

Conception et apprentissage des interactions tactiles : Arpege et le cas des postures multi-doigts

Emilien Ghomi^{1,2}, Olivier Bau^{1,2}, Wendy Mackay^{2,1}, Stéphane Huot^{1,2}

¹LRI, Université Paris-Sud - ²INRIA Saclay

Bât. 490, F-91405 Orsay, France

ghomi, bau, mackay, huot @ lri.fr

RESUME

Malgré la démocratisation des écrans multi-tactiles, les interactions associées restent souvent limitées à quelques gestes n'impliquant qu'un ou deux doigts. Enrichir ces interactions avec de grands vocabulaires de gestes multi-doigts offrirait certainement plus d'expressivité et d'efficacité pour des utilisateurs experts, comme c'est le cas des musiciens ou des sténotypistes. Mais cela soulève deux problèmes majeurs que nous abordons dans cet article : la conception des gestes, et leur apprentissage par un utilisateur novice. Nous proposons tout d'abord des règles de conception de gestes issues d'études mécaniques de la main, ainsi que des résultats montrant l'efficacité de ces gestes pour des utilisateurs experts. Ensuite, nous présentons *Arpege*, un système de guidage pas-à-pas centré sur le geste, qui divise par trois le nombre d'erreurs de saisie au cours de l'apprentissage, comparé à une aide statique standard. Enfin, nous introduisons les widgets multi-doigts, autre approche pour une assistance contextuelle aux gestes multi-doigts, centrée sur la représentation graphique de l'interaction.

MOTS CLES : Gestes, représentation, multi-doigt, apprentissage, postures, feedforward, feedback, widgets.

INTRODUCTION

L'intérêt grandissant pour les interactions tactiles joue un rôle moteur dans le développement des technologies de captation, pour la plupart directement disposées sur des écrans (LCD ou projection). Elles sont aujourd'hui sensibles à plusieurs points ou zones de contact simultanées, ce qui permet de dépasser le paradigme d'interaction mono-pointeur standard hérité des années 70. Après 20 ans d'expérimentations, la première vague d'écrans multi-tactiles¹ grand public est apparue sur le marché dans les années 2000, avec le Lemur de JazzMutant en 2004, la technologie FTIR utilisée par Jeff Han en 2006, et surtout l'iPhone d'Apple en 2007.

¹ Nous qualifierons de « multi-tactile » un dispositif pouvant détecter plusieurs contacts à la fois, et de « multi-doigt » une technique d'interaction nécessitant l'action simultanée de plusieurs doigts.

Malgré ces évolutions matérielles continues, on constate, depuis la large diffusion de l'iPhone, que les techniques d'interaction usuelles n'exploitent qu'une poignée de gestes, et demandent rarement l'utilisation de plus de deux doigts. C'est cette exploration frileuse des capacités multi-tactiles des dispositifs et des degrés de liberté de la main qui a motivé notre recherche sur les interactions multi-doigts, partant du constat que : (i) peu de travaux ont exploré la définition de vocabulaires de gestes multi-doigts, et le choix des gestes est très souvent arbitraire ; (ii) la découverte et l'apprentissage de ces gestes par l'utilisateur nécessitent des aides interactives.

Nous présenterons d'abord les résultats du travail de recherche intitulé *Arpege : Design and Learning of multi-finger chord gestures* [3]. Nous y explorons et proposons des solutions pour la conception et l'apprentissage des « postures multi-doigts », que nous nommerons également « accords ». On les distingue des « gestes multi-doigts » car les points de contacts y sont statiques. D'autre part, même dans le cas des écrans tactiles où le mouvement de la main est contraint par le contact avec la surface de captation, les gestes réalisables restent nombreux. Nous nous sommes donc restreints à l'étude des gestes pour lesquels seul le bout des doigts est en contact avec la surface (figure 1). Le système *Arpege* sera proposé et évalué comme assistance contextuelle à la réalisation de ces accords.

