

加速度センサを用いたジェスチャ認識における 視聴覚フィードバックの影響調査

吉田 楽^{1,a)} 村尾 和哉^{1,b)} 寺田 努^{1,2,c)} 塚本 昌彦^{1,d)}

概要: 加速度センサを用いたジェスチャ認識において、一般的には、事前に採取した数回分のデータを学習データとして登録し、学習データと未知のデータの距離を計算する手法が用いられている。しかし、ユーザの疲労やジェスチャフォームの忘却などによるジェスチャの時間的変化を考慮していないため、使用開始時の数サンプルを学習データとする方法では時間の経過とともにジェスチャが変化し、誤認識を引き起こす可能性がある。筆者らは先行研究において、ジェスチャが安定する箇所をリアルタイムで発見することでジェスチャの時間的変化を考慮した学習データ収集手法を提案している。しかしながら、提案した手法ではジェスチャの安定に時間を要する場合に、採取する学習データのサンプル数が増大することが分かった。そこで本研究では、自身が行っているジェスチャ動作が安定していることを知れば、その状態を継続し、学習データを得るまでのジェスチャの試行回数を削減できるのではないかと考え、学習データ採取時にユーザに視聴覚フィードバックを与え、フィードバックがジェスチャに与える影響を評価した。

1. はじめに

近年のマイクロエレクトロニクス技術の発展によるコンピュータの小型化、高性能化、軽量化に伴い、ユーザがコンピュータを常に身に着けて生活するウェアラブルコンピューティングに対する注目が高まっている。さまざまなセンサやコンピュータを身に着けるウェアラブルコンピューティング環境では、装着型センサを用いてユーザの行動や状況、周りの環境を認識してさまざまなサービスが提供される。特に、ウェアラブルコンピューティングの発展に伴い、加速度センサを用いたジェスチャ認識に関する研究が数多く行われており、その応用が注目されている。携帯電話や家庭用ゲーム機のコントローラにも加速度センサが内蔵され、センサの値から端末の向きや動きなどを検出してユーザの直観的な操作や複雑な入力を実現している。また、手に装着したセンサから日常生活における手の動きを詳細に記録することでリッチなライフログを構築できるなど、今後の日常生活において加速度センサを用いたジェスチャ認識を利用する機会がさらに増えていくことが

予想される。

ジェスチャ認識では、あらかじめ認識対象のジェスチャのデータを学習データとして取得しておき、認識したい未知のジェスチャのデータと学習データを比較することでユーザの行っているジェスチャを認識する。これまでに提案されているジェスチャ認識システムでは、事前に行った数回分のジェスチャ動作のデータを学習データとして使用しているが、ユーザの動作は周囲の環境やユーザの状態によって変化し、学習時の動作から異なると考えられる。例えば、ユーザが学習時に行ったジェスチャの詳細な動作を忘れていたり、ユーザが1日の生活で疲労することで手の振りが小さくなって学習したジェスチャ動作と異なることで認識精度が低下する。しかし、従来の研究の多くはジェスチャ動作の変化を考慮していない。

筆者らは先行研究 [1] において、特定のジェスチャを1日200回繰り返し、7日間のデータ採取からユーザのジェスチャ動作の時間的変化を調査し、ジェスチャが安定する箇所をリアルタイムで発見することでジェスチャ変化の影響を考慮した学習データ収集手法を提案した。結果から、提案した手法を用いて採取した学習データは採取するまでのジェスチャ試行回数は増大するが、以降のデータに対して高い認識精度を示すことを確認した。一方、提案した手法ではジェスチャの安定に時間がかかる場合に、採取する学習データのサンプル数が極めて多くなることがわかった。そこで本研究では、学習データ採取時にユーザに視聴

¹ 神戸大学大学院工学研究科
Graduate School of Engineering, Kobe University, Hyogo,
Kobe 657-8501, Japan

² 科学技術振興機構 さきがけ
PRESTO, Japan Science and Technology Agency

a) g-yoshida@stu.kobe-u.ac.jp

b) murao@eedept.kobe-u.ac.jp

c) tsutomu@eedept.kobe-u.ac.jp

d) tuka@kobe-u.ac.jp

覚フィードバックを与えることで、ユーザが自身のジェスチャの安定度合を知り、ジェスチャがより早く安定し、採取するサンプル数が削減すると考え、実験からフィードバックがジェスチャに与える影響を評価した。

以降、2章では関連研究を紹介し、3章で加速度センサを用いたジェスチャの認識手法について述べる。4章で筆者らの行った先行研究を紹介し、5章で視聴覚フィードバックがジェスチャに与える影響を評価する。最後に6章でまとめと今後の課題について述べる。

