

# 手話動画認識における部位ごとの適正動作判定に関する一考察

## A Consideration on the Discrimination of Appropriate Motion by each Part of the Body in Sign Language Recognition.

笹島 和哉<sup>†</sup> 亀田 凱聖<sup>†</sup> 川口 大夢<sup>†</sup>

梅田 芳護<sup>‡</sup> 田中 博<sup>†</sup> 西村 広光<sup>†</sup>

Kazuya Sasajima<sup>†</sup>, Kaisei Kameda<sup>†</sup>, Hiromu Kawaguchi<sup>†</sup>,  
Yoshimori Umeda<sup>‡</sup>, Hiroshi Tanaka<sup>†</sup>, Hiromitsu Nishimura<sup>†</sup>

<sup>†</sup> 神奈川工科大学 情報学部

<sup>†</sup> Kanagawa Institute of Technology, Faculty of Information and Computer Sciences

<sup>‡</sup> 神奈川工科大学大学院 情報工学専攻

<sup>‡</sup> Graduate School of Kanagawa Institute of Technology, Course of Information and Computer Sciences

E-mail: {s2123026, s2121104, s2121022, s2385004}@cco.kanagawa-it.ac.jp,  
{h\_tanaka, nisimura}@ic.kanagawa-it.ac.jp

### 1. はじめに

画像から骨格情報推定を行い、手話動作を認識する技術開発が進められている[1]。手話認識の技術を身近に応用できる用例を考え、我々は、手話学習者を対象にした学習支援システムについて検討を進めている[2]。具体的には、手話学習者がカメラ前で行った手話単語動作を、手話認識用に構築した識別モデルを利用し尤度を求め、その尤度を手話動作の得点として表示させることにより、手話の自己学習を支援するシステムの試作を進めている。これまでの手話学習支援ツールでは、手話動作全体の認識尤度から動作得点を算出して学習者へのフィードバックとしてきた。しかし、手話動作が正しくない場合に、手話動作のどの部分が正しくないのかを学習者にフィードバックすることができないという課題があった。

この課題に対して本報告では、右腕や左腕といった特定部位の骨格移動情報を固定値にして情報欠損させたデータによって学習させた部位情報欠損識別モデルを複数作成し、それらモデルの尤度分布を利用して正しくない手話動作の部位を推定する技術を提案する。

提案手法により、正しくない手話動作の原因がどこにあるのかをシステム出力できるようにし、その結果を複数被験者で評価・検討した。

### 2. 既存手話学習支援システムの処理内容

手話学習支援システムは下記のような処理を行い、カメラに対して行った動作に対する得点表示を行うことで、学習者へのフィードバックを実現している。

#### 【システム構築時の処理の流れ】

1. 10種の手話単語、4人の被験者、10回の動作情報を機械学習用に収集
2. MediaPipeで骨格移動情報に変換し、両肩と両手首の時系列座標に変換  
※両肩の座標を使い体格差の基準化も実施
3. SVMで機械学習モデル構築

#### 【学習支援時の処理の流れ】

1. カメラから手話動画取得  
※開始、終了は指定姿勢での停止で実現
2. MediaPipeでシステム構築時と同様に、骨格移動情報に変換
3. 学習済みモデルに入力し、尤度算出
4. 尤度から10種手話の総計100点になるように得点化し、得点上位5つを学習者へのフィードバックとして結果表示

### 2.1 対象手話単語

本稿で対象とする手話単語は、日本手話を対象とし、株式会社ケイ・シー・シー様の手話学習者向けの動画辞典アプリ Smart Deaf から、参照数が多い健康医療関連単語かつ両手で動作を行う単語を10単語選定した。対象とした単語を表1に示す。

表1.対象手話単語

1.顔色	2.発達障害	3.ダイエット
4.健康	5.イライラ	6.近視
7.肥満	8.パニック	9.たくましい
10.涙		

### 3. 提案する学習支援の手法

従来の手話学習支援システムにおいて、手話全体の得点で学習者へのフィードバックは行えているものの、得点が低い原因が特定されず、どの部位を修正すべきかについては学習者自身が判断する必要があった。

