

Webベースの実験における 事前タスクを用いたユーザ分類の検討

三山 貴也¹ 中村 聡史¹ 山中 祥太²

概要: Web ベースの実験は、クラウドソーシングを利用することで多くの参加者を素早く募集でき、ユーザインタフェース操作時のデータを大量に収集することができる。しかし、クラウドソーシングを利用して行う Web ベースの実験では、指示を守らない参加者や乱雑な操作を行う参加者も存在するため、実験データの品質が低下してしまう問題がある。そこで本研究では、適切な操作を行うユーザのみが実験に参加できるようにするため、実験で要求される操作と関連する事前タスクを設け、その結果に基づいて適切な操作を行うユーザを抽出するアプローチを検討した。実験では、事前タスクとしてマウス操作を必要とする簡単な操作課題を行い、その後にポインティングタスクを実施した。その結果、事前タスクとポインティングタスクのパフォーマンスには一定の関連性がみられ、事前タスクによってより正確に指示を守るユーザを抽出できる可能性が示唆された。

1. はじめに

Web ベースの実験はインターネットを通じて行われ、ブラウザ上で動作するシステムを使用することで、参加者は場所や時間を問わずに実験に参加できる。そのため、クラウドソーシングを利用して多くの参加者を素早く募集でき、実験を迅速に実行しながらも大規模で多様なサンプルを入手し、多くの参加者から得た実験データによって有意差を検出しやすいという利点がある [1]。一方で、Web ベースの実験では実験実施者が参加者の状況を確認できないため、指示を守らない参加者や不注意な参加者が存在することが知られており、それによって実験データの品質が低下してしまう問題がある [2]。そのため、Web ベースの実験におけるデータの品質管理は重要な課題である。

グラフィカルユーザインタフェース (GUI) に関するユーザ実験を含む研究においても、クラウドソーシングを利用して参加者を募集することが一般的になっている。ここでは、GUI 操作時のデータを大量に収集できることが利点であり、特にポインティングタスクにおけるエラー率の検証のような発生確率の低い事象を扱う場合に恩恵が大きいことが報告されている [3]。一方で、クラウドソーシング実験の参加者の特徴として、実験室実験の参加者に比べてポインティングタスクの操作が不正確である傾向があり、2 倍以

上のエラー率が観察されたという報告も存在する [4]。このように、Web ベースの GUI 実験においても、クラウドソーシングを通じて募集された参加者の性質が実験結果に影響を及ぼし、実験データの品質を低下させる懸念がある。

以上の点を踏まえて本研究では、Web ベースの GUI 実験における実験データの品質向上を目的として、適切な操作を行うユーザのみが実験に参加できるようにするアプローチを提案する。ここでは、実験で要求される操作と関連する事前タスクを設け、その結果に基づいて適切な操作を行うユーザを抽出する。本来目的とする実験をそのユーザ群のみに依頼することで、データ品質を向上させることが狙いである。

本稿では、提案アプローチの初期検討として、事前タスクとしてマウス操作を必要とする簡単な操作課題、実験の主タスクとしてポインティングタスクを実施し、事前タスクとポインティングタスクの操作パフォーマンスの関連性を明らかにする。これにより、事前タスクの操作パフォーマンスからポインティングタスクの操作パフォーマンスを予測し、事前タスクのみの実施によって適切な操作を行うユーザを抽出できるかを検証する。

2. 関連研究

2.1 Web ベースの実験のデータ品質に関する研究

Web ベースの実験におけるデータ品質の懸念に対し、クラウドソーシング実験と実験室実験の結果を比較する研究が行われており、心理学や社会学の実験において多くの場

¹ 明治大学
Meiji University

² LINE ヤフー株式会社
LY Corporation

合で同様の結果が得られているが、一部では異なる結果が報告されている [5][6]。また、実験環境を実験室に近づける研究も行われており、Li ら [7] は、クレジットカードなどの物理カードと画面上のカード画像の大きさを一致させるタスクによってディスプレイの画素密度を計算し、視覚刺激の大きさを統制する手法を提案している。

不注意な参加者に関する研究も行われており、クラウドソーシング実験の参加者のうち 45.9%が何らかの不注意な行動をしたという報告 [8] がある。これに対し、不注意な回答の検出には反応時間や自由記述の分析が効果的だとされている [9]。また、Oppenheimer ら [2] は、参加者が指示を読んでいるかを測定する質問 (Instructional Manipulation Check, IMC) を提案しており、IMC を含む Web アンケートでは 83.8%の回答者が指示を守らずに回答したという報告も存在する [10]。このように、Web ベースの実験ではデータ品質に懸念が残るが、Web ベースの実験を実施した心理学研究の 55%ではデータ品質評価を実施していないという報告 [11] も存在する。

以上のように、Web ベースの実験のデータ品質に関する様々な検証や対策が行われている。本研究は、実験で要求される操作と関連する事前タスクによって適切なユーザを抽出することで、これまでの取り組みをさらに改善する可能性を探るものである。

2.2 Web ベースの GUI 実験に関する研究

Web ベースの GUI 実験においても、クラウドソーシング実験と実験室実験の結果を比較する研究が行われている。Komarov ら [12] は Bubble Cursor[13] の検証を行い、従来のカーソルと比較して操作時間を短縮させるという同様の結果を得たことを報告している。Schwab ら [14] はパン操作とズーム操作のタスクを実施し、PC とモバイルの両環境で従来通りフィッツの法則が成り立つことを報告している。また Findlater ら [4] は、マウス操作とタッチ操作のポインティングタスクにおいて、クラウドソーシング実験の参加者は実験室実験の参加者よりも操作時間が短くエラー率が高いことから、クラウドソーシング実験の参加者は「できるだけ速く正確に」操作するように指示された場合、正確さよりも速さを重視する傾向があるとしている。

