



Démarche centrée utilisateur pour la conception de SIAD basés sur un processus d'ECD, application dans le domaine de la santé

Hela Ltifi, Mounir Ben Ayed, Christophe Kolski, A.M. Alimi

► To cite this version:

Hela Ltifi, Mounir Ben Ayed, Christophe Kolski, A.M. Alimi. Démarche centrée utilisateur pour la conception de SIAD basés sur un processus d'ECD, application dans le domaine de la santé. *Journal d'Interaction Personne-Système, Association Francophone d'Interaction Homme-Machine (AFIHM)*, 2010, 1 (1), pp.1-25. hal-01058947

HAL Id: hal-01058947

<https://hal.inria.fr/hal-01058947>

Submitted on 28 Aug 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Démarche centrée utilisateur pour la conception de SIAD basés sur un processus d'ECD, application dans le domaine de la santé

HELA LTIFI (1,2,3,4)

MOUNIR BEN AYED (1)

CHRISTOPHE KOLSKI (2,3,4)

ADEL M. ALIMI (1)

(1) REGIM : REsearch Group on Intelligent Machines, Ecole Nationale d'Ingénieurs de Sfax, Route Sokra Km 3.5 - BP W - 3038 Sfax, Tunisie

(2) Univ Lille Nord de France, F-59000 Lille, France

(3) UVHC, LAMIH, F-59313 Valenciennes, France

(4) CNRS, FRE 3304, F-59313 Valenciennes, France

Résumé : Cet article s'inscrit dans le cadre de la conception de systèmes décisionnels centrés utilisateur basés sur un processus d'Extraction de Connaissances à Partir des Données (ECD). Ce processus doit déboucher sur un ensemble de modules, il est itératif et interactif. De ce fait il nous semble indispensable de prendre en compte des principes et des méthodes de l'Interaction Homme-Machine dans le développement de tels systèmes. A ce sujet, le développement de systèmes décisionnels interactifs est actuellement abordé suivant deux approches antagonistes : la première est "technocentrée", dans laquelle la technologie est fondamentale ; la deuxième est "centrée utilisateur" plaçant les acteurs humains en position centrale. Bien que la première approche soit encore très présente en entreprise, la tendance actuelle est résolument "centrée utilisateur". Dans ce cadre, nous proposons une approche qui vise à intégrer les étapes du processus ECD dans un modèle de développement enrichi sous l'angle des interactions homme-machine appelé le modèle en U. Notre contexte applicatif est la lutte contre les infections nosocomiales en milieu hospitalier.

Mots clés : Système Interactif d'Aide à la Décision (SIAD), Extraction de connaissances à partir de données (ECD), Interaction Homme-Machine (IHM), Modèle en U.

Abstract: This article concerns the design of Decision Support Systems (DSS) based on a Knowledge Discovery from Data (KDD) process. This process aims at generating a set of software modules; it is iterative and interactive. For this reason, it is essential to us to take Human-Computer Interaction principles and models into account in the development of such systems. The interactive decisional system development is currently approached according to two antagonistic approaches. For the first approach, technology is fundamental; the second is "user-centered", placing the human actors in a central position. Although the first approach is still very present in companies, the current tendency "is user-centered". In this context, we propose an approach which aims at integrating the stages of KDD process in a development model enriched under the human-computer interaction point of view, which is the U-model. Our application context is the fight against the nosocomial infections in the healthcare domain.

Key words: Decision support system (DSS), Knowledge Discovery from Data (KDD), Human-Computer Interaction (HCI), U-model.

1. INTRODUCTION

Parmi les domaines les plus riches en matière de données stockées on peut citer le milieu hospitalier ; en particulier, les services de réanimation représentent une mine de données encore trop peu exploitée actuellement. Les infections nosocomiales représentent un des problèmes majeurs dans un service de réanimation. Ce sont des infections contractées dans un établissement de soins. Pour lutter contre la survenue de ces infections, le service de réanimation a besoin d'un système lui permettant de l'aider à prendre des décisions tout au long du séjour d'hospitalisation des patients [Kawamoto et al. 2005]. Les systèmes d'aide à la décision seraient appropriés au problème des infections nosocomiales [Guigue et Donadey 1999]. On parle aujourd'hui des systèmes **interactifs** d'aide à la décision facilitant l'accès aux données en ouvrant la possibilité à des analyses plus complètes [Pelayo et al. 2009]. Et au vu du nombre croissant de grandes bases de données médicales, organiser les données, les interpréter et en extraire des informations utiles pour l'aide à la décision est en fait un problème ouvert.

De ce fait, nos recherches s'inscrivent dans le cadre de la conception centrée utilisateur de Systèmes Interactifs d'Aide à la Décision (SIAD) basés sur un processus d'Extraction de Connaissances à partir des Données (ECD). La prise de décision permet d'opérer un choix entre diverses alternatives possibles pour résoudre un problème en choisissant le meilleur compromis parmi elles. Le SIAD traite le problème en fonction de connaissances décisionnelles. Certaines de ces connaissances peuvent être extraites à l'aide d'un outil décisionnel qu'est la fouille de données (appelée aussi Data Mining) [Fayyad et al. 1996] [Jambu 1999] [Lefébure et Ventari 2001] qui peut être vue comme une étape d'un processus plus global qui est celui de l'ECD. Il faudra donc interpréter cette information nouvelle, l'évaluer pour enfin la proposer comme élément valide d'aide à la décision.

Les SIAD consistent à assister les décideurs tout au long de ce processus, en particulier au moment du choix des études à effectuer ou des outils à utiliser. La coopération Homme-Machine est primordiale dans l'aide à la décision ; elle concerne l'assistance du décideur humain et donc la possibilité de partager les tâches entre humain et machine [Pacaux-lemoine 1998] [Tricot 2005]. Sous cet angle, l'interaction homme-machine est un aspect crucial au niveau des systèmes interactifs d'aide à la décision, leur conception devant nécessairement s'appuyer sur une démarche centrée utilisateur. L'évaluation des IHM vient afin de valider la qualité de cette modélisation ou d'énoncer les problèmes rencontrés sous l'angle particulièrement de l'utilité et de l'utilisabilité des systèmes [Nielsen 1993].

Dans le cadre de la conception des SIAD, on se rend compte qu'il s'agit de concilier et d'intégrer des méthodes et modèles provenant d'une part de ce domaine et d'autre part de celui de l'IHM [Lepreux 2005] [Lepreux 2007]. Il existe à ce sujet plusieurs modèles enrichis sous l'angle des IHM tels que le modèle en étoile [Hartson et Hix 1989], le modèle nabla [Kolski 1997] [Kolski 1998] ou le modèle en U [Abed 2001]. Dans cet article étendant [Ltifi et al. 2008], l'idée est d'adopter le modèle en U comme démarche de conception et de l'adapter dans les systèmes exploitant un processus d'ECD.

Notre article débute par un bref état de l'art présentant deux concepts fondamentaux que sont les SIAD et l'ECD, et ceci afin de préciser la vision que nous en avons. Par la suite nous présentons le modèle en U. La troisième partie vise à proposer notre approche supportant la conception et le développement des SIAD basés sur un processus ECD. Avant de conclure, nous présentons la réalisation de notre SIAD où nous donnons des

exemples concrets et des extraits d'interfaces Homme-Machine (relatifs à des modules interactifs sous-jacents au processus d'ECD) provenant du déroulement de l'application de notre approche proposée.

2. SYSTEMES INTERACTIFS D'AIDE A LA DECISION BASES SUR UN PROCESSUS D'ECD

Les systèmes décisionnels centrés utilisateurs se situent à la confluence de trois domaines que sont l'aide à la décision, l'Extraction de Connaissances à partir des Données (ECD) et l'Interaction Homme Machine (IHM). Nous pouvons citer dans ce contexte le travail de [Chevrin et al. 2007] qui met l'accent sur la place de l'IHM dans les processus d'ECD.

2.1. Les Systèmes Interactifs d'Aide à la Décision

Il existe dans la littérature, différents systèmes d'aide à la décision dans le domaine médical [Kaplan 2001] [Matheny et Ohno-Machado 2007] [Evans 2007]. Très peu de ces systèmes accompagnent le décideur durant la totalité du processus de décision. Ils sont pour la plupart des systèmes, "intelligents" ou non, cherchant à optimiser des solutions [Kawamoto et al. 2005] sans être réellement interactifs tout en offrant peu de possibilité d'intervention à l'utilisateur. Ces systèmes informatiques pour l'aide à la décision doivent fournir de façon synthétique et simple les éléments nécessaires aux décideurs pour évaluer une situation ; il nous semble indispensable dans ce contexte de mettre en relief l'aspect interactif et de viser à construire des IHM adaptées aux besoins de chaque type d'utilisateur. C'est le **I** de SIAD, objet de cette section.

De ce fait, nous voyons les SIAD comme des systèmes qui, à travers un dialogue **homme-machine**, permettent l'identification, l'exploration et la résolution de problèmes par un utilisateur (décideur) et/ou groupe d'utilisateurs (décideurs). La communication entre les différents acteurs humains d'un système d'aide à la décision (décideurs, consultants, experts, etc.) s'avère primordiale [Turban 1993]. La notion d'interactivité dans un SIAD renvoie au rôle indispensable de l'homme dans son fonctionnement, rôle non passif qui sous-tend le terme "Aide à la Décision", mais aussi à la qualité de l'intégration des différents composants du système et à la nature de l'interface homme-machine jouant "le rôle de collaborateur" avec le décideur [Garlatti 1996] [Lévine et Pomerol 1989]. Un SIAD se déroule selon un processus composé de trois phases [Klein et Methlie 1990] [Lévine et Pomerol 1989] [Simon 1977] [Turban 1993] : (1) la recherche d'information ou renseignement : l'identification du problème à résoudre nécessite de rechercher les informations pertinentes en fonction des besoins du décideur ; cette phase se termine par un énoncé du problème à traiter ; (2) la conception comprenant la génération, le développement et l'analyse des différentes suites possibles de solutions ; (3) lors de la phase de choix, le décideur choisit entre les différentes solutions identifiées pendant la phase précédente. Cette phase inclut la recherche, l'évaluation et la recommandation d'une solution appropriée.

Dans la littérature sur les systèmes d'aide à la décision, le but d'un système est avant tout de nature pratique. Il s'agit en fait d'assister un décideur en mettant à sa disposition les connaissances dont il a besoin pour résoudre certains cas difficiles. Pour la résolution de certains problèmes le processus de décision est lié donc à la connaissance [Lepreux et al. 2003]. Celle-ci peut être extraite par un outil décisionnel, on parle de l'extraction de connaissances à partir des données (ECD).

2.2. L'extraction de connaissances à partir de données

Dans le domaine des SIAD on considère que l'utilisation d'outils pour extraire des connaissances améliore la pertinence des solutions proposées pour une bonne prise de décision [Fayyad et al. 1996] [Jambu 1999] [Sandoval 1997]. On s'intéresse donc à la technologie la plus répandue de découverte de connaissances, qu'est l'Extraction de Connaissances à partir des Données [Fayyad et al. 1996] [Jambu 1999] (ECD = "KDD", Knowledge Discovery in Databases, ou "data mining" en anglais, traduit par fouille de données, FD, en Français).

Le but de l'ECD est de pouvoir extraire un élément : la connaissance. Pour le définir, on peut dire aussi que l'on vise l'"Extraction des connaissances nouvelles, utiles, valides à partir d'une masse de données" [Fayyad et al. 1996]. On considère que l'ECD est un processus interactif et itératif. L'interactivité est liée aux différents choix que l'utilisateur est amené à effectuer. L'itérativité est liée au fait que l'ECD se déroule suivant une suite d'étapes et que l'utilisateur peut décider de revenir en arrière à tout moment si les résultats ne lui conviennent pas (figure 1).

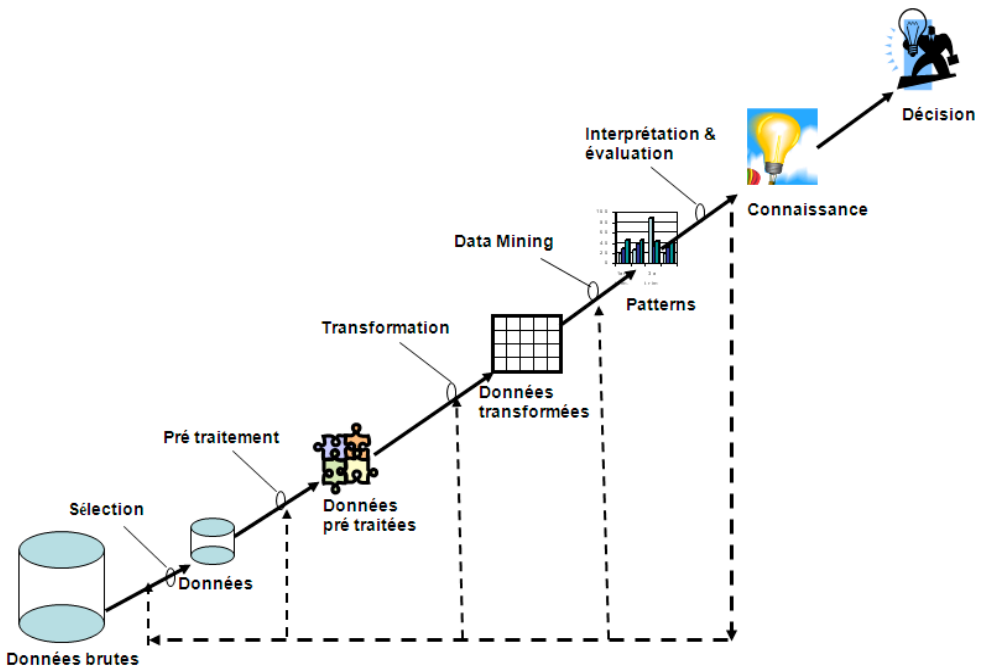


Figure 1 : Le processus d'ECD

Les différentes étapes du processus d'ECD sont les suivantes [Lefebvre et Ventari 2001] : (1) positionnement du problème en cernant les objectifs, définissant les cibles et vérifiant les besoins ; (2) recherche des données (identifier les informations et les sources, vérifier leur qualité ainsi que leur facilité d'accès) ; (3) sélection des données ayant un rapport avec l'analyse demandée dans la base ; (4) nettoyage des données en vue de corriger les inexactitudes ou erreurs de données ; (5) transformation des données dans un format qui les prépare au Mining (transformer dates en durées, ratios, etc.) ; (6) fouille de données, application d'une ou plusieurs techniques (réseaux de neurones, réseaux bayésiens, arbres de décision, etc.) pour extraire des patterns

intéressants ; (7) évaluation du résultat permettant d'estimer la qualité du modèle découvert ; (8) intégration de la connaissance en implantant le modèle ou ses résultats dans le système informatique de l'entreprise.

2.3. Système Interactif d'Aide à la Décision basé sur un processus d'ECD

Un SIAD basé sur l'ECD, sujet de cet article, est un système qui permet de détecter les stratégies de résolution d'un problème de décision par le biais d'un processus de fouille de données. Dans ce processus, l'analyse des besoins des décideurs, les différentes activités réalisées en rapport avec la préparation et la manipulation des données pertinentes, de même que la visualisation des résultats constituent des étapes très importantes. C'est sur de telles étapes que repose l'acceptation ou le refus par l'utilisateur final de l'outil d'aide à la décision visé [Lefébure et Ventari 2001]. Les interactions homme-machine au niveau de ce système devraient permettre de guider les utilisateurs tout au long des étapes d'ECD ; il est important aussi d'adapter au mieux l'IHM à chaque classe de décideur et/ou à chaque décideur [Lajnef et al. 2005].

Nous proposons le schéma visible en figure 2 présentant le déroulement du processus d'un SIAD basé sur un processus ECD, ou encore SIAD basé Data Mining.

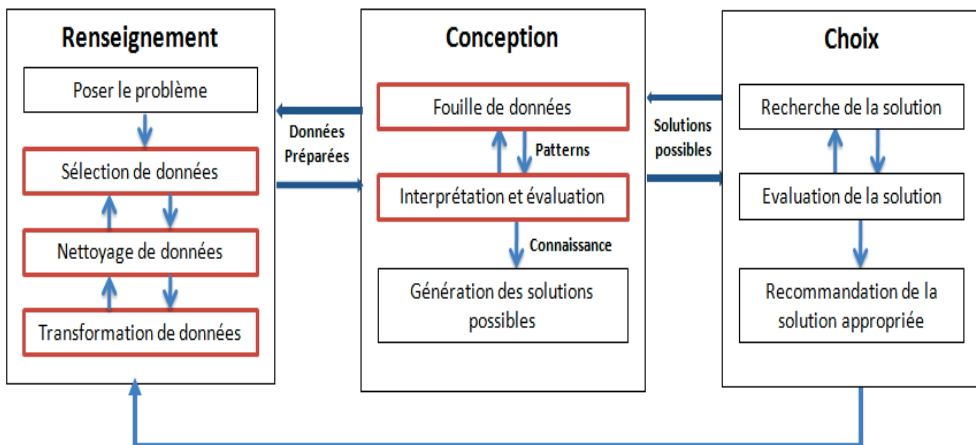


Figure 2 : processus d'un SIAD basé ECD

La figure 2 nous montre l'intégration des étapes de processus d'ECD dans le processus d'aide à la prise de décision. En effet, l'identification du problème, du processus d'ECD, permet de cerner les objectifs et définir les différents objectifs principaux du futur système. Les étapes de prétraitements consistent à construire des corpus de données spécifiques ainsi qu'à faire le nettoyage des données, le traitement des données manquantes, la sélection d'attributs ou la sélection d'instances puis la transformation de ces données. Ces étapes sont cruciales pour la recherche des informations pertinentes du processus de prise de décision [Turban 1993].

La fouille de données peut alors être opérée pour aboutir à des connaissances mises sous la forme de modèles qui doivent être validés. Des post-traitements sont nécessaires pour rendre ces modèles intelligibles soit par un humain soit par une machine [Turban 1993]. D'où la génération, l'analyse et le développement des solutions possibles au problème posé, basées sur les connaissances découvertes par le processus d'ECD.

Les premiers travaux de recherche dans le domaine ECD ont principalement porté sur le développement d'algorithmes automatiques performants. Actuellement, le rôle de l'utilisateur est devenu peu à peu un sujet de préoccupation majeur [Chevrin et al. 2007]. Ce besoin d'intégration a conduit à l'émergence de nombreux outils de visualisation et d'interaction [Fayyad et al. 2002]. Le rôle de l'utilisateur dans un processus d'ECD avait été mis en avant dès le début des années 90 dans l'article fondateur de W.J. Frawley et al. [Frawley et al. 1992]. Ces derniers insistent sur la nécessité d'intégrer l'utilisateur dans la "boucle de découverte" afin de combiner les potentialités humaines de jugement avec les capacités de calcul de la machine.

Dans ce fait, pour la modélisation et la réalisation d'un SIAD basé sur un processus interactif d'ECD, il faut fortement intégrer le facteur humain ; c'est-à-dire suivre une démarche méthodologique centrée utilisateur. Nous nous intéressons aux modèles issus du domaine des IHM qui a pour objectif de mieux prendre en compte l'utilisateur. La question qui se pose donc est la suivante : quel modèle du domaine de l'IHM faut-il choisir ?

3. LE MODELE EN U EN VUE DE SON ADAPTATION POUR LA CONCEPTION DE SIAD BASES SUR UN PROCESSUS ECD

Le développement de systèmes décisionnels interactifs, tels que les SIAD basés sur un processus d'ECD, nous amène naturellement à poser la question de la limite du système et de son interaction avec l'utilisateur. Un premier souci majeur de l'approche anthropocentrée est de travailler "sur le terrain": on n'étudie pas seulement le "travail prescrit" mais aussi le "travail réel" (au sens de [Leplat 1985]). Un autre souci est de se placer dans une logique d'utilisation et de donner à l'évaluation un rôle majeur ; l'évaluation doit pouvoir amener les concepteurs à intégrer, par itérations successives, le nouveau système décisionnel.

Du domaine de Génie Logiciel sont issus des cycles de développements classiques, de même que d'autres plus récents tels que les modèles en cascade [Royce 1970], en V [McDermid et al. 1984], spirale [Boehm 1988], en Y [André 1993] ou encore le processus unifié [Jacobson et al. 1999]. Mais ces modèles demeurent très généraux et les facteurs humains n'y sont pas explicités alors que les équipes de développement devraient être incitées à les prendre en compte. De ce constat, viennent les modèles enrichis sous l'angle de l'IHM positionnant des notions inexistantes (ou peu visibles) dans les modèles de développement qualifiés de classiques, issus du Génie Logiciel et rendant nécessaire la reconsidération de leur structure et organisation [Kolski et al. 2001].

Il existe plusieurs modèles de développement enrichis sous l'angle de l'IHM, synthétisés dans [Kolski et Loslever 1998] [Kolski et al. 2001]. Parmi ces modèles, on peut citer le modèle de [Long et al. 1990], celui de [Curtis et al. 94], en étoile [Hix et al. 1993], celui de [Collins 1995], Nabla [Kolski 1997] [Kolski 1998] ou encore en U [Abed 1990]. De tels modèles ne prétendent pas nécessairement fournir toute l'assurance qu'un projet visant la conception et le développement d'un système interactif soit réalisé avec un succès total. Le souci principal de ces modèles est surtout de souligner, du point de vue méthodologique, des aspects fondamentaux tels que la modélisation des tâches humaines, le développement itératif des prototypes et l'évaluation du système interactif [Kolski et al. 2001].

Le modèle en V fait partie des modèles les plus connus du génie logiciel. Il structure les étapes en deux phases : une descendante pour la conception et la réalisation du

système, et une ascendante pour l'intégration et l'évaluation. Les moyens d'évaluation sont définis (prévus) dans la phase descendante. Ce modèle n'est pas suffisamment spécifique pour permettre un développement efficace d'un SIAD et ne considère pas explicitement les facteurs humains. Ce modèle est un pilier du génie logiciel ; il a souvent été réutilisé et adapté dans d'autres domaines [Lepreux 2006]. Dans le domaine de l'IHM, le cycle en V a été adapté, étendu, revu pour conduire au **modèle en U** [Abed 1990] [Milot 1990] [Abed et al. 1991] [Abed 2001] [Lepreux et al. 2003] [Cathelain 2005]. Ce dernier modèle intègre des étapes centrées sur les facteurs humains et a par la suite été enrichi d'étapes permettant de mieux intégrer les experts du domaine dans la phase descendante.

De ce fait, le modèle enrichi sous l'angle de l'IHM qui nous semble le plus adéquat et qui répond aux besoins décrits dans cette section (facteurs humains, évaluation, etc.) est le modèle en U. Il a été décrit en détail dans ces différents documents ; nous nous contentons donc de le résumer ici. Comme précisé dans [Lajnef et al. 2005], nous avons choisi le modèle en U dans la mesure où il a déjà montré son applicabilité et son adaptabilité dans différents domaines complexes (contrôle aérien, applications chimiques, ferroviaires...).

Ce modèle en U situe les étapes qui n'existent pas dans les modèles classiques du génie logiciel, lesquels demeurent très généraux, tout en partant sur l'hypothèse que les facteurs humains doivent y être considérés par l'équipe de développement. Le modèle en U est structuré en deux phases (figure 3) :

- (i) **une phase descendante** qui commence par une description structurelle et fonctionnelle du système en vue de fournir un cadre structurant pour les futures activités aussi bien que pour les solutions techniques. Le résultat de cette étape est une liste des tâches devant être réparties en tâches automatiques et en tâches interactives (impliquant différents degrés de collaboration entre l'utilisateur et le système), puis analysées et modélisées. Concernant les tâches interactives, il est nécessaire de pouvoir définir l'interface homme-machine pour chacune d'entre elles en se référant au comportement probable de l'utilisateur. Il s'agit de modéliser l'aspect comportemental de l'IHM tout en prenant en compte le modèle des divers utilisateurs en termes de limites et de ressources physiques et cognitives. Le modèle utilisateur correspond à un domaine de recherche vaste et à part entière visant en particulier à comprendre le processus humain de raisonnement. Dans ce cadre, la modélisation des tâches interactives permet de révéler les besoins informationnels des utilisateurs, aussi bien que leurs besoins dans d'éventuels outils d'aide. La modélisation des tâches sert de base pour la spécification des interfaces homme-machine. Une évaluation préliminaire peut être effectuée à ce niveau pour vérifier si le modèle de tâches prescrites et le modèle de tâches réelles sont compatibles. Une fois que les besoins informationnels et les besoins d'aide ont été identifiés, il devient alors possible de définir une architecture pour le système interactif complet. Ses spécifications visent à analyser et définir, en particulier, le comportement de l'interface homme-machine. Il est très important pendant cette étape de prendre en compte un ensemble de critères ergonomiques, bien connus en dans le domaine de l'IHM (cf. à ce sujet [Bastien et Scapin 2001]). La spécification des interfaces homme-machine conduit ensuite à l'implémentation et à l'intégration du système complet.

- (ii) **une phase ascendante** focalisée sur l'évaluation du système global, selon des critères d'efficacité du système et également des critères centrés sur l'être humain. La validation consiste à confronter le modèle des tâches théoriques (prescrites) de la phase descendante avec le modèle des activités réelles mis en évidence en phase ascendante (selon les principes originaux proposés par [Abed et al. 1991]). Le résultat de cette confrontation permet de valider le système Homme-Machine ou bien de mettre en évidence ses carences ainsi qu'affiner progressivement celui-ci, particulièrement au niveau des IHM et des outils d'aide. Le modèle final résultant de la confrontation permet ainsi de généraliser des comportements spécifiques des utilisateurs dans des conditions particulières de travail. Ils sont réutilisables dans des situations relatives à des systèmes similaires. Ce modèle peut être adapté selon les caractéristiques de l'application. Par exemple, dans [Lepreux et al. 2003], il a été adapté pour des types particuliers de SIAD dans des projets dont le point central consiste en l'extraction de l'expertise des experts ; et ceci dans un but de conception de composants logiciels correspondant chacun à des modules d'aide aux activités expertes.

Le modèle en U est axé sur l'IHM, néanmoins, nous pensons qu'il est améliorable. Par exemple, il ne présente pas clairement le développement itératif et incrémental du système interactif à réaliser. Une adaptation au domaine qui nous concerne, associée à quelques améliorations, fait l'objet de la partie suivante.

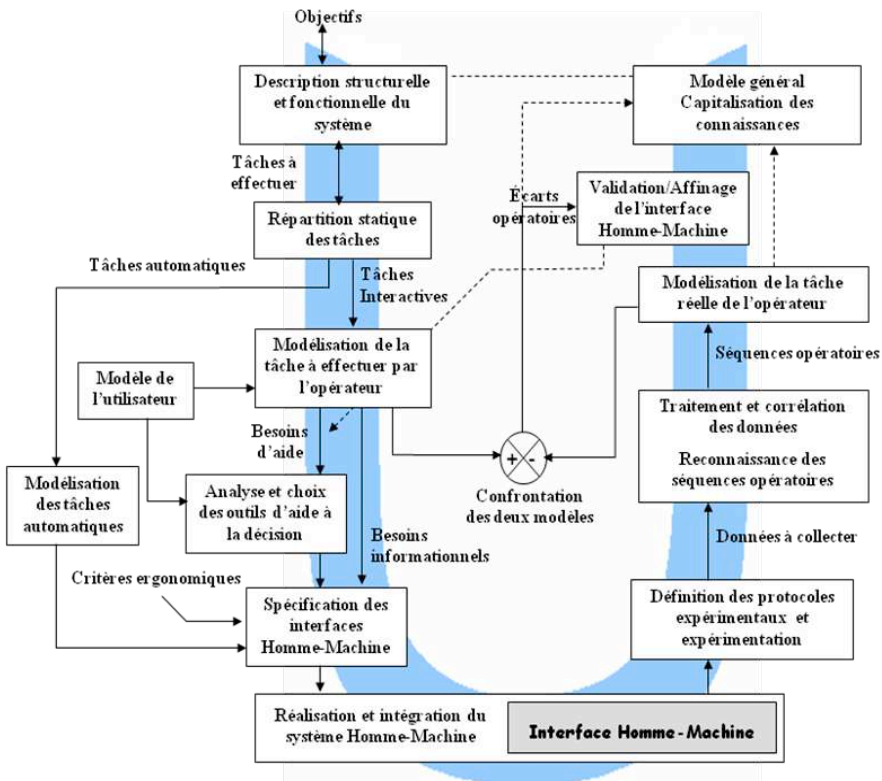


Figure 3 : Le modèle en U (version décrite dans [Abed 2001])

4. APPROCHE PROPOSEE

L'approche proposée consiste en une adaptation du modèle en U [Ltifi et al. 2008] [Ltifi et al. 2009] [Ben Ayed et al. sous presse]. Cette adaptation part du constat suivant : le modèle en U tel qu'il est présenté précédemment (figure 3), se décompose en plusieurs étapes qui montrent clairement l'intégration de l'IHM dans le processus de développement du système. Néanmoins, ce modèle possède des points possibles d'adaptation pour le contexte qui nous intéresse. Ainsi :

- (1) lors de l'étude de l'existant, il est souvent pertinent que le décideur, utilisateur potentiel du système, puisse décrire aussi bien ses caractéristiques que ses besoins fonctionnels, et évaluer et valider des premières maquettes d'interfaces en vue de montrer la manière dont il veut voir les interfaces Homme-Machines de la future application. Certaines de ces informations peuvent être utilisées pour "modéliser" le décideur (caractéristiques, préférences, stratégies...) [Robert 2003] [Seffah et al. 2009].
- (2) De plus, pour les SIAD à base d'ECD, la définition et la répartition des tâches est très importante et doit être bien mise en avant. En outre, le modèle initial ne présente pas clairement l'ordre des activités de conception (capture des besoins, analyse, conception, implémentation et test). Les étapes du modèle en U s'intéressent à la conception et à l'évaluation de la partie interactive du système tout en s'intéressant peu à l'aspect applicatif du système en question, alors que cet aspect est important pour les SIAD à base d'ECD qui concernent cette recherche.
- (3) Nous pensons aussi que l'étape "Modèle général, capitalisation des connaissances" visant à généraliser le comportement humain spécifique de l'utilisateur peut être simplifiée par un bouclage sur l'étape de modélisation de l'utilisateur et de sa tâche réelle. De ce fait nous suggérons d'intégrer ensemble les deux étapes pour une modélisation des tâches réellement effectuées.

En conséquence, le modèle en U adapté que nous proposons pour la conception et la réalisation de SIAD découlant d'un processus d'ECD est visible en figure 4. Sa description globale est la suivante :

Phase descendante du modèle en U adapté : le début de la phase descendante commence par deux étapes essentielles qui ont lieu simultanément et qui marquent le commencement du projet : (1) l'analyse du domaine d'application permet le plus souvent une première description fonctionnelle et structurelle de celui-ci ; (2) le plus précocement possible dans le projet, l'élaboration de premiers prototypes d'interfaces du SIAD visé permettent d'impliquer le plus rapidement possible les futurs utilisateurs en leur donnant un aperçu de solutions possibles (ou alternatives de solutions). Ces deux étapes sont prévues pour fournir un cadre structurant, en ce qui concerne les futures activités aussi bien que les solutions techniques. A l'issue de ces deux étapes, on obtient une définition d'un modèle de processus de travail (rejoignant ce que nous avons précisé en figure 2). Ce qui permet de cerner la liste des tâches à effectuer pour le fonctionnement du futur SIAD à réaliser, que ces tâches soient manuelles, automatiques ou interactives (cf. plus loin). Elles peuvent correspondre à des besoins fonctionnels et non fonctionnels : par exemple la nécessité de facilité d'utilisation du SIAD visé correspond à un besoin non-fonctionnel incontournable ; de même la confidentialité des données relatives aux patients est très importante dans le domaine de la santé. Le travail précédent peut être cyclique, comme le suggèrent les flèches sur le modèle en U adapté.

Les tâches prescrites suite à l'exécution des étapes précédentes, doivent être modélisées [Abed et al. 1991].

Dans cet article, nous adoptons UML (Unified Modeling Language) [Booch et al. 2000] comme langage de modélisation/conception ; ainsi pour définir les tâches du système à l'aide d'UML, nous utilisons d'abord les diagrammes de cas d'utilisation avec leur description détaillée. Après une définition des cas d'utilisation, il faut procéder à l'analyse et la conception centrée tâche du SIAD aussi bien pour la partie IHM que pour les outils d'aide basés sur le Data Mining. Chaque tâche dans le système interactif a un degré d'interactivité. Trois catégories principales de tâches peuvent être identifiées (ces catégories sont classiques en Génie Logiciel et IHM depuis les années 80) [Barthet 1988] [Breedvelt et al. 1997] [Paternò 1999] : (1) les tâches dans lesquelles seul l'utilisateur est impliqué, appelées tâches manuelles, (2) les tâches dans lesquelles seul l'aspect applicatif est représenté, appelées tâches système ou tâches automatiques, (3) les tâches impliquant des changements de degré de collaboration entre l'utilisateur et le système, appelées tâches interactives.

Les diagrammes de collaboration et de séquence d'UML permettent de spécifier et concevoir l'interface pour chaque tâche (tout en associant ces modèles à des représentations de plus en plus précises). Pour ce faire, il s'agit de se référer aux comportements probables des différents types d'utilisateur comme base : il prend la forme d'un modèle de l'utilisateur au sens large (ou « modèle utilisateur ») en termes de limites et ressources physiques et cognitives pour les différents types d'utilisateurs concernés [Yoshikawa 2003] [Robert 2003] ; certaines connaissances sur les utilisateurs sont générales et proviennent de la littérature, d'autres sont spécifiques et proviennent du terrain (interviews, observations...).

En se référant à la figure 2 et selon les spécificités du domaine, il s'agit d'analyser les différents outils d'aide à la décision en vue de définir ceux les plus adaptés au contexte du système à réaliser. La dernière étape de la phase descendante du modèle en U conduit à l'implémentation du SIAD complet ou de son prototype. Cette étape d'implémentation transforme les caractéristiques d'interfaces concrètes (définies dans les diagrammes de séquences associés à d'éventuelles maquettes plus ou moins évoluées) en représentations qui peuvent être implémentée directement par une boîte à outils graphique ou par un générateur d'interface homme-machine. Côté UML, des diagrammes de déploiement et de composants sont élaborés en vue de montrer l'évolution du système de la conception vers sa réalisation.

Phase ascendante du modèle en U adapté : l'évaluation d'un système homme-machine consiste à tester si l'utilisateur (de manière générale rappelons-le, car on peut avoir un ensemble de types d'utilisateur) peut accomplir sa tâche en utilisant l'interface fournie. Deux propriétés sont habituellement explorées pour de telles évaluations : l'utilité et l'utilisabilité, elles-mêmes décomposées en un ensemble de critères bien connus en IHM [Senach 1990] [Nielsen 1993]. Il existe un grand nombre de méthodes utilisables pour la mise en œuvre des évaluations [Senach 1990] [Wilson et Corlett 1996] [Bastien et Scapin 2001] [Ivory et Hearst 2001] [Ezzedine et Kolski 2004] [Huart et al. 2008] : observations, interviews, mouchard électronique, analyse des traces, oculométrie, questionnaires, etc. Dans cette phase d'évaluation, nous nous concentrons généralement sur l'exécution des tâches en lien avec les éléments visibles en figure 2 : (1) d'une part selon le comportement de l'utilisateur pendant l'interaction avec le système (difficultés rencontrées, temps requis pour accomplir une tâche, exactitude du

résultat, nombre et type d'erreurs, opinion, etc.), (2) et d'autre part, selon le système en termes de différences entre les objectifs et résultats obtenus.

Cette phase ascendante commence par la définition rigoureuse des protocoles expérimentaux (sujets, déroulement, situations et tâches concernées, IHM et aides impliquées, données à collecter...) [Abed 2001]. Une fois collectées, les données sont traitées en fonction des principes opérationnels dégagés de l'étape précédente. Il s'agit de mettre en corrélation les données avec l'activité humaine qui a été observée, et ceci pour l'ensemble du processus décisionnel mis en évidence en figure 2. Des séquences opératoires sont ainsi mises en évidence. Ce travail permet de reconstituer (modéliser) progressivement les activités réelles (appelées aussi tâches réelles). L'expérience montre que ces tâches réelles peuvent être sensiblement éloignées des tâches prescrites, prévues initialement dans la phase descendante.

Un principe fondamental du modèle en U est justement la confrontation entre les tâches réelles et les tâches prescrites [Abed 1990] [Millot 1990]. Le résultat de la confrontation conduit, soit à valider le système complet (IHM et aides), soit à révéler ses imperfections pour l'améliorer progressivement. Des retours vers différentes étapes de la démarche descendante sont donc nécessaires, selon l'ampleur des améliorations à apporter au système.

Dans la figure 4, le sens des flèches indique le sens d'envoi d'informations (modèles, prototypes, tâches, etc.) entre les différentes étapes du modèle en U adapté.

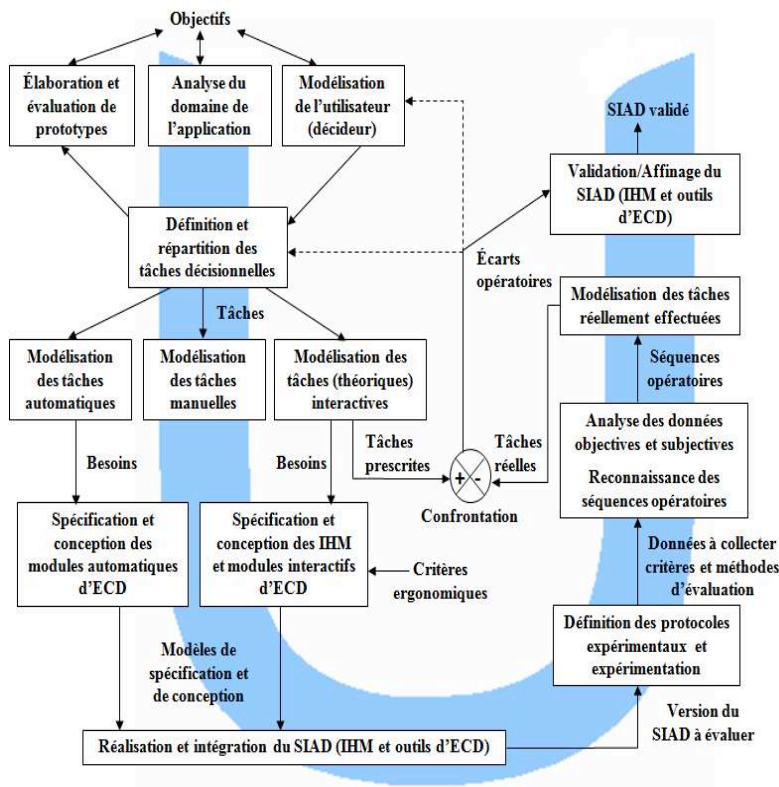


Figure 4 : modèle en U adapté [Ltifi et al. 2008] [Ltifi et al. 2009] [Ben Ayed et al. sous presse]

L'approche de conception du processus d'ECD doit être adaptée dès le début d'un projet afin de satisfaire les besoins des décideurs. En s'inspirant du travail de [Lajnef et al. 2005], nous proposons des similitudes méthodologiques entre le domaine de l'interaction homme-machine et celui de l'Extraction de connaissances (figure 5) :

- (a) Les objectifs principaux du processus ECD sont ceux du futur système décisionnel interactif à concevoir. En effet, il s'agit d'identifier l'ensemble des cas auxquels l'outil de fouille de données vise à répondre : il s'agit de la spécification des exigences du système interactif à concevoir. Une fois les objectifs en termes d'ECD identifiés, il s'agit de les classer puis de décomposer chaque problème en sous-problèmes – en fait, on considère que beaucoup de problèmes complexes peuvent être divisés en sous-problèmes plus simples à résoudre [Garlatti 1996].

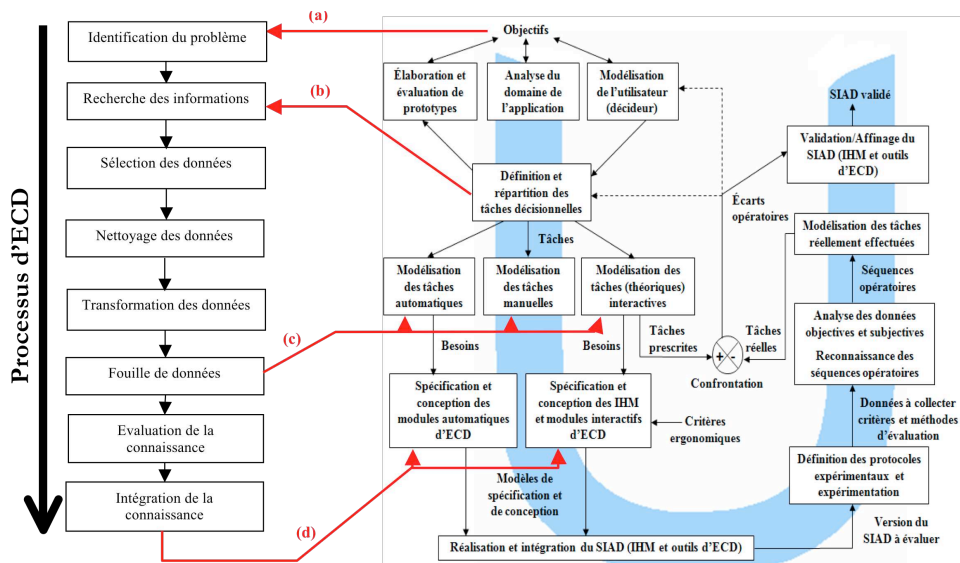


Figure 5 : Le modèle en U adapté au processus d'ECD [Ltifi et al. 2008]

- (b) La définition des tâches décisionnelles permet de donner une idée sur les données pertinentes qu'il faut rechercher, sélectionner (trier, regrouper ou répartir ces données selon les tâches définies), nettoyer et transformer (en réalisant divers calculs tels que : totaux, moyennes, écarts types, comparatif d'une période à l'autre, etc. ; en présentant les résultats d'une manière synthétique ou détaillée, le plus souvent graphique selon leurs besoins ou les attentes des différents types d'acteurs concernés de l'entreprise, etc.).
- (c) Une fois les données bien préparées, elles peuvent être fouillées pour extraire des connaissances. Une ou plusieurs techniques peuvent être utilisées permettant d'extraire les acteurs pertinents et de construire le modèle initial (un réseau de neurones, un arbre de décision, etc.). L'étude de cette tâche est très importante pour le concepteur du système décisionnel. Une telle activité peut se réaliser au niveau de l'étape de modélisation de la tâche afin de délivrer un modèle de la tâche prescrite. Les résultats extraits peuvent aider à spécifier les modules du système à concevoir.

(d) La connaissance extraite et évaluée doit être intégrée dans les modules du SIAD à spécifier et concevoir. La présentation de cette connaissance doit répondre aux critères ergonomiques bien connus dans le domaine de l'IHM afin d'assurer la meilleure interaction possible entre le décideur et la machine [Lepreux et al. 2003].

Le déroulement des différentes étapes du modèle en U et du processus d'ECD se fait d'une manière parallèle. Les flèches de la figure 5 présentent les relations entre les différentes étapes du modèle en U adapté basé sur les étapes du processus d'ECD.

Dans la partie suivante, il est expliqué comment cette démarche est mise en application dans le domaine médical en vue du développement d'un SIAD pour la lutte contre les infections nosocomiales.

5. CONTEXTE APPLICATIF

Dans la partie précédente nous avons présenté notre proposition de démarche centrée utilisateur adaptée au processus d'ECD. Elle a été appliquée à un cas concret dans le domaine médical. Le SIAD en question, basé sur un processus d'ECD s'appuie pour l'étape de fouille de données sur le raisonnement à partir de cas [Riesbeck et Shank 1989] ; ce SIAD est en cours d'utilisation et d'évaluation dans le service de réanimation du Centre Hospitalo-universitaire Habib Bourguiba à Sfax, Tunisie. Il a pour but d'aider les médecins du service, utilisateurs du système, à comprendre, prédire et prévenir les infections nosocomiales.

Les infections nosocomiales (IN) représentent un des problèmes majeurs de la santé publique. Ce sont des infections contractées dans un établissement de soins. Une infection est considérée comme telle lorsqu'elle était absente au moment de l'admission du patient [Garner et al. 1988]. Lorsque l'état infectieux du patient à l'admission est inconnu, l'infection est classiquement considérée comme nosocomiale si elle apparaît après un délai de 48 heures d'hospitalisation. Une infection nosocomiale peut être localisée dans un organe. Les organes les plus touchés sont les reins et le cœur. Ces infections peuvent être directement liées aux soins ou simplement survenir lors de l'hospitalisation indépendamment de tout acte médical. Dans les services de réanimation le problème des infections nosocomiales est d'autant plus préoccupant que les patients qui y sont hospitalisés sont plus fragiles. La lutte contre les IN est un problème complexe. Plusieurs travaux ont proposé des systèmes informatiques de surveillance des IN basés sur les techniques de fouille de données [Brossette et al. 1998] [Brossette et al. 2000]. Les travaux montrent leur efficacité et leur capacité à produire des règles utiles. Mais, telles que ces règles sont décrites dans les articles, leur utilisation par des médecins paraît difficile.

Une enquête a été réalisée dans l'hôpital CHU Habib Bourguiba à Sfax, Tunisie pour mesurer un jour donné (entre le 17 avril 2002 (minuit) et le 18 avril 2002 (minuit)) la prévalence de la survenue d'infections nosocomiales [Kallel et al. 2005]. Cette étude a montré que 17,9 % des 280 patients hospitalisés le jour de la prévalence dans l'ensemble de l'hôpital ont été victimes d'une IN. Les patients hospitalisés dans le service de réanimation sont les plus touchés (77,2 %). Notre projet est en lien avec les médecins du service de réanimation de cet hôpital. Ces médecins sont les experts et futurs utilisateurs du SIAD sur lequel nous travaillons.

5.1. Modules visés couvrant le processus d'ECD

Pour développer le SIAD proposé basé sur l'ECD, nous avons suivi les différentes phases du processus d'ECD (figure 1) ; le système a été subdivisé en modules en se

référant aux étapes de processus d'ECD. Deux modules (applications) ont été conçus et réalisés : un pour le stockage et la préparation des données ; et un pour la fouille de données proprement dite. L'algorithme de fouille que nous avons choisi est une version améliorée de l'algorithme dit des k plus proches voisins (kPPV) utilisé dans la technique du raisonnement à partir de cas (RBC) [Ben Ayed et al. 2006].

Dans ce qui suit, nous allons décrire le déroulement de la démarche relativement à chacun de ces deux modules du SIAD en nous inspirant de [Ben Ayed et al. sous presse].

5.2. Déroulement de la démarche relativement au premier module

Les tâches effectuées dans le cadre du premier module sont décrites dans le tableau 1. Pour celui-ci, nous avons surtout étudié le contexte et défini les objectifs en organisant des réunions fréquentes avec le médecin qui s'occupe du problème des IN, et ayant le double rôle de décideur et d'utilisateur. Parfois il invitait ses collaborateurs à assister aux réunions. Cette collaboration s'est avérée très précieuse, et a permis la définition et la répartition des tâches.

L'implémentation a concerné la mise en place de la base de données ainsi que son prétraitement (nettoyage et transformation des données) et des interfaces homme-machine de saisie des données. Les premiers tests des versions préliminaires des interfaces de saisie ont été très utiles pour le reste du processus de développement. L'intérêt qu'a montré l'utilisateur relativement à notre méthode lui a permis de bien exprimer les évolutions de ses objectifs.

L'interface Homme-Machine de la Figure 6, donne l'accès à l'utilisateur pour saisir les informations relatives à l'hospitalisation du patient (durée de séjour, diagnostics, prise antibiotiques, etc.). Cette interface a été validée par l'utilisateur après plusieurs itérations.

5.3. Déroulement de la démarche relativement au deuxième module

Une fois la réalisation du module de "Stockage et préparation des données" terminée, nous avons entamé celle du module de "Fouille de données" (tableau 2). L'objectif initial de cette application est de fournir au médecin une prédiction sur la probabilité d'apparition d'une IN chez un patient entrant dans le service. Pour la fouille de données, en nous conformant aux besoins exprimés par les experts (utilisateurs), nous avons développé une application pour la prédiction de la survenue d'IN, se basant sur la technique du raisonnement à partir de cas [Kolodner 1996] [Ben Ayed et al. 2006].

	Etapes du Modèle en U adapté	Tâches effectuées
Etape (a) (Fig. 5)	Analyse du domaine d'application	Etude des infections nosocomiales (définition, causes, risques...)
	Premiers prototypes d'interfaces. Description des besoins fonctionnels et structurels du système et définition des tâches	Description structurelle et fonctionnelle du domaine d'application médical déduite après consultation des futurs utilisateurs et élaboration de prototypes d'interfaces Homme-Machine en déterminant leur architecture générale. Les médecins proposent des fenêtres en onglets.
	Modélisation de l'utilisateur	Une modélisation de l'utilisateur (type unique pour le moment) est réalisée : il est expert dans son domaine ; il peut être considéré comme expert aussi concernant l'utilisation des outils informatiques (plus de 3 heures par jour devant son PC).
Etape (b) (Fig. 5)	Définition et répartition des tâches décisionnelles	Une première description des tâches manuelles (remplissage de fiches remplies par les médecins internes, contenant les données nécessaires à l'étude, au chevet des malades), automatiques (calcul de l'âge et d'un paramètre de risque de décès) et interactives (choix des libellés des zones de saisie, ...) est faite à l'aide des cas d'utilisation d'UML.
Etape (c) (Fig. 5)		<i>Cette étape n'entre pas dans le cadre du premier module du SIAD en question</i>
Etape (d) (Fig. 5)		<i>Cette étape n'entre pas dans le cadre du premier module du SIAD en question</i>
Réalisation	Réalisation et intégration du système Homme-Machine	<ol style="list-style-type: none"> 1. Implémentation de la base de données en utilisant le SGBD Oracle® et ses outils (SQLplus,...). 2. Nettoyage de la base de données (valeurs manquantes, répétitions, erreurs de données, etc.) 3. Transformation de la table contenant les informations d'hospitalisation des patients pour les préparer à la fouille (transformation de la date de naissance des patients en âge, etc.) 4. Implémentation des interfaces Homme-Machine de saisie pour entrer les données nécessaires pour appliquer la fouille de données (figure 6). 5. Correction des derniers bugs liés aux derniers tests utilisateurs
Validation	Branche ascendante du modèle en U adapté	<ol style="list-style-type: none"> 1. Mise en œuvre de tests utilisateurs : mise en évidence de problèmes d'utilisabilité au niveau des saisies, mise en évidence d'autres besoins pour les médecins, ajout de boutons, etc. 2. Mise en œuvre de tests utilisateurs selon une démarche de vérification des caractéristiques d'une IHM de qualité (cohérence, prévention des erreurs, ...) et validation finale.

Tableau 1 : Déroulement du développement du module de stockage et préparation de données

Hospitalisation Patient

Hospitalisation des patients

Identifiant: 006833 Nom: nom Prénom: prénom
 Date de naissance: 01/07/1943 Sexe: Femme

Dossier médical du patient

Num Dossier	Date entrée REA	Date sortie REA	Examens effectués	Procédures subies	prise ATB
518/06	01/05/2006 00...	06/05/2006 0...	14	25	6

Antécédents du patient
 Immunosuppresseur

Actualiser

Numéro de dossier: 518/06 Kausus: B
 Date entrée Hospital: mercredi 10 mai 2006
 Date entrée REA: lundi 1 mai 2006
 Date sortie REA: samedi 6 mai 2006
 Durée de séjour: 6
 Age: 63
 Origine: Hospital
 Détail origine: Clinique
 Cause: Médicale

SAPS II: 27
 MP-SAPS II: 15,7280940563622
 Issue: Survivant
 Diagnostic 1:
 Diagnostic 2:
 Prise ATB: Non
 Info nosocomiale: Non

Ajouter
 Modifier
 Supprimer
 Rechercher
 Annuler
 Infection ?

Dossier médical
 Examen infectieux
 Prescription ATB
 Procédure
 Rapport dossier
 Quitter

Figure 6 : Extrait de l'interface Homme-Machine de saisie de données concernant le patient

Le raisonnement à partir de cas est une approche de résolution de problèmes basée sur la réutilisation par analogie d'expériences passées appelées *cas*, au cours d'un cycle de raisonnement [Kolodner 1996]. Un cas est composé de deux parties décrivant un problème et la solution qui lui a été appliquée. Un cas est une expérience passée permettant au système de résoudre des problèmes plus efficacement ou d'éviter les échecs passés : un cas doit donc être utile pour le raisonnement. Dans ce but, un ensemble de caractéristiques est extrait ou sélectionné dans la représentation de chaque cas. Ces caractéristiques déterminent dans quelle situation un cas est applicable et utile.

Ces caractéristiques doivent être prédictives, c'est-à-dire qu'elles doivent être responsables de l'élaboration de la solution mise en œuvre dans le cas et permettre ainsi de déterminer dans quelle mesure une solution est applicable pour un autre problème. Les caractéristiques doivent également être suffisamment abstraites pour permettre la réutilisation des enseignements du cas pour différents problèmes, tout en restant suffisamment concrètes pour être facilement identifiables dans un nouveau problème. Lorsqu'un nouveau problème est à résoudre, un cas cible est construit avec la partie solution inconnue, celle-ci devant être apportée par le raisonnement. Les cas sources représentent des expériences passées qui sont stockées dans une mémoire. Les cas sources et le cas cible partagent le même formalisme de représentation. L'objectif du raisonnement est de transférer des enseignements pertinents des cas sources pour élaborer la solution du cas cible.

	Etapes du Modèle en U adapté	Tâches effectuées
Etape (a)		<i>Cette étape n'entre pas dans le cadre du deuxième module du SIAD en question</i>
Etape (b)		<i>Cette étape n'entre pas dans le cadre du deuxième module du SIAD en question</i>
Etape (c)	Modélisation des tâches automatiques	Utilisation des diagrammes de collaboration d'UML pour modéliser les tâches automatiques relatives à l'application de la technique KPPV de la fouille de données Modélisation de l'intégration de la connaissance extraite dans la prise de décision
	Modélisation des tâches manuelles	<i>Sans objet dans notre cas</i>
	Modélisation des tâches (théoriques) interactives	Modélisation des interactions de l'utilisateur avec la machine (donner la valeur de « K » pour l'application de l'algorithme de fouille, nombre de combinaisons possibles, etc.)
Etape (d)	Spécification et conception des modules auto d'ECD	Utilisation des diagrammes UML pour la spécification: - des modules de calcul (âges, durées de séjour, risques de décès, ...) - des transactions avec le SGBD.
	Spécification et conception des IHM et modules interactifs d'ECD	Spécification des modules interactifs en utilisant les diagrammes de séquence d'UML
Réalisation	Réalisation et intégration du système Homme-Machine	<ol style="list-style-type: none"> 1. Implémentation des interfaces de fouille de données (développement de code source relatif à l'algorithme de KPPV). La fouille se fait en deux étapes suivant deux IHM (figures 7 et 8) 2. Correction des défauts résiduels dégagés par l'utilisateur du système.
Validation	Branche ascendante du modèle en U adapté	<p>En vue de la validation et l'affinage du module de fouille de données et l'évaluation de la connaissance extraite, des tests d'évaluations ont été planifiés et réalisés selon trois critères [Nielsen 1993] :</p> <ol style="list-style-type: none"> 1. Utilité (tests fonctionnels) : il s'agit de vérifier si l'interface homme-machine est utile c'est-à-dire qu'elle fournit les fonctions nécessaires à l'utilisateur. Ce type de test nous permet de vérifier si l'algorithme de KPPV développé répond bien aux besoins de l'utilisateur (validité du résultat, temps d'exécution, etc.). 2. Utilisabilité (tests ergonomiques) : elle concerne la facilité de manipulation et rend compte de la qualité de l'interface en termes de facilité d'apprentissage et d'utilisation. Les médecins ont la possibilité de guider l'application de l'algorithme de KPPV en choisissant la valeur k (1, 3 ou 5) avec un simple clic. En plus ils ont été très intéressés par l'exactitude des résultats présentés (Fig. 7 et 8). 3. Acceptabilité : il s'agit de vérifier si l'interface est compatible avec les valeurs et la culture du service de réanimation de l'hôpital ; on parle dans ce contexte des sigles de la médecine et du service concerné ainsi que des couleurs qui doivent varier entre le bleu et le blanc, etc. <p>En fait, des premiers prototypes de l'application présentent des interfaces homme-machine avec un fond beige qui a été refusé par les médecins. Une réunion a été organisée avec ces derniers pour consulter la palette de couleurs en vue de choisir la couleur qui leur convient le plus ; c'est la couleur courante de l'interface.</p> <p>Des derniers tests ont été réalisés pour l'évaluation du système complet.</p>

Tableau 2 : déroulement du développement du module de fouille de données

Les interfaces homme-machine visibles en figure 7 et figure 8 montrent l'application de l'algorithme de la fouille de données, qui est le classifieur supervisé KPPV (les k-plus-proches-voisins).

