Présentation Multimodale d'Information

Michaël Krus

Septembre 1995

Rapport de stage pour
le *DEA d'Informatique* de l'Université Paris Sud

effectué au LIMSI/CNRS
sous la direction de F. Néel et D. Teil

Table des matières

Table des figures

Liste des tableaux

Remerciements

Je souhaiterais remercier d'abord F. Néel et D. Teil pour la qualité de leur encadrement et leur patience lors de nos longues discussions. Ensuite, Y. Bellik pour la pertinence de ses conseils. Enfin, les permanents et thésards du groupe IMM du LIMSI, pour la saine atmosphère de bonne humeur qu'ils entretiennent.

1 Introduction

Pour communiquer, les humains utilisent naturellement un nombre important de modalités de communication incluant le langage oral, les gestes, les dessins. Ils exploitent des sens tels que la vision, l'ouïe et le toucher. Certains media et modes de communication véhiculent des informations plus efficacement que d'autres, en fonction de la tâche, de l'utilisateur ou du contexte. Mais, si les humains ont des facilités naturelles pour interpréter et produire des énoncés multimodaux, ce n'est pas le cas des ordinateurs.

Par définition, une information doit informer et servir de base à une prise de décision dans un processus. Nous assistons en ce moment à une explosion de la quantité d'information disponible, et les techniques de l'information se doivent de filtrer et raffiner ces données en une forme qui soit expressive, qui reflète notre façon de penser, d'apprendre, de résoudre des problèmes, de prendre des décisions. Il est donc nécessaire d'approfondir notre connaissance des caractéristiques de l'information, de leurs liens avec les modalités de communication et de leur adaptation à la tâche, à l'utilisateur et à l'environnement.

Nous nous intéressons dans cette étude uniquement à la synthèse de présentations multimodales d'information, en complément des études de [Bellik, 1995] qui abordent en détail les problèmes liés à la multimodalité en entrée. Le but est donc de produire un énoncé multimodal qui doit être le reflet des informations présentées et des relations entre les éléments de cette information. Cet énoncé devra être adapté à la tâche, à l'utilisateur et à l'environnement physique du dialogue. La réalisation des interfaces multimodales actuelles étant essentiellement intuitive, nous proposons une méthode pour l'analyse et la synthèse d'une présentation multimodale.

Nous faisons dans la section suivante un survol de l'ensemble des problèmes intervenant dans cette synthèse et nous présentons quelques systèmes existants. La section 3 décrit l'architecture du système que nous avons conçu. La section 4 revient en détail sur la description l'information qui doit être présentée, et la section 5 contient une analyse des modalités qui peuvent être utilisées. Ensuite, la section 6 décrit le processus de production d'un énoncé multimodal. Les problèmes liés à la représentation de toutes les connaissances acquises et de leur mise en oeuvre sont présentés dans la section 7, ainsi que la concrétisation logicielle du système que nous avons réalisé. Nous conclurons par quelques remarques et offrirons des perspectives pour la poursuite de ces travaux.

Cette étude a eu lieu, de Mai à Septembre 1995, au LIMSI (Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur) qui est une unité propre du CNRS située sur le campus de l'Université Paris Sud. Elle a été dirigée par Françoise Néel et Daniel Teil, au sein du groupe Interactions et MultiModalités du département Communication Homme-Machine.

2 Présentation multimodale

Il s'agit de concevoir un *Intelligent MultiModal Presentation System*[1] (IMMPS) au sens défini dans [Roth and Hefley, 1993] : en entrée, un IMMPS reçoit une certaine quantité d'information à communiquer et un ensemble de buts de communication (c'est-à-dire les buts et les intentions signifiés par l'énoncé de ces informations). Il possède des alternatives de modalités et de techniques de présentation. Sa fonction est de concevoir une présentation qui exprime les informations de manière à refléter les buts de communication et à aider les utilisateurs dans l'accomplissement de leur tâche.

[Arens et al., 1993a] désigne par *structure de discours* les données fournies en entrée au IMMPS, et par *structure de présentation* l'énoncé résultant. La structure du discours contient les informations qui doivent être présentées et le but recherché par cette présentation. La structure de présentation décrit les modalités utilisées pour la présentation et leur coordination spatiale, temporelle et sémantique.

2.1 Un art ou une science ?

La présentation multimodale devient une science si elle peut être formalisée par un ensemble de règles. Etant donné un type d'information (voir section 4.2), quelle est la meilleure forme de présentation à utiliser pour communiquer cette information à l'utilisateur ? Il est bien sûr possible d'énumérer toutes les modalités capables de représenter les informations de ce type. Il faudra ensuite utiliser des règles pour choisir la meilleure. Il est important de créer des règles très générales. Ainsi, pour décrire la position d'un bateau (exemple tiré de [Arens et al., 1993b]), on pourrait énoncer la règle suivante :

> *La position d'un bateau est indiquée sur une carte.*

Néanmoins, cette règle n'est ni flexible, ni portable, ni générale. En effet, elle associe directement une information à une modalité. Il serait plus général d'associer des caractéristiques de l'information à des caractéristiques des modalités.

Ainsi, cette règle peut être généralisée en :

> *Les données ayant une connotation spécifique doivent être présentées sur des modalités ayant la même connotation.*

Toutefois, les images, les cartes et le texte pouvant tous véhiculer des connotations spatiales, nous devons avoir d'autres règles qui permettent de les départager. Mais cette sélection ne peut se faire uniquement en fonction de l'information (puisqu'elle peut être représentée de manière presque équivalente dans les trois modalités). Il faut aussi prendre en compte le contexte. Celui-ci est composé du but recherché par la production de l'énoncé, des caractéristiques de l'utilisateur et de l'environnement, de l'historique de l'interaction, et de considérations ergonomiques et pragmatiques.

Cette réflexion nous a permis d'identifier un ensemble de critères qui influencent la synthèse d'une présentation multimodale. Ces critères sont symbolisés dans la figure 1.

1. en français, Système Intelligent de Présentation Multimodale.

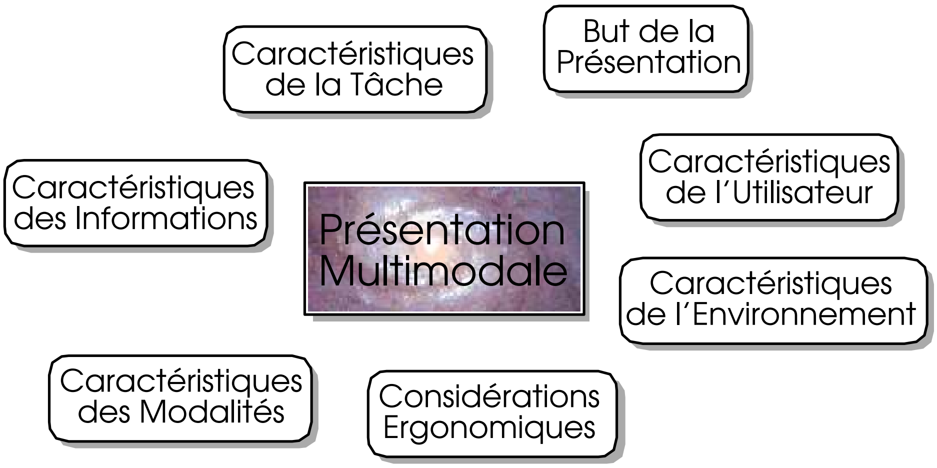


FIG. 1 - *Connaissances influençant la présentation multimodale*

Caractéristiques de l'information : il s'agit d'exprimer les caractéristiques de l'information qui ont une influence sur sa présentation, par exemple, la connotation spatiale.

Caractéristiques de la tâche : celles-ci ont deux volets :

- les aspects *fonctionnels* qui représentent les actions effectuées par la tâche,
- les aspects *intentionnels* qui représentent les buts recherchés par la tâche dans la présentation du résultat.

Nous nous restreignons à la partie de la tâche qui est perceptible par l'utilisateur et qui a donc une influence sur le dialogue.

Caractéristiques de l'utilisateur : celles-ci sont liées d'une part à ses capacités cognitives vis à vis de la tâche et de l'utilisation du système (on ne décrira pas de la même façon le concept de "neige" selon qu'on s'adresse à un esquimau ou à un touareg). D'autre part, il faut aussi considérer ses capacités physiques à percevoir l'énoncé.

Caractéristiques de l'environnement : par exemple, si celui-ci est bruyant, l'utilisateur peut avoir des difficultés à percevoir les messages sonores.

Caractéristiques des modalités : il s'agit à la fois des caractéristiques physiques (mode utilisé, variables perceptibles de la modalité) et sémantiques de l'information énoncée sur cette modalité (nombre de dimensions, aspect dynamique). On pourra aussi prendre en compte l'état des media afin d'éviter la surcharge de l'un d'eux.

Considérations ergonomiques : celles-ci influencent la synthèse de la présentation de plusieurs façons. Par exemple, l'historique de l'interaction peut être prise en compte pour présenter une même information de la même manière au cours du temps. Les caractéristiques de l'information et des modalités peuvent être utilisées pour coordonner les présentations dans le temps et dans l'espace. Ainsi, certaines parties de la présentation pourront faire référence à d'autres parties. Ces considérations sont transversales à toutes les autres et ont une grande influence.

Il faut noter que ces paramètres ne sont pas indépendants. Notamment, les caractéristiques de l'environnement modifient les caractéristiques de l'utilisateur. En effet, le fait que l'environnement soit bruyant rend l'utilisateur sourd.

La quantité des connaissances qui doivent être acquises à cette fin ne doit pas être sous-estimée. Elle relève de plusieurs disciplines.

Les psychologues et les ergonomes travaillent depuis longtemps sur l'étude de l'efficacité d'une présentation (voir [Kolski, 1993]), comme par exemple pour les propriétés de perception des graphes. D'autres travaillent sur l'utilisation de plusieurs modalités – sont-elles toutes utiles et dans quels types de combinaisons. Ces études produisent des données empiriques sur les capacités perceptives de l'homme qui peuvent être utilisées pour choisir la meilleure présentation possible.

L'Intelligence Artificielle (IA) utilise déjà des modèles comme moyen de représenter les caractéristiques d'un utilisateur donné, Notamment pour l'explication des déductions d'un système expert (voir [Blandford and Young, 1993]).

L'IA, pour les systèmes experts, et les techniques de génie logiciel ont beaucoup fait progresser les méthodes d'acquisition et de représentation des connaissances.

Les spécialistes des sciences cognitives travaillent sur des modèles cognitifs de l'utilisateur pour sa perception de l'information et pour sa planification des tâches.

Enfin, dans le domaine des interfaces homme-machine, beaucoup de recherches ont été entreprises pour l'utilisation optimale des techniques de manipulation directe, associées ou non au son. Toutefois peu d'études ont été faites sur l'utilisation simultanée de plusieurs modalités.

Nous présentons dans les sections suivantes les définitions, utilisées pour cette étude, d'un certain nombre de concepts importants. Ensuite, nous reviendrons sur les caractéristiques des structures du discours et de l'énoncé ainsi que sur des considérations ergonomiques. Enfin, nous passerons en revue certains systèmes existants.

2.1.1 Terminologie

Le vocabulaire et les définitions ne sont pas encore établis de manière stable et non-ambiguë dans le domaine de la multimodalité. Afin de clarifier la situation, et de préciser les conventions utilisées dans ce rapport, voici une liste de termes utilisés dans ce rapport et leur définition. Les origines de ces définitions apparaissent dans [Bellik and Teil, 1992, Arens et al., 1993b, Bernsen, 1995a] [2].

Canal d'une modalité [3] : au sens de [Arens et al., 1993b], c'est une dimension indépendante et variable d'un **porteur d'information** donné dans un **substrat** donné.

2. Pour cette section uniquement, les mots dont les définitions sont présentés ici, apparaissent en **gras**.

3. il ne faut pas confondre le *canal d'une modalité* avec un *canal de communication* qui est le lien temporel qui rend possible l'échange d'information entre deux entités communicantes (dans notre cas, un homme et une machine). Un canal de communication transporte un signal, et non pas une sémantique. Dans ce document, "canal" désignera toujours le **canal d'une modalité**.

Le nombre de canaux indique le nombre d'éléments d'information distincts que le porteur peut véhiculer (les *degrés de liberté*). Par exemple, une icône donnée peut véhiculer de l'information au travers de sa taille, sa couleur, mais aussi sa position et son orientation par rapport au **substrat**. Le nombre et la nature des **canaux** dépendent du **porteur d'information** et du **substrat**. L'affectation de valeurs à l'ensemble des **canaux** d'une **modalité** constitue sa **forme de présentation**.

Enoncé : un **énoncé complexe** est une collection, ou composition, d'**énoncés élémentaires**. Un **énoncé élémentaire** est produit par l'invocation d'un **porteur d'information** unique, par exemple, un paragraphe de texte, un graphe, un bip sonore[4]. Un **énoncé** impose la coordination d'un **porteur d'information** dans un **substrat**.

Forme de présentation : c'est la réalisation d'une **modalité** par la détermination de la valeur de ses *canaux* en fonction de l'information et du contexte.

Medium : désigne, dans le contexte du IMMPS, un dispositif physique, un périphérique, qui permet la communication entre l'homme et la machine. Par exemple, un écran, un système de synthèse de parole.

Message : en sortie, c'est un objet actif et dynamique, échangé entre l'application et l'interface, qui véhicule la **structure de discours** devant mener à un **énoncé** multimodal (voir section suivante).

Mode : un mode correspond à un sens physique perceptif de l'utilisateur. Par exemple, le mode visuel, le mode auditif. Chaque mode comprends plusieurs modalités.

Modalité : une **modalité** est définie par la structure des informations échangées, *telle qu'elle est perceptible par l'utilisateur*. Ainsi, la présentation de texte sous forme de caractères ASCII ou sous forme d'image sont en fait la même **modalité**. Par contre, un tableau et un graphe peuvent être considérés comme deux **modalités** différentes car elles peuvent représenter des structures distinctes.

Multimodalité : c'est le fait de coordonner au niveau sémantique, contrairement au multimedia qui ne coordonne qu'au niveau du signal, plusieurs **modalités** pour un **énoncé**.

Porteur d'information : c'est la partie d'un **énoncé** simple qui, pour l'utilisateur, porte la plus grande partie de la sémantique signifiée, comme par exemple, un marqueur sur une carte. Tout **porteur d'information** est une **modalité**.

Substrat : c'est le contexte d'un *énoncé*. Il définit des relations physiques et temporelles, et souvent le contexte sémantique. Une partie de la sémantique d'une information sera souvent définie par la relation du **porteur d'information** avec son **substrat**. Il est possible d'avoir un **substrat** vide mais tout **substrat** non-vide est une **modalité**. Par exemple, une carte est un **substrat** pour des marqueurs indiquant les positions de certains objets.

4. Si rien n'est précisé, le terme “énoncé” désignera un énoncé complexe. Nous parlerons aussi, dans ce cas d'une *présentation multimodale*.

Tâche : c'est l'objectif final recherché par l'utilisateur. Une tâche peut être décomposée en sous-tâches. On distinguera la notion d'application, qui désigne le système informatique (logiciel) mis en oeuvre pour effectuer la tâche. Par exemple, un traitement de texte est une application et la rédaction d'un rapport est une tâche.

Plusieurs de ces définitions sont encore fortement controversées dans la communauté multimodale. Par exemple, est-ce qu'une liste ou un tableau sont des **modalités**? A priori, ils ne semblent véhiculer une sémantique que par les contraintes spatiales qu'ils imposent aux éléments qui les composent. Mais ils peuvent aussi en véhiculer par la forme du cadre et des traits (par exemple en séparant deux zones d'un tableau par un trait plus épais). Ils ont donc leurs propres **canaux** et nous les considérons comme étant des **modalités** à part entière.

Les listes et les tableaux ont la particularité d'être des **substrats** pour des **porteurs d'information**. Il semble que toute **modalité** puisse être un **substrat** pour une autre **modalité** (tant qu'elles sont sur le même **mode**).

