
Conception d'une interface multimodale à partir des comportements utilisateurs

Utilisation du Magicien d'Oz en conception

Stéphanie Buisine* — **Jean-Claude Martin****

* *Laboratoire Conception de Produits et Innovation, ENSAM
151 boulevard de l'Hôpital, F-75013 Paris
stephanie.buisine@paris.ensam.fr*

** *LIMSI-CNRS, BP 133, F-91403 Orsay Cedex
et LINC-IUT de Montreuil, 140 rue de la Nouvelle France, F-93100 Montreuil
jean-claude.martin@limsi.fr*

RÉSUMÉ. Nous proposons une étude de cas de conception d'un système interactif où le comportement des utilisateurs a servi de base au développement informatique. Grâce à une simulation en Magicien d'Oz, nous avons pu modéliser le comportement des utilisateurs avant même d'avoir conçu le système, et nous en avons déduit les spécifications pour l'implémentation. Cette approche va au-delà des critères de la Conception Centrée Utilisateur (norme ISO 13407), puisque l'utilisateur, en plus d'être consulté et observé, nous a également fourni les briques de base du futur système. Nous pensons que cette méthode de conception pourrait être adaptée à d'autres champs d'application, et notamment en conception de produit.

ABSTRACT. We present a design case study of an interactive game system in which users' behavior served as a basis for the implementation. By using a Wizard-of-Oz simulation, we modeled users' behavior although the system was not developed yet, and we could extract the specifications for the development. Such an approach goes beyond the criteria of User-Centered Design (ISO standard 13407) because the user was not only interviewed and observed, but s/he also provided us with elementary specifications of the system. This design process is likely to be useful in other application fields, for example in product design.

MOTS-CLÉS: Interaction Homme-Machine, Magicien d'Oz, Interface Multimodale, Ergonomie, Conception Centrée Utilisateur.

KEYWORDS: Human-Computer Interaction, Wizard-of-Oz, Multimodal Interface, Ergonomics, User-Centered Design.

1. Introduction

L'objectif de cet article est de décrire, sous la forme d'une étude de cas, le processus que nous avons suivi pour la conception d'un système interactif. L'application présentée est un jeu éducatif sur ordinateur dans lequel les utilisateurs interagissent avec des personnages virtuels. En entrée du système, l'interface est dite multimodale, c'est-à-dire qu'elle permet aux joueurs d'utiliser plusieurs modalités de communication : dans notre cas, ces modalités sont la parole et le geste en deux dimensions à la surface de l'écran (écran tactile ou tablette tactile avec stylo). Les utilisateurs peuvent ainsi interagir par l'une ou l'autre de ces modalités (message verbal ou gestuel), mais ils ont également la possibilité de combiner les modalités au sein de messages que nous appelons « constructions multimodales » (par exemple la phrase « Mets ça ici » accompagnée de la désignation d'un objet et d'un emplacement).

Une des problématiques majeures dans le développement d'interfaces multimodales est la fusion des modalités. En effet, le traitement et l'interprétation des constructions multimodales sont un véritable défi pour les systèmes multimodaux. Deux questions principales émergent à ce propos :

– Dans les constructions multimodales, comment les modalités sont-elles intégrées au niveau temporel ? Il faut en effet que le système soit capable de résoudre les références verbales (ex : « ça » et « ici ») avec les éléments gestuels pertinents (la sélection de l'objet, et la localisation spatiale, respectivement).

– Comment les modalités coopèrent-elles au niveau sémantique ? Par exemple, le système doit pouvoir détecter et traiter correctement les constructions redondantes (ex : « déplacer le cube bleu sur la droite » accompagné d'un déplacement du cube bleu par le geste) afin d'éviter de réaliser deux fois la même commande.

Dans les paragraphes suivants, nous allons exposer les méthodes disponibles pour répondre à cette problématique. Puis nous présenterons notre propre démarche, qui repose sur une étude du comportement spontané des utilisateurs. Nous présenterons également la technique qui a permis, au premier chef, de réaliser cette étude : la simulation en Magicien d'Oz. Nous discuterons enfin de l'intérêt de cette démarche de manière plus générale, en conception de produit.

2. Méthodes de conception des interfaces multimodale

2.1. Les interfaces multimodales

La multimodalité en Interaction Homme-Machine caractérise la capacité d'un système à exploiter plusieurs modalités de communication en entrée (de l'utilisateur vers le système) et/ou en sortie (du système vers l'utilisateur). Dans la présente étude, nous traiterons exclusivement de la multimodalité en entrée, c'est-à-dire de

celle qui concerne les informations transmises par l'utilisateur à destination du système.