以上のように、Web ベースの GUI 実験においてもデータ品質に関する様々な検証が行われている。本研究では、GUI 実験特有の操作やインタラクションが要求されることに着目し、それに関連する事前タスクによって適切なユーザを抽出することで、さらなるデータ品質の向上を目指す。

2.3 ポインティングタスクに関する研究

フィッツの法則は、ターゲットまでの距離 A とターゲットの幅 W から最初のクリックまでの時間 MT を予測できるというモデルである [15][16]。

$$MT = a + bID, ID = \log_2 \left(\frac{A}{W} + 1 \right) \quad (1)$$

フィッツの法則に関しては、ターゲットの範囲外をクリックするエラー率 ER を予測する研究も行われており [17][18]、操作が速くなると ER が増加し、操作が慎重になると ER が減少することが知られている。

フィッツの法則に関する実験では、できるだけ速く正確にタスクを行うように指示することが一般的だが、参加者によって速さと正確さのバランスにバイアスがあることも知られている [19][20]。また、このようなバイアスがかかった状態ごとにタスクを行うことで、速さ重視の場合に MT が減少する一方で ER が増加するといったように、状況に応じたポインティングを評価できるという報告もある [21]。

3. 提案アプローチ

3.1 コンセプト

本研究では、Web ベースの GUI 実験における実験データの品質向上を目的として、適切な操作を行うユーザのみが実験に参加できるようにするアプローチを提案する。具体的には、実験の主タスクで要求される操作と関連する事前タスクを設け、その操作パフォーマンスに基づいて適切な操作を行うユーザを抽出することにより、事前タスクの実施によって適切なユーザ群に実験を依頼することを可能とする。これにより、実験参加者の多くが指示を正確に守るユーザとなり、実験データの品質向上が期待できると考える。提案アプローチのイメージ図を図 1 に示す。

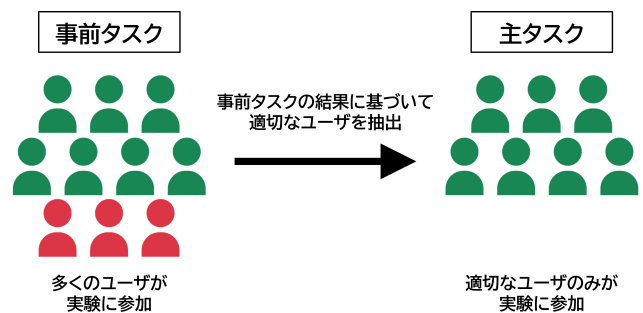


図 1 提案アプローチのイメージ図

3.2 ポインティングタスクに適した事前タスク

本稿では、実験の主タスクとしてポインティングタスクを実施するため、ポインティングと関連のある操作が要求されるタスクを事前タスクとすることで、事前タスクの操作パフォーマンスからポインティングタスクの操作パフォーマンスを予測できると考えられる。ここで我々は、Li ら [7] が提案した物理カードと画面上のカード画像の大きさを一致させるタスク (サイズ調整タスク) に着目した。ここでは図 2 のように、実験参加者が画面上にサイズが標準化された物理カード (クレジットカードなど) を設置し、その大きさと一致するようにカード画像の大きさを調整す

る。そのため、ポインティングタスクと同様にマウス操作が要求され、サイズ調整の所要時間で操作の速さを分析でき、サイズ調整の結果が物理カードの大きさと一致しているかどうかで操作の正確性を分析できる。このような理由から、できるだけ速く正確に操作することが求められるポインティングタスクを主タスクとする場合、サイズ調整タスクが事前タスクとして適していると考えた。

またサイズ調整タスクは、本来は視覚刺激の大きさを統制するために実験前に実施され、我々もこれまでの研究 [22][23] において実験前にサイズ調整タスクを実施してきた。その結果、指示を守らない参加者や乱雑な操作を行う参加者が一定数存在したことから、サイズ調整タスクを事前タスクとすることでそのような参加者を特定し、適切なユーザのみを抽出できると考えた。以上の理由により、本稿ではサイズ調整タスクを事前タスクとして採用する。



図 2 Li ら [7] によるサイズ調整タスクのイメージ図

4. 実験

4.1 実験概要

サイズ調整タスクとポインティングタスクの操作パフォーマンスに関連があるかを調査するため、はじめにサイズ調整タスクを行い、その後にポインティングタスクを行う実験を実施する。ここでは、サイズ調整タスクの操作パフォーマンスからポインティングタスクの操作パフォーマンスを予測し、サイズ調整タスクのみの実施によって適切な操作を行うユーザを抽出できるかを検証するために、サイズ調整タスクとポインティングタスクの両方をすべての参加者に行ってもらおう。本実験では、下記のように仮説を立てた。

- サイズ調整タスクで正確な操作を行うユーザはポインティングタスクのエラー率が低く、不正確な操作を行うユーザはポインティングタスクのエラー率が高い。
- サイズ調整タスクを正確に行うユーザは実験の指示を正確に守るユーザであり、ポインティングタスクにおいて教示通りの操作を行う。

