

## ユーザビリティ定量化手法の提案 (2) チェックリストの有効性検証実験

岡田英彦<sup>†</sup>, 池上輝哉<sup>‡</sup>, 吉坂主旬<sup>‡</sup>, 福住伸一<sup>‡</sup><sup>†</sup>京都産業大学 工学部 情報通信工学科 <sup>‡</sup>NEC 共通基盤ソフトウェア研究所

## 1. はじめに

代表的なユーザビリティ評価手法[1]の1つであるチェックリスト評価法[2]には、開発の上流段階での適用に向くといい利点があるが、一方で、評価結果が評価者のスキルや経験、主観などに依存し、一意で再現性のある評価結果を得ることが難しいという課題があった。この課題に対する研究がこれまでも報告されていたが[3]、評価者間の結果の差の最小化はさらなる課題の1つである。本研究では、チェックリスト評価法に基づくユーザビリティの定量化に向け、評価者の違いによる結果のばらつきをできる限り排除することを目標としたチェックリストの開発を進めている[4]。開発中のチェックリストを用いた評価を複数の評価者で実施し、結果の一意性の検証および改善のための課題抽出を行った。本稿ではこの実験について報告する。

## 2. チェックリスト評価実験

## 2.1 実験方法

本研究では、2007年9月に第1回の実験を行い、12月までに計5回の実験を行ってきた。1回あたりの実験は以下の方法にて行った。

評価するチェックリスト項目

本研究にて開発中のチェックリストに含まれる項目には「基本項目」と「拡張項目」がある。前者はタスクの知識を用いずにチェックが可能と考えられた汎用性の高い項目であり、後者はタスク知識を用いてチェックする項目である。これまでに行った5回の実験はいずれも基本項目を検証対象とし、1回あたり4~5項目を実験者があらかじめ選別して用いた。

評価者

実験1回あたり、5~6名の評価者の協力を得た。この評価者のうち3~4名はチェックリスト評価法の経験がこれまでなくユーザビリティの

知識も十分でない初心者であり、大学生である。一方、残りの1~2名は当該分野の知識・経験を備えた熟練者であり、企業に所属する研究者である。これらの評価者から報告された結果を比較することにより、初心者であっても熟練者と同様の結果を一様に得ることができるかどうか検証できる。5回の実験において、熟練者はのべ3名、初心者のはのべ10名が参加した。

評価対象

ある電子メールソフトのGUIで表示されるウィンドウのなかから、実験1回あたり4~5つのウィンドウを実験者があらかじめ選別し、これらのウィンドウを評価対象の母集団とした。

評価方法

個々のチェックリスト項目について、評価対象とするウィンドウ（母集団に含まれるウィンドウすべてもしくはその一部）、評価すべきGUI部品の種類（メニュー、ボタン、テキスト入力など）、問題点かどうかの判定方法・基準、チェックリストに適合している／いない例、を記載した説明書を実験者があらかじめ作成し、評価者に配布した。評価者はいずれも独立して（他の評価者との協働や情報交換を禁止して）、指定された電子メールソフトを操作しながら、説明書の記載にしたがって評価を行った。また、先行研究[3]とは異なり本研究の実験では、初心者に対してチェックリスト評価法に慣れるための練習評価を設けず、どのチェックリスト項目も初めて評価する項目となるようにした。

評価結果は以下のように報告書に記録させた。あるウィンドウWについて、チェックリスト項目Cに関する問題点が1箇所でも発見された場合は、Wに関するCの評価はその時点で終了とし、「問題あり」と記録させ、問題点の内容も記載させた。一方、問題点がなかった場合には「問題なし」と記録させた。この他に、WにはCで評価すべき対象のGUI部品が存在しない場合には「該当なし」と記録させた。つまり、評価結果は「問題あり」「問題なし」「該当なし」のいずれかである。なお、複数のウィンドウ間の比較に基づいてウィンドウ全体で問題点の有無を判定する必要があるチェックリスト項目では、評価結果はウィンドウごとでなくウィンドウ全体で1つとした。

Proposal of Quantitative Usability Evaluation Method  
(2) Effectiveness Evaluation of ChecklistHidehiko OKADA<sup>†</sup>, Teruya IKEGAMI<sup>‡</sup>, Syuzyun YOSHIZAKA<sup>†</sup> and Shin'ichi FUKUZUMI<sup>‡</sup><sup>†</sup>Faculty of Engineering, Kyoto Sangyo University.<sup>‡</sup>Common Platform Software Research Laboratories, NEC Corporation.

