

凝視後にジェスチャを行うという一連の操作を用いた意図しない操作に堅牢な視線に基づく操作手法

磯本 俊弥*¹ 山中 祥太*² 志築 文太郎*¹

Gaze-based Interaction Technique Robust Against Unintentional Activation using Dwell-then-Gesture

Toshiya Isomoto*¹, Shota Yamanaka*² and Buntarou Shizuki*¹

Abstract – We demonstrate a gaze-based command activation technique that is robust against unintentional command activations using a series of a dwelling and a gesture. We adopted a simple two-level stroke, which consists of a sequence of two orthogonal strokes as a gesture. To achieve robustness against unintentional command activations, we designed and fine-tuned a gesture detection system based on how users move their gaze, as revealed through three experiments.

Keywords : Gaze-based interaction, gaze gesture, eye tracking, gaze movement

1. はじめに

視線を用いてコンピュータを操作する際の課題として意図しない操作が生じることが挙げられる。その1つとして凝視に基づく操作手法におけるミダタッチ問題^[1]が知られている。これは、ある対象へ視線を移した際にユーザの意図に関わらず、その対象を選択するためにユーザは凝視をしていると認識され、選択が行われるという問題である。また、ある特定の視線移動のパターン（以降、ジェスチャ）に基づく操作手法においても、意図しない操作は生じる。例えば、左から右に視線を移動することをジェスチャとした場合、これに紐付けられたコマンドが文章を読んでいる最中に意図せず実行される。

本研究の目的は、意図しない操作に堅牢な視線に基づく操作手法の実現である。意図しない操作は、凝視時間を長くする、あるいは複雑なジェスチャを用いるという工夫により防げるが、これらには操作に要する時間が増える、あるいは操作が難しくなるという問題がある。そこで本研究では、短時間の凝視および「右を見た後に上を見る」などの2段階の視線移動からなるジェスチャによる操作手法を示す。

本手法において、ユーザは凝視後に2段階の視線移動を行うことにより操作を行う(図1)。凝視により対象の選択が可能であり、2段階の視線移動を行うことによりそのジェスチャに紐付けられたコマンドを実行できる。例えば、凝視が認識された点があるアイコンの中であれば、そのアイコンを対象としてコマ

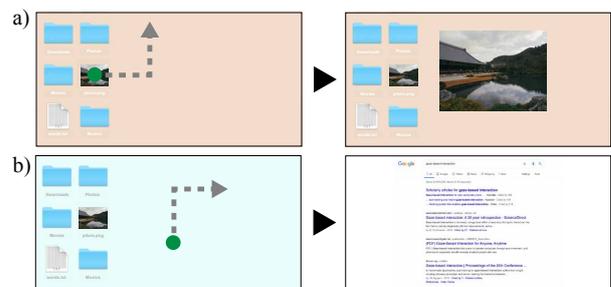


図1 本手法を用いた操作方法。
Fig.1 Manipulations using our technique.

ンドが実行(図1ではファイルを開く)され、外であれば、画面全体を対象としたコマンドが実行(図1では表示画面の変更)される。

以降、本論文では、凝視後にジェスチャを行うという一連の操作を認識するために設計したシステムについて述べる。システム設計は3つの実験を通じて行う。実験1ではシステムの設計を行い、実験2および3では設計したシステムを改良する。その後、改良後のシステムの性能を示す。

なお、本論文は^[2]を和訳し、再構築・修正したものである。

2. 関連研究

本研究に関連する研究を述べた後に、これらの研究に対する本研究の位置づけを示す。

2.1 凝視に基づく操作手法

凝視によるコマンド実行は、少なくとも2度の凝視を要する。例えば、プルダウンメニューやポップアップメニューを用いる場合、ユーザは1度目の凝視にてメニューを表示し、2度目の凝視にて表示された

*1: 筑波大学
*2: ヤフー株式会社
*1: University of Tsukuba
*2: Yahoo Japan Corporation

メニューの中から任意のコマンドを実行する。しかし、ミダスタッチ問題により、意図せず画面にメニューが表示されると、メニューにより画面表示が隠れるという問題が生じる。また、例えばメニュー項目を確認している際にミダスタッチ問題が生じると、誤ったコマンドが実行される。これらを解決するために、ミダスタッチ問題が生じにくいとされている凝視に基づく操作手法^{[3]-[5]}を用いることが考えられるが、これらの手法には制限もある。例えば、文字入力時に次に入力される文字の予測を用いた手法^[3]において、ミダスタッチ問題の軽減は文字入力に制限されている。文字入力に制限されていない手法^{[4],[5]}においても、メニュー項目を探索する場合など対象の内容がユーザにとって未知の場合に適切に動作するかまでは明らかにされていない。加えて、ある程度の視線移動を要する操作手法^[5]では、現在の視線座標から近い位置にある対象への選択時には依然としてミダスタッチ問題が生じる。加えて、凝視に基づく操作手法と比べてジェスチャに基づく操作手法の方がコマンド実行に要する時間および成功率ともに良いことがわかっている^{[6],[7]}。

2.2 ジェスチャに基づく操作手法

ジェスチャによるコマンド実行は、実行したいコマンドに紐付いたジェスチャを行うことにより可能である。このジェスチャとしては、例えば、左から右に視線を移動するような1段階の視線移動^[8]や、1段階の視線移動を複数回組み合わせた多段階の視線移動^[9]が用いられてきた。意図しない操作の発生を防ぐという点においては多段階の視線移動が優れているが、ジェスチャの複雑さは操作の簡単さとトレードオフの関係にある。さらに、多段階の視線移動は、視線移動の難しさから多くの手法(例:^{[10]-[12]})においてガイド(視線移動の目安になるような経路や領域など)が表示されている。このガイドの表示方法は、常に表示する方法^[10]、半透明に常に表示する方法^[11]、凝視を認識した後に表示する方法^[12]に分けられる。

特に凝視を認識した後にガイドを表示するような、凝視とジェスチャを組み合わせた手法では、ユーザは凝視を行った際にコマンドを実行したい対象の選択も同時に行える(これをジェスチャに基づく操作手法のみで行う場合、対象の選択およびコマンドの実行のために2度のジェスチャを行う必要がある)。例えば、パイメニューを用いた手法^[13]では、ユーザはある対象へ短時間の凝視をした後に表示されるパイメニュー上での視線移動により、対象に任意のコマンドを実行できる。また、ある対象へ短時間の凝視をした後に、予め表示されているメニューへ視線を移すことにより操作可能な手法^[14]もある。これらのような手法では、短時間の凝視および1段階の視線移動を用いているため、意図しない対象の選択、ガイドの表示、もしくは

コマンド実行が生じることがある。対して、長い時間の凝視と円滑性追跡眼球運動(例:^[15])に基づく操作を組み合わせた手法^[12]も研究されている。この研究では、ミダスタッチ問題による頻繁なガイド表示(円滑性追跡眼球運動を発生させるためのオブジェクト表示)を防ぐために2sの凝視時間を用いている。

2.3 本研究の位置づけ

本手法では、短時間の凝視後に2段階の視線移動が認識された際にコマンドが実行される。また、これまでの研究において別々の操作として認識されていた凝視およびジェスチャを、凝視後にジェスチャを行うという一連の操作として認識することにより意図しない操作の発生を防ぐ。凝視もしくはジェスチャのみが意図せず認識されたとしても、操作は行われたいため、短い凝視時間およびシンプルなジェスチャを用いることができる。

これまでの凝視とジェスチャを組合せた手法^[12]においては、ミダスタッチ問題によって意図せずガイドが表示される可能性がある。一方、コマンドが紐付いているジェスチャをユーザが確認できるという利点もある。対して、本研究ではシンプルなジェスチャを用いることにより、ガイドの表示を行わずにコマンドを実行できるという利点があるが、コマンドが紐付いているジェスチャを覚える必要がある。ただし、対象の選択に短時間の凝視を用いるため、長時間の凝視(例えば^[12]と同様に2sなど)が認識された際にガイドを表示するように拡張できる。このようにユーザがガイドを必要な際にのみ表示を行うことは、マーキングメニュー^[16]と同様であるが、これまでの視線に基づく操作手法では行われてきていない。そのため、本研究は視線に基づく操作にマーキングメニューを応用するための基礎となる研究ともいえる。

3. 実験1

実験1の目的は、1段階および2段階の視線移動を行う際の視線情報(視線座標およびそのタイムスタンプ)を収集し、どのように視線が移動しているかを調査し、その特徴を見出すことである。

3.1 実験参加者

実験参加者は16名(女性2名)であり、20-24歳(平均22.0歳)であった。6名が眼鏡、2名がコンタクトを着用していた。11名が以前に視線認識機器を用いた実験に参加したことがあったが、いずれも日常的には使用していなかった。

