A paraître dans Actes ERGO-IA96, Biarritz, Oct. 1996

NEIMO, un laboratoire d'utilisabilité numérique : Leçons de l'expérience

L. Aublet-Cuvelier, E. Carraux, J. Coutaz, L. Nigay, N. Portolan*, D. Salber, M.-L. Zanello

CLIPS-IMAG, BP 53, 38041 Grenoble Cedex 9
Tel. 76 51 48 54, fax 76 44 66 75
E.mail : {Laurent.Aublet-Cuvelier, Eric.Carraux, Joelle.Coutaz,
Laurence.Nigay, Daniel.Salber, Marie-Laure.Zanello}@imag.fr

* CCETT, BP 59, 35512 Cesson Sévigné
Tél. 99 12 43 84
email : portolan@ccett.fr

RÉSUMÉ

Dans cet article, nous décrivons NEIMO, un laboratoire d'utilisabilité numérique d'aide à la capture automatique et à l'analyse de données comportementales pour des situations d'interaction innovante (multimodalité, communication homme-homme médiatisée). Au-delà des services d'observation et d'annotation, NEIMO permet la mise en œuvre d'expérimentations de type Magicien d'Oz. Cette plate-forme logicielle multi-station générique a été appliquée à différentes situations. Nous décrivons ici une instanciation de NEIMO pour l'étude d'un terminal multiservice multimodal de télécommunication. Les résultats de cette expérience appellent une réflexion critique et militent pour un nouvel agenda de recherche que nous présentons en conclusion.

ABSTRACT

In this article, we describe NEIMO, a generic and flexible multiworkstation usability lab that supports the observation and analysis of multimodal interaction as well as Wizard of Oz experiments. It captures behavioral data at multiple levels of abstraction from keystroke to high level tasks. NEIMO has been used for different case studies. Here, we describe NEIMO instantiated for the study of the multimodal interaction with a telecommunication terminal. Lessons drawn from this experience call for a new agenda of research.

MOTS-CLÉS : Laboratoire d'utilisabilité, évaluation ergonomique, outil pour l'évaluation, Magicien d'Oz, interaction multimodale.

KEY-WORDS : Usability lab, usability testing, usability lab, evaluation tool, Wizard of Oz, multimodal interaction.

INTRODUCTION

L'évaluation de l'utilisabilité des interfaces homme-machine représente aujourd'hui un courant de recherche actif. Il en résulte un foisonnement de méthodes (et de convictions) qui répondent chacune à des objectifs et des moyens précis [2]. Dans cet article, nous nous intéressons à l'observation expérimentale de sujets mis en situation d'interaction innovante.

Le processus d'observation expérimentale comprend trois étapes : le recueil de données sur l'utilisation du logiciel, l'analyse de ces données et, selon les objectifs de l'expérimentateur, la formulation de recommandations, de règles générales d'ergonomie ou encore l'affinement d'une théorie ou d'un modèle cognitifs. L'une des difficultés de cette démarche est le temps dédié au dépouillement du recueil que les informaticiens (et les décideurs) jugent souvent incompatible avec les délais imposés par le plan de développement.

Une situation d'interaction innovante est synonyme de prospection, d'absence de matériau scientifique éprouvé pour décider de manière prédictive de la compatibilité des choix de conception avec les attentes des futurs utilisateurs. A ce titre, la multimodalité, qui permet notamment l'usage synergique de plusieurs techniques d'interaction comme la parole et le geste, offre un terrain d'étude privilégié [20]. La Communication Homme-Homme Médiatisée (CHHM), qui met en scène plusieurs utilisateurs communicants [23], est un second domaine à défricher.

Dans cet article, nous décrivons NEIMO, un laboratoire d'utilisabilité numérique d'aide à la capture automatique et à l'analyse de données comportementales en situation innovante. La motivation première n'est pas l'évincement du spécialiste mais l'élimination des tâches mécaniques de capture et de dépouillement des données, et le signalement de points singuliers en relation avec les objectifs de l'expérimentateur. Le second élément directeur au développement de NEIMO est son utilisation comme outil d'évaluation formative [11] d'interfaces prospectives partiellement implémentées.

Après une revue de l'état de l'art, nous présentons les principes de la plate-forme NEIMO que nous illustrons ensuite avec l'étude de l'usage de la multimodalité pour un terminal multiservice de télécommunication (vidéophonie, émission et réception de fax intégrées au poste de travail, etc.). Les résultats de cette expérience appellent une réflexion critique et militent pour un nouvel agenda de recherche que nous présentons en conclusion. Le point de vue adopté dans cet article est résolument centré sur les propriétés de la plate-forme NEIMO. Nos recherches sur l'usage des modalités pour des tâches de CHHM ne seront donc pas l'objet de notre discussion.

ETAT DE L'ART

Aujourd'hui, la plupart des industriels soucieux de la qualité ergonomique de leurs logiciels disposent d'un laboratoire d'utilisabilité. De même, la capture de comportement par un dispositif logiciel est une technique connue. La nouveauté tient à la synergie des deux moyens d'observation. Il s'agit, pendant l'expérimentation, de compléter l'enregistrement vidéo analogique par un maximum de données numériques afin d'une part, de minimiser la consultation des bandes vidéo au moment de l'analyse et d'autre part, de produire automatiquement des résultats statistiques via des logiciels du commerce.

Idéalement, les informations numériques comprennent les événements produits par le sujet mais aussi les annotations de l'observateur. Parallèlement, la vidéo analogique est automatiquement estampillée et indexée par les événements sujet et expérimentateur. Ces services de base sont complétés d'outils d'édition et de navigation. Les premiers permettent, en phase de dépouillement, de modifier la "bande numérique" de capture, de définir des filtres pour ne retenir que les séquences utiles. Les seconds aident à positionner la bande vidéo analogique sur un critère donné.

