



HAL
open science

Adaptation en contexte: contribution aux interfaces multimodales et multicanal (mémoire HDR)

José Rouillard

► **To cite this version:**

José Rouillard. Adaptation en contexte: contribution aux interfaces multimodales et multicanal (mémoire HDR). LIFL. 2008. hal-04362108

HAL Id: hal-04362108

<https://hal.science/hal-04362108v1>

Submitted on 22 Dec 2023

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.

HABILITATION À DIRIGER DES RECHERCHES

présentée par

José Rouillard

Discipline : Informatique

Spécialité : Interaction Homme-Machine

Adaptation en contexte : contribution aux interfaces multimodales et multicanal

Le 04/12/2008

devant le jury constitué de :

(par ordre alphabétique)	
Prof. Bertrand DAVID (Université de Lyon)	Rapporteur
Prof. Alain DERYCKE (Université de Lille)	Examineur
Prof. Christophe KOLSKI (Université de Valenciennes)	Examineur
Prof. Philippe MATHIEU (Université de Lille)	Examineur
Prof. Philippe PALANQUE (Université de Toulouse)	Rapporteur
Prof. Jean VANDERDONCKT (Université de Louvain-la-neuve)	Rapporteur

Habilitation à Diriger des Recherches préparée au sein du
Laboratoire d'Informatique Fondamentale de Lille (UMR CNRS 8022)
Université des Sciences et Technologies de Lille
Bât. M3 ext. - 59655 Villeneuve d'Ascq Cedex

A la mémoire de ma mère.

A mon épouse Karine et notre fille Cassandra.

Résumé

Mon domaine de recherche est l'Interaction Homme-Machine (IHM). Mes travaux dans ce domaine s'organisent autour de deux axes.

Le premier axe concerne l'adaptation des interfaces homme-machine. Je m'intéresse en particulier à la manière dont on peut concevoir et réaliser des interfaces qui pourront s'adapter, selon le contexte, aux utilisateurs et à leurs rôles, aux périphériques utilisés ou encore à la tâche à exécuter lors de l'interaction.

Le deuxième axe concerne les interactions multimodales et multicanal¹. J'étudie plus particulièrement les interfaces couplant le vocal (aussi bien en entrée qu'en sortie), avec d'autres moyens d'interaction, notamment via le canal téléphonique.

Ce document présente les problématiques liées à ces deux axes principaux de recherche, les travaux s'y rapportant auxquels j'ai participé depuis septembre 2000 et quelques perspectives de recherche dans ces thématiques.

Mots-clés

Adaptation, contexte, plasticité, multimodal, multicanal, interaction vocale, VoiceXML, SOA, web services, traces, mobilité, QR Codes.

¹ Le terme multicanal est issu du marketing direct. Il est formé sur un nom et quand il devient adjectif, il est invariable. Il convient donc d'utiliser la locution "interactions multicanal". Merci à Christian Boitet de l'Université Joseph Fourier de Grenoble pour ces précisions.

Remerciements

Je remercie Alain DERYCKE pour la confiance qu'il m'a témoignée en m'accueillant dans son équipe.

Je tiens à remercier les rapporteurs des travaux présentés dans ce document : Bertrand DAVID (Professeur à l'université de Lyon), Philippe PALANQUE (Professeur à l'université de Toulouse) et Jean VANDERDONCKT (Professeur à l'université de Louvain-la-Neuve).

Je remercie également les examinateurs, Alain DERYCKE (Professeur à l'université de Lille), Christophe KOLSKI (Professeur à l'université de Valenciennes) et Philippe MATHIEU (Professeur à l'université de Lille) qui ont chaleureusement accepté de participer à ce jury.

Par ailleurs, je souhaite remercier les membres de l'équipe NOCE du laboratoire LIFL avec lesquels j'ai collaboré, et plus particulièrement Jean-Claude TARBY, pour son soutien sans failles et Xavier LE PALLEC pour sa bonne humeur naturelle.

Bien entendu, je n'oublie pas mes coauteurs, notamment Vincent CHEVRIN et Thomas VANTROYS, ainsi que les étudiants que j'ai encadrés et qui ont contribué, de près ou de loin, à l'avancement de ces travaux de recherche.

Enfin, je tiens à remercier du plus profond de mon cœur mon épouse Karine et notre fille Cassandra, pour avoir compris et accepté ma faible disponibilité pendant la préparation de mon HDR. Qu'elles en soient remerciées et reçoivent ici ton mon amour.

Table des matières

RESUME.....	3
MOTS-CLES	3
REMERCIEMENTS	4
1 INTRODUCTION.....	6
1.1 CONTEXTE ET MOTIVATION	6
1.2 L'EQUIPE DE RECHERCHE NOCE.....	6
1.3 PROBLEMATIQUE.....	7
2 ADAPTATION.....	11
2.1 AXE DE RECHERCHE	11
2.1.1 <i>Pourquoi (Why)</i>	12
2.1.2 <i>Quoi (What)</i>	13
2.1.3 <i>Qui (Who)</i>	13
2.1.4 <i>Quand (When)</i>	14
2.1.5 <i>Où (Where)</i>	14
2.1.6 <i>A quoi (To what)</i>	14
2.1.7 <i>Comment (How)</i>	15
2.1.8 <i>Adaptation d'un document numérique</i>	18
2.1.9 <i>De langages basés sur XML pour la plasticité des IHM</i>	19
2.2 CONTRIBUTION A CET AXE DE RECHERCHE	20
2.2.1 <i>Le langage PlasticML</i>	20
2.2.2 <i>Adaptation pour les applications mobiles</i>	22
2.2.3 <i>Les QR Codes Contextuels</i>	25
3 MULTIMODALITE ET MULTICANALITE.....	30
3.1 AXE DE RECHERCHE	30
3.1.1 <i>Terminologie</i>	31
3.1.2 <i>Interactions vocales</i>	35
3.1.3 <i>Le langage VoiceXML</i>	36
3.1.4 <i>X+V</i>	36
3.1.5 <i>EMMA</i>	38
3.2 CONTRIBUTION A CET AXE DE RECHERCHE	40
3.2.1 <i>Référentiel Multi-DMC</i>	40
3.2.2 <i>Multimodalité vocale et graphique</i>	43
3.2.3 <i>Multimodalité vocale et SMS</i>	44
3.2.4 <i>Multimodalité vocale et gestuelle</i>	44
3.2.5 <i>VoiceXML augmenté</i>	46
3.2.6 <i>Des services web pour la multimodalité</i>	48
3.2.7 <i>Instrumentation et mécanisme de traces pour VoiceXML</i>	50
CONCLUSION ET PERSPECTIVES	55
4 REFERENCES.....	60
5 ANNEXES	70
5.1 LE JEU ACTIVE LIFE: ATHLETIC WORLD SUR CONSOLE NINTENDO WII.....	70
5.2 LES COMPOSANTS DE L'OBJET COMMUNICANT CHUMBY	71
5.3 PROTOTYPE DE MACHINE A LAVER PARLANTE.....	71
5.4 LA BOITE A OUTIL DU PLASTICML	72
5.5 ADAPTATION D'UN DOCUMENT PEDAGOGIQUE	72
5.6 UTILISATION DE CODE BARRES 2D	75
5.7 ESPACE DE CLASSIFICATION DE ROUDAUT & LECOLINET	76
5.8 RECONNAISSANCE D'ECRITURE AVEC INKML.....	76
5.9 ETUDE ET RECUEIL DE CORPUS D'UTILISATION DE VOICEXML.....	77
5.10 OUTILS NON INVASIFS POUR LE BCI	80
6 CURRICULUM VITAE DES ACTIVITES DE RECHERCHE ET D'ENSEIGNEMENT	84
7 SELECTION D'ARTICLES DE RECHERCHE	94

1 Introduction

Depuis plusieurs années, on entend couramment parler de nouvelles technologies, de nouvelles interfaces homme-machine, de nouveaux moyens de communiquer avec des systèmes de plus en plus complexes. Bien entendu, chaque génération qui vient remplacer la précédente est forcément nouvelle, d'un point de vue purement historique ; on estime alors que tel nouveau composant est plus performant que tel autre ancien modèle et cela est vérifiable de manière métrique. Cependant, dans le domaine des IHM, depuis que l'on considère que l'utilisateur doit être placé au cœur du système informatique, les avancées scientifiques ne se mesurent pas toujours qu'en terme de rapidité d'exécution, mais par bien d'autres critères, notamment ergonomiques et sociologiques. J'estime, pour ma part, que la principale avancée des approches dites « nouvelles » réside dans la prise en compte du contexte d'utilisation lors de l'interaction. « *Context is Key* » disent les anglo-saxons [Coutaz & al. 2005]. Cette adaptation au contexte est l'un des axes de ma recherche depuis plusieurs années. Les chercheurs impliqués dans cette thématique tentent de fournir des modèles et des outils permettant de relâcher des contraintes afin de gagner en flexibilité au moment de la conception (gain pour le développeur) et de l'utilisation du système (gain pour l'utilisateur final) [Habieb-Mammar 2004], [Laforest 2007]. L'autre axe de recherche de mes travaux va également dans le sens de la flexibilité, puisqu'il embrasse les dimensions multimodales et multicanal des systèmes informatiques. Là encore, l'idée sous-jacente est de permettre une plus grande souplesse d'utilisation en proposant l'usage de nos cinq sens, sans contraindre les choix que feront les utilisateurs, par rapport au contexte d'usage dans lesquels ils se trouveront. La section suivante présente plus en détail le contexte, les motivations et le cadre de mes travaux.

1.1 Contexte et motivation

J'ai soutenu ma thèse d'informatique, intitulée « Hyperdialogue sur Internet. Le système HALPIN », le 11 janvier 2000, à Grenoble, sous la direction de Jean Caelen. J'ai été nommé maître de conférences en informatique en septembre de la même année, en intégrant le laboratoire Trigone de Lille, dirigé par Alain Derycke. Depuis mon arrivée dans l'équipe NOCE de ce laboratoire, mes travaux gravitent autour des nouvelles formes d'interaction homme-machine dans le domaine des interactions multimodales et multicanal pour l'adaptation des services interactifs. Les terrains d'expérimentations choisis sont principalement le commerce électronique et l'apprentissage humain, cependant, les résultats obtenus sont largement transposables.

Ces travaux se sont déroulés dans le cadre de projets régionaux (NIPO², MIAOU³, EUCUE⁴) et nationaux (p-LearNet⁵). La motivation conjointe à ces études scientifiques réside dans la volonté de proposer des idées et/ou des outils permettant de concevoir et de réaliser des interfaces adaptables à des situations réelles d'usage rencontrées par les utilisateurs, et le plus souvent non reproductibles en laboratoire.

1.2 L'Equipe de recherche Noce

La compétence en Interactions Homme-Machine et en nouvelles interfaces au sein de l'équipe NOCE⁶ est déjà assez ancienne et tend à se développer au travers de nos recherches collectives sur les Environnements Informatiques pour l'Apprentissage Humain menées en Collaboration (EIAHC). Il en résulte un enrichissement croisé : (a) le domaine des EIAHC bénéficie des avancées du domaine des IHM, en terme de solutions technologiques, de modélisation et de compréhension des processus d'apprentissage humain instrumentés par l'informatique et la communication ; (b) le domaine des IHM bénéficie de besoins nouveaux, de contextes d'usages particuliers et de champs d'expérimentation écologiquement valides, générés par les nouveaux EIAHC.

Le besoin d'un accès permanent et ubiquitaire à l'information (au bureau, chez soi, dans le train, etc.) se fait de plus en plus sentir avec la venue de nouveaux matériels et le succès des ordinateurs de poche ainsi que des téléphones portables. Ceci a conduit à l'émergence de nouveaux paradigmes en Interfaces Homme-Machine tels que la mobilité, l'informatique ubiquitaire, ou la plasticité des interfaces [Calvary & al. 2006].

² NIPO : Nouvelles Interactions Personnes-Organisations

³ MIAOU : Modèles d'Interaction et Architectures Orientées Utilisateurs

⁴ EUCUE : Evaluation des Usages dans des Contextes d'E-services Ubiquitaires et Evolutifs

⁵ p-LearNet : Pervasive Learning Network : infrastructures et usages pour l'apprentissage humain dans le contexte de l'informatique diffuse (<http://p-learnnet.univ-lille1.fr/>)

⁶ NOCE : Nouveaux Outils pour la Coopération et l'Education

Ceci est à placer dans le contexte des nouvelles approches dites orientées services (Service Oriented Architecture : SOA) et de l'essor des e-Services comme composants essentiels pour la conception des infrastructures, pour le e-Commerce et la e-Formation par exemple.

L'un des objectifs de nos travaux de recherche est la production et la combinaison d'interfaces utilisables par n'importe qui, n'importe où et n'importe quand. Cet enjeu rejoint le thème de la plasticité des IHM telle que définie dans l'AS Plasticité du RTP16 (IHM) tout en y adjoignant nos recherches spécifiques liées d'une part, au multimodal et au multicanal, et d'autre part, à la modélisation évoluée des IHM (co-évolution, méta-modélisation, etc.) dans le contexte des activités humaines collaboratives. Notre intérêt se porte plus spécifiquement sur les questions de couplage/découplage de canaux, mais aussi sur l'articulation Rôles/Tâches/Canaux et les processus de collaboration Humain/Humain au travers de tels dispositifs (par exemple les activités de tuteurage ou de collaboration entre pairs dans le domaine des EIAHC).

Pour mener à bien ces recherches de base, nous nous appuyons sur des compétences liées à nos thèmes de recherche concernant les interactions enrichies (interfaces multimodales, plastiques, ubiquitaires, en situation de mobilité ...) et la modélisation d'IHM. Cependant, les problèmes abordés sont tels que ces compétences nécessitent d'être complétées par une collaboration au sein de l'équipe NOCE, avec :

- d'une part le domaine des systèmes distribués, par exemple au travers des relations entre workflows, tâches, « composants plastiques » et e-Services ;
- et d'autre part, l'axe modélisation, par exemple, modélisation centrée sur les activités, buts, procédures et tâches, méta-modélisation, transformation et équivalence de modèles, établissement de méta-modèles ontologiques des interactions multimodales et multicanal.

1.3 Problématique

L'année 1984 marque l'apparition pour le grand public des interfaces dites WIMP (« Windows, Icon, Menu, Pointing device»). Développé chez Xerox PARC en 1973 puis popularisé par Macintosh, ce type d'interfaces graphiques est encore majoritairement utilisé de nos jours, sur la plupart des ordinateurs. Cependant, depuis quelques années, des nombreuses recherches scientifiques se focalisent sur les interfaces post-WIMP. On ne se limite plus à une seule manière d'interagir avec un système informatique, mais on étudie les différentes solutions envisageables afin de proposer des interfaces utilisateurs qui soient d'usage aussi naturel que possible. On s'intéresse alors, peu ou prou, au domaine de la multimodalité. Mais ce domaine est étroitement lié aux notions d'adaptation des interfaces et il est difficile de circonscrire les limites de chaque domaine.

La multimodalité est un axe des IHM qui est étudié depuis le début des années 80 aux Etats-Unis [Bolt 1980] et depuis 1990 en France [Caelen 1992]. Dès lors, de nombreux travaux ont explorés les possibilités et les limites de ces systèmes [Oviatt 1999], [Nigay 2001], [Bellik 2006], [Martin 2006], [Calvary 2007].

L'idée de pouvoir communiquer avec un ordinateur grâce à nos cinq sens, au lieu de se limiter à une manipulation directe classique (clavier/souris) est en effet fort séduisante. Plus encore, le fait de pouvoir parler (idéalement, dialoguer en langue naturelle) avec un système informatique afin de lui faire faire des actions pour nous ou en notre nom est en soi une motivation très forte qui a donné lieu à des recherches pluridisciplinaires prolifiques. « *Exprimer son problème en langage naturel à une machine, n'est-ce pas le but ultime des recherches en programmation ?* » [Coulon & Kayser 1986]. Nous montrerons plus loin que nos travaux dans ce domaine se sont focalisés sur la dimension vocale couplée à d'autres moyens d'interactions.

Cependant, force est de constater que les promesses des systèmes permettant d'agir naturellement dans nos environnements numériques, « comme on le ferait avec un interlocuteur humain » n'ont pas vraiment été tenues. Si l'évolution vers le multimédia a bien eu lieu, le passage au multimodal reste encore modéré. De véritables applications exploitant pleinement les capacités de la multimodalité n'ont jamais réellement été trouvées. Les seuls domaines dans lesquels des travaux ont montré des résultats significatifs sont les suivants : l'aéronautique, le domaine médical, le domaine militaire, celui des jeux vidéos et les d'utilisations spécifiques par des personnes handicapées.

La substitution d'une manière d'interagir par une autre résulte soit d'un choix de la part de l'utilisateur, soit d'une contrainte. Un chirurgien ayant les mains occupées par la tâche principale qui consiste à

opérer le patient choisira d'utiliser d'autres modalités comme la voix ou un geste (pression du pied sur une pédale notamment) pour effectuer des opérations annexes. En revanche, un utilisateur handicapé (temporairement ou pas), utilisera sa voix plutôt qu'une souris, par exemple, non pas par choix, mais parce qu'il y est contraint.

Au niveau grand public, et depuis quelques années seulement, différentes modalités d'interactions sont parfois proposées aux utilisateurs dans certains domaines particuliers, comme la téléphonie ou bien encore les jeux vidéos⁷.



Figure 1 : Utilisation du toucher sur Iphone pour capturer un code barre 2D et obtenir une information détaillée (localisation d'un restaurant)

Les photos de la Figure 1 montrent l'utilisation du toucher sur téléphone Iphone. D'une pression du doigt, l'utilisateur déclenche l'application chargée de décoder les codes barre 2D. Il vise, avec l'appareil photo intégré dans son téléphone, le code barre présent sur le poster affiché contre le mur et instantanément, une carte s'affiche sur son écran et présente la localisation géographique du restaurant dont il est question sur l'affiche. L'utilisateur peut alors zoomer et naviguer au sein de cette carte. Un lien entre le monde physique et tangible et le monde numérique est ainsi réalisé. Le domaine des loisirs, et celui des jeux vidéos en particulier, bénéficie grandement des avancées dans le domaine de la multimodalité. Le jeu ne se manipule plus uniquement grâce au joystick d'antan, mais avec toute une panoplie de capteurs et d'effecteurs. Les photos de la Figure 2 illustrent des situations d'utilisation du jeu « Family Ski », sur console Nintendo Wii.



⁷ Par exemple, pour le jeu Indiana Jones, sur console Nintendo DS : « Pour aider le docteur Jones dans ses péripéties, il faudra utiliser toutes les capacités de la DS. Ainsi, votre souffle dans le micro permettra d'éteindre des torches, tandis que frotter votre pouce sur l'écran tactile permettra de creuser. ». Sur console Playstation PS2, il est possible de formuler des commandes vocales dans le jeu SOCOM U.S. Navy Seals : « Ce jeu captivant est rendu encore plus exceptionnel par le casque SOCOM qui vous permet d'adresser des ordres verbaux à votre équipe, mais aussi de recevoir des messages radio en réponse, directement dans l'oreille. SOCOM profite également pleinement de l'adaptateur réseau pour PlayStation 2 qui donne aux utilisateurs de l'Internet haut débit l'occasion de prendre part à des combats multijoueurs, avec jusqu'à 16 joueurs au même moment. ».



Figure 2 : Utilisation du jeu Family Ski sur console Nintendo Wii⁸

Pour jouer à ce jeu, l'utilisateur peut choisir d'utiliser la balance « Wii board » qui détecte les mouvements du corps (forces, pressions, inclinaisons, accélérations...) ou bien il peut orienter les deux manettes de jeu (Wiimote et Nunchuk) pour diriger son avatar sur les pistes de ski. En plus des images et des sons émis par le téléviseur, l'utilisateur perçoit des vibrations et des sons qui proviennent de la manette de jeu sans fil (voir aussi en annexe 1 les jeux utilisant des tapis interactifs).

De nouveaux objets communicants comme le lapin Nabaztag⁹ ou le Chumby¹⁰ permettent d'être connecté à Internet via des machines qui ne s'apparentent pas à des ordinateurs classiques.



Figure 3 : L'objet communicant Chumby

Le Chumby, présenté sur la Figure 3, ressemble, au premier abord, à un radio réveil traditionnel. Mais en fait, il possède, entre autres caractéristiques, un système d'exploitation Linux, un accès WIFI à Internet, un écran LCD tactile 320x240, deux haut-parleurs 2 watts stéréo, une sortie audio mini-jack, un micro intégré, un capteur de lumière ambiante, un capteur de pression (*squeeze sensor*), des accéléromètres, deux ports USB, un accès SSH, un gestionnaire de téléchargements de fichiers WGET, et le logiciel principal est un player flash capable d'afficher du contenu multimédia (voir en annexe 2 les détails techniques du Chumby).

Ces objets d'usage courant vont prendre, petit à petit, une place prépondérante dans nos vies car ils permettent d'allier la simplicité des objets quotidiens à la puissance des réseaux de communication. En théorie, chaque objet pourra, dans le futur, posséder sa propre adresse IP¹¹. Ces objets sont parfois qualifiés d'intelligents (*smart*) dans le sens où ils offrent un pont entre le monde physique et le monde numérique.

⁸ Source : http://www.dailymotion.com/video/x46pgu_family-ski-demo-japonaise_videogames

⁹ Nabaztag : <http://www.nabaztag.com/fr/index.html>

¹⁰ Chumby : <http://www.chumby.com/>

¹¹ On passe de 2³² adresses théoriques en IPv4 à 2¹²⁸ adresses en IPv6

Le monde de la domotique profite largement de ces avancées technologiques. En collaboration avec Fujitsu et Haier (spécialiste dans l'électroménager), la société israélienne Santiago, spécialisée dans l'innovation autour des objets de la vie courante, a développé un prototype permettant de communiquer de manière orale avec une machine à laver. Comme le montre la Figure 4, l'utilisatrice est guidée par le système pour chaque tâche à accomplir (voir d'autres illustrations en annexe 3). L'interaction vocale (reconnaitances et synthèses vocales) est utilisée en complémentarité des manipulations directes (presser bouton, ouvrir tiroir lessive, fermer hublot). Les premiers modèles de ce produit devraient être commercialisés en Chine dans quelques mois (Source : VocalNews du 26/09/08).



Figure 4 : Prototype de machine à laver à commandes orales

Ma problématique de recherche depuis septembre 2000 consiste à concevoir, mettre en œuvre et évaluer des interfaces homme-machine multimodales et adaptables au contexte d'utilisation. J'ai choisi de m'intéresser plus spécifiquement à la dimension vocale de la multimodalité, en entrée comme en sortie. Je privilégie donc dans mes travaux les communications téléphoniques et les situations de mobilité des utilisateurs. Des verrous forts liés à la limitation de ce que peut dire un utilisateur et la manière dont il peut le faire (parler avant, en même temps ou après avoir utilisé une autre modalité) m'ont conduit à m'interroger sur les apports possibles des architectures orientées services (SOA) dans ce cadre d'étude. Je me suis spécialisé dans des sous domaines de ces thématiques et j'ai rédigé deux ouvrages ; l'un présentant de manière complète le langage VoiceXML (c'est le seul ouvrage francophone sur le sujet) et l'autre à propos des SOA afin de mieux appréhender le passage de la théorie à la pratique. Des enseignements et des suivis de projets au niveau master professionnel ainsi que des encadrements de masters Recherche, DEA et thèse m'ont permis de parfaire mes connaissances autour de ces thématiques.

Pour favoriser la clarté de la présentation, la suite du document présente une vue synthétique de ces travaux de recherche selon les deux axes « adaptation » et « multimodalité », de manière disjointe, même si, en réalité, ils ont été travaillés conjointement.

En IHM, les chercheurs étudient non seulement les travaux de leur propre domaine, mais également ceux de disciplines connexes à l'informatique que sont la sociologie, la linguistique, l'ergonomie, la psychologie ... pour n'en citer que quelques unes. Je montrerai dans ce document que certaines parties de mes travaux présentent des facettes originales pour l'IHM, en s'inspirant, par exemple, du domaine du marketing direct. La notion d'**adaptation** sera abordée au **chapitre 2** sous plusieurs éclairages différents afin de mettre en évidence les similitudes et les divergences de points de vue selon les disciplines. Le **multimodal** et le **multicanal** seront présentés dans le **chapitre 3**. Mon attirance pour le dialogue homme-machine en général, et pour les interactions vocales en particulier sera révélée tout au long de mon exposé. L'exercice est d'autant plus difficile du fait que les frontières entre ces différents domaines d'étude ne sont pas franches. Un même prototype d'interface vocale téléphonique pourra me servir tantôt pour illustrer la mise en œuvre d'une adaptation à un contexte particulier (profil de l'utilisateur, lieu, type de périphérique utilisé, environnement, etc.), tantôt pour mettre en évidence les facultés multimodales ou multicanal d'un système. Enfin, une conclusion et des perspectives de recherche seront présentées pour conclure cette partie du document. Une sélection de quelques uns de mes articles de recherche scientifique donnera plus de détails sur certains thèmes abordés dans cette HDR.

2 Adaptation

2.1 Axe de recherche

La notion d'adaptation des interfaces homme-machine présente plusieurs facettes. Des travaux généraux et de taxonomies sont disponibles [Malinowski 1992], [Brusilovsky 1996], mais les concepts changent selon les périodes, les domaines d'études et les points de vue des auteurs. [Paramythis 2004] a écrit à ce propos : « *The field of adaptive systems is infamous for its lack of standards, or even commonly accepted approaches* ».

Les premiers travaux autour de l'adaptation furent principalement basés sur les quatre questions suivantes: Quoi, Quand, Pourquoi et Comment s'adapter? [Karagiannidis & al. 1996]. Depuis, de nombreux autres critères ont été utilisés pour la conception des applications adaptables, notamment celles dites « sensibles au contexte », et des ateliers de conception évolués sont maintenant disponibles pour aider à la création et l'étude de ce type d'interfaces [Vanderdonck & al. 2005].

Une première approche de l'adaptation est perceptible en étudiant les séquences temporelles impliquées dans les processus d'adaptation. Les différentes opérations qui se produisent dans les processus d'adaptation peuvent être regroupées en plusieurs étapes. D'un point de vue centré sur le système, les principales étapes citées par [Totterdell & Rautenbach 1990] sont les suivantes: **variation**, **sélection** et **tests**. D'un point de vue centré sur l'utilisateur, les étapes connues de l'adaptation sont l'**initiative** (c'est la décision d'un des agents de proposer une adaptation), la **suggestion** (les différentes alternatives proposées pour l'adaptation), la **décision** (une des alternatives est choisie) et l'**exécution** [Dietrich & al. 1993].

Afin d'atteindre un objectif spécifique, un concepteur crée un système, selon un point de vue personnel et particulier dans un contexte C. Un utilisateur peut avoir un point de vue différent du concepteur, s'il se trouve dans un contexte C'. Par conséquent, une adaptation du système est nécessaire pour passer du contexte C au contexte C'.

Selon [Trigg & al. 1987], il existe quatre façons, pour un système, d'exposer une adaptabilité :

- Un système est dit **flexible** (*flexible*) s'il propose des objets génériques et des comportements qui peuvent être interprétés et utilisés différemment, par différents utilisateurs au cours de différentes tâches.
- Un système est dit **paramétrable** (*parameterised*) s'il offre un éventail de comportements alternatifs parmi lesquels les utilisateurs pourront choisir.
- Un système est dit **intégrable** (*integratable*) s'il peut être interfacé et intégré avec d'autres dispositifs au sein de son environnement ou bien de manière distante.
- Un système est **adaptable** (*tailorable*) s'il permet aux utilisateurs de modifier le système, par facilitation de construction, comportements spécialisés, ou ajout de fonctionnalité.

Pour [Oppermann 1994], les systèmes ne sont pas conçus pour un utilisateur spécifique : « *application systems are not designed for a particular user and a particular task. They are designed and distributed for a class of users and a set of tasks* ». Selon cet auteur, un moyen d'accroître la liberté des utilisateurs sans augmenter la complexité du système est de fournir des systèmes taillés sur mesure (*tailored*). Les deux termes communément utilisés dans la littérature pour qualifier un système d'adaptation sont « adaptable » et « adaptatif ». Un système est dit « **adaptable** » (adaptable) s'il fournit à l'utilisateur des outils permettant de changer les caractéristiques du système. Un système est dit « **adaptatif** » (adaptive) s'il est en mesure de modifier ses propres caractéristiques automatiquement en fonction des besoins des utilisateurs.

Il existe également dans la littérature une différence entre **adaptabilité** et **adaptativité** : « *L'adaptabilité est la capacité du système de s'adapter aux personnalisations expressément demandées par l'utilisateur, tandis que l'adaptativité désigne sa capacité à répondre aux besoins de l'utilisateur sans une intervention explicite de sa part.* » [Moisuc 2006].

Adaptation et adaptativité sont très souvent employées comme synonymes [Khalil 2001]. Mais parfois d'autres termes comme « configurable » (*configurable*) ou « personnalisable » (*customisable*) sont utilisés pour désigner ce type d'adaptation automatique ou semi-automatique [Weibelzahl 2002].

Dans le domaine des hypermédias adaptatifs, les concepts sont sensiblement les mêmes [De Bra & Calvi 1998]. On parle d'**adaptation minimale** lorsque le système permet à l'utilisateur de choisir une version parmi plusieurs définies au préalable, d'**adaptabilité** lorsque le système est capable de réagir à des demandes explicites d'adaptation de la part de l'utilisateur et enfin d'**adaptativité** lorsque le système est capable d'observer le comportement de l'utilisateur et de procéder en conséquence à des adaptations [Balla & al. 2006].

Nous avons représenté les principaux aspects de l'adaptation sur une carte, ci-après. Chaque branche présentée sur la Figure 5 est détaillée dans les paragraphes qui suivent. Cette première représentation n'est pas exhaustive, mais elle permet de clarifier les principales dimensions de l'adaptation.

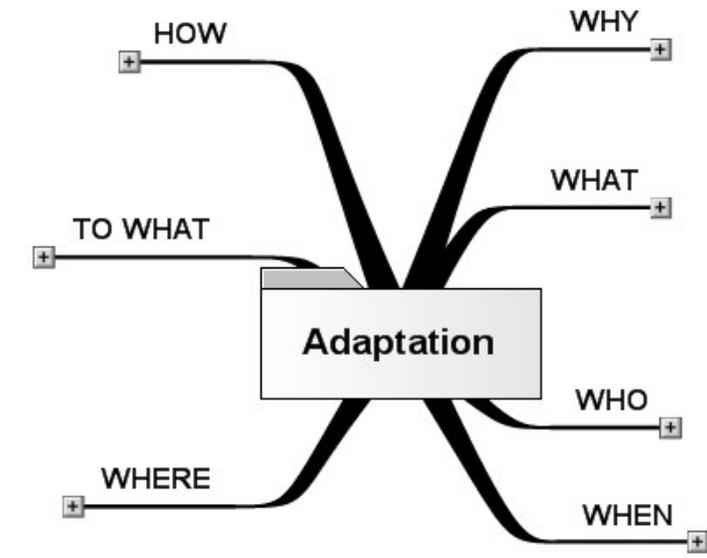


Figure 5 : Les principaux aspects de l'adaptation

2.1.1 Pourquoi (Why)

La première question à propos de l'adaptation est la suivante : « pourquoi adapter ? ». Fondamentalement, l'objectif principal de l'adaptation consiste à accélérer et simplifier l'utilisation des systèmes informatiques, en présentant à l'utilisateur ce qu'il souhaite voir et faire et ainsi rendre des systèmes complexes plus utilisables. Selon [Malinowski 1992], présenter plus facilement et plus efficacement des interfaces est l'objectif principal de l'adaptation.

Mais, pour atteindre ces objectifs, il est nécessaire pour [Dietrich & al. 1993] de disposer d'une interface utilisateur qui convienne à différents groupes hétérogènes d'utilisateurs et qui considère l'expérience croissante de l'utilisateur. Le rôle de l'adaptativité (dans le domaine de documents numériques [Ramel 2006], principalement) est de minimiser l'effort de l'utilisateur consacré à l'exploration des capacités du système afin d'optimiser celui nécessaire pour la résolution d'un problème.

Comme nous le savons, il est difficile de concevoir et de réaliser un logiciel qui conviendra parfaitement à des millions d'utilisateurs [Horgen 2002], mais il est possible de développer des systèmes capables d'ajuster leur interface conformément aux compétences, aux connaissances ou encore aux préférences de chaque utilisateur. Afin de réaliser cette adaptation, nous verrons que les modèles sous-jacents de l'utilisateur et de la tâche sont essentiels, ainsi que la séparation de l'interface utilisateur du reste de l'application [Fischer 2000].

Les systèmes adaptatifs sont utilisés dans de nombreux domaines pour résoudre différentes tâches. La suite de fonctions suivante, établie par [Weibelzahl 2002] et adaptée de [Jameson 2001] donne quelques exemples d'usages : aider l'utilisateur à trouver de l'information ; personnaliser l'information (dite « sur mesure ») pour l'utilisateur ; recommander des produits (par exemple pour le commerce électronique) ; fournir une aide appropriée, selon le contexte d'utilisation ; faciliter l'apprentissage ; conduire un dialogue homme-machine ; favoriser la collaboration entre utilisateurs et/ou avec la machine...

2.1.2 Quoi (What)

Une autre interrogation importante réside dans la question « quoi adapter ? ». Parfois, seulement l'aide, les tutoriels ou les conseils sont adaptés. Dans d'autres cas, le dialogue entre le système et l'utilisateur est modifié par l'adaptation. Il peut également être question d'adaptation au contenu, à la navigation, à la présentation, aux fonctionnalités disponibles, etc. La plupart du temps, c'est uniquement la partie présentation de l'interface qui est adaptée. Dans d'autres cas, la présentation et quelques fonctions du système sont adaptées conjointement. Ce dernier point est appelé « malléabilité » (*malleability or tailorability*) par certains auteurs [Morch 1995], [Morch & Mehandjiev 2000], et demeure assez proche de la notion de « flexibilité » proposée par [Scapin & Bastien 1997]. La malléabilité implique une adaptation par l'utilisateur lui-même : « *Une application malléable est à la fois utilisable et modifiable par ses propres utilisateurs. L'activité de redéfinition correspond elle-même à une des facettes de son utilisation.* » [Bourguin 2000]. Le fait de pouvoir modifier (adapter) les fonctionnalités d'un système sans pour autant l'interrompre [Betbeder-Matibet 2003] est un sujet de recherche complexe, car cela suppose un maintien et une gestion fine de la cohérence globale du système.

2.1.3 Qui (Who)

L'étape suivante nous amène à réfléchir à la question « Qui adapte ? ». Plus précisément, il s'agit de définir qui contrôle l'adaptation et qui la réalise, concrètement. Elle peut être exécutée soit par l'utilisateur, soit par le système, soit encore par les deux, conjointement. Selon [Browne & al. 1990], lorsque le système contrôle l'adaptation, on parle de « personnalisation » (*personalization*) et plus généralement « d'adaptativité » (*adaptivity*). Lorsque l'utilisateur contrôle d'adaptation, on parle alors de « customisation » (*customization*), notamment lorsqu'on a le choix entre plusieurs options, et, plus généralement, « d'adaptabilité » (*adaptability*). A notre connaissance, il n'existe pas de terme spécifique pour désigner le fait que l'adaptation soit à la fois réalisée par l'utilisateur et le système, même si certains auteurs parlent de « contrôle mixte ».

La personnalisation et la customisation visent à répondre aux besoins et aux caractéristiques particulières de chaque utilisateur [Cingil 2000]. En effet, ces mots viennent du domaine du commerce électronique et insistent sur une dimension individuelle de l'adaptation. Rosenberg explique : « *customization involves end users telling us exactly what they want, such as what colors or fonts they like, the cities for which they want to know the weather report, or the sports teams for which they want the latest scores and information. With customization, the end user is actively engaged in telling the content-serving platform what to do; the settings remain static until the end user reengages and changes the user interface* » [Rosenberg 2001]. Selon cet auteur, la personnalisation est spécifique à l'utilisateur final et se base sur ses intérêts implicites durant la session en cours, ainsi que les précédentes.

Certains auteurs comme [Villanova-Oliver 2002], [Habieb-Mammar 2004] et [Samaan 2006] font remarquer que parfois, personnalisation, customisation et adaptation sont synonymes dans les travaux de [Mobasher & al. 2000], [Kappel & al. 2000] ou [Rossi & al. 2001]. Afin de couvrir le processus complet de d'adaptation dynamiques de l'interface utilisateur, le framework ISATINE étend la taxonomie de Dieterich en spécialisant la théorie de l'action de Norman [López-Jaquero & al. 2008]. Pour [Frasincar & Houben 2002], la locution « adaptation dynamique » est employée pour désigner une adaptabilité tandis que la locution « adaptation statique » est plutôt employée pour une adaptativité. Pour Fischer, la différence fondamentale entre les systèmes adaptatifs (adaptive) et adaptable (adaptable) est la suivante : le premier terme désigne une adaptation par le système lui-même, à la tâche en cours et à l'utilisateur, alors que le second terme signifie que l'utilisateur change (avec parfois l'aide substantielle du système) les fonctionnalités du système [Fischer 2000].

Mais « contrôler l'adaptation » ne veut pas forcément dire pour autant « exécuter l'adaptation ». L'adaptation peut être effectuée par le système (adaptativité) ou par les utilisateurs (adaptabilité). Les utilisateurs peuvent être, par exemple, les concepteurs de systèmes ou les administrateurs, des experts ou des utilisateurs finaux [Edmonds 1981]. Une adaptation plus précise peut être proposée en tenant compte des groupes d'utilisateurs (sociaux, culturels...). On s'adapte alors en fonction de stéréotypes (utilisateurs typiques), par rapport aux besoins et aux niveaux de préférences établis pour chaque catégorie, ou bien encore pour des utilisateurs individuels, en fonction de leur personnalité et des connaissances que l'on détient à propos de chaque individu. Ici, c'est donc l'opposition « utilisateur » versus « système » qui entre en considération pour distinguer l'adaptativité de l'adaptabilité. Notre vue globale de l'adaptation, présentée plus loin, en Figure 7, permet de visualiser cela.

2.1.4 Quand (When)

Etudions à présent la question « Quand adapter ? ». L'adaptation peut être effectuée à différents moments. Le temps de la conception (*design time*) est souvent opposé au temps de l'exécution (*runtime*). Certains auteurs opposent ainsi l'adaptation statique (qui s'effectue avant et entre les sessions) à l'adaptation dynamique, qui s'effectue pendant l'interaction [Dietrich & al. 1993]. Au cours d'une session, l'adaptation peut se faire de manière continue, à des moments prédéfinis, avant ou après certaines fonctions prédéfinies, dans des situations particulières ou à la demande de l'utilisateur. [McKinley & al. 2004] distinguent, quant à eux, des temps précis de l'adaptation dynamique. Il peut s'agir d'une adaptation au développement, à la compilation, au chargement ou à l'exécution.

Dans la terminologie de [Stephanidis & al. 1998], l'adaptativité (*adaptive*) désigne le fait que les adaptations se produisent en cours d'exécution du système (peu importe qu'elles soient réalisées par le système ou par l'utilisateur), tandis que l'adaptabilité (*adaptable*) désigne le fait que les adaptations interviennent avant l'exécution du système, par exemple, lors d'une première installation [Kobsa & al. 2001]. Ici, c'est donc l'opposition « dynamique » versus « statique » qui entre en considération pour distinguer l'adaptativité de l'adaptabilité. Là encore, la Figure 7 permet de mieux apprécier ces disparités.

2.1.5 Où (Where)

Passons à présent à la question « Où adapter ? ». L'adaptation peut être interne ou externe. Une adaptation est dite interne (également appelée « fermée ») lorsque les mécanismes d'adaptation sont intégrés au système lui-même alors qu'elle sera dite externe (ou « ouverte ») lorsque l'adaptation se fait en dehors du système, par exemple grâce à un service Web [Oreizy & al. 2004].

Les mécanismes d'adaptation ouverts semblent être préférés dans des projets liés à l'informatique pervasive et ubiquitaire et aux technologies ambiantes, car ils permettent la possibilité de découvrir de nouveaux services au moment de l'exécution, et de faire l'adaptation à la volée, dans le respect de la philosophie des architectures orientées services (SOA). Nous reviendrons sur ce point plus loin dans ce document.

2.1.6 A quoi (To what)

Voyons maintenant la question « A quoi adapter ? ». En d'autres termes, quelle est la cible de l'adaptation ? A quoi le système doit-il s'adapter ? Classiquement, on dit que le système peut s'adapter à l'utilisateur, à la plate-forme, à l'environnement ou à une combinaison des ces trois éléments.

Les utilisateurs peuvent être perçus comme étant des utilisateurs typiques. Pour cela, des types d'utilisateurs auront été définis. Il peut s'agir par exemple d'utilisateurs novices, intermédiaires ou experts. Une personne peut également entrer dans une sous-catégorie d'utilisateurs connus (groupes socio-culturels, capacités physiques ou intellectuelles...). Ces catégories sont très souvent prédéfinies selon la tâche à effectuer. On distingue par exemple un usage « occasionnel », d'un usage « fréquent » ou encore « quotidien » voire « permanent ». En matière d'accès à des ressources ou des données particulières, l'adaptation pourra s'effectuer dès lors que l'utilisateur aura été identifié ou authentifié¹². L'adaptation peut également avoir lieu conformément aux rôles, droits, compétences, aptitudes, préférences, handicaps, culture... de l'utilisateur.

L'adaptation à la plate-forme sous-entend une adaptation aux matériels et logiciels utilisés. Cela signifie qu'une même application peut être adaptée et exécutée sur différents systèmes d'exploitation ou différents périphériques physiques, ou encore que l'application peut s'adapter aux ressources disponibles au moment de l'interaction.

L'adaptation à l'environnement exige des moyens de détection de variables physiques, comme l'emplacement (localisation, coordonnées GPS¹³), la détection de la lumière ambiante, du niveau sonore ou de la température dans laquelle se déroule l'interaction, etc.

¹² L'identification permet de connaître l'identité d'une entité (personne, ordinateur...) alors que l'authentification permet de vérifier cette identité, afin par exemple d'autoriser un accès à une ressource.

¹³ GPS : Global Positioning System

Lorsque les interfaces sont capables de s'adapter au contexte d'utilisation, certains auteurs parlent de plasticité [Thevenin & Coutaz 1999]. Le terme plasticité est inspiré de matériaux qui se dilatent et se contractent sous contraintes sans rupture. Là encore, le contexte d'utilisation d'un système interactif comprend les personnes qui utilisent ce système, la ou les plates-formes utilisée(s) pour interagir avec le système et l'environnement physique dans lequel l'interaction a lieu.

En plasticité, une cible est définie comme un triplet « utilisateur, plate-forme, environnement » envisagé par le concepteur du système. Au moment de l'interaction, ce même triplet sera reconsidéré pour éventuellement être adapté en fonction du contexte.

Dans ce cas, nous sommes très proche de la notion de flexibilité, décrite par [Oppermann 1994] : « *the goal is to fit the system to a specific user and to a specific task not only on the basis of requirements analysis made at the design phase, but also by prolonging the flexibility of the system into the usage phase by its adaptation capability.* ».

2.1.7 Comment (How)

La réponse à la question « comment adapter ? » est traitée dans la littérature selon deux principaux points de vue : les stratégies et les méthodes.

- Les stratégies

Quatre stratégies d'adaptation de base sont proposées par [Cockton 1987] : *enabling*, *switching*, *reconfiguring*, and *editing*. La stratégie dite *enabling* permet une adaptation par activation/désactivation des composants et des fonctionnalités de contrôle des processus des systèmes. La stratégie dite *switching* permet une adaptation par sélection d'une des différentes interfaces utilisateur préconfigurées ou de certains paramètres de l'interface utilisateur (couleurs, police, taille, etc.). La stratégie dite *reconfiguration* permet une adaptation par modification de l'interface utilisateur en utilisant des composants prédéfinis. Enfin, la stratégie dite *editing* permet une adaptation par édition, sans aucune restriction sur la base du modèle de dialogue. D'autres fonctionnalités ont été ajoutées à cette liste : adaptation par ajustement, par transformation, par fusion, par échange ou combinaison [Balint 1995].

- Les méthodes

Concernant les méthodes, il existe quatre principaux moyens identifiés de réaliser une adaptation. Il existe quatre approches possibles pour réaliser une adaptation. [Samaan 2006] et [Tarpin-Bernard 2006] dressent un remarquable état de l'art à ce sujet.

La première approche porte sur *la traduction des interfaces* [Raggett & al. 1999];

La seconde approche est basée sur *la rétroingénierie et la migration des interfaces* [Müller & al. 2000];

La troisième approche est *l'utilisation de langage à balises* [Lie & Bos 1999];

Enfin, la dernière approche se base sur *l'ingénierie dirigée par les modèles (IDM)* [Van den Bergh & al. 2004], [Benyon 1993], [Benyon & Murray 1993].

Les deux premières méthodes (*translation and reverse engineering*) sont considérées comme des adaptations ascendantes (approche bottom-up), car elles utilisent des logiciels ou systèmes préexistants et essaient de les adapter en fonction d'un nouveau contexte, différent de l'original. Cette démarche peut être vue comme une suite de processus qui apportent chacun une partie fondamentale du système que l'on cherche à adapter.

Les deux dernières méthodes (*markup languages and model-based approach*) sont considérées comme des adaptations descendantes (approche top-down) puisqu'elles essaient de créer des interfaces adaptées à un contexte particulier en se basant sur des spécifications. A l'image d'un artiste sculptant une pierre, on utilise ici une matière première brute, que l'on façonne et que l'on transforme étape par étape, tantôt en enlevant des parties indésirables, tantôt en ajoutant des nouvelles parties nécessaires. Bien entendu, une méthode est qualifiée d'hybride lorsqu'elle permet une adaptation faite grâce aux deux approches présentées ci-dessus utilisées conjointement.

La modélisation d'interfaces intelligentes et auto-adaptatives est largement étudiée par la communauté *Intelligent User Interfaces* (IUI) depuis de nombreuses années. Les systèmes ainsi développés se placent entre l'interface et l'application et permettent d'interpréter les actions de l'utilisateur afin de réadapter l'IHM au comportement de l'utilisateur et cela de façon à faciliter ses actions sur l'application [Kolski & al. 1992].

Nous venons de présenter différents aspects de l'adaptation. Comme nous l'avons vu, il n'existe pas de standard ou de définition communément adoptée par tous, permettant de discerner précisément toutes les formes d'adaptation possibles.

Quelques travaux sur l'adaptation 1981 - 2008



Figure 6 : Quelques contributeurs aux travaux sur l'adaptation

La Figure 6 présente une liste restreinte des contributeurs à ce domaine. Elle n'est, bien évidemment, pas exhaustive. Même si les premiers travaux sur les interfaces adaptatives datent des années 70 (systèmes éducatifs et Adaptive Hypermedia Systems), nous présentons un schéma uniquement centré sur les trente dernières années, pour montrer l'engouement pour cette thématique.

Notre vue globale de l'adaptation est proposée sur la Figure 7 ci-après. Elle n'a pas vocation d'être exhaustive, elle non plus. Elle reprend les principaux points exposés précédemment et montre certaines divergences de points de vue à propos des termes adaptabilité et adaptativité, notamment. Pour nous, l'adaptation est principalement appréhendée côté système et nous considérons qu'il s'agit de sa capacité à se reconfigurer lui-même ou tout au moins à réagir en exécutant une ou plusieurs actions en fonction d'un changement de contexte d'utilisation, tout en préservant son utilisabilité.

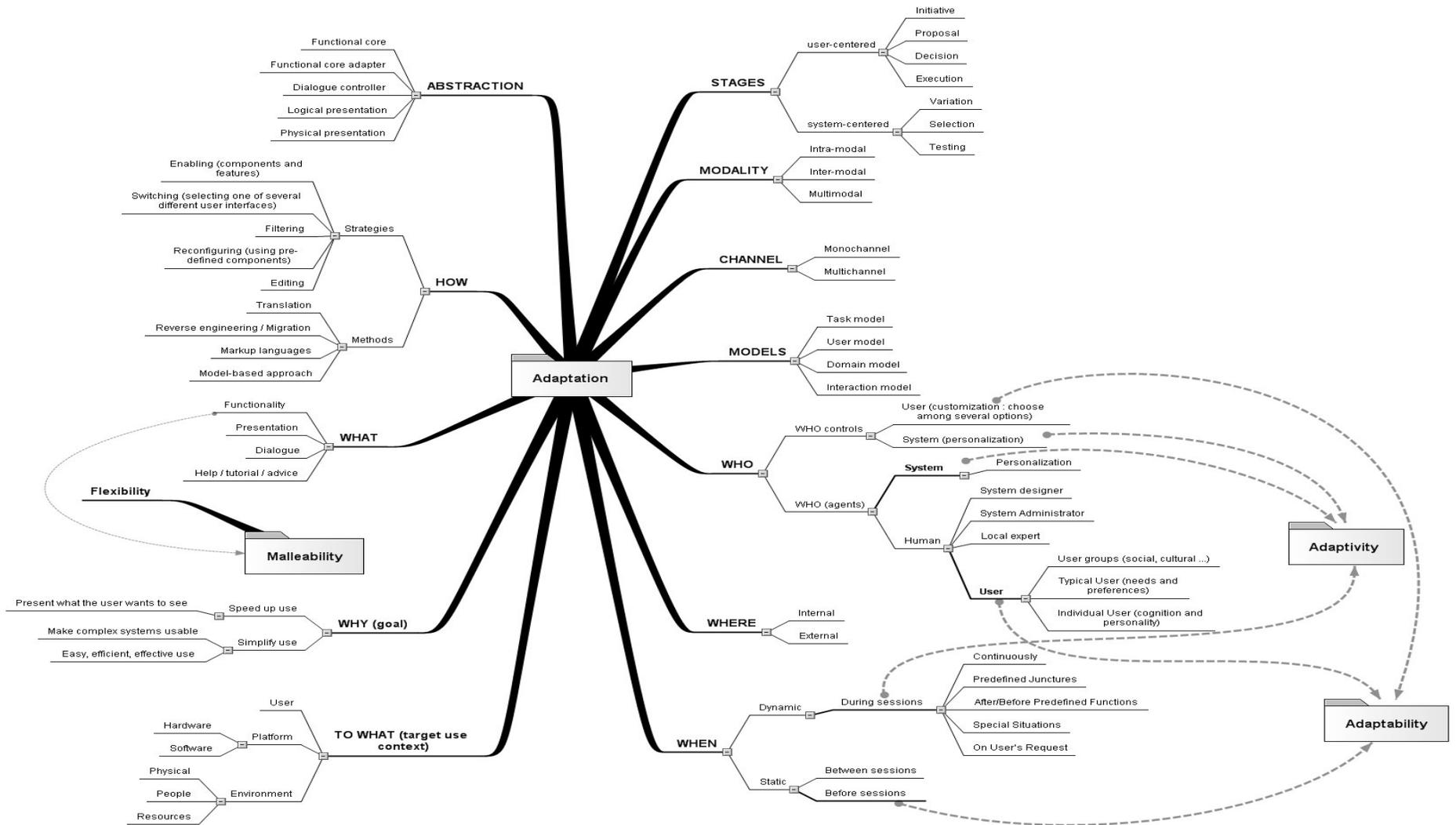


Figure 7 : Notre vue globale de l'adaptation

2.1.8 Adaptation d'un document numérique

Nous venons de voir les principales motivations menant à une adaptation de document numérique. Je vais à présent aborder les différentes manières de réaliser de telles adaptations.

Le besoin d'un accès ubiquitaire à l'information se concrétise notamment par l'apparition constante de nouveaux matériels et le succès des ordinateurs de poche ainsi que des téléphones mobiles. Les capacités de ces matériels qui vont du grand tableau interactifs aux petits téléphones portables et autres smartphones, sont très variables. Ceci appelle une réflexion sur les techniques de spécification d'interfaces indépendantes dont le but est d'éviter des développements spécifiques [Mitrovic & al. 2004]. Certains auteurs pensent que cette situation justifie l'étude scientifique de modèles et de technologies permettant de s'abstraire du périphérique employé par l'utilisateur [Myers & al. 2002].

Idéalement, on souhaite tendre vers des opérations de transformation pour tous et n'importe où (« *transforming the UI for anyone anywhere* » [Lopez & Szekely 2001]). Cette tendance n'est pas uniquement perceptible dans le domaine des IHM, mais également dans d'autres communautés, comme celle du document, au sens large, et des hypermédias, plus particulièrement : ainsi, l'hypermédia [...] peut faire appel à un modèle de tâche, un modèle de l'utilisateur, et un modèle du domaine afin d'élaborer la réponse la plus appropriée [Nanard & Nanard 2001].

La plasticité a été définie en IHM comme la possibilité, pour une interface, de s'adapter à son contexte d'usage dans le respect de son utilisabilité [Thevenin 2001]. Le dictionnaire de l'Académie Française définit la plasticité comme la « *qualité de ce qui peut prendre diverses formes* », et le Larousse la définit en psychologie comme la « *capacité d'un sujet à déstructurer un ensemble perçu et à le restructurer selon une forme différente* ». Un document numérique peut ainsi subir des modifications de son support, de sa forme, de sa structure, et cela en fonction d'un contexte d'usage particulier. Selon la définition classique de Dey, le contexte se définit par toute information qui peut être utilisée pour caractériser la situation d'une entité. Une entité est une personne, un lieu ou un objet considéré pertinent pour l'interaction entre un utilisateur et une application, y compris l'utilisateur et les applications elles-mêmes [Dey & Abowd 2000], [Dey & al. 2001].

Que les éléments du contexte soient données par l'utilisateur ou bien détectés automatiquement, la grande question qui demeure est la suivante : Que fait-on de ce contexte et comment l'utiliser efficacement ? En plasticité des IHM, un contexte d'usage est défini comme un triplet <E,P,U> où E est un des environnements concernés par l'application, P est une des plateformes concernées par l'application et U est un des utilisateurs concernés par l'application. A partir d'une spécification donnée, l'interface pourra ainsi être façonnée de multiples manières pour être utilisée sur des plateformes éventuellement hétérogènes.

Ainsi, classiquement, permettre l'accès à N informations au travers de M périphériques revient, pour le concepteur, à coder N x M pages (ou applications). En lieu et place de ces N x M applications, les chercheurs tentent de proposer un modèle pour plusieurs interfaces [Paternò & Santoro 2002], [Tarby 2004]. On applique alors une séparation fond/forme telle que décrite notamment par Bachimont et Crozat, permettant ainsi une adaptation par reconstruction : « *Tout contenu numérique consiste en une ressource qu'un calcul permet de mettre en forme dynamiquement pour sa consultation. On a donc d'un côté une ressource enregistrée et de l'autre une vue publiée reconstruite.* » [Bachimont & Crozat 2004]. On peut également noter des propositions de méthodes de conception d'IHM plastiques, basées sur des patrons de conception [Hariri 2008].

Ces aspects de génie logiciel concernent principalement la plasticité des interfaces utilisateurs [Calvary & al. 2001a], [Calvary & al. 2001b]. Le projet CAMELEON [CAMELEON] introduit à ce sujet un cadre de travail plus complet (cf. Figure 8) puisqu'il intègre la plasticité depuis le niveau de l'IHM finale jusqu'au niveau des tâches et concepts sous-jacents à l'application plastique associée. Le Cameleon Reference Framework (CRF) a ensuite donné lieu au Similar Adaptation Space (SAS) défini dans [Vanderdonck & al. 2005].

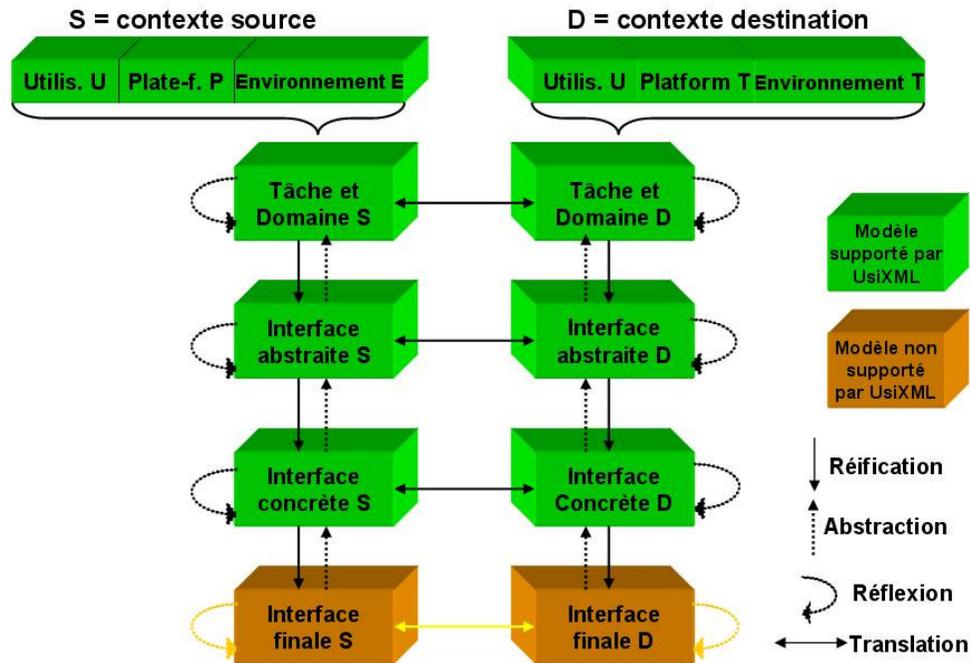


Figure 8. Le cadre de référence du projet CAMELEON

Ce projet s'intéresse aux étapes de développement des IHM d'applications interactives sensibles aux contextes (*context-aware applications*). Dans CAMELEON, en terme d'IHM, le niveau le plus abstrait est celui des tâches et concepts orientés métiers, et le niveau le plus concret est celui de l'interface finale. Or, même si ce cadre de référence semble complet, les travaux qui en ont découlé jusqu'à présent ne concernent que la plasticité des IHM, c'est-à-dire que ces travaux étudient comment une IHM doit s'afficher et s'adapter en fonction du contexte, sans réellement se préoccuper des documents que ces IHM manipulent.

2.1.9 De langages basés sur XML pour la plasticité des IHM

Dans l'optique de la plasticité des IHM, des langages de représentation abstraite ont été conçus depuis quelques années pour permettre aux développeurs de décrire une interface de manière abstraite. Toutefois, la plupart de ces langages restent relativement proches des modes d'interaction graphique, ce qui limite leur application avec d'autres types d'interaction tels que la voix [Lin & al. 2001].

Les nouvelles formes d'interaction envisagées se heurtent à de nombreuses difficultés parmi lesquelles le développement et le maintien des interfaces homme-machine sur chaque système. Pour être utilisable, l'interface utilisateur doit être capable de s'adapter aux dispositifs d'interaction, par exemple à la taille de l'écran, à la présence ou non d'un clavier ou à d'autres contraintes comme les profils ou les préférences de l'utilisateur. Cela a conduit les chercheurs à s'interroger sur les possibilités de mettre en œuvre de nouveaux langages particulièrement bien adaptés à la manipulation de données et aux traitements permettant de s'affranchir des limites des langages traditionnellement utilisés lors de la conception des IHM. Il faut pour cela disposer de langages permettant de dissocier le contenu, d'une part, et la présentation de ces données, d'autre part.

En ce qui concerne le World Wide Web, HTML combine données et présentations dans le même document. Inversement, XML¹⁴[XML] fournit une séparation entre les données et la manière dont on les présente. Dire qu'un langage comme XML est extensible signifie que le concepteur peut ajouter des balises pour ses propres besoins, de manière illimitée. L'utilisation conjointe de données et de feuilles de style XSL¹⁵ [XSL] permet d'afficher les mêmes données sous différents formats et différents aspects.

¹⁴ eXtensible Markup Language

¹⁵ eXtensible Stylesheet Language

De langages tels que UIML¹⁶ [UIML] [Merrick 2001], RIML¹⁷ [Priestersbach & al. 2003], ou encore AUIT¹⁸ [Grundy & Zou 2004], visent le niveau des interfaces concrètes (cf. Figure 8). Ces langages, fondés sur XML, permettent de séparer les données de leur présentation, et de s'abstraire des périphériques en utilisant des objets d'interaction abstraits.

A l'opposé de ces langages, les conceptions basées sur les modèles (« model-based ») [Eisenstein & al. 2001] permettent de s'abstraire des périphériques et de s'intéresser plus particulièrement aux modèles des utilisateurs, aux modèles d'interaction, aux modèles de tâches et aux modèles de contexte d'utilisation [Müller & al. 2001]. Ce type d'approche permet de spécifier les systèmes en se détachant des contraintes matérielles et logicielles. Cependant, le passage de spécifications abstraites à des systèmes réels demande encore beaucoup de travail étant donné la multiplicité des contraintes liées notamment aux périphériques, aux cas d'utilisation, aux types et aux niveaux des utilisateurs. Les langages, associés à ce type de conception, tels que XIML¹⁹ [XIML] ou UsiXML²⁰ [UsiXML], [Vanderdonck & al. 2004], visent avant tout les niveaux les plus abstraits du cadre de référence CAMELEON (tâches et concepts, interfaces abstraites). Pour un état de l'art plus complet des langages de description basés sur XML, le lecteur pourra consulter [Limbourg & al. 2004].

2.2 Contribution à cet axe de recherche

Dans notre équipe de recherche, nous cherchons à modéliser les différents types d'adaptations étudiées et à proposer des solutions pour mettre en œuvre des systèmes capables, pour cela, de considérer des éléments pertinents pour l'interaction.

Typiquement, on s'intéresse :

- au couplage/découplage de canaux : une tâche commencée sur un canal peut-elle être achevée sur un autre ? (ex : Internet puis téléphone). Une tâche exécutée sur deux canaux en parallèle est-elle plus rapidement achevée que sur un seul de ces deux canaux ?
- aux problèmes de rupture de canaux, volontaire ou non : est-il possible de récupérer sur un autre canal le contexte de la tâche en cours ? Facilement ? Rapidement ? Complètement ?
- à la granularité temporelle : l'historique des actions se gère-t-il à la seconde, à l'heure, au mois ?
- aux rôles et profils des utilisateurs : comment exploiter au mieux le profil et le rôle d'un utilisateur dans un tel contexte : un même document ne sera pas présenté de la même manière selon le type d'utilisateur (par exemple : enseignant versus étudiant) ?
- au contexte de la tâche : en quoi le contexte influe-t-il sur une tâche supportant le multicanal : contexte géographique, temporel, émotionnel, matériel, etc. ?

S'interroger à propos de ces questions nous a amené à étudier la combinaison des canaux et des tâches afin d'élaborer un modèle théorique et conceptuel nous permettant de prédire jusqu'à quel point un système multicanal peut répondre de façon pertinente, robuste et efficace à une tâche dans un contexte défini. Ce modèle théorique que nous avons déjà commencé à développer [Chevrin 2006] s'appuie également sur nos connaissances de bases, au sein de l'équipe, en matière de collaboration interpersonnelles et des apports des sciences humaines (par exemple sur la communication et l'intersubjectivité, la théorie de l'activité [Norman & Draper 1986] et les théories instrumentales [Rabardel 1995]) et celles des sciences du langage (linguistique pragmatique et psycholinguistique comme avec le concept de Grounding [Clark & Brennan 1991]).

Pour ma part, j'ai contribué à l'avancement de ces travaux en me focalisant sur la plasticité des interfaces, l'adaptation en situation de mobilité et l'adaptation dynamique en fonction du contexte.

2.2.1 Le langage PlasticML

Bien que l'approche basée sur une abstraction des périphériques soit attrayante, elle nous paraissait trop souvent réductrice car trop ciblée sur l'IHM en terme de boutons, de menus, etc. et ne tenant pas compte des documents manipulés. Nous avons abordé cet aspect de la plasticité avec l'invention du langage PlasticML, développé grâce à l'aide de plusieurs groupes d'étudiants [Rouillard 2003]. Il permet de décrire des éléments d'entrée et de sortie d'interfaces, à un haut niveau, c'est-à-dire sans faire référence au périphérique qui sera utilisé lors de l'interaction.

¹⁶ User Interface Markup Language

¹⁷ Renderer-Independent ML

¹⁸ Adaptable User Interface Technology

¹⁹ eXtensible Interface Markup Language

²⁰ USer Interface eXtensible Markup Language

Grâce à des transformations avec XSLT²¹, les documents en PlasticML sont traduits automatiquement vers des langages idoines (HTML pour le Web, WML pour le WAP, VoiceXML [Anderson & al. 2001] pour le téléphone). PlasticML dispose de balises suffisamment abstraites pour décrire une tâche (ou une action) sans pour autant devoir définir quels objets informatiques (boutons, menu déroulant, cases à cocher, etc.) devront être utilisés pour un périphérique spécifique. Nous avons également proposé un outil permettant aux concepteurs de créer plus facilement des documents plastiques (voir en annexe 4 la boîte à outils qui accompagne le langage PlasticML).

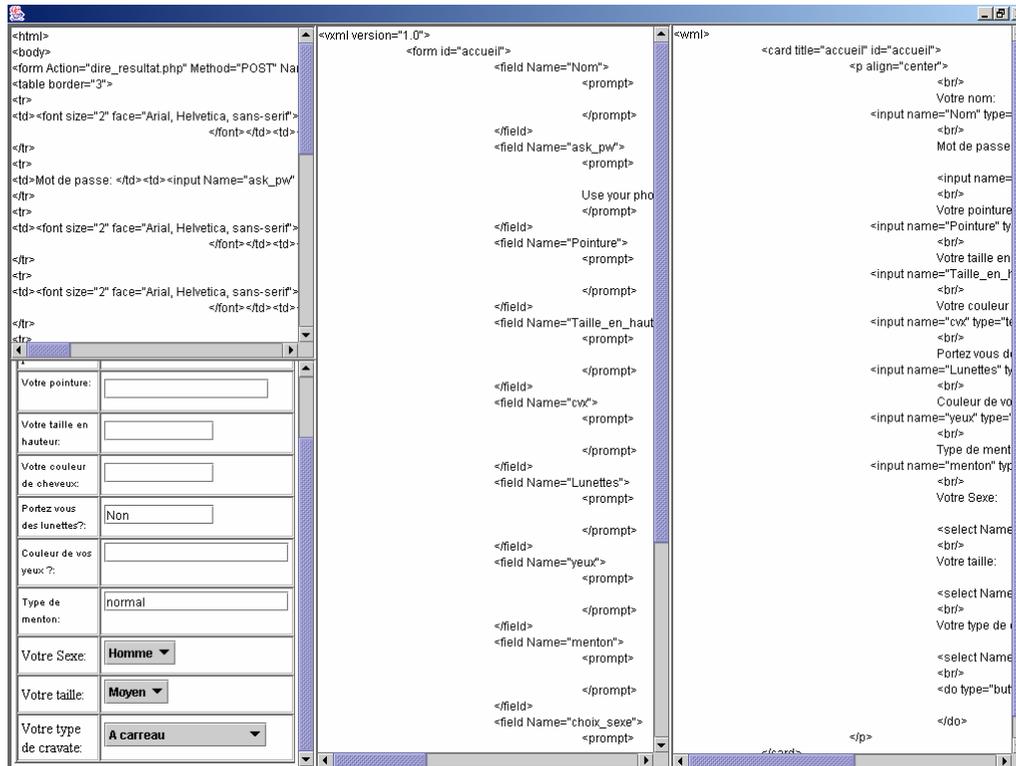


Figure 9 : Un aperçu de la génération vers plusieurs cibles différentes à partir d'une même interface abstraite, codée en Plastic ML

La Figure 9 donne un aperçu de la génération de code automatique dont est capable cet outil. La première colonne correspond au code HTML généré pour une cible graphique (on note que dans la fenêtre du bas, un navigateur présente le résultat final de l'interface). La deuxième colonne correspond au code VoiceXML généré pour une cible vocale. Enfin, la troisième colonne correspond à la génération automatique du code WML pour périphérique supportant le protocole WAP.

Notre approche, vis-à-vis de la plasticité des documents numériques a consisté à identifier et à séparer strictement les éléments de forme des éléments de fond, mais aussi à prendre en compte la portée sémantique [Rouillard 2003b] et pas uniquement syntaxique du document manipulé [Rouillard & al. 2006]. Pour nous, l'intention de l'auteur ne doit pas être dénaturée lors de l'adaptation (voir en annexe 5 l'exemple de la poésie « Le dormeur du Val » d'Arthur Rimbaud).

Nous avons insisté sur l'adaptation par rapport au rôle de l'utilisateur en décrivant dans PlasticML différents niveaux de rôles, hiérarchiques ou pas. De même, nous avons montré l'intérêt d'une adaptation conjointe (au périphérique et à l'utilisateur par exemple), en s'interrogeant sur la manière d'adapter les documents. La méthode « pipeline » est souvent utilisée : on enchaîne les modifications les unes à la suite des autres, mais il n'est pas certain que ce soit la meilleure façon de procéder.

Enfin, pour tenté de mieux prendre en compte l'activité humaine au moment de l'adaptation, nous avons augmenté le niveau d'abstraction du langage PlasticML en permettant aux concepteurs de définir des scripts de manière abstraite pour permettre aux systèmes informatiques de générer, à

²¹ eXtensible Stylesheet Language Transformation

l'exécution, les scripts nécessaires pour chaque cible visée, alors que jusqu'à présent, les scripts accompagnant les langages ciblés n'étaient pas pris en considération.

Dans un cadre appliqué, cela a conduit à une critique de certaines plateformes d'enseignement à distance [Bouyer & al. 2006]. Il était, en effet, impossible de concevoir un objet pédagogique adaptable (respectant la norme SCORM²²), car ces outils ont été pensés et conçus pour une utilisation dans un contexte fixe (sur PC, avec clavier/souris). De nombreux autres terrains d'applications pour l'adaptation ont été explorés. Comme nous l'avons déjà signalé, nos terrains d'expérimentations sont essentiellement tournés vers deux domaines particuliers : le commerce électronique et l'apprentissage. Concernant l'apprentissage, notre contribution se caractérise par une volonté d'assister l'utilisateur tout au long de l'interaction. Cette assistance fait également partie de l'adaptation du système à l'environnement. Elle doit prendre en compte les modèles que nous avons évoqués précédemment (modèle de l'utilisateur, de la tâche, etc.).

Très tôt, nous avons proposé, Jean-Claude Tarby et moi-même, la notion de coach et d'assistant numérique, capable de s'adapter à la tâche (sport, cuisine, réparation mécanique, assemblage de meuble en kit, enseignement, géolocalisation, vente ...), au périphérique, à l'utilisateur, et au contexte d'interaction. Ce dernier point est primordial pour que le système puisse s'adapter en temps réel aux éléments du contexte. Par exemple, il ne faut pas stimuler et solliciter davantage un sportif lorsque l'on décèle qu'il atteint un niveau critique pour ses fonctions vitales (pulsations cardiaques en particulier). Nous poursuivons actuellement nos recherches sur cette piste, en privilégiant le mode d'interaction vocal, couplé avec d'autres moyens d'interactions actifs ou passifs (GPS).

Par ailleurs, j'ai contribué à la réflexion autour du langage UsiXML, en étudiant avec d'autres chercheurs impliqués dans ce projet, des pistes envisageables pour permettre une utilisation multimodale de ce langage d'adaptation. Il a été montré que l'utilisation conjointe du vocal et du gestuel est possible, mais délicate à mettre en œuvre lors d'une adaptation à l'exécution [Lepreux & al. 2007]. Cela m'amène à présenter l'autre axe de recherche de mes travaux, que j'aborde dans le chapitre suivant : la multimodalité et la multicanalité.

2.2.2 Adaptation pour les applications mobiles

J'ai collaboré avec Jean-Claude Tarby (spécialiste des modèles de tâches, notamment), dans le cadre de l'adaptation des applications par rapport au contexte d'utilisation. Nous avons choisi de focaliser nos travaux sur l'adaptation des IHM en situation de mobilité et sur l'assistance aux utilisateurs. En effet, l'évolution des aides et assistances dites « intelligentes », la présence de plus en plus forte de la personnalisation dans les applications interactives, l'omniprésence des périphériques mobiles, et la qualité sans cesse croissante des synthèses vocales, nous ont conduits à proposer la notion de *coach numérique* [Tarby & Rouillard 2006]. Un coach numérique est plus qu'une aide intelligente. Ses caractéristiques sont semblables à celles d'un coach réel, mais il bénéficie en plus des avantages liés à la technologie informatique. Il peut capter des éléments du contexte, par exemple.

Pour valider nos idées, nous avons réalisé un prototype qui se veut générique, dans le sens où il n'est pas dédié à une activité spécifique. Nous pouvons en effet l'utiliser dans différents domaines : coaching sportif, conseiller technique pour assembler un meuble en kit, mécanique auto, aide à la préparation d'une recette de cuisine, etc. Notre prototype est constitué de deux applications. Une première application sur PC génère les fichiers nécessaires à la tâche à réaliser²³, et une autre, embarquée sur périphérique mobile, les exploite. Le prototype a été développé en langage C#, sous Visual Studio 2005. La bibliothèque utilisée pour la synthèse vocale est ActiveTTS²⁴, et celle utilisée pour jouer les fichiers audio est « hekkus »²⁵ pour Pocket PC.

Le fonctionnement du prototype se déroule en deux temps (cf. Figure 10). Dans un premier temps, on fournit à l'application PC un fichier texte, comprenant autant de paragraphes que d'étapes à réaliser, pour une tâche donnée. Dans un second temps, l'application génère N fichiers textuels, au format .txt, ainsi que N fichiers vocaux, au format .wav ou .mp3 par exemple (cf. Figure 11).

²² SCORM : Sharable Content Object Reference Model

²³ A terme, l'utilisateur pourra également choisir de se connecter sur un serveur Web, s'il préfère ne pas installer d'application sur sa propre machine.

²⁴ <http://www.guangmingsoft.net/activetts>

²⁵ Hekkus project, <http://www.shlzero.com>

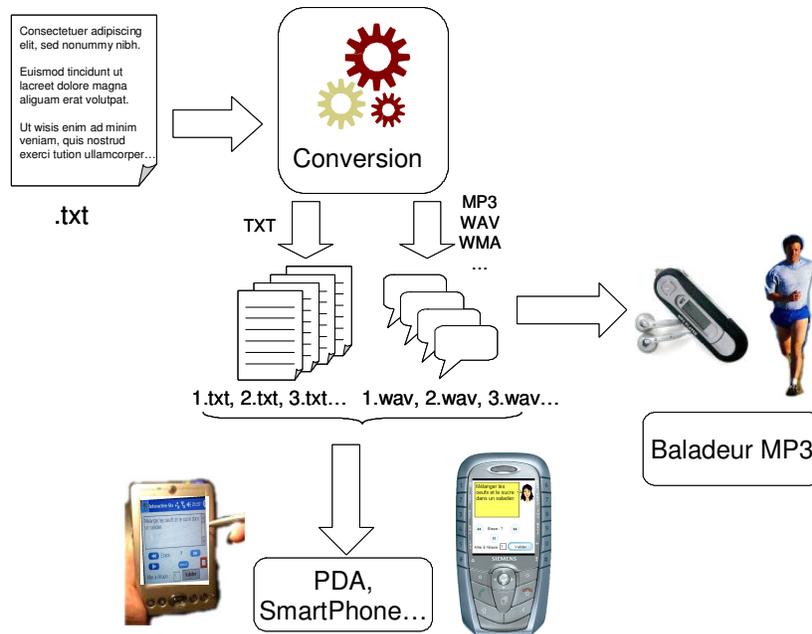


Figure 10 : Principe général du coach numérique

L'utilisateur peut sélectionner les caractéristiques techniques de la voix de synthèse qu'il souhaite appliquer. Dans notre prototype, il est possible de préciser le genre de voix (homme, femme, enfant...), la langue (français, anglais, italien...), le volume, la vitesse d'élocution, la tessiture de la voix, et le format audio voulu (par exemple 8 KHz, 8 bits, mono). A terme, ces choix seront effectués de manière (semi)-automatique, selon la tâche à effectuer et selon certains éléments contextuels.

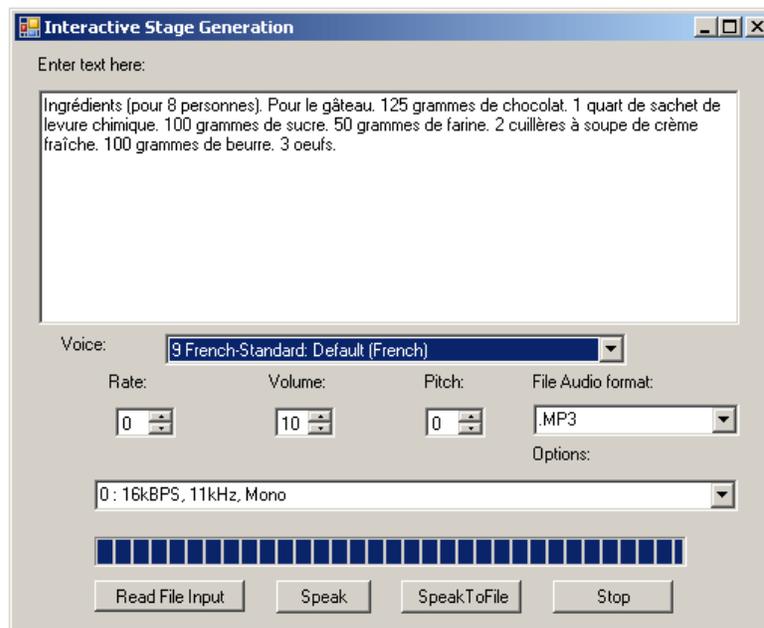


Figure 11 : Interface pour la génération des fichiers textuels et de synthèses vocales.

Les fichiers textuels et sonores obtenus sont ensuite transférés sur le périphérique mobile. Dans notre étude, nous avons utilisé un Smartphone HP iPAQ HW6500. La seconde application, quant à elle, est embarquée dans le périphérique mobile et exploite ces données en affichant pour chaque étape le texte associé, et en diffusant la synthèse vocale qui l'accompagne.

La Figure 12 illustre cette phase d'utilisation. Il sera aisé, dans une prochaine version du prototype, d'ajouter une photo illustrative (des images au format .gif animé, par exemple) ou des vidéos à chaque étape.

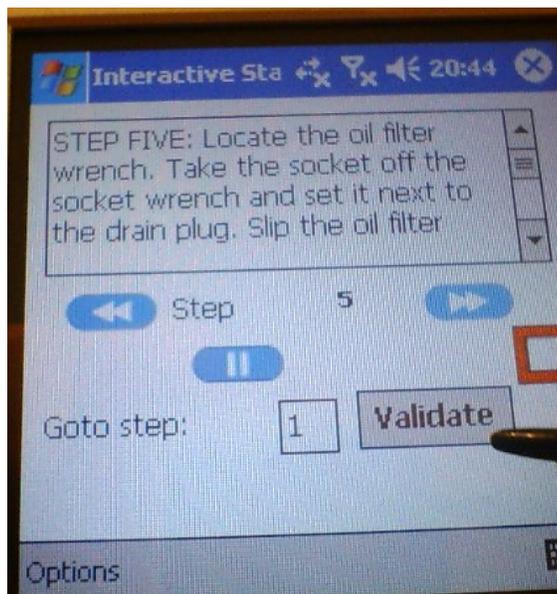


Figure 12 : Visualisation et écoute de l'étape n° 5 du changement du filtre à huile d'une voiture, en anglais, sur Smartphone

Le logiciel embarqué est plus qu'un simple lecteur de fichiers sonores. Il présente simultanément les éléments d'une même étape et permet d'interrompre et de reprendre une synthèse vocale, de la rejouer, de passer à une autre étape, de changer la langue de l'interface, etc.

La Figure 13 montre que l'interface est effectivement adaptable à la langue de l'utilisateur et que les boutons d'actions sont recalculés dynamiquement à chaque événement, pour ne proposer que les actions possibles dans un état particulier (par exemple une « pause » audio ne sera disponible que si un fichier sonore est en cours de diffusion).

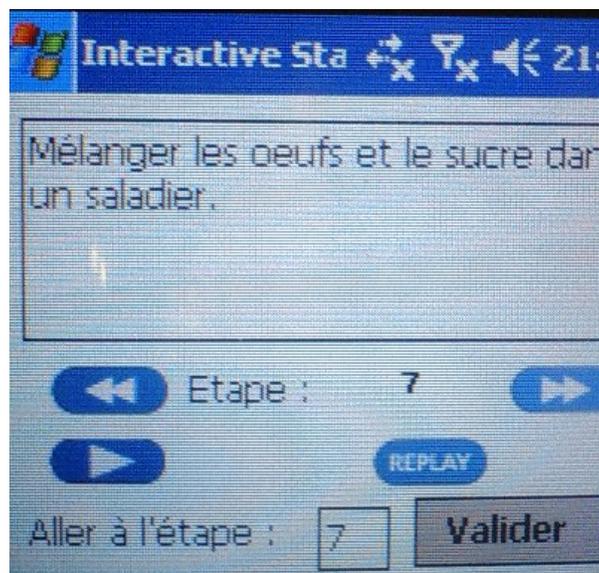


Figure 13 : Visualisation et écoute de l'étape N°7 de la recette de cuisine, en français, sur Smartphone.

