

WWHT : Un modèle conceptuel pour la présentation multimodale d'information

Cyril Rousseau

Yacine Bellik

Frédéric Vernier

LIMSI-CNRS
Université Paris XI, B.P. 133
91403, Orsay, France
Tel. : +33 (0)1 69 85 81 10
Fax : +33 (0)1 69 85 80 88
prenom.nom@limsi.fr

RESUME

La présentation multimodale d'information vise à exploiter plusieurs modalités de communication afin de produire des sorties pertinentes à destination de l'utilisateur. Elle met en jeu divers concepts relatifs à l'information, aux composants d'interaction (modes, modalités, média) et au contexte. Elle requiert la mise en place d'un processus incrémental pour l'élaboration d'une présentation adaptée au contexte d'interaction en cours. Cet article propose un modèle conceptuel pour la présentation multimodale d'information intitulé WWHT, basé sur 4 notions fondamentales : What, Which, How, Then. Ces notions décrivent le cycle de vie d'une présentation multimodale allant de sa naissance à sa disparition en passant par ses évolutions au sein d'un contexte d'interaction en perpétuelle modification. L'application du modèle est illustrée à travers la spécification des sorties d'une application de téléphonie mobile.

MOTS CLES : Interaction Homme-Machine, multimodalité en sortie, présentation multimodale, contexte d'interaction.

LOGICIEL UTILISE : Microsoft Office Word 2003.

FORME DE PARTICIPATION : Article de recherche long.

THEME DE LA SOUMISSION : Interaction Multi-Modales, Interfaces Adaptatives.

WWHT : Un modèle conceptuel pour la présentation multimodale d'information

Cyril Rousseau

Yacine Bellik

Frédéric Vernier

LIMSI-CNRS
Université Paris XI, B.P. 133
91403, Orsay, France
prénom.nom@limsi.fr

RESUME

La présentation multimodale d'information vise à exploiter plusieurs modalités de communication afin de produire des sorties pertinentes à destination de l'utilisateur. Elle met en jeu divers concepts relatifs à l'information, aux composants d'interaction (modes, modalités, média) et au contexte. Elle requiert la mise en place d'un processus incrémental pour l'élaboration d'une présentation adaptée au contexte d'interaction en cours. Cet article propose un modèle conceptuel pour la présentation multimodale d'information intitulé WWHT, basé sur 4 notions fondamentales : What, Which, How, Then. Ces notions décrivent le cycle de vie d'une présentation multimodale allant de sa naissance à sa disparition en passant par ses évolutions au sein d'un contexte d'interaction en perpétuelle modification. L'application du modèle est illustrée à travers la spécification des sorties d'une application de téléphonie mobile.

MOTS CLES : Interaction Homme-Machine, multimodalité en sortie, présentation multimodale, contexte d'interaction.

ABSTRACT

Multimodal presentation of information aims at using several modalities of communication to produce relevant outputs to the user. This presentation involves different concepts regarding to information, interaction components (modes, modalities, media) and interaction context. An incremental process is required to create an adapted presentation. This paper presents a conceptual model for the multimodal presentation of information called WWHT and based on four concepts: What, Which, How, Then. These concepts describe the life cycle of a multimodal presentation from its birth to its disappearance including its evolutions. The application of the

model is introduced through the outputs specification of a mobile phone application.

CATEGORIES AND SUBJECT DESCRIPTORS: H.5.m [Information Interfaces and Presentation (e.g., HCI)]:Miscellaneous.

GENERAL TERMS: Human Factors, Theory.

KEYWORDS: Human-Computer Interaction, output multimodality, multimodal presentation, interaction context.

INTRODUCTION

Les connexions en réseau, la miniaturisation et en particulier la baisse des coûts de production des dispositifs électroniques favorisent leur intégration dans tous les objets du quotidien. Il est ainsi devenu commun de posséder un téléphone portable (71,1 % de la population française) ou bien de croiser dans les espaces publics écrans numériques et bornes électroniques diffusant informations et services. L'informatique est devenue pour une grande majorité de personnes, un élément indispensable à la vie quotidienne, à un point tel qu'elle est parfois qualifiée de pervasive [16].

Les techniques d'interaction et les périphériques de communication sont aujourd'hui divers et performants. Leurs évolutions, sujet à de nombreux travaux ces dernières années, permet aujourd'hui d'envisager une contextualisation de l'interaction.

Le choix des techniques d'interaction d'une application est souvent réalisé en fonction du domaine de l'application et d'un contexte d'interaction prédéfini. Le récent nomadisme de l'utilisateur rend cependant difficile l'unicité du contexte d'interaction.

Les systèmes informatiques de demain devront être en mesure de proposer une interaction adaptée au contexte d'interaction en cours [17]. Il est nécessaire de repenser la communication en fonction de l'utilisateur, des capacités du système et de la nature de l'environnement ambiant. Toutes ces composantes définissent un contexte

d'interaction devant être maintenant considéré et exploité dans sa globalité.

Cet article s'intéresse plus particulièrement au cycle de vie d'une présentation multimodale adaptée au contexte d'interaction en cours. Un modèle conceptuel est introduit afin de permettre la conception de présentations multimodales réactives au contexte. Ces concepts sont par la suite illustrés à travers le domaine de la téléphonie mobile, un téléphone portable représentant un système particulièrement riche en termes de modalités en sortie.

