

# L'interprétation de tableaux dans une situation d'interaction utilisateur-machine

**Luc JULIA, Claudie FAURE**

ENST Dpt SIGNAL - CNRS URA 820, 46 rue Barrault, 75634 PARIS Cedex 13  
e\_mail : {julia, cfaure}@sig.enst.fr

**Résumé** : Dans le cadre des interfaces multimodales d'aide à la conception incrémentale de documents graphiques, nous avons développé un prototype pour l'édition de tableaux par la parole et le geste : TAPAGE. La partie relative à l'interprétation des tableaux à main levée est présentée ici. Les traitements et les représentations des données sont élaborés en fonction des différentes contraintes qu'imposent la qualité variable des tracés manuscrits, la situation d'interaction utilisateur-machine et l'objectif de coopération entre le dessinateur et le système d'interprétation des dessins.

## I / INTRODUCTION

Le travail présenté concerne le noyau applicatif d'un prototype d'interface multimodale pour la conception incrémentale de tableaux par la parole et le geste. L'architecture et les protocoles d'interaction multimodaux du prototype TAPAGE ont été décrits dans : [FAURE C, 1993, 1994 ; JULIA L, 1993]

Les ordinateurs à stylo ouvrent de nouvelles perspectives pour les protocoles d'interaction utilisateur-machine. Ils permettent de recréer la situation papier-stylo où l'utilisateur exprime rapidement ses idées sous forme visuelle. L'interaction directe au stylo évite l'apprentissage de langages ou de logiciels graphiques mais surtout libère l'utilisateur qui n'a plus à recourir à un outil d'expression intermédiaire qui rompt l'intégration naturelle de la gestuelle graphique et de la pensée. Le succès de ce type d'interaction implique que la saisie directe sur papier électronique soit augmentée de procédures de traitements qui interprètent la version encre, c'est-à-dire le brouillon produit au stylo, et reconstruisent une version "idéale" qui servira à sa publication, son archivage ou sa diffusion à

d'autres utilisateurs dans le cadre d'une conception collective du document. On parle alors de coopération entre dessinateur et machine qui travaillent sur le même document afin d'atteindre la version finale désirée.

La conception de documents graphiques dans le contexte d'une coopération utilisateur-machine impose des contraintes sur les traitements. On décrira d'abord la nature de ces contraintes. La présentation de l'interprétation automatique des tableaux fera apparaître les choix qui ont été faits au niveau des procédures et des représentations pour permettre une très grande rapidité d'interprétation et un partage des tâches de correction entre utilisateur et machine au cours de la conception incrémentale des tableaux.

## II / INTERPRÉTATION DANS UN CONTEXTE CONTRAINT

Pour comprendre les problèmes rencontrés pour la réalisation d'un éditeur de tableaux interactif, nous donnons le scénario d'une interaction entre un utilisateur et la machine : un tableau est produit au stylo sur l'écran de l'ordinateur et immédiatement interprété par la machine, la version idéalisée est retournée. Cette version peut ne pas satisfaire l'utilisateur, soit parce qu'une erreur d'interprétation a été faite, soit parce que l'utilisateur a produit un tableau qui ne correspond pas exactement à ses intentions qui évoluent au cours de la conception incrémentale. L'utilisateur a alors la possibilité de corriger son tableau à l'aide d'outils disponibles au niveau de l'interface, les modifications apportées sont interprétées et le tableau mis à jour est affiché instantanément. Une fois la structure satisfaisante, le remplissage et l'impression du tableau peuvent être effectués.

L'interprétation fait intervenir des procédures de traitement et de reconnaissance des données saisies en ligne ainsi que des connaissances liées au domaine d'application. Dans un contexte de coopération homme-machine, l'interprétation de dessins est soumise à plusieurs contraintes.

La première contrainte est celle du style de production dont la qualité est très variable d'un dessinateur à l'autre et même à l'intérieur d'un même dessin. Cette qualité semble liée à la vitesse d'exécution du dessin. Les seuils de tolérance qui sont forcément impliqués dans les décisions de l'interpréteur ne doivent pas être figés, ils seront adaptés à la qualité du dessin.