Nous avons effectué des tests sur nos prototypes d'applications embarquées, avec la synthèse vocale de Loquendo²⁶ (et la voix dite « Juliette »). Cette synthèse vocale nous semble particulièrement adaptée au coaching numérique car elle intègre des éléments prosodiques émotionnels qui, à notre avis, peuvent avoir un impact positif sur le comportement de l'utilisateur. Il est ainsi possible de lui faire jouer différentes « émotions », comme par exemple des étonnements, encouragements, rires, acquiescements, sifflements, soufflements, faibles ou fortes respirations, raclements de gorge, déglutitions, etc.



Figure 14 : Exemple de coach numérique en Flash pour application mobile

Nous avons également réalisé d'autres prototypes, dont celui présenté en Figure 14, qui intègre un « player » Lite 2.0 de Flash 8 dans le but de générer nos coachs directement en Flash pour périphériques mobiles. Cet exemple montre que l'utilisateur peut visualiser son coach (attitudes, expressions faciales...), en même temps qu'il l'entend donner ses instructions, qui sont également reproduites textuellement. Cette redondance volontaire d'informations en sortie permet à l'utilisateur de choisir, selon le contexte, sa modalité d'interaction avec le système. La prochaine étape consistera à installer des capteurs sur le corps de la personne (lors d'un jogging par exemple), afin de disposer d'une large multimodalité en entrée grâce à des entrées au clavier ou au stylet plus une reconnaissance vocale et des capteurs actifs ou passifs. Un capteur passif ne requière pas d'actions particulières de la part de l'utilisateur (une pulsation cardiaque sera automatiquement détectée par exemple), tandis qu'un capteur actif déclenchera un événement à la demande de l'utilisateur (il pourra signifier que le rythme proposé par le coach lui paraît trop intense, ou pas assez soutenu, etc.).

2.2.3 Les QR Codes Contextuels

Toujours à propos d'adaptation, j'ai récemment proposé la notion de QR Codes contextuels, pour permettre, dans un contexte d'informatique ubiquitaire et pervasive, une adaptation en fonction d'un contexte, donné ou détecté. Les QR Codes (pour Quick Response Code) sont des codes barres à 2 dimensions, présentées sous forme d'une matrice. Le QR code a donc l'avantage de pouvoir stocker beaucoup d'informations tout en restant de petite taille (rapide à scanner). Il a été créé par l'entreprise japonaise Denso-Wave [Denso Wave] en 1994.

La notion de QR Codes contextuels repose sur l'idée d'une utilisation des QR Codes en fonction du contexte. En y regardant de plus près, deux sous-domaines d'étude sont ici ouverts. Le premier volet, que j'ai commencé à étudier, mais qui ne sera pas présenté ici, consiste à générer des QR Codes numériques, de manière dynamique, au moment de l'interaction, en fonction d'un contexte. Le second volet d'étude, que je vais détailler, consiste à utiliser des QR Codes existants (fixes), comme support à une adaptation dynamique et contextuelle. Autrement dit, il s'agit, dans le second cas, d'un couplage entre un QR Code et certains éléments du contexte disponibles au moment de l'interaction [Rouillard 2008]. Cette notion est assez proche de celle du « contextual bookmark », qui est une combinaison entre une photo d'un objet physique (une affiche publicitaire, une gare de train, une œuvre d'art dans

²⁶ <http://www.loquendo.com>

un musée...) prise avec un appareil mobile et les meta-informations (date, heure, lieu, température, voire préférences et intentions de l'utilisateur) relative à cet objet physique [Henze & al. 2007].

La Figure 15 présente les parties publiques et privées d'un QR Code contextuel. La partie privée peut contenir une ou plusieurs informations parmi les suivantes: le profil de l'utilisateur, la tâche en cours, le dispositif physique utilisé, le lieu, la date et l'heure, l'environnement de l'interaction. Le dispositif mobile décode le QR code et fusionne ces informations publiques avec les données privées obtenues au cours de l'interaction. Ensuite, le fichier (au format XML) ainsi créé est envoyé à un service web (que j'ai développé en C # sous .NET) qui traite les messages reçus et renvoie des informations adaptées au contexte.

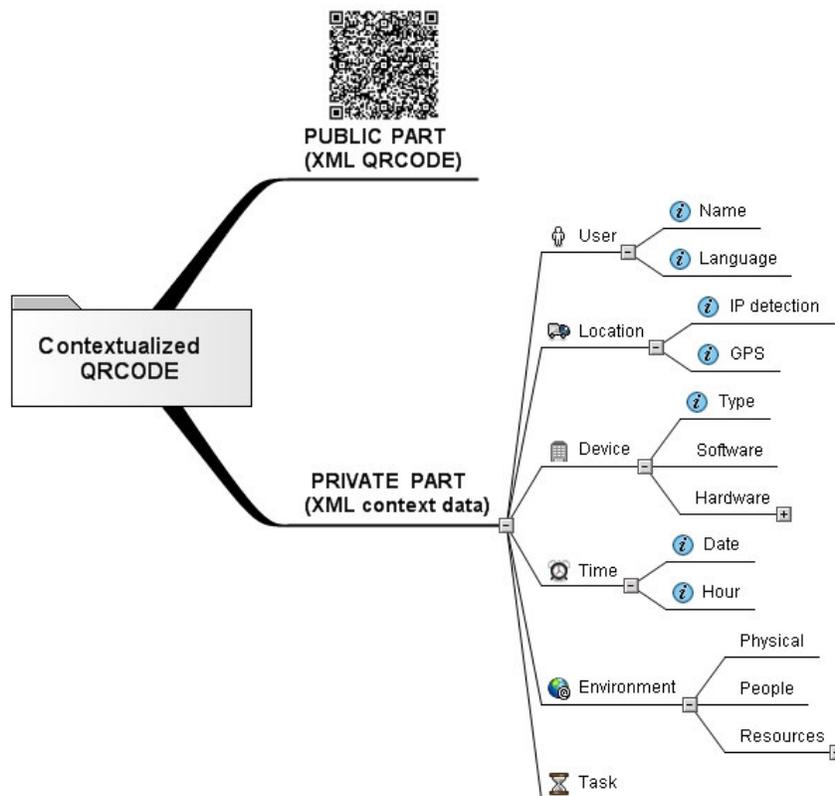


Figure 15 : Parties publique et privée d'un QR Code contextuel

Les exemples d'applications sont nombreux. Le premier scenario est celui du fameux message de bienvenue « Hello World ». Ce message est adapté au profil de l'utilisateur (ici, il s'agit de la langue déclarée et du nom du propriétaire du dispositif mobile), à l'heure de l'interaction et à la localisation géographique. Une utilisatrice pourra obtenir le message « Good morning Jenifer Diaz from Chicago, Illinois, USA. » si elle scanne le QR code à 8 heures du matin et que son adresse IP est la suivante 209.9.235.121 ; alors qu'un autre utilisateur scannant exactement le même code à 22 heures, en France, pourrait obtenir, par exemple, le message « Bonsoir Charles Durand de Lille, Nord-Pas-de-Calais, France. » si son adresse IP²⁷ est la suivante 134.206.10.69.

²⁷ La localisation peut se faire par exemple avec Fraudlabs (<http://www.fraudlabs.com>) :

```
<?xml version="1.0" encoding="utf-8"?>
<IP2LOCATION xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://ws.fraudlabs.com/">
  <countrycode>fr</countrycode>
  <countryname>france</countryname>
  <region>nord-pas-de-calais</region>
  <city>lille</city> <latitude>50.633</latitude> <longitude>3.067</longitude>
  <ispname>universite des sciences et technologies de lille</ispname>
  <domainname>cueep-sallaumines.univ-lille1.fr</domainname>
</ip2location>
```

Un deuxième exemple consiste à utiliser les QR Codes pour, en quelques sortes, émarger une feuille de présence numérique. Lorsqu'un utilisateur scanne le code approprié avec son appareil mobile, une fusion entre données publiques et privées s'effectue, de sorte que l'enregistrement de cette personne à cet événement (une réunion, par exemple) se fait automatiquement. Un autre exemple d'utilisation de QR Codes en contexte est illustré par la Figure 16, avec le système PerZoovasive [Rouillard & Laroussi 2008]. Il s'agit ici d'une enseignante d'école primaire qui cherche à obtenir des informations sur des animaux lors d'une visite au zoo avec sa classe. Elle utilise un smartphone équipée d'une carte²⁸ 3G.



Figure 16 : Un enseignant utilisant le système PerZoovasive au zoo de Lille

Avec notre système perZoovasive, les informations obtenues à propos des animaux rencontrés dépendent de la tâche, de la langue et du niveau de la classe.

La Figure 17 montre le texte obtenu lorsque l'on a scanné le QR Code apposé sur la cage des singes Gibbons, dans le cadre d'une leçon pour des élèves de niveau CM, en langue française.

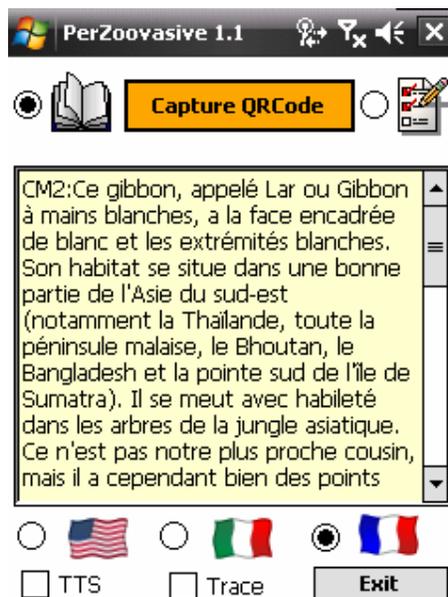


Figure 17 : Exemple de texte généré avec les paramètres suivants : cage=123, level=CM2, language=French, task=lesson

²⁸ Carte 3G aimablement fournie par notre partenaire Orange, du projet P-LearNet.

La Figure 18, quant à elle, montre le texte obtenu lorsque l'on a scanné le QR Code apposé sur la cage des panthères, dans le cadre d'un contrôle des connaissances, pour des élèves de niveau CM1, en langue anglaise.

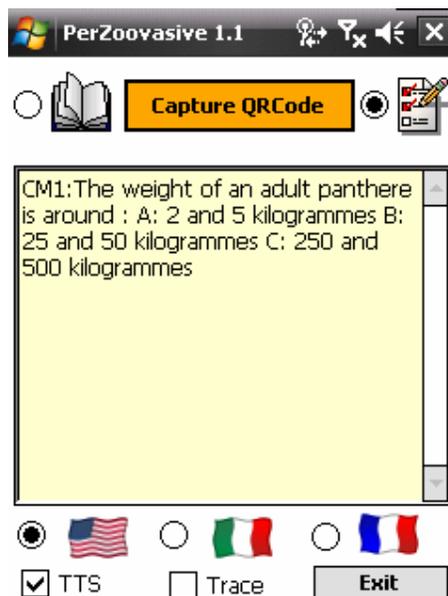


Figure 18 : Exemple de texte généré avec les paramètres suivants :
cage=8245, level=CM1, language=English, task=quiz

En cochant la case « TTS » (en bas à gauche de l'interface graphique) on active la génération d'une synthèse vocale (appel au service web de TTS que j'ai développé pour les partenaires du projet P-Learnnet), qui sera jouée, de manière multimodale redondante, en plus de l'affichage graphique. Cela peut permettre à l'enseignant de faire écouter cette information sonore à l'ensemble de son groupe d'élèves. Nos évaluations ont montrées que si le texte à synthétiser est long (de l'ordre de deux ou trois pages d'écrans de smartphone), le temps d'attente peut aller jusqu'à trois minutes environ.

Globalement, il faut retenir de cette première partie de ma présentation que j'ai tenté, au fil de ces travaux de recherche sur cet axe « adaptation », de travailler au plus près des préoccupations des utilisateurs finaux. Dans un cycle de réflexion et de développement en spirale, j'ai proposé des solutions simples, mais efficaces à des problèmes connus de l'adaptation. Cela m'a permis de proposer des outils (langage et boîte à outils) favorisant le prototypage rapide. A partir de cette démarche, des ouvertures vers d'autres pistes, mêlant multimodalité, informatique pervasive et mobilité des utilisateurs ont émergées. La seconde partie du document montrera comment elles ont été abordées.

Contexte de recherche de l'axe « Adaptation » :

- Thèse de Vincent Chevrin
- Master Recherche d'Antoine Bouyer
- Master Recherche d'Olivier Cremetz
- Master Recherche de Vincent Chevrin
- Projets MIAOU / NIPO / EUCUE

Publications : (les références en gras sont présentées en intégralité en annexes)

[Vantrois & Rouillard 2002] Vantrois Thomas, Rouillard José, Workflow and Mobile Devices in Open Distance Learning, IEEE International Conference on Advanced Learning Technologies, ICALT 2002, IEEE press, 2002, pp. 123-126, Kazan, Tatarstan, Russie.

[Laroussi & Rouillard 2002] Laroussi Mona, Rouillard José, Adaptive communication tools for mobile learning, ABIS'02, 2002, Hannover, Germany.

[Rouillard 2003] Rouillard José, Plastic ML and its toolkit, HCI International 2003, Lawrence Erlbaum Associates, Inc., Volume : 4, 2003, pp. 612-616, Heraklion, Crete, Greece, 2003.

[**Rouillard & al. 2006**] Rouillard José, Chevrin Vincent, Tarby Jean-Claude, Derycke Alain, La plasticité des documents numériques. Revue TSI (Technique et Science Informatique). Numéro spécial, « Document numérique », sous la coordination de Rémy Mullot, n°4, Paris, 2006.

[**Tarby & Rouillard 2006**] Tarby, Jean-Claude, Rouillard, José, Améliorer ses performances grâce au coaching numérique mobile : l'exemple du coaching vocal embarqué, UBIMOB 2006, 2006, Paris.

[Tarby & Rouillard 2006b] Tarby Jean-Claude, Rouillard José, Assistance, advice and guidance with digital coaching, EAM'06 European Annual Conference on Human Decision-Making and Manual Control, Lecture Notes in Computer Science, Springer-Verlag (Berlin), Valenciennes, 2006.

[**Bouyer & al. 2006**] Bouyer Antoine, Rouillard Jose, Derycke Alain, Pourquoi il n'est pas possible actuellement d'utiliser des objets pédagogiques plastiques respectant la norme SCORM, IHM 2006, AFIHM, Montréal, 2006.

[Lepreux & al. 2007] Lepreux Sophie, Hariri Anas, Rouillard Jose, Tabary Dimitri, Tarby Jean-Claude, Kolski Christophe, Towards Multimodal User Interface Composition based on UsiXML and MBD principles, HCI International 2007, Lecture Notes in Computer Science, Springer-Verlag, Berlin, Beijing, Chine, 2007.

[Kaddouci & al. 2008] Kaddouci Sarra, Bouzeghoub Amel, Caron Pierre-Andre, Lecocq Claire, Le Pallec Xavier, Ritaine François-Julien, Rouillard José, Experiments in ubiquitous computing for communities of practice using learning resources, International Conference and Exhibition on Next Generation Mobile Applications, Services, and Technologies (NGMAST 2008), IEEE, pp. 254-259, Cardiff, Wales, UK, 2008.

[**Rouillard 2008**] Rouillard José, Contextual QR Codes, Third International Multi-Conference on Computing in the Global Information Technology (ICCGI 2008), IEEE, Conference Publishing Service, ISBN 978-0-7695-3275-2, pp. 50-55, Athens, Greece, 2008.

[Rouillard & Laroussi 2008] Rouillard José, Laroussi, Mona, PerZoovasive: contextual pervasive QR codes as tool to provide an adaptive learning support, International Workshop On Context-Aware Mobile Learning - CAML'08, IEEE/ACM SIGAPP, Cergy-Pontoise - Paris, 2008.

3 Multimodalité et multicanalité

L'étude de la multimodalité et de la multicanalité constitue notre second axe de recherche. Nous allons montrer que dans nos travaux, les notions d'adaptation présentées dans la première partie de ce document, sont une des composantes en faveur de l'usage de la multimodalité. En fonction du contexte le système personnalise le document et/ou l'interface à présenter à l'utilisateur, tout en utilisant au mieux les canaux et modalités disponibles.

3.1 Axe de recherche

La notion d'interfaces multimodales a été introduite par Richard Bolt [Bolt 1980] dans son système « put-that-there », qui associe commandes vocales et techniques de pointage, comme le montre la Figure 19. Pour déplacer un objet, l'utilisateur pointe du doigt cet objet en prononçant « mets ça », puis sur la destination souhaitée en prononçant « là ».



Figure 19 : Le système « put-that-there » de Bolt

Dragicevic montre dans sa thèse [Dragicevic 2004] que d'autres prototypes ont été construits par la suite afin d'expérimenter l'usage combiné de la manipulation directe et du langage naturel textuel [Cohen & al. 1989], ou des gestes et de la parole [Weimer & Ganapathy 1989].

Des outils de conception et de réalisation d'interfaces multimodales sont disponibles. Il s'agit par exemple des outils du W3C [W3C 2003], de FAME²⁹ [Duarte 2006], des approches à creusets [Nigay 1994], à base de règles [Bellik 1992], ou ceux qui proposent des solutions plus complètes, comme PetShop [Bastide & al. 1998], mais qui demandent un haut niveau de programmation afin de coder les mécanismes propres à chaque agent. Les concepteurs proposent alors des boîtes à outils permettant de prendre en main plus aisément les solutions proposées (par exemple OIDE³⁰ [Serrano & al. 2008] est un outil graphique construit au-dessus du noyau d'OpenInterface [OpenInterface]) et/ou fournissent un système de génération automatique du code (c'est le cas pour l'assemblage des composants dans ICARE [Bouchet & Nigay 2004]).

Bouchet, dans sa thèse, fait un recensement et un classement des approches connues, selon différents critères³¹ et conclut ainsi : « *il n'existe pas d'outil prenant en compte l'intégralité des critères que nous avons définis pour concevoir et développer rapidement, simplement et efficacement n'importe quelle interaction multimodale.* » [Bouchet 2006].

²⁹ FAME : Framework for Adaptive Multimodal Environments

³⁰ OIDE : OpenInterface Interaction Development Environment

³¹ Ces critères sont notamment la couverture du cycle de développement, la cible utilisateur (concepteurs, informaticiens, programmeurs, etc.) le pouvoir d'expression, la nature des combinaisons possibles, etc.

Pour étudier la multimodalité, les chercheurs utilisent communément les propriétés CASE et CARE définies par [Coutaz & Nigay 1994], [Coutaz & al. 1995], [Nigay & Coutaz 1997]. Du point de vue système, les propriétés CASE font référence aux multimodalités Concurrente, Alternée, Synergique et Exclusive, tandis que du point de vue utilisateur, les propriétés CARE désignent la Complémentarité, l'Assignation, la Redondance, et l'Equivalence [Caelen 2000].

L'un des intérêts des interfaces multimodales réside dans le fait d'employer des entrées dites « naturelles ». Cela procure un plus grand confort d'utilisation, mais produit inévitablement des ambiguïtés et des risques d'erreur d'interprétation de la part du système. Pour tenter de les limiter, des techniques de désambiguïsation mutuelle multimodales sont employées afin de conforter ou d'invalider un résultat obtenu de manière monomodale. Ce peut être une reconnaissance vocale couplée à une reconnaissance labiale, des mouvements oculaires accompagnés de commandes vocales [Yang & al. 1998], etc. Ainsi, les points névralgiques dans le domaine de la multimodalité sont la fusion et la fission de données.

Tout comme pour l'axe adaptation que nous avons présenté précédemment, cet axe de recherche à propos de la multimodalité requière un point de terminologie, afin de comprendre les différents positionnements des chercheurs du domaine.

3.1.1 Terminologie

L'IHM est une discipline au carrefour d'autres disciplines. C'est pourquoi les termes de média, multimédia, modes, modalités, multimodal, canal et multicanal revêtent des points de vue différents selon la discipline considérée.

Média :

Dans la vie courante, un média désigne un support d'information (journal, cassette vidéo, CD audio, DVD, etc.). Dans le domaine des IHM, un média désigne un support technique permettant une communication. Plusieurs points de vue sont étudiés :

– sous l'angle technique, un média est un dispositif physique, un procédé technique, servant de support à l'information. Ainsi, un clavier ou une souris sont considérés comme des capteurs, au sens où ils fournissent des informations, en provenance de l'utilisateur et à destination de la machine. Un écran ou une imprimante sont considérés comme des effecteurs, c'est-à-dire qu'ils diffusent de l'information, de l'ordinateur vers l'utilisateur.

– sous l'angle utilisateur, le média peut être perçu comme faisant référence aux éléments sensoriels, perceptuels et cognitifs de l'être humain. On parle alors de média visuel, auditif, audiovisuel, kinesthésique, etc.

Le plus souvent, les chercheurs utilisent cette notion de média, pour faire référence au matériel, support de la communication : « *Nous considérons le média comme un dispositif physique sur lequel plusieurs modalités peuvent coopérer. Ainsi, les modalités sonores verbales et non-verbales peuvent coopérer sur le média "haut-parleur".* » [Truillet 1999].

Multimédia :

La notion de multimédia dénote la diversité des supports techniques d'un système, par lesquels sont véhiculés des informations. Un système multimédia est, selon [Bellik 95], capable d'acquérir, de stocker et de restituer des informations de natures différentes.

Un système multimédia est communément identifié comme pouvant supporter plusieurs entrées et/ou plusieurs sorties ; l'élément discriminant un média d'un autre étant la nature de l'information. Un ordinateur multimédia désigne alors un système qui fusionne plusieurs types de composants tels que le texte, les images, les vidéos... Ce système traite les données au niveau élémentaire du signal, sans interprétation sémantique.

Enfin, une distinction peut être faite entre multi-média et multimédia, notamment dans le domaine de l'enseignement, où la première orthographe désigne l'utilisation de plusieurs médias sur plusieurs machines différentes, tandis que la seconde orthographe indique qu'une unique machine contrôle plusieurs médias, grâce à une numérisation des textes, des sons, des images, etc. [Bruillard 1997].

Mode :

Au sens de Frohlich, les modes d'une interface se définissent comme « *des états dans lesquels différentes actions de l'utilisateur peuvent avoir le même effet.* » [Frohlich 1991]. Du point de vue système, le mode évoque l'état dans lequel le système interactif se trouve à un instant donné, tandis que du point de vue usager, le mode traduit les états au travers desquels les différentes actions de l'utilisateur peuvent produire les mêmes effets. Côté système, l'exemple classique est celui de l'édition de texte : selon que l'on se trouve en mode « saisie » ou en mode « reffappe », l'entrée d'un nouvel élément viendra soit s'ajouter à l'existant, soit remplacer l'existant. Côté utilisateur, l'exemple classique est celui de la fermeture d'une fenêtre : l'utilisateur peut exécuter cette commande soit en cliquant avec la souris sur la croix représentant l'item « Fermer », soit à l'aide des touches ALT et F4 du clavier.

Le mode peut être vu comme la manière dont se fait une action. « *Chaque mode peut comprendre plusieurs modalités : par exemple, le mode auditif comprend les modalités sonores verbales et non verbales.* » [Truillet 1999]. Dans une communication, le mode se rapporte principalement à l'organe (ou au système d'organes) utilisé pour percevoir ou produire des informations. Un mode, dans le domaine de la téléphonie correspond aux normes utilisées par les réseaux. Typiquement, un téléphone devra être bimode pour fonctionner à la fois sur GSM³² et sur UMTS³³ [Weiss 2002]. En français, comme en anglais, le mot « mode » est parfois utilisé pour parler d'une « modalité » : « *Computer systems are being developed that support a variety of modes of interaction including speech, keyboard, handwriting, gesture, facial expression, context and affect.* » [Healey & al. 2002]. En IHM, on fait parfois la distinction entre une interaction en mode « direct » et « indirect ». Cela fait référence au degré d'engagement physique de l'utilisateur. Le premier mode implique que l'utilisateur effectue une manipulation physique directe avec le système tandis que le second mode dénote une interaction via un intermédiaire (interface langagière par exemple). On parle alors aussi de mode « action » et de mode « langage », et plus largement, on oppose ici un « faire » à un « faire-faire ».

Modalité :

Au sens large, une modalité est assimilée à une technique d'interaction, une manière d'interagir avec le système. Ce que l'on nomme une modalité est alors une manière d'utiliser un média. « *Par exemple le média stylo électronique peut être utilisé selon plusieurs modalités : écriture, gestes de commande (dans une application de dessin, faire une croix sur un objet pour supprimer cet objet), gestes de dessin.* » [Martin 2004]. Comme pour le média, la définition de la modalité dépend du point de vue (utilisateur vs système) que l'on adopte. Truillet par exemple, appréhende la modalité en se plaçant toujours du côté de l'être humain, et en se focalisant sur la structure des informations échangées entre les interacteurs. Il explique qu'une structure telle que le braille peut être perçue aussi bien par le toucher que par le visuel.

En IHM, la définition d'une modalité la plus communément admise est celle d'un couple <p,r> où : p désigne un dispositif physique (par exemple, une souris, une caméra, un écran, un haut-parleur), et r dénote un système représentationnel, c'est-à-dire un système conventionnel structuré de signes assurant une fonction de communication (par exemple, un langage pseudo naturel, un graphe, un référentiel de coordonnées GPS...). La notion de modalité active ou passive permet de déterminer si l'utilisateur doit effectuer une action explicite pour recevoir de l'information ou pas (cf. localisation d'un utilisateur par GPS sans intervention particulière de sa part).

Face à cette approche logicielle, orientée concepteurs, définissant avec quoi et comment se déroule l'interaction, grâce au couple <dispositif physique, langage d'interaction>, il existe l'approche des ergonomes et des psychologues qui associent les modalités aux capacités de perception et d'action des humains. L'utilisateur perçoit les informations grâce aux sens visuel, auditif, tactile, olfactif et gustatif, tandis qu'il en émet grâce à ses mouvements et à sa voix³⁴. On parle alors de modalité auditive, de modalité gestuelle [Baudel 1991] ou encore de modalité visio-gestuelle [RTP14 2004]. [Roudaut & Lecolinet 2007] proposent un espace de classification qui combine l'approche système et une l'utilisateur (cf. annexe 7). Nous adhérons à cette proposition et y ferons référence dans la suite de notre exposé.

³² GSM : Global System for Mobile communication : norme européenne pour la téléphonie mobile.

³³ UMTS : Universal Mobile Telecommunication System : norme européenne destinée à remplacer le GSM et permettant un accès haut débit aux services d'Internet mobile.

³⁴ Des travaux sur le BCI (Brain Computer Interface) montrent que l'on peut également interagir avec un ordinateur en émettant des informations via les ondes cérébrales du cerveau qui sont captées et analysées par le système informatique.

Multimodal :

La notion de multimodalité se démarque de celle de multimédia par le fait qu'elle sous-entend une interprétation ou une génération sémantique de l'information. En communication Homme-Machine, les chercheurs distinguent les termes désignant les interactions en précisant l'angle de vue choisi : centré utilisateur ou centré technologie.

Selon [Nigay & Coutaz 1993], la multimodalité est la capacité d'un système à communiquer avec un utilisateur en employant différents types de **canaux** de communication. Elles font une mise au point terminologique importante dans [Nigay & Coutaz 1996] et signalent déjà des ambiguïtés latentes autour de la notion de multimodalité : « *Enfin, l'adjectif modal relatif à la fois aux substantifs mode et modalité souligne l'ambiguïté du terme multimodal : lorsque nous caractérisons une interface de multimodale, s'agit-il de multi-modalité ou de multi-mode ?* ».

Le niveau physique du multimodal fait appel à la perception tandis que le niveau représentationnel fait appel à la cognition. Notons cependant que certains auteurs font référence au multimodal en utilisant le terme « multimédia intelligent » [Neal & Shapiro 1990].

Bouchet explique dans sa thèse, en s'inspirant du modèle de Gaines, que les interactions multimodales se trouvent actuellement au passage du deuxième au troisième niveau d'une évolution qui en compte six. Après la percée technologique qui a fait suite aux travaux de Bolt [Bolt 1980], nous vivons présentement le passage de la phase de réplication à la phase d'empirisme. Elles seront suivies des phases de théorie, d'automatisation et de maturité [Bouchet 2006]. Il explique également que la discipline manque d'outils pour concevoir et mettre en œuvre des systèmes à base d'IHM multimodales, même si certaines pistes commencent à émerger [Vanderdonck & al. 2005].

Canal :

Pour Frolich, un canal se définit « *comme une interface qui opère une transformation d'énergie* » [Frohlich 1991]. Côté utilisateur, il distingue les canaux « voix » et « mouvement » et côté système, il mentionne les canaux « audio », « visuel » et « haptique ». Pour Buxton, plusieurs canaux peuvent appartenir à une ou plusieurs modalités ; par exemple les deux oreilles sont deux canaux appartenant à la même modalité : l'ouïe. [Buxton 1990] (cité par [Nigay & Coutaz 1996]).

Dans le domaine du marketing direct, le canal est un système de distribution homogène. Les trois grands canaux de distribution sont :

- La vente directe : l'entreprise est en contact physique avec sa cible finale – entreprises ou consommateurs – et vend ses produits ou services sans aucun intermédiaire.
- La vente indirecte : l'entreprise commercialise son offre par l'intermédiaire de réseaux externes.
- La vente à distance : elle s'opère de façon indirecte, mais la cible finale est contactée par téléphone (télévente), courrier (mailing) ou Internet.

Par extension, on nomme « canal » un vecteur de communication entre l'utilisateur et l'organisation. On parle alors des canaux web, téléphone, fax, E-mail, SMS, télévision interactive, etc. [Claeyssen & al. 2006]. Nous adoptons cette définition et y ferons référence dans la suite de notre exposé.

Multicanal :

Le terme multicanal³⁵ (*multichannel*) est largement utilisé dans le domaine du marketing direct, en France comme à l'étranger : « *Multichannel marketing is marketing using many different marketing channels to reach a customer* » (Wikipedia). L'opposition entre multimédia et multimodal, liée principalement à la sémantique, que nous avons expliquée précédemment se retrouve aussi dans le marketing avec les termes multisupport et multicanal. Bertrand Frey, président de CRM Company Group³⁶ explique : « *Se contenter de décliner des messages sur différents supports ne correspond pas à une vraie communication multicanal. Décliner, c'est décliner.* » [Frey 2007]. Autrement dit, la réplication des messages sur plusieurs supports, numériques ou traditionnels, n'impliquent pas

³⁵ Le mot « multicanal » comporte aujourd'hui des acceptions multiples et il est important de préciser ce qu'il recouvre pour la Vente à Distance (VAD). Amplifier et améliorer la communication, voilà les deux raisons pour lesquelles une enseigne forte doit se préoccuper du « multicanal » : en termes d'infrastructures certes, mais surtout en termes d'approche client. Car avant même que les entreprises se disent « multicanal » c'est le consommateur lui-même qui est spontanément devenu « multicanal » (Yves Bayart, Direction du développement du groupe 3 Suisses, 29 septembre 2004).

³⁶ <http://www.crmcompanygroup.com/>

forcément une stratégie multicanal. C'est essentiellement la base de données et donc le recoupement d'informations qui sera le vecteur d'une réelle approche multicanal, dans le domaine du marketing direct. La Figure 20 illustre cette démarche avec l'outil e-RIS pour le Management de la Relation Client (ou CRM en anglais, pour Customer Relationship Management) et la gestion du multicanal proposé par l'entreprise HeLITIS.

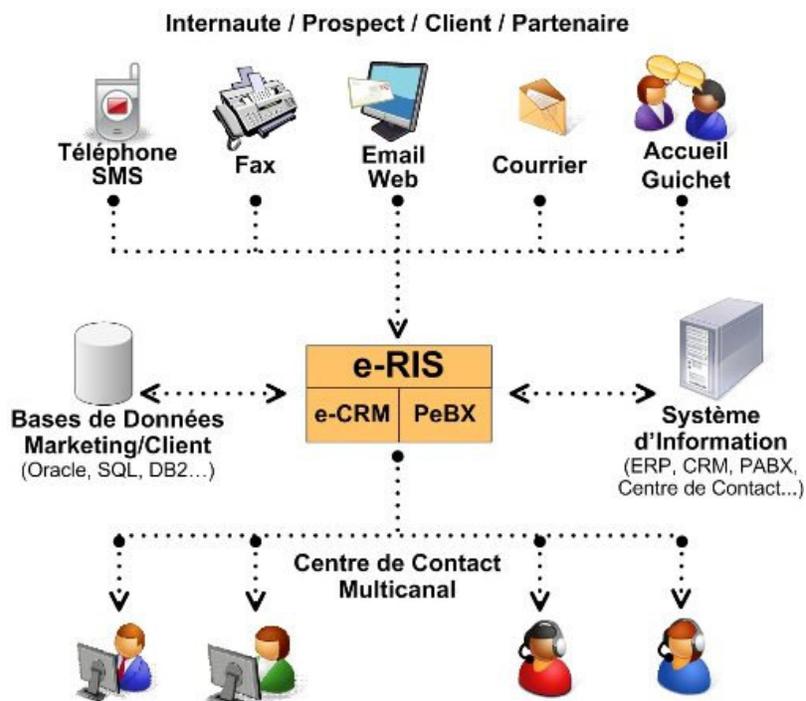


Figure 20 : La solution Multicanal e-RIS (Source <http://www.helitis.com>)

En IHM, il existe peu de références au multicanal à l'exception des travaux en français de [Chevrin 2006] et en anglais de [Healey & al. 2002] : « *Often these modalities require specialized channels to allow access modalities such as cameras, microphones and sensors. A multi-modal multi-channel system faces the challenge of accepting information from any input method and delivering information through the appropriate output methods.* ».

Quelques rares conférences ont lieu sur ce thème précis³⁷, mais le plus souvent, un amalgame est fait avec les notions assez proches de systèmes multi-plateforme ou multi-dispositif. En employant le mot plateforme, on fait généralement référence à une classe de dispositifs qui partagent les mêmes caractéristiques en termes de ressources pour l'interaction. « *They range from small devices, such as interactive watches to very large flat displays. Examples of platforms are the graphical desktop, PDAs, mobile phones and vocal systems. The modalities available have an important role in the definition of each platform.* » [Paternò 2004c].

Un système multi-dispositif quant à lui, permet d'utiliser plusieurs dispositifs (PC, PDA, Smartphone, Kiosque, etc.) pour effectuer le même type de tâche [Calvary & al. 2003]. Par exemple, dans le cadre de la plasticité des interfaces et du « context-aware computing », des travaux montrent des redistributions d'interfaces, de smartphone vers PC [Demumieux & al. 2008]. Cela permet au système de proposer à l'utilisateur, lorsque cela devient possible³⁸, une migration de son environnement de travail d'une plateforme mobile, mais restreinte, vers une plateforme fixe, mais plus puissante, avec ajout de fonctionnalités, la plupart du temps.

³⁷ Voir par exemple le workshop Multi-channel Adaptive Context-sensitive (MAC) Systems, organisé le 15 mai 2006 à l'université de Glasgow, sous l'égide du British HCI group : <http://www.dcs.gla.ac.uk/~rich/mac/>

³⁸ Via une connexion Bluetooth, par exemple.

Sur un même dispositif (*device*), il est possible d'utiliser plusieurs canaux. C'est le cas lorsqu'un utilisateur téléphone, puis surfe sur Internet, puis lit ses mails ou envoie et reçoit des SMS sur un même appareil, de type smartphone, par exemple. Comme toujours, la complexité des systèmes offerts est d'autant plus grande que l'on propose de possibilités de combinaison : «*Finally, multimodality can be combined with multi-platform systems (e.g. the user can switch between a desktop computer and a mobile phone) to offer even richer possibilities [Paternò 2004].*» [Alapetite 2007].

On note qu'ici le terme employé (*switch*) fait plus allusion à un changement de plateforme qu'à un usage multimodal combiné sur les deux plateformes disponibles simultanément. Paternò semble aller dans le même sens. Il explique comment, en fonction du contexte un utilisateur peut utiliser différents environnements de travail pour poursuivre une même tâche, de manière séquentielle. Je cite un large extrait de son article, pour ne pas dénaturer la pensée de l'auteur : «*In migratory interfaces the basic idea is that the interface, to some extent, follows the users moving from one device to another, adapting to the features of the new device without having to restart the session from scratch. This enables multimodal migrating services that, for example, allow the user to start the registration to the service through a desktop system. Then, in the event he realizes that it is getting late, he can carry on the registration on the move using a PDA until he reaches his car where he can use a vocal interface to complete it. All this activity can be performed without having to restart the registration procedure from the beginning at each device change.* » [Paternò 2004b].

Il me semble que la notion d'interface multicanal utilisée par certains chercheurs se rapproche de celle utilisée par d'autres pour faire référence à un système multimodal multi-dispositif. Les travaux de Schaefer et Mueller vont dans ce sens : «*All these renderers are working independently but the states are synchronized as soon as a command is passed to the backend application (the audio player). According to that, we can talk about a multimodal and multi-device remote control, although each modality is processed on a first come first serve basis.* » [Schaefer & Mueller 2008]. L'Open Mobile Alliance (OMA), qui regroupe des opérateurs de téléphonie mobile et des fournisseurs de matériel, de réseaux et de contenus, propose une recommandation³⁹ qui vise le même objectif [OMA 2008].

3.1.2 Interactions vocales

Parmi tous les moyens possibles d'interagir avec un ordinateur, celui qui me paraît le plus intéressant et qui aura donc retenu toute mon attention depuis le début de mes travaux de recherche et lors d'encadrements d'étudiants en informatique est l'interaction vocale.

Là encore, il faut bien avoir conscience que plus l'interaction sera naturelle entre un utilisateur final et un système informatique, plus les modèles utilisés et technologies employées par le concepteur seront complexes. La tâche du programmeur est encore plus difficile lorsqu'il s'agit de proposer à l'utilisateur de s'exprimer librement, sans contraintes, et d'utiliser plusieurs modalités d'interactions séquentielles (comme la voix puis un geste) ou combinées/synergiques (comme la voix et l'utilisation d'un ou de plusieurs gestes en même temps).

Dans la plupart des systèmes multimodaux que nous avons étudiés, la voix n'est utilisable qu'avec un vocabulaire limité et/ou avec un mécanisme d'activation de type « push to talk » : il faut appuyer sur un bouton pour déclencher la reconnaissance vocale, puis relâcher pour terminer. On est donc ici plus proche de la notion de « commandes vocales » que d'un véritable dialogue Homme-Machine en langue naturelle.

J'avais déjà mentionné dans mes perspectives de travaux de thèse que le W3C était sur le point de proposer un langage standard nommé VoiceXML, permettant à un utilisateur d'utiliser un téléphone pour dialoguer oralement avec une machine. J'ai donc étudié plus en détail les spécifications de ce langage et les différentes implémentations proposées.

³⁹ The Multimodal multi-device enabler provides an architecture for the management and synchronization of mixed modalities of interaction on mobile devices. It describes the different logical components required to support multimodal and multidevice interaction and the different deployment configurations that can be supported. (source : OMA Multimodal and Multi-device V1.0).

3.1.3 Le langage VoiceXML

Il existe sur la planète beaucoup plus de téléphones que d'ordinateurs. C'est à partir de cette constatation que les premiers projets de recherche d'accès à Internet par téléphone ont débuté dans les années quatre-vingt-dix. L'idée consiste à proposer l'accès au réseau Internet par le moyen d'un DHM en langue naturelle.

Cette tendance se confirme aujourd'hui avec l'essor considérable que connaît depuis quelques années la téléphonie mobile. Les efforts de standardisation vers un langage non propriétaire, permettant non seulement la gestion des aspects téléphoniques mais également, pour partie, du DHM (relance, aide, reformulation, sous-dialogue, etc.) ont amené à la spécification du langage VoiceXML (pour *Voice Extensible Markup Language*).

Le VoiceXML [VoiceXML 2.0] est donc un langage standard, à balises, permettant d'accéder à Internet grâce à un navigateur vocal via un téléphone fixe ou portable. Il est basé sur XML et s'articule autour de la reconnaissance et de la synthèse vocale, des grammaires de dialogue, et facilite la mise en œuvre de DHM sur serveurs vocaux interactifs.

A l'heure où nous écrivons ces lignes, le W3C annonce la version 3.0 de VoiceXML [VoiceXML 3.0]. Cette version était déjà annoncée depuis 2003 [Dettmer 2003] comme devant supporter les interactions multimodales. Dans cette proposition, des fonctionnalités sont notées comme obligatoires (*must have*) ; c'est le cas de la vérification du locuteur et de la gestion des événements externes⁴⁰, tandis que d'autres sont indiquées comme optionnelles (*nice to have*) ; c'est le cas pour la coordination simultanée d'entrées⁴¹. On notera au passage que le W3C a changé d'avis à propos de la gestion des événements, car en 2004, cette préoccupation n'était pas à l'ordre du jour⁴², alors que dans mes travaux de l'époque (2003), je montrais que c'était un point important à étudier et un verrou indéniable empêchant de faire converger les mondes graphiques et sonores, par exemple. Le W3C travaille cependant sur d'autres langages comme X+V⁴³ ou EMMA⁴⁴ [EMMA] qui partagent certaines balises avec le langage VoiceXML. Ces deux langages vont être succinctement présentés ci-après.

3.1.4 X+V

X+V [X+V] est un langage XML pour le développement d'applications multimodales. Il a été créé face aux demandes croissantes d'applications vocales pour les mobiles et les PDA. Contrairement au langage VoiceXML, X+V utilise à la fois le vocal et le visuel, ce qui offre de nouvelles possibilités de développement d'interfaces pour les périphériques mobiles.

La Figure 21 présente l'architecture du langage X+V. Ce dernier combine le langage XHTML [XHTML] et une sous-partie du langage VoiceXML. Le code des deux langages est séparé, ce qui simplifie le développement et permet de développer le visuel indépendamment du vocal. Rappelons que XHTML est un langage HTML réajusté pour respecter totalement les règles XML. Ensemble, ces deux langages permettent aux développeurs d'ajouter de la voix (en entrée et en sortie) aux traditionnelles pages Web. Cependant, à notre connaissance, il n'est pas possible d'utiliser ces deux modalités (voix et geste) conjointement sur deux dispositifs différents.

⁴⁰ External Event handling while a dialog is in progress (must have) [2.4.1 It MUST be possible for external entities to inject events into running dialogs. The dialog author MUST be able to control when such events are processed and what actions are taken when they are processed], [2.4.2 Among the possible results of processing such events MUST be pausing, resuming, and terminating the dialog. The VoiceXML 3.0 specification MAY define default handlers for certain such external events.], [2.4.3 It MUST be possible for running dialogs to send events into the Multimodal Interaction Framework.].

⁴¹ Co-ordinated, Simultaneous Multi-modal Input (nice to have): The markup language specifies that user input from multiple modalities is interpreted at the same time and that interpretations of the inputs are co-ordinated by the voice browser. For example, in a telephony environment, the user can type 200 on the keypad and say transfer to checking account and the interpretations are co-ordinated so that they are understood as transfer 200 to checking account.

⁴² Question: Events? Answer: Not a W3C topic, probably, more a mobile issue. Current W3C DOM is not enough, it needs mobile extensions. This is probably work for OMA. (source : minutes de la réunion du MMI workshop, 19 July 2004 : <http://www.w3.org/2004/02/mmi-workshop/day-1-am.html>)

⁴³ X+V : XHTML+Voice Profile

⁴⁴ EMMA : Extensible MultiModal Annotation markup language

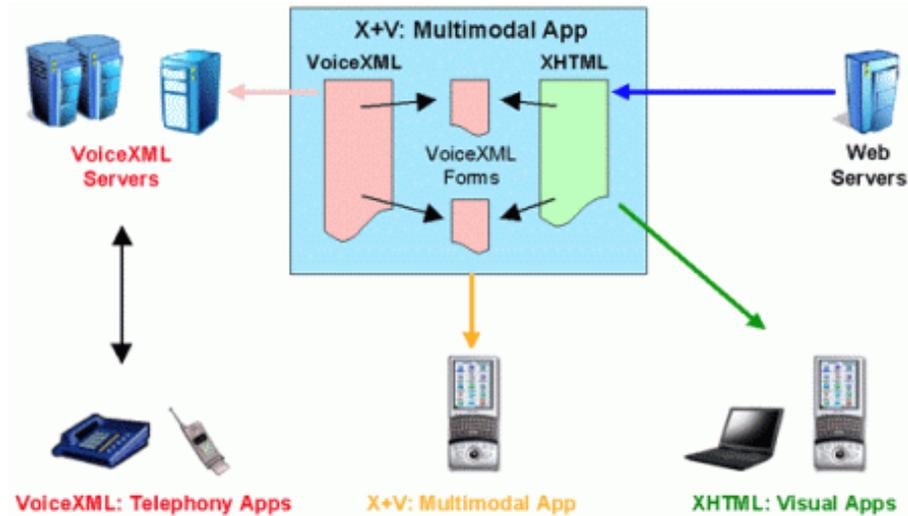


Figure 21 : Architecture du langage X+V (Source : W3C [X+V 2004])

Alexandre Alapetite a montré qu'il est néanmoins possible, grâce à X+V d'implémenter une application multimodale de type complémentaire⁴⁵, sur un même navigateur multimodal supportant X+V et SVG⁴⁶, comme Opera [Alapetite 2007].

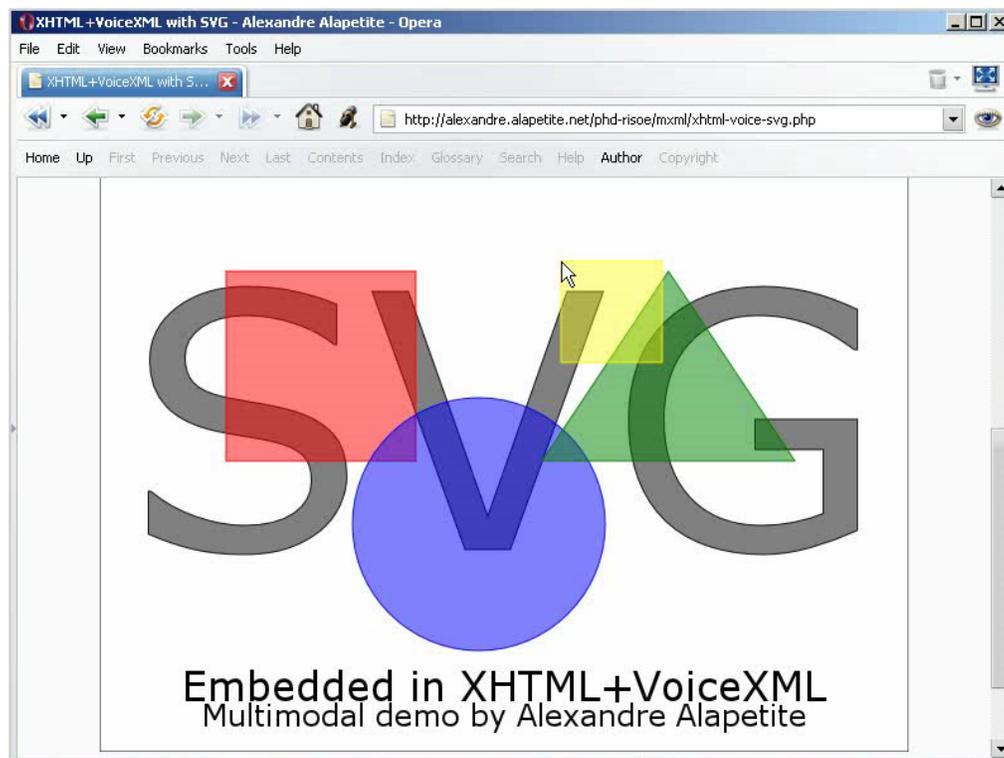


Figure 22 : Un exemple d'utilisation de X+V et de SVG pour une multimodalité complémentaire

La Figure 22 est une capture d'écran effectuée juste après que l'utilisateur ait prononcé la phrase « *I want a new yellow square here* » et cliqué sur une région de la page web. Le résultat est l'affichage d'un carré jaune à l'endroit indiqué par l'utilisateur.

⁴⁵ This minimal "put that here" demo illustrates the combined use of voice and mouse input. (<http://alexandre.alapetite.net/phd-risoe/mxml/xhtml-voice-svg.php>)

⁴⁶ SVG (pour Scalable Vector Graphics) est un format de données spécifié par le World Wide Web Consortium et conçu pour décrire des ensembles de graphiques vectoriels et basé sur XML.

3.1.5 EMMA

EMMA (EMMA: Extensible MultiModal Annotation markup language) [EMMA] est un langage qui permet une représentation sémantique de toutes les informations récupérées en entrée pour les intégrer au sein d'une application multimodale. Dans une architecture multimodale, chacune des entrées est reconnue et interprétée par des modalités spécifiques (reconnaissance vocale, gestuelle, de l'écriture, clavier, souris, ...). Toutes ces informations récupérées, qui sont propres à chacune des modalités, sont ensuite représentées dans un langage commun qui est EMMA. Selon le W3C, EMMA pourrait devenir un langage commun entre les différents composants d'un système multimodal. Ceci permettrait aux développeurs de créer des plates-formes multimodales plus fiables et moins chères pour l'intégration de chaque type de composants, mais à l'heure actuelle aucune implémentation de ses spécifications n'est disponible.

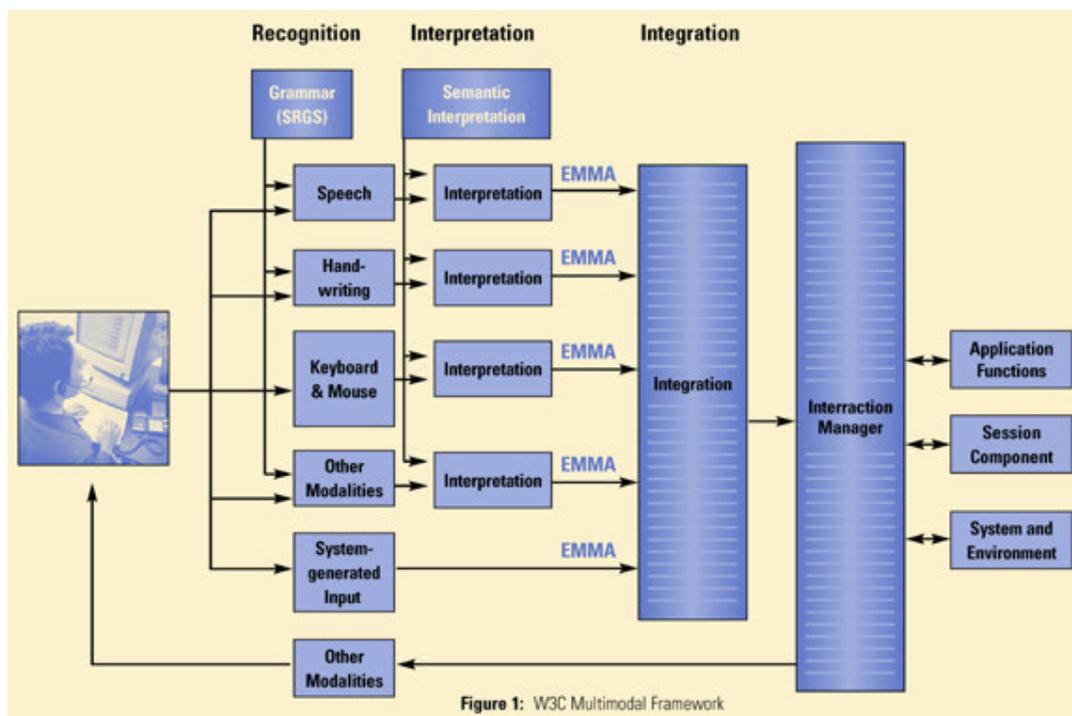


Figure 23 : Framework multimodal proposé par le W3C intégrant le langage EMMA (Source [EMMA2])

La Figure 23 présente l'architecture du framework multimodal proposé par le W3C qui intègre l'utilisation du langage EMMA. Les étapes principales sont la reconnaissance d'une interaction, puis son interprétation, et enfin son intégration. Afin d'illustrer l'utilisation de ce langage, imaginons qu'un utilisateur pointe son stylet ou son stylo électronique sur une carte affichée sur un écran, et prononce la phrase « zoom avant, ici ». Les interactions vocales et gestuelles seront estampillées puis, par exemple, la notation EMMA suivante sera générée :

```
<command>
<action>zoom avant</action>
<location>
<point> 42, 158 </point>
</location>
</command>
```

Cette commande EMMA sera ensuite transférée au manager d'interaction chargé de déterminer la ou les actions à effectuer. Dans notre exemple, il pourrait s'agir de l'envoi d'une commande appropriée au navigateur web puis la génération d'une synthèse vocale annonçant la bonne exécution de la commande de l'utilisateur.

Ce langage est en cours de finalisation et devrait encore évoluer. Tout récemment (octobre 2008), des premiers retours d'implémentation (Implementation Report for EMMA 1.0 Specification) ont été diffusés par quelques industriels comme Loquendo, AT&T, Nuance, Microsoft ou Conversational Technologies et des universitaires, comme l'université de Trento, en Italie ou DFKI (German Research Center for Artificial Intelligence) en Allemagne.

Même si cette approche paraît séduisante, sur le papier, les spécifications du W3C ne décrivent pas la manière de réaliser les implémentations du langage EMMA. Des verrous que nous avons identifiés dans nos travaux demeurent, à notre connaissance, sans réponse, à l'heure actuelle. Il s'agit par exemple des mécanismes de notification d'événements : comment le navigateur web va être notifié qu'il doit effectuer un zoom avant, dans notre exemple ? Il s'agit aussi de la qualité de l'estampillage durant la phase de reconnaissance vocale : Loquendo rapporte que cela est difficile à réaliser avec précision, actuellement. Enfin, de ce que nous en avons compris, rien n'est prévu dans l'architecture proposée par le W3C pour supporter l'interaction dans des systèmes à la fois multimodaux, multidispositifs et multicanaux.

Le Tableau 1 résume les principales caractéristiques de quelques⁴⁷ langages standards permettant de gérer des interactions vocales.

Langages	Type de multimodalité	Atouts	Faiblesses
VoiceXML 2.1 (version actuelle)	Alternée (Voix, puis DTMF, par ex.)	- Gestion d'un dialogue oral H/M sur téléphone - Traces de reconnaissances vocales exploitables - Basé sur XML	- Mono-Dispositif - Trop orienté « formulaire » (fill in the form approach) - Grammaires vocales limitées
VoiceXML 3.0 (future version)	Alternée (Voix puis DTMF, par ex.)	- Reconnaissance du locuteur - Gestion des événements externes - Basé sur XML	- Grammaires vocales limitées
X+V	Alternée (clavier/souris puis Voix, par ex.)	- Basé sur XHTML et VoiceXML	- Mono-Dispositif - Grammaires vocales limitées
EMMA	Synergique (Voix et stylet simultanément, par ex.)	- Principal élément du « W3C Multimodal Interaction Framework » - Basé sur XML	- Mono-Dispositif - Plus orientée vers les entrées que vers les sorties
SMIL 2.1 (Synchronized Multimedia Integration Language)	Multimodalité en sortie. Description temporelle et spatiale des différents composants	- Intégration de contenus multimédias diversifiés (images, sons, textes, vidéo, animations, flux de texte). - Synchronisation permettant la création de présentations multimédias. - Basé sur XML	- Mono-Dispositif - Orientée vers les sorties. Ne gère pas les entrées
SALT (Speech Application Language Tags)	Alternée (Voix, puis DTMF, par ex.)	- Extension HTML avec inclusion parole et téléphonie (DTMF) - Basé sur XML	- Mono-Dispositif - Ne gère pas le dialogue

Tableau 1 : Comparaison des langages du W3C pour les interactions vocales

On remarque que le principal atout commun à tous ces langages est le socle qu'ils partagent (XML) et que la principale faiblesse de ces langages est d'être monodispositifs et trop orientés Web, originellement graphique.

⁴⁷ D'autres langages proposés par le W3C sont utilisés par certains langages décrits dans le Tableau 1. Il s'agit par exemple de SSML (Speech Synthesis Markup Language), de SRGS (Speech Recognition Grammar Specification), ou encore de SISR (Semantic Interpretation for Speech Recognition).

3.2 Contribution à cet axe de recherche

J'ai contribué à cet axe de recherche en prenant en compte les interactions multimodales et multicanal à base d'entrées dites naturelles. Je me suis focalisé sur l'étude des interactions multimodales comportant au moins une dimension vocale (multimodalité vocale et gestuelle, par exemple). J'ai travaillé sur le langage VoiceXML en montrant ses limites (grammaires vocales limitées et traces d'usage peu exploitables) et en proposant des pistes de recherche pour y remédier.

3.2.1 Référentiel Multi-DMC

Il n'est pas toujours facile de distinguer les systèmes en termes de modalités, de canaux utilisés ou de fonctionnalités exploitables sur telle ou telle plateforme. Le référentiel Multi-DMC que je propose permet de positionner un système sur un repère en fonction de trois critères : Dispositif (D), Modal (M) et Canal (C). Il comporte deux positions (Mono ou Multi) pour chacun des trois critères ciblés (DMC). Cela représente donc 2^3 possibilités de positionnement. Ces 8 possibilités sont décrites grâce à la Figure 24 qui présente le référentiel Multi-DMC.

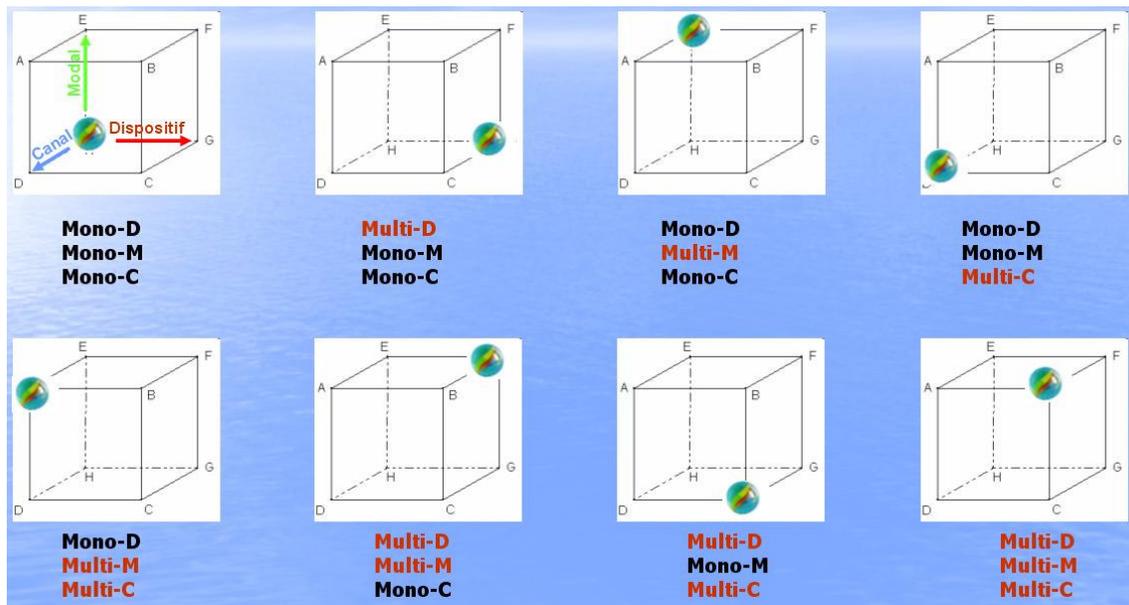


Figure 24 : Référentiel Multi-DMC

Dans ce référentiel, le dispositif, le modal et le canal font respectivement référence au support matériel employé par l'utilisateur, aux modalités (au sens de [Roudaut & Lecolinet 2007]) utilisées durant l'interaction et au support logique de l'échange d'informations entre l'homme et la machine. Précisons que le canal, pour nous, s'entend côté usage et non côté technique. Autrement dit, nous ne considérons pas le support physique permettant le transport du message, mais bien la nature (le type) des messages échangés. Par exemple, le canal SMS et le canal téléphonique sont bien distincts pour nous, alors qu'ils utilisent (en téléphonie mobile) le même réseau de communication.

La position (0 0 0) décrit un système mono-dispositif, mono-modal et mono-canal. Un système fonctionnant sur un unique dispositif (PC, par exemple), avec une seule modalité d'interaction possible (geste, par exemple) et pour un unique canal de communication (le web, par exemple) est de ce type là. La position (1 1 1) décrit un système multi-dispositif, multi-modal et multi-canal. Un système qui est capable de fonctionner sur au moins deux dispositifs (PC et smartphone, par exemple), avec au minimum deux modalités d'interaction possibles (voix et geste, par exemple) et pour au moins deux canaux de communication différents (le web et le téléphone, par exemple) serait de ce type là.

La position (0 0 1) caractérise un système mono-dispositif, mono-modal mais multi-canal. Ce pourrait être l'exemple d'un système fonctionnant uniquement sur smartphone, uniquement de manière auditive (en entrée comme en sortie), mais utilisable aussi bien via le canal web du smartphone que via son canal téléphonique. Ce référentiel n'est, bien entendu, pas complet (on aurait pu y rajouter le critère mono/multi-utilisateur, etc.) mais il permet de mieux préciser les spécificités d'un système et de comparer des systèmes entre eux.

L'exemple que je présente ci-après est un système de type (1 1 1), dans ce référentiel Multi-DMC. La Figure 25 présente, à gauche, une capture d'écran d'une application fonctionnant sur smartphone HTC Kaiser, sous Windows Mobile 6 et, à droite, une capture d'écran d'une application PC, permettant de visualiser une salle, grâce à une webcam installée dans le bâtiment B6 de notre université. Ces applications, codés en C#, permettent à un utilisateur de prendre le contrôle d'une camera de manière distante [Rouillard & Perrault 2008].



Figure 25 : Pilotage d'une caméra à distance, grâce au délégué pour IVY

L'application PC est connectée au bus IVY [IVY] local et écoute les messages lui demandant de déplacer une camera (action=move objet=cam_B6 dans notre exemple). En utilisant l'application sur smartphone, l'utilisateur envoie des données sur le bus IVY, via un délégué, qui agit à sa place, sans limite de distance. On voit dans notre exemple que l'utilisateur a sollicité un déplacement de la caméra vers le haut (complement=up). Cette interaction a été demandée de manière textuelle (cf. textbox au milieu de l'écran du smartphone). Cela implique que l'utilisateur sache quelles commandes utiliser (mot-clef=valeur). De manière équivalente, des boutons directionnels pourraient tout aussi bien être utilisés, sur l'interface du smartphone, à la place des lignes de commandes envoyées au délégué.

Nous avons également validé ce prototype en utilisant un autre canal de communication : l'accès téléphonique. L'utilisateur peut alors prononcer des phrases comme « déplace camera 1 vers la droite » pour obtenir le même résultat que précédemment. Enfin, de manière multimodale, il peut prononcer un ordre incomplet oralement (comme « déplace camera 2 »), puis parachever sa commande en pressant une touche de son téléphone fixe ou mobile. Le résultat de sa commande complète se voit alors sur l'application PC, qui aura commandé le déplacement de la webcam dans le sens désiré. L'utilisateur peut choisir de compléter les ordres donnés soit via le même dispositif, soit via plusieurs dispositifs utilisés conjointement. Une autre application multimodale, multicanal et multi-dispositif a été réalisée, sur la base de la même architecture, pendant le stage de master recherche de Simon Perrault [Perrault 2008] que j'ai encadré. Elle est présentée brièvement ci-après.

La Figure 26 montre un utilisateur disposant de plusieurs périphériques et de plusieurs interfaces différentes pour exécuter une tâche de dessin. Il dispose d'un smartphone HTC connecté en Wifi (dans sa main), d'un écran tactile (à sa gauche), d'un grand écran de visualisation (écran du fond, sur la photo), d'un système de reconnaissance vocale (il porte un micro-casque et la machine face à lui trace (log) les interactions vocales programmées en VoiceXML), et d'un outil de contrôle des opérations (PC portable à sa gauche).



Figure 26 : Utilisation de différentes interfaces pour une tâche de dessin (Painter)

Ce système, de type (1 1 1) dans le référentiel Multi-DMC permet de tracer des formes géométriques simples, en utilisant une ou plusieurs modalités d'interactions, via un ou plusieurs dispositifs disponibles. L'utilisateur peut choisir de n'utiliser que la voix, ou que le smartphone, par exemple, ou bien une combinaison des éléments disponibles.

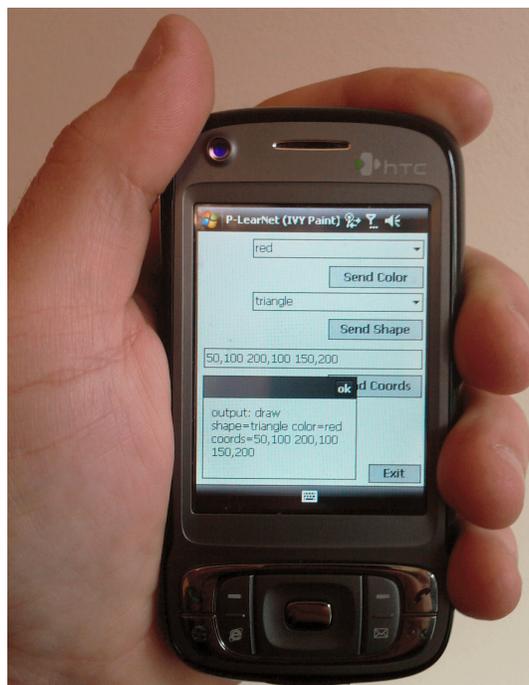


Figure 27 : Utilisation d'un smartphone HTC pour communiquer avec le bus IVY, via le délégué

Par exemple, en pointant sur l'écran tactile deux coordonnées (x1,y1 et x2,y2), puis en demandant vocalement « dessine cercle » ou « dessine rectangle », il fera afficher la forme désirée sur le grand écran. On dispose ainsi d'une application multimodale (manipulation directe, voix, geste...), multicanal (canal téléphonique, canal web) et multi-dispositif (écran tactile, PC, smartphone). Elle peut être utilisée par une seule personne qui dispose d'une multitude de périphériques ou bien par plusieurs utilisateurs qui coopèrent.

L'intergiciel déjà existant (le bus IVY [Buisson & al. 2002]) a été adapté à nos besoins spécifiques pour répondre à des contraintes d'utilisation distante. Nous avons conçu et réalisé un délégué permettant d'agir à distance en injectant des informations sur le bus, utilisable notamment en situation de mobilité. La Figure 27 montre une utilisation sur un smartphone HTC du délégué que nous avons développé à l'occasion de ces travaux. Lorsque l'utilisateur envoie un couleur, une forme ou des coordonnées, il expédie en réalité ses informations au délégué se trouvant sur le réseau associé au bus IVY visé, qui transmet, en son nom, via un web service, ces informations aux autres agents connectés localement.

La Figure 28, quant à elle, présente un des agents que nous avons développé, et qui permet à un humain de prendre connaissance des différents messages qui circulent entre les agents, sur un bus de communication spécifié.

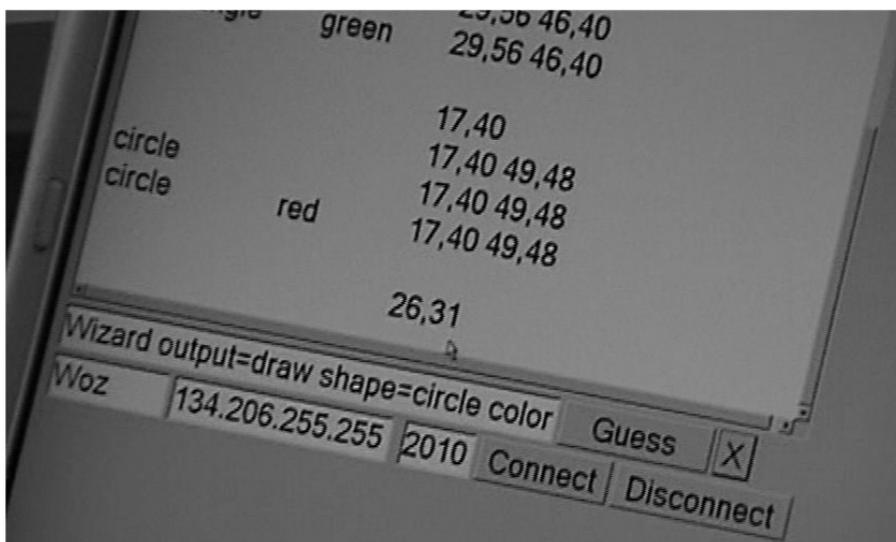


Figure 28 : Interface du Magicien d'Oz

A terme, cet agent devra pouvoir comprendre les intentions des utilisateurs, mais pour l'heure, cet outil fait office de Magicien d'Oz [Kelley 1985]. Plus exactement, il sert d'interface de contrôle pour le compère humain chargé de simuler l'intelligence du système lorsque les ordres collectés paraissent ambigus.

3.2.2 Multimodalité vocale et graphique

J'ai travaillé sur le couplage des aspects graphiques et vocaux au sein d'une même application multimodale dès 2001 [Rouillard 2002], soit plus de trois ans avant que la première spécification du langage X+V soit proposée par le W3C. J'ai montré qu'il était possible de coupler les langages HTML et VoiceXML dans une application de commerce électronique. L'utilisateur pouvait visualiser dans un navigateur web traditionnel les résultats des requêtes vocales effectuées par téléphone. Le dialogue oral téléphonique incorporait alors dynamiquement des informations provenant des manipulations directes effectuées sur la page web. Grâce à une base de données, il était possible de personnaliser les informations et de suggérer des propositions à l'utilisateur. Certes, le mécanisme permettant de synchroniser les deux canaux (web et téléphone) était rudimentaire, puisqu'il s'agissait d'une attente⁴⁸, mais ces premiers travaux ont ouvert des pistes de réflexion pour le couplage du vocal avec d'autres canaux et d'autres modalités. Quelques unes d'entre elles sont évoquées ci-après.

⁴⁸ Plus précisément, un rafraîchissement de la page web dynamique écrite en ASP était effectué toutes les 5 secondes

3.2.3 Multimodalité vocale et SMS

Des travaux en collaboration avec des industriels⁴⁹ m'ont permis de travailler sur la convergence des canaux de communication. Les cibles choisies étaient le téléphone d'une part et le SMS d'autre part. Ils n'utilisent pas les mêmes langages et protocoles de communication, pourtant, nous avons réussi à utiliser un même code écrit en VoiceXML, pour qu'il puisse être compris et utilisé sur deux canaux hétérogènes : le téléphone classique et le SMS. La Figure 29 montre le prototype réalisé, qui permet, en plus d'être utilisable via un serveur vocal interactif supportant le langage VoiceXML, de simuler l'envoi et la réception de flots de messages SMS+ suivant l'algorithme FAI (*Form Interpretation Algorithm*) utilisé par les moteurs d'interprétation de VoiceXML [Rouillard & Dutter 2005].

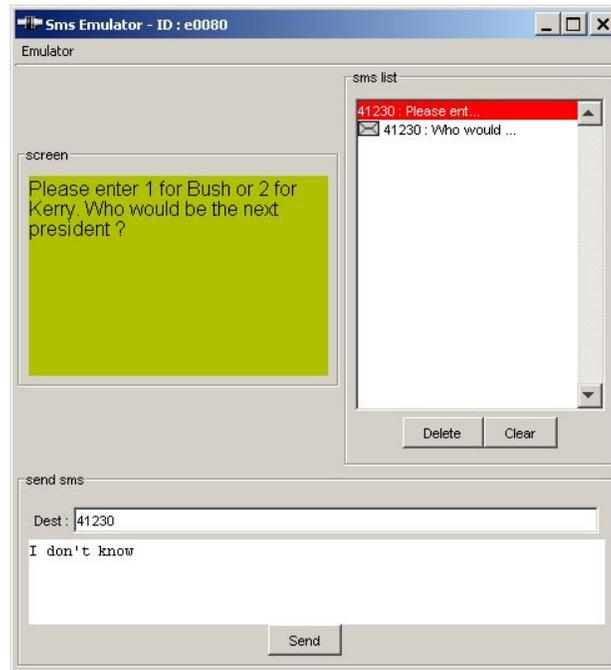


Figure 29 : Emulateur So.Phone SMS

Ces travaux nous ont donc permis d'œuvrer pour la convergence et la réutilisabilité de code sources des applications multimodales et multicanal.

3.2.4 Multimodalité vocale et gestuelle

D'autres travaux dans le domaine de la multimodalité m'ont amenés à utiliser le vocal (en entrée et en sortie) avec d'autres moyens d'interactions comme la reconnaissance de geste. J'ai travaillé vers trois axes : la reconnaissance de geste instrumentée, la reconnaissance d'écriture manuscrite et la reconnaissance de geste par camera.

3.2.4.1 Reconnaissance de geste instrumentée

La Figure 30 montre les résultats de travaux menés par des étudiants en informatique que j'ai encadrés lors d'un projet de 2008. Ils devaient mettre en œuvre un système multimodal autorisant, en entrée une reconnaissance de gestes via les manettes de jeu de la console Wii (Wii mote et Nunchuk) sur PC.

⁴⁹ La société Converge Online en premier lieu, puis d'autres entreprises comme Elan Informatique ou AppLine, toutes spécialisées dans le domaine du vocal.



Figure 30 : L'utilisateur enchaîne des mouvements afin que le système déclenche l'action correspondante à la phrase « Descendre le store de la salle de bain »

Les sorties possibles étaient l'affichage graphique, la synthèse vocale, et le contrôle de phidgets⁵⁰, popularisés notamment par Saul Greenberg [Greenberg & Fitchett 2001]. Comme le montre la Figure 30, l'utilisateur peut enchaîner des mouvements afin de créer des formes ou des lettres, qui assemblées, forment une « phrase » compréhensible par le système. Dans cet exemple précis, l'enchaînement des symboles « flèche vers le bas » + « rectangle comportant deux droites horizontales en son centre » + la lettre « S » a été compris comme l'équivalent de la commande « Descendre le store de la salle de bain ». En réaction, le système a exécuté la tâche demandée, via une animation Flash, et a commandé l'éclairage d'une led rouge (insérée dans un carton, sur la table, dans la 4^{ème} photo).

3.2.4.2 Reconnaissance d'écriture manuscrite

Après avoir discuté avec Max Froumentin du W3C, j'ai voulu mené des travaux sur l'utilisation du langage InkML [InkML] pour la reconnaissance d'écriture manuscrite dans un système multimodal d'enseignement à distance (voir annexe 8). J'ai encadré des étudiants de master professionnel qui ont réalisé un logiciel permettant d'interagir de manière multimodale avec un système d'apprentissage. En sortie, des questions étaient posées à l'utilisateur de manière multimodale (affichage à l'écran et synthèse vocale), et en entrée, l'utilisateur pouvait choisir de répondre soit à l'oral, soit au clavier, soit à l'écrit, c'est-à-dire en écrivant sa réponse de manière manuscrite, avec une souris ou un stylet, dans la zone prévue à cet effet. Cette écriture cursive était alors numérisée puis comparée à la réponse attendue. Les résultats de ces travaux ont montré une bonne utilisabilité du logiciel, pour cette tâche, tant à propos du taux de reconnaissance de l'écriture manuscrite que du laps de temps nécessaire pour obtenir une réponse.

⁵⁰ <http://www.phidgets.com/>

3.2.4.3 Reconnaissance de geste par camera

Dans le cadre d'une coopération avec des chercheurs du CRIL (Centre de Recherche en Informatique de Lens) travaillant sur la recherche d'information et le datamining, nous avons expérimenté le couplage de la voix avec une reconnaissance de geste par camera [Couturier & al. 2008].



Figure 31 : Reconnaissance de geste via une webcam couplé à une reconnaissance vocale

La Figure 31 illustre une situation dans laquelle l'utilisateur, tout en prononçant des commandes vocales, peut gesticuler devant une camera. Il peut pointer du doigt ou bouger la tête devant les pastilles situées aux quatre points cardinaux. Après un certain laps de temps (paramétrable), pendant lequel l'utilisateur reste sur la cible, la commande associée est envoyée au système. L'utilisateur peut également se servir de la souris pour combiner des interactions. La camera sert donc à détecter des gestes utiles pour naviguer au sein de l'espace de travail (avec une vue en œil de poisson), tandis que la voix et/ou la souris servent à explorer en profondeur un axe (zoom avant ou arrière piloté par commande vocale ou molette de la souris).

En résumé, ces travaux autour de la modalité gestuelle restent en marge de mon axe principal de recherche, plus orienté sur le vocal, mais ils ont permis néanmoins, de faire apparaître, outre les problèmes de mise en œuvre de telles applications, des questions importantes à propos de la sémantique. La section suivante relate des travaux plus importants, sur lesquels j'ai longuement cogité. Il s'agit d'une réflexion en profondeur sur les moyens d'étendre les possibilités d'interaction, dans le cadre de l'utilisation du langage VoiceXML.

3.2.5 VoiceXML augmenté

J'ai étudié le langage VoiceXML en détail et je l'enseigne aux étudiants de niveau master. Par ailleurs, j'ai publié le seul livre francophone actuellement disponible dans ce domaine [Rouillard 2004]. J'ai pu donc réellement mettre en œuvre des systèmes multimodaux basés sur VoiceXML et me rendre compte des possibilités et des limites de ce langage. L'un des avantages du VoiceXML réside dans le fait que la communication entre l'homme et la machine s'établit selon le modèle d'une réelle conversation, avec ouverture et clôture de dialogue, relances et reformulations, etc. Techniquement, les balises `<noinput>` et `<nomatch>` sont utilisées pour gérer les erreurs ou incompréhensions. Dans le premier cas, aucune entrée audio n'aura été détectée et dans le second, la phrase prononcée par l'utilisateur ne correspond à aucun élément de la grammaire active.

Un verrou scientifique sur lequel j'ai travaillé consistait à chercher une solution permettant de conserver la souplesse d'un dialogue en langue naturelle tout en essayant d'accroître les possibilités d'expression de la part de l'utilisateur. La tâche n'est pas triviale et quelques projets poursuivent le même objectif (FreeSpeech Project⁵¹, EvalVoiceXML project : Evaluation of Intelligent Component Technologies for VoiceXML Applications, [Mittendorfer & al. 2001]). En fait, depuis la première spécification du langage VoiceXML, le concept de transcription de l'audio vers le texte est évoqué. En effet, ce standard propose en annexe une balise `<transcribe>`, mais malheureusement, aucune implémentation ne respecte cette recommandation. La solution que j'ai proposée repose sur l'utilisation d'une architecture orientée service (SOA). L'idée consiste à conserver un dialogue classique en VoiceXML et de laisser l'utilisateur s'exprimer librement en langue naturelle, lorsque cela est nécessaire (et possible). On travaille alors hors grammaire vocale. On fait appel à un service externe de transcription (voix vers texte), qui pourra fournir un résultat réutilisable au sein du dialogue

⁵¹ <http://www.mperfect.net/freeSpeech>

en cours. Des outils de reconnaissance vocale comme Nuance/Dragon⁵² ou Telispeech⁵³ peuvent aider à atteindre cet objectif. De plus, le résultat de la transcription obtenu peut éventuellement servir d'entrée pour l'appel à un autre service (de traduction d'une langue à une autre par exemple).

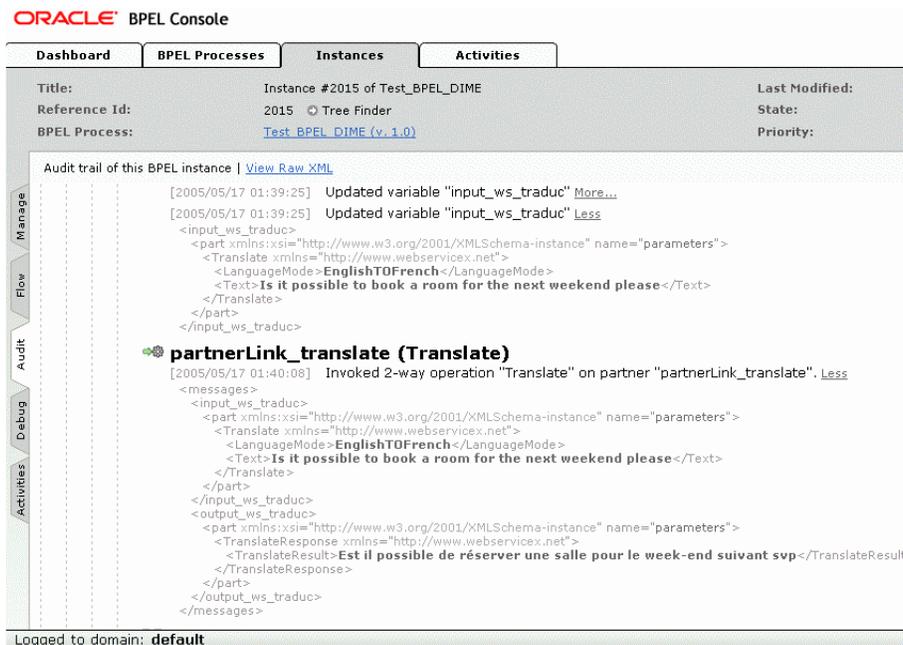


Figure 32 : A partir d'un énoncé oral en anglais, transcrit en texte, une traduction de cette phrase en français est utilisable

Un exemple d'implémentation utilisant le langage BPEL (Business Process Execution Language) d'une transcription (voix vers texte), enchaînée d'une traduction (anglais vers français) est montré sur la Figure 32. Les détails de ces travaux sont donnés dans [Rouillard 2006].

