

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

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ユーザの利用時の品質を示す UX を評価し、製品の改善に繋げるサイクルが重要である。しかし、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指摘されている。ユーザテストの導入ハードルを下げ普及させるために、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必要がある。そこで本研究では、(1) 容易なユーザテストを可能にするための新たな UX 指標として自律神経バランス (ANB) と最大リヤノフ指数 (LLE) を提案し、(2) それらの指標を用いた UX 評価システムを開発する。

(1) テストユーザの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、生理的反応からは機械的に解釈可能なデータを得られるため容易なユーザテスト手法になり得る。そこで、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た新たな UX 指標を提案し有用性を明らかにした。

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、習熟後といった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。そこで、(2) UX 評価システムでは、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より詳細な UX 変化を記録し可視化することを可能にする。

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

In order to design a system that is easy to use, it is important to conduct user tests on actual users to evaluate UX, which indicates the quality of the system when used by the user, and to use this cycle to improve the product. However, it is pointed out that few software development companies conduct UX evaluations, and even if they do, there are few cases where the evaluation results lead to product improvements. In order to reduce the hurdle of user testing, it is necessary to make it easy for users to conduct user testing and find problems in the system without specialized knowledge. In this study, we (1) propose Autonomic Nervous Balance (ANB) and Largest Lyapunov Exponent (LLE) as new UX indicators to enable easy user testing, and (2) develop a UX evaluation system using these indicators.

(1) Although evaluation indices based on the physiological responses of test users are not widely used because there is no established method, physiological responses can provide mechanically interpretable data, which can be an easy user testing method. However, physiological responses can provide mechanically interpretable data that can be easily used as a user testing method. We proposed a new UX index based on the technique of measuring stress from chaotic fluctuations of volumetric pulse waves, and clarified its usefulness.

In the concept of UX, it is important to consider time variation. Although some methods have been proposed to describe changes in UX over time, such as Kurosu's UX graphs, they have been used to roughly indicate the time period from a few weeks to a few months, such as before use, after use, and after mastery. Therefore, the (2) UX evaluation system uses physiological responses, which can be obtained sequentially, as an indicator to record and visualize more detailed UX changes over a period of several minutes to several hours during system use.

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	6
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)	15
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 UX 評価システムの提案	25
4.1 設計	25
4.1.1 使用時 UX グラフの表示	25
4.1.2 使用時 UX グラフとイベントのマッピング	26
4.2 実装	26
4.2.1 脈波解析エンジン	27

4.2.2 フロントエンド	27
4.3 妥当性検討実験の計画	28
4.3.1 実験マテリアルの開発	30
4.3.2 タスクの実施	30
4.3.3 容積脈波 (BVP) の計測と動画の収録	32
4.3.4 アンケートの実施	32
4.4 実験の結果	33
4.5 実験の分析	33
4.6 実験の考察	33
第 5 章 結論	38
5.1 満足性評価のための指標	38
5.2 時系列 UX 評価システム	38
5.3 今後の課題	38
5.3.1 統合的なシステムとしての展望	39
謝辞	41
参考文献	42
付 錄 A 略語表	47
付 錄 B 時系列 UX 評価システムの妥当性調査実験 質問紙	48
B.1 事前アンケート	48
B.2 事後アンケート	48
付 錄 C 時系列 UX 評価システムの妥当性調査実験 自由回答	51
C.1 実験 A	51
C.2 実験 B	53
付 錄 D 時系列 UX 評価システムの妥当性調査実験 選択回答	57
D.1 デジタル機器対応状況	57

図 目 次

1.1 黒須の品質特性図 [1]	3
1.2 UX グラフの例 [2]	5
2.1 SQuaRE を基にした JIS X 25010:2013 の品質モデル	8
2.2 UX のライフサイクル [3]	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 で取得した容積脈波	14
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 での計測中画面	20
3.4 Lyspect での分析画面	21
4.1 脈波の計測画面	27
4.2 容積脈波計の接続画面	28
4.3 RRI, LLE, ANB の推移の表示画面	29
4.4 動画及び BVP の同期画面	29
4.5 イベントタイミングの入力画面	30
4.6 実験 A(テンキー)の画面	31
4.7 実験 B(ドラッグ)の画面	31
4.8 実験の様子	32
4.9 被験者 1 実験 A Bad UI の LLE	35
4.10 被験者 1 実験 A Good UI の LLE	35
4.11 被験者 1 実験 B Bad UI の LLE	35
4.12 被験者 1 実験 B Good UI の LLE	35
5.1 測定画面のプロトタイプ (アプリ用)	40
5.2 測定画面のプロトタイプ (VR 用)	40
5.3 分析画面のプロトタイプ	40
5.4 画面遷移図での可視化画面のプロトタイプ	40

表 目 次

3.1	10 Usability Heuristics for User Interface Design[4]	18
3.2	満足性と ANB, LLE, HR との関係性についての実験結果	22
3.3	ANB, LLE, HR と申告に基づく満足性との関係	23
4.1	時系列 UX 評価検証実験: 被験者の概要	34
4.2	時系列 UX 評価システム検証実験: マテリアルに対する肯定否定の別	35
4.3	時系列 UX 評価システム検証実験: マテリアルに対する肯定的意見の内訳	36
4.4	時系列 UX 評価システム検証実験: マテリアルに対する否定的意見の内訳	36
A.1	略語表	47
D.1	時系列 UX 評価検証実験: 被験者のデジタル機器対応状況	59

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 などのユーティリティアプリケーションも成長が見込まれている [5]. 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 UI 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 [6]. 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 [7].

また、個人のアプリ利用者の年齢層は年々広がっており、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% と特に 60 代以上で利用者数の伸びが顕著である [8][9]. 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、操作の習熟度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[10] 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、ユーザビリティの概念は体系化され、ISO/IEC 25000(別名 SQuaRE) の規格の中

で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に必要な工程として以下が挙げられている [11].

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(6)ヒューリスティック評価、その後に(8)インタフェース評価と(10)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、実装したUI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において、ユーザに直接使用させて評価することは欠かすことの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 [12] 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、この手法では発見できるユーザビリティ上の問題に制限があるだけでなく、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 [13]。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 [6]。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を用いて測定するものである。具体的にはビデオの録画や筋電位センサ、アイトラッキ

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

ング、皮膚伝導率、心拍数などが指標として使われている [6]. 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。

近年話題となっている利用時品質を示す UX(User eXperience) では、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”の重要性を指摘し、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 [14]. 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 [15]. 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etri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う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、ユーザの主観を測定する方法はストレス測定や視線測定など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。

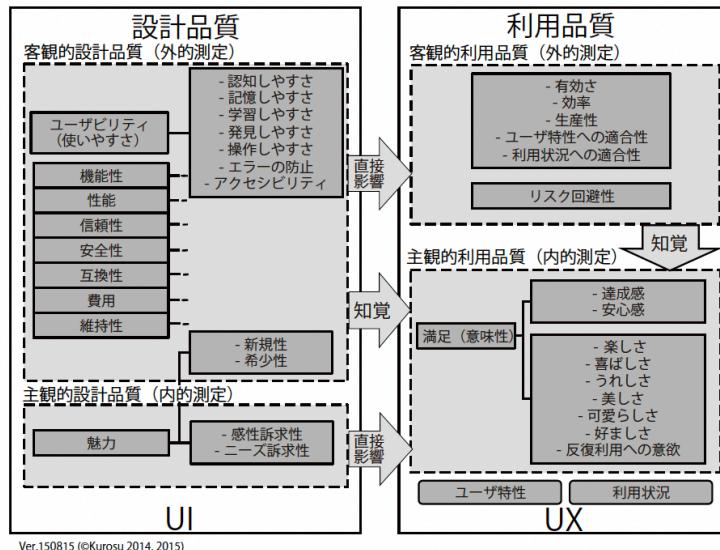


図 1.1: 黒須の品質特性図 [1]

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、その評価手法は様々なものが提案されているが、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、行われても改善に役立てられていないのが現状である。黒須は、

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 [1].

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[16]では、UX向上に取り組んでいると答えた割合が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度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、時間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[15](図1.2)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、縦に満足度、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度を測定する簡便な手法を開発すること、(3)(2)で測定した満足度から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

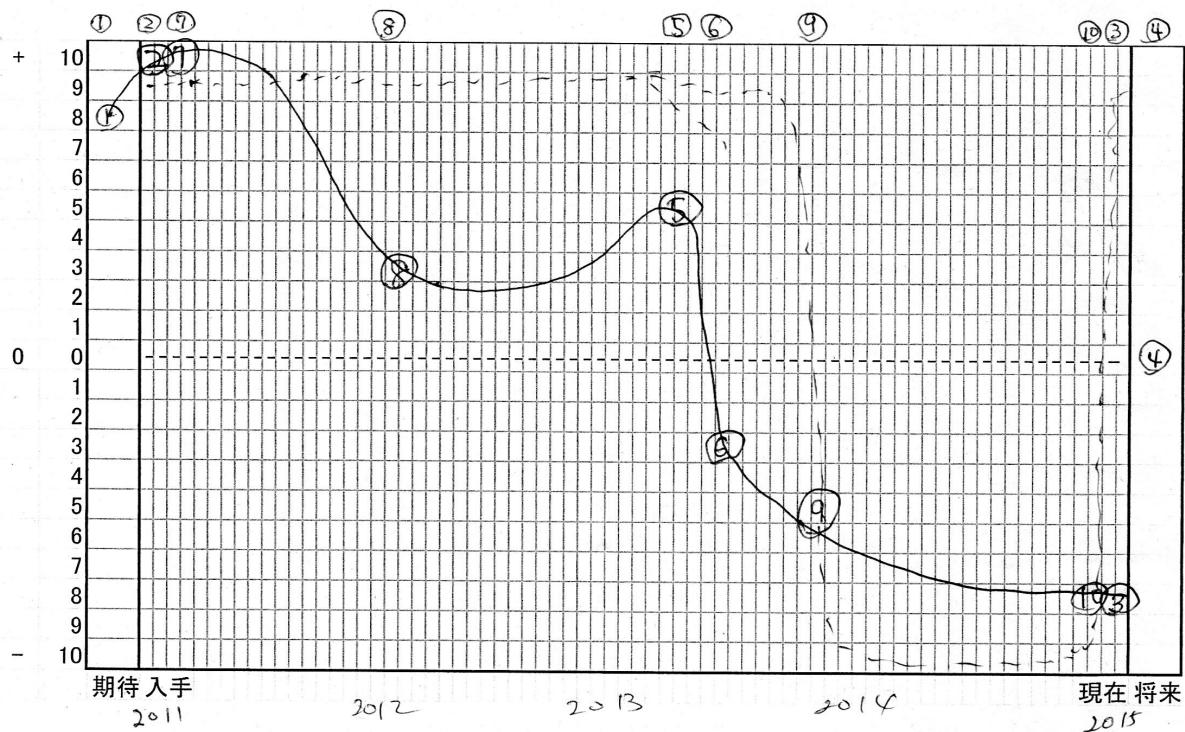
本研究では(1)現在は大規模な開発でしか行われていないユーザテストの導入ハードルを下げて普及させるために、UX評価や統計処理についての専門的な知識が無くても容易にユーザテストを行いシステムの問題点を発見できるようにするシステムの開発を目指す。行動・生理メトリクスは確立された手法が無いため普及していないが、シーケンシャルであることから分析の自動化と相性がよく容易なユーザテスト手法になり得る。ユーザテストでは、被験者は笑ったりそわそわしたり調査票に記入するよりもはるかに多くの行動をするがそれらを記録することでより情報量の大きいデータ入手可能になる[6]。そこで、(2)及び(3)の手法を踏まえた統合的な分析システムを開発する。

(2)満足度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[6]、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[17][18]。さらに、同様の解析により精神状態を測定できる可能性[19]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[20]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

