

ESPACE PROBLEME, FUSION ET PARALLÉLISME DANS LES INTERFACES MULTIMODALES

Laurence Nigay, Joëlle Coutaz

Laboratoire de Génie Informatique-IMAG
BP 53 X, 38041 Grenoble Cedex
Tel. 76 51 44 40, 76 51 48 54
E-mail : nigay@imag.fr, coutaz@imag.fr

Résumé

L'une des nouvelles préoccupations en Communication Homme-Machine est l'extension des facultés sensori-motrices des ordinateurs en vue de répondre aux moyens de communication naturels de l'Homme. Cet article propose un référentiel permettant l'analyse et l'étude des systèmes multi-sensori-moteurs (MSM) actuels et à venir. Pour ce faire, nous adoptons une orientation système tout en nous appuyant sur l'architecture cognitive ICS. Notre espace problème comprend six dimensions. Les deux premières ont trait à la notion de canal de communication : le nombre et le sens des canaux permis pour un système donné. Les autres dimensions caractérisent le degré de sophistication cognitive des fonctions d'interprétation et de restitution du système : niveaux d'abstraction, contexte, fusion et fission des informations échangées, granularité du parallélisme. Nous montrons ensuite comment ces aspects explicités par le référentiel MSM se projettent en terme d'architecture. En particulier, la fusion et le parallélisme font l'objet d'une présentation détaillée dans le cadre de notre modèle conceptuel d'architecture, PAC-Amodeus que nous illustrons au moyen d'un exemple concret.

Abstract

One of the new design goals in HCI is to extend the sensory-motor capabilities of computer systems to better match the natural communication means of human beings. This article proposes a framework that should help reasoning about current and future Multi-Sensory-Motor systems (MSM). To do so, we adopt a system centered perspective although we draw upon the ICS psychological model. Our problem space is comprised of six dimensions. The first two dimensions deal with the notion of communication channel: the number and direction of the channels that a particular MSM system supports. The other four dimensions are used to characterize the degree of built-in cognitive sophistication of the system: levels of abstraction, context, fusion/fission, and granularity of concurrency. We then show how these issues made explicit by the MSM framework can be projected in terms of software architecture. In particular, fusion and parallelism are discussed in detail within PAC-Amodeus, our conceptual architectural model, and illustrated through implementation.

Mots-clés : Modalité, interaction multimodale, espace problème, architecture logicielle, fusion, parallélisme.

Keywords: Modality, multimodal interaction, design space, software architecture, data fusion, concurrency.

1. INTRODUCTION

Les avancées techniques des interfaces graphiques, des systèmes de reconnaissance et de synthèse de la langue naturelle, les progrès de la vision par ordinateur, le son et le geste en 3 dimensions ouvrent, par leur intégration, de nouvelles perspectives pour la communication homme-machine [Krueger, 1985]. L'objectif des "interfaces nouvelles" est d'étendre les facultés sensori-motrices des ordinateurs en vue de répondre aux moyens de communication naturels du sujet humain. Cette extension se manifeste sous différentes formes, depuis la création de nouveaux dispositifs physiques d'interaction jusqu'à la définition de systèmes de représentation symbolique des informations échangées au moyen de ces dispositifs. L'éventail des possibilités et la nouveauté de l'entreprise expliquent la variété des termes, tels "multimédia" et "multimodal".

Dans cet article, nous adoptons une perspective ouverte avec la notion de système multi-sensori-moteur (MSM) qui recouvre à la fois le multimédia et le multimodal. Nous proposons pour cela un espace de référence permettant de raisonner sur le comportement des systèmes MSM. Ce référentiel, présenté au chapitre 2, est une extension des réflexions auxquelles nous avons participé lors des colloques IHM'91 et IHMM'92. Comme le montre Nigay dans [Nigay 93a], cet espace peut servir à classer les systèmes actuels et à venir. Il a aussi l'avantage d'identifier des problèmes pertinents pour la construction logicielle des interfaces. Au chapitre 3, nous rappelons les grandes lignes de notre modèle d'architecture, PAC-Amodeus, applicable à la conception logicielle des interfaces utilisateur. Nous étudions ensuite comment ce modèle intègre deux problèmes importants de notre espace de référence, la fusion et le parallélisme, que nous illustrons au moyen d'un exemple concret.

2. L'ESPACE DE RÉFÉRENCE MSM

Comme le montre la figure 1, l'espace de référence MSM comprend 6 dimensions. Les deux premières ont trait à la notion de canal de communication : le nombre et le sens des canaux permis pour un système donné. Ces notions font l'objet du paragraphe 2.1. Les autres dimensions caractérisent le degré de sophistication cognitive des fonctions d'interprétation et de restitution du système : niveaux d'abstraction, contexte, fusion et fission des informations échangées, granularité du parallélisme. Ces dimensions sont présentées au paragraphe 2.2. Au paragraphe 2.3, nous revenons sur la définition d'interfaces multimédia et multimodale et comparons MSM au référentiel issu des colloques IHM'91 et IHMM'92.

