



Cognitive Walkthrough pour l'évaluation des IHM : synthèse des extensions et évolutions conceptuelles, méthodologiques et technologiques

Thomas Mahatody, Mouldi Sagar, Christophe Kolski

► To cite this version:

Thomas Mahatody, Mouldi Sagar, Christophe Kolski. Cognitive Walkthrough pour l'évaluation des IHM : synthèse des extensions et évolutions conceptuelles, méthodologiques et technologiques. 19e Conférence de l'Association Francophone d'Interaction Homme-Machine (IHM 2007), Nov 2007, Paris, France. pp.143-150, 10.1145/1541436.1541465 . hal-03451162

HAL Id: hal-03451162

<https://uphf.hal.science/hal-03451162>

Submitted on 10 Sep 2022

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Cognitive Walkthrough pour l'évaluation des IHM : synthèse des extensions et évolutions conceptuelles, méthodologiques et technologiques

Thomas Mahatody, Mouldi Sagar, Christophe Kolski
LAMIH – UMR CNRS 8530, Le Mont Houy, F-59313 Valenciennes cedex 9, France
{Thomas.Mahatody, Mouldi.Sagar, Christophe.Kolski}@univ-valenciennes.fr

RESUME

Cet article de synthèse concerne l'évaluation des systèmes interactifs sous l'angle des méthodes d'inspection et plus particulièrement de la méthode Cognitive Walkthrough (CW). Les principes de base de CW sont abordés, ainsi que son évolution et ses extensions. Une synthèse est faite selon les points de vue conceptuel, méthodologique et technologique.

MOTS CLES : Evaluation des systèmes interactifs, Ergonomie des IHM, Cognitive Walkthrough (CW), Méthode d'inspection.

ABSTRACT

This synthesis paper concerns the evaluation of interactive systems using usability inspection methods, especially the Cognitive Walkthrough (CW) method. The background principles of CW followed by its evolution and variants are described. A synthesis from conceptual, methodological and technological points of views is proposed.

KEYWORDS: Interactive systems evaluation, Software ergonomics, Cognitive Walkthrough, Inspection method.

INTRODUCTION

Le développement d'un système interactif est un processus itératif ; l'évaluation est un composant important de ce processus et fait l'objet de nombreuses études et recherches depuis de nombreuses années par la communauté en IHM. Il existe aujourd'hui de nombreuses méthodes et techniques d'évaluation ; plusieurs études ont été faites proposant des revues et synthèses de celles-ci. Citons par exemple : [28,39]. L'inspection de l'utilisabilité est un nom générique d'un ensemble de

méthodes basées sur le fait que un ou plusieurs évaluateurs inspectent ou examinent les aspects de l'utilisabilité de l'interface utilisateur. Selon [32], l'utilisabilité est un concept assez large qui se rapporte à l'apprentissage et à l'utilisation du système par l'utilisateur, de même qu'à la satisfaction de l'utilisateur utilisant le système. Cognitive Walkthrough (CW) fait partie des méthodes d'inspection de l'utilisabilité. C'est une méthode qui associe le parcours (walkthrough) de l'interface avec un modèle cognitif de l'apprentissage par l'exploration. L'évaluateur parcourt l'IHM dans le contexte des tâches qu'un utilisateur typique aura besoin d'accomplir. Les actions et les réponses de l'interface sont comparées aux buts et connaissances de l'utilisateur, et les écarts entre les attentes de l'utilisateur et les étapes requises par l'interface sont notés. Comme les autres méthodes d'évaluation les plus courantes en IHM, CW met l'accent sur les principes de base de l'utilisabilité, mais à la différence des autres, CW se concentre aussi sur les activités cognitives de l'utilisateur, spécialement sur les buts et les connaissances de l'utilisateur exécutant une tâche spécifique. CW a été proposée par Lewis *et al.* [24]. Mais des modifications ont été faites ensuite et des extensions ont été aussi proposées par d'autres auteurs. Nous décrivons dans un premier temps l'évolution de la méthode initiale (trois versions). Ensuite nous passons en revue dix extensions significatives. Une synthèse est proposée en fin d'article.

EVOLUTION DES VERSIONS ORIGINALES

La **première version** [24], destinée à l'évaluation des interfaces de type « walk-up-and-use », c.-à-d. qu'on utilise sans ou avec peu de formation initiale, est basée sur la théorie de l'apprentissage par exploration CE+ [37]. CE+ est une théorie cognitive d'apprentissage de l'utilisation pour la première fois d'une IHM. Des guides de conception appelés « Design Principles for Successful Guessing » sont dérivés de cette théorie pour supporter la conception de système interactif nécessitant peu ou éventuellement pas de formation utilisateur. CE+ est composé de 3 composants principaux (résolution de problème, apprentissage, exécution) et fonctionne comme suit : l'utilisateur du système choisit une action parmi plusieurs alternatives possibles en se référant à la similarité entre son but et la conséquence attendue de l'action ; après l'exécution de l'action sélectionnée, l'utilisateur évalue la réponse du système en utilisant les heuristiques proposées par Lewis [25,26] ; c'est ainsi qu'il évalue sa

progression vers la réalisation de son but. S'il est réalisé, l'apprentissage se produit en mémorisant l'étape qui vient d'être évaluée sous forme de règle au format CCT [18]. Dans le cas contraire, le composant de résolution de problème est invoqué pour découvrir une action appropriée et ainsi de suite. L'évaluation se fait en 2 phases (préparation et évaluation) : d'abord, l'évaluateur ou groupe d'évaluateur spécifie une série de tâches à évaluer et chaque tâche est décomposée en séquence d'actions. L'inspection de chaque action est abordée en répondant aux questions relatives au processus de simulation du modèle CE+ : choix et exécution de l'action puis évaluation de la réponse du système.