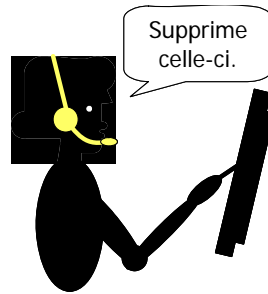


Figure 1. Exemple de style d'interaction avec une interface multimodale (parole et geste à la surface de l'écran tactile). Le système est capable de traiter les commandes verbales seules, les commandes gestuelles seules, et les constructions multimodales comme dans l'exemple ci-dessus.

Les Interfaces Multimodales en entrée permettent donc d'utiliser, simultanément ou pas, de manière combinée ou pas, plusieurs modalités de communication comme la parole et le geste. Ces interfaces visent notamment à fournir une alternative aux systèmes WIMP (Windows – Icons – Menus – Pointing) que nous utilisons tous les jours par l'intermédiaire du clavier et de la souris. Le champ de recherche sur les Interfaces Multimodales repose sur l'hypothèse que celles-ci seraient plus proches de la communication Homme-Homme, par conséquent plus faciles à utiliser et à apprendre, et seraient préférées par les utilisateurs. Les Interfaces Multimodales pourraient également permettre le développement de systèmes interactifs plus sophistiqués, l'élargissement de leur utilisation à une frange plus large de la population, et l'adaptation à des conditions d'utilisation plus variées par rapport aux interfaces classiques (Oviatt *et al.*, 2000).

2.2. Méthodes de conception

2.2.1. Les taxonomies d'intégration des modalités

Comme nous l'avons souligné en introduction, les constructions multimodales sont des messages complexes à traiter par le système, car celui-ci doit interpréter correctement les relations temporelles et sémantiques entre les modalités. Face à un problème de ce type, une démarche classique en Informatique consiste à construire des taxonomies pour guider l'implémentation du système. Leur principe est de proposer différentes classifications, qui se veulent exhaustives, des combinaisons entre les modalités.

Pour les relations temporelles entre deux modalités par exemple, la taxonomie la plus complète semble être celle d'Allen (1983), qui prévoit treize configurations schématisées dans la figure 2.

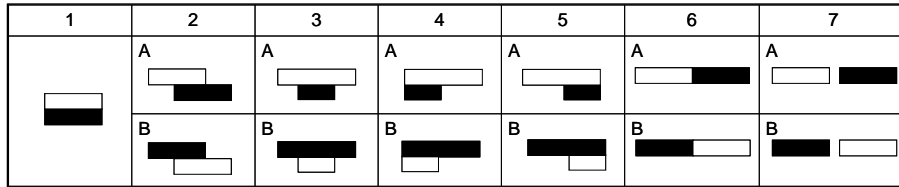


Figure 2. Taxonomie des relations temporelles d'Allen (1983).

Certaines taxonomies spécifiques à la description des Interfaces Multimodales croisent la dimension temporelle avec une dimension sémantique. C'est le cas par exemple de la taxonomie proposée par Nigay et Coutaz (1993). La dimension temporelle comporte deux valeurs (utilisation séquentielle ou parallèle), de même que la dimension sémantique (modalités combinées ou indépendantes). Il en résulte une taxonomie à quatre cas, résumée dans le tableau 1.

		Utilisation des modalités (dimension temporelle)	
		Séquentielle	Parallèle
Fusion (dimension sémantique)	Combinaison	Alterné	Synergique
	Indépendance	Exclusif	Concurrent

Tableau 1. Taxonomie d'intégration des modalités (Nigay & Coutaz, 1993).

Dans la taxonomie précédente (Nigay & Coutaz, 1993), la dimension sémantique est limitée à deux cas, combinaison/indépendance. Cette dimension peut être affinée, en particulier en explorant les différentes relations sémantiques lorsque les modalités sont combinées. C'est ce que Martin *et al.* (2001) ont fait en proposant la taxonomie TYCOON (TYpes de COOpérationN), dans laquelle six types de coopération entre modalités sont définis :

- Complémentarité : au sein d'une même commande, différents éléments sont exprimés par différentes modalités. Les informations doivent être fusionnées pour accéder au message complet.

- Redondance : au sein d'une même commande, le même élément est exprimé par différentes modalités. Les informations doivent être fusionnées pour obtenir un message commun et unique.
- Concurrence : des éléments indépendants sont exprimés par différentes modalités avec un chevauchement temporel. Les modalités sont utilisées en parallèle pour réaliser différentes actions.
- Équivalence : un même élément peut être exprimé par différentes modalités à des moments différents de l'interaction.
- Spécialisation : un élément est toujours exprimé par la même modalité.
- Transfert : un élément exprimé par une modalité est analysé par une autre modalité.