Caractéristiques des cas sources

Ce bouton permet d'appliquer l'algorithme de fouille et de calculer les K plus proches voisins

Ce bouton permet de passer à la deuxième interface pour la prédiction

Ce bouton permet d'appliquer l'algorithme de fouille à un autre patient

Figure 7 : Extrait de l'interface Homme-Machine de l'étape 1 de fouille de données (définitions des KPPV)

Le principe de l'algorithme KPPV consiste à identifier comme son nom l'indique les k plus proches voisins. Pour ce faire, il faut calculer les distances entre les caractéristiques du patient concerné par l'étude et celles des patients déjà hospitalisés et enregistrés dans la base de données du service de réanimation. La distance que nous avons utilisée pour calculer les plus proches voisins est la distance Euclidienne. Le K définit le nombre de plus proches voisins à considérer pour l'élaboration de la décision du classifieur ; dans le but d'avoir le meilleur taux de discrimination, nous avons choisi trois valeurs de K qui sont (1, 3, 5), pour tester les classifieurs 1-PPV, 3-PPV et 5-PPV.

Pour la deuxième interface, il s'agit de calculer la probabilité d'avoir une IN en fonction des k plus proches voisins. La présentation de cette probabilité est faite en fonction de sa valeur. Si elle est supérieure à 50% c'est qu'il existe un grand risque d'avoir une infection nosocomiale ; pour cette raison la valeur de la probabilité se présente en rouge gras et avec une image d'alerte.

Ces interfaces Homme-Machine ont fait l'objet de plusieurs itérations, démonstrations et évaluations avec les utilisateurs.

Caractéristiques du cas cible

Nouveau patient

Dossier médical	Durée de séjour	Age	Catégorie	Kneus	SABS II	Prise ATB	Durée VA	Durée KTC	InNos
518/06	6	63	Médicale	B	27	Oui	7	0	Oui (60%)

Caractéristiques des plus proches voisins

Dossier médical	Durée de séjour	Age	Catégorie	Kneus	SABS II	Prise ATB	Durée VA	Durée KTC	InNos	
Dossier médical 1	549/06	14	74	Médicale	B	43	Oui	13	0	Oui
Dossier médical 2	439/06	15	63	Médicale	B		Oui	15	0	Oui
Dossier médical 3	397/06	17	73	Médicale	B	38	Oui	8	0	Oui
Dossier médical 4	555/06	5	81	Médicale	B	47	Oui	4	0	Non
Dossier médical 5	507/06	4	80	Médicale	B	50	Oui	4	0	Non

Identifiants des voisins

Dossier médical 1	00/45801	A
Dossier médical 2	06/19038	B
Dossier médical 3	06/19563	C
Dossier médical 4	06/28485	D
Dossier médical 5	06/25738	E

Fonction de combinaison

	1	2	3	4	5
Nbre de voisins retenus	1	2	3	4	5
Identifiants des voisins	A	A B	A B C	A B C D	A B C D E
Réponses des voisins	0	0 0	0 0 0	0 0 0 N	0 0 0 N N
Décompte des réponses	Oui:1 Non:0	Oui:2 Non:0	Oui:3 Non:0	Oui:3 Non:1	Oui:3 Non:2
Valeur retenu	Oui	Oui	Oui	Oui	Oui
Évaluation	100%	100%	100%	75%	50%

Combinaison [X] Quitter

Ce bouton permet de calculer la probabilité d'acquisition d'une IN en fonction des caractéristiques des K plus proches voisins (dans ce présent cas: k= 5 patients dont les 3 premiers ont acquis une IN. Donc le calcul des combinaisons donne une probabilité de 60%).

Figure 8 : Extrait de l'interface Homme-Machine de l'étape 2 de fouille de données (calcul de la probabilité d'IN)

6. CONCLUSION

Cet article a proposé une approche centrée utilisateur de conception de système interactif d'aide à la décision basé sur un processus d'ECD. Depuis plusieurs dizaines d'années, une importante masse d'informations est stockée sous forme informatique dans les entreprises. Les systèmes d'information sont destinés à garder la trace d'événements de manière fiable et intègre. Ils automatisent de plus en plus les processus opérationnels, particulièrement, d'aide à la décision. Pour ce faire, l'extraction de connaissances à partir de données (ECD) est utilisée comme outil décisionnel permettant de découvrir, à partir de bases de données, des connaissances auparavant inconnues et potentiellement utiles pour la prise de décision. Un SIAD basé sur l'ECD est hautement interactif. En effet, comme la prise de décision implique l'acteur humain, qu'est le décideur, le processus d'ECD nécessite de sa part une interactivité Homme-Machine pour se réaliser. Une modélisation d'un tel système doit donc selon nous faire appel aux méthodes, modèles et critères de l'IHM. D'où la proposition d'une approche

de développement qui s'appuie sur le modèle en U. Une adaptation de celui-ci au processus d'ECD a fait l'objet de cet article.

L'approche proposée est actuellement mise en œuvre depuis plus de 2 ans. En effet, le SIAD en question est en cours d'évaluation et d'utilisation dans le service de réanimation du Centre Hospitalo-universitaire Habib Bourguiba à Sfax, Tunisie. Il a pour but d'aider les médecins du service, utilisateurs du système, à comprendre (en se référant à l'historique des patients enregistrés dans la base de données de l'hôpital), prédire (par l'application d'un algorithme de fouille de données) et par la suite (en fonction du résultat de prédiction) à prévenir les infections nosocomiales ; infections contractées par les patients pendant leur hospitalisation. Le SIAD développé doit faire l'objet dans les mois à venir d'un ensemble d'expérimentations.

Sous l'angle de l'IHM et en nous basant sur cette expérience dans le domaine médical, on se trouve confronté au besoin de développer une méthodologie spécifique pour l'évaluation de SIAD à base de l'ECD en s'inspirant des critères, méthodes et techniques d'évaluation issues conjointement du domaine de l'IHM et de celui de la fouille visuelle de données (visual data mining) ; c'est-à-dire avec intégration du traitement visuel d'information dans toutes les étapes du processus ECD. Cette fouille visuelle permet de présenter les données et les connaissances dans une forme visuelle permettant à l'utilisateur d'interpréter les données, tirer les conclusions ainsi qu'interagir directement avec ces données. On considère que les techniques de visualisation améliorent les techniques d'ECD courantes en augmentant l'implication de l'utilisateur et sa confiance à propos des constatations découvertes [Keim 2002]. Une telle méthodologie d'évaluation doit permettre d'étudier l'expérience cognitive et affective des utilisateurs du système d'aide à la décision pour la lutte contre les infections nosocomiales [Yoshikawa 2003].

REMERCIEMENTS

Les auteurs tiennent à remercier le Docteur H. Kallel pour l'intérêt qu'il a accordé au projet et le temps passé pour aider à concevoir, implémenter et évaluer le système d'aide à la décision. Ils remercient également les relecteurs anonymes de la revue JIPS pour leurs remarques constructives.

REFERENCES

- Abed, M. 1990. Contribution à la modélisation de la tâche par outils de spécification exploitant les mouvements oculaires: application à la conception et à l'évaluation des interfaces homme-machine. Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambrésis, France.
- Abed, M., Bernard, J.M. et Angué, J.C. 1991. Task analysis and modelization by using SADT and Petri Networks. Proceedings Tenth European Annual Conference on Human Decision Making and Manual Control. Liege.
- Abed, M. 2001. Méthodes et Modèles formels et semi-formels de conception et évaluation des systèmes homme-machine. Mémoire d'HDR, Université de Valenciennes et du Hainaut-Cambrésis, France.
- André, J. 1993. Le cycle de vie en Y. Arche SQL.
- Barthet, M.F. 1988. Logiciels interactifs et ergonomie, Modèle et méthodes de conception, Dunod informatique, Paris.

- Bastien, C. et Scapin, D. 2001. Evaluation des systèmes d'information et Critères Ergonomiques. In Kolski C. (dir.), Environnements évolués et évaluation de l'IHM, Interaction homme-machine pour les SI 2, Hermès, Paris, 53-80.
- Ben Ayed, M., Feki, I. et Alimi, A.M. 2006. Optimisation de la technique de RBC pour la classification dans un processus de data mining. Atelier « Fouille de données complexes », Lille, France.
- Ben Ayed, M., Ltifi, H., Kolski, C. et Alimi, A.M. sous presse. A User-centered Approach for the Design and Implementation of KDD-based DSS: A case Study in the Healthcare Domain. *Decision Support Systems*, accepté et à paraître.
- Booch, G., Rumbaugh, J. et Jacobson, I. 2000. Le guide de l'utilisateur UML. Eds Eyrolles, Paris.
- Boehm, B. 1988. A Spiral Model of Software Development and Enhancement. *IEEE Computer*, 21 (5), 61-72.
- Breedvelt, I., Paternò, F., et Sereriins, C. 1997. Reusable Structures in Task Models, *Proceedings Design, Specification, Verification of Interactive Systems*. Springer Verlag, 251-265.
- Brossette, S.E, Sprague, A.P., Hardin, J.M., Waites K.B., Jones W.T. et Moser S.A. 1998. Association Rules and Data Mining in Hospital Infection Control and Public Health Surveillance. *Journal of the American Medical Informatics Association*, 5, 4, 373-381.
- Brossette, S.E., Sprague, A.P., Jones, W.T., et Moser, S.A. 2000. A data mining system for infection control surveillance. *Methods Inf Med.*, 39, 4-5, 303-310.
- Cathelain, S. 2005. Contribution à la conception de systèmes coopératifs. Application au domaine du contrôle de trafic aérien. Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambrésis, France.
- Chevrin V., Coiturier O., Mephu Nguifo E., et Rouillard J. 2007. Recherche anthropocentrée de règles d'association pour l'aide à la décision. *Revue d'Interaction Homme-Machine*, 8, 2, 117-146.
- Curtis, B., et Hefley, B. 1994. A WIMP No More: The Maturing of User Interface Engineering. *Interactions*, 23-34.
- Collins, D. 1995. *Designing Object-oriented user interfaces*. Readwoods City, CA: Benjamin/Cummings Publishing Inc.
- Ezzedine, H., et Kolski, C. 2004. Démarche d'évaluation d'IHM dans les systèmes complexes, application à un poste de supervision du trafic ferroviaire. *Revue d'Interaction Homme-Machine (RIHM)*, 5, 91-122.
- Evans, R.S. 2007. Case studies in clinical decision support: LDS hospital experience. *Clinical Decision Support*, 6, 143-167.
- Fayyad, U.M., Djorgovski, S.G. et Weir, N. 1996. Automating the Analysis and Cataloging of Sky Surveys. *Advances in Knowledge Discovery and Data Mining*. MIT Press, 471-494.
- Fayyad, U.M., Grinstein, G.G., et Wierse, A. 2002. Information visualization in data mining and knowledge discovery. Morgan Kaufmann.
- Frawley, W.J., Piatetsky-Shapiro, G., et Matheus C.J. 1992. Knowledge discovery in databases: an overview. *AI Magazine*, 57-70.
- Garlatti, S. 1996. Multi-media and interactive decision support systems in complex situation. In 43th meeting of the European working group "Multicriteria Aid for Decisions", Brest.
- Garner, J.S., Jarvis, W.R, Emori, T.G., Hogan, T.C. et Hugues J.M. 1988. CDC definitions for nosocomial infections. *Am. J. of Inf. Contr.*, 16, 3, 128-140.

- Guigue, L., et Donadey, T. 1999. Infections et systèmes d'aide à la décision. *Médecine et maladies infectieuses*, 29 (12), 758-767.
- Hartson, H.R., et Hix, D. 1989. Towards empirically derived methodologies and tools for Human-Computer development. *International Journal of Man-Machine Studies*, 31, 477-494.
- Hix, D., et Hartson, H.R. 1993. *Developing user interfaces: Ensuring usability through product & process*. Wiley professional computing, John Wiley & Sons, inc., New York, USA.
- Huart, J., Kolski, C., Bastien, C. 2008. L'évaluation de documents multimédias, état de l'art. In Merviel S. (Ed.), *Objectiver l'humain ? Volume 1, Qualification, quantification*, Hermes, Paris, pp. 211-250.
- Ivory, M. et Hearst, M. 2001. The State of the Art in Automated Usability Evaluation of User Interfaces. *ACM Computing Surveys*, 33, 4, 173-197.
- Jambu, M. 1999. *Introduction au Data Mining, analyse intelligente des données*. Eds Eyrolles, Paris.
- Jacobson, I., Booch, G., et Rumbaugh J. 1999. *Le processus Unifié de Développement logiciel*. Eds Eyrolles, Paris.
- Kallel, H., Bouaziz, M., Ksibi, H., Chelly, H., Hmida, C.B., Chaari, A., Rekik, N. et Bouaziz M. 2005. Prevalence of hospital-acquired infection in a Tunisian Hospital. *Journal of Hospital Infection*, 59, 343-347.
- Kaplan B. 2001. Evaluating informatics applications-clinical decision support systems literature review. *International Journal of Medical Informatics*, 64, 15-37.
- Kawamoto, K., Houlihan C.A., Balas E.A. et Lobach, D.F. 2005. Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. *British Medical Journal*, 330, 765-768.
- Keim, D. 2002. Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphics*, 8 (1), 1-8.
- Klein, M. et Methlie, L.B. 1990. *Expert Systems: A Decision Support Approach with applications in management and finance*. Eds Addison-Wesley Publishing Company.
- Kolodner, J.L., et Leake, D.B. 1996. *A Tutorial Introduction to Case-Based Reasoning: Experiences. Lessons & Future Directions*, AAAI press, MIT press, 31-65.
- Kolski, C. 1997. *Interfaces Homme-Machine, application aux systèmes industriels complexes (2ème édition revue et étendue)*. Editions Hermès, Paris.
- Kolski, C. 1998. A call for answers around the proposition of an HCI-enriched model. *ACM SIGSOFT Software Engineering Notes*, 23 (3), 93-96.
- Kolski, C., Ezzedine, H. et Abed, M. 2001. Cycle de développement du logiciel : des cycles classiques aux cycles enrichis sous l'angle des interactions homme-machine. In C. Kolski (Ed.), *Analyse et conception de l'I.H.M., Interaction Homme-Machine pour les S.I. 1*, Eds Hermes, Paris, 23-49.
- Kolski, C., Loslever P. 1998. An HCI-enriched model for supporting human-machine systems design and evaluation. *Proceedings 7th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems*, Kyoto, Japan, September.
- Lajnef, M., Ben Ayed, M. et Kolski, C. 2005. Convergence possible des processus du data mining et de conception-évaluation d'IHM : adaptation du modèle en U. *Proceedings of IHM'05, International Conference Proceedings Series*, ACM Press, Toulouse, 243-246.

- Lefébure, R. et Ventari, G. 2001. Data Mining : Gestion de la relation client, Personnalisation des sites Web. Eds Eyrolles.
- Leplat, J. 1985. Erreur humaine, fiabilité humaine dans le travail. Paris, A. Colin.
- Lepreux, S., Abed, M. et Kolski, C. 2003. A human-centred methodology applied to decision support system design and evaluation in a railway network context, *Cognition Technology and Work*, 5, 248-271.
- Lepreux, S. 2005. Approche de Développement centré décideur et à l'aide de patrons de Systèmes Interactifs d'Aide à la Décision. Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambrésis.
- Lepreux, S. 2006. Supports pour la prise en compte des experts et utilisateurs dans le développement de Systèmes Interactifs d'Aide à la Décision. *Ergo-IA'06*, 139-146.
- Lepreux, S. 2007. Réflexions en vue d'une meilleure prise en compte des experts et utilisateurs dans le développement de SIAD : apports des patrons et de la composition d'IHM. *Revue d'Interaction Homme-Machine (RIHM)*, 8, 1, 61-90.
- Ltifi, H., Ben Ayed, M., Kolski, C., et Alimi, A.M. 2008. Prise en compte de l'utilisateur pour la conception d'un SIAD basé sur un processus d'ECD", D. Galaretta, P. Girard, J.C. Tucoulou, M. Wolff (Ed.), Actes de la Conférence ERGO'IA 2008, L'humain au coeur des systèmes et de leur développement (Bidart/Biarritz, 15-17 octobre), ESTIA, 85-92.
- Ltifi, H., Ben Ayed, M., Kolski, C., Alimi, A. 2009. HCI-enriched approach for DSS development: the UP/U approach. *IEEE Symposium on Computers and Communications (ISCC'09, July 5-8)*, IEEE Press, Sousse, Tunisia, 895-900.
- Lévine, P., et Pomerol, J. 1989. Systèmes interactifs d'aide à la décision et systèmes experts. Eds Hermès, Paris.
- Long, J.B., et Denley, I. 1990. Evaluation for practice. Tutorial, Ergonomics society annual conference.
- Matheny, M.E., et Ohno-Machado, L. 2007. Generation of knowledge for clinical decision support: Statistical and machine learning techniques. *Clinical Decision Support*, 10, 227-248.
- McDermid, J., et Ripkin, K. 1984. Life cycle support in the ADA, environment. Cambridge University Press, Cambridge, UK.
- Millot, P. 1990. Coopération homme-machine : exemple de la téléopération. Actes des Journées du GR Automatique. Strasbourg.
- Nielsen, J. 1993. Usability engineering. Academic Press, Boston.
- Pacaux-lemoine M.P. 1998. Coopération Hommes-Machines dans les procédés complexes : modèles techniques et cognitifs pour le contrôle de trafic aérien. Mémoire de Doctorat, Université de Valenciennes et du Hainaut-Cambrésis, Valenciennes.
- Paternò, F. 1999. Model Based Design and Evaluation of Interactive Applications. Springer Verlag, Berlin.
- Pelayo, S., Bernonville, S., Kolski, C., et Beuscart-Zéphir, M. 2009. Applying a human factors engineering approach to healthcare IT applications: example of a medication CPOE project. J.G. McDaniel (Ed.), *Advances in Information Technology and Communication in Health*, IOS Press, 334-339.
- Riesbeck, C. et Shank, R. 1989. Inside Case-Based Reasoning. Lawrence Erlbaum.
- Robert, J.M. 2003. Que faut-il savoir sur l'utilisateur pour réaliser des interfaces de qualité?, In G. Boy (Ed.), *Ingénierie cognitive, IHM et cognition*, Edition Hermes-Lavoisier. Paris, 249-283.

- Royce, W. 1970. Managing the development of large software systems: Concepts and techniques. WESCON technical papers.
- Sandoval, V. 1997. L'informatique décisionnelle. Eds Hermes, Paris.
- Seffah, A., Kolski, C., Idoughi D. 2009. Persona comme outil de design de services interactifs : principes et exemple en e-maintenance. Proceedings of IHM 2009, 21ème Conférence de l'Association Francophone sur l'Interaction Homme-Machine (Grenoble, France, 13-16 octobre 2009), International Conference Proceedings Series, ACM Press, Grenoble, 333-336.
- Senach, B. 1990. Evaluation ergonomique des interfaces homme-machine. Une revue de la littérature. Rapport n°1180, INRIA.
- Simon, H.A. 1977. The new science of management decision. Prentice Hall, New Jersey.
- Tricot, N. 2005. Conception et évaluation de systèmes coopératifs avancés d'aide à la conduite automobile. Mémoire de Doctorat, Université de Valenciennes et du Hainaut-Cambrésis, Valenciennes, France.
- Turban, E. 1993. Decision Support and Expert Systems. Eds New York, Macmillan.
- Wilson, J.R., et Corlett E.N. 1996. Evaluation of human works: a practical ergonomics methodology. Eds Taylor & Francis, London.
- Yoshikawa, H. 2003. Modeling Humans in Human-Computer Interaction. In Jacko, J.A., Sears, A. (eds.), The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Mahwah, New Jersey, Lawrence Erlbaum Associates, 118-149.



Hela Ltifi est actuellement doctorante en Informatique. Elle est membre de deux laboratoires : REGIM (REsearch Group on Intelligent Machines) de l'École Nationale d'Ingénieurs de Sfax Tunisie et LAMIH (Laboratoire d'Automatique, de Mécanique et d'Informatique industrielles et Humaines) de l'Université de Valenciennes et du Hainaut-Cambrésis ; et ceci dans le cadre d'une thèse en cotutelle. Ses activités de recherche s'articulent autour des systèmes interactifs d'aide à la décision basés sur l'extraction de connaissances à partir des données devant être développés selon des démarches centrées utilisateur ; les résultats fournis par ces systèmes devraient permettre d'aider les médecins dans la lutte contre les infections nosocomiales. Elle est co-auteur de plusieurs articles dans des conférences internationales.



Mounir Ben Ayed est docteur en Science. Il est membre de l'équipe REGIM à l'École Nationale d'Ingénieurs de Sfax ENIS, Université de Sfax, Tunisie. Il est actuellement un enseignant des systèmes de gestion des bases de données, des datawarehouses et du data mining. Ses activités de recherche s'articulent autour des systèmes interactifs d'aide à la décision basés sur l'extraction de connaissances à partir des données. Il a dirigé des travaux sur les entrepôts de données, les OLAP et le data mining comme outils pour l'aide à la décision. La plupart de ses travaux sont appliqués dans le domaine médical. Il est co-auteur de plusieurs articles dans des conférences nationales et internationales. Mounir Ben Ayed a été membre du comité d'organisation du premier congrès international de Signaux Circuits et Système (SCS 2004) ; ainsi que coprésident du comité d'organisation de ACIDCA-ICMI en 2005. Il a organisé en plus des séminaires, écoles et workshops.



Christophe Kolski a obtenu le doctorat en 1989 et l'Habilitation à Diriger des Recherches (HDR) en 1995. Il est spécialisé en Interaction Homme-Machine, plus particulièrement en méthodes et modèles pour les systèmes interactifs et en interaction homme-machine intelligente. Il est professeur en informatique à l'Université de Valenciennes et du Hainaut-Cambrésis (UVHC) où il enseigne le génie logiciel et l'interaction homme-machine. Il dirige le groupe de recherche "Raisonnement Automatique et Interaction Homme-Machine" au LAMIH. Christophe Kolski est auteur, co-auteur, éditeur ou co-éditeur de nombreux livres, chapitres de livres, numéros spéciaux de revues, articles dans des revues et conférences nationales et internationales. Il a été co-éditeur de la revue d'Interaction Homme-Machine. Il a été président du comité d'organisation de la conférence CADUI'2002 et co-président des comités scientifiques des conférences IHM'2003 et ERGO-IA'2006.



Adel M. Alimi est professeur en Informatique industrielle titulaire en Génie Electrique (Informatique industrielle) depuis Décembre 2006. Il est fondateur et premier responsable du Groupe de Recherche sur les Machines Intelligentes (REGIM). Adel M. Alimi est le directeur de l'Ecole Nationale d'Ingénieurs de sfax depuis Juin 2005, ENIS, Université de Sfax, Tunisie. Il a été le président de la conférence internationale ACIDCA-ICMI'2005 International Conference on Machine Intelligence en 2005, co-président de la conférence internationale SCS'2004 International Conference on Signals, Circuits & Systems, et président de la conférence internationale ACIDCA'2000, International Conference on Artificial & Computational Intelligence for Decision, Control and Automation. Il a organisé plusieurs sessions spéciales et écoles. Il a été aussi membre de comités de programme de plusieurs conférences internationales. Adel M. Alimi est auteur, co-auteur, éditeur ou co-éditeur de chapitres de livres, numéros spéciaux de revues, articles dans des revues et conférences internationales.

Premiers retours d'expérience sur l'utilisabilité et les usages de systèmes interactifs plastiques

Fatoumata Camara, Gaëlle Calvary, Rachel Demumieux, Vincent Ganneau

► **To cite this version:**

Fatoumata Camara, Gaëlle Calvary, Rachel Demumieux, Vincent Ganneau. Premiers retours d'expérience sur l'utilisabilité et les usages de systèmes interactifs plastiques. Journal d'Interaction Personne-Système, Association Francophone d'Interaction Homme-Machine (AFIHM), 2010, 1 (1), pp.1-37. hal-01058922

HAL Id: hal-01058922

<https://hal.inria.fr/hal-01058922>

Submitted on 28 Aug 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Premiers retours d'expérience sur l'utilisabilité et les usages de systèmes interactifs plastiques

FATOUMATA CAMARA

¹ Laboratoire d'Informatique de Grenoble

² Orange Labs

GAELE CALVARY¹

RACHEL DEMUMIEUX²

VINCENT GANNEAU^{1,2}

Résumé : La propriété de plasticité a été introduite en 1999 en France en réponse à une variabilité nouvelle du contexte d'usage en intelligence ambiante. Une Interface Homme-Machine (IHM) est dite *plastique* lorsqu'elle est douée d'adaptation à son contexte d'usage dans le respect de propriétés centrées utilisateur. Dix ans plus tard, les avancées sont certaines. On recense de nombreux résultats aussi bien théoriques que pratiques. Cependant, l'utilisabilité et l'acceptabilité des IHM plastiques soulèvent de nombreuses questions qui nécessitent la mise en œuvre d'évaluations auprès d'utilisateurs. Cet article présente l'état de l'art en plasticité et en évaluation (méthodes et difficultés). Il décrit trois démonstrateurs et présente les premiers retours d'évaluations menées en laboratoire et sur le terrain.

Mots clés : Plasticité, Adaptation, Contexte d'usage, Evaluation, Expérimentation terrain.

Abstract: The *Plasticity* property has been introduced in 1999 in France to cope with the new variability of the context of use in ambient intelligence. A User Interface (UI) is said to be *plastic* if it is able to adapt to its context of use while preserving human-centered properties. Ten years later, theoretical as well as practical advances have been made. However, several questions related to usability and acceptability of plastic UIs need to be answered through evaluation with users. This paper presents a state of the art in plasticity and evaluation (methods and concerns). It describes three demonstrators and presents first results from evaluations led in laboratory and in the wild.

Key words: Plasticity, Adaptation, Context of use, Assessment, Field test.

1. INTRODUCTION

Avec les avancées des réseaux et la miniaturisation des dispositifs, l'utilisateur n'est plus immobile, assis devant un poste fixe. Il accède à des ressources variées (PC, téléphone, borne publique, services Web) qui apparaissent et disparaissent opportunément. L'interaction homme-machine devient possible en tout lieu, à tout instant, sur tout support. Dès lors, les Interfaces Homme-Machine (IHM) doivent pouvoir se conformer à un contexte d'usage désormais varié, variable et potentiellement imprévisible.

Au fil des années, les IHM ont progressivement évolué pour devenir adaptées, adaptables et adaptatives [Edmonds 1981] :

- *Adaptées* : dans une IHM adaptée, l'adaptation est mise en œuvre par le concepteur pour une cible prédéfinie d'utilisateurs [Benadi 2004]. Ceci constitue un premier niveau d'adaptation, qualifié de statique [Bobillier-Chaumon and al. 2005] ;
- *Adaptables* : une IHM adaptable est une IHM modifiable par l'utilisateur. Il choisit les adaptations qui lui conviennent le mieux en fonction de ses préférences et de ses habitudes (éléments graphiques, raccourcis, etc.) [Benadi 2004]. L'adaptation est dynamique, c'est-à-dire qu'elle a lieu pendant l'exécution du programme, mais demeure entièrement sous le contrôle de l'utilisateur [Bobillier-Chaumon and al. 2005] ;
- *Adaptatives* : dans une IHM adaptative, l'adaptation est prise en charge par le système sans intervention explicite de l'utilisateur. L'adaptation s'appuie sur un processus d'acquisition et d'exploitation d'un modèle utilisateur impliquant des formes d'apprentissage, d'inférence ou d'aide à la décision [Jameson 2003].

En 1999, face à l'intelligence ambiante, la communauté IHM a introduit la propriété de *plasticité*, définie comme étant la capacité d'adaptation d'une IHM à son contexte d'usage dans le respect de son utilisabilité [Thevenin and Coutaz 1999]. Dans cette définition, c'est une variation du contexte d'usage qui est à l'origine de l'adaptation. L'utilisateur peut être impliqué dans le processus d'adaptation couvrant ainsi les deux niveaux IHM *adaptables* et *adaptatives*. Les IHM utilisées peuvent être *adaptées* : plus elles le seront, plus probablement l'utilisabilité sera préservée. Cette préservation de propriétés centrées utilisateur est la raison d'être et le défi de la plasticité.

Dix ans après la définition de cette propriété, les avancées sont notables en conception d'IHM plastiques. Cependant, l'évaluation reste sous étudiée : peu de résultats d'évaluation sont publiés et validés par la communauté scientifique [Gena 2005]. De nombreuses questions restent encore sans réponse, notamment, à quel moment adapter, quelles données sont pertinentes pour le modèle utilisateur, quel niveau de contrôle laisser à l'utilisateur lors du processus d'adaptation ou encore comment intégrer les besoins d'une très large cible d'utilisateurs ?

Cet article apporte des éléments de réponse. Il énonce des critères ergonomiques formulés à partir d'enseignements issus d'évaluations de trois démonstrateurs. Ces démonstrateurs illustrent les leviers de l'adaptation : le remodelage et la redistribution d'IHM. Les plates-formes utilisées sont le téléphone mobile et l'ordinateur. Notre objectif est d'évaluer les performances et le ressenti des utilisateurs vis-à-vis du remodelage et de la redistribution d'IHM afin de déterminer l'intérêt et l'acceptabilité des adaptations. En particulier, il s'agit de comprendre dans quelle mesure les adaptations doivent être placées sous le contrôle des utilisateurs. Les observations

permettent d'alimenter les réflexions sur les méthodes et outils pour l'ingénierie de systèmes interactifs plastiques. Pour cela, nous relatons deux évaluations utilisateurs en laboratoire et une évaluation utilisateurs sur le terrain.

Après un bref état de l'art sur la plasticité et les méthodes et difficultés d'évaluation, cet article présente les démonstrateurs développés et détaille les principaux résultats des évaluations. L'article est structuré en quatre sections. La section 2 est consacrée aux concepts clé de la plasticité : le contexte d'usage, les leviers, le contrôle utilisateur et la valeur pour l'utilisateur. La section 3 traite de l'évaluation. Un ensemble de méthodes y sont présentées, ainsi que les difficultés liées à l'évaluation de systèmes adaptatifs. Les sections 4 et 5 décrivent respectivement les démonstrateurs implémentés ainsi que les évaluations menées (protocoles et résultats). Des recommandations ergonomiques sont, en particulier, formulées à partir des résultats d'évaluation. Enfin, des perspectives sont énoncées.

2. PLASTICITE: CONCEPTS CLE

Cette section présente les concepts clé de la plasticité au fil des évolutions de la définition. Aujourd'hui, la plasticité dénote la capacité d'adaptation d'une IHM à son contexte d'usage dans le respect de la valeur attendue par l'utilisateur [Dâassi 07]. Nous identifions trois mots clé : adaptation, contexte d'usage et valeur. Ils sont ici examinés avec pour l'adaptation deux centres d'intérêt : les moyens de l'adaptation (remodelage, redistribution, multimodalité) ainsi que le degré de contrôle du processus d'adaptation par l'utilisateur.

2.1 CONTEXTE D'USAGE

En plasticité, le contexte d'usage fait aujourd'hui référence au triplet <utilisateur, plate-forme, environnement>. L'utilisateur dénote l'utilisateur du système interactif ; la plate-forme représente l'ensemble des ressources d'interaction disponibles et l'environnement rassemble les caractéristiques physiques et sociales du lieu hébergeant l'interaction. Historiquement, l'effort a été porté sur la plate-forme [Grolaux and al. 2005]. L'environnement a été peu considéré. Il est en général réduit à la localisation de l'utilisateur. Au départ [Thevenin and Coutaz 1999], l'utilisateur n'était pas considéré. Il l'a été plus tardivement à travers la notion de modèle utilisateur.

En 2006, [Tarpin-Bernard 2006] souligne l'importance d'étendre la définition du contexte à la prise en compte de l'activité de l'utilisateur. L'argument donné par l'auteur est le suivant (page 35) : "un même utilisateur utilisant une même plate-forme d'interaction dans un environnement identique peut avoir des besoins très différents en fonction de l'activité qu'il mène à l'instant t". En fait, le besoin de prendre en compte l'activité de l'utilisateur avait été identifié auparavant mais [Tarpin-Bernard 2006] trouve impropre sa dilution dans la dimension utilisateur.

En 2007, la variabilité de l'objectif utilisateur a été intégrée à la réflexion. Jusqu'ici, l'objectif était supposé fixe, connu à la conception. [Gabillon and al. 2007] imagine des objectifs utilisateur émergents. Le cas d'étude est celui de Yoann, malade qui souhaiterait voir un médecin. Il lance son assistant personnel *Compose* et saisit son objectif "je voudrais voir un médecin" (Figure 1a). *Compose* détecte que Yoann est loin de chez lui. Il lui demande alors s'il souhaite consulter un médecin sur place ou rentrer chez lui (Figure 1b). Yoann préfère voir un médecin sur place. Compte tenu de l'heure, *Compose* lui propose quatre solutions (Figure 1c) parmi lesquelles il choisit le médecin de garde (le docteur Mabuse). *Compose* génère une IHM qui répond à l'objectif de

Yoann (Figure 1d). Elle fournit les éléments d'information et d'interaction nécessaires à l'atteinte de l'objectif exprimé : l'adresse du docteur Mabuse, son numéro de téléphone pré-composé, des boutons permettant d'appeler le médecin puis raccrocher, l'itinéraire pour se rendre au cabinet médical ainsi que les coordonnées de la pharmacie de garde la plus proche.

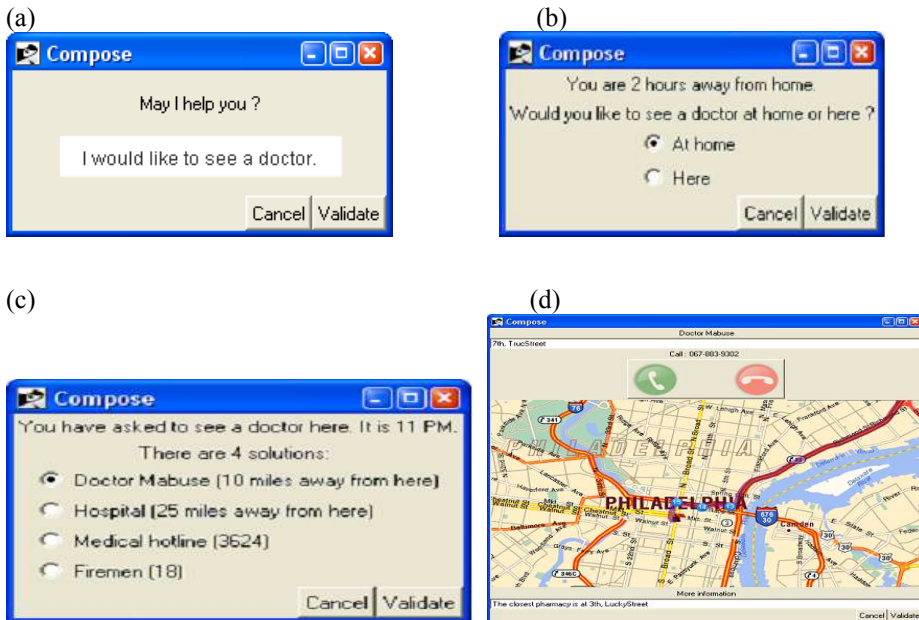


Figure 1: Compose, un système à objectif utilisateur émergent [Gabillon and al. 2007].

En 2008, une dimension supplémentaire est donnée au sujet par l'introduction d'objectifs utilisateur latents, c'est-à-dire non explicitement exprimés par l'utilisateur. EMMA (Embedded Manager for Mobile Adaptation) [Ganneau and al. 2008] s'inscrit dans cette lignée. EMMA est un système sensible au contexte sur téléphone mobile (plate-forme Windows Mobile CE, SPV C100). EMMA maintient un modèle utilisateur bayésien à partir de données contextuelles recueillies en mobilité : la localisation géographique de l'utilisateur, le jour de la semaine, l'heure, l'activité de l'utilisateur sur le téléphone (gestion du profil silencieux/sonore, application utilisée, etc.). L'hypothèse est que l'activité de l'utilisateur est fonction du contexte. En conséquence, lorsqu'un contexte est reconnu et nommé par l'utilisateur comme étant celui correspondant par exemple à la maison, au bureau ou encore aux vacances, EMMA propose à l'utilisateur les activités qu'il effectue en général dans ce contexte. En figure 2a, EMMA reconnaît un contexte clé. Dans ce contexte clé, qui a été nommé "Maison" par l'utilisateur, EMMA propose d'adapter le téléphone : réorganiser les fonctionnalités (Figure 2d) pour rendre saillantes les applications les plus utilisées dans ce contexte ; commuter le profil du téléphone tel que le mode vibreur, silencieux ou encore normal (Figure 2c) ; changer les paramètres de personnalisation comme la couleur du fond d'écran. L'utilisateur peut accepter ou refuser les propositions. Les changements de contexte et les adaptations à l'initiative du système peuvent être placés sous le contrôle de l'utilisateur via des boîtes

de dialogue (Figure 3). L'utilisateur peut choisir entre trois niveaux de contrôle : contrôlé, négocié et observable.

- *Contrôlé* : le système laisse la liberté à l'utilisateur de contrôler certains aspects de l'adaptation. Il s'agit notamment de la gestion des contextes clé (ajout, modification, suppression) ;
- *Négocié* : l'utilisateur est associé aux décisions. Le système propose et l'utilisateur dispose ;
- *Observable* : l'adaptation est perceptible à l'utilisateur mais il ne peut pas la contrôler. Seul un retour d'information lui est donné.

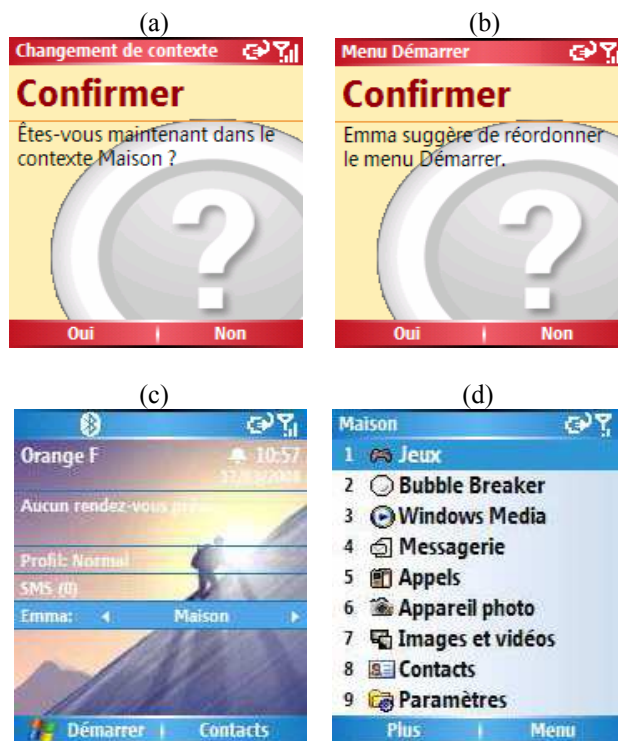


Figure 2: EMMA, un système sensible au contexte à objectif utilisateur latent [Ganneau and al. 2008].



Figure 3: EMMA, illustration des options disponibles pour le contrôle utilisateur [Ganneau and al. 2008].

2.2 LEVIERS DE LA PLASTICITE

Les deux leviers de la plasticité sont le remodelage et la redistribution [Calvary and Coutaz 2002].

Le *remodelage* est une adaptation locale à la plate-forme de l'utilisateur. Elle consiste en un remaniement de l'IHM sans en changer la distribution sur les ressources d'interaction (écran, clavier, souris, etc.). Ce remaniement peut porter sur la présentation (un menu déroulant peut se transformer en un jeu de boutons radio) ou le dialogue (des tâches utilisateur peuvent être ajoutées, modifiées ou supprimées). *FlexClock* [Grolaux and al. 2002] est un exemple typique de remodelage : selon la taille de la fenêtre, la présentation de l'heure change passant d'un texte à un cadran. Un calendrier est éventuellement ajouté selon la place disponible.

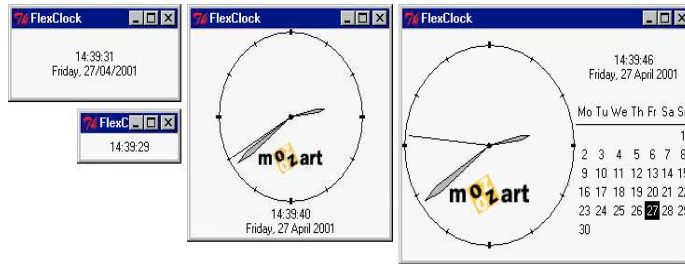


Figure 4: FlexClock, un exemple de remodelage [Grolaux and al. 2002].

Au contraire du remodelage, la *redistribution* change l'état de distribution de l'IHM sur l'ensemble des plates-formes en présence. La redistribution s'opère par migration totale ou partielle du système interactif. Sedan-Bouillon en est un exemple. C'est un site Web plastique pour la promotion des pays de Sedan et Bouillon. La version plastique est une version élaborée à partir d'une version simplifiée du site public (<http://www.bouillon-sedan.com/>). Elle est structurée en trois zones respectivement dédiées au titre, à la navigation et au contenu (Figure 5a). Dans sa version plastique, Sedan-Bouillon requiert l'identification de l'utilisateur (Lionel) (Figure 5b). Si Lionel se connecte à ce même site à partir d'un PDA, alors cette double connexion est détectée et une proposition de redistribution lui est faite : une IHM de contrôle utilisateur apparaît, expliquant à Lionel qu'il peut répartir le site, à son gré, entre le PC et le PDA (Figure 5c). La redistribution se fait au grain des espaces de dialogue. Lionel choisit, par un jeu de cases à cocher, sur quelle(s) plate(s)-forme(s) il souhaite placer ces trois espaces. La redistribution s'opère alors, composant l'IHM à façon.



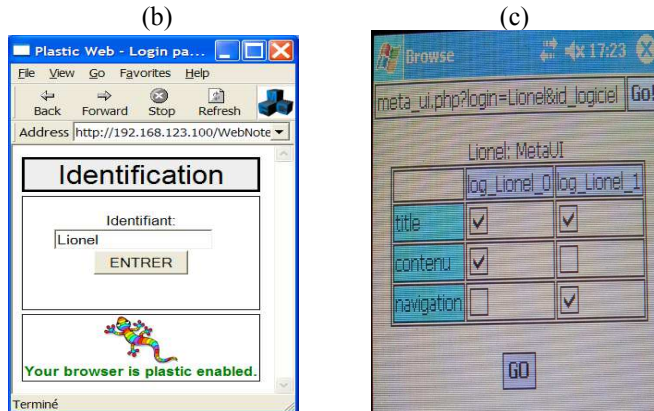


Figure 5: Sedan-Bouillon, un site web plastique (L. Balme et A. Demeure).

La redistribution peut requérir un remodelage. Le remodelage peut être *intramodal*, c'est-à-dire préserver la modalité courante. Il peut être *intermodal*, c'est-à-dire changer de modalité. Il peut être aussi *multimodal*, c'est-à-dire combiner plusieurs modalités. Même si l'adaptation n'est pas au cœur des recherches en multimodalité, elle suscite de l'intérêt dans ce domaine [Bellik 2006] [Huebsch and Kadner 2007] [Hesselman and al. 2008]. Dans tous ces travaux, la multimodalité est utilisée comme moyen d'adaptation. Pour exemple, le scénario suivant a été présenté dans [Hesselman and al. 2008] : Consuelo séjourne dans un hôtel et aimerait regarder un film avant de se rendre en ville pour dîner. Il est abonné au service de vidéo à la demande proposé par son opérateur mobile. Il utilise l'écran tactile disponible dans sa chambre d'hôtel pour sélectionner le film "Casino Royale". Comme une télévision haute définition est également disponible dans sa chambre et qu'il a à portée de main son portable, il reçoit le service en trois parties synchronisées : la partie vidéo du film sur la télévision haute définition, la partie audio sur le système hi-fi de la chambre et l'IHM de contrôle du service sur le mobile (Figure 6a). Lorsque l'heure de dîner approche, Consuelo quitte la chambre et se rend au parking pour prendre sa voiture. Il aimerait bien continuer à regarder son film. Aussi, il utilise les contrôles sur son téléphone mobile pour transférer la partie vidéo du film sur le téléphone (Figure 6b). Une fois installé dans la voiture, son téléphone mobile lui indique que son microphone, haut-parleurs et écran sont désactivés pour des raisons de sécurité. A partir de ce moment, seule la partie audio du film est disponible, et ce, grâce au système audio de la voiture. Comme il n'est pas très intéressant de regarder un film sans image, Consuelo décide de le suspendre à l'aide de commandes vocales. Pour ne pas perdre le fil à son retour, Consuelo annote le film avec quelques mots, vocalement, via le système audio de la voiture (Figure 6c). Enfin, avant de prendre la route, il utilise les commandes présentes sur le volant pour choisir une radio sur Internet qu'il écouterait tout le long du trajet.

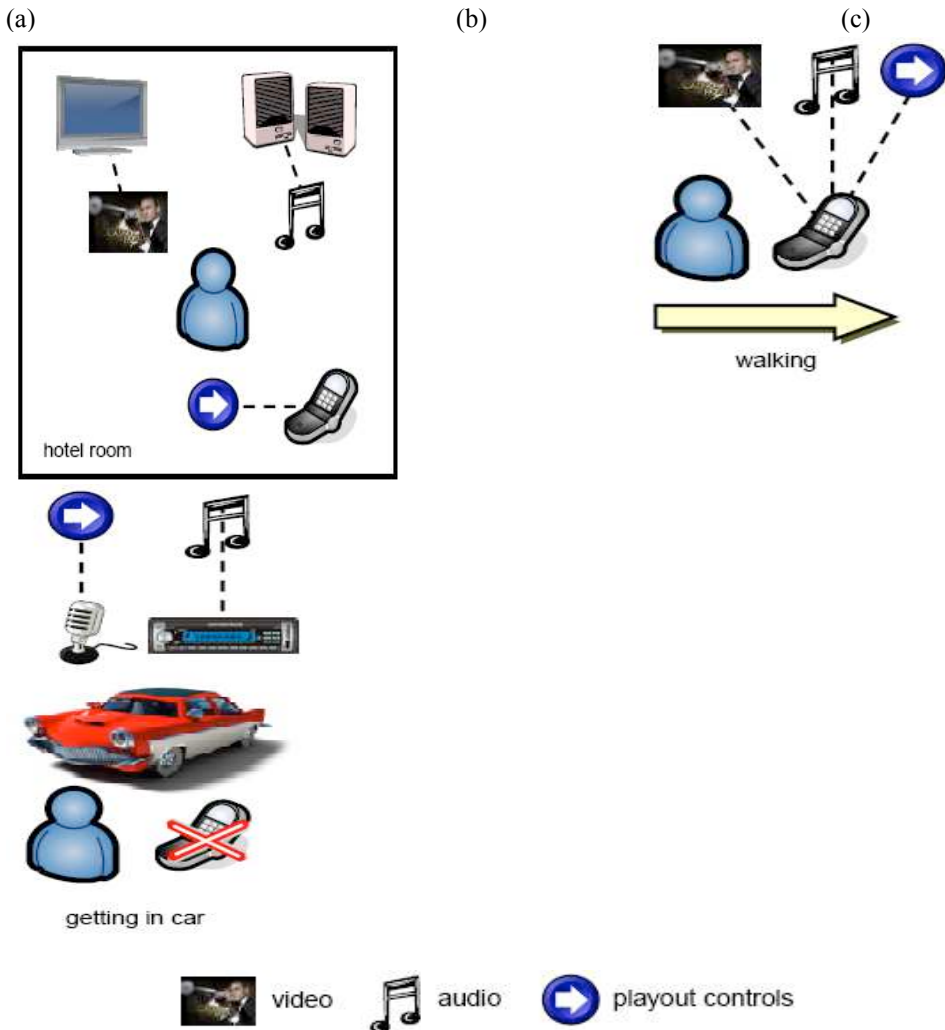


Figure 6: Scénario illustratif de l'adaptation par la multimodalité [Hesselman and al. 2008].

Explorer la multimodalité en plasticité est une perspective d'autant plus intéressante que les modalités se sont multipliées (vocale, tactile, gestuelle 2D, 3D, etc.), les technologies et infrastructures se sont développées et des résultats d'évaluation utilisateur sont disponibles. Par exemple, dans [Holly and al. 2002], les auteurs ont exploré différentes combinaisons de modalités. Le but est d'évaluer l'impact de telles combinaisons sur des sujets dotés d'une vue correcte. Les résultats montrent que :

- Les feedbacks uni-modaux audio, visuel et tactile sont équivalents en termes de charge mentale ;
- Les combinaisons de modalités sont intéressantes pour les retours d'information. Combiner l'audio, le visuel et le tactile réduit la charge mentale ;

- Un retour tactile est efficace seul ou en combinaison avec du visuel.

L'environnement, peu exploré en plasticité, est bien intégré en multimodalité. Dans [Lemmelä and al. 2008], les auteurs développent deux prototypes dans le but d'évaluer différentes modalités dans deux types d'environnement : la rue en situation de marche et la voiture en situation de conduite. Les résultats de ces études démontrent que le contexte d'usage a une influence sur les préférences des utilisateurs en terme de modalité d'interaction. En voiture, par exemple, les utilisateurs préfèrent l'interaction vocale et seraient prêts à utiliser un système entièrement vocalisé. Par contre, en situation de marche, ils ne trouvent pas cette modalité d'interaction utile. Ils estiment plus appropriée l'interaction gestuelle (2D, et surtout, 3D).

2.3 CONTROLE UTILISATEUR

Le contrôle utilisateur caractérise le degré de liberté accordé à l'utilisateur dans le processus d'adaptation. Le processus d'adaptation est un processus de type Perception-Action. A gros grain, il s'agit d'un processus en deux étapes qui consiste successivement à percevoir le contexte (Perception) puis adapter l'IHM en fonction du contexte observé (Action).

Dans la taxonomie de [Dieterich and al. 1994] pour la classification des systèmes interactifs dotés de capacités d'adaptation à l'exécution, les auteurs proposent un découpage du processus d'adaptation en quatre étapes :

- *L'initiative* : elle représente la décision d'un acteur (système ou utilisateur) de suggérer une adaptation ;
- *La proposition* : elle correspond aux différentes possibilités d'adaptation ;
- *La décision* : elle reflète le choix d'une adaptation parmi celles proposées ;
- *L'exécution* : elle consiste en la mise en œuvre effective de l'adaptation choisie.

Ces étapes, augmentées par l'étape de perception du contexte, couvrent l'ensemble du processus d'adaptation (cf. section 2.2). La question du contrôle utilisateur se pose à chaque étape. L'utilisateur peut intervenir aussi bien au niveau de la perception pour fixer les aspects du contexte à observer qu'au niveau de l'action pour orienter les décisions. De façon générale, on distingue quatre niveaux de contrôle :

- *Non observable* : l'adaptation s'opère à l'insu de l'utilisateur ;
- *Observable* : l'adaptation est perceptible par l'utilisateur sans qu'il dispose par contre de moyen de contrôle ;
- *Négocié* : l'utilisateur est associé aux décisions ;
- *Contrôlé* : le système laisse la liberté à l'utilisateur de contrôler le processus d'adaptation.

2.4 VALEUR

Une évolution majeure de la définition de la plasticité a consisté, en 2007, en l'introduction de la notion de *valeur*, en remplacement de l'utilisabilité. Avec la valeur, la plasticité devient la capacité d'une IHM à s'adapter à son contexte d'usage dans le respect de la *valeur* attendue par l'utilisateur [Dâassi 2007].

La notion de valeur a été introduite en IHM en 2004 par Gilbert Cockton. L'auteur estime que l'utilisabilité n'est pas suffisante [Cockton 2004a] [Cockton 2005]. L'exemple donné est celui de la gestion du chauffage [Cockton 2004a]. Les systèmes sont peut-être utilisables au sens où ils sont faciles à utiliser et à apprendre mais aucun ne répond à la véritable motivation de l'utilisateur, à savoir économiser. Ce n'est pas par plaisir que l'utilisateur programme son système de chauffage. C'est pour réduire sa facture. Un système de qualité devrait donc lui rendre observable le montant de l'économie faite.

En 2004, la *valeur* n'était pas définie de manière formelle. Gilbert Cockton disait qu'elle se cernait en discutant avec les gens [Cockton 2004b]. Le premier cadre de développement pour la *conception centrée valeur* (*Value Centered Design* (VCD)) [Cockton 2005] est apparu en 2005. La limite de ces travaux provenait principalement de la confusion autour du mot *value*. Son usage a, par exemple, créé la confusion avec la *conception sensible à la valeur* (*Value Sensitive Design* (VSD)). En effet, en anglais, le mot "*value*" peut avoir plusieurs sens. Il peut être dénombrable ou pas. Dans VSD, "*value*" est non dénombrable. Il est souvent employé au pluriel comme dans la phrase "*human values*". "*Value*" prend un autre sens lorsqu'il est employé au singulier. C'est le sens qui était recherché dans VCD. Avec cette acception, les valeurs font partie de la valeur.

Pour lever toute ambiguïté, Gilbert Cockton propose d'adopter, en 2006, le mot "*worth*" (valeur) à la place de "*value*" [Cockton 2006]. Ainsi, au lieu de parler de "*Value Centered Design*", on parlerait de "*Worth Centered Design*" (WCD). Ce changement de terminologie est allé au-delà d'une substitution de lettre. Par la même occasion, l'auteur a donné une définition de "*worth*". Le manque de définition du mot "*value*" constituait une faiblesse des précédents travaux et contribua certainement à empirer la confusion autour du mot.

La définition de "*worth*" adoptée est issue du dictionnaire "Concise Oxford English Dictionary" [Thompson 1996]. Selon Gilbert Cockton, pour faire court, la valeur ("*worth*") est un motivateur. Pour les concepteurs, il stipule ceci : concevoir pour la valeur consiste à concevoir des choses qui motivent soit à l'achat, l'apprentissage, l'utilisation ou la recommandation d'un produit interactif et idéalement de la plupart d'entre eux ("*designing worth means designing things that will motivate people to buy, learn, use or recommend an interactive product, and ideally most of all these*").

La préservation de la valeur étant le différentiateur entre adaptation et plasticité, il est important d'identifier et d'évaluer les facteurs de valeur.

3. EVALUATION DE SYSTEMES ADAPTATIFS

Cette section traite de l'évaluation des systèmes adaptatifs. Elle s'articule en trois parties. La première relate les difficultés liées à l'évaluation des systèmes adaptatifs. La deuxième présente les conditions possibles d'évaluation (avec ou sans utilisateur, en laboratoire ou sur le terrain). Enfin, la troisième partie expose deux méthodes d'évaluation spécifiques aux systèmes adaptatifs.