UX グラフ記入用紙

対象 スマホ 記入年月日 年 4月 25日 性別 男 年齢 55

① 期待	常時 ネット接続は 期待が 高い	満足度レベル 8
② 入手 2011年	テザリングが 可能になったので 購入決定	満足度レベル 10
③ 現在 2015年	2台目 2年に 近く りせ、トク多く 買い換えて 検討	満足度レベル -8
④ 将来	やはり パソコンの方が良い ネットワーク入力も含め	満足度レベル 0
⑤ 2013年	2台目に変更 WiMax用の機種がなくなるといふ音で	満足度レベル 5
⑥ 2013年	" 今はスピードが遅く 感度も低く 発熱大きい	満足度レベル -3
⑦ 2013年	1台目は楽しく 画面も 美しい	満足度レベル 10
⑧ 2012年	テザリングが 2台までだと PC セグメント等々に 不足	満足度レベル 3
⑨ 2013年	④ 死亡により WiMaxルーターを入手 スマホテザリング不足	満足度レベル -5
⑩ 2015年	WiMaxルーター契約切れ スマホと利用 ひどい(要改善)	満足度レベル -8
⑪ 年		満足度レベル
⑫ 年		満足度レベル
⑬ 年		満足度レベル
⑭ 年		満足度レベル
⑮ 年		満足度レベル



グラフ 満足度
利用の程度

図 1.2: UX グラフの例 [2]

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で容積脈波を取得しストレスを測定、満足性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との関係を示した。

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[21] 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つた。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操作記録映像と同期させた時系列グラフで表示するシステムを開発した。

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4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。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 [22][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であった [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 [10][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 [23]。

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 Bevan 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至った [1]。この定義は現在でも一般の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、ユーザビリティが品質モデルの一部に組み込まれたている。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品質である利用時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 [13] の考え方の普及などにより、ユーザの利用体験に特に

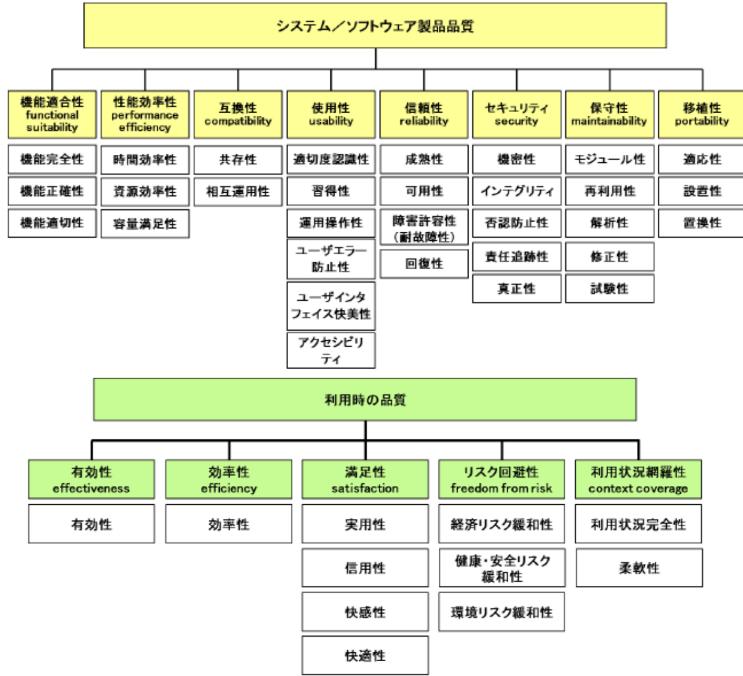


図 2.1: SQuaRE を基にした JIS X 25010:2013 の品質モデル

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、多様なユーザに目を向けているため、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はできない。そこで、SQuaRE の利用時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イシングループで使われており [24]、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 [25]。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 Roto 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 [3]。

UX 白書による UX の定義には「システムと出会うこと」という文言が含まれている。つまり、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 [3]。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、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捉えることができる。つまり、主観的利用品質を客観的に評価することが重要になる。

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 [26]。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なである。

2.2.2 時間相

UX 白書の UX の定義では SQuaRE と異なり時間相が明確になっている。期間の長さに焦点を当てて分類すると一時的 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 [3].

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体験・感情

エピソード的 UX 利用エピソードに関する体験・感情

累積的 UX システムをしばらく使用した後、利用中、利用休止中を通して形成される体験・感情

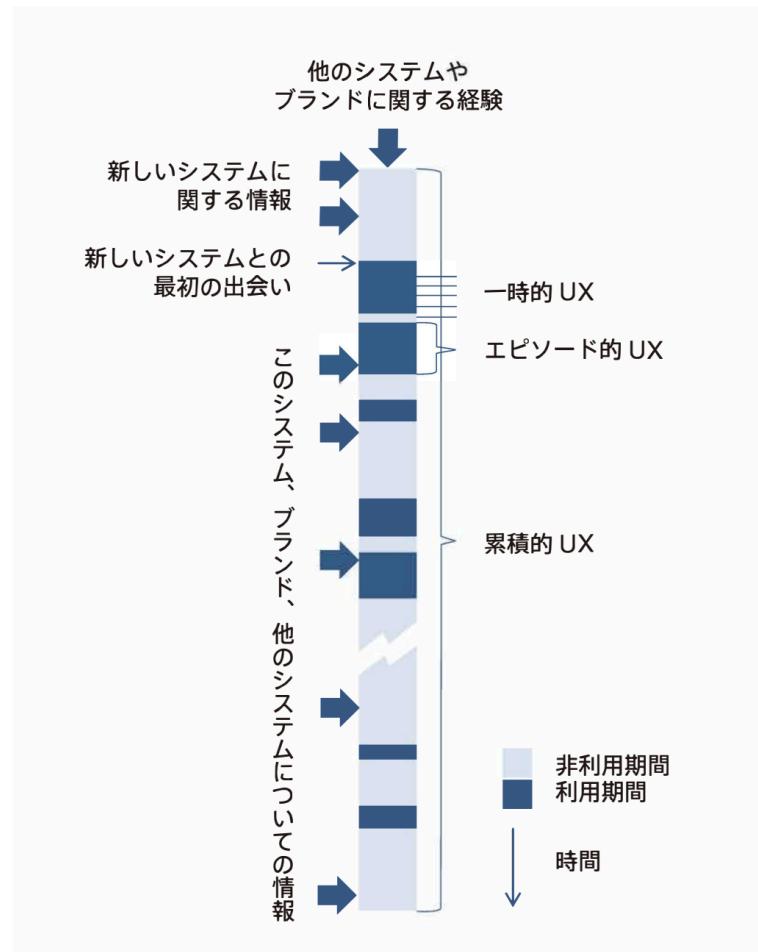


図 2.2: UX のライフサイクル [3]

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 [15] 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。ユーザ自身に記入させることでユーザにとっての UX 変化を累積的 UX のレベルで可視化することができる。一方、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

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手法では正しく評価できない可能性を排除できない。つまり主観的評価と客観的評価の双方に着目した指標が必要である。

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 [6]、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性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、さまざまな手法が提案されている。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観察し、言語行動や非言語行動を抽出することは一般的である。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、目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れる情報の精度にばらつきが出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 [6] 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によって自動的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 [27]。現在であれば表情をビデオで収録し機械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 [28]。

アイトラッキングはtobii社がtobiipro[29]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、見た時間、気づくまでに要した時間、目の動きの距離、注視の時間などがある[6]。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、興味の高まりに相關することがわかっている。しかし、瞳孔反応はさまざま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[6]。

ストレスの測定は満足性に密接に関係すると考えられ、さまざま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。他のさまざま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ストレスを数値化するもので、さまざま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[30]。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[31]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。

心拍数変動(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: Heart 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[32]。光電式容積脈波記録(PPG: Photoplethysmography)を用いて計測する脈拍変動(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[33]。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Apple Watchなどの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(PPG)を用いた手法に期待できる。

脳波記録(EEG: Electroencephalography)は脳の電気信号を脳波計によって記録するもので、さまざま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[34]。

ストレス評価には、これらの他にもさまざま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 [6]。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: Photoplethysmography) 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 [35]。脳波 (EEG) 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。実際の装置を図 2.3 と図 2.4 に示す。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りにくい。



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 BVP を測定する際には PRV を使用することが多かった。しかし、BVP 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 PRV に限らないさまざまな検討されている。そもそも PRV 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 [36]。

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。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 [37]。RRI は BVP 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。

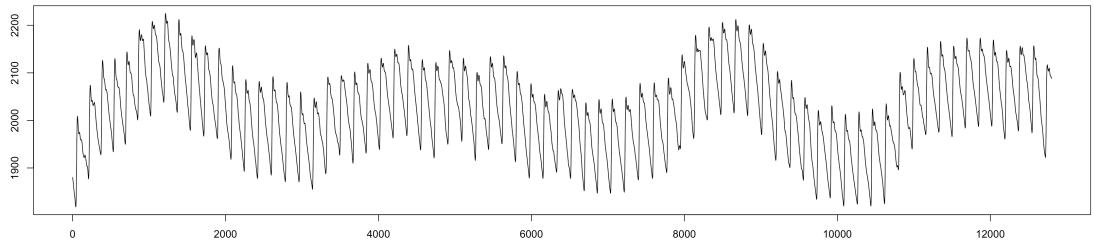


図 2.5: PPG で取得した容積脈波

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 [38]. 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 [39]. LF 成分は血圧変動であるマ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 [40].

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 [38]. 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 PRV は変化す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 [37]、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 や PRV 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い [41][42]。理由として 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想定される。よって、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) は、カオティック¹な情報である 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 [21]. 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の計測が難しい。本研究では一時的 UX[3] 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。

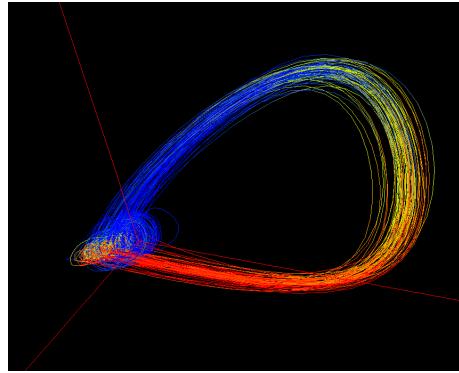


図 2.6: 図 2.5 の脈波から再構成したアトラクタ

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 [43]。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 [44]。

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

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

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

雄山らは、この LLE で PRV のゆらぎを解析する手法を提案している [45]。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 [46]。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まないかたくなな状態・外部適応が困難な状態と表現している [47]。

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さまざま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

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が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 [43]

²空間中の軌跡

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[44]。また、認知症の程度やパーキンソン病の有無とLLEの間に関連があることもわかっている[48][49]。ユーザビリティに関連した分野では、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[50]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[51]。

2.5 問題の所在

前述のように、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測定さ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[15]のような累積的UXではなく、一時的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。理由として、その結果の解釈が困難であることが想定される。

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)一時的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、時系列 UX 評価システムの開発に先立ち、容積脈波 (BVP) 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) と最大リヤプノフ指数 (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、ヒューリスティクスはユーザビリティつまり製品品質の評価であり、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 と避けるべきとされている UI をそれぞれ便宜上 Good UI・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)[12] 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	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バ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こと
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[4]

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.

