



HAL
open science

Pourquoi vouloir interagir avec des ordinateurs ?

Michel Beaudouin-Lafon

► **To cite this version:**

Michel Beaudouin-Lafon. Pourquoi vouloir interagir avec des ordinateurs ?. CNRS Editions. Vers le cyber-monde. Humain et numérique en interaction, pp.18-41, 2020, 978-2-271-13459-2. hal-03121243

HAL Id: hal-03121243

<https://hal.science/hal-03121243v1>

Submitted on 26 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

POURQUOI VOULOIR INTERAGIR AVEC DES ORDINATEURS ?

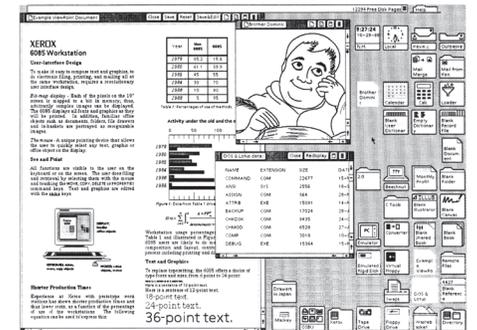
par Michel Beaudouin-Lafon

Les interfaces utilisateurs de nos ordinateurs actuels ont très peu évolué depuis 40 ans et l'avènement des interfaces graphiques. Dans les années 1970, dans les laboratoires américains de Xerox PARC, des équipes de chercheurs ont créé les principes de base des interfaces graphiques actuelles et les premiers prototypes. Le premier produit commercial, le Xerox Star, fut un échec (figure 1). C'est le Macintosh d'Apple qui, reprenant ces mêmes idées mais en visant un marché plus large, a popularisé ces interfaces et véritablement lancé l'informatique personnelle. Aujourd'hui, des milliards de personnes utilisent quotidiennement les interfaces graphiques sur leurs ordinateurs de bureau, leurs ordinateurs portables, leurs tablettes et leurs smartphones. Avec la miniaturisation, la baisse des coûts et le développement des réseaux, les interfaces graphiques ont été l'un des principaux vecteurs de l'avènement de la société de l'information, où l'ordinateur est omniprésent dans notre vie privée et professionnelle.

1 Les interfaces graphiques

Les interfaces graphiques sont fondées sur des principes dits de « manipulation directe », par opposition aux langages de commandes qui étaient, avant leur avènement, la principale forme d'interaction avec les ordinateurs. Alors qu'un langage de commande nécessite de connaître le vocabulaire et la syntaxe du langage, les interfaces à manipulation directe affichent les objets d'intérêt sous forme graphique et permettent à l'utilisateur de les désigner directement avec un dispositif de pointage (souris, surface tactile). Grâce à ce couplage entre représentation graphique et action directe sur l'objet d'intérêt, un ensemble de techniques d'interaction ont été développées pour manipuler tout type de contenu. Fenêtres, menus, boîtes de dialogue, barres de défilement, palettes d'outils sont devenus les briques de base de l'interaction que nous connaissons tous. Avec la puissance graphique croissante des ordinateurs, des techniques de visualisation interactive de plus en plus sophistiquées permettent désormais de manipuler des représentations visuelles de masses de données complexes (cf.).

Pendant l'évolution la plus marquante de ces interfaces au cours des 40 dernières années est l'apparition des écrans tactiles, qui permettent de toucher directement les objets avec le doigt plutôt que par l'intermédiaire d'un dispositif de pointage (souris, pavé tactile) et d'un curseur. Après différentes tentatives infructueuses dans les années 1990 et 2000, avec le Tablet PC et les PDAs (Personal Digital Assistant), c'est à nouveau la société Apple qui a popularisé ces interfaces avec l'iPhone (2007). Contrairement au Tablet PC et aux PDAs, pour lesquels les fabricants avaient simplement transposé les interfaces graphiques des ordinateurs de bureau en remplaçant la souris par un stylet, l'interface de l'iPhone a été conçue d'emblée pour l'interaction tactile, où le doigt cache l'objet avec lequel on veut interagir et est de ce fait peu précise.