3.1 DIFFICULTES LIEES A L'EVALUATION DE SYSTEMES ADAPTATIFS

L'évaluation des systèmes adaptatifs est loin d'être une question triviale [Carbonell and Simonin 2004]. Si l'évaluation demeure encore aujourd'hui un des verrous en adaptation, c'est parce qu'elle est difficile par nature. A ce jour, peu de résultats d'évaluation sont validés par la communauté scientifique. A cela, plusieurs raisons :

- Vaste espace problème de l'adaptation

L'espace problème de l'adaptation est vaste. Les nombreux espaces problème [Thevenin 2001] [Vanderdonckt and al. 2005] [Tarpin-Bernard 2006] [Calvary 2007] en témoignent : ils ajoutent de nouvelles dimensions au fil du temps. En conséquence, l'évaluation peut porter sur différents aspects (contrôle utilisateur, pertinence des adaptations proposées, prévisibilité, transparence de la stratégie d'adaptation, moments de l'adaptation, etc.).

- Manque de critères et de métriques

Si des recommandations ergonomiques ont été formulées, il n'existe pas de référentiel de critères et de métriques pour l'évaluation des systèmes adaptatifs. L'absence de référentiel est en partie due à la spécialisation des systèmes par métiers ou domaines d'application. Très souvent, dans le monde académique, les recherches sont focalisées, nuisant en conséquence à une appréhension globale. Ces considérations cloisonnées complexifient le travail de standardisation. Si les chances de parution d'un tel référentiel dans un avenir proche sont faibles, on peut envisager un référentiel global couvrant les éléments communs de l'évaluation, appuyé de référentiels spécifiques.

- Collecte de données objectives en temps réel

Pour obtenir des résultats objectifs, les évaluations sont à conduire en situations réelles. Cependant, les opinions et ressentis des utilisateurs sont recueillis généralement lors de séances d'entretiens (individuels ou en groupe), postérieures à la phase d'utilisation des systèmes. Ces opinions et ressentis sont, d'une part, subjectifs et, d'autre part, ne constituent pas des données fraîches. Avec le temps qui s'écoule entre la fin de l'utilisation et les entretiens, d'importants « détails » peuvent échapper aux utilisateurs. Pour pallier à ces problèmes, il est nécessaire de mettre en place des outils "*standalone*" ou à embarquer dans le système interactif pour la collecte de données objectives. Le recueil d'avis utilisateurs en temps réel ou rapproché peut se faire par des journaux de bord accessibles via le Web. Cette solution peut s'avérer lourde pour les utilisateurs [Höök 1997].

- Coût de l'évaluation

L'évaluation de systèmes adaptatifs est coûteuse non seulement en ressources mais aussi en temps.

Les adaptations sont motivées par des changements de contexte d'usage. La multiplicité des contextes d'usage et des changements de contexte d'usage peut nécessiter de nombreuses passations de mêmes utilisateurs. La combinatoire augmente encore en faisant varier les participants.

Le temps requis pour mesurer l'impact de l'adaptation peut être long. Il est indispensable de laisser plusieurs jours, semaines, voire mois à l'utilisateur pour découvrir les différentes facettes du système adaptatif en situations réelles. En outre, bien souvent, un temps d'apprentissage est nécessaire au système afin d'acquérir des connaissances plus précises sur l'utilisateur. Cette étape, préalable aux premières adaptations, mais qui se poursuit pendant toute la période d'utilisation, peut être plus ou moins longue selon l'activité de l'utilisateur.

3.2 CONDITIONS POSSIBLES

3.2.1 METHODES D'EVALUATION AVEC *VERSUS* SANS UTILISATEURS

Dans l'évaluation de systèmes adaptatifs, et d'une manière générale, les méthodes d'évaluation peuvent être classées en deux catégories selon qu'elles impliquent ou non l'utilisateur pendant l'évaluation. Nous en citons ici les principales mais sans les développer :

- les tests utilisateurs [Nielsen 1993] ;
- les entretiens [Blanchet and al. 2000] ;
- les questionnaires [Blanchet and al. 2000] ;
- le paradigme du Magicien d'Oz [Amalberti and Valot 1992] [Preece and al. 1994] ;
- les *focus group* [Greenbaum 1998] [Nielsen 1993].

Les méthodes n'impliquant pas l'utilisateur dans l'évaluation ont recours à un ou plusieurs experts dans des domaines tels que l'ergonomie, la sociologie ou la psychologie cognitive. Parmi ces méthodes, on trouve :

- l'évaluation heuristique [Nielsen and Molich 1990] ;
- l'évaluation experte [Gena 2006] ;
- l'analyse des tâches [Diaper 1990] ;
- l'inspection cognitive [Lewis and al. 1990] [Polson and al. 1992].

Pour évaluer nos démonstrateurs, des évaluations expertes ont tout d'abord été conduites en cours de conception. Puis, des évaluations avec des utilisateurs en laboratoire et sur le terrain ont été menées.

3.2.2 EN LABORATOIRE *VERSUS* SUR LE TERRAIN

Deux conditions d'évaluation sont possibles : l'évaluation en laboratoire et l'évaluation sur le terrain [Jambon and al. 2006]. Les termes anglo-saxons consacrés sont *field experimentation* [Goodman and al. 2004], *experimentation in the field* [Kellar and al. 2005] ou encore *experimentation in the wild*.

L'évaluation en laboratoire se décline généralement sous la forme de tests utilisateurs. Cette méthode peut donner des résultats fiables. Cependant, elle ne prend pas en compte tous les facteurs pouvant intervenir dans l'interaction homme-machine et ne permet pas de cerner les usages réels des dispositifs [Norman 1999] [Kjeldskov and Paay 2005].

L'évaluation sur le terrain présente l'avantage de fournir des données sur l'utilisation réelle des systèmes. Cependant, elle n'est pas envisageable pendant les premières phases de la conception car elle requiert un système fonctionnel du point de vue technique et pose les problèmes d'enregistrement des observations et de validité des analyses [Jambon et al. 2006].

Les outils utilisés pour l'évaluation des usages sur le terrain sont variés et dépendent fortement des contraintes d'utilisation du système (comme la mobilité). Sans être exhaustifs, nous pouvons citer les techniques d'auto-documentation utilisées par [Sohn et al. 2008] et [Hulkko et al. 2004]. Ces derniers ont recueilli l'activité des utilisateurs par l'intermédiaire de journaux de bord (diary) ou encore de photographies, de collages et d'échantillonnage d'expérience (ESM). Les utilisateurs devaient décrire leurs activités, leurs besoins et leurs attentes selon différents contextes d'usage, prendre des photographies illustrant leurs contextes ou encore répondre à des questionnaires en ligne servant à décrire leurs activités et leurs ressentis. En outre, l'analyse de logs (automatic logging) peut être aussi utilisée pour identifier les usages réels. Ces outils tracent

l'utilisation d'un dispositif en contexte réel (fréquence d'usage, temps d'utilisation, périodes d'utilisation ou de non utilisation, etc.) et ils sont généralement complétés par des questionnaires ou des entretiens avec des utilisateurs (explicitation de leurs usages).

3.3 METHODES POUR L'EVALUATION DES SYSTEMES ADAPTATIFS

Dans [Chin 2001], l'auteur passe en revue trente-deux publications contenant des évaluations empiriques pour divers systèmes adaptatifs. Il réalise une synthèse des méthodes utilisées pour les évaluer. Il note que les critères d'évaluation peuvent inclure des mesures tant quantitatives (par exemple, le temps de réalisation d'une tâche) que qualitatives (par exemple, le feedback de l'utilisateur sur l'utilité du système). Les données ainsi obtenues sont traitées à l'aide d'outils statistiques comme par exemple l'analyse de la variance (ANOVA). Ce type d'évaluation vise à donner des résultats généralisables car statistiquement valides et fidèles. Deux approches classiques sont utilisées pour l'évaluation empirique des systèmes adaptatifs [Tobar 2003] :

- L'évaluation *comparative* : elle consiste à évaluer, par des tests utilisateurs généralement réalisés en laboratoire, deux versions du système : l'une avec adaptation et l'autre sans. La valeur de l'adaptation est la différence d'utilité entre les deux versions [Höök 1997] [Höök 2000] [Lieberman and al. 2004].
- L'évaluation *structurée* ou *layered evaluation* : cette approche a été introduite par [Totterdell and Boyle 1990]. Elle consiste à évaluer séparément les éléments adaptatifs du système pour identifier l'intérêt et la pertinence des adaptations proposées [Brusilovsky and al. 2001] [Tobar 2003] [Brusilovsky and al. 2004]. La valeur de l'adaptation est évaluée à deux niveaux [Totterdell and Rautenbach 1990] [Benyon and Murray 1993] [Brusilovsky 1996] : au niveau du modèle utilisateur (représentation correcte des caractéristiques, buts, préférences, etc. de l'utilisateur [Brusilovsky 1996]) et au niveau des choix d'adaptation (validité et pertinence des adaptations proposées). L'évaluation structurée n'a pas pour objectif de remplacer les pratiques d'évaluation existantes : chaque adaptation peut précisément être évaluée avec ces techniques.

4. DEMONSTRATEURS

Cette section présente deux démonstrateurs de plasticité : PlasticSMS et MobilePlastic. Le troisième est EMMA : il a fait l'objet de la section 2.1. Les démonstrateurs illustrent les concepts clé de la plasticité : le remodelage, la redistribution et le contrôle utilisateur. La redistribution n'est pas assurée dans EMMA qui, par contre, existe en deux versions : avec ou sans IHM de contrôle. A l'inverse, PlasticSMS et MobilePlastic couvrent la redistribution et intègrent une IHM de contrôle.

4.1 PLASTICSMS

PlasticSMS est un prototype développé en C++. Il offre un service de messagerie plastique permettant de communiquer via SMS depuis un téléphone mobile ou un ordinateur. En situation de mobilité, le téléphone est utilisé de manière traditionnelle. Lorsque l'environnement s'augmente de plates-formes telles qu'un ordinateur, le service offre à l'utilisateur la possibilité de redistribuer l'IHM vers une seconde plate-forme lorsque la proximité de celle-ci est détectée (connexion Bluetooth) (Figures 7 et 8).



Figure 7: Redistribution de l'IHM de PlasticSMS.

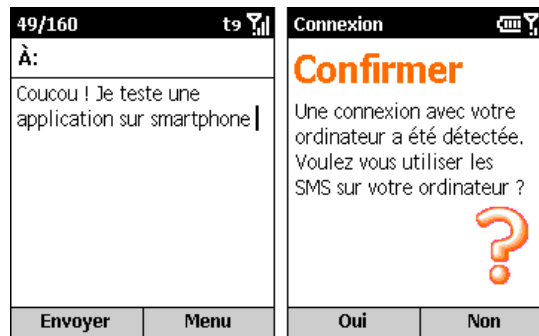


Figure 8: IHM de PlasticSMS sur téléphone.

L'IHM de saisie est alors reproduite sur l'ordinateur, le message pouvant être saisi indifféremment sur le mobile ou l'ordinateur. Les avantages supposés pour l'utilisateur sont un confort visuel accru par la taille de l'écran et une saisie facilitée par la disponibilité d'un clavier en entrée. De plus, l'IHM est dotée de capacité de remodelage. Lorsque l'utilisateur agrandit (respectivement réduit) la taille de la fenêtre, des fonctionnalités supplémentaires apparaissent ou disparaissent (Figure 9). Ainsi, l'utilisateur peut modifier la taille des caractères, envoyer le message ou encore accéder à ses contacts enregistrés dans le répertoire du téléphone.

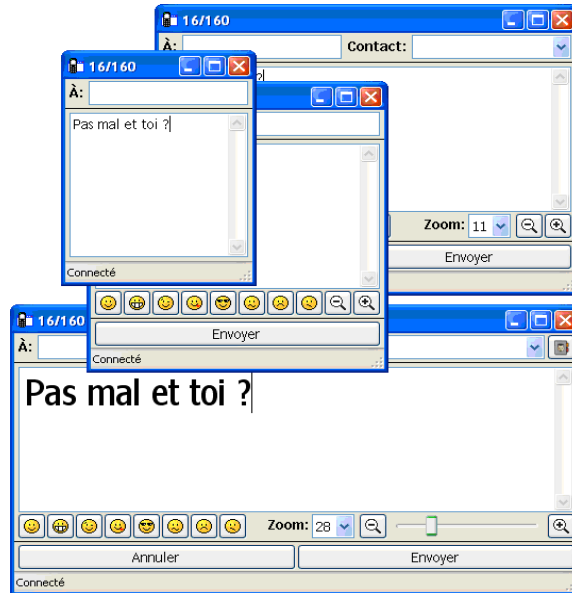


Figure 9: IHM de PlasticSMS sur PC.

L'observabilité et le contrôle des adaptations sont accessibles par l'intermédiaire d'une IHM dite « de contrôle utilisateur » (Figure 10). L'utilisateur dispose d'options permettant d'accepter ou de refuser la redistribution, de demander une confirmation à chaque adaptation proposée par le système, d'adapter automatiquement sans être consulté ou encore de refuser toute adaptation (utilisation classique du SMS).

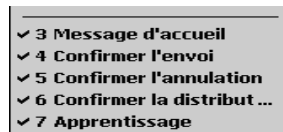


Figure 10: Un exemple d'options accessibles à travers l'IHM de contrôle.

4.2 MOBILEPLASTIC

MobilePlastic est une application développée en C#. Elle propose des fonctions complémentaires à PlasticSMS. Elle permet de gérer les contacts du téléphone mobile, de consulter l'historique des appels entrants et sortants, d'enregistrer ses SMS (fichiers textes) et de consulter le niveau de batterie du téléphone (Figure 11). En terme de plasticité, elle est dotée de possibilités de redistribution (les actions se réalisant sur l'ordinateur) mais pas de remodelage.



Figure 11: MobilePlastic sur PC.

Tout comme dans PlasticSMS, les utilisateurs peuvent contrôler l'adaptation. Des messages de confirmation apparaissent sur le téléphone mobile lorsque l'utilisateur réalise des actions critiques telles que l'ajout, la modification, la suppression d'un contact, l'envoi ou l'annulation d'un SMS. En outre, la liste des contacts du répertoire est réorganisée en fonction de la fréquence des appels.

5. EVALUATION DES DEMONSTRATEURS ET RESULTATS

Cette section décrit le déroulement des évaluations et présente les principaux résultats. Elle s'organise en deux parties. La première est dédiée aux évaluations effectuées en laboratoire ; la deuxième relate les évaluations sur le terrain.

5.1 PLASTICSMS ET MOBILEPLASTIC : EVALUATIONS EN LABORATOIRE

Les évaluations de PlasticSMS et MobilePlastic se sont déroulées en laboratoire. Elles ont donné lieu à un focus group et des tests utilisateurs.

5.1.1 OBJECTIFS

Les principaux objectifs du focus group sont les suivants :

- cerner les attentes des usagers (identifier les principaux besoins pouvant être satisfaits par ce type de services, recueillir des idées nouvelles en termes de fonctionnalités, des propositions de services, etc.),
- appréhender l'intérêt perçu de la plasticité (le ressenti des participants face à un système qui capitalise leur préférences, modifie l'IHM en fonction de leurs actions, etc.),
- appréhender l'intérêt de la négociation avec le système (contrôle utilisateur nécessaire ou non, contraignant ou non, etc.),
- appréhender le ressenti des utilisateurs vis-à-vis du caractère potentiellement intrusif de ce type de système.

Les principaux objectifs des tests utilisateurs sont les suivants :

- évaluer la compréhension du concept de plasticité,
- déterminer l'intérêt d'IHM plastiques (avantages et inconvénients).

Nous indiquons ci-dessous le profil des différents participants (Tableau 1). La section suivante précise les protocoles.

	Focus Group	Tests utilisateurs
Effectif		
Nombre de participants	5 participants	20 participants
Caractéristiques		
Sexe	2 hommes 3 femmes	11 hommes 9 femmes
Age	Moyenne d'âge : 39 ans	Moyenne d'âge : 36 ans

	Age minimum : 20 ans Age maximum : 57 ans	Age minimum : 16 ans Age maximum : 71 ans
Expérience mobile	Aucun participant ne devait avoir ou avoir eu un mobile de type SPV. Ainsi, tous avaient le même niveau d'expérience.	
Echantillon des utilisateurs	Deux catégories d'âge ont été ciblées : un groupe de 16 à 25 ans et un groupe de 45 ans et plus.	
Usage du mobile	Un usage régulier du mobile était demandé.	
Usage de l'ordinateur	Un usage régulier de l'ordinateur était demandé.	

Tableau 1: Profil des participants du focus group et des tests utilisateurs.

5.1.2 PROTOCOLES

Le focus group a duré quatre-vingt dix minutes. La séance a été enregistrée en audio et vidéo et a donc impliqué, en plus des cinq participants, un modérateur et un observateur en régie. La séance s'est déroulée de la façon suivante :

1. Accueil des participants
2. Questionnaire de caractérisation des participants (profils, habitudes, usages des mobiles et de l'ordinateur, etc.)
3. Présentation retro-projetée de la plasticité (illustration de différents exemples de prototypes)
4. Thèmes abordés lors de la discussion :
 - avantages et inconvénients de l'adaptativité,
 - intérêt de l'adaptation aux plates-formes, à l'utilisateur, à l'environnement, pour soi et pour les autres,
 - avantages et inconvénients des démonstrateurs.
5. Clôture et synthèse de la séance.

La passation des tests était individuelle et durait environ deux heures. Les tests utilisateurs ont regroupé vingt utilisateurs. Les séances se sont déroulées de la façon suivante :

1. Accueil des participants
2. Questionnaire de caractérisation des participants (profils, habitudes, usages des mobiles et de l'ordinateur, etc.)
3. Illustration d'exemples de différentes formes d'adaptation
4. Premier entretien permettant d'appréhender l'intérêt perçu des IHM plastiques
5. Tâches de familiarisation avec les démonstrateurs
6. Scénarios d'utilisation sur PlasticSMS :
 - envoi d'un SMS à partir du téléphone avec modification de la taille de la police,
 - envoi d'un SMS à partir de l'ordinateur.
7. Second entretien permettant de recueillir l'avis des utilisateurs sur l'utilisation de PlasticSMS
8. Scénarios d'utilisation sur MobilePlastic :
 - consultation de l'état de la batterie,
 - enregistrement et modification d'un contact,
 - consultation d'un SMS enregistré,
 - suppression d'un contact,
 - émission d'un appel depuis le téléphone mobile.

Le contrôle des adaptations et l'apprentissage des préférences des utilisateurs ont été simulés. Pour cela, tous les participants ont réalisé les scénarios dans le même ordre :

- Pour les premiers scénarios, toutes les adaptations proposées par le système étaient négociées avec l'utilisateur.
- Pour les derniers scénarios, les adaptations étaient réalisées sans négociation auprès de l'utilisateur. De plus, l'évolution des scénarios illustre une réorganisation des contacts après avoir envoyé des messages et appelé les mêmes contacts.

5.1.3 LIMITES

Nous identifions quelques limites aux protocoles :

- Concernant le focus group, les retours utilisateurs sont basés uniquement sur le déclaratif des participants. Par conséquent, ces résultats sont à utiliser avec précaution, les participants ayant tendance à surestimer ou sous-estimer certains usages.
- La mise en situation d'un panel réduit d'utilisateurs (vingt pour les tests utilisateurs) peut paraître restreinte. Toutefois, cette méthode permet d'obtenir une tendance qualitative en mettant en évidence la perception des utilisateurs et les principales difficultés d'utilisation.
- En laboratoire, les utilisateurs sont mis en situations d'utilisation les plus réalistes possible. Néanmoins il est difficile d'approcher un large éventail de contextes d'usage (comme les usages en mobilité par exemple).
- Les vingt utilisateurs du test ont tous passé les scénarios dans le même ordre (étant donné notre volonté d'illustrer l'apprentissage de leurs actions par le système). Un risque sur l'apprentissage système doit être souligné.

5.1.4 PRINCIPAUX RESULTATS

Tout d'abord, nous présentons la perception des utilisateurs vis-à-vis des concepts clé de la plasticité. Puis, nous abordons les performances observées et les difficultés rencontrées par les participants lors des tests utilisateurs.

Perception des utilisateurs

Pour évaluer le ressenti de tous les participants au focus group et aux tests utilisateurs, un premier entretien a été réalisé sur la base d'une présentation rétro-projetée illustrant différents exemples de services et d'applications adaptatives :

- l'adaptation à la plate-forme a été perçue comme étant la plus utile (vingt-sept occurrences) ;
- l'avantage est une optimisation des tâches et une utilisation plus conviviale (vingt-huit) lorsque le dispositif sur lequel l'IHM est déportée possède des caractéristiques plus confortables (écran plus grand ou encore présence d'un clavier) ;
- les principaux inconvénients cités par les participants sont la confidentialité des données (huit) et la crainte de perdre le contrôle de ses actions (douze) ;
- neuf participants ont déclaré qu'ils utiliseraient ce type de services illustrés par les démonstrateurs dans la sphère professionnelle et familiale ;
- les idées de fonctionnalités adaptatives les plus fréquemment citées seraient de pouvoir disposer d'un agenda adaptatif (quatorze occurrences) sur mobile (rappel du contrôle technique de la voiture, de la liste des courses, etc.), une

localisation GPS donnant des informations sur la localisation de son interlocuteur (huit) ou encore des informations circonstanciées en voiture ou en faisant ses courses (quinze).

Performance des participants (tests utilisateurs)

Tous les résultats d'utilisation des deux démonstrateurs ne sont pas exposés dans cet article. Nous avons extrait les performances de certains scénarios et essentiellement ceux relatifs aux leviers de l'adaptation (redistribution, remodelage) ainsi qu'au contrôle utilisateur.

Tâche	Réussite	Temps moyen		Nombre d'assistances
		Jeunes	Plus de 45 ans	
Elargir la fenêtre de l'application	13/20	13,14	22,5	6
Saisir le SMS	20/20	21,1	21,9	1
Envoyer le SMS	20/20	36,9	63,5	13
Consulter le niveau de batterie du téléphone	20/20	3,1	14,8	11
Ajouter un contact	20/20	3,1	44,6	15
Modifier un contact	20/20	34,3	41,2	15

Tableau 2: Performance des utilisateurs (PlasticSMS et Mobile Plastic)

Les utilisateurs ont majoritairement réussi les différents scénarios. En revanche, pour certaines tâches, l'intervention de l'expérimentateur a été nécessaire. De plus, nous avons essentiellement constaté des différences de temps entre le groupe des jeunes et celui des plus de 45 ans. A titre d'exemple, la consultation de l'état de la batterie sur l'ordinateur a posé plus de problèmes aux seniors. Le temps moyen d'exécution des utilisateurs de ce groupe ($M = 14,8$) est considérablement plus élevé que celui des juniors ($M = 3,1$). Les informations ne semblaient pas assez visibles selon les commentaires des testeurs.

Remodelage de l'IHM

Pour le premier scénario d'utilisation de PlasticSMS, nous avons demandé aux testeurs d'envoyer un message texte à partir de l'ordinateur. L'écran était identique à celui du mobile. L'utilisateur devait ensuite élargir la fenêtre pour voir apparaître les options du répertoire et de l'envoi.



Figure 12: Illustration du remodelage de l'IHM.

Nous avons observé des échecs uniquement au niveau du remodelage. En effet, la plupart des participants ont bien vu la possibilité d'agrandir la fenêtre, mais n'ont pas perçu le lien avec l'existence de fonctions supplémentaires. Treize utilisateurs ont jugé l'indication de l'agrandissement de la fenêtre incompréhensible. Notons que parmi les participants qui ont réussi la tâche (treize utilisateurs), sept sont ceux qui ont adopté la stratégie d'agrandir la fenêtre en utilisant l'icône d'agrandissement de Windows et six (sur treize) ont utilisé la flèche.

Contrôle utilisateur

Pour le premier scénario, la confirmation des actions de redistribution de l'IHM sur l'ordinateur et d'envoi du message étaient accessibles sur le téléphone. Douze utilisateurs les ont jugées inutiles et préféreraient qu'elles soient visibles uniquement sur l'ordinateur. Les avis sont identiques pour les autres scénarios ayant nécessité de confirmer les actions sur le mobile.

Perception des éléments adaptatifs

Le dernier scénario d'utilisation consistait à émettre un appel. Les utilisateurs pouvaient alors constater la réorganisation de la liste des contacts en fonction de la fréquence. Notre objectif était de recueillir leur avis sur cet élément adaptatif. Dix-huit participants l'ont jugé utile car cela permet d'éviter des manipulations supplémentaires (gain de temps).

5.1.5 DISCUSSION

Les évaluations en laboratoire nous ont permis de collecter un certain nombre d'observations objectives et déclaratives sur l'utilisation d'IHM plastiques. Néanmoins, les limites des méthodologies que nous avons détaillées plus haut nous conduisent à établir un certain nombre de constats :

- En laboratoire, il est difficile de simuler les adaptations à l'utilisateur et notamment à leurs préférences et à leurs usages. En effet, l'utilisation des dispositifs sur une très courte période (deux heures) rend difficile une réelle prise en compte de ces caractéristiques.
- Pour mettre en évidence le remodelage des IHM du téléphone sur l'ordinateur, nous avons opté pour une IHM similaire au mobile et pouvant s'agrandir pour proposer des fonctionnalités supplémentaires et des modes d'interaction différents. Globalement, tous les participants n'ont pas perçu la pertinence de conserver dans un premier temps un affichage cohérent avec l'IHM du mobile. Il semble préférable d'afficher directement l'IHM dans sa globalité sur l'ordinateur.
- Evaluer la redistribution des IHM sur le terrain auprès d'utilisateurs nécessite de développer des prototypes stables pouvant être installés et utilisés facilement par les utilisateurs chez eux ou en mobilité.
- Nos évaluations et le retour des utilisateurs ont été relativement biaisés par les différences des IHM en termes de design. En effet, les participants ont exprimé des améliorations en termes d'organisation des informations entre les deux démonstrateurs. Il apparaît donc nécessaire de prendre en compte dès la conception les objectifs de l'évaluation et la diversité des scénarios d'usage qui pourront être soumis aux testeurs.
- En laboratoire, il n'est pas possible d'identifier les différences d'usage pour une même application selon les plates-formes utilisées. Ainsi, il serait intéressant d'observer si la fréquence d'envoi des messages varie selon l'utilisation du téléphone mobile ou de l'ordinateur. De la même manière, les usages sont-ils reproduits à l'identique selon le dispositif utilisé ?
- Concernant le contrôle utilisateur (IHM de contrôle), nous avons observé un consensus sur la nécessité d'intégrer cette IHM sur l'ordinateur plutôt que sur le mobile. Nous avons été relativement surpris par ce résultat car il nous semblait important de notifier certaines actions sur le mobile pour rendre observable la répercussion des actions sur ce dispositif.

5.2 EMMA : EVALUATION SUR LE TERRAIN

L'évaluation d'EMMA s'est produite sur le terrain. En effet, EMMA impose des conditions réelles de mobilité pour collecter des données d'usage sur une période de plusieurs semaines. En outre, des évaluations sur le terrain sont nécessaires pour l'apprentissage des contextes clés de l'utilisateur.

5.2.1 PROTOCOLE

Dans le but de recueillir les usages et les impressions sur les deux versions d'EMMA, l'évaluation a été menée auprès de deux groupes différents de participants. Dans la première version, les utilisateurs avaient la possibilité, grâce à l'IHM de contrôle, de paramétrer les adaptations. Ils pouvaient ajouter, modifier et supprimer des contextes d'usage clés à tout moment. Dans la deuxième version, ils étaient tributaires des propositions du système. Le contrôle qu'ils exerçaient était limité à la négociation au fil des adaptations (ils ne disposaient pas de l'IHM de contrôle). Les tableaux ci-après récapitulent le protocole d'évaluation et le profil des participants.

Étapes	Objectifs
Lancement de l'étude	Remise des mobiles équipés d'EMMA aux participants Explication du déroulement de l'étude Remise d'un support expliquant le fonctionnement de l'étude Passation d'un questionnaire de caractérisation
Suivi	Envoi régulier de données d'usage collectées et stockées par e-mail par les participants
Conclusion de l'étude	Récupération des mobiles auprès des participants Entretien individuel avec chaque participant

Tableau 3: Etapes du protocole d'évaluation

Effectifs	
Nombre de participants initial	12 participants (divisés en 2 groupes)
Nombre d'abandons / échecs	1 abandon / 1 échec
Nombre de participants retenus	10 participants
Caractéristiques	
Sexe	5 hommes 5 femmes
Age	Moyenne d'âge : 19 ans Age minimum : 18 ans Age maximum : 23 ans
Equipement	Tous les participants utilisent un téléphone mobile. Ils ont en moyenne eu 2,9 mobiles avant leur mobile actuel.
Expérience mobile	Tous les participants utilisent un mobile (depuis au moins un an à plus de trois ans)
Usage du mobile	
Nombre de contacts enregistrés	8 participant ont entre 50 à plus de 100 contacts 2 participants ont entre 20 et 50 contacts
Usage des SMS	5 participants envoient entre 30 et 200 SMS par mois 5 participants envoient entre 15 et 30 SMS par mois
Usage des MMS	3 participants n'ont jamais envoyé de MMS 7 participants envoient de 2 à 10 MMS par mois
Usage des emails	3 participants envoient occasionnellement des emails
Usage de l'Internet mobile	4 participants utilisent occasionnellement cette fonction

Tableau 4: Profil des participants

5.2.2 LIMITES

Nous identifions quelques limites dans l'évaluation :

- Pour cette évaluation, une démarche qualitative a été mise en place avec un panel réduit d'utilisateurs. En effet, mettre en place une démarche quantitative sans tout d'abord tester l'application avec un petit panel d'utilisateurs est risqué et coûteux. Ainsi, les résultats développés dans cet article s'appliquent uniquement dans le cadre de cette étude, de l'application testée et de la population sollicitée (à savoir des jeunes plutôt technophiles).

- Les résultats relatifs à l'acceptabilité des adaptations proposées s'appliquent uniquement aux adaptations testées au cours de cette étude : réorganisation des fonctionnalités et automatisation de certaines tâches comme le changement de profil de sonnerie.
- Mener une étude sur le long terme est nécessaire pour évaluer des applications adaptatives. Les six semaines de test se sont avérées suffisantes pour le recueil de données, l'apprentissage, les propositions d'adaptation et les révisions des adaptations selon les usages des utilisateurs. Néanmoins, les participants ont eu tendance à oublier certaines informations comme à quel moment les premières adaptations ont été proposées, dans quels contextes étaient-ils ou encore comment ont-ils perçu les adaptations proposées ?
- Pour pallier aux difficultés de recrutement des utilisateurs ayant un mobile bien spécifique comme le SPV C100 et qui par ailleurs n'était plus en vente au moment de l'étude, nous avons prêté les téléphones aux participants. En outre, cela nous permettait de contrôler le niveau d'expérience des participants avec cette gamme de téléphone. Néanmoins, nous avons constaté que, par exemple, les utilisateurs ont très peu personnalisé le mobile (fond d'écran, etc.) alors qu'ils le font en règle générale. Ceci s'explique probablement par le fait qu'ils devaient restituer le matériel en fin d'expérience. Certains usages sont donc à prendre avec précaution car ils peuvent être dus à l'utilisation d'un mobile prêté.

5.2.3 PRINCIPAUX RESULTATS

Les résultats obtenus sont issus d'une part des fichiers de logs et d'autre part des verbatim des participants lors des entretiens individuels. Un programme d'exploitation des logs a été implémenté afin de faciliter le traitement des résultats. Tous les entretiens ont été retranscrits et une analyse du contenu a été conduite.

Fonctionnement de l'application

Les dix participants ont tous utilisé EMMA durant au moins six semaines. Nous avons ainsi collecté vingt cinq mille cinq cent quatre-vingt onze (25591) données tout au long de la durée de cette expérimentation terrain. Ce nombre permet de constater la stabilité fonctionnelle et opérationnelle de l'application mais aussi les usages effectifs de l'application par les testeurs.

Les phases d'apprentissage sont en moyenne au nombre de quatre-vingt trois par participant. Ce nombre nous indique en moyenne la fréquence a priori des sollicitations du système. Malheureusement, nous n'avons pas tracé en parallèle des phases d'apprentissage celles qui donnaient lieu à des propositions système (si le système détectait de nouveaux usages et donc des adaptations) ou celles pour lesquelles le système ne détectait aucun nouvel usage (donc aucune proposition). Or, ces données seraient pertinentes pour mesurer objectivement la fréquence des dialogues entre le système et l'utilisateur. Au cours des entretiens, tous les participants ont jugé les propositions système trop fréquentes mais aucune donnée ne nous permet de quantifier dans quelle mesure.

Contextes clés

Le nombre moyen de contextes clés proposés par le système et/ou ajoutés (pour les participants disposant de l'IHM de contrôle) était de 4,7.

Une participante n'a pas eu de proposition de contexte tout au long de l'étude. En effet, les algorithmes de "clustering" étaient conditionnés à l'usage du changement de profil de sonnerie pour identifier les contextes clés des utilisateurs. Or, cette participante a activé le mode vibreur ou silencieux en passant par le menu "Paramètres" et "Sons" mais jamais en utilisant les profils de sonnerie. Néanmoins le système lui a proposé des adaptations telles que la réorganisation des menus.

La répartition du nombre de contextes clés moyen selon la version d'EMMA (avec ou sans contrôle utilisateur) est la suivante :

	Nombre de contextes clés
Participants avec IHM de contrôle (N=5)	6,4
Participants sans IHM de contrôle (N=5)	3
Ensemble des participants (N=9)	4,7

Tableau 5 : Nombre moyen de contextes selon la version utilisée d'EMMA

Les participants qui disposaient des options permettant de paramétrer et de personnaliser les contextes clés les ont effectivement utilisées. Ils ont en effet ajouté eux-mêmes des contextes en plus de ceux proposés par le système. Pour ce groupe, on observe une répartition du nombre de contextes allant de quatre à dix contextes selon les participants. Quant aux groupes ayant EMMA sans IHM de contrôle, ce nombre varie de zéro à cinq contextes.

Nombre de contextes réellement utiles

En analysant les usages des participants, nous avons constaté que le nombre de contextes réellement utiles, à savoir pour lesquels on observait des usages, pouvait être différent du nombre total de contextes. Deux participants sur neuf ont des contextes vides. Six participants ont des contextes quasiment vides (c'est-à-dire représentant entre 1 et 2% de leurs usages). La répartition des contextes utiles (représentant 80% de leurs usages) est la suivante :

Nombre de contextes utiles	Nombre de participants (N=9)
2 contextes utiles	5
3 contextes utiles	2
1 contexte utile	2
Nombre moyen de contextes utiles	2,11

Tableau 6 : Nombre de contextes réellement utiles (avec des usages constatés)

On obtient ainsi une moyenne de contextes utiles de 2,11 par participant. La majorité des participants a eu des usages distincts dans deux voire trois contextes d'usage. Pour compléter ces données, nous avons calculé le pourcentage de temps d'activité (avec des usages) dans les différents contextes. Il apparaît bien un usage bipolaire voire tripolaire pour la grande majorité des participants (80% des usages sont réalisés dans pas plus de trois contextes).

Ajout et modification de contextes

Pour déterminer l'intérêt de l'IHM de contrôle et son usage réel par les participants, nous avons tracé les actions réalisées dans ce menu et nous avons recueilli l'avis de ces utilisateurs quant aux actions réalisées lors des entretiens.

Tout d'abord, nous constatons des écarts entre ce que déclarent avoir fait les utilisateurs et les données recueillies dans les logs. Etant donné la longueur de l'étude, les participants ont eu dû mal à se souvenir précisément des événements. Ensuite, en analysant les logs des utilisateurs, certaines informations sont manquantes car nous avons tracé au global le nombre de contextes proposés par le système et ceux ajoutés par les utilisateurs et à quel moment (date en cours). Une extraction d'un fichier log illustre le format et les données recueillies concernant les propositions du système :

Ajout de contextes :

10/04/2008 = 2

13/04/2008 = 1

26/04/2008 = 1

Nombre d'ajouts de contextes : 4

Modification de contextes :

10/04/2008 = 1

28/04/2008 = 2

Nombre de modifications de contextes : 3

Suppression de contextes :

28/04/2008 = 2

Nombre de suppressions de contextes : 2

Il aurait été intéressant de collecter ces données sous forme de session indiquant les actions proposées par le système, celles réalisées par l'utilisateur, le moment et son contexte (sa localisation géographique, le jour, la date et l'heure). Ainsi, notre compréhension pourrait être meilleure. En outre, il convient aussi de tracer finement les paramétrages modifiés dans l'IHM de contrôle et ce par contexte clé (cf. illustration de certaines options disponibles à partir de ce menu et que l'on pourrait tracer).

En effet, nous avons tracé le nombre d'occurrences de consultation du menu "EMMA" mais nous n'avons pas collecté les options modifiées dans ce menu.

En terme d'usage, nous avons donc recoupé les verbatim des utilisateurs avec les données objectives des fichiers logs. Parmi les cinq participants disposant de l'IHM de contrôle, on recueille :

- trois ont eux-mêmes ajouté des contextes clés (l'un pour s'amuser et les deux autres ont surtout utilisé ces contextes comme moyen de commutation rapide de profil de sonnerie à partir du plug-in EMMA) ;
- un seul participant a refusé un contexte proposé par le système car il ne comprenait pas pourquoi EMMA lui proposait cet ajout ;
- deux utilisateurs ont tenté de supprimer des contextes qu'ils avaient eux-mêmes ajoutés (mais la suppression des contextes ne fonctionnait pas) ;
- tous les participants ont modifié certaines options accessibles dans le menu comme la suppression de la vocalisation, les fonds d'écran ou encore les sonneries pour un contexte donné.

Tous les utilisateurs étaient satisfaits de disposer de ce menu mais tous ont estimé que certaines propositions étaient difficilement compréhensibles. Ils auraient souhaité disposer de plus d'informations.

Intitulé des contextes

Tous les contextes pour lesquels nous observons des usages ont été nommés par des lieux (comme "maison", "école", "stage" ou encore "Quimper") ou encore par des périodes temporelles comme "week-end" ou "vacances". Il semble donc que, selon le point de vue utilisateur, la localisation géographique ait un poids prépondérant dans la constitution de contextes clés. Cette hypothèse reste quand même à vérifier auprès d'utilisateurs n'ayant reçu préalablement aucun exemple d'intitulé de contexte. En effet, lors de la séance d'explication à propos du déroulement de l'expérimentation, un descriptif contenant des exemples d'intitulés de contexte avait été fourni aux utilisateurs.

Usages des applications par contexte clé

En terme d'utilité de l'application, une des questions qui se pose est la pertinence de la réorganisation des fonctionnalités selon les contextes clés. En effet, il est légitime de vérifier si les utilisateurs ont bien des usages distincts selon les contextes clés. Ainsi, l'application peut-elle faciliter l'accès aux fonctionnalités les plus utilisées dans un contexte donné (ou du moins augmenter la rapidité d'accès) ? Pour répondre à cette question, nous avons relevé pour les neuf participants (ayant eu des contextes clés) quels étaient globalement la composition et l'ordre des applications utilisées selon les différents contextes clés.

Nous avons centré cette analyse sur les neuf premières fonctionnalités du téléphone. Prenons l'exemple d'un testeur, ce dernier a eu trois contextes clés. Pour chaque contexte, la réorganisation est notée de la façon suivante :

- (0) si elle est identique,
- (1) si un menu est différent,
- (2) si deux menus sont différents.

Contexte 1	Contexte 2	Contexte 3	Notation
Paramètres	Paramètres	Paramètres	0
Messages	Messages	Messages	0
Contacts	Contacts	Enr.Video	1
App.photo	App.photo	Photos Vidéos	1
Réveil	Photos Vidéos	App.photo	2
Enr.Video	Enr.Video	WindowsMedia	1
Photos Vidéos	WindowsMedia	Contacts	2
ActiveSync	Journal des appels	Journal des appels	1
Journal des appels	ActiveSync	ActiveSync	1

Tableau 7 : Procédure de calcul des différences de réorganisation entre les contextes clés

Le diagramme ci-dessous résume les différences observées pour l'ensemble des participants sur tous leurs contextes clés utiles.

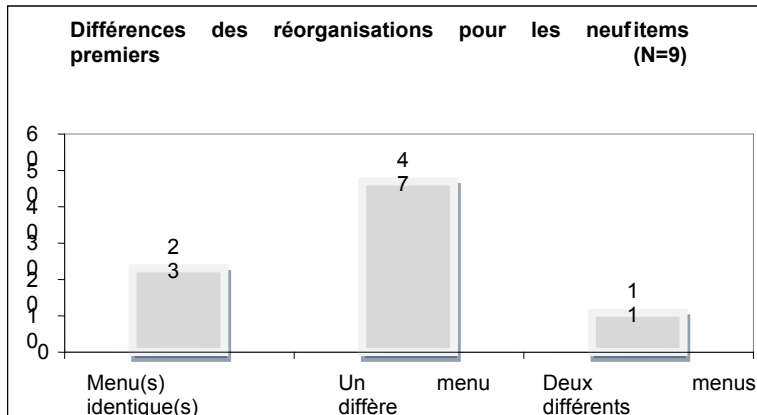


Figure 13 : Différence de réorganisation des neuf premières fonctionnalités (tous les testeurs confondus)

Au travers de ces résultats, nous constatons que globalement les utilisateurs ont bien des usages différents selon les contextes puisque l'ordre des menus varie (dans cinquante huit cas). Dans uniquement vingt trois cas, les menus sont identiques pour différents contextes.

En revanche, si nous regardons plus précisément la réorganisation par contexte clé utile des quatre premières fonctionnalités les plus utilisées, nous observons que dans dix-neuf cas les fonctionnalités sont réordonnées dans le même ordre quel que soit le contexte et dans vingt-deux cas l'ordre des fonctionnalités varie selon le contexte. Il est donc difficile de trancher sur l'utilité de cette adaptation proposée par EMMA. Dans la moitié des cas, la réorganisation de l'ordre des menus n'est pas pertinente, puisque les principaux usages sont les mêmes. Dans l'autre moitié, la réorganisation a permis de remonter des fonctionnalités qui ne sont utilisées que dans certains contextes. Une solution à investiguer consisterait à proposer une partie du menu statique pour les quatre premières fonctionnalités et une partie dynamique pour les autres. Ainsi, les applications a priori plus secondaires pourraient être accessibles plus rapidement.

Décisions utilisateurs

Afin de proposer des adaptations a priori pertinentes du point de vue des utilisateurs, il est nécessaire de les négocier. Pour cela, non seulement l'utilisateur doit pouvoir prendre connaissance des actions réalisées par le système et les comprendre. Mais il doit pouvoir aussi agir sur les adaptations soit en les refusant soit en les acceptant. C'est dans cette optique que nous avons intégré un certain nombre de messages systèmes permettant à l'utilisateur de manifester son accord ou son refus aux propositions d'EMMA. En outre, ces décisions utilisateurs sont tracées par le système et apprises afin soit de ne plus proposer les adaptations toujours refusées soit de les automatiser au-delà d'un certain nombre d'acceptations. Pour finir, ces données nous éclairent aussi sur la pertinence des adaptations proposées par le système.

Nous avons collecté globalement les décisions utilisateurs pour toutes les adaptations proposées par EMMA, c'est-à-dire : les propositions d'ajouts de contextes clés, les réorganisations des menus, les changements de profils de sonnerie, les changements de contextes clés, de sonneries, de fonds d'écran, de page d'accueil et de couleurs et l'automatisation des adaptations.

	Acceptation	Refus	Automatique	Pas pris en compte (time-out)
Toutes adaptations confondues	26%	32%	26%	16%
Ajout de contextes clés	40%	19%	0%	41%
Changement de contextes clés	18%	50%	11%	21%
Réorganisation des menus	46%	19%	22%	13%
Changements de profils de sonnerie	51%	30%	14%	5%

Tableau 9 : Décisions utilisateurs selon le type d'adaptation proposé

Globalement, les participants ont, d'une manière générale, un peu plus rejeté les adaptations proposées qu'ils ne les ont acceptées (32 versus 26%). Par ailleurs, 16% des adaptations proposées n'ont pas été vues par les utilisateurs au moment où elles étaient proposées. En revanche, ils ont automatisé beaucoup d'adaptations proposées, et ce, certainement pour éviter d'être très régulièrement sollicités par le système.

Les propositions d'ajout de contextes clé ont été majoritairement acceptées par les participants. Seulement 19% de ces propositions ont été rejetées.

Les propositions de changements de contexte ont été refusées très majoritairement par les participants. Tous les testeurs ont déclaré que le système les proposait alors qu'ils étaient toujours dans le même contexte. A titre d'exemple, deux participantes ont été obligées d'ajouter un contexte clé chez elle car le système leur proposait d'en ajouter un nouveau alors qu'elles étaient toujours à leur domicile. Pour ne plus être perturbées par les messages, elles ont fini par ajouter un contexte "ordinateur" ou "maison2". Précisons que ces deux participantes ne disposaient pas de l'IHM de contrôle. La position géographique CELL-ID LAC est très sensible sur ce téléphone mobile ce qui explique ces dysfonctionnements constatés par les utilisateurs.

Les participants ont majoritairement accepté les réorganisations des fonctionnalités pour un contexte donné (46% versus 19%). Cependant, les participants n'avaient pas l'impression que leurs usages avaient évolué et ils se demandaient pourquoi le système proposait des réorganisations (verbatim issus des entretiens).

Le changement de mode de profil de sonnerie d'un contexte à un autre a été plébiscité par les utilisateurs. Ils ont jugé cette adaptation comme étant très utile car elle évite un certain nombre de manipulations sur le téléphone pour modifier le mode de sonnerie.

Perception utilisateur

Les entretiens individuels ont permis d'identifier les points forts et les points faibles de l'application, la pertinence ressentie des adaptations ou encore les difficultés rencontrées par les participants.

Avis général

Aucun participant n'a clairement rejeté l'application. Ils ont pour la plupart estimé que le concept était intéressant et pratique (cinq participants). En revanche, les propositions système sont gênantes car trop nombreuses et deux participants ont déclaré que ces propositions semblaient trop rapides (un jour après l'installation d'EMMA voire le jour même). Deux testeurs ont aussi jugé que cette application serait plus utile pour des personnes ayant des usages très variés ou pour des professionnels.

Concernant la réorganisation des menus, cinq participants ont un avis positif pour cette adaptation ("pratique", "bien", "plutôt intéressant"). En revanche une personne a été gênée de ne pas retrouver toujours la même organisation des menus. Trois ont estimé que les réorganisations étaient trop fréquentes et pas toujours justifiées et une personne ne se rendait pas compte des réorganisations car ses usages semblaient trop proches d'un contexte à un autre.

Le basculement de profils de sonnerie a été jugé positivement par six participants. Une seule a clairement indiqué qu'elle préfère le faire elle-même.

Messages systèmes

Comme nous l'avons développé un peu plus haut, tous les participants ont estimé que les sollicitations du système étaient trop nombreuses et certaines incompréhensibles. En effet, les participants ne comprenaient pas toujours pourquoi EMMA proposait de réorganiser les menus par exemple ou d'ajouter un contexte clé. Il semble important voire nécessaire de proposer des messages plus explicites sur les raisons des adaptations proposées par le système.

Disposer ou non de l'IHM de contrôle

Parmi les cinq participants qui utilisaient EMMA sans l'IHM de contrôle, quatre ont jugé qu'il était gênant de ne pas pouvoir contrôler les adaptations. Ils auraient souhaité pouvoir paramétrer certains contextes ou tout simplement les supprimer. Il semble donc nécessaire pour ce type d'application d'ajouter un menu dédié permettant de personnaliser et modifier les adaptations.

5.2.4 ENSEIGNEMENTS

Cette évaluation nous a permis de tirer un certain nombre d'enseignements qui peuvent donner un éclairage sur les bonnes pratiques à respecter au regard de l'évaluation terrain.

- Le recueil des retours utilisateurs doit être pris en compte au plus près du moment de l'adaptation. Au-delà de plusieurs semaines, les participants ont du mal à se souvenir précisément de certaines informations et de leurs impressions. Des méthodes de recueil en ligne sont à investiguer pour la suite des études (journal de bord, SMS de "sondage", etc.).
- Pour éviter toutes pertes de données objectives qui sont précieuses pour l'analyse des comportements du système et de l'utilisateur, il convient de les stocker sur le terminal (si elles ne sont pas trop volumineuses) et de les envoyer automatiquement par GPRS ou GSM à un serveur.
- Le moment de l'adaptation semble être réellement un point critique dans le ressenti et l'acceptation du système. Pour EMMA, les adaptations étaient proposées uniquement lors de la reprise en main du téléphone après une période d'inactivité. Il s'avère que cette limite n'est pas suffisante puisque

majoritairement tous les participants ont estimé que les sollicitations du système étaient trop nombreuses et gênantes. De plus, le clustering et les propositions d'adaptations ont peut-être été réalisés de manière trop systématique. Tester différentes périodes d'adaptation et la fréquence des apprentissages et clustering semble nécessaire pour tenter de déterminer les bons seuils.

- Pour déterminer les options nécessaires à intégrer dans l'IHM de contrôle et les usages de ces fonctionnalités, il convient de tracer précisément les options sélectionnées ou non et à quel moment.
- L'observabilité et la raison des propositions du système semblent aussi être des éléments très importants pour favoriser l'acceptabilité des systèmes adaptatifs. Durant cette étude, tous les utilisateurs se sont plaints de ne pas toujours comprendre pourquoi le système proposait telle modification du menu ou tel ajout de contexte. Il convient donc de travailler sur des modes de dialogue avec l'utilisateur lui permettant de comprendre les actions du système et les propositions faites avant de les accepter ou de les refuser.
- La réorganisation des menus semble être une adaptation pertinente pour certaines cibles d'utilisateurs. Néanmoins, comme nous avons pu le constater, les usages peuvent être très peu variés entre différents contextes clés, les premiers menus étant dans l'ensemble plutôt stables et liés aux fonctions de communication. Une approche consisterait à proposer un menu statique pour les quatre voire cinq premières fonctionnalités et un menu dynamique (adaptatif) pour les fonctionnalités secondaires. Ainsi, les utilisateurs retrouveraient une certaine stabilité des informations et des accès plus rapides pour des tâches réalisées de manière moins fréquente.

5.3 RECOMMANDATIONS ERGONOMIQUES

Les résultats des évaluations nous ont permis de formuler un ensemble de recommandations ergonomiques pour la conception de systèmes interactifs plastiques. Étant donné que ces recommandations sont le fruit des premières évaluations, elles doivent être considérées avec prudence. Elles mettent en évidence quatre critères du référentiel de [Bastien et Scapin 1993] qui méritent une attention particulière. Les quatre critères qui semblent incontournables sont les suivants : contrôle explicite, charge de travail, adaptabilité et guidage.

Ces critères identifient les conditions dans lesquelles la plasticité peut constituer un facteur facilitateur d'usage en téléphonie mobile. Parmi eux, le contrôle explicite est identifié comme un requis par l'ensemble des utilisateurs. Les recommandations énoncées s'articulent autour des deux leviers de la plasticité (remodelage et redistribution) et du contrôle utilisateur.

- Pour le remodelage, les recommandations sont les suivantes :
 - Réduire la navigation (charge de travail).
 - Déléguer au système les tâches répétitives (charge de travail).
 - Prendre en compte les préférences et les besoins de l'utilisateur (adaptabilité).
- Pour la redistribution, les recommandations sont les suivantes :
 - Privilégier la charge de travail à l'homogénéité/cohérence (charge de travail).
 - Améliorer le confort et la convivialité (adaptabilité).

- Réduire la charge de travail (charge de travail).
- Pour le contrôle utilisateur, les recommandations sont les suivantes :
 - Expliquer la raison d'être des propositions (guidage)
 - Limiter les propositions du système (charge de travail).
 - Placer l'adaptation sous le contrôle de l'utilisateur (contrôle explicite).

6. CONCLUSION ET PERSPECTIVES

L'année 2009 marque la première décennie des travaux en plasticité. Autour de ce sujet, traité aujourd'hui aux plans national et international, les avancées sont importantes. Cependant, l'évaluation reste un point faible où les résultats sont globalement pauvres.

Cet article tente d'apporter des réponses à des questions, jusqu'ici sans réponse, sur l'acceptabilité et l'utilisabilité des IHM adaptatives. Pour ce faire, trois démonstrateurs (PlasticSMS, MobileSMS et EMMA) ont été développés et évalués dans le but de recueillir le ressenti des utilisateurs vis-à-vis du remodelage et de la redistribution ainsi que de comprendre l'intérêt du contrôle utilisateur. Les méthodes d'évaluation utilisées impliquent, pour la plupart, l'utilisateur (focus group, tests utilisateur, entretiens). Les évaluations de PlasticSMS et MobileSMS se sont déroulées en laboratoire. EMMA qui nécessite la collecte de données sur des usages réels a été évaluée sur le terrain.

A notre connaissance, les résultats exposés dans cet article figurent parmi les premiers retours utilisateurs sur les IHM plastiques. Nous avons recueilli le ressenti des utilisateurs par rapport aux concepts qui ont fait l'intérêt de ces travaux. L'adaptation à la plate-forme par la redistribution sur une plate-forme plus confortable a été perçue comme la plus utile. Par ailleurs, nous avons observé des échecs uniquement au niveau du remodelage. En effet, il semble préférable d'afficher directement l'IHM, remodelée, dans sa globalité. Nous avons également appris qu'il est préférable de placer l'adaptation sous le contrôle de l'utilisateur. Les enseignements tirés des évaluations ont permis de formuler de premières recommandations ergonomiques sur la base du référentiel de [Bastien et Scapin 1993]. Les critères qui sont apparus comme pertinents au cours des évaluations sont : le contrôle explicite, la charge de travail, l'adaptabilité et le guidage. La charge de travail semble être le critère prépondérant. Cinq des neuf recommandations ergonomiques données dans cet article sont relatives à ce critère.

Toutefois, les résultats présentés dans cet article sont à manipuler avec précaution car les protocoles adoptés présentent des limites. Nous nous sommes inscrits dans une démarche qualitative. Les évaluations ont donc été menées auprès d'un panel réduit d'utilisateurs. Les retours du focus group sont uniquement basés sur le déclaratif. En laboratoire, il est difficile d'approcher un large éventail de contextes d'usage. Pour pallier aux difficultés de recrutement dans l'évaluation d'EMMA, nous avons prêté les téléphones aux participants. Avec ce nouvel équipement, les participants n'ont pas tout à fait adopté leurs pratiques usuelles. Nous avons constaté que, par exemple, ils ont très peu personnalisé le mobile alors qu'ils le font en règle générale. Après six semaines, les participants ont eu tendance à oublier certaines informations sur les propositions d'EMMA lors des premiers jours d'utilisation.

Pour le futur, nous pensons investiguer d'autres concepts clé de la plasticité. La valeur qui vise à aller au-delà de l'utilisabilité est une notion à laquelle nous souhaitons particulièrement nous intéresser. Nous souhaitons, tout d'abord, comprendre ce que représente la valeur d'un système adaptatif aux yeux des utilisateurs. Nous pensons mener cette étude par entretiens puis focus groups avec élaboration de maquettes

visuelles, démonstrateurs et/ou prototypes. Il s'agira ensuite, sur la base de la conception centrée valeur [Cockton 2006], de mettre en place un prototype stable qui sera évalué auprès d'utilisateurs en situations réelles.

REMERCIEMENTS

Nous remercions chaleureusement Dominique Scapin et Jean Vanderdonckt pour l'intérêt porté aux travaux. Nous associons à ces remerciements les relecteurs de cet article qui, par leurs remarques constructives, nous ont permis d'en améliorer l'écriture.

