2018 年度　修士学位請求論文

ポケット内のスマートフォンを利用した両足ジェスチャに関する研究

明治大学大学院先端数理科学研究科

先端メディアサイエンス専攻

田村　柾優紀

Master's Thesis

A Study on Gesture Recogniton by Both Feet with Smartphone in a Pocket

Frontier Media Science Program,

Graduate School of Advanced Mathematical Sciences,

Meiji University

Masayuki Tamura

概要

2019 年現在，目の前の電子機器を操作する状況では，主に手による操作と声による操作が考えられる．しかし，手が汚れておりデバイスに触りたくない状況や，赤子を抱っこしている状況などの手がふさがっている状況では，手による操作は行うことが出来ない．手がふさがっている状況では，音声認識による操作が考えられるが，音声認識では短時間に何回も操作が必要な場合には，何度も同じ言葉を発して操作する手間が存在する．そこで Igarashi ら[1]は，同じ操作を一度に何度も必要とする状況で音声に入力を利用可能とするため，タンギングと呼ばれる管楽器の演奏で舌を用いる奏法を用いた音声認識によるデバイス操作手法を提案し，その手間の問題を解決した．しかし，この手法では，電子書籍を読むための操作などには適していない．そのため，時間をおいてから何度も操作が必要になるような状況では，何度も音声認識をする手間が残っている．

音声認識と手による操作を行わない手法の一つとして，足による操作を行う研究が様々行われている．しかし，これまでに行われてきた多くの研究は専用のデバイスを用意する必要があるため手間が残る．一方で，多くの人が日常的に持ち歩くスマートフォンを用いたジェスチャ認識を検討している研究に，Scott らの研究[5]があげられる．その研究では，ポケット内のスマートフォンを用いて，スマートフォンを入れた側の足によるジェスチャの認識精度を検証している．しかし，前ポケットでの認識精度は 60.2%であり，認識精度は高くない．ここで，スマートフォンは日常において，様々な角度や向きでポケット内に入れられるため，データセット構築時とは違う角度や向きでポケットに入れられた際には，センサの軸の向きが変化する．その結果，認識精度に影響が出ることが考えられるが，Scott らの研究[5]ではその影響を考慮していない．また同研究では，スマートフォンを入れたポケット側の片足のみによるジェスチャ認識であったが，UI 設計において逆方向に同じ操作をすることで，逆の操作ができる UI 設計が多く行われている．例えば，電子書籍では画面の左側をタップすることでページが進み，右側をタップすることでページが戻る UI が実装されている．他にも，Web ページを閲覧する際には，上にスクロールをすることでページをさかのぼり，下にスクロールすることでページを読み進めることができる操作が実装されている．つまり，左足と右足で同じジェスチャをすることで逆の操作が実現できることで，より直感的な操作が可能になると期待される．また両足によるジェスチャ認識が可能となれば，片足によるジェスチャよりも表現できるジェスチャ種類の増加が期待される．

そのため，我々はポケット内のスマートフォンを利用した両足ジェスチャ認識を実現するための手法を提案する．そこで，これまでの我々が行った研究[6]では，まず特定の向きにおける両足ジェスチャの認識精度を検証するため，データセット構築を行い，認識精度を算出した．その結果，両足ジェスチャの認識精度を表す F 値が 0.93 であり，高い認識精度であった．そこで，そのデータセット構築での結果をもとにプロトタイプシステムを開発し，そのシステムを用いて実験協力者に電子書籍を読んでもらい，システムの評価を行っても

らった．しかし，使用実験を通してどのくらいの認識精度と感じたかという体感認識精度についてのアンケートでは，期待したほどの認識精度とは感じられていなかった．その要因としては，これまでの研究で提案した認識手法を用いた際の認識精度がユーザの期待したほどの認識精度ではなかったことがあげられる．他の要因としては，認識までにかかる時間の遅さがあげられる．実際，使用実験を通して，システムが両足ジェスチャを認識するまでにかかる時間がどうだったかについてアンケートで回答してもらったところ，8 人中 7 人の実験協力者が，システムが両足ジェスチャを認識するまでにかかる時間を遅いと感じている．つまり，認識までにかかる時間の遅さを認識精度の悪さと感じていたと考えられる．そこで，本研究では認識精度をより高めるために，特徴ベクトルや分類器を再検討し，認識手法を改善する．そして，認識時間の短縮においては，システム応答時間の限界時間以内に認識を行うため，ジェスチャを開始してからの使用フレーム数の減少を行う．その使用フレーム数の減少によって認識精度が下がることが考えられるため，ジェスチャを開始する直前のジェスチャデータを使用することで認識精度の向上を検討する．

Abstract

As of 2019, major ways to operate electronic devices in front of the eyes are manual operation and voice operation. However, in the situation where hands are dirty and you do not want to touch the device or you are holding a baby, you cannot do the manual operation. Therefore, when hands are occupied, operation by speech recognition can be considered. However, with speech recognition, when operations are required many times in a short period of time, it takes time and effort to say the same words repeatedly. To solve this, Igarashi et al.[1] proposed a device manipulation method with speech recognition using tonguing, a tongue-based rendition style in the performance of a wind instruments. However, even with this method, it is impossible to solve problems such that speech recognition is required for every single page of an e-book with 200 pages. This means that it still takes time and effort to use speech recognition many times when operations are required repeatedly after a certain lapse of time.

Then, as one of the methods in which neither speech recognition nor manual operation are performed, various studies on using feet for performing operations have been conducted. However, many of the previous studies require preparation of dedicated devices, which requires time and effort. On the other hand, Scott et al.[5] studied the gesture recognition using smartphones, which many people carry around on a daily basis. Their study investigated the recognition accuracy of the gesture of the foot on the side of the pocket in which a smartphone was put. However, the recognition accuracy in the front pocket was 60.2%, which was not high enough. Here, since smartphones can be placed at various angles and orientations in a pocket in everyday situations, the direction of the sensor's axis would change when they are placed in a pocket at angles and orientations that are different from those observed when the dataset was built. This may affect the recognition accuracy, but Scott et al 's research[5] did not consider its influence.

Therefore, the present study proposes a two-feet gesture recognition method considering the change in position of the smartphone in a pocket. In our previous research, we constructed the data set and calculated the recognition accuracy in order to verify the recognition accuracy of the two-feet gesture in a specific orientation. As a result, the F-measure representing the recognition accuracy of gestures by both feet was 0.93, which confirmed the recognition accuracy was high. Then, we developed a prototype system based on the result of the data set construction, and conducted an experiment in which the participants are asked to read electronic books using the system and evaluated the system. However, the post-experiment questionnaire reported that the recognition accuracy that the participants felt was not as high as it was expected . The reason for this would be that the recognition accuracy when using the

recognition method proposed in the previous research was not as high as it was expected by the users. Another possible factor is the slowness of the time for recognition. In the past research, we conducted a questionnaire about how long it took the system to recognize gestures of both feet during the experiments. As a result, it was revealed that 7 of 8 participants tended to feel that it took the system long time to recognize the gestures. This seems to suggest that the participants felt the slowness of the recognition as a bad recognition accuracy. Therefore, in this research, in order to further improve the recognition accuracy, we reviewed the feature vectors and classifiers and improved the recognition method. For recognition time reduction, since recognition is performed within the limited system response time, the number of used frames after starting the gesture is reduced. Also, since recognition accuracy may decrease due to a decrease in the number of the used frames, we considered the improvement of recognition accuracy by using gesture data just before starting it.

目次

第 1 章　はじめに

1.1. 現在のデバイス操作における課題

2019 年現在では，目の前にあるスマートフォンや，タブレット，テレビ，照明などのデバイスを操作する方法として，主に手による操作と，音声認識を用いた声による操作が考えられる．しかし，デバイスでレシピを見ながら料理をするときといった，手が汚れていてデバイスを汚さないように操作したい状況や，赤ちゃんを抱っこしていたり，荷物などを持っていて手が塞がっている状況などでは，手によるデバイス操作は向いていない．一方で，音声認識を用いた声による操作では，手を用いないため，手が塞がっている状況においてもデバイスを操作することができる．また音声認識を用いた声による操作は，部屋の電気を消す際や，特定のテレビ局に切り替える際など短期間に何度も行わない操作には有効である．しかし，音声認識は照明の明るさや，テレビの音量を段階的に変更するなど，短期間に繰り返し操作が必要な状況などにおいては，何度も同じ言葉を発して操作する必要があるため，手間が存在する．Igarashi らの研究[1]では，一度に同じ操作を何度も行う必要がある状況における声による操作の手間を解決するため，音声タンギングによるデバイス操作を提案している．しかし，電子書籍の読書中に読み終わってからページをめくりたい場合や，特定のページを探している際やリスニングの勉強中にある文章だけを何度も聞き返したい場合，スライドを用いたプレゼン中に話し終わったタイミングでスライドめくりをしたい場合，Web ページの閲覧において，少しずつページをスクロールしたい場合，見たいテレビ番組を探すためにチャンネルを変えるザッピングをしたい場合といった，一度に操作をするわけではないが，同じ操作が何回も必要となる際には音声認識での操作は手間である．また音声認識は，楽器を演奏している状況や，赤ちゃんを抱っこして寝かしつけている状況，電車に乗っている際にデバイス操作したい状況など，声を発することが困難で音声認識によるデバイス操作が向かない状況が存在する．

1.2. 足によるデバイス操作の研究

手による操作と音声認識による操作を用いずにデバイス操作を行う方法の一つとして，足によるジェスチャを用いたデバイス操作の研究が様々行われている．Matthies らの研究[2][3]では，中敷きに圧力センサを取り付けたデバイスで，足のジェスチャを認識させる手法を提案している．また，Fukahori らの研究[4]では，靴下に圧力センサを取り付けたデバイスにより，足のジェスチャを認識させる手法を提案している．これらの研究は，事前に中敷や靴下にセンサを取り付ける必要があるため，導入に手間がかかる．これに対し，新たにセンサを取り付けずに，身近なデバイスであるスマートフォンを用いて，足のジェスチャ認識

を行う研究が様々行われている． Scott らの研究[5]では，ポケット内のスマートフォンを用いて，スマートフォンを入れた側の足によるジェスチャ認識精度を検証している．しかし，前ポケットでの認識精度は 60.2%であり，認識精度は高くない．また，スマートフォンは日常において，様々な角度や向きでポケット内に入れられるため，ジェスチャのデータセット構築時とは違う角度や向きでポケットに入れられた際には，センサの軸の向きが変化し，認識精度に影響が出ることが考えられるが，その影響についての検証がされていない．また同研究では，スマートフォンを入れたポケット側の片足のみによるジェスチャ認識であったが，UI 設計において逆方向に同じ操作をすることで，逆の操作ができる UI 設計が多く行われている．例えば，電子書籍では画面の左側をタップすることでページが進み，右側をタップすることでページが戻る UI が実装されている．他にも，Web ページを閲覧する際には，上にスクロールをすることでページをさかのぼり，下にスクロールすることでページを読み進めることができる操作が実装されている．つまり，左足と右足で同じジェスチャをすることで逆の操作が実現できることで，より直感的な操作が可能になると期待される．また両足によるジェスチャ認識が可能となれば，片足によるジェスチャよりも表現できるジェスチャ種類の増加が期待される．

1.3. 本研究の目的と課題

我々はこれまでの研究[6]において，ポケット内のスマートフォンを利用した両足ジェスチャ認識手法を提案した．そして，両足ジェスチャのデータセット構築を行い，提案手法による両足ジェスチャの認識精度を検証した．その結果，個人のデータを使って認識させた場合には両足のジェスチャ認識の精度を表す F 値が 0.93 であった．また，その精度評価の結果をもとにプロトタイプシステムを実装し，使用実験ではプロトタイプシステムを用いて電子書籍を読んでもらい，電子書籍を読み終わった後にプロトタイプシステムについてのアンケートに答えてもらった．そのアンケートでは，使用実験中の体感認識精度について回答してもらったが，認識精度の評価実験の結果を基に想定していた程の認識精度とは感じられていなかった．

使用実験中の体感認識精度が期待したほどの精度ではなかった要因としては，3 つの要因が考えられる．

1 つ目の要因は，これまでの我々の研究[6]で提案したジェスチャ認識手法である．これまでの我々の研究[6]で構築したデータセットをもとに算出した両足のジェスチャ認識の精度を表す F 値が 0.93 であったが，この認識精度でもユーザが期待していた程認識精度が高くなかったことが考えられる．

2 つ目の要因は，システムがジェスチャを認識するまでにかかる時間である．これまでの我々の研究[6]では，両足ジェスチャの認識精度を最も重要視しており，ジェスチャの認識までにかかる時間についての考慮を十分行っていなかった．そのため，ジェスチャの認識ま

でにかかる時間についてのアンケート結果において，8 人中 7 人の実験協力者が時間の遅さを感じているため，体感認識精度に影響が出たと考えられる．

3 つ目の要因は，スマートフォンのポケット内での位置変化である．これまでに我々の研究[6]で行った使用実験では，最初にデータセット構築を行い，そのスマートフォンを一度ポケットから取り出してもらい，データセット構築時のデータセットをトレーニングデータとしてプロトタイプシステムに適用させてから，再度ポケットにスマートフォンを入れてもらい，デバイスの操作を行ってもらった．プロトタイプシステムにトレーニングデータを適用させる過程において，スマートフォンを一度取り出したことにより，データセット構築時とはセンサの軸が変わってしまい，認識精度に影響が出たことが考えられる．また，これまでの我々の研究[6]では特定の向きにおける認識精度のみしか検証することができていないため，スマートフォンの向きを変えてポケットに入れた際の認識精度への影響について検証されていない．

そこで，これら 3 つの要因に対して本研究では，データセットの構築について手法を検討するとともに，特徴量と分類器の選定や，特徴ベクトルの生成方法を再検討し，認識精度の改善を図る．次に認識までにかかる時間の短縮においては，望ましいシステムの応答時間以内に認識可能とするための手法の検討を行う．さらに，ポケット内でのスマートフォンの位置変化については，まずポケット内のスマートフォンの位置変化による認識精度への影響を明らかにし，ポケット内のスマートフォンの位置変化による認識精度への影響を低減する手法の検討を行う．

1.4. 本研究の論文構成

本研究の構成は，第 2 章でジェスチャ認識に関する研究と行動推定に関する研究，音声認識に関する研究を挙げ，本研究の位置付けを明確にする．第 3 章ではこれまでの研究で提案した手法と実験，実験結果を説明する．第 4 章では，これまでのデータセット構築での問題点を解決したデータセット構築を行う．第 5 章では，認識手法と特徴量の再選定を行い認識手法の改善を行う．第 6 章では，認識に使用する時間の短縮と，予備動作を利用した認識手法の提案とその手法の有用性を検証する．第 7 章では認識精度をより高めるために，時系列データを活かした認識手法の提案と，その手法の有用性を検証する．第 8 章ではポケット内のスマートフォンの位置変化が認識精度に及ぼす影響について調査を行い，位置変化の影響を低減するための手法の提案と，その手法の有用性を検証する．第 9 章では本研究全体についての考察を行い，認識手法を改めて整理する．第 10 章では本研究のまとめを行う．

