

HI 設計チェックリストとそのユーザビリティ評価

加藤 秀一 堀江 憲一 小川 克彦 木村 重良

NTT ヒューマンインタフェース研究所 NTT 東京技術開発センタ

計算機システムの使いやすさや分かりやすさを、誰もが客観的に手軽に評価できる HI (ヒューマンインタフェース) 設計チェックリストを開発しその有用性を検証した。開発した HI 設計チェックリストは、133 のチェック項目からなり、画面 (9 セクション、69 項目) と操作性 (7 セクション、64 項目) がチェックできる。チェック結果は、先に開発した HI 設計ガイドラインへの適合度を表わす定量的な指数値で得られ、セクション間のバランス評価や他システムとの比較評価に活用できる。本チェックリストによる実システムのチェックデータを分析した結果、利用者間のばらつきは小さく、チェック法の妥当性ととも客観性の高いチェック結果が得られることを確認した。

A Human Interface Design Checklist and Its Usability

Shuichi Kato * Kenichi Horie * Katsuhiko Ogawa * Shigeyoshi Kimura **

* NTT Human Interface Laboratories

** NTT Tokyo Technology and Development Center

1-2356 Take, Yokosuka, Kanagawa 238-03, Japan

1-19-10 Tabata-shinmachi, Kita-Ku, Tokyo 114, Japan

A human interface design checklist, by which anyone can easily and objectively evaluate the usability and the understandability of his or her system, has been developed. Totally 133 check items of two categories are included in this checklist. One is the category for screen check and includes 69 items classified into nine groups. The other is for manipulation check including 64 items and classified into seven groups. Using this checklist, characteristics which represent the conformity degrees to the HI design guideline can be estimated. Analytical results obtained from the experimental evaluators show that there were no significant differences between evaluated data by participants with different experiences. So the usability of the check method adopted in our checklist has been validated statistically.

1. まえがき

計算機システムのユーザ層の広がりやG U I (グラフィカル・ユーザ・インタフェース)の普及ともなっており、システム開発者やソフトウェア設計者からはH I (ヒューマンインタフェース)ソフトウェアの設計ガイドラインが求められていた^[1~4]。一方、計算機ユーザを中心にH Iソフトウェアの使いやすさや分かりやすさを、誰もが客観的に手軽に評価できるH I評価ツールが求められるようになってきた。H Iソフトウェアの開発当初から積極的に活用できる評価ツールは、計算機ユーザだけに限らず、システム開発者やソフトウェア設計者にとっても開発期間やコスト、品質等の面でメリットは大きい。

使いやすさや分かりやすさの評価基準やモデル、ツールはいくつか示されているが^[5~7]、H Iソフトウェアの開発当初から手軽に活用でき、定量的な評価結果が得られるツールとしてチェックリストがある。679項目からなるSmithとMosierのチェックリスト^[8]は、彼らが開発した944項目のガイドライン^[1]から抜粋されており、チェックリストの利用者が判断する各チェック項目毎の重要度(Weighting)とチェック項目/ガイドライン項目への適合度(Rating)により、セクション別(データ入力法、表示法、ユーザガイダンス、等)の定量的評価やシステム全体の評価ができる。ただし、利用者の立場(システム開発者、ソフトウェア設計者、ユーザ)により重要度や適合度の見解が異なる恐れがあることから、客観性の保証や検証まではなされていない。また、実際の画面を見たり操作をしながら網羅的にチェックする方法ではなく、チェックリストだけで評価するため、改善すべきセクションや項目をチェックリスト上で指摘することはできても、実際の画面上で個別に指摘することは難しい。

著者らも、約370項目のH I設計ガイドライン^[3, 4]に続き、H I設計チェックリスト^[9, 10]を開発し、通信システムを始めとするシステムの開発者や設計者、ユーザに提供してきた。これと並行して、H I設計チェックリスト自身のユーザビリティも評価し

てきた。開発したH I設計チェックリストは、133のチェック項目からなり、画面(9セクション、69項目)と操作性(7セクション、64項目)がチェックできる。チェックは実際の画面を見たり操作をしながら網羅的に行い、チェック結果は、H I設計ガイドラインへの適合度を表わす定量的な指数値で得られ、セクション間のバランス評価や他システムとの比較評価に活用できるとともに、改善すべきセクションや項目を実際の画面上で個別に指摘できる。本チェックリストによる実システムのチェックデータを分析した結果、利用者間のチェック結果のばらつきは小さく、チェック法の妥当性ととも客観性の高いチェック結果が得られることを確認した。

