Chapitre 3

Techniques, méthodes et démarches d'évaluation des aides techniques

> « Professionals in ergonomics are experts at designing technologies to best correspond with human anatomy and physiology. The usability of technologies without fatigue or pain, the selection of the best components and materials, have greatly contributed to overall consumer satisfaction with today's technical products. Now ergonomics professionals must also become skilled in assessing user preferences and predispositions to the use of particular technologies. »
>
> [SCH 97]

3.1. Introduction

Un des changements majeurs intervenu au cours de ces dernières décennies porte sur la place que l'on accorde aux utilisateurs dans la conception des systèmes techniques. Pendant longtemps, la technicité des instruments a dominé sur l'activité des opérateurs et sur les préoccupations qui pouvaient dépendre de l'organisation de leur activité, qu'elle soit professionnelle ou domestique. L'utilisateur, *a priori* doté de capacités d'adaptation, était alors considéré comme une variable permettant l'ajustement de l'instrument à la fonction visée par celui-ci. L'utilisateur s'ajustait aux règles de fonctionnement de l'instrument, mais aussi à l'environnement social et économique dans lequel était inséré le système technologique. Tout reposait sur les facultés d'apprentissage des opérateurs, sur la formation et l'acquisition de savoirs et de compétences pour se servir et servir les instruments. « Opérateurs adaptez-vous ! » était alors le mot d'ordre général.

Chapitre rédigé par E. BRANGIER, M.-E. BOBILLIER CHAUMON, J.-P. GAILLARD.

Mais cette conception a très vite trouvé ses limites. D'abord, parce que la complexité croissante des dispositifs technologiques nécessite à la fois de nouveaux modes de collaboration entre les hommes, les machines, les modes d'organisation et les cultures : ce qui était adapté ou ajustable par un individu dans un contexte donné, ne l'était pas forcément pour le même individu dans un autre contexte et a fortiori pour d'autres individus dans d'autres contextes. Ensuite, parce que ces technologies trouvent des applications et des usages dans des domaines toujours plus nombreux de la sphère personnelle et professionnelle. Les utilisateurs se diversifient (par leur profil, leurs compétences, leurs attentes…) et les usages se multiplient (domotique, système automobile embarqué, télémedecine, télétravail, assistant personnel, collecticiel…). Aussi, Aujourd'hui, l'ingénierie de conception se trouve dans l'obligation de prendre en compte cet utilisateur selon une approche symbiotique [BRA 00, BRA 02, BEN 95] qui considère que l'homme vit avec les technologies qui l'entourent et qu'il profite ainsi du travail constructif d'autrui. Avec la technologie, l'homme déplace dans la technologie une partie de son travail, mais aussi de ses souhaits et sources de satisfaction. Bien évidemment une telle approche souligne que les instruments doivent être adaptés aux capacités et aux besoins des personnes.

Un tel adage prend entièrement son sens dans le monde du handicap où se développe des technologies d'assistance qui ne font pas que de faciliter la vie des personnes, mais qui rendent possible certaines choses qui ne l'étaient plus, ou qui ne l'ont jamais été. Du coup, si la facilité d'utilisation devient bel et bien une des conditions de l'usage tout comme l'est également la satisfaction des personnes, ou la qualité de vie que leur apporte ou pas la technologie sensée les aider. Par voie de conséquence, il devient impératif d'évaluer les aides techniques et ainsi de mesurer leur capacité à satisfaire les attentes des personnes handicapées. Ce n'est d'ailleurs pas seulement dans le monde de l'assistance aux handicaps que la nécessité d'évaluer les dispositifs techniques s'est faite ressentir. En effet, elle connaît un essor important, en particulier pour :

– satisfaire des utilisateurs, toujours plus nombreux et variés ;

– optimiser les fonctionnalités des interactions homme-machine (IHM) et simplifier les usages de technologies toujours plus riches et complexes ;

– attirer les consommateurs car l'argument ergonomique est devenu un atout commercial dans la vente d'une technologie (site-web ergonomique, clavier ergonomique, logiciel ergonomique…) ;

– pallier les lacunes des techniques de conception informatique de type *quick and dirty* qui reposent sur un maquettage rapide de l'application via des générateurs d'interfaces et d'applications, mais sans implication réelle de l'utilisateur ;

– assurer l'évolutivité et la pérennité des logiciels et des IHM (Bobillier Chaumon, 1998) ;

– enfin, pour répondre aux besoins d'utilisateurs très particuliers, comme les personnes handicapées, pour lesquels les systèmes techniques conçus prennent souvent la forme de produits sur mesure requérant une évaluation et une adaptation spécifiques.

Dans ce dernier registre, le domaine des aides techniques (AT) pour personnes handicapées représente par nécessité le meilleur des paradigmes. Par définition, la notion de handicap signifie qu'une ou plusieurs fonctions étant défaillantes, les capacités d'adaptation de la personne à son environnement est réduite si ce n'est impossible. L'environnement, social et physique de la personne doit donc lui être adapté de tel sorte que la fonction défaillante puisse être restaurée. A cet égard, l'AT est un instrument ayant comme fonction de servir d'interface entre la personne handicapée et son environnement. Cet AT peut-être un instrument développé et couramment utilisé par des personnes valides, mais dans ce cas elle apparaît souvent comme un élément non adapté de l'environnement physique de la personne handicapée. Elle peut aussi être spécifiquement développée pour répondre au besoins de la personne handicapée, dans ce cas, la spécificité de chaque type de handicap, la spécificité du handicap, en degré et nature, de chaque personne impose naturellement une réflexion sur l'adaptation de l'AT à la personne. Le développement des AT s'inscrit donc tout naturellement dans ce type de conception centrée sur la personne.

En somme, les AT sont une catégorie d'instruments susceptibles de suppléer une fonction naturelle et destinée à réhabiliter la personne. Dans ce chapitre nous présenterons la méthodologie et les différentes techniques permettant d'évaluer leur degré d'adaptation aux capacités des personnes. Cette méthodologie vise plusieurs objectifs d'évaluation :

– des capacités de la personne en situation d'utilisation d'une AT,
– de l'efficacité d'une AT,
– de l'efficience d'une AT,
– de la satisfaction de la personne lors de l'utilisation d'une AT,
– de l'apprenabilité d'une AT,
– de la tolérance aux erreurs d'utilisation.

Efficacité, efficience, satisfaction, apprenabibilité, et tolérance aux erreurs sont cinq critères définissant « l'utilisabilité » d'un instrument, d'un système technique, organisationnel, social ou économique. Définissons ces notions :

La performance d'un système permet d'en estimer son efficacité. Elle est déterminée d'une part par les caractéristiques techniques de l'AT, par les capacités physiques, sensorielles et cognitives de la personne et par la qualité de l'interface entre ces deux composantes du système personne/AT.

L'efficience correspond aux « coûts » générés par l'utilisation de l'AT Ceux-ci sont d'ordre financier, énergétique et pour notre propos, physiologiques et psychologiques en termes de charge de travail, d'apprentissage, de stress.

La satisfaction mesure l'adéquation entre les objectifs fonctionnels de l'AT et les besoins de la personne.

L'apprenabilité détermine la faculté d'apprentissage naturelle et intuitive d'un produit.

La tolérance aux erreurs indique que le système accepte l'erreur, qu'il l'anticipe même et qu'il donne les moyens à l'utilisateur de la comprendre et de la corriger.

Ces notions s'inscrivent donc dans une évolution de l'ergonomie et de son objet d'étude : le passage d'une ergonomie de correction, puis de conception visant à l'amélioration des conditions de travail, à son extension aux conditions de vie et d'utilisation des produits et des services.

D'un point de vue général, l'évaluation d'une AT consiste à porter un jugement sur la valeur de son interface, c'est-à-dire à éprouver sa capacité à satisfaire les besoins de ses utilisateurs (critère de satisfaction), à mesurer sa compatibilité avec l'activité réelle de la personne et leurs cognitions (critère d'apprenabilité et de tolérance aux erreurs) ; en d'autres termes, à garantir un bon niveau de performance humaine et technique (critère d'efficacité). L'évaluation sert également avec faire évoluer l'AT en mettant en évidence ses points faibles et en justifiant les améliorations nécessaires. Mais cette situation d'évaluation n'est jamais neutre : elle dépend de plusieurs facteurs parmi lesquelles on trouve : la subjectivité de l'évaluateur, les influences que peuvent exercer les évalués ou encore les démarches d'évaluation utilisées. Par conséquent, l'évaluation des formes de coopération entre les personnes et les systèmes techniques devient un enjeu considérable pour non seulement satisfaire l'utilisateur ou le client, mais aussi pour garantir la réussite économique des aides techniques.

Dans cette perspective nous présenterons un cadre de compréhension des divers aspects de l'évaluation ergonomique des AT, en soulignant que les démarches utilisées ont un effet direct sur les résultats de l'évaluation réalisée. Nous chercherons ainsi à mettre en évidence les relations qui existent entre les méthodes d'évaluation, le contexte de l'évaluation et les résultats produits. Dans une première partie, nous présenterons les différentes techniques de recueil de données utilisables lors d'évaluation d'AT pour souligner leurs intérêts et leurs limites. La deuxième partie sera consacrée aux méthodes d'évaluation, tandis que la troisième partie abordera les démarches plus intégrées d'évaluation. Il s'agira de montrer que l'efficacité et l'adéquation d'une méthode d'évaluation dépendent de l'objectif qui lui est assigné et de son contexte de mise en œuvre pour, proposer finalement une

grille permettant de sélectionner les méthodes d'évaluation les plus appropriées selon les critères de l'intervention et les finalités de l'évaluation.

3.2. Les techniques de recueil de données pour l'évaluation

Les techniques de recueil de données visent à collecter des informations sur l'AT utilisée ou utilisable dans un contexte particulier. Il s'agit principalement d'une prise d'informations sur le contenu de son interface, mais aussi sur les caractéristiques de l'utilisateur de sa tâche et du contexte organisationnelle qui composent la situation de travail. Ces techniques peuvent se classer en deux grandes catégories : les techniques d'enquête subjectives et les techniques d'enquête objectives [Bainbridge 1990, BAR 95, GRO 90, OPP 97, PER 96, PER 97, PER 98, SCA 93, SEN 93, VAL 93].

3.2.1. *Les techniques subjectives*

Elles concernent toutes les techniques où l'on demande à l'utilisateur de s'exprimer, de donner son avis sur la qualité ergonomique de l'AT et de son interface, et plus particulièrement sur l'adéquation du dispositif à ses caractéristiques intrinsèques (linguistiques, perceptives, opératoires et cognitives) et aux spécificités de son activité. Les techniques subjectives recouvrent les questionnaires, les entretiens et les échelles.

3.2.1.1. *Les questionnaires*

Les questionnaires ont pour objectif de recueillir des données subjectives concernant un grand nombre d'utilisateurs sur la base de questions ouvertes et fermées. Ces dernières peuvent porter sur la modification d'attitude suite à l'utilisation de l'interface, sur la mesure de la satisfaction, sur les besoins et les attentes des utilisateurs… Ces questionnaires permettent également de contrôler la qualité de l'interface (correspondance entre attentes des utilisateurs et possibilités offertes par le logiciel) et de valider les résultats de l'expérimentation [GRO 90, OPP 97, SEN 93]. Généralement, les questionnaires portent sur différents thèmes de l'évaluation :

– évaluation subjective du niveau de charge mentale générée par un dispositif informatique dans le cadre d'une activité professionnelle (Lasfargue, 1999)[1] ;

– évaluation subjective de la qualité ergonomique des interfaces par une grille spécifique [KIR 98, Nielson 1995[2]] ;

1. http://www.ergostressie.com/.
2. http://www.useit.com/papers/heuristic/learning_inspection.html.

– identification des attentes, des besoins et des préférences des utilisateurs sur telle ou telle fonction, information ou structuration de l'interface d'un logiciel ou d'un site internet [BAS 93, LEU 99] ;

– définition des conduites d'interaction et les habitudes de consultation des utilisateurs (fréquence, durée, cible…) selon leur profil (âge, formation, caractéristiques biographiques, loisirs…).

Toutefois, le questionnaire peut poser quelques problèmes relatifs à son étalonnage sur la population cible, surtout lorsqu'il s'agit d'handicaps d'une très grande variabilité. En effet, la validité d'un questionnaire est généralement restreinte à la situation pour laquelle ils ont été mis au point. Du coup, leur transférabilité dans une situation nouvelle peut être très faible, et occultée par exemple une partie des problèmes rencontrés par l'utilisateur. Qui plus est, les questionnaires reposent sur une dissociation de la situation d'usage de la situation d'évaluation. Ce clivage entre le temps des réponses et le temps de l'interaction avec la machine, engendre souvent une altération des réponses (difficultés de remémoration, oubli, occultation, lacune…) de la part de l'utilisateur [SEN 93]. Un risque d'effet de halo existe également [GRO 90] : l'utilisateur a tendance à plus s'appuyer sur son impression générale à l'égard du logiciel que sur les items du questionnaire. Enfin plus largement, le questionnaire, s'il illustre quelques difficultés renseigne en fait très peu sur ce qu'il faut faire pour améliorer un logiciel en ne sachant que ce qui ne va pas. En bref, le questionnaire souligne les points faibles ou forts d'une IHM sans pour autant indiquer les améliorations possibles.

3.2.1.2. *Les entretiens*

Les entretiens sont très utiles pour l'exploration plus ciblée de certains aspects ou de certaines dimensions de l'interaction. A contrario, ils nécessitent plus de temps que le questionnaire [OPP 97]. Les entretiens sont généralement menés de manière semi-directive et ont pour but d'obtenir des informations générales ou ciblées sur la nature, le niveau et la qualité de l'interaction homme-machine. Ils peuvent être réalisés en dehors du cadre de l'interaction. C'est alors une situation de recueil où les utilisateurs fournissent une certaine représentation de la coopération qu'ils pensent ou projettent d'effectuer avec le dispositif. Les évaluations sont alors délivrées en dehors de toute situation d'interaction effective avec l'interface. Ils permettent pourtant d'apprécier rapidement le niveau de satisfaction et d'acceptation général du dispositif. Ils fournissent enfin des pistes d'investigation à explorer dans l'entretien en situation d'interaction. Pour ces derniers, les thématiques abordés peuvent couvrir plusieurs domaines :

– l'avis de l'utilisateur sur le contenu, la forme, l'utilité et l'utilisabilité de l'interface (ce qu'ils aiment ou non dans le dispositif, ce qu'ils ont apprécié ou détesté, ce qui les ont amusé ou irrités) ;

– leurs besoins en termes de données, de fonctions, d'assistance ou d'information particuliers sur des points bien précis de l'interface ;

– le compte rendu, en temps réel, de problèmes d'utilisation ou de compréhension de l'interface liés par exemple, à un manque d'information ou à des erreurs de conception et requérant des corrections ou des aménagements.

Le principal atout de l'entretien par rapport au questionnaire est que l'utilisateur a la possibilité de développer ses réponses en les illustrant, si besoin est, par des exemples pratiques. Cela dit, d'autres techniques de recueil de données permettent de sérier plus « objectivement » la complexité de la situation d'interaction dans des conditions d'utilisation du logiciel spécifiques. C'est en particulier le cas de l'observation et des techniques de verbalisation qui appartiennent au registre des techniques objectives et que nous présenteront plus loin.

3.2.1.3. *Les échelles subjectives*

Des échelles subjectives ont été introduites dans le cadre de l'évaluation de la charge de travail mental. Elles reposent sur le postulat selon lequel les experts d'un domaine ont une connaissance suffisamment précise et stable de la difficulté inhérente à l'exécution de chacune des tâches du domaine. Il est donc théoriquement possible de leur demander d'évaluer cette difficulté sur une échelle, de type échelle de Lickert, dont chacun des échelons aura été précédemment défini. La graduation sur une échelle permettra de procéder à des comparaisons entre des situations différentes. Modifications matérielles, procédurales, environnementales... Chaque personne construisant à priori sa propre échelle subjective, les comparaisons ne devraient donc pas porter entre des personnes différentes. Il ne s'agit donc pas en principe d'échelles dites à intervalles. En conséquence, pour rendre compte des comparaisons, on écartera les traitements statistiques paramétriques comparant des distributions entre elles avec moyennes et écarts types. On aura recours aux techniques d'analyse non paramétriques. Il existe cependant une échelle (échelle SWAT) dont les promoteurs, s'appuyant sur des techniques utilisées pour la construction de tests psychométriques, assurent que les intervalles sont en principe égaux.

Les trois principales échelles utilisées sont issues de l'aéronautique. Elles ont été construites pour répondre aux besoins d'évaluation de la charge mentale des pilotes ou des astronautes. Leur utilisation dans d'autres domaines reste hypothétique et devrait en toute rigueur faire l'objet d'études comparatives de leurs résultats avec d'autres méthodes d'évaluation de la charge de travail mental.