A l'exception de DRUM [15], la capture des événements du sujet se pratique au niveau d'abstraction le plus bas tel que la frappe de caractères [19, 25], ou à un niveau d'abstraction figé comme celui des widgets [10]. En conséquence, la structuration hiérarchique de l'activité ou les entrelacements de tâches sont perdus et doivent être reconstitués à la main. L'espionnage limité aux événements widgets ne couvre que partiellement l'interaction. En particulier, les objets du domaine (comme les figures

géométriques d'un éditeur de dessin) sont réalisées directement au moyen des primitives graphiques de la boîte à outils d'accueil. Limiter le suivi aux widgets facilite la mise en œuvre de l'outil de capture mais une partie sensible de l'interaction spécifique au domaine applicatif est perdue.

Les annotations permettent de traduire des phénomènes attendus et caractéristiques de l'utilisabilité d'un logiciel : entrée et sortie de séquence d'erreur, demande d'aide, consultation de la documentation, recherche d'information, hésitations, etc. La plupart des outils incluent des boutons d'annotation prédéfinis. L'avantage est une prise de note rapide synchronisée automatiquement avec la bande vidéo. L'inconvénient est le manque de souplesse. L'expérience montre que les annotations sont également spécifiques à une expérimentation donnée mais connues par avance (par exemple les propriétés CARE [3] pour caractériser à la volée l'usage de la multimodalité). Elles peuvent aussi émerger en phase de mise au point de l'expérimentation. Elles doivent donc être intégrables dynamiquement au moins avant l'expérimentation proprement dite. DRUM est le seul outil à notre connaisance à offrir des annotations personnalisables qui ne soient pas de simples commentaires.

Certaines interfaces et notamment les interfaces multimodales, ou certaines applications et notamment les collecticiels, ont un coût de développement élevé. Dans ces conditions, il est intéressant d'en tester l'utilisabilité avec une version partiellement réalisée de l'interface. La technique du Magicien d'Oz, bien que discutable si l'on s'en tient à la qualité écologique des données observées, répond au problème de coût de développement et offre une solution raisonnable dans le contexte d'une évaluation formative.

La plupart des systèmes expérimentaux de Magicien d'Oz ont été appliqués à l'étude de l'usage de la langue naturelle pour des systèmes de recherche d'information et de réservation [22, 8, 9, 6], ou encore pour des systèmes experts d'aide et de formation-conseil [12, 29, 5, 21]. Une connaissance bien cernée du corpus offre l'avantage d'ajuster les paramètres du système de reconnaissance de la langue et par là, d'en améliorer les temps de réponse et la robustesse. Récemment, la technique du Magicien d'Oz s'est vue utilisée dans l'étude de la multimodalité [1, 4, 16, 17, 18, 28]. Mais la capture des données comportementales se pratique sans support informatique, laissant à l'analyste la charge de dépouiller manuellement le recueil.

La plate-forme informatique d'expérimentations en Magicien d'Oz s'appuie souvent sur un montage astucieux de logiciels du commerce incluant un système de partage des ressources physiques de l'interaction comme l'écran et la souris [18, 1]. L'avantage est une mise en œuvre rapide de la plate-forme. L'inconvénient est l'absence de support dédié à la tâche du compère. En particulier, on ne peut masquer au sujet les hésitations du compère ou ses erreurs de manipulation de la souris.

Tous les rapports méthodologiques sur la technique du Magicien d'Oz insistent sur la nécessité, pour le compère, d'observer un comportement cohérent en réponse aux diverses formes de déviations opportunistes du sujet (exploration, détournement du scénario, blocages sur erreur fatale, etc.). L'entraînement du compère est nécessaire mais ne suffit pas à éliminer la surcharge cognitive. Dans ces conditions, on convient d'assister le compère par une équipe comme dans un laboratoire d'utilisabilité. Actuellement, les plates-formes d'expérimentation n'offrent pas de support informatique à l'activité des partenaires d'une expérimentation.

Avec NEIMO, réalisé sur stations Apple Macintosh Quadra, nous offrons une première réponse aux limitations identifiées dans la pratique de l'état de l'art.

LES PRINCIPES DE NEIMO

L'environnement NEIMO s'utilise en trois étapes : configuration, expérimentation et analyse. Nous décrirons les dimensions configurables de NEIMO avec la présentation de la phase expérimentale et de l'exemple applicatif introduit plus loin.

En phase d'expérimentation,

- le sujet exécute un ensemble de scénarios sur une station dédiée. La description des scénarios est, pour l'instant, sur support papier (mais sa traduction informatique serait utile à l'outil d'analyse pour détecter les déviations de stratégie du sujet).

- Les expérimentateurs, regroupés dans une même salle mais distincte de celle du sujet, disposent chacun d'un poste de travail. Chaque station est configurée en fonction du rôle de l'expérimentateur : annotateur, simulateur de tâche du domaine, traducteur de parole, etc. Le nombre de stations de travail n'est pas fixe. Il peut se décider à la configuration. Toutes les stations sont reliées sur un réseau local Ethernet.

- Les actions du sujet et celles des expérimentateurs sont enregistrées localement dans un fichier de capture au format ASCII. L'enregistrement local sur chaque station a l'avantage de minimiser l'utilisation du réseau et de là, éviter un ralentissement rédhibitoire de la plate-forme d'expérimentation. La fusion des fichiers de capture locaux en une trace informatique unique a lieu en fin d'expérimentation. NEIMO capte le comportement du sujet à différents niveaux d'abstraction, le choix du grain étant fixé à la configuration : actions physiques sur les dispositifs d'entrée, tâches du domaine au niveau de granularité voulue (tâche élémentaire, tâche composée).

