# Atelier IHM'92

# Interfaces Multimodales et Architecture Logicielle

**Coordinateur**: Joëlle Coutaz (LGI-IMAG).

**Comité de lecture**: Noëlle Carbonell (Inria-Crin), Jean-Claude Martin (LIMSI), Nadine Vigouroux (IRIT).

**Rédacteurs**: F. Azémard (IRIT), T. Bau del (LRI), Y. Bellik (LIMSI), M-L. Bourguet (ICP), N. Carbonell (INRIA-CRIN), J. Coutaz (LGI-IMAG), J-C. Martin (LIMSI), L. Nigay (LGI-IMAG), K. Ouadou (ECL), P. Palanque (LIS), D. Teil (LIMSI), N. Vigouroux (IRIT).

**Participants :** F. Azé mard (IRIT), T. Baudel (LRI), Y. Bellik (LIMSI), S-K. Bennacef (LIMSI), M-L. Bourguet (ICP), N. Carbonell (INRIA-CRIN), R. Caubet (IRIT), J. Coutaz (LGI-IMAG), V. Gaildrat (IRIT), J-C. Martin (LIMSI), L. Nigay (LGI-IMAG), K. Ouadou (ECL), P. Palanq ue (LIS), D. Teil (LIMSI), N. Vigouroux (IRIT).

## 1. Introduction

Ce document concrétise les résultats de l'atelier *interfaces multimodaleset architecture logicielle* organisé dans le cadre de IHM'92, quatrièmes journées sur l'Ingé nierie des Interfaces Homme-Machine.

Une interface homme-machine multimodale (IHM-M) autorise l'utilisation conjointe des différentes modalités sensori-motrices de la communication humaine tels la vision, la langue naturelle et le geste. Elle va au-delà des interfaces multimédia puisqu'elle est capable de "comprendre" les informations échangées sur les différents canaux de communication. La réalisation des IHM-M introduit de nouvelles difficultés, et notamment l'inté gration de modalités multiples au sein d'une architecture appropriée. Notre groupe de travail s'est fixé comme objectif l'analyse de ce processus d'intégration.

Pour cela, nous adoptons comme fondement le modèle d'architecture Arch que nous présentons au paragraphe 2. Nous constatons que l'intégration des modalités se caractérise selon plusieurs dimensions. Ces caracté ristiques font l'objet du paragraphe 3. Elles sont illustrées au paragraphe 4 au moyen de solutions concrètes implémentées par les participants de l'atelier. En conclusion, on trouvera aux paragraphes 5 et 6 une synthèse sur le processus d'intégration ainsi qu'une liste de problèmes non traités pouvant servir d'agenda de recherche à moyen terme.

## 2. Modèle de référence : Arch

### 2.1. Les éléments du modèle Arch

Comme le montre la figure 1, le modèle Arch [UIMS 92] s'appuie sur les composants conceptuels du modèle Seeheim [Pfaff 85]. On y retrouve les notions de noyau fonctionnel, de contrôleur de dialogue et de pré sentation.

**Figure 1**: Les composants du modèle Arch.

Les pieds de l'arche constituent les composants imposés par la réalité : le noyau fonctionnel réalise les concepts du domaine ; le composant d'interaction, en contact direct avec l'utilisateur, est mis en œ uvre au moyen des objets d'interaction d'une boîte à outils. En clé de voûte, le contrôleur de dialogue gère l'enchaînement des tâches ainsi que les liens entre les objets regroupés dans les deux composants voisins : l 'adaptateur de domaine et la présentation.

L'adaptateur de domaine sert, pour l'essentiel, à ajuster les différences de modélisation des objets conceptuels entre le noyau fonctionnel et le contrôleur de dialogue. Le composant de présentation joue un rô le similaire : il permet au contrôleur de dialogue de s'affranchir du fonctionnement de la boîte à outils du niveau interaction. La présentation peut se voir comme une boîte à outils virtuelle qui implémente des objets de pré sentation concrétisés en fin de compte par les objets d'interaction d'une boîte à outil réelle.

La figure 1 doit se voir comme une représentation canonique. En réalité, l'importance relative des cinq composants de Arch varie en fonction du cas à traiter. Cette migration des fonctions entre les composants s' exprime métaphoriquement au moyen du métamodèle *slinky*, du nom de ce jouet qui, une fois mis en mouvement, voit sa masse se déplacer dynamiquement. Le choix de la répartition des fonctions entre les composants relè ve du savoir-faire en ingénierie du logiciel. L'équilibre doit être le résultat d'un dosage analytique des facteurs qualité comme l'efficacité et la portabilité.

