

前腕筋電位によるフィンガージェスチャ識別 — 把持状態での訓練時間短縮を目指して —

鈴木 拓也[†] 武井 彩佳[‡] 山名 早人[§]

[†] 早稲田大学基幹理工学部 〒169-8555 東京都新宿区大久保 3-4-1

[‡] 早稲田大学大学院基幹理工学研究科 〒169-8555 東京都新宿区大久保 3-4-1

[§] 早稲田大学理工学術院 〒169-8555 東京都新宿区大久保 3-4-1

E-mail: {takuya-suzuki-ph, ayaka, yamana}@yama.info.waseda.ac.jp

あらまし 近年広く普及しているボタンやタッチパネルなどのデバイスは、デバイスを持って入力することを想定している。このため、両手が塞がっている状況下ではその入力が困難となる。一方、両手が塞がっている状況においても入力を可能とする方法として、前腕筋電位を用いたフィンガージェスチャがある。しかし、前腕筋電位を用いたフィンガージェスチャ識別に関する既存手法では、ジェスチャを学習するために多くの訓練データを必要とする。ジェスチャの仕方は人や把持する対象物によって変化するため、対象物が変わる度に多くの訓練データの取得が必要となりユーザビリティが悪い。このため、少ない訓練データでの学習が必要とされている。本稿では、訓練として各ジェスチャを1回～数回行なうだけで前腕筋電位を用いたフィンガージェスチャの識別を可能とする手法を提案する。具体的には、マハラノビス距離を用いることで識別困難なジェスチャを棄却すると共に、複数のk近傍法を同時に用いることで精度向上を行なう。実験では、鞆や傘を把持した状態で「人差し指、中指、薬指、小指のいずれかの指をさらに曲げるように力を加えた状態を維持するフィンガージェスチャ」を被験者16人に対して実施した。各ジェスチャ1回ずつ訓練した場合、正解率として80.7%(鞆把持状態)を達成した。ただし、ジェスチャ未検出率は53.0%であり、正解率と検出率をさらに向上させる必要がある。

キーワード 入力インターフェース, 把持状態, フィンガージェスチャ, ジェスチャ識別, 筋電位

1. はじめに

現在、デバイスへの入力方法は多くの種類が存在している。ボタンやタッチパネルなどに代表される「デバイスを直接手で触れたり押したりする入力方法」や、視線入力や音声入力、ジェスチャ入力に代表される「デバイスがユーザの動作などを読み取る入力方法」が存在する。多くの入力方法は使用可能な状況と使用不可能な状況がある。様々な制約がある状況においても利用できる入力方法が必要とされている。

本稿では、ユーザに課される制約を「公共の場において物を把持していること」とする。公共の場では、自宅と異なり、把持した物を置くことができない場合があるため、把持した状態での入力が有効である。音楽プレイヤーなどの音声出力を行なうデバイスやスマートグラスなどのヘッドマウントディスプレイのようなハンズフリーデバイスであれば、把持状態でもユーザはデバイスの出力を受け取ることができる。しかし、把持状態において、ボタンやタッチパネルなどでの操作はできない。把持状態での使用が可能である既存入力手法は音声入力や視線入力、ジェスチャ入力がある。

音声入力は Apple や Google などのサービスが存在

する。視線入力は角膜反射法¹や眼電位²を用いて視線方向を推定する方法が実用化されている。しかし、公共の場では、周囲の騒音や光などのノイズ、静かにすべきといったマナーの問題、プライバシーの問題が存在する。さらに、視線入力では、よそ見による事故の可能性がある。ジェスチャによる入力は、足[1][2][3]や舌[4][5]、指[6]を使用した手法が存在する。しかし、足を用いたジェスチャは足元が不安定な場合や歩行時には使用できない。舌でのジェスチャは、電極を直接貼り付ける必要があり、実用的な段階には至っていない。指を用いたジェスチャは、既存手法では入力を行なう際には入力内容と入力タイミングをシステムに伝えるために両手を必要とする。また、各把持対象物で70秒以上かけて訓練データを得る必要がある[6]。指を用いたジェスチャの仕方は把持対象物によって異なるため、把持している物での訓練データが存在しない場合、すぐジェスチャ入力を行なうことができない。

指を用いたフィンガージェスチャは周囲の環境や歩行などの影響を受けづらい。また、Myo アームバンド³という前腕筋電位を測定するセンサが実用化されている。つまり、前腕筋電位センサを用いたフィンガージェスチャによる入力は、公共の場において物を把持

¹ <https://www.tobiipro.com/ja/product-listing/tobii-pro-glasses-2/>

² <https://jins-meme.com/ja/products/es/>

³ www.myo.com

した状態でも利用可能である。しかし、既存研究[6]では両手を必要とすることと、訓練データを得るための時間が長いことが問題点であり、改善すべき点である。

本稿では、前腕筋電位を用いて片手でのフィンガージェスチャを識別することを対象とし、短時間の訓練で得たデータを用いた場合でも、識別が困難なジェスチャを無視することで高い正解率を達成することを目的とした手法を提案する。誤識別を減らすために、複数の k 近傍法を用いてジェスチャを分類し、マハラノビス距離を用いた棄却判定を行ない、訓練データと特徴量が異なる識別が困難なデータを識別対象から除外する。さらに、直近の複数の棄却判定結果で投票を行ない、システムの出力を決定する。

本稿では次の構成をとる。まず、第2節で関連研究について述べる。第3節で提案手法の説明を行ない、第4節で評価実験とその結果、考察を説明する。最後に、第5節でまとめを行なう。

