

Une approche centrée tâche de la multimodalité

Marie-Laure Zanello ; Jean Caelen
Laboratoire Clips-Imag, Equipe GEOD
Imag bat^t B
Université Joseph Fourier
BP 53
38041 Grenoble Cedex 9
Marie-Laure.Zanello@imag.fr
Jean.Caelen@imag.fr

André Bisseret
INRIA
655 avenue de l'Europe
38330 Montbonnot Saint-Martin

Andre.Bisseret@inria.fr

RESUME

Dans le cadre de la communication homme-machine multimodale un certain nombre de formalismes catégorisant les actes multimodaux existe, mais certains phénomènes ne sont pas pris en compte. L'objectif de cette étude est d'étendre ce formalisme pour en particulier intégrer ces phénomènes. Il permet de prendre en compte les informations conflictuelles d'un même acte multimodal, de différencier les gestes intentionnels des autres, et de faire la distinction entre la multimodalité parallèle (portée par un acte multimodal) et séquentielle (portée par la tâche). Son intérêt est de permettre une classification fine de la multimodalité dans un double but d'évaluation de son usage, et de formulation de recommandations ergonomiques relatives à l'intégration de la multimodalité dans les interfaces homme-machine.

MOTS CLEFS : multimodalité, propriétés CARE, acte multimodal, tâche

INTRODUCTION

L'objectif de cet article est de discuter des propriétés CARE puis de présenter une classification étendue des actes multimodaux intégrant des aspects tels que le conflit dans un acte multimodal, l'intentionnalité d'un geste, l'instanciation parallèle vs séquentielle de la multimodalité, et la gestion de la tâche dans un environnement multimodal.

La multimodalité est tantôt envisagée du côté de la machine tantôt du côté de l'utilisateur. [14] Martin 1995, [7] Carbonell & al 1994. Nous nous plaçons quant à nous sur un troisième axe, complémentaire des deux autres : la tâche que l'utilisateur doit réaliser. Celle-ci nous semble en effet au centre de l'interaction et la cause de la mise en relation de l'utilisateur et de la machine. D'ailleurs c'est bien l'interaction qui est multimodale, l'utilisateur ayant des canaux sensoriels multimodaux et la machine des dispositifs multimédia.

Les définitions que nous utilisons sont les suivantes :

- < Un acte a pour finalité l'atteinte puis la satisfaction d'un but. Lorsqu'il est multimodal, il met en jeu plus d'une modalité sensorielle chez l'homme et plus d'un dispositif chez la machine.
- < La réalisation d'une tâche est multimodale si les actes conduisant à sa réalisation utilisent plus d'une modalité, et ceci que les actions élémentaires soient effectuées de manière alternée ou parallèle.

L'intégration de plusieurs modalités dans la communication homme-machine demeure un problème tant technique qu'ergonomique à l'heure actuelle.

- < au niveau technique les questions se posent sur :
 - * la synchronisation de différentes modalités (geste, parole...) et leur intégration dans la construction du sens de la commande,
 - * la reconnaissance du geste (en deux ou trois dimensions), la performance de la reconnaissance vocale,
 - * etc.
- < au niveau ergonomique se posent les questions
 - * de l'utilisabilité des différentes modalités,
 - * du coût cognitif que leur utilisation engendre,
 - * de leur pertinence vis à vis des buts à atteindre,
 - * etc.

Nous nous intéressons dans cet article à ce second niveau, et plus particulièrement au troisième point qui conditionne l'usage des modalités par l'utilisateur. Ce domaine d'étude n'est pas nouveau [7] Carbonell 1994 ; [11,12] Hauptmann 1989 ; 1994 ; [18] Valot 1993 ; [16] Ozkan 1994. Ces différentes études sont, pour la plupart, des explorations empiriques relatives à l'usage des différentes modalités, et apportent un certain nombre de données sur les gestes, la nature du vocabulaire utilisé en interaction verbale, la synchronie / asynchronie entre parole et geste, etc. Certaines remettent en question la spontanéité de l'usage de la

multimodalité [17] Vacherand-Revel J. & Valot C. 1994, aspect qui est souvent un argument en défaveur de la multimodalité.

Nous ne rentrerons pas dans ce débat concernant l'aspect naturel ou spontané de la multimodalité. Nous avons observé dans plusieurs études, ainsi que dans les nôtres qu'elle existe [8] Catinis & Caelen 1995. Il semble donc important que l'usage de la multimodalité, même si elle est relativement peu fréquente doit se réaliser dans de bonnes conditions, étant peu probable que la technologie retourne en arrière dans la conception d'interfaces et abandonne la multimodalité. De plus l'ergonome ne doit pas nier un fait mais doit l'étudier.

