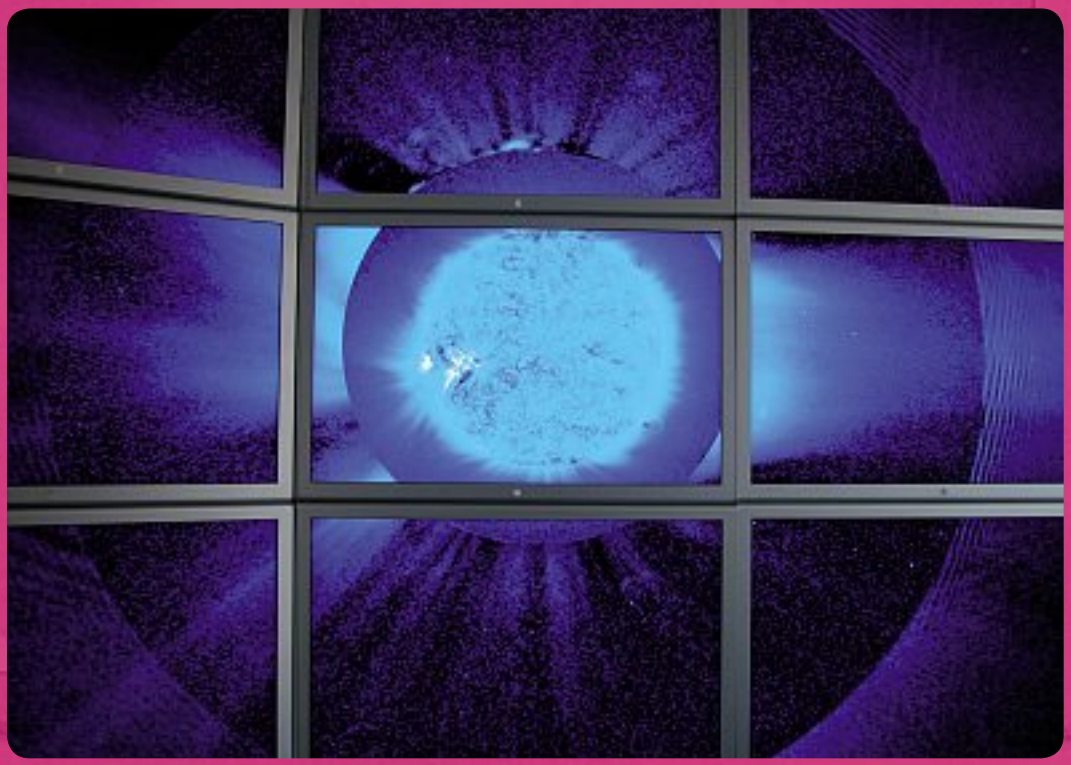
MICHEL BEAUDOUIN-LAFON*,**, EMMANUEL PIETRIGA**,*, WENDY MACKAY**,*,
STÉPHANE HUOT*,**, MATHIEU NANCEL*,**, CLÉMENT PILLIAS**,*, ROMAIN PRIMET**,*

*Laboratoire de Recherche en Informatique (LRI) (UMR 8623, UPS/CNRS)
UFR Sciences
**INRIA Saclay

131 millions de pixels qui font le mur



© M. LECOMPT

Imaginez un mur d'écrans qui affiche des images en haute définition. Imaginez que par des gestes simples, vous puissiez interagir avec lui... Nous ne sommes pas dans le film « Minority Report », mais face à la concrétisation d'un projet unique d'interaction homme-machine (IHM), la plate-forme WILD, qui permet d'interagir avec des masses de données complexes.

Photo : Affichage sur le mur d'écrans WILD d'une image haute résolution du soleil montrant l'effet des éruptions solaires à très grande distance.

