

# Un modèle conceptuel pour une présentation multimodale et contextuelle de l'information

## *A conceptual model for a multimodal and contextual presentation of information*

Cyril ROUSSEAU (1), Yacine BELLIK (1), Frédéric VERNIER (1)

(1) LIMSI-CNRS, Université Paris-Sud XI, B.P. 133, 91403, Orsay, France  
cyril.rousseau@limsi.fr, yacine.bellik@limsi.fr, frederic.vernier@limsi.fr

**Résumé.** La contextualisation de l'interaction Homme-Machine en sortie vise à adapter la présentation de l'information au contexte d'interaction. Cette adaptation requiert la mise en place de nouveaux mécanismes intégrant le contexte dans les processus de conception et d'évolution d'une présentation multimodale. Cet article propose un modèle conceptuel intitulé WWHT permettant une présentation dynamique et intelligente de l'information. Ce modèle s'appuie sur quatre concepts fondamentaux : "What", "Which", "How" et "Then", annotant les événements marquants du cycle de vie d'une présentation multimodale. Inspirée de ce modèle, une plateforme logicielle permettant une conception itérative (spécification, simulation et exécution) des sorties d'un système multimodal est alors présentée. Enfin, nous illustrons l'application du modèle à travers la spécification des sorties de deux systèmes : le premier se rapportant au domaine de la téléphonie mobile et le second à celui du contrôle aérien.

**Mots-clés.** Interaction Homme-Machine, multimodalité en sortie, présentation multimodale, contexte d'interaction.

**Abstract.** The contextualization of the output user interface aims at adapting the information presentation to the context of interaction. This adaptation requires new concepts including the interaction context in the design and evolution processes of a multimodal presentation. This paper presents a conceptual model called WWHT for a dynamic and intelligent multimodal presentation of information. The model is based on four concepts: "What", "Which", "How" and "Then", that describe the steps of the life cycle of multimodal presentations. On the basis of these four concepts, we then propose a software platform that supports an iterative design (specification, simulation and execution) for output multimodal systems. Finally, we illustrate the model with the output presentation design of two applications: a mobile phone simulator and an air traffic control system.

**Keywords.** Human-Computer Interaction, output multimodality, multimodal presentation, interaction context.

## 1 Introduction

Aujourd'hui, les utilisateurs de systèmes informatiques forment une clientèle hétéroclite, allant du jeune adolescent à l'adulte, en passant par les personnes âgées, sans oublier les personnes handicapées. L'informatique est enseignée dès le plus jeune âge comme une matière à part entière. La gestion électronique d'informations personnelles (déclaration de revenus, gestion de comptes bancaires, etc.) destinée à l'ensemble de la population n'est déjà plus au rang des nouveautés technologiques. Cet engouement pour les nouvelles technologies est d'autant plus perceptible si l'on s'attarde sur la gamme des périphériques portables grand public. Cette gamme ne cesse de s'enrichir au point qu'il en devient difficile de recenser l'offre dans son intégralité : ordinateur portable, agenda électronique, téléphone mobile, lecteur multimédia portable, baladeur mp3, console de jeux, etc. La miniaturisation des composants électroniques, la baisse des coûts de production et les connexions réseaux sont en grande partie à l'origine de leur intégration dans notre vie quotidienne. La mobilité conférée par ces périphériques participe à leur popularité. De nouveaux usages jusqu'ici insoupçonnés émergent. Il est devenu commun de téléphoner dans la rue, de travailler dans les transports en commun, ou bien de consulter son courrier électronique à la terrasse d'un café. L'environnement d'interaction jusqu'ici réduit à un espace fixe et fermé s'est brusquement transformé en un espace dynamique et ouvert.

Cette diversification des utilisateurs, des systèmes et des environnements d'interaction entraîne une complexification du contexte d'interaction (Savidis et Stephanidis, 2004). L'interaction doit s'adapter au contexte afin de préserver sa pertinence. En sortie, cette contextualisation de l'interaction requiert la mise en place de nouveaux concepts pour une présentation multimodale de l'information en adéquation avec le contexte courant. La multimodalité est alors exploitée comme un moyen d'adaptation de l'interaction Homme-Machine.

Parmi les modèles conceptuels, SRM (Standard Reference Model) de Bordegoni *et al.* (1997) est l'un des premiers à aborder le problème. Il est ensuite complété pour intégrer le contexte d'interaction dans la conception initiale d'une présentation multimodale (Stephanidis *et al.*, 1997). Cette intégration reste toutefois partielle et ne couvre pas l'intégralité du processus de conception d'une présentation multimodale de l'information. Concernant l'évolution dynamique de l'interaction Homme-Machine, de nouveaux concepts tel que la plasticité (Thevenin et Coutaz, 1999) (Vanderdonck *et al.*, 2005) ont été introduits. Néanmoins, les facteurs d'évolution ne sont pas encore clairement identifiés.

Cet article présente un modèle conceptuel intégrant le contexte d'interaction dans chacune des phases du cycle de vie d'une présentation multimodale. Ce modèle intitulé WWHT, s'appuie sur quatre concepts fondamentaux : « What », « Which », « How », « Then », traitant les problématiques de conception et d'évolution d'une présentation multimodale en adéquation avec le contexte d'interaction. Sur les bases de ce modèle, nous présentons ELOQUENCE, une plateforme logicielle permettant une conception incrémentale et itérative des sorties d'un système multimodal. ELOQUENCE propose un ensemble d'outils pour la spécification, la simulation et l'exécution d'un système multimodal en sortie. Nous illustrons ces travaux par deux applications, la première, en téléphonie mobile sert de fil conducteur, tandis que la seconde, sur le contrôle aérien, nous permet de résumer les concepts introduits.

## 2 Le modèle conceptuel WWHT

Dans cette section, nous introduisons les composantes de la multimodalité en sortie nécessaires à la conception et mise en oeuvre d'une présentation multimodale intelligente de l'information. Ces composantes identifiées, nous présentons ensuite succinctement le modèle conceptuel WWHT, avant de reprendre en détails ses concepts fondamentaux.

### 2.1 Les composantes de la multimodalité en sortie

Nous pensons qu'un système multimodal en sortie ne peut être réduit à un système qui se limite simplement à faire cohabiter plusieurs modalités en sortie. Dans ce cas, nous préférons parler de système multimédia (Coutaz, 1992). Un système multimodal en sortie doit être capable de présenter une information de façon "intelligente" en exploitant ses différentes modalités de communication. Selon les systèmes existants (Kolski *et al.*, 1992), ce concept d'intelligence peut varier. Cependant, tous les systèmes existants partagent le même but : *la présentation de l'information doit être adaptée au contexte d'interaction*. Ce processus de présentation intelligente de l'information s'appuie sur quatre composantes que nous allons détailler au cours des sections suivantes :

- l'information à présenter,
- les composants d'interaction du système,
- le contexte d'interaction,
- la présentation multimodale résultante.

#### ***L'information***

Selon le modèle ARCH (Bass *et al.*, 1992), l'information est l'objet de nature sémantique (en provenance du noyau fonctionnel et transitant par le contrôleur de dialogue) que le système en charge des sorties doit présenter à l'utilisateur. Par exemple en téléphonie mobile, la « réception d'un nouvel appel » ou encore l'indication d'un « niveau de batterie faible » sont deux unités d'information que le système en charge des sorties d'un téléphone portable doit savoir exprimer.

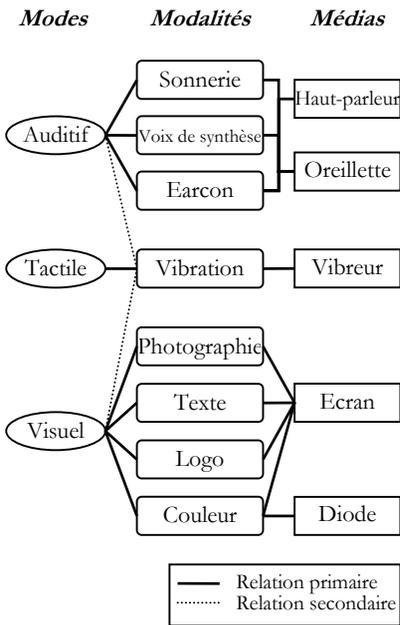
#### ***Les composants d'interaction***

Nous définissons les composants d'interaction d'une application comme les moyens de communication (physiques ou logiques) entre l'homme et la machine. Nous distinguons trois types de composants d'interaction : le *mode*, la *modalité* et le *média*. Selon les auteurs, ces notions peuvent avoir des définitions différentes : (Frohlich, 1991), (Nigay et Coutaz, 1995), (Bordegoni *et al.*, 1997), (Martin, 1998).

Dans notre cas, nos définitions sont orientées utilisateur (Bellik, 1995). Un mode en sortie correspond à un système sensoriel de l'utilisateur (Visuel, Auditif, Tactile, etc.). Une modalité en sortie est définie par la structure d'information telle qu'elle est perçue par l'utilisateur (Texte, Sonnerie, Vibration, etc.) et non pas telle qu'elle est définie au sein de la machine. Enfin un média en sortie est un périphérique de sortie permettant l'expression d'une modalité en sortie (Ecran, Haut-parleur, Vibreur, etc.). Comme l'illustre la Figure 1, les notions de mode, modalité et média en sortie sont interdépendantes : à un mode correspond un ensemble de modalités et à une modalité est rattaché un ensemble de médias permettant son expression. La modalité « Vibration » s'exprime par exemple sur le média « Vibreur » et fait appel au mode « Tactile ». L'ensemble de ces relations tisse un diagramme des composants d'interaction.

Dans ce diagramme, il convient également de distinguer deux types de relations entre les modes et les modalités : principales et secondaires. Une relation principale est définie entre un mode A et une modalité B, si le système sensoriel

correspondant au mode A est celui qui est communément sollicité par les utilisateurs pour percevoir la modalité B. La liaison secondaire résulte quant à elle d'un effet de bord de la relation principale. Dans notre exemple, les vibrations d'un téléphone portable sont destinées à être perçues par l'utilisateur de façon tactile, ce qui se traduit par une relation principale entre le mode « Tactile » et la modalité « Vibration ». Cependant elles peuvent également être entendues, ce qui se traduit par une liaison de type secondaire entre le mode « Auditif » et la modalité « Vibration ».



Critère	Valeurs	Modèle
Personne malvoyante	Sur une échelle de 1 à 10	Utilisateur
Personne malentendante	Sur une échelle de 1 à 10	Utilisateur
Disponibilité de l'écran	Disponible, Indisponible	Système
Disponibilité du Haut-parleur	Disponible, Indisponible	Système
Disponibilité du vibreur	Disponible, Indisponible	Système
Occupation des médias audio	Libre, Occupé	Système
Mode du téléphone	Renforcé, Normal, Silencieux	Système
Niveau de batterie	0-100	Système
Lieu	Cinéma, Rue, Parc, etc.	Environnement
Niveau de bruit	0-130	Environnement

Figure 1. Diagramme des composants d'interaction

Tableau 1. Contexte d'interaction

**Le contexte d'interaction**

Selon Dey et al. (2001), « Le contexte est toute information pouvant être utilisée pour caractériser la situation d'une entité. Une entité est une personne ou un objet considéré comme pertinent pour l'interaction entre l'utilisateur et l'application ».

Nous utilisons une approche à base de modèles (Arens et Hovy, 1995) pour spécifier ces entités (modèle du système, modèle de l'utilisateur, modèle de l'environnement, etc.). A chaque modèle, est associé un ensemble de critères dynamiques ou statiques relatifs à l'entité considérée (disponibilité des médias, préférences de l'utilisateur, niveau de bruit, etc.). Le Tableau 1 présente un exemple de modélisation du contexte d'interaction pour une application téléphonique.