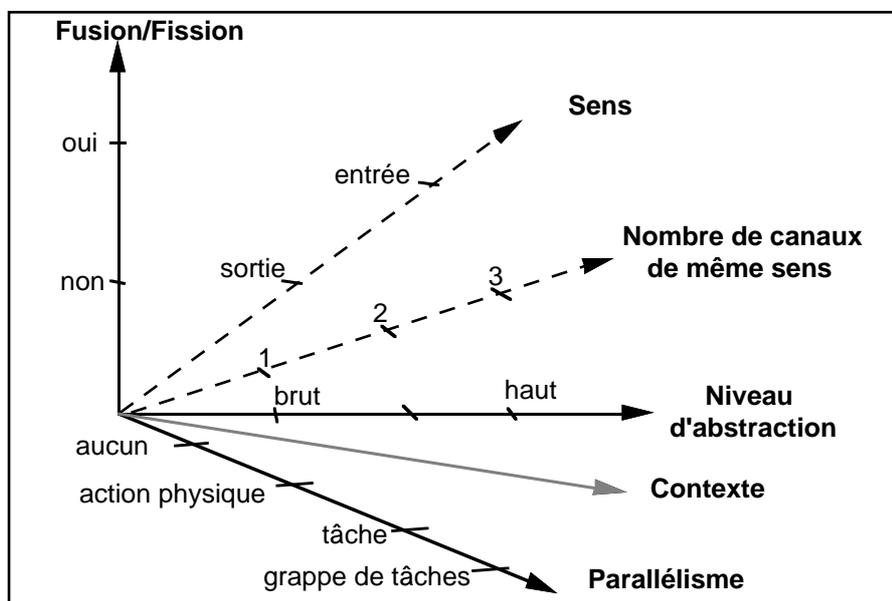


Figure 1 : Le référentiel MSM : un espace de référence pour système multi-sensori-moteur.

2.1. Canal de communication

Un canal de communication d'entrée (inversement, de sortie) regroupe des dispositifs physiques d'entrée ou capteurs (inversement, de sortie ou effecteurs) capables de recevoir (d'émettre) des informations de types donnés sous le contrôle d'une capacité computationnelle ou processus. Inspirée du modèle psychologique de la communication tel ICS [Barnard, 1987], cette définition met l'accent, non pas sur l'aspect liaison, mais sur les intervenants de l'acte de communication : la source et le récipient.

Notre définition s'applique aussi bien aux canaux humains qu'aux canaux digitaux. Par exemple, le canal visuel humain comprend plusieurs capteurs, les rétines, et un sous-système local visuel qui traduit les relations spatio-temporelles des photons en une représentation interne acceptable pour les "sous-systèmes représentationnels" internes [Barnard, 1987]. De même, le processus serveur X-window gère plusieurs dispositifs d'entrée, la souris et le clavier, et traduit les signaux d'interruption en une forme logique, l'événement, compréhensible par les processus internes clients. On trouvera dans [Coutaz, 1993] une description détaillée de la correspondance entre les canaux de communication humains et les sous-systèmes sensori-moteurs d'ICS.

2.2. Fonctions d'interprétation et de restitution

L'information acquise par les canaux d'entrée digitaux est transformée par les processus internes du système. Cette suite d'opérations modélise la fonction d'interprétation. Dans l'autre sens, l'information interne (c.-à-d. l'état interne du système) est transformée jusqu'à ce qu'elle soit acceptable par les canaux de sortie du système. Cette suite d'opérations constitue la fonction de restitution. Interprétation et restitution font intervenir quatre ingrédients en intime relation : niveaux d'abstraction, contexte, fusion et fission, parallélisme.

2.2.1. Niveaux d'abstraction

La notion de niveau d'abstraction exprime le degré de transformations subies par les informations reçues ou émises sur les canaux. Elle couvre également l'éventail des représentations que gère le système depuis les informations brutes (les signaux) jusqu'aux représentations symboliques (le sens). Un canal d'entrée se caractérise par son *pouvoir d'abstraction* depuis les signaux de bas niveau. Un canal de sortie se qualifie par son *pouvoir de matérialisation* depuis les représentations internes. Le niveau d'abstraction doit être analysé pour chacun des canaux de communication. En effet, pour un système donné, certains canaux peuvent être très performants en abstraction ou en matérialisation, d'autres pas.

Par exemple, un acte de parole reçu par le canal digital du microphone peut être enregistré comme un simple signal numérique, comme une suite de mots, ou reconnu comme une phrase significative. Sur le canal audio de sortie, un message oral peut être restitué depuis une représentation du sens (génération de texte), ou bien à partir d'une chaîne de caractères (synthèse vocale), ou encore rejoué à partir d'un pré-enregistrement.