2. H I設計チェックリスト

2.1 開発の狙い

開発に当たっては、H I設計チェックリスト自身のユーザビリティの保証を最重点とした。すなわち、ユーザビリティの高いチェックリストとは、

- (1) 専門家でなくとも利用できる評価手法
- (2) 信頼性・客観性の高い定量的な評価結果
- (3) 改善に直結する評価手法

を提供できるチェックリストであると考えた。これにより、(1)利用者の立場(システム開発者、ソフトウェア設計者、ユーザ)やチェックリストの利用経験による評価結果のばらつきを少なくでき、また、(2)5段階評価のように主観が入りやすい結果ではなく、明確に規定された評価法に基づく定量的な結果により、システムのバランス評価や他のシステムとの客観的な比較が可能となること、更には、(3)評価だけに留まらず、評価結果がH Iの具体的な改善箇所や改善法に直接結びつき、開発期間やコストの削減、品質の保証に活かせる。

ユーザビリティの高いチェックリストの実現に向けて、H I設計チェックリストには以下の評価手法が適用されている。

- (1) チェック内容の品質や根拠をH I設計ガイドラインで保証
- (2) 個々のチェック項目を具体例で説明

(3) すべてのチェック箇所の数え上げ（すべてのチェック箇所の評価結果から、チェック項目やシステム全体の結果を導出）

(4) チェック箇所を Yes / No で評価

これらによる具体的な評価手法を以下の節で示す。

2.2 構成

HI 設計ガイドラインの構成に対応した試作版チェックリストを使用して、実画面や操作のチェックを繰り返し行った経験をもとに、HI 設計チェックリストの構成を決定した。具体的には、チェック

の流れに適合した構成をとった。そのため、HI 設計ガイドラインでは出力方法として一括分類されていた項目を、データ出力、表出力、メニュー、グラフ、等に細分類した。また、ガイドライン特有の定性的な記述や関連するガイドライン項目を、適用場面に応じたより具体的な記述に改め、最終的には133のチェック項目からなるHI 設計チェックリストとした。133のチェック項目は16のセクションに分類され、これらは更に、画面（9セクション、69項目）と操作（7セクション、64項目）に大分類される。図1に構成を、表1にチェック内容を示す。