L'avantage de ces taxonomies est qu'elles permettent au concepteur de n'oublier aucun cas de figure dans l'intégration des modalités par le système. Leur inconvénient est qu'elles restent très abstraites par rapport à un contexte et un contenu d'interaction donné. Par exemple, une taxonomie peut indiquer au développeur que son système doit pouvoir gérer une utilisation *complémentaire* des modalités (comme dans l'exemple de la figure 1, « supprime celle-ci »), mais sans lui donner un aperçu concret des cas qu'il va devoir prendre en compte, et de leur diversité. Un autre inconvénient des taxonomies est qu'elles ne permettent pas distinguer entre les configurations très fréquentes dans les comportements utilisateur, dont la prise en compte va être critique à l'utilisabilité du système, et les cas marginaux, voire exceptionnels, pour lesquels la robustesse du traitement est moins cruciale.

2.2.2. L'étude des comportements utilisateurs

Pour connaître la plausibilité comportementale et la fréquence d'utilisation de chacun des cas prévus dans les taxonomies d'intégration des modalités, il est nécessaire d'observer les utilisateurs en situation.

Par exemple, on a tendance à penser que les entrées gestuelles vont être simultanées aux références verbales dans les constructions multimodales. Or, en observant les utilisateurs, Oviatt (1999) a montré qu'en réalité, les modalités ne se chevauchent que dans la moitié des constructions multimodales. Il arrive régulièrement qu'un utilisateur face un geste (ex : sélection d'un objet), et ne lui associe une commande verbale (ex : « supprime ça ») qu'après un délai de 1 à 3 secondes (Oviatt *et al.*, 1997). Comment le développeur aurait-il pu anticiper ce type de constructions multimodales ? Comment aurait-il déterminé le seuil temporel en deçà duquel les modalités doivent être considérées comme faisant partie d'un message unique ? Les données comportementales, si elles sont utiles au développement de systèmes informatiques en général, semblent encore plus fondamentales dans le cas d'Interfaces Multimodales. Seule une bonne anticipation de la manière dont les utilisateurs vont communiquer permet d'implémenter un système efficace : les patterns extraits des corpus d'observation vont donner de

précieuses indications sur les algorithmes et spécifications informatiques pertinentes à mettre en œuvre.

Il semble par ailleurs que le comportement multimodal des utilisateurs varie selon le domaine d'application (ex : spatial ou conversationnel), le média gestuel employé (ex : souris ou écran tactile), ou encore les caractéristiques des utilisateurs (ex : langue, âge, catégorie socioprofessionnelle) – pour une revue de question voir Buisine (2005). Pour contribuer au développement d'une interface multimodale, il est donc important d'utiliser des données comportementales qui ont été recueillies dans un contexte proche de celui du futur système.

2.2.3. *Problématique pour le projet NICE*

La présente étude a été réalisée dans le cadre du projet européen NICE (IST-2001-35293) qui a pour objet de développer un système interactif dans lequel les utilisateurs peuvent interagir avec des Agents Conversationnels (personnages virtuels) à l'aide d'une interface multimodale en entrée. L'application est destinée à l'interaction conversationnelle et ludique, et le public cible est celui des enfants, adolescents et jeunes adultes de 9 à 18 ans. L'introduction des modalités verbale et gestuelle en entrée semble particulièrement intéressante pour ce type d'application dans lequel le but même de l'interaction est de converser avec des personnages.

Cette combinaison particulière fait du système NICE un type d'application jamais étudié auparavant. Plusieurs facteurs fondamentaux le distinguent en effet des systèmes multimodaux classiques, ce qui empêche de transposer les résultats de la littérature. Parmi ces facteurs, citons :

- La présence d'Agents Conversationnels tout d'abord, dont on peut supposer qu'elle va influencer l'interaction : du point de vue des utilisateurs, on n'interagit peut-être pas avec un personnage virtuel comme on interagit avec une interface non personnifiée.
- Le genre de l'application est différent de celui des systèmes classiquement étudiés dans la littérature : nous nous intéressons en effet à un scénario de jeu, incluant des aspects conversationnels et des aspects orientés tâche avec des objets graphiques. A l'inverse, la plupart des interfaces multimodales existantes concernent des applications spatiales (manipulation de cartes géographiques) ou de type page web, par exemple avec des formulaires à remplir (pour une revue de question sur les domaines d'application des interfaces multimodales, voir Benoit *et al.*, 2000).
- Notre population cible n'est pas non plus celle qui a été la plus testée dans les études antérieures : il existe en effet très peu de données sur le comportement multimodal des enfants. Par ailleurs, notre système étant destiné à des joueurs d'âges variés, il sera nécessaire de comparer le comportement d'enfants et de jeunes adultes sur une même application, ce qui, également, semble n'avoir jamais été fait auparavant.