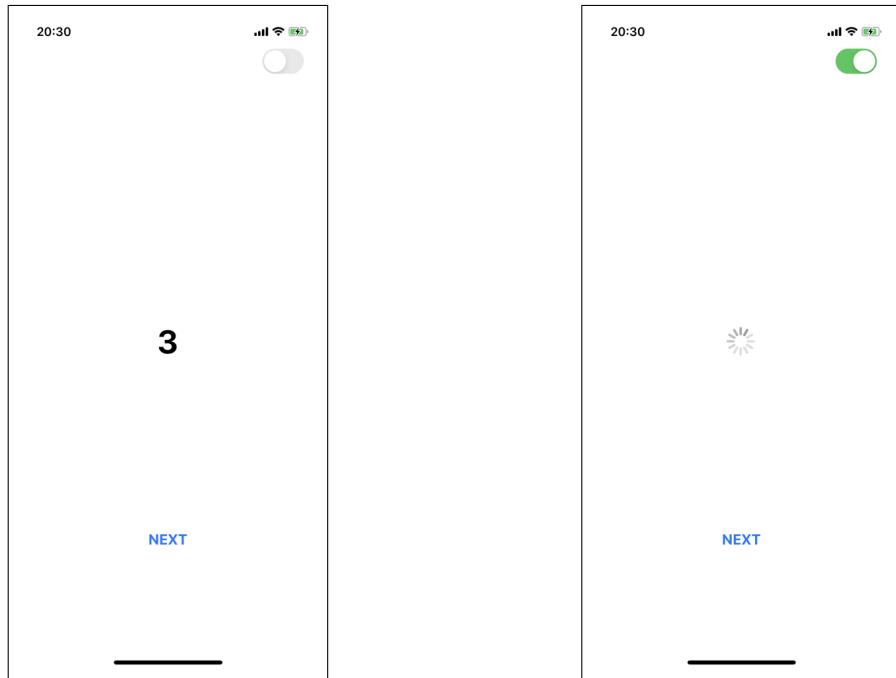


図 3.1: 実験マテリアル: 数字の表示画面

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とした. できるだけ多く書くように指示することで, 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表示された数字を紙に書いてください。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たくさん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[52] 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。

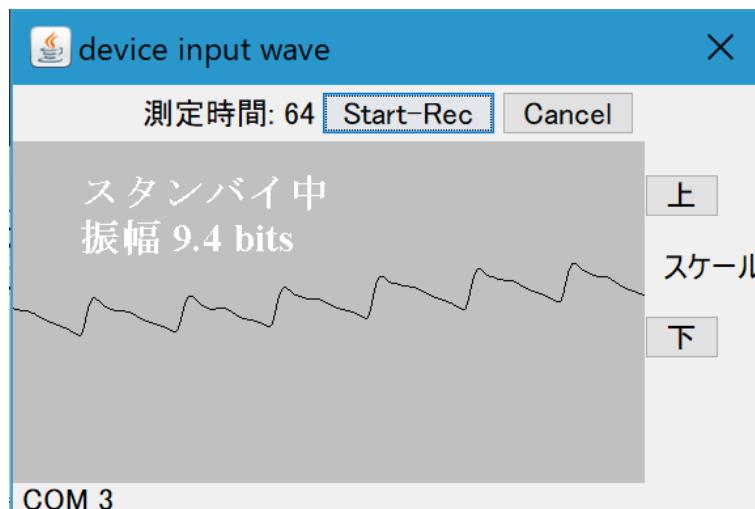


図 3.3: Lyspect での計測中画面

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たため、その場合には BVP をトリミングした後、Lyspect を用いて分析した (図 3.4)。

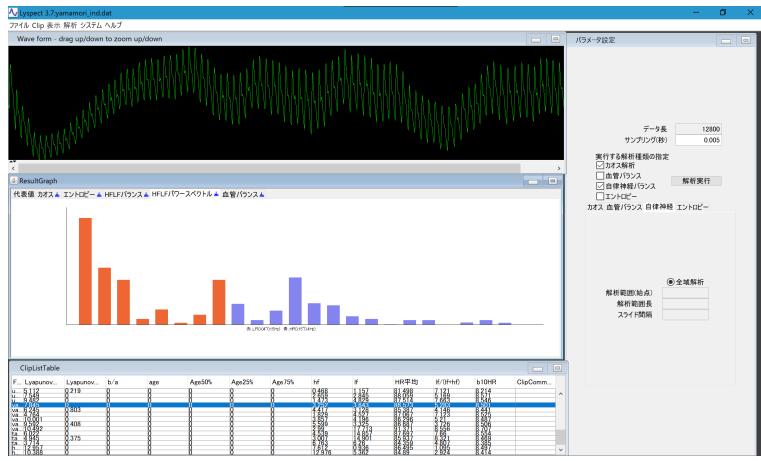


図 3.4: Lyspect での分析画面

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 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 ($M = 6.85, SD = 2.63$) 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, LLE, HR 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, LLE, HR 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 を使用したときと低かった UI を使用したときとの間で中程度 [53] 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つた。
- (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つた。
- (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つた。

(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つた。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つたことから、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 [54] 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つたにも関わらず、HR の平均値には満足性との関係が見られなかつた。このことから、仮に操作後に瞬間的な HR の上昇があつ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UX評価システムの提案

本章では、時系列UX評価システムの設計及び実装と、その妥当性を検討するために行つた実験について述べる。第3章で述べた満足度とANB, LLE, HRの関連性を調査した実験では、満足度とANBとの関連性が明らかになった。そこで、ANBの変化を時系列の満足度変化として捉え、UXを評価するシステムを開発した。また、LLEとHRについては前述の実験では満足度との関係性を明らかにできなかったものの、短いスパンでの代表値として利用することで有効に活用できる可能性がある。そのため、短時間のBVPから計算したLLEと瞬間のHRであるRRIについても計測した。

4.1 設計

時系列UX評価システムのプロトタイプとして主要な部分のみを実装した。本研究において時系列UX評価とは、システムの使用時、使用前、使用後、使用していない期間のUXを、何らかの指標やユーザの自己申告などに基づいてその変化を可視化することを指す。従来手法としてはUXグラフなどがあるが、ここでは特に使用時のUX変化をより短いタイムスパンで表示することを対象とする。

使用時の時系列UX評価において重要なのは、指標を用いて詳細なグラフを表示することと、そのグラフを実際のユーザの動作と対応づけることである。この2点を実現するためのシステムを開発した。

4.1.1 使用時UXグラフの表示

使用時UXグラフを表示するための指標としてANB, LLE, RRIを採用しこれらを同時に表示するシステムを開発した。従来のUXグラフではユーザの自己申告によって手描きで満足度の変化をグラフに書き起こしていた。しかし、使用時の秒単位での変化を使用時や使用後に書いてもらうことは難しい。そこで生体データから得られるストレスに関する指標を満足度を表す値として使用する。具体的には前述のように満足度との関連が見られたANBと、有意差は見られなかったものの短いスパンにすることで活用できる可能性があるLLE, RRIを指標として採用した。

ストレスを測定するためにPPGを使用しBVPを記録する。測定には、耳朶用の光学式容積脈波計であるVital Meter with 3D Accelerometer(製造:TAOS Institute, Inc., 図2.4)を使用し1000Hzでサンプリングした。なお、測定器側で適切なフィルタリングがかけられて

いることを確認し、記録した波形はフィルタリング済みのものである。また、分析に ANB, LLE, RRI の分析にあたっては 200Hz にダウンサンプリングしたデータを使用した。第 3 章で行った研究では指先用のセンサを使用したがここでは耳朶に装着するものを使用した。これにより、手による操作が阻害されることがなくなりより正確なデータが得られるようになった。また、Vital Meter は Bluetooth でコンピュータと接続するため被験者の姿勢や場所に制約が少なくなった。

ANB は 0.15-4Hz の HF と 0.004Hz-0.15Hz の LF 成分から算出される値である。つまり LF 成分の周期は 6.6-250 秒であることから、これを分析するためには最低でもそれを十分に捉えられるだけの長さの波から計算する必要があり、100 拍、約 5 分間以上の測定が標準とされている [38]。そこで、時系列の ANB 分析では、256 秒 (4.26 分) の波形から ANB を算出し、10 秒ずつスライドしながら算出を繰り返すことで、時系列の ANB を取得した。なお、256 秒に満たないデータからは簡易的に 64 秒の波から ANB を算出することとした。

LLE は ANB よりも短い時間の波から有効な値が得られることがわかっており、15 秒の波形から LLE を算出し、ANB と同じく 10 秒ずつスライドしながら算出を繰り返すことで時系列の LLE を取得した。

RRI は ANB の算出で使用する波のピーク位置を元に Peek to Peak の時間から瞬間心拍数を算出し 1 秒毎の RRI とした。

4.1.2 使用時 UX グラフとイベントのマッピング

UX グラフでは、グラフの変化が大きい部分でどのようなイベントが起きていたのかを知ることが重要である。長期間の UX グラフでは、「使用の開始」「他の製品を使ったとき」などのイベントを先に提示し、イベントに対して満足性を記録することでグラフ化していた。しかし、使用時 UX ではグラフを見て問題がある部分のイベントを調査するという流れになると考えられる。

UX グラフと関連付けるイベントの取得には、使用の様子を動画で収録する、スクリーンキャプチャを取る、アプリケーションにログの記録機能を搭載するといった方法が考えられる。今回はプロトタイプであるため、またアプリケーションの種類を限定しないために使用の様子を動画で収録することとスクリーンキャプチャを取ることを採用した。

本研究では、休憩と実験毎をイベントとして分離すればよく、BVP のを 2 つの動画と同期させたうえで波形データをそれぞれのイベント毎に分離するシステムを開発した。

4.2 実装

システムのフロントエンドは Web アプリケーションとして開発し、ANB, LLE, RRI を計算するエンジン部分はサーバーサイドアプリケーションとして開発した。

4.2.1 脈波解析エンジン

脈波の解析エンジンは、第3章の実験で使用した Lyspect の開発元であるセレbralダイナミックスから提供を受け、同一のものをサーバーサイドで実行できるようにした。システムは Lyspect に準じて Java で開発されており、Function as a Service¹である Google Cloud Platform²の Cloud Functions でホスティングした。これにより、数秒から数十秒かかる解析であってもスケールアウト³することで同時に多数の解析を実行することが可能になる。

エンジンには REST API を通じてアクセスできるため、ANB, LLE, RRI それぞれのエンドポイントに 200Hz でサンプリングした BVP を JSON 形式で送ると数秒から数十秒後に解析の結果が返される。

4.2.2 フロントエンド

システムのフロントエンドは Vue.js を用いて Web アプリケーションとして開発した。このシステムでは、BVP の計測、RRI, LLE, ANB による使用時 UX グラフの表示、イベントとのマッピングを行うことができる。なお、プロトタイプであるためブラウザ API のサポート状況の問題で Windows または macOS の Chrome でしか動作を確認していない。

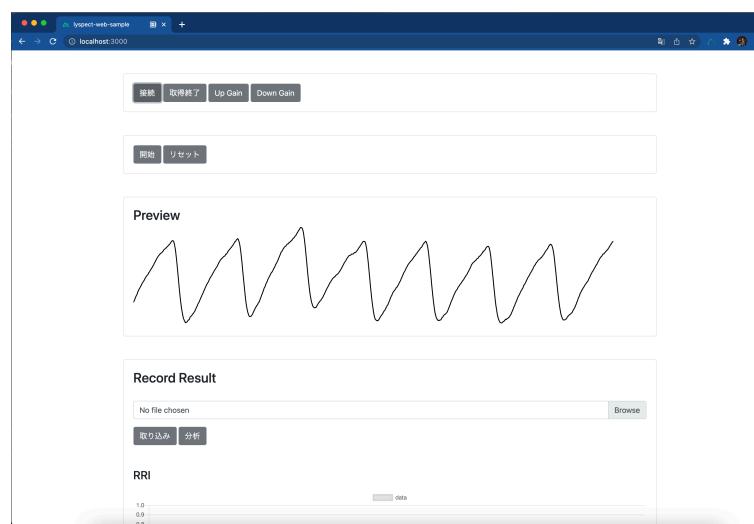


図 4.1: 脈波の計測画面

脈波の計測画面は図 4.1 のようになっている。センサを接続するには左上の接続ボタンをクリックし、ブラウザが表示するダイアログ（図 4.2）から接続したいデバイスを選択する。Vital Meter はシリアル通信を行っているため、コンピュータに専用のドライバが入っていないかったとしても汎用的なシリアルデバイスとして使用が可能である。このシステムでは、