Cette approche a été accueillie favorablement par la communauté, comme en témoigne l'article intitulé « VoiceXML on Steroids » de James A. Larson (chairman of the W3C Voice Browser Working Group), dans lequel il explique : « *Researchers at the Université des Sciences et Technologies de Lille have implemented a <transcribe> tag within VoiceXML that converts speech to text using a dictation recognition engine rather than the conversational speech engines traditionally used in VoiceXML applications. The <transcribe> tag recognizes free form text without a developer-specified grammar. This tag could be very useful in several situations, for example to allow the system to express some utterances not modeled by the application, but pronounced by users and then provide feedback to users by re-using their input.* » [Larson 2005]. Cette manière de procéder est maintenant utilisée pour l'accès à des dictionnaires de manière vocale [Lerlerdthaiyanupap 2008], car bien évidemment, la grammaire nécessaire pour une telle application ne peut pas couvrir toutes les entrées possibles d'un dictionnaire.

Ces travaux rejoignent d'autres préoccupations que nous avons à propos de la réutilisabilité des services proposés aux utilisateurs. On utilise ainsi tout le potentiel des architectures dites SOA (pour Service-Oriented Architecture). Cela ouvre de nombreuses pistes de recherche que nous voulons explorer notamment pour des interfaces embarquées (dans des téléphones ou des smartphones, par exemple) et dans des situations réelles d'utilisation, hors laboratoire. Un ouvrage que j'ai dirigé sur ce sujet (SOA) a été publié en septembre 2007 [Rouillard & al. 2007]. Il a été sélectionné pour la 22^{ème} édition du prix Roberval⁵⁴ et a terminé finaliste pour le prix AFISI⁵⁵ du « meilleur livre informatique de l'année 2008 ».

⁵² Dragon NaturallySpeaking™ Preferred, <http://www.scansoft.com/naturallyspeaking>

⁵³ Telispeech 1.2 de l'entreprise Telisma, <http://www.telisma.com/>

⁵⁴ Le Prix Roberval est un concours international francophone qui récompense des œuvres littéraires, audiovisuelles ou multimédias consacrées à l'explication de la technologie : <http://prixroberval.utc.fr/>

⁵⁵ Association Française d'Ingénierie des Systèmes d'Information

3.2.6 Des services web pour la multimodalité

En poursuivant notre démarche consistant à employer des services dans nos applications, j'ai développé des services web de synthèse et de reconnaissance vocale, codés en C#, qui fonctionnent sous .NET. La composition et l'orchestration de services permettent de contourner, en partie, les problèmes que soulèvent les systèmes monolithiques, qui fonctionnent comme une boîte noire, dont on ne peut pas contrôler finement les mécanismes. L'assemblage de briques logicielles existantes est possible grâce à des langages standards de composition, comme BPEL, par exemple. Cependant, certaines ressources n'existent pas actuellement sous forme de service web. C'est le cas notamment des synthèses et des reconnaissances vocales ou des processus de détection de codes barres à partir d'une image. C'est pourquoi j'ai conçu et réalisé des services web, accessibles sur le réseau Internet et utilisables via le protocole SOAP (Simple Object Access Protocol).

Dans le cadre du projet P-Learnnet, par exemple, nous mettons à dispositions de nos partenaires académiques et industriels des services web (TTS, ASR, codes barre 1D ou 2D) permettant l'intégration facile et robuste de ces ressources dans leurs applications. Ils sont actuellement utilisés dans des applications de coaching (sport, mécanique, cuisine, etc), d'aide à la vente, d'apprentissage, etc. Cela peut paraître trivial, mais encore une fois, il n'en est rien, car, sans entrer dans les détails, il existe plusieurs manières de faire transiter du binaire (image, fichier sonores, vidéos, etc.) sur le réseau Internet, et les outils disponibles ne sont pas encore stabilisés du point de vue des normes et des protocoles (voir DIME : Direct Internet Message Encapsulation, WSE 3.0 : Web Services Enhancements de Microsoft .NET, WS-Attachments ...). L'interopérabilité n'est donc pas toujours garantie.

- La synthèse vocale est l'un des éléments le plus fréquemment utilisé dans le cadre d'applications multimodales. Nous disposons au laboratoire LIFL d'un serveur vocal SIBILO de la société APP-LINE, fournissant une synthèse vocale de bonne qualité (Loquendo). De façon traditionnelle, il n'est pas possible d'utiliser cette synthèse vocale (dite Text-To-Speech ou TTS) depuis l'extérieur du serveur vocal (sauf si l'on dispose d'un client MRCP⁵⁶). Nous avons donc développé un service web en langage C# permettant d'obtenir dans un fichier audio (format .wav) une synthèse vocale, en fournissant à ce service deux paramètres : (a) la phrase à synthétiser et (b) le nom de la voix choisie. Les voix actuellement disponibles sur notre serveur vocal sont indiquées dans le Tableau 2.

Voix	Sexe	Constructeur	Langue
Juliette	Féminin	Loquendo	Français
Mary	Féminin	Microsoft SAPI	Anglais
Mike	Masculin	Microsoft SAPI	Anglais
Sam	Masculin	Microsoft SAPI	Anglais
Mario	Masculin	Loquendo	Italien
Sonia	Féminin	Loquendo	Italien

Tableau 2 : Voix utilisables pour le service web de TTS développé

- L'opération inverse par rapport à la synthèse vocale s'appelle la reconnaissance vocale (ou ASR pour Automatic Speech Recognition). Ici, nous souhaitons pouvoir traiter n'importe quelle phrase, hors contexte. Autrement dit, on ne développe pas un service web auquel on fournirait une phrase (sous forme audio) et une liste de mots (sous forme textuelle) en demandant à la machine de trouver quels sont les mots de la liste qui ont été prononcés par l'utilisateur, mais, l'on fournit une phrase enregistrée sous format audio, que la machine doit tenter de transcrire au mieux, sans aucune indication quant au contexte de cette phrase (locuteur, accent, langue, vocabulaire, tâche, etc.). J'ai également codé ce service web en langage C#. Le prototype fonctionne, mais n'est pas assez robuste en environnement bruité [Rouillard 2006b]. Le traitement de la langue anglaise s'effectue grâce aux outils de Microsoft (SAPI⁵⁷ ASR), tandis que pour la langue française, nous nous basons sur le moteur de reconnaissance vocale du logiciel Dragon.

⁵⁶ MRCP : Media Resource Control Protocol. Le but de ce protocole est de standardiser les dialogues informatiques entre les moteurs d'ASR et de TTS avec les serveurs vocaux (SVI). Il s'agit d'un dialogue client-serveur. Le protocole définit les requêtes, les réponses et le contrôle du média. Il est lui-même basé sur d'autres protocoles comme RTSP (Real Time Streaming Protocol) et RTP (Real Time Protocole).

⁵⁷ SAPI : Speech Application Programming Interface.

- Le service web de détection de code barres que nous avons réalisé requiert en entrée de processus une photo et renvoie en sortie une chaîne de caractères représentant le code barre à une dimension ainsi détecté (ou un message d'alerte lorsque l'opération a échoué). La Figure 33 représente le type de capture d'image que nous pouvons ainsi traiter. Un périphérique léger (PDA, Smartphone ...) équipé d'un appareil photo permet de réaliser de tels clichés. La photo est ensuite envoyée, via le réseau Internet, au service web idoine et le résultat obtenu représente le code barre traduit sous forme textuelle.



Figure 33 : Un code barre (EAN) utilisé en grande distribution.

La Figure 34 est une capture d'écran d'une application cliente fonctionnant sous smartphone (HTC Kaiser), permettant de prendre une photo d'un objet ou d'un produit disposant d'un code barre, puis de l'envoyer au service web que nous avons développé afin de présenter à l'utilisateur le code retrouvé. Nous voyons dans cet exemple que le code barre de la bouteille d'eau présenté en Figure 33 à été retrouvé. Il est ensuite possible de retrouver grâce à ce code, le produit qui lui est associé, au sein d'une base de données (par exemple) et d'effectuer tous les traitements que l'on souhaite, comme la vérification des stocks disponibles, la vérification des allergies alimentaires possibles par rapport à un profil utilisateur connu, etc.

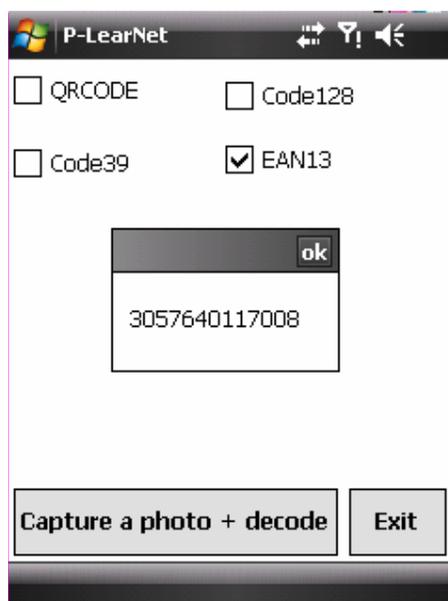


Figure 34 : Détection du code barre de la bouteille d'eau Volvic de 50 cl.

De même, la Figure 35 montre une photo d'une quatrième de couverture d'un ouvrage comportant un code barre. Grâce au même service web que celui que nous venons de présenter, il est possible de détecter l'ISBN (International Standard Book Number) de ce livre.

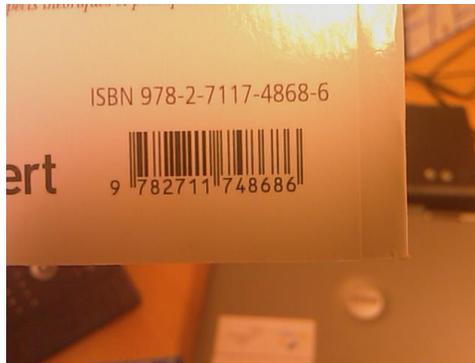


Figure 35 : Un code barre utilisé pour le codage d'un ISBN

Là encore, la Figure 36 présente le code barre textuel ainsi retrouvé, et il serait facile (nous l'avons déjà programmé pour d'autres applications [Rouillard & al. 2007]) d'utiliser le service web d'Amazon afin d'obtenir toutes les informations que l'on souhaite (titre, auteur, prix, avis et commentaires des clients, etc.) à propos de cet ouvrage, sur la base de son code ISBN.

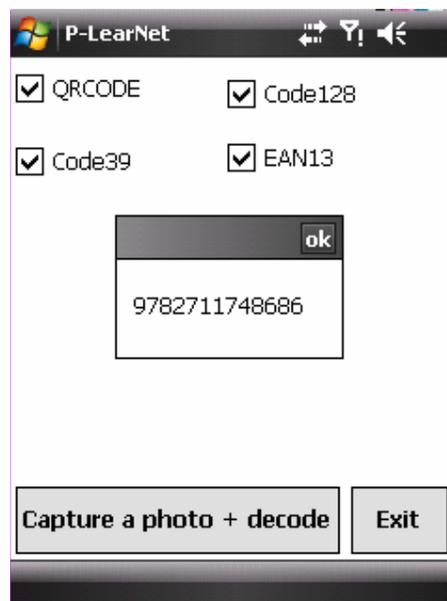


Figure 36 : Détection d'un code barre représentant l'ISBN d'un livre

J'ai également travaillé sur des services web de décodage de code barres à deux dimensions qui ont été utilisés dans les expériences présentées précédemment (QR Codes au zoo, etc.).

3.2.7 Instrumentation et mécanisme de traces pour VoiceXML

La multimodalité permet d'utiliser plusieurs modalités différentes, au sein d'une même application, pour interagir avec le système. Cela peut se faire de manière séquentielle (le téléphone d'abord, puis le Web ensuite), ou bien de manière synergique (Web et téléphone simultanément, par exemple). Des problèmes de synchronisation émergent alors au niveau des interfaces et des bases de données manipulées. La thèse de Vincent Chevrin que j'ai co-encadré montrait en quoi ces aspects sont différents et complémentaires l'un de l'autre : c'est principalement l'aspect temporel qui les relie ; il est donc important de bien choisir la granularité temporelle (milliseconde, seconde, minute, heure, jour, mois, année ...) pour chaque interaction.

La multicanalité, quant à elle, permet aux utilisateurs d'employer plusieurs canaux de communication différents pour entrer en contact avec une organisation (en E-commerce, en E-formation, etc.). Ces canaux peuvent être le support papier, l'E-mail, le téléphone, le Web, etc. Pour chaque interaction, un seul canal est clairement identifié. Des problèmes de cohérence des informations traitées doivent alors être pris en compte.

Des outils et des traces sont donc nécessaires pour évaluer l'utilisabilité et la pertinence de tels systèmes. Cela est particulièrement vrai pour l'évaluation des interfaces vocales et multimodales *in situ* et en contexte. Pour cela j'ai procédé au recueil de corpus de données afin d'établir des constatations vis-à-vis d'hypothèses scientifiques formulées avant étude. Cela permet de modéliser plus finement encore les interactions multimodales et multicanal ciblées dans nos travaux.

Ces travaux s'intègrent dans une réflexion commune des chercheurs de la région Nord, autour des EIAH⁵⁸, à travers le groupe de travail FERNAHTI⁵⁹. Les constatations des communautés scientifiques régionales et nationales demeurent les mêmes : les corpus ainsi que les moyens d'observer les usages *in situ* sont peu nombreux. Il faut donc trouver de nouveaux outils pour modéliser, concevoir et réaliser des systèmes informatiques capables de fournir des traces (à la fois bas niveau et d'un degré de complexité plus élevé) au moment de leur utilisation. Il est aussi très pertinent de capturer des informations relatives aux contextes d'usages (état des réseaux, niveau de stress de l'utilisateur, niveau sonore ambiant, luminosité, etc.), de manière à pouvoir reproduire ces situations, plus tard, si l'on veut comprendre ce qu'il s'est passé au moment de l'interaction. Les résultats de mon étude sont disponibles en annexe 9. Les éléments principaux à retenir sont résumés ici :

Etant donné que le langage VoiceXML ne possède pas de manière endogène un mécanisme permettant de tracer les interactions avec les utilisateurs qui se connectent sur le Serveur Vocal Interactif (SVI), il n'était pas possible jusqu'à la version 2.0 de VoiceXML, d'effectuer une trace réelle de ce qu'avait effectivement prononcé les utilisateurs au cours de leur dialogue avec le système. En revanche, il était possible de consulter la variable chargée de récupérer la valeur supposée correspondre à un élément de la grammaire vocale. Autrement dit, il n'était pas possible, jusqu'ici, d'enregistrer de manière automatique ce que l'utilisateur prononçait réellement lors des différents tours de parole.

Le processus que nous avons mis en œuvre afin d'enregistrer les conversations entre les utilisateurs et notre serveur vocal était externe au système et enregistrait également les autres bruits ambiants. Cette première étape a démontré que ce que disent les utilisateurs n'est pas forcément ce que le système croit entendre. Typiquement, à la question « Quel est votre âge ? », il est arrivé que la machine enregistre la réponse « 33 » dans la base de données, alors que la véritable réponse prononcée par l'utilisateur était « 23 ». L'expérimentateur, en écoute discrète avait noté cette anomalie et a pu réécouter l'enregistrement sonore, plus tard, afin de conforter son opinion.

Depuis la version 2.1 du VoiceXML, relativement récente [VoiceXML 2.1], il est techniquement possible d'enregistrer⁶⁰ ce que prononce l'utilisateur lors d'une interaction vocale. Pour cela, il faut initialiser l'attribut *recordutterance* de la balise *<property>* à la valeur « *true* ». On peut alors obtenir, grâce à certaines variables d'application, les données suivantes :

application.lastresult\$.confidence : Le niveau de fiabilité de l'énoncé de cette interprétation dans l'intervalle 0.0-1.0. Une valeur de "0.0" indique une fiabilité minimale et une valeur de "1.0" une fiabilité maximale ;

application.lastresult\$.inputmode : Le mode selon lequel l'entrée d'utilisateur a été fournie : "dtmf" ou "voice" ;

application.lastresult\$.recording : La donnée vocale correspondant à ce que l'utilisateur a dit ;

application.lastresult\$.recordingduration : La durée (en msec) de la dernière reconnaissance vocale ;

application.lastresult\$.recordingsize : La taille (en octets) de la dernière reconnaissance vocale ;

application.lastresult\$.utterance : La chaîne des mots bruts qui ont été reconnus pour cette interprétation. Dans le cas d'une grammaire DTMF, cette variable contiendra la chaîne numérique reconnue. Nous avons utilisé cette spécificité dont dispose le serveur vocal de notre laboratoire⁶¹ pour mettre en œuvre un mécanisme d'enregistrement global de toutes les traces d'interactions homme-machine possibles en VoiceXML (voix et clavier téléphonique). Cette instrumentation permet donc de collecter des traces qui reflètent l'usage d'une application. Ces informations sont estampillées et enregistrées dans une base de données. Nous avons développé un moyen d'accès à ces traces, pour qu'un expérimentateur, non informaticien, puisse analyser les données

⁵⁸ EIAH : Environnements Informatiques pour l'Apprentissage Humain.

⁵⁹ FERNAHTI : Fédération des chercheurs des Régions Nord pour l'Apprentissage Humain via les Technologies Informatique.

⁶⁰ Voir des exemples d'utilisation de TellMe à l'URL : https://studio.tellme.com/vxml2/ovw/recog_reco.html

⁶¹ En effet, le SVI Sibilo Voice de l'entreprise App-Line supporte partiellement la version 2.1 de VoiceXML.

recueillies [Kampire 2006]. A partir de cette interface, un « expert » du domaine étudié peut prendre connaissance de l'enregistrement vocal (fichier .wav) correspondant à la réponse d'un utilisateur et le comparer à ce que la machine a cru comprendre durant l'interaction.

Nous avons également testé ce processus de traces avec une application multimodale de commerce électronique comportant une reconnaissance et une synthèse vocale, l'usage possible du clavier téléphonique DTMF, mais aussi des clics souris sur des hyperliens d'une page Web, la visualisation des images des produits présentés, etc. La Figure 37 ci-après indique, en première ligne par exemple, que lorsque l'action consistait à choisir une taille de vêtement, la machine a cru reconnaître la réponse « large ». En cliquant sur le fichier .wav de cette même ligne, l'expert peut vérifier que l'utilisateur avait bien prononcé cette information, et le cas échéant rectifier l'information. Les statistiques de bonnes/mauvaises compréhension de la part de la machine sont ainsi mises à jour et permettent d'instrumentaliser les évaluations des interfaces générées.

Time	Task	Machine understanding	Record file name
16h 40m 19s	choose size	large	user_Tue16May_16h_40min_19s_voice.wav
16h 40m 34s	choose color	blue	user_Tue16May_16h_40min_34s_voice.wav
16h 40m 53s	choose product reference	25	user_Tue16May_16h_40min_53s_voice.wav
16h 41m 05s	add to cart	no	user_Tue16May_16h_41min_05s_voice.wav
16h 44m 12s	give client reference	228136	user_Tue16May_16h_44min_12s_voice.wav
16h 44m 30s	choose product set	blue dress	user_Tue16May_16h_44min_30s_voice.wav
16h 44m 40s	choose shelf	woman section	user_Tue16May_16h_44min_40s_voice.wav

Figure 37 : Traces d'enregistrement vocal en VoiceXML 2.1

Le mécanisme mis en œuvre pour cette instrumentalisation est schématisé sur la Figure 38 ci-dessous. En résumé, une page web dynamique récolte les traces du tour de parole précédent, puis sauvegarde ces informations dans une base de données ainsi que des fichiers audio associés, et enfin génère les fichiers VoiceXML pour la prochaine interaction.

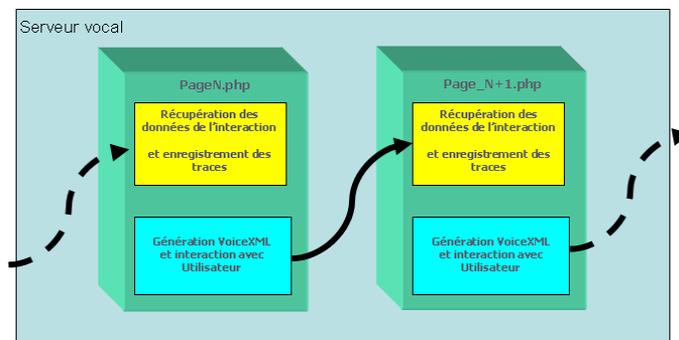


Figure 38 : Mécanisme pour recueillir des traces d'interactions sur SVI

Je travaille actuellement à l'amélioration des serveurs vocaux qui intègrent déjà certaines fonctionnalités du langage VoiceXML 2.1 [VoiceXML 2.1], comme l'enregistrement de ce que prononce l'utilisateur afin d'effectuer des traces de manière standard. Nous sommes en train de tester, en collaboration avec l'entreprise App-Line⁶², certaines balises (dites propriétaires, car ne faisant pas partie de la spécification VoiceXML 2.1 officielle) permettant de débiter et de stopper une trace automatique, et cela à tout moment de l'interaction avec le SVI. Trois solutions sont utilisables : soit grâce à une console de suivi en temps réel des conversations transitant sur le SVI (l'expert clique sur un bouton pour activer/stopper l'enregistrement vocal), soit de manière logicielle (une balise <object> en VoiceXML), soit encore par le biais de Web services supportant le protocole SOAP⁶³ [Chevrin & Rouillard 2008].

Pour résumer ces travaux portant sur la seconde partie de ma présentation, notons que je me suis principalement focalisé sur l'aspect vocal de la multimodalité, en utilisant une démarche modulaire. C'est-à-dire que j'ai tenté de créer des briques logicielles permettant, dans une vision SOA, de réutiliser des composants en les agençant selon les besoins rencontrés, y compris à l'exécution. La section suivante dresse un bilan global de mes recherches et propose quelques éléments de réflexion pour des travaux futurs.

⁶² <http://www.app-line.com>

⁶³ SOAP : Simple Object Access Protocol

Contexte de recherche : de l'axe « Multimodalité et Multicanalité » :

- Thèse de Vincent Chevrin
- Master Recherche Lina Makdessian
- Master Recherche Diane Kampire
- Master Recherche Simon Perrault
- Projets MIAOU / NIPO / EUCUE / pLearNet

Publications : (les références en gras sont présentées en intégralité en annexes)

[Rouillard & Perrault 2008] Rouillard José, Perrault, Simon, Un délégué IVY favorisant les interactions multimodales et multicanal en situation de mobilité, Workshop EMAP (Environnements Mobiles et Apprentissage Pervasif), Paris, 2008.

[Chevrin & al. Rouillard 2008] Chevrin Vincent, Rouillard José, Instrumentation and measurement of multi-channel services systems, International Journal of Internet and Enterprise Management (IJIEEM), Inderscience Publishers, Special Issue on: "Quality in Multi-Channel Services Employing Virtual Channels", Guest Editors: Dr. Rui Sousa, Catholic University of Portugal (Porto), Portugal, Volume 5 - Issue 4, pp. 333 – 352, 2008.

[Rouillard & al. 2008] Rouillard José, Peter Yvan, Tarby Jean-Claude, Vantroys Thomas, Chevrin Vincent, Supporting Mobile Connectivity: from Learning Scenarios to Multichannel Devices, International Journal of Continuing Engineering Education and Life-Long Learning (IJCEELL), Inderscience Publishers, Special Issue on Learning as a Ubiquitous and Continuous Communication Attitude, Guest Editor: Piet Kommers, Volume 18, No. 4., pp.396–410, 2008.

[Rouillard & al. 2007] Rouillard, José, Vantroys, Thomas, Chevrin, Vincent, Les architectures orientées services, Une approche pragmatique des SOA, Vuibert, ISBN : 2-7117-4868-5, 317 pages, Paris, 2007.

[Couturier & al. 2008] Couturier Olivier, Chevrin Vincent, Mephu Nguifo Engelbert, Rouillard José, Vers une convergence entre l'IHM et l'ECD basée sur les facteurs humains pour la conception de systèmes décisionnels performants, Volume 8, numéro 2, pp. 117-146, Revue RIHM - Revue d'Interaction Homme-Machine, Hermes-Lavoisier, Paris, 2008.

[Couturier & al. 2007] Couturier Olivier, Rouillard José, Chevrin Vincent, An interactive approach to display large sets of association rules, HCI International 2007, Lecture Notes in Computer Science, Springer-Verlag, Berlin, Beijing, Chine, 2007.

[Tarby & al. 2007] Tarby Jean-Claude, Ezzedine Houcine, Rouillard José, Tran C.D., Laporte Philippe, Kolski Christophe, Traces using aspect oriented programming and interactive agent-based architecture for early usability evaluation: basic principles and comparison, HCI International 2007, Lecture Notes in Computer Science, Springer-Verlag, Berlin, Chine, 2007.

[Rouillard 2006] Rouillard José, Web services and speech-based applications around VoiceXML, Journal of Networks (JNW, ISSN 1796-2056), Academy Publisher, Volume : 2, Issue : 1, pp. 27-35, Oulu, Finland, 2006.

[Rouillard 2006b] Rouillard, J., Web services and speech-based applications, ICPS'06, IEEE International Conference on Pervasive Services 2006 Lyon, 2006.

[Chevrin & al. 2006b] Chevrin Vincent, Derycke Alain, Rouillard José, Project Ubi-Learn: an Intermediation Infrastructure for Multi-channel Accesses to Future LMS, ELETE (E-learning and mobile learning on telecommunications), Proceedings of the Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services (AICT/ICIW 2006) 0-7695-2522-9/06, IEEE Computer Society Press, Vol n° 4, Guadeloupe, French Caribbean, 2006.

[Chevrin & al. 2006] Chevrin Vincent, Rouillard José, Derycke Alain, Sockeel Sébastien, Multi-channel Adaptive Context-sensitive Intermediation: Reconciling Marketing goals and Users Needs, MAC (Multi-channel Adaptive Context-sensitive), Systems: Building Links Between Research Communities, Glasgow, 2006.

[Couturier & al. 2006] Couturier Olivier, Rouillard Jose, Chevrin Vincent, Une approche hybride pour une meilleure visualisation de grands ensembles de règles d'association, ERGO'IA 2006, Biarritz, 2006.

[Chevrin & al. 2005] Chevrin Vincent, Rouillard José, Derycke Alain, Multi-channel and multi-modal interactions in E-marketing: Toward a generic architecture for integration and experimentation, HCI International 2005, Lawrence Erlbaum Associates, Inc., Las Vegas, Nevada, USA, 2005.

[Chevrin & al. 2005b] Chevrin Vincent, Derycke Alain, Rouillard José, Some issues for the Modelling of Interactive E-Services from the Customer Multi-Channel Interaction Perspectives, IEEE 05, International IEEE international conference on e-Technology, e-Commerce and e-Service, IEEE Press, pp. 256-259, Hong Kong, 2005.

[Rouillard & Truillet 2005] Rouillard, J., Truillet P., Enhanced VoiceXML, HCI International 2005, Lawrence Erlbaum Associates, Inc., Las Vegas, Nevada, USA, 2005.

[Rouillard & Dutter 2005] Rouillard José, Dutter Hervé, Harnessing the power of a VoiceXML engine to SMS based HCI, HCI International 2005, Lawrence Erlbaum Associates, Inc., Las Vegas, Nevada, USA, 2005.

[Derycke & al. 2005] Derycke Alain, Chevrin Vincent, Rouillard José, Intermédiations Multicanales et Multimodales pour l'E-Formation : l'Architecture du Projet Ubi-Learn, EIAH 2005, ATIEF, Montpellier, 2005.

[Chevrin & al. 2004] Chevrin Vincent, Derycke Alain, Rouillard José, L'Architecture Logicielle UMR pour les interactions Multicanaux et Multimodales avec les e-Services, IHM 2004, AFIHM, pp. 199-202, Namur, Belgique, 2004.

[Rouillard 2004] Rouillard J., VoiceXML. Le langage d'accès à Internet par téléphone, éditions Vuibert, ISBN : 271174826X, 197 pages, Paris, 2004.

[Derycke & al. 2003] Derycke Alain, Rouillard José, Chevrin Vincent, Bayart Yves, When Marketing meets HCI: Multi-channel customer relationships and multimodality in the personalization perspective, HCI International 2003, Lawrence Erlbaum Associates, Inc., Volume : 2, pp. 626-630, Heraklion, Crete, Greece, 2003.

[Chevrin & al. 2003] Chevrin Vincent, Derycke Alain, Rouillard José, Un cadre théorique pour la caractérisation des interactions multicanal en E-Marketing, IHM 2003, AFIHM, pp. 97-104, Caen, France, 2003.

[**Rouillard 2003b**] Rouillard José, Multimodalité et sémantique, Chapitre 8 de l'ouvrage "Variation, construction, instrumentation du sens", Hermès-Lavoisier, sous la direction de Maryse Siksou, 378 pages, pp. 177-196, Paris, 2003.

[Derycke & Rouillard 2002] Derycke Alain, Rouillard José, La personnalisation de l'interaction dans des contextes multimodaux et multicanaux : une première approche pour le commerce électronique, IHM 2002, AFIHM, Poitiers, 2002.

[Rouillard 2002] Rouillard, J., A multimodal E-commerce application coupling HTML and VoiceXML, Eleventh International World Wide Web Conference, Waikiki Beach, Honolulu, Hawaii, USA, 2002.

[Rouillard 2000] ROUILLARD José, Dialogue et Multimodalité, Revue RIHM - Revue d'Interaction Homme-Machine, Hermès-Lavoisier, Vol 2 N°1, pp.99-125, Paris, 2000.

Conclusion et perspectives

Dans ce document j'ai exposé mon activité de recherche en IHM effectuée depuis septembre 2000 à l'université des sciences et technologies de Lille. Je me suis intéressé à deux axes principaux qui sont, d'une part, l'adaptation des interfaces homme-machine et d'autre part l'étude des interactions multimodales et multicanal. Ces deux thématiques, présentées de manière disjointe dans cet exposé, sont en réalité conjointes dans mes travaux, au quotidien.

Ma contribution pour l'**axe adaptation** intervient au niveau de la conception et de la réalisation d'interfaces capables de s'adapter, au moment de l'exécution, à des éléments du contexte, fournis ou détectés. Pour étudier la plasticité des interfaces, j'ai tout d'abord montré, grâce aux travaux réalisés autour du développement d'un langage à balises (PlasticML), la possibilité de réaliser des interfaces capables de s'adapter dynamiquement au périphérique utilisé au moment de l'interaction. Ce langage, volontairement orienté vers les interfaces web basées sur XML et les standards du W3C n'avait pas pour vocation de couvrir tout le spectre de la plasticité, du modèle de tâches jusqu'à l'interface finale, mais devait permettre d'identifier les limites éventuelles d'une telle approche. Elles ont été identifiées au niveau conceptuel et au niveau des usages.

Au niveau conceptuel, et en matière de plasticité j'ai mis en exergue la difficulté de concevoir des applications qui pourront/devront être utilisées de manière multimodale. En effet, s'il est relativement aisé de concevoir des interfaces capables de s'adapter à un même type de modalité (graphique pour PC, graphique pour PDA, etc.), en revanche, il est plus complexe de proposer des interfaces plastiques utilisables de manière multimodale alternée ou synergique. Ceci est principalement dû au fait que ces modalités d'interactions ne partagent pas les mêmes repères, du point de vue cognitif, pour l'utilisateur : le graphique est plus lié à l'espace (agencement des objets à l'écran, taille, etc.), tandis que le vocal est plus lié au temps (séquences, tours de parole, etc.).

Au niveau des usages, et plus précisément, dans le cadre de l'apprentissage, j'ai souligné, au cours de mes travaux, la difficulté, voire l'impossibilité dans certains cas (cf. plateforme de formation à distance présentée dans [Bouyer & al. 2006]), de proposer une adaptation dynamique d'interface lorsque le support à cette interaction a été pensé, au moment de sa conception, pour un type de dispositif particulier et une modalité d'interaction particulière (le graphique au détriment des autres modalités, le plus souvent). De plus, si la conception de l'interface générique ne tient pas compte d'un modèle d'interaction, les risques sont grands pour que l'interface finale générée dynamiquement ne soit pas réellement utilisable. Cela se traduit concrètement par un manque de réactivité des interfaces cibles générées. De ce fait, les interfaces riches du Web 2.0 perdent de leur intérêt si elles ne sont pas accompagnées des scripts adéquats (cf. JavaScript et technologie AJAX) permettant des échanges événementiels entre clients et serveurs.

Des adaptations originales, comme notamment celle que j'ai présenté pour les rôles (hiérarchiques ou non) des utilisateurs, montrent l'importance de la sémantique dans les documents abstraits que l'on manipule. Certes, on savait déjà qu'il fallait séparer le fond de la forme, mais plus encore, pour le fond, il faut étiqueter (annoter de manière méta) les ressources pour qu'elles soient manipulables efficacement, lors de l'interaction. Malgré la prolifération, au niveau national et international, de travaux sur les ontologies, par exemple, cet aspect sémantique me semble trop souvent délaissé, car peu exploitable, par la communauté scientifique. J'y vois un verrou important, sur lequel je reviendrai dans mes perspectives.

Toujours sur le même axe de recherche, j'ai exposé mes activités concernant la mobilité des utilisateurs et la prise en compte du contexte lors de l'adaptation des interfaces et des documents manipulés. Dans le cadre de l'informatique pervasive et ubiquitaire, j'ai donné des exemples pour lesquels l'adaptation s'effectue sur un périphérique mobile, via une connexion à Internet de type 3G, à partir d'un code barre à deux dimensions présent dans l'environnement et selon plusieurs éléments du contexte comme la tâche, le rôle de l'utilisateur, le lieu, le moment de l'interaction. Des nombreuses expérimentations ont été menées (coach sportif⁶⁴, coach culinaire, assistance lors d'une réparation mécanique automobile, assistance pour l'enseignement et l'aide à la révision, visite contextuelle d'un

⁶⁴ Dans le domaine du coach sportif, des logiciels commerciaux très performants existent déjà sur le marché. C'est le cas de « Run.GPS » (<http://www.rungps.net/>), qui dispose de nombreuses fonctionnalités, mais qui ne permet pas encore (dans sa version 2.2.7 du 09/10/08) d'interagir avec le système pour lui faire modifier, en temps réel, le coaching proposé (notion de malléabilité).

zoo, chasse au trésor basée sur l'utilisation de QR Codes⁶⁵, etc.) et chaque prototype a permis de faire des avancées en terme de modélisation et d'ingénierie des IHM. Par exemple, en situation de coopération, pour monter un meuble en kit, les instructions données par la machine ne seront pas les mêmes selon que l'on déclare être seul ou accompagné de plusieurs personnes pour accomplir une tâche. Ces travaux, de part leur aspect multimodal rejoignent le second axe de recherche que j'ai présenté dans ce document.

Ce deuxième **axe** concerne les **interactions multimodales et multicanal**. En effet, l'adaptation du système est d'autant plus appréciée par les utilisateurs qu'ils peuvent alterner ou combiner les modalités d'interactions à leur guise. Un professeur des écoles utilisera tantôt une synthèse vocale pour que tout son auditoire puisse entendre (à défaut de lire sur l'écran⁶⁶) les informations proposées par le système, tantôt, il préférera, dans un premier temps, prendre connaissance des données, pour, dans un second temps, les reformuler à ses élèves, selon sa propre approche pédagogique. Ma contribution pour cet axe se situe à plusieurs plans. J'ai tout d'abord proposé un référentiel permettant de positionner un système en termes de canaux, modalités et dispositifs utilisables. Il permet, selon moi, de mieux discerner les difficultés que l'on tente d'adresser lors de la phase de conception du système. J'ai ensuite étudié plus particulièrement les interfaces multimodales couplant le vocal avec d'autres moyens d'interaction, notamment via le canal téléphonique. J'ai longuement étudié le langage VoiceXML proposé par le W3C, et j'ai rédigé un livre sur ce sujet, à destination des étudiants, mais également des enseignants et des chercheurs de ce domaine. Cette démarche se voulait pragmatique, pour permettre aux lecteurs de passer rapidement de la théorie à la pratique, afin qu'ils deviennent concepteurs d'IHM supportant un réel dialogue homme-machine vocal, et qu'ils se rendent compte, par eux-mêmes, des avantages et des inconvénients à utiliser de telles interfaces.

L'utilisation des architectures orientées services (SOA) dans le domaine de la multimodalité m'a permis, sinon de lever, au moins de contourner, un verrou important en matière de limitation des grammaires vocales dans le langage VoiceXML. Derechef, j'ai constaté qu'il y avait un important fossé entre la théorisation autour des SOA et la réelle mise en œuvre de telles architectures. Aucun des ouvrages francophones sur ce domaine n'expliquait réellement et concrètement comment procéder à la réalisation d'applications de ce type⁶⁷. J'ai donc à nouveau rédigé un ouvrage (en collaboration avec Thomas Vantroys et Vincent Chevrin) présentant les facettes théorique et pragmatique des SOA, et en essayant de montrer les bénéfices à utiliser de telles approches.

Par ailleurs, j'ai fourni un effort pour parvenir à une solution permettant de tracer les usages des applications à composantes dialogiques et téléphoniques, de manière endogène, pour lesquels, jusqu'à présent, peu de données quantitatives et qualitatives étaient proposées par la communauté scientifique. Plus spécifiquement, lorsque un problème survenait lors d'une interaction vocale téléphonique entre un humain et une machine, supportée par VoiceXML, il était pratiquement impossible, après coup, de connaître la nature et la cause de l'incident survenu. J'ai décrit une manière possible de pallier à ce problème, mais il reste encore beaucoup d'extensions possibles à ces travaux. Ces derniers n'ont apportés que quelques pierres à l'édifice. Les perspectives de recherche dans ces domaines sont nombreuses et je vais en décrire quelques unes.

Programme de recherche à moyen terme :

Globalement, on note depuis quelques années des avancées en matière de conception des interfaces multimodales et certains outils commencent à apparaître, à l'image du langage EMMA du W3C ou de ceux proposés par la fondation OpenInterface, ou bien encore autour de USIXML et de MultimodalXML [Stanciulescu 2008]. Cependant, à moyen et long terme, l'**informatique ubiquitaire et l'intelligence ambiante** seront prédominantes. Mes travaux de recherche futurs s'orientent vers cette direction.

⁶⁵ Projet de master Informatique : « Mise en œuvre d'un générateur de QR Codes enchaînés pour jeux de piste ». <http://www.qrcoding.free.fr/>

⁶⁶ Pour deux raisons principales : (1) l'écran de périphérique mobile est souvent trop petit pour que tous les élèves d'une même classe puissent consulter simultanément ces informations, et (2) certains enfants ne savent pas encore lire à ce stade de leur apprentissage (niveau CP par exemple).

⁶⁷ C'est d'ailleurs à cette occasion que j'ai adopté la devise de Linus Torvalds : « *Talk is cheap. Show me the code.* ».

Je souhaite me focaliser, durant mon CRCT⁶⁸, obtenu pour l'année 2008-2009, et plus globalement, pour les cinq ans à venir sur différents aspects de ma problématique d'une manière plus profonde et plus systématique. J'aimerais pouvoir cristalliser ma réflexion autour des IHM intelligentes et adaptables au contexte, depuis les tâches jusqu'aux interfaces finales, sans faire l'économie de certains pans de ma recherche qui me semblent déterminants. Je pense notamment aux expérimentations de type **Magicien d'OZ**, qui permettent de simuler le comportement d'un système aux yeux d'un sujet qui croit avoir affaire à un système opérationnel, alors que c'est en réalité un compère qui prend les décisions à la place du système (puisque ce dernier n'existe pas réellement). Cette technique est longue et difficile à mettre en œuvre, mais elle est pourtant très efficace pour pouvoir observer des comportements et des usages avant même que les outils nécessaires ne soient déployés. J'aimerais mettre en œuvre une telle démarche scientifique dans le champ des IHM multimodales et multicanal afin de considérer et d'anticiper les usages qui pourraient être proposés sans contraintes et sans limites conceptuelles ou technologiques, et ceci, particulièrement en situation de mobilité. Des terrains d'expérimentation comme le tourisme ou la domotique, se prêtent bien à ce genre de démarche. Des publics variés (personnes âgées, jeunes enfants, handicapés ...) peuvent être étudiés, dans leur environnement propre. La communauté scientifique doit cependant fournir un effort afin d'extraire les outils d'expérimentations hors des laboratoires de recherche pour aller à la rencontre des « vrais » utilisateurs finaux.

Dans ce sens, je souhaite expérimenter plus avant l'usage de cameras, de gestes des utilisateurs et du Dialogue Homme-Machine (DHM), par exemple. Des projets étudiants que j'ai encadrés en master informatique M1 et M2 (usages de wiimotes⁶⁹ pour des applications multimodales, usage de QR CODEs, de phidgets ...) montrent des perspectives intéressantes, que l'on pourrait mettre en œuvre à court terme. Les pôles de compétitivité, comme celui auquel notre laboratoire est associé (le PICOM, pour les industries du commerce), pourraient sans doute trouver un intérêt dans ces travaux, à condition que les retombées soient tangibles, mutualisables et rapidement transposables à leurs attentes et leurs cœurs de métier. Cette recherche de fond, ancrée sur nos thématiques historiques centrées sur l'utilisateur, se verrait alors colorée par des nouvelles considérations autour de l'intelligence ambiante et des interfaces dites homme-environnement.

Il s'agit plus concrètement de travaux autour de la **vente assistée par ordinateur**. L'assistance pourra bénéficier aussi bien aux vendeurs dans les rayons d'un magasin, équipés d'oreillettes, de PDA, de grands écrans interactifs, qu'aux clients, qui pourront être guidés dans leurs achats et conseillés en temps réel, lors de la comparaison de produits. Les Agents Conversationnels Animés (ACA) sont un des volets de cette thématique, mais bien d'autres pistes sont encore envisageables. Un même agent logiciel sera-t-il alors capable de s'adapter pour conseiller, tantôt un client achetant sur Internet, tantôt un autre, dans un magasin (en cabine d'essayage par exemple) ? On voit, ici encore, que l'adaptation et les couplages entre modalités sont au cœur de ces préoccupations et des travaux de recherche que je compte mener, en renouant avec mes premiers domaines d'études relatifs au commerce. Chacune de ces pistes ouvre à son tour d'autres sous-thèmes de recherche. Dans le cas du DHM en langue naturelle, il faudrait davantage approfondir nos connaissances autour de domaines connexes aux IHM (protocoles SIP⁷⁰ / MRCP⁷¹ / VOIP⁷² / norme VoiceXML 3.0). Je fournirai des efforts dans ce sens et proposerai, si besoin, des supports pédagogiques aux étudiants et chercheurs intéressés par ces aspects.

Afin de faire converger ces travaux, les notions d'ontologies, de moteur de composition de documents numériques, de services web ou encore de traces, par exemple, nous seront très utiles. Il s'agit en tout premier lieu de définir concrètement la manière dont les flux vont transiter, afin que l'information soit successivement (voire parallèlement) traitée. Des API⁷³ doivent être fournies par les différents partenaires d'un projet, afin que l'on puisse effectivement identifier les entrées et les sorties de processus à chaque étape de transformation. Pour ma part, je continuerai à contribuer, comme je l'ai fait jusqu'à présent, à la conception et la mise en œuvre des flux relatifs aux services utiles aux interactions multimodales et multicanal.

⁶⁸ CRCT : Congé pour Recherche et Conversion Thématique

⁶⁹ Manette de jeu de la console Wii de Nintendo.

⁷⁰ Session Initiation Protocol

⁷¹ Media Resource Control Protocol

⁷² Voice Over Internet Protocol

⁷³ API : Application Programming Interface

Ainsi, nous pourrions expérimenter différents mécanismes d'adaptation, car il n'est pas du tout évident que le concept du « pipeline » soit toujours le meilleur, selon le contexte. Des questions de fond demeurent alors : comment garantir la meilleure adaptation possible, selon un contexte déterminé. Est-ce que l'ordre des adaptations successives peut-être calculé de manière optimale, et selon quels critères ? En supposant des adaptations parallèles, comment détecter et éviter les incohérences latentes ?

En fonction des différentes modalités d'interaction utilisables et de leur contexte d'utilisation (activités en cours d'exécution, localisation, profil utilisateur...), on cherchera à réaliser l'orchestration de service la plus pertinente possible. Les problèmes classiques de fusion et de fission des informations demeurent, mais on voit aussi apparaître d'autres aspects du problème, que nous avons déjà commencé à étudier, notamment au travers de la thèse de Vincent Chevrin [Chevrin 2006]. Il s'agit entre autres des notions de ruptures, de transition de réseau à un autre, de découverte et de qualité de services proposés [Chevrin & Rouillard 2008], de couplage/découplage de modalités...

J'identifie ici une première série de **verrous** scientifiques au niveau **conceptuel**. En effet, la communauté scientifique ne propose pas actuellement, compte tenu de l'évolution apparente des tendances vers l'informatique pervasive et ubiquitaire, de solutions stables pour modéliser et concevoir des systèmes informatiques doués de capacités d'adaptation flexibles par rapport au contexte, aux modalités, aux canaux et aux dispositifs. La combinatoire entre ces critères est telle que l'on ne sait pas, à l'heure actuelle, concevoir des interfaces de ce style, autrement que d'une manière *ad hoc*. Les applications connues pour leur efficacité et leurs performances, dans le domaine militaire, de la santé ou des transports, par exemple, sont presque toutes réalisées sur mesure, et coûtent cher à développer et à maintenir. Il n'existe quasiment pas de plateforme de conception permettant de concevoir et de réaliser de manière relativement simple de tels systèmes. Une modélisation plus fine et des progrès sur ce sujet pourraient permettre une meilleure puissance descriptive des systèmes, mais également prédictive. Ainsi des propositions de modalités pourraient plus facilement être calculées et proposées, *in situ*, au moment même de l'interaction et selon des mécanismes autorisant la découverte d'éléments du contexte et des nouveaux services.

D'autre part, la réalisation d'interfaces réellement intuitives nécessite la collecte de données expérimentales permettant l'observation, l'annotation et l'analyse des comportements lors de situations de communication homme-machine ou homme-homme. Des « corpus multimodaux » existent déjà (cf. travaux de Bellik), mais ils utilisent le plus souvent des mécanismes de traces exogènes, comme des caméras vidéo, ou encore des annotations manuelles, effectuées par des expérimentateurs. Ils sont donc très lourds à mettre en œuvre et sont souvent élaborés au niveau européen⁷⁴. Nous avons expliqué notre démarche en vue de l'automatisation de recueils de corpus, via des mécanismes de **traces endogènes au système**, pour le canal téléphonique plus particulièrement. Ces travaux rejoignent ceux auxquels j'ai participé à propos de l'évaluation précoce des systèmes interactifs utilisant la programmation par aspects permettant d'obtenir des informations pertinentes pendant l'exécution du logiciel [Tarby & al. 2007]. Nos efforts collectifs doivent persister dans cette voie prometteuse.

D'ailleurs, il me semble que l'aspect sémantique a un rôle important à jouer à ce niveau de la réflexion. Les choix d'adaptations à effectuer, tout comme les choix des (multi)modalités les plus appropriées pourraient être plus facilement proposés à l'utilisateur si le système pouvait « comprendre » les éléments qu'il manipule. Un travail de convergence, au sein de notre équipe, entre le domaine des IHM et celui des infrastructures nous a amené à étudier l'utilisation de moteurs de workflow pour l'adaptation d'interfaces et de documents numériques. Nous poursuivons nos efforts dans ce sens. La Figure 39 donne un exemple simple de workflow décrivant les tâches qu'un système est capable de réaliser. A partir d'un workflow de ce type, on peut dérouler des activités de manière interactive. Dans notre exemple, il peut s'agir par exemple de déplacer une webcam à distance. Pour une interface graphique, l'avancement dans le workflow permettra de proposer des choix (menu, cases à cocher, bouton radios, etc.) au fur et à mesure des éléments et de la sémantique rencontrés. Pour une interface vocale, des grammaires pourront être générées dynamiquement, en parcourant le graphe dans son intégralité. Dans événements externes pourront influencer les décisions prises, voire les règles elles-mêmes.

⁷⁴ Comme le projet européen IST-ISLE (International Standards for Language Engineering).

Dans notre exemple, l'utilisateur pourra prononcer la phrase « move cam_B6 left » pour faire pivoter la webcam de la salle B6 vers la gauche. Il faudra trouver une flexibilité permettant de répondre à des phrases du style « move up cam_M3 », qui ne suivent pas forcément l'ordre prévu par le concepteur, et qui sont pourtant correctes du point de vue grammatical et sémantique. Plus encore, toujours sur la base de ce même workflow, couplé à une table de décision (identifiant_produit , intitulé_produit), on imagine que l'on pourra solliciter l'identification d'un produit, en prononçant le mot « identify » sur un canal téléphonique, puis en scannant le code barre du produit via un smartphone équipé d'une camera, par exemple.

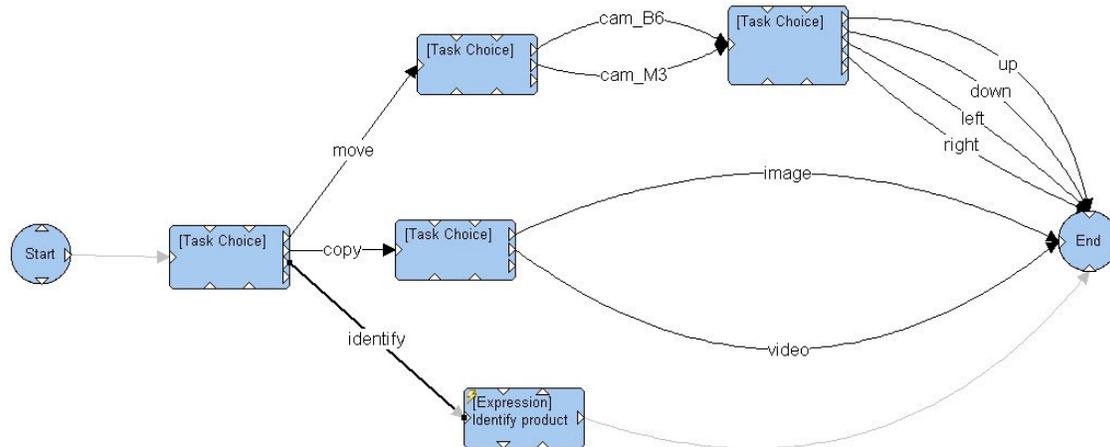


Figure 39 : Workflow décrivant les tâches d'une activité à réaliser

Les connexions à des services web de nature linguistique permettront d'étendre les possibilités de dialogue entre l'homme et la machine. On pourra ainsi utiliser des lemmatiseurs, des dictionnaires, des thesaurus et des traducteurs pour augmenter le vocabulaire disponible ou au contraire le restreindre par désambiguïsation sémantique.

Je perçois à ce niveau une seconde série de verrous scientifiques, plus liés aux **usages** de ces approches. A l'exécution, et dans le but d'éviter qu'une adaptation locale (par rapport à un élément du contexte) n'altère l'intention globale (du concepteur), donc le fonctionnement complet du système, on doit veiller à ce que les adaptations proposées soient cohérentes. Or, sans sémantique associée aux éléments manipulés, il est difficile de parler de cohérence. Pour accompagner ces mécanismes, il me paraît intéressant de solliciter l'utilisateur, dans un souci de transparence et d'honnêteté [Greenfield 2006]. Cette démarche centrée utilisateur devrait en théorie permettre de présenter formellement à ce dernier les mécanismes d'adaptation, pour engendrer une solution concertée et non subie. On devrait alors identifier et présenter toutes les « coutures » (jointures) entre les systèmes (*seamless system*), et non pas les cacher. De ce fait, un dialogue homme-machine en langue naturelle me semble plus que jamais nécessaire. Je considère en effet que le dialogue oral en langue naturelle est l'un des aspects les plus prometteurs des nouvelles interfaces homme-machine. Les téléphones mobiles vont devenir sans aucun doute de plus en plus puissants et permettront de coupler des interactions vocales avec d'autres moyens naturels d'interactions (le toucher principalement). De nombreuses pistes de recherche relatives à ces difficultés sous-jacentes s'ouvrent alors à nouveau.

Enfin, si je suis amené à poursuivre mes travaux au sein de mon équipe de recherche actuelle, je souhaite contribuer au rapprochement de mon laboratoire (le LIFL) avec le LAGIS⁷⁵ (Laboratoire d'Automatique, Génie Informatique et Signal). J'ai œuvré en ce sens, en déposant un dossier scientifique auprès du LIFL, dans le cadre d'actions incitatives internes au laboratoire, sur le thème du BCI. Ce dossier, avec un budget de 5000 euros a été retenu. En effet, une autre perspective de recherche qui m'intéresse grandement, parce que j'y vois un intérêt indéniable pour la société, pourrait être qualifiée d'ultime modalité. Il s'agit du BCI, pour **Brain-Computer Interface** [Hoffmann & al. 2004], ou ICO (Interface Cerveau-Ordinateur) en français. J'ai déjà étudié quelques aspects de cette problématique en utilisant un système capable de capter avec un casque ou de simuler logiquement des activités cérébrales (cf. annexe 10).

⁷⁵ Sous tutelle de l'Ecole Centrale de Lille et de l'Université des Sciences et Technologies de Lille (USTL)

Il apparaît, d'après nos échanges avec le professeur François Cabestaing du laboratoire LAGIS, spécialiste de ce domaine, que les besoins actuels ne portent pas spécifiquement sur le traitement du signal (par des mesures intrusives, par électrodes directement reliées au cerveau du patient et non intrusives, sur le scalp et le cuir chevelu), mais plutôt sur les IHM, et notamment le retour d'information dont dispose l'utilisateur-patient, souvent lourdement handicapé physiquement, lorsqu'on lui demande d'évoquer un mouvement, de manière mentale, pour que le système puisse capter des signaux cérébraux, et les convertir en interactions. On devrait alors pouvoir travailler sur une double adaptation : celle de l'homme à la machine, et conjointement, celle de la machine à l'homme.

4 Références

[Alapetite 2007] Alapetite, Alexandre, On speech recognition during anaesthesia, PhD thesis, Roskilde University, ISSN:0109-9779, 2007

[Anderson & al. 2001] Anderson, E. A., Breitenbach, S., Burd, T., Chidambaram, N., Houle, P., D. Newsome, D, Tang, X., Zhu, X., Early Adopter VoiceXML, Wrox, 2001.

[Bachimont & Crozat 2004] Bachimont, B., Crozat, S., Instrumentation numérique des documents : pour une séparation fonds/forme, Revue Information - Interaction – Intelligence (RI3), numéro spécial sur le document numérique, Volume 4, Numéro 1, 2004.

[Balla & al. 2006] Balla, A., Hidouci, K.W., Khodja, H., Oubraham, A. (2006). *CDPNEUMO: Un hypermédia éducatif pour l'apprentissage des concepts de base de la pneumologie*, colloque international CEMAFORAD3, Sousse'2006, Tunisie.

[Balint 1995] Balint, L. (1995). Adaptive interfaces for human-computer interaction: a colourful spectrum of present and future options. *Systems, Man and Cybernetics, Intelligent Systems for the 21st Century*, IEEE International Conference on Volume 1, Issue , 22-(pp. 292 - 297), vol.1.

[Bastide & al. 1998] Bastide, R., Palanque, P., Le, D., Munoz, J., Integrating rendering specifications into a formalism for the design of interactive systems. Dsv-is'98, Abington, UK, June, 1998, Springer Verlag.

[Baudel 1991] Baudel, Thomas, Spécificités de l'interaction gestuelle dans un environnement multimodal, IHM'91, p. 11-16, Dourdan, 1991

[Bellik 1992] Bellik, Y., Teil, D., Multimodal Dialog Interface. WWDU'92, Berlin, Septembre 1992.

[Bellik 2006] Bellik, Y., Présentation Multimodale de l'Information, HDR, l'Université d'Orsay Paris-Sud, 2006.

[Benyon 1993] Benyon D., (1993). Adaptive systems: A solution to Usability Problems. *User modeling and user adapted interaction*, (pp. 01-22), vol. 1(3).

[Benyon & Murray 1993] Benyon D., Murray, D., (1993). Developing adaptive systems to fit individual aptitudes, *International Conference on Intelligent User Interfaces*, Orlando, Florida, United States.

[Betbeder-Matibet 2003] Betbeder-Matibet, M-L. (2003). Symba : un environnement malléable support d'activités collectives en contexte d'apprentissage. Thèse de doctorat d'informatique, Université du Maine, Le Mans, France.

[Bolt 1980] Bolt, R.A., Put-that-here: voice and gesture at the graphic interface. *Computer Graphics*, 14 , 262-270, 1980.

[Bouchet & Nigay 2004] Bouchet Jullien & Nigay Laurence. ICARE: a component-based approach for the design and development of multimodal interfaces. *CHI'2004 Conference on Human Factors in Computing Systems*, pp 1325-1328, Vienna, Austria.

- [Bourguin 2000] Bourguin, G. (2000). Un support informatique à l'activité coopérative fondé sur la Théorie de l'Activité : Le projet DARE. Thèse de doctorat d'informatique, Université de Lille 1, Laboratoire Trigone, Lille, France.
- [Browne & al. 1990] Browne D., Totterdell P. & Norman M., (Eds.), *Adaptive User Interfaces, Computer and People Series*, Academic Press.
- [Bruillard 1997] Bruillard, E., Les machines à enseigner, Éditions Hermès, Paris, 1997.
- [Brusilovsky 1996] Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia, in *User Modeling and User-Adapted Interaction*, (pp. 87-129), 6 (2-3).
- [Buisson & al. 2002] Buisson, M., Bustico, A., Chatty, S., Colin, F.R, Jestin, Y., Maury, S., Mertz, C., Truillet, P., Ivy: Un bus logiciel au service du développement de prototypes de systèmes interactifs, IHM02, Poitiers, pp. 223-226, ACM, 2002.
- [Buxton 1990] Buxton, W., "Smoke and Mirrors", Byte, p. 202-251, 1990.
- [Caelen 1992] Caelen, J. Compte-rendu du « Workshop » Interfaces Homme-Machine multimodales, GDR-PRC CHM sur le Dialogue, Dourdan, 1992.
- [Caelen 2000] Caelen J. 10 ans de recherches... en multimodalité(s). Colloque sur les Interfaces Multimodales - 09/10 Mai 2000. IRIT. Toulouse.
- [Calvary & al. 2001a] Calvary, G., Coutaz, J., Thevenin, D., Supporting Context Changes for Plastic User Interfaces: Process and Mechanism, HCI-IHM, Lille, 2001.
- [Calvary & al. 2001b] Calvary G., Coutaz J., Thevenin D., A Unifying Reference Framework for the Development of Plastic User Interfaces, EHCI'01, IFIP WG2.7 (13.2) Working Conference, Toronto, May, 2001.
- [Calvary & al. 2003] Calvary, G., Coutaz J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonckt, J. (2003). A Unifying Reference Framework for Multi-Target User Interfaces, *Interacting With Computers*, (pp 289-308). Vol. 15/3.
- [Calvary & al. 2006] Calvary, G., Coutaz, J. Daassi, O., Ganneau, V., Balme, L., Demeure, A., Sottet, J.S. (2006). *Métamorphose des IHM et Plasticité : Article de synthèse*, Ergo'IA 2006, Biarritz, France.
- [Calvary 2007] Calvary, G., Plasticité des Interfaces Homme-Machine, HDR, Université Joseph Fourier de Grenoble, 2007.
- [CAMELEON] Projet Cameleon, <http://girove.cnuce.cnr.it/cameleon.html>
- [Chevrin 2006] Chevrin, V., L'interaction usagers/services, multimodale et multicanale : Une première proposition appliquée au domaine du e-Commerce, thèse de doctorat en informatique, Université de Lille1, 2006.
- [Cingil 2000] Cingil I., Dogac A. & Azgin A., (2000). *A broader approach to personalization*, Communications of the ACM, (pp. 136-141), 43(8).
- [Claeyssen & al. 2006] Yan Claeysen, Anthony Deydier, Yves Riquet, Francis Salerno (Préface), *Le Marketing direct multicanal : Prospection, fidélisation et reconquête du client*, Dunod, Collection : Fonctions de l'Entreprise, 327 pages, ISBN-13: 978-2100502639
- [Clark & Brennan 1991] Clark, H. and Brennan, S. Grounding in Communication. In *Perspectives on Socially Shared Cognition*. Edited by L.B. Resnick, J. M. Levine and S. D. Teasley. 1991.
- [Cockton 1987] Cockton, G. (1987). Some Critical Remarks on Abstractions for Adaptable Dialogue Managers. In D. Diaper, R. Winder (Eds.): *People and Computers III*. Proceedings of the *Third Browne*

Conference of the British Computer Society, Human Computer Interaction Specialist Group, (pp. 325-343). University of Exeter.

[Cohen & al. 1989] P. R. Cohen, M. Dalrymple, D. B. Moran, F. C. Pereira, et J. W. Sullivan, Synergistic use of direct manipulation and natural language, Proceedings of the SIGCHI conference on Human factors in computing systems, pages 227–233. ACM Press, 1989.

[Coulon & Kayser 1986] Coulon, D., Kayser, D., *Informatique et langage naturel : Présentation générale des méthodes d'interprétation des textes écrits*, Revue TSI : Technique et Science Informatique, Vol. 5. n°2, 1986.

[Coutaz & al. 2005] J. Coutaz, J. L. Crowley, S. Dobson, and D. Garlan, "Context is Key", Communications of the ACM, Special issue on the Disappearing Computer, Vol 48, No 3, pp 49-53 March 2005.

[Coutaz & al. 1995] Joëlle Coutaz, Laurence Nigay, Daniel Salber, Ann Blandford, Jon May, and Richard M. Young. Four easy pieces for assessing the usability of multimodal interaction : the CARE properties. In INTERACT, pages 115-120. Chapman & Hall, 1995.

[Coutaz & Nigay 1994] Coutaz J. and Nigay L. Les propriétés "CARE" dans les interfaces multimodales. In IHM'94, pages 7-14, Lille, Décembre 1994.

[De Bra & Calvi 1998] De Bra, P., Calvi, L. (1998). AHA: a Generic Adaptive Hypermedia System, Proc. of the 2nd *Workshop on Adaptive Hypertext and Hypermedia*, pp. 5-12, Pittsburgh, USA. [Http://wwwis.win.tue.nl/ah98/DeBra.html](http://wwwis.win.tue.nl/ah98/DeBra.html)

[Demumieux & al. 2008] Demumieux, Rachel, Ganneau, Vincent, Calvary, Gaëlle, Gegovska, Elisa, Les interfaces plastiques premiers retours utilisateurs : évaluations en laboratoire, Ergo-IA 2008, L'humain au cœur des systèmes et de leur développement, 2008, Biarritz.

[Denso Wave] Denso Wave QR Code and QR Code features, online document. <http://www.denso-wave.com/qrcode/qrcodefeature-e.html> et <http://www.denso-wave.com/qrcode/aboutqr-e.html>

[Dettmer 2003] Dettmer, R., It's good to talk, speech technology, for on-line services access, IEE Review, Volume: 49, Issue:6, June 2003, pp 30-33.

[Dey & Abowd 2000] Dey, A.K., Abowd, G.D., (2000). Towards a Better Understanding of Context and Context-Awareness, in Workshop on the What, Who, Where, When and How of Context-Awareness, at the 2000 Conference on Human Factors in Computing Systems.

[Dey & al. 2001] Dey, A.K., Salber, D. & Abowd, G.D. (2001). A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications, Human-Computer Interaction, (pp. 97-166), 16(2, 3, 4).

[Dietrich & al. 1993] Dieterich, H., Malinowski, U., Kühme, T. & Schneider-Hufschmidt, M. (1993). State of the Art in Adaptive User Interfaces. In M. Schneider-Hufschmidt, T. Kühme, U. Mallinowski (Ed.), *Adaptive User Interfaces, Principles and Practice* (pp.13-48). Included in series Human Factors in Information Technology, 10.

[Dragicevic 2004] Dragicevic, P., Un modèle d'interaction en entrée pour des systèmes interactifs multi-dispositifs hautement configurables, thèse de doctorat d'informatique, Nantes, 2004.

[Duarte 2006] Duarte, C., Carrico, L. A conceptual framework for developing adaptive multimodal applications. Proceedings of IUI 2006, ACM Press, p. 132-139.

[Edmonds 1981] Edmonds, E.A. (1981), *Adaptive man-computer interfaces*. In Coombs and Alty (Ed), (pp 389-426). CompuLing Skills and the User Interface.

[Eisenstein & al. 2001] Eisenstein J., Vanderdonck J., Puerta A. Applying Model-Based Techniques to the Development of UIs of Mobile Computers, in Proceedings of Intelligent User Interfaces 2001, Santa Fe NM, 14-17, pp. 69-76, 2001.

[EMMA] Extensible MultiModal Annotation markup language, W3C, <http://www.w3.org/TR/2005/WD-emma-20050916/>

[Fischer 2000] Fischer, G. (2000). User Modeling in Human Computer Interaction, *User modeling and user-adapted interaction*, (pp. 65-86), 11, Kluwer Academic Publishers.

[Frasincar & Houben 2002] Frasincar F. & Houben G-J, Hypermedia Presentation Adaptation on the Semantic Web, in de Bra P., Brusilovsky P. & Conejo R., (Eds.), Proceedings of the 2nd *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH 2002)*, Malaga, Spain, (pp. 133-142), LNCS 2347.

[Frey 2007] Frey, B., Ne confondons pas multicanal et multisupport, magazine *Stratégies*, 2007. http://www.strategies.fr/archives/1473/page_44546/ne-confondons-pas-multicanal-et-multisupport.html

[Frohlich 1991] Frohlich, D., The design space of interfaces, multimedia systems, *Interaction and Applications, 1st Eurographics workshop*, Stockholm, Suède, Springer Verlag, p. 53-69, 1991.

[Greenberg & Fitchett 2001] Greenberg, S., and Fitchett, C., Phidgets: easy development of physical interfaces through physical widgets, Proceedings of the 14th annual ACM symposium on User interface software and technology, pp. 209 – 218, Orlando, Florida, 2001.

[Greenfield 2006] Greenfield, A., *Everyware: The Dawning Age of Ubiquitous Computing (Voices That Matter)*, 2006, ISBN 0-321-38401-6, 267 pages.

[Grundy & Zou 2004] Grundy, J., and Zou, W., AUIT: Adaptable User Interface Technology, with extended Java Server Pages. In A. Seffah & H. Javahery (eds). *Multiple User Interfaces*. John Wiley & Sons, New York, pp.149-167, 2004.

[Habieb-Mammar 2004] Habieb-Mammar, H., (2004), *EDPHA : un Environnement de Développement et de Présentation d'Hyperdocuments Adaptatifs*. Thèse de doctorat d'informatique, Institut National des Sciences Appliquées de Lyon, France.

[Hariri 2008] Hariri, A. (2008). Contribution à une méthode de conception et génération d'interface homme-machine plastique. Mémoire de Doctorat, Université de Valenciennes et du Hainaut-Cambrésis, Valenciennes.

[Healey & al. 2002] Healey, J., Hosn, R., Maes, S.H, Adaptive Content for Device Independent Multi-modal Browser Applications, Lecture Notes In Computer Science; Vol. 2347, Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, pp. 401-405, ISBN: 3-540-43737-1, 2002.

[Henze & al. 2007] Henze, N., Lim, M., Lorenz, A., Mueller, M., Righetti, X., Rukzio, E., Zimmermann, A., Magnenat-Thalmann, N., Boll, S., Thalmann, D., Contextual bookmarks, Joint Workshop Mobile Interaction with the Real World, MIRW 2007 and the 5th Workshop on "HCI in Mobile Guides", MGuides 2007, Singapore, pp. 51-54, 2007

[Hoffmann & al. 2004] Hoffmann, U., Garcia, G.N., Vesin J.-M. et Ebrahimi, T., Application of the Evidence Framework to Brain-Computer Interfaces, in Proceedings of the IEEE Engineering in Medicine and Biology Society Conference, 2004.

[Horgen 2002] Horgen, S., A. (2002). *A Domain Model for an Adaptive Hypertext System based on HTML*, Master degree in Computer Science, Available on-line from <http://www.aitel.hist.no/~svendah/ahs.html>

[InkML] Ink Markup Language (InkML), W3C Working Draft 23 October 2006 <http://www.w3.org/TR/InkML/>

[IVY] Bus IVY : <http://www.tls.cena.fr/products/ivy/>

[Jameson 2001] Jameson, A. (2001). *Systems That Adapt to Their Users: An Integrative Perspective*. Saarbrücken: Saarland University.

[Khalil 2001] Khalil, C.I.J. (2001). *The Design of a Multimedia Adaptive Interface for Process Control Using a Multi-Agent Approach*. PhD thesis, Computer Science, Loughborough University, UK.

[Kappel & al. 2000] Kappel G., Retschitzegger W., & Schwinger W. (2000), *Modeling Customizable Web Applications – A Requirement's Perspective*, in *Proceedings of the International Conference on Digital Libraries, (ICDL 2000)*, Kyoto, Japan.

[Kampire 2006] Kampire, Diane, *Instrumentation des Suivis d'Usages dans le Contexte de l'Informatique Ubiquitaire*, Mémoire de master Recherche, Université des Sciences et Technologie de Lille, 2006.

[Karagiannidis & al. 1996] Karagiannidis, C., Koumpis, A., and Stephanidis, C.(1996). *Deciding 'What', 'When', 'Why', and 'How' to Adapt in Intelligent Multimedia Presentation systems*", *ECAI'96 Workshop on Towards a standard reference model for intelligent multimedia presentation systems*, Budapest, Hungary.

[Kelley 1985] Kelley, J.F., *CAL - A Natural Language program developed with the OZ Paradigm: Implications for Supercomputing Systems*. First International Conference on Supercomputing Systems, St. Petersburg, Florida, ACM, pp. 238-248, 1985.

[Kobsa & al. 2001] Kobsa, A., Koenemann J. & Pohl W. (2001). *Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships*, (pp. 111-155). *The Knowledge Engineering Review*, 16(2).

[Kolski & al. 1992] Kolski C., Tendjaoui, M., Millot P., (1992). *A process method for the design of "intelligent" man-machine interfaces: case study: "the Decisional Module of Imagery"*. *International Journal of Human Factors in Manufacturing*, 2, pp. 155-175.

[Laforest 2007] Laforest, F., *De l'adaptation à la prise en compte du contexte. Une contribution aux systèmes d'information pervasifs*, HDR, Université de Lyon, 2007.

[Larson 2005] Larson J.A., *VoiceXML on Steroids*, *Speech Technology Magazine*, 2005. <http://www.speechtechmag.com/Articles/Column/Forward-Thinking/VoiceXML-on-Steroids-29935.aspx>

[Lerlerdthaiyanupap 2008] Lerlerdthaiyanupap, Thanyaphorn, *Speech based dictionary application*, M.Sc. Thesis, Université de Tampere, 2008 (<http://tutkielmat.uta.fi/pdf/gradu03206.pdf>)

[Lie & Bos 1999] Lie H. W., Bos B., *Cascading Style Sheets, Designing for the Web*. Addison Wesley, 2nd edition, 1999, ISBN 0-201-59625-3.

[Lin & al. 2001] Lin J., Sinha A., Landay J., *Universal Access Through Multimodal Applications*, in *Workshop on Transforming the UI for Anyone, Anywhere, CHI 2001, Workshop "Transforming the UI for anyone anywhere"*, Seattle, WA, April 2001.

[Limbourg & al. 2004] Limbourg, Q., Vanderdonckt, J., Michotte, B., Bouillon, L., Florins, M., Trevisan, D., *UsiXML: A User Interface Description Language for Context-Sensitive User Interfaces*, in *Proceedings of the ACM AVI'2004 Workshop "Developing User Interfaces with XML: Advances on User Interface Description Languages"* (Gallipoli, May 25, 2004), K. Luyten, M. Abrams, Q. Limbourg, J. Vanderdonckt (Eds.), Gallipoli, pp. 55-62, 2004.

[Lopez & Szekely 2001] Lopez J., Szekely P., *Automatic web page adaptation*, *CHI2001-Workshop "Transforming the UI for anyone anywhere"*, Seattle, April 2001.

- [López-Jaquero & al. 2008] López-Jaquero, V., Vanderdonckt, J., Montero, F., González, P., Towards an Extended Model of User Interface Adaptation: the ISATINE framework, Proc. of Engineering Interactive Systems 2007 (IFIP WG2.7/13.4 10th Conference on Engineering Human Computer Interaction jointly organized with IFIP WG 13.2 1st Conference on Human Centred Software Engineering and DSVIS - 14th Conference on Design Specification and Verification of Interactive Systems) EIS'2007 (Salamanca, 22-24 March 2007), J. Gulliksen, M.B. Harning, Ph. Palanque (eds.), Lecture Notes in Computer Science, Vol. 4940, Springer-Verlag, Berlin, 2008, pp. 374-392.
- [Malinowski 1992] Malinowski, U., Kuhme, T., Dietrich, H., Schneider-Hufschmidt, M. (1992). A taxonomy of adaptive user interfaces. In Monk, Diaper, Harrison, editors, *People and Computers VII*, (pp. 391-414), Cambridge University Press.
- [Martin 2004] Martin, J.-C., Introduction aux Interfaces Homme-Machine Multimodales. Conférence invitée, Actes des entretiens de l'Institut de Garches. Innovations technologiques et Handicap, Issy-les-Moulineaux, 2004
- [Martin 2006] Martin, J.-C., Multimodal human-computer interfaces and individual differences: Perception, representation and generation of situated multimodal behaviours, HDR, Université Paris XI, 2006.
- [McKinley & al. 2004] McKinley, Philip K., Sadjadi, Seyed Masoud, Kasten, Eric P., Cheng, Betty H. C., Composing Adaptive Software, *IEEE Computer* 37(7): 56-64 (2004).
- [Merrick 2001] Merrick, R. A., Device Independent User Interfaces in XML, *BelCHI 2001*.
- [Mitrovic & al. 04] Mitrovic, N., Royo, J.A., Mena, E., ADUS: Indirect Generation of User Interfaces on Wireless Devices, Fifteenth International Workshop on Database and Expert Systems Applications (DEXA 2004), Seventh International Workshop Mobility in Databases and Distributed Systems (MDDS 2004), Zaragoza (Spain), IEEE Computer Society, ISBN 0-7695-2195-9, ISSN 1529-4188, 2004.
- [Mittendorfer & al. 2001] Mittendorfer, M., Winiwarer, W., Niklfeld, G., Making the VoiceWeb smarter Integrating Intelligent Component Technologies and VoiceXML, Proc. WISE 2001, Kyoto, Japan
- [Mobasher & al. 2000] Mobasher B., Cooley R. & Srivastava J. (2000), *Automatic Personalization Based on Web Usage Mining*, (pp. 142-151). *Communications of the ACM*, 43(8).
- [Moisuc 2006] Moisuc, B. (2006). Adaptation dans les systèmes d'information spatio-temporelle interactifs, *INFORSID 2006*, 1er-3 juin, Hammamet, Tunisie.
- [Morch 1995] Morch, A. (1995). Three levels of end-user tailoring: customization, integration, and extension. *Third Decennial Aarhus Conference*, (pp 41-51). Aarhus, Denmark.
- [Morch & Mehandjiev 2000] Morch, A., & Mehandjiev, N. D. (2000). Tailoring as Collaboration: The Mediating Role of Multiple Representation and Application Units. *CSCW'2000*, (pp. 75-100). Philadelphia, Pennsylvania, USA.
- [Müller & al. 2000] Müller H., Jahnke J., Smith D., Storey M-A., Tilley S., Wong K., Reverse Engineering: A Roadmap. In 22nd International Conference on Software Engineering, Proceedings of the Conference on The Future of Software Engineering, Limerick, Ireland, 2000, pp. 47 – 60.
- [Müller & al. 2001] Müller A., Forbrig P., Cap C., Using XML for Model-based User Interface Design. CHI2001-Workshop "Transforming the UI for anyone anywhere", Seattle, 2001.
- [Myers & al. 2002] Myers, B., Hudson, S., Pausch, R. Past, Present, Future of User Interface Tools. *Transactions on Computer-Human Interaction*, ACM, 7(1), 2000, pp. 3-28.
- [Nanard & Nanard 2001] Nanard J., Nanard M., "Principes hypermédias et systèmes d'information", In *Environnements Evolués et Evaluation des IHM, Interface Homme-Machine pour les Systèmes d'Information (Vol. 2)*, C. Kolski (Dir.), Hermès, ISBN : 2-7462-0243-3, pp. 147-173, 2001.

[Neal & Shapiro 1990] Neal, J. G., Shapiro, S. C., Intelligent Multi-Media Interface Technology, in Intelligent User Interfaces, J. W. Sullivan and S. W. Tyler, eds., ACM Press, 1990.

[Nigay & Coutaz 1993] Nigay, L. and Coutaz, J., A design space for multimodal systems: Concurrent processing and data fusion. In Proceedings of ACM INTERCHI'93 Conference on Human Factors in Computing Systems, Voices and Faces, pages 172-178, 1993.

[Nigay 1994] Nigay, L., Conception et modélisation logicielles des systèmes interactifs. Thèse de l'Université Joseph Fourier, Grenoble, 1994.

[Nigay & Coutaz 1996] Nigay, L. & Coutaz, J. (1996). "Espaces conceptuels pour l'interaction multimédia et multimodale". *Technique et science informatiques (TSI)*, vol. 15, 9. pp 1195-1225.

[Nigay & Coutaz 1997] Nigay, L., Coutaz, J. Multifeature Systems: The CARE Properties and Their Impact on Software Design. *Intelligence and Multimodality in Multimedia Interfaces*, AAAI Press (1997).

[Nigay 2001] Nigay, L., Modalité d'interaction et multimodalité, HDR, Université Joseph Fourier, Grenoble, 2001.

[Norman & Draper 1986] Norman D.A , Draper, S., User Centered System Design : New Perspectives on Human-Computer Interaction, Lawrence Erlbaum Associates, 1986.

[O'Hara & al. 2007] O'Hara, Kenton; Kindberg, Tim; Glancy, Maxine; Baptista, Luciana; Sukumaran, Byju; Kahana, Gil; Rowbotham, Julie, Collecting and Sharing Location-based Content on Mobile Phones in a Zoo Visitor Experience, *Computer Supported Cooperative Work (CSCW)*, Springer, Volume 16, Numbers 1-2, April 2007 , pp. 11-44(34).

[OMA 2008] OMA : Open Mobile Alliance, OMA Multimodal and Multi-device V1.0, Requirement, 2008. http://www.openmobilealliance.org/Technical/release_program/mmmd_v1_0.aspx

[OpenInterface] OpenInterface European project. IST Framework 6 STREP funded by the European Commission (FP6- 35182). <Http://www.oiproject.org>.

[Oppermann 1994] Oppermann, R. (1994). Adaptively supported adaptability, *International Journal of Human-Computer Studies*, Volume 40, Issue 3 (pp. 455 - 472). Academic Press, Inc. Duluth, MN, USA.

[Oreizy & al. 2004] Oreizy, P., Gorlick, M.M., Taylor, R., N, Heimbigner, D., Johnson, G., Medvidovic, N., Quilici, David, A., Rosenblum, S. & Wolf. A., L. (1999). An Architecture-Based Approach to Self-Adaptive Software, *IEEE Intelligent Systems*, vol. 14, no. (pp. 54-62).

[Oviatt 1999] Oviatt, S. L. Ten myths of multimodal interaction, *Communications of the ACM*, 42, 11, (1999), 74-81

[Oviatt & Cohen 1992] Oviatt, S. L. & Cohen, P. R. (1992). Spoken language in interpreted telephone dialogues. *Computer Speech and Language*, 6 (3) 277-302.

[Paramythis 2004] Paramythis, A. (2004). Towards Self-Regulating Adaptive Systems. In Weibelzahl, S., & Henze, N. (Eds.), *Proceedings of the Annual Workshop of the SIG Adaptivity and User Modeling in Interactive Systems* of the German Informatics Society (ABIS04), (pp. 57-63), Berlin,.

[Paternò & Santoro 2002] Paternò, F., Santoro, C, One Model, Many Interfaces, CADUI'02, Valenciennes, 2002.

