

ユーザビリティ評価方法の実践的拡張および適用

河野 哲也[†] TAN LIPTONG[†] 岩本 善行[†] 白井 明[‡] 居駒 幹夫[‡]

[†] 株式会社 日立製作所 ITプラットフォーム事業本部 プラットフォームQA本部 ソフト品質保証部

[‡] 株式会社 日立製作所 ITプラットフォーム事業本部 ソフトウェア本部 生産技術部

[†] [‡] 〒244-0817 横浜市戸塚区吉田町292番地

E-mail: [†] [‡] {tetsuya.kouno.cb, liptong.tan.rk, yoshiyuki.iwamoto.jp, akira.shirai.ws, mikio.ikoma.yh}@hitachi.com

あらまし 現在、ユーザビリティ評価はソフトウェアテストにおいて大きな関心の一つである。従来のユーザビリティ評価方法にはいくつかの制約があるため、テスト・品質保証部門で適用するには課題がある。そこで本研究では、NEMという従来のユーザビリティ評価方法を取り上げ、実践的な拡張を行う。そして、拡張したユーザビリティ評価方法の適用例を示す。

キーワード 非機能要求, 使用性, ユーザビリティ評価, NEM(Novice Expert ratio Method)

Practical Enhancement and Application of a Usability Evaluation Method

Tetsuya KOUNO[†] Liptong TAN[†] Yoshiyuki IWAMOTO[†] Akira SHIRAI[‡] and Mikio IKOMA[‡]

[†] Software Quality Assurance Department, IT Platform R&D Management Division, Hitachi, Ltd.

[‡] Software Engineering Department, IT Platform R&D Management Division, Hitachi, Ltd.

[†] [‡] 292, Yoshida-cho, Totsuka-ku, Yokohama, 244-0817 Japan

E-mail: [†] [‡] {tetsuya.kouno.cb, liptong.tan.rk, yoshiyuki.iwamoto.jp, akira.shirai.ws, mikio.ikoma.yh}@hitachi.com

Abstract Today the usability evaluation is one of important interests in software testing. There are several issues for application of conventional usability evaluation methods to quality assurance organizations because these methods have some restrictions. In this paper, we focus on NEM (Novice Expert ratio Method), which is one of the conventional methods, and describe a practical enhancement of NEM. And this paper shows an example application of the enhanced method.

Keyword Non-Functional Requirements, Usability, Usability Evaluation, NEM(Novice Expert ratio Method)

1. はじめに

近年、ソフトウェアテストの重要性は広く認知され、多くのソフトウェア開発現場ではテスト設計の前倒しやテスト技法の導入など様々な取組みが行われている。

従来のテスト技法の多くは、テスト対象の機能が仕様通りに動くかといった機能要求に関するテストに焦点を当てたものが多く、効率性や使用性など非機能要求に関するテストに焦点を当てたものは充分ではない。一方、多くのソフトウェア開発現場では機能要求通りに動くものの性能面や使いやすさといった非機能要求に関わる問題に関心がシフトしつつある。すなわち、現在、非機能要求に関するテストは大きな関心の一つとなっている。

現在、非機能要求は、その関心の大きさに伴い様々なところで検討・議論されている[1][2]。例えば、非機能要求仕様定義ガイドライン[2]では、機能性や信頼性、使用性、保守性などの10種類の特性を定義し、それらに関する副特性および測定項目を定めている。本研究では、このガイドラインでも取り上げられているよ

うに非機能要求の中で重要な要素の一つである使用性（以降、ユーザビリティと呼ぶ）に焦点をあて、ユーザビリティの評価方法の検討を進める。

従来、ユーザビリティ評価に関して、ユーザビリティテストやヒューリスティック評価法など様々な評価方法が提案されている。これら従来の評価方法は、「ユーザビリティに関する専門的な知識を有する専門家が必要であること」や「一般ユーザを用意する必要があること」などの制約があることから、専門家や専門組織が担当し実施することが多いと考えられる。またテスト・品質保証部門では、そのような専門的な知識を有しない立場でユーザビリティに関する定性的な改善要望や問題点を開発組織にフィードバックしたとしてもそれらが反映してもらえないことも珍しくない。

