système PROP. Dans le sous-système IMPLIC a lieu la combinaison d'informations perçues et sémantiques. La particularité de la combinaison à ce niveau est que les informations perçues parvenant à IMPLIC arrivent directement des sous-systèmes perceptifs sans passer ni par OBJ, ni par

MPL, ni par PROP. C'est le cas des informations qui sont ressenties mais pas interprétées : un bruit violent, par exemple, mais aussi toutes les informations que nous percevons "inconsciemment". C'est le cas en particulier des attitudes physiques de nos interlocuteurs, des gestes qui accompagnent les discours, bref d'une grande partie de la communication non-verbale.

Enfin un cas particulier de combinaison peut également se produire dans le sous-système OBJ : des informations sémantiques issues de PROP vont être combinées avec les informations visuelles afin de guider l'interprétation. C'est le cas par exemple lorsque nous sommes dans un environnement familier plongé dans la pénombre et que nous devinons un objet parce que nous savons qu'il est là.

En appliquant ICS à la perception d'informations complexes, il est donc possible de déterminer des critères afin que des informations combinées puissent être perçues et interprétées correctement. ICS montre aussi que la combinaison d'informations est un phénomène complexe qui peut intervenir à différents niveaux du système cognitif. A travers quelques exemples liés à la perception d'informations visuelles et auditives, nous avons montré qu'ICS peut aider à réfléchir sur la présentation de ces informations en générant des principes et ainsi guider des choix de conception.

Au paragraphe suivant, nous présentons l'expérimentation Garden Movie qui a permis d'étudier un cas de combinaison d'informations au niveau du sous-système PROP.

3.5. Expérimentation avec ICS : Garden Movie

J'ai participé à l'expérimentation Garden Movie lors d'un séjour de trois mois au MRC-Applied Psychology Unit à Cambridge (Grande-Bretagne), en collaboration avec Jon May et Phil Barnard. J'ai pris part à la conception et au dépouillement des résultats de l'expérimentation. J'ai aussi réalisé l'application support (avec MacroMedia Director [MacroMedia 1993]) ; j'ai également testé la plupart des sujets.

Pour ma part, la motivation de cette collaboration était double :

- je souhaitais étudier les processus psychologiques mis en œuvre dans la coopération entre utilisateurs d'un système permettant la communication audio/vidéo,

- je souhaitais aussi me familiariser avec la démarche expérimentale en psychologie, dans le but de l'appliquer aux tests d'utilisabilité avec la plate-forme NEIMO (voir chapitre 5).

La découverte d'un nouvel environnement et l'appréhension "de l'intérieur" d'une discipline dans laquelle j'étais naïf a été une expérience extrêmement enrichissante. Ce paragraphe décrit les motivations de l'expérimentation, le dispositif expérimental mis en place. Il présente brièvement les résultats obtenus, les discute et les explique, et enfin tire les leçons de cette collaboration.

3.5.1. Motivations de l'expérimentation

Dans la communication humaine face à face, une grande partie des informations échangées l'est sous forme non-verbale : gestes, attitudes, contact visuel, etc. Mais lorsque la communication se fait via un lien audio/vidéo, des difficultés apparaissent [Heath 1992]. Les gestes par exemple sont moins bien compris que dans la communication face à face. Une raison en est l'absence d'un cadre de référence commun. L'usage traditionnel de la vidéo, qui simule la communication face à face, semble laisser penser aux utilisateurs que leur cadre de référence est parallèle à celui des autres participants alors qu'il est en fait en vis-à-vis. Les systèmes de vidéoconférence commerciaux comme ceux décrits dans [Reinhardt 1993] offrent aux utilisateurs une disposition face à face et la possibilité de partager des vues sur des documents. Dans ce type de systèmes, un télépointeur ou des annotations partagées sont souvent utilisés pour remplacer la communication gestuelle.

Les systèmes de dessin partagé sont un champ d'investigation intéressant pour l'étude de la communication gestuelle via vidéo. Commune [Minneman 1991] permet aux utilisateurs de dessiner sur une surface commune horizontale et de se voir de face sur des écrans vidéo. Ce système utilise des télépointeurs mais les auteurs ont constaté que des gestes étaient aussi échangés par les utilisateurs. Dans ClearBoard [Ishii 1992], deux utilisateurs peuvent dessiner simultanément sur un espace partagé en voyant l'image de l'autre participant superposée sur le dessin. Une caractéristique originale de ClearBoard est qu'il renverse l'image du participant. Les deux utilisateurs ont ainsi une référence commune de la droite et de la gauche du dessin et peuvent utiliser des gestes pour montrer une partie du dessin. VideoWhiteboard [Tang 1991], précurseur de ClearBoard, adoptait une approche similaire mais projetait l'ombre des participants au lieu de leur image. De même que dans ClearBoard, l'ombre projetée était inversée. Cependant, pour ces deux systèmes, l'efficacité de cette technique comparée à la disposition classique face à face n'a pas été étudiée.

Enfin, [Gaver 1993] propose une approche originale pour une tâche où deux utilisateurs coopèrent pour arranger des objets d'une maison de poupées miniature. La disposition choisie utilise plusieurs caméras qui donnent aux utilisateurs le choix de plusieurs vues de la maison de poupées et de leur partenaire. L'expérimentation montre que les possibilités de choix des caméras rend difficile l'établissement d'un cadre de référence commun entre les utilisateurs. Mais un résultat intéressant de cette étude indique que les utilisateurs préfèrent une vue qui montre l'espace de travail commun plutôt que la vue face à face. Ils peuvent ainsi créer un espace de référence partagé centré sur la tâche.