¹HTTP リクエストなどのイベントに応じて関数が実行されるサーバレスアーキテクチャ

²Google が提供するクラウドサービスの総称

³コンピュータの台数を増やし処理を並列化・分散化させる手法。FaaS では実際にコンピュータの台数が増えるとは限らないがインスタンスという形で仮想敵に台数が増えているといえる。

Chrome ベースのブラウザで実装されている Web Serial API[55] を使用してアクセスしている。この API ではセキュリティ上の問題から Web アプリケーションが自由にシリアルデバイスにアクセスすることを許しておらず、ブラウザが表示するダイアログを経由することが必須となっている。

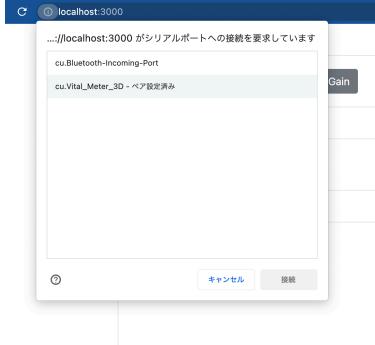


図 4.2: 容積脈波計の接続画面

センサが接続されると Preview のフィールドに波形が表示される。波形が安定したことを確認して開始ボタンをクリックすると記録した BVP を保存するファイルを選択するダイアログが表示される。このシステムでは、BVP の計測データが大量になることからメモリ上ではなくファイルに直接書き出す方式を採用している。通常の Web アプリケーションではファイルへの逐次書き出しが許可されていないため、いくつかのブラウザで実験的に実装されている File System API[56] を使用してこれを可能にした。

BVP の記録を終了した後、分析ボタンをクリックすると脈波解析エンジンにデータが送られ、結果が図 4.3 のように時系列で可視化される。RRI と ANB(HFLF) は高いほどストレスが高く、LLE(Lyapunov) は低いほどストレスが高いことを表す。

BVP の計測と同時に操作画面のスクリーンキャプチャや第三者視点での動画を収録し、それらの動画とストレスの値を関連付けることで、どの操作をしているときにストレスが高くなったのか、あるいはストレスが低かったのかを明らかにすることができる。動画と脈波のタイミングを合わせ、実験の開始、終了、休憩の開始、終了などのイベントと関連付ける機能の画面を図 4.4 図 4.5 に示す。このシステムでは、簡易的に各動画の収録開始時刻と脈波の記録開始時刻を入力することでこれらを同期できるようになっている。また、表示される動画のタイムバーを操作するももう一方の動画の再生時刻も同期された時刻に変化するようになっており、第三者視点の動画とスクリーンキャプチャの双方からイベントのタイミングを探すことができる。

4.3 妥当性検討実験の計画

実験マテリアルとして 2 種類のアプリケーションを開発し、それぞれ動きがスムーズでない Bad UI とスムーズに動作する Good UI を用意した。そして、これらの計 4 種類のアプ

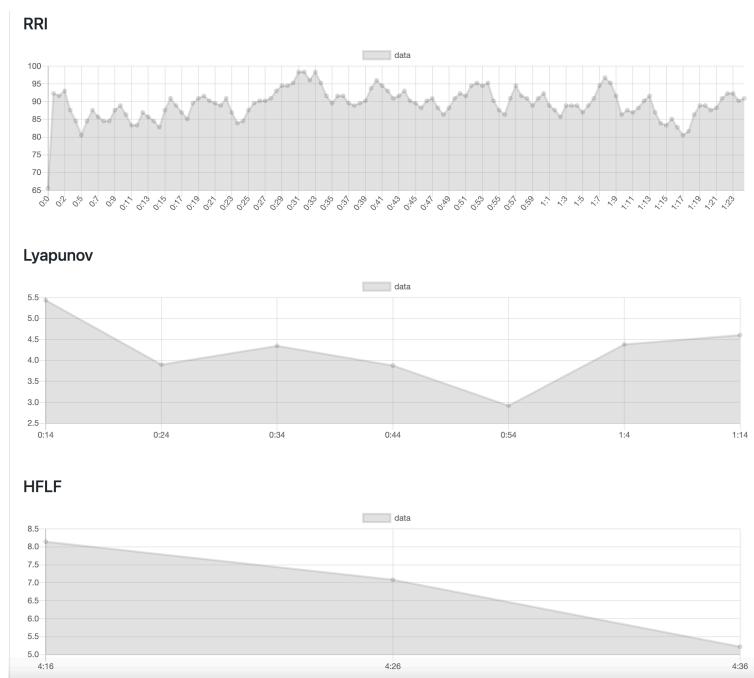


図 4.3: RRI, LLE, ANB の推移の表示画面

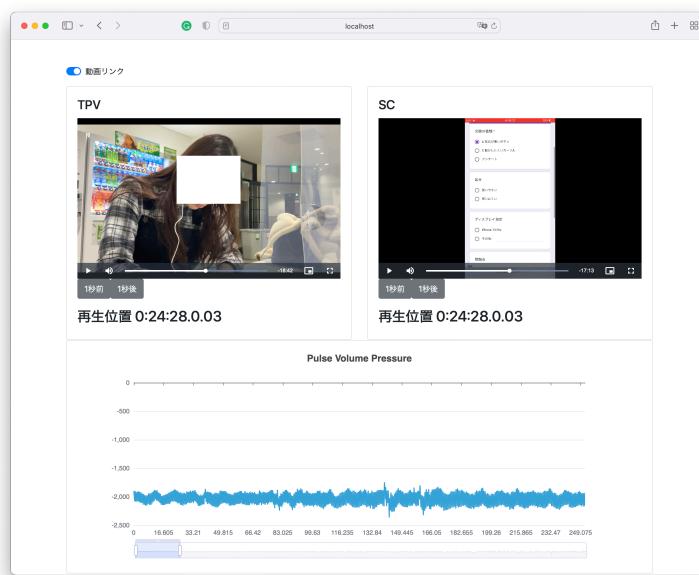


図 4.4: 動画及び BVP の同期画面

図 4.5: イベントタイミングの入力画面

リケーションを使っている間の UX 变化を前述のシステムを使用して分析しシステムの有用性を調査した。実験は、2021 年 11 月から 12 月に 48 名の大学生を対象に行い、被験者の平均年齢は 20.5 歳 ($SD = 1.66$)、生物学的性の内訳は男子 23 名、女子 25 名だった。

4.3.1 実験マテリアルの開発

実験マテリアルとしてテンキーで数字を入力するアプリ (実験 A, 図 4.6) とボールをドラッグして移動させるゲーム (実験 B4.7) を開発した。それについて動作がスムーズな Good UI と、スムーズでない Bad UI を用意した。また、いずれも「できるだけ多く正解したい」という意欲を持ってもらうためにカウントダウンのタイマーを表示したうえで正解数のカウンターを配置した。

4.3.2 タスクの実施

実験 A Good UI, 実験 A Bad UI, 実験 B Good UI, 実験 B Bad UI をカウンターバランスを取って順番に実施した。実験はそれぞれ 3 分間で行い、各実験の間と前後に 3 分間の休憩を設けた。また、実験の指示に差がないように以下のような教示文を印刷して提示した。

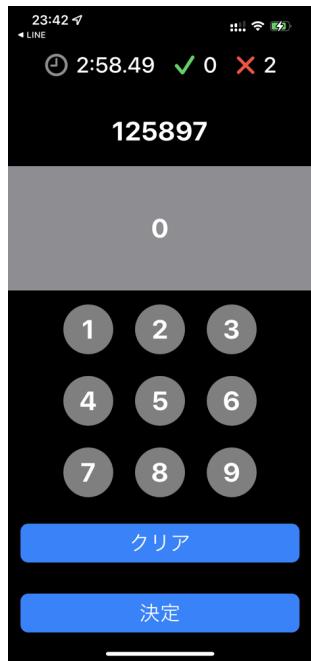


図 4.6: 実験 A(テンキー) の画面

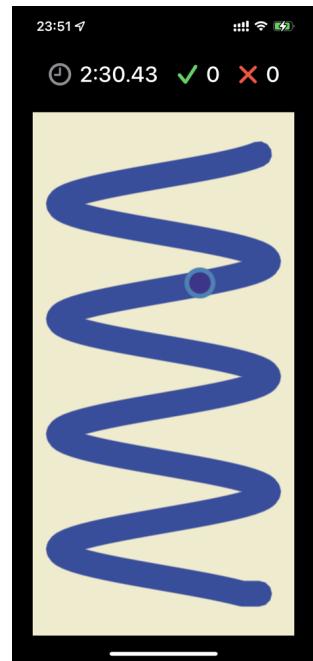


図 4.7: 実験 B(ドラッグ) の画面

実験 A の教示文

この実験は数学の基礎的な能力を測るもので、実験を開始すると右のような画面が表示されます。

上に数字、その下に入力画面、その下にテンキーが表示されます。上に表示されている数字を小さい順に入力画面に打ち込んで、打ち込み終わったら決定ボタンを押してください。

入力を間違えた場合は、クリアボタンを押すことで入力を取り消すことができます。正解すると正解のアラート、不正解だと不正解のアラートが表示されます。制限時間は3分間です。できるだけたくさん正解するように頑張ってください。

操作方法で何かわからないことはありますか？

課題を開始したらもし課題に問題点や疑問点があっても課題を継続してください。それでは始めてください。

実験 B の教示文

この課題は細かい操作の速さと正確さの能力を測るものです。実験を開始すると右のような画面が表示されます。

ブルーの通路の右上にブルーのボールが配置されています。このボールを指でスライドするとスライドした方向に移動します。ブルーの道からはみ出さないように一番下までボールを持っていってください。はみ出すとボールは最初に戻り最初からやり直しになります。正解すると正解のアラート、不正解だと不正解のアラートが表示されます。制限時間は3分間です、できるだけ多くの回数成功させてください。

操作方法で何かわからないことはありますか？

課題を開始したらもし課題に問題点や疑問点があっても課題を継続してください。それでは始めてください。

4.3.3 容積脈波 (BVP) の計測と動画の収録

実験では BVP を耳介用のセンサーによって測定し、合わせて実施の様子を第三者視点、スクリーンキャプチャの形式で動画記録した。実験の様子を図 4.8 に示す。



図 4.8: 実験の様子

4.3.4 アンケートの実施

実験の前後にウェブフォームの形式でアンケートを実施した。実験の前には被験者を特定するための情報、精神状態やストレスに影響を与える可能性のある就寝時刻と起床時刻、前

提知識に影響を与える趣味や職種の有無を尋ねた。実験の後には、それぞれの実験についての所感を自由記述で求めた。併せて前提知識の影響を調べるためにスマートフォンやコンピュータの使用歴を、ストレス耐性を調べるためにストレス反応尺度である SRS-18[57] の質問紙に基づく項目を尋ねた。

4.4 実験の結果

実験では実験 A, 実験 B それぞれの Good UI, Bad UI について、BVP と動画を記録し、アンケートの回答が得られた。アンケートによると、Good UI, Bad UI として想定したデザインについて全ての被験者から前者を肯定的に、後者を否定的に感じたという回答が得られた。アンケートで得られた被験者の概要を表 4.1 に、各実験に対する所感を集計した結果を表 4.2, 表 4.3, 表 4.4 に示す。自由記述で得られた所感は QDA ソフトの NVivo を用いて分類し集計した。但し、一人の被験者が複数の意見を記述している場合があるため回答の件数の合計は被験者数と一致しない。実験 A, 実験 B それぞれについて Good UI には肯定的な意見、Bad UI には否定的な意見が多かった。特に実験 B の Bad UI では、使いやすさとしては Good UI を良いとしたものの Bad UI について肯定的な反応を付け加える例があった。それらの意見では実験 B がゲーム性のあるアプリケーションであったことを挙げており、成功した際の達成感などについて述べていた。

