ユーザの自然なインタラクションに基づく操作ミス推測

一居 和毅[1,a)]　池松 香[2,b)]　礒本 俊弥[2,c)]　加藤 邦拓[3,d)]　杉浦 裕太[1,e)]

概要：近年，スマートフォンやタブレットなどのタッチスクリーンデバイスが日常生活に深く浸透している．しかし，誤ったタッチ操作や選択ミスによって，ユーザ体験や操作効率が損なわれることが課題となっている．本研究では，ユーザの操作ミスからの迅速な復帰を可能にする UI の設計に向けて，センサデータを活用した入力タスクにおける操作ミスの判定手法を提案する．9 軸センサ，視線座標，表情，タップ座標，タップした要素の中心座標，画面の注視度などのデータを収集し，機械学習によって正誤判定を行うシステムを構築した．実験では，センサデータに基づく判定モデルにより，個人モデルを用いて，平均 87.3%の精度でタスクの正誤判定ができた．また，本システムを応用することで，操作ミスからの迅速な復帰が可能となる．さらに今後は，このシステムを日常のアプリケーションに組み込み，ユーザが操作ミスからスムーズに復帰できるようにすることで，操作効率とユーザ体験の向上に寄与することを目指す．

1. はじめに

ユーザインタフェースの設計において，操作ミスから迅速に復帰できる仕組みは，ユーザエクスペリエンスを向上させるために不可欠である．特に，現代の複雑なデジタル環境では，多くのデバイスやアプリケーションが多様な入力手段を必要としているため，操作ミスのリスクが高まっている．このため，ユーザが誤って操作した際でも簡単に修正できる UI の開発が求められている．従来の研究では，エラー防止のためのガイドライン [1], [2] や，ユーザが操作ミスしにくい UI デザインの提案 [3]–[6] が行われてきた．しかし，すべてのエラーを防ぐことは現実的ではなく，むしろエラー発生後にいかに迅速に回復できるかが重要な課題となっている．この点において，ユーザの操作履歴や行動をリアルタイムで把握し，適切な支援を提供するアプローチが有効とされている．

本研究では，スマートフォンを対象とし，9 軸センサ，視線，表情，タップ座標など，標準搭載されるセンサから取得可能なデータを活用して，ユーザの入力の正誤を判定し，操作ミスからの迅速な復帰を支援する UI の設計に焦点を当てる．図 1 にシステムの概要を示す．このシステムでは，リアルタイムで取得可能なデータを基に機械学習モデルを使用してユーザ操作の正誤を判定し，使用しているアプリケーションや機能に応じたフィードバックを提供することが目標である．このような UI により，単にエラーを検出するだけでなく，操作状況に応じてエラー回復プロセスを最適化し，意図した操作に素早く戻ることを支援することが可能になると考えられる．本論文では，センサデータを用いた正誤判定システムの設計とモデルの性能評価実験について報告する．

具体的には，機械学習ベースで正誤判定をするシステムを作成し，表情データの有無により正誤判定精度が向上するかどうかを調査した．その結果，タスクの種類にかかわらず，表情データがある場合に精度が高くなり，表情データが入力タスクの正誤判定に寄与していることが確認された．また，一般モデルでは精度が低かったものの，個人ごとのモデルを用いることで，平均 87.3%の精度，0.89 の AUC（Area Under the Curve）が得られた．これは，特に 9 軸センサのデータが個人差を多く含むことが影響していると考えられ，個別調整によるモデル精度の向上が示唆される結果であった．今後は，我々の発見を通じてユーザが操作ミスから容易に復帰できる UI の構築に向けた新たな知見を提供することを目指す．

2. 関連研究

ユーザが操作ミスをした際に迅速かつ効果的に回復できるインタフェースの設計は，HCI 分野における重要な課題である．特に，現代の複雑なインタフェース環境では，操作ミスが生じる可能性が高く，ユーザが誤った操作を行った場合にどのように復帰できるかが，ユーザエクスペリエ

[1] 慶應義塾大学
[2] LINE ヤフー株式会社
[3] 東京工科大学
[a)] kazukichi956@keio.jp
[b)] k-ikematsu@acm.org
[c)] r.t.isomoto@gmail.com
[d)] kkunihir@acm.org
[e)] sugiura@keio.jp

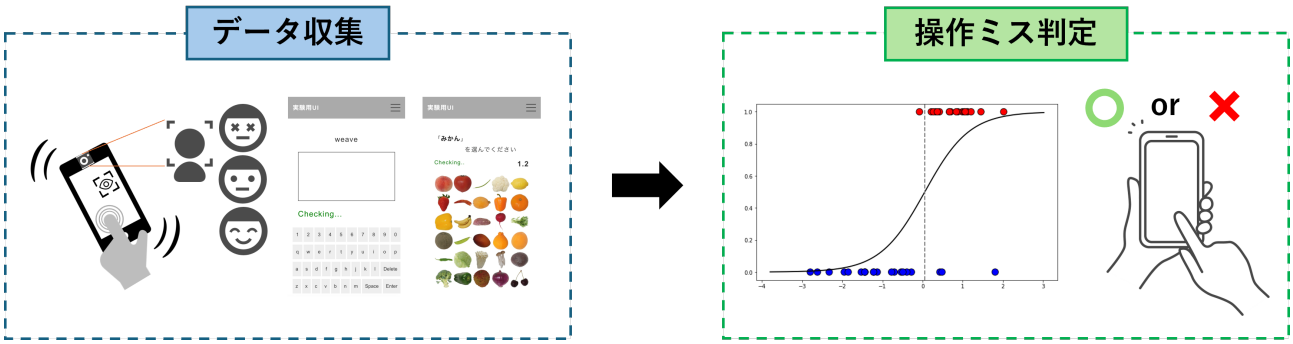

図 1 システム概要.