La **seconde version** est basée sur une extension du modèle CE+ [37]. Ce nouveau modèle est lié à la théorie de l'action de Norman [34] et au modèle de construction-intégration de Kintsch [19]. La théorie de Norman dit que la réalisation d'une tâche met en jeu sept activités : l'établissement d'un but, la formulation d'une intention, la spécification d'une suite d'actions, l'exécution des actions, la perception et l'interprétation de l'état du système et enfin l'évaluation de l'état du système par rapport au but considéré. A l'instar de la première version, celle-ci s'exécute en deux phases : préparation et évaluation. La phase d'évaluation est guidée par trois formulaires. Les éléments des formulaires traitent les problèmes et les échecs qui peuvent se produire lors du processus d'exploration du modèle. Le premier formulaire traite de la relation entre les buts nécessaires pour manipuler l'interface et les buts courant de l'utilisateur. Le deuxième s'occupe des problèmes liés à la sélection de l'action en supposant que les buts sont appropriés. Et enfin le troisième formulaire concerne l'évolution des buts de l'utilisateur en prenant en compte la réponse du système après l'exécution de l'action.

Dans la **troisième version** [27,44], l'accent est mis sur la motivation de l'utilisateur à choisir la bonne action. L'évaluateur est invité à relater pourquoi l'utilisateur choisira l'action ou pourquoi il ne la choisira pas. En définitive le choix de l'utilisateur est motivé par les réponses aux quatre questions suivantes : Est-ce que l'utilisateur tente de produire ce que l'action doit produire ? Est-ce que l'utilisateur s'aperçoit que l'action correcte est disponible ? Est-ce que l'utilisateur associe l'action correcte avec l'effet désiré ? Si l'action correcte est exécutée, est-ce que l'utilisateur voit qu'il progresse vers la solution ? (Pour plus de détails, voir [23]).

EXTENSIONS SIGNIFICATIVES DE CW

Dans cette partie, nous passons en revue chronologiquement dix extensions de CW. Le niveau de détail dépend de l'originalité de l'extension en question, tout en tenant compte de la place disponible.

SHIVA [46] (Structured Human Interface VALidation Technique) est une méthode de développement et d'évaluation d'IHM orientée tâche et objet. Selon ces

auteurs, pour avoir une évaluation efficace, il est nécessaire de construire une méthode d'évaluation liée à la méthode de développement en question. Bien qu'elle se démarque des extensions qui suivent du point de vue conceptuel (dans la mesure où SHIVA ne se base pas sur un modèle cognitif explicite), cette méthode d'inspection nous semble intéressante : en effet, elle est basée sur une méthodologie associant une modélisation des tâches et une modélisation de la navigation dans le système [2], et on utilise ce modèle de navigation pour évaluer l'utilisabilité. L'évaluation se déroule en 2 cycles. Dans le premier, on inspecte les différentes vues du système en répondant à un ensemble de questions dérivées des principes d'utilisabilité précisées dans la norme ISO 9241-10 [12]. Dans le second, on inspecte les scénarios issus de l'analyse de tâches en analysant la séquence de vues nécessaire pour exécuter chaque scénario. Les vues nécessaires pour exécuter chaque tâche simple sont ainsi identifiées. Pour chaque étape l'évaluateur répond à un ensemble de questions concernant le changement possible de buts, en mettant en évidence les problèmes. Les résultats obtenus lors des deux cycles doivent être intégrés dans des tableaux montrant quelles vues sont nécessaires pour quelles tâches pour obtenir une image cohérente de l'utilisabilité du système.

Heuristic Walkthrough est une méthode d'évaluation combinant les avantages des trois méthodes : Heuristic Evaluation [31,33], CW et Usability Walkthrough [15]. Pour Heuristic Walkthrough, l'évaluation se fait en deux phases. La phase 1 est une évaluation guidée par une liste priorisée de tâches, et une liste de questions provenant de CW. Pendant la phase 2, l'évaluateur est libre d'explorer n'importe quel aspect du système en l'utilisant. Néanmoins, il est guidé par la connaissance obtenue dans la phase 1, la liste des tâches, une liste d'heuristiques (par exemple celles fournies dans [31,33]) et la liste de questions utilisées dans la phase 1.

Norman's CW est certes basé sur le modèle de Norman [34], mais celui-ci est modifié par [40] selon la Figure 1. Selon Hutchins *et al.* [11], une distance cognitive indique la quantité et la qualité de traitement d'information nécessaire pour remplir l'espace entre les états (intention, évaluation, etc.). La notion de distance cognitive peut être à la fois liée à l'exécution de l'action et aussi à l'évaluation du résultat. Dans le premier cas elle se rapporte à la quantité de traitement de l'information nécessaire pour réduire le fossé entre l'intention et l'action physique sur laquelle l'intention est communiquée au système. En d'autres termes, elle se rapporte à l'acte de traduction des pensées et des buts de l'utilisateur dans le langage du système. Dans le dernier cas, la distance cognitive se rapporte à l'effort mental nécessaire à la traduction de l'information affichée par le système dans les termes du modèle conceptuel adopté par l'utilisateur. L'évaluation consiste à mettre à l'épreuve les activités de l'utilisateur exécutant une ou plusieurs tâches à l'intérieur d'un scénario donné.

L'évaluateur explore le système en cherchant les actions qui doivent contribuer à l'exécution de la tâche. Il choisit des actions dont la description correspond avec ce qu'il essaie de faire. Ensuite il interprète la réponse du système en vue d'évaluer si on progresse vers l'achèvement de la tâche ou s'il faut reconsidérer le but. Ce qui permet donc d'identifier d'une part si la signification et la forme de l'interface sont bien interprétées par l'utilisateur et d'autre part s'il est capable de fixer un but faisable et d'exécuter l'action correcte sur l'objet adéquat. A chaque étape de l'interaction, des questions doivent être renseignées (Cf. la localisation des questions Q1 à Q6 sur la figure 1).

