



HAL
open science

Méthodes d'évaluation des interfaces homme-machine, utilisation a priori et/ou a posteriori

Christophe Kolski, Martial Grislín

► **To cite this version:**

Christophe Kolski, Martial Grislín. Méthodes d'évaluation des interfaces homme-machine, utilisation a priori et/ou a posteriori. Journées Supervision et Coopération Homme-Machine, Jan 1995, Paris, France. <hal-03383135>

HAL Id: hal-03383135

<https://uphf.hal.science/hal-03383135v1>

Submitted on 11 Apr 2025

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

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



HAL Authorization

Méthodes d'évaluation des interfaces homme-machine, utilisation a priori et/ou a posteriori

Christophe KOLSKI et Martial GRISLIN

*Laboratoire d'Automatique et de Mécanique Industrielles et Humaines (LAMIH)
UA CNRS 1775, Le Mont Houy, BP 311, 59304 Valenciennes cedex*

Résumé — Dans le cycle de développement d'applications de supervision, l'évaluation des interfaces homme-machine est souvent négligée, voire ignorée. Pourtant, de nombreuses méthodes sont actuellement susceptibles d'améliorer leur qualité ergonomique. Notre objectif est de recenser ces méthodes et de discuter de leur utilisation a priori et/ou a posteriori dans le cycle de développement d'applications de supervision.

Mots clés — applications de supervision, interfaces homme-machine, méthodes d'évaluation.

Introduction

Quel que soit le domaine d'application (supervision, systèmes d'information, bureautique, etc), on constate que la conception des interactions homme-machine et l'évaluation de leur qualité posent depuis longtemps de nombreux problèmes pratiques, théoriques, et méthodologiques. Il n'est d'ailleurs pas étonnant de constater que la prise en compte des facteurs humains n'apparaît pas explicitement dans les étapes du cycle de vie des applications interactives : voir à ce sujet les modèles cascade (BOEHM, 1981), spirale (BOEHM et al., 1984) ou encore le modèle en V préconisé par l'AFCIQ. De même, malgré l'effervescence de méthodes issues du génie logiciel (Cf. BOOCH, 1994 ; SOMMERVILLE, 1994 ; etc), l'utilisateur est souvent peu pris en compte dans les méthodes de développement les plus connues. Pourtant, et comme l'expliqueront plusieurs conférenciers lors de ces journées, il a été depuis longtemps démontré que la qualité de l'interface homme-machine est cruciale pour l'efficacité et la fiabilité du système homme-machine complet.

C'est donc par rapport à la problématique des applications interactives de supervision, que nous souhaitons discuter, lors de ces journées, de l'existence de méthodes d'évaluation utilisables a priori et/ou a posteriori lors du cycle de développement de telles applications. Nous nous inspirons pour cela (i) de certaines évaluations menées au LAMIH ces dernières d'années dans des situations industrielles réelles et/ou simulées, et (ii) de réflexions récentes sur les principes d'évaluation a priori et a posteriori (KOLSKI, 1993 ; 1995; GRISLIN et al., 1993 ; 1995). Après une brève présentation du principe de l'évaluation des interfaces homme-machine, nous différencierons les notions d'évaluation a posteriori et a priori. Puis, nous recenserons des méthodes d'évaluation nous paraissant utilisables pour l'évaluation d'imageries de supervision, en indiquant si elles sont utilisables a priori et/ou a posteriori.

I. Principe de l'évaluation des interfaces homme-machine

L'évaluation ergonomique d'une interface homme-machine 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 du dispositif évalué à un modèle de référence permettant d'établir des conclusions" (SENACH, 1990). Le modèle de référence est proportionné par les objectifs qui ont été spécifiés et nécessite la définition de variables cibles avec des techniques de recueil de données (figure 1). La construction et la validation du modèle de référence constituent les principales

sources de difficulté pour une bonne évaluation. En effet, une interaction homme-machine est difficile à formaliser de manière simple et il existe souvent un écart entre la théorie formalisée et la réalité. En outre, il est important de bien délimiter le niveau de précision du modèle afin de pouvoir observer et faire ressortir les aspects pertinents vis-à-vis de l'évaluation. Il s'agit donc de déterminer ce qui permettra de porter un jugement sur l'interface homme-machine concernée.

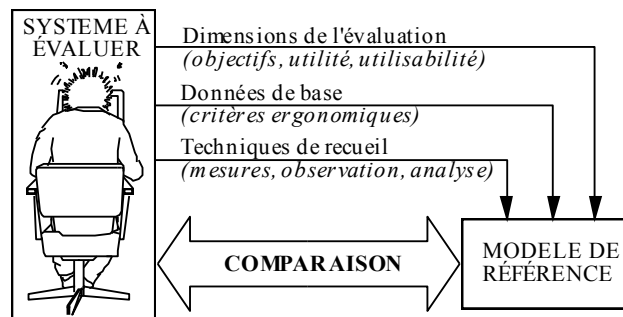


Figure 1 : Principe de l'évaluation

La richesse des fonctionnalités (l'utilité) et la facilité d'utilisation (l'utilisabilité) d'une interface sont deux propriétés importantes pour le diagnostic de la qualité ergonomique. *L'utilité* détermine si l'interface permet à l'utilisateur d'atteindre ses objectifs de travail. Elle correspond par exemple 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 par exemple de facilité d'apprentissage et d'utilisation et de qualité de la documentation, figure 2 (NIELSEN, 1993).

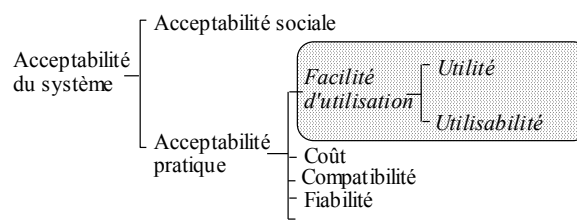


Figure 2 : Utilité et utilisabilité

II. Evaluation a priori, évaluation a posteriori

Une évaluation rigoureuse nécessite de suivre une démarche avec des méthodes et des techniques. Ce recours à une telle aide dépend des caractéristiques de la situation à évaluer et, en particulier de l'état de développement de la solution à évaluer. Lorsque l'interface 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 généralement à réaliser des tests d'utilisabilité avec des utilisateurs habituels ou des sujets sélectionnés pour l'expérience. Ce type d'évaluation s'effectue **a posteriori** sur des interfaces déjà utilisées ou prototypées et aboutit à l'émission de jugements relatifs à son utilisation. Cette approche d'évaluation est la plus répandue pour l'instant.

