
ESPACE DE CONCEPTION POUR LES INTERFACES MULTIMODALES.

Frédéric Vernier et Laurence Nigay

frederic.vernier@imag.fr, laurence.nigay@imag.fr

Laboratoire CLIPS-IMAG. Equipe IIHM

BP 53 38041 Grenoble cedex 9

Résumé — Dans cet article nous proposons un espace de conception qui clarifie l'analyse des interfaces multimodales en sortie et la mise en œuvre de nouvelles interfaces multimodales. Après avoir présenté dans quel contexte nous utilisons les termes liés à notre recherche nous présentons notre approche qui consiste à utiliser les relations de Allen [1] appliquées aux compositions spatiales, temporelles, articulatoires, syntaxiques et sémantiques de modalités. Chacune des différentes formes de multimodalité que nous avons identifiées est brièvement discutée et illustrée afin de mettre en évidence l'étendue des problèmes d'ingénierie couverts par notre réflexion.

Introduction

La multimodalité est un domaine vaste qui a déjà fait l'objet de nombreuses études. Du point de vue utilisateur de multiples questions se posent sur l'utilisation des interfaces multimodales. De notre point de vue, celui du concepteur, nous nous intéressons à la mise au point des composants logiciels liés à la multimodalité. Dans cette optique nous nous intéressons à l'interaction en sortie (du système vers l'utilisateur) bien que le terme multimodalité soit rarement utilisé dans ce cas. Si le terme multimédia est plus souvent retenu, il désigne davantage les supports de l'information échangée (médiats) que son contenu. Notre étude se consacre à l'interaction multimodale en sortie, qui peut être analysée selon le dispositif mis en œuvre et le langage utilisé [5]. Le concept de dispositif est largement répandu et reconnu car il correspond à des objets physiquement reliés aux systèmes informatiques. Le concept de langage d'interaction est plus discuté car il implique la reconnaissance des capacités d'un système informatique à communiquer avec un être humain. Cette communication est parfois basée sur un langage presque naturel mais elle repose la plupart du temps sur des langages plus simples et plus efficaces tels que les langages graphiques ou les langages de commande. L'intérêt du concepteur consiste à séparer la programmation du pilote de celle de l'interaction qui se trouve à un plus haut niveau d'abstraction. Aussi, dans la suite de cet article, nous considérons qu'une modalité, en entrée comme en sortie, désigne un moyen de communication qui met en œuvre un dispositif physique et un langage d'interaction.

Multimodalité en sortie

En sortie, l'interaction multi-dispositifs est un domaine assez pauvre à cause du manque de variété dans les dispositifs de sortie. Les manettes à retour d'effort ne sont utilisées que pour les simulateurs et les jeux. Les haut-parleurs constituent la seule alternative classique aux écrans.

Les modalités sonores (exploitant des haut-parleurs) soulèvent néanmoins des problèmes :

- Le volume des haut-parleurs peut être réglé (éventuellement à 0) par l'utilisateur, sans que le système en soit informé.
- Le volume de l'environnement sonore de l'utilisateur peut être plus élevé que celui de la modalité sonore.
- Le contexte d'interaction de l'utilisateur (lieu de travail par exemple) peut lui imposer un volume très bas, voire nul.

Pour toutes ces raisons, les modalités sonores sont rarement exploitées dans l'interaction, excepté dans certains types de logiciels comme la création musicale où elles sont inévitables. Nous constatons que, dans les interfaces traditionnelles, les modalités sonores sont généralement utilisées pour répéter l'information graphique (redondance et insistance).

Nous nous efforcerons, dans la suite de cet article, de nous rappeler que la multimodalité peut être appliquée aux systèmes multi-langages afin de ne pas restreindre nos résultats aux trop rares systèmes combinant utilement un écran et des haut-parleurs.

Le contexte de notre recherche étant posé, nous commencerons la présentation de notre étude par la distinction des étapes de sélection et de composition. Nous aborderons chacune de ces deux étapes avant de présenter notre espace de conception des systèmes multimodaux.