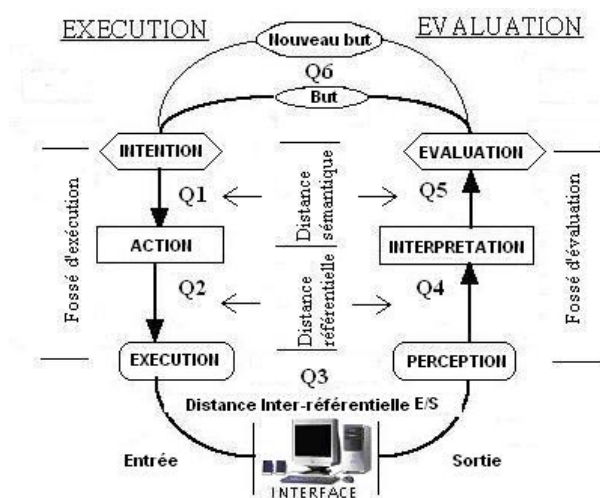


Figure 1 : Modèle de Norman modifié

Streamlined CW est une variante de la 3ème version de CW, et comporte 5 phases. La première est celle de définition de l'entrée du parcours d'inspection. La deuxième consiste en une réunion des évaluateurs, pour définir ce qu'il faudra faire et ce qu'il ne faudra pas faire pendant l'évaluation, ce qu'il faudra éviter pour que les concepteurs ne soient pas sur la défensive vis-à-vis de leur conception, pour définir le rôle de chaque membre de l'équipe. Pendant la phase d'inspection (phase 3), l'évaluateur procède de la même manière que dans CW en répondant aux deux questions suivantes : (1) Est-ce que l'utilisateur sait ce qu'il doit faire à cette étape ? (2) Si l'utilisateur réalise correctement l'action, est-ce qu'il saura qu'il a réalisé la bonne action et qu'il progresse vers le but en question ? La quatrième phase consiste à enregistrer les problèmes détectés. La dernière consiste à les corriger. Cette méthode insiste beaucoup sur la rigueur au niveau de la conduite de l'évaluation.

CW for the Web [4] (CWW) est une extension basée sur une théorie appelée CoLiDeS (Comprehension-based Linked model of Deliberate Search) [21]. CoLiDeS est en fait une extension du modèle LICA (Linked model of Comprehension-based Action planning and Instruction taking), modèle d'exploration basée sur la compréhension [20]. CWW permet de simuler un utilisateur naviguant sur une page avec un but en tête.

L'exploration consiste à choisir une action (clic sur une icône, un lien, etc.) puis à évaluer le résultat en fonction du but. CoLiDeS suppose que la sélection d'une action est un processus à deux phases : d'intention, de sélection. Pendant la première, l'utilisateur segmente la page en un ensemble de zones et génère une brève description de chaque zone à partir de sa connaissance de la convention de structuration de page-écran. L'utilisateur s'occupe de la zone dont la description est perçue comme similaire à son but courant. Pendant la phase de sélection de l'action, l'utilisateur génère les descriptions de tous les outils graphiques dans la zone concernée et agit sur celui dont la description est la plus proche de son but. Les questions d'évaluation sont les suivantes : Est-ce que l'action correcte est suffisamment mise en évidence pour l'utilisateur ? Est-ce que l'utilisateur relie la zone correcte de la page avec le but et ceci en utilisant l'information sur le titre et la compréhension de la convention de structuration de la page ? Est-ce que l'utilisateur relie le but avec les outils graphiques corrects dans la zone en question et ceci en utilisant des étiquettes des liens et d'autres informations descriptives ? Est-ce que l'utilisateur interprète correctement la réponse du système avec l'action choisie ? CWW transforme cette approche en utilisant l'Analyse Sémantique Latente (ASL) [22] : au lieu de faire juger subjectivement par des évaluateurs les réponses à ces questions, on utilise l'ASL permettant d'estimer la liaison sémantique des textes basés sur l'analyse statistique d'un large corpus. Donc ici, l'ASL est utilisée pour estimer la similarité sémantique entre les descriptions des buts de l'utilisateur et les descriptions des zones des pages ainsi que les buts et les descriptions des actions possibles sur la page web. L'évaluation se déroule en 4 phases. Dans la phase 1 on formule un ensemble de buts utilisateurs réalistes composé de 100 à 200 mots, puis on identifie la sélection correcte prévue pour chaque but. La phase 2 consiste à calculer les similarités sémantiques entre les buts, les titres et les étiquettes des liens. La phase 3 a pour objet l'identification des titres et étiquettes de liens problématiques en supposant que selon CoLiDeS : un titre cause un problème s'il n'est pas familier, ou s'il est en confusion avec d'autres. Un titre avec une valeur ASL de moins de 0.8 est supposé non familier. Une paire de titres ayant une valeur ASL de plus de 0.6 est supposée confuse. La phase 4 consiste à rechercher les problèmes spécifiques au but.

Groupeware Walkthrough (GW) [35], méthode d'inspection pour Groupeware, est une modification profonde de CW visant à inclure des facteurs de complexité liés au fonctionnement d'une équipe de travail. GW intègre une modélisation de la tâche et un processus de type walkthrough. Un modèle de tâche permet d'identifier et analyser les tâches collaboratives du monde réel. Le modèle de groupe de tâches doit permettre d'exprimer la variabilité des actions dans le groupe de travail. Un processus Walkthrough vise à évaluer le support du système vis-à-

