

ユーザの生体情報を用いたシステムの使いやすさ評価

奈良工業高等専門学校 情報工学科

上野 秀剛

Usability Evaluation of System using User's Biological Information

HIDETAKE UWANO

Department of Information Engineering, Nara National College of Technology

Abstract: 本稿では著者がソフトウェア工学の分野において研究している生体情報を用いた定量的評価について紹介する。工学の分野においては眼球運動や筋電図、脳波といった生体データから被験者の行動や心理状態を定量的に分析する方法が提案されており、様々な研究に利用されている。本稿ではソフトウェアの使いやすさを計測するために、している研究と、熟練者の行動や思考から暗黙知を抽出する研究について紹介する。また、著者がそれぞれの分野で著者が行ってきた脳波を用いたソフトウェアの使いやすさ評価、および、視線移動の計測による熟練ソフトウェア開発者の行動測定に関する実験について詳細に説明する。

key words: Biological information, Human factor, Usability, Performance measurement, Software engineering

1. はじめに

現代社会ではあらゆる所でソフトウェアが使われている。各種工場の機械制御や在庫管理、オフィスの業務支援、携帯電話などあらゆる人を対象とした、あらゆるソフトウェアが存在している。ソフトウェア工学と呼ばれる分野では、ソフトウェアの開発と利用に関する様々な研究が行われているが、中でも著者は以下の2つのテーマについて興味を持っている。

- 利用者にとってどのようなソフトウェアが使いやすいのか
- 開発者がより効率的に、高品質なソフトウェアを作るにはどうすればよいのか

ソフトウェアが使いにくいためにイライラしたり、操作を誤ったために作業途中の情報が消えてしまったという経験は、コンピュータを日常的に使用している人の多くが経験していることだろう。ソフトウェアやハードウェアの使いやすさ（ユーザビリティ: Usability）に関する問題は、特に家電メーカーにおいて古くから取り組まれており、新製品を被験者に利用してもらい、インタビューやアンケートによる評価が行われてきた。ソフトウェア開発の現場においても同様の取組が行われている。ソフトウェアのユーザビリティは売上に影響を及ぼすだけではなく、サポートの費用を下げることもできるため、利用者の視点に立った開発が重要である。

一方で、ソフトウェア開発の現場では、より高品質なソフトウェアをより短時間で作成することが求められている。近年のソフトウェアの多機能化、複雑化に伴い、

ソフトウェアの不具合による製品回収やサービスの停止といったトラブルが増加している。そのため、作業手順の標準化といったシステム面の改善と共に、開発者が高いスキルを取得すること求められおり、熟練者がどのように開発を行っているか分析し、熟練者のノウハウを生かした教育・支援を行う必要がある。

これら2つの活動はいずれも人間に焦点を当てており、行動や心理状態を計測するためにアンケートやインタビュー、ビデオ撮影による分析が行われていた。しかし、定量的な分析が難しい、無意識に行っている暗黙知に従った行動などは把握が難しいといった問題があった。

これらの問題を解決するための方法として生体データから被験者の行動や心理状態を分析する方法が提案されている。現在利用されている生体情報には心電図、皮膚表面温度、眼球運動、筋電図、脳波、発汗量などがある。生体データを用いた評価方法には以下の利点がある。

- アンケート・インタビューによる評価と異なり定量的な計測が可能
- 評価方法が確立すれば、専用ソフトウェアを開発することで自動的な評価が可能であるため、専門家以外でも評価が可能
- 評価者の主観が入りにくい
- 熟練者とそれ以外の人の行動を比較することで暗黙知の顕在化が可能

本稿ではソフトウェア工学分野で生体情報を用いた研究について紹介すると共に、著者がこれまでに行った生体情報を用いた2つの実験について詳説する。また、同

様の手法を医療分野におけるシステム、環境の評価に適用することの有用性について考察する。

2, 3 章では生体情報をユーザビリティ評価に用いている研究について紹介する。4, 5 章では熟練者の行動測定を行っている研究を紹介する。6 章では本稿のまとめを行う。

2. ユーザビリティ評価

ソフトウェアの多機能化・複雑化に伴い、機能的な利便性だけではなくユーザビリティの観点から、システムを利用している際の快・不快や心理的作業負荷(Mental Workload:MWL) など、ユーザの心理状態を考慮することが重要となっている。従来、対象となるシステムのユーザビリティを評価するための方法として、インタビューを用いる方法やシステム利用後に質問票に回答してもらう方法、対象システムを利用している際の思考状態をリアルタイムに発話してもらう発話プロトコル法[1]などが用いられている。ユーザビリティの評価に用いられる質問票として代表的なものには SD 法[2], QUIS [3], NASA-TLX [4] がある。また、NASA-TLX に関しては芳賀らによって日本語化されている[5]。これらの手法は計測のために特別な機器を必要とせず、比較的容易に計測できるためユーザビリティの評価によく用いられている。一方で、得られたデータの分析や評価に多大な時間がかかることや、再現性が低く、定性的・主観的な評価になってしまうといった欠点も存在する。

このような主観的な評価を補うために人間の心理状態を定量的・客観的に評価する方法として、人間の生理状態を外部から測る生体計測についての研究が行われている。生体計測では心拍変動や脈波、発汗、筋電、脳波といった人間の生体情報を専用の機器を用いて計測し、その変化や違いを分析する。生体情報は人間の心理状態と密接な関係があるといわれており[6]、評価対象となるシステムを利用している際の生体情報を用いることで定量的なデータによる心理状態の計測を行うことができる。このような生体情報による定量的な分析と、従来行われている定性的な分析結果についての関係を調査し、蓄積していくことで、定量的なデータによるユーザビリティの評価が可能になると期待される。

