

オンラインアンケートの回答信頼性検証に向けた 回答時画面操作ログ取得システム

後上 正樹^{1,a)} 松田 裕貴^{1,b)} 荒川 豊^{2,c)} 安本 慶一^{1,d)}

概要: 近年、オンラインアンケートは社会調査に広く用いられている。しかし、同一の回答者に反復的に依頼すると、飽きや心理的バイアスによって丁寧な回答を得辛くなる。また、回答によって報酬を付与する形式の一般的な依頼構造においては、単発の実施であっても努力の最小限化によって結果の信頼性が侵される可能性がある。本研究では、回答時の迷い等がタッチジェスチャーに表出するという仮説を立て、スクロール量や質問単位の回答時間、選択肢の変更等を時系列データとして記録するシステムを開発した。本システムを適用するにあたり、回答者側は追加のソフトを導入する必要がない。本システムで得られたデータについて分析し、有用性を確認した結果について報告する。

キーワード: オンラインアンケート, 努力の最小限化, 回答信頼性, タッチジェスチャー

Screen operation acquisition system for verification of answer reliability of online survey

MASAKI GOGAMI^{1,a)} YUKI MATSUDA^{1,b)} YUTAKA ARAKAWA^{2,c)} KEIICHI YASUMOTO^{1,d)}

1. はじめに

近年、クラウドソーシングによるオンラインアンケート調査が社会科学関連の研究にも用いられている。これは、インターネットを介して全世界の参加希望者を対象としてアンケートの回答を依頼することができる仕組みである。これを用いることで、一度に大量のデータを取得できる上に、比較的容易にデータの代表性を保障できる等の利点が挙げられる。しかし、アンケート調査では回答者が必ずしも正確に回答するとは限らない。適切な認知コストを払わずに回答された結果から導かれる推論は真実とは異なるものになってしまう可能性がある。そこで、satisficingを検

出することができれば、それを含む回答に対して適当な処理を施すことでより真実に近い知見を得ることができる。

Simonら [1] は、人間の認知的資源には限りがあることによって生じる、アンケート調査において回答要求に対する努力を最小化しようとする傾向を努力の最小限化 (satisficing) と定義した。また、Krosnickら [2] は satisficing の分類として、選択肢を十分に検討せず回答する場合を弱い satisficing、全く認知コストを払わずあてずっぽうに回答するような場合を強い satisficing と分類した。三浦ら [3] は、IMC (Instructional Manipulation Check) や DQS (Directed Question Scale) という、いわばひっかけ問題のようなものを設問文や選択肢に混入したアンケートを作成し、回答結果から強い satisficing の検出を試みた。しかし、この方法ではひっかけ問題を配置した箇所以外での satisficing が検出できない。また、設問数が増えるため、回答者の負担が増加するという欠点も考えられる。

近年、オンラインアンケートの回答に用いられるデバ

¹ 奈良先端科学技術大学院大学,
Nara Institute of Science and Technology

² 九州大学, Kyushu University

a) masaki.gogami@ubi-lab.com

b) yukimat@is.naist.jp

c) arakawa@ait.kyushu-u.ac.jp

d) yasumoto@is.naist.jp



払われる認知コスト：少
→ satisficing：あり
→ 結果の信頼性：低

払われる認知コスト：多
→ satisficing：なし
→ 結果の信頼性：高

(a) 払われる認知コストが少なく、結果の信頼性が低い例
(b) 払われる認知コストが多く、結果の信頼性が高い例

図1 スマートフォンの操作状況がアンケート結果の信頼性に影響を与える例

スとして、スマートフォンがPCに取って代わってきている。これを受けて、Rogerら[4]は、PCおよびタブレットとスマートフォンでアンケート結果の質にどのような違いがあるのかを調査し、結果の信頼性についてどのデバイスについても特に差はないと結論づけている。したがって、スマートフォンは他のデバイスと比べてオンラインアンケートを回答するデバイスとして不適切でないと言える。また、その高いポータビリティにより今後さらにオンラインアンケートの回答に用いられると予測される。

そこで本研究では、図1のようなオンラインアンケートをスマートフォンで回答する際に払われる認知コストが少ないことによって発現するsatisficingを検出することを目指とする。satisficingの検出ができれば、信頼性が低い回答結果を抽出して排除することができ、アンケート結果を用いた推論の精度向上に貢献することができる。