vis des tâches qui sont modélisées comme un ensemble de chemins possibles pour aboutir à un résultat voulu. Chaque chemin possible est exploré pendant le walk-through, et le support pour chaque chemin dans l'interface est évalué. La collaboration implique deux types distincts d'activités : les activité-tâches (actions qui doivent se produire pour accomplir la tâche), les activités du groupe (actions que les membres du groupe doivent exécuter pour accomplir ensemble une tâche). Notons que l'approche traditionnelle d'évaluation se focalise sur l'activité tâche. L'activité de groupe peut être analysée suivant une abstraction de haut niveau et une de bas niveau. Le haut niveau incluant les aspects sociaux et organisationnels est traité en s'appuyant sur GTA (Groupware Task Analysis) [45]. GTA est un moyen efficace pour rassembler des informations sur le contexte de travail, mais son niveau d'analyse est trop élevé pour être utilisé dans une situation d'inspection. Pour capturer le niveau des détails nécessaires pour l'inspection, il faut considérer l'activité du groupe au plus bas niveau : celui des processus de collaboration. Néanmoins l'analyse de haut niveau précède celle de bas niveau. Les processus de collaboration [9] sont les activités de base du travail collaboratif, les actions et interactions à petite échelle que les membres du groupe doivent effectuer pour réaliser une tâche de façon collaborative. Les composants principaux d'un modèle de tâches sont les scénarios, tâches, sous-tâches (individuelles ou collaboratives), actions. Le scénario est une description de haut niveau des activités liées à l'accomplissement d'un résultat spécifique. L'analyse de tâche commence par la collecte des données observées sur le travail, et continue avec l'identification des épisodes d'interaction collaborative. Les scénarios et les tâches peuvent donc être spécifiés et les sous-tâches collaboratives peuvent être analysées selon le processus de collaboration. Pour chaque scénario collaboratif, GW peut ensuite être exécutée de la façon suivante : (1) se familiariser avec les contextes, l'environnement, (2) pour chaque tâche du scénario, essayer d'exécuter chaque tâche alternative, enregistrer comment chaque tâche est exécutée, enregistrer les problèmes détectés, après l'exécution de chaque tâche, se poser un ensemble de questions (est-ce que la tâche peut être exécutée efficacement, est-ce que l'interface fournit le moyen de le faire et correctement ? Est-ce que la tâche peut être effectuée de façon efficiente ? Est-ce que la tâche est exécutée avec satisfaction ?), (3) après l'exécution de toutes les tâches, il faut vérifier si l'IHM permet au groupe de réaliser toutes les tâches entendues.

Activity Walkthrough [3] est une version modifiée de CW, basée sur la théorie de l'activité, outil descriptif puissant qui d'une part s'intéresse à la compréhension de l'activité humaine et d'autre part incorpore les notions d'intentionnalité, d'histoire, de médiation, de collaboration et de développement [30]. Selon cette théorie, l'unité de l'analyse est l'activité entière. Une activité est composée d'un sujet, d'un objet et des outils qui servent

de médiation. Un sujet est une personne ou un groupe de personnes engagées dans une activité. Un objet est tenu par le sujet et motive l'activité. La médiation peut se produire à travers l'utilisation de différents types d'outils : outil matériel, outil mental (y compris la culture), moyens de penser et langage. Selon [14], l'ordinateur est considéré comme un outil de médiation. L'activité est réalisée à travers un objectif conscient dirigé par des actions qui à leur tour sont réalisées à travers des opérations inconscientes. L'évaluation se fait selon 6 phases. La phase 1 (préparation) permet d'identifier les tâches typiques à analyser, en se basant sur la spécification des besoins. La phase 2 (contextualisation) permet d'identifier les activités dans lesquelles l'application est utilisée comme médiatrice. Pour chaque activité, il faut identifier les actions à travers l'application qui contribuent à la réalisation de l'activité, et les objets et résultats de ces actions. Il faut considérer aussi les autres moyens de réalisation de l'activité sans l'application, c'est-à-dire d'autres artefacts utilisés comme médiatrice de l'activité. Il faut considérer aussi l'horizon d'attente de l'utilisateur, par ex. son expérience en utilisant des applications ou outils similaires. La phase 3 (vérification des tâches) consiste à évaluer sur la base d'une contextualisation de l'application, dans quelle mesure chaque tâche correspond aux actions déterminées dans les activités dans lesquelles l'application sera implantée. La phase 4 (analyse des tâches) consiste à décomposer chaque tâche en séquence d'opérations atomiques comme dans CW. La phase 5 (évaluation) consiste à répondre à 3 questions relatives au contexte perceptuel, aux réponses du système et à l'apprentissage d'utilisation du système. La phase 6 (vérification) consiste à analyser la tâche en se basant sur la phase précédente. Un rapport de synthèse doit être établi pour conclure l'évaluation.

Interaction Walkthrough [42] est une version modifiée de CW, basée sur une théorie dite de l'interaction cyclique (Fig. 2), dont l'idée a initialement été introduite par [5] dans leur cycle de reconnaissance-action. Le modèle en 7 étapes de Norman [34] imagine aussi un cycle d'interaction ; néanmoins, aucun de ces auteurs n'énonce formellement comment l'environnement et les contextes influencent l'interaction de l'utilisateur. Le moyen à la disposition de l'utilisateur pour manipuler le système est centré sur l'action (« chemin but-action »). Une action avec le dispositif d'entrée déclenche des effets système (« chemin action-effets »). Dès la fin de la phase d'exécution, celle d'évaluation commence, c'est ainsi que le système rentre dans un nouvel état. Le « chemin effets-buts » s'occupe des changements sur ceux qui sont perçus et continue sur des nouveaux buts dans le contexte d'interaction. Les 3 chemins d'interaction donnent ainsi 3 classes de problèmes de bas niveau (Cf. [42]) : de but-action, d'action-effet, d'effet-but. L'évaluation concerne les 3 chemins du cycle d'interaction. La phase de préparation consiste à localiser

ser les problèmes. Celle d'inspection permet de détecter les problèmes en répondant à des questions. Celle de vérification consiste à réexaminer les problèmes trouvés précédemment pour vérifier s'ils suivent les chemins suivis par le concepteur.