L'échelle de Cooper-Harper (1969) fait référence à la manœuvrabilité des avions. Elle est estimée par le pilote au moyen de descripteurs correspondant à un arbre de décision comportant trois niveaux : manœuvre impossible, difficile et facile. Chaque niveau de l'arbre se subdivise en trois autres niveaux. Chacun de ces niveaux

représente un descripteur répondant à des critères de difficultés rencontrées par le pilote, de l'effort mental et du nombre d'erreurs commises. Le niveau 1 représente une tâche ayant été réalisée très facilement le niveau 10 une tâche impossible à réaliser. Cette échelle a été adaptée avec succès à d'autres contextes (Wierville & Casali, 1983).

L'échelle SWAT (*Subjective Workload Assessment Technic*) repose sur une technique de construction psychométrique. Les auteurs de cette échelle (Reid & Nygren, 1988) postulent que la charge de travail mental comporte trois facteurs : temporel, effort mental et astreinte psychologique (stress). L'échelle postule un modèle additif entre ces trois facteurs. Chacun de ces facteurs comporte trois niveaux : faible, moyen, fort. La charge de travail mental est ainsi représentée par un espace de dimension 3 (les 3 facteurs) avec chacun 3 niveaux, soit 27 cellules. La procédure d'application de l'échelle consiste dans un premier temps à demander à chaque opérateur d'ordonner les 27 cellules en référence à des situations préexistantes au domaine et parfaitement connues de lui. Une fois ce travail préalable terminé, on considère que l'opérateur a construit son échelle de charge de travail mental. Les écarts entre chacun des intervalles de l'échelle sont considérés comme égaux. Ultérieurement, l'opérateur utilisera cette échelle pour estimer la charge de travail mental générée par une tâche ou une nouvelle situation. Cette technique a été validé par ses promoteurs sur des tâches représentatives de grandes fonctions cognitives et motrices : mémoire de travail, perception visuelle, traitement de l'information symbolique, arithmétique, habileté manuelle, ainsi que lors de simulations de tâches de pilotage d'avions. Les résultats obtenus avec l'échelle SWAT sont corrélés avec des indicateurs physiologiques de charge de travail, des mesures de charge obtenus par des techniques de concurrence cognitive et a démontré une bonne validité prédictive dans le domaine du pilotage. Cette échelle a été validée en France, toujours dans le domaine de l'aéronautique par Amalberti *et al*. (1986) qui estiment ses résultats pertinents et significatifs.

L'échelle TLX (*Task Load Index*) est représentative des techniques mettant en œuvre des échelles bipolaires pondérés. Il s'agit de demander à l'opérateur d'évaluer subjectivement l'importance d'un facteur de charge de travail en lui donnant une note comprise entre deux extrêmes. La note obtenue est ensuite pondérée en fonction de son importance relative par rapport aux autres facteurs concourant à la charge de travail. TLX comprend six descripteurs considérées comme des facteurs de charge de travail : la pression temporelle, l'activité mentale, l'activité physique, la performance, la fatigue, le stress. Chaque descripteur est précisément défini de sorte que l'opérateur puisse en évaluer l'importance dans la tâche. A l'issue de celle-ci, on lui demande de cocher sur une feuille de papier des segments de droite représentant chacun un des six descripteurs. Par exemple, le segment pression temporelle avec l'origine du segment signifiant aucune pression temporelle à une extrémité et signifiant extrêmement élevée à son autre extrémité. Dans un second temps le

psychologue présente à l'opérateur les descripteurs deux par deux et lui demande de choisir pour chacun des couples le descripteur le plus important. Les valeurs estimées sur les descripteurs sont ensuite mesurées et pondérées par le nombre de fois ou l'opérateur aura retenu le facteur comme plus important dans les couples. TLX a été validé en France dans le domaine de l'aéronautique, pilotage d'avions et contrôle de la navigation aérienne. Enfin, il convient de noter que l'échelle TLX évalue la charge de travail globale comprenant la charge de travail physique et la charge de travail mental.

3.2.2. *Les techniques objectives*

Elles regroupent toutes les techniques d'observation dont le but est de recueillir des faits, des conduites et des verbalisations, individuels ou collectifs, pouvant être qualifiés, décrits et quantifiés. C'est à partir de ces données qu'il est possible de mesurer et d'analyser le niveau de compatibilité des dispositifs techniques avec les spécificités de l'utilisateur et les caractéristiques de son activité. Parmi les différentes démarches, on distinguera les techniques d'observation directes des indirectes. Les premières permettent de suivre *in situ* le déroulement de l'observation, sur une unique période. Les techniques indirectes donnent la possibilité d'enregistrer des microconduites de l'interaction ou d'évaluer des états de l'opérateur au moyen de critères physiologiques.

3.2.2.1. *Les techniques d'observation directes*

Ces techniques nécessitent que l'utilisateur réalise un travail réel. Elles se focalisent sur la fréquence et la nature des incidents rencontrés (erreurs, problèmes, hésitations), sur la stratégie des utilisateurs (comparaison et justification de l'écart entre le prescrit et le réel), sur le temps d'apprentissage, la rapidité de performance à une tâche et le facteur de rétention dans le temps [SEN 93, VAL 93, VAN 94]. Parmi ces techniques, on retiendra :

– l'observation directe : un expert observe une situation réelle et collecte toutes les informations relatives à l'activité réelle de l'utilisateur et à son usage de l'interface : difficultés rencontrées, informations recherchées, utilisées ou ignorées, nature et déroulement des séquences opératoires, nature et fréquence des interactions homme-machine et homme-homme, outils et documents utilisés, types d'erreurs commises et les raisons, etc. Les aspects retenus sont ceux qui apparaissent, sinon comme des facteurs explicatifs, du moins comme permettant de donner un sens aux conduites d'interaction de l'utilisateur. Le but est d'arriver à une définition et à une compréhension au plus juste de l'activité réelle de l'opérateur avec le dispositif technique (Mazoyer & Salembier, 1987) ;

– les verbalisations simultanées à l'utilisation de l'interface : elles sont des indicateurs de l'activité mentale en temps réel. De manière pratique, il s'agit pour l'utilisateur de penser à « haute voix » durant son interaction avec le dispositif. La verbalisation est en effet une situation où le contenu de la mémoire de travail de

l'utilisateur s'exprime assez clairement. Et c'est à partir de ces éléments que des inférences peuvent être faites sur les représentations mentales et les procédures mobilisées par l'usager durant sa collaboration avec le dispositif (Caverni & Bastien, 1988). Concernant ces techniques de verbalisation, de nombreuses critiques ont été émises sur l'utilisation de la verbalisation comme moyen d'identifier les processus cognitifs en œuvre durant l'interaction homme-machine. La verbalisation suppose d'abord un minimum de capacité d'abstraction et de formalisation des expériences, des sensations éprouvées…, et suppose surtout la possibilité de « mettre à mots » ; c'est-à-dire d'avoir à sa disposition « les mots pour le dire » (Teiger, 1993). Ce qui veut dire que des déficiences verbales ou cognitives rendent impossible l'explicitation des conduites mentales, et de ce fait ces dernières seraient inaccessibles à soi comme aux autres, car non verbalisées. Enfin, d'un point de vue méthodologique, on critique le fait de demander à l'individu de se concentrer sur une tâche de verbalisation, alors que, dans le même temps, toute son attention est retenue dans l'exécution d'une autre activité. On crée ainsi les conditions d'une surcharge de travail par ajout d'une tâche supplémentaire à l'activité principale. Cette surimpression de tâches fait qu'on ne disposerait que d'une couche très superficielle des processus mis en œuvre. En dépit de toutes ces critiques, ces techniques restent les meilleures pour rendre compte ou tout au moins, pour s'approcher des mécanismes cognitifs déployés durant l'interaction. En cela, nous rejoignons le point de vue théorique sur le cours d'action formulé par Pinsky vu comme « l'activité d'un acteur déterminé, engagé activement dans un environnement physique et social déterminé et appartenant à une culture déterminée, activité qui est significative pour ce dernier, c'est-à-dire montrable, racontable et commentable par lui à tout instant de son déroulement à un observateur-interlocuteur » (Cité par Theureau & Jeffroy, 1994, p. 19) ;

– l'évaluation coopérative croisée : les utilisateurs effectuent des évaluations concomitantes sur l'interface, durant ou après son utilisation. Cette confrontation de point de vue crée une sorte d'émulation critique bénéfique au diagnostic [DUM 93]. Toutefois, un biais peut se révéler lorsque la personnalité de certains évaluateurs de l'échantillon inhibe les réactions des autres, voire qu'oriente les thèmes du discours ;

– techniques des incidents critiques : à partir d'entretiens et d'observations (en situation de travail), cette technique cherche à identifier la nature et l'origine des problèmes et des difficultés que rencontrent l'utilisateur au cours son interaction avec le dispositif. On obtient ainsi une vue générale des dysfonctionnements et des anomalies du dispositif imputable soit au système, soit à l'utilisateur ou au deux. Aussi, pour plus de rigueur dans le diagnostic, on préconise de recourir à des questionnaires ou à des mouchards électroniques.

3.2.2.2. *Les techniques d'observation indirectes*

L'objectif de ces techniques est de recueillir, parfois à l'insu de l'utilisateur, des données sur l'utilisation du dispositif. Toutefois, par déontologie, les utilisateurs sont, bien évidemment, informés que leurs interactions pourront être enregistrées.

Les enregistrements vidéos permettent de revenir sur certains points précis qui sont impossibles à relever lors d'une observation directe. Ils peuvent aussi servir de support pédagogique pour la communication avec les concepteurs afin d'illustrer, par exemple, les implications regrettables de certains choix de développement. Ces enregistrements sont aussi employés comme support d'expression puisqu'on demande aux utilisateurs filmés de réagir et commenter ce qu'ils font à l'écran : pourquoi ont-ils choisi telles fonctions ? pourquoi ces soupirs ou cet étonnement ? Ces verbalisations consécutives peuvent compléter les verbalisations simultanées.

Le contrôle de l'interaction par des mouchards électroniques (*Log File*) consiste à enregistrer les interactions à l'écran et au clavier pour réaliser ensuite des analyses statistiques [APE 85]. En obtenant les fréquences et l'ordre des différentes séquences d'actions, il est par exemple possible d'évaluer l'évolution de l'apprentissage des opérateurs (Brangier & Pino, 2000). Mais ces mouchards ou traceurs sont de plus en plus utilisés pour déterminer le taux d'audience des sites interfaces Internet [DUB 00]. Dans ces cas, des chiffres de consultations médiocres sont souvent interprétés comme symptomatiques d'une mauvaise ergonomie de l'interface ou d'une incapacité de l'application à satisfaire les besoins de l'utilisateur.

Les indicateurs physiologiques peuvent également être utilisés. Pour évaluer la charge de travail physique l'enregistrement du rythme cardiaque reste l'indicateur le plus accessible. Consommation d'oxygène et pression artérielle lui sont corrélées. L'arythmie sinusale, variation de l'intervalle entre chaque pulsation cardiaque, est très difficilement interprétable. Elle semble plutôt concernée la tension émotionnelle. Les mesures de débit sanguin cérébral, les critères endocriniens tels que les mesures de taux de stéroïdes et de catécholamines (adrénaline et noradrénalines) sont réservés aux études physiologiques du rythme circadien et de l'astreinte physique imposée par l'activité de travail. Ces taux varient aussi en fonction de la tension émotionnelle et non en fonction de la charge de travail mental. Enfin les critères neurophysiologiques, électroencéphalogrammes, potentiels évoqués, sont surtout de bons indicateurs du niveau de vigilance et de ses variations au cours de longues périodes de travail. En revanche, l'évaluation de la charge de travail mental paraît peu accessible à ce type d'indicateurs.

La double tâche est essentiellement utilisée pour estimer la charge de travail mental des opérateurs, en plaçant l'utilisateur en situation de concurrence cognitive. Celui-ci reçoit pour consigne de réaliser deux tâches en même temps. La première, dont on cherche à estimer la valeur de la charge de travail mental qu'elle induit, est explicitement considérée comme étant la tâche principale. La seconde a pour fonction de saturer les capacités cognitives de l'opérateur. Partant des performances mesurées à la tâche secondaire, on en infère la charge de travail mental requise pour exécuter la tâche principale. Si la tâche principale mobilise peu de ressources cognitives, la performance à la tâche secondaire sera élevée. En cas contraire, la

performance à la tâche secondaire sera faible. Cette technique permet en théorie de comparer entre elles différentes tâches et différentes aides techniques pour les exécuter. De nombreuses critiques ont été émises à son encontre. Rien n'assure que l'opérateur n'exécute prioritairement la tâche principale. Dans les situations de concurrence cognitives l'opérateur peut être amené à développer de nouvelles stratégies d'exécution de la tâche principale, cognitivement plus économiques. L'introduction d'une tâche secondaire, en modifiant les conditions d'exécution de la tâche principale, modifierait la stratégie d'exécution de la tâche principale.

3.3. Les méthodes d'évaluation des aides techniques

La taxonomie choisie pour classer ces méthodes reprend et développe celle employée par Scapin [SCA 93] et Senach [SEN 93] dans le domaine de l'évaluation des logiciels. Elle répond à des critères d'exhaustivité (toutes les méthodes d'évaluation peuvent y être classées) et de discrétion (une méthode ne peut faire partie que d'une catégorie à la fois). On distinguera ainsi :

Les méthodes prédictives : il s'agit de concevoir la relation entre la personne et l'AT comme une coopération entre elle et un système technique. Cette coopération nécessite un dialogue entre la personne handicapée et l'AT future probable qui est encore au stade du développement, et de prévoir sa compatibilité avec les futurs usagers et leur activité.

Les méthodes analytiques : il s'agit d'inspecter un aspect (graphique, sémantique, structurel…) ou une dimension (perceptive, linguistique, logique…) de l'interface et de l'interaction entre la personne et l'AT (P/AT) dans des conditions d'utilisation particulières, qui peuvent être artificielles. Elles peuvent aussi s'appuyer sur différents guides (modèles, recommandations, critères) pour déterminer « le caractère approprié ou non de l'interface du point de vue ergonomique » [SCA 93, p.19]. Les méthodes prédictives et analytiques sont constructives : on cherche à évaluer les difficultés rencontrées par l'utilisateur dans le but d'y remédier. Ce type d'évaluation s'inscrit dans un cycle « constat des difficultés/impossibilités/échec dans l'exécution d'une tâche/remédiation ». La production d'erreurs et l'échec sont réhabilités comme sources d'informations utiles pour développer la compétence d'un système technique.

Les méthodes empiriques : il s'agit d'évaluations contextualisées, c'est-à-dire effectuées en situation réelle de coopération P/AT Elles « ont pour objectif de définir, selon diverses dimensions comportementales, la qualité ergonomique d'une interface telle qu'elle peut être établie à partir de son utilisation » [SCA 93, p.18]. Elles correspondent à un dernier type d'évaluation de type pronostique, en rapport avec la seconde boucle itérative évoquée précédemment. Elles sont donc mises en œuvre lors du cycle de la validation du prototype.

3.3.1. *Les méthodes prédictives*

Nous distinguerons les méthodes qui s'emploient à spécifier et à évaluer la structuration du dialogue personne/AT et celles qui s'intéressent aux performances attendues de l'utilisateur final.

3.3.1.1. *Les modèles de définition des interactions opérateur-machine*

Parmi l'arsenal méthodologique disponible pour évaluer l'articulation du dialogue opérateur-machine, certaines démarches permettent de prédire le niveau d'adéquation entre la logique d'action de l'utilisateur et la logique de fonctionnement de l'AT (lorsque ce dernier comporte un logiciel de commande et de contrôle).

3.3.1.1.1. Diagramme à transition d'états

Un état est une partie du programme informatique et du travail de l'utilisateur qui requiert une saisie, une désignation ou une sélection de la part de l'opérateur. Ces états sont reliés entre eux par les actions de l'opérateur.

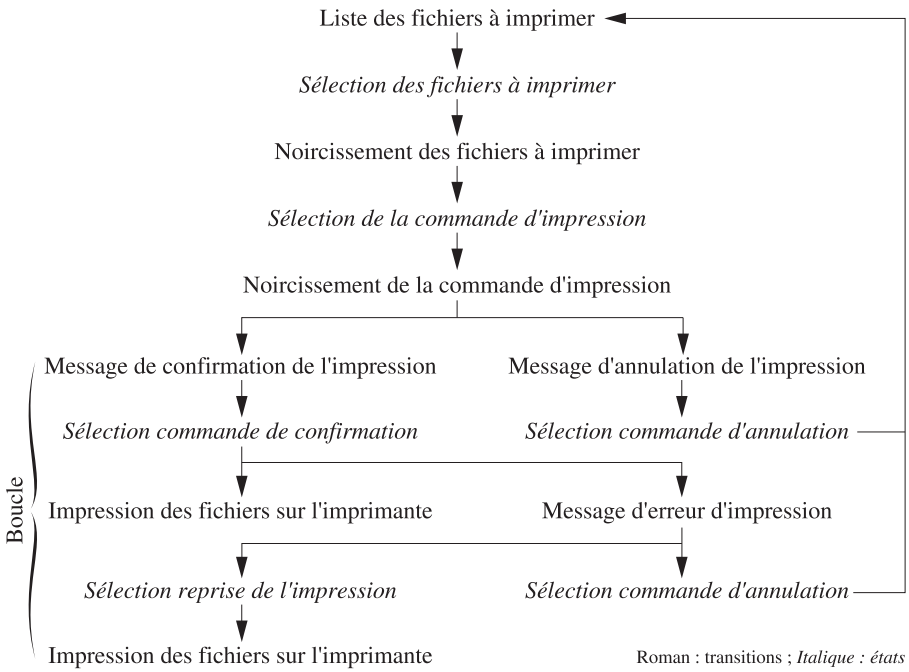


Figure 3.1. *Diagramme à transition d'état*
« l'exemple d'un dialogue d'impression de fichiers »

L'objectif d'un diagramme à transition d'état est de mettre en évidence les états équivoques, les incompatibilités, les boucles, les chemins qui ne mènent à rien dans le but de pallier les insuffisances de l'interaction (figure 3.1). Ce diagramme sert à définir la dynamique du dialogue opérateur-logiciel. Il décrit les règles de passage d'un état à un autre, et définit le nombre d'états et de sous-états possibles, selon la complexité de la tâche ou les types d'utilisateurs (novices, expérimentés). Chaque état initie des sous-états qui autorisent des actions. L'ensemble des actions possibles définit d'une part les états autorisés ou interdit à l'utilisateur ; et d'autre part les actions auxquelles elles mènent et les actions irréalisables. En résumé, le diagramme à transition d'état définit la cinématique des interactions. A titre d'exemple, cette figure correspond à un diagramme à transition d'état d'un dialogue d'impression d'un fichier.