Nous appelerons *instance d'Arch*, un exemplaire de modèle Arch produit en appliquant slinky à un système interactif particulier. Nous conviendrons aussi que l'*interface homme-machine*(IHM) d'un système regroupe tous les composants d'une instance d'Arch à l'exception du noyau fonctionnel.

### 2.2. Utilisation du modèle Arch dans l'atelier

Le modèle Arch et slinky, son compagnon, nous permettent de préciser les notions d'événement et d'énoncé. Ils mettent également en évidence des niveaux privilégiés pour l'inté gration des modalités.

#### *2.2.1. Evénement*

Un événement est l'unité d'information qui transite entre les composants interaction et présentation du modèle Arch. On le caractérise en géné ral par un type, une date, et un contenu informationnel (ou valeur).

Le niveau d'abstraction du contenu informationnel dépend de l'instance d'Arch considérée. Par exemple, selon les outils graphiques employés au niveau interaction, l'évé nement traduit un simple clic souris ou bien véhicule l'identité de l'objet désigné au moyen de la souris ; ou encore, selon le système de reconnaissance de la parole du composant d'interaction, l'évé nement contient un mot ou une phrase.

#### *2.2.2. Enoncé*

Un énoncé est l'unité d'information qui transite entre le contrôleur de dialogue et l'adaptateur de noyau fonctionnel. Il traduit un changement d'état de l'IHM dont l' effet doit se poursuivre dans le noyau fonctionnel ou, dans le sens inverse, un changement d'état du noyau fonctionnel qui doit être reflété dans l'IHM.

La nature de l'énoncé dépend de l'instance d'Arch. Dans les exemples de réalisation de notre groupe de travail, il modélise généralement une commance complète dont l'effet sémantique doit ê tre calculé par le noyau fonctionnel. Cependant, nous n'avons pas de réponse claire sur la nature des énoncés émanant du noyau fonctionnel.

#### *2.2.3. Niveaux d'intégration*

La fusion des événements en énoncés, et le phénomène inverse, est un processus procédant par étapes. Les composants du modèle Arch définissent des niveaux de référence pour ces étapes.

## 3. Intégration des modalités : les dimensions

Notre analyse sur l'intégration des modalités s'appuie sur une taxonomie des interfaces multimodales ébauchée lors des journées de travail IHM' 91. On trouvera dans [Bourguet 92, Coutaz 92, Nigay 93], un affinement de cette proposition initiale. Il ressort de cet effort de classification deux composantes intéressantes pour la mise en œuvre des interfaces multimodales : l' usage temporel (c.-à-d. parallèle ou séquentiel) des dispositifs d'entrée/sortie, et l'usage d'une seule ou de plusieurs modalités.

Les systèmes synergiques, qui permettent l'usage parallèle de plusieurs modalités, nécessitent de "fusionner" (intégrer) les événements en vue de construire des énoncés. Dans l' atelier, nous nous sommes intéressés à ce processus de construction. Par manque d'expérience, nous n'avons pas étudié le phénomène inverse qui, partant d'un énoncé produit par le noyau fonctionnel, aboutit à une restitution multimodale.

Nous présentons ci-dessous les dimensions susceptibles d'orienter le processus d'intégration des modalités pour les entrées.

### 3.1. Le type de système multimodal

Notre taxonomie de référence identifie quatre points singuliers : les multimodalités alternée, exclusive, synergique ou concurrente.

Un système donné peut n'autoriser qu'un type de multimodalité. Un autre peut être hybride et permettre, selon les contextes, plusieurs types de multimodalité. Dans ce cas, le modèle d'inté gration doit être capable de gérer plusieurs stratégies de fusion et de décider dynamiquement de la stratégie idoine.

### 3.2. Existence d'une modalité dominante

Une modalité est dominante lorsqu'elle détermine la forme et la structure de l'interaction. Elle constitue le véhicule d'expression privilégié. Les autres modalités disponibles sont utilisé es plutôt pour enrichir la modalité dominante.

Une modalité dominante peut servir de pivot d'intégration. En général, un pivot spécialisé accélère les calculs. On pourrait espérer ainsi améliorer les performances du processus d'intégration.