Il faudra distinguer le **substrat** du *contexte perceptif* de l'utilisateur qui est lié à sa façon d'interpréter un énoncé. Prenons le cas de la présentation d'un graphe décrivant l'évolution de la valeur d'une action boursière (exemple tiré de [Arens et al., 1993b]). Pour une personne qui suit cette valeur depuis longtemps, l'information pertinente sera la dernière valeur mesurée, le reste du graphe étant simplement le contexte. Par contre, pour une personne qui ne connaît pas cette action, toutes les valeurs sont importantes et les axes du graphe forment le contexte. Enfin, pour un utilisateur qui ne connaît rien au fonctionnement de la bourse, c'est l'ensemble du graphe qui sera pertinent et le contexte sera vide.

La **structure du discours** et la **structure de l'énoncé** sont définies dans les sections suivantes.

2.1.2 Structure du discours

La structure du discours possède un contenu sémantique (les informations) et une intention, c'est-à-dire le but recherché lors de la présentation. Elle correspond ici à ce que les linguistes nomment le "*langage en situation*".

Il est important de mettre en avant les caractéristiques de l'information qui ont une influence sur sa présentation, et pas uniquement les aspects représentationnels liés au domaine de l'application.

La structure du discours contient les éléments suivants :

- *le type sémantique de l'information*: cela permet, au minimum, de savoir si le système cherche à afficher un nombre, un nom. Mais le réel intérêt réside dans la connaissance des propriétés sémantiques de l'information, c'est-a-dire les relations entre ses éléments, son comportement dynamique, ...

- *le but de l'énoncé*: une même information pourra être présentée différemment en fonction du but recherché par la présentation. Par exemple des tableaux permettent d'afficher une liste d'informations afin d'en connaître la valeur exacte. Par contre, pour comparer ces valeurs entre elles, il vaut mieux utiliser une représentation, par exemple, sous forme de camembert.

> Ainsi, une liste de trains pourra être affichée sous forme de tableau s'il s'agit de connaître les horaires précis de départ. Par contre, s'il s'agit de comparer la régularité des départs pendant la journée, on pourra les présenter sous forme de graphe en fonction de l'heure (voir [Maybury and Hovy, 1994]). Dans les deux cas, la liste de trains peut être obtenue de plusieurs façons.

Cette structure devrait être neutre vis à vis des modalités, de leur synchronisation et répartition.

Afin de connaître les valeurs possibles de ces caractéristiques, il faut procéder à une analyse précise de la tâche. Ceci permettra de connaître la nature exacte des informations échangées mais aussi des actes de communication nécessaires à cet échange.

La section 2.2.1 aborde les problèmes liés aux actes de dialogue.

2.1.3 Structure de l'énoncé multimodal

Ainsi, en se fondant sur la structure du dialogue et les contraintes définies ci-dessus, la synthèse de la présentation multimodale devra déterminer les éléments suivants :

- *une liste de modalités* : pour chacune d'elles il faudra déterminer les paramètres suivants :
  - l'énoncé de l'information, c'est à dire la concrétisation d'une partie de l'information dans la modalité donnée.
  - La forme de présentation, c'est-à-dire un ensemble de paramètres propres à chaque canal de chaque modalité (comme la couleur pour les modalités visuelles, ou la prosodie pour la parole).
- *une synchronisation temporelle des modalités* : il s'agit de coordonner les modalités afin de mettre en valeur certaines relations entre les données.
- *une répartition spatiale pour les modalités* : cette répartition doit mettre en valeur certains liens sémantiques et pragmatiques entre les informations (comme le classement, l'équivalence).

Dans la synthèse de cette structure, il est important de coordonner les modalités. Cette coordination porte d'abord sur le choix d'une ou plusieurs modalité(s) pour la communication d'un certain type d'information et la répartition de ces informations à travers les modalités. Mais elle doit aussi utiliser des connaissances sur l'information présentée sur une modalité pour influencer la génération sur une autre modalité. Ainsi, un texte pourra faire référence à une partie d'une image (comme un objet sélectionné). Il n'est donc pas suffisant de savoir quelle information est présente dans chaque modalité, il faut aussi savoir comment elle est reliée aux informations présentées dans les autres modalités.

La coordination elle-même est porteuse d'information : une coordination spatiale et temporelle peut véhiculer des informations sur les relations entre les différentes parties des données présentées. Par exemple, on pourra utiliser une liste pour refléter un ordre parmi les données.

2.2 Qu'est-ce qu'une *bonne* présentation multimodale?

Une information est constituée de valeurs brutes mais aussi de relations entre ces différentes valeurs, relations qui elles-mêmes véhiculent une signification. De plus, le but recherché par la présentation de ces informations a une influence significative sur cet énoncé.

Nous présentons ici quelques considérations qui permettent de mettre en évidence et de représenter ces relations.

2.2.1 De l'influence du dialogue

Dans [Maybury, 1993b], l'auteur décrit la communication comme étant, au plus haut niveau, une façon d'agir sur les croyances, les connaissances, et les buts de la personne à qui l'énoncé est destiné. Cela peut être obtenu par des actes rhétoriques, qui sont des ensembles d'actes de communication indépendants des modalités. A leur tour, ces actes peuvent être réalisés par des objets visuels, sonores ou linguistiques spécifiques à une modalité.

(a) (b) (c)

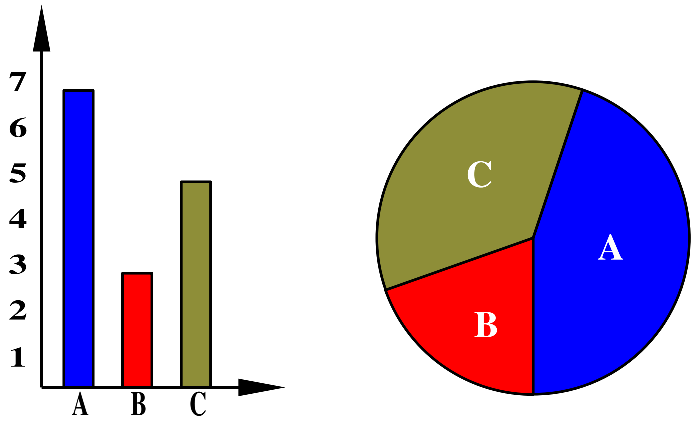
FIG. 2 - *Différentes façons de représenter Paris*

Une même information peut-être présentée différemment et conduire à des interprétations différentes voir contradictoires. Par exemple, la figure 2 utilise différentes possibilités de représenter la ville de Paris, en tant que destination pour un avion. 2(a) utilise une carte, mais il reste à faire le choix de l'échelle et l'identification de l'information pertinente sur cette carte. En effet, les cartes contiennent beaucoup d'informations, comme les noms des villes, des régions, des rivières, leur localisation l'une par rapport à l'autre. Il faut identifier l'information pertinente. 2(b) utilise une image qui, dans la culture commune, est associée à l'information signifiée. Toutefois, cette représentation utilise des connaissances que tous les utilisateurs n'ont pas à priori. Enfin, si 2(c) présente simplement le nom de la ville, il est alors difficile de savoir ce qui est pertinent parmi la somme de connaissances que l'utilisateur peut avoir sur la ville de Paris. Il faut trouver un dosage entre l'expressivité d'une présentation et les capacités d'interprétation et de perception de l'utilisateur.

De même, les relations entre les parties d'une information mises en évidence par une forme de présentation, peuvent être différentes en fonction de cette présentation. Par

Type	**Valeur**
A	7
B	3
C	5

(a)



(b) (c)

FIG. 3 - *Différentes façons de présenter les mêmes données*

exemple, dans la figure 3, les mêmes données sont présentées avec trois modalités différentes. En 3(a), un tableau permet de connaître avec précision les valeurs pour chaque élément. En 3(c), l'utilisation d'un camembert facilite la comparaison des valeurs mais il est plus difficile de les évaluer. Enfin en 3(b), l'histogramme permet à la fois une estimation des valeurs et une comparaison entre elles, mais avec moins d'efficacité que pour chacun des cas 3(a) et 3(c).

Il est important de représenter l'intention signifiée par la présentation. Des exemples de buts dans la présentation d'informations sont :

- Lecture précise de valeurs.
- Consultation d'un ensemble de valeurs (par exemple pour en évaluer la moyenne, l'étendue ou la somme).
- Comptage (par exemple déterminer la fréquence d'apparition de différentes valeurs dans un ensemble).
- Comparaison de valeurs.
- Estimation de la corrélation.
- Localisation (par exemple, pour l'indexation selon un ou plusieurs attributs).
- L'expression de la causalité (effet/cause, action/résultat, problème/solution).

Pour chaque but de présentation, il faudra concevoir une présentation capable de le représenter. [André and Rist, 1993] présente certaines façons d'atteindre ce but, mais il se limite à l'utilisation de modalités textuelles et graphiques.

Cela doit conduire à une étude précise du dialogue mis en oeuvre pendant le déroulement d'une tâche. Cette étude a pour but de déterminer la représentation formelle des actes de dialogue de l'application. Dans [Maybury, 1993b], l'auteur représente ces actes sous forme d'une hiérarchie d'actes rhétoriques (par exemple *identifier*, *définir*, *comparer*) qui sont décomposés en actes illocutoires (*informer*, *demander*, *avertir*) puis en actes locutoires (par l'utilisation du mode indicatif ou impératif dans l'énoncé). Ces actes sont ensuite utilisés dans des opérateurs de planification pour déterminer la présentation multimodale désirée.

La section 3.1 présente plusieurs manières de représenter ces actes et ces opérateurs.

2.2.2 De l'influence du temps

Le temps intervient dans les présentations multimodales dans plusieurs cas :

1. Lorsque l'information doit être présentée dynamiquement, comme par exemple dans le cas d'un film vidéo. Les images qui composent le film sont stockées statiquement mais leur présentation doit se faire dans un ordre précis et à une fréquence donnée.

2. Lorsque l'information est dynamique, c'est-à-dire que sa valeur évolue au cours du temps. Ainsi, la mesure de température ou de l'attitude [5] d'un avion varie continuellement dans le temps. Il faudra mettre en place un protocole entre l'application et le IMMPS pour que ce dernier puisse remettre à jour la présentation de l'information.

3. Si la modalité est dynamique, c'est-à-dire que la forme de l'énoncé varie dans le temps (comme la prosodie d'un message vocal, ou le clignotement d'un élément visible), ou que son signal n'est pas persistant (comme toutes les modalités sonores).

4. Quand des relations entre les informations peuvent être exprimées par la coordination temporelle de modalités. Par exemple, si un message sonore indique "*cliquez sur ce bouton*", il sera possible de faire clignoter le bouton en question au moment où le mot "*ce*" est prononcé.

5. Lorsque l'utilisateur doit pouvoir agir sur la présentation. Une partie de la présentation ne sera réalisée que lorsque l'utilisateur aura effectué une interaction avec cette présentation, par exemple en cliquant sur un bouton. Cela correspond à la définition de *présentation active* donnée dans [André and Rist, 1994]. La réalisation de telles présentations nécessite une coopération entre le module de présentation multimodale (pour la sortie) et celui de fusion multimodale (pour l'entrée) afin que le premier puisse décrire au second le type d'interaction défini pour une présentation.

Un cas particulier, pour le quatrième cas, concernent les *fonctions réflexes*. Ce terme désigne une réaction d'un élément d'une présentation qui est traitée directement dans l'IMMPS sans avoir à attendre la retour de l'application. Dans les interfaces traditionnelles, le fait de cliquer sur un bouton provoque une animation qui simule l'enfoncement de ce bouton. Un tel comportement peut être modélisé par des fonctions réflexes.

La section 5.3.1 aborde les problèmes de synchronisation. La structure présentée dans la section 5.5 permet de représenter ces synchronisations au sein de l'énoncé.

5. valeur des angles de roulis, tangage et lacet.

2.2.3 De l'influence de l'espace

Les relations spatiales sont très efficaces pour véhiculer des relations sémantiques entre les éléments d'un énoncé multimodal. L'ordre des éléments dans une liste ordonnée pourra être reflété spatialement. De même, des éléments d'un même type pourront être regroupés.

Les relations spatiales sont particulières quand elles concernent un porteur d'information et son substrat. Un tableau ne véhicule de l'information que par l'organisation spatiale qu'il impose aux éléments qu'il contient. De même, un marqueur sur une carte véhicule des informations non seulement par sa forme, sa couleur, mais aussi par sa position vis à vis de la carte et de sa topologie.

Il est possible de concevoir un grand nombre de métaphores pour l'ordonnancement spatial des informations. Les interfaces traditionnelles sont à base de fenêtres, de menus, de boutons et d'icônes. Il est toutefois possible d'envisager d'autres représentations, par exemple, en utilisant des techniques de réalité virtuelle ou augmentée.

2.2.4 Heuristiques de sélection

Certaines heuristiques peuvent être utilisées pour effectuer la sélection des modalités :

1. des conseils sur le choix des modalités en fonction du domaine des données, et des caractéristiques des modalités. Par exemple, "*présenter des données qui ont une certaine connotation sur des modalités qui ont les mêmes connotations*";

2. des heuristiques d'adaptation à l'utilisateur et à l'environnement;

3. des heuristiques générales comme par exemple, "*présenter la même information sur deux modalités différentes*" ou "*utiliser du texte et des images fixes pour les informations importantes*";

4. des heuristiques de validation pour vérifier qu'une combinaison de modalités attire correctement l'attention de l'utilisateur et qu'elle ne surcharge pas le travail perceptif. Par exemple, "*ne pas présenter de gros volumes de données sur une modalité non persistante*".

Les études ergonomiques sont une source d'informations non négligeable pour formuler ces heuristiques. Par exemple, [Kolski, 1993] propose une étude ergonomique sur la qualité des interfaces dans différents domaines d'application.

C'est ce type d'heuristiques qui permet l'énonciation des règles décrites dans ce document.

2.3 Systèmes existants

Il existe plusieurs systèmes qui possèdent des approches différentes pour résoudre le problème qui nous concerne. Des aspects théoriques sur les problèmes liés à la présentation multimodale et des connaissances nécessaires à sa réalisation sont présentés dans [Roth and Hefley, 1993, Arens et al., 1993b], et des travaux sur les taxinomies de l'information et des modalités sont présentés dans [Bernsen, 1995a, Maybury and Hovy, 1994, Bellik et al., 1995].

• COMET, décrit dans [Feiner and McKeown, 1993], utilise des caractéristiques rhétoriques pour sélectionner les informations à présenter, puis les assigne à des modalités (textes et images 3D), pour générer des explications sur le mode d'emploi d'une radio. Cette assignation est conduite en parallèle pour toutes les modalités. Ceci permet à une modalité de faire référence à des informations présentées sur une autre modalité. Ceci est accompli en fournissant à tous les générateurs une description complète de l'état courant de la présentation, par l'intermédiaire d'un tableau noir.

• TEXPLAN, dans sa version multimodale présentée dans [Maybury, 1993b], définit des présentations en termes d'actes de dialogue pour l'aide à la navigation en utilisant du texte et des cartes. Le système possède des opérateurs de planification organisés hiérarchiquement. Ceux-ci précisent les contraintes et les pré-conditions pour l'application de l'opérateur, les effets de cette application et la décomposition de l'acte en sous-actes. La génération est incrémentale à partir d'un acte principal. Les contraintes portent sur le type sémantique de l'information et les pré-conditions sur l'état courant de la présentation et du dialogue. Contrairement à COMET, TEXPLAN combine les phases de sélection de l'information et de choix des modalités, ce qui permet de choisir le contenu en fonction des modalités. TEXPLAN est capable de produire des descriptions actives au sens où elles sont composées d'étapes et où il existe un mécanisme pour naviguer dans ces étapes.

• Le système WIP, décrit dans [André et al., 1993], réalise des présentations à base de textes et de graphiques pour la description du mode d'emploi d'une cafetière. Les données connues à priori par le système sont contenues dans une base de faits concernant le domaine d'application et les buts de communication. Ces données sont utilisées par un générateur de plan pour créer un arbre de présentation. Comme TEXPLAN, WIP possède des opérateurs de planification pour mettre en oeuvre les contraintes. Et comme COMET, il est capable de faire des coordinations sémantiques entre les modalités et, par exemple, de faire référence dans un texte à un objet de la représentation graphique. Toutefois, ceci est fait par un coordinateur central car seul une petite portion de la structure de présentation est visible par chaque générateur de modalité. WIP fait un travail important pour identifier le contenu des différentes parties du document, et les relations entre elles. De plus, étant donné qu'il ne présente que des données sur le mode visuel, il est très focalisé sur des considérations d'ordonnancement spatial. Néanmoins, il ne fait pas de coordination temporelle des différentes parties de la présentation.

