



HAL
open science

Caractériser la transition des menus vers les raccourcis claviers

Gilles Bailly, Emmanouil Giannidakis, Marion Morel, Catherine Achard

► **To cite this version:**

Gilles Bailly, Emmanouil Giannidakis, Marion Morel, Catherine Achard. Caractériser la transition des menus vers les raccourcis claviers. 30eme conférence francophone sur l'interaction homme-machine, Oct 2018, Brest, France. pp.30-41. hal-01898856

HAL Id: hal-01898856

<https://hal.science/hal-01898856>

Submitted on 27 Oct 2018

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.

Caractériser la transition des menus vers les raccourcis claviers

Characterize the transition from menus to hotkeys

Gilles Bailly

Sorbonne Université, CNRS, ISIR
gilles.bailly@upmc.fr

Marion Morel

Sorbonne Université, CNRS, ISIR
mmorel@isir.upmc.fr

Emmanouil Giannidakis

Télécom ParisTech Université Paris-Saclay
emmanouil.giannidakis@telecom.paristech.fr

Catherine Achard

Sorbonne Université, CNRS, ISIR
catherine.achard@upmc.fr

ABSTRACT

This paper aims at better understanding the transition from menus to shortcuts. We first discuss the limitations and the opportunities of the theoretical and empirical characterizations of this transition. We then consider keyboard shortcuts as a case study and manually annotate empirical data to estimate several behavioral markers such as the initial switch, the transition duration or the performance dip. These markers serve to precisely characterize and compare three interaction techniques. Finally, we compare two methods to automatically characterize the transition from menus to shortcuts.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**;

KEYWORDS

keyboard shortcuts, menus, skill acquisition

RÉSUMÉ

Ce papier vise à mieux comprendre la transition des menus vers les raccourcis. Nous discutons d'abord les limitations et les opportunités des caractérisations, théoriques et empiriques, de cette transition. Nous considérons ensuite les raccourcis claviers comme cas d'étude où nous annotons manuellement une collection de données pour estimer des marqueurs comportementaux comme le début de la transition, sa durée ou la chute de performance. Ceci permet de caractériser finement la transition de trois techniques d'interaction. Finalement, nous comparons deux méthodes pour caractériser automatiquement la transition des menus vers les raccourcis.

MOTS-CLEFS

raccourcis claviers, menus, expertise

1 INTRODUCTION

De nombreux utilisateurs continuent d'utiliser les menus [14], alors que les raccourcis claviers (ou gestuels) sont plus efficaces. Il résulte que de nombreuses techniques d'interaction ont été proposées pour faciliter la *transition* des menus vers les raccourcis claviers [22, 26, 30, 36, 37, 44, 46], gestuels [4, 6, 9, 31, 49] ou adaptatifs [12]. Pour autant, on connaît encore mal les caractéristiques de cette transition.

Dans ce papier, nous visons à mieux caractériser la transition entre les menus et les raccourcis (claviers ou gestuels) afin de pouvoir répondre à des questions comme :

- Quand est ce que l'utilisateur commence la transition ?
- Quelle est la durée de la transition ?
- Comment évolue la performance (temps de sélection, erreur) avant, pendant et après la transition ?

À cette fin, nous mettons d'abord en évidence le fossé entre les caractérisations théoriques de la transition menus-raccourcis et les caractérisations empiriques. Nous proposons ensuite plusieurs marqueurs comportementaux à estimer pour caractériser plus finement cette transition.

Nous considérons comme cas d'étude celui des raccourcis claviers et la collection de données issues de [26]. Nous étudions en particulier les micro-dynamiques [24] de la transition au niveau de chaque commande de chaque utilisateur. Ce niveau de granularité plus fin est nécessaire pour capturer certains comportements qui sont masqués lorsque les données sont agrégées [24].

En particulier, nous avons annoté manuellement 588 séquences de sélections afin de déterminer le début et la fin de chaque transition. Nous exploitons ces données pour estimer par exemple la durée de la transition, la chute de performance ou le gain de performance. Nous utilisons ces marqueurs comportementaux pour comparer les trois techniques d'interaction présentée dans [26]

Cette analyse raffine les résultats précédents. Elle suggère en particulier que (1) la fréquence n'a pas d'effet sur le début de la transition et sur la durée de la transition ; (2) les techniques d'interaction ont un effet sur le début de la transition, mais pas sur la durée ; (3) la chute de performance est d'environ 3s et ne dépend pas de la technique ; (4) les utilisateurs commencent la transition avant d'avoir atteint le palier de performance.

Enfin, pour réduire le coût de l'annotation manuelle de la transition, nous présentons une méthode automatique s'appuyant sur les réseaux de neurones, l'évaluons sur notre collection de données et la comparons au modèle de Markov caché (HMMs).

2 ÉTAT DE L'ART

Ce projet est ancré dans le domaine des techniques d'interaction pour la sélection de commandes. Nous discutons aussi les travaux sur le développement de l'expertise. Avant cela, nous fixons la terminologie et discutons le problème que nous adressons.

Terminologie. Les termes stratégie, choix (économie), méthode (psychologie) ou modalité (IHM) sont souvent utilisés avec un sens proche mais dépendant de la communauté. Nous utilisons ici le terme *méthode* pour le choix de la technique d'interaction (menu ou raccourci) et *stratégie* pour différencier quatre comportements liés à une utilisation différente des méthodes :

- L'utilisateur a utilisé la méthode menu uniquement ;
- Il a sélectionné une commande dans le menu mais a enfoncé un modificateur (e.g. CTRL ou SHIFT) avant ou pendant l'ouverture du menu ;
- Il a ouvert le menu mais a exécuté un raccourci ;
- Il a utilisé uniquement la méthode raccourci.

Utilisation d'interactions sous-optimales. Plusieurs théories ont été proposées pour expliquer le fait que les utilisateurs interagissent de manière sous-optimale avec les systèmes interactifs, même avec plusieurs années d'expérience [10, 11, 40]. Par exemple, l'effet d'Einstellung [35] stipule que les utilisateurs préfèrent utiliser les stratégies qu'ils ont déjà expérimentées plutôt que d'essayer d'en découvrir des plus efficaces. Carroll et Rosson [13] suggèrent que ce comportement est lié à deux biais cognitifs : le biais d'*assimilation* qui est la tendance d'un utilisateur à interpréter une nouvelle situation à partir de ses expériences passées et le biais de *production* qui est le fait d'être concentré à terminer la tâche courante et donc ne pas prendre le temps d'apprendre d'autres méthodes pour réaliser leur objectif les prochaines fois. C'est le paradoxe de l'utilisateur actif : favoriser la productivité à court terme plutôt que la performance sur le long terme [13]. Enfin, Gray et ses collègues [19, 23, 25] se sont concentrés sur les micro-stratégies pour terminer une tâche et suggèrent que les utilisateurs sont prédisposés à choisir la stratégie nécessitant le moins d'efforts mentaux pour cela. Ils ont également remarqué que les utilisateurs préfèrent généralement les stratégies (1) bien assimilées et génériques pour être appliquées dans différentes situations ou (2) décomposables en sous-stratégies offrant des retours d'information rapides et incrémentaux.

Techniques d'interaction. De nombreuses techniques d'interaction ont été proposées pour encourager l'utilisation des raccourcis claviers [7, 22, 26, 29, 36, 37, 41, 43] ou gestuels [6, 9, 27, 31, 49]. Ces techniques reposent sur l'utilisation de retours visuels [22, 26, 29, 37, 43] de retours auditifs [26], de méthodes de guidage renforçant la répétition moteur [6, 9, 22, 27, 31, 36, 49], de pénalité [26, 36] ou de nouveaux vocabulaires d'entrée [6, 7, 49]. Dans cet article, nous revisitons deux techniques d'interaction [26] : la première indique de manière sonore le raccourci clavier associé à une commande lorsque celle-ci est sélectionnée dans le menu. La seconde désactive le menu. L'utilisateur peut naviguer dans le menu, mais ne peut plus cliquer les items le forçant à utiliser les raccourcis claviers.

Développement de l'expertise. Le cadre conceptuel décrit dans [44] analyse quatre domaines de développement de l'expertise : (1) le développement intramodal ; (2) le développement intermodal ; (3) l'extension du vocabulaire et (4) l'association de tâche. Bien que cet article se concentre principalement sur le développement intermodal, nous décrivons aussi les résultats liés à l'amélioration intramodal.

1. Développement intramodal. La courbe d'apprentissage ou courbe de performance (Power Law of Practice en anglais) [39] peut être écrite sous la forme : $y = ax^b + c$, où y est typiquement le temps de sélection, x est la quantité de pratique et a, b, c trois paramètres. Bien que cette courbe s'ajuste bien à la performance au niveau d'un groupe d'utilisateurs, la fonction exponentielle $y = ae^b(x-1) + c$ offre un meilleur ajustement au niveau de l'individu. Conceptuellement, il est fréquent de distinguer trois phases [14] : (1) la performance initiale : à ce stade, l'utilisateur a une connaissance limitée de l'application ; (2) l'apprentissage : l'utilisateur devient de plus en plus efficace mais l'interaction nécessite un certain effort et (3) la performance ultime : l'utilisateur a atteint un plateau de performance. Ces trois phases peuvent être mises en relation avec celles du modèle des stades de l'apprentissage : Cognitive, Associative et Automatique [3, 18, 45].