Sélection et composition de modalités

Comme nous l'avons annoncé, nous considérons deux éléments dans la conception d'une interface multimodale :

- la composition de modalités,
- la caractérisation d'une modalité simple ou composée en vue de sa sélection.

Bien que ces deux éléments de conception soient fortement liés, nous les considérerons néanmoins séparément à des fins analytiques. Ils sont essentiels à

la clarification du travail du concepteur. Pour chacun de ces deux points, nous proposons des cadres de conception.

Sélection

La sélection de la modalité est une étape particulière dans le cas de l'interaction en sortie car cette étape est liée à l'acteur initiant la communication. En entrée, le choix des modalités à mettre en œuvre est souvent laissé à l'utilisateur alors qu'en sortie cette étape peut être engagée par trois acteurs différents :

- Le concepteur qui programme le système pour utiliser la ou les modalités qu'il a choisies.
- Le système qui, au travers d'un algorithme plus ou moins évolué, met en œuvre la ou les modalités adéquates.
- L'utilisateur qui paramètre son système afin de recevoir les informations de ce dernier par le canal qui lui convient.

Un système multimodale implique forcément la sélection, par l'acteur concerné, de plusieurs modalités. En plus des caractéristiques inhérentes aux modalités qui vont guider ce choix, la forme que va prendre la composition est très importante. L'étape suivante consiste donc à caractériser les différentes compositions envisageables.

Composition

Dans un système interactif multimodal, plusieurs modalités sont mises en œuvre et leur composition est un point central de conception. Nous considérons cinq aspects de composition : les compositions temporelle, spatiale, articulatoire, syntaxique et sémantique des modalités.

La composition spatiale reflète l'agencement dans l'espace (ici l'environnement de l'utilisateur) des modalités. Nous considérons ensuite la composition temporelle qui correspond aux enchaînements possibles de modalités mises en œuvre. La composition articulatoire des modalités s'exprime à un niveau d'abstraction plus bas, entre les flux d'information transitant par les dispositifs associés aux modalités. Nous situons la composition syntaxique au niveau des langages d'interaction associés aux modalités. Enfin la composition sémantique s'attache à la composition des informations véhiculées par les modalités (niveau langage).

Nous avons choisi d'étudier l'ensemble des aspects de la composition (temporel, spatial, articulatoire, syntaxique et sémantique) selon les cinq schémas issus des relations de base de Allen [1] et reproduits graphiquement dans les titres de colonne du Tableau 1.

L'application des schémas de composition aux cinq aspects énumérés auparavant nous conduit à établir le Tableau 1 et nous permet de proposer un nom pour les 25 cas obtenus.






Composition					
Temporelle	Anachronique	Séquentielle	Concomitante	Coïncidente	Parallèle / Simultanée
Spatiale	Disjointe	Adjacente	Intersectée	Imbriquée	Recouvrante
Articulatoire	Indépendante	Fissionnée	Fissionnée et dupliquée	Partiellement Dupliquée	Dupliquée
Syntaxique	Différente	Complétion	Divergente	Extension	Jumelle
Sémantique	Concurrente	Complémentaire	Complémentaire et redondante	Partiellement redondante	Totalement redondante

Tableau 1- Application des schémas de composition aux cinq aspects de la composition : temporel, spatial, articulatoire, syntaxique et sémantique.

Composition temporelle de modalités

La séquentialité et le parallélisme dans l'usage des modalités ont été largement étudiés [2 et 4]. Au contraire, les trois autres schémas de composition temporelle du Tableau 1 sont nouveaux pour la composition de modalité et constituent un affinement important pour l'étude de l'interaction.

Deux modalités anachroniques se distinguent de la séquentialité par un intervalle de temps entre les modalités. La durée minimale de l'intervalle pour passer d'une composition anachronique à une composition séquentielle doit être fixée par le concepteur en fonction de ses besoins. Il peut par

exemple considérer deux modalités comme anachroniques si l'intervalle dépasse la limite de perception de causalité (en dessus d'une seconde).