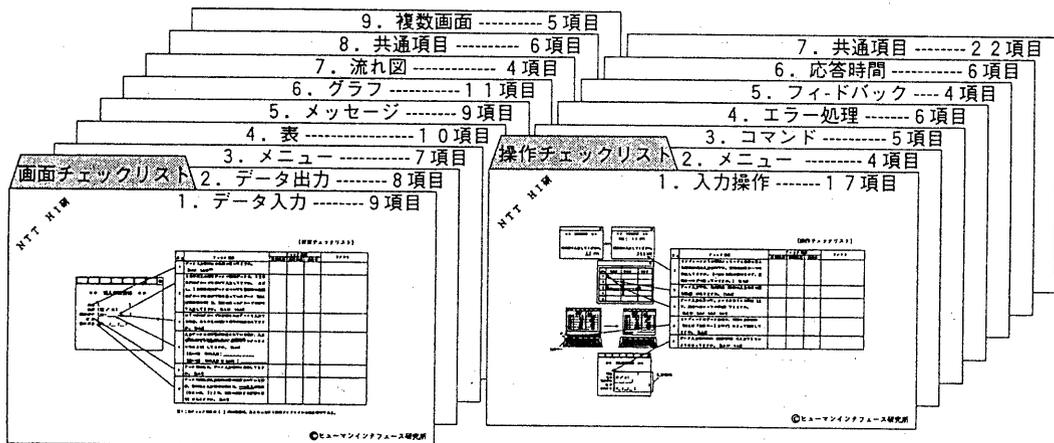


図1. HI 設計チェックリストの構成

表1. HI 設計チェックリストのチェック内容

(a) 画面チェックリスト		(b) 操作チェックリスト	
セクション名	チェック内容	セクション名	チェック内容
データ入力	入力領域名、入力項目のグループ化、書式、プロンプト表示法、等	入力操作	質問形式、先頭機能、保護領域、確認、自動桁揃え、ゼロやスペースの自動挿入、等
データ出力	表示領域名、出力項目のグループ化、書式、単位、等	メニュー	初期メニューや直前メニューへの復帰、選択肢の数、等
メニュー	選択肢のグループ化、表示形式、属性のサンプル表示、等	コマンド	区切り記号、空白数、訂正法、連続起動、等
表	表示項目名、単位、分割法、見やすさ、等	エラー処理	繰り返しエラーへの対処法、メッセージの表示契機、等
メッセージ	肯定形や能動態の使用、操作順序との対応、色、エラー対処法、等	フィードバック	進行状況表示、完了表示、誤操作への対処、等
グラフ	注釈文、ラベル、単位、属性、強調表示法、等	応答時間	応答時間、エラーメッセージ表示契機、スクロール速度、等
流れ図	処理順序と流れの方向、強調表示法、等	共通項目	操作の確認、取消機能、中断/継続機能、ヘルプ、ポインティング領域の大きさ、操作モード表示、画面遷移、等
共通項目	画面名、色の数、番号付与基準、アイコン、等		
複数画面	画面の識別番号、対話やメニュー履歴、強調表示、等		

2.3 評価指標

Smith と Mosier のチェックリスト^[8]では、チェックの最小単位をチェック項目とし、各チェック項目毎の重要度 (Weighting) と適合度 (Rating) の積を基準に、セクション毎の評価指標やシステム全体の評価指標を算出しており、重要度や適合度の判定に主観が入り込む余地がある。本 H I 設計チェックリストではより客観性を高めるため、チェック項目に該当する箇所をすべて数え上げると同時に、各チェック箇所の適合度を Yes (適合) または No (不適合) のいずれかで判定し、全チェック箇所数に対する適合箇所数の割合 (チェック項目別適合率と呼ぶ) を、チェック項目の評価指標としている。各セクションの評価指標は、該当するセクションに含まれる全チェック項目のチェック箇所数の総和に対する適合箇所数の総和の割合とし、セクション別適合率と呼んでいる。画面および操作性全体の適合率も同様に定義し、これらは特に、画面/操作の適合指数と呼んでいる。従って、各チェック項目のチェック箇所数が Smith と Mosier の重要度に、チェック項目別適合率が適合度に対応する。本 H I 設計チェックリストでは、あるチェック項目でチェックした箇所数が多ければ多いほど、そのチェック項目は重要であると考え、チェック箇所数を、セクション別適合率や画面、操作の適合指数算出時の重みとしている。従って、チェック対象システム/タスク実行時の画面や操作の出現頻度を反映した、よりダイナミックな評価指標となっている。

H I 設計チェックリストには、各チェック項目別、セクション別に、チェック箇所数、適合箇所数、適合率を記入する欄が設けてある (図 1 のリストは左側に具体例、右側の表は左から順にチェック項目、チェック箇所数、適合箇所数、適合率、コメント欄となっている)。適合率や適合指数は、H I 設計ガイドラインに適合していない項目の検出、セクション間のバランス比較や他のシステムとの客観的比較の指標となり、適合率が低いチェック項目やセクションは重点的に改善する余地があると言える。H I 設計チェックリストには、対応する H I 設計ガイドライン項目番号がチェック項目単位に記述されており、不適合箇所の改善が容易な構成となっている。例えば、図 2 の評価結果が得られたとすると、画面のグラフと複数画面セクション、操作のエラー処理と応答時間セクションが重点的改善候補と判定できる。なお、図 2 のレーダーチャートの真ん中の数値は適合指数を意味し、画面は全チェック箇所の 83% が、操作は 79% が適合していることを表わす。

2.4 評価と改善のステップ

チェック対象システムの仕様やタスクの背景知識を持ち合わせていない、中立の立場の部外者が H I 設計チェックリストで評価し改善案を作成するまでの典型的なステップを示す。システム開発者、ソフトウェア設計者、または、ユーザが自システムの評価・改善を実施する場合にはステップ①または②は省略できる。