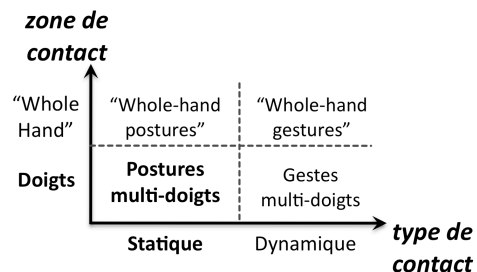


Figure 1 : Catégorisation des interactions avec une surface multi-tactile, et place des postures multi-doigts. Une technique d'interaction peut résulter de la combinaison de plusieurs de ces catégories.

Pour conclure, nous présenterons nos travaux en cours et nos perspectives de recherche sur la conception de « widgets multi-doigts », représentations graphiques

centrées sur l'interaction et son contexte, plutôt que sur le geste abstrait et son apprentissage comme avec Arpege.

POSTURES MULTI-DOIGTS

Travaux de référence

La plupart des interactions proposées sur les surfaces multi-tactiles sont dites « naturelles » [12], et reposent sur des affordances ou des comportements facilement appréhendables et des vocabulaires restreints. Cela permet à l'utilisateur novice d'obtenir rapidement un résultat et à l'utilisateur expert d'internaliser et de mémoriser un minimum de connaissances. En effet, les clefs pour l'interaction sont presque intégralement comprises dans le comportement visuel de l'interface (représentation graphique).

A l'inverse, plusieurs artefacts permettent une expressivité incontestable malgré des règles d'interaction invisibles a priori et que l'on ne découvre que par l'apprentissage. C'est le cas par exemple de la trompette, dont la morphologie ne montre pas que la note jouée dépend aussi de la configuration de la cavité buccale. De nombreux exemples de pratiques expertes soutiennent cette démarche tournée vers une interaction moins triviale et nécessitant une expertise plus poussée, mais qui peut s'avérer adaptée à des besoins spécifiques en efficacité ou en expressivité : la virtuosité des musiciens, l'utilisation experte de raccourcis clavier ou de « *chorded keyboards* », la rapidité des sténotypistes, etc. Étonnamment, cette approche reste marginale dans le domaine des interactions multi-tactiles.

Des interactions par « accords » étaient déjà présentes dans la démonstration de Douglas Engelbart en 1968 [5] (clip 13). Cette démarche avait également mené la société FingerWorks, fondée en 1998, au développement de dispositifs permettant de déclencher des commandes par un riche vocabulaire de postures et gestes multi-doigts, avant d'être intégrée à Apple. Enfin, ce type d'interactions expertes multi-doigts a été exploré sur les surfaces tactiles notamment par Wu et al. [13], mais un problème reste sans réponse : quelles sont les règles pour définir les accords du vocabulaire, et comment assister l'utilisateur dans son apprentissage ?

Conception de vocabulaires de postures multi-doigts

Le point de départ pour la définition d'un vocabulaire riche fut l'étude des capacités motrices de la main [9,7] afin d'identifier les postures à éviter pour minimiser l'inconfort de l'utilisateur.

Tout d'abord, les doigts ne sont pas égaux en termes de mobilité : l'index a plus de force que l'auriculaire, le pouce est plus indépendant que l'annulaire, etc. Notre point de départ fut la position « naturelle » de la main : la paume parallèle à la table tactile et chaque doigt dans sa position de repos. Ensuite, ce premier vocabulaire fut

enrichi d'autres accords, en déplaçant certains doigts de leur position de repos. Les doigtés ainsi créés permettent de saisir les accords aussi confortablement que possible. Nous avons finalement extrait deux règles de cette étude des contraintes mécaniques de la main (figure 2) :

- Exclure les postures pour lesquelles l'annulaire ou le majeur est levé de la surface tactile sans qu'au moins un des doigts voisins ne soit levé,
- Exploiter l'indépendance du pouce, de l'index et de l'auriculaire pour définir des gestes supplémentaires.



Figure 2 : Exemples de postures à trois doigts, vues du dessus et représentées sous la forme de « *cheat-sheets* ». Les deux premières respectent la première règle de conception, les deux suivantes tirent en plus parti de la deuxième règle.