4.5 実験の分析

得られた脈波は開発したシステムを使用し、実験 A Good UI, 実験 A Bad UI, 実験 B Good UI, 実験 B Bad UI 及びその前後の休憩時で分割、それぞれ ANB, LLE, RRI を算出した。本システムでは、第 3 章の実験と異なり、前述のように 10 秒毎の時系列変化としてそれらの値が表示される。表示される結果の例を図 4.9-4.12 に示す。これらのグラフは縦軸が LLE、横軸が 10 秒刻みの時間となっている。LLE のほかにも RRI, ANB について動揺にグラフにプロットしたが、いずれも大きな揺れがあった。また、1 つのグラフで表されるイベント期間中の代表値として平均値をとり Good UI と Bad UI を比較したが関連を見出すことができなかった。

4.6 実験の考察

本実験では、被験者の満足度について Good UI と Bad UI で明確に差があったにも関わらず、現時点で LLE, ANB, RRI と満足度の間の関連性を明らかにすることはできなかった。しかし、第 3 章の研究では、少なくとも ANB について満足度との関連性が明らかになっている。そのため、満足度とこれらの値に全く関連性がないわけではなく、時系列のストレス指標の分析方法を改めて検討する必要があると考える。

ID	A-1	A-2	B-1	B-2	C	年齢(歳)	性別	就寝時刻	起床時刻	(1)		最近のストレス反応			
										(1)	(2)	(3)	(4)	(5)	(6)
1	1	2	3	4	1	22	F	1:30	8:00	1	0	中	低	中	中
2	1	2	4	3	1	20	F	23:00	6:40	1	0	中	低	低	低
3	1	3	2	4	1	24	M	23:00	10:00	1	1	稍高	低	中	中
4	1	4	2	3	1	22	M	3:00	10:00	0	0	低	低	低	低
5	1	3	4	2	1	18	M	23:30	8:30	0	0	低	低	稍高	中
6	1	4	3	2	1	22	M	3:00	8:00	1	0	中	中	中	中
7	2	1	3	4	1	22	F	2:00	7:30	0	0	稍高	低	稍高	中
8	2	1	4	3	1	22	M	1:30	11:30	0	0	低	低	低	低
9	3	1	2	4	1	21	M	1:30	9:30	0	0	低	低	低	低
10	4	1	2	3	1	21	M	0:30	9:30	0	0	低	低	低	低
11	3	1	4	2	1	19	M	1:00	4:50	0	0	稍高	中	低	中
12	4	1	3	2	1	21	M	3:30	10:55	0	0	低	低	低	低
13	2	3	1	4	1	18	M	1:22	7:30	0	0	中	低	低	低
14	2	4	1	3	1	19	M	1:00	9:00	1	0	低	中	低	低
15	3	2	1	4	1	19	F	1:30	9:00	1	0	低	低	低	低
16	4	2	1	3	1	19	M	6:00	12:15	0	0	稍高	稍高	高	高
17	4	3	1	2	1	22	F	1:00	7:00	1	0	低	低	低	低
18	2	3	4	1	1	18	M	1:30	7:15	0	0	低	低	中	低
19	2	4	3	1	1	22	F	0:30	7:30	0	0	中	低	低	低
20	3	2	4	1	1	19	M	0:00	6:30	0	0	中	低	中	中
21	4	2	3	1	1	18	F	12:00	9:30	0	0	低	低	中	低
22	3	4	2	1	1	19	M	2:20	11:00	0	0	中	中	低	低
23	4	3	2	1	1	18	F	1:00	11:00	0	0	中	中	中	中
24	1	2	3	4	2	19	M	2:00	8:30	0	0	高	中	高	高
25	1	2	4	3	2	22	F	1:40	10:30	0	0	低	低	低	低
26	1	3	2	4	2	21	F	4:30	10:00	0	0	低	低	低	低
27	1	4	2	3	2	22	F	3:00	11:00	1	0	中	低	低	低
28	1	3	4	2	2	22	M	23:30	13:30	1	0	低	中	中	中
29	1	4	3	2	2	22	F	2:30	9:30	0	0	低	低	中	低
30	3	4	1	2	2	21	F	3:30	11:00	0	1	稍高	低	低	低
31	2	1	3	4	2	18	M	2:00	9:00	1	1	低	低	中	中
32	2	1	4	3	2	19	M	1:00	8:00	0	0	低	低	中	低
33	3	1	2	4	2	22	F	0:30	8:00	0	0	高	中	低	稍高
34	4	1	2	3	2	21	F	0:45	8:30	1	0	中	稍高	低	中
35	3	1	4	2	2	21	M	1:00	9:00	0	0	稍高	中	中	稍高
36	4	1	3	2	2	23	F	19:00	3:00	0	0	高	低	稍高	稍高
37	2	3	1	4	2	22	F	1:00	7:20	1	0	低	低	低	低
38	2	4	1	3	2	21	F	2:00	7:30	1	0	低	低	中	低
39	3	2	1	4	2	21	M	1:00	8:00	0	0	低	低	中	低
40	4	2	1	3	2	22	F	1:15	8:10			高	高	高	高
41	3	4	1	2	2	18	M	1:10	7:40	0	0	低	低	中	低
42	4	3	1	2	2	19	M	0:00	7:10	0	0	中	低	中	低
43	2	3	4	1	2	19	F	12:00	6:00	1	0	中	低	中	低
44	2	4	3	1	2	22	F	4:30	9:30	0	0	中	低	低	低
45	3	4	2	1	2	18	F	1:00	8:30	0	0	中	低	稍高	中
46	4	2	3	1	2	19	F	12:00	7:15	0	0	低	低	低	低
47	3	2	4	1	2	22	F	23:50	8:00	1	0	中	低	高	中
48	4	3	2	1	2	21	F	23:30	7:30	0	0	中	低	低	低

A-1, A-2, B-1, B-2はそれぞれ実験 A Bad UI, Good UI, 実験 B Bad UI, Good UI. 番号は実験の実施順を示す。

Cはアンケートの形式を示す(問題毎ページ遷移は1, セクション毎ページ遷移は2)

(1):職業・専攻・趣味がデザインや芸術に関連している(あてはまる場合は1)

(2):職業・専攻・趣味がソフトウェアのフロントエンドの開発に関連している(あてはまる場合は1)

(3):抑うつ・不安 (4):不機嫌・怒り (5):無気力 (6):合計

表 4.1: 時系列 UX 評価検証実験: 被験者の概要

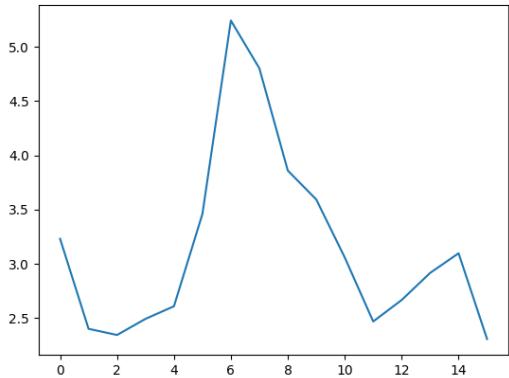


図 4.9: 被験者 1 実験 A Bad UI の LLE

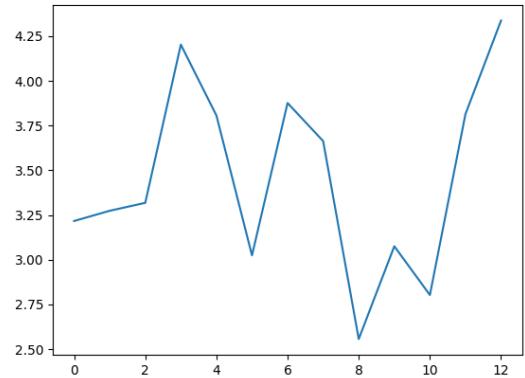


図 4.10: 被験者 1 実験 A Good UI の LLE

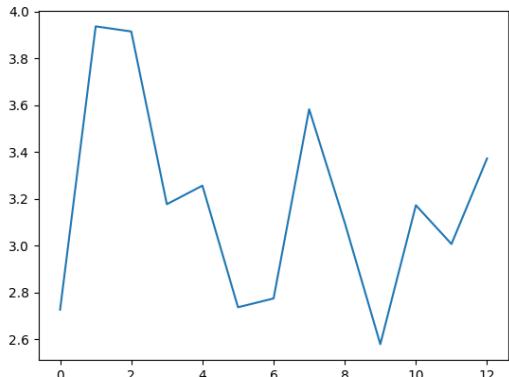


図 4.11: 被験者 1 実験 B Bad UI の LLE

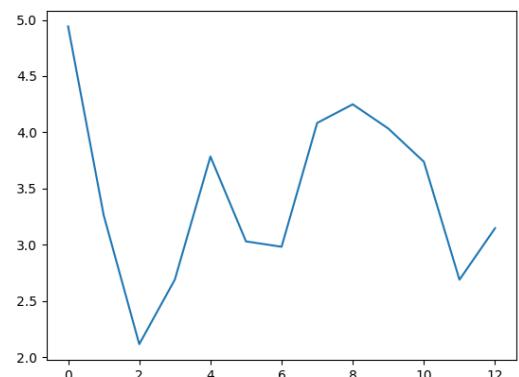


図 4.12: 被験者 1 実験 B Good UI の LLE

	実験 A		実験 B	
	Good UI(件)	Bad UI(件)	Good UI(件)	Bad UI(件)
肯定的	16	1	22	3
否定的	34	1	1	34

表 4.2: 時系列 UX 評価システム検証実験: マテリアルに対する肯定否定の別

	実験 A		実験 B		合計(件)
	Good UI(件)	Bad UI(件)	Good UI(件)	Bad UI(件)	
スムーズ	7		8		15
反応が良い	3		5		8
快適だった	3		5		8
正常だった	2		1		3
ゲーム性があった			1		1
達成感があった					2
順調だった				2	2
ストレスを感じなかった	1		1		2
挑戦したくなった	1				1
ミスが無かった				1	1
多く正解できた	1				1
楽しかった	1				1

表 4.3: 時系列 UX 評価システム検証実験: マテリアルに対する肯定的意見の内訳

	実験 A		実験 B		合計(件)
	Good UI(件)	Bad UI(件)	Good UI(件)	Bad UI(件)	
反応が悪かった		25		23	48
難しかった		4		8	12
ストレスを感じた		5		5	10
ミスがあった		8		2	10
焦った		6		1	7
悔しかった				5	5
作業のように感じた			1	1	2
動搖した		2			2
煩わしかった		1		1	2
モヤモヤした				1	1
何がダメかわからなかった				1	1
困った				1	1
慣れなかった		1			1
疲労を感じた				1	1

表 4.4: 時系列 UX 評価システム検証実験: マテリアルに対する否定的意見の内訳

今回の実験では、それぞれの実験を1つのイベントとして捉え、まとめて解析していたが、実際には実験内でも待つ時間、ボタンを押す時間、押し方を検討する時間などより詳細なイベントが存在している。それらのイベントとストレスの変化の関連性を改めて詳細に分析していく必要がある。また、ストレスがかかるイベントが発生してから各指標に反映されるまでの時間についても研究していく必要があると考える。

加えて、Good UI と Bad UI についての被験者の所感を分析した結果、特に実験 2において Bad UI として用意したものがポジティブに評価される例があった。ゲームのような機能よりもアプリケーションの使用自体が目的となる場合においてはユーザが感じる UX の基準がユーティリティアプリケーションとは異なり、ユーザビリティ(設計上の品質)の基準も異なっていなければならないということが示唆された。

