TAPAGE : un éditeur de TAbleau par la PArole et le GEste

E. Anquetil*, S. Bercu*, B. Delyon*, C. Faure, L. Julia**, G. Ménier*, R. Meneu*, G. Lorette*, J. Lemoine***, H. Oulhadj***, F. Poirier**, S. Rossignol**, H. Wehbi*****

* IRISA/URA CNRS 227
Campus de Beaulieu
Avenue du Général Leclerc
35042 Rennes Cedex
lorette@irisa.fr

** URA CNRS 820
Télécom Paris-SIG
46 rue Barrault
75634 Paris cedex 13
cfaure@sig.enst.fr

*** LERISS
Université Paris XII Val de Marne
Avenue du Général de Gaulle
94010 Créteil Cedex
lemoine@p12vx1.dnet.circe.fr

Introduction

Ce projet porte sur la réalisation d'une interface multimodale pour la conception de documents graphiques. Les modalités impliquées dans cette interface sont : le geste 2D, la parole et la vision.
La tâche envisagée est l'élaboration de documents graphiques de manière incrémentale, c'est-à-dire en plusieurs phases impliquant des opérations de production et de correction ou mise à jour. Le document résultant pouvant être intégré à des documents préexistant, en particulier des documents textuels produits sur des éditeurs de textes classiques. Les différentes opérations impliquées dans cette tâche sont effectuées à l'aide d'une interface qui exploite les moyens naturels de la communication humaine (soit la parole, le geste graphique et leur combinaison). La réalisation d'un démonstrateur permet de dégager les premiers éléments d'une modélisation de l'interaction multimodale, de rendre compte de l'existence et du caractère opératoire de méthodes en architecture logicielle et en traitement des données graphiques, ainsi que de tester en situation d'utilisation le bien fondé d'une démarche visant à transférer les modes de communication humains à l'interaction homme-machine.

Le parti pris de produire des documents graphiques à l'aide d'un stylo a pour but de recréer une situation papier-stylo où les idées sont rapidement exprimées sous leur forme visuelle par des moyens naturellement disponibles. Les données graphiques sont produites à l'aide d'un stylo sur un papier réactif et constituent leur version encre (électronique). Les ordinateurs à stylo permettent aujourd'hui de pouvoir exploiter les avantages de la production directe (manuscrite) de données graphiques telles que l'écriture ou le dessin et les avantages offerts par un traitement automatique de ces données pour transformer la version encre initiale en une version propre à la publication, à l'archivage intelligent du document, ou au travail collectif. La reconnaissance et l'interprétation des formes vont transcrire l'écriture manuscrite en caractères ASCII, reconstruire "idéalement" les dessins et déclencher des actions par des commandes gestuelles. La convivialité des interfaces utilisant le stylo en entrée repose sur la préférence d'une gestuelle éprouvée dans le monde papier aux apprentissage et utilisation de logiciels graphiques. Le clavier reste un outil particulièrement bien adapté à la production de textes longs, mais dans le cadre de certains domaines d'activités, en particulier la production de textes courts ou de symboles dans des documents graphiques ou la correction de documents imprimés, l'entrée manuscrite permet l'insertion directe sur la surface de travail, c'est-à-dire de positionner et d'exprimer une information simultanément. La convivialité de ces interfaces nécessite que la reconnaissance des données graphiques soit fiable, que le scripteur ou le dessinateur ne soit pas soumis à des contraintes et que les limitations technologiques (erreurs de reconnaissance essentiellement) soient prises en compte dans les protocoles d'interaction. L'interprétation des tableaux et la reconnaissance de certaines commandes gestuelles sont décrites en partie I. La partie II développe la reconnaissance de l'écriture. En tant qu'une forme d'expression de la langue, l'écriture devrait recevoir autant d'attention que la parole. Le rapport insiste sur les différentes formes que peut prendre l'écriture.

La complémentarité parole-geste a souvent été mise en évidence, nous soulignerons l'importance de la vision comme modalité de retour ou comme modalité implicite au moment du choix des commandes. Les opérations qui permettent de produire un document graphique (saisie, corrections par effacement, déplacement, changement d'aspect et ajout) font l'objet de commandes multimodales dont le protocole permet de réaliser une combinaison synergique du geste de pointage

et de la parole pour les deictiques et une substitution entre ces modes de communications. L'interface et l'application se déterminent mutuellement, surtout dans notre cas où les signaux qui produisent les commandes et les données de l'application sont de même nature. De plus, sur un plan ergonomique, la nature de la tâche, où des activités, doit trouver sa trace dans les protocoles d'interaction proposés à l'utilisateur (en entrée et en sortie). L'architecture logicielle adoptée pour réaliser TAPAGE est à base d'agents spécialisés qui simulent un parallélisme fonctionnel. Le protocole d'interaction et l'architecture sont décrits en partie I.