2. 関連研究

本節では、把持状態やフィンガージェスチャに関する研究を紹介する。

2.1. 把持状態におけるジェスチャ入力

把持状態におけるジェスチャ入力には、足や舌、指を用いたジェスチャが存在する。

足を用いたジェスチャ入力では、ズボンのポケットに入れたスマートフォンの加速度センサで足首やつま先を支点とした足の動作を識別する手法[1]や、脚に加速度センサを取り付けて、足を蹴り出す動作や円を描く動作を識別する手法がある[2]。また、靴下の下部に圧力センサを取り付けて、足への体重のかけ方を識別する手法がある[3]。[3]では、予め日常動作のデータを収集することで、歩行などの動作で誤って識別されることを回避できる。しかし、歩行時に入力することはできないため、入力を行なうには立ち止まる必要がある。

舌を用いたジェスチャ入力では、下顎の下部に筋電位センサを取り付けることで、舌を動かした時の筋肉の状態を識別する手法があり[4][5]、90%以上の正解率を達成している。しかし、センサの取り付け位置が目立つことや、電極を直接貼り付けることが欠点である。

指を用いたジェスチャ入力では、コップや鞆を把持した状態でのフィンガージェスチャの識別手法がある[6]。フィンガージェスチャは歩行時でも入力が可能であり、前腕筋電位を測定するセンサは既に実用化されている。足や舌を用いたジェスチャと比較し、フィンガージェスチャはより多くの状況で使用可能である。

2.2. フィンガージェスチャの識別手法

フィンガージェスチャを識別する手法で用いられるセンサは、加速度センサなどの慣性センサ[7]やカメ

ラ[8][9]、曲げセンサ[10]、前腕筋電位センサ[6][11][12]がある。これらのセンサの比較を述べる。

慣性センサを用いる手法では、フィンガージェスチャを識別するためには、指先に付ける必要がある。把持状態で入力を行なうためには、予めセンサを付けておく必要がある。指先にセンサを付けることで、非把持状態における指の動きを阻害することになる。

カメラを用いる手法では、閉塞と呼ばれる把持対象物によって指が映らない問題や、ユーザの周辺にいる他者のプライバシーの問題が生じる。

曲げセンサを用いる手法では、手の甲に曲げセンサを取り付けることで、指の曲げ具合を識別することができる。しかし、慣性センサと同様に、非把持状態において手の動きを阻害することなどの問題が生じる。

前腕筋電位センサを用いる手法では、慣性センサや曲げセンサと異なり、手の動きを阻害しない。また、カメラと異なり、閉塞や他者のプライバシーの問題が生じない。把持状態でのフィンガージェスチャの識別に適したセンサは、筋電位センサである。

2.3. 比較対象となる手法の説明

Saponas らが提案した把持状態におけるフィンガージェスチャの識別手法[6]について説明する。右手でジェスチャを行ない、左手で入力タイミングを「握るジェスチャ」により指定する。システムは常に右手のジェスチャを分類し、左手のジェスチャが行なわれたら、直近の3回連続して分類されたジェスチャラベルを出力とする。右手のジェスチャセットは、人差し指、中指、薬指、小指のいずれかで持ち手を押す動作である。

予めボールを握った時とリラックスした時のデータを取得、各チャンネルの値が0~1の範囲に収まるように正規化する。実験では把持対象物をコップや1kg程度の鞆とし、各ジェスチャを25回ずつ行なって訓練データ及びテストデータを得る。コップでは65%、鞆では86%の正解率である。鞆を把持した時のデータを用いて、訓練回数を減らした時の評価も実施している。各ジェスチャ10回ずつ、約70秒間の訓練データでの正解率は80%であり、各ジェスチャ1回ずつ、約7秒間の訓練データでの正解率は60%程度であった。

この手法の問題点は以下の2点である。

1. 片手しか使えない場合は入力を行なえない。
2. 把持対象物で70秒以上かけて訓練データを得る必要がある。

上記2つの問題点を解決することで、Saponas らの手法[6]と比べて、さらに多くの状況で使用できるようになる。

3. 提案手法

本節では提案手法について説明する。

表 3.1 使用する用語の定義

用語	定義
生データ	Myo アームバンドの 1 回のサンプリングで得た 8 チャネルの筋電位データ。
生データ列	時系列を保ったままの連続した複数の生データ。
セグメント	時系列を保ったままの連続した一定個数の生データで、特徴量を抽出する単位。
訓練用データ	評価において、訓練に用いるジェスチャラベル付きのセグメントの集合。
訓練用データセット	1 個以上の訓練用データのまとまり。
テスト用データ	評価において、テストに用いるセグメントの集合と 1 個の正解で構成されるジェスチャラベルの組。
テスト用データセット	1 個以上のテスト用データのまとまり。

本研究では、Myo アームバンド³の 8 チャネルの 200[Hz]の筋電位センサと、Bluetooth Low Energy(BLE)を使用する。表 3.1 に用語の定義を示す。

3.1. 解決すべき問題点と解決方法

短時間の訓練で得た訓練データを用いて、把持状態における片手でのフィンガージェスチャを識別することが提案手法の目的である。課せられる制約は、訓練を行なう時間が短いこととシステムが常にジェスチャを識別対象とすることである。

以上を実現する上での問題点を以下に示す。

- 問題 1. 訓練時間が短く、ジェスチャの実施回数が少ないため、他のジェスチャと誤識別する
- 問題 2. ジェスチャ実施タイミングが与えられず、ジェスチャ未実施時に誤識別する