RÉFÉRENCES

- Amalberti, R., Valot, C. Le Magicien d'Oz. CERMA, journée du PRC Rhône-Alpes.1992.
- Bastien, J.M.C., Scapin D. Ergonomic Criteria for the Evaluation of Human-Computer Interfaces, Rapport technique INRIA, N°156, Juin 1993.
- Bellik, Y. Présentation Multimodale de l'Information. Habilitation à Diriger des Recherches. Université d'Orsay Paris-Sud, 2006, 90 pages.
- Benadi, S. Structuration des données et des services pour le télé-enseignement. Thèse de Doctorat de l'Institut National des Sciences Appliquées de Lyon, Septembre 2004.
- Benyon, D., Murray, D. Adaptive Systems : From Intelligent Tutoring to Autonomous Agents. In Knowledge-Based Systems, 197–219, 1993.
- Bobillier-Chaumon, M.E., Carvallo, S., Tarpin-Bernard, F., Vacherand-Revel, J. Adapter ou uniformiser les interactions personnes-systèmes? Revue d'Interaction Homme-Machine, *Europia*, Vol. 6, 2, 2005, 91-129.
- Blanchet, A., Ghiglione, R., Massonat, J., Trognon, A. Les techniques d'enquête en sciences sociales : observer, interviewer, questionner. Dunod, Paris, France, 2000.
- Brusilovsky, P. Methods and Techniques of Adaptive Hypermedia. In User Modeling and User-Adapted Interaction, 87–129, juillet 1996.
- Brusilovsky, P., Karagiannidis, C., Sampson, D. The Benefits of Layered Evaluation of Adaptive Applications and Services. In Proceedings of the Workshop on Empirical Evaluations of Adaptive Systems, 8th International Conference on User Modeling (UM'01), S. Weibelzahl, D. N., G. Weber, éditeurs, 1–8, juillet 2001.
- Brusilovsky, P., Karagiannidis, C., Sampson, D. Layered Evaluation of Adaptive Learning Systems. In International Journal of Continuing Engineering Education and Life-Long Learning, 14(4), 402–421, 2004.
- Calvary, G., Coutaz, J. Plasticité des Interfaces : une nécessité ! Information-Interaction-Intelligence, Actes des deuxièmes Assises nationales du GDR I3, Cépaduès Editions, J. Le Maître (Ed), Nancy, France, Décembre 2002, 247-261.
- Calvary, G. Plasticité des Interfaces Homme-Machine. Habilitation à Diriger des Recherches. Université Joseph Fourier, Novembre 2007, 97 pages.
- Carbonell, N., Simonin, J. Interfaces adaptatives – Modèles de l'utilisateur. Journée d'étude Interfaces adaptatives, Laboratoire Paragraphe, Université Paris 8, juin 2004.
- Chin, D. N. Empirical evaluation of user models and user-adapted systems. *User Modeling and User-Adapted Interaction* 11(1-2), 2001, 181–194.
- Cockton, G. From Quality in Use to Value in the World. In ACM Proc. CHI 2004, Late Breaking Results, 2004a, 1287-1290.
- Cockton, G. Value-centred HCI. Proceedings of the third Nordic conference on Human-computer interaction, 2004b, 149-160.
- Cockton, G. A development Framework for Value-Centred Design. In ACM Proc. CHI 2005, Late Breaking Results, 2005, 1292-1295.

- Cockton, G. Designing worth is worth designing. In Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles, 2006, 165-174.
- Dâassi, O. Les comets : une nouvelle génération d'Interacteurs pour la Plasticité des Interfaces Homme-Machine. Thèse de Doctorat de l'Université Joseph Fourier, Grenoble I, Janvier 2007.
- Demumieux, R., Ganneau, V., Calvary, G., Gegovska, E. Les interfaces plastiques premiers retours utilisateurs : évaluations en laboratoire. In Actes du colloque Ergo-IA 2008, L'humain au cœur des systèmes et de leur développement, 2008, 117-124.
- Diaper, D. Task Analysis for Human-Computer Interaction. Ellis Horwood, Chichester, UK, 1990.
- Dieterich, H., Malinowski, U., Kühme, T., Schneider-Hufschmidt, M. State of the Art in Adaptive User Interfaces. In Adaptive User Interfaces: Principles and Practices, Schneider-Hufschmidt & Al. (ed.), 1994, 13-48.
- Dittrich, K., Gatzui, S., Geppert, A. The Active Database Management System Manifesto: A Rulebase of ADBMS Features. In Proceedings of the Second International Workshop on Rules in Database Systems (RIDS'95), pages 3–20. Springer-Verlag, London, UK, 1995.
- Edmonds, E.A. Adaptive Man-Computer Interfaces. In Coombs, M. J. and Alty, J. L. (Eds), Computing Skills and the User Interface. London: Academic Press, 1981.
- Florins, M. Graceful Degradation, A Method for Designing Multiplatform Graphical User Interfaces, PhD Université Catholique de Louvain, Canada, Juillet 2006.
- Gabillon, Y., Calvary, G., Fiorino, H. Composing interactive systems by planning. In Proceedings of the 4th French-speaking conference on Mobility and ubiquity computing, ACM International Conference Proceeding Series, 2007, 37-40.
- Ganneau, V., Demumieux, R., Calvary, G. EMMA: modèle utilisateur pour la plasticité des interfaces homme-machine en mobilité. In Proceedings of the 4th French-speaking conference on Mobility and ubiquity computing, ACM International Conference Proceeding Series, 2008, 1-8.
- Gena, C. Methods and techniques for the evaluation of user-adaptive systems. In The Knowledge Engineering Review, Volume 20, 2005, 1-37.
- Gena, C. A User-Centered Approach for Adaptive Systems Evaluation. In Proceedings of the 5th Workshop on User-Centred Design and Evaluation of Adaptive Systems, 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'06), juin 2006, 430–439.
- Goodman, J., Brewster, S., Gray, P. Using Field Experiments to Evaluate Mobile Guides. In Proceedings of the 3rd Annual Workshop on HCI in Mobile Guides. 2004.
- Greenbaum, T. The handbook for Focus Group Research. Seconde édition, 1998.
- Grolaux, D., Vanderdonckt, J., Van Roy, P. FlexClock, a Plastic Clock Written in Oz with the QtK toolkit. In Proceedings of the First International Workshop on Task Models and Diagrams for User Interface Design, 2002, 135 – 142.
- Grolaux, D., Vanderdonckt, J., Van Roy, P. Attach Me, Detach Me, Assemble Me Like You Work, In proceedings of INTERACT 2005, M.F. Costabile and F. Paternò (Eds.), 2005, 98 – 212.
- Hesselman, C., Cesar, P., Vaishnavi, I., Boussard, M., Kernchen, R., Meissner, S., Spedalieri, A., Sinfreu, A., Räck, C. Delivering interactive multimedia services in dynamic pervasive computing environments. In Proceedings of the 1st international conference on Ambient media and systems, Ambi-sys, 2008.

- Holly, S.V., Jacko, J.A., Emery, V.K. Multimodal feedback: establishing a performance baseline for improved access by individuals with visual impairments. In Proceedings of the fifth international ACM conference on Assistive technologies, ACM SIGACCESS Conference on Assistive Technologies, 2002, 49-56.
- Höök, K. Evaluating the Utility and Usability of an Adaptive Hypermedia System. In Proceedings of the 2nd International Conference on Intelligent User Interfaces, IUI 97.
- Höök, K. Steps to take before intelligent user interfaces become real. *Interacting with Computers* (2000) 12(4), 409-426.
- Huebsch, G., Kadner, K. Multimodal information access across multiple devices. In Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology, International Conference On Mobile Technology, Applications, And Systems, 2007, 736-742.
- Hulkko, S., Mattelmäki, T., Virtanen, K., Keinonen, T. Mobile probes. In Proceedings of the third Nordic conference on Human-computer interaction, 43 – 51, 2004.
- Jameson, A. Adaptive Interfaces and Agents. *Human-Computer Interface Handbook*, J.A. Jacko & A. Sears (Eds), Chapter 15, 2003, 305-330.
- Jambon, F., Golanski, C., Pommier, P.J. Évaluation des dispositifs mobiles : sur le terrain ou en laboratoire ? In Actes de la 18^{ème} Conférence Francophone sur l'Interaction Homme-Machine (IHM'06), ACM, New York, NY, USA, 35-42, 2006.
- Kellar, M., Reilly, D., Hawkey, K., Rodgers, M., MacKay, B., Dearman, D., Ha, V., MacInnes, J., Nunes, M., Parker, K., Whalen, T., Inkpen, K. It's a Jungle Out There : Practical Considerations for Evaluation in the City. In CHI'05 extended abstracts on Human Factors in Computing Systems (CHI'05), ACM, New York, NY, USA, 1533–1536, avril 2005.
- Kjeldskov, J., Paay, J. Just-For-Us : A Context-Aware Mobile Information System Facilitating Sociality. In Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services (MOBILEHCI'05), ACM, New York, NY, USA, 23-30, 2005.
- Lemmelä, S., Vetek, A., Mäkelä, K., Trendafilov, D. Designing and Evaluating Multimodal Interaction for Mobile Contexts. In Proceedings of the 10th international conference on Multimodal Interfaces, ACM International Conference on Multimodal Interfaces, 2008, 265-272.
- Lewis, C., Polson, P.G., Wharton, C., Rieman, J. Testing a Walkthrough Methodology for Theory-Based Design of Walk-Up-and-Use Interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'90), ACM, New York, NY, USA, 235–242, 1990.
- Lieberman, H., Faaborg, A., Espinosa, J., Stocky, T. Common Sense on the Go: Giving Mobile Applications an Understanding of Everyday Life. In *British Telecom Technology Journal*, Vol. 22, No. 4, October 2004
- Mariage, C. Espace de classification des méthodes et outils d'évaluation de l'ergonomie des systèmes à technologie Web. Mémoire en vue de l'obtention du Diplôme Interuniversitaire d'Etudes Approfondies en Informatique (DEA). Université Catholique de Louvain, 2001, 59 pages.
- Nielsen, J., Molich, R. Heuristic Evaluation of User Interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'90), ACM, New York, NY, USA, 1990, 249–256.

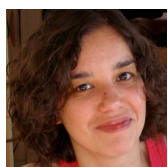
- Nielsen, J. Usability Engineering. Morgan Kaufmann, San Francisco, CA, USA, 1993.
- Evaluation des EIAH : une nécessaire diversité des méthodes. In proceedings of TICE 2004, Compiègne, France, 2004, 265-271.
- Norman, D.A. Affordance, Conventions, and Design. In *Interactions*, 38–43, 1999.
- Polson, P.G., Lewis, C., Rieman, J., Wharton, C. Cognitive Walkthroughs : A Method for Theory-Based Evaluation of User Interfaces. In *International Journal of Man-Machine Studies*, 741–773, mai 1992.
- Preece, J., Rogers, I., Sharp, H., Benyon, D., Holland, S., Carey, T. Human-Computer Interaction. Addison-Wesley, Essex, UK, 1994.
- Senach, B. L'évaluation ergonomique des interfaces homme-machine : une revue de la littérature. In *L'ergonomie dans la conception des projets informatiques*, J.-C. Sperandio, éditeur. Octares Éditions, 69–122, 1993.
- Sohn, T., Li, K.A., Griswold, W.G., Hollan, J.D. A diary study of mobile information needs. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, 433 – 442, 2008.
- Tarpin-Bernard, F. Interaction Homme-Machine adaptative. Habilitation à Diriger des Recherches. Institut National des Sciences Appliquées de Lyon, Université Claude Bernard Lyon I, Décembre 2006, 99 pages.
- Thevenin, D. and Coutaz, J. Plasticity of User Interfaces: Framework and Research Agenda. In: Edinburg, A.S., Johnson, C. (eds.) *Proc. Interact 1999*, IFIP IOS Press Publ., Amsterdam, 110–117.
- Thevenin, D. Adaptation en Interaction Homme-Machine : le cas de la Plasticité. Thèse de l'Université Joseph Fourier, Grenoble I, 2001, 234 pages.
- Thompson, D. Concise Oxford English Dictionary, 1996.
- Tobar, C. Yet Another Evaluation Framework. In *Proceedings of the 2nd Workshop on Empirical Evaluation of Adaptive Systems*, 9th International Conference on User Modeling (UM'03), S. Weibelzahl, A. Paramythis, éditeurs, 15–24, juin 2003.
- Totterdell, P., Boyle, E. The Evaluation of Adaptive Systems. In *Adaptive User Interfaces*, D. Browne, P. Totterdell, M. Norman, éditeurs. Academic Press, 161–194, 1990.
- Totterdell, P., Rautenbach, P. Adaptation as a Problem Design. In *Adaptive User Interfaces*, D. Browne, P. Totterdell, M. Norman, éditeurs, Academic Press, 59-84, 1990.
- Vanderdonckt, J., Grolaux, D., Van Roy, P., Limbourg, Q., Macq, B., Michel, 93 B., A Design Space for Context-Sensitive User Interfaces, *Proc. of ISCA 14th Int. Conf. on Intelligent and Adaptive Systems and Software Engineering IASSE'2005*.



Fatoumata Camara est actuellement étudiante en deuxième année de doctorat. Son travail se déroule dans les Orange Labs, centre de Lannion et dans le Laboratoire d'Informatique de Grenoble (LIG), équipe IIHM (Ingénierie de l'Interaction Homme-Machine). Sa thèse s'inscrit dans le domaine de l'Informatique, plus précisément, en Interaction Homme-Machine (IHM). Elle s'intéresse au problème de l'adaptation, selon deux axes : Information et Interaction. Il s'agit d'étudier, conjointement, ces deux formes d'adaptation avec un barycentre porté sur la qualité de vie.



Gaëlle Calvary est Professeur en Informatique à Grenoble INP. Elle effectue ses recherches au Laboratoire d'Informatique de Grenoble. Son thème de recherche est la Plasticité des Interfaces Homme-Machine (IHM), c'est-à-dire l'étude de la capacité d'adaptation des IHM à leur contexte d'usage dans le respect de la valeur attendue par l'utilisateur cible. L'approche qu'elle a le plus explorée est l'Ingénierie Dirigée par les Modèles. Elle étudie la conservation et l'exploitation à l'exécution des modèles de conception pour la plasticité des IHM et, de façon, plus générale, leur qualité. Elle est membre du projet européen ITEA UsiXML et de l'ANR MyCitizSpace. Elle est membre du comité directeur du GDR I3 et co-fondatrice de l'atelier CESAME (Conception et Exécution de Systèmes interactifs doués d'Adaptation dans un monde Mixte en Evolution).



Rachel Demumieux est ergonomiste senior à Orange Labs (Direction Recherche et Développement de France Télécom). Elle travaille dans le domaine de la conception et l'évaluation d'interfaces de dispositifs utilisés en mobilité. Ses travaux de recherche portent sur les problématiques de collecte de données d'usage en mobilité et sur l'adaptation des interfaces en fonction des profils utilisateurs et des contextes d'utilisation. Elle a co-encadré la thèse de Vincent Ganneau et actuellement celle de Fatoumata Camara traitant de l'adaptation des interfaces et des modes d'interaction selon les contextes d'usage.



Vincent Ganneau est Ingénieur de Recherche et Développement chez Haploid, une start-up spécialisée dans la conception et la réalisation d'applications pour téléphones mobiles. Il est titulaire d'un Doctorat en Informatique sur l'Ingénierie de l'Interaction Homme-Machine. Sa thèse traite d'un Modèle Utilisateur pour la Plasticité des Interfaces Homme-Machine (IHM) en Mobilité. Ses travaux examinent l'adaptation des IHM du point de vue de l'usage. Ils explorent les Réseaux Bayésiens pour modéliser et exploiter les données collectées en mobilité afin d'identifier les contextes d'usage clés et proposer la meilleure adaptation.



Vérification et validation formelles de systèmes interactifs fondées sur la preuve : application aux systèmes Multi-Modaux

Yamine Aït-Ameur, Idir Aït-Sadoune, Mickael Baron, Jean-Marc Mota

► To cite this version:

Yamine Aït-Ameur, Idir Aït-Sadoune, Mickael Baron, Jean-Marc Mota. Vérification et validation formelles de systèmes interactifs fondées sur la preuve : application aux systèmes Multi-Modaux. Journal d'Interaction Personne-Système, Association Francophone d'Interaction Homme-Machine (AFIHM), 2010, 1 (1), pp.1-30. hal-00634186

HAL Id: hal-00634186

<https://hal.archives-ouvertes.fr/hal-00634186>

Submitted on 28 Aug 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vérification et validation formelles de systèmes interactifs fondées sur la preuve : application aux systèmes multi-modaux

YAMINE AIT-AMEUR, IDIR AIT-SADOUNE et MICKAEL BARON

(*) Laboratoire LISI-ENSMA, Poitiers

JEAN-MARC MOTA

(**) Thales Rail Signalling Solutions, Vélizy

Résumé : Cet article s'intéresse à la validation et à la vérification formelles d'IHM Multi-Modales (IHM3). Il décrit une partie des résultats obtenus dans le cadre du projet RNRT VERBATIM, dont l'objet est la VERification Biformelle et Automatisation du Test des Interfaces Multimodales. Ce projet s'intéresse, entre autres, à la mise en œuvre d'une technique formelle : la méthode B événementiel. Cette technique fondée sur la preuve ne souffre donc pas du problème d'explosion du nombre d'états rencontré généralement par les techniques de vérification sur modèles. Nous discutons les apports de cette technique pour la conception d'IHM3, en particulier, sa capacité à exprimer et à vérifier des propriétés de la famille CARE. Notre approche utilise des notations et techniques semi-formelles issues du domaine des IHM et propose de les formaliser. Enfin, nous appliquons notre approche sur une étude de cas appelé "Pages Jaunes CLIPS".

Mots clés : IHM multi-modales, technique formelle fondée sur la preuve, vérification, validation, propriétés CARE, modèle de tâches.

Abstract: This paper focuses on the formal validation and verification of multi-modal human computer interfaces. It describes part of the obtained results of the French RNRT VERBATIM project whose purpose is the Multimodal Interfaces Biformal Verification and Test Automation. This project focuses on the application of a formal technique, namely the event B method. This approach is based on a proof technique and therefore it does not suffer from the state number explosion problem occurring in classical model checking. We outline the capability of this technique to support the design of multi-modal human computer interfaces, in particular, the capability to support the expression and the verification of properties issued from the CARE family. The proposed approach uses notations and semi-formal techniques issued from the HCI design area. We apply our approach on a case study called "CLIPS Yellow Pages".

Key words: Multi-modal HCI, proof based technique, verification, validation, CARE properties, task model.

(*) Laboratoire LISI (EA 1232), ENSMA, Téléport 2, 1 avenue Clément Ader, BP 40109, 86961 Futuroscope Cedex.

(**) Thales Rail Signalling Solutions, 20-22 rue grange Dame rose 78140 Vélizy.

1. INTRODUCTION

La diversité des interfaces homme-machine (IHM) ainsi que les progrès réalisés dans la définition de nouveaux dispositifs d'interaction ont conduit à une complexité dans la conception et la mise en œuvre des IHM. Le recours à des modèles de conception et à des notations de description des IHM devient indispensable pour maîtriser cette complexité. Dans le but d'améliorer la flexibilité et l'utilisabilité des IHM, de nombreux travaux ont proposé des techniques, des notations et des méthodes pour les différentes étapes de développement d'une IHM. Toutefois leur application aux IHM multi-modales (IHM3) reste toujours un sujet de préoccupation pour les chercheurs.

L'étude de l'utilisabilité des IHM multi-modales implique la prise en considération de nouvelles notions absentes dans les IHM classiques. Parmi les propriétés propres aux IHM multi-modales nous trouvons les propriétés CARE (Complémentarité, Assignment, Redondance et Équivalence [Coutaz et al. 1995]) que le système doit assurer [Nigay et al. 1995].

Par ailleurs, il est bien établi que les méthodes formelles participent à l'augmentation de la qualité des développements de systèmes interactifs et d'interfaces homme-machine [Hix and Hartson 1993]. De nombreux travaux mettant en œuvre ce type de technique ont été menés dans ce domaine. Ils se fondent soit sur les techniques formelles orientées modèles (model checking) ou bien sur la preuve (theorem proving).

Nous proposons d'utiliser une technique formelle fondée sur la preuve (la méthode B événementiel [Abrial 2010]) afin de prendre en compte les propriétés CARE dès les premières étapes de conception, à un niveau de développement qui fait abstraction aussi bien des détails d'implantation que des éléments du noyau fonctionnel. Plus précisément, nous nous intéressons, dans cet article, à la mise en œuvre des techniques formelles fondées sur la preuve pour la conception, la validation et la vérification formelles de propriétés d'utilisabilité, en l'occurrence les propriétés de la famille CARE, des IHM3. Cet article étend [Ait-Ameur et al. 2006] et décrit une partie des résultats obtenus dans le cadre du projet RNRT VERBATIM¹ s'intéressant à la validation formelle des propriétés liées à la multimodalité en télécommunications mobiles.

Cet article est structuré de la façon suivante. La section qui suit décrit les modèles et notations utilisées dans le processus de spécification et de conception de systèmes interactifs. Nous détaillons également les propriétés qui doivent être établies et vérifiées pour les systèmes interactifs classiques et multi-modaux. Une deuxième section effectue un état de l'art sur l'utilisation des techniques formelles dans le domaine de l'interaction homme-machine. Nous présentons rapidement nos travaux antérieurs qui ont consisté à l'utilisation de la méthode B dans sa version classique et justifions le passage à la version B événementiel que nous décrivons dans la section suivante. Puis nous présentons successivement notre approche pour la modélisation de tâches utilisateurs et de conception de l'architecture logicielle nous permettant ainsi de modéliser le contrôleur de dialogue. Dans une dernière section, nous présentons l'étude de cas des pages jaunes développée au CLIPS-IMAG et mise en œuvre pour illustrer notre approche. Cette section montre comment les propriétés de la famille CARE sont validées sur une conception B événementiel. Les interactions multi-modales sont introduites progressivement grâce au raffinement en

¹Réseau National de Recherche en Télécommunications. VERification Biformelle et Automatisation du Test des Interfaces homme machine Multimodales.

Site du projet <http://iihm.imag.fr/nigay/VERBATIM/>

codant un modèle de tâches utilisateurs. Enfin, nous terminons par un bilan de ce travail ainsi que quelques perspectives.

2. CONCEPTION DES INTERFACES HOMME-MACHINE

De manière générale, deux scénarii de développement différents aboutissant à des processus de développement différents peuvent être définis. Soit (1) étant donné un ensemble de propriétés, la spécification et la conception sont établies conformément à ces propriétés, soit (2) étant donnée une spécification, les propriétés doivent être vérifiées en adéquation avec la spécification établie. Dans les deux cas, la validation de propriétés est une activité fondamentale. Les propriétés doivent être validées par n'importe quelle spécification de la modélisation traduisant les exigences du cahier des charges.

2.1 Spécification / Conception

Comme pour les développements de programmes classiques, le domaine de l'IHM a vu un grand nombre de notations et de techniques pour la conception et la description. Par ailleurs, la finalité des différentes compétences comme le génie logiciel, l'ergonomie et la psychologie a fait que ces notations sont hétérogènes. Le besoin d'intégration est un problème majeur. La plupart de ces techniques restent semi-formelles et l'aspect graphique est considéré comme une caractéristique importante puisque qu'elle représente facilement la notation dans son ensemble. Nous avons choisi de diviser ces techniques et notations en deux parties :

- Les techniques et les notations orientées description permettent de décrire les besoins utilisateurs en terme d'utilisabilité d'une interface utilisateur. Elles sont souvent proposées par des non-informaticiens (ergonomie et psychologie par exemple). La majorité de ces notations sont centrées utilisateur et sont donc loin des implantations informatiques. Les besoins utilisateur sont exprimés par un ensemble de tâches qui permettent d'atteindre un but donné à partir d'un état initial. En voici quelques exemples : MAD [Scapin and Pierret-Golbreich 1990; Scapin and Pierret-Golbreich 1989], XUAN [Gray et al. 1994], CTT [Paternò 2001; Paternò et al. 2001] et K-MAD [Lucquiaud 2005; Baron et al. 2006].
- Les techniques et les notations orientées conception permettent d'exprimer la structure statique du logiciel implémentant l'interface utilisateur décrite et/ou du système. Ces techniques et notations séparent l'interface utilisateur du noyau fonctionnel. En voici quelques exemples : le modèle Seeheim [Pfaff 1985], le modèle PAC [Coutaz 1987], le modèle ARCH [Bass et al. 1991], le modèle MVC [Burbeck 1992] et les modèles hybrides [Guittet 1995; Depaulis et al. 2006]. Voir aussi [Kolski et al. 2009] pour une synthèse récente sur les modèles d'architecture à base d'agents.

2.2 Propriétés

Les propriétés dans les systèmes interactifs doivent non seulement vérifier les propriétés propres aux systèmes informatiques classiques (sûreté, équité) mais aussi doivent prendre en compte l'utilisateur (psychologie, comportement, aspects ergonomiques, etc.) ainsi que l'environnement (système d'exploitation, systèmes d'entrée/sortie, etc).

De nombreuses formulations et classifications des propriétés inhérentes aux IHM ont été proposées dans [Harrison and Thimbleby 1990; Dix et al. 1993; Duke and Harrison 1995; Roche 1998].

Nous pouvons les résumer suivant deux grandes classes de propriétés :

- Les propriétés de validité qui caractérisent un fonctionnement attendu ou voulu par un utilisateur. Dans cette catégorie de propriétés, nous pouvons distinguer la complétude (l'utilisateur doit atteindre un objectif donné de différentes manières), la flexibilité pour la représentation de l'information, le déroulement de tâches (atteignabilité, non-préemption, interaction à plusieurs fils), les propriétés liées au temps, les aspects psychologiques et ergonomiques.
- Les propriétés de robustesse qui sont relatives à la sûreté de fonctionnement de l'interface homme-machine et au système en général. Parmi ces propriétés, nous trouvons des propriétés liées à la visualisation comme l'observabilité, l'insistance et l'honnêteté.

À ces propriétés s'ajoutent celles qui sont relatives à un type précis d'IHM et en l'occurrence dans cet article nous traitons les propriétés inhérentes aux systèmes multi-modaux. De nombreuses études ont permis de classifier les actes multi-modaux [Bellik 1995] et nous retiendrons les propriétés CARE (Complémentarité, Assignation, Redondance et Équivalence [Coutaz et al. 1995]). Ces propriétés définissent les propriétés de robustesse et de flexibilité des systèmes mutli-modaux.

3. LES TECHNIQUES FORMELLES EN IHM

Parmi les techniques formelles les plus largement utilisées en génie logiciel, les approches orientées modèles jouent un rôle majeur dans le domaine de l'interaction homme-machine. Ces méthodes sont basées sur la description de modèles au moyen d'un ensemble de variables qui sont modifiées par les opérations et les événements des modèles. Généralement, ces techniques sont divisées en deux catégories : la preuve automatique par vérification sur modèles et les systèmes fondées sur la preuve. Les deux techniques ont été appliquées au domaine de l'IHM.

La première catégorie est basée sur l'évaluation de propriétés logiques d'un système à transitions obtenu à partir des valeurs des variables. Parmi ces techniques, nous citons les logiques temporelles et les réseaux de Petri. Dans le domaine de l'IHM, ces techniques ont été utilisées la première fois pour la vérification formelle de propriétés de systèmes interactifs [Campos and Harrison 1997]. Par exemple, [Abowd et al. 1995] vérifie des interfaces utilisateur avec SMV (Symbolic Model Verifier) utilisant CTL (Computational Tree Logic), tandis que [Faconti et al. 1992] utilise LOTOS pour écrire des spécifications d'interacteur. [Brun 1997] développe un formalisme basé sur une logique temporelle, nommé XTL (eXtended Temporal Logic). La vérification sur modèles (model checking) est également utilisée par [Palanque et al. 1995; Navarre et al. 2005] où l'utilisateur et le système sont modélisés au moyen de réseaux de Petri orientés objets appelés ICO. Dans la même catégorie, [D'Ausbourg 1998] a utilisé un langage à flots de données pour la validation automatique de systèmes interactifs.

La seconde catégorie est basée sur des systèmes de preuve où le modèle est décrit par des variables, des opérations, des événements, des propriétés temporelles et des invariants. Les opérations doivent préserver ces invariants et une famille de propriétés (sûreté, non-blocage, vivacité, etc.). Pour assurer la conformité de ces spécifications, des obligations de preuve sont générées et doivent être prouvées. Selon l'implémentation qui en est faite, le système de preuve peut prouver automatiquement les obligations de preuve. Parmi ces techniques, nous trouvons Z, basée sur la théorie des ensembles [Spivey 1992], VDM

[Bjorner 1987], basée sur le calcul de pré/post-conditions [Hoare 1969], et B, basée sur le calcul de la plus faible pré-condition [Abrial 1996b; Abrial 1996a]. Dans le domaine de l'IHM, VDM et Z ont été utilisées pour la définition de structures atomiques comme les interacteurs [Duke and Harrison 1993a; Duke and Harrison 1993b], et Z et Object-Z sont utilisées maintenant plus intensément, enfin HOL (a Higher Order Logic Theorem Prover) a été utilisée pour la vérification de spécifications d'interfaces utilisateur [Bumbulis et al. 1996].

4. CONCEPTION À BASE DE MODULES

La première approche que nous avons développée exploite la méthode B [Abrial 1996b]. Cette démarche est fondée sur une approche de conception ascendante. Elle a permis la spécification, la vérification et le raffinement de spécifications formelles B. De plus, des programmes écrits en ADA et TCL-TK ont été produits au moyen de cette technique, démontrant ainsi que notre approche peut être utilisée à toutes les étapes du développement de systèmes interactifs [Aït-Ameur et al. 1998; Jambon et al. 2001; Aït-Ameur et al. 2003; Aït-Ameur et al. 2004].

Cette approche est appelée composition à base de modules puisque la décomposition des applications interactives étudiées a été représentée en B suivant le modèle d'architecture imposé par ARCH. Elle est ainsi entièrement basée sur une structuration modulaire des machines abstraites pour composer des sous-systèmes afin d'aboutir au système complexe.

Par ailleurs, cette approche a permis d'apporter une solution pour construire et valider des tâches utilisateurs. La validation de tâches s'appuie sur une démarche comparable aux traces explicites ou aux diagrammes de séquence UML dans le sens où des traces d'opérations sont explicitement construites et validées. L'objectif est d'obtenir une trace d'opérations atomiques (du contrôleur de dialogue) par décomposition en sous tâches où le seul opérateur de contrôle est la séquence. Si ce développement est valide (toutes les obligations de preuve générées ont été prouvées), alors la tâche est valide. Elle permet donc la validation a priori de tâches (faisabilité), c'est-à-dire qu'il existe une séquence d'opérations qui implémente la tâche de plus haut niveau. Elle permet aussi la validation a priori de la conception (complétude), c'est-à-dire que toutes les obligations de preuve de la modélisation de la conception sont prouvées. Cette approche ne nécessite pas de modification de la partie conception. Il y a donc homogénéité des langages de modélisation entre la partie conception et validation.

Toutefois, deux critiques de l'approche à base de modules peuvent être formulées :

- (1) la première concerne la phase de conception. L'approche à base de modules ne permet pas de décrire l'état du dialogue d'une application interactive. En effet, elle ne définit qu'un ensemble d'opérations mais pas les événements qui permettent de les déclencher ni leur ordre. La description de systèmes concurrents n'est donc pas possible. Cette lacune est due principalement à la sémantique de B classique qui ne permet pas de décrire des systèmes réactifs comme les IHM. Par ailleurs, la conception ascendante oblige à reprouver à chaque composition des obligations de preuve plus complexes ;
- (2) la seconde concerne la phase de validation de tâches. Nous avons montré dans nos travaux que l'établissement de scénarii se fondait sur l'utilisation exclusive de l'opérateur de séquence. Par conséquent l'établissement des traces d'opérations à valider est fastidieux à mettre en place. En effet, en comparaison avec les formalismes possédant une forte capacité à structurer les tâches utilisateur, comme CTT par exem-

ple, notre approche de validation à base de traces doit énumérer tous les cas possible par un seul opérateur alors que l'utilisation de différents opérateurs (interruption, dés-activation, entrelacement et itérations) rendrait la validation de tâches plus expressive et complète.

L'approche que nous proposons dans cet article propose une réponse aux deux critiques formulées ci-dessus au moyen de la mise en œuvre de B événementiel.

5. LA MÉTHODE B ÉVÉNEMENTIEL

Nous utilisons la méthode formelle B événementiel [Abrial 2010] pour montrer qu'il est possible de traiter un développement formel complet d'un système interactif. À la différence des autres techniques formelles comme Z [Spivey 1992], la méthode B événementiel est outillée puisqu'il existe des outils comme l'Atelier B² et la plateforme Rodin³. L'Atelier B a été utilisé dans le cadre de ces travaux. Il permet de générer automatiquement toutes les obligations de preuve et est doté à la fois d'un prouveur automatique et interactif.

5.1 Modèle

Un modèle B événementiel décrit un système de transitions étiquetées (STE) par un ensemble d'états, un ensemble d'actions, un état initial et une relation de transition. Les actions effectuées par le système permettent la transition entre états. Le STE n'est pas manipulé explicitement. Il n'est décrit que via des événements dont le développeur donne la condition de déclenchement (garde de l'événement) et l'action instantanée associée (corps de l'événement). Un modèle est dit clos au sens où il décrit non seulement le système logiciel d'intérêt mais aussi son environnement d'exécution (ou une partie). La sémantique associée est une sémantique à base de traces d'événements. Un modèle est caractérisé par l'ensemble des séquences (traces) d'événements autorisés sous couvert des propriétés énoncées.

Un modèle est constitué de deux parties : une partie dite *statique* correspondant à la définition de l'état et de ses propriétés invariantes ; une partie dite *dynamique* décrivant les transitions.

En résumé, un modèle est composé des clauses suivantes (cf. figure 1) :

- MODEL correspondant à la déclaration du nom du modèle ;
- SETS introduisant des ensembles abstraits ou énumérés d'un modèle ;
- VARIABLES énonçant une liste de noms permettant la déclaration des variables (état) ;
- INVARIANT correspondant à la définition des propriétés des variables : il s'agit de propriétés logiques qui expriment en logique du premier ordre les propriétés invariantes d'un modèle ;
- ASSERTIONS permettant de définir des lemmes : il s'agit d'une liste de prédicats, portant sur les variables, déductibles de l'invariant. Ils correspondent à des lemmes utiles pour la preuve ou pour la compréhension du modèle ;
- INITIALISATION permettant de spécifier, en utilisant des substitutions, l'initialisation des variables (état initial) ;

²L'Atelier B est un outil qui permet une utilisation opérationnelle de la méthode B : <http://www.atelierb.eu>

³La plateforme Rodin est un environnement de développement basé sur Eclipse pour la méthode B événementiel : <http://www.event-b.org/platform.html>

<p>MODEL <i>m</i> SETS <i>s</i> VARIABLES <i>v</i> INVARIANT <i>I</i> ASSERTIONS <i>A</i> INITIALISATION <i>i</i> EVENTS <i>Spécification des événements du modèle</i> END</p>	<p>REFINEMENT <i>r</i> REFINES <i>m</i> SETS <i>t</i> VARIABLES <i>w</i> INVARIANT <i>J</i> VARIANT <i>V</i> ASSERTIONS <i>B</i> INITIALISATION <i>i'</i> EVENTS <i>Spécification des événements du raffinement</i> END</p>
(a)	(b)

Fig. 1. (a) Forme générale d'un modèle, (b) forme générale d'un raffinement

- **EVENTS** regroupant une liste d'éléments proches des actions de Back [Back and Sere 1989; Back et al. 2003] et des commandes gardées de Dijkstra [Dijkstra 1976] nommées événements (transitions entre états).

5.2 Les substitutions généralisées

L'initialisation ainsi que les événements sont décrits grâce aux substitutions généralisées. Ces dernières sont fondées sur le principe de plus faible précondition de Dijkstra [Dijkstra 1976]. Étant donné une substitution S et un prédicat P alors $S[P]$ représente la plus faible précondition qui satisfait P après l'exécution de S . Les substitutions intervenant dans les modèles B événementiel sont définies par les expressions suivantes.

$$[\text{SKIP}] P \Leftrightarrow P \quad [1]$$

$$[S_1 \parallel S_2] P \Leftrightarrow [S_1] P \wedge [S_2] P \quad [2]$$

$$[x := E] P \Leftrightarrow P(x/E) \quad [3]$$

$$[x \in T] P \Leftrightarrow \forall x'. (x' \in T \Rightarrow P(x/x')) \quad [4]$$

$$[\text{SELECT } G \text{ THEN } S \text{ END}] P \Leftrightarrow G \Rightarrow [S] P \quad [5]$$

$P(x/E)$ représente le prédicat P où toutes les occurrences libres de x sont remplacées par l'expression E .

Les substitutions [1], [2], [3] et [4] représentent respectivement l'événement nul, la substitution parallèle exprimant que les substitutions S_1 et S_2 sont réalisées en parallèle, l'affectation, l'affectation d'un élément quelconque appartenant à l'ensemble T . La substitution [5] est une substitution gardée où S est réalisée sous couvert de la garde G .

5.3 Événements

L'événement correspond à un changement d'état dénotant une transition dans le système modélisé. Il est composé de trois parties (cf. figure 2) : un nom ; une condition appelée

garde notée G ; une action (corps de l'événement) exprimée à l'aide d'une substitution généralisée qui dénote un changement d'état instantané notée S .

```

nom_événement =
SELECT  $G$ 
THEN
   $S$ 
END

```

Fig. 2. La forme d'un événement

Notons que l'action d'un événement n'est déclenchée que si sa garde est satisfaite. Une fois qu'un événement s'est déclenché, il peut, si sa garde est encore satisfaite, se déclencher à nouveau. Lorsque les gardes de plusieurs événements sont satisfaites simultanément, seul l'un d'entre eux se déclenchera.

```

événement_simple =
BEGIN
   $S$ 
END

événement_gardé =
SELECT  $G$ 
THEN
   $S$ 
END

événement_non_déterministe =
ANY  $l$ 
WHERE  $G$ 
THEN
   $S$ 
END

```

Fig. 3. Les 3 formes d'événements

Un événement peut se présenter sous trois formes différentes (cf. figure 3)

- (1) un événement est dit *simple* lorsque celui-ci n'est pas gardé (ou autrement dit sa garde est la proposition *vraie*). Il peut toujours se déclencher. En cas de déclenchement, il modifie l'état du système selon la substitution S ;
- (2) un événement est dit *gardé* lorsqu'il doit satisfaire une condition G , appelée garde, définie uniquement à partir de variables d'état, pour se déclencher ;
- (3) un événement est dit *non déterministe* lorsque celui-ci se déclenche s'il existe des valeurs pour les variables l satisfaisant la condition G . La garde de cet événement est donc de la forme $\exists(l).(G)$ où l est une liste de variables locales distinctes. L'action de l'événement est modélisée par la substitution généralisée S qui peut utiliser les variables l .

Type	Obligation de preuve
(1) Initialisation	$[i]I$
(2) Événement gardé	$I \wedge G \Rightarrow [S]I$
(3) Assertion	$I \Rightarrow A$

Table I. Obligations de preuve pour la consistance d'un modèle

Pour s'assurer qu'un modèle est consistant (voir les obligations de preuve du tableau I), il faut montrer, d'une part, que l'initialisation i établit l'invariant I (1) et d'autre part,

que les modifications engendrées sur l'état par les substitutions S de chaque événement maintiennent l'invariant sous la garde G de l'évènement (2). Si le modèle comporte des assertions, il faudra établir que chaque assertion A est une conséquence de l'invariant I (3).

5.4 Un exemple de modèle B événementiel

Considérons la spécification d'une horloge présentée dans [Cansell 2003] de la figure 4. La spécification abstraite *Clock* décrit une variable h pour les heures de l'horloge. Deux événements sont décrits. Le premier (événement *incr*) permet d'augmenter la variable h et se déclenche tant que $h \neq 23$. Le second événement est *zero*. Il est déclenché quand $h = 23$ pour réinitialiser la variable h .

5.5 Le raffinement

Un modèle B événementiel peut être raffiné et enrichi par de nouveaux événements et de nouvelles propriétés [Abrial and Hallerstede 2007]. Le processus de raffinement conduit le développeur à la conception finale de l'interface utilisateur, réalisée à la suite de plusieurs étapes de raffinement qui fournissent différents niveaux d'abstraction.

Un raffinement, comme le montre la figure 1, est défini par un nom de modèle déclaré dans la clause REFINEMENT et par sa filiation grâce à la clause REFINES qui indique le nom du modèle raffiné : r est appelé modèle concret et m modèle abstrait. En tant que modèle à part entière, il comporte une partie statique ainsi qu'une partie dynamique. Il est constitué donc des clauses présentées précédemment en section 5.1. A celles vues précédemment s'ajoute la clause VARIANT qui permet de définir les variants décrémentés par les événements : nous aurons l'occasion d'y revenir plus précisément en section 5.5.2. Le raffinement en tant que modèle doit prouver l'établissement et la conservation des propriétés énoncées comme invariantes. Mais il doit aussi assurer que tous les comportements du système concret sont des comportements du système abstrait.

5.5.1 Le raffinement d'évènement. Chaque événement d'un modèle abstrait est raffiné en un événement plus concret. On retrouve au sein d'un raffinement tous les événements présents dans l'abstraction. Par contre ces mêmes événements peuvent subir des modifications que ce soit au niveau de la garde et/ou au niveau de l'action, pour réduire le non-déterminisme et accroître la précision de la description. Un événement abstrait est dit raffiné par un événement concret lorsque la garde concrète est contrainte plus fortement que la garde abstraite ou lorsque l'invariant concret est préservé par l'action conjointe des deux événements. Étant donné deux événements e et e' , ayant respectivement les gardes G , G' et les substitutions S et S' , l'obligation de preuve assurant que e' raffine correctement e , sachant que J correspond à l'invariant concret et I à l'invariant abstrait, s'écrit : $I \wedge J \wedge G' \Rightarrow G \wedge [S'] \neg [S] \neg J$.

5.5.2 Ajout de nouveaux événements. Lors du raffinement, il est possible d'ajouter de nouveaux événements. Il faut donc prouver que les nouveaux événements raffinent bien l'évènement vide *SKIP* de l'abstraction en démontrant l'obligation de preuve suivante : $I \wedge J \wedge G' \Rightarrow [S'] J$. De plus, une preuve supplémentaire doit être validée afin de montrer que le nouvel événement ne peut prendre le contrôle indéfiniment. Pour cela il faut montrer que chaque nouvel événement fait décroître un variant, défini au préalable dans la clause VARIANT du raffinement. Étant donné un événement concret e' , défini par une garde G' et une substitution S' , un invariant abstrait I , un invariant concret j ; sachant que V corre-

<p>MODEL <i>Clock</i></p> <p>VARIABLES <i>h</i></p> <p>INVARIANT $h \in 0..23$</p> <p>ASSERTIONS $h < 100$</p> <p>INITIALISATION $h := 13$</p> <p>EVENTS <i>incr</i> = SELECT $h \neq 23$ THEN $h := h + 1$ END; <i>zero</i> = SELECT $h = 23$ THEN $h := 0$ END END</p>	<p>REFINEMENT <i>ClockWMinute</i></p> <p>REFINES <i>Clock</i></p> <p>VARIABLES <i>h, m</i></p> <p>INVARIANT $m \in 0..59$</p> <p>ASSERTIONS $(h \neq 23) \vee (h = 23) \Rightarrow$ $(h \neq 23 \wedge m = 59) \vee (h = 23 \wedge m = 59) \vee (m \neq 59)$</p> <p>VARIANT $59 - m$</p> <p>INITIALISATION $h := 13 \parallel m := 14$</p> <p>EVENTS <i>incr</i> = SELECT $h \neq 23 \wedge m = 59$ THEN $h := h + 1 \parallel m := 0$ END; <i>zero</i> = SELECT $h = 23 \wedge m = 59$ THEN $h := 0 \parallel m := 0$ END;</p>	<p><i>ticTac</i> = SELECT $m \neq 59$ THEN $m := m + 1$ END</p>
---	--	--

Fig. 4. Modèle B événementiel d'une horloge et son raffinement

spond au variant et V_0 au variant avant le déclenchement de e' , il faut montrer l'obligation de preuve suivante : $(I \wedge J \Rightarrow V \in \mathbb{N}) \wedge (I \wedge J \wedge G' \Rightarrow [S'](V < V_0))$.

5.5.3 Le raffinement de modèle. Un modèle concret raffine correctement un modèle abstrait s'il ne se bloque pas plus que le modèle abstrait. Pour cela nous devons montrer que la disjonction des gardes abstraites implique la disjonction des gardes concrètes. Sachant que les G_i représentent les gardes abstraites et les G'_i les gardes concrètes, l'obligation de preuve à montrer sera donc de la forme : $I \wedge J \wedge (G_1 \vee \dots \vee G_m) \Rightarrow G'_1 \vee \dots \vee G'_m$.

5.6 Un exemple de raffinement B événementiel

Dans la spécification du raffinement *ClockWMinute* (cf. figure 4), nous introduisons une nouvelle variable m , dénotant les minutes, et un nouvel événement *ticTac*. Ce dernier incrémente la variable m tant que $m \neq 59$. Avant de démontrer qu'il ne prend pas le contrôle indéfiniment, nous montrons que le variant $59 - m$ décroît. L'introduction des minutes a un impact sur les événements *incr* et *zero*. Nous enrichissons leurs gardes en ajoutant la contrainte $m = 59$ et aussi leurs actions en ajoutant la réinitialisation de la variable m . Une condition de non blocage est exprimée dans la clause **ASSERTIONS** qui, comme nous l'avons dit précédemment, consiste à vérifier que la disjonction des gardes abstraites implique la disjonction des gardes concrètes.

Task ::=	$Task \gg Task$	-- Sequence
	$Task [] Task$	-- Choix
	$Task Task$	-- Parallèle
	$Task \parallel Task$	-- Ordre indépendant
	$[Task]$	-- Tâche optionnelle
	$Task [>] Task$	-- Désactivation
	$Task > Task$	-- Interruption
	$Task^* [>] Task$	-- Désactivation d'une tâche infinie
	$Task^N$	-- Tâche itérative finie
	$Task_{At}$	-- Tâche atomique

Fig. 5. Grammaire de la notation CTT.

6. LA VALIDATION DE TÂCHES

Les insuffisances de l'approche de validation à base de traces explicites évoquées en section 4 nous ont conduit à proposer une extension où nous utilisons un langage de modélisation de tâches utilisateurs défini par la communauté IHM, en l'occurrence ConcurTaskTrees (CTT), qui permet :

- de décrire des tâches complexes par des expressions combinées à des opérateurs temporels ;
- d'éviter la définition de traces pour chaque tâche grâce à l'introduction d'autres opérateurs de contrôle. Il sera inutile d'énumérer toutes les séquences possibles de tâches à valider. Elles sont simplement caractérisées par le modèle de tâches CTT.

6.1 La notation de modélisation de tâches CTT

CTT (ConcurTaskTrees [Paternò 2001]) a été conçu pour décrire l'activité de l'utilisateur jusqu'au niveau de l'interaction. Il est résolument orienté vers la description de systèmes interactifs. Il permet de décrire des tâches utilisateurs en les combinant avec des opérateurs temporels. Un modèle de tâches CTT est basé sur une hiérarchie de tâches semblable à une structure d'arbre. Chaque tâche d'un même niveau est composée par un opérateur temporel qui détermine l'ordonnancement de ce niveau.

Sur la figure 5, la grammaire décrivant la syntaxe de la notation CTT est donnée. Elle présente des opérateurs temporels d'une algèbre de processus classique à la CCS [Milner 1980].

Une représentation graphique d'un arbre CTT est également associée à cette notation. Par exemple, une représentation graphique de l'expression donnée ci-dessous est présentée sur la figure 6. Elle a été produite par l'outil CTTE [Paternò et al. 2001] pour la tâche T_0 décomposée par la suite.

Dans les prochaines sections, nous montrons comment la sémantique de la notation CTT peut être formellement décrite en B événementiel. Des règles de traduction génériques sont proposées pour chaque opérateur.

6.2 Règles génériques pour la traduction des opérateurs de CTT

Cette approche utilise la capacité du raffinement de B événementiel. Pour toutes les règles de traduction, var_i représente l'état des variables, T_i le processus (la tâche), G_i la garde

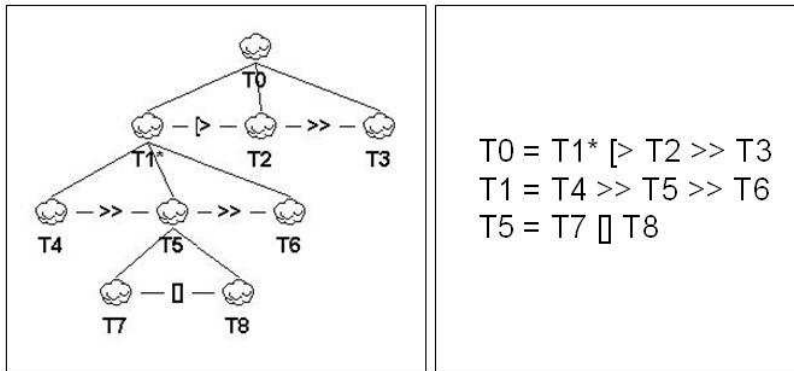


Fig. 6. Une représentation graphique d'une expression CTT obtenue avec l'outil CTTE

des événements et S_i une substitution généralisée correspondant aux actions effectuées par le processus T_i .

Chaque opérateur de CTT est représenté par un raffinement. Ainsi une règle BNF de la forme $T_0 ::= T_1 op T_2$ est décomposée en deux machines : la première contient l'événement T_0 et la seconde raffinant la première contient les événements T_0 , T_1 et T_2 et effectue la décomposition de T_0 en $T_1 op T_2$ et où op est un opérateur CTT.

Les règles de traduction en B événementiel des opérateurs basiques (séquence, choix, parallèle, itération finie et tâche atomique) sont données dans cette section. Ces opérateurs sont utilisés pour coder les opérateurs restants (ordre indépendant, tâche optionnelle, désactivation, désactivation d'une tâche itérative et interruption).

```

MODEL  $T_0$ 
INVARIANT  $I(var_i)$ 
EVENTS
 $EvtT_0 =$ 
SELECT
   $G_0$ 
THEN
   $S_0$ 
END;

```

Fig. 7. Modélisation de la tâche racine T_0

Pour ces règles de traduction, nous utilisons une tâche T_0 représentée par l'événement $EvtT_0$ (figure 7), considérée comme la tâche racine. Cette tâche est décomposée pour traduire les opérateurs basiques. L'ensemble des variables var_i décrit tous les états des variables qui caractérisent le processus de la tâche T_0 . D'autres variables peuvent être ajoutées au niveau du raffinement, si besoin, pour observer des nouveaux éléments introduits par la décomposition. S_0 est une substitution, sous couvert que la garde G_0 est vraie, qui exprime le changement d'état de var_i pour le processus de la tâche T_0 .

RAFFINEMENT <i>RefSequenceT₀</i>		
REFINE <i>T₀</i>		
INVARIANT		
$J(var_i, var_j) \wedge EtatAct \in \{0, 1, 2\} \wedge J'(var_j)$		
ASSERTIONS		
$G_0 \Rightarrow ((EtatAct = 2 \wedge G_1) \vee (EtatAct = 1 \wedge G_2) \vee (EtatAct = 0 \wedge G'_0))$		
VARIANT		
<i>EtatAct</i>		
INITIALISATION		
$EtatAct := 2$		
EVENTS		
$EvtT_1 =$	$EvtT_2 =$	$EvtT_0 =$
SELECT	SELECT	SELECT
$EtatAct = 2 \wedge G_1$	$EtatAct = 1 \wedge G_2$	$EtatAct = 0 \wedge G'_0$
THEN	THEN	THEN
$EtatAct := 1 \parallel$	$EtatAct := 0 \parallel$	S'_0
S_1	S_2	END;
END;	END;	

Fig. 8. Codage en B événementiel de l'opérateur séquence

6.2.1 *Séquence ">>".* Considérons la tâche suivante $T_0 ::= T_1 \gg T_2$ qui exprime l'activation de la tâche T_1 suivie de la tâche T_2 . La traduction en B événementiel est donnée par un modèle (cf. figure 8) avec deux événements $EvtT_1$ et $EvtT_2$ correspondant respectivement aux tâches T_1 et T_2 . La traduction utilise un variant *EtatAct* initialisé à 2 et qui décroît jusqu'à la valeur 0. Si la garde de $EvtT_1$ est vraie, l'événement est déclenché, sa substitution généralisée est exécutée et le variant décroît. Ensuite, l'événement $EvtT_2$ se déclenche si sa garde est vraie et si la valeur du variant est mise à 1 par $EvtT_1$. La disjonction des gardes est donnée dans la clause ASSERTIONS. La variable var_j définit l'état des variables du raffinement. Elles sont liées aux variables abstraites du modèle par l'invariant de collage $J(var_i, var_j)$. Cet invariant de collage est défini dans tous les raffinements suivants.

Notez que l'événement $EvtT_0$ termine l'opération de séquence des tâches T_1 et T_2 . L'invariant de collage assure que l'événement $EvtT_0$ du raffinement raffine celui de l'abstraction.

6.2.2 *Choix "[]".* Considérons la tâche $T_0 ::= T_1 [] T_2$ définissant un choix non déterministe entre la tâche T_1 et T_2 , c'est-à-dire que T_1 ou T_2 est déclenchée. La traduction en B événementiel est donnée par un modèle (cf. figure 9) avec deux événements gardés $EvtT_1$ et $EvtT_2$. Le variant *EtatChoix* est arbitrairement initialisé soit à 1 soit à 2. Selon la valeur de la garde de chaque événement, $EvtT_1$ ou $EvtT_2$ est déclenché. Chaque événement fait décroître la valeur du variant à la valeur 0 interdisant les autres événements de se déclencher. L'événement $EvtT_0$ finalise le codage de l'opérateur choix entre les deux tâches. Il décrit le raffinement de l'événement correspond à la tâche T_0 .

6.2.3 *Tâche itérative finie "T^N".* Considérons la tâche itérative suivante $T_0 ::= T_1^N$. Le principe de codage d'une boucle en B événementiel consiste à déclencher N fois les événements associés à la tâche T_1 . Un variant est utilisé. Il décroît de N vers 0. La traduction d'une boucle codée en B événementiel (cf. figure 10) nécessite trois événements : un

```

RAFFINEMENT RefChoixT0
REFINE T0
INVARIANT
   $J(\text{var}_i, \text{var}_j) \wedge \text{EtatChoix} \in \{0, 1, 2, 3\} \wedge J'(\text{var}_j)$ 
ASSERTIONS
   $G_0 \Rightarrow ((\exists p).(p \in \{1, 2\} \wedge \text{EtatChoix} = 3)) \vee (G_1 \wedge \text{EtatChoix} = 1) \vee$ 
   $(G_2 \wedge \text{EtatChoix} = 2) \vee (\text{EtatChoix} = 0 \wedge G'_0)$ 
VARIANT
  EtatChoix
INITIALISATION
  EtatChoix := 3
EVENTS
EvtInitChoix =      EvtT1 =      EvtT2 =      EvtT0 =
ANY p      SELECT      SELECT      SELECT
WHERE      EtatChoix = 1  $\wedge$  G1      EtatChoix = 2  $\wedge$  G2      EtatChoix = 0
  p  $\in$  {1, 2}  $\wedge$       THEN      THEN      THEN
  EtatChoix = 3      EtatChoix := 0 ||      EtatChoix := 0 || S2       $\wedge$  G'0
THEN      S1      END;      S'0
  EtatChoix := p      END;      END;
END;

```

Fig. 9. Codage en B événementiel de l'opérateur choix

```

RAFFINEMENT RefIterationFinieT0
REFINE T0
INVARIANT
   $J(\text{var}_i, \text{var}_j) \wedge \text{EtatLoop} \in \text{NAT} \wedge \text{Debut} \in \{0, 1\} \wedge J'(\text{var}_j)$ 
ASSERTIONS
   $G_0 \Rightarrow (\text{Debut} = 1 \vee (\text{EtatLoop} > 0 \wedge \text{Debut} = 0 \wedge G_1) \vee (\text{EtatLoop} = 0 \wedge \text{Debut} = 0 \wedge G'_0))$ 
VARIANT
  EtatLoop + Debut
INITIALISATION
  Debut := 1
EVENTS
EvtInitLoop =      EvtLoop =      Evt0 =
SELECT      SELECT      SELECT
  Debut = 1      EtatLoop > 0  $\wedge$       EtatLoop = 0  $\wedge$ 
THEN      Debut = 0  $\wedge$  G1      Debut = 0  $\wedge$  G'0
  EtatLoop := NAT ||      THEN      THEN
  Debut := 0      EtatLoop := EtatLoop - 1 ||      S'0
END;      SLoop      END;
      END;

```

Fig. 10. Codage en B événementiel de la caractéristique itération finie

premier événement $Evt_{InitLoop}$ pour l'initialisation du variant $EtatLoop$, un deuxième événement Evt_{Loop} pour exécuter le corps de la boucle et décroître le variant, et enfin un troisième événement $EvtT_0$ pour terminer la boucle. Le variant $EtatLoop$ modélisant le nombre d'itérations est initialisé par l'événement $Evt_{InitLoop}$. L'événement Evt_{Loop} est alors déclenché $EtatLoop$ fois. Quand l'itération est terminée, Evt_0 se déclenche. Le variant $EtatLoop$ décroît de sa valeur arbitraire jusqu'à 0. La clause ASSERTIONS assure qu'au moins une garde d'un des événements est toujours respectée. La valeur initiale du variant est arbitrairement fixée par l'opérateur $:\in$. L'avantage d'une telle approche est la possibilité d'encoder un nombre arbitraire d'une boucle sans augmenter la complexité du processus de preuve. À l'inverse, les techniques formelles dites par vérification sur modèles souffriraient du problème d'explosion combinatoire des états lors de l'augmentation du nombre d'itérations.