[Paternò 2004] Fabio Paternò, Multimodality and Multiplatform Interactive Systems. Proceedings of WCC'2004, the 18th IFIP (International Federation for Information Processing) World Computer Congress, Toulouse, August 2004, Kluwer Academic Publishers, René Jacquart (ed.), "Building the Information Society", pp. 421-426, ISBN:1 4020 8156 1

- [Paternò 2004b] Fabio Paternò, Multimodality and Multi-device interfaces, W3C Workshop on Multimodal Interaction, Sophia Antipolis, France, 2004.
<http://www.w3.org/2004/02/mmi-workshop/paterno-cnr.pdf>
- [Paternò 2004c] Fabio Paternò, Multimodality and Multi-Platform Interactive Systems, Book Series IFIP International Federation for Information Processing, Springer, Boston, ISSN 1571-5736 (Print) ,1861-2288 (Online), Volume 156/2004, Building the Information Society, ISBN 978-1-4020-8156-9, pp. 421-426, Collection Computer Science, 2004.
- [Paternò & al. 2008] F. Paternò, C.Santoro, J.Mäntyjärvi, G.Mori, S.Sansone, Authoring pervasive multimodal user interfaces, in International Journal of Web Engineering and Technology 2008, Vol. 4, pp. 235-261, Inderscience Enterprises Ltd., 2008.
- [Perrault 2008] Perrault, Simon, Modélisation et implémentation d'un assistant vocal interactif pervasif, Mémoire de master Recherche, Université des Sciences et Technologie de Lille, 2008.
- [Rabardel 1995] Rabardel P., Les hommes et les technologies, approche cognitive des instruments contemporains, Armand Colin, 1995.
- [Raggett & al. 1999] Raggett D., Le Hors A., Jacobs I., HyperText Markup Language (HTML) 4.01 Specification. W3C Recommendation, December 1999. (<http://www.w3.org/TR/html4/>)
- [Ramel 2006] Ramel, J-Y. (2006), *Propositions pour la représentation et l'analyse de documents numériques*, HDR, Université François Rabelais De Tours, France.
- [Rosenberg 2001] Rosenberg M., *The personalization story*, ITworld.com, 11th may 2001, <http://www.itworld.com/Man/2676/ITW010511rosenberg/>
- [Rossi & al. 2001] Rossi G., Schwabe D. & Guimarães R. (2001), Designing Personalized Web Applications, 10th International World Wide Web Conference (WWW10), Hong Kong.
- [Roudaut & Lecolinet 2007] Roudaut A. et Lecolinet E., Un espace de classification pour l'interaction sur dispositifs mobiles. IHM'07, Paris, 2007
- [RTP14 2004] CNRS-STIC AS 49 Interaction Gestuelle, Rapport final, RTP 14, « Communication et Dialogue », Action Spécifique n° 49, Interaction Gestuelle, Etude et développement de la communication Homme-Machine visuo-gestuelle, Rapport de fin d'action, 2004.
- [Samaan 2006] Samaan, K. (2006), Prise en compte du modèle d'interaction dans le processus de construction et d'adaptation d'applications interactives, Thèse de doctorat d'informatique, Ecole Centrale de Lyon, France.
- [Scapin & Bastien 1997] Scapin, D.L. & Bastien, J.M.C. (1997). Ergonomic criteria for evaluating the ergonomic quality of interactive systems (pp. 220-231). *Behaviour & Information Technology*, 6(4-5).
- [Schaefer & Mueller 2008] Schaefer, Robbie, Mueller, Wolfgang, Assessment of a multimodal interaction and rendering system against established design principles, Journal on Multimodal User Interfaces, Springer Berlin / Heidelberg, ISSN 1783-7677 (Print) 1783-8738 (Online), Volume 2, Number 1, pp. 25-41, 2008
- [Serrano & al. 2008] M. Serrano, D. Juras, M. Ortega, L. Nigay, OIDE: un outil pour la conception et le développement d'interfaces multimodales, Quatrième Journées Francophones : Mobilité et Ubiquité 2008, Saint Malo, France, 28-30 mai 2008, ACM Press, pp. 91-92.
- [Spriestersbach & al. 2003] Spriestersbach, A., Ziegert, T., Grassel, G., Wasmund, M., Dermler, G., A single source authoring language to enhance the access from mobile devices to Web enterprise applications, WWW2003 Developers Day Mobile Web Track, 12th World Wide Web Conference WWW'03, Budapest, 2003.

[Stanciulescu 2008] Stanciulescu, Adrian, A Methodology for Developing Multimodal User Interfaces of Information System, Ph.D. thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgium, 2008.

[Stephanidis & al. 1998] Stephanidis, C, Paramythis, A, Sfyarakis, M, Stergiou, A, Maou, N, Leventis, A, Paparoulis, G & Karagiandidis, C. (1998), Adaptable and adaptive user interfaces for disabled users in AVANTI Project in Triglia, S., Mullery, A., Campolargo, M., Vanderstraeten, H. and Mampaey, M. (Eds) *Intelligence in Services and Networks: Technology for Ubiquitous Telecom Services*, (pp. 153–166), Springer.

[Tarby 2004] Tarby, J.-C., One Goal, Many Tasks, Many Devices: From Abstract User Task Specification to User Interfaces, in *The Handbook of Task Analysis for Human-Computer Interaction*, Dan Diaper (ed.) and Neville Stanton (ed.), Lawrence Erlbaum Associates, 2004, pp. 531-550.

[Tarpin-Bernard 2006] Tarpin-Bernard, F., *Interaction Homme-Machine Adaptative*, HDR, ICTT, Lyon, France

[Thevenin & Coutaz 1999] Thevenin, D. & Coutaz, J. (1999) Plasticity of User Interfaces: Framework and Research Agenda. Proceeding of IFIP *Conference on Human Computer Interaction*, INTERACT.

[Thevenin 2001] Thevenin, D. (2001) *Adaptation en Interaction Homme-Machine : le cas de la Plasticité*. Thèse de doctorat d'informatique de l'université Joseph-Fourier Grenoble I.

[Totterdell & Rautenbach 1990] Totterdell, P., Rautenbach, P., (1990). Adaptation as a Problem of Design. In: Browne, D., Totterdell, P., Norman, M., (Eds.): *Adaptive User Interfaces*. Academic say, Press, London, (pp. 59-84).

[Trigg & al. 1987] Trigg, R., Moran, T. & Halasz, F. (1987). Adaptability and Tailorability in NoteCards, in Bullinger and Shackel (Eds.) Proc. INTERACT '87, North-Holland.

[Truillet 1999] Truillet P., Modélisation de coopérations intermodales : application à l'interaction non visuelle, thèse de doctorat en informatique, université Paul Sabatier, Toulouse, 1999.

[UIML] <http://www.uiml.org>

[UsiXML] USeR Interface eXtensible Markup Language, <http://www.usixml.org>

[Van den Bergh & al. 2004] Van den Bergh J., Coninx K., Model-Based Design of Context- Sensitive Interactive Applications: a Discussion of Notations. In Proceedings of 3rd International Workshop on TAsk MOdels and DIAGrams for user interface design TAMODIA'2004. Prague, Czech Republic, November, ACM 2004, pp. 43-50.

[Vanderdonckt & al. 2004] Vanderdonckt, J., Limbourg, Q., Michotte, B., Bouillon, L., Trevisan, D., Florins, M., UsiXML: a User Interface Description Language for Specifying Multimodal User Interfaces, in Proc. of W3C Workshop on Multimodal Interaction WMI'2004 (Sophia Antipolis, 19-20 July 2004).

[Vanderdonckt & al. 2005] Vanderdonckt, J., Grolaux, D., Van Roy, P., Limbourg, Q., Macq, B., Michel, 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 (Toronto, 20-22 July 2005), International Society for Computers and their Applications, Toronto, 2005, pp. 207-214.

[Villanova-Oliver 2002] Villanova-Oliver, M. (2002). Adaptabilité dans les systèmes d'information sur le web : Modélisation et mise en oeuvre de l'accès progressif. Thèse de doctorat d'informatique, Institut National Polytechnique de Grenoble, France.

[VoiceXML 2.0] VoiceXML 2.0., W3C Recommendation (16/03/04) <http://www.w3.org/TR/voicexml20>

[VoiceXML 2.1] VoiceXML 2.1, Recommendation, (19/06/07) <http://www.w3.org/TR/voicexml21/>

[VoiceXML 3.0] VoiceXML 3.0, W3C Working Draft (08/08/2008),<http://www.w3.org/TR/vxml30reqs/>

[Weibelzahl 2002] Weibelzahl, S. (2002). *Evaluation of Adaptive Systems*, Graduate Programme Human and Machine Intelligence, Kunzenweg 21, University of Education Freiburg, Germany.

[Weimer & Ganapathy 1989] D. Weimer et S. K. Ganapathy, A synthetic visual environment with hand gesturing and voice input, Proceedings of the SIGCHI conference on Human factors in computing systems, pages 235–240. ACM Press, 1989.

[Weiss 2002] Weiss, S., *Handheld Usability*, John Wiley & Sons, 2002.

[XIML] <http://www.ximl.org>

[XML] <http://www.w3.org/XML>

[XSL] <http://www.w3.org/Style/XSL>

[X+V] XHTML+Voice Profile 1.0, (W3C Note 21 December 2001). <http://www.w3.org/TR/xhtml+voice/>

[Yang & al. 1998] Jie Yang, Rainer Stiefelhagen, Uwe Meier, et Alex Waibel, Visual tracking for multimodal human computer interaction, Proceedings of the Conference on Human Factors in Computing Systems (CHI-98) : Making the Impossible Possible, pages 140–147, New York, avril 18–23, 1998. ACM Press.

5 Annexes

5.1 Le jeu *Active Life: Athletic World* sur console Nintendo Wii



Figure 40 : Utilisation mono-utilisateur du tapis interactif pour console Nintendo Wii



Figure 41 : Utilisation multi-utilisateur du tapis interactif pour console Nintendo Wii

(Source : <http://www.gameguru.in/nintendo-wii/2007/21/active-life-athletic-world-for-nintendo-wii-revealed-by-namco-bandai/>)

5.2 Les composants de l'objet communicant Chumby

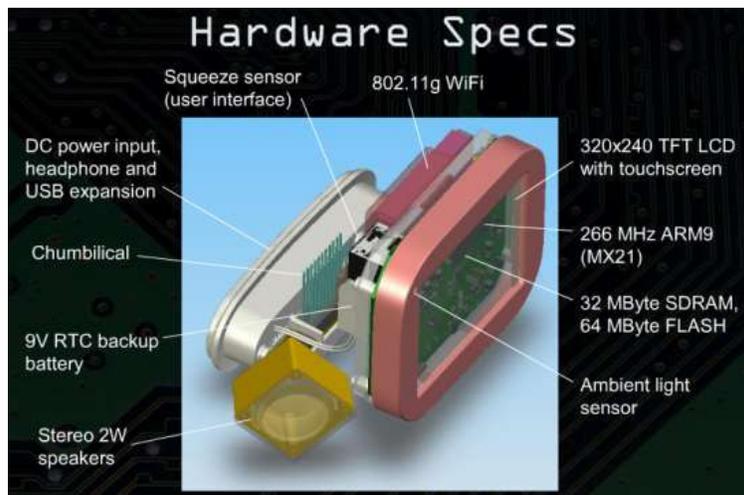


Figure 42 : Les composants du «Chumby»
(source : image extraite de la présentation de Bunnie Huang à Toorcon 2006)

5.3 Prototype de machine à laver parlante

Voici quelques illustrations de la machine à laver parlante de la société Santiago :



Figure 43 : La machine à laver interroge oralement l'utilisateur à propos du type de linge à laver



Figure 44 : L'utilisateur répond vocalement à la machine à laver



Figure 45 : La machine à laver guide l'utilisateur (ici elle lui demande de fermer le hublot) afin de continuer le processus

5.4 La boîte à outil du PlasticML

La Figure 46 présente la boîte à outil qui accompagne Plastic ML, et qui permet à un concepteur de générer, de manière relativement guidée l'interface abstraite de son application.

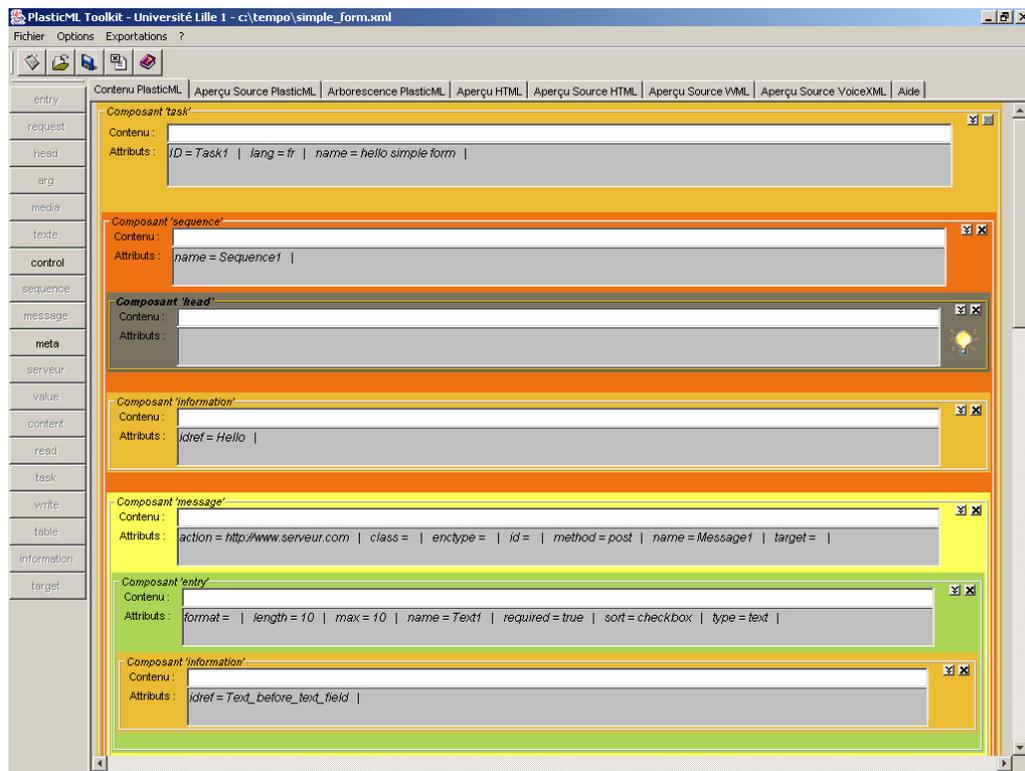


Figure 46 : La boîte à outil permettant de développer une interface en Plastic ML

5.5 Adaptation d'un document pédagogique

Notre approche, vis-à-vis de la plasticité des documents numériques, consiste à identifier et à séparer strictement les éléments de forme des éléments de fond. La Figure 47 ci-après est un exemple de document pédagogique décrit grâce au langage PlasticML. Il s'agit de la poésie « Le dormeur du Val » d'Arthur Rimbaud. Dans cet exemple, nous présentons, certes la poésie, mais également d'autres éléments descriptifs qui permettront une mise en forme ultérieure, lors de la prise de connaissance du document par l'utilisateur final (visualisation, écoute, transformation en alphabet braille, etc.).

Cependant, ces éléments descriptifs doivent être d'un niveau d'abstraction relativement élevé, afin de ne pas se focaliser, lors de la création du document, sur un mode privilégié d'accès à l'information (graphique au détriment des autres modes, le plus souvent).

```
<?xml version="1.0" encoding="iso-8859-1"?>
<plasticML version="2.0">

  <interface title="Le Dormeur du Val">

    <block role="ALL" align="center" hierarchy="NO" line_return="NO">
      <output>
        <view src="dormeur.jpg" alt="image du dormeur du val"
          width="400" height="213" />
      </output>
    </block>

    <block role="ALL" align="center" hierarchy="NO" line_return="YES">
      <output>C'est un trou de verdure où chante une
        <important1>rivière</important1>,
      </output>
    </block>

    <block role="ALL" align="center" hierarchy="NO" line_return="NO">
      <output>Accrochant follement aux herbes des haillons</output>
    </block>

    (...)
    <block role="ALL" align="center" hierarchy="NO" line_return="YES">
      <output>Tranquille. Il a deux trous
        <important2>rouges</important2> au côté droit.
      </output>
    </block>

    <block role="ALL" align="center" hierarchy="NO">
      <output>
        <view src="Arthur_Rimbaud.gif" alt="Portrait Arthur Rimbaud"
          width="99" height="143" />
        <link url="http://www.poetes.com/rimbaud/">
          Arthur Rimbaud
        </link>
      </output>
    </block>

  </interface>
</plasticML>
```

Figure 47 : Extrait d'un document PlasticML : la poésie « Le dormeur du Val » d'Arthur Rimbaud

Ainsi, dans notre exemple, chaque strophe est représentée par un bloc (ce qui est une représentation possible parmi d'autres). Une fois le document décrit en PlasticML, il est alors possible de mettre en forme le document final, au moment de sa consultation par l'utilisateur (c'est-à-dire dans le contexte d'usage), en l'adaptant dynamiquement selon les terminaux d'accès à l'information disponibles (contexte d'utilisation).

La Figure 48, ci-après, présente le document numérique obtenu, à partir du document PlasticML (Figure 47), en vue d'une visualisation par l'utilisateur sur un navigateur graphique pour le Web. Les mots « rivière » et « rouges » qui avaient été encapsulés (respectivement) par les balises⁷⁶ <important1> et <important2>, dans le document de description abstraite de l'interface, se trouvent mis en forme (italique, gras, etc.) au moment de la génération en HTML. La photo de l'auteur du poème est affichée, ainsi qu'un hyperlien pointant vers un site consacré à Rimbaud.

⁷⁶ Ces balises ne sont pas hiérarchisées. Autrement dit, <importantN> n'est pas considérée comme plus importante que <importantN-1>.

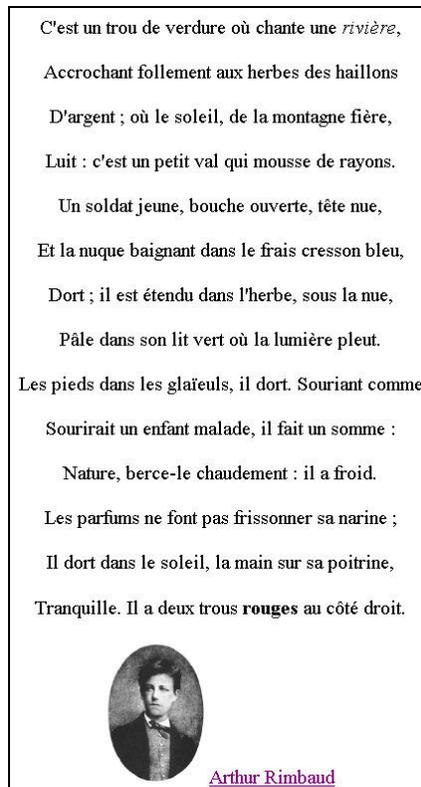


Figure 48 : Adaptation du document numérique en vue d'une visualisation par l'utilisateur via un navigateur graphique pour le Web

La Figure 49, ci-après, présente les documents numériques obtenus, à partir du document PlasticML de la Figure 47, lorsque l'adaptation se fait vers un téléphone supportant le WAP. Les mots « rivière » et « rouges » ont été respectivement mis en forme en italique et gras, comme pour la version Web. Mais ici, les strophes du poème sont visibles bloc par bloc (notion de card du WML), et l'utilisateur doit manipuler les touches de navigation de son téléphone pour prendre connaissance du poème dans son intégralité. La photo de l'auteur n'est pas présentée dans la version WAP (bien que la transformation de l'image vers le format wbmp soit techniquement possible).



Figure 49 : Adaptation du document numérique en vue d'une visualisation par l'utilisateur via un navigateur graphique pour le WAP

5.6 Utilisation de code barres 2D

Voici quelques illustrations d'utilisation de code barre 2D (Datamatrix) :



Figure 50 : Code barre 2D installé au zoo de Londres (Datamatrix grande taille)



Figure 51 : Code barre 2D installé au zoo de Londres (Datamatrix petite taille)



Figure 52 : Datamatrix permettant de composer un numéro de téléphone automatiquement (BBC)

5.7 Espace de classification de Roudaut & Lecolinet

La Figure 53 est tirée de l'article [Roudaut & Lecolinet 2007]. Elle présente un espace de classification des techniques d'interaction en entrée et en sortie sur dispositifs mobiles.

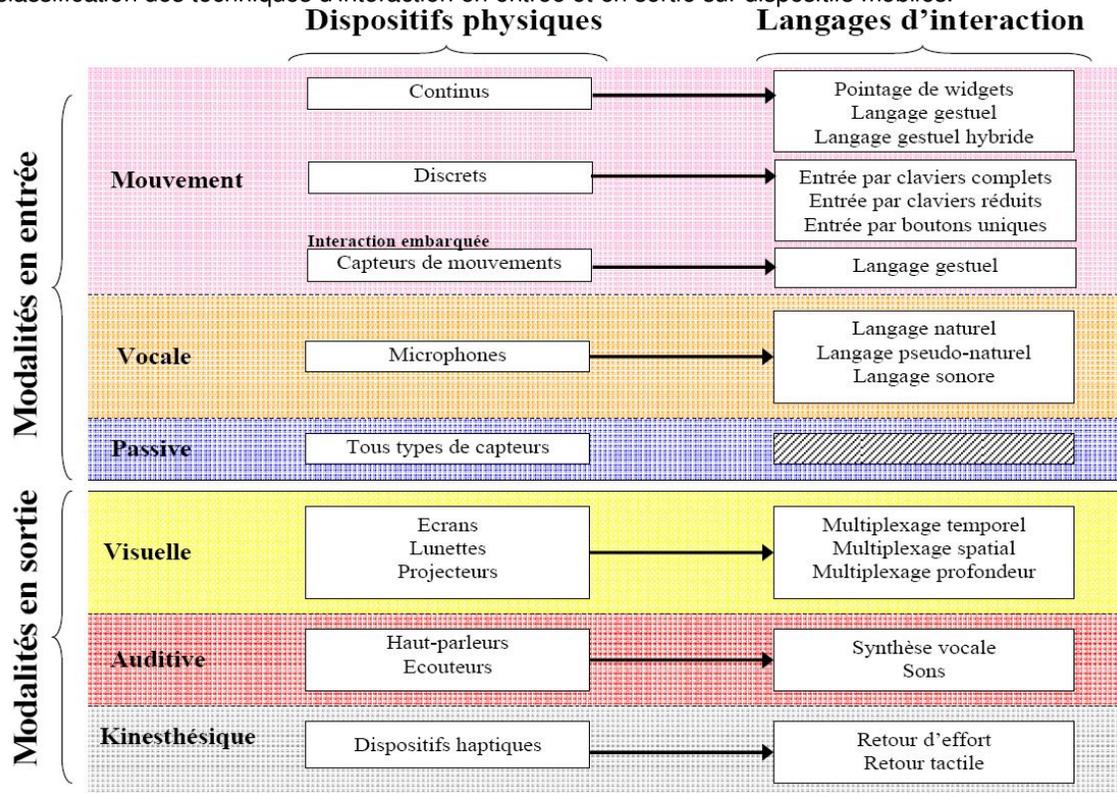


Figure 53 : Espace de classification de [Roudaut & Lecolinet 2007]

5.8 Reconnaissance d'écriture avec InkML

La Figure 54 montre l'utilisation de langage InkML pour la reconnaissance d'écriture manuscrite dans un système multimodal d'enseignement à distance.

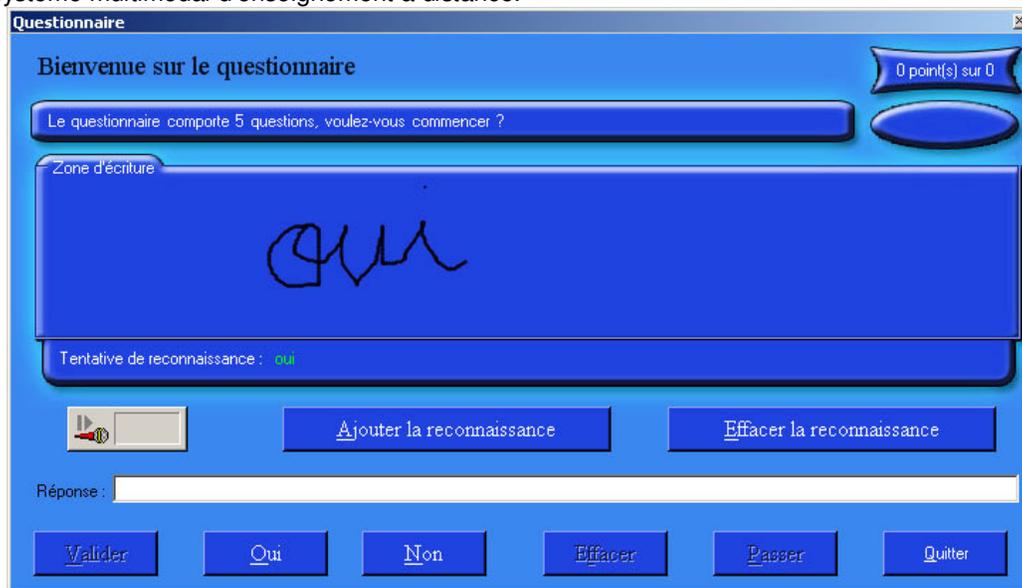


Figure 54 : Utilisation de la reconnaissance d'écriture manuscrite dans un système multimodal d'enseignement à distance

5.9 Etude et recueil de corpus d'utilisation de VoiceXML

Dans l'étude que nous présentons ci-après, l'objectif était double :

1) Nos voulions, dans un premier temps mettre en place un système de DHM permettant d'obtenir de manière semi-automatique des réponses de la part des personnes contactées, afin de livrer les réponses obtenues (ainsi les statistiques qui les accompagnent) à la région Nord Pas-de-Calais, commanditaire de cette étude. L'ensemble des questions et des réponses possibles était connu à l'avance. Les questions posées n'étaient pas forcément les mêmes pour tous les personnes, car l'ordre et l'enchaînement des questions dépendait des réponses aux questions précédentes. L'automate suivait donc un algorithme particulier, pour ne poser que les questions adéquates à la situation. Par exemple, il ne fallait pas poser la question « Suivez-vous vos cours en école ou en université ? » si la personne venait juste de répondre, à la question précédente, qu'elle avait arrêté ses études.

L'ensemble des réponses possibles était également connu à l'avance, puisque sans cela, il ne serait quasiment pas possible de récolter les réponses des personnes interrogées. En effet, en VoiceXML, l'utilisateur peut répondre à une question de deux manières possibles : soit en prononçant un mot ou une phrase prévue par le concepteur du système (on parle alors de grammaire vocale), soit en utilisant les touches de son clavier téléphoniques (on parle alors de grammaire DTMF77).

Dans un cas comme dans l'autre, si l'on n'indique pas à l'avance à la machine ce que l'utilisateur est susceptible de répondre, la poursuite du dialogue est difficilement réalisable, sauf si l'on enregistre ce que dit l'utilisateur, et qu'on le soumet à un système de reconnaissance vocale indépendant de celui utilisé sur la plate-forme VoiceXML. Dans d'autres travaux, nous avons étudié cette dernière approche [Rouillard & Truillet 2005], mais les temps de réponses ainsi que les résultats obtenus sont encore insuffisants pour pouvoir réellement utiliser des systèmes se passant totalement de grammaires vocales.

2) Dans un second temps, nous souhaitions étudier le corpus de données ainsi recueilli et l'analyser pour valider des hypothèses de travail. Nous formulons les hypothèses suivantes :

H1 : l'interaction vocale sera plus utilisée que la manipulation directe, lorsque les deux usages seront possibles, car selon la littérature scientifique relative à la multimodalité, ce mode est en général, plus naturel, plus facile à utiliser et reste moins contraignant que tout autre mode [Oviatt & Cohen 1992].

H2 : le type de téléphone (fixe ou portable) qui sera utilisé par les personnes appelées n'aura pas d'influence sur la qualité audio de la communication, et plus particulièrement sur la qualité de la reconnaissance vocale.

H3 : le type de synthèse vocale générée (Homme versus Femme) n'influencera pas les résultats obtenus. C'est-à-dire que nous envisageons le même taux de « bonnes » interactions, quel que soit le type de voix utilisé pour la génération des phrases que devra prononcer le serveur vocal en TTS (Text To Speech).

- Le corpus Questions/ Réponses SEC

Le serveur vocal utilisé pour le recueil du corpus SEC (Suivi des Etudiants du CUEEP) était le système Phonic, d'Idylic⁷⁸, supportant le langage VoiceXML version 1.0, avec une reconnaissance vocale de Philsoft de *Telisma* et une synthèse vocale *Tempo* d'*Elan Informatique*, et DTMF (touches du téléphone). Trois pré-tests ont été effectués afin de vérifier le bon fonctionnement du matériel et des logiciels. Cela portait notamment sur la génération automatique du fichier VoiceXML contenant les informations relatives à la personne à contacter, l'aboutement⁷⁹ téléphonique vers des téléphones fixes et mobiles, l'enregistrement audio des conversations Homme-Machine, et la sauvegarde des réponses obtenues dans la base de données.

Le corpus a été recueilli entre le 17/01/2006 et le 10/02/2006. Le nombre de personnes à contacter dans le fichier était de 39. Le nombre de personnes effectivement contactées est de 23 (58,97 %). Sur les 16 autres contacts, 5 étaient absents à chaque tentative d'appel, 4 étaient de faux numéros (soit réattribués, soit inactifs), et 7 ont commencé le questionnaire mais ne l'ont pas terminé. La part des communications effectuées vers des téléphones fixes représente 47,83 % des appels (respectivement 52,17 % vers des téléphones portables).

⁷⁷ DTMF : Dual Tone Multi Frequency

⁷⁸ <http://www.idylic.com/>

⁷⁹ Dans les spécifications du langage VoiceXML, il est possible de paramétrer le type de transfert à l'aide de l'attribut *bridge* de la balise *transfer*. S'il s'agit du mode *blind*, l'aboutement s'effectue sur le réseau de l'opérateur, tandis que s'il s'agit du mode *bridge*, l'aboutement est réalisé localement, sur le serveur vocal interactif. De plus, si l'on aboute plus de deux lignes entre elles, on obtient une conférence téléphonique.

L'enquête était donc destinée à obtenir de manière semi-automatique des données statistiques, comme par exemple, la situation professionnelle des personnes interrogées, au moment de l'appel. Les réponses attendues pour cet exemple de question, et pour lesquelles une grammaire vocale (et DTMF) avait été préparée, étaient les suivantes : *en activité, en contrat de travail, en formation, à la recherche d'un emploi.*

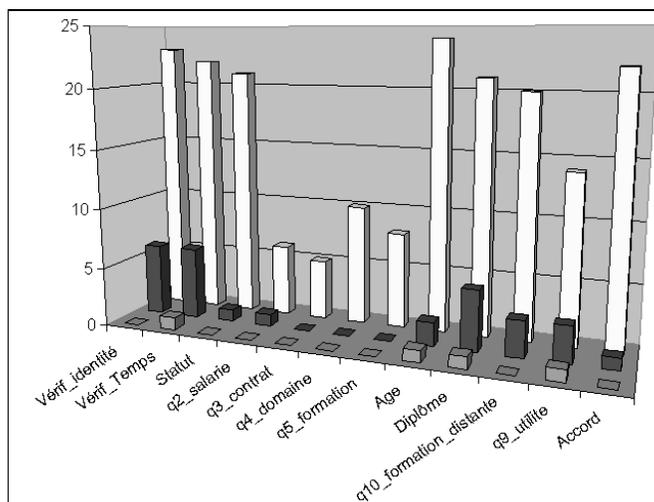


Figure 55 : Rangées V3, V2, V1, représentant les réponses vocales pour quelques questions posées par le SVI.

La Figure 55 présente le nombre d'interventions vocales pour quelques questions du corpus. La rangée la plus éloignée donne le nombre d'intervention vocale au rang numéro 1, c'est-à-dire lorsque l'utilisateur a choisi ce mode pour répondre lors de sa première intention. La rangée intermédiaire représente le rang numéro 2, et la plus proche donne le rang numéro 3. Ainsi, pour la ligne 1 (étiquette *verif_identité*), 23 fois⁸⁰ les utilisateurs ont tenté de parler pour exprimer leur première réponse à cette question ; puis 6 fois lors d'un deuxième essai, toujours pour cette même question. On note que c'est pour donner leur âge que les utilisateurs ont le plus utilisé leur voix.

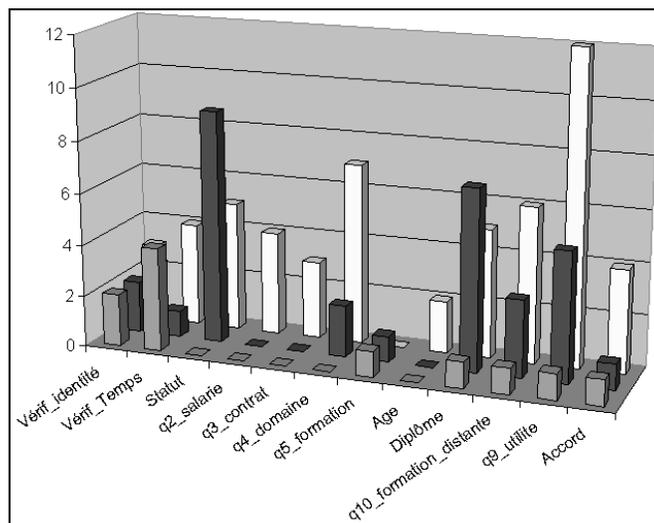


Figure 56 : Rangées D3 D2, D1, représentant les réponses par DTMF pour quelques questions posées par le SVI

Sur la Figure 56, on voit la répartition, pour les mêmes questions, mais en ce qui concerne le mode DTMF, cette fois-ci. On observe par exemple que c'est pour la question « *Q9_utilite* : Pourquoi avez

⁸⁰ Le nombre d'observations est de 26 (et non pas 23), car une personne a recommencé deux fois le processus, avant de le valider, et une autre a recommencé une fois.

vous suivi cette formation ? Satisfaction personnelle ? Poursuite d'étude ? ou prétendre à un emploi ? » que les utilisateurs ont utilisé le plus leur clavier téléphonique. La Figure 57, ci-après, résume les erreurs (rangée la plus éloignée), demande d'aide (rangée intermédiaire) et demande de répétition (rangée la plus proche). On voit donc que c'est en tentant de recueillir l'âge des utilisateurs que la machine a fait le plus d'erreur, et aussi que les utilisateurs ont demandé le plus d'aide. C'est, en revanche, à propos du diplôme obtenu « DAEU A » ou « DAEU B » que les personnes interrogées ont demandé le plus souvent à la machine de répéter sa question.

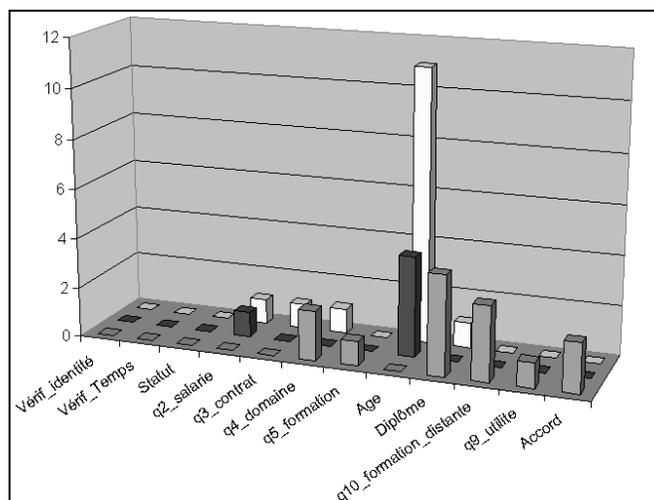


Figure 57 : Rangées Aide, Répéter et Erreur

- Le corpus audio SEC

Le corpus audio représente au total 4888 secondes d'enregistrement, soit 1h35 minutes. En moyenne chaque conversation dure 213 secondes (écart type de 90). Le dialogue le plus court dure 132 secondes, et le plus long dure 574 secondes. Ces durées ont été automatiquement calculées par l'automate qui estampillait le début et la fin de chaque conversation. Il s'agit ici des durées effectives des DHM, c'est-à-dire que l'on n'intègre pas dans ce calcul le laps de temps dans lequel l'expérimentateur explique le principe de fonctionnement à la personne appelée.

Il apparaît très nettement au vu des résultats d'interaction avec le serveur vocal, que l'hypothèse H1 que nous avons formulé était correcte, dans le cadre de cette étude. En effet, sur les 325 interactions, 228 (soit 70,15%) étaient orales, contre 97 par manipulation directe (DTMF). Plus précisément, ces 228 interactions vocales ont été réalisées en première intention à 196 fois (étiquette notée V1 dans le corpus), en deuxième intention à 28 reprises (V2), et en troisième intention à 4 reprises (V3). Pour le DTMF, les valeurs D1, D2 et D3 sont respectivement de 55, 31 et 11. Les résultats relatifs aux types de téléphones (11 téléphones fixes contre 12 téléphones mobiles dans l'étude), semblent aller dans le sens de l'hypothèse que nous avons formulé (H2) : il n'y a pas plus d'erreur avec des téléphones mobiles qu'avec des téléphones fixes, dans notre étude (respectivement 8 erreurs de compréhension, avec des téléphones fixes, corrigées après réécoute par l'expérimentateur, de manière asynchrone, contre 7 erreurs pour les portables).

- Exemples d'erreurs tirées du corpus

Voici quelques exemples, tirés du corpus SEC, que nous donnons pour illustrer les éléments que nous venons de citer, ci-dessus.

- Question : « Pourquoi avez-vous suivi cette formation ?⁸¹ satisfaction personnelle ? poursuite d'étude ? ou prétendre à un emploi ? »

- Réponse : « Au début satisfaction personnelle, et ensuite »

⁸¹ Même si la phrase n'est pas très correcte, syntaxiquement, nous y avons laissé plusieurs points d'interrogations afin d'obtenir une meilleure prosodie (ton montant pour une interrogative).

La machine n'a pas compris cette phrase, car la grammaire vocale ne s'attendait pas à une telle amorce de la part de l'utilisateur. Sans le mécanisme de trace et d'écoute discrète que nous avons mis en œuvre, il n'aurait pas été possible, après coup, de se rendre compte d'où provenait l'erreur.

- Question : « Etes-vous en activité, à la recherche d'un emploi, en formation, ou en contrat de travail ? »

- Réponse : retraitée.

Cette réponse n'ayant pas été prévue par les commanditaires de l'étude, la machine n'a pas compris cette phrase. La même question a donc été posée plusieurs fois à l'utilisateur, qui, après avoir insisté et changé de ton, a finalement décidé de raccrocher, non sans avoir prononcé une dernière phrase de mécontentement : « oh, hein, ça va bien hein ! ».

- Question : « Etes-vous en activité, à la recherche d'un emploi, en formation, ou en contrat de travail ? »

- Réponse : « Oui ».

L'utilisateur a cru qu'on lui demandait s'il était dans l'une des situations évoquées, et n'a pas compris qu'il s'agissait d'énoncer la situation précise dans laquelle il se trouvait.

- Question : « Votre formation est-elle littéraire ou technique et scientifique ? »

- Réponse : « Non »

- Réponse : « Scientifique »

Ici, le découpage demandé par le commanditaire de l'étude était d'une part « littéraire » et d'autre part « technique et scientifique ». L'utilisateur n'a pas su, à l'écoute, faire ce distinguo.

Ces quelques exemples illustrent bien le besoin de traces, d'instruments et d'outils dont les chercheurs ont besoin pour mener à bien leurs études scientifiques.

5.10 Outils non invasifs pour le BCI

La société Emotiv (emotiv.com) commercialise le casque Emotiv EPOC (cf. Figure 58). « Le casque en question, baptisé EPOC, réunit un gyroscope à son sommet (capteur de mouvement), des électrodes, et une batterie sur le côté. Les signaux du cerveau sont perçus par une technologie maintenant bien rodée : l'électroencéphalographie. ».

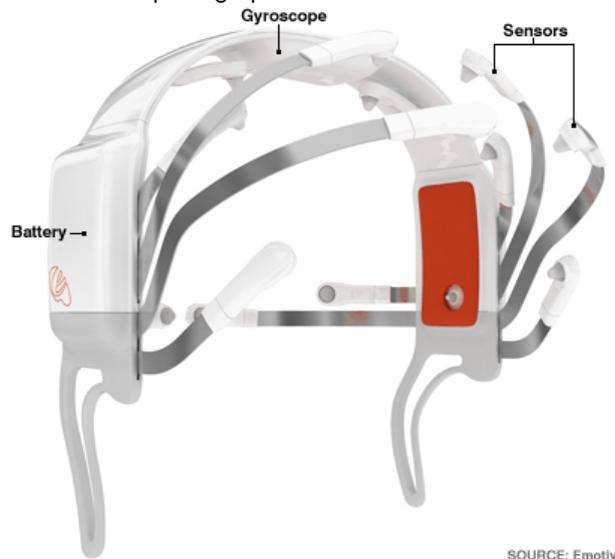


Figure 58 : Le casque EPOC de Emotiv, pour le BCI

La Figure 59 et la Figure 60 illustrent l'utilisation du logiciel Emotiv Control Panel. Il sert à détecter des signaux cérébraux ainsi que les expressions du visage de l'utilisateur.

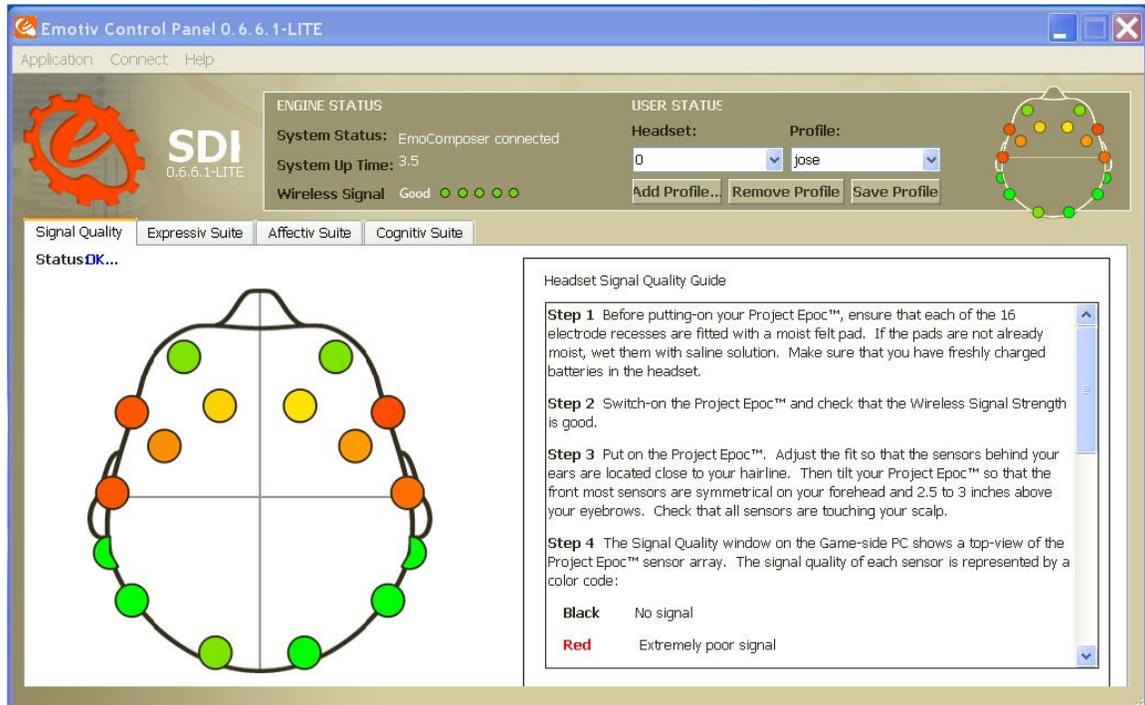


Figure 59 : Le logiciel Emotiv Control Panel (pour détecter des signaux cérébraux)

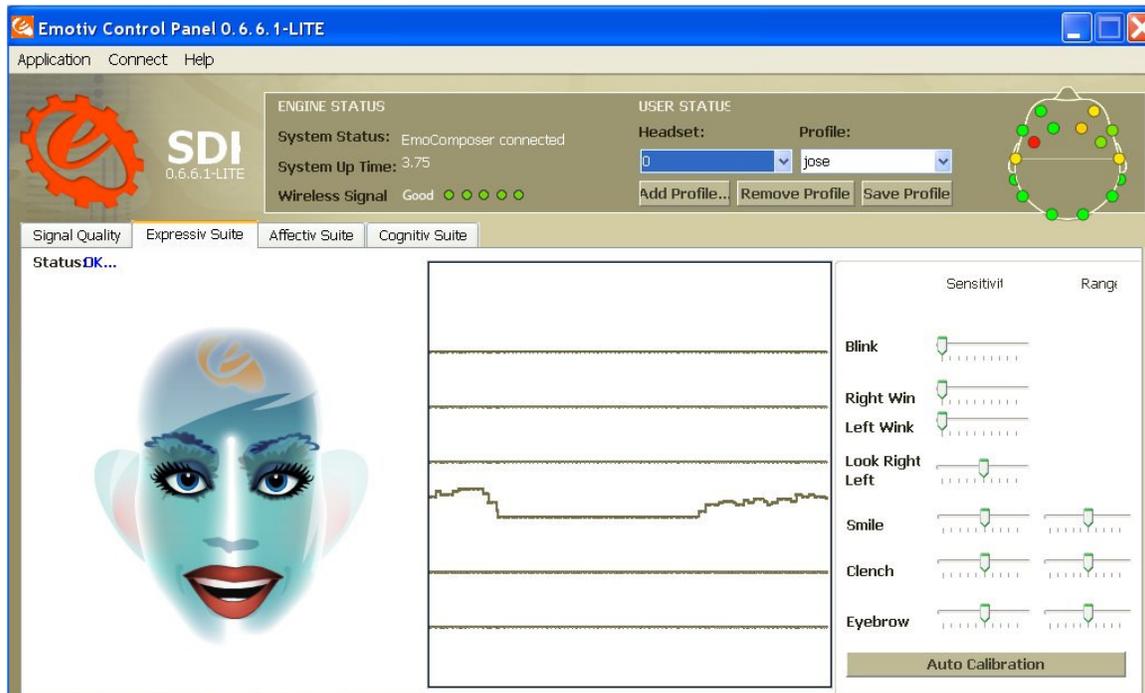


Figure 60 : Le logiciel Emotiv Control Panel (pour détecter des expression du visage)

La Figure 61 montre le logiciel EmoComposer, qui permet de simuler des comportements, lorsqu'on ne possède pas le casque EPOC (qui n'est vendu qu'aux USA pour l'instant).

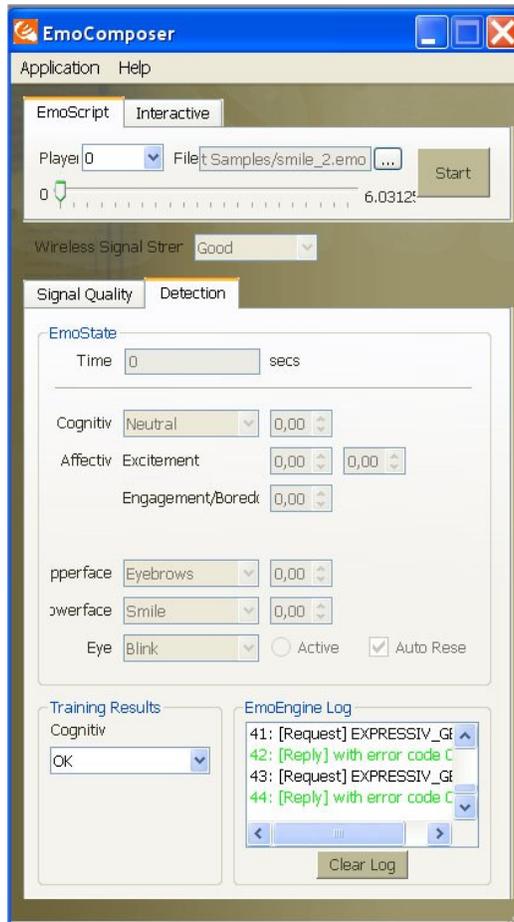


Figure 61 : Le logiciel EmoComposer, pour simuler des comportements

Un exemple de code source permettant de générer des événements avec Emotiv EmoComposer est donné ci-dessous (Figure 62).

```

<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE EML>
<EML version="1.0" language="en_US">
<sequence>
  <time value ="0s">
    <expressiv_lowerface event="smile" value="0" />
  </time>
  <time value ="0s21t">
    <signal_quality
value="1,1,1,1,0.909091,0.727273,1,1,1,1,0.909091,0.818182,1,0.818182,1,1,1,1" />
  </time>
  <time value ="0s29t">
    <expressiv_lowerface event="smile" value="0.03" />
  </time>
  <time value ="1s">
    <expressiv_lowerface event="smile" value="0.2" />
  </time>
  <time value ="1s11t">
    <expressiv_lowerface event="smile" value="0.66" />
  </time>
  <time value ="1s14t">
    <expressiv_lowerface event="smile" value="0.65" />
  </time>

```

```

    <time value = "1s17t">
      <expressiv_lowerface event="smile" value="1" />
    </time>
    <time value = "1s21t">
      <signal_quality
value="1,1,1,1,0.909091,0.545455,1,1,1,1,0.727273,0.818182,1,0.818182,1,1,1,1" />
    </time>
    <time value = "2s8t">
      <expressiv_lowerface event="smile" value="0.86" />
    </time>
    <time value = "2s13t">
      <expressiv_lowerface event="smile" value="0.5" />
    </time>
    <time value = "2s16t">
      <expressiv_lowerface event="smile" value="0.24" />
    </time>
    <time value = "2s21t">
      <expressiv_lowerface event="smile" value="0.01" />
    </time>
    <time value = "2s22t">
      <signal_quality
value="1,1,1,1,0.909091,0.727273,1,1,1,1,0.818182,0.727273,1,0.727273,1,1,1,1" />
    </time>
    <time value = "2s24t">
      <expressiv_lowerface event="smile" value="0" />
    </time>
    <time value = "3s21t">
      <signal_quality
value="1,1,1,1,0.909091,0.818182,1,1,0.909091,1,0.909091,1,1,0.818182,1,1,1,1" />
    </time>
  </sequence>
</EML>

```

Figure 62 : Code source du fichier smile2.emo pour Emotiv EmoComposer

6 Curriculum vitae des activités de recherche et d'enseignement

ROUILLARD José
Né le 23/10/70 à Cavaillon (84)
Nationalité française
Marié, un enfant

Adresse personnelle :
9, place de la solidarité
59790 Ronchin

Adresse professionnelle :
LIFL (Laboratoire d'Informatique Fondamentale de Lille)
UMR CNRS 8022
Université des Sciences et Technologies de Lille
Bâtiment M3 extension
59655 Villeneuve d'Ascq Cedex
Téléphone (bureau 234) : 03.28.77.85.80
E-mail : jose.rouillard@univ-lille1.fr

Cursus universitaire

- **Depuis 2000 : Maître de Conférences** en Informatique. Université des Sciences et Technologies de Lille (USTL). Laboratoire d'Informatique Fondamentale de Lille (LIFL). Spécialité : Interaction Homme-Machine. Nomination : 01/09/2000.
 - Actuellement en CRCT (Congé pour Recherche et Conversion Thématique), au titre de l'université USTL, pour deux semestres (2008-2009).
 - Titulaire de la PEDR (Prime d'Encadrement Doctoral et de Recherche) depuis octobre 2008.
- **1999-2000** : **ATER** en Informatique, à l'UFR de Sciences Économiques de l'Université Pierre Mendès France de Grenoble II.
- **1996-1999** : **Doctorat d'Informatique**, Laboratoire CLIPS (Communication Langagière et Interaction Personne-Système), équipe GEOD (Groupe d'Étude sur l'Oral et le Dialogue), fédération IMAG, sous la direction de Jean Caelen. Titre de la thèse : « **Hyperdialogue sur Internet. Le système HALPIN** », soutenue le 11 janvier 2000, avec la mention Très Honorable.
- **1995-1996** : DEA d'Informatique Systèmes et Communications (spécialité Interface Homme-Machine). Stage d'initiation à la recherche : « Une composante orale de dialogue pour Internet », au laboratoire CLIPS, équipe GEOD à l'IMAG de Grenoble, CNRS-INPG-UJF.
- **1994-1995** : Service National Militaire en tant que Scientifique du Contingent (12 mois) : Chercheur et développeur informatique au CATA Méditerranée - Base Aérienne d'Aix-en-Provence. Informatisation de la Division Personnel du Centre Administratif Territorial de l'Armée de l'air. Formation informatique du personnel militaire et civil de la Région Aérienne Méditerranée.
- **1992-1994** : Maîtrise MIAGE à Aix en Provence.

Activité d'enseignement et responsabilités

J'ai commencé à enseigner en janvier 1993, durant des vacances en informatique, à l'IUT d'Aix en Provence. J'ai continué à former des étudiants de différents niveaux (jeunes appelés du contingent, personnel administratif adultes) durant mon armée (1994-1995) en tant que Scientifique du Contingent : formation informatique des agents du CATA (Centre Administratif Territorial de l'Armée de l'air) lors de la phase d'informatisation des positions du personnel militaire et civil de la Région Aérienne Méditerranée.

J'ai ensuite enseigné à Grenoble, en 1997, pour des vacances à l'université Pierre Mendès-France, puis lors de mon monitorat du CIES, à partir 1998, durant ma thèse de doctorat. Les matières enseignées étaient principalement la bureautique et les SGBD (word, excel, powerpoint, Access).

Grâce à mon poste d'ATER, j'ai pu dispenser des cours d'algorithmique, de systèmes d'exploitation, de multimédia, et de bases de données. Ensuite, dans le cadre de mon poste de maître de conférences, à Lille, j'ai enseigné les interfaces homme-machine, le génie logiciel, la programmation des sites web couplés aux bases de données, sous forme de cours magistraux, de travaux dirigés et travaux pratiques, notamment au CUEEP (Centre Université-Economie d'Education Permanente) et à l'IAE (Institut d'Administration des Entreprises).

Je conduis également des enseignements plus proches de ma recherche, avec notamment des cours sur les IHM évoluées, la multimodalité, le langage VoiceXML (cours magistral et travaux pratiques en master 2 Recherche et Professionnel, école d'ingénieur, aussi bien en présentiel qu'en EAD : Enseignement à Distance).

Pour tous ces enseignements, j'ai monté entièrement ou partiellement (je me suis grandement inspiré des cours de Jean-Claude Tarby lors de ma première année en tant que maître de conférences) les supports de cours, TD et TP, ainsi que les sujets d'examen de ces formations. Le tableau ci-dessous dresse un bilan de mes enseignements, par cycles de formation.

Année \ Cycle	00-01	01-02	02-03	03-04	04-05	05-06	06-07	07-08
1 ^{er} (%)	33,51	46,07	53,67	34,23	33,09	36,91	21,56	20,54
2 ^{ème} (%)	49,21	26,18	15,60	20,72	23,53	20,04	29,05	23,06
3 ^{ème} (%)	17,28	27,75	30,73	45,05	43,38	43,05	49,39	56,40
Total	192 h	192 h	272 h	252 h	272 h	244 h	228 h	258 h

Tableau 3 : Répartition de mes enseignements

Par ailleurs, j'ai été responsable pédagogique du « DESS MICE » (Multimédia et Internet pour le Commerce Electronique) de l'université Lille1, de 2001 à 2004. Ce DESS MICE proposait une formation Bac+5 aux dimensions technologiques et socio-économiques du Commerce Electronique (Business-to-Consumer, Business-to-Business, Place de Marché Electronique...). Dans le cadre de la réforme LMD, j'ai accompagné le DESS MICE vers sa migration en « Master 2 Professionnel E-Services », qui, comme son nom l'indique, couvre un spectre plus large que le E-commerce.

Animations Scientifiques

Membre de comités de programme :

- Relecteur de chapitre pour l'ouvrage : « Multiplatform E-Learning Systems and Technologies: Mobile Devices for Ubiquitous ICT-Based Education », édité par Tiong-Thye Goh, Idea Group Inc, 2010.
- Relecteur de chapitre pour l'ouvrage : « Multimodality in Mobile Computing and Mobile Devices: Methods for Adaptable Usability », édité par Stan Kurkovsky, Idea Group Inc, 2008.
- Relecteur de chapitre pour l'ouvrage : « Intelligent User Interfaces: Adaptation and Personalization Systems and Technologies », édité par Constantinos Mourlas et Panagiotis Germanakos, Idea Group Inc, 2008.
- Relecteur pour la revue *RIHM* (Revue d'Interaction Homme-Machine)
- Relecteur et correspondant Europe pour la revue *BIAA* (Bulletin d'Informatique Approfondie et Applications)
- Responsable des articles courts pour la conférence nationale *IHM 2005*.
- Relecteur pour les conférences suivantes :

Années	Conférences Internationales	Conférences Nationales
2009	- CHI'09 (Boston) - ICCGI'09 (The Fourth International Multi-Conference on Computing in the Global Information Technology, Cap Esterel) - MobileHCI'09 (Bonn)	IHM'09 (Grenoble)
2008	- MobileHCI'08 (Amsterdam)	IHM'08 (Metz)
2007	- MobileHCI'07 (Singapore) - KCC (International Symposium on Knowledge Communication and Conferences, Orlando)	IHM'07 (Paris)
2006	- MobileHCI'06 (Espoo, Finland) - ICHSL.6 (6th IEEE International Conference on Human-System Learning, Toulouse)	IHM'06 (Montréal)
2005	- Common Innovation in E-learning, Machine Learning and Humanoid approaches (Marrakech)	IHM'05 (Toulouse)
2004		IHM'04 (Namur)
2003		- IHM'03 (Caen) - RéciTAL'03 (Rencontre des étudiants chercheurs en informatique pour le Traitement Automatique de la Langue, Batz-sur-Mer)
2002		RéciTAL'02 (Nancy)
2001	ESSLLI'01 (European Summer School in Logic, Language and Information, Helsinki, Finland).	RéciTAL'01 (Tours)

Membr e de comités d'organisation :

- Président du comité d'Organisation de la conférence Ubimob'09 (Lille, Juillet 2009).
- Responsable du Comité d'organisation RÉCITAL'00 (Lausanne).
- Responsable des étudiants volontaires pour la conférence IHM-HCI'01 (Lille).

Projet MIAOU-NIPO-TACT :

Responsable scientifique de la tâche T4 (Interaction multimodale et multicanal pour des services interactifs) du groupe MIAOU (Modèles d'Interaction et Architectures Orientées Usages) : Projet de recherche scientifique faisant suite à NIPO (Nouvelles Interactions Personnes/Organisations) en collaboration avec le laboratoire LAMIH-RAIHM de l'université de Valenciennes, et le LIFL de Lille. Nous travaillons conjointement sur les thèmes de la personnalisation des interfaces, des modèles de tâche, des évaluations et de l'ergonomie, et des nouveaux langages d'IHM. Dans ce cadre, deux conférences internationales ont été organisées, à Lille et à Valenciennes : IHM-HCI2001 et CADUI'2002 (4th International Conference on Computer-Aided Design of User Interfaces).

Relations avec le monde industriel ou socio-économique

La cité Numérique :

Partenariat industriel avec la cité Numérique, représentée par Yves Bayart, Responsable R&D du groupe 3 Suisses International. Collaboration sur un projet de personnalisation et de multimodalité, au sein de la relation avec le client. Financement d'une bourse de thèse cofinancée par la région Nord-Pas-de-Calais. Co-encadrement de ce thésard (Vincent Chevrin) pour 2002-2005.

APP-Line (<http://app-line.com>) :

Accord de partenariat et R&D avec les entreprises App-Line et Loquendo, lors des négociations pour l'acquisition d'un serveur vocal téléphonique Siblio du laboratoire TRIGONE.

Convergeonline (<http://www.convergeonline.com>) :

Partenariat et recherches scientifiques communes dans le domaine du langage VoiceXML et du SMS+. Rédaction d'un article accepté à la conférence internationale HCII 2005 : *Harnessing the power of a VoiceXML engine to SMS based HCI*, José Rouillard, Hervé Dutter, HCII 2005, Las Vegas, USA.

Rayonnement

P-LearNet : Pervasive Learning Network

Membre du projet « P-Learnnet » : infrastructures et usages pour l'apprentissage humain dans le contexte de l'informatique diffuse, qui regroupe sept partenaires, parmi lesquels certains accueillent les expérimentations des dispositifs mis en place afin de réaliser des études d'usages (Auchan, La Poste, Université Médicale Virtuelle Francophone). Ce projet a obtenu un financement dans le cadre de l'appel à projet ANR Télécoms 2006 et la labellisation du pôle de compétitivité « Industries du commerce ». Son budget global est d'environ 730k€ sur la période 2007-2009.

PICOM :

Membre du pôle de compétitivité « PICOM » : industries du commerce. Ce cluster (pôle) regroupe environ 12 acteurs du monde de la recherche, 36 entreprise de distribution (basées principales dans le Nord de la France), 27 entreprises de technologie, 6 représentants de l'Etat et des collectivités, 7 organisation professionnelles ainsi que 5 structures d'aide au développement. Contact avec des partenaires de l'axe Nouveaux Univers Sensoriels (Auchan, 3 Suisses International, La Redoute, La Poste, Finaref, Youngs, Leroy-Merlin, Devianne, Chronodrive, Ciel Innovation échange ...). <http://www.picom.fr/>

Conférence sur les SOA (Architectures Orientées Services) : animée par José ROUILLARD, Maître de Conférences en informatique, LIFL / Université de Lille 1 ; Patrick MAHU, Direction Technique SOGETI, Responsable Offre Architecture / SOA ; Patrice BERNARD, Consultant, CONIX CONSULTING et Michel PETIT, Consultant Senior, RENAULT, 28 janvier 2007, Telecom Lille 1.

Conférence publique lors du salon Vocal Expo, sur invitation d'APP-LINE et de Loquendo.

Intitulé de notre présentation commune : « Les Standards, un vrai bénéfice client ». Concorde Lafayette, Paris, 20/06/2006 (<http://www.vocalexpo.com>).

École d'été ARCO CNRS : Animateur de l'atelier scientifique « Dialogue personnes-machines », lors de École d'été ARCO CNRS, organisée par l'Association pour la Recherche Cognitive et le CNRS (<http://users.info.unicaen.fr/~beust/Atelier>), île de Tatihou, 10-18 juillet 2002.

Groupes de Travail :

Membre de l'Action Spécifique AS Plasticité des Systèmes Interactifs, du RTP 16 ; du GT sur les Agents Conversationnels Animés (ACA : <http://www.limsi.fr/aca>), du GT CESAME (Conception et Evaluation de Systèmes interactifs Adaptables et/ou MixtEs, GT 4.6 du GDR-I3, <http://www.irit.fr/CESAME>), et du GT Mobilité et Ubiquité (GT 4.2 du GdR I3, <http://iihm.imag.fr/nigay/GTMOB/>).

Projet EUCUE :

Membre du projet EUCUE : Evaluation des Usages dans des Contextes d'E-services Ubiquitaires et Evolutifs. Ce projet est une réponse à la ligne Action 1 « Nouveaux Usages » du programme recherche TAC 2004-2006, sur les Sciences et Technologies de l'information et de la Communication.

Encadrement d'étudiants en Informatique

- 1 co-encadrement de thèse (Vincent Chevrin)
- 7 encadrements de Master Recherche et DEA (Simon Perrault, Olivier Cremetz, Diane Kampire, Antoine Bouyer, Vincent Chevrin, Lina Makdessian, Maria Victoria DUQUE)
- Plus d'une trentaine d'encadrement de stage de Master 2 Professionnel et DESS.
- Tuteur de nombreux projets et de stages dans différentes formations : DEUST BCEE, projets de DESS MICE (Multimédia Internet et Commerce Électronique) en présentiel et en EAD (Enseignement A Distance), Master E-services, Master IPM (Ingénierie et Pédagogique Multimédia), Polytech'Lille...

Nom du diplômé **CHEVRIN VINCENT** Diplôme ⁽¹⁾ **THESE**
Titre du travail : L'interaction usagers/services, multimodale et multicanale : une première proposition appliquée au domaine du e-Commerce
Date début : **01/09/2002** Date fin : **07/04/2006** % encadrements : **50 %**
Noms et % des co-directeurs : **Alain Derycke (50%)**

Nom du diplômé **PERRAULT SIMON** Diplôme ⁽¹⁾ : **MASTER 2 RECHERCHE**
Titre du travail : Modélisation et implémentation d'un assistant vocal interactif pervasif
Date début : **01/01/08** Date fin : **01/07/08** % encadrements : **100 %**
Noms et % des co-directeurs :

Nom du diplômé **CREMETZ OLIVIER** Diplôme ⁽¹⁾ : **MASTER 2 RECHERCHE**
Titre du travail : Modélisation des mécanismes d'adaptation dynamique et contextuelle d'une interaction multimodale
Date début : **20/12/06** Date fin : **07/07/07** % encadrements : **33 %**
Noms et % des co-directeurs : **Alain Derycke (33 %), Vincent Chevrin (33 %)**

Nom du diplômé **DIANE KAMPIRE** Diplôme ⁽¹⁾ : **MASTER 2 RECHERCHE**
Titre du travail : Instrumentation des Suivis d'Usages dans le Contexte de l'Informatique Ubiquitaire
Date début : **20/12/05** Date fin : **07/07/06** % encadrements : **33 %**
Noms et % des co-directeurs : **Alain Derycke (33 %), Yvan Peter (33 %)**

Nom du diplômé **ANTOINE BOUYER** Diplôme ⁽¹⁾ : **MASTER 2 RECHERCHE**
Titre du travail : Plasticité des interfaces Homme-Machine et Objets Pédagogiques structurés
Date début : **20/12/04** Date fin : **01/07/05** % encadrements : **50 %**
Noms et % des co-directeurs : **Alain Derycke (50 %)**

Nom du diplômé : **CHEVRIN VINCENT** Diplôme (1) : **DEA**
Titre du travail : Multimodalité et personnalisation dans les systèmes interactifs évolutifs : adaptation dynamique au contexte de la tâche des utilisateurs.
Date début : **01/01/02** Date fin : **30/06/02** % encadrements : **50 %**
Noms et % des co-directeurs : **Alain Derycke (50%)**

Nom du diplômé : **MAKDESSIAN LINA** Diplôme (1) : **DEA**
Titre du travail : Utilisation de VoiceXML pour un Hyperdialogue sur le World Wide Web
Date début : **01/01/01** Date fin : **30/06/01** % encadrements : **50 %**
Noms et % des co-directeurs : **Alain Derycke (50%)**

Nom du diplômé : **DUQUE Maria Victoria** Diplôme (1) : **DEA**
Titre du travail : L'interaction multimodale pour la recherche d'information
Date début : **01/01/00** Date fin : **30/06/00** % encadrements : **50 %**
Noms et % des co-directeurs : **Jean Caelen (50%)**

Publications Scientifiques (vue synoptique)

Publications	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	Total Général
Revue Internationale						1						1	2	4
Revue Nationale		1		1	1	1					1		1	6
Ouvrage International														0
Chapitre d'ouvrage International													2	2
Ouvrage National									1			1		2
Chapitre d'ouvrage National							1	1						2
Conférence Internationale		1	1	6			3	2		4	4	4	4	29
Conférence Nationale		1	5	3	1		1	1	1	1	3		2	19
Communications sans actes		1	2				1		1					5
Autres (rapports, thèse, etc.)	1				1									2
Total	1	4	8	10	3	2	6	4	3	5	8	6	11	71
Encadrement doctoral														
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	Total Général
Thèse										1				1
MASTER 2 RECHERCHE (ex-DEA)					1	1	1			1	1	1	1	7

Publications Scientifiques (références)

↳ Références :

1. ROUILLARD José, Plastic interfaces for ubiquitous learning, Book chapter: Mobile Devices for Ubiquitous ICT-Based Education. Idea Group. Inc.
2. ROUILLARD José, Multimodal and Multichannel issues in pervasive and ubiquitous computing, Book chapter: Multimodality in Mobile Computing and Mobile Devices: Methods for Adaptable Usability, Idea Group. Inc.
3. ROUILLARD JOSÉ, PERRAULT Simon, Un délégué IVY favorisant les interactions multimodales et multicanales en situation de mobilité, Workshop national EMAP'08, Paris, 2008.
4. ROUILLARD JOSÉ, LAROUCSI, MONA, PerZoovasive: contextual pervasive QR codes as tool to provide an adaptive learning support, International Workshop On Context-Aware Mobile Learning - CAML'08, IEEE/ACM SIGAPP, Cergy-Pontoise - Paris, pp. 542-548, 2008.
5. ROUILLARD José, Multimodal and multichannel interactions for pervasive issues, International ETSI Workshop on Multimodal Interaction on Mobile Devices, Sophia Antipolis - Nice, 2008.
6. ROUILLARD José, Contextual QR Codes, The Third International Multi-Conference on Computing in the Global Information Technology, ICCGI 2008, Athens, Greece, pp. 50-55, 2008.
7. BOUZEGHOUB Amel, CARON Pierre-André, LECOCQ Claire, LE PALLEC Xavier, ROUILLARD José, L'informatique omniprésente pour les communautés de pratique autour de contenus pédagogiques 4èmes journées Francophones Mobilité et Ubiquité (UBIMOB 2008), ACM, Saint Malo, France, 2008.
8. KADDOUCI Sarra, BOUZEGHOUB Amel, CARON Pierre-andre, LECOCQ Claire, LE PALLEC Xavier, RITAINE François-julien, ROUILLARD José, Experiments in ubiquitous computing for communities of practice using learning resources, International Conference and Exhibition on Next Generation Mobile Applications, Services, and Technologies (NGMAST 2008), IEEE, pp. 254-259, Cardiff, Wales, UK, 2008.
9. ROUILLARD José, PETER Yvan, TARBY Jean-Claude, VANTROYS Thomas, CHEVRIN Vincent, Supporting Mobile Connectivity: from Learning Scenarios to Multichannel Devices,

International Journal of Continuing Engineering Education and Life-Long Learning (IJCEELL), Inderscience Publishers, Special Issue on Learning as a Ubiquitous and Continuous Communication Attitude, Guest Editor: Piet Kommers, Volume 18, No. 4., pp.396–410, 2008.

10. CHEVRIN Vincent, ROUILLARD José, Instrumentation and measurement of multi-channel services systems, International Journal of Internet and Enterprise Management (IJIEEM), Inderscience Publishers, Special Issue on: "Quality in Multi-Channel Services Employing Virtual Channels", Guest Editors: Dr. Rui Sousa, Catholic University of Portugal (Porto), Portugal, Volume 5 - Issue 4, pp. 333 – 352, 2008.
11. ROUILLARD José, VANTROYS Thomas, CHEVRIN Vincent, Les Architectures Orientées Services, Une approche pragmatique des SOA, Vuibert, ISBN : 2-7117-4868-5, 2007, 316 Pages, Paris. [Sélectionné pour le Prix Roberval 2008], [Finaliste pour le prix AFISI (Association Française d'Ingénierie des Systèmes d'Information) du « meilleur livre informatique de l'année 2008 ».]
12. EZZEDINE Houcine, TRABELSI Abdelwaheb, TARBY Jean-Claude, LAPORTE Philippe, ROUILLARD José, Deux approches d'évaluation précoce de systèmes interactifs : application à la supervision du transport urbain, Workshop International : Logistique and Transport 2007 (LT 2007), IEEE/SMC, 2007, Sousse, Tunisie.
13. COUTURIER Olivier, CHEVRIN Vincent, MEPHU NGUIFO Engelbert, ROUILLARD José (ordre alphabétique), Vers une convergence entre l'IHM et l'ECD basée sur les facteurs humains pour la conception de systèmes décisionnels performants, Revue RIHM - Revue d'Interaction Homme-Machine, Hermes-Lavoisier, 2008, Paris.
14. ROUILLARD José, *Web services and speech-based applications around VoiceXML*, Journal of Networks (JNW, ISSN 1796-2056), Academy Publisher, Volume 2, Issue 1, 2006, pp. 27-35, Oulu, Finland.
15. LEPREUX Sophie, HARIRI Anas, ROUILLARD José, TABARY Dimitri, TARBY Jean-Claude, KOLSKI Christophe, Towards Multimodal User Interface Composition based on UsiXML and MBD principles, HCI International 2007, Lecture Notes in Computer Science, Springer-Verlag, Berlin, Beijing, Chine.
16. TARBY Jean-Claude, EZZEDINE Houcine, ROUILLARD José, TRAN C.D., LAPORTE Philippe, KOLSKI Christophe, Traces using aspect oriented programming and interactive agent-based architecture for early usability evaluation: basic principles and comparison, HCI International 2007, Lecture Notes in Computer Science, Springer-Verlag, Berlin, Chine.
17. COUTURIER Olivier, ROUILLARD José, CHEVRIN Vincent, An interactive approach to display large sets of association rules, HCI International 2007, Lecture Notes in Computer Science, Springer-Verlag, Berlin, 2007, Beijing, Chine.
18. ROUILLARD José, CHEVRIN Vincent, TARBY Jean-Claude, DERYCKE Alain, *La plasticité des documents numériques*. Revue TSI (Technique et Science Informatique). Numéro spécial « Document numérique », sous la coordination de Rémy Mullot, n°4/2006.
19. TARBY Jean-Claude, ROUILLARD José, *Assistance, advice and guidance with digital coaching*, EAM'06 European Annual Conference on Human Decision-Making and Manual Control Lecture Notes in Computer Science, Springer-Verlag, Berlin, 2006, Valenciennes.
20. ROUILLARD José, *Web services and speech-based applications*, ICPS'06, IEEE International Conference on Pervasive Services, 2006, Lyon.
21. CHEVRIN Vincent, ROUILLARD José, DERYCKE Alain, SOCKEEL Sébastien, *Multi-channel Adaptive Context-sensitive Intermediation: Reconciling Marketing goals and Users Needs*, MAC (Multi-channel Adaptive Context-sensitive), Systems: Building Links Between Research Communities, 2006, Glasgow.

22. CHEVRIN Vincent, DERYCKE Alain, ROUILLARD José, *Project Ubi-Learn: an Intermediation Infrastructure for Multi-channel Accesses to Future LMS, ELETE (E-learning and mobile learning on telecommunications) / AICT 2006 (Advanced International Conference on Telecommunications)*, Proceedings of the Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services (AICT/ICIW 2006) 0-7695-2522-9/06, IEEE Computer Society Press, Vol n° 4, 2006, Guadeloupe, French Caribbean.
23. TARBY Jean-Claude, ROUILLARD José, *Améliorer ses performances grâce au coaching numérique mobile : l'exemple du coaching vocal embarqué*, UBIMOB 2006, 2006, Paris.
24. COUTURIER Olivier, ROUILLARD José, CHEVRIN Vincent, *Une approche hybride pour une meilleure visualisation de grands ensembles de règles d'association*, ERGO'IA 2006, 2006, Biarritz.
25. BOUYER Antoine, ROUILLARD José, DERYCKE Alain, *Pourquoi il n'est pas possible actuellement d'utiliser des objets pédagogiques plastiques respectant la norme SCORM*, IHM 2006, AFIHM, 2006, Montréal.
26. CHEVRIN Vincent, ROUILLARD José, DERYCKE Alain, *Multi-channel and multi-modal interactions in E-marketing: Toward a generic architecture for integration and experimentation*, HCI International 2005, Lawrence Erlbaum Associates, Inc., 2005, Las Vegas, Nevada, USA.
27. ROUILLARD José, TRUILLET Philippe, *Enhanced VoiceXML*, HCI International 2005, Lawrence Erlbaum Associates, Inc., 2005, Las Vegas, Nevada, USA.
28. ROUILLARD José, DUTTER Hervé, *Harnessing the power of a VoiceXML engine to SMS based HCI*, HCI International 2005, Lawrence Erlbaum Associates, Inc., 2005, Las Vegas, Nevada, USA.
29. CHEVRIN Vincent, DERYCKE Alain, ROUILLARD José, *Some issues for the Modelling of Interactive E-Services from the Customer Multi-Channel Interaction Perspectives*, IEEE 05, International IEEE international conference on e-Technology, e-Commerce and e-Service, IEEE Press, 2005, pp. 256-259, Hong Kong.
30. DERYCKE Alain, CHEVRIN Vincent, ROUILLARD José, *Intermédiations Multicanales et Multimodales pour l'E-Formation : l'Architecture du Projet Ubi-Learn*, EIAH 2005, ATIEF, 2005, Montpellier.
31. ROUILLARD José, *VoiceXML, Le langage d'accès à Internet par téléphone*, Paris, éditions Vuibert, 2004, ISBN : 2-7117-4826-X, 197 pages.
32. CHEVRIN Vincent, DERYCKE Alain, ROUILLARD José, *L'Architecture Logicielle UMR pour les interactions Multicanaux et Multimodales avec les e-Services*, IHM 2004, AFIHM, 2004, pp. 199-202, Namur, Belgique.
33. TARBY Jean-Claude, ROUILLARD José, *Plasticité des interfaces*, AS Plasticité, RTP 16, Méthodes et outils pour l'interaction homme-machine, 2004, Grenoble, Communications sans actes.
34. ROUILLARD José, *Multimodalité et sémantique*, Chapitre 7 de l'ouvrage "Variation, construction, instrumentation du sens", Hermès-Lavoisier, sous la direction de Maryse Siksou, 378 pages, 2003, pp. 177-196, Paris.
35. ROUILLARD José, *Plastic ML and its toolkit*, HCI International 2003, Lawrence Erlbaum Associates, Inc., Volume : 4, 2003, pp. 612-616, Heraklion, Crete, Greece.
36. DERYCKE Alain, ROUILLARD José, CHEVRIN Vincent, BAYART Yves, *When Marketing meets HCI: Multi-channel customer relationships and multimodality in the personalization*

- perspective*, HCI International 2003, Lawrence Erlbaum Associates, Inc., Volume : 2, 2003, pp. 626-630, Heraklion, Crete, Greece.
37. CHEVRIN Vincent, DERYCKE Alain, ROUILLARD José, *Un cadre théorique pour la caractérisation des interactions multicanal en E-Marketing*, IHM 2003, AFIHM, 2003, pp. 97-104, Caen, France.
 38. ROUILLARD José, L'hyperdialogue, Chapitre 7 de l'ouvrage "Interfaces Homme-Machine et Recherche d'information", Hermès-Lavoisier, sous la direction de Céline Paganelli, 334 pages, 2002, pp. 229-259, Paris.
 39. LAROUSSE Mona, ROUILLARD José, Adaptive communication tools for mobile learning, ABIS 2002, 2002, Hannover, Germany.
 40. VANTROYS Thomas, ROUILLARD José, Workflow and Mobile Devices in Open Distance Learning, IEEE International Conference on Advanced Learning Technologies (ICALT 2002), IEEE press, 2002, pp. 123-126, Kazan, Tatarstan, Russie.
 41. ROUILLARD José, A multimodal E-commerce application coupling HTML and VoiceXML, WWW 2002, The Eleventh International World Wide Web Conference, ISBN 1-880672-20-0, 2002, Waikiki Beach, Honolulu, Hawaii, USA.
 42. DERYCKE Alain, ROUILLARD José, La personnalisation de l'interaction dans des contextes multimodaux et multicanaux : une première approche pour le commerce électronique, IHM 2002, AFIHM, 2002, Poitiers.
 43. PETER Yvan, VANTROYS Thomas, ROUILLARD José, Support de l'activité coopérative dans un cadre mobile, Colloque sur la mobilité, GDR I3, 2002, Nancy.
 44. ROUILLARD José, CAELEN Jean, Le système HALPIN : Recherche documentaire en langue naturelle, Revue RIHM - Revue d'Interaction Homme-Machine, Hermès-Lavoisier, Vol 2 N°2, 2001, pp.55-82, Paris.
 45. ROUILLARD José, Quelques spécificités du dialogue Homme-Machine sur Internet, Revue Internationale BULAG. (Bulletin de Linguistique Appliquée et Générale), "Traitement Automatique des Langues et Internet", Coordonné par Henri Madec, Vol N°26, 2001, pp.219-234, Besançon.
 46. ROUILLARD José, Dialogue et Multimodalité, Revue RIHM - Revue d'Interaction Homme-Machine, Hermès-Lavoisier, Vol 2 N°1, 2000, pp.99-125, Paris.
 47. ROUILLARD José, Hyperdialogue avec un agent animé sur le Web, ERGO-IHM'2000, ILS/CCI Bayonne Pays Basque, 2000, Biarritz.
 48. ROUILLARD José, HYPERDIALOGUE SUR INTERNET. Le système HALPIN, Thèse d'Informatique, université Joseph Fourier, 2000, Grenoble.
 49. ROUILLARD José, Les enjeux d'un dialogue Homme-Machine sur Internet - L'Hyperdialogue, Bulletin d'informatique approfondie et applications, BIAA, Université de Provence, n° 52, 1999, pp. 03-20, Marseille.
 50. VAUFREYDAZ Dominique, AKBAR Mohammad, ROUILLARD José, Internet Documents: A Rich Source for Spoken Language Modeling, Automatic Speech Recognition and Understanding (ASRU'99), 1999, Keystone, Colorado, USA.
 51. ROUILLARD José, HALPIN : A natural language information retrieval system for a digital library on the World Wide Web, Webnet'99, 1999, Waikiki Beach, Honolulu, Hawaii, USA.
 52. ROUILLARD José, CAELEN Jean, Multimodal Information Seeking Dialogues on the World Wide Web, EUROSPEECH'99, ICASA, 1999, pp. 2151-2154, Budapest, Hungary.