- NEIMO a été conçu dès le départ pour l'étude de la multimodalité. Les propriétés CARE (Complémentarité, Assignation, Redondance, Equivalence des modalités) sont utilisées comme élément de structuration des observations : à la configuration, les expérimentateurs décident que telle tâche pourra être effectuée par le sujet de manière équivalente avec la modalité M1 (par exemple la parole) ou la modalité M2 (par exemple, la manipulation directe) ; pour telle autre, des expressions redondantes seront permises (par exemple accepter la sélection du téléphone direct de Paul conjointement à l'énoncé oral "appeler Paul"), ou encore permettre la complémentarité et notamment l'usage d'expressions déictiques comme dans "appeler cette personne".

Dans sa version actuelle, NEIMO ne capture pas le contenu des actes de parole du sujet mais leur occurence. Pour s'adresser oralement au système, le sujet doit appuyer sur un bouton réservé. Cette option, apparemment contraignante, est néanmoins la seule réaliste dans l'état actuel de la technologie en reconnaissance de la parole. La nature des modalités (par exemple, parole, manipulation directe) est détectée par le système mais, en l'absence de systèmes de compréhension des énoncés langagiers, l'outil d'analyse n'est pas en mesure d'identifier le type exact de relation entre modalités. En conséquence, cette relation est repérée et enregistrée à la volée par annotation. Nous reviendrons sur ce point à propos de notre agenda de recherche.

Dans la troisième phase, les données comportementales sont restituées à l'analyste. Dans sa forme actuelle, l'outil d'analyse est essentiellement un restituteur "objectif" du recueil dont il facilite le parcours :

- Comme le montre la figure 1, une fenêtre (en haut à gauche) pilotable à la manière d'un magnétoscope, rejoue numériquement le contenu de l'écran du sujet. Les services de défilement usuels d'un magnétoscope sont ici augmentés de capacités de déplacements motivées par la tâche de l'analyste : passage au scénario suivant ou précédent, positionnement sur un point singulier, par exemple "se placer à la prochaine erreur de type lexical" ou "aller à l'heure 1h33'".

- L'outil d'analyse fournit (en bas à gauche) des renseignements quantitatifs sur la durée des scénarios, de même une indication sur leur réussite. L'icône du pouce

relevé dénote la satisfaction de l'expérimentateur tandis que l'icône à visage humain traduit l'impression du sujet. Ces informations proviennent d'une annotation enregistrée en phase d'expérimentation en fin de scénario. Toute information quantitative est exportable vers Excel pour une analyse statistique approfondie.

- Pour chaque scénario, les relations temporelles entre les tâches sont restituées sous forme d'un diagramme de gantt projeté sur un mur en perspective [14] (fenêtre en bas à droite). Cette technique permet de montrer sur la face centrale du mur le détail des informations faisant l'objet de l'attention de l'analyste sans perdre le contexte de l'ensemble des informations présentées sur les faces latérales. Le centre d'intérêt peut être changé par une opération de défilement latéral. Le diagramme de gantt est augmenté de bulles dont la désignation fait jaillir une annotation enregistrée pendant l'expérimentation ou dont le double clic positionne la fenêtre "magnétoscope" sur la scène correspondante.

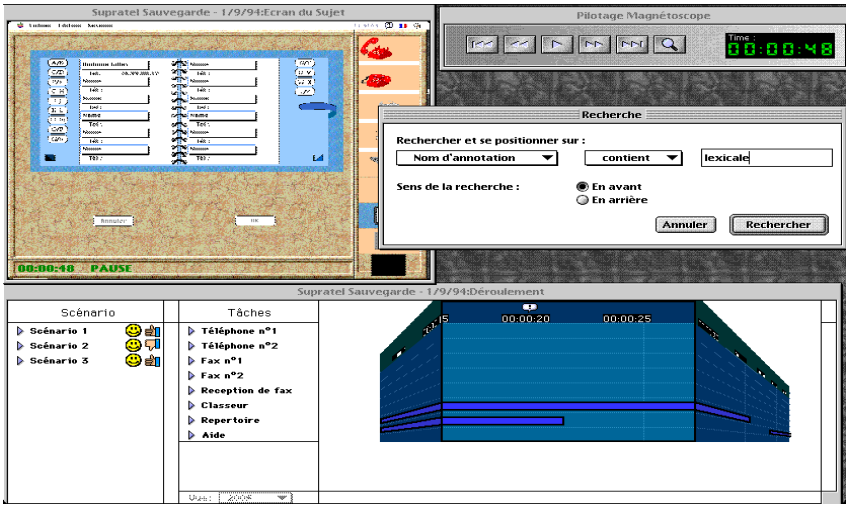

Figure 1 : Un exemple d'écran de l'outil d'analyse.

Les qualités de généricité et d'extensibilité du noyau logiciel de NEIMO nous ont permis de développer plusieurs plates-formes d'expérimentation : l'étude des stratégies de linéarisation à partir des figures de Levelt et l'usage des modalités dans des tâches de télécommunication. Nous illustrons ci-dessous les conditions expérimentales pour ce deuxième exemple : Supratel.

SUPRATEL : UN EXEMPLE D'UTILISATION DE NEIMO

Supratel est une maquette de terminal de télécommunication multiservice (fax, téléphone, répertoire, etc.) qui s'inscrit dans un processus de conception itérative. Par conséquent et en l'état, ses qualités ergonomiques ne sont pas l'objet de notre discussion mais l'utilisation de NEIMO comme outil d'étude constitue ici notre axe de réflexion. La figure 2 illustre la configuration matérielle utilisée.

La station du sujet

L'écran de la figure 3 montre le principe de l'interface du terminal utilisé par le sujet : à droite, un bandeau d'icônes correspond aux services de télécommunication ; en bas, un répertoire d'adresses est ouvert en permanence. Plusieurs tâches de télécommunication peuvent être menées de front de manière entrelacées. La zone du haut est réservée à la tâche courante, par exemple une communication vidéophonique.