Étude d'un vocabulaire et résultats obtenus

Avec un vocabulaire de 12 gestes créés grâce à ces règles [3], nous avons mené une expérimentation contrôlée visant à mesurer l'efficacité de ces postures multi-doigts pour déclencher des commandes, comparées à une simple palette de boutons tactiles. Dans le premier cas, une importante partie du temps nécessaire pour saisir la commande est due à l'interprétation de la représentation graphique, et sa traduction en mouvements à effectuer. Dans le second cas, la contrainte temporelle vient de l'obligation de se déplacer jusqu'à la palette de boutons.

En moyenne, les accords ont permis aux sujets d'être plus rapides qu'avec la palette de boutons. En revanche, le nombre d'erreurs de saisie est beaucoup plus important pour les accords. Ce dernier résultat suggère, comme nous le supposions, que même avec un vocabulaire a priori « bien conçu », l'utilisation de tels gestes requiert un apprentissage interactif.

Nous avons alors envisagé deux alternatives pour assister l'utilisateur dans la réalisation de postures multi-doigts. Premièrement, une approche centrée sur le vocabulaire nous a amené à concevoir un système de guidage et d'apprentissage adapté aux postures multi-doigts : *Arpege*. La seconde approche, explorée dans nos travaux actuels que nous présenterons à la fin, repose sur la conception de « widgets multi-doigts » qui renseignent à l'avance sur les postures ou gestes pris en compte.

REPRESENTATION ET GUIDAGE

Travaux de référence

Pour apprendre quels sont les gestes disponibles et les commandes associées, l'utilisateur peut être aidé en voyant : (i) ce qui lui est accessible à tout instant et

comment le faire («*feedforward*»); (ii) son avancée dans la réalisation du geste prise en compte par le système, ou le résultat de son action («*feedback*»).

Différentes approches ont été envisagées pour traiter le problème de l'apprentissage des gestes multi-doigts. Certains systèmes, comme les TouchGhosts [10] ou l'outil de configuration des trackpads d'Apple, consistent en une simple vidéo et misent tout sur l'apprentissage par observation. Les systèmes qui impliquent l'utilisateur sont toutefois plus communs et c'est sur ces derniers que se concentre notre étude. Tout d'abord, il existe quelques systèmes « hors-contexte » : l'utilisateur apprend les gestes dans une application à part avant de les réaliser en contexte [11].

Parmi les méthodes « en contexte », la plus répandue est le « cheat-sheet » (figure 2). C'est une collection de représentations iconographiques (statiques) de gestes. Peu de systèmes à la fois « en contexte » et « dynamiques » ont été développés pour les écrans multi-tactiles. Le récent ShadowGuides [6] fournit un *feedforward* et un *feedback* à l'utilisateur, mais la représentation dynamique est déportée de la main et une importante partie des informations permettant de réaliser le geste reste contenue dans un cheat-sheet statique.

Les techniques d'aides à la fois dynamiques, « en contexte » et « in-situ » (à l'endroit où l'interaction est réalisée) permettent une transition fluide entre les stades novice et expert. Elles sont accessibles à la demande de l'utilisateur (économisant ainsi l'espace d'affichage que requiert un cheat-sheet) et jouent le rôle de guides lors de la réalisation du geste. Une fois la décomposition du geste assimilée, l'utilisateur est libre de faire le geste à une vitesse plus élevée, sans invoquer l'aide du guide. C'est le cas des Marking Menus [8], du Multitouch Menu [1] ou d'Octopocus [2]. C'est cette approche que nous avons transposée au cas des postures multi-doigts.

Arpège : Feedforward et feedback

Le principe d'Arpege est donc de fournir un guide contextuel dynamique, à la demande de l'utilisateur. Lorsqu'il invoque Arpege, cinq points sont affichés, représentant les cinq doigts de la main. Au dessus de chacun, on peut lire la liste des commandes dont « l'arpège » commence par ce doigt (figure 3.a). Le premier doigt est celui de l'accord qui est le plus éloigné du pouce, pour limiter l'occlusion d'Arpege par la main.