2. 関連研究

Murao らの研究 [2] では、9 個の加速度センサと 9 個の角速度センサを格子状に配置した盤面を用いて 27 種類のジェスチャの認識精度を求め、センサの数、センサの位置およびセンサの種類による認識精度の違いを示している。Agrawal ら [3] は加速度センサを搭載した携帯電話をペンのように持って空中に書いたアルファベットを認識している。加速度のデータを幾何学的な動きに変換し、文字として認識する。いくつかの制約に準拠することにより、83% の認識精度を達成している。空中に書いた文章や図形をメールで送信することもできる。このように、加速度センサを用いたジェスチャ認識に関する研究が行われているが、いずれの研究においてもジェスチャ動作がユーザの状態によって変化することは考慮していない。

これに対して、Liu ら [4] は日々変化するジェスチャ動作に対応する手法を提案している。評価では 8 名の被験者から 3 軸加速度センサ 1 個を用いて、Nokia 研究所が提唱する直線や円などの 8 種類の図形を描くジェスチャを 4000 サンプル以上収集している。認識アルゴリズムは Dynamic Time-Warping (DTW)[5] を採用し、学習したデータを日々更新することによって、高い認識精度を達成している。しかし、学習データの更新にはシステムへの正解情報のフィードバックを行う必要があり、日々学習データを採取しなければならない。また、学習データの更新は一日 1 回であり、日々のジェスチャ動作の変化には対応できるが、一日の中のジェスチャ動作の変化に対応できるかは明らかにされていない。

金ら [6] は聴覚フィードバックが筆記作業に与える影響を調査している。漢字のなぞり書きにおいて、ヘッドホンを装着した被験者に筆記音を聞かせた場合、無音状態の場合と比較して作業量が上昇したことを確認している。聞かせた筆記音をモノラル形式にした場合とステレオ形式にした場合で結果に差が出ており、フィードバックの方法によってジェスチャに与える影響が変わることが明らかとなっている。Brewster ら [7] は、聴覚フィードバックとジェスチャの認識について調査している。Personal Digital Assistant (PDA) を操作する際に、被験者の指の動きを認識して被験者の指の位置に合わせて効果音を聞かせること

により、効果音がない場合と比較して、指で「円を描く」、「N を書く」などのジェスチャの認識精度が向上することを確認している。また、Andersen ら [8] は、視聴覚フィードバックが筆記動作に与える影響に関する実験を行い、タブレットデジタイザでの 20 種類の図形の描き写しにおいて、フィードバックを行わない場合、描いた軌跡を表示した場合、ペンの位置と速度に応じた効果音を聞かせた場合、またその両方を組み合わせた場合の結果から、フィードバックがある方が正確に図形を描けることを確認している。

このように視聴覚フィードバックがジェスチャ動作に影響を与えることが確認されているが、ジェスチャを繰り返す過程においてフィードバックがジェスチャ動作に与える影響を調査した研究は筆者らの知る限り行われていない。

3. 加速度センサを用いたジェスチャ認識

加速度センサを用いたジェスチャ認識では、事前に採取した数回分のジェスチャデータを学習データとして登録しておき、認識したい未知のジェスチャデータと学習データとの距離（類似度）を計算する。距離が最も小さい（類似度が最も大きい）学習データのラベルを未知のデータの認識結果として出力する。

学習データと未知データの 2 つの時系列波形の距離計算の最も簡単な手法はユークリッド距離であるが、時間的歪みの影響を受けやすく、比較するデータのサンプル数が同じでなければならないという欠点がある。そこで本研究では、Dynamic Time Warping (DTW) を用いる。DTW はユークリッド距離のもつ欠点を克服した時系列データの類似度計算手法である。DTW の特徴として、時間的非線形圧縮距離を計算することで、時系列のスケールや部分的な速度に違いがあるデータ間の類似度やサンプル数の異なるデータ間の類似度が求められる。

DTW の具体的なアルゴリズムを以下に示す。長さ m と n の 2 つの離散波形 $X = (x_1, \dots, x_m)$ と $Y = (y_1, \dots, y_n)$ を比較する際、 $m \times n$ 行列 $d(x_i, y_j) = (x_i - y_j)^2$ を定義する。次に、 X および Y のインデックスのペアから成るワーピングパス $W = (w_1, \dots, w_k)$ を探索する。ワーピングパスは以下の 3 つの条件を満足する。

- 境界条件
 $w_1 = (1, 1), w_k = (m, n)$
- 連続性
 $w_k = (a, b), w_{k-1} = (a', b') \Rightarrow a - a' \leq 1 \wedge b - b' \leq 1$
- 単調性
 $w_k = (a, b), w_{k-1} = (a', b') \Rightarrow a - a' \geq 0 \wedge b - b' \geq 0$

距離最小のパスを探索するために、以下の手順を適用する。

(1) Initialization:

$$DTW(0, 0) = 0$$

$$DTW(i, 0) = \infty \text{ for } i = 1, \dots, m$$

$$DTW(0, j) = \infty \text{ for } j = 1, \dots, n$$

(2) Do for $i = 1, 2, \dots, m$:

Do for $j = 1, 2, \dots, n$:

$$DTW(i, j) = d(x_i, y_j) + \min \begin{cases} DTW(i-1, j-1) \\ DTW(i-1, j) \\ DTW(i, j-1) \end{cases}$$

(3) Output:

Return $DTW(m, n)/(m+n)$

上式によって得られたコスト $DTW(m, n)/(m+n)$ が X と Y の距離となる。ここで $m+n$ で除している理由は、DTW 距離はサンプル数に応じて大きくなるためである。コンテキストを認識する際は、ジェスチャの種類を付与した時系列データを学習データとして保持しておき、入力データとすべての学習データとの距離を計算し、最短距離の学習データに付与されているジェスチャを結果として出力する。

4. 先行研究

筆者らはこれまでに疲労や忘却などによるジェスチャ動作の時間変化について調査し、ジェスチャの時間変化を考慮した学習データの決定手法を提案している。本節では筆者らが行った先行研究について述べる。

4.1 予備実験

手首に 3 軸加速度センサを装着した 22~23 歳の男性 5 名の被験者から、「ボールを投げる」ジェスチャと「星を描く」ジェスチャをそれぞれ 3 秒程度の間隔で 200 回行ったものを 1 セットとして、1 日 1 セットを 7 日間採取した。200 回繰り返すことで繰り返しにより変化するデータを、7 日間継続することで日々変化するデータを採取できると考えた。採取した 7 日分のデータのうち 1 日目の 200 回の試行の 1 回目のデータを学習データとして、図 1 のように 7 日分の各データとの距離を DTW を用いて被験者ごとに計算した。同様に、1 日目の 2 回目から 200 回目までの各サンプルを学習データとして残りのデータとの DTW 距離を計算し、この作業を学習データとする日を 1 日目から 7 日まで変えて行った。

実験結果の一例を図 2 に示す。図 2 は、ある被験者の「ボールを投げる」ジェスチャにおいて、1 日目のデータを学習データとしたときの結果である。グラフの縦軸は DTW 距離の平均、横軸は学習データとしたデータの位置を示している。グラフ中の day2 のグラフは 1 日目の i 回目のデータと 2 日目の 200 回分のジェスチャとの DTW 距離の平均を示しており、 i を 1 から 200 まで変化させている。ただし、学習データを選択した日と距離計算を行う対象のジェスチャの試行日が同じ場合のグラフについては、 i 回目のデータを距離計算する対象のデータは i 回目以外の 199 サンプルとしている。ここで、平均値が大きいほど

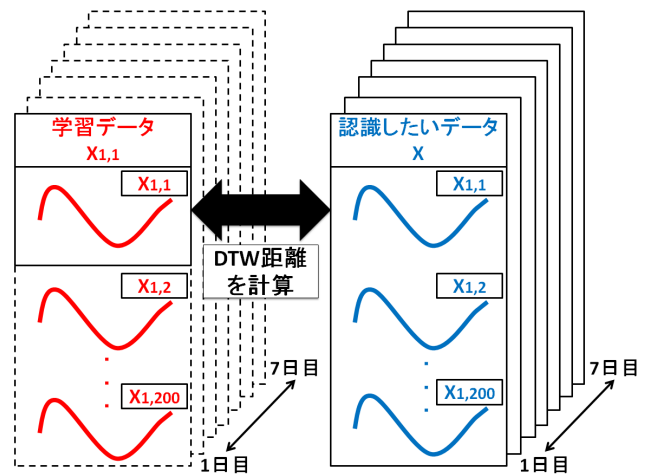


図 1 DTW 距離の計算方法

学習データとの距離が大きく、行われたジェスチャは学習データと似ていないことを意味している。また、グラフの上下の振れ幅が大きいほど、その周辺の学習データのジェスチャが安定していないことを意味している。

図 2 から、学習データとして使用するデータのサンプルの位置が後半になるにつれて、DTW 距離が小さくなっていくことがわかる。これはジェスチャ動作が実験の序盤と終盤で変化しており、開始からの数回のデータを学習データとする従来の手法では試行回数を重ねるにつれて認識するデータとの距離が大きくなることを示している。その結果、他のジェスチャとの差異が相対的に小さくなるため誤認識が起りやすくなると考えられる。また、学習データとして使用するデータのサンプルの位置が後半になるにつれて、DTW 距離が収束していることがわかる。これは被験者がジェスチャを繰り返すうちにジェスチャに慣れてフォームが安定した、もしくは疲労によりフォームが体が疲れにくいものに変化し安定したと考えられる。他の被験者および「星を描く」ジェスチャにおいても同様の結果が見られた。

このことから、ジェスチャの初めの数回を学習データとして用いる従来手法では誤認識を起こす可能性が明らかとなった。そこで「ボールを投げる」および「星を描く」ジェスチャ以外に「円を描く」、「パンチする」、「縦に線を描く」など 8 種類のジェスチャを各被験者から採取し、それぞれ 10 回分のデータ（「ボールを投げる」および「星を描く」ジェスチャは 1 日目の初めの 10 回分のデータ）を学習データとして、1 日目~7 日目のデータを DTW を用いて認識精度を調べた。また「ボールを投げる」および「星を描く」ジェスチャにおいては、図 3 のように、1 日目の 200 回目のデータを学習データとしたときの DTW 距離の平均値を終値として、終値 +10% の値を十分小さく安定している値と考え、その値を初めて下回る点を収束点と定義し、収束点におけるジェスチャを試行回数および DTW の観点から理想的なジェスチャとして、そのジェスチャを学