第 2 章　　関連研究

2.1. 行動推定に関する研究

センシング機器を取り付けたユーザのセンサデータから行動推定を行う研究は多く行われている．Bao らの研究[7]では，身体の異なる部分に 5 つの小型 2 軸加速度計を装着し，取得したデータを用いた身体動作検出アルゴリズムを提案した．その結果，決定木分類器を使用した際には日常の活動を 84%の精度で認識可能であることを明らかにした．また，村尾ら[8][9]は手首，腰，足首の 3 か所に加速度センサを装着し，各箇所で取得した加速度の値を用いて，SVM と DPMatching を組み合わせることで行動推定を行う手法を提案している．大内ら[10]はセンシングデバイスを胸ポケットに入れた状態で，加速度センサの値を取得し，「歩行」「作業」「安静」の 3 状態を 95%以上の精度で推定可能とする手法を提案している．河内ら[11]は，携帯電話が身体上の特定の 5 か所のどこに格納されているかを，歩行中に判定可能とするシステムと，それを応用したシステムを実装している．米田らの研究[12]では，気圧センサを用いて，屋外状態では軌道上移動の気圧推移比較を用いた現在地推定を，屋内状態では気圧値を用いた階層推定手法を提案している．その結果，現在地推定については 75%の精度で推定可能であり，階層推定は高精度で認識可能であった．

一方，本研究は推定を行うだけでなく，スマートフォンで取得したセンサの値をリアルタイムに用いてジェスチャを認識し，デバイス操作する点でこれらの研究と目的が異なっている．

2.2. システム用のデバイスを用いたジェスチャ認識に関する研究

センシングデバイスを用いてジェスチャ認識を行う研究は多く行われている．FingerPad[13]は，親指で人差し指に書いた文字やジェスチャを認識させ，その認識結果を基にデバイス操作可能なインタラクションシステムを提案した．その結果，座った状態で操作を行った場合，横幅 1.2mm のターゲットを 93%の精度で操作可能であること明らかになった．Humantenna[14]は，実世界の自分の全身をアンテナとして用いてリアルタイムなインタラクションを行い，自分の全身を使ったジェスチャ認識を実現するシステムを提案している．Han らの研究[15]では，モバイルインタラクションにおいてキックジェスチャが効果的に識別可能であることを明らかにした．Cyclops[17]では，魚眼レンズを身体の中央に取り付け，その魚眼レンズで得られた情報によりジェスチャ認識を可能とした．MimiSense[18]は口を開ける，口を閉じる，及び顎を左に動かすなどの下顎運動によりジェスチャを行い，デバイスを操作可能とするシステムを提案している．Ubi-Finger[19]は，手指のジェスチャを用いて情報機器や情報家電機器の操作を実現する指装着型のウェアラブルデバイスを提案し，実際に Ubi-Finger を用いて実世界機器を操作する実験を通して，システムに対する評価実

験を行っている．山本ら[20]は，ジョギング時における入力インタフェースとして，足ステップによるメニュー制御システムの提案と実装を行っており，使用実験から提案システムの有用性を検証した．また山本ら[21]は，自然で目立ちにくいジェスチャを用いて日常生活のさまざまな状況の変化に応じて入力方法を変えることのできる状況依存のジェスチャ入力手法を提案した．その手法を実現するために，3 軸の加速度センサを靴とリストバンドに取り付け，ジェスチャの認識精度を検証した．澤田らの研究[22]では，手の甲に取り付けた 3 次元加速度センサから運動量として, 加速度の変化, 回転力, 回転方向の分布などを求め，標準パターンとのマッチングを行う事により，10 種類程度のジェスチャを 100%近い識別率で認識できることが明らかになった．また，データグローブ，ポジションセンサ及び，小型 3 次元センサを用いて, 5 指の屈曲，空間的な手の配置，胸のダイナミカルな動きを検出し，それらを統合することによってジェスチャを認識する手法を提案した．管家ら[23]は，実ドラムの運搬コストの問題と，ドラムの打楽器全てが仮想的な打楽器の場合に演奏性や表現力が著しく低下する問題を解決するため，実ドラムと仮想ドラムとの併用を実現する Airstic Drum を提案している．具体的には，仮想ドラムと実ドラムの殴打動作を，ドラムスティックに取り付けた加速度センサおよび角速度センサから識別することにより併用を実現している．Lv らの研究では[24]では，ユーザが手首や膝の上にスマートフォンを装着し，「空中」の対話ジェスチャを実行する，スマートフォン用の新しいウェアラブルなバイブレーション対話フレームワークを提案している．Gong らの研究[25]では，PIR センサを利用して，親指による 6 つのジェスチャを識別可能な手法を提案している．そして冷たいものを触って手が冷えた場合や，熱いものを触って手が熱い場合などの状況においても正確に認識している．Lin らの研究[26]では手の甲に筋電位を測定するセンサを取り付け，そのセンサから得られる情報からハンドジェスチャの認識を行っている．Multi Toe[27]では卓上型アプリケーションが机のサイズの制限を受けてしまうため，数十個を超える画面上のオブジェクトを表示できない問題を，高解像度のマルチタッチ入力を背面投影型の床に統合することで解決している. Marin らの研究[28]ではリープモーションで取得した手の角度や位置などの情報と，キネクトで取得した深度と色の情報を組み合わせることで，それぞれのデバイスでジェスチャを認識するよりも高い精度でハンドジェスチャを認識できることを明らかにしている．

しかし，これらの研究ではシステム用のデバイスを使用しているため，手軽にシステムを利用することができない．一方本研究では，システム用の装置を用いることなく，既存のデバイスを用いることにより，ジェスチャ認識を行っている点でこれらの研究とは異なる．

2.3. カメラ映像を用いたジェスチャ認識に関する研究

カメラで撮影した動画像からジェスチャ認識や行動推定をしている研究は様々存在する．小荒らの研究[29]では，照明などの環境条件の影響を受けなやすい状況下でもロバストに動

作し，計算量の少ない画像ベースの身振り認識手法を実現するため，動作パターンの 1 次元符号化に基づく身振り認識手法の提案を行っている．高橋らの研究[30]では，カメラから得られた画像を低解像度化した濃淡画像の各画素に対して，時間軸方向の FFT を行うことにより，得られる振幅スペクトルと位相スペクトルを用いることで，手を用いた周期ジェスチャを認識可能な手法の提案を行っている．Davis らの研究[31]では，これまでの研究では背景環境や照明環境に影響を受けてしまい，認識精度が悪化してしまうことを．受けない行動を識別するための手法として，画像シーケンス内のどこで動きが発生したかを表す 2 値の motion-energy 画像と，強度が動きの新しさを表すスカラー値画像である motion-history 画像を用いて認識する手法を提案している．Molchanov らの研究[32]では，広範囲に変化する照明条件下で実行される異なる被写体からのハンドジェスチャをカラーカメラおよび深度カメラで認識する場合には，未だに正確な認識が困難という問題を解決するため，3D 畳み込みニューラルネットを用いた動的手振り認識手法を提案している．Vogler らの研究[33]では，アメリカ手話文を認識するため，データグローブなどを使わずに，隠れマルコフモデル（HMM）を使用してカメラ映像から位置の特定を行い，ハンドジェスチャの認識を行っている．Cutler らの研究[34]では，フレーム間差分とオプティカルフローからハンドジェスチャを認識する手法の提案を行っている．Sun らの研究[35]では，スマートフォンのフロントカメラを利用して，ユーザの口の動きを spatial transformer network と CNN，RNN を組み合わせたモデルによる認識手法を提案し，その手法の性能評価を行った実験では，20 の無音音声コマンドを正確に認識できることを実証した．呉らの研究[36]では肌色領域，髪色領域の一次二次モーメントに基づいて入力画像にある頭部の 3 次元姿勢を推定し，連続 DPMatching により，頭部ジェスチャのスポッティング認識を実現し，良好な認識結果が得られた． 西村らの研究[37]では動作者にデータグローブ等の接触型センサやマーカを装着させることなく，人間の身振り手振りをカメラにより取得した動画像を基にジェスチャ認識をおこなう手法に着目し，モデルの自動切り出し法と認識判定法を提案し，オンライン教示可能なシステムを作成することにより動作者に適応可能とした．玉城らの研究[38]ではカメラ動画による 3 次元ハンドジェスチャ認識には前腕回旋動作時に認識精度が低下するという問題点があり実用的ではなかったが，手指の輪郭特徴量情報に爪の位置情報を追加する事で，その問題を解決した．

これらの研究におけるジェスチャ認識では，手を用いる必要や，目や口を認識用のデバイスに向ける必要がある．そのため，荷物などを持っているために手がふさがっている状況，サングラスやマスクをしている状態などではカメラ画像によるジェスチャ認識が向かない状況や状態が存在する．また，カメラを搭載してないデバイスを操作する場合には，認識用のデバイスに顔を向けて操作する手間が存在することが考えられる．しかし本研究では，ズボンのポケットにスマートフォンを入れている状態で，手と目，口を用いずに両足によるジェスチャでデバイス操作を行う点で異なる．

2.4. 携帯型情報機器を用いた行動分析に関する研究

携帯型情報機器のセンサを用いて，人の行動を分析している研究は様々存在する．Negulescu らの研究[39]では，一般的な大型モバイルタッチディスプレイにおけるユーザのタッチの正確性が十分でない問題と，そういったデバイスの操作では，片手でターゲットに到達するために，ユーザは手でデバイスを傾けて指を移動させている特性に着目した．そして，スマホの傾きの情報を用いることで，ユーザが触ろうとしていた場所を予測し，ユーザによるタッチ位置の精度を向上させ，エラー率が 41%減少することを明らかにした．Eardley らの研究[40][41][42]では，スマートフォンの持ち方や，スマートフォンを操作する際の体の姿勢によりスマートフォンがどのように動くかを検証し，持ち方毎の最適な画面設計のガイドラインを提案している．

これらの研究は携帯型情報機器における手を使った UI 操作を拡張するための方法として，ジェスチャ入力を検討している．一方で，我々の研究における目的は UI 操作を拡張することではなく，手による操作が向かない状況における操作を実現する手法の提案を目的としている点で異なる．

2.5. 情報機器におけるジェスチャ認識に関する研究

情報機器における操作性を拡張するためにジェスチャを利用した操作を検討している研究は様々行われている．CommandBoard[43]では，スマートフォンで主に使われているソフトキーボードでは，文字色の変更や，太文字に変更などの書式の変更を行いたい場合には手間がかかってしまう点に着目した．その手間を解決するために，ジェスチャによるコマンド入力で書式の変更ができる手法を提案し，入力時間，選択時間，入力ミス率の全てにおいて既存の手法より優れていることを明らかにした．加藤らの研究[44]では，携帯型情報機器におけるペンジェスチャ入力 UI において，ディジタルインク入力とペンジェスチャ入力の区別が難しいという問題に対し，画面上にボタンを配置する手法とペンの停留を用いる手法の 2 つの手法を用いることで，ディジタルインク入力とペンジェスチャ入力の区別を実現する手法の提案を行った．石原らの研究[45]では，手に持った状態の携帯端末から取得した加速度の値に対して DPMatching を用いることにより手軽かつ高精度の個人認証を可能とする 3D 動作認証を提案し，そのシステムを 1 か月以上にわたって動作を行った場合，本人拒否率を 10%以下にできることを確かめた．

これらの研究は情報機器における手を使った UI 操作を拡張するための方法として，ジェスチャ入力を検討している．一方で，我々の研究における目的は UI 操作を拡張することではなく，手による操作が向かない状況における操作を実現する手法の提案を目的としている点で異なる．

2.6. スマートウォッチにおけるジェスチャ認識に関する研究

スマートウォッチにおける操作性を拡張するためにジェスチャを利用した操作を検討している研究が様々行われている．

Gong らの研究[46]ではスマートウォッチによる主な操作が手による操作であり，荷物などを持っている際には操作できない問題を解決するため，6 つの方向に腕時計の手首を回すことでテキスト入力可能な WrisText を提案している．Xu らの研究[47]では，スマートウォッチに搭載される加速度と角速度のデータから，スマートウォッチを腕につけている方の人差し指で表面に書いた図形の書き込みを 95%の精度で正確に認識できる手法の提案をしている．

これらの研究はスマートウォッチにおける手を使った UI 操作を拡張するための方法として，ジェスチャ入力を検討している．一方で，我々の研究における目的は UI 操作を拡張することではなく，手による操作が向かない状況における操作を実現する手法の提案を目的としている点で異なる．

第 3 章　両足ジェスチャ認識手法

本章ではまず，我々が過去の研究で取り組み本研究のベースとなっている両足ジェスチャによる操作手法[6]と，実験結果の問題点について整理する．

3.1. 提案手法

Scott らの研究[5]では，スマートフォンを入れたポケット側の片足のみによるジェスチャ認識であったが，UI 設計において逆方向に同じ操作をすることで，逆の操作ができる UI 設計が多く行われている．例えば，電子書籍では画面の左側をタップすることでページが進み，右側をタップすることでページが戻る UI が実装されている．他にも，Web ページを閲覧する際には，上にスクロールをすることでページをさかのぼり，下にスクロールすることでページを読み進めることができる操作が実装されている．そのため，左足と右足で同じジェスチャをすることにより逆の操作が可能になると，より直感的な操作ができると期待される．また両足によるジェスチャ入力が可能となれば，片足によるジェスチャよりも表現できるジェスチャ数の増加が期待される．

Scott らの研究[5]では，加速度センサのみで特徴ベクトルを生成し，ナイーブベイズの分類器を用いてジェスチャの動きを識別することを試みた結果，前ポケットでの認識精度 60.2%と高くなかった．しかし，スマートフォンには加速度以外にも様々なセンサが搭載されているため，加速度センサ以外にもセンサを用いることでよりジェスチャの動きを詳細に取得することができ，その結果認識精度が向上することが考えられる．例えば，角速度センサはスマートフォンが 1 秒の間にどのくらい回転したかという情報を取得することができる．そのため，ジェスチャをした際のポケット内でのスマートフォンの回転具合について取得できることが期待される．また，マイクセンサにより，ジェスチャをした際のポケットとスマートフォンの接触音が取得できるため，ポケット側の足が動いたのかそうでないのかを取得できると期待される．

そこで，我々は加速度以外にも角速度，マイクセンサのそれぞれから取得した最大値と最小値の特徴量を使って，ポケット内のスマートフォンを用いた両足ジェスチャを認識する手法を提案してきた[6]．

3.2. 実験

ポケット内のスマートフォンを用いた両足ジェスチャ認識手法の有用性を検証するため，性能評価実験と使用実験の 2 つの実験を行った．

3.2.1. 性能評価実験

ユーザが行った両足ジェスチャをどの程度正確にシステムが認識できるかを検証した．その検証を行うため，我々はまずデータセット構築用の Android 用センシングアプリを開発した．本センシングアプリは，随時加速度や角加速度，マイクセンサにより収集した音声データに対してフーリエ級数展開を行って得た最大音圧とその周波数，UNIX 時間などの記録を行うものである．

