

クラウドソーシングを活用したGUI実験における 参加者スクリーニング手法の検証

三山 貴也¹ 中村 聡史¹ 山中 祥太²

概要: クラウドソーシングを活用した GUI 実験では、多くの参加者を素早く募集でき、UI 操作時のデータを大量に収集できる。一方で、指示を守らない参加者や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する懸念がある。我々はこれまで、実験前の事前タスクによって参加者をスクリーニングする手法を検討してきた。しかしこれまでの研究では、スクリーニングが実験結果に与える影響の検証が不十分であり、手法の適切な評価がなされていなかった。そこで本稿では、不適切な参加者が混入した場合の実験結果への影響をシミュレーションすることにより、手法の有用性の検証を行った。その結果、不適切な参加者の割合を少なくするほど GUI 操作のパフォーマンスモデルの推定精度が向上する傾向がみられ、事前タスクによるスクリーニング手法の有用性が示唆された。

1. はじめに

クラウドソーシング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、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、クラウドソーシング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、クラウドソーシング実験におけるデータ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実験が一般的になっている。利点の例として、GUI 操作時のデータを大量に収集できることから、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適切に扱える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は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な傾向があり、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、クラウドソーシングを活用した GUI 実験においても、

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。

我々はこれまでの研究 [5] において、クラウドソーシングを活用した GUI 実験におけるデータ品質向上を目的として、実験前の事前タスクによって参加者のスクリーニングを行い、適切なユーザ群のみに本来目的とする実験を依頼するアプローチを提案してきた。ここでは、事前タスクとして簡単なマウス操作課題を用意し、その後にポインティングタスクを実施した結果、教示を正確に遵守して操作するユーザを抽出できる可能性が示唆されている。しかしこれまでの研究では、事前タスクによるスクリーニングが実験結果に及ぼす影響の検証が不十分であり、スクリーニング手法の適切な評価がなされていなかった。

そこで本稿では、スクリーニングの結果として不適切とされた参加者が混入した場合の実験結果についてシミュレーションを行い、スクリーニング手法の有用性を検証する。具体的には、参加者全体に含まれる不適切な参加者の割合を変化させたときの実験結果において GUI 操作のパフォーマンスモデルの推定精度を算出し、スクリーニングによってモデル推定精度が向上するかどうかを評価する。

2. 関連研究

2.1 クラウドソーシング実験のデータ品質

クラウドソーシング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いる。心理学や社会学分野の研究テーマでは、両方の実験環境において多くの場合で同様の結果が

¹ 明治大学
Meiji University

² LINE ヤフー株式会社
LY Corporation

得られるものの、一部では異なる結果になると報告されている [6][7]。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 [8] 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、視覚刺激の物理サイズ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 [9] 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 [10]。また Oppenheimer ら [2] は指示を読んでいるかを測定する質問 (Instructional Manipulation Check, IMC) を提案し、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がある [11]。

以上のように、クラウドソーシング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、実験前の事前タスクによって参加者のスクリーニングを行うアプローチにより、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 クラウドソーシングを活用した GUI 実験

GUI 実験についても、クラウドソーシング実験と実験室実験の結果の比較が行われている。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実験室実験の結果が、クラウドソーシング実験でも再現されたと報告している。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、マウスおよびタッチ操作のポインティングタスクで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、クラウドソーシングを活用した GUI 実験においても実験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に適した事前タスクによって参加者をスクリーニングすることで、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

ポインティングタスクの代表的なモデルとしてフィッツの法則があり、ターゲットまでの距離 A とターゲットの幅 W に基づいて、最初のクリックまでの時間 MT を予測できる [15][16]。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 に関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。また、

ER を予測する研究も行われており、長方形や円形のターゲットについてモデルが提案されている [19][20]。

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 [21][22]。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 [23]。

3. 提案アプローチ

3.1 コンセプト

本研究では、クラウドソーシングを活用した GUI 実験における実験データの品質向上を目的として、実験前の事前タスクによって参加者のスクリーニングを行い、本来目的とする実験 (主タスク) を適切なユーザ群のみに依頼するアプローチを提案する。具体的には、主タスクの操作と関連のある操作を含む事前タスクを設け、その操作パフォーマンスに基づいて適切なユーザを抽出することで、事前タスクのみの実施によって適切なユーザ群に実験を依頼可能とする。これにより、参加者の多くが適切な操作を行うユーザとなり、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。