Dans le cadre d'une expérimentation NEIMO, cette interface, implémentée en HyperCard s'exécute sur la station du sujet. Cette station comprend un clavier, une souris, un écran et une caméra pour les tâches de visiophonie. L'interface est actionnable par le sujet au

moyen de la manipulation directe ou bien de manière équivalente par la parole. La reconnaissance de la parole est assurée par un expérimentateur. Dans la version actuelle, la voix n'est pas véhiculée par Ethernet mais par une liaison directe analogique. La vidéo, quant à elle, est transmise de manière numérique via Ethernet.

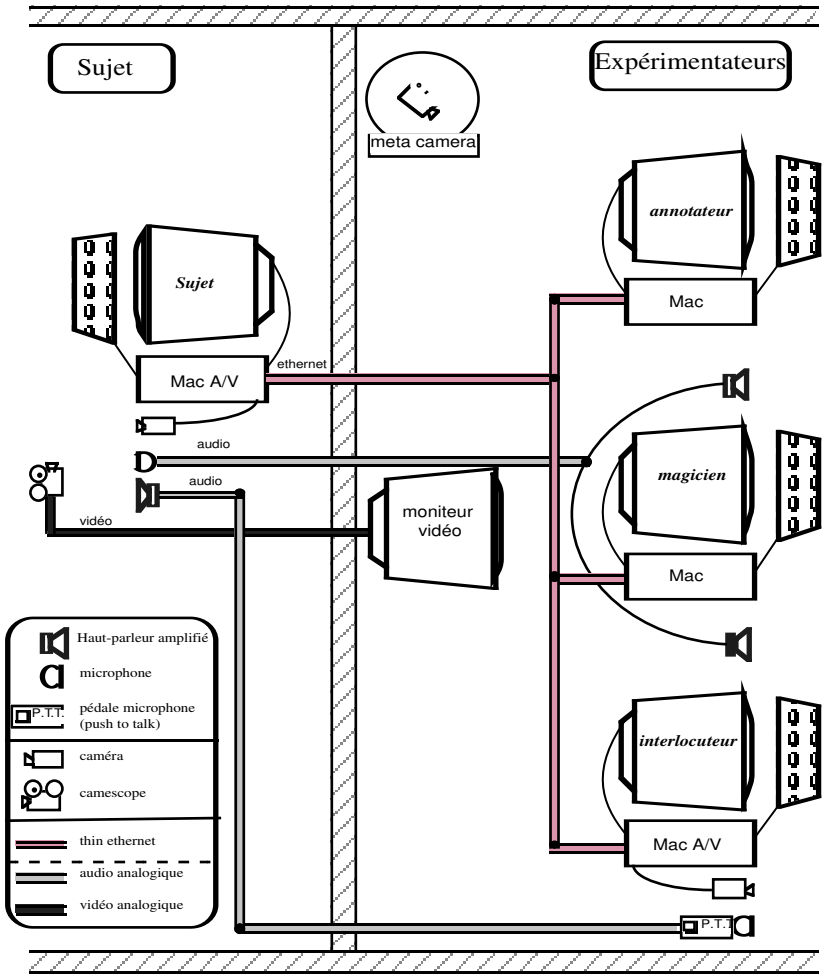

Figure 2 : Configuration matérielle de la plate-forme NEIMO pour l'étude de l'usage de la multimodalité dans Supratel.

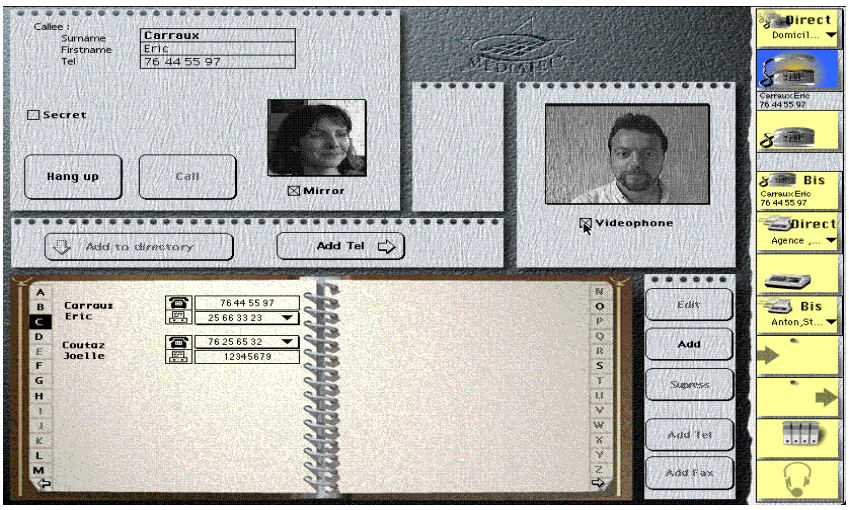

Figure 3 : L'interface multimodale du terminal multiservice Supratel utilisée par nos sujets.

L'équipement des expérimentateurs

Pour cette expérience, nous avons configuré NEIMO en 3 postes d'observation : une station pour l'annotateur, une station de Magicien d'Oz pour la reconnaissance de la parole et une station pour l'interlocuteur, ce dernier jouant le rôle d'un correspondant dans une tâche de télécommunication. Egalement, au centre de la salle mais en surélévation, un moniteur vidéo analogique offre une vue complémentaire 3/4 arrière sur le sujet (écran, visage et mains).

