

卒業論文 2021年度（令和3年度）

脈波を使った時系列UX評価手法に関する研究

慶應義塾大学 環境情報学部

佐々木 雄司

全世界インターフェースデザイン（増井研究会）

2022年1月

卒業論文 2021 年度（令和 3 年度）

脈波を使った時系列 UX 評価手法に関する研究

論文要旨

人々にとって身近なサービスがオンライン化されるにつれ、誰にとっても使いやすいシステムを設計する重要性が指摘されている。近年では、機能性よりもむしろ使いやすさや体験を広告する製品も見られ、UX の向上がビジネスに直結する課題となっている。使いやすいシステムを設計するには、実ユーザに対するユーザテストによって UX を評価し、製品の改善に繋げるサイクルが重要である。さらに、子供から高齢者まで利用者層が広がることで想定すべきユーザの種類は多くなりユーザテストはより困難になる。そもそも、UX 評価を行っているソフトウェア開発企業は少なく、行っていたとしても評価した結果が製品の改善に繋がっているケースがほとんど無いと言われている。

そこで本研究では、現在は大規模な開発でしか行われていないユーザテストの導入ハードルを下げ普及させるために、UX 評価や統計処理についての専門的な知識が無くても容易にユーザテストを行いシステムの問題点を発見できるようにするシステムの開発を行った。テストユーザの行動や生理的反応に基づく評価指標は確立された手法が無いため普及していないものの、分析の自動化と相性が良いため手法さえ確立できれば容易なユーザテスト手法になり得る。そこで簡便であり高精度に測定できる可能性のある、容積脈波のカオス揺らぎからストレスを測定する技術を応用し、新たな UX 指標を提案し有用性を明らかにした。

また、UX の概念では時間変化を考慮することが重要である。これまでに黒須の UX グラフなど UX の時間変化を記述する方法が提案されてきたが、使用前、使用後、習熟後までの数週間から数ヶ月レベルの期間で大まかに示すものだった。本研究では、シーケンシャルにデータを取得できる生理的反応を指標として採用することでシステム使用中の数分から数時間程度の間の変化を記録し可視化するシステムを開発し有効性を明らかにした。

キーワード

UX, 脈波, UI, ユーザテスト, ユーザビリティ, 人間中心設計

慶應義塾大学 環境情報学部

佐々木 雄司

Abstract Of Graduation Thesis Academic Year 2021

Studies on the UX Evaluation Method Using Pulse Waves

Summary

As more and more services become online, the importance of designing user-friendly systems has been pointed out. In recent years, some products advertise UX rather than functionality, and improving UX has become an issue directly related to business. In order to design a user-friendly system, it is important to evaluate UX through user testing on real users, and to use this cycle to improve the product. Furthermore, as the range of users expands from children to the elderly, the number of user types increases, making user testing more difficult. To begin with, it is said that few software development companies conduct UX evaluation, and even if they do, there are few cases where the evaluation results lead to product improvement.

Therefore, in order to familiarize user testing, which is currently only conducted in large-scale effect, we developed a system that enables users to easily conduct user testing and find problems in the system without any specialized knowledge of UX evaluation or statistics. Although evaluation metrics based on the physiological reactions of test users are not widely used because there is no established method, they have the potential of the easy way because they are compatible with the automation of analysis. Therefore, we proposed a new UX evaluation method using a technology to measure stress from chaotic fluctuations of the volumetric pulse wave of a fingertip, which is simple and can be measured with high accuracy and clarified its usefulness.

In addition, it is important to consider the time variation in the concept of UX. Some methods have been proposed to describe the temporal change of UX, such as Kurosu's UX graph, but they show the time period roughly from a few weeks to a few months. In this study, we developed a system to record and visualize the changes from a few minutes to a few hours during using a system by adopting physiological responses, which can be obtained sequentially, as an index and clarified its effectiveness.

Keywords

User Experience, Pulse Wave, User Interface, User Test, Usability, Human Centered Design

Faculty of Environment and Information Studies
Keio University

Sasaki, Yuji

目 次

第 1 章 序論	1
1.1 背景	1
1.2 目的	4
1.3 本論文の構成	5
第 2 章 関連研究と諸概念の整理	7
2.1 ユーザビリティ	7
2.2 UX	8
2.2.1 満足性	9
2.2.2 時間相	9
2.3 UX メトリクス	11
2.3.1 センサを用いた生理メトリクスの測定	11
2.3.2 ストレスの測定	12
2.4 容積脈波 (BVP)	13
2.4.1 自律神経バランス (ANB)	14
2.4.2 最大リヤブノフ指数 (LLE)	14
2.5 問題の所在	16
第 3 章 満足性と ANB 及び LLE との関係性	17
3.1 実験方法	17
3.1.1 実験マテリアルの開発	17
3.1.2 タスクの実施	19
3.1.3 容積脈波 (BVP) の計測	20
3.2 結果	20
3.3 分析	21
3.4 考察	24
第 4 章 時系列ユーザビリティ評価システムの提案	25
4.1 評価システムの設計・開発	25
4.1.1 時系列ストレスグラフ	25
4.1.2 観察ビデオとのマッピング	25
4.2 評価システムの有効性についての実験	25
4.2.1 対象と手続き	25

4.2.2 結果と考察	25
第5章 結論	27
5.1 満足性評価のための指標の提案	27
5.2 評価システムの提案	27
5.3 今後の課題	27
謝辞	28
参考文献	30
付録A 付録	37

図 目 次

1.1 黒須の品質特性図 [1]	3
2.1 SQuaRE を基にした JIS X 25010:2013 の品質モデル 画像入れ直すこと	8
2.2 UX のライフサイクル [2]	10
2.3 指尖用光電式容積脈波計	13
2.4 耳朵用光電式容積脈波計	13
2.5 PPG で取得した脈波	13
2.6 図 2.5 の脈波から再構成したアトラクタ	15
3.1 実験マテリアル: 数字の表示画面	19
3.2 実験マテリアル: 数字の読み込中画面	19
3.3 Lyspect での分析画面	20
3.4 Lyspect での分析画面	21
4.1 実験マテリアル: テンキー	25
4.2 実験マテリアル: ドラッグ	25
4.3 実験の様子	26

表 目 次

3.1	10 Usability Heuristics for User Interface Design[3]	18
3.2	満足性と ANB 及び LLE との関係性についての実験結果	22
3.3	ANB 及び LLE と申告に基づく満足性との関係	23

第1章 序論

1.1 背景

近年、一般市民にとって身近なサービス機能がオンライン化されるにつれ、ソフトウェアのユーザビリティの重要性が認識されるようになっている [1]. 総務省の情報通信白書によると 2020 年までに世界のモバイル向けアプリ市場は売上高で 1,924 億ドルとなっており、今後も拡大が予測されている。これらの市場はモバイルゲームが牽引してきたが、今後はそれに加えて学習や翻訳、健康管理、SNS などのアプリケーションも成長が見込まれている [4]. これらはいずれも消費者向けのサービスであり、業務用サービスに比べ習熟を求めることができず利用者数も多くなる傾向にある。このようなモバイル向けアプリ市場ではデザインや使い勝手が製品の購買の決め手になるためユーザビリティがより重要になる。近年では、機能だけではなく、むしろ使いやすさのコンセプトを全面に押し出して製品を広告したり、製品やブランドの魅力として使いやすさを訴える企業が増えている [5]. 家電メーカーの Balmuda 社長の寺尾は次のように述べており、機能性よりも製品から受ける体験を重視した製品を目指している。

現代を生きる私たちは、家電や携帯電話、クルマなど、さまざまな便利な道具に囲まれて暮らしています。しかし、便利であればそれで良いのでしょうか？人生に本当に必要なのは、驚きや感動、うれしくなるような体験なのだと思います [6].

また、個人のアプリ利用者の年齢層が年々広がっており、幅広い年齢層にとって使いやすいサービス設計が必要になってきている。インターネットの利用率を年代別に見ると平成 19 年末時点で 6-12 歳が 68.7%，50-59 歳が 81.2%，60-64 歳が 63.0%，65-69 歳が 36.9% だったのでに対し、令和 2 年では 6-12 歳が 80.7%，50-59 歳が 94.7%，60-69 歳が 82.7% と利用者数の伸びが顕著である [7][8]。利用者の年齢層が広がることで嗜好や身体の状態、前提知識に大きな幅が生まれることが考えられる。サービス開発者には様々な利用者を想定して設計することが求められ、ユーザビリティの高いデザインの開発がより困難になる。

ユーザビリティについては古くから検討されており、Shackel[9] はユーティリティを高めることと同程度にユーザビリティを高めることが重要であると主張している。ユーティリティは機能性 (functionality) とも言い換えられ、「機能があっても使いにくいコンピュータ」に対して「使いこなせるか」ということを示すためにユーザビリティの概念を提唱した [1]. その後、ユーザビリティは体系化され、ISO/IEC 25000(別名 SQuaRE) の規格の中に取り込まれ、ソフトウェアの品質基準のひとつとなっている。

Nielsen のユーザビリティエンジニアリング原論ではユーザビリティに配慮した開発では次の工程が例として挙げられている [10].