その目標に際して本稿では、オンラインアンケートを回答する際に発生するsatisficingを検出するための指標について検討し、その指標を定量的に記録するシステムを実装・評価した結果について報告する。

本稿の構成は次の通りである。2章では関連研究について述べる。3章でsatisficingを検出するための指標について述べる。4章では、3章で検討した指標を取得するアンケートシステムについて述べる。5章で評価実験の設計および実験方法について説明し、6章に結果を記す。最後に、7章で本稿のまとめと今後の展望を述べる。

2. 関連研究

Rogerら[4]は、PCおよびタブレットとスマートフォンでアンケート結果の質にどのような変化があるのかを調査した。この際、評価基準としたのは回答時間や未回答率および題意に沿わずある一定の規則に沿って回答する行為を指すストレートライニングである。結果として、スマートフォンはPCおよびタブレットに比べて回答時間が長い傾向が観察された。しかし、結果の信頼性についてはどのデ

バイスについても特に差はないと結論づけている。

Janら[5]は、選択肢がどの設問でも固有で賛成と反対の要素で構成されるA/D (Agree/Disagree)形式と選択肢が設問ごとに対応付けて設定されているIS (Item-Specific)形式のアンケートを比較する実験を実施した。アイトラッキングによって回答時の視線を記録したところ、IS形式の方が注視回数が多く、注視時間が長く、再注視回数も多かったと報告した。

加藤ら[6]は、スマートフォンにおけるwebブラウジング中の平均スワイプ速度及びユーザの注視時間を取得することで、閲覧中の記事への興味度合いを推定した。結果として、これらの特徴量と興味度合いを尋ねるアンケート結果に強い相関が確認された。これより、Janら[5]の結果から、A/D形式よりもIS形式の方がアンケートに対する興味度合いが高くなる可能性がある。

平部ら[7]は、スマートフォンに内蔵されている種々のセンサに加えて、タッチ操作をセンサとして活用するシステムをAndroid端末において構築した。8種類のタッチ操作を識別する実験において、識別精度は85%であったと報告している。また、例えば片手で持つ場合や両手で持つ場合等の8種類のスマートフォンの持ち方が96%の精度で推定できたことを報告している。

Marioら[8]は、30種類の操作特徴セットを用いて、上下左右の基本的なスワイプによって個人の操作特徴を分類及び識別することで新しいバイオメトリック認証の可能性を検討した。実験では、個人の操作を登録してから1週間後の認証テストにおいて平均誤り率が4%未満であったと報告している。平部らとMarioらの技術を用いれば、アンケートの回答者ごとに丁寧に回答した時の操作を記録し、それ以降の回答時の操作と比較することでsatisficingを検出することができる可能性が考えられる。

3. satisficingを検出するための指標の検討

satisficingを検出するための指標として考えられる事項を表1に示す。ここで、5章で説明する評価実験に用いるアンケートシステムであるLimeSurvey[9]について、表中では「LS」と略記する。また、「ネイティブアプリ」とは、ながら操作を検知するためにデバイス内蔵の各種センサを用いるためのアプリケーションのことを指している。これは現段階では実装しておらず、実環境下の実験においてLimeSurveyのシステムで取得する指標のみでsatisficingの検出が困難であると判断した際に検討する予定である。

アンケートの回答時間は設問文および選択肢の文章量と難易度および回答者の傾注度合いに依存すると考えられる。ここで、文章量と難易度は回答者に依存しない要素であると仮定すると、「設問単位の回答時間」が設問固有のsatisficingと関連していると考えられる。すなわち、モデル化した値よりも速い回答について、回答者の傾注度合い

表1 satisficing が検出に寄与する可能性が考えられる指標の例

指標カテゴリ	指標	取得システム
回答方法	アンケート全体の回答時間	LS 標準機能
	設問単位の回答時間	LS プラグイン
	回答の変更	LS プラグイン
	スクロール速度	LS プラグイン
	スクロール長	LS プラグイン
回答環境	配信から回答までの時間	LS 標準機能
	回答時間帯	LS 標準機能
	ながら操作	ネイティブアプリ