53. VAUFREYDAZ Dominique, ROUILLARD José, AKBAR Mohammad, A network architecture for building applications that use speech recognition and/or synthesis, EUROSPEECH'99, ICSSA, 1999, pp. 2159-2162, Budapest, Hungary.
54. ROUILLARD José, CAELEN Jean, A Multimodal and conversational application in natural language for information seeking on the World Wide Web : The HALPIN system, 4th International Conference on Applications of Natural Language to Information Systems, 1999, pp. 239-244, Klagenfurt, Austria.
55. ROUILLARD José, CAELEN Jean, HALPIN : a multimodal and conversational system for information seeking on the World Wide Web, ESCA ETRW workshop : Accessing information in spoken audio, 1999, pp. 20-24, Cambridge, (UK).
56. ROUILLARD José, Navigation versus dialogue sur le Web. Une étude des préférences, IHM'99, AFIHM, 1999, Montpellier.
57. ROUILLARD José, Le système HALPIN : Hyperdialogue avec un Agent en Langage Proche de l'Interaction Naturelle, TALN'99, ATALA, 1999, pp. 137-146, Cargèse, Corse.
58. CAELEN Jean, ROUILLARD José, Le système HALPIN : recherche documentaire en langue naturelle, Workshop NTE-SPI-Sante, 1999, Lyon.
59. ROUILLARD José, CAELEN Jean, Etude du dialogue Homme-Machine en langue naturelle sur le Web pour une recherche documentaire, Deuxième Colloque International sur l'Apprentissage Personne-Système, CAPS'98, 1998, pp. 99-119, Caen.
60. ROUILLARD José, Hyperdialogue Homme-Machine sur le World Wide Web : Le système HALPIN, ERGOIA'98, 1998, pp. 343-345, Biarritz.
61. ROUILLARD José, Contribution à l'étude du dialogue Homme-Machine à travers le Web : la personnalisation, RECITAL'98, 1998, pp. 127-134, Le Mans.
62. VAUFREYDAZ Dominique, MEIYE Jean-Philippe, ROUILLARD José, Webcognition, expérimentations cognitives sur Internet, 6ème journée d'étude de l'association Accion, CREPCO, 1998, Aix-en-Provence.
63. MEIYE Jean-Philippe, ROUILLARD José, VAUFREYDAZ Dominique, Webcompletion - protocole de normes associatives sur Internet Ecole thématique CNRS – Sémantique, CNRS, 1998, Caen (Asnelles-sur-mer).
64. CAELEN Jean, ROUILLARD José, Interface vocale pour la recherche documentaire, Workshop NTE-SPI-Sante, 1998, Lyon.
65. ROUILLARD José, Dialogue oral Personne-Système sur le Web, Contribution pour l'atelier thématique "Interaction sur le WWW" de la conférence IHM'98, 1998, Nantes.
66. ROUILLARD José, Une composante orale de dialogue pour Internet, Rapport d'activité scientifique à la région Rhône-Alpes pour le projet ORION (2ème année), 1998, Grenoble.
67. ROUILLARD José, CAELEN Jean, Etude de la propagation au sein du Web à travers les liens hypertextes, Numéro spécial de la revue Hypertextes et Hypermédiat - 4ème conférence Internationale Hypertextes et Hypermédiat, Hermès, 1997, pp. 87-100, Paris.
68. ROUILLARD José, CAELEN Jean, A multimodal browser to navigate and search information on the Web, Fourteenth International Conference on Speech Processing (ICSP97) - IEEE Korea Council, IEEE Korea signal processing society, 1997, Seoul, Korea.
69. ROUILLARD José, Vers un dialogue oral sur le World Wide Web, 2èmes Rencontres Jeunes Chercheurs en Parole, 1997, La Rochelle.

70. ROUILLARD José, Une composante orale de dialogue pour Internet, Rapport d'activité scientifique à la région Rhône-Alpes pour le projet ORION (1ere année), 1997, Grenoble.

71. ROUILLARD José, Une composante orale de dialogue pour Internet. FACOPI : Feuilletteur A Commandes Orales Pour Internet, Rapport de DEA en Informatique Systèmes et Communications, Université Joseph Fourier de Grenoble, 1996, Grenoble.

7 Sélection d'articles de recherche

J'ai sélectionné huit articles pour compléter ce dossier d'HDR. Les quatre premiers abordent plutôt le thème de l'adaptation, et les quatre derniers, celui du multimodal et du multicanal.

A1 : [**Rouillard & al. 2006**] Rouillard José, Chevrin Vincent, Tarby Jean-Claude, Derycke Alain, La plasticité des documents numériques. Revue TSI (Technique et Science Informatique). Numéro spécial, « Document numérique », sous la coordination de Rémy Mullot, n°4, Paris, 2006.

Type : Revue nationale avec comité de lecture

A2 : [**Tarby & Rouillard 2006**] Tarby, Jean-Claude, Rouillard, José, Améliorer ses performances grâce au coaching numérique mobile : l'exemple du coaching vocal embarqué, UBIMOB'06, 2006, Paris.

Type : Conférence nationale avec comité de lecture

A3 : [**Bouyer & al. 2006**] Bouyer Antoine, Rouillard Jose, Derycke Alain, Pourquoi il n'est pas possible actuellement d'utiliser des objets pédagogiques plastiques respectant la norme SCORM, IHM 2006, AFIHM, Montréal, 2006.

Type : Conférence nationale avec comité de lecture

A4 : [**Rouillard 2008**] Rouillard José, Contextual QR Codes, Third International Multi-Conference on Computing in the Global Information Technology (ICCGI 2008), IEEE, Conference Publishing Service, ISBN 978-0-7695-3275-2, pp. 50-55, Athens, Greece, 2008.

Type : Conférence internationale avec comité de lecture

M1 : [**Rouillard 2003b**] Rouillard José, Multimodalité et sémantique, Chapitre 8 de l'ouvrage "Variation, construction, instrumentation du sens", Hermès-Lavoisier, sous la direction de Maryse Siksou, 378 pages, pp. 177-196, Paris, 2003.

Type : Chapitre de livre

M2 : [**Rouillard 2006**] Rouillard José, Web services and speech-based applications around VoiceXML, Journal of Networks (JNW, ISSN 1796-2056), Academy Publisher, Volume : 2, Issue : 1, pp. 27-35, Oulu, Finland, 2006.

Type : Revue internationale avec comité de lecture

M3 : [**Rouillard & al. 2008**] Rouillard José, Peter Yvan, Tarby Jean-Claude, Vantroys Thomas, Chevrin Vincent, Supporting Mobile Connectivity: from Learning Scenarios to Multichannel Devices, International Journal of Continuing Engineering Education and Life-Long Learning (IJCEELL), Inderscience Publishers, Special Issue on Learning as a Ubiquitous and Continuous Communication Attitude, Guest Editor: Piet Kommers, Volume 18, No. 4., pp.396-410, 2008.

Type : Revue internationale avec comité de lecture

M4 : [**Chevrin & al. Rouillard 2008**] Chevrin Vincent, Rouillard José, Instrumentation and measurement of multi-channel services systems, International Journal of Internet and Enterprise Management (IJIEM), Inderscience Publishers, Special Issue on: "Quality in Multi-Channel Services Employing Virtual Channels", Guest Editors: Dr. Rui Sousa, Catholic University of Portugal (Porto), Portugal, Volume 5 - Issue 4, pp. 333 - 352, 2008.

Type : Revue internationale avec comité de lecture

La plasticité des documents numériques

José Rouillard, Jean-Claude Tarby, Vincent Chevrin, Alain Derycke

Laboratoire TRIGONE, Institut CUEEP, Bâtiment B6

Université des Sciences et Technologies de Lille

59655 Villeneuve d'Ascq cedex – France

Jose.Rouillard@univ-lille1.fr, Jean-Claude.Tarby@univ-lille1.fr,

V.chevrin@ed.univ-lille1.fr, Alain.Derycke@univ-lille1.fr

RÉSUMÉ. Nous présentons dans cet article la notion de plasticité des documents numériques. Il s'agit d'étudier les multiples manières de présenter des documents numériques, en les adaptant, selon différents critères, en vue d'une personnalisation de forme et de fond. Nous nous intéressons plus particulièrement à l'intérêt et aux modalités d'adaptation des documents en fonction des contextes d'utilisation et des contextes d'usage dont fait partie l'utilisateur et la tâche à réaliser. Cette différence entre les « formes » que peut prendre un même document numérique s'apparente à la notion de plasticité qui peut alors s'appliquer sur un document à différents stades de sa vie (création, lecture, modification...).

ABSTRACT. This article presents the notion of plasticity for digital documents. We study the multiple manners to present an electronic document, in order to personalize it according to the form and to the content. This adaptation of the document can be made according to several criteria of personalization. We are interested more particularly in the methods of adaptation for the documents according to the context of use to which belong the user with her/his tasks and the device allowing to read this document. This difference between the "forms" that the same digital document can take connects with the concept of plasticity which can then be applied to a document at various stages of its life (creation, reading, modification...).

MOTS-CLÉS : Document, document numérique, plasticité, interface adaptative.

KEYWORDS: Document, digital document, plasticity, adaptative user interface.

1. Introduction

L'information qui circule aujourd'hui sur la planète est majoritairement stockée sous forme numérique¹. Par ailleurs, le besoin d'un accès ubiquitaire à l'information (au bureau, chez soi, dans le train, etc.) se concrétise d'autant plus avec la venue de nouveaux matériels et le succès des ordinateurs de poche ainsi que des téléphones mobiles [Rouillard 04]. Qu'ils soient grands (tableaux interactifs) ou petits (téléphones portables, PDA²), les capacités de ces matériels sont variables. Ceci appelle une réflexion sur les techniques de spécification d'interfaces indépendantes dont le but est d'éviter des développements spécifiques [Mitrovic & al. 04]. Certains auteurs pensent que cette situation justifie l'étude scientifique de modèles et de technologies permettant de s'abstraire du périphérique employé par l'utilisateur [Myers & al. 02].

Dans la pratique, tous les moyens d'accès ne se valent pas. Avec la multiplicité des moyens de connexion à Internet, les nouvelles applications permettant d'accéder à ce réseau se doivent d'être plastiques, c'est-à-dire capables de s'adapter à la plateforme, à l'utilisateur et au contexte d'interaction [Calvary & al. 01a]. Il faut pour cela concevoir des interfaces abstraites et des mécanismes de transformation pour obtenir des interfaces concrètes qui soient adaptées à chaque « Device Utilisateur » (DU)³. Le W3C⁴ a d'ailleurs lancé une activité dans le domaine du « Device Independence »⁵ dès le début de l'année 2001.

Idéalement, on souhaite tendre vers des opérations de transformation d'interfaces pour tous et n'importe où (« *transforming the UI for anyone anywhere* ») [Lopez & Szekely 01]. Cette tendance n'est pas uniquement perceptible dans le domaine des Interfaces Homme-Machine (IHM), mais également dans d'autres communautés, comme celle du document, au sens large, et des hypermédias, plus particulièrement : *Ainsi, l'hypermédia [...] peut faire appel à un modèle de tâche, un modèle de l'utilisateur, et un modèle du domaine afin d'élaborer la réponse la plus appropriée* [Nanard & Nanard 01].

Dans cet article, nous nous intéressons à la notion de plasticité des documents numériques. Pour cela, nous étudierons tout d'abord la notion de document (numérique ou non), puis nous aborderons la propriété de plasticité sous l'angle des

¹ Un projet à l'Université de Berkeley a estimé à un exa-octet (c'est-à-dire à 1 milliard de giga-octets) la quantité de données générées annuellement de par le monde. Parmi ces données, 99,997 % sont disponibles sous forme numérique. [Keim 01] cité par [Nigay 01].

² Personal Digital Assistant

³ Nous utilisons la définition du « Device Utilisateur » (DU) exposée dans (Chevrin 06). Elle fait référence à la plateforme logicielle, à la plateforme matérielle, mais également aux aspects physiques (facteurs de taille, mobilité, portée).

⁴ World Wide Web Consortium

⁵ <http://www.w3.org/2001/di>

IHM. Ensuite, nous expliquerons la notion de plasticité des documents numériques par elle-même, avant de conclure dans la dernière partie.

2. Document et document numérique

2.1. Notion de document

De manière globale, la notion de document demeure assez floue, et la définition même du document peut varier d'un champ d'application à l'autre [Buckland 98]. Les chercheurs du domaine ne confrontent et ne discutent d'ailleurs que rarement leurs points de vue concernant la nature du document, au sens large. De ce fait, des aspects sensiblement différents émergent lors de discussions entre « chercheurs STIC⁶ », plutôt « orientés technologies », et « chercheurs SIC⁷ » qui ont une sensibilité plus « informationnelle ». Ainsi, quelques nuances apparaissent lorsque l'on distingue les documents par rapport à des critères tels que la forme, le signe ou le médium [RTP-DOC 03]. En s'intéressant à la forme du document, on considère le document en tant qu'objet et on tente de qualifier sa nature et sa structure : est-ce un objet réel, palpable ou bien immatériel, virtuel ? En se focalisant sur le signe, on tente d'identifier la portée sémantique (messages, intentions...) que véhicule le document. En observant la facette médium d'un document, on s'interroge davantage sur la manière dont ce document devient un vecteur de communication.

Nous concilions dans nos travaux ces trois dimensions en distinguant particulièrement :

- les formes, structures et normalisations des documents (notamment via la séparation stricte entre données et présentation, facilitée par le langage XML) ;
- les manipulations, utilisations, et réutilisations possibles (stockage, indexation, référencement...) ;
- les contextes d'utilisation (outils, mobilité des utilisateurs, granularité de l'information fournie, dimensions spatiales et temporelles, etc.) ;
- le statut du document dans les relations sociales (notions de rôles, d'autorités, de preuves, de confidentialité) ;

Un document peut également être vu comme constitué selon trois dimensions : le contenu, le contenant, et le contexte de production [Chabin 97] [Stern 97].

- Le *contenu* est l'information contenue dans le document. L'auteur a conçu et représentée cette information de façon à la rendre compréhensible par un lecteur. Cette écriture a influencé la structure du document sur le plan

⁶ STIC : Sciences et Technologies de l'Information et de la Communication

⁷ SIC : Sciences de l'Information et de la Communication

organisationnel (plan du document) et sur le plan matériel (polices de caractères, couleurs...). Par conséquent, le contenu d'un document contient de façon implicite les intentions de l'auteur (document servant de ressource pédagogique, de preuve juridique...), et par conséquent le mode attendu d'utilisation et de lecture du document par le lecteur.

- Le *contenant* est le support de stockage de l'information. Ce support peut être du papier, une microfiche, un disque dur, une plaque de verre, etc. Le support et le contenu sont intimement liés, d'une part pour la phase de création du document (écrire sur du papier et écrire sur une microfiche sont très différents), et d'autre part pour la phase d'utilisation du document (lecture d'un texte papier ou d'un texte sur microfiche).
- Le *contexte de production* est le contexte dans lequel l'information a été consignée sur le support. Ce contexte inclut le contexte de l'activité de production. On saura par exemple si le document a été produit lors d'un procès, d'une réunion historique, de travaux pratiques, etc. On pourra éventuellement savoir où le document a été écrit, quand, par qui, etc. Cette dimension recoupe la notion de « document comme signe » de [RTP-DOC 03] qui dit que « le document est indissociable du sujet en contexte qui le construit ou le reconstruit et lui donne sens ».

Notons que nous pouvons ajouter à la dimension «support» ci-dessus la dimension associée à la phase de reproduction du document. Cette phase est très liée au support puisque celui-ci conditionne le mode de reproduction du document. En effet, reproduire à partir d'une feuille papier ou reproduire à partir d'une microfiche sont deux activités différentes. A ce titre, la notion de document original tend à devenir obsolète avec le numérique. Par exemple, un document écrit avec un traitement de texte peut être imprimé des milliers de fois de la même façon ; il n'y a, de fait, plus de document original comme cela aurait été le cas avec une écriture manuelle ou une machine à écrire. L'apparition du numérique a donc changé la notion de document original, mais le numérique a surtout apporté d'autres aspects qui font évoluer la notion de document.

2.2. Document numérique

Dans notre propos, un document numérique est un document créé à partir de logiciels ou de systèmes d'information, ou bien un document numérisé à partir d'un document tangible, et pouvant être présenté sous une forme compréhensible par un lecteur, humain ou non.

Si l'on reprend les trois critères « forme, signe et médium » d'un document, un document numérique se voit alors attribuer des définitions particulières [RTP-DOC 03] :

- Pour le critère médium, l'équation *document traditionnel = inscription + sens* devient avec le numérique *document numérique = texte informé + connaissances*, ce qui dans le cas du Web sémantique débouche sur *document numérique avec le Web sémantique = texte informé + ontologies*.
- Pour le critère forme, l'équation *document traditionnel = support + inscription* devient avec le numérique *document numérique = structures + données, et document XML = données structurées + mise en forme*.
- Pour le critère signe, l'équation *document traditionnel = inscription + sens* devient avec le numérique *document numérique = texte informé + connaissances*. « Le remplacement d'*inscription* par *texte informé* voudrait signifier que le texte (pris au sens large, y compris audiovisuel) a été soumis ou pourrait être soumis à un traitement permettant d'en repérer les unités d'information. Le remplacement de *sens* par *connaissances* voudrait introduire la notion de personnalisation pour un lecteur ou un usager donné » [RTP-DOC 03]. Cet aspect de personnalisation du document numérique est un des piliers de notre travail, et sera abordé un peu plus loin.

2.3. Discussion

Plusieurs aspects importants sont à retenir au sujet des documents numériques. Premièrement, avec le numérique, le support est devenu accessoire. L'important est que le sens associé au contenu ne soit pas altéré. Or, ce sens a été donné par le créateur du document dans le *contexte de création*, et ce sens sera (re)construit par rapport au *contexte de production et de diffusion* lors de l'utilisation du document, ce qui va conditionner l'interprétation du contenu.

Deuxièmement, on peut noter qu'un document numérique n'existe qu'à deux moments, celui de sa conception par son auteur et celui de sa (re)construction par le lecteur. « Il est peu probable que le document sera toujours identique dans l'un et l'autre cas. [...] Une communauté de chercheurs réfléchit à cette question, dans le contexte du Web sémantique, en termes de *documents virtuels personnalisables* » [RTP-DOC 03]. De nouveau, nous retrouvons ici la notion de *personnalisation*. Cette notion est très fortement liée à la notion de *plasticité* détaillée dans la partie 3.

Troisièmement, des notions associées au document traditionnel sont fortement modifiées avec l'intégration de la dimension numérique. Nous avons vu précédemment que le numérique a fait disparaître la notion de document original. Par ailleurs, le contexte de production se découpe dorénavant en deux : contexte de production pour la création et contexte de production pour la (re)construction/restitution du document au lecteur. Le contexte de production inclut dans les deux cas l'environnement informatique permettant l'élaboration du document, c'est-à-dire principalement le(s) logiciel(s) utilisé(s) et la plateforme logicielle. Contrairement au document traditionnel qui bénéficie intrinsèquement de nombreux types de supports, le support d'un document numérique, avant de devenir

tangible, est uniquement informatique (disque dur, CDROM, clé USB...) de même que son format (.doc, .txt, .pdf, base de données...). De plus, le support d'un document numérique n'est pas fixe comme celui d'un document traditionnel; il est aisé de passer du format .doc au format .pdf par exemple. Enfin, le contenu d'un document traditionnel est fixe, alors qu'il ne l'est pas toujours avec un document numérique. On peut en effet facilement ajouter, modifier, supprimer de l'information dans les documents traditionnels et numériques, mais ces derniers fournissent la possibilité d'effacer toutes ces étapes de mises à jour pour ne garder que la dernière version, ce qui n'est pas possible avec un document original traditionnel. Notons au passage que cette disparition des modifications peut s'avérer très fâcheux dans certains cas où la traçabilité est cruciale (dans le monde juridique par exemple).

« Un document numérique est un ensemble de données organisées selon une structure stable associée à des règles de mise en forme permettant une lisibilité partagée entre son concepteur et ses lecteurs » [RTP-DOC 03]. Cette définition appelle des questions telles que : comment mettre en forme le document pour être sûr de cette lisibilité ? Est-ce que cette mise en forme est dynamique, c'est-à-dire qu'elle dépend du lecteur, du moment où l'information est lue, du lieu où elle est lue, du contexte d'usage dans lequel elle est lue, etc. ? Toutes ces questions se rattachent à l'aspect dynamique du contexte d'usage du document. Ce *contexte d'usage* est constitué du lecteur du document, de ses activités au moment où il lit le document, y compris son activité de lecture du document, ainsi que de ses intentions vis-à-vis du document. Ce contexte d'usage est donc centré sur le lecteur, mais il est également attaché à un *contexte d'utilisation* qui, lui, est centré sur l'aspect informatique, c'est-à-dire la plateforme et l'environnement logiciel (le DU) permettant la lecture du document. Cette lecture peut être immédiate (afficher un pdf par exemple) ou bien être le résultat d'un traitement informatique (par exemple une extraction d'informations depuis une base de données).

Dans la suite de cet article, nous abordons la problématique de l'accès, l'archivage et l'utilisation des documents numériques, sous l'éclairage d'un champ de recherche relativement récent en IHM (Interface Homme-Machine), celui de la plasticité.

3. Propriété de plasticité

Le dictionnaire de l'Académie Française définit la « plasticité » comme la « qualité de ce qui peut prendre diverses formes », et le Larousse la définit en psychologie comme la « capacité d'un sujet à déstructurer un ensemble perçu et à le restructurer selon une forme différente ». Ces deux définitions sont proches de la notion de documents numériques telle que nous l'avons vu dans la partie précédente. En effet, un document numérique peut subir des modifications de son support, de sa forme, de sa structure, et cela en fonction du contexte d'usage et du contexte d'utilisation.

Etant donné que les documents numériques, pour être manipulés (affichés, transformés...), ont besoin de l'informatique, et le plus souvent d'une interface homme-machine (IHM), il nous paraît essentiel de se pencher sur la plasticité des IHM.

3.1. Plasticité des IHM

Dans le domaine informatique, la *plasticité* a été définie pour les interfaces homme-machine (IHM) [Thevenin 01] comme la possibilité, pour une interface, de *s'adapter* à son contexte d'usage dans le respect de son utilisabilité. En plasticité des IHM, un *contexte d'usage* est défini comme un triplet $\langle E, P, U \rangle$ où E est un des environnements concernés par l'application, P est une des plateformes concernées par l'application et U est un des utilisateurs concernés par l'application. A partir d'une spécification donnée, l'interface pourra ainsi être façonnée de multiples manières pour être utilisée sur des plateformes éventuellement hétérogènes. Plusieurs types d'adaptations peuvent être observées ; il s'agit d'adaptations dites au développement, à l'exécution, ou bien à l'installation. Ceci est particulièrement visible avec Internet. En effet, l'évolution récente de ce réseau, couplée à l'évolution technologique des périphériques tels que les téléphones portables (technologie WAP [WAP] et I-Mode entre autres), ainsi que l'apparition de nouveaux matériels et le succès des ordinateurs de poche (PDA), ont obligé les concepteurs et les réalisateurs de sites Internet à produire de plus en plus de pages différentes permettant pourtant d'obtenir la même information. Ainsi, classiquement, permettre l'accès à N informations au travers de M périphériques revient, pour le concepteur, à coder $N \times M$ pages (ou applications). En lieu et place de ces $N \times M$ applications, les chercheurs tentent de proposer un modèle pour plusieurs interfaces [Paternò & Santoro 02] [Tarby 04].

Ces aspects de génie logiciel concernent principalement la plasticité des interfaces utilisateurs [Calvary & al. 01b]. Le projet CAMELEON [CAMELEON] introduit à ce sujet un cadre de travail plus complet (cf. Figure 1) puisqu'il intègre la plasticité depuis le niveau de l'IHM finale jusqu'au niveau des tâches et concepts sous-jacents à l'application plastique associée.

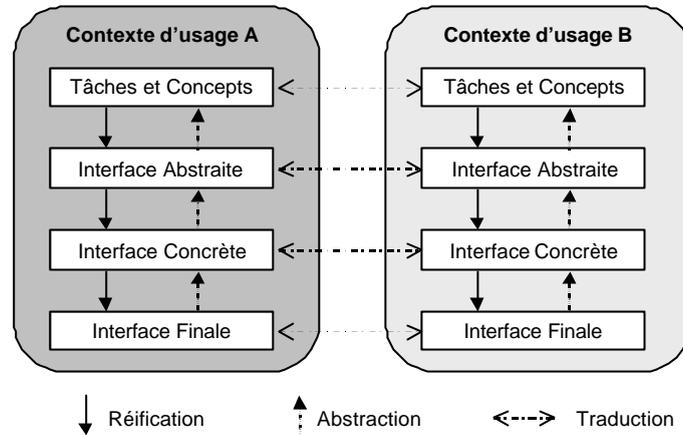


Figure 1. Le cadre de référence du projet CAMELEON

Ce projet s'intéresse aux étapes de développement des IHM d'applications interactives sensibles aux contextes (*context-aware applications*). Dans CAMELEON, en terme d'IHM, le niveau le plus abstrait est celui des tâches et concepts orientés métiers, et le niveau le plus concret est celui de l'interface finale. Or, même si ce cadre de référence semble complet, les travaux qui en ont découlé jusqu'à présent ne concernent que la plasticité des IHM, c'est-à-dire que ces travaux étudient comment une IHM doit s'afficher et s'adapter en fonction du contexte, sans se préoccuper des documents que ces IHM manipulent.

3.2. De nouveaux langages informatiques basés sur XML pour la plasticité des IHM

Dans l'optique de la plasticité des IHM, des langages de représentation abstraite ont été conçus depuis quelques années pour permettre aux développeurs de décrire une interface de manière abstraite. Toutefois, la plupart de ces langages restent relativement proches des modes d'interaction graphique, ce qui limite leur application avec d'autres types d'interaction tels que la voix [Lin & al. 01]. On parle alors de systèmes proposant un contenu « taillé sur mesure ». Ce *sur mesure* peut être par rapport au périphérique visé, au contexte informatisé (dans le cadre de la mobilité par exemple), à l'utilisateur lui-même : son rôle, ses droits et privilèges, ses capacités physiques ou cognitives, son niveau de stress, le temps imparti pour la tâche à effectuer ; ou encore à l'environnement physique (bruité, lumineux ...).

Les nouvelles formes d'interaction (voix, geste, multimodal, multicanal [Chevrin 06]...) se heurtent à de nombreuses difficultés parmi lesquelles le développement et le maintien des interfaces homme-machine sur chaque système.

Pour être utilisable, l'interface utilisateur doit être capable de s'adapter aux dispositifs d'interaction, par exemple à la taille de l'écran, à la présence ou non d'un clavier ou à d'autres contraintes comme les profils ou les préférences de l'utilisateur. Cela a conduit les chercheurs à s'interroger sur les possibilités de mettre en œuvre de nouveaux langages particulièrement bien adaptés à la manipulation de données et aux traitements permettant de s'affranchir des limites des langages traditionnellement utilisés lors de la conception des IHM. Il faut pour cela disposer de langages permettant de dissocier le contenu, d'une part, et la présentation de ces données, d'autre part.

En ce qui concerne le World Wide Web, HTML combine données et présentations dans le même document. Inversement, XML⁸[XML] fournit une séparation entre les données et la manière dont on les présente. Dire qu'un langage comme XML est extensible signifie que le concepteur peut ajouter des balises pour ses propres besoins, de manière illimitée. L'utilisation conjointe de données et de feuilles de style XSL⁹ [XSL] permet d'afficher les mêmes données sous différents formats et différents aspects.

De nouveaux langages tels que UIML¹⁰ [UIML] [Merrick 01], RIML¹¹ [Spriestersbach & al. 03], ou encore AUIT¹² [Grundy & Zou 04], visent le niveau des interfaces concrètes (cf. Figure 1). Ces langages, fondés sur XML, permettent de séparer les données de leur présentation, et de s'abstraire des périphériques en utilisant des objets d'interaction abstraits. Par exemple, on ne parlera pas de « listbox », ou de « menu », mais de « liste d'éléments » et de « groupes d'actions ». Bien que cette approche basée sur une abstraction des périphériques soit attrayante, elle nous paraît encore trop réductrice car trop ciblée sur l'IHM (bouton, menu...) et ne tenant pas compte des documents manipulés. Notons cependant une exception avec le langage PlasticML [Rouillard 03] qui permet de décrire des éléments d'entrée et de sortie d'interfaces, à un haut niveau, c'est-à-dire sans faire référence au périphérique qui sera utilisé lors de l'interaction. Grâce à des transformations avec XSLT¹³, les documents en PlasticML sont traduits automatiquement vers des langages idoines (HTML pour le Web, WML pour le WAP, VoiceXML pour le téléphone). PlasticML dispose de balises suffisamment abstraites pour décrire une tâche (ou une action) sans pour autant devoir définir quels objets informatiques (boutons, menu déroulants, etc.) devront être utilisés pour un périphérique spécifique.

A l'opposé de ces langages, les conceptions basées sur les modèles (« model-based ») [Eisenstein & al. 01] permettent de s'abstraire des périphériques et de

⁸ eXtensible Markup Language

⁹ eXtensible Stylesheet Language

¹⁰ User Interface Markup Language

¹¹ Renderer-Independent ML

¹² Adaptable User Interface Technology

¹³ eXtensible Stylesheet Language Transformation

s'intéresser plus particulièrement aux modèles des utilisateurs, aux modèles d'interaction, aux modèles de tâches et aux modèles de contexte d'utilisation [Müller & al. 01]. Ce type d'approche permet de spécifier les systèmes en se détachant des contraintes matérielles et logicielles. Cependant, le passage de spécifications abstraites à des systèmes réels demande encore beaucoup de travail étant donné la multiplicité des contraintes liées notamment aux périphériques, aux cas d'utilisation, aux types et aux niveaux des utilisateurs. Les langages, associés à ce type de conception, tels que XIML¹⁴ [XIML] ou UsiXML¹⁵ [UsiXML], visent avant tout les niveaux les plus abstraits du cadre de référence CAMELEON (tâches et concepts, interfaces abstraites).

Pour un état de l'art plus complet des langages de description basés sur XML, le lecteur pourra consulter [Limbourg & al. 04].

3.3. Discussion

Les documents numériques peuvent être plastiques sans que l'IHM qui les affiche le soit nécessairement. La Figure 2 donne un exemple de ce type de situation avec le cas d'un cours sur le fonctionnement des moteurs affichant ou non une vidéo et des images en complément du texte. Nous pouvons donc distinguer la notion de plasticité des IHM de celle de plasticité des documents numériques. De plus, la plasticité des IHM, comme son nom l'indique, ne s'occupe que des IHM et non des documents que ces IHM affichent. Nous avons vu que les documents numériques intègrent des informations, telles que les intentions de l'auteur ou bien encore la structure logique du contenu, que la plasticité des IHM ignore totalement. Pour être clairement définie, la plasticité des documents numériques ne peut donc pas se contenter des critères énoncés dans la plasticité des IHM.

Nous considérons que la plasticité des documents numériques se trouve à l'intersection de deux domaines correspondant l'un aux IHM et l'autre aux documents numériques. Pour s'adapter aux contextes de production, d'utilisation et d'usage, un document numérique doit donc être plastique. [Gallezot 02] a défini la plasticité d'un document comme « [...] la capacité [...] à être polymorphe, à recomposer des unités informationnelles selon des usages, des besoins, des situations différentes ».

Le contexte d'usage est présent dans les deux domaines (IHM et documents numériques), mais sa définition dans la plasticité des IHM, c'est-à-dire $\langle E, P, U \rangle$, est différente de notre définition au sujet des documents numériques. En effet, le contexte d'usage, tel que nous l'avons défini au §2.3, est composé uniquement de l'utilisateur avec ses tâches et ses intentions, et n'inclut donc pas la plateforme et l'environnement. Rappelons cependant que notre contexte d'usage est en lien direct

¹⁴ eXtensible Interface Markup Language

¹⁵ USer Interface eXtensible Markup Language

avec le contexte d'utilisation (cf. §2.3) qui, lui, est l'équivalent de E+P dans le contexte d'usage de la plasticité des IHM.

Chaque adaptation d'un document numérique devra être qualifiée par rapport à un contexte d'usage et d'utilisation précis. Ainsi, le contexte sera constitué d'informations caractérisant une personne, un lieu, un objet, ou tout autre élément considéré comme pertinent pour une interaction entre un utilisateur et une application, y compris l'utilisateur et l'application eux-mêmes [Dey & Abowd 00] [Dey 01]. Le contexte peut alors dépendre du périphérique employé par l'utilisateur, du profil de cet utilisateur, de sa localisation géographique au moment de l'utilisation de l'interface, de l'historique sur une ou plusieurs sessions, etc. [Signer & al. 02]. Cependant, les adaptations aux contextes ne sont pas de même niveau ; celle des IHM concerne le contenant (c'est-à-dire le support, cf. §2.1) des documents numériques alors que celle des documents numériques concerne leurs contenus (c'est-à-dire l'information, cf. §2.1).

4. Contenu sur mesure et plasticité des documents

Après avoir vu les notions de document numérique et de plasticité, nous abordons à présent celles de plasticité des documents et d'adaptation aux différents contextes.

4.1. Plasticité des documents numériques

La Plasticité des documents numériques va être appréhendée dans les lignes qui suivent selon deux approches : la plasticité du support et la plasticité du contenu.

4.1.1. Plasticité du support

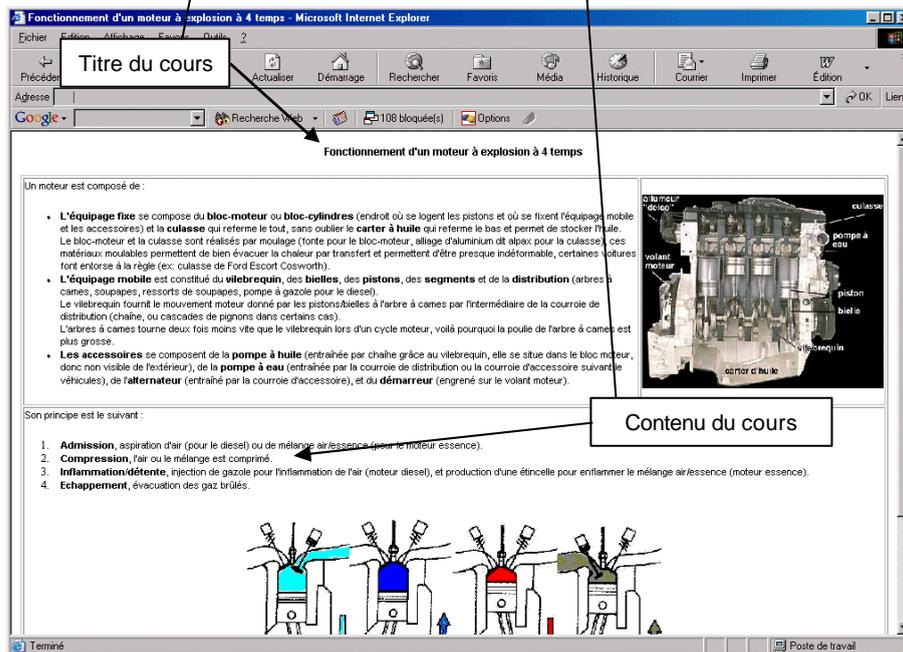
La plasticité du support (cf. §2.1) d'un document numérique correspond à modifier l'apparence de ce document sans toucher à sa structure. Cette modification d'apparence se fait, bien évidemment, au travers de l'application qui diffuse le document, cette application pouvant elle-même avoir une IHM plastique ou non.

Un document numérique peut donc posséder des propriétés intrinsèques de plasticité, par exemple en associant des informations redondantes dans des formats différents (cf. Figure 2 avec l'explication de l'explosion dans un moteur à essence sous forme textuelle pure et sous la forme d'une vidéo, par exemple).

Cependant, même si un document possède ces propriétés, *a priori*, il ne décide pas seul de la modification de son apparence. Celle-ci est pilotée par l'application qui diffuse le document en se basant sur des informations prenant en compte le contexte d'utilisation (PDA/PC, environnement de travail bruité/calme...), et le contexte d'usage (utilisateur avec ses droits d'accès, ses rôles, ses tâches...).



a. Affichage sur un téléphone wap



b. Affichage sur une page web

Figure 2. Affichage du fonctionnement d'un moteur sur deux DU différents. Dans (a) et (b) le texte est strictement identique, mais (b) permet un affichage du texte augmenté d'images et de vidéos.

4.1.2 Plasticité du contenu

La plasticité du contenu (cf. §2.1), c'est-à-dire de l'information, d'un document numérique revient à modifier la structure même de ce document. On touche donc ici au fond et non plus à la forme du document.

Un document peut être « balisé » (au sens XML) pour fournir des informations sur sa structure, mais également sur son usage, sur sa sémantique, etc. Un document de cours peut ainsi être balisé en une partie de cours présentant des concepts, puis des détails sur ces concepts, suivis de questions/réponses associés à ces concepts. Ces dernières peuvent à leur tour être balisées en fonction du niveau de complexité ou d'approfondissement des concepts vus. Ce type de balisage permet tout d'abord d'accéder à la structure du document pour éventuellement en modifier le contenu ; par exemple, dans le cadre de l'EAD¹⁶, un enseignant peut modifier en temps réel le document (correction orthographique, ajout de liens hypertextes, commentaires...) et ainsi le mettre à jour automatiquement sur les postes des étudiants en EAD. Ensuite, ce type de balisage permet de modifier la structure du document ; par exemple, dans un EIAH¹⁷, la plateforme d'enseignement pourra réarranger la structure du document en fonction du parcours et de l'activité de chaque étudiant. De cette façon, dans le cadre d'une révision, on affichera plus de questions/réponses qu'à l'accoutumée, ou bien on affichera d'abord des questions, en proposant seulement ensuite des concepts en cas de blocage, ou bien encore on ajoutera des informations complémentaires telles qu'une webographie ou des commentaires de l'enseignant.

4.2 Plasticité et adaptation aux contextes

Notre approche, vis-à-vis de la plasticité des documents numériques, consiste à identifier et à séparer strictement les éléments de forme des éléments de fond. La Figure 3 ci-après est un exemple de document pédagogique décrit grâce au langage PlasticML. Il s'agit de la poésie « Le dormeur du Val » d'Arthur Rimbaud. Dans cet exemple, nous présentons, certes la poésie, mais également d'autres éléments descriptifs qui permettront une mise en forme ultérieure, lors de la prise de connaissance du document par l'utilisateur final (visualisation, écoute, transformation en alphabet braille, etc.). Cependant, ces éléments descriptifs doivent être d'un niveau d'abstraction relativement élevé, afin de ne pas se focaliser, lors de la création du document, sur un mode privilégié d'accès à l'information (graphique au détriment des autres modes, le plus souvent).

¹⁶ EAD : Enseignement à distance

¹⁷ EIAH : Environnement Informatique pour l'Apprentissage Humain

```

<?xml version="1.0" encoding="iso-8859-1"?>
<plasticML version="2.0">

  <interface title="Le Dormeur du Val">

    <block      role="ALL"      align="center"      hierarchy="NO"
line_return="NO">
      <output>
        <view src="dormeur.jpg" alt="image du dormeur du val"
width="400" height="213" />
      </output>
    </block>

    <block      role="ALL"      align="center"      hierarchy="NO"
line_return="YES">
      <output>C'est un trou de verdure où chante une
        <important1>rivière</important1>,
      </output>
    </block>

    <block      role="ALL"      align="center"      hierarchy="NO"
line_return="NO">
      <output>Accrochant      follement      aux      herbes      des
haillons</output>
    </block>

    (...)
    <block      role="ALL"      align="center"      hierarchy="NO"
line_return="YES">
      <output>Tranquille. Il a deux trous
        <important2>rouges</important2> au côté droit.
      </output>
    </block>

    <block role="ALL" align="center" hierarchy="NO">
      <output>
        <view src="Arthur_Rimbaud.gif" alt="Portrait Arthur
Rimbaud"
width="99" height="143" />
        <link url="http://www.poetes.com/rimbaud/">
          Arthur Rimbaud
        </link>
      </output>
    </block>

  </interface>

</plasticML>

```

Figure 3. *Extrait d'un document PlasticML : la poésie « Le dormeur du Val » d'Arthur Rimbaud*

Ainsi, dans notre exemple, chaque strophe est représentée par un bloc (ce qui est une représentation possible parmi d'autres). Une fois le document décrit en PlasticML, il est alors possible de mettre en forme le document final, au moment de sa consultation par l'utilisateur (c'est-à-dire dans le contexte d'usage), en l'adaptant dynamiquement selon les terminaux d'accès à l'information disponibles (contexte d'utilisation).

4.2.1. Adaptation par rapport au DU

La Figure 4, ci-dessous, présente le document numérique obtenu, à partir du document PlasticML (Figure 3), en vue d'une visualisation par l'utilisateur sur un navigateur graphique pour le Web. Les mots « rivière » et « rouges » qui avaient été encapsulés (respectivement) par les balises¹⁸ <important1> et <important2>, dans le document de description abstraite de l'interface, se trouvent mis en forme (italique, gras, etc.) au moment de la génération en HTML. La photo de l'auteur du poème est affichée, ainsi qu'un hyperlien pointant vers un site consacré à Rimbaud.

¹⁸ Ces balises ne sont pas hiérarchisées. Autrement dit, <importantN> n'est pas considérée comme plus importante que <importantN-1>.

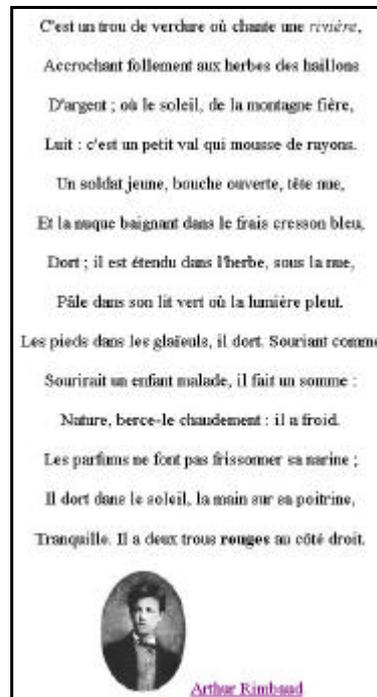


Figure 4. Adaptation du document numérique en vue d'une visualisation par l'utilisateur via un navigateur graphique pour le Web

La Figure 5, ci-après, présente les documents numériques obtenus, à partir du document PlasticML de la Figure 3, lorsque l'adaptation se fait vers un téléphone supportant le WAP. Les mots « rivière » et « rouges » ont été respectivement mis en forme en italique et gras, comme pour la version Web. Mais ici, les strophes du poème sont visibles bloc par bloc (notion de *card* du WML), et l'utilisateur doit manipuler les touches de navigation de son téléphone pour prendre connaissance du poème dans son intégralité. La photo de l'auteur n'est pas présentée dans la version WAP (bien que la transformation de l'image vers le format wbmp soit techniquement possible).



Figure 5. *Adaptation du document numérique en vue d'une visualisation par l'utilisateur via un navigateur graphique pour le WAP*

Ce premier niveau d'adaptation par rapport aux contraintes technologiques peut être raffiné si l'on prend en compte d'autres éléments comme par exemple le modèle de l'utilisateur. Nous abordons cette notion dans les lignes qui suivent.

4.2.2. Adaptation par rapport à l'utilisateur

Un aspect difficile à prendre en compte dans le domaine de la plasticité est sans aucun doute l'utilisateur. Cette difficulté est due d'une part au côté « humain » de l'utilisateur, et d'autre part à son ou ses rôles, et leurs évolutions, dans le contexte d'exécution de la tâche. Ces deux aspects sont à présent détaillés.

4.2.2.1. Aspect « humain » de l'utilisateur

Chaque être humain est sujet à de nombreuses perturbations (bonnes ou mauvaises) qui font évoluer continuellement son état (par exemple son niveau de fatigue ou de stress). Par conséquent, un utilisateur peut tout à fait se retrouver dans l'obligation d'exécuter la même tâche (par exemple la consultation de documents) dans le même contexte logiciel et matériel (y compris le lieu où se déroule la tâche) en étant dans deux états, de fatigue par exemple, différents d'un jour à l'autre. La plasticité à ce niveau se traduira, par exemple, au travers d'une adaptation du contenu des documents en fonction de l'état de fatigue, ou bien encore une adaptation de l'IHM au niveau des services offerts afin d'optimiser le temps nécessaire à la réalisation de la tâche demandée.

L'« *affective computing* » [Affect] commence à donner des résultats significatifs quant à la détection et à la création d'émotions chez l'humain. Il est devenu possible de détecter le niveau de stress de l'utilisateur, uniquement par la voix. Des compagnons logiciels, détectant et gérant l'intérêt d'un utilisateur pour la tâche qu'il réalise, commencent à voir le jour. Par ailleurs, les interacteurs se diversifient de plus en plus, et une partie d'entre eux est anthropocentrée (lecteur d'empreintes digitales, capteur de pression, etc.). Il est donc tout à fait envisageable de penser que les applications de demain prendront en compte cet aspect pour aider les utilisateurs dans la manipulation des documents numériques. On peut imaginer par exemple que ces applications pourront évaluer la satisfaction de l'utilisateur lorsqu'elles lui afficheront des documents, et ainsi proposer éventuellement d'autres mises en forme plus satisfaisantes. Dans un avenir plus lointain, et compte-tenu des nouvelles avancées au sujet du BCI¹⁹ (Brain Computer Interface) [Hoffmann & al. 04], on peut imaginer que compte tenu du contexte d'utilisation, et surtout des intentions de l'utilisateur (contexte d'usage), les applications pourront déduire quels documents, avec leurs contenus et leurs contenants, il faudra afficher à l'utilisateur.

4.2.2.2. Rôles de l'utilisateur

Chaque utilisateur joue un ou plusieurs rôles au sein de ses activités que ce soit dans son entreprise (directeur, chef du personnel, commercial...), dans sa famille (père, mère ...), dans ses activités associatives (animateur...), au sein de l'Etat (électeur, salarié...) ou dans tout autre contexte. Or, même si la notion de rôle a déjà été étudiée dans différents domaines, par exemple en psychologie ou en informatique (TCAO, e-learning, modélisation objet...), cette notion prend des sens différents en fonction des domaines d'étude. Si l'on regarde de plus près le domaine informatique, on constate que la définition des rôles dans RM-ODP²⁰ [Putman 01] est l'une des plus complètes dans le domaine de la modélisation objet puisqu'un rôle peut être dynamique et qu'il peut être associé à une politique de rôles. De plus, un rôle est associé à un utilisateur, ce dernier pouvant être un objet, une personne, un groupe de personnes ou une organisation. La spécification XACML²¹ [XACML] continue dans cette voie puisqu'elle définit des règles d'accès à des ressources, dont les documents numériques, en se basant sur la notion de rôles.

En informatique, les rôles sont généralement étudiés soit du côté purement humain, soit du côté purement objet, laissant ainsi de côté tous les liens entre ces deux mondes. La conséquence immédiate de cette séparation est l'importante perte d'informations lors du passage des modèles représentant des activités humaines aux modèles objets les implémentant. Or, ceci est d'une importance capitale si l'on veut que la plasticité des documents numériques s'applique parfaitement. En effet,

¹⁹ BCI : A Brain Computer Interface is a communication system, that implements the principle of "think and make it happen without any physical effort" [Hoffmann & al. 04].

²⁰ Reference Model for Open Distributed Processing

²¹ eXtensible Access Control Markup Language

comment adapter correctement un document à des contextes d'usage et d'utilisation différents si l'on a perdu les liens reliant l'utilisateur, avec ses rôles et ses tâches, à ces contextes ?

Hormis la définition de la notion de rôle, il reste encore beaucoup à faire sur l'étude de la dynamique des rôles. Cette dynamique se retrouve, lors des contextes d'usage, dans l'émergence et l'évolution des rôles, ces deux aspects influençant très fortement la plasticité des documents qui sont associés à ces rôles. Nous nous intéressons plus particulièrement ici aux relations entre les rôles (hiérarchiques ou non) et à l'évolutivité des rôles (changement de rôles, émergence de rôles, etc.). Au niveau des documents numériques, la notion de rôle peut se traduire de multiples manières. Dans PlasticML, nous associons les rôles de l'utilisateur au sein de chaque bloc d'informations, grâce à l'attribut *role* de la balise *block* (cf. Figure 3).

Les rôles hiérarchiques permettent d'indiquer, grâce à des valeurs numériques, le niveau de visibilité d'une information à un niveau N. Ainsi, une visibilité de niveau N inclut les informations de niveau inférieur (N-1, N-2...). Par exemple, si l'on attribue le nombre 1 au rôle étudiant, et le nombre 2 au rôle enseignant, on indique que dans ce document électronique, l'enseignant (N=2) voit également les informations du niveau étudiant (N=1).

Les rôles non hiérarchiques peuvent être utilisés pour signifier qu'une information sera présentée de manière différente selon le rôle de celui qui consulte le document. Par exemple, dans le domaine de la e-santé, le dossier électronique d'un patient pourra être consulté selon un rôle particulier (membre de la famille, infirmière, médecin, pharmacien, ...). Il n'y a pas ici de notion de hiérarchie. Un médecin n'aura pas forcément un droit d'accès, au dossier d'un patient, supérieur aux membres de la famille du patient, mais l'information devra être présentée de manière adaptée : « traitement contre la douleur » pour certains, « traitement sous morphine » ou « Morphine P.O.²² : 60 mg » pour d'autres, voire « Morphine Per Os à libération immédiate à la dose de 30 mg x 2 par 24 h ».

Cette notion de rôle peut donc être liée aux fonctions des utilisateurs, mais également, de manière plus pragmatique, aux besoins des utilisateurs. On rejoint ici la notion d'assemblage ou de (re)construction de briques et de fragments documentaires (type Lego), expliquée dans [RTP-DOC 03]. L'adaptation sera nécessairement plus efficace lorsqu'elle sera établie en fonction du Device Utilisateur (DU)²³ et de l'utilisateur, simultanément.

²² P.O. est l'abréviation médicale de « par voie orale » (Per Os)

²³ Nous utilisons la définition du « Device Utilisateur » (DU) exposée dans (Chevrin 06). Elle fait référence à la plateforme logicielle, à la plateforme matérielle, mais aussi aux aspects physiques (facteurs de taille, mobilité, portée).

4.2.3. Adaptation conjointe au DU et à l'utilisateur

Dans le cadre du projet UMR²⁴ et de notre collaboration avec le groupe 3 Suisses International, nous avons développé des applications qui adaptent les informations en fonction du DU et de l'utilisateur. On applique donc ici la séparation fond/forme telle que décrite notamment par Bachimont et Crozat, permettant ainsi une adaptation par reconstruction : « *Tout contenu numérique consiste en une ressource qu'un calcul permet de mettre en forme dynamiquement pour sa consultation. On a donc d'un côté une ressource enregistrée et de l'autre une vue publiée reconstruite.* » [Bachimont & Crozat 04].



Figure 6. Adaptation conjointe au DU et à l'utilisateur

Les captures de la Figure 6 montrent une adaptation simultanée pour le DU (ici, un ordinateur personnel versus un assistant PDA) et pour l'utilisateur (ici, une cliente française versus une cliente anglaise). Pour le PC, toutes les colonnes d'informations sont affichées, tandis que pour le PDA, seules certaines colonnes le sont (pour un gain évident de place). Le contenu lui aussi est modifié en fonction du profil de l'utilisateur (texte en différentes langues, monnaies différentes, etc.).

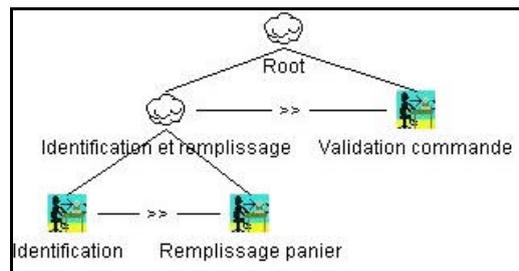
Nous avons jusqu'à présent développé l'idée que la plasticité des documents numériques peut s'effectuer en fonction de critères (relativement) connus à l'avance²⁵ comme le DU permettant l'accès à l'information ou le profil de l'utilisateur, voire le couplage des deux. Nous montrons dans les lignes qui suivent, qu'il est également possible d'intégrer des éléments plus dynamiques, liés à l'activité même, et qui ne sont connus, la plupart du temps, qu'au moment de la prise de connaissance effective du document.

²⁴ Ubiquitous Marketing Relationship

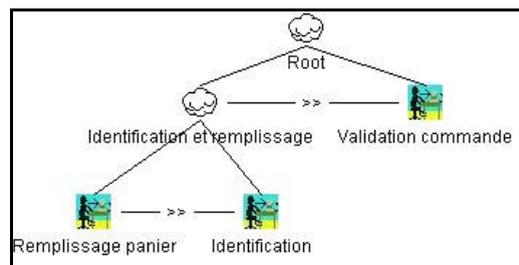
²⁵ Il est vrai que l'on peut, grâce à CC/PP [<http://www.ccpp.org>] découvrir au moment de l'exécution quelles sont les caractéristiques techniques du matériel employé par l'utilisateur.

4.2.4. Adaptation par rapport à l'interaction

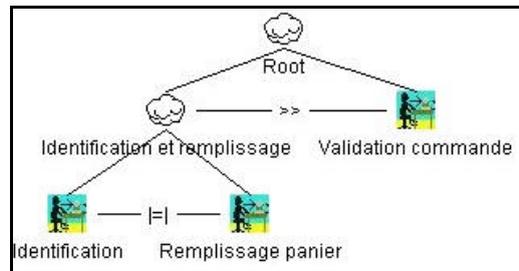
L'utilisation des documents est, bien entendu, fortement liée aux activités humaines (« Je pense, donc j'archive ») [Chabin 99]. Nous estimons, tout comme Beaudoux, que l'interaction (qui fait partie du contexte d'usage) est également une composante importante à prendre en compte pour l'adaptation des documents numériques [Beaudoux 04].



cas 1



cas 2



cas 3

Légende :

|| : choix — » — : séquence ☁ : tâche composite 🌳 : tâche interactive

Figure 7. Vue schématisée de trois cas de composition de services

Au cours de nos travaux sur la plasticité des documents numériques, nous avons essayé d'augmenter le niveau d'abstraction du langage que nous utilisons. En effet, nous avons rajouté à PlasticML des balises permettant de définir des scripts de manière abstraite pour permettre aux systèmes informatiques de générer, à l'exécution, les scripts nécessaires pour chaque cible visée, alors que jusqu'à présent, les scripts accompagnant les langages ciblés n'étaient pas pris en considération.

De même, pour pouvoir rester flexible, notre approche nous amène à avoir une vision axée sur des e-services interactifs (ESI) pour gérer les bases d'informations et de connaissances. Il existe de multiples définitions du terme e-service. Dans notre approche, nous définissons un ESI comme un groupe de tâches utilisateur formant un tout cohérent, de manière à avoir un sens pour son activité, du point de vue de l'utilisateur et du point de vue de son domaine d'utilisation (par exemple, en marketing, nous nous intéressons aux grandes fonctions de l'entreprise dans sa relation avec le client). Un ESI va fournir un flux d'interaction que l'on peut décomposer en un flux de contrôle (boutons, liens, etc.) et un flux de données.

Dans le cadre de cette contribution, nous nous focalisons sur la gestion des documents numériques par le biais d'ESI. C'est l'interaction même qui va guider la composition des ESI entre eux pendant l'exécution. Le document numérique que va percevoir l'utilisateur n'est donc pas figé. C'est au cours même de l'interaction qu'il va s'enrichir (ou s'appauvrir) de différents éléments jugés pertinents (ou non) par le système de composition.

Pour illustrer cela, nous présentons trois cas de composition de services dans le domaine du e-commerce : l'identification, le remplissage du panier et la validation de la commande. La Figure 7 expose de manière schématique ces deux exemples à l'aide de l'outil CTTE [CTTE].

Pour le cas 1, la succession des tâches est la suivante : « Identification », puis « Remplissage du panier » et enfin « Validation de la commande ». Pour le cas 2, l'identification n'est faite qu'après remplissage du panier. Ces deux cas sont typiques de la composition de services interactifs « audiotel » actuels (Redoutel, Chouchoutel, Blantel, etc.).

Un troisième cas, plus flexible, peut laisser l'utilisateur choisir de s'identifier d'abord puis de remplir son panier ensuite (ou bien l'inverse), et non pas l'imposer comme dans les cas 1 et 2. Ce cas reflète la composition de services interactifs rencontrés habituellement sur les sites Web de commerce électronique.

La composition des différents services a donc un impact sur la présentation des documents numériques générés. Dans le cas 1, le système pré-remplit certains champs du document (les réductions, etc.) puisque l'utilisateur a dû s'identifier avant de remplir son panier. En effet, dès lors qu'il y a identification, toutes les informations

issues de la GRC²⁶ peuvent être exploitées (date anniversaire du client donnant droit à 25% de réduction, ...).

Ces cas mettent en évidence les choix possibles en ce qui concerne la composition de services en vue d'adapter les documents numériques en fonction de l'interaction.

Après avoir effectué une première implémentation *ad hoc* à l'aide d'EJB²⁷ et de servlets, nous avons réalisé une deuxième implémentation en utilisant le langage BPEL²⁸ [BPEL] [OASIS].

Au cours de cette deuxième implémentation, nous avons utilisé un moteur d'orchestration de services Web développé par Oracle pour transformer des documents en fonction de plusieurs critères fixes dont les valeurs ne seront connues qu'à l'exécution (exemple : critère « Périphérique », valeur « Téléphone », « PDA », etc.).

5. Conclusion et perspectives

Une des réflexions émergentes faisant suite aux recherches initiées au niveau national par le collectif RTP-DOC nous a incité à approfondir nos études concernant l'adaptation des documents aux contextes d'usage et d'utilisation pour envisager une meilleure utilisation des documents numériques, en fonction de critères établis.

Nous nous sommes plus particulièrement intéressés aux modalités d'adaptation via les terminaux utilisés, la tâche à réaliser, ainsi que les profils, rôles et privilèges des utilisateurs (cf. contextes physique, humain, social et organisationnel [Beaudouin-Lafon 00]).

Nous avons montré dans cet article l'intérêt de disposer de langages de programmation capables de décrire un document à un haut niveau d'abstraction, de manière à pouvoir composer (et recomposer) un document adapté au contexte. Parmi les multiples adaptations possibles d'un document numérique, nous nous sommes concentrés ici sur trois aspects qui nous semblent importants à prendre en compte :

- (1) le DU employé par l'utilisateur pour prendre connaissance du document ;
- (2) le profil et les différents rôles que peut jouer l'utilisateur, pour une activité donnée ;
- (3) la tâche à réaliser par l'utilisateur, qui permet de contextualiser le document et de le rendre plus pertinent au sein d'une interaction globale.

²⁶ Gestion de la Relation Client

²⁷ Enterprise JavaBeans

²⁸ Business Process Execution Language

La notion de plasticité vue sous l'angle de l'IHM est relativement récente. Il s'agit de permettre à l'application de s'adapter au contexte, à une situation particulière, parfois même inédite, aux usages, aux utilisateurs, etc.

Nous n'avons pas détaillé ici les règles de transformation mises en œuvre, mais nous avons présenté brièvement quelques exemples de génération automatique de documents numériques, à partir du langage générique PlasticML. Certains langages basés sur XML permettent de réaliser ce type d'opération, mais il demeure quelques contraintes fortes lors de la mise en œuvre de ces outils. De plus, la conception du document dit abstrait (ou générique) reste encore souvent un travail que le concepteur effectue à la main. Ce dernier doit être expérimenté et connaître suffisamment les langages cibles sélectionnés, le code source généré n'étant pas toujours utilisable tel quel (erreurs de transcription, incompatibilités, non respect de la norme en vigueur, ...).

Nous avons vu qu'il est parfois nécessaire d'améliorer les outils et les langages abstraits employés pour qu'ils prennent en compte des fonctionnalités importantes lors de l'interaction. Ce fut le cas avec l'ajout dans PlasticML de balises gérant les scripts, afin de pouvoir obtenir des objets pédagogiques plastiques.

Nos travaux actuels adressent les principes de multimodalité et de couplage de multiples canaux de communication. Nous avons élaboré des prototypes, dans le domaine du commerce électronique, et nous avons dû résoudre, comme c'est souvent le cas lors de l'utilisation de plusieurs canaux de communications, des problèmes de synchronisation, pour maintenir la cohérence globale des documents présentés.

Il nous reste à étudier plus en profondeur les avantages et inconvénients des différents modes de composition (juxtaposition, séquence, concurrence ..) des canaux d'interaction, afin de déterminer les couplages optimaux en fonction du contexte d'usage, tel que nous l'avons défini dans cet article. De plus, il est envisagé d'étudier les éventuelles perturbations que pourraient occasionner la plasticité des documents numériques sur les intentions de départ de l'auteur. Par exemple, l'intention pédagogique du créateur est-elle visible ou non, conservée ou modifiée, voire perdue, lors de l'utilisation du document ?

6. Remerciements

Les auteurs tiennent à remercier le groupe TAC MIAOU (Modèles d'Interaction et Architectures Orientées Usages) ainsi que le FEDER pour son soutien dans ces travaux de recherche, ainsi que les relecteurs de cet article, pour leurs critiques, intéressantes et constructives.

7. Bibliographie

- [Affect] MIT Media Lab, Affective Computing Group, <http://affect.media.mit.edu>
- [Bachimont & Crozat 04] Bachimont, B., Crozat, S., « Instrumentation numérique des documents : pour une séparation fonds/forme », *Revue Information - Interaction - Intelligence (RI3), numéro spécial sur le document numérique*, Volume 4, Numéro 1, 2004.
- [Beaudouin-Lafon 00] Beaudouin-Lafon, M., « Ceci n'est pas un ordinateur. Perspectives sur l'Interaction Homme-Machine ». *Technique et Science Informatique*, 19(1-2-3):69-74, 2000.
- [Beaudoux 04] Beaudoux, O., « Un modèle de composants (inter)actifs centré sur les documents », *Revue Information - Interaction - Intelligence (RI3), numéro spécial sur le document numérique*, Volume 4, Numéro 1, 2004.
- [BPEL] Business Process Execution Language for Web Services, Version 1.1, 5 May 2003, <ftp://www6.software.ibm.com/software/developer/library/ws-bpel11.pdf>
- [Buckland 98] Buckland, M., « What is a digital document ? », *Document numérique*, 02/1998.
- [Calvary & al. 01a] Calvary, G., Coutaz, J., Thevenin, D., « Supporting Context Changes for Plastic User Interfaces: Process and Mechanism », *HCI-IHM*, Lille, 2001.
- [Calvary & al. 01b] Calvary G., Coutaz J., Thevenin D., «A Unifying Reference Framework for the Development of Plastic User Interfaces », *EHCI'01, IFIP WG2.7 (13.2) Working Conference*, Toronto, May, 2001.
- [CAMELEON] <http://giove.cnuce.cnr.it/comeleon.html>
- [Chabin 97] Chabin, M.-A., « La cinquième dimension de l'archive numérique », *Document numérique*, Vol. 1, No. 2, 1997, pp. 205-216.
- [Chabin 99] Chabin, M.-A., « *Je pense donc j'archive. L'archive dans la société de l'information*, L'Harmattan, 1999.
- [Chevrin 06] Chevrin, V., « *L'Interaction usagers/services, multimodale et multicanale : une première proposition appliquée au domaine du e-Commerce* », Thèse en informatique de l'université de Lille1, 2006.
- [Coutaz & Thévenin 99] Coutaz J., Thevenin D. (1999), « Plasticity of User Interfaces : Framework and Research Agenda », *Actes du colloque Human-Computer Interaction - INTERACT'99*, 1999
- [CTTE] <http://giove.cnuce.cnr.it/concurtasktrees.html>
- [Dey 01] Dey, A.K., « Understanding and Using Context », *Personal and ubiquitous computing*, Volume 5, February 2001, pp 4-7.
- [Dey & Abowd 00] Dey, A.K., Abowd, G. D., « Towards a Better Understanding of Context and Context-Awareness », in *Proceedings of CHIA'00 workshop on Context-Awareness*, 2000.

- [Eisenstein & al. 01] Eisenstein, J., Vanderdonckt, J., Puerta, A. « Applying Model-Based Techniques to the Development of UIs of Mobile Computers », in *Proceedings of Intelligent User Interfaces*, 2001, Santa Fe NM, 14-17 January 2001, pp. 69-76.
- [Gallezot 02] Gallezot G. : http://www.unice.fr/urfist/DOCNUM/Gabriel_Gallezot.html
- [Grundy & Zou 04] Grundy, J., et Zou, W., « AUIT: Adaptable User Interface Technology, with extended Java Server Pages ». in A. Seffah & H. Javahery (eds). *Multiple User Interfaces*. John Wiley & Sons, New York, 2004, 149-167.
- [HARMONIA] <http://www.harmonia.com>
- [Hoffmann & al. 04] Hoffmann, U., Garcia, G.N., Vesin J.-M. et Ebrahimi, T. « Application of the Evidence Framework to Brain-Computer Interfaces », in *Proceedings of the IEEE Engineering in Medicine and Biology Society Conference*, 2004.
- [Keim 01] Keim, D., « Visual Exploration of large data Sets », *Communications of the ACM*, Vol. 44., N. 8, 2001, p. 39-44.
- [Limbourg & al. 04] Limbourg, Q., Vanderdonckt, J., Michotte, B., Bouillon, L., Florins, M., Trevisan, D., « UsiXML: A User Interface Description Language for Context-Sensitive User Interfaces », in *Proceedings of the ACM AVI'2004 Workshop "Developing User Interfaces with XML: Advances on User Interface Description Languages"* (Gallipoli, May 25, 2004), K. Luyten, M. Abrams, Q. Limbourg, J. Vanderdonckt (Eds.), Gallipoli, 2004, pp. 55-62.
- [Lin & al. 01] Lin J., Sinha A., Landay J., « Universal Access Through Multimodal Applications », in *Workshop on Transforming the UI for Anyone, Anywhere, CHI 2001*, Seattle, WA, April 2001.
- [Lopez & Szekely 01] Lopez J., Szekely P., « Automatic web page adaptation », in *Workshop on Transforming the UI for Anyone, Anywhere, CHI 2001*, Seattle, WA, April 2001.
- [Merrick 01] Merrick, R. A., « Device Independent User Interfaces in XML », *BelCHI*, 2001.
- [Mitrovic & al. 04] Mitrovic, N., Royo, J.A., Mena, E., « ADUS: Indirect Generation of User Interfaces on Wireless Devices », *Fifteenth International Workshop on Database and Expert Systems Applications (DEXA 2004)*, *Seventh International Workshop Mobility in Databases and Distributed Systems (MDDS 2004)*, Zaragoza (Spain), IEEE Computer Society, ISBN 0-7695-2195-9, ISSN 1529-4188, September 2004/2004.
- [Müller & al. 01] Müller A., Forbrig P., Cap C., « Using XML for Model-based User Interface Design » in *Workshop on Transforming the UI for Anyone, Anywhere, CHI 2001*, Seattle, WA, April 2001.
- [Myers & al. 02] Myers, B., Hudson, S., Pausch, R. « Past, Present, Future of User Interface Tools ». *Transactions on Computer-Human Interaction*, ACM, 7(1), 2000, pp. 3-28.
- [Nanard & Nanard 01] Nanard J., Nanard M., « Principes hypermédias et systèmes d'information », in *Environnements Evolués et Evaluation des IHM, Interface Homme-Machine pour les Systèmes d'Information* (Vol. 2), C. Kolski (Dir.), Hermès, ISBN : 2-7462-0243-3, Avril 2001, pp. 147-173.

- [Nigay 01] Nigay, L., « *Modalité d'Interaction et Multimodalité* », Habilitation à Diriger des Recherches, spécialité Informatique de l'Université Joseph Fourier - Grenoble I, 2001.
- [OASIS] <http://www.oasis-open.org/>
- [Paternò & Santoro 02] Paternò, F. and Santoro, C « One Model, Many Interfaces », *CADUI'02*, Valenciennes, 2002.
- [Putman 01] Putman, J. R., « *Architecting with RM-ODP*, Prentice Hall, 2001.
- [Rouillard 03] Rouillard, J., « PlasticML and its toolkit », *HCI International 2003*, Lawrence Erlbaum Associates, Inc, Heraklion, Greece, 2003.
- [Rouillard 04] Rouillard, J., « *VoiceXML, Le langage d'accès à Internet par téléphone* », éditions Vuibert, 2004, 197 pages.
- [RTP-DOC 03] Roger T. Pédaque (pseudonyme pour le collectif RTP-DOC), « Document : forme, signe et médium, les re-formulations du numérique », *STIC-CNRS*, version 3, du 08/07/03.
- [Signer & al. 02] Signer, B., Norrie, M. C., Geissbuehler, P., Heiniger, D., « Telephone Interface for Avalanche Warnings based on Information Server for Adaptable Content Delivery », *Pervasive 2002*, International Conference on Pervasive Computing, Zurich, 2002.
- [Spriestersbach & al. 03] Spriestersbach, A., Ziegert, T., Grassel, G., Wasmund, M., Dermler, G., « A single source authoring language to enhance the access from mobile devices to Web enterprise applications », *WWW'03 Developers Day Mobile Web Track, 12th World Wide Web Conference WWW'03*, Budapest, 2003.
- [Stern 97] Stern, Y., « Les quatre dimensions des documents électroniques », *Document numérique*, Vol. 1, No. 1, 1997, pp. 55-60.
- [Tarby 04] Tarby, J.-C., « One Goal, Many Tasks, Many Devices: From Abstract User Task Specification to User Interfaces », in *The Handbook of Task Analysis for Human-Computer Interaction*, Dan Diaper (ed.) and Neville Stanton (ed.), Lawrence Erlbaum Associates, 2004, pp. 531-550.
- [Thevenin 01] Thevenin, D., « *Adaptation en Interaction Homme-Machine : le cas de la Plasticité* ». Thèse pour l'obtention du titre de docteur en informatique, Université Joseph Fourier, Grenoble I, 2001.
- [UIML] User Interface Markup Language, <http://www.uiml.org>
- [UsiXML] USeR Interface eXtensible Markup Language, <http://www.usixml.org/>
- [VoiceXML] VoiceXML Forum, <http://www.voicexml.org>
- [Vanderdonckt & al. 04] Vanderdonckt, J., Limbourg, Q., Michotte, B., Bouillon, L., Trevisan, D., Florins, M., « UsiXML: a User Interface Description Language for Specifying Multimodal User Interfaces », in *Proc. of W3C Workshop on Multimodal Interaction WMI2004*, Sophia Antipolis, 2004.
- [WAP] <http://www.WAPforum.org>

[XACML] eXtensible Access Control Markup Language, <http://www.oasis-open.org>

[XIML] eXtensible Interface Markup Language, <http://www.ximl.org>

[XML] <http://www.w3.org/XML>

[XSL] <http://www.w3.org/Style/XSL>

8. Biographies

José Rouillard est Maître de Conférences à l'Université des Sciences et Technologies de Lille et membre de l'équipe NOCE du laboratoire TRIGONE. Il a soutenu sa thèse en informatique, préparée au laboratoire CLIPS-IMAG de Grenoble, en 2000, sur le thème du Dialogue Homme-Machine multimodal et en langue naturelle sur Internet. Il est actuellement impliqué dans divers projets et groupes de travail régionaux (MIAOU, FERNAHTI, EUCUE) et nationaux (GT ACA, GT CESAME). Il a codirigé la thèse de Vincent Chevrin. Ses domaines de recherche actuels s'articulent autour de la plasticité des IHM (il est à l'origine du langage PlasticML), de la mobilité, des traces et des évaluations des interfaces, de la multimodalité/multicanalité, du DHM et des interfaces vocales. Il est d'ailleurs l'auteur du seul ouvrage francophone publié à ce jour à propos du langage VoiceXML (le langage d'accès à Internet par téléphone).

Jean-Claude Tarby est Maître de Conférences à l'Université des Sciences et Technologies de Lille et membre de l'équipe NOCE du laboratoire TRIGONE. Ses travaux de thèse présentés en 1993 à Toulouse ont trait à l'Interaction Homme-Machine, et plus particulièrement aux méthodes de conception orientées tâches. Il est membre de l'Association Francophone d'Interaction Homme-Machine (AFIHM) depuis sa création en 1996. Il est impliqué dans divers projets et groupes de travail régionaux (MIAOU, FERNAHTI) et nationaux (GT CESAME). Ses domaines de recherche actuels sont centrés autour des méthodes de conception : modèles de tâches, approches dirigées par les modèles, plasticité des IHM, traces et usages, TCAO.

Vincent Chevrin est docteur en informatique, membre de l'équipe NOCE du laboratoire Trigone de l'Université des Sciences et Technologies de Lille et ATER à l'Ecole Polytechnique Universitaire de Lille. Ses travaux de thèse présentés en 2006 ont trait aux interactions multicanales usagers/services et s'inscrivent notamment dans le cadre du e-commerce. Il a participé à des programmes de recherche collaboratives menés sur le plan régional et national.

Alain Derycke est professeur à l'université des sciences et technologies de Lille et codirecteur du laboratoire pluridisciplinaire TRIGONE, qui possède deux équipes de recherche: l'une en informatique et l'autre en sciences de l'éducation. Après une première activité de recherche en télécommunication, dans une unité du CNRS, l'ayant conduit à une thèse d'état en sciences physique en 1986, Alain Derycke Derycke a évolué vers les sciences et technologies de l'information et de la

communication, avec un intérêt marqué pour les Interfaces Hommes-Machines, le Travail et l'Apprentissage Collaboratifs Assistés par Ordinateur, les technologies éducatives et le e-commerce. Il a participé à de nombreux contrats de recherche sur le plan français et européen, et il est l'un des membres fondateurs, en 2003, du Réseau d'Excellence Européen Kaléidoscope sur les technologies avancées pour l'apprentissage humain. Réseau dont il a fait partie de l'instance de décision de sa création à décembre 2005. Il a eu des responsabilités scientifiques dans des programmes de recherches collaboratives menées au niveau régional, et il porte actuellement des projets de recherche et de développement, en lien avec les entreprises aussi, autour des usages des technologies dans les contextes ubiquitaires et mobiles, dans le cadre du Pôle de Compétitivité « Industries du Commerce ».

Améliorer ses performances grâce au coaching numérique mobile : l'exemple du coaching vocal embarqué

Jean-Claude Tarby, José Rouillard

Laboratoire Trigone, Institut CUEEP

Université Lille 1

59655 Villeneuve d'Ascq Cedex, France

tel. : (33)3 20 43 32 70 fax : (33)3 20 43 32 01

{ jean-claude.tarby, jose.rouillard } @univ-lille1.fr

RESUME

Nous présentons dans cet article la notion de *coach numérique*. Plus qu'une assistance traditionnelle, le coach numérique permet à l'utilisateur d'améliorer ses performances, en le guidant, étape par étape, et en le stimulant, notamment de manière vocale, grâce à des informations personnalisées préparées spécifiquement en fonction de chaque utilisateur (textes et voix de synthèse dans sa langue). Ces informations sont exploitées par une application fonctionnant sur un périphérique mobile, par exemple un SmartPhone. L'utilisateur est ainsi coaché numériquement et individuellement pendant l'activité, et non plus au préalable et de manière générique. Les domaines d'applications sont multiples (sport, cuisine, bricolage, culture, tourisme, ...) et les moyens d'accroître les performances individuelles, voire collectives, sont encore nombreux : GPS, capteurs sensoriels, contexte de l'activité... Une définition des coaches numériques ainsi qu'une première taxinomie sont proposées dans cet article. Le concept est ensuite illustré grâce à un prototype de coach embarqué sur périphérique mobile.

Mots clés

Coach numérique, personnalisation, assistance, performance humaine, synthèse vocale, application embarquée.

ABSTRACT

We present in this paper the concept of *digital coach*. More than a classical assistance, the digital coach allows the user to improve her/his performances by guiding her/him step by step, and by stimulating her/him, in particular with voice, thanks to personalized information specifically prepared according to the user (texts and speech synthesis in her/his mother tongue). This information is used by an application running on a mobile device, for example a SmartPhone. Then the user is coached individually during the activity and not before the activity or in a generic way. This principle can be applied to a lot of domains (sport, cooking,

do-it-yourself, culture, tourism...) and the means of increasing the individual and collective performances are numerous: GPS, sensors, activity context... A definition and a first taxonomy of digital coach are proposed in this article. The concept is then illustrated with a prototype of digital coach on a mobile device.

Categories and Subject Descriptors

C.2.4 [Computer Systems Organization]: Computer-Communication Networks – *Distributed Systems*

General Terms

Design, Reliability, Experimentation.

Keywords

Digital coach, personalization, assistance, human performance, speech synthesis, embedded application.

1. INTRODUCTION

Il est difficile aujourd'hui d'avoir « avec soi » une application qui sache nous conseiller, nous guider, nous assister, nous motiver, etc., dans nos activités. Ces actes sont l'apanage des coaches auxquels le recours est en forte progression depuis quelques années aussi bien dans le milieu professionnel (gestion d'équipe par exemple) qu'à titre individuel (gestion de la confiance en soi, conseils pour l'alimentation, etc.). Ces coaches adaptent leurs expertises pour *accompagner leurs clients de façon personnalisée*. La notion de personnalisation est aujourd'hui plus forte que jamais en informatique : personnalisation par « skin », par feuille de style CSS (Cascade Style Sheet), etc., mais aussi personnalisation des informations par accès réservé (login/mot de passe), par abonnement (choix de thèmes sur les sites d'information tel que Le Monde), par « syndication » avec les flux RSS (Really Simple Syndication), par espionnage (cookie, suivi de l'historique des visites), etc. La personnalisation touche également le domaine des appareils mobiles : logo, sonnerie, fond d'écran, coque d'habillage... tout est personnalisable.

Ces dix dernières années, l'informatique a connu des progrès considérables dans trois domaines qui nous concernent plus particulièrement dans cet article. Le premier est celui de la mobilité grâce à qui l'informatique est devenue pervasive et ubiquitaire. Le second est celui de l'assistance à l'utilisateur, et plus particulièrement les « aides intelligentes » basées sur des agents dits « intelligents », par exemple le « trombone » de Microsoft®. Le troisième concerne les voix de synthèse et les commandes vocales qui peuvent être embarquées maintenant sur

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

UbiMob'06, September 5-8, 2006, Paris, France.

Copyright 2006 ACM 1-59593-467-7/06/0009�\$5.00.

des périphériques mobiles. L'évolution des aides et assistances dites « intelligentes », la présence de plus en plus forte de la personnalisation dans les applications interactives, l'omniprésence des périphériques mobiles, et la qualité sans cesse croissante des synthèses vocales, nous ont conduit à proposer la notion de *coach numérique*. Notre conviction est que les techniques de mobilité couplées aux techniques d'aide intelligente et de personnalisation permettront de réaliser les *applications mobiles coachantes de demain*. Un coach numérique est plus qu'une aide intelligente. Ses caractéristiques sont semblables à celles d'un coach réel, mais il bénéficie en plus des avantages liés à la technologie informatique.

Cet article présente tout d'abord le concept de coach numérique. Ensuite, nous présentons le prototype que nous avons développé pour montrer la faisabilité de nos idées. Puis, nous concluons et donnons nos perspectives.

2. LE CONCEPT DE COACH NUMÉRIQUE

2.1 Scénario d'usage : un coach de « footing »

Après une longue période d'interruption, et en accord avec son médecin, Marc a décidé de se remettre en forme par le footing. Il se connecte sur un site web de remise en forme avec coaching numérique. Là, il s'enregistre et remplit ses données personnelles telles que sexe, âge, poids, taille... mais aussi des données sur sa pratique antérieure du sport, des données sur sa qualité de vie (tabac, boisson...), etc. Ensuite, il indique le type d'entraînement qu'il désire : durée souhaitée de la remise en forme, niveau à atteindre à la fin (remise en forme légère, course de 10km dans 3 mois...), etc. Il indiquera également l'endroit où il habite (un entraînement à Lille ou à Chamonix n'aura pas le même contenu !). D'autres renseignements pourront lui être demandés, par exemple s'il préfère la voix d'un homme ou d'une femme (échantillons à l'appui) et quels genres de musiques le relaxent ou le stimulent. Cette phase de renseignements terminée, le système génère un programme d'entraînement composé d'un planning détaillé, ainsi que d'un ensemble de fichiers MP3, associés à chaque jour, que Marc téléchargera depuis le site sur son baladeur MP3. Avant chaque séance, Marc consulte son planning d'entraînement et les activités proposées. Lorsque l'heure d'une séance approche, il endosse sa tenue de sport, se rend sur le lieu qui lui a été indiqué dans le planning, chausse les écouteurs de son baladeur MP3, et met en route ce dernier. Voici un extrait de ce que Marc entendra le premier, prononcé par la voix qu'il a choisie : « Bonjour Marc. Votre but aujourd'hui est de reprendre un rythme cardiaque suffisant pour passer aux étapes des jours suivants. Nous allons commencer par un échauffement. Marchez en inspirant profondément pendant 100 mètres, puis accélérez le pas progressivement pendant environ 1km, soit 15minutes. Pour vous y aider, vous allez entendre une musique durant le temps nécessaire. Elle vous aidera à rythmer votre pas, et accélérera pour vous amener à la vitesse voulue ».

