



HAL
open science

Inspection d'interfaces et critères ergonomiques

Dominique L. Scapin, Christian Joseph Maurice Bastien

► **To cite this version:**

Dominique L. Scapin, Christian Joseph Maurice Bastien. Inspection d'interfaces et critères ergonomiques. [Rapport de recherche] RR-2901, INRIA. 1996. inria-00073791

HAL Id: inria-00073791

<https://inria.hal.science/inria-00073791v1>

Submitted on 24 May 2006

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.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Inspection d'interfaces et critères ergonomiques

Dominique Louis Scapin
Joseph Maurice Christian Bastien

N° 2901

Mai 1996

THÈME 3



R
apport
de recherche



Inspection d'interfaces et Critères Ergonomiques

Dominique L. Scapin et J. M. Christian Bastien

Thème 3

Projet Psycho-Ergo

Rapport de recherche n° 2901 - Mai 1996 - 14 pages

Résumé : Les méthodes d'évaluation analytiques d'interfaces utilisateurs (par opposition aux méthodes d'évaluation empiriques, comme les "tests utilisateurs") ne reposent pas directement sur des performances ou des opinions d'utilisateurs recueillies en situation réelle ou en laboratoire, mais reposent sur l'examen, plus ou moins guidé, plus ou moins automatisé, d'interfaces spécifiées, simulées, ou opérationnelles, à partir de théories, de modèles de l'interaction homme-ordinateur, de guides, de check-lists, d'heuristiques, etc. Un sous-ensemble de ces méthodes peut être regroupé sous le vocable de "méthodes d'inspection ergonomique", il s'agit de diverses méthodes qui consistent en un examen d'interfaces par des spécialistes (évaluation experte) ou des non-spécialistes au moyen de diverses techniques d'inspection, essentiellement basées sur diverses dimensions de l'utilisabilité. L'examen de ces dimensions est au centre des études décrites ici. En particulier, on décrit la problématique et les grandes étapes de recherche menées pour établir un ensemble particulier de dimensions de l'utilisabilité: les *Critères Ergonomiques*. Quatre étapes sont décrites: la construction itérative des critères et trois expériences de validation. La première expérience traite notamment de la validité intrinsèque des critères ; la seconde expérience concerne l'efficacité et l'utilisabilité des critères ; la troisième expérience à également trait à l'efficacité et à l'utilisabilité des critères, mais au travers d'une comparaison avec un autre ensemble de dimensions, et pour une population de non-experts. La dernière section décrit l'état actuel de la méthode critères et les perspectives envisagées.

Mots-clés : ergonomie ; interfaces homme-ordinateur ; logiciels interactifs ; méthodes d'évaluation ; dimensions de l'utilisabilité ; critères ergonomiques.

Interface inspection and Ergonomic Criteria

Abstract: Analytic methods for the evaluation of user interfaces (as opposed to empirical evaluation methods, such as "user testing" do not rely directly on users' performances or opinions gathered in context or in a laboratory situation, but rely on the examination, more or less guided, more or less automatized, of interfaces which can be either specified, simulated or operational, using theories, human-computer interaction models, guides, check lists, heuristics, etc. A sub-set of these methods can be labelled "ergonomic inspection methods" : these are the various methods which consist of an examination of interfaces by specialists (expert evaluation) or by non-specialists by means of various inspection methods based essentially on various usability dimensions. The study of these dimensions is the at the core of the experiments briefly described here. More specifically, this paper describes the orientation and the main research steps established to define a particular set of usability dimensions: the *Ergonomic Criteria*. Four steps are described: the iterative design of the criteria and three experiments aimed at validating them. The first experiment concerns the intrinsic validity of the criteria; the second experiment deals with the efficiency and usability of the criteria; the third experiment concerns the efficiency and usability of the criteria as well, but through a comparison with another set of dimensions, and for a population of non-experts. The last part of the paper describes the present state of the criteria-based method and the perspectives for the future.

Keywords: ergonomics; human-computer interfaces; interactive software; evaluation methods; usability dimensions; ergonomic criteria.

1 Introduction

L'incorporation du savoir ergonomique dans le processus de conception des systèmes informatiques est au centre des préoccupations de recherche du projet de Psychologie Ergonomique de l'INRIA (pour un résumé des travaux, voir Scapin, 1993). Les thèmes étudiés ont été la modélisation formelle des tâches (Scapin & Pierret, 1990 ; Sebillotte, 1991 ; Hammouche, 1993 ; Sebillotte & Scapin, 1994), l'étude de l'expertise ergonomique de diagnostic (Pollier, 1992), la modélisation ergonomique d'interfaces (Gibbons, 1992 a & b ; Cybis, 1993), ainsi que le développement des Critères Ergonomiques ("C.E."). C'est sur ce dernier thème que porte ce texte.

