

SYNTHESE

Évaluation des Interfaces Homme-Machine lors du développement des systèmes interactifs

Martial Grislin, Christophe Kolski

LAMIH - CNRS 1775
Le Mont Houy - BP 311
59304 Valenciennes cedex
e-mail {grislin, kolski}@univ-valenciennes.fr

RESUME. Dans le cycle de développement d'applications interactives, l'évaluation des interfaces homme-machine est souvent négligée, voire ignorée par les spécialistes du génie logiciel. Pourtant, de nombreuses méthodes et techniques sont actuellement susceptibles d'améliorer la qualité des interfaces homme-machine. Cet article constitue une revue de synthèse de celles-ci.

ABSTRACT. In the software development cycle of interactive applications, man-machine interface evaluation is often neglected or unknown by the software engineering specialists. Nevertheless many evaluation methods and techniques can contribute to the improvement of the man-machine interface quality. This article constitutes a review of such methods and techniques.

MOTS-CLES : évaluation, interface homme-machine, cycle de développement, génie logiciel.

KEY WORDS : evaluation, human-computer interface, software development cycle, software engineering.

1. Introduction

Le génie logiciel est actuellement en pleine effervescence et on assiste à l'avènement de nombreuses méthodes, modèles, techniques qui visent à couvrir l'ensemble des étapes [JAU 90][JAC 93][BOO 94][DEN 94][CUR 94]. Mais, dans la majeure partie des cas, la conception des IHM et l'évaluation de leur qualité n'apparaissent pas explicitement et systématiquement dans ces étapes [BAR 88][BAL 94][KOL 95]. Pourtant, les applications informatiques interactives sont de plus en plus répandues. Du point de vue des facteurs humains, il est évident que leur qualité est cruciale pour leur efficacité, leur utilisabilité¹ et leur acceptabilité auprès des utilisateurs et ceci particulièrement pour les systèmes homme-machine complexes où les contraintes d'utilisation peuvent être sévères en termes de sécurité, économie ou production.

Cet article constitue une revue de synthèse des techniques utilisées aujourd'hui et utilisables demain en génie logiciel pour l'évaluation d'IHM. La première partie discute de la notion d'évaluation des IHM. La seconde consiste en une discussion autour des classifications actuelles, visant à mieux introduire la troisième partie où sont recensées les méthodes et techniques d'évaluation. Pour chacune d'entre elles, nous en donnons le principe et les critères d'évaluation concernés, ainsi qu'un positionnement dans le cycle de développement de systèmes interactifs.

2. L'évaluation des IHM

L'IHM doit être considérée comme composant du système homme-machine global. Dans ces conditions, son évaluation consiste à le *vérifier* et à chercher à le *valider*. Le système homme-machine est vérifié s'il correspond aux spécifications issues de la définition des besoins. Il est validé s'il correspond aux besoins en respectant les contraintes du domaine d'application ; sinon, ses insuffisances par rapport à des critères identifiés a priori doivent être mises en évidence.

2.1. Principe de l'évaluation

L'évaluation ergonomique d'une IHM consiste à s'assurer que l'utilisateur est capable de réaliser sa tâche au moyen du système de communication qui lui est proposé. "*Toute évaluation consiste à comparer un modèle de l'objet évalué à un modèle de référence permettant d'établir des conclusions.*" [SEN 90]. Lors de l'évaluation de système

¹ Les anglo-saxons utilisent les termes "usability" ou "user-friendliness" pour qualifier le caractère ergonomique d'un dispositif, le terme "ergonomics" étant réservé à la discipline ergonomique. Le français ne fait pas cette distinction, mais il est courant d'utiliser les barbarismes "utilisabilité" ou "convivialité".

interactif, le modèle que l'on peut qualifier d'observé (ou d'analysé) est donc comparé à un modèle de référence. Ce modèle doit être représentatif de l'adéquation de l'IHM évaluée par rapport aux besoins spécifiques définis par le concepteur (fig. 1).

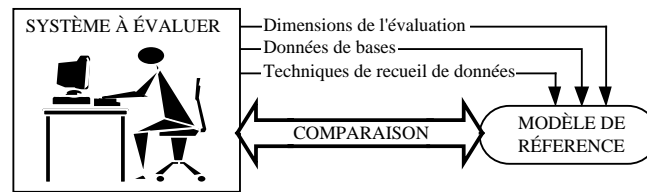


Figure 1. Principe global de l'évaluation

La construction du modèle de l'IHM est dirigée par le contexte ; c'est-à-dire que la sélection des propriétés pertinentes pour le diagnostic de la qualité ergonomique dépend non seulement des objectifs de l'évaluation mais aussi des caractéristiques de la population et des exigences des tâches. Ces propriétés sont *l'utilité* et *l'utilisabilité* :

- *L'utilité* détermine si l'IHM permet à l'utilisateur d'atteindre ses objectifs de travail. Elle correspond aux capacités fonctionnelles, aux performances et à la qualité de l'assistance technique fournie à l'utilisateur par le système.

- *L'utilisabilité* rend compte de la qualité de l'interaction homme-machine, en termes de facilité d'apprentissage et d'utilisation, ainsi que de qualité de la documentation. C'est en quelque sorte une frontière entre l'utilité potentielle d'un produit et son utilité réelle. En effet, fréquemment, l'ajout à l'IHM de fonctionnalités censées faciliter l'exploitation a pour effet pervers de compliquer la tâche de l'utilisateur.

Pour l'évaluation, il est important de bien délimiter le niveau de précision du modèle afin de pouvoir analyser et faire ressortir les aspects considérés comme pertinents vis-à-vis de l'évaluation. Il s'agit donc de déterminer ce qui permettra de porter un jugement sur l'IHM concernée. C'est à ce niveau qu'interviendront les critères d'évaluation sur lesquels se focalise la partie suivante.

2.2. Critères ergonomiques pour l'évaluation des IHM

L'analyse de l'IHM doit mettre à jour un ensemble de variables cibles, c'est-à-dire les données de base à recueillir. La figure 2 présente quelques unes des principales variables dépendantes utilisées pour l'évaluation des IHM. Deux grandes familles se distinguent : l'acceptabilité sociale (des systèmes inacceptables socialement seraient par exemple des systèmes posant des questions indiscretes aux utilisateurs, ou visant à identifier dans un but de licenciement des opérateurs insuffisamment productifs) et l'acceptabilité pratique. La seconde classe comprend les contraintes courantes de production, de coût, de fiabilité... mais aussi la facilité d'utilisation qui concerne principalement l'IHM. C'est dans ce critère d'utilisation que l'on retrouve les deux propriétés citées précédemment. Notons que l'article se focalise par la suite sur l'utilisabilité.

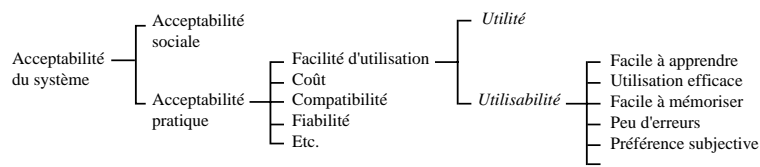


Figure 2. Principales variables cibles pour l'évaluation (traduit de [NIE 93])

La littérature propose un très grand nombre de recommandations ergonomiques sous forme de guides ergonomiques [SMI 86][SCA 87][BRO 88], des standards (BSI 91/40677, ISO 9241), mais aussi de guides de style accompagnant généralement des environnements graphiques (Toolbox du Macintosh, boîtes à outils de Windows et Presentation Manager du monde IBM-Microsoft, Motif et Open Look du monde Unix, etc). Pour l'évaluation, il s'agit soit de les utiliser comme dimension à observer ou mesurer, soit d'en extraire des critères d'évaluation.

En se basant sur un grand nombre de travaux de recherche, une classification a été proposée par Bastien et Scapin [BAS 93] pour aboutir à un jeu de critères et sous-critères qui prennent en compte ces recommandations. Dix-huit critères élémentaires sont ainsi énumérés et explicités dans le tableau 1 (nous recommandons vivement au lecteur de lire le document d'origine ou [VAN 94] afin de prendre connaissance de recommandations possibles associées à chaque critère). L'ensemble de ces critères va aider l'évaluateur à estimer la qualité ergonomique de l'IHM au sens de **l'utilisabilité** et à prendre si nécessaire des décisions de modification et/ou d'amélioration. Notons qu'il existe des relations de causalité entre critères (une baisse de la performance peut être la conséquence d'une surcharge de travail) ; c'est pourquoi ils ne doivent pas être étudiés séparément.

<i>Critères</i>	<i>Éléments de l'IHM concernés</i>
Guidage : ensemble des moyens mis en œuvre pour conseiller, orienter, informer et conduire l'utilisateur lors de ses interactions avec l'ordinateur.	
Incitation (1)	Informations fournies à l'utilisateur sur l'état dans lequel il se trouve, actions possibles ou attendues et moyens de les mettre en œuvre.
Groupement par localisation (2)	Positionnement des items les uns par rapport aux autres pour indiquer leur appartenance ou non à une même classe, ou en montrer la distinction.
Groupement par format (3)	Indices graphiques des items les uns par rapport aux autres pour indiquer leur appartenance ou non à une même classe.
Retour informatif (4)	Réponses de la machine après actions de l'utilisateur ("Feed-back" immédiat).
Clarté (5)	Caractéristiques lexicales de présentation des informations pouvant entraver ou faciliter la lecture de ces informations.
Charge de travail : ensemble des éléments ayant un rôle dans la réduction de la charge perceptive ou mnésique des utilisateurs et dans l'augmentation de l'efficacité du dialogue.	
Concision (6)	Charge de travail perceptive et mnésique en rapport à des éléments individuels d'entrée/sortie.
Actions minimales (7)	Charge de travail au niveau des options ou moyens utilisés pour atteindre un but.
Densité d'information (8)	Charge de travail perceptive et mnésique pour des ensembles d'éléments et non pour des items.
Contrôle explicite : prise en compte par le système à la fois des actions explicites des utilisateurs et du contrôle qu'ils ont sur le traitement de leurs actions.	
Actions explicites (9)	Relations explicites entre le fonctionnement de l'application et les actions des utilisateurs. Le système doit exécuter seulement les opérations demandées par l'utilisateur, au moment où il les demande.
Contrôle utilisateur (10)	L'utilisateur doit toujours avoir la main et donc pouvoir contrôler le déroulement des traitements en cours.
Adaptabilité : capacité du système à réagir selon le contexte, les besoins et les préférences de l'utilisateur.	
Flexibilité (11)	Moyens mis à la disposition de l'utilisateur pour personnaliser l'IHM selon ses stratégies ou habitudes de travail et les exigences des tâches.
Expérience utilisateur (12)	Moyens mis en œuvre pour permettre au système de respecter le niveau d'expérience de l'utilisateur.
Gestion des erreurs : moyens permettant d'éviter ou réduire les erreurs et les corriger.	
Protection (13)	Moyens mis en place pour détecter les erreurs d'entrées de données ou de commandes ou d'actions aux conséquences néfastes.
Qualité des messages (14)	Pertinence, facilité de lecture et exactitude de l'information fournie sur la nature des erreurs commises et les actions correctives à entreprendre.
Correction des erreurs (15)	Moyens mis à la disposition de l'utilisateur pour lui permettre de corriger ses erreurs.
Homogénéité et cohérence (ou consistance) (16) : façon dont les choix de conception de l'IHM sont conservés pour des contextes identiques, et différents sinon.	
Signifiante des codes et dénominations (17) : Adéquation entre l'objet ou l'information affichée ou demandée et son référent.	
Compatibilité (18) : Accord existant entre les caractéristiques des utilisateurs (mémoire, perception, habitudes, etc.) et les tâches d'une part, et l'organisation des sorties, des entrées et du dialogue d'une application donnée, d'autre part.	

Tableau 1. *Jeu de critères et sous-critères pour l'évaluation de l'utilisabilité de l'IHM [résumé de BAS 93]*

Pour ce qui est de l'**utilité**, il est possible de l'évaluer par une analyse de la tâche et/ou des activités en se basant sur deux critères principaux : *l'adéquation à la tâche et la répartition du travail*. L'adéquation à la tâche consiste à vérifier si les procédures cognitives développées par l'utilisateur sont similaires à celles initialement développées par le concepteur et donc à estimer si la tâche (réelle) redéfinie par l'utilisateur est en accord avec la tâche (prescrite) à effectuer. Concernant le critère de répartition du travail, il s'agit d'analyser la façon dont est organisé le travail entre l'homme et la machine [MIL 88][DEB 93].