が低いとみなすことができる。ゆえにその回答は、適切な認知コストが払われず、satisficing が発現したと解釈できる可能性が考えられる。

「回答の変更」は、少なくとも意図して雑な回答をしたようにした場合には発生しないと考えられる。したがって、少なくとも1回以上の変更が見られた回答はsatisficing が発現していないとみなすことができると考える。ここで、「回答の変更」とは、選択形式においては「選択肢の変更」、自由記述形式の設問においては「記述の変更」とする。ただし、変更されなかった回答に払われた認知コストが必ずしも低いとは言えないため、この指標のみで的確にsatisficing の検出が実現できるわけではないことに注意されたい。satisficing がある回答結果も含めて、選択肢を変更しなかった回答を除外すれば高純度な結果を得ることはできるが、データ数の大幅な減少が予想されるため、選択肢が変更される割合についても分析する。

本稿では、「スクロール速度」は一回のスクロール操作によって移動した画面移動の速度と定義する。また、「スクロール長」は一回のスクロール操作によって移動した画面の移動量と定義する。これらは設問間の移動という挙動のパラメータとして捉えることができる。satisficing が発現している場合は、早く終わらせたい思いから設問間の移動が粗くかつ速くなると考えられる。

「配信から回答までの時間」は回答者のアンケートに対する積極性を反映する可能性が考えられる。この指標には、「回答なし」という場合も含む。これに関してはもはやアンケートの回答を達成していない点でsatisficing ではないが、該当するアンケートに対するモチベーションの低さが伺える。

「回答時間帯」については、特に回答時間帯を指定するアンケートにおいて重要な指標となる可能性が考えられる。例えば、平日の朝の時間帯であれば、仕事に行く準備をしていてアンケートに十分な時間が割けず、結果的にsatisfice が発現し易い傾向が見られる可能性がある。一方で、仕事が終わって、予定に追われることなく落ち着いた状態でアンケートに取り組むことができる夜には、satisfice が発現しにくい傾向がみられる可能性が考えられる。

「ながら操作」については、回答中に取り得る回答者の

様々な状態がsatisfice に影響する可能性が考えられる。例えば、食事中に回答する場合は食べ物とアンケートの両方に意識を向けることになるため、アンケートを答える行為のみを行う場合よりもsatisfice は発現しやすい可能性があると考えられる。このように、オンラインアンケートにおいては何か他のことをしながら回答するケースは大いにあり得るため、ながら操作を考慮しなければならないと考える。そこで、平部ら [7] と Mario ら [8] の結果を応用すれば、スマートフォンの持ち方とタッチスクリーン操作から個人ごとにながら操作の有無を推定できる可能性が考えられる。

ただし、本稿ではシステムの評価実験を目的として理想環境において実施したため、回答環境に関する指標は考慮しないこととする。

4. 提案するアンケートシステム

4.1 システムのアプローチ検討

3章において議論した指標を取得するシステムを実現するにあたり、次に示す手法が考えられる。

- OS 側でのタッチジェスチャー取得
- カメラによる動画撮影
- アンケートアプリケーションの自作
- 既存アンケートシステムの活用

これらの利点と欠点について検討し、本稿で提案するシステムの採用理由を述べる。

OS 側でタッチジェスチャーを取得する例として、Touch Analyzer [7] がある。これは Android 端末の OS が出力するイベントデバイスファイルを逆解析することで、使用中のアプリケーションに依存せずタッチジェスチャーを取得する。この手法であれば特別なアプリケーションのインストールが不要であるが、スマートフォンを PC と USB で接続する必要があり、広範なアンケートの実施においてスケーラビリティの問題が発生する。また、Android 限定であり、iOS では使用できない点もスケーラビリティを損なう要素であり、本研究においては不適な点として挙げられる。

カメラで回答操作を録画して、ムービーの解析によって回答操作を抽出するアプローチも考えられる。この手法もアプリケーションに非依存であるが、一方でアンケートシステムとは別にカメラを用意する環境が必要となり、Touch Analyzer 同様、スケーラビリティの観点で問題がある。

専用のアンケートアプリケーションを全て自分で作るという手法も考えられる。自作するため設計および実装の自由度が高く、理論上どの OS についても所望の挙動を実現することができる。また、回答者がアプリケーションをインストールするという以外の特異なオペレーションは不要であり、スケーラビリティの問題もクリアできる。一方で、

アンケート実施者は広く使われている既存のアンケートシステムではなく、操作に慣れないアンケートアプリケーションを導入する必要がある。この点は、実施者側に起因するスケーラビリティに影響を与える要因となり得る可能性がある。また、筆者についてもアンケートシステムをゼロから自作するのは骨が折れる作業であり、本来フォーカスすべき目的ではなく手段に時間と労力を割くことになりかねないため、この手法は得策とは言い難い。