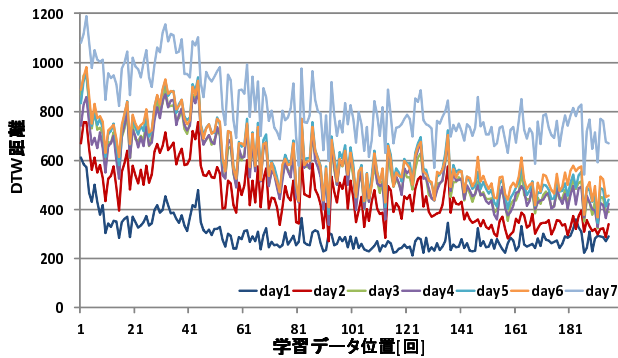


図 2 被験者 1 の「投げる」ジェスチャの結果

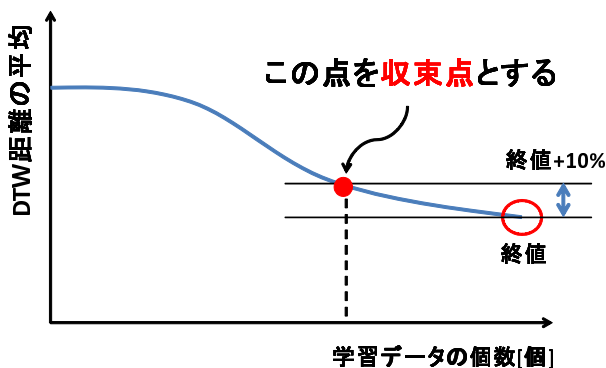


図 3 収束点の位置

表 1 7 日間の最低認識精度 [%]

ジェスチャ	投げる		星を描く	
	従来	理想	従来	理想
被験者 1	100	100	100	100
2	98.5	99.5	61.5	91.5
3	100	100	1.50	87.0
4	100	100	93.5	100
5	100	100	99.5	100

習データとしたときの認識精度を調べた。その結果から、各被験者ごとに 7 日間のうち最も低くなった認識精度を表 1 に示す。「ボールを投げる」および「星を描く」ジェスチャで 1 日目の初めの 10 回分のデータを学習データとして用いた場合の結果を「従来」の列に、収束点までの理想的な学習データを用いた場合の結果を「理想」列に示す。表 1 より、「ボールを投げる」ジェスチャでは、学習データとして 1 日目の初めの 10 回分のデータを用いた場合および 1 回目から収束点までを用いた場合、ともに 98.5% 以上の高い認識精度を達成した。一方、「星を描く」ジェスチャにおいては初めの 10 回分のデータを学習データとした場合、被験者 3 は認識精度が 1.50% まで低下した。収束点までを学習データとした場合は 87.0% 以上の認識精度を達成しており、従来手法のように初めの数回を学習データとしたすると誤認識を起し、適切な位置までのデータを学習データとすることで認識精度を改善できることがわかった。

4.2 学習データ探索手法

予備実験から、適切な学習データを用いることで高い認識精度を達成できることがわかったが、収束点は 200 回分のデータを用いて得られた結果であり、実際にはジェスチャデータの収集中に収束点を判定する必要がある。そこで、リアルタイムで適切な学習データを探索する二つの手法を提案した。

4.2.1 手法 1

予備実験の結果より学習データとして適切な点は、他のデータとの DTW 距離および、その振れ幅が小さくなることが分かった。ジェスチャが安定するとそれ以降のデータの変化が少なくなることから、安定した点以降のデータを学習データとして採取する必要はないと考えられる。そこで、手法 1 はジェスチャを n 回行った時点で $n - T$ 回目のデータを学習データとして、 $n - T + 1$ 回目から n 回目までのデータとの DTW 距離の平均

$$\mu_n = \frac{\sum_{i=0}^{T-1} DTW(x_{n-T}, x_{n-i})}{T},$$

および分散

$$\sigma_n^2 = \frac{\sum_{i=0}^{T-1} \{DTW(x_{n-T}, x_{n-i}) - \mu_n\}^2}{T}$$

を計算する。 μ_n と σ_n の値が以下のすべての条件を満たしたとき、その点を収束点と判断し、学習データの採取を終了する。

- $\mu_n < \mu_T$
- $\sigma_n < \sigma_T$,
- $\mu_n < \mu_{n-i} \times (1 + \alpha)$, for $i = 1, \dots, l$
- $\sigma_n < \sigma_{n-i} \times (1 + \alpha)$, for $i = 1, \dots, l$

予備実験で収集したデータより $\alpha = 0.01$, $T = 10$, $l = 9$ とした。平均および分散の計算に 10 回分のデータを必要とし、直近 10 個の平均および分散の値から収束の判定を行うため、手法 1 では最低 20 回のジェスチャを行う必要がある。