L'un des enjeux majeurs du XXIe siècle sera la maîtrise de l'information et plus précisément du déluge d'information qui nous assaille un peu plus chaque jour, des messages qui s'accumulent dans nos boîtes aux lettres aux données massives produites par les capteurs qui sondent le monde qui nous entoure. Le défi pour la science informatique est immense, tant les quantités de données croissent de manière encore plus rapide que les capacités de traitement des ordinateurs. Mais le défi est aussi humain, car nos capacités cognitives pour appréhender ces données restent, elles, quasiment stables.

Permettre à des êtres humains d'appréhender, de manipuler et de comprendre des mondes d'information de plus en plus complexes, tel est l'objet de l'*interaction homme-machine* (IHM), un domaine de recherche pluridisciplinaire à l'intersection de l'informatique, de la psychologie et du design. La recherche en IHM a produit il y a une trentaine d'années la souris, le clavier et les interfaces graphiques qui sont devenus la principale lorgnette à travers laquelle nous pouvons contempler et façonner les mondes d'information. Ces outils sont aujourd'hui insuffisants, car ils ne passent pas à l'échelle des quantités de données auxquelles nous sommes désormais confrontés. De nouveaux instruments sont donc nécessaires qui doivent concilier l'ampleur des masses de données avec les limitations de nos capacités cognitives.

La plateforme WILD présentée dans cet article est un instrument de ce type, un *macroscope* destiné à visualiser et interagir avec des données massives. Le domaine d'application privilégié, mais non limitatif, de WILD est la découverte scientifique. L'activité scientifique est en effet l'une des plus grosses pourvoyeuses de données : les grands instruments comme les satellites ou les accélérateurs de particules fournissent des tera-octets, voire des peta-octets* de données ; les modèles de processus chimiques ou biochimiques font intervenir des dizaines ou des centaines de milliers d'atomes et des millions ou des milliards d'interactions. Ce domaine d'application nous permet également d'avoir aisément accès à de nombreux utilisateurs potentiels, grâce à l'immense richesse des laboratoires de l'Université Paris-Sud et du plateau de Saclay.

Un macroscope pour explorer les données massives

La plateforme WILD est un environnement multi-surfaces interactif **(encadré 1)**. Elle combine plusieurs surfaces d'affichage permettant d'interagir avec les données : un mur d'image de 5m50 x 1m80 affichant l'équivalent de 100 écrans d'ordinateurs standard, une table multi-tactile et plusieurs dispositifs portables (iPhone et iPad). On peut également y connecter des ordinateurs portables classiques.

Une telle plateforme redéfinit radicalement les conditions d'interaction auxquelles nous ont habitué les interfaces graphiques traditionnelles. Il n'est, par exemple, pas question d'utiliser la souris pour manipuler les objets sur un écran d'une telle taille. Aussi l'objectif principal de nos travaux de recherche est de comprendre comment exploiter les facultés sensori-motrices humaines pour définir de nouvelles méthodes d'interaction à la fois intuitives et puissantes. A travers les trois scénarios suivants, issus de nos collaborations avec les laboratoires associés au projet, nous illustrons le type d'utilisations envisagées.

Zoomer dans la galaxie - Les images astronomiques produites par les satellites sont de plus en plus grandes et de plus en plus précises. Ainsi, le télescope SPITZER a permis de produire une image du centre de la galaxie à partir de plus de 800 000 clichés, dont l'affichage à la résolution d'un écran normal nécessiterait une surface de la taille d'un terrain de football ! Même la très grande résolution du mur d'image de WILD ne permet pas de visualiser de telles données de façon statique : il faut pouvoir combiner la vue d'ensemble que l'on aurait depuis les gradins avec la possibilité de voir les détails en marchant sur le terrain. La solution est de combiner différentes surfaces d'affichage et de permettre de zoomer sur certaines parties sans perdre le contexte **(figure 1)**.

FIGURE 1

L'image du centre de la galaxie produite par le téléscope SPITZER, à deux niveaux d'échelle.

© IN SITU / CNRS PHOTOTHEQUE / CYRIL FRESILLON / ESA / PLANCK HFI & LFI CONSORTIA

FIGURE 2

Affichage et manipulation des images de 64 cerveaux.