1. L'interface originale du Xerox Star en 1981 n'a rien à envier à nos interfaces actuelles.

Cependant, malgré leur succès, les interfaces graphiques exploitent très peu nos capacités d'interaction. L'interaction avec la souris est fondée sur trois actions élémentaires: pointer (déplacer le curseur vers l'objet d'intérêt), cliquer, et glisser-déposer. L'interaction sur écran tactile, elle, est fondée sur quelques gestes simples: « *tap* » pour désigner un objet, « *swipe* » pour balayer l'écran, en général pour la navigation, et « *pinch* » pour agrandir/réduire ou ouvrir/fermer un objet.

De plus, les écrans tactiles ne le sont pas vraiment: s'ils détectent le contact d'un ou plusieurs doigts, ils ne fournissent pas de retour d'information tactile (ou « haptique ») — on ne touche jamais qu'une surface de verre lisse et froide, quels que soient les objets affichés à l'écran. Cette interaction haptique serait particulièrement utile pour les personnes aveugles et malvoyantes, ainsi que pour réaliser des manipulations simples sans regarder l'écran, par exemple lorsque l'appareil est dans la poche. Quelques prototypes expérimentaux permettent de produire des sensations haptiques lorsque l'on touche l'écran, en utilisant des matériaux piézoélectriques, ou l'énergie électrostatique, ou encore des dispositifs microfluidiques qui injectent un liquide pour rendre saillantes des parties de l'écran. Ils sont cependant encore loin d'être commercialisables.

Face aux limitations des interfaces actuelles, la recherche a depuis longtemps exploré de nouvelles formes d'interaction, plus riches ou plus adaptées à certains contextes d'usage. Deux d'entre elles commencent à avoir un succès commercial: les interfaces vocales et la réalité virtuelle et augmentée.

2 Les interfaces vocales

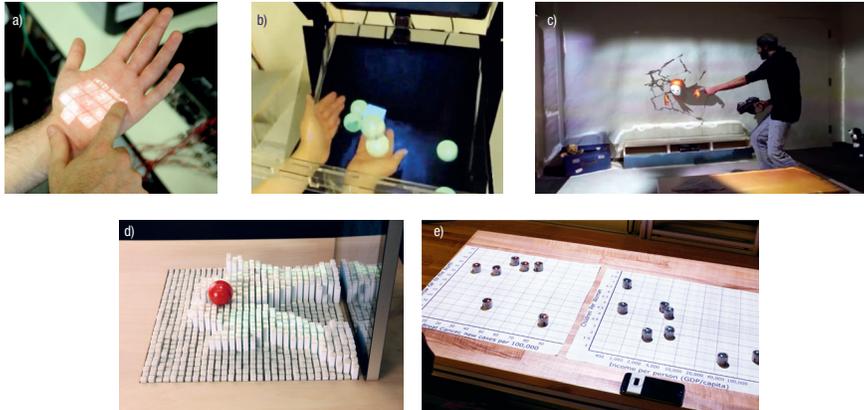
Les travaux sur la reconnaissance et la synthèse de la parole ont commencé dès les années 1950, mais ce n'est que récemment que les assistants vocaux tels que Amazon Echo, Google Home ou Apple Siri ont trouvé un large public. L'interaction vocale est particulièrement utile dans les situations où l'on ne peut manipuler physiquement un ordinateur, par exemple pour permettre à un chirurgien de contrôler certaines fonctions d'un robot chirurgical. Elle est également très utile pour les aveugles et malvoyants en complément ou à la place d'une interface visuelle, à la fois en sortie, avec la lecture du contenu de l'écran, et en entrée, pour fournir des commandes. L'interaction vocale est particulièrement efficace pour des commandes simples et non ambiguës. Mais les systèmes actuels ont du mal à gérer des dialogues complexes, et certaines interactions peuvent être fastidieuses. En dehors de

certaines classes d'utilisateurs et certains types d'usages, les interfaces vocales sont donc, en général, un complément, et non un substitut, aux interfaces graphiques.