En rapport avec l'utilité, l'aspect **performance** du système homme-machine doit être considéré avec attention lors de l'évaluation, pour certains systèmes interactifs. L'IHM est conçue pour assurer la communication entre l'homme et sa machine certes, mais le couple homme-machine doit souvent produire des résultats ; c'est le cas par exemple dans les systèmes industriels complexes. C'est pourquoi un (ou plusieurs) critère de performance du système global est indispensable pour déterminer si les objectifs sont atteints. La difficulté réside dans le choix d'une donnée représentative, pertinente et surtout mesurable.

De la même manière, les **préférences** de l'utilisateur sont à prendre en compte également dans toute évaluation ("c'est lui qui va travailler avec le machine, il a donc son mot à dire"). C'est un critère subjectif, qui est étudié par exemple à l'aide de questionnaires, et qui permet de connaître le jugement et les suggestions de l'utilisateur sur la

qualité de l'IHM. Nielsen [NIE 93] émet toutefois une réserve concernant l'avis des utilisateurs : ceux-ci commencent souvent par rejeter une modification conseillée par le concepteur si elle s'avère complètement différente de ses habitudes, pour —si elle est bonne— ne plus s'en passer ensuite.

Enfin, l'évaluation de l'IHM fait partie de celle du système global. Aussi l'analyse de l'ensemble de la **documentation**, notamment les aides utilisateur doit être prise en compte de façon explicite dans les critères, même si celles-ci ne seront quasiment jamais consultées pour une IHM qui respecte toutes les recommandations ergonomiques et qui s'avère adaptée à l'utilisateur (facile d'utilisation, rapide à apprendre, intuitive...), ou alors pour laquelle l'utilisateur est parfaitement formé.

Comme le montrera la revue de la section 4, chaque technique et méthode d'évaluation permet de collecter des données. Ainsi, les mesures effectuées sur le système homme-machine vont regrouper un ensemble de variables observées par rapport aux propriétés citées précédemment. Elles forment en quelque sorte une base de données regroupant aussi bien des informations objectives (par exemple correspondant à des taux d'erreurs de manipulation de l'IHM enregistrées automatiquement) que subjectives (par exemple issues de questionnaires ou d'interviews). Pour les données objectives, différents traitements spécifiques pourront être appliqués par les évaluateurs, certaines étant basées sur l'utilisation de méthodes d'**analyse de données** issues des statistiques descriptives [VOL 81]. Pour des évaluations fines, il est clair qu'une bonne pratique est nécessaire aux évaluateurs pour interpréter les résultats, qui pourront être aussi bien qualitatifs que quantitatifs. A ce sujet, il s'agit de préciser qui est intervenant de l'évaluation.

2.3. Les intervenants de l'évaluation

L'évaluation d'une IHM nécessite une compétence pluridisciplinaire que se partage un certain nombre d'intervenants : spécialiste des facteurs humains (ergonomes, psychologues du travail...) connaissant la tâche, concepteurs (souvent des informaticiens), mais on préférera l'intervention d'une équipe pluridisciplinaire d'évaluateurs. En fait, on considère maintenant de plus en plus que l'évaluation n'est véritablement efficace que si la démarche adoptée est participative [LAN 91][KOL 93][CAC 93]. Par exemple, dans les applications complexes, la notion de groupe de travail est primordiale. Ces groupes qui composent l'équipe de développement, s'organisent donc autour de trois types principaux de connaissances : celles du spécialiste des facteurs humains, de l'utilisateur et du concepteur, auxquels il s'agit dans certaines applications complexes de rajouter celles des experts de l'application.

- **Les connaissances du spécialiste des facteurs humains** : il doit être familier des exigences physiques et cognitives, ainsi que des limites et des ressources théoriques d'un utilisateur réalisant une tâche donnée dans un environnement socio-technique, ... mais doit aussi être largement sensibilisé au fonctionnement du système technique.

- **Les connaissances de l'utilisateur** : elles se basent sur la pratique et l'expérience de différentes situations. L'utilisateur connaît l'efficacité de certaines interventions ; dans le meilleur des cas, il a déjà utilisé différents outils informatiques. Il est donc théoriquement en mesure de recenser une grande partie des tâches qui lui ont été assignées et d'en décrire les contraintes.

- **Les connaissances du concepteur** : il apporte ses connaissances liées aux méthodes et outils de développement. Il doit en outre faciliter la liaison entre les données recueillies et leur modélisation.

- **Les connaissances de l'expert de l'application** : dans certains domaines d'application (par exemple les systèmes industriels complexes), des experts connaissent l'application, son comportement physique, l'interaction entre ses variables principales constitutives, etc. Il est aussi capable de recenser les informations utiles pour le maintien du fonctionnement normal du système et de contribuer à l'analyse des modes opératoires qu'effectuent les utilisateurs.

Toute évaluation consiste à identifier ou à prévoir les difficultés que rencontrent les utilisateurs, à détecter les points forts et les points faibles du système, à en comprendre les raisons pour ensuite trouver des solutions correctives. Comme nous venons de le voir, avec toujours un même objectif, les techniques employées sont très dépendantes du contexte. Aussi, avant toute analyse, il est nécessaire de bien choisir la méthode à employer, liant à la fois le critère à évaluer à la situation concrète [GRI 93]. C'est un des points sur lesquels nous insistons dans la partie suivante, en nous basant sur les classifications actuelles.

3. Discussion préalable sur les classifications actuelles en matière d'évaluation d'IHM

Plusieurs classifications des méthodes et techniques d'évaluation d'IHM existent dans la littérature. On trouve souvent la distinction entre des approches dites prédictives et des approches dites expérimentales, c'est le cas par exemple de la classification de Nielsen et Molich [NIE 90]. Proche de ces auteurs, Senach [SEN 90], qui a constitué pour nous une source importante d'inspiration, fait une distinction entre approches analytiques et empiriques. Sweeney et Maguire [SWE 93] proposent trois grandes familles de méthodes où l'évaluation est basée sur l'utilisateur, la théorie ou alors sur une expertise. Whitefield et ses collègues [WHI 91] organisent leur classification en fonction de la présence ou non de l'utilisateur d'une part, et du système interactif à évaluer d'autre part. Balbo et Coutaz [BAL 94][COU 94] classifient les techniques en fonction de plusieurs paramètres, tel leur degré d'automatisation, la prise en compte de l'utilisateur, ou encore le savoir utilisé pour l'évaluation.

Une évaluation rigoureuse des IHM nécessite de suivre des méthodes et des techniques d'évaluation. Ce recours pour les évaluateurs (ne faut-il pas plutôt parler dans de nombreux cas de recours pour *l'équipe de développement* ?) à une aide dépend des caractéristiques de la situation à évaluer et, en particulier de l'état de développement du système interactif à évaluer : réalisé concrètement, ou hypothétique mais potentiellement envisageable.

A ce sujet, les classifications existantes sont très importantes, et semblent souvent en théorie très pertinentes, puisqu'elles permettent d'appréhender de manière générale les conditions principales d'évaluation. Toutefois, en tirant les leçons de nombreuses évaluations réalisées ces dernières années sur des systèmes homme-machine complexes réels et/ou simulés, nous nous sommes aperçus que chaque cas d'évaluation est un cas particulier, où l'évaluation relève d'un état d'esprit qui doit se retrouver tout au long des étapes du cycle de vie du système interactif concerné, et où les conditions d'évaluation peuvent varier d'une semaine à l'autre, sous les contraintes industrielles. Par exemple des utilisateurs représentatifs peuvent d'un jour à l'autre ne plus être disponibles ; une situation de référence peut être imposée à d'autres fins en fait qu'une évaluation de l'IHM ; les décideurs peuvent imposer que l'évaluation ne dépasse pas un nombre très limité de journées de travail (peut-on dans ce cas encore parler d'évaluation ? ou plutôt d'évaluation à moindre coût [NIE 93]) ; ou alors ils peuvent exiger que les méthodes utilisées soient les plus "simples" possibles de manière à être réutilisée ensuite par des ingénieurs en interne indépendamment de spécialistes de la communication homme-machine (au risque d'en faire une utilisation non rigoureuse) ; les premières évaluations peuvent mettre en évidence des problèmes de fond concernant l'utilité du système à évaluer, etc.

C'est pourquoi, dans cet article, notre objectif n'est pas de proposer une nouvelle classification, mais de décrire des ensembles de méthodes qui *pourront et devront être combinées* pour arriver à une évaluation pertinente. Dans un souci de clarté, il nous a été nécessaire d'effectuer des regroupements de méthodes et techniques. Nous en avons défini trois que l'on retrouve d'ailleurs explicitement ou implicitement dans la plupart des classifications existantes : centrées sur l'utilisateur, centrées sur une expertise, centrées sur une modélisation de l'IHM et/ou de l'interaction homme-machine. Nous justifions ces trois regroupements ci-dessous.

Lorsque *l'IHM à évaluer existe*, des méthodes d'analyse de données comportementales recueillies au moyen d'observations ou de mesures, sur le terrain ou en laboratoire sont préconisées. Ces méthodes consistent à réaliser des tests d'utilisabilité avec des utilisateurs habituels ou des sujets sélectionnés pour l'expérience. Ce type d'évaluation s'effectue le plus souvent sur des IHM déjà utilisées ou maquettées et aboutit à l'émission de jugements relatifs à leur utilisation. Dans ce premier cas, on distingue un certain nombre de méthodes et techniques que l'on peut regrouper de la manière suivante :

- (1) des approches centrées sur les avis et/ou les activités d'utilisateurs représentatifs ;
- (2) des approches qualifiées d'expertes centrées sur le jugement *d'experts en communication homme-machine* ou sur l'utilisation de *grilles d'évaluation* ou de questionnaires listant les qualités d'une bonne IHM.

Lorsque *l'IHM à évaluer est inexistante*, des méthodes d'évaluation analytiques sont utilisables. Elles permettent de fournir une appréciation de l'IHM avant son utilisation réelle. L'évaluation repose alors sur la comparaison des informations concernant l'interaction, à un stade particulier de son développement, avec un modèle de référence. Ce modèle décrit une IHM qui répond aux besoins nécessaires à l'utilisateur dans la réalisation de sa tâche en terme d'utilité et d'utilisabilité. Dans ce second cas, des méthodes issues des approches suivantes sont utilisables :

- (2) des approches qualifiées d'expertes (déjà citées précédemment) qui comparent l'IHM à un modèle informel de la communication homme-machine, en visant à prédire certains aspects liés aux interactions homme-machine ;
- (3) des approches qualifiées d'analytiques, centrées sur une modélisation de l'IHM et/ou de l'interaction homme-machine. Celles-ci consistent le plus souvent à effectuer l'évaluation à l'aide de métriques objectives à partir d'un modèle descriptif des tâches humaines, ou à partir d'une description des pages-écrans.

Toutefois, il nous paraît irréaliste de considérer ces regroupements de manière indépendante. En effet, des relations transversales entre méthodes sont souvent très utiles. Par exemple, lors de l'analyse de situations de référence avec des utilisateurs (1), il est souvent utile de procéder à une modélisation des tâches (3) dans un but de spécification ; mais il est aussi possible de réutiliser cette modélisation lorsque le système interactif est réalisé, afin de la confronter à une modélisation des activités réelles des utilisateurs (1) du système réalisé [ABE 90]. Insistons donc simplement sur le fait que *les mêmes méthodes peuvent contribuer différemment à l'évaluation au début et en fin de projet*, point très important que l'on ne retrouve que très peu traité dans les classifications existantes (cf. à ce sujet [BAL 94]).

4. Recensement des méthodes et techniques d'évaluation d'IHM

Les méthodes et techniques utilisables pour l'évaluation des IHM proviennent de domaines variés, tels que l'ergonomie ou la psychologie, le génie logiciel et l'ingénierie de la connaissance. Nous en recensons une quarantaine regroupées selon les trois approches d'évaluation citées précédemment (fig. 3). Leurs principes de base sont résumés dans des tableaux, en indiquant les critères ergonomiques qui nous paraissent potentiellement concernés, ainsi que quelques références bibliographiques.