Dans la version actuelle, nous l'avons vu, le canal audio entre sujet et expérimentateurs est purement analogique :

- La communication audio dans le sens sujet-expérimentateurs est active en permanence. Elle sert à la fois aux tâches de communication téléphonique entre le sujet et l'interlocuteur et aux tâches d'observation du comportement vocal du sujet : actes de parole à destination du système comme "Appeler Paul" et métacommentaires du type "Bon, j'ai fini le scénario". Les premiers sont traités par le magicien tandis que les métacommentaires concernent l'annotateur voire l'interlocuteur pour le cas où il faudrait remettre subtilement le sujet sur les rails du scénario. Ici encore, on note l'intérêt d'une configuration multi-station pour la répartition des tâches.

- La communication dans le sens expérimentateurs-sujet est activée sous le contrôle de l'interlocuteur via un bouton poussoir : pour répondre au principe de l'engagement de l'utilisateur dans la réduction des erreurs de mode [24], la communication est active tant que le bouton poussoir est enfoncé (c'est fatigant mais efficace). L'activation du canal audio en direction du sujet doit être observable par tous les expérimentateurs afin que ceux-ci, qui ont tendance à faire des commentaires oraux sur la situation, puissent observer le silence et adopter le mode de communication gestuel. A l'évidence, une communication audio permanente entre les deux salles aurait permis au sujet d'entendre les commentaires privés des expérimentateurs. Il faut noter que la nécessité de ce bouton n'a été identifiée que suite à la phase d'entraînement !

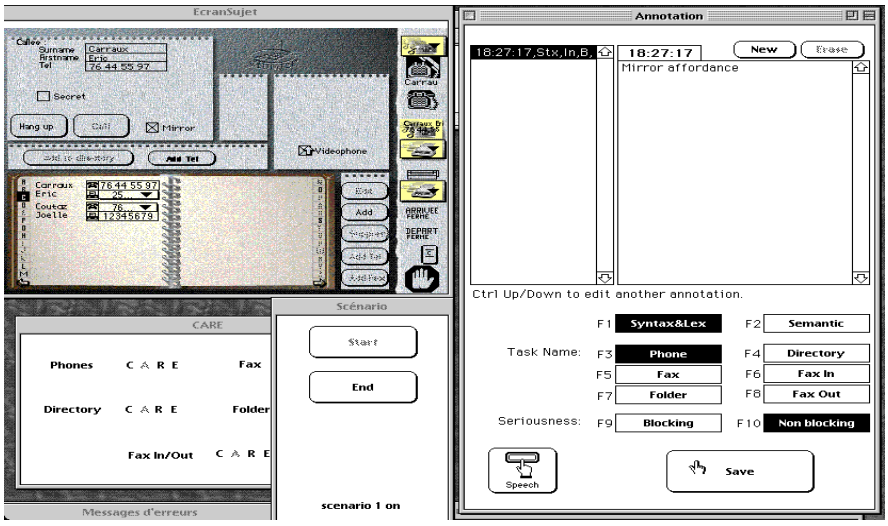

Figure 4 : L'interface de l'annotateur.

L'interface de l'annotateur

La figure 4 montre un exemple de configuration de la station de l'annotateur. En haut à gauche, une reproduction miniature de l'écran du sujet assure un espionnage fidèle des actions du sujet. Tous les déplacements et clics souris du sujet, de même que les actions claviers sont perceptibles à l'annotateur. Ce niveau de détail permet de détecter les hésitations, les recherches infructueuses et de les noter. Grâce à des boutons dédiés,

l'expérimentateur indique pour la tâche en cours s'il s'agit d'une erreur de type lexicale, syntaxique ou sémantique et son niveau de gravité.

Toutes les annotations antérieures sont listées en haut dans une fenêtre éditable à défilement. Dans l'exemple, on relève une seule annotation produite à 18h27'17" de type lexicale et concernant un bouton (symbole B). Les annotations peuvent être accompagnées d'un commentaire saisi au clavier (ou, dans une prochaine version, de manière vocale via le bouton "Speech"). Ici, l'annotateur vient de noter un problème d'affordance en liaison avec la fonction miroir du vidéophone. Toute annotation peut être éliminée ou éditée.

L'interface du magicien

Le magicien a deux tâches essentielles : vérifier la correction de la phrase énoncée par le sujet et agir en conséquence. Si l'énoncé est valide, le magicien en traduit l'effet par manipulation directe sur l'écran miniature. Dans le cas contraire, un message d'erreur doit être exprimé au sujet. Dans la version actuelle, les messages sont restitués dans une boîte de dialogue mais une version vocale est simple à réaliser. Parallèlement, le message est automatiquement enregistré par NEIMO.

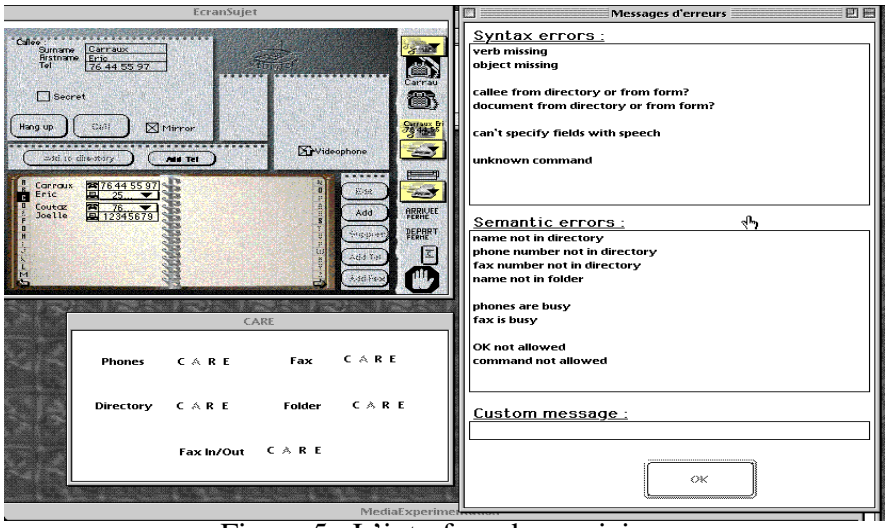

Figure 5 : L'interface du magicien.