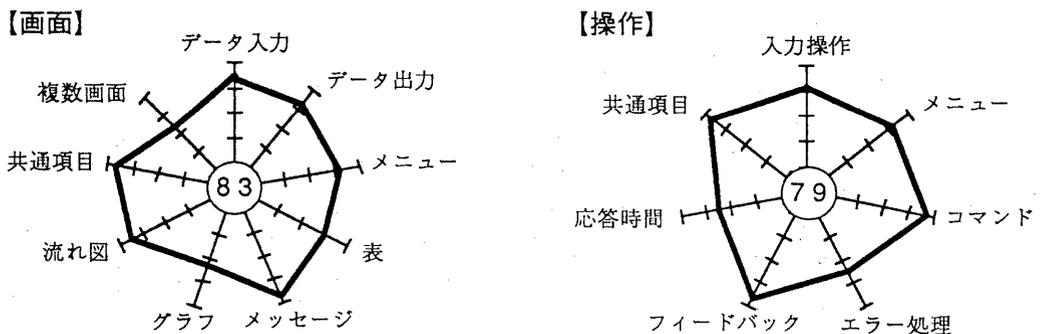


図 2. 評価結果の例 (適合指数とセクション別適合率)

① チェック対象システム／タスクの選定と調査

チェック実施に当たっては、ある程度の業務関連知識が必要となる。部外者がチェックする場合には、システムやタスクの用途、ユーザ層や業務知識を調査する。

② チェック対象画面と操作の選定

すべての画面や操作をチェックできない／する必要がない場合には、代表的なタスクや、改善が望まれている画面や操作を選定する。通常、連続した複数の画面や操作を選定する。

③ H I 設計チェックリストによるチェック

通常は、選定した全画面と操作の該当する箇所をすべてチェックし、評価指標を求めるが、必要ならば（例えば、複数の設計者が各々の思想に基づいてバラバラに設計した場合には）、タスク単位、画面単位に評価指標を集計する。

④ 評価（適合指数、レーダーチャート、等）

評価結果（チェック項目別適合率、セクション別適合率）をもとにレーダーチャートの作成、適合指数の算出を行う。また、必要ならば他のシステムや標準値との比較を行い、チェック箇所数が多くて適合率の低いチェック項目やセクションが存在しないかをチェックする。存在すれば、それらは重点的に改善すべき項目である。

⑤ 改善案の作成

重点的改善項目の改善案を作成する。改善法はチェック項目から容易に判断できる（チェック項目はガイドラインの裏返しとなっており、ガイドラインに立ち返らなくとも改善は容易である）。

⑥ 必要ならばドキュメントを作成

システム開発者やソフトウェア設計者、ユーザにドキュメントで評価結果や改善案を提示する意義は大きい。評価だけに終わらず、具体的に不適合箇所や改善案までを提示すると、納得が得られ易い。

3. ユーザビリティ評価

H I 設計チェックリストが、（1）専門家でも利用できるか、（2）客観性の高い評価結果を提供できるか、を検証するための実験を実施した。

実験では、画面チェックリストだけを使用した。操作チェックリストは機能の有無や操作の是非のチェックが中心であり、ほとんどのチェック項目はチェック箇所数が1となり、試行実験の結果からみればらつきは少ないと予測したためである。

3.1 実験手法

3.1.1 被験者

専門家でなくとも H I 設計チェックリストで正しい評価が可能か否かを検証するため、H I 知識や H I 設計ガイドラインおよび H I 設計チェックリストの使用経験を有する経験者 1 名（男性、経験 2 年）と、未経験者 3 名（男性 1 名、女性 2 名）を被験者とした。年齢はいずれも 20 代である。

3.1.2 タスク

未経験者 3 名には、実験タスクに先立って学習タスクを実施した。学習用画面（約 10～15 画面、2～3 システム）により、H I 設計チェックリストの使用経験を積ませた（約 1 カ月）。

