

実践的アプローチによるH C I 評価手法の解析

山岸 典子¹

東 基衛²

1. 日本電気(株) ソフトウェア生産技術開発本部
2. 早稲田大学理工学部工業経営学科

H C I (Human-Computer Interaction) の評価手法として4手法を取り上げ、実験的に適用した。各手法のメリット、デメリットを報告する。取り上げた評価手法は、「思考口述法(thinking-aloud protocol method)」、「インタビュー」、「評価シート」、「データの自動収集」である。実験の結果、「思考口述法」や「インタビュー」はデータの解析に時間がかかるが、問題点の洗い出しやその原因の追求といったシステムの改善に直接役立つ情報がとれる方法であることがわかった。「評価シート」や「データの自動収集」は定量的なデータが容易に得られるが、システム間やバージョン間の比較、ユーザと設計者の意識の比較というような限られた目的にのみ有効であることがわかった。

EVALUATION TECHNOLOGIES ASSESSMENT FOR HUMAN-COMPUTER INTERACTION

NORIKO YAMAGISHI¹

MOTOEI AZUMA²

1: Software Engineering Development Laboratory, NEC Corporation, 11-5, Shibaura 2-chome, Minatoku, Tokyo 108, Japan

2: Dept. of Industrial Engineering & Management, School of Science and Engineering, Waseda University, 4-1, Okubo 3-chome, Shinjuku, Tokyo 160, Japan

SUMMARY

This paper reports experiments on human-computer interaction evaluation. Four techniques are applied to the same HCI design, and the results are compared and analyzed. The techniques applied were protocol analysis, interview, evaluation questionnaire and logging data analysis. Both strength and shortcomings for each technique were isolated. Results indicated protocol analysis and interview are time consuming, but useful in identifying problem areas for a system. Questionnaire and logging data analysis, both of which can produce some sorts of quantitative results are useful only for limited purposes, such as for comparative analysis among systems, versions or categories of users.

1. はじめに

HCI (Human-Computer Interaction) の設計に当たって利用できるガイドラインは存在するのだろうか。また、HCI設計の良否を一般的に評価することは可能なのだろうか。

HCIの良否の判断基準は、個人の経験、好み、あるいは考え方などにより、必ずしも一定の評価を与えることは困難である。このことが冒頭の疑問に対する回答をためらってしまう最大の原因となっている。

日本電気(以下「当社」という)では1983年からソフトウェアの品質計測/保証技術であるSQMAT (Software Quality Measurement and Assurance Technology)の研究開発を行うとともに社内での適用を推進している。SQMATの概念ならびに効果については第8回のICSE他で報告している(砂塚, 1985, 東, 1986)。SQMATの適用を通じて明らかにされた問題点として、SQMATの中の「使いやすさ」の評価は限られた目的に対しては有効であるが、十分な評価手法ではなく、一般的に適用できる評価手法の必要性が認識された。

本研究の目標はHCI設計の評価手法を確立することにある。ここでの大きな問題は設計者が良いと思っているHCI設計とユーザが利用して満足を感じるHCI設計が必ずしも一致しないという点である。

HCI設計の評価に関する報告として次のようなものがある。評価シートを用いたHCIの定量的評価を実施したもの(菊田他, 1984)。評価するシステムとは別のシステムがユーザインタフェースのパフォーマンスを自動的に収集する方法でシステムの使いやすさを定量的に評価したもの(Neal et al., 1983)。思考口述法を用いてHCI評価を行ったもの(加藤, 1985, Lund, 1985, Neal et al, 1983, 和西他, 1986)。

本論文では、人間工学の分野で利用されている4つの評価手法をHCI設計に対する評価手法として比較した結果を報告する。具体的にはソフトウェア開発支援システムの一つであるPGENのHCI設計の評価に対して思考口述法、インタビュー、評価シート、データの自動収集の4手法を適用し、各手法のHCI評価手法としての有効性を分析した。

2. 実験方法

2.1 対象システム

当社のソフトウェア開発支援システムPGENをHCI評価の対象とした。PGENには旧バージョンと旧バージョンを改善した新バージョンがあり、両バージョンをHCI評価の対象とした。

PGENはバンキングシステム開発の支援システムであり、自然言語で書かれた仕様書を入力すると、COBOLのソースコードを出力するシステムである(岩元他, 1986, Haavind, 1986)。ソフトウェア開発者は、日本文エディタ、図エディタ、表エディタを使用して仕様書の入力を行う。主なHCIはメニュー(プルダウンメニュー)選択である。メニュー項目はx-yタレットとライトペンで選択する。日本文エディタの中で、テキストの節や段落の構造を示す時に、コマンド言語を利用する。