既存の web アンケートシステムとして広く使われているものに、Google Form, SurveyMonkey, LimeSurvey がある。これらを用いて機能要件を実現できれば、ゼロからアプリケーションを作る必要はない。このうち Google Form と SurveyMonkey はサービスとして運用されており、利用者が自作のプログラムを組み込む余地がなく、要件を満たすシステムが実現できない。一方で、LimeSurvey はオープンソースであり、プラグインとして独自に開発した機能を組み込む仕組みがある。本研究で必要となる Javascript でタッチジェスチャーや回答内容を取得する技術を導入することもできる。

これら各アプローチの利点と欠点を考慮し、要件の実現性、システムの拡張性や拡散性の観点から、本研究では LimeSurvey をベースとしたシステムを提案することとした。

4.2 提案システムの構成

システムの要件としてアンケートの実施と指標のセンシングの2点が挙げられる。本研究では、プラグインとして自作プログラムを組み込むことができるオープンソースのアンケートシステムである LimeSurvey を採用し、図2に示すシステムを構築した。LimeSurvey の標準機能では回答時のタッチジェスチャーを記録することができないため、その機能をプラグインとして開発した。

プラグインは2つの Javascript (以下 js) ファイルと1つの PHP ファイルから成る。1つ目の js スクリプトで被験者のトークンを取得し、2つ目の js スクリプトで回答操作を記録する。2つ目の js で取得した回答操作ログデータを PHP 側に Ajax 通信で POST し、PHP 側からデータベースに格納する。データベースのテーブルはメインテーブルとテキストテーブルの2つがある。メインテーブルにはタッチジェスチャーおよび選択形式設問の回答記録が格納され、テキストテーブルには自由記述形式設問の回答記録が格納される。

本プラグインの導入方法は、これら3つのファイルを LimeSurvey サーバのルートディレクトリに配置し、質問設定画面のヘルプ文章設定欄にてファイルパスを js で読み込む設定をするのみである。トークンを使用する場合は標準機能として設定されるトークン入力画面とは別に、質問グループを1つ作成し、再度トークンを入力させる質問グ

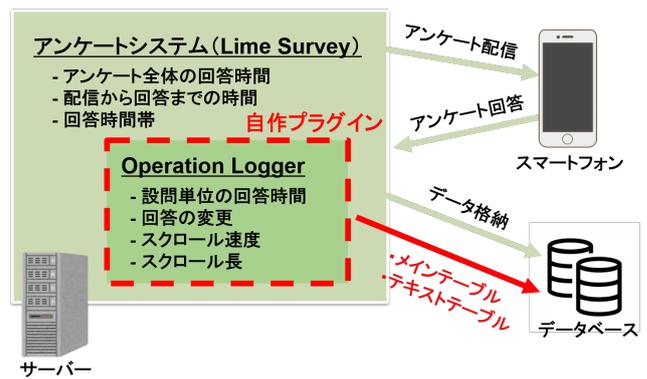


図2 システムおよび取得するデータ

ループを設定する (LimeSurvey は質問グループの下に質問を作成する構造となっており、質問グループがなければ質問が作成できない)。その質問のヘルプ欄にてトークン取得用の js スクリプトを呼び出す。それ以降の質問グループでは、質問グループ内の最後の質問のヘルプ欄にて回答操作取得用 js ファイルを呼び出す。なお、回答者側はソフトウェアの追加や設定の変更をする必要がない。

4.3 システムの詳細

開発したシステムの詳細について述べる。データを取得するイベントはタッチイベント、選択肢のタップ、自由記述の入力の3種類である。

タッチイベントの種類は、touchstart, touchmove, touchend の3種類であり、それぞれ操作開始、スワイプ、操作終了のタイミングで発火する。これらのタイミングにおける時刻、画面内の座標、ページ上端からの移動量、タッチイベントの種類を取得する。これにより、「全体の回答時間」、「スクロール速度」、「スクロール長」を検出することができる。

選択肢をタップしたタイミングでは、時刻、選択肢の id、設問の id を取得する。これにより、「選択形式の設問単位の回答時間」と「選択肢の変更」を検出することができる。

自由記述を入力したタイミングでは、時刻、文字列、設問の id を取得する。これにより、「記述の変更」や「自由記述形式の設問単位の回答時間」が検出できる。入力内容を記録する仕様として、1文字入力される度に取得する方法と、ある程度まとまったテキスト単位で取得する方法について検討した。前者の方法は、膨大な量のテキストデータを取得することになり、多大な記憶容量を必要としてしまう。そのため、本システムでは後者の仕様を採用した。さらに、記録するテキスト単位の程度を検討した結果、「確定キーの押下」および「変換候補の選択」の挙動によって入力されるテキストを記録単位とした。