Nous nous retrouvons donc devant la problématique suivante : pour optimiser la conception de l'interface multimodale, nous souhaitons connaître les caractéristiques des comportements des utilisateurs. Or, les résultats antérieurs sur l'utilisation des interfaces multimodales ne semblent pas transférables au système que nous étudions ; il est donc plus sûr de mettre en place un recueil de données spécifique pour le projet NICE.

Mais comment recueillir le comportement des utilisateurs face à un système qui n'existe pas encore ? Dans le paragraphe suivant, nous présentons la technique du « Magicien d'Oz », qui est une méthode générique de simulation et qui nous a permis, dans le projet NICE, d'étudier les comportements utilisateurs alors même que le système n'était pas encore développé.

3. Simulation en Magicien d'Oz

Le Magicien d'Oz est une technique ancienne de simulation des systèmes techniques avant leur mise en service effective. Elle est couramment utilisée en conception d'Interfaces Homme-Machine incluant notamment la reconnaissance vocale (Dahlbäck *et al.*, 1993). La technique consiste à charger un compère humain de contrôler à distance l'interface de l'utilisateur de sorte que l'interaction soit fluide et que le système semble réellement fonctionnel. Les utilisateurs ne sont informés qu'a posteriori que le système était simulé (pour des raisons éthiques il faut toujours, en fin d'expérimentation, dévoiler le dispositif utilisé et les buts de l'expérimentation).

Le compère humain (ou magicien) réalise diverses fonctionnalités à la place du système, mais cela ne signifie pas que l'on puisse mettre en œuvre un Magicien d'Oz en partant de rien. Cela suppose au contraire l'existence d'une véritable plateforme de simulation, comprenant au minimum :

- L'interface utilisateur en sortie (affichage graphique, messages sonores produits par le système...),
- Le moyen, pour le magicien, d'avoir connaissance en continu du comportement de l'utilisateur (transmissions par caméras, par micros, par événements système...),
- L'interface magicien permettant de contrôler l'interface utilisateur, et placée en réseau avec celle-ci. Pour que la simulation soit plausible, les réactions du magicien doivent être aussi rapides que possible. Un soin particulier doit donc être apporté à la conception de son interface, en y incluant par exemple des réponses pré-encodées et des accès directs aux fonctionnalités (Oviatt *et al.*, 1992 ; Dahlbäck *et al.*, 1993). Ceci suppose d'avoir au préalable élaboré un modèle de réalisation de la tâche ou du scénario. Certains systèmes complexes peuvent nécessiter l'intervention de plusieurs magiciens pour gérer l'interaction (un pour simuler la reconnaissance vocale, un autre pour les actions gestuelles...), ce qui implique

autant d'interfaces à développer et à coordonner (voir par exemple la plateforme développée par Coutaz *et al.*, 1996). Pour augmenter la crédibilité du dispositif, il est recommandé de simuler quelques erreurs de traitement (Oviatt *et al.*, 1992).

De telles plateformes permettent ainsi de recueillir de grandes quantités de données comportementales, sous différentes formes : corpus audio-visuel, événements système (log files), données subjectives fournies par les utilisateurs à l'issue du test (interviews, questionnaires). La simulation permet en outre une évaluation préliminaire, non pas du système, mais par exemple de l'application visée (utilité potentielle...), du modèle de dialogue, ou de divers éléments de design.

4. Description de l'étude réalisée

Nous avons donc mis en place un recueil de données en Magicien d'Oz dès le début du projet NICE. Les graphismes des Agents 3D n'avaient pas encore été créés, et le scénario exact n'avait pas encore été fixé par le consortium. Nous avons donc imaginé un scénario simple, impliquant principalement des aspects orientés tâche, et quelques aspects conversationnels : le jeu consistait à rendre service à des Agents en allant trouver l'objet qui manquait à chacun et en le leur apportant. Nous avons développé pour cela des Agents 2D. Les graphismes du jeu comportaient quatre décors, quatre Agents et dix-huit objets déplaçables (ex : lampes, plantes...).