3 La réalité virtuelle et augmentée

La réalité virtuelle est elle aussi une technologie ancienne: le premier casque de réalité virtuelle date de 1967 (Ivan Sutherland). Là aussi, ce n'est que récemment que le coût et la performance de cette technologie ont permis de la rendre accessible au plus grand nombre. La réalité virtuelle immerge l'utilisateur dans un monde synthétique dans lequel ce qu'il perçoit (images, sons et, idéalement, sensations tactiles) est produit en temps réel par l'ordinateur, qui capte et interprète ses actions (gestes, mouvements, déplacements), voire ses commandes vocales. Elle est particulièrement utile pour appréhender des phénomènes complexes et des objets tridimensionnels, comme des simulations de phénomènes physiques et des maquettes numériques de véhicules, d'avions ou de bâtiments. Elle est utilisée depuis un certain temps dans l'industrie. Son (relatif) succès récent auprès du grand public est plutôt lié au domaine des loisirs, avec le cinéma et les jeux. L'immersion totale dans un univers synthétique est à la fois une force et une limitation de la réalité virtuelle, car elle coupe l'utilisateur du monde physique. La « réalité augmentée » tente donc de réconcilier les deux en superposant au monde physique des objets virtuels. Cette technologie, également utilisée dans l'industrie depuis un certain temps, commence à voir des applications grand public: géolocalisation et guidage qui superposent les informations à une vidéo temps réel de l'espace environnant; jeux tels que Pokemon Go qui fonctionnent sur le même principe; catalogues interactifs qui permettent de visualiser immédiatement un meuble dans son appartement, etc.

À ces trois grandes catégories d'interfaces – graphiques, vocales et réalité virtuelle et augmentée – s'ajoutent de nouvelles perspectives liées à l'avènement de l'intelligence artificielle (IA). Grâce à l'IA, et au prix d'une collecte massive de données bien souvent à caractère personnel, de nouveaux systèmes sont désormais capables de réaliser des tâches à la place de l'être humain, comme identifier des lésions cancéreuses dans une image médicale ou conduire une voiture. Avec les interfaces cerveau-ordinateur (cf.), il ne sera peut-être même plus besoin de spécifier explicitement des commandes mais simplement de les « penser ». Les interfaces que nous connaissons aujourd'hui ne sont-elles pas vouées à disparaître, la machine devenant de plus en plus autonome et nos comportements de plus en plus prévisibles?



2. (a) SkinPut utilise le corps comme interface, www.chrisharrison.net/index.php/Research/Skinput ; (b) Holodesk permet d'interagir directement avec des objets virtuels, <http://people.inf.ethz.ch/otmarh/HoloDesk.html> ; (c) RoomAlive transforme une pièce en jeu vidéo, www.microsoft.com/en-us/research/project/roomalive ; (d) InForm permet d'agir sur des objets physiques à distance, <http://tangible.media.mit.edu/project/inform> ; (e) Zooids est une armée de micro-robots pour la visualisation physique, <https://aviz.fr/swarmui>.

4 Comment éviter l'asservissement aux machines ?

Les technologies issues de l'IA peuvent certes nous libérer de certaines tâches, mais c'est au prix de la création d'un nouveau prolétariat au service de ces machines. Les « travailleurs du clic » alimentent les algorithmes pour des salaires de misère, les livreurs et chauffeurs répondent aux ordres des algorithmes d'entreprises comme Uber ou Deliveroo et les modérateurs doivent nettoyer les réseaux sociaux des pires contenus. Par ailleurs, ces technologies ne sont pas infaillibles: la voiture « presque » autonome exige du conducteur d'être en mesure de réagir instantanément en situation d'urgence, alors qu'il a désappris la conduite; et les employés des entrepôts d'Amazon doivent porter un gilet de géolocalisation pour que les robots ne les heurtent pas.

Ce futur annoncé de l'intelligence artificielle tend à nous asservir aux machines: l'expression « l'humain dans la boucle », qui est utilisée pour montrer que l'humain est pris en compte, décrit en fait un humain au service des algorithmes. L'alternative consiste à utiliser la puissance de l'ordinateur pour augmenter nos capacités et non pas nous en déposséder, ce que Wendy

Mackay, directrice de recherche à Inria, appelle « l'ordinateur dans la boucle ». Douglas Engelbart, l'un des visionnaires de l'informatique, envisageait dès les années 1960 l'ordinateur comme le moyen d'« augmenter l'intellect humain ». Soixante ans plus tard, cette vision est encore d'actualité. L'ordinateur est « une bicyclette pour l'esprit », comme le disait Steve Jobs, un magnifique outil qui peut, et doit, nous rendre plus humains. Dans l'imaginaire de Hollywood, l'exosquelette d'Iron Man qui décuple ses capacités est un futur plus excitant que celui des humains qui se sont enfuis sur l'Axiom dans le film Wall-E, et passent leurs journées dans un parc à thème, abreuvés de vidéos et de boissons sucrées sur leurs fauteuils autonomes.