### 3.3. Stratégie d'intégration

La stratégie d'intégration précise les niveaux d'abstraction de Arch concernés par l'intégration. Elle peut conduire à une intégration *progressive*le long des diffé rents niveaux d'abstraction de Arch, ou bien *unique*(dans un seul composant). Elle peut être *précoce*("eager") ou *différée*("lazy") :
- une stragégie précoce a lieu dès les niveaux bas de Arch : elle tente de fusionner les événements issus du composant d'interaction. Elle conduit généralement à une fusion progressive ;
- une stratégie retardée a lieu dans le contrôleur de dialogue : elle est remise au niveau d'abstraction le plus haut. La fusion a donc lieu dans un composant unique.

Une stratégie précoce semble offrir les moyens de produire les retours d'information réactifs de la manipulation directe qui ne relèvent pas de la sémantique du domaine. A l'inverse de la stratégie retard ée, une stratégie précoce peut être amenée à défaire des fusions.

### 3.4. Critères d'intégration

Un critère d'intégration définit une condition pour fusionner ou non des informations. Nous avons identifié les critères suivants :

- la proximité temporelle sert à mettre en correspondance des événements issus de modalités différentes mais produits en des instants très proches.

- la fenêtre temporelle a plusieurs utilités :
  . un intervalle pour la détection des proximités temporelles,
  . une limite d'attente d'événements mutlimodaux.

- la complémentarité logique (structurelle) des événements permet dans certains cas de fusionner, au sein d'une même commande, des événements distants temporellement.

- la complétude d'une structure de données d'intégration peut constituer une condition de passage entre niveaux d'abstraction. Cette complétude peut servir également à limiter d'attente d'événements.

- les contextes de dialogue et historique d'interaction interviennent dans la résolution des coréférences, anaphores et ellipses.

- l'incompatibilité des modalités épargne au processus des tentatives d'intégration de modalités ne pouvant être utilisées conjointement.

### 3.5 Stratégie d'exploration

Le processus d'intégration élabore un espace de possibilités de relations entre modalités. Cet espace est parcouru soit *en largeur d'abord*soit en *profondeur d'abord*:

- En largeur d'abord, il s'agit d'identifier les relations possibles entre les modalités à un niveau d'abstraction donné et choisir la meilleure avant de passer au niveau suivant (et ceci de manière répétitive jusqu 'à atteindre le noyau fonctionnel).

- En profondeur d'abord, le processus de fusion exploite la première relation (souvent temporelle) et l'enrichit à travers les niveaux d'abstraction en fonction des relations structurelles et sé mantiques ; c'est seulement au niveau sémantique que sont validées ou réfutées les interprétations. Les retours arrière dans l'espace des possibilités se font progressivement à travers les niveaux.

### 3.6. Technique de représentation

La technique de représentation précise la nature des structures de données employées comme support aux informations fusionnées.

Le formalisme de représentation peut être *unique*(quel que soit le niveau d'abstraction) ou *multiple*(adapté au niveau d'abstraction)

Les structures de données peuvent être *dynamiques*ou *statiques*.

### 3.7. Commentaires

Notre liste de dimensions n'est pas exhaustive et ses termes manquent encore de rigueur. Toutefois, elle nous a permis lors de l'atelier, de clarifier le processus d'intégration et de mettre en é vidence les points communs, les divergences et les aspects complémentaires des solutions proposées.

## 4. Intégration des modalités : exemples de solutions

Les expériences et solutions présentées ici sont le fruit d'études menées par les membres de l'atelier. Chaque approche se situe dans la grille d'analyse de la section précédente et montre commen t est réalisée la fusion pour le cas d'école "mets ça ici" et pour un exemple privé mettant en évidence ses spécificités.

**4.1. L'architecture de référence: PAC-AMODEUS [L. Nigay & J. Coutaz, LGI]**
**4.2.**
**4.3. La fusion dans l'interface du modeleur déclaratif [N. Vigouroux & al, IRIT]**

**(sera inséré mercredi 13)**

## 4.4. L'intégration à un niveau conceptuel [Azémard - IRIT]

Dans notre modèle, l'intégration événementielle ne se limite pas à la seule mise en correspondance des événements pour une fusion. Il s'agit aussi de prendre en compte d'autres connaissances ... L'inté gration est donc réalisée au sommet de la pyramide ARCH par le contrôleur de dialogue. Lorsque les événements sont fusionnées dès le premier niveau, ils n'existent plus en tant qu'événements dans les niveaux supé rieurs. Ils sont phagocytés, et il faut déstructurer si les associations ne sont pas bonnes. Nous pensons que l'événement doit exister du plus bas niveau jusqu'au contrôleur de dialogue.