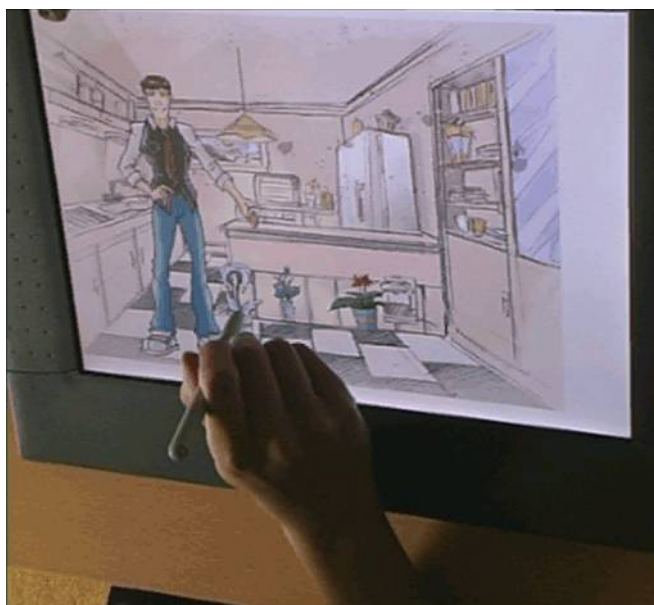


Figure 3. Notre système simulé.

4.1. Méthode

4.1.1. Utilisateurs

Dix-sept utilisateurs ont participé à cette expérience. Ils se répartissent en deux groupes :

- Sept adultes : 3 hommes (âge moyen de 24 ans 6 mois, $\sigma = 2,5$ ans) et 4 femmes (âge moyen de 26 ans 9 mois, $\sigma = 7,5$ ans).
- Dix enfants : 7 garçons (âge moyen de 11 ans et 6 mois, $\sigma = 1,9$ ans) et 3 filles (âge moyen de 9 ans et 4 mois, $\sigma = 0,6$ ans).

4.1.2. Dispositif Magicien d'Oz

Le dispositif de Magicien d'Oz était composé de deux ordinateurs (figure 4) : le PC2 assurait la présentation du jeu aux utilisateurs et le PC1 la gestion de l'interaction par le Magicien. L'utilisateur et le Magicien se trouvaient dans des pièces différentes.

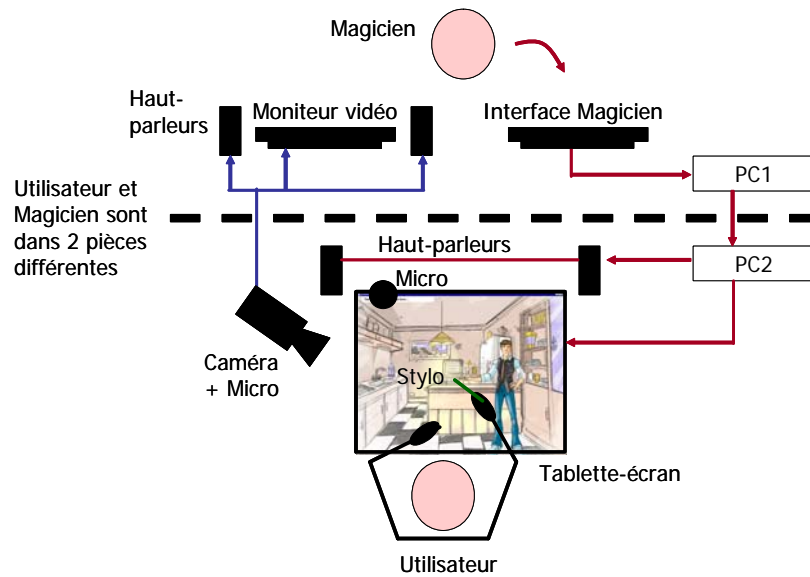


Figure 4. Schéma du dispositif Magicien d'Oz utilisé.

Le PC2 était relié à une tablette-écran permettant une interaction directe sur l'écran par l'intermédiaire d'un stylo. Deux haut-parleurs assuraient la diffusion de la voix des Agents ; toutes les répliques verbales des Agents étaient également transcrites en haut de l'écran de l'utilisateur. Un micro était disposé derrière l'écran pour inciter les utilisateurs à parler. La reconnaissance vocale, l'interprétation du

langage, des gestes et de leurs combinaisons multimodales étaient cependant simulées par le Magicien.

Une caméra numérique, installée au-dessus de la tablette-écran, permettait un enregistrement audio-visuel du comportement de l'utilisateur en conservant son anonymat (le visage n'était pas filmé).

Un moniteur et des enceintes reliés à la caméra numérique permettaient au Magicien de suivre en temps réel les paroles et les actions de l'utilisateur. Pour y répondre, il disposait d'une interface (figure 5) lui permettant de modifier l'environnement du jeu (changement de lieu, déplacement d'un objet) ou le comportement des Agents (déclenchement de paroles et de comportements non-verbaux).