-
- 4.1. *Les approches centrées sur les utilisateurs*
 - 4.1.1. Les approches empiriques
 - 4.1.2. L'estimation de la charge de travail
 - 4.1.3. Les tests de conception avec les utilisateurs
 - 4.2. *Les approches centrées sur une expertise humaine ou "papier"*
 - 4.3. *Les approches centrées sur une modélisation de l'IHM et/ou de l'interaction H-M*
 - 4.3.1. Modèles formels dits prédictifs
 - 4.3.2. Modèles formels dits de qualité de l'IHM, approches logicielles
 - 4.3.2.1. Systèmes d'évaluation automatique de l'affichage
 - 4.3.2.2. Systèmes de génération automatique d'affichage
-

Figure 3. *Recensement des méthodes et techniques selon les trois approches*

Bien entendu, l'objectif n'est pas de détailler chacune des méthodes et techniques, un article même volumineux ne suffirait pas (!). Cette revue de synthèse vise avant tout à fournir des points de repère au lecteur qui se référera aux auteurs cités afin d'approfondir l'étude des méthodes qui l'intéressent. Le choix de l'une ou l'autre de ces méthodes dépend surtout des contraintes d'application de la technique proprement dite (temps, disponibilité de personnel, accès, coût...). A ce sujet, la mise en place de critères de choix des techniques est un travail difficile et constitue un thème de recherche à part entière, dont nous ne discuterons pas dans cet article [GRI 93][BAL 94].

4.1. Les approches centrées sur les utilisateurs

Ces approches sont fondées sur le recueil et l'analyse des données comportementales provenant de l'utilisation des IHM par des utilisateurs représentatifs de la population finale. L'évaluation souvent expérimentale doit être effectuée dans un environnement le plus proche possible de la réalité.

4.1.1. Les approches empiriques

Les techniques et méthodes associées à cette première classe (Cf. tableau 2 ; exemples en figure 4) peuvent être utilisées pour l'évaluation de dispositifs existants lorsque des utilisateurs expérimentés sont disponibles.

Certaines de ces méthodes exigent une interaction directe des évaluateurs avec les utilisateurs, c'est le cas du "recueil de l'expertise", de "l'analyse des protocoles", des "incidents critiques" et de la construction "d'arbre des causes". Par contre, d'autres méthodes consistent à recueillir des données représentatives de l'interaction à l'aide d'observations, puis à analyser ces traces de l'activité des utilisateurs, c'est le cas du "mouchard électronique", des traces écrites du travail et de l'analyse des tactiques visuelles. Pour chacune des approches empiriques, il s'agit après analyse des données recueillies, d'en tirer les conclusions avec les utilisateurs, dans un but de spécification du système interactif, d'amélioration de l'existant, ou pour préciser des besoins en aide ou en formation.

<i>Technique ou méthode</i>	<i>Principe / Objectif</i>	<i>Critères concernés</i>
Recueil de l'expertise [DIÉ 90] [BENY 93]	- Rendu de l'activité et des problèmes de l'utilisateur expérimenté en langue naturelle (entretiens, verbalisations spontanées ou provoquées) ou à l'aide de graphiques. - Mise en évidence à l'aide de techniques issues de l'ingénierie de la connaissance des exécutions particulières (ce qui a été fait), de procédures (comment procéder), des connaissances sous-jacentes aux procédures (propriétés du système), de l'organisation et des stratégies de raisonnement...	2, 3, 5-7, 9-12, 16, 18
Analyse des protocoles [ERI 84] [KOU 91]	- Recueil pendant une session de travail des activités de l'utilisateur en interaction avec le système : stratégies, objectifs, problèmes types, besoins en informations... - L'utilisateur doit penser à haute voix (enregistrements simultanés audio et vidéo). Biais possibles dus à ces conditions inhabituelles de travail pour l'utilisateur. - Méthode appartenant au recueil de l'expertise (Cf. ci-dessus).	2, 3, 5-7, 9-12, 16, 18
Vidéo [YOU 87] [MAG 89] [THE 89]	- Enregistrement objectif et non intrusif (au contraire de l'analyse des protocoles) de l'activité de l'utilisateur pour estimer ses stratégies, les problèmes rencontrés, les informations prélevées... - Nécessité d'explications complémentaires de la part des utilisateurs lors de sessions rejouées en différé.	1-5, 9, 10, 14, 17, 18
Traces écrites du travail [SPE 91]	- Recueil et analyse des traces écrites du travail de l'utilisateur (notes écrites, documents de conduite, cahiers de doléances...). - Prises de notes par l'utilisateur centrées sur des points précis ou des problèmes particuliers au fur et à mesure de son travail.	critères "en vrac"
Questionnaires [SIN 90]	- Questionnaires d'utilisation, portant sur la qualité de l'IHM, remplis à tête reposée par les utilisateurs. - Analyse des données objectives recueillies par des méthodes statistiques classiques, mais appréciation subjective des besoins pour certaines classes d'utilisateurs. - Biais dus au fait que les questionnaires sont remplis en dehors des situations de travail normales et anormales.	tous les critères sont potentiellement concernés
Arbre des causes [VIL 88] [KIR 90]	- Description graphique d'un enchaînement causal sous la forme d'une arborescence (avec connecteurs ET et OU) en partant de l'événement indésirable et en recherchant ses causes possibles. - Mise en évidence d'erreurs humaines dues à des problèmes d'utilisabilité du système interactif. - Nécessité préalable de recueil de données sur l'utilisateur et ses activités avec des méthodes comme celles citées précédemment.	2, 3, 7, 13-18
Incidents critiques [SEN 85] [KOE 94]	- Recueil systématique des dysfonctionnements à partir d'entretiens avec les utilisateurs et d'observations en situation. - Démarche suivie : description des incidents, mise en évidence de classes de situations, proposition d'améliorations.	tous les critères
Mouchard électronique (monitoring) [NEA 83] [BIA 93]	- Recueil automatique et non intrusif, en situation de travail, des actions des utilisateurs et de leurs répercussions. - Pratique de laboratoire possible sur site lorsque le système est ouvert (insertion d'un mouchard qui recueille les données). - Prise en compte et étude de la dynamique des situations. - Apparition de systèmes de capture automatique de données (actions clavier, clavier-souris...) associés progressivement à des logiciels d'analyse [NEA 83][THE 89][HAM 92].	1, 4, 6, 9-11, 13-15
Tactiques visuelles [ABE 90] [BER 92]	- Observation de l'activité oculaire au moyen d'un oculomètre. - A partir de l'étude de l'activité oculomotrice : étude de la manière dont l'utilisateur recherche et localise les informations utiles en fonction des différentes situations de fonctionnement, mise en évidence des stratégies utilisées et localisation des lacunes dans l'IHM lors de situations de crise. - Instrumentation lourde mais suffisamment maîtrisée pour être utilisée de plus en plus lors de l'analyse de situations de contrôle de systèmes complexes (multi-écrans).	2, 5-8, 14, 17

Tableau 2. *Approches empiriques*

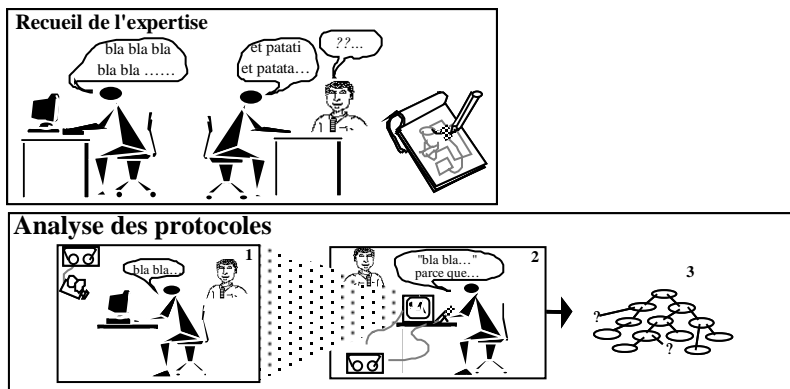


Figure 4. Exemples de méthodes issues des approches empiriques

Positionnement dans le cycle de développement : nous avons souligné que ces approches nécessitent une expérience d'utilisation, c'est-à-dire un site de référence existant, ce qui est difficile pour la conception d'un système complètement nouveau. Toutefois, le développement d'un produit démarre rarement de rien et vise généralement l'amélioration d'une situation existante, déjà informatisée ou non. Dans ce cas, toutes ces techniques d'évaluation peuvent être envisagées **a priori** dès les étapes d'analyse et de spécification et introduire un état d'esprit d'évaluation, en visant à tirer les leçons des points forts et points sensibles du ou des sites de référence.

Bien entendu, on les utilise aussi — le plus souvent d'ailleurs — **a posteriori**, c'est-à-dire après les phases d'implémentation, de tests et d'exploitation, lorsque l'IHM existe concrètement, et ceci dans le but de l'évaluer et de l'améliorer.

4.1.2. L'estimation de la charge de travail

Une mesure qualitative ou quantitative du niveau d'activité d'un utilisateur effectuant une tâche est pertinente pour évaluer la difficulté à utiliser l'IHM. L'évaluation de cette charge de travail est basée sur l'observation du travail cognitif de l'utilisateur.

De nombreuses méthodes sont proposées par des spécialistes provenant d'horizons divers : psychologie, physiologie, ergonomie, sciences pour l'ingénieur. La littérature à leur sujet abonde, que ce soit concernant leur développement ou leur mise en œuvre [WIE 93]. Les principales "familles" de méthodes généralement citées sont listées dans le tableau 3 (des exemples sont visibles en figure 5).

Technique ou méthode	Principe / Objectif	Critères concernés
Échelles subjectives [WIE 83] [AMA 86] [HAR 88]	- Questionnaires d'opinion remplis par l'utilisateur après la réalisation de sa tâche afin d'estimer son niveau d'activité mentale et physique. - Plusieurs questionnaires existent, par exemple : COOPER HARPER scale, SWAT (Subjective Workload Assessment Technique), TLX (Task Load index).	6-8
Mesure de la performance [BEN 87] [SPE 88]	- Estimation de la charge mentale en mesurant une donnée représentative du niveau de difficulté de l'utilisateur lorsqu'il réalise sa tâche (temps de réponse, erreurs par rapport à la consigne). - Fondée sur l'hypothèse que toute augmentation de la charge de travail engendre une dégradation de sa performance sur la tâche principale (voir par exemple : la méthode de la double tâche).	6-8
Mesures physiologiques [WIL 88] [SPE 91]	- Évaluation de la charge en quantifiant les changements au niveau de certaines fonctions physiologiques (activité électrique du cerveau, des muscles, clignement de l'oeil, diamètre de la pupille, fréquence cardiaque ou respiratoire...) - Donne un indice sur la capacité à faire face à une tâche.	6-8
Méthode temporelle [MIL 88] [RIE 93]	- Estimation de la charge grâce à un modèle normatif simulé, basé sur différentes hypothèses (canal de transmission d'information limité, lois de la théorie de l'information...) - Calcul en temps réel de l'indice de charge, dépendant du temps disponible et du temps requis pour réaliser la tâche.	6-8

Tableau 3. Méthodes d'estimation de la charge de travail (d'après [SIM 93])

L'analyse des méthodes amène à dire que l'estimation de la charge de travail doit reposer sur l'utilisation de plusieurs indices. Santucci [SAN 87] propose le trépied de la figure 5 qui associe une description du travail réalisée par l'utilisateur à un estimateur subjectif et un indice physiologique. L'analyse de la tâche effectuée renseigne sur les

différentes stratégies de l'utilisateur, un changement de stratégie pouvant signifier une variation de charge. L'indice subjectif rend compte du "vécu conscient" de l'utilisateur alors que l'indice physiologique par un "câblage" de l'utilisateur informe sur sa capacité et son niveau de vigilance.

Positionnement dans le cycle de développement : Comme pour les approches empiriques, les techniques d'estimation de la charge de travail nécessitent une expérience d'utilisation. C'est pourquoi elles sont utilisables **a priori** lors de l'analyse d'une situation de référence.

Lorsque l'IHM est réalisée, ou lorsqu'il est possible d'utiliser un prototype adapté à l'évaluation de la charge de travail [RIE 93], l'utilisation de ces méthodes **a posteriori** devient également possible, que ce soit en site réel ou en simulation.

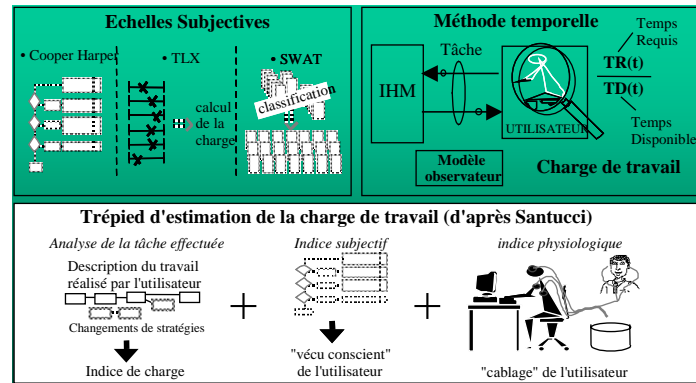


Figure 5. Exemples de méthodes d'estimation de la charge de travail

4.1.3. Les tests de conception avec les utilisateurs

Ce type d'évaluation, bien connu en génie logiciel [BOE 84], peut être mis en œuvre lorsqu'il n'existe pas encore d'expérience d'utilisation du système. L'évaluation doit être réalisée selon un cycle itératif tout au long du processus de conception, avec des utilisateurs potentiels (tableau 4).