**La présentation multimodale**

L'expression d'une information se fait à l'aide d'une présentation multimodale. Cette présentation se compose d'un ensemble de couples (modalité, média) liés par des propriétés de redondance et/ou de complémentarité (Coutaz et al., 1995). La réception d'un appel sur un téléphone portable peut par exemple s'exprimer par une présentation multimodale se composant de deux couples (modalité, média) : un premier couple (Sonnerie, Haut-parleur) signale l'arrivée d'un appel, un second (Texte, Ecran) présente le numéro de la personne à l'origine de l'appel.

## 2.2 Un aperçu du modèle

A partir des quatre composantes (information à présenter, composants d'interaction, contexte d'interaction et présentation multimodale), nous proposons un modèle conceptuel décrivant le cycle de vie d'une présentation multimodale. Ce modèle intitulé WWHT s'appuie sur quatre questions fondamentales :

- *What* : Quelle information présenter ?
- *Which* : Quelle présentation multimodale choisir ?
- *How* : Comment instancier cette présentation ?
- *Then* : Comment faire évoluer cette présentation ?

Les trois premiers concepts (What, Which et How) font référence à la phase de construction d'une présentation multimodale (Bordegoni *et al.*, 1997) tandis que le dernier concept (Then) s'intéresse au devenir de la présentation (Calvary *et al.*, 2003). La Figure 2 et les sections suivantes présentent le processus de conception d'une présentation multimodale adaptée au contexte d'interaction. Le devenir de la présentation est traité dans la section 6.

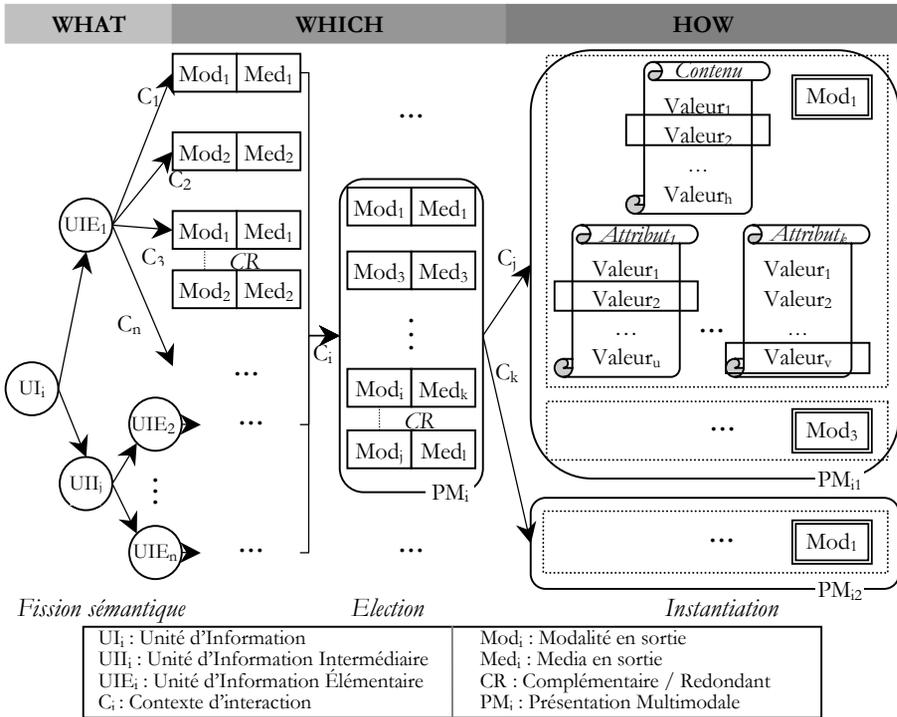


Figure 2. Conception d'une présentation multimodale adaptée au contexte d'interaction

## 3 What : Quelle information présenter ?

En premier lieu, il est nécessaire de décomposer l'information sémantique (Figure 2, UI) en unités d'information élémentaires (Figure 2, UIE<sub>i</sub>) à présenter à l'utilisateur. Cette décomposition formalise la communication entre le contrôleur de dialogue (au sens de ARCH) et le module en sortie, tout en limitant la complexité du processus d'expression d'une information. Prenons l'exemple de la réception d'un appel sur un téléphone portable. L'unité d'information « appel en provenance

de X » est décomposable en deux unités d'information élémentaire : l'évènement (appel) et l'identité de l'appelant (X).

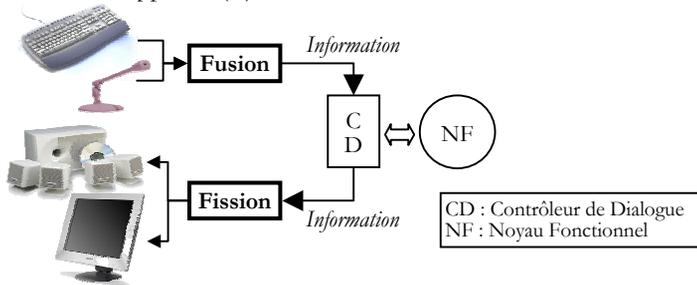


Figure 3. Le parallèle Fusion / Fission

Le terme « fission » est communément employé par opposition au terme « fusion » (Figure 3) pour qualifier le processus de sélection des modalités en sortie. Il ne nous semble pas pertinent d'utiliser ce terme pour désigner un processus de sélection. Une fission en sortie doit effectivement avoir lieu mais à un niveau sémantique. C'est pourquoi nous préférons parler de fission sémantique lors de la décomposition d'une unité d'information en unités d'information élémentaire et d'allocation lors de la sélection des modalités.

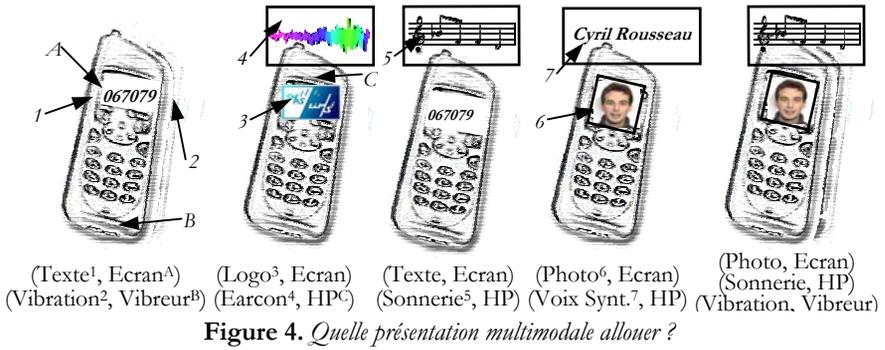
La fission sémantique extrait les informations élémentaires à présenter à l'utilisateur. Cette extraction peut s'accompagner d'une perte relative à la structure syntaxique de l'information. En contrepartie, la décomposition en informations élémentaires autorise une « personnalisation » de la présentation de l'information. Les composants d'interaction alloués sont alors dédiés à une ou plusieurs informations élémentaires facilitant leur perception par l'utilisateur.

Généralement, la fission sémantique est réalisée de façon manuelle par les concepteurs de l'application lors de la spécification des sorties. Son automatisation nécessite des mécanismes d'analyse sémantique de l'information, rendant le problème particulièrement délicat. Le sujet reste néanmoins intéressant comme perspective de recherche à long terme.

#### 4 Which : Quelles présentations multimodales choisir ?

Suite à la décomposition de l'information, une présentation doit être allouée pour celle-ci (Karagiannidis *et al.*, 1997). L'allocation consiste à sélectionner pour chaque unité d'information élémentaire, une présentation multimodale (Figure 2, [Mod., Med;]) adaptée à l'état courant du contexte d'interaction (Figure 2, C<sub>i</sub>) puis à les regrouper en une même présentation (Figure 2, PM<sub>i</sub>) symbolisant l'expression de l'unité d'information initiale. La présentation résultante se compose alors d'un ensemble de couples (modalité, média) liés par des propriétés de redondance et/ou de complémentarité.

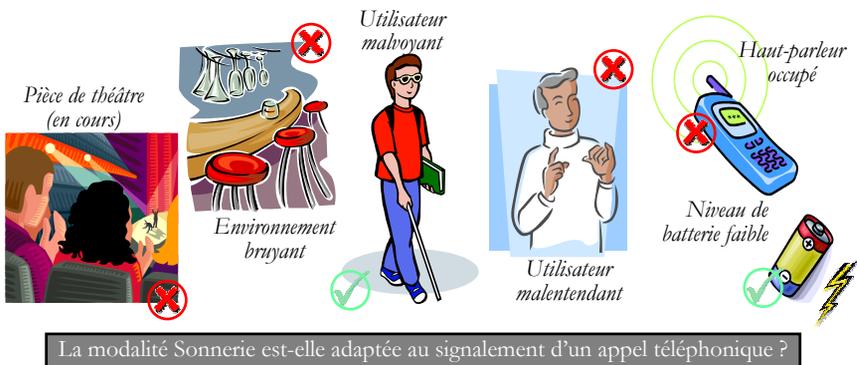
Cette phase d'allocation ne se résume donc pas à une simple recherche de modalités en sortie mais bien à l'identification de couples (modalité, média) adaptés au contexte d'interaction courant. Le contexte d'interaction pris dans sa globalité apporte alors une nouvelle dimension à notre problématique. Il s'agit de prendre en compte une description donnée de l'état du contexte comme facteur de décision dans le choix du ou des couples (modalité, média) à déployer. Ceci induit une complexité non négligeable, en particulier dans le cadre d'applications à fort potentiel de communication ou d'un contexte d'interaction sujet à évolution.



Reprenons l'exemple de la réception d'un appel téléphonique. Le téléphone portable est un parfait sujet d'application disposant d'un grand nombre de composants d'interaction. Cette capacité de communication en sortie accentue le nombre de présentations multimodales envisageables pour exprimer l'information (Figure 4). Ainsi, pour un contexte d'interaction donné, l'information initiale de réception d'un appel téléphonique s'exprimerait par exemple à travers une présentation multimodale composée de deux couples (modalité, média) : un premier couple (Sonnerie, Haut-parleur) signale l'arrivée de l'appel (première UIE) tandis qu'un second (Texte, Écran) présente le numéro de l'appelant (seconde UIE).

#### 4.1 Le modèle comportemental

Le processus de sélection des composants d'interaction en fonction du contexte s'appuie sur un modèle comportemental. Le modèle comportemental spécifie les composants d'interaction (modes, modalités et médias) adaptés à un état donné du contexte. Poursuivons l'exemple de la réception d'un appel sur un téléphone portable. Six états du contexte d'interaction variant selon l'environnement (théâtre / bar), l'utilisateur (malvoyant / malentendant) ou le système (Haut-parleur occupé / niveau de batterie faible) sont présentés en Figure 5. Le modèle comportemental identifie alors les composants d'interaction compatibles avec ces états de contexte. Selon ce modèle, la modalité Sonnerie est inadaptée dans quatre des six contextes déclarés, pour des raisons d'éthique (théâtre), de perception (bar et malentendant) ou d'indisponibilité d'un composant d'interaction (Haut-parleur occupé).