本実験は、我々が作成した Web システムにアクセスすることで参加できるものであり、クラウドソーシングを通じて参加者を募集することで大規模な実験を行う。実験は

図 3 に示す流れで行われ、まず実験システムにアクセスすると実験の説明を行い、その後にサイズ調整タスクを 2 回、ポインティングタスクの練習を 1 セット、ポインティングタスクの本番を 4 セット行う。ここでポインティングタスクの練習とは、ポインティング操作に慣れてもらうためのタスクであり、できるだけ速く正確に操作を行うように教示したうえでタスクを行ってもらおう。これにより、本番で「できるだけ速く」または「できるだけ正確に」と教示された場合に速さと正確さのバランスを意識して操作できるようにすることを目指している。実験は Yahoo!クラウドソーシング*1を通じて参加者を募集し、実験を完了した参加者には 200 円の報酬を支払った。

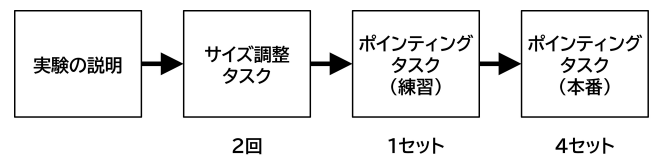


図 3 実験の流れ

4.2 サイズ調整タスク

本実験におけるサイズ調整タスクは、3.2 節で述べたように、実験参加者が画面上に物理カードを設置してその大きさと一致するようにカード画像の大きさを調整するタスクである。ここで、クラウドソーシング実験では参加者ごとに使用しているディスプレイの画素密度が異なり、1 回のサイズ調整タスクでは参加者が正確に調整を行ったかを判断することができない。そこで本実験では、サイズ調整タスクを 2 回実施し、2 回の調整結果の誤差を分析することによって各ユーザの操作の正確性を評価する。また、カード画像の初期サイズが物理カードの大きさと近い値であったため参加者がサイズ調整を行う必要がないという状況を避けるため、カード画像の初期サイズは極端に小さいまたは大きい値が望ましい。そのため本実験では、カード画像の初期サイズ条件として 100 px と 900 px を用意し、2 つの条件が 2 回のサイズ調整タスクで 1 回ずつ順番はランダムに割り当てられるようにした。ここでは、物理カードとして使用可能な例として、クレジットカード、Suica や PASMO などの交通系 IC カード、健康保険証、運転免許証などを挙げた (ISO/IEC 7810 の ID-1 規格：縦 53.98 mm × 横 85.60 mm)。

本実験におけるサイズ調整タスクは図 4 のような画面で行われ、カード画像右下をドラッグする操作によって大きさを調整できるようになっている。この動作は Li ら [7] の研究をもとに開発された jsPsych[24] のプラグイン*2とほぼ同様である。なお、本来はサイズ調整タスクは視覚刺激の大きさを統制するためのものであるが、本実験ではサイ

*1 <https://crowdsourcing.yahoo.co.jp>

*2 <https://www.jspsych.org/v7/plugins/virtual-chinrest/>

ズ調整タスクの結果によってポインティングタスクのターゲットの大きさを統制することはしない。これは、サイズ調整タスクにおいて不適切な操作をする参加者が存在する想定であり、サイズ調整タスクの結果得られた画素密度に基づいて視覚刺激を統制すると、過度に大きいまたは小さい視覚刺激を表示してしまう可能性があるためである。そのため、本実験ではサイズ調整タスクをあくまでも適切な操作を行うユーザを抽出するための事前タスクとして扱う。

サイズを調整するタスク 1/2



図 4 実験におけるサイズ調整タスク

4.3 ポインティングタスク

本実験におけるポインティングタスクは、教示に従って画面上に表示される円形のターゲットをクリックしてもらうタスクである。ここでは、図 5 のように 1280×720 px の領域内にターゲットが表示され、ターゲットをクリックすると条件に従って次のターゲットが表示されるという試行を指定された回数だけ繰り返す。本実験では、ターゲット間の距離 A は 510 px の 1 条件、ターゲットの直径 W は $8, 38, 78$ px の 3 条件とした。これにより、 ID の値は $6.02, 3.85, 2.91$ bits となり、ミスクリックが発生しやすい条件とミスクリックが発生しにくい条件に加えてその中間の条件を扱う。なおターゲットの出現位置は、ターゲット間の距離 A の条件を満たす以外はランダムである。また、ポインティングタスクの本番では参加者が指示に従って操作を行うかを検証するため、下記の 2 つの教示を用意して指定された教示に従ってタスクを行ってもらおう。なお、ポインティングタスクの練習における教示は下記の 2 つを組み合わせた「できるだけ速く正確に」であり、ポインティングタスクの本番において参加者が教示通りの操作を行うことができない状況を可能な限り防いでいる。

- できるだけ速く：ターゲットが表示されてからできるだけ速くクリックしてください。ただし、ターゲットを狙わずに乱雑な操作を行うことはしないでください。
- できるだけ正確に：ミスクリック（ターゲットの範囲外をクリック）をしないようにしてください。ただし、意味もなく必要以上に時間をかけることはしないでください。

本実験ではポインティング 15 試行を 1 セットとしており、15 試行のターゲットのうち W の 3 条件が 5 試行ずつ順番はランダムで設定されるようにした。また、ポインティングタスクの本番では合計 4 セットを実施し、1 セットごとに 2 つの教示のいずれかをランダムに割り当て、全 4 セットのうち各教示ごとに 2 セットずつが割り当てられるようにした。以上の設定により、ポインティングタスクの本番において、参加者 1 人あたりの総試行回数は $3W \times 5$ 試行 $\times 2$ 教示 $\times 2$ セット = 60 試行であり、1 つの ($W \times$ 教示) 条件について、 5 試行 $\times 2$ 教示 = 10 試行分のデータが記録される。