そのような背景を踏まえ、テスト・品質保証部門においてユーザビリティに関する専門的な知識を有しないテスト・品質保証技術者でも適用できるようなユーザビリティの評価方法の検討・拡張を行うことを本研

究の目的とする。

以降、2章では従来のユーザビリティ評価方法を概観する。3章では従来のユーザビリティ評価方法のNEMを取り上げ検討・拡張を行い、4章ではその適用を行う。

2. 従来のユーザビリティ評価方法

本章では従来のユーザビリティ評価法として、ユーザビリティテスト・ヒューリスティック評価法・チェックリスト評価法・認知的ウォークスルー・NEM(Novice Expert ratio Method)の5つの方法を取り上げ、概観する。

ユーザビリティテスト[3]は、ユーザテストとも呼ばれ、当該製品が対象とするユーザに実際にテスト対象を使ってもらいながらユーザビリティ上の問題点を洗い出す評価方法であり、開発者が見落としがちなテスト対象の問題点を効率的に洗い出すことができる。

ヒューリスティック評価法[3]は、評価のためのガイドラインを事前に用意しユーザビリティに関しての知識を習得した専門家が自らの知識と経験からガイドラインに従いテスト対象の評価を行う方法であり、専門的な知見に照らし合わせた問題点を洗い出すことができる。

チェックリスト評価法[3]は、事前に設定したチェックリストに基づきテスト対象がその確認内容を満たしているかどうかを評価する方法であり、チェックリストの質に依存するもののテスト対象の問題点を広く洗い出すことができる。

認知的ウォークスルー[3]は、システムを利用するユーザにとってどのような目標があり、それを実現するためにどのような手続きが必要で、そのために何が手が必要となるのか、そしてその結果はどのようにフィードバックされるのか、という流れで評価する方法である。具体的には、評価者が想定されるタスクに沿って実際にテスト対象を利用しながらユーザの思考過程や操作を推測して問題点を洗い出す方法である。

NEM[4][5]は、Novice Expert ratio Methodの略でありユーザと開発者とのギャップに関する問題を操作時間の比較によって明らかにする方法である。具体的には、ユーザと設計者が事前に用意されたシナリオに従いテスト対象を操作し、一定の大きさに分割したシナリオごとに操作時間を計測しユーザの操作時間と設計者の操作時間との比をNE(Novice Expert)比として表し、開発者にとって操作しやすいがユーザにとっては操作しにくいといった問題を定量的に分析し洗い出す方法である。

以上、従来のユーザビリティ評価方法においてユー

ザビリティテスト、ヒューリスティック評価法、チェックリスト評価法、認知的ウォークスルーは定性的な評価といった側面が強いが、一方NEMは操作時間およびNE比といった定量的な評価の側面が強い。ここで本研究では定量的な評価という側面を重視し、NEMを取り上げ実践的な検討・拡張を行う。

3. ユーザビリティ評価方法の検討と拡張

本章では、まず文献[4]、[5]に従いNEMの概要を示し、次にNEMの検討および拡張を行う。なお、NEMの詳細については、文献[4]、[5]を参照されたい。

3.1. NEMの概要

NEMは、操作時間を計測しNE比を分析することで問題箇所を特定することができる方法であり、次の一連の手順によって実施される。なお、本手順では、ユーザビリティ評価を企画する企画者とテスト対象を操作する評価者の2つの立場があるため、必要に合わせて企画者と評価者を示すようにする。

(1) 評価対象の選定

テスト対象全体に対してユーザビリティ評価を行うことは現時的ではないので、特定の機能群もしくはユーザの目的などユーザビリティを評価する対象を決定する。

(2) シナリオの設定

(1)で選定した対象に対してどのようなシナリオを評価したいかを検討し、一連のシナリオを設定する。ここで設定するシナリオは操作時間を計測するためのものであり、操作時間に影響が出るため評価者には詳しいシナリオは提示せず、操作目標といった比較的抽象度の高い課題を与える。