La définition et la validation des "C.E." pour la conception et l'évaluation d'interfaces utilisateurs poursuivait plusieurs objectifs :

- définir le concept même d'utilisabilité en identifiant ses composantes,
- structurer les résultats du domaine de l'ergonomie des logiciels en proposant des dimensions autorisant l'organisation de bases de données ou de bases de règles (cf. leur utilisation dans la compilation de recommandations de Vanderdonck (1995)),
- contribuer aux standards (les "C.E." ont été repris en majeure partie dans la norme AFNOR Z 67-133-1),
- enfin, contribuer à l'amélioration du diagnostic ergonomique, notamment dans son exhaustivité, sa fiabilité et sa robustesse.

C'est ce dernier objectif qui nous intéresse particulièrement, notamment dans une perspective de future utilisation, par des non-spécialistes, de méthodes d'évaluation ergonomique. En effet, même si nombre de travaux fondamentaux et appliqués demeureront toujours de la responsabilité des ergonomes, il nous semble que fournir des méthodes utilisables par le plus grand nombre (i.e., les divers intervenants en matière de conception/évaluation d'interfaces : acheteurs, vendeurs, concepteurs, informaticiens, graphistes, utilisateurs, etc.) contribuera à améliorer globalement la qualité ergonomique des interfaces du futur.

Dans ce texte, on évoquera la problématique des recherches, puis les divers travaux de construction et de validation des "C.E." ; enfin, à partir de l'état actuel de l'approche "C.E.", la discussion fournira des propositions d'études pour le futur.

2 Problématique

Quand une personne (experte ou non) a pour objectif d'évaluer une interface, elle peut faire appel, en fonction de son degré d'expertise ergonomique, de la complexité des problèmes, du coût des méthodes, de la disponibilité des utilisateurs, etc., à diverses méthodes et techniques utilisées dans le domaine de l'ergonomie des logiciels (observations, interviews, tests utilisateurs, méthodes d'inspection, etc.). Elle dispose notamment de nombreuses sources d'information du type guides généraux de conception (Mayhew, 1992) ; ensemble de recommandations (Scapin, 1986 ; Smith & Mosier, 1986 ; Vanderdonckt, 1995) ; check-lists (Ravden & Johnson, 1989) ; normes (e.g. AFNOR, ISO).

Bien que tous ces documents aient été rédigés dans le but de concevoir des interfaces ergonomiques, peu d'entre eux ont été évalués afin de s'assurer de leur qualités intrinsèques et de leur utilisabilité lors de la conception et lors de l'évaluation. Smith et ses collaborateurs (Smith et Aucella, 1983 ; Mosier & Smith, 1986) ont montré que leur compilation de recommandations de conception était difficile à utiliser : difficultés pour repérer et choisir les recommandations, pour établir des priorités parmi les recommandations sélectionnées et traduire les recommandations en règle de conception spécifiques. Il a été établi que seulement 41% des répondants au questionnaire, utilisaient la compilation pour évaluer une conception proposée ; 25% pour évaluer une conception complète, et 18% pour évaluer un système opérationnel. La dernière version du guide de recommandations de Smith et Mosier (1986) n'a pas été testée avec des utilisateurs. Une autre étude (de Souza & Bevan, 1990) sur un projet de norme comportant des recommandations de conception sur les dialogues avec menus, a montré que des concepteurs qui utilisaient cette norme faisaient des erreurs, i.e., prenaient des décisions de conception qui enfreignaient certaines des recommandations. Les erreurs sont apparues comme résultant du manque d'information au sujet des objectifs de la conception et des avantages à utiliser des recommandations, des conditions selon lesquelles elles devraient être appliquées, de la nature précise de la solution proposée, et des procédures à suivre pour les appliquer.

L'expérience des concepteurs a montré avoir plus d'influence sur la conception d'une interface qu'un document normatif consistant en seulement deux pages de recommandations relatives à l'utilisation de touches fonction, et au partage de l'écran en champs fixes (Thovtrup & Nielsen, 1991). Dans une autre étude (Tetzlaff & Schwartz, 1991), des participants ont eu des difficultés dans l'interprétation de recommandations et d'exemples, et se sont basés principalement sur les illustrations graphiques au point de ne pas prendre en compte le texte associé.

Pour alléger certaines de ces difficultés rencontrées dans la consultation des compilations de recommandations, des outils d'aide pour les concepteurs travaillant avec des recommandations ont été développés (voir Vanderdonckt, 1993, pour une revue de ces outils). Cependant, peu d'outils ont été soumis à des tests d'utilisabilité (e.g., Ogawa & Yonemura, 1992).