La seconde contrainte est caractéristique d'une situation d'interaction utilisateur-machine, c'est la contrainte de temps : que ce soit lors de la création, ou lors des modifications, l'interprétation doit être immédiate, les structures de représentation et les calculs doivent être optimisés. La négociation entre la qualité et le temps de l'interprétation devra être menée au plus juste, avec un impératif : l'utilisateur ne doit pas attendre plus d'une seconde un résultat.

A cela il faut ajouter la contrainte de coopération qui impose un partage des tâches entre machine et dessinateur lors de la modification du dessin. La machine doit propager automatiquement des modifications locales dans la structure du tableau et effectuer les modifications qui sont suggérées par celles de l'utilisateur pour maintenir une cohérence globale du tableau.

Une dernière contrainte, d'avantage liée à l'interface qu'au noyau applicatif est signalée. Le stylo utilisé à la fois pour produire les données (les dessins et du texte) et pour effectuer des commandes gestuelles fait apparaître la contrainte d'ambiguïté des tracés graphiques propre aux interactions au stylo. Cette ambiguïté est gérée dans ce prototype au niveau de l'interface : l'utilisateur informe le système du statut des tracés graphiques qui vont être produits. Cette solution nous semble plus fiable qu'une discrimination automatique des commandes et des données dans la mesure

où le tracé ne contient pas souvent l'information discriminante (un O ou un encerclement par exemple sont identiques). Le recours à une décision contextuelle serait souvent souhaitable mais pose les problèmes de la définition de ces contextes et de leur manipulation qui risque de ralentir l'interaction sans garantir une parfaite fiabilité. Les commandes gestuelles font l'objet de procédures spécialisées pour leur reconnaissance [ROSSIGNOL S, 1993].

### III / LES SOLUTIONS ADOPTÉES DANS TAPAGE

Nous présentons l'interprétation automatique de tableaux qui a pour but de donner une forme idéale au dessin à main levée. Les données fournies par le stylo sont des suites de points correspondant aux divers tracés (fig. 1). Ces données ne seront interprétées qu'à la fin du dessin afin d'exploiter une vue globale du tableau et afin de ne pas perturber le dessinateur en lui retournant des informations visuelles pendant son activité de production de dessin, voir à ce propos une argumentation sur la reconnaissance différée dans [NAKAGAWA M, 1993]. Nous nous mettons donc dans les mêmes conditions que la reconnaissance en-line moyennant quelques conventions propres à notre méthode portant sur l'ordre et le sens des tracés.

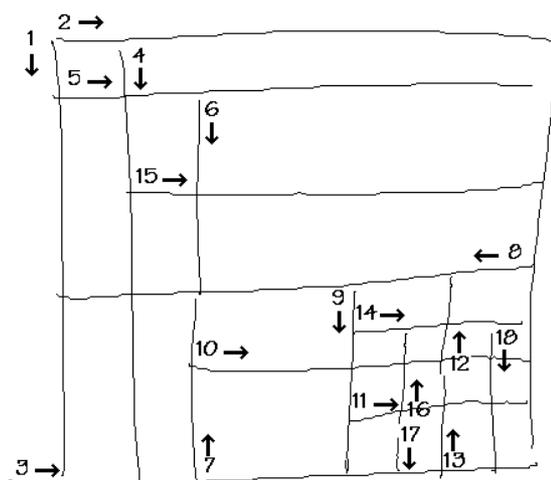
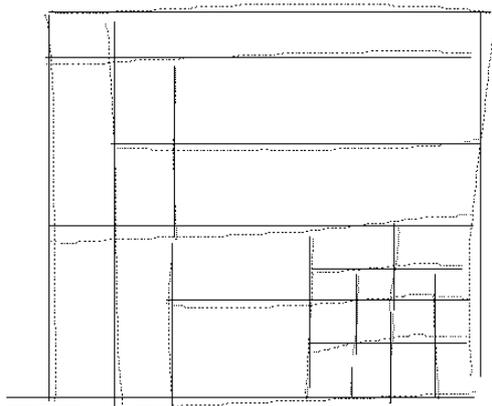