ターゲットをできるだけ速くクリックしてください (あと15回)



図 5 実験におけるポインティングタスク

5. 結果と考察

Yahoo!クラウドソーシングを通じて 500 人が実験を完了した。このうち、実験説明ページにおいてマウスではなくトラックパッドを使用していると回答した 29 人、実験データに欠損が確認された 16 人を除外し、 455 人 (男性 354 人、女性 101 人) を分析対象とする。

5.1 サイズ調整タスク

図 6 は 2 回のサイズ調整タスクにおける調整結果の分布である。ここでサイズ調整結果とは、実験参加者がサイズ調整を行った後のカード画像の横幅の値であり、本来はこの値をもとに視覚刺激の大きさを統制する。各参加者のサイズ調整結果は 1 回目が平均 359.45 px (標準偏差 144.17 px)、2 回目が平均 359.41 px (標準偏差 121.20 px) であり、 390 人 (86%) は 1 回目と 2 回目のどちらも 200 px 以上 600 px 未満であった。また図 7 は 2 回のサイズ調整タスクにおける調整誤差 (絶対値) の分布である。ここでは、参加者の多くは調整誤差が小さく、 354 人 (78%) は 20 px 未満であるが、調整誤差の大きい参加者も存在し、 73 人 (16%) は 50 px 以上であった。また各参加者の操作時間は、1 回目が平均 8.03 s (標準偏差 6.13 s)、2 回目が平均 6.27 s (標準偏差 5.49 s) であるが、1 回目に 38 人、2 回目に 25 人が操作時間 0 s であり、操作を全く行わない参加者も存在した。

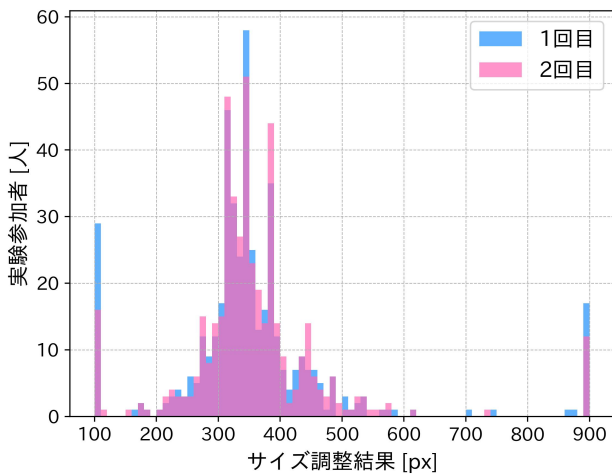


図 6 サイズ調整結果の分布

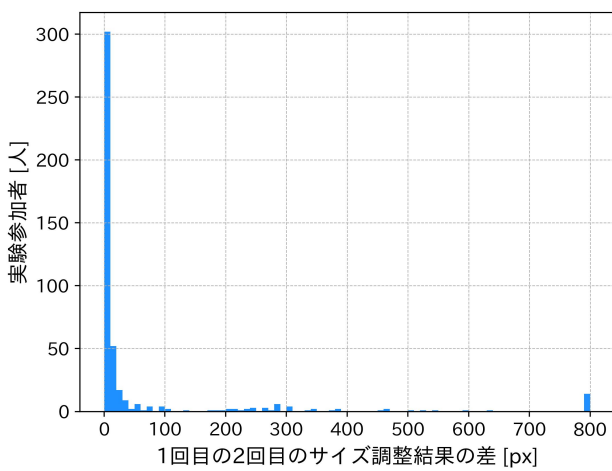


図 7 サイズ調整誤差の分布

5.2 ポインティングタスク

60 試行 × 455 人 = 27,300 試行 について、分析前に外れ値の除外を行った。まず、各試行の最初のクリックまでの移動距離がターゲット間の距離 A の半分未満の場合、1 つ前のターゲットをダブルクリックしたなど明らかな操作ミスとして除外した。次に、教示 × W それぞれで 1 人あたり 10 試行であるが、その 10 試行のなかで MT が外れ値となる試行を除外する。ここでは、四分位範囲 (IQR) 法を用いて、 MT が第 1 四分位数より $3 \times IQR$ 以上短いまたは第 3 四分位数より $3 \times IQR$ 以上長い試行を除外した。さらに、各参加者の全 60 試行の平均 MT とエラー率 ER について同様に $3IQR$ 法を用い、外れ値となる参加者の試行をすべて除外した。以上の処理により、544 試行と 4 人の参加者が外れ値として検出され、除外後の 26,596 試行を分析対象とした。

ポインティングタスクの独立変数を教示と W 、従属変数を MT と ER として、データ分布を Anderson-Darling 検定 ($\alpha = 0.05$) で検証したところ、 MT と ER のいずれも正規分布していない条件を含んでいた。したがって aligned rank transform を用いたノンパラメトリック ANOVA[25]

で分散分析を行った。多重比較には Tukey の方法を用いた。

図 8 は各 W について教示ごとの MT を比較したものである。ここでは、 W の主効果 ($F_{2,2250} = 10731.82, p < 0.001, \eta_p^2 = 0.91$)、教示の主効果 ($F_{1,2250} = 364.34, p < 0.001, \eta_p^2 = 0.14$)、教示 × W の交互作用 ($F_{2,2250} = 47.87, p < 0.001, \eta_p^2 = 0.04$) がみられた。また、すべての W で「できるだけ速く」では「できるだけ正確に」より MT が有意に短い結果となった (すべてのペアで $p < 0.001$)。