La composition de concomitance de modalités intervient quand une modalité en remplace une autre, avec un certain intervalle de temps pendant lequel les deux modalités sont présentes. Comme pour les modalités anachroniques, la durée de cet intervalle de transition dépend des besoins du concepteur. Une telle composition est importante pour aider et accompagner l'utilisateur dans la transition d'une modalité à une autre. Elle nécessite cependant que le ou les dispositifs puissent rendre en parallèle ces deux modalités pendant quelques instants.

Enfin, la composition coïncidente implique que l'une des deux modalités n'apparaisse que lorsque la seconde est déjà présente. Cette même modalité se termine avant que la seconde soit achevée. Une telle relation est nécessaire lorsque le concepteur met en œuvre une modalité qui fait sens seule et une seconde modalité qui ne fait sens qu'au regard de la première. Si l'analyse identifie le besoin d'utiliser une telle composition, la modalité englobante ne doit pas disparaître tant que la seconde est encore présente (fenêtre modale par exemple).

Composition spatiale de modalités

L'aspect spatial de la composition est un trait plus particulier aux modalités de sortie. En effet si les aspects temporel et sémantique de la composition des modalités d'entrée sont importants, l'aspect spatial est souvent négligé car les dispositifs d'entrée sont souvent différents pour chaque modalité. En sortie, le problème est différent car les modalités ne partagent essentiellement que deux dispositifs : l'écran et les haut-parleurs. Si nous considérons que deux modalités peuvent recouvrir le même lieu (comme la surface d'un écran ou l'espace de propagation d'un son), c'est que les sens perceptifs humains (vision et ouïe) nous permettent de percevoir plusieurs éléments au même endroit. La transparence par exemple nous permet de percevoir deux informations au même emplacement d'un écran. De même, lorsque le son n'est pas spatialisé, tous les sons perçus au travers des haut-parleurs nous semblent venir du même endroit.

Deux modalités disjointes signifient que l'utilisateur peut percevoir une distance entre les modalités. A contrario, deux modalités adjacentes signifient que l'utilisateur perçoit ces modalités en des lieux différents mais comme côte à côte. Deux modalités intersectées, imbriquées ou recouvrantes dénotent trois types de transparence. La composition imbriquée définit par exemple le fondement des lentilles magiques ("Magic lenses") [6] : une lentille translucide recouvre une partie de l'espace d'information. Ces mêmes modalités de lentilles magiques s'intersectent lorsque le concepteur les considère les unes par rapport aux autres et non plus par rapport au fond sur lequel l'opération de la lentille s'applique. Enfin le système des "Pixels Miroir" [7] est un exemple de modalités recouvrantes où la totalité de l'écran est utilisée par deux modalités différentes : le document et le reflet de l'utilisateur.

Ces cinq types de composition sont importants pour le concepteur d'une interface graphique, généralement limitée par l'espace disponible à l'écran. Les modalités recouvrantes ou imbriquées utilisent moins d'espace, mais induisent généralement un effort de perception et d'interprétation de la part de l'utilisateur (problème de continuité visuelle).

Composition articulatoire de modalités

La composition des flux d'information qui transitent par les périphériques du système est un aspect important des systèmes multimodaux. Pour l'interaction en entrée, la conception de moteurs de fusion soulève des problèmes d'ingénierie. Les auteurs de [4] proposent le terme de fission qui paraît plus adapté à l'interaction en sortie.

Les cinq cas de composition articulatoires affinent ce terme de fission. Nous appelons l'application des

schémas présentés aux extrêmes du tableau "composition indépendante" et "composition dupliquée". Ces compositions traduisent respectivement l'absence de lien entre les couches bas niveau des modalités ou au contraire une duplication de toute l'information entre les deux modalités. La composition d'une modalité sonore et d'une modalité graphique sera forcément indépendante car deux dispositifs sont mis en jeu. A l'opposé les modalités véhiculées par chaque haut parleur d'un système monophonique sont simplement dupliquées.