(3) 操作時間の計測単位の設定

(2)で設定したシナリオを分節化もしくは集約化し、操作時間の計測単位を決める。

(4) 評価者の選定

一般ユーザと開発者を選定する。なお、評価結果は平均値で処理するため、多くの評価者を集めることが重要となる。

(5) 評価実施

(2)で設定したシナリオに対応する操作課題を評価者に与え、評価者はテスト対象を操作する。また、企画者は(3)の計測単位に従い、操作時間を計測・収集する。

(6) 評価結果の整理と分析

(5)で収集した評価結果に対して、一般ユーザと開発者ごとに計測単位にあわせて操作時間の平均値を算出する。さらに、一般ユーザと開発者の操作時間の平均値に対して比を算出する。この比をNE比と呼ぶ。こ

ここで、操作時間の平均値は折れ線グラフで表し、NE比は棒グラフで表し、一つのグラフで表現することとする。そして、NE比に基づき分析を行い、問題のあるシナリオを特定する。

以上、NEMによるユーザビリティ評価は、操作時間および一般ユーザと開発者との対比といったユーザビリティの部分的側面しか評価できていないことに注意されたい。

3.2. NEMの検討と拡張

本節では、NEMをテスト・品質保証部門で適用することを前提に検討・考察を進める。また、それらの前提と合わせて、より良い分析、すなわちユーザビリティに関する問題の特定についての拡張も考慮に含める。

まず、先に述べた前提を考えると、(4)評価者の選定において一般ユーザを集めることは難しい。文献[4]、[5]では、NEのNoviceを一般ユーザとしているが、元々の意味を考えると必ずしもユーザである必要はない。すなわち、テスト対象に対しての専門的な知識を有しない非熟練者であることが重要であると考え、例えば、品質保証技術者でも担当製品以外の技術者であれば非熟練者とも考えることも可能である。よって本研究では、Noviceを非熟練者と呼ぶようにする。

加えて、NEのExpertを設計者（開発者）としているが、必ずしも開発者である必要なく、同様に熟練者であることが重要であり、例えば品質保証技術者でも熟練者とみなすことも可能である。よって本研究では、Expertを熟練者と呼ぶことにする。

次に、(6)評価結果の整理と分析において、操作時間の平均値を表すとしているが、操作時間のばらつきも考慮に含めた方がより良い分析が行えると考える。例えば、同じ操作時間平均値30秒でも、{25秒/30秒/35秒}と{5秒/30秒/55秒}では後者が何らかの問題があると考えられるからである。

よって、本研究では操作時間のばらつきも分析にとって重要な要素であると考え、熟練者と非熟練者ともにシナリオごとにばらつきが分かる箱ひげ図で表し、そのグラフを追加で評価結果の整理に含めることとする。これにより、熟練者・非熟練者いずれの操作時間において、ばらつきを大きくさせる問題についても分析が行えるようになる。

以上より、(4)評価者の選定において非熟練者と熟練者と捉えるようにすること、および(6)評価結果の整理と分析において操作時間に対して箱ひげ図で表現したグラフを追加することの2点を拡張点とする。次章では、その拡張点を含めたユーザビリティ評価方法の適用例を示す。

4. 適用例

4.1. 適用の流れ

本章では、ミドルウェア製品を対象として3.2で拡張した評価方法の適用を行う。なお、企画者は筆者を初めとした品質保証技術者5名であった。以降、3章で示した手順に従って適用例を示す。