Pour ce, tous les événements sont amenés jusqu'à un même niveau de représentation. Le formalisme adopté est très proche des graphes conceptuels [Sowa 84]. L'interprétation est ensuite le fait de règles à d éclenchement réflexe. Ces règles sont activées par la présence de sommets non-instanciés dans un graphe. Un graphe traduit ce qui a été compris d'un événement. Nous parlons de graphe de compréhension.

Il convient de préciser que dans les applications visées, la parole sert de pivot à l'interprétation. Les mots d'un événement parole ont au travers d'un lexique une ou plusieurs repré sentations conceptuelles. A un mot qui n'est pas concept de l'application, correspond un concept générique complété d'un marquage lexical. Détailler les mécanismes de **jointure**et **filtrage**qui nous permettent d'obtenir le graphe d'un énoncé serait trop long ici, nous renvoyons le lecteur à [Azémard 92-a] et [Azémard 92-b].

Nous présentons au travers des exemples suivants les étapes successives qui conduisent des premiers graphes jusqu'au graphe final de compréhension d'un énoncé.

| |
|---|
| Les graphes doivent se lire de la manière suivante :<br>[*xclasse] : est une variable qui capture uniquement une instance de la descendance de la classe spécifiée, et x est un chiffre permettant de distinguer plusieurs variables d'une même classe.<br>{marquage} : marquage lexical de premier niveau, pronom, article défini, déictique ou adverbe ... |

### 4.4.1. Notre exemple imposé : "mets ça là" accompagné de deux clics souris

a - Les premiers graphes conceptuels
- Modalité souris : A un clic souris correspond un objet de l'application, ici les graphes sont des sommets isolés. Soient [Objet14] pour le clic1 et [Lieu21] pour le clic2.
- Modalité parole : le lexique fournit les graphes suivant :

mets : [Mettre0] -
(objet) -> [*1objet]
(lieu) -> [*1lieu]
ça : [*2objet{déictique}]
la : [*2lieu{adverbe}] ou [*3objet{pronom}] ou [*4objet{article défini}]

b - Construction du graphe de compréhension de l'événement parole

[*4objet{article défini}] est écarté par une règle syntaxique spécifiant qu'un article doit précéder un nom. Dans la hiérarchie des classes, OBJET est ancê tre de LIEU. Plusieurs associations sont donc possibles. Nous observons quatre cas :

1 : [*1objet] ∫ [*2objet{déictique}] et [*1lieu] ∫ [*2lieu {adverbe}]
2 : [*1objet] ∫ [*2objet{déictique}] et [*1lieu] ∫ [*3objet{pronom}]
3 : [*1objet] ∫ [*2lieu{adverbe}] et [*1lieu] ∫ [*2objet{déictique}]
4 : [*1objet] ∫ [*3objet{pronom}] et [*1lieu] ∫ [*2objet{déictique}]

Pour choisir, nous employons une mesure d'éloignement sémantique. Cet éloignement est la somme des spécialisations à réaliser sur l'ensemble des sommets concernés par la jointure dirigé e pour rendre notre graphe conceptuel connexe. Ici, le cas 1 est retenu, l'éloignement est égal à zéro. Pour tous les autres, il est supérieur ou égal à 1. En effet, il y a au moins une spécialisation à faire pour passer de objet à lieu. Le graphe de compréhension de l'événement parole, après jointure dirigée, est donc le suivant :

[Mettre0] -
(objet) -> [*1objet ∫ *2objet{déictique}]
(lieu) -> [*1lieu ∫ *2lieu {adverbe}]

c - Intégration

L'intégration consiste à associer une valeur à chaque sommet non-instancié du graphe de compréhension. L'ordre de passage est donné par la position des mots dans l'événement parole. C'est au travers de rè gles qu'est réalisée cette intégration.

Exemple de règle : **SI** *élément de File_Evénement = *évt_clic_souris*
*élément de File_Sommet = *Sommet* **déclencheur**
*type de *Sommet = non_instancié*
*marquage de *Sommet = "déictique"*
*(mode_synergique)*
**ALORS** *(tentative_intégration *évt_clic_souris *Sommet)*
*type de *Sommet <- instancié*

Graphe de compréhension de l'énoncé renvoyé par le contrôleur de dialogue :

[Mettre0] -
(objet) -> [Objet14]
(lieu) -> [Lieu21]

L'objet désigné par le premier clic souris sera placé à l'endroit pointé par le deuxième clic.

### 4.4.2. Un exemple, où Intégration n'est pas toujours Fusion :

Il faut parfois observer le contexte pour lever des ambiguïtés. Par exemple, dans un même événement parole, un terme ambigu peut être associé à un ou plusieurs clics souris et un autre terme verra son ambiguït é levée par une recherche dans l'historique du dialogue.