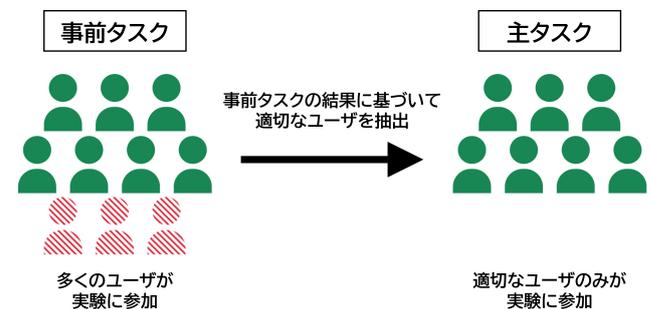


図 1 提案アプローチのイメージ図

3.2 参加者スクリーニングのための事前タスク

本稿では、主タスクとしてポインティングタスクを扱うため、ポインティングと関連のある操作を含む事前タスクを用意することで、適切な参加者をスクリーニングできると考えられる。ここで我々は、Liら [8] 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、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、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。以上の理由から、正確な操作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、サイズ調整タスクを事前タスクとして利用できると考えた。

またサイズ調整タスクは、本来は視覚刺激の物理サイズを統制するために実験前に実施されるが、我々がこれまでに実施したサイズ調整タスクで不適切な操作や雑な操作を行う参加者が存在した [24]。そのため、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、適切なユーザを抽出できると考えた。以上の理由から、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 [8] によるサイズ調整タスクのイメージ図

4. 分析データ

本稿で分析に用いるのは、これまでの研究 [5] で得られた実験データである。本章では、その実験について説明する。

4.1 実験概要

実験はすべての参加者に対して同様の流れで実施され、実験の説明の後、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 セット行った。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、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。455 人（男性 354 人、女性 101 人）のデータが分析対象となっている。

4.2 サイズ調整タスク

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 Li ら [8] が提案したタスクである。ここで、クラウドソーシング実験では参加者ごとにディスプレイの画素密度が異なり、サイズ調整タスク 1 回では参加者が正確に調整を行ったかを判断できない。そこで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もとに各参加者の操作の正確性を評価できるようにした。ここでは、クレジットカードと同じサイズの物理カード（ISO/IEC 7810 の ID-1 規格：縦 53.98 mm × 横 85.60 mm）を使用可能とした。なお、サイズ調整タスクで不適切な操作を行った場合でもポインティングタスクで過度に大きいまたは小さい視覚刺激が表示されることを防ぐため、実験ではサイズ調整タスクの結果をもとに視覚刺激の物理サイズは統制していない。

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

4.3 ポインティングタスク

実験におけるポインティングタスクは、円形のターゲットをクリックすると次のターゲットが表示される試行を繰り返すタスクである。ここでは、ターゲット間の距離 A は 510 px の 1 条件、ターゲット直径 W は 8, 38, 78 px の 3 条件とした。また、素早い操作を要求する「速く」とミスクリックをしないことを要求する「正確に」の 2 つの指示を用意し、指定された指示に従って操作を行ってもらった。

実験では 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ている。また、1 セットごとに 2 つの指示のいずれかをランダムに割り当て、全 4 セットのうち各指示が 2 セットずつ割り当てられている。以上により、参加者 1 人あたりの総試行回数は $3W \times 5$ 試行 $\times 2$ 指示 $\times 2$ セット = 60 試行であり、1 つの（指示 $\times W$ ）条件について、5 試行 $\times 2$ 指示 = 10 試行分のデータが記録されている。

5. シミュレーション

5.1 分析概要

4 章で述べた分析データを用いて、事前タスクによるスクリーニングの結果として不適切とされた参加者が混入した場合の実験結果についてシミュレーションを行い、スクリーニング手法の有用性を検証する。ここでは、参加者全体に含まれる不適切な参加者の割合を変化させたときのポインティングタスクの結果について GUI 操作のパフォーマンスモデルの推定精度を算出し、スクリーニングによってモデル推定精度が向上するかどうかを検証する。

本稿で扱う GUI 操作のパフォーマンスモデルを式 (1)、式 (2) に示す。式 (1) は、ターゲットまでの距離 A とターゲット直径 W から操作時間 MT を予測するモデルである [16]。また式 (2) は、直径 W のターゲット領域 D におけるクリック成功率 $P(D)$ を予測するモデルであり、実験で用いた各 W についてターゲットの中心を原点とするクリック座標の標準偏差を求め、回帰分析によって任意の W における標準偏差 σ_{reg_x} , σ_{reg_y} を得ることにより、成功率を予測できる [20]。

