TAPAGE : UNE INTERFACE POUR L'AIDE A L'EDITION DE TABLEAUX PAR LA PAROLE ET LE GESTE

Claudie Faure et Luc Julia
URA CNRS 820
Télécom Paris - Dép. SIG
46 rue Barrault
75634 Paris cedex 13

Résumé : Le projet TAPAGE s'inscrit dans le cadre de la conception d'interfaces multimodales pour l'aide à l'élaboration, l'édition et la manipulation de données textuelles et graphiques. Les modes d'interaction retenus sont la parole et les gestes, ce qui implique des modèles pour l'interprétation des signaux qui supportent la communication humaine. L'interface est construite à partir d'un ordinateur à stylo. Nous présentons les modes gestuels et parlé en nous limitant à ce qui concerne ce matériel. Le système TAPAGE est décrit dans sa version actuellement opératoire, illustrée de résultats. Les données et les commandes sont engendrées par des signaux de même type (graphique, textuel) pour les applications envisagées. Ceci constitue une difficulté spécifique pour la mise au point des protocoles d'interaction.

Mots clés : interface multimodale, ordinateur à stylo, aide à la conception graphique

1. Introduction

TAPAGE est un système d'aide à l'édition de tableaux qui est conçu à partir d'un ordinateur à stylo et d'une carte de reconnaissance de parole. La multimodalité concerne donc les modes liés au stylo et la parole. L'apparition des ordinateurs à stylo s'inscrit dans une évolution de l'informatique qui tend à proposer à l'utilisateur un environnement de plus en plus convivial, où ses efforts pour apprendre à se servir d'un outil informatique et pour mémoriser des commandes sont d'autant plus réduits que les modes d'interaction homme-machine lui sont "naturels", conférant ainsi une qualité anthropomorphique à l'interface. Le projet TAPAGE à pour but d'analyser les rôles des modes gestuels et parlé dans la définition des protocoles d'interaction, et de tester sur des utilisateurs ce type d'interface à propos d'une application précise. L'application développée concerne l'édition de tableaux mais plus généralement il s'agit de concevoir des modules d'idéalisation automatique de tracés graphiques produits à main levée. Ce problème relève de la reconnaissance automatique de formes pour le traitement des tracés numériques et de la sémiotique graphique pour l'idéalisation proprement dite. Le champ applicatif visé est l'aide à l'édition de documents et à la conception graphique.

2. Les modes gestuels

Nous nous intéressons dans ce paragraphe à la communication gestuelle indirecte où le message est véhiculé par l'inscription visible que laisse la main munie d'un instrument de traçage sur la surface du papier réactif. Pour une vision plus large voir [ARGENTIN 89], [EKMAN et FRIESEN 72], [GREIMAS 70]. Nous parlons de différents modes gestuels pour insister sur la possibilité que donne le geste d'utiliser plusieurs formes d'expression :
- linguistique, par l'écriture. C'est un mode de communication trés naturel, son usage se voit pénalisé par la difficulté à reconnaître l'écriture "naturelle" (c'est-à-dire le cursif lié). Ce goulet d'étranglement des interfaces gestuelles constitue un pôle d'étude dans le projet

TAPAGE qui ne sera pas développé ici (voir : [BENNACER et al 92], [BERCU et al 92], [MENIER et LORETTE 92]),
- symbolique, on regroupe ici des signes qui ont une signification conventionnelle : les chiffres, les symboles mathématiques, chimiques, électriques, les commandes gestuelles qui apparaissent comme des signes de type typographique. Ces commandes peuvent désigner d'un seul geste l'action à effectuer, l'objet et la position concernés par l'action. C'est le cas de l'action d'effacement représentée graphiquement par un trait qui barre un objet,
- graphique où la structure spatiale va jouer un rôle prépondérant dans l'interprétation des signaux. C'est le cas des diagrammes en réseaux, des tableaux, des schématisations d'arrangements spatiaux (plans, maquettes) et aussi de certains langages visuels de programmation,
- désignation où existe un mimétisme fonctionnel du stylo et de la souris. Le clic de localisation est rendu par un posé de stylo, on peut étendre des fonctions comme "attrape", "étire", "déplace" au stylo. La localisation peut se faire directement sur les données (localiser un mot, une figure) ou sur un menu, dans ce cas le geste est un déclencheur d'action.

3. Gestes et parole

Nous nous limitons ici à la complémentarité du geste et de la parole en ne retenant que les gestes pratiques, intentionnellement produits dans le but de compléter ou de préciser un message parlé. Le choix du mode d'expression est déterminé par une recherche d'économie qui porte sur plusieurs facteurs dont : la rapidité, la précision, la simplicité. La rapidité concerne le temps d'émission (de réception) du message, la précision vise à produire des messages non-ambigus et la simplicité d'un message relève de la complexité des mécanismes nécessaires à la construction (à l'analyse) de son expression.
L'exemple typique que l'on donne souvent de la complémentarité économique du geste et de la parole est :

(C1) /met ça là/ accompagné d'une désignation spatiale du /çà/ et du /là/.