**Figure 5.** *La catégorisation d'un composant d'interaction selon le contexte*

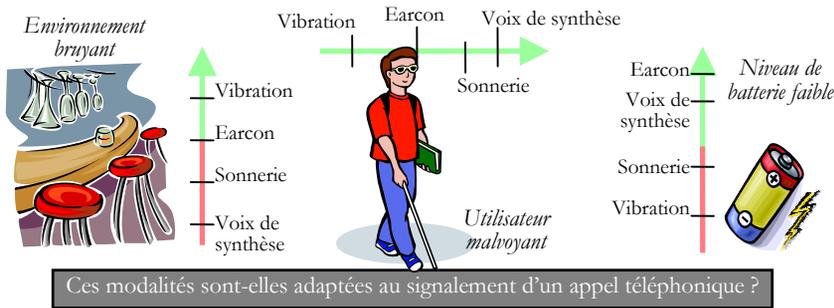


Figure 6. La classification des composants d'interaction selon le contexte

**Le poids contextuel**

Le modèle comportemental ne se contente pas de catégoriser (adapté / inadapté) les composants d'interaction de l'application. Ce dernier les classe selon leur poids contextuel, une valeur entière (positive ou négative) qui dénote le niveau d'adéquation du composant au contexte courant. La Figure 6 présente un exemple de classification de modalités dans trois contextes pour le signalement d'un appel téléphonique.

**Expression du modèle comportemental**

La formalisation du modèle comportemental se réalise de diverses façons selon les systèmes : règles (Stephanidis *et al.*, 1997), matrices (Duarte et Carriço, 2006), automates (Johnston et Bangalore, 2005), réseau de Pétri (Navarre *et al.*, 2005), etc. Dans nos travaux, nous avons choisi une formalisation à base de règles (Rousseau *et al.*, 2004). Ce formalisme a l'avantage de proposer un raisonnement simple et intuitif, limitant le coût d'apprentissage du désigner. Ce choix soulève néanmoins des problèmes de cohérence, de complétude et d'évolutivité du modèle comportemental, pour lesquels des outils de vérification et de visualisation de la base de règles doivent être proposés (section 7).

Une règle du modèle comportemental se compose de prémisses décrivant un état du contexte d'interaction (exemple : un niveau de bruit supérieur à 90 dB) et de conclusions spécifiant en fonction de ces prémisses la structure de la présentation (exemple : utiliser la modalité Vibration). La forme des conclusions détermine la typologie de la règle. Trois types de règles sont ainsi distingués :

- la règle contextuelle,
- la règle « critérielle »,
- la règle de composition.

La règle contextuelle se compose de conclusions ayant pour cible les composants d'interaction du système. Celle-ci a en effet pour fonction de spécifier à travers ses conclusions le poids contextuel (l'intérêt vis-à-vis du contexte décrit en prémisses) des composants d'interaction cibles de la règle. Le poids contextuel se traduit par une note pouvant être positive comme négative attribuée au composant. Cette évaluation du composant est strictement locale (propre à la règle) et peut être par la suite complétée par les conclusions d'autres règles contextuelles.

Plutôt que de cibler une modalité donnée, la règle critérielle permet de favoriser ou de défavoriser toutes modalités répondant à un critère donné (langagier, analogique (Bernsen, 1996), spatialité, confidentialité, etc.). Elle permet ainsi une sorte de factorisation des règles avec pour conséquence positive une diminution du nombre de règles à spécifier.

	Description en langage naturel	Type
R1	Si l'utilisateur est malvoyant <b>Alors</b> ne pas utiliser le mode Visuel	Contextuel
R2	Si le Haut-parleur est occupé <b>Alors</b> ne pas utiliser la modalité Sonnerie	Contextuel
R3	Si le niveau sonore est supérieur à 90 dB <b>Alors</b> ne pas utiliser le média Haut-parleur	Contextuel
R4	Si l'UIE est un « appel » <b>Alors</b> favoriser la modalité Vibration <b>Et</b> recommander la modalité Sonnerie	Contextuel
R5	Si l'UIE est un « appel » <b>Et</b> le téléphone est en mode renforcé <b>Alors</b> utiliser une redondance	Composition
R6	Si l'UIE est un « appelant » <b>Alors</b> favoriser les modalités fortement Analogiques	Critériel
R7	Si l'UI est un « appel de X » <b>Et</b> le niveau de batterie est faible <b>Alors</b> éviter le média Vibreur	Contextuel

**Tableau 2.** Exemples de règles du modèle comportemental

Enfin, la règle de composition dispose de conclusions relatives aux propriétés de combinaison des modalités en sortie (Vernier et Nigay, 2000). Elle permet ainsi l'allocation de présentations composées de plusieurs couples (modalité, média) sur des critères de redondance et/ou de complémentarité.

Continuons l'exemple de la réception d'un appel sur un téléphone portable. Le Tableau 2 présente sept règles dont : cinq de type contextuel (R1, R2, R3, R4 et R7), une de type composition (R5) et une de type critériel (R6).

Dans une situation nominale, seules les règles R4 et R6 sont appliquées pour présenter la réception d'un nouvel appel. La présentation est alors composée de deux paires : (Vibration, Vibreur) pour signaler l'évènement appel (première UIE) et (Photographie, Écran) pour présenter l'appelant (seconde UIE). Dans un contexte différent tel qu'un niveau de batterie faible, la règle R7 adapte la forme de la présentation en préférant la modalité Sonnerie à la modalité Vibration dans un souci d'économie d'énergie. De même, dans le cadre d'un téléphone en mode renforcé, la règle R5 intervient dans la présentation de l'UIE « Appel » et entraîne l'ajout du couple (Sonnerie, Haut-parleur) à la présentation initiale.

#### 4.2 Le processus d'élection

En utilisant une analogie avec le monde politique, nous appelons le processus de sélection des modalités : *élection*. Notre processus d'élection s'appuie sur un modèle comportemental (électeurs) formalisé par une base de règles qui une fois appliquées ajoutent ou retirent des points (voix) à certains modes, modalités ou médias de sortie (candidats) selon l'information à exprimer (thématique du scrutin) et l'état courant du contexte d'interaction (situation économique et sociale, etc.).

Le processus d'élection se divise en deux phases distinctes : *l'élection pure* et *l'élection composée* (Figure 7). L'élection pure choisit le meilleur couple (modalité, média) exprimant une unité d'information élémentaire donnée pour un état fixé du contexte d'interaction. L'élection composée enrichit la présentation en sélectionnant de nouveaux couples redondants ou complémentaires au premier. Les mécanismes et les concepts rattachés à chacune de ces phases sont explicités dans ce qui suit.

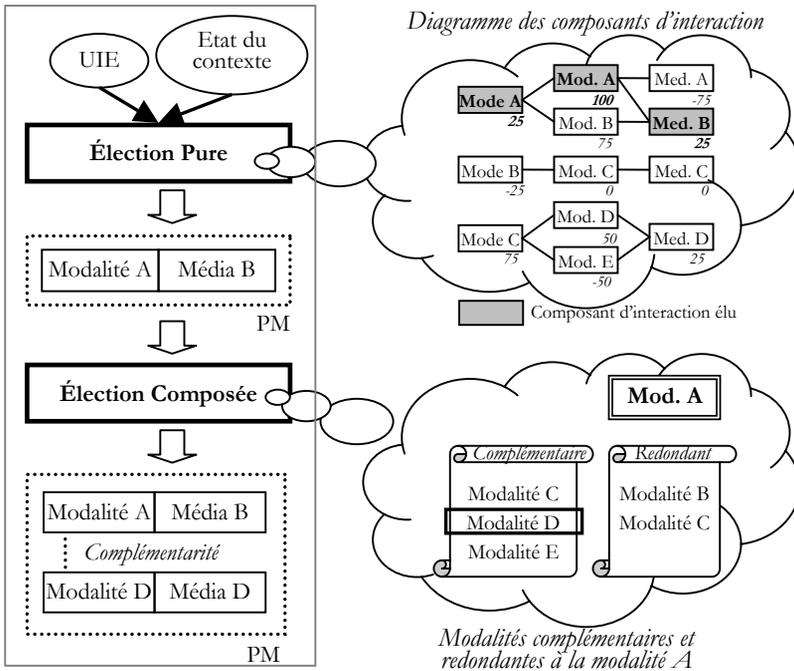


Figure 7. Le processus d'élection

**L'élection pure**

La première phase d'élection parcourt le diagramme des composants d'interaction à la recherche du meilleur triplet (mode, modalité, média). L'application du modèle comportemental (règles contextuelles et critérielles) définit le poids contextuel de chacun des composants d'interaction du système. Ce poids contextuel décrit l'adéquation du composant au contexte d'interaction courant et se positionne ainsi comme un critère de sélection. La Figure 8 présente l'algorithme d'élection employé. Notons qu'il est possible de pondérer l'effet des poids contextuels à l'aide de coefficients multiplicateurs s'appliquant de façon globale selon le type du composant d'interaction.

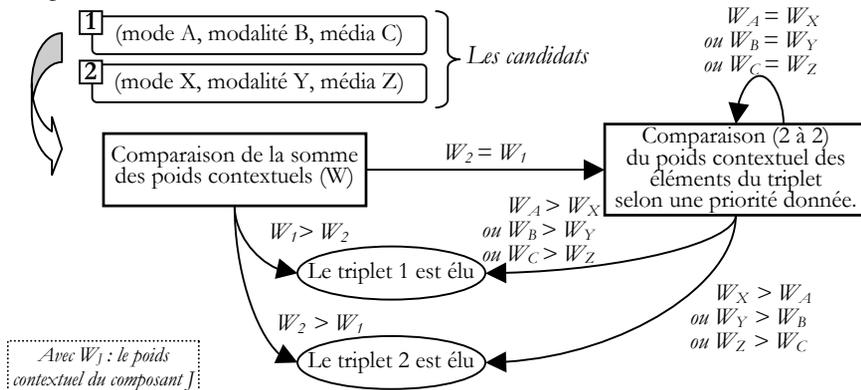


Figure 8. L'algorithme d'élection

Dans le cadre de la Figure 7, les deux triplets (mode A, modalité A, média B) et (mode C, modalité D, média D) se démarquent avec un poids contextuel de 150. Cette égalité entraîne une comparaison deux à deux des éléments du triplet en suivant l'ordre spécifié par le concepteur : modalité, mode, média. Cet ordre favorise la modalité au mode et permet ainsi l'élection du couple (modalité A, média B) au détriment du couple (modalité D, média D). Il est à noter que le triplet élu est réduit à un couple (modalité, média), le mode étant déductible.

### ***L'élection composée***

En l'état actuel, la présentation est composée d'un unique couple (modalité, média) issu de l'élection pure. L'élection composée poursuit la construction de la présentation en y ajoutant de nouveaux couples redondants ou complémentaires au premier (Coutaz *et al.*, 1995). Le modèle comportemental (les règles de composition) spécifie selon l'information à exprimer et le contexte d'interaction courant, la structure (le nombre de couples) et la typologie des liaisons (redondance / complémentarité) de la présentation. Il est à noter que cette spécification est une recommandation qui selon les cas pourra être partiellement ou totalement suivie.

Les couples candidats à l'élection composée sont déduits de la première modalité élue. Cette dernière dispose d'une description des modalités potentiellement redondantes et complémentaires. Une classification selon le poids contextuel des couples issus de cette description complète la phase d'élection. Les couples de plus fort poids contextuel et répondant à la typologie demandée sont élus et ajoutés à la présentation.