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

$$P(D) = \iint_D \frac{1}{2\pi\sigma_{\text{reg}_x}\sigma_{\text{reg}_y}} \exp \left(-\frac{x^2}{2\sigma_{\text{reg}_x}^2} - \frac{y^2}{2\sigma_{\text{reg}_y}^2} \right) dx dy \quad (2)$$

5.2 前処理

シミュレーション前にポインティングタスクにおける外れ値の除外を行った。まず、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 の半分未満の試行を除外した。次に、指示 $\times W$ それぞれで 1 人あたり 10 試行であるが、そのなかで操作時間 MT が外れ値となる試行を除外した。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 IQR 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。

5.3 分析手順

参加者をスクリーニングする場合、事前タスクの結果によって参加者を「合格群」と「不合格群」に分け、適切なユーザー群である合格群に実験を依頼する。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に分類して分析を行う。ここでは、これまでの研究 [5] の結果をもとに、2 回のサイズ調整の結果が適正範囲内かつ調整結果の誤差が閾値 T (px) より小さい参加者を合格群、それ以外の参加者を不合格群とする。

以上を踏まえてシミュレーションでは、まず閾値 T によって分析データに含まれる参加者を合格群と不合格群に分類し、参加者数が合計 N の状況において不合格群の割合 $X\%$ を変化させる検証を行うため、不合格群から $N \times X\%$ 人、合格群から $N \times (100 - X)\%$ 人をランダムに抽出する。これにより、参加者数が N のとき、閾値を T としたところ、不合格群の割合は $X\%$ であったという状況を再現する。次に、抽出した参加者群におけるポインティングタスクの結果について、式 (1) と式 (2) のモデル推定精度を算出する。以上の操作を 1,000 回繰り返してモデル推定精度の平均をとることにより、参加者数 N 、閾値 T 、不合格群の割合 X の場合におけるモデル推定精度とする。

本シミュレーションでは N 、 T 、 X を変化させ、様々な場合におけるモデル推定精度を算出することで、不適切な参加者の混入が実験結果に及ぼす影響を分析する。ここでは、 N を 10, 20, 40, 80 人、 T を 5 px から 50 px まで 5 px ずつ区切った値、 X を 10% から 100% まで 10% ずつ区切った値とした。 N を最大 80 人、 T を最大 50 px としたのは、参加者数 N を大きくするほど多くの不合格群のデータが分析に必要となり、閾値 T を緩く設定すると不合格群に分類される参加者が減少するため、分析に必要なデータを十分に確保できないためである。閾値 T を変化させたときの合格群と不合格群の人数を表 1 に示す。

表 1 閾値の変化による合格群と不合格群の人数

閾値 T (px)	5	10	15	20	25	30	35	40	45	50
合格群 (人)	204	294	319	344	353	361	366	369	371	371
不合格群 (人)	251	161	136	111	102	94	89	86	84	84

6. 結果と考察

6.1 操作時間 MT について

図 3、図 4 は、閾値 T と不合格群の割合 X を変化させ、

各条件におけるポインティングタスクの操作時間 MT の結果について、式 (1) に対するモデル推定精度 R^2 を評価した結果を示したものである。ここでは、ターゲット直径 W の各値を 1 つずつ除外しながら LOOCV (Leave-One-Out クロスバリデーション) を行っている。図 3 はポインティングタスクの教示が「速く」の場合、図 4 は「正確に」の場合で、シミュレーションにおける参加者数 N ごとに結果をヒートマップで示している。

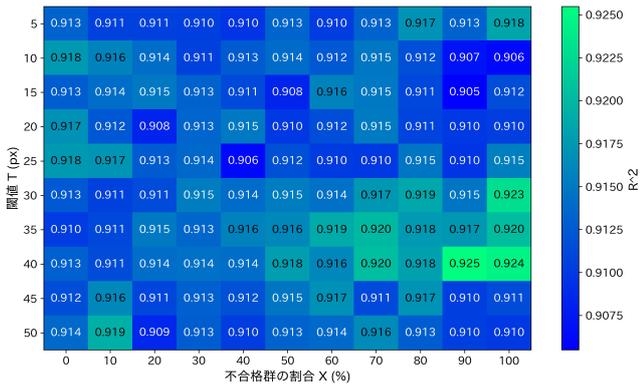
図 3、図 4 において、閾値 T や不合格群の割合 X の値によって推定精度は大きく変化せず、参加者数 N の増加にしたがって推定精度は微増するものの、大きくは変化しない。この要因として、スクリーニングを操作の正確性をもとに行ったことや、 MT のモデルは ER のモデルと比較して少人数の少ない反復回数でも十分なモデル推定精度を示す [3] ことが考えられる。閾値 T は 2 回のサイズ調整タスクの調整結果の誤差であり、誤差の大きい参加者は操作が雑な参加者と考えられるが、式 (1) は操作時間 MT に関するモデルであるためエラー率 ER に関するモデルと比較して操作の正確性の重要度が低く、操作が雑な参加者が多い場合でも実験結果に与える影響は小さいと考えられる。以上より、式 (1) のような MT のモデルをクラウドソーシング実験によって検証する場合は、本稿におけるスクリーニング手法の有用性は小さく、必要性は低いと考えられる。

