



Projet exploratoire VERBATIM

VERification Biformelle et Automatisation du Test des Interfaces Multimodales
(thèmes: faciliter les usages , nouvelles interfaces, ergonomie, vérification formelle)

Sous-Projet 5 Test in situ d'interfaces et d'interactions de services mobiles multimodales.

Lot 1 Aspects théoriques appliqués à l'interaction multimodale.

Auteur : Laurence Pasqualetti

AQL Groupe Silicomp, Clearsy, France Telecom, IMAG-CLIPS, IMAG-LSR, LISI



INTRODUCTION	3
DEFINITIONS ET CRITERES DE LA MULTIMODALITE	3
QU'EST-CE QU'UNE MODALITE ?	3
APPROCHES PSYCHOLOGIQUE ET ERGONOMIQUE DES THEORIES ET DONNEES EXPERIMENTALES SUR LA COMMUNICATION, APPLIQUEES A L'INTERACTION MULTIMODALE	4
1. Identification des concepts fondamentaux des théories de la communication	4
Un processus de traitement de l'information	4
Un processus orienté (sortie-entrée)	4
Un processus parfois contraint parfois appauvri (homme-machine)	5
La communication non-verbale.	6
L'attention	6
2. Les travaux sur la performance en situation de double tâche	7
3. L'étude des modalités	8
Du point de vue des modalités d'entrée	8
Du point de vue des modalités de sortie	9
IDENTIFICATION DE QUESTIONS SOULEVEES PAR LES TRAVAUX EN ERGONOMIE SUR L'UTILISATION DE LA MULTIMODALITE	10
4. Combinaisons des modalités dans les interfaces multimodales	10
5. Les actes multimodaux	10
6. Modélisation de la multimodalité	12
Les quatre propriétés récurrentes	12
Des grilles d'analyse pour l'évaluation ergonomique de la multimodalité	16
7. Utilisation de modalités redondantes	16
Fréquence du recours à la redondance	16
Utilisation de modalités complémentaires	17
Libre choix des modalités	18
Les conséquences de l'interaction multimodale sur la performance	19
CONCLUSION	21
BIBLIOGRAPHIE	22

Le présent document est rédigé dans le cadre du sous projet 5 du projet RNRT VERBATIM, dont l'objet est la VERification Biformelle et Automatisation du Test des Interfaces Multimodales.

Introduction

Depuis quelques années, les progrès réalisés en matière de technologies informatiques ouvrent de nouvelles perspectives dans le domaine de l'interaction homme-machine. L'une de ces perspectives, explore l'idée de permettre aux utilisateurs et aux systèmes d'interagir de manière multimodale, comme on peut l'observer dans les situations d'interaction homme-homme, c'est-à-dire en utilisant différentes modalités communicationnelles (parole, texte, gestes, etc.). L'exploration de cette perspective commence à se concrétiser par la conception de systèmes capables de développer des interactions « multimodales » avec les utilisateurs. Cependant, ces systèmes reposent pour l'essentiel sur une démarche de conception "intuitive" et guidée par la technologie plutôt que par une analyse systématique des apports de l'interaction multimodale et des problèmes associés.

Ce document vise donc à traiter l'interaction Homme/Machine, en mettant l'accent sur les aspects cognitifs de l'interaction, ceci afin de contribuer à l'élaboration d'une méthode validation de recommandations ergonomiques pour la conception de systèmes multimodaux.

Définitions et critères de la multimodalité

De nombreuses définitions de la multimodalité sont proposées dans la littérature. Ces définitions dépendent bien souvent des objectifs d'étude fixés par les auteurs, comme par exemple les problèmes de conception d'un système multimodal ou ceux d'évaluation de l'efficacité d'une tâche multimodale. La définition de Balazinska (1997), ci-après, nous paraît la plus apte à être prise en compte car elle est à la fois suffisamment précise (toutes les composantes de l'interaction y sont décrites : l'information, l'émetteur, le récepteur et le canal) ; et générale (la vue système et la vue utilisateur sont toutes deux considérées).

L'interaction multimodale se caractérise par la possibilité d'utilisation, pour l'homme comme pour la machine, de plusieurs modalités. Ainsi le " système multimodal se compose de deux ou de plusieurs voies de communication humain vers machine et/ou machine vers humain " (Balazinska, 1997, p.11). Par exemple, un système qui permet d'utiliser à la fois la modalité 'verbale' et la modalité 'tactile', que ce soit de façon parallèle et/ou séquentielle, est défini comme un système proposant une communication (interactionnelle) multimodale.

Qu'est-ce qu'une modalité ?

Une modalité est la combinaison de plusieurs éléments. Il s'agit d'un processus, orienté dans un sens déterminé, plus ou moins riche, qui reçoit, traite et produit de l'information.

Approches psychologique et ergonomique des théories et données expérimentales sur la communication, appliquées à l'interaction multimodale

1. Identification des concepts fondamentaux des théories de la communication

L'objectif de cette partie est de préciser les différentes théories de la communication et de dresser un repérage des résultats expérimentaux car nous considérons que la communication est le fondement existentiel dans la relation humaine.

Un processus de traitement de l'information

Tout comme Martin (1996), nous définissons une modalité comme relevant d'un processus de saisie et de traitement de l'information. Lorsqu'une information est produite, comme dans le cas d'une commande, elle provient d'un émetteur qui code cette information dans un mode d'expression particulier : une information conceptuelle est traitée, transformée en une information concrète. Balazinska (1997) précise que "une modalité n'est pas un simple processus de traitement de signaux. Il s'agit plutôt d'un processus de synthèse et d'analyse". En effet, l'information produite est synthétisée, codée, sous une forme particulière par l'émetteur, une forme qui sera apte à être véhiculée par un canal de communication. Le récepteur va capturer le signal diffusé par le canal de communication, l'analyser, et le décoder afin de produire une information interprétée (cf. figure 1).

Un processus orienté (sortie-entrée)

Deux types de modalités peuvent ainsi être distinguées : les modalités de sortie et les modalités d'entrée. Celles qui traduisent l'information abstraite en signal concret et celles qui traduisent le signal concret en information abstraite. Ces deux types de modalités sont des traducteurs d'information conceptuelle en signal physique.

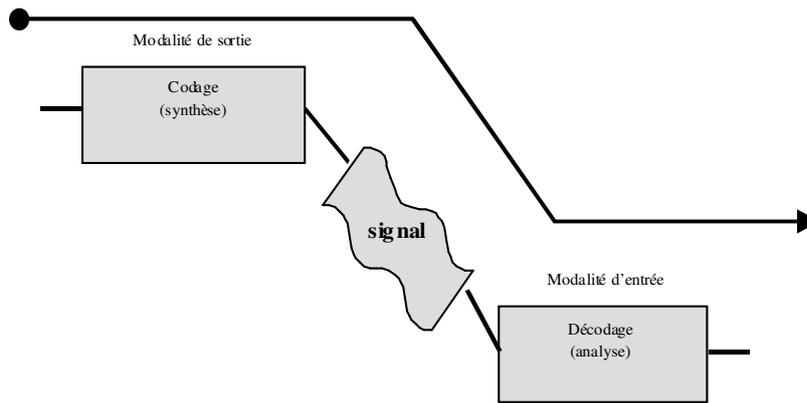


Figure 1 : la modalit e comme un processus ou de d codage

Un processus parfois contraint parfois appauvri (homme-machine)

Dans ce cadre, l'interaction homme-machine est consid r e comme une interaction humain-humain appauvrie, il est  galement utile de distinguer, entre ces modalit es d'entr e et de sortie, celles qui sont appauvries (Machine) de celles qui sont simplement contraintes (Humain). Nous distinguons ainsi, d'une part, les modalit es de sortie humaines (Mod_s_H), des modalit es de sortie machines (Mod_s_M), et d'autre part, les modalit es d'entr e machines (Mod_e_M), des modalit es d'entr e humaines (Mod_e_H).