3.3.1.1.2. Graphe des commandes

Ce graphe vise à définir l'arborescence, la hiérarchisation des commandes et la compatibilité de cette hiérarchisation avec la réalisation des tâches réelles. Il constitue un moyen de mettre en évidence la disponibilité des commandes les unes par rapport aux autres. Une commande est déclarée disponible si, compte tenu de l'état où se trouve l'opérateur, il peut l'utiliser. Le graphe des commandes indique également les enchevêtrements, les ambiguïtés ou les itérations de commandes qui risquent de ne pas satisfaire les utilisateurs dans la réalisation de leur tâche. De plus, il montre avec force que l'exécution d'une commande implique une modification de la signification de la dite commande. De fait, ce graphe met en évidence les différences de statut entre les états initiaux et les états finaux associés aux interactions (figure 3.2).

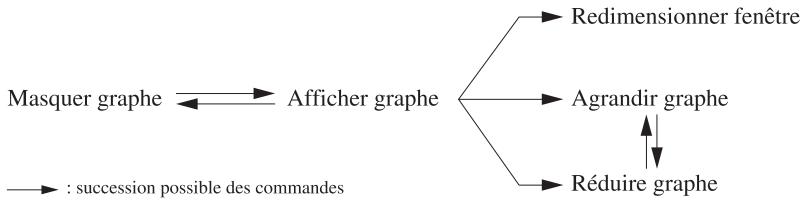


Figure 3.2. *Graphe des commandes de visualisation d'un graphique dans un logiciel de contrôle d'un processus informatique (une commande n'est disponible qu'à la condition où l'utilisateur se trouve dans l'état précédent la commande)*

Bien évidemment, les graphes de commandes peuvent être complétés par l'adjonction des états de l'interface associés aux commandes, pour former ainsi des graphes états-commandes.

3.3.1.1.3. Le Diagramme de fiabilité

Il correspond à une représentation des états du logiciel : le concepteur est alors chargé de spécifier les procédures attendues et inattendues de la part de l'opérateur tout comme les actions attendues et inattendues du logiciel (figure 3.3). Son but est à la fois de diminuer, voire de rendre impossible les cheminements non désirés ; de concevoir et valider les messages d'erreurs ; de définir les comportements du logiciel lorsque l'opérateur l'engage dans les actions inattendues, et donc bien souvent non prévues lors de la conception.

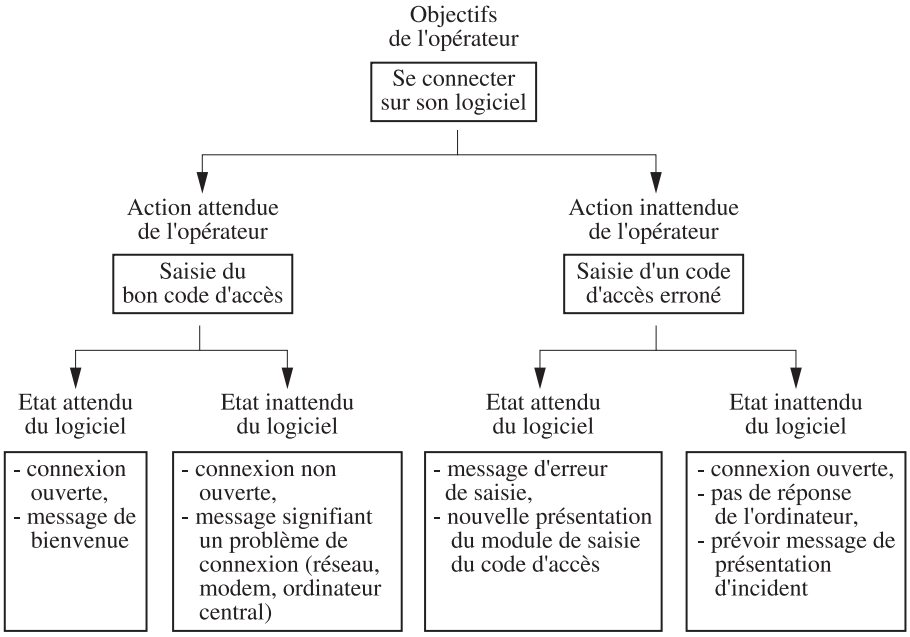


Figure 3.3. *Diagramme de fiabilité d'une tâche de connexion d'un terminal sur un site central*

Cette méthode a le gros avantage de faire émerger les comportements inattendus de l'opérateur, et les réponses inattendues du système. Par contre, les réponses inattendues du système sont souvent bien pauvres en contenu, tant il est difficile de prévoir la non-fiabilité technique. De fait, celle-ci se résume la plupart du temps à des faits possibles du type : « serveur ne répond pas », « communication réseau interrompue », et s'adresse en conséquence plus à la maintenance du système qu'à la prévention des erreurs de l'utilisateur.

En définitive, bien que ces trois graphes aient été conçues à l'origine comme des méthodes de spécification du dialogue opérateur-machine, ils peuvent aussi diagnostiquer et surtout prévoir les difficultés futures de l'interaction personne-AT De fait, ces méthodes visent en quelque sorte à anticiper l'erreur d'utilisation pour la prévenir dans le processus de conception.

3.3.1.2. *Les modèles prédictifs des performances de l'utilisateur*

Ces modèles ont pour vocation de modéliser de façon générale la connaissance de la conception [INN 91, VAN 91] et de prédire les performances d'utilisation (durées d'exécution, occurrence des erreurs) sans avoir de détails précis sur l'utilisation. Ils se découpent en trois modèles complémentaires : modèle d'analyse des tâches, modèle linguistique, modèle de complexité cognitive de l'interaction.

3.3.1.2.1. Les modèles d'analyse des tâches

Les modèles d'analyse des tâches ont été développés sur la base d'analyse de tâches interactives élémentaires concernant des éditeurs de texte. Il s'agit d'abord de segmenter des tâches complexes en tâches élémentaires. Puis de les évaluer quantitativement selon plusieurs critères (nombre et vitesse de frappe pour effectuer une tâche, temps de réponse du dispositif et de préparation mentale pour agir) pour obtenir finalement, par agrégation de ces résultats, une prédiction de la performance de l'utilisateur [COU 90, SEN 93]. Deux modèles sont représentatifs de ce courant méthodologique : il s'agit de KLM (*Keystroke Level Model*) et GOMS (*Goals*, *Operator*, *Methods*, *Selection rules*) :

– le KLM permet de prédire le temps de réalisation d'une tâche habituelle d'un opérateur expérimenté qui ne fait pas d'erreur et dont la préparation de l'action est constante. Même si les prédictions obtenues sont assez précises, un problème se pose. En effet, rares sont les utilisateurs qui, en situation réelle, ne commettent aucune erreur et qui préparent leurs actions sur une durée fixe ;

– le modèle GOMS consiste à formaliser les stratégies d'action des utilisateurs pour la réalisation de leurs tâches. Celles-ci sont représentées sous la forme d'un arbre hiérarchique descendant composé de buts et de sous-buts. Plusieurs critiques posent cependant le problème de son efficacité : d'abord, elles visent son aspect extrêmement réducteur, en particulier en ce qui concerne la prise en compte des erreurs [COU 90, BRA 91, SCA 93, SEN 93]. Elles touchent aussi à la valeur des prédictions qui ne sont précises que pour des tâches macroscopiques. Enfin, après plusieurs années d'existence, il semble que son usage (du fait de son coût) se cantonne aux seules universités [Naël 96].

3.3.1.2.2. Les modèles linguistiques

Dans les modèles linguistiques, l'accent est mis sur une structure linguistique impliquée dans l'interaction et représentée sous la forme d'une grammaire. Il existe

deux modèles principaux l'ALG (*Action Language Grammar*) de Reisner [REI 81] et le CLG (*Command Language Grammar*) de Moran [MOR 81].

– l'ALG consiste à décomposer les buts de l'utilisateur, mis en parallèle avec les actions à mener selon les règles d'une grammaire d'interaction ;

– le CLG sert à décomposer un dispositif technique en plusieurs niveaux d'abstraction qui sont à leur tour décomposés. A partir d'une description des tâches, on obtient une description sémantique puis une description syntaxique pour aboutir finalement à une description de l'interaction. Cette démarche séquentielle et assez complexe en fait un outil davantage destiné aux informaticiens qu'aux ergonomes [SEN 93].

3.3.1.2.3. Modèle de la complexité cognitive de l'interaction

Un dernier type de modèle prédictif est celui de la complexité cognitive de l'interaction de Kieras et Polson (1985). Ce modèle repose sur le postulat selon lequel la complexité cognitive d'un dispositif est étroitement liée aux connaissances dont l'utilisateur dispose pour manipuler ce même dispositif (représentation de la tâche à réaliser et du dispositif). Malgré des apports indéniables (prédiction des risques cachés d'utilisation, anticipation des difficultés d'interaction, définition de critères et de normes de performance, …), certains auteurs contestent leur utilité. En particulier John et Marks (1997) qui affirment que la valeur prédictive de ces méthodes a été très largement surestimée. Pour le montrer, ils ont mis au point une expérience qui permettant d'évaluer la valeur prédictive[3] de différentes techniques : analyse des besoins des utilisateurs, cheminement cognitif (*cognitive walkthrough* sur lequel on reviendra ultérieurement), GOMS, évaluation heuristique, notation des actions de l'utilisateur et enfin, simple lecture des spécifications de l'interface. La méthodologie employée a consisté à demander à 6 sujets novices dans le domaine de l'évaluation des interfaces opérateurs/machine d'utiliser une des ces méthodes pour rechercher des problèmes d'utilisabilité. Les problèmes ainsi identifiés ont été transmis au concepteur qui décidait ou non d'apporter une modification à l'interface. Les résultats obtenus ont ensuite été comparés avec ceux obtenus grâce à des tests d'utilisation. Bien qu'il faille nuancer les conclusions (qui ne s'appliquent qu'à un échantillon limité de sujets inexpérimentés), il ressort trois points significatifs :

– les méthodes d'évaluation s'avèrent être moins efficaces qu'on ne le pense habituellement. En effet, seul un nombre de problèmes assez restreint aboutit à une réelle amélioration de l'interface. Qui plus est, les modifications apportées à certains problèmes prédits engendrent de nouvelles difficultés que les méthodes n'avaient bien sûr par prévues ;

– on ne peut pas vraiment établir de différences significatives entre les résultats obtenus : les méthodes ont donc toutes sensiblement la même valeur prédictive ;

3. Il ne s'agit pas seulement de trouver les problèmes de compatibilité mais aussi de voir dans quelle mesure la communication des résultats des investigations incite les concepteurs à effectuer des changements dans le design.

– ce qui implique que... la simple lecture des spécifications de l'interface aboutit au même résultat que l'utilisation de méthodes plus élaborées (et surtout plus coûteuses).

3.3.2. *Les méthodes analytiques*

Ces méthodes se centrent sur l'analyse de dimensions particulières des interactions et de l'adéquation de ses dimensions avec l'utilisateur et ses besoins.

3.3.2.1. *L'évaluation itérative*

La conception d'une aide technique ne correspond pas à une procédure figée, mais à un processus dynamique qui évolue au fur et à mesure de l'avancée du projet. L'évaluation itérative (figure 3.4) se propose d'évaluer le produit tout au long de sa conception : le produit est corrigé, si nécessaire, après chaque évaluation jusqu'à ce qu'il donne entière satisfaction. [MON 91, Michard 1993, Bobillier Chaumon 1999].

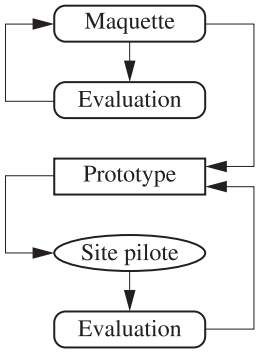


Figure 3.4. *L'évaluation itérative*

Concrètement, ces évaluations itératives cherchent à faire émerger les dysfonctionnements par des mises en situation d'utilisation. Les simulations peuvent être réalisées sur maquette[4], prototype[5] et recette (correspond au produit fini) ou site pilote. Les apports à chacune des deux boucles itératives de l'évaluation sont différents : une maquette renseigne sur les modes de dialogue, l'adéquation du

4. Une maquette correspond à un enchaînement d'écrans avec des données truquées. Elle comporte un minimum de codes (assez pour déclencher les événements) et une présentation graphique relativement complète de l'interface.

5. Un prototype est une version du logiciel réalisant l'ensemble des fonctionnalités de la version définitive, mais réalisé en utilisant des outils (langages de programmation en particulier) permettant une programmation rapide, quitte à sacrifier quelques unes des caractéristiques exigées du produit opérationnel : portabilité, robustesse, voire rapidité.

vocabulaire, l'absence d'information et la présence d'informations superflues. Un prototype, permet en plus de la maquette, de faire des opérations de saisies (mais pas sur la globalité du logiciel), de contrôler les rubriques et les messages. Le produit fini, une recette ou un site pilote permettent la réalisation de tests dans une situation réelle de travail avant la diffusion du produit [BAR 95, VAL 93].

La démarche d'évaluation itérative présente un atout majeur : la spécification d'un ensemble de composants à un niveau donné peut entraîner la remise en cause de certains choix faits au niveau précédent. Cette propriété répond à deux exigences : d'une part, réduire les délais de production du logiciel (Hoyos, Gstalter, Strube & Zang, 1987) et, d'autre part, repenser le cycle de vie d'un projet en intégrant, dans les différentes phases de conception et de fabrication du logiciel, le point de vue de l'utilisateur final (Harker, 1987). L'évaluation et la conception itérative garantit « une meilleure concordance entre les spécifications du logiciel et les besoins réels des utilisateurs. Cet effort de conception permet ensuite de diminuer les coûts de maintenance » [BAR 95, p.84]. Les nouveaux outils graphiques de conception (générateurs d'interfaces) contribuent d'ailleurs largement à ce mode d'évaluation. Ils permettent d'effectuer les simulations des solutions retenues en réalisant des maquettes et de les tester auprès des utilisateurs (Chao, 1987).

Pourtant, malgré le nombre important de contrôles effectués à différents moments du cycle de développement, l'évaluation reste encore insuffisante parce que les mesures et les indicateurs obtenus (fréquence d'erreurs commises, nombre de fonctions inadaptées…) ne sont pas très traités de manière cohérente, ni comparer à d'autres résultats de diagnostic logiciel. C'est pour palier, en partie, ce type de défaut que se développe des méthodes comparatives.

3.3.2.2. *L'évaluation comparative*

Ce type d'évaluation a pour but de faire émerger les qualités et défauts d'une AT par deux modes de comparaisons :

– le premier consiste à comparer le système à d'autres dispositifs qui proposent le même type de service ; les critères retenus peuvent être l'utilisabilité et la performance technique. Rauterberg (1996) a par exemple tenté de mesurer les qualités ergonomiques de différentes interfaces de conception dotées respectivement d'un dispositif de commandes : 1) par menu ; 2) par icônes et manipulation directe (interface graphique) ; 3) par codification (caractéristiques d'une interface textuelle) ; 4) par touches-fonctions et raccourcis-clavier (proche d'une interface textuelle). Il montré que l'interface graphique « à manipulation directe » s'avère la plus ergonomique (en offrant une meilleure visibilité et accessibilité) et la mieux employée par les utilisateurs novices. A l'inverse, les interfaces « à codification » et « à commande » se révèlent nettement moins conviviales pour les débutants ;

– le second mode d'évaluation compare l'usage normal du logiciel à une situation de référence qui correspond à une utilisation optimale du logiciel. Cette situation de référence est établie par la « spécification de performances d'usage ». On élabore des techniques d'analyse de l'impact afin de classer des solutions permettant de corriger le dispositif [SEN 93]. Concrètement, cela démarre par la définition des caractéristiques désirées de l'interface (fonctionnalités et implémentation correspondant aux exigences ergonomiques), puis par le recensement des solutions possibles, et enfin par la sélection des meilleures solutions par des évaluations avec les utilisateurs.