Pendant, lorsque l'interface à évaluer est inexistante, des méthodes d'évaluation que l'on peut qualifier d'analytiques sont également utilisables. Elles permettent de réaliser une analyse **a priori**, i.e. de fournir une appréciation de l'interface avant son utilisation réelle. L'évaluation repose alors sur la comparaison des informations concernant l'interface —à un stade particulier de son développement— avec un modèle de référence décrivant les propriétés d'une bonne interface. Ce référent est issu des connaissances en ergonomie ou d'une architecture de ce que doit être le dialogue homme-machine.

Certaines méthodes analytiques préconisent le respect de recommandations. Ces méthodes peuvent alors être utilisées au cours de la conception d'interfaces, en vue d'un contrôle a priori de la qualité ergonomique de l'interface prévue. Les principes d'évaluation a priori font l'objet de recherches de plus en plus actives.

III. Revue des méthodes d'évaluation

SENACH (1990a ; 1990b) a classé les méthodes d'évaluation selon deux approches : (i) *l'approche empirique* se base sur la mesure des performances d'utilisateurs représentatifs de la population finale, lors d'expérimentations, alors que (ii) *l'approche analytique* prend en compte des modèles de l'interface et de l'interaction homme-machine, l'évaluation étant réalisée en comparant l'interface à un modèle de référence. Cette section résume le chapitre 7 de l'ouvrage de KOLSKI (1993) et met à jour la classification de SENACH. Pour chaque méthode (ou classe de méthode), son principe de base est résumé et discuté brièvement par rapport au domaine d'application visé.

III.1. L'approche empirique

L'approche empirique permet d'évaluer l'ergonomie de l'interface à partir du recueil et de l'analyse de données provenant de son utilisation par des utilisateurs représentatifs de la population finale, et ceci dans un environnement d'évaluation le plus proche possible de celui d'utilisation. Deux approches sont distinguées : (i) les tests de conception, utilisables lorsqu'il n'existe pas d'expérience d'utilisation du système, et (ii) le diagnostic d'usage, dans le cas contraire, figure 3.

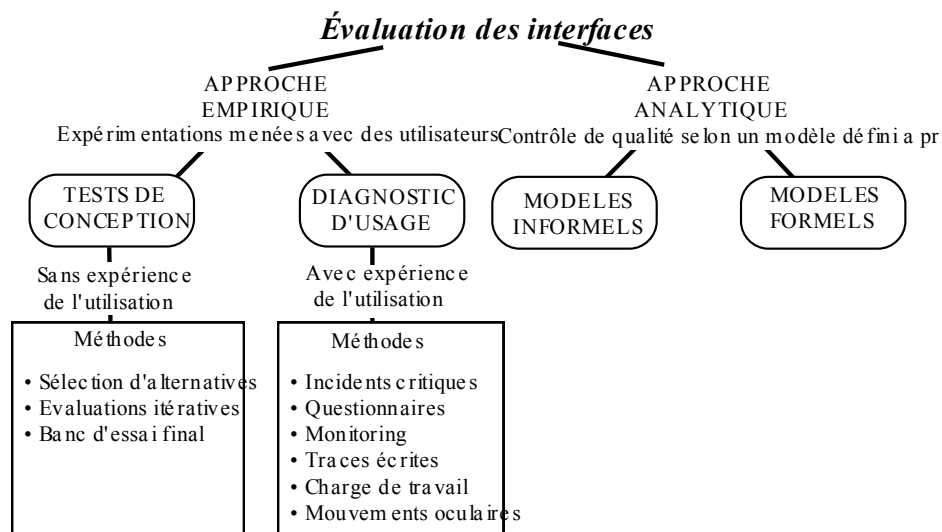


Figure 3 : Approche empirique d'évaluation de l'interface homme-machine

III.1.1. Les tests de conception

Ce type d'évaluation peut être mis en oeuvre **a priori** lorsqu'il n'existe pas encore d'expérience d'utilisation du système. Des tests sont alors réalisés tout au long du processus de conception. Trois démarches seront successivement décrites (figure 4).

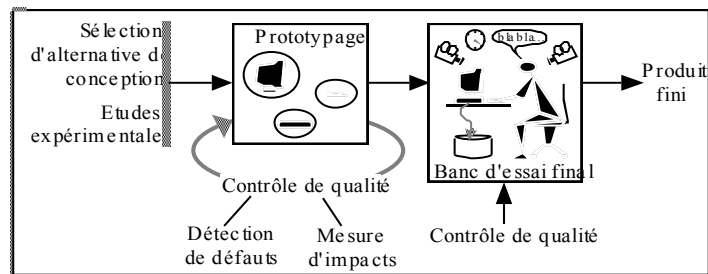


Figure 4 : Tests de conception

La méthode des **tests de conception par sélection d'alternatives** est utilisable lorsqu'il n'existe pas de critère de choix évident entre plusieurs possibilités, et le recueil des données empiriques doit alors permettre de hiérarchiser les solutions envisagées au départ. Cette méthode est la plupart du temps conduite en amont du prototypage (voir ci-dessous). Des alternatives sont donc évaluées en collaboration avec les utilisateurs par rapport aux objectifs du système final, les résultats servant de base pour les étapes suivantes de la démarche de conception. Une des limitations à ce type d'évaluation est qu'elle est effectuée souvent en dehors du contexte réel de la tâche, de ses contraintes, de son environnement. En effet, une alternative obtenant de médiocres résultats en phase d'évaluation peut s'avérer meilleure dans une situation de travail, en se rattachant à d'autres informations suivant la logique de l'utilisateur pour la tâche courante. Cette technique peut être utilisée avantageusement dans un contexte de supervision où la complexité des tâches et la multiplicité des supports d'information utilisés entraînent rapidement la mise en évidence de nombreuses alternatives.

Une autre méthode concerne **l'évaluation itérative de l'interface (prototypage)**. L'apparition actuellement sur le marché de nombreux environnements graphiques facilite la réalisation et la modification d'interface homme-machine par la réduction de l'écriture de code et la réutilisabilité des programmes. La plupart de ces outils facilitent la réalisation itérative de plusieurs versions plus ou moins avancées de l'interface, qu'il est donc possible d'évaluer au fur et à mesure du projet. Le prototypage a pour objectif de minimiser les coûts de développement de l'interface tout en visant l'optimisation progressive de sa qualité selon des critères ergonomiques généraux, ainsi que des critères spécifiques à l'application. L'efficacité de cette démarche dépend directement des possibilités de *réutilisation* des composants logiciels et de *réaménagement* rapides en fonction des résultats de l'évaluation. La dernière version du prototype est celle considérée comme satisfaisante vis-à-vis des besoins des utilisateurs pour les différentes tâches à réaliser. La rentabilité du prototypage est totale si cette dernière version peut être utilisée en tant que produit final. Le prototypage rend nécessaire la participation des utilisateurs à la démarche de conception, tout en servant de support de dialogue entre ces derniers et les concepteurs.