COMPOSANTES DE LA MULTIMODALITE EN SORTIE

Nous pensons qu'un système multimodal en sortie ne peut être réduit à un système qui se limite simplement à exploiter plusieurs modalités en sortie. Dans ce cas, nous préférons parler de système multimédia. Un système multimodal en sortie doit en outre être capable de produire des présentations qui soient adaptées au contexte d'interaction en cours. Ce processus s'appuie sur quatre composantes :

- L'information à présenter
- Les composants d'interaction du système
- Le contexte d'interaction en cours
- La présentation multimodale résultante

L'information

En référence au modèle ARCH [3], on peut dire que 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. La « réception d'un nouvel appel » ou encore l'indication d'un « niveau de batterie faible » sont par exemple deux unités d'information à exprimer par le système en charge des sorties d'un téléphone portable.

Les composants d'interaction

Nous définissons les composants d'interaction comme étant les moyens de communication (physiques ou logiques) entre l'homme et la machine. Trois types de composants d'interaction peuvent être distingués: Mode, Modalité et Médium. Selon les auteurs, ces trois notions peuvent avoir des définitions différentes mais néanmoins équivalentes selon le point de vue adopté.

Dans notre cas, nous faisons référence à nos définitions considérées comme orientées utilisateur [5]. 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 médium en sortie est un périphérique de sortie permettant l'expression d'une modalité en sortie (Ecran, Haut parleur, etc.). Ces trois notions sont dépendantes les unes des autres. En effet, à un mode correspond un

ensemble de modalités et à une modalité est rattaché un ensemble de média permettant son expression. La modalité « Vibration » s'exprime par exemple sur le médium « Vibreur » et fait appel au mode « Tactile ».

Enfin, il convient de distinguer deux types de relations entre les composants d'interaction : principales et secondaires. La relation principale fait référence à un lien voulu entre composants. La liaison secondaire résulte quant à elle d'un effet de bord. Dans le cas de 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 « TPK (Tactilo-Proprio-Kinesthésique) » 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 ». L'ensemble de ces relations tisse un diagramme des composants d'interaction en sortie.

Le contexte d'interaction

Nous nous référons à la définition de Anind Dey [10] concernant la notion de contexte d'interaction : « 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 ». Ces entités sont modélisées à l'aide de modèles [4] (modèle du système, modèle de l'utilisateur, etc.). A chaque modèle est associé un ensemble de critères dynamiques ou statiques relatif à l'entité considérée (disponibilité des média, préférences de l'utilisateur, niveau de bruit, etc.).

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-médium en sortie liés par des propriétés de redondance et/ou de complémentarité. La réception d'un appel sur un téléphone portable peut par exemple s'exprimer à travers une présentation multimodale de la forme (« modalité Sonnerie », « médium Haut parleur ») signalant l'arrivée d'un appel, complétée par (« modalité Texte », « médium Ecran ») présentant le numéro de l'appelant.

LE MODELE WWHT : WHAT-WHICH-HOW-THEN

Nous allons maintenant proposer un modèle conceptuel intitulé WWHT présentant le cycle de vie d'une présentation multimodale. Ce modèle est basé sur quatre concepts, les trois premiers faisant référence à la conception de la présentation [7] et le dernier à son devenir [8] :

- *What* : Quelle information présenter ?
- *Which* : Quelles modalités choisir pour la présenter ?
- *How* : Comment présenter cette information à travers ces modalités ?
- *Then* : Comment faire évoluer cette présentation ?