ンスを大きく左右する要因となっている．そのため，操作ミスからの復帰を支援するユーザインタフェースの設計に関する研究 [7] が進んでおり，多くのアプローチ [8], [9] が提案されている．以下に紹介する研究では，操作ミスの検知や，ユーザが素早く操作を修正できる仕組みを導入することによる，ユーザ負担の軽減が提案されている．

2.1 視線データとタッチ操作に基づく操作ミス検出

Sendhilnathan らの研究 [10] では，仮想現実環境における視線データを分析し，入力認識エラーやユーザエラーを検出する手法が提案されている．この研究では，VR の没入感が高いが故に，ユーザの注意や視線動態が重要な指標となることを指摘している．視線動態をリアルタイムで監視し，ユーザが操作ミスを行った際に特有の視線パターンが現れることを確認し，これをエラー検出に活用している．視線が止まる位置や移動の速度は，操作ミスのサインとして有効であることが実証された．この研究は，VR インタフェースにおいて迅速な操作ミス回復支援が可能であることを示唆しており，視線データの分析を他のインタフェースにも応用できる可能性を示している．さらに，Peacock らの研究 [11] においても，視線データが操作ミス検出の主要な指標となることが示されている．この研究では，タッチ操作や他の入力方法と組み合わせることで，より正確な操作ミス検出が可能であることを提案しており，特に視線の滞留時間や操作のズレに基づいてエラーを迅速に特定することができることを示している．近年は，視線追跡に基づく入力方法 [12] や意思決定 [13] や，入力認識エラーの検出手法に関する研究 [14] も進められており，視線情報を活用したインタラクションの最適化が期待される．

2.2 タッチスクリーンにおける強化学習を用いた操作精度向上手法

Lafreniere らの研究 [4] では，タッチスクリーン上での操作ミスを減らすために，強化学習を利用したターゲット選択精度向上の手法を提案している．ユーザの過去の操作データを学習し，特定のターゲットを選択する際の精度を向上させるアルゴリズムを開発しており，これにより特に操作ミスの発生が多い，細かいターゲットや複雑な操作においても，精度の高い選択が可能となる．強化学習によるアプローチは，ユーザの操作パターンに適応し，操作ミスを防ぐためのリアルタイムなフィードバックを可能にするが，新規ユーザや異なる操作環境での適応に関する課題も指摘されている．

2.3 表情認識を利用した操作ミス検出

Zhang ら [3] は，スマートスピーカーとの対話において，ユーザの表情を解析することで操作ミスや不適切な応答を中断する FrownOnError を提案している．この手法では，ユーザが困惑や不満を感じた際の表情変化に着目し，デバイスの応答を即座に調整する仕組みを実現している．特に，顔の表情はリアルタイムで検出でき，ユーザが意図しない応答をすぐに修正できる点で効果的である．ただし，表情認識の精度は環境要因や個人差に大きく依存するため，誤認識を防ぐためには他のデータと統合することが望ましい．

本研究では，これらの既存の研究の課題を克服するために，視線，タッチ，表情，9 軸センサデータを組み合わせた複合的なアプローチを提案している．各データソースが持つ個々の限界を補完し合うことで，より高精度な操作ミス検出が可能となる．この統合的アプローチは，操作ミスからの復帰を支援する UI デザインの新たな方向性を示すものであり，ユーザエクスペリエンスの向上に寄与すると期待される．

3. 提案手法

我々はスマートフォンを対象に，操作ミスから容易な復帰を目的とした，ユーザの操作ミスを操作決定後に検出する手法を提案する．図 2 に提案手法の処理概要を示す．提案手法では，システムがタッチ入力 (touch-up event) を検出すると，タッチイベント前後の 350ms (合計 700ms) のセンサデータから，二値分類モデルを用いて入力の正誤を推論する．操作ミスと判定された場合，システムはアプリケーションや機能に応じたフィードバック処理を行うことを想定している．なお，提案手法で扱う操作ミスは，タッチ入力により決定されるもの (具体的には，ソフトウエア

キーボードにおけるキー選択や，画像・リンクなどのターゲット選択操作を想定) のうち，ユーザの意図に合致しない操作を指す．

本手法は機械学習による予測を基に処理を行うため，誤った予測が発生することを考慮したシステム設計を行う必要がある．そのため，提案手法では即座に復帰処理を実行するのではなく，ユーザに対してレコメンドの形でフィードバックを提供する設計とする．例えば，ブラウザのリンク選択の場面であれば，選択されたページへの遷移後に，他の選択候補であったリンクをポップアップで表示するといったユーザが選択可能な形式でのフィードバックを想定している．これにより，偽陽性（実際にはユーザの意図に沿う操作をしたが，操作ミスとして検出された）の場合は，意図しない復帰処理の自動実行を回避することが可能である．なお，モデルの判定時に算出される信頼度スコア[*1]に対して閾値を設けることで，確度の高い判定の場合にのみレコメンドを提示する調整も可能である [15]．また，偽陰性（実際には操作ミスであったが，操作ミスとして検出されなかった）の場合はフィードバックを行わないため，従来の復帰操作と同様に，ユーザは手動操作を行う．この場合，提案手法は復帰プロセスの改善に寄与しないものの，操作性の劣化は生じないと考えられる．

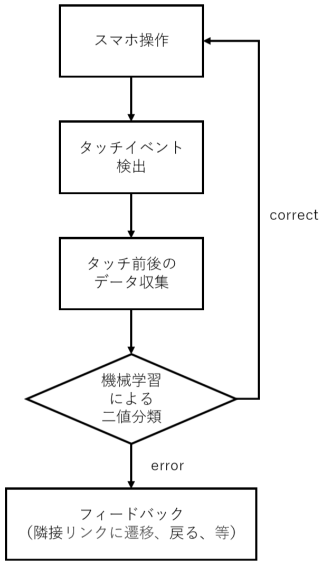

図 2 ミス検出の処理概要.

3.1 取得データ

本システムは，ユーザが入力タスクを行う際のセンサデータを取得し，それらを特徴量として用い，二値分類を行うことを目的としている．取得するデータには，以下の項目が含まれる：

加速度 ユーザがデバイスをどのように動かしているか，デバイスの移動速度の変化を検知する．デバイスを不安定に持っている場合や，素早く操作した場合の操作ミスを検知する．

角速度 デバイスの回転運動を測定する．デバイスの向きが急激に変わる場合，意図しない操作が行われやすいため，角速度データを使用して操作ミスのリスクを評価可能である．

地磁気 操作時にデバイスが意図しない方向に傾いていたり，ユーザが不適切な姿勢でデバイスを操作している状況を特定する．

タップ座標およびタップした要素の中心座標 タップ座標はユーザが画面上で行ったタッチ操作の位置を指し，要素の座標はユーザがタップした UI 要素の位置情報を指す．このデータは，ユーザがボタンのどれだけ中央付近をタップしたかという位置関係を示す．

表情 ユーザの表情状態を把握するために，表情認識技術を適用する．FrownOnError[3] を参考に，操作ミス時に起こるであろう表情変化を検知する．

視線座標 視線追跡技術により，ユーザが画面上でどこを見ているかを座標として取得する．これにより，ユーザの注意の方向が分析できる．

画面の注視度 ユーザが特定の UI 要素や画面上の領域に対してどの程度の注意を払っているかを測定する．ユーザの要素に対する集中度と操作ミスの関係を分析する．

データ収集には Apple iPhone 14 Pro を使用した．加速度，角速度，地磁気はスマートフォン内蔵のセンサを通じて取得した．視線座標および画面の注視度の取得は EyeDid SDK[*2]を使用した．表情認識には Python の FER（Face Emotion Recognizer）ライブラリ[*3]を用いた．