Une autre méthode consiste à mettre en place **un banc d'essai final**. A ce sujet, un système interactif doit respecter de nombreux critères avant de pouvoir être considéré comme *utilisable* (GOULD, 1988 ; NIELSEN, 1993). Avant toute commercialisation, il importe d'évaluer globalement chacun de ces composants et d'en faire ressortir les points forts et les points faibles. Dans ce but, des *stations d'évaluation*, faisant office de *banc d'essai final*, font leur apparition. Il en existe par exemple dans les locaux des constructeurs de logiciels LOTUS, MICROSOFT ou WORDPERFECT. Spécialement aménagés, elles permettent le recueil empirique de données objectives et subjectives à l'aide de méthodes d'observation élaborées (WILSON et CORLETT, 1990). Dans le domaine de la supervision de système complexe, les stations d'évaluation prennent la forme de simulateurs réalistes (LEQUESNE, 1991). La simulation permet aux utilisateurs de se familiariser avec le système, de l'évaluer dans des conditions proches de l'exploitation tout en étant plus économiques, et de proposer des améliorations visant à résoudre les difficultés liées à l'accomplissement de certaines tâches. Bien entendu, la simulation entraîne la nécessité de développer une plate-forme spécifique où l'environnement dynamique de l'application cible doit être reconstitué. Dans les procédés industriels complexes, la mise en place d'un banc d'essai final peut être d'un coût prohibitif.

III.1.2. *Le diagnostic d'usage*

Cette évaluation est effectuée le plus souvent **a posteriori** lorsqu'il existe une expérience d'utilisation du système dans sa globalité. Six techniques, présentées successivement, permettent de diagnostiquer des fonctions ou des modes de représentation défaillants, inutiles, difficiles à exploiter, etc.

La méthode des incidents critiques a été utilisée pour la première fois par FLANAGAN (1954, cité par SINCLAIR, 1990) pour étudier des accidents d'avions évités de justesse. Elle consiste à recueillir systématiquement les dysfonctionnements du système homme-machine à partir d'entretiens avec les utilisateurs et d'observations effectuées sur leur poste de travail. Chaque incident est décrit sous la forme de court récit. Les récits font l'objet d'une classification regroupant des incidents instances d'un même problème. Les problèmes sont ensuite regroupés en classes plus générales. Elles peuvent être ensuite utilisées pour définir les premières fonctionnalités d'un nouveau système, réaliser des aménagements, préciser les objectifs de formation, etc. Cette méthode permet rapidement d'effectuer un diagnostic global des principaux dysfonctionnements. Elle est employée dans la plupart des situations d'analyse et d'évaluation, prenant éventuellement la forme de variantes.

Le questionnaire d'utilisation permet à l'évaluateur de recueillir des informations subjectives sous une forme sûre et structurée, propice à l'analyse. Elle a l'avantage de permettre à l'utilisateur de travailler "à tête reposée", même si ceci est susceptible d'introduire un biais dans ses réponses puisqu'il n'est pas toujours en mesure de se souvenir des difficultés rencontrées lors des différentes situations de travail. Plusieurs travaux ont été consacrés à la manière de réaliser un questionnaire (voir GRAY, 1975 ou SINCLAIR, 1990). En supervision de système complexe, la réalisation d'un questionnaire peut exploiter des travaux spécifiques à ce domaine : GERTMAN et al. (1982) ont mis au point un questionnaire, dont on trouvera un exemple d'utilisation dans COCKERILL-SAMBRE (1984). Il a été ensuite largement étendu par les auteurs (voir GILMORE et al., 1989).

Le monitoring, ou "mouchard électronique", permet de recueillir automatiquement des données objectives à l'aide de l'ordinateur, figure 5. Cette méthode peut renseigner l'évaluateur sur les activités et les performances d'un utilisateur, que ce soit en situation réelle ou en simulation. Elle est généralement associée à des questionnaires et des verbalisations. Elle n'est pas intrusive, ne perturbant pas la réalisation des tâches, et n'apportant donc pas de biais dans les données recueillies. Celles-ci sont généralement traitées automatiquement afin d'obtenir des informations variées : stratégies utilisées, séquençement réel des tâches, performance obtenue, identification des erreurs d'utilisation, etc (MAGUIRE et SWEENEY, 1989 ; DRURY, 1990). Par les nombreuses données temporelles fournies, cette technique peut contribuer à l'amélioration des interfaces et des modules d'assistance.

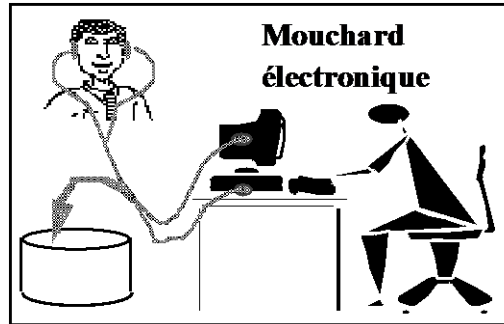


Figure 5 : Principe du monitoring

L'analyse des traces écrites par l'utilisateur lors du travail est susceptible de faciliter la mise en évidence de problèmes d'utilisation. Dans les salles de supervision, on retrouve le plus souvent des rapports qui décrivent les événements survenus, les défauts, les arrêts, les redémarrages, etc. Certaines d'entre elles peuvent mettre en lumière des lacunes dans l'interface. Dans certaines salles, on retrouve aussi des cahiers de doléances, surtout dans les premiers mois faisant suite au lancement d'une nouvelle installation ou imagerie. Ces cahiers peuvent être d'un apport précieux pour l'amélioration de l'interface, en servant d'outil de dialogue entre utilisateurs et développeurs.

La charge de travail est une grandeur utile pour l'évaluation des interfaces homme-machine. En raison de la complexité des tâches à effectuer par les utilisateurs dans le domaine de la supervision, il est particulièrement instructif d'évaluer dans quelles mesure les outils graphiques influent sur leur charge de travail. Il existe de nombreuses méthodes d'estimation de la charge. Celles-ci peuvent être soit subjectives, consistant à demander à l'utilisateur d'indiquer a posteriori son sentiment de charge, le plus souvent au moyen d'échelles subjectives, soit objectives, utilisant des paramètres physiques ou physiologiques mesurés pour estimer la charge de travail (MILLOT, 1988). Les résultats obtenus permettent de mettre en évidence des points sensibles pour les outils graphiques utilisés, et de procéder ensuite à des aménagements de ceux-ci en cas de difficultés lors de l'exécution de certaines tâches (BERGER, 1992 ; RIERA, 1993).