3.5.2. Dispositif expérimental

Le but de notre expérimentation "Garden Movie" est d'étudier l'influence de plusieurs paramètres sur la réalisation coopérative d'une tâche de dessin par deux utilisateurs dans un environnement permettant des communications audio et vidéo. Pour des raisons pratiques et de rigueur expérimentale, l'un des utilisateurs est simulé, l'autre est le sujet testé. Le sujet voit sur son écran un espace de travail (similaire à MacDraw) de trois cases sur trois qu'il est censé partager avec un autre utilisateur. A côté de cet espace de travail, une fenêtre vidéo montre l'utilisateur avec lequel il coopère (figures 3.4 et 3.5).

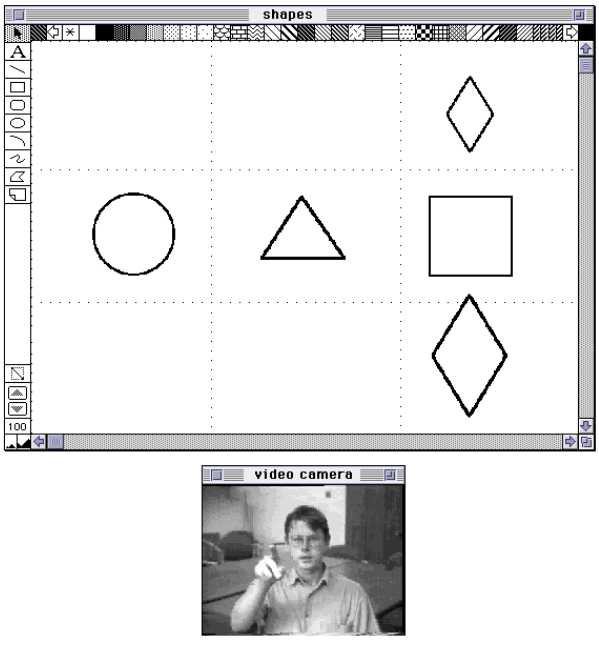

Figure 3.4. Une copie d'écran de l'expérimentation Garden Movie. On présente ici la vue de face et l'espace de travail est en configuration verticale (c'est-à-dire sans perspective).

Le collègue présenté dans la fenêtre vidéo (en fait un enregistrement) demande au sujet de déplacer un objet de l'espace de travail vers une nouvelle position. Il accompagne sa

demande de gestes désignant l'objet à déplacer et sa nouvelle position. Le collègue est censé voir sur son écran exactement le même espace de travail. Un *essai* se déroule de la façon suivante :

- au début de l'essai, le sujet a sur son écran une configuration analogue à celle de la figure 3.4,
- la séquence vidéo est présentée au sujet ; le collègue demande au sujet de déplacer deux objets et accompagne la désignation des objets et de leurs nouvelles positions de gestes de la main droite,
- lorsque la séquence est terminée, elle est remplacée par un bouton “Play again”. Ce bouton permet au sujet de revoir la séquence vidéo,
- le sujet exécute la demande du collègue en déplaçant les objets avec la souris et signale qu’il est prêt à passer à l’essai suivant en appuyant sur un bouton “OK”. Le bouton “OK” n’apparaît que lorsque la séquence vidéo est entièrement jouée.

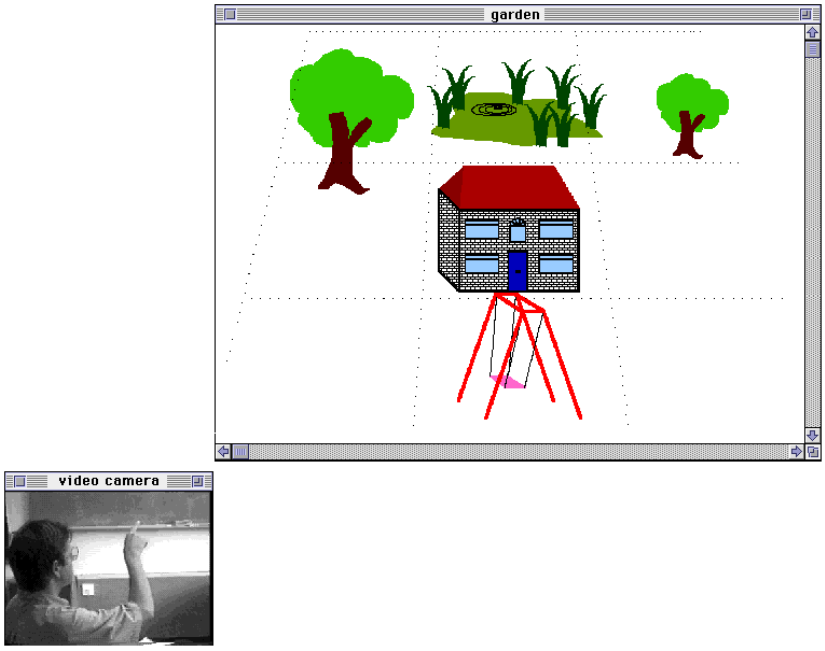

Figure 3.5. Une copie d'écran de l'expérimentation Garden Movie. On présente cette fois la vue “par-dessus l’épaule” et l’espace de travail est en configuration horizontale (c’est-à-dire avec perspective).

Dans l’expérimentation, les effets de trois variables ont été étudiés : la position de la caméra, la disposition de l’espace de travail, et l’usage des déictiques. Trois positions de

caméra ont été étudiées : une vue de face, une vue de face “miroir” et une vue “par-dessus l’épaule”. La vue de face telle que celle présentée sur la figure 3.4 reproduit la disposition de deux personnes face à face, par exemple de part et d’autre d’une table. Elle reproduit donc une situation réelle ; cependant, les deux utilisateurs ayant la même vue du document, un objet situé à gauche dans l’espace de travail sera désigné par le collègue sur sa propre gauche, donc à droite sur l’image vidéo observée par le sujet. La vue miroir est une première tentative pour résoudre ce problème de latéralisation : la vue miroir présente la vue de face transformée par une simple symétrie horizontale. Un objet situé à gauche de l’espace de travail sera désigné par le collègue par un geste sur la gauche de l’image vidéo. Enfin la troisième vue étudiée est donnée par une caméra située derrière l’épaule de l’utilisateur. Cette vue est une autre tentative de solution au problème de latéralisation : les deux utilisateurs regardent la figure selon un point de vue commun ; on souhaite ainsi créer un espace de travail partagé. Afin de renforcer cette illusion d’espace partagé, la fenêtre vidéo est placée dans ce cas en bas à gauche de l’espace de travail en sorte que les gestes du collègue semblent désigner l’espace de travail du sujet.