3.2 実験機器および実験環境

視線認識機器としてTobii EyeX(サンプリングレート:60Hz)をディスプレイ下部に取り付け使用した。ディスプレイには、画面における背景などの反射を防ぐためにノングレアタイプの画面(DELL社製G2410。

画面サイズ：24 インチ（530×299 mm），ディスプレイサイズ：570×375 mm，解像度：1,920×1,080 ピクセル）を用いた．実験に関係ない物体が実験参加者の視界に入らないように，ディスプレイを無地の壁の前に配置した．また，実験中の実験参加者の動きによる視線認識精度の低下を防ぐために，頭部固定機（株式会社ナモト貿易，TKD-UK1）を用いて実験参加者の頭部位置を固定した．実験参加者からディスプレイまでの距離は，視線認識機器が実験参加者の視線を適切に認識できるように約 65 cm とした．なお，本実験環境下の視野角の 1° は画面上における約 11 mm に相当する．視線認識機器が正しく実験参加者の視線を認識できるように，ディスプレイの位置および高さを実験参加者に応じて調節した．人工的な照明による明かりのみとし，実験中の部屋の明るさを一定に保った．

3.3 実験内容

実験 1 は，実験参加者に 1 段階の視線移動を促す条件 1 および 2 段階の視線移動を促す条件 2 の 2 条件からなる．条件 1 では，実験参加者に RU ↗， LU ↖， RD ↘，および LD ↙ の 4 ジェスチャを行うよう指示した．例えば， RU を指示する際は，「右上を見てください」のように指示した．条件 2 では，実験参加者に $R \rightarrow U$ ↗， $L \rightarrow U$ ↖， $R \rightarrow D$ ↘， $L \rightarrow D$ ↙， $U \rightarrow R$ ↗， $U \rightarrow L$ ↖， $D \rightarrow R$ ↘，および $D \rightarrow L$ ↙ の 8 ジェスチャを行うよう指示した．例えば， $L \rightarrow D$ を指示する際は，「左を見て下を見てみてください」のように指示した．

なお，条件 2 において 1 段階目の方向から 2 段階目の方向へ視線移動の方向を変える際は，視線を止める必要はないと指示した．さらに，各条件において，視線移動を行う際に画面上のどこを見るかまでは指示しなかった．また，視線認識結果と実際に見ている点の誤差による影響をなくすために，視線認識の結果に関する情報を実験参加者に通知しなかった^[17]．実験参加者のうち 8 名は条件 1 から実験を開始し，残りの 8 名は条件 2 から実験を開始した．

各条件を開始する前に視線認識機器のキャリブレーションを行った．実験 1 の各条件における 1 試行は，1 度指示に従い視線を移動させることである．各試行における視線移動の開始位置を固定するために，各試行前に画面の中心に表示されている円（直径 27.9 mm）を見るよう指示した．その後，キーボードのいずれかのキーを押下すると音声にて再生される各条件の指示に従い視線移動を行うように伝えた．1 度目のキーの押下が行われると，システムは視線情報の収集を開始する．実験参加者には，試行を終了したとき（指示通りのジェスチャを行った後）に再度キーを押すように指示した．この 2 度目のキーの押下が行われると，システムは視線情報の収集を終了する．その後，実験参加者が再度キーを押すことにより次の試行が開始する．

実験参加者 1 人あたり各条件の各指示それぞれを 16 試行ずつ行った．つまり，条件 1 において 1,024 試行（4 ジェスチャ × 16 試行 × 16 人），条件 2 において 2,048 試行（8 ジェスチャ × 16 試行 × 16 人）分の視線情報を収集した．各条件において指示されるジェスチャを条件 1 において 4 ジェスチャ，条件 2 において 8 ジェスチャずつ出現順をランダムとした．疲れの影響を減らすために，試行間に任意の休憩，各条件 32 試行ごとに少なくとも 5 分間の休憩を設けた．実験に要した時間は 1 人あたり約 52 分であり，実験終了後，実験参加者は謝金として 5,000 円を受け取った．

3.4 事後処理

実験参加者がどのように視線を移動したかを解析する前に外れ値となる試行を取り除いた．まず視線座標のばらつきを除くために各試行の視線座標に対しローパスフィルタ（ $P_i = 0.25p_i + 0.75P_{i-1}$ ； P_i はローパスフィルタ適用後の i 番目の視線座標， p_i は i 番目に収集された視線座標）を適用した．

次に，ローパスフィルタを適用した後の各試行に対して，指示と実際に行ったジェスチャの方向が異なる試行を取り除いた．例えば，条件 1 において，指示が RU である試行では，視線移動の方向が第 2-4 象限の方向である試行を外れ値として取り除く．条件 2 において，指示が $L \rightarrow D$ である試行では，1 段階目もしくは 2 段階目の視線移動の方向が右か上である試行を外れ値として取り除く．結果として，条件 1 では 3.7%，条件 2 では 9.8% が外れ値として取り除かれた．その後指示と行ったジェスチャ終了点の方向（例えば $L \rightarrow D$ および $D \rightarrow L$ の際は左下）は適切であるが，指示と異なるように視線が移動していた試行を手作業にて取り除いた．例えば，指示が $L \rightarrow D$ である試行において， $D \rightarrow L$ のように視線が移動していた試行である．加えて条件 1 の際に条件 2，条件 2 の際に条件 1 のように視線が移動していた試行も取り除いた．例えば，条件 1 では，指示が RU である試行において $R \rightarrow U$ や $U \rightarrow R$ のように視線が移動していた試行である．結果として，条件 1 では 1.9%，条件 2 では 2.0% が外れ値として取り除かれた．

以降の解析には全体から外れ値を取り除いた試行（以降，正解試行）のみを用いる．つまり，条件 1 において全体の 94.4% である 967 試行，条件 2 において全体の 88.2% である 1,806 試行が正解試行である．なお，条件 2 において，外れ値となった試行が 11.8% と多くなった理由として，実験参加者から「視線を移動させる方向を間違えた」とコメントを得ていたことから，実験タスクが単純故に実験参加者が慎重さを欠いたことが考えられる．



図2 実験1の条件2の正解試行における視線移動の軌跡。
Fig.2 Gaze trajectories of successful trials in Condition 2 of Experiment 1.

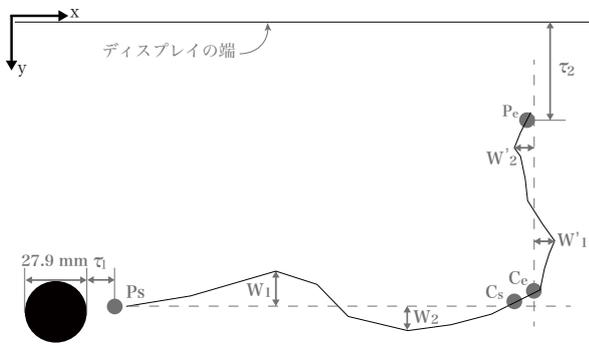


図3 実験1の解析方法。
Fig.3 Analysis used in Experiment 1.

3.5 解析

図2に、条件2の正解試行における視線移動の軌跡を示す。条件2の視線移動の軌跡を観察したところ、これらの軌跡には2つの特徴が見られた。1つは、視線移動の方向を変える際（例えば、 $R \rightarrow U$ において右を見た後に上を見始めるまで）に視線が留まることであり、もう1つは、視線移動時に視線は直線的には移動せず、ブレている（ジグザグや斜めに移動している）ことである。

これら2つの特徴を定量化するために、条件2の各正解試行の視線移動に対して、視線が移動を始めた座標 (P_s)、1段階目の方向から2段階目の方向へ視線移動の方向が変わり始めた座標 (C_s)、1段階目の方向から2段階目の方向へ視線移動の方向が変わり終えた座標 (C_e)、および視線が移動を終えた座標 (P_e) という4つの座標を検出することとした(図3)。これらの検出方法は以下の通りである。

P_s 1段階目の方向が右の場合、視線が画面中央に表示した円から外に出ている、かつ d_x (1フレーム前の視線の x 座標から現在の視線の x 座標までの距離) が閾値 τ_1 より大きくなった座標。1

段階目の方向が上の場合は、 d_x ではなく d_y (1フレーム前の視線の y 座標からの距離) が $-\tau_1$ より小さくなった座標。

C_s P_s が検出された後に d_x が τ_1 より小さくなった座標。ただしその後、 d_x が τ_1 より大きくなった場合は C_s の検出を再度行う。

C_e d_y が C_s から垂直方向に τ_1 より大きい、かつ d_y が d_x より大きい座標。

P_e C_e が検出された後に、 x および y 座標が画面右上端から τ_2 以内に移動した座標。

この検出方法において、ある点を見ている際の視線座標のばらつきは5mmであること^[18]から、視線がある点に留まっているとみなすための閾値 τ_1 を5mmとした。また、 τ_2 を55.5mmとした。これは、正解試行における視線移動の軌跡の多くが画面端まで移動していたことから、全正解試行において P_e を決定できるような値である。