RAFFINEMENT $RefParalleleT_0$		
REFINE T_0		
INVARIANT		
$J(var_i, var_j) \wedge EtatConc_1 \in \{0, 1\} \wedge EtatConc_2 \in \{0, 1\} \wedge J'(var_j)$		
ASSERTIONS		
$G_0 \Rightarrow ((G_1 \wedge EtatConc_1 = 1) \vee (G_2 \wedge EtatConc_2 = 1) \vee (EtatConc_1 = 0 \wedge EtatConc_2 = 0 \wedge G'_0))$		
VARIANT		
$EtatConc_1 + EtatConc_2$		
INITIALISATION		
$EtatConc_1 := 1 \parallel EtatConc_2 := 1$		
EVENTS		
$EvtT_1 =$	$EvtT_2 =$	$EvtT_0 =$
SELECT	SELECT	SELECT
$G_1 \wedge EtatConc_1 = 1$	$G_2 \wedge EtatConc_2 = 1$	$EtatConc_1 = 0 \wedge$ $EtatConc_2 = 0 \wedge G'_0$
THEN	THEN	THEN
$EtatConc_1 = 0 \parallel$ S_1	$EtatConc_2 = 0 \parallel$ S_2	S'_0
END;	END;	END;

Fig. 11. Codage en B événementiel de l'opérateur parallèle

6.2.4 *Parallèle "||"*. Considérons la tâche suivante $T_0 ::= T_1 || T_2$. La sémantique du parallélisme par entrelacement impose de décrire tous les comportements possibles en énumérant donc toutes les traces. Cet opérateur exploite l'entrelacement mis en avant par la sémantique du B événementiel, i.e. si deux événements ont leur garde à vraie, ils sont déclenchés en parallèle d'une manière entrelacée (cf. figure 11). Deux événements $EvtT_1$ et $EvtT_2$ correspondant aux tâches T_1 et T_2 sont définis. Une fois que les événements $EvtT_1$ et $EvtT_2$ sont déclenchés, ils ne peuvent pas être déclenchés de nouveau. Leur variant (respectivement $EtatConc_1$ et $EtatConc_2$) décroît de un à zéro.

6.3 Traduction des autres opérateurs temporels

Les opérateurs de séquence, de parallélisme, de choix et la caractéristique itération finie ont été codés dans des modèles B événementiels. Il existe d'autres opérateurs (or-

dre indépendant, tâche optionnelle, désactivation). Leur traduction est obtenue en utilisant les opérateurs basiques définis précédemment en exploitant également la sémantique d'entrelacement de la méthode B événementiel. Nous invitons les lecteurs à parcourir la référence suivante de façon à trouver la description exhaustive de la traduction de la grammaire CTT [Baron 2003; Aït-Ameur et al. 2009].

6.4 Méthodologie de conception

La conception d'un modèle B événementiel à partir d'une expression de tâche avec CTT par l'application des règles proposées dans cet article permet de vérifier et de valider un système interactif d'une manière incrémentale. Notre objectif est de représenter chaque tâche abstraite de CTT réalisée par un ensemble de tâches concrètes ou abstraites (ajout de nouvelles tâches), ordonnées par les différents opérateurs CTT, par l'opération du raffinement de B événementiel (ajout de nouveaux événements).

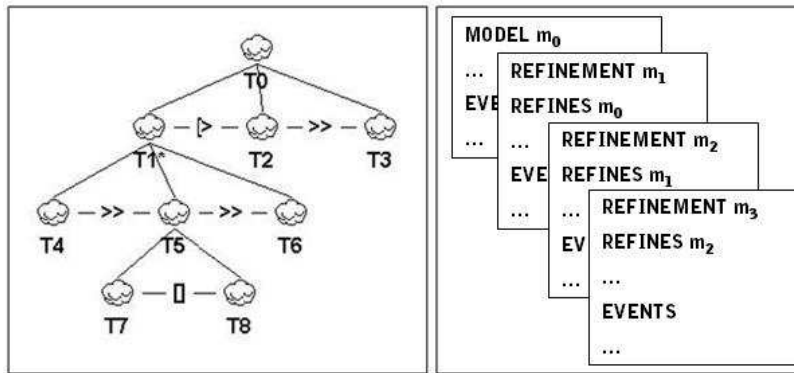


Fig. 12. Différents modèles B événementiel obtenus à partir d'une tâche CTT

Sur l'exemple de la figure 12, chaque tâche T_i est spécifiée par un événement nommé Evt_{T_i} . L'application de notre méthodologie sur l'exemple aboutit au départ à un modèle abstrait nommé m_0 contenant l'évènement Evt_{T_0} . En appliquant les règles définies dans la section 6.2, l'évènement Evt_{T_0} est raffiné par les évènements Evt_{T_1} , Evt_{T_2} et Evt_{T_3} dans un modèle m_1 lié au modèle m_0 par raffinement. Par la suite, l'évènement Evt_{T_1} est raffiné par les évènements Evt_{T_4} , Evt_{T_5} et Evt_{T_6} dans le modèle m_2 lié également au modèle m_1 par raffinement. De la même manière, ce processus se déroulera jusqu'à ce qu'il n'y ait plus de tâche CTT qui se décompose en sous tâches.

7. L'ARCHITECTURE LOGICIELLE

La grande majorité des modèles d'architecture de systèmes interactifs distingue trois composants essentiels : le noyau fonctionnel, la présentation et le contrôleur de dialogue. Ces modèles séparent ces trois composants pour des raisons de modularité et pour permettre une validation et vérification compositionnelles. Cette partie s'intéresse à la modélisation et à la vérification formelles du composant contrôleur de dialogue. Plusieurs scénarios de développement de ce composant sont possibles.

7.1 Architecture logicielle et B événementiel

Nous présentons en premier la modélisation, en B événementiel, des composants noyau fonctionnel et présentation. Chacun de ces composants est un système de transitions. Ce système est modélisé en B par un état déclaré dans la clause **VARIABLES** et un ensemble d'événements dans la clause **EVENTS**. Ces événements décrivent les changements d'état (les transitions). Les clauses **INVARIANT** et **ASSERTIONS** comportent les propriétés pertinentes de ces systèmes. Enfin, ces systèmes sont construits par une suite de raffinements qui préservent les propriétés. On notera $MODEL_{NF}$ le modèle B décrivant le noyau fonctionnel et $MODEL_{IHM}$ celui qui décrit la présentation (cf. figure 13).

MODEL <i>MODEL_{NF}</i>	MODEL <i>MODEL_{IHM}</i>
VARIABLES <i>Var_{NF}</i>	VARIABLES <i>Var_{IHM}</i>
INVARIANT <i>I(Var_{NF})</i>	INVARIANT <i>I(Var_{IHM})</i>
ASSERTIONS <i>A(Var_{NF})</i>	ASSERTIONS <i>A(Var_{IHM})</i>
INITIALISATION <i>Var_{NF} := ...</i>	INITIALISATION <i>Var_{IHM} := ...</i>
EVENTS <i>Evt =</i> SELECT <i>G_{NF}</i> THEN <i>S_{NF}</i> END	EVENTS <i>Evt =</i> SELECT <i>G_{IHM}</i> THEN <i>S_{IHM}</i> END
END	END

Fig. 13. Modèles décrivant respectivement le noyau fonctionnel et la présentation

Le contrôleur de dialogue est représenté par le produit synchronisé des deux systèmes de transitions représentant respectivement le noyau fonctionnel et la présentation. L'état du composant obtenu est une paire $Var = (Var_{NF}, Var_{IHM})$ composée des variables du noyau fonctionnel et de celles de la présentation et les événements sont de la forme présentée sur la figure 14.

G_{NF} et S_{NF} correspondent respectivement à la garde et à la substitution de la partie noyau fonctionnel. À l'inverse, G_{IHM} et S_{IHM} correspondent à la garde et à la substitution de la partie interface utilisateur. Notons que des événements d'attente peuvent apparaître dans le cas où S_{NF} et S_{IHM} valent SKIP.

La preuve des modèles B obtenus dépend de la façon de construire ce produit. Cette construction exploite le raffinement. Nous avons identifié quatre scénarios de conception d'IHM. Chacun décrit une implémentation de la composition des composants noyau fonctionnel et de la présentation pour définir le contrôleur de dialogue :

- (1) description des raffinements B du noyau fonctionnel puis introduction des événements de la présentation par raffinement ;
- (2) description des raffinements B de la présentation puis introduction des événements du noyau fonctionnel par raffinement ;

- (3) composition des événements du noyau fonctionnel et de la présentation dans le même modèle B ;
- (4) description des raffinements B de la présentation avec un noyau fonctionnel abstrait.

$Evt =$ <pre> SELECT $G_{NF} \wedge G_{IHM}$ THEN $S_{NF} \parallel S_{IHM}$ END; </pre>

Fig. 14. Forme d'un événement du contrôleur de dialogue

Dans la suite nous montrons uniquement le scénario quatre que nous avons appliqué à notre étude de cas. Une description exhaustive des scénarios peut être trouvée dans les références suivantes [Aït-Ameur et al. 2006] et [Aït-Ameur et al. 2008].

<pre> MODEL $MODEL_{NF}$ VARIABLES Var_{NF} INVARIANT $var_{NF} \in \{0, 1\}$... $Evt_{NF} =$ SELECT $var_{NF} = 1$ THEN $var_{NF} := 0$ END; END </pre>	<pre> MODEL $MODEL_{IHM}$ VARIABLES Var_{IHM} INVARIANT $I(Var_{IHM})$... $Evt_{IHM} =$ SELECT G_{IHM} THEN S_{IHM} END; END </pre>
--	---

Fig. 15. Modèles décrivant respectivement l'abstraction du noyau fonctionnel et la présentation pour le scénario 4

7.2 Abstraction du noyau fonctionnel

Ce scénario est obtenu par une description des raffinements B de la présentation. Nous considérons la partie noyau fonctionnel comme étant abstraite. Nous ne nous intéressons pas aux activités effectuées par le noyau fonctionnel ni à leur spécification. Nous enregistrons, au niveau du contrôleur de dialogue, le fait qu'un événement du noyau fonctionnel a été déclenché. Pour cela, nous employons une variable entière avec un comportement booléen, qui détermine si l'état du noyau fonctionnel a été modifié ou pas par le déclenchement d'un événement.

Nous donnons sur la figure 15 deux modèles concernant les parties noyau fonctionnel et interface utilisateur. Sur la gauche, nous utilisons la variable Var_{NF} pour abstraire l'état du noyau fonctionnel. Si l'événement Evt_{NF} est déclenché, c'est-à-dire quand Var_{NF} est égal à un, l'état du noyau fonctionnel est modifié à zéro.

Nous montrons sur la figure 16 le résultat de la composition les deux modèles précédents obtenus après raffinement.

```

MODEL
  MODELCD
VARIABLES
  VarNF, VarIHM
INVARIANT
  VarNF ∈ {0, 1} ∧ I(VarIHM)
  ...
  Evt =
    SELECT GIHM ∧ VarNF = 1
    THEN
      SIHM || VarNF := 0
    END;
END

```

Fig. 16. Modèle décrivant la composition de l'abstraction du noyau fonctionnel et de la présentation pour le scénario 4

8. APPLICATIONS AUX IHM3

La modélisation d'une IHM3 en B événementiel suit une démarche de conception descendante fondée sur le raffinement. Des événements de haut niveau sont déclarés dans le modèle abstrait racine. Ce dernier est raffiné par l'introduction de nouveaux événements qui précisent le modèle abstrait initial. Nous avons appliqué nos travaux à l'étude de cas "Pages Jaunes", décrite ci-après, et avons retenu le scénario 4. Nous avons utilisé une abstraction du noyau fonctionnel permettant une vérification formelle modulaire.

Nous détaillons dans la section qui suit une modélisation de l'IHM3 représentant un modèle de l'étude de cas présentée ci-dessus correspondant au scénario 4 qui ne prend en compte que la multi-modalité en entrée.

8.1 Étude de cas : "Pages Jaunes CLIPS"

L'étude de cas "Pages Jaunes CLIPS", développée par l'équipe IHM du laboratoire CLIPS-IMAG, est une application multi-modale implémentée en utilisant la plate-forme de composants d'interactions multi-modales ICARE (Interaction-Care [Bouchet et al. 2004]). Cette application (cf. figure 17) permet de rechercher un contact en spécifiant le nom (Name) d'une personne et une adresse (Address) de localisation. Suite à l'envoi de la requête de recherche de la personne (Search), un plan est affiché. Sur ce plan, l'utilisateur peut naviguer et cibler sa recherche. L'application "Pages Jaunes CLIPS" regroupe un ensemble de commandes permettant à l'utilisateur d'accomplir la tâche (spécifier nom, spécifier adresse, envoyer requête, déplacement haut/bas/gauche/droite sur la carte, agrandissement avant/arrière, ...). Il peut déclencher ces commandes en utilisant différentes modalités d'interaction : la voix, la souris et le clavier, ou une combinaison de ces modalités, en s'appuyant sur l'utilisation individuelle et synergique des différentes modalités d'entrées.

De cette manière l'utilisateur peut remplir ces champs en utilisant uniquement la voix en prononçant le nom de la personne et le lieu, ou en combinant de façon complémentaire les modalités de la voix, de la souris et du clavier (sélection du champs puis prononciation du nom ou alors en choisissant oralement le champs et en saisissant la valeur par le clavier). Les modalités peuvent également être utilisées de manière redondante dans le cas par exemple où l'utilisateur spécifie par la voix la commande "agrandissement" et en

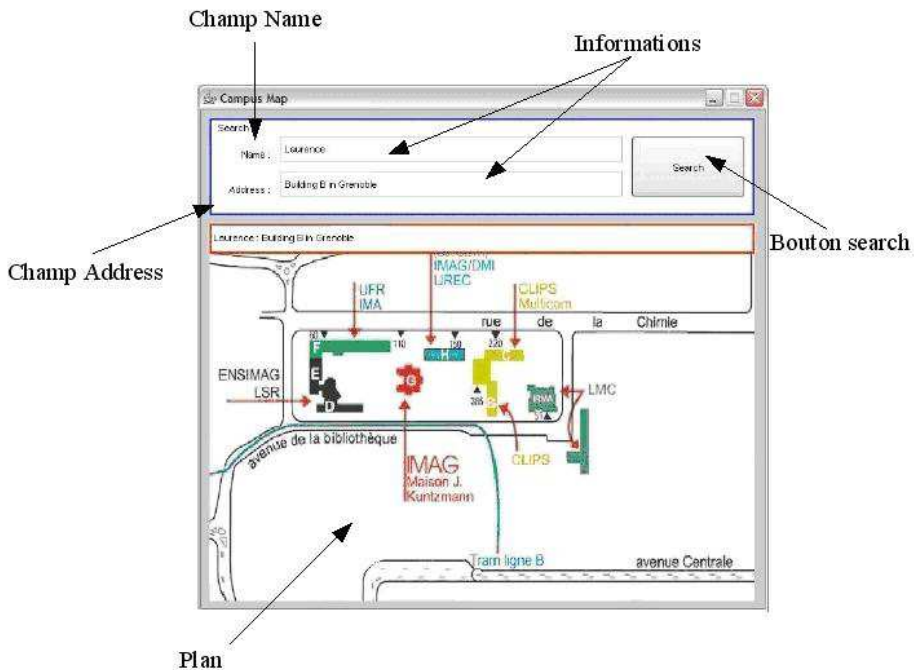


Fig. 17. IHM de l'étude de cas "Pages Jaunes Clips".

parallèle appuie sur la touche du clavier correspondante à cet effet, une seule commande d'agrandissement sera alors transmise au système.

Notons que nous ne nous intéressons dans cet article qu'à la multi-modalité en entrée (de l'utilisateur vers le système "Pages Jaunes CLIPS"), c'est pourquoi, nous ne détaillerons pas la multi-modalité en sortie.

8.2 Modélisation par raffinement

Trois modèles ont été définis. Ils sont reliés par une relation de raffinement. Le premier introduit les événements de haut niveau. Le deuxième présente les événements de la présentation ainsi que leur synchronisation avec le CD. Enfin, le troisième introduit les interactions multi-modales avec les différentes possibilités d'interaction. Ces modèles font tous abstraction du noyau fonctionnel. Ils ne font que rendre compte du fait que des événements abstraits du noyau fonctionnel ont été déclenchés.

Dans la suite, nous décrivons une représentation simplifiée du développement de l'IHM3 associée à l'étude de cas "Pages Jaunes CLIPS". Nous avons volontairement réduit les modèles B événementiel afin d'en faciliter la lecture. Le code complet peut être demandé directement aux auteurs.

8.2.1 Le modèle racine. Le modèle racine est un modèle abstrait représentant la synchronisation entre les événements de saisie et de déclenchement du noyau fonctionnel (cf. figure 18). Deux événements *Query* et *Result_query* décrivent ce modèle. Le premier indique ($Query_sending := 1$) que la requête est prête à être envoyée au noyau fonction-

```

MODEL IMAPPY
VARIABLES
  Query_sending, Query_ready
INVARIANT
  Query_sending ∈ 0..1 ∧ Query_ready ∈ 0..1 ∧
  Query_sending ≠ Query_ready
ASSERTIONS
  (Query_ready = 1) ∨ (Query_sending = 1)
INITIALISATION
  Query_sending || Query_ready := 0, 1

EVENTS
Query =
  SELECT Query_ready = 1
  THEN
    Query_sending := 1 || Query_ready := 0
  END;

Result_query =
  SELECT Query_sending = 1
  THEN
    Query_sending := 0 || Query_ready := 1
  END
END

```

Fig. 18. Modèle racine représentant la synchronisation entre les événements de saisie et déclenchement du noyau fonctionnel

nel alors que le second indique que le résultat de la requête est prêt ($Query_ready := 1$). Les deux variables $Query_sending$ et $Query_ready$ décrivent la synchronisation de ces deux événements. Notons que les événements décrits ici sont abstraits (aucun élément relatif aux différentes modalités d'interaction) et n'indiquent que l'ordre correct de déclenchement. Une première propriété est établie à ce niveau : le noyau fonctionnel n'est sollicité qu'après avoir déclenché l'événement de l'interface.

La suite de cet article s'intéresse au raffinement de l'événement $Query$ qui décrit la saisie du nom et de l'adresse selon différentes modalités.

8.2.2 Identification des interactions abstraites. L'événement $Query$ est décomposé de sorte à prendre en compte les différentes possibilités de saisie du nom et de l'adresse de la personne à rechercher. Ainsi, on pourra :

- saisir le nom un nombre arbitraire de fois : $Input_Name^*$;
- saisir l'adresse un nombre arbitraire de fois : $Input_Address^*$;
- lancer ensuite la recherche : $Search$.

Notons que l'on peut entrelacer la saisie du nom et de l'adresse (saisir l'un puis l'autre dans n'importe quel ordre et retour). Pour décrire les différentes tâches interactives que l'utilisateur peut effectuer sur le système des "Pages Jaunes CLIPS", nous utilisons les opérateurs temporels de la notation CTT. Ainsi la tâche CTT qui décrit les différentes possibilités de saisie est définie par :

$$Query = (Input_Name^* || Input_Address^*) \gg Search \gg Result_query \quad (1)$$

La tâche *Query* est implantée dans le raffinement *IMAPPY_Ref_1* décrit sur les figures 19 et 20. Nous nous concentrons sur les événements introduits pour raffiner les deux événements principaux du niveau abstrait.

<p>REFINEMENT IMAPPY_Ref_1 REFINES IMAPPY SETS <i>string</i> VARIABLES <i>Query_sending, EV1, Nn, Na,</i> <i>map, name, address</i> INVARIANT $EV1 \in 0..3 \wedge Nn \in N \wedge Na \in N \wedge$ $map \in BOOL \wedge name \subseteq string \wedge address \subseteq string \wedge$ $(Query_ready = 1 \Rightarrow EV1 \neq 0)$ INITIALISATION $Query_sending := 0 \parallel map, EV1 := FALSE, 3 \parallel name, address, Nn, Na := 0, 0, 1, 1$</p>
--

Fig. 19. Raffinement de la tâche Query

Nous avons introduit de nouvelles variables pour la description de l'interaction en entrée. Tout d'abord, nous décrivons les variables *name* et *address* de type *string* qui récupèrent respectivement le nom et l'adresse. Les variables entières *Nn* et *Na* désignent le nombre arbitraire d'itérations * de CTT. Elles sont initialisées, dans l'événement *Name_Address*, par un nombre entier arbitraire (grâce à l'opérateur \in) qui indique le nombre de fois que cet événement est déclenché. Enfin, la variable *EV1* (variable entière décroissante) initialisée à 3 est un invariant qui décrit l'ordre de déclenchement de chaque événement.

Les événements de la figure 20 décrivent la tâche CTT précédente. L'événement *Name_Address* initialise les variants *Nn* et *Na* alors que les trois événements *Input_Name*, *Input_Address* et *Search* décomposent (raffinent) l'événement abstrait *Query*. Lorsque ces trois événements ont été déclenchés, ils rendent le contrôle à l'événement abstrait *Query* qui ne fait que le transmettre aux autres événements.

Notons qu'à ce niveau, nous avons pris en compte les déclenchements des interactions sans entrer dans les détails de leurs définitions. Cela sera abordé dans les prochains raffinements et nous permet de nous concentrer sur la correction des synchronisations et des ordres de déclenchement.

Comme nous l'avions déclaré précédemment, nous ne nous intéressons pas à la multi-modalité en sortie ni au noyau fonctionnel. C'est pour cette raison que les événements associés (*Result_Query*) n'ont pas été raffinés. Cette démarche montre qu'il est possible de développer les différents aspects d'une IHM3 de façon modulaire afin de simplifier le processus de preuve. En effet, les obligations de preuve engendrées sont allégées du fait que certains événements restent abstraits.

8.2.3 Identification des modalités. La dernière étape de raffinement consiste à identifier et à introduire les différentes interactions multi-modales en entrée mises en jeu (cf. figure 21 et figure 22). Ainsi, on décrit les ensembles *SPEECH*, *KEYBOARD* et *MOUSE* indiquant et typant les différentes interactions multi-modales possibles. Notons que ces ensembles peuvent être enrichis ou réduits sans remettre en cause la modélisation

```

EVENTS
Name_Address =
  SELECT EV1 = 3
  THEN
    EV1 := EV1 - 1 || Nn :∈ N || Na :∈ N
  END;

Input_Name =
  SELECT EV1 = 2 ∧ Nn ≠ 0
  THEN
    Nn := Nn - 1 || name :∈ P(string)
  END;

Input_Address =
  SELECT EV1 = 2 ∧ Na ≠ 0
  THEN
    Na := Na - 1 || address :∈ P(string)
  END;

Search =
  SELECT EV1 = 2 ∧ Nn = 0 ∧ Na = 0
  THEN
    EV1 := 1
  END;

Query =
  SELECT EV1 = 1
  THEN
    Query_sending, EV1 := 1, 0
  END;

Result_query = ...

END

```

Fig. 20. Raffinement de la tâche Query

effectuée. La variable *modality* décrit la modalité activée à un instant donné. Elle est initialisée à *NONE*. La variable *EV1* quant à elle indique que le focus est sur la saisie du nom.

Nous procédons à la décomposition (raffinement) de l'événement *Input_Name* et *Input_Address* qui font intervenir les modalités voix et clavier/souris. Pour cela, il est nécessaire d'enrichir le modèle B par la déclaration de nouveaux événements.

Le nom est entré en deux étapes : saisie du champ à saisir (le nom) par la voix ou le clavier/souris puis entrée du nom (valeur du nom) par la voix ou bien par clavier/souris. Ces deux événements peuvent être itérés un nombre arbitraire de fois. Les suffixes *field* et *value* indiquent respectivement les événements associés au choix du champ à remplir et à la valeur entrée pour ce champ. Dans la suite, seuls les événements raffinant l'événement *Input_Name* sont présentés selon la tâche :

$$Input_name^* = (Input_Name_Fied \gg Input_Name_Value)^* \quad (2)$$

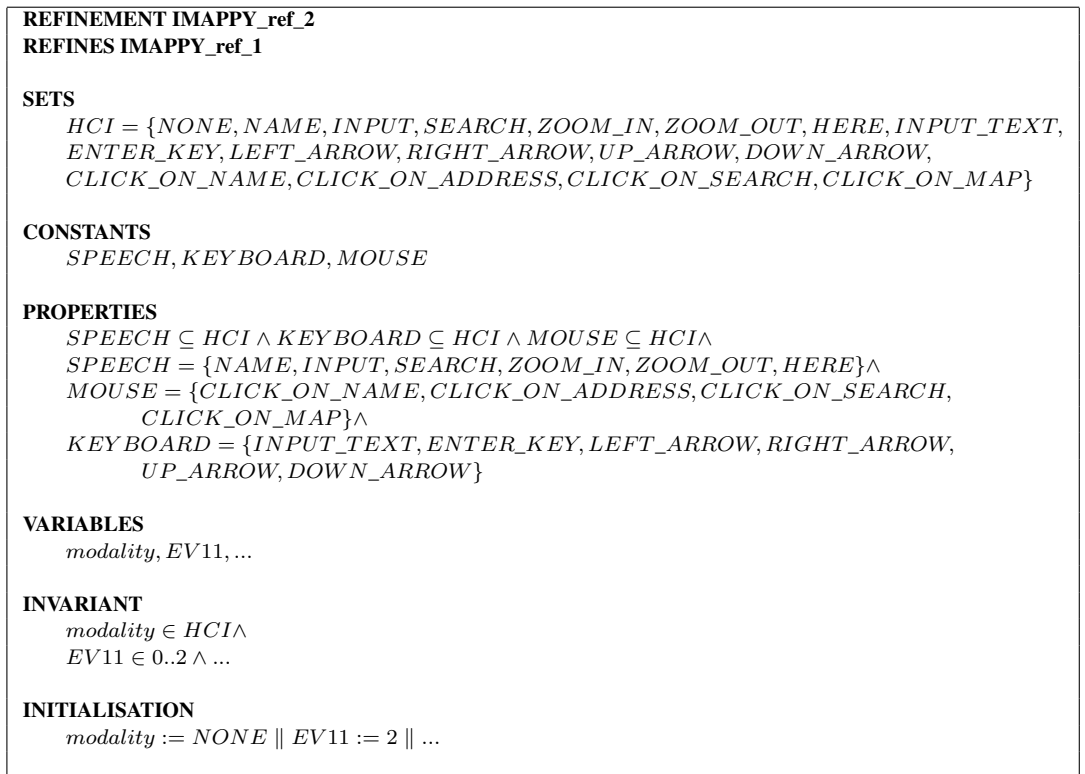


Fig. 21. Introduction des différentes interactions multi-modales

À ce niveau, il faut observer que tous les événements peuvent être déclenchés, c'est-à-dire, qu'il est possible de saisir le nom par la voix ou par le clavier/souris uniquement, ou bien par une combinaison des deux. On peut restreindre ces choix en fonction des propriétés CARE que l'on souhaite établir. Ce point est discuté dans la section suivante. Nous avons donc une description de toutes les possibilités.

8.3 Vérification des propriétés CARE par construction

La démarche développée dans cet article est fondée sur la notion de tâche. Nous avons représenté la totalité de l'IHM3 par des systèmes de transitions. Nous avons veillé à ne pas représenter le système de transitions explicitement, mais nous nous sommes appuyés sur la définition de tâches utilisateurs en prenant pour langage de tâches, la notation CTT. Nous avons montré que le modèle de tâches peut être représenté par un modèle B événementiel et que le modèle de tâches dirige le développement de modèles B événementiel.

Les tâches implantées en B événementiel permettent de valider des propriétés CARE grâce à la construction des modèles B événementiel. La représentation d'une interaction multi-modale satisfaisant les propriétés CARE est possible par le codage d'une tâche CTT décrivant le type de propriété que l'on souhaite représenter. Si une tâche interactive multi-modale $T0$ est réalisée par l'expression $T1_{m_i} \gg T2_{m_j}$ avec m_i et m_j qui représentent

```

EVENTS
Name_address = ...

Input_Name_Field =
  SELECT EV1 = 2 ∧ Nn ≠ 0 ∧ EV11 = 2
  THEN
    EV11 := EV11 - 1 ||
    modality :∈ {CLICK_ON_NAME, NAME}
  END;

Input_Name_Value =
  SELECT EV1 = 2 ∧ Nn ≠ 0 ∧ EV11 = 1
  THEN
    EV11 := EV11 - 1 ||
    modality :∈ {INPUT, INPUT_TEXT} ||
    name :∈ P(string) ||
    Nn := Nn - 1
  END;

Input_Name =
  SELECT EV1 = 2 ∧ Nn ≠ 0 ∧ EV11 = 0
  THEN
    EV11 := 2
  END;

Input_address = ...
Search = ...
Query = ...
Result_query = ...
END

```

Fig. 22. Introduction des différentes interactions multi-modales

la même modalité si $i = j$, ou deux modalités différentes si $i \neq j$. Pour la tâche $T0$ on obtient :

(1) une équivalence de modalités entre m_1 et m_2 avec l'implantation de la tâche :

$$\begin{aligned} & -((T1_{m_1} \gg T2_{m_1}) [] \\ & (T1_{m_2} \gg T2_{m_2}))^* ; \end{aligned}$$

(2) la redondance avec l'implantation de la tâche :

$$\begin{aligned} & -((T1_{m_1} \gg T2_{m_1}) || \\ & (T1_{m_2} \gg T2_{m_2}))^* ; \end{aligned}$$

(3) la complémentarité avec l'implantation de la tâche :

$$\begin{aligned} & -((T1_{m_1} \gg T2_{m_2}) [] \\ & (T1_{m_2} \gg T2_{m_1}))^* \end{aligned}$$

Dans notre étude de cas, le modèle B événementiel *IMMAPY_ref2* offre toutes les combinaisons d'interactions possibles. Par exemple pour coder la redondance de modalités avec la voix exclusivement, il suffit dans un nouveau raffinement d'affecter à la variable *modality* la valeur *NAME* dans l'événement *Input_Name_Field* et la valeur *INPUT* dans l'événement *Input_Name_Value*. De même pour coder la redondance avec le clavier/souris exclusivement, il suffit dans un raffinement d'affecter à la variable

modality la valeur *CLICK_ON_NAME* dans l'événement *Input_Name_Field* et la valeur *INPUT_TEXT* dans l'événement *Input_Name-Value*. Enfin pour coder la complémentarité de la voix et du clavier/souris exclusivement, il suffit, dans l'événement *Input_Name_Field*, d'affecter à la variable *modality* la valeur *NAME*, et la valeur *INPUT_* dans l'événement *Input_Name-Value*.

Pour illustrer ceci, les tâches faisant intervenir les modalités voix sont post-fixées par *V* et les tâches faisant intervenir les modalités clavier/souris sont post-fixées par *CS*. Ainsi pour la tâche *Input_Name*, on obtient :

- (1) une équivalence de modalités entre clavier-souris et voix avec l'implantation de la tâche :

$$-((Input_Name_Field_CS \gg Input_Name_Value_CS) [] \\ (Input_Name_Field_V \gg Input_Name_Value_V))^* ;$$

- (2) la redondance avec l'implantation de la tâche :

$$-((Input_Name_Field_CS \gg Input_Name_Field_CS) [] \\ (Input_Name_Field_V \gg Input_Name_Value_V))^* ;$$

- (3) la complémentarité avec l'implantation de la tâche :

$$-((Input_Name_Field_V \gg Input_Name_Field_CS) [] \\ (Input_Name_Field_CS \gg Input_Name_Field_V))^*$$

8.4 Bilan des obligations de preuve

Le tableau II illustre les résultats obtenus sur l'étude de cas. 432 obligations de preuve sont automatiquement générées. Certaines d'entre elles sont dites triviales (Obv). En effet, ces dernières sont "si évidentes" qu'elles sont validées par l'outil dès leur génération. Toutefois, toutes les obligations de preuve ne sont pas aussi simples à démontrer. L'Atelier B propose deux modes pour les valider : un mode automatique et un mode interactif. Dans notre cas les 59 obligations de preuve "non triviales" n'ont pas toutes été démontrées dans le mode automatique. Seules 3 d'entre elles ont nécessité l'utilisation d'une preuve interactive i.e. le concepteur prend en main la preuve. Dans notre cas, le concepteur n'a fourni aucun effort puisque celui-ci s'est contenté de faire appel au démonstrateur de prédicats. Ce démonstrateur, uniquement disponible dans l'environnement de preuve interactive, est néanmoins entièrement automatique. Ces 3 obligations de preuve ayant été démontrées uniquement par l'utilisation du démonstrateur de prédicats, nous pouvons donc conclure que 100% des preuves ont été démontrées automatiquement.

Modèle	Obv	nOP	Auto	Interactif	%Pr
IMappy	55	10	10	0	100
IMappy_ref1	176	22	22	0	100
IMappy_ref2	142	27	24	3	100
Total	373	59	56	3	100

Table II. Résultat du nombre de preuves effectuées pour l'étude de cas

8.5 Discussion entre techniques fondées sur la preuve et vérification sur modèles

Le raffinement utilisé par B événementiel permet une modélisation par décomposition qui offre l'avantage d'introduire de nouvelles descriptions à la spécification. La totalité du système est ainsi bâtie de proche en proche. Le raffinement permet d'une part de définir les propriétés à prouver et d'autre part de conserver les propriétés déjà établies.

Notre approche se démarque des techniques de Model Checking qui procèdent par composition (SMV [McMillian 1992; Abowd et al. 1995], réseaux de Petri [Palanque et al. 1995; Navarre et al. 2005], XTL [Brun 1997], flots de données [D'Ausbourg 1998]) où les propriétés sont validées sur des systèmes complexes (parcours exhaustifs, graphes de marquage).

De plus l'utilisation d'un nombre arbitraire permet de réaliser des preuves inductives dont la complexité ne dépend pas de la valeur de ce nombre. Cette démarche évite le problème de l'explosion combinatoire que rencontrent des techniques fondées sur les parcours exhaustifs où la construction de graphes de marquage. Cependant cette approche n'est pas complètement automatique du fait de la présence possible de preuves interactives. C'est pourquoi nous préconisons l'utilisation conjointe des deux types de technologies dans une démarche intégrée.

9. CONCLUSION

Cet article a montré qu'il était possible de conduire le développement formel d'une IHM3. Des modèles événementiels formels sont décrits et raffinés de proche en proche. Chaque raffinement introduit de nouvelles informations et de nouvelles propriétés. Ce travail est fondé sur la notion de tâches avec la notation CTT ainsi que les propriétés CARE permettant de qualifier une IHM3. Nous n'avons pas proposé de nouvelle notation ni de nouvelle technique. Nous nous sommes contentés de fournir une assistance formelle à un développeur d'IHM3 souhaitant établir a priori des propriétés de l'interface en cours de conception.

Par ailleurs, cet article a montré qu'à l'image des architectures logicielles proposées pour les IHM3, il est également possible de respecter le critère de modularité dans une conception utilisant une méthode formelle. Nous avons décrit le système abstrait dans sa globalité, à un haut niveau d'abstraction (deux événements seulement), ensuite nous avons dirigé nos développements vers la multi-modalité en entrée. Cette modularité permet d'alléger le processus de développement et de preuve. La définition de schémas de représentation d'opérations de CTT permet à des non spécialistes de produire ces modèles ; néanmoins la preuve interactive restera à la charge des concepteurs.

Actuellement, nous étudions la possibilité de tester et d'animer les modèles B obtenus afin de pouvoir conduire une expérimentation à un haut niveau d'abstraction. Cela permettrait de compléter la validation des tâches par une activité d'animation [Aït-Sadoune and Aït-Ameur 2008]. Nous nous intéressons également à la modélisation de la multi-modalité en sortie [Mohand-Oussaïd et al. 2010].

RÉFÉRENCES

- ABOWD, G., WANG, H.-M., AND MONK, A. 1995. A Formal Technique for Automated Dialogue Development. In *Proceedings of DIS'95*, G. Olsan and S. Schuon, Eds. 219–226.
- ABRIAL, J.-R. 1996a. Extending B without changing it (for developing distributed systems). In *First B Conference, Putting Into Practice Methods and Tools for Information System Design*, H. Habrias, Ed. Nantes, France.
- ABRIAL, J.-R. 1996b. *The B Book. Assigning Programs to Meanings*. Cambridge University Press.

- ABRIAL, J.-R. 2010. *Modeling in Event-B: System and Software Engineering*, Cambridge ed. Cambridge University Press.
- ABRIAL, J.-R. AND HALLERSTEDTE, S. 2007. Refinement, Decomposition, and Instantiation of Discrete Models: Application to Event-B. *Fundamenta Informaticae* 77, 1–28.
- AÏT-AMEUR, Y., AÏT-SADOUNE, I., AND BARON, M. 2006. Etude et comparaison de scénarios de développements formels d’interfaces multimodales fondés sur la preuve et le raffinement. In *6ème Conférence Francophone de Modélisation, Optimisation et Simulation des Systèmes : Défis et Opportunités (MOSIM’06)*, Lavoisier, Ed. Rabat, Maroc, 578–588.
- AÏT-AMEUR, Y., AÏT-SADOUNE, I., BARON, M., AND MOTA, J.-M. 2006. Validation et vérification formelles de systèmes interactifs multimodaux fondés sur la preuve. In *18^e Conférence Francophone sur l’Interaction Homme-Machine (IHM’2006)*. Vol. 1. ACM Press, Montréal, Canada, 123–130.
- AÏT-AMEUR, Y., AÏT-SADOUNE, I., BARON, M., AND MOTA, J.-M. 2008. Développements formels d’interfaces multimodales fondés sur la preuve et le raffinement. Scénarios de développement. *RSTI série ISI Ingénierie des Systèmes d’Information : Modélisation multiple, Formalisme et Modèles* 13, 2, 127–154.
- AÏT-AMEUR, Y., BARON, M., AND GIRARD, P. 2003. Formal validation of hci user tasks. In *IEEE - The 2003 International Conference on Software Engineering Research and Practice - SERP 2003*, A.-A. Ban, A. H.R., and M. Youngsong, Eds. Vol. 2. CSREA Press, Las Vegas, Nevada USA, 732–738.
- AÏT-AMEUR, Y., BARON, M., KAMEL, N., AND MOTA, J.-M. 2009. Encoding a process algebra using the Event B method. *Software Tools and Technology Transfer. Springer Verlag* 11, 3, 239–253.
- AÏT-AMEUR, Y., BRÉHOLÉE, B., GIRARD, P., GUITTET, L., AND JAMBON, F. 2004. Formal verification and validation of interactive systems specifications. from informal specifications to formal validation. In *Conference of Human Error, Safety and Systems Development, HESSD, Toulouse*.
- AÏT-AMEUR, Y., GIRARD, P., AND JAMBON, F. 1998. Using the b formal approach for incremental specification design of interactive systems. In *Engineering for Human-Computer Interaction*, S. Chatty and P. Dewan, Eds. Vol. 22. Kluwer Academic Publishers, 91–108.
- AÏT-SADOUNE, I. AND AÏT-AMEUR, Y. 2008. Animating Event B Models by Formal Data Models. In *ISOLA 2008 - 3rd International Symposium on Leveraging Applications of Formal Methods, Verification and Validation*. Porto Sani, Greece.
- BACK, R.-J., FAN, X., AND PREOTEASA, V. 2003. Reasoning about pointers in refinement calculus. In *Proceedings of the Tenth Asia-Pacific Software Engineering Conference Software Engineering Conference*. IEEE Computer Society, 425.
- BACK, R.-J. AND SERE, K. 1989. Stepwise refinement of action systems. In *Proceedings of the International Conference on Mathematics of Program Construction, 375th Anniversary of the Groningen University*. Springer-Verlag, 115–138.
- BARON, M. 2003. Vers une approche sûre du développement des interfaces homme-machine. Ph.D. thesis, Université de Poitiers.
- BARON, M., LUCQUIAUD, V., AUTARD, D., AND SCAPIN, D. 2006. K-MADE : un environnement pour le noyau du modèle de description de l’activité. In *18ème Conférence Francophone sur l’Interaction Homme-Machine (IHM’2006)*, A. Press, Ed. Montréal, 287–288.
- BASS, L., PELLEGRINO, R., REED, S., SHEPPARD, S., AND SZCZUR, M. 1991. The Arch Model : Seeheim Revisited. In *CHI 91 User Interface Developer’s Workshop*.
- BELLIK, Y. 1995. Interfaces Multimodales : concepts, modèles et architecture. Ph.D. thesis, LIMSI- Université d’Orsay.
- BJORNER, D. 1987. Vdm a formal method at work. In *Proc. of VDM Europe Symposium’87*, S.-V. LNCS, Ed.
- BOUCHET, J., NIGAY, L., AND GANILLE, T. 2004. Icare software components for rapidly developing multimodal interfaces. In *Proceedings of ACM-CHI 2004, Extended Abstracts*. ACM Press, Vienna, Austria, 1325–1328.
- BRUN, P. 1997. *XTL: a temporal logic for the formal development of interactive systems*. Springer-Verlag, 121–139.
- BUMBULIS, P., ALENCAR, P., COWAN, D., AND LUCENA, C. 1996. Validating Properties of Component-Based Graphical User Interfaces. In *Eurographics Workshop on Design, Specification, and Verification of Interactive Systems (DSV-IS’96)*. Springer Verlag, 347–365.
- BURBECK, S. 1992. Application programming in Smalltalk-80: How to use the Model-View-Controller (mvc). Report, University of Illinois in Urbana-Champaign (UIUC).

- CAMPOS, J. AND HARRISON, M. 1997. Formally Verifying Interactive Systems: A Review. In *Eurographics Workshop on Design, Specification, and Verification of Interactive Systems (DSV-IS'97)*. Springer Verlag, 109–124.
- CANSELL, D. 2003. Assistance au développement incrémental et à sa preuve. Ph.D. thesis, Habilitation à diriger les recherches, Université Henri Poincaré.
- COUTAZ, J. 1987. Pac, an implementation model for the user interface. In *IFIP TC13 Human-Computer Interaction (INTERACT'87)*. North-Holland, Stuttgart, 431–436.
- COUTAZ, J., NIGAY, L., SALBER, D., BLANDFORD, A., MAY, J., AND YOUNG, R. 1995. Four easy pieces for assessing the usability of multimodal interaction: the CARE properties. In *Proceedings of Human Computer Interaction - Interact'95*. Chapman and Hall, 115–120.
- D'AUSBOURG, B. 1998. Using Model Checking for the Automatic Validation of User Interface Systems. In *Eurographics Workshop on Design, Specification, and Verification of Interactive Systems (DSV-IS'98)*. Springer Verlag, 242–260.
- DEPAULIS, F., JAMBON, F., GIRARD, P., AND GUITTET, L. 2006. Le modèle d'architecture logicielle H4 : Principes, usages, outils et retours d'expérience dans les applications de conception technique. *Revue d'Interaction Homme-Machine (RIHM)* 7, 1, 93–129.
- DIJKSTRA, E. W. 1976. *A Discipline of Programming*. Prentice-Hall, INC.
- DIX, A., FINLAY, J., ABOARD, G., AND BEALE, R. 1993. *Human-Computer Interaction*. Prentice Hall.
- DUKE, D. AND HARRISON, M. 1993a. Abstract Interaction Objects. In *Proceedings of Eurographics conference and computer graphics forum*. Vol. 12. 25–36.
- DUKE, D. AND HARRISON, M. 1993b. Towards a Theory of Interactors. Tech. rep., Amodeus Esprit Basic Research Project 7040, System Modelling/WP6.
- DUKE, D. AND HARRISON, M. D. 1995. Event Model of Human-System Interaction. *IEEE Software Engineering Journal* 10, 1, 3–10.
- FACONTI, G. P., ZANI, N., AND PATERNÒ, F. 1992. The input model of standard graphics systems revisited by formal specification. *Comput. Graph. Forum* 11, 237–251.
- GRAY, P., ENGLAND, D., AND MCGOWAN, S. 1994. XUAN: Enhancing the UAN to Capture Temporal Relation Among Actions. Tech. rep., Department of Computing Science, University of Glasgow.
- GUITTET, L. 1995. Contribution à l'ingénierie des interfaces homme-machine - théorie des interacteurs et architecture h4 dans le système nodao. Ph.D. thesis, Université de Poitiers.
- HARRISON, M. AND THIMBLEBY, H. 1990. *Formal Methods in Human-Computer Interaction*. Series on Human computer interaction. Cambridge University Press.
- HIX, D. AND HARTSON, H. R. 1993. *Developing User Interfaces : Ensuring Usability Through Product and Process*. Wiley Professional Computing.
- HOARE, C. 1969. An Axiomatic Basis for Computer Programming. *CACM* 12, 10, 576–583.
- JAMBON, F., GIRARD, P., AND AÏT-AMEUR, Y. 2001. Interactive system safety and usability enforced with the development process. In *Engineering for Human-Computer Interaction (8th IFIP International Conference, EHCI'01, Toronto, Canada, May 2001)*, R. M. Little and L. Nigay, Eds. Lecture Notes in Computer Science, vol. 2254. Springer, Canada, 39–55.
- KOLSKI, C., FORBRIG, P., DAVID, B., GIRARD, P., CHI DUNG, T., AND EZZEDINE, H. 2009. Agent-based architecture for interactive system design: current approaches, perspectives, evaluation. In *Human-Computer Interaction, Part I, HCI 2009 HCI International 2009*. Number LNCS 5610. Springer-Verlag, San Diego États-Unis d'Amérique, 806–815.
- LUCQUIAUD, V. 2005. Sémantique et outil pour la modélisation des tâches utilisateur : N-MDA. Ph.D. thesis, Université de Poitiers / ENSMA / INRIA.
- MCMILLIAN, K. 1992. The smv system. Tech. rep., Carnegie Mellon University.
- MILNER, R. 1980. *A Calculus of Communicating Systems*. Springer Verlag.
- MOHAND-OUSSAÏD, L., KAMEL, N., AÏT-SADOUNE, I., AÏT-AMEUR, Y., AND AHMED-NACER, M. 2010. *Un cadre formel pour la conception et la validation de systèmes interactifs multimodaux dans le secteur du transport*, In *IHM et Transports terrestres*, Kolski, C ed. Hermes, Paris., à paraître.
- NAVARRÉ, D., PALANQUE, P., BASTIDE, R., SCHYN, A., WINCKLER, M., NEDEL, L., AND FREITAS, C. 2005. A formal description of multimodal interaction techniques for immersive virtual reality applications. In *Proceedings of INTERACT 2005*. Lecture Notes in Computer Science, Springer Verlag, Roma, Italy.

- NIGAY, L., F.JAMBON, AND J.COUTAZ. 1995. Formal specification of multimodality. In *CHI'95 Workshop on Formal Specifications of User Interfaces*. ACM Press, Denver (Colorado), USA.
- PALANQUE, P., BASTIDE, R., AND SENGÈS, V. 1995. Validating interactive system design through the verification of formal task and system models. In *IFIP TC2/WG2.7 Working Conference on Engineering for Human-Computer Interaction (EHCI'95)*, L. J. Bass and C. Unger, Eds. Chapman & Hall, Grand Targhee Resort (Yellowstone Park), USA, 189–212.
- PATERNÒ, F. 2001. *Model-Based Design and Evaluation of Interactive Applications*. Springer.
- PATERNÒ, F., MORI, G., AND GALIMBERTI, R. 2001. Ctte: An environment for analysis and development of task models of cooperative applications. In *ACM CHI 2001*. Vol. 2. ACM/SIGCHI, Seattle.
- PFAFF, G. 1985. *User Interface Management Systems*. Eurographics seminars. Springer Verlag - Berlin.
- ROCHE, P. 1998. Modélisation et validation d'interface homme-machine. Ph.D. thesis, École Nationale Supérieure de l'Aéronautique et de l'Espace.
- SCAPIN, D. L. AND PIERRET-GOLBREICH, C. 1989. Mad : Une méthode analytique de description des tâches. In *Colloque sur l'Ingénierie des Interfaces Homme-Machine (IHM'89)*. Sophia-Antipolis, France, 131–148.
- SCAPIN, D. L. AND PIERRET-GOLBREICH, C. 1990. Towards a Method for Task Description : MAD. In *Work with display units*. Elsevier Science Publishers, North-Holland.
- SPIVEY, J. M. 1992. *The Z notation: A Reference Manual*. Prentice Hall International (UK) Ltd.

Using K-MADe for learning task modeling: interests and difficulties

Sybille Caffiau, Dominique L. Scapin, Loé Sanou, Laurent Guittet

► **To cite this version:**

Sybille Caffiau, Dominique L. Scapin, Loé Sanou, Laurent Guittet. Using K-MADe for learning task modeling: interests and difficulties. Journal d'Interaction Personne-Système, Association Francophone d'Interaction Homme-Machine (AFIHM), 2010, 1 (1), pp.1-28. hal-01058930

HAL Id: hal-01058930

<https://hal.inria.fr/hal-01058930>

Submitted on 28 Aug 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Using K-MADe for Learning Task Modeling: Interests and Difficulties

SYBILLE CAFFIAU

LISI, ENSMA Futuroscope Chasseneuil

DOMINIQUE L. SCAPIN

INRIA Domaine de Voluceau

LOE SANOU

LAURENT GUITTET

LISI, ENSMA Futuroscope Chasseneuil

Résumé : L'enseignement de la conception de logiciel centrée utilisateur concerne de nombreux aspects. Un de ces aspects les plus importants est la modélisation des tâches. La modélisation des tâches contribuant fortement à la qualité ergonomique et à l'acceptation du logiciel résultant, il est indispensable qu'elle soit la mieux comprise possible par les futurs concepteurs de logiciels. Dans ce but, cette étude cherche à évaluer un modèle de tâche particulier (K-MAD) et son outil associé (K-MADe) dans une démarche d'enseignement de la modélisation des tâches. Cet article décrit la problématique de l'étude, le modèle et le logiciel enseigné, les différentes phases de l'enseignement, et le retour d'expérience d'utilisation concrète du logiciel. Ce retour d'expérience identifie les apports et les inconvénients liés à cette utilisation, lors de son enseignement, et propose des pistes d'évolution des modèles obtenus et d'édition de ces modèles.

Mots clés : Interaction Homme-Machine, Conception centrée-utilisateur, Modèles de tâches, Analyse hiérarchique des tâches, Outils de modélisation, Enseignement, Évaluation.

Abstract: Teaching user-centred software design covers many aspects. One of the most important ones is task modeling. Since task modeling contributes largely to the ergonomic quality and acceptance of the resulting software, it is essential for task modeling concepts to be well understood by future software designers. To this end, this study aims at evaluating a task model (K-MAD) and its associated tool (K-MADe) as regards the task modeling training. This article describes the rationale of the study, the model and the software tool used, the teaching steps, and the experience feedback on the practical use of the software. The latter identifies the benefits and disadvantages of this practical use from the teaching point of view, but also from a practical standpoint, including in terms of evolution of the models obtained and of their edition.

Key words: Human-Computer Interaction, User-centered design, Task model, Hierarchical task analysis, Modelled tools, Teaching, Evaluation.

1. INTRODUCTION

Nowadays, taking the user into account is an essential need for interactive application design [Diaper 2004]. Satisfying this need can be achieved partly by offering appropriate methods for training students (future HCI designers) particularly in the area of task modeling. Students have been taught how to model the human activities for several years. However, the design of task models that express activities within software application is a complex task. Moreover, the duration usually assigned to this specific training is limited. These two reasons require looking for techniques that are both non time-consuming and pedagogically beneficial for students.

In order to conduct such teaching activities, dedicated software has been used for several years in order to support task model training in our university. Among the user-oriented task models that are supported by freely available tools, two models have been used: CTT [Paternò, et al. 1997] (with CTTE as the tool); and since 2007, K-MAD [Lucquiaud 2005] (with K-MADe as the tool [Baron, et al. 2006]). The goal of this paper is neither to explain this choice, nor to compare the use of these two formalisms but to present an assessment of the use of one particular task model tool to support the learning of task modeling. This study was conducted using the K-MADe software. The main reason for that choice is that K-MAD is a task model *kernel* composed of all main concepts essential to design a complete task model (i.e., common to most existing task models). The other reason is that current research on K-MADe is the assessing of the model (e.g. according to new application domains) and the developing of the tool. Thus, this study, through the assessment of the task modeling learning process, aims at producing data useful for improving the current support software.

In this paper, the first section describes the model being taught as well as its associated tool. The following sections present the goals of the study, the participants, the experimental phases, and the evaluation method. Then, the results are presented with data collection and analysis. Finally, the conclusion summarizes the evaluation and identifies future work for the evolution of K-MADe.

2. THE MODEL AND THE TOOL

This first section describes the K-MAD model and the K-MADe tool, which were used during the learning sessions.

2.1 The K-MAD model

The K-MAD task model is hierarchical. It represents the user's activity as a tree of tasks, from the most general level (root) to the most detailed level (elementary actions). In order to express activities, the K-MAD model consists of components and scheduling characteristics.

In the K-MAD model, a task is identified with a name, a goal and an identification number. In addition, a task is characterized by several elements: duration, feedback, executant (user, system, interactive or abstract), modality (when it is user-performed), frequency, importance level, triggered events and additional comments, when applicable.

The task scheduling characteristics are: interruptability, mandatory (or not), iteration (number or expression that sets the number of times the task is performed), pre condition, event (which constrains the task execution) and object modifications, named post conditions (which express the object dynamics).

The objects that are defined are used to characterize the pre conditions, the iterations and the object manipulation expressions.

The scheduling decomposition of parent tasks is also specified. The task schedules its subtasks either sequentially, concurrently, alternatively or without any order. When a task is not split further, it is defined as an elementary task.

2.2 The K-MADe Tool

K-MADe [Baron, et al. 2006] allows to describe, modify and query the task models (queries are under development).

K-MADe implements the above characteristics of the model as well as additional characteristics such as the *unknown* executant type for more tool flexibility.

Although all kinds of profiles can use this tool, it is particularly intended for ergonomists and specialists in HCI. As there are many skill variations in such populations, the tool allows different levels of description, from simple graphical drawings to detailed mathematical expressions using several available tools:

- A graphic editor of the K-MAD task model, which uses direct manipulation techniques for constructing, handling and canceling tasks (label 1 in Figure 1).
- A task characteristic editor. The editor has three forms. A reduced form with only the main task characteristics; (Figure 2 a) a form which lists all task characteristics in tables (label 2 in Figure 1); and finally a printable form with all task characteristics (Figure 2 b).
- An editor of abstract objects, users, events and concrete objects. Objects can be added, modified and removed. The modification and the removal of an object imply modifications of all associated objects. *Sheets* (label 3 in Figure 1) allow to access these object definition spaces.
- A tool for expressing the pre conditions, post conditions and iterations. The tool checks whether the grammar of an expression is correct or not (Figure 3; and later on Figure 11).
- A simulator for running and animating models of tasks (Figure 8).
- Various tools for analyzing task models (statistical (Figure 7a), coherence (Figure 7b), queries,...).
- A tool for printing task trees and detailed task characteristics.