2.2.2. Contexte

La capacité d'un canal à abstraire ou à matérialiser peut dépendre de *variables contextuelles* ou *contexte*. Un contexte comprend un ensemble de paramètres d'état utilisés par les processus internes pour contrôler l'interprétation ou la restitution. Il agit comme un filtre cognitif. Par exemple, dans les systèmes MSM actuels, nous observons deux contextes d'interprétation intéressants : les commandes et les concepts du domaine. Les informations liées aux commandes font l'objet d'une interprétation à un haut niveau d'abstraction tandis que les informations, objets de la tâche, sont laissées inchangées. Par exemple, dans les systèmes de traitement de texte, les saisies au clavier ou à la souris qui correspondent à des commandes sont "comprises" par le système tandis que ces mêmes saisies à destination du contenu des documents ne sont pas interprétées.

Ainsi, le pouvoir d'abstraction ou de matérialisation d'un système est propre à chaque canal, mais pour un canal donné, ce pouvoir dépend du contexte.

2.2.3. Fusion et fission d'information

La fusion est la combinaison de plusieurs unités d'information pour former de nouvelles unités. La fission correspond au processus inverse. L'une et l'autre traduisent deux phénomènes importants du processus d'interprétation et de restitution.

En interprétation, la fusion peut intervenir à un bas niveau d'abstraction entre des informations pouvant provenir de plusieurs canaux d'entrée. Elle peut aussi s'effectuer à un plus haut niveau d'abstraction pour des informations issues de différents contextes. Par exemple, le paradigme du "mets ça ici" nécessite la fusion de l'acte de parole reçu sur le canal du microphone avec les signaux reçus sur le canal lié à la souris. Plus usuelle est la fusion de clics souris répartis sur une palette et une zone de dessin pour dessiner une figure géométrique. Ces informations issues du même canal transitent selon des contextes différents pour être regroupés à un haut niveau d'abstraction [Nigay 91].

En restitution, la fusion intervient elle aussi à de multiples niveaux d'abstraction. Au niveau le plus haut, elle intervient dans l'adaptation des informations du noyau fonctionnel aux besoins conceptuels de l'interface [Coutaz 91]. Au niveau le plus bas, elle se manifeste par exemple sous forme d'incrustations (vidéo et graphique).

La fission en interprétation traduit le besoin de décomposer une information issue d'un canal ou d'un contexte pour franchir un niveau d'abstraction. Par exemple, l'acte de parole "dessine un cercle dans une nouvelle fenêtre" fait référence à deux domaines de discours : les figures géométriques ("dessine un cercle") et l'interface homme-machine ("nouvelle fenêtre"). Cette phrase dont le sens a pu être identifié le long d'un canal unique doit être décomposée en deux primitives de haut niveau du système : "créer fenêtre" et "créer cercle" dans la nouvelle fenêtre.

La fission en restitution revêt plusieurs formes. La plus courante, est la représentation multiple d'un même concept sur un canal donné (par exemple le concept de température sous forme d'un thermomètre et d'un réel) [Coutaz 88]. La représentation multiple peut aussi s'effectuer en coréférence sur des canaux distincts tel le message oral "attention à cette température" accompagné de l'affichage en rouge du thermomètre à surveiller.

2.2.4. Parallélisme

Nous limitons la discussion au parallélisme perceptible à l'interface. Le parallélisme interne du système est l'affaire des modèles d'architecture. Au niveau de l'interface, le parallélisme se manifeste selon trois niveaux de granularité : action physique, tâche élémentaire, espace de tâches.

Au niveau physique, le parallélisme en entrée autorise l'utilisateur à agir simultanément sur plusieurs dispositifs d'entrée. Si ces dispositifs sont répartis sur différents canaux, alors plusieurs canaux d'entrée peuvent être sollicités en parallèle comme dans l'exemple du "mets ça ici". En sortie, parallélisme signifie que plusieurs primitives de sortie peuvent être produites simultanément sur un même canal (comme en animation graphique) ou sur plusieurs canaux (comme pour l'exemple "attention à cette température"). Notons que la fusion et la fission à bas niveau d'abstraction dépendent de l'existence du parallélisme au niveau physique.

Au niveau tâche élémentaire, le parallélisme en entrée autorise l'utilisateur à construire plusieurs commandes de manière concurrente : en parallélisme vrai si le parallélisme est géré au niveau physique ou, en l'absence de parallélisme physique, de manière entrelacée (dialogue à plusieurs fils [Coutaz 88]). En sortie, une tâche élémentaire recouvre l'ensemble des primitives de sortie pour exprimer un changement d'état du système.