Technique ou méthode	Principe / Objectif	Critères concernés
Prototypage [JAM 91] [MIL 91] [POU 93]	- Évaluation précoce avec des utilisateurs potentiels de certains aspects particuliers pris un par un (entrées, sorties, qualité des messages...) - Nécessité d'outils de prototypage permettant des modifications rapides de l'IHM. - Biais possibles puisque les tests se font en dehors du contexte réel des tâches.	Tous les critères
Scénari-mages (story-board) [WIL 90] [NIE 93]	- Présentation aux utilisateurs d'un enchaînement d'écrans dessinés sur papier ; premier support de dialogue pour l'équipe de développement. - En raison de l'apparition d'outils graphiques de plus en plus évolués, possibilité de les dessiner directement sur écran. - Méthode s'inscrivant dans les premières étapes d'une démarche de prototypage (Cf. ci-dessus) ; utilisée depuis longtemps intuitivement par de nombreux concepteurs.	1-5, 9, 11, 12, 16-18
Analyse des menus (en tant que sélection d'alternative) [MCD 86] [PAL 87]	- Mise en évidence et description du schéma mental de la hiérarchie des menus que doit avoir l'utilisateur de l'IHM, par la "technique du tri conceptuel". Mise en relation de la description ainsi identifiée avec le modèle mental utilisateur. - Comparaison possible des modèles mentaux d'utilisateurs expérimentés et novices. - Technique utilisée surtout en ingénierie de la connaissance.	1, 7-9, 11, 12, 16, 18
Groupes de discussion [CAC 93]	- Discussion des différents aspects de l'IHM, entre des experts en communication H-M, des utilisateurs et des concepteurs. - Revue possible de problèmes, de besoins, etc. - Caractéristique des démarches dites participatives.	tous les critères potentiellement
Banc d'essai final [NOV 87] [WAL 90] [ROW 94]	- Contrôle final avant commercialisation ou intégration sur site de la qualité par une mesure d'utilisabilité globale du système interactif (performance, fonctionnalités, IHM utilisateur, documentation, traduction, apprentissage, etc.). - Recueil empirique de données objectives et subjectives sur des stations d'évaluation spécialement aménagées : caméras, microphones, magnétophones, chronomètres, glaces sans tain, mouchards électroniques... (Cf. stations d'évaluation de LOTUS, MICROSOFT, WORDPERFECT...).	Tous les critères

Tableau 4. Méthodes de tests de conception avec les utilisateurs

A chaque phase de développement, les solutions envisageables pour un problème spécifique sont hiérarchisées. Puis, des données sont recueillies au moyen d'une batterie de tests qui engendrent des aménagements mis à l'épreuve à leur tour dans une nouvelle version, et ainsi de suite jusqu'à satisfaction. Les alternatives sont évaluées par rapport aux objectifs du système final, les résultats servant de base pour les étapes suivantes de la démarche de conception. Il est dans ce cas possible d'évaluer l'IHM au fur et à mesure du projet, et on parle alors d'évaluation itérative (fig. 6). Des expérimentations peuvent être mises en place pour tester des alternatives de conception avec les utilisateurs (Cf. l'analyse des menus, fig. 7).

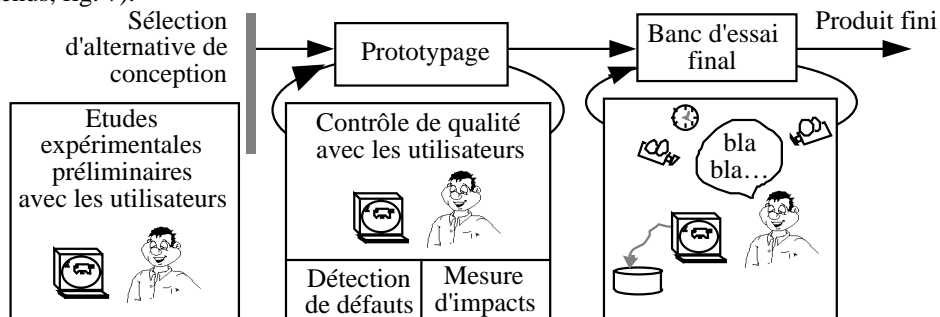


Figure 6. Démarche globale de l'évaluation itérative [SEN 90]

Positionnement dans le cycle de développement : Le prototypage est utilisé **a priori** lors de la conception proprement dite, tout en pouvant s'accompagner de groupes de discussion. Tout au début du projet, des sélections d'alternatives peuvent être effectuées, et des maquettes papier (technique des scénarimages) peuvent être proposées aux utilisateurs.

Les bancs d'essais concernent quant à eux l'évaluation **a posteriori** du système interactif qui vient d'être réalisé ; ils interviennent donc lors de l'étape de *tests*, avant l'exploitation.

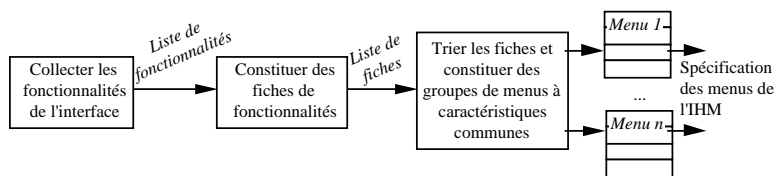


Figure 7. Adaptation de la "technique du tri conceptuel" pour l'identification de menus adaptés à l'utilisateur

4.2. Les approches centrées sur une expertise humaine ou "papier"

L'évaluation des aspects de communication homme-machine d'un système interactif nécessite souvent une analyse qualifiée d'experte. Cette approche vient en complément des approches empiriques pour vérifier qu'un maximum de critères ont été pris en compte. Cette approche sera également obligatoire s'il n'est pas possible de faire intervenir des utilisateurs expérimentés et/ou potentiels dès le début du projet. On distingue plusieurs méthodes présentées dans le tableau 5. Certaines préconisent le respect de recommandations ergonomiques centralisées dans des grilles (fig. 8). D'autres sont fondées sur l'implication personnelle et le jugement d'experts en communication homme-machine.

Technique ou méthode	Principe / Objectif	Critères concernés
Expertise humaine [MOL 90] [POL 92]	<ul style="list-style-type: none"> - Intervention d'un ou plusieurs experts en communication homme-machine pour juger de la qualité ergonomique d'une IHM et proposer des améliorations. - Revue systématique de l'IHM, en la confrontant avec les points de vue des utilisateurs si ceux-ci existent. - Inconvénient : chaque spécialiste se focalise sur des aspects particuliers et fonde son évaluation sur une démarche globale qui lui est propre. Pour avoir une vision complète des problèmes, nécessité de recourir à plusieurs spécialistes. 	Tous les critères
Cognitive Walk-throughs [LEW 90] [POL 92]	<ul style="list-style-type: none"> - Méthode particulière d'intervention d'un expert humain (Cf. ci-dessus) qui suit les étapes suivantes : recensement des tâches humaines à évaluer, des séquences d'actions que l'utilisateur devra effectuer, remplissage d'un questionnaire spécifique pour chacune de ces actions. - L'analyse des résultats s'appuie une théorie cognitive de l'apprentissage par exploration [KIE 85] (Cf. tableau 9). - Méthode permettant de prédire des aspects de l'IHM facilitant la résolution des problèmes rencontrés par l'utilisateur, ainsi que le processus d'apprentissage. 	1-4, 6-10, 14, 17, 18
Expérience de l'analyste [OMB 55]	<ul style="list-style-type: none"> - Étude du système interactif et de ses difficultés d'utilisation par un évaluateur qui se met à la place des utilisateurs et apprend leur métier. - Méthode préconisée seulement si les tâches sont simples. - Biais car l'évaluateur se prend pour modèle, alors qu'il ne deviendra sans doute jamais un utilisateur expérimenté. 	1-5, 9-13, 16, 17
Grille d'évaluation [RAV 89]	<ul style="list-style-type: none"> - Revue de qualité de l'IHM par notation systématique par rapport à une liste de critères ("check-list"). - L'analyse des données recueillies permet la mise en évidence de points forts et points sensibles de l'IHM. - Les recommandations de la littérature peuvent composer les entrées d'une grille (Cf. Guide de 944 conseils ou principes [SMI 86], aide-mémoire des questions ergonomiques à se poser [SCA 87], guide de 3700 principes [VAN 94]). - Difficulté d'utilisation par un non spécialiste de l'ergonomie, questions à se poser parfois trop générales ; la présence du spécialiste est donc indispensable. 	Tous les critères

Tableau 5. Approches "expertes" d'évaluation

GRILLE D'ÉVALUATION		Note				
		1	2	3	4	5
Prise en compte de l'expérience utilisateur, respect du niveau d'expertise						
1	L'information est-elle présentée dans le même état d'esprit que l'utilisateur s'en fait ?					
2	L'affichage graphique est-il compatible avec la représentation que se fait l'utilisateur du système ?					
3	La structure du système est-elle adaptée à la perception de la tâche de l'utilisateur ?					
4	Le déroulement des activités exigées pour effectuer une tâche est-il en accord avec les attentes de l'utilisateur ?					
5	Le système travaille-t-il dans la même voie que la pensée de l'utilisateur ?					
6	Estimation globale du système en terme de prise en compte de l'expérience de l'utilisateur ?					
Clarté visuelle <i>Caractéristiques lexicales de présentation des informations (cognitif, perceptif)</i>						
Groupement par localisation, positionnement des items, appartenance-distinction.						
1	Chaque écran est-il clairement identifié (titre informatif ou description) ?					
2	L'organisation des informations sur l'écran est-elle toujours logique ?					
3	La séparation entre les différents informations est-elle...					

Figure 8. Extrait d'une grille d'évaluation (inspiré de [RAV 89])

Positionnement dans le cycle de développement : Lorsque les données relatives à l'utilisation ne peuvent être enregistrées, ou que l'IHM n'existe pas encore, l'évaluation doit se faire avec un modèle de référence. Le modèle peut être construit **a priori** au moyen de l'expertise humaine, de la méthode "cognitive walkthroughs" et des grilles d'évaluation dès l'étape de *spécification*. En effet, les spécifications qui concernent la définition des besoins, de l'architecture de l'IHM, des enchaînements graphiques, la présentation d'information et des traitements doivent être évaluées avant que l'IHM ne soit conçue. Puis, l'équipe de développement peut disposer de ces trois méthodes "expertes" **a posteriori** jusqu'à l'exploitation.

Concernant l'expérience de l'analyste, celle-ci n'est envisageable avec précaution **a posteriori** qu'une fois l'IHM réalisée (ou **a priori** avec un prototype).

4.3. Les approches centrées sur une modélisation de l'IHM et/ou de l'interaction homme-machine

Ces approches, qualifiées également d'analytiques, sont basées sur des modèles formels —informatisés ou non—, ainsi que sur la mise en œuvre de métriques dans la mesure du possible objectives. Afin de mettre en évidence les approches existantes possibles pour l'évaluation, nous reprendrons ici la distinction de Senach [SEN 90] concernant les modèles qualifiés de prédictifs et ceux dits de qualité.

4.3.1. Modèles formels dits prédictifs

La modélisation des tâches et/ou des activités humaines est particulièrement importante car elle facilite la communication entre les intervenants de l'équipe de développement. Les besoins des utilisateurs en termes de fonctionnalités, d'informations ou de formation complémentaire peuvent ainsi être plus facilement mis en évidence. Cette modélisation peut être effectuée au moyen de différents modes de description généralement inspirés d'une décomposition en buts/sous-but, graphique ou non selon une forme plus ou moins propice à l'analyse (fig. 9). Les tableaux 6 à 9 listent les plus connus parmi ceux applicables aux applications interactives.

De manière générale, les modèles formels dits prédictifs peuvent être utilisés pour prédire certains aspects liés à l'interaction homme-machine (hiérarchie des tâches, cheminement des actions de l'utilisateur, alternatives, temps requis pour réaliser une tâche...) et faciliter la mise en évidence de problèmes potentiels. Des métriques sont progressivement mises en place par les chercheurs pour mesurer objectivement des aspects associés à la qualité de l'interaction (cohérence, compatibilité avec l'image mentale de l'utilisateur, qualité des écrans...).