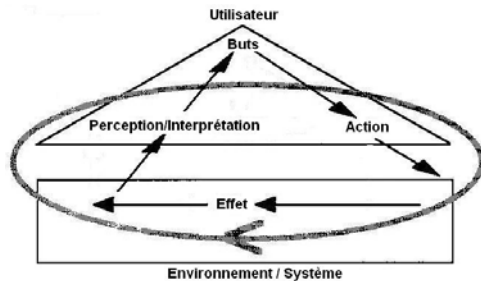


Figure 2 : La théorie de l'interaction cyclique adaptée de [29]

CW with users [7] est une version incluant explicitement les utilisateurs au cours du processus. Elle se déroule en 3 phases. La première consiste à effectuer CW de manière traditionnelle. La deuxième incorpore les utilisateurs : il faut d'abord recruter des utilisateurs représentatifs des profils des futurs utilisateurs finaux. Après une explication introductive, chaque utilisateur est invité à accomplir l'ensemble des tâches correspondant à leur profil. Ils sont invités à exprimer à haute voix leurs pensées, sensations et opinions sur n'importe quel aspect du système ou du prototype pendant qu'ils interagissent avec. Chaque utilisateur accomplit les tâches sans aucune explication complémentaire et à la fin de chaque tâche il doit noter les principaux défauts et imperfections détectées. Une fois les tâches terminées, il est invité à commenter les problèmes identifiés pour cerner son point de vue. La 3ème phase invite les experts à revoir *a posteriori* les doutes ressentis lors à la seconde phase.

Extended CW [16,17] est une extension de la 3ème version de CW, basée sur une extension du modèle de Norman (fig. 3).

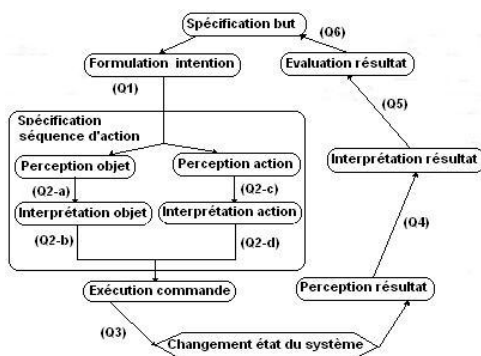


Figure 3 : Modèle de l'action de Norman étendu

Dans celui-ci, la modification concerne la 3ème étape en l'éclatant pour qu'on puisse percevoir et interpréter séparément l'objet et l'action. Pendant la phase d'évaluation, on doit répondre à 9 questions au lieu de 4

pour CW, se rapportant directement au modèle étendu (Fig. 4).

SYNTHESE

Au début, CW était destinée à évaluer les problèmes d'utilisabilité des systèmes de type « Walk-up-and-use », c'est-à-dire qu'on utilise sans ou avec peu de formation initiale. Son succès est tel que les professionnels et les chercheurs s'intéressent à cette méthode pour l'améliorer davantage ou pour l'adapter à d'autres types d'application plus ou moins spécifiques. Dix versions modifiées ou étendues de CW ont été présentées. Les modifications et les extensions effectuées concernent dans la plupart des cas les aspects conceptuels ou théoriques, d'autres concernent en plus les aspects méthodologiques et deux versions seulement concernent en partie les aspects technologiques.

La première version nécessite implicitement une connaissance de base en science cognitive pour bien comprendre la terminologie utilisée dans les formulaires. La compréhension de la distinction entre but et action est une condition cruciale pour la réussite de la méthode de la part des évaluateurs. Pour faire face à ce problème, la seconde version, plus formelle, entraîne les évaluateurs à faire des analyses détaillées pour éviter les ambiguïtés. Le modèle d'action de Norman et le modèle de construction-intégration de Kintsch permettent à cette méthode de faire des analyses détaillées et structurées selon les buts et les actions concernés. Mais avec le nombre important de formulaires à remplir et les listes de questions auxquelles il faut répondre, la méthode devient plus complexe et plus lourde à appliquer. Des outils ont été développés [38] pour aider les évaluateurs à remplir les différents formulaires et à se concentrer sur l'évaluation. Néanmoins, même avec cette aide précieuse, les auteurs de la méthode ont constaté que la méthode est coûteuse en temps et en énergie dépensée pour son application. La 3ème version tente de réduire la complexité et les coûts d'exploitation en introduisant des modifications sur l'aspect méthodologique. Néanmoins la méthode nous semble loin d'être parfaite dans sa troisième version. De manière plus générale, notons aussi que la plupart des auteurs des différentes extensions citées ont émis des remarques, critiques et suggestions en lien avec les points faisant l'objet de leur extension.

SHIVA utilise un modèle de navigation comme base de l'évaluation. Si CW utilise ses propres heuristiques appelées « Design Principles for Successful Guessing », SHIVA utilise les heuristiques proposées par la norme ISO 9241-10. Du point de vue méthodologique, SHIVA adopte le Walkthrough selon deux cycles. Le premier permet d'évaluer chaque vue dans une perspective locale, tandis que le second permet de vérifier globalement l'adéquation de l'évaluation pour la réalisation d'une tâche ou plusieurs séquences de tâches. Nous pensons que cette méthode est intéressante dans la mesure où elle procède en deux phases (évaluation à base de liste heu-

ristique, et à base de scénarios). Elle nous semble toutefois critiquable car l'évaluation ne repose plus sur une modélisation cognitive aussi explicite que dans n'importe laquelle des versions initiales de CW. Nous pensons aussi que les expérimentations effectuées avec cette méthode ne permettent pas de conclure clairement à sa validité puisque les auteurs n'apportent pas véritablement de résultats tangibles issus de leurs expérimentations.

En ce qui concerne **Heuristic Walkthrough**, les modifications ont été apportées au niveau méthodologique. En effet après la première phase qui suit un processus CW entier, la seconde phase pourrait être considérée comme une redondance en termes d'évaluation, c'est-à-dire qu'on évalue une seconde fois le système ce qui devient assez lourd. Mais selon l'expérimentation de Sears [43], cette redondance permet à Heuristic Walkthrough de trouver plus de problèmes d'utilisabilité que dans CW. C'est un résultat prometteur qui appelle selon nous à des études complémentaires.