Ce type d'évaluation a par la suite été amélioré par Good *et al*. (1986). Leur méthode d'analyse d'impact a été employée afin d'optimiser l'organisation des fenêtres d'un écran d'ordinateur. Leur approche se basait sur l'observation du comportement des utilisateurs face à l'utilisation d'un prototype, en suivant le protocole suivant :

– mesure du niveau effectif d'utilisabilité,

– recherche des sources de difficultés,

– prédiction des possibilités d'amélioration,

– classement des difficultés par rang (selon l'effet sur la vitesse de travail par exemple).

Une fois l'analyse terminée, les principaux résultats furent soumis aux ingénieurs qui devaient estimer les avantages et les inconvénients de chaque solution proposée. Les critères de sélection retenus étaient le coût et l'impact des corrections à apporter sur l'interface, et la facilité d'implémentation de ces modifications.

3.3.2.3. *L'évaluation par le contrôle qualité*

Le contrôle qualité est une méthode d'évaluation qui porte principalement sur le produit fini, et dont les critères d'appréciation sont fournies par un banc d'essai final. On tâche d'isoler les variables en rapport avec l'utilisabilité et d'élaborer tout un ensemble de scénarii composés des tâches qui serviront à l'évaluation finale. Il existe deux méthodes possibles :

– la méthodologie « universelle » de contrôle qualité au cours de laquelle on établit le profil des utilisateurs pour les entraîner, les tester et les classer. On recueille aussi leurs opinions vis-à-vis du dispositif (par exemple, ce qu'ils aiment, ce qu'ils n'apprécient pas, ce qu'ils souhaiteraient trouver, ce qu'ils n'ont pas trouvé…) ;

– la méthodologie utilisant des « stations d'évaluation » dans lesquelles on dispose d'un dispositif technique élaboré (vidéo, glace sans tain, microphone : Pergolab, voir plus avant) permettant le recueil des données.

Le contrôle qualité peut également évaluer les aspects cognitifs des interfaces. L'objectif est alors de rendre compatible le fonctionnement de l'interface avec les représentations mentales des utilisateurs par la détermination des modèles mentaux employés par les utilisateurs et des conditions de leur apprentissage. On pense ainsi faciliter la recherche et l'exploitation de l'information dans un monde virtuel qui se rapproche le plus possible des représentations conceptuelles et fonctionnelle de l'utilisateur. Schnaider et Hufschmidt (1991) ont par exemple montré qu'une interface graphique, à la différence d'une interface textuelle, réduisait considérablement la charge mentale d'un utilisateur en minimisant le recours à la mémorisation des commandes et des séquences d'action. Les utilisateurs sont plus enclins à manifester des comportements spontanés d'utilisation (appelés *directness*) avec ce type interface car ils opèrent dans ce qui leur semble être leur propre monde (Nanard, 1991 ; Burnett, Baker, Bohus, Carlson, Yang & Zee, 1995)

3.3.2.4. *L'évaluation par experts*

Les évaluations par expert se décomposent en trois catégories [OPP 97] : le rapport de spécialistes ou *expert walkthrough*, le *cognitive walkthrough* et enfin, les « heuristiques d'évaluation ». « L'expert walkthrough » traduit le fait que l'expert parcourt un certain nombre de tâches représentatives. Le « cognitive walkthrough », plus rigoureux que la précédente, consiste à évaluer une interface par rapport à une ou plusieurs tâches. Cette deuxième catégorie d'évaluation experte nécessite donc une description de l'interface (maquette papier, prototype...), un scénario et des hypothèses sur les utilisateurs et le contexte. Les « heuristiques d'évaluations » (check-lists, guides, principes) ont été étudiées par Nielsen [NIE 94]. Cet auteur a notamment réalisé une analyse factorielle de 249 problèmes d'utilisabilité, et a ainsi dégagé 10 heuristiques pour l'évaluation : dialogue simple et naturel, utiliser le langage de l'utilisateur, minimiser la charge mnésique de l'utilisateur, cohérence, feed-back, sorties marquées clairement, raccourcis clavier, bons messages d'erreur, prévenir les erreurs, aide et documentation. En général, l'évaluation par expert se fonde sur les connaissances et l'expérience personnelles de celui-ci. Les appréciations varient donc d'un expert à un autre selon la formation qu'il a reçue et les cas qu'il a eu à analyser. C'est pourquoi il est conseillé, afin d'assurer une évaluation exhaustive, de recourir à une expertise croisée [PER 98, SCA 93].

D'ailleurs, Nielsen [NIE 94] estime que le nombre optimal d'évaluateurs doit être compris entre 3 et 5. Au-delà de ce nombre le rapport coût/bénéfice augmente souvent de façon injustifiée. Sur 6 projets la proportion des problèmes d'utilisabilité trouvée par chacun des évaluateurs n'est que de 35 %. Cependant, les évaluateurs ayant tendance à détecter des problèmes différents, il est alors possible d'agréger les problèmes détectés par eux. Pour déterminer plus précisément le nombre d'évaluateurs souhaitables, Nielsen et Landauer (1993) proposent un modèle fondé sur la formule suivante :

$$\text{Nombre de problèmes trouvé (I)} = N\,(1 - (1 - l)^{i})$$

(I) est une estimation du nombre de problèmes d'utilisabilité trouvé en agrégeant les différents problèmes trouvés par les évaluateurs. i (exposant) est le nombre d'évaluateurs indépendants. N indique le nombre total de problèmes d'utilisabilité et l est la proportion de problèmes trouvés par un seul évaluateur. Sur six études d'utilisabilité la valeur moyenne de l est de 34 % et celle de N de 33. En prenant ces valeurs moyennes on obtient la courbe suivante (figure 3.5) :

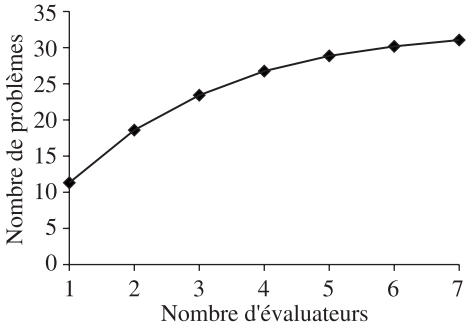


Figure 3.5. *Nombre de problèmes d'utilisabilité trouvé en fonction du nombre d'évaluateurs (d'après Nielsen et Landauer, 1993)*

Avec trois évaluateurs la proportion de problèmes détectés s'élève en moyenne à 71 %. Cette fonction permet aussi de déterminer le rapport financier coût/bénéfice optimal évoqué plus loin dans la troisième partie de ce chapitre et concernant l'impact de l'évaluation. Une étude de Naël [NAE 96] confirme par ailleurs que pour obtenir une validité de 80 % des problèmes détectés, il faut au moins trois experts.

3.3.2.5. *L'évaluation par des check-lists de critères ergonomiques*

Les résultats des expériences d'évaluation d'interfaces menées en laboratoire et sur le terrain ont contribué à la confection de guides pour la conception des systèmes : recommandations ergonomiques, check-lists, ou guidelines. Ces critères accompagnent et assistent les concepteurs dans l'élaboration d'un logiciel [SCA 93]. Ils fournissent ainsi des repères aux concepteurs et aux ergonomes pour évaluer et améliorer la qualité ergonomique des dispositifs.

A titre d'exemple, Bastien *et al.* [BAS 93] proposent huit catégories de recommandations : le « guidage », la « charge de travail », le « contrôle explicite », « l'adaptabilité », la « gestion des erreurs », « l'homogénéité/cohérence », la « signifiance des codes et dénominations » et la « compatibilité ». Chacun de ces critères se décomposent en un ensemble de sous-critères, et ainsi de suite… Nous ne

développerons ici que la première catégorie pour illustrer notre propos. Le Guidage concerne « l'ensemble des moyens mis en œuvre pour conseiller, orienter, informer, et conduire l'utilisateur lors de ses interactions avec l'ordinateur (messages, alarmes, labels, etc.), y compris dans ses aspects lexicaux ». Quatre sous-critères participent au Guidage. Il s'agit de l'incitation, du groupement/distinction entre items, du feedback immédiat et de la lisibilité [BAS 93, p. 45]. L'incitation consiste « à permettre aux utilisateurs de réaliser des actions spécifiques, de faire connaître les possibilités d'action et de savoir où ils se positionnent dans l'interface. Un exemple de recommandation concernant l'incitation est de donner des indices sur la longueur autorisée des entrées dans un champ ».

De telles recommandations, suffisamment détaillées, se révèlent par ailleurs être d'excellents supports de formation à l'ergonomie des logiciels pour des spécialistes et pour des non-spécialistes de l'IHM [SCA 93]. Une étude dirigée par Bastien *et al.* [BAS 94] a d'ailleurs indiqué qu'elles permettaient d'améliorer les performances d'évaluation des experts.

Néanmoins, certains points sont sujets à controverse. En effet, les recommandations sont générales et ne tiennent compte, ni du contexte d'utilisation, ni de la situation d'interaction. Ce qui peut dès lors conduire à quelques contradictions dans l'évaluation du logiciel [KAR 95, SCA 93, SEN 93]. Par exemple, selon ces recommandations, le nombre d'informations affiché simultanément sur un même écran devrait être limité à 7 +/– 2 items pour favoriser la perception et le traitement cognitif. Ce qui se comprend pour l'activité de superviseurs de processus dangereux (central nucléaire, aiguilleur du ciel). Toutefois, certaines opérateurs exigent au contraire un affichage maximum de données sur un minimum d'écran (cas fréquent de malades tétraplégiques, qui avec un contacteur en tout ou rien peuvent accéder à une grande quantité d'informations ; dans ces cas on privilégie l'accès aux données plutôt que l'organisation des données). Pour réduire ces risques, Senach [SEN 93] propose de donner des exemples d'application concrets pour chaque critère. Dans cet ordre d'idées, [OPP 97] développent un test comprenant 300 items qui consiste à mettre en relation les critères ergonomiques avec le dispositif technique (AT). Son principal atout réside dans la prise en compte du contexte d'utilisation et des tâches.

L'application de ces critères ergonomiques pose un autre problème que sont les caractéristiques de l'évaluateur. En effet, ces critères peuvent être interprétés et appliqués de différente manière selon le niveau d'expertise (du novice inexpérimenté au professionnel chevronné, tous n'ont pas la même perception ni la même utilisation de ces critères), le domaine d'activité et l'intérêt personnel (motivation, implication plus ou moins forte dans le projet…) de l'évaluateur. En d'autres termes, on se saurait garantir la qualité ergonomique d'un logiciel par l'application de ces seuls critères tant les facteurs personnels peuvent affecter les conditions de contrôle.

Pollier (1991) a évalué la performance d'évaluation d'une interface (application de gestion de communications multimédias) par différents ergonomes selon deux niveaux : les diagnostics effectués et les stratégies mises en œuvre pour organiser l'examen de l'interface. Pour ce dernier aspect, les résultats indiquent que le processus d'évaluation est très complexe et cinq stratégies d'évaluation sont très imbriquées : stratégies selon les objectifs d'évaluation, la structure de l'interface, les niveaux d'abstraction, les objets de l'interface et les critères ergonomiques. Selon ces stratégies, les démarches des ergonomes varient.

Un autre problème rencontré est le manque de connaissances des évaluateurs pour expliquer l'origine des problèmes ergonomiques identifiés grâce aux critères : il ne sert à rien en effet de proposer des recommandations si, par ailleurs, les motifs des maladresses ergonomiques persistent. Les mêmes causes produisant les mêmes effets, la prochaine version du logiciel reproduira sans nulle doute les mêmes erreurs

Enfin, d'aucuns critiquent le manque d'adéquation des recommandations ergonomiques avec l'avancée des technologies de l'information et de la communication. Baudel [BAU 96] soutient en effet que les critères ergonomiques qui sont habituellement utilisées pour concevoir et évaluer les interfaces opérateur-machine sont inadaptées pour de nouvelles technologies (interface à réalité augmentée, interface de navigation par liens hypertextes, par commande vocale...). Pour lui, le dialogue opérateur-machine serait mieux défini par les caractéristiques technologiques elles-mêmes que par des critères ergonomiques, pour deux raisons :

– première raison : « Evaluer pour qui, par rapport à quoi ? ». Comment et par rapport à quel cadre de référence l'ergonome peut-il évaluer une nouvelle interface, alors que les apprentissages et les modes d'interaction que l'individu déploient dans cette situation originale sont totalement inédits et non pas d'équivalent jusqu'ici ? A cette première remarque, on peut objecter que l'individu essaie d'obtenir le maximum d'effet pour le minimum d'effort (principe d'économie cognitive). Il sera ainsi tenter de minimiser l'apprentissage de la nouveauté en optimisant les transferts de compétences issues de situations antérieures jugées équivalentes. L'individu ne créé donc rien *ex-nihilo*, mais remanie et enrichit ses acquis antérieurs, et plus précisément ses modèles d'interaction développés pour d'autres dispositifs. Dans ces conditions, la plus-value de l'évaluation ergonomique est alors d'éviter des ruptures d'apprentissage entre deux systèmes de travail, et de concevoir des interfaces favorisant les transferts de connaissances entre ces différents univers.

– seconde critique : « Limites de l'utilisabilité des nouveaux outils ». Et quand bien même ces outils ne portent pas de label ergonomique, l'aspect high-tech suffit à séduire les utilisateurs les plus réticents et à masquer l'absence de convivialité des outils (malgré des temps de réponse médiocres par exemple). Certes, on ne peut nier l'effet de séduction que produit certaines innovations technologiques sur les individus, même en dépit de leur piètre qualité et performance. Mais cet

enchantement technologique ne saurait perdurer. Il existe en effet un seuil de tolérance que l'individu ne peut dépasser sans risquer d'augmenter sa propre charge mentale. Autrement dit, plus les ressources de l'individu seront sollicitées pour comprendre, utiliser et mémoriser le fonctionnement de ce dispositif, plus la coopération opérateur -machine s'en trouvera contrariée, avec des risques de sous-utilisation et même de rejet de l'application. La contribution de l'ergonomie est donc de réfléchir, dès la conception, à la meilleure interface possible, afin de rendre accessible ces technologies au plus grand nombre, y compris aux personnes âgées et à celles qui souffrent d'handicaps.

En définitive, si les critères ergonomiques contribuent activement et efficacement à la conception des dispositifs, il n'en demeure pas moins qu'ils doivent évoluer :

– pour s'ajuster au contexte d'utilisation particulier des systèmes (en effet, une l'interface destinée à une tâche de supervision d'un réacteur nucléaire n'est pas la même que celle conçue pour écouter sa messagerie lorsque l'on est déficient visuel) ; aussi les critères retenues pour d'évaluation ne seront pas les mêmes ;

– et pour intégrer pleinement les contraintes et les opportunités d'usage liés aux évolutions technologiques de ces dispositifs. A titre d'exemples, on peut citer l'adaptation des critères ergonomiques de Bastien et Scapin aux nouvelles interfaces Web [BAS 98] ou encore la formulation de critères et de méthode d'évaluation spécifique pour les interfaces virtuelles [KAL 99].

3.3.3. *Les méthodes empiriques*

Ces méthodes se centrent sur l'évaluation des conduites en situations réelles d'interactions.

3.3.3.1. *Le diagnostic d'usage*

Cette méthode d'évaluation concerne les dispositifs déjà intégrés à un contexte de travail et se focalise sur l'expérience acquise des opérateurs. On cherche à déterminer la façon dont l'instrument est utilisé en situation réelle de travail (par exemple, l'usage détournée de certaines fonctions des outils : catachrèse [RAB 95] ; la sous-utilisation ou non utilisation d'autres fonctions…) ainsi que le niveau de maîtrise par l'utilisateur (novice, expert, intermédiaire). Ces mesures se réalisent par le biais de différentes techniques dont le choix dépendra de la situation analysée [SEN 93] : méthode des incidents critiques, questionnaires, mouchards électroniques, entretiens…

Des méthodes proposent par exemple d'appréhender l'interaction opérateur-machine comme un processus continu de cycle d'interaction avec l'environnement. L'action d'une personne change l'état du monde. Il faut alors mener l'évaluation en

respectant et en s'inspirant des objectifs courants des utilisateurs. Cette évaluation conduit à la reformulation de but et d'action à venir, l'ensemble débouchant sur un nouvel état de l'environnement.

3.3.3.2. *Les tests d'usages*

Dans le cycle de vie d'un dispositif technique, on trouve deux étapes où celui-ci doit être impérativement testé auprès des utilisateurs. Le premier test doit être effectué lorsque l'interface utilisateur a été implémentée. Le second apparaît uniquement si le dispositif est modifié. Dans ce dernier cas, les modifications doivent être testées afin de s'assurer si les performances sont correctes et si les nouvelles fonctionnalités correspondent bien aux attentes des opérateurs. Ces tests sont également nécessaires pour détecter les éventuels effets secondaires des modifications sur quelques aspects du dialogue.