Une des contributions du domaine de l'ergonomie des logiciels a été de fournir des dimensions, des principes organisateurs, notamment pour l'évaluation d'interfaces. Intéressons-nous plus spécifiquement à ces notions de dimensions de l'utilisabilité (heuristiques, principes, critères, etc.), qui sont supposées guider, résumer, faciliter le diagnostic ergonomique. Il apparaît tout d'abord que les ensembles de dimensions actuellement disponibles varient d'un auteur à l'autre en nombre et niveau de précision. Il apparaît également que pour définir des dimensions ergonomiques de manière satisfaisante, i.e., pour pouvoir les utiliser et pour qu'elles contribuent à des résultats identifiables, il est nécessaire de s'assurer de leurs caractéristiques intrinsèques et de leur utilisabilité. En effet, on devrait s'interroger sur :

- le lien entre de telles dimensions et les recommandations prises individuellement,
- leur indépendance et leur compréhensibilité,
- leur fiabilité (l'utilisation de dimensions par divers utilisateurs conduit-elle aux mêmes résultats ?) et leur applicabilité selon le contexte.

C'est seulement en répondant à ce type de questions sur la validité interne, l'homogénéité, l'indépendance, le caractère explicite, et la validité externe que l'utilisabilité et l'impact de ces dimensions pourront être établis. De cette façon, les dimensions correspondant à ces exigences pourront être utilisées à grande échelle tant pour l'organisation des connaissances ergonomiques que pour le transfert de ces connaissances vers les concepteurs, et leur utilisation pour la formation en ergonomie, pour la conception de grilles d'évaluation, pour la structuration des rapports d'évaluation, pour la conception de métriques, pour l'indexation de bases de données et pour la conception d'outils d'évaluation informatisés.

Trois types de dimensions ont été étudiés, à des degrés divers : les "heuristiques" de Molich et Nielsen (1990) ; les "principes de dialogues" (ISO 9241, Part 10, 1994) ; et les "critères ergonomiques".

Les "heuristiques" ont été utilisées, pour évaluer différents systèmes, dans plusieurs expériences, par des spécialistes de l'utilisabilité aussi bien que par des non-spécialistes. Les évaluateurs individuels n'ont trouvé que de 20% à 51% des problèmes d'utilisabilité

que les interfaces contenaient (Nielsen & Molich, 1990). Par ailleurs, les regroupements des rapports d'évaluation de plusieurs évaluateurs ne donnent des résultats satisfaisants qu'à condition de réunir les rapports d'évaluation de trois à cinq personnes. Il est apparu aussi que, dans une évaluation heuristique, les problèmes majeurs d'utilisabilité avaient une plus grande probabilité d'être détectés, mais aussi qu'au total, environ deux fois plus de problèmes mineurs sont détectés ; par ailleurs, certains problèmes (e.g., le manque de sorties clairement indiquées) semble être plus difficiles à découvrir que d'autres (Nielsen, 1992).

Les "Principes de dialogue" (ISO 9241-Part 10) ont été testés en termes d'acceptabilité auprès d'experts en utilisabilité de plusieurs pays par le biais d'une enquête à base de questionnaires. Il a été conclu que ces principes fournissaient "un cadre pour la conception et l'évaluation de systèmes de dialogue" (Beimel et al., 1994). Il reste que cet ensemble de principes devrait être testé en termes d'utilisabilité, d'une part avec des mesures de performance, et d'autre part, auprès de la population réelle de futurs utilisateurs, qui ne seront pas forcément des experts internationaux. En particulier, si les questions d'indépendance et de caractère distinct peuvent être éludées pour un guidage général en matière d'ergonomie des interfaces, il n'en va pas de même si de telles dimensions sont envisagées comme procédures normalisées d'évaluation ou même simplement comme méthode d'indexation de recommandations (du type de celles qui sont contenues dans les autres parties de la norme ISO 9241, notamment les parties 12 à 17).

Les Critères Ergonomiques ont quant à eux suivi un processus de construction empirique et ont donné lieu à plusieurs expériences qui ont tenté de répondre à ces questions. La section suivante résume les données disponibles sur ces questions et sur les travaux centrés sur l'évaluation des "C.E."

3 Construction et validation des Critères Ergonomiques

3.1 Construction

Le moteur initial des travaux sur les "C.E." a été de formaliser et de structurer les connaissances ergonomiques, notamment telles qu'elles apparaissent sous la forme de recommandations. Un modèle générique de traduction des recommandations, exprimé pour la conception et pour l'évaluation, a été proposé (Scapin, 1990 a & b). La notion essentielle de ce modèle est celle de "C.E." : ceux-ci représentent d'abord un moyen de classification des règles (e.g., dans une base de données), mais surtout, ils représentent

les dimensions ergonomiques majeures selon lesquelles une interface peut-être évaluée ou spécifiée. Un ensemble important de recommandations a été collecté à partir de guides ergonomiques et d'un examen d'Ergonomics Abstracts. Les recommandations ont été sélectionnées puis caractérisées à partir de divers descripteurs. Cette approche a permis de répertorier environ 800 recommandations (Aschehoug & Scapin, 1991) sur lesquelles les divers descripteurs ont été utilisés pour indexation. A partir de ces travaux, une première version des "C.E." a été produite (Scapin, 1990 b & c).

La construction des "C.E." a suivi une démarche itérative de classification et d'accord inter-juges. Il a cependant semblé également nécessaire de valider la qualité et l'utilisabilité de ces critères de façon expérimentale, ce qui représente une des rares validations de telles dimensions dans la littérature.