Concernant **Norman's CW**, le modèle de l'action est adopté comme fondement théorique, ce qui permet de traiter le problème à un haut niveau d'abstraction et de localiser l'origine des problèmes lors de l'évaluation. Au niveau méthodologique, les questions sont focalisées à l'échelle des distances cognitives facilitant ainsi en principe la localisation des problèmes. Dans leur document de référence [40], les auteurs ne fournissent pas d'information pour juger la validité de cette méthode.

Les modifications incarnées par **Streamlined CW** se situent au niveau des aspects méthodologiques et consistent d'abord à bien préparer l'équipe d'évaluation, ensuite de bien encadrer le processus en imposant des règles de base, et enfin à reformuler les quatre questions initiales en deux questions. Ces modifications permettent en principe de mieux maîtriser la contrainte temporelle, puis d'éliminer les longues discussions concernant la conception pendant l'évaluation et enfin d'empêcher les concepteurs de défendre à tort leurs points de vue. Il nous semble que le fait de généraliser les questionnaires au lieu de les détailler peut compromettre l'efficacité de la méthode pour des évaluateurs novices. C'est un point qui pourrait être étudié lors d'expérimentations futures.

Sur le plan de l'aspect conceptuel, **CW for the Web** adopte le modèle CoLiDeS. De plus c'est la seule extension qui incarne une modification sur les aspects technologiques. En effet elle est destinée spécifiquement à évaluer les problèmes d'utilisabilité des sites Web ; les 4 questions sont focalisées sur la navigation et la recherche d'information. Par ailleurs, pour contourner les jugements subjectifs concernant les réponses aux 4 questions, une ASL est utilisée. Pour cela on utilise une description riche des buts permettant d'incorporer plus d'informations concernant la compréhension des utilisateurs de leurs tâches et arriver ainsi à obtenir des buts réalistes. Le choix des espaces de connaissances sémanti-

ques appliqué à l'ASL permet de tenir compte de la population cible (par ex. selon l'âge) lors de l'évaluation. Il nous semble que l'efficacité de cette méthode dépendra donc de la qualité de l'espace de connaissances sémantiques choisi et de son interprétation.

Groupeware Walkthrough est la seule extension destinée à évaluer des applications de type CSCW. Basée sur un modèle de processus de collaboration pour faire face à la complexité et aux spécificités induites par ces applications, entre autres les problèmes d'utilisabilité liés à la collaboration, elle permet aussi d'étudier en détail les contextes de chaque tâche se situant dans le cadre de l'évaluation grâce à GTA. Néanmoins, elle ne permet pas de détecter des problèmes d'utilisabilité classiques. Il semble également qu'elle demanderait des améliorations pour permettre de détecter et prendre en compte les différents aspects socio-organisationnels d'utilisation dans un contexte réel d'un système collaboratif [35].

Activity Walkthrough est basée sur la théorie de l'activité, ce qui permet aux évaluateurs utilisant cette méthode d'inclure plus de contexte dans l'évaluation, puisque l'élément de base de l'analyse est l'activité réelle des utilisateurs. Cependant, il nous semble que si on aborde CW par la théorie de l'activité qui exige la prise en compte de nombreux facteurs notamment liés à la variabilité humaine, mais aussi aux contextes d'utilisation, alors on complique indéniablement l'évaluation en terme de lourdeur et de difficulté d'interprétation (ce qui exige aussi une grande expertise en terme d'évaluation). Ce point va d'ailleurs dans le sens de l'auteur de la méthode [3], ainsi que de celui d'autres auteurs tel Ryu [41] qui insistent sur la complexité et la lourdeur de la méthode.

Interaction Walkthrough est basée sur le modèle dit de l'interaction cyclique. Par opposition aux autres méthodes, qui sont censées détecter les problèmes plus ou moins compliqués en les localisant et en expliquant leurs origines, Interaction Walkthrough détecte plus facilement les problèmes de bas niveau. Cependant, si on se réfère à Ryu et Monk [42] qui pensent que le comportement de l'utilisateur est dépendant de la technologie utilisée et que cette technologie doit être vue à l'intérieur d'une situation d'activité ayant un sens pour l'utilisateur et dans un large contexte, on peut douter de la validité des résultats de cette méthode vue la complexité liée à l'analyse de l'interaction.

CW With Users correspond à un élargissement vers une vraie intervention des utilisateurs dans le sens le plus large du terme (celle en particulier de l'ergonomie francophone). C'est une variante intéressante, sous l'angle de l'ergonomie, mais n'a-t-elle pas pour conséquence de faire sortir la méthode de la catégorie des méthodes d'inspection destinée à des experts uniquement, puisqu'elle fait intervenir des experts et des utilisateurs ? Ceci nous amène à nous poser une autre ques-

tion fondamentale : les méthodes d'intervention et d'analyse choisies sont-ils les plus adaptés (par ex. utilisation de la verbalisation simultanée à haute voix) ?

Du point de vue conceptuel, **Extended CW** est une modification de CW adoptant un modèle étendu de la théorie de l'action de Norman ; ce qui permet d'évaluer l'accessibilité de l'information en tenant compte de tous les aspects cognitifs dès la perception, jusqu'à l'exécution de l'action tout en passant par la compréhension. Sur l'aspect méthodologique, les questions sont focalisées sur les distances cognitives comme dans Norman's CW. Kato et Hori [17] concluent après avoir expérimenté la méthode que Extended CW détecte plus de problèmes d'utilisabilité que CW dans un temps comparable. Bien que les fondements théoriques de la méthode nous semblent solides, la spécificité des conditions expérimentales (type des participants, rémunération, types d'IHM évaluées) décrites dans [16] ne permet pas de conclure sur une généralisation quant à l'efficacité de cette extension dans d'autres conditions (selon les critères de validité proposées par Gray et Salzman [8]).