La Figure 7 illustre cette seconde phase en y listant les modalités redondantes et complémentaires à la modalité A (première modalité élue). Dans le cadre de notre exemple, la présentation « doit » se composer d'un deuxième couple (modalité, média) complémentaire au premier. Trois modalités complémentaires à la modalité A sont candidates : les modalités C, D et E. Par comparaison du poids contextuel des triplets respectifs, il s'avère que le triplet (mode C, modalité D, média D) dispose du meilleur poids. C'est pourquoi, la présentation finale se compose des couples (modalité A, média B) et (modalité D, média D).

## **5 How : Comment instancier cette présentation ?**

Une fois la présentation calculée, celle-ci est instanciée. L'instanciation consiste à déterminer, selon l'état du contexte d'interaction (Figure 2, C<sub>i</sub>) les contenus lexico-syntaxiques et les attributs morphologiques des modalités de la présentation (Figure 2, PM<sub>i</sub>). Dans un premier temps, un contenu concret à exprimer à travers les modalités de la présentation est choisi (Colineau *et al.*, 2004). Puis en cohérence avec ce contenu, les attributs de la présentation (attributs des modalités (Layāida *et al.*, 2005), paramètres spatiaux (Graf, 1995) et temporels (Dalal *et al.*, 1996)) sont fixés.

Cette phase du modèle aborde un sujet particulièrement délicat se référant à la génération multimodale. Les choix à réaliser au niveau du contenu et des attributs de la présentation sont nombreux et doivent s'effectuer en étroite collaboration sous peine d'incompatibilités. Parallèlement à cette première difficulté, l'état courant du contexte d'interaction doit être pris en considération afin de préserver l'adaptabilité de la présentation élue. Tout ceci esquisse une problématique complexe dans laquelle nous allons tâcher de mettre en place de nouveaux concepts mêlant le contexte d'interaction au processus d'instanciation d'une présentation multimodale.

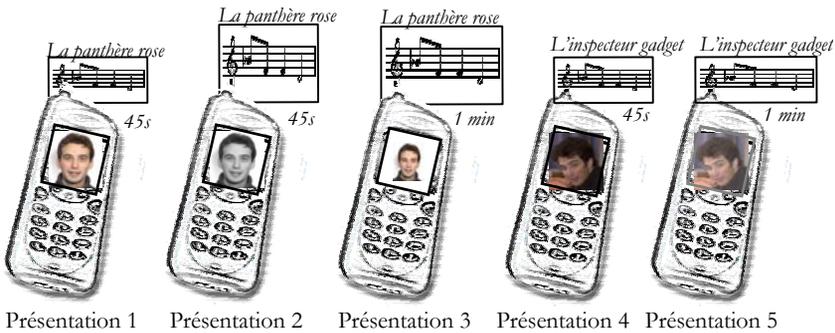


Figure 9. *Quelle présentation multimodale instancier ?*

Reprenons l'exemple de la réception d'un appel sur un téléphone portable. Supposons que la présentation soit constituée de deux couples (modalité, média) : un premier couple (Sonnerie, Haut-parleur) signale l'appel (première UIE) tandis que le second (Photographie, Écran) introduit la personne à l'origine de l'appel (deuxième UIE). Une des instanciations du contenu concret de cette présentation pourrait être la musique de la panthère rose (modalité Sonnerie) et un portrait (modalité Photographie) de la personne à l'origine de l'appel (Figure 9, Présentations 1, 2 et 3). L'instanciation des attributs morphologiques de la présentation pourrait être une durée de 45 secondes associées à un volume moyen pour la sonnerie et l'affichage en couleur et en plein écran du portrait (Figure 9, Présentation 1).

### 5.1 La sélection du contenu

Dans l'idéal, la génération du contenu devrait être automatique mais ceci dépasse très largement le cadre de nos recherches. La génération automatique de contenu est un problème à part entière pour chaque modalité considérée, et constitue le sujet de nombreuses recherches telles que la génération de langage naturel (André, 2003) ou la synthèse de gestes (Brafport *et al.*, 2004). Notre problématique s'établit sur la sélection d'un contenu de modalité ajusté aux informations élémentaires à exprimer, formaté au média associé et adapté au contexte d'interaction courant. La Figure 10 illustre cette problématique dans le cadre de la modalité Photographie. Il s'agit de sélectionner l'un des portraits référencés afin de présenter la personne à l'origine d'un appel téléphonique. Chacun des portraits proposés est associé à un numéro de téléphone donné (travail, domicile et portable), connotant la provenance de l'appel.

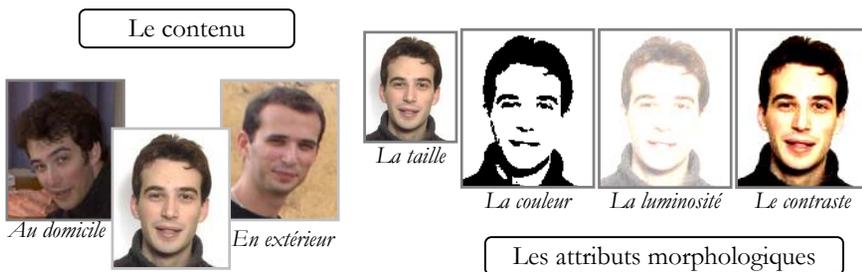


Figure 10. *Quels contenu et attributs pour la modalité Photographie ?*

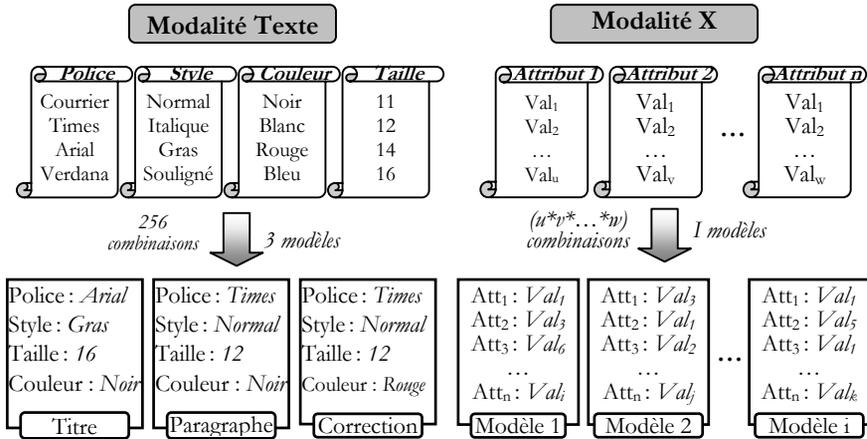


Figure 11. Les modèles d'instanciation des attributs de modalités

## 5.2 L'instanciation des attributs

La Figure 10 s'intéresse également aux attributs morphologiques à utiliser pour instancier la modalité Photographie (Rist et Brandmeier, 2002). Les quatre attributs présentés donnent un aperçu de la complexité des choix possibles dans cette seconde phase de l'instanciation.

On constate que le nombre d'instances utilisées pour une modalité donnée est souvent peu significatif au regard des possibilités. C'est pourquoi, nous pensons qu'il est préférable de spécifier un ensemble de modèles d'instanciation d'attributs au lieu de l'ensemble des instances possibles pour chaque attribut de la modalité. La Figure 11 illustre ce concept dans le cadre de la modalité Texte avant de généraliser le raisonnement à l'ensemble des modalités en sortie.

## 5.3 Le modèle d'instanciation d'une modalité

Nous proposons de formaliser ces mécanismes de sélection du contenu et des attributs au sein d'un *modèle d'instanciation* de la modalité. Ce modèle référence l'ensemble des instances possibles de la modalité et leurs contextes d'utilisation. Il se compose d'un ensemble de feuilles de style décrivant les contenus et attributs à employer selon le contexte. L'ensemble des attributs et/ou contenus de la modalité ne sont pas obligatoirement référencés par une même feuille de style. Plusieurs feuilles de style peuvent ainsi se combiner afin de compléter l'instanciation de la modalité.

Le concepteur définit d'abord une feuille de style par défaut, répertoriant les valeurs par défaut. Puis, il spécifie les conditions d'application des feuilles de style partielles. Cette condition d'application régleme l'utilisation d'une feuille de style. Elle se formalise dans une logique propositionnelle prenant pour composantes : les critères du contexte d'interaction, les propriétés de l'information et les attributs du média. Ces éléments représentent l'ensemble des facteurs pouvant influencer le processus d'instanciation d'une modalité en sortie, permettant ainsi d'adapter le contenu et les attributs d'une modalité selon le contexte.

Reprenons l'exemple de la modalité Photographie dans la présentation de l'appelant. La feuille de style par défaut pourrait être de la forme : taille à 128\*128, couleur en 65536, luminosité à 50%, contraste à 50 %. Trois feuilles de style partielles complètent le modèle d'instanciation en précisant le contenu (le portrait) devant être utilisé selon le numéro de téléphone entrant (condition d'application).

## 6 Then : Comment faire évoluer cette présentation ?

Nous venons de voir comment présenter une information et plus précisément comment adapter une présentation multimodale à un état donné du contexte. Cependant le contexte d'interaction peut être sujet à évolution ce qui soulève un problème de péremption de la présentation. La présentation est en effet adaptée lors de sa conception mais risque de ne plus l'être dans le cas d'une évolution du contexte. Ce problème de péremption affecte principalement les présentations persistantes, une évolution du contexte étant peu probable (mais pas impossible) dans le cas de présentations ponctuelles. Une présentation multimodale doit donc être adaptée au contexte d'interaction tout au long de son cycle de vie. Cette contrainte nécessite la mise en place de mécanismes et de concepts garantissant une évolution de l'interaction en sortie en accord avec l'évolution du contexte.

### 6.1 Les facteurs d'évolution

Les facteurs d'évolution d'une présentation sont particulièrement diffus mais peuvent être classés en quatre grands groupes :

- le facteur informationnel,
- le contexte d'interaction,
- les actions de l'utilisateur,
- les facteurs spatio-temporels.

Le facteur informationnel représente le principal facteur d'évolution. Le noyau fonctionnel de l'application actualise un paramètre en cours de présentation, entraînant son évolution. Prenons l'exemple du niveau de batterie d'un périphérique portable, la présentation du niveau évolue selon son état (Figure 15).

Concernant le contexte, toutes ses évolutions ne nécessitent pas nécessairement une vérification de la péremption des présentations en cours. L'évolution doit ainsi porter sur les éléments du contexte susceptibles d'influer directement sur la présentation. Par exemple, l'augmentation du niveau de bruit n'a aucune incidence sur une présentation visuelle. Ces éléments sont nécessairement des critères intervenant dans les prémisses des règles du modèle comportemental ou dans les conditions d'application du modèle d'instanciation, ayant donné naissance à la présentation. Une évolution du contexte sur un de ces critères peut alors changer l'application du modèle comportemental et/ou du modèle d'instanciation. La présentation est alors périmée (inadaptée au contexte d'interaction en cours) et doit donc évoluer en accord avec le nouveau contexte.

Les actions de l'utilisateur sur une présentation multimodale sont également un facteur d'évolution (Figure 12). Le survol d'une icône par un dispositif de pointage peut par exemple provoquer l'affichage d'un texte rappelant la sémantique de l'icône. L'interaction avec cette dernière peut pareillement avoir des répercussions sur la forme et le contenu de la présentation.

De même, le facteur temporel peut participer à l'évolution d'une présentation multimodale. Les systèmes de planification sont un exemple d'applications adaptant la présentation d'une information (rendez-vous, tâche, etc.) en fonction de la date et de l'heure. La Figure 13 présente un rendez-vous à trois instants distincts (1 jour avant, 1h30 avant et 1h après) sous l'application Agenda Fusion pour Pocket PC. Selon la proximité temporelle du rendez-vous, ce dernier se voit plus ou moins mis en valeur afin de rappeler à l'utilisateur son emploi du temps. Parallèlement, la présentation du rendez-vous une fois passé, est actualisée pour se distinguer des tâches à venir.