本システムを用いてデータセットを構築し，関連研究で用いられていた SVM と DP Matching の 2 つの認識手法によるジェスチャ認識精度をそれぞれ検証し，左右どちらの足でジェスチャが行われたかをどの程度識別可能かについて調査を行った．なお，この実験では 1 種類のジェスチャを左足と右足で行い，ジェスチャの識別ができるか調査すると同時に，本研究に適している認識手法を明らかにする．そこで，19〜23 歳の実験協力者 8 人にデータセット構築用のセンシングアプリを起動中のスマートフォンを右足の前ポケットに入れた状態で，タブレット端末を用いて図 1 のように実験協力者にジェスチャ課題を提示し，ジェスチャを行ってもらった．それらのデータセットを用いて両足によるジェスチャの認識精度を検証した．

その結果，特徴量に角速度 3 軸（最大値・最小値）と加速度 3 軸（最小値）を使用し，閾値を超えた時間より前の 9 フレーム分(約 300[ms])と，後の 16 フレーム分(約 540[ms])の計 25 フレームを使用して生成した特徴ベクトルをもとに算出した認識精度が 0.93 と高い認識精度であった．本実験では角速度と加速度以外にもマイクセンサで収集した音の特徴も使用したが，認識精度が最大の組み合わせに音の特徴量は入らなかった．しかし，マイクセンサにより収集した音声データに対してフーリエ級数展開を行って得た最大音圧のみで分類した時の認識精度は 0.53 であり，ランダムで 3 値分類したときの認識精度よりは認識精度が高いため，最大音圧の特徴も両足ジェスチャの認識に寄与している可能性がある．

3.2.2. 使用実験

これまでの我々の研究[6]で行った使用実験の目的は，実験協力者にプロトタイプシステムを使用してもらい，認識精度，ページめくりまでにかかる時間などの観点からシステムの有用性を検証することが目的である．その検証を行うため，性能評価実験における結果をもとにプロトタイプシステムを実装し，そのプロトタイプシステムを用いて電子書籍を 1 冊読んでもらった．また，読書が終了した直後にアンケートに回答してもらった．読んでもら

（a）　待機画面

（b）　タスク提示画面

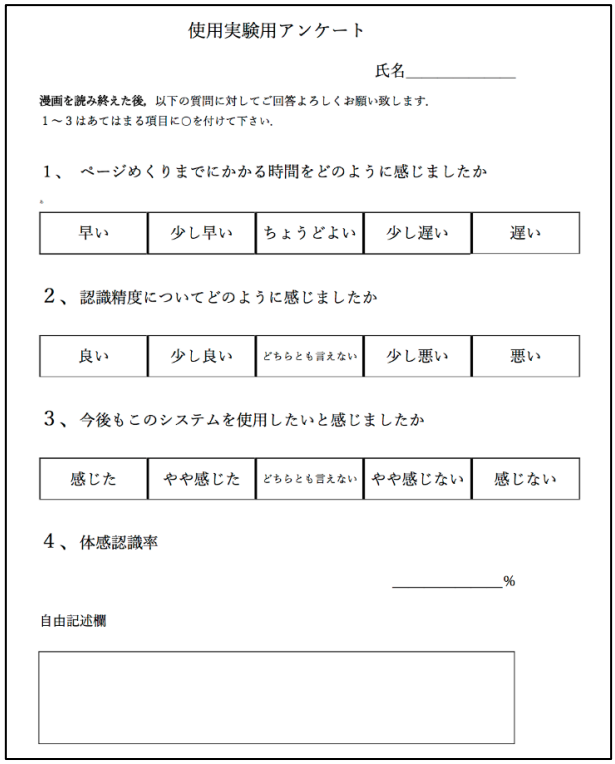
図 1　これまでの我々の研究[6]におけるタスク提示方法

った漫画については，約 1000 冊の漫画の中からジャンルを問わず，読みたい漫画を一巻選択してもらった．これは著者自身が本を選んだ場合，ユーザ間でその本の好き嫌いが分かれることでアンケートに影響が表れる可能性を考慮し，その影響を取り除くためである．アンケートではシステムに対する意見と感想を調査するため，アンケート項目には以下の 4 つを選定した．図 2 が使用したアンケート用紙の画像である．

- 質問 1 ： ページめくりまでにかかる時間をどのように感じましたか
- 質問 2 ： 認識精度についてどのように感じましたか
- 質問 3 ： 今度もこのシステムを使いたいと感じましたか
- 質問 4 ： 体感認識率

質問 1〜3 に対しては，それぞれ 5 段階のリッカート尺度（-2〜+2）により評価してもらった．そして，それぞれのアンケートで「早い」「良い」「使いたいと感じた」と答えた場合に最大値，逆の答えの場合に最小値となるようにした．「体感認識率」の質問に対しては，使用実験を通してジェスチャの認識精度が体感で何%位と感じたかを記入してもらった．また自由記述欄を用意し，システムに対する意見や感想を記述してもらった．

表 1 が実験協力者 8 名(P 1〜P 8)によるアンケート結果である．質問 1 においては，ページめくりまでの時間が少し遅いと感じている傾向にあり，質問 2 では認識精度についてはどちらとも言えないという傾向にある．質問 3 の今後も使用したいかどうかについてもどちらとも言えないという傾向にある．質問 4 の体感認識率の結果としては，73.1%であった．実験協力者 P1〜P7 の実験協力者は 70%〜90%の範囲で回答しているが，P8 のみ 25%と極端に低い回答をしていた．実験協力者の自由記述には，「フィードバックが欲しかった」，「連続入力できないようになっていてやりにくかった」，「人によって読むペースが違うと思うので、それを踏まえたフィードバック時間の設定ができると良い」などといった回答がみられた．


使用実験用アンケート

氏名＿＿＿＿＿＿＿＿

漫画を読み終えた後，以下の質問に対してご回答よろしくお願い致します．
1～3はあてはまる項目に○を付けて下さい．

1、 ページめくりまでにかかる時間をどのように感じましたか

早い	少し早い	ちょうどよい	少し遅い	遅い

2、認識精度についてどのように感じましたか

良い	少し良い	どちらとも言えない	少し悪い	悪い

3、今後もこのシステムを使用したいと感じましたか

感じた	やや感じた	どちらとも言えない	やや感じない	感じない

4、体感認識率

＿＿＿＿＿＿%

自由記述欄

図 2　アンケート用紙

3.3. これまでの研究のまとめと問題点

これまでの我々の研究[6]では両足のジェスチャ認識の精度を検証した結果，F 値が 0.93 であった．またその結果を基に，使用実験を行った結果，体感認識率が 73.1%と期待したほど高くなかった．体感認識精度が期待したほどの精度ではなかった要因としては，3 つの要

表 1　アンケート結果

実験協力者	質問 1	質問 2	質問 3	質問 4
P 1	- 1	1	1	85%
P 2	- 1	-1	1	80%
P 3	- 2	-1	-1	75%
P 4	0	-1	1	70%
P 5	- 1	1	1	90%
P 6	- 1	1	0	90%
P 7	- 1	1	1	70%
P 8	- 1	-2	-2	25%
全体	- 1	-0.125	0.250	73.1%

因が考えられる．

1 つ目の要因は，これまでの我々の研究[6]で提案したジェスチャ認識の手法である．データセット構築で算出した両足のジェスチャ認識の精度を表す F 値が 0.93 であったが，この認識精度でもユーザが期待していた程認識精度が高くなかったことが考えられる．

2 つ目の要因は，システムがジェスチャを認識するまでにかかる時間である．これまでの我々の研究[6]では，両足ジェスチャの認識精度を最も重要視しており，ジェスチャの認識までにかかる時間についての考慮を十分行っていなかった．そのため，ジェスチャの認識までにかかる時間についてのアンケート結果において，8 人中 7 人の実験協力者が時間の遅さを感じているため，体感認識精度に影響が出たと考えられる．

3 つ目の要因は，ポケット内のスマートフォンの位置変化である．使用実験では最初にデータセット構築を行い，そのスマートフォンを一度ポケットから取り出してもらい，データセット構築時のデータセットをトレーニングデータとしてプロトタイプシステムに適用させてから，再度ポケットにスマートフォンを入れてもらい，デバイスの操作を行ってもらった．プロトタイプシステムにトレーニングデータを適用させる過程において，スマートフォンを一度取り出したことにより，データセット構築時とはセンサの軸が変わってしまい，認識精度に影響が出たことが考えられる．

1 章でも述べた通り，本研究ではこの 3 つの要因を解決するための研究を行った．

第 4 章　　データセットの再構築

第 3 章で示した使用実験のアンケート結果から，使用時における認識精度や認識までにかかる時間に問題点があることが明らかとなった．これは，認識精度を最重要の項目に位置付けていたため，閾値を超えた時点の前後のフレームを用いてジェスチャ認識を行っていたことが原因であり，ジェスチャ認識に使用する時間は，閾値を超えてから約 500[ms]前後経過したときの時間までのセンシングデータを使用していたためである．その結果，少なくともジェスチャを開始した時間から，実際に操作が反映されるまで約 500[ms]前後は経過していた．また，8 人中 7 人の実験協力者が認識までにかかる時間が遅いと感じていることが明らかとなったため，500[ms]前後かかる認識時間はユーザの要求するレベルに達していなかったと考えられる．ここで，ある動作をしてから反映されるまでの最適な時間を検証している研究の一つに Miller らの研究[49]があげられる．Miller らの研究[49]では，キーボードを押してから 100[ms]から 200[ms]以内がシステム応答時間の限界であることを実験により明らかにしている．この研究から，ジェスチャを行ってから認識するまでにかかる時間は最低でも 200[ms]以内に認識させる必要があると考えられる．

そのため，ジェスチャ認識に使用する時間の短縮を検討することにより，ポケット内のスマートフォンを利用した両足ジェスチャ認識手法の改善を行う．具体的には，あるユーザがジェスチャを開始した時間から，どのくらいの時間のセンシングデータを使用することで，ジェスチャを正しく認識することができるかを検証する．そのために，ジェスチャ開始タイミングから，少しずつフレーム数を増やした際のそれぞれの使用フレームにおけるジェスチャ認識の精度を検証する．しかし，これまでの我々の研究[6]で行ったデータセット構築では，図 1 のように待機画面では，待機画面からタスク実行画面に切り変わるまでの時間を 1 秒単位の整数で表示していた．そのため，実験協力者はタスク画面に切り替わるタイミングを正確に推測することが難しいため，実験協力者がジェスチャを開始するタイミングが，システムにより指示したジェスチャ開始のタイミングからずれていることが考えられる．つまり，これまでのデータセットを用いずに，システムにより指示したジェスチャ開始タイミングと，実験協力者のジェスチャのタイミングが正確に一致したデータセットを構築する必要がある．

そこで，本章ではまず正確にタイミングが一致するデータセット構築手法について検討し，データセットの再構築を行う．

4.1. データセット構築用システム

ここで，人によるジェスチャを開始するタイミングのぶれをなくすための方法として，我々が着目したのは，音楽ゲームをプレイしている際の UI である．音楽ゲームの UI ではリズムに合わせて流れてくるノーツがタップ位置に完全に重なったタイミングでタップす

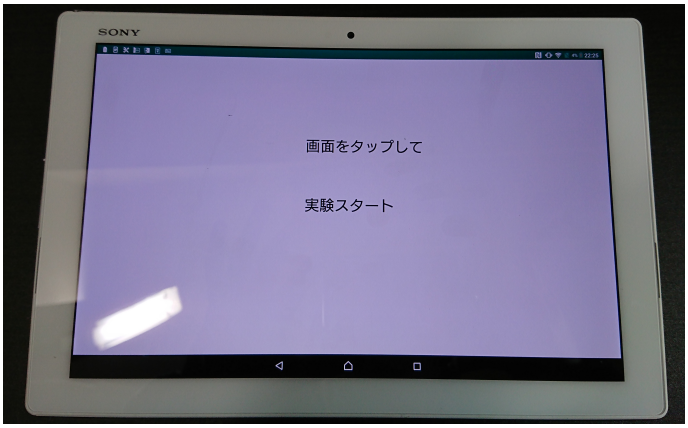
る UI が実装されている．そこで我々は音楽ゲームにおける UI を参考にし，図 3 のようにタブレットでジェスチャの開始タイミングを，実験協力者が正確に把握できるように統制を図った．ノーツとは音楽ゲームなどで用いられるリズムに合わせて流れる音符である．

本システムでは Xperia X Performance のスマートフォンを用いて，足のジェスチャをセンシングし，Xperia Z4 Tablet のタブレットを用いてタスクの提示を行った．ここで，これまでに行ったデータセット構築では，データセット構築用のセンシングアプリを，起動中のスマートフォンを右足の前ポケットに入れた状態で，図 3 のようにタブレット端末を用いて実験協力者にジェスチャ課題を提示したうえで，両足によるジェスチャを行ってもらい，データセット構築を行った．そのデータセット構築中に，スマートフォンで収集していたデータのセンシング周波数は約 30[Hz]であったが，サンプリング周波数を約 50[Hz]に変更することにより，ジェスチャ時の細かい動きを収集することができると考えられる．そこで，サンプリング周波数は約 50[Hz]とした．センシングした情報は，これまでの我々の研究[6]をもとに 3 軸の加速度$[m/s^2]$，3 軸の角速度$[rad/s]$，マイクセンサにより収集した音声データに対してフーリエ級数展開を行って得た最大音圧[db]と，最大音圧の周波数[Hz]を使用した．

4.2. データセット構築時の手順

実験協力者には，最初にかかとの上げ下げのジェスチャの仕方を著者自身が実演し，教示した．その次に，ジェスチャを開始するタイミングについては図 3 の(b)の灰色の部分と，水色のノーツの部分が完全に重なったタイミングでジェスチャを開始するように著者自身が実演し，実験協力者にジェスチャの開始タイミングを教示した．実験協力者が実験についての理解ができたことを確認した後，スマートフォンとタブレット用のそれぞれのデータセット構築システムを用いて，かかとの上げ下げのジェスチャを 200 回してもらった．ここで，スマートフォンを入れるポケットと向きは，右足の前ポケットで，マイク部分が下になり，ディスプレイが足に触れない向きになるように実験の統制を図った．また，実験協力者が約 5000[ms]に 1 回の間隔でジェスチャを行うように，ノーツが約 5000[ms]に一回の間隔で灰色の部分と完全に重なるようにシステムで制御した．ジェスチャを行う時間間隔の設定は，前のジェスチャが完全に終了しており，次のジェスチャに影響を与えない間隔となるように設定している．約 5000[ms]という時間は，筆者自身が実際にジェスチャを行った際にかかった時間をもとに設定した．ジェスチャの合間の姿勢については，特に指示をしなかった．

なお，このデータセット構築では 18〜25 歳の大学生 15 人の実験協力者に 200 回の踵の上げ下げのジェスチャを行ってもらった．15 人(A〜O)の実験協力者の中で利き足が右足の実験協力者が 14 人(A〜K，M〜O)であり，左足の実験協力者が 1 人(L)だった．また，200 回のジェスチャの内訳は，ジェスチャの種類による偏りを防ぐため，右足によるジェスチャを 100 回，左足によるジェスチャを 100 回とした．