Nous consid rons  galement, dans cette  tude, que les concepts de charge mentale et d'effort cognitif ne doivent pas  tre n glig s.

En effet, pour  tre per u et compris, un message doit pr senter un  quilibre entre la relative pr visibilit  et la totale pr visibilit . Dans le cas d'une totale pr visibilit , il y a impossibilit  de comprendre (le message est trop complexe). La th orie de l'information pr cise que la quantit  d'information en tant que r duction de l'incertitude du message est constitu e par la quantit  d'originalit  de ce message. Plus le r cepteur s'attend   la suite du message, plus le message est intelligible. Cette attente est appel e "accommodation informationnelle". La perception sera donc optimale non seulement si tous les signes sont communs au r cepteur et   l' metteur mais aussi si la fr quence d'utilisation de ces signes est voisine de celle attendue par le r cepteur. L'id e est qu'il existe un seuil d'originalit  du message, correspondant au taux d'originalit  acceptable par unit  de temps ou d'espace dans un message. Si le seuil est

dépassé, le récepteur doit faire un effort, il fera répéter ou relira ce qui représente un coût. S'il y a une limite à la capacité d'intégration de l'information nouvelle chez le récepteur, une règle fondamentale est de ne pas dépasser cette capacité. Il suffit alors de diminuer la quantité de signes ou de diluer cette information dans un grand nombre de signes, par exemple en usant de la redondance.

Miller définit la redondance comme " un excès de signaux sémantiques par rapport à ce qui est théoriquement nécessaire et suffisant pour l'intellection ou la transmission d'un message ". C'est l'inverse de la concision, la redondance consiste à délayer ou répéter. Deux raisons la justifient, (i) l'existence d'un seuil d'originalité du message en liaison avec le seuil d'intelligibilité et la limite de saturation du récepteur, et (ii) la possibilité de bruits. Les bruits sont variés, il s'agit de perturbations dans la transmission. Ils proviennent d'interférence, des défaillances sensorielles, des attitudes de l'émetteur ou du récepteur, de l'ambiguïté des messages (sémantique), des erreurs de codage. La redondance fait échec aux bruits, cependant il y a un seuil au-delà duquel les effets sont contraires. Plus, il y a de bruit plus il faut être redondant mais aussi lent, explicite, répétitif, clair et situé au maximum dans un répertoire connu du récepteur.

La communication non-verbale.

En effet, les signes verbaux et non-verbaux entretiennent une relation privilégiée qui est à l'origine de trois phénomènes :

La concordance et la convergence de ces deux systèmes a un effet clarificateur et renforçateur.

La discordance entre ces systèmes provoque chez les récepteurs des troubles émotionnels et brouille les significations du message.

La perception des signes non-verbaux si peu consciente soit-elle, a des effets étonnants. Certains auteurs ont montré qu'ils avaient 4 fois plus d'effets que les signes verbaux sur les attitudes des récepteurs.

Le jugement général d'auditeurs sur le contenu d'un message est influencé par les signes non-verbaux associés dans une proportion variant selon la personnalité de chaque auditeur.

L'attention

Dans la littérature psychologique et ergonomique, deux concepts entretiennent une relation étroite, il s'agit de l'effort cognitif et de la charge mentale. L'effort cognitif est pensé comme un coup cognitif, comme une allocation de ressources au système de traitement à un moment donné lors de la réalisation d'une tâche. La charge mentale est dépendante à la fois des contraintes de la tâche et des capacités de traitement du sujet humain (Kanheman, 1973). La puissance de ces dernières évolue avec le degré d'expertise du sujet.

Le système cognitif humain disposerait de ressources mentales en quantité limitée ce qui aurait des conséquences sur la qualité, l'efficacité et la profondeur de traitement. Pendant la réalisation d'une tâche, il y a répartition de ces ressources dans les différentes étapes de traitement en fonction de la tâche. Cette répartition n'est pas forcément équitable.

En situation de double tâche, on observe une distribution inégale des ressources, ce déséquilibre est accentué lorsque les deux informations (stimulus) sollicitent le même canal.

En situation d'attention focalisée sur une cible, si un distracteur intervient, il ne peut être traité efficacement que si les ressources résiduelles (de la focalisation) peuvent lui être attribuées.

En situation de partage d'attention, les ressources sont partagées entre les différents

traitements. La distribution des ressources permet un certain niveau d'opérationnalité donc la capacité à faire deux choses en même temps. En focalisation, la priorité est donnée au traitement principal, le canal non prioritaire subit un déficit.

Dans son modèle, Kahneman souligne le caractère intensif de l'attention. Il parle entre autres des qualités éveillantes d'une stimulation (ou composante spécifique d'un stimulus) qui est déterminante pour l'intensité relative des traitements qui vont être effectués. Il distingue un stimulus éveillant (ou alertant) d'un stimulus pertinent, leurs caractéristiques sont différentes. L'attention permet la sélection d'un stimulus pertinent indépendamment de ses caractéristiques alertantes. L'effort cognitif varie selon les situations. En situation d'alerte, les mécanismes sous-jacents en jeu sont pré-câblés, ils ne relèvent pas d'analyse consciente du sujet. Dans ce cas, l'effort cognitif (quantité de ressources dédiées au traitement) sera moindre. Dans un premier temps le stimulus n'est pas analysé entièrement, son analyse complète nécessite un surcoût de traitement donc un effort supplémentaire qui est la marque de l'investissement attentionnel.

L'interférence dans les théories de l'attention

Les études sur l'évolution de la performance lors de la réalisation de deux tâches effectuées simultanément ont permis d'envisager l'interférence selon deux points de vue :

- soit de façon capacitaire, la dégradation de la performance est expliquée par un dépassement des demandes attentionnelles par rapport à la capacité totale des ressources indifférenciées (Kahneman, 1973),
- soit de façon structurale, l'interférence dépend de la nature de la tâche.

Wickens (1984) propose un modèle attentionnel de ressources multiples basé sur 4 dimensions dichotomiques : la modalité sensorielle de présentation (visuelle ou auditive), le code de traitement (spatial ou verbal), les modalités de sorties (verbale ou manuelle) et l'étape de traitement (central ou tardif). Selon ce modèle, 2 tâches interféreront d'autant moins qu'elles auront des demandes de ressources séparées ; une tâche audito-verbale et une tâche visuo-spatiale seront moins dégradées que deux tâches audito-verbales.

Ces théories ne permettent pas de rendre compte des phénomènes d'interférence en double tâche. Il arrive que par leur similitude et indépendamment de la quantité de ressources qu'elles demandent, deux tâches engendrent deux phénomènes opposés, une augmentation ou une diminution de la dégradation. Le premier phénomène, appelé confusion, se produit quand les 2 tâches ont en commun leur matériel. Ce conflit apparaît quand le traitement de l'une des tâches produit un résultat nuisible au traitement de l'autre. Le second phénomène, appelé coopération, se produit lorsque les deux tâches partagent la même routine de traitement ou le même rythme, la dégradation est moins importante (ex. le suivi d'un rythme dans les 2 tâches, le batteur d'un orchestre qui a en parallèle la 'frappe grosse caisse' et la 'frappe caisse claire') Klapp (1979).