L’influence de la disposition de l’espace de travail a également été étudiée : deux configurations ont été utilisées. La première (cf. figure 3.4) présente des figures géométriques sur le plan vertical de l’écran. La seconde (cf. figure 3.5) présente des objets que l’on peut trouver dans un jardin (d’où le nom de l’expérimentation !) et grâce à une illusion de profondeur, donne l’impression que les objets sont disposés sur un plan horizontal.

Enfin, l’utilisation de références déictiques a été étudiée : les instructions données par le collègue peuvent soit être explicites, soit comporter des déictiques. Une instruction explicite comporte le nom des objets à déplacer et indique l’endroit où les déplacer relativement à un autre objet. Par exemple, pour la figure 3.4, “déplace le petit losange sous le cercle” est une instruction explicite. Les déictiques ont été utilisés pour référencer les objets à déplacer (comme dans “*cet* objet”, “*ceci*”) et les relations spatiales dans l’espace de travail (comme “de *ce* côté du cercle”). La figure 3.6 récapitule les variables de l’expérimentation.

Vue	**Espace de travail**	**Déictique**
de face	horizontal (jardin)	non (explicite)
par-dessus l’épaule	vertical (MacDraw)	oui (objet ou spatial)
de face miroir		

Figure 3.6. Les variables de l’expérimentation Garden Movie. Les noms des variables sont indiqués en haut de chaque colonne et les valeurs possibles en-dessous.

Une expérimentation avec un sujet est divisée en trois phases qui correspondent aux trois positions de caméra possibles (face, face miroir, derrière l'épaule) (figure 3.7).

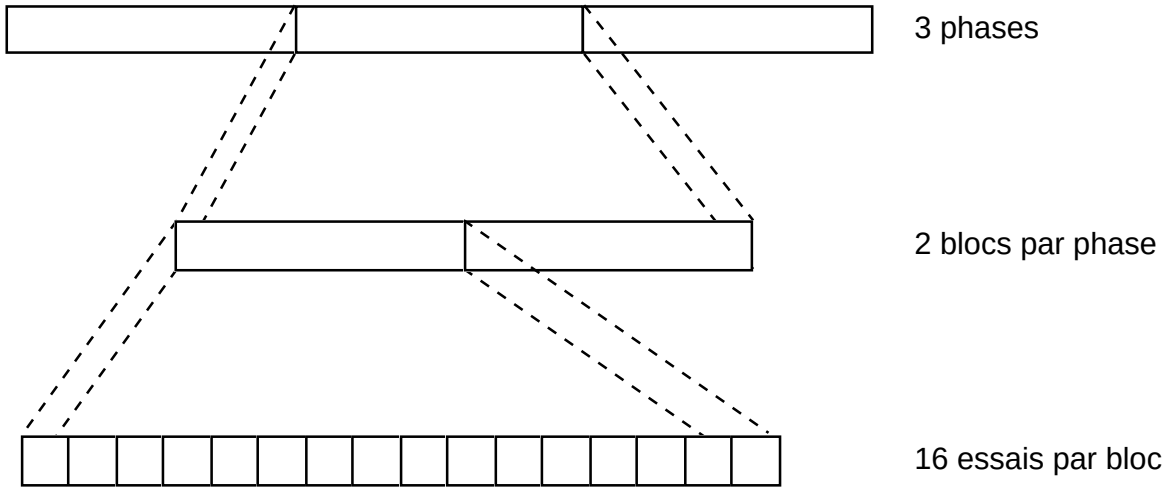

Figure 3.7. La structure d'une session d'expérimentation Garden Movie. Un essai correspond à une configuration comme en montrent les figures 3.4 et 3.5.

Lors d'une phase, la position de la caméra est maintenue constante. Chaque phase est découpée en deux blocs de seize essais chacun. Chaque bloc correspond à une configuration différente de l'espace de travail (verticale, horizontale). Enfin, dans chaque bloc la moitié des essais fait intervenir des instructions explicites, l'autre moitié fait intervenir des instructions comportant des déictiques (dans chaque bloc, quatre déictiques portant sur les objets et quatre déictiques portant sur les relations spatiales).

Lors d'une expérimentation, trois types d'informations ont été collectées : l'utilisation du bouton "Play again", les erreurs dans l'exécution des instructions, et le temps d'exécution de chaque essai. L'information concernant l'utilisation du bouton "Play again" permet de déterminer combien de fois chaque séquence vidéo a été vue par le sujet. (Nous indiquions aux sujets qu'ils pouvaient utiliser le bouton "Play again" autant de fois qu'ils le voulaient tant que les instructions du collègue "n'étaient pas claires".) Les erreurs dans l'exécution des instructions sont des erreurs de placement des objets (par exemple, objet placé à droite alors que l'instruction donnée par le collègue demandait de le placer à gauche). Le temps d'exécution de l'essai est calculé à partir du moment où la dernière visualisation de la séquence s'arrête et jusqu'au moment où le sujet indique qu'il a terminé l'essai en appuyant sur le bouton "OK".

Lors de l'expérimentation, vingt-quatre sujets choisis au hasard dans le "panel" de sujets de MRC-APU ont été testés. L'expérimentation comportant trois variables pouvant prendre douze valeurs différentes (3x2x2) et les différents blocs de l'expérimentation

étant présentés dans un ordre différent pour chaque sujet, douze sujets suffisent pour couvrir tous les cas possibles. Nous avons testé deux séries de douze sujets pour augmenter la fiabilité de nos résultats, soit treize femmes et onze hommes d'âge moyen 31,2 ans. Les instructions lues aux sujets sont présentées en annexe B.