Tools to check the task model grammar (Figure 7b), to simulate (Figure 8) and to analyze the task models are triggered from the *tool* menu (encircled in Figure 1). More information about the K-MADe tool may be found in the user manual [K-MADe].

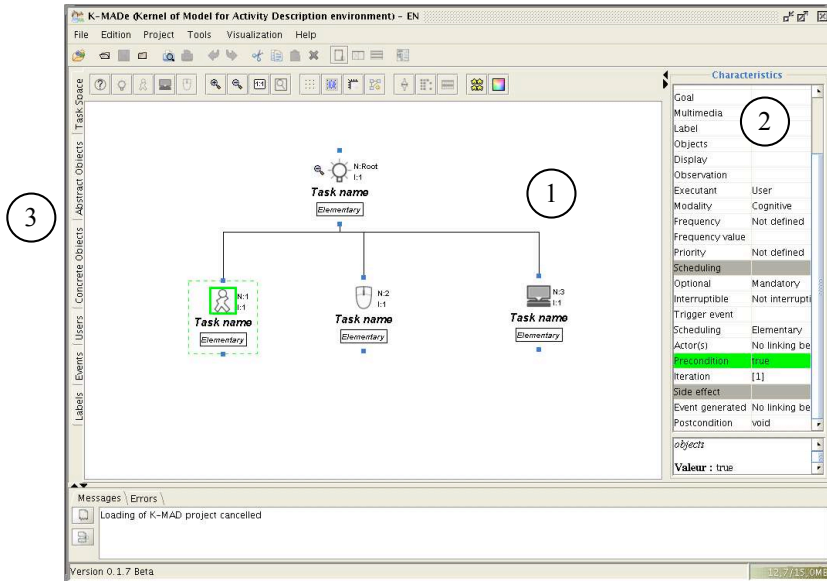


Fig. 1. The main window of K-MADE tool

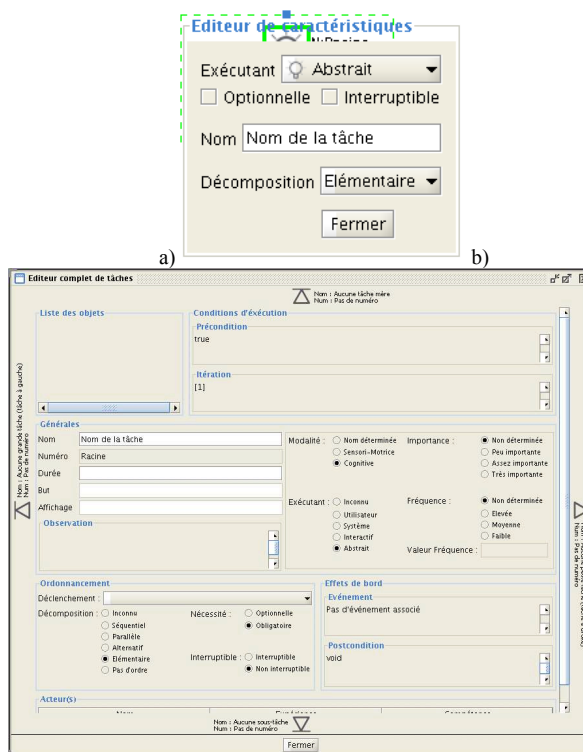


Fig. 2. Two other task editors

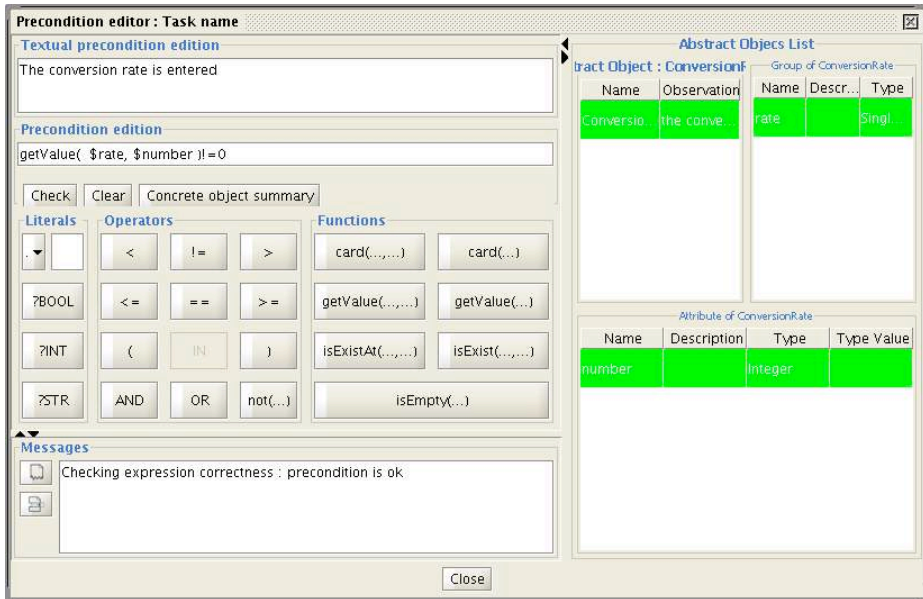


Fig. 3. The window to edit a task precondition

3. OBJECTIVES

Data gathered in this evaluation aims at analyzing the role played by the K-MADe tool during the K-MAD model learning process. We identified three steps that characterize the learning process: the *modeling discovery*, the *model appropriation* and, once the model is learned, the *model use* via K-MADe (the tool).

For the *modeling discovery* phase, we wished to analyze the students' motivation and willingness to use the software when performing task modeling. We then wished to study the consequences of using K-MADe on the model appropriation in terms of concepts. In particular, the issue was whether the software adds complexity to the task analysis, or helps with the learning.

Lastly, once students learned the model, we wanted to evaluate how they use K-MADe in order to design their own task models; which K-MADe functions are used; how they define concepts (when, what and why), which ones are assimilated and to what extent K-MADe does help in that regard.

4. PARTICIPANTS

Two groups of students took part in the study. All subjects were students in their fourth year at a French university. They were all bio-informatics students (Master Génie Physiologique et Informatique: <http://www-gphy.univ-poitiers.fr>). The first group was composed of the 48 students (31 males and 17 females) from the 2007/2008 school-year, and the second one of the 50 students of the 2008/2009-school-year (28 males and 22 females). Participants were aged from 21 to 25 years old. Both years were taught the same HCI program. Participant characteristics are presented in Table I.

Table I. Participant characteristics

	2007/2008 school-year group	2008/2009 school-year group
Male	31	28
Female	17	22
Number	48	50
Formation	bio-informatics students in 4 th year at a French University	
Aged	from 21 to 25 years old	

5. STUDY SCHEDULE

The study of the K-MADe role in the learning of task modeling took place in two years. Each year, the HCI course was composed of two phases: the lecture and the practical sessions.

The lecture part presented the basic principles of user-centered design and task modeling. During this lecture, the K-MAD model was detailed and illustrated with K-MADe tool captures. Then, during the practice courses, the students used the K-MADe tool to express several activities (i.e. sets of tasks). Each year, students followed two practice training sessions. In addition, in the first year, a session was added in order to perform a specific evaluation step. During this session, students were observed when they practiced task modeling without any help.

Figure 4 summarizes the different sessions for each group according to the task modeling steps and the evaluation year: the initial lecture (Figure 4), the practicing following the HCI course composed of training course sessions (Figure 4) and the *usage* evaluation session (Figure 4).

Evaluation data was only collected in practice course sessions. In terms of timing, the evaluations of both student groups' studies were one year distant: the contents of the second-year learning phase was modified in order to complete the first year data.

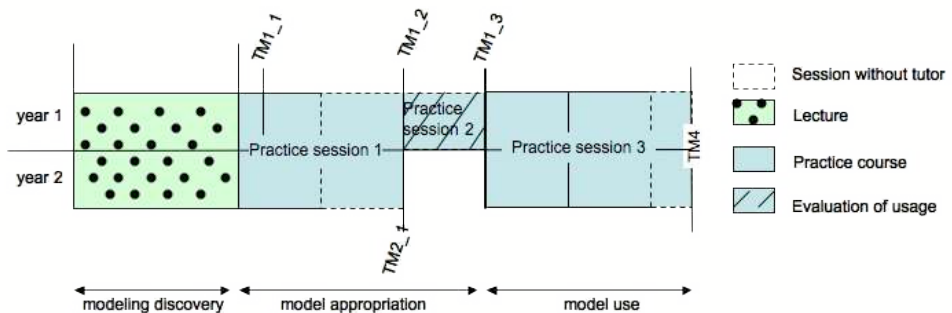


Fig. 4. Schedule of task model learning steps

5.1 The Lecture

The HCI course focused on user-centered design, and task modeling was a part of it. Task models [Balbo, et al. 2004, Limbourg and Vanderdonck 2004] may be used to analyse systems, and express the different activities that the user wishes to perform when using the software [Diaper and Stanton 2004]. These models are designed in order to reach several goals [Balbo, et al. 2004], such as to help during the design step, to help

during the validation step [Mori, et al. 2004], or to generate an application component. The lecture on the task modeling process did not aim at teaching all task models; it only focused on the K-MAD model. This model was initially explained in details during the lecture (nearly 4 hours), then the students practiced task modeling using K-MADe during the training courses that followed (§ 5.2). Even if they were not modeling experts, they were more extensively trained on how to use a task model notation than on actual ergonomics expertise [Couix 2007].

The second part of the lecture was focused on various evaluation concepts and on the main methods used [Nielsen 1993]. This survey of concepts and methods served also as background for the students of the first school-year, who played the role of observers in the *usage* evaluation session. Even though their participation was limited to observing and taking notes (the notes were completed by other data sources), the presentation of evaluation concepts aimed at facilitating the student evaluation work (this study was their first practical evaluation).

Both year courses concerned the same concepts and followed the same schedule.

5.2 Practical Training Step

The practical training step was composed of two training courses sessions (without any pattern in Figure 4) and one session of usage evaluation (with the hachured pattern in Figure 4). The first and third sessions were similar for both groups, except for the number of task models to design. However, the second practice session (the *usage* evaluation) was only performed by the first school-year group. In this section, we describe the exercises performed during these three sessions as well as the corresponding instructions.

5.2.1 First practice session, first school-year group. During the first practice training session, the first year group had to design two different task models. The first one described their activity before arriving at the university each morning (TM1_1). Few concepts were used to design this task model. They had to decompose this activity without defining objects or conditions. The training target of this exercise was to show students the capacity of K-MAD to model the different scenarios that were anticipated by each student. In order to design this first task model, students could use either the K-MADe tool or simply paper and pencil. After this first modeling, students used K-MADe as a task model editor.

The second task model described the activity of a clerk in a car rental office. In the design of that second task model (TM1_2), all K-MADe concepts were used: tasks, objects, events and users.

Tasks. In a car rental office, the clerk welcomes visitors, establishes contracts, receives cars that come back from rental or from repairing and last does what is needed for the damaged cars. Concurrently to these activities, the clerk answers the phone when it rings. Tasks performed by a car rental clerk may be expressed in the form of task trees such as in Figure 5.

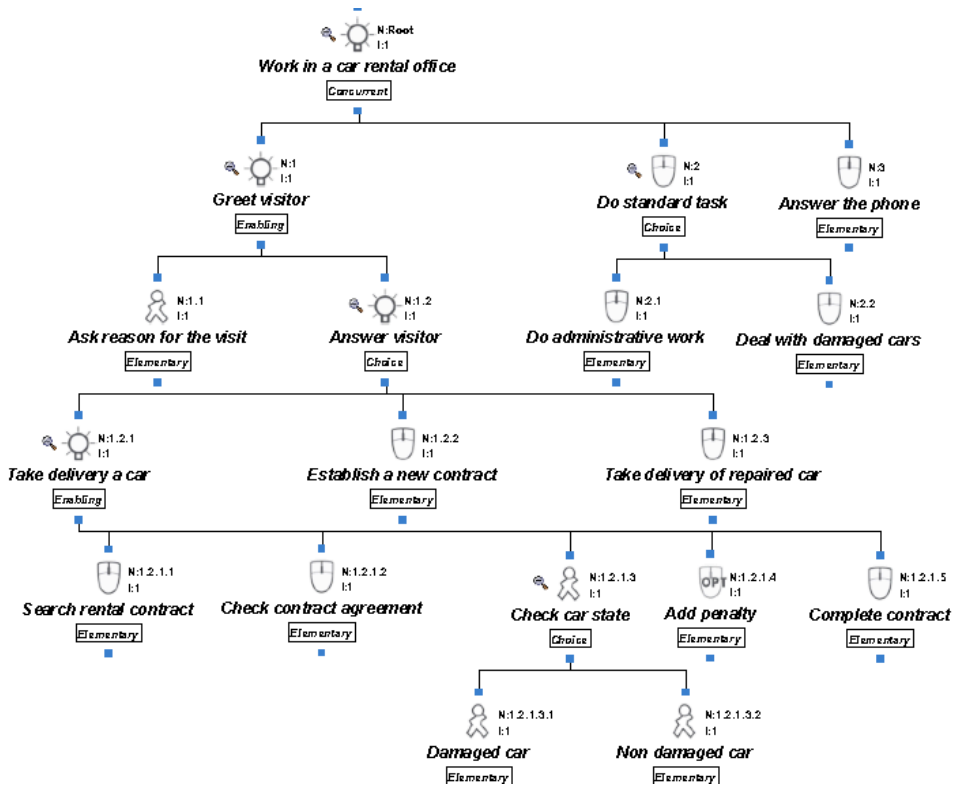


Fig. 5. K-MAD task tree of car rental clerk work

Objects. Two objects were handled in this activity: *car* and *contract*. A car is defined by a color (String) and its production year (Integer) and a contract concerns a client (String) and a car (String), and contains a price (Integer) and a rental time (Integer). K-MAD objects are stored into groups. A car and a contract may be member of different groups according to their state: available, rent, damaged, repaired for a car and new, current or past for a contract. Table II summarizes the characteristics of objects and of groups that store them.

Table II. Car rental Objects

Object Name	Object attributes		Object groups		
	Name	Type	Name	Description	Type
Car	color	String	AvailableCar	Cars that can be rented	Set
			RentCar	Cars not available	Set
			CarInReception	The car currently observed	Singleton
	year	Integer	DamagedCarTo	Damaged cars	List

				that need to be repaired	
			DamagedCarAtGarage	Cars at garage	Set
Contract	client	String	CurrentContract	The contract currently in edition	Singleton
	car	String	NewContract	Contracts for a new client	Singleton
	time	Integer	Store	Contracts stored in the car rental office	Set
	price	Integer			

The defined objects are used to express computable expressions. For example, the post-condition of the *Search rent contract* task is that a stored contract comes back to be edited (*replace(\$Store,\$CurrentContract)*) and its computing allows the *Complete contract* task execution whose pre-condition is “a contract is in edition” (*Card(\$CurrentContract)*).

Events. Only one event is modeled for this activity: the phone ringing. This event triggers the *Answer the phone* task.

Users. Three different users are associated to the task tree presented in Figure 5: the clerk woman, the garage and a client.

To design that task model, the students were paired and they did not have any time limit, after the end of the practice step (the duration averaged about two hours). During the first hour and half, a tutor was present and answered the student’s questions. When designing this task model, the students were able to explore the K-MADE tool.

5.2.2 First practice session, second school-year group. During this first task modeling session, the second group of students had to perform only one task model (TM2_1). This task model expresses the activities they might perform with a calendar (such as editing an event, synchronizing a calendar on a computer and on a phone, for example). All K-MAD concepts were mandatory. In order to limit the initial difficulties of the K-MADE use for the students, they began their task model without K-MADE (by using paper-pencil) during two hours, in order to focus on the concepts rather than on the tool. After that, they completed their task model using K-MADE (nearly two hours with the K-MADE tool). As for the first year group session (described in 5.2.1), students were only helped at the beginning of the task model design process and they completed their models alone.

5.2.3 Second practice session (first school-year group). During the *usage* evaluation session (the second practice session), students had to model the activity of completing a volley-ball game marking sheet (TM1_3) alone, without any tutor help. Instructions for this activity were given at the beginning of the session. The instructions were extracted from the official instructions from the French Federation of Volley Ball (FFVB), and two examples of marking sheets (completed and non-completed ones) were provided. This task model design was limited in time: one hour and thirty minutes.

5.2.4 Third practice session (both school-year groups). Finally, to design the last task model (TM4), students could ask questions to tutors, they modeled activities of members of a genetic analysis laboratory (technicians, secretary, manager...). From this task model, they had to design an application for the laboratory

(<http://www.genindexe.com/>). They worked by groups of three or four students to perform this last modeling exercise.

6. EVALUATION METHOD

This evaluation aimed to analyze the use of K-MADe during the whole task model learning process. In order to perform this assessment, we used two classical evaluation techniques [Nielsen 1993]: observation of subjects when using the tool and consultation of users by analyzing questionnaires.

A two level observation was conducted: an observation of a whole group (global observation) and an observation of the behavior of each student (individual observation). The individual observation of all students focused on evaluating the usage of the K-MADe formal components.

In the next section, we describe the experimental procedure.

6.1 Global Observation

In the first year, after each session, the tutors noted their observations: the difficulties in understanding (wordings and/or concepts) were precisely recorded. The tutors also noted the student behavior (reactions/ comments/ questions). These *global notes* aimed at helping the interpretation of other gathered data (such as the task models produced). They contained two types of information: the exercise (or the question) and the comments about it.

During the whole learning process, the tutors noted also the individual queries. Those concerned the individual activities such as checking the models or difficulties in using the tool. They were either orally asked or sent by email. As during the training, each individual query and correction was explained to the whole group, *individual note* is considered as global data.

In addition to the common note data (observation reference and comment), this type of notes (*individual notes*) included the reference of the observation author (student name or group). Due to individual information, the *personal notes* were added to the *global observation* and to the *individual observation*. The tutors' notes played an important information role and completed the understanding of other gathered data. Moreover, during the sessions, tutors collected specific data on evaluation topics.

6.2 Individual Observations

The individual observations of student behavior when using K-MADe to design their task models required a particular method. These observations were set up during the second practice session (Figure 4). All students of the first school year were paired. During the first half part of the session, one student acted as a task model designer (using K-MADe, named *user*), while the second acted as the evaluation expert (named *observer*). They reversed roles during the second part. The tutors assigned roles played by each student in order to be sure that all the student levels were represented. Each session lasted one hour and a half with a fifteen minute break between sessions. The activity that was to be modeled was the same for all students and it was described at the beginning of the sessions.

6.2.1 Observers' notes. During the modeling, the *observer* ensured that the user verbally described his/her modeling process, and noted what s/he observed concerning the use of the tool by his/her *user* (hesitations, exploration of several parts of the

software without actions, etc.). In order to help observers in their evaluation, they were given observation sheets (illustrated in Table III). These sheets consisted mainly in a three-column table corresponding to the three types of information recorded for each observation:

- The type of the observation among a set of defined categories (user goal (G), tool functionalities (F), functionality utilization (FU) and information (I)).
- The observation in textual form.
- The time indication about the observation.

In order to complete the observers' notes and the designed task model, another type of data was used: the user-logs.

Table III. Observation sheet example

Type	Observation	Time
FU	The main window is not accessible ("simulation" is noted but the simulation window is not accessible either). => launch K-MADE again	14h32
G	looking for the object definition	14h34 14h37
F	user does not understand the meaning of the button with shell-hole	14h40

6.2.2 User-Logs. To complete the observations carried out during the *usage* evaluation (practice session 2 of the first school year), the students used a version of K-MADE that produced user-logs. This version allowed keeping track of the user's actions using timestamps and produced a user-log as a text-file. Specifically, the log indicated when the user entered and exited each K-MADE space: task space, abstract objects, condition edition windows (pre, post and iteration), etc. Figure 6 shows an example of such information. User-logs were not generated in order to be used to replay user actions. Evaluators used them in order to deduce a mental representation of student task modeling (in order to complete observation notes).

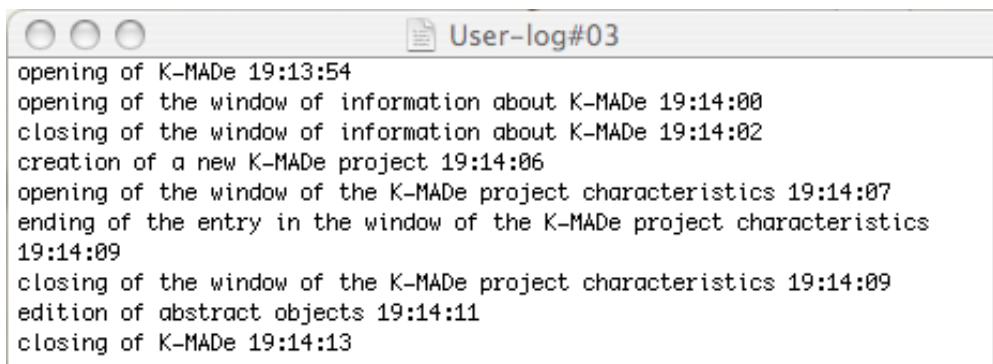


Fig. 6. A user-log produced using K-MADE

6.3 Questionnaires

Results of the first school-year evaluation allow us to point out several questions on the task modeling learning process. In order to answer these specific questions, we defined two questionnaires. The students of the second school-year were asked to complete them as regards the use of the tool. They completed the first one at the end of the *model appropriation* step (once TM2_1 was designed) and the second one at the end of the *model use* step (once TM4 was designed). These questionnaires (presented in the Annex part) were composed of questions on the use of the K-MADE tool as a way to learn task modeling. Three questions were common to both questionnaires in order to analyze the progression of students' answers.

7. EVALUATION DATA

Since each session tried to reach different goals, we did not collect the same data for each experiment. In this section, we present the results for each session. Table IV summarizes the data gathered according to the session. In the second part of this section, the selection of data included in the analysis (Table V) is explained.

7.1 Data collection

What are the benefits of using a tool at the beginning of the model appropriation? During the first practice session of the first year, the goal was to understand the benefits of the tool at the beginning of the *model appropriation*. In order to analyze this first point, we counted the cases where the students chose to use the K-MADE tool to design their first task model (TM1_1). Then the tutors noted the distribution between students.

How did students design their task models? What K-MADE functionalities were used to perform task modeling? The knowledge of the task modeling process requires to collect data for the design of the three task models (TM1_2, TM2_1 and TM1_3) as well as the answers to the questionnaires (by focusing on specific items). The TM1_2 task model allowed the students of the first year to become familiar with the K-MADE tool in order to guarantee that the design of the TM1_3 task model is independent from the tool discovery step.

The second practice session (designing TM1_3) aimed at analyzing the task modeling process, particularly when the K-MADE entities were defined and when they were manipulated. In order to obtain this information, we used user-logs and notes from the observers. These two types of information allowed the collection of two complementary sets of data. While the observer was focused on the user usage, on her/his goals and on his/her conceptual models, the user-logs gave information on how the K-MADE components were used. By using timestamps on both data, we can establish how users used the tool components.

Moreover, we requested each student to exploit their notes and the user-log in order to produce an evaluation report. This report included the modeling process of the observed user, his/her use of the tool, and an analysis of the resulting model. Even though the documents produced were readable and quite organized, models, observer notes and user-logs were also collected for expert analysis.

How did the students' learning progress? This *usage evaluation* session helped to understand the learning of the task model. However, as the first year evaluation needed to be completed, additional data was collected on the learning process of the second

year students. The main point that needed additional information was the learning progression. The collected task models were designed without any time limitation and thus provide a first indication about the learning progression. Two other task models performed in the second year (TM2_1 and TM4) completed the data.

Finally, in order to complete the data about the evolution of K-MADe use, questionnaires were submitted to the second school year students. These questionnaires were composed of questions about specific points raised from the exploitation of the first-year evaluation data.

Lastly, during all the sessions, when the tutors were present, they took notes (including for the last session). These notes were used to complete the answers of all questions.

Table IV. Data collected

Learning step	Practice session	Data	Year 1	Year 2
Model appropriation	session 1	tutor notes	X	X
		designed task model TM1_1	X	
		designed task model TM1_2	X	
		designed task model TM2_1		X
		questionnaires		X
	session 2	user-logs	X	
		observer-student notes	X	
		student exploitation document	X	
designed task models TM1_3		X		
Model use	session 3	designed task models TM4	X	X
		tutor notes	X	X
		questionnaires		X

7.2 Selection of data

Eleven models were assigned to the first group as students were paired to perform the second task model (TM1_2). However, one task model was not completed due to a virus and another one was not returned at all. Then 9 TM1_2 were analyzed.

During the *usage evaluation* session (practice session 2) we collected one folder (set of information) per student (48 folders). It included: observer notes, observer exploitation document, task model, user-log. The second practice session aimed at gaining some understanding of the modeling process. The data used to infer the user modeling process was mainly taken from user-logs. This file was automatically generated. However, we did not wish to use these user-logs without taking the context into account (described in the observer-student notes and the exploitation documents). Four of the folders were not complete and therefore were not included in the analysis. We ended up considering 44 out of the 48 folders in our analysis. However, the TM1_3 folders were only used when we considered that they had a strong impact: for example, when we analyzed the impact of K-MADe on the training progression, we only considered the 18 completed folders corresponding to the folders of the students that designed the collected TM1_2 task models.

Three specific data sources were gathered from the second year. Firstly, we gathered their TM2_1 task models. To design these task models, students worked in groups of three or four people (16 groups). All 16 groups designed task models that were returned and used.

Secondly, after the first use of K-MAD and K-MADe to design the TM2_1 task models, we collected the first questionnaires. The answers to the questionnaire (43 questionnaires were returned while 7 were not) were analyzed according to their different items.

Thirdly when the students completed their last task models (TM4), we collected the second questionnaires. 37 questionnaires were returned. As for the first questionnaire, we selected questionnaires according to the potential impact of the answers. To study the learning process, we analyzed the students who answered both questionnaires (questionnaire 1 and questionnaire 2). From all questionnaires, we collected 35 pairs of questionnaires.

Finally we collected the designed TM4 of both evaluation groups. To design the TM4 task models, students were put into groups. Each group designed one or several task models. All the groups returned all their task models (regrouped into folders).

Table V. Selection of the data

year/group	data	collected	used
1	TM1_2	11	9
1	first group TM1_3 folders	48	44
2	second group TM2_1	16	16
2	first questionnaire	43	according to their relevance
2	second questionnaire	37	
2	both questionnaires	35	

8. RESULTS

The data collected during all the learning sessions allowed us to illustrate the task modeling learning process. We classified our results according to three lines: the training point of view; the autonomy of the students to correct their models; and the evaluation of the completed models.

8.1 The use of the tool from a training point of view

During the first session of the first year, 22 students were present; among these students, 16 chose to design their first task models (TM1_1) with K-MADe. This choice can be explained by two reasons: the major part of students chose to use K-MADe before they were asked to use it to design the other models (in order for them to factorize their learning, i.e., learn at the same time both the model and the tool); others explained that they thought the use of K-MADe would help them to learn a new formalism. The students had a bioinformatics background so they had no apprehension about using the software.

The four students that chose paper and pencil at the beginning of the TM1_1 design, ended up using K-MADe, even though the tutors imposed the use of K-MADe only for the second task model (TM1_2). They changed their choice (to use the K-MADe tool) after a discussion with the tutors about the validation of their models. After this discussion, as they had the feeling that their models were less readable (deletions,

adding of hierarchical level) than the K-MADE user models, they preferred re-editing their task models and completing them using K-MADE.

In the second year, in order to complete this first analysis, we modified the instructions to design the task models. The design of their first task models (TM2_1) was divided into two steps. Firstly, they expressed the task decomposition and informal objects without K-MADE (using paper and pencil). Secondly, they used K-MADE to complete the design of their TM2_1 task models. In the questionnaires, we asked students what learning process they preferred: the use of K-MADE only (from the start of task modeling) - a two step process with a first shorter paper-pencil step - or the first step for as long as it took in the second year (nearly 2 hours for the paper-pencil step) - or a longer paper-pencil step. The results are presented in Table VI.

Table VI. The students' preferences regarding the task modeling learning step

Task modeling learning step	Number
K-MADE use only	5
Shorter paper-pencil step	10
Steps as in the second school year training schedule	26
Longer paper-pencil step	2

Students most predominantly (84% or 38 out of 43 expressed) indicated that they found the splitting of the model learning into two steps more adapted to the task. The duration of the paper-pencil step was not precisely defined. While 60% judged two hours as an appropriate duration, 23% considered it too long.

Three task modeling processes may be defined: (1) K-MADE not used; (2) modeling by using paper and pencil prior to using K-MADE; (3) the use of K-MADE only.

At the end of the task modeling learning, 17% of students indicate that they began the task modeling on paper before completing it by using K-MADE (second task model process). All these students specified that the paper step focused more on task decomposition. All other students (29/35) designed their task models only by using K-MADE (third task model process). However, some students changed their task modeling processes. These processes were evaluated from the beginning to the end of the task modeling learning. Table VII presents the migration of the students' processes.

The migration of students from the process without K-MADE (first task model process) or mixed (second task model process) to the design using only K-MADE (third task model process) may mean that all students would have designed their task models by following the third process if they had practiced long enough.

Table VII. The migration of student processes

Initial task model process	Final task model process	Number of students
first task model process	third task model process	4
second task model process	third task model process	8
third task model process	second task model process	3
		Total: 15/35 expressed

Due to the small size of the first task model (TM1_1), the benefits of K-MADE for navigating task models is not highlighted by this first modeling exercise. In order to facilitate such navigation, K-MADE provides two functionalities: the zoom on the whole

task model and the capacity to display or not the subtasks of a task. The data gathered during practice session 2 shows that 93.75% of students used at least one of these functionalities in order to navigate within their task models.

While the K-MADE software provides several functionalities to facilitate the edition of task models, the paper-pencil method offers a greater freedom in personalization. An assessment of five experts in the ergonomics domain [Couix 2007] shows that they mainly use a personal formalism (4/5). The use of software tools cannot be adapted to this type of users. However, K-MADE allows users to personalize the task model presentation (without any formalism implication). Even though tutors did not present this personalization functionality, 37.6% of the first school-year students modified their preferences during the design of their TM1_3 model. Thus, the personalization of their design tool seems to be a key functionality for the students. Moreover, to be appropriate for task modeling learning, the use of a model that they have never used before needs strict application of the grammar rules in order to allow tutors to identify the student misunderstandings. The use of software tools (such as K-MADE or CTTE) that support training, increases the need to comply with formalism requirements.

K-MADE offers two tools that help to observe modeling rules: the grammar-checking tool (Figure 7a) and the simulation tool (Figure 8). The grammar-checking tool detects mistakes such as the fact that an elementary task cannot be composed of subtasks. Since the beginning of K-MADE use, the tutors have indicated that the presence of both tools in K-MADE has helped to teach, but that it has done so only as regards the task modeling grammar difficulties.

The students detected and corrected their grammar mistakes alone. In addition to this planned use, the grammar-checking tool highlighted the students' misunderstanding of the task composition concept (nearly 1/3 of the first school-year students). The students mixed up the notion of decomposition and inheritance and thus defined tasks composed of only one subtask. This misunderstanding was explained and corrected by the tutors as soon as it was detected (during the first practice session).

8.2 Correction of task models by the students

The formal aspects of task modeling allow the detection of discrepancies (by using the grammar checking tool for example, Figure 7a) prior to allowing the access to the simulation tool (Figure 8). The simulation tool presents a dynamic simulation of the task description (label 1 in Figure 8). Once the available task set (label 2 in Figure 8) is calculated, the designer selects a task that will be added to the current scenario (label 3 in Figure 8). According to the described context (label 5 in Figure 8), the selected task is performed and expressions are computed (label 4 in Figure 8). In addition to the tools that may be used to ensure task model consistency with formalism requirements, another tool indicates information characteristics on the designed task models (Figure 7b). This informational tool was never used by students which shows that during the learning process, students only used tools that helped them to ensure consistency with formalism requirements.

Some observations in this evaluation concerned the formal verification tools for the model design. The reasons why students used K-MADe correction tools and their timing (grammar-checking tool and simulation tool) in the task model processes can be found below.

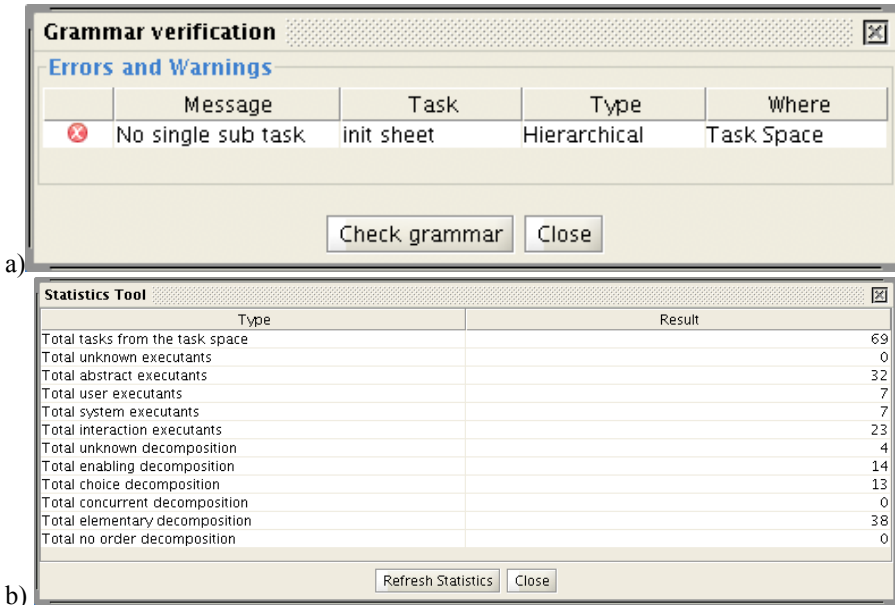


Fig. 7. Views of analysis of task models: a) the grammar-checking tool and b) the statistics tool

When students designed task models without any help (TM1_3), 62.5% of them used at least one correction tool while 12.5% indicated that they did not because of lack of time.

The study of user-logs (from practice session 2) shows that users triggered the correction tools, especially after they defined their task trees (37.5%) and when they considered the task modeling as complete (75%). These uses correspond to the errors detected by the grammar-checking tool. This tool detects at least one error in 90% of all cases indicating that users need this type of tools to design their models. Moreover this grammar-checking tool is the most used correction tool (87.5% of students closed the simulation tool as soon as any coherence error is detected). This last observation introduces the idea that the students triggered the simulation tool in order to grammatically check their task models. Moreover, the formal conditions were systematically checked (sometimes many times) when they were written in order to verify their syntax.

While data from the *usage evaluation* (practice session 2) indicates that users take advantage of the grammar-checking tool, it does not allow to evaluate the full benefits of the simulation tool. During the first year, students asked a lot of questions about the simulation tool functioning and use. To design their TM1_3 task models, only half of students successfully opened the simulation tool and only 6.25% created at least one scenario. During the next sessions (to design the TM4 task model), tutors explained the use of the K-MADe simulation tool again. From tutors' notes, two main reasons

explained that students did not understand the simulation tools: first, when the tool was presented (during the lecture), the students did not understand the relevance of this tool and thus they did not memorize the simulation tool functioning; secondly, when students were ready to verify their models, the tutors showed where the simulation tool could be triggered but did not show how to create scenarios.

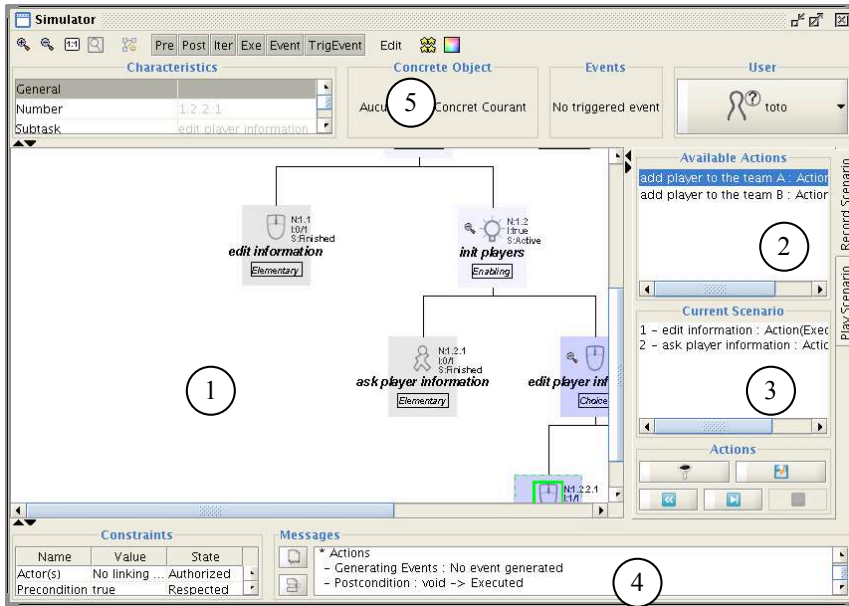


Fig. 8. View of the simulation tool

In order to make sure that the misunderstandings on the simulation tool were due to a teaching problem, during the second year, a part of the lecture focused on the task model checking by using tools. The independent presentation of the simulation and grammar-checking tools showed that both tools could be used in order to validate task models. In the questionnaires, the students were asked to explain in their own words what these tools actually were. All 35 pairs of returned questionnaires were considered. For the students that defined tools (*define* in Table VIII), the main part provided a correct definition (*correct* in Table VIII).

Table VIII. Number of students that defined verification tool

	From the first questionnaire		From the second questionnaire	
	Define	Correct	Define	Correct
grammar-checking	21	18	18	18
simulation	19	19	20	19

The definition of the verification tools and their use allows a good understanding of these two tools and their benefits (at the end of the learning process 100% of students

correctly explained the grammar-checking tool and 95% of students correctly defined the simulation tool).

In order to analyze this observation more precisely, in the second year, the tutors presented the triggering of the grammar-checking tool independently from the simulation tool. Table IX presents the results of the questionnaire answers about the use of the grammar-checking and simulation tools by the second year students. In the middle of the learning process (once TM2_1 task models were designed), 3 questionnaires (from the 35 pairs of returned questionnaires) were not used for the analysis of this point as they contained inconsistencies or as student answers were unclear. For the 32 other questionnaires, 23 separately used both tools. The others (9) used only one of the verification tools (4 the grammar-checking tool and 5 the simulation tool).

At the end of the learning, from the 32 selected questionnaires, 30 separately used both tools and 2 used only one of the tools (1 the grammar-checking tool and 1 the simulation tool). No student that used both tools launched the grammar-checking tool via the simulation one. Thus, the first school-year students performed the grammar-checking via the simulation tool because it was presented according to this process. Both tools were used in 93.75% of the student designs.

Table IX. Results of questionnaire answers about tool uses

	Questionnaire 1	Questionnaire 2
Separate use of grammar-checking and simulation tools	23	30
Use of grammar-checking tool only	4	1
Use of simulation tool only	5	1
Use of grammar-checking tool via simulation tool	0	0
No significant answers		3

Finally, 77.8% of the students in the first year who produced task models with grammar-errors modified their task models to correct them. All of these students corrected their task model until there were no grammar errors. The user-logs showed that the students who did not correct their TM1_3 task models understood the functioning of the grammar-checking tool later (five minutes prior to the end of the session). We may make the hypothesis that they did not correct their task model grammar because of lack of time. The other reason may be the misunderstanding of the grammar-checking tool messages (Figure 9). From the questionnaires, we analyzed the understanding level of the grammar-checking messages. This analysis highlights that, at the end of the task modeling training, 77% of students claimed that they understood the grammar-checking messages. However, the message understanding requires time (50% of students declare they did not understand the messages after their first modeling task (TM2_1) and this number decreases to 17% after their TM4 modeling task). The TM4 task models supported this idea. All the collected task models were grammatically correct.

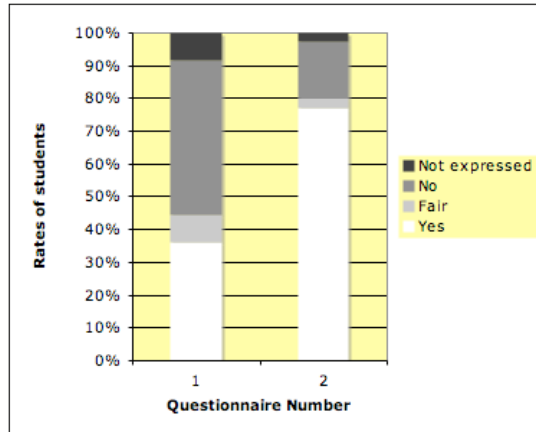


Fig. 9. Understanding of the grammar-checking messages

8.3 Evaluation of models

In this section, we exploit task models that were designed without any time limit in order to analyze time-independent data.

During the first practice session, in order to design their second task model (TM1_2), the first school-year students had to define abstract objects, concrete objects, events, users and to use them to write pre, post and iteration conditions. All these concepts were presented during the lecture and the tutors illustrated the editing of a pre condition. They also corrected the students' mistakes. We did not evaluate whether students correctly used the entities, but only whether they used them or not. Figure 10 presents the percentage of first school-year students who defined formal entities to design the second task model (TM1_2). In addition to the entity definition shown in Figure 10, all TM1_2 task models were composed of at least two abstraction levels.

As shown in Figure 10, the abstract and concrete objects and their use to define *post conditions* seem to be well understood. The definition and use of objects require the manipulation of strings (to name), set types (to store objects) and basic computer types (to define attributes). All the students have a computer science background and therefore no difficult technical issue was raised. Whilst the need for defining entities is well understood by students, the roles and links do not seem to be as well assimilated. Concrete objects are typically used to define conditions. However, 12.5% of students did not instantiate them when defining abstract objects, and 12.5% did not use objects at all to express conditions. 25% of students thus defined abstract objects without any link to any other concept.

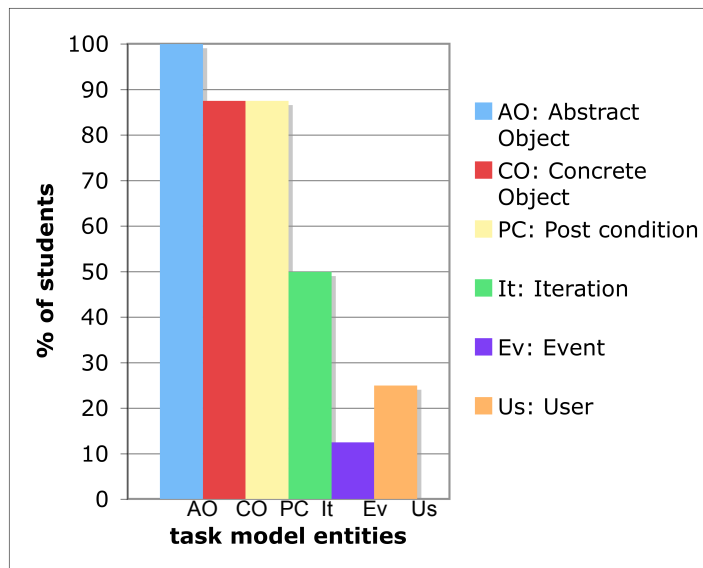


Fig. 10. Percentage of students that defined formal entities during the first session

Concerning the formal **conditions**, 87.5% of students defined at least one post condition from the first session onwards. We did not take into account the definition of pre conditions because tutors used them to illustrate the use of the K-MADE calculator (see Figure 3). When students did not succeed to formally define their conditions (based on the B language [Lano 1996]) using the K-MADE calculator, they did it literally by using textual definition (Figure 11). Even though the definition of the K-MAD formal entities and conditions when using K-MADE caused some difficulties, the necessity of their use in the task modeling process is natural for students in order to complete the task scheduling.

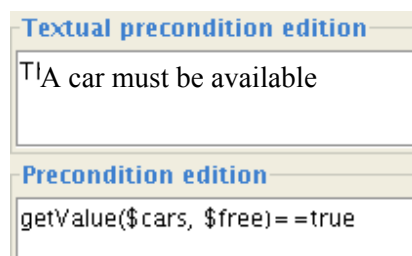


Fig. 11. Edition of a pre condition

The post condition concept is known by the students as Boolean conditions. However, K-MADE defines post conditions as the action(s) performed by task execution. This definition may imply that students misunderstand and that tutors ought to clarify the post condition definition in order for the students to correctly use K-MADE (as they did for the following year students).

Figure 10 shows that few students used the *iteration* conditions, *events* and *users*. Iterations are defined as pre and post conditions (using the calculators); events as well as users are defined as strings. The lack of use of these concepts is not due to the teaching. Whilst objects, pre and post conditions were presented (and illustrated) by tutors, iteration, user and event were only presented in the lecture (without any example) during the first year. Conversely, in the second school year, tutors explicitly presented all concepts. There were no significant changes in the rates of iteration, user and event definitions for the same activities (TM4).

In the first year, we studied the student task modeling process during practice session 2 and we especially observed the impact of formal entities in these processes. Prior to identifying the impact of formalising in the task modeling process, we observed that some students did not define objects. Indeed, 26% of first school-year students (12/46) in the second practice session did not try to define (or use) any K-MADe formal entities. However, we cannot precisely identify why. Two reasons may explain the absence of these elements in the task model process: the limited duration of the experiment, or the non-assimilation of object concepts. Student notes and reports did not allow us to identify the main reason. Six participants indicated that the sessions were not long enough but others (6 out of 12) did not give any relevant information on the subject. The task models of the second school-year students (that are designed without any time limitation) support this idea (only 2 of the 43 TM2_1 are not composed of objects).

From the 34 remaining TM1_3 folders (folders with task model integrating formal entities), we identified two main processes followed by students to define objects and conditions. Table X presents the repartition of students for each process. The most used one (71.5% of processes integrating formal objects and conditions) separately defined objects and conditions. For this student group, 7 out of 10 did not define their objects again once the conditions were defined. These observations give us some understanding about the role objects have in the task modeling process. As an example, the concurrent definition of objects and task tree composition indicates that the user associates objects and tasks. In contrast, when the definition and the use of objects are separated from the task tree composition, we can deduce that the user defines objects only to use them for condition expressions. Therefore, objects associate properties and tasks for some students.

Table X. Repartition of definition processes

Did not use any formal entities	Separately	Conjointly
12	25	9

9. CONCLUSION AND FUTURE WORK

In this paper, we presented an assessment based on the exploitation of data gathered during the use of K-MADe to support the learning of task modeling by bioinformatics students. This evaluation concerned the use of K-MAD by students in a design phase. The data gathered highlights that students find the functionalities of task model navigation (93.75% of students use at least one of the navigation functionalities) useful. Moreover, they indicate that, after being introduced to task modeling (60% of students think that the first two hours of task modeling without K-MADe are mandatory), they prefer using the K-MADe software because the designed task models are more readable

(all students that chose to use paper-pencil method to design their first task model changed their choice after the first modifications). From a training point of view, the use of K-MAde provides a syntactic help (90% of students that launched the simulation tool had at least one syntactic error). Moreover, this analysis shows that students use computing K-MAD concepts only (*user*, *event* and *iteration* concepts are less used than others even though they are presented) because they have some feedback (by using the simulation tool for example). So, this evaluation highlights that the presence of task model tools is essential for learning modeling as it is for editing task model [Giese, et al. 2008].

Overall, this study shows that the use of K-MAde is attractive for task modeling learning. It is particularly beneficial when students are trained to use the conditions in K-MAD. In order to increase these learning benefits, the tool must be modified. These modifications aim at better integrating the grammar-checking tool and facilitating the simulation tool use. For example, Figure 12 shows a grammar error detection for a task (this task is the first task of the marking sheet activity), which is modified in order to integrate the grammar-checking tool. Students can thus see the grammar-checking of their task models as soon as they edit them.

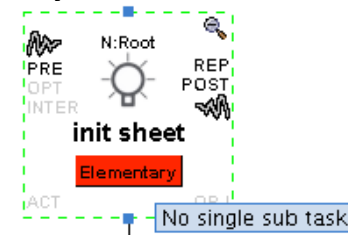


Fig. 12. Example of grammar-checking for a task

During this training evaluation, we also observed that the definitions of objects and conditions are difficult for students. However, both concepts are mandatory to completely express activities [Dix 2008] and to make verifications to validate the models. Another assessment was performed in order to specifically study object and condition definitions [Caffiau, et al. 2008].

Finally, this study focuses on a specific use of the task modeling process (to design interactive applications) for specific user profiles (bioinformatics students). In order to complete this study the same assessment ought to be performed with different user profiles (e.g. computer scientists or ergonomics experts on HCI design). However, this first study allows us to establish some guidelines for task modeling teaching. These guidelines are not general (as they result from a study in which participants have specific knowledge and backgrounds, and in which only the K-MAde tool is studied), but we use them in order to improve our task modeling teaching for students with computer science knowledge and to propose the same teaching for students with different skills (I.T. students, bio-informatics and biology students).

The first identified guideline is issued from the observation of the benefit of K-MAde use during the learning process. The tool helps students to design task models that are consistent with model formalism requirements and therefore facilitates tutor work and students' learning (by increasing student autonomy) during task modeling training.

It must be noted that the use of the tool for task modeling learning is beneficial only if tutors are expert users. As they can anticipate usage difficulties and implementation bugs (for our training, we used a K-MADe version that is currently being developed).

The second identified guideline concerns the activities that are modeled during the training. If the students are familiar with the first modeled activity, the activity discovery step does not take a long time and the students can thus focus only on the modeling process. However, the use of an activity for which the number of available scenarios may be important (such as going to university) is prohibited.

Moreover, to provide exercises for teaching task modeling to students with different skills, activities that include applications (such as playing an on-line game or sending emails) must be limited because computer scientists will describe an activity from a system viewpoint. However, in order to illustrate the use of task modeling in user-centered design, we propose to use the task modeling of an activity for which students will design an interactive application (such as marking a volley-ball marking sheet).

To conclude, even if we define three task-modeling processes for task modeling training, we recommend enforcing the concept definition order. All concepts will thus be used. From our teaching experience, before the learning step, some students modify the concept definition order and adapt their modeling process.

RÉFÉRENCES

- BALBO, S., OZKAN, N. AND PARIS, C. 2004. Choosing the Right Task-modeling Notation: A Taxonomy. In *The Handbook of Task Analysis for Human Computer Interaction*. Lawrence Erlbaum Associates. 445-466.
- BARON, M., LUCQUIAUD, V., AUTARD, D. AND SCAPIN, D. 2006. K-MADe : un environnement pour le noyau du modèle de description de l'activité. In *Proceedings of IHM'06*. ACM. 287-288.
- CAFFIAU, S., GIRARD, P., SCAPIN, D. L., GUITTET, L. AND SANOU, L. 2008. Assessment of Object use in Task Modeling. In *Proceedings of Engineering Interactive Systems (HCSE 2008 and TAMODIA 2008)*. Springer (LNCS 5247). 14-28.
- COUX, S. 2007. Usages et construction des modèles de tâches dans la pratique de l'ergonomie : une étude exploratoire. Conservatoire National des Arts et Métiers, Université Paris Descartes, Université Paris 8, Université Victor Segalen Bordeaux 2. Available at: <http://www.biomedicale.univ-paris5.fr/taskmodelsurvey/accueil/>.
- DIAPER, D. 2004. Understanding Task Analysis for Human-Computer Interaction. In *The Handbook of Task Analysis for Human Computer Interaction*. Lawrence Erlbaum Associates, Inc. 5-48.
- DIAPER, D. AND STANTON, N. A. 2004. The Handbook of Task Analysis for Human Computer Interaction. *Lawrence Erlbaum Associates*.
- DIX, A. 2008. Tasks = Data + Action + Context: Automated Task Assistance through Data-Oriented Analysis (invited paper). In *Proceedings of Engineering Interactive Systems (HCSE 2008 and TAMODIA 2008)*. Springer (LNCS 5247). 1-13.
- GIESE, M., MISTRZYK, T., PFAU, A., SZWILLUS, G. AND DETTEN, M. v. 2008. AMBOSS: A Task Modeling Approach for Safety-Critical Systems. In *Proceedings of Engineering Interactive Systems (HCSE 2008 and TAMODIA 2008)*. Springer (LNCS 5247). 98-109.
- K-MADE electronic reference with download of the software and the manuel (french and english). <http://kmade.sourceforge.net/>.

- LANO, K. 1996. The B Language Method: A guide to practical Formal Development. In *Formal Approaches to Computing and Information Technology (FACIT)*. Springer.
- LIMBOURG, Q. AND VANDERDONCKT, J. 2004. Comparing Task Models for User Interface Design. In *The Handbook of Task Analysis for Human Computer Interaction*. Lawrence Erlbaum Associates, Inc. 135-154.
- LUCQUIAUD, V. 2005. Sémantique et Outil pour la Modélisation des Tâches Utilisateur: N-MDA. Thesis. Poitiers. 285.
- MORI, G., PATERNÒ, F. AND SANTORO, C. 2004. Design and Development of Multidevice User Interfaces through Multiple Logical Descriptions. *IEEE Transactions on Software Engineering*. 507-520.
- NIELSEN. 1993. Usability Engineering. *Academic Press*.
- PATERNÒ, F., MANCINI, C. AND MENICONI, S. 1997. ConcurTaskTrees: A Diagrammatic Notation for Specifying Task Models. In *Proceedings of IFIP TC13 human-computer interaction conference (INTERACT'97)*. 362-369.

BIOGRAPHIES

Sybille Caffiau a obtenu son doctorat en 2009 sur le thème de la prise en compte de l'humain en conception d'applications interactives. Ces recherches couvrent les disciplines de l'ergonomie, l'Ingénierie Dirigée par les Modèles (IDM) et l'Interaction Homme Machine (IHM). Elle est actuellement Attaché Temporaire d'Enseignement et de Recherche (ATER) à l'Ecole Nationale Supérieure de Mécanique et d'Aérotechnique (ENSMA) de Poitiers et membre du Laboratoire d'Informatique Scientifique et Industrielle (LISI). Elle est co-auteur de plusieurs articles dans des conférences nationales et internationales et d'articles de revue.



Dominique L. Scapin est Directeur de Recherche à l'I.N.R.I.A., Rocquencourt, France où il a été Responsable Scientifique des projets MERLin (*Méthodes pour l'Ergonomie des Logiciels Interactifs*) et PsychoErgo. Il mène depuis 1977 des recherches en ergonomie des logiciels, notamment sur les méthodes d'évaluation et de conception des IHM : par ex., sur les langages de commande ; les critères ergonomiques ; les méthodes et outils de description des tâches ; sur le multimedia et la multimodalité. Auteur de nombreuses publications scientifiques en ergonomie des IHM, il est membre du comité de rédaction des revues BIT (et éditeur associé), IwC, UAIS, IJHCI, RIHM, JIPS ; lecteur pour la plupart des journaux scientifiques du domaine ; membre du comité scientifique de la plupart des conférences du domaine. Il a été président des démos pour INTERCHI'93 ; président d'IHM'94 ; co-président des articles de recherche pour INTERACT'95 ; vice-président d'IHM'99 ; co-président d'ERGO-IHM 2000 ; co-président du comité de programme d'IHM'02. Il a géré une vingtaine de contrats de recherche nationaux, internationaux, ou industriels ; encadré une quinzaine de thèses de 3^{ème} cycle en ergonomie et en informatique. Il est expert pour diverses organisations nationales et internationales (par exemple : MESR, ANR, EU, ISO) ; consultant auprès de diverses entreprises et organisations. Il est un ancien président de l'Human Factors Society Europe ; et de l'AFIHM.



Loé Sanou est actuellement Attaché Temporaire d'Enseignement et de Recherche (ATER) à l'Université de Poitiers et rattaché au Laboratoire d'Informatique Scientifique et Industrielle (LISI) de l'Ecole Nationale Supérieure de Mécanique et d'Aérotechnique (ENSMA). Il a obtenu son doctorat en 2008 sur le thème de la Programmation sur Exemple (PsE) en définissant une boîte à outils d'implémentation d'application intégrant la PsE, avec application dans le domaine des tests d'IHM (conformité de l'IHM par rapport au modèle de tâches prescrit). Il est auteur et co-auteur de plusieurs articles de conférences nationales et internationales. Il est membre de l'AFIHM depuis 2002.