実験タスクでは、社内で運用または開発中のシステム（主に通信システム）から、代表的な連続画面（平均 6 画面／システム）を選定し、1 システム毎に、節 2.4 で述べた評価と改善の 6 ステップを実施し、合計 10 システムを順に評価した。各被験者は、ステップ①、②を共同で、ステップ③を個別に実施し、ステップ③終了時点で各自のチェック結果を持ち寄り、合議の上で正しいチェック結果を判定した。経験者と言えどもチェック箇所の見落としや数え間違い、等があるためである。従って、各被験者は 1 システム毎にチェック結果の正誤が分かる。ステップ④、⑤、⑥は各システムの評価責任者 1 名（いずれも未経験者が順に担当）だけが実施した。

3.1.3 収集データ

各システム毎に、ステップ③のチェックに要した時間（被験者の申告値）と、セクション別正解率、システム全体の正解率を収集した。各正解率はチェック項目単位の正否から算出した。チェック項目単位の正否は、各システムの全チェック対象画面に対し、そのチェック項目でチェックすべきチェック箇所数と適合箇所数がともに正しい場合にのみ正

解とした。従って、1カ所でもチェック箇所を見落としたり適合／不適合の判定を誤ると不正解となる。セクション別正解率は、そのセクションに含まれる全チェック項目数に対する正解チェック項目数の割合であり、システム正解率は、全チェック項目数に対する正解チェック項目数の割合である。ただし、セクション全体がチェック対象とならなかった場合には、そのセクションの全チェック項目（例えば、チェック対象画面に流れ図が含まれていなかった場合には流れ図セクションの4項目）を正解率の算出対象から除外した。10システムのチェック対象画面数は4～9画面（平均は6画面）、平均チェック箇所数は約240、平均チェック項目数は約43である。これより、平均チェック箇所数／画面は約40、チェック項目当たりの平均チェック箇所数は約5.6となる。なお、実験期間は約半年間である。

3.2 解析結果

3.2.1 正解率

経験を主要因、10種類のシステムを被験者内要因とした、分散分析によるシステム正解率の有意差検定では、経験の効果に有意差はない。一方、システムの効果には有意差があるが、両者の交互作用に有意差はない。

経験 $F(1, 2) = 8.876, p = 0.097$

システム $F(9, 18) = 16.925, p < 0.001$

交互作用 $F(9, 18) = 1.811, p = 0.136$

図3にシステム正解率の平均値を示す。システムによる差が大きいのは、各システムの個性の現われであり、経験者と言えども初めてチェックするような

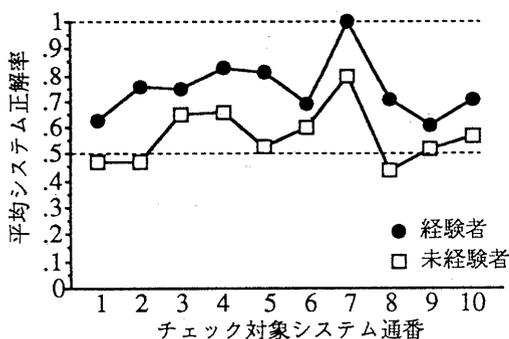


図3. 平均システム正解率

画面が含まれていたことによる。

経験者の正解率が必ずしも100%ではない（10システム平均で75%）のは、大雑把に言えば5.6箇所（チェック項目当たりの平均チェック箇所数）すべてを完璧にチェックできる確率が75%であることを意味し、1箇所当たりの正確さでは95%に相当する（ $0.95^{5.6} \approx 0.75$ ）。未経験者の10システム平均は57%であるが、1箇所当たりの正確さでは90%に相当する。

セクション別正解率（ただし、グラフ、流れ図セクションはデータ不足につき解析不能）の4名の個人差の検定結果では、共通項目と複数画面セクションに個人差がある。

データ入力 $F(3,24) = 0.624, p = 0.605$

データ出力 $F(3,24) = 2.167, p = 0.109$

メニュー $F(3,24) = 2.578, p = 0.077$

表 $F(3,24) = 0.719, p = 0.549$

メッセージ $F(3,24) = 0.075, p = 0.973$

共通項目 $F(3,24) = 6.014, p = 0.002$

複数画面 $F(3,24) = 11.511, p < 0.001$

個人差のあるセクションは、チェック項目、チェック箇所数とも少ないセクションである。一方、システム正解率同様、データ出力以外の全セクションのシステムの効果に有意差がある。