加速度，角速度，地磁気はそれぞれ x，y，および z の 3 軸に対する値で表現され，60 fps にて取得した．タップ座標は，ユーザが要素にタップした際の x および y のピクセル座標値で表現される．従って，本実験で用いた iPhone 14 Pro では，x 座標が 0 から 1179 pix，y 座標が 0 から 2556 pix の間となる．なお画面左上隅を原点とし，画面右方向，下方向がそれぞれ，x 軸，y 軸の正方向である．視線座標もタップと同様の座標で表現され，こちらも 60 fps にて取得した．表情は，FER から取得できる恐怖，驚き，中立（無表情），幸せ，悲しみ，怒り，および嫌悪の 7 つの表情を使用した．FER は 7 表情それぞれである確率を提示し，もっとも確率の高い表情をその時のユーザの表情として示す．

[*1] scikit-learn developers: LogisticRegression(2007-2025), https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

[*2] VisualCamp: Simple, Easy, EyeTracking SDK SEESO (2024). https://docs.eyedid.ai/

[*3] Foundation, P. S.: fer・PyPI (2023). https://pypi.org/project/fer/

3.2 データ処理

本節では，ユーザの表情やセンサデータを統合し，分析を行うための一連の処理について述べる．まず，記録したセンサデータのうち FER で取得した 7 つの表情を，3 つの特徴量に削減する．操作ミスが行われた時に値が上昇すると予想される「恐怖」，「驚き」，「悲しみ」，「怒り」，「嫌悪」を「ネガティブ」として統合し，「幸せ」を「ポジティブ」として扱い，「ネガティブ」，「ポジティブ」，および「ニュートラル」の 3 つを特徴量として使用した．また，その後の分析や統計処理がスムーズに行われるように数値データに変換するプロセスも実行され，Python の pandas[*4] ライブラリを用いて各変数を適切なデータ型に変換している．計測しているタイムスタンプに基づいて，操作ミスや正しい操作が行われた時刻のデータと，その前後のセンサデータが抽出される．この段階で，ウィンドウサイズ（時間幅）を設定し，そのウィンドウ内の 9 軸センサの値や視線座標，注視度，表情ラベルの最大値と最小値の差分が新たな特徴量として生成される．これにより，エラー発生前後の動態的な特徴が捉えられる．

以下に，本研究において使用した 28 個の特徴量を示す．

- タップイベントが発生した時刻におけるデータを使用した特徴量（計 16 個）
  - タップ座標 / Tap (x, y)
  - 要素の座標 / Element (x, y)
  - 視線位置 / Gaze (x, y)
  - 注視度 / Attention
  - 加速度 / Accel (x, y, z)
  - 角速度 / Gyro (x, y, z)
  - 地磁気 / Mag (x, y, z)
- ウィンドウサイズ内のデータに対して最大値と最小値の差分をとった特徴量（計 12 個）
  - 視線位置 / Gaze_max_min_diff (x, y)
  - 注視度 / Attention_max_min_diff
  - 加速度 / Accel_max_min_diff (x, y, z)
  - 角速度 / Gyro_max_min_diff (x, y, z)
  - 地磁気 / Mag_max_min_diff (x, y, z)

3.3 解析

収集したセンサデータを元に計算した特徴量を用いて，ロジスティック回帰モデルによる二値分類を行う．また，5 分割交差検証を実施し，モデルの汎用性を評価する．この手法により，データセットを 5 つの部分に分割し，各部分を検証用データとして使用することで，モデルの過学習を防ぐ．本研究では，入力タスクと画像選択タスクを行い，これらのタスクそれぞれについて，leave-one-user-out 交差検証による一般モデルと，各個人のデータのみを用いた 5 分割交差検証による個人モデルの評価を行う．加えて，FER ライブラリによる表情情報がモデルの性能にどの程度寄与するかを確かめるため，「表情あり」と「表情なし」それぞれの条件にて解析を行った．

[*4] NumFOCUS, I.: pandas documentation - pandas 2.2.3 documentation (2024). https://pandas.pydata.org/docs/index.html

4. 実験

4.1 実験手順

本実験は，参加者が iPhone 14 Pro にインストールされたアプリケーションを用いて，特定のタスクを実行する形式で進められる．最初に，参加者に対して実験の目的や進行方法について説明し，実験参加が任意であること，また参加を辞退する権利があることを明示する．この説明が終わると，参加者はアプリケーションを起動し，タスク画面にアクセスする．タスクは 2 種類あり，文字入力タスクと画像選択タスクから構成される．まず，文字入力タスクが実施され，参加者は図 3(a) のように指定された英単語の入力を行う．次に，図 3(b) のような画像選択タスクが実施され，参加者は 2 秒の制限時間内に画像を選択する．この制限時間は，参加者の操作ミスを誘発するために設定している．実験中はアプリケーションを通じて 3.1 節に記述したデータの取得を行った．本実験を通じ，入力タスクでは，783 (試行) × 10 (参加者) = 7830 試行のデータを，画像選択タスクでは，117 (試行) × 10 (参加者) = 1170 試行のデータを収集した．参加者はタスク間に適宜休憩をとった．なお，本実験の実施について，著者の所属する研究機関である慶應義塾大学の研究倫理委員会から番号：2024-094 として承認を得ている．

4.2 参加者

本実験は，21–23 歳の男女の健常者を 10 名（P1–P10）

(a) (b)

図 3 (a) 文字入力タスク．画面上部に表示される文字列（この例では「weave」）を入力する．(b) 画像選択タスク．5×6 に配置された画像から指示された画像を選択する．

募った．参加者は，18 歳以上を対象とし，また健康状態に関する基準が設けられ，視覚に障がいのある場合や，自身の身体データを一般公開することに抵抗がある場合は除外された．

5. 結果

収集したデータにタッチ入力 (touch-up event) は合計で 9,002 件含まれ，そのうち操作ミス (タスク指示と合致しない入力) のデータは 1,099 件であった．入力タスクと画像選択タスクにおいて，一般モデルと個人モデル，そして表情データの有無それぞれについて，実験データをロジスティック回帰により解析した結果の概要を表 1 および表 2 に示す．個人モデルにおいては，一定の精度および AUC が達成されている一方，一般モデルではチャンスレートとほとんど変わらない結果が示された．従って，後続する 5.1 節および 5.2 節では，個人モデルの結果の詳細を報告する．

表 1 入力タスクの結果概要.