Des études ont permis de classifier les actes multimodaux [5] Caelen & Coutaz 1991 ; [2,3] Bellik & Teil 1992 ; [13] Martin & Béroule 93 ; [14] Martin 1995, [9] Coutaz et Nigay 1994. Ces classifications sont devenues de plus en plus précises dans le temps.

Une partie de ces modèles considèrent les modalités du point de vue du système et les autres d'un point de vue de l'utilisateur. Nous retiendrons ce type de classification et plus particulièrement le découpage en propriétés CARE des actes multimodaux.

Ces propriétés dénombrent quatre types d'utilisation des modalités :

- < la complémentarité : chaque modalité utilisée est nécessaire à la compréhension de la totalité de l'acte,
- < l'assignation : la même modalité (ou sous-ensemble de modalités) est choisie pour une action donnée (appelée aussi spécialisation),
- < la redondance : les mêmes informations sont portées par les différentes modalités utilisées dans l'acte,
- < l'équivalence : les différentes modalités permettent d'obtenir le même résultat (équivalence de résultat), et peuvent être d'un coût cognitif différent (équivalence fonctionnelle). [18] Valot 1993.

L'utilisation de ces propriétés pour caractériser la réalisation de la tâche n'est guère envisageable du fait du manque de caractérisation concernant la temporalité, l'enchaînement des énoncés, leur contribution à la tâche, etc. Nous trouvons dans la littérature [6] Caelen 1992, [15] Nigay & Coutaz 1993 des notions de ce type, appliquées aux systèmes.

Les auteurs distinguent quatre types de multimodalités intégrant la notion de temporalité.

- < Multimodalité alternée : une seule modalité est utilisée à un instant donné, mais plusieurs sont utilisées pour un même énoncé,
- < Multimodalité exclusive : une seule modalité est utilisée à un moment donné, et pour un énoncé,

< Multimodalité synergique : plusieurs modalités sont utilisées par énoncé et en même temps,

< Multimodalité concurrente : plusieurs énoncés sont produits en même temps en utilisant des modalités différentes.

Ces notions sont intéressantes pour effectuer une analyse de la réalisation de la tâche dans un environnement multimodal.

Nous proposons une classification des actes et des tâches en environnement multimodal afin de rendre compte des différents usages de la multimodalité et permettre une évaluation et un retour sur la conception. Pour cela nous complétons d'une part les propriétés CARE et d'autre part nous leur intégrons les propriétés CASE (concurrence, alternance, synergie, exclusivité).

RECUEIL DES ACTES MULTIMODAUX

L'extension que nous proposons se base sur l'analyse de deux types de corpus en Communication Homme-Machine multimodale : une interface de téléphonie, et une interface de dessin.

Les expériences se déroulent selon une variante du paradigme du magicien d'Oz : la reconnaissance vocale est simulée (reproduisant les caractéristiques d'un système de reconnaissance à mots clefs), mais les sujets sont mis au courant avant l'expérimentation de la présence de compères effectuant la simulation. Il leur est donc demandé de "jouer le jeu" en respectant un vocabulaire limité (60 mots).

Les dialogues sont finalisés, les sujets ayant des tâches précises à réaliser. Les modalités autorisées pour la réalisation des tâches sont :

- < la parole
- < le geste avec le curseur de la souris (désignation et action)
- < les raccourcis claviers

Les sujets sont des salariés d'origines variées : cadres et secrétaires pour l'essentiel, utilisant de manière quotidienne l'outil informatique. Nos observations portent sur 14 corpus (7 de chaque situation expérimentale).

Nous n'explicitons pas plus avant la méthodologie employée, s'agissant ici de présenter notre classification, et non le détail des résultats. Cf. [10] Aublet-Cuvelier & al 1996 pour une description plus précise du dispositif expérimental.

MOTIVATIONS POUR UNE EXTENSION DES PROPRIETES CARE

Nous présentons ci dessous nos remarques sur les propriétés CARE ainsi qu'une extension, et la prise en compte de l'enchaînement des modalités lors de la réalisation d'une tâche multimodale.

L'acte multimodal

La première remarque porte sur les gestes en interaction Homme-Machine. D'une manière générale ils ont été étudiés, que ce soit dans le cadre de la communication homme-homme ou dans celui de la communication Homme-Machine.

Par exemple, [1] Bellalem 1995 analyse le geste de désignation en communication Homme-Machine. L'auteur fait la distinction entre les gestes accidentels (qui ne sont pas accompagnés d'acte de langage ou sans objet associé), des gestes de désignation ou de manipulation.