(1)評価対象の選定

本適用の操作の目的は、テスト対象の中でも基本的な機能を使用する「システムを構築する」に選定した。

(2)シナリオの設定

システムを構築するための一連のシナリオを次のように設定した。

①ID・パスワード入力

提示された管理ユーザIDとパスワードを入力する。

②ログイン

ログインボタンを押下する。

③初期設定

所定のリンクをたどり初期設定入力を行う。

④追加設定

追加設定画面を開き所定の設定を入力する。

⑤セットアップ設定

セットアップ画面を開きセットアップを行う。

⑥環境設定

環境設定画面を開き所定の設定を行う。

⑦起動

起動画面を開き起動ボタンを押下する。

⑧ログの確認

起動状況のログを表示し問題がないことを確認する。

(3)操作時間の計測単位の設定

「②ログイン」は、操作時間が極端に短いために③に含め、②と③で一つの計測単位とした。加えて、⑦も同様の理由で、⑥と⑦で一つの計測単位とした。

(4)評価者の選定

本テスト対象は、コンシューマ製品ではなく、ミドルウェアに関して一定の専門知識を有する技術者がユーザであることが想定されたため、当該製品担当外の同程度の知識を有する品質保証技術者4名を非熟練者として選定した。加えて、当該製品の担当の品質保証技術者4名を熟練者として選定した。