3.5.1 視線が留まった時間

視線移動の方向を変える際に視線が留まった時間を C_e と C_s におけるタイムスタンプの差とした。結果、視線が留まった時間の平均時間は133ms ($SD = 86$)であった。

3.5.2 視線移動時のブレ

視線移動時のブレを、1段階目の視線移動 ($P_s - C_s$) 間および2段階目の視線移動 ($C_e - P_e$) 間に視線移動の方向と直交する方向に視線が移動した量とした。例えば図3では、1段階目の視線移動の方向が水平方向であるため、上方向と下方向へのブレの最大値の和 ($W_1 + W_2$) を計算し、2段階目の視線移動の方向が垂直方向であるため、左方向と右方向へのブレの最大値の和 ($W'_1 + W'_2$) を計算する。その結果、視線移動時のブレの平均は、1段階目の視線移動間において17.0mm ($SD = 25.9$) であり、2段階目の視線移動間において17.8mm ($SD = 24.1$) であった。

凝視後にジェスチャを行うという一連の操作を用いた意図しない操作に堅牢な視線に基づく操作手法

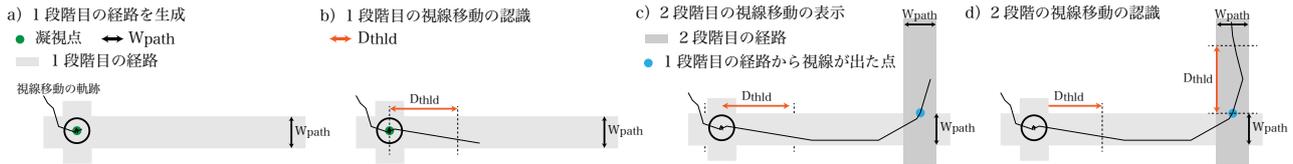


図4 本手法の認識方法.

Fig. 4 Method of detecting our technique.

4. 認識システム

実験1から、視線移動の方向を変える際に視線が133ms ($SD = 86$) 留まり、視線移動時に移動方向と直交して17.8mm程度ブレることがわかった。この結果を用いて、斜めの視線移動(実験1の条件1)を認識せず、2段階に視線が移動したこと(実験1の条件2のように2段階に視線が移動するジェスチャ)の認識が可能なシステムを設計する。本手法の認識方法を以下に示す(図4)。

- ある点に対して凝視が認識された時、凝視点を中心に幅 W_{path} の1段階目の経路を生成する。本手法において、凝視は視線が5mm以内に T_{dwell} 間留まると認識される。
- 1段階目の視線移動は、水平(垂直)方向に凝視点からある閾値 D_{thld} 以上、視線が移動することにより認識される。この時視線が1段階目の経路から外に出るとジェスチャの認識を終了する。
- 1段階目の視線移動が認識された後、視線が1段階目の経路から外に出た点(P)を中心に幅 W_{path} の2段階目の経路が生成される。
- 2段階目の視線移動は、垂直(水平)方向に閾値 D_{thld} 以上Pから移動した場合認識される。2段階目の視線移動が認識されると、その方向に対応したコマンドが実行される。

なお、この認識方法では各経路の表示は行われない。a)の凝視の認識は、b-d)いずれでも行われ、凝視が認識されるとb)から認識を再開する。以降、 T_{dwell} 、 W_{path} および D_{thld} の3つのパラメータを求め、認識システムを実装する。

4.1 認識システムに用いるパラメータの決定

認識システムに用いるパラメータは大きすぎず、小さすぎない値とする必要がある。それぞれのパラメータの決定方法を述べる。

4.1.1 凝視時間 T_{dwell}

短時間の凝視を用いることにより、ジェスチャを行う際に凝視が認識された場合はコマンドを実行するために視線を移動していないとし、意図しない操作を防げる。しかし、1段階目の視線移動から2段階目の視線移動へ方向を変えるときにも視線が留まることから、ある程度 T_{dwell} を大きくする必要がある。

そこで、 T_{dwell} を3.5.1節の視線が留まった時間の平均と標準偏差の和、つまり219(133+86)msとした。これは、条件2における視線移動時に視線が留まった時間の85.9パーセント値である。

4.1.2 経路の幅 W_{path}

W_{path} が小さい場合、ユーザは操作を行うために、より直線的に視線を移動する必要がある。これにより、直線的に視線が移動していない場合は、ユーザは意図的に2段階の視線移動を行っていないとし、意図しない操作を防げる。しかし、視線移動を行う際はブレが生じるため、視線移動の容易さという点においてはある程度 W_{path} を大きくする必要がある。

そこで、 W_{path} を3.5.2節の視線移動時のブレの平均と標準偏差の和をもとに決定した。ウィルコクソンの符号順位検定($\alpha = 0.05$)の結果、1段階目と2段階目の視線移動のブレの平均と標準偏差の和に有意差はなかった。そこで、値が大きかった42.9(17.0+25.9)mm(視野角3.9°)を W_{path} とした。これは、条件2における視線移動時のブレの93.2パーセント値である。

4.1.3 視線移動を認識するための閾値 D_{thld}

D_{thld} も、 W_{path} と同様な理由から、適当な値とする必要がある。 D_{thld} が小さい場合、視線の移動量が少なくとも1段階目および2段階目の視線移動が認識されるため、意図しない操作が生じやすくなる。対して、 D_{thld} が大きい場合、意図しない操作の発生を防げる。しかし、視線の移動量が増えると、例えば画面に右端付近にて凝視が認識された場合、さらに右側へ視線が移動するようなジェスチャを認識できない。

ここで、実験1の条件2での視線移動の軌跡を調べたところ(図2)、多くの視線移動が画面の端まで移動していた。そのため、視線が移動した距離の平均および標準偏差から D_{thld} を求めた場合、必要以上に D_{thld} が大きくなる可能性があった。そこで、既に求めた T_{dwell} および W_{path} を用いて、 D_{thld} の値を変えたときに、どれだけ条件1の正解試行が2段階の視線移動と認識されるか(条件1が2段階の視線移動と認識された場合、意図しない2段階の視線移動が認識され意図せずコマンドが実行されることになる)をもとに決定した。結果として、条件1および条件2の正解試行が2段階の視線移動と認識された割合は、 D_{thld}

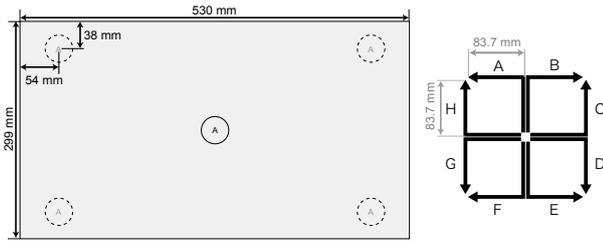


図5 実験2に用いた表示.

Fig. 5 Displays used in Experiment 2.

が 27.9 mm のときそれぞれ 9.9% および 74.5%, D_{thld} が 41.9 mm のとき 0.5% および 15.8%, および D_{thld} が 55.8 mm のとき 0.0% および 5.4% であった. この結果から, 条件 1 が 2 段階の視線移動と認識される割合が 0.0% であった 55.8 mm (視野角 5.1°) を D_{thld} として用いることとした. なお, 条件 2 において 2 段階の視線移動と認識された割合が 5.4% と低いが, これは, 実験時に視線の移動方法を指示していなかったことが原因であると考えられる.

5. 実験 2

実験 2 の目的は, 設計した認識システムを用いた際の本手法の性能を調査し, 性能をより高くするためのパラメータを得ることである. 調査する本手法の性能は, 意図した操作が正しく認識される割合およびコマンド実行に要した時間である. 実験環境は実験 1 と同じであるが, 実験 2 では頭部固定機を使用していないが, 実験 1 と同様に実験参加者とディスプレイまでの距離が約 65 cm となるようにした. また実験 2 では, 視線認識機器から収集される視線座標に対して実験 1 と同じローパスフィルタを適用した.

5.1 実験参加者

実験参加者は男性 16 名であり, 21-24 歳 (平均 23.0 歳) であった. 2 人が眼鏡を着用していた. 13 名が以前に同様な実験において視線認識機器を使用したことがあったが, いずれも日常的には使用していなかった.

5.2 実験手順

実験 2 において用いた表示を図 5 に示す. 実験 2 ではコマンド実行の対象を, 図 5 左に示す画面中央 (C) および画面四隅, すなわち左上 (TL), 右上 (TR), 左下 (BL), および右下 (BR) の 5 箇所いずれかに表示した. 実験参加者が選択すべき対象は実線の円として表示される. 対象となった円の中心には, 図 5 右に示される, 8 ジェスチャそれぞれに対応する 'A' から 'H' までの文字のいずれかが黒色に表示される. また, 実験参加者へ凝視が認識されたことを通知するために, 対象への凝視が認識されると文字を赤色にした.