(a)　実験開始前

(b) 実験中

図 3 データセット構築システム

第 5 章　　特徴量の再選定

これまでの我々の研究[6]では，SVM を用いた認識精度が最も高い認識精度手法であったが，使用実験のアンケート結果から，その認識精度もユーザが期待しているほどの認識精度ではないことが明らかとなった．そこで，本研究では，近年機械学習で使用されている分類器である Random Forest を用いた際の認識精度と，SVM を用いた際の認識精度の比較を行い，Random Forest による認識の有用性についての検討を行う．また，両足のジェスチャ認識において，どの特徴がどのくらい両足ジェスチャ認識に寄与しているのかを明らかにする．

5.1. 特徴量の検討

これまでの研究では，最大値・最小値のみを用いて認識精度の算出を行っていため，波形の最大値と最小値の情報のみしか用いることができていなかった．ここで，波形の特徴をさらに活かすことにより，両足ジェスチャの認識精度改善が期待される．例えば，波形の平均値はその時間帯にどのくらいの時間をかけて動いたのかという情報を取得することができる．また波形の中央値もその時間帯にどのくらいの時間をかけて動いたのかという情報を取得することができると考えられる．

そこで最大値と最小値以外にも平均値と中央値を使ってポケット内のスマートフォンを用いた両足ジェスチャを認識する．

5.2. ジェスチャデータの抽出

本章では第 4 章で構築したデータセットを使用して認識精度の検証を行う．そこで，まず特徴ベクトルを生成するための区間を決定するため，センシングデータをジェスチャ時と待機時のものに分類後，それぞれ一定数のフレームを抽出する．これまでに我々が行った研究では，センシング周波数は 30[Hz]であり，特徴ベクトルは閾値を超えた時間より前の 9 フレーム分(約 300[ms])と，後の 16 フレーム分(約 540[ms])の計 25 フレームを使用して特徴ベクトルを生成し，ジェスチャを認識していた．ここで，第 4 章で行ったデータセット構築では，サンプリング周波数を 30[Hz]から 50[Hz]に変更している．そこで，本研究においても閾値を超えた時間よりも前の 15 フレーム（約 300[ms]）のセンシングデータと，後の 27 フレーム（約 540[ms]）の計 42 フレームのセンシングデータを使用する．

右足によるジェスチャと左足によるジェスチャデータの抽出では，まずタスク開始時間に一番近いセンシングデータから 100 フレーム分（約 2000[ms]分）のデータを抽出する．そして，その 100 フレーム分のデータを基に，我々が設定した閾値を超えた最初のフレームを特定し，また，待機状態のセンシングデータの抽出では，ジェスチャを開始するタイミング

より前の時間帯のセンシングデータを使用する，そのため，ジェスチャ開始タイミングから 102 フレーム前（約 2000[ms]前）から 61 フレーム前(約 1250[ms]前)までの計 42 フレーム分のセンシングデータを使用した．

5.3. 性能評価

ジェスチャ認識の性能評価では，従来手法である最大値・最小値により生成した特徴ベクトルとそれぞれの分類器の組み合わせにおける認識精度と，最大値・最小値・平均値・中央値により生成した特徴ベクトルとそれぞれの分類器における認識精度を比較することにより，両足によるジェスチャ認識手法を改善するための認識手法を検討する．また，両足によるジェスチャ認識のどのセンサがどのくらい認識に寄与しているのかについても明らかにする．

今回使用する特徴量はこれまでの表 2 の特徴を選定した．分類器については，Random Forest と，SVM を用いる．今回使用した SVM と Random Forest の認識に使用したライブラリは，scikit-learn[48]を使用した．SVM と Random Forest におけるパラメータはそれぞれデフォルトのパラメータを使用した．本研究では，生成した特徴ベクトルにおいて，センサの値の大きさによる偏りにより認識精度に与える影響を考慮し，その特徴ベクトルの値を標準化した．また，第 4 章で行ったデータセット構築では右足によるジェスチャ 100 回，左足によるジェスチャ 100 回の合計 200 回ジェスチャを行ってもらっており，本研究ではジェスチャが行われる前の時間帯を待機状態として収集しているため，待機状態のデータが 200 回分存在する．しかし，200 回分の待機状態を全て使った場合には，ジェスチャのデータ数と待機状態のデータ数に偏りが生じてしまい，待機状態での認識精度がジェスチャのデータよりも認識精度へ与える影響が大きくなってしまうため，待機状態のデータが 100 回分になるようにアンダーサンプリングを行った．ここで，認識精度の算出の際には，300 回分のデータ（左足によるジェスチャデータ 100 回分，右足によるジェスチャデータ 100 回分，

表 2 ジェスチャ認識に使用した特徴一覧

特徴
X 軸・Y 軸・Z 軸の加速度値
X 軸・Y 軸・Z 軸の角速度値
マイクで得た最大音圧の周波数
マイクで得た最大音圧

表 3 それぞれの認識手法における認識精度（Previous:従来手法, New:提案手法）

	SVM		RandomForest	
	Previous	New	Previous	New
Precision	0.9820	0.9838	0.9848	0.9863
Recall	0.9811	0.9833	0.9844	0.9856
F-measure	0.9816	0.9836	0.9846	0.9859

待機状態のデータ 100 回分）から，それぞれ 8 割のデータ(240 回分のデータ)を用いて学習を行い，残りの 2 割のデータ（60 回分のデータ）を用いて認識精度の算出を行った．60 回分のデータの抽出では，データの偏りを考慮し，左足によるジェスチャデータが 20 回分，右足によるジェスチャデータが 20 回分，待機状態のデータが 20 回分になるようにランダムでそれぞれ 100 回分のデータから抽出した．

5.4. 結果・考察

表 3 が各クラスにおける Precision（適合率）の平均値，Recall（再現率）の平均値，それらの平均値をもとに算出した F-measure（F 値）の表である．表 3 において Previous と書かれた列が従来手法である最大値・最小値で生成した特徴ベクトルを用いた時の認識精度を求めたものであり，New と書かれた列が提案手法である最大値・最小値・平均値・中央値で生成した特徴ベクトルを用いた時の認識精度である．表 4〜表 7 はそれぞれの手法と特徴ベクトルの組み合わせを用いて算出した 15 人分の混同行列を合算した混同行列である．縦軸と横軸は，それぞれ実際に行われたジェスチャの種類と，認識されたジェスチャの種類を表している．表 8 は個人ごとにそれぞれの手法と特徴量を用いて算出した混同行列である．表 9 は，表 8 で最も認識精度が低かった実験協力者 O の Random Forest と提案手法で生成した特徴ベクトルで算出した混同行列である．表 10 が提案手法により生成した特徴ベクトルと Random Forest を用いて算出した際の認識精度への特徴量の寄与度を表した表である．

表 3 の結果から，Random Forest と SVM のそれぞれの手法で，従来手法で生成した特徴ベクトルを用いた際の認識精度を，提案手法により生成した特徴ベクトルを用いた際の認識精度が上回っている．表 10 からも，マイク以外のセンサにおいては平均値・中央値が認識精度の改善に寄与できているため改善していると考えられる．その結果，特徴ベクトルについては，期待した通りに従来手法による特徴ベクトルではなく，提案手法により特徴ベクトルを生成することで認識精度が改善したと考えられる．マイクセンサにおいて，音圧の最大値以外の寄与度が低かった要因としては，マイクの特徴量は音の強さであるため，最小値が他のセンサとは違い，ジェスチャ時とジェスチャをしていない時で同じ情報となってしまったため，寄与度が下がったと考えられる．また，最小値の寄与度が低いために，平均値と中央値の特徴量に影響が出たため，平均値と中央値の寄与度も下がったと考えられる．

本章では，ジェスチャの認識精度を改善することを目的として，近年機械学習で用いられている Random Forest による認識手法を検討した．その結果としては，表 3 から分かるように，最大値と最小値により生成した特徴ベクトルを用いた際と，提案手法により生成した特徴ベクトルを用いた際の認識精度は，Random Forest を用いた際の認識精度の方が高かったが，認識精度の差がわずかであるため，認識手法による差は見受けられなかった．

表 4　SVM と従来の特徴ベクトルを用いて算出した 15 人分の混同行列

	認識結果		
	左	右	待機
左	294	2	4
右	2	295	3
待機	4	2	294

表 5　SVM と提案手法の特徴ベクトルを用いて算出した 15 人分の混同行列

	認識結果		
	左	右	待機
左	295	1	4
右	1	298	1
待機	5	3	292

表 6　RF と従来の特徴ベクトルを用いて算出した 15 人分の混同行列

	認識結果		
	左	右	待機
左	296	2	2
右	3	296	1
待機	6	0	294

表 7　RF と提案手法の特徴ベクトルを用いて算出した 15 人分の混同行列

	認識結果		
	左	右	待機
左	298	0	2
右	2	297	1
待機	8	0	292

表 8　それぞれの認識手法における認識精度（F 値）
（Previous:従来手法, New:提案手法）

実験協力者	SVM		RandomForest	
	Previous	New	Previous	New
A	0.9837	0.9837	1.0000	0.9837
B	0.9837	0.9837	0.9837	0.9837
C	0.9837	0.9837	0.9837	1.0000
D	0.9500	0.9837	1.0000	1.0000
E	1.0000	1.0000	0.9837	1.0000
F	0.9837	0.9675	0.9837	0.9682
G	1.0000	1.0000	1.0000	1.0000
H	1.0000	1.0000	1.0000	0.9837
I	1.0000	0.9667	0.9837	1.0000
J	0.9837	0.9837	0.9837	0.9837
K	1.0000	1.0000	1.0000	1.0000
L	1.0000	1.0000	0.9837	0.9837
M	1.0000	1.0000	1.0000	1.0000
N	1.0000	1.0000	1.0000	1.0000
O	0.8548	0.9010	0.8833	0.9024

表 9　RF と提案手法の特徴ベクトルを用いて算出した
実験協力者 O の混同行列

	認識結果		
	左	右	待機
左	18	0	2
右	1	18	1
待機	2	0	18

表 10　Random Forest の認識精度への特徴量の寄与度(%)

加速度	X 軸	最小値	3.427%
		最大値	5.349%
		平均値	1.851%
		中央値	0.949%
	Y 軸	最小値	9.650%
		最大値	8.395%
		平均値	1.639%
		中央値	0.388%
	Z 軸	最小値	5.972%
		最大値	4.747%
		平均値	0.588%
		中央値	1.434%
角速度	X 軸	最小値	9.674%
		最大値	1.272%
		平均値	6.360%
		中央値	2.356%
	Y 軸	最小値	1.982%
		最大値	2.309%
		平均値	1.267%
		中央値	0.534%
	Z 軸	最小値	4.584%
		最大値	10.136%
		平均値	8.655%
		中央値	4.656%
マイク	周波数	最小値	0.016%
		最大値	0.008%
		平均値	0.054%
		中央値	0.019%
	最大音圧	最小値	0.022%
		最大値	1.472%
		平均値	0.134%
		中央値	0.100%

第 6 章　予備動作を用いた認識手法

第 4 章でも述べた通り，認識にかかる時間に問題があった．そこで，本章では最低でも200[ms]以内に両足ジェスチャの認識を可能とする手法の検討を行う．

6.1. 認識手法

そこで，ジェスチャ認識に使用する時間を短縮する手法を検討する． 200[ms]以内にジェスチャを認識するためには，ユーザがジェスチャを開始した時間から 200[ms]以降のデータを使うことができないため，少ない情報からジェスチャの認識を行うことになる．その場合には，約 500[ms]を用いて認識させていたこれまでの手法と比べると，ジェスチャの認識精度が低下することが考えられる．ここで，200[ms]以降のジェスチャデータは使用できないが，ジェスチャを始める直前に無意識に行われる構えの動作である予備動作のセンシングデータを利用することで，予備動作を用いない時より認識精度が高くなることが考えられる．その結果として，これまでの手法と同じ認識精度を出したい場合には，予備動作のジェスチャデータと，これまでの手法で用いたジェスチャデータよりも短い時間のジェスチャデータを利用することで，同じ位の認識精度でジェスチャを認識することができるようになるため，予備動作を利用した手法を用いることで，認識時間を短縮することができると考えられる．

6.2. 性能評価

予備動作を用いた認識手法の性能評価では，予備動作のセンシングデータを利用することで，利用しないで生成した特徴ベクトルを用いた認識精度よりも認識精度が高くなるかを検証する．今回使用するセンサは第 5 章と同じであり，使用する特徴量は第 5 章の結果から最大値と最小値，平均値，中央値を使用した．使用する分類器については，第 5 章で用いた Random Forest と，SVM を使用し，特徴ベクトルの値は標準化している．

6.3. 実験結果・考察

表 11 が Random Forest と SVM において，ジェスチャ開始タイミングより前のフレームと，ジェスチャ開始タイミングのフレームと，それ以降のフレームを含んだフレームを用いて特徴ベクトルを生成し，認識精度を算出した F-measure（F 値）である．ジェスチャ開始タイミングは水色のノーツが完全に灰色の部分と一致したタイミングである．今回使用した特徴量は，最大値・最小値・平均値・中央値であるため，使用フレームの総数が 2 フレーム未満の場合は，最大値と最小値が一致してしまうため，認識精度の算出は行わなかった．

表 11 の赤い背景で表示されている箇所は，ジェスチャ開始タイミングのフレームと，それ以降のフレームを含む使用フレーム数が同一の特徴ベクトルの中で，最も認識精度が高いことを表している．表 12 は予備動作を用いない組み合わせの中で最も認識精度が高かった組み合わせで算出した混同行列である．表 13 は予備動作を用いた組み合わせの中で，最も認識精度が高かった組み合わせで算出した混同行列である．表 14 は最も認識精度が高かった組み合わせにおける実験協力者ごとの認識精度を表した表である．表 15〜表 18 は最も認識精度が高かった組み合わせを用いて算出した認識精度の中で，ジェスチャの認識精度が特に低かった 4 人の表である．

本研究では望ましいシステムの応答時間(200[ms])以内でジェスチャを認識することを目材しているため，ユーザがジェスチャをし始めてから最低でも 10 フレーム(約 200[ms])以内にジェスチャを認識する必要がある．そのため，その条件を満たした中で最も高い精度だったのは，ジェスチャをする前の 10 フレームと，ジェスチャをし始めたフレームとそれ以降のフレームを含む 10 フレームの組み合わせが最も高い認識精度であった．またジェスチャをし始めたフレームと，それ以降のフレームを含む 10 フレーム以外の組み合わせにおいても，全ての組み合わせにおいて，ジェスチャを開始する前のフレームを含むことにより認識精度が向上した．表 12 と表 13 の混同行列から，予備動作を利用することで右足によるジェスチャと待機状態をより正確に認識することが出来るようになったことが認識精度の向上に寄与していたことが明らかとなった．

表 11 Random Forest と SVM における F 値
（背景が赤いセルは同一の後フレームを使用している中で最も認識精度が高い組み合わせを表している，
縦軸：前フレーム使用数，横軸：後フレーム使用数）