モデル	表情データ	平均 Acc.(SD)	平均 AUC (SD)
一般モデル	なし	56.6% (0.040)	0.60 (0.022)
一般モデル	あり	59.8% (0.045)	0.64 (0.044)
個人モデル	なし	79.0% (0.040)	0.83 (0.055)
個人モデル	あり	88.2% (0.055)	0.85 (0.055)

表 2 画像選択タスクの結果概要.

モデル	表情データ	平均 Acc.(SD)	平均 AUC (SD)
一般モデル	なし	48.5% (0.030)	0.49 (0.073)
一般モデル	あり	54.4% (0.046)	0.56 (0.092)
個人モデル	なし	89.9% (0.108)	0.92 (0.078)
個人モデル	あり	92.1% (0.090)	0.95 (0.061)

表 3 文字入力タスクの個人モデル（表情なし）の解析結果.

	Precision	Recall	F1 Score	Accuracy	AUC
P1	0.77	0.77	0.77	0.77	0.82
P2	0.84	0.83	0.83	0.83	0.83
P3	0.89	0.86	0.85	0.86	0.89
P4	0.74	0.74	0.74	0.74	0.81
P5	0.74	0.74	0.74	0.74	0.75
P6	0.80	0.79	0.79	0.79	0.74
P7	0.82	0.81	0.81	0.81	0.84
P8	0.75	0.75	0.75	0.75	0.81
P9	0.80	0.80	0.80	0.80	0.91
P10	0.82	0.81	0.81	0.81	0.87

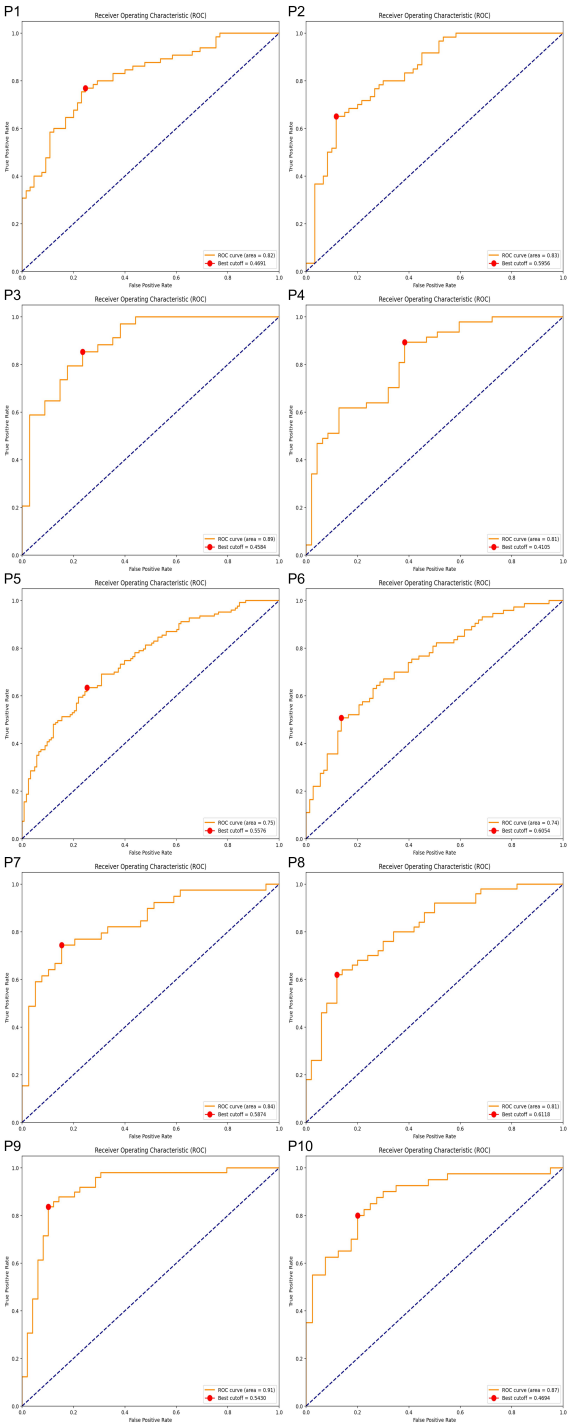

図 4 文字入力タスクの個人モデル（表情なし）の ROC 曲線.

5.1 文字入力タスク

表 3 および図 4 に，文字入力タスクにおける各ユーザに対する個人モデル（表情なし）の性能を示す．精度（Accuracy）は，全体として 0.74 から 0.86 の範囲で安定した値を示しており，AUC スコアも 0.74 から 0.91 と比較的高い結果を示している．また，Precision や Recall，F1 スコアもユーザごとに一貫して高い値であった．特に，参加者 P3 および P9 においては AUC スコアが 0.89 以上と高く，正誤の識別性能が優れていることが確認できる．

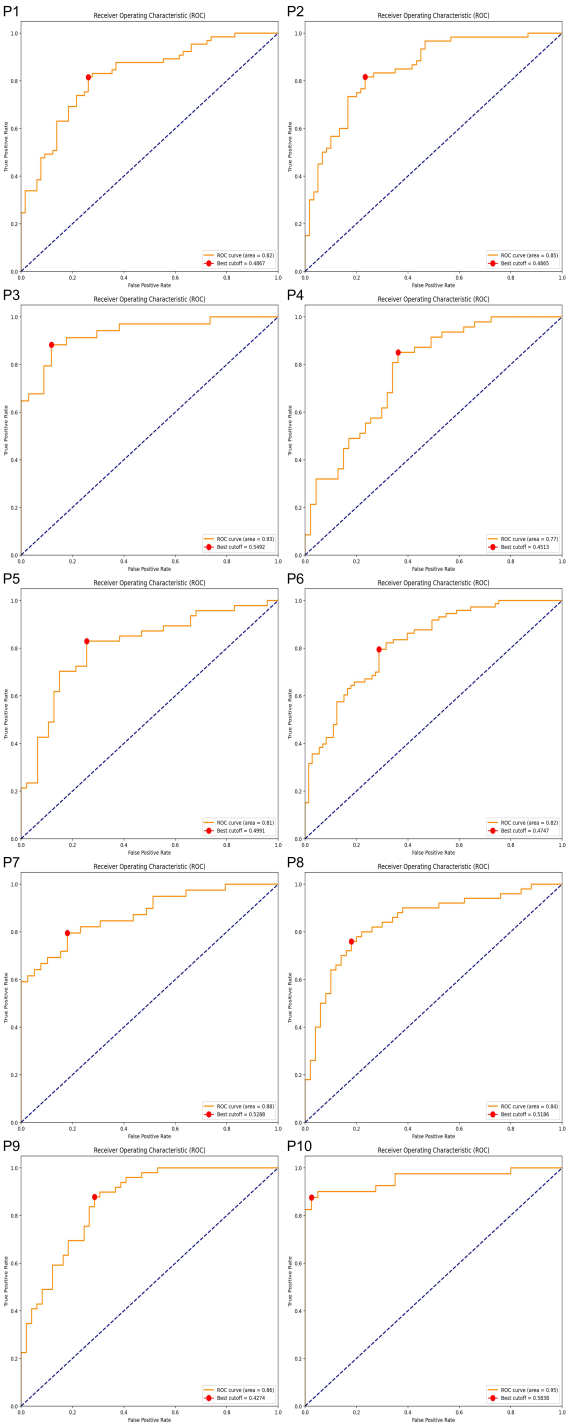

図 5 文字入力タスクの個人モデル（表情あり）の ROC 曲線.