1. ユーザー調査
2. 競合製品との比較分析
3. パラレルデザイン
4. ユーザー参加型デザイン
5. トータルインターフェースのコーディネートデザイン
6. ガイドライン・ヒューリスティック評価
7. プロトタイピング
8. インタフェース評価
9. 反復デザイン
10. インストールしたシステムのフォローアップ調査

この工程では、実際の開発であるプロトタイピングの前にヒューリスティック評価、その後インタフェース評価とフォローアップ調査と3つのポイントでユーザビリティ評価を行っている。このように、ユーザビリティの高いシステムを開発するには、ヒューリスティクス（経験則）に基づいてユーザビリティを検討しデザインするのに加え、実際に開発したシステムのインターフェースのユーザビリティを評価し、修正する反復的な作業が必要になる。さらに、システムの導入後も調査を行い、さらに反復的に改善していくことが重要である。

ユーザビリティの評価では、ユーザに直接使用させて評価することは欠かすことができない重要な工程である。Nielsen のヒューリスティック評価 [11] では、多くのユーザが共通して使いやすいと感じるであろう項目を挙げ、それらを満たしたシステムを開発することでユーザビリティの向上を目指している。しかし、この手法では適用できる範囲に限界があるだけでなく、そもそもユーザの多様性を考慮して設計することができない。ここでの多様性とは、障害者や高齢者というだけでなく、年齢や性別などの特性、嗜好や価値観などの指向性、精神状態や物理的環境などユーザのあらゆる違いを含んでいる [12]。これらのユーザそれぞれがシステムの利用に際して起こることについてはヒューリスティクスでは網羅できていない。そこで、製品のターゲットとなるユーザを被験者として集め、実際にシステムを操作してもらうユーザテストが行われる。

ユーザテストでは、ユーザビリティに関連する様々な指標を測定することでユーザビリティを評価する。測定手法はパフォーマンスマトリクス、自己申告メトリクス、行動・生理メトリクスに分類することができる [5]。パフォーマンスマトリクスとは、タスク成功率、タスク時間、エラー頻度、効率、学習可能性などで、定量的に計測できるため測定が容易であ

る。自己申告メトリクスはリッカート尺度¹による質問紙や自由記述などのアンケートであり一般的に行われている。行動・生理メトリクスは、言語行動と非言語行動を観察したりセンサー等を用いて観測するものである。具体的にはビデオの録画や筋電位センサー、アイトラッキング、皮膚伝導率、心拍数などが指標として使われている[5]。これらの測定手法は、複数の手法を組み合わせて補完的に活用していく必要がある。

近年、UXが話題になっていることに代表されるように、ユーザの主観的側面が注目されており、ユーザテストへの導入も検討されている。前述の測定手法では、自己申告メトリクスと行動・生理メトリクスが主観的側面に注目した手法といえる。Jordanは前述の機能性とユーザビリティに加えて“嬉しさ”的重要性を述べており、ユーザは機能がありユーザビリティが満たされると嬉しさを要求すると述べている[13]。黒須はISO/IEC 25010:2011の品質特性の図を改良し図1.1のような品質特性図を提案している[14]。この中では、設計品質をUI、利用品質をUXと便宜上分けており、さらに利用品質の中に定量的に測定可能な客観的利用品質と直接測定できない主観的利用品質を置いている。利用品質のうち、客観的利用品質についてはパフォーマンスマトリクスといった手法で測定が可能になっており、主観的利用品質については自己申告メトリクスや行動・生理メトリクスが活用できる。自己申告メトリクスだけではユーザが全てを申告できなかったりバイアスがかかりやすいことから十分とはいえない、行動・生理メトリクスなど他の測定手法を併用する必要がある。しかし、ユーザの主観を正確に測定する方法は様々なものが提案されているものの一般に普及しているとはいえない。

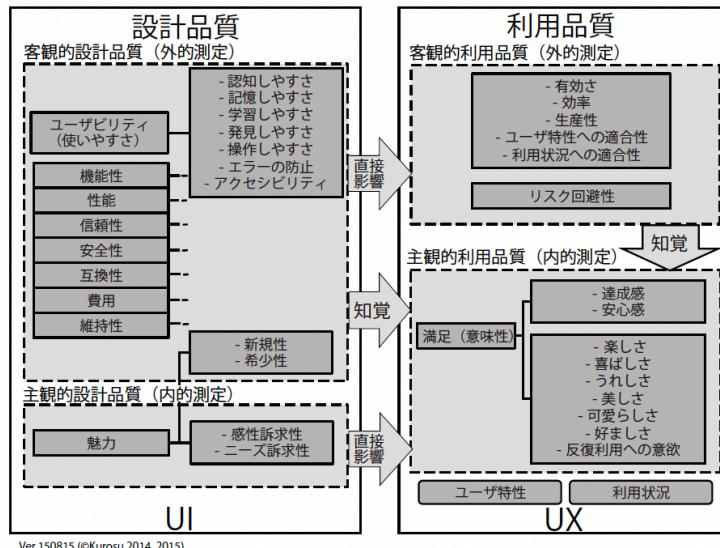


図 1.1: 黒須の品質特性図 [1]

このように、ユーザビリティ及びUXの評価はユーザにとって使いやすいシステムの開発に不可欠なものであり、その評価手法は定着しているもの、定着していないものを含めて

¹項目について合意または非合意を5段階等で回答させるアンケート手法

様々に存在しているが、実際の開発では UX の評価は大規模な消費者向けサービスを除いてほとんど行われていないか、行われていても改善に役立てられていないのが現状である。黒須 [1] は、

著者が企業における実態を調べると、そもそも UX 調査を実施していることが稀というほど少なく、さらに実施していても、それを担当した部署と企画や分析を担当する部署との間の連携が取れておらず、せっかく取得した実利用に関する情報が企画やユーザ理解にもとづく具体化に役に立っていないことがわかった。

と述べている。また、UX 品質保証サービスを提供する企業が 2020 年にソフトウェア開発企業の社員 490 名を対象に行った調査では、UX 向上に取り組んでいると応えた割合は 34% に留まっており、UX の評価を行っている割合は更に少ないと考えられる。実ユーザの体験を調査せずにユーザの満足性が高いシステムを開発することはできないため、小規模の開発でも UX 評価を行えるようにする必要がある。

また、UX はユーザ毎に異なるだけでなく、使用の前、使用中、習熟後などのステージや折々の出来事によって変化するため、変化を考慮して評価する必要がある [1]。黒須はこのような UX の変化を記述、視覚化するために UX グラフ [14] を提案している。この UX グラフはユーザが製品を数週間から数ヶ月といった長期間にわたって使用した体験を、縦に満足性、横を時間軸にしてカーブで表すものである。一方で、システムを使用している数分間から 1 時間程度の期間の変化についてはユーザに変化を記述させることが難しく UX グラフは使われていない。しかし、UX の概念では、時間軸の変化は重要な要素であり、短期間の使用時であってもそれを評価する方法が必要であると考えられる。

以上のことから、(1) 現在は大規模な開発でしか行われていないユーザテストの導入ハードルを下げ普及させること、(2)UX の重要な部分である満足性を測定する簡便な手法を開発すること、(3)(2) で測定した満足性から UX の時間変化を記録し可視化する手法を開発することが求められる。

1.2 目的

本研究では(1) 現在は大規模な開発でしか行われていないユーザテストの導入ハードルを下げ普及させるために、UX 評価や統計処理についての専門的な知識が無くても容易にユーザテストを行いシステムの問題点を発見できるようにするシステムの開発を目指す。行動・生理メトリクスは確立された手法が無いため普及していないものの、分析の自動化と相性が良いため手法さえ確立できれば容易なユーザテスト手法になり得る。ユーザビリティテストでは、被験者は笑ったりそわそわしたり調査票に記入するよりもはるかに多くの行動を行うがそれらを記録することでより情報量の大きいデータ入手可能になる [5]。そこで、(2) 及び(3) の手法を用いて統合的な分析システムを開発する。

(2) 満足性を測定する新たな簡便な手法として、容積脈波でストレスを測定する手法の満

満足度評価への応用を提案する。行動・生理メトリクスは満足度と関係があると考えられ、これまでにも皮膚伝導率や心拍数を用いてストレスで満足度を評価する手法が提案されてきた。容積脈波は、カオス解析することで鬱やパーキンソン病の診断に活用できる可能性が指摘されている[15][16]。さらに、同様の解析により精神状態を測定できる可能性[17]が指摘されていることから、従来の皮膚伝導率や心拍数などのシンプルな方法よりも高精度なストレス測定が期待できる。また、脳波を用いてストレスを測定しユーザビリティを評価する手法[18]も提案されているが、装置が大がかりになるために導入にコストがかかったり、測定器を装着すること自体がストレスになる可能性がある。そのため、耳朶に装着する小型のセンサーで容積脈波を取得しストレスを測定、満足度を評価する方法を提案する。これに基づき、容積脈波を解析して得たストレス指標と従来のユーザビリティ指標の関係を示した。

短期間のUX変化を黒須のUXグラフで表記させることは難しいが、容積脈波から得られるストレス指標のような生理メトリクスであればシーケンシャルであるため(3)使用中のUXの時間変化を記録し可視化することが可能になる。容積脈波を解析しストレスを算出するシステムであるLyspect[19]では、ストレス指標の時系列変化を出力することができなかった。そこで容積脈波をリアルタイムで取得し、解析して得られるストレス変化を時系列グラフで表示するシステムを開発した。そして、開発したシステムの有効性を示した。

1.3 本論文の構成

本論文の構成を示す。

第1章では本研究の背景について述べた。第2章では関連研究と諸概念を整理する。第3章では脈波によるストレスチェックを応用したユーザビリティ評価手法を提案し、第3章ではそれを発展させた時系列ユーザビリティ評価手法を提案する。そして、それらの有効性を示す。最後に、第5章の結論では本研究を総括し、考察と展望を述べる。付録として、本研究で行った実験で得られたデータを添付する。

第2章 関連研究と諸概念の整理

2.1 ユーザビリティ

英単語としての usability は use(使う)+able(できる)+ity(こと)から構成されており、「使うのに便利で実用的な」という意味で 14 世紀から使われていた [20][1]。しかし、概念としてのユーザビリティがアカデミアで取り上げられるようになったきっかけはコンピュータの登場と普及だった [1]。コンピュータはトレーニングを受けた専門家が使用するものとして作られたが、一般への普及が図られるなかで使いやすさを検討する必要性が出てきたのだと考えられる。

Shackel によると彼以前に Mirror や Bennet がユーザビリティに言及しているが、彼らはユーザビリティを「使いやすさ (ease of use)」とほぼ同義として扱っていた [9][1]。Nielsen は、ユーザビリティを複数の要素の上位概念であるとし、下位概念として学習可能性 (Learnability), Efficiency(効率性), Memorability(記憶のしやすさ), Errors (エラー), Satisfaction(満足度) を挙げている [21]。