Comme le montre la figure 5 (partie droite de l'écran), les erreurs sont organisées par catégorie (lexicale, syntaxique, sémantique) et listées selon un libellé lapidaire (par exemple "verb missing"). A ce libellé correspond, pour le sujet, un texte explicatif "amical".

L'interface de l'interlocuteur

Dans l'exemple de Supratel, l'interlocuteur gère deux lignes de communication vidéophoniques avec le sujet et a la charge de lui envoyer des fax et d'en recevoir. La figure 6 montre les fenêtres qui correspondent chacune à une tâche simulée de télécommunication.

Dans cette configuration, l'interlocuteur ne dispose pas de la reproduction miniature de l'écran du sujet. Des contraintes techniques de performance ont imposé ce choix : la station de l'interlocuteur est lourdement chargée par la vidéophonie. Cette absence de retour d'information a pu être compensée avec l'écran analogique complété par la participation active des autres expérimentateurs (notamment pour détecter les instants où le sujet changeait de ligne de communication).

Notre expérience avec NEIMO et son application en vraie grandeur au terminal multimodal de télécommunication multiservice appelle quelques remarques sur le support apporté aux expérimentateurs.

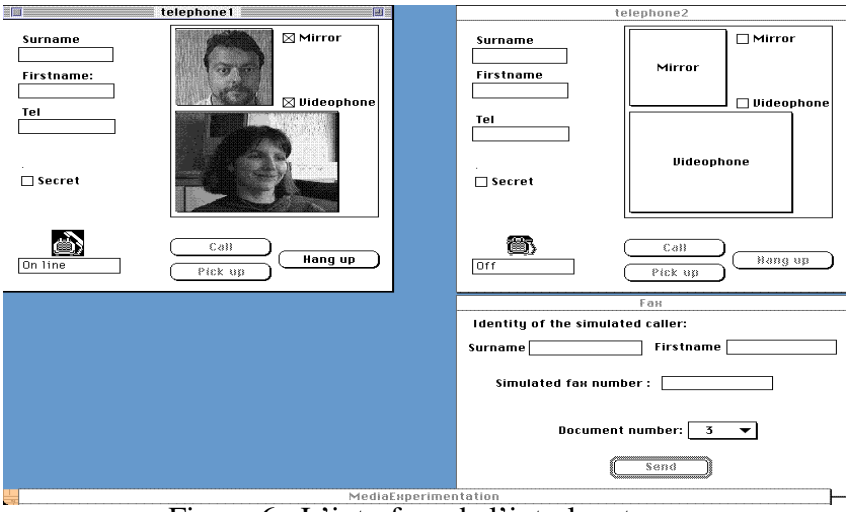

Figure 6 : L'interface de l'interlocuteur.

ANALYSE CRITIQUE

Il convient de confirmer des résultats connus mais aussi d'apporter des éléments nouveaux applicables, nous le pensons, à d'autres environnements de test d'utilisabilité.

En expérimentation contrôlée, les scénarios doivent être mis au point par des séances d'entraînement significatives (deux jours, dans notre cas pour 8 expérimentateurs différents et une durée d'expérimentation de 2 semaines consécutives). Une fois au point, les scénarios doivent être clairement énoncés tant pour le sujet que pour les expérimentateurs [27].

Les expérimentateurs chargés des tâches de simulation doivent être parfaitement entraînés et concentrés. Nous avons noté que, suite à des coupures d'une journée, il était bon de faire une session à blanc pour retrouver les bons réflexes d'une coopération efficace (comme la première crêpe que l'on jette pour avoir une poêle en parfaite condition !).

Un sens collectif émerge naturellement entre les expérimentateurs. Le groupe s'invente une gestuelle de communication lorsqu'il convient de respecter le silence ; les expérimentateurs libres aident celui qui est au fourneau. Dans notre cas, nous n'avions pas désigné de supercompère : ce rôle est apparu naturellement de manière implicite à la faveur de l'expérimentation.

On cite volontiers la nécessité pour le magicien parole d'adopter un comportement cohérent dans ses "capacités de traduction". Plus nouveau, est le cas de l'expérimentateur en contact direct avec le sujet comme l'interlocuteur de Supratel. Il est certain que ces expérimentateurs doivent avoir des aptitudes d'acteur : garder le sérieux pendant la conversation (certains sujets sont tyranniques!), remettre subtilement le sujet sur les rails du scénario et éviter ses pièges (le sujet: "Bon, ben, ça marche le multimodal, hein?" ; un interlocuteur averti: "Pardon, vous dites ? C'est à quel sujet ?")

Le problème de la charge cognitive est un fait connu. Nous pensons que la répartition dynamique de cette charge peut être favorisée par une topologie étudiée des postes d'observation. Dans notre cas, et comme le montre la figure 2, les trois stations sont rapprochées et alignées côte à côte. Au centre, l'expérimentateur qui a la charge la plus lourde. (Pour Supratel, il s'agit du magicien.) Son encadrement par deux collègues facilite ses prises de décision en cas d'hésitation (voire de panique!). Cette possibilité d'assistance humaine non perceptible par le sujet est essentielle.

L'usage de moniteurs vidéo analogiques est habituel en test d'utilisabilité. Nous en confirmons l'utilité comme référentiel commun et comme palliatif aux insuffisances du support numérique. Dans notre exemple :

- la transmission du réseau Ethernet "hoquette" parfois et nuit à l'exigence temps réel de l'expérience. Alors, l'écran miniature du sujet affiché sur les postes d'observation est légèrement désynchronisé avec la réalité. En "cas de crise", le recalage sur le moniteur vidéo est indispensable.