次に，表 4 および図 5 に表情データを含む個人モデルの性能を示す．精度は全体的に 0.79 から 1.00 の範囲で，特に P3 においてはすべての指標が 1.00 と最高値を示している．また，P3，P6，P8，P9 の 4 人で，0.9 以上と安定して高い精度を示している．AUC スコアも 0.77 から 0.95 の範囲で推移し，各ユーザごとに高い識別性能を示している．表情ありのモデルは表情なしのモデルと比較し，平均精度は 9.2%高く，他の指標においても高い性能であった．

また，図 6 に入力タスクにおける表情データを含む個人モデルの特徴量とその寄与率について示す．文字入力タスクにおいて，視線位置や地磁気，角速度などの寄与率が高くなっていた．表情については，差分をとることで，処理を加える前に比べて寄与率が向上した．

表 4 文字入力タスクの個人モデル（表情あり）の解析結果.

	Precision	Recall	F1 Score	Accuracy	AUC
P1	0.88	0.88	0.88	0.88	0.82
P2	0.90	0.88	0.87	0.88	0.85
P3	1.00	1.00	1.00	1.00	0.93
P4	0.79	0.79	0.79	0.79	0.77
P5	0.82	0.82	0.82	0.82	0.81
P6	0.90	0.90	0.90	0.90	0.82
P7	0.87	0.87	0.87	0.87	0.88
P8	0.92	0.90	0.90	0.90	0.84
P9	0.92	0.90	0.90	0.90	0.86
P10	0.88	0.88	0.88	0.88	0.95

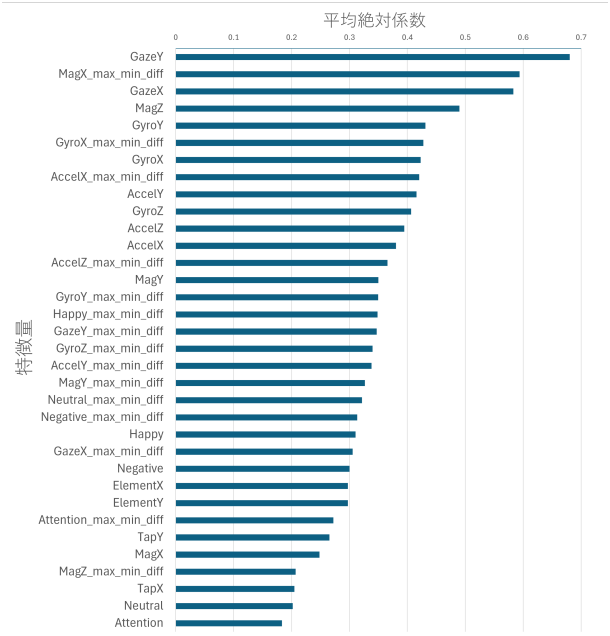

図 6 文字入力タスク個人モデル（表情あり）の特徴量と平均寄与率.

5.2 画像選択タスク

表 5 および図 7 に画像選択タスクにおける表情データを用いない個人モデルの性能を示す．精度は 0.76 から 1.00 の範囲であり，参加者の半数（P4，P5，P6，P8，P10）ではすべての指標が 1.00 を達成している．特に，P4 と P5 は，Precision，Recall，F1 スコアのすべてが 1.00 であることから，非常に高い識別性能を示している．また，AUC スコアは全体的に高く，0.78 から 1.00 の範囲で推移し，特に P4，P5，P6，P8 では 1.00 を記録していることから，全体的に高い性能が確認できる．P1，P2，P3，P7，および P9 も高い精度を示しているが，他のユーザと比較するとやや劣っている．

次に，表 6 および図 8 に表情データを用いた個人モデル

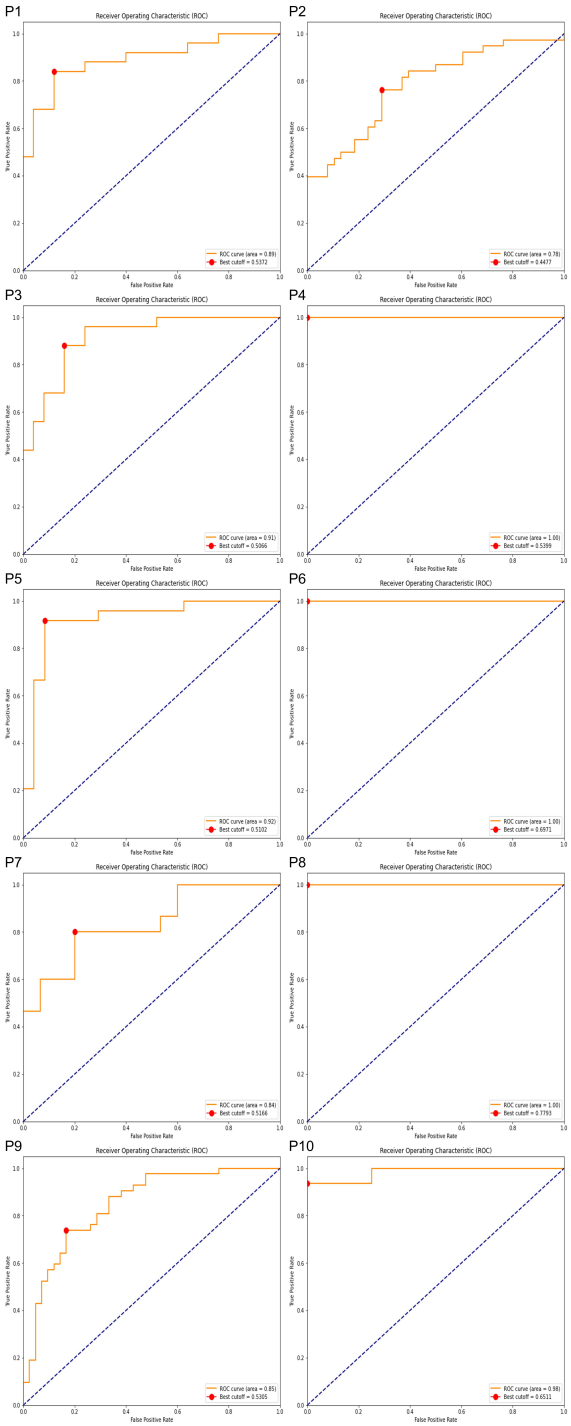

図 7 画像選択タスクの個人モデル（表情なし）の ROC 曲線.