3.5.3. Résultats

On trouvera dans [Barnard 1994] la description détaillée des résultats de l'expérimentation. Nous nous contenterons ici d'en donner la synthèse.

Le taux d'erreur est défini comme le pourcentage d'essais dans lesquels les sujets ont fait une erreur de placement d'objet. Le taux de "replay" est le pourcentage d'essais dans lesquels les sujets ont redemandé à voir la séquence au moins une fois. Ces deux taux ont été étudiés en fonction des variables de l'expérimentation. Les résultats font apparaître que la disposition de l'espace de travail a une influence limitée sur ces taux ou sur le temps nécessaire à la réalisation de l'essai. En revanche, la position de la caméra et l'utilisation des déictiques ont une influence significative. Nous donnons d'abord les résultats globaux, puis nous approfondissons les résultats concernant l'intelligibilité des déictiques. Lorsque les instructions données sont explicites (figure 3.8), le taux d'erreur et le taux de replay ont des valeurs négligeables (proches de 5%), quelle que soit la vue présentée. Pour les instructions contenant des déictiques, le taux d'erreur moyen est proche de 30% et le taux de replay proche de 25%.

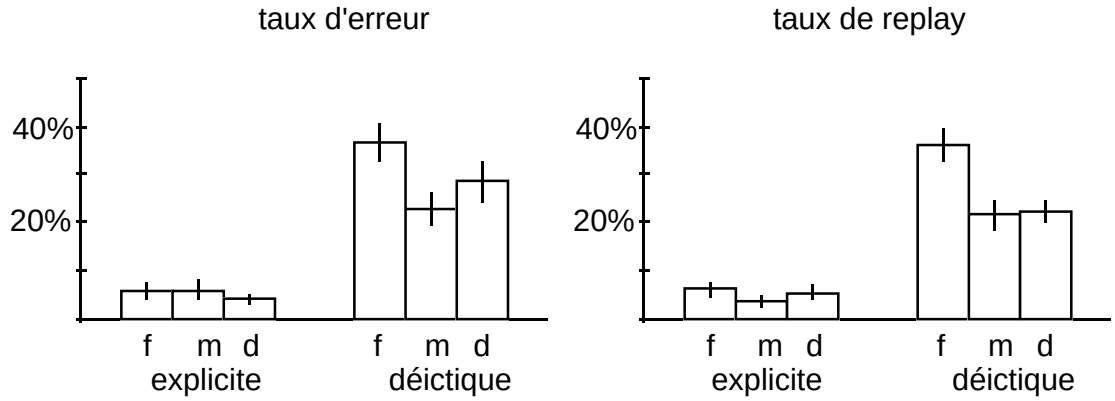

Figure 3.8. Taux d'erreur et de replay suivant les instructions (explicites ou déictiques) et les positions de caméra (f pour face, m pour miroir, d pour derrière l'épaule).

L'analyse des temps d'exécution d'un essai montre également une nette différence suivant que les instructions comportent des déictiques ou non : le temps moyen d'exécution d'un essai est de 8,27 secondes lorsque les instructions sont explicites, et de 9,20 secondes avec des déictiques.

Nous nous intéressons maintenant uniquement aux instructions comportant des déictiques (figure 3.9).

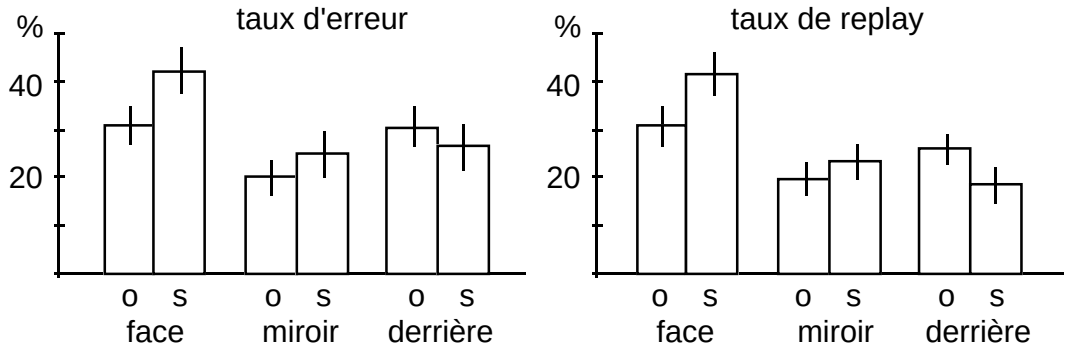

Figure 3.9. Taux d'erreur et de replay suivant les positions de la caméra et les types de déictiques (o pour objet, s pour spatial).

En fonction de la vue présentée, le taux d'erreur et de replay est plus important pour la vue de face (proche de 40%) que pour les deux autres (20% à 30%). Dans les cas de la vue de face et de la vue de face miroir, les taux d'erreur et de replay sont moins importants lorsque le déictique désigne un objet que lorsqu'il désigne une relation spatiale. Dans le cas de la vue derrière l'épaule, les déictiques spatiaux donnent lieu à de plus faibles taux que les déictiques d'objets.

En ce qui concerne les temps d'exécution d'un essai, un essai comportant un déictique spatial est effectué en sensiblement plus de temps (9,49 secondes) qu'un essai avec un déictique d'objet (8,89 secondes).