PARTIE I

Une interface multimodale pour la conception incrémentale de documents graphiques

C. Faure, L. Julia, F. Poirier, S. Rossignol

1. Introduction

Les dessins sont produits à l'aide d'un stylo sous une forme que l'on peut qualifier de "brouillon". Ils sont traitées afin de faire apparaître automatiquement leur forme "idéalisée" c'est-à-dire reconstruite dans un style qui est celui de documents publiables. Cette conception de document est de type incrémental : l'utilisateur collabore avec la machine pour obtenir un dessin final qui s'élabore au cours d'un processus intégrant la production graphique, la vision et la pensée créatrice. Ceci implique que le système dispose, en plus des traitements des données graphiques du type reconnaissance de formes, d'une interface élaborée et de procédures spécifiques pour permettre à l'utilisateur d'avoir un outil d'aide à la pensée dans sa tâche de conception.

La production de dessin au stylo recrait une situation déjà connue par le dessinateur. L'interface se devait de poursuivre dans l'esprit d'une exploitation maximale des avantages que présente l'utilisation des signaux de la communication humaine pour l'interaction homme-machine en intégrant plusieurs canaux de communication : le geste 2D, la parole et la vision. La nature de l'application conditionne le choix du matériel utilisé : la saisie graphique se fait sur un ordinateur à stylo (NotePad 3130 de NCR) auquel est associée une carte de reconnaissance de parole (DATAVOX de VECSYS).

La partie I introduit d'abord les objectifs de l'étude, les modalités impliquées dans l'interface développée et une description des tâches menées par l'utilisateur. Il présente ensuite les étapes de la réalisation de cette partie du projet, les choix qui ont été faits pour l'architecture et les protocoles d'interaction, les méthodes de traitement des données graphiques. Il se termine par une discussion sur le bilan et les perspectives de cette étude. On trouvera dans [Julia, 1992 ; Faure & Julia, 1992, Faure & Julia, 1993 ; Rossignol, 1993 ; Julia & Faure, 1993 ; Poirier & al, 1993] des présentations complémentaires de TAPAGE.

2. Objectifs

Le projet a été mené dans le but d'étudier les différents facteurs qui participent à la conception de ce type d'interface. Il a conduit à la réalisation du démonstrateur TAPAGE qui illustre, à propos de la conception incrémentale de tableaux, la réalisation des premiers objectifs :

- l'interaction synergique de la parole et du geste 2D,
- la possibilité pour l'utilisateur de choisir à chaque instant ses modes d'interaction,
- la métaphore du papier-stylo augmenté des capacités d'interprétation de la machine,
- la rapidité et les performances des traitements des tracés graphiques,
- la définition d'une architecture adaptée à ce type d'interface.

Ce démonstrateur a aussi pour fonction d'être un outil expérimental pour étudier en situation d'interaction homme-machine les comportements du système et des utilisateurs, non pas seulement

pour évaluer les performances de TAPAGE mais plus fondamentalement pour établir des modèles pour la conception d'interfaces utilisant les signaux de la communication humaine.

3. Les problèmes

3.1. La multimodalité : geste, parole, vision

L'acteur humain communicant dispose de plusieurs canaux pour émettre et recevoir de l'information qu'il va combiner notamment pour économiser du temps, des efforts de constructions syntaxiques ou de recherche de vocabulaires précis, ou pour gagner en précision. L'idée qui sous-tend l'utilisation des signaux de la communication humaine pour l'interaction homme-machine est de faire glisser les avantages de la communication multicanaux à l'interaction homme-machine. Outre que l'humain ne se comporte pas spontanément avec les machines comme avec ses congénères, les limitations technologiques viennent contraindre et perturber ces modes naturels de transmission d'information. Néanmoins, l'évolution technologique, tant au niveau des performances de reconnaissance des signaux qu'au niveau du matériel, permet d'envisager l'interaction multimodale comme une question actuelle pour les interfaces homme-machine.

Dans cette étude, le geste enregistré est ramené à la trace d'un contact entre l'extrémité d'un stylo et un papier électronique. Ce type de geste, qualifié de 2D, permet de transmettre des informations linguistiques (par l'écriture), graphiques (par le dessin), et spatiales (par des pointages). Nous avons développé ce thème dans [Faure & Julia, 1993]. La qualité des interfaces multimodales repose sur de nombreux facteurs dont celui relatif aux performances de la machine pour comprendre les signaux qu'elle reçoit. Ce projet inclus des études sur la lecture automatique de l'écriture manuscrite [Bennacer & al, Bercu & al, Ménier & Lorette, 1992] décrites par ailleurs et l'interprétation du dessin qui fait apparaître la nécessité de combiner des méthodes de reconnaissance et des modèles d'interprétation pour la communication visuelle.

La littérature insiste sur les commandes gestuelles soit pour en souligner les avantages par rapport à des commandes textuelles [Morrel-Samuels, 1990, Kurtenbach & Hulteen, 1990], soit pour montrer l'intérêt d'une synergie des gestes et de la parole qui s'illustre par des commandes de type "Mets ça là" [Bellik & Teil, 1992]. A la combinaison synergique des modes, nous avons ajouté une multimodalité de type substitutif qui permet à l'utilisateur de disposer de plusieurs modalités d'interaction (la parole ou le geste de pointage) pour obtenir les mêmes effets.