Laurent Guittet est Maître de Conférences au Laboratoire d'Informatique Scientifique et Industrielle (LISI). Il enseigne à l'école nationale supérieure de mécanique et d'aérotechnique (ENSMA). Après un diplôme d'ingénieur de cette école en 1981, et un doctorat d'informatique en 1992 sur l'architecture des systèmes interactifs, ses activités de recherche s'articulent autour de la conception et la validation des systèmes interactifs. Il est co-auteur de nombreux articles dans des revues et conférences nationales et internationales.

10. ANNEX: THE QUESTIONNAIRES

Nom : _____ Prénom : _____ Spécialité : _____

Point 1 : Approche de la modélisation des tâches

1.1 – Les premiers modèles de tâches (pour le calendrier) ont été réalisés sur papier. Aurais-tu préféré les réaliser directement avec l'outil K-MADe ?

- oui, dès le premier modèle à réaliser
- oui, à la première séance mais pas dès le début
- non, la première séance sur papier était une étape nécessaire
- non, et j'aurai souhaité rester sur papier plus longtemps

Pourquoi ? _____

1.2 – À présent, lorsque tu réalises un modèle de tâches, comment procèdes-tu ?

- je le fais sur papier.
- je le commence sur papier puis je le termine en utilisant K-MADe.
Précises, ce que vous faites sur papier et ce qui est fait sur K-MADe : _____
- _____
- je n'utilise que K-MADe.

Point 2 : Outil d'aide à la conception

2.1 – K-MADe dispose d'outils pour aider à la conception des modèles de tâches. Parmi ceux-ci, tu utilises :

- l'outil de vérification de la cohérence
Indiques avec tes mots en quoi consiste cet outil :
Comment le lances-tu ?
Quand l'utilises-tu ?
- avant le début de la conception
- pendant le processus, très régulièrement
- après certaines étapes du processus. Lesquelles ? _____
- _____
- une fois que je pense le modèle de tâche terminé
- autre. Précises : _____
- _____
- l'outil de simulation
Indiques avec tes mots en quoi consiste cet outil :
Comment le lances-tu ?
Quand l'utilises-tu ?
- avant le début de la conception
- pendant le processus, très régulièrement
- après certaines étapes du processus. Lesquelles ? _____
- _____
- une fois que je pense le modèle de tâche terminé
- autre. Précises : _____
- _____

2.2 – Comprends-tu toujours les erreurs soulevées par l'outil de vérification de cohérence ?

- oui non

Comment fais-tu pour les rectifier ?

- je me réfère au(x) message(s)
- je cherche dans tout l'arbre (sans lire le message, ce qui m'importe c'est qu'il y ait des erreurs de détectées)
- j'accède directement à la tâche concernée en double-cliquant sur le message
- autre. Précises : _____
- _____

Merci.

Nom :

Prénom :

Spécialité :

Point 1 : Approche de la modélisation des tâches

1.1 – Lorsque tu réalises un modèle de tâches, comment procèdes-tu ?

 je le fais sur papier. je le commence sur papier puis je le termine en utilisant K-MADe.Précises, ce que vous faites sur papier et ce qui est fait sur K-MADe : _____
_____ je n'utilise que K-MADe.**Point 2 : Outil d'aide à la conception**

2.1 – K-MADe dispose d'outils pour aider à la conception des modèles de tâches. Parmi ceux-ci, tu utilises :

 l'outil de vérification de la cohérence

Indiques avec tes mots en quoi consiste cet outil :

Comment le lances-tu ?

Quand l'utilises-tu ?

 avant le début de la conception pendant le processus, très régulièrement après certaines étapes du processus. Lesquelles ? _____ une fois que je pense le modèle de tâche terminé autre. Précises : _____ l'outil de simulation

Indiques avec tes mots en quoi consiste cet outil :

Comment le lances-tu ?

Quand l'utilises-tu ?

 avant le début de la conception pendant le processus, très régulièrement après certaines étapes du processus. Lesquelles ? _____ une fois que je pense le modèle de tâche terminé autre. Précises : _____

2.2 – Comprends-tu toujours les erreurs soulevées par l'outil de vérification de cohérence ?

 oui non

Comment fais-tu pour les rectifier ?

 je me réfère au(x) message(s) je cherche dans tout l'arbre (sans lire le message, ce qui m'importe c'est qu'il y ait des erreurs de détectées) j'accède directement à la tâche concernée en double-clicant sur le message autre. Précises : _____**Si tu as des remarques sur K-MADe (problèmes récurrents, propositions d'amélioration...), n'hésites pas à les ajouter au dos de la feuille. Merci.**

Evaluation des systèmes mobiles et ubiquitaires: proposition de méthodologie et retours d'expérience

Francis Jambon, Nadine Mandran, Brigitte Meillon, Christian Perrot

► To cite this version:

Francis Jambon, Nadine Mandran, Brigitte Meillon, Christian Perrot. Evaluation des systèmes mobiles et ubiquitaires: proposition de méthodologie et retours d'expérience. Journal d'Interaction Personne-Système, Association Francophone d'Interaction Homme-Machine (AFIHM), 2010, 1 (1), pp.1-34. hal-01058933

HAL Id: hal-01058933

<https://hal.inria.fr/hal-01058933>

Submitted on 28 Aug 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Évaluation des systèmes mobiles et ubiquitaires : proposition de méthodologie et retours d'expérience

FRANCIS JAMBON, NADINE MANDRAN,
BRIGITTE MEILLON ET CHRISTIAN PERROT
Laboratoire d'Informatique de Grenoble / équipe MultiCom
Université de Grenoble et UMR CNRS 5217

Résumé : L'évaluation des systèmes interactifs mobiles et/ou ubiquitaires par l'intermédiaire des tests utilisateurs semble a priori plus pertinente sur le terrain qu'en laboratoire d'utilisabilité. Pourtant, les résultats de la littérature apparaissent comme contradictoires. Notre objectif dans cet article est d'en expliciter les raisons et de proposer une méthodologie minimisant les biais. Les expérimentations décrites dans la littérature et nos propres travaux nous ont amenés à définir le concept d'environnement interactif et trois approches expérimentales possibles : laboratoire, terrain et situation réelle. Nous proposons ensuite une méthodologie et une technique (le cheval de Troie) adaptées à l'évaluation en situation réelle. Enfin, nous illustrons notre approche théorique par trois expérimentations et en donnons des retours d'expérience. Nous concluons ensuite sur les limites de notre approche.

Mots clés : Systèmes mobiles, informatique ubiquitaire, évaluation en situation réelle, technique du cheval de Troie.

Abstract: The evaluation of mobile and/or ubiquitous interactive systems via user testing seems a priori more relevant in the field than in a usability laboratory. However, the results of the literature are contradictory. In this article, we aim at explaining the reasons why, and we propose a methodology that could minimize biases. The experiments described in the literature and our own experiments lead us to define the interactive environment concept and three possible experimental approaches: laboratory, field and reality testing. Then, we propose a methodology and a technique –the Trojan horse– adapted to the evaluation in reality testing. At last, we illustrate the theoretical approach by three experiments and give experience feedbacks on them. We conclude on the limits of our approach.

Key words: Mobile devices, ubiquitous computing, reality testing, Trojan horse technique.

1. INTRODUCTION

Les systèmes mobiles ont franchi, il y a plusieurs années déjà, le seuil de la maturité technique et ont trouvé un réel usage auprès de nombreux utilisateurs. Parmi les dispositifs largement diffusés, citons par exemple le système TomTom® de géolocalisation et de routage pour véhicule ou les téléphones mobiles BlackBerry® permettant de lire le courrier électronique. Évolution de ces systèmes mobiles, les systèmes dits ubiquitaires, que l'on regroupe souvent sous le terme plus général d'intelligence ambiante, sortent progressivement de l'état de maquette de laboratoire pour être mis en œuvre dans le monde réel, par exemple dans les musées [Sermet & Millet, 2007].

Ces deux catégories de systèmes permettent la mobilité de l'utilisateur, que ce soit **avec** le système (cas des systèmes mobiles) ou **dans** le système (cas des systèmes ubiquitaires). En outre, le principal intérêt de ces systèmes est d'être sensibles à leur environnement (géolocalisation, disponibilité des réseaux de communication, identification de l'utilisateur, etc.). Ces deux caractéristiques, qui font l'essence même de ces systèmes, ouvrent de nouvelles perspectives concernant les méthodologies adaptées à leur évaluation [Kjeldskov & Graham, 2003], [Hagen, Robertson, Kan, & Sadler, 2005]. En effet, l'évaluation par l'intermédiaire des tests utilisateur de tels systèmes impose de se placer dans un ensemble de situations qui donne tout le sens à leur usage. Ce n'est donc pas seulement un « système interactif » que l'on doit évaluer, mais plus généralement un « environnement interactif ». Toute la difficulté est de définir puis créer cet environnement.

Dans cet article, nous explicitons la problématique et les principaux travaux du domaine qui nous ont amenés à l'évaluation des systèmes interactifs en situation réelle. Nous détaillons ensuite la méthodologie que nous proposons pour leur évaluation, ainsi que ses limites. Enfin, à la lumière des expérimentations que nous avons menées, nous décrivons notre retour d'expérience à la fois méthodologique et technique, puis concluons sur la faisabilité de telles expérimentations. Cet article constitue une synthèse revue et corrigée des précédentes publications de notre équipe sur ce thème : [Jambon, 2006, 2009; Jambon, Golanski, & Pommier, 2006, 2007; Jambon, Mandran, Meillon, & Perrot, 2008; Jambon, Mandran, & Perrot, 2007; Jambon & Meillon, 2009].

2. LIMITES DES APPROCHES ACTUELLES

Afin d'évaluer un système mobile ou un ensemble de systèmes participant à de l'intelligence ambiante en présence de l'utilisateur, il est nécessaire de placer cet utilisateur dans l'environnement interactif tout en gardant assez de contrôle et de capacité d'observation de cet environnement pour pouvoir analyser l'interaction entre l'utilisateur et le système ou l'ensemble de systèmes.

2.1. Environnement interactif

En premier lieu, il est nécessaire de déterminer quels sont les éléments de cet environnement qui seront pertinents vis-à-vis des objectifs de l'évaluation. Puis, en second lieu, il convient de définir comment l'utilisateur peut être placé en interrelation avec ces éléments, soit en les simulant, soit en faisant appel à des éléments issus du monde réel. Du point de vue de l'évaluation, nous proposons de regrouper les éléments de l'environnement interactif selon quatre catégories :

- **l'utilisateur** ou les utilisateurs dans le cas d'un environnement collaboratif ;

- le ou **les dispositifs** matériels et logiciels qui sont l'objet direct de l'évaluation ;
- **les tâches** liées aux fonctionnalités du ou des dispositifs que l'on souhaite évaluer ;
- **le contexte**, que nous définissons comme le complémentaire des trois éléments précédents : les autres utilisateurs, les autres dispositifs au sens large (qu'ils soient informatisés ou non) et l'activité de l'utilisateur sans rapport direct avec le dispositif évalué.

L'une des caractéristiques saillantes du contexte est qu'il peut varier en fonction du déplacement de l'utilisateur. De plus, faire la distinction entre ce qui est contextuel et ce qui ne l'est pas est parfois difficile, car par définition ces systèmes s'intègrent dans leur environnement. En effet, certains éléments du contexte participent directement à l'interaction, alors que d'autres font partie de « l'ambiance générale » et ont un rôle bien moindre, sans toutefois pouvoir être considérés comme négligeables. C'est par exemple le cas des conditions météorologiques dans une station de ski, de l'affluence dans un musée.

2.2. Laboratoire ou terrain ?

La question que l'on doit se poser est : quel degré de réalisme doit-on apporter au contexte de l'environnement interactif pour assurer la validité des analyses ? Faut-il privilégier le laboratoire ou le terrain ? Cette question n'est pas nouvelle en ergonomie. En effet, elle a été abordée dès les années soixante, notamment par Chapanis [Chapanis, 1967]. Dans le cadre des systèmes mobiles et ubiquitaires, l'importance de la mobilité, laquelle influençant la notion de contexte et sa variabilité, ainsi que les progrès techniques accomplis concernant les systèmes d'acquisition (permettant ainsi leur usage dans de nombreux contextes), nous ont incités à revisiter la question.

Dans la littérature, nous distinguons principalement deux approches. Une première approche consiste à simuler l'ensemble des éléments du contexte avec plus ou moins de réalisme. Il s'agit là d'expérimentations en laboratoire d'utilisabilité, dont la méthodologie est bien connue. Pour cette approche, l'un des aspects importants du protocole expérimental est le moyen utilisé pour simuler la mobilité de l'utilisateur. Une seconde approche consiste à placer l'utilisateur dans le monde réel, où les éléments du contexte sont déjà « naturellement » présents. Ce sont les évaluations dites « de terrain ». Pour ce second type d'approche, un élément important est le relatif éloignement de l'utilisateur par rapport aux observateurs, qui rend plus délicate l'acquisition des données expérimentales. Sur ce point, cette approche est similaire aux méthodes d'évaluation à distance.

Globalement, il pourrait sembler « évident » que les évaluations sur le terrain aient une plus grande validité expérimentale que celles effectuées en laboratoire, car nous pouvons être certains que tous les éléments attendus du contexte sont présents, du fait même que l'expérimentation se situe dans la réalité. Nous allons voir que la littérature n'est pas unanime à ce sujet.

Simulation de la mobilité

La mobilité de l'utilisateur peut être simulée de différentes manières. La simulation la plus simple consiste à faire se déplacer l'utilisateur selon un parcours prédéfini, souvent circulaire, autour d'obstacles, et dans un espace restreint : autour d'un bâtiment [Brewster & Walker, 2000], dans un couloir [Pirhonen, Brewster, & Holguin, 2002], ou

dans une salle d'expérimentation [Kjeldskov & Stage, 2004]. Une autre technique consiste à détourner de leur usage prévu (l'entraînement des sportifs) certains systèmes simulant la marche comme les mini-steppers [Pirhonen et al., 2002] ou les tapis motorisés [Kjeldskov & Stage, 2004]. Ces techniques permettent de recréer les mouvements et vibrations que peut ressentir un utilisateur en marchant, et ainsi tester notamment la lisibilité et la facilité avec laquelle il utilise les dispositifs d'interaction dans ces conditions.

Remarquons que la seconde technique est la moins contraignante pour l'instrumentation car elle n'oblige pas l'utilisateur à se déplacer physiquement. Cela permet d'utiliser des caméras fixes ainsi que divers câbles comme l'alimentation électrique du prototype ou le réseau. La première n'a pas ces facilités, mais requiert, de la part de l'utilisateur, une attention particulière pour suivre son itinéraire, ce qui le met dans des conditions plus proches de la marche dans un environnement réel. De plus, elle est mieux adaptée à l'évaluation des systèmes ubiquitaires qui imposent de fait un déplacement physique de l'utilisateur dans le système. Kjeldskov et Stage [Kjeldskov & Stage, 2004] ont comparé ces deux techniques de simulation avec une évaluation statique (utilisateur assis à une table) et une évaluation de terrain (utilisateur laissé libre de ses déplacements). Il en ressort des résultats a priori surprenants : toutes les techniques ont à peu près le même pouvoir de détection, sauf l'évaluation statique qui détecte nettement plus de problèmes d'utilisabilité que les autres. Ceci se manifeste notamment par un plus grand nombre de problèmes de niveau cosmétique¹ détectés. Les auteurs ont cependant identifié des biais possibles concernant le contexte expérimental, sur lesquels nous reviendrons.

Évaluation à distance

La mobilité de l'utilisateur, si elle n'est pas simulée par un appareillage, implique nécessairement un déplacement de celui-ci et donc l'augmentation (même transitoire) de la distance entre lui et les observateurs. Ainsi, l'évaluation d'un système mobile ou ubiquitaire relève aussi de problématiques très proches de l'évaluation à distance [Hammontree, Weiler, & Nayak, 1994]. Cette distance n'est pas forcément importante. Dans une rue, l'observateur peut être seulement à quelques mètres de l'utilisateur. Dans un laboratoire d'utilisabilité, l'utilisateur va se déplacer, se tourner, et donc rendre très difficile son suivi à travers une glace sans tain ou même avec des caméras orientables télécommandées. Dans tous les cas, la visualisation des interactions est délicate.

Lorsque la distance augmente, à partir de quelques dizaines de mètres, on ajoute à cela des difficultés pour collecter l'information. On a alors recours à des caméras de type col-de-cygne et des micros sans fil, systèmes comparables à ceux utilisés lors de certaines évaluations à distance [Scholtz, 2001]. Lorsque la distance augmente encore, dès que l'observateur perd le contact visuel direct avec l'utilisateur, la situation devient alors très similaire à une évaluation à distance, tant du point de vue technique que méthodologique. En effet, le rôle du facilitateur² de l'expérimentation s'en trouve modifié significativement car il doit utiliser des systèmes de communication (téléphone,

¹ La définition utilisée pour le terme « cosmétique » est celle de Molich [Molich, 2000].

² Le *facilitateur* est chargé d'accompagner l'utilisateur au cours de l'expérimentation, de lui donner des consignes, de le questionner, de résoudre d'éventuels petits problèmes techniques. Il se distingue de l'*observateur* qui a pour rôle unique d'observer la situation, de prendre des notes, de gérer les systèmes d'enregistrement, mais qui n'interagit pas directement avec l'utilisateur.

talkie-walkie, etc.) afin de transmettre les consignes et répondre aux questions de l'utilisateur.

Le domaine de l'évaluation ergonomique à distance (« Remote Usability Testing » pour les anglo-saxons) [Hammtree et al., 1994] est principalement représenté dans la littérature par des expérimentations destinées à l'évaluation de sites web. En effet, ces systèmes, assez simples à mettre en œuvre, sont bien adaptés à ce type de protocole expérimental. Nous n'avons trouvé dans la littérature qu'un seul exemple de test à distance d'un système mobile, mais là encore il s'agit d'un navigateur web sur PDA [Waterson, Landay, & Matthews, 2002]. La question centrale qui occupe bon nombre d'études sur l'évaluation à distance concerne la validité des analyses effectuées. Cette question est apparue très tôt [Hartson, Castillo, Kelso, & Neale, 1996] [Castillo, Hartson, & Hix, 1998] et continue d'alimenter la littérature avec des résultats contradictoires. Pour la majorité des auteurs, peu de différences sont constatées [Bartek & Cheatham, 2003] [Brush, Ames, & Davis, 2004] [Thompson, Rozanski, & Haake, 2004], pour d'autres au contraire, il y a des différences significatives sur les problèmes ergonomiques détectés [Tullis, Fleischman, McNulty, Cianchette, & Bergel, 2002]. De notre point de vue, ces contradictions sont issues du fait que, plus que la distance, ce sont les conditions expérimentales du terrain qui déterminent ces différences. Nous nous sommes alors posé la question de définir ce qu'est une expérience de « terrain ».

Définition de la notion de « terrain »

Dans la littérature, la notion de « terrain » n'est pas définie précisément, mais plutôt vue comme l'opposé du laboratoire d'utilisabilité. Il y a un consensus de fait à présenter les expérimentations de terrain comme revenant à placer l'utilisateur dans l'environnement « naturel » de l'usage attendu du système étudié. On parle parfois de situation « écologique ». Par exemple, un PDA destiné à des infirmières sera évalué à l'intérieur d'un hôpital [Kjeldskov, Skov, Als, & Høegh, 2004]. Dans le vocabulaire anglo-saxon, plusieurs termes sont utilisés pour décrire ce type de protocole expérimental. Les plus usités sont « field experiments » (e.g. [Goodman, Brewster, & Gray, 2004]), « field tests » (e.g. [Hertzum, 1999]), « field trials » (e.g. [Jensen & Larsen, 2007]), ou expérimentations « in the wild » (e.g. [Waterson et al., 2002]). Certaines appellations plus anecdotiques sont parfois utilisées, comme « quasi-experimentations » [Roto, Oulasvirta, Haikarainen, Lehmuskallio, & Nyysönen, 2004]. La communauté scientifique qui s'intéresse régulièrement à la question depuis quelques années, notamment via des ateliers, utilise également un vocabulaire varié. On parle ainsi de « reality testing »³, d'expérimentations « in situ »⁴ ou plus récemment de « mobile living labs »⁵. Il n'existe pas à notre connaissance de tentative d'uniformisation de ces appellations. Nous utiliserons le terme de « terrain » dans la première partie de l'article, puis nous le préciserons en introduisant la notion de « situation réelle », et enfin nous proposerons des appellations plus précises dans la partie discussion.

³ CHI'2006 workshop « Reality Testing: HCI Challenges in Non-Traditional Environments » (<http://www.cs.indiana.edu/surg/CHI2006/>)

⁴ MobileHCI'07 workshop « In-Situ 2007: Using Mobile Devices and Emergent Technology for In-Situ Evaluations » (<http://insitu2007.freeband.nl/>)

⁵ MobileHCI'09 workshop « Mobile Living Labs 09: Methods and Tools for Evaluation in the Wild » (<http://mll09.telin.nl>)

Quelle est la valeur ajoutée du terrain ?

De nombreux travaux ont cherché à déterminer si le fait de se placer dans le contexte réel apportait véritablement une plus grande validité aux évaluations [Fields, Amaldi, Wong, & Gill, 2007]. Parfois, il s'agit plus simplement de réduire les coûts logistiques de l'expérimentation [Rowley, 1994]. Nous avons précédemment réalisé un état de l'art de ces travaux [Jambon et al., 2006] et [Jambon et al., 2008]. Cependant, une recherche plus approfondie des publications existantes, de récents résultats issus de la littérature, ainsi que les retours d'expérience de nos propres expérimentations, nous ont amenés à remettre en cause certains points importants de cet état de l'art. Dans la littérature, la comparaison entre les protocoles expérimentaux s'effectue principalement selon deux critères : le nombre de problèmes détectés et éventuellement leur impact, généralement selon la définition de Molich [Molich, 2000]. Les résultats de ces travaux peuvent être classés en trois catégories : ceux qui ne détectent pas de différence notable, ceux qui estiment que le laboratoire apporte plus de détection de problèmes d'utilisabilité que le terrain, et ceux qui affirment le contraire.

Concernant la première catégorie, Kjeldskov et al. indiquent détecter très peu de différences entre le laboratoire et le terrain lors de l'évaluation d'un dispositif mobile destiné à des infirmières [Kjeldskov et al., 2004]. En effet, tous les problèmes d'utilisabilité, sauf un, ont été détectés en laboratoire. Une deuxième publication de Kjeldskov et al. concernant l'évaluation d'un guide mobile pour un réseau de transport en commun, confirme ce résultat quantitatif [Kjeldskov et al., 2005]. Cependant, les auteurs précisent qu'il y a peu de corrélation entre les problèmes détectés lorsqu'ils sont de niveau cosmétique. Les résultats quantitatifs sont similaires à ceux de Betiol et Cybis [Betiol & Cybis, 2005] ainsi que ceux de Kaikkonen et al. [Kaikkonen, Kekäläinen, Cankar, Kallio, & Kankainen, 2005]. Cependant, alors que Betiol et Cybis détectent une plus grande sévérité des problèmes détectés en laboratoire, Kaikkonen et al. au contraire, indiquent que les problèmes détectés sur le terrain sont légèrement plus sévères.

D'autres auteurs, parfois les mêmes que précédemment, affirment détecter plus de problèmes d'utilisabilité en laboratoire que sur le terrain. Par exemple, Kjeldskov et Stage ont réalisé une étude poussée concernant six méthodes : avec utilisateur assis à une table, en simulant la mobilité en laboratoire (4 façons différentes), et sur le terrain [Kjeldskov & Stage, 2004]. Ces travaux montrent que la technique la plus simple (utilisateur assis à une table) détecte plus de problèmes que toutes les autres, notamment le terrain. Remarquons cependant que le différentiel est principalement dû à des problèmes de niveau cosmétique. Ces résultats sont cohérents avec ceux de Baillie et Schatz [Baillie & Schatz, 2005]. Notons qu'ils sont aussi en accord avec les travaux de Hertzum sur l'évaluation à distance de système non mobiles [Hertzum, 1999].

Enfin, plus récemment, Duh et al. [Duh, Tan, & Chen, 2006] ont remis en cause ces études en affirmant qu'au contraire, le terrain apporte plus de détection de problèmes ergonomiques que le laboratoire d'utilisabilité. Le travail de Duh et al. est d'autant plus intéressant que les auteurs utilisent la même classification [Molich, 2000] de la sévérité des problèmes ergonomiques que Kjeldskov et al. [Kjeldskov & Stage, 2004] avec un protocole expérimental très similaire à celui de Kaikkonen et al. [Kaikkonen et al., 2005] qui concluaient au contraire que les différences étaient inverses ou minimales... Ces résultats sont cohérents avec les travaux de Po et al. concernant l'évaluation heuristique. Po et al. ont détecté que l'évaluation heuristique sur le terrain, c'est-à-dire lorsque les experts réalisent leur évaluation dans l'environnement d'usage prévu du dispositif,

permet de révéler de plus nombreux problèmes d'utilisabilité que lorsque cette évaluation est réalisée en laboratoire, notamment vis-à-vis des problèmes liés au contexte [Po, Howard, Vetere, & Skov, 2004].

Ces travaux mènent à la conclusion que non seulement il n'y a pas de consensus concernant l'intérêt des expérimentations sur le terrain, mais aussi que les résultats expérimentaux des comparaisons entre laboratoire et terrain sont difficilement reproductibles. Nous en avons conclu que les variables indépendantes permettant de différencier les expériences de terrain de celles en laboratoire ne sont pas encore bien identifiées ou que certains biais non identifiés faussent les résultats.

Biais possibles

En marge des résultats de leurs articles, les auteurs indiquent quelques remarques ou anecdotes qui peuvent donner des pistes intéressantes. Kjeldskov et al. [Kjeldskov & Stage, 2004] par exemple, indiquent qu'une explication possible du faible nombre de problèmes détectés sur le terrain peut venir du fait que les utilisateurs n'explorent pas tous les aspects de l'interface. En effet, ils n'effectuent que les tâches qui sont pertinentes vis-à-vis de leur activité sur le terrain.

Dans le même article, les auteurs ont remarqué que, lorsqu'un sujet se déplace dans une rue en compagnie d'un facilitateur et d'un caméraman, un effet de groupe se produit et l'utilisateur se retrouve isolé du monde, comme dans une bulle protectrice : les personnes de l'environnement s'écartent sur le passage et les relations sociales s'en trouvent limitées car les autres personnes n'osent pas s'immiscer dans l'expérimentation. Nous avons constaté le même effet lors de nos expérimentations, où les personnes connaissant un utilisateur ont eu des réticences à le saluer lors de son passage.

Baillie et Schatz ont détecté que les utilisateurs effectuaient leurs tâches plus rapidement et en faisant moins d'erreurs sur le terrain [Baillie & Schatz, 2005]. Ils supposent que les utilisateurs se sentaient plus « détendus » hors du contexte du laboratoire d'utilisabilité. De manière similaire, mais pour un système non mobile, Schulte-Mecklenbeck et al. [Schulte-Mecklenbeck & Huber, 2003] indiquent que, dans une tâche de recherche d'information via une interface web, les utilisateurs présents en laboratoire plutôt qu'à distance abandonnent moins facilement l'expérimentation et récupèrent plus d'informations. Les auteurs supposent que la « pression psychologique » de la présence de l'observateur en est la principale raison.

Une autre anecdote, rapportée par Isomursu et al. [Isomursu, Kuutti, & Väinämö, 2004] nous donne une autre piste. Dans son expérimentation, les utilisateurs sont recrutés dans la rue et par groupe de deux. L'un est chargé d'utiliser le dispositif (un PDA sensible au contexte) tandis que l'autre est chargé de filmer, à l'aide d'un téléphone mobile disposant d'une caméra, la situation vécue lorsqu'une difficulté intervient. Cette vidéo est nommée « experience clip ». Les auteurs ont constaté que lorsque les sujets étaient recrutés seuls, et que l'un des chercheurs servait de caméraman, la qualité, la durée et la pertinence des vidéos diminuait significativement. Plus surprenant, les sujets ont tous indiqué que cette présence ne les gênait en rien dans l'évaluation...

Limites des protocoles

Ces remarques et anecdotes nous ont amenés à étudier plus précisément les protocoles expérimentaux utilisés en laboratoire et sur le terrain. Dans toutes les études

citées concernant des systèmes mobiles, les protocoles sont quasiment identiques en laboratoire et sur le terrain. Notamment, les tâches effectuées par l'utilisateur sont prescrites. De plus, afin de faciliter le déroulement de l'expérimentation et l'observation, les évaluations se déroulent en présence d'un facilitateur, parfois aussi en présence d'un caméraman ou d'observateurs. L'instrumentation des expérimentations se base généralement sur l'usage d'une caméra filmant le contexte et parfois d'une seconde caméra col-de-cygne fixée au dispositif permettant de visualiser clairement les interactions entre l'utilisateur et le dispositif mobile.

L'intérêt d'utiliser des protocoles identiques est de pouvoir comparer les deux approches en ne considérant pas le protocole expérimental comme une variable indépendante. Cependant, du point de vue méthodologique, cela revient à utiliser le terrain comme une extension du laboratoire au sens où l'utilisateur, le dispositif, le facilitateur et l'instrumentation nécessaires à l'évaluation sont simplement « déplacés » dans un contexte réel. Seul le « décor » a changé. En conséquence, l'utilisateur se retrouve dans un contexte qui est loin d'être réel. En effet, l'utilisateur met en œuvre un dispositif qui est parfois très impressionnant et encombrant, notamment si une caméra col-de-cygne est utilisée. De plus, il subit, même sans en être conscient, une certaine influence due à la présence des personnes chargées de gérer ou d'observer l'expérimentation.

Au final, notre hypothèse est que sur le terrain, l'utilisateur évolue dans une sorte de « bulle » où les tâches qu'il doit réaliser, son activité générale, ses relations sociales, etc. sont contraintes par le protocole expérimental. Nous supposons que les bénéfices attendus du terrain sont en partie annihilés par ces contraintes, et qu'en conséquence, les différences détectées entre les deux configurations sont en partie des artefacts.

3. PROPOSITION D'UNE NOUVELLE APPROCHE

3.1. Laboratoire, terrain et situation réelle

Notre hypothèse initiale est qu'il n'y a pas deux configurations possibles mais plutôt trois : le laboratoire, le terrain et la **situation réelle**. Cette dernière configuration se différencie du terrain par une absence quasi totale des contraintes liées à l'expérimentation, permettant ainsi d'en limiter les artefacts. En effet, nous supposons que les différences entre le laboratoire et le terrain sont peu déterminantes car le protocole expérimental est finalement très similaire, du fait notamment de la présence de personnels et dispositifs d'observation, mais aussi du fait que la tâche est prescrite. Sur ces points, nos hypothèses sont en accord avec celles de Thomas et Kelloggs, exprimées dans les années 80, concernant la minimisation du « fossé écologique » [Thomas & Kellogg, 1989]. Ainsi, nous proposons de définir les approches « terrain » et « situation réelle » selon les critères suivants :

Comparée au laboratoire, la situation est dite **sur le terrain** lorsque :

- Le contexte d'usage du dispositif est réel, c'est-à-dire que l'évaluation doit avoir lieu dans le contexte où le dispositif est censé être utilisé, ou à défaut, un contexte équivalent ;
- Le dispositif et les données à disposition des utilisateurs sont perçus comme réels par l'utilisateur, même si des techniques de simulation comme celle du Magicien d'Oz sont utilisées.

La situation est dite **en situation réelle** si, de plus :

- La tâche de l'utilisateur n'est pas prescrite, mais découle de la situation vécue par l'utilisateur en fonction de ses propres aspirations et des événements planifiés ou opportunistes issus du contexte ;
- Les personnels liés à la gestion de l'expérimentation et à l'observation, ainsi que les dispositifs d'acquisition sont hors du champ de perception de l'utilisateur.

De nombreuses expérimentations sur le terrain ont été réalisées grâce à des systèmes d'acquisition techniquement élaborés, notamment via des systèmes d'enregistrement audio-vidéo multi-sources portables (par exemple [Lyons & Starner, 2001], [Roto et al., 2004], [Zouinar, 2004] et [Calvet, Salembier, Kahn, & Zouinar, 2005]), grâce à des systèmes d'analyse de traces et de questionnaires proactifs (par exemple [Froehlich, Chen, Consolvo, Harrison, & Landay, 2007]) ou grâce à une combinaison de ces systèmes [Riegelsberger & Nakhimovsky, 2008]. Cependant, peu d'expérimentations en situation réelle sont relatées dans la littérature concernant les systèmes mobiles ou ubiquitaires. Les rares expérimentations concernant les téléphones mobiles se focalisent sur des statistiques d'usage comme celle de Demumieux et Losquin [Demumieux & Losquin, 2005] ou celle de Jensen et Larsen [Jensen & Larsen, 2007]. Citons également l'utilisation d'une technique dérivée du magicien d'Oz par Consolvo et al. pour la simulation in-situ d'un système ubiquitaire [Consolvo et al., 2007]. Cette rareté peut s'expliquer par le fait que ces expérimentations sont relativement difficiles à mettre en œuvre à cause de l'imprévisibilité des événements du contexte réel, un point souligné par Kellar et al. [Kellar et al., 2005], et aux difficultés rencontrées pour observer l'interaction sans introduire dans le contexte des systèmes d'acquisition visibles.

3.2. Principe d'incertitude et technique du cheval de Troie

Principe d'incertitude

En effet, en l'absence de facilitateur, d'observateur, de caméraman ou de dispositifs visibles d'acquisition de données, l'analyse des interactions entre l'utilisateur et le système devient difficile. On doit se limiter le plus souvent aux traces enregistrées par les dispositifs (actions de l'utilisateur sur l'interface, géolocalisation, capteurs, etc.) et aux entretiens postérieurs à l'expérimentation. Sont également exclus du protocole l'usage d'un journal de bord, la technique des incidents critiques [Hagen et al., 2005], ou d'autres techniques similaires comme celle des « emoticons » [Arhipainen, Rantakokko, & Tähti, 2004] car ils modifient l'activité de l'utilisateur et donc perturbent le caractère « écologique » de la situation.

Si, via les traces, la détection des principaux problèmes ergonomiques est possible, nous avons montré que leur explicitation reste un problème [Jambon et al., 2006]. Les entretiens qualitatifs à la fin de l'expérimentation sont bien adaptés à des expérimentations courtes où les objectifs sont relatifs à l'usage des dispositifs, mais ils sont beaucoup moins adaptés à des expérimentations de longue durée ou si l'on se focalise sur les détails de l'interaction, car l'utilisateur risque d'oublier une partie des difficultés rencontrées.

Nous nommons cette contrainte « Principe d'Incertainitude » du fait de sa proximité sémantique avec celui énoncé par Heisenberg concernant la physique quantique, et qui dans notre contexte, peut s'énoncer ainsi : « ***il n'est pas possible à la fois d'observer précisément une situation d'interaction homme-machine sans, par effets de bord, la perturber*** ». Un principe similaire peut se rencontrer en ethnologie [Malaurie, 2002]. Ce principe a pour conséquence de placer les expérimentateurs devant un difficile

dilemme : soit ils choisissent d'observer avec précision en acceptant de nombreux biais, soit ils minimisent les biais, mais ils ne disposeront alors que d'observations très limitées, avec pour conséquence de minimiser l'intérêt de l'étude.

Même s'il n'est pas a priori possible de transgresser ce principe d'incertitude, notre objectif a été de définir un protocole expérimental permettant de minimiser les biais tout en garantissant que nous disposerions d'assez d'informations pour appréhender la situation et ainsi réaliser l'évaluation. Pour des raisons déontologiques évidentes, il n'est bien entendu pas possible de faire abstraction du fait même que l'utilisateur ait connaissance d'être dans le cadre d'une expérimentation. Nous avons donc tout d'abord cherché à tirer le maximum des traces issues des dispositifs, puis nous nous sommes intéressés au moyen d'introduire des dispositifs d'observation sans qu'ils soient vus comme tels.

Technique du cheval de Troie

En situation réelle, il est ainsi nécessaire de se passer d'un observateur et a fortiori d'un caméraman. Pourtant, les enregistrements vidéos sont une source extrêmement riche et utile pour interpréter les événements non anticipés. Les caméras col-de-cygne, permettant de capturer l'interaction avec les dispositifs mobiles, sont un exemple de dispositif très utile mais trop intrusif. L'usage de caméras de très petite taille portées par l'utilisateur lui-même pourrait être une solution. Cette instrumentation reste cependant problématique car, aujourd'hui encore, ces caméras ne sont pas assez performantes et miniaturisées pour se faire oublier. De plus, la présence de la caméra doit être indiquée à l'utilisateur pour d'évidentes raisons déontologiques.

C'est pourquoi, plutôt que de chercher à masquer les dispositifs d'instrumentation, notre idée a été de les faire passer pour autre chose que ce qu'ils sont. Cette technique est de la même inspiration que celle du magicien d'Oz, au sens où l'on masque certains aspects du protocole d'évaluation à l'utilisateur. Nous l'avons nommée « Cheval de Troie », car elle est basée sur un double usage des dispositifs liés à l'instrumentation : le premier usage est voyant, mais sa présence est justifiée, le second et véritable usage est masqué. Ainsi, pour un dispositif d'instrumentation donné, nous présentons à l'utilisateur ce dispositif comme faisant partie du système ou de l'environnement étudié. Mais en fait, l'intérêt principal du dispositif est l'acquisition des informations nécessaires à l'évaluation.

Par exemple, pour l'évaluation d'un dispositif destiné à des skieurs [Jambon, 2006], la mini-caméra disposée sur le casque de chacun des skieurs avait pour premier usage d'enregistrer le vécu du skieur afin de pouvoir le lui rejouer sur son smartphone. Cette mini-caméra permettait aussi de filmer l'interaction du skieur avec son smartphone, et de ce fait, elle permettait de disposer d'informations très utiles pour l'évaluation. Cette astuce nous avait semblé être une solution ponctuelle. Nous l'avons néanmoins reproduite, sous une forme différente, lors de l'expérimentation au Muséum de Lyon : la carte distribuée aux visiteurs faisait partie de la scénographie et permettait, en même temps, l'étude du parcours des visiteurs [Jambon, Mandran et al., 2007].

Nous nous sommes ainsi aperçus que la technique du cheval de Troie était généralisable dans de nombreux cas. Deux approches sont possibles :

- Si le système ou l'environnement dispose déjà d'un dispositif ayant potentiellement les capacités d'acquisition des données souhaitées, il s'agit de le compléter, par exemple par un système d'enregistrement ou par l'installation

d'équipements complémentaires. Dans ce cas, la justification de la présence du dispositif est aisée.

- Si le système ou l'environnement ne contient pas ce dispositif, il faut l'y ajouter en le faisant passer pour une fonctionnalité supplémentaire, mise à disposition de l'utilisateur, même si elle ne sert à rien du point de vue de l'expérimentation. Par exemple, si l'on souhaite étudier l'usage géolocalisé des fonctions d'un téléphone mobile, il faudra que le téléphone dispose aussi d'un GPS mis à disposition de l'utilisateur et présenté comme l'une des fonctionnalités testées, même si l'expérimentation ne s'intéresse qu'à l'envoi de SMS.

Il existe néanmoins certains cas où cette technique n'est pas utilisable. De manière générale, elle ne peut être mise en œuvre dans les cas où la présence du dispositif d'acquisition n'est pas justifiable ou semble incongrue. Par exemple, il sera délicat de justifier l'usage d'une mini-caméra portée par l'utilisateur si l'on souhaite évaluer un dispositif grand public destiné à donner les horaires de bus.

La technique du cheval de Troie pose également un sérieux problème déontologique. En effet, si les informations données à l'utilisateur ne sont pas complètes, cette technique revient à enregistrer l'activité de l'utilisateur à son insu, ce qui est n'est pas en accord avec les règles d'éthique des expérimentations [Johnson, Solso, & Beale, 1997]. Ainsi, une attention particulière doit être portée à l'aspect déontologique du protocole expérimental. Il est notamment impératif d'informer les utilisateurs de l'ensemble des données recueillies et enregistrées, et de l'utilisation qui va en être faite. Cela est délicat, car afin de maintenir l'aspect « écologique » de la situation, il serait préférable de cacher aux utilisateurs le véritable objectif de l'acquisition de données. Sur le plan déontologique, ce n'est pas admissible sauf dans certains cas très particuliers d'expérimentations en psychologie expérimentale. C'est pourquoi nous proposons une voie médiane, où nous informons les utilisateurs de l'utilisation des données en « remarques » lors des consignes. L'information complète leur est donnée en fin d'expérimentation, moment auquel leur consentement pour l'utilisation réelle de ces données peut être recueilli. Notons que si des données personnelles sont nécessaires, une déclaration à la CNIL doit être effectuée.

3.3. Motivations et agenda de recherches

Notre hypothèse est qu'il existe trois configurations pour l'évaluation des dispositifs mobiles. Si la première, le laboratoire, est aujourd'hui bien connue et validée, les deux autres restent encore à défricher aussi bien du point de vue technique que méthodologique. De plus, il n'y a pas unanimité concernant le gain attendu à utiliser une configuration sur le terrain ou en situation réelle vis-à-vis d'une configuration en laboratoire. Les configurations laboratoire et terrain ayant déjà été bien étudiées dans la littérature, nous nous sommes intéressés plus particulièrement à la situation réelle. Nous souhaitons en effet déterminer si ces expérimentations en situation réelle peuvent apporter certaines informations impossibles à obtenir en laboratoire ou sur le terrain. Notre étude s'intéresse aux configurations expérimentales, et plus particulièrement à l'aspect recueil. Nous ne nous sommes pas intéressés aux autres caractéristiques du protocole comme la sélection des sujets ou les techniques d'évaluation.

Nous nous sommes ainsi engagés dans une série de trois expérimentations destinées à valider les difficultés techniques et méthodologiques des expérimentations en situation réelle. Une des difficultés de cette démarche est l'impossibilité de reproduire, dans les

trois configurations, le même protocole expérimental. En effet, en situation réelle, les systèmes d'acquisition classiques utilisés en laboratoire sont proscrits. C'est pour cette raison que, dans un premier temps, notre objectif a été d'évaluer le pouvoir d'analyse des traces que peuvent laisser les dispositifs comparés aux techniques d'observation classiques (enregistrements audio-vidéo principalement). Pour cela, nous avons réalisé une première expérimentation faisant appel à deux techniques d'évaluation distinctes : MapMobile [Jambon et al., 2006]. Dans un second temps, nous avons cherché à valider les protocoles expérimentaux concernant la configuration en situation réelle peu explorée dans la littérature. Pour cela, deux expérimentations, dont le but était de tester la capture de données en situation difficile (activité sportive en extérieur) ou sur une longue durée (quelques semaines) et un grand nombre d'utilisateurs (plusieurs centaines de personnes), ont été mises en place : E-Skiing [Jambon, 2006] et Muséum [Jambon, Mandran et al., 2007].

4. EXPERIMENTATIONS ET RETOURS D'EXPERIENCE

Les trois expérimentations décrites ci-après ont été effectuées dans le cadre des projets ADAMOS (RNTL-PROACT Franco-Finlandais) et IMERA (Région Rhône-Alpes Emergence). Ces deux projets avaient pour objectif l'étude du processus de conception de services pouvant réagir de manière proactive vis-à-vis des utilisateurs. Ces services s'appuient sur des dispositifs mobiles, ubiquitaires, communicants, et sensibles au contexte d'usage, permettant ainsi aux services d'être adaptatifs. Au cours de ces projets, trois campagnes d'expérimentation des dispositifs ont été effectuées.

4.1. Expérimentation « MapMobile »

MapMobile est un assistant numérique personnel permettant à son utilisateur de se géolocaliser et d'être guidé à l'intérieur et à l'extérieur d'un bâtiment. Le prototype a été développé conjointement par France Télécom R&D et le CEA-Leti. Dans le cadre du projet ADAMOS, des fonctions d'adaptativité (en fonction des centres d'intérêts de l'utilisateur) et de proactivité (en fonction de la géolocalisation et d'événements extérieurs) ont été ajoutées. L'objectif de l'expérimentation était triple : (1) évaluer l'utilisabilité du prototype, (2) valider les méthodes et techniques d'évaluation et (3) servir d'illustration à une étude sociologique. Ce sont seulement les deux premiers points qui nous intéressent ici.

Protocole expérimental

L'expérimentation était située dans un environnement réel : les bâtiments de France Télécom R&D à Meylan pendant les heures de travail. Le dispositif était totalement fonctionnel. Une tâche de haut niveau était imposée à l'utilisateur : rendre visite à un contact professionnel dans un lieu inconnu de lui. Au cours de l'expérimentation, l'utilisateur était accompagné d'un facilitateur et d'un caméraman (cf. images 1 et 2). Il s'agissait donc d'une expérience de terrain sans être une expérimentation en situation réelle. En effet, nous avons cherché des conditions expérimentales permettant de disposer à la fois des données classiques disponibles lors d'évaluations sur le terrain et une simulation des données disponibles lors d'évaluations en situation réelle.



Images 1 et 2 : sujet (au centre en bleu) avec le facilitateur (à droite en blanc) et le caméraman (à gauche en orange) filmant l'écran du système MapMobile (image de droite) au cours de l'expérimentation.

L'expérimentation a concerné une douzaine de sujets dont seulement dix ont été effectivement inclus dans les analyses du fait de problèmes techniques ou organisationnels. Les sujets ont été sélectionnés de manière à être répartis selon les critères sociologiques de la méthode CAUTIC [Forest, Guilloux, Mallein, & Panisset, 1998]. Dans les faits, même si cela n'était pas recherché, l'expérience des sujets dans l'usage des nouvelles technologies ainsi que leur genre ont été également bien répartis. Les sujets étaient rémunérés. Les bâtiments de France Télécom Meylan R&D ont la forme de carrés, vides en leur centre, se joignant par les sommets sans marques distinctives. L'orientation à l'intérieur des bâtiments est particulièrement difficile, ce qui rendait le contexte d'expérimentation tout à fait réaliste⁶.

Le protocole était le suivant : Les sujets étaient accueillis au poste de garde situé à l'entrée du site, puis accompagnés jusqu'à l'accueil du bâtiment. À ce moment, le PDA destiné à les guider dans le bâtiment leur était remis et ils étaient équipés d'un micro-cravate sans fil. Les consignes leur étaient alors données. Ensuite, les sujets se déplaçaient librement dans le bâtiment, guidés par le dispositif, tout en étant régulièrement interrompus par celui-ci du fait de son comportement proactif. Il était notamment fait mention d'un retard de la personne visitée, d'un document à aller chercher, d'informations profilées disposées sur un panneau d'affichage, etc. L'expérimentation durait environ une demi-heure. Tout au long de l'expérimentation, le facilitateur encourageait les sujets à exprimer leurs difficultés et ils étaient filmés par un caméraman. À l'issue de l'expérimentation, les sujets étaient reçus en entretien par un sociologue.

Évaluation et méta-évaluation

L'objectif de l'expérimentation pour notre équipe était donc double : évaluer l'utilisabilité du dispositif et comparer deux méthodes d'évaluation selon les types de données disponibles. Nous avons classé ces données disponibles selon deux types. Les traces dites « terrain » se réfèrent aux données habituellement prises en compte sur le terrain, c'est-à-dire les enregistrements audio-vidéo du sujet et du dispositif ainsi que le film de son écran. Ces données sont caractérisées par leur grande richesse, mais aussi par le fait qu'il est très difficile de les analyser automatiquement. Au contraire, les traces

⁶ Pour l'anecdote, les expérimentateurs se sont parfois eux aussi perdus dans les couloirs...

dites « situation réelle » regroupent tous les événements issus de l'interaction entre l'utilisateur et le système (appuis sur les boutons, sélections de commandes, affichages, sons). Ces données, peu riches et le plus souvent de faible niveau d'abstraction, sont peu lisibles manuellement mais plus aisées à analyser automatiquement.

Nous avons cherché à déterminer le pouvoir d'analyse de chaque type de trace, mais aussi et surtout à déterminer si les problèmes ergonomiques découverts à l'aide des traces « situation réelle » peuvent être confirmés par les traces « terrain » dont la méthodologie est bien connue. Notre objectif était de déterminer s'il était possible de s'abstraire des traces « terrain », difficiles à obtenir en situation réelle, sans pour autant perdre la validité des résultats obtenus. Pour cela, nous avons suivi une démarche en deux phases :

1. **Analyses en aveugle :** Nous avons analysé les traces « terrain » (film d'écran, notes et fichiers audio-vidéo) et les traces « situation réelle » (actions de l'utilisateur sur le système et retours d'information du système) de manière indépendante sans avoir connaissance des informations issues de l'autre type de traces. L'objectif était de déterminer ce que l'on peut déduire de chacun des deux types de traces pris isolément, c'est-à-dire leur réel pouvoir d'analyse.
2. **Croisement des résultats :** À partir des conclusions des deux analyses précédentes, nous avons croisé les résultats obtenus afin de détecter les cohérences, les incohérences, et les résultats complémentaires. Nous avons ensuite catégorisé les résultats selon les fonctions étudiées, les critères ergonomiques analysés, etc.

En pratique, les deux types de traces étaient recueillis tout au long de l'expérimentation (à l'exception des images du caméscope) dans une régie improvisée dans un bureau du bâtiment, via des technologies sans fil (HF et WiFi). Cette régie n'était pas montrée aux sujets de l'expérimentation (cf. images 3, 4 et 5).

Évaluations « terrain » et « situation réelle »

Aucune consigne particulière n'avait été donnée aux évaluateurs concernant le fond ou la forme des résultats attendus. Il leur était simplement demandé de détecter le plus de problèmes d'utilisabilité possible, en se limitant impérativement à l'usage d'un seul des deux sous-ensembles de données disponibles. Les résultats dits « terrain » ont été obtenus à partir des données suivantes :

- Les enregistrements audio et vidéo du contexte réalisés par le caméraman à l'aide d'un caméscope grand public ;
- L'enregistrement audio des commentaires du sujet obtenus via un micro-cravate sans fil et enregistrés sur un ordinateur portable situé dans la régie ;
- L'enregistrement du film d'écran du PDA utilisé par le sujet, obtenu via une liaison réseau sans fil et enregistré sur un ordinateur portable situé dans la régie.

Les commentaires du sujet étaient écoutés (via un casque) et annotés au fil de l'eau par une ergonome qui disposait en plus du film d'écran du PDA en temps réel. Cette ergonome se trouvait dans la régie et donc hors de la vue du sujet. Un rapport a ensuite été rédigé à partir de ces notes en se basant notamment sur les critères ergonomiques classiques [Bastien & Scapin, 1995].



Images 3, 4 et 5 : sur l'image en haut à gauche, régie permettant le recueil des traces « terrain » (à gauche) et « situation réelle » simulée (à droite). Sur l'image en bas à gauche, ergonomiste effectuant les annotations au fil de l'eau. Sur l'image de droite, copie d'écran des traces « situation réelle » recueillies pendant l'expérimentation.

Les résultats dits « situation réelle » ont été obtenus à partir des traces informatiques collectées sur le dispositif. Les traces étaient transmises en temps réel par le réseau sans fil via un bus de données développé spécifiquement (Usybus basé sur Ivy [Buisson et al., 2002]). Les traces collectées étaient :

- Les actions de l'utilisateur sur l'interface du dispositif ;
- Les réactions de l'interface du dispositif perceptibles par l'utilisateur ;
- La géolocalisation du dispositif par triangulation WiFi.

Un ingénieur était chargé de surveiller le bon déroulement de l'acquisition et de l'enregistrement des données. Ces traces ont été stockées puis analysées a posteriori, principalement en construisant des matrices de transition et en utilisant des algorithmes de détection de différences en les comparant à une expérimentation de référence (sans erreur) effectuée par un membre de l'équipe.

Évaluation croisée « terrain » versus « situation réelle »

De manière globale, nous avons constaté, en première lecture, que le nombre de problèmes ergonomiques détectés grâce aux traces « terrain » est bien plus important que le nombre de problèmes détectés par l'analyse des traces « situation réelle ».

Les problèmes détectés à partir des traces « terrain » sont de nature absolue au sens où ces problèmes se réfèrent aux critères ergonomiques ou heuristiques connus. À

l'inverse, les traces « situation réelle » détectent des problèmes le plus souvent par comparaison entre différentes passations ou par rapport à une passation de référence. On peut donc les considérer comme plus relatives.

Le niveau d'abstraction des problèmes détectés par les traces « situation réelle » peut sembler au premier regard d'un faible niveau (cosmétique). En fait, ces traces permettent de détecter des anomalies au niveau des interactions, mais les problèmes détectés sont souvent de plus haut niveau (par exemple : des concepts mal compris). Ce faible niveau d'abstraction est aussi lié au fait que les traces enregistrées ne possédaient pas d'informations sur les tâches en cours (limite connue des traces « situation réelle »).

Les deux méthodes ont clairement montré que les utilisateurs n'ont pas compris le principe de devoir indiquer au système la présence d'un obstacle sur le chemin prévu (sur le terrain grâce aux commentaires / en situation réelle grâce à la détection d'une durée d'accomplissement de la tâche trop importante). De même, l'utilisation d'un point de passage lors de la spécification du routage est peu utilisée par les sujets (sur le terrain grâce aux commentaires / en situation réelle grâce à la détection d'actions non réalisées). L'incompréhension par les utilisateurs du concept de message géolocalisé a été clairement mis en évidence par les deux méthodes (sur le terrain grâce aux commentaires / en situation réelle grâce à la détection d'actions non réalisées).

Les traces « terrain » ont permis, principalement grâce aux commentaires des utilisateurs, de détecter de nombreux problèmes difficiles à identifier par d'autres moyens, par exemple : la signification parfois ambiguë des icônes, l'identification incorrecte de la modification du routage, l'orientation de la carte, la compréhension des messages affichés, etc. De manière générale, ces problèmes ont trait à l'interprétation de la part des sujets d'informations qui, même si elles ne sont pas correctement interprétées, ne provoquent pas de modification notable de la séquence d'interaction. Ceci est en partie lié au biais introduit dans le protocole par le rôle du facilitateur, qui, une fois le problème exprimé par le sujet, aidait celui-ci à poursuivre son interaction de manière correcte. Ainsi, c'était uniquement cette interaction correcte qui était enregistrée dans les traces « situation réelle », masquant de ce fait le problème rencontré par le sujet.

Un faux négatif a été découvert sur l'analyse de l'utilisation du guidage. En effet, trois des utilisateurs ont confondu la commande « modifier guidage » avec la commande « reprendre guidage ». Cette variation a été détectée, mais finalement jugée normale dans le processus naturel d'exploration du logiciel. Or, la confusion était bien réelle. Cette interprétation trop optimiste a été favorisée par le grand nombre de variations inter-utilisateurs détectées dans l'utilisation de la page concernée.

À l'opposé, trois types de faux positifs ont été mis à jour. Le premier type concerne la détection d'un temps anormalement long lors de l'indication d'un obstacle sur le parcours. En effet, un sujet a mis quasiment trois minutes pour acquiescer le message d'avertissement. Un problème d'utilisabilité sérieux a donc été soupçonné. En fait, le message est arrivé un peu trop tôt dans le scénario tandis que le sujet et le facilitateur étaient engagés dans une discussion à propos de la précédente étape du scénario. Ici, le biais vient de la conjonction d'une limitation du prototype et du rôle du facilitateur. Le deuxième type de faux positif a été provoqué par la seule limite du prototype. En effet, certaines actions affichées n'étaient pas suivies des actions attendues, par exemple pour la commande « annuler ». Ainsi, nous avons détecté des appuis répétés sur cette commande, ce qui nous a fait soupçonner un problème de libellé mal compris ou de lenteur de réaction du dispositif. En fait, l'utilisateur pensant qu'il n'avait pas appuyé la première fois, a répété son action. Le troisième type de faux positif concerne la

détection d'actions manquantes. On a supposé, dans ce cas, que l'utilisateur n'avait pas suivi la trajectoire d'interaction prévue. En fait, un problème de transmission sur le réseau en était responsable.