Pour terminer, nous tenons à insister sur le fait que les variantes de CW, bien qu'elles ne soient pas homogènes, modélisent ce qui se produit quand les utilisateurs exécutent leurs tâches avec un système particulier (aspect conceptuel), en vue d'évaluer ce système en fournissant des questions spécifiques (aspect méthodologique) et en utilisant différents outils et techniques (aspect technologique). Faute de place nous ne pouvons pas détailler ici les questions posées par chaque méthode mais en général, elles ont pour but d'évaluer à chaque interaction si les utilisateurs arrivent à formuler des buts, suivre et exécuter les actions appropriées pour atteindre les buts, interpréter correctement les réponses du système, pour finalement atteindre correctement les buts.

CONCLUSION ET PERSPECTIVES

CW est applicable dans les phases de développement et conception, pour déceler les problèmes liés à l'utilisabilité du système dès sa modélisation et la spécification de ses fonctionnalités. Elle peut être utilisée également *a posteriori* pour déterminer les difficultés d'utilisation lors de la réalisation de scénarios spécifiques et de mises en situations particulières. CW est largement reconnue dans la communauté scientifique en IHM. Ses fondements théoriques associés à son intérêt pratique font qu'elle a, depuis son apparition, fait l'objet de nombreuses études et extensions. Mais pour démontrer l'efficacité et la performance de chaque méthode citée dans cet article, leurs auteurs respectifs se contentent le plus souvent de une ou un nombre restreint d'expérimentations ou études et tirent des conclusions souvent subjectives. Pour déterminer quelle méthode est meilleure, aucune étude n'a été faite pour les comparer entre elles, c.-à-d. entre les différentes versions et les différentes extensions. Il faut remarquer aussi que les auteurs de chaque version ou extension ont adopté diffé-

rentes méthodologies et techniques pour juger de sa validité. Il est intéressant de souligner que quelques études ont été menées pour comparer CW avec d'autres types de méthodes d'évaluation [6,13,15]. Cependant, Gray et Salzman [8] ont fait le point sur les études comparatives existantes impliquant aussi bien CW (dans une de ses versions) que d'autres méthodes, et ont critiqué le fait que ces études présentent beaucoup d'anomalies. Notons aussi que certaines études ont montré qu'une méthode peut être efficace pour un système donné et qu'elle ne l'est pas forcément pour d'autres systèmes [10]. Ainsi, CW a beaucoup évolué, notamment sur les aspects conceptuels, mais toutefois sans avoir pu prouver systématiquement la validité et l'efficacité de chaque évolution ou extension. Sur les aspects méthodologiques et technologiques, nous constatons qu'il existe encore des imperfections. Sur le plan méthodologique, nous constatons des lacunes aux niveaux de la préparation de l'évaluation (formation des évaluateurs, prise en compte des contextes, analyse des tâches...) et de l'exploitation des résultats. Sur le plan technologique, nous pensons qu'une automatisation partielle est nécessaire dans la mesure où elle pourrait contribuer à sa facilité d'utilisation et à la réduction du temps d'évaluation. Un défi consisterait à s'inspirer de ces variantes pour former une méthode plus générale, capable de s'appliquer à des contextes et des types d'IHM variés [1], ce qui fait l'objet d'une de nos perspectives de recherche.

REMERCIEMENTS

Les auteurs remercient la Région Nord-Pas de Calais, l'état et le FEDER qui ont contribué à financer ces recherches (projets MIAOU, EUCUE).

BIBLIOGRAPHIE

1. Bastien, J.M.C. L'inspection ergonomique des logiciels interactifs : intérêts et limites. In Psychologie ergonomique: tendances actuelles, Collection Le Travail Humain, Hoc J.M. & Darses F (Eds.), Paris, 2004.
2. Beck, A., Janssen, C., Weisbecker, A., Ziegler, J. Integrating Object-Oriented and Graphical User Interface Design. In Proc. of the SE/HCI Workshop, Sorrento, Italy, 1994.
3. Bertelsen, O.W. The activity walkthrough: an expert review method based on activity theory. In Proc. NordiCHI, Tampere, Finland, 2004.
4. Blackmon, M.H., Polson, P., Kitajima, M., Lewis, C. Cognitive Walkthrough for the Web. In Proc. ACM CHI, 2002, pp. 463-470.
5. Card, S.K., Moran, T.P., Newell, A. The Psychology of human-computer interaction. Hillsdale, NJ: Erlbaum, 1983.
6. Desurvire, H.W., Kondziela, J.M., Atwood, M.E., What is gained and lost when using evaluation methods other than empirical testing. In People and Computers VII, A. Monk, D. Diaper & M.D. Harrison (Eds.), 1992, pp. 89-102.
7. Granollers, T., Lorés, J. Cognitive Walkthrough With Users: an alternative dimension for usability methods. In Proc. HCI International, Las Vegas, 2005.