6.2 エラー率 ER について

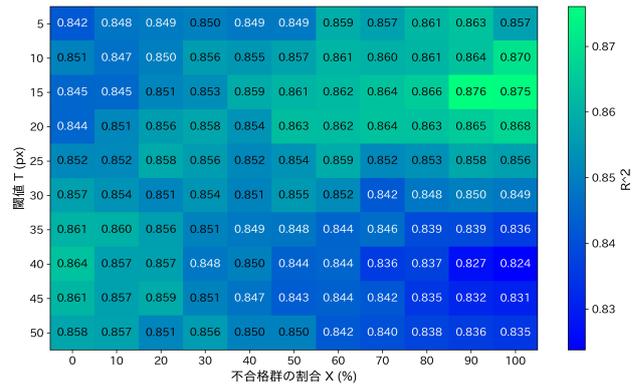
6.2.1 スクリーニング手法の有用性

図 5、図 6 は、エラー率 ER について操作時間 MT と同様の分析を行い、式 (2) に対するモデル推定精度 R^2 を評価した結果を示したものである。図 5、図 6 において、各ヒートマップの左上から右下にかけてモデル推定精度が低下しており、閾値を緩く設定した場合に不合格群の割合が多くなると、モデル推定精度が低下する傾向があることがわかる。ここで、閾値を緩く設定した場合は、サイズ調整タスクにおける誤差がかなり大きく、操作が非常に雑な参加者が不合格群となるため、そのような参加者の割合が増加することでモデル推定精度が低下していると考えられる。

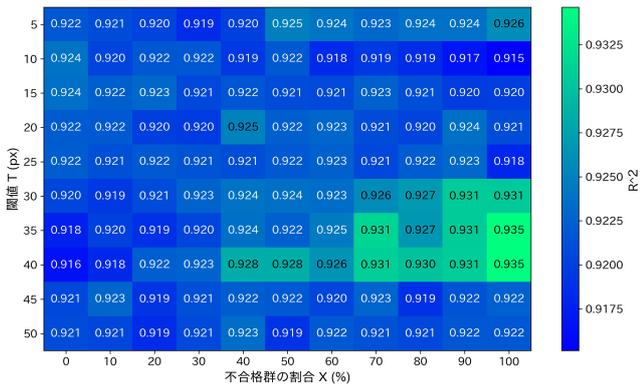
この結果は、参加者群によってモデル評価の結果が異なる可能性があることを示しており、 $N = 40$ の場合に閾値 T と不合格群の割合 X の値によって R^2 は約 0.75 から約 0.95 までの値をとることからも、参加者群の性質がモデル評価に及ぼす影響は大きいと考えられる。そのため、式 (2) のような ER のモデルをクラウドソーシング実験で検証する場合は、参加者のスクリーニングを行い、適切な参加者に実験を依頼することが重要だと考えられる。また、本稿におけるサイズ調整タスクを利用したスクリーニング手法は、操作の正確性をもとに適切な参加者を抽出するため、正確な操作が求められる ER の検証においては効果を発揮すると考えられる。不合格群の割合を少なくするほどモデ