Les travaux sur l'apprentissage de tâches complexes, nous montrent qu'il ne faut pas entraîner les sujets sur les deux tâches séparément (Lintern & Wickens, 1991). Il faut amener les sujets à s'entraîner directement sur les deux ainsi la coopération sera favorisée, le contrôle plus efficace et les risques de conflits réduits.

2. Les travaux sur la performance en situation de double tâche

Il faut savoir que les causes de l'interférence sont subtiles et parfois totalement évidentes. Il est par exemple impossible d'utiliser une modalité pour donner deux réponses (parler et

chanter ou siffler et parler en même temps). Une série de travaux nous permettent d'illustrer les causes de l'interférence.

Dans une recherche de Kelso, Southard & Goodman (1979), où les mains sont en jeu, les sujets doivent atteindre des cibles plus ou moins grosses à des distances variables avec la main droite ou gauche. Le temps mis pour atteindre la cible croît avec la petitesse de la cible et son éloignement. Que se passe-t-il quand on demande d'atteindre une petite cible éloignée avec la main gauche et une grosse cible proche avec la main droite ? La main droite ralentit de façon à ce que les deux mains atteignent les cibles au même moment. Il est vrai que les mains travaillent typiquement en coordination plutôt que comme des systèmes indépendants.

En fait, les mains ne sont pas un cas particulier. Il ne faut pas conclure trop vite et dire que des tâches qui mettent en œuvre des systèmes de réponse bien distincts seront soumises à des interférences similaires. Il se trouve que généralement il va ainsi. C'est le phénomène qu'on appelle d'indocilité. En général, les sujets ont tendance à avoir des difficultés à émettre deux réponses à la fois, et si cela est nécessaire, une des réponses est mise en attente jusqu'à ce que l'autre soit achevée. Ce délai dépend de la charge relative aux deux processus de traitement en concurrence.

Le délaissement d'une tâche dépend du moment de l'apparition de la deuxième tâche. Si la charge attentionnelle est importante alors on a tendance à différer la tâche secondaire, l'interférence est d'autant plus importante et marquée que le processus de réponse utilise la même modalité (manuelle).

L'automatisme diminue l'interférence entre des tâches concurrentes, cependant il semblerait qu'il y ait toujours un minimum d'interférence. Même lorsque les systèmes de réponse sont séparés et que les sujets sont très entraînés, on peut s'attendre à une chute des performances résultant de la nécessité de gérer simultanément les 2 tâches.

3. *L'étude des modalités*

Du point de vue des modalités d'entrée

Les résultats font état d'une meilleure efficacité du temps partagé en présentation bi-modale qu'en présentation unimodale. En situation d'attention partagée entre deux modalités Visuel/Auditif, on constate une meilleure performance que s'il y a un partage pour une même modalité (V/V ou A/A). Les résultats ne permettent pas de dire s'il y a véritablement un partage ou plus grande rapidité à basculer d'une modalité vers l'autre.

Le son (signal auditif) est traité plus rapidement que le signal visuel. C'est une meilleure alerte surtout quand l'attention du sujet est orientée dans une autre direction (situation typique de surveillance). Mais quand les stimuli sont discrets (un son bref et/ou une lumière à détecter), c'est la réponse à la lumière qui prédomine, le son n'est pas entendu. Il y a une capture visuelle mais aussi intéroceptive (Wickens, 1987).

Le couple auditivo-visuel joue un rôle important dans la perception des signaux parole, il y a une coopération des modalités. Il y a une orientation d'une modalité vers l'autre (sorte de couplage auditivo-visuel). Pour les modalités d'entrée, il existe des processus d'interruption, de capture ou de couplage des modalités.

Du point de vue des modalités de sortie

Il y a plus d'interférence quand les modalités de réponses sont les mêmes que lorsqu'elles sont différentes (Wickens, 1980). Certains travaux montrent que les réponses verbales souffrent moins de la situation de double tâche que les réponses manuelles. Mais par ailleurs, les réponses verbales sont plus sensibles que les réponses manuelles à la variation de difficulté de la tâche principale. En situation partagée, on a un meilleur fonctionnement quand les modalités de réponses diffèrent que lorsqu'elles sollicitent le même système. Une réponse discrète verbale interfère moins avec une tâche continue de poursuite qu'une tâche motrice impliquant l'autre main. Et quand la réponse manuelle est discrète, celle-ci est prioritaire sur une réponse verbale (au détriment de la précision) (MacLeod, 1977).

Selon le modèle de Wickens les étapes de traitement, encodage, traitement et sélection de la réponse, ne sont pas toutes consommatrices dans les mêmes proportions. Posner et Boies (1971) spécifient que l'encodage et l'exécution des réponses sont les moins consommatrices.

Identification de questions soulevées par les travaux en ergonomie sur l'utilisation de la multimodalité

L'objectif de cette section est de recenser, à travers la présentation de plusieurs études expérimentales, différentes questions que soulève l'ergonomie à propos des interactions multimodales. Cette approche doit permettre de repérer les acquis de la discipline et les limites de ces acquis. En matière de nouvelles interactions, nous verrons que les caractéristiques des recherches menées posent de nombreuses difficultés de généralisation des conclusions apportées (nombre de sujets observés, indicateurs, type de tâche demandée aux sujets, durée de la recherche).

Parfois, l'utilisateur dispose de plusieurs modalités pour exprimer un ordre, une commande. Ces modalités peuvent être *redondantes* (plusieurs sont utilisées, mais une seule est nécessaire). Elles sont parfois non pas redondantes mais *complémentaires*, dans le sens où chaque modalité transmet à la machine différents éléments sémantiques d'un même message. Les résultats d'analyse d'activité présentés ci-après suivent ces deux catégories d'interaction : modalités redondantes versus complémentaires. Le troisième point aborde les dispositifs expérimentaux qui laissent au sujet le choix de la modalité, et donc parfois le choix entre multimodalité ou monomodalité.

4. Combinaisons des modalités dans les interfaces multimodales

L'approche la plus courante (Carbonell, Valot, Mignot et Dauchy, 1995 ; Martin, 1995 ; Mignot, 1995) consiste à identifier et définir les possibilités multimodales à travers deux conceptions, l'une issue de l'ergonomie cognitive, centrée sur l'acte (à l'initiative de l'utilisateur), l'autre relevant davantage de l'analyse informatique, centrée système. Un troisième courant (Bisseret, Caelen et Zanella, 1996) préfère envisager la multimodalité du point de vue de la tâche que l'utilisateur doit accomplir lors de l'utilisation d'un système/environnement multimodal, la tâche représentant la seule unité pertinente d'analyse des interactions, cette dernière établissant le lien entre les modalités perceptives de l'homme et les possibilités du dispositif.

5. Les actes multimodaux

“ *Un acte a pour finalité l'atteinte puis la satisfaction d'un but* ” (Bisseret et al., 1996, p.99), qui s'opérationnalise par la combinaison de plusieurs ressources sensorielles (dans le cas de la multimodalité) permettant une communication inter-agents.