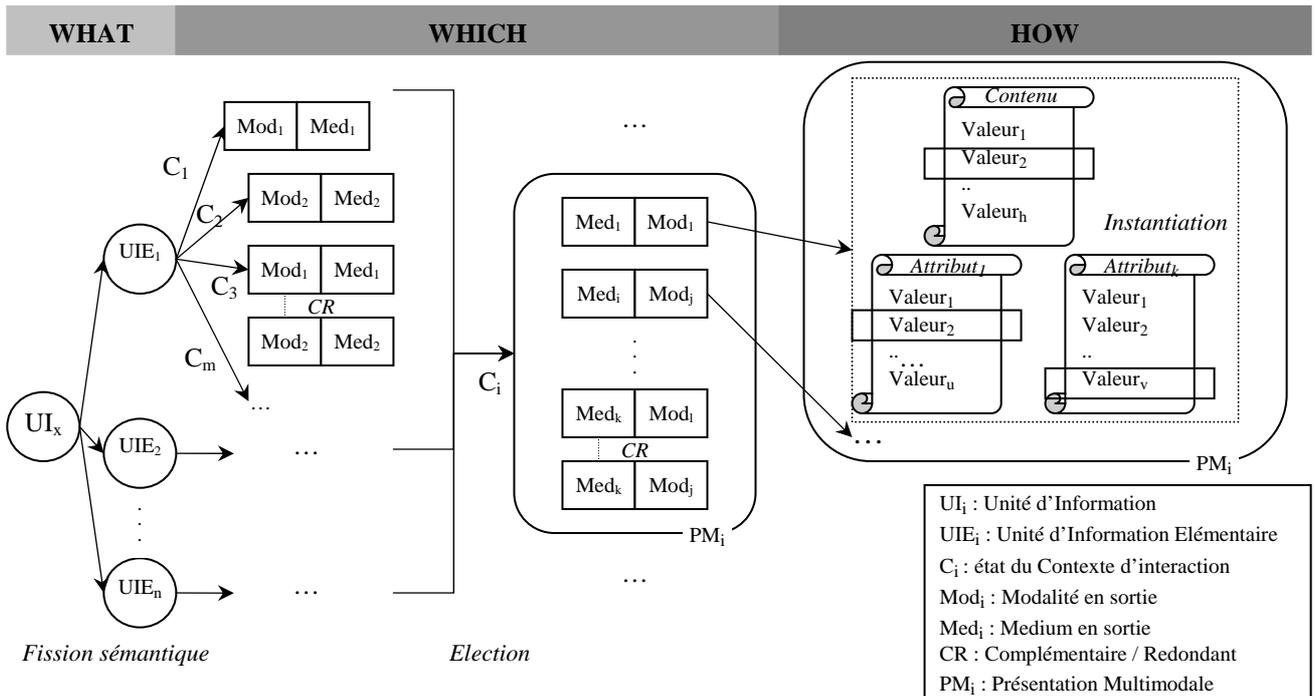


Figure 1 : Conception d'une présentation multimodale adaptée au contexte d'interaction.

La figure 1 présente le processus de conception d'une présentation multimodale adaptée au contexte d'interaction. Les différents points et la terminologie associée sont explicités dans les sections suivantes.

What : Quelle information présenter ?

Il est tout d'abord nécessaire de décomposer l'information sémantique (Figure 1, UI_i) en unités d'information élémentaires (Figure 1, UIE_i) devant être présentées à l'utilisateur. Prenons l'exemple de la réception d'un appel sur un téléphone portable. L'unité d'information est alors « appel de X » pouvant être décomposée en deux unités d'information élémentaires : l'évènement (appel) et l'identité de l'appelant (X).

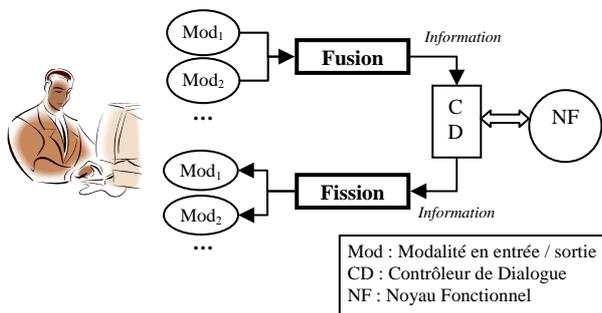


Figure 2 : Le parallèle Fusion / Fission.

Certains auteurs emploient le terme « fission » par opposé au terme « fusion » (Figure 2) pour qualifier le processus de sélection des modalités en sortie. Il ne nous semble pas pertinent d'utiliser ce terme pour désigner le processus de sélection. Une fission en sortie a effectivement 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 élémentaires et d'*allocation* lors de la sélection des modalités (Figure 3).

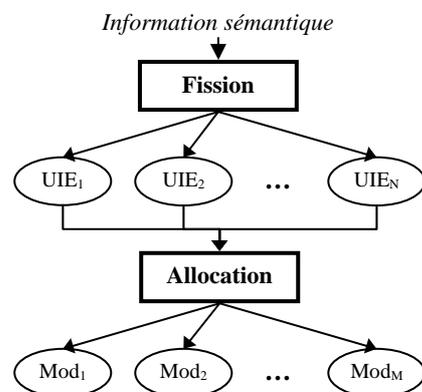


Figure 3 : La fission sémantique.

A l'état actuel, la fission sémantique est réalisée de façon manuelle par les concepteurs. Son automatisation est un problème délicat mais néanmoins intéressant comme perspective de recherche à long terme.

Which : Quelles modalités choisir ?

Suite à la décomposition de l'information, une présentation doit être allouée [11]. L'allocation consiste à sélectionner pour chaque unité d'information élémentaire, une présentation multimodale (Figure 1, PM_i) adaptée à l'état courant du contexte d'interaction (Figure 1, C_i) puis à les regrouper en une même présentation symbolisant l'expression de l'unité d'information initiale.

Ce processus de sélection en fonction du contexte d'interaction s'appuie sur un *modèle comportemental*. Ce modèle spécifie les composants d'interaction (modes, modalités et média) adaptés à un état donné du contexte. Le mode Auditif est par exemple peu adapté à un environnement bruyant à l'inverse du mode Visuel.

La formalisation du modèle comportemental peut être réalisée de plusieurs façons : arbre de décision, règles d'adaptation [11], automate, réseau de Pétri, etc. Dans nos travaux nous avons choisi une formalisation à base de règles. Ce formalisme a l'avantage de proposer un raisonnement simple limitant le coût d'apprentissage. Ce choix introduit néanmoins des problèmes de cohérence et de complétude de la base de règles.