L'utilisateur pose le premier doigt, et Arpege affiche en blanc la position du prochain à placer, pour chacune des commandes encore accessibles. Ces positions sont affichées en opaque, et le reste des accords encore accessibles est en transparence (figure 3.b). À chaque étape, les commandes et les gestes qui ne sont plus accessibles disparaissent (figure 3.c où l'on ne voit plus la commande « copy »). Lorsque tous les doigts requis pour la commande souhaitée ont été posés aux bonnes

positions, le nom de la commande qui va être déclenchée est signalé (figure 3.d). Si un doigt posé est levé de la table, le système retourne à l'état dans lequel il était avant que ce doigt ne soit posé.

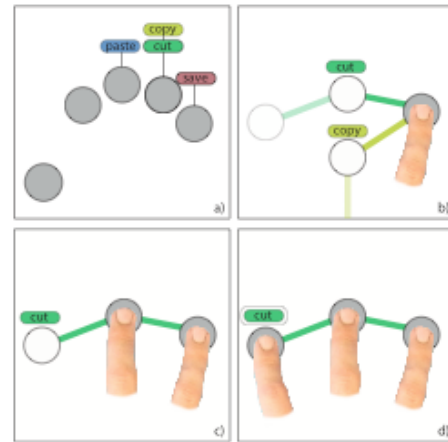


Figure 3 : Étapes de l'utilisation du guide dynamique Arpege pour saisir la commande « Cut ».

Pour permettre une transition fluide entre le mode novice et le mode expert, le déclenchement de la commande est finalement effectué par l'utilisateur comme il le ferait sans Arpege : une fois tous les doigts posés, il doit donc les relever de la table et les reposer tous en même temps, pour saisir l'accord lui-même plutôt que son « arpège ».

Deux gestes peuvent partager des positions de doigts (figure 4.a). Il est également possible qu'un doigt ne fasse pas partie d'un accord. Dans ce cas, Arpege affiche un point de la même couleur que le tracé, et la position du doigt suivant en blanc (figure 4.b).

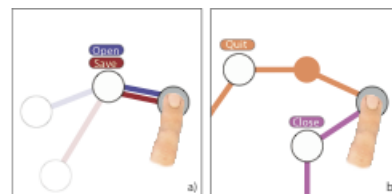


Figure 4 : Cas particuliers représentés par Arpege : a) deux accords ont des points en commun. b) un doigt ne doit pas être posé sur la surface tactile.

Évaluation d'Arpege et résultats obtenus

Nous avons réalisé une expérimentation contrôlée pour comparer Arpege à la méthode d'apprentissage la plus répandue : les cheat-sheets. Les résultats ne montrent pas de différence significative entre les deux méthodes pour le temps de saisie, la fréquence d'utilisation de l'aide, et l'efficacité de l'apprentissage. En revanche, Arpege permet de diviser le nombre d'erreurs par trois. Dans le cas des cheat-sheets, l'utilisateur adopte une stratégie par essais-erreurs qui peut être très pénalisante lors d'un apprentissage en contexte. Avec Arpege, il va découvrir le geste progressivement mais évitera les erreurs.

La seconde approche, qui fait l'objet de nos travaux en cours, consiste en la conception de « widgets multi-doigts », une solution intermédiaire entre la manipulation directe et les vocabulaires de postures multi-doigts abstraites.

REPRESENTATION ET AFFORDANCES

Travaux de référence

Arpege étant un système centré sur le geste, il ne fournit pas d'informations supplémentaires sur le résultat de l'action et son contexte. Afin de proposer d'autres éléments de représentation, nous étudions la conception de « widgets multi-doigts ». Comme défini dans l'interaction instrumentale [4], leur rôle est de recevoir l'action, de représenter leurs propres réactions, et de transmettre ensuite la commande à l'objet sur lequel on veut agir. Le multitouch menu [1] et le menu multi-doigt de Wu et al. [13] sont de tels widgets.

Nous explorons également les widgets qui ne déclenchent pas que des commandes discrètes. Le *Lemur* de *JazzMutant* propose un widget multi-doigt appelé *Multiballs*. Il consiste en une zone rectangulaire dans laquelle des cercles peuvent être créés, déplacés, et lancés (avec une inertie simulée). Ensuite, des contraintes masse-ressort sont créées entre les points de contacts et les cercles existants, de manière à avoir un contrôle fin en position ou créer des mouvements d'oscillation. Les positions des cercles sont autant de paramètres de contrôle pour l'application-cible.