Certaines recherches (Caelen et Coutaz 1991 ; Bellik et Teil, 1992 ; Martin et Bérroule, 1993 ; Martin, 1995 ; Coutaz et Nigay, 1994) dénombrent quatre propriétés (CARE, caractérisé par les trames grisées) définissant les actes multimodaux (cf.figure)

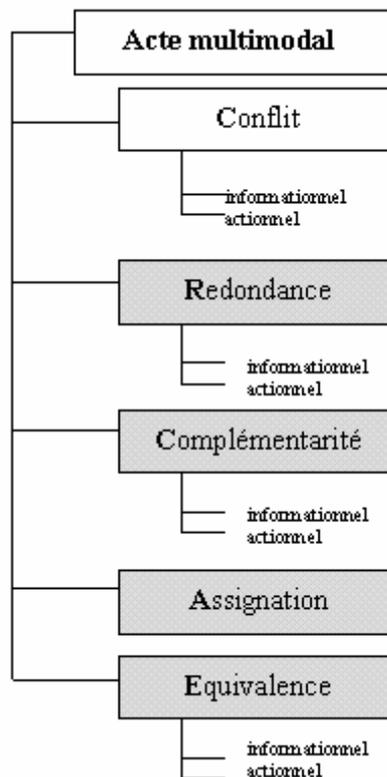


Figure : L'acte multimodal à travers les propriétés CARE (trame grise) versus CCARE

a- Assignment

La même modalité (ou sous-ensemble de modalités) est choisie pour une action donnée (appelée aussi spécialisation). Un exemple concret concerne l'énoncé de contraintes liées à un contexte particulier (Barès et Pastor, 1996) qui peuvent limiter les modalités de l'emploi d'un canal (comme par exemple, l'inhibition du canal vocal pour des ordres de tir activables seulement par boutons-poussoirs observables dans les avions de combat 'aérien'). Savoir que dans telle ou telle situation il est incorrect et cognitivement chargé d'avoir recours à la multimodalité semble tout autant nécessaire que de disposer d'informations précises sur les degrés de saturation de la quantité d'information transmise et sur ses modalités de diffusion.

b- Equivalence

Les différentes modalités permettent d'obtenir le même résultat (équivalence de résultat), et peuvent être d'un coût cognitif différent (équivalence fonctionnelle), (Valot, 1993).

c- Complémentarité

Chaque modalité utilisée est nécessaire à la compréhension de la totalité de l'acte.

d- Redondance

Les mêmes contenus informationnels sont exprimés par des modalités différentes utilisées pour l'accomplissement de l'acte. La redondance permet un renforcement du message.

e- Conflit

Cette définition d'acte multimodal peut être complétée par la notion de conflit (Bisseret et al., 1996). Le conflit correspond au changement de planification, à une mauvaise représentation

du système ou encore à un manque ou à un raté de coordination perceptive et/ou actionnelle (CARE devient CCARE).

Bellalem (1995) complète ces propriétés avec le critère d'intentionnalité qui permet de différencier les gestes accidentels (inconscients) des gestes de désignation, de manipulation (conscients). Les premiers relevant d'un message '**informationnel**' (manipulation accidentelle, fortuite), les seconds qualifiés d'**'actionnels**' (intentionnels). Ces critères transversaux ne seraient pas attribués à la propriété 'assignation' car elle ne permettrait pas le développement d'une réflexion sur l'utilisation ou la non-utilisation d'une modalité. Elle est attribuée d'office à une tâche précise (la réciproque 'une tâche précise pour une modalité particulière' étant réalisable) ou à un acte particulier.

6. Modélisation de la multimodalité

Les diverses expérimentations qui portent sur la multimodalité montrent rapidement que l'analyse des simples possibilités combinatoires au niveau fin, relatif à l'acte multimodal, ne permettent pas de relater la complexité de l'utilisation des interfaces multimodales ; il importe alors d'étendre cette analyse à la tâche multimodale, référent pertinent et utile du point de vue des échanges communicationnels mis en œuvre entre l'homme et le système.

Vigouroux et al. présentent les six axes définissant le système/espace Multi-Sensori-Moteur (MSM) proposés par Coutaz (1993) :

- le sens d'interaction
- le nombre de canaux
- la fonction d'interprétation ou de fusion en entrée versus la fonction d'interprétation ou de fission en sortie
- le niveau d'abstraction
- l'usage des modalités
- le contexte
-

Ces six axes se réduisent à trois pour définir l'espace référentiel d'étude de la multimodalité, espace présenté plus haut dans notre travail, qui engendrent la caractérisation de quatre classes d'interfaces multimodales (approche du point de vue du système) identifiées par Bourguet et Nigay (1992).

Les quatre propriétés récurrentes

Certains auteurs (Bourguet, 1992 ; Caelen, 1992 ; Nigay et Coutaz, 1993) proposent une typologie particularisée de ces propriétés à travers les spécificités du système ; indications intéressantes lors d'analyses de la réalisation de la tâche dans un environnement multimodal. Dans cette première approche la typologie est orientée système, celle-ci pouvant être déclinée en fonction des actes multimodaux ou encore de la tâche.

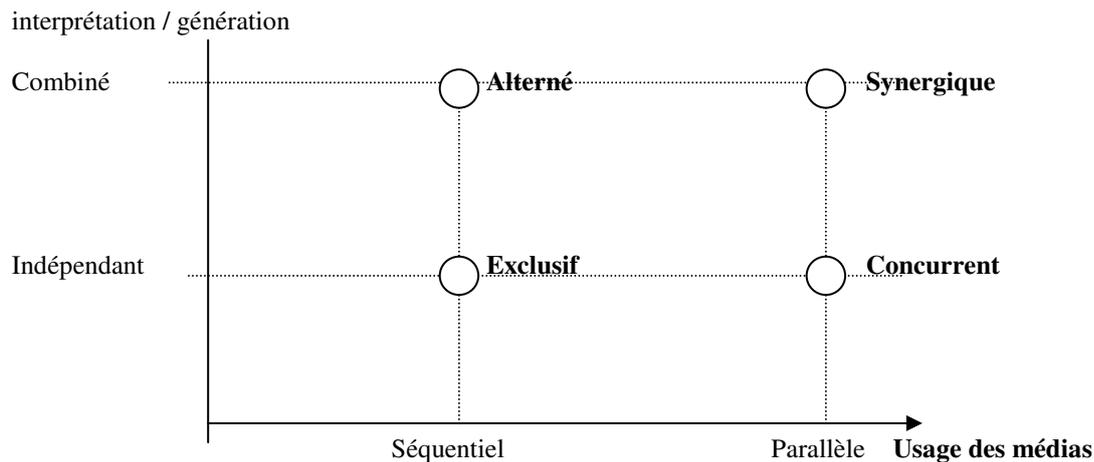


Figure : D'après Bourguet, 1992 – “ Conception et réalisation d'une interface de dialogue personne-machine multimodale ” INPG

Cette taxinomie décrit les capacités de la machine dans l'interprétation ou la génération d'expressions multimodales. L'ensemble des systèmes multimodaux sont composés de deux axes distincts :

- d'une part en fonction de l'usage des différents modes d'expression : **parallèle** (les deux modes se recouvrent) **ou séquentiel** (les deux actes ne sont pas situés dans un même intervalle de temps),
- et d'autre part, concernant la relation entre les deux modes : **indépendance ou combinaison**.

Parallèlement à cette approche, Coutaz (1992) introduit une analyse tri-dimensionnelle à même d'expliquer les différences entre les applications.

(i) le **niveau d'abstraction**, qui permet de distinguer les applications multimédia (pas de niveau sémantique) des applications multimodales, qui distinguent plusieurs niveaux sémantiques et induisent l'abstraction.