Les règles d'adaptation [11] sont centrées sur la notion de but de conception et proposent des mécanismes de composition de modalités implicites limitant leur réutilisation. Nous avons distingué trois types de règles : contextuelles, CARE et critérielles. Les prémisses d'une règle décrivent un état du contexte d'interaction (par exemple un niveau de bruit supérieur à 90 dB). Les conclusions d'une règle contextuelle définissent le poids contextuel (l'intérêt vis-à-vis du contexte décrit en pré-misse) des composants d'interaction cibles de la règle (par exemple une valeur négative pour le mode Auditif). Les règles de type CARE permettent l'allocation de présentations composées de plusieurs couples modalité-média sur des critères de redondance et/ou de complémentarité [9]. Enfin les règles critérielles permettent la sélection d'un ensemble de modalités à partir d'un critère global de modalité donné (langagier, analogique [6], confidentialité, etc.).

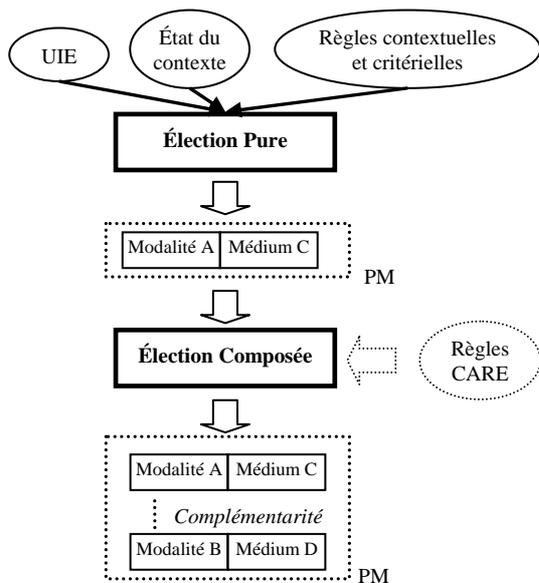


Figure 4 : Le processus d'élection.

En utilisant une analogie avec le monde politique, nous appelons le processus de sélection des modalités : *élec-*

tion. Notre processus d'élection s'appuie sur une base de règles (électeurs) qui une fois appliquée ajoute ou retire des points (voix) à certains modes, modalités ou média en sortie (candidats) selon l'état courant du contexte d'interaction (situation économique, sociale, etc.).

L'application des règles contextuelles et critérielles définit ce que l'on appelle l'élection pure alors que l'application des règles de type CARE correspond à l'élection composée (Figure 4). L'élection pure choisit le meilleur couple modalité-média exprimant une UIE donnée tandis que l'élection composée enrichit la présentation en sélectionnant de nouveaux couples redondants ou complémentaires au premier.

How : Comment présenter cette information ?

Une fois la présentation allouée, celle-ci est instanciée [1]. L'instanciation consiste à déterminer les contenus lexico syntaxiques et les attributs morphologiques des modalités. Dans un premier temps, un contenu concret à exprimer à travers les modalités de la présentation est choisi. Puis en cohérence avec ce contenu, les attributs de la présentation (attributs des modalités, paramètres spatial et temporel, etc.) sont fixés.

Reprenons l'exemple de la réception d'un appel sur un téléphone portable. Supposons par exemple 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-Ecran introduit la personne à l'origine de l'appel (deuxième UIE). Une instanciation possible du contenu concret de cette présentation pourrait être la musique de la panthère rose (modalité Sonnerie) et un portrait de la personne à l'origine de l'appel (modalité Photographie). L'instanciation des attributs morphologiques de la présentation serait par exemple une durée de 45 secondes associée à un volume moyen pour la sonnerie et l'affichage du portrait en plein écran.

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 [2] ou la synthèse de gestes [12].

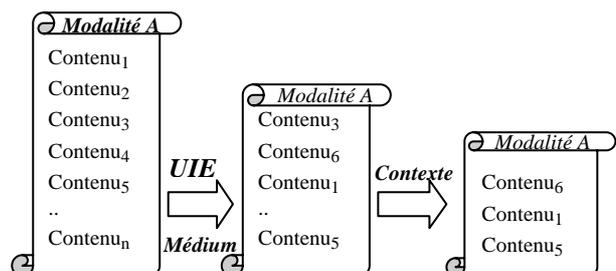


Figure 5 : Sélection du contenu d'une modalité A.

Un processus de sélection du contenu utilisé dans nos travaux est présenté en figure 5. A chaque modalité est associé l'ensemble des contenus possibles. La ou les unités d'informations élémentaires à exprimer, le médium élu ainsi que l'état du contexte d'interaction [13] permettent de réduire et de classer cette liste afin de déterminer le contenu le plus adapté à cette modalité.

Concernant les attributs morphologiques, 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'instanciations d'attributs au lieu de l'ensemble des instances possibles pour chaque attribut de la modalité. Il reste alors à sélectionner le meilleur modèle d'instanciation des attributs selon le contenu choisi, la ou les information(s) élémentaire(s) à présenter, le médium élu et l'état du contexte. Ces mécanismes s'appuient sur un système de classification des possibilités. Une formalisation similaire à la formalisation employée pour le problème de la sélection des modalités peut être envisagée.