なお、本システムは表2の環境で所望の動作を確認した。

表2 提案システムの動作確認済みモバイル端末環境

OS	機種名	ブラウザ
Android 9.0	Google Pixel 3	Google Chrome
		Firefox
		Opera
Android 7.12	Xiaomi Redmi 4X	Google Chrome
iOS 13.2.2	iPhone8	Safari

表3 評価実験におけるアンケート内容

ページ番号	設問ID	設問内容	回答形式
1	Q1	性別を選択してください。	ラジオボタン (2 択)
	Q2	「My name is Taro.」を日本語訳してください。	自由記述
	Q3	現在の職業を選択してください。	ドロップダウン (15 択)
	Q4	夏か冬、どちらが好きですか？ あなたがやりたい仕事を小学生が読んで分かるように簡潔に説明してください。	ラジオボタン (2 択)
2	Q5	あなたが小学生の時に将来なりたかった職業はなんですか？	自由記述
	Q6	穏やかである	ラジオボタン (7 択)
	Q7	失敗を引きずることがある	ラジオボタン (7 択)
3	Q8	優しい	ラジオボタン (7 択)
	Q9	朝型である	ラジオボタン (7 択)
	Q10	はつらつとした	ラジオボタン (7 択)
4	Q11	暗い	ラジオボタン (7 択)
	Q12	気がかりな	ラジオボタン (7 択)
	Q13	嬉しい	ラジオボタン (7 択)
	Q14	嫌な	ラジオボタン (7 択)
	Q15	不安な	ラジオボタン (7 択)
	Q16	楽しい	ラジオボタン (7 択)
	Q17	沈んだ	ラジオボタン (7 択)
	Q18	心配な	ラジオボタン (7 択)
	Q19	先ほどの問題で、「失敗を引きずることがある」という質問に対して選んだ選択肢を回答してください。	ラジオボタン (7 択)

eventName	xAxis	yAxis	dateTime	msTime	selectedAnswer	selectedQuestion	scrollY
hci_android_1_2	0	0	20191205163323	1575531203252	start	start	0
touchstart	258	461	20191205163326	1575531206463	0	0	0
touchmove	258	452	20191205163326	1575531206525	0	0	0
touchmove	259	449	20191205163326	1575531206553	0	0	0
touchmove	269	397	20191205163326	1575531206739	0	0	0
touchmove	277	397	20191205163326	1575531206939	0	0	21
touchmove	280	397	20191205163327	1575531207146	0	0	30
touchend	280	397	20191205163327	1575531207151	0	0	30
touchend	52	392	20191205163327	1575531207837	0	0	33
touchend	52	392	20191205163327	1575531207881	A2	574166X6X17	33
touchstart	254	536	20191205163328	1575531208444	0	0	33
touchmove	254	525	20191205163328	1575531208522	0	0	33
eventName	xAxis	yAxis	dateTime	msTime	selectedAnswer	selectedQuestion	sc
touchmove	256	520	20191205163328	1575531208545	0	0	33
touchmove	304	525	20191205163328	1575531208745	0	0	151
touchmove	308	526	20191205163328	1575531208771	0	0	157
touchend	308	526	20191205163328	1575531208772	0	0	157
touchstart	262	634	20191205163329	1575531209788	0	0	173
touchmove	262	624	20191205163329	1575531209886	0	0	173
touchmove	263	626	20191205163329	1575531209909	0	0	181
touchmove	293	619	20191205163330	1575531210075	0	0	247
touchend	619	619	20191205163330	1575531210075	0	0	247
touchend	691	691	20191205163337	1575531217678	0	0	287
touchend	239	691	20191205163337	1575531217735	focus1	574166X6X18	287
touchstart	50	775	20191205163411	1575531251278	0	0	570