2. Développement intermodal. Scarr et al. [44] identifient trois raisons expliquant pourquoi les utilisateurs ne font pas la transition d'une méthode à une autre. (1) Ils peuvent simplement ne pas savoir qu'une méthode alternative (et plus efficace) existe. (2) Ils peuvent sous-estimer les bénéfices de cette méthode. (3) Ils peuvent ne pas être prêts à faire face à la chute de performance *temporaire* lors du changement de méthode.

Gray et Lindstedt [24] généralisent ces deux concepts (développement intra et intermodal) à différents types de stratégies avec le cadre Plateau, Chute et Rebond. Les utilisateurs sont dans des plateaux de performance jusqu'à ce qu'ils expérimentent d'autres stratégies mises en évidence par des chutes et des rebonds de performance. Ils peuvent avoir plusieurs *plateaux* de performance avant d'atteindre une *asymptote* de performance, c'est à dire une performance ultime pour une tâche et un environnement donné.

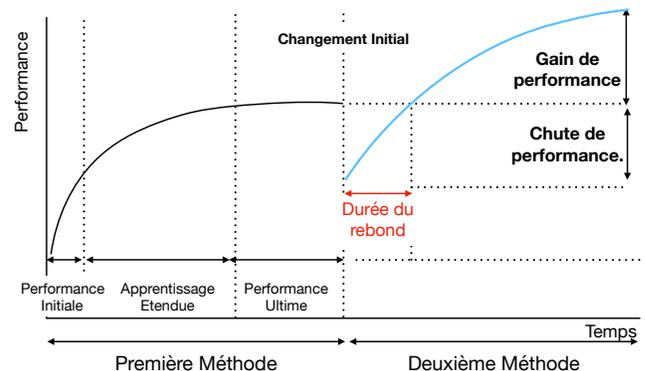


FIGURE 1: Cadre conceptuel [44] décrivant le développement d'expertise intra- et inter-méthode. Les principaux marqueurs comportementaux sont indiqués en gras. Nous en introduisons un additionnel : Durée du rebond (rouge).

3 TRANSITION MENUS - RACCOURCIS

Comment décrire la transition entre deux méthodes ? Nous illustrons ici le fossé entre les caractérisations théoriques issues de la littérature et celles utilisées dans les études empiriques pour comparer différentes techniques d'interaction.

3.1 Caractérisations théoriques

Les cadres conceptuels comme celui de l'expertise [44] ou DPL [24] offrent plusieurs marqueurs comportementaux permettant de caractériser la transition entre deux méthodes (voir Figure 1) :

- **Changement initial.** Ce marqueur indique la date à laquelle l'utilisateur change de modalité.
- **Chute de performance.** Ce marqueur estime la chute de performance lors du changement de méthode.
- **Gain de performance.** Le gain de performance peut être estimé comme la différence de performance ultime entre la première et la seconde méthode.
- **Durée du rebond.** En s'appuyant sur la Figure 1, nous introduisons la durée de rebond comme nouveau marqueur comportemental. Ce marqueur indique le temps nécessaire pour retrouver sa performance d'avant transition.

Ces cadres conceptuels offrent une approximation utile pour caractériser la transition entre les menus et les raccourcis. Cependant, ils restent des approximations théoriques qu'il faut considérer avec précautions (e.g. est ce que l'utilisateur initie la transition qu'une fois le plateau de performance atteint ?). Ensuite, la *perception* de la performance future est probablement plus pertinent que le gain de performance *réel* sur la décision de changer de méthode [14]. Plus important, ces cadres conceptuels suggèrent que la transition peut être définie comme une date séparant l'utilisation de la première méthode (avant) de la seconde méthode (après). Cette approche est totalement valide lorsque les utilisateurs sont *forcés* de changer de méthodes. Par exemple, lorsqu'ils remplacent leur clavier AZERTY pour un nouveau clavier QWERTY. Cependant, dans le cas des commandes¹, les menus et les raccourcis coexistent et les utilisateurs peuvent facilement alterner entre ces deux méthodes. En particulier, après une erreur avec un raccourci, ils peuvent réutiliser les menus pendant un temps afin d'éviter de refaire une erreur.

Dans ce papier, nous argumentons que la transition ne doit pas être définie comme une date, mais comme une période. Nous remplaçons alors le marqueur comportemental "changement initial" par deux nouveaux marqueurs comportementaux :

- **Début de la transition.** Nous définissons ce marqueur comme la date de la fin de la période de stabilité² avec la première méthode (c.à.d menus).
- **Durée de la transition.** C'est la durée de la période d'instabilité des méthodes utilisées et/ou de la réussite. Elle se termine lorsque l'utilisateur est de nouveau stable, mais cette fois-ci avec la seconde méthode (c.à.d les raccourcis). Nous verrons dans la section 6 que cette période est singulière avec un entremêlement des stratégies

1. Il est aussi possible d'avoir en permanence à sa disposition deux types de claviers.

2. La notion de stabilité est discutée dans la section 5

3.2 Caractérisations empiriques

De nombreuses études ont comparé des techniques d'interaction favorisant l'usage des raccourcis claviers [22, 26, 36, 37] ou gestuels [4, 27]. La principale **mesure** utilisée consistait à estimer le pourcentage de raccourcis utilisés durant l'étude expérimentale (Figure 2-gauche). Cependant, nous avons identifié des variations dans la définition dans cette mesure liées à la manière d'agréger les données. Nous présentons d'abord les deux principales variantes utilisées dans les travaux existants. Nous discutons ensuite deux autres dimensions pour lesquelles les données ont été agrégées mais qui mériteraient davantage d'attention.

Précision. Nous avons identifié comme mesure le pourcentage d'utilisation des raccourcis utilisés [26, 36] ou le pourcentage d'utilisation *réussie* des raccourcis [4, 6]. La première mesure traduit davantage l'intention de l'utilisateur à adopter les raccourcis. Cependant elle n'informe pas s'il a réellement réussi à les utiliser et donc s'il a réussi la transition.

Dans une évaluation contrôlée pénalisant peu ou pas les erreurs, le participant peut être incité à utiliser les raccourcis, même sans les connaître pour gagner du temps sur l'expérience. Dans ce cas, il semble plus pertinent d'utiliser les raccourcis réussis, car ils montrent que l'utilisateur s'investit dans la tâche. Sinon, considérer les raccourcis (réussis ou pas) fournit plus d'information sur le début de la transition : abandon des menus (même avec des erreurs).

Temporalité. Une alternative est de décrire l'évolution du pourcentage d'utilisation (réussi ou pas) des raccourcis au cours du temps [4, 36, 37]. Cette représentation est moins concise, mais reflète davantage la dynamique du comportement des utilisateurs et peut montrer des interactions subtiles entre Technique et Temps. Par exemple, deux techniques d'interaction peuvent avoir le même pourcentage global d'utilisation des raccourcis, mais des dynamiques différentes (e.g. évolution lente vs. progression rapide mais tardive).

Utilisateur. Il est courant en IHM d'agréger les participants pour diminuer la variabilité inter-participants. Cependant, cette pratique est déconseillée lorsque l'on étudie des phénomènes liés à l'apprentissage [15, 24]. En effet, agréger les données de plusieurs utilisateurs permet d'extraire une courbe d'apprentissage "lisse", plus facile à interpréter, mais qui masque plusieurs comportements informatifs comme des séquences de plateaux, chutes et rebonds de performances [15, 24].

Conditions : commande et fréquence. De la même façon, il est courant d'agréger les commandes avant analyse dans la mesure où nous ne sommes pas intéressés par les différences inter-commandes. Cependant, la plupart des protocoles expérimentaux étudiant la transition vers les raccourcis [4, 26, 36] assignent des fréquences différentes à chaque commande, un facteur pouvant avoir un impact important sur l'apprentissage [34]. La manière d'agréger ou non ces différents facteurs peut amener à différentes interprétations (voir section 6).