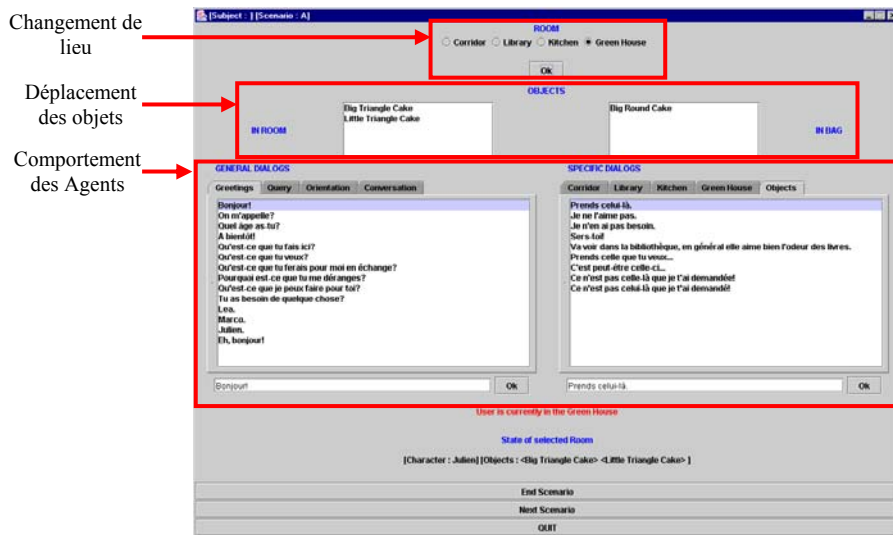


Figure 5. Interface Magicien.

Quatre-vingt trois séquences comportementales possibles pour les Agents avaient été anticipées et pré-encodées afin de guider l'utilisateur, de dialoguer avec lui et de répondre à ses requêtes les plus probables compte tenu du scénario de jeu. Chaque séquence comportementale était composée d'une réplique verbale et d'une animation de l'Agent en conséquence (animation de la tête de l'Agent, de ses yeux, de la direction de son regard, de sa bouche et de ses bras). Les séquences de comportements multimodaux configurées dans des fichiers XML pouvaient être déclenchées depuis l'interface du Magicien, où elles étaient organisées en plusieurs onglets dans deux grandes catégories (dialogues généraux, dialogues spécifiques). Outre ces séquences comportementales pré-encodées, le Magicien avait également

la possibilité, au cours de l'expérience, de taper une réplique non prévue et de l'associer au comportement non verbal de la réplique de son choix dans la base.

4.1.3. Analyses

A l'issue de cette expérience, nous avons obtenu un corpus vidéo d'interaction multimodale entre les utilisateurs et les Agents. Ce corpus a été annoté à l'aide des outils Praat (<http://www.praat.org/>) et Anvil (Kipp, 2001). Praat est un logiciel d'analyse du son, que nous avons utilisé pour transcrire les verbalisations des utilisateurs. Le signal sonore provenant des vidéos a été importé dans Praat, puis segmenté manuellement, et les mots associés à chaque segment ont été retranscrits.

Le logiciel Anvil a ensuite été utilisé afin de compléter les annotations (figure 6). Le corpus contenait 117 constructions multimodales, que nous avons pu analyser précisément, notamment du point de vue des relations temporelles et sémantiques entre modalités.

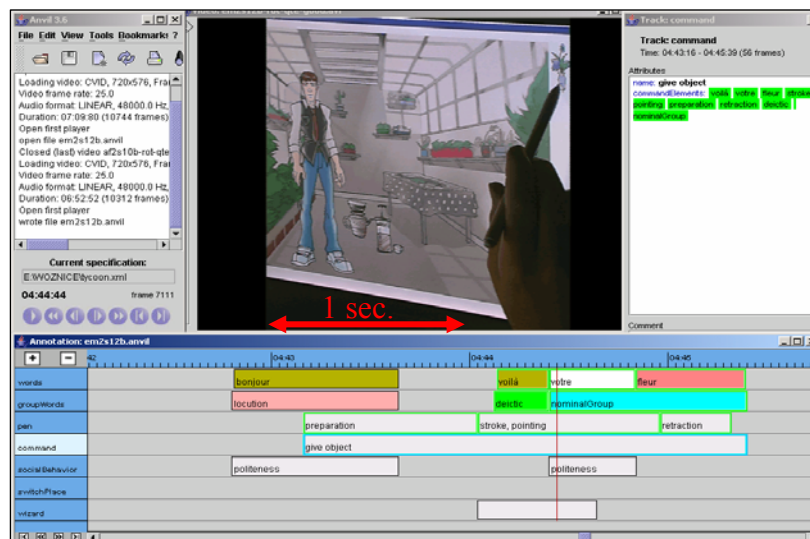


Figure 6. Copie d'écran du logiciel Anvil : un extrait vidéo est visible dans la partie supérieure de la fenêtre, et la partie inférieure présente les annotations réalisées sur cet extrait.