• L'approche de [Arens et al., 1993a], est plus évoluée en termes de modélisation du discours et en raison de l'importance donnée pour refléter celle-ci dans la présentation. Elle définit une structure de discours et une structure de présentation (comme TEXPLAN). La conception de la présentation est divisée en quatre étapes :

- sélection de modalités,
- redondance de modalités,
- techniques de présentation propres à chaque modalité, et
- adaptation à l'utilisateur.

Toutefois, cette approche possède une définition de la modalité différente de la nôtre. Pour elle, les histogrammes et les camemberts sont deux techniques différentes d'une même modalité. Mais nous estimons que les propriétés expressives de ces techniques sont radicalement différentes, et nous les considérons donc comme étant des modalités distinctes (même si elles ont des caractéristiques communes, comme le fait d'être graphique).

- Le système de Han et Zuckerman, décrit dans [Han and Zukerman, 1995], vise le même but de la conversion d'une structure de discours en une structure de présentation. Mais c'est un système multi-agents qui planifie la présentation pour chaque modalité avec un mécanisme de négociation entre modalités. Ainsi, la résolution des contraintes se fait de façon coopérative et distribuée entre plusieurs agents. Cela facilite l'intégration de nouvelles modalités.

Discussion

Cette approche multi-agents diffère de l'utilisation de la logique sous contraintes (sous différentes formes) de COMET, TEXPLAN et WIP. Si elle peut sembler plus modulaire, elle nécessite un travail important de la part d'un contrôleur principal, au cas où l'un des agents ne pourrait remplir son rôle correctement.

De plus, les agents étant des "*boîtes noires*", il leur est impossible de savoir quelle information est représentée dans une autre modalité. Il en va de même pour WIP où les coopérations entre différentes modalités doivent être faites par un contrôleur principal. Inversement, les modules de COMET ont accès à toute la présentation et ils peuvent agir en fonction des autres modules.

Il faut noter que tous ces systèmes ont pour application des tâches d'explication d'un processus (mode d'emploi, navigation, ...). On peut se demander si cette orientation n'a pas biaisé les efforts de modélisation, et si d'autres domaines d'application n'apporteraient pas de nouveaux problèmes et de nouvelles solutions.

Aucun des systèmes rencontrés ne permet de créer des présentations actives, c'est-à-dire avec laquelle l'utilisateur peut interagir. Ils se contentent de présenter des documents passifs suite à une requête. Dans [André and Rist, 1994], les auteurs proposent une extension de WIP où l'utilisateur pourrait prendre part à la présentation multimodale. Mais cela se rapproche plus de l'hypertexte que d'un réel dialogue coopératif entre la machine et l'homme.

Par ailleurs, très peu d'entre eux utilisent une coopération temporelle entre les modalités afin d'exprimer des relations sémantiques entre les objets présentés sur différentes modalités. Dans [Feiner et al., 1993], les auteurs utilisent des effets d'animation (zoom et translation) dans des modalités graphiques et les coordonne temporellement avec des éléments de modalités textuelles. La coordination temporelle est également abordée dans les travaux effectués sur les données multimedia (voir [Little et al., 1991]).

Les capacités de ces systèmes à s'adapter à l'utilisateur sont, en général limitées au choix d'un utilisateur novice ou expert. [Gargan et al., 1988] décrit un modèle hiérarchique de l'utilisateur permettant de créer des classes d'utilisateurs par spécialisation.

Enfin, à notre connaissance, aucun système existant n'utilise l'environnement de l'utilisateur comme paramètre de la présentation. Néanmoins, on pourra faire référence aux études sur la capture du contexte (voir [Beaudouin-Lafon, 1993] pour cette approche dans les mediaspaces) et sur le traitement du signal pour la détection du bruit.

Le système que nous voulons concevoir se place résolument dans l'optique d'un dialogue homme-machine où les messages peuvent être interrompus par de nouveaux messages plus urgents et où l'historique du dialogue est pris en compte. De plus, les éléments présentés doivent pouvoir être actifs afin que l'utilisateur puisse répondre à cette présentation.

2.4 Cadre de l'étude

Nous nous sommes intéressés, pour cette étude, aux problèmes de la synthèse de présentations multimodales à partir de la structure du dialogue et des contraintes décrites ci-dessus.

La création de la structure du dialogue n'est pas du ressort de ce rapport. Seuls les différentes facettes de cette structure seront évoquées. Il est ainsi vain de vouloir faire une énumération exhaustive des types d'informations et des actes de discours. Seule une étude poussée pour chaque domaine d'application permettra de le faire.

De même, il n'est pas question d'énumérer toutes les modalités possibles mais seulement de fournir une approche suffisamment générale pour les décrire, les différencier et mesurer la qualité de leur coordination multimodale. Nous nous focaliserons sur certaines modalités significatives.

L'évaluation des capacités de l'utilisateur à interpréter un énoncé multimodal est très complexe et relève du domaine de la cognition humaine. Nous sommes ici à la frontière entre la représentation externe présentée et la représentation interne que l'utilisateur se fait de l'information. Les caractéristiques de cette représentation interne n'ont pas été abordées ici.

Il est très important que tous les résultats s'intègrent dans les travaux précédemment effectués au LIMSI sur l'architecture d'un système de dialogue multimodal. Nous nous restreignons pour la présentation à l'usage de la métaphore habituelle de manipulation directe des interfaces traditionnelles, augmentée de l'utilisation du son et de la voix.

Il convient d'identifier des critères permettant de faire le lien entre les contraintes imposées par l'utilisateur, la tâche, l'environnement d'une part, et les possibilités offertes par les modalités disponibles d'autre part.

3 Architecture et modèles

Nous présentons ici l'architecture globale du IMMPS proposée dans le cadre de cette étude. Cette architecture doit se plier à deux contraintes :

- Décrire et contenir l'ensemble des connaissances disponibles ainsi que les mécanismes pour les exploiter.

- Etre conforme à l'architecture définie antérieurement par le LIMSI.

3.1 Représentation des connaissances

Il s'agit de décrire d'une part les connaissances recueillies ci-dessous et les mécanismes permettant de faire des inférences à partir de ces connaissances afin d'obtenir une présentation multimodale. Il n'existe évidement pas d'algorithme connu pour créer des présentations multimodales. De plus, les usages et les techniques évoluent constamment, ce qui rend leur modélisation difficile.

Il conviendra de définir un *langage unifié* pour décrire les messages, les taxinomies de l'information, des modalités, des actes de dialogues, et les règles de sélection. La spécification de ce langage dépasse le cadre de ce rapport. Toutefois, nous abordons dans cette section et la suivante les problèmes liés à la représentation des connaissances et la présentation multimodale. Nous décrivons ensuite l'implémentation réalisée pendant cette étude.

Beaucoup de systèmes existants (voir [Maybury, 1993b, André et al., 1993]) utilisent des logiques modales et des opérateurs de planification pour représenter les connaissances. La figure 4 montre un exemple d'opérateur.

NAME	`Identify-location-linguistically`
HEADER	`Identify(S, H, entity)`
CONSTRAINTS	`Entity?(entity)`
PRECONDITIONS	`WANTS(S, KNOW(H, Location(entity)))`
EFFECTS	`KNOW(H, Location(entity))`
DECOMPOSITION	`Assert(S, H, Location(entity))`

FIG. 4 - *Exemple d'un opérateur de planification (tiré de [Maybury, 1993b])*

La logique modale permet de décrire les actions en terme de modification des croyances de l'utilisateur par le système. C'est une extension du calcul des prédicats du premier ordre qui comporte des prédicats (par exemple, `Entity?`), des fonctions et des actes de communication (par exemple `Identify`, `Assert`). Maybury définit des opérateurs intentionnels comme `WANT`, `BELIEVE` ou `KNOW`.

La technique des opérateurs de planification s'apparente à celle des schémas (*frame* en anglais). Cette technique permet de structurer les informations dans un ensemble de schémas contenant des champs, qui permettent de focaliser la recherche. Chaque schéma contient des informations sur la façon de remplir les champs, sur leur valeur par défaut. Les schémas peuvent, de plus, être organisés dans une hiérarchie qui permet de procéder par raffinements successifs de l'information. Enfin, il est possible d'adjoindre des règles à ces schémas. Ceux-ci forment un contexte explicite qui permet de guider le raisonnement et regroupent les règles naturellement par domaine.

[Han and Zukerman, 1995] présentent une architecture à base d'agents intelligents, dérivée du domaine de l'*Intelligence Artificielle Distribuée*, pour la réalisation de présentations multimodales. L'expertise concernant une modalité ou une technique de présentation est regroupée dans un agent. Mais il faut un agent central qui est responsable de la coordination des autres agents. C'est là que réside la faiblesse de ce système car les agents sont incapables de coopérer entre eux puisqu'ils n'ont aucune connaissance des autres. La coopération doit toujours se faire par l'agent central qui doit donc contenir beaucoup de connaissances sur chacun des agents. Ces connaissances sont donc dupliquées.

Dans [Werkman, 1994], l'auteur présente une nouvelle architecture, toujours à base d'agents, pour la coordination de présentations multimedia. Il développe une technique appelée *Knowledge-Based Negociation* où chaque agent possède une *perspective partageable*, qui est visible par les autres agents. Ainsi les agents peuvent utiliser des connaissances sur les autres agents dans leurs traitements. De plus, la communication entre agents se fait par l'utilisation d'un langage basé sur la théorie des actes de dialogue abordée ci-dessus, dans la section 2.2.1.

Cette dernière approche nous semble donc être la plus à même de modéliser les connaissances et de produire les présentations multimodales souhaitées. Elle devra s'intégrer dans l'architecture décrite dans la section suivante mais n'a toutefois pas été mise en oeuvre.

3.2 Architecture du système multimodal

L'architecture principale du système (voir figure 5) est décrite dans [Bellik et al., 1994, Bellik et al., 1995]. Elle est à rapprocher des modèles *Arch* et *Seeheim*.

Le système est organisé le long d'un pipeline entre l'application et l'utilisateur. En entrée, ce pipeline effectue la fusion des événements produits par l'utilisateur en un message qui sera transmis à l'application. En sortie, il a pour fonction de synthétiser la structure de dialogue à partir des directives de l'application et de la transformer en structure de présentation. Ce système comprend actuellement trois contrôleurs.

Le *contrôleur de dialogue* est au centre du IMMPS. Il gère le contenu sémantique de l'échange qui a lieu entre l'utilisateur et la machine. En ce sens, il se veut totalement indépendant de la tâche et de l'interface utilisée pour mettre en oeuvre ce dialogue. Il dispose, par l'intermédiaire du modèle du dialogue (voir section 3.4.2), de connaissances lui permettant de contrôler la validité du dialogue en cours (et éventuellement d'engager un méta-dialogue pour clarifier les informations échangées), la priorité de communication des messages, ...

Le *contrôleur du flot application* traduit le langage de l'interface en celui de l'application les commandes de l'utilisateur, et inversement, il traduit les informations liées à l'application en langage codé compréhensible par l'IMMPS. Dans les deux cas, il peut éventuellement être chargé d'effectuer des tâches de vérification des messages.

Enfin, en sortie, le *contrôleur d'Interaction* effectue la traduction d'un message conceptuel (la structure du discours) en un énoncé multimodal. En entrée, il effectue la fusion des événements monomodaux pour produire des messages destinés à l'application. Il est décrit plus en détail dans la section suivante.

3.3 Le contrôleur d'interaction

La figure 6 schématise l'architecture du contrôleur d'interaction. Celui-ci contient notamment deux générateurs, l'un pour l'entrée (utilisateur vers machine), l'autre pour la sortie.

Le *Générateur de Messages en Entrée* reçoit des événements depuis les périphériques en entrée. Il est responsable de la datation de ces événements. Ensuite, il effectue leur

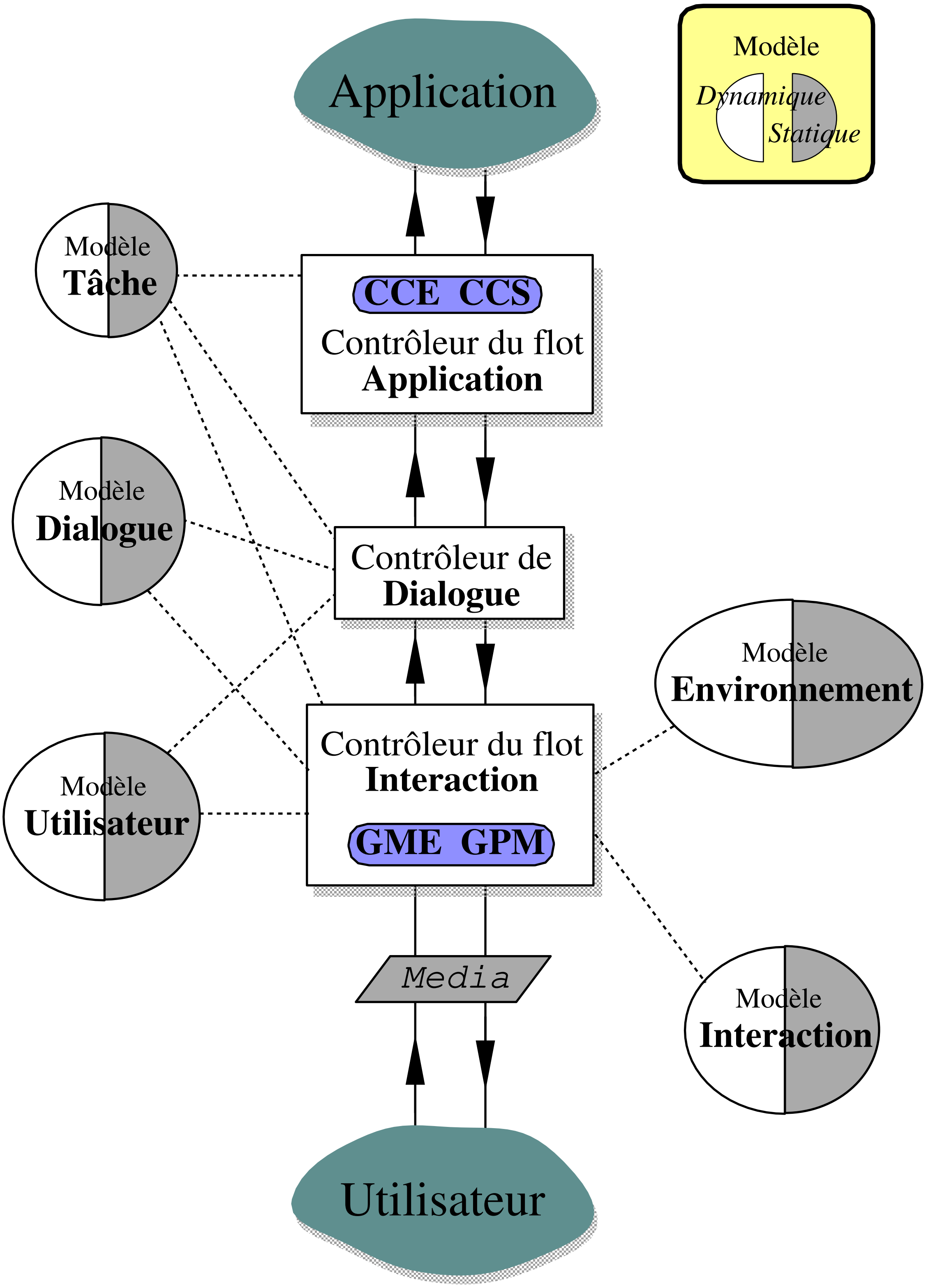


FIG. 5 - *Architecture du système*

fusion pour créer des messages multimodaux. Les critères et les mécanismes de cette fusion sont très variés. Pour plus d'informations, voir [Bellik, 1995, Martin et al., 1995].

Le *Générateur de Présentations Multimodales* est chargé de faire la synthèse de la présentation multimodale à partir d'un message décrivant la structure du discours. Il a accès à la taxinomie des modalités, et aux gestionnaires de tous les périphériques. De plus, il peut consulter tous les modèles décrits dans la section suivante.