本報告ではこの課題を改善するため、どの部位の動作がおかしいのかを特定する技術について検討を行った。本報告の提案手法は、下記の2段階の手法である。

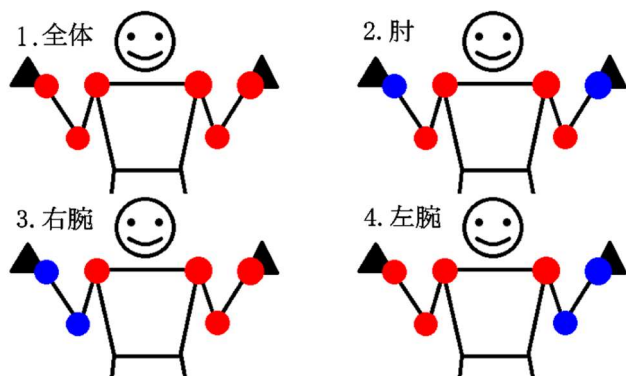
1. 部位情報欠損識別モデルを利用し尤度を求める
2. 求めた尤度による得点と分散の関係を表した散布図を用いた手話動作評価

#### 3.1 特定部位座標の動作情報削除方法

特定部位の手話動作判定を行うため、全体の手話動作から骨格推定を行い、その後取得した特定骨格点情報を固定値(0,0)にすることで、その部位の動作情報を欠損させることとした。特定部位の情報欠損させたデータを作成し、学習・判定に利用することとした。今回、特定部位の動作情報欠損処理としては、該当骨格推定部位の座標を(0,0)とすることとした。

今回は動作情報を削除するパターンとして、図1に示すように4種のパターンを考え、それぞれ識別モデルを構築することとした。図1において、赤色は特徴量計算に使用する部位の骨格情報、青色は座標情報を(0,0)にした部位の骨格情報である。以下に4種のパターンを示す。

1. 図1左上：両肘・手首の4点の座標情報を全て動作情報とした従来の手話認識と同じ「全体動作判定モデル」
2. 図1右上：両手首の座標情報を(0,0)とした「肘動作判定モデル」
3. 図1左下：画面に向かって左腕の肘と手首の座標情報を(0,0)とした「右腕動作判定モデル」
4. 図1右下：画面に向かって右腕の肘と手首の座標情報を(0,0)とした「左腕動作判定モデル」



○手話動作評価項目 (赤：特徴量計算座標, 青：固定値(0,0)部位座標)

図1. 特徴量計算座標、固定値(0,0)座標部位関係

#### 3.2 判定尤度の分散を考慮した類似判定法

特定部位の手話動作識別に関して、異なる手話がほぼ同一の動きとして判定される可能性を考慮し、求めた尤度から分散を求めた。分散が大きい場合、類似した動作の単語が少なく、得点が低い=誤った動作。逆に分散が小さい場合、類似した動作の単語が多く、得点が低い≠誤った動作として、正しい動作をした可能性があると考えた。横軸を分散、縦軸を評価スコアとした散布図を作成し、手話動作の類似性を考慮した上で手話動作評価を行うこととした。

#### 3.3 散布図を用いた手話動作評価

分散は識別する10単語中における手話動作の類似度を表す。このことから、識別する10単語において特定部位の手話動作が全く同じである場合、求める分散が最大値かつ10単語の得点は均一に10点となり、識別する10単語における特定部位の手話動作が全く異なる場合、求める分散が最小値かつ得点は手話学習者が行った単語が100点、以外が0点になると仮定できる。この分散と得点の関係が1次関数の関係であるという仮定のもと、 $x$ を分散、 $y$ を得点とし、最大値( $x,y$ )=(900,100)、最小値( $x,y$ )=(0,10)の2点を通る直線として

$$\text{直線式: } y = \frac{1}{10}x + 10 \quad (1)$$

の関係が見いだせるとして、結果判定閾値条件に利用するため、図2以降の青色の直線を入れて結果を示すこととする。そして、学習データの分布から求められる近似線と(1)式の直線の2つの直線を基準に、3段階で手話動作を評価した。図2に得点と分散の関係、近似線と(1)式の直線を表した散布図例を示し、表2に結果判定閾値条件と3段階評価項目を示すこととした。