2.2 Définition du coach numérique

« Le coaching [...] se définit comme une relation continue qui permet au client d'obtenir des résultats concrets et mesurables dans sa vie professionnelle et personnelle. *A travers le processus de coaching, le client approfondit ses connaissances, améliore ses performances.* Grâce à l'interaction avec le coach, le client clarifie ses objectifs et s'engage dans l'action. L'accompagnement permet

au client de progresser plus rapidement vers les résultats visés, car la relation de coaching l'invite à se centrer sur ses priorités et à être conscient de ses choix » [7]. Notre travail consiste à implanter la réplique d'un coach réel sous la forme d'un coach numérique dans un système informatique, ce système pouvant être une application, un site web, ou tout autre type de réalisation informatique. Le coaching que nous mettons en œuvre dans notre travail est la traduction fidèle de la définition donnée précédemment. Le coach et son client (c'est-à-dire l'utilisateur qui est coaché) se « connaissent » et se « rencontrent ». Suivant le type de coaching demandé (embarqué, intrusif, interactif...), l'utilisateur aura des contacts plus ou moins réguliers avec son coach. Ceci peut se faire de façon explicite (l'utilisateur se connecte à son coach et fait le point avec lui pour une éventuelle mise à jour de son coaching) ou implicite (les actions de l'utilisateur sont renvoyées automatiquement au coach qui prend alors des décisions quant à des modifications éventuelles du coaching). Le coaching numérique, tel que nous le concevons dans nos travaux, peut s'appliquer à des domaines aussi variés que le sport, la cuisine, le bricolage, la culture, le tourisme, etc. Nous englobons dans notre définition du coaching numérique la notion de coaching telle que définie précédemment, mais aussi les notions de guide, d'assistant, de conseiller, d'encadrant, et de relais (voir l'exemple du voyage aux USA plus loin).

2.3 Caractéristiques du coach numérique

Expert du domaine. Comme un coach réel, le coach numérique doit connaître parfaitement son domaine. Cette caractéristique n'est pas spécifique au coach numérique. Toutes les techniques de gestion de connaissances peuvent être utilisées pour élaborer cette expertise.

Relation personnelle avec son client. Le coach et son client sont liés par une relation personnelle privilégiée. Cette relation permet au coach d'adapter son discours tant sur la forme (tournure de phrase, intonation, style de musique...) que sur le fond en s'adaptant au contexte réel du client (lieu nommé pour un exercice sportif, matériel personnel pour du bricolage...), mais aussi à ses goûts, ses préférences, etc.

Embarqué/centralisé. Un coach centralisé est un coach hébergé sur un serveur web ou toute autre type d'application centrale. A l'opposé, un coach embarqué est un coach que l'on emporte facilement avec soi grâce aux périphériques mobiles.

Réactif. Puisqu'il est possible de stocker l'historique des contacts entre le coach et son client et par conséquent de connaître l'évolution du coaching lui-même (progrès et régressions de l'utilisateur, taux de réussite, etc.), le coach peut encore mieux adapter son discours et ses actions vis-à-vis de l'utilisateur.

Statique/dynamique. Un coach statique génère ses recommandations avant qu'elles ne soient utilisées, par exemple dans le scénario d'usage cité précédemment. Un coach dynamique au contraire les choisit, les adapte et/ou les génère en temps réel, éventuellement en se basant sur des données dynamiques telles que la position GPS de l'utilisateur, le nombre de personnes dans une salle de musée, etc.

Intrusif ou non. Suivant la volonté de l'utilisateur, le coach peut être intrusif ou non. Dans le premier cas, le coach peut intervenir sans que l'utilisateur l'ait demandé, alors que dans le second cas l'utilisateur doit explicitement appeler le coach pour que celui-ci intervienne.

Interactif. L'interactivité avec le coach est de deux types. Tout d'abord, dans certaines situations, il est intéressant que le coach propose des choix. Dans ce cas, l'utilisateur choisira en fonction de ses *desiderata* ou du contexte au moment de l'activité. Par exemple, pour un coaching sur l'assemblage d'un meuble en kit, le coach peut dire : « Si vous êtes seul, passez à la piste suivante. Si vous êtes deux, passez à la piste 5 ». Un second type d'interactivité peut être ajouté par des commandes vocales, y compris avec un PDA¹, permettant ainsi à l'utilisateur de commander son coach à la voix avec des ordres génériques tels que : « lis tout » (lecture automatique sans toucher le clavier... pratique lorsqu'on a les mains occupées), « lis l'étape 5 » (accès à une étape par son numéro), « lis l'étape parlant de... » (accès à une ou plusieurs séquences contenant un mot-clé²), « stop/pause/lecture », etc. Ces commandes vocales peuvent être complétées par une IHM.

2.4 Avantages du coaching numérique

Dans le coaching numérique, *l'utilisation majoritaire de la voix* (par synthèse vocale ou à partir d'enregistrements réels) est primordiale. Entendre des consignes plutôt que de les lire *libère la vue et les mains*, ce qui permet de les utiliser pour d'autres tâches (très intéressant pour le sport, le bricolage, ou la cuisine). De plus, la synthèse vocale a fait d'énormes progrès, et les voix de synthèse sont aujourd'hui capables de *produire des émotions*, presque aussi bien que des voix humaines [9][11]. Entendre une voix « humaine » qui *nous parle personnellement* produit plus d'émotions qu'un texte donnant des consignes impersonnelles.

Nous avons développé un système à base de synthèse vocale qui est capable de s'exprimer dans plusieurs langues. Notre coaching numérique est donc *multilingue*³, et *polyglotte*⁴. Il est multilingue car nous disposons de voix de synthèse dans différentes langues étrangères, permettant ainsi de faire prononcer un texte comme le ferait un natif du pays. Notre système est polyglotte car parce qu'il peut prononcer des phrases mélangeant différentes langues. Ceci est intéressant par exemple dans le cas où l'utilisateur ne maîtrise pas la langue du pays dans lequel il doit se rendre lors d'un voyage. Ainsi, le coach numérique peut aider cette personne à préparer son voyage en lui donnant des informations sur les formalités administratives à accomplir (visa, passeport, etc.), des conseils divers et variés, etc., mais le coach numérique peut faire beaucoup plus en aidant l'utilisateur à communiquer lors de phases critiques de son voyage, par exemple pour prendre un taxi ou pour obtenir la chambre qu'il a réservée. Le coach numérique peut dans ce cas adopter différentes stratégies. Tout d'abord, le coach numérique peut servir de « personne relais » en *parlant à la place de l'utilisateur*. Par exemple, en entrant dans le taxi, l'utilisateur fait entendre la voix de son coach qui dira au conducteur « Take me to the Hilton hotel, 3000 Paradise Road » avec une prononciation impeccable. De même, à l'hôtel, l'utilisateur fera dire au coach : « Hi ! My name is Marc Denis. I've reserved a room. The confirmation number is XNK078 », etc.

¹ www.microsoft.com/windowsmobile/downloads/voicecommand

² Ceci est réalisable car la synthèse vocale est produite à partir de fichiers textes sur lesquels on peut lancer une indexation, et une recherche de mots ou d'expressions.

³ Multilingue : qui parle trois langues ou davantage, apprises en tant que langues maternelles.

⁴ Polyglotte : qui parle plusieurs langues.

Une seconde stratégie de la part du coach numérique consiste à *encourager l'utilisateur à parler*. Le coach peut, par exemple, proposer un entraînement sur plusieurs jours avant le départ. Ce programme d'entraînement est à rapprocher de l'exemple du footing car il sera basé sur une connaissance approfondie du niveau de départ de l'utilisateur et du niveau à atteindre. Avec la dernière stratégie, le coach *accompagne l'utilisateur au moment opportun*. Dans le cas où l'utilisateur a suivi l'entraînement évoqué ci-dessus, ou bien s'il estime que son niveau est insuffisant, le coach peut jouer un rôle d'accompagnateur au moment fatidique. Ainsi, lorsque l'utilisateur devra prononcer une phrase clé (par exemple « Take me to... »), il entendra son coach lui *murmurer* la phrase, ceci afin d'avoir un repère sonore non perturbateur.

3. REALISATION ACTUELLE

Pour valider nos idées, nous avons réalisé un prototype générique, dans le sens où il n'est pas dédié à une activité spécifique. Ce prototype correspond à la partie vocale (production et utilisation) du coach. Nous pouvons utiliser ce prototype dans différents domaines tels que le sport, la mécanique auto, la cuisine, etc. Notre prototype est constitué de deux applications. Une première application sur PC génère les fichiers nécessaires à la tâche à réaliser⁵, et une autre, embarquée sur périphérique mobile, les exploite. Le prototype a été développé en langage C#, sous Visual Studio 2005. La bibliothèque utilisée pour la synthèse vocale est ActiveTTS⁶, et celle utilisée pour jouer les fichiers audio est Hekkus [6] pour Pocket PC. Le fonctionnement du prototype se déroule en deux temps (cf. Figure 1). Dans un premier temps, on fournit à l'application PC un fichier texte, comptant autant de paragraphes que d'étapes à réaliser. L'application génère alors N fichiers textuels, au format .txt, ainsi que N fichiers vocaux, au format .wav ou .mp3 par exemple. L'utilisateur peut sélectionner les caractéristiques techniques de la voix de synthèse qu'il souhaite appliquer. Dans notre prototype, il est possible de préciser le genre de voix (homme, femme, enfant...), la langue (français, anglais, italien...), le volume, la vitesse d'élocution, la tessiture de la voix, et le format audio voulu (par exemple 8 Khz, 8 bits, mono). A terme, ces choix seront effectués de manière (semi)-automatique, selon la tâche à effectuer et selon certains éléments contextuels. Les fichiers textuels et sonores obtenus sont ensuite transférés sur le périphérique mobile. Dans notre étude, nous avons utilisé un Smartphone HP iPAQ HW6500. La seconde application, quant à elle, est embarquée dans le périphérique mobile et exploite ces données en affichant pour chaque étape le texte associé, et en diffusant la synthèse vocale qui l'accompagne. Il sera aisé, dans une prochaine version du prototype, d'ajouter une photo illustrative (des images au format .gif animé, par exemple) ou des vidéos à chaque étape. Le logiciel embarqué est plus qu'un simple lecteur de fichiers sonores. Il présente simultanément les éléments d'une même étape et permet d'interrompre et de reprendre une synthèse vocale, de la rejouer, de passer à une autre étape, de changer la langue de l'interface, etc.

⁵ A terme, l'utilisateur pourra également choisir de se connecter sur un serveur Web, s'il préfère ne pas installer d'application sur sa propre machine.

⁶ <http://www.guangmingsoft.net/activetts>

Dans le cadre de nos applications embarquées prototypes, nous avons effectué des tests avec la synthèse vocale de Loquendo⁷ (avec la voix dite « Juliette »). Cette synthèse vocale nous semble particulièrement adaptée au coaching numérique car elle intègre des éléments prosodiques émotionnels qui, à notre avis, peuvent avoir un impact positif sur le comportement de l'utilisateur. Il est ainsi possible de lui faire jouer différentes « émotions », comme par exemple des étonnements, encouragements, rires, acquiescements, sifflements, soufflements, faibles ou fortes respirations, raclements de gorge, déglutitions, etc.

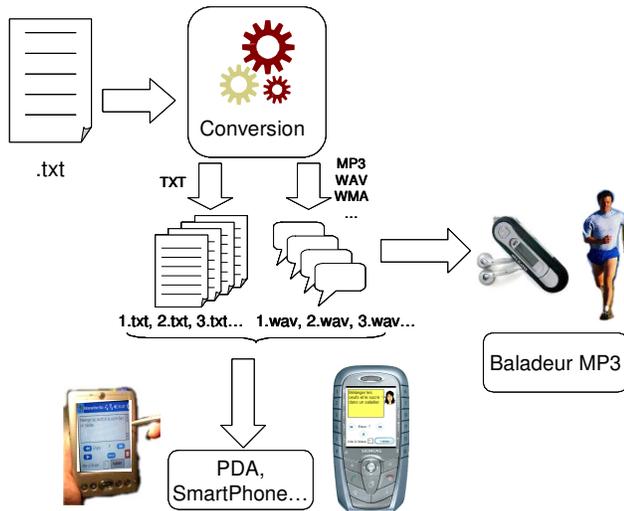


Figure 1 : Principe général du coach numérique

4. CONCLUSION ET PERSPECTIVES

Nous avons proposé dans cet article la notion de coach numérique. Nous avons montré que le coach numérique, et plus particulièrement le coach numérique vocal, permet de disposer d'une assistance personnelle que l'on peut emporter avec soi sur des périphériques mobiles. Cette assistance englobe des capacités d'aide, de motivation, de guide, de conseils, etc. Nous avons dressé une première liste de propriétés caractérisant les coachs numériques, ainsi que la liste des avantages de ce type de coaching. Nous avons donné des exemples d'applications et expliqué le fonctionnement de notre prototype actuel qui montre la faisabilité technique du projet sur le plan vocal.

Nous étudions actuellement l'implantation de commandes vocales ainsi que l'indexation textuelle et vocale en temps réel. Pour cela, nous allons réaliser d'autres prototypes utilisant un serveur vocal (Sibilo d'App-Line⁸) qui supporte VoiceXML 2.0 et qui dispose de reconnaissance et de synthèse vocales Loquendo. Par ailleurs, nous testons actuellement le « player » Lite 2.0 de Flash 8 dans le but de générer nos coachs directement en Flash pour périphériques mobiles.

A terme, on peut imaginer que le coach numérique sera en relation avec des capteurs (pouls, lumière, chaleur, etc.) pour s'adapter en temps réel au contexte d'usage et aux réactions de l'utilisateur. Les travaux sur les agents intelligents [8] et sur le dialogue

homme-machine [13] devraient permettre, dans le futur, de faire ressembler de plus en plus un coach numérique à un coach réel.

5. REMERCIEMENTS

Les auteurs remercient pour leurs supports financiers partiels le programme MIAOU du contrat de plan Etat Région Nord Pas de Calais et le Fonds Européen de Développement Régional (FEDER).

6. BIBLIOGRAPHIE

- [1] Cassell, J., Sullivan, J., Churchill, E. *Embodied Conversational Agents*, MIT Press avril 2000.
- [2] Dufau, H., Perdriset, F. Le coaching, enjeux, paradoxes et perspectives. *La revue Communication et Organisation*, n°28, 2005.
- [3] Eisenstein, J., Rich, C. Agents and GUIs from Task Models, In *Proceedings of Intelligent User Interfaces 2002*, San Francisco, CA, 2002.
- [4] Finin, T. et al. KQML as an Agent Communication Language, *3rd International Conference on Information and Knowledge Management (CIKM94)*, ACM Press, December 1994.
- [5] Gonzales, C., Wuillemin, P.H. Réseaux Bayesiens en modélisation d'utilisateurs, *Sciences et Techniques éducatives Vol.5 n° 2*, pp. 173-198, 1998.
- [6] Hekkus project, <http://www.shlzero.com>
- [7] International Coach Federation France, <http://www.icffrance.org>
- [8] Kolski, C., Grislin-Le Strugeon, E., Adam E., Mathieu P. Conception des systèmes multi-agents : pistes de réflexion en vue de futures coopérations entre ergonomes et informaticiens. In *Actes du congrès ERGO-IA'2004 "Ergonomie et Informatique Avancée"*, ESTIA, Biarritz, pp. 147-154, novembre 2004.
- [9] Loquendo, http://vhost.oddcast.com/vhost_minisite/support/ssml/expres_sive_cues.php?voice=4
- [10] Louart, P. Le coaching : son intérêt, ses méthodes. *Les Cahiers de la Recherche*. Claree, IAE de Lille, 2002.
- [11] Miriam et Léon, <http://www.timespace.com/french/miriam.asp>
- [12] Nicholson, A., Korb, K. *Bayesian Artificial Intelligence*, Addison-Wesley, juin 1984.
- [13] Sadek, D., Nouvelles perspectives pour l'ergonomie des interactions personne-machine : dialogue naturel et agents intelligents, In *Actes du congrès ERGO-IA'2004 "Ergonomie et Informatique Avancée"*, ESTIA, Biarritz, novembre 2004.
- [14] Wooldridge, M. J., and Jennings, N. R. Intelligent agents: Theory and practice, *Knowledge Engineering Review 10(2)*, 1995.

⁷ <http://www.loquendo.com>

⁸ <http://www.app-line.com>

Pourquoi il n'est pas possible actuellement d'utiliser des objets pédagogiques plastiques respectant la norme SCORM

Antoine Bouyer

José Rouillard

Alain Derycke

Laboratoire Trigone

CUEEP – Bat B6

USTL de Lille

59655 Villeneuve d'Ascq Cedex

antoine.bouyer@gmail.com

jose.rouillard@univ-lille1.fr

alain.derycke@univ-lille1.fr

RESUME

Dans un contexte de multimodalité et d'interfaces abstraites, cet article fait émerger des verrous liés à la dynamicité des interfaces, à travers une étude de cas. Celle-ci se situe dans le domaine de l'enseignement à distance, qui est très orienté vers l'Internet classique. Nous montrons pourquoi, aujourd'hui, il n'est pas possible de déployer des interfaces concrètes, à la fois basées sur un modèle générique, et utilisables sur des plateformes d'enseignement à distance respectant la norme SCORM.

MOTS CLES : Plasticité, E-learning, objets pédagogiques, norme SCORM.

ABSTRACT

In a context of multimodality and abstract interfaces, this paper shows problems related to the dynamicity of interfaces, through a case study. This one is applied to the distance learning field, which is very directed towards the traditional Internet. We show why, today, it is not possible to deploy concrete interfaces, at the same time based on a generic model, and usable on learning management systems respecting SCORM.

CATEGORIES AND SUBJECT DESCRIPTORS: H.5 [Information Interfaces and presentation]

GENERAL TERMS: Design, Experimentation

KEYWORDS: Plasticity, E-learning, pedagogical objects, SCORM.

INTRODUCTION

Le laboratoire Trigone travaille depuis quelques années avec les interfaces adaptatives. Celles-ci permettent de

concevoir une application indépendamment du support sur lequel elle va être exécutée (ordinateur de bureau, PDA, téléphone, ...). Néanmoins, pour être utilisables, ces applications nécessitent d'être dynamiques, plutôt que constituées de « pages » statiques.

Travaillant aussi dans le domaine des EIAH (Environnements Informatiques pour l'Apprentissage Humain), notre équipe s'est intéressée à l'étude de la plasticité des objets pédagogiques (OPs). La transformation de ces éléments – qui sont la base de l'apprentissage à distance – afin de pouvoir les utiliser dans un contexte de mobilité et de multimodalité nous a permis de découvrir certains verrous associés à la dynamicité des interfaces abstraites. Cela rend cette étude de cas intéressante de par les problèmes généraux qu'elle soulève, plutôt que par les résultats directs obtenus. Ainsi, cet article reviendra d'abord sur la définition et l'utilisation des OPs, puis sur les possibilités existantes pour les rendre plastiques ainsi que les problèmes scientifiques et techniques rencontrés. Enfin, nous présenterons les verrous plus généralement liés au domaine.

QU'EST-CE QU'UN OBJET PÉDAGOGIQUE ?

Avec le développement d'Internet, l'intérêt pour l'enseignement à distance a considérablement augmenté. Dès lors, des plateformes d'e-formation (dites LMS pour *Learning Management System*) [4], [6] ont été créées et distribuées, librement ou non. Pour chaque plateforme, l'enseignant peut créer des modules (leçon de cours, QCM ou exercices à trous, par exemple), qui seront ensuite assemblés pour réaliser un cours complet. Comme il n'existait pas de standard structurant ces modules, chaque plateforme était indépendante et possédait son propre protocole de communication.

Lorsque l'on souhaitait changer de plateforme, tous les modules existants devaient donc être re-codés.

Pour remédier à ce problème, et permettre la mise en commun de matériel pédagogique, la norme SCORM (Sharable Content Object Reference Model) [14] a été proposée par ADL (<http://www.adlnet.org/>), à l'initiative du département de la défense Américaine (*Departement of Defense*). Son objectif est de permettre aux systèmes d'apprentissage en ligne d'importer, de partager, de réutiliser, et d'exporter des contenus d'apprentissage, de manière normalisée. Pour ce faire, SCORM regroupe un ensemble de spécifications déjà existantes ([1], [2], [8]), définissant ainsi l'agrégation d'objets pédagogiques (SCO), la navigation de l'un à l'autre, ainsi que la communication entre ceux-ci et les plateformes d'apprentissage à distance *Scorm-compliant*.

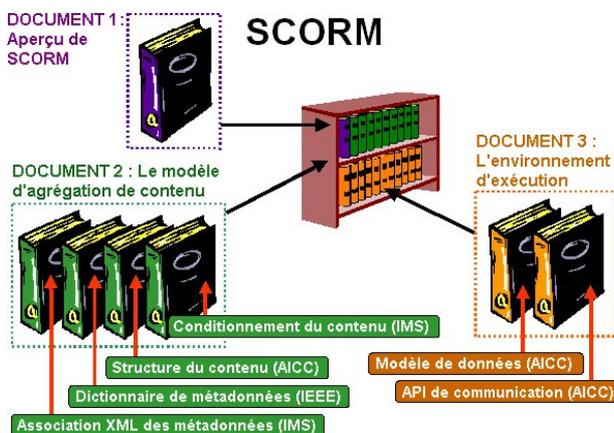


Figure 1 : Vue synthétique de la norme SCORM

Concrètement, SCORM peut être découpé en plusieurs parties différentes (cf. figure 1). Tout d'abord, le « modèle d'agrégation du contenu » assure la promotion de méthodes cohérentes en matière de stockage, d'identification, de conditionnement d'échange et de repérage du contenu. Il définit des méta-données décrivant le contenu (spécification LOM : <http://ieeeltsc.org/>), ainsi que l'empaquetage de l'ensemble des fichiers et documents joints. Ensuite, l'« environnement d'exécution » décrit les exigences nécessaires du système d'apprentissage quant à la gestion de l'environnement d'exécution. Il s'agit d'une API Javascript permettant une communication normalisée dans les deux sens, et indépendante des outils utilisés. Elle permet principalement à l'OP d'utiliser ou d'enregistrer des variables sur la plateforme, à l'aide de quelques fonctions Javascript. Enfin, le « modèle de séquençement et de navigation » permet une présentation dynamique du contenu. Il décrit comment le système interprète les règles de séquençement exprimées par un développeur de contenu, ainsi que les événements de navigation déclenchés par l'apprenant ou par le système.

Ces plateformes et la norme SCORM étant pensées pour produire un résultat graphique, cela pose des problèmes d'IHM. Par exemple, il peut apparaître un effet mosaïque

[15], [9], c'est-à-dire qu'il n'y a pas d'unité graphique entre la plateforme et les différents cours visualisés pourtant en même temps sur l'écran. De plus, dans le cas d'étudiants déficients (visuels, par exemple), aucun outil ou méthode ne permet de transformer ou de réadapter l'interface. Par la suite, ce sont les transformations des interfaces vers d'autres canaux qui vont être étudiées.

POURQUOI VOULOIR OBTENIR DES OBJETS PÉDAGOGIQUES PLASTIQUES ET MULTICANAUX ?

Avec l'essor de l'informatique ubiquitaire (ordinateurs portables, PDA, téléphones portables, ...), les utilisateurs doivent pouvoir utiliser leurs applications sur les différents modes de communications qui leur sont proposés à un instant donné (Internet, wap, téléphonie, ...). Cependant, pour développer une application de N pages destinée à M périphériques, il faut en principe coder N*M pages. Pour éviter ces redondances, le laboratoire Trigone a défini un langage de description d'IHM abstraites : PlasticML [12]. Le principe, comme pour d'autre langage ressemblant à USIXML [17], est de concevoir l'interface de façon unique, indépendamment de son contexte d'utilisation [16], [19]. Il n'y a alors plus que N pages à développer. Ensuite, une transformation XSL permet de traduire celles-ci vers le format voulu. Dans le cas de PlasticML, il s'agit de HTML [7], WML (wap) [20] et VoiceXML (téléphonie) [18], [13]. Etant donné le potentiel de l'informatique ubiquitaire pour le futur de l'e-Formation, en particulier le *mobile-learning* [5], nous nous intéressons aux OPs présentant une IHM « plastique ». De plus, cela permet d'envisager à terme des interfaces multimodales pour les OPs, et pas seulement sur d'autres périphériques graphiques [10] (PDA, Smartphone, ...).

Néanmoins, cette création a rapidement été limitée par la norme SCORM elle-même. D'une part, l'API de communication se fait grâce à des fonctions Javascript. Or les différents langages utilisés (HTML, WML et VoiceXML) possèdent chacun leur propre langage de script (respectivement Javascript, WMLScript et ECMA Script.) Nous ne pouvons donc pas créer *a priori* d'OPs SCORM accessibles vocalement ou via le WAP. D'autre part, étant donné que SCORM est orienté vers l'Internet « classique », les systèmes d'apprentissages compatibles SCORM ([4], [6]) ne proposent pas de portail Wap ou vocal. A l'inverse OpenUSS [11] permet une navigation sur le wap, mais n'est pas *Scorm-compliant*.

COMMENT RENDRE UN OBJET PÉDAGOGIQUE PLASTIQUE ? MÉTHODES ET PROBLÈMES ASSOCIÉS.

Plusieurs possibilités d'utilisation d'OPs plastiques ont été identifiées. Elles sont décrites ci-dessous, et les difficultés liées à chacune d'elles sont analysées.

La Rétro-engineering

Tout d'abord, la première transformation qui peut venir à l'esprit consisterait à utiliser un phénomène de rétro-engineering. Il faut dans ce cas faire transiter le flux HTML, envoyé par le système pour l'apprenant, à travers

un proxy. Celui-ci peut alors transformer en temps réel le code HTML vers un langage abstrait, qui sera ensuite retransformé vers le langage supporté par l'utilisateur [3].

Aussi, cette méthode a été écartée de nos objectifs dès le départ, notamment à cause des problèmes qu'elle engendre. La première difficulté apparaît à travers la richesse du code HTML et du CSS, par rapport aux autres langages. Ainsi, une page Internet peut devenir rapidement très complexe (cadres, tableaux imbriqués, largeur des éléments imposés, ...). Cela peut freiner, voire empêcher sa transposition vers des interfaces pour d'autres périphériques (petits écrans, interfaces vocales, etc.). Un autre verrou réside dans le fait que certaines pages Web ne sont absolument pas transformables. En effet, les animations flash ou les applets java, par exemple, ne pourront pas être abstraits, ni ré-exécutés telles quelles.

Utilisation D'objets Pédagogiques Plastiques Sur Des Plateformes « Classiques »

Rappelons que notre objectif est de créer des OPs plastiques, dans le but de les transformer par la suite vers un format cible voulu. Pour cela, nous utilisons le langage d'interfaces abstraites PlasticML, conçu par le laboratoire Trigone [12].

Dans un premier temps, le langage PlasticML a été étendu afin qu'il supporte l'utilisation d'événements et de scripts. Dans un souci de simplicité, il a été décidé de ne pas abstraire le langage de script. Les scripts sont ainsi stockés dans autant de fichiers externes que de langages cibles visés. Le nom des fichiers est le même pour tous les langages, seule l'extension change. De la même

```
Code PlasticML :
<plasticml version="2.0">
  <interface title="SCORM" load="open()"
    unload="close()">
    <script source="sco"/>
    ...
    <usevent name="Valid" action="check()"/>
    ...
  </interface>
</plasticml>
```

```
Code HTML après la transformation :
<html>
  <head>
    <title>SCORM</title>
    <script language="JavaScript" src="sco.js">
    </script>
  </head>
  <body onload="javascript:open();"
    onunload="javascript:close();">
    ...
    <input type="button" value="Valid"
      onClick="javascript:check();"/>
    ...
  </body>
</html>
```

Code 1 : Les scripts et les événements en PlasticML

façon, les noms des fonctions définies dans chaque fichier et utilisés par les événements sont les mêmes, et seule l'implémentation diffère (voir Code 1). Cela permet de ne pas s'occuper de l'implémentation des scripts dans le code de l'interface abstraite. Après réflexion, nous avons retenu cette solution, car les fonctions définies par les scripts, dans le cadre de SCORM, sont peu nombreuses, et une fois les fichiers pour chaque script créés, ils peuvent être réutilisés d'un objet à l'autre.

Plus globalement, les modifications apportées à PlasticML autorisent plus de dynamique, à travers l'utilisation de scripts prédéfinis pour un canal donné. Cette possibilité de gérer des scripts, lors de l'exécution des pages Web, permet de réaliser des tâches en fonction des événements rencontrés. Ceux-ci peuvent être déclenchés soit par l'ouverture ou la fermeture d'une page, soit par l'action d'un utilisateur (qui clique sur un bouton par exemple). Dans notre étude de cas, cet événement peut, par exemple, lancer une fonction qui initialise la communication entre l'OP situé chez l'apprenant et la plateforme sur le serveur, ou bien qui corrige un QCM proposé sur la page.

Un ensemble de pages PlasticML a été créé. Elles ont été regroupées dans un fichier zip, avec le fichier php permettant d'effectuer la transformation et un fichier xml décrivant la structure de ce *package*. Puisque ce paquet respectait la norme SCORM, il a pu être déployé sans problèmes. Ensuite, les pages possédaient des fonctions de scripts similaires à celles de l'API SCORM, si bien que la navigation des différentes pages jusqu'à un QCM de compréhension, s'est déroulée sans encombre sur un navigateur "classique".

Un souci de compatibilité survient au moment d'utiliser toutes les possibilités de PlasticML au travers d'un système d'apprentissage et de prendre connaissance du cours sur du wap ou au travers d'un téléphone normal. Deux solutions sont envisageables. La première consiste à accéder à cet objet directement et sans passer par un système d'apprentissage (comme la plateforme Claroline, développée en PHP, par exemple). La transformation interne à PlasticML fonctionne alors parfaitement, mais la puissance de la plateforme n'est pas utilisée. En particulier, il n'y a plus d'identification, ni de persistance des données (résultats aux QCM ou temps passé par un étudiant sur un cours). La seconde consiste à utiliser une plateforme qui apporte ces fonctionnalités ; mais celle-ci n'affiche pas seulement l'OP. En effet, l'écran est divisé en cadres (frame HTML), et l'OP n'est affiché que dans l'un d'eux, les autres permettant d'afficher la bannière du système ou le sommaire des chapitres du cours parcouru. Ainsi, même si cet objet peut être transformé vers du WML ou du VoiceXML, le cadre principal reste toujours en HTML.

Pour récapituler, le sujet reste ouvert, et le verrou technique présenté dans cet article demeure, puisqu'à notre connaissance, personne n'a montré qu'il était possible d'utiliser une plateforme standard sur des

canaux autres que ceux de l'Internet « classique ». La norme SCORM et donc les plateformes existantes sont tournées vers ce canal de manière relativement contraignante, notamment à cause de l'utilisation d'une API Javascript.

CONCLUSION

Tout au long de cet article, notre approche visant à permettre l'exécution d'un objet pédagogique sur un canal différent de l'Internet classique a été expliquée. Nous avons dû pour cela faire évoluer le langage PlasticML afin qu'il puisse supporter plus de dynamisme, grâce à la prise en compte de scripts.

La méthode présentée ici n'a pas aboutie. En effet, SCORM a été pensée pour être exécuté seulement via un navigateur Web. Ainsi une utilisation graphique, quel que soit le périphérique sous-jacent (PC, PDA, Smartphone, ...) fonctionne. En revanche, l'intégration de techniques, comme les frames HTML et le Javascript, l'empêche d'être utilisée sur d'autres modes de navigation, comme le vocal par exemple. Voilà pourquoi, selon nous, le format actuel de cette norme ne permet pas l'usage d'objets plastiques sur les LMS. De plus, la version SCORM 2004, qui commence à être implémentée par les éditeurs de plateformes, propose des améliorations d'ordre pédagogique. Néanmoins, aucune évolution des aspects interfaces n'est annoncée (dynamisme, multicanalité, plasticité...).

Plus globalement, cette étude de cas, nous a permis de faire émerger un verrou lié aux scripts dans la conception d'une application abstraite dynamique. En effet, chaque langage résultant d'une transformation possède son propre langage de script, qui possède lui-même ses propres caractéristiques techniques. Ainsi, il est difficile pour l'instant de définir à l'avance le comportement d'une application, disjointement du support d'exécution.

BIBLIOGRAPHIE

1. Aviation Industry CBT Committee. (<http://www.aicc.org/> *)
2. Ariadne (<http://www.ariadne-eu.org/> *)
3. Bouillon, L. and Vanderdonck, J. *Retargeting Web Pages to other Computing Platforms with VAQUITA*. Proceedings of the 9th Working Conference on Reverse Engineering WCRE'02 (October 29 - November 01, 2002, Richmond), IEEE, 2002, pp. 339 – 348.
4. Claroline. (<http://www.claroline.net/> *)
5. Derycke, A., Chevrin, V. et Rouillard, J. *Intermédiations Multicanales et Multimodales pour l'E-Formation : l'Architecture du Projet Ubi-Learn*. Conférence EIAH'2005, ATIEF (25-27 Mai 2005, Montpellier), INRP (eds), 2005, 6p.
6. Ganesha. (<http://www.anemalab.org/ganesha/> *)
7. HTML (HyperText Markup Language) : langage support d'Internet. (<http://www.w3.org/MarkUp/> *)
8. IMS Global Learning Consortium, Inc. (<http://www.imsglobal.org/> *)
9. Ip, A., Radford, A. and Canale, R. *Overcoming the presentation Mosaic Effect of Multi-Use Sharable Content Objects*. Proceedings of the 20th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education ASCILITE'03 (7-10th Décembre 2003, Adelaide, Australie), ASCILITE, 2003, pp. 256 – 262.
10. Lin, N. H., Shih, T. K., Hsu, H., Chang, H., Chang, H., Ko, W. C., and Lin, L. J. 2004. *Pocket SCORM*. Proceedings of the 24th international Conference on Distributed Computing Systems Workshops - W7: EC (Icdcs'04) - Volume 7 (March 23 - 24, 2004). ICDCSW. IEEE Computer Society, Washington, DC, 274-279.
11. OpenUSS. (<http://openuss.sourceforge.net/> *)
12. Rouillard, J. *Plastic ML and its toolkit*. Proceeding of the 10th International Conference on Human - Computer Interaction HCI'03 (22-27th June 2003, Heraklion, Crete, Greece), 2003.
13. Rouillard, J. *VoiceXML - Le langage d'accès à Internet par téléphone*, Editions Vuibert, 2004, ISBN : 2-7117-4826-x, 197 pages.
14. Sharable Content Object Reference Model. (<http://www.adlnet.org/scorm/> *)
15. St-Pierre, R. and Hope, P. *Dynamic Appearance Model Analysis and Alternatives*. Canadian Department of National Defence, Publié par Sun Microsystems Inc, 2003.
16. Thévenin, D. *Adaptation en Interaction Homme-Machine : Le cas de la Plasticité*. Thèse en informatique, Université Joseph Fourier, Grenoble, 21 Décembre 2001. (<http://iihm.imag.fr/thevenin/these/TheseDavidThevenin.pdf> *)
17. USIXML (User Interface eXtensible Markup Language). (<http://www.usixml.org/> *)
18. VoiceXML : langage permettant d'accéder à Internet via un téléphone. (<http://www.voicexml.org/> *)
19. Vanderdonck, J., Limbourg, Q., Michotte, B., Bouillon, L., Trevisan, D., Florins, M. *UsiXML: a User Interface Description Language for Specifying Multimodal User Interfaces*. Proc. of W3C Workshop on Multimodal Interaction WMI'2004 (Sophia Antipolis, 19-20 July 2004)
20. WML : langage support du wap. (<http://www.wapforum.org/> *)

* Tous les liens Internet ont été visités le 20 Février 2006

Contextual QR Codes

José Rouillard

Laboratoire LIFL - Université de Lille 1

59655 Villeneuve d'Ascq Cedex - France

jose.rouillard@univ-lille1.fr

Abstract— QR Codes (2D barcodes) are used to encode and decode data at a rapid rate. Using camera phones to read two dimensional barcodes for various purposes is currently a popular topic in both research and in practical applications. But until now, the information provided by QR Codes was solely static. What you see is exactly what was encoded. This paper proposes the notion of contextual QR Codes that merge a public QR Code and private information, in order to provide data related to a particular context. For example, a public tag <Hello> will be decoded and translated into “Good evening John Smith” or “Bonjour Jean Dupond”, if the system is able to obtain data such as the name of the user, the language used on the machine (English or French for instance) and the moment when the interaction takes place.

Keywords— QR Codes, Two-Dimensional barcode, Context-aware, Pervasive system, Ubiquitous computing, Ambient intelligence.

I. INTRODUCTION

Computers are widespread and many everyday-objects come equipped with computer technology. Mobile phones are equipped with high-resolution color displays, wireless access to the Internet, and respectable processing power and memory. Nowadays, different kinds of codes are used in order to store, retrieve and manage information. Mark Weiser's vision of Ubiquitous Computing underlines the need of seamlessly unifying computers and humans around the notion of rich environment. He explained: “The most profound technologies are those that disappear.” [26], [27]. Indeed, with the purpose of being minimally intrusive, pervasive systems have to deal with interaction issues in context-aware intelligent environments. This can be achieved through seamless interaction of environment and user, application and service adaptation according to user preferences and expectations, and efficient utilization of available resources [16].

My work is part of an exploratory project on adaptative services and usages for human learning in the context of pervasive communications, called P-LearNet (Pervasive Learning Networks). The aim of this paper is to present contextual QR Codes: instead of always exposing the same information to everybody, a contextualized QR Code is computed in order to provide specific information depending on a particular context.

This document is structured as follows: Section two presents related work and explains the background and motivation of this project. Section three gives an overview of QR Codes covering technology, usage, and development approaches. Section four describes the notion of context used in section five in order to present contextual QR Codes. Applications scenarios are given in section six before a conclusion with a roadmap for future work.

II. RELATED WORK

In the domain of object identification, RFID¹ and NFC² technologies are considered as the latest generation, but, barcodes are still regarded as an interesting option, because of the basic technology and simplicity of the concept. Indeed, any inexpensive printer may be used, but creating an RFID requires a special dedicated device [7]. Previous studies show the relevance of using barcodes with phones equipped with a camera. Prototypical applications hyperlinking the real world were studied: allergy-assistant, country of origin of a given product, price comparison, instant eBay-auctions, etc. The Allergy-Check application [1], for instance, is based on the recognition of one-dimensional (1D) barcodes. *Once the user defines a profile that contains all substances he or she is allergic to, holding the mobile phone in front of a product's barcode gives the user a simple answer to the question Is that product compatible with my allergies?* [2].

With this approach, one can access data related to billions of products carrying EAN-13 barcodes (ISO/IEC 15420:2000) [12], if available in online specialized databases (Amazon, Wiki food, SINFOS...). Classical 1D barcodes are widely popular and universally recognized because of their reading speed, accuracy, and functional characteristics. However, the need to store more information in codes printed on small spaces leads to the emergence of two-dimensional barcodes (2D Codes). They can be used to access data and services like bus time tables, product information, etc. Figure 1 shows the same information encoded in a QR Code (left) and in a EAN-13 (European Article Numbering) barcode (right).

¹ Radio-Frequency Identification (RFID) is an automatic identification method, relying on storing and remotely retrieving data using devices called RFID tags or transponders.

² Near Field Communication (NFC) is a new standard for mobile phones that allows them to both act as an RFID reader and be read by other RFID readers.



Figure 1: QR Code (2D) and EAN-13 (1D) barcodes encoding the same data (3254565174110).

There is a wide variety of 2D code representations available and specifically designed to simplify camera-based recognition. Some of them are open standards while others are proprietary such as Semacodes, Spotcodes, Rohs' VisualCodes, ColorCode, CyberCode, MobileTag, VeriCode, ShotCode, eZcode, HotScan, Codablock F, Aztec, FP Code (Fine Picture Code - Fujitsu) and BeeTagg (ConnVision). PDF417 (Portable Data File) and MaxiCode are used under AIM International ISO Standardization. The two most well known 2D barcode standards are DataMatrix (ISO/IEC 16022:2000)[13] and QR Code (ISO/IEC 18004:2000)[14]. There is no license fee to be paid to use neither DataMatrix nor QR Codes.

Even though a study comparing them [4] quoted by [7] explained the superiority of DataMatrix (excepted for Japanese Kanji symbols encoding), QR Codes are most common in Asia and particularly popular in Japan.

III. QR CODES

A QR Code is a two-dimensional barcode introduced by the Japanese company Denso-Wave in 1994. This kind of barcode was initially used for tracking inventory in vehicle parts manufacturing and is now used in a variety of industries. QR stands for "Quick Response" as the creator intended the code to allow its contents to be decoded at high speed.

A. Technology

A QR Code is a matrix code developed and released primarily to be a symbol that is easily interpreted by scanner equipment. It contains information in both vertical and horizontal directions, whereas a classical barcode has only one direction of data (usually the vertical one). Compared to a 1D barcode, a QR Code can hold a considerably greater volume of information: 7,089 characters for numeric only, 4,296 characters for alphanumeric data, 2,953 bytes of binary (8 bits) and 1,817 characters of Japanese Kanji/Kana symbols. QR Code also has error correction capability. Data can be restored even when substantial parts of the code are distorted or damaged.

In the QR Code standard, corners are marked and estimated so that the inside-code can be scanned [18]. The barcode recognition process has 5 steps: (1) edge detection, (2) shape detection, (3) identification of barcode control bar, (4) identification of the barcode orientation, dimensions and bit density using the control bar, and (5) calculating the value of the barcode [21].

B. Usage

Without a machine, it's impossible for a human to manually decode QR Codes but they are easily processed by scanning equipment. In Japan, many cellular phones are now natively equipped with a QR Code-reading software. Users photograph QR Codes and the software integrated into their phones decodes the messages and displays, manipulates, or stores the information on their mobile devices. Depending on the type of data recognized and the nature of the application, alternative actions can follow the decoding stage: a phone number can be automatically dialed, a short text message can be sent, a web page corresponding to the decoded URL (Uniform Resource Locator) can be displayed in a mobile browser, or a definite application can be executed.

QR Codes are part of daily life in Japan, Korea, Taiwan, Hong Kong and China. A study published in January 2005 by MRI showed that out of 2053 Japanese mobile phone users, 90% have recognized a QR Code. McDonalds uses codes to inform users about the nutritious value of its burgers. Apple advertised the new i-Pod on billboards with QR codes. QR Codes used in a Nike advertising campaign allows direct access to a dedicated mobile site³. In Japan, some teachers are using QR codes to distribute resources to learners [9]. QR codes now appear in magazines, advertisements, product wrappings, T-shirts, passports, business cards and on subway billboards in Japan. But, at a consumer market level, QR Codes are virtually unknown outside of Asia. It is believed that usage of QR Codes in the rest of the world will not gain momentum until mobile telephone operators begin pre-installing QR Code-reading software on mobile phones. Today, in Europe few phones come with the software installed. Nevertheless, several companies, mainly in France, the UK, and Switzerland are starting to use QR codes to promote goods or services such as in Swiss online newspaper, or to inform customers about daily rates, for instance (European Central Bank).

C. Development (DIY)

Users can scan existing QR Codes or they can generate their own. Creating QR Codes online is very easy and many web sites (Kaywa, Snapmaze, Activeprint...) can be used to encode and print out such codes.

For camera phones and PDAs (Personal Digital Assistant) that are not equipped with QR Code readers, there are some add-on tools that decode QR Codes simply by positioning the device in front of the code. This is done automatically within the streaming flow and the user doesn't have to take a picture of the QR Code. QuickMark [19] and I-nigma [11] readers are good examples of free tools using this technique that are available for many manufactured models and devices. QuickMark provides extension functionalities to QR Codes, by allowing partial or entire encryption of codes.

³ <http://mobile.nikefootball.com/qrcode.jsp>

Another interesting feature is the “Magic Jigsaw”: this option encodes binary data (a picture for example) as a chain of QR Codes that the user can scan to retrieve the original content.

Developers that want to use QR Codes in their applications have to first decide whether the computation of the code will be made locally or not. Obviously, a network connection is required when the treatment is done remotely, through a web service for example, such as those provided by Richard Jones [15]. Another option is to send an E-mail to r@qry.jp [20] with the picture of the QR Code as an attachment. The system will instantly send back the converted data.

Alternatively, if no network connection is needed or available, the code management will have to be done by the mobile device, in an autonomous way. If the final user only needs to scan codes and see the result messages, the softwares mentioned above are sufficient. But for developers who have to manage QR Codes, some SDKs (Software Development Kit) are announced and some are already available.

Microsoft Windows Live Barcode project, OpenNETCF QRCode Library for .NET Compact Framework and Google ZXing (Zebra Crossing) project will be available soon. Twit88 [24] provides an open source project on QR codes. We tried this solution and also some commercial other tools, like Tasman [23] (Tasman.Bars Java and .Net SDK).

Figure 2 presents the mobile application that we have developed in our laboratory. It was written in C# language and runs, with Tasman library, on a smartphone HTC TyTN II (Kaiser) supporting Windows Mobile 6. As the user clicks on the bottom left button the camera manager is invoked. Then, according to the selected representation (QR Code, for instance on Figure 2), the decoded information is presented to the user. Obviously, this is the classical and minimal management that our prototype is able to do. We will see in the following, how this application can become context aware by using various information.

IV. CONTEXT AWARENESS

Current literature shows a growing interest in the creation of context-aware systems, but there is no consensus on the definition of context. However some definitions are quoted several times and become de facto standard [3]. For example, Anind Dey's definition, is one such standard: *A context-aware system is a computing system using context to provide relevant information and/or services to the user, where relevancy depends on the user's tasks.* [5].

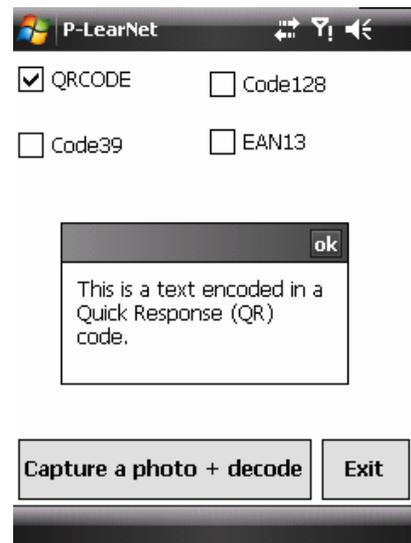


Figure 2: Our mobile application decoding QR Codes

Indeed, one of the main goals of ubiquitous computing is to provide relevant information, at the right time and place and under the right form. Context-aware systems should help filter information. This allows users to concentrate on the task rather than the technology.

Our current works is inspired by the AMULETS project, in which researchers used metadata and particularly location-based data (GPS coordinates) embedded in semacode URL, in order to determine the location of the device scanning the barcode that the students used [22], [17]. With this method, some parameters are passed in the URL that is generated dynamically.

V. USING QR CODES IN CONTEXT

Recently, some researchers introduced the notion of “contextual bookmark” which is a “*combination of a snapshot of a physical object taken with a mobile device and meta information about the content related to this physical object*” [10]. Our definition of a contextual QR Code is close to this one and is the following: it’s the result of a fusion between a public part of information (QR Code) and a private part of information (the context) provided by the device that scanned the code.

Figure 3 shows the public and private parts of a contextual QR Code. The private part can be one or more information among the subsequent: user’s profile, current task, device used, location, time and environment of the interaction. The machine decodes the QR Code and merges it with private data obtained during the interaction. Then, the XML (Extensible Markup Language) resulting file is sent to a web service (created in our laboratory) that computes the code and returns personalized messages. This web service was developed in C# .Net and is able to retrieve, according to particular tags, the right module to invoke (for example “Hello World” or “Meeting” Applications).

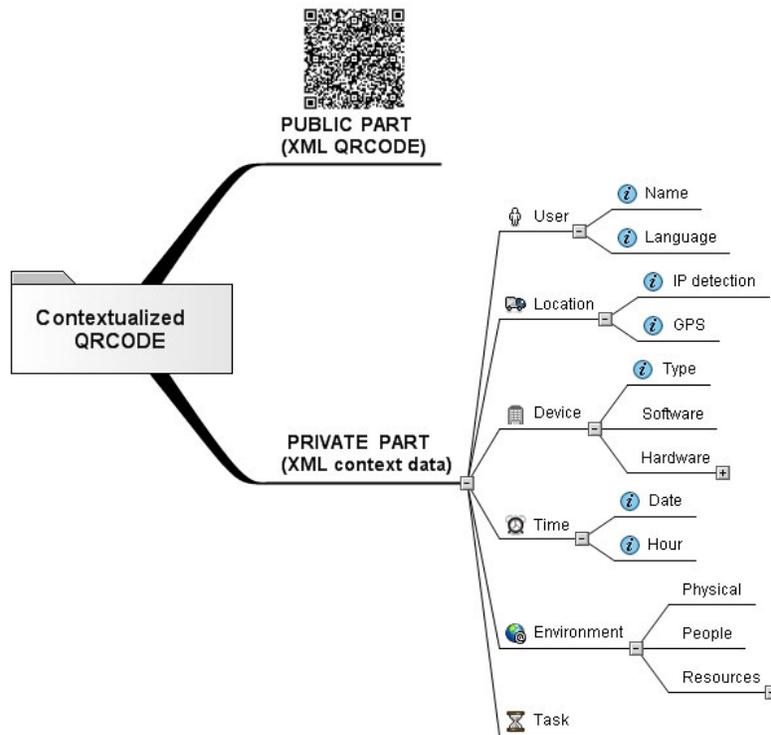


Figure 3: Pubic and private parts of a contextual QR Code

VI. APPLICATION SCENARIOS

At present, we use a HTC smartphone for our tests. To illustrate our approach, we developed a prototype able to run basic scenarios. In those scenarios, a QR Code represents data formatted in XML language and corresponds to the public part of the message. The second part of the message is generated when the user takes a picture of the QR Code, taking into account private information.

The functional decomposition of the program is the following: the “capture photo” module and “decode data” module are used in order to obtain a public XML message. Then, the “retrieve private data” and “create private XML message” modules are called. In a third time, the “merge private and public XML” module is used just before the “call Web services” and “display results” modules are invoked.

A. Hello World example

The first scenario is the venerable “Hello World” greeting. Imagine that the following message is encoded in a QR Code: `<public> <tag> Hello </tag> </public>`. If a user tries to scan this QR Code with a traditional reader (say Quickmark [19] or I-nigma [11] for instance), she will discover exactly this message on her mobile screen. Now, if she uses our contextual QR Code application, she will see a personal greeting message on her screen.

For example, if she scans the code at 8 o’clock A.M., connected with the IP address 209.9.235.121, the message will

be, for example: “Good morning Jenifer Diaz from Pompano beach, Florida, USA.”.

This is possible with IP detection [25] and thanks to an “ip to location” web service [8]. The name of the user is automatically detected in the mobile registry (ControlPanel\Owner).

It will take around 5 seconds to complete this task. Less than 2 seconds are dedicated to the decoding operation (this is almost the same time as this needed by QuickMark and I-Nigma) and 3 to 5 seconds are needed to reach the different web services used by our application. Of course that time will be dependent on the amount of encoded information and the network’s speed.

B. Meeting example

The second proposed scenario is the meeting example. We often use a classical sheet of paper to write down our name and affiliation during a meeting. Sometimes, this is also used as a proof of presence of a person to a meeting or a brainstorming. Of course, a paper can’t be used simultaneously by different people that have to sign it. But, a paper containing a QR Code can be scanned by many users, and even in different orientation. The exact time is not mentioned on the paper, because, it’s too long and boring to do.

We think that the idea to use a QR Code that can be scanned during a meeting could be interesting if the users are correctly equipped with devices supporting the kind of task. Figure 4

represents an example of an XML file. The first part of the document concerns public information in relation to a meeting, and the second part, captured at the runtime, is about personal information of the user than scanned the QR Code.

VII. CONCLUSION

We have presented in this paper the technology, usage and way of programming two-dimensional barcodes, called QR Codes. They offer an interesting way to capture and distribute a tremendous amount of information in a simple, quick and efficient manner. There are plenty of potential uses for this technology, but we think that future usages will surely be based on context-aware features.

We have proposed the notion of contextual QR Codes, which merges a public QR Code and some private information, in order to provide data related to a particular context. In the first example, the user scans a QR Code that contains a generic “Hello World” XML message, and obtains a personalized and contextual message. In the second example, users are invited to scan a QR Code that contains public information about a meeting (title, date, hour, room, etc.). The fusion of those data with the ones given during the interaction (such as the name of the person, the location, the language used) is then sent to a web service able to manage the provided information. Finally, the user receives on his mobile appliance the result of the interpretation of the contextualized QR Code.

We think that contextual QR Codes can be easily integrated into pervasive and ubiquitous applications and we believe that the power of Quick Response Codes coupled with context-aware information will provide an important impact on Human Computer Interaction.

Further works will lead us to work with other parameters of the context not yet used in our prototype. As we mentioned in Figure 3, the private part of the message could be related to user, location and time, but also to device, environment and task. We used the first three parameters in our prototype, but not the three last. In future version of our prototype, we will certainly add the possibility to personalize the responses according to the kind of device detected, the environment (light, noise, etc.) and user’s task, in order to become more and more context-aware. Of course, some evaluations will be done to know how to improve the concept and the software in order to satisfy user’s needs.

ACKNOWLEDGMENT

We are grateful to ANR P-LearNet project for providing support for this research, to Tasman, QuickMark, and Eric Metois (Eym barcode SDK) for special tools provided and to Eugène Van Den Bulke and Rachel Mathews for their help in improving this paper.

REFERENCES

- [1] Adelman, R., Langheinrich, M., Floerkemeier, C., A Toolkit for Bar-Code-Recognition and -Resolving on Camera Phones – Jump Starting the Internet of Things. Proceedings of the workshop on Mobile and Embedded Interactive Systems (MEIS’06) at Informatik 2006. GI

```

<public>
  <meeting>
    <id>1234</id>
    <title>P-LearNet Brainstorming</title>
    <date>10/02/08</date>
    <room>326</room>
  </meeting>
</public>
<private>
  <device_owner>Alex Harasymczuk
</device_owner>
  <date> <day>10</day>
  <month>10</month>
  <year>2008</year>
</date>
  <time>
    <hour>17</hour>
    <minute>39</minute>
    <second>27</second>
  </time>
  <language>english</language>
</private>

```

Figure 4: Public and private parts of an XML file sent to a web service for treatment

The QR Code can be printed on a sheet of paper and/or can be generated on a digital screen, in order to be scanned at a relatively long distance. Figure 5 shows the result obtained by this user (Alex) after he scanned the code. Additional features have been added to our prototype: a picture of each person identified is displayed on the screen of the meeting. An automatic report is generated with the name of each participants and the time of their registration, etc.

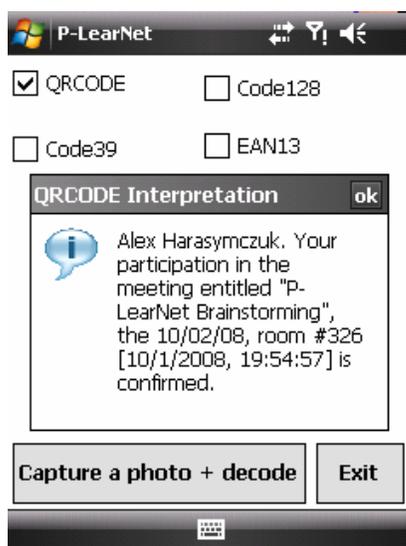


Figure 5 : Contextual QR Code interpretation

- Lecture Notes in Informatics Series (LNI). Dresden, Germany, 2006.
- [2] Adelman, R., Langheinrich, M., Rapid Prototyping Platform for Mobile Phone Based Services on Retail Products., Demo at the International Symposium on Ubiquitous Computing Systems (UCS 2007), Akihabara, Tokyo, Japan, 2007.
- [3] Clerckx, T., Model-based development of context-aware interactive applications in ambient intelligence environments, Ph.D thesis, Hasselt University, 2007.
- [4] Consumer Electronics Association, Comparison of Data Matrix and QR Code. Technical report, 2001.
- [5] Dey. A., Providing architectural support for building context-aware applications. Ph.D thesis, College of computing, Georgia institute of technology, 2000.
- [6] Denso Wave QR Code and QR Code features, online document. Retrieved March 2008 from <http://www.denso-wave.com/qrcode/qrcodefeature-e.html> and <http://www.denso-wave.com/qrcode/aboutqr-e.html>
- [7] Falas, T., Kashani, H., Two-Dimensional Bar-Code Decoding with Camera-Equipped Mobile Phones, Fifth Annual IEEE International Conference Pervasive Computing and Communications Workshops (PerCom), pp. 597 – 600, 2007.
- [8] Fraudlabs tools. Retrieved March 2008 from: <http://www.fraudlabs.com/ip2location.aspx>
- [9] Fujimura, N., Doi, M., Collecting students' degree of comprehension with mobile phones, Proceedings of the 34th annual ACM SIGUCCS conference on User services table of contents, Edmonton, Alberta, Canada pp. 123 – 127, ISBN:1-59593-438-3, 2006.
- [10] Henze, N., Lim, M., Lorenz, A., Mueller, M., Righetti, X., Rukzio, E., Zimmermann, A., Magnenat-Thalmann, N., Boll, S., Thalmann, D., Contextual bookmarks, Joint Workshop Mobile Interaction with the Real World, MIRW 2007 and the 5th Workshop on "HCI in Mobile Guides", MGuides 2007, Singapore, pp. 51-54, 2007.
- [11] I-Nigma : Retrieved March 2008 from: <http://www.i-nigma.com/>
- [12] ISO/IEC 15420:2000 - Information Technology – Automatic Identification and Data Capture Techniques - Bar Code Symbology - EAN/UPC, 2000.
- [13] ISO/IEC 16022:2000 - Information Technology – International Symbology Specification - Data Matrix, 2000.
- [14] ISO/IEC 18004:2000 - Information Technology – Automatic Identification and Data Capture Techniques - Bar Code Symbology - QR Code, 2000.
- [15] Jones, R., QR Code web service, Retrieved March 2008 from: <http://www.binaryrefinery.com/qrweb/webservice.aspx> (User and password need to be both set to br0wncow! for now).
- [16] Khan, M.T., Zia, K., Daudpota, N. Hussain, S.A., Taimoor, N., Integrating Context-aware Pervasive Environments, IEEE International Conference on Emerging Technologies, ICET 2006, Peshawar, Pakistan, 2006.
- [17] Kurti, A., Milrad, M., Alserin, F., & Gustafsson, J., Designing and Implementing Ubiquitous Learning Activities Supported by Mobile and Positioning Technologies. Proceedings of the IASTED CATE 2006 Conference held in Lima, Peru, pp 193-199, 2006.
- [18] Ohbuchi, E., Hanaizumi, H., Hock, L.A., Barcode Readers using the Camera Device in Mobile Phones, in Proc. of 2004 International Conference on Cyberworlds, pp. 260-265, 2004
- [19] Quickmark. Retrieved March 2008 from: <http://www.quickmark.com.tw/>
- [20] QRY. Retrieved March 2008 from: <http://qry.jp/>
- [21] Reilly, D., Smolyn, G. and Chen, H., Toward fluid, mobile, and ubiquitous interaction with paper using recursive 2D barcodes. Pervasive Mobile Interaction Devices 2007 (PerMID 2007), workshop at Pervasive 2007, Toronto, Canada, 2007.
- [22] Svensson, M., Pettersson, O., Case study: Extending content metadata by appending user context, Växjö University, Faculty of Mathematics/Science/Technology, School of Mathematics and Systems Engineering, 2006.
- [23] Tasman SDK.. Retrieved March 2008 from: <http://www.tasman.co.uk/>
- [24] Twit88. Retrieved March 2008 from: <http://www.twit88.com/home/opensource/qrcode>
- [25] What is my IP : Retrieved March 2008 from: <http://wimip.fr/>
- [26] Weiser, M., The Computer for the 21st Century, Scientific American 265, No. 3, 94-104, 1991.
- [27] Weiser, M., Some Computer Science Problems in Ubiquitous Computing, Communications of the ACM 36, No. 7, 74-83, 1993.

Chapitre 8

Multimodalité et sémantique

Résumé

Ce chapitre traite de la sémantique et de la multimodalité dans le domaine des interactions Homme-Machine. Nous exposons un bref rappel des notions de média, multimédia, mode, modalité et multimodalité, avant de montrer en quoi la sémantique est une notion essentielle qui différencie un système informatique multimédia d'un système informatique multimodal. Du fait que plusieurs approches sont possibles en conception d'interfaces Homme-Machine, selon que l'on se place côté machine, côté utilisateur, on encore de manière hybride (machine/ utilisateur), la multimodalité s'inspire de modèles pluridisciplinaires et bénéficie des avancées en informatique, sociologie, psychologie, économie et théorie des jeux, voire des théories du marketing (le multicanal étant effectivement étudié lors de communication personne/organisation).

Nous illustrons notre propos grâce à la présentation d'un système de dialogue Homme-Machine orienté apprentissage, dans lequel un internaute interagit avec un agent animé. L'apport et l'enjeu de la multimodalité dans ce contexte est expliqué et commenté : l'agent est perçu visuellement, puisqu'il se déplace sur l'écran, mais aussi auditivement, car une voix de synthèse accompagne ses actions, et apporte des indications complémentaires ou redondantes à l'utilisateur. Cette multimodalité véhicule également des éléments supplémentaires pour l'acquisition de sens : geste de désignation, comportements anthropomorphiques (humeur, émotions, aides personnalisées), animations permettant de comprendre ce qu'il se passe au niveau du noyau fonctionnel du système informatique, etc.

2 Variation, construction et instrumentation du sens

8.1 Les interactions Hommes-Machine

La majeure partie des tâches que nous confions à nos ordinateurs, n'est pas d'ordre purement calculatoire, comme on le pense souvent, mais plutôt d'ordre computationnel. En effet, le calcul, au sens mathématique du terme, n'est qu'une partie des traitements qu'effectuent nos machines. Quotidiennement, nous les sollicitons pour des tâches répétitives (traitement de texte, envoi de courrier électronique, recherche d'information, etc.), pour lesquelles nous manipulons des symboles. La plupart du temps, l'ordinateur n'est pas le destinataire de cette information, mais simplement un médiateur. Nous nous trouvons donc dans une situation de dialogue entre humains, instrumenté grâce à une chaîne d'outils.

Parfois, la machine peut devenir un interlocuteur, capable de "comprendre" certaines commandes ou certains éléments d'un dialogue. Dans ce cas, l'ordinateur n'est pas simplement utilisé pour transmettre des informations d'un humain vers un autre. Il doit les traiter en y donnant du sens. D'ailleurs, pour évaluer les capacités d'une machine, une possibilité – parmi d'autres – consiste à tester non pas seulement la puissance calculatoire de celle-ci, mais bien plus encore, ses facultés à entretenir une conversation avec un interlocuteur humain (voir à ce propos le fameux « test de Turing »). Le problème le plus délicat, que d'autres avant nous ont déjà mis en lumière, c'est sans doute que l'homme veut tenter de transmettre quelque chose (une intelligence ?) à la machine, alors même qu'il ne maîtrise pas complètement la nature et les phénomènes qui régissent son fonctionnement naturel. Dans son « Histoire universelle des chiffres », Georges Ifrah livre sa réflexion sur ce sujet, avec un paragraphe qu'il intitule « Lorsque la métaphore fut identifiée à la réalité », et dont voici un large extrait :

« Dans leur enthousiasme, ces savants crurent d'abord que toute opération de la pensée humaine était exclusivement de nature calculatoire (c'est-à-dire algorithmique), et donc que tout processus intellectuel était exécutable sur une machine de type ordinateur.

Et comme ils désignèrent sous le nom de neurones et de synapses des modules qui n'eurent en réalité qu'une très lointaine parenté avec les composants réels, beaucoup plus complexes, d'un encéphale vivant, ils établirent inévitablement, selon un anthropomorphisme simpliste, et forcément réducteur, un parallèle, très étroit entre les circuits d'un calculateur électronique avec les cellules nerveuses et les composants neurobiologiques du cerveau vivant.

Et c'est ainsi que l'organe central d'un simple ordinateur électronique fut désormais considéré comme l'équivalent du cerveau humain, le public assimilant dès lors les activités d'un calculateur à celles, éminemment supérieures, de l'esprit humain.

D'où, par identification abusive de la métaphore avec la réalité, et par une sorte de convergence avec le mythe du « cerveau électronique », le développement de l'idée d'une machine prétendument douée de pensée inductrice, et créatrice, munie de la faculté de prendre toutes sortes d'initiatives, et à laquelle l'homme pourrait confier tous ses problèmes sans exception pour obtenir d'elle toutes les solutions voulues quasi instantanément.» [IFR 94].

On a donc tendance à faire, de manière trop hâtive, un parallèle entre les possibilités d'une machine et celles de l'homme, avec la fausse idée qu'une grande puissance calculatoire puisse être suffisante pour compenser de faibles capacités à donner du sens aux informations traitées.

8.2 Qu'est ce que computer ?

Depuis quelques décennies, nous sommes passés de systèmes technocentrés à des systèmes plus anthropocentrés, pour lesquels l'utilisateur est au cœur du processus d'interaction. Or, par nature, l'homme a la faculté de computer¹, parfois sans même savoir quels sont les mécanismes en jeu dans ses raisonnements. L'esprit humain est vu, par certains chercheurs, comme un système de manipulation de symboles (H. Simon), ou comme un système de traitement de l'information (A. Newell). Jean-Louis Le Moigne explique « *Computer, ce n'est pas seulement calculer arithmétiquement des nombres, c'est, très généralement, manipuler et traiter des symboles.* » [LEM 86].

Le contexte est très souvent un facteur permettant de désambiguïser un énoncé. Par exemple, une personne qui voit un panneau mentionnant « St. John St. » dans une ville anglaise n'aura guère de mal à traduire cela en « Saint John Street ». Henriette Walter montre que l'aspect géographique et territorial influence grandement le sens des mots et des expressions d'une langue [WAL 98]. Elle prend l'exemple de conversation suivant :

- remettez-vous donc !
- non merci, je reste droit.
- mais si, j'insiste, vous allez resquiller et vous ruiner la jambe.
- ça vaut mieux que de partir de cinq en cinq.

Un dialogue apparemment incohérent, et pourtant parfaitement logique (lorsque l'on est de Sète) :

¹ À prononcer à la française : c'est ici le verbe français, et non le substantif anglais. Du latin *computus* «compte». Littéralement «computer» signifie calculer. Plus exactement, calculer la date des fêtes mobiles. Ainsi le «comput ecclésiastique» dresse le calendrier de la date de Pâques.

4 Variation, construction et instrumentation du sens

- asseyez-vous donc !
- non merci, je reste debout.
- mais si, j’insiste, vous allez glisser et vous casser la jambe.
- ça vaut mieux que de décliner lentement.

Il faut donc computer les informations pour y trouver du sens. Au niveau langagier, l’émetteur d’un message fait parfois l’hypothèse que ses interlocuteurs vont décoder le message de manière approprié. C’est ce que Violaine Prince appelle les principes d’économie et d’expansion [PRI 96]. Par exemple, lorsque l’on dit « Ne jetez rien dans cette poubelle » et que l’on souhaite faire comprendre à l’interlocuteur « Ne jetez rien d’important, ou de confidentiel, dans cette poubelle ». D’autres chercheurs appellent cela l’effet paillason [CHA 02] : en effet, lorsque l’on vous demande de vous essuyer les pieds sur un paillason, il est évident que l’on vous suggère d’essuyer les semelles de vos chaussures ; on ne souhaite pas vous voir retirer vos chaussures et vos chaussettes, pour frotter la plante de vos pieds sur le paillason. Ces traitements d’informations sont quasiment instinctifs pour les humains. Il n’en va pas de même pour les ordinateurs.

Cependant, la communication humaine ne s’appuie généralement pas sur un seul canal de communication, mais bien sur plusieurs, et le plus souvent, de manière simultanée. Dans un dialogue, le geste accompagne fréquemment le discours. [MCN 92] a suggéré que le geste et la parole étaient générés par la même région du cerveau, et d’autres pensent que les premiers langages humains étaient basés sur des gesticulations (Cf. Zimmer 95, cité par [THO 96]). Beaucoup de classifications ont été utilisées pour décrire les types de gestes que l’on utilise lors de dialogues ou de discours [RIM 91], [POY 80]. La plupart sont dérivées des travaux de [EFR 41]. Globalement, on retrouve entre quatre et six classes de gestes différentes. Les plus étudiées et les plus utilisées pour la communication homme-machine sont les suivantes : gestes non-descriptifs, descriptifs (idéographiques), iconographiques, pantomimiques, déictiques, emblématiques (symboliques) et métaphoriques. Certains chercheurs parlent également d’une autre catégorie (en anglais « self adaptors », [MCN 92], [EKM 69]), pour référencer les actions « personnelles » comme se passer la main dans les cheveux, se gratter, etc. Les gestes, mimiques et attitudes du visage ont également beaucoup été étudiés en communication homme-homme et homme-machine [EKM 75]. Et, là encore, des taxonomies de ces gestes ont été proposées : on note par exemple les gestes symboliques, émotionnels, affectifs, conversationnels, ponctuateurs, régulateurs, manipulateurs, etc. [EKM 78].

Au regard de ces travaux de recherches et des différentes observations effectuées en laboratoire, il nous semble que si nous voulons doter nos machines de facultés à

interpréter nos actes et notre langage, nous devons pousser plus avant les travaux dans le domaine de la sémantique couplée à la multimodalité.

8.3 La multimodalité en IHM

Nous avons vu précédemment que les ordinateurs ne servent pas uniquement à calculer des opérations, mais qu'ils permettent également de traiter, stocker et indexer des informations de différentes natures. Le problème principal que nous rencontrons lors des processus de recherche d'informations n'est plus la disponibilité de celles-ci mais plutôt la capacité de sélection d'une information qui réponde à nos besoins [NIG 01]. De plus la quantité d'information numérique disponible ne fait que croître ; cette augmentation ne peut pas être maîtrisée uniquement par la croissance en puissance des calculateurs. Un projet à l'Université de Berkeley a estimé à un exa-octet² (1 million de téra-octets) la quantité de données générées annuellement de par le monde. Parmi ces données, 99,997 % sont disponibles sous forme numérique [KEI 01].

Dès lors, quelles sont les idées novatrices les plus prometteuses pour permettre l'accès à ces informations par un large public ? Les Interfaces Homme/Machine (IHM) utilisent de plus en plus de multimodalité. Le dialogue ne se fait plus seulement avec des canaux de communication traditionnels (clavier/souris en entrée, et écran en sortie), mais avec plusieurs comme par exemple des observations par caméra, gestes de désignation, objets à retour d'efforts, etc. [ROU 01]. L'homme souhaite communiquer de manière naturelle et intelligente avec les machines. Pour cela, les chercheurs dotent les ordinateurs de capteurs et d'effecteurs. Mais l'efficacité de l'interaction avec l'utilisateur n'est pas seulement fonction du matériel et des données, il faut encore que le système « donne du sens » aux éléments qu'il manipule, or « un ordinateur multimédia n'est pas forcément multimodal » [COU 95]. La conception des Interfaces Homme-Machine (IHM) se centre principalement sur l'utilisateur et les usages. De ce fait, le domaine de recherche des IHM se situe à l'intersection de nombreuses autres disciplines, comme l'informatique, l'ergonomie, la linguistique, la psychologie sociale et cognitive, etc.

Historiquement, c'est en 1983 que Nicholas Negroponte fonde le MediaLab au MIT avec l'objectif d'étudier l'intégration de nouvelles technologies comme la synthèse de la parole, la vision par ordinateur, la synthèse d'image, les nouveaux dispositifs d'interaction comme le visiocasque et les transducteurs gestuels rétroactifs. En France, c'est en 1990 à Grenoble, qu'émerge le concept de

² Un téra-octet = 2 puissance 40 octets ; un exa-octet = 2 puissance 60 octets.

6 Variation, construction et instrumentation du sens

multimodalité en informatique. L'apport de la multimodalité va se concrétiser notamment à travers la robustesse, la flexibilité, l'adaptabilité³ et l'adaptativité⁴ des interactions.

8.3.1 Terminologie

Étant donné que l'IHM est pluridisciplinaire, chaque discipline amène une définition particulière des concepts manipulés. Nous savons qu'une définition claire et précise de chaque terme n'est pas toujours possible, et que cela donne lieu à des interprétations multiples, voire des ambiguïtés lors de travaux ou de projets comportant des équipes pluridisciplinaires. Ainsi le terme « canal » par exemple, ne revêt pas exactement le même sens selon l'approche (marketing, psychologique...). Au sein même d'une discipline comme l'informatique, un canal n'aura pas le même sens pour un spécialiste des réseaux ou pour un expert des interactions hommes-machines. Il nous semble donc pertinent de donner quelques points de repère, afin de fixer la terminologie que nous emploierons dans ce chapitre.

8.3.1.1 Média

Dans la vie courante, un média désigne un support d'information (journal, cassette vidéo, CD audio, DVD, etc.). Dans le domaine des IHM, un média désigne un support technique permettant une communication. Plusieurs points de vue sont étudiés :

- sous l'angle technique, un média est un dispositif physique. Ainsi, un clavier ou une souris sont considérés comme des capteurs, au sens où ils fournissent des informations, en provenance de l'utilisateur et à destination de la machine. Un écran ou une imprimante sont considérés comme des effecteurs, c'est-à-dire qu'ils diffusent de l'information, de l'ordinateur vers l'utilisateur.

- sous l'angle utilisateur, le média peut être perçu comme faisant référence aux éléments sensoriels, perceptuels et cognitifs de l'être humain. On parle alors de média visuel, auditif, audiovisuel, kinesthésique, etc.

8.3.1.2 Multimédia

Un système multimédia est communément identifié comme pouvant supporter plusieurs entrées et/ou plusieurs sorties ; l'élément discriminant un média d'un autre étant la nature de l'information. Un ordinateur multimédia désigne alors un système ou un logiciel informatique qui fusionne plusieurs types de composants tels que le texte, les images, les vidéos... Ce système, capable d'acquérir, de stocker et

³ Adaptation statique des préférences d'usage, des capacités perceptives de l'utilisateur, etc.

⁴ Adaptation dynamique, au cours de l'interaction, des préférences d'usage, des capacités perceptives, etc.

de restituer des informations de natures différentes traite les données au niveau élémentaire du signal, sans interprétation sémantique.

Enfin, une distinction peut être faite entre multi-média et multimédia, notamment dans le domaine de l'enseignement, où la première orthographe désigne l'utilisation de plusieurs médias sur plusieurs machines différentes, tandis que la seconde orthographe indique qu'une unique machine contrôle plusieurs médias, grâce à une numérisation des textes, des sons, des images, etc. (cf. chapitre d'Eric Bruillard dans cet ouvrage et [BRU 97]).

8.3.1.3 Mode

Communément, le mode désigne la manière générale dont un phénomène se présente ou dont une action se produit [TRU 00]. Par exemple, « le mode désigne en linguistique la manière dont le verbe exprime l'état ou l'action (impératif, subjonctif, ...) » [BEL 92] BELLIK, Y., TEIL, D., Les types de multimodalités. Actes IHM92. 4èmes Journées sur l'ingénierie des interfaces Homme-Machine, Paris, 30 Nov - 2 Déc 1992.

[BEL 95]. Dans une communication, le mode de communication se rapporte principalement à l'organe (ou au système d'organes) utilisé pour percevoir ou produire des informations. Un mode, dans le domaine de la téléphonie correspond aux normes utilisées par les réseaux. Typiquement, un téléphone devra être bimode pour fonctionner à la fois sur GSM⁵ et sur UMTS⁶ [WEI 02].

Dans le domaine des IHM, le mode correspond à un état particulier, pour un instant donné, dans lequel se trouve le système interactif. Concrètement, une même action de la part de l'utilisateur sera interprétée de manière différente de la part de la machine, selon le mode en vigueur à un moment donné. Truillet donne l'exemple bien connu de l'éditeur de texte dénommé « vi » du système Unix, qui selon le mode dans lequel il se trouve (mode commande ou mode frappe) n'exécute pas la même action. Un appui sur la touche « x » du clavier permet de taper la lettre « x » de l'alphabet, ou bien cela permet de supprimer le caractère courant, selon le mode dans lequel on se trouve.

Certains équipements sont qualifiés de « multimode ». Ainsi, un boîtier multimode vendu dans le commerce permet aux secrétaires d'utiliser le même micro-casque pour plusieurs tâches au choix (dictée vocale ou conversations téléphoniques). Il comprend un commutateur à trois positions, pour la dictée seule, la conversation seule, ou bien la dictée couplée avec conversation.

⁵ Global System for Mobile communication : norme européenne pour la téléphonie mobile.

⁶ Universal Mobile Telecommunication System : norme européenne destinée à remplacer le GSM et permettant un accès haut débit (2Mbps) aux services d'Internet mobile. Permettra notamment la visiophonie.

8 Variation, construction et instrumentation du sens

8.3.1.4 Modalité

En IHM, la modalité fait référence à la manière d'utiliser les media. Selon [NIG 96] le terme modalité n'est pas toujours utilisé à bon escient. Pour une même définition donnée, certains appliquent le terme modalité, d'autres le terme média. Globalement, on accepte de dire qu'une modalité est une certaine manière d'utiliser un media. Par exemple, écrire, faire un geste, dessiner sont autant de modalités possibles pour transmettre une information en utilisant le media « stylo ». Comme pour le média, la définition de la modalité dépend du point de vue (utilisateur vs système) que l'on adopte. Truillet par exemple, appréhende la modalité en se plaçant toujours du côté de l'être humain, et en se focalisant sur la structure des informations échangées entre les interacteurs. Il explique qu'une structure telle que le braille peut être perçue aussi bien par le toucher que par le visuel.

Sous l'approche hybride, couplant les deux angles d'étude de la conception d'IHM, l'utilisateur et le système interactif, Nigay et Coutaz proposent *la définition d'une modalité comme le couple $\langle p, r \rangle$ où :*

p désigne un dispositif physique (par exemple, une souris, une caméra, un écran, un haut-parleur),

r dénote un système représentationnel, c'est-à-dire un système conventionnel structuré de signes assurant une fonction de communication (par ex., un langage pseudo naturel, un graphe, une table).

L'expression suivante fournit une définition complète de la notion de modalité : modalité ::= $\langle p, r \rangle$ | \langle modalité, $r \rangle$ [NIG 01].

Dans ce contexte, \langle clavier, langage naturel \rangle , \langle microphone, langage naturel \rangle , \langle souris, formulaire graphique \rangle ou encore \langle ordinateur de poche, langage gestuel \rangle sont des modalités. En désignant une modalité comme étant définie par \langle modalité, $r \rangle$ on parle également de composition récursive. C'est le cas où une modalité sert uniquement à définir des unités informationnelles pour un autre système représentationnel, le tout formant alors une modalité. Exemple : Pseudo Langue Naturelle écrite ::= \langle \langle souris, menu \rangle , Pseudo Langue Naturelle \rangle . La modalité \langle souris-menu \rangle consiste à sélectionner des bouts de phrases dans un menu. Le tout forme un dispositif logique qui, combiné avec « Pseudo Langue Naturelle », permet d'obtenir la modalité « Pseudo Langue Naturelle écrite ».

8.3.1.5 Multimodal

De manière courante, une installation dite multimodale permet de faire cohabiter des modes qui d'ordinaire sont utilisés séparément. Par exemple, des infrastructures comme la gare de Lisbonne, ou l'aéroport de Lyon St Exupéry,

forment un pôle multimodal, dans le sens où elles accueillent, simultanément des trains, des bus, des métros, des avions, des voitures, etc. Cela offre (en théorie) une efficacité accrue pour l'utilisateur, qui peut bénéficier de la synergie des modes regroupés dans une même infrastructure.

A la différence du multimédia, le concept de multimodal est employé, en informatique, lorsque l'on donne du sens aux informations provenant (ou en direction) de plusieurs médias. Le niveau physique du multimodal fait appel à la perception tandis que le niveau représentationnel fait appel à la cognition. Notons cependant que certains auteurs font référence au multimodal en utilisant le terme multimédia intelligent [NEA 90].

8.3.2 *La Multimodalité*

Ce terme est important dans le développement de notre chapitre, c'est pourquoi nous lui consacrons une partie importante au sein de ce chapitre. La multimodalité fait référence à l'usage de plusieurs modalités pour la réalisation de la même tâche. Cela doit permettre, en théorie, une *meilleure* interaction et faciliter l'utilisabilité des systèmes proposés. Les critères ergonomiques d'évaluation de cette multimodalité sont souvent le caractère naturel de l'interaction, la robustesse de l'interaction, et sa flexibilité selon le contexte et les types d'utilisateurs.

Depuis les années 80, la multimodalité est étudiée dans le cadre des interactions entre l'homme et la machine. Les chercheurs s'intéressent à la multimodalité dite « en sortie », c'est-à-dire, en variant les canaux de présentation de l'information, en provenance de la machine et à destination de l'utilisateur, mais également, « en entrée », en permettant à l'utilisateur d'interagir de plusieurs manières différentes avec le système, voire de coupler les canaux de communication. Depuis les travaux de Bolt, et son paradigme du « Put that here » [BOL 80], où il combine des entrées vocale et gestuelle, de nombreux travaux ont été menés pour étudier l'utilité et l'utilisation réelle de la multimodalité en IHM.

Les approches théoriques pour décrire et concevoir ces systèmes multimodaux essayent de prendre en compte les visions utilisateur et système dans leurs modélisations. Les usages de la multimodalité, du point de vue de l'utilisateur, sont souvent présentés grâce aux propriétés CARE [COU 94], [COU 95], tandis que les propriétés CASE font référence au point de vue système.