実験 2 における 1 試行は, 実験参加者が 219 ms (T_{dwell}) 間対象を凝視した後 (文字が赤色になった

後) に, 対象内に書かれている文字に対応するジェスチャを行うことである. 実験開始時は画面に何も表示されていない. 実験参加者が 'Enter' キーを押すと画面に対象と文字が表示され, 本手法の認識が始まる. このとき, 文字とジェスチャの対応がわからない場合は 'Space' キーを押すよう指示した. 'Space' キーが押されるとガイド (図 5 右) が画面中央に表示され, 再度 'Space' キーが押されると表示が消える. ガイドが表示されている間は凝視およびジェスチャの認識は行われぬ. 実験参加者が 1 試行を終えた後に 'Enter' キーを押すと, 指示通りに本手法が認識されたかをビープ音にて通知するとともに, その試行において収集した視線座標を画面に表示する. このとき, 実験参加者はどのように視線が移動したため認識された (されなかった) かを確認できる. その後, 'Enter' キーを押すと画面表示が消え, 任意の休憩をした後, 再度押すことにより次の試行が開始する.

40 試行 (8 ジェスチャ \times 5 箇所) 行うことを 1 セッションとし, 実験参加者 1 人あたり 6 セッション行った. 前半 3 セッションを練習セッションとし後半 3 セッションを本番セッションとした. セッションごとに対象の表示位置およびジェスチャの指示をランダムとした. 練習セッションのみ, 実験監督者が各試行においてなぜ認識された (されなかった) かを実験参加者に説明した. 最初のセッションの開始前に視線認識機器のキャリブレーションを行った. また, 視線認識結果に関する情報を実験参加者に通知しなかった. 練習セッションおよび本番セッションそれぞれにおいて, 合計 1,920 試行 (40 試行 \times 3 セッション \times 16 人) の視線情報を収集した. 全セッション終了後, 実験参加者にアンケートへの回答を指示した. 実験に要した時間は約 85 分であり, 実験終了後, 実験参加者は謝金として 5,000 円を受け取った.

5.3 実験結果

コマンド実行の認識率として, 全試行のうち本手法が認識されかつ指示とジェスチャが合っている試行数の割合を求めた. また, コマンド実行に要した時間は対象を凝視し始めた時間からジェスチャが認識されるまでの時間である. 以降, 検定にはウィルコクソンの符号順位検定を用いた ($\alpha = 0.05$).

本番セッションにおける認識率の平均は, C において 77.6%, TL において 50.8%, TR において 61.7%, BL において 70.1%, および BR において 67.5% であった. コマンド実行に要した時間の平均は, それぞれ 813 ms, 894 ms, 895 ms, 815 ms, および 826 ms であった. 対象の位置によらず認識率が低かったのは T_{dwell} を 219 ms としたことが原因であると, 次の 2 つの理由から考えられる. 1 つは, 練習セッション時に, 実験参加者へ視線移動の方向を変えるときに視線

凝視後にジェスチャを行うという一連の操作を用いた意図しない操作に堅牢な視線に基づく操作手法

を止めないよう頻繁に指示していたことであり、もう1つは、実験参加者が「視線移動時に視線を留めないようにすることが難しい。」とアンケートに回答したことである。なお、 W_{path} および D_{thld} に関する回答はなかった。そのため、パラメータの調整は T_{dwell} のみ行う。

また、画面四隅における認識率に関して、検定の結果 C-BL 間以外は有意差がみられた ($p < 0.05$)。この理由として、5名の実験参加者は「画面角から画面の外に向かって視線を移動することは難しかった。」とアンケートに回答していたことが挙げられる。対して、11名の実験参加者は「画面角から画面中央へ向かって視線を移動することは、ディスプレイの縁に沿って視線を移動することができたため容易だった。」とアンケートに回答した。そこで、画面四隅から画面中央へ視線を移動した場合（例えば、対象の位置が TR である際にジェスチャの指示が $L \rightarrow D$ および $D \rightarrow L$ の試行）の認識率を計算したところ、平均 71.4%であった。これらの結果から、実験 1 の結果をもとに設計した認識システムを画面四隅に使用することは難しいと考えられ、対象の位置ごとにパラメータを調整する必要があることがわかった。

5.4 パラメータ調整

T_{dwell} の影響を調査するために、 T_{dwell} を 219 ms から 1,000 ms まで 10 ms ずつ増加させながら、本セッションにおける認識率をシミュレートした。認識率は T_{dwell} の増加に伴い向上し、 T_{dwell} が 490 ms の時に 88.0%となった。さらに、実験 1 における視線が留まった時間の平均 (133 ms) と $1SD (= 86.4)$, $2SD$, $3SD$, および $4SD$ の和、つまり、219 ms, 306 ms, 392 ms, 479 ms を T_{dwell} とした際の、認識率もシミュレートした。なお、これらは、視線が留まった時間の 85.9, 95.5, 98.6, および 99.5 パーセント値である。結果、認識率はそれぞれ 77.6%, 84.9%, 87.0%, 87.5%であり、検定の結果、306–392 ms 間および 392–479 ms 間以外に有意差が見られた。また、219 ms, 306 ms, 392 ms, 479 ms を T_{dwell} とした際のコマンド実行に要した時間はそれぞれ、810 ms ($SD = 100$), 913 ms ($SD = 97$), 1002 ms ($SD = 101$), および 1087 ms ($SD = 101$) であった。これらの結果から、認識率を高めること、および意図しない操作を防ぐこと両方の観点から、306 ms を T_{dwell} として扱う。

6. 実験 3

実験 3 の目的は、認識システムを意図しない操作に対して堅牢にするために、どのような改良をする必要があるかを調査することである。この調査には、凝視に基づく対象選択手法を用いたアプリケーションの操作および、本手法を用いたアプリケーションの操作を

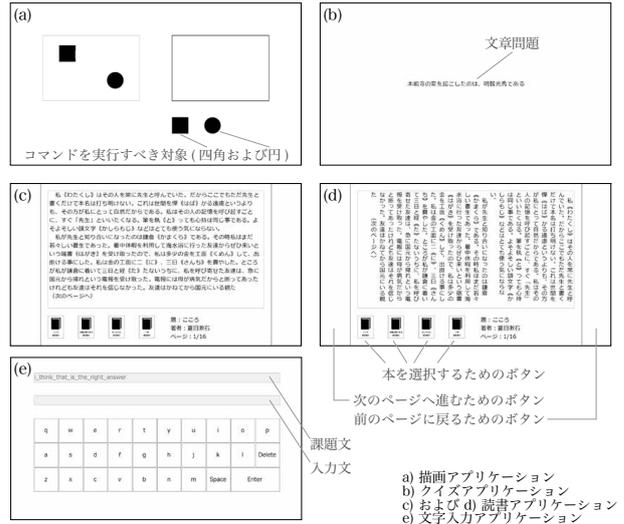


図 6 実験 3 に用いたアプリケーション。
 Fig. 6 Applications used in Experiment 3.

する際の視線情報を用いることにした。実験 1 および 2 から実験 3 までに約 6ヶ月の期間が空いており、実験 3 を実施する時には Tobii EyeX が製造中止となっていたため、実験 3 では視線認識機器として当時 Tobii 社が推奨していた視線認識機器である Tobii Eye Tracker 4C (サンプリングレート: 90 Hz) を用いた。なお、Tobii Eye Tracker 4C の性能は Tobii EyeX と比べ向上しているため、結論には影響はないと考えられる。視線認識機器以外の使用機器および実験環境は実験 2 と同一である。実験 3 において使用するパラメータは、 T_{dwell} が 306 ms, W_{path} が 42.9 mm, および D_{thld} が 55.8 mm である。

6.1 実験参加者

実験参加者は男性 16 名であり、21–25 歳 (平均 22.5 歳) であった。3 名が眼鏡、2 名がコンタクトを着用していた。13 名が以前に実験参加者として視線認識機器を使用したことがあったが、いずれも日常的には使用していなかった。実験参加者 16 名のうち実験 2 に参加していた 8 名を A 群とし、残りの 8 名を B 群とした。

6.2 実験に用いたアプリケーション

実験 3 に用いたアプリケーション (図 6) は、描画アプリケーション (図 6a), クイズアプリケーション (図 6b), 読書アプリケーション (図 6c および d), および文字入力アプリケーション (図 6e) である。描画およびクイズアプリケーション (以降、ジェスチャアプリケーション) は、本手法を用いて操作が行われる。また読書および文字入力アプリケーション (以降、凝視アプリケーション) は凝視に基づく対象選択を用いて操作が行われる。実験 3 における凝視に基づく対象選択の凝視時間は 1.0 s とした。

描画アプリケーションのタスクでは、実験参加者に

表 1 実験 3 におけるジェスチャとコマンドの対応.

Table 1 Mappings of commands and gestures used in Experiment 3.