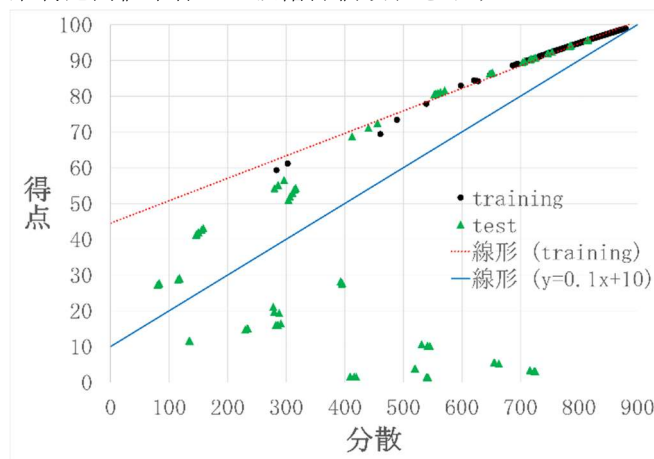


図2. 得点と分散の関係と学習データ分布の近似線、(1)式の直線を表した散布図の一例

図2において、黒色のデータポイントは、学習データ分布、緑色のデータポイントは、テストデータ分布、

赤色の点線は、学習データ分布から求められる近似線、青色の直線は、(1)式の直線である。

表 2.結果判定閾値条件と 3 段階評価項目

データポイントの位置	評価
近似線以上	正しい動作
近似線と(1)式の直線の間	正誤判断が難しい動作
直線より下	誤っている動作

## 4. 実験評価・結果

手話動作の取得に関して、筆者らは Web カメラでの撮影を行った。撮影対象は手話初学者 1 人、過去に常時手話を使用している方の指導のもとで撮影した学習データ用の手話動画を参考に、手話動作を行った。手話者は立った状態で、10 単語 3 回の手話動作を行い、動画を取得した。撮影した動画は手動で 1 単語ずつに切り分けて評価実験を行った。

### 4.1 識別結果による平均点

MediaPipe を用いて得られる節点座標に関して、読み込む動画ファイル数や順序によって微細な差が発生するため、動作評価に公平性を持たせるため同一の動画ファイルを 3 つに増加し、ランダムな順序で識別処理を行った。表 3 に単語、部位ごとにおける得点の平均値を示す。得点の平均値は小数点第 2 位で四捨五入している。表 3 において、赤の部分は何が 70% 以上、青の部分は得点が 20% 以下である。

表 3.得点の平均値

単語・部位	全体	右腕	左腕	肘
顔色	14.2	0.6	31.4	2.2
発達障害	2.3	0.7	0.1	0.5
ダイエット	63.1	76.1	67.0	11.6
健康	37.7	1.5	1.5	2.4
イライラ	90.0	99.9	97.3	99.9
近視	6.3	51.7	0.2	38.7
肥満	44.6	74.0	83.7	7.5
パニック	95.6	47.9	56.3	75.4
たくましい	91.0	97.2	96.3	27.7
涙	48.7	47.7	3.2	26.6

### 4.2 3 段階評価で手話動作を評価した結果

散布図を用いた手話動作の 3 段階評価を適用するため、部位ごとを取得したテストデータの識別結果から、手話動作を行った手話者が意図した単語の得点を縦軸、識別を行い求めた尤度の分散を横軸とし、散布図を作成した。図 3.1, 3.2, 3.3, 3.4 にそれぞれ全体、右腕、左腕、肘の得点と分散の関係、近似線と(1)式の直線を記載した散布図を示す。図 3.1, 3.2, 3.3, 3.4 において、黒色のデータポイントは、学習データ分布、緑色のデータポイントは、テストデータ分布、赤色の点線

は、学習データ分布から求められる近似線、青色の直線は、(1)式の直線、青色の領域は、3 段階評価の「正しい動作」、黄色の領域は、3 段階評価の「正誤判断が難しい動作」、赤色の領域は、3 段階評価の「誤っている動作」とした。