Au niveau espace de tâches, le parallélisme exprime les possibilités d'entrelacement entre des grappes de tâches. Par exemple, pour un robot mobile de surveillance, les tâches de l'utilisateur peuvent être structurées en trois espaces : description de la géographie, spécification de missions, et contrôle d'exécution de mission. Les relations entre ces trois espaces peuvent avoir une incidence sur l'utilisation temporelle des canaux de communication et donc sur les fonctions d'interprétation et de restitution du système.

2.3. Interfaces multimédia et multimodale

Les interfaces multimédia et multimodale requièrent l'une et l'autre au moins deux canaux de communication, soit d'entrée, soit de sortie. En ce sens, ces interfaces sont toutes deux multi-sensori-motrices. Une interface multimodale se caractérise par la capacité supplémentaire de disposer d'au moins deux canaux (soit d'entrée, soit de sortie) qui, fonctionnant de manière combinée ou non, sont doués d'un haut pouvoir (soit d'abstraction, soit de matérialisation). Plus précisément, dans le référentiel MSM, une interface multimédia se caractérise par : au moins 2 canaux pour une direction donnée, les autres paramètres étant laissés libres. Une interface multimodale se définit par : au moins 2 canaux pour une direction donnée et un haut niveau d'abstraction pour au moins 2 de ces canaux, les autres paramètres étant laissés libres.

La figure 2, qui représente le référentiel développé à IHM'91 et IHMM'92, s'obtient par projection de l'espace MSM en fixant la direction et le nombre des canaux étudiés. Toutefois, on observe que ce premier référentiel ne considérait pas le phénomène de fission ni la notion de contexte d'interprétation/de restitution. Le parallélisme ne faisait pas de distinction explicite entre les niveaux de granularité (action physique, commande, etc.). En particulier, les interfaces de type "alterné" et "synergique" supposaient la fusion séquentielle ou parallèle des informations au niveau des actions physiques. Inversement, les interfaces de type "exclusif" et "concurrent" sous-entendaient l'absence de fusion au niveau bas seulement ainsi que la construction (séquentielle ou parallèle) de commandes.

		USAGE DES MODALITES	
		Séquentiel	Parallèle
FUSION	Combiné	ALTERNE	SYNERGIQUE
	Indépendant	EXCLUSIF	CONCURRENT
		Sens	Sens
		Pas de Sens	Pas de Sens
NIVEAUX D'ABSTRACTION			

Figure 2 : Premier référentiel d'analyse des interfaces multimodales.

Avec le référentiel MSM, nous pensons avoir contribué à l'amélioration de l'espace problème pour l'étude des nouvelles interfaces. Nous allons illustrer son application à la construction de ces interfaces à partir de notre modèle conceptuel d'architecture PAC-Amodeus.

3. PAC-AMODEUS, MODELE D'ARCHITECTURE DE REFERENCE

La figure 3 présente les composants logiciels de PAC-Amodeus. Une description détaillée de ce modèle peut être consultée dans [Nigay 91a]. Nous en rappelons les éléments essentiels. Dans ses grandes lignes, PAC-Amodeus reprend les composants du modèle ARCH [Arch 91] dont il affine le contrôleur de dialogue en termes d'agents PAC [Coutaz 87].

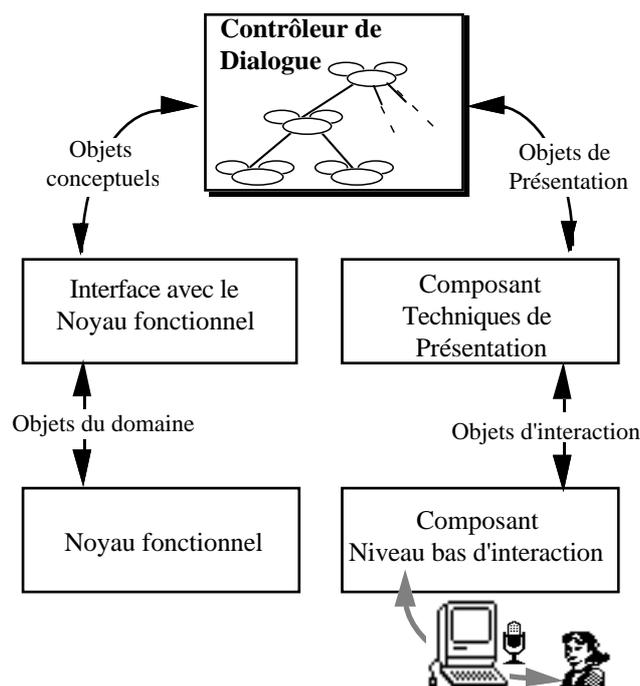


Figure 3 : Composants du modèle d'architecture PAC-AMODEUS.

Le *noyau fonctionnel* réalise les concepts du domaine.

L'*interface avec le noyau fonctionnel* sert d'adaptateur entre le monde des objets du domaine et les objets conceptuels exportés vers l'utilisateur. Il est conçu pour absorber les changements entre ses voisins directs : le *noyau fonctionnel* et le *contrôleur de dialogue*.