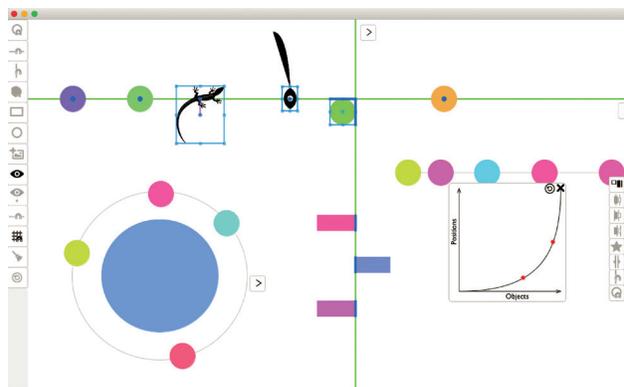
Les laboratoires de recherche en interaction humain-machine explorent depuis longtemps des technologies d'augmentation. Ils visent à créer une « réalité mixte » qui combine le monde physique et le monde numérique de manière encore plus intégrée que la réalité augmentée (figure 2). Ainsi, Chris Harrison, du Human Computer Interaction Institute, a développé SkinPut, qui combine la projection d'images sur son propre corps, comme l'image d'un clavier de téléphone sur la main, et la détection d'interactions tactiles avec cette image: le corps devient alors l'interface. Otmar Hilliges, de Microsoft, a créé Holodesk, une interface qui permet de manipuler des objets virtuels avec les mains comme s'ils étaient réels, grâce à un miroir semi-transparent et un système de suivi des gestes de haute précision. Microsoft a également développé RoomAlive, qui scanne le contenu d'une pièce entière et affiche ensuite, par projection, un monde virtuel adapté à la géométrie et au mobilier de la pièce. Sean Follmer, du MIT Media Lab, a créé InForm, une matrice de « pixels physiques » pouvant coulisser verticalement de façon à créer des formes tridimensionnelles et agir sur des objets posés dessus. Mathieu Le Goc, de l'Inria, a créé les Zooids, une armée de micro-robots qui se déplacent sur le bureau et peuvent servir à visualiser de l'information sous forme physique, mais aussi à déplacer des objets physiques sur le bureau.

5 La réconciliation des mondes physique et numérique

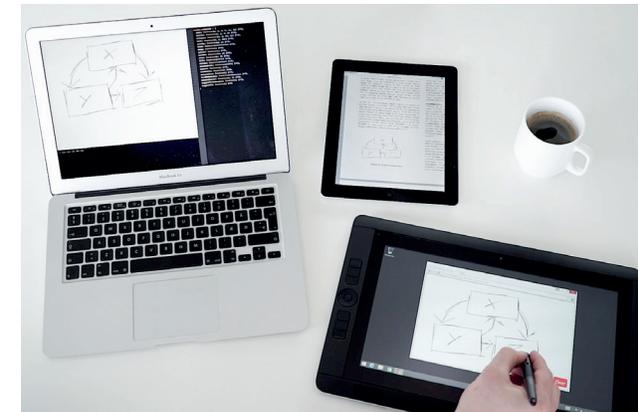
Ces recherches exploratoires ont pour but de réconcilier le monde physique et le monde numérique, que nos interfaces actuelles ont tendance à trop séparer. Pour être menée à bien, cette réconciliation nécessite de mieux

comprendre et exploiter nos capacités d'interaction et de communication, afin de créer des systèmes au service de l'humain. Nos modes d'interaction et de communication reposent sur trois grands paradigmes. Tout d'abord, pour interagir avec le monde physique, nous créons des outils qui permettent d'augmenter nos capacités. Ce sont des interfaces « à la première personne » (je fais). C'est principalement l'approche étudiée par le domaine de l'Interaction Humain-Machine. Ensuite, nous déléguons une tâche à une personne qui peut la faire à notre place. Ce sont donc des interfaces dites « à la deuxième personne » (je fais faire). C'est principalement l'objet de l'Intelligence Artificielle. Enfin, pour interagir avec d'autres personnes afin d'effectuer une tâche collective, ou simplement pour échanger de l'information, nous utilisons des médias de communication, comme la voix ou l'écrit. Il s'agit d'interfaces « à la troisième personne » (je fais avec un/une autre). C'est principalement l'objet de la Communication Médiatisée et du Travail Coopératif Assisté par Ordinateur.

