

# モーションキャプチャデータを用いた 日本手話の非手指信号の基本要素抽出

安江 謙太 酒向 慎司

名古屋工業大学

## 1 はじめに

手話は、手指信号(Manual Signals, 以下 MS)と非手指信号(Non-Manual Signals, 以下 NMS)によって表出される視覚言語の一つである。MS は手の動きによって表され、主に語彙の意味を表す。NMS は手以外の部位の動きによって表され、語彙の意味に加え文法標識や副詞的な意味を表す。日本手話における代表的な NMS としては、眉上げ、うなずき、首振り、まばたきなどが挙げられている[1]。

近年では手話の MS に着目してコンピュータで手話単語を認識する研究が盛んに行われている。その研究の一つとして、有賀らの研究[2]が挙げられ、様々な手話単語の中で現れる MS の共通要素に着目し、それらを集約するようなモデル化手法である。しかし、MS のみを用いた認識では、手話を単語単位で認識することはできても、手話文全体の意味を確定することは難しく、文法や副詞的な意味を表す NMS の認識が不可欠である。

先に述べたように日本手話で表出される NMS には様々なものがあるが、個々に表出されるパターンが特定の部位の形状によるものや、その動き方によるもの、動く方向によるものなどがあり、NMS の認識手法には定まった知見が十分に得られていないのが現状である。

本研究では、NMS の自動認識を念頭に有賀らが用いた MS の共通要素を集約する手法を NMS に適用し、NMS の主要な部位の動きを計測したモーションキャプチャデータに対して各部位で表出される形状や動きの固有のパターンの抽出を試みる。

## 2 手話の言語的特性を考慮した手話認識

有賀らの手法[2]では、画像から抽出された手の動きや形状の特徴データを用いて、統計モデルの一つである隠れマルコフモデル(HMM)により手話単語をモデル化した。そして、様々な手話単語の中で現れる共通要素を抽出するため、全単語の全状態パラメータ空間をクラスタリングし、得られた類似区間をサブユニットとした。このサブユニットを MS で表出されるパターンの

基本要素とし、様々な手話単語をサブユニットの遷移によって近似し、単語間でサブユニットモデルを共有したシステムを実装した。これにより、単語単位で学習されたモデルよりも高い認識率が得られることを示した。

NMS にも MS と同様に、様々な単語の中で現れる共通要素が存在すると考えられるので、有賀らの手法を NMS に適用することを検討する。

## 3 HMM クラスタリングによる基本要素抽出

### 3.1 対象とする部位

本研究では、表1に示される NMS として重要度の高いと考えられる基本的な部位と動作に着目して、部位ごとに基本要素の抽出を試みる。

### 3.2 特徴量抽出

NMS として表出される顔の部位は、形状や動き方の双方で自由度が高いため、画像データからそれらの情報を取得する場合、高度な画像処理の技術が必要となる。そのため、NMS の各要素の特徴を正確に取得できるモーションキャプチャデータが有効である。本研究では、後述する手話単語を対象に NMS として表出される顔部位も含めて詳細なデータが計測されたデータベースを用いる。それぞれの部位に設置したマーカーの3次元座標データの軌跡より、部位別の特徴量を取得する。

### 3.3 手話単語のモデル化

部位別に取得された特徴量を用いて、手話単語ごとに部位ごとの HMM を学習する。HMM は学習データを時系列に沿った複数の状態の遷移により表すため、HMM の各状態は、単語動作中の定常的な特徴区間を近似したものとなる。

### 3.4 クラスタリングによる基本要素抽出

学習された HMM の全単語の全状態をクラスタリングすることで NMS の各部位ごとの共通要素を抽出する。クラスタリング手法としては、有賀ら

表1:対象の部位

部位	代表的な動作	意味
頭部動作	うなずき 首振り	並列化 否定
眉	眉上げ	疑問詞
目	まばたき	平叙文
口形	ボ, バ	副詞的意味

A study on fundamental element extraction using motion capture data for non-manual signals in JSL, Yasue Kenta, Sako Shinji, Nagoya Institute of Technology.

が基本要素を抽出する際に用いた手法を用いる。HMM の全状態を個別のクラスタとし、クラスタ間距離が最小になるものから順次統合してゆく。統合されたクラスタを NMS の基本要素とする。

#### 4 評価実験

実験により、提案法が NMS の基本要素抽出に有効であることを示す。本実験では、長嶋らの研究[3]で収録された手話単語データセットを用いる。このデータセットは、400 単語の手話単語データで構成され、それぞれの単語ごとに Kinect センサーによる深度データ、モーションキャプチャデータ、映像データが同期収録されている。このデータセットのうち、異なる動作を含まない 281 単語のモーションキャプチャデータを用いる。