Le tableau 6 s'intéresse aux modèles formels permettant une représentation graphique des tâches, indépendamment (c'est le cas pour HTA, GOMS, KLM, MAD et SADT-Petri) ou non (c'est le cas pour JSD* et DIANE) d'une méthode de développement issue du génie logiciel. Un exemple de description d'une tâche à l'aide de MAD est visible en figure 10.

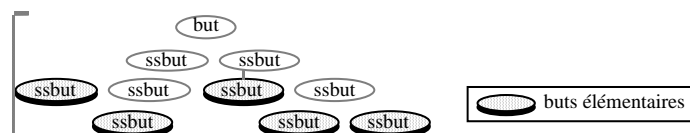


Figure 9. Description hiérarchique de la tâche sous la forme de buts et sous-but

<i>Technique ou méthode</i>	<i>Principe / Objectif</i>	<i>Critères concernés</i>
<i>Modèles sans liaison avec une méthode de développement issue du génie logiciel.</i>		
HTA [ANN 67] [DUN 81]	- Décomposition hiérarchique graphique de la tâche globale en sous-tâches élémentaires pour mettre en évidence les compétences et informations nécessaires à l'utilisateur pour chaque but ou sous-but, compte tenu de la structure du système. - "Ancêtre" des modèles de tâches, a servi au départ pour mettre en évidence des besoins en formation, ainsi que comme moyen d'analyse de la fiabilité humaine.	1, 7, 8, 11, 12
MAD [SCA 89] [SEB 94]	- Description graphique des tâches (principes de planification hiérarchique) à partir des données de l'analyse du travail, en insistant sur la logique de traitement de l'information. - Chaque tâche est associée à une fiche textuelle descriptive de l'état du système homme-machine. - Des opérateurs spécifiques permettent de décrire des tâches parallèles ou séquentielles, des bouclages et des alternatives. - Issue de l'INRIA, elle fait l'objet d'extensions actuellement, par exemple dans un but de spécifications d'IHM [HAM 93].	1, 7, 8, 11, 12
GOMS et KLM [CAR 83] [IRV 94]	- Description de l'activité cognitive d'un utilisateur engagé dans la réalisation d'une tâche de routine sans erreur. - Modélisation en termes de buts organisés hiérarchiquement, d'opérateurs ou actions élémentaires, de méthodes ou procédures de réalisation d'un but et de règles de sélection des méthodes lorsque plusieurs solutions permettent d'atteindre un même but. - Le modèle KLM associe à GOMS des méthodes de calcul de temps de réalisation de tâches élémentaires [HAU 94]. - Prédiction des performances, du langage de commande, des capacités motrices de l'utilisateur, des temps de réponse du système et description de la méthode de réalisation de la tâche. - Sans doute le plus étudié actuellement (courant nord-américain) - Il fait l'objet d'extensions considérant les aspects temporels des tâches (CPM-GOMS, [JOH 92]).	1, 7, 8, 11, 12
SADT-Petri [ABE 90] [BEN 93] ou TOOD [MAH 95]	- Représentation graphique du système H-M basée sur un modèle de tâche caractérisant le comportement de la machine, de l'utilisateur et de leurs interactions. - Décomposition fonctionnelle du système de façon descendante, modulaire, hiérarchique et structurée par des diagrammes SADT ; puis description pour chaque actigramme SADT du comportement décisionnel humain face à une situation donnée avec des réseaux de Petri synchronisés. - Issue du LAMIH, la méthode a été validée lors de l'analyse des tâches prescrites et des activités humaines dans des salles de contrôle de systèmes industriels complexes.	SADT : 1, 7, 8, 11, 12 Petri : 4, 9, 10, 13
<i>Modèles en liaison avec une méthode de développement issue du génie logiciel</i>		
DIANE [BAR 88] et DIANE+ [TAR 93]	- Étend la méthode MERISE [TARD 91] en y explicitant de nouvelles notions : tâches humaines, procédures, logique d'utilisation du système... Description graphique des tâches selon le principe de planification hiérarchique. - Couvre surtout la phase d'analyse des besoins, en partie celle de spécification de l'IHM. - L'extension DIANE+ introduit une méthode de génération automatique d'IHM.	1, 7, 8, 11, 12
JSD* [LIM 92]	- Étend la méthode Jackson System Development [JAC 83] avec globalement les mêmes objectifs que DIANE pour MERISE. - Vise à couvrir en continu trois phases : analyse des besoins, "synthèse" et spécification de l'IHM, à l'aide de différents formalismes graphiques et textuels. - Articulation importante avec les étapes de développement avec JSD du système complet.	1-5, 7-9, 11-14, 16, 18

Tableau 6. *Approches analytiques d'évaluation, exemples de modèles de tâche*

Les modèles linguistiques (listés dans le tableau 7) explicitent la structure de l'interaction homme-machine au moyen de grammaires : les plus connus sont ALG, CLG et TAG, utilisés surtout pour des IHM à langage de commande, même si sur la figure 10, on trouve une description par CLG d'une image de supervision. Ces modèles sont utilisables pour décrire les tâches d'interaction entre l'utilisateur et l'IHM, ce qui sous-entend des possibilités

d'évaluation à l'aide de métriques objectives. Les modèles linguistiques sont à retenir pour l'idée de formaliser l'interaction homme-machine comme un langage de programmation.. et donc à terme de pouvoir automatiser l'évaluation. C'est le cas également des modèles formels listés dans le tableau 8, qui sont fondés sur une théorie concernant l'activité cognitive de l'utilisateur.

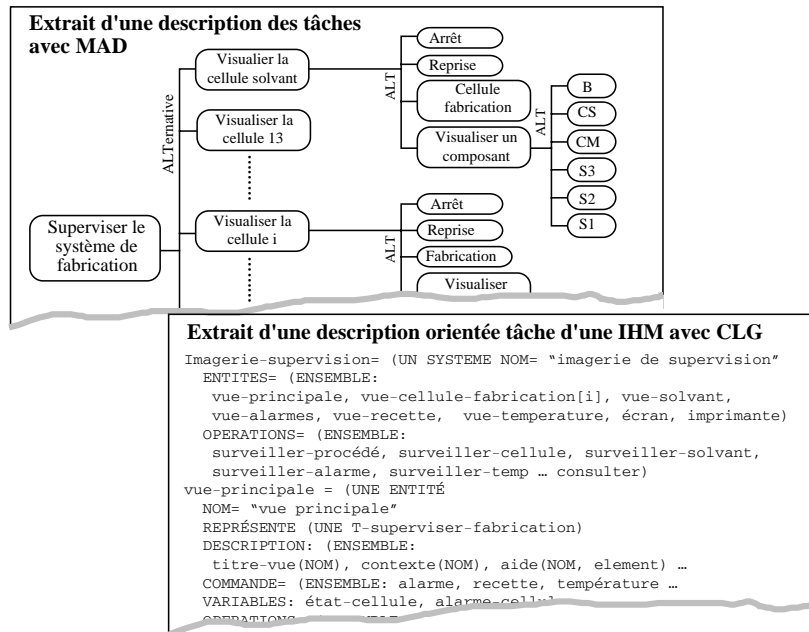


Figure 10. Exemples de descriptions par modèles formels

<i>Technique ou méthode</i>	<i>Principe / Objectif</i>	<i>Critères concernés</i>
ALG [REI 84]	<ul style="list-style-type: none"> - Construction d'un modèle des actions mises en jeu par l'utilisateur, sous la forme de règles de production. - Décomposition récursive des buts des utilisateurs : un symbole initial pour le but général, des phrases pour les procédures et des mots pour les actions élémentaires. - Mise en évidence de 3 "indices" ergonomiques : la complexité du langage, la simplicité des procédures, la cohérence au niveau des séquences d'action. - Utilisable surtout pour les IHM à langage de commande.. 	1, 7, 8, 11, 12
CLG [MOR 81] [COU 90]	<ul style="list-style-type: none"> - Description orientée tâche d'un système interactif, selon six niveaux d'abstraction : tâches et sémantique pour les buts et sous-but, syntaxe et interaction pour la communication entre l'utilisateur et le système, présentation et unités d'entrée-sortie pour la composante physique. - Proposition de métriques relatives à la complexité des tâches, aux composants d'IHM utilisés et à certaines erreurs humaines possibles [SHA 87]. 	1, 4, 8-10, 12, 13, 15, 16, 18
TAG et ETAG [PAY 89] [TAU 90]	<ul style="list-style-type: none"> - TAG, selon globalement les mêmes principes que ALG et CLG, permet une décomposition à l'aide d'une grammaire des tâches en sous-tâches simples routinières pouvant être exécutées par l'utilisateur sans prise de décision. - Proposition de métriques concernant les problèmes de cohérence dans l'utilisation de l'IHM, ainsi que les difficultés et inconsistances liées à l'apprentissage. - ETAG étend TAG en affinant la description servant de base à l'évaluation, selon trois aspects : sémantique de chaque tâche, dictionnaire des tâches, règles de la grammaire d'action. 	1, 7, 8, 11, 12
UAN [HAR 92]	<ul style="list-style-type: none"> - Description orientée action d'IHM à manipulation directe à l'aide d'une notation spécifique. - Dans des tableaux sont exprimés des suites temporelles d'actions de l'utilisateur, les retours d'information fournis par l'IHM et les changements d'état de variables de l'IHM. - La notation a été récemment étendue pour l'expression d'IHM multimodales [COU 93]. 	1, 4, 9, 10, 18

Tableau 7. *Approches analytiques d'évaluation, exemples de modèles linguistiques, sans liaison avec une méthode issue du génie logiciel*

<i>Technique ou méthode</i>	<i>Principe / Objectif</i>	<i>Critères concernés</i>
CCT (Cognitive Complexity Theory) [KIE 85]	<ul style="list-style-type: none"> - Fondé sur une théorie visant à formaliser la notion de complexité du point de vue de l'utilisateur du système informatique. - Description sous forme de règles du modèle de l'utilisateur, description des tâches selon les principes de GOMS. - Utilisation de métriques pour prédire les connaissances requises pour utiliser l'IHM et certaines difficultés d'utilisation. 	1-5, 7, 9-12, 15, 16, 18
ICS (Interacting Cognitive Subsystems) [BAR 87]	<ul style="list-style-type: none"> - Fondé sur une théorie concernant la modélisation cognitive de l'utilisateur sous forme de sous-systèmes sensoriels, de traitement de l'information, et moteurs. - Description de tâches par l'identification des sous-systèmes impliqués. - Utilisation de métriques pour prédire la charge cognitive de l'utilisateur pour une tâche donnée. 	1, 4, 5-8, 14, 17, 18

Tableau 8. *Approches analytiques d'évaluation, exemples de modèles formels fondés sur une théorie, sans liaison avec une méthode issue du génie logiciel*

Positionnement des modèles prédictifs dans le cycle de développement : Tous les modèles prédictifs énumérés s'intéressent à l'interaction homme-machine au sens des tâches et de leur organisation. Ils concernent le niveau conceptuel du développement, c'est-à-dire que l'on peut les utiliser **a priori** dès les étapes *d'analyse* et *de spécification*. Ils sont de plus très importants en tant que support de dialogue entre chaque intervenant du projet.

Un travail de recherche très important doit encore être mené pour faciliter leur articulation avec les méthodes du génie logiciel ; rappelons que seuls les modèles DIANE et JSD* peuvent s'intégrer naturellement dans, respectivement, les méthodes MERISE et JSD. On oublie trop souvent que de tels modèles peuvent en plus être utilisés **a posteriori** pour modéliser les activités des utilisateurs, et comparer cette modélisation à celle effectuée **a priori** vis-à-vis des tâches prévues par les concepteurs, afin d'en tirer les conséquences en terme d'aménagement des IHM et/ou de formation [ABE 90].

4.3.2. Modèles formels dits de qualité de l'IHM, approches logicielles²

Ces modèles formels sont, dans l'état actuel de la recherche, complémentaires des modèles formels dits prédictifs décrits en 4.2.1. En effet, ils s'intéressent aux propriétés mesurables de l'IHM, et ceci selon des critères d'utilisabilité formalisés. On distingue des approches d'évaluation automatique de l'affichage, ainsi que des approches de génération automatique des affichages ou des spécifications d'IHM.