En résumé, il existe plusieurs caractérisations empiriques pour décrire la capacité de l'utilisateur à faire la transition des menus

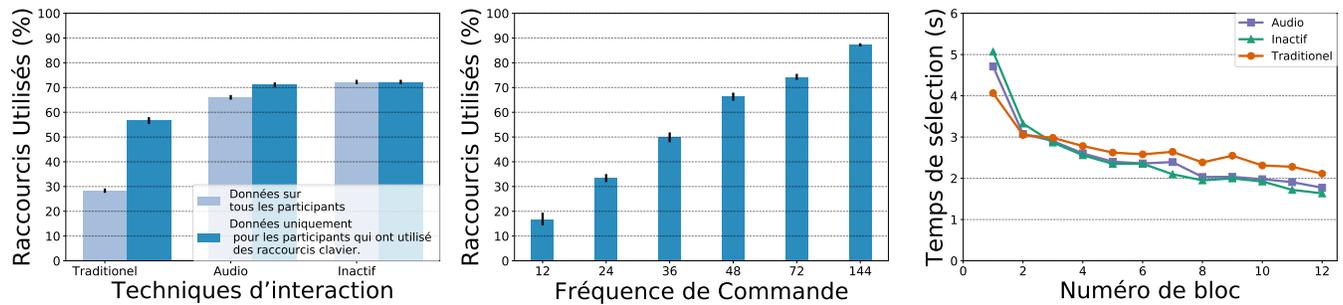


FIGURE 2: Résultats de l'étude présentée dans [26]. Utilisation des raccourcis claviers par technique et fréquence avec des intervalles de confiance à 95% ainsi que le temps de sélection par bloc et technique.

vers les raccourcis (claviers ou gestuels). Aujourd'hui, la principale caractérisation consiste à reporter l'évolution de l'utilisation (réussie ou non réussie) des raccourcis en fonction du temps (ou bloc). Bien que cette caractérisation soit utile, elle masque plusieurs phénomènes de la transition qui ne peuvent être observés qu'au niveau de l'individu comme la durée de la transition ou du rebond.

Une autre mesure utilisée dans ces études est l'évolution du temps de sélection au cours du temps (bloc) (voir Figure 2-droite). Cette caractérisation est particulièrement utile car elle couvre un aspect important de l'utilisabilité (vitesse). Cependant, elle ne permet pas de capturer des phénomènes liés à la transition comme la chute temporaire de performance.

3.3 Discussion

De cette étude sur la caractérisation de la transition entre les menus et les raccourcis, nous tirons les enseignements suivants :

- (1) *Le fossé entre les caractérisations théoriques et empiriques.* Au niveau des marqueurs comportementaux, les premières se concentrent sur la date de changement de méthode, la chute de performance et le gain de performance. Les secondes se concentrent sur l'évolution de l'utilisation des raccourcis. Au niveau de la définition de la transition, les premières laissent penser que la transition peut être perçue comme un point dans le temps, alors que les secondes laissent penser que l'utilisateur a une progression continue du taux d'utilisation des raccourcis.
- (2) *Le besoin d'analyser les données avant de les agréger.* Agréger les données d'un groupe d'utilisateurs / commandes supprime du bruit lié à la variabilité des utilisateurs, mais masque des phénomènes liés à l'apprentissage comme des séquences de plateaux, chutes et rebonds de performance [24].
- (3) *Le manque d'outils pour caractériser la transition.* Il apparaît que l'évolution du taux d'utilisation réussie des raccourcis est insuffisante pour décrire précisément la transition. Bien que les caractérisations théoriques présentent des marqueurs comportementaux, elles n'indiquent pas comment les estimer à partir de données empiriques. C'est l'objet des sections 5 et 7.1. Mais avant cela, nous présentons la collection de données sur laquelle nous allons utiliser notre méthode.

4 CAS D'ÉTUDE : RACCOURCIS CLAVIERS

Pour notre cas d'étude sur les raccourcis claviers, nous utilisons la collection de données présentée dans [26]. Nous résumons ici brièvement l'étude correspondante.

Plan expérimental. La tâche consistait à sélectionner une commande cible dans l'un des six menus déroulants (non hiérarchiques) de la barre de menus ou à l'aide d'un raccourci clavier le plus rapidement et le plus précisément possible. Le stimulus était une image affichée en bas de l'écran représentant la commande à exécuter. Une commande était un animal, un fruit, un légume, etc. La lettre utilisée dans un raccourci clavier n'appartenait pas au nom de la commande pour éviter des effets de familiarité. Si l'utilisateur ne sélectionnait pas la bonne commande, il devait de nouveau exécuter la commande cible après une pénalité de 3s.

Trois techniques étaient comparées : (1) TRADITIONNEL. Le menu déroulant traditionnel avec un renforcement visuel : l'item (nom de la commande + le raccourci) sélectionné disparaissait progressivement alors que le reste du menu se masquait immédiatement ; (2) AUDIO. Le menu traditionnel avec un renforcement audio : le raccourci clavier était joué oralement par un synthétiseur vocal une fois la commande exécutée à partir du menu ; (3) INACTIVE. Le menu traditionnel existait mais les clics dans le menu étaient désactivés afin de forcer les utilisateurs à utiliser les raccourcis claviers. 42 participants (14 par technique) ont participé à cette étude.

Les commandes (14 au total) avaient des fréquences différentes et n'apparaissaient donc pas le même nombre de fois. Elles pouvaient apparaître 144 fois (2 commandes), 72 fois (2), 48 fois (2), 36 fois (4), 24 fois (2) ou 12 fois (2) fois au total.

L'étude était divisée en 12 blocs de 30 sélections. L'ordre de présentation des commandes était pseudo-aléatoire afin que la fréquence de chaque commande soit respectée à chaque bloc (e.g les deux plus fréquentes commandes apparaissaient 12 fois par bloc alors que les deux moins fréquentes apparaissaient 1 fois).

Données collectées au format CSV. Pour chaque essai (ou sélection), les variables indépendantes étaient l'identifiant du participant, de la technique, du bloc, de l'essai au sein du bloc ainsi que les informations de la commande cible (libellé, position, fréquence, raccourci clavier). Les variables dépendantes étaient le temps (ms), la réussite (0/1), la stratégie utilisée (voir le paragraphe Terminologie).

Résumé de l'analyse des données. La figure 2 illustre les résultats de cette expérience comme présentés dans le papier original [26]. Elle représente en particulier à gauche, le pourcentage d'utilisation des raccourcis claviers utilisés (sans aide du menu) pour chaque technique d'interaction différenciant l'ensemble des utilisateurs de ceux qui ont utilisé au moins une fois les raccourcis claviers; au milieu, le pourcentage d'utilisation des raccourcis claviers en fonction de la fréquence des commandes; à droite, l'évolution du temps de sélection en fonction de la technique et du bloc.

Discussion. Cette étude reflète bien les expériences comparant des techniques d'interaction pour l'utilisation des raccourcis claviers ou gestuels. Les marqueurs comportementaux discutés dans les cadres théoriques (*changement initial*, *chute de performance*, *gain de performance*) ainsi que les trois autres que nous avons introduit (début et durée de la transition, durée de rebond) ne sont pas utilisés pour caractériser la transition. Dans la suite de cet article, nous approfondissons cette analyse. Pour ceci, il est d'abord nécessaire d'identifier le début et la fin de la transition.

5 ANNOTATION MANUELLE

L'objectif de cette annotation est d'identifier le début et la fin de la transition (le début peut coïncider avec la fin) afin d'étudier les différents marqueurs comportementaux sur les données de [26]. À notre connaissance, il n'existe pas de méthodes *automatiques* qui ont été validées pour le type de données considérées (transition des menus vers des raccourcis) [15, 24, 47, 48]. Nous réalisons donc cette annotation *manuellement*. Les données annotées pourront ensuite servir de vérité terrain pour développer/tester des méthodes d'annotation automatique (voir section 7).

5.1 Plan expérimental

Participants. Deux des auteurs ont annoté les 588 séquences (3 techniques \times 14 participants \times 14 commandes) d'abord de manière indépendante puis collectivement pour résoudre les conflits.

Tâche. La tâche consistait, si **présents**, à identifier le début et la fin de la transition pour chaque commande, représentée comme une séquence de sélections. En accord avec la section 3.1 :

- le début de la transition est défini comme la fin de la période de stabilité avec le menu.
- la fin de la transition est définie comme le début de la période de stabilité avec les raccourcis claviers.

À ce stade, il n'existait pas de définition précise du terme "stabilité" dans ce contexte. Elle était donc laissée à l'appréciation des annotateurs. Les annotateurs s'accordent cependant sur le fait qu'elle doit être entendue en termes de choix de méthode et précision. De plus, pour faciliter une annotation cohérente entre les annotateurs, ils s'accordent sur le fait que le début de la transition ne peut pas être une sélection réussie avec le menu et que la fin de la transition doit être une sélection réussie avec le raccourci clavier. L'impact du choix et de la compréhension des instructions par les annotateurs est discuté à la fin de cette section.

Interface. L'interface WEB d'annotation développée en python (serveur) et html/javascript (client) est téléchargeable à git.io/fNY7e

et est illustrée Figure 3. Elle comprend une page par technique et commande. Sur chaque page, une séquence de sélections est représentée par des points (entre 12 et 144 en fonction de la fréquence de la commande). La couleur des points encode la stratégie utilisée : Bleu (Menu seulement), Vert (raccourcis claviers seulement), Orange (Menu avec un modificateur enfoncé), Violet (Raccourci claviers ayant ouvert le menu). La croix indique une sélection incorrecte.

L'annotateur clique sur un point (c.à.d. une sélection) pour indiquer le début ou la fin de la transition. Il peut annuler ou changer son annotation en re-cliquant sur le point correspondant. Des traits verticaux représentent les transitions qui sont résumées dans la table à droite. Celle-ci est éditable et permet de différencier facilement une transition instantanée (début == fin) du cas où l'utilisateur commence la transition, mais sans la terminer. La table permet aussi d'indiquer le cas où l'utilisateur a commencé une transition, mais subitement abandonne et retourne définitivement aux menus.