		0	1	2	3	4	5	6	7	8	9	10	15	20
0	RF			0.758	0.775	0.800	0.827	0.837	0.850	0.869	0.870	0.881	0.925	0.970
	SVM			0.739	0.750	0.775	0.801	0.808	0.831	0.850	0.862	0.877	0.931	0.956
1	RF		0.737	0.753	0.780	0.820	0.839	0.837	0.858	0.869	0.883	0.888	0.939	0.972
	SVM		0.723	0.745	0.783	0.787	0.807	0.812	0.831	0.844	0.864	0.879	0.938	0.956
2	RF	0.707	0.730	0.771	0.791	0.802	0.811	0.838	0.869	0.859	0.876	0.885	0.929	0.961
	SVM	0.702	0.716	0.750	0.766	0.797	0.813	0.808	0.835	0.849	0.869	0.889	0.936	0.957
3	RF	0.715	0.762	0.776	0.789	0.820	0.834	0.847	0.842	0.876	0.875	0.871	0.941	0.965
	SVM	0.711	0.734	0.755	0.784	0.793	0.814	0.814	0.838	0.847	0.874	0.887	0.940	0.956
4	RF	0.744	0.744	0.792	0.780	0.811	0.833	0.838	0.851	0.873	0.884	0.893	0.933	0.973
	SVM	0.727	0.734	0.760	0.779	0.800	0.816	0.822	0.837	0.854	0.871	0.879	0.933	0.960
5	RF	0.728	0.745	0.782	0.786	0.810	0.834	0.830	0.854	0.866	0.869	0.891	0.943	0.970
	SVM	0.734	0.739	0.758	0.778	0.792	0.821	0.827	0.840	0.855	0.865	0.878	0.934	0.960
6	RF	0.735	0.747	0.769	0.802	0.807	0.825	0.832	0.845	0.876	0.887	0.891	0.935	0.969
	SVM	0.736	0.756	0.761	0.785	0.800	0.816	0.827	0.842	0.852	0.860	0.882	0.937	0.962
7	RF	0.709	0.762	0.783	0.789	0.822	0.825	0.833	0.839	0.866	0.892	0.888	0.939	0.967
	SVM	0.745	0.751	0.764	0.788	0.799	0.812	0.829	0.837	0.851	0.863	0.878	0.936	0.960
8	RF	0.718	0.759	0.775	0.805	0.809	0.824	0.841	0.856	0.864	0.886	0.879	0.936	0.970
	SVM	0.754	0.757	0.772	0.787	0.797	0.816	0.826	0.838	0.853	0.862	0.880	0.935	0.960
9	RF	0.740	0.753	0.765	0.777	0.814	0.830	0.831	0.846	0.853	0.878	0.894	0.939	0.967
	SVM	0.734	0.751	0.771	0.788	0.807	0.817	0.824	0.842	0.854	0.868	0.877	0.935	0.959
10	RF	0.725	0.757	0.773	0.797	0.803	0.821	0.843	0.842	0.860	0.878	0.897	0.937	0.962
	SVM	0.754	0.747	0.772	0.794	0.805	0.823	0.829	0.844	0.859	0.870	0.882	0.935	0.959

表 12 予備動作を用いない組み合わせの中で
最も認識精度が高かった組み合わせで算出した 15 人分の混同行列

	判定結果		
	左	右	待機
左	266	14	20
右	12	268	20
待機	26	16	258

表 13 予備動作を用いた組み合わせの中で最も認識精度が高かった
組み合わせで算出した 15 人分の混同行列

	判定結果		
	左	右	待機
左	266	7	27
右	11	272	17
待機	23	10	267

表 14 最も認識精度が高かった組み合わせにおける
実験協力者ごとの認識精度

実験協力者	Random Forest		
	Precision	Recall	F-measure
A	0.9349	0.9333	0.9340
B	0.8101	0.8000	0.8050
C	0.9516	0.9500	0.9508
D	0.7472	0.7333	0.7402
E	0.9365	0.9333	0.9349
F	0.8376	0.8333	0.8354
G	0.9841	1.0000	0.9920
H	1.0000	1.0000	1.0000
I	0.9379	0.9333	0.9356
J	0.8865	0.8833	0.8849
K	0.9333	0.9333	0.9333
L	0.9349	0.9333	0.9341
M	0.8088	0.8000	0.8044
N	0.9127	0.9000	0.9063
O	0.8541	0.8500	0.8520

表 15 最も認識精度が高かった組み合わせの特徴ベクトルを用いて算出した実験協力者 B の混同行列

	判定結果		
	左	右	待機
左	17	1	2
右	3	16	1
待機	4	1	15

表 16 最も認識精度が高かった組み合わせの特徴ベクトルを用いて算出した実験協力者 D の混同行列

	判定結果		
	左	右	待機
左	15	2	3
右	3	14	3
待機	5	0	15

表 17 最も認識精度が高かった組み合わせの特徴ベクトルを用いて算出した実験協力者 M の混同行列

	判定結果		
	左	右	待機
左	16	0	4
右	1	16	3
待機	1	3	16

表 18 最も認識精度が高かった組み合わせの特徴ベクトルを用いて算出した実験協力者 O の混同行列

	判定結果		
	左	右	待機
左	17	0	3
右	1	17	2
待機	1	2	17

第 7 章　一連の動作を生かした認識手法

第5章では波形の特徴をさらに生かすために, 特徴量の選定を新たに行い, 認識手法の精度改善を行った．そして，第 6 章では認識にかかる時間を短縮するために，ジェスチャを始める直前に無意識に行われる構えの動作である予備動作を用いた認識手法の検討を行った．その結果, 予備動作のジェスチャデータを用いることにより認識精度が向上していた．具体的にはジェスチャ開始タイミングより前の 10 フレームと，ジェスチャ開始タイミングのフレームと，それ以降のフレームを含む 10 フレームの合計 20 フレームでの認識精度が最も高い認識精度となった．しかし，その認識精度はまだまだ高くないため，認識精度の改善を行う必要がある．

第 6 章までの手法では 20 フレーム（約 400[ms]）のジェスチャデータ全体から最大値と最小値，平均値，中央値の特徴量を算出し，特徴ベクトルの生成を行ってきた．そのため，ジェスチャを行う際の一連の動作が活かしきれていない．例えば，かかとの上げ下げのジェスチャの場合には，予備動作と足を上げる動作，足を下げる動作で構成されているが，これまでの手法では，それら全ての動作が含まれているジェスチャデータから，特徴量の算出を行っていた．そこで，波形を細かく分割することで，ジェスチャの一連の動作を取得することが可能となり，その分割したジェスチャデータから生成した特徴ベクトルを用いることにより，両足ジェスチャの認識精度がさらに高くなることが考えられる．

7.1. 認識手法

そこで我々の研究では，ジェスチャ中の一連の動作を活かした認識手法として，認識に使用するセンシングデータを等分割し，分割したそれぞれのジェスチャデータにおいて，最大値，最小値，中央値，平均値の値を算出し，それらすべての特徴量を組み合わせて生成した特徴ベクトルを用いた認識手法を提案する．

7.2. 性能評価

提案手法の性能評価では，予備動作と足を上げる動作，足を下げる動作を区別せずに特徴ベクトルを生成し，その特徴ベクトルを用いて認識する第 6 章までの認識手法と，ジェスチャ中の一連の動作を活かした提案手法における両足ジェスチャの認識精度を比較することで提案手法の有用性を検証する．ここで認識精度の算出においては，第 5 章での実験結果から最大値・最小値・平均値・中央値を元に生成した特徴ベクトルと，それぞれの分類器の組み合わせによって認識精度を算出する．今回使用するセンサはこれまでと同じく加速度，角速度，マイクセンサにより収集した音声データに対してフーリエ級数展開を行って得た最大音圧と，最大音圧の周波数を使用した．分類器については，第 6 章で用いた Random Forest

と，SVM を用いる． なお SVM と Random Forest におけるパラメータはそれぞれデフォルトのパラメータを使用しており，特徴ベクトルの値は標準化した．

第 6 章の結果から予備動作を含んだフレームでの認識精度が最も高い結果であり，ジェスチャ開始タイミングのフレームと，それ以降のフレームの使用数の増加に比例して認識精度が向上する結果となった．そこで，ジェスチャデータの抽出の際には，ジェスチャ開始タイミングより前のフレームを使用し，ジェスチャ開始タイミングのフレームとそれ以降のフレームを含んだ 10 フレーム前までのフレームを使用した際のジェスチャデータを使用して特徴ベクトルの生成を行い，認識精度を算出する．ここで，第 6 章で最も認識精度が高かった手法と比較をするため，学習データとテストデータは第 6 章と同一のデータを選定した．

7.3. 実験結果・考察

表 19 が時系列データを活かした認識手法を用いた際の認識精度の表である．例えば前フレームが 2 フレーム，後フレームが 10 フレーム，分割数が 6 の場合における認識精度は，前フレームと後フレームの合計 12 フレームを 6 分割し，それぞれ 2 フレームから最大値，最小値，平均値，中央値を算出し，それぞれ 6 個分の特徴量から生成した特徴ベクトルを用いて SVM と Random Forest それぞれの分類器で Precision（適合率），Recall（再現率），F-measure（F 値）を算出した結果である．表 20 が時系列の細分化による認識手法を用いた際に最も認識精度が高い組み合わせの特徴ベクトルで算出した混同行列である．表 21〜表 23 が第 6 章で最も認識精度が高い組み合わせだった前フレーム 10 で，後フレーム 10 の合計 20 フレームを 2 等分，5 等分，10 等分して生成した特徴ベクトルで算出した混同行列である．表 24 が最も認識精度が高かった組み合わせの特徴ベクトルで算出した実験協力者ごとの認識精度の表である．表 25〜表 28 が表 24 で認識精度が特に低かった 4 名の混同行列である．

表 19 の結果から，前フレームが 2 フレーム，後フレームが 10 フレーム，分割数が 2 で生成した特徴ベクトルと Random Forest を組み合わせて算出した認識精度が最も高い認識精度が高い組み合わせであった．また本章では，全体のフレームから算出した特徴ベクトルでの認識精度をより向上させることを期待して，時系列データを活かした認識手法の提案を行った．その結果としては，ほとんどの場合で前の章で算出した前のフレームと後のフレームが同じ組み合わせで生成した特徴ベクトルを用いた際の F 値よりも低い結果となった．また，前の章において，ジェスチャを開始したタイミングと，それ以降のフレームが 10 フレーム以内の最大認識精度である 0.896 を下回る結果となった．そのため，ジェスチャ中の一連の動作を活かした認識手法を用いても認識精度の改善にはつながらなかったと考えられる．

表 19　時系列の細分化による認識手法を用いた際の認識精度
（背景が赤いセルは提案手法を用いたことで認識精度が向上した組み合わせ）

前フレーム	後フレーム	分割数	分類器	Precision	Recall	F-measure
2	10	6	RF	0.8896	0.8844	0.8870
			SVM	0.8650	0.8378	0.8511
3	9	4	RF	0.8653	0.8589	0.8621
			SVM	0.8543	0.8244	0.8391
4	8	3	RF	0.8505	0.8422	0.8464
			SVM	0.8515	0.8233	0.8372
	10	7	RF	0.8635	0.8600	0.8617
			SVM	0.8548	0.8278	0.8411
5	10	3	RF	0.8689	0.8633	0.8661
			SVM	0.8670	0.8444	0.8556
6	6	2	RF	0.8201	0.8144	0.8173
			SVM	0.8216	0.7944	0.8078
	9	5	RF	0.8632	0.8556	0.8593
			SVM	0.8518	0.8278	0.8396
	10	8	RF	0.8645	0.8633	0.8639
			SVM	0.8544	0.8311	0.8426
7	7	2	RF	0.8486	0.8422	0.8454
			SVM	0.8431	0.8167	0.8297
8	8	2	RF	0.8547	0.8522	0.8535
			SVM	0.8541	0.8300	0.8419
	8	4	RF	0.8557	0.8478	0.8517
			SVM	0.8567	0.8344	0.8454
	10	9	RF	0.8691	0.8656	0.8673
			SVM	0.8527	0.8311	0.8418
9	9	2	RF	0.8545	0.8511	0.8528
			SVM	0.8635	0.8411	0.8521
	9	6	RF	0.8516	0.8478	0.8497
			SVM	0.8487	0.8278	0.8381
10	10	2	RF	0.8783	0.8722	0.8753
			SVM	0.8785	0.8589	0.8686
	10	4	RF	0.8855	0.8811	0.8833
			SVM	0.8683	0.8478	0.8579
	10	10	RF	0.8813	0.8767	0.8790
			SVM	0.8537	0.8322	0.8428

表 20 時系列の細分化による認識手法を用いた際に
最も認識精度が高い組み合わせの特徴ベクトルで算出した 15 人分の混同行列

	判定結果		
	左	右	待機
左	271	9	20
右	12	276	12
待機	32	19	249

表 21 前フレーム 10，後フレーム 10，分割数 2 の
特徴ベクトルで算出した 15 人分の混同行列

	判定結果		
	左	右	待機
左	261	11	28
右	8	271	21
待機	33	14	253

表 22 前フレーム 10，後フレーム 10，分割数 5 の
特徴ベクトルで算出した 15 人分の混同行列

	判定結果		
	左	右	待機
左	273	4	23
右	11	272	17
待機	33	19	248

表 23 前フレーム 10，後フレーム 10，分割数 10 の
特徴ベクトルで算出した 15 人分の混同行列

	判定結果		
	左	右	待機
左	271	10	19
右	9	270	21
待機	36	16	248

表 24 最も認識精度が高かった組み合わせにおける
実験協力者ごとの認識精度

実験協力者	RandomForest		
	Precision	Recall	F-measure
A	0.9565	0.9500	0.9532
B	0.7768	0.7500	0.7632
C	0.9516	0.9500	0.9508
D	0.7852	0.7833	0.7842
E	0.9042	0.9000	0.9021
F	0.9507	0.9500	0.9504
G	1.0000	1.0000	1.0000
H	0.9841	0.9833	0.9837
I	0.8538	0.8500	0.8519
J	0.9286	0.9167	0.9226
K	0.9060	0.9000	0.9030
L	0.8868	0.8833	0.8850
M	0.7970	0.7833	0.7901
N	0.9000	0.9000	0.9000
O	0.7631	0.7667	0.7649

表 25 時系列の細分化による認識手法を用いた際に最も認識精度が
高い組み合わせの特徴ベクトルで算出した実験協力者 B の混同行列

	判定結果		
	左	右	待機
左	18	0	2
右	1	18	1
待機	2	0	18

表 26 時系列の細分化による認識手法を用いた際に最も認識精度が
高い組み合わせの特徴ベクトルで算出した実験協力者 D の混同行列

	判定結果		
	左	右	待機
左	16	0	4
右	0	18	2
待機	2	5	13

表 27 時系列の細分化による認識手法を用いた際に最も認識精度が
高い組み合わせの特徴ベクトルで算出した実験協力者 M の混同行列

	判定結果		
	左	右	待機
左	15	0	5
右	1	16	3
待機	1	3	16

表 28 時系列の細分化による認識手法を用いた際に最も認識精度が
高い組み合わせの特徴ベクトルで算出した実験協力者 O の混同行列

	判定結果		
	左	右	待機
左	17	0	3
右	0	17	3
待機	5	3	12

第 8 章　位置変化に適応する手法

8.1. 実験目的

これまでに我々が行ってきた研究[6]では，ポケットにスマートフォンを入れる向きや位置が変わった際の認識精度への影響についての考慮を行わずに，特定の向きにおける認識精度を算出していた．しかし，スマートフォンを入れる際には，特定のポケットや向き，位置にスマートフォンを入れる状況は稀である．そのため，ポケットにスマートフォンを入れる際に，データセット構築時とはスマートフォンを入れたポケットや向き，位置を変えて入れた場合には，データセット構築時とは加速度や角速度などの軸の正方向の向きが変化することにより，ジェスチャデータがデータセット構築時のデータから変化し，認識精度に影響を及ぼすことが考えられる．そのため，ポケット内でのスマートフォンの位置変化による認識精度への影響について検証する．