図 9 は各教示ごとのフィッツの法則への適合を表しており、「できるだけ速く」で $R^2 = 0.9891$ 、「できるだけ正確に」で $R^2 = 0.9941$ となり、両条件とも高い適合度を示した。そのためいずれの教示においても、455 名のデータの平均 MT を用いれば、実験参加者たちは典型的なポインティング操作の挙動を示しているといえる。

図 10 は各 W について教示ごとのエラー率 ER を比較したものである。ここでは、 W の主効果 ($F_{2,2250} = 755.09, p < 0.001, \eta_p^2 = 0.40$)、教示の主効果 ($F_{1,2250} = 566.20, p < 0.001, \eta_p^2 = 0.20$)、教示 × W の交互作用 ($F_{2,2250} = 130.58, p < 0.001, \eta_p^2 = 0.10$) がみられた。また、すべての W で「できるだけ正確に」では「できるだけ速く」より ER が有意に低い結果となった (すべてのペアで $p < 0.001$)。

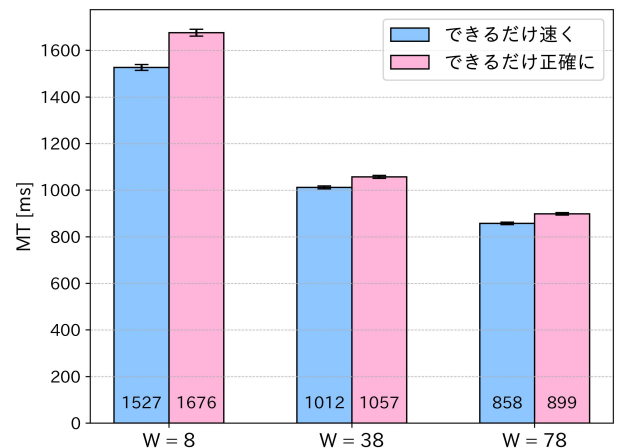


図 8 各条件ごとの MT

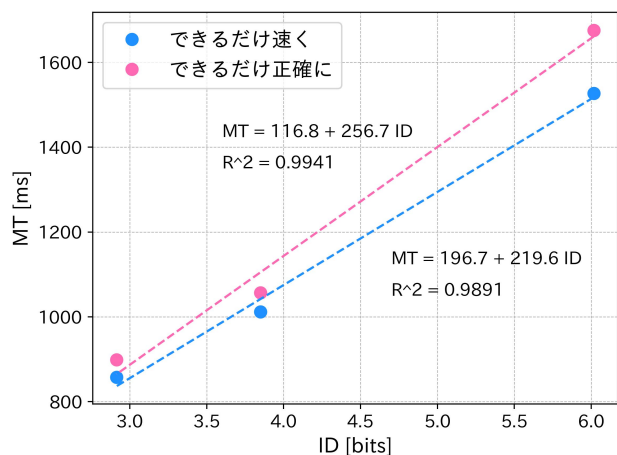


図 9 フィッツの法則への適合

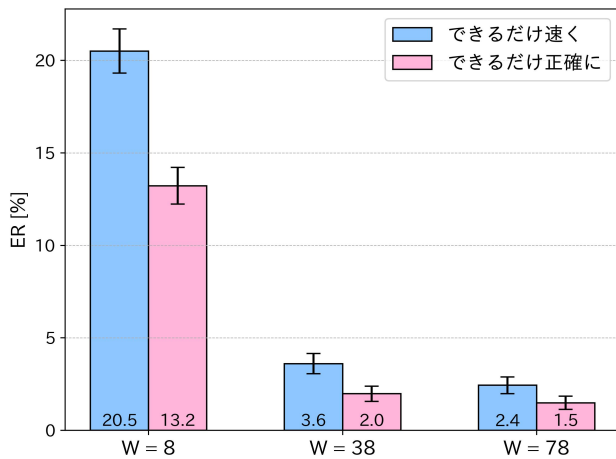


図 10 各条件ごとの ER

5.3 サイズ調整タスクとポインティングタスクの関連性

サイズ調整タスクとポインティングタスクの関連性を分析して仮説を検証するために、サイズ調整タスクの結果に基づいてユーザ分類を行い、それぞれの群についてポインティングタスクの分析を行う。なお、5.2 節におけるポインティングタスクの分析で実施した外れ値の除外は行わず、各参加者がポインティングタスクで行う操作を分析する。これは外れ値を示した参加者をサイズ調整タスクの結果に基づいて特定できるかを調査し、仮説を検証するためである。

5.3.1 2 群間比較による正確な操作を行うユーザの抽出

適切な操作を行うユーザのみに実験を依頼するためには、事前タスクの結果によってユーザを「合格群」と「不合格群」の 2 群に分け、合格群を実験対象とする必要がある。そこで、サイズ調整タスクの結果が正確な参加者を合格群、不正確な参加者を不合格群として、それぞれの群ごとにポインティングタスクの分析を行う。ここでは、5.1 節の結果をもとに、1 回目と 2 回目の調整結果がどちらも 200 px 以上 600 px 未満かつ調整誤差が 20 px 未満の参加者を合格群、それ以外の参加者を不合格群とした。分類基準に調整結果の範囲を使用したのは、1 回目と 2 回目の調整結果がどちらも上限値や下限値付近の不適切な値である参加者や、物理カードを使用せずにタスクを行ったにもかかわらず調整誤差が小さい参加者を除外するためである。分類の結果、合格群 344 人、不合格群 111 人であった。