人間の心理状態を定量的に計測するために、従来さまざまな生体情報を利用した研究が行われている。工学分野でよく使われている生体情報は大きく末梢神経系、中枢神経系の2つに分類される[7]。

末梢神経系の活動を観察する指標として、心電図、皮膚表面温度、眼球運動、筋電図などがある[8]~[11]。水野らは暗算タスクにおける MWL の継続による情動ストレスの変化を測定するため、顔面熱画像の鼻部額部差分温

度、脳波、心電図の継続的計測を行い、鼻部額部差分温度変動の最大温度変異による MWL の定量的評価の可能性を示唆している[9]。Nakayama らは暗算などのタスク遂行時における眼球運動を計測し、タスクの難易度が高くなるにつれて瞳孔径と瞬目の回数が増加することを示している[11]。

中枢神経系の活動として計測することができる生体反応には脳波(EEG)、脳磁気図(MEG)、機能的核磁気共鳴画像(fMRI)、コンピュータトモグラフィ(CT)などがある。この中でも脳波は他の方法と比べて計測時の姿勢・動作が比較的自由的、コンピュータ利用時にも計測が可能、計測装置が安価であることから、ディスプレイ上に表示されたソフトウェアを利用している際の心理状態を計測するのに適している。

人間の心理状態を観察するための脳波計測においては脳波に対して離散フーリエ変換することで得られた α 波、 β 波のパワースペクトルや、 α 波、 β 波の脳波全体に対する割合、 α 波と β 波の比率である β/α が指標としてよく用いられている[7,12,13]。松永らは脳波により人間の満足感を評価する満足感計測システムを開発し、実験により脳が扱う情報処理量が小さいと快が生じ、反対に情報処理量が大さいと不快が生じるという仮説を検証した[14]。林らは高品位映像の客観的評価指標として脳波を計測し、 α 波のパワースペクトルが主観的評価と相関があることを示している[15]。Oohashi らは可視聴域を超える高周波成分を含む音を聞いた際に α 波成分が増大し、高周波成分を除去するとそれが弱まることを示している[13]。

3. 脳波計測実験

この章ではソフトウェアのユーザビリティを定量的に計測するための方法として、著者らが行った、ソフトウェア利用後におけるユーザの脳波を計測した実験について述べる。人間の中枢神経系を非侵襲に計測する方法の中でも脳波計測は比較的安価な装置を用いて行うことができ、他の計測方法と比べてユーザの姿勢に制限が少ないため、コンピュータ利用後における脳の活動を計測するのに適している。

3.1. 概要

本実験では Microsoft 社製のスプレッドシート作成ソフトウェアである Microsoft Office Excel 2003 及び Microsoft Office Excel 2007 を用いて被験者に4種類のタスクを行ってもらい、タスク終了後の脳波を計測した。ほぼ同様の機能を持ちながら異なるインタフェースを持ったソフトウェアを比較することで、機能の違いによる

影響を受けずにインタフェースの違いによるユーザビリティの違いを調べることができる。

実験で使用した全てのタスクは Excel 2003 または Excel 2007 のどちらでも実行が可能であり、全てのタスクで使用するデータファイルは同一のもの（架空の成績表）とした。また、タスクを行う順序については学習効果を考慮し、カウンターバランスを行った。

全タスク終了後に、Excel の利用頻度や、各タスクで利用した機能についてのアンケートを行った。アンケートの一部を図 1 に示す。アンケートは Questionnaire for User Interaction Satisfaction (QUIS)[3] を参考に、各機能や Excel の各バージョンにおけるユーザビリティについての主観的評価や各バージョンの利用頻度を測るための項目を設定した。各アンケート項目は利用頻度について 4 段階、ユーザビリティについて 7 段階の選択式とした。実験で計測する被験者の脳波とアンケート結果の相関を見ることで、主観評価の脳波の関係を確認する。

3.2. 実験環境

脳波計測には脳機能研究所製の感性スペクトル解析装置 ESA-16 を用い、サンプリング周波数を 200Hz としてタスク終了後の 2 分間、閉眼安静状態の脳波を記録した。本実験でタスク後に閉眼安静状態で脳波を計測した理由として(1) α 波は閉眼状態においては減衰し、閉眼状態において増加する[7]、(2)マウスやキーボードを操作することによって脳波にアーチファクトが混入する、という 2 点があげられる。大橋らはタスクによって変化した脳波が 60 秒から 100 秒の間持続することを報告しており[27]、タスク後に脳波を測ることでアーチファクトを含まない、タスクによる脳波の変化を測ることができると考えられる。したがって、本実験では脳波の計測しやすい閉眼安静状態の脳波を計測する。

脳波計測においてはグラウンド電極を額の中心(Fpz)とし、導出法については基準電極導出法（片側耳朶法）を

用いて、基準電極を右耳(A2)とした。計測用電極は後頭部(Pz)に配置した。これは、後頭部には筋電位など脳波以外の生体現象によるアーチファクトが入りにくいためである。また、心電(ECG)を計測するために両前腕の内側にも電極を配置した。その他のアーチファクトに対しては、脳波計測に不要な電子機器の電源を抜く、電極の装着不良を防ぐため被験者に頭部用ネット包帯を着用してもらうなどの対策を行った。

被験者はヘッドレスト付きの椅子に座り、デスクトップ PC 上で Excel 2003 と Excel 2007 をマウスとキーボードを用いて操作してもらった。事前にキーボードやマウスの位置、椅子の高さなどを調整してもらい、筋電位が入りにくいようにした。