Le principe de ces tests repose sur la ratification de l'interface utilisateur par de petits protocoles expérimentaux, destinés à mesurer quantitativement et/ou qualitativement la compatibilité opérateur-instrument. Différents types de tests (performance, utilisation, anticipation, formation, évaluation) mesurent ainsi l'adéquation du dispositif conçu avec les caractéristiques de la tâche et le fonctionnement mental de l'utilisateur.

Ici, l'objet de l'évaluation est plutôt une maquette qui va subir quelques modifications. Cela dit, une interface finale déjà implémentée peut tout à fait être soumise à ces tests, en particulier lorsqu'il s'agit d'estimer son efficacité dans un nouveau contexte d'activité. Idéalement, ces tests doivent déboucher sur une validation de l'interface. Mais généralement ils spécifient plutôt les évolutions de l'interface utilisateur que devront prendre en compte les concepteurs pour améliorer la compatibilité opérateur-instrument. Sans vouloir être exhaustif, il convient tout de même de préciser quelques points communs à tous les tests. L'ingénieur voulant tester la qualité d'une interface doit :

- constituer un échantillon représentatif des utilisateurs,
- enregistrer les interactions, puis les analyser,
- catégoriser les échecs d'utilisation,
- et les interpréter.

Nous allons présenter rapidement plusieurs méthodes de test permettant de vérifier l'adaptation d'une interface aux besoins des utilisateurs.

3.3.3.2.1. Test de performance

Il s'agit d'effectuer des mesures quantitatives des temps d'utilisation pour parvenir à l'objectif déclaré, du nombre de touches utilisées, du nombre d'erreurs

effectuées, des temps de réponse des utilisateurs. Ce protocole de test nécessite de disposer d'un nombre d'utilisateurs relativement important (environ une vingtaine). Les résultats obtenus sont généralement limités mais ont l'avantage d'être fiables. Ce genre de test nécessite du temps et des dispositifs onéreux d'enregistrement des comportements des utilisateurs (caméra vidéo et magnétoscope), ce que peu d'entreprises sont prêtes à accorder.

3.3.3.2.2. Test d'utilisation

Ici, le problème n'est plus d'obtenir des résultats quantitatifs mais des données qualitatives sur l'interaction opérateur-instrument. Au moyen d'observations et d'entretiens menés à partir de plusieurs scénarii d'utilisation, l'expérimentateur s'attache à comprendre l'activité de l'utilisateur, c'est-à-dire à la manière dont il s'y prend pour atteindre son objectif. Les observations doivent mettre en évidence le langage du corps de l'utilisateur. Ses soupirs, froncements de sourcils, hochements de la tête sont les signes d'un obstacle qu'il rencontre. Parfois, l'utilisateur aura tendance à occulter ses difficultés en pensant qu'elles sont dues à un manque d'intelligence de sa part. Or, il n'en est rien. Les difficultés de l'utilisateur sont avant tout liées à un manque de compatibilité opérateur-instrument. Les entretiens individuels ou collectifs peuvent compléter les observations. En outre, le test d'utilisation présente l'avantage d'être peu coûteux, riche en informations, informel, non protocolaire. Il pose néanmoins le problème de la généralisation des résultats obtenus. Il apparaît comme un moyen d'envisager les aménagements ergonomiques pouvant améliorer la compatibilité entre le modèle d'interaction et le modèle de l'utilisateur. Il permet encore de prévoir la maintenance du dispositif technique et représente, à ce titre, un moyen de recueillir des informations sur les nouveaux besoins des utilisateurs.

3.3.3.2.3. Test d'anticipation

Dans ce test, il s'agit de comprendre si les caractéristiques de l'interface permettent aux utilisateurs d'anticiper le déroulement de l'organigramme des fonctionnalités. Ce test repose sur l'idée qu'un modèle facilement utilisable pourrait être celui où l'utilisateur peut anticiper les procédures, les actions, les buts futurs offerts par le logiciel quelle que soit sa position au niveau d'un nœud de l'organigramme. Au cours d'interactions, l'expérimentateur cherche donc à comparer les anticipations de l'utilisateur avec sa démarche effective sur l'interface. Ce test permet de corriger l'interface en l'ajustant à la planification effectuée par l'opérateur. Le but recherché est donc une augmentation de la compatibilité la structuration de l'interface avec les caractéristiques mentales de l'opérateur et de sa tâche.

3.3.3.2.4. Test de la formation

Ce test a pour objectif de mesurer l'adéquation des dispositifs d'aide (système d'aide en ligne, manuel d'utilisation, formation) aux besoins des utilisateurs. Des

entretiens menés auprès des utilisateurs doivent permettre de définir les situations de travail conduisant à des erreurs d'utilisation. Une fois ces situations catégorisées, il est impératif pour l'ergonome de mesurer la capacité des dispositifs d'aide à parfaire la formation de l'utilisateur. Il doit alors faire une analyse ergonomique des dispositifs d'assistance et chercher à définir le contenu et la durée de la formation pour qu'un groupe d'utilisateurs donné atteigne un niveau de performance optimal. Ce test de la formation a un double intérêt. En premier lieu, il met en évidence des défauts ergonomiques des dispositifs d'assistance et donc permet de les améliorer. En second lieu, il indique les erreurs de conception du dispositif technique, et dans ce cas il peut conduire les concepteurs à réviser une partie de l'interface.

3.3.3.2.5. Test instrumenté d'une grille d'évaluation

Ce test se présente sous la forme d'un questionnaire d'évaluation des caractéristiques de l'interface (comme par exemple Shneiderman, 1987). Dans ce cas, l'opérateur est prié de répondre aux questions, d'indiquer les difficultés qu'il a rencontrées avec l'interface…. Indirectement, il évalue la qualité de l'interface développée par les concepteurs. Les résultats du questionnaire guident d'une part, les corrections à apporter sur l'interface et d'autre part, le type de formation particulière à délivrer sur les points que les utilisateurs ont identifiés comme délicats.

3.3.3.2.6. Vers des tests automatisés ?

De récents travaux ont permis de développer des systèmes techniques capables de générer automatiquement des interfaces respectant les critères ergonomiques, mais aussi de procéder à l'évaluation automatique de l'interaction opérateur-instrument [ELL 99, JAM 99]. Grâce aux technologies agents[6], des interfaces ont la capacité de s'adapter dynamiquement au contexte de travail et à l'utilisateur [DUV 00]. Mais pour ce qui est du test automatique des interfaces, Ella, Kolski, Jacques et Yim [ELL 99] ont développé un outil d'évaluation qui intègre non seulement les critères d'utilité et d'utilisabilité traditionnelles, mais qui sera aussi capable d'apprécier des critères d'acceptabilité socioculturelle (dimension sociale, éthique et culturelle sur la couleur, la langue, le sens et la sémiologie de certains termes…).

En définitive, s'il est établi que les tests permettent de témoigner de l'adéquation des systèmes techniques aux besoins des utilisateurs (c'est-à-dire ce qui marche), ces tests doivent aussi et très paradoxalement s'attacher à démontrer le contraire, c'est-à-dire que ces systèmes ne marchent pas. Mais ces tests d'évaluation présentent également un tout autre intérêt. Ils sont le lieu où l'utilisateur final a la possibilité de s'exprimer dans le processus de conception : il peut demander voire exiger des

6. Agent : Petit programme qui fournit un service particulier à l'utilisateur comme identifier ses principales erreurs ou préférences afin de configurer une interface adaptée.

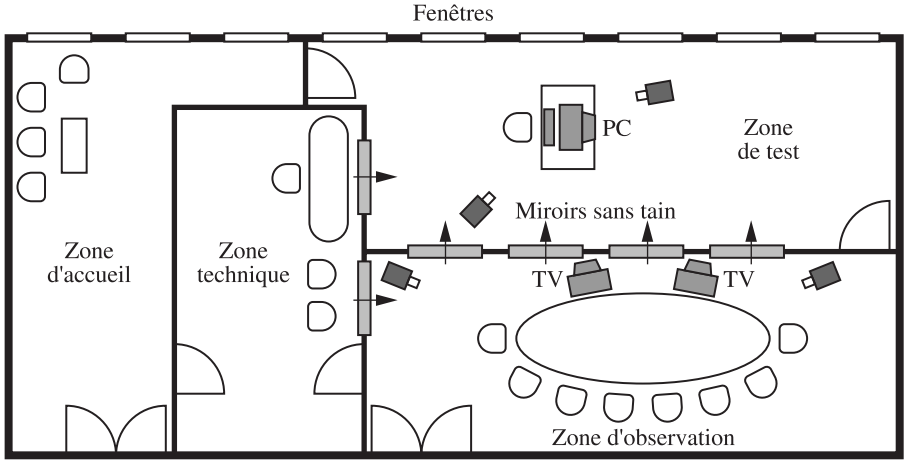
modifications significatives sur sa future interface. Le test devient dès lors l'occasion de formuler de nouveaux besoins. Pour cette raison, c'est aussi un moment privilégié de dialogue et donc de collaboration entre les concepteurs et les utilisateurs.

3.3.3.3. *Le laboratoire d'utilisabilité (Pergolab :* Psycho-ergonomie-laboratoire *ou* User Lab*)*

Physiquement, le laboratoire d'utilisabilité correspond à un lieu regroupant quatre salles séparées par des miroirs sans tain, qui isolent le(s) sujets (utilisateurs) des observateurs (concepteurs, chef de projets, informaticiens, responsable utilisateur, ergonome… en fait toutes personnes impliquées dans la conception du dispositif).

Les premiers sont dans une salle recréant leur environnement de travail, les seconds sont regroupés autour d'une table, face à la glace dans tain et à des moniteurs. Par un circuit fermé d'enregistrement vidéo, plusieurs observations vont être menées simultanément : des caméras enregistrant le comportement du sujet lors de l'interaction avec l'interface, d'autres saisissant la dynamique des échanges entre les observateurs, enfin une dernière caméra joue le rôle de convertisseur pour présenter le contenu et les manipulations de l'utilisateur sur l'écran informatique.

La figure 3.6 donne une illustration du plan d'organisation du laboratoire d'utilisabilité (Pergolab de l'université de Metz)



3e étage bâtiment C

Figure 3.6. *Architecture spatiale du Pergolab (plan vue de dessus)*

La structuration assez particulière de ce laboratoire répond à une triple préoccupation :

– primo, évaluer grâce à des scenarii et des méthodes d'observations fiables (enregistrements vidéos) l'utilisatibilité et l'utilité des dispositifs. Par exemple, l'*Usability Methods Research Laboratory* du Centre d'ingénierie des facteurs humains (*the Human Factors Engineering Center*) de la Virginia Tech permet d'évaluer les dialogues, l'aide en ligne, les messages d'erreurs, les dispositifs d'entrée et de sortie et la présentation [WILL 96]. Cela dit, ce laboratoire s'attache aussi à révéler la contrepartie verbale et non verbale des impressions de l'utilisateur, et de manière plus générale la perception effective du sujet lors de l'expérimentation de l'application et de son intégration dans sa logique et dans sa vision de la tâche (Reconnu, 1996) ;

– secundo, il s'agit aussi de faire prendre conscience aux observateurs, responsables du projet ou de l'organisation, que les conduites d'interaction attendues chez l'individu ou les usages prévus du logiciel ne correspondent pas à ceux qui sont réellement mis en œuvre par le sujet. Tous les acteurs du projets observent le déroulement du test et assume le vécu de l'utilisateur sans pouvoir intervenir par des questions, commentaires, attitudes et comportements. C'est sur la base de ces observations qu'ils vont échanger et discuter pour comprendre l'origine de ces dysfonctionnements et prendre des décisions afin de corriger et d'améliorer le dispositif ;

– tertio, grâce à un jeu de caméras tournés vers cette salle d'observateurs, l'intérêt du laboratoire ne réside plus seulement dans l'analyse « microscopique » des comportements de l'utilisateur. Pergolab permet d'appréhender les tâches de test et d'observation comme un système dynamique organisé autour d'un jeu de relations, de négociations, de compromis dans le processus de conception.

Notons aussi que des variantes portables existent à ce type de laboratoire. Rowley (1994) décrit comment on peut amener le laboratoire chez l'utilisateur en installant une caméra vidéo chez celui-ci, ce qui n'est pas sans poser d'autres problèmes scientifiques et déontologiques.

3.3.4. *Synthèse : validité et pertinence des méthodes d'évaluation des interactions opérateur-instrument*

Dans cette partie, nous avons présenté diverses méthodes d'évaluation des interfaces. Celles-ci, nombreuses, ont été classées d'après leur objet d'étude : évaluation des interactions en situation de travail (démarche empirique), évaluation de différentes dimensions de l'interaction par la validation de critères (démarche analytique) évaluation des interactions et des performances attendues dans le futur dispositif (démarches prédictives). Elles n'ont pas les même fonctions, ni les mêmes atouts.

Si, les méthodes empiriques peuvent se prévaloir d'être la référence incontournable pour évaluer les qualités ergonomiques d'une interface, cela s'explique par le fait qu'elles seules permettent de prendre véritablement en considération les utilisateurs dans leur contexte de travail et d'interaction réel. Mais des critiques persistent néanmoins ; en particulier les question de coût et de mise en œuvre. En effet, ces techniques sont souvent onéreuses et difficiles à utiliser (on paralyse une partie du temps de travail).

De même, l'accès aux utilisateurs n'est pas toujours aisé et les biais dans l'observation et l'expérimentation relativisent souvent la généralisation des résultats (échantillon non significatif...) [BAR 95, SCA 93]. Cependant, l'approche empirique demeure l'une des meilleures méthodes lorsque l'on se trouve confronté à de nouvelles problématiques. Les méthodes analytiques présentent, quant à elles, l'avantage d'avoir lieu pendant les différents moments du processus de conception, et donc de réajuster en permanence le dialogue opérateur-instrument. Elles donnent également la possibilité d'avoir des points de vue microscopiques, ou du moins spécifiques sur l'interaction opérateur-instrument (sur les aspects syntaxiques ou graphiques de l'interface par exemple). Le principal défaut de cette méthode réside toutefois dans l'absence d'implication de l'utilisateur dans certaines étapes du processus d'évaluation. Enfin, les modèles prédictifs permettent à la fois de définir et de prédire les conduites d'interaction futures probables des usagers dans le dispositif final. Elles fournissent ainsi des indicateurs assez précis pour construire le dialogue opérateur-instrument. Mais ici, c'est la situation d'interaction réelle qui n'est pas prise en compte, mais seulement simulée avec des scénarii de travail.

Le tableau 3.1 présente une synthèse critique de ces différentes méthodes. La grille de lecture est la suivante : pour chaque démarche d'évaluation exposée (Quoi ?), on définit le moment de sa mise en œuvre (durant le maquettage, le prototypage, sur le produit fini ou tout au long du processus de conception : Quand ?), le lieu où elle se déroule (en situation réelle de travail ou en situation hors travail – laboratoire par exemple – : Où ?), la méthode et les techniques de recueil de données employées (comment ?), les personnes susceptibles d'employer cette démarche (ergonome et/ou concepteur : Par qui ?), et enfin les avantages/inconvénients de chacune d'entre elles.

DEMARCHES EMPIRIQUES : DIAGNOSTIC D'USAGE			
Quand ?	**Où ?**	**Comment ?**	**Par qui ?**
Tests sur le produit fini ou site pilote	Lieu de travail	Possibilité d'utiliser les différentes techniques subjectives et objectives de recueil de données	Ergonome

Avantages	Inconvénients
– référence primordiale, – identification des problèmes sérieux et récurrents, – évaluation des impacts effectifs de la situation de travail sur l'interaction opérateur-instrument, – prise en compte de la diversité des acteurs et de la pluralité des situations de coopération opérateur-instrument, – observation des usages réels, détournés ou inadaptés de l'interface, – scénarii comprenant des tâches réelles, – propositions de recommandations très complètes, – prise en compte de tous les aspects du dispositif technique, – possibilité d'ajuster la formation et la documentation, – implication réelle des utilisateurs dans l'évaluation et adhésion au projet, – permet de contrôler les performances avant l'exploitation ou la commercialisation (cas des tests d'usage), – valide les corrections réalisées à partir de la maquette et complète les informations (saisie, contrôle des rubriques et messages).	– coûts élevés des corrections, – complexité technique et fonctionnelle pour réaliser les modifications, – mise en place assez lourde (mobilise les acteurs et une partie de leur activité).