© CNRS PHOTOTHEQUE / CYRIL FRESILLON

FIGURE 3

Contrôler l'affichage avec un objet physique, ici un cerveau.

© CNRS PHOTOTHEQUE / CYRIL FRESILLON

Comparer des centaines de cerveaux - Dans d'autres domaines scientifiques, c'est la capacité de comparer un grand nombre de données similaires qui est importante. Par exemple, les chercheurs en neurosciences de NeuroSpin étudient la variabilité du cerveau humain, et doivent confronter leurs algorithmes d'analyse d'images cérébrales à l'expertise humaine. Pour cela ils souhaitent afficher, comparer et classifier des dizaines, voire des centaines de cerveaux. Ici les surfaces de WILD permettent de juxtaposer un grand nombre d'images similaires que les chercheurs peuvent manipuler globalement ou individuellement **(figure 2)**. Par exemple, un modèle de cerveau tenu dans la main peut servir à contrôler simultanément l'orientation de tous les cerveaux affichés sur le mur, tandis qu'on peut examiner un détail ou comparer précisément deux images sur la table.

Comprendre la synthèse des protéines - Une troisième activité typique est de pouvoir confronter des données de natures différentes sur un même phénomène. Par exemple, les biologistes de l'Institut de Génétique et de Microbiologie d'Orsay souhaitent étudier simultanément un segment de génome, un réseau d'interaction de protéines et la visualisation 3D des molécules impliquées. Cette activité est typiquement collaborative, puisqu'elle nécessite de combiner l'expertise de différents spécialistes. Chacune de ces données est en soi complexe et bénéficie des capacités d'affichage de WILD, mais la clé est de permettre de corréler les observations entre les différentes vues, de pouvoir aisément incorporer de nouvelles données, et de garder une trace intelligible des observations réalisées.

Affordance et participation

Le défi de recherche posé par ces scénarios est de rendre les interactions aussi transparentes que possible pour les utilisateurs, c'est-à-dire de leur permettre d'effectuer des manipulations très diverses de façon simple. Mais faire simple s'avère compliqué !

FIGURE 4

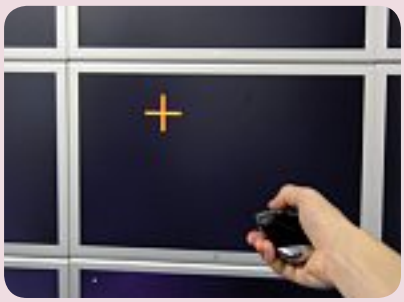
Transférer des objets entre surfaces d'affichage.

© IN SITU / CNRS PHOTOTHEQUE

Le cinéma nous a habitué à des scènes spectaculaires, par exemple dans le film Minority Report où un acteur contrôle un affichage mural complexe avec des gestes libres de la main et des bras. Nous avons cependant montré que dans une utilisation réelle, les utilisateurs préfèrent des interactions utilisant des dispositifs physiques qu'ils peuvent tenir dans la main et manipuler, comme une télécommande. En fait, l'interaction à mains libres est moins précise, moins rapide, plus difficile à apprendre et plus limitée dans ses fonctionnalités que l'interaction via un dispositif physique. La raison n'est pas seulement technologique, comme la difficulté à reconnaître de façon fiable les gestes. Il est également probable que l'évolution n'a pas favorisé chez l'être humain la faculté de contrôler des objets par le langage, fut-il oral ou gestuel. Au contraire, la théorie dite écologique de la perception de J. J. Gibson postule que les animaux et donc l'être humain sont capables de percevoir de façon immédiate les capacités d'action offertes par les objets qui les entourent (ce que Gibson appelle *affordances*). Alors que tout langage nécessite un code connu et partagé par les interlocuteurs et met en œuvre des capacités cognitives supérieures, nos fonctions perceptives élémentaires sont capables d'interpréter spontanément la taille, la forme et l'aspect d'un objet en termes de fonctionnalités possibles : actionner un bouton, faire rouler une molette, glisser le doigt sur une surface **(encadré 2)**.
Nos travaux visent à créer des interfaces simples