(ii) le niveau de **fusion**

(iii) les **contraintes temporelles**

Dans sa classification des interfaces, qui résulte de la prise en compte des deux dernières dimensions citées ci-dessus, Coutaz traite des mêmes aspects que Bourguet, à l'exception du fait qu'il n'introduit pas la notion de sémantique de l'action propre aux combinaisons multimodales, mais se limite à l'identification des moyens permettant de définir correctement les types d'événements que le système aura à gérer. Nous ne définirons pas dans le détail ces classifications, celles-ci étant très proches de la première, la différence s'établissant réellement au niveau des libellés et de la nature des expressions qui peuvent être des commandes (Sortie) comme des réceptions de commandes (Entrée).

type a – interface séquentielle et exclusive (exclusive)

dans ce cas l'interface ne produit ou ne reçoit uniquement qu'une seule expression (commande ou réception), cette application ne s'effectue que par l'intermédiaire d'une seule modalité.

type b – interface concurrente et exclusive (concurrente)

type c – interface séquentielle et synergique (alternée)

type d – interface concurrente et synergique (synergique)

Cependant, ces différents types d'interfaces dépendent également de la nature des combinaisons des éléments de communication, se construisant suivant des critères relatifs à la fusion des événements multimodaux :

- “ **la complémentarité logique ou structurelle** ” signifie que le système va pouvoir intégrer des événements distants dans une même expression
- la “ **complétude de la structure de données** ” principe qui permet au système de modifier son niveau d'abstraction en fonction des besoins de la situation
- le “ **contexte de dialogue** ” utilisant l'historique proche de l'interaction lui permettant d'associer des éléments par coréférence*(signifie qu'une interaction d'une modalité donnée ne peut être correctement comprise par le système sans tenir compte d'informations sur une ou plusieurs autres modalités)
- l'“ **incompatibilité des modalités** ” qui va permettre d'éviter des situations conflictuelles dues à la fusion d'événements monomodaux ne pouvant pas être utilisés conjointement.
- la “ **proximité temporelle** ” relative à l'analyse systémique d'une situation “ synergique ” lors de laquelle le système va devoir identifier le lien sémantique et syntaxique des informations produites ou reçues dans un bref délai. C'est cette activité qui va permettre la constitution d'unités de sens établissant une véritablement communication.

L'approche centrée tâche de la multimodalité

Bisseret et al. (1996) proposent (avec pour principe) une déclinaison de ces propriétés par rapport à l'ensemble des éléments présentés auparavant qui permet d'inclure les séquences de l'interaction multimodale suivant les trois axes pré-définis que sont l'utilisation des modalités et les possibilités du système par le filtre de la temporalité nécessaire à la compréhension intégrale des échanges.

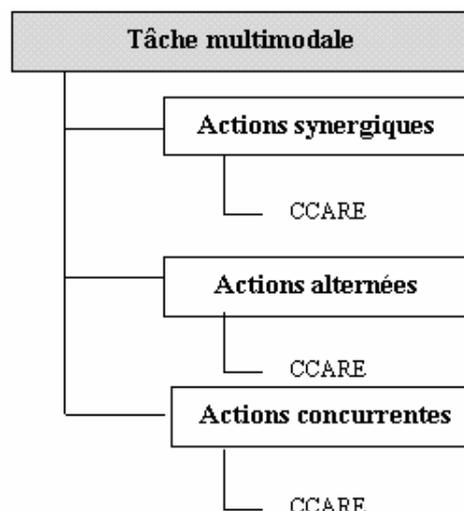


Figure : Propriétés T-CCARE

Nous remarquons ici que cette taxinomie intègre les propriétés sémantiques du point de vue utilisateur (CCARE) qui représentent les caractéristiques intrinsèques des types d'action. La limite de cette classification réside dans le fait que les auteurs “ intègrent ” l'ensemble de ces

propriétés à toutes les actions, alors qu'en mode appliqué (dans la réalité communicationnelle multimodale) des particularités combinatoires vont apparaître et donc pondérer cette attribution systématique. Nous nous proposons d'évaluer ces exceptions dans la suite de notre travail.

Classification des modalités de sortie/ système en fonction de modalités de base (Bernsen, 1994, 1993)

Bernsen préfère d'appréhender les modalités à travers leur structure d'information et de leurs caractéristiques. L'auteur s'est attaché à rendre possible l'évaluation des structures d'information intervenant dans la communication homme-machine, en présentant certains critères pouvant être pertinents pour la construction d'une typologie précise. Les travaux sont considérés comme les plus complets sur la théorie des modalités en sortie ; il centre sa recherche sur la représentation des expressions qui permet au concepteur d'interface d'identifier les modalités qui conviennent à la représentation des concepts. Pour ce faire, il définit la combinatoire [média-profit], le média représentant le support d'expression en relation avec les capacités sensorielles de l'utilisateur alors que le profit est déterminé par l'extraction de quatre types de critères, de propriétés.

En d'autres termes, et de façon plus systématique, la définition et la classification d'une modalité pure s'établit à travers cinq critères, quatre critères attachés au profil de la modalité et un critère précisant le média d'expression.

1. Langagière vs Non langagière

précise s'il y a "préexistence" d'un lexique, d'une syntaxe, d'une sémantique et d'une pragmatique.

2. Analogique vs Non analogique

précise si la modalité véhicule les propriétés de l'entité concernée (profondeur des métaphores, sémantique des icônes représentées)

3. Arbitraire vs Non arbitraire

indique si la sémantique introduite dans la modalité relève d'une attribution particulière ou s'appuie sur des représentations collectives/ sociales, émanant d'un consensus (non arbitraire), d'une acceptation générale

4. Statique vs Dynamique

indique l'influence du temps sur la structure des informations

5. Média d'expression

indique si la modalité est d'origine graphique, sonore ou tactile

Partant du constat de certaines incohérences de l'approche de Bernsen, Bellik et al. (1996) proposent une classification des modalités dans un espace multidimensionnel couvrant des situations de communication, y compris les entrées. Il tentent par leur présentation de réduire les incompatibilités de certains critères, ce qui permet alors l'application des critères indépendamment de la contiguïté des modalités.

	Critère temporel	Critère spatial	Critère langagier	Critère d'analogie	Critère de prosodie
Mode visuel	*animations/clics/notations	*2D/3D/3Dvirtuel	*valeurs continues	*photo/synthèse/peinture/graphique	
Mode auditif		*mono/stéréo/3D	*valeurs continues	*naturel/synthèse	*chant/parole
Mode oral			*valeurs continues		*chant/parole

Mode gestuel	*statique/ dynamique	*1D(clavier)/2Dso uris ou stylo/3D	*valeurs continues		
Mode Tactile Proprioceptif Kinetique	* animations/clig notement/rotati ons		*valeurs continues	*	

Tableau 2 : Tableau de synthèse concernant la classification de Bellik et al.

Ce type de taxinomie est intéressante, mais présente, au même titre que les autres, une évaluation de l'interaction multimodale centrée système. L'intérêt d'un tel exercice est de permettre d'identifier les particularités attachées à l'utilisation de telle ou telle modalité en situation d'interaction, tout en relevant les combinaisons qui restent faisables, malgré des actions déjà en cours mises en œuvre par l'utilisateur.