DEMARCHES EMPIRIQUES : TESTS D'USAGE			
Quand ?	**Où ?**	**Comment ?**	**Par qui ?**
Tests sur un produit fini, une évolution de version ou en maintenance	Situation de travail, situation hors travail (laboratoire)	Ratification de l'interface utilisateur par de petits protocoles expérimentaux, Possibilité d'utiliser les différentes techniques subjectives et objectives de recueil de données	Ergonome éventuellement assisté du concepteur

<table>
<tr><th colspan="2">Avantages</th><th colspan="2">Inconvénients</th></tr>
<tr><td colspan="2">Idem diagnostic d'usage</td><td colspan="2">Idem diagnostic d'usage</td></tr>
<tr><th colspan="4">DEMARCHES EMPIRIQUES : LABORATOIRE D'UTILISABILITE</th></tr>
<tr><th>Quand ?</th><th>Où ?</th><th>Comment ?</th><th>Par qui ?</th></tr>
<tr><td>Tests sur le produit fini, un prototype ou une évolution de version</td><td>Situation hors travail (laboratoire) et lieu de travail (avec un système d'observation embarqué)</td><td>Situation d'interaction opérateur-instrument filmés intégralement et méticuleusement, et observée par les principaux acteurs du projet derrière une glace sans tain. Possibilité d'utiliser les différentes techniques subjectives et objectives du recueil de données</td><td>Ergonomes et toutes les personnes impliquées dans la conception du projet (chef de projets, concepteurs, responsable utilisateur…)</td></tr>
<tr><th colspan="2">Avantages</th><th colspan="2">Inconvénients</th></tr>
<tr><td colspan="2">– le laboratoire est une sorte de loupe grossissante,
– identification des problèmes microscopiques,
– évaluation des impacts effectifs de la situation de travail sur l'interaction opérateur-instrument,
– prise en compte de la diversité des acteurs et de la pluralité des situations de coopération opérateur-instrument, notamment par les jeux de triangulation rendues possibles par les miroirs sans tain,
– mise en relation des concepteurs (derrière le miroir) et des utilisateurs (en zone d'observation,
– possibilité de comprendre les aspects psychosociaux liés à l'usage de l'aide technique,
– scénarii comprenant des tâches réelles,
– propositions de recommandations très complètes,
– restitution du test par montage de séquences vidéos,</td><td colspan="2">– nécessite de bien sérier les variables étudiées,
– importance de l'animateur (psychologue ergonome) pour mettre en œuvre les expériences.</td></tr>
</table>

<table>
<tr><td colspan="2">– prise en compte de tous les aspects du dispositif technique,
– possibilité d'évaluer également la formation et la documentation,
– permet de contrôler les performances avant l'exploitation ou la commercialisation,
– permet de prévoir des moyens pour accompagner la mise en place de l'aide technique (soutien, formation, résistance au changements…),
– valide les corrections réalisées à partir de la maquette et complète les informations (saisie, contrôle des rubriques et messages).</td><td colspan="2"></td></tr>
<tr><td colspan="4">DEMARCHES ANALYTIQUES : EVALUATION ITERATIVE</td></tr>
<tr><td>Quand ?</td><td>Où ?</td><td>Comment ?</td><td>Par qui ?</td></tr>
<tr><td>Evaluation tout au long du cycle de conception du système technique</td><td>Situation hors travail</td><td>Evaluation sur la maquette, le prototype ou le produit fini par des mises en situation d'utilisation (simulés)
Possibilité d'utiliser toutes les techniques subjectives et objectives de recueil de données</td><td>Ergonome éventuellement assistés des concepteurs</td></tr>
<tr><td colspan="2">Avantages</td><td colspan="2">Inconvénients</td></tr>
<tr><td colspan="2">– identification des problèmes sérieux et récurrents,
– « objectivité »,
– effet sur la conception grâce à sa force démonstrative,
– permet aux utilisateurs de se projeter plus facilement dans la future situation de travail,
– évalue l'interface de façon plus riche que sur papier,
– permet des remarques sur l'insertion du logiciel par rapport à la situation actuelle de travail,
– évaluation renseigne sur les modes de</td><td colspan="2">– coût élevé pour obtenir des maquettes fiables et réalistes,
– dans le cas de tests sur la maquette, ne permet pas de tester les modes de saisies, les contrôles de rubriques, les messages, les types de représentation...
– difficulté de différencier les problèmes d'apprentissage et ceux d'utilisation car les utilisateurs ne sont pas encore expérimentés,
– difficulté ou impossibilité d'intégrer les variables de la situation réelle de travail (variables extérieures),
– pas de possibilité de réutiliser la maquette dans le développement du produit réel,</td></tr>
</table>

dialogue, l'adéquation du vocabulaire, les présentation des informations, l'enchaînement des fonctionnalités… – réduit les coûts (il est plus économique et facile de réaliser des corrections sur une maquette ou un prototype que sur le produit fini), – permet d'introduire progressivement dans la conception des éléments d'interface validé, – permet de contrôler les performances avant l'exploitation ou la commercialisation, – augmenter la qualité du produit, – favorise l'adhésion au projet par l'implication des utilisateurs dans la procédure d'évaluation, – réutilisation possible pour concevoir le dispositif, – valide les corrections réalisées à partir de la maquette et complète les informations (saisie, contrôle des rubriques et messages), – distinction entre les problèmes d'apprentissage et ceux d'utilisation, – dans le cas du prototype, simulation plus proche de la réalité que la maquette (cas du prototype), – projection plus facile pour les utilisateurs dans leur futur tâche, – plus d'éléments pour préparer à la formation, concevoir des aides en lignes et ajuster l'organisation du travail.	– ne prend en compte que des données brutes (limitées aux indicateurs comportementaux spécifiés), – intervient parfois trop tard dans le processus, – contrôles et messages ne sont pas complets car l'accès à la base de données et les temps de réponse ne peuvent être évalués.

DEMARCHES ANALYTIQUES : EVALUATION COMPARATIVE			
Quand ?	**Où ?**	**Comment ?**	**Par qui ?**
Evaluation du produit final	Situation de travail Situation Hors travail (laboratoire)	Comparaison du système technique avec d'autres systèmes présentant les mêmes services Comparaison de la situation d'interaction avec une situation d'interaction	Ergonome assisté des concepteurs

		optimale de référence Possibilité d'utiliser toutes les techniques subjectives et objectives de recueil de données	
Avantages		**Inconvénients**	
– évalue la qualité, les défauts et la performance de l'interface par rapport à d'autres instruments, – permet de classer et de catégoriser les corrections à apporter par ordre de priorité, de coûts ou de performance, – conduit à faire une certaine veille technologique, – instaure un dialogue entre concepteur et ergonome pour évaluer la pertinence des corrections à apporter.		– absence de l'utilisateur et non prise en compte du contexte de travail (dans le cas de l'évaluation comparative des logiciels), – de même, ne prend pas en compte les conditions de développement des systèmes (toujours dans l'étude comparative des logiciels). C'est le concepteur et non l'utilisateur qui détermine le classement des modifications à apporter sur des critères davantage techniques et fonctionnels qu'ergonomiques.	
DEMARCHES ANALYTIQUES : EVALUATION PAR LE CONTROLE DE QUALITE			
Quand ?	**Où ?**	**Comment ?**	**Par qui ?**
Evaluation du produit final	Situation hors travail	Bancs d'essai sur les aspects formels et logique de l'interface Possibilité d'utiliser toutes les techniques subjectives et objectives de recueil de données	Ergonome
Avantages		**Inconvénients**	
– permet de cerner et de classer le profil des utilisateurs, – permet d'obtenir des opinions générales sur la qualité de l'interface, – permet de tester rapidement l'acceptabilité d'un nouveau dispositif.		Ne donne pas de préconisations d'amélioration spécifiques, juste une impression de satisfaction générale	
DEMARCHES ANALYTIQUES : EVALUATION PAR EXPERTS			
Quand ?	**Où ?**	**Comment ?**	**Par qui ?**
Evaluation du produit final	Situation hors travail et travail	Confronter et croiser les évaluations de plusieurs experts.	Ergonome

		Possibilité d'utiliser toutes les techniques subjectives et objectives de recueil de données	
Avantages		**Inconvénients**	
– identifient de nombreux problèmes, – identifient des problèmes sérieux, – coût faible, – prédisent plus que les besoins de l'évaluation, – tolérance des « bugs », – effets sur la conception, – bon marché.		– requiert une certaine expertise et plusieurs évaluateurs, – « subjectivité » : dépend de la qualité et de l'expérience de l'expert ; scénarios limités, – sous estimation des blocages des utilisateurs finaux.	
DEMARCHES ANALYTIQUES : EVALUATION PAR DES CHECK-LISTS DE CRITERES ERGONOMIQUES			
Quand ?	**Où ?**	**Comment ?**	**Par qui ?**
Evaluation du produit final	Situation hors travail	Validation d'une liste de critères ergonomiques Grille de critères ergonomiques	Ergonome, concepteur
Avantages		**Inconvénients**	
– tôt dans le processus, – identifient des problèmes généraux, – peuvent être utilisés par des non-spécialistes, – applicables à tous les niveaux du design, – évaluation rapide, – bon support de formation.		– omettent certains problèmes graves, – peuvent être mal appliqués, – difficiles à suivre, – parfois trop généraux, – interprétations contradictoires, – pas de prise en compte de la tâche réelle, du contexte, – ne touche que les aspects de surface, ne s'intéresse pas aux principes de fonctionnement ni à la logique de l'interface, – recommandations très superficielles.	
DEMARCHES PREDICTIVES : MODELES DE DEFINITION DES INTERACTIONS OPERATEUR-INSTRUMENT			
Quand ?	**Où ?**	**Comment ?**	**Par qui ?**
Evaluation d'un produit en phase de conception, d'un	Situation hors travail	Modélisation de l'interaction opérateur-instrument	Ergonome, concepteur

maquette, d'un prototype		future probable par le biais de diagrammes de conception	
Avantages		**Inconvénients**	
– permet de spécifier, dès la conception du logiciel le dialogue opérateur-instrument, – détermine la logique et la cinématique de la future interface, – permet d'anticiper les difficultés d'interaction possibles.		– non prise en compte de l'utilisateur final et des caractéristiques de son activité, – modèle assez rigide.	
DEMARCHES PREDICTIVES : MODELES PREDICTIFS DES PERFORMANCES DE L'UTILISATEUR			
Quand ?	**Où ?**	**Comment ?**	**Par qui ?**
Evaluation d'un produit fini	Situation de travail	Modéliser la connaissance de conception et prédire les performances d'utilisation	Ergonome
Avantages		**Inconvénients**	
– tôt dans le processus, – analyse quantitative, – donnent des renseignements inattendus, – « Objectivité », – un bon modèle de prédiction de la performance.		– extrême complexité, – tendent à se focaliser sur une seule dimension, – lourds à mettre en œuvre et requiert une certaine expertise, – peu utilisés sur le terrain, – assez complexe à utiliser.	

Tableau 3.1. *Résumés synthétiques des différentes démarches d'évaluation*

3.4. Perspectives pour l'organisation de l'évaluation ergonomique

Dans le paragraphe précédent, par souci de clarté, nous avons pris le parti de distinguer volontairement les démarches d'évaluation dans trois grandes classes. Il est toutefois nécessaire de préciser que ces méthodes ne s'opposent pas, bien au contraire, elles se complètent et peuvent même être intégrées dans une seule et unique démarche d'évaluation des systèmes techniques. Une autre approche consiste à sélectionner la méthode qui correspond le mieux aux objectifs et au contexte de l'évaluation ergonomique [CHR 95, DOW 89, Dowell 1997, SAW 96, SEN 93].

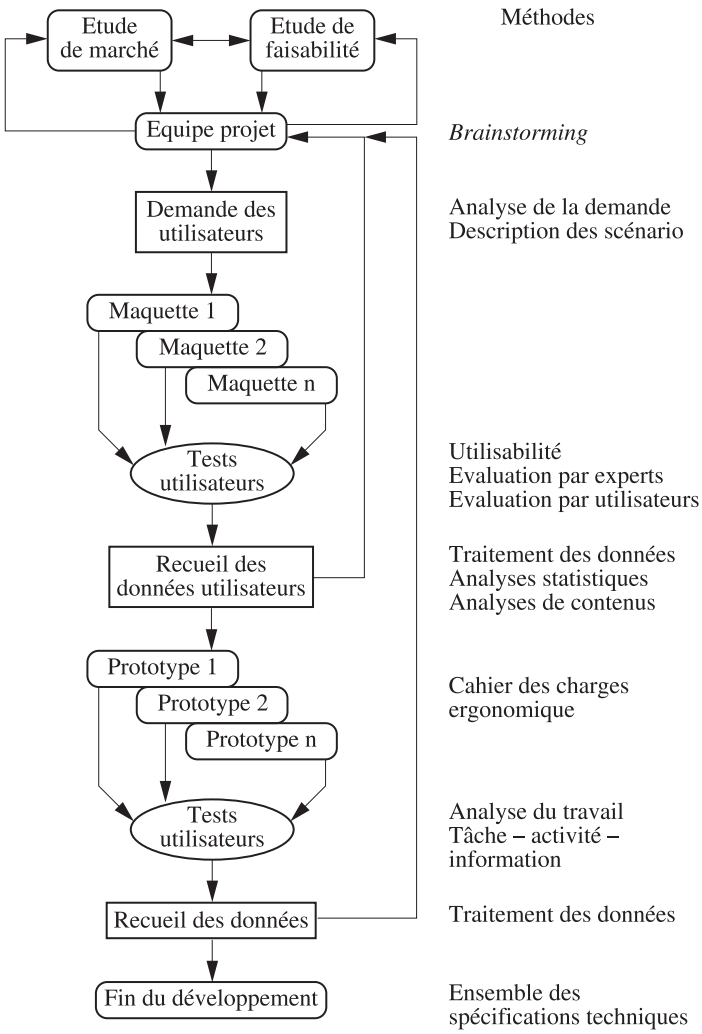


Figure 3.7. *Processus itératif « centré utilisateur » d'utilisabilité d'un produit ou d'un service (d'après Crerar et Benyon, 1998)*

3.4.1. *Vers une démarche intégrée d'évaluation des aides techniques ?*

Par comparaison avec les situations classiques d'analyse du travail, l'évaluation de nouveaux produits et services destinées à un public non professionnel requiert une méthodologie particulière. Dans la mesure ou l'AT est en cours de conception,

l'activité des personnes handicapées ne peut être observée en situation réelle. Il convient donc d'instaurer une démarche reposant sur deux principes :

- initier une démarche itérative (figure 3.7),
- décrire l'activité future probable.

A ce titre un projet de conception d'une AT devra mettre l'utilisateur futur au cœur de la démarche de conception. Par comparaison avec les situations plus classiques en matière de développement de nouvelles technologies dites « centrées technologies », l'évolution actuelle vise à piloter des projets « centrés utilisateur ». L'utilisateur y est observé *a priori* et non *a posteriori*.

La première boucle, correspondant aux maquettes successives, doit être courte. Le cycle « évaluation/recommandations/modifications de la maquette » doit mettre en œuvre des techniques et une méthode d'évaluation simples et rapides. La seconde boucle du processus, correspondant à la validation du prototype, justifie un cycle plus long. Il s'agit à ce stade, non plus d'évaluer certains facteurs de l'utilisabilité de l'AT, mais plutôt de satisfaire, en situation simulée, puis réelle, à l'ensemble des critères d'utilisabilité. A cet égard, et pour répondre aux exigences particulières de ce processus itératif à double boucle, nous avons précédemment décrit les différentes méthodes d'évaluation et les techniques utilisables. A titre d'exemple on peut citer les travaux de Valentin *et al*. [VAL 93] dans le domaine du secteur tertiaire. La démarche d'évaluation itérative a été mise en œuvre à partir d'une douzaine d'évaluations sur le terrain. Elle s'appuie sur des entretiens avec les concepteurs, des évaluations d'expert et des check-lists et des tests sur les maquette, prototypes ou logiciels réalisés avec les utilisateurs (comprenant des observations et questionnaires). Cette démarche se déroule en cinq étapes[7] :

- prise de connaissance du logiciel,
- définition de la population et des scénarios,
- recueil des données,
- analyse des données recueillies,
- synthèse et analyse des résultats.

3.4.2. *Vers une sélection justifiée des techniques d'évaluation*

D'autres auteurs ont proposé des méthodologies ayant pour objectif de sélectionner la méthode d'évaluation idoine. Parmi celles-ci, on pourra citer l'approche de Senach [SAN 93] qui propose de :

7. Pour une description plus détaillée de cette démarche, on pourra se référer à l'ouvrage de Valentin [VAL 93].

– définir les objectifs de l'évaluation, qui s'expriment en réponse à une question (par exemple, « est-ce qu'un dispositif est utilisé ou non ? ») ;

– identifier les dimensions d'analyses pertinentes : s'agit-il de mesurer l'utilité (c'est-à-dire déterminer si le produit satisfait à des contraintes fonctionnelles et opérationnelles) ou l'utilisabilité (la qualité des interactions opérateur-machine) ?

– identifier les variables cibles : suivant le contexte, doit-on plutôt se centrer sur des mesures de performance ou s'intéresser aux opinions et attitudes des utilisateurs ?

– choisir les techniques de recueil des données et définir les modalités et la méthode d'évaluation (en fonction du contexte) ;

– et enfin de recueillir les données et les analyser pour en tirer les conclusions qui s'imposent.

Christie *et al*. [CHR 95] préconisent quant à eux de :

– évaluer les buts : des buts précis doivent être assignés à chaque phase de l'évaluation (par exemple, niveau d'implication des concepteurs et des managers) ;

– sélectionner les outils et méthodes à employer suivant plusieurs critères : validité de la méthode, ressources disponibles (financières, matérielles...), temps disponible, considérations logistiques et éthiques ;

– définir les membres de l'équipe d'évaluation en essayant de rapprocher concepteurs, évaluateurs et utilisateurs ;

– communiquer les résultats. Cette phase s'inscrivant en réponse aux questions : où ? quand ? et à qui ?