5.2 Accord entre les annotateurs

Après avoir annoté chaque séquence de manière indépendante, nous avons étudié la fiabilité des annotations [38].

Début de la transition. les annotateurs avaient les mêmes résultats pour 543 des 588 séquences annotées (92.4%). Le coefficient de Cohen-Kappa correspondant est égal à 0.90³ Le consensus entre les deux annotateurs est donc tout à fait satisfaisant.

Fin de la transition. Les annotateurs avaient les mêmes résultats pour 534 des 588 séquences annotées (90.8%). Le coefficient de Cohen-Kappa est égal à 0.88. Bien que légèrement moins bon que celui pour le début de transition, il reste tout à fait satisfaisant.

Résolution des conflits. Pour résoudre les conflits, les annotateurs ont analysé ensemble chaque séquence ayant des annotations différentes. Nous avons identifié deux types de résolutions : Immédiate et après discussion.

Résolution immédiate. L'un des annotateurs estime immédiatement que l'annotation de l'autre annotateur est plus pertinente. Ce type de résolution représente 10/45 cas pour le début de la transition et 10/54 cas pour la fin de la transition.

Résolution après discussion. Chaque annotateur explique son choix et tente de convaincre l'autre. Les sources de conflits étaient multiples. Par exemple, est ce qu'un raccourci clavier entouré de plusieurs sélections par menu devait être considéré comme le début d'une transition ou juste comme un essai isolé ? Il pouvait y avoir également des effets de bords, spécialement pour les séquences courtes : comment considérer les dernières sélections d'une séquence ? Pour résoudre ces types de conflits, les annotateurs examinaient plus attentivement les autres séquences du même utilisateur afin de mieux comprendre son comportement. Ceci était suffisant pour s'accorder sur une décision.

3. Landis et al. [8, 32] considèrent que les valeurs supérieures à 0.81 reflètent un consensus presque parfait et Altman et al. [1, 20] considèrent que les valeurs supérieures à 0.80 reflètent un très bon consensus.

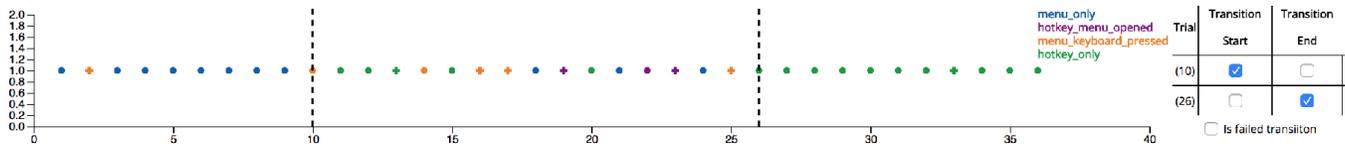


FIGURE 3: Interface d'annotation d'une commande d'un utilisateur. Une sélection est représentée par point. La couleur indique la stratégie et le symbole indique la réussite. Les lignes verticales indiquent le début et la fin de la transition

5.3 Discussion

Comme de nombreuses tâches d'annotation, il existe une part de subjectivité dans l'interprétation des instructions. Un point important est que bien qu'aucune définition n'ait été donnée pour le terme "stabilité", il y a un fort consensus sur le résultats des annotateurs. Un autre point important est le choix du nombre de phases. Nous avons eu de nombreuses discussions entre les auteurs et un expert (non-auteur) en sélection de commandes pour déterminer le nombre de phases à annoter. Il est en effet possible d'imaginer que la phase de transition puisse être divisée en plusieurs phases distinctes. Parmi celles envisagées, nous pouvons citer une phase "dominée" par l'usage des menus et une autre "dominée" par les raccourcis. Il a également été envisagé de différencier une phase où l'utilisateur "explore" la méthode alternative de la phase où l'utilisateur est davantage actif dans l'apprentissage du raccourci. Dans cet article, nous avons fait le choix de considérer la transition comme une simple entité laissant comme travail futur une décomposition plus fine de cette phase. Nous discuterons cependant dans la section suivante la présence d'une phase de pré-transition.

Cet exercice a permis d'identifier le début et la fin de la transition pour chaque commande de chaque utilisateur, c'est à dire avant d'agréger les données par commandes, utilisateurs, fréquences ou techniques. Ceci permet maintenant d'étudier plus finement la transition entre les menus et raccourcis claviers. Dans la section 7.1, nous présentons une méthode d'annotation automatique.

6 ANALYSE FINE DE LA TRANSITION

Méthodes statistiques. Suite aux récentes critiques concernant la p-value, nous préférons utiliser les intervalles de confiance et les tailles d'effets pour analyser nos données [17]. Pour chaque marqueur comportemental, nous représentons les résultats sous deux formes : (a) un diagramme en bar par technique ; (b) un diagramme comparant la différence des moyennes entre techniques. Une valeur positive indique que la technique de droite est plus rapide que la technique de gauche alors qu'une valeur négative indique le contraire. Une valeur qui intersecte l'axe 0 indique que même s'il y avait une différence entre les techniques, il n'est pas possible de la mesurer. Dans les deux cas (a et b), nous représentons la moyenne avec un intervalle de confiance (CI) à 95% indiquant la plage des valeurs possibles pour la population moyenne [17]. Tous les intervalles de confiance sont des 95% BCa Bootstrapé CIs [16].

Données. Nous ne considérons pas les commandes avec une fréquence égale à 12 car il n'y a pas suffisamment de séquences où l'utilisateur commence (et donc termine) la transition (Table 1). Pour les marqueurs comportementaux nécessitant que l'utilisateur

TABLE 1: Nombre de séquences où l'utilisateur a commencé et terminé la transition par technique et fréquence.

| Freq. | Traditionnel | | Audio | | Inactif | | Total |
|--------------|--------------|-------------|------------|------------|------------|------------|-------|
| | Debut | Fin | Debut | Fin | Debut | Fin | |
| 12 | 4 (14.3%) | 0 (0%) | 14 (50%) | 3 (10.7%) | 28 (100%) | 8 (28.6%) | 28 |
| 24 | 21 (37.5%) | 10 (17.86%) | 38 (67.9%) | 24 (42.9%) | 28 (100%) | 33 (58.9%) | 56 |
| 36 | 11 (39.3%) | 11 (39.3%) | 24 (85.7%) | 18 (64.3%) | 28 (100%) | 19 (67.7%) | 28 |
| 48 | 12 (42.9%) | 9 (32.1%) | 26 (92.9%) | 24 (85.7%) | 28 (100%) | 25 (89.3%) | 28 |
| 72 | 12 (42.9%) | 12 (42.9%) | 26 (92.9%) | 24 (85.7%) | 28 (100%) | 27 (96.5%) | 28 |
| 144 | 12 (42.9%) | 12 (42.9%) | 26 (92.9%) | 26 (92.9%) | 28 (100%) | 28 (100%) | 28 |
| Total | 68 | 44 | 140 | 92 | 140 | 99 | |

termine la transition (e.g. durée de la transition), nous ne considérons pas non plus les commandes avec la fréquence 24 pour la même raison. Les temps de sélection n'inclue pas la pénalité de 3s. Enfin, nous utilisons une transformation logarithmique des données pour le calcul des CIs pour réduire l'influence des valeurs extrêmes et l'asymétrie des distributions [17].

6.1 Transition

Début de la transition. La Figure 4-A-B compare les différentes techniques pour le marqueur "début de transition. On observe un effet important de la technique sur le début de transition : les utilisateurs commencent plus tôt la transition avec AUDIO (en moyenne après 6 sélections) que TRADITIONNEL (12 sélections). Dans le cas de INACTIF, la technique est conçue afin que l'utilisateur soit forcé de commencer la transition dès la première sélection. La Figure 4-c suggère que la fréquence n'a pas d'effet sur le début de la transition (larges barres d'erreurs et intersections avec l'axe 0).

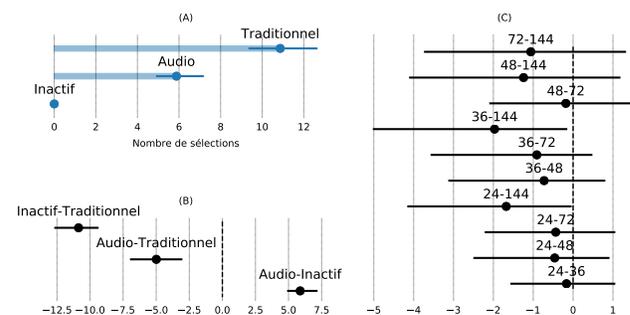


FIGURE 4: (A) Nombre de sélections avant transition. Différences des moyennes entre techniques (B) et fréquences (C).

Durée de la transition. La Figure 6-A-B compare la durée de la transition par technique. Elle suggère que les techniques n'ont pas d'effets sur la durée de la transition avec une moyenne de 17 (CI=[+/-2]) sélections. La Figure 6-c suggère également que la fréquence n'a pas d'effet sur la durée de la transition.