L'ensemble de ces propositions offrent une vue relativement exhaustive des possibilités techniques, mais manque de précisions sur des notions comme le croisement de modalités, le seuil de saturation pour l'usage des modalités, la compatibilité de sens entre modalités, etc.

Des grilles d'analyse pour l'évaluation ergonomique de la multimodalité

Nous avons vu que la description d'une interaction multimodale devait prendre en compte un certain nombre d'éléments afin d'être exhaustive et efficace : la nature des modalités mises en jeu, leur orientation (entrée ou sortie), leur type (humain ou machine), la disponibilité des outils et des canaux relatifs aux modalités, les contraintes environnementales pouvant influencer sur leur disponibilité, et enfin les intentions et unités de sens qu'elles véhiculent (et entre autres les relations action-référent et les relations entre les buts et les sous-buts).

Nous proposons de croiser ces différents éléments afin de fournir des grilles repères pour l'analyse de l'interaction (nous nous centrerons ici sur le type humain).

Contrairement aux grilles descriptives de systèmes multimodaux développées dans la littérature, nous allons axer les nôtres sur le point de vue utilisateur.

7. Utilisation de modalités redondantes

L'utilisateur peut avoir recours à des modalités différentes de manière redondante. Par exemple, Siroux, Guyomard, Multon et Remondeau (1998) observent l'utilisation de la confirmation suivante. L'utilisateur énonce "Are there any beaches on Lannion?" et appuie sur l'écran sur le mot "Lannion". Le lieu est ainsi désigné de manière redondante.

Fréquence du recours à la redondance

Martin, Julia et Cheyer (1998) nous indiquent que la redondance est rarement observée dans les études. Ils citent pourtant des résultats expérimentaux qui montrent que les utilisateurs adoptent une stratégie fortement redondante lorsqu'il y a possibilité de faire référence à un objet par une désignation courte (un caractère). Le pointage et la "modalité linguistique" sont alors utilisés.

La redondance n'est pas toujours parfaite. L'interaction est plus ou moins exactement redondante, du fait des variations du niveau de précision. En effet, le niveau de précision peut

varier en fonction de la modalité utilisée. Par exemple, on observe parfois, lorsqu'il s'agit de désigner un lieu, que le recours à une modalité de type commande vocale s'avère relativement peu précis ou encore que le recours à des graphiques est assez précis.

Plusieurs configurations peuvent être regroupées sous cette rubrique.

On observe (Mignot et Carbonell, 1996, cité par Martin et al., 1998) la situation suivante : un utilisateur, tout en employant une commande vocale, prend conscience qu'il a omis de commander l'exécution d'une action préalable. Il a alors recours au geste pour faire exécuter cette action, tout en poursuivant l'interaction vocale.

Les gestes illocutoires divergent des gestes référentiels. Siroux et al. (1998, p. 103) observent deux cas. Le premier est celui d'un geste tactile qui diffère de la parole : lors d'une action de pointage, l'utilisateur parle de "zone". Le second concerne une localisation désignée oralement par "zone" et gestuellement par une ligne (un segment).

Ces études qui relatent d'activités d'utilisateurs posent plusieurs questions, notamment :

- Pourquoi les utilisateurs ont-ils recours aux modalités redondantes puisqu'elles n'apportent guère d'informations supplémentaires au dispositif ?
- Pourquoi les utilisateurs expriment des commandes contradictoires (cas des "contradictions") ? Quand apparaissent-elles ? Quelle est leur signification pour l'utilisateur : sont-elles utilisées consciemment ? S'agit-il d'erreur, de problème de planification ? Sont-elles révélatrices de situations d'incompréhension de l'utilisateur ? Signifient-elles que l'utilisateur a pris conscience d'une erreur de commande et qu'il revient sur son action ? Comment doivent-elles être traitées par le dispositif ? Par quelle modalité le dispositif doit-il répondre ?

Utilisation de modalités complémentaires

Face à un mode de dialogue permettant d'utiliser plusieurs modalités lors d'une même commande, l'utilisateur a le choix. Il peut, pour une commande donnée, combiner ou non les modalités. Par exemple, pour connaître les hôtels existants sur une zone géographique donnée, l'utilisateur pourra poser sa question par oral et désigner la zone, non pas à l'oral mais avec sa main, sur une carte présentée sur écran tactile.

Quelques données empiriques transmises par Martin et al. (1998) montrent qu'il existe différentes activités inventées par les utilisateurs, pour combiner les modalités, notamment lors d'emploi de déictiques (i.e. dont le sens dépend du contexte d'énonciation, comme "ici et maintenant" qui ne peuvent être correctement interprétés qu'en connaissance du moment et du lieu dans lequel a été énoncé la phrase, transmis le geste, etc.).

- On relève des phrases énoncées en langage naturel, avec recours à des pronoms ou des adjectifs démonstratifs. Par exemple, le sujet dira "quels sont les hôtels existants sur cette zone ?" tout en désignant de la main la zone en question.

- On relève des phrases qui éludent le déictique. Ce dernier est désigné par la main. Par exemple, le sujet dira "quels sont les hôtels existants à ?" tout en désignant de la main la zone en question.

Ce type de phrases aurait un taux de fréquence peu élevé.

Les utilisateurs ont également recours aux anaphores (reprise d'une partie précédemment explicitée).

En effet, les possibilités d'ambiguïtés, liées à l'emploi de déictiques et d'anaphores, sont importantes. Par exemple, imaginons la suite d'interaction suivante. L'utilisateur pose

oralement la question suivante "quel est le contenu du courriel 1 ?", puis il ouvre, par modalité tactile, le "courriel 2". Il donne alors l'ordre oral suivant "répondre à ce courriel". De quel courriel s'agit-il alors ?

Libre choix des modalités

Les utilisateurs ont parfois la possibilité de choisir d'avoir recours à une modalité exclusive ou d'utiliser des modalités complémentaires.

Une expérience de Huls et Bos (1998) montre que les sujets ont plutôt recours au mode langagier (requête exprimée par des phrases) qu'au mode multimodal (saisie d'une commande plus sélection graphique). Elle montre également qu'un facteur intervient dans la décision des utilisateurs : il s'agit de la *longueur des noms des objets* sur lesquels la tâche porte. Plus précisément, les fichiers à manipuler dans la tâche proposée aux sujets sont soit désignés par des noms longs, soit par des noms courts. Quand les noms de fichiers sont longs, les sujets choisissent moins souvent le mode d'interaction langagier (demande écrite) que lorsque les mots sont courts. Le pointage est préféré lorsque les mots sont longs. Par contre, la *visibilité des objets* sur l'écran n'a pas d'effet sur les modalités choisies (les objets sont parfois non visibles sur la fenêtre : il faut utiliser une barre de défilement pour les percevoir).

Une autre expérience présentée par les mêmes auteurs confirme que le mode multimodal est peu choisi par les sujets. Les sujets utilisent préférentiellement la modalité "manipulation d'objet" que les modalités "phrases saisie au clavier", "menu" et "multimodale".

Par ailleurs, les *contraintes temporelles* sont un facteur décisif dans le choix de la modalité (Norman, 1991). Une recherche montre que sous pression temporelle, les sujets choisissent la modalité "mode de commande" (l'utilisateur donne des instructions par lignes de code) plus souvent que la modalité de manipulation directe. Il semblerait que le langage de commande requiert moins d'effort que la manipulation directe. Notons que 60 % des sujets de cette recherche ont changé de modalités (langage de commande vs. manipulation directe) en fonction des contraintes temporelles de la situation.