図 11 は 教示 × W それぞれについて 2 群の ER を比較したものである。ここでは、群の主効果 ($F_{1,453} = 13.22, p < 0.001, \eta_p^2 = 0.03$), 教示 × 群の相互作用 ($F_{1,2265} = 38.70, p < 0.001, \eta_p^2 = 0.02$), W × 群の相互作用 ($F_{2,2265} = 21.25, p < 0.001, \eta_p^2 = 0.02$) がみられた。また、「できるだけ正確に」における W = 8 の条件では合格群は不合格群と比較して ER が 4.0% 低く、より慎重な操作が求められる場面では合格群と不合格群の差が大きくなっている。さらに

図 12 は、「できるだけ正確に」におけるエラー数の分布を 2 群間で比較したものである。ここでは、エラー数が少ない場合は合格群の人数比が大きく、エラー数が多くなるにつれて不合格群の人数比が大きい傾向がみられる。そのため、合格群ではポインティングタスクで正確な操作を行うユーザが多く、不合格群では不正確な操作を行うユーザが多いと考えられる。

また図 13 は、教示による MT の差を 2 群間で比較したものであり、すべての W で合格群のほうが教示による MT の差が大きいがわかる。ここで、実験において教示を正確に守れば「できるだけ速く」では素早い操作、「できるだけ正確に」では慎重な操作を行い、教示による MT の差は大きくなると想定される。各群において「できるだけ速く」の平均 MT が「できるだけ正確に」の平均 MT より短い参加者は、合格群 76% (260 人)、不合格群 66% (73 人) であることから、合格群は実験の指示をより正確に守るユーザが多いと考えられる。

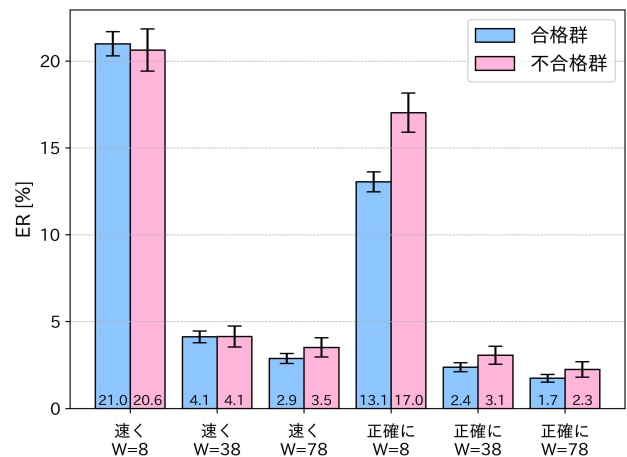


図 11 合格群と不合格群の ER の比較

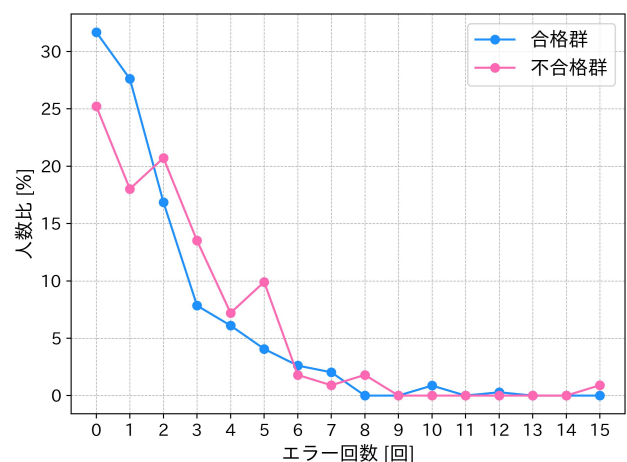


図 12 「できるだけ正確に」におけるエラー数の人数分布

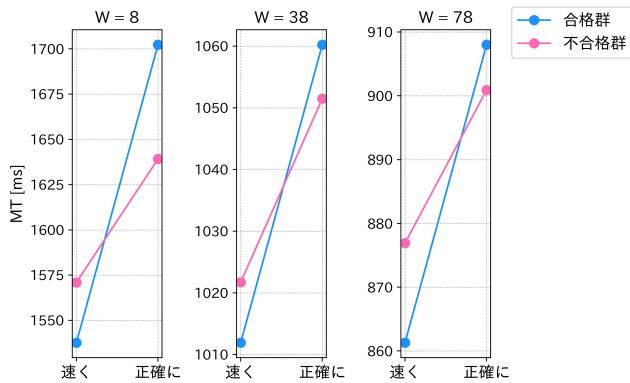


図 13 教示による MT の差

以上のように、各ユーザのサイズ調整タスクにおける操作の正確性がポインティングタスクでの操作の正確性に影響を及ぼし、さらにサイズ調整タスクを正確に行うユーザは実験の指示をより正確に守るといった傾向がみられている。このことから、事前タスクとしてサイズ調整タスクを実施してその結果をもとにユーザ分類を行うことで、より正確に指示を守って操作を行うユーザを抽出できると考えられる。また、正確に直径 8 px のターゲットをクリックするという慎重な操作が要求される場面とその傾向が顕著であったことから、事前タスクによる適切なユーザの抽出は、精密な操作が求められる実験や負荷の高い実験においてより効果を発揮すると考えられる。