4.2.2 手法 2

学習データと変化していくデータとの距離を小さくするためには、同じジェスチャでもさまざまなパターンのジェスチャデータを学習しておくことが効果的である。予備実験からジェスチャを繰り返し行うことにより、ジェスチャのフォームが次第に安定していくことが分かっている。すなわち、ジェスチャが安定するまでのデータはさまざまなパターンのデータを含んでいると考えられる。そこで、学習データを採取する過程で類似するジェスチャデータが連続して現れたとき、それ以降ジェスチャの変化が少なくなると仮定し、その時のデータを学習データとする。具体的には、 n 回目のジェスチャを試行した時点で、1 回目から T' 回目までのデータを学習データとして n 回目のジェスチャデータとの DTW 距離を計算し、その平均値

$$\mu'_n = \frac{\sum_{i=0}^{T'-1} DTW(x_i, x_n)}{T'}$$

を得る。 μ'_n が以下の条件を満たすとき、1回目から n 回目までのデータを学習データとする。

- $|\mu'_n - \mu'_{n-i}| < \mu_{n-i} \times (1 + \beta)$, for $i = 1, \dots, l'$

予備実験で収集したデータより $\beta = 0.1$, $T' = 5$, $l' = 4$ とした。平均の計算に5回分のデータを必要とし、直近5個の平均の値から収束の判定を行うため、手法2では最低10回のジェスチャを行う必要がある。

4.3 評価

予備実験で得られた各被験者の1日目のデータから、手法1および手法2を用いて学習データの位置を探索し、その点までに採取したすべてのデータを学習データとして2日目から7日目のデータとのDTW距離を計算した。比較手法として、各被験者の1日目の1回目から5回目までのデータを学習データとする従来手法も評価した。「ボールを投げる」ジェスチャの結果を表2に示す。()内の数値は各手法の学習データ採取回数である。表2から分かるように、手法1および手法2を用いることで、従来手法と比較してDTW距離を小さくすることができた。しかし、手法1および手法2を用いた場合、従来手法と比較して学習データのサンプル数が極端に多くなる場合もあった。そこで、本研究では先行研究で提案した手法を用いて学習データを採取する際、被験者に視聴覚フィードバックを与えることで、ジェスチャ動作を早期に収束させ、学習データ採取までに要する試行回数の削減を試みる。

5. 視聴覚フィードバックによる影響調査

先行研究の結果から、提案した手法を用いることで最初の数回を学習データとする従来手法より、以降に行われるジェスチャに近い学習データを採取できることを確認した。しかし、提案した手法で学習データを採取すると、試行するジェスチャの回数が多くなることも明らかとなった。ジェスチャの試行回数を削減するためには、少ない試行回数でジェスチャ動作が安定し、提案した手法の条件を満足する必要がある。そこで、本節では手法1を用いて学習データを採取する際、被験者に視聴覚フィードバックを与えることで、ジェスチャ動作を早期に収束させ、学習データ採取までに要する試行回数の削減を試みる。

5.1 実験環境

3.1節で述べた予備実験と同様に、手首に3軸加速度センサを装着した21~25歳の被験者10名(男性8名、女性2名)から、「パンチする」ジェスチャ、「アッパーカットする」ジェスチャ、「チョップする」ジェスチャおよび「ビンタする」ジェスチャのデータを採取した。「パンチする」ジェスチャおよび「アッパーカットする」ジェスチャ

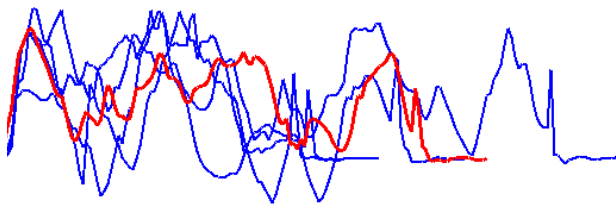
表2 「投げる」ジェスチャの結果

被験者		手法1	手法2	従来手法
1	採取回数	(48)	(20)	(5)
	2日目	395	395	560
	3日目	557	557	730
	4日目	525	525	662
	5日目	579	579	738
	6日目	608	608	747
	7日目	817	817	974
2	採取回数	(41)	(15)	(5)
	2日目	750	790	839
	3日目	903	916	997
	4日目	741	768	821
	5日目	754	768	834
	6日目	746	748	832
	7日目	760	760	893
3	採取回数	(123)	(165)	(5)
	2日目	285	282	387
	3日目	331	324	444
	4日目	335	330	448
	5日目	362	358	431
	6日目	367	357	480
	7日目	400	391	482
4	採取回数	(96)	(21)	(5)
	2日目	427	500	658
	3日目	474	541	699
	4日目	439	495	628
	5日目	475	531	611
	6日目	452	531	676
	7日目	428	491	603
5	採取回数	(58)	(13)	(5)
	2日目	520	587	615
	3日目	496	584	618
	4日目	538	595	673
	5日目	524	646	695
	6日目	564	655	702
	7日目	569	706	736