国際規格としてユーザビリティが定義されたのは ISO9241 シリーズ「人間とシステムの相互作用の人間工学 (Ergonomics of Human-System Interaction)」であり、ISO9241-11:1998 ではユーザビリティを「特定のユーザーが、特定の使用状況において、特定の目標を達成するために、製品を有効、効率的かつ満足に使用できる度合い。 (Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.)」と定義している。この定義は ISO の関係者やベヴァンらが普及活動に力を入れた結果標準としての地位を獲得するにいたった [1]。この定義は現在でも一般的な技術者に浸透していると考えられる。しかし黒須は、客観的に計測可能な有効性や効率性と主観的なものである満足性を同列に扱うことについて疑問を呈しており、有効かつ効率的であれば満足性が上がるため満足性はより上位の概念であると主張している [1]。

ISO/IEC9126-1:2001 や ISO/IEC25010:2011 SQuaRE(System and Software Quality Requirements and Evaluation) では体系化が進められており、ユーザビリティが品質モデルの一部に組み込まれた形となっている。SQuaRE では、図 2.1 のように品質モデルが製品品質モデルと利用時の品質モデルに分けられており、ユーザビリティは製品品質の下位概念に位置づけられている。そして、有効性、効率性、満足性は利用時の品質の中に置かれている。

SQuaRE の改訂では、設計上の品質である製品品質とユーザが使用した際のユーザにとっての品質である利用時の品質とに分けた点が重要である。人間の思い通り予想通りに動いてくれるシステムを設計しようとする人間中心設計 [12] の考え方の普及などにより、設計時に

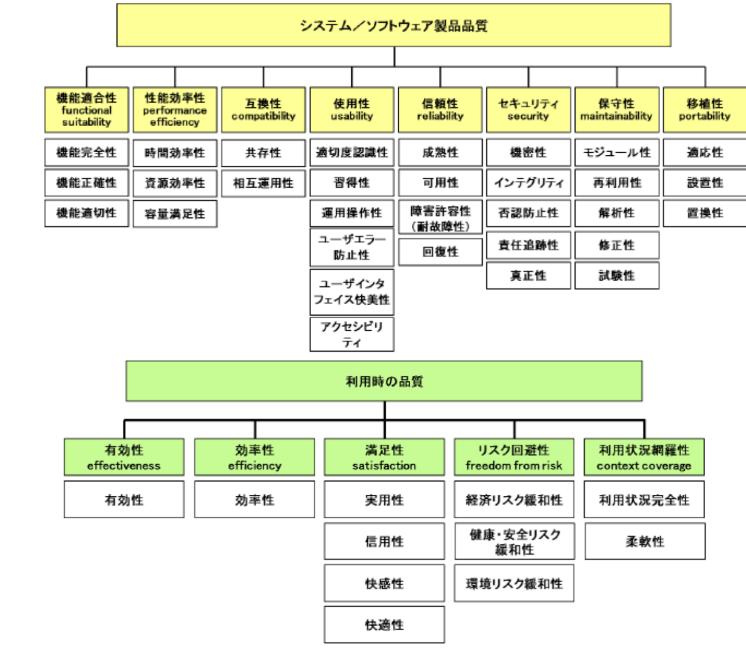


図 2.1: SQuaRE を基にした JIS X 25010:2013 の品質モデル 画像入れ直すこと

ユーザの存在をこれまで以上に重要視する動きの中でユーザの利用時に特に注目する必要が出てきたと考えられる。この考え方は後述の UX の概念に関わっており、一方でユーザビリティはユーザに目を向けたものではなく製品設計レベルの概念だということになる。

本研究では、ユーザがシステム利用時にどのような生理的反応を見せるかということに注目しており、SQuaRE の定義でいうところの利用時の品質やその下位概念の満足性を測定することを目指しているといえる。第 1 章では、Nielsen のヒューリスティック評価はユーザや利用シーンの多様性を考慮しておらず不十分であると述べた。確かにニールセン自身が定義したユーザビリティには満足度が含まれていたためにユーザによってユーザビリティが異なることが考えられた。しかし SQuaRE の定義では、ユーザビリティは実ユーザの存在から離れ特定ユーザ・特定の使用状況・特定の目的において満たされていればよく、Nielsen のヒューリスティック評価が有効になると考えられる。一方、本研究で扱うユーザテストでは、ユーザに目が向けられており、SQuaRE の定義におけるユーザビリティという概念は当てはめることができない。そこで、SQuaRE の利用時の品質などの概念を拡張した概念である UX について検討する。なお、本論文ではユーザビリティという用語は基本的に使用せず、使用する場合は「使いやすさ」という一般的な意味でのみ使用する。

2.2 UX

User Experience という言葉は、1995 年には Apple のヒューマンインターフェースグループで使われており [22]、ヒューマンインターフェースやユーザビリティといった言葉ではカバーできない工業デザインのグラフィック、インターフェース、物理的なインタラクション、マニュ

アルなど人がシステムを利用する際のあらゆる側面を含む言葉として発明したと Norman がインタビューで述べている [23]. しかしその後、UX という言葉の普及に伴ってあらゆる事象に使われるようになり本来の意味が不明瞭になりつつあった. そこでロトらによって UX 白書で UX の定義が定められた. UX 白書では、UX とは “「一般的な意味における経験」とは異なり、システムと出会うことにおける経験” であるとしている [2].

UX 白書による UX の定義に「システムと出会うこと」という文言が含まれている事からわかるとおり、UX は製品の設計やシステム自体についての概念ではなくユーザと製品との間に生まれるものであり、前述の SQuaRE の利用時品質に関連が深いことがわかる. また、UX 白書では UX に影響する要素 (factor) として、ユーザとシステムに加えて文脈を挙げている [2]. UX はこれらの 3 つの掛け合わせで表現されるものでどれかが変化すれば UX も変化するといえる.

UX は製品が最終的にユーザにどのように受け取られるかという重要な概念ではあるが、一方で製品設計においては「UX を向上させる」ということが必ずしも正しいとは限らない. UX はシステムに対してユーザの数 × 文脈の数だけ存在するためそれら全てを向上させることは不可能だからである. 実ユーザの UX を考慮して、ビジネス上の目標あるいは技術上の制約に沿うように製品設計に落とし込み、その上で製品品質を向上させる必要がある.

2.2.1 満足性

黒須の品質特性図 (図 1.1) にあるように、UX は SQuaRE の利用時品質に対応しており、それには客観的利用品質と主観的利用品質が含まれる. 客観的利用品質は測定可能な指標であり既に評価手法の研究は十分なされているため本研究では取り上げない. また、客観的利用品質は主観的利用品質の原因であり、主観的利用品質は客観的利用品質の要素を知覚した結果が含まれたものであると見ることができる. そのため主観的利用品質の評価がより重要になる.

主観的利用品質には、達成感、安心感、楽しさ、喜びしさ、嬉しさなどの感情が含まれており、これらと従属関係を持ち最高位に位置する言葉として満足が挙げられている [24]. つまり、満足性を測定することで主観的利用品質を評価することができ、知覚という形で客観的利用品質も満足性の度合いに含まれるため UX 全体を評価することに繋がる可能性がある. このように、何らかの指標によって満足性を測定することが UX を評価するうえで重要である.

2.2.2 時間相

ISO の SQuaRE に対して UX 白書の UX の定義では時間相が明確になっている. どれくらいの期間に焦点を当てるかという視点で分類すると一時的 UX、エピソード的 UX、累積的 UX に分けられる. また、これに使用前の期待などを指す予期的 UX を含め図 2.2 のようにライフサイクルとして捉えることもできる [2].

一時的 UX システム使用中の特定のインタラクションに対する感情の変化

エピソード的 UX 利用エピソードに関する評価

累積的 UX システムをしばらく使い利用中や利用していない時間を通して形成される評価

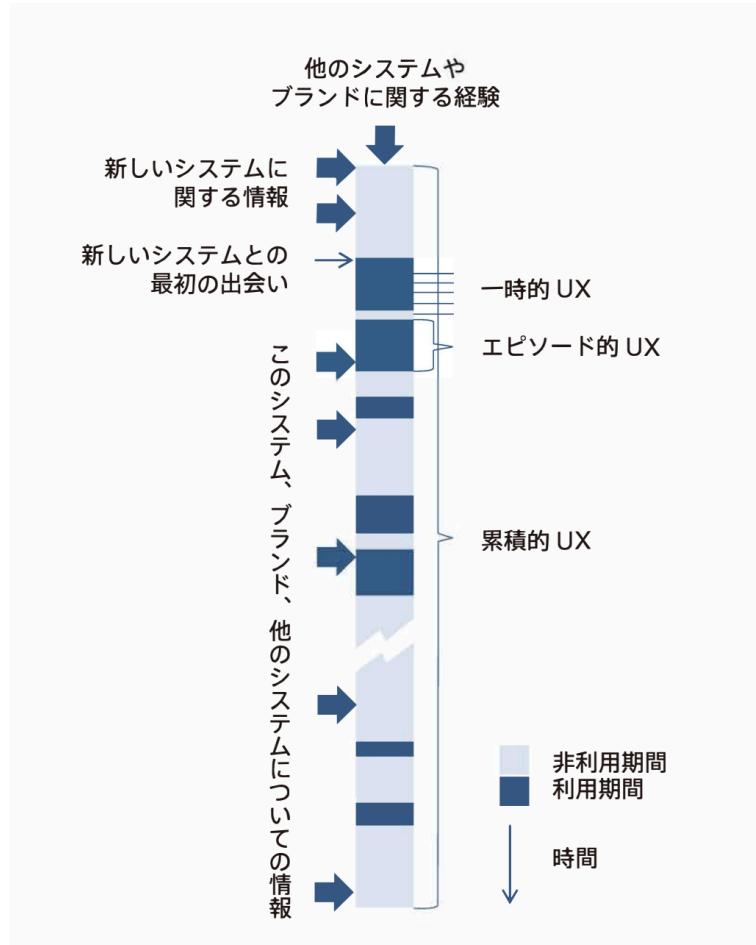


図 2.2: UX のライフサイクル [2]