Dans les cas des instructions contenant des déictiques désignant des relations spatiales, l'analyse fait apparaître une différence significative suivant le type de relation spatiale. Suivant la configuration de l'espace de travail (vertical plan ou horizontal avec effet de perspective), les relations spatiales sont exprimées différemment. Si les relations horizontales "à gauche de" et "à droite de" sont les mêmes dans les deux cas, les relations spatiales verticales "au-dessus de" et "au-dessous de" sont utilisées dans l'espace plan alors que ces relations deviennent "derrière" et "devant" pour l'espace avec effet de perspective. Les résultats montrent que les taux d'erreur et de replay sont nettement plus importants pour les relations spatiales verticales que pour les relations spatiales horizontales (figure 3.10).

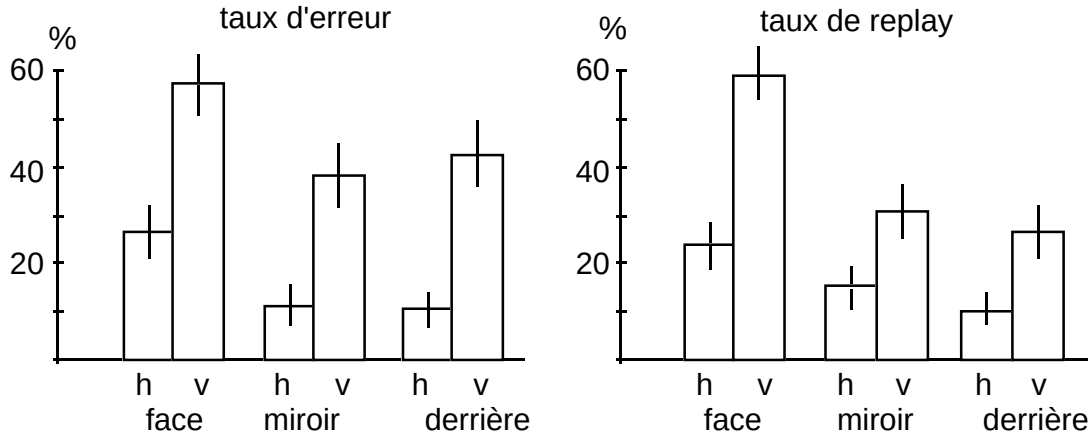

Figure 3.10. Taux d'erreur et de replay suivant les positions de caméra et les types de relations spatiales (h pour horizontale, v pour verticale).

L'analyse des temps d'exécution d'un essai montre que le temps d'exécution est plus important pour les relations spatiales verticales (9,87 secondes) que pour les relations spatiales horizontales (9,06 secondes).

3.5.4. Discussion

Les résultats de l'expérience montrent que les différentes variables de l'expérimentation ont une influence sur les erreurs, les replays et le temps d'exécution. Nous analysons ci-après les résultats obtenus et donnons une interprétation avec le modèle ICS des mécanismes cognitifs mis en œuvre.

Les résultats font clairement apparaître que les instructions explicites donnent lieu à moins d'erreur et sont exécutées plus rapidement que les instructions comportant des références déictiques. Cependant, les taux d'erreur et de replay, bien que faibles, ne sont pas nuls dans le cas des instructions explicites. Ce fait met en évidence que la tâche demandée aux sujets est complexe même sans déictique. Le sujet doit en effet partager son attention visuelle entre la séquence vidéo et l'espace de travail. On peut se demander si des instructions uniquement orales auraient facilité la tâche du sujet et donné de meilleurs résultats pour les essais sans déictique.

Sur l'ensemble de l'expérimentation, le positionnement de la caméra "de face" donne lieu au plus grand nombre d'erreurs, de replays et au plus long temps d'exécution. Ce résultat était attendu : avec cette position de caméra, il y a systématiquement conflit entre les gestes du collègue et la disposition des objets dans l'espace de travail. Ce conflit peut être résolu par le sujet par un renversement mental des gestes du collègue. Dans notre contexte expérimental, l'utilisation de la vue de face pour la réalisation d'une tâche déjà complexe semble déborder les capacités cognitives. Les sujets ne semblent pas toujours parvenir à

effectuer la transformation nécessaire pour interpréter correctement les gestes du collègue. Cette supposition est confortée par l'analyse des processus mis en jeu dans le cadre de l'architecture ICS (figure 3.11).