図3 取得するデータの例 (メインテーブル)

id	dateTime	msTime	selectedAnswer	selectedQuestion
hci_android	20191210191812	1575972092774	start	start
focus1	20191210191824	1575973104635	私の	574166X5X15
focus1	20191210191825	1575973105279	私の名前は	574166X5X15
focus1	20191210191830	1575973110447	私の名前は太郎です	574166X5X15
hci_android	20191210191841	1575973121511	start	start
focus1	20191210191903	1575973143845	みんなが持っている	574166X6X18
focus1	20191210191912	1575973152626	みんなが持っているAmazonなどの	574166X6X18
focus1	20191210191917	1575973157990	みんなが持っているAmazonなどのネットショッピングの	574166X6X18
focus1	20191210191919	1575973159288	みんなが持っているAmazonなどのネットショッピングのデータを使って	574166X6X18
focus1	20191210191924	1575973164176	みんなが持っているAmazonなどのネットショッピングのデータを使って	574166X6X18
focus1	20191210191929	1575973169210	みんなが持っているAmazonなどのネットショッピングのデータを使って	574166X6X18
focus1	20191210191929	1575973169744	みんなが持っているAmazonなどのネットショッピングのデータを使って	574166X6X18
focus1	20191210191930	1575973170971	みんなが持っているAmazonなどのネットショッピングのデータを使って	574166X6X18
focus1	20191210191939	1575973179746	みんなが持っているAmazonなどのネットショッピングのデータを使って	574166X6X18
focus1	20191210191942	1575973182955	みんなが持っているAmazonなどのネットショッピングのデータを使って	574166X6X18
focus1	20191210191944	1575973184188	みんなが持っているAmazonなどのネットショッピングのデータを使って	574166X6X18
focus1	20191210191945	1575973185224	みんなが持っているAmazonなどのネットショッピングのデータを使って	574166X6X18
focus1	20191210191950	1575973190980	みんなが持っているAmazonなどのネットショッピングのデータを使って	574166X6X18
focus1	20191210191954	1575973194139	みんなが持っているAmazonなどのネットショッピングのデータを使って	574166X6X18

図4 取得するデータの例 (テキストテーブル)

5. 評価実験

4章で述べたシステムがsatisficingの検出に有効であるかを検証する評価実験を行った。

5.1 アンケート設計

ページ番号, 設問内容, 回答形式を表3に示す。回答形式の種類は自由記述, ラジオボタン, ドロップダウンがある。自由記述の文字数も一つの指標になり得ると考えたため, 文字数指定はなしとした。なお, ページ番号3と4の設問においては, ページ全体に対して次のような設問文を提示した。「以下の4項目の性質について, 自分がどの程度あてはまるかを『0. 全くあてはまらない』から『6. 非常によくあてはまる』までの7段階の中からよくあてはまる数字を1つだけ選んで下さい。」

ページ番号1と2では, それぞれ回答が回答者によらず普遍的なものと回答者によって変わるものを設定した。ここでの狙いは, 提案システムで取得できる指標が普遍的設問か否かで異なる傾向を示す可能性を模索することである。ページ番号3では, 選択肢が二値でなく, 回答が一意に定

まらない設問を設定した。ここでは, 回答に迷う人と迷わない人に対する指標の振る舞いを観察する。ページ番号4では, DAMS (Depression and Anxiety Mood Scale) [10] という抑うつ気分, 不安気分, 肯定的気分を測定する質問票に対する指標の振る舞いを観察する。これは, 様々な統計分析の結果, 実用性があると認められた質問票に対して本システムが有効であるか否かを検討することが狙いである。

5.2 実験方法

本実験に使用する端末は表2の1行目に記す, Google Pixel 3 (Android 9.0) とし, ブラウザはGoogle Chromeとした。被験者は男性5名 (右利き, 22-23歳) である。各被験者に普段の日本語入力で用いているキーボードを聴取し, それに合わせてQWERTYキーとテンキーのいずれかを設定した。被験者の操作状況を後で確認できるように別のスマートフォンで録画した。アンケート開始画面にて各被験者はトークンを入力することで回答を開始する。