La complémentarité du geste et de la parole implique un autre mode de communication qu'est la vision. Il est évident que le geste reçu est avant tout saisi comme une image 3D ou celle de sa trace 2D sur un support. Pour ce qui est de la production, le geste, la parole ou la combinaison geste-parole sont fortement liés au contexte visuel qui caractérise la situation de production. Nous avons illustré la contextualisation des commandes sur des exemples de désignation d'objets dans [Faure & Julia, 1993].

3.2. Les tâches

La tâche principale est une conception de dessins avec ou sans texte. On insiste ici sur la différence entre la copie d'un dessin (préexistant sous sa forme définitive sur un autre document ou mentalement) et la conception qui s'élabore au cours de la production gestuelle, en plusieurs étapes, d'où sa qualification d'incrémentale.

Dans une telle situation, l'interface joue un rôle prépondérant pour assurer à l'utilisateur une sensation de collaboration effective avec son partenaire artificiel dont le rôle ne se réduit pas à traiter des données saisies en une fois et à restituer le résultat du traitement. Il doit pouvoir d'une part gérer l'évolution temporelle du dessin et d'autre part fournir les moyens pour le faire évoluer. On peut considérer que les procédures de mise à jour que nécessite l'évolution temporelle des données font parties de l'application et que les moyens dont dispose l'utilisateur pour réaliser cette évolution font parties de l'interface. Ces moyens sont des contextes d'interaction qui peuvent correspondre à des opérations d'évolution rendues disponibles et/ou à des états antérieurs des données (la version brouillon par exemple). Les opérations permettent des suppressions, des transformations (par déplacement ou changement d'aspect) et des insertions.

Il faut noter que ces opérations d'évolution sont aussi utilisables lorsque l'interprétation du dessin retournée par la machine ne satisfait pas l'utilisateur. Quelle que soit les performances des traitements graphiques, ils ne peuvent pas répondre aux attentes de l'utilisateur dans toutes les situations, d'autant plus que les données graphiques produites au stylo (que ce soit l'écriture ou le dessin) conduisent à des cas d'illisibilité (même pour un interpréteur humain) dont il vaut mieux tenir compte au moment de la conception, c'est-à-dire qu'il faut intégrer dans le protocole d'interaction les limitations technologiques et les ambiguïtés toujours possibles. On a donc choisi de fusionner dans un même protocole les corrections que l'utilisateur apporte aux erreurs d'interprétation et celles qui sont de l'ordre de sa propre démarche de pensée.

Il semble difficile d'admettre, sinon pour des raisons de présentation, une séparation entre application et interface dans la conception d'une IHM. La nature de la tâche introduite ci-dessus montre à quel point elles se déterminent l'une l'autre : les procédures de l'application et de l'interaction sont liées (voire confondues), les contextes d'interaction et les opérations qui sont effectuées sont en partie relatifs à des situations créées par l'application (erreurs d'interprétation, visualisation des données évolutives). Mais un autre facteur nous pousse à lier application et interaction, c'est le fait que l'utilisateur mène ces deux tâches simultanément : l'interaction étant le moyen de réaliser la tâche principale. L'influence réciproque de ces deux tâches est discutée dans [Faure & Arnold, 1993].

Nous retenons l'importance d'alléger la charge cognitive impartie à l'interaction et de ne pas créer de rupture dans le processus de conception de l'utilisateur. L'utilisation des signaux de la communication humaine va favoriser les comportements spontanés, les automatismes dans l'interaction et éviter à l'utilisateur l'élaboration de commandes dans un langage artificiel. La multimodalité de type substitutif qui permet à l'utilisateur de préférer un mode d'interaction à un autre et de faire ce choix de manière opportuniste conduit à adapter l'interaction à la situation, par exemple en maintenant son attention visuelle sur le dessin par l'usage de la parole ou en restant dans le mode gestuel, homogène du point de vue des organes sollicités, pour le dessin et les commandes.

La vitesse des traitements et des actions est un facteur important, il évite la chute d'attention qu'introduirait une situation d'attente. Le manque de familiarité avec l'interface vient perturber cet idéal d'interaction ainsi que les limitations technologiques qui nécessitent de contraindre la spontanéité de l'utilisateur par des vocabulaires (de gestes ou de mots) et des syntaxes qui ne sont pas analogues à ceux de la communication humaine. Néanmoins, c'est cet effet de "naturel" qui justifie l'utilisation des signaux de la communication humaine pour l'interaction avec entre autre argument celui de limiter les effets perturbateurs de l'interaction sur la tâche principale.