では、聴覚フィードバックの有無による影響を、「チョップする」ジェスチャおよび「ビンタする」ジェスチャでは、視覚フィードバックによる影響を調査する。加速度センサはワイヤレステクノロジー社製 WAA-006[9]を使用し、サンプリング周波数は50Hzでデータを取得した。

5.2 クイズの正解音を用いた聴覚フィードバック

被験者10名のうち、5名の被験者からは「パンチする」ジェスチャと「アッパーカットする」ジェスチャのデータを手法1のフィードバックなしで、残りの5名の被験者からは手法1のフィードバックありでデータを採取した。3.1節より、同じジェスチャを繰り返し行うことによってジェスチャが安定し、実験に影響を及ぼすことを防ぐため、同じジェスチャでの実験は一度とする。また被験者の個人差による影響を調査するために、フィードバックのある場



- ・n回目のジェスチャデータ
- ・n-i 回目のジェスチャデータ

図 4 視覚フィードバックの方法

合とない場合の実験を一度ずつ行ってもらった。フィードバックは n 回目のジェスチャ終了時に、 $n-10$ 回目のジェスチャと $n-i$ 回 ($0 \leq i \leq 9$) のジェスチャとの DTW 距離の平均

$$\mu_n'' = \frac{\sum_{i=0}^9 DTW(x_{n-10}, x_{n-i})}{10},$$

と分散

$$\sigma_n''^2 = \frac{\sum_{i=0}^9 \{DTW(x_{n-10}, x_{n-i}) - \mu_n''\}^2}{10}$$

を計算して、 μ_n'' と σ_n'' が以下の条件を満たすたびにクイズの正解のような音を再生した。

- $\mu_n'' < \mu_{n-1}'' \times (1 + \gamma)$
- $\sigma_n'' < \sigma_{n-1}'' \times (1 + \gamma)$

今回の実験では $\gamma = 0.1$ とした。 n 回目のジェスチャ終了時に音が再生された場合、 n 回目のジェスチャのデータが $n-1$ 回目のデータと近似している旨を事前に被験者に伝えて、聴覚フィードバックが採取されるサンプル数に与える影響を調査した。ただし、手法 1 の収束条件と今回の音が発生する条件は異なるため、音が再生されない場合でも手法 1 の収束条件を満足して実験が終了する場合もある。

5.3 ジェスチャの波形表示による視覚フィードバック

聴覚フィードバックの実験と同様に、10名の被験者のうち、5名の被験者からは「チョップする」ジェスチャと「ピンタする」ジェスチャのデータを手法 1 のフィードバックなしで、残りの 5名の被験者からは手法 1 のフィードバックありでデータを採取した。採取方法も聴覚フィードバックの実験と同様に、同じジェスチャでの実験は一度とし、また被験者の個人差による影響を調査するために、フィードバックのある場合とない場合の実験を一度ずつ行ってもらった。視覚フィードバックは図 4 に示すように、 n 回目のジェスチャ終了時に、 n 回目のジェスチャデータの x 軸、 y 軸、 z 軸の 3 軸の合成波形を赤色、 $n-i$ 回目 ($1 \leq i \leq 4$) のジェスチャデータの合成波形を青色で表示した。ジェスチャ回数が 5 回未満の時は、 n 回目のジェスチャデータの合成波形を赤色、それ以外のジェスチャデータの合成波形を青色で表示した。

5.4 結果と考察

フィードバックありとなしのときに手法 1 が収束点と判断するまでに要したジェスチャ試行回数をジェスチャごとに表 3, 表 4, 表 5 および表 6 に示す。表 3 および表 4 は聴覚フィードバックを用いた「パンチする」ジェスチャおよび「アッパーカットする」ジェスチャの結果、表 5 および表 6 は視覚フィードバックを用いた「チョップする」ジェスチャおよび「ピンタする」ジェスチャの結果を示している。表中の「-」は、ジェスチャを 200 回行った時点で手法 1 の条件を満たすことができず、データの採取が終了しなかった場合を示している。

5.4.1 聴覚フィードバックに関する考察

まず聴覚フィードバックに関する考察を行う。ジェスチャごとに結果を見ると、表 3 より、「パンチする」ジェスチャでは、全体の平均が 43.3 回、フィードバックがある場合では平均 34.4 回、フィードバックがない場合では平均 52.2 回でデータの採取が終了しており、フィードバックがある場合にジェスチャが収束するまでに要する試行回数が減少している。一方、表 4 より、「アッパーカットする」ジェスチャでは、200 回で終了しなかった被験者を除くと、全体の平均が 42.9 回、フィードバックある場合では平均 45.8 回、フィードバックがない場合では平均 40.0 回となり、フィードバックの有無による大きな差異は見られなかった。このことから、「パンチする」ジェスチャではフィードバックが有効であり、「アッパーカットする」ジェスチャではフィードバックの効果は小さいと考えられるが、被験者 6～被験者 10 は被験者 1～被験者 5 と比較してジェスチャが収束が遅く、フィードバックありの場合でもフィードバックなしの被験者 1～被験者 5 より採取サンプル数が大きくなってしまった可能性も考えられる。