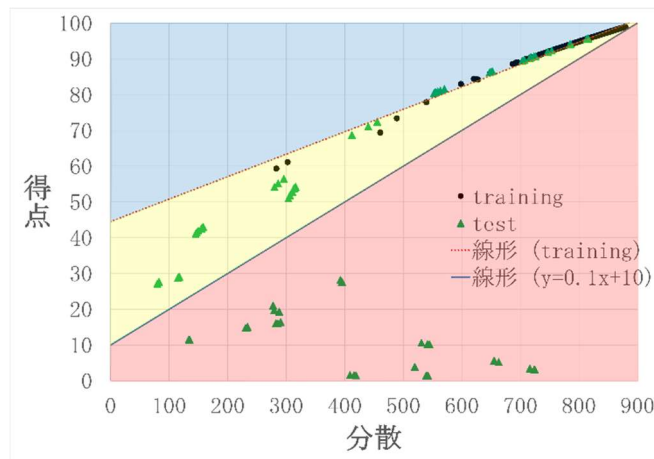


図 3.1.全体動作判定モデルにおける 10 クラス得点分散と得点の分布

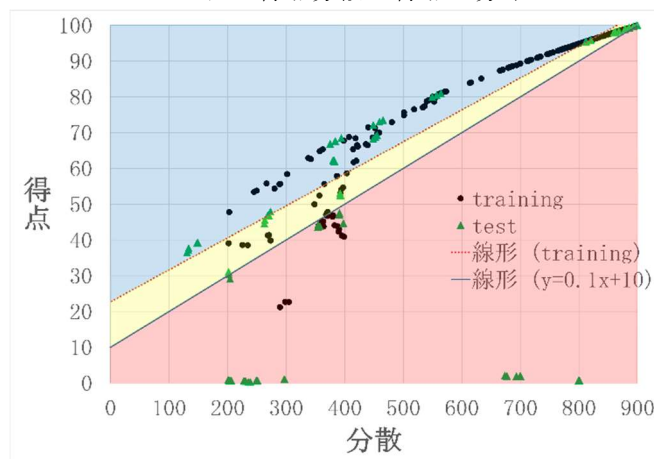


図 3.2.右腕動作判定モデルにおける 10 クラス得点分散と得点の分布

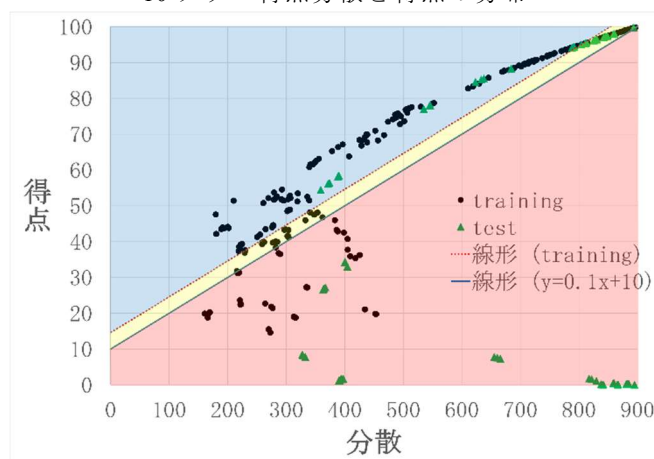


図 3.3.左腕動作判定モデルにおける 10 クラス得点分散と得点の分布

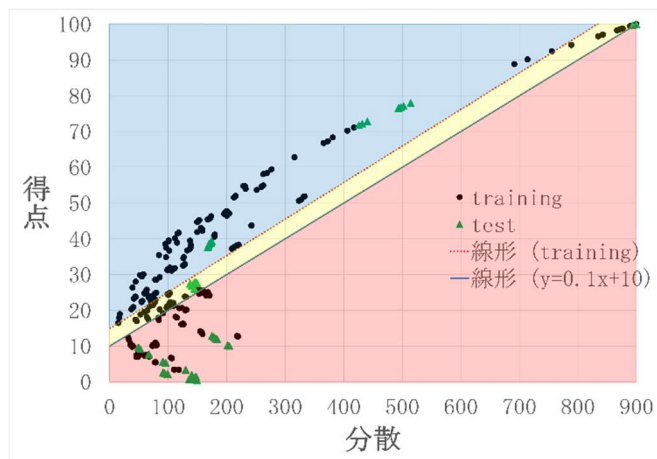


図 3.4.肘動作判定モデルにおける  
10 クラス得点分散と得点の分布

表 2 の評価を単語、部位ごとに適用した結果、全体動作判定モデルでの評価において、「正誤判断が難しい動作」や「誤っている動作」が該当する単語は、顔色や発達障害、健康、近視、肥満、涙となった。これらの単語の特定部位の動作判定モデルでの評価に関して、顔色や発達障害、健康は右腕、左腕、肘のどの項目でも「誤っている動作」に該当したため、全体的に手話動作の見直しが必要と判断した。それ以外の単語、近視や肥満、涙については、特定部位の動作判定モデルでの評価で「正しい動作」、「正誤判断が難しい動作」、「誤っている動作」それぞれ評価が異なり、手話動作を見直すべき部位の特定に繋げることができた。

### 4.3 システム評価と人の評価比較

上記のシステム評価に対して、人による手話動作の評価では差異があるのか検討するため、複数人で今回取得した手話動作を評価した。本報告の目的である「誤った手話動作に対して、どの部位が悪いのかを判定すること」に合わせ、システムによる全体動作判定モデルでの手話動作評価が「誤っている動作」と評価された単語の手話動作のみを選定した。さらに、手話動作の評価者に対して、「誤っている動作」と評価された部位を伝えた上で、手話動作を観察させ、特定部位の動作判定モデルでのシステム評価について、「そう思う」、「そう思わない」、「わからない」の中から回答させた。今回、手話初学者 4 人から全体動作判定モデルでのシステム評価が「誤っている動作」になった 11 回の動作について、システム評価の不適当な動作部位の特定が適切に行われているかヒアリング調査を行った。表 4 に全 44 回のヒアリング回答の割合を示す。