Figure 1 : dessin original et numérotation des tracés dans l'ordre de production

### III 1 / Interprétation des tracés

L'interprétation automatique est présentée par étapes, leur description souligne que les choix effectués pour les traitements et les représentations sont liés aux contraintes énoncées ci-dessus. Ces étapes sont illustrées de figures où nous rappelons en pointillés l'allure du tableau avant chaque étape pour augmenter la visibilité des changements intervenus.

La première étape de la remise en forme automatique dans TAPAGE (fig. 2) consiste à identifier chaque tracé à une verticale (V) ou à une horizontale (H) par comparaison des abscisses et ordonnées d'un ensemble de points repères du tracé. Cette opération utilise un seuil fixé *a priori*, qui a été déterminé en fonction de la résolution de l'écran. L'algorithme de segmentation fonctionne par dichotomie récursive, il est assimilable à un algorithme de la corde qui sera l'étape *split* d'une méthode *split and merge* [PAVLIDIS T, 1977]. Pour chaque tracé, le critère d'alignement de points repères est évalué. Les points repères d'un tracé de longueur L sont situés aux positions : 0, L/4, L/2, 3L/4, L. Quand ce critère n'est pas vérifié (c'est le cas du tracé n° 2), le tracé est découpé en deux au niveau d'un point repère et l'algorithme est appliqué sur les tracés résultants.

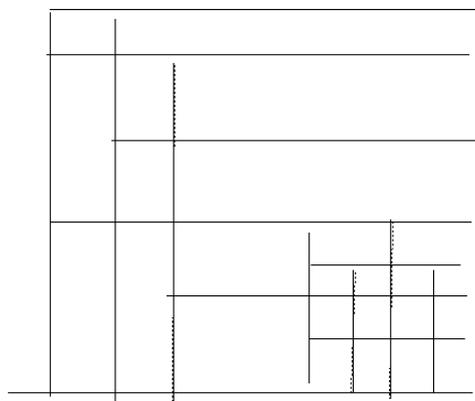


**Figure 2 :** Recherche des segments localement verticaux ou horizontaux

Le choix de points repères pré-définis rend l'algorithme de la corde sous-optimal mais

évite de calculer les distances à la droite d'approximation pour chaque point et par suite permet un gain de temps. Ce choix n'affecte pas de manière visible les performances de la segmentation en V et H. De plus, les informations relatives aux sens des tracés sont perdues intentionnellement afin d'uniformiser la notion de verticale (segment allant de haut en bas) et celle d'horizontale (segment allant de gauche à droite), ce qui permet d'utiliser une structure de données plus légère dans les étapes suivantes.

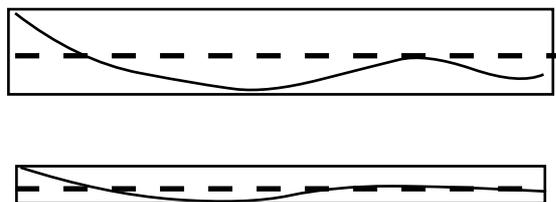
Lors de la seconde étape il y a recherche d'alignements des segments H et V, c'est le *merge* de la méthode *split and merge*. Les segments obtenus ne sont pas forcément verticaux ou horizontaux, c'est l'étape trois qui va les redresser. C'est au cours des étapes deux et trois (dont le résultat est présenté sur la figure 3) que l'on va commencer à maintenir des listes correspondant à une grille d'aimantation variable localement.



**Figure 3 :** Fusion des horizontales et des verticales adjacentes et ajustement des directions

Une grille d'aimantation (fig. 4) englobe chaque tracé interprété H ou V. Sa taille est calculée pour chacun des tracés en fonction de la dispersion des points initiaux du tracé brouillon autour du segment H ou V reconstruit par l'algorithme de segmentation. Plus la dispersion de ces points est grande (plus le dessin est brouillon) plus la taille de la grille sera importante. Cette grille définie

des seuils locaux qui seront utilisés pour la suite de l'interprétation.