Vérification de la cohérence de la présentation

A ce processus de conception d'une présentation multimodale (*What-Which-How*), des mécanismes de vérification de la cohérence sont associés. La figure 6 présente les différents types d'incohérence pouvant être soulevés durant l'application du modèle.

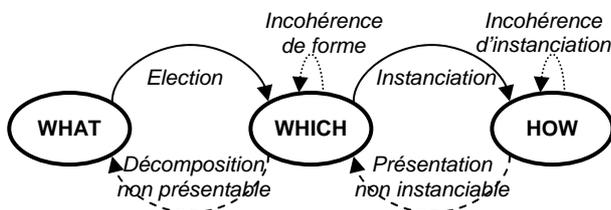


Figure 6 : Cohérence de la présentation multimodale.

Une première incohérence intitulée incohérence de forme, indique un problème d'ordre structurel sur la présentation multimodale fraîchement élue (*Which*). Certaines modalités peuvent en effet être élues pour exprimer plusieurs unités d'information élémentaires en une même présentation, ce qui n'est pas toujours réalisable. Un compromis doit alors être établi afin de garantir la cohérence de la présentation globale. Ce compromis peut aller d'une réélection des UIEs incriminées à une remise en cause des résultats de la fission sémantique (*What*).

Une seconde incohérence intitulée incohérence d'instanciation, indique un problème d'instanciation de la présentation (*How*). La sélection du contenu et/ou du modèle d'instanciation des attributs de la présentation n'a pu être menée à terme faute de choix. Ce blocage se traduit par une remise en cause du contenu choisi voir de la forme de la présentation élue. Prenons le cas d'un utilisateur non voyant. Les résultats des deux élections ex-

primant les UIEs « appel » et « appelant » sont dans les deux cas la modalité Sonnerie. Si aucune sonnerie personnalisée n'est associée à la personne à l'origine de l'appel, nous sommes dans le cas où la présentation ne peut être instanciée faute de contenu adéquat. On parle alors d'incohérence d'instanciation remettant en cause la présentation élue (*Which*).

Then : Comment faire évoluer la présentation ?

Nous venons de voir comment présenter une information et plus précisément comment adapter une présentation multimodale à un état donné du contexte d'interaction. Cependant le contexte d'interaction peut être sujet à évolution ce qui soulève un problème de préemption 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 affecte principalement les présentations persistantes, une évolution du contexte étant peu probable (mais pas impossible) dans le cas de présentations ponctuelles. De plus, toutes les évolutions du contexte ne nécessitent pas nécessairement une vérification de la préemption des présentations en cours. L'évolution doit porter uniquement sur les éléments du contexte susceptibles d'influer directement sur la présentation en cours. Ces éléments sont nécessairement des critères intervenant dans les prémisses des règles du modèle comportemental ayant données naissance à la présentation en cours. Une évolution du contexte sur un de ces critères peut alors changer l'application du modèle comportemental. La présentation est alors périmée (inadaptée au contexte d'interaction en cours) et doit être invalidée pour subir une mise à jour.

Cette invalidation ne concerne pas toujours la présentation dans sa globalité mais peut affecter un ou plusieurs éléments de la présentation. Une réélection partielle permet alors de mettre à jour la présentation en fonction du nouveau contexte à moindre coût.

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 de gestion des présentations persistantes.

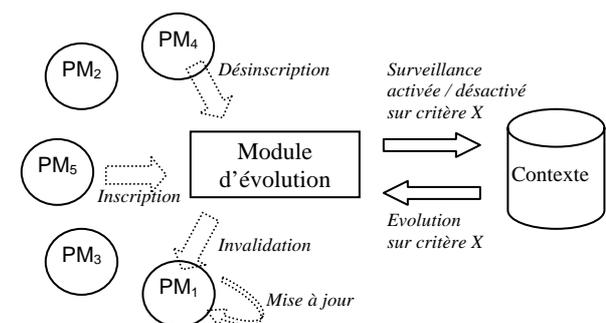


Figure 7 : Evolution des présentations multimodales.

Dans nos travaux, nous avons utilisé une première approche que nous qualifions de « centralisée ». Celle-ci consiste à établir un module d'évolution des présentations (Figure 7) [14]. Toute nouvelle présentation est alors inscrite à ce module. Cette inscription enclenche une surveillance des critères du contexte garantissant sa validité. L'évolution d'un de ses critères provoque une invalidation de la présentation et l'application des mécanismes de mise à jour. Enfin, la fin d'une présentation se traduit par une désinscription du module d'évolution.

Une seconde approche dite « distribuée » est en cours d'étude. Selon cette approche, l'expression d'une information ne produit plus une présentation multimodale mais un modèle de présentation (Figure 8). Ce modèle de présentation est un automate à transitions à deux niveaux (global et local) dans lequel chaque état global symbolise une présentation et où chaque transition (globale ou locale) est étiquetée par un facteur d'évolution. L'état courant (Figure 8, PM_1) représente la présentation initiale soit l'expression originelle de l'information. Les autres états (Figure 8, PM_2 et PM_3) représentent toutes les autres alternatives de présentation.