の性能を示す．精度は全ての被験者で高く，P4，P6，P7，P8，P10 はすべての指標で 1.00 を達成しており，特に高い識別性能を示している．P1 と P2 も高い精度を示すが，他の被験者に比べるとやや劣る結果となっている．AUC も全体的に高く，非常に信頼性の高いモデルであることが確認できる．表情ありのモデルは表情なしのモデルと比較し，平均精度は 2.2%高く，他の指標においても高い性能であった．

また，図 9 に画像選択タスクにおける表情データを含む個人モデルの特徴量とその寄与率について示す．画像選択タスクにおいては，注視度や角速度，加速度などの寄与率が高くなっている．表情については，文字入力タスクと比較して寄与率が高く，またどちらのタスクにおいても，喜びの表情が特に寄与している．

6. 議論

6.1 表情特徴量の利用

文字入力タスクおよび画像選択タスクにおいて，表情特徴量の有無による精度の違いが現れた．特に，表情の特徴量を含めたモデルでは，両タスクにおいて，精度，AUC 共に高くなる傾向が確認された．これは，表情が操作ミスやユーザの状態を大きく反映おり，操作状況に応じた表情変化を適切に捉えることが，操作ミスの検出精度を向上させる鍵となることを示している．しかしながら，本研究における「ポジティブ」，「ネガティブ」，「ニュートラル」の 3 分類が最適であるかどうかは確かめきれていない．操作ミス時に「恐怖」，「驚き」，「悲しみ」，「怒り」，「嫌悪」が上昇するだろうと考え「ネガティブ」として統合したが，これは個人によっても差が生じると考えられる．

また，表情ではなく，より直接的な顔の特徴点に基づく分析も有効であると考える．特に，FrownOnError[3] などの先行研究が示すように，顔のランドマークやアクションユニットなどの情報を活用し，眉や口角といった特定の部位の動きを捉えることで，操作ミスの検出精度が向上する

表 5 画像選択タスクの個人モデル（表情なし）の解析結果.

	Precision	Recall	F1 Score	Accuracy	AUC
P1	0.86	0.80	0.79	0.80	0.89
P2	0.81	0.80	0.80	0.80	0.78
P3	0.86	0.80	0.79	0.80	0.91
P4	1.00	1.00	1.00	1.00	1.00
P5	1.00	1.00	1.00	1.00	0.92
P6	1.00	1.00	1.00	1.00	1.00
P7	0.88	0.83	0.83	0.83	0.84
P8	1.00	1.00	1.00	1.00	1.00
P9	0.78	0.76	0.76	0.76	0.85
P10	1.00	1.00	1.00	1.00	0.98

表 6 画像選択タスクの個人モデル（表情あり）解析結果.

	Precision	Recall	F1 Score	Accuracy	AUC
P1	0.86	0.80	0.79	0.80	0.96
P2	0.81	0.80	0.80	0.80	0.88
P3	0.92	0.90	0.90	0.90	0.89
P4	1.00	1.00	1.00	1.00	1.00
P5	0.92	0.90	0.90	0.90	0.92
P6	1.00	1.00	1.00	1.00	1.00
P7	1.00	1.00	1.00	1.00	1.00
P8	1.00	1.00	1.00	1.00	1.00
P9	0.82	0.81	0.81	0.81	0.84
P10	1.00	1.00	1.00	1.00	0.99

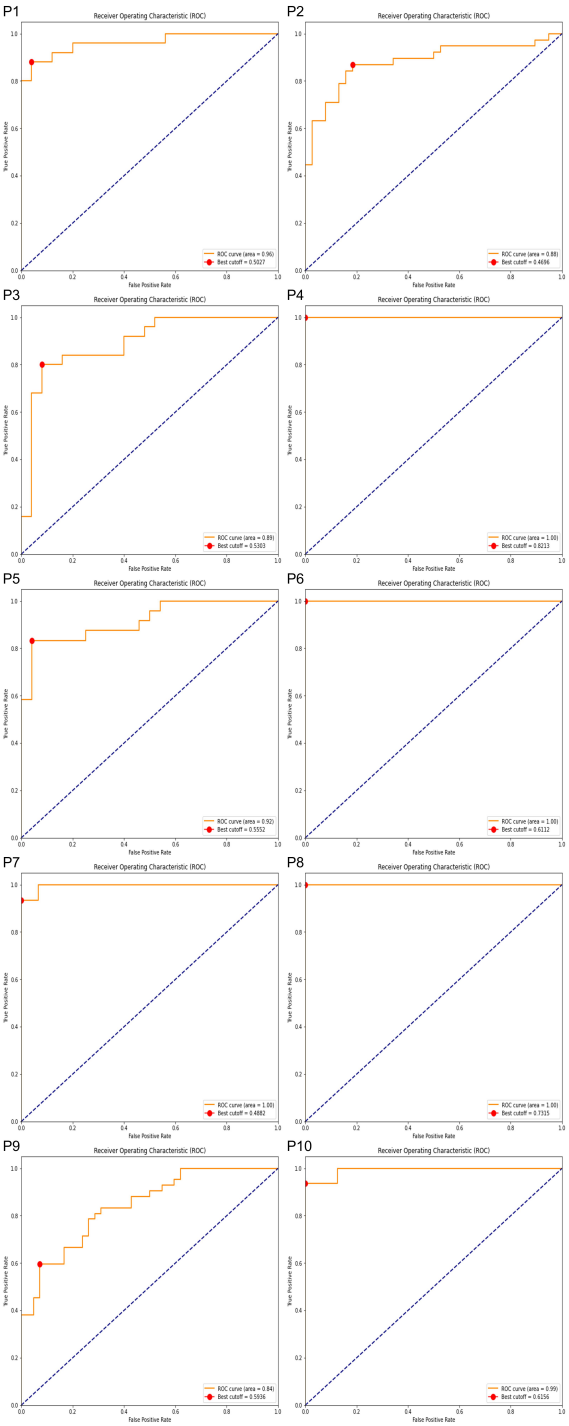

図 8 画像選択タスクの個人モデル（表情あり）ROC 曲線.