	コマンド	1 回目	2 回目
描画	コピー	$U \rightarrow R$	$U \rightarrow L$
	ペースト	$D \rightarrow L$	$D \rightarrow R$
クイズ	正	$R \rightarrow D$	$R \rightarrow U$
	誤	$L \rightarrow U$	$L \rightarrow D$

画面右に表示されている対象（四角・円）のコピーを行い、画面左に配置されている対象と同じ位置にペーストするよう指示した。タスクを実施するためには、各コマンドが紐付けられたジェスチャを2度ずつ（四角・円のコピー・ペースト）行う必要がある。クイズアプリケーションのタスクでは、実験参加者に画面に表示されている文章問題に回答するよう指示した。問題は正・誤にて回答可能な文章問題4問であり、回答は正・誤それぞれ2度ずつ出現する。タスクを終えるには、正・誤の回答が紐付けられたジェスチャを2度ずつ行うことになる。読書アプリケーションのタスクでは、本の選択および次のページへ進む・前のページに戻るといった操作を行いながら読書するよう指示した。なおこのタスクでは、縦方向および横方向への視線移動を偏りなく取得するため、横書き（図6c）および縦書き（図6d）での文章表示を行った。文字入力アプリケーションのタスクでは、画面上部に表示されている課題文¹と同じ文章を作成するよう指示した。スペースの表示を“_”（アンダーバー）とし、文章の入力を終えた際には、‘Enter’キーを押すよう指示した。文章が正しい場合は次の課題文が表示される。課題文はスペースを含めた32文字から構成される。

6.3 実験手順

実験は、描画およびクイズアプリケーションそれぞれを用いたタスク（以降、それぞれ描画タスクおよびクイズタスク）、横書きおよび縦書きの読書アプリケーションを用いて読書を行うタスク（以降、読書タスク）、文字入力アプリケーションを用いた入力を行うタスク（以降、文字入力タスク）から構成される。本手法への慣れが意図しない操作の発生に影響するかを調査するために、A群の実験参加者はジェスチャアプリケーションを用いたタスク（ジェスチャタスク）から、B群の実験参加者は凝視アプリケーションを用いたタスク（凝視タスク）から実験を開始した。実験3に要した時間は約55分であり、実験終了後、実験参加者は謝金として5,000円を受け取った。

6.3.1 ジェスチャタスク

ジェスチャタスクは、本手法を用いてコマンドを実行する練習の後に実施された。この練習は、実験参加

者の意図した8ジェスチャが連続して認識されるまで行われ、約212s要した。その後、各タスクを2度ずつ行うよう指示した。タスクの実施順は、描画（1回目）→描画（2回目）→クイズ（1回目）→クイズ（2回目）である。各タスクの各回数目における、コマンドとジェスチャの組み合わせを表1に示す。例えば描画タスクの1回目において四角の対象にコピー・ペーストを行う場合、実験参加者はまずコピーするために四角の対象を凝視した後に $U \rightarrow R$ を行う。その後ペーストを行うために、図6a左の四角と同じ位置を凝視した後に $D \rightarrow L$ を行う。また、例えばクイズタスクの1回目では、実験参加者は問題が正解だと思ふ場合、画面内の任意の位置を凝視した後に $R \rightarrow D$ 行う。なお、ジェスチャタスクでは、凝視が認識された点を描画し、かつジェスチャが認識された場合はビーブ音による通知を行った。タスクによる疲れの影響を取り除くため、各タスク終了後に5分間の休憩を設けた。タスクの終了後に、実験参加者には、アンケートへの回答を指示した。ジェスチャタスクでは結果として合計256回（8ジェスチャ×2回×16名）の本手法に基づく操作を行う際の視線情報を収集した。

6.3.2 凝視タスク

凝視タスクは、実験参加者に選択したい対象を1.0s間見続けると選択できることを説明し、実際にどのように選択できるかを試した後に実施した。タスクの実施順は、読書（横書き）→読書（縦書き）→文字入力である。なお、凝視タスクの際は、凝視が認識されたことに対する通知を行っていない。実験による疲れの影響を取り除くため、各タスク終了後に5分間の休憩を設けた。読書タスクでは実験参加者に、横書きおよび縦書きそれぞれ5分間読書をするように指示した。文字入力タスク（5文章の入力）の終了に要した時間は約7分であった。凝視タスクでは結果として合計272分間（17（=5+5+7）分×16名）の凝視に基づく操作を行う際の視線情報を収集した。

6.4 実験結果

本実験では、意図しない操作の認識回数を、凝視タスク時に認識された本手法の回数とした。なお、凝視タスク時に本手法が認識されてもコマンドは実行されない。結果、意図しない操作の認識回数は26回であった（読書（横書き）：5回、読書（縦書き）：4回、文字入力：17回）。マンホイットニー検定の結果、A-B群（実験2に参加しているかどうか）間に有意な差は見られなかった。

ジェスチャタスク実施後のアンケートには、3名が2段階に視線を移動させることは難しいと回答し、6名がジェスチャが正しく認識されなかったと回答した。しかし、全実験参加者は全ジェスチャタスクを正しく終わることができていた（コマンド実行ができな

1: MacKenzie らが提案したフレーズセット^[19]から抽出

凝視後にジェスチャを行うという一連の操作を用いた意図しない操作に堅牢な視線に基づく操作手法

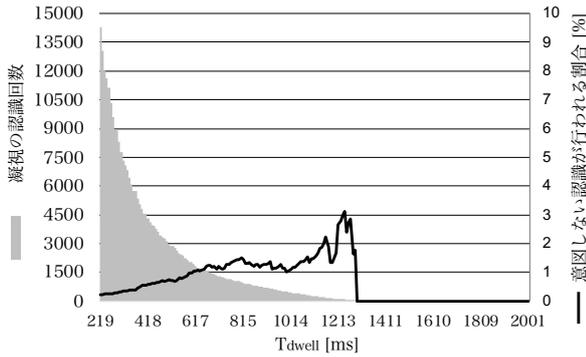


図 7 T_{dwell} ごとの凝視の認識回数と凝視が認識された際に意図しない操作が認識される割合。

Fig. 7 Number of dwell detections and unintentional detection rate against the T_{dwell}

めに実験を中止する必要はなかった)。これらのことから、実験参加者はコマンドを実行するために、本手法を 2 回以上実施していたと考えられる。

6.5 解析

凝視およびジェスチャタスクを通じて収集した視線情報をもとに、本手法の認識システムを改良するための解析を行う。

6.5.1 パラメータを変えた際の意図しない操作の認識回数

T_{dwell} , W_{path} および D_{thld} の値を以下のように変えた際に意図しない操作がどの程度検出されるかをシミュレートする。

T_{dwell} 219 ms から 2,000 ms まで 10 ms ごとの値。

W_{path} 42.9 mm から 4.29 mm まで 4.29 mm ごとの値。

D_{thld} 55.8 mm から 535.7 mm まで 5.58 mm ごとの値。

結果、多くの組み合わせにおいて認識回数が 0 回となった。以降、これらの組み合わせの中から、コマンド実行が容易になると考えられるパラメータを探す。

まず、 W_{path} および D_{thld} を変えずに (それぞれ 42.9 mm および 55.8 mm) T_{dwell} のみを増加させると、意図しない操作の認識回数は 306 ms から 1,290 ms までで 26 回から 0 回へ減少した。このことから、大きい T_{dwell} を用いることにより意図しない操作の発生を抑えられることがわかった。この理由として、大きい T_{dwell} を用いる場合、単に凝視の認識回数が少なくなるために、その後のジェスチャの認識が行われなかったと考えられる。図 7 に T_{dwell} ごとの、凝視の認識回数および凝視が認識された際にジェスチャが認識された割合 (意図せず本手法が認識された割合) を示す。 T_{dwell} の増加に伴い、意図しない操作が認識される割合も増加する傾向が見られた。つまり、 T_{dwell} を大き

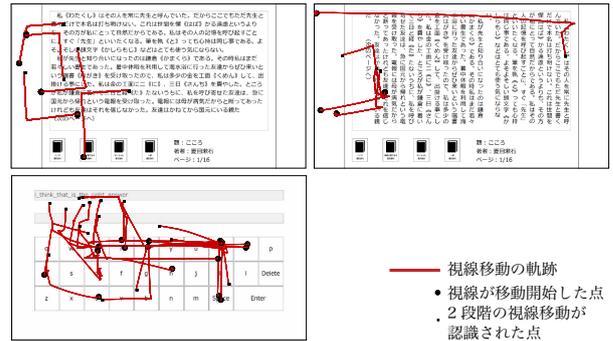


図 8 本手法が認識された際の視線移動の軌跡。

Fig. 8 Trajectories of two-level strokes caused when our technique were unintentionally detected.