1 WILD : une plateforme de visualisation interactive très haute résolution

La plateforme WILD est installée depuis 2009 dans les locaux du Laboratoire de Recherche en Informatique sur le Campus d'Orsay. Elle a été financée par la Région Île-de-France, Digiteo, l'Université Paris-Sud, le CNRS, l'INRIA, le laboratoire commun INRIA-Microsoft Research et l'ANR. Elle est constituée de cinq principaux éléments.



© IN SITU

Pointage à distance sur le mur de WILD

Le plus imposant est un mur d'écrans, constitué de 32 écrans de 30 pouces de diagonale organisés en une matrice 8 x 4. Cette surface d'affichage a une taille de 10 m² (5,50 m x 1,80 m), soit le mur d'une belle pièce, et une résolution de 20 480 x 6400 pixels, soit 131 millions de pixels. A titre de comparaison, cela est équivalent à 100 écrans de résolution standard (1280 x 1024) ou 64 images de télévision haute définition (Full HD). Seule une visite permet de réellement juger l'intérêt de combiner un affichage de très grande taille et de très haute résolution : on peut en quelques secondes embrasser du regard l'ensemble du mur, et se rapprocher pour examiner un détail ou lire un texte de quelques millimètres de haut.
L'inconvénient principal de l'usage d'écrans est leurs bords, qui occupent une part non négligeable de l'espace d'affichage. Ce problème est cependant voué à disparaître, car de nouvelles technologies d'écrans vont pratiquement éliminer les bords. L'alternative, qui consisterait à utiliser un système de projection, ne permet en aucun cas d'atteindre une telle résolution. Par exemple, l'Université de Toronto dispose d'un mur de même taille physique, constitué des images de 18 projecteurs vidéo haute définition. Mais il affiche à peine un dixième du nombre de pixels de WILD, ce qui le rend peu utilisable de près. WILD n'est cependant pas le plus grand mur d'écrans actuel : à l'Université de San Diego, le projet HiperSpace a mis en place un mur de 70 écrans (14 x 5) de même taille que ceux de WILD, soit près de 290 millions de pixels. L'originalité de WILD réside dans le fait que c'est le seul mur d'image de cette taille, actuellement muni des capacités d'interaction décrites ci-dessous.

Le second composant de WILD est une table multitactile de la société IntuiLab. Cette table affiche une image de 1920 x 1080 pixels par le biais d'un projecteur placé dessous, et peut suivre en temps réel jusqu'à 20 points de contacts simultanés, typiquement les doigts des utilisateurs qui désignent et manipulent des objets à l'écran, mais aussi des objets posés sur la table. Cette table est complétée d'autres dispositifs interactifs portables que l'on peut utiliser dans la pièce : surfaces tactiles de type iPad et iPhone Apple, souris gyroscopiques, etc.
Le troisième composant de WILD est un système VICON de suivi de mouvements utilisant dix caméras infrarouge disposées au plafond, autour de la pièce. Ce système est capable de suivre en temps réel, avec une précision du demi-millimètre dans l'ensemble de la pièce, la position dans l'espace de petits marqueurs rétro-réfléchissants que l'on pose sur les objets ou personnes dont on veut suivre les mouvements. Il est utilisé notamment pour repérer la position des dispositifs portables, et, grâce à des casquettes et des gants, la position et les gestes des utilisateurs dans la pièce.