Bien que ces démarches aient le mérite de planifier, de manière assez précise, des axes de réflexion avant d'utiliser une méthode d'évaluation, on peut regretter leurs niveaux d'abstraction et le manque de données quant à leurs implications concrètes.

C'est dans le souci de remédier à ces difficultés que Denley et Long [DEN 97] ont conçu une démarche d'aide à la planification de l'évaluation des logiciels. Cette méthodologie a une portée suffisamment générale pour qu'elle puisse être, nous semble-t-il, appliquée à l'évaluation des systèmes techniques et plus spécifiquement des AT Elle est destinée en première instance à des évaluateurs inexpérimentés. Elle a été développée avec l'appui de techniques de recueil d'expertise. Ces auteurs ont interviewé des spécialistes du domaine afin de dégager les règles à respecter dans le développement d'un plan d'évaluation. Cette approche accorde une grande importance à la notion de performance [DOW 89], c'est-à-dire au rapport entre la qualité de l'évaluation (informations quantitatives ou qualitatives, types de données recueillies...) et son coût (temps, nombre d'utilisateurs concernés...). Elle comporte 4 phases distinctes auxquelles seront associées une méthode destinée à déterminer l'impact effectif de l'évaluation (exemple du tableau 3.2).

Etape	Sous-objectifs
1. Identifier les facteurs de performance	1. Formuler le problème concernant l'I.H.M. et son utilisation 2. Identifier les objectifs de développement du client 3. Identifier la phase du cycle de vie de l'I.H.M. 4. Identifier la performance désirée de l'I.H.M. 5. Identifier les buts que le client assigne à l'évaluation 6. Identifier les sources d'informations relatives à la performance 7. Créer une liste de facteurs de performance 8. Valider la liste
2. Identifier les facteurs organisationnels	1. Identifier le statut du projet et son historique 2. Identifier l'opinion du client sur l'ergonomie 3. Identifier les compétences du client 4. Identifier les implications de l'évaluation au niveau du management 5. Identifier le niveau d'exigence requis pour le système 6. Identifier les sources d'informations relatives à l'organisation 7. Créer une liste des facteurs organisationnels 8. Valider la liste
3. Spécifier le niveau de performance désirée	1. Rédiger un document spécifiant le niveau de performance 2. Valider le document auprès du client

Tableau 3.2. *Les différentes étapes de la définition du niveau de performance, d'après Denley et Long [DEN 97]*

3.4.2.1. *Phase 1 : définition du niveau de performance*

La première étape relative à la définition de niveau de performance intègre l'étape d'analyse de la demande formulée par le client. Cette phase revêt une importance particulière car c'est à partir de cette définition qu'il sera possible de choisir une méthode appropriée. En l'occurrence, il s'agit d'effectuer le rapport entre la qualité de l'évaluation et son coût. En d'autres termes, il faut trouver un compromis acceptable entre ce qu'il faudrait évaluer et ce qu'on peut effectivement évaluer (compte tenu des contraintes matérielles). Il est donc nécessaire de prendre en compte les pressions des réalités, qu'elles soient financières, temporelles ou humaines [NAE 96]. Très concrètement, la tâche de l'évaluateur consiste à recueillir des informations se rapportant aux facteurs de performance et aux facteurs organisationnels. Le niveau de performance attendu se matérialise sous la forme d'une liste d'objectifs prioritaires à atteindre. L'identification des facteurs

organisationnels se rapporte à l'analyse de la demande du client. Celle des facteurs de performance au niveau de performance demandée. A chaque étape correspond un certain nombre de sous-objectifs. Ceux-ci se subdivisent eux-mêmes en tâches précises à effectuer (par exemple, identifier l'environnement d'utilisation), en techniques à employer (par exemple, entretiens, questionnaires...), en entrées (par exemple, documentation), en sorties (les résultats du sous-objectif) et en conseils (par exemple, vérifier la cohérence du système, de la tâche et du groupe d'utilisateurs). Une fois le document rédigé, il va se révéler utile aux deux parties en présence : l'évaluateur peut garder une trace de ce qui a été convenu au départ (le client ne peut donc pas changer les objectifs en cours d'évaluation), le client peut vérifier à tout moment que l'évaluateur répond à sa demande. Cette phase achevée, il est alors possible de choisir le type de méthode convenant à l'évaluation.

3.4.2.2. *Phase 2 : sélection des méthodes d'évaluation*

Long et Dowell [DOW 89] classent les méthodes d'évaluation des logiciels selon trois catégories :

1) une discipline artisanale (étayée par des connaissances implicites),

2) une science appliquée,

3) une discipline d'ingénierie[8].

Puis ils développent une procédure de sélection en s'appuyant sur les travaux de Hill *et al*. [HILL 91]. Pour ces derniers, la tâche d'évaluation comporte cinq dimensions essentielles :

1) but : les développeurs peuvent assigner deux buts à une évaluation : soit proposer des changements dans le design, soit diagnostiquer les difficultés de l'utilisateur. En conséquence, les méthodes correspondantes sont soit prescriptives, soit descriptives ;

2) sortie : les données constituant la sortie de l'évaluation sont soit quantitatives, soit qualitatives ;

3) entrée : cet aspect se rapporte aux moyens dont dispose l'évaluateur. En l'occurrence, on prend en compte les ressources en personnel, le fait que le système soit utilisable ou non et la présence éventuelle d'utilisateurs ;

4) rapport entrée/sortie : ce rapport concerne la manière dont sont analysées les données. Celles-ci peuvent être traitées subjectivement ou objectivement ;

5) procédure : pour configurer les méthodes.

8. Si cette classification n'était pas apparue pertinente dans le cadre de la présentation des méthodes d'évaluation de la partie précédente, elle s'avère néanmoins indispensable dans le processus de sélection élaboré par Denley & Long.

Par ailleurs, le cycle de vie du logiciel doit également être pris considération pour choisir la méthode appropriée.

Ce processus de sélection des méthodes d'évaluation s'opère en deux temps : (i) l'évaluateur doit tout d'abord définir les dimensions pertinentes à l'évaluation, puis (ii) choisir la catégorie de méthodes qui correspond le mieux à ces dimensions.

Pour l'aider dans ses choix, l'évaluateur dispose d'une série de tableaux synthétisant des analyses effectuées auprès d'évaluateurs confirmés (tableau 3.3).

		CLASSES DE METHODES		
DIMENSIONS DE L'EVALUATION		**Artisanales**	**Science appliquée**	**Ingénierie**
But	Prescriptif	√	√	√
	Descriptif	√	√	√
Rapport entrée/sortie	Subjectif	√		
	Objectif		√	√
Sortie	Quantitative		√	√
	Qualitative	√		
Phase du cycle de vie	Analyse			√
	Spécification		√	√
	Implémentation	√	√	
	Test	√	√	

Tableau 3.3. *Sélection de la classe de méthodes, d'après Denley et Long [DEN 97]*

Après avoir sélectionné la classe de méthodes, l'évaluateur choisira une technique d'évaluation spécifique. Pour ce faire, il a la possibilité de s'inspirer de toute une série de tableaux (3.4 et 3.5) qui précisent les critères et les exigences de mise en œuvre de chaque démarche pour une classe déterminée.

Par exemple, si les ressources en personnel sont moyennes, que le système existe et que des utilisateurs sont disponibles, alors on pourra choisir l'observation (tableau 3.4).

		METHODES ARTISANALES			
DIMENSIONS DE L'EVALUATION		*Walkthrough*, Guides, méta-évaluation	Observations	Entretiens	Verbalisations
Ressource en personnel	Evaluation	Importante	Moyenne	Moyen/bas	Moyen/bas
	Connaissance du domaine	Importante	Moyenne	Moyen/bas	Moyen/bas
Système existant	Oui	√	√	√	√
	Non	√		√	
Utilisateurs présents	Oui	√	√	√	√
	Non	√			

Tableau 3.4. *Choix d'une méthode artisanale*

Si, en revanche, l'évaluateur doit sélectionner une technique dans la classe des sciences appliquées avec, pour contraintes ou dimension d'évaluation : des ressources en expertise moyennes, un système déjà existant et des utilisateurs disponibles, alors le choix se portera sur l'expérimentation (tableau 3.5).

		SCIENCES APPLIQUEES				
DIMENSIONS DE L'EVALUATION		*Walkthrough* guides, méta-évaluation	Expérimentation	Observations	Questionnaires	Modèles de l'utilisateur
Ressource en expertise	Evaluation	Importante	Moyenne	Moyen/élevé	Moyen/élevé	Importante
	Connaissance du domaine	Importante	Moyenne	Moyen/élevé	Moyen/élevé	Importante
Système existant	Oui	√	√	√	√	√
	Non	√		√		
Utilisateurs présents	Oui	√	√	√	√	√
	Non	√				

Tableau 3.5. *Choix d'une méthode issue des sciences appliquées*

Cette procédure de sélection présente l'avantage d'être étayée par une approche rationnelle. Elle permet d'éviter les biais inhérents aux habitudes de travail ou à la formation suivie par les évaluateurs. En effet, certains évaluateurs peuvent avoir

tendance à privilégier un certain type de méthode sans tenir compte du contexte. Après cette phase de sélection de la méthode, il reste encore à la configurer.

3.4.2.3. *Phase 3 : configuration des méthodes*

Configurer la méthode signifie qu'il faut la faire correspondre aux critères de performance qui ont été établis auparavant. Concrètement, l'évaluateur se doit de construire une procédure qui soit adaptée à l'outil.

Selon Timmer [TIM 90], huit paramètres interviennent dans la configuration de la méthode :

1) le temps : on doit prendre en compte le temps disponible ;

2) le contexte : on doit prendre en considération les caractéristiques du contexte permettant d'améliorer les résultats de l'évaluation ;

3) les caractéristiques du client : le profil du financeur doit être défini afin de permettre d'optimiser les résultats (par exemple, quels sont ses a priori vis-à-vis de l'ergonomie ?) ;

4) l'accessibilité : quelles sont les possibilités d'accès aux utilisateurs ?

5) les ressources : par exemple, le matériel disponible pour les enregistrements vidéo ou audio ;

6) le recueil des données : quelle métrique va-t-on choisir ?

7) l'analyse des données : quels sont les moyens à disposition pour analyser les données (par exemple, logiciel de traitement statistique) ?

8) L'environnement physique : les caractéristiques de l'environnement doivent être prises en compte pour garantir le succès de l'évaluation.

En fonction de chacun de ces paramètres, Denley et Long proposent une série de règles de production provenant de leur recueil d'expertise. A titre d'exemples :

– R1 : si des ressources financières sont disponibles, alors proposer le recrutement d'un deuxième évaluateur, car cela évitera les biais dus au style d'un seul évaluateur ;

– R2 : si un sujet ne se sent pas capable de verbaliser, alors l'observer pendant qu'il parle avec un collègue, car la collaboration permet de réduire l'inhibition...

En appliquant l'ensemble de ces règles, on peut réussir à élaborer une procédure adéquate au contexte.

3.4.2.4. *Phase 4 : évaluation*

La méthode étant choisie et configurée, il ne reste plus alors qu'à pratiquer l'évaluation proprement dite. Nous ne nous attarderons pas sur ce point qui a déjà été suffisamment développé dans la partie consacrée à la description des méthodes. En

revanche, une fois l'évaluation réalisée, l'évaluateur doit, *a posteriori*, appréhender l'impact que les recommandations ont effectivement eu sur l'utilisateur et son activité. C'est dans cette perspective que Sawyer *et al.* [SAW 96] ont élaboré une méthode d'évaluation de l'impact. L'impact d'une action ergonomique peut succinctement se définir par le nombre de problèmes d'utilisabilité résolus. Face à la nécessité d'évaluer précisément cet impact sur le produit final, Sawyer *et al.* [SAW 96] ont créé une mesure nommée « ratio de l'impact » :

> Ratio de l'impact = (Nombre de problèmes effectivement résolus/Total des problèmes identifiés)*100

Cette mesure permet de comparer différentes méthodes et différents contextes d'évaluation. Sawyer *et al*. [SAW 96] ont d'ailleurs déterminé divers facteurs susceptibles d'influencer l'impact d'une action ergonomique :

– l'utilisation d'un rapport écrit peut avoir un effet positif s'il est communiqué à tous mais les recommandations ne sont pas forcément suivies ;

– des évaluations multiples (des méthodes telles que l'évaluation heuristique combinées à des tests utilisateurs peuvent améliorer l'impact) ;

– une réponse écrite du client permet d'identifier ceux qui sont susceptibles d'implanter les améliorations ;

– la facilitation du processus de réponse (par le biais de l'organisation de réunions) ;

– la formulation des recommandations doit être précise (1 ou 2 par problèmes à résoudre) ;

– la cotation des problèmes d'utilisabilité (en terme de « faible », « moyen » ou « important ») permet d'augmenter la crédibilité du rapport car celui-ci se rapproche des techniques employées par les ingénieurs ;

– l'implication le plus tôt possible dans le projet favorise l'augmentation de l'impact ;

– hormis ces facteurs, certaines variables demeurent incontrôlables. Par exemple le temps disponible pour implanter les changements ; ou une annulation pure et simple du projet.

Pour terminer, ces auteurs proposent d'autres recommandations pour améliorer l'impact :

– mieux connaître le produit (par exemple en lisant les modes d'emploi),

– prendre en compte le fait que le produit évalué fait partie ou non d'une gamme,

– utiliser des tests d'utilisabilité afin de faire prendre conscience des problèmes aux concepteurs,

– avoir dans l'équipe un informaticien pour éviter que la résolution des problèmes soit décrétée impossible à implémenter.

C'est en respectant ces quelques principes qu'on peut réussir à augmenter la portée des interventions ergonomiques.

3.4. Conclusion

Ce chapitre a cherché à répertorier, de la façon la plus exhaustive possible, les différentes méthodes d'évaluation des interfaces, méthodes devenues essentielles pour garantir la qualité ergonomique d'une interface. Nous nous sommes inspirés de la classification de Senach [SEN 93] pour développer une taxonomie qui distingue les méthodes empiriques, des méthodes analytiques et prédictives. Respectivement, les premières impliquent la présence d'utilisateurs pour procéder à une étude de l'utilisabilité, les deuxièmes permettent d'analyser des dimensions plus spécifiques de l'interface, alors que les dernières cherchent à formaliser et à prédire la qualité de l'interaction homme-machine.

Les méthodes empiriques sont indispensables, et en particulier dans le monde des aides techniques pour handicapés, puisque l'implication des utilisateurs dans le processus d'évaluation est une des conditions nécessaires, mais non suffisante, pour garantir l'adéquation de l'interface à leurs besoins moteurs, sensoriels, cognitifs et affectifs.

Ce chapitre a également présenté une méthodologie qui fourni une aide précieuse aux évaluateurs inexpérimentés afin de choisir le mode d'évaluation le plus approprié selon la situation [DEN 97]. Il est ainsi possible, à partir de guides de recueils de données et de tableaux synthétiques, de déterminer relativement facilement pour chaque phase de la conception l'outil qui convient le mieux.

Enfin, nous souhaitons insister sur un dernier point trop rarement abordé dans l'évaluation. Il s'agit de l'aspect financier, qui sont souvent opposés aux recherches menées sur les aides techniques pour handicapés : le coût des recherches s'avère élevé et le nombre d'utilisateur très faible. A cet argument nous pourrions bien évidemment souligné que la vie humaine n'a pas de prix, et que les AT ne font pas que simplifier la vie des personnes handicapées ; mais elles les rendent possibles ! Mais répondons simplement à cette question en utilisant les mêmes arguments financiers.

Depuis le début des années 1990 sont apparues plusieurs estimations des coûts fondées sur l'observation d'évaluations réalisées dans le domaine industriel et portant sur des produits informatiques à usage domestique [Bias & Mayhew, 1993, NIE 94]. Sur la base des valeurs publiées à partir de plusieurs projets, Nielsen [NIE 94] fait état d'un coût fixe pour une évaluation estimée entre 3 700 et 4 800 dollars et entre 410 et 900 dollars en coût variable pour chaque évaluateur. Ces

valeurs dépendent bien évidemment de la complexité du produit et de la structure de coût de l'entreprise. Le bénéfice, quant à lui, est estimé par l'accroissement de la productivité de l'utilisateur et par une satisfaction accrue du client (réduction du temps d'apprentissage, convivialité, facilité d'utilisation) générant un accroissement du volume de vente du produit. Sur ces exemples de projets, le bénéfice a été estimé à 15 000 dollars pour chaque problème d'utilisabilité détecté. En faisant le produit du nombre de problèmes détectés en fonction du nombre d'évaluateurs par le bénéfice estimé généré par chaque problème détecté il est possible d'en calculer le fructueux bénéfice total… Mais dans cette logique économique quelle place pour l'évaluation des aides techniques pour handicapés, qui sont d'abord des produits spécifiques, voire conçus sur mesure, et rarement promis à une diffusion massive ? Dans un monde, qui malheureusement, fait encore trop souvent dominer la technologie sur l'homme, et l'économie sur le handicap, ceci tiendra lieu de remarque finale.