5.3.2 4 群間比較によるタスク間の関連性の分析

サイズ調整タスクの操作とポインティングタスクの操作の関連性をより詳細に分析するため、各参加者をサイズ調整タスクの操作が正確な「高精度群」、ある程度正確な「中精度群」、不正確な「低精度群」、不適切な「不適正群」の 4 群に分類して分析を行う。ここでは、1 回目と 2 回目の調整結果がどちらも 200 px 以上 600 px 未満の参加者のうち、調整誤差 10 px 未満を高精度群、調整誤差 10 px 以上 20 px 未満を中精度群、20 px 以上 50 px 未満を低精度群として、それ以外の参加者を不適正群とした。分類の結果、高精度群 294 人、中精度群 50 人、低精度群 27 人、不適正群 84 人であった。

図 14 は 教示 × W それぞれについて 4 群の ER を比較したものである。ここでは、ER について群の主効果 ($F_{3,451} = 8.77, p < 0.001, \eta_p^2 = 0.06$)、教示 × 群の相互作用 ($F_{3,2255} = 12.13, p < 0.001, \eta_p^2 = 0.02$)、W × 群の相互作用 ($F_{6,2255} = 15.29, p < 0.001, \eta_p^2 = 0.04$) がみられた。また、「できるだけ正確に」における W = 8 の条件では他の群と比較して高精度群の ER が低い傾向がみられ、高精度群と中精度群のペアでは 5.8% の差、高精度群と不適正群のペアでは 5.4% の差がみられた。この結果から、より慎重な操作が要求される場面において、高精度群では操作が正確な傾向、不適正群では操作が不正確な傾向があ

り、サイズ調整タスクの正確性がポインティングタスクの正確性に影響を及ぼすことが考えられる。一方で、中精度群は不適正群よりも ER が高く、低精度群は高精度群よりも ER が高いが中精度群よりも ER が低いことから、タスク間の操作に関連がみられない参加者も存在することが考えられる。

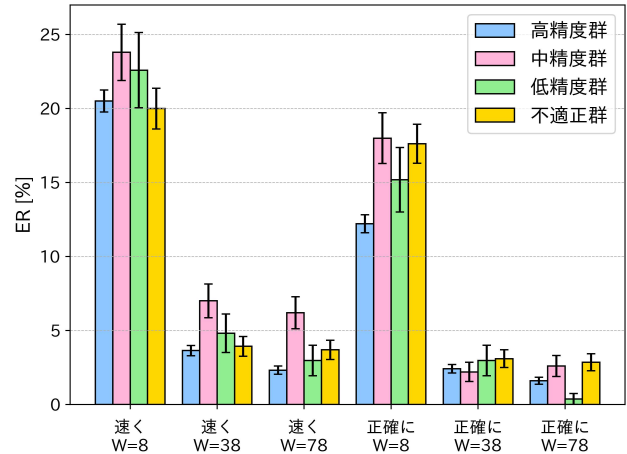


図 14 4 群間の ER の比較

以上のように、4 群間の比較による分析では、高精度群と不適正群において各ユーザのサイズ調整タスクの操作の正確性とポインティングタスクの操作の正確性に関連がみられるが、中精度群と低精度群においてはタスク間の操作に関連がみられない可能性が示された。ここでは、中精度群であるがポインティングタスクではエラー率が高い参加者や、低精度群であるがポインティングタスクではエラー率が低い参加者も存在すると想定され、事前タスクの実施のみでは適切なユーザを完全に抽出することができないことが考えられる。また、5.3.1 項で行った 2 群間比較において「合格群」に分類されたユーザをさらに「高精度群」と「中精度群」に分類した結果、正確に直径 8 px のターゲットをクリックする慎重な操作が要求される場面で 2 群間の ER には 5% 以上の差がみられた。このことから、事前タスクとしてサイズ調整タスクを実施することでより正確に指示を守るユーザを抽出できるが、その分類基準を主観的な閾値とする場合には慎重に設定を行う必要があることが示唆されており、今後の課題であるといえる。

6. おわりに

本稿では、Web ベースの GUI 実験におけるデータ品質向上を目的として、適切な操作を行うユーザのみに実験を依頼するアプローチの初期検討を行った。実験では、事前タスクとしてサイズ調整タスク、その後主タスクとしてポインティングタスクを行い、事前タスクによって GUI 実験において適切な操作を行うユーザを抽出することを試みた。その結果、サイズ調整タスクの操作の正確性はポインティングタスクの操作の正確性と関連があり、サイズ調整

タスクにおいて正確な操作を行うユーザはポインティングタスクにおいてより正確に指示を守る傾向がみられた。このことから、事前タスクにおいて抽出された正確な操作を行うユーザのみに実験を依頼することで、より高品質な実験データを収集できる可能性が示唆された。また、実験ではより慎重な操作が要求される場面において事前タスクと主タスクの操作の関連性が強く、事前タスクによる適切なユーザの抽出は精密な操作が求められる実験や負荷の高い実験において効果的である可能性も示唆された。一方で、タスク間の操作に関連がみられないユーザが存在することや、ユーザ分類の基準を主観的な閾値によって決定することによる課題も示唆されており、事前タスクによる適切なユーザの抽出方法には改善が必要である。そのため今後は、実施する実験に適した事前タスクの探究を行うとともに、データ量を確保しつつも適切なユーザを抽出できる閾値の設定方法を検討する予定である。