Une autre méthode consiste en **l'analyse des mouvements oculaires**. Il s'avère que dans la supervision, il est fait énormément appel à la perception visuelle des informations. Dans ces conditions, la direction absolue du regard devient un élément utilisable en parallèle avec d'autres moyens pour analyser l'activité d'un utilisateur devant un écran ou un ensemble de supports d'informations. Il est possible, à partir de l'étude de son activité oculo-motrice, d'étudier la manière dont il recherche et localise les informations utiles en fonction des différentes situations, de mettre en évidence des stratégies utilisées et d'identifier des lacunes liées aux outils lors de certaines situations de crise. Il existe de nombreuses techniques de mesure des mouvements oculaires (voir ACIER, 1990 et DERICQUEBOURG, 1991). Bien que leur mise en place pose certains problèmes techniques ou ergonomiques, celles-ci paraissent désormais suffisamment maîtrisées pour contribuer à l'évaluation des stratégies visuelles de l'utilisateur en salle de supervision, et donc à l'aménagement des interfaces. Voir à ce sujet les expérimentations menées par ABED et ANGUE (1990), BERGER (1992) ou encore SIMON (1993).

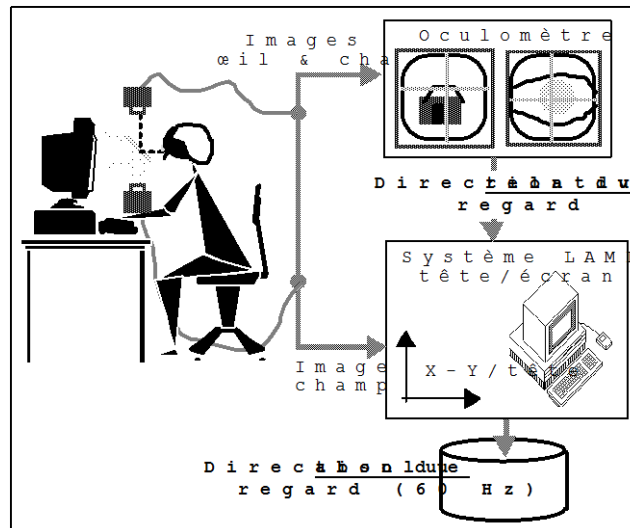


Figure 6 : Principe d'évaluation par analyse des mouvements oculaires

III.2. L'approche analytique

L'approche analytique vise à contrôler la qualité de l'interface selon un modèle défini *a priori* : ce modèle peut être soit formel, soit informel, figure 7.

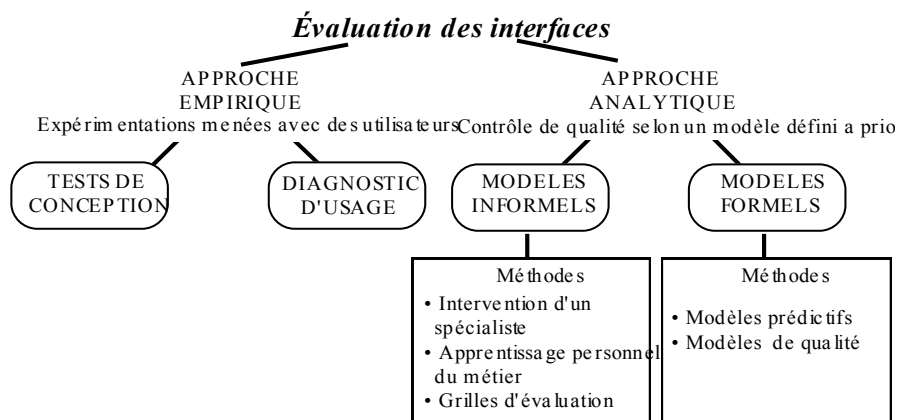


Figure 7 : Approche analytique d'évaluation de l'interface homme-machine

III.2.1. Les modèles informels

Les modèles dits informels (figure 7) peuvent être mis en oeuvre **a priori**, ou **a posteriori**.

Le premier type d'évaluation consiste à **faire intervenir un spécialiste en communication homme-machine** pour juger de la qualité d'une interface et proposer des améliorations. Cette méthode est sans doute celle qui donne les meilleurs résultats. Une des critiques les plus courantes adressées toutefois à ce type de méthode est que chaque spécialiste se focalise sur des aspects particuliers de l'interface et fonde son évaluation sur une démarche globale qui lui est propre (HAMMOND et al., 1984 ; MOLICH et NIELSEN, 1990). L'idéal consisterait donc à faire intervenir plusieurs spécialistes, à confronter puis synthétiser leurs résultats. Malgré ces limitations, pour l'évaluation d'interfaces de supervision, cette méthode est à préconiser dans la mesure où elle permet de diagnostiquer rapidement un certain nombre d'erreurs de conception.

L'apprentissage personnel du métier consiste pour l'évaluateur à se mettre pour un temps à la place d'utilisateurs et à apprendre leur métier. SPERANDIO (1991) ne recommande pas cette méthode

: elle est à exclure lorsque les tâches sont compliquées, elle est rarement utile, l'évaluateur a tendance à se prendre comme modèle, ne peut que devenir un utilisateur sans expérience, donc éprouver des difficultés à analyser objectivement le travail. Cependant, pour des tâches bien où l'évaluateur est en fait un utilisateur potentiel (par exemple en bureautique), la mise en oeuvre de la méthode entraîne moins de difficultés et en tout cas permettre au moins l'évaluation de la prise en main du système. Dans le domaine qui nous préoccupe, cette méthode pose un énorme problème : il est hors de question d'utiliser l'imagerie en ligne au risque de modifier des paramètres de régulateurs, des seuils d'alarmes ou des valeurs de consigne. Afin d'améliorer la compétence de dialogue avec les utilisateurs et de mieux appréhender l'utilisation de fonctionnalités de l'imagerie, il peut être toutefois possible pour l'évaluateur d'utiliser l'imagerie hors ligne, sans conséquence sur le système.

Des **grilles d'évaluation** ont pour objectif d'assister l'évaluateur en recensant des paramètres caractérisant l'ergonomie d'une interface. Pour chacun de ces paramètres, l'interface est notée systématiquement selon une échelle comportant plusieurs points. Par exemple, une méthode d'évaluation par "check-list" a été mise au point par RAVDEN et JOHNSON (1989) dans le cadre d'un projet ESPRIT. Elle est utilisable par des concepteurs, des spécialistes des facteurs humains et des utilisateurs finaux du système. Dans le domaine spécifique de la supervision, GILMORE et al. (1989) fournissent une liste très complète de questions facilitant l'évaluation. Malgré les progrès restant à accomplir dans la définition et l'optimisation de telles grilles, l'application de cette méthode peut d'ores et déjà s'appuyer sur des travaux ayant obtenu des résultats prometteurs. Elle peut être d'un apport très bénéfique pour l'évaluateur puisqu'elle l'oblige à ne négliger aucun des aspects de l'interface.