3.3. 実験結果

3.3.1. データの解析

PC に取り込まれたサンプリング周波数 200Hz の脳波に対してパワースペクトル解析を行った。始めにそれぞれの被験者について、各タスク終了後の閉眼安静状態における 81.96 秒間の脳波を切り出した。安静開始直後は姿勢が変化し筋電位の影響を受けやすいため、脳波の切り出しは安静開始 10 秒後からとした。次に切り出した脳波に対してフィルタ処理を行った。フィルタはハイパスフィルタ（遮断周波数 3Hz, 減衰傾度 +6dB/oct）、ローパスフィルタ（遮断周波数 60Hz, 減衰傾度 -6dB/oct）、バンドエリミネーションフィルタ（中心周波数 60Hz, 遮断域 47.5Hz~72.5Hz, 次数 2）である。バンドエリミネーションフィルタは交流電流による影響を取り除くために用いた。これらのフィルタ処理を行った脳波に対してハミング窓をかけ、高速フーリエ変換(FFT)を施し、パワースペクトルを求めた。

得られたパワースペクトルから α 波と β 波の脳波全体に対する比率（以降、単に α 波、 β 波とする）と、 β 波を α 波で割った β/α を求めた。 α 波と β 波の周波数成分は国際脳波学会による分類にあわせて、それぞれ 8Hz 以上 13Hz 未満および 13Hz 以上 30Hz 未満とした。また、脳波全体の範囲を 3Hz 以上 30Hz 未満とした。

α 波と β 波の比率は脳の活動を見るための指標としてよく用いられており[17]~[19]、本実験でも β/α をタスク後の心理状態を測るための指標として用いる。脳波における α 波や β 波の比率やその強さは個人差が大きく、絶対値による比較は有用ではない。本実験では各被験者の脳波を個人の平均値で正規化し、比較を行う。

3.3.2. アンケート結果と脳波の相関

アンケートで得られた被験者の主観的評価と脳波の関

3. Excel2007 について									
●利用頻度	全く利用していない	年に数回	月に数回	週に数回以上					
●各機能の使い方は理解しやすい	まったく思わない	1	2	3	4	5	6	7	強く思う
●作業効率が高い（生産性が高い）	まったく思わない	1	2	3	4	5	6	7	強く思う
●表示される情報（ポップアップやメニュー名など）が明確で分かりやすい	まったく思わない	1	2	3	4	5	6	7	強く思う
●インタフェースが好ましい	まったく思わない	1	2	3	4	5	6	7	強く思う
●全体として使いやすい	まったく思わない	1	2	3	4	5	6	7	強く思う
●全体として満足している	まったく思わない	1	2	3	4	5	6	7	強く思う

図 1 実験で使用したアンケート

係を見るため、Excel に対するアンケートの各項目と脳波の相関を分析した。

表 1 にアンケート結果について平均値と標準偏差、および t 検定による検定結果を示す。この表はアンケート項目「作業効率」、「インタフェース」、「使いやすさ」、「満足感」において Excel 2003 と Excel 2007 で有意な差があったことを示しており、Excel 2007 の方がアンケートによる評価が低かったことを示している。

表 2 にアンケート項目に対する回答と脳波の相関を示す。表は脳波の各指標とアンケート項目「利用頻度」、「作業効率」、「インタフェース」、「使いやすさ」、「満足感」に有意な相関があることを示している。特に、「インタフェース」というアンケート項目との相関が高く、 α 波との相関が 0.558、 β/α との相関が 0.510 を示している。

3.3.3. バージョンによる脳波の違い

異なるバージョンの Excel を利用した際における脳波の違いを分析するために、被験者の α 波と β 波、 β/α についてバージョンごとの平均を求めた。

バージョンごとの脳波の違いを図 2 に示す。縦軸は正規化後のパワーの大きさを示しており、1 を超えた値は脳波が各個人の平均よりも大きいことを表し、1 を下回った値は平均よりも小さいことを示している。

この図は Excel 2003 を利用した際の α 波が Excel 2007 を利用している際に比べて大きく、 β 波と β/α は Excel 2007 を利用しているときよりも小さいことを示している。これらの差が有意であるかを調べるために t 検定を行った結果、いずれの差も有意であった(いずれも $p < 0.001$)。

3.4. 考察

本実験で得られた脳波と実験後のアンケートにおける主観的な評価項目「作業効率」「インタフェース」「使いやすさ」「満足感」の間に有意な相関関係が見られた。また、脳波と「利用頻度」の間にも有意な相関関係が見られた。加えて、過去に行われている脳波を用いた研究においても、精神的負荷が多い時や不快に感じている際に α 波が減少し、 β 波と β/α が減少することが観察されており、本実験の結果と一致していることから、タスク終了後の脳波はソフトウェアのユーザビリティ評価に用いることが可能だと考えられる。