Le processus qui conduit à la création et à la réalisation d'un énoncé multimodal est décrit dans la section 6. Brièvement, trois agents sont chargés respectivement de sélectionner les modalités, adapter les canaux de chaque modalité pour la présentation de l'information, coordonner temporellement, spatialement et sémantiquement les modalités. Ces opérations se font coopérativement dans le sens où, par exemple, une coordination sémantique de plusieurs modalités pourra avoir une influence sur la valeur de l'un des canaux de l'une de ces modalités.

Les messages à présenter sont organisés au sein du contrôleur d'interaction dans une file d'attente. Ils sont classés dans cette file par ordre de criticité décroissante (voir 4.1).

3.4 Les modèles

Tous les modèles comportent deux parties, l'une statique, l'autre dynamique. La partie statique contient toutes les connaissances connues a priori. La partie dynamique contient des informations sur l'évolution des modèles au cours du dialogue.

Nous ne décrivons ici que la partie des modèles utilisée par la synthèse de présentation multimodale et donnons des exemples de règles en relation notament avec les modèles de l'utilisateur et de l'environnement. Les règles relatives au dialogues et à l'intéraction sont abordées plus loin.

3.4.1 Modèle de l'interaction

La partie statique du modèle de l'interaction contient les règles modélisant les considérations ergonomiques de communication à la disposition du IMMPS. Ces règles contrôlent la validité d'un énoncé vis à vis de la coopération entre modalités, de la coordination temporelle et spatiale, etc (voir section 6.2.2). Elle contient aussi les connaissances sur les contraintes liées aux media.

La partie dynamique contient un historique des présentations, qui permet de savoir si une information est déjà présente, et les modalités qui ont été utilisées au cours du dialogue écoulé pour présenter certaines informations. Elle contient aussi des informations sur l'état courant des media.

3.4.2 Modèle du dialogue

Ce modèle contient la taxinomie des actes de dialogue. Ceux-ci ont été abordés dans la section 2.2.1 et le lecteur trouvera plusieurs commentaires relatifs à leur implémentation dans la section 3.1. La partie dynamique contient un historique du dialogue.

Il pourra aussi contenir des règles de résolution d'anaphores et de méta-dialogues. Par exemple, s'il y a une incohérence entre une valeur saisie par l'utilisateur et la valeur attendue par le système, le contrôleur de dialogue devra engager un méta-dialogue en

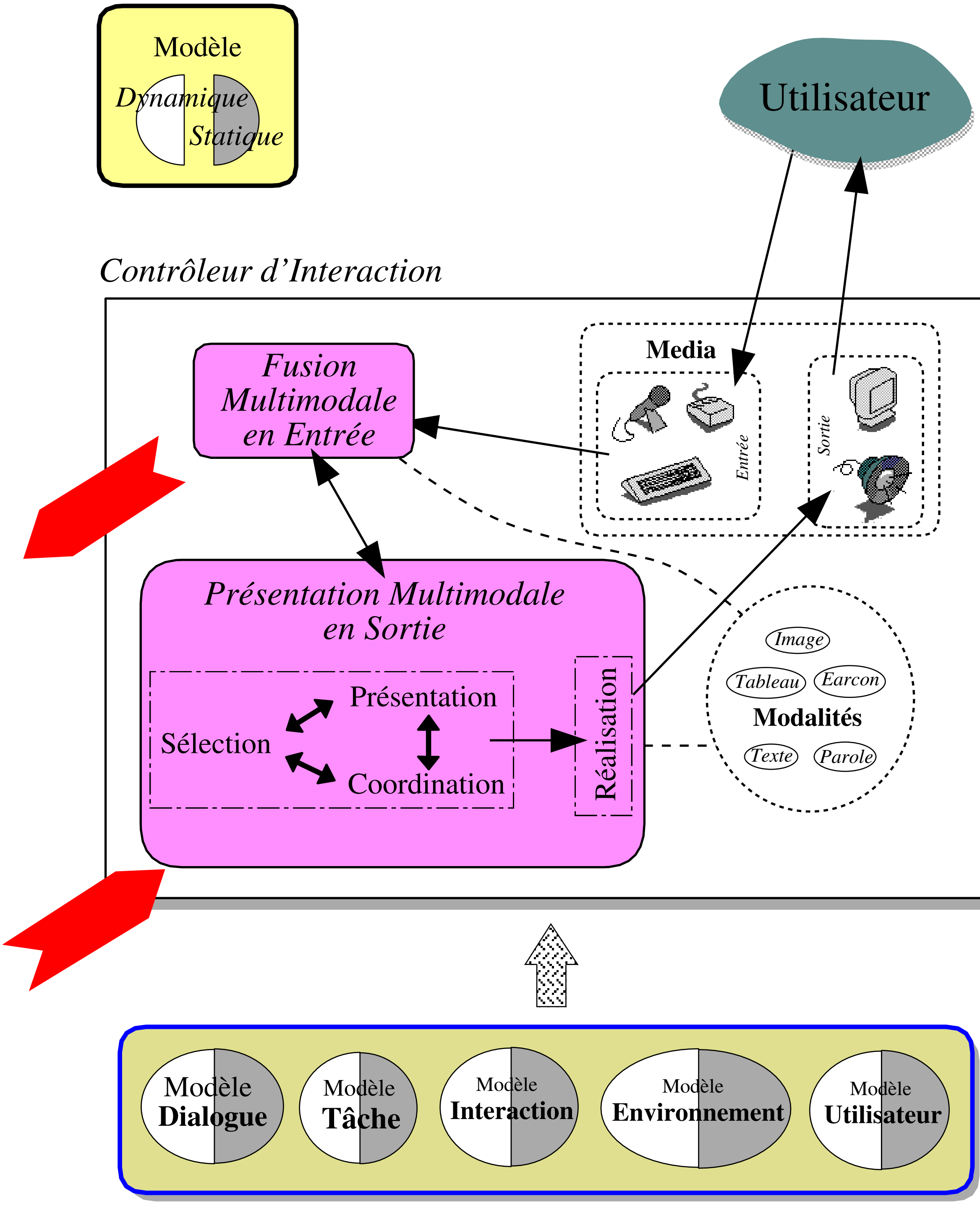


FIG. 6 - *Contrôleur d'interaction*

utilisant certaines règles pour indiquer à l'utilisateur qu'il a commis une erreur et lui demander de saisir une valeur correcte.

3.4.3 Modèle de la tâche

Le modèle de la tâche dépasse le contexte de cette étude. On trouvera toutefois une présentation détaillée dans [Ferrari, 1994, Bellik et al., 1994]. Ce modèle décrit les caractéristiques fonctionnelles de la tâche.

Les tâches se répartissent en *tâches complexes*, qui sont décomposables en un certain nombre de sous-tâches. Les tâches indivisibles sont appelées *tâches élémentaires*.

Il existe des règles pour la composition des tâches. Les sous-tâches peuvent être organisées séquentiellement, parallèlement ou sans contraintes temporelles.

Comme pour les actes de dialogue, il est intéressant d'obtenir un formalisme qui soit indépendant de l'application. Pour cela, les tâches élémentaires sont réparties entre :

- le flot interface-utilisateur (par exemple, demande de saisie ou de confirmation d'informations),
- le flot interne du IMMPS (par exemple les accès aux différents modèles), et
- le flot interface-application (par exemple, les commandes envoyées à l'application).

Afin de décrire ces tâches élémentaires et leurs compositions pour former des tâches complexes, [Bellik et al., 1994] définit un langage formel. Les tâches sont organisées dans une hiérarchie de classes. Enfin, pour déterminer le contexte d'une tâche élémentaire, les auteurs définissent la PPTCC, la *Plus Petite Tâche Complexe Courante*, et la PGTCC, la *Plus Grande Tâche Complexe Courante*.

3.4.4 Modèle de l'utilisateur

Caractéristiques de l'expertise. Il s'agit de modéliser le niveau d'expertise dans l'utilisation des interfaces d'une part et dans la tâche d'autre part. Nous pouvons, par exemple, différencier les utilisateurs experts et novices. Il faut noter que ces caractéristiques ont une influence sur ce qui est présenté (il faudra fournir des explications plus claires pour un novice) et sur la forme de la présentation.

Dans [Gargan et al., 1988], les auteurs définissent un modèle hiérarchique des utilisateurs permettant de spécialiser certaines caractéristiques. Ces modèles sont :

- le modèle *canonique* qui contient des informations conceptuelles propres à tous les utilisateurs,
- le modèle *stéréotypique* qui contient des informations sur des classes d'utilisateurs dans un domaine d'application donné, et
- le modèle *individuel* de l'utilisateur.

Caractéristiques physiques. Il faut prendre en compte les aspects physiques de l'utilisateur qui ont une influence sur ses capacités perceptives vis à vis de l'interface. Par exemple, pour un utilisateur sourd, il est inutile de produire des énoncés à destination du mode auditif.

Plusieurs paramètres jouent un rôle. Par exemple, simplement la présence devant l'écran, la distance par rapport à celui-ci, la direction de son regard, peuvent être prises en compte. Cela nécessite l'utilisation d'un certain nombre de capteurs. Par exemple, une caméra pourra être utilisée pour déterminer si l'utilisateur est présent ou absent, et éventuellement sa distance par rapport à l'écran. De même, un oculomètre pourrait évaluer la direction du regard.

Si l'utilisateur ne regarde pas l'écran, il est important d'attirer son attention surtout si l'énoncé est critique vis à vis de la tâche. On pourra alors utiliser des modalités sonores qu'il pourra percevoir.

D'autre part, si l'utilisateur parle pour donner une commande vocale (ou parle simplement avec un autre interlocuteur), il faut éviter d'utiliser des modalités sonores qui pourraient perturber la reconnaissance vocale.

Nous nous proposons d'utiliser les règles suivantes, à titre d'exemple, en ce qui concerne le modèle de l'utilisateur :

- ▷ R1 *si l'utilisateur parle, interrompre la synthèse en cours, ne pas utiliser la synthèse vocale. Des* earcons *peuvent toutefois être utilisés.*
- ▷ R2 *si l'utilisateur utilise un gant numérique pour une manipulation, éviter les présentations qui nécessitent une désignation gestuelle.*
- ▷ R3 *si l'utilisateur ne regarde pas l'écran, si possible diffuser par synthèse vocale, sinon attirer son attention par des* earcons.

Bien sûr, beaucoup d'autres règles peuvent être énoncées, en relations notament avec le type de tâche que l'utilisateur doit accomplir ou contrôler.

3.4.5 Modèle de l'environnement

L'environnement est un facteur important qui peut influencer la perception d'un énoncé par l'utilisateur. Par exemple, le niveau sonore ambiant peut perturber l'identification d'un message sonore[6]. De même, des vibrations excessives peuvent gêner l'utilisateur lors de la lecture de données sur l'écran.

Afin de pouvoir prendre en compte ces paramètres, il est indispensable d'en faire une étude. En effet, ce qui peut être considéré comme bruyant dans un environnement peut ne pas l'être dans un autre. De même, la parole a une structure particulière qu'il convient d'identifier afin de ne pas la confondre avec un bruit. Une étude ergonomique du déroulement de la tâche et de son environnement est donc nécessaire.

6. De même que cela peut gêner la reconnaissance vocale en entrée.

Nous nous proposons d'utiliser les règles suivantes en ce qui concerne le modèle de l'environnement :

- ▷ R4 *si des vibrations ont lieu, éviter les présentations nécessitant des désignations précises. Par exemple : augmenter la taille des fenêtres, boutons, polices de caractères, ...*
- ▷ R5 *ajuster la luminosité de l'écran (si celle-ci est programmable) à celle ambiante.*
- ▷ R6 *s'il y a trop de bruit ambiant qui ne correspond à aucun modèle, éviter les commandes vocales.*

Là aussi, beaucoup d'autres règles peuvent également être énoncées.

4 Structure du discours

Un segment du discours est véhiculé par un message. Nous parlerons ici exclusivement des messages de sortie (dans le sens machine vers homme).

Un message de sortie est un objet actif envoyé par l'application pour que le gestionnaire d'interaction le présente à l'utilisateur. Le message est actif au sens où la valeur de certains paramètres est évaluée dynamiquement et de façon répétée au cours de la synthèse de la présentation.

Le message comporte une dualité entre sa valeur communicative (la forme, l'intention et l'importance de la présentation) et sa valeur informative (le fond, le contenu sémantique).

4.1 Attributs des messages

La notion de message a été discutée dans [Bellik et al., 1994]. Toutefois, ces travaux n'incluaient pas d'étude des types sémantiques. De même, l'influence des caractéristiques dynamiques de la présentation n'avait pas été évaluée.

Nous proposons de caractériser les messages avec les attributs suivants :

Label : identifie le message.

Type du message : indique le but recherché par la présentation du message, en relation avec les actes de dialogue, comme décrit précédemment.

Criticité : indique l'importance du message dans le contexte de la tâche courante. Les valeurs possibles, identifiées pour l'instant, sont : *Vitale*, *Urgente*, *Pertinente*, et *Nominale*. Cet attribut permettra de choisir les modalités les plus perceptibles et d'accentuer certains de leurs paramètres (clignotement, prosodie) en fonction du degré de criticité. Cette criticité est relative à la tâche et sera fixée par celle-ci.

Type sémantique et valeur de l'information : définissent le contenu de ce qui doit être présenté (voir section suivante). Il est à noter que le type est indépendant de la modalité choisie pour le présenter. Il faudra établir une équivalence entre celui-ci et les différentes formes de présentation possibles. C'est cet aspect en particulier qui est au coeur de cette étude.

Priorité de communication : indique l'importance du message lorsque plusieurs messages de même criticité doivent être présentés simultanément. C'est la seconde clef pour le classement des messages dans la file d'attente (après la criticité, voir section 3.3). C'est le contrôleur de dialogue qui affecte cette valeur au message en fonction de son importance dans le contexte du dialogue en cours. Nous définissons, pour cette étude, une priorité *haute*, *moyenne* ou *faible*.

Durée de présentation : indique la durée pendant laquelle le message doit être perceptible pour l'utilisateur. Ce paramètre a une influence différente en fonction du caractère transitoire ou non-transitoire (voir section 5.1) des modalités qui formeront l'énoncé du message.

En effet, les modalités transitoires ont une durée de présentation qui leur est intrinsèquement liée (par exemple pour la parole). Il est alors impossible de prolonger la présentation de ces modalités au delà de cette durée sinon en répétant le message. A l'opposé, il serait erroné d'interrompre la présentation d'un énoncé avant qu'il ne soit complet sur ce type de modalité (sauf dans le cas où un message plus critique doit être présenté sur cette modalité). Il est, par contre, possible de choisir de répéter un message dont la durée d'énonciation serait plus courte que la durée de présentation désirée.

Les modalités non-transitoires pourront être supprimées, au-delà de cette durée de présentation.

Celle-ci est déterminée par la tâche pour que son déroulement soit optimal.

Durée de validité : c'est la durée pendant laquelle le contenu sémantique du message est considéré comme valide. Au delà, le contrôleur d'interaction préviendra le contrôleur de la tâche que la valeur doit être réactualisée. Celui-ci devra alors envoyer un nouveau message. Ainsi apparaît l'importance du déterminisme de la synthèse de la présentation pour qu'un même message avec les mêmes paramètres soit affiché de la même façon s'il apparaît plusieurs fois. C'est aussi un des aspects de la nécessaire coopération entre le contrôleur d'interaction et l'application, mise en valeur dans la section 2.1.

Comme pour la durée de présentation, cette valeur est déterminée par l'application.

Ensemble de modalités : permet de spécifier (complètement ou en partie) les modalités à utiliser pour la présentation.

Il faut noter que les messages étant des objets dynamiques, la valeur de certains de leurs attributs est évaluée à chaque fois qu'elle est lue. Ainsi, lorsque plusieurs messages sont dans la file d'attente du contrôleur d'interactions, leur criticité et leur priorité peuvent évoluer, et leur place dans la file peut donc changer. Il est nécessaire de la réordonner avant de choisir le nouveau message à présenter.

4.2 Taxinomie de l'information

Par *information*, nous désignons le contenu sémantique du message à présenter. La représentation de cette information est indépendante de toute représentation des données au sein de l'application. Ainsi, si l'IMMPS utilise un formalisme de base de données, de schéma (au sens des *frames* de l'Intelligence Artificielle) ou de théorie des graphes, il aura du même coup les caractéristiques de ces représentations. Par exemple, le modèle relationnel des bases de données est utile pour représenter les relations entre des ensembles de données. Toutefois, il est peu adapté à la représentation des données singulières comme un total ou une moyenne.