ここで、問題 1 は、多種多様な把持対象物に対応するために訓練時間の短縮が求められることに起因する問題である。また、問題 2 は、[6]の手法のようにジェスチャタイミングを与える手とジェスチャを行なう手を別々にせず、片手のみで行えるようにしてユーザビリティを高めることに起因して発生する問題である。

この 2 つの問題に対して、本提案手法では以下の 3 つの手法を用いる。

1. 複数の k 近傍法を用いた分類
問題 1 を解決する

2. マハラノビス距離を用いた棄却判定
問題 1 と問題 2 を解決する
3. 直近の複数の棄却判定結果を用いた投票
問題 1 と問題 2 を解決する

3.2. ジェスチャセットとラベル

ジェスチャセットは、「人差し指」、「中指」、「薬指」、「小指」のいずれかの指をさらに曲げるように力を加えた状態を維持することで構成される。各ジェスチャにそれぞれラベルを設定する。また、把持対象物を自然に把持した状態を「リラックス」とする。識別結果においては、「不明」ラベルを用いる。「不明」ラベルは、分類や棄却判定、直近の棄却判定結果の投票で、結果が棄却される場合に使用される。

3.3. 全体の流れ

訓練データを取得する処理とユーザのジェスチャを識別する処理までの流れを図 3.1 に示す。行なわれたジェスチャを識別する処理の流れを図 3.2 に示す。

訓練データを取得する処理は、システムがユーザに行なうジェスチャの指示を各ジェスチャ 1 回ずつ、それぞれ 2 秒間ずつ連続して伝える。ユーザは指示に従ってジェスチャを行なう。システムは、行なわれたそれぞれジェスチャの末尾 1 秒間の生データを対象として正規化後の特徴量を抽出する。

入力されたジェスチャを識別する処理では、得られた生データから常に正規化後の特徴量を抽出する。抽

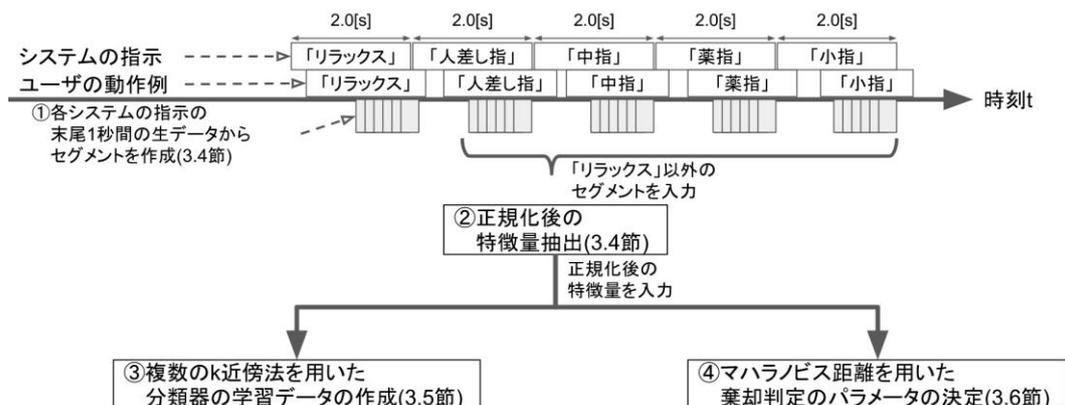


図 3.1 訓練データを取得する処理の流れ

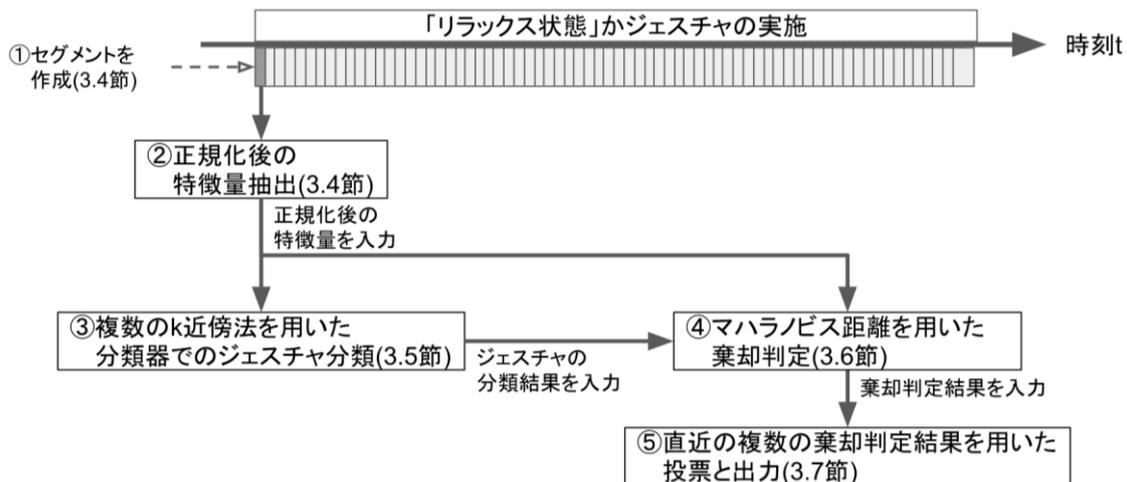


図 3.2 ジェスチャを識別する処理の流れ

出した特徴量をもとに k 近傍法での分類を行なう。また、マハラノビス距離を用いて棄却判定をする。棄却される場合は、「不明」ラベルを付与する。最後に、直近の棄却判定結果を用いてシステムの最終的な識別結果を決定する。