Le *contrôleur de dialogue*, clé de voûte du système interactif, prend en charge l'enchaînement des tâches, gère chaque fil de dialogue au moyen d'une hiérarchie d'agents PAC, et assure la correspondance entre objets conceptuels et objets de présentation. On trouvera dans [Nigay 92a], un ensemble de règles heuristiques permettant d'identifier ces agents en fonction de l'interface à gérer. Chaque niveau dans la hiérarchie d'agents PAC correspond à des niveaux d'intervention dans les transformations successives de l'information (processus de fusion et de fission).

Le composant *techniques de présentation* définit les règles de correspondance entre les objets de présentation et les objets d'interaction. Par exemple, la classe de présentation "choix multiple" est concrétisée sous forme d'un "menu de boutons radio". Ce composant constitue le plus haut niveau d'abstraction où intervient la notion de modalité.

Le composant *bas niveau d'interaction* désigne la plate-forme d'accueil logicielle et matérielle. Ce niveau regroupe les services d'acquisition, d'estampille et de répartition des événements mais aussi les objets d'interaction des boîtes à outils et les systèmes de reconnaissance spécialisés. Par exemple, un message vocal est transmis au composant de présentation sous forme d'un événement estampillé dont le contenu est une chaîne de caractères. Certains événements peuvent ne pas être transmis au niveau supérieur. Par exemple, le déplacement d'une fenêtre est résolu par le gestionnaire de fenêtre qui appartient au niveau bas d'interaction.

Nous avons appliqué le modèle PAC-Amodeus à la réalisation de trois systèmes multimodaux :

- VoicePaint [Gourdol 91], un éditeur de dessin sur Macintosh, qui offre à l'utilisateur la possibilité de dessiner avec la souris tout en modifiant le contexte graphique par commandes vocales. (Par exemple, l'utilisateur, tout en traçant un trait, modifie l'épaisseur du pinceau en disant <plus épais>.)
- NoteBook [Nigay 91b], un bloc-notes électronique, développé sur machine NeXT [Webster 89] avec le système de reconnaissance de la parole continue Sphynx [Lunati 91]. Il permet des énoncés de commandes tels que la phrase dite <insère une page ici> tout en spécifiant le lieu d'insertion avec la souris.
- MATIS (Multimodal Airline Travel Information System) [Nigay 93b], un système d'information sur les voyages aériens, développé sur machine NeXT avec le système Sphynx. Ce système autorise des énoncés de commandes tels que la phrase dite <I would like a Usair flight from this city to this city> conjuguée à deux sélections-souris pour spécifier les villes de départ et d'arrivée.

Ces trois systèmes sont de type "synergique" (voir figure 2). Au travers de ces trois expériences, nous avons montré la validité du modèle PAC-Amodeus pour la réalisation des interfaces multimodales. Nous avons cependant travaillé uniquement sur le traitement des informations en entrée. Le problème de la fusion des informations en sortie par exemple n'a pas été abordé pour la réalisation de ces systèmes.

Après avoir présenté le modèle de référence, nous montrons maintenant où le parallélisme des traitements et la fusion des informations en entrée, deux caractéristiques principales des systèmes multimodaux, interviennent dans une telle architecture. VoicePaint, NoteBook et MATIS servent d'exemples pour étayer cette discussion.

4. PARALLELISME ET FUSION DES DONNEES

4.1. Le parallélisme

Des traitements parallèles sont possibles dans chacun des composants de PAC-Amodeus. Cependant, il n'est pas nécessaire d'avoir des traitements en parallèle (comme par exemple plusieurs processus) pour obtenir du parallélisme perceptible à l'interface. Ainsi dans MATIS, deux processus gèrent l'acquisition des événements en entrée mais un seul processus gère l'analyse syntaxique de tous les événements. MATIS supporte cependant du parallélisme au niveau de la tâche (voir axe Parallélisme de l'espace MSM) en autorisant la spécification de plusieurs commandes indépendantes en parallèle.

Dans le composant "niveau bas d'interaction" de PAC-Amodeus, les événements peuvent être acquis en parallèle à raison d'un traitant d'événements par dispositif d'entrée/sortie. Dans le système MATIS, un processus traite les événements parole (reconnaissance de la phrase dite) tandis qu'un autre processus traite les événements en provenance du clavier et de la souris. MATIS dispose donc de deux canaux digitaux d'entrée.

Le niveau "techniques de présentation", qui reçoit les événements issus des objets d'interaction, est le siège d'activités d'abstraction à raison d'une activité par modalité. Comme exemples d'activités d'abstraction, un clic souris en provenance du niveau bas est transformé en un message de sélection, ou bien une chaîne de caractères issue du système de reconnaissance de la parole est représentée sous forme d'arbre syntaxique. C'est le cas du système MATIS, où c'est le lieu de l'analyse syntaxique des événements parole. Ces événements sont d'abord traités au

niveau du composant "niveau bas d'interaction" et transmis sous forme d'une chaîne de caractères au niveau "techniques de présentation".