Le dernier composant de WILD est le cluster d'ordinateurs qui alimente le mur d'écrans. Ce cluster se compose de 16 machines Apple Mac Pro multicœurs, chacune pilotant deux écrans et munie de deux cartes graphiques haute performance. Le cluster est complété par deux ordinateurs frontaux de même type. Au total, ce cluster de visualisation dispose de 144 cœurs de calcul, 192 giga-octets de mémoire centrale, 18 giga-octets de mémoire graphique et 36 téra-octets de stockage sur disque. La puissance totale est de 1,5 téraflops pour les processeurs de calcul et 6 téraflops pour les processeurs graphiques. Il est de plus connecté au cluster de calcul du LRI par un réseau à très haut débit (10 giga-bits/s).

sans être simplistes, car plus naturelles. Pour cela, outre l'approche expérimentale évoquée ci-dessus, qui nous permet de comparer objectivement différentes techniques d'interaction pour des tâches données, nous utilisons la *conception participative*. Cette approche, née en Scandinavie dans les années 70, consiste à faire appel aux capacités d'innovation des utilisateurs en les impliquant dans le processus de conception. Il ne s'agit pas de faire des utilisateurs des concepteurs, mais de tirer parti de leur connaissance intime de leurs pratiques de travail pour comprendre quels aspects peuvent être améliorés par l'ordinateur. Cette approche met bien souvent en évidence le décalage entre l'activité telle qu'elle est

2 Pointer du doigt

Le geste de pointage est l'interaction élémentaire qui consiste à désigner un objet sur une surface d'affichage. La théorie sous-jacente à ce geste banal d'atteinte d'une cible est fondée sur la théorie de l'information : pointer un objet, c'est le choisir parmi l'ensemble de tous les objets possibles, c'est-à-dire transmettre à l'ordinateur l'information spécifiant ce choix. Par exemple, si l'objet visé est un carré de 10 x 10 pixels dans un écran de 1000 x 1000 pixels, un clic au hasard a une chance sur 10 000 d'atteindre la cible. On dit alors qu'il faut transmettre $\log_2$ (10 000) = 13 bits d'information pour spécifier cette cible. Sur le mur de WILD, qui affiche plus de 130 millions de pixels, la quantité d'information pour cette même cible passe à 20 bits. Pour l'image de la taille d'un terrain de football évoquée dans le texte, on arrive à près de 30 bits.

Ces nombres semblent ridicules comparés aux giga-octets de nos ordinateurs, mais ici ce sont les capacités humaines de traitement de l'information qui sont en jeu. Selon la Loi de Fitts*, le temps de pointage est directement proportionnel à la quantité d'information à fournir : dans l'exemple ci-dessus, s'il faut 2 secondes pour pointer la cible sur l'écran, il en faudrait 3 sur le mur de WILD et il en faudrait 4,5 sur le terrain de football. Malheureusement, la loi de Fitts ne s'applique plus dès lors que l'amplitude du mouvement exige de se mouvoir plutôt que de simplement bouger le bras : il faut en réalité plus de 5 secondes pour aller d'un bout à l'autre du mur de WILD, et une bonne minute pour traverser le terrain de football.

Pour cette raison, nous développons des techniques de pointage qui s'affranchissent de ces limitations et s'approchent de la limite théorique de la loi de Fitts. D'une part nous expérimentons le pointage à distance, qui permet de parcourir de grandes distances sans se déplacer, comme lorsque l'on utilise un pointeur laser. Il faut cependant régler le problème de la précision : plus on est loin du mur, plus le pointeur laser est sensible aux mouvements involontaires de la main. D'autre part nous étudions des interfaces zoomables* qui permettent d'agrandir toute ou partie de l'image affichée. La difficulté alors est d'éviter de se perdre et de contrôler simultanément l'échelle de l'affichage et la zone visée. Enfin nous exploitons les situations où le système peut aider le pointage par sa connaissance des données affichées. Ainsi, si l'on cherche une ville sur une carte, le système peut automatiquement accrocher le curseur à la ville la plus proche, ce qui revient à augmenter artificiellement la taille des cibles en absorbant l'espace libre entre elles.