3.4. 使用する特徴量

Myo アームバンドのサンプリング周波数は 200[Hz] であり、毎秒 200 個の 8 チャンネルの生データが得られる。得られる生データからスライディングウィンドウを用いて生データを切り出す。1 つのウィンドウをセグメントと呼び、セグメント毎に特徴量を抽出する。各セグメントは N 個の生データで構成され、ウィンドウ長は $5 \times N$ [ms] とする。セグメント間のウィンドウのスライド幅は w 個分の生データで、 $5 \times w$ [ms] とする。

1 つのセグメントに含まれる生データ $x_i (1 \leq i \leq N)$ から、RMS (二乗平均平方根) と振幅スペクトル、各チャンネルの絶対値の最大値を特徴量として使用する。RMS、振幅スペクトル、各チャンネルの絶対値の最大値ごとに訓練時に得た特徴量を用いて 0~1 の範囲に収まるように正規化を行なう (これはユーザ毎に行う)。

RMS (2 乗平均平方根) は式 (3.1) で定義される。

$$\text{RMS} = \sum_{i=1}^N \sqrt{\frac{x_i^2}{N}} \quad (3.1)$$

振幅スペクトルは、セグメントの末尾 8 つの生データを用いて、式 (3.2) のハニング窓で前処理を行なってから、25[Hz] ビンで FFT を行ない、各チャンネル 4 個のビンを使用する。

$$w(i) = 0.5 - 0.5 \cos\left(2\pi \frac{i}{7}\right), (0 \leq i \leq 7) \quad (3.2)$$

セグメントのウィンドウ長は 250[ms]、50 個分の生データとした。RMS は N が小さいと値が安定しづらい。しかし、 N を大きくするとより過去の生データを利用することになり、ジェスチャを実施してから識別され

るまでの時間が長くなる。1 回のジェスチャに 1 秒間かけて行なうこととしたため、ウィンドウ長を 250[ms] とした。ウィンドウのスライドの幅は、訓練データを取得する処理では、より多くの訓練データを得るために 5[ms]、 $w = 1$ とした。ユーザのジェスチャを識別する処理におけるウィンドウのスライド幅は、システムの識別速度などを考慮し 50[ms]、 $w = 10$ とした。

3.5. 複数の k 近傍法を用いた分類

各ジェスチャを 1 回ずつ行なって得た訓練データでは、訓練回数が少ないため、SVM やランダムフォレストなどの分類器では正しい識別境界を得にくい。そこで、提案手法では k 近傍法を用いる。距離指標はコサイン類似度を使用する。

コサイン類似度を使用するため、「リラックス」の状態の特徴量は識別率を低下させる原因になる可能性がある。これを防ぐため、「リラックス」のセグメントは除外し、ジェスチャセットのジェスチャのセグメントを用いて訓練データを作成する。k 近傍法のパラメータ k は 1~151 とし、151 個の分類器の分類結果が一致した場合のみ、分類結果のラベルを付与し、1 つでも分類結果が異なれば「不明」ラベルを付与する。複数の分類器を使用する理由は、複数のジェスチャラベルの学習データの分布が重複している場合などにおいて、誤識別を回避するためである。

3.6. マハラノビス距離を用いた判定

マハラノビス距離を用いた棄却判定は、セグメントから抽出した特徴量の RMS が、各ジェスチャラベルの学習データが持つ特徴量 RMS と近いかどうかを判定し、誤識別の可能性があると判定した場合に「不明」ラベルを出力して棄却するための処理である。

複数の k 近傍法を用いた分類では、入力セグメントの特徴量が最も近い学習データのジェスチャラベルを選択するため、常に認識させた場合、ジェスチャを行っていない時にもジェスチャを行なったと識別して

しまう．そこで，入力セグメントの特徴量が学習データから近いかどうか判定し，ジェスチャの実施の有無を判定する．

各セグメントの特徴量の RMS はジェスチャごとに正規分布に従うと仮定し，式(3.3)に示す RMS のマハラノビス距離を 2 乗した値 $l_m(g, i)$ を用いる．ただし， g はジェスチャラベルを表し， i はジェスチャラベルごとの訓練回数の番号であり，訓練回数 n に対して $1 \leq i \leq n$ を満たす．また， \mathbf{x} は入力セグメントの RMS を表し， $\boldsymbol{\mu}(g, i)$ 及び $\Sigma(g, i)$ は使用する学習データでの RMS の平均ベクトル及び共分散行列を表す．

$$l_m(\mathbf{x}, g, i) = (\mathbf{x} - \boldsymbol{\mu}(g, i))^T \Sigma(g, i)^{-1} (\mathbf{x} - \boldsymbol{\mu}(g, i)) \quad (3.3)$$

次に棄却判定の条件について説明する．算出した $l_m(g, i)$ に対して許容閾値 θ と棄却閾値 $\tilde{\theta}$ を用意する．複数の k 近傍法での分類結果のジェスチャラベルを G とした時，以下の 2 つの条件のどちらかを満たす場合，「不明」ラベルを出力し，結果を棄却する．ただし， \widetilde{GS} はジェスチャセットから G を除外したジェスチャの集合である．

1. $\forall i, l_m(G, i) > \theta$ を満たす．
2. $\exists g \in \widetilde{GS}, \exists i, l_m(g, i) < \tilde{\theta}$ を満たす．