4.2. Résultats

Cette expérience nous a permis de mettre en évidence des résultats originaux vis-à-vis de la littérature antérieure sur les interfaces multimodales. Ces résultats

sont décrits en détail ailleurs (Buisine & Martin, 2003, 2005). Pour synthétiser, parmi les 117 constructions multimodales observées, nous avons relevé autant de constructions redondantes que complémentaires. Nous avons également trouvé un pourcentage non négligeable (16%) de constructions concurrentes dans notre corpus – résultat surprenant par rapport à la littérature sur l'Interaction Homme-Machine multimodale, dans laquelle ce type de coopération n'a jamais été décrit. Ces constructions concurrentes ont presque toutes été produites par des enfants. Aucune des études antérieures que nous avons recensées (Buisine, 2005) ne rapporte d'exemples de constructions concurrentes. Nous avons également été amenés à définir de nouveaux types de coopération qui n'existaient pas dans les taxonomies de comportements multimodaux (Buisine & Martin, 2005). Ces résultats sur les coopérations sémantiques entre modalités ont contribué à l'élaboration de l'algorithme de fusion multimodale.

Au niveau temporel, les modalités ont majoritairement été intégrées de façon simultanée (73%), c'est-à-dire avec un chevauchement temporel entre parole et geste. Ce résultat est conforme aux données de la littérature (Oviatt *et al.*, 2003). Nous avons observé autant de cas de précedence du geste sur la parole que l'inverse. Pour le futur système, nous avons déduit les paramètres temporels de fusion multimodale suivants :

- Lorsque les deux modalités sont séparées d'un délai inférieur ou égal à 1,8 sec (entre la fin de la première modalité et le début de la seconde), elles doivent être fusionnées. Au delà de 2 sec de délai, il est probable que la succession des modalités soit une répétition ou des commandes indépendantes.
- Cette fenêtre de fusion peut également être exprimée comme un délai de 2,5 sec entre la fin de la première modalité et la fin de la seconde.
- Dans la fenêtre temporelle de fusion, les modalités doivent être fusionnées quel que soit leur ordre de succession (parole puis geste ou geste puis parole).

Enfin, cette expérience a permis pour la première fois de comparer le comportement multimodal de jeunes adultes à celui d'enfants sur la même application. Nos résultats indiquent que les adultes semblent utiliser plus de constructions multimodales (27,5%) que les enfants (18,3%). Néanmoins, il y a eu relativement peu de spécificités du comportement multimodal des enfants dans notre corpus. Concernant l'intégration sémantique, nous avons relevé une tendance des enfants à produire des commandes concurrentes : les enfants ont parfois dissocié le canal verbal et le canal gestuel pour réaliser deux actions à la fois. Le traitement de ce type de constructions est un défi pour le futur système NICE. Concernant l'intégration temporelle, notre corpus contenait une majorité de constructions simultanées, mais les constructions séquentielles provenaient principalement du groupe des enfants. Nous avons noté d'autres différences entre adultes et enfants, qui relèvent moins de la multimodalité que de leur comportement de jeu.

Enfin, signalons que notre dispositif expérimental a globalement été efficace, puisqu'il a permis de mener à bien le recueil de corpus. Tous nos utilisateurs ont pu

réaliser le scénario avec succès, et ils ont cru interagir avec un système fonctionnel (un seul utilisateur, exclu de l'analyse par la suite, avait deviné qu'il s'agissait d'une simulation).

5. Discussion

La procédure adoptée pour la conception de l'interface multimodale du système NICE avait pour objectif d'optimiser la phase de développement, en anticipant les cas de constructions multimodale, l'intégration temporelle et sémantique des modalités. En amont du développement de l'interface, nous avons simulé le futur système par la méthode du Magicien d'Oz, afin de recueillir un corpus de comportements multimodaux des utilisateurs.

Le principe de cette approche – guider le développement grâce au comportement des utilisateurs – semble à première vue correspondre à une démarche de Conception Centrée Utilisateur. En réalité, elle va au-delà. En effet, la Conception Centrée Utilisateur (norme ISO13407, 1999) consiste à considérer les utilisateurs et leurs besoins tout au long du processus de développement d'une application informatique. Elle impose que le développement du produit soit guidé par les besoins des utilisateurs plutôt que par les possibilités technologiques.

La Conception Centrée Utilisateur aboutit à des spécifications systèmes dans la mesure où elle apporte des contraintes au développement. Ici au contraire le recueil des données utilisateur a facilité la réalisation technique en fournissant des patterns de combinaison des modalités. Dans notre étude, l'utilisateur n'a pas seulement été consulté et observé, mais il a fourni les briques de base du futur système. Nous avons ainsi doublement optimisé la conception : d'une part en permettant, dans une certaine mesure, que l'interface soit en adéquation avec les comportements des futurs utilisateurs, et réponde donc à leur besoin (dans l'esprit de la Conception Centrée Utilisateur), et d'autre part en simplifiant la phase de programmation informatique (critère qui ne figure pas dans la norme de Conception Centrée Utilisateur, et qui lui est même souvent contraire).