3.2 Expérience 1

Dans une première étude (Bastien & Scapin, 1992), une expérimentation a été menée avec l'objectif de vérifier si les critères ainsi définis correspondaient bien aux problèmes ergonomiques dont ils sont supposés rendre compte. En particulier, il s'agissait de vérifier si les définitions, justifications et exemples de recommandations accompagnant les critères permettaient un bon appariement entre "C.E." et problèmes ergonomiques.

Pour ce faire, vingt-quatre participants (12 spécialistes et 12 non-spécialistes de l'ergonomie) ont pris part à une tâche d'identification de concept ("C.E.") à partir d'exemples. Cette tâche consistait plus précisément à identifier, pour un problème de conception d'une interface donnée, le critère ergonomique correspondant. La plupart des problèmes de conception étaient tirés d'une application réelle permettant la gestion d'une base de données bibliographiques. Les résultats ont montré que les deux groupes de participants ne se distinguaient ni par le nombre d'identifications correctes ni par le temps consacré à la lecture des définitions ou à la tâche d'identification. Les identifications correctes variaient en moyenne de 59,85 à 74,3% selon qu'il s'agissait des critères élémentaires ou principaux (voir plus loin le Tableau 1). De plus, l'examen détaillé des données a permis de déterminer des classes de critères bien définis et des classes de critères dont les définitions bénéficieraient d'améliorations. Chacune de ces classes comportait la moitié des critères élémentaires. Suite à ces résultats, un certain nombre de modifications ont été apportées aux définitions ; de plus, des exemples et des commentaires ont été ajoutés afin de mieux expliciter les différences entre critères. Cette étude a permis d'envisager une méthode d'évaluation basée sur des "C.E." définis de façon explicite et non ambiguë ; au total, 18 Critères élémentaires (voir Tableau 1)

ont été définis (Bastien et Scapin, 1993a). Ce document a été utilisé dans les expérimentations suivantes.

Tableau 1. Liste des Critères Ergonomiques
(les critères élémentaires sont affectés d'un astérisque)

| |
|---|
| 1. Guidage |
| 1.1 Incitation* |
| 1.2 Groupement/Distinction entre items |
| 1.2.1 Groupement/Distinction par la localisation* |
| 1.2.2 Groupement/Distinction par le format* |
| 1.3 Feed-back immédiat* |
| 1.4 Lisibilité* |
| 2. Charge de travail |
| 2.1 Brièveté |
| 2.1.1 Concision* |
| 2.1.2 Actions minimales* |
| 2.2 Densité informationnelle* |
| 3. Contrôle explicite |
| 3.1 Actions explicites* |
| 3.2 Contrôle utilisateur* |
| 4. Adaptabilité |
| 4.1 Flexibilité* |
| 4.2 Prise en compte de l'expérience de l'utilisateur* |
| 5. Gestion des erreurs |
| 5.1 Protection contre les erreurs* |
| 5.2 Qualité des messages* |
| 5.3 Correction des erreurs* |
| 6. Homogénéité / Cohérence* |
| 7. Signifiante des codes et dénominations* |
| 8. Compatibilité* |

3.3 Expérience 2

Une deuxième expérience (Bastien & Scapin, 1993b, 1995) a été conduite afin de tester l'efficacité et l'utilisabilité des "C.E."

Deux groupes de spécialistes en ergonomie des logiciels (Contrôle ($n = 10$), Critères ($n = 9$)) ont été invités, au cours de deux phases, à évaluer l'interface (programmée intentionnellement avec des erreurs ergonomiques de conception) d'un système de gestion d'une base de données musicales "Hypermanip" (Scapin, 1992) développée sur Macintosh avec HyperCardTM. Dans la première phase de l'étude, tous les participants ont évalué l'interface en ne se basant que sur leur propre expertise. Au cours de la deuxième phase de l'étude, les participants ont évalué l'interface une deuxième fois mais cette fois-ci en visionnant l'enregistrement de leurs interactions précédentes. Les participants du groupe Critères, contrairement à ceux du groupe Contrôle ont utilisé pour ce faire le jeu de "C.E.". Au cours de la première phase, aucune différence n'a été observée entre les groupes quant au nombre de problèmes d'utilisabilité détectés et

quant à la proportion de problèmes détectés et à la proportion de problèmes détectés en commun selon le nombre d'évaluations réunies. Au cours de la deuxième phase toutefois, les performances des participants du groupe Critères se sont avérées supérieures à celles des participants du groupe Contrôle. Les premiers ($M = 46.1$; $E.T. = 11.47$) ont détecté plus de nouveaux problèmes que les seconds ($M = 32.3$; $E.T. = 6.91$) et la proportion des problèmes détectés de même que la proportion des problèmes détectés en commun s'est avérée supérieure en fonction du nombre d'évaluations mises en commun. Par exemple, pour détecter environ 50% des problèmes ergonomiques, il faut 5 participants dans le groupe Critères et 6/7 dans le groupe Contrôle. En résumé, les critères ont amélioré les performances d'évaluation des experts.