Il est donc important de mettre en avant les caractéristiques des informations qui ont une influence sur leur présentation. Il n'est pas question de déterminer tous les types d'information possibles, mais seulement de fournir un certain nombre de critères qui permettront de caractériser ces types dans le cadre de la création d'une interface pour une application.

Pour décrire les informations à présenter, nous utilisons deux catégories de paramètres :

- un identificateur de type sémantique qui décrit les caractéristiques statiques de l'information,
- un ensemble de paramètres décrivant l'état dynamique de l'information.

4.2.1 Les types sémantiques

Il s'agit de caractériser l'appartenance d'une information à un type, une classe ou un domaine. Pour cela, nous définissons les critères suivants :

Critère de complexité : permet de distinguer une information simple (comme la valeur d'un angle) d'une information composite, c'est-à-dire qui exprime une relation entre plusieurs autres informations. Les types de ces sous-informations peuvent être homogènes (comme des coordonnées dans un espace 3D) ou hétérogènes (comme dans le cas de la description d'un horaire de train, qui est constituée d'un identifiant de train, des heures et des lieux de départ et d'arrivée, ...).

La complexité peut être vue comme une relation entre plusieurs types de données hétérogènes. Celle-ci a une arité[7] qui peut être prise en compte. Par convention, nous appellerons *type d'information pur*, un type qui ne sera pas *composite*.

Dans le cas des types complexes, il est nécessaire d'avoir une liste de sous-types indiquant les types sémantiques de chacun des composants du type. De même, nous

7. Nombre de dimensions de la relation.

associons à chaque sous-type un ensemble de règles qui permettent de spécialiser le comportement de ce sous-type.

Critère de domaine : permet de regrouper différents types de valeurs par catégorie sémantique. Par exemple, des domaines possibles sont la date, la position dans l'espace, la masse, la température, la monnaie. Il est intéressant de noter que ce type de caractérisation est utilisé depuis longtemps dans les tableurs, essentiellement dans le but de déterminer le format de présentation. De plus, ce critère pourra être utilisé afin de regrouper pour l'affichage les informations portant sur un même domaine.

Critère de densité : indique si le domaine est dense[8], discret, ou même énuméré ou booléen. En effet, pour une information de type dense (ou *continu*), on utilisera de préférence une représentation également continue.

Les critères ci-dessus permettent de définir un ensemble de types sémantiques abstraits (ou génériques), c'est-à-dire non-directement exploitables par l'application. A partir de ces types, il est possible de définir plusieurs hiérarchies de types sémantiques concrets. Ils héritent des caractéristiques de leur supertype et peuvent en modifier certains aspects. Le lecteur trouvera dans la section 4.2.3 un certain nombre d'exemples de types sémantiques abstraits et concrets.

D'autres critères sont proposés dans la littérature (voir [Maybury and Hovy, 1994, Roth and Hefley, 1993]). Cependant, en raison de leur spécificité, il semble difficile de les intégrer dans la taxinomie dans le cadre de l'étude. Ils sont toutefois mentionnés ici pour mémoire.

Critère d'identification : précise si l'information est une référence à un objet plutôt que l'objet lui-même. Par exemple, pour l'affichage d'une liste d'avions, ce sont les identifiants des avions qui seront affichés, alors que pour une liste de températures, c'est la valeur même de ces températures qui sera affichée. Il est également possible de ne le considérer que comme une spécialisation du critère de domaine.

Critère de Mesure : il peut être aussi intéressant de distinguer les *coordonnées* des *quantités*. Une coordonnée est définie suivant un ou plusieurs axes (comme un point dans l'espace, une température). Ainsi, cela n'a pas de sens de faire un rapport de proportion entre deux données de ce type, seule une différence (ou un intervalle) a un sens (on peut faire un intervalle entre deux points de l'espace, et non un rapport). A l'opposé, entre deux quantités, il est possible de faire un rapport et une différence (comparer deux prix et faire leur différence).

4.2.2 Les paramètres dynamiques

Nous présentons ici un ensemble d'attributs définis pour toutes les instances d'information quel que soit leur type. Ces attributs ne portent donc pas sur les messages mais bien sur les informations.

Critère de qualité : indique, dans le cadre de la tâche courante, l'état *nominal*, *critique*, ou *fatal* de l'information à présenter.

8. Quels que soient deux éléments de l'ensemble, il existe un troisième élément qui est compris entre les deux premiers.

Il ne faut pas confondre ce critère avec la criticité. En effet, un message peut-être critique sans que la valeur de l'information le soit. Par exemple, dans une phase de maintien d'attitude d'une navette spatiale, il est toujours important de connaître l'angle de roulis, mais, si la manoeuvre est conduite correctement, la valeur de celui-ci sera toujours nominale.

Cet attribut pourrait être soit un indicateur, soit un intervalle de valeurs pour chaque type de qualité. Cette seconde solution peut être peu commode dans le cas général, car pour certains types sémantiques il peut être difficile de construire de tels intervalles. Il est ainsi préférable d'utiliser un indicateur et de déléguer au modèle de la tâche l'évaluation de la qualité de cette information.

Ce critère permet de mettre en valeur la partie du message concernant cette information. De plus, nous pourrions associer à chaque qualité un énoncé, qui permettra d'alerter l'utilisateur.

Critère de Couverture[9] : permet de déterminer la couverture du domaine par les éléments à présenter, indiquant ainsi s'il y en a un seul, peu, beaucoup, ou s'ils sont tous présents. Cela permet, par exemple, de choisir une présentation sous forme de table pour une liste, et de texte pour un seul élément. De plus, pour présenter un ensemble d'éléments, le système peut simplement faire une liste d'objets dans une modalité donnée, toutefois, l'utilisateur serait incapable de savoir si cette liste contient tous les éléments ou seulement une partie. La présentation pourra varier pour mettre en évidence les différentes possibilités de quantification. La valeur de ce critère peut être *singleton*, *faible*, *grande* ou *totale*.

Tous les types concrets obtenus par dérivation, contiennent les attributs définis ci-dessus. Ils peuvent, de plus, posséder des attributs propres.

4.2.3 Exemples de type d'information

Voici deux exemples pour illustrer les différentes caractéristiques et les critères de l'information.

1. **Angle de roulis** : C'est la valeur, en degrés-minutes-secondes ou en radians, de l'angle d'un avion ou d'une navette spatiale (voir section 7.2 et [Bellik et al., 1994]) autour de son axe longitudinal. C'est un type *pur*, son domaine est celui des *angles*, et il est *dense*.

   Sa qualité dépend de la tâche en cours. Si celle-ci consiste à arrimer la navette à une station orbitale, il est toujours critique de connaître la valeur de cet angle. De plus cette valeur doit être très précise et si elle s'éloigne même faiblement de la valeur nominale, elle devient alors *critique*.

   De cela, nous pouvons déduire plusieurs façons de représenter une valeur de ce type. Nous pourrons utiliser soit du texte, soit de la parole, soit un synoptique représentant la valeur. Ce sont les règles liées à l'utilisateur, au dialogue et à l'environnement qui permettront de choisir une combinaison de ces modalités. Toutefois, l'angle de roulis

9. également appelé critère de quantification.

est souvent associé aux angles de tangage et de lacet. Leur présentation est combinée dans un synoptique (appelé *horizon artificiel*) qui permet au pilote d'accéder aux trois valeurs simultanément. C'est un exemple où le domaine d'application impose le choix de la présentation.

2. **Liste d'horaires de trains** : un horaire de train est une valeur *composite*, constituée d'un numéro de train, d'une ville et d'une heure de départ, d'une ville et d'une heure d'arrivée, et éventuellement d'autres informations comme le kilométrage. C'est un type *énuméré*.

   Il doit contraindre les présentations des sous-types qu'il contient. Par exemple, si l'on utilise des synoptiques du type utilisé dans les bornes interactives SNCF, il est souhaitable de contraindre l'affichage de l'heure pour que celle-ci soit toujours affichée sous forme textuelle (`HH:MM`) plutôt que sous forme analogique (en utilisant une horloge).

5 Structure de la présentation multimodale

La structure de présentation est composée d'un ensemble de modalités, pour lesquelles il faut déterminer la forme de présentation et les coordinations spatiale et temporelle. Nous présentons dans cette section une taxinomie des modalités et les problèmes liés à leur coordinations. Enfin, nous étudierons les contraintes que ces problèmes imposent aux media.

5.1 Taxinomie des modalités

Cette taxinomie est basée sur les travaux présentés dans [Bernsen, 1994, Bernsen, 1995a, Bellik et al., 1995, Bellik, 1995, Arens et al., 1993b]. Toutefois, elle diffère de ces travaux originaux en plusieurs points (voir ci-dessous).

Cette taxinomie doit faire le lien entre les représentations externes des modalités et la représentation interne élaborée par l'utilisateur qui la perçoit. Il s'agit donc de décrire les propriétés représentationnelles des modalités.

Les modalités sont classées dans un espace multidimensionnel en fonction des critères suivants :

Critère de variabilité temporelle : indique le caractère *statique* ou *dynamique* de la modalité. Les modalités visuelles peuvent par exemple faire appel à des opérations de clignotement de déplacement, de déformation. Les modalités vocales peuvent utiliser des critères de prosodie pour faire varier l'intonation au cours de l'énoncé d'un message.

Critère d'endurance temporelle : nous avons ajouté au paramètre précédent une indication, absente chez Bernsen et Bellik, de la nature *persistante* ou *non-persistante* d'un énoncé, comme par exemple pour l'utilisation de la parole ou de bips sonores. Ces modalités peuvent être statiques ou dynamiques, mais elles ont un caractère transitoire. A l'opposé, un message visuel est toujours présent et peut être consulté plusieurs fois, même si sa forme change dans le temps.

Bellik et Bernsen ne font pas apparaître ce critère. Il nous a semblé opportun de faire une distinction supplémentaire entre les modalités dont la forme change au cours du temps et celles dont le signal est simplement éphémère.

Critère dimensionnel : indique le nombre de dimensions sémantiques qu'une modalité peut représenter[10]. Par exemple, un tableau est exactement de dimension deux, alors qu'une carte peut avoir plus de dimensions. Une image et un texte (vocal ou écrit) en ont un nombre infini.

Il faut noter que chaque dimension est elle aussi une modalité. Un tableau peut en effet contenir du texte, des images. Toutefois, cela impose certaines contraintes sur l'étendue du choix de ces modalités. Elles doivent toutes correspondre au même mode perceptif de l'utilisateur. Il est, par exemple, impossible d'utiliser exclusivement un message sonore pour décrire les éléments d'un tableau (on pourra par contre utiliser la redondance des présentations). Cela est du à ce que le tableau véhicule de l'information par l'organisation spatiale des modalités qu'il contient.

Critère langagier : indique la nature plus ou moins complexe des structures lexicale, syntaxique, grammaticale et sémantique sur lesquelles sont basées les modalités. Par exemple le texte (vocal ou écrit) repose sur une forte base langagière, contrairement à une image.

Contrairement à Bernsen, mais comme Bellik, nous considérons que ce critère doit être une échelle de valeurs plutôt qu'un descripteur booléen. Il sera ainsi possible de caractériser la complexité du langage sous-jacent.

Critère d'analogie : indique si une modalité véhicule des informations dont la représentation est directement liée à l'entité signifiée[11]. Nous pouvons ainsi connaître le degré de similitude entre la modalité et ce qu'elle représente. Si cette similitude est faible, il faut utiliser des connaissances sur les conventions de représentation utilisées afin de décoder l'information décrite dans la modalité.

Par exemple, les sons du monde réel, et les images diagrammatiques (icônes) sont des représentations analogiques. Le langage écrit et parlé, les graphes et les sons arbitraires ne le sont pas.

Bernsen considère ce critère comme étant booléen. Nous pensons qu'il faut le rendre continu afin de mesurer le degré de ressemblance. Bernsen est conduit à distinguer le cas des images *diagrammatiques* de celui des images du *monde réel*. Toutes les deux sont analogues, mais les représentations diagrammatiques manipulent la représentation de façon variable (par exemple en faisant abstraction de certains détails

10. Il ne s'agit pas des dimensions physiques visibles de la modalité.
11. On dira aussi que ces modalités sont iconiques, ou isomorphiques.

non pertinents pour la tâche). Bernsen note qu'il est difficile de faire la distinction entre ces deux extrêmes et qu'il existe donc un "*continuum de représentation*" entre eux. Il semble souhaitable d'utiliser une échelle continue de valeurs pour qualifier le degré d'analogie.

Critère de détectabilité : indique le degré d'aisance de perception de la modalité par l'utilisateur. Par exemple les sorties sonores sont très détectables (sauf dans un environnement bruité ou pour un utilisateur sourd), ainsi que les modalités dynamiques en général. A l'opposé, un texte simple est typiquement peu détectable.

Il est ainsi possible de produire des modalités abstraites en fonction de ces critères. Il faut noter que, contrairement à la taxinomie de Bernsen, celle définie ici n'est pas générative au sens où il n'est pas concevable de faire l'énumération de toutes les modalités avec la combinaison des différents critères[12]. Le lecteur trouvera dans la section 6.2.2 une discussion sur les propriétés expressives liées aux critères décrits ci-dessus.

Ces modalités sont abstraites et ne peuvent pas être utilisées directement. Des modalités concrètes sont créées par dérivation pouvant ainsi particulariser et enrichir les modalités originales. Les modalités abstraites ne sont pas affectées à un mode particulier. Il est donc ainsi théoriquement envisageable d'avoir des modalités avec les mêmes caractéristiques pour des modes différents.

Dans la taxinomie, il faut distinguer les modalités qui peuvent être utilisées comme substrat pour autre modalité. En pratique, il semblerait que toutes les modalités puissent être utilisées comme substrat d'une autre modalité. Il faut toutefois que ces modalités soient sur le même mode (par exemple, un texte ne peut servir de substrat à un son même si ces deux modalités peuvent être utilisées de manière redondante). Inversement, certaines modalités ne sont que des substrats et ne véhiculent de l'information que par les contraintes qu'elles imposent aux éléments dont elles sont le substrat.

Par exemple, une carte peut à la fois véhiculer sa propre sémantique (comme des représentations des rivières, des forêts, de l'altitude) et servir de substrat à des étiquettes textuelles qui indiquent le nom des villes. Un tableau ou une liste ne comportent toutefois pas de sémantique en eux-mêmes mais en véhiculent par l'organisation spatiale qu'ils imposent aux éléments qui composent leurs cellules.

5.2 Exemples de modalités

Voici quelques modalités décrites dans cette étude. Pour chacune d'elle figure la valeur des critères décrits ci-dessus ainsi que la liste des paramètres (canaux) susceptibles d'être modifiés.

1. *Earcons (bip sonores)* : les earcons sont des messages sonores synthétique (les messages sonores naturels sont appelés *icônes auditives*). Ils peuvent être arbitrairement évolués. C'est une modalité dynamique (le signal peut être modulé dans le temps) et non-persistante.

12. En effet, les critères langagier et analogique ne sont plus binaires, mais continus.

Les earcons sont typiquement des sons synthétiques qui n'existent pas dans la vie courante. Ils sont donc non-analogiques. Toutefois, ils ne sont pas régits par un langage, mais peuvent être arbitrairement évolués et représenter des structures complexes. Cette structure peut aller du simple bip pour indiquer une erreur jusqu'au compteur Geiger qui indique de façon sonore la mesure d'un phénomène.

Canaux : volume, durée, pitch, ...

2. *Synthèse de parole* : la synthèse de parole est une modalité dynamique et transitoire. Elle est basé sur un système de langage très évolué et possède donc un nombre de dimentions sémantiques infinie. Enfin, elle est non-analogique puisqu'elle ne peut que décrire les choses, pas leur ressembler.

   Canaux : prosodie (rythme, intensité, variation du fondamental), ...

3. *Texte écrit* : un texte écrit est à priori statique, persistant et non-analogique. Il est basé sur un langage évolué et possède donc un nombre infini de dimensions sémantiques.

   Toutefois, le texte peut être enrichi par des attributs dynamiques, par exemple en faisant clignoter une mot, ou en manipulant les couleurs.

   Canaux : orientation, taille, police, couleur, ...