La recherche menée par Mignot et Carbonell (1996), montre "que les sujets privilégient l'usage de la parole et de la multimodalité synergique plutôt que celui du geste". Elle montre cependant, comme les études citées précédemment, que les interventions multimodales synergiques¹ ne sont pas le mode de dialogue le plus couramment utilisé : il représente 41 % des commandes formulées.

Une expérimentation de France Telecom R&D (Calvet & al.2001) ont mis des sujets en situation de réalisation d'une tâche familière, il leur était demandé d'utiliser une application de consultation de courrier électronique au moyen d'un ordinateur de poche (PDA) qui a été connecté pour les besoins de l'expérimentation à leur messagerie personnelle.

Quatre modalités étaient étudiées et implémentées :

- **la modalité vocale**, qui est présentée comme un mode d'interaction en langage naturel sans contraintes particulières (pas d'énoncés prédéfinis à utiliser par les sujets).
- **la modalité tactile**, qui nécessite un stylet pour pointer les boutons de commandes.
- **la modalité gestuelle**, qui repose sur un code prédéfini qui impose de conserver le PDA dans une main et d'effectuer le geste de la main libre devant l'appareil.

¹ Ces interventions sont définies comme une intervention d'expressions orale(s) et gestuelle(s) comprise entre deux réactions successives du système simulé"

- *la modalité* qualifiée improprement (et temporairement...) d'"*embodied*", qui consiste à associer des commandes à certaines manipulations prédéfinies de l'artefact (changements d'orientation, mouvements divers, etc.)

Seule la modalité tactile était réellement implémentée, les autres modalités étaient simulées par le compère grâce à une interface identique et un système de vidéo numérique.

Les premiers résultats obtenus sur l'utilisation des modalités, indique que l'ensemble des modalités a été utilisé par les sujets au cours des différentes sessions. Il n'est pas observé de domination d'une modalité commune à l'ensemble des sujets pour effectuer une ou plusieurs commandes (spécialisation) mais la spécialisation se situe à un niveau intra-individuel.

Les facteurs repérés influençant les choix de modalité sont :

- *Contexte de l'activité en cours (procédure opératoire récurrente)*. Ces séquences sont fortement liées au contexte de l'activité en cours.
- *Changement de but dans l'activité*. Les changements de modalité sont souvent associés à des phases de réorientation locale de l'activité.

Correction d'erreurs. De nombreux changements de modalités sont apparus dans des situations dysfonctionnelles afin de corriger des erreurs provenant soit du sujet, du système ou bien du compère.

Si quelques réponses sont déjà apportées (e.g. la multimodalité n'est pas le mode d'interaction privilégié des utilisateurs), des questions subsistent : comment s'effectuent le choix et les changements de choix de l'utilisateur (quels sont les critères qui amènent un utilisateur à abandonner une modalité d'interaction au profit d'une autre ? quel est le niveau de conscience des changements effectués ? ...), est-ce que la possibilité de choisir est un facteur de surcharge de l'activité ?

Les conséquences du choix de la modalité sur la performance seront abordées dans la partie suivante de ce document.

Les conséquences de l'interaction multimodale sur la performance

Est-ce que laisser le choix de la multimodalité aux utilisateurs améliore la performance ? Quel est l'impact de cette liberté sur le temps de réalisation de la tâche ? Est-ce que l'utilisateur utilise la modalité la plus efficace ?

Des résultats expérimentaux (Huls et Bos, 1998) montrent que laisser le choix de la modalité d'interaction aux sujets s'accompagne d'une augmentation du temps pris pour réaliser la tâche. Plus précisément, lors d'une tâche de gestion de fichiers, on compare le temps mis pour réaliser la tâche dans quatre conditions expérimentales :

- modalité imposée "sélection et manipulation des objets"
- modalité imposée "saisie d'une commande écrite"
- modalité imposée "multimodale" (saisie d'une commande écrite et sélection graphique)
- modalité libre.

Dans la modalité imposée "sélection des objets", la tâche est plus vite réalisée que dans la modalité libre. La modalité imposée "multimodale" est la condition dans laquelle la tâche est effectuée le plus lentement.

Ces données laissent à penser que laisser la liberté de la modalité ou imposer une modalité multimodale n'est pas à recommander aux concepteurs de dispositifs, dès lors que la durée de la réalisation de la tâche est importante.

Une expérimentation du CENA (Mertz,1997) place les sujets en situation multimodale. Les sujets doivent simultanément utiliser leurs deux mains et percevoir des informations diffusées visuellement et auditivement.

Les sujets doivent réaliser jusqu'à trois tâches, simultanément :

- sélectionner des références d'avions au moyen d'un pavé numérique ou d'une souris,
- mémoriser des mots énoncés oralement
- appuyer avec leurs pieds sur une des deux pédales à leur disposition, en fonction de l'état d'un signal qui s'affiche régulièrement.

Les résultats montrent notamment que plus on demande aux sujets de réaliser de tâche (jusqu'à trois), moins leurs saisies sont précises et rapides. Quand les sujets sont en surcharge cognitive, ils ne parviennent plus à viser correctement la cible (accéder un menu, positionner le pointeur sur des valeurs ...) et procèdent alors par aller-retour et ajustement successifs. L'auteur note que, dans certains cas, les sujets n'étaient plus capables de choisir la bonne direction du curseur, ils le déplaçaient dans le sens opposé. Il note également un problème de peu de conscience des dégradations, résultat typique de ce paradigme de la double tâche.

On note que la méthode de la double tâche est intéressante pour notre projet, et ce à double titre. Elle correspond à une utilisation imposée de plusieurs modalités pour accomplir des tâches. Elle est considérée comme à la fois un amplificateur d'appréciation subjective et un amplificateur de problèmes techniques. C'est pourquoi il nous semble important de ne pas négliger les résultats empiriques obtenus lors des études en double tâche.

Dans le sens "Mod_s_M -> Mod_e_H", la multimodalité s'accompagne d'une augmentation du nombre d'erreurs. C'est le sens des résultats obtenus par Huls et Bos (1998) par une expérience qui compare quatre conditions :

- feed-back écrit (le dispositif commente les actions de l'utilisateur de manière écrite)
- feed-back parlé
- feed-back écrit et parlé
- sans feed-back.

La tâche expérimentale consiste à gérer des fichiers. Les données recueillies montrent que la condition la plus rapide est l'écrit, ensuite "écrit et parlé", puis "sans feed-back", puis enfin "parlé".

Les conditions qui font le moins l'objet d'erreur sont "parlé", puis "écrit", puis "écrit et parlé" et enfin "sans feed-back".

En résumé, on peut dire que ces résultats montrent que le feed-back apporte une aide. Ils montrent également les résultats suivants pour les canaux suivants :

- Canal visuel seul : interaction rapide et peu d'erreur
- Canal auditif seul : interaction lente mais très peu d'erreur
- Canal "double" : plus d'erreur

Retenons de cette partie que la multimodalité ne s'accompagne pas d'amélioration de la performance, ni en temps de réalisation de la tâche, ni en nombre d'erreurs. On peut même observer une dégradation de la performance. On relativisera cependant ces résultats qui portent sur seulement quelques dispositifs.