6. 結果と考察

表1の指標のうち, 回答方法の指標カテゴリに属する指標5つに関して, 評価実験において取得した結果を本節で記す。

回答時間

アンケート全体の回答時間を図5, 設問ごとの回答時間を図6に示す。全体の回答時間を見ると, おおよそ2分30

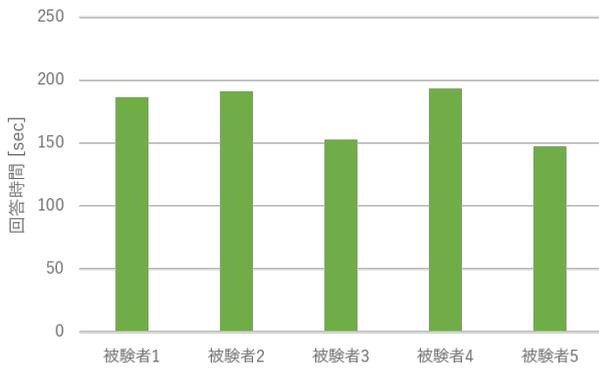


図5 アンケート全体の回答時間

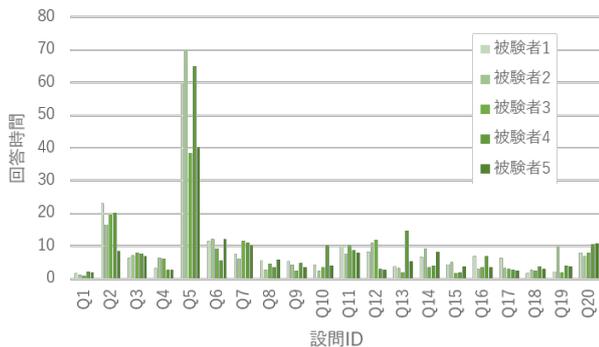


図6 設問単位の回答時間

秒から3分の間に収まっていることが分かる。被験者3と5について、他の意見者より30秒ほど早く終えているのは、Q5の回答時間の影響が大きいことが6から見て取れる。Q5は3に示している設問であり、「小学生が読んで分かるように簡潔に」回答するという制約から、このような個人差が生まれたのだと推測される。また、5章で述べた設問設定の狙い通り、ページ番号1と2の設問を回答形式を対応付けて比較すると、ページ番号2の回答の方が回答時間が長い傾向が見られる。

スクロール長およびスクロール速度

被験者ごとの平均スクロール長および平均スクロール速度を図7に示す。被験者1, 2は大きなストロークでゆっくりと操作していることが分かる。一方、被験者5は大きなストロークでありながら、素早い操作をしていることが見て取れる。

ページ番号4の設問は、ページの上端にそのページの設問全体に関する説明文が記載されている状況であった。その設問に関して、数問回答してからその説明文を読み直す挙動が被験者全員において観測された。回答にあたって設問の条件を確認するためであったとヒアリングにより理由が判明したが、このような挙動も *satisficing* に寄与する指標として考えられる可能性を見出した。

回答の変更

選択肢の変更回数を図4に、記述の変更回数を図5に示

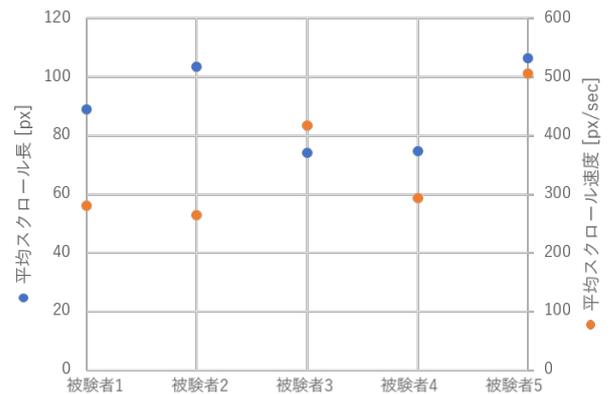


図7 平均スクロール長と平均スクロール速度

表4 選択肢の変更回数

被験者 ID	選択肢変更回数	誤選択回数
1	0	0
2	3	1
3	3	1
4	3	1
5	2	2

表5 記述の変更回数

被験者 ID	記述変更回数
1	1
2	2
3	0
4	4
5	1

す。本実験における選択式、自由記述式の設問数はそれぞれ18問、2問である。図2の「選択肢変更回数」とは、システムで取得した回数であり、そのうち誤操作であったものを「誤操作回数」に示している。

選択形式の設問においては、誤操作による選択に対して誤りであることをシステムが認識できないという問題点が挙げられた。*satisficing* が高ければ高いほど誤操作は増加すると予想されるため、この問題は重大であると捉える。Firefoxにおいて、同様の実験を実施した際に、選択肢のタップに対する当たり判定が他のブラウザよりも厳しく、偽陽性の誤操作が大幅に減少できる可能性を感じた。これより、対応策の一つとして当たり判定の厳しさをコントロールすることが考えられる。