L'intention de déplacement est rendu par un verbe d'action (/met/) qui implique un objet présent dans la séquence parlé (/ça/) mais non spécifié, le démonstraitf économise un vocabulaire précis voire des constructions linguistiques, un seul geste renvoit à l'objet physique concerné. De même la position (/là/) évite de décrire verbalement une position spatiale désignée par un seul geste.

Plus généralement, l'information spatiale est difficile à traduire avec précision sur le plan linguistique. Le geste est plus efficace pour exprimer des caractéristiques spatiales quantitatives comme la position d'un point dans un espace ou la taille d'un objet (dans une commande comme /agrandir/).

Le meilleur choix des modes, au sens d'une économie, est toujours contextuel. On introduit une discussion à propos de la désignation d'objets. Dans un image graphique composée d'un ensemble de figures géométriques on peut vouloir effacer une ou plusieurs de ces figures. On considère une suite d'exemples de commandes :
(C2) /efface çà/ et geste ; (C3) /efface les carrés/ ; (C4) /efface les figures de gauche/ ; (C5) /efface les figures/ et gestes.

(C2), conforménent à (C1) permet d'effacer un objet désigné par un geste. L'économie est du type de celle décrite pour (C1). (C3) permet d'effacer les carrés présents sur l'image, le pluriel et l'attribut "carré" spécifique des figures à éliminer évitent les gestes répétitifs de désignation qui seraient nécessaires. (C4) permet de désigner un groupement perceptif et de le traiter comme un seul objet. Dans ce cas l'expression linguistique de ce groupement n'est pas ambiguë. Pour (C5) on doit considérer que les figures à effacer ne sont pas identifiables par un attribut ou une combinaison de quelques attributs (de forme, de couleur, de taille), qu'elles sont réparties dans l'espace de telle manière qu'aucun groupement perceptif n'apparaît ou que, s'il apparaît, la verbalisation de sa position pose

problème (ni le plus à gauche, ni le plus haut ...). La nature des gestes sera fonction de la répartition des figures dans l'espace. Un groupement perceptif fait apparaître un objet virtuel localisé qu'un seul geste peut désigner par encerclement de ces composantes, dans ce cas (C2) peut aussi être utilisée pour effacer des groupes d'objets. En l'absence de cette structuration, les gestes pointeront sur chaque objet à effacer.

4. La réalisation

Le projet TAPAGE s'attache à réaliser dans un premier temps une maquette de démonstration sur une application adaptée à la technologie et à la vocation des ordinateurs à stylo. Cette maquette est évolutive, elle progresse en fonction des modules de traitement intégrés et des architectures qui définissent les protocoles d'interaction multimodale. Dans une première version, sa fonction consiste à idéaliser automatiquement des dessins manuels de tableaux, à les remplir de chiffres ou de mots, à les corriger. Les modes d'interaction sont la parole, l'écrit, le dessin, les commandes gestuelles (pointage, symboles).

L'idéalisation des tableaux

Ce module du noyau comprend plusieurs étapes de traitement que nous décrivons briévement. Les signaux se présentent sous forme de suite de points (figure 1). Dans une première étape les tracés sont segmentés de manière à trouver les ensembles de verticales (V) et d'horizontales (H) qui les approximent. Ils sont d'abord polygonalisés par une méthode de type *SPLIT and MERGE* [PAVLIDIS 82] qui les découpe en segments V et H compte tenue d'un paramètre de tolérance, puis regroupe les segments de même direction sur un critère de voisinage. Les segments obtenus sont alors redressés suivant les directions H et V. Dans une deuxième étapes, les jonctions en L et T sont détectées et reconstruites idéalement (prolongement des segments pour une mise en contact effective au niveau des points de jonction, effacement des segments parasites).

Figure 1 : Tracé à main levée d'un tableau

La troisième étape concerne un niveau de traitement qui s'appuie sur les valeurs numériques des données et sur une connaissance du domaine. Les tableaux sont des structures spatiales qui suivent les principes de la communication visuelle. Ainsi, les différences physiques portant sur un ou plusieurs attributs (taille, forme, couleur …) signalent des différences sur le plan de l'interprétation. La communication visuelle tend à éviter les différenciations qui introduiraient des informations non souhaitées ([MARKS et REITER 90]). Suivant ce principe, des groupes de colonnes, de lignes et de cases auront des tailles identiques dans un tableaux. Or, le dessin manuscrit ne produit jamais des grandeurs exactement égales. Cette égalisation résultera d'une reconstruction automatique, d'autant plus difficile à faire que les sous-ensembles de colonnes, de lignes et de cases devant être égalisées ne sont pas connus à priori. Une recherche dynamique de motifs répétitifs est effectuée en parcourant la structure du tableau par un algorithme

récursif suivant l'axe V, puis l'axe H. Les séparateurs de motifs sont fixés chaque fois qu'une séquence de motifs est détectée. Les zones de recherche et l'ordre dans lequel elles sont examinées ne sont pas définis a priori, ils dépendent des positions des séparateurs fixés aux étapes précédentes de l'analyse. La figure 2 illustre le résultat obtenu.