Conclusion

En comparaison avec les modalités uniques, la multimodalité présente, *a priori*, un mode d'interaction plus naturel avec les dispositifs. Pour Huls et Bos, 1998, les interfaces multimodales possèdent encore d'autres avantages :

- en multipliant les modes d'interaction, elles permettent de guider l'attention des utilisateurs,
- elles permettent également d'augmenter le volume d'information transmises,
- elles permettent au dispositif, comme à l'utilisateur, de faire référence à des objets qui ne font pas partie du modèle du monde en cours,
- l'augmentation du feed-back des actions (par exemple, le système indique par la voix et par un affichage : "vous allez effacer le fichier Picasso"), liée à la multiplication des modalités utilisées dans l'interaction, donne plus de possibilités aux utilisateurs de prendre conscience de leurs éventuelles erreurs.

Pourtant, les quelques résultats expérimentaux que nous venons d'exposer pointent de nombreuses limites et problèmes posés par l'utilisation de dispositifs multimodaux. Les concepteurs de dispositifs multimodaux se heurtent à la faiblesse des effectifs de données empiriques sur l'activité des utilisateurs. Par exemple, Stock et al. (1994) ont posé les bases de modes de navigation dans un système d'information (ces bases reposent sur la flexibilité de l'initiative du dialogue et la flexibilité des types de recherche de l'utilisateur). Ils souhaiteraient alors obtenir des données issues d'études du comportement des sujets, pour orienter leurs développements informatiques.

Lorsque des résultats expérimentaux existent, il est légitime de se poser la question de leur pouvoir de généralisation. En effet, le type de média utilisé (type d'écran, type de microphone, etc.) pourrait déterminer les résultats obtenus.

Bibliographie

- Balbo, S., J. Coutaz, & D. Salber. *Towards automatic evaluation of multimodal user interfaces*. in *Conference on Intelligent User Interfaces*. 1993.
- Benoit, C., et al., *Audio-Visual and Multimodal Speech Systems*, in *Handbook of Standards and Resources for Spoken Language Systems - Supplement Volume.*, D. Gibbon, Editor. 2000.
- Bellik, Y., Ferrari S., Néel F., Teil D., Pierre E., Tachaires V. (1996). Interaction multimodale : concepts et architecture, Actes des 5èmes Journées Internationales de Montpellier, L'interface des mondes réels et virtuels.
- Calvet, G., Kahn, J., Salembier, P., Zouinar, M., Pasqualetti, L., Nigay, L., Rey, G. & Briois, J.C. (2001). Etude empirique de l'usage de la multimodalité avec un ordinateur de poche. *Actes de la conférence "IHM-HCI 2001"*, 10-14 septembre, Lille, France.
- Cole, R.A., et al., *Survey of the State of the Art in Human Language Technology*, . 1995, Center for Spoken Language Understanding CSLU
- Dahlbäck, N., A. Jönsson, and L. Ahrenberg. Wizard of Oz-studies - why and how. in *Proceedings of the International Workshop on Intelligent User Interfaces*. 1992. New York, NY, USA: ACM Press.
- Hayes., W.L., *Statistics for the Social Sciences*. 1993, New-York, USA: Holt, Rinehard, and Winston. 79.
- Huls, C., Bos, E. (1998). Studies into Full Integration of Language and Action in H. Bunt, R.J. Kanheman, D. (1973). *Attention and effort*, Londres, Prentice Hall.
- Kelso, J.S., Southard, D.L. & Goodman, D. (1979). On the coordination of two-handed movements. *Journal of Experimental Psychology : Human Perception and Performance*, 5, 229-238.
- Klapp, S. (1979). Doing two things at once : The role of temporal compatibility. *Memory et cognition*, 7, 375-381.
- Lintern, G. & Wickens, C.D. (1991). Issues for acquisition and transfer of timesharing and dual-task skills, in D.L. Damos (Ed.), *Multi-task performance*, London : Taylor & Francis, pp123-138.
- MacLeod, P. (1977). A dual-task response modality effect : support for a stimuli process models of attention, *Quarterly Journal of Experimental Psychology*, 29, 651-667.
- Martin, J. Cl., Julia, L., Cheyer, A. (1998). *A theoretical Framework for Multimodal User Studies*, <http://www.limsi.fr/Individu/martin/publications/download/cmc98-2.ps>
- Mellor, B. and C. Baber. *Modelling of speech-based user interfaces*. in *Eurospeech'97*. 1997.
- Mertz, C. (1997) *Should we kill the mouse ? comparing mental load induced by 6 input dialogs*, (Proceeding of the ACM CHI'97), CENA.
- Mignot, C., & Carbonell, N. (1996). Commande orale et gestuelle : étude empirique. *Technique et Science Informatiques*, 15(10), 1399-1428.
- Norman, K. (1991). *The Psychology of Menu Selection: Designing Cognitive Control at the Human/Computer Interface*, Ablex Publishing Corporation.
- Oviatt, S. and R. VanGent. *Error resolution during multimodal human-computer interaction*. in *International Conference on Spoken Language Processing*. 1996. Philadelphia (PA).
- Oviatt, S., A. DeAngeli, and K. Kuhn. *Integration and synchronization of input modes during multi-modal human-computer interaction*. in *International Conference on Computer-Human Interaction*. 1997. Atlanta (GA): ACM
- Posner, M.I. & Boies, S. (1971). Components of attention, *Psychological Review*, 78, 391-408.

- Rosenthal, R. and R.L. Rosnow, *Essentials of Behavioral Research: Methods and Data Analysis*. Psychology, ed. McGraw-Hill. 1991: McGraw-Hill.
- Shneiderman, B., *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, ed. A. Wesley. 1997, Menlo Park.
- Siroux, J., Guyomard, M., Multon, F., Remondeau, C. (1998). Modeling and Processing of Oral and Tactile Activities in the GEORAL System, in H. Bunt, R.J. Beun, T. Tijn, Borghuis (Eds.). *Multimodal Human-Computer Communication, Systems, Techniques, and Experiments*, Springer.-verlag, Berlin, 101-110.
- Suhm, B. *Empirical evaluation of interactive multimodal error correction*. in *IEEE Workshop on Automatic Speech Recognition and Understanding*,. 1997. Santa Barbara (CA).: IEEE Signal Processing Society.
- Suhm, B., *Multimodal Interactive Error Recovery for Non-Conversational Speech User Interfaces*, . 1998, Karlsruhe University: Karlsruhe, Germany.
- Sweeney, M., M. Maguire, and B. Shackel, *Evaluating user-computer interaction: A framework*. *International Journal of Man-Machine Studies*, 1993(38): p. 689-711.
- Vo, M.T. and C. Wood. *Building an application framework for speech and pen input integration in multimodal learning interfaces*. in *International Conference on Acoustics, Speech and Signal Processing*,. 1996. Atlanta (GA).
- Waibel, A., *Interactive translation of conversational speech*. *Computer*, 1996. 29(7).
- Walker, M.A., *et al.* *A general framework for evaluating spoken dialogue agents*. in *35th Annual Meeting of the Association of Computational Linguistics*. 1997. Madrid
- Wickens, C.D. (1980), the structure of attentional resources, in R.S. Nickerson (ed.) *Attention and performance*, VIII, Hillsdale, Lawrence Erlbaum.
- Wickens, C.D. (1984), Processing resources in attention, in R. Parasuraman et D.R. Davies (eds), *Varieties of attention*, Londres, Academic Press, 63-102.
- Wickens, C.D. (1987), Attention, in P.A. Hancock (ed.) *Human Factors Psychology*, North-Holland, Elsevier Publishers, 29-79.
- Wiener – « cybernétique et société », traduction française. Ed. des Deux rives, paris