Enfin, le facteur spatial fait référence à la position et à l'espace alloué à une présentation multimodale. L'application FlexClock (Grolaux *et al.*, 2002) adaptant la présentation d'une horloge en fonction de la taille de la fenêtre support est un exemple de système exploitant ce facteur d'évolution (Figure 14). L'horloge se voit ainsi composée d'une modalité graphique et d'un calendrier dans le cadre d'une fenêtre de grande taille et d'une simple modalité Texte dans une situation inversée.

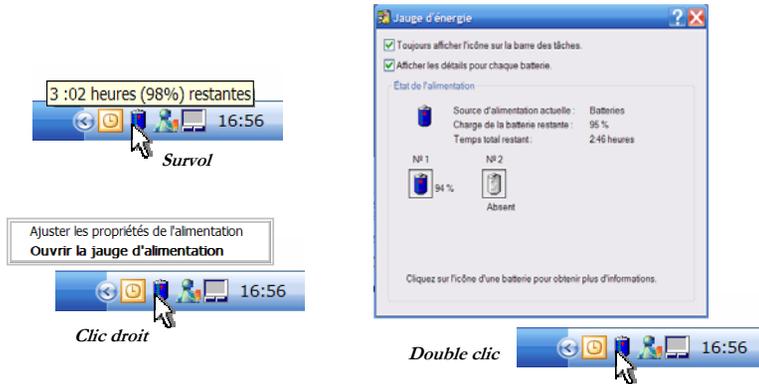


Figure 12. L'influence des actions de l'utilisateur dans l'évolution d'une présentation

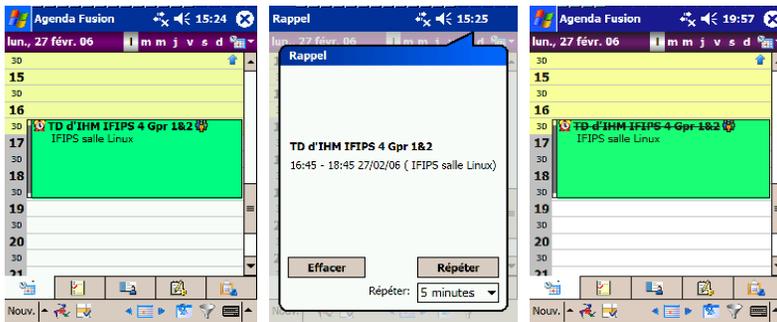


Figure 13. L'influence du facteur temporel dans l'évolution d'une présentation

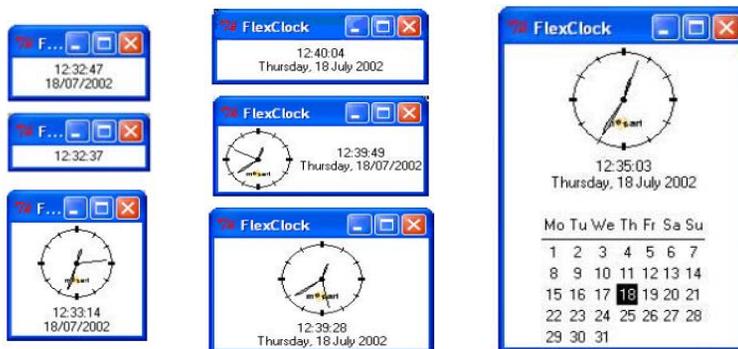


Figure 14. L'influence du facteur spatial dans l'évolution d'une présentation FlexClock de Grolaux *et al.* (2002)

### 6.2 Les types d'évolution

Nous distinguons deux types d'évolution. Prenons l'exemple d'une présentation multimodale indiquant le niveau de batterie d'un périphérique portable quelconque. La Figure 15 propose quatre évolutions de la présentation initiale en l'adaptant à l'état de la batterie. La présentation à 37% (Figure 15, 3<sup>ème</sup> état) évolue en ajustant entre autres les couleurs de ses modalités. De même, la présentation d'une coupure imminente (Figure 15, 5<sup>ème</sup> état) se voit complétée d'une modalité auditive de type Earcon. Une distinction apparaît alors entre les évolutions portant sur des attributs de la présentation (couleur, style, taille) et les évolutions relatives aux modalités de la présentation (Texte, Bulle, Earcon).

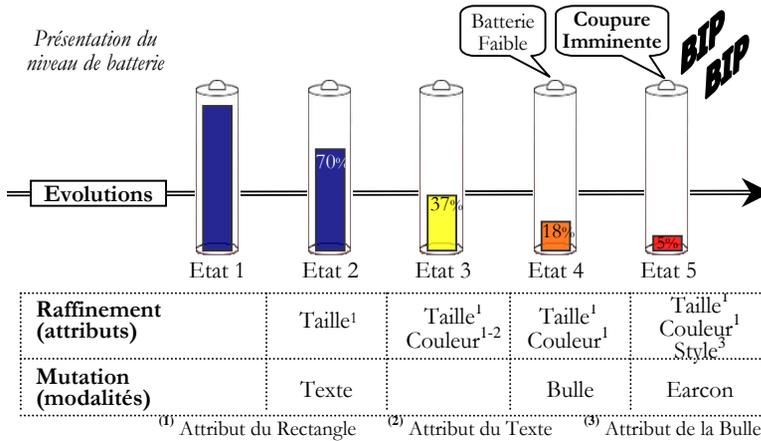


Figure 15. Les types d'évolution d'une présentation

Nous définissons deux types d'évolution (non exclusifs) selon la nature des répercussions sur la présentation multimodale : le raffinement et la mutation. Le premier type d'évolution intitulé « raffinement » ne change pas les composants d'interaction (modalités, médias) utilisés par la présentation mais leurs instanciations (Figure 15, états 2-3-4-5). En revanche, le second type intitulé « mutation » est un changement radical des modalités et/ou médias composant la présentation (Figure 15, états 2-4-5).

Les types d'évolution identifiés, nous allons maintenant nous intéresser aux mécanismes de gestion des présentations multimodales persistantes. Les paragraphes suivants présentent deux approches encapsulant les mécanismes d'évolution à des niveaux architecturaux différents.

### 6.3 Une approche centralisée

Cette première approche dite « centralisée » regroupe l'ensemble des mécanismes d'évolution au sein d'un module commun aux présentations multimodales courantes : le module d'évolution (Figure 16). Toute nouvelle présentation est alors inscrite à ce module. Cette inscription référence la présentation et active le suivi des critères du contexte garantissant sa validité (sa non péremption). L'évolution d'un de ces critères entraîne une invalidation de la présentation et l'application de mécanismes de mise à jour. La fin d'une présentation se traduit par sa désinscription du module d'évolution.

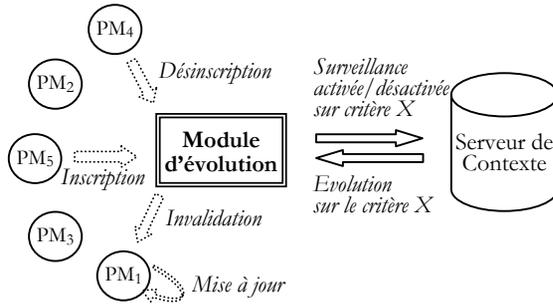


Figure 16. Le module d'évolution des présentations multimodales

Les critères du contexte garantissant la validité d'une présentation sont déductibles lors de la conception de la présentation (Figure 17). Les règles d'élection (Figure 17,  $R_i$ ) et les conditions d'application (Figure 17,  $Cond_i$ ) évaluées lors de la phase de conception pour présenter une unité d'information élémentaire (Figure 17,  $UIE_i$ ) sont connues. Il est ainsi possible d'identifier les critères du contexte (et leurs instances) à l'origine de la présentation. Cet ensemble de critères instanciés (Figure 17,  $C_i$ ) définit le cadre de validité de la présentation : son contexte d'origine. Le suivi de ces critères consiste alors à vérifier que le contexte courant reste conforme au contexte d'origine. Cette conformité se traduit par l'inclusion du contexte d'origine (A) dans le contexte courant (B), soit : tout élément (critère instancié) de A est élément de B. Une inclusion partielle souligne en revanche une non-conformité du contexte d'origine entraînant l'invalidation de la présentation.

En cas d'invalidation d'une présentation multimodale, des mécanismes de mise à jour sont appliqués. Ces mécanismes adaptent la présentation au nouveau contexte d'interaction. Pour ce faire, deux procédés sont envisageables (Figure 18). Le premier procédé consiste à réappliquer dans son intégralité le processus de conception d'une présentation multimodale (Figure 18, A). Une fois le processus appliqué, la nouvelle présentation remplace la présentation périmée. Cette solution simple à mettre en œuvre reste cependant particulièrement coûteuse en temps d'exécution.