III.2.2. Les modèles formels

Dans l'approche analytique, l'évaluation de l'interface homme-machine peut être assistée de démarches et d'outils qualifiés de *formels* : on distinguera alors les modèles prédictifs et les modèles de qualité de l'interface (figure 7) visant une évaluation **a priori** des interfaces homme-machine.

III.2.2.1. Les modèles formels prédictifs

Plusieurs **modèles formels dits prédictifs** sont progressivement mis au point, partant de l'hypothèse que certaines performances de l'utilisateur peuvent être prédites, et donc considérées lors de la conception de l'interface. Par exemple :

- *Le modèle GOMS* (Goal, Operator, Method, Selection), défini par CARD et al. (1983), permet de modéliser le comportement de l'utilisateur à différents niveaux d'abstraction, depuis la tâche jusqu'aux actions physiques en termes de buts, d'actions élémentaires, de méthodes pour réaliser un but et de règles de sélection de méthodes. Bien qu'orienté vers la conception, il peut servir pour l'évaluation. Les prédictions concernent la durée de réalisation des tâches, le choix des méthodes et le choix des opérateurs. La démarche consiste à décrire hiérarchiquement l'utilisation d'un dispositif, puis à prédire les performances, et enfin à valider celles-ci dans un contexte expérimental. Ce modèle est très réducteur puisqu'il ne permet de prédire les performances que d'un utilisateur expert ne commettant pas d'erreur.
- *Le modèle Keystroke*, issu également de CARD et al. (1983), est dérivé de GOMS. Il concerne les actions physiques d'un utilisateur expert lorsqu'il effectue une tâche. Il permet de prédire ses performances à partir de : (i) la description préalable de la tâche, (ii) du langage de commande du système, (iii) de paramètres caractéristiques des capacités motrices de l'utilisateur et des temps de réponse du système, (iv) et de la méthode de réalisation de la tâche. Il vise la prédiction du temps d'exécution d'une tâche. En règle générale, il hérite des avantages et des limitations de GOMS. Ses principes possèdent toutefois l'intérêt d'être facilement compréhensibles et de fournir des valeurs numériques simples, bien que sa mise en oeuvre pose rapidement de nombreux problèmes.

- *Le modèle ALG* (Action Language Grammar), provenant de REISNER (1981 ; 1984), permet de représenter sous forme de grammaire les actions de l'utilisateur de l'interface, sous la forme de règles de production du type : *POUR effectuer-telle-action FAIRE telles-opérations*. ALG vise un double objectif : permettre la comparaison entre des alternatives de comparaison selon un critère de facilité d'utilisation, et faciliter l'identification des choix de conception susceptibles d'engendrer des erreurs d'utilisation. ALG permet donc de ramener en quelque sorte l'évaluation lors de la conception de l'interface. Cependant, ce langage ne prend pas en compte la dynamique et les types de modes d'entrées/sorties, ceux-ci ayant une influence indéniable sur la complexité des tâches en fonction du contexte opérationnel.
- Bien qu'orienté vers la conception, *le modèle CLG* (Command Language Grammar) proposé par MORAN (1981) permet comme le précédent une évaluation prédictive de la facilité d'utilisation de l'interface. Un système est décrit en CLG à différents niveaux d'abstraction (tâche, sémantique, syntaxique, interaction, matériel). Le modèle permet de localiser les décisions d'évaluation qui seront critiques pour l'utilisateur. Il propose d'ailleurs différentes métriques permettant de prédire l'efficacité de l'interface. CLG possède comme limitations l'absence d'aide à la structuration et l'absence de moyens d'expression du parallélisme et des interruptions, éléments indispensables de modélisation de l'accomplissement d'une tâche.

De manière générale, ces modèles formels demandent encore à être affinés pour être utilisés dans un contexte de supervision, dans la mesure où ils ont fait surtout l'objet de validations pour des tâches simples, par exemple de bureautique. Notons également l'existence de la méthode MAD issue de travaux menés à l'INRIA par SCAPIN et PIERRET-GOLBREICH (1989). Cette méthode a d'abord été testée sur des tâches de bureau pour ensuite être utilisée pour décrire des tâches plus complexes de contrôle de la navigation aérienne (EL FAROUKI, 1991).

Le LAMIH a mis au point une méthode d'analyse et de modélisation des tâches humaines basée sur l'utilisation de deux méthodes bien connues des informaticiens et automaticiens : SADT et les réseaux de Petri. Elle a été développée au LAMIH, pour répondre initialement à un besoin d'analyse de tâches humaines de contrôle du trafic aérien (ABED, 1990). Cette méthode repose sur une décomposition hiérarchique descendante de plus en plus précise. A l'aide d'actigrammes SADT, il est possible de décrire l'enchaînement des tâches, les données en entrée, en sortie, celles contrôlant la tâche, ainsi que les mécanismes ou supports d'activité exprimant les moyens utilisés pour exprimer la tâche. A l'aide de SADT, il n'est pas possible de représenter la composante dynamique des tâches, c'est pourquoi l'utilisation des réseaux de Petri s'avère indispensable et complémentaire de SADT. Rappelons qu'un réseau de Petri est un graphe constitué de transitions et de place reliés par des arcs orientés. La description sous la forme de réseau de Petri intervient lorsque le niveau de description sous la forme de SADT est suffisamment fin pour faire apparaître une tâche exécutable par l'utilisateur. Un exemple de description à l'aide d'un réseau de Petri d'une tâche terminale de supervision d'un réseau ferroviaire est visible en figure 8 (BENAISSA, 1993).¹

¹ Cette méthode a récemment évolué vers TOOD, qui introduit l'idée de description orientée objet des tâches. Dans cette méthode, notons que le modèle de tâche résultant se présente sous la forme d'un ensemble d'entités coopératives concernant l'application, l'interface et l'opérateur humain. La méthode TOOD a fait l'objet de premières validations (MAHFOUDI et al., 1995).

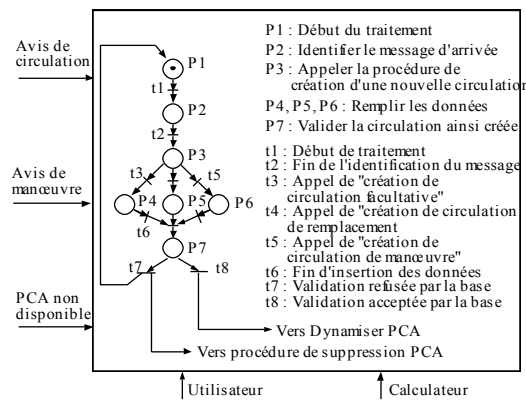


Figure 8 : Modélisation par SADT d'une tâche humaine terminale

Ces modèles formels dits prédictifs sont étudiés avant tout pour faciliter l'évaluation **a priori** des interfaces hommes-machines. Toutefois, il est important de souligner qu'ils peuvent être également utilisés **a posteriori**. Par exemple, ABED et ANGUE (1990) confrontent a posteriori un modèle des tâches prescrites et un modèle des activités réelles d'opérateurs afin de mettre en évidence des lacunes ou des problèmes relatifs à l'utilisation des interfaces homme-machine.