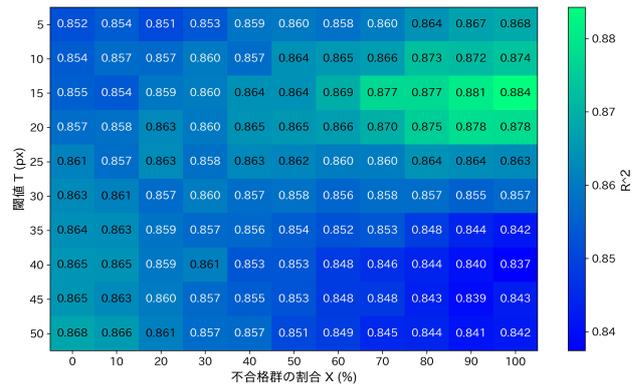
(a) $N = 10$



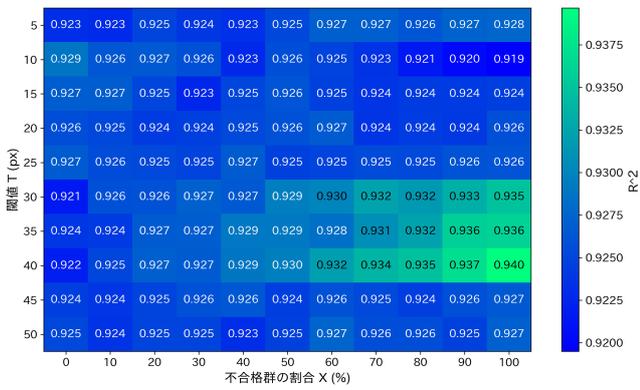
(a) $N = 10$



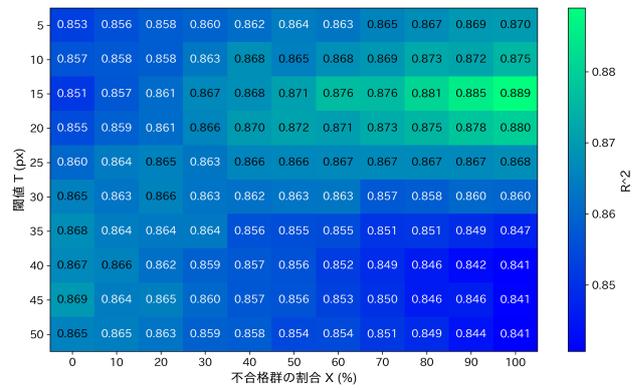
(b) $N = 20$



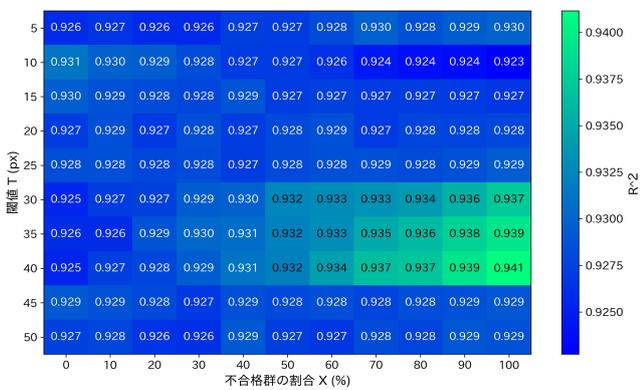
(b) $N = 20$



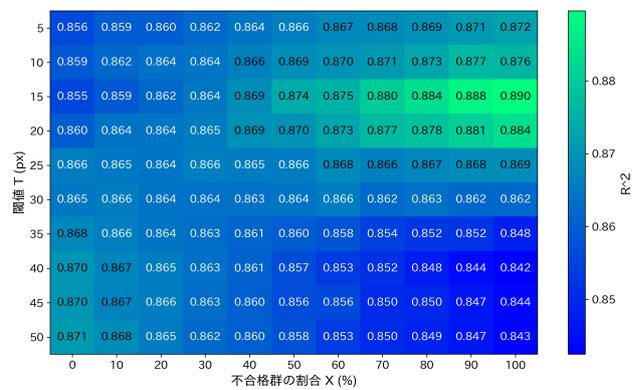
(c) $N = 40$



(c) $N = 40$

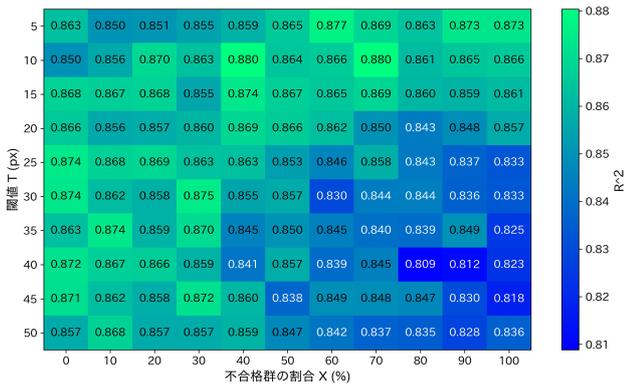


(d) $N = 80$

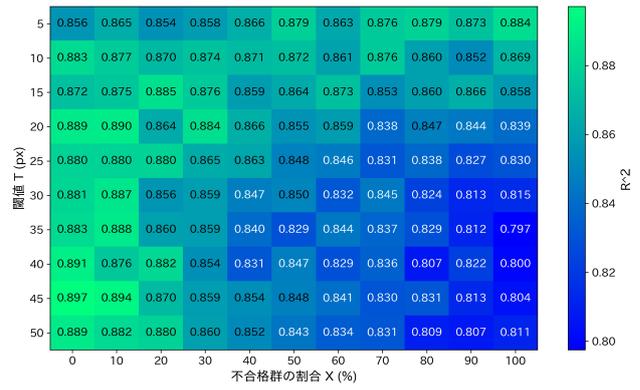


(d) $N = 80$

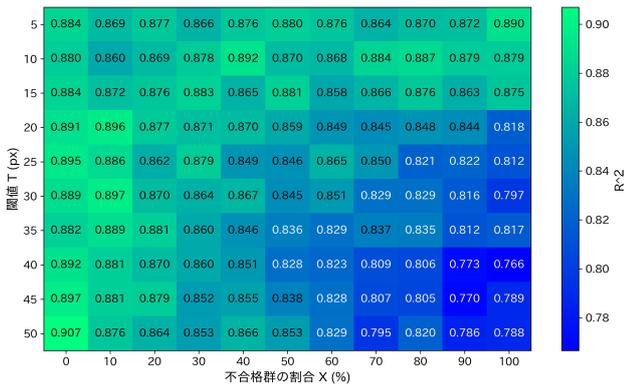
図3 MTについて式(1)に対するモデル推定精度(速く) 図4 MTについて式(1)に対するモデル推定精度(正確に)