4.3.2.1. Systèmes d'évaluation automatique de l'affichage

Cette approche consiste en la mise en œuvre d'outils informatiques capables d'effectuer automatiquement des mesures ergonomiques sur des pages-écrans évaluées une par une, et ceci indépendamment de leur contexte d'utilisation (Cf. tableau 9). Ils utilisent pour cela des algorithmes spécifiques focalisés sur plusieurs critères d'utilisabilité, comme dans le système DAP ; ou alors ils prennent la forme de systèmes à base de connaissance ergonomiques en rapport avec l'utilisabilité, c'est le cas de SYNOP et de KRI/AG (figure 11). On retrouve aussi une telle approche dans [GIB 92].

<i>Technique ou Principe / Objectif méthode</i>	<i>Critères concernés</i>
DAP (Display Analysis Program) [TUL 88]	2, 3, 5, 8
- Système capable d'estimer automatiquement la qualité des affichages en fonction de six paramètres : densité globale d'information, densité locale, nombre de groupe d'informations distincts, taille moyenne des groupes, nombre d'items, complexité de l'affichage. - Évaluation une à une des pages-écrans alphanumériques, sans référence aux tâches humaines, et proposition de recommandations pour améliorer les pages-écrans.	
SYNOP [KOL 89] [KOL 91]	2, 3, 5, 8, 16
- Évaluation automatique "statique" de synoptiques industriels à partir de recommandations ergonomiques. - Critères liés essentiellement à la présentation des informations (lisibilité, format d'écran, modes de représentation...) - Évaluation une à une des pages-écrans graphiques, sans référence aux tâches humaines, puis modifications automatiques, et propositions de recommandations lorsque les modifications sont impossibles.	
KRI/AG [LOW 92]	2, 3, 5, 8, 16
- Système à base de connaissance d'évaluation automatique d'une IHM de dialogue. - Création de l'écran avec TeleUSE sous X-Windows et traduction en fichiers UIL, analysés à l'aide de règles issues de l'ergonomie des logiciels et de guides de style Motif. - Évaluation une à une des pages-écrans graphiques, sans référence aux tâches humaines, et proposition de recommandations pour améliorer les pages-écrans.	

Tableau 9. *Approches analytiques d'évaluation, exemples de modèles formels de qualité : la complexité perceptive*

² Dans les modèles formels dits de qualité, Senach [SEN 90] distingue également des approches qualifiées de cognitives visant à prendre en compte certains traitements effectués sur l'information ; la plupart de ces approches sont issues de la psychologie cognitive. Pour notre part, nous avons choisi de nous intéresser essentiellement aux modèles formels dits de qualité ayant fait l'objet d'implantations logicielles significatives. En effet, ceux-ci font l'objet d'un courant de recherche de plus en plus important en informatique.

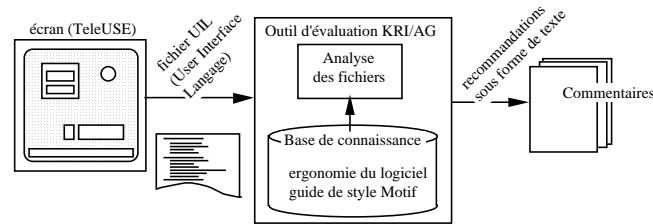


Figure 11. Principe de KRI/AG [LOW 92]

Dans la mesure où ces outils interviennent hors du contexte d'utilisation, ils doivent être obligatoirement associés à d'autres méthodes considérant les tâches humaines. De tels systèmes démontrent la possibilité de rendre opérationnelles les nombreuses recommandations ergonomiques actuellement proposées sous une forme souvent rébarbative (manuels et guides).

Positionnement dans le cycle de développement : Il faut considérer que ces systèmes sont à l'état de prototypes de recherche, même s'ils ont été validés sur des cas concrets. Ils sont censés intervenir **a posteriori** sur des pages-écrans réalisées ou **a priori** sur des pages-écrans en cours de prototypage afin d'apporter une aide à l'évaluateur dans sa démarche globale, en se focalisant sur des critères d'utilisabilité. De tels systèmes devraient bientôt être intégrés dans des environnements graphiques de développement.

4.3.2.2. Systèmes de génération automatique d'affichage

Cette approche consiste en des systèmes à base de connaissances ergonomiques visant la génération automatique d'affichage (c'est le cas d'APT) ou de spécification d'IHM (c'est le cas d'Ergo-Conceptor, figure 12). Ils ont pour objectif de conduire à des affichages respectant **a priori** des concepts ergonomiques de base (Cf. Tableau 10). L'intérêt de ces systèmes consiste avant tout en l'automatisation de plusieurs étapes de la conception et de la réalisation d'IHM. Les vues obtenues doivent être validées ensuite avec les utilisateurs. Dans cette approche à base de connaissance, l'évaluation est donc en fait *sous-jacente* et vise à rejoindre la conception en orientant certains choix (en terme d'utilisabilité) le plus en amont possible.

Ces systèmes sont actuellement à l'état de maquette de laboratoire, mais font l'objet de recherches actives ; voir à ce sujet également les recherches menées autour des systèmes UIDE [FOL 91] et son dérivé USAGE [BYR 94], SAGE [ROT 94], les travaux menés autour de DIANE+ [TAR 93], ou SIROCO [NOR 92].

Technique ou méthode	Principe / Objectif	Critères concernés
APT (A Presentation Tool) [MAC 86]	<ul style="list-style-type: none"> - Automatisation de la présentation de l'information indépendamment de l'application. - Utilisation d'un langage comprenant des primitives graphiques et des opérateurs de composition, qui permettent la génération de modes de représentation de plus haut niveau. - Tests des alternatives de conception en fonction de critères statiques ; représentation des données numériques à l'aide de fonctions de type histogramme, barre-graphe, etc. 	2, 3, 5, 6, 8, 16, 18
Structure logique de l'affichage [PER 87]	<ul style="list-style-type: none"> - Description de la structure de l'affichage, selon différents niveaux d'abstraction (écran en 5 zones, 4 éléments dans la première zone, etc.), puis sous la forme d'un réseau sémantique. - Évaluation selon une approche à base de règles concernant un ensemble restreint de critères d'utilisabilité. - Peu d'information sur les domaines d'application possibles. 	2, 3, 5, 6, 8, 16, 18
ERGO-CONCEPTOR [MOU 92]	<ul style="list-style-type: none"> - Génération automatique des spécifications d'IHM à partir d'une description de l'application (de type procédé industriel à supervisor) selon différents niveaux d'abstraction, à l'aide d'une base de connaissances ergonomiques. - Spécifications d'IHM directement utilisables par le concepteur qui peut réaliser interactivement chaque vue graphique. 	2, 3, 5, 6, 8, 16, 17, 18

Tableau 10. Approches analytiques d'évaluation, exemples de modèles formels de qualité : la Génération automatique d'affichage

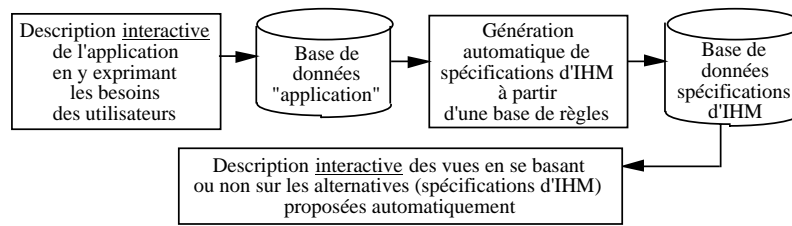


Figure 12. Principe global d'utilisation du système Ergo-Conceptor

Positionnement dans le cycle de développement : lorsque de tels systèmes seront disponibles sur le marché, c'est **a priori** lors des étapes de conception et de codage qu'il devrait être possible d'évaluer l'IHM. Toutefois, il sera important que les futurs systèmes de ce type s'intègrent de manière naturelle dans les méthodes du génie logiciel.

Cette idée de ramener l'évaluation a priori le plus haut possible dans le développement, en la rendant opérationnelle sous la forme le plus souvent d'un système à base de connaissance, augure bien des environnements graphiques à venir.

5. Synthèse sur les techniques et méthodes

Il est maintenant important de situer, par rapport aux différentes étapes d'un cycle de développement de système interactif, l'ensemble des méthodes et techniques recensées. Le cycle en V est celui qui nous paraît à ce sujet le plus parlant. C'est pourquoi nous l'avons utilisé en figure 13. Ainsi, il apparaît d'abord que les *approches empiriques*, les méthodes *d'estimation de la charge de travail* et les méthodes basées sur une *expertise humaine ou papier* auront énormément d'importance lors de l'analyse des besoins si une ou plusieurs situations de référence peuvent être analysées.

On s'aperçoit donc que l'évaluation peut rejoindre la conception, et ceci dans les phases amont du cycle de vie. Ainsi, par la suite, de l'étape de spécification fonctionnelle à celles de codage jusqu'aux tests d'intégration, des *tests de conception* peuvent être mis en œuvre. C'est tout au long de ces étapes que devraient apparaître progressivement des systèmes de *génération automatique* tels que nous les avons décrits (section 4.3.2.2), prenant en compte les caractéristiques des tâches humaines à effectuer. Si l'on se focalise sur l'étape de spécification fonctionnelle, les *modèles formels prédictifs* y prennent de plus en plus d'importance. On peut de plus leur associer une *expertise humaine ou papier*. Pour certains modèles actuels (tel CLG), l'étape de conception globale est également concernée. Dans l'état actuel des recherches, les *systèmes d'évaluation automatique* (hors-contexte d'utilisation de l'IHM) se situent dans les étapes de conception détaillée/codage/tests unitaires. Ils devraient à terme progressivement remonter dans les étapes supérieures en prenant en compte les tâches humaines. Notons à ce sujet le système EMA [BAL 94] encore à l'état de maquette mais qui va dans ce sens en prenant en compte les tâches humaines.

Lorsque le système est réalisé, il est possible de l'évaluer par *banc d'essai final*. Il est également possible d'utiliser les *approches empiriques*, les méthodes *d'estimation de la charge de travail* et *l'expertise humaine ou papier*, qui ont déjà été citées pour l'analyse de situations de référence.

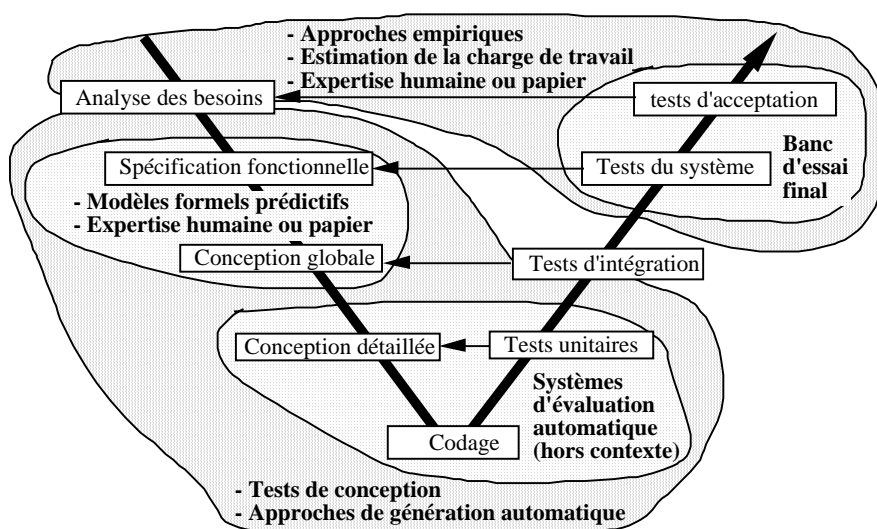


Figure 13. Positionnement des méthodes et techniques d'évaluation

Sur la figure 13, on s'aperçoit que des méthodes et techniques peuvent répondre potentiellement aux besoins de l'équipe d'évaluation pour toutes les étapes du cycle de développement. Pourtant, la réalité est tout autre. Quand elles sont utilisées lors d'un projet, c'est ponctuellement, de manière non systématique, et pas toujours au moment le plus adéquat, et souvent sans couvrir l'ensemble des critères liés à l'utilité et à l'utilisabilité.

6. Conclusion

Cet article a proposé une revue de synthèse des méthodes et techniques d'évaluation des IHM. Malgré leur richesse, actuellement, on constate aisément que ce sont surtout des évaluations à tendance empirique qui sont effectuées, et ceci lors de la phase ascendante —selon le cycle en V— de développement d'une IHM. L'IHM n'est évaluée que lorsqu'elle a été codée. Aussi, des lacunes introduites lors des étapes de spécification/conception ne sont détectées que tardivement alors qu'une évaluation a priori les aurait sans doute détectées plus tôt. Pourtant, des moyens d'évaluation a priori sont maintenant disponibles. Nous en avons donné un aperçu dans l'article. Les raisons pour lesquelles elles ne sont pas utilisées sont de plusieurs ordres.