Le paradigme de l'ordinateur comme outil est fondé sur l'observation que l'être humain a une capacité extraordinaire à créer et utiliser des outils pour toutes sortes de tâches : crayons pour dessiner, ustensiles pour cuisiner, bicyclette pour se déplacer, etc. Des travaux en neurosciences ont démontré que le simple fait de tenir un outil en main modifie notre perception de l'environnement : tenir un bâton nous fait percevoir des objets éloignés comme étant plus accessibles, comme si notre bras était soudain plus long. Ainsi, notre



3. StickyLines permet d'attacher des objets graphiques à des guides afin de les maintenir alignés, d'ajuster l'alignement en fonction de la forme des objets, comme le lézard ou le point d'exclamation, et de contrôler la répartition des objets (à droite).



4. Webstrates permet de partager en temps réel des documents complexes et de les éditer avec des outils adaptés à chaque support (tablette, ordinateur...) et à chaque utilisateur.

cerveau est en quelque sorte « câblé » pour l'usage d'outils. Nous sommes également capables de nous approprier des objets pour en faire des outils, même si ce n'est pas leur fonction première. Qui n'a pas utilisé un crayon comme règle pour tracer un trait droit, ou une tasse pour dessiner un cercle.

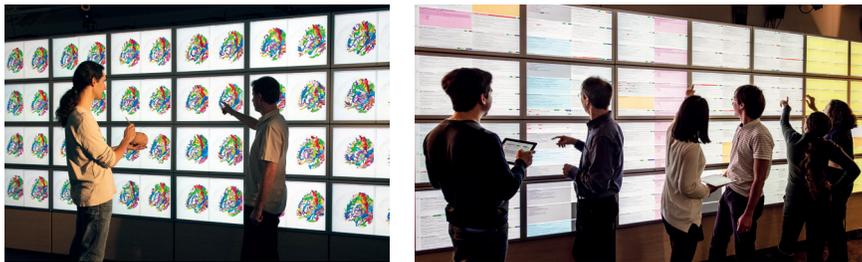
Ces capacités de création et d'appropriation sont malheureusement très peu présentes dans nos interfaces actuelles : chaque application propose un ensemble fixe de fonctionnalités, limitées à cette application. Un traitement de texte permet d'entrer et de modifier du texte, mais pas d'en faire soudain une partition musicale ou une surface de dessin, alors qu'une simple feuille et un crayon permettent ces trois usages (et aussi de plier la feuille pour en faire un avion, ce qu'un traitement de texte ne saura jamais faire!). De même, le sélecteur de couleur utilisé dans une application de retouche photographique est confiné dans cette application, et ne peut pas être utilisé dans le traitement de texte, et vice-versa.

Pour réellement exploiter le paradigme de l'ordinateur-outil, il faut créer des logiciels plus flexibles et plus appropriables. Notre équipe de chercheurs de l'Université Paris-Sud et du CNRS (LRI) a exploré cette approche et a ainsi créé StickyLines (figure 3), un outil permettant d'aligner et de répartir des formes graphiques. Au lieu de la traditionnelle commande d'alignement, un outil permet d'aligner des objets en les attachant au guide d'alignement et

de maintenir les objets alignés en déplaçant simplement le guide. Le résultat est une méthode plus simple et plus puissante pour positionner les objets. En présentant cet outil à des graphistes professionnels, nous avons identifié leur besoin de « tricher » avec l'alignement strictement géométrique, par exemple lorsqu'ils doivent aligner des logos dont la forme est telle que leur centre visuel n'est pas leur centre géométrique. Nous avons alors complété l'outil d'alignement avec un outil d'ajustement qui permet de décaler le centre géométrique de l'objet, permettant au graphiste un contrôle fin de la mise en page. Cet outil a été aussi utilisé de manière imprévue pour grouper sur un même guide des objets avec des ajustements extrêmes, qui ne sont plus du tout alignés. Cet exemple montre qu'il est possible de créer des outils numériques simples et puissants, que les utilisateurs s'approprient de manière inattendue.