Les modalités fissionnées proviennent, comme leur nom le suggère d'une fission de deux flux d'information qui ont été séparés. Les systèmes tels que les murs d'images mettent en œuvre une fission à bas niveau d'abstraction d'un flux vidéo découpé en sous-parties contiguës.

Les modalités "dupliqués et fissionnés" sont rarement mises en œuvre et nécessitent un contexte spécial. Nous considérons par exemple comme dupliquées et fissionnées deux vidéo issues d'une même source mais dont chacune est sous-titrée dans une langue différente.

Les modalités partiellement dupliquées présentent par contre un plus grand intérêt. On constate en effet que certains dispositifs ont la capacité de présenter des informations identiques mais en quantité ou qualité différentes. L'écran d'une station de travail et d'un ordinateur de poche peuvent par exemple présenter la même image mais celle de l'ordinateur de poche sera tronquée par rapport à celle de la station de travail. Un document Postscript sorti sur imprimante peut également avoir une meilleure résolution que son double affiché à l'écran.

Composition syntaxique de modalités

Comme la composition articulatoire se réfère au dispositif de la modalité, la composition syntaxique de modalités implique le langage d'interaction. Nous distinguons, grâce aux cinq aspects, cinq niveaux de composition syntaxique qui vont de la composition de modalités dont le langage d'interaction est le même (composition jumelle) jusqu'à la composition de langages d'interaction différents (sans rapport). La composition de complétion dénote l'utilisation de deux langages au pouvoirs d'expression complémentaires. Le langage naturel complète par exemple très bien celui des dates afin de produire des énoncés tel que "*la réunion aura lieu le 10/03/1000*". La composition de modalités dont les syntaxes sont divergentes indique l'utilisation de langages d'interaction ayant une base commune mais dont certaines parties divergent. C'est le cas par exemple du langage écrit et du langage parlé. Enfin nous proposons de considérer la composition d'extension qui dénote des langages aux pouvoirs d'expression imbriqués. Par exemple, le texte formaté par des attributs tels que italique, gras ou encore souligné est un langage plus étendu que le texte sans formatage.

Composition sémantique de modalités

L'aspect central à considérer dans une composition de modalité est la relation entre la sémantique des informations véhiculées par les deux modalités. Les formes de multimodalité qui ont été le plus étudiées sont sans nul doute la complémentarité et la

redondance [2, 3]. La concurrence entre modalités, identifiée dans l'espace de conception de [4], n'est généralement pas considérée comme un cas de multimodalité car une analyse séparée des deux modalités est préférable. Par contre les modalités que nous appelons "partiellement redondantes" et celles que nous appelons "complémentaires et redondantes" nécessitent un complément d'explications.

Les modalités "complémentaires et redondantes" véhiculent des informations dont une partie seulement se répète. Une telle composition de modalités graphiques est souvent nécessaire dans les vues multiples qui présentent des informations différentes, mais dont une partie est commune pour permettre à l'utilisateur de faire un lien entre les informations de chaque vue.

Nous appelons une modalité M1 "partiellement redondante" par rapport à une autre modalité M2 lorsque M1 reprend une partie du sens véhiculé par M2 sans en reprendre la totalité. C'est un aspect très courant dans le domaine de la visualisation d'une grande quantité d'information car les vues synthétiques de l'espace, proposées dans ne nombreux systèmes, n'offrent en général qu'une redondance partielle avec la vue principale de l'espace.

Aspects et schémas : un espace unificateur

Une composition de modalité de sortie fait intervenir l'une des cinq compositions pour les aspects temporels, spatiaux, syntaxiques et sémantiques. Ces quatre axes indépendants impliquent donc 5^4 relations différentes entre modalités bien qu'il soit rare que le concepteur se préoccupe des quatre aspects à la fois. Nous relevons la particularité de la troisième ligne du tableau (composition articulatoire) qui n'est pas indépendante des deux dernières lignes (compositions syntaxique et sémantique). Certaines combinaisons ne nous apparaissent pas possibles. Ainsi, si une composition repose sur un flux d'information dupliqué vers deux dispositifs de sortie, les deux modalités ne sont pas sémantiquement complémentaires ou de langages qui se complètent. La composition "fissionnée et dupliquée" et la composition "partiellement dupliquée" sont également restrictives. Les deux compositions "indépendantes" et "fissionnées" par contre autorisent toutes les compositions au niveau syntaxique et sémantique.