可能性がある．感情の総体的な変動を分析するのではなく，特定の操作行動や操作ミスに関連する表情パターンをより詳細に検出できると考える．ランドマーク情報から，操作中の微細な表情変化，特に眉や目元，口角のわずかな変化をリアルタイムで把握することにより，ユーザが意図しない操作を行ったタイミングを特定しやすくなる．これにより，迅速なフィードバックやカスタマイズされたサポートの提供が可能となり，操作ミスからの迅速な復帰を促す UI 設計において有効な手段となるだろう．このように，特定

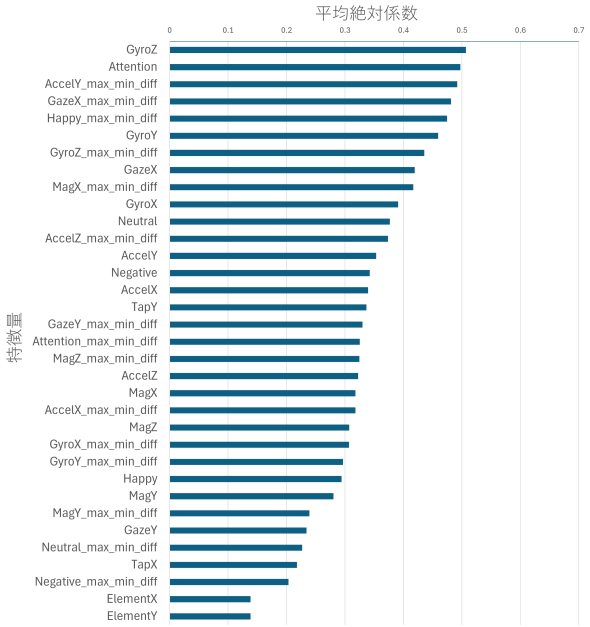

図 9 画像選択タスク個人モデル（表情あり）の特徴量と平均寄与率.

のランドマークに基づいた表情データの分析は，ユーザ体験の向上に貢献する可能性があり，UI 設計において重要な方向性といえる．

6.2 結果の信頼性

本実験では，被験者に対して実験の目的や進行方法を事前に説明しており，被験者が操作ミスに対する反応を意識的に調整する可能性が考えられる．この点について，実験結果への影響を検討する必要がある．しかし，本研究で使用したデータは，タッチ操作の前後 350ms のセンサデータに限定しており，この短い時間枠では意図的な表情変化の抑制が反映される可能性は低いと考えられる．特に，表情データとして使用した特徴量は，瞬間的かつ無意識的な反応を主に捉えるものであり，被験者がミスを意識的に隠そうとした場合でも，その影響は限定的であると判断される．

6.3 ユースケース

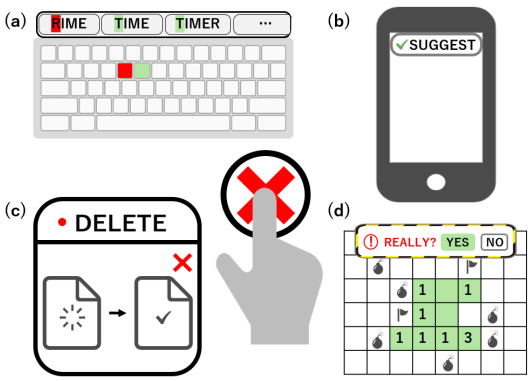

図 10 ユースケース.

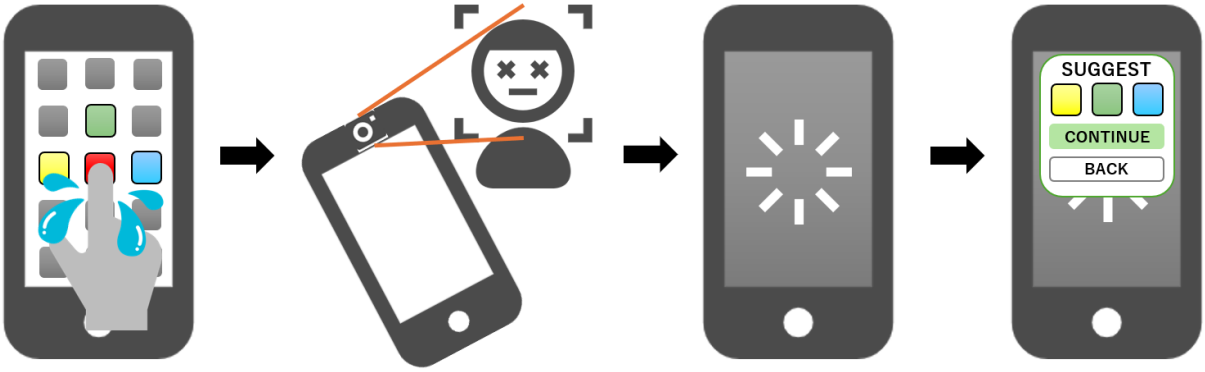

図 11 フィードバックシステム.

文字の入力時に操作ミスが発生した場合，単語を入力し終えた後にミス箇所までキャレットを移動して修正する必要がある．本システムを利用することで，図 10 (a) のように，サジェスト一覧にタップされたキーの周辺キーを考慮した候補を表示するなど，ミスを前提とした補助が可能になると考えられる．

画像選択タスクにおいては，複数の使用例が想定される．例えば，Web ブラウザでのリンク選択やホーム画面でのアイコン選択である．Web ブラウザでは，誤操作により意図しない画面遷移が発生すると，通信や読み込みによる遅延が生じる可能性がある．このような場合，図 10 (b) に示す通り，本システムがミスを検知し，他の候補リンクやアイコンを提示することで，より快適な操作が実現できるだろう．

また，ファイルアップロード時のファイルビューアにおいて，誤って選択した場合に取り消し操作が間に合わないことがある．図 10 (c) のように，仮にアップロードが完了してしまっても，誰かに閲覧される前に迅速に削除する仕組みを提供することが求められる．

他にも，意図しない入力を識別できるようになれば，マインスイーパのようなゲームにおいても，図 10 (d) のように，誤タップによるゲームオーバーを回避することができるようになると考える．

上記のように，様々な場面において図 11 のようなシステムを導入することで，これまで待つことしかできなかったローディング時間を使用し，多様な操作ミスへのフィードバックが可能となると考えられる．

ただし，画像選択タスクは実際の使用状況とは異なっているため，本研究の結果が直接的に適用可能であるとは限らない．画像選択タスクの汎用性については，今後，自由にブラウジングをしてもらうなどのより現実的なタスクをテストデータとして使用することで検証していく予定である．