くすることにより、意図しない操作の認識は行われやすいといえる。しかし、 T_{dwell} を大きくすることはコマンド実行の容易さにおいて有効であることから、適度に大きい値の使用を検討する。

次に、 W_{path} と D_{thld} を変えることの影響を調べるために、 T_{dwell} を 306 ms として、 W_{path} と D_{thld} を変えた場合の意図しない操作の認識回数を調べた。結果、意図しない操作の認識回数が 0 回となる (W_{path} , D_{thld}) の組み合わせが多くあった。これらの組み合わせの中から、操作が容易であると考えられる W_{path} が最も大きく D_{thld} が最も小さい組み合わせを求めた。結果、(13.7 mm, 100.0 mm) の組合せがみつかった。しかし、依然として W_{path} が小さくかつ D_{thld} が大きく、実際に著者がこの組み合わせにおいてコマンド実行をしようと試みたが失敗することが多かった。

そこで、全 26 回の意図しない操作の認識が行われた際の視線移動をプロットすることによりさらなる調査を行った。意図しない操作の認識が行われた際の視線移動の軌跡を図 8 に示す。図 8 が示すように、多くの視線移動はジグザグではなく直線的であった。この結果から、 W_{path} を小さくすることの、意図しない操作を防ぐ効果は小さいことがわかる。さらに、多くの視線移動の移動量は、水平方向に比べ垂直方向が少ないこと、および意図しない認識の 1 段階目の方向は水平方向が多い (水平方向: 17 回, 垂直方向: 9 回) こともわかる。そこで、 W_{path} および T_{dwell} を変えずに、 D_{thld} を水平方向 (D_h) と垂直方向 (D_v) に分けた場合の影響を調査した。結果、 D_h および D_v それぞれが小さくなるのは D_h が 116.0 mm, D_v が 89.3 mm の組み合わせのときであった。実験に用いたディスプレイのアスペクト比が 16:9 であったことも影響していると考えられるが、 D_h に比べ D_v を小さくすることも、意図しない認識を防ぐことだけではなく、コマンド実行の容易さに貢献すると考えられる。また、 D_{thld} を D_h と D_v に分けることは、人間の視野角が水平方向

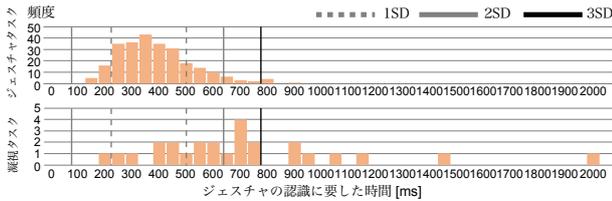


図9 ジェスチャの認識に要した時間の頻度。
Fig.9 Frequencies of time required for gesture recognition.

と垂直方向にて異なる^[20]ことから妥当である。さらに、Møllenbach からも、垂直方向の1段階の視線移動は水平方向の1段階の視線移動に比べ、時間がかかると述べている^[8]ことから、 D_h より D_v を小さくすることは妥当である。

6.5.2 意図しない操作時のジェスチャの認識に要した時間

26回の意図しない操作の認識の内、いくつかのジェスチャの認識には長い時間を要していた。そこで意図的にジェスチャを行った際、および、意図しない操作の認識が行われた際のジェスチャに要した時間を調査した。ジェスチャに要した時間は、視線が移動し始めた時間からジェスチャが認識されるまでの時間である。意図的に行われたジェスチャの認識に要した時間はジェスチャタスク、意図せず操作が認識された際に要した時間は凝視タスクの際に収集した視線情報から求めた。図9に、意図的に行われたジェスチャの認識に要した時間（上）および、意図せず操作が認識された際に要した時間（下）の50msごとの頻度を示す。意図的に行ったジェスチャの認識に要した時間の平均は362ms ($SD = 137$)であり、1試行を除いて100–800msの範囲に収まっていた。対して、意図せずジェスチャが認識された際に要した時間はばらついていることがわかる。

この結果より、ジェスチャを認識するための時間の範囲に制限を設けることにより、意図しない操作を防ぐことが可能であると考えられる。そこで、実際に認識を行うための時間を、意図的に行われたジェスチャに要した時間の平均 $\pm 1SD$ (225–499ms)、平均 $\pm 2SD$ (88–636ms)、および平均 $\pm 3SD$ (0–773ms)に制限した。その結果、それぞれ20回(76.9%)、13回(50.0%)、および7回(26.9%)の意図しない操作の認識を防げることがわかった。また、認識を行うための時間を0–773msに制限した際に、ジェスチャタスク時に行われた全256回の本手法のうち3回(1.2%)が認識されなかった。これらのことから、ジェスチャを認識するための時間を制限することは、意図しない操作を防ぐ点において有効であると考えられる。

7. 認識システムの改良

実験3の解析結果をもとに、意図しない操作の認識を防ぐことが可能であり、かつ容易にコマンドを実行できるようにシステムを改良する。その後、改良したシステムを用いた場合、どの程度の認識率にてコマンドを実行可能かを調査する。

7.1 最適なパラメータの決定

ジェスチャを認識するための時間を0–773msに制限し、 T_{dwell} 、 W_{path} 、 D_h 、および D_v を変えながらシミュレーションを行う。容易にコマンドを実行できるかの判定には、ジェスチャタスク時に行われた合計256回の本手法が何回認識されるかという、意図した本手法の認識率を用いる。なお、実験3に用いたパラメータをさらに変更するため、シミュレーションの結果では、意図した本手法の認識率は低くなると考えられる。

シミュレーションの結果、 T_{dwell} を506ms、 W_{path} を34.6mm(視野角 3.1°)、 D_h を116.0mm(視野角 10.5°)、および D_v を66.9mm(視野角 6.1°)としたパラメータの組み合わせが最適となることがわかった。このパラメータの組み合わせでは、意図しない認識は生じず、かつ意図した操作の認識率は40.6%(256回中104回)である。この40.6%という認識率は、意図しない認識が生じないパラメータの組み合わせの中では最大の認識率である。 T_{dwell} として用いる506msは、ミダスタッチ問題を軽減するための手法であれば約100–400ms^{[3]–[5]}、また凝視とジェスチャを組み合わせた手法では2s^[12]が凝視時間として使用されていることから、大きすぎない凝視時間であるといえる。また、 D_h として用いる116.0mmは実験に用いたディスプレイの横幅(530mm)の21.8%の大きさであり、 D_v として用いる66.9mmは実験に用いたディスプレイの縦幅(299mm)の22.3%である。そのため、視線を移動する必要のある絶対距離は水平方向と垂直方向において異なるが、ディスプレイ上の相対的な距離(つまり、各方向ディスプレイ上の22.0%程度の距離)は近いいため、コマンド実行の観点においても有効であると考えられる。

7.2 最適化したシステムを用いた本手法の性能

シミュレート結果のパラメータを使用した際の性能を調査した。

男性10名を実験参加者として募った。実験参加者は著者が所属する研究室内の学生を含む21–25歳(平均23.1歳)である。4名が眼鏡を着用していた。全員以前に実験参加者として視線認識機器を使用したことがあったが、日常的には使用していなかった。調査に用いた機器、および環境は実験3と同じである。

調査に用いたアプリケーションは実験2に用いたア

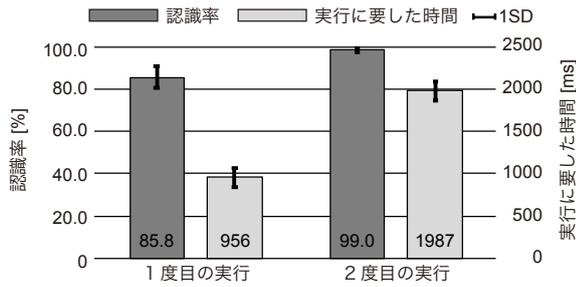


図 10 本手法を 1 度および 2 度行った際の、認識率およびコマンド実行に要した時間。
Fig. 10 Success rate and activation time in attempting once and twice.