Excel 2003 と Excel 2007 を利用した後の閉眼安静状態

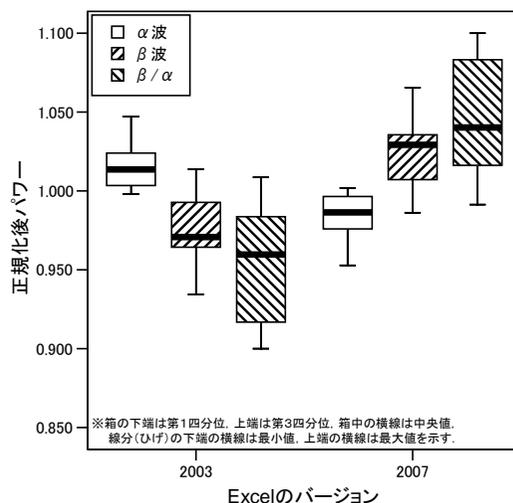


図 2 Excel のバージョンによる脳波の違い

表 1 アンケート結果

		利用頻度	理解が容易	作業効率	情報が明確	インタフェース	使いやすさ	満足感
Excel2003	平均値	3.3	5	5.4	4.8	4.9	5.3	5
	標準偏差	0.82	1.33	1.26	1.32	1.45	1.34	1.15
Excel2007	平均値	1.8	3.5	3.5	4	3.1	3.4	3.3
	標準偏差	1.14	2.17	1.72	1.63	2.13	1.9	1.95
p < 0.05		yes	no	yes	no	yes	yes	yes

表 2 アンケート結果と脳波の相関

		利用頻度	理解が容易	作業効率	情報が明確	インタフェース	使いやすさ	満足感
α 波	pearson's r	0.436	0.438	0.510*	0.297	0.558*	0.496*	0.455*
	p 値	0.054	0.054	0.021	0.204	0.011	0.026	0.044
β 波	pearson's r	-0.493*	-0.31	-0.405	-0.118	-0.437	-0.419	-0.364
	p 値	0.027	0.183	0.077	0.62	0.054	0.066	0.115
β/α	pearson's r	-0.475*	-0.374	-0.462*	-0.212	-0.510*	-0.464*	-0.415
	p 値	0.034	0.104	0.04	0.37	0.021	0.04	0.069

*は危険率 5%で有意差があることを示す。

における脳波に有意な差が見られ、 α 波は Excel 2007 を利用した際に減少し、 β 波と β/α については Excel 2007 を利用した際に増加した。これは Excel 2003 を利用したタスクよりも Excel 2007 を利用したタスクのほうが精神的な負荷が高いためと考えられる。アンケート結果についても Excel 2003 より Excel 2007 において、「利用頻度」「作業効率」「インタフェース」「使いやすさ」「満足感」の項目で評価が低く、脳波による結果と一致している。この結果は被験者の半数が Excel 2007 を初めて利用したためにインタフェースに慣れておらず、操作に戸惑ったり各メニュー項目の構成や名称について考える必要があったためと考えられる。

また、Excel 2003 と Excel 2007 とともに利用頻度が「週に数回以上」と回答していた被験者については α 波、 β 波、 β/α のいずれについても Excel 2003 と Excel 2007 の間で大きな差はみられなかった（正規化後の値でそれぞれ 0.004, 0.011, 0.002 の差）。一方で Excel 2003 の利用頻度が「週に数回以上」で、かつ Excel 2007 を初めて利用した被験者の場合、Excel 2003 と Excel 2007 の間で α 波、 β 波、 β/α のいずれも差が大きかった（それぞれ 0.039, 0.068, 0.106）。このような結果は双方のバージョンに慣れている被験者の場合、どちらを利用していても精神的負荷に違いがないのに対して、Excel 2007 を初めて利用した被験者の場合、インタフェースなどの違いから大きな精神的負荷を受けたためだと考えられる。

本実験で得られた結果から、ソフトウェア利用後における閉眼安静状態の脳波とユーザの主観的評価には関係があり、ユーザビリティの定量的評価のための指標としての脳波が有用である可能性が示された。脳波によるユーザビリティの評価が可能になれば、定量的なデータを元にした分析が可能になり、分析手順の自動化やモデル検証などが可能になるとと思われる。従来行われている定性的評価では分析に時間がかかり、また分析者によって結果が異なる可能性があるが、脳波を用いた定量的評価では分析方法を自動化することで、分析の再現性を高め、分析時間を大幅に短縮することができると考えられる。

本実験においては脳波の指標として α 波、 β 波、 β/α の 3 つを用いて分析を行った。実験の結果、 β/α についてアンケートの各項目と相関が見られ、一部のタスクにおいて、 β/α でバージョン間の違いが見られた。また、Excel 2003 をよく利用しているが Excel 2007 を初めて利用した被験者でも β/α が最も差が大きかった。この結果はニオイが人に及ぼす生理的影響の計測[28]や運転中の覚醒度の計測[29]に用いられている α 波と β 波の比率が、ソフトウェア利用者の状態計測にも用いることができる可能性を示している。

3.5. 本章のまとめ

本章ではソフトウェアのユーザビリティを定量的に評価するための方法として生体情報のひとつである脳波に着目し、その実験手法の確認とアンケートによる主観的評価と脳波の関係を分析するための実験について説明した。実験の結果、バージョンの異なるソフトウェアを利用したタスク後の脳波に有意な差があることが分かった。この結果は従来行われてきた脳波を利用した研究と一致しており、また、実験後のアンケートからもその有用性が確かめられた。

しかしユーザビリティを構成する様々な要因について、それぞれがどれだけ脳波に反映されているかについては今回の実験では明らかにしていない。

今後の実験によってユーザビリティの構成要素について個別に脳波との関係を調べ、脳波によるユーザビリティ評価の有効性を確認していきたい。

また、本実験で設定した実験方法を用いることで、ソフトウェア以外にも Web サイトのユーザビリティ評価や、利用時の快・不快を測ることも可能と思われる。近年インターネット上に公開される Web アプリケーションが増えてきており、それらのユーザビリティを定量的に評価することも有意義な研究であると考えられる。