第 4 章のデータセット構築では，ポケットにスマートフォンを入れる際の向きを，マイク部分が下になり，ディスプレイ部分が足に触れない向きになるように実験の統制を図っていた．しかし，本章ではポケット内でのスマートフォンの位置変化による認識精度への影響を調査する必要があるため，新たにデータセット構築を行う．

8.2. 位置変化のあるデータセット構築

データセット構築では，20 歳〜25 歳の大学生 7 人の実験協力者に，両足による踵の上げ下げのジェスチャを，第 4 章で提案したデータセット構築用システムを用いて行ってもらった．本実験ではポケットにスマートフォンを入れる際のスマートフォンの向きのパターンは以下の 4 パターンである．そのため，本データセット構築では右足と左足の 2 パターン，スマートフォンをポケットに入れる際の 4 パターンのすべての組み合わせ(計 8 パターン)のそれぞれで 40 回ずつ両足のかかとの上げ下げのジェスチャを行った際の合計 320 回のジェスチャを 1 セットとした．そして，実験協力者には 2 セットの合計 640 回かかとの上げ下げのジェスチャをしてもらい，データセットを構築した．

パターン①.　マイク部分が下になり，ディスプレイ部分が足に触れない向き
パターン②.　マイク部分が下になり，ディスプレイ部分が足に当たる向き
パターン③.　マイク部分が上になり，ディスプレイ部分が足に触れない向き
パターン④.　マイク部分が上になり，ディスプレイ部分が足に当たる向き

8.3. 位置変化の影響検討

位置変化による認識精度への影響調査では，ポケット内のスマートフォンの向きや位置が

認識精度へ与える影響について検証する．本章では，まずポケット内でのスマートフォンの向きが与える認識精度への影響を調査するため，まずスマートフォンを入れるポケットと向きが，右足の前ポケットで，マイク部分が下になり，ディスプレイが足に触れない向きでのジェスチャデータを基に，ほかの向きでの認識精度の算出を行う．次に，右足すべての向きで学習した分類器を用いた際の認識精度の算出を行う．

認識精度の算出に用いる分類器と前フレームと後フレームの組み合わせは前の章を参考に，分類器には Random Forest を用いて，前フレームを 10 フレーム使用し，後フレームを 10 フレーム使用する．Random Forest におけるパラメータはそれぞれデフォルトのパラメータを使用し，特徴ベクトルの値は標準化した．また，前の節で行ったデータセット構築では，合計 40 回のジェスチャ（右足によるジェスチャを 20 回，左足によるジェスチャを 20 回）を 1 セットとし，それぞれのパターンで 2 セット行ってもらった．第 5 章と同じように，ジェスチャのデータ数と待機状態のデータ数の偏りを考慮し，同じデータ数となるようにアンダーサンプリングを行った．ここで，学習に使用するパターンの認識精度の算出の際には，120 回分のデータ（左足によるジェスチャデータ 40 回分，右足によるジェスチャデータ 40 回分，待機状態のデータ 40 回分）から，8 割のデータ(32 回分のデータ)を用いて学習を行い，残りの 2 割のデータ（8 回分のデータ）を用いて認識精度の算出を行った．学習に使用しないパターンの認識精度の算出ではアンダーサンプリングは行わずに 120 回分（右足 40 回分，左足 40 回分，待機 40 回分）のデータを用いて認識精度の算出を行った．今回使用するセンサは第 6 章と同じく加速度，角速度，マイクセンサにより収集した音声データに対してフーリエ級数展開を行って得た最大音圧と，最大音圧の周波数を使用した．また今回使用する特徴量は第 5 章の結果から，最大値と最小値，平均値，中央値を用いる．

8.4. 位置変化の影響実験の結果と考察

まず，表 29 がスマートフォンを入れるポケットと位置と向きが右足の前ポケットで，パターン①のみで学習した分類器を用いた時に，ほかのパターンのジェスチャを分類した際の認識精度を算出した表である． 表 30 がスマートフォンをポケットに入れる位置と向きが右足の前ポケットで，パターン①〜④の計 4 パターンで学習した分類器を用いた時に，ほかのパターンのジェスチャを分類した際の認識精度を算出した表である． 表 31 がスマートフォンを入れる位置と向きが右足と左足の前ポケットのパターン①〜④の計 8 パターンで学習したモデルを用いた時に，それぞれのパターンのジェスチャを分類した際の認識精度を算出した表である．表 32 がスマートフォンを入れるポケットと位置と向きが左足の前ポケットで，パターン①のみで学習したモデルを用いた時に，ほかのパターンのジェスチャを分類した際の認識精度を算出した表である． 表 33 がスマートフォンを入れる位置と向きが左足の前ポケットで，パターン①〜④の計 4 パターンで学習したモデルを用いた時に，ほかのパターンのジェスチャを分類した際の認識精度を算出した表である．

表 29 の結果から，向きが一致しない場合には，認識精度が悪化した．つまり，ポケット内でのスマホの向きが大きく変わった場合には認識精度が悪化する可能性が示唆されている． 表 29 と表 32 の結果から分かるように同じパターンであっても，右足のポケットに入っているか左足のポケットに入っているかというポケットの違いが認識精度に影響していた．表 30 と表 33 はそれぞれの足のポケットにおける全ての向きのパターン（4 パターン）で学習したモデルを用いてそれぞれの組み合わせのテストデータを用いて認識精度を算出した表であるが，それぞれの足で学習したモデルは反対の足のポケットにスマートフォンを同じ向きで入れた場合にも認識精度が悪化していることが分かる．また，4 パターンで学習させたモデルを用いて，スマートフォンを入れている足のパターン①の認識精度を算出したが，1 パターンのみで学習させたモデルで算出した認識精度から低下していることが分かる．つまり，スマートフォンのポケット内での向きと位置によって認識精度が悪化することが明らかとなった．

表 29　右足のパターン①のみで学習させたときの
ポケット内のスマートフォンの位置と向きによる認識精度への影響

ポケット	向き	Precision	Recall	F-measure
右足	パターン①	0.9196	0.9107	0.9151
	パターン②	0.2882	0.3417	0.3127
	パターン③	0.5372	0.5405	0.5388
	パターン④	0.4226	0.4488	0.4353
左足	パターン①	0.5921	0.5536	0.5722
	パターン②	0.3949	0.4286	0.4110
	パターン③	0.4557	0.4845	0.4697
	パターン④	0.5248	0.5071	0.5158

表 30　右足の 4 パターンで学習させたときの
ポケット内のスマートフォンの位置と向きによる認識精度への影響

ポケット	向き	Precision	Recall	F-measure
右足	パターン①	0.8742	0.8631	0.8686
	パターン②	0.8641	0.8512	0.8576
	パターン③	0.9223	0.9167	0.9195
	パターン④	0.8886	0.8810	0.8848
左足	パターン①	0.4675	0.4702	0.4689
	パターン②	0.4459	0.4369	0.4413
	パターン③	0.4117	0.3774	0.3938
	パターン④	0.4580	0.4655	0.4617

表 31　右足と左足の 4 パターンで学習させたときの
ポケット内のスマートフォンの位置と向きによる認識精度への影響

ポケット	向き	Precision	Recall	F-measure
右足	パターン①	0.7903	0.7798	0.7850
	パターン②	0.8213	0.8036	0.8124
	パターン③	0.8403	0.8393	0.8398
	パターン④	0.8370	0.8155	0.8261
左足	パターン①	0.8774	0.8452	0.8610
	パターン②	0.8459	0.8333	0.8396
	パターン③	0.8517	0.8333	0.8424
	パターン④	0.8428	0.8155	0.8289

表 32　左足のパターン①のみで学習させたときの
ポケット内のスマートフォンの位置と向きによる認識精度への影響

ポケット	向き	Precision	Recall	F-measure
右足	パターン①	0.6137	0.5833	0.5981
	パターン②	0.3838	0.3945	0.3891
	パターン③	0.4871	0.5048	0.4958
	パターン④	0.3648	0.4167	0.3890
左足	パターン①	0.9883	0.9881	0.9882
	パターン②	0.3180	0.3417	0.3294
	パターン③	0.5587	0.5393	0.5488
	パターン④	0.6378	0.6119	0.6246

表 33　左足の 4 パターンで学習させたときの
ポケット内のスマートフォンの位置と向きによる認識精度への影響

ポケット	向き	Precision	Recall	F-measure
右足	パターン①	0.4532	0.4583	0.4558
	パターン②	0.4307	0.4313	0.4310
	パターン③	0.4665	0.4488	0.4575
	パターン④	0.4113	0.3833	0.3968
左足	パターン①	0.8982	0.8929	0.8955
	パターン②	0.9019	0.8929	0.8974
	パターン③	0.8959	0.8750	0.8853
	パターン④	0.9090	0.8988	0.9039

8.5. 位置変化に適応した認識手法

前の節での結果から位置変化が認識精度へ悪影響を与えていることが明らかとなった．そのため，位置変化による認識精度への影響を軽減する必要がある．そこで，ポケット内でのスマートフォンの状態を表す特徴量を用いることで，スマートフォンの位置変化による認識精度への影響を軽減する．ここでは，これまでに用いた特徴量以外にも，地磁気と加速度センサから算出されるスマートフォンの傾きの特徴量を利用することで，位置変化に適応した認識手法を提案する．

8.6. 性能評価

位置変化に適応した認識手法を用いることで，ポケット内のスマートフォンの向きや位置が認識精度へ与える影響を軽減することができるかを検証する．具体的には，これまでに用いていたセンサ以外に傾きのセンサを用いることで，認識精度へ与える影響がどのように変化するのか調査する．本章では，まずポケット内でのスマートフォンの向きが与える認識精度への影響を調査するため，まずスマートフォンを入れるポケットと向きが，右足の前ポケットで，マイク部分が下になり，ディスプレイが足に触れない向きでのジェスチャデータを基に，ほかの向きでの認識精度の算出を行う．次に，スマートフォンを入れる位置と向きが右足の前ポケットのパターン①〜④の計 8 パターンで学習した分類器で，他の向きでの認識精度の算出を行う．

認識精度の算出に用いる分類器と前フレームと後フレームの組み合わせは前の章を参考に，分類器には Random Forest を用いて，前フレームを 10 フレーム使用し，後フレームを 10 フレーム使用する．Random Forest におけるパラメータはそれぞれデフォルトのパラメータを使用し，特徴ベクトルの値は標準化した．また，前の節で行ったデータセット構築では，合計 40 回のジェスチャ（右足によるジェスチャを 20 回，左足によるジェスチャを 20 回）を 1 セットとし，それぞれのパターンで 2 セット行ってもらった．第 5 章と同じように，ジェスチャのデータ数と待機状態のデータ数の偏りを考慮し，同じデータ数となるようにアンダーサンプリングを行った．ここで，学習に使用するパターンの認識精度の算出の際には，120 回分のデータ（左足によるジェスチャデータ 40 回分，右足によるジェスチャデータ 40 回分，待機状態のデータ 40 回分）から，8 割のデータ(32 回分のデータ)を用いて学習を行い，残りの 2 割のデータ（8 回分のデータ）を用いて認識精度の算出を行った．学習に使用しないパターンの認識精度の算出ではアンダーサンプリングは行わずに 120 回分（右足 40 回分，左足 40 回分，待機 40 回分）のデータを用いて認識精度の算出を行った．今回使用するセンサは第 6 章と同じく加速度，角速度，マイクセンサにより収集した音声データに対してフーリエ級数展開を行って得た最大音圧と，最大音圧の周波数と，地磁気と加速度から産出される傾きを使用した．また今回使用する特徴量は第 5 章の結果から，最大値と最小値，

平均値，中央値を用いる．

8.7. 位置変化に適応した手法の実験結果・考察

まず，表 34 と表 37 ではそれぞれ右足のポケットと左足のポケットにパターン①のみのジェスチャデータで学習したモデルを用いた時に，それぞれのパターンのジェスチャを分類した際の認識精度を算出した表である．表 35 と表 38 ではそれぞれ右足のポケットと左足のポケットに，パターン①〜④の計 4 パターンの入れ方をした際のそれぞれのジェスチャデータで学習したモデルを用いた時に，それぞれのスマートフォンの入れ方のパターンにおけるジェスチャを分類した時の認識精度を算出した表である． 表 36 がスマートフォンを入れる位置と向きが，右足と左足の前ポケットのパターン①〜④の計 8 パターンのジェスチャデータで学習したモデルを用いて，それぞれのパターンのジェスチャを分類した時の認識精度を算出した表である．背景が赤い箇所は傾きの特徴量を使用したことで認識精度が向上した組み合わせを表している．表 39 と表 40 が右足の 4 パターンのトレーニングデータで学習させたモデルを用いて，右足と左足の前ポケットでのそれぞれの向きのテストデータを分類した際の 7 人分の混同行列を表した表である．表 41 と表 42 が右足の前ポケットの 4 パターンのジェスチャデータで学習させ，左足の 4 パターンのトレーニングデータで学習させたモデルを用いて，右足と左足の前ポケットでのそれぞれの向きのテストデータを分類した際の 7 人分の混同行列を表した表である．表 43 と表 44 が右足と左足の前ポケットの 4 パターン（計 8 パターン）のジェスチャデータで学習させたモデルを用いて，右足と左足の前ポケットでのそれぞれの向きのテストデータを分類した際の 7 人分の混同行列を表した表である．左足の前ポケットの 4 パターンのジェスチャデータで学習させたモデルを用いたときのそれぞれの組み合わせにおける認識精度を表した表である．表 47 がスマートフォンを入れる位置と向きが右足と左足の前ポケットのパターン①〜④の計 8 パターンで学習した分類器を用いた時に，ほかのパターンのジェスチャを分類した際の認識精度を算出した表である．表 45 と表 46 がそれぞれの足の前ポケットのパターン①のみで学習させたときの Random Forest の認識精度への特徴量の寄与度(%)を表した表である．そして，表 47 と表 48 がそれぞれの足の前ポケットの 4 パターンで学習させたときの Random Forest の認識精度への特徴量の寄与度(%)を表した表である．また，表 49 が右足と左足それぞれ 4 パターンを使用した計 8 パターンで学習させたときの Random Forest の認識精度への寄与度(%)を表した表である．