UX の変化を時系列で捉えるための方法として黒須が UX グラフ [14] を提唱している。これは、入手前の期待、入手時の感情などを含めて製品を使用した際のあるいは使用していないときの出来事とその満足度レベルを列挙したうえで、縦軸に満足度レベル、横軸に時間をとってグラフ化するものである。これはユーザ自身に記入させるものでユーザにとっての UX 変化が累積的 UX のレベルで可視化することができる。一方、エピソード的 UX の中で、個別のインタラクションに対する一時的 UX まではユーザに思い出しながら書いてもらうことが困難であり全て書き切れるとは言えない。また、評価対象は満足度という主観的なものではあるものの、ユーザに自己申告させる方式では正しく評価できない可能性を排除できない。

2.3 UX メトリクス

UX メトリクスとは、UX を何らかの数的指標として測定するものであり、UX を製品設計に活用するためには欠かせない手法である。UX メトリクスには、パフォーマンスマトリクスや自己申告メトリクス、行動・生理メトリクスなどがあり [5]、これらを組み合わせて評価することが必要であると前に述べた。利用時の品質特性に当てはめるとパフォーマンスマトリクスは客観的利用品質の測定であり、自己申告メトリクスや行動・生理メトリクスは主観的利用品質の測定であるといえる。パフォーマンスマトリクスはよく知られており有用であるものの、システムの全ての部分を評価することはできないため、問題がありそうな部分や変更を加えようとする一部分のみを切り出して測定することになる。自己申告メトリクスは質問紙や記述式アンケートに代表される手法であり、前述の UX グラフもその一例である。主観的利用品質つまり満足度を評価するもう一つの手法である行動・生理メトリクスは質問紙では把握しきれない細かな情報を生かすことができると期待される。

行動・生理メトリクスはユーザのあらゆる行動や生理反応を記録しようとするもので、様々な手法が研究されている。まず、システムを使用しているユーザを直接あるいはビデオで記録したものを観察し、言語行動や非言語行動を抽出することは一般的な方法である。言語行動としては、システムを褒めたり否定したりする直接的なコメントに加えて、「どこを押せばいいんですか?」などの質問などが考えられる。非言語行動としては、顔の表情や集中力、いらだち、眼を細めて読もうとするなどの行動から読み取ることができる。このような方法は特別なシステムが必要なくいつでも始められる一方で観察者の能力によって読み取る情報の精度にばらつきができることが考えられる。

2.3.1 センサを用いた生理メトリクスの測定

観察するための機器が必要な指標としては、顔の表情、アイトラッキング、瞳孔の直径、ストレス [5] などがある。

顔の表情は直接観察することもできるが、システムを使うことで自動的にかつ取りこぼしなく計測しようとする研究がなされている。Hazlett らは 2007 年に顔面筋電位を用いて表情を分類する研究を行っている [25]。現在であれば表情をビデオで収録し機会学習によって分類すれば筋電位センサのような大がかりな装置を装着することなく表情を分析することが可能であると考えられる。Staiano らはシステム使用時の映像に対しインタラクション毎に表情の注釈をつけるシステムを提案しており、映像の表情分類には機械学習を使用している [26]。

アイトラッキングは tobi 社が tobiipro[27] というアイトラッカー及びその分析ツールを発売している。事前のキャリブレーションが必要なほか、機器が高額であるため通常のシステム開発で活用されているとは言えないが UX 評価での実績はある。アイトラッキングで行える分析には特定の要素やエリアを見たユーザの割合、特定の要素やエリアを見た時間、特定の要素に気づくまでに要した時間、眼の動きの距離、注視の時間などがある [5]。瞳孔の直径についての情報は多くのアイトラッカーで得ることができる。瞳孔反応は明るさのレベ

ルに反応する以外に認知処理や興奮、興味の高まりに相関することがわかっている。しかし、瞳孔反応は様々な心理状態に反応するためその理由が肯定的なのか否定的なのかを分類することが難しいと言われている[5]。

ストレスの測定は満足性に密接に関係すると考えられ、様々な手法が検討されている。本研究で扱う容積脈波もストレスの測定の一種であるといえる。他の様々なストレス評価手法については後述する。

2.3.2 ストレスの測定

ストレスの測定とは、ストレスと相関があるとされている生理反応を測定しその度合いでストレスを判断するもので、様々なセンサとそれを用いた指標が提案されている。

皮膚伝導率 (GSR: Galvanic Skin Response) はストレスが上がり発汗量が増えることで皮膚の電気抵抗が変化することを用いたもので、タスクパフォーマンスとの相関が明らかになっている[28]。皮膚伝導率計は手に装着するため自由な指の動きや手での操作を阻害しユーザビリティ評価に使うことができなかった。そこで、手のひらの一部分のみを覆い、指の動きを阻害しない皮膚伝導率計である Galvactivator[29] が提案されている。しかし、ユーザビリティを評価するうえで少しであっても手に異物を装着することは避けるべきと考えられる。

心拍変動 (HRV: Heart Rate Variability) は心電図 (ECG: Electrocardiogram) を用いて測定するものでストレスと関連があることが知られている。これをユーザビリティ評価に活用する研究は古くから行われており、Rowe らの研究では心拍数 (HR: Hert Rate) の変化が目標処理能力の低下する時点を示す可能性を指摘している[30]。光電式容積脈波記録法 (PPG: Photoplethysmography) を用いて計測する脈拍数変動 (PRV: Pulse Rate Variability) でも心拍変動と同様の結果が得られることがわかっている[31]。医療機関や研究機関では ECG の使用が可能であるが、機器が大がかりになるため通常の開発では使用することができなかった。そこで最近ではスマートウォッチにも搭載され日常的に使用が可能な光電式容積脈波を用いた手法に期待できる。

脳波 (EEG: Electroencephalography) は脳の電気信号を脳波計によって記録するもので、様々な刺激に対して反応することが知られている。Amaral らの研究では簡単とラベリングされたタスクと難しいとラベリングされたタスクについて F,C,P,O 領域の 8 つの電極から得られた 7 つの周波数帯の平均パワースペクトル密度を特徴量としてサポートベクターマシン分類器で評価したところ有意な差が見られたとしている[32]。

ストレス評価には、これらの他にも様々な手法が検討されているが、一般的なユーザビリティ評価でストレスを測定することはほとんど無い。ストレスの評価に使用する機器が大がかりであったり高価であったりするうえ、センサを装着すること自体が被験者のストレスになり正常な測定ができなくなる可能性があるからである[5]。しかし、容易に計測できるようになればシステムの主観的品質・満足性を他の方法に対して比較的ダイレクトに近い情報を取りこぼしなく得られる可能性がある。

2.4 容積脈波 (BVP)

光電式容積脈波 (PPG) は血中のヘモグロビンが吸収する波長の光を照射し反射光や透過光の変化からヘモグロビン量を測る手法である [33]. 脳波 (ECG) と比較して装置が小型かつ使用が容易であり、脳波や皮膚伝導率 (GSR) のような使用時の違和感が少ないことが特徴である。実際の装置は図 2.3 や図 2.4 である。図 2.3 は指先に装着するカフ式のものであり、中央のクリップを人差し指に装着する。図 2.4 は耳朶に装着するもので黒い部分が耳朶に装着するクリップになっており、右下に移っている本体は Bluetooth でコンピュータに接続する。耳朶用を使用すればシステムの操作時に邪魔にならない。



図 2.3: 指尖用光電式容積脈波計



図 2.4: 耳朶用光電式容積脈波計

光電式容積脈波 (PPG) から得られる容積脈波 (図 2.5) を BVP(Blood Volume Pulse) といい、前述のように脈拍変動 (PRV) を取得することができる。ユーザビリティ評価で容積脈波を測定する際には脈拍変動を使用することが多かった。しかし、容積脈波にはそれ自体に多くの情報が含まれており、心理学系分野では脈拍変動に限らない様々な検討されている。そもそも脈拍変動は自律神経系 (ANS: Autonomic Nervous System) の影響を受けており、脈拍変動ではその変化の一部を捉えることができる [34].

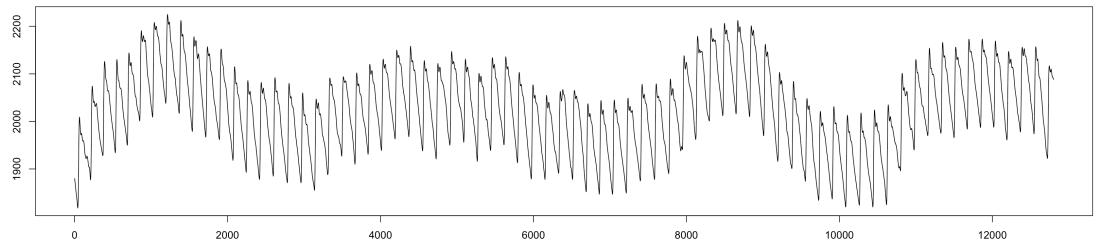


図 2.5: PPG で取得した脈波

自律神経系 (ANS) はヒトの精神状態を反映していることが知られており、これを詳細に分析することでより正確にストレスを測定できる可能性がある。心拍変動 (HRV) は自律神経指標として知られているが、心電図の R 波と R 波の間隔 (RRI: R-R Interval) のことで

あり，RRI のゆらぎは安静時に大きくなり，ストレス時に低減する [35]. RRI は容積脈波の Peak to Peak と概ね同期するため，ここでは容積脈波から得るものも RRI と表現する.

2.4.1 自律神経バランス (ANB)

RRI のゆらぎは 0.15-0.4Hz の高周波 (HF: High Frequency) と 0.004-0.15Hz の低周波 (LF: Low Frequency) の成分に分けられ，それぞれ副交感神経活動，交感神経活動の指標として考えられている [36]. HF 成分の 0.15-0.4Hz は呼吸周波数 (多くの人の呼吸数は秒間 9-24 回) であり，呼吸による RRI の変動に対応している. この値は副交感神経が活性化すると高くなる [37]. LF 成分は血圧変動であるメイヤー波に対応しており，交感神経と副交感神経の両方の活動を反映している. LF に対する副交感神経の影響は 0.15Hz 以下の深呼吸をしている間に現れる. そのためリラックスしている際は副交感神経の増大によって LF の値が高くなる [38].