第5章 結論

5.1 満足性評価のための指標

第3章では新たな満足性評価指標として ANB と LLE を提案し、次の結果が得られた。

1. ANB は満足性が高い UI を使用したときと低い UI を使用したときとの間で有意な差が見られ、満足性が高い UI を使用したときのほうが低くなる傾向が見られた。
2. LLE 及び HR では有意な差が見られなかった。

(1)ANB は満足性の評価指標として、UX 評価に有用である可能性が示唆された。また、いまだ手法が確立されていない UX 評価において、UX の重要な概念のひとつである満足性を定量的に評価する新たな手法として PPG を用いた BVP の測定と ANB の活用を提案した。

(2) 従来研究用途で活用されることがあった HR は用途や測定方法によっては活用が難しい可能性を指摘した。LLE については、数分レベルのまとまった時間の代表値として UX 評価に利用することは難しいことが明らかになった。

5.2 時系列 UX 評価システム

時系列 UX 評価システムとして、ANB, LLE, RRI を十数秒から数分程度の短いスパンで区切り一定秒毎にスライドさせながら繰り返し計算することで時系列でそれらの値を取得できるようにするシステムを開発した。また、実施時の状況を記録した動画やスクリーンキャプチャを脈波と同期させることでどのようなアクションを行った際にストレス指標が変化したのかを表示できるようにした。

システムの有用性を調査するため行った実験では、使用中に各指標の変化があったものの、Good UI と Bad UI の間でのストレス指標の変化を明確に示すことができなかった。

5.3 今後の課題

本研究では、PPG と ANB の活用の可能性を示すことができた。一方で残されている課題について述べる。

第3章で行った実験では、ANB の有用性が示唆された。しかし、実験で用いた Bad UI の例は限られており、UI の小さな差異など様々な UI の問題点を検出できるかどうか検討する

必要がある。これにより、実際の開発において ANB がどこまで信頼できるのかを明らかにできると考える。

時系列 UX 評価システムはプロトタイプであるため実用上の課題があり、改善していく必要がある。まず、BVP と動画をそれぞれ別に記録し後からシステムに読み込むのではなくシステム上で一元的に記録できるようにする必要がある。同時に、自動的に BVP と動画の同期を行うようにより同期精度を高め正確な測定が可能になる。次に、入力した BVP のエラー処理を改善する必要がある。現時点では BVP は目視でセンサーのずれによるノイズが無いかどうかを確認しているが、見落としをなくし簡便にするためには機械学習を用いて自動的に削除するなどの方法が考えられる。以上のような改善を通じてより正確なストレス評価を可能にする必要がある。

今回の実験では、各操作時のストレス指標の変化をグラフ化したもの、その解釈に検討の余地があった。今回のように、実験毎の分析ではなく、タップなどの操作レベルで変化を分析する必要がある。さらに、操作によって生じたストレス変化が指標に反映されるまでの時間についても検討する必要がある。

5.3.1 統合的なシステムとしての展望

本システムは、今後統合的な UX 評価システムとして発展させたいと考えている。その内容について述べる。

システムは BVP を測定する画面(図 5.1, 図 5.2), 被験者毎に分析する画面(図 5.3), 複数の被験者をまとめてそのストレス変化を画面遷移図の上に表示する画面(図 5.4)で構成される。

測定画面では、今回開発したシステムが備える動画と BVP の同期に加えて発言や操作ログなど様々な満足性指標を記録できるようにする。そして、分析画面で被験者を個別に分析し、そのデータを集約して画面遷移図上にストレスが高いと感じている被験者が多い部分を表示する。

このシステムによって、UX 評価が誰でも簡単に実施できるようになる。さらにユーザの個別の反応である UX を、複数人分まとめて画面遷移図上に可視化することで製品設計に落とし込むところまでを補助できるようにすることを目指している。



図 5.1: 測定画面のプロトタイプ(アプリ用)

図 5.2: 測定画面のプロトタイプ(VR用)

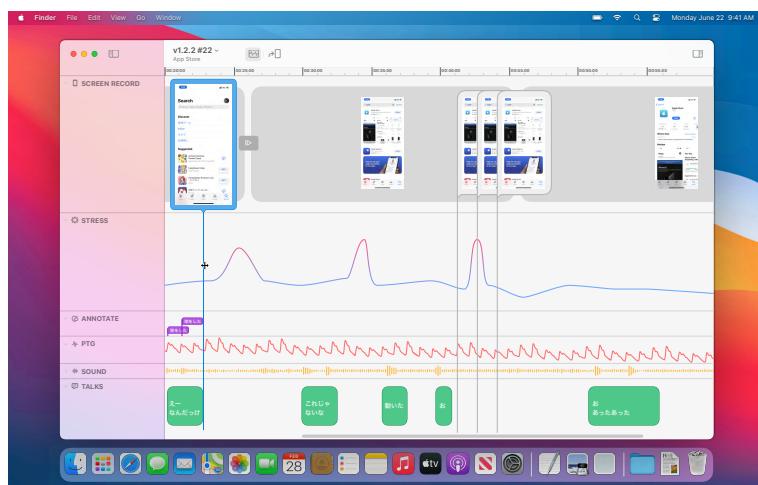


図 5.3: 分析画面のプロトタイプ

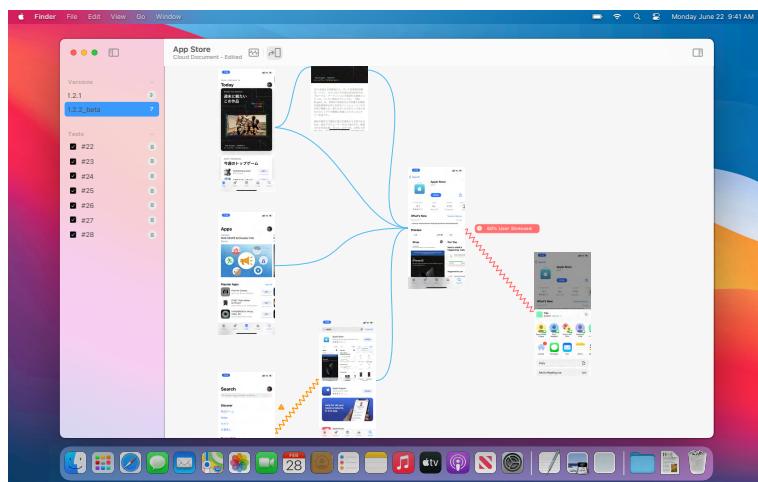


図 5.4: 画面遷移図での可視化画面のプロトタイプ

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、博士課程の大和比呂志氏、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] 黒須正明. 黒須教授のユーザ工学講義 UX グラフ 最新版. <https://u-site.jp/lecture/ux-graph-2>. Accessed: 2022-01-25.
- [3] 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.
- [4] Nielsen Norman Group. 10 usability heuristics for user interface design. <https://www.nngroup.com/articles/ten-usability-heuristics/>. Accessed: 2022-01-13.
- [5] 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>. Accessed: 2022-01-13.
- [6] 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.
- [7] Balmuda. バルミューダについて. <https://www.balmuda.com/jp/about/>. Accessed: 2022-01-13.
- [8] 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. Accessed: 2022-01-13.
- [9] 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. Accessed: 2022-01-13.
- [10] Brian Shackel and Simon J Richardson. *Human factors for informatics usability*. Cambridge university press, 1991.
- [11] Jakob Nielsen. 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.
- [12] 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.

- [13] 黒須正明. 人間中心設計の基礎. HCD ライブライバー. 近代科学社, 2013.
- [14] Patrick W Jordan. *Designing pleasurable products: An introduction to the new human factors*. CRC press, 2000.
- [15] 黒須正明. 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.
- [16] SHIFT. UX 開発の実態調査. <https://topics.shiftinc.jp/ux/download/>. Accessed: 2022-01-25.
- [17] 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, *BIO SIGNALS 2013 - Proceedings of the International Conference on Bio-inspired Systems and Signal Processing, Barcelona, Spain, 11-14 February, 2013*, pp. 298–301. SciTePress, 2013.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] Oxford University Press. usability. <https://www.oed.com/view/Entry/276161?redirectedFrom=usability>. Accessed: 2022-01-13.

- [23] J. Nielsen. *Usability Engineering*. Interactive Technologies. Elsevier Science, 1994.
- [24] 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.
- [25] 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>. Accessed: 2022-01-14.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] Tobii. <https://www.tobiipro.com/ja/product-listing/>. Accessed: 2022-01-14.
- [30] 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.
- [31] Rosalind W Picard and Jocelyn Scheirer. The galvactivator: A glove that senses and communicates skin conductivity. In *Proceedings 9th Int. Conf. on HCI*, 2001.
- [32] 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.
- [33] Foroohar Foroozan and Jiang Wu. 手首からの光電式容積脈波記録信号に基づく脈拍数変動解析のための高信頼性の心拍検出アルゴリズム. <https://www.analog.com/jp/technical-articles/robust-beat-to-beat-detection-algorithm-for-pulse-rate-variability-analysis.html>. Accessed: 2021-01-14.

- [34] 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.
- [35] Kirk Shelley and S. Shelley. *Pulse Oximeter Waveform: Photoelectric Plethysmography*, pp. 420–423. 01 2001.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] Paul M Lehrer and Richard Gevirtz. Heart rate variability biofeedback: how and why does it work? *Frontiers in psychology*, Vol. 5, p. 756, 2014.
- [40] TRYTECH. チェックマイハート. <https://www.trytech.co.jp/checkmyheart/glossary.html>. Accessed: 2022-01-13.
- [41] 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.
- [42] Thomas Wollmann, Farhad Abtahi, Aboozar Eghdam, Fernando Seoane, Kaj Lindencrantz, 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.
- [43] P. Berg'e and Y. Pomeau. カオスの中の秩序 亂流の理解へ向けて. 産業図書, 1992.
- [44] 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.
- [45] 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.
- [46] 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.

- [47] 雄山真弓. 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.
- [48] 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.
- [49] 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.
- [50] 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.
- [51] 小孫康平. 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.
- [52] 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>. Accessed: 2022-01-13.
- [53] Jacob Cohen. *Statistical power analysis for the behavioral sciences*. Academic press, 2013.
- [54] 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.
- [55] Mozilla and individual contributors. Web Serial API. https://developer.mozilla.org/en-US/docs/Web/API/Web_Serial_API. Accessed: 2022-01-13.
- [56] Mozilla and individual contributors. File System API. <https://developer.mozilla.org/ja/docs/Web/API/FileSystem>. Accessed: 2022-01-13.
- [57] 鈴木伸一, 嶋田洋徳, 三浦正江, 片柳弘司, 右馬埜力也, 坂野雄二. 新しい心理的ストレス反応尺度 (srs-18) の開発と信頼性・妥当性の検討. 行動医学研究, Vol. 4, No. 1, pp. 22–29, 1997.