評価実験のために、コマンドの履歴と使用頻度を記録する機能をシステムに追加した。

2.2 評価シート

米国メリーランド大学シュナイダーマン教授の評価シートの項目(Shneiderman, 1986)および独自の評価項目を利用し、対象システムのHCIに合った評価シートを作成した。

当実験で用いた評価シートの概要は以下のとおりである。

- (1) 評価項目を次の3つのカテゴリに分類した。

「学習性」「対話性」「使用感」

- (2) 対象システムのHCIに合わせて評価項目を設定した。

項目数は、「学習性」16、「対話性」67、「使用感」17。

- (3) 各項目に対するコメント記入欄を設けた。

2.3 被験者

対象システムに対する開発者とユーザの意識の違いを知ることができるように被験者を以下のように選択した。

24才から29才までの日本人ソフトウェア技術者7名（男性）。

- | | |
|---|------|
| (1) 対象システムの開発者 | : 3名 |
| (2) 対象システムの開発者と同レベルのスキルを持っている者で、
対象システムのH C I設計の概念を知らない技術者 | : 4名 |

2.4 実施手順

(1) 実験準備

被験者に対して実験の説明会を開く。その際、実験の目的、方法、手順の説明および対象システムの操作方法の説明を行う。また、実験時の課題を記した資料と対象システムの操作マニュアルを配布する。被験者には対象システムに対する習熟度が実際にシステムを利用するユーザのレベルにできるだけ近いことが要求される。被験者は操作マニュアルを見ないで対象システムを利用できるようになるまで学習を行う。

(2) 実験

被験者を1グループ4人の3グループに分ける。各グループは開発者2名と開発者以外の被験者2名で構成される。1つのエディタ（日本文エディタ、図エディタまたは表エディタ）は1つのグループで評価される。順序効果を消すために、2人の開発者の内、1人が先に新バージョンのエディタを利用し、次に旧バージョンのエディタを利用する。もう1人の開発者は先に旧バージョンのエディタを利用し、次に新バージョンのエディタを利用する。開発者以外の被験者2名も同様の利用順序をとる。各被験者は対象システムの典型的な操作を含む課題を遂行する。1被験者1実験あたりの対象システム利用時間は約30分である。

被験者に対象システムを利用する際、思っていることをすべて口に出して言うように要求する（思考口述法）。実験者は被験者が発話を続けるように促す。実験者は被験者が課題を全く遂行できなくなる時以外は、被験者の課題遂行を助けることはしない。被験者が課題を遂行する様子をVTRにより録音、録画する。

コマンドの履歴と使用頻度は対象システムによって自動的に収集する（データの自動収集）。

被験者が課題を遂行した後、各被験者に対してインタビューを行う（インタビュー）。インタビュー時間は1被験者1回あたり約20分である。最初に対象システムのH C I全般に対するコメントを求め、次に評価シートの項目についてのコメントを求める。インタビューの結果を記録する。最後に被験者が評価シートによる評価を行う（評価シート）。評価項目に対する意見をコメント記入欄に書く。

(3) データの解析

実験結果の解析は次に示す3つの実験目的に合わせて行う。

- 各エディタのH C I上の問題点の明確化
- 各エディタのH C Iに関する新、旧バージョンの比較
- 各エディタに対する開発者と開発者以外の被験者のH C I評価結果の比較

(4) 結果のフィードバック

データの解析がすべて終わった時点で報告会を開き、対象システムの設計グループに対して評価結果を報告する。

3. 実験データの解析

3.1 思考口述法

実験時の被験者の発話を解析するために、問題票とコメントカードを用意した。

VTRの録音、録画記録をレビューし、トラブル1件あたり1枚の問題票を作成した（図1）。VTRの記録から被験者が陥ったトラブルの内容とその時の発話、被験者がとった行動を、それぞれ問題票の「(ト

ラブルの)内容」、「被験者の発話内容」、「回復手順」の欄に記入した。「(トラブルの)原因」と「防止策」については、実験者および対象システムの開発者で会議を開催し話し合った。

この解析の結果30枚の問題票が作成され、設計者の予想と異なる被験者の行動が具体的にわかった。トラブルの防止策については様々な方法が会議の参加者から出され、1つの方法に決定することはできなかった。