**Figure 4 :** Exemples de grilles d'aimantation. En trait plein, le tracé initial. En pointillé le segment reconstruit.

Comme on le voit sur la figure 5, la quatrième étape de correction automatique permet d'éliminer les petits dépassements et de remplir les espaces au niveau des jonctions en T ou en L. Les coordonnées des segments V et H reconstruits aux étapes précédentes, associées aux tailles des grilles d'aimantation, vont permettre la détermination des références (notées *Ref*) et la reconstruction des jonctions T et L.

Algorithme de recherche des références :

- Soit  $X_1$ , la première abscisse traitée, alors  $RefX_1=X_1$  est la première référence pour les abscisses.

- Soient  $X_i$  une autre abscisse et  $j$  le nombre de références d'abscisses définies à cet instant.

- Si  $X_i \approx RefX_k$  avec  $k \in [1, j]$

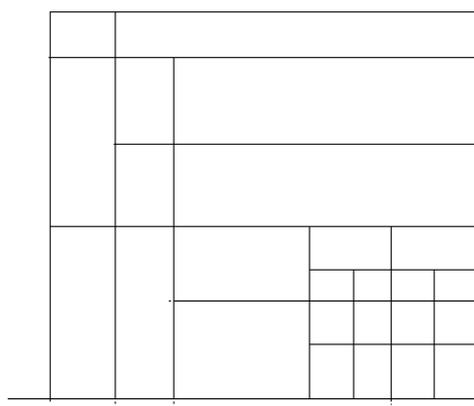
$\approx$  signifie "est proche de" et utilise les seuils locaux définis à partir des grilles d'aimantation.

Alors  $RefX_k = (RefX_k + X_i)/2$

Sinon  $RefX_{j+1}=X_i$  est une nouvelle référence pour les abscisses.

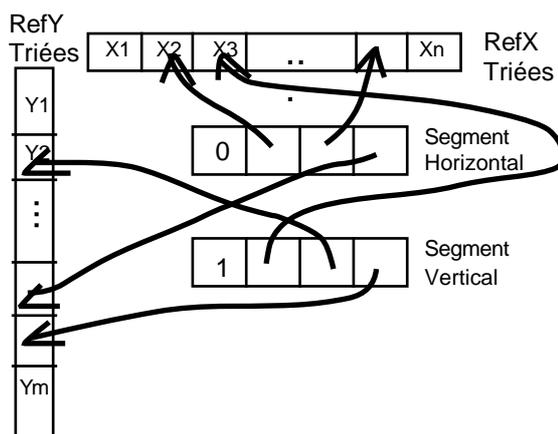
De même pour les ordonnées.

Tous les segments sont alors définis grâce à ces références. L'effet d'aimantation est une égalisation de la valeur des abscisses (resp. ordonnées) des points qui sont à l'intérieur d'une même grille d'aimantation définie pour un segment V (resp. H). La conséquence sera une reconstruction de la plupart des jonctions T et L, comme on peut le constater sur la figure 5.



**Figure 5 :** Reconstruction des jonctions par le calcul des références.

La structure de représentation est illustrée sur la figure 6. Chaque segment est représenté par un entier indiquant sa classe de direction, H ou V, et des pointeurs sur des listes de références qui spécifient la position du segment (pointage sur une référence X, resp. Y, pour un segment V, resp. H) et les points extrémités. Par exemple, quand un segment H est en contact avec deux segments V à chacune de ses extrémités, ce qui est toujours vrai dans le cas des tableaux à cellules fermées, les références X pointées par H sont aussi les références de position de ces segments V.



**Figure 6 :** Structure des données

Cette organisation des données permettra de propager rapidement des modifications locales dans l'ensemble du tableau par simple changement d'une valeur de référence

et héritage de cette valeur pour les segments qui pointent sur elle.