4. Les étapes de l'étude

1. Définition globale du cadre et des objectifs de l'étude tels qu'ils ont été énoncés ci-dessus.
2. Création de l'environnement informatique permettant la réalisation d'une interface intégrant la parole et le geste 2D. Il faut préciser que les ordinateurs à stylo étaient quasiment introuvables quand nous avons démarré l'étude et qu'ils ne sont pas prévus pour faire du développement. Le "détournement" d'un ordinateur à stylo en vue de notre étude, quoique ne relevant pas d'une recherche fondamentale, a nécessité des compétences en informatiques et beaucoup de temps.
3. Premier interpréteur des données graphiques.
4. Intégration de la parole et du geste : définition et réalisation d'une architecture multi-agents.
5. Exploitation de l'architecture pour définir des protocoles d'interaction multimodaux pour une tâche de coopération homme-machine.
6. Intégration des commandes gestuelles pour manipuler les données graphiques.
7. Tests d'évaluation en vue d'améliorer ou de corriger l'interface et l'interpréteur graphique.
8. Intégration de la reconnaissance de l'écriture manuscrite (en cours, cf. le texte dans ce même rapport sur la reconnaissance de l'écriture). Actuellement la reconnaissance est effectuée par le système attaché à l'ordinateur à stylo et nécessite un geste de commande pour faire apparaître une fenêtre d'écriture gênante pour l'utilisateur qui demande à écrire directement sur la page qui contient le dessin. De plus elle est monoscripteur et n'opère que sur des caractères séparés.

5. La réalisation de TAPAGE

5. 1. Protocole d'interaction

Les protocoles d'interaction permettent de combiner de manière synergique le geste et la parole et d'établir une équivalence entre des signaux issus de divers capteurs au niveau de l'interprétation des commandes. Nous appelons unité d'information (UI) un signal reçu ou émis par un canal de transmission. Un événement est construit à partir d'une seule UI ou de plusieurs. Les événements sont des données ou des commandes. Les données sont les dessins produits au stylo ou l'écriture. La machine peut réagir après interprétation des commandes par des actions réflexes à la réception des UI. Les commandes sont des événements interprétés qui obéissent à une syntaxe de type :

Verbe [Objet …] [Variable], soit $\mathbf{<V_1\ O^*\ V_2>}$

Le verbe V_1 est le seul élément indispensable à ce triplet, c'est lui qui décide de la nécessité des deux autres éléments et qui fournit le noyau de la commande.

La figure 1 décrit la forme générale du protocole de commande.

Figure 1

L'instantiation des éléments composant une commande peut être réalisée à partir des informations correspondant à des signaux saisis par le stylo et/ou par le micro. La forme figée de la syntaxe est reportée au niveau le plus abstrait qui est indépendant de la nature physique des UI qui servent à instantier les commandes.

On distingue deux types de commandes, les commandes d'état et les commandes d'action. Les commandes d'état ne sont formées que d'un verbe (par exemple : /Dessine/) et agissent sur l'environnement de l'application en spécifiant à l'utilisateur le mode dans lequel il travaille. Dans le domaine de la conception de documents graphiques, les signaux de données (le dessin ou l'écriture) correspondent à des traces sur le papier réactif qu'il faut conserver (sous leur forme initiale ou idéalisée) alors que les signaux de commandes produits au stylo doivent disparaître après réalisation de l'action commandée. Le fait que le même médium puisse fournir des données et des commandes soulève le problème de leur catégorisation. A cela s'ajoute des ambiguïtés d'interprétation de symboles qui présente une identité de forme (cercle et zéro par exemple).

Les commandes d'états permettent de discriminer a priori les signaux de commande, le dessin et l'écriture. Les états correspondent aux modes "correction", "dessin", "écrit". Ils constituent des contextes d'interaction dans lequel l'utilisateur se place en pointant le bouton correspondant à cet état ou en prononçant son nom. Une fois le contexte sélectionné, l'environnement s'adapte en faisant apparaître des menus spécifiques et (de manière transparente pour l'utilisateur) en désactivant des

procédures relatives aux autres contextes. Les commandes d'état sont aussi responsables du lancement des procédures dans l'application. Chacune de ces commandes a un rôle qui est décrit par un scénario d'actions qui vont transformer les représentations internes et les informations affichées.

Les commandes d'action permettent d'agir sur les objets de l'application. On y spécifie O^* qui est le complément d'objet direct du verbe (/Efface **ça**/) et éventuellement V_2 qui peut être une position (/Mets ça **là**/) ou un attribut (/Colorie ce trait **en rouge**/). O^* peut désigner dans certains cas plusieurs objets : si l'on utilise une sélection par entourage ou un pointage successif avant d'appeler la commande (/ça, ça et ça Efface le/).