LF と HF との比をとることによって HF 成分と LF 成分のバランスを表すことができ，このバランスを自律神経バランス (ANB: Autonomic Nervous Balance) という. 副交感神経は HF と LF 両方に影響するため ANB は交感神経の活性度と考えられる [36]. 一般には数値が高いと交感神経優位，低いと副交感神経優位を表すため，高いほどストレスが高いといえる.

しかし，HF，LF，ANB で示される値はいくつかの算出方法があり，算出方法によって値が変化する. そして，自律神経以外の要因でも心拍変動は起きるためこの値だけで判断することはできない. また，この値を計算するには 3 分から 5 分程度の計測が推奨されており [35]，ユーザビリティ評価での活用には制約がある. 本論文では ANB を次のように定義する.

$$ANB = \frac{LF}{HF} \quad (2.1)$$

HRV を使ったユーザビリティ評価に関する研究は多いが，ANB に言及している論文は少なかった [39][40]. HR の測定に比べ ANB の分析はシステムの精度や専門性の高さから利用へのハードルが高い割に ANB の優位性が十分に説明されていないことが原因ではないかと考える. ANB と他指標の比較検討を行い有用性を検証する必要がある.

2.4.2 最大リヤプノフ指数 (LLE)

最大リヤプノフ指数 (LLE: Largest Lyapunov Exponent)，カオティック¹な情報である脈波変動 (BVP) について計算するとそのゆらぎの度合いを示す指標となる [19]. な自律神経バランス (ANB) は長年にわたって研究されており心理学系の研究等で実績があるものの，周波数領域の解析であるため 3 分以上の脈波計測データの使用が推奨されるように短時間で

¹カオス (混沌) とも言い，パワー・スペクトルがいくつかの鋭いピークを持ちながらも，それとは別に連続的成分をある幅にわたって持つときこれをカオティックな運動と呼ぶ [41]

の計測が難しい。本研究では一時的 UX[2] の変化を捉えようとしており、この制約が障害になる可能性がある。そこで、時間領域で解析を行う LLE に着目した。

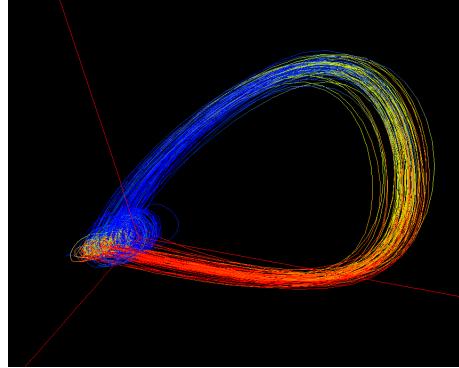


図 2.6: 図 2.5 の脈波から再構成したアトラクタ

リヤプノフ指数は、トラジェクトリー²の分離の速さを表すことで、カオティックな流れの時間発展の追跡を容易にすることができる [41]。例えば脈波のトラジェクトリーは図 2.6 のようになる。実験データにおいては時間成分が含まれるため、期間内のリヤプノフ指数の最大値である LLE を用いる。リヤプノフ指数は実験データから次のように導出される [42]。

$$LE = \lim_{t \rightarrow \infty} \lim_{\epsilon \rightarrow 0} \frac{1}{t} \log \frac{|\delta X_\epsilon(t)|}{|\epsilon|} \quad (2)$$

$$\delta X_\epsilon(t) = X(t) - X_\epsilon(t) \quad (3)$$

$$\epsilon = X(0) - X_\epsilon(0) \quad (4)$$

雄山らは、この LLE で脈波変動 (BVP) のゆらぎを解析する手法を提案している [43]。そして、その指標が不安、恐怖、安堵といった精神状態と関係があることを明らかにしている [44]。雄山は著書の中で、LLE が大きいのは肯定的には行動的で積極的な状態・否定的には不安定で危なっかしい状態であり、小さいのは変化を好みいかたくなな状態・外部適応が困難な状態と表現している [45]。

LLE による容積脈波解析に関する研究では、LLE が精神状態の分類のほかにもストレス度の評価や様々な疾患の診断に使用できる可能性が示唆されている。Hu らの研究では、学生の 4-6 時間の学習時に LLE と ANB の変化を測定しそれらの関連を調べた結果、ストレスを示すと考えられる LLE は低下し、ANB は変動することがわかった [42]。また、認知症の程度やパーキンソン病の有無と LLE に関連があることもわかっている [46][47]。ユーザビリティに関する分野ではタスクの判断ミスや操作ミスと LLE の関連について調べた研究 [48] があるが、ユーザの集中力を調査することに主眼が置かれておりシステムのユーザビリティの調査を目的としたものではない。また、小孫は論文で LLE を UX 評価に活用することを提案しているが実装には至っていない [49]。

²空間中の軌跡

2.5 問題の所在

上述のように、UX評価では満足性を測定することが重要である。満足性を測定しうる手法として機器を用いた生理メトリクスの測定があるが、通常の開発でそれらが行われている例は稀である。自己申告メトリクスの補完として生理メトリクスが有効であるにもかかわらず活用が浸透していない理由として次が考えられる。

- 測定機器が高価かつ大がかりであり用意することが難しい
- 測定機器の装着自体にストレスが伴う
- 測定機器の装着によって手の動きなど動作が阻害される
- 測定や分析に特別な知識や技術が必要
- 測定の精度が不十分または明瞭な結果が得られない

また、UX評価を行っている場合であっても時間相を考慮した評価はほとんど行われていない。研究としてもUXグラフ[14]のような累積的UXではなく、エピソード的UXの時間変化を分析することはあまり行われてこなかった。

以上のことから(1)容易に測定可能で専門知識がなくても結果を解釈しやすい新たな指標の提案、(2)エピソード的UXレベルでのUX変化を記録し分析できるようにすることの2点が求められる。そこで、(1)新たな指標として容積脈波による自律神経バランス(ANB)及び最大リヤプノフ指数(LLE)を提案、(2)提案指標の変化を指標としてUX変化を記録するシステムの開発を行った。

第3章 満足性とANB及びLLEとの関係性

本章では、後述の時系列 UX 評価システムの開発に先立ち、容積脈波によるストレス指標と満足性の関連を把握するために行った予備実験について述べる。新たなストレス指標としては自律神経バランス (ANB) と最大リヤブノフ指数 (LLE) を検討した。また、従来の指標として心拍数 (HR) も合わせて計測した。

3.1 実験方法

実験では、Nielsen のヒューリスティクスに基づいて作成した 2 つの UI を使ったタスクを被験者に実施してもらい、それぞれを使用している際の ANB と LLE、HR を比較した。ヒューリスティクスで Bad UI とされているものは多くの場合満足性が低くなると考えられるものの、はユーザビリティ評価に使用するものであり、それがそのまま満足性に直結するとはいえない。そこで自己申告メトリクスを併用し、どちらの UI がより好ましかったかについて回答を求めた。

実験は 2020 年 3 月に 17 歳から 74 歳までの 16 名 (男性 13 名、女性 3 名) を対象に行った。

3.1.1 実験マテリアルの開発

UX 評価のマテリアルとして、Nielsen のヒューリスティクスで推奨されている UI と避けべきとされている UI をそれぞれ便宜上 Good UI と Bad UI と名付け、サンプルとなるアプリケーションを開発した。今回対象としたヒューリスティクスは “Visibility of system status” (システムの状態の可視性)[11] である。10 のヒューリスティクス (表 3.1) のなかでアプリケーションの機能に関係ないものや学習が必要なもの、ユーザの予備知識に関係するものなど今回の実験に適さないものを排除して選択した。

原題	邦訳	例
Visibility of system status	システムの状態の可視化	プログレス表示、選択状態表示
Match between system and the real world	システムと実世界の対応	コントロール操作部とパーナーの配置の一一致
User control and freedom	ユーザによる制御と自由度	Undo と Redo のサポート
Consistency and standards	一貫性と標準	OS メーカーによるガイドラインへの準拠
Error prevention	誤操作の防止	破壊的操作時にキャンセルをデフォルトにする
Recognition rather than recall	思い出さずとも認識できること	必要な機能がメニューに並んでいること
Flexibility and efficiency of use	フレキシブルで効率的	ショートカットキー
Aesthetic and minimalist design	美しいミニマムなデザイン	ほとんど必要とされない情報を表示しないこと
Help users recognize, diagnose, and recover from errors	エラーの認識、診断、回復の支援	適切なエラーメッセージの表示
Help and documentation	ヘルプとドキュメント	ヘルプの充実

表 3.1: 10 Usability Heuristics for User Interface Design[3]

具体的にはプログレス表示があるものを Good UI, 無いものを Bad UI とした. マテリアルの画面は図 3.1 の通りである. 画面にはノイズにならないように数字を表示するラベルと数字を変更するボタンのみを配置した. また, 右上のスイッチでスタッフが Good UI と Bad UI を切り替えられるようにした. アプリケーションの機能は NEXT ボタンをタップするとランダムで 0-5 秒後にランダムな数字の表示が切り替わるだけのシンプルなものである. Good UI では NEXT ボタンをタップすると図 3.2 のようにプログレスを表示するインジケータが表示され, システムが次の数字を読み込み中であるという状態を表示する. 一方で Bad UI では NEXT ボタンを押してもその時点では何の反応もなく, 0-5 秒後に数字が切り替わるまで上手くタップできたかどうかがわかりにくくなっている.