(5)評価実施

評価者の操作の一連の流れを観察し問題の所在箇所を確認することも目的の一つとしたため、操作時間はストップウォッチで計測した。

なお、純粋なデータが採取できるようにするために、計8名の評価者に対しては事前にユーザビリティ評価を行っていることは伝えずに操作を行ってもらった。

(6)評価結果の整理と分析

評価結果の詳細および分析は次節で示す。

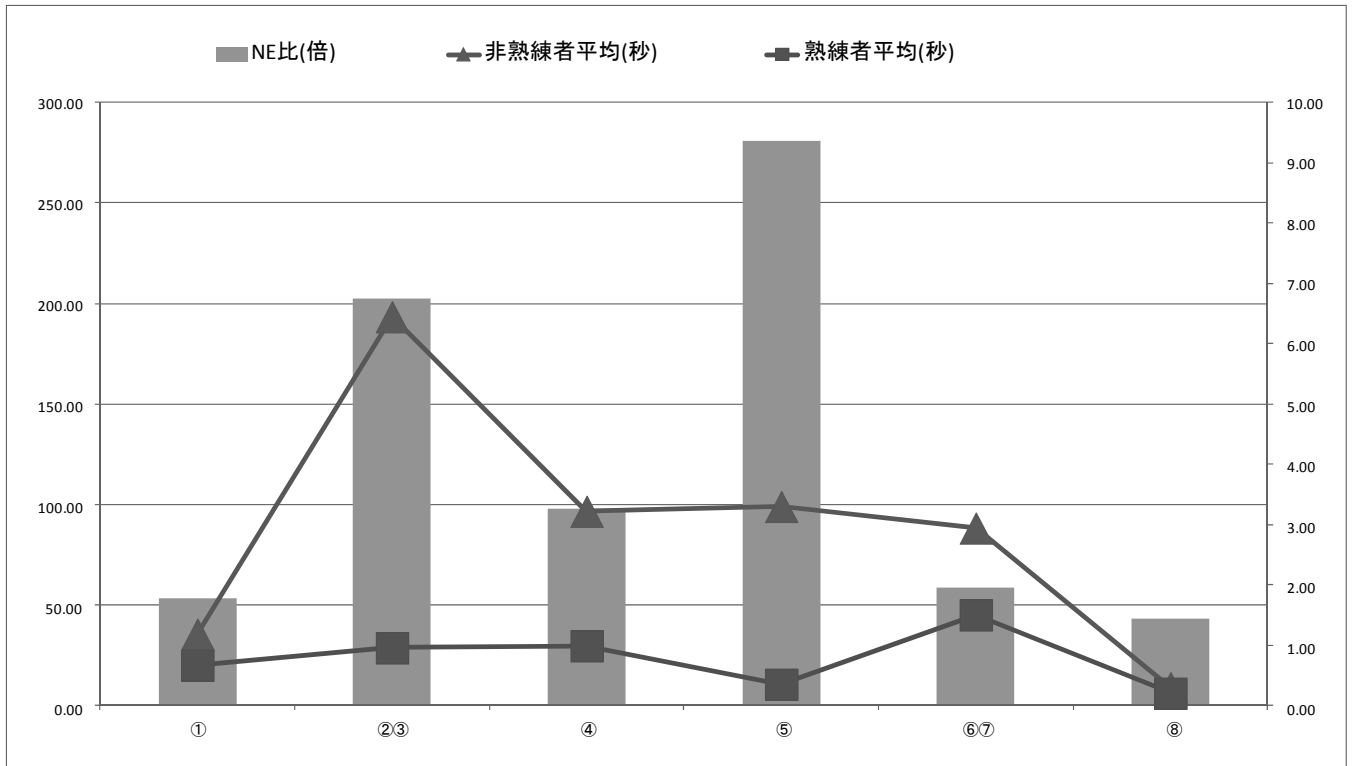


図1 平均操作時間とNE比

4.2. 適用結果と分析

本節では、平均操作時間とNE比のグラフ、および非熟練者・熟練者の操作時間のばらつきのグラフを示し、各グラフに対して分析を行い、続けて適用結果の分析を行う。

まず、平均操作時間とNE比のグラフを図1に示す。

図1が示すように「⑤セットアップ設定」が大きなNE比となった。その理由を操作の観察結果およびセットアップ設定画面に対して企画者で分析を行った結果、セットアップ設定画面のデザインが他の画面と整合性が取れていないことが原因であることが分かった。そのため、設定先を見つけるのに時間がかかってしまったと想定される。この点については、当該製品の担当の品質保証技術者は当該画面のインターフェースになれてしまっていたため、特に問題とっておらず今回の適用で初めて分かった問題であった。

次に、非熟練者の操作時間のバラツキを図2に示し、熟練者の操作時間のバラツキを図3に示す。

図2が示すように、「②ログイン」と「③初期設定」で大きなばらつきがみられる。その理由は、「③初期設定」において一名の評価者が設定操作を間違えたために多くの時間を要したことが原因であることが操作の観察結果より分かった。また、図1の「②③」で非熟練者の平均操作時間が大きくなっていることもそれが原因で平均を吊り上げていることが分かった。

一方、NE比が大きかった「⑤セットアップ設定」では、特定の評価者だけが多くの時間を費やし平均値を吊り上げたわけではないことが分かる。つまり、先に述べた問題が特定の非熟練者に当てはまるわけではなく、総じて画面が分かりにくいといえる。

図3が示すように、熟練者の操作時間では、大きなばらつきはみられなかった。「⑥環境設定」と「⑦起動」で若干のばらつきが見られるが、それは「⑥環境設定」において、企画側が用意したシナリオと熟練者の通常の作業とが若干食い違ったため、操作に戸惑いが生じたことが原因のようである。