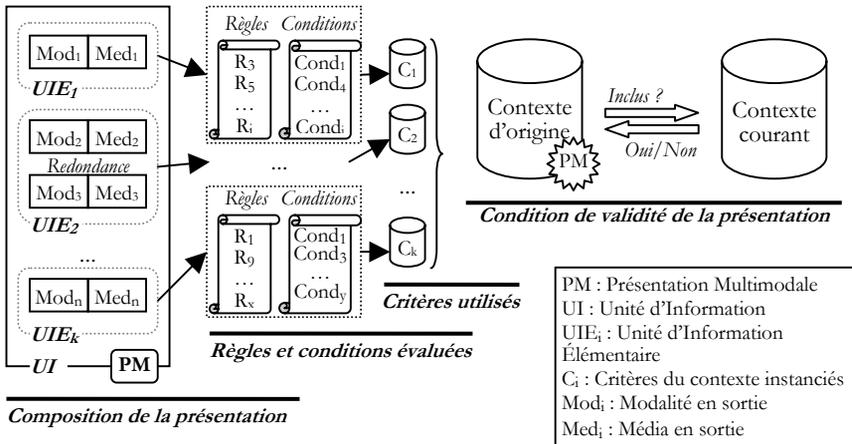


Figure 17. La condition de validité d'une présentation multimodale

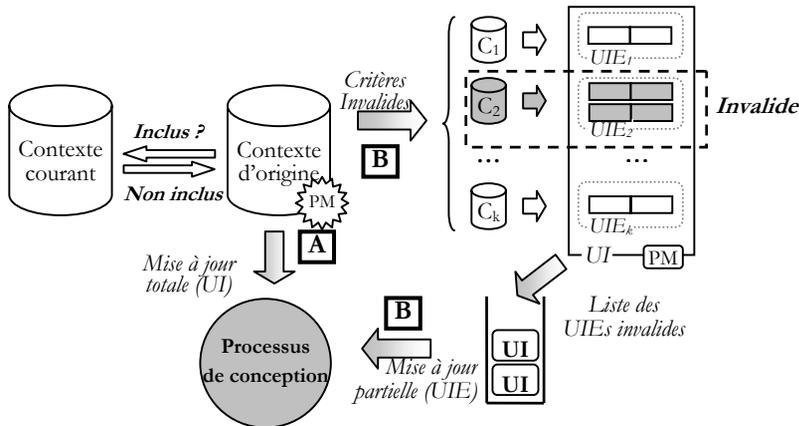


Figure 18. La mise à jour d'une présentation multimodale

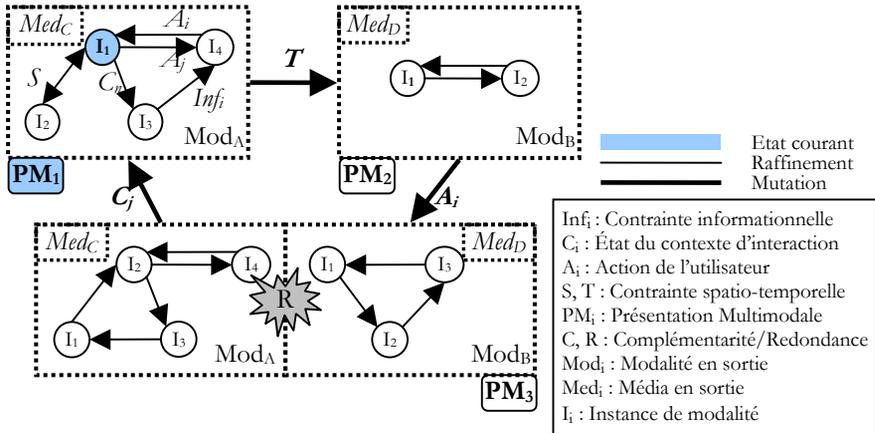
Le second procédé tente de limiter le coût de mise à jour d'une présentation en effectuant une application incrémentale du processus de conception (Figure 18, **B**). Il s'agit de distinguer au sein de la présentation (Figure 18, PM), les éléments valides des éléments invalides. Cette distinction est réalisée au niveau des unités d'information élémentaires exprimées par la présentation. Chaque unité d'information élémentaire (Figure 18, UIE<sub>i</sub>) dispose de son propre contexte d'origine (Figure 18, C<sub>i</sub>) garantissant sa validité. Les unités d'information élémentaires dont le contexte d'origine est non conforme au contexte courant sont considérées comme invalides. Il en va de même pour les composantes (modalités et médias) de la présentation associée. Le processus de conception est ainsi appliqué spécifiquement pour les informations élémentaires invalidées. L'application du processus s'accompagne d'une vérification de la cohérence de la présentation.

La mise à jour effectuée, le nouvel état de la présentation est consigné dans l'historique de présentation de l'information associée. Cet historique retrace l'évolution de la présentation de l'information (les présentations passées et leurs contextes d'origine) et se modélise sous la forme d'un automate à états. Il est alors envisageable de réutiliser cet historique afin d'améliorer la mise à jour de la présentation. L'évolution de la présentation pourrait ainsi s'effectuer selon ses états passés, limitant les calculs de transition. En l'état actuel, il semble difficile d'automatiser l'extraction de recommandations de manière efficace. Ce sujet reste néanmoins une perspective de recherche à court terme.

#### 6.4 Une approche distribuée

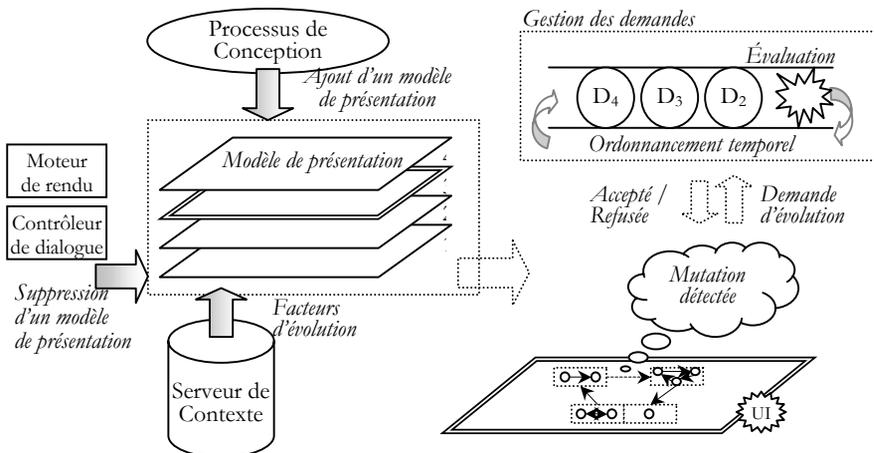
Dans l'approche « distribuée », l'expression d'une information ne produit plus une présentation multimodale mais *un modèle de présentation* (Figure 19). Ce modèle de présentation est un automate à transitions à deux niveaux (global et local) dans lequel chaque état symbolise une présentation et où chaque transition est étiquetée par un facteur d'évolution.

L'état global courant (Figure 19, PM<sub>1</sub>) représente la présentation initiale de l'information. Les autres états globaux (Figure 19, PM<sub>2</sub> et PM<sub>3</sub>) représentent toutes les autres alternatives de présentation. De même, le ou les états locaux courants (Figure 19, I<sub>1</sub>) précisent l'instanciation des modalités composant la présentation courante (l'état global courant). Les autres états locaux (Figure 19, I<sub>2</sub>, I<sub>3</sub> et I<sub>4</sub>) indiquent les alternatives à suivre en cas d'évolution.



L'étiquette de transition (locale ou globale) identifie le facteur d'évolution à l'origine d'un changement d'état : informationnel (Figure 19, Inf<sub>i</sub>), contexte d'interaction (Figure 19, C<sub>i</sub>), action de l'utilisateur (Figure 19, A<sub>i</sub>) ou paramètre spatio-temporel (Figure 19, S-T). La nature du changement d'état (local ou global) retranscrit la typologie de l'évolution. Ainsi, le raffinement d'une présentation se traduit par un changement d'état local (Figure 19, I<sub>1</sub>→I<sub>1</sub>) tandis que la mutation d'une présentation se traduit par un changement d'état global (Figure 19, PM<sub>1</sub>→PM<sub>1</sub>).

L'exécution du modèle de présentation est assurée par le serveur de contexte (un problème complexe en soi). Ce dernier mis à jour par les différents modules de l'application, indique l'état courant du contexte d'interaction et recense les informations relatives aux autres facteurs d'évolution. Le modèle est ainsi en mesure de détecter une évolution et de mettre en place sa « stratégie » de présentation. Cette stratégie de présentation ne peut être appliquée sans une vérification préalable de la cohérence de l'interaction en sortie (Figure 20).



L'évolution d'un modèle de présentation peut en effet entraîner une modification du contexte d'interaction et ainsi influencer les autres modèles de présentation en cours. L'utilisation de la modalité Earcon (suite à une mutation) provoque par exemple une occupation du Haut-parleur et limite l'usage de nouvelles modalités auditives (les sujets potentiels d'une évolution parallèle). Il est donc nécessaire de synchroniser les évolutions en cours afin de préserver la cohérence de l'interaction. Cette synchronisation s'effectue à l'aide d'une demande d'évolution, émise lors de la détection d'une évolution au sein d'un modèle de présentation. Ces demandes sont collectées puis ordonnancées afin de vérifier les recouvrements entre les modèles de présentation impliqués.

### 6.5 Discussion : centraliser ou distribuer ?

Suite à la présentation de ces deux approches, une question subsiste : *Faut-il centraliser ou distribuer les mécanismes d'évolution d'une présentation ?* Au regard des apports et inconvénients de chacun, il semble difficile d'affirmer qu'une approche soit plus pertinente qu'une autre.

Ainsi, l'approche centralisée apparaît comme une solution rapide, et efficace. Cependant, les concepts rattachés à l'évolution (les facteurs, les types et les mécanismes) sont représentés de façon implicite, limitant la compréhension de l'évolution. Cette faible lisibilité lui confère une vision locale et temporaire du problème de l'évolution.

Le modèle de présentation offre une vision globale du problème faisant de lui un cadre théorique pour modéliser l'évolution désirée. Il souffre néanmoins d'une complexité non maîtrisée dans le cadre d'un large contexte d'interaction ou d'une application sujette à évolution.

## 7 La plateforme ELOQUENCE

Sur les bases du modèle WWHT, nous avons développé une plateforme logicielle permettant une conception itérative de systèmes multimodaux en sortie (Flippo *et al.*, 2003). Cette plateforme intitulée ELOQUENCE comprend deux outils qui permettent respectivement de spécifier et de simuler les sorties d'un système multimodal, et d'un noyau permettant l'exécution du système final.

### 7.1 L'outil de spécification

L'outil de spécification accompagne le concepteur dans la définition des éléments nécessaires à l'application du modèle : les composants d'interaction, le contexte d'interaction, les unités d'information, le modèle comportemental et le modèle d'instanciation. Cette spécification s'effectue à l'aide de cinq éditeurs dédiés. La Figure 21 présente l'éditeur des composants d'interaction dans le cadre de l'application téléphonique. Celui-ci comprend une zone d'édition du diagramme et d'une zone de propriétés relative au composant sélectionné (attributs, critères et commentaires). L'éditeur du modèle comportemental présenté en Figure 22 spécifie les prémisses d'une règle d'élection sous la forme d'un arbre et les conclusions à l'aide d'une table. Dans l'exemple, la règle en cours d'édition propose de favoriser la modalité Vibration au détriment de la modalité Sonnerie pour présenter un appel téléphonique dans un cinéma. Enfin, la spécification résultante est exportée en MOXML (Multimodal Output eXtended Markup Language), un formalisme de représentation de données basé sur XML. (Rousseau *et al.*, 2004).