許容閾値 θ と棄却閾値 $\tilde{\theta}$ の値によっては，多くのセグメントを棄却してしまい，識別対象となるセグメントの数が減少するため，入力に長い時間がかかる．学習データが限られていることから，許容閾値 θ を小さい値で固定とすると，力の加え方が訓練時と少しだけ異なる場合でも棄却される可能性がある．そこで，識別対象とするセグメントの数を減らし過ぎないようにするために，状況に応じて許容閾値 θ を緩和する必要がある．

具体的には， k 近傍法の分類器が連続して同じジェスチャラベルを分類結果とする場合は，許容閾値 θ を緩和する．この閾値の緩和倍率を変数 θ_m とする． θ の初期値を θ_0 とすると， $\theta = \theta_m \theta_0$ となる．緩和倍率 θ_m は，複数の k 近傍法の分類結果が連続して同じラベルであれば線形増加し，複数の k 近傍法の分類結果が 1 秒間同じジェスチャラベルである場合，最大緩和倍率 θ_M となるように設定した．ただし， $\theta_M > 1$ とし， $1 \leq \theta_m \leq \theta_M$ とした．複数の k 近傍法の分類結果が直前のセグメントの分類結果と異なる場合は， θ_m を 1 に戻す．

許容閾値 θ の初期値 θ_0 と棄却閾値 $\tilde{\theta}$ は，カイ二乗分布を用いて学習データの 99.9% を占める値とした．特徴量の RMS は 8 次元であるため $\theta_0 = \tilde{\theta} = 26.12$ とした．最大緩和倍率 θ_M は実験で得たデータの一部を用いてグリッドサーチを行ない，20 とした．

3.7. 分類結果の投票とシステムの出力の決定

ジェスチャ切り替え時などに瞬間的にジェスチャ

を誤識別することがある．このような誤識別を回避するために，直近の複数の分類結果を用いて 2 段階の投票を行なう．投票は直近 16 個のセグメントの棄却判定結果を使用し，過半数を占めるジェスチャラベルを出力とする．ただし，過半数を占めるジェスチャラベルが 1 つも存在しない場合は棄却するため，「不明」ラベルを出力とする．

4. 評価方法

本節では，提案手法と従来手法の比較および評価，その結果と考察を述べる．

4.1. 比較対象の既存手法

提案手法との性能比較のために，ベースラインとして Saponas ら [6] の手法を用いる．ただし，使用するセンサが異なるため，生データからの切り出し方など，以下で述べる変更を加え実装した．

使用するセンサは，[6] はサンプリング周波数が 2048[Hz]，AD 変換の分解能が 24bit である BioSemi Active Two である．今回の実装は，サンプリング周波数が 200[Hz]，AD 変換の分解能が 8bit である Myo アームバンドである．右前腕のセンサの電極数は [6] では 6 チャンネルで，実装では 8 チャンネルである．[6] での左手のジェスチャは，実装ではキーボードを用いたキー入力とした．この変更による [6] の手法の性能低下は生じない．特徴量の抽出では，[6] ではセグメントのウィンドウ長とスライド幅は約 31[ms] であり，毎秒 32 個のセグメントが作成される．実装では，セグメントのウィンドウ長を 40[ms]，スライド幅を 30[ms] とした．FFT のビン幅は，[6] は 13 個のビン幅 32[Hz] のビンを使用し，0~416[Hz] まで使用している．しかし，Myo アームバンドのサンプリング周波数は 200[Hz] であるため，100[Hz] までしか使用できない．[6] でのビン幅に近づけるためにビン幅を 25[Hz] とし，ビン数は 4 個である．

4.2. 実験内容

被験者は大学生および大学院生の男女 16 名 (女性：5 名，男性：11 名) である．年齢は 19~34 歳で，平均は 22.4 歳，標準偏差は 3.8 歳である．右前腕の筋電位を測定するために，Myo アームバンドを使用した．生データの記録や被験者への指示を出すために PC (OS : Windows7, CPU : Intel Core i3-4000M, 2.4[GHz], 2[Core], RAM : 4[GB]) と 24 インチディスプレイを用いた．この PC と Myo アームバンドは Bluetooth Low Energy で接続する．

被験者が把持する物は，総重量 1.5kg の鞆と 0.5kg 程度の傘である．傘は差した状態とした．鞆や傘の持ち方は被験者が持ちやすい持ち方とした．実験風景を図 4.1 に示す．鞆などの把持の仕方や指の曲げやすさなどが人によって異なるため，各ジェスチャは被験者が

無理なく行なえる範囲で、可能な限り対象の指のみ力を入れて行なってもらった。

実験を行なう前に、Myo アームバンドを右前腕の肘から手首への 1/4 あたりの位置に取り付けた。取り付けた後は Myo アームバンドの位置を変更しないように指示をした。実験は、以下に示す 3 種類の実感から構成される。実験 1 は正規化処理を行うための予備実験であり、以降の 2 つの実験により 2 種類の把持対象物(靴、傘)を対象とした評価を行った。なお、どちらで先に実験を行なうかは、被験者間で均等になるようにした。また、把持の仕方やジェスチャの仕方は可能な限り変えないように指示した。図 4.2 に被験者に対する指示画面例を示す。

実験 1(正規化処理用データの収集)：把持しない状態と把持対象物の持ち手を強く握ったまま状態のデータを 4 秒間取得する実験である。本実験で取得したデータは、従来手法での各チャンネルの値を 0~1 の範囲に収める正規化の処理に用いる。