6.4 一般モデルの精度向上

一般モデルの精度が低かった点については，9 軸センサのデータに大きな個人差が存在することが要因の一つと考えられる．センサデータが個人の動きや環境によって異なるため，一般モデルでは十分な汎用性を持たせることが難しかったと考察される．したがって，より精度の高い一般モデルを構築するためには，より汎用的な特徴量の選定が課題であると考える．特に，個人差が大きいデータを基にしたモデルでは，特徴量が個人に特化してしまうため，より一般的に適用可能な特徴量の探索が必要である．また，今回はロジスティック回帰モデルを使用したが，今後は異なるアルゴリズムでの調査も行い，それぞれのアルゴリズムがどのような特徴量や条件下で優れているかを検討していく．また，異なるアルゴリズムでの結果を比較することで，今回の手法の限界や改善点を明確にし，モデルのさらなる最適化を図る．

6.5 今後の展望

今後，実際に使用するにあたっては，正確な正誤判定を行うためのキャリブレーションが必要である．個人ごとに異なるデータ特性に対応するため，最適なキャリブレーションを導入することで，モデルの精度向上が期待される．具体的には，利用者が日常的に収集するデータ等を用いて個人モデルを構築する手法が有効であると考えられる．データの収集方法としては，日常的なスマートフォン操作における「戻るボタン」や「バックスペースキー」操作の直前の入力を，操作ミスとしてデータを収集することが有効であると考える．長期的なデータ収集により，個々のユーザの動作や表情のパターンを捉え，それに基づいたモデルの最適化を行うことで，より高精度のモデルが構築できると考える．また，より広範なユーザ層に適用できるよう，データの取得方法や特徴量の一般化を目指した研究も進めていくべきである．たとえば，個人差の影響を減らすために，センサデータに対して標準化手法や補正を行い，共通の特徴量を抽出するアプローチが考えられる．さらに，ユーザご

とに異なる表情の反応や行動傾向を考慮し，個別の特徴量モデルを構築することで，各ユーザのセンサデータに基づいた操作ミス検出を最適化できると考えられる．したがって，キャリブレーションの手法を検討し，個人の表情の変化に対応するシステムを実現していく．これにより，一般モデルの精度向上と汎用性の高いモデルの実現を試みる．

7. 結論

本研究では，9 軸センサ，表情，視線，タップ座標などのデータを活用し，ユーザの入力タスクにおける誤認識や操作ミスを検出する方法を探求した．その結果，表情を考慮することで，入力および画像選択タスクの精度が向上することが確認された．一方で，一般モデルの精度が低かったのは，9 軸センサのデータに個人差が大きく影響していると考えられる．今後は，機械学習技術を用いた個別モデルの構築を進め，表情の変化に応じたインタフェースの最適化と同時に，特徴量の一般化による一般モデルの精度向上を目指していく．

参考文献

[1] Yamanaka, S., Shimono, H. and Miyashita, H.: Towards More Practical Spacing for Smartphone Touch GUI Objects Accompanied by Distractors, *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, ISS '19, New York, NY, USA, Association for Computing Machinery, p. 157–169 (online), DOI: 10.1145/3343055.3359698 (2019).

[2] Usuba, H., Yamanaka, S. and Miyashita, H.: Touch Pointing Performance for Uncertain Touchable Sizes of 1D Targets, *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '19, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/3338286.3340131 (2019).

[3] Yan, Y., Yu, C., Zheng, W., Tang, R., Xu, X. and Shi, Y.: Frownonerror: Interrupting responses from smart speakers by facial expressions, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14 (2020).

[4] Li, Z., Zhao, M., Das, D., Zhao, H., Ma, Y., Liu, W., Beaudouin-Lafon, M., Wang, F., Ramakrishnan, I. and Bi, X.: Select or suggest? Reinforcement learning-based method for high-accuracy target selection on touchscreens, *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–15 (2022).

[5] Bi, X. and Zhai, S.: Bayesian touch: a statistical criterion of target selection with finger touch, *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, New York, NY, USA, Association for Computing Machinery, p. 51–60 (online), DOI: 10.1145/2501988.2502058 (2013).

[6] Yu, D., Desai, R., Zhang, T., Benko, H., Jonker, T. R. and Gupta, A.: Optimizing the Timing of Intelligent Suggestion in Virtual Reality, *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/3526113.3545632 (2022).

[7] Xu, X., Yu, C., Wang, Y. and Shi, Y.: Recognizing Unintentional Touch on Interactive Tabletop, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 4, No. 1 (online), DOI: 10.1145/3381011 (2020).

[8] Jung, K. and Jang, J.: Development of a two-step touch method for website navigation on smartphones, *Applied ergonomics*, Vol. 48, pp. 148–153 (2015).

[9] Huang, Z., Gao, C., Wang, H., Deng, X., Lai, Y.-K., Ma, C., Qin, S.-f., Liu, Y.-J. and Wang, H.: SpeciFingers: Finger Identification and Error Correction on Capacitive Touchscreens, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 8, No. 1, pp. 1–28 (2024).

[10] Sendhilnathan, N., Zhang, T., Lafreniere, B., Grossman, T. and Jonker, T. R.: Detecting Input Recognition Errors and User Errors using Gaze Dynamics in Virtual Reality, *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/3526113.3545628 (2022).

[11] Peacock, C. E., Lafreniere, B., Zhang, T., Santosa, S., Benko, H. and Jonker, T. R.: Gaze as an Indicator of Input Recognition Errors, *Proc. ACM Hum.-Comput. Interact.*, Vol. 6, No. ETRA (online), DOI: 10.1145/3530883 (2022).

[12] Namnakani, O.: Gaze-based Interaction on Handheld Mobile Devices, *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, ETRA '23, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/3588015.3589540 (2023).

[13] Vazquez-Alvarez, Y. and Brewster, S.: Investigating background & foreground interactions using spatial audio cues, *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, New York, NY, USA, Association for Computing Machinery, p. 3823–3828 (online), DOI: 10.1145/1520340.1520578 (2009).

[14] Ratwani, R. M., McCurry, J. M. and Trafton, J. G.: Predicting postcompletion errors using eye movements, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, New York, NY, USA, Association for Computing Machinery, p. 539–542 (online), DOI: 10.1145/1357054.1357141 (2008).

[15] Nishida, N., Ikematsu, K., Sato, J., Yamanaka, S. and Tsubouchi, K.: Single-tap latency reduction with single- or double- tap prediction, *Proc. ACM Hum. Comput. Interact.*, Vol. 7, No. MHCI, pp. 1–26 (online), DOI: 10.1145/3604271 (2023).