実験時の被験者の発話の内、システムに対する意見をコメントカードに書き出した(1コメント/1カード)。コメントカードはKJ法(川喜田, 1967)で整理した。

61枚のコメントカードが作成された。KJ法で整理した結果、最終的に21の次期バージョンの設計に直接役立つ情報が得られた。

3.2 インタビュー

インタビューで得られた意見をコメント集計用紙にまとめた。図2にコメント集計用紙の例を示す。被験者からの多角的なコメントを得るために、インタビュー項目はHCIを多角的に捉えている評価シートの項目と一致させた。

インタビューのような実験後の被験者の発話は実験中の発話と比較すると、説明的な情報が抜けていることをEricssonらは指摘している(Ericsson, 1984)。当実験では、対象システムを利用している時の被験者の発話とインタビュー時の発言は意味的に一致しており、インタビューが被験者の意見に対してその理由を促すような質問をすることで説明的な情報も取ることができた。評価シートのスコアの意味(なぜスコアが良いのか、悪いのか)を理解するためにインタビューより得られたコメントが役立った。

< 問 題 票 > 発行番号 F-1-2

対象システム	PQEN 表示ソフト	被験者	A	実験者	B	日付	3/20/85									
内容(現象)																
右の画面を最初「9分割」しようとしたができなかった。																
<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> </table> →ここを分割しようとした																
被験者発話内容																
- 何か、間違えた		- 今、2の4、僕の論理的な話でした														
- 何か、おかしい		- これも														
- 何か、どうしたんだろう		- 何か、もう一度														
- 終了押しをもう一回																
- もう一度「9分割」																
回復手順																
一度「水平線」9分割」を終了した。																
次に「もう一度「水平線」9分割」を始める。																
左の画面から先に分割した。																
<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> </table> ←後																
原因																
1. 表の作成手順が表ソフトにおいて論理的な矛盾が生じないように制限がなかった。																
2. 被験者の上記のコメントを知らず、20時Xの20分を知らなかった。																
3. 20時、表示ソフトの通常動作のガイダンスが表示できなかった。																
防止策																
1. 表の作成手順を自由にする。論理的な間違えは、表が作り終わると警告するようになる。																
2. ユーザーが「表」の作成手順を自由にする、制限のあるものにすべからざるようになる。																
3. ガイダンスを改善する。																

図1. 問題票

コメント集計用紙

被験者: Aグループ
ページ: 2/4

【図エディタに対する意見】

一対話のしやすさについて

1. 文字について
 - 一般的に細かい
 - 見にくい
3. メニュー項目のことは
 - まあまあOK
4. 用語の一貫性について
 - 世の中の一般的なものを使う必要がある
5. システムからの指示について
 - キー入力の際、ガイド(またはヘルプ)機能が欲しい
 - n分割についての指示が不足
7. エラーメッセージについて
 - エラーメッセージはほとんど考慮されていない
9. 画面表示について
 - 文字入力の時、下の表がかくれてどこに入れているのかわからなくなる
 - エコーバックの位置を考慮し直す必要がある
 - コマンドの終了が離れたところになり不便
 - 図、表名入力のエコーバックは不要
10. 画面表示と作業の流れについて
 - モードがなんであるかときき返す
 - 自分の位置、状態をもう少し明確にユーザに示して欲しい

図2. コメント集計用紙

3.3 評価シート

評価シートの各項目のスコアの平均と範囲を図3のように示した。

図3は新バージョンの図エディタに対する開発者と開発者以外の被験者の評価点を示している。ここでは「画面表示と作業の流れについて」という項目に対して開発者は高い評価を与えており、開発者以外の被験者は低い評価を与えていることがわかる。しかし、この結果だけでは、なぜ開発者以外の被験者の評価が低いのかを知ることができない。この理由についてはインタビューより得られた「画面表示と作業の流れについて」に対する開発者以外の被験者の意見より知ることができる。ここでは「モードが何であるかときどき迷う」という開発者以外の被験者の意見が得られている。このことからシステムを利用する際、自分がどのモードにいるかわからなくなるといことが開発者以外の被験者にとって大きな問題であったことがわかる。

同様に対象システムの新バージョンと旧バージョンに対するスコアも平均と範囲を示した。

評価シートは、新旧バージョンのHCI比較、開発者と開発者以外の被験者のHCIに対する意識の相違の把握に有効であった。しかし、評価点の意味を理解するためには、定性的なデータを利用する必要があった。

3.4 データの自動収集

対象システムが自動的に収集したコマンドの履歴から、特定の一連の作業に対するコマンドの使用パターンを抽出した(図4)。

コマンドの使用パターンを検討することで、設計者が予想したユーザの行動が実際の行動と合っているかどうかを知ることができた。また、他の手法から得られた言葉のデータを裏付ける情報が得られた。