付録A 略語表

略語	英語表記	日本語表記	備考
ANS	Autonomic Nervous System	自律神経系	
BVP	Blood Volume Pulse	容積脈波 / 脈波	
ECG	Electrocardiogram	心電図	
EEG	Electroencephalography	脳波記録	
GSR	Galvanic Skin Response	皮膚伝導率	
HF	High Frequency	高周波	脈波の高周波成分
HR	Heart Rate	心拍数	
HRV	Heart Rate Variability	心拍数変動	
LF	Low Frequency	低周波	脈波の低周波成分
LLE	Largest Lyapunov Exponent	最大リヤプノフ指数	
PPG	Photoplethysmography	光電式容積脈波記録	
PRV	Pulse Rate Variability	脈拍変動	
RRI	R-R Interval	R波感覚	瞬間の心拍数
SQuaRE	Software Product Quality Requirements and Evaluation	システム及びソフトウェア製品の品質要求及び評価に関する国際規格	ISO/IEC25000
UX	User Experience	ユーザ体験	

表 A.1: 略語表

付録B 時系列UX評価システムの妥当性調査実験 質問紙

B.1 事前アンケート

1. 年齢
2. 性別 (Biological)
選択肢
 - 男性
 - 女性
 - その他
 - 無回答
3. 昨日の就寝時刻
4. 今日の起床時刻
5. 以下の条件で当てはまるものにチェックを入れてください
選択肢
 - 職業・専攻・趣味がデザインや芸術に関連している
 - 職業・専攻・趣味がソフトウェアのフロントエンドの開発に関連している

B.2 事後アンケート

1. 「数字を入力する実験」を2回行ったときに気づいたこと・感じたことを何でも自由に記述してください
2. 「ボールをなぞって移動させる実験」を2回行ったときに気づいたこと・感じたことを何でも自由に記述してください
3. 普段次のデバイスを使用する頻度を教えてください
 - Androidのスマートフォン
 - iPhone
 - その他のスマートフォン

- Android のタブレット
- iPad
- その他のタブレット PC
- Windows の PC (タブレット以外)
- Mac
- その他の PC

選択肢

- 1 時間に 1 回以上
- 1 日に数回
- 1 日に 1 回
- 週に数回
- 週に 1 回未満
- 使わない (所有していない)

4. 次のデバイスを日常的に使用している (使用していた) 期間を教えてください

- Android のスマートフォン
- iPhone
- その他のスマートフォン
- Android のタブレット
- iPad
- その他のタブレット PC
- Windows の PC (タブレット以外)
- Mac
- その他の PC

選択肢

- 1 年未満
- 3 年未満
- 5 年未満
- 使っていたことはない

5. スマートフォンで文字を入力する際によく使う入力方法は何ですか

選択肢

- フリック入力 (テンキー)
- ローマ字入力
- 音声認識
- 手書き入力

6. 次の項目にそれぞれ最もあてはまるものを選んでください
選択肢

- アプリのアップデートでデザインが変わるとワクワクする
- アプリのデザインが変わると憶えるのが面倒くさい
- アップデートが来たら直ぐにアップデートするほうだ

7. 以下にあげる質問は、あなたのここ 2, 3 日の感情や行動の状態にどのくらい当てはまりますか。最も当てはまる数字を 1 つだけ〇で囲んで下さい

- 怒りっぽくなる
- 悲しい気分だ
- 何となく心配だ
- 怒りを感じる
- 泣きたい気持ちだ
- 感情を抑えられない
- くやしい思いがする
- 不愉快だ
- 気持ちは沈んでいる
- いらいらする
- いろいろなことに自信がない
- 何もかもいやだと思う
- よくないことを考える
- 話や行動がまとまらない
- なぐさめて欲しい
- 根気がない
- ひとりでいたい気分だ
- 何かに集中できない

選択肢

- 全く違う
- いくらかそうだ
- まあそうだ
- その通りだ

付 錄C 時系列UX評価システムの妥当性調査実験 自由回答

C.1 実験A

- どちらも正解だと思った答えを入力しても、正解にならなかったのは不思議だと思いました。1回目の方は入力するたびに焦りとイラつきがありましたが、2回目の方はずっと不正解でも何度もトライしようという気になりました。
- 1回目の実験のとき、何度タップしても反応がなかったので、少し焦った。
- 1回目について、まず初めに入力したはずの数字が入力されず、困惑した。その後何回かランダムに押せば入力すればいいと気づくも、課題通りに入力したはずなのに不正解となり、始めに聞いた課題を書き間違えたのかと思って焦った。何回か小さい順から入力後、違うアルゴリズムで正解になるパターンだと考え、大きい順から入力するも不正解となり、表示されたまま入力すると正解したので意図を理解した。1回目はその理解までに1分半くらいは使ったと思う。誤入力もありストレスだった。2回目は念のため言われた通りの課題をこなしたが不正解となつたので、そのまま入力するのに徹した。タップしたらそのまま数字が入力されるので快適だった。
- 1回目ではテンキーの入力判定が悪く、2回目では普段利用している時と変わらない判定であった。
- ”1回目はボタンの感度が悪く、6回押さないと数字が入力できないときがあったが、2回目は一回押しただけでボタンが反応した。スマホの反応の速さの違いによる感情の変化を調べる実験だったのか？”
- 2回目の数字が押しにくかったです。
- 素直に反応してくれないとイライラする
- 1回目のときは、スムーズに数字を当てたが、2回目のときは、何度も押しても反応しない時があったので動搖しました。
- クリアボタンの位置が悪魔的
- はじめは反応悪いだけかなと思ったけど、ひとつ前の実験から察するに、「そういうやつね」となった。
- 1回目の実験では1度のタップで反応しない場合もあったため、入力に煩わしさ対して2回目の実験では全ての操作が1度のタップで行えたためスムーズな入力感だった。
- 1回目に行った時は数字の打ち込みがとても難しく、スマホの画面に問題があるのでないかと思うほど打ち込みにくかった。しかし、2回目はとてもスムーズにいつもスマホを触る時と同じように打ち込むことができた。

- 先にスムーズに操作できる回からだったため、2回目の異常さにはすぐに気がつきました。初めは、わざととは思わず、何か実験に不備が出てしまうのではないかと心配になりました。スムーズに操作できないと、押し間違いのミスをしてしまった時のあつ。という気持ちが大きくなり、自分の焦りに繋がるように思いました。2回目の方が、スピードよりも正確性重視で操作を行なっていたように思いますし、小さなストレスも多かったと感じます。
- 2回目に数字を入力するのが、うまく行く場合と行かない場合が不規則だったので、最後までなれなかった。
- 1回目はスムーズに入力できたのに対して、2回目はとても押し辛かったです。
- 2回目の方が反応がよかったです。1回目の方は何回も数字を押さないと反応しなかった。
- タップした時に自分が思ったように数字を押せていなかったら焦ってミスが生れた。また、何回も繰り返していくと最初より精度が落ちた気がした。
- 2度目はテンキーの反応が悪く苛立ちがあったため、クリア・決定の操作も誤ることがあった
- 2回目の方が入力しづらかった。
- 1回目の何度も押さないといけない方は、煩わしかった。
- 2回目の時に一度数字を選択しても反映されなくて、連打しているうちに2回押しちゃつてまた1番最初からやり直しということが何回かあって、少し焦った。
- 1度目は何度か押しても反応しない場合があった2度目はスムーズに行けた
- 1回目は数字のボタンが押しにくく何回も入力し直した。2回目はスムーズに入力できストレスフリーで行うことができた。
- 1回目に数字を押した時は反応しづらかった。直前のアンケートでは普通に反応していたため、何かそういうアプリかなあと思いました2回目は普通に打てたので楽しかったです
- 初めの方は動きが悪くて、やりづらいなと思いました。
- 1回目の実験で入力する際に、押しても反応しない時があったのがとてもいらいらしました。2回目は押しやすかったです。
- 一回目の実験の操作は、数字ボタンが反応しないことがあり難しかった。しかし、クリアと決定ボタンには不具合がないのが不思議だった。二回目の実験の操作は正常に反応したため快適に感じた。
- 2回目の実験の際に、ボタンの反応が悪くなり、入力ミスを起こしてしまった。
- 1回目は一度入力すれば文字が反映されたのに対し、2回目はランダムで1~5回タップしないと文字が反映されなかった
- バラバラに数字を並べられると意外と難しい。2回目はわざと一発で入力できないようになされているのだなと感じた。自分のめんどくさがりで短気なところから何回もタップして同じ数字を2回入力してしまったりした。単純にゲーム感覚で面白かった。
- 1回目の実験ではレスポンスが早く感じたが2回目の実験ではレスポンス以前の問題で数回押してもなぜか反応しないことがあったため意図的に反応させないとしたら結構悪用が可能なのではと考えている。

- 2回目では画面の反応が悪くストレスを感じた。また、なぜか問題解答の正確さも落ちてしまった。
- 1回目は普通だったけれど、2回目は数回に1回だけ反応していた。
- 1回目は数字が打ちづらかった。
- 一度目は反応が鈍く反映されにくく、二度目は普通に押せば反映されるように感じた
- 入力のしやすさに差があった。
- 2回目の実験は入力しづらかった。
- 1回目の時は少し反応が鈍かったので、2回目の時の方が多く正解することが出来ました。
- 2回目にストレスがかかるように設計されたゲームであると気づいた。反応が悪いほうをするときの方が慎重になってゲームしていた気がする。反応しないと思って連打するときもあったためミスも増えたように感じた。
- 1回目は何度かタップしないと文字入力ができなかった。2回目はスムーズにできた。
- 初め、文字が打ちにくくて機械の不具合かと思った。
- 6つの数字を、3つずつに分けて考えると素早く判断することができると感じた。何回か繰り返して慣れてくると、逆に入力ミスや小さい数字の見落としなどのミスが増えてきたように感じた。
- 数字のボタンを押しても反応しないことがある。
- 1.2回目共、数字のキーをなぞるように打っていくとミスが少ないと感じ、そのように行つた。1回目と比較して2回目では数字を打つのが難しく、その分ミスマッチが多くなったため慎重に行う方がいいと判断し、そうした
- 2回目は1回目と違って、テンキーが反応しない時があった。

C.2 実験B

- 1回目の方は時間内に1回だけしかクリア出来なかつたので、ゲーム性があつて面白かった。2回目の方はスムーズに操作できたが、何をしてもクリアしたので、ゲーム感覚よりかは作業のように感じた。
- 2回目の実験のとき、ボールが思ったように動かなかつた。半分以上進んだときにやり直しになったときは、少し悲しかつた。
- 数字の入力を経験した後なので1回目のボールの実験も面倒なやつだらうなと思って取り組んだ。最後まで行けなかつたのが悔しい。2回目はとても快適だつた。
- 1回目ではボールを動かす際の移動距離はランダムであるような気がした。2回目では指の動きとボールの動きがリンクしており、ゴールが容易であった。また、2回目ではみ出した判定が線に対してボールの1番離れた所となつていた。
- 1回目は指でなぞった通りにボールが動いたが、2回目はボールの動きを操れなかつた。スマホの反応の速さの違いによる感情の違いを調べる実験なのか？
- 一回目が楽だったので、二回目はめんどくさくなると予測していたが想像以上だったので、かなりのストレスだつた。

- 1度目のものは何がだめで振り出しに戻されてるのかわからなかったです。2回目もなぜ戻されているのかわかりませんでしたが、スムーズに動く分やりやすかったと思います。
- 2回目反応が悪く球を動かしにくかった
- ひたすら同じことを繰り返すと飽きてくる。
- 1回目のときは、なかなかボールが動かなくて困りましたが、2回目のときはスムーズに動くようになったので一瞬さっきの動かし方が悪かったのかと思いました。
- 2回目が特に難しい
- はじめは反応悪いだけかなと思ったけど、時間が経っても改善されないので「そういうやつね」となった。1回目が順調だっただけに、2回目が一回も成功できずモヤモヤした。
- 1回目の実験では指での入力に対して不自然な反応だったため、少しいらだちを感じて1度しか成功しなかった。対して2回目の実験では、入力に対して素直な反応だったため、成功回数が一回目に比べて多くなった。
- 数字を打ち込む作業と同じくとてもやりづらく、指先を繰り返し小刻みになぞる動作だった。とてももどかしく感じた。2回目はなぞることがとても楽で、あまり神経質にならずにスピードを速めながらなぞることができた。
- 先に、上手く操作ができない回からおこなったため、この感度が正常なのかと無理やり思い込んで頑張ろうとしていたように思います。そのため、うまく操作できない自分への焦りを感じました。身近な機器がスムーズに操作ができるることは、勉強や自由時間のクオリティを格段に上げてくれていることを実感したように思います。
- 1回目は、あまりにもタッチの反応が悪かったので、最後までうまく行く方法がわからなかった。2回目は1回目と比べて驚くほど反応が甘かった。
- 一回目全くボールが動かなくて、最初の坂で止まってしまった。2回目はスムーズでやりやすかった。
- 1回目の方が2回目よりも反応が良かった。しかし、どちらも数字なら実験に比べると反応が悪かった気がした。
- こちらも、最初の方はすぐにゴール出来たが、繰り返していくうちに思った位置までボールがきていないことが増えミスが増えた。一回ミスすると繰り返してしまった
- 2回目は反応が悪く思った通りに動かなかったがむしろ集中できた
- 2回目の方が反応が悪く、1回もクリアできなくて残念だった。
- 携帯が、ちゃんとと思い通りに動いているのは、嬉しい。感謝の気持ちが芽生えた。
- 1回目は一度もミスすることなく何回も成功できたが、2回目は全然指に合わせてボールが動かなくて少しイララした。一度は成功できたものの、他はすぐに一番最初からになったので悔しかった。
- 1度目は進むスピードが遅く、少しのズレで最初からのやり直しだったため、難しかった
- 1回目はスムーズにボールを動かすことができ、はみ出してもまたやり直せばいいという考え方から素早く操作できたが、2回目は動かしにくかったため慎重に動かしすぎて時間内にゴールまで辿り着けなかった。