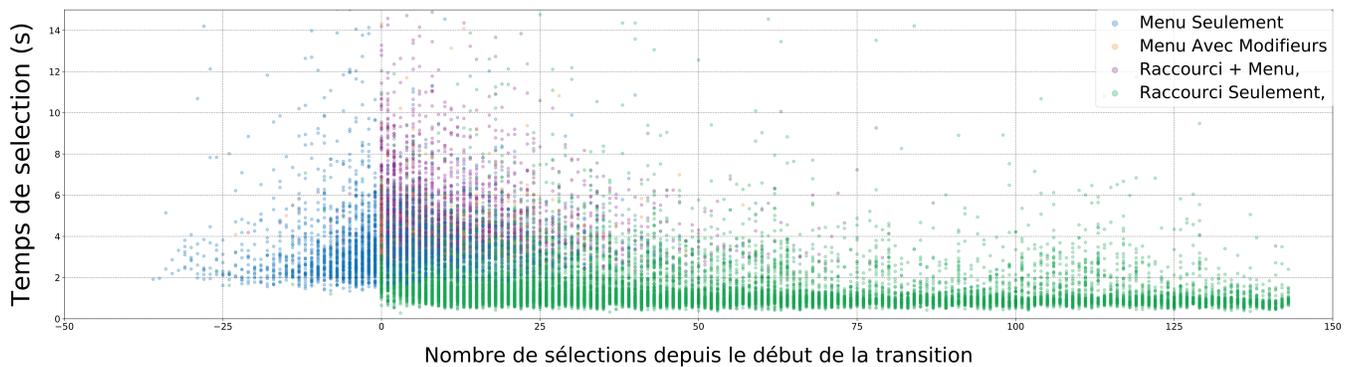


FIGURE 5: Représentation de toutes les séquences de commandes où l'utilisateur a entamé une transition. Les séquences sont alignées sur la date de transition. La couleur encode la stratégie

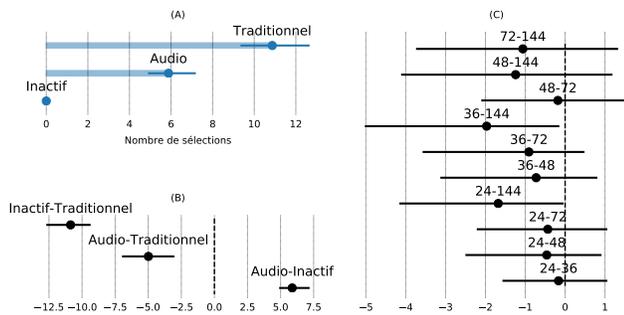


FIGURE 6: (A) Nombre de sélections pendant la transition par technique. Différence des moyennes entre (B) techniques et (C) fréquences.

Degré de stabilité. Nous estimons le degré de la stabilité de la phase de transition comme le nombre de changements de stratégies et le pourcentage d'erreur. Les changements de stratégies représentent 31.5% (CI=[-1.2%,+2.0%]) des sélections pendant la phase de transition contre 2.3% (CI=[-0.6%,+0.8%]) avant la transition et 1% (CI=[-0.3%,+0.4%]) après la transition. De la même manière, le pourcentage d'erreurs est de 14.1% (CI=[-1.4%,+1.5%]) pendant la transition contre 2.6% (CI=[-0.6%,+0.8%]) avant la transition et 2.2% (CI=[-0.3%,+0.4%]) après la transition.

6.2 Temps de sélection

La figure 5 représente le temps de sélection en fonction du nombre de sélections depuis le début de la transition. La couleur encode la stratégie utilisée. Nous détaillons cette figure à travers 2 marqueurs comportementaux.

Gain de performance. La figure 7 représente la distribution du temps de sélection (A) avant et après la transition ainsi que (B) pendant la transition. Pour le temps de sélection avec les menus dans la figure A, nous distinguons les utilisateurs ayant commencé la transition de ceux qui n'ont pas commencé la transition.

En dehors de la transition (A), on observe un faible recouvrement entre les raccourcis claviers d'une part (*median* = 1.2; CI = +/- 0.2) et les menus (*median* = 3.7s, CI = [-0.4, +0.2]) d'autre part,

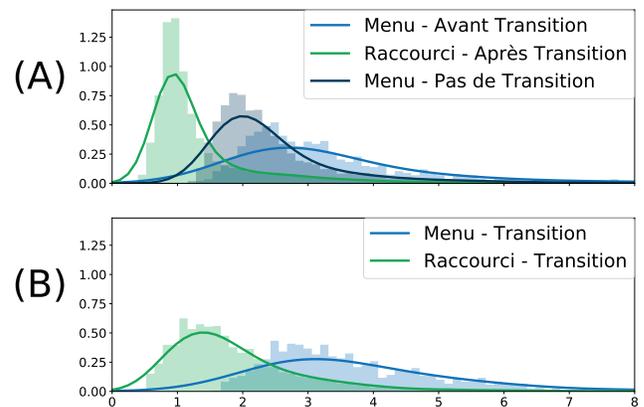


FIGURE 7: Distribution du temps de sélection par phases et par méthodes.

même pour les utilisateurs ayant pratiqué uniquement les menus (*median* = 2.4, CI = +/- 0.2). Ce recouvrement est plus important pendant la transition (B), principalement car la distribution des raccourcis claviers s'étend davantage vers la droite.

On observe également que les utilisateurs n'ayant pas fait de transition sont plus rapides avec les menus que ceux qui ont commencé la transition, probablement parce qu'ils ont pu pratiquer davantage cette méthode (courbe d'apprentissage).

Pour estimer plus précisément le gain de performance, nous proposons de mesurer la différence du meilleur temps de sélection avant et après la phase de transition pour chaque séquence. À cause de la variabilité d'une sélection à l'autre, nous avons observé qu'il était plus pertinent de considérer les deux sélections consécutives les plus rapides et de prendre la moyenne. La figure 8 montre que le gain de performance est 1.3s (CI= +/- 0.3s) en moyenne et ne dépend ni de la technique ni de la fréquence.

Chute de performance et pré-transition Alors que le temps de sélection avec les menus devrait diminuer avec la pratique (courbe de

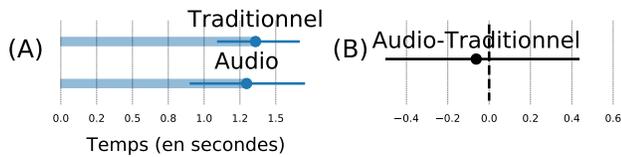


FIGURE 8: (A) Gain de performance par technique. (B) Différence des moyennes.

performance [39]), la figure 5 suggère une augmentation progressive du temps de sélection en approchant du début de la transition. Nous proposons alors de mesurer la différence entre le temps de sélection au début de la transition (moyenne du temps des deux premières sélections) et le meilleur temps consécutifs. La Figure 9-A-B indique une chute de performance de 1.4 secondes ($CI=[-0.5,+0.6]$) qui ne semble pas dépendre de la technique d'interaction (Inactif n'est pas considéré ici car elle n'a pas de phase Avant transition).

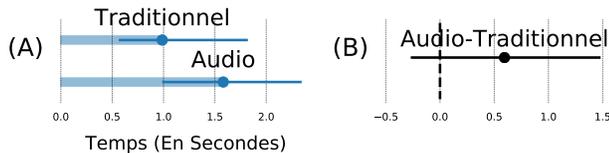


FIGURE 9: (A) Chute de performance par technique. (B) Différence des moyennes.

Nous raffinons l'étude de la chute de performance en analysant les quelques sélections précédant le début de la transition. Plus précisément, nous analysons la période séparant la meilleure sélection avec le menu et la première sélection de la phase de transition que nous appelons la phase de *pre-transition* (Figure 10). La durée de cette période est en moyenne 2.9 ($CI=[-0.4,+0.5]$) sélections avec AUDIO et 5.4 ($CI=[0.9,1.2]$) avec TRADITIONNEL.

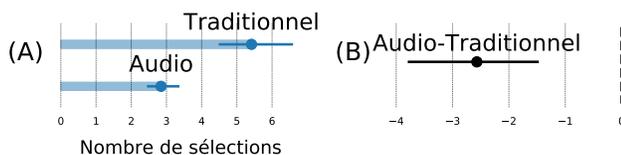


FIGURE 10: (A) : Nombre de sélections entre la meilleure performance avant la transition et le début de la transition. (B) La différence des moyennes. Les barres d'erreur représentent la CIs à 95%

6.3 Durée du rebond

Nous avons défini la durée du rebond comme le temps nécessaire pour retrouver la performance d'avant transition. Nous la mesurons comme le nombre de sélections séparant la meilleure sélection avant la transition et la première sélection plus rapide après la transition⁴

⁴ Dans les deux cas, nous considérons la moyenne du temps de deux sélections consécutives

(Figure 11). La durée du rebond est de 18 sélections ($CI=+/- 2$). Une analyse visuelle de chaque séquence suggère que la fin de la période de rebond coïncide souvent avec la fin de la transition. En effet, le nombre de sélections entre la fin de la période de transition et la fin de la période de rebond est égale à 0.6 sélections ($CI=[-0.3,+0.6]$) pour TRADITIONNEL et 0.3 sélections ($CI=+/- 0.1$) pour AUDIO.