- Par erreur de conception de l'IHM, un élément important n'est pas observable ou difficilement détectable sur les stations des expérimentateurs. Le support vidéo analogique permet de compenser ces lacunes. Dans notre cas, l'enfoncement par le sujet du bouton dédié au signal de début de parole n'est pas assez perceptible sur la reproduction miniature mais bien visible sur l'écran analogique. Grâce au référentiel vidéo commun, un expérimentateur peut, s'il en est besoin, aider le magicien à entrer en action ou encore, nous l'avons vu dans la section précédente, aider l'interlocuteur à détecter un changement de ligne de communication.

La redondance et la complémentarité informationnelles entre les supports numériques et analogiques sont donc essentielles mais il convient d'intégrer ces deux supports sans provoquer de discontinuité visuelle. Une incrustation vidéo sur chaque poste d'observation réduit la discontinuité au risque de perdre le référentiel commun. Nous pensons qu'il doit être maintenu en redondance tant les expérimentateurs le désignent en commun pour des analyses à la volée.

Les capacités de configuration de la plate-forme d'expérimentation sont, à l'évidence, un facteur important. Par exemple, l'organisation et la liste des messages d'erreurs que le magicien destine au sujet est une source récurrente de discussions. Certains magiciens préfèrent une classification, non pas par type, mais par fréquence d'occurrence. Dans NEIMO, cette personnalisation est possible sans programmer mais n'est pas aisée pour un non expert de notre environnement. La même remarque tient pour les prises d'annotation.

Enfin, nous recommandons au moins une "méta-caméra" tournée vers les expérimentateurs pour auto-évaluer le travail d'équipe et ajuster les termes de l'expérimentation. En ce qui nous concerne, la caméra nous a été utile comme témoin en vue d'améliorer le support informatique des expérimentateurs. Dans la version actuelle, les postes d'observation ne capturent pas toutes les actions des expérimentateurs.

AGENDA DE RECHERCHE

L'évaluation expérimentale n'est pas, peu s'en faut, un domaine opérationnel, réutilisable et abouti. Les méthodes et les outils sont encore assemblés à façon. Il conviendrait de viser davantage de généricité sans perdre les avantages de la souplesse. A la lumière de notre expérience avec la maquette NEIMO, nous tenons à souligner les points suivants.

La littérature et nos contacts avec des spécialistes avertis indiquent que toute méthode d'évaluation expérimentale est nécessairement adaptée au contexte (type de produit, étape dans le processus de développement, environnement culturel). Pour un support comme NEIMO, il convient donc d'étudier l'équilibre entre l'excès de méthode intégrée et l'absence de structuration à la conduite expérimentale. Si l'outil doit être instanciable, visons la définition d'un *métalaboratoire* d'utilisabilité numérique. De la sorte, nous pourrions satisfaire les méthodes structurées inspirées du Génie Logiciel tout comme les expérimentations exploratoires.

En l'état, nous l'avons vu, la capture numérique doit être complétée par l'enregistrement analogique. L'avantage du numérique est le développement potentiel d'outils d'aide à l'analyse. La conversion numérique a posteriori des bandes vidéo (c'est-à-dire entre la

phase d'expérimentation et l'analyse) est aujourd'hui praticable. Mais l'interprétation du contenu sémantique de ces vidéos reste purement manuelle. Avec les avancées de la vision par ordinateur, on peut envisager à moyen terme l'extraction de traits pertinents comme la reconnaissance de l'expression faciale [26, 7]. Il en va de même pour l'interprétation des méta-commentaires oraux. Pour ce type d'application "off line", le traitement en temps réel n'est plus un critère essentiel ce qui permettrait, en vision comme en reconnaissance de la parole, de développer des algorithmes robustes adaptés à notre problème d'extraction de comportement. Dans le cas de notre domaine d'étude, la détection de l'usage redondant ou complémentaire de modalités est aujourd'hui à la charge de l'annotateur. Désormais, connaissant le contenu d'un énoncé oral, le système devient capable comme l'atteste l'algorithme de fusion développé par Nigay [20] de détecter le type de composition de modalités. Ainsi, peut-on migrer certaines tâches d'annotation vers l'outil.

Le dernier point que nous souhaiterions souligner est la généricité d'un outil comme NEIMO vis-à-vis de la variété des interfaces observables. Aujourd'hui, le branchement de l'interface à tester sur un outil de capture est une affaire de programmeurs. Il faudrait que demain cette activité soit transparente. A cette fin, nous suggérons aux développeurs d'outils de réalisation d'interface (boîtes à outils, squelettes d'application, générateurs) d'intégrer sur demande mais automatiquement, une instrumentation à différents niveaux d'abstraction. En complément, il conviendrait de définir un format normalisé des fichiers de capture de façon à les brancher ensuite sur différents outils d'analyse. Nous y travaillons.

REMERCIEMENTS

Cette recherche a reçu les soutiens financiers du projet ESPRIT BR Amodeus EU 7040, de CNET-France Télécom et de l'Université Joseph Fourier.

RÉFÉRENCES

[1] L. CATINIS , J. CAELEN, Analyse du comportement multimodal de l'usager humain dans une tâche de dessin. *Actes IHM'95*, Cepadues Eds: Toulouse, 1995, p. 123-129.

[2] J. COUTAZ, S. BALBO, Evaluation des interfaces utilisateurs : taxonomie et recommandations.*Actes IHM'94*, 1994, p. 211-218

[3] J. COUTAZ, L. NIGAY, D. SALBER, A. BLANDFORD, J. MAY, R. YOUNG, R.M., Four Easy Pieces for Assessing the Usability of Multimodal Interaction: The CARE Properties. *INTERACT'95*, Nordby, K., Helmersen, P. H., Gilmore, D., & Arnesen, S., (Eds). Chapman & Hall, 1995, p. 115-120.