L'étape précédente a éliminé la plupart des extrémités libres des segments uniquement par un choix de structure de données, il peut toutefois en rester d'autres, le seuil d'aimantation s'avérant trop faible. La cinquième étape se divise en deux parties : la recherche d'une extrémité libre puis la détermination du segment perpendiculaire le plus proche pour effectuer une connexion en T ou en L. Ces opérations sont particulièrement coûteuses en temps de calcul (surtout la seconde). L'intérêt de l'étape 4 est très visible sur cet exemple : il ne reste à l'étape 5 qu'une seule jonction à traiter (jonction L en bas à gauche sur la figure 5) alors que l'étape 4 a déjà traité 21 jonctions T ou L.

### III 2 / Interprétation des structures

Le tableau a désormais une allure correcte, il n'a toutefois pas encore la structure exacte que souhaitait lui donner son concepteur : certaines différences de tailles entre des lignes et des colonnes ne sont pas significatives, elles résultent de l'imprécision introduite par la production manuscrite. Une analyse sémantique du dessin obtenu est nécessaire pour éliminer ces différences.

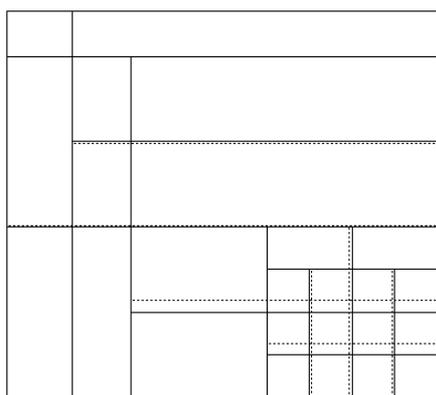


Figure 7 : Égalisation des motifs

La figure 7 illustre la différence entre le tableau avant et après cette analyse. Nous sortons ici du module d'interprétation des tracés en termes de segments H et V et de jonctions T et L pour rentrer dans le module

d'interprétation de la structure qui constitue l'étape sémantique, nous verrons plus loin le pourquoi de cette séparation.

Cette correction sémantique est faite dynamiquement par un algorithme récursif qui recherche les motifs qui permettent la remise en forme finale du tableau. Un motif est défini comme une zone entre deux segments H ou V successifs, il est caractérisé par sa largeur s'il s'agit d'un motif horizontal (dans l'exemple de la figure 8, ce sont des motifs horizontaux qui sont recherchés) et par sa hauteur pour un motif vertical. Les motifs horizontaux sont d'abord recherchés, puis les motifs verticaux. On introduit la notion de *segment fixe* et de *segment mobile*. Au départ seuls les segments les plus à l'extérieur (c'est à dire le cadre du tableau) sont fixes. Les motifs H (respectivement V) sont recherchés entre deux segments V (resp. H) successifs fixes qui forment l'espace de recherche le long d'un segment H (resp. V) qui constitue la droite de recherche. L'ordre de recherche est fixé d'une part par les positions des segments dans le tableau, de gauche à droite pour les horizontaux, de haut en bas pour les verticaux, et d'autre part dynamiquement à mesure que les segments sont fixés. Sur la figure 8, la suite des espaces de recherche des motifs verticaux est (a, e) (a, c) (c, e). Les segments sont fixés dans l'ordre : a, e, c (motifs m1), b (motifs m2), d (motifs m3).

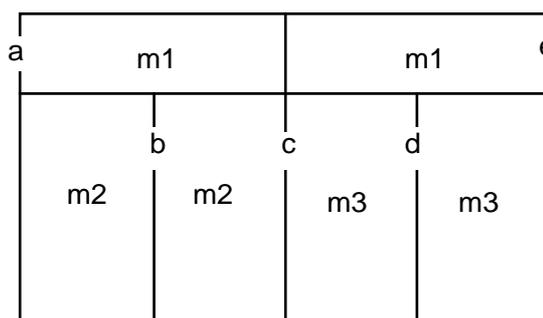


Figure 8: Exemple : l'algorithme récursif de recherche de motifs identiques détermine les espaces de recherche (a, e) (a, c) (c, e).