加えて、本実験で用いた計測方法はソフトウェアに対する「慣れ」自体を定量的に計測する方法としても有用であると考えられる。ユーザビリティの評価では、被験者の慣れによる影響を排除するために事前に訓練や学習を行ってもらい、システムに対する学習状態や知識を均一にする。しかし、実際にどの程度被験者の状態が統一されているのかについては計測する方法がなく、実験者の主観によって学習時間や方法が決められているのが現状である。本実験で用いた方法によって事前学習の過程における脳波の変化を時系列に観察し、その時間ごとの差分を見ることで、被験者の対象システムに対する慣れの変化を定量的に評価し、慣れの影響を排除した状態でのユーザビリティの評価が可能になるとと思われる。

4. 熟練者のノウハウ獲得

専門的な技能が必要とされる分野では熟練者が持つノウハウを解析し、文書化することで教育・訓練に活用することができる。そのため、熟練者の行動を観察したり、インタビューすることでノウハウを獲得する試みが行われている。しかし、外部から見ることでできる行動からは、その行動に至った意図や動機を理解する事は難しい。熟練者のノウハウは、本人の自覚無しに用いられる暗黙知であることが多く、ある行動に至った意図をインタビ

ユーなどで獲得するのは困難である。

このような問題に対して、作業している際の視線移動から熟練者の意図や判断の仕方を計測しようとする試みが行われている。一般的に「目は心の窓」と言われているが、何らかの作業を遂行しようとする人間の判断や意図は、視線の動きに現れるとされている[20]。

視覚から得た情報を元に判断が行われる作業においては、判断に先立って必要な情報を収集するために視線の移動が起こる。そのため、熟練者の視線移動を計測することで熟練者のノウハウを計測できると期待される。

視線を用いた研究の例としては以下の文献が挙げられる[21]～[23]。Lawらは腹腔鏡手術の訓練装置を使用している際の初心者と熟練者の視線を分析している[21]。分析の結果、熟練者は器具を操作している時であっても視線が患部に集中している一方、初心者は器具を見ていることが明らかになっている。Kasarskisらはシミュレータ上で飛行機の着地を行っている際のパイロットの視線を計測している[22]。その結果、熟練者が高度計、速度計、滑走路を短時間ずつ、交互に見るのに対して、初心者は高度計をより見ていることが明らかになっている。CrosbyとStelovskyは誤りを含んだプログラムのソースコードと、正常に動作するプログラムを被験者に示し、ソースコードを理解する過程の視線を計測した[23]。その結果、初心者はソースコード上に書かれたコメントに視線が集中するのにに対して、熟練者はコメントをあまり見ておらず、ソースコードを良く見ていることが明らかになった。これは、コメントが必ずしもソースコードを正しく説明していないことがあるため、熟練者は動作を確実に理解できるようソースコード自体を読んでいると考えられる。

これらの研究はいずれも熟練者と初心者の視線移動を比較し、情報収集の仕方が異なっていることを明らかにしている。熟練者の情報収集の仕方を把握することで、初心者のための行動指針の作成や、教育に役立てることができる。

5. ソフトウェアレビューにおける視線計測実験

この章ではソフトウェア開発における品質向上手法であるソフトウェアレビュー（以下単にレビュー）を対象として著者らが行った視線計測実験について述べる。

レビューはソフトウェア開発時に作成される様々な文書（ソースコード、仕様書、設計書）を開発者が精読することで誤り（バグ）を検出する作業である。開発現場で最も用いられている品質向上手法であり、他の手法に比べても効率よくバグを検出できるとされている。一方、成否が開発者の能力に強く依存するため熟練者の行動を詳細に理解する事は教育・訓練の点からも重要である。

5.1. 概要

実験ではプログラミングの経験が3年以上ある博士前期過程の学生5名を対象とした。いずれの被験者も最低1度のレビュー経験があった。C言語で書かれた12行～23行のソースコードをレビュー対象とした。いずれのソースコードにもコメントは含まれておらず、1つだけ不具合が埋め込まれている。

ソースコードはディスプレイ上に20ポイントのフォントで表示され、スクロールをすることなく全ての行読めるようにした。

5.2. 視線計測環境

本実験ではソースコードを注視する視線を行単位で識別する必要がある。そこで本環境では高解像度で高精度の計測ができるnac Image Technology社¹の非接触型視線計測装置、EMR-NCを採用した。EMR-NCの視線計測頻度は30Hzであり、計測誤差はディスプレイ上の5.4ピクセルに相当する。これは20ポイントのフォントで文書を表示したときの0.25行に相当する。

ソースコードを表示するために21インチ液晶ディスプレイ（EIZO FlexS-can771）を解像度1024×768に設定し使用した。また実験中に作業者の体が動くことにより発生する測定誤差を最小にするため、固定式の椅子を使用した。

EMR-NCにより計測された視線データCSVファイルとして保存される。データはディスプレイ上の絶対座標値で表された視線位置とデータの取得時間である。また、付属ソフトウェアでは一定期間、一定の範囲にとどまった視線である停留点（Fixation）を計算する。停留点は一般に作業者がその位置に注目していると解釈される。

視線計測装置が保存したデータは著者の開発したソフトウェアCrescent（Code Review Evaluation System by Capturing Eye movemeNT）によって行単位のデータに変換される。CrescentはJavaとSWT(Standard Widget Tool)を用いて作成された約4000行のプログラムである。図3にCrescentによる視線表示の例を示す。このウィンドウはレビューが行われたソースコード（左側）と作業者の視線の動きを棒グラフで示したもの（右側）を並べて表示する。また視線の時系列情報に合わせてコードと対応する棒をハイライトすることによって視線の再生が可能である。このとき、任意の時間から再生を行うことや、スロー再生が可能である。また、他のツールを用いた分析が行えるように、行単位の時系列情報をCSVファ