Par une scrutation continue des agents périphériques, l'interpréteur saisi au vol les informations qui lui permettent d'instantier les commandes, en associant à un même symbole les informations qui sont équivalentes du point de vu de la syntaxe (verbe parlé ou verbe “pointé”). Il gère la construction d'une commande en utilisant des connaissances sur les classes de verbes. Suivant la classe, la commande complète prendra la forme V_1 ou V_1 O* ou V_1 O* V_2. Quand la forme associée au verbe est complètement instantiée, la commande est syntaxiquement correcte. Le rôle dominant du verbe ne doit pas faire supposer que l'instantiation des symboles doit obligatoirement commencer par le verbe. L'utilisateur peut en effet commencer par sélectionner un ou des objets avant de fournir le verbe. Dans le cas où un objet est d'abord sélectionné suivi d'un verbe ne nécessitant pas d'objet, l'objet sélectionné est ignoré et la commande est valide.

Ce mode d'interprétation par instantiation de formes prédéfinies est combinée à une utilisation de la carte de reconnaissance vocale en détecteur de mots. Ceci permet de ne pas figer les productions parlées de l'utilisateur, /déplace ce gêneur dans le coin/ ou /mets ça là/ correspondent à la même UI : un verbe de la classe /mets/. Les productions verbales spontanées qui accompagnent une pensée ou un geste, sans toutefois vouloir signifier quelque chose à la machine, sont possibles … dans la mesure où les mots prononcés ne sont pas trop proches du vocabulaire de commande ce qui peut créer des reconnaissances parasites.

Le vocabulaire de TAPAGE est d'environ 70 mots. Ces mots ne sont généralement que des verbes. TAPAGE accepte jusqu'à 5 synonymes pour une commande, par exemple "Mets", "Déplace", "Bouge", "Amène", "Place" sont synonymes.

Chaque classe de synonymes est associé à un bouton des menus visualisés. Il est donc possible de pointer ou de nommer les éléments des menus. Il y a en permanence coexistence des menus visualisés et d'un menu parlé. Les menus visuels sont courts et spécialisés, ce qui facilite l'accès aux contenu. Le menu parlé est la concaténation de tous les menus visuels. On peut ainsi appeler n'importe quel élément des menus, en particulier ceux qui sont invisibles, sans quitter des yeux le dessin en cours de conception et sans effectuer un ou plusieurs pointages sur les menus visuels.

L'intégration de la vision dans l'interaction est sensible au niveau des retours où les éléments sélectionnés (dans les menus ou sur le dessin) sont différenciés par leur aspect visuel. Des actions réflexes déclenchent des signaux visuels à la réception d'UI pour faire savoir à l'utilisateur si son message a été bien reçu. Par exemple, les éléments sélectionnés sur le dessin s'affichent en pointillés, les encres de correction ne sont pas de la même couleur que les encres de dessin et d'écriture. D'autre part, il existe plusieurs commandes gestuelles de sélection (pointage multiple, encerclement ou surlignement d'objets spatialement proches), ce qui permet de choisir celle qui s'adapte le mieux à la répartition spatiale des objets à sélectionner.

5.2. L'architecture

Nous avons retenu une architecture à base d'agents. Actuellement la plupart des architectures d'interfaces Homme-Machine sont élaborées à partir de ce concept d'agents, particulièrement illustré par PAC [Coutaz, 1990] qui permet de dégager une vision structurée de l'architecture, tant au niveau de sa description que de sa réalisation. Ceci permet aussi d'envisager le dialogue multimodal en terme d'autonomie et de coopération de ces agents et par suite d'optimiser la répartition du travail entre agents en distribuant certaines fonctions.

Les agents sont organisés de manière hétérarchique (figure 2). On distingue un niveau périphérique, un niveau d'interprétation, un niveau de sémantisation où se fait le lien entre l'application et l'interface. A cette structure en niveaux, s'ajoute la possibilité de communication directe entre agents indépendamment de leur niveau.

Figure 2

Au niveau périphérique, les agents assurent la relation avec le monde extérieur. Cinq media, associés à des agents spécialisés, sont utilisés dans TAPAGE pour communiquer avec le monde extérieur : un microphone, un stylo, un clavier, un haut-parleur et un écran. Ils engendrent six modalités : quatre en entrée (parole, écriture, dessin et geste) et deux en sortie (son et affichage), chacune étant prise en charge par un agent spécifique, dit de présentation, qui récolte en permanence des UI en scrutant les faits et gestes de l'utilisateur. Ces UI sont les constituants des événements. Au niveau interprétation les événements sont construits, en effectuant éventuellement une fusion des modalités, et prennent la forme d'une représentation symbolique a-modale. Cette représentation est interprétée dans les termes de l'application par l'agent qui établit le lien avec l'application et active les procédures relatives aux actions spécifiés par les événements.

Les agents possèdent des propriétés qui seront exploitées dans le système. L'autonomie et la spécialisation des agents assure un parallélisme qui rend possible la simulation de la perception multi-canaux. Chaque agent périphérique ne réagit qu'aux stimuli pour lesquels il est programmé et ceci indépendamment de ce qui peut être reçu par ailleurs. C'est ainsi que peut être réalisé les deux types de multimodalité : synergique ou substitutive.