Une évolution du contexte, une action de l'utilisateur sur la présentation ou bien encore un paramètre temporel (durée), autrement dit tout élément pouvant nécessiter une mise à jour de la présentation peuvent être directement pris en compte par le modèle de présentation par l'intermédiaire des étiquettes de transition (locale ou globale). La présentation courante peut alors subir deux types d'évolution. La première intitulée « raffinement » (changement d'état local) ne change pas les composants d'interaction (modalités, média) utilisés par la présentation mais juste leurs instanciations (Figure 8, I_1). L'augmentation du volume d'un son est un exemple de raffinement de la modalité Sonnerie. La seconde intitulée

« mutation » est un changement radical des modalités et/ou média de communication se traduisant par un changement d'état global. Le passage d'une modalité Icône aux modalités Texte et Earcon pour indiquer un niveau de batterie faible est un exemple de mutation.

A la différence de l'approche « centralisée », une présentation ne connaît pas l'état global de l'interaction, plus précisément l'état courant des autres présentations. Cette ignorance peut se traduire par un problème d'incohérence entre les présentations. Un modèle diffusant les informations relatives à l'état courant de chaque présentation et accessible par tous doit être mis en place afin de garantir la cohérence de l'interaction en sortie.

Cette approche a l'avantage de rendre autonome la présentation d'une information. Elle a cependant un coût particulièrement important sachant que toutes les présentations quelque soit le contexte d'interaction doivent être élues en même temps. De plus la complexité de la modélisation du contexte influence directement sur la taille du modèle de présentation.

APPLICATION

Nous allons maintenant reprendre l'exemple du téléphone portable et expliciter de manière plus approfondie l'application de notre modèle.

Spécification des sorties

La spécification des sorties d'un système multimodal peut être divisée en quatre phases :

- Spécification des composants d'interaction
- Modélisation du contexte d'interaction
- Spécification des unités d'information
- Conception d'un modèle comportemental.

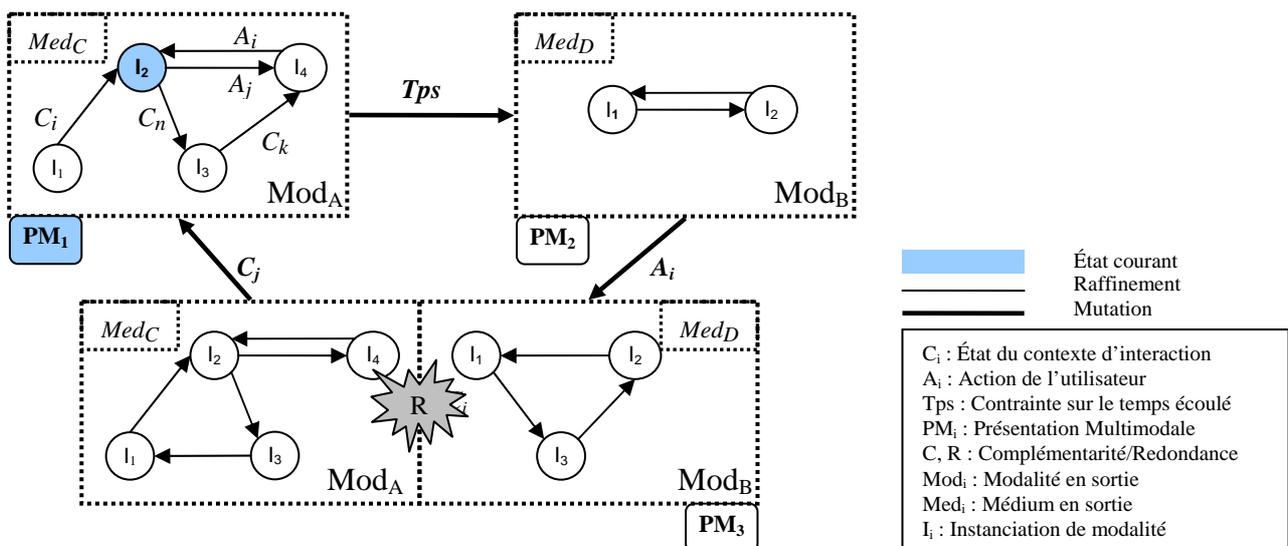


Figure 8 : Modèle de présentation d'une Unité d'Information.

La figure 9, présente le diagramme des composants d'interaction (modes, modalités et média en sortie) gérés par un téléphone portable. Le contexte d'interaction est restreint à quatre critères : « utilisateur malentendant » (modèle de l'utilisateur), « disponibilité de l'oreillette », « mode du téléphone » (modèle du système) et « niveau de bruit ambiant » (modèle de l'environnement).