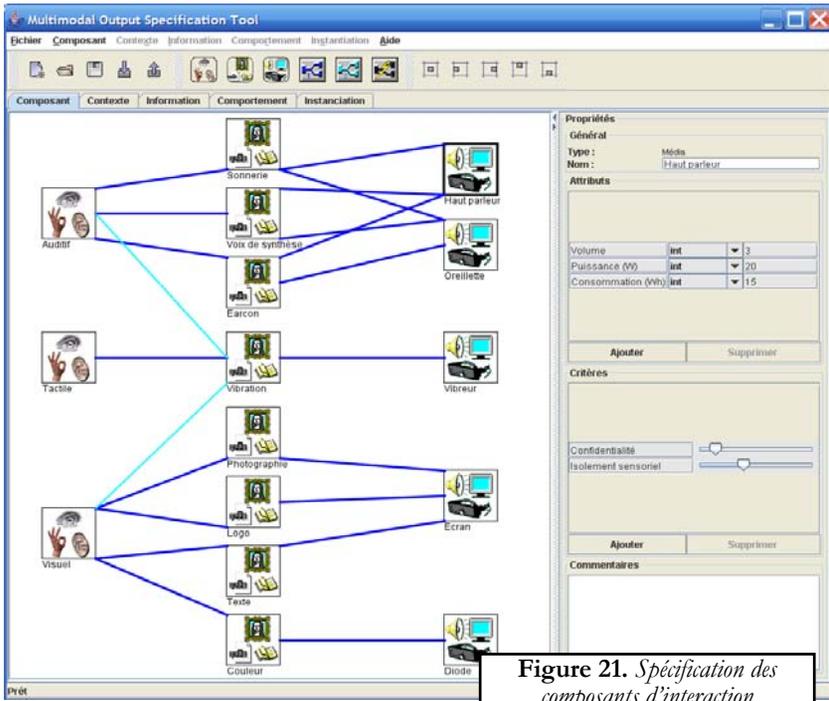


Figure 21. Spécification des composants d'interaction

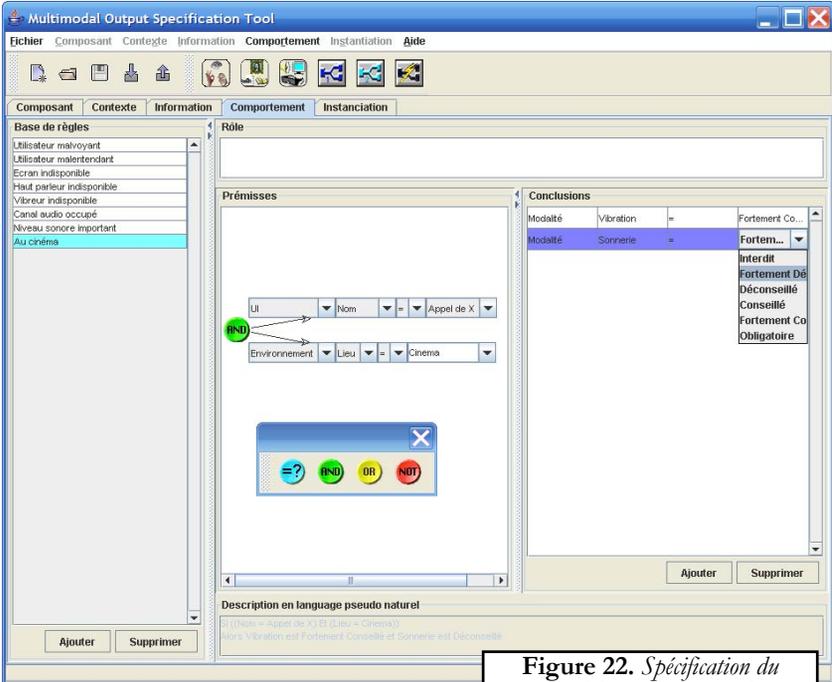


Figure 22. Spécification du modèle comportemental

### 7.2 L'outil de simulation

L'outil de simulation (Rousseau *et al.*, 2005) se compose de quatre services avec leur interface (Figure 23). Un premier service (Figure 23, A) permet de configurer les paramètres de simulation : les fichiers XML décrivant la spécification et le niveau de simulation (allocation, instanciation, rendu). Un second service (Figure 23, B) simule le contrôleur de dialogue. Il permet de déclencher la présentation d'une information de notre choix. Le troisième service (Figure 23, C) simule le serveur de contexte, ce qui permet au concepteur d'éditer l'état du contexte d'interaction dans laquelle la présentation de l'information va se dérouler. Enfin, les résultats de la simulation sont présentés sous forme textuelle dans une quatrième et dernière interface utilisateur (Figure 23, D).

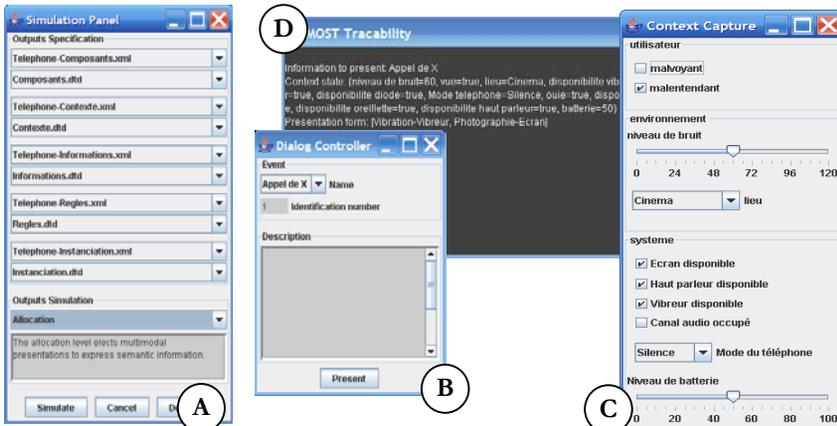


Figure 23. Simulation de la réception d'un appel téléphonique

### 7.3 Le noyau d'exécution

L'outil de simulation se compose d'un noyau indépendant, utilisable pour l'exécution d'un système final. Ce noyau d'exécution dispose d'une architecture modulaire et d'une couche de communication en Java RMI (Figure 24). Les trois principaux modules de l'architecture (moteurs d'allocation, d'instanciation et d'évolution) implémentent les concepts fondamentaux du modèle WWHT. Un module de gestion des présentations multimodales complète l'architecture en centralisant les ressources et les communications.

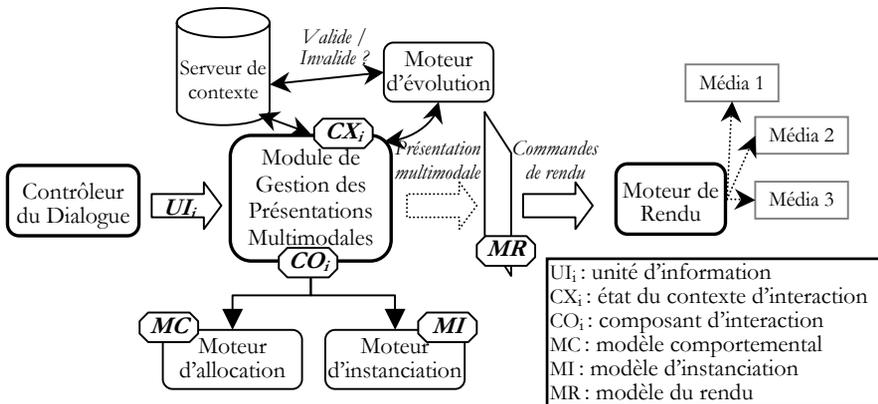


Figure 24. Modèle d'architecture du noyau d'exécution

## 8 Application

Nous allons maintenant reprendre les concepts introduits et expliciter leur application dans la conception des sorties d'un simulateur de contrôle aérien.

### 8.1 Le simulateur de contrôle aérien

Cette application a été développée dans le cadre du projet INTUITION (Interaction mUltimodale Intégrant les Technologies InnOvaNtes). Ce projet financé par la DGA (Délégation Générale pour l'Armement) regroupe trois laboratoires français (LIMSI-CNRS, CLIPS-IMAG et LIHIS-IRIT) et un partenaire industriel (THALES-Avionics). L'objectif du projet est la mise en place d'une plateforme générique d'adaptation des nouvelles technologies dans les interfaces Homme-Machine (Bastide *et al.*, 2005).

Le simulateur se compose principalement d'un pseudo-pilote, d'un moteur de rendu et d'un moteur d'interaction multimodale comprenant les systèmes : ICARE, Petshop et ELOQUENCE. Le système ICARE (Bouchet *et al.*, 2004) capture et interprète les événements en entrée du système. Ces événements sont transmis au contrôleur de dialogue formalisé en réseaux de Pétri grâce à l'outil Petshop (Navarre *et al.*, 2005). Enfin, l'information est présentée à l'utilisateur par l'intermédiaire de la plateforme ELOQUENCE.

L'application a été développée avec des techniques d'interaction multimodale non nécessairement ergonomiques, l'objectif étant de démontrer l'apport de nos outils sans se préoccuper du niveau d'utilisabilité de l'interface.

### 8.2 Spécification des sorties

La Figure 25 présente un aperçu du diagramme des composants d'interaction de l'application, comprenant des modalités élémentaires : Texte, Avion (modèle 3D d'avion), etc., et une modalité composée : Etiquette (les paramètres d'un vol). Le contexte d'interaction est ici restreint à quatre critères décrits en Tableau 3.

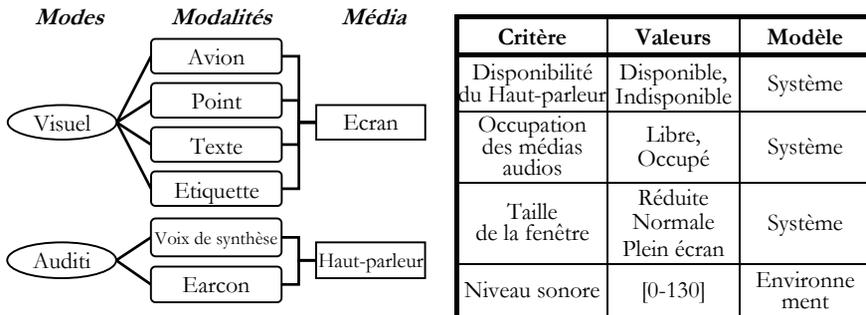


Figure 25. Diagramme des composants d'interaction Tableau 3. Contexte d'interaction

Une trentaine d'unités d'information : entrée d'un vol, sélection d'un vol, édition d'un paramètre de vol, sortie d'un vol, etc., sont gérées par l'application. Dans cet exemple, nous étudierons plus précisément l'unité d'information « Entrée du vol X », signalant l'entrée d'un vol dans le secteur aérien et se décomposant en deux unités d'information élémentaires : l'entrée et le vol.

Enfin, huit règles d'élection du modèle comportemental sont décrites dans le Tableau 4. Le modèle d'instanciation est partiellement illustré dans la section suivante. Ainsi, la modalité Avion dispose d'un contenu (un objet 3D modélisant un avion) et de quatre attributs : taille, texture, coordonnées et direction. De même, la modalité Etiquette (de vol) comprend quatre contenus (identifiant, altitude, vitesse, cap) et des attributs de couleur, de taille et de positionnement.

	Description en langage naturel
R1	Si le Haut-parleur est indisponible <b>Alors</b> ne pas utiliser le média Haut-parleur
R2	Si les médias audios sont occupés <b>Alors</b> ne pas utiliser la modalité Voix de Synthèse
R3	Si le niveau sonore est compris entre 70 et 90 dB <b>Alors</b> défavoriser la modalité Voix de Synthèse
R4	Si le niveau sonore est supérieur à 90 dB <b>Alors</b> ne pas utiliser le mode Auditif
R5	Si la fenêtre est de taille réduite <b>Alors</b> ne pas utiliser la modalité Avion
R6	Si la fenêtre est en plein écran <b>Alors</b> ne pas utiliser la modalité Point
R7	Si l'UIE est une « entrée » <b>Alors</b> favoriser la modalité Earcon
R8	Si l'UIE est un « vol » <b>Alors</b> favoriser la modalité Avion <b>Et</b> utiliser une complémentarité

Tableau 4. Règles d'élection du modèle comportemental

8.3 Simulation des sorties

La Figure 26 présente un aperçu de l'application du modèle WWHT pour l'unité d'information « Entrée du vol X ». Dans une situation nominale, seules les règles R7 et R8 sont appliquées. L'entrée du vol (première UIE) est ainsi signalée par le couple (Earcon, Haut-parleur) tandis que les couples (Avion, Ecran) et (Etiquette, Ecran) introduisent le vol (seconde UIE). Dans un contexte différent, la présentation de l'information peut être d'une toute autre forme. Supposons que la fenêtre de l'application est de taille réduite, la modalité Avion est alors remplacée par la modalité Point afin d'optimiser la lisibilité de l'interface (règle R5).

L'exemple d'évolution proposé en Figure 26 est centré sur la seconde unité d'information élémentaire : le « vol ». Cette évolution a pour but de réduire progressivement la taille de la présentation suite à l'entrée du vol dans le secteur aérien. Un premier raffinement réduit la taille du modèle 3D d'avion et le nombre de paramètres de vol présentés par l'étiquette (de 4 à 2). Il est suivi d'une mutation remplaçant le modèle d'avion par un point. Enfin, un dernier raffinement restreint l'étiquette à un unique paramètre, l'identifiant de vol.