(1) D'abord, on constate une méconnaissance de la part des concepteurs de l'importance de la prise en compte des facteurs humains lors du développement de systèmes interactifs. D'ailleurs, il est clair que les chefs de projets éprouvent des difficultés à chiffrer le coût et le retour d'investissement d'une évaluation ergonomique.

Le rapprochement progressif des communautés de chercheurs en informatique et en automatique d'une part, et dans les sciences cognitives d'autre part (sur le thème de la Communication Homme-Machine en particulier), de même que l'enseignement de l'ergonomie dans de plus en plus de formations techniques, devraient permettre à terme d'améliorer la situation.

(2) Un autre problème réside dans l'inadéquation entre les méthodes du génie logiciel et celles utilisées en ergonomie et en psychologie cognitives.

On s'aperçoit cependant que des tentatives prometteuses de rapprochement existent (Cf. DIANE et JSD*). D'autres sont en cours.

(3) Il faut aussi souligner l'insuffisance des formalismes et des métriques objectives dans les méthodes d'évaluation les plus utilisées par les spécialistes de la communication homme-machine. On peut aussi insister sur les insuffisances concernant les systèmes supports à l'évaluation (très peu sont pour l'instant sortis des laboratoires).

Bien qu'encore peu répandus, les modèles qualifiés de formels font l'objet d'un courant de recherche important et devraient à notre avis déboucher rapidement sur des systèmes opérationnels.

(4) Les méthodes et techniques cités dans cet article sont souvent considérées indépendamment les unes des autres, et paraissent d'ailleurs pour la plupart rébarbatives aux spécialistes des développements logiciels.

Il est clair qu'un travail de mise en cohérence des méthodes et techniques actuelles devrait être mené, de manière à faciliter leur articulation vis-à-vis des étapes du cycle de développement d'un système interactif.

Un objectif global de la recherche dans le domaine est actuellement de s'orienter vers des démarches d'évaluation positionnées tout au long du cycle de vie des systèmes interactifs : l'évaluation doit être réalisée à chaque étape. Par rapport à un cycle en V, nous avons insisté sur le fait que ceci peut être effectué **a priori** dès les phases d'analyse des besoins, spécification/conception, implémentation et ceci jusqu'à la validation de l'IHM et son exploitation, c'est-à-dire **a posteriori**. Notons que d'autres cycles étendant les plus connus du génie logiciel sont actuellement envisagés. Par exemple, Hix et Hartson proposent un modèle en étoile centrant le développement logiciel sur l'évaluation [HIX 93] ; Kolski propose un modèle appelé \square (nabla) explicitant la prise en compte de facteurs humains lors du développement des systèmes interactifs [KOL 95].

Malgré son importance, il reste encore un long chemin à parcourir pour que l'évaluation d'IHM entre dans les "us et coutumes" du génie logiciel, bien que de nombreux chercheurs s'y attellent pourtant depuis plusieurs années. L'évaluation des IHM constitue un domaine de recherche extrêmement riche, en pleine évolution au niveau des méthodes et techniques, et en conclusion de cet article, il nous paraît pertinent de citer un commentaire de Senach [SEN 90] qui restera longtemps d'actualité : "*L'évaluation [...] constitue un processus parallèle à la conception, [...] ne peut pas être réduite à une simple étape ponctuelle [...] une équipe d'évaluation doit pouvoir utiliser un ensemble d'outils — allant des techniques cliniques aux méthodes formelles — à chaque étape de la vie d'un produit.*"

6. Bibliographie

- [ABE 90] ABED M., Contribution à la méthodologie de la tâche de conception par des outils de spécification exploitant les mouvements oculaires. Application à la conception et l'évaluation des IHM, *Thèse de doctorat*, université de Valenciennes, 1990.
- [AMA 86] AMALBERTI R. et al., Charge de travail mental, validation en France d'une technique d'évaluation subjective américaine : SWAT. *Rev. Méd. Aéro. Spat.*, tome xxv, n° 98, p. 158-163, 1986.
- [ANN 67] ANNET J.F., K.D. DUNCAN, Task analysis and training design. *Occup. Psychology*, 41, p. 211-221, 1967.
- [BAL 94] BALBO S., Evaluation ergonomique des interfaces utilisateur : un pas vers l'automatisation. *Thèse de doctorat*, Université de Grenoble I, 1994.
- [BAR 87] BARNARD P.J., Cognitive resources and the learning of human-computer dialogs, in *Interfacing Thought*, J.M. Carroll (Ed.), MIT Press, pp. 112-158, 1987.
- [BAR 88] BARTHET M.F., *Logiciels interactifs et ergonomie. Modèles et méthodes de conception*. Dunod informatique Paris, 1988.
- [BAS 93] BASTIEN J.M.C., D.L. SCAPIN, Ergonomic criteria for the evaluation of Human-computer Interfaces. Rapport Technique INRIA n°156, Juin 1993.
- [BEN 87] BENSOUA S., Modèle de description pour l'évaluation de la charge mentale par la méthode de la double tâche. *Thèse de docteur-ingénieur*, université de Saint-Etienne, 1987.

- [BEN 93] BENAÏSSA M.L., Une démarche de conception, réalisation et évaluation d'IHM : application au projet ferroviaire ASTREE. *Thèse de doctorat*, université de Valenciennes, 1993.
- [BENY 93] BENYSH D.V. et al., A comparative review of knowledge structure measurement techniques for interface design. *Int. J. of Human-Computer Interaction*, 5 (3), p. 211-237, 1993.
- [BER 92] BERGER T., Contribution à l'étude de l'activité cognitive de l'opérateur et à l'évaluation de sa charge de travail. *Thèse de doctorat*, université de Valenciennes, 1992.
- [BIA 93] BIAS R.G. et al., Three Usability Enhancements to the Human Factors-Design Interface. *Proc. of the HCI International'93*, vol 2, Salvendy & Smith (Eds), Elsevier, p. 169-174, Orlando, Florida, August 8-13, 1993.
- [BOE 84] BOEHM B.W. et al., Prototyping versus specifying : a multiproject experiment. *IEEE transactions on Software Engineering*, 10 (3), May 1984.
- [BOO 94] BOOCH G., *Conception orientée objets et applications*, 2ème édition, Addison-Wesley, 1994.
- [BRO 88] BROWN C.M., *Human-Computer Interface Design Guidelines*. NJ: Ablex, 1988.
- [BYR 94] BYRNE M.D. et al., Automating interface evaluation. *CHI'94 Conf. Proc.*, ACM Press, Addison-Wesley.
- [CAC 93] Special issue about Participatory Design. *Communication of the ACM*, 36 (4), June, 1993.
- [CAR 83] CARD S. et al., *The psychology of human-computer interaction*. Lawrence Erlbaum, 1983.
- [COU 90] COUTAZ J., *Interfaces homme-ordinateur : conception et réalisation*. Dunod, Paris, 1990.
- [COU 93] COUTAZ J., FACONTI G., PATERNO F., NIGAY L., SALBER D., *MATIS : a UAN Description and Lesson Learned*, Amodeus Report, SM/WP14, 1993.
- [COU 94] COUTAZ J., BALBO S., *Evaluation des interfaces utilisateur : taxonomie et recommandations*. Actes des sixièmes journées sur l'Ingénierie des Interfaces Homme-Machine, IHM'94, 211-218, Lille, 8-9 décembre, 1994.
- [CUR 94] CURTIS B., HEFLEY B., A Wimp no more, the maturing of user interface engineering. *Interactions*, ACM, January, p. 22-34, 1994.
- [DEB 93] DEBERNARD S., Contribution à la répartition dynamique des tâches entre opérateur et système automatisé : application au contrôle du trafic aérien. *Thèse de doctorat*, université de Valenciennes, 1993.
- [DEN 94] DENNING P.J., P.A. DARGAN, A Discipline of software architecture. *Interactions*, ACM, January, p. 55-65, 1994.
- [DIE 90] DIENG R., Méthodes et outils d'acquisition des connaissances. *Rapport de Recherche INRIA n°1319*, novembre 1990.
- [DUN 81] DUNCAN K.D., Training for fault diagnosis in industrial process plant. In *Human Detection and Diagnosis of System Failures*, J. Rasmussen and B. Rouse (Eds.), Plenum Press, New-York, 1981.
- [ERI 84] ERICSSON A.K., H.A. SIMON, *Protocol analysis: verbal reports as data*. Cambridge, MA: MIT Press, 1984.
- [FOL 91] FOLEY J.D. et al., UIDE: an intelligent user interface design. In *Architectures for intelligente interfaces: elements and prototypes*, Sullivan, Tyler (Eds.), p. 339-385, MA: Addison-Wesley, 1991.
- [GIB 92] GIBBONS S.C., Towards an expert system based menu interface evaluation tool. *Rapport de recherche INRIA n°1581*, janvier 1992.
- [GRI 93] GRISLIN M., C. KOLSKI, J-C. ANGUE, Towards an Organization of Man-Machine Interface Evaluation Techniques using a Usability Criteria Grid. *XII European Annual Conf. on Human Decision Making and Manual Control*, university of Kassel, Germany, June 22-24, 1993.
- [HAM 92] HAMMONTREE M.L., HENDRICKSON J.J., HENSLEY B.W., Integrated data capture and analysis tools for research and testing on graphical user interfaces, in *CHI'92 Conf. Proc.*, Monterey, 3-7, May, ACM Press, pp. 431-432, 1992.
- [HAM 93] HAMMOUCHE H., De la modélisation des tâches à la spécification d'interfaces utilisateur. *Rapport INRIA n° 1959*, juillet 1993.
- [HAR 88] HART S.G., L.E. Staveland, Development of NASA-TLX (task load index): results of empirical and theoretical research. In *Human mental workload*, Hancock and Meshkati (Eds.), North-Holland, Advances in psychology, p. 139-184, 1988.
- [HAR 92] HARTSON H.R., GRAY P.D., Temporal aspects of tasks in the user action notation, *Human-Computer Interaction*, vol. 7, pp. 1-45, 1992.
- [HAU 94] HAUNOLD P, W. KUHN D.F., A Keystroke level analysis of a graphics application: manual map digitizing. *CHI'94 Conf. Proc.*, ACM Press, Addison-Wesley, 1994.
- [HIX 93] HIX D., HARTSON R., *Developing user interfaces, ensuring usability through product & Process*, Wiley, 1993.
- [IRV 94] IRVING S. et al., A GOMS analysis of the advanced automated cockpit. *CHI'94 Conf. Proc.*, ACM Press, Addison-Wesley, 1994.
- [JAC 83] JACKSON M., *System Development*. Prentice Hall, 1983.
- [JAC 93] JACOBSON I. et al., *Le génie logiciel orienté objet, une approche fondée sur les cas d'utilisation*. Addison-Wesley, 1993.
- [JAM 91] JAMES M.G., PRODUSER: PROcess for Developing USER Interfaces. In *Taking Software Design Seriously*, J. Karat (Ed.), Academic Press, 1991.
- [JAU 90] JAULENT P., *Génie logiciel : les méthodes*. 2° édition. Armand Colin, Paris, 1990.
- [JOH 92] JOHN B.E., A.H. VERA, A GOMS analysis of graphic, machine-paced, highly interactive task. *CHI'92 Conf. Proc.*, Bauersfeld et al. (Eds), Monterey, California, May 3-7, p. 251-258.
- [KIE 85] KIERAS D., P.G. POLSON, An approach to the formal analysis of user complexity. *Int. J. of Man-Machine Studies*, 22, 365-394, 1985.
- [KIR 90] KIRWAN B., Human reliability assessment. In *Evaluation of Human works: a practical ergonomics methodology*. Wilson and Corlett (Eds.), Taylor & Francis, 1990.
- [KOE 94] KOENEMANN-BELLIVEAU J. et al., Comparative usability evaluation: critical incidents and critical threads. *CHI'94 Conf. Proc.*, ACM Press, Addison-Wesley.
- [KOL 89] KOLSKI C., Contribution à l'ergonomie de conception des interfaces graphiques homme-machine dans les procédés industriels : application au système expert SYNOP, *Thèse de Doctorat*, janvier, 1989.
- [KOL 91] KOLSKI C., P. MILLOT, A rule-based approach to the ergonomic "static" evaluation of man-machine graphic interface in industrial processes. *Int. J. of Man-Machine Studies*, 35, p. 657-674, 1991.
- [KOL 93] KOLSKI C., *Ingénierie des IHM, conception et évaluation*. Editions Hermès, Paris, 372 pages, 1993.
- [KOL 95] KOLSKI C., Méthodes et modèles de conception des interfaces homme-machine, *mémoire d'Habilitation à Diriger des Recherches*, université de Valenciennes, 1995.