3.4 Expérience 3

Dans une troisième expérience (Bastien, Scapin, & Leulier, 1996), on s'est intéressé à l'utilisation des "C.E." par des non-spécialistes en ergonomie. De plus, les "C.E." ont été comparés à d'autres dimensions, en l'occurrence les Principes de Dialogue de la norme ISO 9241- Part 10, du point de vue de la performance de diagnostic.

Dix-sept étudiants en ergonomie ont été répartis de façon aléatoire dans trois groupes ("Contrôle" ($n = 6$), "Critères" ($n = 6$) et "ISO" ($n = 5$)). Ils ont été invités à évaluer une partie de l'interface du prototype d'application de gestion d'une base de données musicales utilisé lors de l'étude précédente. Environ une semaine avant la session expérimentale, les participants des groupes "Critères" et "ISO" ont respectivement reçu les "C.E." (accompagnés de leur définition, justifications, exemples et commentaires) et le document ISO. Ils ont été invités à prendre connaissance des documents avant de se présenter à la session expérimentale. Cette dernière s'est déroulée en trois phases : au cours de la première phase (10 minutes), les participants assistaient à une démonstration des fonctionnalités de l'application qu'ils allaient évaluer ; cette première phase terminée, les participants ont pris part à une phase d'exploration du système d'une durée de 10 minutes afin de se familiariser avec l'application ; au cours de la troisième phase de l'étude, il a été demandé aux participants du groupe "Critères" de relire le document et plus particulièrement les parties "définitions", "justifications" et "exemples". De la même manière, chacun des participants du groupe "ISO" a été invité à relire le document sur les principes de dialogue ISO. Les participants des deux groupes ont ensuite été invités à évaluer l'interface de façon exhaustive. La consigne précisait qu'ils devaient procéder pas à pas : pour chaque page écran, boîte de dialogue ou fenêtre à l'écran, il leur était demandé de reprendre un à un chacun des Critères ou principes de

dialogue décrits dans le document et d'indiquer les problèmes de conception s'y rapportant. Les participants appartenant au groupe "Contrôle", n'ayant pas de documents, devaient s'en remettre à leur seul jugement lors de la phase d'évaluation.

L'analyse des résultats indique que globalement l'effet groupe est statistiquement significatif. Des tests à posteriori montrent que le nombre de problèmes détectés en moyenne par les participants du groupe Critères ($M = 86,2$; $E.T. = 12,7$) diffère de façon significative à la fois du groupe Contrôle ($p = .0101$) et du groupe ISO ($p = .0144$). Par contre, le nombre de problèmes détectés en moyenne par les participants des groupes Contrôle ($M = 61.8$; $E.T. = 15.8$) et ISO ($M = 62.2$; $E.T. = 13.8$) ne diffère pas de façon significative.

En résumé, du point de vue des diagnostics, le groupe disposant du document ISO ne se distingue pas du groupe ayant effectué son évaluation sans aucun document. En revanche, les participants du groupe Critères détectent plus de problèmes, et ce de façon significative, que les participants des groupes Contrôle et ISO. De plus, la proportion des problèmes détectés, de même que la proportion des problèmes détectés en commun, est apparue supérieure en fonction du nombre d'évaluations mises en commun.

4 Etat actuel et Perspectives

Les "C.E." ont été bâtis à partir d'une base de recommandations importante, ce qui leur assure une certaine exhaustivité ; ils ont été testés du point de vue de leur indépendance, de la stabilité de leur affectation, de leur utilisabilité et de leur rôle comme aide au diagnostic.

Les critères dans leur forme actuelle apparaissent donc valides, fiables et utilisables. Plus important, leur utilisation augmente la performance d'évaluation. L'ensemble des "C.E." a été utilisé comme une grille pour le diagnostic et pour des rapports d'évaluations d'interfaces, notamment dans l'industrie.

Cependant, on pense que l'efficacité de l'aide apportée par les "C.E." peut être améliorée en définissant plus avant une méthode associée. En effet, dans l'état actuel des choses, l'ébauche de méthode (telle qu'elle a été utilisée par exemple dans les expérimentations précédentes) se résume à l'examen d'une interface en utilisant systématiquement chacun des critères et sous-critères, selon l'ordre dans lequel ils sont décrits, à partir d'un document présentant la version actuelle des "C.E." (Bastien et Scapin, 1993a).

Ce document (à partir duquel l'évaluation peut être mise en oeuvre) fournit, pour les 18 "C.E." élémentaires (cf. Tableau 1) :

- une définition qui décrit chaque critère,
- des justifications qui avancent les raisons (notamment psychologiques) sur lesquelles chaque critère repose,
- des exemples de recommandations correspondant au critère considéré,
- des commentaires et des contre-exemples permettent de bien différencier les critères les uns des autres

Cependant il ne s'agit pas encore d'une méthode d'évaluation complètement développée : par exemple, le document ne précise ni les façons d'explorer systématiquement l'interface, ni ne donne de méthode d'application détaillée de chaque critère. Il reste donc beaucoup à faire pour définir une méthode complète et des outils précis d'évaluation.