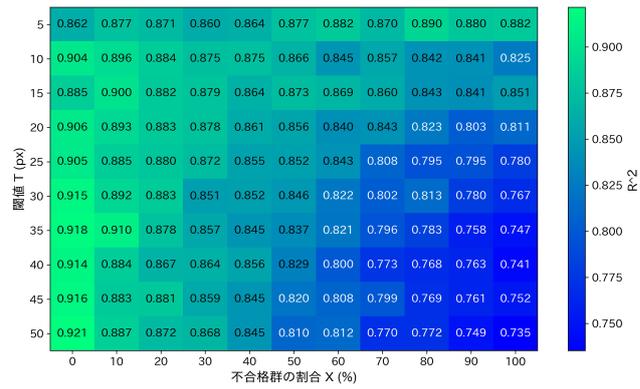
(a) $N = 10$



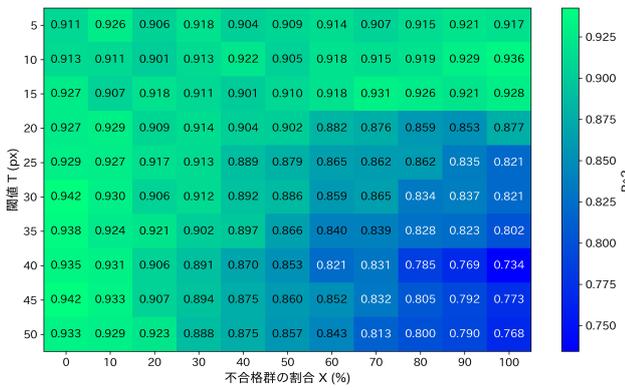
(a) $N = 10$



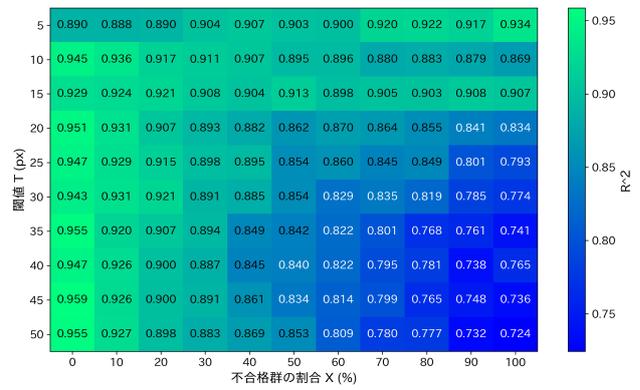
(b) $N = 20$



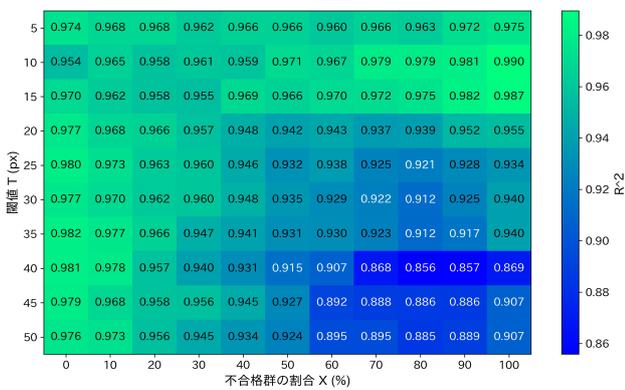
(b) $N = 20$



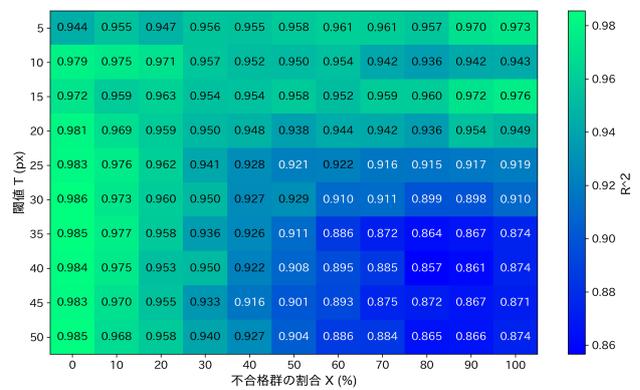
(c) $N = 40$



(c) $N = 40$



(d) $N = 80$



(d) $N = 80$

図 5 ER について式 (2) に対するモデル推定精度 (速く) 図 6 ER について式 (2) に対するモデル推定精度 (正確に)

ル推定精度が向上する傾向があることから、スクリーニング手法の有用性が示唆されたといえる。

また、参加者数 N の増加にしたがって各ヒートマップの左上部分のモデル推定精度は向上しており、 N を増やすことがモデル評価に有用であるというクラウドソーシング実験の利点が示されている。不合格群の割合の増加による影響についても、 $N = 80$ のヒートマップの右下部分におけるモデル推定精度は他の N と比較して高く、さらに参加者数を増やすことで不適切な参加者の混入による影響は最小限に抑制できると考えられる。しかし $N = 20$ や $N = 40$ においては、 R^2 の値が 0.9 を超える場合と 0.75 付近の場合が混在しており、 $N = 80$ でも 0.98 を超える場合と 0.85 付近の場合が混在している。そのため、参加者数がモデル評価には十分な場合でも、不適切な参加者の存在がモデル評価に影響を及ぼす可能性がある。本稿におけるスクリーニング手法は、不適切な参加者の混入による影響を抑えるには参加者数が十分でない場合にその影響を抑制し、モデル推定精度の向上に寄与できると考えられる。