Ces propriétés d'autonomie et de spécialisation vont aussi permettre une répartition des tâches au sein des agents et des notions de distribution de certaines fonctions (comme l'interprétation ou l'affichage). Les effets les plus visibles de cette organisation du travail se manifestent par la rapidité des réactions de la machine.

Les agents périphériques possèdent une vue des représentations internes et de la surface d'affichage. Ils peuvent ainsi savoir si le signal produit par l’utilisateur les concernent en fonction de l'état de l'application (dessin ou écrit ou correction). Il est aussi possible à ce niveau de discriminer les pointages d'objets et de position, et d'identifier l'objet ou le lieu sélectionné. Cette information transitera par l'interpréteur qui ne fera que la transporter et la fournir à l'agent assurant la liaison avec l'application. L'interpréteur doit savoir qu'il s'agit d'un objet ou d'un lieu mais il n'a pas à connaître sur quel objet et en quel lieu précis l'action doit se produire, son monde est purement symbolique, sans aucune vision de l'application. Cette discrimination précoce (avant le niveau propre à l'interprétation) va faciliter l'instantiation des commandes en réduisant l'interprétation à

leur forme purement symbolique. On a alors une distribution de l'interprétation sur différents niveaux du système.

La communication directe entre agents, indépendamment de leur niveau, permet des actions réflexes suite à des événements qui ne sont pas des commandes (ceux ci doivent transiter par le niveau interprétation). Par exemple quand l'agent spécialisé dans les commandes gestuelles reçoit un ordre de sélection, il prévient l'agent d'affichage pour qu'il mette en pointillés les objets ainsi sélectionnés. Cette action se fait au niveau du dialogue inter-agents, sans perturber l'application. Ceci implique bien entendu que les agents possèdent un minimum d'informations sur les données et le contexte de l'application. De la même façon, l'application peut afficher des informations à l'écran sans recourir à l'agent d'affichage. La fonction d'affichage est donc distribuée entre l'interface et l'application et permet de meilleures performances du programme.

5. 3. Le graphique

L'idéalisation de tableaux

Les tableaux dessinés au stylo sont traités en plusieurs étapes dont nous ne rappelons ici que les grandes lignes :

- les tracés sont ramenés à des segments de direction verticale ou horizontale,
- des fenêtres d'aimantation définies à partir des coordonnées des tracés brouillon sont utilisées pour déterminer des seuils locaux de fusion des suites de segments suivant une direction verticale ou horizontale,
- un analyseur parcourt successivement les directions horizontale et verticale pour déterminer de manière récursive les colonnes et les lignes du tableau qui seront reconstruites avec des tailles égales.

Un exemple de dessin et de reconnaissance brute est donné en figure 3.

Figure 3

Lors des modifications apportées au tableau, la transformation (de suppression, d'insertion ou de déplacement) va activer une mise à jour globale de manière à respecter les dispositions conventionnelles. La figure 4 illustre une suite d'actions sur des tableaux dont les caractéristiques numériques sont légèrement différentes.

Figure 4

L'application ainsi que l'interaction gestuelle a été testée dans une tâche de recopie de 7 tableaux différents de structure plus ou moins complexe. Ce test a été fait avec un expert, deux initiés et trois novices qui n'avaient jamais utilisé ce type de matériel. Tous les sujets ont su réaliser la tâche, malgré quelques abandons et reprise à zéro. Ces abandons sont dus essentiellement à un débordement du dessin de la surface de travail (de dimension trop réduite) du papier réactif ou à des erreurs importantes dans la structure du tableau qui engage l'utilisateur dans un cycle de correction qu'il préfère arrêter en effaçant le dessin pour recommencer. L'expert est en général (mais pas toujours) le plus rapide pour réaliser la tâche. On ne note pas de différence notoire entre novices et initiés du point de vue des temps. Le temps passé à corriger (par déplacement ou effacement) peut être nul : au moins un des dessinateurs a obtenu le résultat attendu sans avoir recours aux corrections.

5.4. Les commandes gestuelles

La concision est un des avantages des commandes gestuelles. Il est possible de signaler par un seul geste la nature de l'action et l'objet (ou la position) où elle s'applique. Les commandes comme les flèches de déplacement et les ratures d'effacement illustrent cette qualité (figure 5). Nous avons donc retenu ces deux commandes. A celles-ci s'ajoutent celles qui vont permettre des désignations multiples et des désignations partielles d'objets.

La désignation multiple est assurée de différentes manières : entourage des objets à sélectionner, pointages successifs (par un tracé sous forme de point ou de trait superposé à l'objet), pointage continu (un trait de sélection se superpose à plusieurs objets). La désignation partielle est obtenue par entourage d'une partie de l'objet. Les objets de TAPAGE sont les segments verticaux ou horizontaux du tableau, la désignation partielle sélectionne la portion de l'objet correspondant à une case.