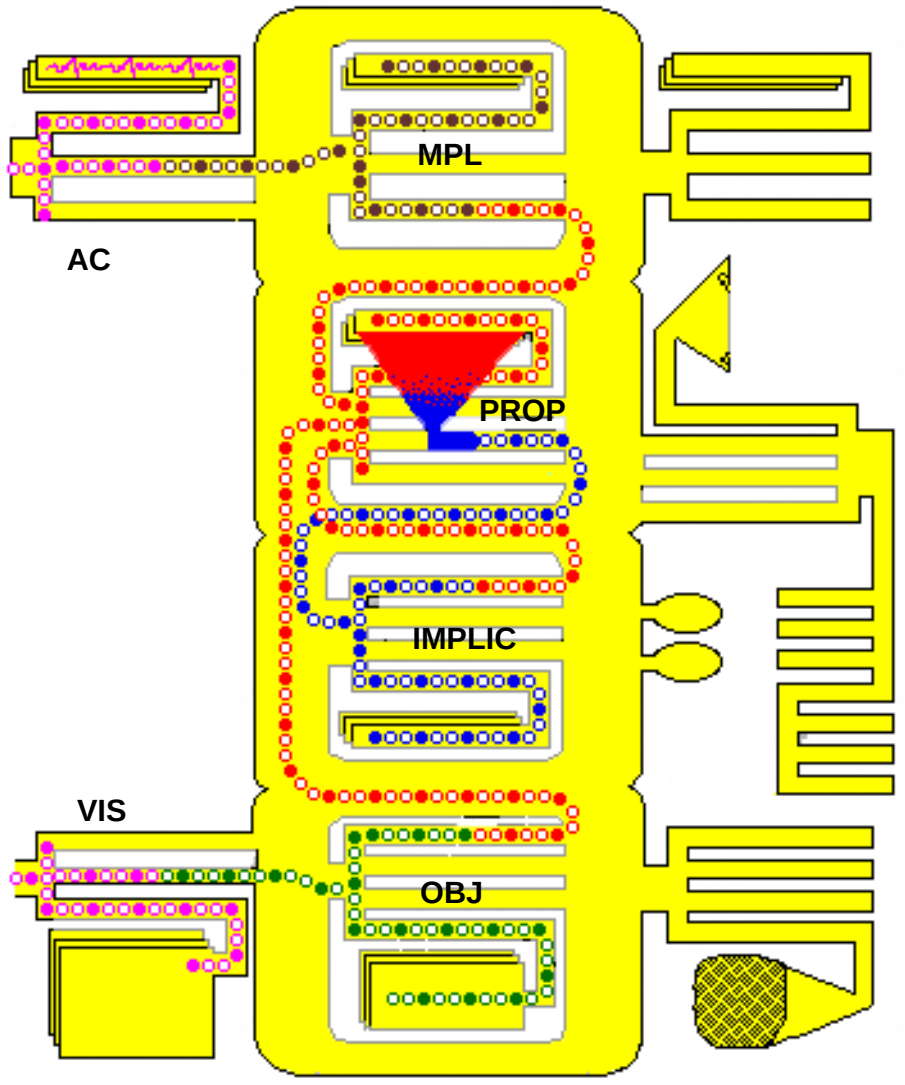

Figure 3.11. Fusion des informations au niveau du sous-système PROP dans l'architecture ICS.

La résolution des déictiques nécessite plusieurs combinaisons d'informations : tout d'abord, la séquence vidéo requiert la combinaison d'informations auditives (acquises via AC et MPL) et visuelles (via VIS et OBJ). Ces informations sont combinées au niveau du sous-système PROP et la séquence vidéo est comprise après passage de l'information dans le sous-système IMPLIC. La compréhension d'un déictique demande ensuite la combinaison de l'information élaborée par IMPLIC avec celle correspondant à l'interprétation de la disposition des objets dans l'espace de travail. Cette interprétation est fournie par OBJ via VIS. A ce moment du traitement des informations, deux flots

d'information sont disponibles à l'entrée de PROP : celui correspondant à la séquence vidéo et celui correspondant à l'espace de travail. Or la séquence vidéo étant renversée par rapport à l'espace de travail, les deux flots d'information ne sont pas cohérents et ne remplissent donc pas le critère de cohérence exposé en 3.4.2.3. La combinaison peut cependant s'effectuer, mais a de fortes chances de donner une information incohérente et donc de conduire le sujet à mal interpréter le déictique. Une étape supplémentaire de transformation (le renversement mental) est nécessaire pour que la compréhension du déictique soit correcte. Il y a alors cohérence entre les deux informations et la combinaison peut s'effectuer correctement.

ICS permet de détailler la complexité des traitements cognitifs nécessaires à l'interprétation de la vue de face et d'expliquer à quel niveau cognitif et pourquoi une surcharge cognitive peut intervenir. En revanche, ICS fournit peu d'indications pour expliquer les autres résultats. En particulier, la meilleure adéquation de la vue miroir par rapport à la vue par derrière ne peut pas être justifiée simplement. Les traitements cognitifs mis en œuvre sont similaires dans les deux cas et rien dans ICS ne tranche de façon claire en faveur de l'une ou l'autre configuration. De même les différences observées dans les résultats entre les déictiques spatiaux et d'objets et entre les déictiques spatiaux horizontaux et verticaux ne sont pas éclaircies par ICS.

La vue de derrière permettant de créer un point de vue commun entre le sujet et le collègue de la séquence vidéo, on aurait pu s'attendre à ce que la vue de derrière donne lieu à moins d'erreurs que les autres vues. En fait la vue miroir se révèle être souvent plus adéquate que la vue de derrière. Une exception notable est la résolution des déictiques spatiaux et d'objets (cf. figure 3.8) : la vue miroir et la vue de derrière donnent des résultats équivalents pour les déictiques spatiaux, mais la vue de derrière est moins performante pour les déictiques d'objets. La vue de derrière ne permet donc pas une meilleure résolution des déictiques.

Les résultats montrent aussi que, quelle que soit la position de la caméra, les sujets ont plus de difficultés à résoudre les déictiques spatiaux verticaux que horizontaux. Une explication possible est que, comme mentionné plus haut, les déictiques spatiaux font intervenir des formulations différentes suivant la disposition de l'espace de travail (horizontal/vertical) et ces différences ont pu perturber l'interprétation de ces déictiques.

Comme avec tous les résultats d'expérimentation, il convient d'être extrêmement prudent dans l'interprétation des résultats. Les résultats obtenus montrent la meilleure adéquation de la vue miroir *pour le dispositif expérimental considéré.* Il serait abusif d'en conclure que la vue miroir est mieux adaptée à des tâches coopératives dans un environnement

audio/vidéo. Cependant, les résultats montrent que, pour certaines tâches coopératives et pour la résolution de déictiques, la possibilité d'avoir une vue miroir peut améliorer l'intelligibilité des déictiques. Nous examinons maintenant cette constatation d'un point de vue technique.

La position des caméras dans un environnement de communication audio/vidéo peut donc contribuer à l'utilisabilité du système. Cependant, la position des caméras est également soumise à des contraintes qui rendent certaines configurations difficiles à mettre en œuvre. La plupart des systèmes de vidéoconférence, par exemple, placent la caméra face à l'utilisateur et à proximité de l'écran. Nous avons vu que cette solution est utilisable dans un contexte proche de celui de l'expérimentation Garden Movie à condition de fournir une vue miroir. La vue miroir peut être fournie par manipulation informatique de l'image captée par la caméra : une transformation simple (symétrie verticale) appliquée systématiquement aux images captées est suffisante. Nous remarquons avec intérêt que CU-SeeMe [CU-SeeMe 1995], un outil de vidéoconférence largement diffusé dans la communauté Internet, dispose précisément d'une fonction miroir.