6.2.2 モデル推定精度が低下する要因

ポインティングタスクの教示が「正確に」の場合に ER のモデル推定精度が低下する要因として、不合格群には操作が雑な参加者が多いことが考えられる。図 7 は参加者数とターゲット直径をそれぞれ $N = 80$ と $W = 8$ に設定したシミュレーションにおける、エラー率の実測値の平均を示したものである。ここでは、不合格群の割合が多くなるとエラー率が上昇する傾向があり、5%以上の差がある場合もある。ここで閾値 T は 2 回のサイズ調整タスクの調整結果の誤差であり、誤差が大きく不合格群に分類される参加者は操作が雑だと考えられる。そのため、不合格群の割合が増加すると観測されるエラー率が上昇し、結果的にモデル推定精度が低下する可能性がある。

また、不合格群には教示に沿った操作を行わない参加者が多いことも要因として考えられる。実験において教示を正しく守れば「速く」では素早い操作、「正確に」では慎重な操作を行い、操作時間 MT の差は大きくなる。図 8 は、 $N = 80$ における $W = 8$ の MT について「正確に」の場合と「速く」の場合の差を示したものである。ここでは、不合格群の割合が多くなると操作時間の差が小さくなる傾向がある。そのため、不合格群の割合が増加すると教示に沿った操作を行わない参加者によるノイズが混入しやすく、結果的にモデル推定精度が低下する可能性がある。

7. 課題と今後の展望

本稿では、不適切な参加者の混入が GUI 実験の結果に及ぼす影響をシミュレーションにより分析し、サイズ調整タスクを利用したスクリーニングはエラー率 ER に関するモデルの推定精度を向上させることを示した。一方で、シ

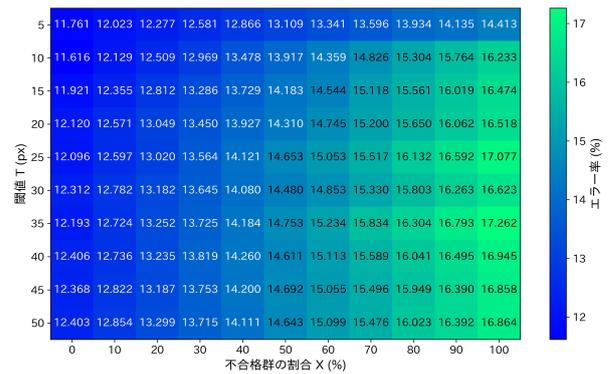


図 7 $N = 80$ における $W = 8$ のエラー率 (正確に)