使用するモーションキャプチャデータは、被験者の体に設置されたマーカー112 個を用いて計測したものであり、顔表面にも 36 個のマーカーが用いられている。この中から NMS の動作を取得するために必要となるマーカーの座標データを用いて、それぞれの部位について HMM の特徴データ表現について検討した。部位ごとに用いるマーカーと特徴表現を表 2 に示す。

作成した学習用データを用いてそれぞれの部位ごとに HMM を学習し、クラスタリングを行った。口形について、1 単語 5 状態で HMM を学習し、学習された全単語の全 1,405 状態を、終了条件を調整することで 26 個と 50 個の 2 通りのクラスタ数にクラスタリングした。それぞれの場合について、全単語におけるクラスタごとの出現回数を図 1 に示す。

得られたクラスタそれぞれは、手話動作中の口形パターンを近似したものである。図 1 より、クラスタの出現回数は指数関数的に減少し、頻出する特定の口形パターンが存在することがわかる。図 2 にクラスタ数 26 における手話単語中の口形パターンの遷移を表した。生成された“口形 10”は図 1 の(A)において最も出現回数が多い口形パターンである。本研究で使用したデータベースでは、様々な手話単語で手話動作を行うタイミングで口を大きく開く表現が見られたため、この口形の出現回数が多くなっていると考えられる。ほかの部位についても、同様に頻出するパターンを抽出することができた。

#### 5 むすび

本研究では、様々な手話単語の中で現れる NMS の基本要素を自動的に抽出する手法について検討した。HMM 学習時の状態数やクラスタリング時のクラスタ数を変化させ基本要素の抽出を試み、

単語中で共通に表出される NMS の基本要素が得られていることが確認できた。

本研究では単語データを用いて NMS の抽出を行ったが、文法や副詞的な機能を含む NMS は、対話の中でより顕著に現れると考えられる。今後は提案手法を対話データに適用することで、より多様な NMS の基本要素が抽出できると考えられる。

#### 参考文献

- [1] 神田和幸, 谷千春, 植村英晴, 木村勉, 長嶋祐二, 原大介, “基礎から学ぶ手話学”, 2009.
- [2] 有賀光希, 酒向慎司, 北村正, “日本手話の音韻構造を考慮した HMM に基づく手話認識”, 電子情報通信学会技術研究報告, vol. 110, no. 221, pp. 127-132, 2010.
- [3] 長嶋祐二, 酒向慎司, 渡辺桂子, 原大介, 堀内靖雄, 市川薫, “手話の語彙構造・文法解明に供する 3D 超高精度 DB の開発”, 日本音響学会 2018 年秋季研究発表会, pp. 1471-1474, 2018.

表 2: 各部位で使用したマーカーと特徴量

部位	マーカー	特徴表現法
頭部動作	側頭部 2 点と鼻 1 点	3 点のなす平面の法線ベクトル
眉	両眉 3 点ずつの 6 点	側頭部 2 点の midpoint を原点とした座標
目	両目上瞼と下瞼の 4 点	上瞼と下瞼の距離
口形	上唇と下唇, 左右口角の 4 点	口の縦と横の距離

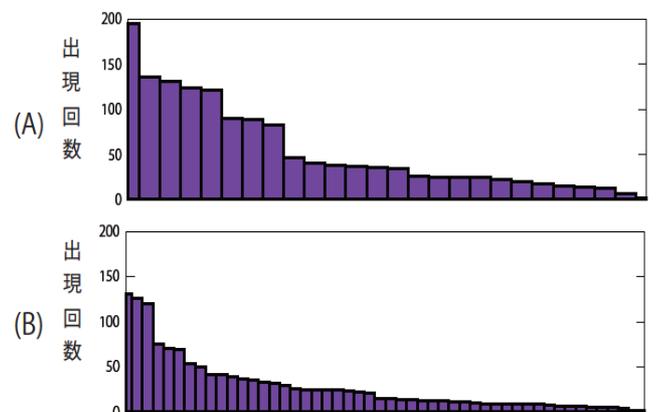


図 1: クラスタ数 26 (A) と 50 (B) における出現回数

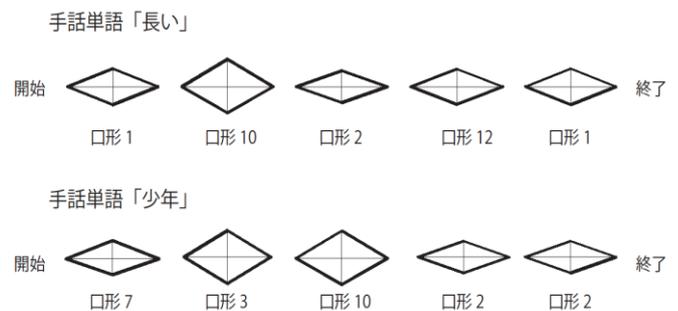


図 2: 口形 HMM (5 状態) における基本要素遷移の例