6. Conclusion

Etant donné que de nombreux produits, de nos jours, intègrent une Interface Homme-Machine multimédia (téléphone mobile, PDA, domotique...) et nécessitent un développement informatique, nous pensons que le processus de conception suivi au cours de notre étude est susceptible d'être appliqué avec succès à d'autres types de projets. En outre, le développement du prototypage d'interfaces et du prototypage virtuel, et le fait qu'ils soient aujourd'hui de plus en plus accessibles aux équipes de conception, allègent les contraintes de mise en place d'une plateforme de simulation en Magicien d'Oz.

L'essor pris par les interfaces numériques et les systèmes informatiques s'étant bien souvent fait au détriment de la facilité d'utilisation, l'adoption d'une telle démarche offre la perspective d'augmenter l'acceptabilité d'un produit et donc son succès. Notre étude suggère par ailleurs que ce type de processus de conception peut même faciliter la réalisation technique d'un produit.

Cette étude a été en partie financée par le projet européen NICE, IST-2001-35293 (<http://www.niceproject.com/>).

7. Bibliographie

- Allen J., "Maintaining knowledge about temporal intervals", *Communication of the ACM*, vol. 26, 1983, p. 832-843.
- Benoit C., Martin J.C., Pelachaud C., Schomaker L., & Suhm B., "Audio-visual and multimodal speech-based systems". In D. Gibbon, I. Mertins & R. Moore (Eds.), *Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation*, 2000, p. 102-203. Boston: Kluwer Academic Publishers.
- Buisine S., Conception et Evaluation d'Agents Conversationnels Multimodaux Bidirectionnels. Thèse de Doctorat en Psychologie Cognitive - Ergonomie, Université de Paris 5, 2005.
- Buisine S., & Martin J.C., "Experimental evaluation of bi-directional multimodal interaction with conversational agents". *Proceedings of Interact'2003*, 2003, p. 168-175.
- Buisine S., & Martin J.C., "Children's and adults' multimodal interaction with 2D conversational agents". *Proceedings of CHI'2005*, 2005, p. 1240-1243.
- Coutaz J., Salber D., Carraux E., & Portolan N., "NEIMO, a multi-workstation usability lab for observing and analyzing multimodal interaction". *Proceedings of CHI'96*, 1996, p. 402-403.
- Dahlbäck N., Jönsson A., & Ahrenberg L., "Wizard of Oz studies - Why and how". *Proceedings of IUI'93*, 1993, p. 193-200.
- ISO13407. Human-centred design processes for interactive systems, 1999, Genève: International Organization for Standardization.
- Kipp M., "Anvil - A generic annotation tool for multimodal dialogue". *Proceedings of Eurospeech'2001*, 2001, p. 1367-1370.

- Martin J.C., Grimard S., & Alexandri K., "On the annotation of the multimodal behavior and computation of cooperation between modalities". *Proceedings of International Conference on Autonomous Agents Workshop on Representing, Annotating, and Evaluating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents*, 2001, p. 1-7.
- Nigay L., & Coutaz J., "Design space for multimodal systems: Concurrent processing and data fusion". *Proceedings of InterCHI'93*, 1993, p. 172-178.
- Oviatt S.L., "Ten myths of multimodal interaction", *Communications of the ACM*, vol. 42, 1999, p. 74-81.
- Oviatt S.L., Cohen P.R., Fong M.W., & Frank M.P., "A rapid semi-automatic simulation technique for investigating interactive speech and handwriting". *Proceedings of ICSLP'92*, 1992, p. 1351-1354.
- Oviatt S.L., Cohen P.R., Wu L., Vergo J., Duncan L., Suhm B., Bers J., Holzman T., Winograd T., Landay J., Larson J., & Ferro D., "Designing the user interface for multimodal speech and gesture applications: State-of-the-art systems and research directions", *Human Computer Interaction*, vol. 15, 2000, p. 263-322.
- Oviatt S.L., Coulston R., Tomko S., Xiao B., Lunsford R., Wesson M., & Carmichael L., "Toward a theory of organized multimodal integration patterns during Human-Computer Interaction". *Proceedings of ICMI'03*, 2003, p. 44-51.
- Oviatt S.L., De Angeli A., & Kuhn K., "Integration and synchronization of input modes during multimodal human-computer interaction". *Proceedings of CHI '97*, 1997, p. 415-422.