Figure 2 : Le tableau aprés idéalisation, désignation et remplissage d'une case

La multimodalité

TAPAGE est développé à partir du NCR 3125 Notepad qui comprend un microprocesseur 80386, 3Mo de mémoire vive, un écran VGA haute résolution et un disque dur de 20 Mo. Le stylo est sans fil, son contact avec l'écran est très proche d'un feutre sur papier, sa gestion informatique est assimilable à celle d'une souris. Les différents modes impliquent des systèmes de reconnaissance spécifiques. Il a d'abord fallu adapter la saisie des données à notre application. Quand le plein écran permet de visualiser en temps réel la trace du stylo, le système de reconnaissance de caractères (PENOS) intégré est actif. Ce module a du être désactivé (ce qui n'était pas prévu par les constructeur) pour pouvoir permettre une saisie de dessins puis leur analyse par le module décrit ci-dessus. Le système de reconnaissance de caractères peut être réactivé à la demande formulée par une commande gestuelle qui fait apparaître une zone d'écriture. La reconnaissance de la parole est faite par DATAVOX commercialisé par VECSYS. Ce système est prévu pour reconnaitre des mots connectés avec des vocabulaires de plusieurs centaines de mots et permet de définir une grammaire.

A cette étape, nous avons essentiellement testé la faisabilité de l'intégration de différents modes au niveau du matériel, l'architecture de l'interface demande des études complémentaires. Dans cette version, la voix permet de déclancher des actions (traitements, sortie du programme) désignées par des mots d'un vocabulaire de commande. Le stylo permet de désigner une case du tableau qui une fois sélectionnée change d'aspect pour conforter l'utilisateur par un retour visuel. Le stylo permet les signes des commandes gestuelles qui visualisent des "outils" comme le clavier virtuel ou la fenêtre d'écriture manuscrite qui peuvent être utilisés pour remplir la case désignée. La case remplie de la figure 2 illustre la difficulté à verbaliser certaines positions. La case peut aussi être directement remplie au clavier attaché au Notepad, sans que ce mode d'entrée soit spécifié au préalable. Les mots du vocabulaire de commande sont acceptés sous forme écrite ou parlé. On donne l'exemple d'une des séquences d'événements qui conduisent à la figure 2 :

mode :	dessin →	parlé →	gestuel →	clavier →	parlé
signaux :	tableau	/remet en forme/	pointeur	215	/sortir/
classes :	donnée	commande	commande	donnée	commande

5. Commentaires

La définition des protocoles d'interaction multimodale et de l'architecture associée est la question qui doit retenir notre attention maintenant que nous avons les moyens d'évaluer les possibilités d'intégration des matériels et leur fiabilité. Les applications envisagées ont une caractéristique importante qu'il ne s'agit pas de sous estimer lors de la définition de ces protocoles. Elle concerne l'identité des modes (ou des type de signaux) pour les données et les commandes. Nous travaillons en permanence dans le monde de la communication, soit pour envoyer des informations à la machine soit pour concevoir des documents qui portent des informations (textuelles ou graphiques). Un signal sonore, ou graphique peut être une commande, dans ce cas sa durée de vie est provisoire, il est "effacé" dés que la commande s'exécute. Il peut aussi être une donnée qui doit persister sur l'écran qui figure le document en construction, et ne peut disparaître qu'après une commande d'effacement. La discrimination des signaux en données et commandes est un problème spécifique à notre projet, très délicat à régler si l'on ne veut pas perdre la spontaneïté de l'utilisateur dans la définition du protocole d'interaction.

6. Conclusion

Cette maquette a permis de tester la faisabilité de l'intégration des modes gestuels et parlé dans une interface réalisée à partir d'un ordinateur à stylo et d'une carte de reconnaissance vocale. Cette phase "d'intégration informatique" réussie ainsi que la réalisation d'un module d'idéalisation de tableaux, nos efforts s'orientent sur la définition de protocoles d'interaction plus évolués qui serviront à construire une maquette opératoire pour des utilisateurs naïfs. Les qualités ergonomiques du système seront ainsi testées en situation, en particulier pour établir des critères d'acceptabilité et de préférence pour les protocoles multimodaux.

Références

[ARGENTIN 89]
ARGENTIN G. (1989) *Quand faire c'est dire*. Pierre Mardana, éditeur.

[BENNACER et al 92]
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 et al 92]
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.

[EKMAN et FRIESEN 72]
EKMAN P., FRIESEN W.V. (1972) Hand movements. *The Journal of Communication.* pp. 353-374.

[GREIMAS 70]
GREIMAS A.J. (1970) *Du sens*. Editions du Seuil.

[MARKS et REITER 90]
MARKS J., REITER E. (1990) Avoiding Unwanted Conversational Implicatures in Text and Graphics. *Proc. Eight National Conference on Artificial Intelligence*, The MIT Press, pp. 450-456.

[MENIER et LORETTE 92]
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.

[PAVLIDIS 82]
PAVLIDIS T. (1982) *Structural Pattern Recognition*. Springer Verlag.