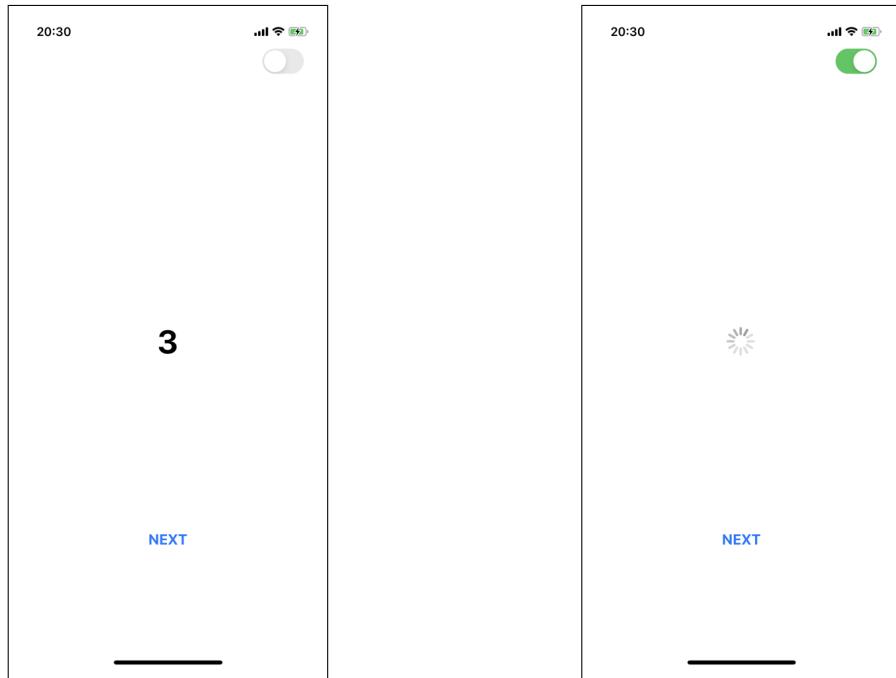


図 3.1: 実験マテリアル: 数字の表示画面

図 3.2: 実験マテリアル: 数字の読み込み中画面

3.1.2 タスクの実施

被験者にマテリアルを操作してもらう際には, バイアスがかからないようにダミータスクを用意したうえで Good UI と Bad UI それぞれを順番に実施してもらった. なお, Good UI と Bad UI の間には 3 分間の安静時間を設け, それぞれのマテリアルを操作する順番はカウンターバランスをとった.

ダミータスクは, 1 分間で NEXT ボタンをタップして表示される数字をできるだけ多く書き取るものにした. できるだけ多く書くようにと指示することで, 競争意識によって手を抜きにくくなるほか, 早く表示してほしいと思わせるように設計した.

タスクの指示

今から1分間のテストを2回行ってもらいます。このアプリでは、NEXTボタンを押すと数字が読み込まれますので紙に表示された数字を書いてください。書き終わったらもう一度NEXTボタンを押して次の数字を読み込んで紙に書いてください。制限時間は1分ですのでできるだけ沢山書くようにしてください。ただし、NEXTボタンを押してから数字が読み込まれるまで少し時間がかかりますので待ってください。

3.1.3 容積脈波 (BVP) の計測

BVPの計測は光学式指尖容積脈波計(PPG)のLifescore Quickプローブ(型番LQ-11, 図2.3)と専用の分析システムであるLyspect[50]を使用した。Lyspectでは300HzでBVPを記録し、LLEの平均とHF及びLFの値、HRの平均を取得した(図3.3)。被験者には実験の開始から終了まで利き手と逆の人差し指にセンサを装着してもらい休憩中も外さないようにした。なお、計測はマテリアルを使用している1分間の間のみとし1人当たり計2回記録した。

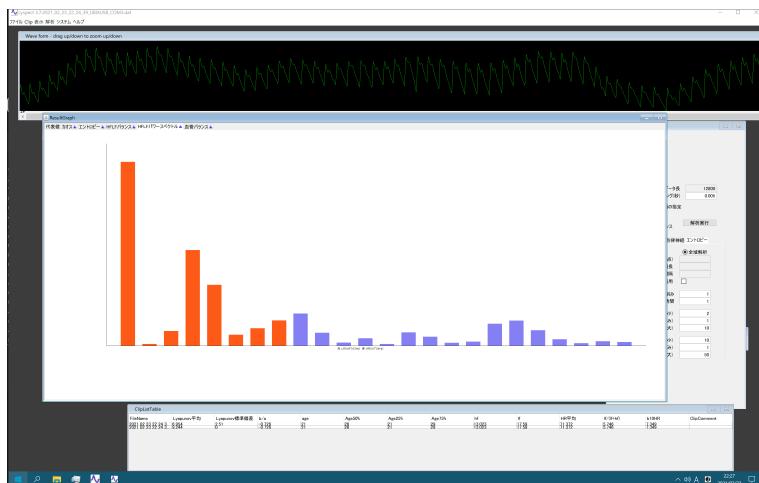


図 3.3: Lyspect での分析画面

3.2 結果

実験の結果を表3.3に示す。カウンターバランスを考慮し、半数の8名はGood UI - Bad UIの順で実施、残りの被験者には逆の順番で実施した。なお一部の被験者には実験開始直後にセンサのズレを直したことが原因と考えられるBVPの大きな乱れが見られ、その場合にはBVPをトリミングしてLyspectを用いて分析した(図3.4)。

簡単な自己申告メトリクスとして「どちらのUIのほうが使いやすいと感じたか」と質問して得られた結果を表の申告列に表示した。今回ヒューリスティクスに基づいて設定した

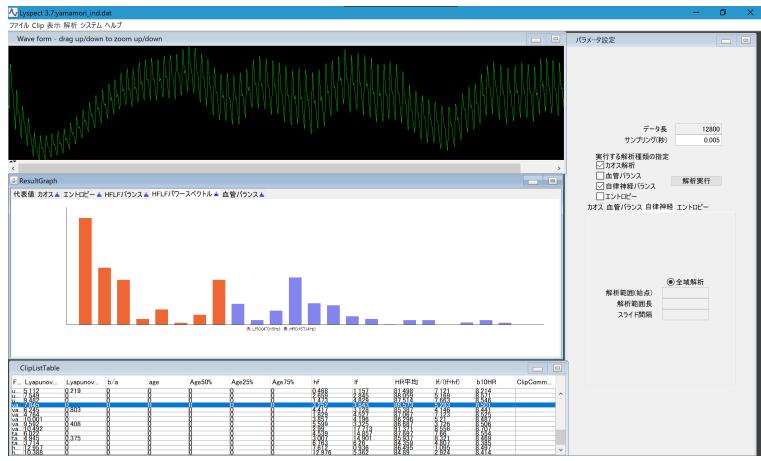


図 3.4: Lyspect での分析画面

Good UI は設計としてはユーザビリティが高いと考えられるが、本研究では満足性との関連を調査するため客観的利用品質であるユーザビリティよりも主観的利用品質を優先した。そのため「使いやすいと感じた」と申告された UI をその被験者にとっての満足性の高い UI とする。

3.3 分析

ANB と LLE, HR を申告に基づいた満足性の高かった UI と低かった UI それぞれに対応させ、満足性との一致を調べた。前述のように ANB は大きいほうがストレスが高く、LLE はゆらぎの大きさを示し小さい方がストレスが高く、HR は大きい方がストレスが高いと考えられる。これに基づいて ANB と LLE, HR それぞれについて満足性が高かった UI のほうがストレスも低かった場合に一致したとみなした。その結果、表 3.3 のように ANB は 16 件のうち 2 件を除いて一致し、LLE は 16 件中 9 件しか一致せず、HR は 6 件しか一致しなかった。

ANB, LLE, HR それぞれについて満足性の高い UI と満足性の低い UI の値に差が無いかどうか検討するために、満足性を独立変数とし ANB, LLE, HR をそれぞれ従属変数とした対応のある t 検定を実施した。その結果、ANB では、満足性の高かった UI ($M = 1.84, SD = 1.56$) と満足性の低かった UI ($M = 2.91, SD = 2.20$) の間で ANB に有意な差が見られた ($t(15) = -2.14, p = .02, d = .54$)。一方、LLE では、満足性の高かった UI ($M = 6.85, SD = 62.64$) と満足性の低かった UI ($M = 6.61, SD = 2.68$) の間で LLE に有意な差は見られなかった ($t(15) = -.67, p = .25, n.s.$)。また、HR でも、満足性の高かった UI ($M = 79.33, SD = 8.96$) と満足性の低かった UI ($M = 78.30, SD = 8.97$) の間で HR に有意な差は見られなかった ($t(15) = 1.00, p = .17, n.s.$)。

No.	年齢	性別*	実験順†	申告‡	Good UI				Bad UI				
					HF	LF	ANB	LLE	HR	HF	LF	ANB	LLE
1	44	M	G - B	B	1.30	3.19	2.46	5.26	68.80	2.21	1.87	0.84	6.12
2	18	M	G - B	B	6.76	6.26	0.93	3.71	88.19	3.01	14.90	4.96	4.95
3	48	F	G - B	G	1.22	1.13	0.93	3.25	77.177	0.43	1.32	3.09	4.37
4	55	M	G - B	G	2.39	9.10	3.82	2.74	67.70	1.09	4.33	3.99	2.51
5	26	M	G - B	G	5.60	3.33	0.93	9.59	86.69	2.99	17.71	5.92	10.49
6	24	M	G - B	G	18.00	13.38	0.74	9.44	63.23	28.57	9.55	0.33	9.62
7	26	M	G - B	G	10.10	8.21	0.81	7.49	63.35	6.50	9.04	1.39	6.29
8	19	F	G - B	G	4.72	3.55	0.75	7.52	79.023	10.42	9.04	0.87	10.43
9	74	F	B - G	G	1.27	1.49	1.17	5.08	85.39	0.47	1.16	2.47	5.11
10	17	M	B - G	G	0.91	2.19	2.40	9.50	79.55	1.12	2.88	2.57	7.87
11	24	M	B - G	G	4.42	3.13	0.71	6.25	87.68	1.83	4.53	2.48	4.76
12	38	M	B - G	G	2.87	8.41	2.93	4.31	79.79	3.81	12.68	3.33	4.51
13	35	M	B - G	G	0.61	3.04	4.98	6.05	82.89	0.73	6.55	8.95	5.78
14	20	M	B - G	G	3.14	7.10	2.26	6.88	84.36	4.46	18.10	4.06	5.17
15	22	M	B - G	G	7.61	0.94	0.12	12.96	86.50	12.98	5.36	0.41	10.39
16	20	M	B - G	G	2.66	2.85	1.07	7.55	88.06	1.47	4.83	3.28	9.48