Enfin, le contrôleur de dialogue organisé en agents est également le siège d'activités parallèles. Chaque agent traite les informations reçues en parallèle. C'est le niveau techniques de présentation qui distribue les informations aux agents, ceux-ci ayant exprimé leurs intérêts pour certaines informations.

4.2. La fusion des données

Un système transforme les objets d'interaction en objets de présentation, puis en objets conceptuels, enfin en objets du domaine et vice versa. Lors de chaque phase de transformation, des objets peuvent être combinés ; cette fusion se fait donc à différents niveaux. Nous avons identifié trois catégories de fusion : lexicale, syntaxique et sémantique.

La *fusion lexicale*, réalisée dans le composant de bas niveau, traduit une synchronisation entre des informations de bas niveau. Par exemple, dans le Macintosh, l'enfoncement conjoint de la touche "shift" et d'un clic souris est fusionné par le système en un seul événement.

Les *fusions syntaxique et sémantique* sont à la charge du contrôleur de dialogue.

La *fusion syntaxique* consiste à combiner des informations pour obtenir une commande complète comme par exemple la commande "insérer une note" du système NoteBook.

La *fusion sémantique* combine des commandes pour aboutir à une nouvelle fonction. Par exemple, dans VoicePaint, deux commandes (dessiner une droite et changer l'épaisseur) sont combinées en une seule pour obtenir une droite à plusieurs épaisseurs.

Les fusions syntaxique et sémantique s'appuient sur une représentation uniforme : le creuset. Le creuset sert à représenter les informations à tous les niveaux d'abstraction. Ainsi de la fusion de deux creusets, résulte un creuset. Comme le montre la figure 4, un *creuset* est un objet structuré à deux dimensions. L'un des axes définit les composants structurels de l'information (par exemple les éléments d'une commande). L'axe des temps identifie l'ordre de remplissage des composants structurels. La présence de plusieurs valeurs pour un même composant peut révéler une redondance ou des incohérences. Le creuset doit être rempli à l'intérieur d'une fenêtre temporelle, sinon il est abandonné.

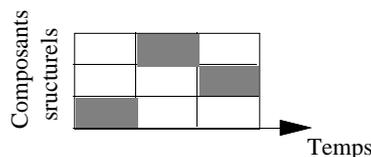


Figure 4 : L'objet creuset comme structure de représentation commune pour la fusion.

L'organisation non séquentielle et hiérarchique du contrôleur de dialogue est un support à la réalisation des fusions syntaxique et sémantique. Les données sont combinées en parallèle et de façon incrémentale à chaque niveau dans la hiérarchie d'agents. L'exemple d'un éditeur graphique supportant la commande multimodale "Mets cà là" est détaillé dans [Nigay 92b]. Cet exemple montre l'intérêt de la hiérarchie d'agents PAC.

La partie Contrôle de chaque agent contient un moteur de fusion. De plus un agent a la possibilité de rajouter des données à combiner issues de son propre état interne. Ainsi le mécanisme de fusion se décompose en un ensemble de fusions effectuées par chaque agent. La fusion est effectuée sur la base de trois critères : la structure des objets à combiner, le temps et le contexte courant. Le moteur de fusion ne se met jamais en attente d'information complémentaire. Aussi certaines fusions incorrectes sont parfois effectuées et il est donc nécessaire de pouvoir dé-fusionner. Le formalisme unique pour représenter l'information facilite cette activité de "dé-fusion". Le moteur de fusion met en place trois types de fusion, utilisant des critères de fusion différents :

- La Micro Fusion se base sur le temps et combine des informations, structurellement complémentaires et ayant des intervalles de temps qui s'entrelacent (parallélisme ou pseudo-parallélisme).
- La Macro Fusion se base sur le temps aussi, et combine des informations structurellement complémentaires qui appartiennent à une fenêtre temporelle (proximité temporelle).
- La Fusion Contextuelle utilise uniquement le contexte courant et fusionne des informations structurellement complémentaires. Dans le système MATIS par exemple, le contexte courant correspond à la requête en cours de construction.

Il est à noter que la représentation sous forme de creuset est indépendante du domaine de l'application. Il en est de même pour les critères de fusion. Le moteur de fusion, qui effectue les trois types de fusion précédemment cités, est donc indépendant du domaine de l'application. Il combine des informations sans en connaître leurs sémantiques.