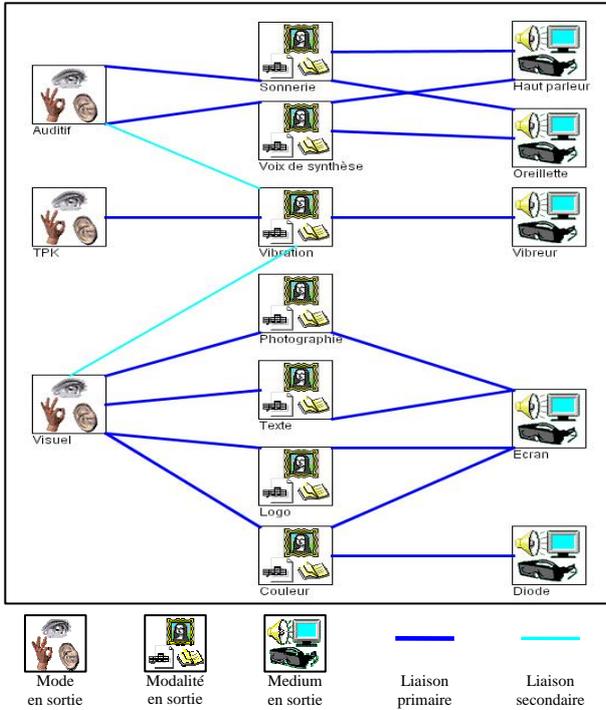


Figure 9 : Diagramme des composants d'interaction (réalisé sous MOSTe [15]).

L'unité d'information étudiée est « Appel de X » (réception d'un nouvel appel). Enfin, le tableau 1 présente six règles du modèle comportemental dont quatre de type contextuel (R1-R4), une de type CARE (R5) et une de type critériel (R6).

Id	Description en langage naturel
R1	Si l'utilisateur est malentendant Alors ne pas utiliser le mode Auditif
R2	Si l'oreillette est disponible Alors ne pas utiliser le médium Haut-parleur
R3	Si le niveau de bruit est supérieur à 80 dB Ou que le téléphone est en mode silencieux Alors éviter le mode Auditif
R4	Si l'UIE est un « appel » Alors favoriser les modalités Sonnerie et Vibration
R5	Si l'UIE est un « appel » Et le téléphone est en mode augmenté Alors utiliser de la Redondance
R6	Si l'UIE est un « appelant » Alors favoriser les modalités fortement Analogiques

Tableau 1 : Exemple de règles du modèle comportemental.

Simulation des sorties

La figure 10 présente un aperçu des résultats de l'application du modèle pour l'information « Appel de X ». L'exemple d'évolution proposé est centré sur l'unité d'information élémentaire « Appel ». Cette évolution a pour but de renforcer de manière progressive et au fil du temps le signalement de l'information à l'utilisateur. Un premier raffinement accentue la puissance de la vibration. Il est suivi d'une mutation permettant le passage de la vibration à la sonnerie. Enfin, un dernier raffinement augmente le volume de la sonnerie.

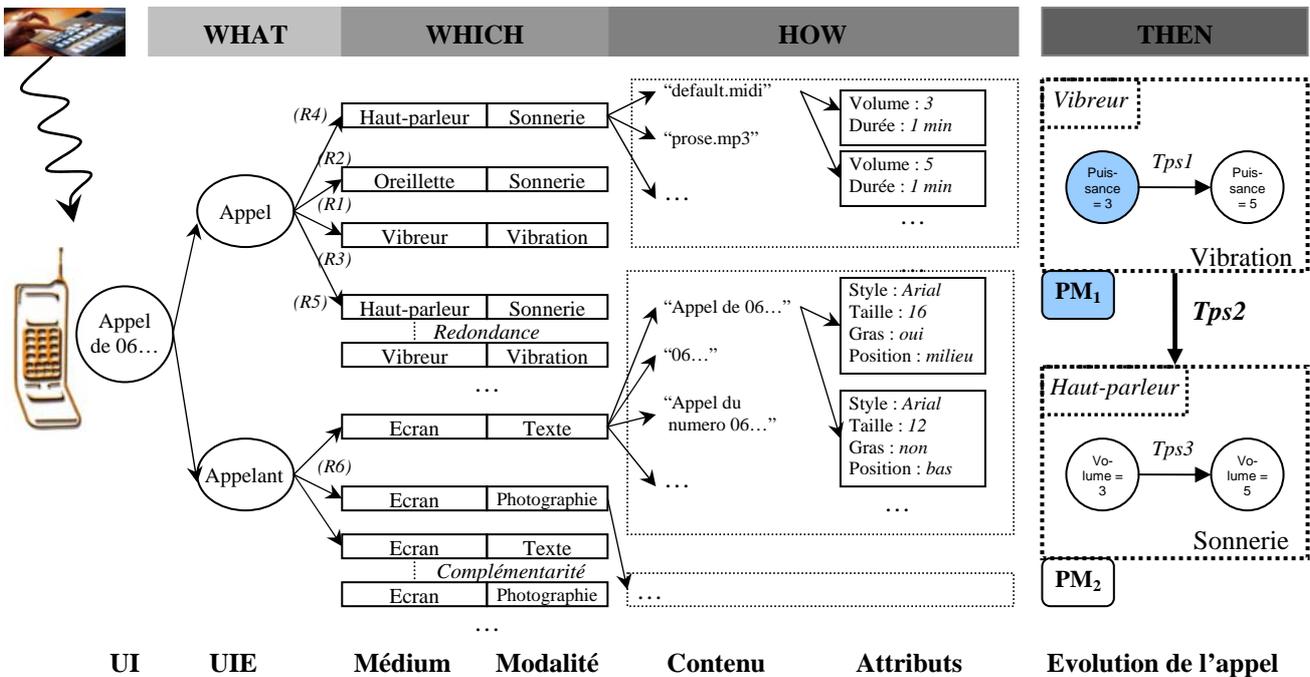


Figure 10 : Application du modèle dans le cadre de la réception d'un appel téléphonique.