- 数字の方と交互にやったため、2回目は触りやすくなっているのかなと寧ろ3分間あの単調な作業をする方が頭おかしくなりそうです
- 初めの方は一筆ではかけず、2回目はスイスイかくことができました。
- 2回目の実験で、ボールをうまく進められなったり、なんとかして進めようとしたが最初に戻ったりして、とてもイライラしました。1回目はスムーズにできました。
- 一回目の実験の操作は、なぞった際に進む速度がバラバラで難しかった。また、途中でボールの上に指を合わせる必要がないことに気づいた。二回目の実験の操作は、ボールの上をなぞった通りに動いたため快適だった。一回目よりもアウト判定がゆるいように感じた。
- 1回目の実験ではボールの反応が悪かったものの、2回目では回復したため、少し驚いた。
- 1回目はスムーズにすらすらとボールが動いたのに対し、2回目は徐々にしか動かず、1回目よりも2回目の方が「線から外れた判定」が厳しいと感じた
- ボールをスライドさせるという説明文では指をボールに置くのか進行方向に指を動かすのかわからなかった。一回目は非常にカクカクしていてやりづらかった。これは数字と反対でわざと一回目がやりづらくされていたと気づいた。一回目がカーブのところでコースアウトしてしまったのでスムーズに動く二回目でも慎重になった。
- 一回目は明らかにレスポンスが悪く実験どころではない状況であった。ゲームであるをやられたらレビューが悲惨なことになること間違いなしである。2回目は大凡のレスポンスが改善されておりあの動作ならまだ問題ないと思う。
- 2回目では思った通りにボールが動かずより正確になぞれるように注意した。一回しかうまくいかなかったがうまく行った時は達成感を感じた。
- 一回目は普通だったけれど、2回目は滑りづらくて、私の手が人間じゃなくなったのかとおもった。そういう実験なのかなぁどっちなのかなあと思っていたら2回目の数字もへんだったからそういう実験なのだなと思った。
- 1回目はボールが動かしづらかった。
- 一度目は有効な範囲が狭く、触っても反応しているかわからないようなほどしか進まなかつたが、二度目はさらさら出来たため、どこまではみ出して動かしてもいいか試したりしながら進めた。だいぶはまだしても大丈夫だったように感じた。
- 2回目の実験が動かし易かった。
- 1回目の実験はなぞりにくかった。
- 1回目はボールが少しずつしか進まなかつたので、2回目の方がスムーズに行うことができました。
- 1回目の方がボールの反応が悪く自分が思った通りに動かなくストレスがあった。2回目と比べた後の疲労感という意味でも1回目の方がしんどかった。1回目は1回もクリアできていなかったため、最後になればなるほど焦りが出てきてまたミスも増えたのかもしれない
- ボールが1回目に比べてとても動かしにくかった。少しでもずれると最初に戻ってしまうので難しかった。

- 二回目はボールが動きづらくなっていたため、非常にやりづらく、多少の苛立ちがあった。
- コツを掴むと素早くなぞることができたが、一度変な癖がついてしまって何回か連続してミスしてしまった。途中から綺麗になぞれてるのかよく分からない感じがした。
- 指でなぞっている方向にボールが移動しなかった。スムーズに移動しなかったためやりにくかった。
- 数字の実験と同様に、1回目と比較して2回目はなぞりにくかった。そのため、1回目は一筆書きであまり道からずれることを考えずに行っていたが、2回目は何度もタップし直し、道からそれないように慎重になった。その分1回ゴールした際の達成度が高かった。
- 2回目は指を動かしてもボールがついてこなかった。1回目の曲がり角まで行くとスタート位置に戻って、ボールが一瞬赤くなった。

付 錄D 時系列UX評価システムの妥当性調査実験 選択回答

D.1 デジタル機器対応状況

デバイスの使用期間・使用頻度の項目19は次の通り。

0. Android スマートフォン
1. iPhone
2. その他のスマートフォン
3. Android タブレット
4. iPad
5. その他のタブレット
6. Windows PC
7. Mac
8. その他のPC

使用頻度の選択肢は次の通り。

0. 使わない
1. 週に1回未満
2. 週に数回
3. 1日に1回
4. 1日に数回
5. 1時間に1回以上

使用期間の選択肢は次の通り。

0. 1年未満

1. 3年未満
2. 5年未満
3. 使っていたことはない

括弧で示した質問項目は次の通り。

(1) スマートフォンで文字を入力する際によく使う入力方法は何ですか

選択肢

0. フリック入力(テンキー)
1. ローマ字入力
2. 音声認識
3. 手書き入力

(2) アプリのアップデートでデザインが変わるとワクワクする(あてはまると回答:1)

(3) アプリのデザインが変わると憶えるのが面倒くさい(あてはまると回答:1)

(4) アップデートが来たら直ぐにアップデートするほうだ(あてはまると回答:1)

ID	デバイスの使用頻度									デバイスの使用期間												
	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	(1)	(2)	(3)	(4)
1	0	5	0	0	3	0	4	0	0	3	2	3	3	2	3	1	3	3	0	1	0	0
2	4	2	2	2	2	2	5	2	2	2	3	3	2	3	3	1	3	3	0	1	0	0
3	0	5	0	0	5	0	1	5	1	3	2	3	3	2	3	2	2	0	0	0	1	0
4	4	0	0	0	0	0	5	0	5	0	3	3	3	3	3	3	0	0	1	0	1	0
5	4	0	0	0	0	0	4	0	0	2	3	3	3	1	3	2	3	3	0	0	1	0
6	5	0	0	0	5	0	5	0	0	2	3	3	1	1	3	2	3	3	1	0	1	0
7	0	5	0	0	4	0	5	0	0	3	1	3	3	2	0	1	3	3	1	1	1	0
8	0	5	0	0	0	0	4	0	0	0	2	3	3	3	1	3	3	0	0	1	0	0
9	0	5	0	0	4	0	2	1	0	3	2	3	3	2	3	2	2	3	0	1	0	1
10	0	5	0	0	4	0	0	4	0	3	2	3	3	1	3	3	2	3	0	1	0	1
11	0	4	0	0	5	0	1	0	0	3	2	3	3	2	3	2	3	3	0	1	0	0
12	0	5	0	0	0	0	2	0	0	3	2	3	3	3	1	3	3	0	1	0	0	0
13	0	5	0	0	4	0	2	0	0	3	2	3	3	1	3	1	3	3	0	1	0	1
14	0	5	0	0	1	0	0	4	0	3	2	3	3	2	3	1	1	3	0	1	0	1
15	0	5	0	0	4	0	4	0	0	3	2	3	3	2	3	0	3	3	0	0	1	0
16	0	0	0	0	4	0	0	5	0	3	2	3	3	2	3	3	2	3	1	1	0	0
17	0	5	4	0	0	0	4	0	0	1	0	3	3	3	1	3	3	0	0	1	0	0
18	0	4	0	0	4	0	4	0	0	3	2	3	3	2	3	1	3	3	0	0	0	1
19	0	5	0	0	0	0	2	0	0	1	2	3	3	3	2	3	3	0	0	1	0	0
20	5	0	0	0	2	0	5	0	0	1	3	3	3	2	3	1	3	3	0	1	0	0
21	1	5	0	0	0	0	4	0	0	0	1	3	3	1	3	0	3	3	0	0	0	1
22	0	5	0	0	1	0	2	0	0	2	2	3	3	2	3	0	3	3	0	0	1	0
23	0	5	0	0	5	0	5	0	0	1	2	3	3	1	3	0	3	3	0	1	0	0
24	0	5	0	0	4	0	3	0	0	1	0	3	3	0	3	0	3	3	0	0	1	0
25	0	5	0	0	1	1	0	0	0	3	2	3	3	2	2	3	3	3	0	0	1	0
26	1	4	0	0	0	0	3	0	0	0	2	3	3	3	2	3	3	3	0	1	0	0
27	0	5	0	0	0	0	2	0	0	3	2	3	3	3	2	3	3	3	0	0	0	1
28	5	0	0	0	0	0	2	0	0	2	3	3	3	3	2	3	3	3	0	0	0	1
29	5	5	5	5	5	5	5	5	5	0	0	0	0	0	0	0	0	0	4	0	0	0
30	0	5	0	0	5	0	4	5	0	3	2	3	3	2	3	3	2	2	0	0	1	0
31	0	5	0	0	2	0	1	4	0	3	2	3	3	2	3	0	1	3	0	1	0	1
32	0	5	0	0	4	0	4	5	0	0	2	3	3	2	3	2	1	3	0	1	0	1
33	0	5	0	0	0	0	4	0	0	3	2	3	3	3	2	3	3	3	0	0	1	0
34	5	0	0	0	0	0	4	0	0	2	3	3	3	3	2	3	3	3	0	1	1	0
35	5	0	0	3	0	0	3	0	0	2	3	3	1	3	3	2	3	3	0	1	0	1
36	2	5	2	2	5	2	2	5	2	3	0	3	2	0	1	2	0	3	0	0	0	1
37	0	5	0	0	0	0	2	0	0	3	0	3	3	3	0	3	3	3	0	0	0	1
38	0	4	0	0	0	2	2	0	0	3	2	3	3	3	0	2	3	3	0	0	0	1
39	0	4	0	0	4	0	0	4	0	3	2	3	3	1	3	3	1	3	0	0	1	0
40	0	5	0	0	0	0	4	0	0	3	2	3	3	3	2	3	3	3	0	0	0	1
41	0	5	0	0	4	0	4	0	0	1	2	3	3	2	3	0	3	3	0	0	0	1
42	0	5	0	0	4	0	4	0	0	3	1	3	3	2	3	0	3	3	0	0	0	1
43	0	5	0	0	1	0	0	4	0	3	2	3	3	1	3	3	0	3	0	0	1	0
44	0	5	0	0	0	0	4	0	0	3	2	3	3	3	3	2	3	3	0	0	0	1
45	0	5	0	0	5	0	4	0	0	3	2	3	3	2	3	0	3	3	0	1	1	0
46	0	5	0	0	1	0	0	5	0	3	2	3	3	2	3	3	0	3	0	0	0	1
47	2	5	2	2	2	2	5	2	2	3	2	3	3	3	2	3	3	3	0	0	0	1
48	0	5	0	0	2	0	4	0	0	3	1	3	3	2	3	1	3	3	0	1	0	0

表 D.1: 時系列 UX 評価検証実験: 被験者のデジタル機器対応状況