対象システムが自動的に収集したコマンドの使用頻度からは、設計者がユーザの行動を予測して作成したコマンドでもユーザが全く使用しないコマンドがあることがわかった。使用頻度の高さからだけでは、コマンドの重要性やユーザの満足感を知ることができなかった。

3.5 所要工数

図5に実験及び実験結果の解析に要した工数の内訳を示す。思考口述法に最も工数がかかっている。特に実験後のデータ解析に工数が必要であった。

1 実験あたり被験者が対象システムを利用した時間は平均32.8分、インタビューに要した時間が平均15.7分、被験者が評価シートに回答するのに要した時間は平均19分であった。各手法のデータ解析にかかった工数は、思考口述法が平均6.3人時/1実験、インタビューが1人時/全実験、評価シートが4人時/全実験、データの自動収集が1人時/全実験であった。

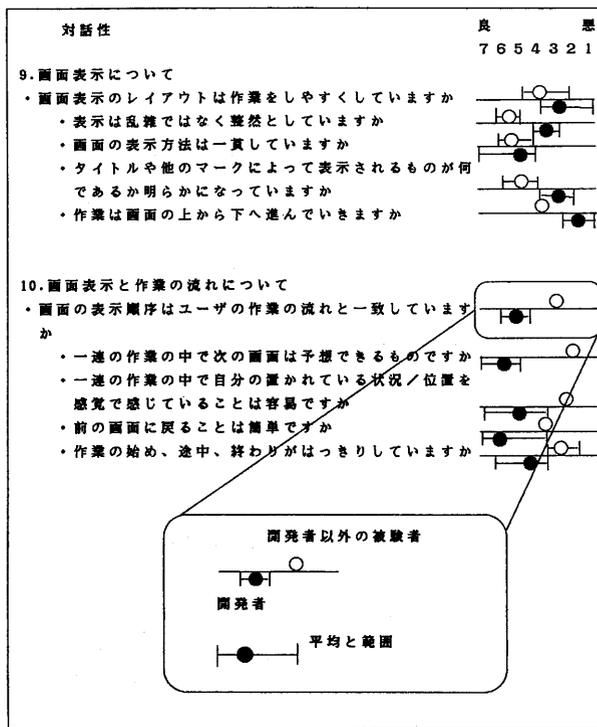


図3. 評価シート

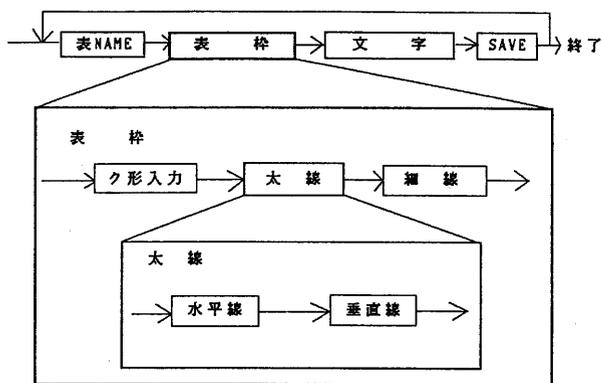


図4. コマンド使用パターン (表作成)

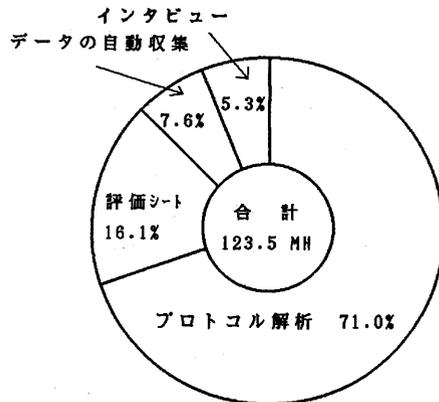


図5. 所要時間

4. 考 察

定性的なデータの得られる手法、特に思考口述法がHCI設計上の問題点を洗い出したり、改善に直接役立つ意見を得るのに予想以上に役立つことがわかった。また、インタビューは評価シートから得られる定量的なデータの意味を知ることのできる重要な手段である。

思考口述法において日本人被験者は思っていることを口に出さない傾向があるとしばしば報告されている。これは対話型コンピュータシステムに対するユーザの行動を理解するために質問記録分析法が提案された大きな理由の一つである(加藤, 1986)。当実験においては、この傾向はなかった。被験者数が少数(7名)なので一般的なことはいえないが、実験に先立つ説明会で被験者に対して思考口述法や実験の目的、対象システムの操作方法等について説明をしたことでこの傾向をなくすことができたようである。