Dans nos études les gestes sont bidimensionnels, et limités à l'usage de la souris. L'utilisateur peut donc effectuer un geste de désignation, sans clic-souris, donc sans modification de son environnement de travail.

L'utilisateur peut aussi effectuer des modifications sur l'interface, par manipulation directe des objets de l'interface, ou désigner les objets par clic-souris. Nous considérons, comme N. Bellalem qu'un geste est "accidentel" ou fortuit s'il n'est pas accompagné d'un acte de langage et/ou s'il ne porte pas sur un objet.

Nous ajoutons ainsi une distinction pour chaque propriété, portant sur l'intentionnalité. L'ajout de cette notion permet de lever l'ambiguïté dans le cas ou geste et acte de langage sont cohérents, et permet aussi de ne pas faire d'interprétation inconsidérée.

Un geste fortuit peut en effet être un indice mnésique pour le raisonnement de l'utilisateur ou être aléatoire. Nous appellerons "actionnel" toutes les actions intentionnelles (clic-souris dans notre cas) et "informationnel" tous les autres gestes (accidentels ou fortuits).

Deuxièmement, nous avons observé que les informations portées par un acte multimodal sont parfois contradictoires. Il y a donc conflit entre les différentes modalités. Les conflits dans un acte multimodal peuvent être expliqués de plusieurs manières :

- < il peut s'agir d'un changement de planification de la part du sujet ; il commence à réaliser la commande avec la souris puis se ravise ou la rectifie et continue avec la parole,
- < il peut s'agir aussi d'une mauvaise représentation des capacités de compréhension du système (les surestimations ou sous-estimations sont connues en communication homme-machine : [11] Hauptmann 1989 ; [4] Bisantz & Sharit 1993.),

- < il peut s'agir tout simplement d'une mauvaise coordination des gestes et de la parole (l'exemple typique est lorsqu'on dit « tourne ici à droite » en désignant une direction sur sa gauche).

Quelle que soit la raison de ce conflit il est nécessaire que le système l'identifie comme tel pour qu'il puisse le gérer et prendre les décisions adéquates (action par défaut, demande de précision...)

Troisièmement, si l'on considère les différentes propriétés, on peut remarquer qu'elles ne sont pas toutes de même niveau. Le conflit, la complémentarité, et la redondance caractérisent l'acte multimodal lui même, sa structure, les éléments informationnels et les modalités qui le composent. Ces propriétés contribuent à l'analyse des actes multimodaux, leur aboutissement pouvant être la formulation de recommandations ergonomiques.

L'assignation, et l'équivalence se placent dans une autre dimension. L'usage des différentes modalités étant plus ou moins restreint par l'interface, l'observation d'assignation ou d'équivalence va permettre d'évaluer en termes de coûts cognitifs cette interface.

Ces diverses remarques nous conduisent au schéma suivant :

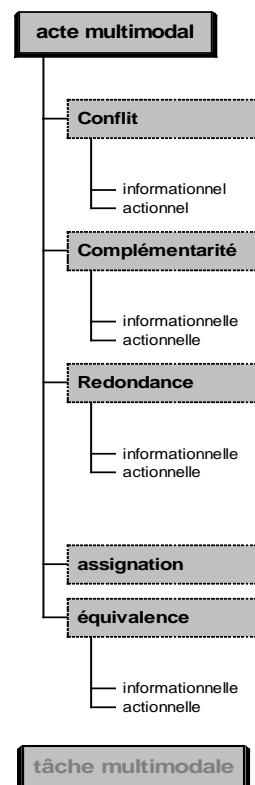


Fig. 1
Extension des propriétés CARE pour l'acte multimodal : les propriétés CCARE

La tâche multimodale

L'étude d'un acte multimodal permet l'étude fine des phénomènes intervenants pour une commande, mais l'utilisation des différentes modalités au cours d'une tâche peut varier à chaque commande. Pour réaliser son but l'utilisateur peut enchaîner les commandes les unes après les autres ou bien les entrelacer, changer de modalité suite à un échec, ou encore changer de stratégie selon le type de tâche. Pour rendre compte de ces différents phénomènes et permettre une analyse quantitative, il est nécessaire d'étudier l'usage de la multimodalité sur un axe temporel.