表 4.システム評価妥当性評価回答の割合

そう思う	そう思わない	わからない
18/44	24/44	2/44

回答を集計した結果、4 割で提案手法による不適切な手話が妥当であるという評価を得た。しかし、「そう思わない」の回答割合が最も高かった。この結果の要因について、システム評価において手話動作が「誤っている動作」と評価された単語や部位ごとに、どの単語に誤認識されたのかを確認、手話動作を比較したところ、本報告における識別手法では左右の肘、手首座標のみで手話動作を識別しているため、視覚的には異なる手話動作に見えても、座標の変化具合では類似していることが挙げられた。

### 5. まとめ

本報告では、特定部位の骨格移動情報を欠損させ、学習させた認識モデルを複数作成することによって部位における手話動作の認識尤度で評価スコアを求め、評価スコアと分散の関係から手話動作を評価する手法を提案し、有効性を確認した。実験から部位ごとの認識尤度による評価スコアと分散の関係を表す散布図において、分割した 3 領域に基づいた手話動作評価から誤った手話動作の 4 割に対して適切な誤り部位の判定ができた。同時に、左右の肘、手首座標だけでの識別手法では、識別が不十分になる可能性が見られた。そのため、手話動作の手の形も識別対象とするなど、識別精度向上に繋げることが今後の課題である。

### 謝 辞

本研究開発は、公益財団法人電子通信普及財団の助成を受けたものです。また、手話動作をご指導いただいた株式会社ケイ・シー・シーの関係各位に感謝いたします。

### 文 献

- [1] 若尾 吏, 川喜多 佑介, 西村 広光, 田中 博, 三次 仁, “加加速度データと座標データを用いたマルチモーダル情報による手話動作認識,” 画像電子学会, Vol.52, No.2, pp.348-356, 2023.
- [2] 亀田 凱聖, 川口 大夢, 西村 広光, 田中 博, “手話学習支援を目的とした手話復習ツールの拡張機能,” 情報・システムソサイエティ特別企画, 予稿集, ジュニア&学生ポスターセッション, March.2024.