*Mは男性, Fは女性

†GはGood UI, BはBad UI

‡申告は聞き取りに対して「使いやすかった」と答えたもの

表 3.2: 満足性と ANB 及び LLE との関係性についての実験結果

No.	申告	ANB			LLE			HR		
		満足性高	満足性低	一致	満足性高	満足性低	一致	満足性高	満足性低	一致
1	Bad UI	0.84	2.46		6.12	5.26		68.10	68.80	不一致
2	Bad UI	2.40	2.57		9.50	7.87		88.19	84.42	不一致
3	Good UI	0.93	3.09		3.25	4.37	不一致	77.18	82.88	
4	Good UI	3.82	3.99		2.74	2.51		67.70	67.39	不一致
5	Good UI	0.93	5.92		9.59	10.49	不一致	86.69	91.37	
6	Good UI	0.74	0.33	不一致	9.44	9.62	不一致	63.23	64.09	
7	Good UI	0.81	1.39		7.49	6.29		63.35	63.69	
8	Good UI	1.17	2.47		5.08	5.11	不一致	79.02	81.50	
9	Good UI	0.71	2.48		6.25	4.76		85.39	87.07	
10	Good UI	2.93	3.33		4.31	4.51	不一致	79.55	76.40	不一致
11	Good UI	4.98	8.95		6.05	5.78		87.68	81.64	不一致
12	Good UI	2.26	4.06		6.88	5.17		79.79	69.02	不一致
13	Good UI	0.75	0.87		7.52	10.43	不一致	82.89	77.78	不一致
14	Good UI	4.96	0.93	不一致	4.95	3.71		85.94	84.36	不一致
15	Good UI	0.12	0.41		12.96	10.39		86.50	84.89	不一致
16	Good UI	1.07	3.28		7.55	9.48	不一致	88.06	87.51	不一致

表 3.3: ANB 及び LLE と申告に基づく満足性との関係

3.4 考察

本実験の結果を以下に示す.

- (1) ANB は、被験者の満足性が高かった UI を使用した時と低かった UI を使用した時との間で中程度 [51] に有意な差が見られ、ほとんどの被験者で満足性が高かった UI の使用時のほうが低くなった。
- (2) LLE は、被験者の満足性が高かった UI を使用した時と低かった UI を使用した時との間で有意な差が見られなかった。
- (3) HR は、被験者の満足性が高かった UI を使用した時と低かった UI を使用した時との間で有意な差が見られなかった。

(1)について考察する。

ANB はシステム使用時の満足性と関係があり UX を評価するための指標として有用である可能性が示された。ANB が高ければストレスが高いと考えられ UX が良くないことを示し、逆に低ければストレスも低く UX が良いと言える。また、一方で、今回の実験では被験者の申告に頼っていることや、明らかにストレスがかかりやすい UI をマテリアルとして採用していること、同じ操作を何度も繰り返させることでその変化が数値に表れやすくしていることなどの制約があった。実用化に向けてより多くの種類の UI の問題をこの指標で捉えられるかどうかを明らかにする必要がある。

(2)(3)について考察する。

今回の実験では LLE, HR と満足性との間に有意な関係性は見られなかった。さらに、LLE と HR それぞれ約半数の標本で、被験者が申告した満足性の高い UI と LLE, HR の値が一致していなかったことから、今回の計測方法では LLE と HR が活用できる可能性が低い。HR を使用したユーザビリティ評価の先行研究 [52] では、計測したい一時的 UX のイベント直後に ECG によって計測した RRI から HR を算出しているため、今回よりも短い時間で計測していた可能性がある。そのため、RRI レベルの瞬間的な HR を計測すれば満足性を計測できる可能性がある。LLE についても同様に計測時間を短くしたときの満足性との関係性について検討する必要がある。

また、この実験の結果から、これまで使用してきた HR は一定の条件下で上手く機能せず、また HR が機能しない場合でも ANB が活用できる可能性が示唆された。今回の実験では、1 分の間で継続的に操作を行ったにも関わらず、HR の平均値には満足性との関係が見られなかった。このことから、仮に操作後に瞬間的な HR の上昇があったとしても持続しないかノイズが多くなっている可能性がある。操作のタイミングと HR の変化の関係性や変化の持続時間などについて検討する必要がある。一方、今回の実験のように待機時間のストレスを測るなど長時間の代表値を得たい場合には HR ではなく ANB が有効になる可能性がある。

第4章 時系列ユーザビリティ評価システムの提案

4.1 評価システムの設計・開発

4.1.1 時系列ストレスグラフ

4.1.2 観察ビデオとのマッピング

4.2 評価システムの有効性についての実験

4.2.1 対象と手続き

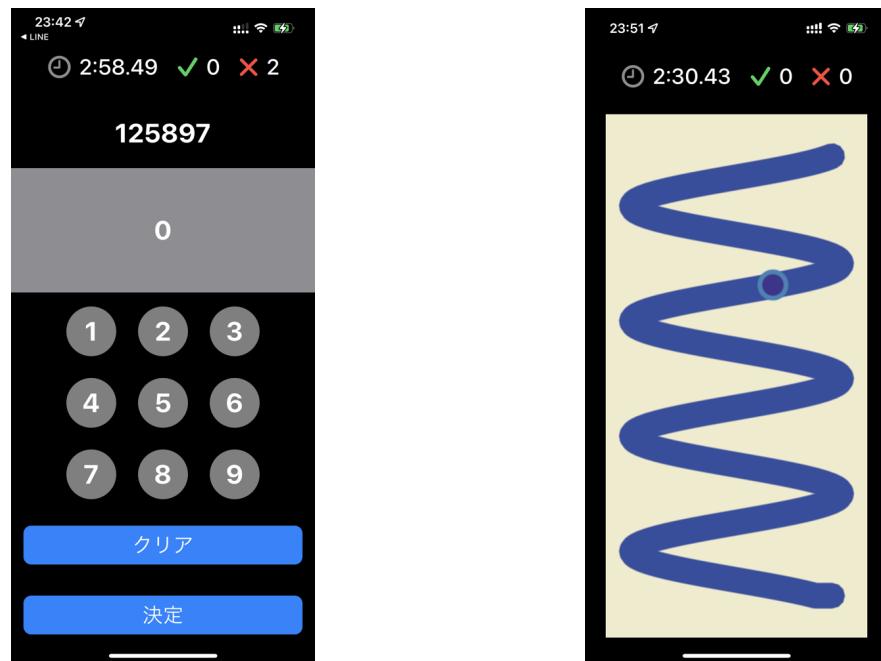


図 4.1: 実験マテリアル: テンキー

図 4.2: 実験マテリアル: ドラッグ

4.2.2 結果と考察



図 4.3: 実験の様子

第5章 結論

5.1 満足性評価のための指標の提案

5.2 評価システムの提案

5.3 今後の課題

謝辞

本研究の遂行にあたり、指導教員である増井俊之教授には1年次から温かいご指導をいただきました。本研究のみならず、様々な開発のアイデアに対して学術的あるいは製品化の観点から本質的なご意見をいただき議論させていただきましたことに心から感謝申し上げます。増井研究会では、博士課程の田中優氏、修士課程の左治木隆成氏、同輩の尾崎正和氏、後輩の佐々木雅斗氏には様々なご意見をいただき充実した研究会活動をさせていただきました。卒業生、研究会メンバー諸氏にも多くのご支援をいただきました。

本研究の中心となった、Lyapunov 指数を用いた指尖容積脈波の解析は、関西学院大学名誉教授であった故雄山真弓博士にご教授いただいたものです。高校時代に「活用の方法を提案してほしい」と言われた内容が学部でのメインテーマになったことは感慨深いものです。指尖容積脈波の活用について知見とシステムを提供していただいた関西学院高等部の丹羽時彦先生、セレブラルダイナミックス株式会社の海津成男氏に深く御礼申し上げます。

共同研究者である徳島大学デザイン型 AI 教育研究センターの福井昌則准教授、関西学院大学理工学研究科の萩倉丈氏には日頃から研究含め様々な相談に乗っていただきました。お二人とこれまでに様々な研究を行い、発表してきた経験が生きていると思います。本研究では、関西学院大学経済学部の前田慶士郎氏、関西学院大学文学部総合心理科学科の上田花菜氏、東北大学工学部の村上聰氏に実験計画や実施において多大なご支援をいただき、公私ともにお世話になりました。また、Bridge UI 株式会社には研究のバックアップをいただきました。

筆者の研究の原点といえば、関西学院高等部の宮寺良平先生、同中学部の河野隆一先生のご指導だと思います。先生方には在学中から今に至るまで大きな影響を受けました。

本研究は、孫正義育英財団の支援を受けて実施されたものです。ご支援のおかげで自由な研究をさせていただきました。また、財団生諸氏からも良い刺激をいただきました。

公私にわたりご指導、ご支援をいただきました皆様に心から御礼申し上げます。

2022年1月 吉日

佐々木 雄司



参考文献

- [1] 黒須正明. UX 原論：ユーザビリティから UX へ. 近代科学社, 2020.
- [2] Virpi Roto, Effie Lai-Chong Law, Arnold Vermeeren, and Jettie Hoonhout. 10373 Abstracts Collection – Demarcating User eXperience. In Jettie Hoonhout, Effie Lai-Chong Law, Virpi Roto, and Arnold Vermeeren, editors, *Demarcating User eXperience*, No. 10373 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2011. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany.
- [3] Nielsen Norman Group. 10 usability heuristics for user interface design. <https://www.nngroup.com/articles/ten-usability-heuristics/>.
- [4] 総務省. 令和 3 年版情報通信白書. <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r03/html/XXXXXX.html>.
- [5] T.S. Tullis, W. Albert, ソシオメディア. ユーザーエクスペリエンスの測定: UX メトリクスの理論と実践. 情報デザインシリーズ. 東京電機大学出版局, 2014.
- [6] Balmuda. バルミューダについて. <https://www.balmuda.com/jp/about/>.
- [7] 総務省. 令和 2 年通信利用動向調査. https://www.soumu.go.jp/johotsusintokei/statistics/data/210618_1.pdf.
- [8] 総務省. 平成 20 年通信利用動向調査. https://www.soumu.go.jp/johotsusintokei/statistics/data/090407_1.pdf.
- [9] Brian Shackel and Simon J Richardson. *Human factors for informatics usability*. Cambridge university press, 1991.
- [10] ニールセンヤコブ. ユーザビリティエンジニアリング原論: ユーザーのためのインターフェースデザイン. 情報デザインシリーズ. 東京電機大学出版局, 2002.
- [11] Jakob Nielsen and Rolf Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, p. 249–256, New York, NY, USA, 1990. Association for Computing Machinery.
- [12] 人間中心設計の基礎. HCD ライブライター. 近代科学社, 2013.
- [13] Patrick W Jordan. *Designing pleasurable products: An introduction to the new human factors*. CRC press, 2000.