Un autre exemple est le système Webstrates créé par des collègues de l'Université d'Aarhus au Danemark avec des chercheurs du LRI et de Télécom Paris pour explorer ce que pourrait être un substrat d'interaction beaucoup plus riche et flexible que les documents numériques actuels. Webstrates (figure 4) permet de partager n'importe quel contenu Web en temps réel, entre dispositifs divers (ordinateur de bureau, tablette...) et d'éditer ce contenu avec les outils de son choix. Les outils eux-mêmes sont des documents que les utilisateurs peuvent échanger ou partager, voire créer ou modifier. Webstrates montre que l'on peut s'affranchir de la notion d'application et des barrières infranchissables qu'elle implique, et créer au contraire des environnements beaucoup plus ouverts et flexibles, plus à l'image de notre expérience du monde physique.

Les chercheurs expérimentent également l'interaction « en grand » dans des salles interactives munies d'immenses écrans interactifs (cf.) et de dis-

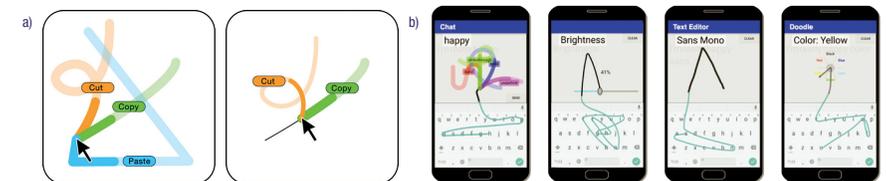


5. L'interaction avec de grands murs d'écrans, ici la plateforme WILD du Laboratoire de Recherche en Informatique (projet DIGISCOPE), permet de manipuler des données complexes, comme une banque de données de cerveaux (à gauche) ou des données textuelles (à droite), <http://digiscope.fr>.

positifs d'interaction divers (figure 5). Ces environnements préfigurent ce que pourraient être le bureau, la salle de réunion ou la salle de classe du futur, où l'information numérique ne serait plus enfermée dans les petits écrans de nos ordinateurs, tablettes et smartphones, mais au contraire disponible en grand, sous différentes formes et, surtout, de manière interactive. Ces environnements se prêtent naturellement au travail coopératif, sur place mais aussi à distance, ce qui oblige à repenser les interfaces qui actuellement font généralement l'hypothèse d'un unique utilisateur. Nous avons ainsi développé des interfaces tangibles permettant à l'utilisateur de manipuler une banque de données d'images du cerveau, des gestes multi-utilisateurs pour manipuler des données sur des écrans de la taille d'un mur, et des dispositifs de vidéo-communication adaptés à ce type d'espace.

6 Développer l'expertise des utilisateurs

Une caractéristique des outils physiques est qu'ils permettent à leurs utilisateurs de développer une expertise au cours du temps, de la même façon qu'un instrumentiste peut devenir un virtuose de son instrument de musique. Par comparaison, même si l'on peut parfaitement maîtriser un logiciel, cette expertise est à la merci des changements effectués par l'éditeur du logiciel dans sa prochaine version. Le résultat est que peu d'utilisateurs exploitent les fonctions avancées. Peut-on imaginer des interfaces qui au contraire encouragent le développement de l'expertise et permettent d'atteindre la virtuosité ? Revenons à l'exemple des écrans tactiles des smartphones et tablettes, qui n'exploitent qu'un vocabulaire réduit de gestes (« *tap*, *swipe*, *pinch* »). L'expertise



6. (a) Octopocus permet de découvrir les commandes gestuelles, www.olivierbau.com/octopocus.php; (b) CommandBoard permet d'entrer des commandes gestuelles complexes en combinant le clavier gestuel et la zone au-dessus de celui-ci, <https://ex-situ.lri.fr/projects/commandboard>.