Mur d'image affichant un ensemble de documents à grande résolution

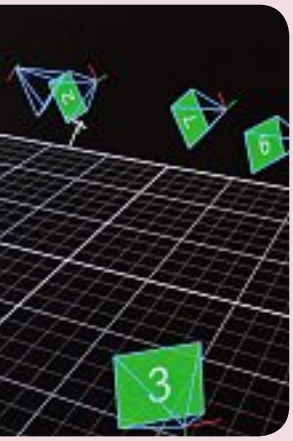
Cablage et structure support des écrans, permettant différentes configurations

Système VICON à base de caméras pour le suivi de marqueurs passifs

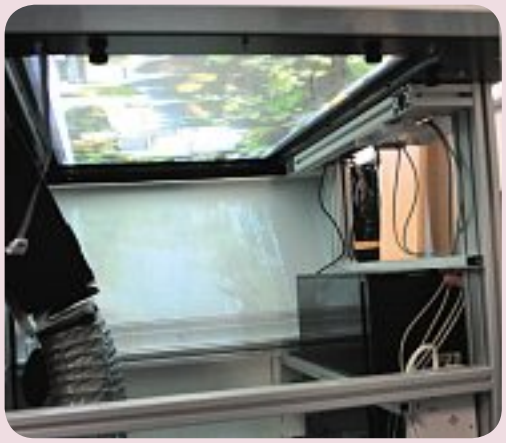
imaginée par les concepteurs et telle qu'elle est réellement pratiquée. Elle mène aussi bien souvent les utilisateurs à produire spontanément des idées innovantes. Ainsi, lorsque nous avons montré un prototype de notre application de visualisation d'images du cerveau aux chercheurs de NeuroSpin, l'un d'entre eux a spontanément eu l'idée d'utiliser un modèle de cerveau en plastique pour mettre en évidence un sillon sur les images affichées **(figure 3)**. De la même façon, lors d'une autre séance de prototypage, l'idée est apparue d'utiliser un dispositif mobile tel qu'un iPhone pour y glisser des objets affichés sur la table, se déplacer vers le mur et, par un geste inverse, transférer ces objets sur le mur **(figure 4)**.

Détail illustrant la précision de l'affichage haute résolution

Table multitactile et iPhone



Cluster de visualisation



Intérieur de la table multitactile

© PHOTOS : IN SITU

De telles idées d'interaction ne sont que le point de départ d'un long processus itératif au cours duquel les idées sont raffinées, combinées et mises à l'épreuve. Pour le chercheur en interaction homme-machine, il s'agit moins d'apporter une solution satisfaisante à un problème concret que de mettre à jour des régularités et d'identifier des concepts unificateurs qui pourront être appliqués à d'autres contextes. Ainsi, nous explorons depuis plusieurs années l'*interaction instrumentale*, qui conçoit toute interaction comme médiatisée par un instrument directement contrôlé par l'utilisateur. Inspiré par l'analyse de nos interactions avec le monde réel, où l'on utilise de nombreux outils et appareils pour manipuler des objets physiques, l'interaction instrumentale permet d'augmenter la puissance des interfaces informatiques en réutilisant les mêmes instruments dans des contextes divers. Par exemple, l'usage évoqué ci-dessus d'un iPhone comme véhicule pour transporter des données entre deux surfaces d'affichage est un instrument qui peut être utilisé avec toutes sortes d'objets.

Desktop, palmtop, tabletop, wall ...side

La plateforme WILD pose par ailleurs des problèmes spécifiques d'ingénierie logicielle : comment développer des applications qui exploitent les capacités de la plateforme, et comment réutiliser des applications existantes dans ce nouveau contexte. En effet, il n'est pas viable de devoir redévelopper complètement les applications spécialisées utilisées par les astronomes ou les biochimistes.

Chaque surface d'affichage étant gérée par un ou plusieurs ordinateurs distincts, aucun logiciel existant ne peut être utilisé tel quel. C'est particulièrement le cas pour le mur d'image dont l'affichage est contrôlé par 16 ordinateurs travaillant en parallèle. Mais c'est aussi le cas pour la table interactive. En effet, au lieu d'utiliser une souris et un clavier comme sur nos ordinateurs de bureau ("desktop"), nous utilisons des périphériques d'entrée spécifiques comme la surface multitactile de la table, le système de suivi des mouvements ou les dispositifs mobiles comme l'iPhone. De plus, nous souhaitons permettre à plusieurs utilisateurs de travailler simultanément, alors que les applications sont habituellement conçues pour un seul utilisateur. Il nous faut donc développer des gammes d'applications : pour la paume de la main (palmtop), pour la table (tabletop) et pour le mur (wallside).