## 2.2 実験結果

前節の「評価方法」において述べた通り、本実験の評価結果は、「問題あり」「問題なし」「該当なし」のいずれかとなる。この結果を初心者・熟練者間で比較することにより、初心者であっても熟練者と同じ結果を得ることができたかどうか、また複数の評価者全体で一意的な評価結果を得ることができたかどうかを調べられる。例えば、表 1(a)に示すような結果であった場合には(初 1~4 は 4 名の初心者、熟 1・2 は 2 名の熟練者を表す)、初心者 4 名の結果がすべて熟練者と一致しており、このチェックリスト項目が一意的な結果の導出に有効であることを示している。一方、表 1(b)に示すような結果であった場合には、初心者 4 名のうち 1 名が熟練者とは異なる結果を報告しており、このチェックリスト項目が一意的な結果の導出において課題を含んでいることを示している。この課題としては例えば、評価基準の定義にあいまいさがある、などが考えられる。

表 1: 評価結果の例

	初1	初2	初3	初4	熟1	熟2
問題あり						
問題なし	●	●	●	●	●	●
該当なし						

(a)

	初1	初2	初3	初4	熟1	熟2
問題あり	●		●	●	●	●
問題なし		●				
該当なし						

(b)

ここで、初心者の結果が熟練者の結果とどの程度一致しているかを表す指標として、一致率を以下のように定義する。

$$\text{一致率} = 100 * \frac{\text{熟練者と一致した初心者数}}{\text{初心者数}} (\%)$$

例えば、表 1(a)(b)の結果では一致率はそれぞれ 100%, 75%となる。

一致率の平均を 5 回の実験別に求めると、表 2 の通りであった。

## 3. 考察

表 2 より、一致率の値は現時点では必ずしも高いとは言えないが、不一致の原因として、チェック漏れや勘違いなどのヒューマンエラー[4]と考えられたものが含まれていた。そこで、1・2 回目の実験のなかから該当すると考えられた(C,W)の組合せ 8 件について、評価に参加した複数の初心者による協議[4]を行わせ、最終的な結

表 2: 一致率の平均 (実験回数別)

1 回目	2 回目	3 回目	4 回目	5 回目
62.5%	77.6%	72.0%	61.9%	66.7%

果を定める実験を行った。この結果、8 件のうち 6 件は一致率がすべて 100%に向上した(協議前の段階では一致率 25%~70%であった)。このことから、複数評価者が参加して協議により最終的な評価結果を定めることで、ヒューマンエラーによる不一致は解決できる見込みが高いことがわかった。今回は時間的な制約から協議を行ったのは 1・2 回目のみとなったが、3~5 回目においても同様の効果が期待できる。さらに、1~3 回目の実験において一致率の低かった項目に対してはチェックリスト項目の改良を行い、改良後の項目を 4~5 回目の実験において別の初心者に評価させたところ、一致率の改善が確認できた。例えば「配色のコントラスト」に関するチェックリスト項目では、初版では評価対象の定義の不備により一致率 0%であったが、定義の改良後は一致率 83%まで改善された。このような改良は今後も継続して行ってゆく予定である。

## 4. まとめ

著者らが開発しているユーザビリティ定量評価のためのチェックリストについて、評価結果の一意性を実験により検証した。複数の評価者の参加と協議による最終的な評価結果の決定を行うことで一致率が高くなったことから、同様の方法論にて当該チェックリストを用いることで、一定の信頼性のある結果を得ることが期待できる。全 126 項目のチェックリストのうち、未検証の項目についても今後、実験による検証とチェックリストへのフィードバックを進めてゆく。

## 参考文献

- [1] Nielsen, J.: Usability Engineering; Academic Press (1993)
- [2] Raviden, S., Johnson, G.: Evaluating Usability of Human-Computer Interfaces: A Practical Method; Prentice Hall (1989)
- [3] 加藤, 堀江, 小川, 木村: HI 設計チェックリストとその有用性の評価; 情報処理学会論文誌, Vol.36, No.1, pp.61-69 (1995)
- [4] 池上, 岡田, 吉坂, 福住: ユーザビリティ定量化手法の提案 (1)評価者による結果のぶれを排除するチェックリストの構築; 情報処理学会第 70 回全国大会 (2008)