表 29 と表 32 が傾きを用いなかった際，表 34 と表 37 が傾きの特徴量を用いた際のそれぞれ右足のポケットと左足のポケットにパターン①の向きでスマートフォンを入れたジェスチャデータで学習させた時に，それぞれのポケット内のスマートフォンの位置と向きにおける認識精度への影響を調査した表である．その結果，傾きの特徴量を用いることで両足ジェスチャの認識精度が改善していた．また，それぞれの足の前ポケットに入れる際の 4 パ

ターン全ての向きのトレーニングデータを使用し，表 30 と表 33 が傾きの特徴量を用いずに，表 41 と表 43 が傾きの特徴量を用いて学習させたモデルを用いた際のそれぞれの組み合わせのテストデータを分類した際の認識精度を表した表である．その結果，1 パターンで学習させた場合と同様に傾きの特徴量を用いることで認識精度が改善される結果となった．

表 34　右足のパターン①のみで学習させたときの
ポケット内のスマートフォンの位置と向きによる認識精度への影響

ポケット	向き	Precision	Recall	F-measure
右足	パターン①	0.9263	0.9226	0.9245
	パターン②	0.3378	0.3560	0.3467
	パターン③	0.6217	0.5548	0.5863
	パターン④	0.5549	0.5048	0.5287
左足	パターン①	0.5429	0.5560	0.5494
	パターン②	0.3901	0.3857	0.3879
	パターン③	0.4417	0.4381	0.4399
	パターン④	0.5934	0.5857	0.5895

表 35　右足の 4 パターンで学習させたときの
ポケット内のスマートフォンの位置と向きによる認識精度への影響

ポケット	向き	Precision	Recall	F-measure
右足	パターン①	0.8999	0.8929	0.8964
	パターン②	0.8861	0.8810	0.8835
	パターン③	0.9170	0.9107	0.9138
	パターン④	0.8983	0.8929	0.8955
左足	パターン①	0.4339	0.4440	0.4389
	パターン②	0.3839	0.3845	0.3842
	パターン③	0.4479	0.4381	0.4429
	パターン④	0.4452	0.4417	0.4434

表 36　右足と左足の 4 パターンで学習させたときの
ポケット内のスマートフォンの位置と向きによる認識精度への影響

ポケット	向き	Precision	Recall	F-measure
右足	パターン①	0.8890	0.8750	0.8819
	パターン②	0.8508	0.8452	0.8480
	パターン③	0.8762	0.8631	0.8696
	パターン④	0.8762	0.8631	0.8696
左足	パターン①	0.9091	0.9048	0.9069
	パターン②	0.8957	0.8869	0.8913
	パターン③	0.8630	0.8571	0.8601
	パターン④	0.9221	0.9107	0.9164

表 37　左足のパターン①のみで学習させたときの
ポケット内のスマートフォンの位置と向きによる認識精度への影響

ポケット	向き	Precision	Recall	F-measure
右足	パターン①	0.4689	0.4643	0.4666
	パターン②	0.3501	0.3583	0.3541
	パターン③	0.4579	0.4750	0.4663
	パターン④	0.3952	0.4262	0.4101
左足	パターン①	0.9885	0.9881	0.9883
	パターン②	0.3101	0.3190	0.3145
	パターン③	0.5104	0.4845	0.4971
	パターン④	0.5776	0.5917	0.5846

表 38　左足の 4 パターンで学習させたときの
ポケット内のスマートフォンの位置と向きによる認識精度への影響

ポケット	向き	Precision	Recall	F-measure
右足	パターン①	0.4458	0.4500	0.4479
	パターン②	0.4874	0.4588	0.4727
	パターン③	0.4108	0.3988	0.4047
	パターン④	0.3796	0.3560	0.3674
左足	パターン①	0.8909	0.8869	0.8889
	パターン②	0.9107	0.9048	0.9077
	パターン③	0.9068	0.8929	0.8998
	パターン④	0.9303	0.9167	0.9234

表 39 右足の 4 パターンのトレーニングデータで学習させたモデルを用いて，右足の前ポケットでのそれぞれの向きのテストデータを分類した際の 7 人分の混同行列

(a) 右足の前ポケットにパターン①の向きで入れた場合

	判定結果		
	左	右	待機
左	49	0	7
右	4	51	1
待機	3	3	50

(b) 右足の前ポケットにパターン②の向きで入れた場合

	判定結果		
	左	右	待機
左	50	1	5
右	3	52	1
待機	7	3	46

(c) 右足の前ポケットにパターン③の向きで入れた場合

	判定結果		
	左	右	待機
左	48	4	4
右	2	52	2
待機	2	1	53

(d) 右足の前ポケットにパターン④の向きで入れた場合

	判定結果		
	左	右	待機
左	54	1	1
右	1	52	3
待機	10	2	44

表 40 右足の 4 パターンのトレーニングデータで学習させたモデルを用いて，左足の前ポケットでのそれぞれの向きのテストデータを分類した際の 7 人分の混同行列

(a)　左足の前ポケットにパターン①の向きで入れた場合

	判定結果		
	左	右	待機
左	86	184	10
右	225	38	17
待機	20	11	249

(b)　左足の前ポケットにパターン②の向きで入れた場合

	判定結果		
	左	右	待機
左	51	226	3
右	220	31	29
待機	22	17	241

(c)　左足の前ポケットにパターン③の向きで入れた場合

	判定結果		
	左	右	待機
左	70	204	6
右	201	65	14
待機	40	7	233

(d)　左足の前ポケットにパターン④の向きで入れた場合

	判定結果		
	左	右	待機
左	100	177	3
右	237	25	18
待機	22	12	246

表 41 左足の 4 パターンのトレーニングデータで学習させたモデルを用いて，右足の前ポケットでのそれぞれの向きのテストデータを分類した際の 7 人分の混同行列

(a) 右足の前ポケットにパターン①の向きで入れた場合

	判定結果		
	左	右	待機
左	64	176	40
右	203	57	20
待機	10	13	257

(b) 右足の前ポケットにパターン②の向きで入れた場合

	判定結果		
	左	右	待機
左	80	181	18
右	176	99	6
待機	31	43	206

(c) 右足の前ポケットにパターン③の向きで入れた場合

	判定結果		
	左	右	待機
左	44	225	11
右	208	51	21
待機	8	32	240

(d) 右足の前ポケットにパターン④で入れた場合

	判定結果		
	左	右	待機
左	32	219	29
右	225	40	15
待機	7	46	227

表 42 左足の 4 パターンのトレーニングデータで学習させたモデルを用いて，左足の前ポケットでのそれぞれの向きのテストデータを分類した際の 7 人分の混同行列

(a) 左足の前ポケットにパターン①の向きで入れた場合

	判定結果		
	左	右	待機
左	54	2	0
右	8	46	2
待機	3	4	49

(b) 左足の前ポケットにパターン②の向きで入れた場合

	判定結果		
	左	右	待機
左	53	2	1
右	2	50	4
待機	2	5	49

(c) 左足の前ポケットにパターン③の向きで入れた場合

	判定結果		
	左	右	待機
左	53	0	3
右	6	47	3
待機	1	5	50

(d) 左足の前ポケットにパターン④の向きで入れた場合

	判定結果		
	左	右	待機
左	56	0	0
右	4	49	3
待機	2	5	49

表 43 右足と左足の 4 パターンのトレーニングデータで学習させたモデルを用いて，
右足の前ポケットでのそれぞれの向きのテストデータを分類した際の 7 人分の混同行列

(a) 右足の前ポケットにパターン①の向きで入れた場合

	判定結果		
	左	右	待機
左	48	3	5
右	7	47	2
待機	3	1	52

(b) 右足の前ポケットにパターン②の向きで入れた場合

	判定結果		
	左	右	待機
左	44	7	5
右	2	51	3
待機	5	4	47

(c) 右足の前ポケットにパターン③の向きで入れた場合

	判定結果		
	左	右	待機
左	48	4	4
右	7	47	2
待機	2	4	50

(d) 右足の前ポケットにパターン④の向きで入れた場合

	判定結果		
	左	右	待機
左	53	3	0
右	7	47	2
待機	5	6	45

表 44 右足と左足の 4 パターンのトレーニングデータで学習させたモデルを用いて，左足の前ポケットでのそれぞれの向きのテストデータを分類した際の 7 人分の混同行列

(a) 左足の前ポケットにパターン①の向きで入れた場合

	判定結果		
	左	右	待機
左	52	4	0
右	7	47	2
待機	3	0	53

(b) 左足の前ポケットにパターン②の向きで入れた場合

	判定結果		
	左	右	待機
左	51	3	2
右	4	51	1
待機	1	3	52

(c) 左足の前ポケットにパターン③の向きで入れた場合

	判定結果		
	左	右	待機
左	47	6	3
右	7	46	3
待機	2	3	51

(d) 左足の前ポケットにパターン④の向きで入れた場合

	判定結果		
	左	右	待機
左	54	0	2
右	7	48	1
待機	4	1	51

表 45　右足のパターン①のみで学習させたモデルを用いた際の
Random Forest の認識精度への特徴量の寄与度(%)

加速度	X 軸	最小値	3.810%	**マイク**	周波数	最小値	0.003%
		最大値	1.415%			最大値	0.084%
		平均値	9.747%			平均値	0.348%
		中央値	5.655%			中央値	0.241%
	Y 軸	最小値	2.798%		最大音圧	最小値	0.230%
		最大値	5.139%			最大値	0.544%
		平均値	4.100%			平均値	0.601%
		中央値	1.606%			中央値	0.363%
	Z 軸	最小値	4.511%	**傾き**	X 軸	最小値	0.820%
		最大値	4.220%			最大値	1.876%
		平均値	2.642%			平均値	1.305%
		中央値	3.164%			中央値	0.675%
角速度	X 軸	最小値	5.686%		Y 軸	最小値	2.927%
		最大値	3.379%			最大値	0.935%
		平均値	4.658%			平均値	2.157%
		中央値	1.857%			中央値	0.582%
	Y 軸	最小値	2.850%				
		最大値	2.017%				
		平均値	1.503%				
		中央値	1.280%				
	Z 軸	最小値	2.991%				
		最大値	4.910%				
		平均値	4.061%				
		中央値	2.299%				

表 46　左足のパターン①のみで学習させたモデルを用いた際の
Random Forest の認識精度への特徴量の寄与度(%)

加速度	X 軸	最小値	5.046%	**マイク**	周波数	最小値	0.147%
		最大値	4.229%			最大値	0.111%
		平均値	6.873%			平均値	0.215%
		中央値	3.818%			中央値	0.122%
	Y 軸	最小値	2.566%		最大音圧	最小値	0.195%
		最大値	8.373%			最大値	0.242%
		平均値	2.227%			平均値	0.282%
		中央値	2.861%			中央値	0.171%
	Z 軸	最小値	3.010%	**傾き**	X 軸	最小値	0.251%
		最大値	2.629%			最大値	2.550%
		平均値	1.514%			平均値	0.848%
		中央値	1.773%			中央値	0.423%
角速度	X 軸	最小値	2.626%		Y 軸	最小値	1.247%
		最大値	5.709%			最大値	0.615%
		平均値	7.668%			平均値	0.415%
		中央値	1.854%			中央値	0.730%
	Y 軸	最小値	1.639%				
		最大値	3.095%				
		平均値	2.167%				
		中央値	1.505%				
	Z 軸	最小値	4.166%				
		最大値	4.881%				
		平均値	6.125%				
		中央値	5.083%				

表 47　右足の 4 パターンで学習させたモデルを用いた際の
Random Forest の認識精度への特徴量の寄与度(%)

加速度	X 軸	最小値	3.473%	**マイク**	周波数	最小値	0.253%
		最大値	3.738%			最大値	0.119%
		平均値	2.802%			平均値	0.617%
		中央値	1.858%			中央値	0.232%
	Y 軸	最小値	4.165%		最大音圧	最小値	0.397%
		最大値	4.515%			最大値	0.549%
		平均値	3.821%			平均値	0.656%
		中央値	2.576%			中央値	0.493%
	Z 軸	最小値	4.689%	**傾き**	X 軸	最小値	1.628%
		最大値	4.309%			最大値	1.311%
		平均値	3.107%			平均値	1.494%
		中央値	1.909%			中央値	1.211%
角速度	X 軸	最小値	5.226%		Y 軸	最小値	1.437%
		最大値	5.598%			最大値	1.543%
		平均値	4.517%			平均値	1.201%
		中央値	1.920%			中央値	1.283%
	Y 軸	最小値	4.610%				
		最大値	4.188%				
		平均値	1.600%				
		中央値	0.961%				
	Z 軸	最小値	4.589%				
		最大値	5.741%				
		平均値	3.493%				
		中央値	2.171%				

表 48　左足の 4 パターンで学習させたモデルを用いた際の
Random Forest の認識精度への特徴量の寄与度(%)

加速度	X 軸	最小値	3.336%	**マイク**	周波数	最小値	0.147%
		最大値	3.232%			最大値	0.123%
		平均値	1.885%			平均値	0.527%
		中央値	1.994%			中央値	0.274%
	Y 軸	最小値	4.306%		最大音圧	最小値	0.435%
		最大値	3.695%			最大値	0.442%
		平均値	3.199%			平均値	0.608%
		中央値	2.125%			中央値	0.453%
	Z 軸	最小値	4.007%	**傾き**	X 軸	最小値	1.583%
		最大値	5.411%			最大値	1.304%
		平均値	2.479%			平均値	1.361%
		中央値	1.505%			中央値	1.562%
角速度	X 軸	最小値	5.875%		Y 軸	最小値	2.070%
		最大値	6.043%			最大値	1.676%
		平均値	4.912%			平均値	1.307%
		中央値	2.133%			中央値	1.387%
	Y 軸	最小値	4.375%				
		最大値	3.565%				
		平均値	1.346%				
		中央値	0.961%				
	Z 軸	最小値	5.117%				
		最大値	7.164%				
		平均値	3.600%				
		中央値	2.475%				

表 49　右足と左足の 4 パターンで学習させたモデルを用いた際の
Random Forest の認識精度への特徴量の寄与度(%)

加速度	X 軸	最小値	3.485%	**マイク**	周波数	最小値	0.292%
		最大値	3.563%			最大値	0.160%
		平均値	2.658%			平均値	0.685%
		中央値	2.062%			中央値	0.415%
	Y 軸	最小値	2.501%		最大音圧	最小値	0.517%
		最大値	2.479%			最大値	0.592%
		平均値	1.995%			平均値	0.884%
		中央値	1.769%			中央値	0.644%
	Z 軸	最小値	4.310%	**傾き**	X 軸	最小値	1.995%
		最大値	4.335%			最大値	2.192%
		平均値	2.655%			平均値	2.018%
		中央値	1.804%			中央値	2.109%
角速度	X 軸	最小値	4.369%		Y 軸	最小値	2.476%
		最大値	4.422%			最大値	2.090%
		平均値	2.666%			平均値	2.438%
		中央値	1.833%			中央値	2.356%
	Y 軸	最小値	4.285%				
		最大値	4.123%				
		平均値	2.257%				
		中央値	1.897%				
	Z 軸	最小値	5.414%				
		最大値	6.065%				
		平均値	4.466%				
		中央値	2.723%				