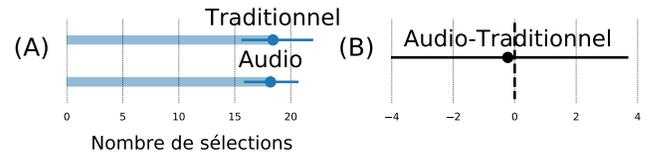


FIGURE 11: (A) Durée du rebond par technique. (B) Différence des moyennes.

6.4 Discussion

Notre analyse présente une nouvelle caractérisation de la transition des menus vers les raccourcis claviers et offre un nouveau regard sur le développement de l'expertise inter-méthode. Elle suggère en particulier que :

- (1) Il existe clairement une phase de transition pour la plupart des séquences de sélection d'une commande. Cette transition est rarement instantanée comme suggérée par les caractérisations théoriques, et ne s'étale pas tout au long de l'expérience comme suggérée par les caractérisations empiriques.
- (2) Les techniques d'interaction ont un impact sur le début de transition (AUDIO : 6 sélections ; TRADITIONNEL : 12 sélections), mais pas sur sa durée de la transition (17 sélections de moyenne).
- (3) La variable NOMBRE DE SÉLECTIONS semble mieux expliquer le début et la durée de la transition que la FRÉQUENCE. La fréquence est en un facteur complexe [34] : il modifie non seulement le nombre de sélections pour une commande cible, mais aussi le nombre de commandes intercalé entre deux commandes cibles consécutives. Nous pensions initialement, que l'utilisateur serait davantage motivé à commencer la transition si une même commande apparaissait plusieurs fois dans un court intervalle de temps. Par exemple, nous pensions que si une même commande apparaissait trois fois à la suite, l'utilisateur commencerait une transition. Ça ne semble pas être le cas. Au contraire, le nombre de sélections semble être davantage explicatif.
- (4) Le GAIN DE PERFORMANCE observé dans cette étude est de 1.3s suggérant un bénéfice à utiliser les raccourcis claviers. Cependant le lecteur peut questionner la mesure utilisé pour ce marqueur comportemental : la différence temporelle entre la sélection la plus rapide AVANT et APRÈS la transition. Rien n'indique que l'utilisateur a pu atteindre le plateau de performance avec les deux méthodes. En fait, (1) il apparaît que les utilisateurs de menus sont plus rapides s'ils n'effectuent jamais de transitions, suggérant que de nombreux utilisateurs commencent la transition avant d'avoir atteint le plateau de performance des menus. Ceci est intéressant car ce n'est pas

l'absence de progression avec une méthode qui a motivé les utilisateurs à commencer la transition; (2) Concernant les raccourcis claviers, il semble également que les utilisateurs n'ont pas eu le temps d'atteindre le plateau de performance, en particulier pour les séquences avec peu de sélections (correspondant aux commandes avec des petites fréquences). En résumé, notre mesure ne reflète donc pas exactement le marqueur comportemental comme décrit dans les caractérisations théoriques. Cependant, celles-ci ne semblent pas refléter le comportement des utilisateurs; les utilisateurs commençant fréquemment la transition avant d'avoir atteint un palier de performance. Notre mesure est donc plus adaptée à nos observations, mais devrait être renommée *gain de performance avant-après transition*.

- (5) La chute de performance a souvent été évoquée comme raison expliquant pourquoi les utilisateurs ne faisaient pas l'effort d'apprendre les raccourcis claviers, mais à notre connaissance n'avait jamais été estimée. Nos résultats suggèrent une chute d'environ 1.4 secondes. Cette analyse a aussi mis en évidence la présence d'une phase de pré-transition.
- (6) la transition est précédée d'une phase de PRÉ-TRANSITION caractérisée par le fait que le temps de sélection avec les menus augmentent. Nous pensons que l'utilisateur prépare sa transition durant cette phrase en passant plus de temps dans les menus pour lire/apprendre les raccourcis clavier. La durée de cette phase semble dépendre de la technique d'interaction : 2.9 sélections pour AUDIO contre 5.4 pour TRADITIONNEL. Ceci suggère que l'utilisateur doit lire/apprendre davantage les raccourcis claviers avec TRADITIONNEL à cause de l'absence de retours d'information avancé. Une étude oculométrique est nécessaire pour mieux comprendre les intentions des utilisateurs. Elle permettrait non seulement d'analyser l'évolution du locus d'attention au cours de l'expérience, mais aussi d'estimer l'effort cognitif. Nous envisageons également comme travail futur d'utiliser l'imagerie cérébrale (EEG) et d'utiliser des méthodes pour inciter les utilisateurs à exprimer leur comportement pour capturer par exemple les hésitations.
- (7) Enfin, nous avons estimé la DURÉE DU REBOND à 18 sélections. Il est intéressant d'observer que la fin de la transition coïncide avec la fin de la période de rebond.

7 ANNOTATION AUTOMATIQUE

Nous présentons maintenant une méthode d'annotation automatique réalisée avec les bibliothèques TensorFlow et Keras pour (1) s'affranchir de l'annotation manuelle qui est chronophage et potentiellement dépendante des annotateurs et (2) vérifier la cohérence des annotations proposées dans la section 5.

7.1 Formulation du problème et état de l'art

L'annotation automatique peut être formulée comme un problème de détection (segmentation + reconnaissance) consistant, à partir d'une série temporelle à deux variables discrètes (stratégie et précision), à déterminer la phase de chaque instant parmi les trois phases : Avant, Pendant et Après la transition.

Pour ce type de problèmes, plusieurs solutions ont été proposées dans la littérature parmi lesquelles les chaînes de Markov cachées

(HMM) [28, 42], les algorithmes de changement de points (Change Point Algorithm) [2], les méthodes de transfert d'étiquettes suite à un alignement de type Dynamic Time Warping (DTW) [21] ou les méthodes utilisant les réseaux de neurones profonds [5, 33]. Tandis que les algorithmes de changements de points sont entièrement non supervisés et reposent sur de la détection de ruptures, les méthodes de transfert d'étiquettes mettent en place un alignement entre deux séquences pour transférer les étiquettes d'une séquence sur l'autre, ce qui n'est pas très utile pour notre problème où beaucoup de séquences étiquetées sont disponibles.

Vu les gains en performance des réseaux de neurones profonds pour l'étiquetage de séquences temporelles, nous avons choisi de les utiliser et plus particulièrement, un encodeur-décodeur avec une couche convolutive 1D [5, 33].

7.2 Encodeur-décodeur de signaux temporels avec une couche convolutive 1D

7.2.1 Extraction de phases : l'encodeur-décodeur. Chaque séquence, codée à chaque instant par deux valeurs (stratégie et précision), constitue l'entrée de l'encodeur/décodeur. La sortie aura la même dimension que la séquence d'entrée (même nombre de points) mais chaque instant sera représenté par une seule valeur codant la phase : -1 pour *Avant la transition*, 0 pour *Pendant la transition* et 1 pour *Après la transition*.

Le but de l'apprentissage est d'apprendre au réseau à décoder la phase à partir de la séquence d'entrée.

La figure 12 résume le fonctionnement global de l'encodeur/décodeur par réseau de neurones convolutifs temporels.

L'encodeur a pour but de coder la partie utile de l'information dans un vecteur latent de petite dimension. Il est constitué de deux couches convolutionnelles⁵ composée de N_F filtres de taille Δ_F , suivie d'une fonction d'activation RELU (unité linéaire rectifiée) et d'une couche de max-pooling réalisant un sous-échantillonnage en conservant la valeur maximale sur une fenêtre. Ces trois couches peuvent être répétées N fois avant d'obtenir le vecteur latent.

Le décodeur a pour but de créer la sortie désirée à partir du vecteur latent. Il est composé d'une couche de sur-échantillonnage suivie d'une couche convolutionnelle et d'une couche RELU. De la même manière que pour l'encodeur, ces trois couches peuvent être répétées N fois. Une couche softmax qui transforme la sortie en probabilité d'appartenance aux phases termine le réseau.

L'apprentissage consiste à apprendre les coefficients des filtres de convolution qui permettent au mieux de prédire les séquences de phase à partir des séquences d'entrée sur une base d'apprentissage. Nous avons utilisé $N=2$, $N_F = (6, 12)$ et $\Delta_F = 5$. En effet, compte tenu de la longueur des signaux (entre 12 et 144) et de la taille du descripteur (2), ces valeurs sont un bon compromis entre un signal ni trop résumé ni trop détaillé de façon à faire ressortir clairement les transitions.