Deux motifs sont "identiques" quand les grandeurs qui les caractérisent (largeur ou hauteur) sont égales, à la tolérance près du seuil d'égalisation. La remise en forme fixe les segments mobiles qui délimitent les

suites de motifs "identiques" dans des positions où l'égalité des grandeurs est stricte. On note que l'identité des motifs m2 et m3 résulte d'un héritage de propriétés et non d'une connaissance explicitement représentée dans la structure de données. C'est parce que le motif m1 a permis de séparer le tableau en deux parties égales que m2 et m3 se retrouvent être identiques. L'algorithme exploite des propriétés structurelles pour limiter son calcul d'identité de motifs à ce qui est strictement nécessaire. La figure 9 est le tableau définitif tel qu'il est affiché sur l'écran, moins d'une demi seconde a été suffisante, à partir de la fin de la saisie du dessin, pour obtenir ce résultat.


*Figure 9: Tableau final*

#### IV / LA COOPÉRATION DESSINATEUR-MACHINE

L'interprétation des dessins est effectuée à chaque étape de la conception incrémentale. L'utilisateur peut faire évoluer son dessin par des modifications : il peut ajouter, déplacer, dupliquer ou effacer des composantes graphiques. Ces modifications impliquent des mises à jour où le tableau est interprété en tenant compte de la structure préexistante et des changements apportés. Les ajouts sont de nouvelles données encre qui viendront s'intégrer aux données déjà interprétées. Pour des modifications comme l'effacement ou le déplacement, la tâche est partagée entre l'utilisateur et la machine. Par exemple, si un segment est déplacé, les jonctions entre ce segment et d'autres segments du tableaux sont conservées : il y a

déplacement du sous-ensemble de la structure qui maintient une cohérence structurelle du tableau sans que l'utilisateur ait à préciser chacune des opérations. Cette mise à jour se fait automatiquement sans qu'il soit besoin d'aller rechercher l'information de jonction qui est déjà présente dans la représentation interne du tableau grâce à la structure adoptées sous forme de pointeurs entre segments.

Le tableau affiché peut ne pas correspondre aux intentions de l'utilisateur : une erreur d'interprétation s'est produite ou bien les intentions de l'utilisateurs ont changés, ce qui est caractéristiques de la conception incrémentale. Dans ces deux cas, l'utilisateur dispose d'outils pour modifier son tableau. Ces outils peuvent être appelés à l'aide de commandes multimodales qui exploitent la tendance humaine naturelle à combiner geste et parole [HAUPTMANN A, 1993], ou l'utilisation de pointage au stylo sur des menus, des objets et des positions dans le dessins, ou encore des commandes purement gestuelles. Les traitements pour la reconnaissance des commandes gestuelles (zigzag d'effacement, sélection par encerclement d'objet et flèche de déplacement) sont décrites dans [ROSSIGNOL S, 1993].

La résolution de ce type de matériel est assez grossière, de l'ordre de 30 points par cm, et la parallaxe rend difficile une localisation exacte du stylo sur l'écran. A cela s'ajoute une difficulté due au comportement "intelligent" du système qui égalise les lignes ou les colonnes qui sont sensiblement de même taille dans la version encre. En l'absence de cette égalisation automatique, qui correspond à l'interprétation de structure (cf. § III 2), il serait impossible d'obtenir des lignes et des colonnes égales compte tenu de la résolution et de la parallaxe. Par contre, si le dessinateur désire créer intentionnellement des lignes ou des colonnes adjacentes de taille légèrement différentes, il verra le système égaliser à chaque interprétation. La séparation entre le module d'interprétation des données et celui de l'interprétation des structures peut être exploitée pour empêcher des égalisations et permettre des différences minimales entre

tailles de motifs en désactivant ce dernier module à la demande de l'utilisateur.