Il faudra utiliser également des considérations ergonomiques pour ces modalités. Par exemple, la synthèse vocale est encore peu satisfaisante dans la plupart des systèmes existants, et elle peut entrainer une certaine lassitude chez l'utilisateur. De plus, elle n'est pas confidentielle ce qui être un problème pour certaines informations.

5.3 Coordination des modalités

Comme nous l'avons évoqué dans la section 2.1.3, il faut coordonner sémantiquement, temporellement et spatialement les modalités. Des considérations ergonomiques sont utilisées pour guider cette coordination. Elles sont abordées dans la section 6.2.2.

5.3.1 Synchronisation temporelle des modalités

Les représentations temporelles sont basées sur des *instants* et des *intervalles*. Un intervalle est défini par deux instants et la durée entre eux. Il existe treize relations possibles entre deux intervalles de temps (voir figure 7, où seules sept relations sont présentées, les autres étant les relations inverses[13]). Ces intervalles représentent les composants temporels nécessaires à la synchronisation des modalités, leurs positions relatives indiquant les dépendances temporelles.

Puisque les présentations ont pour vocation d'être interactives, une importante part de la synchronisation peut venir de cette interaction. Par exemple, on pourra choisir de ne diffuser une vidéo que lorsque l'utilisateur aura cliqué sur un bouton donné. Nous n'avons pas intégré, pour l'instant, de solutions pour ce genre d'interaction dans notre architecture; toutefois la coopération entre les générateurs de messages en entrée et en

13. l'égalité et son inverse étant équivalents.

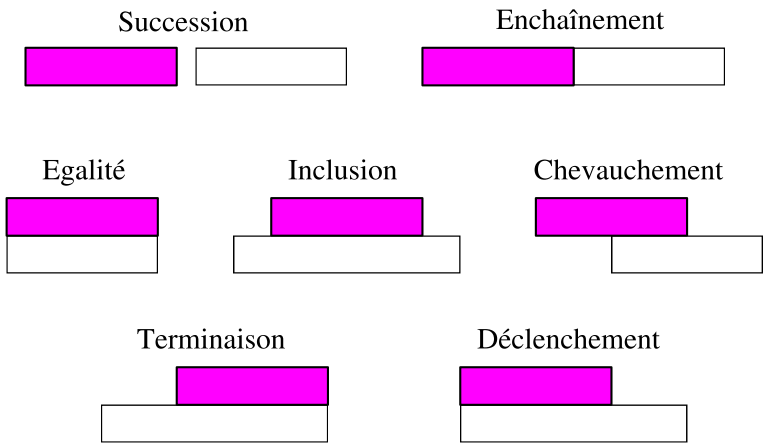


FIG. 7 - *Relations entre intervalles de temps (d'après [Little et al., 1991])*

sortie en facilitera certainement la mise en oeuvre. Dans [Little et al., 1991], l'auteur propose une modélisation sous forme de réseaux de Petri qui sont parfois utilisés dans les systèmes hypertexte. [André and Rist, 1994] présente une extension de WIP pour les présentations actives.

Une description plus détaillée des problèmes de synchronisation est présentée dans [Little et al., 1991]), basée sur des approches langagières et schématiques pour la description des relations temporelles.

5.3.2 Organisation spatiale des modalités visuelles

Comme la synchronisation temporelle, la répartition spatiale des modalités doit refléter les relations sémantiques et pragmatiques entre les données. C'est un problème de type combinatoire. Deux sources de contraintes sont à considérer : les contraintes sémantiques et pragmatiques des éléments et les contraintes géométriques et topologiques des objets affichés. Les premières correspondent à des relations rhétoriques comme la séquence ou le contraste. Les secondes correspondent aux paramètres géométriques constants (comme les coordonnées) et aux relations géométriques entre deux objets.

Ainsi, WIP utilise une approche fondée sur la résolution de contraintes dans un espace de recherche discret et fini. Ces opérateurs de contraintes sont organisés dans une hiérarchie afin que certains puissent être plus prioritaires que d'autres.

Toutefois, il est possible d'utiliser d'autres métaphores que celles employées dans les interfaces graphiques traditionnelles. Par exemple, la représentation des informations au sein d'un monde virtuel en trois dimensions pourrait permettre à des utilisateurs non-expérimentés de conserver les repères qu'ils ont dans le monde réel.

5.4 Contraintes sur les media

Afin d'être utilisés dans le cadre d'une présentation multimodale, les media doivent être capables d'effectuer certaines opérations. Celles-ci sont particulièrement importantes pour la synchronisation des modalités.

Pour pouvoir contrôler l'enchaînement de messages de priorité différente (les plus prioritaires pouvant interrompre les autres), il faut pouvoir suspendre la diffusion d'un message pour diffuser un message plus critique et ensuite éventuellement reprendre ou poursuivre celle du message interrompu. Toutefois, il faudra dans l'avenir faire une étude pour déterminer les critères qui permettront de décider de l'interruption d'une modalité, la granularité de cette interruption (par exemple, à quel moment doit-on interrompre la diffusion d'un message vocal : à la fin d'une phrase, entre deux mots), et les actions qui doivent être entreprises après la diffusion du nouveau message (est-ce qu'il faut reprendre la diffusion du message original depuis le début, ou simplement poursuivre à partir du moment de l'interruption). Ces choix seront à déterminer pour toutes les modalités.

De plus, il est important, lors du choix des modalités, de pouvoir contrôler les media afin de ne pas surcharger l'utilisateur dans sa perception de cette modalité. Une mesure de la charge doit être définie. Cette mesure devra prendre en compte à la fois la capacité du medium à présenter une nouvelle information et celle de l'utilisateur à la percevoir.

Afin de représenter les media et de contraindre leurs caractéristiques, nous avons utilisé la représentation objet décrite dans le modèle d'architecture du LIMSI. Une classe `Device` contient toutes les caractéristiques et fonctionnalités propres à l'ensemble des media. Ensuite, il est possible de créer des media par dérivation. La figure 8 représente la hiérarchie des media. Il faut distinguer les media abstraits (qui ne correspondent pas à des périphériques physiques réels) des media concrets (dont la représentation, dans la figure, est en grisée).

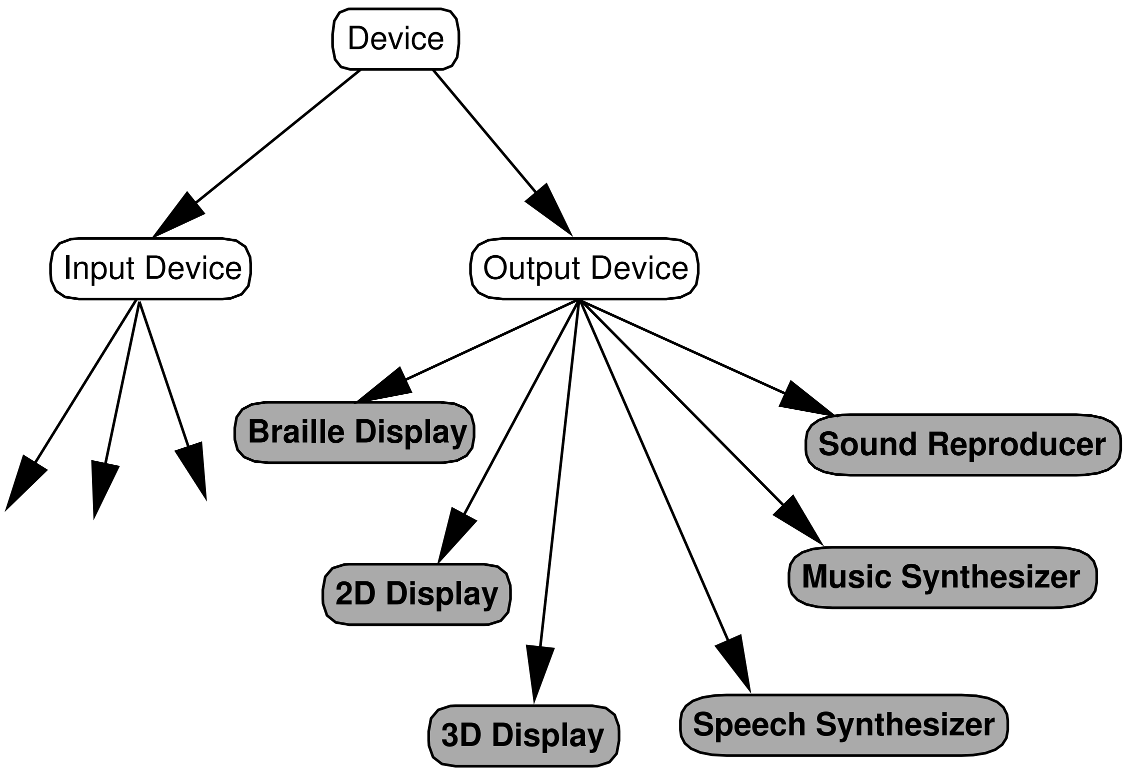


FIG. 8 - *Hiérarchie des media (d'après [Bellik et al., 1994])*

5.5 Représentation de l'énoncé

Il faudra créer une structure qui soit capable de représenter toutes les caractéristiques d'un énoncé multimodal, dont notamment, la liste des modalités, la valeur de leurs canaux, de leurs coordinations spatiale et temporelle.

Les structures utilisées dans les systèmes multimedia peuvent également être utilisées ici. Toutefois, elles s'intéressent surtout aux problèmes de synchronisation temporelle et n'abordent pas les problèmes de répartition spatiale (car celles-ci sont en général prédéterminées).

6 Synthèse de présentations multimodales

Le modèle décrit dans les sections précédentes permet de définir une bonne théorie des modalités. Néanmoins, il reste à construire une bonne théorie de la multimodalité.

Quelle est la *meilleure* présentation multimodale pour un message donné ? C'est celle qui répond à tous les critères décrits dans la section 2.1. Nous allons dans cette section détailler toutes les étapes nécessaires pour atteindre ce but.

Le passage de la structure du discours à la structure de l'énoncé peut se faire comme ceci :

1. *Sélection des modalités possibles en fonction du type de l'information* ;
2. *Application des règles de filtrage* : issues des types d'information, du modèle de la tâche, de l'utilisateur et de l'environnement, de l'historique de l'interaction, de coopération entre modalités ;
3. *Sélection des valeurs pour les canaux* de chaque modalité (par exemple, pour le texte, le choix de la police de caractère, de la couleur) ;
4. *Synchronisation spatiale et temporelle* ; et
5. *Réalisation.*

Les étapes 2 à 4 sont très fortement liées et il doit être possible de passer de l'une à l'autre afin de pouvoir faire des co-références entre modalités.

Il n'est évidemment pas possible d'envisager toutes les façons possibles de combiner un nombre quelconque de modalités. Notre approche vise plutôt à mesurer la qualité d'une combinaison donnée par certains critères objectifs.

6.1 Sélection des modalités

Il s'agit de distinguer les modalités en fonction des types d'information qu'elles sont les plus à même de représenter.

La détermination des types d'information étant essentiellement un travail fait avant la réalisation de l'application par l'analyse détaillée de la tâche et des actes de communication s'y rattachant, il nous semble naturel de vouloir représenter la relation entre l'information et les modalités de façon statique.

Ainsi, pour chaque type d'information, nous créons une liste de modalités qui peuvent le représenter. Les types d'information étant organisés en une hiérarchie[14], il s'agit ainsi de liens entre des noeuds de cette hiérarchie et ceux de la hiérarchie des modalités. Ces liens sont hérités par les descendants du noeud, qui peuvent également les modifier.

Il faut néanmoins établir un certain nombre de règles empiriques pour faciliter la création de ces liens. Certaines de ces règles sont décrites dans [Arens et al., 1993b]. Nous reprenons ici les règles que nous avons incorporé dans notre étude, mais il en existe beaucoup d'autres. Ces règles s'appliquent à plusieurs critères du type d'information, et elles doivent toutes être prises en compte pour la création des liens. Ces règles permettent de plus, dans certains cas, de sélectionner à la fois le porteur d'information et le substrat.

Voici des règles concernant certains critères du type sémantique de l'information et des caractéristiques dynamiques de celles-ci :

- Critère de Complexité de l'information :
  - ▹ R7 *Si le type sémantique est* pur*, utiliser une modalité de dimension sémantique égale à un; pas de restriction sur le substrat.*
  - ▹ R8 *Si le type sémantique est de dimension deux, utiliser un substrat de même dimension. Si toutes les dimensions sont denses, utiliser un substrat continu. Comme porteur d'information, utiliser une modalité simple (icône ou étiquette textuelle).*
  - ▹ R9 *Si le type sémantique est multiple, utiliser un substrat continu si toutes les dimensions sont denses; sinon utiliser un substrat discret; comme porteur d'information, utiliser une modalité avec une dimension supérieure ou égale à un.*
  - ▹ R10 *Si le type sémantique est complexe, vérifier s'il existe une représentation spécialisée.*
- Critère Temporel de l'information :
  - ▹ R11 *Si l'information est dynamique, utiliser une modalité dynamique si sa fréquence de mise à jour est comparable à celle de l'information; sinon utiliser une modalité statique.*
  - ▹ R12 *Si l'information est statique, utiliser une modalité statique.*

14. En fait, il s'agit plutôt d'une forêt que d'un arbre, chaque arbre ayant pour racine un type abstrait.

- Critère de Couverture de l'information :
  - ▷ R13 *S'il y a beaucoup d'informations, ne pas utiliser une modalité avec une faible endurance temporelle que se soit pour le porteur d'information ou le substrat.*
  - ▷ R14 *S'il y a peu d'informations, on peut utiliser des modalités dynamiques.*

[André and Rist, 1993] abordent d'autres problèmes pour la représentation d'informations *covariantes* (c'est à dire plusieurs informations sémantiquement liées, par exemple, par une relation de cause à effet), de la quantification et de la négation.

6.2 Coopération entre modalités

Après avoir obtenu la liste des modalités capables de représenter un type d'information, il faut déterminer la combinaison d'un certain nombre de ces modalités qui formera la meilleure présentation possible. Pour cela, nous pouvons utiliser des techniques de coopération entre modalités.

6.2.1 Types de coopération

Il est possible d'utiliser conjointement plusieurs modalités en fonction de certains critères de coopération décrits dans [Martin et al., 1995, Bellik, 1995] :

- *Equivalence* : Ceci correspond au cas où deux modalités permettent de véhiculer la même sémantique. Bien sûr, les signaux sont différents, mais leur interprétation par l'utilisateur est équivalente (elle ne pourra cependant jamais être exactement la même).
- *Spécialisation* : consiste à utiliser de préférence une modalité pour un type d'énoncé particulier et à ne pas l'utiliser pour d'autres types de modalités.
- *Redondance* : nous pouvons utiliser ce critère afin de clarifier un énoncé en diffusant une même partie de l'information sur deux modalités différentes (voir ci-dessous). Il est préférable d'utiliser des modalités correspondant à des modes différents afin de ne pas surcharger l'utilisateur dans la perception d'un même mode.
- *Complémentarité* : nous pouvons présenter des aspects différents d'une même information sur plusieurs modalités. Par exemple, pour décrire un carré rouge, il est possible de présenter l'image de ce carré et de l'annoter avec un commentaire "*carré rouge*".

6.2.2 Aspects ergonomiques et pragmatiques

Nous présentons ici un certain nombre de règles sur l'utilisation conjointe de plusieurs modalités, sur la répartition des données à travers ces modalités, et sur des techniques pour les rendre plus expressives.

Modalités langagières et modalités analogiques. Les modalités langagières ont la particularité d'être peu *expressives* mais très *directives*, c'est-à-dire qu'elles fournissent un *focus* (une identification) sur l'information pertinente sans donner d'indications spécifiques relatives à cette information. Par exemple, le nom d'une ville désigne cette ville sans ambiguïté mais ne véhicule pas d'informations relatives à celle-ci (voir section 2.2.1). Le champ d'interprétation est donc très large. Inversement, les modalités analogiques ont la particularité d'être très *expressives* mais peu *directives*. Un plan peut véhiculer beaucoup d'informations de nature différente mais aucune d'entre elles n'est explicitement désignée par la représentation. L'expressivité des modalités analogiques vient du fait qu'elles reposent sur un système à plusieurs dimensions, comme par exemple l'espace 2D.