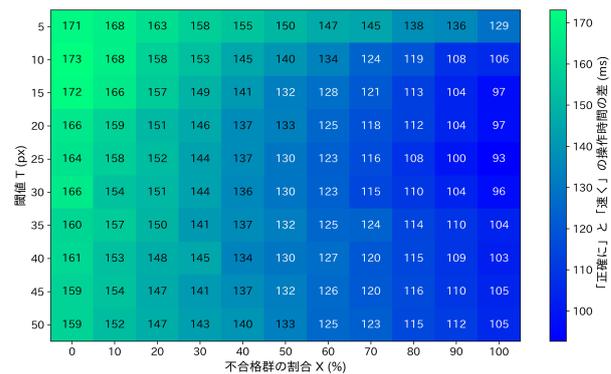


図 8 $N = 80$ における $W = 8$ の教示による MT の差

ミュレーションの結果からスクリーニング手法には課題もあることがわかり、今後さらなる検証が求められる。

まず、図 5、図 6 の各ヒートマップにおいて最も左上に位置する閾値 $T = 5$ 、不合格群の割合 $X = 0$ の場合にモデル推定精度が最良ではない場合が多い。 $T = 5$ は最も厳しい閾値の条件であり、 $X = 0$ では不合格群に属する参加者は存在しないため、実験実施者にとって理想的な参加者群だと考えられる。そのような参加者群にもかかわらずモデル推定精度が最良とならない要因として、スクリーニング手法の制約が考えられる。本稿におけるスクリーニングでは、事前タスクでは適切な操作を行ったが、その後の主タスクでは適切な操作を行わなかった参加者を特定できない制約がある。そのため、スクリーニングの結果として理想的とされる参加者群でもノイズとなる参加者を含む可能性がある。また、実験のポインティングタスクではターゲット直径 W が 3 条件、教示 $\times W$ における 1 人あたりのデータは 10 試行とやや少ない設定となっていた。そのため、今後は W の条件数や 1 人あたりの試行数を増やしてモデル推定精度をより厳密に評価する実験を行う予定である。

また、サイズ調整タスクによるスクリーニングでは閾値を設定する必要があるが、図 5、図 6 のように閾値や不合格群の割合が変化するとモデル推定精度は段階的に変化するため、閾値を一意に決定することは難しいと考えられる。閾値を厳しく設定すれば、より適切な操作を行う参加者をスクリーニングできると考えられるが、参加者数が減少し

てしまう可能性があるため、実験実施者が実験に合わせて閾値を設定することが求められると考えられる。

本稿における参加者スクリーニングでは、事前タスクとして設定したサイズ調整タスクにおける操作の正確性をともに適切な参加者を抽出した結果、主タスクであるポインティングタスクの操作の正確性（エラー率 ER ）に関するモデルの推定精度を向上させることができた。ここで、サイズ調整タスクはポインティングタスクとは異なるタスクであるため、この結果は事前タスクが主タスクと異なるタスクの場合も適切にスクリーニングを行うことができることを示している。そのため、主タスクを変更するたびに事前タスクを変更する必要はなく、ステアリングの法則など他の GUI 実験にもサイズ調整タスクによるスクリーニングが適切に適用できると考えられる。今後はサイズ調整タスクによるスクリーニング手法が他の GUI 実験にも適用可能かどうかの検証も行う予定である。

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] 三山貴也, 中村聡史, 山中祥太: 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, 情報処理学会 研究報告ヒューマンコンピュータインタラクション (HCI), Vol. 2025-HCI-211, No. 14, pp. 1–8 (2025).
- [6] 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).
- [7] 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).
- [8] 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).
- [9] 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).
- [10] 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).
- [11] 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).
- [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] Bi, X. and Zhai, S.: Predicting finger-touch accuracy based on the dual Gaussian distribution model, *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pp. 313–319 (2016).
- [20] Yamanaka, S. and Usuba, H.: Rethinking the dual gaussian distribution model for predicting touch accuracy in on-screen-start pointing tasks, *Proceedings of the ACM on Human-Computer Interaction*, Vol. 4, No. ISS, pp. 1–20 (2020).
- [21] 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).
- [22] 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).
- [23] Yamanaka, S., Kinoshita, T., Oba, Y., Tomihari, R. and Miyashita, H.: Varying subjective speed-accuracy biases to evaluate the generalizability of experimental conclusions on pointing-facilitation techniques, *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–13 (2023).
- [24] 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).