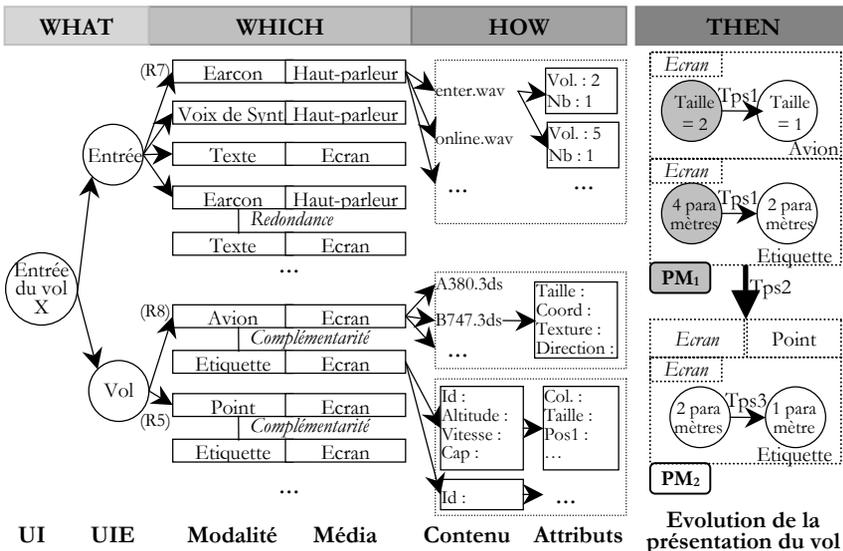


Figure 26. Application du modèle dans le cadre de l'entrée d'un vol dans un secteur aérien

## 9 Conclusion

La contribution de cet article porte sur la complétion et la contextualisation du cycle de vie d'une présentation multimodale. Le modèle WWHT rassemble les concepts nécessaires à la description de ce cycle de vie. Ces concepts au nombre de quatre : « What », « Which », « How » et « Then », structurent les processus de conception et d'évolution de la présentation. Le processus de conception se voit ainsi décomposé en trois phases : la fission sémantique (What), l'élection (Which) et l'instanciation (How), intégrant le contexte d'interaction dans leurs mécanismes de décisions grâce à l'exploitation d'un modèle comportemental et d'un modèle d'instanciation. La gestion d'une présentation multimodale ne se limite pas à sa conception initiale. Il convient de garantir sa pérennité dans un contexte d'interaction en perpétuelle évolution (Then). Les facteurs et les types d'évolution possibles ont été identifiés et deux approches centralisant ou distribuant les mécanismes de gestion des présentations multimodales persistantes ont été proposées puis brièvement comparées.

Fondée sur ce modèle, la plateforme logicielle ELOQUENCE implémente les concepts identifiés et outille le processus de conception de systèmes multimodaux en sortie. Deux outils sont proposés pour spécifier et simuler les sorties d'un système multimodal. Un noyau d'exécution complète la plateforme en lui conférant la possibilité d'exécuter le système final.

Ces concepts et outils ont été appliqués avec succès dans le cadre de deux applications portant sur les domaines de la téléphonie mobile et du contrôle aérien, apportant un début de validation logicielle.

Enfin, des travaux sont en cours sur l'identification et la résolution des différentes incohérences pouvant être levées au sein du modèle WWHT. Un système de négociation entre les différentes phases du modèle doit être mis en place afin d'éviter les situations de blocage. De nouvelles fonctionnalités sont également en prévision pour l'outil de spécification. Nous souhaitons utiliser un second formalisme pour le modèle comportemental afin de proposer une vue alternative à la base de règles et donner une vue globale du modèle.

## Remerciements

Le travail présenté dans cet article a été financé en partie par le contrat n°00.70.624.00.470.75.96 de la DGA (Délégation Générale pour l'Armement).

## Références

André, E. (2003). Natural Language in Multimedia/Multimodal Systems. In *Computational Linguistics*, R. Mitkov (Ed.), Oxford University Press, 650-669.

Arens, Y., Hovy, E.H. (1995). The Design of a Model-Based Multimedia Interaction Manager. *Artificial Intelligence Review*, vol. 9, num. 3, 167-188.

Bass, L., Faneuf, R., Little, R., Mayer, N., Pellegrino, B., Reed, S., Seacord, R., Sheppard, S., Szczur, M.R. (1992). A Metamodel for the Runtime Architecture of an Interactive System, The UIMS Tool Developers Workshop, *SIGCHI Bulletin*, vol. 24, num. 1, 32-37.

Bastide, R., Bazalgette, D., Bellik, Y., Nigay, L., Nouvel, C. (2005). The INTUITION Design Process: Structuring Military Multimodal Interactive Cockpits

Design According to the MVC Design Pattern. In *Proceedings HCI International 2005*, Las Vegas, Nevada, USA, July 2005.

Bellik, Y. (1995). *Interfaces Multimodales: Concepts, Modèles et Architectures*. Thèse de l'Université Paris XI, LIMSI, mai.

Bernsen, N.O. (1996). A reference model for output information in intelligent multimedia presentation systems. In *Proceedings ECAI'96 Workshop on Towards a Standard Reference Model for Intelligent Multimedia Presentation Systems*, Budapest.

Bordegoni, M., Faconti, G., Maybury, M.T., Rist, T., Ruggieri, S., Trahanias, P., Wilson, M. (1997). A Standard Reference Model for Intelligent Multimedia Presentation Systems. *Computer Standards and Interfaces*, vol. 18, num 6-7, 477-496.

Bouchet, J., Nigay, L., Ganille, T. (2004). ICARE Software Components for Rapidly Developing Multimodal Interfaces. In *Proceedings of ICMI 2004*, State College, Pennsylvania, USA, October 2004, 251-258.

Braffort, A., Choisier, A., Collet, C., Dalle, P., Gianni, F., Lenseigne, B., Segouat, J. (2004). Toward an Annotation Software for Video of Sign Language, Including Image Processing Tools and Signing Space Modelling. In *Proceedings Language Resources and Evaluation, LREC'04*, Lisbonne, Portugal, May 2004.

Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonck, J. (2003). A Unifying Reference Framework for Multi-target User Interfaces. *Interacting With Computer*, vol. 15, num. 3, 289-308.

Colineau, N., Paris, C., Wu, M. (2004). Actionable Information Delivery. *Artificial Intelligence Review, Special Issue on Tailored Information Delivery*, vol. 18, num. 4, 549-576.

Coutaz, J. (1992). Multimedia and Multimodal User Interfaces: A Software Engineering Perspective. In *Proceedings Workshop on Human Computer Interaction*, St Petersburg, Russia.

Coutaz, J., Nigay, L., Salber, D., Blandford, A., May, J., Young, R.M. (1995). Four Easy Pieces for Assessing the Usability of Multimodal Interaction: the CARE Properties. In *Proceedings INTERACT'95*, Lillehammer, Norway, 115-120.

Dalal, M., Feiner, S., McKeown, K., Pan, S., Zhou, M., Höllerer, T., Shaw, J., Feng, Y., Fromer, J. (1996). Negotiation for Automated Generation of Temporal Multimedia Presentations. In *Proceedings ACM Multimedia'96*, Boston, USA, November 1996, 55-64.

Dey, A.K., Salber, D., Abowd, G.D. (2001). A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. *Human-Computer Interaction*, 16 (2-4), 97-166.

Duarte, C., Carriço, L. (2006). A Conceptual Framework for Developing Adaptive Multimodal Applications. In *Proceedings Intelligent User Interfaces, IUI'06*, Sydney, Australia, January 2006, 132-139.

Flippo, F., Krebs, A., Marsic, I. (2003). A Framework for Rapid Development of Multimodal Interfaces. In *Proceedings international conference on Multimodal interfaces, ICMI'03*, Vancouver, Canada, November 2003, 109-116.

Frohlich, D.M. (1991). The Design Space of Interfaces. In *Multimedia principles, systems and applications*, Kjelldahl, L. (Ed.), Springer Verlag éditions, 53-69.

- Graf, W. H. (1995). The Constraint-based Layout Framework LayLab and its Applications. In *Proceedings ACM Multimedia Workshop on Effective Abstractions in Multimedia Layout, Presentations, and Interaction*, San Francisco, U.S.A.
- Grolaux, D., Van Roy, P., Vanderdonckt, J. (2002). FlexClock: A Plastic Clock Written in Oz with the QtK Toolkit. In *Proceedings TAMODLA Workshop on Task Models and Diagrams for User Interface Design*, Bucharest, Romania, July 2002.
- Johnston, M., Bangalore, S. (2005). Finite-state Multimodal Integration and Understanding. In *Natural Language Engineering*, vol. 11, num. 2, 159-187.
- Karagiannidis, C., Koumpis, A., Stephanidis, C. (1997). Adaptation in Intelligent Multimedia Presentation Systems as a Decision Making Process. *Computer Standards and Interfaces*, vol. 18, num. 2-3, 509-514.
- Kolski, C., Tendjaoui, M., Millot, P. (1992). A process method for the design of "intelligent" man-machine interfaces: case study: "the Decisional Module of Imagery". *International Journal of Human Factors in Manufacturing*, vol. 2, 155-175.
- Layaïda, N., Lemlouma, T., Quint, V. (2005). NAC : une architecture pour l'adaptation multimédia sur le Web. *Technique et Science Informatiques*, vol. 24, num. 7, 789-813.
- Martin, J.C. (1998). TYCOON: Theoretical Framework and Software Tools for Multimodal Interfaces. In *Intelligence and Multimodality in Multimedia Interfaces*, Lee, J. (Ed.), AAAI Press.
- Navarre, D., Palanque, P., Bastide, R., Schyn, A., Winckler, M. A., Nedel, L., Freitas, C. (2005). A Formal Description of Multimodal Interaction Techniques for Immersive Virtual Reality Applications. In *Proceedings INTERACT'05*, Rome, Italy, September 2005.
- Nigay, L., Coutaz, J. (1995). A Generic Platform for Addressing the Multimodal Challenge. In *Proceedings CHI'95*, Denver, Colorado, USA, May 1995, 98-105.
- Rist, T., Brandmeier, P. (2002). Customizing Graphics for Tiny Displays of Mobile Devices. *Personal and Ubiquitous Computing*, vol. 6, num. 4, 260-268.
- Rousseau, C., Bellik, Y., Vernier, F., Bazalgette, D. (2004). Architecture Framework for Output Multimodal Systems Design. In *Proceedings OZCHI'04*, Wollongong, Australia, November 2004.
- Rousseau, C., Bellik, Y., Vernier, F. (2005). Multimodal Output Specification / Simulation Platform. In *Proceedings ICMI'05*, Trento, Italy, October 2005.
- Savidis, A., Stephanidis, C. (2004). Unified User Interface Design: Designing Universally Accessible Interactions. *Interacting with Computers*, vol. 16, num. 2, 243-270.
- Stephanidis, C., Karagiannidis, C., Koumpis, A. (1997). Decision Making in Intelligent User Interfaces. In *Proceedings Intelligent User Interfaces*, IUI'97, 195-202.
- Thevenin, D., Coutaz, J. (1999). Plasticity of User Interfaces: Framework and Research Agenda. In *Proceedings INTERACT'99*, Edinburgh, UK, 110-117.
- Vanderdonckt, J., Grolaux, D., Van Roy, P., Limbourg, Q., Macq, B., Michel, B. (2005). A Design Space for Context Sensitive User Interfaces. In *Proceedings, LASSIE'05*, Toronto, Canada, July 2005.

Vernier, F., Nigay, L. (2000). A framework for the Combination and Characterization of Output Modalities. In *Proceedings Workshop on Design, Specification and Verification of Interactive Systems*, DSV-IS'00, Limerick, 35-50.