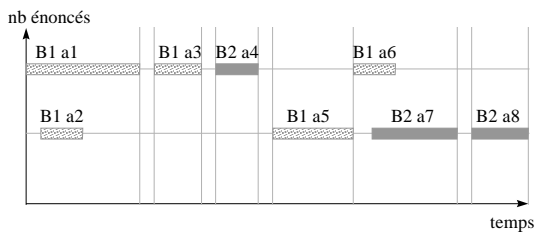


Fig. 3

Exemple d'enchaînement de buts et d'actions lors d'une tâche (B : but en cours, a : acte multi ou monomodal)

Dans l'exemple ci dessus, la tâche est multifils, elle doit satisfaire deux buts. Si l'on suppose que le graphe ci dessus est le résultat des actes nécessaires à la réalisation d'une tâche, on peut faire les observations suivantes :

- a1 et a2 sont synergiques, ils contribuent au même but et se déroulent durant la même fenêtre temporelle.
- a3, a5 sont alternés,
- a5 et a6 sont séquentiels.
- Les actes a6 et a7 sont concurrents puisqu'ils se déroulent dans la même fenêtre temporelle et qu'ils portent sur des buts différents.
- Nous ne retenons pas la notion d'énoncés parallèles, les actes synergiques et concurrents le sont par définition.

Nous obtenons ainsi trois types d'enchaînement des actes, ceci permettant d'examiner précisément la réalisation de chaque tâche, la gestion de la multimodalité et les changements de stratégies en fonctions des interruptions, des erreurs, des reprises, etc.

Ces remarques nous amènent à inclure les propriétés CCARE portant sur des actes isolés, dans un schéma plus large concernant la tâche et réhabilitant la temporalité. Cela conduit au schéma suivant qui définit les propriétés T-CCARE.

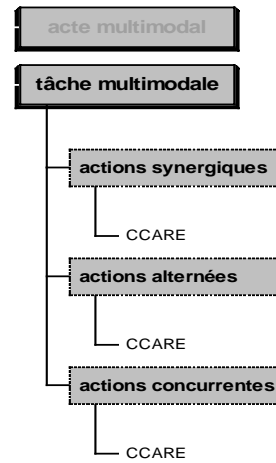


Fig. 2

Propriétés T-CCARE

EXEMPLES D'ACTES MULTIMODAUX

Le conflit

Le conflit actionnel : Dans l'exemple qui suit le conflit vient du fait que l'action et la commande orale ne concerne pas le même objet.

00:00:18	action	clic-cercle
00:00:22	parole	<i>tout sélectionner</i>

Tableau 1. Conflit au niveau de l'objet concerné par l'acte multimodal (sujet 05 dessin)

Le conflit informationnel : ici le conflit est de même nature que celui décrit précédemment, mais l'acte multimodal est accompagné d'un geste et non d'un clic-souris.

0:00:56	geste	geste-cercle
0:00:57	parole	<i>réduire l'ensemble</i>

Tableau 2. Conflit informationnel (sujet 04 dessin)

La complémentarité

La complémentarité actionnelle : l'exemple qui suit est typique de l'utilisation de la complémentarité.

0:00:03	action	clic-cercle
0:00:06	parole	<i>à droite</i>

Tableau 3. Complémentarité actionnelle (sujet 03 dessin)

La complémentarité informationnelle :

00:33:05	geste	déplace-répertoire
00:33:11	parole	<i>alors, ajouter</i>

Tableau 4. complémentarité informationnelle (sujet 11 téléphonie)

La redondance

La redondance actionnelle : celle-ci porte rarement sur l'acte entier. Nous avons identifié essentiellement deux sortes de redondance : celle qui porte sur l'objet de l'action et celle qui porte sur l'action elle-même. Dans l'exemple ci-dessous la redondance porte sur l'objet.

0:00:04	action	clic-barre
0:00:49	parole	<i>barre horizontale</i>

Tableau 5. Redondance sur l'objet de l'action (sujet 03 dessin)

La redondance informationnelle : la même distinction que précédemment peut être faite quand le geste est informationnel.

02:17:55	action	déplace-répertoire-bouton et bouton éditer
02:18:09	parole	<i>... dans répertoire, modifier la fiche de mireille Parran</i>

Tableau 6. Redondance informationnelle sur l'action (sujet 06 téléphonie)

EXEMPLES DE TACHES MULTIMODALES

Nous présentons ci-dessous une illustration des différentes informations que permet d'identifier le formalisme T-CCARE. Dans le tableau ci-dessous, la tâche du sujet est composée de deux buts : agrandir un objet et le déplacer.

	mm:ss	code
1	00:00:00	
2	00:00:04	clic-cercle droit + parole <i>plus grand</i>
3	00:00:06/00:00:12	compère réalisation
4	00:00:12	parole <i>plus grand</i>
5	00:00:13/00:00:19	compère réalisation
6	00:00:22	clic-cercle droit
7	00:00:25	déplace-cercle droit
8	00:00:27	déplace-ajustement-cercle droit