実験 2(訓練用データの収集)：提案手法における訓練データの収集の方法と同様の方法でデータを収集する。以下に示す内容(①~⑤)を 1 セットとし、10 セット行なう。ただし、ジェスチャの順番はできるだけ偏りがないように予め決めておき、各セットの開始前にそのセットでのジェスチャの順番を提示する。

- ①被験者は、「リラックス」の状態になり、左手でエンターキーを入力する。③における指示が表示されるまで、「リラックス」の状態を維持する。
- ②システムは、エンターキーの入力を受け付けてから 1 秒間何もせず待つ。その後、③へ進む。
- ③システムは、被験者に次に行ってもらおうジェスチャを 2 秒間指示する。ただし、既に全てのジェスチャを 1 回ずつ指示しており、次に指示するジェスチャが存在しない場合は終了する。
- ④被験者はシステムからの指示を確認し次第、指示されたジェスチャを 1 回行ない、再び③における指示が表示されるまでその状態を維持する。
- ⑤システムが③において被験者に指示を出してから 2 秒間経過した後、再び③から続行する。

実験 3(テスト用データの収集)：リラックス状態と各ジェスチャを交互に行なう内容を(①~⑤)を 1 回行



図 4.1 実験風景(左：靴、右：傘)

なう。ジェスチャを行なう順番は、同じジェスチャが連続せず、できるだけ偏りのないように、異なるジェスチャを 1 回ずつ交互に行なうような順番を予め決めておいて実施した。具体的には、『人差し指』、『中指』、『小指』、『薬指』、『中指』、『薬指』、『小指』、『人差し指』、『…』といった順番である。また、各ジェスチャ 24 回ずつ行なう。

- ①システムが被験者に「リラックス」の状態になるように指示する。
- ②被験者は「リラックス」の状態になり、0.5 秒程度待ってからエンターキーを入力する。システムは、エンターキーが入力されてから 0.5 秒後に③に進む。ただし、次の③でのシステムの指示があるまで被験者は「リラックス」の状態を維持する。
- ③システムは、次に行なってもらおうジェスチャセットの 4 種類のジェスチャの内の 1 種類のジェスチャを指示する。ただし、既に 24 回指示を出したジェスチャは除外する。また、既にジェスチャセットの各ジェスチャを 24 回ずつ指示した場合は終了する。
- ④被験者は③でのシステムの指示を確認し次第、ジェスチャを行ない、0.5 秒程度待ってからエンターキーを入力する。次の①でのシステムの指示があるまで被験者はジェスチャを行なった状態を維持する。
- ⑤システムは、④におけるエンターキーの入力を受け付けてから 1 秒後に、再び①から続行する。

なお、「被験者が指示と異なるジェスチャを行なった場合」や「実験 2 において、指示が表示されてから被験者がジェスチャを行なうのが遅れた場合」は、実験 1 では実施していたジェスチャから、実験 2 では実施中のセットから、実験 3 では直前の「リラックス」の指示から、それぞれ取得し直した。

4.3. 評価用のデータセットの作成方法と評価指標 訓練用データセット作成方法

訓練回数が少ない場合について評価する時、使用する訓練データによって結果に違いが生じる。訓練データによる結果への影響を小さくするために、各訓練回数について、実験 2 の各セットで得た訓練データの組み合わせ方法を変更して複数回評価を行なう。ただし、全ての組み合わせで評価を行なうと評価に要する時間が膨大となるため、組み合わせ方を限定した。



図 4.2 実験参加者への指示画面
(図では「中指」のジェスチャを指示)

$i(1 \leq i \leq 10)$ セット目の訓練で得た、各指示の末尾 1 秒間の生データ列を $D_{Tr}(i, g)$ と定義する。ただし、 GS_{Tr} を {「リラックス」, 「人差し指」, 「中指」, 「薬指」, 「小指」} とし、 $g \in GS_{Tr}$ とする。また、訓練用データセットを $DS_{Tr}(n, S(n))$ と定義する。 n は使用する訓練データの訓練のセット数を表す変数で、 $n \in \{1, 2, 3, 5, 7, 10\}$ である。 $S(n)$ は使用する訓練データのセット番号を示す変数 $s_k (1 \leq k \leq n)$ の集合である。訓練用データセットは $DS_{Tr}(n, S(n)) = \{D_{Tr}(s_k, g) | s_k \in S(n), g \in GS_{Tr}\}$ である。

評価に要する時間が軽減するために、各セット数について、 $S_j(n), (1 \leq j \leq 10, n \in \{1, 2, 3, 5, 7\})$ 及び $S_1(10)$ を用いる。ただし、 $S_j(n)$ 及び $f(x)$ は以下に示す通りである。

$$S_j(n) = \{D_{Train}(f(j+t)) | 0 \leq t \leq n-1\}$$

$$f(x) = \begin{cases} x & (1 \leq x \leq 10) \\ x-10u & (10u < x \leq 10(u+1), 1 \leq u) \end{cases}$$

テスト用データセット作成方法

ジェスチャの実施タイミングをシステムに与えるテスト用データセット $DS_{part}(i)$ と、与えないテスト用データセット $DS_{whole}(i)$ について説明する。ただし、 i は実験 3 でのジェスチャの指示の番号を示し、 $1 \leq i \leq 96$ である。 $DS_{part}(i)$ は実験 3 における 1 回の③から⑤の終了までの期間の生データで作成し、 $DS_{whole}(i)$ は 1 回の③から次の②の終了までの期間の生データで作成する。ただし、 $DS_{whole}(i), (1 \leq i \leq 8, 89 \leq i \leq 96)$ を提案手法のパラメータ決定に用いたため、 $DS_{part}(i), (9 \leq i \leq 88)$ と $DS_{whole}(i), (9 \leq i \leq 88)$ を評価で使用する。