3.5. Bibliographie

[AND 97] ANDRIESSEN J.H.E., ARNOLD B., « Evaluation-Methodology for Telematic Application Systems: Quality for Users and Context », *CHI 97 Electronic Publications: Organizational Overviews*, http://www.acm.org/sigchi/chi97/proceedings/tutorial/gp.htm, 1997.

[APE 85] APERLEY M.D., FIELD G.E., « A comparative evaluation of menu-based interactive human computer interaction dialogue technique », dans B. Schackel (dir.), *Proceedings of human Computer Interaction, Interact'84*, Elsevier Science Publishers, North-Holland, p. 296-300, 1985.

[ASS 97] Association nationale pour l'amélioration des conditions de travail (ANACT), « Les incontournables en ergonomie », http://www.anact,fr/publications/index.html, 1997.

[BAR 95] BARTHE M., *Ergonomie des logiciels : une nouvelle approche des méthodologies d'informatisation*, Masson, Paris, 1995.

[BAS 93] BASTIEN J.M., SCAPIN D.L., Ergonomic Criteria for the Evaluation of human-Computer Interfaces, Rapport technique n° 156, INRIA, Rocquencourt1, 1993.

[BAS 94] BASTIEN J.M., SCAPIN, D.L., Evaluating a user interface with ergonomic criteria, Rapport de recherche n° 2326, INRIA, 1994.

[BAS 98] BASTIEN J.M.C., LEULIEUR C., SCAPIN D.L., « L'ergonomie des sites web », dans J.-C. Le Moal, B. Hidoine (dir.), *Créer et maintenir un service Web,* Cours INRIA, Pau, p. 111-173, 1998.

[BAU 96] BAUDEL T., « Limites de l'évaluation ergonomique pour la conception de nouveaux paradigmes d'interaction », *Actes du Colloque des huitièmes journées sur l'interaction Homme/Machine* (IHM'96), Grenoble, 1996.

[BEN 95] BENDER J., DE HAAN J., BENNETT D., *The symbiosis of work and technology*, Taylor & Francis, Londres, 1995.

[BRA 00] BRANGIER E., Psychologie ergonomique de l'assistance à l'opérateur. Approche symbiotique de la relation homme-technologie-organisation. Habilitation à Diriger des Recherches, Université Paris 5, 2000.

[BRA 02] BRANGIER E., VALLERY G., « Aspects psychologiques et organisationnels du développement des nouvelles technologies de la communication et de l'information », dans E. Brangier, C. Louche, A. Lancry (dir.), *Psychologie du travail et des organisations*, PUN, Nancy, 2002.

[BRA 90] BRANGIER E., « Ergonomie des logiciels : approche psycho-ergonomique de l'interaction homme-ordinateur », *Cahiers de notes documentaires INRS*, 139, 2, p. 391-404, 1990.

[BRA 91] BRANGIER E., « Comment les recherches concernant les dialogues homme-machine abordent-elles le problème de l'interaction ? », *Connexion*, n° 57, vol. 1, 149-161, 1991.

[CAE 96] CAELEN J., « Définition et caractéristiques des interfaces homme machine », dans *Observatoire Français des Techniques Avancées, Nouvelles Interfaces homme-machine* (ARAGO 18), p. 31-44, OFTA, Paris, 1996.

[CAR 83] CARD S.K., MORAN T.P., NEWELL A., *The psychology of human-computer interaction*, LEA, Hillsdale, New Jersey, 1983.

[CHR 95] CHRISTIE B., SCANE R., COLLYER J., « Evaluation of human-computer interaction at the user interface to advanced I.T. systems », dans J.R. Wilson, E.N. Corlett (dir.), *Evaluation of human work*, Taylor and Francis, Londres, p. 310-356, 1995.

[COU 90] COUTAZ J., *Interfaces homme-ordinateur*, Dunod, Paris, 1990.

[COU 96] COUTAZ J., « Ingénierie de l'interaction homme-machine », dans *Observatoire Français des Techniques Avancées, Nouvelles Interfaces homme-machine* (ARAGO 18) OFTA, Paris, p. 163-174, 1996.

[DEN 97] DENLEY I., LONG J., « A planning aid for human factors evaluation practice », *Behaviour and information technology*, 16, p. 203-219, 1997.

[DOW 89] DOWELL J., LONG J.B., « Towards a conception for an engineering discipline of human factors », *Ergonomics*, 32, p. 1513-1535, 1989.

[DUB 00] DUBOIS J.M., DAO-DUY J.M., ELDIKA S., « L'analyse des traces informatiques des usages : un outil pour valider la conception d'un site web », *Actes du colloque RJC'IHM*, Ile de Berder, p. 85-89, 2000.

[DUB 93] DUMAS J.S., REDISCH J.C., *A practical Guide to usability testing*, Albex Publishing, Norwood, New jersey,1993.

[DUV 00] DUVALLET C., BOUKACHOUR H., « Modélisation par agents d'une interface intelligente », *Actes du colloque RJC'IHM*, Ile de Berder, p. 91-94, 2000.

[ELL 99] ELLA A.N., KOLSKI C., JACQUES C., YIM P., « Vers un système d'aide à la décision pour l'évaluation des systèmes interactifs », *Actes du Colloque IHM'99*, Montpellier, tome I, p. 126-131, 1999.

[GOO 97] GOOD M., SPINE T.M., WHITESIDE J., GEORGE P., « User derived impact analysis as a tool for usability engineering », dans M. Mantei, P. Orbeton (dir.), *Human Factors in Computing Systems-III*, A.C.M, Amsterdam, p. 241-246, 1986.

[GRO 90] GROSJEAN V., JAVAUX D., « L'évaluation ergonomique des logiciels : une approche décompositionnelle », *Technologie de l'informatique et Société*, 3, p. 125-145, 1990.

[HIL 91] HILL B., DENLEY I., LONG J.B., « Towards an evaluation planning aid: classifying and selecting evaluation methods », dans E.J. Lovesey (dir.), *Contemporary Ergonomics 1991*, Taylor and Francis, Londres, p. 355-362, 1991.

[INN 91] INNOCENT P.R., « Usability Testing: A Practical Use for Mental Model Construction », dans M.J. Tauber, D. Ackermann (dir.), *Mental Models and Human-Computer Interaction 2*, North-Holland, Pays-Bas, p. 271-288, 1991.

[ISO 97] ISO, International Standards Organisation DIS 9241-11, Ergonomic requirements for office work with visual display terminals (VDTs), Part 11: Guidance on usability, 1997.

[JAM 99] JAMBON F., BOISDRON Y., « Vers une validation du dialogue homme-machine pour le application graphiques interactives de conception technique », *Actes du Colloque IHM'99*, Montpellier, tome I, p. 1-4, 1999.

[JOH 97] JOHN B.E., MARKS S.J., « Tracking the effectiveness of usability evaluation methods », *Behaviour and information technology*, 16, p. 188-202, 1997.

[KAL 99] KALAWSKY R., BEE S., NEE S., « Human factors evaluation techniques to aid understanding of virtual interfaces », *BT Technology Journal*, 17(1), p. 128-141, 1999.

[KAR 88] KARAT J., « Software evaluation methodologies », dans M. Helander (dir.) *Handbook of human-computer interaction*, Elsevier Science Publishers, Amsterdam, p. 891-903, 1988.

[KAR 92] KARAT C., CAMPBELL R., FIEGEL T., *Comparison of empirical testing and walkthrough methods in user interface evaluation*, Paper presented at CHI'92: Human Factors in Computing Systems, A.C.M., New York, p. 397-404, 1992.

[KAR 95] KARAT J., DAYTON T., « Practical Education for Impoving Software Usability », dans I.R. Katz, R. Mack, L. Marks (dir.), *Human Factors in Computing Systems CHI95 Conference Proceedings*, A.C.M., New York, p. 162-169, 1995.

[KIE 85] KIERAS D.E., POLSON P.G., « An approach to the formal analysis of the user compexity », *Int. J. Man-Machine Studies*, 22, p. 365-394, 1985.

[KIR 98] KIRAKOWSKI J., CIERLIK B., « Mesasuring the usability of web sites », *Preprint of paper for Human Factors and Ergonomics Society Annual Conference*, Chicago, 1998.

[LAM 90] LAMBERT J., COLIN B., JEFFROY F., HARADJDJI Y., « Articulation entre différentes méthodes d'analyse en ergonomie du logiciel », *Actes du 26e Congrès de la SELF*, Montréal, 1990.

[LEU 99] LEULIEU C., BASTIEN C.., SCAPIN D.L., « L'évaluation des sites Web : apports respectifs des questionnaires de satisfaction et l'inspection ergonomique », *Actes du Colloque IHM'99*, Montpellier, tome II, p. 49-53, 1999.

[LIN 97] LIN H.Y., CHOONG Y.Y., SALVENDY G., « A proposed Index of usability: a method for comparing the relative usability of different software systems », *Behaviour and information technology*, 16, p. 267-278, 1997.

[LON 89] LONG J.B., DOWELL J., « Conceptions of the discipline of H.C.I.: Craft, Applied Science and Engineering », *Proceedings of the fifth conference of the BCS HCI SIG*, p. 9-32, Cambridge University Press, Cambridge, 1989.

[MAC 94] MACK R., NIELSEN J., « Executive summary », dans J. Nielsen et R. Mack (dir.), *Usability inspection methods*, p. 1-24) , John Wiley and Sons, New York, 1994.

[MAC 95] MAC CLELLAND, « Product assessement and user trials », *Evaluation of Human Work*, p. 249-285, Taylor and Francis, Londres, 1995.

[MAC 97] MAC LEOD M., BOWDEN R., BEVAN N., CURSON I., « The MUSIC performance mesurement method », *Behaviour and information technology*, 16, p. 279-293, 1997.

[MEI 91] MEINADIER J.P., *L'interface utilisateur : pour une informatique plus conviviale*, Dunod, Paris, 1991.

[MON 87] MONK A.F., DIX A., « Refining early design decisions with a black box model », dans D. Diaper, R. Winder (dir.), *People and Computer III*, p. 147-158, Cambridge University Press, Grande-Bretagne, 1987.

[MON 91] MONTERO S., Prototypage : méthodes et outils d'aide à la conception et à la réalisation des systèmes d'information, Thèse de doctorat, INSA, Lyon, 1991.

[MOR 81] MORAN T.P., « The Command Language Grammar: A representation of the user interface of interactive computer systems », *Int. J. Man-Machines Studies*, 15, 3-50, 1981.

[NAE 96] NAËL M., « L'évaluation des nouvelles interfaces homme-machine interactives », dans *Observatoire Français des Techniques Avancées*, Nouvelles Interfaces homme-machine (ARAGO 18) (p. 323-340). Paris : OFTA. (1996).

[NIE 92] NIELSEN J., « Evaluating the thinking aloud technique for use by computer scientists », dans H. Hartson, D. Hix (dir.), *Advances in human-computer interaction*, volume 3, p. 69-82, Ablex Publishing Corp. , Norwood, 1992.

[NIE 94] NIELSEN J., « Heuristic evaluation », dans J. Nielsen et R. Mack (dir.), *Usability inspection methods*, p. 25-62, John Wiley and Sons, New York, 1994.

[NOR 86] NORMAN D.A., « Cognitive Engineering », dans D.A. Norman, S.W. Draper (dir.), *User Centered system design*, p. 31-61, Lawrence Erlbaum associates, Hillsdale, New-Jersey, 1986.

[NOW 97] Nowicki J., Henry S.L., « Usability Services at Compuware-Madison: Bringing Usability to Data Proceeding », dans *CHI 97 Electronic Publications : Organizational Overviews, World Wide Web*, http//www.acm.org/sigchi/chi97/proceedings/tutorial/gp.htm, 1997.

[OPP 97] OPPERMAN R., REITERER H., « Software evaluation using the 9241 evaluater », *Behaviour and information technology*, 16, p. 232-245, 1997.

[PER 96] PERLMAN G., « Curricula for Human-Computer Interaction », *World Wide Web*, http//www.acm.org/sigchi/cdg, 1996.

[PER 97] PERLMAN G., « Practical Usability Evaluation », dans *CHI 97 Electronic Publications : Organizational Overviews. World Wide Web*, http//www.acm.org/sigchi /chi97/proceedings/tutorial/gp.htm, 1997.

[PER 98] PERLMAN G., « Practical Usability Evaluation », dans *Professional Development Workshop, World Wide Web*, http//www.acm.org/chapters/buckchi/workshop.html, 1998.

[RAB 95] RABARDEL P., *Les hommes et les technologies : approche cognitive des instruments contemporains*, Armand Colin, Paris, 1995.

[RAU 92] RAUTERBERG M., « A method of a quantitative measurement of cognitive complexity », dans G. Van der Veer *et al.* (dir.), *Human-Computer interaction: tasks and organisation*, p. 295-307, CUD Publ., Rome, 1992.

[RAU 94] RAUTERBERG M., « Usability Engineering Now! », *Computer information*, 7-8, p. 57-61, 1994.

[RAU 97] RAUTERBERG M., « About a method to mesure ergonomic quality of user interfaces in a task independant way », dans *Workshop on emerging technologies in human engineering testing and evaluation*, Nato Research Study Group 24, Bruxelles, 1997.

[REI 81] REISNER P., « Formal grammar and human factors design of an intercative graphics System », *IEEE Transactions on Software Engineering*, SE.7, 2., 1981.

[SAW 96] SAWYER P., FLANDERS A., WIXON D., « Making a difference: the impact of inspections », dans *Proceedings of the 1996 ACM conference on human factors in computing systems*, ACM Press, Vancouver, 1996.

[SCA 93] SCAPIN D.L., « Situation et perspectives en ergonomie du logiciel », dans J.C. Sperandio (dir.), *L'ergonomie dans la conception des projets informatiques*, p. 7-68, Octarès, Toulouse, 1993.

[SCH 97] SCHERER M.J., GALVIN J.C., « Assistive technology », dans Shrawan Kumar (dir.), *Perspectives in Rehabilitation ergonomics*, Taylor and Francis, 1997.

[SEN 93] SENACH B., « Evaluation ergonomique des interfaces hommes-machines : une revue de la littérature », dans J.C. Sperandio (dir.), *L'ergonomie dans la conception des projets informatiques*, p. 69-123, Octarès, Toulouse , 1993.

[SPE 93] SPERANDIO J.C., « Introduction », dans J.C. Sperandio (dir.), *L'ergonomie dans la conception des projets informatiques*, p. 1-6, Octarès, Toulouse 1993.

[TIM 90] TIMMER P., *The elicitation, representation and structuring of generic and specific control knowledge for configuration of evaluation methods*, University College, Londres, 1990.

[VAL 93] VALENTIN A., VALLERY G., LUCONGSANG R., *L'évaluation ergonomique des logiciels. Une démarche itérative de conception*, ANACT, Montrouge, 1993.

[VAN 91] VAN DER VELDEN J.M., ARNOLD A., « Mental Models and the Evaluation of User Interfaces: A Case-Study of a Library System », dans M.J. Tauber, D. Ackermann (dir.), *Mental Models and Human-Computer Interaction 2*, p. 179-189, North-Holland, Pays-Bas, 1991.

[VAN 94] VANDERDONCKT J., *Guide ergonomique des interfaces homme-machine*, Presses Universitaires de Namur, Namur, 1994.

[VIR 92] VIRZI R., « Refining the test phase of usability evaluation: how many subjects is enough? », *Human Factors*, 34, 4, p. 457-468, 1992.

[VIR 96] VIRZI R.A., SOKOLOV J.L., KARIS D., « Usability problem identification using both low and high fidelity prototypes », dans *Proceedings of the 1996 A.C.M. conference on human factors in computing systems*, A.C.M. Press, Vancouver, 1996.

[WIL 87] WILLIGES R.C., WILLIGES B.H., ELKERTON J., « Software interface design », dans *Evaluation of Human Work*, Taylor and Francis, Londres, 1987.

[WIL 92] WILLIGES R.C., WILLIGES B.H., HAN S.H., « Developing quantitaive guidelines integrated data from sequential experiments », *Human Factors*, 34, p. 399-408, 1992.

[WIL 93] WILLIGES R.C., WILLIGES B.H., HAN S.H., « Sequential experimentation in human-computer interface design », dans H.R. Hartson, D. Hix (dir.*), Advances in human-interaction*, volume 4, p. 1-30, Ablex Publishing Corp, Norwood, 1993.

[WIL 95] WILSON J.R., RAJAN J.A., « Human-machine interfaces for systems control », dans *Evaluation of Human Work*, p. 357-406, Taylor and Francis, Londres, 1995.

[WIL 96] WILLIGES R.C., « Usability Methods Research Laboratory (U.M.R.L.) », *World Wide Web*, http://www.hci.ise.vt.edu/hfec/hci.html, 1996.

[WIT 91] WITHEFIELD A., WISON F., DEWELL J., « A framework for human factors evaluation », *Behaviour and information technology*, 10, p. 65-79, 1991.