Tableau 7. Exemple de tâche multimodale (Sujet 05 dessin)

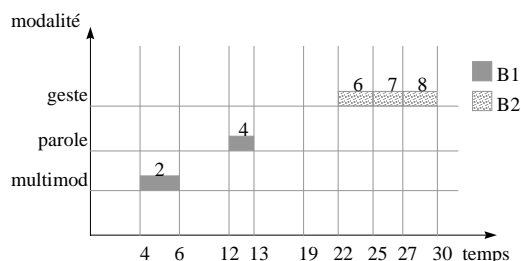


Fig.4 Exemple de réalisation de tâche multimodale (Sujet 05 dessin)

On peut voir que toutes les actions sont effectuées de manière séquentielle, qu'une seule action est

multimodale, etc. Le but de ce formalisme n'est cependant pas de donner une représentation visuelle de la réalisation de la tâche, mais de permettre une évaluation statistique de l'usage des modalités et du mode de gestion de celles-ci.

CONCLUSION

Nous avons vu comment certaines informations nous semblaient manquantes dans les formalismes existants pour caractériser les actes multimodaux, nous avons exposé pourquoi il nous semblait pertinent d'étendre ces formalismes à la tâche. Cette classification doit nous permettre de quantifier finement les actes multimodaux, de les caractériser, ainsi que d'analyser les différentes stratégies de gestion de la multimodalité lors de la réalisation de la tâche, dans différentes situations expérimentales.

BIBLIOGRAPHIE

- [1] Bellalem N. ; Romary L. (1995). Langue et geste pour le dialogue Homme-Machine finalisé, 01Design'95, p115-125
- [2] Bellik Y. & Teil D. (1992). Les types de multimodalité, IHM'92, p22-28
- [3] Bellik Y. & Teil D. (1992). définitions terminologiques pour la communication multimodale, IHM'92 p229-232
- [4] Bisantz A.M. ; Sharit J. (1993). The effects of feedback on performance and retention of skill for a natural language interface, Behavior and Information Technology, 12 (1), p32-47
- [5] Caelen J. ; Coutaz J. (1991). Interaction multimodale Homme-Machine : quelques problèmes généraux., IHM'91, Dourdan, p41-57
- [6] Caelen J. (1992). Compte-rendu du workshop IHMM. GDR-PRC Communication Homme Machine éd. Dourdan
- [7] Carbonell N. ; Valot C. ; Mignot C. ; Dauchy P. (1994). Etude empirique : usage du geste et de la parole en situation de communication homme-machine, ERGO'IA'94, p128-139
- [8] Catinis L. ; Caelen J. (1994). Analyse du comportement multimodal de l'utilisateur humain dans une tâche de dessin, IHM'95, p123-129
- [9] Coutaz J. ; Nigay L. (1994). Les propriétés CARE dans les interfaces multimodales, IHM'94, p7-14
- [10] Cuvelier L. ; & al (1996). NEIMO, un laboratoire d'utilisabilité numérique : leçons de l'expérience, Ergo'IA'96, à paraître

- [11] Hauptmann A.G. (1989). Speech and gesture for graphic image manipulation, CHI'89, p241-145
- [12] Hauptmann A.G. & McAvinney P., (1993). Gestures with speech for graphic manipulation. International Journal of Man-Machine Studies, 38 (2), 231-250.
- [13] Martin J.C. et Béroule D. (1993). Types et buts de coopérations entre modalités, IHM'93, p17-22
- [14] Martin J.C. (1995). Coopération entre modalités et liage par synchronie dans les interfaces multimodales, Thèse de troisième cycle, Paris
- [15] Nigay L. ; Coutaz J. (1993). A design space for multimodal system : concurrent processing and data fusion. Interchi'92, p172-178
- [16] Ozkan N. (1994). Analyses communicationnelles de dialogues finalisés, thèse, Grenoble.
- [17] Vacherand-Revel J. ; Valot C. (1994). la multimodalité "naturelle" dites vous ? Ergo'IA'94, p297-305
- [18] Valot C. (1993). Interface multimodale, Projet Grenoblois, Rapport d'activité du pôle Interfaces Homme-Machine Multimodales, GDR-PRC Communication Homme-Machine, p61-123
- [19] synthèse de l'atelier "interface multimodaux, IHM'93 Système d'analyse des interactions homme-ordinateur, (sous-groupe : formes de multimodalité en situation d'interaction utilisateur-machine)