Widgets multi-doigt

Nous travaillons actuellement à la catégorisation de tels widgets afin de proposer un cadre de conception adapté. Les dimensions envisagées sont par exemple le type de contraintes spatiales amenées par le widget, l'évolution de leur morphologie, le type de contrôle accessible, etc. Enfin, à partir de cette catégorisation, nous proposons d'étudier :

- La réification d'interactions multi-doigts déjà répandues (« *pinch* », « *rotate* », « *zoom* », etc.)
- La conception et l'évaluation de widgets multi-doigts adaptés : (i) aux degrés de liberté de la main ; (ii) aux vocabulaires générés grâce à nos règles

Avec les contraintes morphologiques liées aux widgets, il sera difficile d'atteindre la richesse des vocabulaires présentés précédemment, mais le niveau d'expertise nécessaire sera moins élevé car l'utilisateur sera guidé en amont par la morphologie du widget. Ce guidage est à mi-chemin entre les interactions dites « naturelles » où les actions sont limitées, et les gestes totalement libres.

CONCLUSION ET TRAVAUX FUTURS

L'étude menée avec Arpege nous a permis de vérifier que les postures multi-doigts étaient efficaces, et qu'un système de guidage contextuel pouvait réduire le nombre d'erreurs générées. Nous envisageons d'étendre ce

travail à tous les gestes multi-doigts, en combinant *Arpege* et *Octopocus*.

La deuxième étude présentée vise à la représentation graphique non seulement du geste à apprendre, mais aussi de l'interaction. Ces deux travaux ouvrent la voie à des alternatives à la manipulation directe souvent proposée sur les écrans multi-tactiles. Dans le premier cas, l'indirection vient du fait que les postures sont abstraites et libres dans l'espace, dans le second il y a en plus un intermédiaire, comme pour l'interaction instrumentale [4].

BIBLIOGRAPHIE

1. Bailly, G., Demeure, A., Lecolinet, E., and Nigay, L. 2008. Multitouch Menu (MTM). Proc. IHM '08.
2. Bau, O. and Mackay, W. E. 2008. Octopocus: a dynamic guide for learning gesture-based command sets. In Proc. UIST '08. ACM, 37–46.
3. Bau, O., Ghomi, E. and Mackay, W. E. 2010. Arpege: Design and learning of multi-finger chords gestures. Submitted to ACM TOCHI, 2010.
4. Beaudouin-Lafon, M. Instrumental interaction: an interaction model for designing post-wimp user interfaces. In proc. CHI '00: ACM, 446–453.
5. Engelbart, D. 1968. A research center for augmenting human intellect (The Demo). At <http://sloan.stanford.edu/MouseSite/1968Demo.html>
6. Freeman, D., Benko, H., Morris, M. R., and Wigdor, D. Shadowguides: Visualizations for in-situ learning of multi-touch and whole-hand gestures. In Proc. ITS'09.
7. Hager-Ross, C. and Schieber M. Quantifying the independence of human finger movements: Comparisons of digits, hands, and movement frequencies. Journal of Neuroscience. 2000.
8. Kurtenbach, G. P. The design and evaluation of marking menus. PhD thesis. 1993.
9. Lee, J. and Kunii, T. L. 1995. Model-based analysis of hand posture. IEEE Computer Graphics and Applications 15, 5, 77–86.
10. Vanacken, D., Demeure, A., Luyten, K., and Coninx, K. 2008. Ghosts in the interface: Meta-user interface visualizations as guides for multi-touch interaction. Proc. TABLETOP '08, 81–84.
11. Westerman, W. 2008. Gesture learning. freepatentsonline.com.
12. Wilson, A. D., Izadi, S., Hilliges, O., Garcia-Mendoza, A., and Kirk, D. 2008. Bringing physics to the surface. In Proc. UIST '08. ACM, 67–76
13. Wu, M. and Balakrishnan, R. 2003. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In Proc. UIST '03. ACM, 193–202.