8. Gray, W.D., Salzman, M. C., Damaged Merchandise? A review of experiments that compare usability evaluation methods. *Hum.-Comp. Interaction*, 11, 1998, pp. 203-261.
9. Gutwin, C., Greenberg, S. The mechanics of collaboration: developping low cost usability evaluation methods for shared workspaces. *Proc. IEEE Wetice*, 2000, pp. 92-103.
10. Huart, J., Kolski, C., Sagar, M. Evaluation of multimedia applications using inspection methods: The CW case. *Interacting with Computers*, 16, 2004, pp. 183-215.
11. Hutchins, E.L., Hollan, J.D., Norman, D.A. Direct manipulation Interface. *Hum.-Comp. Inter.*, 1, 1985, pp. 311-388.
12. ISO 9241, Ergonomic requirements for office work with visual display terminals (VDTs)- Part 10, 1994.
13. Jeffries, R., Miller, J.R., Wharton, C., Uyeda, K.M. User interface evaluation in the real world : a comparison of four techniques. *Proc. of the ACM CHI'91 Conf. on Human Factors in Computing Systems*, pp. 119-124, ACM Press.
14. Kaptelinin, V. Computer-Mediated Activity : Functional Organs in Social and Developmental Contexts". In Nardi B. (ed.), *Context and Consciousness: Activity Theory and Human-Computer Interaction*, MIT Press, 1996, pp. 45-68.
15. Karat, C.M., Cambell, R., Fiegel, T. Comparison of empirical testing and walkthrough methods in user interface evaluation. *Proc. ACM CHI*, 1992, pp. 397-404.
16. Kato, T., Hori, M. Articulating the cognitive walkthrough based on an extended model of HCI. *Proc. HCI International*, Las Vegas, 2005.
17. Kato, T., Hiro, M. Beyond Perceivability: Critical Requirements for Universal Design of Information. 8th Annual ACM Conf. on Assistive Technologies, 2006, pp. 287-288.
18. Kieras, D. An approach to the formal analysis of user complexity. *IJHCI*, 22, 1985, pp. 365-394.
19. Kintsch, W. The role of knowledge in discourse comprehension: A construction-Integration model. *Psychological Review*, 95, 2, 1988, pp. 163-182.
20. Kitajima, M., Polson, PG. A comprehension based model of comprehension. *HCI*, 12, 1997, pp. 345-389.
21. Kitajima, M., Blackmon, M.H., Polson, P. A Comprehension-based model of Web navigation and its application to Web usability analysis. In *People and Computer XIV*, 2000, pp. 357-373.
22. Landauer, T.K., Dumas, S.T. A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 1997, pp. 211-240.
23. Lavery, D., Cockton, G. Cognitive Walkthrough - Usability Evaluation Materials. Technical Report TR-1997-20, Dep. of Computing Science, University of Glasgow, 1997.
24. Lewis, C., Polson, P., Wharton, C., Rieman, J. Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. *Proc. ACM CHI*, 1990, pp. 235-242.
25. Lewis, C. Why and How to learn why: analysis-based generalization of procedure. *Cognitive Science*, 12, 1988, pp. 211-225.
26. Lewis, C., A model of Mental Model Construction. In *Proc. ACM CHI*, 1986, pp. 306-313.
27. Lewis, C., Wharton, C. Cognitive Walkthrough. In *Handbook of Human-Computer Interaction*, M. Helander, T.K. Landaeur, P. Prabhu (eds.), Elsevier, 1997, pp. 717-732.
28. Mariage, C. MetroWeb : logiciel de support à l'évaluation de la qualité ergonomique des sites web. Thèse de Doctorat, UCL, Louvain-la-Neuve, 2005.
29. Monk, A. A Cyclic Interaction: a unitary approach to interaction, action and environment. *Cognition*, 68 (2), 1998, pp. 95-110.
30. Nardi, B.A. *Context and Consciousness: Activity Theory and Human-Computer Interaction*. M.I.T. Press, 1996.
31. Nielsen, J. Finding Usability Problems Through Heuristic Evaluation. In *Proc. ACM CHI*, 1992, pp. 373-380.
32. Nielsen, J., Mack, R. *Usability Inspection Methods*. John Wiley & Sons, New York, 1994.
33. Nielsen, J., Molich, R. Heuristic Evaluation of User Interfaces. In *Proc. ACM CHI*, 1990, pp. 249-256.
34. Norman, D. A. Cognitive engineering. In D. A. Norman & S.W. Draper (Eds.), *User centered systems design: New perspectives in human-computer interaction*, Lawrence Erlbaum Associates, Hillsdale, 1986, pp. 31-61.
35. Pinelle, D., Gutwin, C. Groupware walkthrough: Adding context to groupware usability. In *Proc. ACM CHI*, 2002, pp. 455-462.
36. Polson, P.G., Lewis, C., Rieman, J., Wharton, C. Cognitive walkthroughs: A method for theory-based evaluation of user interfaces. *IJMS*, 36 (5), 1992, pp.741-773.
37. Polson, P.G., Lewis, C.H. Theory-based design for easily learned interfaces. *Human-Computer Interaction*, 5(5), 1990, pp.191-220.
38. Rieman, J., Davies, S., Hair, D.C., Esemplare, M., Polson, P., Lewis, C. An Automated Walkthrough: Description and Evaluation. In *Proc. ACM CHI*, 1991, pp. 427-428.
39. Riiahio, S. Experiences with usability evaluation methods. Licentates thesis, Laboratory of information processing science, Helsinki Univ. of Tech., May, 2000.
40. Rizzo, A., Mandrigiani, E., Andreadis, A., The AVANTI Project: Prototyping and Evaluation with a Cognitive Walkthrough Based on the Norman's Model of Action. In *Proc. DIS'97: Designing Interactive Systems: Processes, Practices, Methods, & Techniques*, 1997, pp. 305-309.
41. Ryu, H. Collective Web Usability analysis. *International Journal of Web Engineering and Technology*, in press.
42. Ryu, H., Monk, A. Analysing interaction problems with cyclic interaction theory: Low-level Interaction Walkthrough. *Psychology Jour.*, 2(3), 2004, pp. 304-330.
43. Sears, A. Heuristics Walkthroughs: Finding the problems without the noise. *IJHCI*, Vol. 9, 1997, pp. 213-234.
44. Spencer, R. The Streamlined Cognitive Walkthrough method, working around social constraints encountered in a software development company. In *Proc. ACM CHI*, 2000, pp. 353-359.
45. Van der Veer, G.C., Lenting, B.F., Bergevoet, B.A.J., GTA: Groupware Task Analysis – Modeling Complexity. *Acta Psychologica*, 91, 1996, pp. 297-322.
46. Ziegler, J., Burmester, M. Structured Human Interface Validation Technique – SHIVA. In *Symbiosis of Human and Artefact*, Y. Anzai, K. Ogawa, H. Mori (Eds.), Elsevier Science B.V., pp. 899-90.