La reconnaissance de ces symboles demande un prétraitement qui polygonalise les tracés en éliminant les petites variations locales. Le tracé est représenté par une double suite temporelle de segments codés sur seize seuils de quantification de direction. Chaque suite de ce doublet correspond à une rotation des directions de quantification de $2\pi/32$. Ceci a pour effet d'éviter d'obtenir le codage d'une droite sur deux directions de quantification quand sa direction est à la limite des frontières de quantification : les deux suites sont comparées et celle qui minimise le nombre de segments du tracé est acceptée comme son représentant.

Un test de fermeture et la mémorisation des points caractéristiques des tracés (départ et fin de la flèche, coordonnées minimales et maximales) complètent les informations utiles pour la reconnaissance de la forme et de la localisation du symbole de commande.

Un test a été fait pour 20 tracés du même scripteur, le nombre de reconnaissances correctes est donné en comparant les résultats obtenus avec un seul ensemble de quantification et avec deux ensembles. La généralisation à d'autres scripteurs fait apparaître que la reconnaissance des flèches demande à être améliorée.

	1 jeu	**2 jeux**
Droite	17	20
2 Droites	16	20
Triangle	8	17
Rectangle	10	17
Flèche simple	13	16
Flèche double	17	20
Patatoïde	18	18

Figure 5

6. Bilan et perspectives

Dans le cas de TAPAGE, la question générale de la conception de documents graphiques à l'aide d'une interface multimodale demande à être décomposée pour mener une analyse susceptible de faire évoluer nos connaissances sur cette question. Cette décomposition s'appuiera d'abord sur les différents modules de traitement qui composent le système et dont certains sont des produits commercialisés, et sur les composantes du problème que nous avons voulu étudier prioritairement.

Cette analyse se compose d'un premier volet qui est relatif à la démonstration par TAPAGE d'une mise en œuvre effective des objectifs décrits dans ce rapport :

- la réalisation d'un environnement informatique hétérogène intégrant un ordinateur à stylo et une carte de reconnaissance vocale,
- la synergie des modes gestuel et parlé,
- la substitution de signaux physiquement différents dans l'instantiation des commandes,
- la variété des catégories de traces graphiques (commandes — pointage(s), signes —, dessin, écriture)
- la rapidité d'action par une combinaison des notions de distribution et de spécialisation dans l'architecture,
- la rapidité des traitements graphiques,
- la gestion d'un univers graphique évolutif et imparfait en fusionnant les contextes de correction d'erreur et de conception incrémentale.

Sur le plan pratique, il s'est avéré que la parole ne garantit pas toujours un parfait fonctionnement du système, en particulier dans des situations de démonstration en milieu bruité. Le papier électronique présente une parallaxe qui gène la sélection par pointage des objets pour un utilisateur non familier avec ce matériel et de toute façon empêche d'accéder à la précision que l'on peut avoir en travaillant sur un support papier (on peut évaluer cette imprécision sur la position désirée comme étant de l'ordre du mm).

Un deuxième volet concerne l'analyse de TAPAGE en situation d'utilisation. Des tests préliminaires et spécifiques à des composantes de TAPAGE ont été présentés au cours du rapport. Pour la suite, il devient nécessaire d'établir des scénari qui devront prendre en compte le fait que la carte vocale, tout en étant très tolérante, est mono-utilisateur de même que la reconnaissance de l'écriture actuellement implantée et que nous ne pouvons pas demander à chaque sujet de faire un apprentissage. La combinaison geste/parole dans le cas des déictiques et des choix dans les menus est un résultat acquis du point de vue du fonctionnement du système, il reste à tester l'usage de ces modes par des utilisateurs. Pour l'analyse des fonctions graphiques, il est possible de remplacer l'écriture manuscrite par un clavier connecté. Les sujets sont répartis en plusieurs catégories (des novices aux experts). Le scénario proposé porte sur une tâche de copie d'un tableau (test des traitements graphiques liés à ces données) et une tâche de correction d'un tableau déjà affiché à l'écran. Cette dernière a un double objectif : tester l'efficacité des moyens dont dispose TAPAGE pour effectuer cette tâche et acquérir des connaissances sur les comportements spontanés des sujets.

L'enregistrement des sujets à qui l'on demande de verbaliser leur actions nous indiquera quelles sont les extensions qu'il serait souhaitable d'envisager pour l'interaction par la parole.

Cette première réalisation sera étendue à la conception de diagrammes en réseaux qui constitue un langage visuel et se présentent comme des configurations spatiales de type objets (géométriques)-relations. Ce problème combine des modèles relevant de la perception et de la reconnaissance de formes ainsi que de la sémiotique visuelle. On insistera sur la notion d'organisation spatiale des objets et l'influence du contexte sur la reconstruction idéale des figures géométriques. Un travail expérimental préliminaire a été effectué [Faure, 1993]. Les traitements des données graphiques sont en cours de développement. Le protocole d'interaction, tout en reprenant celui développé pour TAPAGE, sera affiné dans le cadre de cette application.