¹ <http://www.nacinc.jp/>

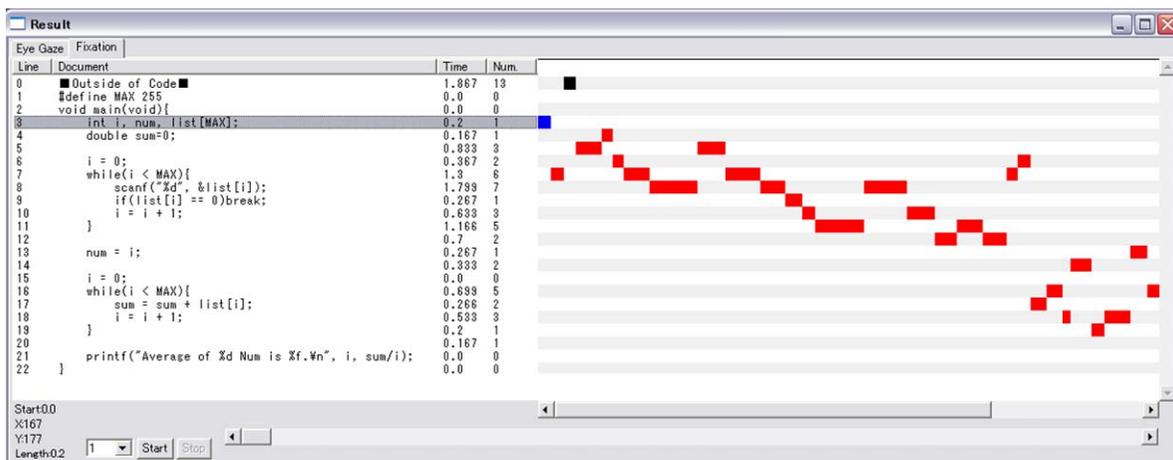


図 3 Crescent による視線表示の例

イルに出力することができる。

5.3. 結果と考察

図 4 および、図 5 に Crescent が出力した時系列情報をグラフにしたものを示す。図の左側に被験者がレビューを行ったソースコード、右側にレビュー時の視線移動を示している。グラフの縦軸はソースコードの各行、横軸は視線の留まった順序を示しており、被験者がどのような順でソースコードの各行を注視したかを示している。図 4 で見られるように被験者 E はソースコード全体を 2 回スキャンした後にソースコードの中ほどに位置している while ループに集中している。図 5 では被験者 C はレビュー開始直後に 2 つある関数のヘッダー部分 (1 行目と 13 行目) を読んでいる。その後、被験者は関数 makeSum(), main() をそれぞれ読み、その後は関数 makeSum() に集中している。このようなレビュー開始時にコード全体を上から下に眺めたり、構造に沿って読んだ後に、特定の範囲を集中して読むという動作が良く見られた。視線情報を分析した結果、平均してレビュー時間のうち初めの 30% の間にコード全体の 72.8% を読んでいることがわかった。また、コード全体のスキャンにかかる時間と、不具合の検出にかかる時間に有意な負の相関が見られた ($r = -0.568$, $p = 0.002$)。この結果はソースコードをより長い時間をかけてスキャンをすることがより効率の良いバグ発見につながっていることを示している。これは以下のように解釈することができる。ソースコードを注意深くスキャンするレビュー作業者はスキャン中によりプログラムの構造を正しく理解し、バグが埋まっていると思われる行を識別することができる。一方、スキャンを十分に行っていない作業者はプログラムの構造を正確に把握することができず、ソースコード中の重要な行を見落としたり、バグとは関係のない行に集中したりするため、レ

ビューの効率を落としている。

これ以外にも初めて変数が現れた時にその変数の宣言文を確認する動作や、処理が行われた変数が前に使われた行を確認する動作などが確認できた。このような視線の動きは被験者が変数の値を思い出す様子や、変数の値を再計算する様子が反映されていると考えられる。

加えて、定量的な結果ではないが、実験後に行ったインタビューで被験者自身の視線データを見せることで、何を考えて特定の行を長時間確認したか、どのように不具合のある位置を特定したのかといった、通常のインタビューで得られない詳細なコメントが得られた。

この結果はレビュー中の視線の動きが作業者の思考に関する情報を多く含んでいることを示している。したがって視線情報を用いることによって、レビュー実験において従来用いられてきたインタビューの効果をより高めることができる。また、レビュー熟練者の視線を記録し初心者に見せることで、教育・訓練に用いることができると考えられる。

5.4. 本章のまとめ

本実験ではソースコードレビューにおけるレビュー作業間での性能差を明らかにするために (1) レビュー中の視線を記録・再生する実験環境の構築、および (2) 構築した環境を用いた実験を行った。

実験では被験者の視線からレビュー開始時にコード全体を眺める動作 (スキャン) が確認でき、レビュー開始時に十分なスキャンを行った被験者はコード内のバグを早く見つける傾向があることがわかった。

今回の実験結果から、作業者の視線からコードレビューの様子が分析できることがわかった。被験者に作業中の思考について発言してもらう発話法や、作業中に定期的にインタビューを行うことで被験者の様子を観察する