VTRで録音、録画される情報は大変多く、特定の目的に絞ってデータを収集する必要がある。心理学の分野では、思考口述法は主に認知モデルをつくるために利用されている。一般的にこの目的のためには、被験者の全ての発話をディクテーションし、コード化するという手順が取られている(Ericsson et al., 1984)。しかし、当実験の結果からシステムのHCI設計上の問題点を洗い出すためには、発話をディクテーションしたりコード化する必要はないことがわかった。

評価シートを利用することのメリットの1つは、定量的なデータが比較的簡単に得られることである。しかし、具体的に何をどのように改善したらよいかについての情報はあまり得ることができない。対象システムにより自動的に収集したデータからはユーザの満足度を知ることができなかった。思考口述法やインタビュー、評価シートのコメント記入欄から得られる定性的なデータを利用することで、定量的なデータの意味を知ることができることがわかった。

当実験で最も工数を要したのは、実験者による自然言語データの意味の解析である。実験の効率化のためには、文書および発話から得られる自然言語データを自動的に解析する方法が必要である。

5. ま と め

以下のような各手法のメリット、デメリットが明らかになった。

(1) 思考口述法、インタビュー

これらの手法は自然言語のデータを解析する必要があるため、時間がかかる。しかし、対象システムのHCI設計上の問題点を洗い出すのに非常に有効な手法である。

(2) 評価シート、データの自動収集

これらの手法からは比較的容易に定量的な結果を出すことができる。対象システムによるデータの自動収集は被験者に気づかれずに行うことができる。しかし、バージョン間の比較や設計者とユーザの意識の比較というような目的が限られている場合に有効であり、評価の前に予想できない問題点を洗い出すようなことはできない。また、定量的なデータの意味を解析するためには定性的なデータが必要である。

今後は、VTRの記録を被験者自身に見せてからのインタビューの実施や、実験者と被験者の1対1のインタビュー後のグループ討論の実施を試みたい。これらの実施により、客観性のあるコメントや一般性のあるコメントが得られると予想される。

謝 辞

本研究に対し実験対象システムを提供して下さったソフトウェア生産技術開発本部の岩元部長（現在、日本電気ホームエレクトロニクス（株）開発研究所）ならびに最新の評価シートを提供していただいた米国メリーランド大学ヒューマン/コンピュータ・インタフェース研究所長シュナイグマン教授に感謝の意を表するとともに研究を進めるにあたり有益な助言をいただいたHoneywell-NEC Supercomputers Inc. の御前部長ならびにソフトウェア生産技術開発本部の砂塚氏、土田氏に深謝致します。また、被験者として協力していただいた方々に感謝致します。

参考文献

- Ericsson. K. A, Simon. H. A, Protocol Analysis: Verbal Reports as Data, The MIT Press, 1984
- Haavind. R, Tools for Compatibility, High Technology, August 1986, pp. 42.
- Kawakita. J, A Heuristic Method of Abduction from Field Data (a lecture in manuscript form delivered at the 8th International Congress of Anthropological and ethnological Sciences).
- Lund. M. A, Evaluating the User Interface: The Candid Camera Approach, CHI'85, 1985, pp.107-113
- Neal. A. S, Simons. R. M, Playback: A method for evaluating the usability of software and its documentation, CHI'83 Human Factors in Computing Systems, Boston, December 12-15, 1983, pp. 78-82
- Shneiderman. B, Designing the User Interface: Strategies for Effective Human-Computer Interaction, Addison-Wesley, 1986.
- Sunazuka. T, Azuma. M, Yamagishi. N, Software Quality Assessment Technology, 8th Int. Conference on Software Engineering, London, UK, August 28-30, 1985, pp. 142-148
- 東、砂塚、ソフトウェア品質計測/保証技術 (SQMAT)、品質、Vol. 16, No. 1, pp. 79-84, 日本品質管理学会
- 岩元、西谷、和田他、PGEN-1:仕様書理解をベースとしたプログラム自動生成システム、情報処理学会第32回全国大会、1986, pp. 1271-1272
- 加藤、計算機ユーザの認知的行動原理を探るための一手法、情報処理、Vol. 26, No. 9, 1985, pp. 1106-1109
- 加藤、質問記録分析法—計算機ユーザの行動の理解をはかる、2rd Symposium on Human Interface, October 29-30, 1986, pp. 211-216
- 川喜田、発想法、中央公論社、1987
- 菊田、花岡、エンドユーザ向け製品の"使い易さ"の定量的評価、情報処理学会第28回全国大会、1984, pp. 649-650
- 和西、須永、原田、映像情報機器のためのインタフェース・デザインに関する研究1、2、3、デザイン学研究 No55, 1986, pp. 62-64