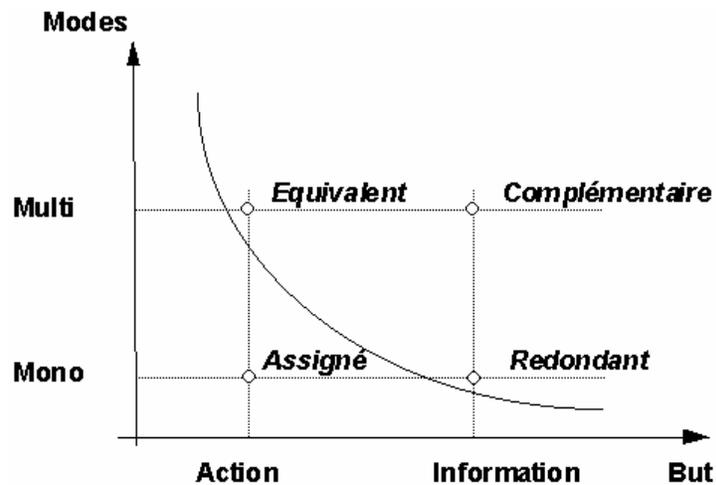


Figure 8.1. CARE : espace de conception centré utilisateur

La Figure 8.1 ci-dessus présente l'espace de conception CARE.

- **C** = Complémentarité. Chaque mode est nécessaire et contribue à la compréhension de l'action.
- **A** = Assignation. L'utilisateur choisit un mode récurrent particulier, ou un sous-ensemble de modes, pour s'exprimer.
- **R** = Redondance. L'utilisateur utilise simultanément plusieurs modes à travers lesquels les informations sont redondantes.
- **E** = Équivalence. L'utilisateur choisit indifféremment tel ou tel mode, ou un sous-ensemble de modes.

La Figure 8.2 ci-après présente l'espace de conception CASE. Jean Caelen explique que cette classification repose sur deux critères : l'usage des médias d'une part qui peut s'effectuer de manière séquentielle ou parallèle, et d'autre part, l'usage des modes (pour l'interprétation ou la génération) qui peut se réaliser de manière indépendante ou combinée [IHM 91].

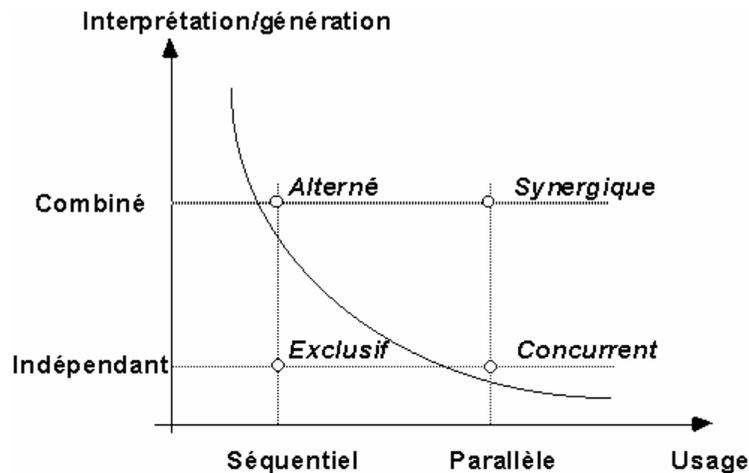


Figure 8.2. CASE : espace de conception centré système

– **C** = Concurrent. Cela fait référence à deux tâches distinctes, en parallèle, sans coréférence. En mode multimodal concurrent, l'usage des médias peut donc se faire de façon parallèle, mais les informations circulant sur ces médias sont indépendantes. Il peut y avoir redondance d'information (exemple : affichage à l'écran et synthèse de la parole) ou conflit si deux commandes contradictoires surviennent en même temps.

– **A** = Alterné. Une tâche est effectuée, avec entrelacement temporel, en coréférence de modalités. En mode multimodal alterné, l'usage des médias est séquentiel et le traitement des informations peut combiner différents modes (exemple : "mets ça là" [BOL 80] en désignant l'objet et le lieu après la fin de la phrase).

– **S** = Synergie. Une tâche est effectuée, en parallèle, en coréférence de modalités. En mode multimodal composé (ou synergique), l'usage des médias s'effectue en parallèle et les traitements sont combinés (exemple : "mets ça là" en parlant et manipulant simultanément). C'est l'aspect le plus proche de la communication naturelle, mais il est complexe à interpréter (cf. ambiguïtés temporelles : dire avant de faire, faire avant de dire).

– **E** = Exclusif. Une tâche unique est accomplie à la fois. En mode multimodal exclusif, deux médias ne peuvent pas être utilisés en même temps. Les informations sont véhiculées par 2 modes indépendants (exemple : clavier, puis souris).

Bellik et Teil [BEL 92] complètent l'approche de Caelen et Coutaz en précisant que les types de multimodalités dépendent de trois paramètres :

- production des énoncés (séquentielle ou parallèle)
- nombre de médias dans un énoncé (un ou plusieurs)
- usage des médias (exclusif ou simultané)

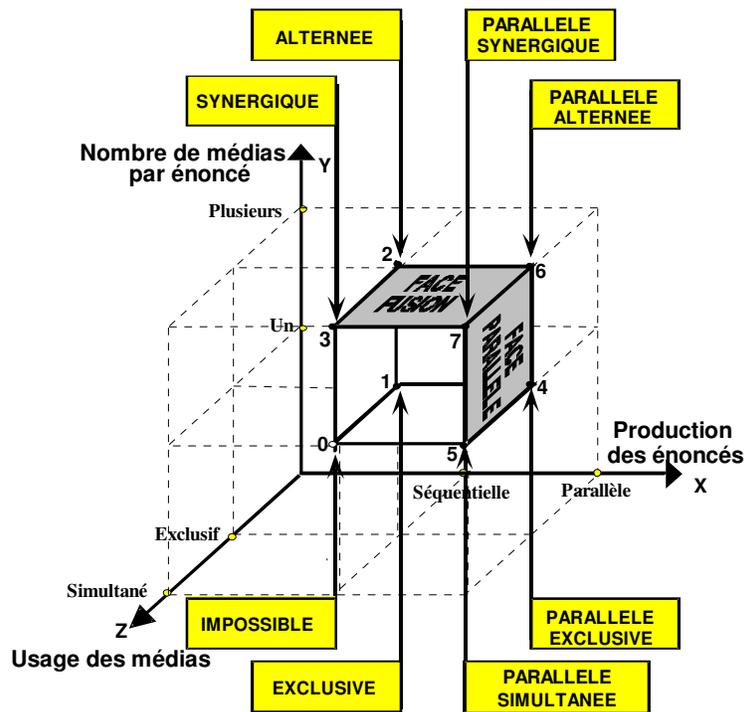


Figure 8.3. Les différents types de multimodalités selon Bellik

Avec cette approche, on a donc sept types de multimodalités possibles ; et non pas huit, puisqu'un sommet du cube représenté en Figure 8.3 est une combinaison impossible : on ne peut pas avoir d'usage simultané des médias si la production des énoncés est séquentielle, avec utilisation d'un seul média par énoncé.

D'autres chercheurs vont encore plus loin. En utilisant les relations de Allen, appliquées aux compositions spatiales, temporelles, articulatoires, syntaxiques et

sémantiques des modalités, ils obtiennent un espace conceptuel représentée par une matrice à 25 positions possibles [VER 00].

Les travaux en multimodalité tendent à montrer que tous les modes de communications ne se valent pas forcément. Différents facteurs seront déterminants pour choisir de concevoir et développer une interface multimodale. Ce choix peut être guidé par la tâche à accomplir : un contrôle aérien, par exemple, pourra éventuellement, être mieux géré si l'humain peut utiliser plusieurs sens simultanément (vision, ouïe, toucher). Un chirurgien qui opère un malade, et qui a déjà ses deux mains occupées pourra utiliser d'autres dispositifs pour piloter une caméra endoscopique (pédales, commande vocale). Le type particulier d'utilisateurs visés est bien entendu un facteur déterminant du choix de la modalité (personnes handicapées, personnes âgées, enfants...). Le lieu, le contexte, le type de périphériques disponibles sont encore autant d'éléments devant être pris en compte (milieu bruité/lumineux, utilisation debout/assis).

D'autres pistes de recherche suggèrent l'emploi de la multimodalité pour pallier les problèmes rencontrés par les interactions traditionnelles basées sur la manipulation directe et le pointage. Cela est valable lorsque l'utilisateur a besoin d'explicitier ses besoins, lors d'une recherche d'information, par exemple, et que les systèmes classiques ne sont plus d'aucune utilité [ROU 01]. Ce sera certainement le cas également, lorsqu'il y aura négociation entre des agents, dans le cadre de l'enseignement à distance ou du commerce électronique. On fait alors référence aux 5 sens humains : montrer le produit, le faire toucher (par systèmes à retour d'effort), le faire sentir (parfums de synthèse proposés par les 3 Suisses à des œnologues, ou expériences dans des cinémas allemands), faire entendre (le claquement d'une portière, gage de fiabilité et de sécurité), etc. Cependant, le fait de concevoir une application de manière multimodale ne présage en rien de la bonne utilisation (et utilisabilité) de cette application. En effet, il est illusoire de croire qu'un système, parce qu'il est multimodal, sera forcément utilisé de manière multimodale par les utilisateurs [OVI 99]. Sharon Oviatt présente d'ailleurs à ce propos les dix mythes de la multimodalité :

- 1) ce n'est pas parce qu'une interface est multimodale que les utilisateurs vont utiliser la multimodalité.
- 2) le pattern parole-pointage n'est pas le plus intéressant.
- 3) la multimodalité ne signifie pas obligatoirement « parallélisme ».
- 4) la parole n'est pas un mode « de base » dans un système multimodal.
- 5) le langage multimodal ne diffère pas du langage unimodal.
- 6) l'interaction multimodale ne favorise pas la redondance.
- 7) les erreurs sur un mode ne sont pas compensées par un autre mode.

14 Variation, construction et instrumentation du sens

- 8) les utilisateurs n'organisent pas « leur » multimodalité de la même manière.
- 9) les modes ne sont pas équivalents.
- 10) un système multimodal n'est pas plus efficace qu'un autre.

8.3.3 *Dialogue et multimodalité*

La multimodalité couplée au dialogue en langage naturel demeure une piste de recherche intéressante, et donne lieu à de nombreux travaux. Plus encore, des interactions multimodales en relation directe avec des activités de communications humaines sont offertes dans bon nombre d'outils que nous utilisons quotidiennement, mais nous avons peu conscience qu'elles le sont effectivement. Nous nous focalisons sur quelques domaines particuliers comme les jeux vidéo (cf. manettes vibrantes, volants et pédales pour simulateurs), ou quelques logiciels de bureautique⁷, mais nous ne prenons pas conscience immédiatement que bon nombre de nos interactions courantes avec les machines qui nous entourent sont multimodales. Quelques exemples suffiront à étayer notre propos :

- (a) l'ascenseur qui annonce vocalement des informations (« vous arrivez à tel étage » ou bien, « attention, vous êtes devant la cellule » (sous-entendu vous empêchez la porte d'ascenseur de se refermer) ;
- (b) le signal sonore de votre véhicule qui indique que vous avez oublié d'éteindre vos phares lorsque vous ouvrez votre portière et que le contact est coupé ; dans les véhicules haut de gamme, tel le Xsara Picasso de Citroën, ce signal sonore et couplé à l'affichage d'un message digital sur le tableau de bord, en langue naturelle, du type « phares allumés » ;
- (c) le téléphone portable qui permet d'attribuer à tel appelant une sonnerie particulière : cognitivement, on obtient plus d'informations car le sens de la sonnerie n'est plus simplement « vous avez un appel » mais plus encore « vous avez un appel de la part de telle personne » ; libre à vous de décrocher ou pas compte tenu de la valeur ajoutée de cette information.
- (d) panneau digital et lumineux comme le plan de la ville dans une gare, que l'on peut toucher — en entrée — et qui allume des diodes pour former un itinéraire et qui parle — en sortie — (cf. Figure 8.4) ;
- (e) son 3D dans les jeux vidéo, palonnier, volant vibrant, manettes à retour d'effort, tapis de combat (cf. Fighting Arena sur la Figure 8.5).

⁷ Le pédalier-secrétariat proposé par l'entreprise Dragon permet de piloter au pied la réécoute des textes transcrits dans Dragon NaturallySpeaking. Il propose trois fonctions : écoute à partir de la position du curseur (pédale centrale), déplacement au paragraphe précédent (pédale de gauche), déplacement au paragraphe suivant (pédale de droite).

Tous ces exemples sont autant de mise en œuvre de dispositif qui peuvent être considérés comme multimodaux, dans le sens où ils autorisent l'utilisateur à choisir un moyen particulier (parmi plusieurs possibles) de communiquer avec son environnement. Ce moyen d'interagir avec un dispositif n'est pas forcément le meilleur, mais il est une alternative possible.



Figure 8.4. Plan digital tactile et lumineux



Figure 8.5. Dispositif pour coups de pied et de poing pour console de jeu Playstation

Nous avons proposé dans [ROU 00] un dialogue multimodal sur Internet permettant d'effectuer une recherche d'information documentaire de plusieurs manières possibles : par manipulation classique (clavier/souris), par interaction vocale en langue naturelle, ou par couplage des deux. De même en sortie, nous avons étudié les apports d'une interaction vocale dialoguée (avec synthèse vocale calculée à la volée) et présentation graphique des résultats en vue dite « fish-eye-

view » (vue en œil de poisson). Notre constat était un manque de langage des programmations pouvant permettre la réalisation de dispositifs multimodaux. De tels langages de programmation commencent aujourd'hui à voir le jour. Le W3C a proposé par exemple EMMA⁸ (Extensible MultiModal Annotation language - <http://www.w3.org/TR/EMMAreqs>) et Ink Markup Language⁹ (<http://www.w3.org/TR/inkreqs>) qui prennent en compte la mise en œuvre et la gestion d'interfaces multimodales sur le réseau Internet.

8.4 Mise en œuvre

L'utilisation de la multimodalité en IHM permet, entre autres, de donner à l'utilisateur plus de liberté quant à la manière d'interagir avec un système informatique. Le fonctionnement du système demeure le plus souvent inchangé, mais la manière de récolter (respectivement « de restituer ») les données en provenance (respectivement « à destination ») de l'utilisateur diffère selon les contextes d'utilisation, les périphériques disponibles, les profils d'utilisateurs identifiés, etc. Dans les lignes qui suivent, nous présentons des exemples de systèmes informatiques multimodaux. Bien qu'il soit possible d'utiliser simultanément de la multimodalité en entrée et en sortie, nous exposons volontairement les deux cas séparément, pour faire apparaître les apports de chacun.

8.4.1 Multimodalité en entrée

Prenons le cas d'un logiciel consacré à la recherche d'informations cinématographiques. L'interface graphique d'une application permettant de sélectionner des films en fonction de leur durée pourrait ressembler à celle proposée en Figure 8.6. Sur un ordinateur personnel, il est aisé de saisir ces données au travers d'un formulaire.

⁸ EMMA is a target data format for the semantic interpretation specification being developed in the Voice Browser Activity, and which describes annotations to speech grammars for extracting application specific data as a result of speech recognition. EMMA supersedes earlier work on the natural language semantics markup language in the Voice Browser Activity.

⁹ The Ink Markup Language will serve as the data format for representing ink entered with an electronic pen or stylus in a multimodal system. The markup will allow for the input and processing of handwriting, gestures, sketches, music and other notational languages in web-based multimodal applications.

Choix d'un film selon sa durée

Durée minimale choisie : heure(s) et minutes

Durée maximale choisie : heure(s) et minutes

Figure 8.6. Interface graphique permettant de sélectionner un film selon sa durée

Sur la majorité des périphériques actuels et dans un proche avenir, on peut envisager une saisie vocale en langue naturelle. Avec une reconnaissance vocale on aurait par exemple un dialogue tel que celui de la Figure 8.7.

(...)

Homme : Je veux effectuer une sélection selon la durée du film

Machine : Quelle est la durée minimale souhaitée pour ce film ?

Homme : Pardon ?

Machine : Vous devez préciser la durée minimale des films recherchés (par exemple 1 heure 40 minutes)

Homme : une heure trente

Machine : Quelle est la durée maximale souhaitée pour ce film ?

Homme : deux heures

Machine : J'ai trouvé 16 films ayant une durée comprise entre 1h30 et 2h. Souhaitez-vous connaître les titres de ces films ?

(...)

Figure 8.7. Exemple de dialogue oral permettant de sélectionner un film selon sa durée

Ici, la multimodalité ne change rien au service rendu par le système informatique, puisque dans un cas comme dans l'autre, seize films répondant à la requête de l'utilisateur ont été trouvés. Idéalement, on doit être en mesure d'utiliser indifféremment l'un ou l'autre des moyens d'interaction proposés, sans surcharge cognitive supplémentaire ou besoin d'apprentissage particulier. En pratique, bien entendu, chaque modalité a un coût propre, qui n'est pas forcément équivalent à celui des autres modalités disponibles. De même, l'utilisation conjointe de modalités ne garantit en rien une meilleure efficacité.

8.4.2 Multimodalité en sortie

Le cas d'étude que nous présentons ci-après traite de l'utilisation de la multimodalité dans le cadre d'une interaction dialoguée (pour partie). Une situation de jeu est mise en œuvre. Le support utilisé est le réseau Internet et la technologie MSA (Microsoft Agent). Microsoft propose, en effet, des agents animés que l'on intègre aux pages Web, en insérant du code dans l'entête d'une page HTML. Ceci est censé améliorer l'interaction entre l'homme et la machine, puisque l'agent peut adopter des attitudes explicites (comme la surprise, l'explication, l'incompréhension, ...), se déplacer sur l'écran, désigner des parties de l'écran (à gauche, à droite, etc.). Ces MS-Agent (<http://msdn.microsoft.com/msagent>) disposent de logiciels de synthèse et de reconnaissance vocale (en anglais, en français, etc.), qui s'installent sur la machine cliente, lors du chargement de la page Web.

Dans notre étude, purement descriptive, il s'agit pour la machine de tirer un nombre au hasard entre deux bornes (par défaut entre 0 et 100). L'utilisateur doit entrer une proposition numérique dans une boîte de dialogue et valider cette proposition en cliquant sur un bouton. Bien que nous nous placions dans une situation « d'apprentissage », dans laquelle l'utilisateur (un enfant par exemple) pourra éventuellement découvrir le mécanisme de dichotomie, à travers le jeu, et à l'aide d'une application multimodale, nous présentons ce dispositif informatique uniquement à titre d'illustration, en ne nous intéressant qu'aux possibilités techniques qu'il apporte. Nous ne traitons en aucun cas du contenu pédagogique du logiciel, et ne nous prononçons pas sur le bien fondé de cette démarche sur un plan éducatif.

Deux profils ont été programmés : novice et expert. En mode expert, le système se contente de répondre « c'est plus que X » ou bien « c'est moins que X » (X mis pour la valeur entrée par l'utilisateur). Dans ce mode, Merlin ne fait pas de gestes. On présente des informations textuelles à l'écran, et l'on diffuse, en parallèle, une voix de synthèse qui prononce exactement la même chose que ce qui a été affiché. En mode novice, l'utilisateur est aidé par l'agent animé, qui lui rappelle les bornes en vigueur, et qui fait des gestes pour expliciter ces bornes.

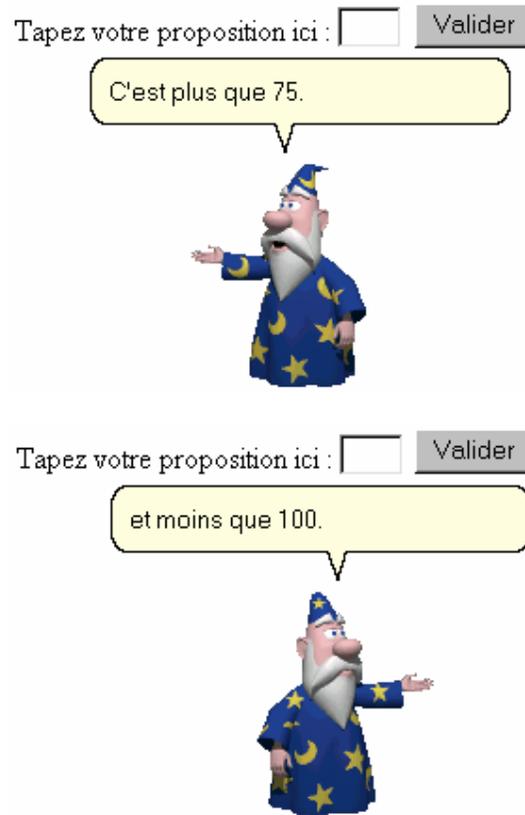


Figure 8.8. L'agent animé Merlin utilise des gestes pour aider l'utilisateur

Dans l'exemple de la Figure 8.8, nous observons une interaction entre les interlocuteurs. L'utilisateur novice vient d'entrer la valeur 75. L'ordinateur compare la valeur donnée avec la valeur à trouver, et calcule les éléments à présenter en réponse. Nous avons programmé trois types d'informations simultanées données à l'utilisateur par l'agent animé Merlin. En effet, il énonce oralement, grâce à une voix de synthèse « c'est plus que 75 et moins que 100 ». D'autre part, ces informations sont également affichées sur l'écran, dans des bulles de dialogue (dont on peut éventuellement paramétrer la taille de la police de caractères). Enfin, Merlin accompagne ses énoncés de gestes de la main, orientés vers la gauche (de l'écran), pour indiquer la borne inférieure, puis vers la droite, pour donner la borne supérieure. Notons que si ces informations sont considérées

20 Variation, construction et instrumentation du sens

comme redondantes, ici, dans le cadre précis de cet exercice, nous ne considérons pas qu'elles sont strictement équivalentes, d'un point de vue sémantique, étant donné qu'un découpage de l'énoncé est effectué, pour distinguer la borne inférieure de la borne supérieure.

Dans l'exemple de la Figure 8.9, l'utilisateur a demandé de l'aide au système. Merlin lui rappelle la valeur des bornes de la session en cours (80 et 50) et explique comment procéder par dichotomie pour proposer une valeur à tester optimale (65 dans notre exemple, c'est-à-dire $50+15$ ou $80-15$).



Figure 8.9. Merlin explique comment procéder à une dichotomie

L'utilisateur est aidé par le fait qu'il ne doit pas forcément mémoriser les bornes, puisqu'elles sont recalculées et redonnées à chaque intervention de l'agent animé. D'autre part, on entend et voit simultanément l'explication qu'il donne.

Un apport important de cette multimodalité en sortie semble être la possibilité de fournir des éléments pertinents pour une meilleure compréhension, de la part de l'utilisateur, de ce que fait l'ordinateur. Les animations (Merlin qui écrit dans un livre, qui se frotte le menton, qui pense, etc.) sont autant de points permettant de comprendre ce que fait le système à chaque instant (cf. théorie de l'activité de Donald Norman [NOR 86]). L'agent animé est alors perçu, semble-t-il, comme la machine elle-même ; l'utilisateur fusionnant mentalement les capacités de la machine à celles de son interlocuteur. Cela reste cependant à vérifier de manière scientifique, en procédant à des mesures et des évaluations (en laboratoire et en situation réelle d'utilisation) qui jusqu'à présent n'ont pas encore été effectuées.

8.5 La multimodalité au cœur des IHM

Nous avons procédé à un rappel des notions importantes utilisées en communication homme-machine dans le cadre de la multimodalité. La souplesse et le naturel escomptés laissent à penser qu'à terme, toute interface sera utilisable de manière multimodale. C'est déjà le cas pour bon nombre de tâches quotidiennes mais nous n'y portons pas forcément attention. Hormis certains cas particuliers, pour lesquels des contraintes sont fortes (handicap de l'utilisateur), les apports de la multimodalité au sens des IHM résident, à notre sens, dans la liberté de choix laissée à l'utilisateur lors de la manipulation de l'interface. Une des critiques classiques, à l'égard de la multimodalité, que l'on entend parfois de la part de personnes qui ne sont pas spécialistes du domaine, consiste à s'interroger sur le bien fondé d'une telle approche. Le néophyte tente maladroitement de montrer que l'approche multimodale n'est pas convaincante, sous prétexte qu'une approche dite « traditionnelle » peut en faire autant.

L'apport de la multimodalité est tout autre. Cela permet à l'utilisateur de partager du sens avec la machine par plusieurs moyens d'interactions différents (complémentaires, redondantes, mutuellement exclusives, etc.). Chaque utilisateur choisit le meilleur type d'interaction possible selon certains critères : la tâche à accomplir, le contexte, les habitudes, l'environnement (bruité, lumineux ...), le niveau de stress, le temps imparti pour mener à bien cette tâche, etc.

Actuellement, les travaux en multimodalité évoluent vers les modalités dites sensorielles, de présentation, et d'action. On s'intéresse de plus près aux interactions en grand (salle où l'utilisateur se trouve immergé, tableau « magique » avec lequel il peut interagir dans un environnement de réalité augmentée [COU 02]) et en petit (dispositifs miniaturisés, dispositifs mobiles, sans fil, etc.). Dans le cadre de leurs recherches sur les gestes médicochirurgicaux assistés par ordinateur,

22 Variation, construction et instrumentation du sens

des chercheurs grenoblois du laboratoire TIMC¹⁰ ont mis au point un système multimodal, permettant à un chirurgien d'utiliser sa langue comme transmetteur¹¹ d'information. Lors de la manipulation d'un instrument dans le corps du patient, le médecin reçoit à distance de très faibles impulsions électriques qui lui indiquent toute déviation de sa trajectoire. Il peut ainsi corriger son geste, sans avoir à consulter les traditionnels moniteurs (écrans de contrôle). Ce sont donc d'autres informations que celles conventionnellement utilisées qui font sens ici.

L'exemple que nous avons présenté utilisait la technologie MSA pour illustrer notre propos, et faire apparaître les apports de la multimodalité, tant en entrée qu'en sortie. D'autres recherches dans cette voie, utilisant des agents animés, sont engagées au MIT par exemple, dans lesquelles le contexte joue un rôle important pour la compréhension des situations. Le « context aware assistant » de [YAN 00] est en effet un agent animé capable de gérer une situation de filtrage à l'entrée d'un bureau par exemple, en utilisant des capteurs sensoriels qui lui permettent de savoir combien de personnes sont déjà présentes dans la pièce. Il cherche ensuite la meilleure façon de gérer l'activité en fonction des contraintes : faire patienter le visiteur, lui proposer un autre rendez-vous selon les disponibilités de chaque personne (visiteur / visité), envoyer un message électronique à la personne travaillant dans son bureau pour lui signifier une urgence, etc.

Par ailleurs, on observe que certaines nouvelles interfaces personnes/organisation sont disposées à intégrer des composants multimodaux dans leurs architectures. C'est le cas notamment pour les grands groupes de vente à distance, qui, par l'intermédiaire de systèmes vocaux interactifs gérés dynamiquement, pourraient résoudre leurs problèmes de maintenance de messages oraux préenregistrés. On voit également apparaître des travaux d'étude sur le couplage entre les notions de multicanal (utiliser plusieurs canaux : e-mail, fax, SMS, etc.) et de multimodalité, notamment dans les domaines du commerce électronique [DER 02], de l'apprentissage et de l'enseignement à distance. Enfin, la personnalisation va croître, nous semble-t-il, pour répondre de manière beaucoup plus appropriée aux attentes des utilisateurs. Encore une fois, les avancées significatives ne peuvent pas se faire dans un seul champ d'étude. Si les technologies informatiques ont bien entendu un rôle primordial à jouer, le couplage avec les domaines connexes de la cognition devraient créer des conditions favorables pour faire émerger des nouveaux outils informatiques utiles, puissants, robustes et malléables.

¹⁰ <http://www-timc.imag.fr/gmcao/index.html>

¹¹ La langue transmet les informations au cerveau. C'est une multimodalité en sortie : la langue joue le rôle de capteur et non pas d'effecteur, dans ce cas précis.

Bibliographie

- [BEL 92] BELLIK, Y., TEIL, D., Les types de multimodalités. Actes IHM'92. 4èmes Journées sur l'ingénierie des interfaces Homme-Machine, Paris, 30 Nov - 2 Déc 1992.
- [BEL 95] Y. BELLIK, D. BURGER, The Potential of Multimodal Interfaces for the Blind: an Exploratory Study, RESNA'95, Vancouver, Canada, 9-14 June 1995.
- [BOL 80] BOLT, R.A., Put-that-here: voice and gesture at the graphic interface. *Computer Graphics*, 14, 262-270, 1980.
- [BRU 97] BRUILLARD, E., *Les machines à enseigner*, Éditions Hermès. Paris, 1997.
- [CHA 02] CHARPAK, G., BROCH, H., *Devenez sorciers devenez savants*, éditions Odile Jacob, Paris, 2002.
- [COU 94] COUTAZ, J., NIGAY, L., Les propriétés "CARE" dans les interfaces multimodales, IHM'94, Lille 1994.
- [COU 02] COUTAZ, J. LACHENAL, C. BERARD, F. BARRALON, N., Quand les surfaces deviennent interactives, in *Les cahiers du numérique*, Lavoisier, Vol. 3, Numéro 4-2002, pp.101-126.
- [COU 95] COUTAZ, J., NIGAY, L., SALBER, D., BLANDFORD, A., MAY, J., Young R., Four Easy Pieces for Assessing The Usability of Multimodal Interaction: The CARE properties, InterAct'95, Lillehammer (Norway), June 1995, pp. 115-120.
- [DER 02] DERYCKE, A., ROUILLARD, J., La Personnalisation de l'Interaction dans des Contextes Multimodaux et Multicanaux : une Première Approche pour le Commerce Electronique, IHM 2002, Poitiers, 2002.
- [EFR 41] EFRON, D., *Gesture, Race and Culture*. The Hague: Mouton & Company, 1972. Reprinted from *Gesture and Environment*, New York: King's Crown Press, 1941.
- [EKM 69] EKMAN, P., FRIESEN, W., The repertoire of Non-verbal behavior: Categories, Origins, Usage, and Coding. *Semiotica*, 1, 49-98, 1969.

- [EKM 75] EKMAN, P., FRIESEN, W., *Unmasking the face*. New Jersey: Prentice-Hall, 1975.
- [EKM 78] EKMAN, P., FRIESEN, W., *Facial action coding system*. Palo Alto, CA: Consulting Psychologists Press, 1978.
- [IFR 94] IFRAH, G., *Histoire universelle des chiffres*, Robert Laffont, Tome 1 et 2, Paris, 1994.
- [IHM 91] *Production des participants en ateliers, IHM91, Troisièmes Journées sur l'Ingénierie des Interfaces Homme-Machine*, Dourdan, 11-13, 1991.
- [KEI 01] KEIM, D., *Visual Exploration of large data Sets*. *Communications of the ACM*, Vol. 44., N. 8, 2001, p. 39-44.
- [LEM 86] LE MOIGNE, J.L., *Genèse de quelques nouvelles sciences : de l'intelligence artificielle aux sciences de la cognition*, *Intelligences des mécanismes - Mécanismes de l'intelligence*, Fayard, 1986.
- [MCN 92] MC NEILL, D., *Hand and Mind: What gestures reveal about thought*, Chicago, IL, University of Chicago Press, 1992.
- [NEA 90] Neal, J. G., Shapiro, S. C., *Intelligent Multi-Media Interface Technology*, in *Intelligent User Interfaces*, J. W. Sullivan and S. W. Tyler, eds., ACM Press, 1990.
- [NIG 96] NIGAY, L., COUTAZ, J., *Espaces de conception des interfaces multimédia et multimodales*. *Revue : TSI, numéro spécial Multimédia et Collecticiel*, Volume 15, N° 9, 1996, AFCET & HERMES Publ, pp. 1195-1225.
- [NIG 01] NIGAY, L., *Modalité d'Interaction et Multimodalité*, Habilitation à Diriger des Recherches, spécialité Informatique de l'Université Joseph Fourier - Grenoble I, 2001.
- [NOR 86] NORMAN, D., *User Centered System Design, New Perspectives on Human-Computer Interaction*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1986.
- [OVI 99] OVIATT, S., *Ten myths of multimodal interaction*. *Communications of the ACM*, Vol. 42, N. 11, 1999, p. 74-81.
- [PEK 02] PEKKOLA S., HIEKKILÄ J., TUUNAINEN V., *Launching Multi-Modal Interaction on Ec-site*. *Proceedings of the 35th Hawaii International Conference on System Sciences*, IEEE press, 2002.
- [POY 80] POYATOS, F., *Interactive functions and limitations of verbal and Nonverbal behaviors in natural conversation*, *Semiotica*, 30-3/4, 211-244, 1980.
- [PRI 96] PRINCE, V., *Vers une informatique cognitive dans les organisations*, *Collection Sciences cognitives*, Masson, 1996.
- [RIM 91] RIMÉ, B., SCHIARATURA, L., *Gesture and speech*. In R.S. Feldman & B. Rimé, *Fundamentals of Nonverbal behavior*, 239-281. New-York : Press syndicate of university of Cambridge, 1991.
- [ROU 00] ROUILLARD, J., *Hyperdialogue sur Internet. Le système HALPIN*, Thèse de doctorat d'informatique, Université Grenoble I, 2000.

- [ROU 01] ROUILLARD, J., Dialogue et Multimodalité. Revue RIHM - Revue d'Interaction Homme-Machine. Vol 2, N°1, pp.99-125, 2001.
- [ROU 02] ROUILLARD, J., A multimodal E-commerce application coupling HTML and VoiceXML, Eleventh International World Wide Web Conference, Waikiki Beach, Honolulu, Hawaii, USA, 2002.
- [THO 96] THORISSON, K. R., Communicative Humanoids A Computational Model of Psychosocial Dialogue Skills, Ph D, MIT, 1996.
- [VER 00] VERNIER, F. NIGAY, L., Espace de Conception pour les Interfaces Multimodales, Acte du colloque sur les interfaces multimodales, Grenoble, 2000.
- [TRU 00] TRUILLET, P., Support de cours d'IHM de DESS SIGMA, Université de Toulouse, 2000.
- [WAL 98] WALTER, H., *Le français d'ici, de là, de là-bas*, 1998.
- [WEI 02] WEISS, S., *Handheld Usability*, John Wiley & Sons, 2002.
- [YAN 00] YAN, H., SELKER, T., *Context-Aware Office Assistant*, proceedings of 2000 International Conference on Intelligent User Interfaces, New Orleans, Louisiana, 2000.

Index

C

Communication, 1, 4, 5, 6, 7, 8, 9, 11,
13, 14, 21, 24

D

Dialogue, 1, 2, 3, 4, 5, 14, 15, 17, 18, 19,
25

I

Interface, 1, 5, 7, 13, 16, 17, 21, 22, 23,
24, 25
Internet, 7, 15, 18, 24

L

Langage, 4, 5, 8, 13, 14, 16

M

Machine, 1, 2, 3, 4, 5, 6, 7, 9, 14, 17, 18,
21, 23, 24, 25

O

Ordinateur, 2, 4, 5, 6, 8, 16, 19, 21

S

Sens, 1, 2, 3, 4, 5, 6, 9, 13, 14, 15, 21,
22

U

Usage, 5, 6, 9, 10, 11, 12, 23
Utilisateur, 1, 3, 5, 6, 7, 8, 9, 10, 13, 14,
15, 16, 17, 18, 19, 20, 21, 22

Table des matières

8.1	Les interactions Hommes-Machine.....	2
8.2	Qu'est ce que computer ?.....	3
8.3	La multimodalité en IHM.....	5
8.3.1	Terminologie.....	6
8.3.1.1	Média.....	6
8.3.1.2	Multimédia.....	6
8.3.1.3	Mode.....	7
8.3.1.4	Modalité.....	8
8.3.1.5	Multimodal.....	8
8.3.2	La Multimodalité.....	9
8.3.3	Dialogue et multimodalité.....	14
8.4	Mise en œuvre.....	16
8.4.1	Multimodalité en entrée.....	16
8.4.2	Multimodalité en sortie.....	18
8.5	La multimodalité au cœur des IHM.....	21

Web services and speech-based applications around VoiceXML

José Rouillard

Laboratoire LIFL (Trigone) - CUEEP - Bat B6
 Université des Sciences et Technologies de Lille
 59655 Villeneuve d'Ascq Cedex - France
 Email: jose.rouillard@univ-lille1.fr

Abstract—VoiceXML applications are described by context-free grammars. Then, recognized vocabulary is limited. We show in this paper one possible approach in order to use VoiceXML applications and speech web services, together. The idea is to use capabilities of speech-dictation systems when the input value is not recognized in an active grammar. A prototype of VoiceXML application using speech web services is presented. A user can speak a free sentence, in English for instance, and receive a French translation, on the same modality (phone) or via another one (PC screen, for example).

Index Terms—Web services, Speech interfaces, VoiceXML, Context free speech

I. INTRODUCTION

Telephones are much more numerous than computers on the planet. That assertion can explain why vocal technologies and interfaces are an important part of Human-Computer Interaction (HCI) area. Using natural language within interaction is supposed to facilitate exchanges between humans and machines. That's why simple and efficient vocal interactions are awaited in many domains such as E-health, E-learning, E-trade, M-trade or E-administration. These factors pushed the World Wide Web Consortium (W3C) to work on this direction and to publish a recommendation concerning a vocal interaction language, based on XML, which allows describing and managing vocal interactions on the Internet network.

VoiceXML is becoming the standard language used for developing interactive voice response and speech-enabled applications. The key idea is to use the same networks, architectures and protocols as previous applications deployed on the Web. More precisely, VoiceXML is a programming language, designed for Human-Computer audio dialogs that feature synthesized speech, digitized audio, recognition of spoken, DTMF (Dual Tone Multi Frequency) key input, recording of spoken input, telephony, and mixed initiative conversations. Its major goal is to bring the advantages of web-based development and content delivery to interactive voice response

applications [32], [33], [34]. Since a few years, VoiceXML is used to conceive and develop vocal but also multimodal solutions [1], [24].

VoiceXML 1.0 was published in March 2000, and since March 2004, a W3C recommendation of the version 2.0 is available. The changes which appear between the two versions are relatively light. In the short run, new VoiceXML 3.0 version should allow multimodal interactions [6]. Additionally, X+V language [36] (XHTML+Voice Profile 1.0) is available since December 2001 and is designed for Web clients that support visual and spoken interaction at the same time. Users can choose to enter information through a traditional Web browser (Opera for instance), or to speak to the computer, via a microphone connected to the PC.

But VoiceXML applications are described by context-free grammars. Thus, recognized vocabulary is limited. We will see in this paper the potential power to couple VoiceXML applications and context free grammar speech-based web services. The idea is to use capabilities of speech-dictation systems, seamlessly connected to VoiceXML applications using web services.

The paper is organized as follows. Section 2 describes VoiceXML platform features. Section 3 illustrates the proposed solution. Section 4 discusses this solution.

II. VOICEXML PLATFORM FEATURES

A VoiceXML platform is the foundation for developing and operating voice automation applications. During the Human-Computer interaction, it executes the commands and logic specified by applications written in VoiceXML. It also provides the speech processing capabilities (speech recognition, speech synthesis, voice authentication...).

As mentioned in Fig. 1, VoiceXML platform architecture is based on HTTP protocol, and uses both phone and Internet networks. The Web server is often connected to a database which the user can question and update.

A zoom on the VoiceXML gateway shows that this computer is equipped with a telephone card, able to manage incoming and outgoing calls, a VoiceXML interpreter charged to carry out all the orders programmed in this language, and a connection with Internet network.

Based on "Web services and speech-based applications", by José Rouillard, which appeared in the Proceedings of the IEEE International Conference on Pervasive Services (ICPS'06), 26-29 June 2006, Lyon, France © 2006 IEEE.

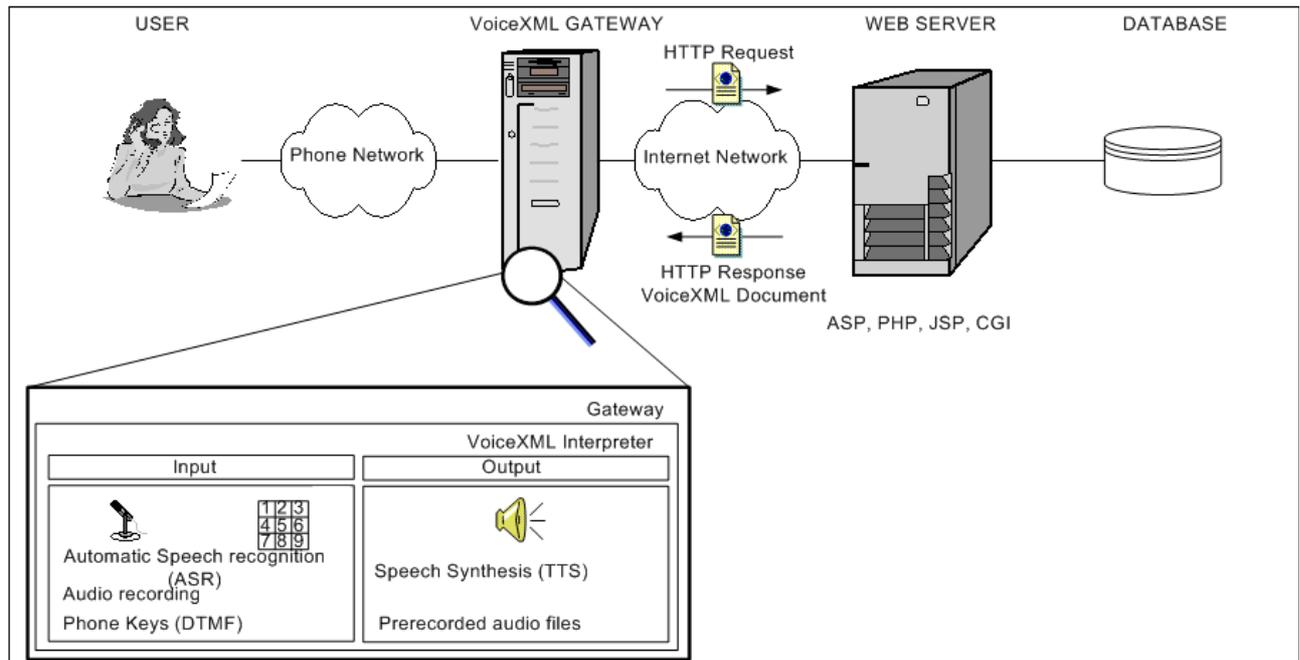


Figure 1. VoiceXML platform architecture

Automatic Speech Recognition (ASR) engine, recording of audio files and Dual Tone Multi Frequency (DTMF) telephone keyboard are the inputs users interact with. Text To Speech (TTS) voice synthesis and restitution of pre-recorded audio files are the outputs the machine can use.

Some VoiceXML features are based on the observation of Human-Human behaviour. The *bargein* attribute, for example, specifies whether or not the caller will be able to interrupt the TTS/audio output with a DTMF keypress or voice utterance. A value of true indicates that the user is allowed to interrupt, while a value of false forces the caller to listen to the entire prompt before being allowed to give input to the application.

A. VoiceXML grammar

In VoiceXML, recognition grammars directly affect the language model used in vocal applications. It is possible to decide if only the system will lead the conversation (system initiative dialogs) or if the user will have the opportunity to anticipate some information (mixed initiative dialogs). This is illustrated by the well known weather forecast vocal application example, in which a user could have the possibility to say "Los Angeles, California", instead of waiting for the two questions: "which city?" and "Which state?".

In a mixed initiative dialog, users have some flexibility in choosing the sequence of interactions. Technically, it means that it is possible to have more than one input field active. It's a choice among concurrently active recognition grammars. But, in a case as in the other, the system must perfectly know what the user is supposed to say. Indeed, a VoiceXML grammar identifies different words or phrases that a user might say: the recognized

vocabulary and syntax are decided *ab initio* by the programmer.

A vocal grammar is, in a certain way, the core of a VoiceXML application, since it determines what the user connected to the server is able to say, or not. Grammars can range from a simple list of possible words to a complex set of sentences. Fig. 2 presents a VoiceXML version 1.0 code example, in which an external grammar is invoked (see Fig. 3).

```
<?xml version="1.0" encoding="iso-8859-1"?>
<vxml version="1.0">
<form>
  <field name="choice">
    <prompt>Tell me your name,
please.</prompt>
<grammar type="application/x-jsgf"
src="users_names.gram"/>
    <filled>Your name is <value
expr="choice"/>.
      <clear/>
    </filled>
  </field>
</form>
</vxml>
```

Figure 2. A VoiceXML code example

A vocal grammar could be given in the body of the VoiceXML script (inline grammar), or in separate file (external grammar). Different kind of grammar could be used such as Nuance GSL [21] or Java Speech Grammar Format [11]. Figure 3 shows the external vocal grammar used in our example. It uses a BNF (Backus Naur Form¹)

¹ John Backus and Peter Naur introduced for the first time a formal notation to describe the syntax of a given language.

notation [19]. The two firsts rules are private, while the last one is public. The sign "|" symbolizes an alternative (a logical "or"). The sign "*", named Kleene star, indicates that the element before this symbol can appear zero or several times. With the square brackets it is possible to declare an element as optional. The brackets define a regrouping. Finally, the sign "+" indicates that element before this symbol can appear one or several times.

```
#JSGF V1.0;
grammar the_names;
<connector> = and|or|but not;
<names>=john|paul|georges|ringo;
public <sentence> = (<names>
[<connector>*])+;
```

Figure 3: A Java Speech Grammar Format (JSGF) code example

We can see on Figure 4 that user can only answer (to the question pronounced by the machine "Tell me your name, please") a sentence following the defined vocal grammar.

```
Computer: Tell me your name, please.
Human: ringo
C: Your name is ringo.
C: Tell me your name, please.
H: georges and ringo
C: Your name is georges and ringo.
C: Tell me your name, please.
H: john but not paul
C: Your name is john but not paul.
C: Tell me your name, please.
H: paul but not georges paul
C: Your name is paul but not georges paul.
C: Tell me your name, please.
H: ringo and georges but not paul and john
C: Your name is ringo and georges but not paul and john.
```

Figure 4: Example of human-machine dialogue obtained with a VoiceXML application

Here, the important thing to understand is that user can only pronounce a sentence predefined in the VoiceXML application's grammar. In our example, it means that it will be impossible for the computer to understand "Peter" or "Karen", for instance, because those utterances are not available in this grammar. It will be exactly the same thing with a VoiceXML application where the user is supposed to give the name of a shape, a color or an action. To understand a sentence like "move the red triangle", this verb, qualifier and noun must be defined, by the programmer, in grammar rules.

B. Related works

The W3C MultiModal Interaction Working Group (MMIWG) works on standards that will allow users to

interact with applications with a combination of input modalities that include speech, pen, keyboard, and pointing, as well as future input modalities [19].

The MMIWG is developing a XML specification called Extensible MultiModal Annotation (EMMA) for representing the intentions reflected in a user's input, in any modality, along with other information about the circumstances of the utterance [9]. In VoiceXML, the <noinput> tag allows the developer to assign event handlers when the application expects voice or DTMF input, and received nothing from the caller. The <nomatch> tag is useful when the caller inputs a value that is not recognized by any of the active grammars.

In the same way, with EMMA's specifications, the notation for uninterpretable input can refer to any possible stage of interpretation processing engine, including raw transcriptions. For instance, if input speech cannot be correctly recognized or the spoken input is not matched by a grammar, it can be tagged as emma:uninterpreted.

More generally, in voice applications, the performances of voice synthesis can appear disappointing when the system must face ambiguities, when it meets unknown words or proper names. The voice recognition employed is supposed multi-speaker, without enrolment phase, and able to be executed in noisy environment. Vocal grammars are inevitably limited: the speaker will not be able to say all he/she wants, but only what have been planned in design phase. This last point is a significant limit that needs particular efforts.

Some related works were driven by Casey Chesnut, in early 2003, with the FreeSpeech Project [8]: the challenge was to speak freely with a vocal application, running on a Pocket PC, thanks to a speech web service connected to SAPI [27] Voice Recognition engine. In the EvalVoiceXML project (Evaluation of Intelligent Component Technologies for VoiceXML Applications), Mittendorf and other authors proposed to explore possible interfaces between VoiceXML and NLU (Natural Language Understanding) modules.

The first idea was to use a trivial grammar that recognizes an arbitrary string of words, disregarding any grammatical or semantic constraints, and to pass the resulting string of words on to a NLP (Natural Language Processing) system.

The second approach was to use raw speech data. Our works are following this suggestion: "As an extreme approach, the VoiceXML speech recognition mechanisms could be bypassed completely by recording a user utterance and passing the unprocessed audio data to an NLU with its own speech recognition capabilities." [16].

Since the first version of VoiceXML, the concept of transcription from audio to written text was evoked. Indeed, the standard proposed in appendix a <transcribe> element to add into the next version of VoiceXML. However, none of the successive versions of the VoiceXML language support this tag. Moreover, <transcribe> element is not implemented by any manufacturer, and will not be available either in the future version of VoiceXML 3.0.

III. POSSIBLE SOLUTIONS

A possible solution to this issue consists in employing a large vocabulary voice recognition system (such as Nuance - Dragon Naturally Speaking [7] or Telispeech from Telisma [29]), independent of that used by the VoiceXML platform, which allows the transcription of text not awaited by the grammar used in the vocal application. This speech system is used on demand by the VoiceXML Server. An advantage of this method is that all limitations of VoiceXML (e.g. handling of unknown words) are resolved.

A. Scenarios

In this section we give some propositions that could bring interesting solutions.

a) A user connects to a vocal server with its (wired or mobile) phone. The computer suggests recording a free speech. Therefore, the user speaks a free sentence, in English for instance, and receives a French translation, on the same mode (its phone) or via another mode (a screen for example). This is normally unconceivable, in classical VoiceXML, because the platform's speech recognition engine can't do prediction;

b) Some users call a vocal server and record free vocal messages. An electronic white board, located in a "public" place of this organisation (cafeteria, meeting room, etc.) is in charged to display the textual messages to everybody, with the name of the person, date, time, priority, etc. It could be, for example "*Martin, 3:05 pm: Do not forget the lab meeting tomorrow, 10 o'clock, room 206*"; "*Tina, 9:17 pm: A new book concerning VoiceXML is available*" ...

c) A user pronounces a word or a sentence in the phone, and relevant links are obtained thanks to a web service provided by a search engine like Google for example [12]. According to the user's choices, results can be sent by mail, fax, SMS, or synthesised by phone;

d) A person is reading a book while she is waiting for her train. She encounters an unknown word and wants to get a definition. A first web service gives a plain text (i.e. the transcription) when it receives a recorded speech. A second web service provides a complete dictionary definition of a word. The synthesized (TTS) definition is given to the user. She will also receive the complete definition in her mailbox. Figure 5 shows a sample example of E-mail obtained by a user that pronounced the word "medicine".

```

medicine
  n 1: the branches of medical science
that deal with nonsurgical
      techniques [syn: {medical
specialty}]
  2: (medicine) something that treats or
prevents or alleviates
      the symptoms of disease [syn:
{medication}, {medicament},
      {medicinal drug}]
  3: the learned profession that is
mastered by graduate training
      in a medical school and that is
devoted to preventing or
      alleviating or curing diseases and
injuries; "he studied
      medicine at Harvard" [syn:
{practice of medicine}]
  4: punishment for one's actions; "you
have to face the music";
      "take your medicine" [syn: {music}]
  v : treat medicinally, treat with
medicine [syn: {medicate}]

```

Figure 5: Sample of E-mail containing the definition of the word "medicine", that the user pronounced freely

All those scenarios need more than a traditional VoiceXML platform to be implemented. Indeed, they are mainly based on a speech-to-text translation feature that is not available for the moment in basic VoiceXML specifications.

B. Architecture

The decomposition of the proposed architecture (see Fig. 6) can be described as the following:

- (1) A VoiceXML application, with no defined grammar, activated by the user, asks to record a free sentence. This vocal recording is converted into an audio file;
- (2) It is transmitted to a traditional server via HTTP;
- (3) This audio file is then treated by an independent speech recognition tool;
- (4) The obtained transcription is transmitted to the vocal server;
- (5) The VoiceXML application can use the textual version of the user input.

C. Technical feasibility

In order to prove the technical feasibility of our project, we have implemented different version of the presented scenarios (synchronous and asynchronous). As we will see on Fig. 8, the chosen architecture is based on the notion of service. We used SOAP [28] to supply speech web services.

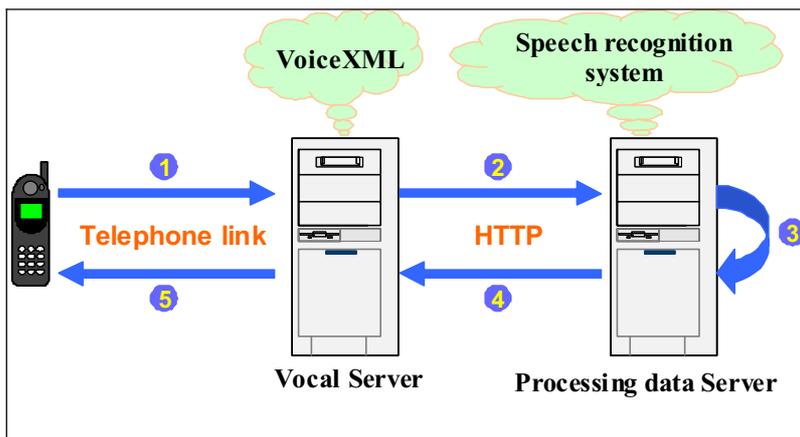


Figure 6. Global architecture

A Service-Oriented Architecture (SOA) is essentially a collection of services [20]. Some ways of connecting services to each other is needed, because a service is seen as an action carried out by a component "supplier" to the attention of a "consuming" component, possibly based on another system [13].

The detailed architecture given on Fig. 8 shows that a vocal server managed in VoiceXML can be beneficially coupled with web services, in order to transmit and obtain useful information. Our contribution lies mainly in a particular web service, able to transcribe free speech into plain text. Technically, this could be done by different means: we present briefly the two principal possibilities chosen: by URL and by DIME attachment. In our project, the first solution was considered as an asynchronous solution while the other one was seen as a synchronous one.

Asynchronous solution

It is possible to provide to our web service an audio file, previously recorded. The file is processed asynchronously (some seconds) and we get a written version.

We decided to set an URL as input parameter for this web service (see the top of Fig. 8). Fig. 7 gives an example of Man-Machine dialogue obtained with a vocal server that used VoiceXML coupled to a web service providing a Wave to text transcription.

```

05:19:30 C: At the tone, please record
your message.
05:19:30 C: (audio clip)
05:19:34 A: recording started
05:19:42 A: recording stopped
05:19:42 C: Please wait while I'm
analysing your message.
05:20:00 C: I think you said: Have you
got a swimming pool in this hotel
    
```

Figure 7. Result of the signal processing: in a VoiceXML application, the machine gives back the sentence that the user said, without any grammar.

It is clear that, to be relevant, the dialogue manager will have to be connected to a robust and powerful

Natural Language Understanding system, because, even if the transcription is correct, on a syntactic level, it will not necessarily be easy, on a semantic level, to really understand what the users' expectations were. This work goes beyond the boundary of this article.

Synchronous solution

In synchronous mode, we coupled a traditional VoiceXML application to our web service of transcription. The vocal application allows the recording of a free message from any telephone. This message, converted into audio file, is sent to the web service, in SOAP attachment. The obtained transcription is synthesized and presented in oral form to the user. We also tested possible sequences with other web services of translation (French/English for example) or of dictionary, which run correctly.

Instead of passing an URL to the web Service, it is also possible to use a WS-Attachments protocol, proposed by Microsoft and called DIME (Direct Internet Message Encapsulation). This mechanism allows sending directly an attachment (picture, sound, etc.) to a web services. We used C# language and WSE, within Visual Studio, for our developments. Web Services Enhancements 2.0 for Microsoft .NET (WSE) is a .NET class library for building web services using the latest Web services protocols including WS-Security, WS-SecureConversation, WS-Trust, WS-Policy, WS-SecurityPolicy, WS-Addressing, and WS-Attachments².

The way web Services interact with each other at the message level, including the execution order and the data flow, is called orchestration [23], [3]. Since a few years, an important number of web Service orchestration languages or specifications are emerging, just like BPEL (Business Process Execution Language) [2] WSCI (Web Service Choreography Interface), and BPML (Business Process Management Language) for example [4]. We worked in this project with BPEL and the Oracle BPEL Process manager [22].

² A new version of Windows Communication Foundation is available in its release version as part of the .NET Framework 3.0.

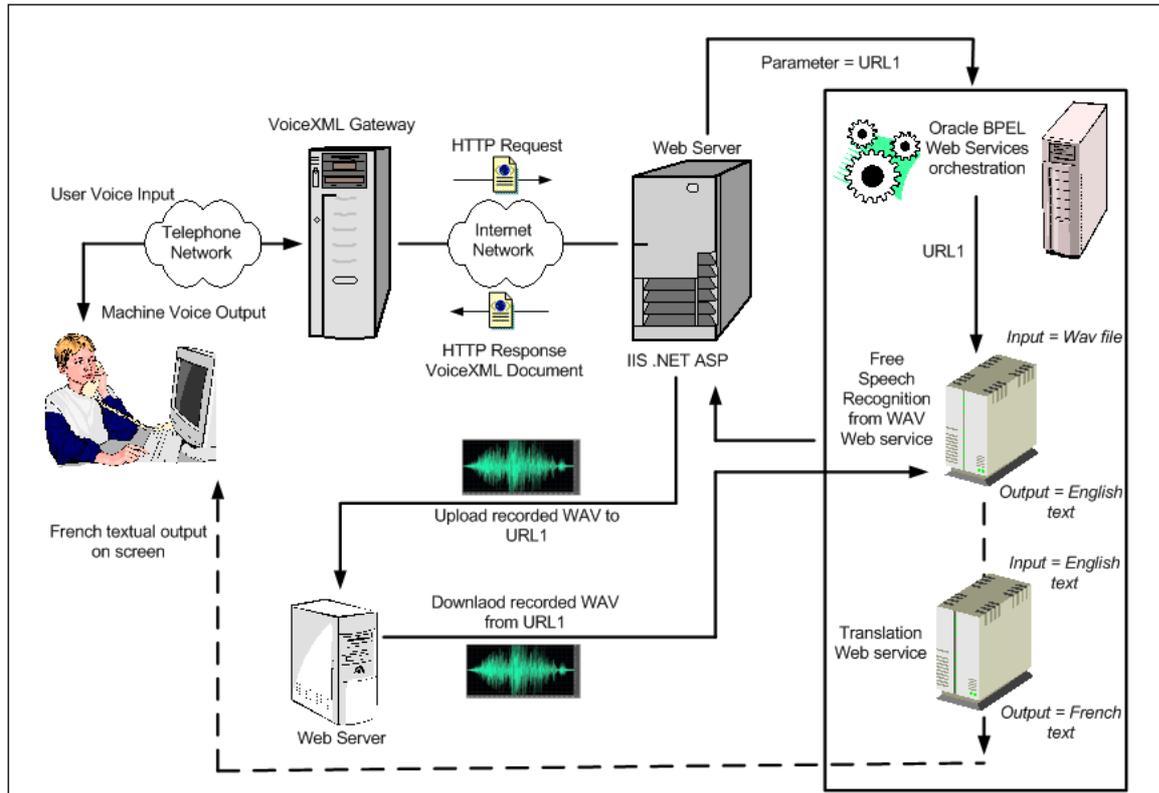


Figure 8. Detailed architecture

An example of Collaxa/Oracle BPEL design view is given on Fig. 9. It shows the web services orchestration used for the scenario number 3 in which the user wants to receive an E-mail containing the definition of the word pronounced on the phone.

Our “Wav_From_URL” web service is invoked and returns the calculated transcription. Therefore, the partner link “Word_definition” is called, with in input parameter the result of the voice recognition. And finally, a “Mailing” web service is called in order to send the definitive answer to the user.

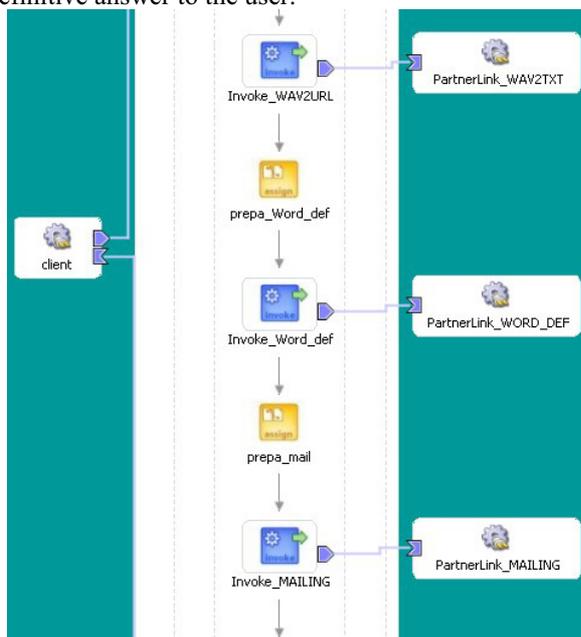


Figure 9. BPEL Oracle Designer view

Fig. 10 is a capture of the Oracle BPEL console used for this project. It provides different views and tools (audit, flow, processes, instances, activities...).

With this view, for example, we can see how the services are managed: the output of one service is the input of the following, and the pronounced sentence “*Is it possible to book a room for the next weekend please*” will be translated in French as “*Est-il possible de réserver une salle pour le week-end suivant svp*”. Here, we are not working on the robustness of the speech translation given by the partner, but we are focusing on the web service’s orchestration. Indeed, in this context, a better English-French translation should give “*chambre*” instead of “*salle*”, and “*prochain*” instead of “*suivant*”, for example. A version entirely implemented with the .NET framework was tested. It increases speed and deployment of the web services by using WSDL [35] for its description and UDDI [30] for its referencing in the directories of web services.

IV. DISCUSSION

The idea of intermixing grammar-based and free-form recognizers is intuitive and everybody working on the subject certainly meets this idea. It could be considered at the same time a classical suggestion (yet another ...) or an important step toward improvement of voice application’s usability. Our discussion with Max Froumentin, from the W3C, in the early 2005 [10] led us to the observation that, at least, two ways were possible in order to work in this direction.

Figure 10. Oracle BPEL results. A French sentence is proposed, based upon an English recorded speech

The first one was trying to improve the VoiceXML language itself while the second one was to work around VoiceXML and the notion of Web services, with the intention of giving new feature to speech-based applications, including those supporting VoiceXML, but not limited to this only language.

Thus, we made a first submission to the community [26], presenting briefly our main suggestions. James A. Larson³ was the first to give a feedback with his paper entitled « VoiceXML on Steroids » [15], where he explained “ *Researchers at the Université des Sciences et Technologies de Lille have implemented a <transcribe> tag within VoiceXML that converts speech to text using a dictation recognition engine rather than the conversational speech engines traditionally used in VoiceXML applications. The <transcribe> tag recognizes free form text without a developer-specified grammar. This tag could be very useful in several situations, for example to allow the system to express some utterances not modeled by the application, but pronounced by users and then provide feedback to users by re-using their input.*

Some others reviews were less optimistic and criticized the idea this way: if the application is going to do anything more than just store the free-form segment for human use, the problem of recognition accuracy, ambiguity, etc. need to be deal with in the overall system. In short, their argument explaining the reason why this was not already in the VoiceXML protocol is that there are too many complications to deal with and keep the protocol simple enough to be usable.

More generally, researchers and practitioners are adding new tags to VoiceXML (2.0 and 2.1). Many of these tags will find their way into VoiceXML 3.0, announced multimodal. From a technical and scientific point of view, using these new tags can result in the generation of higher quality speech applications faster and more efficiently, but this have to be balanced with the economic point of view, and the problems of portability (for example platforms that do not support new tags or functionalities).

We searched solutions close to SOA philosophy and proposed to use Web services to bring complementary features to speech-based applications [25]. According to us, this approach allows the integration, not only in applications that use VoiceXML, but also in multimodal speech-based, a possibility, for the user, to use natural language, across Web services mechanisms.

One possible limitation is related to the time needed to join remote services together. In our first evaluations, the range of elapsed time was from 3.66 to 46 seconds.

³ Dr. James A. Larson is chairman of the W3C Voice Browser Working Group that is standardizing VoiceXML and related markup languages for developing speech applications. Jim works for Intel Corporation, in which he is Manager of Advanced Human Input/Output. He also teaches courses in developing speech applications at Portland State University and the Oregon Graduate School in Oregon Health and Sciences University. His Web site is <http://www.larson-tech.com>

Here we have to face the famous usage/usability conflict, well known in Human-Computer Interfaces sciences. This particularly technical point is essential within the acceptance or not of services by end-users. The time needed to compute raw speech on distant machines and give back the N-best hypothesis solution can be considered for many people as a serious drawback. In another hand, some users seem to be ready to wait a little bit more, if the given solution allows using a large vocabulary instead of a limited grammar.

V. CONCLUSION

We explained that a strong constraint curbs the employment of VoiceXML to the use of a predetermined grammar. By coupling a traditional VoiceXML platform with an independent system of voice recognition, we showed that it is possible to increase its capacities of understanding, since a user can pronounce a sentence initially not configured, at the VoiceXML grammar level. We implemented a few number of the presented scenarios, in order to prove that the solution is technically feasible.

The solution presented in this article, based on a web service able to process a signal can be used both in synchronous and asynchronous way. As we showed, it can be integrated in a web services orchestration, with BPEL, for example. According to us, that opens many directions for multimodal human-computer interface development.

A version entirely implemented with the .NET framework was tested. It increases speed and deployment of the web services. Of course, this service will be more robust if the speech recognition software associated is more powerful.

Our future works are oriented in this direction. Commercial systems (speech recognition software) usually used [7], [31] based on Hidden Markov Models incorporate an acoustic model, a large vocabulary, and a linguistic model. One of the software version [7] used for the test phase is equipped with a vocabulary gathering more than 250000 current and specialized words. Ideally, the use of a voice recognition system, really speaker independent, like the CMU Janus system [14], should improve the robustness of the whole application.

On the top of this, the possibility to change on the fly the desired resources (ASR, most of the time) using MRCP⁴ (Media Resource Control Protocol) [17] will be a way to be more efficient. Best results, according to a context, will be chosen among different speech recognition engines, working at the same time.

ACKNOWLEDGMENT

The present research work has been supported by the French « Région Nord Pas-de-Calais » and the FEDER (Fonds Européen de Développement Régional) during the

MIAOU and EUCUE projects. The authors gratefully acknowledge the support of these institutions. The author also wishes to thank Dr Philippe Truillet, from the IRIT laboratory (Toulouse, France) which is at the origin of a part of this work for his fruitful collaboration, and Rémi Thomas for his help in improving this paper.

REFERENCES

- [1] Anderson, E. A., Breitenbach, S., Burd, T., Chidambaram, N., Houle, P., D. Newsome, D. Tang, X., Zhu, X., Early Adopter VoiceXML, Wrox, 2001.
- [2] Andrews, T., Curbera, F., Dholakia, H., Golland, Y., Klein, J., Leymann, F., Liu, K., Roller, D., Smith, D., Thatte, S. and Trickovic, I, Business Process Execution Language for Web Services version 1.1., Technical report, BEA, IBM, Microsoft, SAP, Siebel Systems, May 2003.
- [3] IBM Web Service BPEL: <http://www-106.ibm.com/developerworks/webservices/library/ws-bpel/>
- [4] Courbis, C., Finkelstein, A., Toward aspect weaving applications, 27th International Conference on Software Engineering (ICSE'05), St. Louis, Missouri, USA, 2005.
- [5] Courbis, C., Finkelstein, A., Weaving Aspects into Web Service Orchestrations, 3rd IEEE International Conference on Web Services (ICWS 2005), Orlando, Florida, 2005.
- [6] Dettmer, R., It's good to talk, speech technology for on-line services access, IEE Review, Volume: 49, Issue:6, June 2003, pp 30-33.
- [7] Dragon NaturallySpeakingTM Preferred, <http://www.scansoft.com/naturallyspeaking>
- [8] Chesnut, C., FreeSpeech project, <http://www.mperfect.net/freespeech>
- [9] EMMA: Extensible MultiModal Annotation markup language, W3C Working Draft (16 September 2005), <http://www.w3.org/TR/emma/>
- [10] Froumetin, Max, The W3C and its Multimodal Interaction Activity, March, 2005. University of Lille, <http://www.w3.org/2005/Talks/0321-maxf-w3c>
- [11] JSGF: Java Speech Grammar Format, <http://www.w3.org/TR/jsgf>
- [12] Google API : <http://www.google.com/apis/>
- [13] Krafzig, D., Banke, K., Slama, D., Enterprise SOA: Service Oriented-Architecture Best Practices, Prentice Hall PTR, 2004, 416 pages.
- [14] Levin L. , Lavie A. , Woszczyna M. , Gates D. , Ga-valda M. , Koll D., Waibel A., The JANUS-III translation system: Speech-to-speech translation in multipledomains, Machine translation, 2000 , vol. 15, no 1-2 , pp. 3 – 25.
- [15] Larson, James A, "VoiceXML on Steroids ", speech technology magazine, November/December 2005, online version available : <http://www.larson-tech.com/Writings/Steroids.htm>
- [16] Mittendorfer, M., Winiwarter, W., Niklfeld, G., Making the VoiceWeb smarter Integrating Intelligent Component Technologies and VoiceXML, Proc. WISE 2001, Kyoto, Japan
- [17] Media Resource Control Protocol (MRCPv1 (RFC 4463)): <http://www.ietf.org/rfc/rfc4463.txt>
- [18] MMIWG, MultiModal Interaction Working Group, <http://www.w3.org/2002/mmi/>
- [19] Naur, P., Revised Report on the Algorithmic Language ALGOL 60, Communications of the ACM, Vol. 3 No.5, pp. 299-314, 1960.
- [20] Newcomer, E., Lomow, G., Service-Oriented Architecture with web services, Addison Wesley, 2005, 444 pages.
- [21] GSL: Nuance Grammar Specification Language, <http://studio.tellme.com/grammars/gsl>

⁴ MRCP is a communication protocol which allows speech servers to provide various speech services (such as speech recognition and speech synthesis) to its clients. See www.ietf.org for details.

- [22] Oracle BPEL Process manager home page: <http://www.oracle.com/technology/products/ias/bpel/index.html>
- [23] Peltz, C., Web Services Orchestration - a review of emerging technologies, tools, and standards, Technical report, HP, January 2003. Technical white paper, http://devresource.hp.com/drc/technical_white_papers/WSOrch/WSOrchestration.pdf
- [24] Rouillard, J., VoiceXML, Le langage d'accès à Internet par téléphone, Paris, Vuibert, 2004.
- [25] Rouillard, J., Web services and speech-based applications, ICPS'06, IEEE International Conference on Pervasive Services 2006 Lyon, 2006.
- [26] Rouillard, J., Truillet, P., Enhanced VoiceXML, HCI International 2005, Las Vegas, 2005.
- [27] <http://www.microsoft.com/speech/download/sdk51/>
- [28] SOAP: Simple Object Access Protocol, <http://www.w3.org/TR/soap>
- [29] Telispeech 1.2 from Telisma, <http://www.telisma.com/>
- [30] UDDI, Universal Description, Discovery and Integration protocol, <http://www.uddi.org>
- [31] ViaVoice (IBM), <http://www.scansoft.com/viavoice>
- [32] VoiceXML 1.0., W3C Recommendation, <http://www.w3.org/TR/voicexml10>
- [33] VoiceXML 2.0., W3C Recommendation, <http://www.w3.org/TR/voicexml20>
- [34] VoiceXML 2.1, W3C Candidate Recommendation <http://www.w3.org/TR/2005/CR-voicexml21-20050613/>
- [35] WSDL: Web Services Description Language, <http://www.w3.org/TR/wsdl>
- [36] X+V, XHTML + Voice Profile, <http://www.voicexml.org/specs/multimodal/x+v/12>

José Rouillard was born in Cavaillon, France, on October 23, 1970. He received his PhD degree in Computer Science from University of Grenoble (France) in 2000.

Then, he joined the Science and Technology University of Lille (USTL) as a lecturer, the same year.

Dr. Rouillard has written the only French speaking book talking about VoiceXML and his research interests include HCI plasticity (see PlasticML for example), multi-modality and multi-channel interfaces. He has written more than 45 articles and refereed conference papers and journals. He is now engaged in research on mobility and pervasive/ubiquitous computing.

Supporting mobile connectivity: from learning scenarios to multi-channel devices

José Rouillard*, Yvan Peter,
Jean-Claude Tarby,
Thomas Vantroys
and Vincent Chevrin

LIFL, Université des Sciences et Technologies de Lille,

59655 Villeneuve d'Ascq cedex, France

E-mail: jose.rouillard@univ-lille1.fr

E-mail: yvan.peter@univ-lille1.fr

E-mail: jean-claude.tarby@univ-lille1.fr

E-mail: thomas.vantroys@univ-lille1.fr

E-mail: vincent.chevrin@univ-lille1.fr

*Corresponding author

Abstract: The introduction of distance learning does not only bring a wider audience, but also much more diversity among the learners: first, because it can be integrated more easily into a Life-long Learning strategy; secondly, because the learners are not restricted to a single area and thus learners from different countries and with different cultures follow the curriculum. We have observed this in various DL diplomas in which we participate. In this article, we will shed some light on the difficulties and challenges arising from these multi-cultural settings. Based on our research work, we would like to insist on two particular points which are the necessity to adapt the pedagogical settings (e.g. pedagogical scenarios) according to the learners' behaviour to overcome unforeseen problems due to cultural differences and the importance of considering mobile technologies to overcome limited access to the technology in developing countries and to ensure continuous interaction among learners and with tutors.

Keywords: Distance Learning; DL; multi-modal and multi-channel communication; pedagogical scenario; workflow.

Reference to this paper should be made as follows: Rouillard, J., Peter, Y., Tarby, J-C., Vantroys, T. and Chevrin, V. (2008) 'Supporting mobile connectivity: from learning scenarios to multi-channel devices', *Int. J. Continuing Engineering Education and Life-Long Learning*, Vol. 18, No. 4, pp.395–409.

Biographical notes: José Rouillard is an Associate Professor in Computer Science at the University of Lille 1. He obtained his PhD in 2000 from the University of Grenoble (France) in the field of Human-Computer Interfaces. He is interested in HCI plasticity, multi-modality and multi-channel interfaces. He has written more than 60 papers and is now engaged in research on mobility and pervasive/ubiquitous computing.

Yvan Peter is an Associate Professor in Computer Science at the University of Lille 1. He obtained his PhD in 1998 from the University of Franche-Comté, France in the field of Middleware Engineering. He is interested in the development of flexible software to support learning and collaboration. He is a member of the scientific committee of the European Network of Excellence Kaleidoscope in the field of Technology Enhanced Learning and responsible for the national project p-LearNet about pervasive learning environments.

Jean-Claude Tarby received his PhD in Computer Science from the University of Toulouse I in 1993. Since 1995, he is an Assistant Professor at the LIFL Laboratory of University Lille 1. His research is focused on the modelling and the design of cooperative applications based on user's tasks. His interests include multimodality, meta-modeling, user interface design methods, groupware and traces of activities.

Thomas Vantroys is an Associate Professor in Computer Science at the University of Lille 1, France. He obtained his PhD in 2003 from the University of Lille 1 in the field of flexible workflow engine for e-learning. He is interested in the SOA and more particularly in service orchestration in the context of pervasive-learning.

Vincent Chevrin is Computer Science PhD from the University of Lille in France. His PhD has largely focused on multi channels aspects in the context of communications and transactions between organisations and users in the e-Commerce area. Currently, he is an Engineer of application and works in the field of marketing. He also continues to work on some challenges close to the multi-channel and pervasive applications issues.

1 Introduction

For a number of curriculum, Distance Learning (DL) has brought not only a wider audience, but also much more diversity among the learners. This is due to two reasons. First, DL can be integrated more easily into a Life-long Learning (LFL) strategy bringing experimented people to assist to the courses. For this reason, the curriculum is attended by learners which have a more diverse experience.

Secondly, the learners are not restricted to a local area where they attend the courses and thus learners from different countries and with different cultures follow the curriculum.

We have observed this in various DL diplomas in which we participate. The first one is a Master in Computer Science, including people from France and from French speaking countries (Tunisia, Algeria, etc.). The second diploma is a Master in Pedagogical Engineering.

In this article, we will shed some light on the difficulties and challenges arising from these multi-cultural settings. We would like to insist on two particular points which are the necessity to adapt the pedagogical settings (e.g. pedagogical scenarios) according to the learners' behaviour to overcome unforeseen problems due to cultural differences and the importance of considering mobile technologies to overcome limited access to the technology in developing countries and to ensure continuous interaction among learners and with tutors.

Based on the illustration of these two points drawn from our practical experience, we describe some research work that we have done towards providing flexibility of the pedagogy and of the platform access modes. On the former, we present an approach based on the modelling of pedagogical scenario that illustrate the tutors' and learners' activities and a solution for the enactment and revision of these scenarios at runtime.

On the latter, we give an overview of the potential of multi-channel communication to provide ubiquitous access to the learning management platforms. To take into account a large range of potential access devices, we need to support different kinds of interaction, from simple modes based on widespread devices, such as mobile phone, for instance to very rich modes (several devices and channels at the same time).

To conclude we link these different solutions in a coherent manner to provide a comprehensive and flexible solution to support the kind of flexibility needed to manage the diversity provided by the cultural and technical differences among learners.

2 Practices and issues of transnational DL

Since more than 30 years, our Institute has a long experience in DL and E-Learning (EL) with a much diversified population composed of people who are workless and need to have a reorientation, and people who look for diploma. For the former, we practice DL by 'classical' ways such as fax, phone, postal mail, etc. For the latter, we opened two professional masters in EL, one in 1998 in computer science, and one in 2003 in computer science and learning science. These EL diplomas are open for French-speaking people. Until now, the students were from Tunisia, Rwanda, Syria and obviously France.

After several years of experience in these two EL diplomas, and beside the classical problems for EL like the lack of motivation mainly due to technical problems and isolation, we have noticed some particular points which are relevant to the multi-cultural dimensions. These dimensions may be factorised into five categories which are now detailed in the remainder of this part 2.

2.1 *Academically-Based Education vs. Life-long Learning*

Our two masters are composed of persons coming from the academic world (Academically-Based Education; ABE) and from the business world (LFL). For the moment, ABE people are less concerned by EL form although the demand for ABE could be more important in the next years because of the evolution of teaching in the French University. Consequently, the LFL people are the most concerned by our EL diploma.

Since our diplomas must be the same in presence and in EL (to comply with legal obligations), we had to adapt our teaching to these two very different populations. This adaptation is based onto five criteria:

- *Age*. The ABE people are about 23 years old, but the LFL people are about 40. Therefore, we cannot talk to the LFL people in the same manner as to the ABE people. Often LFL are older than their teachers and tutors.
- *Culture*. The centres of interest are very different between the ABE and LFL people. In the same way, the ABE students have not the same experience about (professional) life as the LFL learners.

- *Motivation.* Even if both want to pass their degree their, motivation is very different. The ABE students want a diploma to obtain a work when the LFL learners want to evolve in their company, to have their work officially recognised by their boss, or to move to another work or company. Therefore, they do not ask the same type of teaching and competencies.
- *Way of life.* The LFL learners are usually married and have children opposite to the ABE students who are most often single. When we ask to the ABE students to work during 8–12 hours per day, we cannot do it for the LFL learners, because they must manage their family life.
- *Work style.* An ABE student is available during all the week, but an LFL learner is not. Depending on their work and their responsibilities, the LFL students may work on their lessons and homework every day (i.e. one hour per day), once per week (during the week end for instance), etc. More, some LFL people can study at work, but others cannot. Some of them prefer to work off-line (week-end in the country), others on-line. Some of them want only the PDF documents, whereas others request multimedia documents.

For all of these differences, we have to be able to adapt our teachings to satisfy the two-learner populations. For the moment, these adaptations are made *ad hoc* depending upon the people, the diploma, etc.

3 Local and distance students

3.1 Problems about time lag

Some of our LFL learners are living in France, some others are living close to France, and others are living very far from France. These distances can be very harmful for real time examinations when the time lag is too high (until seven hours at the moment). That means that we either must avoid to do real time examinations or to plan such examination with regard to the time lags. The first solution implies a modification of teaching, the second solution implies to look after the duration of examinations and the start and finish time; we cannot ask to a student to have an examination at 3.00am or at 11.00pm!

3.2 Problems about cultural differences

At the beginning, we had some difficulties to write examination subjects comprehensible by everybody. This problem was due to cultural differences. Actually you cannot write subject which are very specific to your country or countries around you. For instance, people from Africa may have some difficulties to understand an examination about ski station, and people from Peru may not know what Société Nationale des Chemins de Fer is (which is the national rail company in France). The cultural difference must be identified at the beginning and the teaching must be adapted consequently.

3.3 Problems about life way

We encountered some unexpected difficulties because some countries have political or economical problems.

Concerning economical limitations, we had for instance a student in Rwanda who had to drive 100 km each time that he wanted to be connected to our EL platform (he had to go to the next big town to find a ‘cyber coffee’). Consequently, he was unable to follow the teaching in the same manner than people who are permanently online. On top of that, the work of this student was liable to water... Why? Because when the rains were not sufficient, there was not enough water in barrages and so electricity in towns and villages! One solution to this problem of lack of water could be the \$100 PC (cf. Figure 1). This solution is very interesting to be independent of electricity. More of that, this new computer should imply deep thinking for teachers about a ‘new EL for this material...’ Another solution is to adapt our teaching to other channels (for instance SMS, phone, etc.).