アプリケーションと同じである。実験参加者には、このアプリケーションを用いて、画面中心 (C) を始点とした 8 ジェスチャが正しく認識されるまで繰り返し行う (1 セッション) ように指示した。実験開始前に視線認識機器のキャリブレーションを行った。その後、実験参加者には本手法がどのように認識されるかに慣れるための練習を少なくとも 5 分間行い、練習として 1 セッション行うよう指示した。その後、本番として 5 セッションを行う。各セッション間に少なくとも 1 分の休憩を設けた。調査を通じて行われた本手法は合計 400 回 (8 ジェスチャ × 5 セッション × 10 人) である。この調査に要した時間は実験参加者ごとに約 25 分であった。

改良したシステムを用いた際の本手法の性能を図 10 に示す。認識率は、本手法を 1 度行った場合 (左) において平均 85.8% ($SD = 5.1$) であった。このとき実験参加者ごとの認識率の最大は 90.0% であり、最小は 72.5% であった。本手法を 2 度行った場合 (右) における平均認識率は 99.0% ($SD = 1.2$) であった。また、実験参加者ごとの認識率の最大は 100.0% であり、最小でも 97.5% と高くなった。ジェスチャごとでは $L \rightarrow U$ および $R \rightarrow D$ の認識率がそれぞれ 76.0% および 72.0% と低くなったが、2 度行うことによりそれぞれ 98.0% および 100% の認識率となった。これらの結果より、ユーザは本手法を行った際にコマンドが実行されなかったとしても、再度行うことによりコマンドを実行できると言える。

ジェスチャに要して良い時間の範囲 (0–773 ms) に関して、意図的に行われた 4 回 (1.0%) のジェスチャがこの範囲を超えたために認識されなかった。ジェスチャに要した時間は、1 度目に行ったときに 450 ms ($SD = 119$) であり、2 度目に行ったときに 525 ms ($SD = 112$) であった。1 度目と 2 度目で要した時間が異なった理由としては、実験参加者が 2 度目の視線移動をより慎重に行ったためであると考えられる。この範囲は実験 3 (つまり、本調査とは異なるパラメー

タを用いた実験) の結果から求めたものであるが、有効であることがわかった。

8. 課題

実験 1–3 を通じて、意図しない操作の認識を防ぐことが可能であり、かつ容易にコマンドを実行できるようにシステムを設計・改良した。結果として、図 4 に示す認識システムにおけるパラメータに、 T_{dwell} として 506 ms, W_{path} をとして 34.6 mm 視野角 (3.1°), D_h として 116.0 mm 視野角 (10.5°), および D_v として 66.9 mm 視野角 (6.1°) を用いた場合に実験にて収集した視線情報において意図しない操作は発生しないことがわかった。また、意図して本手法を行おうとした時の成功率は 85.8% であることがわかった。本章において、本研究から明らかとなった今後の課題および本研究の結果を議論する。

8.1 実験条件による制限

実験 1 の参加者 1 名は、「生まれつき眼振が生じる」と述べていた。そこで、眼球振盪がジェスチャの認識に与える影響を調べた。収集した視線情報を観察したところ、視線移動開始前に見よう指示した円付近 (図 2) において視線が左右に移動していた。さらに、実験 1–3 の実験参加者は全員若く普段からコンピュータを扱っていた。眼球運動は年齢によって変化する (例えば、小児 (12 歳以下) および高齢者 (50 歳以上) の記憶誘導性サッカード (ある刺激へのサッカード (視覚誘導性サッカード) ではなく、記憶していた刺激位置へのサッカード) の潜時は、若年者に比べ長くなる^[21]) ため、本研究の実験結果を年齢問わず使用することは難しいとも考えられる。眼球振盪の影響は、凝視に基づく対象選択手法においても影響すると考えられるため、本手法のみに影響する問題ではないが、眼球 (もしくは眼) に何らかの症状を患っているユーザ、または、小児および高齢のユーザにおいても、本研究において実装したシステムおよびそのパラメータが適切かを調査する必要がある。

また、実験に用いたアプリケーションは、実験 1 および 2 では背景が無地のアプリケーションであり、実験 3 では、描画、クイズ、読書、および文字入力アプリケーションの 4 つであった。意図しない操作を防ぐためには、より多くの種類のアプリケーションにて視線情報を収集する必要がある。

実験時は、実験に関係のないアイコンなどを画面に表示しなかった。しかし、実際にコンピュータを操作する際は、アイコンや文章、画像などが表示され、これらが視線移動に影響すると考えられる。実際に、実験 2 では 2 名、実験 3 では 3 名が、視線を移動する際の参考となるアイコンなどが画面に表示されていないため、視線移動が難しかったとコメントした。さら

に、実験2の1名が、画面の縁に沿って視線を移動すると、視線移動が容易に感じたコメントした。これらのコメントから、画面背景の表示が視線移動の容易さに影響し、本手法の認識率に良い影響を与えると考えられる。対して、例えば、ユーザが意図的にジェスチャを行おうとしている際に、画面に表示されているアイコンに視線が移り、視線移動を失敗するとも考えられる。そのため、画面の表示がどのような影響を与えるかも調査する必要がある。

実験2と3において、凝視が認識されたことの通知を実験参加者に行った。この通知により、実験参加者はジェスチャが認識されなかった原因として、2段階の視線移動中に意図せず凝視が認識されたためであるとわかる。加えて、凝視が認識されているにもかかわらず、長く凝視し続ける必要がなくなるという利点が考えられる。一方、通知を行うことによりユーザが不快に感じる可能性もあるが、実験中にそのようなコメントはなかった。凝視が認識されたことの通知以外にも、視線座標の通知を行うべきかどうかは、視線に基づく操作全体での検討すべきことでもあるため、通知の有無、方法（視覚や聴覚）などを今後調査する必要がある。

8.2 2段階の視線移動に関して

本研究にて収集した視線移動は、既存研究において提案されてきた1段階の視線移動に基づくコマンド実行手法^[8]の設計や用途に関する議論にも役立つと考えられる。まず、1段階の視線移動は、2段階の視線移動より日常的に生じる視線移動である^[9]ため、このシステムを実装する場合、パラメータをより制限する必要がある。そこで、実験3の凝視タスクにて収集した視線情報を用いたシミュレーションを行った。結果、 T_{dwell} を506msとした場合に意図しない1段階の視線移動の発生を0回にするには、 W_{path} を5.3mmより小さく、 D_h を368mm（実験に用いたディスプレイの横幅の69.3%）より大きく、かつ D_v を167mm（実験に用いたディスプレイの縦幅の55.9%）より大きくする必要があることがわかった。このパラメータの組み合わせにおいてユーザは、視線を直線的にかつ、水平・垂直方向ともにディスプレイの半分以上移動する必要がある。また、視線を移動する必要がある距離が大きいと、対象の位置によっては実行可能なコマンドの数が制限される。つまり、1段階の視線移動は、アプリケーションやシステム全体へのコマンドを実行するには向いていると思われるが、複数あるアイコンの中から1つのアイコンにコマンドを実行することには不向きである。

本研究では既存のジェスチャ認識システムを用いずに、2段階の視線移動をジェスチャとし、これを認識するシステムを設計した。一方、手指に基づく操作手

法においては、\$1 Recognizer^[10]や動的時間伸縮法を用いたジェスチャ認識が行われてきている。\$1 Recognizerを用いた場合のジェスチャ認識の一例として、マウスの左ボタンを押して指を離すまでのマウスの軌跡を用いたジェスチャ認識がある²。つまり、マウスの左ボタンを押した点がジェスチャ認識を開始点となり、ボタンから指を離した点がジェスチャ認識の終了点となる。本研究においてこのような認識システムを用いなかった理由は、視線に基づく操作においてジェスチャ認識の開始点と終了点を決定することが難しいためである。例えば、ジェスチャ認識の開始点と終了点を凝視が認識された点とすると、凝視が2度認識された時点で何らかのジェスチャが認識されコマンドが実行される。結果として、ミダスタッチ問題に直面している凝視に基づく操作手法を用いた場合、意図せずジェスチャ認識の開始点および終了点が認識され、意図しないコマンドが実行される可能性がある。そのため、これらの認識システムを使用せず、2段階の視線移動を認識するシステムを実装した。なお、ユーザの意図した2段階の視線移動の認識が可能な本手法においては、ユーザが2段階の視線移動をし終えたときにジェスチャの認識は終了する。

8.3 マーキングメニューへの応用

本手法では、ガイドの表示を行わないため、慣れていないユーザは、どのコマンドがどのジェスチャに紐付けられているかがわからなくなる可能性がある。この課題は、マーキングメニュー^[16]の概念を取り入れることにより解決できると考えられる。

本手法ではガイドの表示を行わないため、実行したいコマンドが紐付いているジェスチャを既に知っているユーザを対象としたシステムであると言える。そこで、長い時間凝視を続けるとガイドを表示するようにシステムを改良する。しかし、ミダスタッチ問題の発生により、意図せずガイドが表示されてしまう問題が生じる。そのため、ガイドを表示するための凝視時間も、認識システムの設計と同様に実験の結果から決定することを予定している。実行したいコマンドが紐付いているジェスチャを既に知っているユーザは、短時間（例えば、506ms）の凝視を行った後にジェスチャを行うことによりコマンドを実行でき、そうでないユーザは、長時間の凝視を行い、表示されるガイドを参考にしながらコマンドを実行できる。なお、凝視時間に加え、どのようにガイドを表示すべきかの検討が必要であるため、ガイドの表示方法を検討している研究（例えば、^[12]、^[22]など）を参考に行う。加えて、長時間の凝視を行った後にガイドを表示するのであれば、^[12]にて行われているように、円滑性追跡眼球運動を用いる