Notre espace de compositions est à mettre en rapprochement avec les espaces existants comme CARE [2], TYCOON [3] et MSM [5]. Ainsi dans l'espace TYCOON et CARE, les concepts de complémentarité et de redondance relèvent de composition sémantique. Dans l'espace MSM [5] les aspects articulatoires et sémantiques sont identifiés par un axe "niveau d'abstraction" pour différencier Multimédia et Multimodale. Enfin dans ces trois espaces, la distinction entre modalités séquentielles et parallèles relève de l'aspect temporel de la composition de modalités. Cet aspect de la composition est sans doute le mieux documenté puisque les concepteurs disposent depuis longtemps des relations de Allen [1]. L'aspect spatial de la composition est peu abordé dans les aides à la composition de modalités en entrée car chaque modalité étant associée à un dispositif, la position des modalités est laissée à la discrétion de l'utilisateur.

L'espace de conception que nous proposons complète et unifie les espaces TYCOON, CARE et MSM. Notre espace est plus complet puisqu'il propose 25 types de compositions rangées derrière les cinq aspects fondamentaux que nous avons conjugués avec cinq schémas de composition communs.

Conclusion

Dans cet article, nous avons considéré une interface multimodale en sortie selon deux angles d'étude : la composition de plusieurs modalités et la caractérisation d'une modalité. Nous avons défini un espace de conception formé de deux axes (aspect et schéma) pour identifier 25 compositions différentes. La caractérisation d'une modalité est importante pour permettre le choix d'une modalité, que celui-ci soit fait par le concepteur, le système ou laissé à l'utilisateur. L'une des contributions importantes de nos espaces est d'unifier et de compléter les approches existantes, que ce soit pour les interfaces en entrée ou en sortie.

Pour la complétude de l'étude, il convient aussi d'étudier les relations entre une modalité d'entrée et de sortie, puisque c'est de leur couplage qu'émerge l'interaction. Peut-on appliquer notre espace de conception pour la composition d'une modalité d'entrée avec une modalité de sortie ?

Bibliographie

- [1] Allen J. *Maintaining Knowledge about Temporal Intervals*. Communication of the ACM. Volume 26. N°11. Novembre 1983. pp. 832-843.
- [2] Coutaz J., Nigay L., Salber D., Blandford A., May J. et Young R., *Four easy pieces for assessing the usability of multimodal interaction: The CARE properties*. Actes de la conférence Interact'95 (Lillehammer, Norway), 1995, pp. 115-120.
- [3] Martin J.C., *Six primitive types of cooperation for observing, evaluating and specifying cooperations*. Actes de la conférence AAAI Fall 1999 (November 5-7th, 1999, Cape Cod, North Falmouth, Massachusetts, USA), 1999.
- [4] Nigay L. et Coutaz J., *A Design Space For Multimodal Systems: Concurrent Processing And Date Fusion*. Actes de la conférence INTERCHI'93 (24-29 avril, Amsterdam, Hollande), 1993, pp. 172-178.
- [5] Nigay L. et Coutaz J., *Espaces conceptuels pour l'interaction multimédia et multimodale*, TSI, spécial Multimédia et Collecticiel, AFCET &Hermes Publ., Vol 15(9), 1996, pp. 1195-1225.
- [6] Stone M., Fishkin K. et Bier E. *The Movable filter as a user interface tool*. , Actes de la conférence CHI'94 (24-28 Avril, Boston, Massachusetts), 1994, pp. 306-312.
- [7] Vernier F., Lachenal C., Nigay L. and Coutaz J. *Interface Augmentée par effet Miroir*, Actes de la conférence IHM'99 (22-26 Novembre, Montpellier, France), 1999, pp. 158-165.