III.2.2.2. Les modèles formels de qualité de l'interface

Il existe d'autres modèles dits **modèles formels de qualité de l'interface**. Ceux-ci s'intéressent aux propriétés mesurables de l'interface, d'un point de vue moins orienté tâche que les modèles précédents : ils sont beaucoup plus focalisés sur la présentation de l'information.

Ainsi, certaines recherches s'intéressent à l'évaluation "automatique" de la présentation de l'information, telle qu'elle est perçue par l'utilisateur. Elles peuvent se baser sur les nombreuses règles ergonomiques recensées dans ce domaine depuis une vingtaine d'années. Par exemple, TULLIS (1984, 1988) s'est orienté sur une démarche automatique d'évaluation d'écran, son programme étant capable de mesurer la densité globale d'information, la densité locale, le nombre de groupes d'informations distincts, la taille moyenne des groupes, le nombre d'items et la complexité de l'affichage. STREVELER et WASSERMAN (1984) ont suivi globalement la même démarche. Notons également les systèmes experts d'évaluation d'interfaces à base de menu en cours de réalisation par GIBBONS (1992) et LOWGREN et NORDQVIST (1992).

SYNOP, développé au LAMIH et actuellement sous forme de maquette de laboratoire, s'intéresse quant à lui à la présentation d'information pour des interfaces de supervision (KOLSKI et MILLOT, 1991) ; L'architecture du système à base de connaissance SYNOP est visible en figure 9. Ce système possède de nombreuses limitations liées à l'incomplétude des connaissances et aux difficultés d'interfacing avec le logiciel de création et d'animation graphique. Il a toutefois démontré la possibilité de prendre en compte certaines tâches de bas niveau d'évaluation et de correction systématiques.

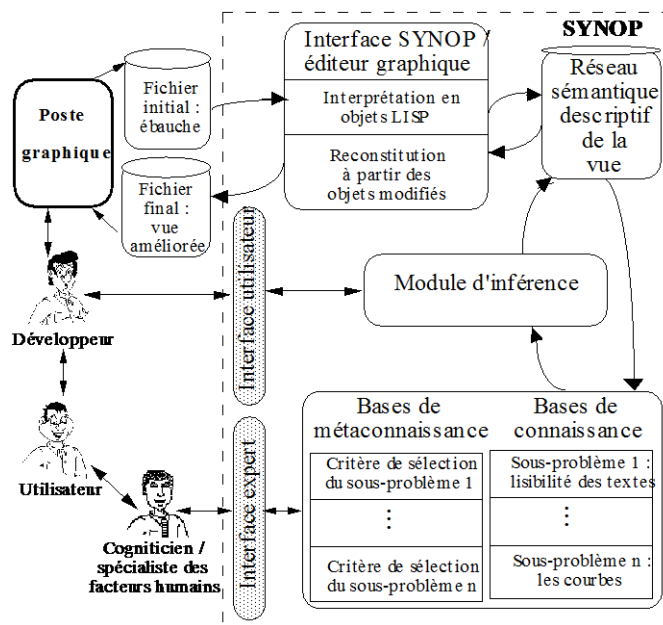


Figure 9 : Architecture du système SYNOP

Il s'agit de noter que selon cette approche **d'évaluation "automatique" de la présentation de l'information**, les vues répondent **a priori** à un ensemble de critères ergonomiques de base. Mais, la principale limitation de tous les systèmes d'évaluation "automatique" est que l'évaluation s'effectue écran par écran, sans s'intéresser aux aspects sémantiques de l'application. Cependant, au vu des résultats obtenus, elles peuvent conduire à terme à des outils opérationnels pour l'ingénierie d'évaluation. Toutefois, de nombreuses recherches doivent encore être menées pour aboutir à leur opérationnalisation.

D'autres recherches visent la **génération automatique d'affichage (ou de spécifications d'interface)** respectant des concepts ergonomiques de base. Comme pour les trois systèmes précédents, on peut s'attendre à terme à des réalisations industrielles concrètes selon les principes cités. Par exemple, le système développé par MACKINLAY (1986) a pour objectif d'automatiser la présentation de l'information, indépendamment de l'application. Les différentes présentations graphiques correspondent à des phrases de langages graphiques s'apparentant à des langages formels. Le formalisme comporte un ensemble de primitives graphiques et d'opérateurs de composition, permettant la génération de modes de représentation de plus haut niveau. Le système développé par PERLMAN (1987) suit globalement la même démarche. Ces systèmes visent des applications purement statiques, loin de la problématique de la supervision de systèmes industriels complexes.

Dans ce domaine spécifique, nous avons mis en oeuvre une maquette de laboratoire appelée ERGO-CONCEPTOR (KOLSKI et MOUSSA, 1991 ; MOUSSA, 1992). Sa démarche suit trois étapes : (i) une étape de description de l'application à superviser orientée par les résultats de l'analyse du système, description en terme de hiérarchie d'abstraction "moyens-objectifs", et de description structurelle et fonctionnelle, (ii) des spécifications d'imagerie sont ensuite automatiquement générées, à l'aide de connaissances ergonomiques, différents niveaux d'abstraction, (iii) ces spécifications sont mises à la disposition du concepteur par l'intermédiaire d'un éditeur dédié. La structure du système est visible en figure 10. Une première validation en laboratoire a démontré la faisabilité de l'approche, puisque les vues répondent **a priori** à un premier ensemble de critères ergonomiques, mais plusieurs lacunes apparaissent, par exemple au niveau de l'incomplétude des bases de connaissance.

Pour toutes les approches actuelles de **génération automatique d'affichage (ou de spécifications d'interface)**, de nombreuses recherches doivent encore être menées avant d'envisager

leur industrialisation, sous forme de connection avec des progiciels de création et d'animation d'imageries industrielles, ou d'intégration dans ceux-ci.