- [KOU 91] KOUBEK R.J., G. SALVENDY, Cognitive performance of super-experts on computer program modification tasks. *Ergonomics*, 8, p. 1095-1112, 1991.
- [LAN 91] LANNING T.R., Let the users design ! In *Taking Software design seriously*, Karat (Ed.), Academic Press, p. 127-135, 1991.
- [LEW 90] LEWIS C., POLSON P., WHARTON C., RIEMAN J., Testing a walkthrough methodology for theory-based design of Walk-Up-and-Use interfaces, *CHI'90 Conf. Proc.*, Seattle, ACM New-York, pp. 235-242, 1990.
- [LIM 92] LIM K.Y., LONG J.B., SILCOCK N., Instanciation of task analysis in a structured method for user interface design, in *Proc. of the 11th Interdisciplinary Workshop on Informatics and Psychology : Task Analysis in Human-Computer Interaction*, Scarding, June, 9-11, 1992.
- [LOW 92] LOWGREN J., T. NORDQVIST, Knowledge-based evaluation as design support for graphical user interfaces. *CHI'92 Conf. Proc., Bauersfeld et al. (Eds)*, Monterey, California, May 3-7, p. 181-188, 1992.
- [MAC 86] MACKINLAY J., Automating the design of graphical presentation of relational information. *ACM Transactions on Graphics*, 5 (2), p. 110-141, 1986.
- [MAG 89] MAGUIRE M., M. SWEENEY, System Monitoring: Garbage Generator or Basic for comprehensive Evaluation system? *Proc. of the fifth conf. of the British Computer Society HCI*, University of Nottingham, 5-8 september, p. 375-394, 1989.
- [MAH 95] MAHFOUHI A. et al., A new methodology for Man-Machine Systems analysis and design Design. *Industrial Ergonomics and Safty Conf.*, Seattle, Washington USA, June 13-16, 1995.
- [MCD 86] MCDONALD J.E. et al., A formal design methodology based on user knowledge. In *Human factors in Computing systems-III*, Mantei and Orbeton (Eds.), ACM, North-Holland, Amsterdam, p. 285-290, 1986.
- [MIL 88] MILLOT P., *Supervision des procédés automatisés et ergonomie*. Hermès, Paris, 1988.
- [MIL 91] MILLER-JAKOBS H.H., Rapid Prototyping: An Effective Technique for System development. In *Taking Software Design Seriously: Practical Techniques for Human-Computer Interaction Design*, J. Karat (Ed.), Boston: Academic Press, p. 273-286, 1991.
- [MOL 90] MOLICH R., J. NIELSEN, Improving a human computer dialogue. *Communications of the ACM*, 33, p. 338-348, 1990.
- [MOR 81] MORAN T.P., The Command Language Grammar: a representation of the user interface of interactive computer system. *Int. J. of Man-Machine Studies*, 15, p. 3-50, 1981.
- [MOU 92] MOUSSA F., KOLSKI C., Vers une formalisation d'une démarche de conception de synoptiques industriels : application au système ERGO-CONCEPTOR. *Colloque ERGO-IA "Ergonomie et Informatique Avancée"*, 7-9 Octobre 1992, Biarritz.
- [NEA 83] NEAL A.S., R.M. SIMON, Playback : a method for evaluating the usability of software and its documentation. In *Human Factors in Computing Systems-I*, A. Janda (Ed.), ACM, North-Holland, Amsterdam, pp. 78-82, 1983.
- [NIE 90] NIELSEN J., MOLICH R., Heuristic evaluation of user interfaces, *CHI'90 Conf. Proc.*, Seattle, ACM New York, pp. 349-356, 1990.
- [NIE 93] NIELSEN J., *Usability Engineering*. Academic Press, Boston, 1993.
- [NOR 92] NORMAND V., Le modèle SIROCO : de la spécification conceptuelle des interfaces utilisateurs à leur réalisation. *Thèse de doctorat*, université Joseph Fourier, Grenoble I, 1992.
- [NOV 87] NOVARA F. et al., The evaluation of products using the usability methodology with proposals for product development. *Working paper A5.3a, ESPRIT Project 385 - HUFIT/04-OLI-11/87*.
- [OMB 55] OMBREDANE A., J.M. FAVERGE, *L'analyse du travail*. Paris, PUF, 1955.
- [PAL 87] PALMER J.E. et al., The design and evaluation of on line help for UNIX EMACS : access mechanisms. In *Human-Computer Interaction, INTERACT'87, Bulinger and Shackel (Eds.)*, Elsevier Science Publishers B.V., North-Holland, 461-466, 1987.
- [PAY 89] PAYNE S.J., T.R.G. GREEN, Task-Action Grammar: the model and developments. In *Task analysis for HCI*, Diaper (Ed), John Wiley and Sons, p. 75-107, 1989.
- [PER 87] PERLMAN G., An axiomatic model of information presentation. *Proc. of the Human Factors Society Meeting*, New York, p. 1223-1229, 1987.
- [POL 92] POLLIER A., Evaluation d'une interface par des ergonomes : diagnostics et stratégies. *Le Travail Humain*, tome 55, n°1/1992, p. 71-96.
- [POLS 92] POLSON P.G., LEWIS C., RIEMAN J., WHARTON C., Cognitive walkthroughs : a method for theory-based evaluation of user interfaces, *Int. J of Man-Machine Studies*, 36, pp. 741-773, 1992.
- [POU 93] POULAIN T. et al., Un atelier de génie logiciel pour la spécification, la réalisation et l'évaluation de synoptiques industriels embarqués. *Génie logiciel & Systèmes experts*, n°31, p. 58-70, juin 1993.
- [RAV 89] RAVDEN S., G. JOHNSON, *Evaluating Usability Of Human-Computer Interfaces*. Chichester: Ellis Horwood, 1989.
- [REI 84] REISNER P., Formal grammars as a tool for analysing ease of use : some fundamental concepts. In *Human Factors in Computer Systems, Thomas and Schneider (eds)*, Norwood, N.J. : Ablex, 1984.
- [RIE 93] RIERA B., Méthodologie d'évaluation des interfaces Homme/Vehicule automobile. *Thèse de doctorat*, université de Valenciennes, 1993.
- [ROT 94] ROTH S.T. et al., 1994, Interactive graphic design using automatic presentation knowledge. *Proceedings Human Factors in Computing systems (CHI'94)*, ACM Press, Addison-Wesley.
- [ROW 94] ROWLEY D.E., Usability testing in the field: bringing the laboratory to the user. *Proceedings Human Factors in Computing systems (CHI'94)*, ACM Press, Addison-Wesley.
- [SAN 87] SANTUCCI G., Evaluation de la charge de travail en aéronautique (deuxième partie). *Médecine aéronautique et spatiale*, tome xxvi, n° 102, p. 146-152, 1987.
- [SCA 87] SCAPIN D.L., Guide ergonomique de conception des IHM. *Rapport Technique INRIA n°77*, 1987.
- [SCA 89] SCAPIN D.L., C. PIERRET-GOLBREICH, MAD : Une méthode analytique des descriptions des tâches. *Actes du colloques sur l'ingenierie des IHMs*, p.131-148, Sophia Antipolis, 24-26 mai 1989.
- [SEB 94] SEBILLOTTE S., Méthodologie pratique d'analyse de la tâche en vue de l'extraction de caractéristiques pertinentes pour la conception d'interface. *Rapport INRIA n° 163*, mars 1994.
- [SEN 85] SENACH B., P. Alengry, Evaluation d'un dispositif d'assistance à la conduite et au dépannage de chaînes de fabrication automatisées. *Rapport INRIA*, Février 1985.

- [SEN 90] SENACH B., Evaluation de l'ergonomie des IHM, *Actes du Congrès ERGO-IA'90, ergonomie et informatique avancée*, Biarritz, p. 19-21..
- [SHA 87] SHARATT B., The incorporation of early interface evaluation into command language grammar specification. In *People and Computers III*, D. Diaper and R. Winder (Eds.), Cambridge University Press : Cambridge, 11-28, 1987.
- [SIM 93] SIMON P., Contribution de l'analyse des mouvements oculaires à l'évaluation de la charge de travail mental. *Thèse de doctorat*, université de Valenciennes, 1993.
- [SIN 90] SINCLAIR M.A., Subjective assessment. In *Evaluation of Human works : a practical ergonomics methodology*. Wilson and Corlett (Eds.), Taylor & Francis, p. 890, 1990.
- [SMI 86] SMITH S.L., J.N. MOSIER, Guideline for designing user interface software. *Report MTR-10090*, ESD-TR-86-278, Bedford, MA: The MITRE Corporation, 1986.
- [SPE 88] SPERANDIO J.C., *L'ergonomie du travail mental*. Masson Paris, Edition II, 1988.
- [SPE 91] SPERANDIO J.C., Les méthodes d'analyse du travail en psychologie ergonomique. *La recherche en Psychologie (Domaine et méthodes) de J.P. Rossi*, Dunod, 1991.
- [SWE 93] SWEENEY M. et al., Evaluating user-computer interaction: a framework. *Int. J. Man-Machine Studies*, 38, 689-711, 1993.
- [TARD 91] TARDIEU H., ROCHFELD O., COLLETTI R., *La méthode Merise, principes et outils, 2ème édition*. Editions d'Organisation (tome 1), Paris, 1991.
- [TAR 93] TARBY J.C., Gestion automatique du dialogue homme-machine à partir de spécifications conceptuelles. *thèse de l'université de Toulouse I*, 1993.
- [TAU 90] TAUBER M., ETAG: Extended Task Action Grammar - A language for the description of the user's task language. In *Human-Computer Interaction - INTERACT'90*, Diaper et al. (Eds.), p. 163-168, North-Holland.
- [THE 89] THEAKER C.J. et al., HIMS : a tool for HCI evaluations. In *People and computers V, The BCS HCI Special Interest Group*, Sutcliffe and Macaulay (Eds.), p. 427-442, Cambridge University Press, 1989.
- [TUL 88] TULLIS T.S., Screen design. *Handbook of Human-Computer Interaction*, M. Helander (ed.), Elsevier Science Publishers B.V. (North-Holland), 377-411, 1988.
- [VAN 94] VANDERDONCKT J., *Guide ergonomique de la présentation des applications hautement interactives*, Presse Universitaires de Namur, 1994.
- [VIL 88] VILLEMEUR A., *Sûreté de fonctionnement des systèmes industriels : fiabilité, facteur humain, informatisation*. Eyrolles, Paris, 1988.
- [VOL 81] VOLLE, *Analyse des données*. Deuxième édition, Economica, Paris, 1981.
- [WAL 90] WALSH P.A., J.V. LAWS, Methods and tools in industry for mobility testing : the STL usability evaluation workbench. *BCS HCI Special Interest Group Meeting*, Camden Town, London, 21 May 1990.
- [WHI 91] WHITEFIELD A. et al., A framwork for human factors evaluation. *Behaviour & Information Technology*, vol 10, n°1, p. 65-79, 1991.
- [WIE 83] WIERWILLE W.W., CASALI J.G., Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator. *Human Factors*, 25 (1), p. 1-16, 1983.
- [WIE 93] WIERWILLE W.W., F. EGGENEIER, Recommendation for mental workload measurement in a test and evaluation environment. *Human Factors*, 35 (2), 263-281, 1993.
- [WIL 88] WILSON G.F., R.D. O'DONNELL, Measurement of operator workload with the neuropsychological workload test battery. In *Human Mental Workload*, Hancock and Meshkati (Eds.), North-Holland, Advances in Psychology, 1988.
- [WIL 90] WILSON J.R., E.N. CORLETT, Evaluation of human work, a practical ergonomics methodology. Taylor & Francis, 1990.
- [YOU 87] YOUMANS D.M., Using video in the design process. *Proc. of IEE colloquium on Evaluation Techniques for Interactive Systems Design*, Organized by Professional Group C5 (MMI), 2 october, Savoy Place, London, p. 2/1-2/3, 1987.

Remerciements. Nous tenons à remercier le Professeur Jean-Claude ANGUE pour ses remarques pertinentes. Merci également aux rapporteurs de la revue pour leurs critiques constructives ; nous pensons plus particulièrement à Mme Jocelyne NANARD pour ses compléments d'information précieux.