3.2.2 チェック時間

経験を主要因、10種類のシステムを被験者内要因とした、分散分析によるチェック時間の有意差検定では、システム正解率同様、経験の効果に有意差はなく、システムの効果に有意差がある。両者の交互作用に有意差はない。

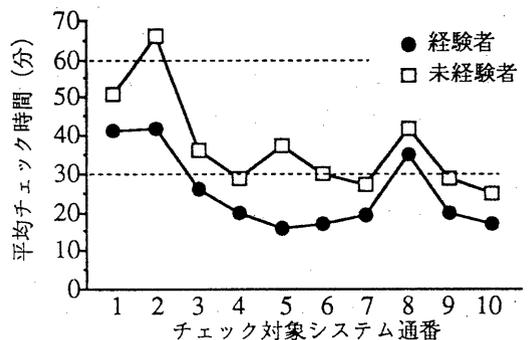


図4. 平均チェック時間

経験 $F(1, 2) = 5.969, p = 0.135$

システム $F(9, 18) = 9.631, p < 0.001$

交互作用 $F(9, 18) = 0.452, p = 0.888$

図4に平均チェック時間を示す。システムの効果の有意差は、システムによりチェック画面数やチェック箇所数に差があるためであり、チェック時間の大小はチェック箇所数にほぼ比例している。なお、平均チェック時間は、10システム平均で、経験者が約25分、未経験者が約37分である。

4. 考察

H I 設計チェックリストの開発に当たってはユーザビリティの保証を最重点課題とした。すなわち、ユーザビリティの高いチェックリストの条件は、

- (1) 専門家でなくとも利用できる評価手法
- (2) 信頼性・客観性の高い定量的な評価結果
- (3) 改善に直結する評価手法

の提供であるとし、そのために、H I 設計ガイドラインで保証されたチェック内容を具体例で説明した体裁をとり、Yes / No で評価したチェック箇所をすべて数え上げる手法をとった。

(1) 専門家でなくとも利用できる評価手法

解析結果が示すように、システム正解率やチェック時間に対し、H I 知識やH I 設計ガイドライン、H I 設計チェックリストの使用経験による差は認められない。システム正解率は全画面（平均6画面）に対する正解率であり、経験者が75%、未経験者が51%～61%（平均57%）であるが、画面単位では両者とも90%以上の正解率でチェックできると推定される。実験とは別の被験者（システム開発者）による自システムのチェック事例からも、学習タスクを全く実施しなくとも、約80%の画面単位の正解率が得られており、専門家でなくともH I 設計チェックリストを手軽に（画面当たり5～6分のチェック時間で）利用できると期待している。

一方、システム正解率、チェック時間両者とも、システムの効果に有意差があるが、チェック時間の場合は、チェック箇所数とチェック時間の大小が対応しており、自然な結果である。システム正解率の場合には、各システムの業務特性や、画面構成や表

示法の個性による差が、被験者間の微妙な解釈の差となって表われており、経験者と言えども100%の正解率を常には得られない一因となっている。対応策としては、(1) 適合率や適合指数の算出単位を小さく（例えば、画面単位に）し、それらの結果からシステム/タスク全体の評価指標を算出する、

(2) 何人かの混成チームでチェックし、評価の精度を高める、等が効果的と考えている。

(2) 信頼性・客観性の高い定量的な評価結果

信頼性はH I 設計ガイドラインで、客観性はYes / No で評価したチェック箇所をすべて数え上げる手法で保証した。実際に、2000箇所以上のチェック箇所はすべてYes/No で評価できており、判定基準は明確である。5段階評価や、SmithとMosierの重要度、適合度の判定のように、利用者の立場による主観が入り込む余地は殆どなく、信頼性、客観性の高い定量的評価手法が提供できたと考えている。

一方、H I 設計チェックリストで満足できる結果を得ることは、必要条件ではあるが十分条件ではない。H I 設計チェックリストでは、初めてのユーザにもやさしいインタフェースか否かを中心に、いくつかの異なる観点からチェックしており、単純そうな画面でも延べ20～30箇所、複雑な画面では100箇所以上をチェックし、不適合箇所を指摘できるが、最終的なH I 設計には、実ユーザによる評価実験結果、等をも考慮することが望まれる。