## V / CONCLUSION

L'interaction utilisateur-machine fait apparaître la nécessité de traiter rapidement les données produites au stylo par le dessinateur. Les traitements et surtout la représentation interne du tableau ont été choisis pour minimiser les temps de calcul. La représentation choisie est basée sur la notion de *références* numériques pour organiser les données. La structure n'est pas explicite par opposition à une représentation par graphe (par exemple) qui serait plus analogue aux données visualisées. L'économie réalisée par cette forme de représentation est particulièrement visible quand le dessin subit des modifications locales qui sont propagées automatiquement dans la structure du tableau.

Pour s'affranchir de la contrainte de style, les seuils sont définis en fonction d'une qualité estimée du tracé et d'un seuil de base de faible valeur. Chaque segment horizontal ou vertical reconstruit est associé à une fenêtre d'aimantation déterminée à partir des valeurs numériques des tracés originaux qui le composaient.

La combinaison des grilles d'aimantation et des représentations par références permet de reconstruire la plupart des jonctions par des ajustement numériques aveugles, au sens où l'événement graphique traité (la jonction) n'est ni recherché ni spécifiquement traité par le système.

La séparation du module d'interprétation des tracés et du module d'interprétation des structures sera maintenu pour les futures applications. Le module sémantique pose le problème fondamental de la collaboration d'un dessinateur et d'une machine "intelligente" qui n'a à sa disposition que les règles et les conventions générales du domaine. La question de la personnalisation des interfaces implique que cette intelligence standardisée puisse aussi accepter des cas spécifiques, rares, que l'utilisateur veut produire avec l'intention de ne pas se conformer au modèle le plus général du domaine.

Des tests de validation ont été faits avec des utilisateurs experts, initiés et novices qui devaient recopier des tableaux plus ou moins complexes. La structure idéale des tableaux a toujours été obtenue après des temps d'interaction et des recours aux corrections variables (en moyenne 1mn 3s d'interaction) mais ne faisant pas apparaître de différence significative entre les initiés et les novices. Les abandons de certains sujets étaient dus à un tracé de tableau qui dépassait les limites de la surface de l'écran, ou à un cycle de corrections-mises à jours qui durait trop longtemps. Dans ces deux cas, le sujet recommençait son dessin. Dans les cas de tableaux complexes, certains sujets demandaient l'affichage du tableau interprété en cours de dessin pour continuer à travailler sur un tableau propre.

Les applications graphiques pouvant tirer parti des ordinateurs à stylo et des interfaces multimodales sont très nombreuses. Nous pensons en particulier à des outils pour les architectes et nous avons commencé à élaborer un éditeur de diagrammes en réseaux (DERAPAGE : DEssin RAPide par la PARole et le GEste) qui pour la reconnaissance des objets graphiques combine les méthodes d'interprétation et de représentation des tableaux et celles de la reconnaissance des commandes gestuelles.

### **Bibliographie :**

FAURE C., JULIA L. (1993) Interaction homme-machine par la parole et le geste pour l'édition de documents : TAPAGE. *L'interface des mondes réels & virtuels : Montpellier*, pp. 171-180.

FAURE C., JULIA L. (1994) An agent based architecture for a multimodal interface. Proc. AAAI Spring Symposium, pp. 82-86.

HAUPTMANN A.G., McAVINNEY P. (1993). Gestures with speech for graphic manipulation. *International Journal of Man-Machine Studies*, 38, pp. 231-249.

JULIA L., FAURE C. (1993) A Multimodal Interface for Incremental Graphic Document Design. *HCI International'93 : Orlando*, p. 237.

NAKAGAWA M., KATO N., MACHII K., SOUZA T. (1993) Principles of Pen Interface Design for Creative Work. *Proc. 2nd ICDAR*, pp.718-721.

PAVLIDIS T. (1977) *Structural Pattern Recognition*. Springer Verlag.

ROSSIGNOL S. (1993) Reconnaissance de gestes sur ordinateur sans clavier. *Rapport de stage d'option B1b*, Ecole Polytechnique.