[4] N. DAHLBÄCK, A. JÖNSSON AND L. AHRENBERG, Wizard of Oz studies — why and how. *Third Conference on Applied Natural Language Processing*, Trento, Italy, 31 march—3 April, 1992.

[5] D. DIAPER, The Wizard's Apprentice: A Program to Help Analyse Natural Language Dialogues. *People and Computers V, 5th Conference of the British Computer Society*, 1989.

[6] L. DYBKJAER, H. DYBKJAER, *Wizard of Oz Experiments in the development of the Dialogue Model for P1*. Rapport CCI, Roskilde University, Denmark, ISBN 87-7349-191-8, 1993.

[7] I. ESSA, T. DARRELL, A. PENTLAND, *Tracking Facial Motion.* MIT Media Laboratory Perceptual Computing Section. Technical Report no 272, 1994.

[8] J.-M. FRANCONY, E. KUIJPERS AND Y. POLITY, Towards a methodology for Wizard of Oz experiments. *Third Conference on Applied Natural Language Processing*, Trento, Italy, 31 march—3 April, 1992.

[9] N. FRASER, N. GILBERT AND C. MCDERMID, The Value of Simulation Data. *Third Conference on Applied Natural Language Processing*, Trento, Italy, 31 march—3 April, 1992.

[10] M.L. HAMMONTREE, J.J. HENDRICKSON, B.W. HENSLEY, Integrated Data Capture and Analysis Tools for Research and Testing on Graphical User Interfaces. *Actes CHI'92 Conference Proceedings*, ACM Press Publ., 1992, p. 431-432.
[11] D. HIX & R. HARTSON, *Developing User Interfaces, Ensuring Usability through Product & Process*. Wiley, 381 pages, 1993.
[12] A. JÖNSSON AND N. DAHLBÄCK, Talking to A Computer is Not Like Talking To Your Best Friend. *Proceedings of the Scandinavian Conference on Artificial Intelligence '88*, p. 53-68.
[13] J. LÖWGREN AND T.NORDQVIST, A Knowledge-Based Tool for User Interface Evaluation and its Integration in a UIMS. *Proc.Human-Computer Interaction—INTERACT '90*, p. 395-400.
[14] J. D. MACKINLAY, G. G. ROBERTSON, S. K. CARD, The Perspective Wall : detail and context smoothly integrated. *Proceedings of CHI'91 Human Factors in Computing Systems*, Addison Wesley, 1991, p. 173-179.
[15] M. MACLEOD, DRUM, *Diagnostic Recorder for Usability Measurement*. NPL, DITC HCI group, Teddington, Middlesex, TW11 OLW, UK, 1993
[16] D. MAULSBY, S. GREENBERG, R. MANDER, Prototyping an Intelligent Agent through Wizard of Oz. *InterCHI'93 Proceedings*, p. 277-284.
[17] C. MIGNOT, C. VALOT, N. CARBONELL, An experimental study of future "natural" multimodal human-computer interaction. *In InterCHI'93 Companion Proceedings*, ACM Publ., Amsterdam, 1993, p. 67-68.
[18] C. MIGNOT, *Usage de la parole et du geste dans les interfaces multimodales - Etude expérimentale et modélisation*. Thèse de doctorat de l'Université Henri Poincaré, Nancy I, juillet 1995, 229 pages.
[19] A. NEAL, R. SIMONS. PLAYBACK, A Method for Evaluating the Usability of Software and its documentation. *CHI'83 Proceedings*, ACM Publ., 1983, p. 78-82.
[20] L. NIGAY, *Conception et modélisation logicielles des systèmes interactifs : application aux interfaces multimodales*. Thèse de doctorat Informatique préparée au Laboratoire de Génie Informatique (IMAG), Université Joseph Fourier, 28 janvier 1994, 315 p.
[21] Y. POLITY, J.-M. FRANCONY, R. PALERMITI, P. FALZON, S. KAZMA, *Recueil de dialogues homme-machine en langue naturelle écrite*. Les Cahiers du CRISS, n° 17, 1990.
[22] M. RICHARDS & K. UNDERWOOD, How Should People and Computers Speak to Each Other. *Proceedings of Interact '84 — First IFIP Conference on Human-Computer Interaction*, p. 268-273.
[23] D. SALBER, *De l'interaction homme-machine individuelle aux systèmes multi-utilisateurs. L'exemple de la Communication Homme-Homme-Médiatisée*. Thèse de doctorat Informatique préparée au Laboratoire de Génie Informatique (IMAG), Université Joseph Fourier, 8 septembre 1995, 305 pages.
[24] A. SELLEN, G.P. KURTENBACH, W. BUXTON, The prevention of Mode error through Sensory Feedback. *In Human Computer Interaction*, Lawrence Erlbaum Publ., 7(2), 1992, p. 141-164.
[25] C.J. THEAKER, R. PHILLIPS, T.M.E. FROST, W.R. LOVE. HIMS, A Toll for HCI Evaluations. *In HCI'89, People and Computer V*, A. Sutcliffe & L. Macaulay Eds., Cambridge University Press Publ., 1989, p. 428-439.
[26] M. TURK AND A. PENTLAND, Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, p. 71-86, 1991.
[27] A. VALENTIN, G. VALLERY & R. LUCONGSANG, L'évaluation ergonomique des logiciels, une démarche itérative de conception. *ANACT*, Collection Outils et Méthode, 1993, 151 p.
[28] C. VALOT, *Interface Multimodale*. Projet Grenoblois, Rapport d'activité du Pôle Interfaces Homme-Machine Multimodales, 1993.
[29] S. WHITTAKER AND P. STENTON, User Studies and the Design of Natural Language Systems. *Fourth Conference of the European Chapter of the ACL proceedings*, 1989.