CONCLUSION ET PERSPECTIVES

Dans ce papier, nous avons présenté un modèle conceptuel pour la présentation multimodale d'information. Ce modèle intitulé WWHT s'articule autour de quatre notions (*What-Which-How-Then*) présentant le cycle de vie d'une présentation multimodale réactive au contexte d'interaction en cours. Les trois premiers points (*What-Which-How*) sont relatifs à la conception d'une présentation adaptée. Le dernier point (*Then*) s'attarde sur le devenir de cette présentation, à ses évolutions au cours du temps.

Ce modèle offre un cadre pour la conception et la gestion des sorties d'un système multimodal en fonction du contexte d'interaction, mais indépendamment des entrées du système. La continuité de l'interaction ne pourra être obtenue sans l'incorporation des entrées dans le processus de conception de la présentation en sortie.

Des travaux sont également en cours sur l'évolution des présentations (*Then*) afin de comparer les deux approches identifiées (« centralisée » / « distribuée ») sur ce sujet et d'en tirer des recommandations quant aux conditions de l'usage d'une approche plutôt que l'autre.

Enfin, une plate forme sur les bases de ce modèle est en cours d'implémentation. Cette plate forme intitulée MOST (Multimodal Output Specification / Simulation Tool) permet à ce jour de spécifier des sorties d'un système multimodal et d'en simuler l'élection et la gestion.

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).

BIBLIOGRAPHIE

1. André, E. The generation of multimedia presentations. R. Dale, H. Moisl and H. Somers (eds.). In *Techniques and Applications for the Processing of Language as Text*, 2000, pp. 305-327.
2. André, E. Natural Language in Multimedia/Multimodal Systems. Mitkov, R. (ed.). *Handbook of Natural Language Processing*, 2003, pp. 650-669, Oxford University Press.
3. A Metamodel for the Runtime Architecture of an Interactive System. *SIGCHI Bulletin*, The UIMS Tool Developers Workshop, Vol. 24, 1992, pp. 32-37.
4. Arens, Y, Hovy, E. H. The Design of a Model-Based Multimedia Interaction Manager. *AI Review*, Vol. 9, No. 3, 1995.
5. Bellik, Y., Interfaces Multimodales : Concepts, Modèles et Architectures, *Thèse de Doctorat*, Université Paris XI (Orsay), 1995.
6. Bernsen, N.O. A Reference Model for Output Information in Intelligent Multimedia Presentation Systems. In *Proceedings of the ECAI'96*, 1996.
7. Bordegoni, M., Faconti, G., Maybury, M.T., Rist, T., Ruggieri, S., Trahanias, P. and Wilson, M. A Standard Reference Model for Intelligent Multimedia Presentation Systems. *Computer Standards & Interfaces*, Vol. 18, No. 6-7:, 1997, pp. 477-496.
8. Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonckt, J. A unifying reference framework for multi-target user interfaces. *Journal of Interacting With Computer*, Elsevier Science B.V, Vol. 15, No. 3, 2003, pp. 289-308.
9. Coutaz, J., Nigay, L., Salber, D., Blandford, A., May, J., and Young, R.M. Four Easy Pieces for Assessing the Usability of Multimodal Interaction: the CARE Properties. In *Proceedings of INTERACT'95*, Lillehammer, Norway, 1995.
10. Dey, A. K., Salber, D. and Abowd, G.D. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. Moran, T.P. and Dourish, P. (eds.). *Context-Aware computing: a special triple issue of Human-Computer Interaction*, 2002.
11. Karagiannidis, C., Koumpis, A., and Stephanidis, C. Deciding 'What', 'When', 'Why', and 'How' to adapt in intelligent multimedia presentation systems. In *Proceedings of ECAI workshop on Towards a standard reference model for intelligent multimedia presentation systems*, 1996.
12. Lejeune, F. Analyse sémantico-cognitive d'énoncés en Langue des Signes Française pour une génération automatique de séquences gestuelles. *Thèse de Doctorat*, Université Paris XI (Orsay), 2004.
13. Lemlouma, T., Layaida, N. Context-Aware Adaptation for Mobile Devices. *Mobile Data Management*, Berkeley, California, USA, 2004.
14. Rousseau, C., Bellik, Y., Vernier, F. and Bazalgette, D. Architecture framework for output multimodal systems design. In *Proceedings of OZCHI'04*, Australia, Wollongong, 2004.
15. Rousseau, C., Bellik, Y., Gruaz, R., Etienne, F. et Vernier, F. Fourniture F12.3 : Manuel Utilisateur de MOSTe. *PEA INTUITION – PHASE 2*, Réf.: CDS/ET/TT/04-158 I.R.01.
16. Satyanarayanan, M. Pervasive Computing: Vision and Challenges. *IEEE Personal Communications*, Vol. 8, No. 4, 2001, pp. 10-17.
17. Thevenin, D., Adaptation en Interaction Homme-Machine : le cas de la Plasticité. *Thèse de Doctorat*, Université Joseph Fourier, Grenoble I, 2001.