² <http://depts.washington.edu/acelab/proj/dollar/index.html>
参照

ことによりタッチでの操作に用いられている不規則なジェスチャ (例えば^[23]) も使用でき, さらに多くのコマンドの選択肢をユーザに提示できる可能性がある。

8.4 対象の位置による影響

実験2の結果から, 画面四隅にある対象 (TL, TR, BL, およびBR) では, 認識率が低下することがわかった。この原因としては, 画面端付近では視線認識機器の精度が低くなることや, 画面四隅からジェスチャを行う場合は, 既に視線が画面四隅まで移動しているため, さらに視線を移動させることが難しくなることが考えられる。これは, 画面に取り付けて使用する視線認識機器では, 頭部運動の影響を受けることが原因であり, 頭部装着型の視線認識機器を用いることにより解決できる可能性がある。これらの機器を用いると, ユーザが頭部を移動したとしても高精度に視線認識が可能であるため, 対象の位置によらず画面中央と同じように操作が可能となると考えられる。

9. おわりに

本論文において, 凝視後にジェスチャを行うという一連の操作によるコマンド実行手法を示した。本手法を用いることによりユーザは, 506 ms の凝視を行った後に, 2段階の視線移動を行うことによりコマンドを実行できる。この一連の操作を認識するためのシステムを実験を通じて実装したことにより, 意図しない操作を防ぐことを可能とした。認識システムの実装は3つの実験を通じて収集したユーザの視線移動の軌跡をもとに行った。まず実験1において, ユーザがどのように視線を移動するか調査しシステムの設計を行った。その後の実験2および3を通じて, 意図しない操作が生じないかつコマンド実行が容易となるようにシステムの改良を行った。さらに, 設計および改良した認識システムを用いた場合, ユーザはコマンド実行を85.8%の精度にて成功でき, このときに要した時間は956 msであることを示した。また, ユーザは一度コマンド実行を失敗したとしても, 再度2段階の視線移動を行なうことにより, 99.0%の精度にてコマンドを実行できる。視線に基づく操作での課題である意図しない操作を防げること, および実際に操作が可能であることから, 凝視後にジェスチャを行うという一連の操作を用いる本手法は, 視線に基づく操作手法における基礎となるだろう。

参考文献

- [1] Jacob, R. J. K.: The Use of Eye Movements in Human-computer Interaction Techniques: What You Look at is What You Get; In *ACM Transaction on Information Systems*, Vol. 9, No. 2, pp. 152-169 (1991).
- [2] Isomoto, T., Yamanaka, S. and Shizuki, B.: Gaze-based Command Activation Technique Robust Against Unintentional Activation using Dwell-then-Gesture; In *Proceedings of Graphics Interface 2020*, GI '20, Canadian Human-Computer Communications Society / Société canadienne du dialogue humain-machine, pp. 256-266 (2020).
- [3] Mott, M. E., Williams, S., Wobbrock, J. O. and Morris, M. R.: Improving Dwell-Based Gaze Typing with Dynamic, Cascading Dwell Times; In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, ACM, pp. 2558-2570 (2017).
- [4] Nayyar, A., Dwivedi, U., Ahuja, K., Rajput, N., Nagar, S. and Dey, K.: OptiDwell: Intelligent Adjustment of Dwell Click Time; In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, IUI '17, ACM, pp. 193-204 (2017).
- [5] Isomoto, T., Ando, T., Shizuki, B. and Takahashi, S.: Dwell Time Reduction Technique Using Fitts' Law for Gaze-based Target Acquisition; In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, ETRA '18, ACM, pp. 26:1-26:7 (2018).
- [6] Dybdal, M. L., Agustin, J. S. and Hansen, J. P.: Gaze Input for Mobile Devices by Dwell and Gestures; In *Proceedings of the 2012 ACM Symposium on Eye Tracking Research & Applications*, ETRA '12, ACM, pp. 225-228 (2012).
- [7] Hyrskykari, A., Istance, H. and Vickers, S.: Gaze Gestures or Dwell-based Interaction?; In *Proceedings of the 2012 ACM Symposium on Eye Tracking Research & Applications*, ETRA '12, ACM, pp. 229-232 (2012).
- [8] Møllenbach, E., Lillholm, M., Gail, A. and Hansen, J. P.: Single Gaze Gestures; In *Proceedings of the 2010 ACM Symposium on Eye-Tracking Research & Applications*, ETRA '10, ACM, pp. 177-180 (2010).
- [9] Drewes, H. and Schmidt, A.: Interacting with the Computer Using Gaze Gestures; In *Proceedings of the 11th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part II*, INTERACT '07, Springer-Verlag, pp. 475-488 (2007).
- [10] Wobbrock, J. O., Wilson, A. D. and Li, Y.: Gestures Without Libraries, Toolkits or Training: A \$1 Recognizer for User Interface Prototypes; In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, UIST '07, ACM, pp. 159-168 (2007).
- [11] Istance, H., Hyrskykari, A., Immonen, L., Mansikkamaa, S. and Vickers, S.: Designing Gaze Gestures for Gaming: An Investigation of Performance; In *Proceedings of the 2010 ACM Symposium on Eye-Tracking Research & Applications*, ETRA '10, ACM, pp. 323-330 (2010).
- [12] Delamare, W., Han, T. and Irani, P.: Designing a Gaze Gesture Guiding System; In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17, ACM, pp. 26:1-26:13 (2017).
- [13] Urbina, M. H., Lorenz, M. and Huckauf, A.: Pies with EYEs: The Limits of Hierarchical Pie Menus in Gaze Control; In *Proceedings of the 2010 ACM Symposium on Eye-Tracking Research & Applications*, ETRA '10, ACM, pp. 93-96 (2010).

- [14] Ware, C. and Mikaelian, H. H.: An Evaluation of an Eye Tracker As a Device for Computer Input; In *Proceedings of the 1987 CHI/GI Conference on Human Factors in Computing Systems and Graphics Interface*, CHI '87, ACM, pp. 183–188 (1987).
- [15] Vidal, M., Bulling, A. and Gellersen, H.: Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets; In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, ACM, pp. 439–448 (2013).
- [16] Kurtenbach, G.: The Design and Evaluation of Marking Menus, PhD Thesis, Toronto, Ont., Canada, Canada (1993). UMI Order No. GAXNN-82896.
- [17] Jacob, R. J. K.: Eye Movement-Based Human-Computer Interaction Techniques: Toward Non-Command Interfaces; Vol. 4, pp. 151–190 (1993).
- [18] Feit, A. M., Williams, S., Toledo, A., Paradiso, A., Kulkarni, H., Kane, S. and Morris, M. R.: Toward Everyday Gaze Input: Accuracy and Precision of Eye Tracking and Implications for Design; In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, ACM, pp. 1118–1130 (2017).
- [19] MacKenzie, I. S. and Soukoreff, R. W.: Phrase Sets for Evaluating Text Entry Techniques; In *Proceedings of the 2003 SIGCHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '03, ACM, pp. 754–755 (2003).
- [20] Rush, T. C. and Fulton, J. F.: *Medical Physiology and Biophysics*, W. B. Saunders & Co (1961).
- [21] 彦坂 興秀: 発達と加齢に伴う眼球運動の変化とその異常; Vol. 29, No. 3, 一般社団法人 小児神経学会, pp. 213–219 (1997).
- [22] Henderson, J., Malacria, S., Nancel, M. and Lank, E.: Investigating the Necessity of Delay in Marking Menu Invocation; In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, ACM, pp. 169:1–169:13 (2020).
- [23] Alvina, J., Griggio, C. F., Bi, X. and Mackay, W. E.: CommandBoard: Creating a General-Purpose Command Gesture Input Space for Soft Keyboard; In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, ACM, pp. 17–28 (2017).

(2020年8月10日受付)

著者紹介

磯本 俊弥 (学生会員)



1995年生。2018年筑波大学情報学群情報メディア創成学類卒業。2020年同大学大学院システム情報工学研究科コンピュータサイエンス専攻博士前期課程修了。同年同大学院システム情報工学研究群情報理工学位プログラム博士後期課程在学中。ヒューマンインタフェースに関する興味を持ち、現在は視線に基づく操作手法に関する研究を行う。ACM, IEEE 各学生会員。

山中 祥太



2016年に明治大学大学院にて博士号を取得(工学)。2015年より日本学術振興会特別研究員DC2, 2016年より日本学術振興会特別研究員PDおよび明治大学総合数理学部客員研究員を勤めたのち、2017年にヤフー株式会社に入社(Yahoo! JAPAN 研究所, 主任研究員), 2018年より同社上席研究員, 現在に至る。ユーザインタフェース研究, 特に運動性能のモデル化に興味を持つ。

志築 文太郎 (正会員)



1971年生。1994年東京工業大学理学部情報学科卒業。2000年同大学大学院情報理工学研究科数理・計算科学専攻博士課程単位取得退学。博士(理学)。現在、筑波大学システム情報系教授。ヒューマンインタフェースに関する研究に興味を持つ。日本ソフトウェア科学会, ACM, IEEE Computer Society, 情報処理学会, ヒューマンインタフェース学会各会員。