(3) 改善に直結する評価手法

H I 設計チェックリストは、H I 設計ガイドラインへの適合度をチェックするものであり、根拠となったガイドラインとの対応関係が明記されている。従って、不適合と判定されたチェック箇所の改善策は容易に設計できる。現に、実験タスクのチェック結果と、被験者自らが設計した不適合箇所の改善策は、開発者や設計者、ユーザに問題なく受け入れられ、バージョンアップ時の改善項目として反映されている。改善策の設計工数も、改造工数に比較して、非常に小さく、プロトタイプや設計レビュー時にチェックする効果は大きい。

5. 結論

計算機システムの使いやすさや分かりやすさを、誰もが客観的に手軽に評価できるH I 設計チェックリストを開発しそのユーザビリティを評価した。H I 設計チェックリストは、133のチェック項目からなり、画面（9セクション、69項目）と操作性（7セクション、64項目）がチェックできる。チェック結果は、先に開発したH I 設計ガイドラインへの適合度を表わす定量的な指数値で得られ、セクション間のバランス評価や他システムとの比較評価に活用できる。本チェックリストによる実システムのチェックデータを分析した結果、利用者間のばらつきは小さく、チェック法の妥当性ととも客観性の高いチェック結果が得られることを確認し、ユーザビリティの高い、すなわち、

- (1) 専門家でなくとも利用できる評価手法
- (2) 信頼性・客観性の高い定量的な評価結果
- (3) 改善に直結する評価手法

を具備するH I 設計チェックリストが実現できた。

今後、H I 設計チェックリストの有用性の検証を更に進めるとともに、システム構成やサービス特性、ユーザ特性（システムを使用するユーザの年齢層、経験度、使用頻度、等）をも反映できる評価体系へと発展させたい。また、ISOで進められているH C Iの国際標準化^[6]の進展にあわせ、それらとも整合をとって行きたい。

<謝辞>

本研究の機会を与えていただいたヒューマンインタフェース研究所 高野所長、釜江前所長、マルチメディア処理研究部 栗原部長、遠藤前部長に深謝いたします。また、評価にあたり非常に多くの方にご協力いただきました。心よりお礼申し上げます。

<参考文献>

- [1] Smith, S., L., Mosier, J., N. : Guidelines for Designing User Interface Software, Technical Report ESD-TR-86-278, MITRE, MA, [1986].
- [2] Shneiderman, B. : Designing the User Interface,

Addison-Wesley Publishing Co. Inc., [1987]. (東, 井関 監訳: ユーザ・インタフェースの設計, 日経マグローヒル社, [1987].)

- [3] 米村, 小川: 通信システムにおけるヒューマンインタフェース設計指針の考察, NTT R&D, 39 (2), pp. 249-256, (1990).
- [4] 米村, 小川: ブックメタファを用いたH I 設計ガイドラインデータベース, 情報処理学会, 第36回ヒューマンインタフェース研究会, 36-2, [1991].
- [5] Whiteside, J., Bennett, J., Holtzblatt, K. : Usability Engineering : Our Experience and Evolution, in M. Helander (Ed.), Handbook of Human-Computer Interaction, pp. 791-817, (1988).
- [6] 浜田, 森川, 小川: H C Iの国際標準化〜その1: 標準化の視点とユーザビリティ〜, 情報処理学会, 第45回ヒューマンインタフェース研究会, 45-14, [1992].
- [7] 旭, 岡田, 井関: 「使いやすさ」の計測: U I テスタ, 計測自動制御学会H I 部会, 第8回ヒューマン・インタフェース・シンポジウム論文集, pp. 287-290, [1992].
- [8] Smith, S., L., Mosier, J., N. : A Design Evaluation Checklist for User-System Interface Software, Technical Report ESD-TR-84-358, MITRE, MA, [1984].
- [9] 木村, 米村, 小川: ソフトウェアのユーザビリティチェックリスト, 日本人間工学会, 第21回関東支部大会, A1-16, pp.63-64, [1991].
- [10] 堀江, 加藤, 小川, 木村: H I 設計チェックリストとその有用性, 計測自動制御学会H I 部会, 第9回ヒューマン・インタフェース・シンポジウム論文集, 投稿中, [1993].