La fonction miroir peut cependant avoir des effets indésirables et il est souhaitable qu'elle puisse être désactivée par l'utilisateur. En effet, puisque l'image est renversée, l'environnement de l'utilisateur observé est également renversé ; par exemple un environnement familier, tel un bureau, peut être surprenant vu "à l'envers". Dans certains cas, voir une image à l'envers peut être plus déplaisant : si l'utilisateur observé sur l'image vidéo porte un tee-shirt avec une inscription, cette inscription apparaîtra en miroir à l'image.

La vue derrière l'épaule est plus difficile à réaliser en pratique que les deux autres : placer la caméra derrière soi est contraignant, surtout si cette position inhabituelle n'est pas utilisée constamment. Bien sûr, on peut utiliser un système à plusieurs caméras et activer la caméra située derrière soi en cas de besoin ; d'autres problèmes subsistent néanmoins. La caméra doit être placée de façon précise afin que l'utilisateur et ses gestes soient bien visibles : pour un utilisateur de station informatique, placer une caméra derrière soi pour que les gestes soient visibles, et en évitant de préférence de filmer l'écran de sa station impose des contraintes contradictoires.

L'expérimentation Garden Movie nous a fourni un nouveau principe lié à l'interprétation cognitive des expressions déictiques. Nous déduisons de ce principe la propriété de réversibilité du média vidéo présentée au chapitre 4.

3.5.5. Leçons de l'expérimentation Garden Movie

Un des objectifs de la réalisation et de la conduite de cette expérimentation était d'étudier la démarche expérimentale et de l'appliquer aux tests d'utilisabilité que nous serions amenés à réaliser avec la plate-forme NEIMO. L'expérimentation Garden Movie nous a effectivement permis de découvrir et surtout d'appliquer la démarche expérimentale. Et nous nous sommes rendus compte que la psychologie expérimentale et les tests d'utilisabilité, même s'ils ont en apparence des points communs, sont deux mondes bien différents. La mise en œuvre de tests d'utilisabilité peut cependant bénéficier des méthodes de la psychologie expérimentale.

La différence majeure entre les deux démarches réside dans les postulats de départ. Une expérimentation psychologique est conçue pour étudier le comportement humain. Partant d'une hypothèse de départ sur un comportement humain, éventuellement étayée par une théorie psychologique, l'expérimentation permet de confirmer ou d'infirmer cette hypothèse, et éventuellement fait apparaître de nouvelles questions. Un test d'utilisabilité vise à étudier une interface homme-machine. Partant d'un aspect de l'interface que l'on souhaite étudier, le test d'utilisabilité permet de mettre en évidence les défauts éventuels liés à cet aspect de l'interface. On voit ici se dégager deux différences importantes : la psychologie expérimentale s'intéresse au comportement humain, les tests d'utilisabilité étudient une interface homme-machine. D'autre part, en psychologie expérimentale, l'expérimentateur a en général une hypothèse *a priori* qu'il cherche à valider ; au contraire, les ergonomes qui réalisent un test d'utilisabilité s'efforcent de ne pas avoir d'*a priori* sur le système qu'ils testent.

Dans la préparation du test, les approches sont aussi différentes. Pour préparer une expérimentation psychologique, l'expérimentateur identifie un ensemble de variables indépendantes ; l'expérimentation a pour but de présenter au sujet des stimuli qui combinent les valeurs possibles de cet ensemble de variables et de mesurer les réactions du sujet en réponse à ces stimuli. Afin de limiter son champ d'investigation et pour ne pas perturber les résultats, l'expérimentateur cherche à limiter le nombre de variables ; une expérimentation psychologique fait rarement intervenir plus de quatre ou cinq variables. Pour préparer un test d'utilisabilité, l'ergonome prépare un scénario représentatif des tâches pour lesquelles l'interface a été conçue. Le test consiste en la réalisation de ce scénario par le sujet. L'ergonome observe le comportement du sujet et en déduit d'éventuels défauts de l'interface, en s'appuyant sur son expérience et sur des règles heuristiques.

La conduite d'une expérimentation psychologique est une école de rigueur. Comme les variables de l'expérimentation sont clairement recensées, il est essentiel que d'autres variables parasites ne viennent pas perturber l'expérimentation. Par exemple, les instructions que l'expérimentateur donne au sujet doivent être strictement les mêmes pour tous les sujets, et le comportement de l'expérimentateur doit être identique avec tous les sujets pendant l'expérimentation. Un test d'utilisabilité est beaucoup moins strict ; l'ergonome peut par exemple se permettre d'intervenir et d'aider le sujet si le besoin s'en fait sentir.

Dans l'analyse et l'utilisation des résultats enfin, les approches varient. Les données à recueillir lors d'une expérimentation psychologique sont déterminées à l'avance et sont ensuite analysées avec des outils statistiques afin de déterminer des constantes de réactions indépendamment des individus. Dans un test d'utilisabilité classique, les données recueillies le sont de façon très informelle : il s'agit des notes des expérimentateurs, de bandes vidéo, etc. D'autres données peuvent également être recueillies, tel le temps nécessaire à la réalisation d'une tâche donnée. Contrairement aux expérimentations psychologiques, il n'y a pas d'analyse statistique poussée de façon à gommer les comportements individuels. Un défaut de l'interface mis en évidence par un seul sujet sera pris en compte. En psychologie expérimentale, un sujet fournissant des résultats très en-dehors de la moyenne des sujets sera minimisé par l'analyse statistique, voire écarté de l'analyse. Un test d'utilisabilité numérique comme ceux que nous pouvons réaliser avec la plate-forme NEIMO se rapprocherait plus de la méthode expérimentale en psychologie : les données recueillies peuvent être traitées avec des outils statistiques, mais le but reste différent. Il ne s'agit pas d'identifier des constantes du comportement humain, mais de détecter les difficultés que peut poser une interface aux utilisateurs.