Pour illustrer les différents types de fusion, nous avons choisi un exemple d'énoncés accepté par MATIS. Les figures 5-1 et 5-2 présentent les étapes de fusion et de dé-fusion de cet exemple. Ce processus de fusion syntaxique se fait au niveau du contrôleur de dialogue et est déclenché par des agents distincts. L'utilisateur sélectionne avec la souris Usair (liste dans une palette) et prononce ensuite la phrase <Flights from Boston serving a meal>. Ces deux informations sont fusionnées puisqu'elles appartiennent à une même fenêtre temporelle (Macro Fusion). L'utilisateur prononce ensuite la phrase <Two flights from > tout en sélectionnant la ville <Pittsburgh>. La sélection est traitée en première (son temps de traitement étant inférieur à la celui de la parole). La sélection <Pittsburgh> est donc fusionnée avec la requête courante (Fusion Contextuelle). Lorsque l'événement parole <Two Flights from> est disponible, il est alors nécessaire de dé-fusionner la sélection <Pittsburgh>, qui peut alors être fusionnée avec l'événement Parole, émis en parallèle de la sélection (Micro Fusion).

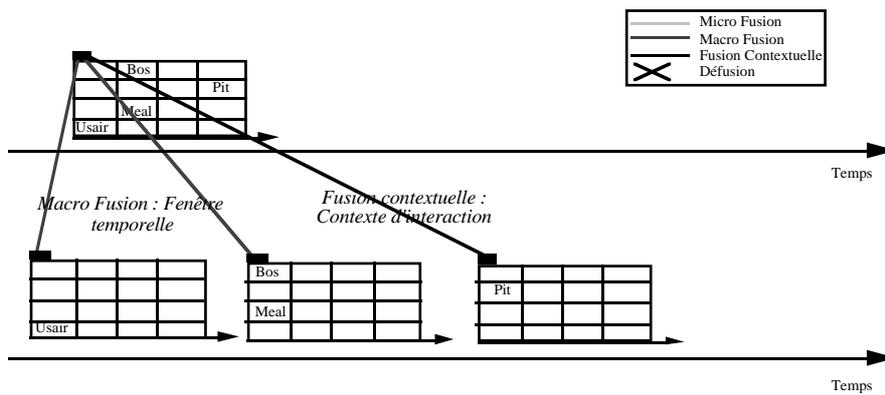


Figure 5-1 : 1ère étape : Macro Fusion puis Fusion Contextuelle.

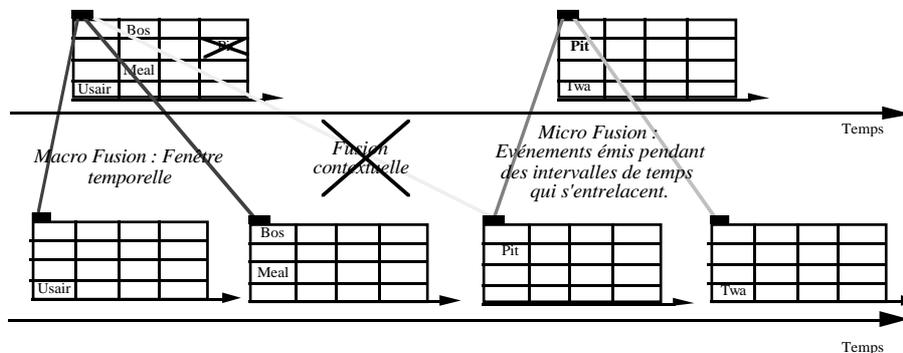


Figure 5-2 : 2ème étape : Dé-fusion et Micro Fusion.

Dans notre exemple, des agents distincts reçoivent les informations à fusionner à chaque étape. Ces agents fournissent avant même que la commande soit complète des retours d'informations spécifiques au niveau d'abstraction. Par exemple l'agent qui reçoit la sélection Pittsburgh produit un retour d'information immédiat traduisant la sélection, en grisant la case sélectionnée. Comme le montrent les figures 5, chaque information est rangée dans les cases structurales idoines d'un creuset. Le creuset résultant de la Macro Fusion contextuelle et de la Fusion Contextuelle, de la figure 5-1, est reçu par un agent, père des agents ayant reçu les messages contenant l'information originale. Lors de la deuxième étape, figure 5-2, la Fusion Contextuelle est dé-faite et le creuset contenant l'information Pittsburgh est fusionné avec le nouveau creuset.

Nous avons montré que notre modèle d'architecture peut supporter des traitements en parallèle à tous les niveaux et fournit un cadre de réalisation pour la fusion de données. Ce sont deux conditions nécessaires à la réalisation de systèmes synergiques. De plus notre modèle offre un support intéressant pour fournir un retour d'information immédiat. En effet un retour d'information peut être généré par chaque agent dans la hiérarchie. Ainsi il est possible de définir des retours

d'information partiels et spécifiques au niveau d'abstraction, avant même que le mécanisme de fusion soit fini, pour une commande donnée. Ce modèle offre aussi une grande flexibilité. En effet il est possible, par exemple, de changer ou de rajouter une modalité d'interaction en ne modifiant que les composants Niveau Bas d'interaction et Techniques de Présentation, le Contrôleur de Dialogue étant indépendant des modalités.