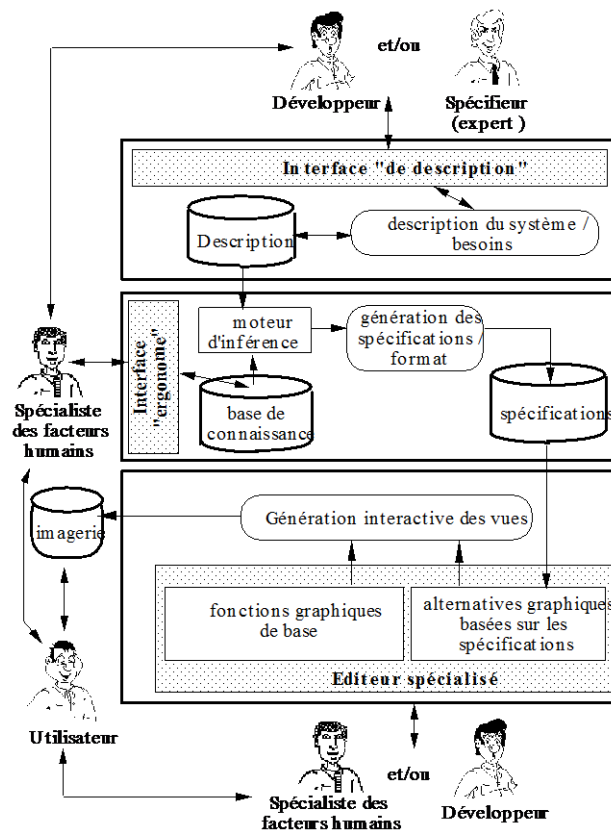


Figure 10 : Architecture du système ERGO-CONCEPTOR

Conclusion

Dans cette revue des méthodes d'évaluation, nous avons délibérément omis de présenter celles qui ne sont utilisables que pour des domaines d'application trop simples. Nous nous sommes focalisés sur celles susceptibles de contribuer à l'évaluation d'interfaces de supervision, tout en insistant sur le fait que certaines de ces méthodes demandent encore de nombreuses recherches et validations.

Nous avons mis en évidence le fait qu'il existe actuellement une grande variété de méthodes contribuant à l'évaluation des interfaces homme-machine. Celles-ci sont issues de nombreuses recherches menées en informatique, en automatique "humaine" et dans les sciences cognitives. A l'aide particulièrement des tests de conception, et des modèles formels et informels, l'évaluation devient même de plus en plus possible au niveau même de la spécification/conception. Cette constatation rejoint le constat fondamental de SENACH énoncé en 1990 : *“Cette organisation [des techniques d'évaluation des interfaces homme-machine] va à l'encontre des idées reçues en matière d'évaluation : elle montre que celle-ci constitue un processus parallèle à la conception, qu'elle ne peut pas être réduite à une simple étape ponctuelle et qu'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.”*

Plusieurs axes de recherche en ce qui concerne les aspects méthodologiques de l'évaluation peuvent être listés sans souci d'exhaustivité : (i) faire le lien entre les modèles et méthodes de développement de systèmes interactifs d'une part, et l'évaluation d'autre part, (ii) viser la mise en place de méthodologies d'évaluation basées sur des batteries de méthodes complémentaires, (iii) étudier le

passage de l'étape d'analyse et de modélisation du système homme-machine à celles de spécification et de conception, étapes dans lesquelles l'évaluation serait menée en parallèle.

Références Bibliographiques

- ABED M. (1990).** *Contribution à la modélisation de la tâche par outils de spécification exploitant les mouvements oculaires : application à la conception et à l'évaluation des interfaces homme-machine.* Thèse de doctorat, université de Valenciennes, Septembre.
- ABED M., ANGUE J.C. (1990).** *Using the measure of eye movements to modelize an operator's activity.* Ninth European Annual Conference on "Human decision making and manual control", Varese, Italy, September 10-12.
- ACIER B. (1990).** *Contribution à la conception et à l'évaluation d'interfaces Homme-Machine par l'analyse du comportement visuel d'un superviseur humain.* Thèse de doctorat, université de Valenciennes, Janvier.
- BASTIEN J.M. (1993).** *Ergonomic Criteria for the Evaluation of Human-Computer Interface.* Rapport Technique INRIA n°156.
- BENAISSA M. (1993).** *Une démarche de conception, réalisation et évaluation d'IHM : application au projet ferroviaire ASTREE.* Thèse de doctorat, université de Valenciennes, décembre.
- BERGER T. (1992).** *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, mars.
- BOEHM B.W. (1981).** *Software Engineering Economics.* Englewoods Cliffs N.J. Prentice Hall.
- BOEHM B.W., GRAY T.E., SEEWALDT T. (1984).** *Prototyping versus specifying : a multiproject experiment.* IEEE transactions on Software Engineering, 10 (3), May.
- BOOCH G. (1994).** *Conception orientée objets et applications.* 2ème édition, Addison-Wesley.
- CARD S.K., MORAN T.P., NEWELL A. (1983).** *The psychology of human-computer interaction.* Hillsdale, NJ : Erlbaum.
- COCKERILL-SAMBRE (1984).** *Prise en compte des facteurs humains dès la conception et l'installation de la coulée continue de la S.A. Cockerill-Sambre-Chertal,* Action Communautaire Ergonomique, Recherche 107 247-11-026, Liège.
- DERICQUEBOURG G. (1991).** *Contribution à l'élaboration d'une méthodologie de conception des interfaces opérateur incluant une commande visuelle.* Thèse de doctorat, université de Valenciennes, Octobre.
- DRURY C.G. (1990).** *Methods for direct observation of performance.* In : Evaluation of Human works : a practical ergonomics methodology. Wilson J.R. and Corlett E.N. (Eds.), Taylor & Francis, pp. 890.
- EL FAROUKI L., SCAPIN D., SEBILLOTTE S. (1991).** *Prise en compte des tâches du contrôleur pour l'ergonomie des interfaces.* Rapport final CENA/N.91223, Convention entre le CENA et l'INRIA, août.
- FLANAGAN C. (1954).** *The Critical incident technique.* Psychology Bulletin, 51, p. 327-386.
- GERTMAN D.I., BLACKMAN H.S., BANKS W.W., PETERSEN R.J. (1982).** *CRT display evaluation : the multidimensional rating of CRT generated displays.* Report for the US Nuclear Regulatory Commission, Washington.
- GIBBONS S.C. (1992).** *Towards an expert system based menu interface evaluation tool.* Rapport de recherche, n. 1581, INRIA, janvier.
- GILMORE W.E., GERTMAN D.I., BLACKMAN H.S. (1989).** *User-computer interface in process control. A Human Factors Engineering Handbook,* Academic Press.