また、被験者ごとのフィードバックの有無に注目すると、被験者 10 名のうち 7 名において、フィードバックがある場合はない場合と比較して採取するサンプル数が減少している。その中でも、被験者 4, 被験者 5 および被験者 6 はフィードバックがない場合は、そのジェスチャ全体の平均採取回数より採取回数が大きかったが、フィードバックがある場合には、そのジェスチャ全体の平均採取回数より小さくなり、フィードバックが有効であったと考えられる。

さらに、被験者 3 および被験者 8 が 200 回ジェスチャを行っても採取が終了しなかった原因としては、ジェスチャデータがジェスチャ開始当初から安定していたため、繰り返しジェスチャを行っても手法 1 の条件を満たすことができなかったと考えられる。

加えて、被験者 1 および被験者 9 はフィードバックありの場合に採取するサンプル数が大きくなっている。これはフィードバックありで採取を行った場合、音が再生されなかったときに音が再生されるようにジェスチャの軌道を意識的に変化させており、これによりジェスチャが安定する

までに回数が増えてしまったと考えられる。

5.4.2 視覚フィードバックに関する考察

次に視覚フィードバックに関する考察を行う。ジェスチャごとに結果を見ると、表5より、「チョップする」ジェスチャでは、全体の平均が50.5回、フィードバックがある場合では平均72.8回、フィードバックがない場合では平均28.2回でデータの採取が終了しており、フィードバックがある場合にジェスチャが収束するまでに要する試行回数が増加している。表6より、「ビンタする」ジェスチャでは、全体の平均が42.2回、フィードバックがある場合では平均51.4回、フィードバックがない場合では平均33.0回でデータの採取が終了しており、「ビンタをする」ジェスチャにおいてもフィードバックがある場合にジェスチャが収束するまでに要する試行回数が増加している。このことから、どちらのジェスチャに関してもフィードバックありの場合の試行回数が大きくなる結果となり、波形を表示する視覚フィードバックを行うことでジェスチャの収束が遅くなってしまったと考えられる。

また、被験者ごとのフィードバックの有無に注目すると、被験者10名のうち9名において、フィードバックがある場合はない場合と比較して採取するサンプル数が増加している。その中でも、被験者1、被験者2、被験者3、被験者5および被験者7はフィードバックがない場合は、そのジェスチャ全体の平均採取回数より採取回数が小さかったが、フィードバックがある場合には、そのジェスチャ全体の平均採取回数より大きくなり、フィードバックによりジェスチャの収束が遅くなったといえる。ほとんどの被験者において、フィードバックありの場合にジェスチャが収束するまでに要する試行回数が増えた原因として、ジェスチャの波形が過去4回の波形と差異が現れた場合に、その差異を小さくしようとジェスチャ軌道を意識的に変化させることでジェスチャが安定するまでに回数が増えてしまったと考えられる。特に、フィードバックありの場合で試行回数が多くなった被験者2、被験者3、被験者5および被験者7にその傾向が強く現れたと考えられる。図5に被験者2、被験者3、被験者5および被験者7の視覚フィードバックありのときのDTW距離の平均のグラフを示す。縦軸がDTW距離の平均、横軸が行ったジェスチャの試行回数を表す。この図からもわかるように、ジェスチャを繰り返し行ってもDTW距離が初めの値から小さくならず、振幅も大きくなっており、ジェスチャの安定までに時間がかかったことがわかる。

さらに、フィードバックありの場合に試行回数が少なくなった被験者6やフィードバックの有無によって試行回数の変化があまりなかった被験者1、被験者4、被験者7、被験者8、被験者9および被験者10に関しては、データ採取中に波形の大きな差異がみられなかったり、差異が出てもその後差異がすぐに小さくなったため、フィードバックの

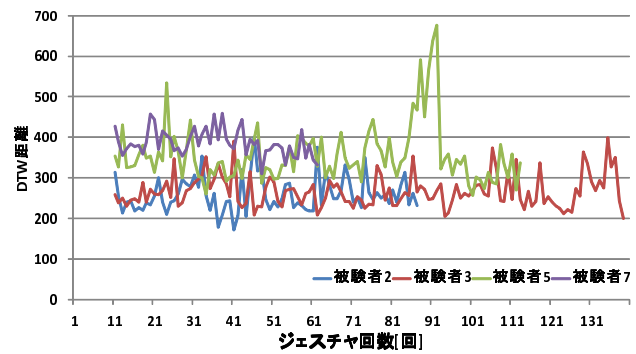


図5 視覚フィードバックの手法において結果が悪かった被験者のDTW距離の平均の変化

表3 「パンチする」ジェスチャの結果

ジェスチャ	フィードバック	被験者	結果 [回]
パンチ	有	1	27
		2	30
		3	55
		4	39
		5	21
	無	6	47
		7	83
		8	61
		9	32
		10	38