5. CONCLUSION

Notre contribution à l'étude de l'interaction multimodale est multiple :

- nous avons affiné l'espace de référence ébauché lors des journées IHM'91, en proposant l'espace de référence MSM,
- nous avons intégré dans un modèle d'architecture les phénomènes de fusion et de parallélisme, deux caractéristiques essentielles des systèmes synergiques,
- nous proposons un mécanisme de fusion indépendant du domaine de l'application du système,
- nous avons mis en pratique notre modèle à la mise en œuvre de plusieurs systèmes multimodaux.

Nous avons cependant à approfondir le mécanisme de fusion et ses critères. Notre travail futur s'oriente sur l'intégration d'un réel modèle du dialogue au sein de l'architecture logicielle et l'exploitation du contexte de l'interaction comme critère pour le mécanisme de fusion. Notre support pour ce travail sera le système MATIS. De plus une étude doit être conduite sur le traitement des informations en sortie, et en particulier sur les processus de fusion et de fission des informations en sortie.

6. REMERCIEMENTS

Ce travail a été partiellement financé par le projet ESPRIT BR AMODEUS 3066 et 7040, et par le pôle IHM-Multimodale du PRC Communication Homme-Machine. Nous tenons à remercier Jean Caelen (ICP, Grenoble) pour des échanges fructueux et A. Rudnicky (CMU, Pittsburgh, USA) pour nous avoir permis de réaliser NoteBook et MATIS en utilisant Sphinx.

7. REFERENCES

[Arch 91]

The UIMS Workshop Tool Developers A Metamodel for the Runtime Architecture of an Interactive System. SIGCHI Bulletin, 24, 1 (Jan. 1992), pp. 32-37.

[Barnard 87]

Barnard P., "Cognitive Resources and the Learning of Computer Dialogs", in *Interfacing Thought, Cognitive aspects of Human Computer Interaction*, J.M. Carroll Ed., MIT Press Publ., pp. 112-158.

[Coutaz 87]"

Coutaz J., "PAC: an Implementation Model for Dialog Design", In Proc. Interact'87, (Stuttgart, Sept. 1987), H-J. Bullinger, B. Shackel ed., North Holland, pp. 431-436.

[Coutaz 90]

Coutaz J., "Interface Homme-Ordinateur: Conception et Réalisation"; Dunod Informatique, 1990.

[Coutaz 91]

Coutaz J., Balbo S., "Applications: A Dimension Space for User Interface Management Systems". In Proc. CHI'91, ACM Publ., May, 1991, pp. 27-32.

[Coutaz 93]

Coutaz J., Nigay L., Salber D., "The MSM Framework for Multi-Sensory-Motor Systems", ESPRIT BR 7040 Amodeus Document, System Modelling/WP4, January, 1993.

[Krueger 85]

Krueger M.W., Gionfrido T. and Hinrichsen K. Videoplacement, "An Artificial Reality". In Proc. CHI'85 Human Factors on Computing Systems (April 1985), ACM Press, pp. 35-40.

[Lunati 91]

Lunati J.M. and Rudnicky A.I., "Spoken Language Interfaces: The OM system". In Proc. CHI'91 Human Factors on Computing Systems (New Orleans, April 27-May 2 1991), ACM Press, pp. 453-454.

[Nigay 91a]

Nigay L., Coutaz J., "Building user interfaces : Organizing software agents"; Conference ESPRIT'91, Bruxelles, novembre 1991.

[Nigay 91b]

Nigay L., "An example of Multimodal Interactive System: a Voice Enabled Note Book", rapport, technique, October 1991.

[Nigay 92a]

Nigay L. and Coutaz J., "PAC-Expert: Towards an Automatic Generation of Dialogue Controllers", Rapport Technique IMAG RT-83, mai, 1992. Paru également comme deliverable D18, projet ESPRIT BR AMODEUS 3066.

[Nigay 92b]

Nigay L. et Coutaz J. Interfaces multimodales et architecture : Fusion et parallélisme. Dans les actes de IHM'92 Quatrièmes journées sur l'ingénierie des interfaces homme-machine, 30 Nov., 1 et 2 Déc. 1992, Paris, pp. 29-36.

[Nigay 93a]

Nigay L., Coutaz J., "A design space for multimodal interfaces: concurrent processing and data fusion", Interchi'93, Amsterdam, May, 1993.

[Nigay 93b]

Nigay L., "MATIS: a Multimodal Airline Travel Information System", rapport technique, en cours de rédaction, Janvier 1993.

[Webster 89]

Webster B.F. : "The NeXT Book". Addison Wesley, New York, 1989.