Une fois le réseau appris, il peut associer à toute séquence d'entrée $S_k = \{s_1, s_2, \dots, s_{M_k}\}$ la séquence de sortie $P_k = \{p_1, p_2, \dots, p_{M_k}\}$ où $\forall i = 1 \dots M_k, p_i \in \{-1, 0, 1\}$ correspond aux phases *Avant*, *Pendant* et *Après* la transition.

5. La convolution est un opérateur multiplicatif terme-à-terme à fenêtre glissante qui permet d'extraire un descripteur de dépendances locales tout en conservant les relations temporelles entre les instants

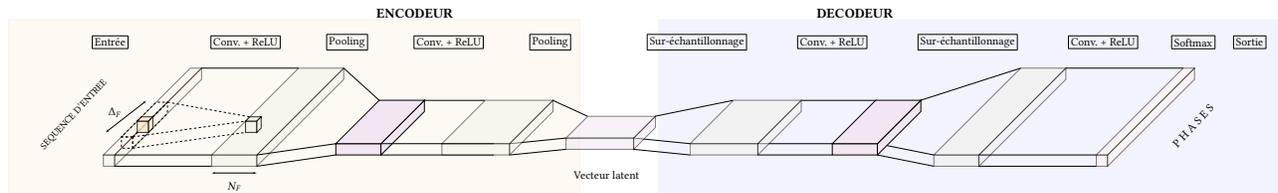


FIGURE 12: Encodeur-décodeur de séries-temporelles à partir de réseaux de neurones convolutifs.

7.2.2 Estimation du début et de la fin de la transition.

Idéalement, la séquence P_k est composée d'un seul passage entre -1 et 0 et d'un seul passage entre 0 et 1 fournissant directement le début et la fin de la transition recherchée. Comme ce n'est pas toujours le cas, une étape supplémentaire est nécessaire. Elle est basée sur des heuristiques simples :

- Le début est le premier passage entre -1 et 0 .
- La fin est le premier passage entre 0 et 1 .

Ces instants sont affinés de manière à ce que le début de la transition corresponde à une sélection qui n'est pas un menu seul et la fin de la transition corresponde à un raccourci clavier sans erreur.

7.3 Résultats et discussion

Toutes les séquences ont été considérées en test. Néanmoins, seules les séquences de taille supérieure à 36 ont été utilisées en entraînement du réseau, les séquences courtes étant trop peu informatives pour que le réseau puisse en tirer des informations.

Les résultats ont été obtenus en validation croisée "un contre tous" (LOO : Leave-one-out Cross Validation) et en "un utilisateur contre tous" (LOSO : Leave-one-subject-out Cross Validation). Ils sont présentés à la Table 2.

La Table 2 présente les taux reconnaissance obtenus pour chacune des transitions. Ils sont estimés d'une part avec les annotations après résolution des conflits (accord absolu) et d'autre part, avec celles avant résolution des conflits (accord avec au moins un annotateur). Les résultats sont meilleurs dans ce dernier cas, laissant penser que la négociation n'a pas toujours amené à la solution répandant le mieux à notre modélisation.

Nous avons comparé nos résultats à ceux d'un HMM discret à 8 symboles (4 stratégies combinées à erreur/non erreur) et 3 états (avant/pendant/après transition) suivi du même affinement que pour l'encodeur/décodeur. Les résultats obtenus avec l'encodeur/décodeur sont meilleurs que ceux des HMM : 92% de reconnaissance absolue pour le début de transition et 82% pour la fin contre 83% et 78%. D'autre part, les validations "un contre tous" et "un utilisateur contre tous" amènent à des scores similaires, laissant penser que la détection de transition est relativement indépendante de l'identité de l'utilisateur.

Les erreurs de détection sont souvent dues à des comportements singuliers que l'algorithme n'a pas modélisé comme par exemple un utilisateur qui tente de passer aux raccourcis mais se ravise. Dès lors, il existe une période de transition, mais celle-ci est suivie à nouveau de sélections par menu.

Une étude plus approfondie du taux de détection en fonction de la taille des signaux et de la technique utilisée montre que l'apprentissage est plus performant pour TRADITIONNEL que AUDIO

| Validation | Encodeur/Décodeur | HMM |
|---|-------------------|-------------|
| <i>Accord absolu</i> | | |
| LOO | 91.16/82.99 | 82.99/77.72 |
| LOSO | 92.52/82.14 | 82.99/77.89 |
| <i>Accord avec au moins un annotateur</i> | | |
| LOO | 92.69/85.20 | 85.54/83.33 |
| LOSO | 93.71/84.86 | 85.71/83.33 |

TABLE 2: Taux de détection de transition (en %) en validation croisée "un contre tous" (LOO) et "un utilisateur contre tous" (LOSO) avec une modélisation par encodeur/décodeur ou HMM. Les résultats sont donnés sous la forme *début de transition / fin de transition*.

ou INACTIF pour les fins de transitions. TRADITIONNEL et AUDIO amènent à des résultats similaires pour le début de la transition. Par ailleurs, il semble que les signaux de taille intermédiaire amènent aux meilleurs résultats (excepté pour une longueur de 12, mais ces signaux sont majoritairement des signaux sans transition).

Dans de futurs travaux, nous proposons d'utiliser les réseaux convolutionnels de manière à prédire directement les début et fin de transition et ce, afin d'éviter d'avoir à utiliser des heuristiques pour les obtenir. Par ailleurs, une optimisation du paramétrage (nombre de couches, nombre de filtres, taille des filtres) pourrait être établie sur une base de validation. Plus généralement, les résultats pourraient être améliorés en augmentant le nombre de données.

8 CONCLUSION

Dans cet article, nous discutons la caractérisation de la transition des menus vers les raccourcis avec comme cas d'application, les raccourcis claviers. Nous avons mis en évidence : (1) le fossé entre les caractérisations théoriques et empiriques et (2) le besoin d'analyser les données au niveau de la séquence de sélections par commande et utilisateur, plutôt que de les analyser après avoir agrégé les données. Ceci a permis par exemple de montrer que la transition n'est pas instantanée, que les utilisateurs n'atteignent pas le plateau de performance des menus avant de commencer la transition où que le nombre de sélections est plus explicatif que la fréquence pour expliquer le début de la transition. Et (3) le manque d'outils pour caractériser la transition. Nous avons effectué une annotation manuelle de données et proposée une annotation automatique s'appuyant sur les réseaux de neurones.

9 REMERCIEMENTS

Ce projet est financé par l'ANR, avec la référence ANR-16-CE33-0023 (GESTURE). Nous remercions S. Malacria et P. Dragicevic.