識別結果の決定方法と評価指標

$DS_{part}(i)$ では、従来手法および提案手法ともに、最後のシステムの出力を識別結果とする。出力が 1 つも無い場合は、「不明」ラベルとする。 $DS_{whole}(i)$ では、全ての識別したジェスチャラベルの集合を識別結果とする。

各テスト用データの識別結果に対して、「正解」、「誤識別」、「未検出」のいずれか 1 つを付与する。付与する基準を以下に示す。ただし、システムが被験者に指示するジェスチャを「正解ジェスチャ」、ジェスチャセットから「正解ジェスチャ」を除外したジェスチャ集合を「不正解ジェスチャ」とする。

- 「正解」: 「正解ジェスチャ」を識別し、「不正解ジェスチャ」のジェスチャを識別しなかった場合
- 「誤識別」: 「不正解ジェスチャ」のいずれかのジェスチャを識別した場合
- 「未検出」: 「不明」のみ識別した場合

テスト用データセット $DS_{part}(i)$ 及び $DS_{whole}(i)$ において、「正解」の個数を C 、「誤識別」の個数を M 、「未検出」の個数を U 、評価に使用したテスト用データの個数を $N = 80$ とすると、 $C + M + U = N$ が成立する。行なったジェスチャに対する正解率と誤識別率、未検出率、「未検出」を除外した時の正解率を用いて評価する。

4.4. 結果と考察

鞆及び傘での「未検出」を除外した時の正解率を図 4.3 及び図 4.4 に、未検出率を図 4.5 及び図 4.6 図 4. に示す。また、鞆及び傘での「未検出」を除外した時の正解率から分かることを表 4.1 に示す。