Afin d'atteindre cet objectif, i.e., de développer une méthode d'évaluation plus détaillée qui pourrait augmenter davantage les performances d'évaluation, et de concevoir des outils (hypertexte, système expert, etc.) supportant une telle méthode, le travail futur devrait s'effectuer dans les directions suivantes :

- enrichir le contenu de chaque critère i.e., inclure un ensemble plus complet de "règles" spécifiques pour chacun des critères ; identifier les manières de mettre en oeuvre chacun de ces critères,
- définir un ensemble de priorités pour l'évaluation (e.g., des ordres d'examen des critères.). De telles priorités devraient être définies en fonction des objectifs d'évaluation (e.g. le *Guidage* devrait avoir une priorité sur la *Flexibilité* pour des utilisateurs inexpérimentés ; si une performance élevée est requise, alors la *Charge de Travail* devra avoir une priorité élevée, etc.),
- définir les façons d'examiner systématiquement tous les états et éléments d'interface (écran, fenêtres, séquences d'inputs, etc.). Ceci requiert la détermination des contextes d'interaction et la définition de modèles adéquats de l'interface (composants et comportement, i.e., également le dialogue) sur lesquels appliquer les critères,
- définir les pré-requis à l'évaluation, i.e., toutes les caractéristiques liées à la tâche et à l'utilisateur qui sont nécessaires pour appliquer certains critères

(notamment pour les critères *Compatibilité* et *Adaptabilité*); des liens devront être établis avec des méthodes formelles d'analyse de tâche.

Evidemment, la fiabilité de la méthode avec certains de ces développements devra être évaluée avec de vrais utilisateurs (concepteurs et utilisateurs finaux). L'utilisation de la méthode et son influence sur les logiciels produits, en termes de coût/efficacité, aura également besoin d'être évaluée. L'approche critères devra enfin être comparée à d'autres méthodes d'évaluation pour estimer son efficacité relative, et identifier sa contribution au processus global d'évaluation d'interface.

L'objectif des travaux futurs sera donc de définir et de valider une méthode d'Évaluation des Interfaces à partir de Critères Ergonomiques ("ÉVINCE"), selon deux étapes : une étape de construction et une étape de validation.

- L'objectif de l'étape de construction sera essentiellement d'opérationnaliser plus encore les Critères en augmentant leur domaine de référence (les recommandations associées), en proposant des heuristiques et des outils associés à chacun d'entre eux et en définissant une structure méthodologique d'examen des interfaces ainsi que des moyens de rendre compte des évaluations (grille d'analyse).
- L'objectif de l'étape de validation sera de tester l'efficacité de la méthode de manière expérimentale. Par ailleurs, l'extension des critères au multimédia est envisagée.

Rappelons les caractéristiques souhaitables d'une telle méthode :

- être basée sur l'examen d'une interface (par opposition au test utilisateurs),
- permettre dans un temps relativement court, le diagnostic du plus grand nombre de problèmes ergonomiques,
- être systématique et explicite,
- être utilisable par des non-spécialistes en ergonomie.

Corollairement, cela signifie plusieurs choses :

- cette méthode entre dans la catégorie des méthodes dites "méthodes d'inspection" qui sont définies comme des méthodes informelles d'analyse de l'utilisabilité consistant en l'examen d'une interface spécifiée, prototypée ou existante avec l'objectif d'identifier les défauts ergonomiques de conception. Ces dernières reposent sur l'expertise des évaluateurs (spécialistes des facteurs humains, concepteurs de système, informaticiens)

et/ou sur une certaine connaissance des facteurs humains telle qu'elle est disponible dans les documents habituels cités plus haut (guides, check-lists, etc.). Dans notre cas, il s'agira d'inspection ergonomique d'interfaces par des évaluateurs non nécessairement experts utilisant les "C.E." enrichis comme guide d'évaluation.

- une telle méthode concernera essentiellement les aspects de l'interface (et plus généralement de la situation) pour lesquels une réponse ergonomique existe, c'est à dire les problèmes ergonomiques pour lesquels des travaux de référence sont disponibles. Pour des problèmes plus complexes (e.g., liés à une expertise) ou nouveaux (e.g., interfaces virtuelles), il faudra faire appel à d'autres techniques (modélisation formelle fine, évaluations empiriques avec des utilisateurs, etc).

- une telle méthode ne se substituera pas, mais sera un complément à d'autres méthodes d'évaluation (e.g., les méthodes basées sur des modèles, les interviews, les tests utilisateurs, etc.). Les critères devraient être vus plutôt comme le moyen de garantir la conformité avec les recommandations de conception d'interface, comme un moyen d'évaluer, relativement rapidement les erreurs ergonomiques, pour découvrir les défauts de conception repérables avant les tests utilisateurs.