Figure 1 The ‘\$100’ PC (see online version for colours)



Source: <http://laptop.media.mit.edu>.

Concerning political limitations, Syria decided to block lots of internet connections. The students can send e-mails, but they were unable to surf on all the websites that we provided, neither to upload files onto our platform. We had to find urgently some solutions to resolve these problems, CDROM sent by postal mail for instance.

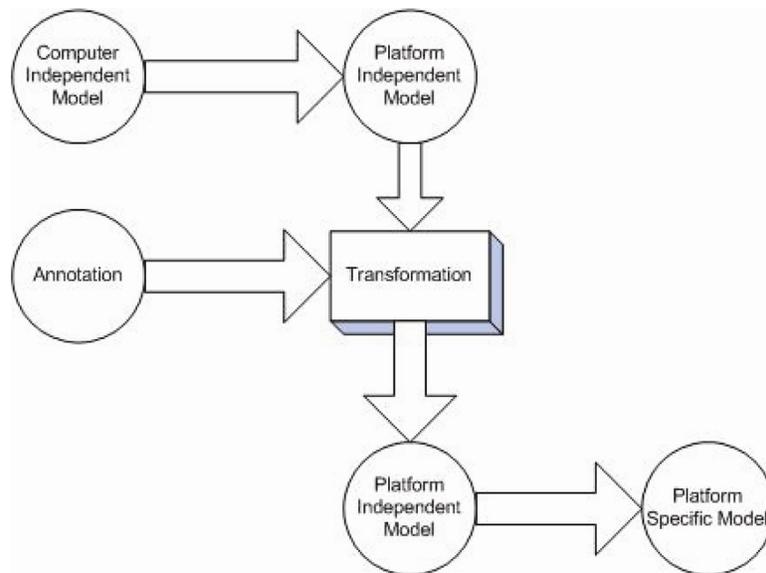
4 Supporting flexible pedagogical settings

As Fischer (2001) pointed out, the challenge about pedagogical flexibility is not only to make information available to people at any time, at any place, and in any form, but also specifically to say the right thing at the right time in the right way.

From the experience presented before, we can see that we need adaptation of the pedagogical setting. For this, we can have two approaches. First, at design time we can abstract from the concrete setting so as to keep the pedagogical objectives in mind. We can then deploy the abstract setting onto a concrete one. The second approach is more concerned with runtime issues and relies on the platform capability to support runtime adaptation of the activities. These two approaches are presented hereafter.

4.1 Modelling and deployment of the pedagogical scenarios

In order to support multi-cultural scenarios, we have to place the model in the middle of our system. In fact, the model driven approach is gaining momentum. We can note that by the fact the Model Driven Architecture (MDA, 2006) promoted by the Object Management Group (OMG, 2006) aims to separate the model of a system from its implementation. The main reason is that a change in the technology field, like changing from Java to C#, should not have consequences on the model because models are persistent and implementation are transient. The same approach can be conducted in Learning Management Systems (LMS). This step resting on successive transformations of models is particularly well adapted to take into account the multi-cultural aspects. It is also well integrated with an approach based on pedagogical scenarios. Indeed, teaching that it carried out in a traditional way or by means of an LMS is not summarised by a simple access to pedagogical resources but through a progress into various coordinated activities (Koper, 2001) (Dillenbourg, 2002). These platforms should not be limited any more to the management of modules and contents, but must support the possibility of really creating complete courses of formation, i.e. organising and accompanying the pedagogical learning paths between various modules. The definition and exchange of these courses requires languages called Educational Modelling Language (EML), around whose efforts of standardisation are in hand (Pernin, 2006). These approaches by scenarios are close to processes handled within the workflow systems (Lin et al., 2001; Vantroys, 2003). The transformation process that allows carrying out our teaching scenarios is presented in Figure 2. Initially, the teacher describes in an abstract way (i.e. non-comprehensible by a computer) the various activities which compose the curriculum as well as the pedagogical aims and the necessary knowledge in order to follow the lessons. That corresponds in the MDA terminology to a Computer Independent Model. In a second time, a teaching engineer transcribes this representation in a more formal version. In our case, we use for that the language IMS-LD (IMS, 2003). That corresponds in the MDA terminology to a Platform Independent Model (PIM). This model has the advantage of being independent of any execution system and of any context. Thus, it is reusable and adaptable according to specific contexts' of use. In a third time, the teacher and the teaching engineer define specific rules to take account of the multi-culturality, like, e.g. execution of the scenario starting from the server or downloadable on the client according to the possible problems related to the infrastructure network. These rules are used for the transformation from PIM to PIM (see Figure 2). In our case, we make a transformation from IMS-LD to XML Process Definition Language (XPDL) (Workflow Management Coalition, 2002). We particularly use the extensibility properties of XPDL to take account of the rules. The last stage consists in deploying the model towards an execution system.

Figure 2 Model transformation process (see online version for colours)

4.2 Flexible support for the execution of pedagogical scenarios

In our works, we focus at each step on the user (teacher, teaching engineer, learner, etc.) and we largely agree with Dourish (2004) when he says that

“[...] the focus of the design is not simply ‘how can people get their work done,’ but ‘how can people create their own meanings and uses for the system in use’; and in turn, this suggests an open approach in which users are active participants in the emergence of ways of working.”

The Cooperative Open Workflow (COW) platform (Peter and Vantroys, 2005), developed in our laboratory aims at producing a bridge between EML centred on the concept of activity and the systems of workflows in order to allow the execution of teaching scenarios. The three principal objectives are:

- to design a platform centred on the user (Bourguin and Derycke, 2005) where he/she acts throughout cycle of life to adapt and modify the system with the various contexts of use
- to relocate the models in the system kernel so they are constantly available and easy to handle
- to install mechanisms of flexibility allowing the realisation of the previous objectives.

This work is focused on a platform of teaching scenarios execution. We do not try to recreate a new LMS but rather to provide a new type of service which could be integrated in existing LMS. One of the keys of integration is to respect standards not only technically, but also pedagogically (use of standardised description languages).

For technical interoperability, we chose a development based on standards J2EE particularly by supporting EJB. In order to extend integration in heterogeneous platforms, our system has a simple object access protocol interface in order to be integrated as a teaching scenarios management web service.

For the languages interoperability, we chose a meta-model approach and we weave bonds between the technical model and the model of the target teaching language. In the following, we will see the various properties which we implemented within our workflow engine in order to answer the identified needs concerning teachers and learners. The main characteristic relates to flexibility during the execution of the scenarios. Most of human activities are not predictable. Thus, it is utopian to believe that we can produce a platform meeting all the needs (present and future). Indeed, it is difficult for a teacher to express and formalise his/her scenario. It will often comprise fuzzy parts which will be cleared up during the execution. Another case can be that of students presenting learning difficulties. In this case, it will be necessary to modify the model in order to add for example additional modules. An alternative to the writing of the scenarios *ex nihilo* is to do reengineering from the execution logs in order to rebuild the used scenario and to formalise it by using for example the IMS-LD language (Barré, El-Kechaï and Choquet, 2005). It is then possible to replay this scenario via our platform. The flexibility brought by COW appears through two ways: the first relates to the possibility of modifying the scenario model during the execution, and the second relates to the possibility of modifying the interpretation of the scenario (i.e. the behaviour of the engine) on the fly.

The first axis is based on introspection and intercession mechanisms and also on the open implementation principle (Kiczales, des Rivières and Bobrow, 1991) which make possible to handle the model in order to add/delete/modify activities and to change paths between activities.

The second axis of flexibility relates to the execution engine itself. Indeed, it is impossible for the programmer to envisage all the use cases of the system. He can, however, use mechanisms of introspection and intercession to modify the behaviour of the platform and to adapt it to particular contexts of use. The way of managing the time constraints is a typical example of behaviour which one must be able to adapt. Indeed, the management of time is one of the other properties of COW. Taking into account of time makes it possible for example to limit the access to resources and tasks and avoids a too significant drift between fastest and slowest learners. The progress of the group is thus more homogeneous.

There are two temporal concepts: duration (minimal and/or maximum time) and dead lines. In order to manage them, we can use various strategies, such as for example taking into account the weekends or not. We can also dynamically adapt calculation according to contexts of use and users wishes. To allow various behaviours when constraints violation appears, we use the pattern strategy, for example, to send an e-mail to the student or the tutor, or to validate in an automatic way the activity.

One of the parameters to be taken into account is the individualisation of the process of training, i.e. the possibility for the students of a group to go at their own rhythm on certain sequences of the scenario. The teacher must specify at the modelling time the different kinds of sequences (individual or collective). A collective sequence does not mean that there is necessarily a collaborative task. Indeed, the teacher can just wish that all the students move ahead at the same time even for an individual work. To support these various sequences, we use the concept of sub-process, i.e. of subsets of scenario

which will be activated only once during the execution for collective activities and once for each learner in the case of individual activities.

5 Supporting mobile and ubiquitous learning

In *Democracy and Education*, Dewey (1916) writes “Education is not an affair of ‘telling’ and being told, but an active and constructive process”. For Dewey, education must be grounded in experience and active inquiry. Supporting mobile and ubiquitous learning is probably an important key of the problem.

Pervasive (or ubiquitous or ambient) learning relies on the ‘always on’ concept (see pervasive games). With an ‘always on’ education, available 24 hours a day, 7 days a week, anywhere, at any time, we are close to the notion of pervasive learning in a social aspect that connects learners not only to communities of people, but also devices and situations. Hence, learners can construct relevant and meaningful learning experiences, in locations and at times that they find relevant.

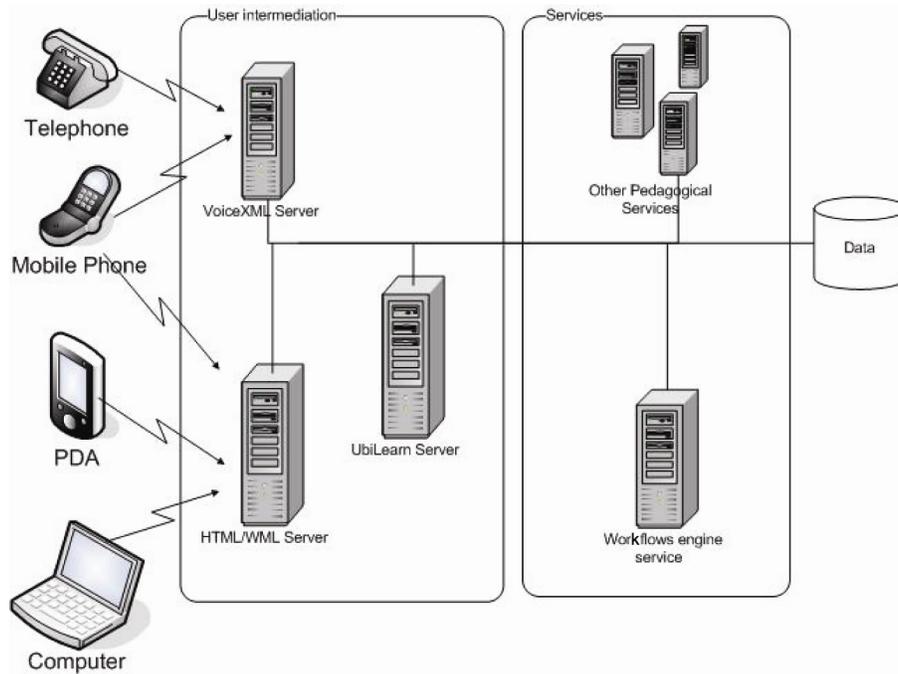
Pervasive learning is related to four key elements: community, relationality, autonomy and locationality. But as Thomas (2005) mentioned: “Pervasive learning is a field still in its infancy”. This line of research goes by a number of names – ubiquitous computing (Weiser, 1991), context-aware computing (Dey, Abowd and Salber, 2001), pervasive computing (Ark and Selker, 1999), embodied interaction (Dourish, 2001), etc.

According to Bomsdorf (2005) (referencing Ogata, Yin and Yano, 2004), the main characteristics of ubiquitous learning are:

- *Permanency*. Learners can never lose their work unless it is purposefully deleted.
- *Accessibility*. Learners have access to their documents, data or videos from anywhere.
- *Immediacy*. Wherever learners are, they can get any information immediately.
- *Interactivity*. Learners can interact with experts, teachers, or peers in the form of synchronous or asynchronous communication.
- *Adaptability*. Learners can get the right information at the right place in the right way.
- *Situating of instructional activities*. The learning could be embedded in our daily life.

Supporting user mobility is for us to offer to learners the possibility to access the platform with mobile devices, for example. This access supposes document adaptation to the peripheral. Our aim is to propose a generic method to transform documents without making one document for each device.

As we address mobile device with few capacity (few memory, few persistent storage size), we wanted to have thin client. The most common used technology for thin client is at present time Hypertext Transfer Protocol (HTTP) and Hypertext Markup Language (HTML). With the concept of thin client, students can access their learning environment with any computer which has a web browser. However, learners use mobile systems as Personal Digital Assistant (PDA) and cellular phone. To support these devices, we build a platform based on a N-tiers architecture (see Figure 3).

Figure 3 Intermediation architecture for user access (see online version for colours)

The first tier is composed by the devices used by learners to access the system. For the moment, we address the following devices:

- *Personal Computer (PC/MAC)*. These devices offer the most interaction possibility (text, graphic, sound). As we wanted thin client, the student access the system with a web browser.
- *PDA*. These devices have smaller screen, graphics must be resized. The student can access the system with a classical web browser or by using WAP (2006).
- *Cellular phone*. These devices offer two types of interactions. The first is the Wireless Application Protocol (WAP) which allows the visualisation of simple page web written in Wireless Markup Language (WML). The second interaction is the use of voice. For vocal interaction we use the VoiceXML language (VoiceXML, 2006) (Rouillard, 2004).
- *Wired phone*. The interactions are restricted to voice and telephone keypad. The student can access the system by calling a vocal server. The second tier is constituted of 'presentation' servers. These servers interact with the workflow server (third tier) to obtain the resources and apply the transformation to transmit document suitable to device used. The last tier is constituted of the database where resources and workflow state are stored.

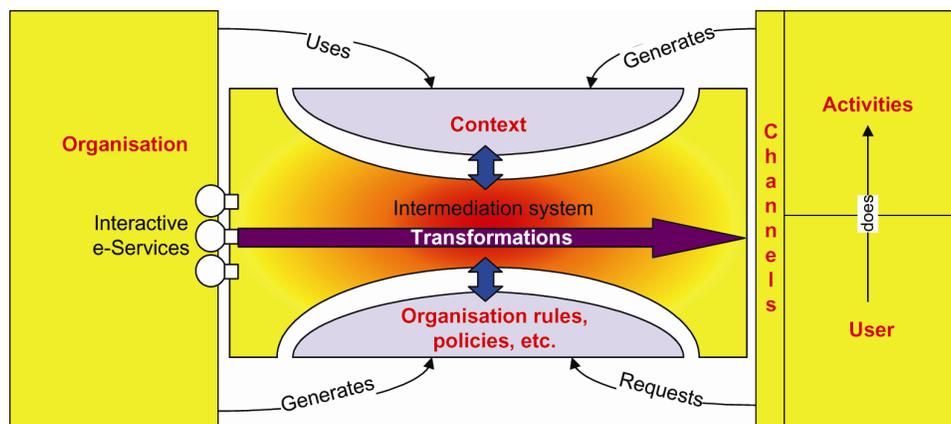
5.1 Integration of any channels for seamless interaction between organisations and learners

The possibility to use multiple channels when it is possible is really important in order to support mobile connectivity within learning scenarios. Moreover, this integration must take into account the issue of the rupture in the interaction, as for example, due to a network failure, etc.

For the moment, the only way of re-establishing the lost interaction depends on the learner's initiative of re-contacting the organisation. We sustain that this is a task that should be best done by the system itself, by detecting the rupture and calling automatically the learner, using the best adapted channel taking into account the contexts. Here, there can be incompatibility between the context of the user (his/her profile, etc.) and the context of the organisation (from the pedagogical point of view, with the organisation rules, policies, etc.).

Figure 4 shows the conceptual view of our software architecture and the various kinds of context that we try to support and manage.

Figure 4 Conceptual view of our software architecture (see online version for colours)



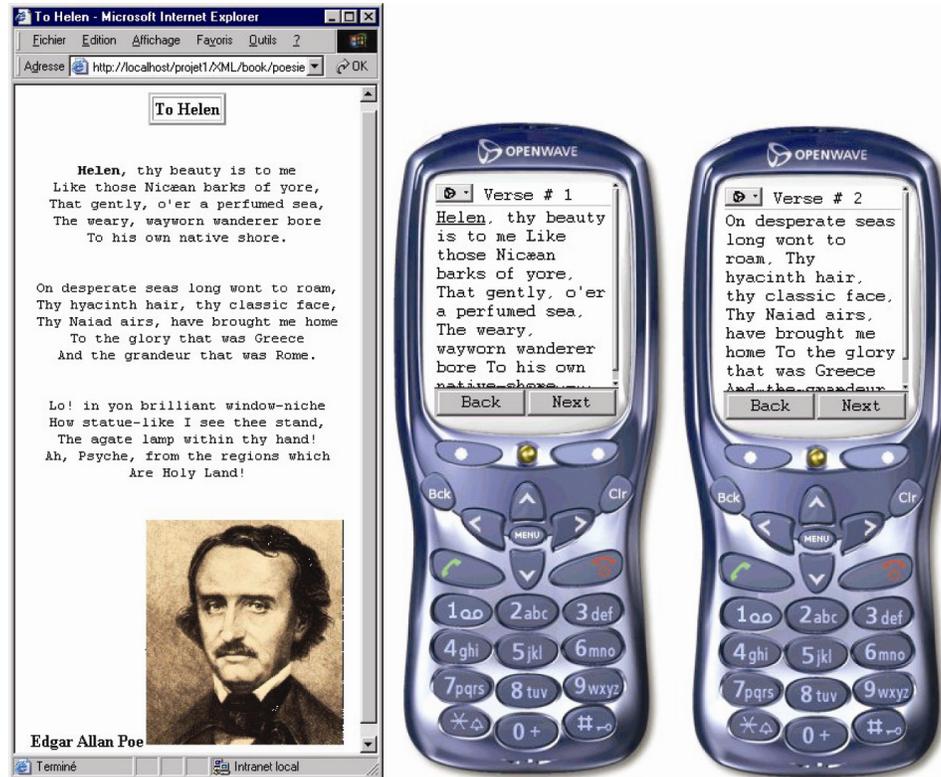
We can see in Figure 4 that we take account of two kinds of context:

- The 'policies' of the pedagogical organisation.
- The context provided (directly or not) by the user and used by the organisation in order to adapt the different services. In our prototype, we considered as a context, user preferences, channels in use, etc., but in a more ambitious perspective, we could have added geo-localisation, for example.

These two kinds of context are not always compatible. For example, if a learner has the habit of using services via her/his cellular phone, and the organisation does not deliver this service via this channel, the two contexts are not compatible. Therefore, our main interrogation is: how to manage these two contexts at the same time and in a coherent way?

In order to illustrate the matter of contexts incompatibility, we present in Figure 5 a poetry example. The ‘abstract pedagogical document’ (an XML file) is transformed into a suitable document, according to the detected device, network capabilities, etc. Thus, this poetry will be readable not only on a traditional web browser (with in addition a picture of the author, for example), but also on a cellular phone using the WAP protocol.

Figure 5 The same pedagogical document presented through different channels (Web Browser and Wap browser) (see online version for colours)



If the system cut the document technically (example 12 lines per screen) on the small phone screen, there is a risk to lose the pedagogical intention of the author that specially cut this poem with appropriate strophes. It will be exactly the same problem on an aural mode (VoiceXML language for example) if the Text-To-Speech (TTS) do not respect the punctuation and the pauses determined by the author.

5.2 The inter-dependence between service composition and channels

We will focus, now, more precisely on the nature of the dependence of service composition in relation to the channels used during the interaction.

In Chevrin, Derycke and Rouillard (2006), we have shown several scenarios where the composition of different e-services is conditioned by the nature of the channel and their interaction styles (direct manipulation vs. speech dialogue). Indeed from our previous evaluation of first prototypes (Derycke et al., 2003; Chevrin, Derycke and

Rouillard, 2005; Chevrin, Rouillard and Derycke, 2005), combining different interaction channels, we have discovered that the concurrent accesses to several e-services conducted to different interleaving of the e-services primitives, depending of the channel in use. For example, reading and choosing from a table of items using a voice-guided interaction is different from a web interaction. In the former case, we cannot force the user to remember all the made choices and on the latter case it is possible to see the whole table on the screen. Therefore, the fragmentation of the source document (e.g. a form) must be adapted to the channel characteristics and to the interaction elements (such as widgets, vocal forms, etc.).

Moreover, we agree in affirming that the learner should be able to choose the channel that she/he wants and to execute the service that she/he wants. Obviously, this is only possible if the service is available for the selected channel.

In the field of software engineering, with the Service-Oriented Architecture (SOA) and in the business field (Rust and Kannan, 2003), we can show that SOA is also a good framework to solve this issue. We introduced the Interactive e-Service (e-SI) concept in Chevrin, Rouillard and Derycke (2005). Therefore, our model-driven architecture builds, so as to say, a bridge between three domains:

- web-services (the e-SI side, we can see in Figure 4)
- learning rules (the policies, pedagogical rules, in Figure 4)
- human-computer interface, for context-aware aspect (see Figure 4).

6 Conclusions

In this article, we have presented the main problems that we have to face considering learning as cultural migration through mobile connectivity. Adaptation is needed not only between academically-based education and life-long learning, but also between local and distant learners (time, cultural differences, life way, etc.).

In order to minimise these matters and to adapt our lessons to different kinds of people, situations and contexts, we proposed to work on the modelling and deployment of flexible pedagogical scenarios. Thus it is possible, thanks to EMLs to describe an activity on a higher level and then to concretely transform it, at the runtime, thanks to different sort of treatment, depending on the technical possibilities available *in situ*. We achieve this by using our COW platform which plays the role of an intermediation platform between teachers and learners. We also take into account the user device, his/her student profile, and particular contextual data given or detected during the interaction. We predominantly tried to offer mobile communications possibilities to our students, using standards frameworks and languages (VoiceXML, WML, XHTML) in the deployment phase. Further work will lead us to improve connections between technical feasibility and pedagogical intentions. We are already able to provide, from a high-level document, multiple final documents, to be used on different channels (even in the same time and in a synergic way) and our next step is to insure that the pedagogical intention is not deteriorated within the different transformations on the learning platform.

Acknowledgements

The authors are thankful to the MIAOU and MOSAIQUES programmes (French *Nord Pas-de-Calais* Region) and the UE funds (FEDER) for providing support for this research.

References

- Ark, W. and Selker, T. (1999) 'A look at human interaction with pervasive computers', *IBM Systems Journal*, Vol. 38, pp.504–507.
- Barré, V., El-Kechaï, H. and Choquet, Ch. (2005) 'Re-engineering of collaborative e-learning systems: evaluation of system, collaboration and acquired knowledge qualities', *AIED'05 Workshop: Usage Analysis in Learning Systems*, Amsterdam.
- Bomsdorf, B. (2005) Adaptation of Learning Spaces: Supporting Ubiquitous Learning in Higher Distance Education, Mobile Computing and Ambient Intelligence: The Challenge of Multimedia, Dagstuhl Seminar Proceedings.
- Bourguin, G. and Derycke, A. (2005) 'Systèmes interactifs en co-evolution: réflexions sur les apports de la théorie de l'activité au support des pratiques collectives distribuées', *Revue d'Interaction Homme-Machine, AFIHM Europa*, Vol. 6, p.29.
- Chevrin, V., Derycke, A. and Rouillard, J. (2005) 'Some issues for the modelling of interactive e-services from the customer multi-channel interaction perspectives', *EEE 05 International. IEEE International Conference on E-Technology, E-Commerce and E-Service* (pp.256–259). Hong Kong: IEEE Press.
- Chevrin, V., Derycke, A. and Rouillard, J. (2006) 'Project UBI-learn: an intermediation infrastructure multi-channel access to future LMS', *Conference IEEE, AICT/ICIW 2006*. 19–24 February, La Guadeloupe, France.
- Chevrin, V., Rouillard, J. and Derycke, A. (2005) 'Multi-channel and multi-modal interactions in e-marketing: toward a generic architecture for integration and experimentation', *HCI International Conference*, Las Vegas, Lawrence Erlbaum editors.
- Derycke, A., Rouillard, J., Chevrin, V. and Bayart, Y. (2003) 'When marketing meets HCI: multi-channel customer relationships and multimodality in the personalization perspective', *HCI International 2003*. Heraklion, Crete, Greece, Vol. 2, pp.626–630, Lawrence Erlbaum Associates.
- Dewey, J. (1916) *Democracy and Education*. New York, NY: MacMillan.
- Dey, A., Abowd, G. and Salber, D. (2001) 'A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications', *Human-Computer Interaction*, Vol. 16, pp.97–166.
- Dillenbourg, P. (2002) 'Over-scripting CSCL: the risks of blending collaborative learning with instructional design', in P.A. Kirschner (Ed.), *Three Worlds of CSCL. Can We Support CSCL* (pp.61–91). Heerlen: Open Universiteit Nederland.
- Dourish, P. (2001) *Where the Action Is: The Foundations of Embodied Interaction*. Cambridge: MIT Press.
- Dourish, P. (2004) 'What we talk about when we talk about context', *Personal and Ubiquitous Computing*, Vol. 8, pp.19–30.
- Fischer, G. (2001) 'User modeling in human-computer interaction', *Journal of User Modelling and User-Adapted Interaction (UMUAI)*, Vol. 11, pp.65–86.
- IMS (2003) 'Global learning consortium', *IMS Learning Design Information Model*, version 1.0 Final Specification.
- Kiczales, G., des Rivières, J. and Bobrow D. (1991) *The Art of the Metaobject Protocol*. Cambridge, MA: The MIT Press, ISBN: 0-262-61074-4.

- Koper, R. (2001) 'Modeling units of study from a pedagogical perspective: the pedagogical meta-model behind EML', *Technical Report, Educational Technology Expertise Centre, Open University of the Netherlands*. Available at: <http://eml.ou.nl/introduction/docs/ped-metamodel.pdf>.
- Lin, J., Ho, C., Sadiq, W. and Orłowska, M.E. (2001) 'On workflow enabled e-learning services', in Paper presented in the Proceedings of the *IEEE International Conference on Advance Learning Technologies, ICALT2001*, 6–8 August, Madison, USA.
- MDA (2006) 'OMG model driven architecture'. Available at: <http://www.omg.org/mda> last visited June.
- Ogata, H., Yin, C. and Yano, Y. (2004) 'Context-aware support for learning japanese polite expressions', in Paper presented in the Proceedings of the *IEEE International Workshop on Wireless and Mobile Technologies in Education, WMTE'04*.
- OMG (2006) 'Object management group'. Available at: <http://www.omg.org> last visited June.
- Pernin, J-Ph. (2006) 'Normes et standards pour la conception, la production et l'exploitation des EIAH', chapter 9 in M. Grandbastien and J-M. Labat (Eds), *Environnements informatiques pour l'apprentissage humain*, Hermes, ISBN 2-7462-1171-8.
- Peter, Y. and Vantroys, T. (2005) 'Platform support for pedagogical scenarios', *Revue Educational Technology and Society*, Vol. 8, pp.122–137 (International Forum of Educational Technology and Society – ISSN 1176-3647).
- Rouillard, J. (2004) *VoiceXML. Le langage d'accès à Internet par téléphone*, aux éditions Vuibert, ISBN: 271174826X, p.197, Paris.
- Rust, R. and Kannan, P.K. (2003) 'E-service: a new paradigm for business in the electronic environment', *Communication of the ACM*, Vol. 46, pp.37–42.
- Thomas, S. (2005) 'Pervasive, persuasive e-learning: modeling the pervasive learning space', in Paper presented in the Proceedings of the *Third IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOMW'05)*, pp.332–336.
- Vantroys, T. (2003) 'Du langage métier au langage technique, une plate-forme flexible d'exécution de scénarios pédagogiques', *PhD Thesis*. France: University of Sciences and Technologies of Lille.
- VoiceXML (2006) 'Forum'. Available at: <http://www.voicexml.org> last visited June.
- WAP (2006) 'Forum'. Available at: <http://www.wapforum.org> last visited June.
- Weiser, M. (1991) 'The computer for the 21st century', *Scientific American*, Vol. 265, pp.94–104.
- Workflow Management Coalition (2002) 'Workflow process definition interface – xml process definition language', WfMC-TC-1025, version 1.0. Available at: http://www.wfmc.org/standards/docs/TC-1025_10_xpdl_102502.pdf.

Instrumentation and measurement of multi-channel services systems

Vincent Chevrin and José Rouillard*

Laboratoire LIFL – Université des Sciences
et Technologies de Lille – 59655 Villeneuve d’Ascq cedex, France
E-mail: Vincent.Chevrin@univ-lille1.fr
E-mail: Jose.Rouillard@univ-lille1.fr
*Corresponding author

Abstract: This paper focuses on the mediation between Interactive e-Services (IeS) and channels used during human-organisation interactions. This mediation consists in a succession of service adaptations, changing from an abstract structure to a concrete one. Our contribution is mainly devoted to aspects concerning the Quality of Service (QoS), particularly about instrumentation and measurement issues, and it provides clues to answer the question “How can multi-channel services be adequately monitored and controlled?”.

Keywords: multi-channel; context-sensitive; pervasive computing; Quality of Service; QoS; instrumentation; measurement.

Reference to this paper should be made as follows: Chevrin, V. and Rouillard, J. (2008) ‘Instrumentation and measurement of multi-channel services systems’, *Int. J. Internet and Enterprise Management*, Vol. 5, No. 4, pp.333–352.

Biographical notes: Vincent Chevrin has a Computer Science PhD from the University of Lille in France. His PhD has largely focused on multi-channel aspects in the context of communications and transactions between organisations and users in the e-Commerce area. He is currently an engineer of application and works in the field of marketing. He also continues to work on some challenges close to the multi-channel and pervasive applications issues.

José Rouillard is an Associate Professor in Computer Science at the University of Lille 1. He obtained his PhD in 2000 from the University of Grenoble (France) in the field of Human-Computer Interfaces. He is interested in HCI plasticity, multi-modality and multi-channel interfaces. He has written more than 60 papers and is now engaged in research on mobility and pervasive/ubiquitous computing.

1 Introduction

With the rapid development of telecommunication technologies, offering convergence (through IP protocols) and universal roaming for wireless networks, from GSM/UMTS to WiFi, and those of the mobile communicating devices (PDAs, Origami, etc.), new models of organisation and infrastructure for services and usages can be envisioned

and must be carefully designed and evaluated in a usage and application directed approach.

Our research focuses on the mediation between Interactive e-Services (IeS) and channels used during an interaction between a human being and an organisation. This mediation consists in a succession of service adaptations from an abstract structure toward a concrete one, deliverable to a user. These adaptations are closely linked to different contexts both coming from the user and the organisation context. The user context (which will be explained later in this paper) can be seen such as defined by Dey et al. (2001). The organisation context concerns the rules but also the strategies (marketing aspects) used by that organisation.

Our contribution is mainly devoted to aspects concerning the QoS. This means that we are investigating the different manners to deliver services to the user with suitable QoS. The question of instrumentation and measurement is an important issue for us. We consider it is valuable to understand how multi-channel services can be adequately monitored and controlled. That is why we decided to develop a tool able to catch several traces of the key elements during interaction sessions. Analysing those traces could help us to determine whether the QoS is adequate. Thus, we could intervene to modify the system adequately. Here, the notion of QoS is subjective, and another question that naturally ensues is “*How can we define and measure the quality of a service component provided through multimodal/multi-channel interface or via mobile devices?*”. In scientific literature, there is no rule (i.e., like those given by Nielsen and Tahir (2001) concerning the web pages characteristics) to decide if the quality of a service is sufficient or not. In our paper, we will try to answer both questions. We will begin with the second one, because the possible answers to this question will be useful to answer the first one.

This paper is structured as follows. First, Section 2 will introduce our multi-channel services approach. Then, in Section 3, we will deal with the quality of service in interactive e-services. Section 4 will be a presentation of a technical solution we have built aimed at managing the mediation between IeS and channels and that we called Ubi-Learn. This includes the presentation of experimental results concerning the QoS of customer telephone interactions in the context of a multi-channel service system. Finally, we close with the main conclusions and suggestions for future research.

2 Our multi-channel services approach

For some years now, we have been witnessing a strong growth of services provided through multiple channels. As we said before, this is linked with the proliferation of ‘virtual channels’, such as the internet, telephone, mobile devices, etc. In this section, we present our multi-channel services approach and we explain the possible connections between services and channels.

In Direct Marketing (DM) area, for instance, customers have been used to the multi-channel approach for a long time. Indeed, remote sales through phone or mail are an old practice. In other domains, some researchers employed the word ‘channel’ to talk about communication between humans, or between human and organisations, etc. For instance, the overall service provided by many remote sales organisations such as *3 Suisses International*, which includes several service components offered through virtual channels (e.g., the web, phone and mobile phones), in parallel with or as an alternative to physical facilities. According to several researchers, a channel of service

delivery can be defined as the means of communication through which a service is delivered to (or reaches) the customer. Virtual channels can be defined as the means of communication using advanced telecommunications, information, and multimedia technologies. Nowadays, the trend for multi-channel environments is growing (Sousa and Voss, 2006). This report gives arguments showing that new technologies afford an increasing number of virtual channels of service delivery.

We hold another vision of a channel. For us, an interaction channel with an organisation is the *union* of:

- a *means of routing* (of information/of documents, of tangible goods, of services, etc.) *toward a specific location* (fixed, mobile), *with specific features* (delay, temporality, quality of services—availability, faithfulness, costs, etc.)
- an *interaction device* with the user (customer, citizen, learner, etc.) with some «*modals*» *properties* (using of natural language or not, direct action, etc.), some *temporal properties*, and some «*affordance*» (semiotic and metaphoric aspect, guidelines, etc.).

The organisation could be a Direct Marketing company, an educational centre, etc. And this organisation would give the set of IeS available to the customers, employees, partners, learners, etc., i.e., all users who could be interested in those IeS. In addition, we note that this evolutionary definition allowed us to build a bridge between two communities: Direct Marketing (DM) and Human Computer Interaction (HCI). Thus, all channels, virtual or physical, can be mapped to this definition.

2.1 Toward a channel modelling

In Chevrin (2006), we presented our taxonomy of mediation between the channels used during a personalised interaction with a user and IeS. We consider mediation as the management of the link between the channels and the IeS during the interaction.

Figure 1 shows the four main notions we need to qualify mediation: network, user's device, person and interaction.

On the right side of Figure 1, we can see that a channel is composed of both a network and a user's device:

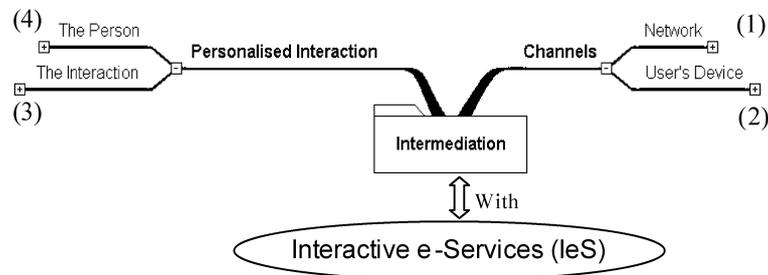
- Networks and access protocols have some norms and constraints. It is part (1) on Figure 1. Naturally, this part of the channel is subject to strong constraints of QoS, according to the user's task. Lots of works deal with this subject. Concerning that point, we were inspired from the Multi-channel Adaptive Information System (MAIS) (MAIS, 2006) project.
- In the user's device part (Interaction Device (ID), part (2) on Figure 1), the quality must be determined during the interaction, according to the contexts. This quality consists in determining if the ID's characteristics are sufficient to manage the IeS while keeping both the intention of the author and the ergonomic aspects in mind.

On the left side of Figure 1, we can see that a personalised interaction is composed of both a person and an interaction:

- The interaction mechanism knowledge: semiotic aspect, ergonomics, HCI design, conversational aspect with the grounding (Chevrin et al., 2006), adequacy to the tasks/medias, etc. (see part (3) on Figure 1). This part is used to detect intrinsic characteristics of a channel, thus we can find the link between the needs of the IeS and the author's intention.
- All the information related to the user (profile, etc.). The data linked to him or her will be provided, in part, by the Customer Relationship Management (CRM) systems (see part (4) on Figure 1). (Chevrin et al., 2005b). This can allow us to determine the user's need.

Those explanations helped us to propose the channel definition given in the previous section. The next section is dedicated to the IeS issues and the links with QoS.

Figure 1 The four main elements of our mediation taxonomy



2.2 Interactive e-Services (IeS) issues

Our needs concerning the description of IeS invocation and composition are slightly different from the main works about web services. This is due to our different needs of dynamicity. With standard web services (WSDL and UDDI), the goal is to discover (at run time) an IeS on the fly, for a particular need (rent a car, for example) with the possibility to delegate the service to a brokering service. But the workflow is static and the model is done at design time. In the case of the Ubi-Learn project (that will be presented in Section 4), we don't really need to discover new IeS on the fly, because the external IeS (delivered by a third party) are known at design time (Chevrin et al., 2005a).

New technologies and ways of using those technologies are affording new 'communication' opportunities in several areas. For example, in business, we can see the growing of e-Commerce then m-Commerce for mobile, then u-Commerce for ubiquitous, and soon p-Commerce for pervasive. Of course, there are different economic models and according to the economical model, the IeS proposed by the organisation will be different both from design and business process levels. The same report can be done in e-Learning (Chevrin et al., 2006). The MultiModal Interaction (MMI) W3C group has been working a lot on the different issues and interests of multimodal interactions. Those works show a strong link between the modality(ies) used and the user activity (task). Our IeS approach is, among other things, well-founded by our overall (multi-channel) vision of an interaction between a user and an organisation that offers several IeS (sets of tasks). Even if the web services technology can be used (not necessarily, though), for the realisation of the IeS, it is necessary to distinguish them from classic web services.

The goal of our IeS approach, with a complex mediation, is to allow the re-use of IeS in a great number of devices settings (channel or channel coupling) for different targets (people, audience (De Troyer, 2001)) and for different organisations (generics). Following the example of the W3C, we are convinced that there is a strong dependence between the medium (modality for the W3C and also the channel for us) used during the interaction and the organisation and the making-up of the IeS deliver to the user. Our theoretical framework (Chevrin, 2006) defines more precisely what a channel is and proposes some relevant properties that allow us to characterise communicational devices (channel or set of channels).

One recurrent problem that we must provide an answer for, is the management of the continuity of interaction. In HCI, it is called seamless interaction. This property is really important in DM. Indeed, a break down during an interaction between an organisation and a customer is often harmful for the organisation (loss of customer confidence, loss of sales, etc.) (Chevrin et al., 2005a, 2005b). With our approach (that will be presented in Section 4), this property of ‘no break down’ is really taken into account with three distinct points:

- the restoring of both the break down and the loss toward the customer
- the recovery of the interaction at the break down time (memory of all the actions done before the break down, such as the IeS finished, the data exchanged, etc.)
- the channel used for the interaction restoring is not necessarily the same as the one used before the break down (considering the different parameters).

Moreover, our approach must follow the three next directions:

- The IeS will have to be suited into the mediation layer because they need to be context-aware and adapted to the communication channels. Nevertheless, it is important that this adaptation be made from descriptions supplied by the IeS (auto-supplying, same as introspection). It would be also interesting if the IeS presented some properties that could allow us to alter, from outside, several elements modifying its behaviour from the customer point of view.
- It is also useful to take into account the potential of the organisation to support the self-service adaptable (My-Organisation) by the customer himself.
- The IeS must keep a minimal quality to be delivered to the user. In particular, the author intention must be respected, and the channel must be adapted to deliver it.

Those works about IeS are still in progress. One of our main goals is to place some QoS parameters inside the different IeS. Then, when an IeS will be called, a first test of QoS will be done from the IeS itself (such as the possible channels target).

In the next section, we show that IeS management (composition, adaptation, etc.) is closely linked to the channel used during the interaction.

2.3 The inter-dependence between IeS management and channels

We consider that the dependence between IeS (tasks/activities) and channel is a key parameter to take into account for the QoS management. The channel used will influence the IeS management, adaptation and presentation to the end user. In the same way, according to the IeS, the different channels available for the user will be used differently.

For example, if a video must be displayed, the system will suggest to the user, if possible, to switch for a while, from his or her cellular phone (without enough capabilities) to his or her web browser, in order to refer to this document better.

In Chevrin et al. (2006) we have shown several scenarios where the composition of different IeS is conditioned by the nature of the channel and their interaction styles (direct manipulation vs. speech dialogue). Indeed, from our previous evaluation of the first prototypes combining different interaction channels (Derycke et al., 2003; Chevrin et al., 2005a, 2005b), we have discovered that the concurrent accesses to several IeS conducted to different interleaving of the IeS primitives, depending on the channel in use. The fragmentation of the source document (for example, a form) must be adapted to the channel characteristics and to the interaction elements (such as widgets, vocal forms, etc.). Moreover, we agree that the learner should be able to choose the channel that she or he wants and to carry out the service that she or he wants. Obviously, this is only possible if the service is available for the selected channel.

3 QoS and Interactive e-Services

For a long time, there have been interactions between people and organisations using telecommunication networks. For example, traditional distance learning organisations use telecommunication networks to support distant learners activities: tutoring, counselling are mostly done by telephone in a one to one relationship, etc. Many examples can be given in other domains, like e-Commerce, e-Health, e-Governance, etc.

The evolution and the emergence of new technologies has led to new uses and obviously, new opportunities to interact with users. There are two relevant examples of those opportunities:

- Mobile commerce (m-Commerce), which appeared a few years ago with the emergence of small devices, such as cellular phones, PDAs, etc. Even if designers and programmers must take into account the technical characteristics of this kind of devices in order to adapt their content, m-Commerce goes beyond the web pages miniaturisation and the services offered are not necessarily the same as those suggested in e-Commerce.
- The same observation can be made about mobile Learning, which is not only an adaptation of e-Learning systems to access through wireless network, but increases the accessibility (from the learner's location) and the 'reachability' (to the learner's location). Several research projects (Attewell, 2005; McLean, 2003) have shown that in spite of the current limitation of the mobile devices put in the users' hands (i.e., a basic cellular phone), it is possible to consider new learning activities, less focused on rich document interactions, but more communication oriented, which can be compared to Computer Supported Collaborative Learning (CSCL) mode of education.

In addition, in the last years, a new concept has appeared: the Ubiquitous or Pervasive Computing also called Ambient Intelligence. The next two sections deal with that concept and are followed by the description of our vision for QoS management and the discussion of related works in QoS measurement and management.

3.1 Pervasive computing or ambient intelligence

The concept of Ambient Intelligence (AmI) or pervasive computing was developed by the ISTAG advisory group to the European Commission's DG Information Society and the Media. According to ISTAG (2001), Ambient Intelligence (AmI) is "*a stable yet evolving vision*". AmI is not the solution for social problems, it represents a new paradigm for how people can work and live together.

The concept of AmI is a vision in which humans are surrounded by computing and networking technology embedded in their surroundings. AmI puts the emphasis on user-friendliness, efficient and distributed services support, user empowerment, and support for human interactions. This vision assumes a shift away from PCs to a variety of devices which are unobtrusively embedded in our environment and which are accessed via intelligent interfaces. In order for AmI to become a reality, a number of key technologies are required:

- unobtrusive hardware (miniaturisation, nano-technology, smart devices, sensors, etc.)
- a seamless mobile/fixed web-based communication infrastructure (interoperability, wired and wireless networks, etc.)
- dynamic and massively distributed device networks
- natural feeling human interfaces (intelligent agents, multi-modal interfaces, models of context awareness, etc.)
- dependability and security (self-testing and self-repairing software, privacy ensuring technology, etc.)

The IST Advisory Group (ISTAG, 2001) has already given some relevant scenarios of pervasive computing in different areas. Here, we provide a simple scenario in e-Learning. This allows us to illustrate some QoS issues in a multi-channel situation.

"A teacher has made contents for a distance learning promotion course. This document is available on a learning platform, accessible by both students and teachers, and is composed by a video accompanied with some reports. Different channels (such as web browser, PDA, phone, etc.) are supported by the platform."

We consider QoS measurement to be a difficult task. In a multi-channel interaction, this task is even more difficult because of the lack of references or rules. In that scenario, we can show that the author intention can be important to determine the QoS of a service. For example, if the teacher considers the video to be a central point of the course, students will have to use a channel able to broadcast that video. Obviously, there are some other parameters, which could be used to measure the QoS. We will see some of them in the rest of the document. Here, with this simple scenario, we just wanted to show that the QoS measurement is a large and difficult issue.

3.2 Context-aware adaptation for pervasive computing

Studies about pervasive computing put the users' tasks at the central point (Banavar et al., 2000), and we think that it is still valid today.

In this section, we want to focus our argumentation on the context-aware computing issues. In fact, there is no consensus about the definition of context in the context-aware computing research field (Abowd and Mynatt, 2000). Several searchers have suggested definitions of context, such as Dey et al.

“any information that can be used to characterise the situation of entities (i.e., whether a person, place or object) that are considered relevant to the interaction between a user and an application, including the user and the application themselves. Context is typically the location, identity and state of people, groups and computational and physical objects,” (Dey et al., 2001)

or in Chalmers et al. (2004) and Brezillon (2003), where a distinction is made between relevant and irrelevant contexts:

“You can distinguish relevant context from that which is not at any step of a decision-making process or a performing task. Irrelevant context is called external knowledge. Relevant contextual knowledge depends on the agent, the current situation, and the decision at hand.”

Here, context and knowledge are strongly linked, and each interaction situation has its own relevant context.

Recent works, such as Henricksen (2003), aim to better formalise context. All those researches about Aml, pervasive, ubiquitous or context-aware deal with four questions:

- capturing the contextual variables
- normalising and exchanging them, perhaps through shared models based on ontology
- using dedicated software architectures to take into account the context evolving elements, like events broadcasted to the concerned applications
- better formalise the rules for decision about the adaptation of the services or documents to the context.

We remember from the previous works on context-aware computing that there is a real need for better models of contexts and models of adaptation of services which are context-sensitive.

3.3 *QoS of an IeS*

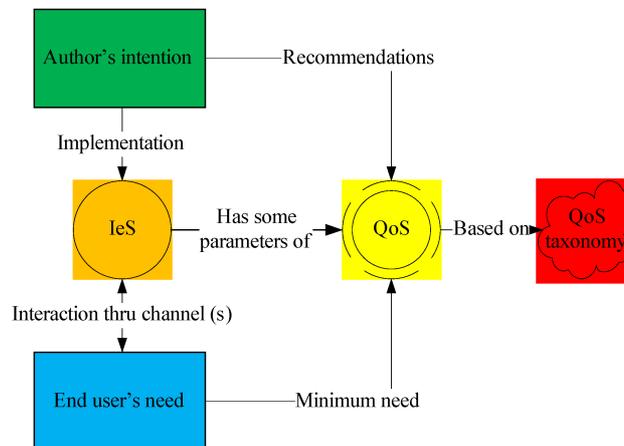
The QoS of an IeS can take place at different levels of interaction. We have found several levels in the life cycle of an IeS, which are relevant to control the quality of these IeS in our framework. Obviously, the following list is not exhaustive:

- The specification of the IeS must be abstracted in order to be mapped (adapted) to the different channels. For IeS delivered by an outsider, we cannot be sure that the specifications will be abstract enough; that is the reason why we are working on a proxy which filters the IeS, and adapts it in a compliant way. Here, one of the main difficulties is to keep the author’s intention. Nevertheless, the users’ needs must be taken into account. In our scenario, if the learner has previously seen the video and would just like to remember one part of it, he or she can use a device which has a poorer QoS, even if the author’s intention is not respected.

- Concerning the concrete IeS interface for the user, it is necessary to follow heuristics,¹ such as those proposed by Nielsen and Tahir (2001) for the web. Nevertheless, this kind of heuristics does not exist yet in the concerned channels, such as phones with audio interactions, PDAs, etc. That is why we added traces to our system (see Section 4.3) in order to find our own heuristics, or at least try to improve the interaction quality according to the context.
- The intention of the IeS author must be respected during the different levels of adaptation. If we consider the scenario previously discussed, a teacher makes an IeS composed by a video together with explanations: if a learner accesses this IeS through a small device, which is not able to play the video, the intention of the author is not respected, because for him or her, the video and documents cannot be separated, and a single picture, for example is not enough to replace that video. In addition, the user's intention should also be taken into account.

Figure 2 summarises our vision of the QoS management. We consider that the QoS could often be subjective information. Mapping QoS taxonomy (a lot of works deal with that subject, and our aim is not to propose another one) and both author's intentions and user's needs will be helpful to detect if an IeS can be proposed for a particular user or not. Thus, the context of the interaction will be fundamental to make such a choice. In fact, an IeS could be identified as 'available to the end user' according to the mapping between some QoS parameter and both the author's intention and the end user's need. An aspect of our works consists in trying to perform the general quality of the interaction (high level of the IeS), proposing to manage composition of the different IeS according to the context, and particularly according to the channels used during the interaction.

Figure 2 Our vision of the QoS management (see online version for colours)



3.4 Related works in QoS measurement and management

The QoS and related issues have been the subject of many researches and modelling efforts across several communities for a few years. In our case of study, it is relevant to get acquainted with the progresses of the web and web service community (e.g., Ran, 2003), the network and internet communities (e.g., Crawley et al., 1998) and also the

middleware community (e.g., Zinky et al., 1999). In spite of the differences between those communities, they have the same objective concerning quality. According to Marchetti et al. there are two major issues:

“(i) to identify the relevant measurable characteristics affecting the quality of the services provided by a given “object” (e.g., a web-service, a network infrastructure, a middleware platform) and (ii) to define means (e.g., architectures, paradigms, components, and protocols) to implement an ‘object’ whose values of its measurable characteristics satisfy some quantitative constraints.” (Marchetti et al., 2004)

Those authors name the measurable characteristics of objects QoS parameters (Frolung and Koistinen, 1998).²

For example, concerning the web-service community, works identify QoS parameters for better characterising services (Marchetti et al., 2004). Zeng et al. (2003) propose a methodology enabling the evaluation of the QoS of a composite service. In relation to the works of Ran (2003), they suggest to extend service discovery on the basis of QoS-related information. The main issue is the difficulty to obtain, define, and evaluate several of the QoS parameters, such as availability, performance, and so on. That is reinforced by the fact that we work in an end-to-end manner since the channels used to let a service and users interact are sometimes beyond the control of the provider. A solution will be to have services labelled by providers with sets of QoS parameters. However, this solution is limited because of the user’s perception to the QoS, such as provider, network or device availabilities, are not under provider control. van Moorsel (2001) proposes approaches to reduce those issues. Several contributions of MAIS (2006) discuss those works.

In a communication, Marchetti et al. present their own solution based on a ‘Quality Model’:

“The quality model consists of (i) a system model defining objects and actors, and of (ii) a set of roles and rules enabling the association of quality information to objects. Quality information is expressed using quality parameters and quality sets.” (Marchetti et al., 2004)

This system is based on the Service Oriented Architecture (SOA) with an extension to a channels representation. These works have in several points, the same orientation as we have.

In fact, the MAIS project (MAIS, 2006) is largely oriented by QoS issues. Some of the works developed in that project have been used as a starting point of our research. More details on the subject (relation between MAIS project and our works) are available in Chevrin (2006).

4 The Ubi-Learn project

In this section we will describe both the conceptual and practical solution we have built aimed at managing the mediation between IeS and channels and that we called *Ubi-Learn*. Then, we will present how we see the QoS management in our infrastructure.

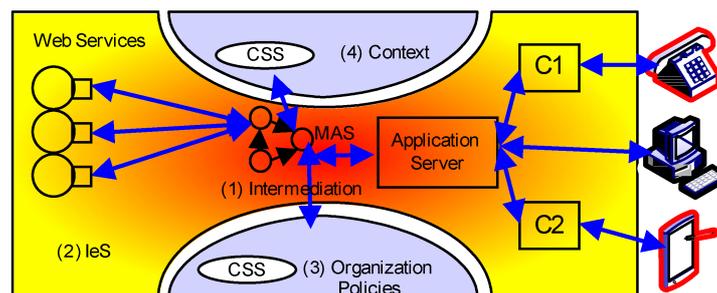
4.1 Overview

Our Ubi-Learn project aims to support more mobile and flexible learning processes by integrating recent progresses and development of software engineering, pervasive computing and context-aware computing domains of research. We have started the design and implementation of this new infrastructure for the intermediation between the learners, with fixed or mobile locations, accessing a system through various devices and value-added networks, and a collection of learning services, which can be extensible, depending on the context. The problem of the intermediation is complex, due to the number of settings that can be encountered, and to the needs of flexibility and dynamic adaptations. This implies that some decisions about the composition (of medium, channels, services, etc.) must be taken at run-time and not at design-time. Our technological infrastructure is designed with those constraints, and provides capabilities to be extended in order to support new communications channels or new user devices in the future.

4.1.1 Software architecture

Figure 3 shows the mediation between IeS delivered by the organisation and the user via the use of different channels (multi-channel interaction) synchronously or asynchronously. This entails issues, such as ruptures in an interaction when a user switches channels, etc. This mediation is context-sensitive. We considered the Context (4) as the context-aware defined by Dey et al. (2001). The other type of context is the organisation policies (3), which concern the different rules and policies of the organisation. Those two contexts (3) and (4) are achieved by a Collection of Context Services (CSS). Those services could be delivered by an outsider, as Liberty Alliance³ in (3). This context management is the key of the adaptation of the IeS delivered to the users. The management of such pervasive e-Services (Chevrin et al., 2005b) is based on the transformation chain that we have implemented with a Multi-Agents System (MAS) middleware presented in Figure 4 and described in more detail in Section 4.1.3.

Figure 3 Simplified view of our software architecture (Ubi-Learn) (see online version for colours)



In the area of pervasive e-Services (our IeS), we claim that a good management of the mediation entails two major issues. First, using two contexts (3) and (4) in Figure 3 enables both adapting and composing of the different IeS for a good presentation to the user. Secondly, from our previous evaluation of early prototypes (Chevrin et al., 2005b), combining different interaction channels, we discovered that the concurrent accesses to

several IeS led to different interleavings of the IeS primitives, depending on the channel in use. The user's entry point of Ubi-Learn is an Application Server where our Java application is run, which delivers various types of pages (displayed or spoken). One goal of Ubi-Learn was to build an application able to deal with multiple channels and we considered the following ones: telephone, Web browser and Personal Digital Assistant (PDA). To interact with our prototype through a telephone, a converter (named C1 in Figure 3) is needed to translate the VoiceXML pages generated into speech and also to interpret what the users say into data sent to the application. The same kind of converter is needed for PDA (named C2 in Figure 3), but Web browsers can access the Application Server as long as the latter is able to deliver native Web pages. If we go from that extreme to the other, there are the pervasive services, which require a front end to be implemented; that front end's action is to call Web Services and produce XML flows on the fly.

4.1.2 Collaboration between the main entities of Ubi-Learn

The entry of the system is a java servlet called DirectoryServlet which is linked to the MAS system through the Proxy Agent. This Agent creates and runs a majority of the Agents, except Display, Session and User Agents. An example of the link between the MAS and a servlet is given in Gandon (2006).

The number of Display Agents is the same as the number of users' requests and the number of User Agents is the same as the number of connected users. CRM Agent accesses to the different organisation rules. This agent knows rules such as *if user's birthday 20% discount or if student's mark lower than 10/20, then give him two extra exercises*. Those rules depend on the organisation business type. Rules Agent contacts the service to be run according to the services already ran and finished, and static or dynamic rules. The Display Agent is able to transform an XML flow in a Markup Language flow such as XHTML, VoiceXML, etc. understandable by the diverse User Devices. Finally, each service is represented by an agent. When the service is called for the first time, its agent sends a XML flow to the Interface Agent and warns the Session Agent whether the service is actually finished (this allows the break down into the interaction management; see Chevrin (2006) for details).

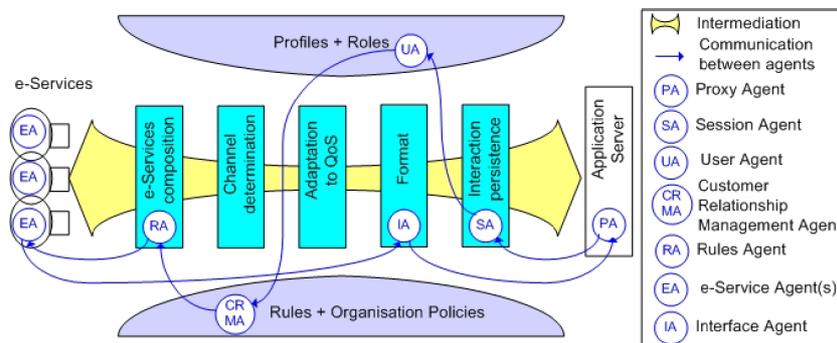
4.1.3 The mediation middleware

Figure 4 shows that a mediation uses several levels: the IeS Composition, the Channel prediction, the QoS, the format that we call the Quality of Interaction (QoI) by analogy with the QoS, and finally the persistence of data. Those different levels influence the mediation and obviously, the composition and the adaptation of the IeS, but also the capacity to choose the best channel for a particular IeS or composition of IeS. It is important to notice that, in Figure 4, all agents performing the actions are represented as a single agent whereas, to implement one agent of the figure, there could be a hierarchy of concrete Jade agents (Jade, 2006) (e.g., a factory, a manager, etc.). Now we describe more technically the different levels which take part in a mediation into our middleware.

The first level we describe is the composition of IeS (see Figure 4). It consists in determining the temporal composition (i.e., the orchestration) of the IeS depending on the context. The composition of IeS is performed by an appropriate agent in the MAS, called the RA (see Figure 4). When a user accesses Ubi-Learn through a telephone channel,

the RA will decide not to access the e-Service which enables to skim through the catalogue, although this e-Service is accessible through a Web channel. In our prototype, the orchestration rules are static and thus difficult to enrich. We are still working on another evolution of the prototype including a flexible workflow manager based on previous works (Hadjouni et al., 2005). Moreover, several relevant researches are carried out in this area. For example, Blake's works (Blake, 2004) on a workflow based on software agents for business orchestration. Other works as Maamar et al. (2005) study web services orchestration with software agents.

Figure 4 Conceptual and technical view of our middleware (see online version for colours)



The second level, the 'Channel determination', means that the context influences the channel used for the customer-organisation interaction. For example, simple rules of an organisation could force the use of a fax, to validate the customer order (e.g., in a travel agency). To manage that, the UA and the CRMA (see Figure 4) are used to request the different contexts to predict the channels available to carry out an e-Service. Both marketing and communication rules (coming from organisation policies) which predict the right channels could be more complex and are based on our theoretical framework on the communicational features of the channels (we also achieved a taxonomy formalising a mediation, explained in Chevrin et al. (2005a)). For the time being, this feature is achieved by UA and the CRMA, but we intend to design dedicated agents for it. The use of two contexts (respectively used by the UA and the CRMA in Figure 4) can lead to incompatibilities between them. So, a major question is: how to bring a 'cross-fertilisation' between context-aware and policies of the organisation? This complex and open problem is not discussed in this paper.

The third level, 'Adaptation to QoS' is closely linked to the 'Channel determination' level. For example, a customer using a 3G mobile phone will be able to receive high-quality videos whereas a GSM telephone user will only get a low-quality or no video at all. In this area, there are some researches with some relations to our approach. The MAIS project (MAIS, 2006) addresses a large effort on the QoS. This project is also particularly oriented to networks and studies adaptivity on each layer in the information systems, from the application to the networks and devices. Several layers of adaptivity are considered and particularly the needs of multi-channel information systems and the systems for the disabled users are addressed.

The fourth level, 'Format' indicates the data presentation to the user according to the channel used. In other words, Format is the manner in which the data from the IeS

are formatted to be displayed or spoken to the customer. Thus, the IeS are adaptable to the channels used. Indeed, every e-Service is represented by an EA (see Figure 4), which produces an XML flow and an appropriate agent called the IA aims to transcribe it into the relevant language (XHTML for a Web browser, WML for a PDA and VoiceXML for a telephone). To perform that, the agent uses XSL transformations. Some researches are achieved in this area, such as the recent project AMACONT (Hinz and Zoltan, 2004), and also in the Web Application area, relatively close to our approach from the IeS adaptation point of view. In this project, the authors suggest a step by step electronic document generation via a pipeline such as the Cocoon framework (<http://cocoon.apache.org/fr>) in the Apache suite. Nevertheless, their notion of device does not take into account the nature of the different networks. Another recent project, Thales (<http://www.thalesgroup.com/home/home/>), offers this kind of adaptation with the Human-System Interaction Container (HIC) (Lard et al., 2004).

The fifth level of the prototype is the 'Interaction persistence' (see Figure 4). This is a kind of layer which manages interaction ruptures; two kinds of ruptures are to be considered: the unintended ones (e.g. due to a communication breakdown) and the intended ones (e.g., a user naturally finishes using the application, but not all the interaction). This level is achieved by the Session Agent SA (see Figure 4) and implements an 'enhanced' session management which integrates multiple communications through different channels.

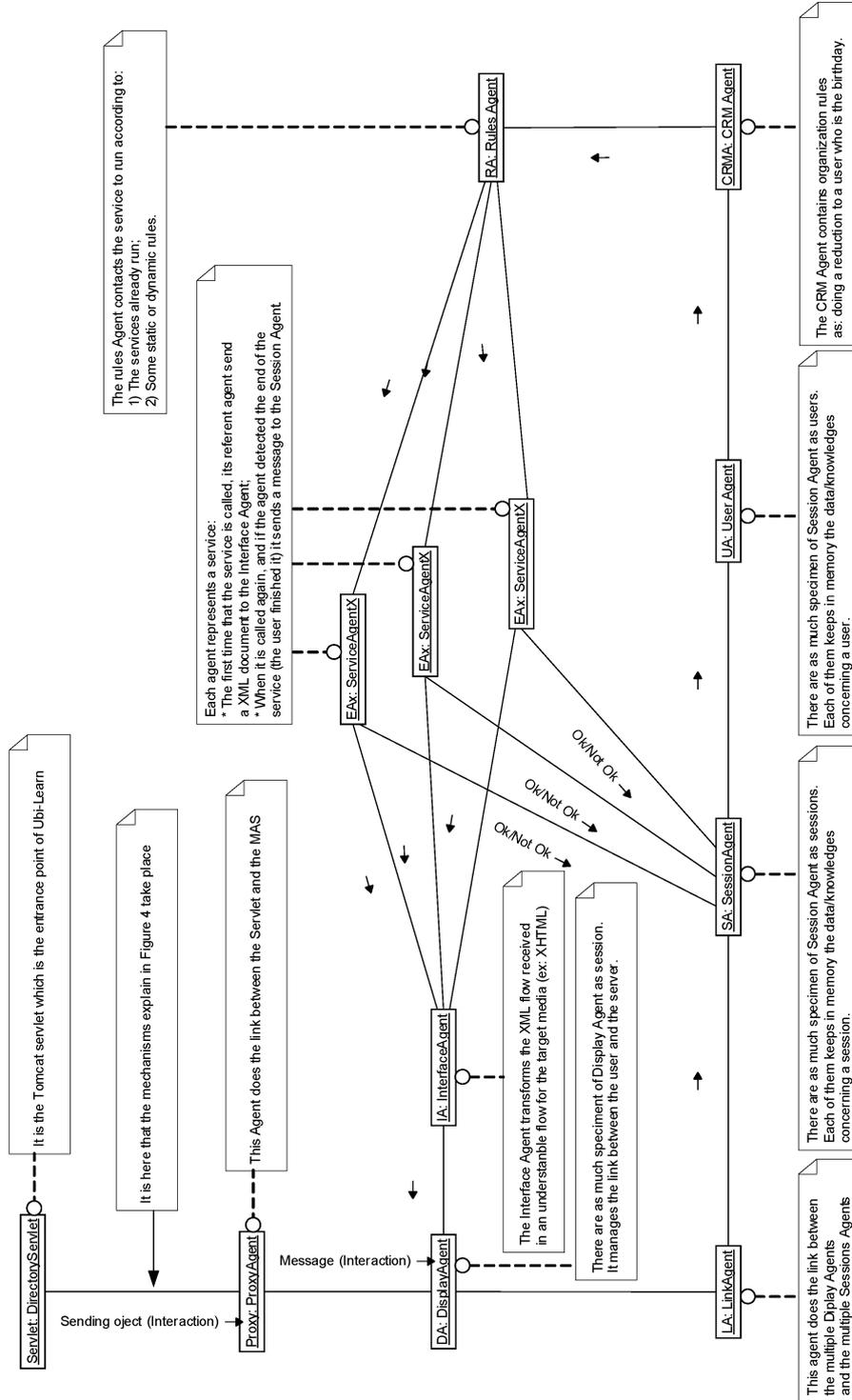
For example, if a customer is using Ubi-Learn through a telephone and a rupture occurs during the interaction, he would be able to re-access Ubi-Learn from his or her computer via a Web channel and then be re-identified; and he or she would recover everything he or she was doing.

Figure 5 shows the UML Collaboration diagram of Ubi-Learn, and the links between several major Agents of the MAS system. Some comments have been added on Figure 5, in order to make its reading easier. To summarise the global running of the middleware let us imagine that a customer accesses the Application Server through several channels. This server sends the user request to the *PA*. This agent creates a *SA* specific both to the user and the channel and it is responsible for the persistence of each mediation. Then, the request passes through *UA* (manages the user preferences, roles, etc.) and *CRMA* (manages the organisation policies). Those agents take the contexts into account. Afterwards, the *RA* achieves a particular composition of the IeS according to the channels in use. The *RA* sends a request to the appropriate *EA*, and the latter sends an XML flow (representing an abstract representation of the e-SI) to the *IA*, which sends the Markup Language flow (XHTML, WML or VoiceXML) to the *PA*. Then the data is sent back to the Application Server. And the latter transmits the Markup Language flow to the right channel. In the next section, we will look further into the part of the QoS management.

4.2 *Management and measurement*

In this section we will present what we mean by QoS and how we propose to manage it through our infrastructure. There are several distinct kinds of QoS. They can be related to the networks, to the cost (money, time, etc.), ergonomics, etc. Whatever they are, we have to manage them for multi-channel, context-sensitive interactions.

Figure 5 UML collaboration diagram of Ubi-Learn



In this way, we consider that there are several distinct ways to manage QoS. Here we present three of them. First, as already stated, the IeS itself must contain some parameters of QoS (such as the channels where it is possible to carry it out, the author's intention, etc.). Then, an adaptation level could be spread out in the mediation middleware. In that case, there are two possibilities:

- We create a special adaptation level dedicated to the QoS management. It's the last step of the IeS adaptation. If it is possible, this level tries to adapt the IeS and the channels used else the system gives back to the user a message that explains why it is not possible to access this IeS.
- Each adaptation level has its own QoS measurement system.

Finally, we must be able to modify the system to improve the QoS of the IeS if it is possible, or give some recommendations to the user. To achieve that, we must set up a system able to register the different interactions between the user and the system. This idea to save some traces raises several problems. The most serious of them is to decide if instruments should be inside or outside the system. We have already done several experiments on this part, and we present one of them in the next section.

4.3 *Experimental results*

A part of the Ubi-Learn project is dedicated to the instrumentation and the measurement of interactions performed in multi-channel services systems. One of those channels is the telephone. Experimental results presented in this section are mainly dedicated to that channel. Text To Speech (TTS) synthesis and Automatic Speech Recognition (ASR) used in this project are considered as services. Measuring the quality of this kind of service is not trivial. If it's interesting to propose a natural dialogue between users and machines, based on ASR and TTS, it is much more interesting to have tools that check the effectiveness and the usability of the used equipments and interfaces.

Concerning that last point, the observations of all scientific communities remain the same: there are very few corpus and tools provided to observe the use of multi-channel systems. It is necessary to find new ways to model, conceive and carry out information processing systems able to provide logs (at a low level but also for a higher degree of complexity) situated in time and space. It will also be very relevant to capture information about the use contexts (state of the networks, level of the user's stress, ambient noise level, light brightness, etc.), in order to be able to reproduce those situations, later, if one wants to understand what occurred at runtime.

The process we implemented in order to record conversations between the users and our vocal server was external to the system itself. It recorded conversations but also surrounding noises. This first step showed that what the system believed to hear was not necessarily what the users uttered. Typically, with the question 'How old are you?' it happened that the machine recorded '33' in the database, whereas the true answer pronounced by the user was '23'. The human tutor, in quiet listening can note this anomaly down and listen to the recorded sentence again to make up his or her opinion.

The audio corpus represents 4888 seconds, that is to say 1 hour and 35 minutes of speech. On average, each conversation lasts 213 seconds (standard deviation of 90). The shortest dialogue lasts 132 seconds, and the longest lasts 574 seconds. Those durations were automatically calculated by the machine which noted down the

beginning and the end of each conversation. After the results obtained during this first experiment, we can say that the VoiceXML language (VoiceXML) did not have, in an internal way, a mechanism which gave the possibility to follow the interactions with the users connected on the Interactive Voice Response (IVR). Until version 2.0 of VoiceXML (Rouillard, 2004), it was not possible to carry out a real trace of what the users had actually pronounced during their dialogue with the system. But it was possible to consult the variable supposed to recover the value corresponding to vocal grammar elements. In other words, it was not possible, up to now, to record automatically what the user really pronounced at runtime.

Since the relatively recent version of VoiceXML 2.1, which is yet only a 'working draft' of the W3C, it is technically possible to record what the user pronounces during the vocal interaction. For that, it is necessary to initialise the *recordutterance* attribute of the *<property>* tag with the value 'true'. One can then obtain the following data thanks to certain application variables (see Table 1):

Table 1 Variables of the application, available in VoiceXML 2.1

application.lastresult\$.confidence	Confidence level (between 0.0 and 1.0). A value of '0.0' indicates a low confidence and '1.0' a high confidence
application.lastresult\$.inputmode	Input mode used ('dtmf' ⁴ or 'voice')
application.lastresult\$.recording	Vocal data corresponding to that the user said
application.lastresult\$.recordingduration	Duration (in msec) of the last voice recognition
application.lastresult\$.recordingsize	Size (in bytes) of the last voice recognition
application.lastresult\$.utterance	The utterance recognised for this interpretation

We used this specificity (available in our laboratory IVR-vocal server) to develop a recording mechanism of the possible man-machine interactions in VoiceXML (voice and telephone keyboard). Thanks to this instrumentation, it is possible to collect traces which reflect the use of an application. During each interaction, the related piece of information is recorded in a database. We have developed tools allowing experimenters (not necessarily data processing specialists) to analyse the collected data. From this graphical user interface, an 'expert' of the domain (a pedagogical engineer, for instance) can listen to the vocal recording (.wav) corresponding to the user's answer and compare it to what the machine believed it had understood during the interaction.

We tested this process with a multimodal e-Commerce application using speech recognition and voice synthesis, providing the visualisation of the products pictures, and also allowing the use of telephone keyboard DTMF (Dual-Tone Multi-Frequency), and the possibility to click on web page hyperlinks. For example, on the first row of Figure 6, we can see that when the task was about choosing a cloth size, the machine believed to recognise the answer 'large' from the user voice interaction.

By clicking on the file (.wav) of that very row, the expert can check that the user has really pronounced this word or sentence, and can correct the information if necessary. The good/bad understanding statistics (of the machine) are updated during each interaction. This allows to measure and check the interface quality of service.

Figure 6 Recorded traces (VoiceXML 2.1) (see online version for colours)

Time	Task	Machine understanding	Record file name
16h 40m 19s	choose size	large	user_Tue16May_16h_40min_19s_voice.wav
16h 40m 34s	choose color	blue	user_Tue16May_16h_40min_34s_voice.wav
16h 40m 53s	choose product reference	25	user_Tue16May_16h_40min_53s_voice.wav
16h 41m 05s	add to cart	no	user_Tue16May_16h_41min_05s_voice.wav
16h 44m 12s	give client reference	228136	user_Tue16May_16h_44min_12s_voice.wav
16h 44m 30s	choose product set	blue dress	user_Tue16May_16h_44min_30s_voice.wav
16h 44m 40s	choose shelf	woman section	user_Tue16May_16h_44min_40s_voice.wav

5 Conclusion

In this paper, we have presented our approach concerning the measurement of the quality of services in systems supporting multi-channel services. We showed that the interdependence between IeS management and channels is very important in order to provide relevant solutions according to contexts, tasks, user profiles, network traffic, etc.

With the e-Learning example developed in this paper, it's easy to understand that the quality of the service is not just a network/infrastructure quality issue. Indeed, in multi-channel architectures, we have to face the possibility that the same information can be given across different and heterogeneous networks and devices.

We have worked for many years on that issue concerning the quality of service and proposed a mediation middleware. This framework, based on a multi-agent system, takes into account, at the same time, rules and organisation policies, profiles and user's roles together with IeS capabilities.

Some interesting results about the instrumentation and the measurement of multi-channels systems let us realise that if we want to fathom the quality of a service, we have to reproduce the context of the interaction (with the same elements as those captured during the interaction). We particularly worked on the possibility to record telephone conversations while a machine and a human are talking to each other. In these circumstances, we provided tools for users that make it easier to assess whether the service was valuable. We believe that, in pervasive computing situations, particularly, it will provide robust and rich information that could be used to improve the quality of multi-channel services.

Now, we want to improve our system of trace capture. Currently, we catch traces with suitable systems and the data obtained are not organised according to a dedicated formalisation. So, first, we want to make a generic system allowing to catch different useful traces (vocal, textual, etc.). This system will help us to determine if the quality of the interaction proposed to the end user is convenient or not. And secondly, we work on a specification of traces in order to map the information logging on a formal model of traces. This system of trace-logging will also allow us to carry out a study on the use of our system (using of mobile device, etc.).

Acknowledgements

The authors are thankful to the MIAOU and EUCUE programs (French *Nord Pas-de-Calais* Region), the ANR P-LearnNet project and the UE funds (FEDER) for providing support for this research. Special thanks to Rui Sousa, Cesar Olavo and Mariette Bessac for their reviews and comments.

References

- Abowd, G. and Mynatt, E. (2000) 'Charting past, present, and future research in ubiquitous computing', *ACM Transactions on Computer-Human Interaction*, Vol. 7, No. 1, March, pp.28–59.
- Attewell, J. (2005) *Mobile Technologies and Learning: A Technology Update and M-Learning Project Summary*, Document from the Learning and Skills Development Agency, UK.
- Banavar, G., Beck, J., Gluzberg, E., Munson, J., Sussman, J. and Zukowski, D. (2000) 'Challenges: an application model for pervasive computing', *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking*, ACM Press, Boston, Massachusetts, United States, pp.266–274, ISBN:1-58113-197-6.
- Blake, M.B. (2004) 'Forming agents for business process orchestration', *Proceedings of the 37th Annual Hawaii International Conference on System Sciences (Track on Workflow and Web Services)*. ISBN: 0-7695-2056-1.
- Brezillon, P. (2003) 'Focusing on context in human-centred computing', *IEEE Intelligent Systems*, May–June, pp.62–66.
- Chalmers, D., Dulay, N. and Sloman, M. (2004) 'A framework for contextual mediation in mobile and ubiquitous computing applied to the context-aware adaptation of maps', *Personal and Ubiquitous Computing*, Springer-Verlag Publisher, Vol. 8, pp.1–18.
- Chevrin, V. (2006) *L'Interaction Usagers/Services, multimodale et multicanale: une première proposition appliquée au domaine du e-Commerce*, PhD in Computer Science, USTL, France.
- Chevrin, V., Derycke, A. and Rouillard, J. (2005a) 'Some issues for the modelling of interactive e-services from the customer multi-channel interaction perspectives', *International conference on e-Technology, e-Commerce and e-Service*, IEEE Press, Hong Kong, pp.256–259.
- Chevrin, V., Derycke, A. and Rouillard, J. (2006) 'Project UBI-Learn: an intermediation infrastructure multi-channel access to future LMS', *Conference IEEE, AICT/ICIW 2006*, 19–24 February, La Guadeloupe, France.
- Chevrin, V., Rouillard, J. and Derycke, A. (2005b) 'Multi-channel and multi-modal interactions in e-marketing: toward a generic architecture for integration and experimentation', *HCI International Conference*, Lawrence Erlbaum, Las Vegas, p.10.
- Crawley, E., Nair, R., Rajagopalan, B. and Sandick, H. (1998) 'A framework for QoS-based routing in the internet', *Internet Engineering Task Force–RFC 2386*, August.
- De Troyer, O. (2001) 'Audience-driven web design', in Rossi, M. and Siau, K. (Eds.): *Information Modelling in the New Millennium*, IDEA Group Publishing, ISBN 1-878289-77-2.
- Derycke, A., Rouillard, J., Chevrin, V. and Bayart, Y. (2003) 'When marketing meets HCI: multi-channel customer relationships and multimodality in the personalization perspective', *HCI International Conference*, Lawrence Erlbaum Associates, Heraklion, Crete, Greece, Vol. 2, pp.626–630.
- Dey, A., Salber, D. and Abowd, G. (2001) 'A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications', *Human-Computer Interaction*, Vol. 16 Nos. 2–4, pp.97–166.
- Frolung, S. and Koistinen, J. (1998) *QML: A Language for Quality of Service Specification*, Technical Report HPL98-10, HP Labs, HP Software Technologies Laboratory, February.
- Gandon, F. (2006) *INRIA – Sophia Antipolis*, <http://JADE.tilab.com/>
- Hadjouni, M., Tirellil, I. and Laroussi, M. (2005) 'Collaborative platform for mobile learning based on e-Services', in Isaias, P., Borg, C., Kommers P. and Bonamo, P. (Eds.): *Conference IADIS International 'Mobile Learning 2005'*, Qawra, Malta, 28–30 June, pp.226–230.
- Henricksen, K. (2003) *A Framework for Context-Aware Pervasive Computing Applications*, PhD in Computer Science, University of Queensland, Australia, September, p.201.

- Hinz, M. and Zoltan, F. (2004) 'AMACONT: a system architecture for adaptive multimedia web applications', *Workshop 'XML Technologien für das Semantic Web' (XSW 2004)*, Berliner XML Tage, Berlin, 13–15 October, pp.65–74.
- ISTAG (2001) *Scenarios for Ambient Intelligence in 2010*, Final Report, Ducatel *et al.* (Eds.): European Commission, IST Program, Brussels, February, 54 pages.
- Jade (2006) <http://jade.tilab.com/>
- Lard, J., Sedogbo, C. and Bisson, P. (2004) 'Advanced in software architecture design applied to human computer interaction', *Processing. SIM'04, Semantic Intelligent Middleware for the Web and the Grid Proceeding of ECAI – Workshop*, Valencia – Spain, 23–27 August.
- Maamar, Z., Mostéfaoui, S.K. and Yahyaoui, H. (2005) 'Toward an agent-based and context-oriented approach for web services composition-appendices', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 5.
- MAIS (2006) Project, <http://www.mais-project.it/index.php>
- Marchetti, C., Pernici, B. and Plebani, P. (2004) 'A quality model for multichannel adaptive information', *International World Wide Web Conference. Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers and Posters*, New York, NY, USA, pp.48–54.
- McLean, N. (2003) *The M-Learning Paradigm: An Overview*, A Report for the Royal Academy of Engineering and the Vodaphone Group Foundation, Australia, November.
- Nielsen, J. and Tahir, M. (2001) *Homepage Usability: 50 Websites Deconstructed*, New Riders Publishing, Indianapolis, ISBN 0-73571-102-X.
- Ran, S. (2003) *A Model for Web Services Discovery with QoS*, ACM, SIGecom Exchanges, Spring.
- Rouillard, J. (2004) *VoiceXML. Le langage d'accès à Internet par téléphone, aux éditions Vuibert*, ISBN: 271174826X, 197 pages, Paris.
- Sousa R. and Voss, C.A. (2006) 'Service quality in multichannel services employing virtual channels', *Journal of Service Research*, Vol. 8, pp.356–371.
- van Moorsel, A. (2001) *Metrics for the Internet Age: Quality of Experience and Quality of Business*, Technical Report, HP Labs, July, appears also in the *Proceedings of the 5th Performability Workshop*, Erlagen, Germany, 16 September.
- Zeng, L., Benattallah, B., Dumas, M., Kalagnanam, J. and Sheng, Q.Z. (2003) 'Quality driven web services composition', *Proceedings of the Twelfth International Conference on World Wide Web*, ACM Press, pp.411–421.
- Zinky, J.A., Bakken, D.E. and Schantz, R.E. (1997) 'Architectural support for quality of service for CORBA objects', *Theory and Practice of Object Systems*, Vol. 3, No. 1, pp.1–20.

Notes

¹Here, we mean both usability and ergonomic rules.

²Other names have been given, e.g. dimension, attributes, etc. (Frolung and Koistinen, 1998).

³<http://www.projectliberty.org/>

⁴Dual-tone multi-frequency.

Website

VoiceXML: VoiceXML Forum <http://www.voicexml.org>