7. Conclusion

L'évaluation d'un démonstrateur ou plus généralement d'une réalisation faite dans le cadre d'une recherche ne peut pas s'aligner sur des protocoles utilisés pour évaluer un produit ou un prototype industriel. Un démonstrateur n'a pas pour but d'illustrer la mise en œuvre d'un modèle mais de participer à sa définition. On préférera voir l'évaluation comme une analyse et non comme un calcul brut de performances. TAPAGE a atteint un niveau de réalisation qui nous fait envisager son exploitation pour faire progresser notre connaissance sur le domaine des interfaces multimodales. Il est à la fois la réalisation de base qui fournit un cadre informatique pour aborder de nouveaux domaines et l'outil expérimental qui nous permet d'expérimenter les comportements humains en situation d'interaction multimodale.

La complexité des processus de conception intégrant des constructions abstraites et des expérimentations se retrouve au moment où il s'agit d'évaluer la réalisation.

Diffusion et collaborations

Le projet a fait l'objet d'une collaboration avec l'IRISA - Université de Rennes I (G. Lorette) et le LERISS - Université de Paris-Val de Marne (J. Lemoine).

Le démonstrateur TAPAGE a été présenté lors de diverses manifestations dans ses versions successives et a fait l'objet de communications :

- Séminaire AFCET/ G.R.C.E. (Communication écrite), Paris (08.10.92).
- Congrès IHM'92, Paris (02.12.92).
- Congrès IMRV'93, Montpellier (24,25&26.03.93).
- Forum des Recherches en Informatique, Palaiseau (02&03.06.93).
- Congrès ICOHD'93, Paris (06.07.93).
- Congrès HCI'93, Orlando (12.08.93).
- Congrès IHM'93, Lyon (19.10.93).

En relation avec d'autres équipes ou d'autres projets nous avons crée à partir du NodePad une interface de saisie de caractères manuscrits en vue de constituer une base de données pour la reconnaissance et mené des études complémentaires sur la multimodalité, plus particulièrement avec le projet MUNIX [Poirier & Lefebvre, 1993] et le projet Image-LangageS du GDR Sciences Cognitives de Paris [Faure & Arnold, 1993].

Références

Bellik Y., Teil D. (1992). Les types de multimodalités. Actes des 4ème journées sur l'ingienérie des interfaces homme-machine, IHM'92, pp. 22-28.

Bennacer L., Lemoine J., Petit E. (1992). Une méthode en ligne de reconnaissance d'écriture par double balayage. Actes de CNED'92, BIGRE n° 80. pp. 333- 338.

Bercu S., Delyon B., Lorette G. (1992). Segmentation pour une méthode de reconnaissance d'écriture cursive en-ligne. Actes de CNED'92, BIGRE n° 80. pp. 144- 151.

Coutaz J. (1990). Interfaces Homme-Ordinateur. Dunod.

Faure C. (1993). From hand-sketched diagrams to their ready-to-publish version: an experimental study. ICOHD'93.

Faure C., Arnold M. (1993). L'interaction homme-machine du point de vue des principes d'économie. Actes des 5ème journées sur l'ingienérie des interfaces homme-machine, IHM'93.

Faure C., Julia L. (1992). TAPAGE : une interface pour l'aide à l'édition de tableaux par la parole et le geste. Actes des 4ème journées sur l'ingienérie des interfaces homme-machine, IHM'92.

Faure C., Julia L. (1993). Interaction homme-machine par la parole et le geste pour l'édition de documents : TAPAGE. Actes Interface des mondes réels et virtuels. Montpellier. pp. 171-180.

Julia L. (1992). Vers une interface multimodale. Rapports de DEA ITCP, Université P. & M. Curie - Paris 6. pp. 112-147

Julia L., Faure C. (1993). A multimodal interface for incremental graphic document design. Actes des posters HCI International'93, Orlando. p. 237.

Kurtenbach G., Hulteen E.A. (1990). Gestures in Human-Computer Communication. The art of human-computer interface design. Adison Wesley, pp. 309-317.

Menier G., Lorette G. (1992). Segmentation et reconnaissance en ligne d'écriture cursive à l'aide de plusieurs niveaux d'information contextuelle. Actes de CNED'92, BIGRE n° 80, pp. 318-332.

Morrel-Samuels P. (1990). Clarifying the distinction between lexical and gestural commands. International Journal of Man-Machine Studies, 32, pp. 581-590.

Poirier F., Lefebvre P. (1993). Une interface multimodale de dialogue homme-machine sous Unix. Informatique 93 - Interface des mondes réels et virtuels, Montpellier, pp. 161-170.

Poirier F., Julia L., Rossignol S., Faure C. (1993). TAPAGE : Edition de tableaux sur ordinateur à stylo vers une désignation naturelle. Actes des 5ème journées sur l'ingienérie des interfaces homme-machine, IHM'93.

Rossignol S. (1993). Reconnaissance de gestes sur un ordinateur sans clavier. Rapport de stage d'option, Ecole Polytechnique.