7. まとめと今後の展望

本稿では、オンラインアンケートの回答信頼性を毀損する可能性がある *satisficing* の検出に向けた操作ログ取得システムを提案した。提案システムは拡散性や導入コストを考慮した上で、オープンソースのアンケートシステムである LimeSurvey をベースとし、そのプラグインとして回答操作を取得するものとした。現状では、LimeSurvey の回答

形式のうちラジオボタン, ドロップダウン, 自由記述に対応しており, 一般的なアンケート内容に対して広く適用できると考えられる. 評価実験においては, システムの要件として挙げた項目の取得が行えていることを確認した. 選択形式の設問において, 誤操作で選択肢を選んてしまった場合に, それが誤りであることをシステムが認識できない等の問題点が明らかになった. 今後はそのような問題点に対する解決を図りながら, 本システムを用いた実験に向けて *satisficing* を検出するために適切なアンケートの設計に取り組む.

今後は方向で提案したシステム上で *satisficing* の有無を分析するためのアンケートを設計し, 実験を行う予定である. アンケート設計に際して, 基礎的なアンケート設計法 [11] から, 各種ヒューマンインタラクションに関する学会が示している再現性に対する考え方 [12] を参考にする.

また, 本稿では理想的な環境を想定して実験を行ったが, 実用の環境では *satisficing* を検出することに寄与する可能性がある指標が存在する. 今回考慮していないものとして, ながら操作の有無が考えられる. これらの発生を阻害しない, より実用の環境における実験を計画する. その際, 平部ら [7] の技術により, スマートフォンの持ち方とタッチスクリーン操作から個人ごとにながら操作の有無を推定できる可能性が考えられる.

さらに, 本研究では, アンケートの回答結果の信頼性が損なわれる原因として *satisficing* を取り上げているが, 他にも *satisficing* に影響を与える要因として反応バイアスというものがある. これは, 正確に回答する意思の有無に関わらず, ヒトの心理的性質に起因する統計誤差である. 本研究で取得する予定の操作ログを用いて反応バイアスの検出を実現し, さらなるデータの信頼性向上に貢献することが期待される.

謝辞

本研究の一部は, JST さきがけの支援のもと実施されている.

参考文献

- [1] Herbert A. Simon. Rational choice and the structure of the environment. 1956.
- [2] Jon A. Krosnick. Response strategies for coping with the cognitive demands of attitude measure in surveys. pp. 213–236, 1991.
- [3] 三浦麻子, 小林哲郎. オンライン調査における努力の最小限化が回答行動に及ぼす影響. 行動計量学, Vol. 45, No. 1, pp. 1–11, 2018.
- [4] Roger Tourangeau, Hanyu Sun, Ting Yan, Aaron Maitland, Gonzalo Rivero, and Douglas Williams. Web surveys by smartphones and tablets: Effects on data quality. *Social Science Computer Review*, Vol. 36, No. 5, pp. 542–556, 2018.
- [5] Jan Karem Höhne and Timo Lenzner. New Insights on the Cognitive Processing of Agree/Disagree and Item-Specific

Questions. *Journal of Survey Statistics and Methodology*, Vol. 6, No. 3, pp. 401–417, 10 2017.

- [6] 加藤勇太, 岩本健嗣, 松本三千人. タッチ操作ログを用いた web コンテンツ閲覧時における興味度合い推定の研究. 情報処理学会論文誌, Vol. 59, No. 2, pp. 508–518, 2018.
- [7] Yuko Hirabe, Hirohiko Suwa, Yutaka Arakawa, and Keiichi Yasumoto. Touchanalyzer: A system for analyzing user's touch behavior on a smartphone. *International Journal of Computer Science and Mobile Computing*, Vol. 7, pp. 25–38, 2018.
- [8] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE Transactions on Information Forensics and Security*, Vol. 8, No. 1, pp. 136–148, Jan 2013.
- [9] Limesurvey: the online survey tool - open source surveys. <https://www.limesurvey.org/>. (Accessed on 07/12/2019).
- [10] 福井至. Depression and anxiety mood scale(dams) 開発の試み. 行動療法研究, Vol. 23, No. 2, pp. 83–93, 1997.
- [11] Bernard C. K. Choi and Anita W. P. Pak. A catalog of biases in questionnaires. *Preventing Chronic Disease*, Vol. 2, pp. 51–63, 2005.
- [12] 水口充. エンタテインメントコンピューティング研究における価値基準の枠組みの提案. エンタテインメントコンピューティングシンポジウム 2018 論文集, 第 2018 巻, pp. 57–64, sep 2018.