第 9 章　総合考察

本研究では，ジェスチャ認識においてどの特徴量がどのくらい認識に寄与するかの度合いを明らかにした．その結果，特に認識に寄与したのが「角速度 Z 軸の最大値」「加速度 Y 軸の最小値」「角速度 X 軸の最小値」の 3 つの特徴量であった．これらが寄与した理由としては，ポケットに入れたスマートフォンは動きが制限されているため，足の動きと角速度の軸が一致しやすかったと考えられる．これまでのポケット内のスマートフォンでジェスチャ認識を行う研究[5]では，加速度のみを使用することが主であったが，これらの研究において角速度も用いることで認識精度の向上が期待できることが考えられる，

また，ジェスチャ中の一連の動作を活かした認識手法を提案したが，結果としては認識精度の向上にあまり寄与しなかった．これはユーザによるジェスチャのスピードが異なっていたためであると考えられる．一連の動作を活かした認識手法は，本研究のようなポケット内のスマートフォンを用いたジェスチャ認識に限らず，様々なジェスチャ認識研究に応用できると考えている．その場合には，ユーザによるジェスチャのスピードを考慮して，ジェスチャデータを分割することで一連の動作を活かした認識手法によって認識精度が改善されると考えられる．

さらに，スマートフォンの位置変化による認識精度への影響を軽減するため，傾きの特徴量を用いた．その結果傾きの特徴量は，スマートフォンの位置変化の影響を軽減することが出来た．そのため，スマートフォンを利用したジェスチャ認識を行う際には，スマートフォンの向きや位置の変化を考慮するために傾きを使うことで，認識精度への影響を軽減可能であると考えられる．

他にも，予備動作のデータを利用して算出した認識精度の方が，予備動作のデータを利用しないで算出した認識精度よりも高かった．そのため，予備動作を利用することでユーザがジェスチャを開始してから早い時間でジェスチャを認識することができるため，ユーザの待ち時間を減らすことができ，快適にシステムを利用可能になると考えられる．しかし，本研究では実際に予備動作を用いたプロトタイプシステムを実装することが出来ていないため，実際にシステムに組み込むことは今後の課題である．

第１０章　まとめ

本研究では,これまでに我々が行ってきたポケット内のスマートフォンを用いた両足ジェスチャによる研究において明らかにした問題について述べ，それぞれの問題を解決するための手法を提案した．これまでの我々の研究[6]で行った使用実験では，プロトタイプシステムを使用した後に，体感認識精度について自由記述で 0〜100%までの認識精度を回答してもらった．その結果，期待したほどの認識精度とは感じられていなかった．その要因としては，プロトタイプシステムでの認識精度がユーザの期待したほどの認識精度ではなかったことがあげられる．次に，その要因としては，プロトタイプを使用した後に，認識までにかかる時間がどうだったかを回答してもらったアンケートにおいて 8 人いる実験協力者のうち 7 人が認識までにかかる時間が遅いと感じている傾向にあることが明らかになった．そのため，認識までにかかる時間の遅さを認識精度の悪さと感じていたと考えられる．

まず認識精度の改善を行うために認識手法の再検討を行い認識精度の改善を目指した．これまでの我々の研究[6]では波形の形状を表す最大値と最小値を用いて特徴ベクトルの生成を行ってきたが，さらに波形の形状を表す特徴量を使用することで認識精度の改善を図った．そのため，特徴量の再選定を行い，その特徴量により生成した特徴ベクトルを用いた際の認識精度と，これまでの我々の研究[6]において使用していた特徴量により生成した特徴ベクトルを用いた際の認識精度の比較を行い，波形の形状を表す特徴量をさらに取り入れることで認識精度が向上するかを検証した．その結果，最大値と最小値で生成した特徴ベクトルによる認識精度より，最大値と最小値，平均値，中央値の特徴量を用いて生成した特徴ベクトルによる認識精度の方が高い傾向にあった．また，認識精度に寄与した度合いを算出した結果からも，平均値や中央値を使用することで認識精度の改善がみられたと考えられる．

次に認識時間の短縮を行うために認識に使用するジェスチャデータの再検討を行うため,ジェスチャを始める直前に無意識に行われる構えの動作である予備動作を用いた認識手法を提案した．その結果，すべてのパターンにおいて予備動作のジェスチャデータを用いることにより認識精度の改善がみられた．

また，これまでの我々の研究[6]においてはポケット内でのスマートフォンの位置変化による認識精度への影響を考慮していなかった．そこで，まずポケット内のスマートフォンの位置変化が認識精度に与える影響について検証した．その結果，ポケット内のスマートフォンの位置変化により認識精度が悪化することが明らかとなった．そこで，その位置変化による認識精度の悪化を低減するため，傾きの特徴量を使用した特徴ベクトルを用いることで位置変化による認識精度への影響を低減できるかを検証した．その結果，傾きの特徴量を用いることで，認識精度が改善されることが明らかとなった．

謝辞

Web 公開版からは謝辞を削除しました.

参考文献

[1] Igarashi, T., Hughes. J.: Voice as Sound using non-verbal voice input for interactive control, In Proc. of UIST'01, pp.155-156(2001).

[2] Matthies, D.J.C., Muller, F., Anthes C. and Kranzmuller D.: Shoe Sole Sense:Proof of Concept for a Wearable Foot Interface for Virtual and Real Enviroments, In Proc of VRST '13, pp.93-96(2013).

[3] Matthies, D.J.C., Roumen, T.: CapSoles: who is walking on what kind of floor?, In Proc MobileHCI '17, p.9(2017).

[4] Fukahori, K., Sakamoto, D,. Igarashi, T.: Exploring Subtle Foot Plantar-based Gestures with Sock-placed Pressure Sensors, In Proc CHI'15, p.10(2015).

[5] Scott, J., Dearman, D., Yatani and K. Truong, N.K.: Sensing Foot Gestures from the Pocket, In Proc. of UIST '10, pp.199-208(2010).

[6] 田村柾優紀, 中村聡史: ポケット内のスマートフォンによる両足ジェスチャ認識手法の提案と分析, 研究報告グループウェアとネットワークサービス, p.8(2017).

[7] Bao, L., Intille S.S.: Activity Recognition from User-Annotated Acceleration Data, In Proc. of Pervasive '04, Vol.LNCS3001, pp.158-175(2004).

[8] 村尾和哉, Laerhoven, K.. 寺田努, 西尾章治郎: センサのピーク値を用いた状況認識手法とその評価. 除法処理学会研究報告, Vol.51, No.3, pp.1068-1077(2010).

[9] 村尾和哉, 寺田務. 加速度センサの定常性による動作認識手法. 情報処理学会論文誌, Vol. 52, No.6, pp.1968-1979(2011).

[10] 大内一成, 土井美和子: Activity Analyzer：携帯電話搭載センサによるリアルタイム生活行動認識システム, 情報処理学会研究報告, Vol.2011-UBI-30 No.3, p.8(2011).

[11] 河内智志, 薛媛, 藤波香織: 携帯端末の身体上格納場所判定機能のスマートフォンへの実装. インタラクション 2011, pp.531-534(2011).

[12] 米田圭祐, 望月祐洋, 西尾信彦: 気圧センシングを用いた行動認識手法. 研究報告モバイルコンピューティングとユビキタス通信, Vol.14, p. 8(2014).

[13] Chan, L., Liang, R., Tsai, M., Cheng, K., Su, C., Mike Y., Chen, Cheng W., Chen B.. FingerPad: Private and Subtle Interaction Using Fingertips. In Proc. ACM UIST '13 , pp.255-260(2013).

[14] Cohn, G., Morris, D., Patel, S., Tan, D.: Humantenna : Using the Body as anAntena for Real-Time Whole-BodyInteration. In Proc Of CHI 2012, pp.1901-1910(2012).

[15] Han, T., Alexander, J., Karnik, A., Irani, P., Subramanian, S.. Kick: investigating the use of kick gestures for mobile interactions. In Proc Of ACM UIST '11 , pp.29-32(2011).

[16] Chan, L., Hsieh C., Chen Y., Yang, S., Huang D., Liang, R., Chen B.. Cyclops: Wearable and

Single-Piece Full-Body Gesture Input Devices, In Proc of CHI '15 Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp.3001-3009(2015).

[17] 安藤宗孝, 久保勇貴, 志築文太郎, 高橋伸: Mimi Sense:：外耳道内の気圧変化を利用した下顎運動認識認識システム. In Proc of WISS 2016, p.6(2016).

[18] 塚田浩二, 安村通晃: Ubi-Finger：モバイル指向ジェスチャの研究. 情報処理学会論文誌, Vol.43, No.12, pp.3675-3684(2002).

[19] 山本哲也, 寺田務, 塚本昌彦, 義久智樹: ジョギング時における情報機器利用のための足ステップ入力方式, 情報処理学会論文誌, Vol.50, No.12, pp.2881-2888(2009).

[20] 山本哲也, 寺田務, 塚本昌彦: ウェアラブルコンピューティングのための手足を使った状況依存ジェスチャ入力手法, ヒューマンインターフェース学会論文誌, Vol.14, No.2, pp.21-30(2012).

[21] 澤田秀之, 橋本周司: 加速度センサを用いたジェスチャ認識と音楽制御への応用. 電子情報通信学会論文誌, 基礎・境界 J79-A(2), pp.452-459(1996).

[22] 澤田秀之, 橋本周司, 松島俊明: 運動特徴と形状特徴に基づいたジェスチャ認識と手話認識への応用, 情報処理学会論文誌, Vol.39, No.5, pp.1325-1333(1998).

[23] 管家浩之, 竹川佳成, 寺田努, 塚本昌彦: Airstic Drum：実ドラムと仮想ドラムを統合するためのドラムスティックの構築, 情報処理学会論文誌, Vol. 54, No.4, pp.1393-1401(2013).

[24] Lv, Z.: Wearable Smartphone: Wearable Hybrid Framework for Hand and Foot Gesture Interaction on Smartphone, In Proc of 2013 IEEE International Conference on Computer Vision Workshops, pp.436-443(2013).

[25] Gong, J., Zhang, Y., Zhou, X., Yang, X. D.: Pyro: Thumb-Tip Gesture Recognition Using Pyroelectric Infrared Sensing, In Proc of UIST'17, p.11(2017).

[26] Lin, J.W., Wang, C., Huang, Y.Y., Chou K.T.: BackHand: Sensing Hand Gestures via Back of the Hand, In Proc Of UIST'15, pp.557-564(2015).

[27] Augsten, T., Kaefer, K., Meusel, R., Fetzer, C., Kanitz, D., Stoff, T., Becker T.: Multi Toe: high-precision interaction with back-projected floor based on high-resolution multi-touch input, In Proc Of UIST'10,(2010).

[28] Marin, G., Dominio, F., Zanuttigh, P.: Hand Gesture Recognition With Leap Motion And Kinect Devices, 2014 IEEE International Conference on Image Processing, p.7(2014).

[29] 小荒健吾, 西川敦, 進戸健太郎, 石井薫, 安井陽一, 宮崎丈夫: 動きの変化に着目した動作パターンの１次元符号化に基づく身振り認識, 情報処理学会論文誌, Vol.44, No.2, pp.466-477(2003).

[30] 高橋真人, 入江耕太, 寺林賢司, 梅田和昇: 周波数検知に基づくジェスチャ認識, 日本ロボット学会誌, Vol.28, No.6, pp.756-765(2010)..

[31] Davis, J.W., Bobick, A. F.: The Representation and Recognition of Human Movement Using Temporal Templates, In Proc. 1997 IEEE Computer Society Conference, pp.928-934(1997).

[32] Molchanov, P., Gupta, S., Kim, K., and Kautz, J.: Hand Gesture Recognition with 3D Convolutional Neural Networks, In Proc of IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.38 , pp.1583-1597(2016).

[33] Vogler, C., Metaxas, D.: ASL Recognition Based on a Coupling Between HMMs and 3D Motion Analysis, In Proc of IEEE Sixth International Conference on Computer Vision, pp. 363-369(1998).

[34] Cutler, R., Turk, M.: View-based Interpretation of Real-time Optical Flow for Gesture Recognition, In Proc of International Conference on Face & Gesture Recognition, p.6(1998)

[35] Sun, K., Yu, C., Shi, W., Lu, L. and Shi, Y.: Lip-Interact: Improving Mobile Device Interaction with Silent Speech Commands, InProc of UIST'18, pp.581-593(2018).

[36] 呉海元, 小林弘知, 陳謙, 塩山忠義, 島田哲夫: 色彩動画像からの頭部ジェスチャ認識システム. 情報処理学会論文誌, Vol.37, No.6, pp.1234-1242(1996).

[37] 西村拓一, 向井理朗, 野崎俊輔, 岡隆一: 動作者適応のためのンライン教示可能なジェスチャ動画像のスポッティング認識システム. 電気情報通信学会論文誌, 1998 II–情報処理 J81-D-2(8), pp.1820-1830(1998).

[38] 玉城絵美, 味八木崇, 暦本純一: インタラクションシステムのための高次元な 3 次元ハンドジェスチャ認識手法. 情報処理学会文誌, Vol.51, No.2, pp.229-239(2010).

[39] Negulescu, M., McGrenre, J.: Grip Change as an Information Side Channel for Mobile Touch Interaction, CHI'15, pp.1519-1522(2015).

[40] Eardley, R., Roudaut, A., Gill, S., and Thompson, S. J.: Understanding Grip Shifts: How Form Factors Impact Hand Movements on Mobile Phones, In Proc of CHI'17, pp.4680-4691(2017).

[41] Eardley, R., Roudaut, A., Gill, S. and Thompson, S.: Investigating How Smartphone Movement is Affected by Body Posture, In Proc of CHI'18, p.8(2018).

[42] Eardley, R., Roudaut, A., Gill, S. and Thompson, S.: Designing for Multiple Hand Grips and Body Postures within the UX of a moving Smartphone, In Proc of DIS'18, pp.611-621(2018).

[43] Alvina, J., CarlaF.Griggio, C. F., Bi X. and Mackey W. E.: CommandBoard: Creating a General-Purpose Command Gesture Input Space for Soft Keyboards, In Proc of UIST'17, pp.17-28(2017).

[44] 加藤直樹, 大美賀かおり, 中川正樹: 携帯型ペン入力情報機器におけるペンジェスチャ入力指示インタフェース, 情報処理学会論文誌. Vol.41, No.9, pp.2413-2422(2000).

[45] 石原進, 太田雅敏, 行方エリキ, 水野忠則. 端末自体の動きを用いた携帯端末向け個人認証. 情報処理学会論文誌, Vol.46, No.12, pp.2997-3007(2005).

[46] Gong, J., Xu, Z., Guo, Q., Seyed, T., Chen, X.A., Bi, X., Yang X. D.: WrisText: One-handed Text Entry on Smartwatch using Wrist Gestures, In Proc of CHI'18, p.14(2018).

[47] Xu, C., Pathak, H. P., Mohapatra, P.: Finger-writing with Smartwatch: A Case for Finger and Hand Gesture Recognition using Smartwatch, In Proc of HotMobile'15, pp.9-14(2015).

[48] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine Learning in Python, The Journal of Machine Learning Research, pp. 2825-2830(2011).

[49] Miller, R.B.: Response Time in Man-computer Conversational Transactions, Fall Joint Computer Conferences, pp.267-277(1968).