En conclusion, la démarche de la psychologie expérimentale est plus rigoureuse parce qu'elle sert un but différent. Au même titre qu'une expérience de physique, une expérimentation psychologique relève d'une démarche scientifique classique : on veut vérifier ou infirmer une hypothèse de départ ; l'objet de l'expérimentation, c'est-à-dire ses variables sont clairement identifiées, et l'expérimentation est reproductible. Un test d'utilisabilité en revanche vise à mettre en évidence les défauts d'une interface. Les ergonomes, sans idée préconçue sur l'interface, observent le comportement du sujet et, à partir de leur expérience et de règles heuristiques, détectent les faiblesses de l'interface.

3.6. Synthèse

La psychologie cognitive apporte une contribution majeure à l'étude des systèmes de communication homme-homme médiatisée dans le cadre des systèmes multi-utilisateurs.

Contrairement aux approches traditionnelles, représentées par le modèle du processeur humain, nous avons préféré privilégier les aspects qualitatifs plutôt que quantitatifs de l'interaction de l'utilisateur avec ces systèmes. Il y a à cela plusieurs raisons : tout d'abord, les données fournies par les modèles quantitatifs ne sont que des valeurs moyennes ; certains de ces résultats sont régulièrement contestés et dépendent fortement de l'expérience de l'utilisateur. Dans le cas des systèmes multi-utilisateurs, où l'adaptation de l'interface à chacun des utilisateurs du système est une nécessité, l'utilisation de valeurs moyennes comme base de travail n'est pas pertinente. D'autre part, l'apparition de nouveaux médias pour la communication humaine informatisée, tels le son et la vidéo, pose des questions nouvelles en lien direct avec la perception. Dans la communication humaine, informatisée ou non, il peut être difficile de déterminer la tâche et encore plus d'évaluer quantitativement un échange entre participants. L'utilisation d'un modèle qualitatif et prédictif tel ICS, appliqué à la perception d'informations complexes, permet d'évaluer l'utilisation des ressources cognitives de l'utilisateur et ainsi de déterminer la validité des choix de conception. Comme pour toute théorie psychologique, un tel modèle trouve idéalement sa place en début de conception, et permet de critiquer des choix et d'en suggérer de nouveaux. Dans le cas d'un système existant, il peut aussi servir à expliquer des difficultés d'utilisation. Il faut cependant noter que, même si le modèle ICS peut être présenté assez simplement, son utilisation effective dans un processus de conception requiert une certaine expertise et la participation de psychologues.

Références

[Andler 1992] D. Andler. *Introduction,* in *Introduction aux sciences cognitives.* D. Andler, (ed.) Gallimard, Paris, France, 1992.

[Barnard 1985] P. J. Barnard. *Cognitive Resources and the Learning of Computer Dialogs,* in *Interfacing Thought, Cognitive aspects of Human Computer Interaction.* J. M. Carroll, (ed.) MIT Press, 1985. pp. 112-158.

[Barnard 1992] P. J. Barnard et J. May. *Real time blending of data streams: a key problem for the cognitive modelling of user behaviour with multimodal systems,* Amodeus Project, Working Paper, UM/WP 26, 1992.

[Barnard 1994] P. J. Barnard, J. May et D. Salber. *Deixis and Points of View in Media Spaces,* Amodeus Project, Working Paper, UM/WP 19, 1994.

[Card 1983] S. K. Card, T. P. Moran et A. Newell. *The Psychology of Human-Computer Interaction,* Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1983.

[CU-SeeMe 1995] *CU-SeeMe* . Logiciel pour Apple Macintosh. Cornell University, 1995.

[Gaver 1993] W. W. Gaver, A. Sellen, C. Heath et P. Luff. *One is not Enough: Multiple Views in a Media Space,* InterCHI'93, ACM/IFIP Conference on Human Factors in Computing Systems, Amsterdam, Pays-Bas, 1993. pp. 335-341.

[Heath 1992] C. Heath et P. Luff. *Media Space and Communicative Asymmetries: Preliminary Observations of Video-Mediated Interaction,* in *Human-Computer Interaction,* 7(3), 1992. pp. 315-246.

[Ishii 1992] H. Ishii et M. Kobayashi. *ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact,* CHI'92, ACM Conference on Human Factors in Computing Systems, Monterey, California, USA, 1992. pp. 525-532.

[MacroMedia 1993] *Director 3.1.1.* Logiciel pour Apple Macintosh. MacroMedia Inc., 1993.

[McGurk 1976] H. McGurk et J. MacDonald. *Hearing lips and seeing voices,* in *Nature,* numéro(264), 1976. pp. 746-748.

[Minneman 1991] S. L. Minneman et S. A. Bly. *Managing a trois: A Study of a Multi-User Drawing Tool in Distributed Design Work,* CHI'91, ACM Conference on Human Factors in Computing Systems, New Orleans, Louisiana, USA, 1991. pp. 217-224.

[Nigay 1994] L. Nigay. *Conception et réalisation des systèmes interactifs: Application aux Interfaces Multimodales.* Thèse de doctorat, Université Joseph Fourier Grenoble I, 1994.

[Reinhardt 1993] A. Reinhardt. *Video Conquers the Desktop,* in *BYTE,* 18(9), septembre 1993. pp. 64-80.

[Smolensky 1992] P. Smolensky. *IA connexionniste, IA symbolique et cerveau,* in *Introduction aux sciences cognitives.* D. Andler, (ed.) Gallimard, Paris, 1992. pp. 77-106.

[Tang 1991] J. C. Tang et S. L. Minneman. *VideoWhiteboard: Video Shadows to Support Remote Collaboration*, CHI'91, ACM Conference on Human Factors in Computing Systems, New Orleans, Louisiana, USA, 1991. pp. 315-322.