- GOULD J.D. (1988).** *How to design usable system.* In "Handbook of Human-Computer Interaction", M. Helander (ed.), Elsevier Science Publishers B.V., North-Holland.
- GRAY M. (1975).** *Questionnaire typography and production.* Applied Ergonomics, 6, pp. 81-89.
- GRISLIN M., C. KOLSKI, J.C. ANGUE (1993).** *Towards an Organization of Man-Machine Interface Evaluation Techniques using a Usability Criteria Grid.* XII European Annual Conference on Human Decision Making and Manual Control, Kassel, Germany, June 22-24.
- GRISLIN M., KOLSKI C., ANGUE J.C. (1995).** *Human-computer interface evaluation in industrial complex systems : a review of usable techniques.* 6th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems, M.I.T., Cambridge, USA, June 27-29.
- HAMMOND N., HINTON G., BARNARD P., MacLEAN A., LONG J., WHITEFIELD A. (1984).** *Evaluating the interface of a document processor : a comparison of expert judgement and user observation.* In Human-Computer Interaction, INTERACT'84, B. Schackel (Ed.), Elsevier Science Publishers B.V. : North-Holland, IFIP, 725-729.
- KOLSKI C. (1993).** *Ingénierie des interfaces homme-machine, conception et évaluation.* Editions Hermès, Paris.
- KOLSKI C. (1995).** *Méthodes et modèles de conception et d'évaluation des interfaces homme-machine.* Mémoire d'Habilitation à Diriger des Recherches, université de Valenciennes, 4 janvier 1995.
- KOLSKI C., MILLOT P. (1991).** *A rule-based approach to the ergonomic "static" evaluation of man-machine graphic interface in industrial processes.* International Journal of Man-Machine Studies, 35, pp. 657-674.
- KOLSKI C., MOUSSA F. (1991).** *Une approche d'intégration de connaissances ergonomiques dans un atelier logiciel de création d'interfaces pour le contrôle de procédé.* Quatrièmes Journées Internationales : Le génie logiciel et ses applications, EC2, Toulouse, 9-13 décembre.
- LEQUESNE J. (1991).** *Les simulateurs dans le domaine de l'énergie.* Revue Générale de l'Electricité, n°5/91, Mai.
- LÖWGREN J., NORDQVIST T. (1992).** *Knowledge-based evaluation as design support for graphical user interfaces.* CHI'92 Conference Proceedings, ACM, Addison-Wesley.
- MACKINLAY J. (1986).** *Automating the design of graphical presentation of relational information.* ACM Transactions on Graphics, 5 (2), 110-141.
- MAGUIRE M., SWEENEY M. (1989).** *System monitoring : garbage generator or basis for comprehensive evaluation system ?* In People and Computers V, A. Sutcliffe et L. Macaulay (Eds.), Proceedings of the Fifth Conference of the British Computer Society Human-Computer Interaction Specialist Group, university of Nottingham, 5-8 September.
- MAHFOUDI A., ABED M., ANGUE J.C. (1995).** *TOOD : Task Object Oriented Description for Man Machine Systems Design.* 6th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems, M.I.T., Cambridge, USA, June 27-29.
- MILLOT P. (1988).** *Supervision des procédés automatisés et ergonomie.* Editions Hermes, Paris, Décembre.
- MOLICH R., NIELSEN J. (1990).** *Improving a human computer dialogue.* Communications of the ACM, 33, pp. 338-348.
- MORAN T.P. (1981).** *The Command Language Grammar : a representation of the user interface of interactive computer system.* International Journal of Man-Machine Studies, 15, pp. 3-50.
- MOUSSA F. (1992).** *Contribution à la conception ergonomique des interfaces de supervision dans les procédés industriels : Application au système ERGO-CONCEPTOR.* Thèse de Doctorat, soutenue le 15 juillet 1992 à l'université de Valenciennes.
- NIELSEN J. (1993).** *Usability engineering.* Academic Press.

- PERLMAN G. (1987).** *An axiomatic model of information presentation.* Proceedings of the 1987 Human Factors Society Meeting, New York, Human Factors Society, 1229-1223.
- RAVDEN S.J., JOHNSON G.I. (1989).** *Evaluating usability of human-computer interfaces : a practical method.* Ellis Horwood, Chichester.
- REISNER P. (1981).** *Formal grammar and human factors design of an interactive graphics system.* IEEE Transactions on Software Engineering, 7 (2), March.
- REISNER P. (1984).** *Formal Grammar as a tool for analyzing ease of use, some fundamental concepts.* In "Human Factors in Computer Systems", J.C. Thomas, M.L. Schneider (Eds.), Ablex, New Jersey.
- RIERA B. (1993).** *Methodologie d'évaluation des interfaces Homme/Vehicule automobile.* Thèse de Doctorat, université de Valenciennes, janvier.
- SCAPIN D., PIERRET-GOLBREICH C. (1989).** *MAD : une Méthode Analytique de Description des Tâches.* Actes du colloque sur l'ingénierie des interfaces Homme-Machine. Sophia-antipolis, 24-26 Mai.
- SENACH B. (1990a).** *Evaluation ergonomique des interfaces Homme-machine : une revue de la littérature.* Rapport de recherche, INRIA, Sophia Antipolis, n° 1180, Mars, 70p.
- SENACH B. (1990b).** *Evaluation de l'ergonomie des interfaces homme-machine : du prototypage aux systèmes experts.* Actes du Congrès ERGO-IA '90 : Ergonomie et Informatique avancée, Biarritz, septembre.
- SIMON P. (1993).** *Contribution de l'analyse des mouvements oculaires à l'évaluation de la charge de travail mental.* Thèse de doctorat, université de Valenciennes, décembre.
- SINCLAIR M.A. (1990).** *Subjective assessment.* In : *Evaluation of Human works : a practical ergonomics methodology.* Wilson J.R. and Corlett E.N. (Eds.), Taylor & Francis, pp. 890.
- SPERANDIO J.C. (1991).** *Les méthodes d'analyse du travail en psychologie ergonomique.* Dans La recherche en psychologie (domaines et méthodes), J.P. Rossi (Ed.), Bordas, Paris.
- TULLIS T.S. (1984).** *A Computer based tool for evaluating alphanumeric displays.* In Human-Computer Interaction : INTERACT'84, B. Shackel (Ed.), Amsterdam, North-Holland, 719-723.
- TULLIS T.S. (1988).** *A system for evaluating screen formats : research and application.* In "Advances in Human-Computer Interaction", H.R. Harston et D. Hix (Eds.), volume 2, Ablex Norwood, N.J., pp. 214-286.
- SOMMERVILLE I. (1994).** *Le génie logiciel.* 4ème édition, Addison-Wesley.
- STREVELER D.J., WASSERMAN A.I. (1984).** *Quantitative measures of the spatial properties of screen design.* In Human Computer Interaction : INTERACT '84, B. Shackel (Eds.), Amsterdam, North-Holland, pp. 81-89.
- WILSON J.R., CORLETT E.N. (Eds.) (1990).** *Evaluation of human works : a practical ergonomics methodology.* Taylor & Francis.