「未検出」を除外した、検出できたジェスチャに対

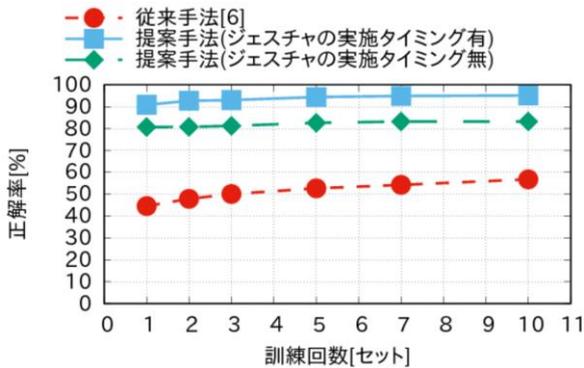


図 4.3 鞆での「未検出」を除外した正解率

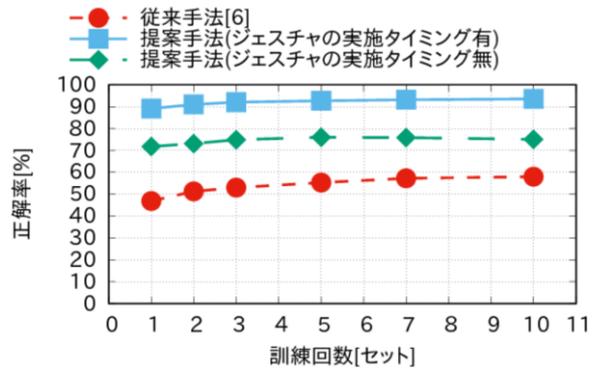


図 4.4 傘での「未検出」を除外した正解率

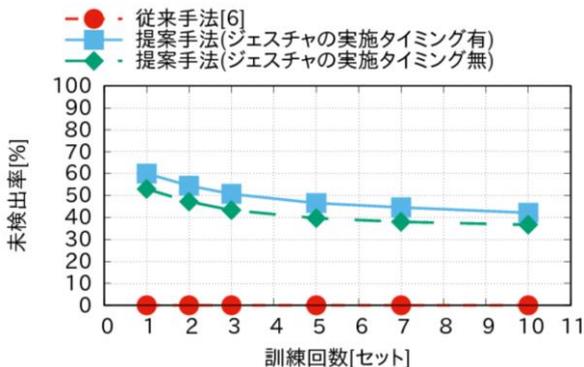


図 4.5 鞆での未検出率

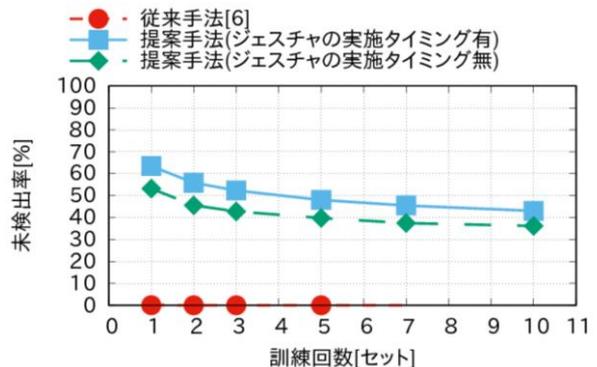


図 4.6 傘での未検出率

表 4.1 「未検出」を除外した正解率に関する結果から分かること

観点	分かること
従来手法について	● 訓練回数が増えるにつれて靴では最大 12.3%, 傘では最大 11.1%上昇する.
ジェスチャの実施タイミングを与えた時の提案手法について	● 把持対象物によらず, 訓練回数が増えるにつれて上昇する. ● 訓練回数 1 セットでは靴は 90.9%, 傘は 89.1%であり, 10 セットでは靴は 95.1%, 傘は 93.6%である.
ジェスチャの実施タイミングを与えない時の提案手法について	● 把持対象物によらず, 訓練回数が増えるにつれて上昇する. ● 訓練回数 1 セットでは靴は 80.7%, 傘は 71.8%であり, 10 セットでは靴は 83.2%, 傘は 75.0%である.
ジェスチャの実施タイミングを与えた時の従来手法と提案手法の比較	● 訓練回数によらず, 提案手法は従来手法よりも靴では 38.3~46.3%, 傘では 35.6~42.3%高い. ● 訓練回数が増えるほど差が小さくなる.
提案手法におけるジェスチャの実施タイミングの有無についての比較	● 訓練回数によらず, ジェスチャの実施タイミングを与えた方が靴では 10.2~11.9%, 傘では 16.7~18.5%高い. ● 訓練回数が増えるほど差が大きくなる.

する正解率は, ジェスチャの実施タイミングを与えない場合でも, 提案手法は従来手法を上回っている. しかし, さらに正解率を向上させる必要がある.

提案手法の未検出率は 50~60%である. 実際に提案手法のシステムを用いる場合, システムが識別できるようにユーザが適応することも考えられる. しかし, 検出率が低いと, 入力を行なうのに必要なジェスチャの実施回数が増えてしまう. したがって, 検出率を向上させる必要がある.

5. まとめ

把持状態ではボタンやタッチパネルを用いて入力を行なうことができない. これに対して本稿では, 把持状態において, 短時間で得た訓練データに対しても高い精度で識別することを目指し, 片手で行なうフィンガージェスチャの識別手法を提案した. 具体的には, 把持状態におけるフィンガージェスチャの誤識別を減らすために, 複数の k 近傍法を用いた分類と, マハラノビス距離を用いた棄却判定, 直近の複数の棄却判定結果を用いた投票を適用した. 評価実験の結果, 訓練回数 1 回の時, ジェスチャの実施タイミングを与えない場合, 正解率 80.5%(靴を把持した場合)を達成し, 従来手法よりも高い結果となった.

主な今後の課題として, 1) 靴や傘以外の物体を把持した時や歩行時における識別精度を評価すること, 2) 誤識別率や未検出率を低下させ, 正解率を向上させる手法を追加すること, 3) 訓練時間の短さを維持しつつ, オンライン学習を用いて学習データを増やすこと, などが考えられる.

参考文献

- [1] J. Scott, D. Dearman, K. Yatani and K. N. Truong, "Sensing Foot Gestures from the Pocket", Proc. of ACM UIST'10, pp. 199-208.
- [2] J. Alexander, T. Han, W. Judd, P. Irani and S. Subramanian, "Putting Your Best Foot Forward:

Investigating Real-world Mappings for Foot-based Gestures", Proc. of CHI'12, pp. 1229-1238.

- [3] K. Fukahori, D. Sakamoto and T. Igarashi, "Exploring Subtle Foot Plantar-based Gestures with Sock-placed Pressure Sensors," Proc. of ACM CHI'15, pp. 3019-3028.
- [4] M. Sasaki, T. Arakawa, A. Nakayama, G. Obinata and M. Yamaguchi, "Estimation of tongue movement based on suprahyoid muscle activity", Int. on IEEE MHS'11, pp. 433-438.
- [5] Q. Zhang, S. Gollakota, B. Taskar and R. P. N. Rao, "Non-intrusive Tongue Machine Interface," Proc. of CHI'14, pp. 2555-2558.
- [6] T. S. Saponas, D. S. Tan, D. Morris, R. Balakrishnan, J. Turner and J. A. Landay, "Enabling Always-available Input with Muscle-computer Interfaces", Proc. of ACM UIST'09, pp. 167-176.
- [7] K. Wolf, R. Schleicher, S. Kratz and M. Rohs, "Tickle: A Surface-independent Interaction Technique for Grasp Interfaces", Proc. of ACM TEI'13, pp. 185-192.
- [8] D. Kim, O. Hilliges, S. Izadi, A. Butler, J. Chen, I. Oikonomidis and P. Olivier, "Digits: Freehand 3D Interactions Anywhere Using a Wrist-worn Gloveless Sensor", Proc of ACM UIST'12, pp. 167-176.
- [9] L. Chan, Y. Chen, C. Hsieh, R. Liang and B. Chen, "CyclopsRing: Enabling Whole-Hand and Context-Aware Interactions Through a Fisheye Ring", Proc of ACM UIST'15, pp. 549-556.
- [10] J. Lin, C. Wang, Y. Huang, K. Chou, H. Chen, W. Tseng and M. Chen, "BackHand: Sensing Hand Gestures via Back of the Hand," Proc. of ACM UIST'15, pp. 557-564.
- [11] C. Amma, T. Krings, J. Böer and T. Schultz, "Advancing Muscle-Computer Interfaces with High-Density Electromyography", Proc. of ACM CHI'15, pp. 929-938.
- [12] F. Kerber, M. Puhl and A. Krüger, "User-independent Real-time Hand Gesture Recognition Based on Surface Electromyography", Proc. of ACM MobileHCI'17, p. 36:1--36:7.