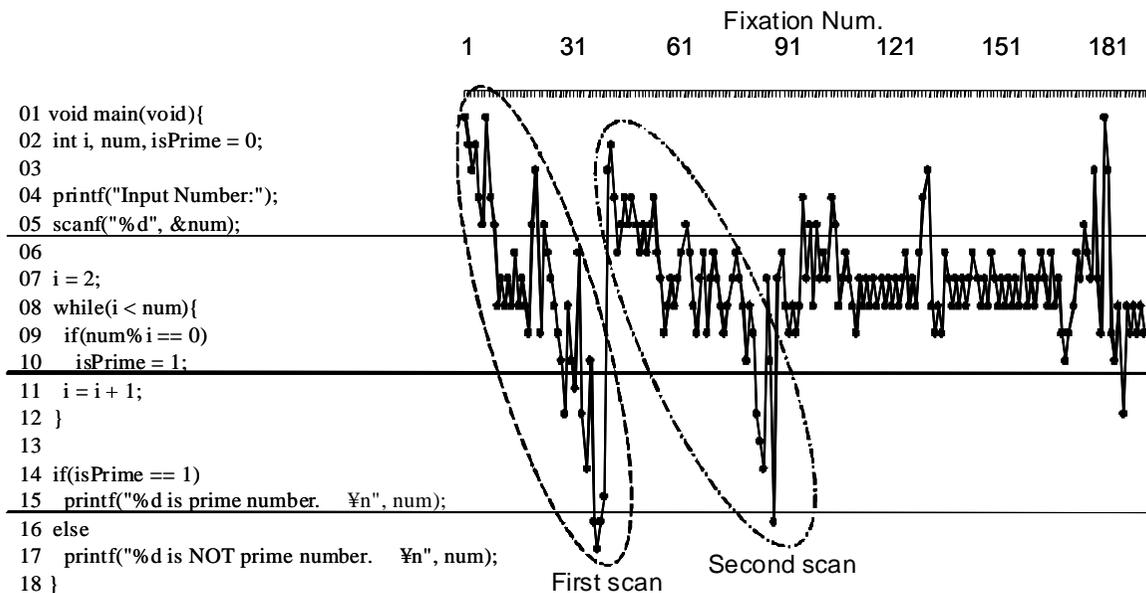


図4 スキャンパターンを含む視線 (被験者 E : Prime)

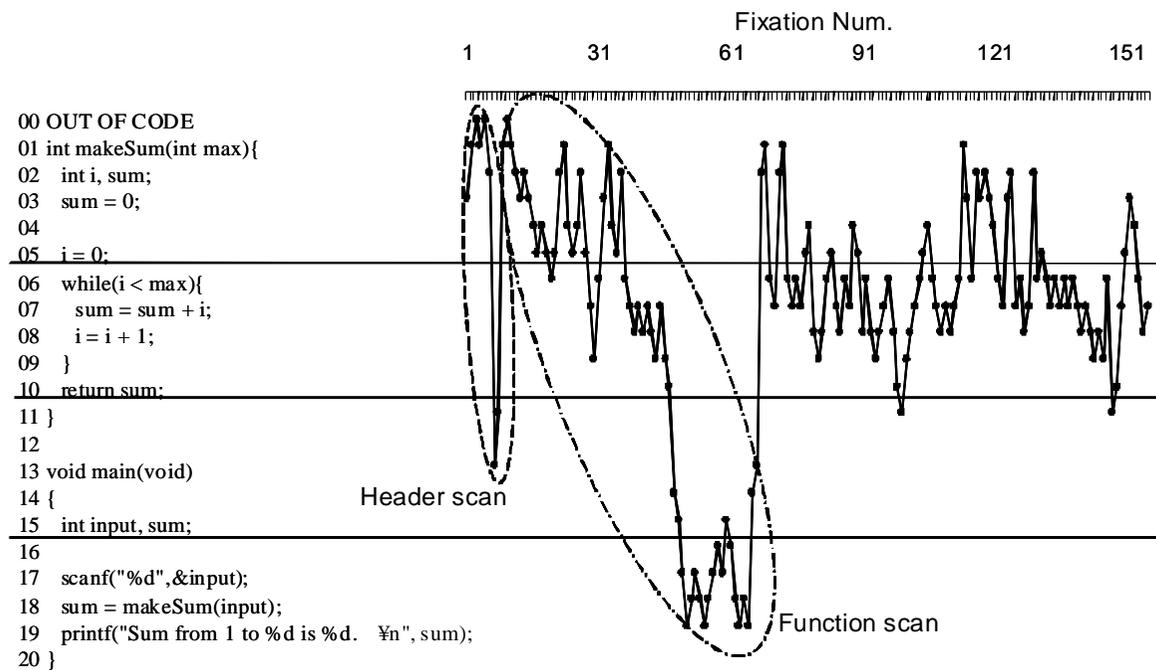


図5 スキャンパターンを含む視線 (被験者 C : Accumulate)

方法に比べ、視線を用いた観測は作業者の行動を妨げず、より自然な状態での記録が可能であり、レビューの観察に適した方法であるといえる。また、被験者が意識していない動作についても記録・分析が可能のため、熟練者が持つ暗黙知の顕在化に特に有用であると考えている。

6. おわりに

本稿ではソフトウェア工学の分野で行われている、生体情報を用いた研究の紹介を行うと共に、著者がこれま

で行ってきた実験について詳説した。生体情報を用いた研究は今までの盛んに行われており、視線計測とその応用に関する国際会議も存在している²。生体情報を利用するためには専用の装置が必要であるが、視線計測については数千円の Web カメラで計測が可能なソフトウェアが公開されるなど[24]、利用しやすい環境が整いつつあると

² ETRA Eye Tracking Research & Applications, 隔年の開催で最新は 2010 年。 <http://http://etra.cs.uta.fi/>