Il est intéressant de combiner ces deux types de modalité pour une présentation, en utilisant par exemple à la fois du texte et de l'image. Dans le cas de l'angle de roulis, on pourra utiliser le synoptique traditionnel et la valeur de l'angle en chiffre, le texte étant une annotation du graphique et le graphique une illustration du texte. Cela est mis en évidence dans les interfaces à base d'icônes, car la diversité de ceux-ci nécessite l'utilisation complémentaire de texte pour les identifier. Ils sont plutôt utilisés pour désigner une classe.

Une autre possibilité est l'utilisation de l'image dynamique ou enrichie, où un effet de couleur ou de clignotement permettrait de spécifier l'objet du message dans une image (voir [Maybury and Hovy, 1994, Goodman, 1993]). Par exemple, dans la figure 9, un cache translucide est utilisé pour mettre en valeur une partie de l'image et attirer ainsi l'attention de l'utilisateur.

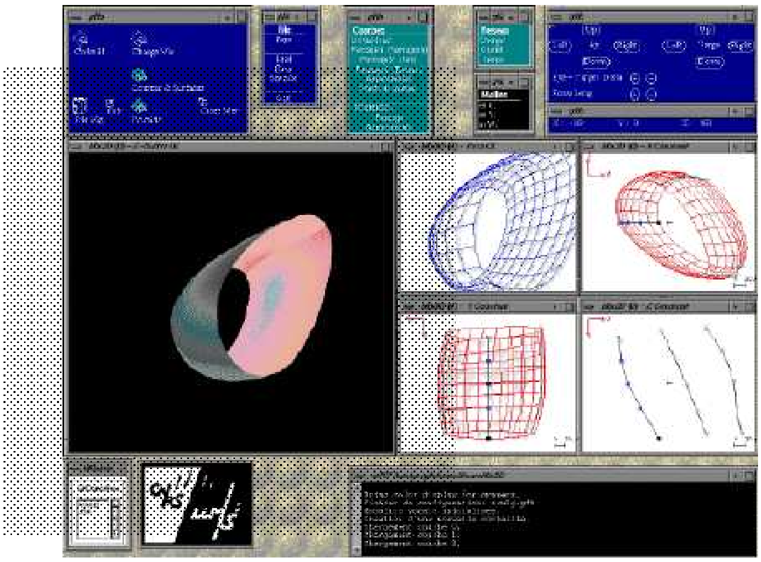


FIG. 9 - *Mise en valeur d'une partie d'une image (image tirée de [Bourdot et al., 1995])*

Nouvelles modalités et métaphores. De nouvelles techniques de présentations peuvent être envisagées. L'utilisation d'un cache est étendue dans [Harrison et al., 1995] pour créer des objets semi-transparents qui permettent de guider l'attention de l'utilisateur sur une information donnée.

Dans [Bederson and Hollan, 1995], l'interface est basée sur une technique de zoom. Les informations sont présentées dans un monde de taille quelconque et des mécanismes de zoom permettent d'accéder à certaines parties de la présentation.

Dans [Mynatt and Edwards, 1992], les auteurs proposent d'augmenter les interfaces visuelles de retours sonores afin de faciliter leur utilisation par des personnes non-voyantes. Ainsi, certains sons peuvent être associés aux différents composants d'une interface visuelle (bouton, menu, etc) et des paramètres de ces sons peuvent indiquer l'état de ces composants (comme par exemple un son plus ou moins grave selon que l'option d'un menu est active ou pas). Comme pour les interfaces visuelles, il faudra utiliser une convention de représentation, un *Hear and Feel*. Il faut noter que cette approche est directement utilisable dans notre modèle par l'utilisation de fonctions réflexes.

L'utilisation du son 3D est aussi très prometteuse (voir [Beaudouin-Lafon, 1994]). En effet, plusieurs informations sonores peuvent être véhiculées simultanément en positionnant leur source dans différents points de l'espace. Le cerveau humain est capable, dans une certaine mesure, d'assimiler plusieurs signaux auditifs simultanément (phénomène connu sous le nom de *cocktail party effect*).

6.3 Filtrage

Ce filtrage a pour but d'obtenir, à partir de la liste de modalités capables de représenter un type d'information donné, une liste de modalités pour l'énoncé. Il est nécessaire d'opérer ce filtrage à cause de la multiplicité des représentations possibles pour une même information (voir section 6.1). Il est inutile (et souvent préjudiciable à la compréhension) d'utiliser toutes les modalités possibles pour un énoncé.

Afin de différencier des modalités qui sont toutes capables de représenter la même information (comme le texte, l'image et les cartes pour la position d'un bateau, voir section 2.1), nous pouvons utiliser des règles ergonomiques. Par exemple :

- ▷ R15 *Si plusieurs modalités peuvent être utilisées, et s'il existe déjà une présentation, préférer les modalités qui sont déjà présentes dans la présentation.*
- ▷ R16 *Si plusieurs modalités peuvent être utilisées, et s'il y a d'autres informations qui doivent également être présentées, préférer les modalités qui peuvent accommoder toutes ces informations.*

Le filtrage peut se faire par l'application des règles de sélection au sein d'un système expert ou d'un générateur de plan (voir section 7). Les règles proviennent de plusieurs sources présentées dans ce rapport. Nous avons tenté, pour ce rapport, de représenter ces règles au sein d'un tableau. Ces tableaux permettent de sélectionner la valeur de certaines caractéristiques des modalités en fonction des caractéristiques de l'information et des modèles. Le tableau 1 contient les règles relatives aux messages et aux caractéristiques dynamiques de l'information. Celles qui concernent les caractéristiques des modèles de l'utilisateur et de l'environnement sont regroupées dans le tableau 2.

Dans notre système, nous avons choisi d'appliquer d'abord les règles de filtrages liées à l'information et au message, puis celles destinées à l'adaptation à l'utilisateur et à l'environnement. Les mécanismes exacts de la mise en oeuvre de ces règles restent toutefois à préciser.

Modalité			**Mode**
Variabilité	Endurance	Détectabilité	
D : Dynamique S : Statique	P : Persistant T : Transitoire	H : Haute M : Moyenne B : Basse	• : imposer ○ : interdire ⋆ : indifférent

	Modalité			**Mode**		
	Variabilité	Endurance	Détectabilité	Sonore	Visuel	Tactile
Information						
Qualité						
Fatale	**D**	**P**	**H**	•	•	
Critique	S - D	**P**	H - M - B	•	⋆	
Nominale	S - D	T - P	H - M - B	⋆	⋆	
Couverture						
Singleton	S - D	T - P	H - M - B	⋆	⋆	
Faible	S - D	T - P	H - M - B	⋆	⋆	
Grande	S - D	**P**	H - M - B	○	⋆	
Totale	S - D	T - P	H - M - B	⋆	⋆	
Message						
Criticité						
Vitale	S - D	**P**	H - M - B	⋆	•	
Urgente	S - D	**P**	H - M - B	⋆	⋆	
Pertinente	S - D	T - P	H - M - B	⋆	⋆	
Nominale	S - D	T - P	H - M - B	⋆	⋆	
Priorité						
Haute	*Nous n'avons pas utilisé ce critère à l'heure actuelle*					
Moyenne						
Faible						

Exemple de lecture :

- Si l'information est *Fatale*, utiliser des modalités dynamiques, persistantes et très détectables, sur les modes sonore et visuel.
- Si la couverture de l'information est grande (il y a beaucoup d'éléments), ne pas utiliser une modalité transitoire, ou qui est sur le mode sonore.

TAB. 1 - *Règles de filtrage pour le message et l'information*

	Modalité			Mode		
	Variabilité	Endurance	Détectabilité	Sonore	Visuel	Tactile
Utilisateur						
Localisation						
Absent	S - D	**P**	**H**	•	•	
Présent	S - D	T - P	H - M - B	⋆	⋆	
Proche	S - D	T - P	H - M - B	⋆	⋆	
Eloigné	S - D	T - P	**H**	•	•	
Regard						
Attentif	S - D	T - P	H - M - B	⋆	⋆	
Distrait	**D**	**P**	H - M - B	•	⋆	
Environnement						
Niveau sonore						
Calme	S - D	T - P	H - M - B	⋆	⋆	
Bruyant	S - D	T - P	H - M - B	∘	•	

TAB. 2 - *Règles de filtrage pour les modèles*

L'analyse de ces tableaux révèle que les critères de dimension, de langage et d'analogie des modalités sont peu sollicités dans ces tableaux. En effet, ces critères sont liés essentiellement aux caractéristiques statiques de l'information et font donc intervenir les caractéristiques sémantiques de la modalité. Ils sont utilisés pour établir la liste originale des modalités capables de représenter un type d'information. De plus, certains critères des messages (comme la priorité de communication) et de l'information (comme la couverture) n'affectent pas de manière significative le choix des modalités mais seulement la sélection de leur forme de présentation.

Ces tableaux sont simplement une tentative pour représenter les règles et la combinaison de critères entre eux. Ils ne sont probablement pas satisfaisants en tant que représentation interne des connaissances mais ils peuvent mettre en évidence les relations entre les différents critères.

6.3.1 Répartition des informations

Pour les données de type composite, c'est-à-dire représentant une relation entre des informations de type hétérogène, la présentation peut être très complexe. Les domaines composant ces données peuvent être distribués à travers :

- des attributs multiples d'une même modalité (par exemple, les couleurs, formes, tailles et/ou positions d'un ensemble d'objets dans un graphe),
- des ensembles d'objets d'une modalité (par exemple, un groupe de mots ou de symboles graphiques sur un plan), et
- des modalités différentes (par exemple, un ensemble de tableaux et de graphes, chacun se rapportant à un domaine unique).

Nous n'avons pas étudié ces problèmes de répartition en profondeur. Leur étude produira de nouvelles règles pour la sélection des modalités. Celles-ci devront être ajoutées à celles décrites ci-dessus. Toutefois, il faudra prendre soin à l'ordre d'application de ces règles car, le fait qu'une information fasse partie d'une donnée composite pourra contraindre sa présentation. Par exemple, pour afficher l'heure de départ d'un train, l'usage veut que l'on utilise un affichage sous forme digitale (par exemple 20:30) plutôt qu'analogique (une pendule).

6.3.2 Validation

Au cours de la conception d'une présentation, une série de règles de validation sont appliquées. Elles tentent de résumer certains problèmes cognitifs relatifs à la présentation multimodale. Les règles suivantes sont tirées de [Sutcliffe and Faraday, 1994].

- ▹ R17 *si des modalités visuelles et verbales sont utilisées, vérifier que les informations de chaque modalité sont liées sémantiquement.*
- ▹ R18 *utiliser du texte ou des images fixes pour des informations clés – cette information serait perdue sur des modalités non-persistantes.*
- ▹ R19 *utiliser le mode sonore pour des avertissements.*
- ▹ R20 *éviter l'utilisation de plusieurs techniques de focalisation de l'attention (clignotement et mise en valeur par un cadre transparent par exemple).*

On remarquera que certaines de ces règles concernent les *modes* et s'appliquent donc à plusieurs modalités.

7 Mise en oeuvre et application

Nous avons présenté, au fil de ce rapport, des règles et des conseils pour l'analyse des informations, des actes de dialogues et des modalités afin de créer une présentation multimodale satisfaisante. Ainsi, avant toute mise en oeuvre du système pour une application donnée, il faudra faire une étude complète des données manipulées par cette application et des buts recherchés dans le dialogue avec l'utilisateur.

Les types d'information et les buts de communication sont définis pour un domaine d'application particulier. Il en va de même pour les règles qui permettent de filtrer les différentes modalités possibles pour une présentation en fonction de ces buts. Il s'agit d'analyser les aspects structurels, fonctionnels et intentionnels de la tâche, comme décrit dans [Maybury, 1993b, Sutcliffe and Faraday, 1994]. Les aspects fonctionnels et intentionnels doivent être considérés car, en général, il y a plusieurs manières de réaliser une tâche et plusieurs façons d'en communiquer le résultat. Sans une description du but d'un

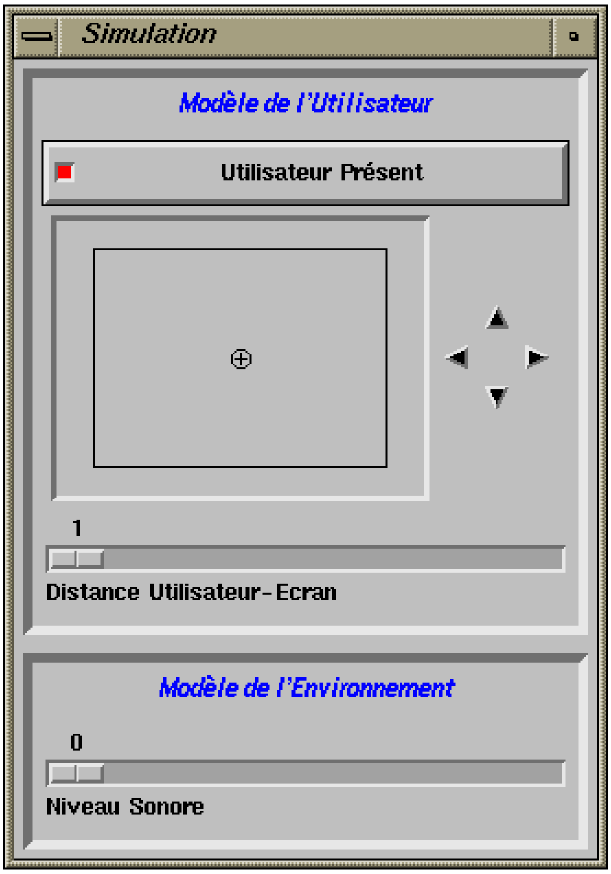


FIG. 10 - *Simulation des modèles de l'utilisateur et de l'environnement*

énoncé, le IMMPS ne pourrait comprendre les équivalences entre ces méthodes. Ce sont deux aspects distincts mais complémentaires de la description d'une tâche.

Afin de pouvoir mettre en oeuvre toutes ces règles, il faudra concevoir une représentation des connaissances et des mécanismes pour réaliser des inférences afin de produire l'énoncé souhaité.

7.1 Système réalisé

Afin de valider les critères identifiés précédement, nous avons réalisé un outil permettant de faire varier les critères retenus et d'observer l'influence sur la présentation de l'information.

Cet outil est constitué de trois modules :

- un module de manipulation des modèles de l'utilisateur et de l'environnement,
- un module de manipulation des caractéristiques des messages et de l'information
- un module pour la réalisation de la présentation correspondante.

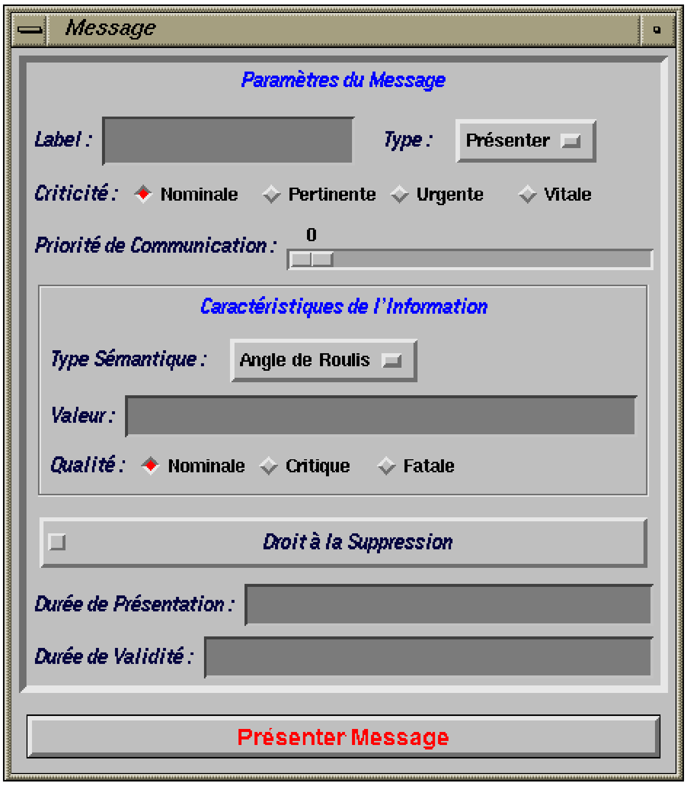


FIG. 11 - *Manipulation des messages*

7.1.1 Simulation des modèles

La figure 10 représente la fenêtre du simulateur qui permet de manipuler les caractéristiques des modèles de l'utilisateur et de l'environnement. Le cadre tracé à l'intérieur du synoptique représente l'écran que l'utilisateur regarde (la taille du cadre change avec la distance de l'utilisateur par rapport à l'écran), et le marqueur symbolise la direction de son regard (il est ainsi possible de savoir si l'utilisateur regarde l'écran ou non).