Considérons l'exemple suivant, où l'utilisateur vient juste d'ouvrir une fenêtre. Il prononce alors "mets ce couple dans la fenêtre" et l'accompagne de deux clics. Sur des bases morphologiques, le premier clic serait associ é à couple et le second à fenêtre. Or il faut comprendre "mets-y ce couple" . Au niveau conceptuel il apparaît qu'un couple est formé de deux éléments. Ce sont ces deux éléments que l'utilisateur désigne successivement à l'écran.

Les graphes conceptuels des clics souris identifient deux objets [Objet17] et [Objet23]. Le graphe de compréhension de l'événement parole est le suivant :

```
[Mettre4] -
        (objet) -> [*1couple{déictique}] -
                        (élement1) -> [*1objet{déictique}]
                        (élément2) -> [*2objet{déictique}]
        (lieu) -> [*1fenêtre{article_défini}]
```

Le marquage lexical "déictique" s'est étendu aux composants du couple. L'emploi des déictiques est généralement associé à la modalité synergique. De même, la référence à un objet à l'aide d'un article défini indique très souvent qu'il s'agit d'un objet précédemment cité.

Le graphe conceptuel complètement instancié sera le suivant :

```
        [Mettre4] -
                (objet) -> [couple12] -
                        (élement1) -> [Objet17]
                        (élément2) -> [Objet23]
                (lieu) -> [Fenêtre3]
```

Les deux objets désignés à l'écran par l'utilisateur sont donc déplacés dans la fenêtre qu'il venait d'ouvrir.

### 4.5. Fusion et modélisation connexionniste [J.C. Martin & D. Béroule, LIMSI]

### 4.6. [M.L. Bourguet, ICP]

## 5. Intégration des modalités : synthèse

## 6. Problèmes et perspectives

## 7. Références


[Azémard 92-a]
F. Azémard, U. Arronategui, T. Huguet, J.P. Macchion, F. Mieulet : Interprétation et gestion du dialogue sous LOIR. Séminaire Dialogue, Pôle Parole et Langage Naturel du GDR-PRC CHM, Dourdan (Avr. 1992).

[Azémard 92-b]
F. Azémard, J.P. Macchion, F. Mieulet: Multimodal : le réflexe pour une stratégie d'interprétation dans IHM'92, Telecom Paris (déc. 1992)

[Bourguet 92]
M.L. Bourguet, J. Caelen : Interfaces Homme-Machine multimodales : gestion des événements et représentation des connaissances" dans ERGO-IA'92, pp. 124-134.

[Bellik 92a]
Y. Bellik, D. Teil : Les types de multimodalité dans IHM'92, Telecom Paris (déc. 1992) pp. 22-28.

[Bellik 92b]

Y. Bellik, D. Teil : Définitions terminologiques pour la communication multimodale, dans IHM'92, Annexes, pp. 229-232.

[Caelen 92]
J. Caelen : Compte-Rendu du "Workshop" IHM-M organisé par le GDR-PRC "Communication Homme-Machine" à Dourdan les 13 & 14 avril 1992, dans IHM'92, Annexes, pp. 213-228.

[Coutaz 92]
J. Coutaz ?????????

[Gourdol 91]
A. Gourdol : Architecture des Interfaces Homme-Machines multimodales, méméoire de DEA Informatique, IMAG, Grenoble, Juin, 1991.

[Nigay 92]
L. Nigay, J. Coutaz : Interfaces multimodales et architecture : fusion et parallélisme, dans IHM'92, Telecom Paris (déc. 1992) pp. 29-35.

[Pfaff 85]
G. Pfaff, P.J.W. Ten Hagen : Seeheim workshop on User Interface Management Systems. Berlin : Springer-Verlag.

[Sowa 84]
J.F. Sowa, Conceptual Structures, information processing in mind and machine. Addison Wesley, (1984).

[UIMS 92]
The UIMS Tool Developers Workshop : A Metamodel for the Runtime Architecture of an Interactive System. In SIGCHI bulletin, ACM Press, January, 1992, 24 (1), pp.32-37.

[Vigouroux 92]
N. Vigouroux, V. Gaildrat, R. Caubet, G. Pérennou : Une architecture d'interface pour l'interprétation d'informations multimodales : application à un modeleur déclaratif de scènes dans IHM'92, Telecom Paris (déc. 1992) pp. 49-56.