いえる。

本稿で紹介した生体情報の計測によるユーザビリティの評価および熟練者のノウハウ獲得は医療分野においても応用が可能であると考えている。例えば、ベテランのノウハウが重要である業務内容について、生体情報やアンケート、行動観察によるノウハウの抽出とその結果に基づいた教育方法の確立や、業務システムのユーザビリティ評価に基づいた、改善などが挙げられる。

また、病院全体をシステムととらえ、システム利用者である患者の行動分析による、病院環境の定量的評価と改善などにも応用可能だと思われる。医療分野を全く知らない専門外からの意見ではあるが、何らかの参考になれば幸いである。

参考文献

- [1] Ericsson, K. A., Simon, H. A.: Protocol Analysis: Verbal Reports as Data, MIT Press, 1993.
- [2] Osgood, C. E., Suci, G. J., Tannenbaum, P. H., The Measurement of Meaning, University of Illinois Press, 1957.
- [3] Chin, J. P., Norman, K. L., Shneiderman, B.: Subjective User Evaluation of CF PASCAL Programming Tools, Technical Report (CAR-TR-304), 1987.
- [4] Hart, S. G., Staveland, L. E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, Human Mental Workload, Elsevier, 139-183, 1988.
- [5] 芳賀繁, 水上直樹: 日本語版 NASA-TLX によるメンタルワークロード測定, 人間工学, 32(2), 71-79, 1996.
- [6] 岡田謙一, 西田正吾, 葛岡英明, 中谷美江, 塩澤秀和: ヒューマンコンピュータインタラクション, オーム社, 2002.
- [7] 宮田洋, 藤澤清, 柿木昇治, 山崎勝男: 新生理心理学-生理心理学の基礎, 北大路書房, 1998.
- [8] 村田厚生: 情報検索作業における精神的な作業負担の測定, 電子情報通信学会論文誌. A, 基礎・境界, 74(4), 706-714, 1991.
- [9] 水野統太, 野村収作, 野澤昭雄, 井出英人: メンタルワークロードにおける情動ストレスの評価, 信学技報, HIP2007-27, 107(60), 143-147, 2007.
- [10] Hazlett, R. L.: Measurement of User Frustration: A Biologic Approach, CHI '03: Extended Abstracts on Human Factors in Computing Systems, 734-735, 2003.
- [11] Nakayama, M., Takahashi, K., Shimizu, Y.: The Act of Task Difficulty and Eye-Movement Frequency for the 'Oculo-Motor Indices, ETRA '02: Proceedings of the 2002 Symposium on Eye Tracking Research & Applications, 37-42, 2002.
- [12] 松永久, 中沢弘: 満足感計測のための基礎的研究-主観的な満足感と前頭部双極誘導による脳波の関係, 人間工学, 34(4), 191-201, 1998.
- [13] Oohashi, T., Nishina, E., Honda, M., Yonekura, Y., Fuwamoto, Y., Kawai, N., Maekawa, T., Nakamura, S., Fukuyama, H., Shibasaki, H.: Inaudible High-Frequency Sounds Affect Brain Activity: Hypersonic Effect, The Journal of Neurophysiology, 83(6), 3548-3558, 2000.
- [14] 松永久, 中沢弘: メンタルワークロードと脳波により評価される満足感との関係, 日本機械学会論文集. C編, 66(648), 2884-2890, 2000.
- [15] 林秀彦, 國藤進, 宮原誠: 高品位映像の評価: 脳波を指標とする客観評価法, 映像情報メディア学会誌: 映像情報メディア, 56(6), 954-962, 2002.
- [16] 大橋力: マルチメディアと脳, 電子情報通信学会論文誌. D-II, 情報・システム, II-情報処理, 79(4), 468-475, 1996.
- [17] 寺内文雄, 久保光徳, 大釜敏正, 青木弘行: 針葉樹材のニオイが随伴性陰性変動(CNV)に及ぼす影響, 材料, 45(4), 397-402, 1999.
- [18] Hong J. Eoh, Min K. Chung, Seong-Han Kim: Electroencephalographic Study of Drowsiness in Simulated Driving with Sleep Deprivation, Industrial Ergonomics, 35(4), 307-320, 2005.
- [19] Karel A. Brookhuis, Dick D. Waard: The Use of Psychophysiology to Assess Driver Status, Ergonomics, 36(9), 1099-1110, 1993.
- [20] 荻阪良二, 中溝幸夫, 古賀一男: 眼球運動の実験心理学, 名古屋大学出版社, 1993.
- [21] B. Law, M. S. Atkins, A. E. Kirkpatrick, A. J. Lomax, and C. L. Mackenzie: Eye Gaze Patterns Differentiate Novice and Expert in a Virtual Laparoscopic Surgery Training Environment, In Proceedings of ACM Symposium of Eye Tracking Research & Applications (ETRA), 41-48, 2004.
- [22] P. Kasarskis, J. Stehwen, J. Hichox, A. Aretz, and C. Wickens: Comparison of Expert and Novice Scan Behaviors During VFR Flight, In Proceedings of the 11th International Symposium on Aviation Psychology, 2001.
- [23] M. E. Crosby, and J. Stelovsky: How Do We Read Algorithms? A Case Study, IEEE Computer, 23(1),

24-35, 1990.

[24] J. San Agustin, H. Skovsgaard, E. Mollenbach, M. Barret, M. Tall, D.W. Hansen, and J.P. Hansen:

Evaluation of a Low-Cost Open-Source Gaze Tracker, ETRA '10: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, 77-80, 2010.