Comme décrit dans [Beaudouin-Lafon, 1993], ces mesures pourraient être réalisées à l'aide de capteurs de contexte.

7.1.2 Manipulation des messages

La figure 11 représente la fenêtre du simulateur qui permet de manipuler les caractéristiques du message et de saisir la valeur de l'information qui doit être présentée.

Lorsque l'utilisateur a saisi tous les paramètres nécessaires, il peut envoyer le message au système de synthèse de présentation multimodale.

7.2 Application

Le simulateur réalisé pour cette étude est capable de présenter des informations concernant uniquement la valeur de l'angle de roulis d'un avion ou d'une navette, et d'adapter cette présentation aux caractéristiques du message, de l'information, de l'utilisateur et de l'environnement.

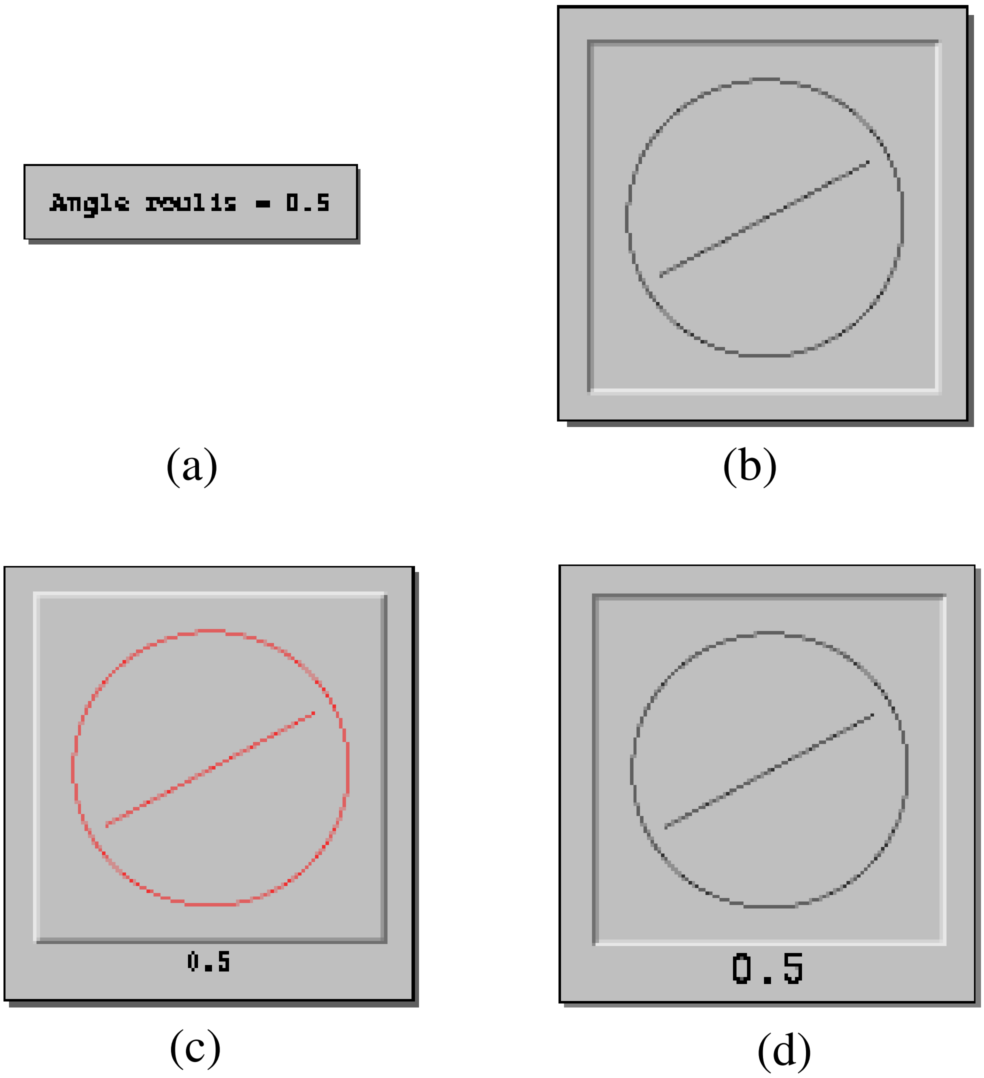


FIG. 12 - *Résultat de la présentation d'un angle de roulis*

A chaque présentation, le système indique les règles qu'il a utilisé (cette trace n'est pas reprise dans ce rapport). La figure 12 contient plusieurs présentations réalisées par le simulateur :

- En 12(a), la criticité du message est la plus faible possible. L'utilisateur est bien présent, proche et attentif. La présentation résultante utilise donc simplement un texte.
- En 12(b), la criticité est maintenant pertinente, la représentation utilise donc un synoptique qui est plus grand et plus facilement lisible.
- En 12(c), la criticité est maximale. Le synoptique est donc présenté au centre de l'écran et la valeur de l'angle est affichée textuellement. L'information est maintenant fatale, donc la couleur du synoptique change.
- En 12(d), la criticité est redevenue urgente et l'information est simplement critique. Le synoptique est donc déporté sur le côté de l'écran. Mais l'utilisateur est maintenant loin, la taille de la police de caractère est donc augmentée.

Certaines adaptations sont purement sonores. Ainsi, le système émet un bip sonore si l'information est critique ou fatale (le bip est différent pour les deux cas). De même, si l'utilisateur ne regarde pas l'écran (pour attirer son attention).

7.2.1 Implémentation de l'architecture

Dans cette implémentation, les règles sont simplement codées sous forme d'expressions conditionnelles en C++. De plus, elles sont regroupées pour chaque type sémantique, qui adapte la présentation en fonction des différents paramètres.

Les modèles, les messages et les modalités sont implémentés par des classes C++. La dérivation est utilisée pour créer de nouveaux messages ou de nouvelles modalités. L'accès aux paramètres des messages et des modalités se fait par des *requêtes*, chacune d'elles nécessitant un nombre variable de paramètres. Ainsi, si un objet reçoit une requête à laquelle il est incapable de répondre, il la fait suivre à son parent dans la hiérarchie. Cette approche, similaire au passage de message de Smalltalk, a été utilisée pour circonvenir le problème du typage strict du C++, et afin de faciliter la future intégration de ces classes dans un système expert.

8 Conclusion

Nous avons présenté dans ce rapport les problèmes liés à la synthèse de présentations multimodales.

Nous avons d'abord abordé les difficultés liées à la modélisation des informations et des fonctions dans lesquelles celles-ci sont impliquées.

Ensuite, nous avons présenté une taxinomie des modalités afin de caractériser leurs propriétés expressives.

Puis, nous avons combiné les connaissances acquises de manière à produire une présentation multimodale qui reflète le sens du dialogue, et qui soit adaptée à l'utilisateur et à l'environnement.

Enfin, nous ébauchons l'intégration de ces résultats dans un système pour la mise en oeuvre d'un dialogue multimodal. La représentation sous forme de table est encore exploratoire et l'outil que nous avons crée doit permettre de mesurer la qualité de nos hypothèses.

Toutefois, il reste à approfondir bon nombre de problèmes. Nous n'avons fait qu'une évaluation de ces problèmes et une analyse des solutions possibles. L'agenda des recherches nécessaires à la réalisation d'un système complet pour la mise en oeuvre d'un dialogue homme-machine multimodal reste très chargé.

Il faut d'abord valider la taxinomie de l'information et des actes de dialogue, en intégrant l'analyse de nombreux domaines d'application et en s'inspirant des travaux issus de l'ergonomie et du traitement du langage. On pourra notament se reporter aux développements dans le domaine de l'hypermedia, où la représentation et la manipulation des informations sont très importantes (voir [CACM 38(8), 1995]).

Parallèlement, l'évaluation et la validation des techniques de présentation devront être entreprises. Ainsi, la gestion de l'interaction avec l'utilisateur doit être soigneusement intégrée afin de se rapprocher d'un véritable dialogue. Par exemple, la gestion de l'interruption des présentations (par l'utilisateur ou par d'autres présentations) pose de nombreux problèmes, notamment pour la granularité de cette interruption et la gestion de la reprise.

Il faudra enfin choisir un formalisme pour la représentation des connaissances et les mécanismes du raisonnement sur celles-ci. Une partie importante de cette tâche consistera en l'élaboration du langage unifié abordé dans cette étude.

Toutes ces connaissances et ces systèmes devront être intégrés dans une *boîte à outils* pour la réalisation d'interfaces multimodales, en s'inspirant éventuellement des travaux de Bernsen qui propose un outil logiciel pour l'évaluation de modalités et l'attribution de ces modalités à des types d'information(voir [Bernsen, 1995b]).

Références

[André et al., 1993] André, E., Finkler, W., Graf, W., Rist, T., Schauder, A., and Wahlster, W. (1993). WIP: The Automatic Synthesis of Multimodal Presentations. In [Maybury, 1993a], pages 75–93.

[André and Rist, 1993] André, E. and Rist, T. (1993). The Design of Illustrated Documents as a Planning Task. In [Maybury, 1993a], pages 94–116.

[André and Rist, 1994] André, E. and Rist, T. (1994). Multimedia Presentations: The Support of Passive and Active Viewing. In *AAAI Spring Symposium on Intelligent Multi-Media Multi-modal Systems*, pages 22–29.

[Arens et al., 1993a] Arens, Y., Hovy, E., and van Mulken, S. (1993a). Structure and Rules in Automated Multimedia Presentation Planning. In *IJCAI'93, the Thirteenth International Joint Conference on Artificial Intelligence*, pages 1253–1259.

[Arens et al., 1993b] Arens, Y., Hovy, E., and Vossers, M. (1993b). On the Knowledge Underlying Multimedia Presentations. In [Maybury, 1993a], pages 280–306.

[Beaudouin-Lafon, 1993] Beaudouin-Lafon, M. (1993). L'Usage de Capteurs de Contexte dans les Systèmes Interactifs. In *IHM'93, Cinquièmes Journées sur l'Ingénierie des Interfaces Homme-Machine.*

[Beaudouin-Lafon, 1994] Beaudouin-Lafon, M. (1994). L'usage du son dans les systèmes interactifs. In *Journées d'Informatique Musicale.*

[Bederson and Hollan, 1995] Bederson, B. B. and Hollan, J. D. (1995). Pad++: A Zoomable Graphical Interface System. In *CHI'95, Conference on Human Factors in Computing Systems.*

[Bellik, 1995] Bellik, Y. (1995). *Interface multimodales : concepts, modèles et architectures.* PhD thesis, Université Paris-Sud, Orsay.

[Bellik et al., 1995] Bellik, Y., Ferrari, S., Néel, F., Teil, D., Pierre, E., and Tachoires, V. (1995). Interaction Multimodale : Concepts et Architecture. In *4ème Journées Internationales L'Interface des Mondes Réels et Virtuels*, pages 37–45.

[Bellik et al., 1994] Bellik, Y., Ferrari, S., Néel, F., Teil, D., and Tachoires, V. (1994). Communication Multimodale. Rapport Final Contrat DRET No 92/358.

[Bellik and Teil, 1992] Bellik, Y. and Teil, D. (1992). Définitions Terminologiques pour la Communication Multimodale. In *IHM'92, Quatrièmes Journées sur l'Ingénierie des Interfaces Homme-Machine.*

[Bernsen, 1994] Bernsen, N. O. (1994). Modality theory in support of multimodal interface design. In *AAAI Spring Symposium on Intelligent Multi-Media Multi-modal Systems*, pages 37–44.

[Bernsen, 1995a] Bernsen, N. O. (1995a). Foundations of Multimodal Representations. Amodeus, ESPRIT BRA 7040. delivrable 14, Section B.

[Bernsen, 1995b] Bernsen, N. O. (1995b). The taxonomy workbench : A multimedia database system for analysing representational modalities. WPCS-95-2, Working Papers in Cognitive Science and HCI, Centre for Cognitive Science, Roskilde University.

[Blandford and Young, 1993] Blandford, A. and Young, R. M. (1993). Developing runnable user models: Separating the problem solving techniques from the domain knowledge. Amodeus, ESPRIT BRA 7040.

[Bourdot et al., 1995] Bourdot, P., Krus, M., and Gherbi, R. (1995). Management of Non-Standard Devices for Multimodal User Interfaces under UNIX / X11. In *CMC'95, International Conference on Cooperative Multimodal Communication.*

[CACM 38(8), 1995] CACM 38(8) (1995). Designing hypermedia applications. Communications of the ACM, 38(8).

[Feiner et al., 1993] Feiner, S. K., Litman, D. J., McKeown, K. R., and Passonneau, R. J. (1993). Towards Coordinated Temporal Multimedia Presentations. In [Maybury, 1993a], pages 139–147.

[Feiner and McKeown, 1993] Feiner, S. K. and McKeown, K. R. (1993). Automating the Generation of Coordinated Multimedia Explanations. In [Maybury, 1993a], pages 117–138.

[Ferrari, 1994] Ferrari, S. (1994). Modèle de la tâche dans le cadre d'une interface de dialogue multimodal. In *IHM'94, Sixième Journées sur l'Ingénierie des Interfaces Homme-Machine.*

[Gargan et al., 1988] Gargan, R. A., Sullivan, J. W., and Tyler, S. W. (1988). Multimodal Response Planning: An Adaptive Rule Based Approach. In *CHI'88 Conference on Human Factors in Computing Systems*, pages 229–234.

[Goodman, 1993] Goodman, B. (1993). Multimedia Explanations for Intelligent Training Systems. In [Maybury, 1993a], pages 148–171.

[Han and Zukerman, 1995] Han, Y. and Zukerman, I. (1995). A Cooperative Approach for Multimodal Presentation Planning. In *CMC'95, International Conference on Cooperative Multimodal Communication*, pages 145–159.

[Harrison et al., 1995] Harrison, B. L., Ishii, H., Vicente, K. J., and Buxton, W. A. (1995). Transparent layered user interfaces: An evaluation of a display design to enhance focused and divided attention. In *CHI'95, Conference on Human Factors in Computing Systems.*

[Kolski, 1993] Kolski, C. (1993). *Ingéniérie des interfaces homme-machine. Conception et valuation.* Hermès.

[Little et al., 1991] Little, T., Ghafoor, A., Chen, C., C.S., C., and Berra, P. (1991). Multimedia Synchronization. *IEEE Data Engineering Bulletin*, 14(3):26–35.

[Martin et al., 1995] Martin, J., Veldman, R., and Béroule, D. (1995). Towards Adequate Representation Technologies for Multimodal Interfaces. In *CMC'95, International Conference on Cooperative Multimodal Communication*, pages 207–224.

[Maybury, 1993a] Maybury, M. (1993a). *Intelligent Multimedia Interfaces.* AAAI Press, California.

[Maybury, 1993b] Maybury, M. (1993b). Planning Multimedia Explanations Using Communicative Acts. In [Maybury, 1993a], pages 59–74.

[Maybury and Hovy, 1994] Maybury, M. and Hovy, E. (1994). Intelligent multimedia interfaces. In *Tutorial notes for AAAI Spring Symposium on Intelligent Multi-Media Multi-modal Systems.*

[Mynatt and Edwards, 1992] Mynatt, E. and Edwards, W. K. (1992). Mapping GUIs to auditory interfaces. Technical Report 92-08, Graphics, Visualization and Usability Centre, Georgia Institute of Technology, USA.

[Roth and Hefley, 1993] Roth, S. F. and Hefley, W. E. (1993). Intelligent Multimedia Presentation Systems : Research and Principles. In [Maybury, 1993a], pages 13–58.

[Sutcliffe and Faraday, 1994] Sutcliffe, A. and Faraday, P. (1994). Designing Presentation in Multimedia Interfaces. In *CHI' 94, Conference on Human Factors in Computing Systems*, pages 92–98.

[Werkman, 1994] Werkman, K. J. (1994). A DAI Architecture for Coordinating Multimedia Applications. In *AAAI Spring Symposium on Intelligent Multi-Media Multi-modal Systems*, pages 93–97.