- [14] 正明黒須. Ux の概念—プラクティス論文としての解説—. デジタルプラクティス, Vol. 6, No. 4, pp. 249–254, oct 2015.
- [15] Tuan D. Pham, Truong Cong Thang, Mayumi Oyama-Higa, Hoc X. Nguyen, Saji Hameed, and Masahide Sugiyama. Chaos and nonlinear time-series analysis of finger pulse waves for depression detection. In Sergio Alvarez, Jordi Solé-Casals, Ana L. N. Fred, and Hugo Gamboa, editors, *BIOSIGNALS 2013 - Proceedings of the International Conference on Bio-inspired Systems and Signal Processing, Barcelona, Spain, 11-14 February, 2013*, pp. 298–301. SciTePress, 2013.
- [16] Mayumi Oyama-Higa, Tokihiko Niwa, Wenbiao Wang, and Yoshifumi Kawanabe. Identifying characteristic physiological patterns of parkinson's disease sufferers using sample entropy of pulse waves. In Reyer Zwiggelaar, Hugo Gamboa, Ana L. N. Fred, and Sergi Bermúdez i Badia, editors, *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018) - Volume 5: HEALTHINF, Funchal, Madeira, Portugal, January 19-21, 2018*, pp. 189–196. SciTePress, 2018.
- [17] Shun Arai, Kazuhisa Ohira, Niwat Thepvilajanapong, Nobuji Tetsutani, Yoshito Tobe, Mayumi Oyama-Higa, and Yoshikatsu Ohta. A design of software adaptive to estimated user's mental state using pulse wave analysis. In *Ninth International Conference on Networked Sensing, INSS 2012, Antwerp, Belgium, June 11-14, 2012*, pp. 1–4. IEEE, 2012.
- [18] Vagner do Amaral, Leonardo A. Ferreira, Plinio T. Aquino, and Maria Claudia F. de Castro. Eeg signal classification in usability experiments. In *2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*, pp. 1–5, 2013.
- [19] Mayumi Oyama-Higa, Tiejun Miao, Shigeo Kaizu, and Junji Kojima. Mental health self-check system using “lyspect”. In *Proceedings of Sixth International Symposium on e-Health Services and Technologies and the Third International Conference on Green IT Solutions*, pp. 9–18, 2012.
- [20] usability.
- [21] J. Nielsen. *Usability Engineering*. Interactive Technologies. Elsevier Science, 1994.
- [22] Don Norman, Jim Miller, and Austin Henderson. What you see, some of what's in the future, and how we go about doing it: Hi at apple computer. In *Conference companion on Human factors in computing systems*, p. 155, 1995.
- [23] Sheena Lyonnais. Where did the term “user experience” come from? <https://blog.adobe.com/en/publish/2017/08/28/where-did-the-term-user-experience-come-from#gs.ls8wnt>.

- [24] Masaaki Kurosu. Descriptive words for expressing the user experience. In Masaaki Kurosu, editor, *Human Centered Design*, pp. 81–90, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
- [25] Richard L. Hazlett and Joey Benedek. Measuring emotional valence to understand the user’s experience of software. *International Journal of Human-Computer Studies*, Vol. 65, No. 4, pp. 306–314, 2007. Evaluating affective interactions.
- [26] Jacopo Staiano, Maria Menendez, Alberto Battocchi, Antonella De Angeli, and Nicu Sebe. Ux mate from facial expressions to ux evaluation. In *Proceedings of the Designing Interactive Systems Conference*, p. 741–750, New York, NY, USA, 2012. Association for Computing Machinery.
- [27] Tobii. tobiipro. <https://www.tobiipro.com/ja/product-listing/>.
- [28] Tao Lin, Masaki Omata, Wanhua Hu, and Atsumi Imamiya. Do physiological data relate to traditional usability indexes? In *Proceedings of the 17th Australia conference on computer-human interaction: Citizens online: Considerations for today and the future*, pp. 1–10. Citeseer, 2005.
- [29] Rosalind W Picard and Jocelyn Scheirer. The galvactivator: A glove that senses and communicates skin conductivity. In *Proceedings 9th Int. Conf. on HCI*, 2001.
- [30] Dennis W. Rowe, John Sibert, and Don Irwin. Heart rate variability: Indicator of user state as an aid to human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’98, p. 480–487, USA, 1998. ACM Press/Addison-Wesley Publishing Co.
- [31] Foroohar Foroozan and Jiang Wu.
- [32] Vagner do Amaral, Leonardo A. Ferreira, Plinio T. Aquino, and Maria Claudia F. de Castro. Eeg signal classification in usability experiments. In *2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*, pp. 1–5, 2013.
- [33] Kirk Shelley and S. Shelley. *Pulse Oximeter Waveform: Photoelectric Plethysmography*, pp. 420–423. 01 2001.
- [34] Erik Peper, Richard Harvey, I-Mei Lin, Hana Tylova, and Donald Moss. Is there more to blood volume pulse than heart rate variability , respiratory sinus arrhythmia , and cardiorespiratory synchrony ? 2007.
- [35] 中川千鶴. 特集3 人間工学のための計測手法. 人間工学, Vol. 52, No. 1, pp. 6–12, 2016.
- [36] 山口勝機ほか. 心拍変動による精神負荷ストレスの分析. 研究紀要. 志學館大学, Vol. 31, No. 1, pp. 1–10, 2010.

- [37] Paul M Lehrer and Richard Gevirtz. Heart rate variability biofeedback: how and why does it work? *Frontiers in psychology*, Vol. 5, p. 756, 2014.
- [38] TRYTECH. チェックマイハート. <https://www.trytech.co.jp/checkmyheart/glossary.html>.
- [39] Zhigao Lei, Shuhang Han, and Qijie Zhao. User experience evaluation based on ppg for human-vehicle haptic interaction. In *2020 Chinese Automation Congress (CAC)*, pp. 1810–1813, 2020.
- [40] Thomas Wollmann, Farhad Abtahi, Aboozar Eghdam, Fernando Seoane, Kaj Lindecrantz, Martin Haag, and Sabine Koch. User-centred design and usability evaluation of a heart rate variability biofeedback game. *IEEE Access*, Vol. 4, pp. 5531–5539, 2016.
- [41] ベルジェP, ポモワイヴェ. カオスの中の秩序 亂流の理解へ向けて. 産業図書, 1992.
- [42] Yuyu Hu, Zhe Li, Mayumi Oyama-Higa, and Emako Miyoshi. Exploring the use of pulse waves as psychological indicators in learning. In *2017 International Conference of Educational Innovation through Technology (EITT)*, pp. 307–310, 2017.
- [43] Mayumi Oyama-Higa, Teijun Miao, and Yuko Mizuno-Matsumoto. Analysis of dementia in aged subjects through chaos analysis of fingertip pulse waves. In *2006 IEEE International Conference on Systems, Man and Cybernetics*, Vol. 4, pp. 2863–2867, 2006.
- [44] Akira Imanishi and Mayumi Oyama-Higa. On the largest lyapunov exponents of finger plethysmogram and heart rate under anxiety, fear, and relief states. In *2007 IEEE International Conference on Systems, Man and Cybernetics*, pp. 3119–3123, 2007.
- [45] 心の免疫力を高める「ゆらぎ」の心理学. 祥伝社新書. 祥伝社, 2012.
- [46] Mayumi Oyama-Higa, Teijun Miao, and Yuko Mizuno-Matsumoto. Analysis of dementia in aged subjects through chaos analysis of fingertip pulse waves. In *2006 IEEE International Conference on Systems, Man and Cybernetics*, Vol. 4, pp. 2863–2867, 2006.
- [47] Mayumi Oyama-Higa, Tokihiko Niwa, Fumitake Ou, and Yoshifumi Kawanabe. Application of sample entropy of pulse waves in identifying characteristic physiological patterns of parkinson ’ s disease sufferers. In *International Joint Conference on Biomedical Engineering Systems and Technologies*, pp. 447–463. Springer, 2018.
- [48] Akira Imanishi and Mayumi Oyama-Higa. The relation between observers’ psychophysiological conditions and human errors during monitoring task. In *2006 IEEE*

International Conference on Systems, Man and Cybernetics, Vol. 3, pp. 2035–2039, 2006.

- [49] 康平小孫. ビデオゲームの特性と教育利用. 皇學館大学教育学部研究報告集, Vol. 3, pp. 43–55, 03 2011.
- [50] セレbralダイナミックス. カオテック lyspect 紹介. <http://www.chaotech.org/jigyo/jigyo.html>.
- [51] Jacob Cohen. *Statistical power analysis for the behavioral sciences*. Academic press, 2013.
- [52] Michael Trimmel, Monika Meixner-Pendleton, and Sandra Haring. Stress response caused by system response time when searching for information on the internet. *Human Factors*, Vol. 45, No. 4, pp. 615–622, 2003.

付録A 付録