- une telle méthode peut être vue également comme un moyen aisé de faire entrer dans les habitudes des concepteurs/évaluateurs la nécessité de préoccupations ergonomiques, puis de faire appel à des ergonomes professionnels pour avancer dans la conception. En effet, il ne s'agit pas de considérer que la totalité de l'approche ergonomique sera du ressort du non-spécialiste en ergonomie, mais plutôt que certains de ses aspects les plus flagrants pourront être traités relativement rapidement, sous la forme d'une inspection ergonomique systématique. Les ergonomes pourraient ainsi se consacrer à des problèmes plus difficiles requérant les connaissances et les méthodes qu'ils ont à leur disposition. L'approche critères devrait aussi aider certains d'entre eux à optimiser leur démarche évaluative.

D'un point de vue plus général, notre perspective (cf. Scapin, 1994) est de considérer que l'ergonomie est un domaine scientifique qui se doit de proposer et d'appliquer des méthodes génériques et non pas se contenter de jugements individuels (très variables selon l'expertise ; cf. Pollier, 1992). Ainsi on peut espérer une certaine consolidation des résultats du domaine et une meilleure inclusion de l'ergonomie dans le processus de conception de logiciels.

5 Références

- AFNOR Z 67-133-1 (1991). "Traitement de l'information - Evaluation des produits logiciels ; Partie 1 : Définition des Critères Ergonomiques de conception et d'évaluation des interfaces utilisateurs".
- Aschehoug, F. & Scapin, D. L. (1991). *Base de recommandations ergonomiques pour l'évaluation et la conception d'interfaces utilisateurs*. Document interne, non diffusable.
- Bastien, J. M. C., & Scapin, D. L. (1992). A validation of ergonomic criteria for the evaluation of human-computer interfaces. *International Journal of Human-Computer Interaction*, 4(2), 183-196.
- Bastien, J. M. C., & Scapin, D. L. (1993a). *Ergonomic criteria for the evaluation of human-computer interfaces* (Report No. 156). INRIA Rocquencourt.
- Bastien, J. M. C., & Scapin, D. L. (1993b). *Preliminary findings on the effectiveness of ergonomic criteria for the evaluation of human-computer interfaces*. Short paper presented at the Conference on Human Factors in Computing Systems, INTERCHI '93, Amsterdam, The Netherlands.
- Bastien, J. M. C., & Scapin, D. L. (1995). Evaluating a user interface with ergonomic criteria. *International Journal of Human-Computer Interaction*, 7, 105-121.
- Bastien, J. M. C., Scapin, D. L., & Leulier, C. (1996). Looking for usability problems with the Ergonomic Criteria and with the ISO 9241-10 dialogue principles. In M. J. Tauber (Eds.), *Conference on Human Factors in Computing Systems. CHI'96 Conference Companion* (Vancouver, Canada, April 13-18)(pp. 77-78). New York, NY: ACM.
- Beimel, J., Schindler, R., & Wandke, H. (1994). Do human factors experts accept the ISO 9241 Part 10 - Dialogue principle - standard? *Behaviour & Information Technology*, 13, 299-308.
- Cybis, W. (1993). Vers un modèle d'objets d'interactions abstraits pour le raisonnement ergonomique sur l'interface homme-machine. *XXIX Congrès de la SELF*. — 21-23 Septembre 1994.
- de Souza, F., & Bevan, N. (1990). The use of recommendations in menu interface design: Evaluation of a draft standard. In D. Diaper, D. Gilmore, G. Cockton, & B. Shackel (Eds.), *Proceedings of the IFIP TC 13 Third International Conference on Human-Computer Interaction: INTERACT'90* (pp. 435-440). Cambridge, U.K.: Elsevier Science Publishers.
- Gibbons, S. C. (1992a). *Towards an expert system based menu interface evaluation tool* (Rapport de recherche N° 1581). INRIA Rocquencourt.
- Gibbons, S. C. (1992b). Organizing the domain of menus. *Proceedings of The Human Factors Society*, October 12-16, Atlanta, GA, USA
- Hammouche, H. (1993). *De la modélisation des tâches à la spécification d'interfaces utilisateur*. Rapport de recherche N°1959, INRIA Rocquencourt.
- International Standards Organisation (1994). *ISO 9241 Part 10. Ergonomic requirements for office work with visual display terminals - Part 10 Dialogue Principles ; Draft International Standard* (Report No.
- Mayhew, D. J. (1992). *Principles and recommendations in software user interface design*. Englewood Cliffs: Prentice Hall.
- Molich, R., & Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of the ACM*, pp. 338-348.
- Mosier, J. N., & Smith, S. L. (1986). Application of recommendations for designing user interface software. *Behaviour and Information Technology*, 5, 39-46.
- Nielsen, J. (1992). Finding usability problems through heuristic evaluation. In P. Bauersfeld, J. Bennett, & G. Lynch (Eds.), *Proceedings of ACM CHI'92 Conference on Human Factors in Computing Systems* (pp. 373-380). Monterey, California: ACM.

- Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In J. Carrasco & J. Whiteside (Eds.), *Empowering people: CHI'90 Conference Proceedings* (pp. 249-256). Seattle, Washington: ACM.
- Ogawa, K., & Yonemura, S.-I. (1992). Usability analysis of design guideline database in human-computer interface design. In *Proceedings of the Human Factors Society 36th Annual Meeting* (pp. 433-437).
- Pollier, A. (1992). Évaluation d'une interface par des ergonomes : Diagnostics et stratégies *Le Travail Humain*, 55(1), 71-96.
- Ravden, S. J., & Johnson, G. I. (1989). *Evaluating usability of human-computer interfaces: A practical method*. Chichester, England: John Wiley & Sons.
- Scapin, D. L. (1986). *Guide ergonomique de conception des interfaces homme-machine [Ergonomic guidelines for the design of human-computer interfaces]* (Tech. Rep. No. 77). Rocquencourt, France: Institut National de Recherche en Informatique et en Automatique.
- Scapin, D. L. (1990a). Aiding mechanisms for the design of user interfaces: some research issues, *First International Conference on Automation Technology*, July 4-6, 1990, Taipei, R.O.C.
- Scapin, D. L. (1990b). Organizing human factors knowledge for the evaluation and design of interfaces. *International Journal of Human-Computer Interaction*, 2(3), 203-229.
- Scapin, D. L. (1990c). Des critères ergonomiques pour l'évaluation et la conception d'interfaces utilisateurs. In Actes du XXVI Congrès de la SELF. Montréal, Canada.
- Scapin, D. L. (1992). "HyperManip", prototype d'application expérimentale de gestion d'une base de données musicales. (HyperCardTM), INRIA Rocquencourt.
- Scapin, D. L. (1993). Situation et perspectives en ergonomie du logiciel. In J.-C. Sperandio (Eds.), *L'Ergonomie dans la conception des projets informatiques* (pp. 7-68). Toulouse, France : Octares.
- Scapin, D. L. (1994). *For a psycho-engineering approach to HCI* (Rapport de recherche N° 2215). Rocquencourt, France : Institut National de Recherche en Informatique et Automatique.
- Scapin, D. L. & Pierret, C. (1990) Towards a method for task description. In *Working with Display Units*, D. Berthelette (Ed.), Elsevier.
- Sebillotte, S., (1991). Décrire des tâches selon les objectifs des opérateurs : de l'interview à la formalisation. *Le travail Humain*, 54, (3).
- Sebillotte, S. & Scapin, D.L. (1994). From users' task knowledge to high level interface specification. *International Journal of Human-computer Interaction*, 6, 1-15.
- Smith, S. L., & Aucella, A. F. (1983). *Design recommendations for the user interface to computer-based information systems* (Report No. ESD-TR-83-122). Hanscom Air Force Base, MA: U.S.A.F. Electronic Systems Division.
- Smith, S. L., & Mosier, J. N. (1986). *Recommendations for designing user interface software* (Report No. ESD-TR-86-278). Mitre Corporation.
- Tetzlaff, L., & Schwartz, D. R. (1991). The use of recommendations in interface design. In S. P. Robertson, G. M. Olson, & J. S. Olson (Eds.), *Proceedings of ACM CHI'91 Conference on Human Factors in Computing Systems* (pp. 329-333). New Orleans, Louisiana: ACM.
- Thovtrup, H., & Nielsen, J. (1991). Assessing the usability of a user interface standard. In S. P. Robertson, G. M. Olson, & J. S. Olson (Eds.), *Proceedings of ACM CHI'91 Conference on Human Factors in Computing Systems* (pp. 335-341). New Orleans, Louisiana: ACM.
- Vanderdonckt, J. (1993). Hypermedia on human-computer interaction principles and recommendations: A survey. *ACM SIGLINK Newsletter*.
- Vanderdonckt, J. (1995). *Guide ergonomique des interfaces homme-machine*. Namur, Belgique : Presses Universitaires de Namur.

Table des matières

| | |
|---|-----------|
| 1 Introduction | 1 |
| 2 Problématique | 2 |
| 3 Construction et validation des Critères Ergonomiques | 4 |
| 3.1 Construction..... | 4 |
| 3.2 Expérience 1 | 5 |
| 3.3 Expérience 2 | 6 |
| 3.4 Expérience 3 | 7 |
| 4 Etat actuel et Perspectives | 8 |
| 5 Références | 11 |



Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Lorraine - Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)

Unité de recherche INRIA Rennes - IRISA, Campus universitaire de Beaulieu 35042 Rennes Cedex (France)

Unité de recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 Grenoble Cedex 1 (France)

Unité de recherche INRIA Sophia Antipolis - 2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur

INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

ISSN 0249 - 6399



★ R R - 2 9 8 1 ★