RÉFÉRENCES

- [1] Douglas G Altman. 1990. *Practical statistics for medical research*. CRC press.
- [2] Samaneh Aminikhanghahi and Diane J. Cook. 2017. A survey of methods for time series change point detection. *Knowledge and Information Systems* 51, 2 (2017), 339–367.
- [3] John R Anderson. 1982. Acquisition of cognitive skill. *Psychological review* 89, 4 (1982), 369.
- [4] Caroline Appert and Shumin Zhai. 2009. Using Strokes As Command Shortcuts : Cognitive Benefits and Toolkit Support. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 2289–2298. DOI : <http://dx.doi.org/10.1145/1518701.1519052>
- [5] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. 2017. SegNet : A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (2017), 2481–2495.
- [6] Gilles Bailly, Eric Lecolinet, and Laurence Nigay. 2008. Flower menus : a new type of marking menu with large menu breadth, within groups and efficient expert mode memorization. In *Proceedings of the working conference on Advanced visual interfaces*. ACM, 15–22.
- [7] Gilles Bailly, Thomas Pietrzak, Jonathan Deber, and Daniel J Wigdor. 2013. Métamorphe : augmenting hotkey usage with actuated keys. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 563–572.
- [8] Ram Bajpai and Himanshu Chaturvedi. 2015. Evaluation of Inter-Rater Agreement and Inter-Rater Reliability for Observational Data : An Overview of Concepts and Methods. *Journal of the Indian Academy of Applied Psychology* 41, 3 (2015), 20–27.
- [9] Olivier Bau and Wendy E. Mackay. 2008. OctoPocus : A Dynamic Guide for Learning Gesture-based Command Sets. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology (UIST '08)*. ACM, New York, NY, USA, 37–46. DOI : <http://dx.doi.org/10.1145/1449715.1449724>
- [10] Suresh K Bhavnani, Bonnie E John, and Ulrich Flemming. 1999. The Strategic Use of CAD : An Empirically Inspired, Theory-based Course. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* May (1999), 183–190. DOI : <http://dx.doi.org/10.1145/302979.303036>
- [11] Suresh K. Bhavnani, Frederick Reif, and Bonnie E. John. 2001. Beyond command knowledge : identifying and teaching strategic knowledge for using complex computer applications. (2001), 229–236. DOI : <http://dx.doi.org/10.1145/365024.365107>
- [12] Sara Bouzit, Gaëlle Calvary, Denis Chêne, and Jean Vanderdonck. 2016. A Comparison of Shortcut and Step-by-step Adaptive Menus for Smartphones. In *Proceedings of the 30th International BCS Human Computer Interaction Conference : Fusion! (HCI '16)*. BCS Learning & Development Ltd., Swindon, UK, Article 26, 12 pages. DOI : <http://dx.doi.org/10.14236/ewic/HCI2016.26>
- [13] John M Carroll and Mary Beth Rosson. 1987. Paradox of the Active User. (1987). DOI : <http://dx.doi.org/10.1017/CBO9781107415324.004> arXiv:arXiv:1011.1669v3
- [14] Andy Cockburn, Carl Gutwin, Joey Scarr, and Sylvain Malacria. 2014. Supporting Novice to Expert Transitions in User Interfaces. *ACM Comput. Surv.* 47, 2, Article 31 (Nov. 2014), 36 pages. DOI : <http://dx.doi.org/10.1145/2659796>
- [15] Marc Destefano and Wayne D Gray. 2016. Where should researchers look for strategy discoveries during the acquisition of complex task performance ? The case of space fortress. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. 668–673.
- [16] Thomas J DiCiccio and Bradley Efron. 1996. Bootstrap confidence intervals. *Statistical science* (1996), 189–212.
- [17] Pierre Dragicevic. 2016. Fair Statistical Communication in HCI. In *Modern Statistical Methods for HCI*. Springer, 291–330.
- [18] Paul Morris Fitts and Michael I Posner. 1967. Human performance. (1967).
- [19] Wai Tat Fu and Wayne D. Gray. 2004. Resolving the paradox of the active user : Stable suboptimal performance in interactive tasks. *Cognitive Science* 28, 6 (2004), 901–935. DOI : <http://dx.doi.org/10.1016/j.cogsci.2004.03.005>
- [20] Ovande Jr. Furtado and Jere D Gallagher. 2017. Collecting inter and intra-rater reliability for the Furtado-Gallagher Computerized Observational Movement Pattern Assessment System. (2017). DOI : <http://dx.doi.org/10.1177/0031512518769205>
- [21] Y. Gao, S. Vedula, G. Lee, M. Lee, S. Khudanpur, and G. Hager. 2016. Unsupervised surgical data alignment with application to automatic activity annotation. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*. 4158–4163.
- [22] Emmanouil Giannidakis, Gilles Bailly, Sylvain Malacria, and Fanny Chevalier. 2017. IconHK : Using Toolbar Button Icons to Communicate Keyboard Shortcuts. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4715–4726. DOI : <http://dx.doi.org/10.1145/3025453.3025595>
- [23] Wayne D Gray and Deborah A Boehm-Davis. 2000. Milliseconds matter : An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology : Applied* 6, 4 (2000), 322.
- [24] Wayne D. Gray and John K. Lindstedt. 2016. Plateaus, Dips, and Leaps : Where to Look for Inventions and Discoveries During Skilled Performance. *Cognitive Science* (2016), n/a–n/a. DOI : <http://dx.doi.org/10.1111/cogs.12412>
- [25] Wayne D Gray, Chris R Sims, Wai-Tat Fu, and Michael J Schoelles. 2006. The soft constraints hypothesis : a rational analysis approach to resource allocation for interactive behavior. *Psychological review* 113, 3 (2006), 461.
- [26] Tovi Grossman, Pierre Dragicevic, and Ravin Balakrishnan. 2007. Strategies for Accelerating On-line Learning of Hotkeys. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 1591–1600. DOI : <http://dx.doi.org/10.1145/1240624.1240865>
- [27] Carl Gutwin, Andy Cockburn, Joey Scarr, Sylvain Malacria, and Scott C. Olson. 2014. Faster Command Selection on Tablets with FastTap. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2617–2626. DOI : <http://dx.doi.org/10.1145/2556288.2557136>
- [28] Sanjay Krishnan, Animesh Garg, Sachin Patil, Colin Lea, Gregory Hager, Pieter Abbeel, and Ken Goldberg. 2017. Transition state clustering : Unsupervised surgical trajectory segmentation for robot learning. *The International Journal of Robotics Research* 36, 13-14 (2017), 1595–1618.
- [29] Brian Krisler and Richard Alterman. 2008. Training Towards Mastery : Overcoming the Active User Paradox. (2008), 18–22.
- [30] Brian Krisler and Richard Alterman. 2008. Training Towards Mastery : Overcoming the Active User Paradox. In *Proceedings of the 5th Nordic Conference on Human-computer Interaction : Building Bridges (NordicCHI '08)*. ACM, New York, NY, USA, 239–248. DOI : <http://dx.doi.org/10.1145/1463160.1463186>
- [31] Gordon Paul Kurtenbach. 1993. *The design and evaluation of marking menus*. Ph.D. Dissertation. University of Toronto.
- [32] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (1977), 159. DOI : <http://dx.doi.org/10.2307/2529310> arXiv:NIHMS150003
- [33] Colin Lea, René Vidal, Austin Reiter, and Gregory D. Hager. 2016. Temporal Convolutional Networks : A Unified Approach to Action Segmentation. *14th European Conference on Computer Vision* (2016), 47–54.
- [34] Wanyu Liu, Gilles Bailly, and Andrew Howes. 2017. Effects of Frequency Distribution on Linear Menu Performance. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1307–1312. DOI : <http://dx.doi.org/10.1145/3025453.3025707>
- [35] Abraham S. Luchins. 1942. Mechanization in problem solving : The effect of Einstellung. *Psychological Monographs* 54, 6 (1942), i–95. DOI : <http://dx.doi.org/10.1037/h0093502>
- [36] Sylvain Malacria, Gilles Bailly, Joel Harrison, Andy Cockburn, and Carl Gutwin. 2013. Promoting Hotkey Use Through Rehearsal with ExposeHK. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 573–582. DOI : <http://dx.doi.org/10.1145/2470654.2470735>
- [37] Sylvain Malacria, Joey Scarr, Andy Cockburn, Carl Gutwin, and Tovi Grossman. 2013. Skillometers : Reflective Widgets That Motivate and Help Users to Improve Performance. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 321–330. DOI : <http://dx.doi.org/10.1145/2501988.2501996>
- [38] Marry L. McHugh. 2012. Interrater reliability : the kappa statistic. *Biochemia Medica* (2012), 276–282. DOI : <http://dx.doi.org/10.11613/BM.2012.031> arXiv:arXiv:gr-qc/9809069v1
- [39] A. Newell and P. S. Rosenbloom. 1993. *The Soar Papers (Vol. 1)*. MIT Press, Cambridge, MA, USA, Chapter Mechanisms of Skill Acquisition and the Law of Practice, 81–135. <http://dl.acm.org/citation.cfm?id=162580.162586>
- [40] Erik Nilsen, HeeSen Jong, Judith S. Olson, Kevin Biolsi, Henry Rueter, and Sharon Mutter. 1993. The Growth of Software Skill : A Longitudinal Look at Learning & Performance. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. ACM, New York, NY, USA, 149–156. DOI : <http://dx.doi.org/10.1145/169059.169126>
- [41] Thomas Pietrzak, Sylvain Malacria, and Gilles Bailly. 2014. CtrlMouse Et TouchCtrl : Duplicating Mode Delimiters on the Mouse. In *Proceedings of the 26th Conference on L'Interaction Homme-Machine (IHM '14)*. ACM, New York, NY, USA, 38–47. DOI : <http://dx.doi.org/10.1145/2670444.2670447>
- [42] L.R. Rabiner. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* 77, 2 (Feb. 1989), 257–286.
- [43] Joey Scarr, Andy Cockburn, Carl Gutwin, and Philip Quinn. 2011. Dips and Ceilings : Understanding and Supporting Transitions to Expertise in User Interfaces. (2011), 2741–2750.
- [44] Joey Scarr, Andy Cockburn, Carl Gutwin, and Philip Quinn. 2011. Dips and Ceilings : Understanding and Supporting Transitions to Expertise in User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2741–2750. DOI : <http://dx.doi.org/10.1145/1978942.1979348>
- [45] Richard M Shiffrin and Walter Schneider. 1977. Controlled and automatic human information processing : II. Perceptual learning, automatic attending and a general theory. *Psychological review* 84, 2 (1977), 127.

- [46] Susanne Tak, Piet Westendorp, and Iris Van Rooij. 2013. Satisficing and the use of keyboard shortcuts : Being good enough is enough? *Interacting with Computers* 25, 5 (2013), 404–416. DOI : <http://dx.doi.org/10.1093/iwc/iwt016>
- [47] Caitlin Tenison and John R Anderson. 2016. Modeling the distinct phases of skill acquisition. *Journal of Experimental Psychology : Learning, Memory, and Cognition* 42, 5 (2016), 749.
- [48] Caitlin Tenison, Jon M. Fincham, and John R. Anderson. 2016. Phases of learning : How skill acquisition impacts cognitive processing. *Cognitive Psychology* 87 (2016), 1 – 28. DOI : <http://dx.doi.org/10.1016/j.cogpsych.2016.03.001>
- [49] Shengdong Zhao and Ravin Balakrishnan. 2004. Simple vs. Compound Mark Hierarchical Marking Menus. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04)*. ACM, New York, NY, USA, 33–42. DOI : <http://dx.doi.org/10.1145/1029632.1029639>