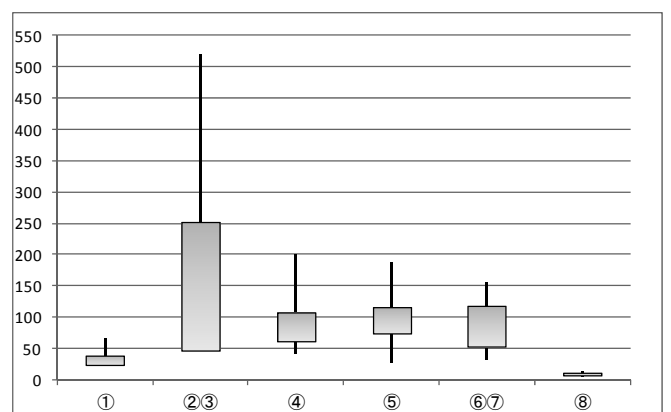


図2 非熟練者の操作時間のばらつき

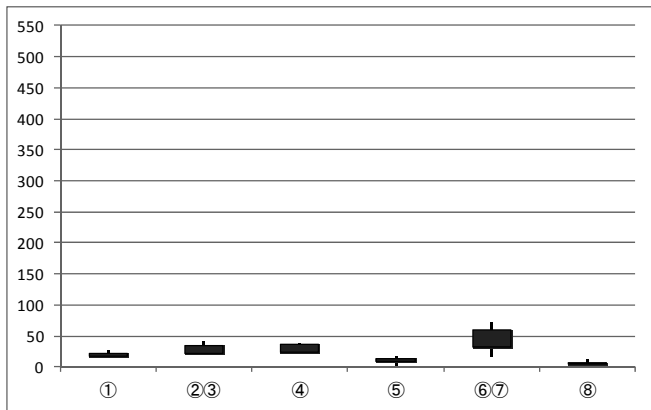


図3 熟練者の操作時間のばらつき

以上、本適用により、今まで経験的にわかっていたいなかったユーザビリティに関する問題が明らかになった。これは、NE比で表現したことにより問題の所在箇所が特定できたことが大きく寄与する結果となった。さらに、非熟練者および熟練者ごとに操作時間のばらつきをあわせて示し、NE比が妥当であることも確認できた。

なお、開発の制約によりテスト対象の画面仕様に対して改良を行うことが許されなかったため、本適用で明らかになったユーザビリティに関する問題は次バージョンで解決するように検討を進めている

4.3. 考察

本研究では、NEM本来の一般ユーザが評価を行うという制約を拡張し、非熟練者であれば評価が行えると考え、その適用を行った。適用結果より、大きなNE比を特定し、その問題を明らかにすることができた。この結果より、一般ユーザを用意できないような制約の元、テスト対象に関しての非熟練者を評価者として用意することができれば適用可能であることが示唆される。

またNEMは、ユーザビリティという枠組みのなかで操作時間に着目して評価を行う方法である。そのため、問題の所在箇所は特定することができるが、何が問題であるのかは他の評価方法の併用、および評価者へのヒアリングなど多岐にわたる検討が必要である。

本研究では、一般ユーザを非熟練者と捉えたが、一般ユーザという観点で他の従来のユーザビリティ評価方法の対応を考えると、ユーザビリティテストに操作時間といった定量的な視点を組み込んだ方法とも言える。あわせて、ヒューリスティック評価法・認知的ウォークスルーの専門家によって実施される評価方法は、NEMの評価に対して補完的な役割で実施していくことが望まれる。

NEMは一般ユーザ（非熟練者）と開発者という対比

で分析することを目的としているが、非熟練者に絞った評価としては、一回目の操作時間と二回目の操作時間とが、操作の学習によってどの程度操作時間が短縮できるかといった習得性といった観点での拡張も可能である。

一方、本適用において非熟練者の操作時間のばらつきで分かった操作の間違いに関しては、NEMの拡張の一アプローチとして考えられる。具体的には、NEMは操作時間という指標に対して、エラー数という指標を追加しNE比を表すような拡張が挙げられる。

5. おわりに

本研究では、非機能要求におけるユーザビリティに焦点をあて、ユーザビリティの専門組織ではないテスト・品質保証部門でもユーザビリティ評価方法が適用できるようにすることを狙って検討を進めた。具体的には、NEMという従来のユーザビリティ評価方法の検討・拡張を行った。くわえて、その適用例を示し、効果を確認した。

今後の課題は、本手法の適用のためのガイドラインの整備や適用事例の拡充、新たな評価指標の検討および評価方法の構築が考えられる。

文 献

- [1] 独立行政法人 情報処理推進機構 ソフトウェア・エンジニアリング・センター (2010) : 「非機能要求グレード」 (<http://sec.ipa.go.jp/reports/20100416.html>) .
- [2] (社)日本情報システム・ユーザー協会編(2008):「非機能要求仕様定義ガイドライン」.
- [3] 黒須正明 他(2000):「ユーザ工学入門 -使い勝手を考える・ISO13407への具体的アプローチ」
- [4] 鱗原晴彦 他(1999):設計者と初心者ユーザの操作時間比較によるユーザビリティ評価手法, ヒューマンインタフェースシンポジウム'99.
- [5] 鱗原晴彦 他(2001):定量的ユーザビリティ評価手法:NEMによる操作性の評価事例およびツール開発の報告, ヒューマンインタフェースシンポジウム'01.