影響が少なかったと考えられる。

5.4.3 調査結果のまとめ

ジェスチャの収束には聴覚フィードバックが有効であることがわかった。逆に、ジェスチャの波形を表示する視覚フィードバックはジェスチャの収束を遅らせることがわかった。また、被験者によってフィードバックの有効性に差が出る結果となりさらなる調査が必要である。今回は、聴覚フィードバックとして音を用い、視覚フィードバックとしてジェスチャ波形を表示する方法を用いたが、他の聴覚なフィードバックや視覚フィードバック、触覚フィードバックの利用、またはフィードバックの組合せやフィードバックを行う条件の調整によりさらに有効なフィードバック方法を提案できると考える。

6. おわりに

本研究では、先行研究で提案した学習データ採取法において採取するサンプル数を削減するための手法として、クイズの正解音を用いた聴覚フィードバックおよびジェスチャ波形を表示する視覚なフィードバックを提案し、その影響を評価した。実験から、一部の被験者において聴覚フィードバックの有効性が見られた。また、視覚フィードバックを行うことで、フィードバックを行わない場合と比較してジェスチャの収束が遅くなることがわかった。

今後は他の視聴覚フィードバックや触覚なフィードバック

表 4 「アッパーカットする」ジェスチャの結果

ジェスチャ	フィードバック	被験者	結果 [回]
アッパー カット	有	6	27
		7	72
		8	-
		9	48
		10	36
	無	1	22
		2	40
		3	-
		4	44
		5	54

表 5 「チョップする」ジェスチャの結果

ジェスチャ	フィードバック	被験者	結果 [回]
チョップ	有	2	87
		3	139
		6	32
		7	62
		8	44
	無	1	38
		4	21
		5	23
		9	38
		10	21

表 6 「ピンタする」ジェスチャの結果

ジェスチャ	フィードバック	被験者	結果 [回]
ピンタ	有	1	47
		4	34
		5	112
		9	40
		10	24
	無	2	40
		3	28
		6	37
		7	29
		8	31

クによる影響の調査を行う。また、評価するジェスチャ数の増加、フィードバックを行う条件の調整やさまざまなフィードバックを組み合わせることで、被験者やジェスチャの違いによる影響を受けないフィードバック方法を構築していく。さらに、新たなフィードバック方法を用いて先行研究で提案した学習データ採取法と組み合わせることで、少ないサンプル数でより手軽で高精度な学習データ採取法を提案する。

謝辞 本研究の一部は、科学技術振興機構戦略的創造研究推進事業(さきがけ)、文部科学省科学研究費補助金基盤研究(A)(20240009)および若手研究(B)(24700066)によるものである。ここに記して謝意を表す。

参考文献

- [1] 吉田 楽, 村尾和哉, 寺田 努, 塚本昌彦: 加速度センサを用いたジェスチャ認識に疲労および忘却が与える影響の評価, 情報処理学会研究報告, Vol. 2012-UBI-33, No. 19, pp. 1-8 (Mar. 2012).
- [2] K. Murao, T. Terada, A. Yano, and R. Matsukura: Evaluating Gesture Recognition by Multiple-Sensor-Containing Mobile Devices, *Proc. of the 15th IEEE International Symposium on Wearable Computers (ISWC 2011)*, pp. 55-58 (Oct. 2011).
- [3] S. Agrawal, I. Constandache, S. Gaonkar, R. Choudhury, K. Caves, and F. DeRuyter: Using Mobile Phones to Write in Air, *Proc. of the 9th International Conference on Mobile Systems, Applications, and Services (MobiSys 2011)*, pp. 15-28 (June/July 2011).
- [4] J. Liu, Z. Wang, L. Zhong, J. Wiekramasuriya, and V. Vasudevan: uWave: Accelerometer Based Personalized Gesture Recognition and Its Applications, *Proc. of the 7th IEEE Pervasive Computing and Communication (PerCom 2009)*, pp. 1-9 (June 2009).
- [5] C. Myers and L. R. Rabiner: A Comparative Study of Several Dynamic Time-warping Algorithms for Connected Word Recognition, *The Bell System Technical Journal*, Vol. 60, pp. 1389-1409 (Oct. 1981).
- [6] 金ジョンヒョン, 橋田朋子, 大谷智子, 苗村 健: 筆記音のフィードバックが筆記作業に与える影響について, インタラクシオン 2012, 1EXB-52, No. 19, pp. 445-450 (Mar. 2012).
- [7] S. Brewster, J. Lumsden, M. Bell, M. Hall, and S. Tasker: Multimodal 'Eyes-Free' Interaction Techniques for Wearable Devices, *In Proc. of the 21st Conference on Human Factor in Computing Systems (CHI 2003)*, pp.473-480 (Apr. 2003).
- [8] T. H. Andersen and S. Zhai: "Writing with Music": Exploring the Use of Auditory Feedback in Gesture Interfaces, *ACM Transactions on Applied Perception (TAP)*, Vol. 7, No. 3, pp. 1-24 (June 2010).
- [9] Wireless Technologies, Inc.: <http://www.wireless-t.jp/>.