謝辞 本研究の一部は JSPS 科研費 JP22K12135 の助成を受けたものです。

参考文献

- [1] Reips, U.-D.: Standards for Internet-based experimenting., *Experimental psychology*, Vol. 49, No. 4, p. 243 (2002).
- [2] Oppenheimer, D. M., Meyvis, T. and Davidenko, N.: Instructional manipulation checks: Detecting satisficing to increase statistical power, *Journal of experimental social psychology*, Vol. 45, No. 4, pp. 867–872 (2009).
- [3] Yamanaka, S.: Utility of crowdsourced user experiments for measuring the central tendency of user performance to evaluate error-rate models on guis, *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 9, pp. 155–165 (2021).
- [4] Findlater, L., Zhang, J., Froehlich, J. E. and Moffatt, K.: Differences in crowdsourced vs. lab-based mobile and desktop input performance data, *Proceedings of the 2017 CHI conference on human factors in computing systems*, pp. 6813–6824 (2017).
- [5] Crump, M. J., McDonnell, J. V. and Gureckis, T. M.: Evaluating Amazon’s Mechanical Turk as a tool for experimental behavioral research, *PLoS one*, Vol. 8, No. 3, p. e57410 (2013).
- [6] Horton, J. J., Rand, D. G. and Zeckhauser, R. J.: The online laboratory: Conducting experiments in a real labor market, *Experimental economics*, Vol. 14, pp. 399–425 (2011).
- [7] Li, Q., Joo, S. J., Yeatman, J. D. and Reinecke, K.: Controlling for participants’ viewing distance in large-scale, psychophysical online experiments using a virtual chinrest, *Scientific reports*, Vol. 10, No. 1, p. 904 (2020).
- [8] Brühlmann, F., Petralito, S., Aeschbach, L. F. and Opwis, K.: The quality of data collected online: An investigation of careless responding in a crowdsourced sample, *Methods in Psychology*, Vol. 2, p. 100022 (2020).
- [9] Curran, P. G.: Methods for the detection of carelessly invalid responses in survey data, *Journal of Experimental Social Psychology*, Vol. 66, pp. 4–19 (2016).
- [10] 三浦麻子, 小林哲郎: オンライン調査モニタの Satisfice に関する実験的研究, 社会心理学研究, Vol. 31, No. 1, pp. 1–12 (2015).
- [11] Gottfried, J.: Practices in Data-Quality Evaluation: A Large-Scale Review of Online Survey Studies Published in 2022, *Advances in Methods and Practices in Psychological Science*, Vol. 7, No. 2 (2024).
- [12] Komarov, S., Reinecke, K. and Gajos, K. Z.: Crowdsourcing performance evaluations of user interfaces, *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 207–216 (2013).
- [13] Grossman, T. and Balakrishnan, R.: The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor’s activation area, *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 281–290 (2005).
- [14] Schwab, M., Hao, S., Vitek, O., Tompkin, J., Huang, J. and Borkin, M. A.: Evaluating pan and zoom timelines and sliders, *Proceedings of the 2019 chi conference on human factors in computing systems*, pp. 1–12 (2019).
- [15] Fitts, P. M.: The information capacity of the human motor system in controlling the amplitude of movement., *Journal of experimental psychology*, Vol. 47, No. 6, p. 381 (1954).
- [16] MacKenzie, I. S.: Fitts’ law as a research and design tool in human-computer interaction, *Human-computer interaction*, Vol. 7, No. 1, pp. 91–139 (1992).
- [17] Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E. and Keith Smith, J.: Optimality in human motor performance: ideal control of rapid aimed movements., *Psychological review*, Vol. 95, No. 3, p. 340 (1988).
- [18] Wobbrock, J. O., Cutrell, E., Harada, S. and MacKenzie, I. S.: An error model for pointing based on Fitts’ law, *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1613–1622 (2008).
- [19] Sharif, A., Pao, V., Reinecke, K. and Wobbrock, J. O.: The reliability of Fitts’ s law as a movement model for people with and without limited fine motor function, *Proceedings of the 22nd international Acm Sigaccess conference on computers and accessibility*, pp. 1–15 (2020).
- [20] Yamanaka, S.: Test-Retest Reliability on Movement Times and Error Rates in Target Pointing, *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, pp. 178–188 (2022).
- [21] 木下大樹, 大場洋介, 富張瑠斗, 山中祥太, 宮下芳明: 速さと正確さへの主観的なバイアスがポインティング手法評価の一般化可能性に与える影響, インタラクシオン 2023 論文集, 情報処理学会, pp. 78–87 (2023).
- [22] 金谷一輝, 徳原真彩, 三山貴也, 木下裕一朗, 中村聡史: 画像選択肢の段階的表示速度の違いが選択に及ぼす影響, *HCS2024-50*, Vol. 124, No. 161, pp. 90–95 (2024).
- [23] 宮本快士, 三山貴也, 中村聡史, 山中祥太: 待機画面におけるプログレスバーの進行速度変化が離脱に及ぼす影響, *HCS2024-52*, Vol. 124, No. 161, pp. 102–107 (2024).
- [24] De Leeuw, J. R.: jsPsych: A JavaScript library for creating behavioral experiments in a Web browser, *Behavior research methods*, Vol. 47, pp. 1–12 (2015).
- [25] Wobbrock, J. O., Findlater, L., Gergle, D. and Higgins, J. J.: The aligned rank transform for nonparametric factorial analyses using only anova procedures, *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 143–146 (2011).