7. (a) FieldWard permet de créer de nouvelles commandes gestuelles en s'assurant qu'elles sont reconnues par le système <https://ex-situ.lri.fr/projects/pathward-and-fieldward>; (b) Expressive Keyboard permet de varier la couleur du texte selon la façon dont les mots sont tracés sur le clavier gestuel <https://ex-situ.lri.fr/projects/expressive-keyboards>.

ne peut s'exprimer que par une plus grande rapidité d'exécution. Pourquoi ne pourrait-on pas réellement « jouer » du smartphone, grâce à des gestes plus sophistiqués qui, sans supprimer les gestes de base, permettraient d'atteindre des fonctions plus avancées ou de créer des raccourcis pour les fonctions favorites de l'utilisateur ?

Wendy Mackay de l'Inria a développé une famille de techniques d'interaction gestuelle à cet effet. Octopocus (figure 6a) permet à l'utilisateur de découvrir les gestes disponibles : lorsque l'on marque une hésitation, en posant le doigt sans le bouger ou en s'arrêtant au cours d'un geste, Octopocus affiche un guide dynamique montrant les gestes disponibles directement à l'écran. L'utilisateur n'a plus qu'à suivre le geste souhaité. En pratiquant régulièrement un geste, l'hésitation disparaît et le guide ne s'affiche plus, permettant une transition fluide du mode novice au mode expert. CommandBoard (figure 6b) permet d'exploiter la richesse d'un langage de commande en écrivant le nom de la commande sur un clavier gestuel et en continuant le geste vers le haut de l'écran, où les paramètres de la commande peuvent être spécifiés.

Enfin, Fieldward (figure 7a) permet à l'utilisateur de créer ses propres gestes en s'assurant qu'ils seront reconnus de manière fiable par l'ordinateur : pendant que l'utilisateur trace le geste, l'écran se colore de zones allant du rouge au bleu. Si l'utilisateur termine le geste dans une zone rouge, celui-ci est ambigu pour la machine ; s'il termine dans une zone bleue, il sera reconnu de façon fiable. L'ordinateur indique ainsi à l'utilisateur l'information pertinente de son fonctionnement interne afin de guider l'utilisateur.

Une dernière technique, Expressive Keyboards (figure 7b), est destinée à mieux exploiter l'expressivité naturelle de nos gestes. Alors que les systèmes de reconnaissance de gestes classiques ont pour objectif premier de reconnaître le geste en le classifiant, et donc en éliminant la variabilité entre gestes similaires, Expressive Keyboard extrait cette variabilité pour l'exploiter de manière créative. Par exemple, la couleur du texte ou la police de caractères peut varier en fonction des caractéristiques du geste telles que sa forme (plus ou moins arrondie ou anguleuse), sa taille globale, ou sa vitesse d'exécution. L'expérience montre que les utilisateurs sont capables de maîtriser ces paramètres d'exécution afin de contrôler l'effet en sortie, ce qui ouvre des perspectives nouvelles pour les interfaces gestuelles.

Ces quatre techniques sont destinées à augmenter la puissance d'expression des interfaces gestuelles sur écran tactile. Elles sont compatibles avec les interactions de base, et destinées à des utilisateurs qui souhaitent développer leur expertise. Elles illustrent ce que pourrait être un réel outil informatique au service des capacités humaines.

7 Conclusion

En conclusion, même si les interfaces actuelles sont encore relativement primitives, elles ne seront pas supplantées, mais plutôt complétées par de nouvelles formes d'interaction comme l'interaction vocale ou la réalité virtuelle, ou par les avancées de l'intelligence artificielle. Les travaux de recherche démontrent que l'on peut développer des techniques d'interaction selon le paradigme de l'ordinateur-outil qui exploitent mieux nos capacités pour les augmenter. Le paradigme de l'ordinateur-partenaire ne doit pas pour autant être ignoré. Au-delà du modèle de délégation de tâche qui lui est souvent dévolu, on peut imaginer de véritables partenariats humain-machine dans lesquels la machine apporte sa part d'expertise pour stimuler l'intelligence humaine au lieu de la remplacer. Certains travaux ont commencé à explorer cette voie, par exemple dans la production musicale où un ordinateur accompagne en temps réel un instrumentiste qui improvise, sans connaissance *a priori* de ce qui va être joué. C'est en combinant les trois paradigmes de l'objet, du partenaire et du média que l'on pourra donner vie à la vision de Douglas Engelbart d'utiliser les ordinateurs pour augmenter l'intellect collectif humain.