Synthèse

En résumé, les traces obtenues en situation réelle ne permettent pas de détecter autant de problèmes d'utilisabilité que les traces obtenues sur le terrain. Ceci est particulièrement vrai en ce qui concerne les difficultés d'interprétation. Cette affirmation est à modérer du fait des biais introduits par le protocole (rôle du facilitateur et scénario très linéaire). Cependant, on peut se douter que les traces obtenues en situation réelle, même si elles détectent un nombre significatif de problèmes, ne permettront pas d'explicitier de manière fiable la cause de ces problèmes. En effet, le manque d'informations sur le contexte, que l'on obtient généralement grâce aux enregistrements audio-vidéo, reste un problème délicat qui peut engendrer à la fois des faux négatifs et des faux positifs.

Or, se limiter aux traces obtenues en situation réelle est très intéressant dans les situations où l'enregistrement de vidéos du contexte est difficile ou même non souhaitable. C'est pourquoi les informations de contexte qui font défaut doivent être reconstituées par d'autres moyens. Pour cela, nous proposons de récupérer le maximum de données du contexte par des capteurs additionnels (centrale d'attitude, géolocalisation, état du noyau fonctionnel de l'application, film d'écran) de manière à combler ce manque informationnel et ainsi reconstituer des bribes de contexte.

4.2. Expérimentation « E-skiing »

E-skiing est un service mobile destiné à enrichir la pratique du ski via l'envoi proactif aux skieurs, sur le terrain et peu de temps après leurs descentes, de données concernant celles-ci. Ces données permettent aux skieurs de revivre leurs précédentes descentes, via une vidéo et un ensemble de grandeurs numériques liées à leurs performances (trajectoire, vitesse, accélération, etc.). Pendant l'expérimentation, les informations issues des capteurs portés par le skieur étaient enregistrées pour chaque descente, traitées en fin de descente, puis les skieurs, de retour sur les pistes, étaient prévenus de manière proactive que de nouvelles informations sur leurs précédentes descentes étaient disponibles. Ils pouvaient alors consulter individuellement, grâce au smartphone (cf. images 6 à 10) : le chemin parcouru, la distance et le temps de descente, la vitesse maximale, le coefficient d'engagement (lié à l'accélération) et la vidéo de leur descente.

L'objectif de cette deuxième expérimentation était triple : (1) tester l'ergonomie de l'interface de restitution, (2) tester l'usage fait par un groupe de skieurs d'un service mobile proactif et (3) valider les aspects techniques et méthodologiques d'une évaluation en situation réelle.

Protocole expérimental

Pour cette expérimentation, nous avons recruté un groupe de skieurs ayant l'habitude de pratiquer ensemble, ceci afin de permettre également une approche sociologique des interactions dans le groupe. Ces skieurs étaient habitués à l'usage des nouvelles technologies et n'ont pas eu de difficulté à utiliser le matériel. Après une présentation des capteurs et des possibilités du service par une démonstration, les sujets

étaient invités à aller skier comme ils en avaient l'habitude, puis à revenir au local en bas des pistes après chaque descente (pour le chargement des données). Il s'agissait d'une expérimentation en situation réelle car les skieurs étaient libres d'effectuer les activités qu'ils désiraient et n'étaient pas accompagnés d'observateurs. L'expérimentation a mobilisé cinq skieurs dont quatre étaient équipés du service. Dans une seconde partie de l'expérimentation, qui n'est pas l'objet de cet article, les skieurs ont été conviés à une réunion sous forme de soirée conviviale de type « focus-group » où la version sédentaire du service leur était présentée, celle qui est destinée à être utilisée à la maison. Cette réunion avait également pour objectif d'obtenir un retour sur les aspects ergonomiques du système et sur l'usage qu'ils avaient fait du service proposé.



Images 6, 7 et 8 : Casque équipé d'une mini-caméra et d'un accéléromètre (à gauche), téléphone mobile implémentant l'interface de restitution mobile (à droite) et groupe de skieurs équipés (au centre).



Images 9 et 10 : Écran du téléphone mobile implémentant l'interface de restitution (à gauche) et copie d'écran de cette interface (à droite).

Techniquement, le service E-skiing était composé d'un ensemble de capteurs associés à des enregistreurs autonomes (vidéo, accélération, et position géographique) portés par les skieurs et d'une interface de restitution (téléphone mobile de type « smartphone ») qu'ils emmenaient aussi avec eux (cf. images 6 à 10). En back-office, nous disposons de deux ordinateurs portables destinés à traiter les données et d'un serveur de données permettant de mettre à disposition les données une fois traitées. La récupération des données s'effectuait à la fin de chacune des descentes car les enregistreurs autonomes n'avaient pas la capacité de se connecter sans fil. Par contre,

les messages proactifs et l'accès aux performances par les skieurs s'effectuaient via le réseau mobile (sauf les vidéos pour une raison de coût). Les détails des aspects techniques de l'expérimentation peuvent être consultés dans [Jambon, 2006] et [Jambon & Meillon, 2009].

Nous n'avons testé qu'un seul groupe de skieurs car l'objectif principal n'était pas d'obtenir des résultats statistiquement pertinents, mais s'approchait plus du test d'un démonstrateur de concept. Nous avons mis en œuvre la technique du « cheval de Troie » pour les caméras et les GPS portés par les skieurs. En effet, chacune des caméras permettait à la fois de filmer la descente des skieurs et leurs interactions avec les téléphones mobiles. Chacun des GPS permettait de calculer le parcours, mais aussi de localiser les lieux d'usage du service.

Principaux résultats

Le premier résultat de cette expérimentation est d'ordre technique. Cette expérimentation mettait en œuvre en situation réelle des technologies très innovantes et donc parfois peu fiables... Nous avons anticipé les problèmes techniques en effectuant de nombreux tests en laboratoire, en extérieur, puis sur les lieux. Malgré ces tests, les enregistrements vidéos se sont révélés quasiment inexploitable pour l'analyse des interactions du fait d'un mauvais cadrage : le faible angle de vue des mini-caméras et les déplacements des casques au cours des sauts effectués par les skieurs ont eu pour conséquence de placer hors champ les écrans des smartphones. Cet aspect n'avait pas été détecté lors des tests, car ceux-ci s'effectuant selon des tâches prescrites, les skieurs n'avaient pas fait de sauts comme ils en avaient pourtant l'habitude...

C'est pourquoi l'utilisabilité du service n'a pas pu être étudiée par ce moyen. Elle a été évaluée dans la seconde partie de l'expérimentation via des questions informelles posées au cours du focus-group. Nous avons été très surpris de constater, que, malgré une interface homme-machine développée très rapidement et sans attention particulière à sa qualité ergonomique, les utilisateurs n'ont fait aucune critique. Or, de manière évidente, l'ergonomie était largement perfectible. C'est en soi un résultat intéressant : les utilisateurs se sont focalisés sur le service et ont occulté les aspects ergonomiques qu'ils considéraient comme secondaires. Ainsi, il faut garder à l'esprit qu'une évaluation ergonomique réalisée en situation réelle, via des questionnaires, peut conduire à minimiser les problèmes d'utilisabilité réels du dispositif.

Malgré l'absence d'enregistrements vidéo utilisables, la géolocalisation et les logs des envois et réceptions de données nous ont permis d'obtenir des informations très intéressantes concernant l'usage du service proposé. L'activité des skieurs a été reconstruite à partir de ces informations qui avaient été enregistrées en premier lieu pour faciliter la mise au point du système. Nous avons ainsi pu déterminer les moments où les skieurs étaient prévenus par les relevés de SMS et où ils accédaient à l'information via les logs du serveur de données. Les enregistrements GPS nous ont ensuite permis d'en déduire les lieux. Nous nous sommes ainsi aperçus du très faible usage du dispositif en général, et de son utilisation quasi exclusivement hors des pistes. Ce résultat a priori surprenant a été explicité lors du focus group. Les skieurs étant venus pour skier, c'est effectivement ce qu'ils ont fait ! Ensuite, lors de pauses, ils ont parfois pris le temps de regarder si le service pouvait être intéressant. Ils ont également détourné l'usage de l'enregistrement vidéo, en se filmant entre eux au lieu d'utiliser la caméra pour filmer leur propre descente. Ces résultats mettent en lumière l'intérêt de l'approche en situation

réelle : les utilisateurs vivent leur vie et n'utilisent le système à leur disposition que s'ils en sentent le besoin. Nous sommes très loin des évaluations sur le terrain.

Les difficultés les plus importantes que nous avons rencontrées ont été d'ordre logistique et méthodologique. Le traitement des données en cours d'expérimentation a transitoirement saturé l'équipe d'observateurs stationnée au chalet. En effet, nous n'avions pas anticipé le fait que les skieurs reviendraient très régulièrement et ensemble au chalet. L'important volume de données à traiter rapidement, ajouté au fait que les enregistreurs ont parfois été mélangés entre les skieurs, a fait régner une certaine confusion. En outre, le protocole imposait la mise en route des enregistreurs par les skieurs eux-mêmes. Or, les consignes étant réduites, tous les skieurs n'ont pas toujours compris la procédure et ont parfois oublié les mises en route. De plus, ils n'ont pas osé utiliser les téléphones mobiles pendant certaines phases d'acquisition de données, pensant que ceux-ci n'étaient pas disponibles. Ce dernier point a été identifié comme un biais, mais les skieurs nous ont indiqué que cela n'avait eu aucune conséquence sur leur usage du service.

Synthèse

Ces expérimentations ont montré que la mise en œuvre d'une évaluation en situation réelle est véritablement très complexe aussi bien sur le plan technique que méthodologique. Nous avons appris à cette occasion qu'il est impératif de ne faire aucune concession sur ces deux aspects. En d'autres termes, le dispositif testé, comme les systèmes d'acquisition, doivent fonctionner de manière fiable, robuste, être autonomes et avoir une gestion automatique. Il ne faut pas déléguer une partie, même faible, du protocole aux utilisateurs. Ceci est une contrainte très forte, car les systèmes aujourd'hui disponibles pour les tests sont souvent des prototypes peu robustes et peu intégrés.

Nous avons également montré que la situation réelle est peu propice à la détection des problèmes d'utilisabilité, mais que de très intéressants résultats concernant l'usage réel des systèmes peuvent être aisément obtenus à partir des traces. Les résultats détaillés de cette expérimentation peuvent être consultés dans [Jambon, 2006]. Suite à cette expérimentation, nous avons cherché à valider un protocole expérimental adapté à l'évaluation en situation réelle d'un système ubiquitaire sur une longue durée et avec un grand nombre d'utilisateurs, afin de pouvoir disposer de données statistiquement intéressantes.

4.3. Expérimentation « Muséum »

L'expérimentation « Muséum » s'est déroulée pendant l'exposition temporaire « ni vu – ni connu »⁷ au Muséum du département du Rhône. Cette exposition, dont le thème était le camouflage, se proposait d'enrichir le parcours du visiteur d'une expérience sur les concepts de vie publique et privée vis-à-vis de paparazzi virtuels. Au cours de son parcours dans l'exposition, le visiteur était amené à fournir des informations personnelles et à être photographié. À la fin de l'exposition, il lui était présenté la première page d'un journal à scandales fictif fusionnant les informations recueillies à son insu (cf. figure 1). Le visiteur était ensuite informé des techniques utilisées pour le « traquer » et des problématiques liées au respect de la vie privée.

⁷ http://www.museum-lyon.org/expo_temporaires/ni_vu_ni_connu/

La demande d'évaluation de l'exposition était issue en premier lieu du scénographe de l'exposition, lequel désirait connaître l'affluence et plus précisément l'utilisation effective des bornes interactives ainsi que la fiabilité des techniques mises en œuvre pour implémenter la scénarisation. Plus généralement, les responsables des musées se préoccupent de connaître le comportement de leurs visiteurs selon les expositions proposées [Gob & Drouguet, 2006] et désirent également savoir comment les dispositifs interactifs installés sont utilisés. Or ces informations, pourtant basiques, sont très rarement disponibles et fiables. En effet, même si cela semble a priori simple, il n'est pas aisé d'analyser le parcours de visiteurs, car il ne suffit pas d'effectuer des comptages aux points de passage, il faut également pouvoir distinguer les visiteurs les uns des autres.

Notre objectif, par rapport à la précédente expérimentation E-skiing, était de prouver la possibilité de passage à l'échelle des expérimentations en situation réelle. Nous avons donc augmenté la durée de l'expérimentation : quelques semaines au lieu d'une seule journée pour les skieurs. Nous avons aussi augmenté le nombre d'utilisateurs : plusieurs centaines de visiteurs au lieu d'un seul groupe de quatre skieurs. Cela nous a permis d'obtenir des résultats statistiquement exploitables.



Figure 1 : Étapes du traçage des visiteurs dans l'exposition « ni vu – ni connu ».

Protocole expérimental

Le fonctionnement de cet environnement ubiquitaire se basait sur la technologie des étiquettes radiofréquence (RFID). Lors de son entrée dans le musée, le visiteur se voyait remettre, en même temps que son ticket d'entrée, une carte avec un texte à trous. Une étiquette RFID était dissimulée dans cette carte et des lecteurs d'étiquettes étaient dissimulés dans l'exposition, notamment à la borne de jeu, au passage devant des paparazzi fictifs (prise de photo), et à la borne de sortie. Reliés au système d'information centralisé du musée, ils permettaient, en croisant les données, de faire ressentir au visiteur un réel sentiment d'avoir été « traqué ».

L'objectif pour notre équipe étant le passage à l'échelle, nous avons choisi d'utiliser des technologies moins innovantes, mais beaucoup plus robustes que celles mises en œuvre pour l'expérimentation E-skiing. Nous avons réutilisé l'infrastructure disponible pour l'exposition, en la complétant. Pour cela, nous avons ajouté un lecteur d'étiquettes RFID à l'entrée de l'exposition, de manière à obtenir un horodatage fiable du point d'entrée. Un autre lecteur aurait dû se trouver en sortie de l'exposition et encore d'autres à certains points de passage. Cependant, les contraintes techniques et les aspects liés à la sécurité des personnes (largeur minimum des sorties) ne nous ont pas permis de les installer. Le système d'information du musée a également été modifié de manière à permettre l'enregistrement et le téléchargement journalier des données. Les consignes données pour cette expérimentation étaient réduites au minimum. Les caissières à

l'entrée du Musée devaient simplement donner une carte par visiteur en même temps que le ticket d'entrée. La carte elle-même portait juste l'indication : « gardez cette carte jusqu'à la sortie de l'exposition... elle vous servira. ».

La technologie des étiquettes RFID avait déjà été mise en œuvre dans plusieurs musées afin d'enrichir la visite d'une exposition (par exemple [Hsi & Fait, 2005]). Cependant, à notre connaissance, cette technologie n'avait pas encore été utilisée dans le but d'évaluer de manière systématique l'usage fait par le visiteur de l'environnement ubiquitaire mis à sa disposition. Ainsi, nous avons, tout comme dans l'expérimentation E-skiing, appliqué la technique du cheval de Troie en détournant un dispositif présent dans l'environnement pour s'en servir comme dispositif de capture d'information sur le comportement de l'utilisateur. Des données à caractère personnel étant enregistrées, une déclaration à la CNIL a été effectuée. La CNIL était également partenaire de l'exposition.

Principaux résultats

L'expérimentation s'est déroulée sans trop de difficultés d'ordre technique. Précisons néanmoins que l'évaluation, qui devait durer initialement trois mois, a été réduite aux trois dernières semaines de l'exposition du fait de retards dus aux réglages des lecteurs et à l'impression des cartes contenant les étiquettes RFID. Cela nous a confirmé que les aspects logistiques en situation réelle sont un véritable défi. La limitation technique la plus contraignante a été la faible distance de détection pratique des étiquettes RFID, loin des performances supposées.

L'analyse des données s'est tout d'abord focalisée sur le filtrage des données incohérentes dues soit à des défaillances techniques, soit au non-respect de consignes lors de la distribution des cartes. Ensuite, un premier niveau d'analyse nous a permis de déterminer l'affluence de l'exposition : sur la période des trois semaines de l'expérimentation, 492 cartes ont été distribuées à des visiteurs. Un second niveau d'analyse a eu pour objectif de déterminer les temps de parcours et les types de parcours. Nous avons ainsi pu déterminer que près de la moitié des visiteurs ont effectué le parcours complet tel qu'il était prévu par le scénariste (cf. figure 2). Les résultats détaillés de cette expérimentation peuvent être consultés dans [Jambon, Mandran et al., 2007].

Les biais potentiels liés au protocole expérimental ont été plus délicats à gérer, car nous n'avons pu nous apercevoir de leur existence qu'au moment de dépouiller les données. Dans le protocole expérimental en situation réelle, l'utilisateur ne reçoit pas de consignes spécifiques à l'évaluation, il n'y a donc pas de biais lié à cette consigne. Cependant, les personnels du musée peuvent apporter involontairement des biais. Par exemple, une des personnes de l'accueil n'a pas compris les consignes et n'a pas distribué de carte à tous les visiteurs. Ainsi, les données obtenues certains jours ne recensent pas l'ensemble des comportements des visiteurs. L'échantillon final que nous avons analysé peut donc éventuellement être biaisé, mais nous n'avons pas de moyen de contrôle ni d'estimation de ce biais.

Lors de l'analyse des données, l'absence d'enregistrements vidéo a rendu délicate l'interprétation des anomalies dans les parcours. Dans certains cas, ce sont les commentaires des personnels du musée, notamment les gardiens, qui nous ont permis de découvrir les raisons des comportements insolites détectés. Ce problème est générique à toute expérimentation en situation réelle lorsque l'on ne dispose pas de vidéo. En effet, le faible niveau sémantique des traces recueillies peut rendre difficile l'interprétation

des comportements, car ce travail d'interprétation repose sur la capacité à inférer un comportement à partir des actions de bas niveau. Ceci se produit notamment dans les cas où ce comportement de l'utilisateur n'a pas du tout été anticipé.

Synthèse

Cette expérimentation a montré qu'obtenir des informations statistiquement exploitables en situation réelle est tout à fait envisageable, même sur une large échelle. Elle a également confirmé la complexité technique et logistique de ce type d'expérimentation.

Il était initialement prévu, dans le protocole d'expérimentation, d'analyser l'utilisabilité des bornes, grâce aux traces d'interactions enregistrées en même temps que les parcours, et selon une méthode proche de celle utilisée pour MapMobile. Le musée était principalement intéressé par les parcours, et faute de temps, cette étude n'a été que partiellement réalisée.

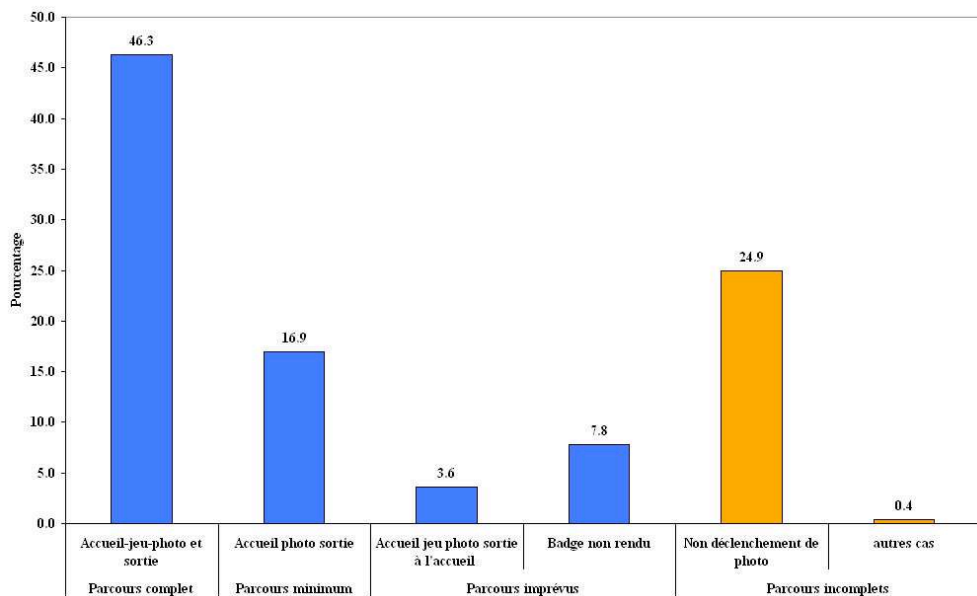


Figure 2 : Types de parcours empruntés par les visiteurs de l'exposition « ni vu – ni connu ».

5. DISCUSSION ET PROPOSITION DE TAXONOMIE

5.1. Retour d'expérience concernant la situation réelle

Les expérimentations en situation réelle nous ont apporté des résultats parfois surprenants, au sens où ils n'étaient pas anticipés, mais de ce fait aussi très pertinents. C'est probablement là leur principal intérêt : mettre en lumière des comportements non attendus des utilisateurs liés à leur activité réelle ou au contexte réel. Ces résultats ont été parfois aussi déroutants et même décevants comparés à l'investissement réalisé. Par exemple, le service E-skiing avait demandé des développements complexes destinés à le rendre compatible avec le réseau de téléphonie mobile, alors qu'au final, une simple borne WiFi associée à un portable situé près du bar aurait suffi... Mais personne, dans

le projet, n'avait anticipé cet usage a priori, bien que cela puisse paraître évident a posteriori.

Nous avons mis en lumière que la détection des problèmes est relativement aisée, mais que leur explicitation requiert toujours la recherche d'informations complémentaires, soit auprès des utilisateurs eux-mêmes en leur faisant revivre une partie de l'expérimentation, soit auprès des témoins naturellement présents dans le contexte lors des expérimentations. A contrario, les consignes sont souvent très aisées à définir car elles sont réduites à leur plus simple expression : une explication du fonctionnement du système.

Nous nous sommes aperçus d'une contrainte du protocole : une fois l'expérimentation lancée, il est très difficile de récupérer les défaillances techniques ou les erreurs commises dans le protocole. Si quelque chose a été oublié, c'est souvent trop tard car on ne peut pas modifier un dispositif ou le protocole sans risquer de perturber le contexte de l'expérimentation et ainsi introduire un biais. Il faut soit laisser faire en acceptant une dégradation des conditions expérimentales, soit prendre la décision d'arrêter toute l'expérimentation. Ce point est important car les expérimentations en situation réelle sont souvent complexes, et donc requièrent d'importants moyens matériels et humains.

En outre, les aspects techniques et logistiques sont complexes, mais tout à fait gérables à condition d'avoir une organisation adaptée et une bonne gestion du temps. Dans tous les cas, les expérimentations en situation réelle requièrent beaucoup plus d'efforts que des expérimentations en laboratoire, avec un différentiel que nous estimons à un facteur deux à dix selon le contexte. Bien entendu, de nombreux tests, dont au moins un en situation quasiment identique, sont nécessaires. Il est illusoire de croire qu'il est possible de tout prévoir sur le papier, car l'intérêt de la situation réelle est justement l'occurrence des imprévus.

Enfin, il est primordial de toujours enregistrer le plus de données possible, même a priori redondantes ou secondaires. En effet, il se produira probablement des pannes des systèmes d'enregistrement ou plus probablement il apparaîtra des situations inédites, et il faudra reconstituer ces situations à partir d'indices issus des données que l'on avait jugées, au départ, secondaires.

5.2. Généralisation : les quatre approches

En prenant comme point de départ la distinction entre laboratoire et terrain, usuelle dans la littérature, nous avons proposé la notion de situation réelle comme une spécialisation du terrain. Or, la recherche d'une classification de ces configurations nous a amenés à définir une autre configuration, la « simulation réaliste ». En outre, nous avons associé à l'ensemble de ces configurations, de nouvelles appellations, basées sur une terminologie inspirée de la biologie, plus précise et, de plus, cohérente avec les propositions de Kjeldskov et al. [Kjeldskov & Skov, 2007].

Le laboratoire (in-vitro)

Le laboratoire d'utilisabilité est la configuration de référence. Il se caractérise notamment par un environnement artificiel contrôlé. Ainsi, les problèmes d'utilisabilité détectés seront directement dépendants des tâches effectuées par l'utilisateur, et donc des scénarios choisis. Il est donc peu probable que des éléments liés à l'usage réel seront découverts.

Nous nommons désormais cette configuration « in-vitro » en accord avec les propositions de Kjeldskov et al. [Kjeldskov & Skov, 2007].

Le terrain (in-situ)

Comparé au laboratoire d'utilisabilité, l'intérêt principal du terrain est d'apporter les éléments du contexte qui vont rendre l'évaluation plus pertinente. Le protocole expérimental étant peu différent du laboratoire d'utilisabilité, notamment la tâche de l'utilisateur n'étant le plus souvent pas libre, les problèmes ergonomiques détectés dépendront directement du choix des scénarios donnés dans la consigne. Ainsi, il n'est pas surprenant que les problèmes détectés sur le terrain soient similaires en type et en nombre dans ces deux situations. De plus, la présence éventuelle du facilitateur apporte un biais connu dans le déroulement de l'activité du sujet.

Si l'environnement est peu contraint, le terrain peut apporter son lot d'événements permettant de rendre l'évaluation un peu plus réaliste, même si le sujet est souvent isolé du monde extérieur par la « bulle » formée par la présence des observateurs. C'est donc avant tout des problèmes d'utilisabilité que l'on va détecter, mais resitués dans le contexte d'usage réel du système. Par exemple, si un écran de téléphone mobile est inadapté à la luminosité extérieure, ce problème ne sera pas détecté en laboratoire d'utilisabilité, mais le sera sur le terrain, à condition qu'il fasse beau... C'est là l'une des limitations de cette configuration. En effet, les éléments du contexte étant par nature non prévisibles, rien ne garantit qu'ils se produiront. Le risque est alors d'obtenir des résultats biaisés car certaines variables indépendantes ne sont pas prises en compte.

Nous nommons désormais cette configuration « in-situ » en accord avec les propositions de Kjeldskov et al. [Kjeldskov & Skov, 2007].

La situation réelle (in-vivo)

Concernant les expérimentations en situation réelle, le protocole expérimental change radicalement. En effet, le facilitateur ne peut plus intervenir et l'utilisateur n'a, a priori, que des tâches libres. Il va donc « vivre sa vie » avec le système. Deux types de résultats seront obtenus : des statistiques d'usage du dispositif, c'est-à-dire les tâches que l'utilisateur a souhaité effectuer, et l'utilisabilité « filtrée par l'usage », c'est-à-dire les problèmes ergonomiques liés aux tâches effectuées dans le contexte d'usage. Ces tâches étant un sous-ensemble des tâches possibles, il est cohérent que le nombre de problèmes ergonomiques détectés en situation réelle soit bien moindre qu'en laboratoire d'utilisabilité ou que sur le terrain. Néanmoins, les problèmes ergonomiques détectés gagnent en pertinence car ils ont été mis en lumière par l'usage réel du dispositif.

Cependant, les résultats doivent être pris avec circonspection, car les tâches effectuées par les sujets et les événements du contexte ne se produisent pas de manière déterministe, et constituent ainsi des variables indépendantes masquées. Pour minimiser ce risque, il peut être intéressant, si cela est réalisable, d'augmenter significativement le nombre de sujets et le temps d'expérimentation, en espérant que les événements statistiquement probables se produiront. De plus, le fait même que l'utilisateur ait connaissance de faire partie d'une évaluation apporte a priori un biais, connu sous le nom de « effet Hawthorne » [Macefield, 2007]. Cet effet, bien que très controversé, suggère que la performance des utilisateurs s'améliore du fait même qu'ils aient conscience de faire partie d'une expérimentation.

Nous nommons désormais cette configuration « in-vivo » afin d'être cohérent avec les autres appellations. Remarquons que Kjeldskov et al. ne mentionnent pas cette configuration dans leur article [Kjeldskov & Skov, 2007].

La quatrième configuration : la simulation réaliste (in-simu)

Dans l'inspiration du « Truman Show »⁸, il existe une quatrième configuration. Celle-ci consiste à avoir un contexte artificiel, mais réaliste, et des tâches pouvant être libres, prescrites, ou encore basées sur des scénarios réalistes. Cette configuration ne peut pas être adaptée à toutes les situations car il est nécessaire de pouvoir simuler correctement le contexte, et la durée des tâches doit être compatible avec l'usage du simulateur. Deux auteurs ont exploré cette configuration sous des appellations très différentes. Kjeldskov et al. sous l'appellation « in-sitro » [Kjeldskov & Skov, 2007] et Hertzum sous l'appellation « workshop » [Hertzum, 1999]. Dans ces deux publications, les résultats obtenus en simulation se rapprochent de ceux obtenus sur le terrain.

Cette approche est à rapprocher des nouvelles techniques de type LOFT (« Line Oriented Flight Training ») utilisées en formation et entraînement des pilotes de l'aviation civile, consistant à simuler non pas uniquement des séries de situations d'urgence, mais des parties de vol réalistes où se mélangent situations normales, situations exceptionnelles et situations d'urgence. D'un certain point de vue, cela peut également se rapprocher de la notion de « living laboratory » dont la définition exacte n'est pas encore bien stabilisée. Pour nous, nous considérons cette configuration comme une situation réelle dégradée où le contexte est en partie contraint.

Nous nommons désormais cette configuration « in-simu » afin d'être cohérent avec les autres appellations. Kjeldskov et al. mentionnent cette configuration sous le nom de « in-sitro » [Kjeldskov & Skov, 2007], car elle y est présentée comme intermédiaire entre « in-vitro » et « in-situ ». Nous n'avons pas conservé ce terme car nous l'avons jugé peu explicite, et de plus, notre classification est différente et ne situe pas cette configuration à la même place.

5.3. Proposition de classification

Considérant les quatre configurations possibles, la tâche (prescrite, scénario réaliste ou libre) et le contexte (artificiel, simulé ou réel) semblent se dégager comme les deux principales grandeurs orthogonales. Nous avons masqué les aspects liés au réalisme du dispositif et à la présence des observateurs et dispositifs d'observation car, liés aux deux autres, ils nous ont semblés secondaires. En effet, le réalisme du dispositif n'est véritablement intéressant que lorsqu'il va de pair avec le réalisme du contexte, et la présence des observateurs n'est pas souhaitable lorsque le contexte est réel ou les tâches libres.

Nous déduisons ainsi neuf situations théoriquement possibles (figure 3). Nous les avons regroupées selon les quatre configurations définies précédemment: in-vitro, in-simu, in-situ et in-vivo. Globalement, c'est le réalisme du contexte qui détermine la configuration, avec pour le contexte réel, une distinction selon que les tâches sont libres ou non.

⁸ « The Truman Show » est un film relatant une émission de télé-réalité où le sujet vit, à son insu, dans un monde totalement artificiel où toutes les personnes qu'il rencontre sont des acteurs et où tous les événements se produisant sont déclenchés par un réalisateur.

A l'intérieur de chaque configuration, toutes les situations sont théoriquement possibles. Cependant, nous avons distingué les situations qui nous paraissaient les plus intéressantes (cercles noirs), celles intermédiaires (cercles gris) et celles qui nous semblaient moins intéressantes (cercles blancs).

Les situations les moins intéressantes sont les situations extrêmes où le protocole expérimental nous a semblé peu réalisable. Ainsi, l'usage de tâches libres en contexte artificiel nous a semblé peu pertinent car il peut être difficile pour un utilisateur d'agir de manière « naturelle » dans le contexte très contraint d'un laboratoire d'utilisabilité. De même, demander à un utilisateur d'exécuter scrupuleusement un ensemble de tâches prescrites en contexte réel, compte tenu des sollicitations externes, nous a semblé peu réaliste.

A l'opposé, nous pensons que les situations les plus intéressantes sont les situations qui présentent une bonne adéquation entre la prescription des tâches et le réalisme du contexte. En effet, le contexte très artificiel d'un laboratoire s'adapte bien à des tâches prescrites, le contexte simulé permet aux utilisateurs de suivre des scénarios réalistes dans un contexte proche de la réalité, tandis que les tâches libres ne se justifient vraiment qu'en contexte réel.

Les situations intermédiaires sont un compromis entre les deux précédentes. Remarquons que la configuration in-situ ne comporte pas de situation parmi les plus intéressantes. En effet, nous pensons que cette configuration n'est pas intéressante en elle-même, mais peut être un bon compromis lorsque les configurations in-simu ou in-vivo ne peuvent être mises en œuvre.

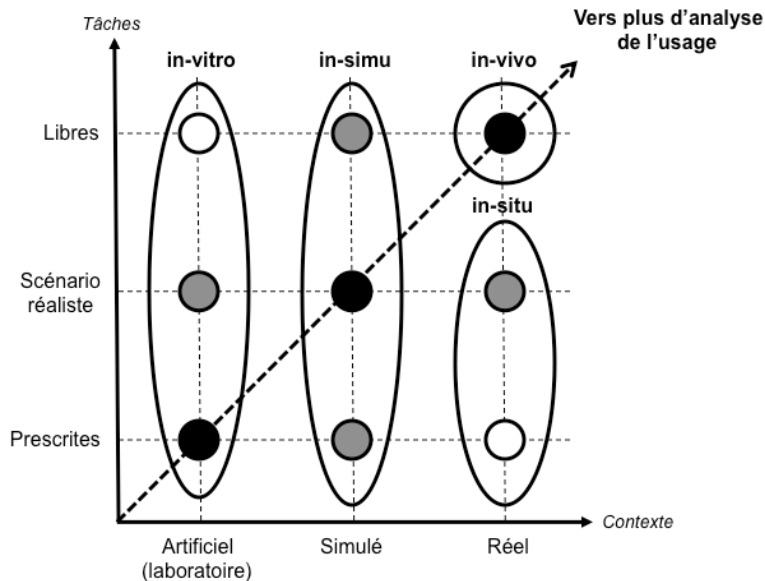


Figure 3 : Typologie des configurations en fonction du contexte et des tâches.

5.4. Quelle démarche adopter ?

Les quatre configurations définies ci-dessus nous semblent plus complémentaires qu'opposées. De notre point de vue, la configuration in-vitro permet d'étudier en détail les problèmes ergonomiques posés par le système interactif, selon des scénarios choisis.

Cette configuration se focalise sur l'utilisabilité. La configuration in-simu permet de détecter les fonctionnalités qui font sens pour l'utilisateur ainsi que certains problèmes d'utilisabilité, mais dans un environnement contrôlé. Enfin, la configuration in-vivo permet de finaliser l'étude en insistant sur les fonctionnalités qui font sens pour l'utilisateur dans le contexte réel où il est. Cette dernière configuration se focalise sur l'usage. Cas particulier, la configuration in-situ peut se substituer aux configurations in-simu ou in-vivo si, respectivement, le contexte est difficile à simuler ou si laisser l'utilisateur effectuer des tâches libres n'a pas de sens. Le choix d'une configuration pourra aussi être déterminé par les objectifs de l'évaluation ou par les limites techniques du système testé. En effet, in-vitro, il est possible d'utiliser des maquettes (par exemple via la technique du Magicien d'Oz), ce qui est quasiment impossible in-vivo où le système doit être perçu comme pleinement opérationnel (à de rares exception près [Consolvo et al., 2007]).

L'ordre « naturel » que nous proposons pour un dispositif donné consiste à effectuer les trois types d'évaluation dans l'ordre {in-vitro → in-simu → in-vivo} de manière à disposer progressivement d'un système dont les problèmes ergonomiques auront été éliminés avant de s'intéresser à son usage. Cette approche présente néanmoins un défaut : si certaines des fonctionnalités ne font pas sens à l'utilisateur, elles auront néanmoins été mises au point, à grands frais, lors des deux étapes précédentes. Il est possible aussi de prendre le contre-pied de cet ordre « naturel » et de suivre un ordre inverse : {in-vivo → in-vitro}. L'idée ici est de détecter les usages qui font sens en amont, avant de se concentrer ensuite sur leur mise au point. Le défaut de cette approche est qu'un problème ergonomique bloquant pour l'utilisateur va donner de faux résultats négatifs sur l'usage, c'est-à-dire qu'une fonctionnalité inutilisable ne sera tout simplement pas utilisée... De manière générale, il est également possible d'utiliser les quatre configurations de manière opportuniste, selon un cycle de développement plus général comme nous l'avons proposé sous la forme des « Moments de la conception » [Caelen, Jambon, & Vidal, 2005]. Ici, chaque type d'évaluation est un « Moment ». Libre aux concepteurs, en fonction de leurs besoins et de l'évolution du projet, de choisir les configurations et l'ordre dans lequel elles seront utilisées.

6. CONCLUSION ET PERSPECTIVES

Ces travaux nous ont permis de mettre en lumière la pertinence des expérimentations in-vivo, notamment lorsqu'elles sont associées à la technique du « cheval de Troie » qui permet la capture de données comportementales en minimisant les biais liés à l'observation. Cette configuration nous semble plus adaptée à détecter les problèmes liés à l'usage du dispositif qu'à son utilisabilité, où d'autres techniques, comme les expérimentations in-vitro, in-simu ou in-situ, moins complexes à mettre en œuvre, peuvent être utilisées. Nous avons également proposé une taxonomie des configurations d'expérimentation qui peut être vue selon deux dimensions principales : la tâche et le contexte. Ces travaux ont été effectués dans le cadre des études sur les systèmes mobiles et ubiquitaires, car ces systèmes requièrent une immersion dans leur contexte d'usage, mais ils sont probablement applicables dans le cadre des systèmes sédentaires plus classiques.

Actuellement, les perspectives de notre travail s'orientent vers deux directions. Premièrement, nous cherchons à affiner notre connaissance des différences entre les types de configurations, en s'intéressant à élargir la liste de nos critères de comparaison. Des pistes intéressantes peuvent être trouvées dans les travaux comparant les conditions

expérimentales, issus du domaine de l'analyse de comportements en économie (par exemple les travaux de Fiedler et al. [Fiedler & Haruvy, 2009]) ou encore du domaine des systèmes d'apprentissage basés sur la simulation (par exemple les travaux Dahl et al. [Dahl, Alsos, & Svanæligs, 2010]). Nous nous intéressons également à formaliser et rendre plus reproductible les comparaisons entre les configurations expérimentales, par exemple en se basant sur les critères de comparaison proposés par Hartson et al. [Hartson, Andre, & Williges, 2003], inspirés du domaine de la recherche d'information. En corolaire, les expérimentations avec tâches libres sont très souvent synonymes de très grandes quantités de données à traiter, et, qui plus est, des données de faible niveau d'abstraction qu'il faut analyser. Notre seconde perspective est d'automatiser l'analyse de ces données. Les projets TELEOS et TRACES qui s'intéressent respectivement à l'analyse des gestes chirurgicaux et à l'analyse des déplacements des personnes, en sont actuellement les supports.

REMERCIEMENTS

Ces expérimentations ont été financées conjointement par les projets RNTL franco-finlandais ADAMOS et Région Rhône-Alpes IMERA. Outre les auteurs de l'article, de nombreuses personnes se sont investies dans la mise en œuvre des expérimentations, à la fois du point de vue technique et du point de vue logistique. Les auteurs tiennent ainsi à remercier toutes les personnes ayant contribué au succès de ces trois expérimentations.

Pour MapMobile : D. David du CEA-Leti, A. Martin de France Télécom R&D, F. Forest de la MSH-Alpes, C. Golanski, M. Pernière et P.-J. Pommier du LIG/MultiCom (ex. CLIPS-IMAG). Pour E-skiing : P. Schermesser, A. Vidal et D. David du CEA-Leti, F. Forest de la MSH-Alpes, M. Léger et D. Maréchal de Rossignol ainsi que la station de l'Alpe d'Huez pour son accueil. Pour l'exposition « ni vu – ni connu » au Muséum du département du Rhône : C. Sermet, le scénographe de l'exposition, J.-F. Salmon le responsable informatique du Muséum, ainsi que le centre ERASME et la société Tagproduct qui en ont réalisé la mise en œuvre technique.

REFERENCES

- Arhippainen, L., Rantakokko, T., & Tähti, M. 2004. *Mobile Feedback Application for Emotion and User Experience Collection*. Paper presented at the Proactive computing workshop - PROW 2004, Helsinki, Finland.
- Baillie, L., & Schatz, R. 2005. *Exploring multimodality in the laboratory and the field*. Paper presented at the Proceedings of the 7th International Conference on Multimodal Interfaces, Toronto, Italy.
- Bartek, V., & Cheatham, D. 2003. *Experience remote usability testing, Part 1 & 2* (URL: "<http://www-106.ibm.com/developerworks/web/library/wa-rmustr1/>"): IBM's Pervasive Computing Division.
- Bastien, C., & Scapin, D. 1995. Evaluating a user interface with ergonomic criteria. *International Journal of Human-Computer Interaction*, 7(2), 105-121.
- Betioli, A. H., & Cybis, W. d. A. 2005. *Usability Testing of Mobile Devices: A Comparison of Three Approaches*. Paper presented at the IFIP TC13 International Conference on Human-Computer Interaction - Interact'2005, Rome, Italy.
- Brewster, S., & Walker, A. 2000. *Non-Visual Interfaces for Wearable Computers*. Paper presented at the IEE workshop on Wearable Computing (IEE, 00/145), London.
- Brush, A. J. B., Ames, M., & Davis, J. 2004. *A comparison of synchronous remote and local usability studies for an expert interface*. Paper presented at the CHI'04

- conference on Human factors in computing systems extended abstracts, Vienna, Austria.
- Buisson, M., Bustico, A., Chatty, S., Colin, F.-R., Jestin, Y., Maury, S., et al. 2002. *Ivy: un bus logiciel au service du développement de prototypes de systèmes interactifs*. Paper presented at the Proceedings of the 14th French-speaking conference on Human-computer interaction (Conférence Francophone sur l'Interaction Homme-Machine), Poitiers, France.
- Caelen, J., Jambon, F., & Vidal, A. 2005. Conception participative : des "Moments" à leur instrumentation. *Revue d'Interaction Homme-Machine (RIHM)*, 6(2), 1-29.
- Calvet, G., Salembier, P., Kahn, J., & Zouinar, M. 2005. *Étude empirique de l'interaction multimodale en mobilité : approche méthodologique et premiers résultats*. Paper presented at the 17e Conférence Francophone sur l'Interaction Homme-Machine - IHM'05, Toulouse, France.
- Castillo, J. C., Hartson, H. R., & Hix, D. 1998. *Remote usability evaluation: can users report their own critical incidents?* Paper presented at the CHI'98 conference on Human factors in computing systems, Los Angeles, California, United States.
- Chapanis, A. 1967. The Relevance of Laboratory Studies to Practical Situations. *Ergonomics*, 10(5), 557 - 577.
- Consolvo, S., Harrison, B., Smith, I., Chen, M. Y., Everitt, K., Froehlich, J., et al. 2007. Conducting In Situ Evaluations for and With Ubiquitous Computing Technologies. *International Journal of Human-Computer Interaction*, 22(1), 103-118.
- Dahl, Y., Alsos, O. A., & Svanæligs, D. 2010. Fidelity Considerations for Simulation-Based Usability Assessments of Mobile ICT for Hospitals. *International Journal of Human-Computer Interaction*, 26(5), 445-476.
- Demumieux, R., & Losquin, P. 2005. *Collecter les usages réels des clients de téléphonie mobile (un outil embarqué)*. Paper presented at the 17e Conférence Francophone sur l'Interaction Homme-Machine - IHM'05, Toulouse, France.
- Duh, H. B.-L., Tan, G. C. B., & Chen, V. H.-H. 2006. *Usability evaluation for mobile device: a comparison of laboratory and field tests*. Paper presented at the Proceedings of the 8th conference on human-computer interaction with mobile devices and services, Helsinki, Finland.
- Fiedler, M., & Haruvy, E. 2009. The lab versus the virtual lab and virtual field – An experimental investigation of trust games with communication. *Journal of Economic Behavior & Organization*, 72(2), 716-724.
- Fields, B., Amaldi, P., Wong, W., & Gill, S. 2007. In Use, In Situ: Extending Field Research Methods *International Journal of Human-Computer Interaction*, 22(1&2), 1-6.
- Forest, F., Guilloux, V., Mallein, P., & Panisset, J. 1998. La conception assistée par l'usage dans les bibliothèques publiques. In M.-H. Kœnig (Ed.), *Connaître ses publics : savoir pour agir (La Boîte à outils)* (Vol. 8). Villeurbanne: Institut de Formation des Bibliothécaires - IFB.
- Froehlich, J., Chen, M. Y., Consolvo, S., Harrison, B., & Landay, J. A. 2007. *MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones*. Paper presented at the Proceedings of the 5th International Conference on Mobile Systems, Applications and Services (MobySys'07), San Juan, Puerto Rico.
- Gob, A., & Drouguet, N. 2006. *La muséologie. Histoire, développement, enjeux actuels*: Armand Colin.

- Goodman, J., Brewster, S., & Gray, P. 2004. *Using Field Experiments to Evaluate Mobile Guides*. Paper presented at the 3rd Annual Workshop on HCI in Mobile Guides, Glasgow, UK.
- Hagen, P., Robertson, T., Kan, M., & Sadler, K. 2005. *Emerging research methods for understanding mobile technology use*. Paper presented at the OzCHI'2005, Canberra, Australia.
- Hammontree, M., Weiler, P., & Nayak, N. 1994. Remote Usability Testing. *interactions*, 1(3), 21-25.
- Hartson, H. R., Andre, T. S., & Williges, R. C. 2003. Criteria For Evaluating Usability Evaluation Methods. *International Journal of Human-Computer Interaction*, 15(1), 145-181.
- Hartson, H. R., Castillo, J. C., Kelso, J., & Neale, W. C. 1996. *Remote evaluation: the network as an extension of the usability laboratory*. Paper presented at the Conference on human factors in computing systems (CHI'96), Vancouver, British Columbia, Canada.
- Hertzum, M. 1999. *User Testing in Industry: A Case Study of Laboratory, Workshop, and Field Tests*. Paper presented at the 5th ERCIM Workshop on "User Interfaces for All" – Special Theme "User-Tailored Information Environments", Dagstuhl, Germany.
- Hsi, S., & Fait, H. 2005. RFID enhances visitors' museum experience at the Exploratorium. *Communications of the ACM*, 48(9), 60-65.
- Isomursu, M., Kuutti, K., & Väinämö, S. 2004. *Experience clip: method for user participation and evaluation of mobile concepts*. Paper presented at the 8th conference on Participatory design: Artful integration: interweaving media, materials and practices, Toronto, Ontario, Canada.
- Jambon, F. 2006. *Retours d'expérience lors d'évaluation de dispositifs mobiles en situation réelle*. Paper presented at the 3e Journées Francophones Mobilité et Ubiquité - Ubimob'06, Paris, France.
- Jambon, F. 2009. User Evaluation of Mobile Devices: In-Situ versus Laboratory Experiments. *International Journal of Mobile Computer-Human Interaction (IJMHCI)*, 1(2), 56-71.
- Jambon, F., Golanski, C., & Pommier, P.-J. 2006. *Évaluation des dispositifs mobiles : sur le terrain ou en laboratoire ?* Paper presented at the 18e Conférence Francophone sur l'Interaction Homme-Machine - IHM'06, Montréal, Canada.
- Jambon, F., Golanski, C., & Pommier, P.-J. 2007. *Meta-Evaluation of a Context-Aware Mobile Device Usability*. Paper presented at the International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM'07).
- Jambon, F., Mandran, N., Meillon, B., & Perrot, C. 2008. *Évaluation des systèmes mobiles et ubiquitaires : proposition de méthodologie et retours d'expérience*. Paper presented at the Ergo-IA'08 "L'humain au coeur des systèmes et de leur développement", Biarritz, France.
- Jambon, F., Mandran, N., & Perrot, C. (2007). La RFID au service de l'analyse du parcours muséal des visiteurs. *La lettre de l'Office de Coopération et d'Information Muséographiques (OCIM)*, 11-17.
- Jambon, F., & Meillon, B. 2009. *User Experience Evaluation in the Wild*. Paper presented at the CHI'09 Extended Abstracts on Human Factors in Computing Systems, Boston, MA, USA.
- Jensen, K. L., & Larsen, L. B. 2007. *Evaluating the usefulness of mobile services based on captured usage data from longitudinal field trials*. Paper presented at the 4th

- international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology, Singapore.
- Johnson, H. H., Solso, R. L., & Beale, K. 1997. Ethics of Experimental Research. In *Experimental Psychology: A Case Approach (6th Revised edition)* (pp. 400): Addison Wesley Longman.
- Kaikkonen, A., Kekäläinen, A., Cankar, M., Kallio, T., & Kankainen, A. 2005. Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing. *Journal of Usability Studies*, 1(1), 4-17.
- Kellar, M., Reilly, D., Hawkey, K., Rodgers, M., MacKay, B., Dearman, D., et al. 2005. *It's a jungle out there: practical considerations for evaluation in the city*. Paper presented at the CHI'05 extended abstracts on Human factors in computing systems, Portland, OR, USA.
- Kjeldskov, J., & Graham, C. 2003. *A Review of Mobile HCI Research Methods*. Paper presented at the Mobile Human-Computer Interaction – MobileHCI'2003, Udine, Italy.
- Kjeldskov, J., Graham, C., Pedell, S., Vetere, F., Howard, S., Balbo, S., et al. 2005. Evaluating the Usability of a Mobile Guide: The Influence of Location, Participants and Resources. *Behaviour & Information Technology*, 24(1), 51-65.
- Kjeldskov, J., Skov, M., Als, B., & Høegh, R. 2004. *Is It Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field*. Paper presented at the Mobile Human-Computer Interaction - MobileHCI'2004, Glasgow, UK.
- Kjeldskov, J., & Skov, M. B. 2007. Studying Usability In Situ: Simulating Real World Phenomena in Controlled Environments. *International Journal of Human-Computer Interaction*, 22(1), 7-36.
- Kjeldskov, J., & Stage, J. 2004. New Techniques for Usability Evaluation of Mobile Systems. *International Journal of Human-Computer Studies - IJHCS*, 60(5-6), 599-620.
- Lyons, K., & Starner, T. 2001. *Mobile Capture for Wearable Computer Usability Testing*. Paper presented at the Fifth International Symposium on Wearable Computers (ISWC'01), Zurich, Switzerland.
- Macefield, R. 2007. Usability studies and the Hawthorne Effect. *Journal of Usability Studies*, 2(3), 145-154.
- Malaurie, J. (Ed.). 2002. *De la vérité en ethnologie...* (Polaire ed.): Economica.
- Molich, R. 2000. *Usable Web Design (In Danish)*: Ingeniøren|bøger.
- Pirhonen, A., Brewster, S., & Holguin, C. 2002. *Gestural and audio metaphors as a means of control for mobile devices*. Paper presented at the ACM-SIGCHI Conference on Human Factors in Computing Systems (CHI'2002), Minneapolis, Minnesota, USA.
- Po, S., Howard, S., Vetere, F., & Skov, M. B. 2004. *Heuristic Evaluation and Mobile Usability: Bridging the Realism Gap*. Paper presented at the Mobile Human-Computer Interaction (MobileHCI'2004), Glasgow, UK.
- Riegelsberger, J., & Nakhimovsky, Y. 2008. Seeing the bigger picture: a multi-method field trial of google maps for mobile, *CHI'08 extended abstracts on Human factors in computing systems* (pp. 2221-2228). Florence, Italy: ACM Press.
- Roto, V., Oulasvirta, A., Haikarainen, T., Lehmuskallio, H., & Nyysönen, T. 2004. *Examining mobile phone use in the wild with quasi-experimentation* (Technical

- Report No. 2004-1). Helsinki, Finland: Helsinki Institute for Information Technology (HIIT).
- Rowley, D. E. 1994. *Usability testing in the field: bringing the laboratory to the user*. Paper presented at the Conference on Human Factors in Computing Systems (CHI'94), Boston (MA), United States.
- Scholtz, J. 2001. *Adaptation of Traditional Usability Testing Methods for Remote Testing*. Paper presented at the 34th Annual Hawaii International Conference on System Sciences - HICSS'01, Maui, Hawaii.
- Schulte-Mecklenbeck, M., & Huber, O. 2003. Information search in the laboratory and on the Web: With or without an experimenter. *Behavior Research Methods, Instruments & Computers*, 35(2), 227-235.
- Sermet, C., & Millet, M. 2007. "Ni vu, ni connu", une scénographie de camouflages. *La lettre de l'OCIM (Office de Coopération et d'Information Muséographiques)*(113), 4-10.
- Thomas, J. C., & Kellogg, W. A. 1989. Minimizing Ecological Gaps in Interface Design. *IEEE Software*, 6(1), 78-86.
- Thompson, K. E., Rozanski, E. P., & Haake, A. R. 2004. *Here, there, anywhere: remote usability testing that works*. Paper presented at the 5th conference on Information Technology Education, Salt Lake City, UT, USA.
- Tullis, T., Fleischman, S., McNulty, M., Cianchette, C., & Bergel, M. 2002. *An Empirical Comparison of Lab and Remote Usability Testing of Web Sites*. Paper presented at the Usability Professionals Association Conference - UPA'2002, Orlando, FL, USA.
- Waterson, S., Landay, J. A., & Matthews, T. 2002. *In the lab and out in the wild: remote web usability testing for mobile devices*. Paper presented at the CHI'02 conference on Human Factors in Computing Systems extended abstracts, Minneapolis, Minnesota, USA.
- Zouinar, M., Relieu, M., Salembier, P. & Calvet, G. 2004. *Observation et capture de données sur l'interaction multimodale en mobilité*. Paper presented at the Premières journées francophones Mobilité et Ubiquité - UbiMob'04, Sophia-Antipolis.



Francis JAMBON est Maître de Conférences en informatique à Polytech Grenoble (Université Joseph Fourier) où il enseigne la modélisation des systèmes d'information, l'interaction homme-machine et la conception des systèmes mobiles. Il fait partie de l'équipe MultiCom du LIG (Laboratoire d'Informatique de Grenoble) depuis 2002. Auparavant, et depuis 1998, il a fait partie de l'équipe IHM et PSE du laboratoire LISI/ENSMA à Poitiers. Ses travaux de recherche portent principalement sur les méthodes d'évaluation et de validation des systèmes interactifs, et d'analyse de l'activité humaine, notamment dans le cas des systèmes mobiles, ubiquitaires ou critiques. Il est membre de l'ACM-SIGCHI et de l'AFIHM.



Nadine MANDRAN est ingénieur d'étude CNRS au Laboratoire d'Informatique de Grenoble (LIG). Après une formation en statistique et en sociologie, elle a exercé ses activités dans plusieurs domaines de recherche. Aujourd'hui, son activité se déroule dans le cadre de la plateforme expérimentale du LIG Marvelig. Elle travaille à la mise en œuvre de méthodologies et de protocoles expérimentaux pour les équipes de recherche dont le domaine est lié aux comportements des utilisateurs face à des nouveaux

concepts. Cet axe permet d'assurer conseil et partenariat méthodologique pour la production et le traitement qualitatif ou statistique des données. Un autre axe est le déploiement de la démarche qualité pour la plateforme expérimentale. Dans ce cadre elle met en place différents processus pour la capitalisation des prototypes du laboratoire et pour le suivi des expérimentations. Elle est membre du Réseau Qualité du CNRS.



Brigitte MEILLON est ingénieur CNRS au LIG (Laboratoire d'informatique de Grenoble), dans l'équipe MultiCom. Elle est responsable technique de la plateforme d'expérimentation située au CTL (Centre des technologies du Logiciel), sur le campus de Grenoble. Son activité est centrée sur l'instrumentation des expérimentations et sur la capture de comportements d'utilisateurs. En fonction des objectifs des différents projets de recherche et contrats, elle propose des outils existants ou développés en interne pour produire des données pertinentes et pour les traiter. Elle fait évoluer le matériel et les outils de la plateforme, en fonction des différents projets et contrats de l'équipe (domotique, géolocalisation, RFID). Elle s'intéresse plus particulièrement à l'oculométrie d'un point de vue technique depuis une dizaine d'années et a développé des logiciels d'acquisition et de traitement de données oculométriques.



Christian PERROT est ingénieur de recherche au CNRS, chargé de valorisation dans l'équipe MultiCom du LIG. Il conduit des projets à caractère industriel autour des RFID. Il est responsable technique du projet de table interactive mettant en jeu des objets Tangibles dans le cadre du projet ANR « TTT ». Il contribue également au déploiement de solutions technologiques robustes dans les musées dans le domaine des outils de médiation fixes ou mobiles associés à des technologies de traçage du comportement des visiteurs.