Afin d'unifier le traitement des dispositifs d'entrées, nous avons développé WILDInputServer, un logiciel qui rassemble les différentes sources présentes dans la plateforme et les redistribue aux applications intéressées. Cela nous permet notamment de modifier la configuration des dispositifs d'entrée sans changer les applications. Pour l'affichage d'images zoomables, nous avons adapté notre logiciel ZVTM au mur d'image : chaque ordinateur exécute la même application, mais affiche une partie différente de l'image. L'ensemble des ordinateurs sont coordonnés par un maître qui s'exécute sur l'ordinateur frontal. Il s'avère que cette approche s'applique également à certains logiciels existants : nous avons ainsi adapté le logiciel PyMol d'affichage de molécules, qui est très largement utilisé par les chimistes et biochimistes, pour qu'il fonctionne sur le mur d'image.

FIGURE 5

ZVTM (à gauche) et WILDMol, notre version de PyMol (à droite)

© IN SITU

Enfin nous avons développé un ensemble d'outils permettant de partager aisément des données issues des ordinateurs portables des participants sur les différentes surfaces de WILD. Ainsi, Scotty et Metisse permettent de transporter pratiquement n'importe quelle fenêtre d'une application sous Mac OS ou Linux vers WILD, et d'interagir avec son contenu directement. Nous avons également une imprimante virtuelle qui peut être utilisée de n'importe quel ordinateur apporté dans la pièce et qui affiche le document imprimé sur le mur d'image ou sur la table. Bien qu'exploratoires, tous ces logiciels sont fonctionnels et fournissent de bonnes performances, validant ainsi les choix techniques de la plateforme.

La plateforme WILD est unique au monde : aucun autre laboratoire ne dispose actuellement d'un tel ensemble de surfaces d'affichages associées à de telles capacités d'interaction. Notre conviction est que, dans dix ou vingt ans, les descendants de WILD seront devenus monnaie courante, non pas seulement dans quelque laboratoire public ou privé de pointe, mais aussi dans de nombreux lieux de la vie quotidienne : à l'école, dans les salles de réunion, voire chez soi.

Le fondateur d'IBM, Thomas J. Watson, est réputé pour avoir considéré à l'aube de l'informatique qu'il n'y avait de la place dans le monde que pour 4 ou 5 ordinateurs. Aujourd'hui, il est courant pour une personne d'avoir un ordinateur au travail, un autre à la maison, un portable et un smartphone. WILD préfigure non seulement la diversification des tailles des surfaces d'affichage et des modes d'interaction qui nous permettront de maîtriser des masses de données toujours plus grandes, mais aussi une bien meilleure intégration entre ces différentes surfaces interactives. ■

Glossaire

Interface zoomable :

Interface permettant de combiner plusieurs niveaux d'échelles. Google Maps et Google Earth sont des interfaces zoomables simples. D'autres types d'interfaces zoomables utilisent des lentilles qui permettent de grossir une partie de l'image en conservant le contexte autour.

Loi de Fitts :

Loi empirique issue des travaux de Paul Fitts en 1954, qui prédit le temps de mouvement *TM* pour atteindre une cible de largeur *L* à une distance *D* selon la formule suivante :
$TM = a + b \log_2(1 + D / L)$, où *a* et *b* sont des constantes qui dépendent du contexte de la tâche, comme par exemple le type de dispositif de pointage.

Peta-octet :

10^{15} octets, soit un million de milliards.
(Pour mémoire : Téra-octet = 10^{12} et Giga-octet = 10^{9})

© IN SITU