

手話の調動認識における HMM 最適化に関する検討

1 A B - 3

西田 昌生

長嶋 祐二

工学院大学

1 まえがき

聾者のコミュニケーションに主に用いられる手話は、手指の運動を 3 次元空間上で展開し表現される視覚的な言語である。手話の調動は手形を表す N 形態、運動を表す V 形態、非手指動作を表す S および G 形態により構成される [1]。

筆者らは、これまで手話の大局的な運動をトラッキングする方法として単眼カメラ [2]、および磁気センサーによる手法について検討した。入力されたデータから V 形態を認識する方法は HMM [3] を用いていた。しかし、HMM の形状は、手話の調動を十分に反映していなかった。また、調動の量子化は掌の移動方向に対してのみラベリングを行っていたため、必ずしも適当なラベル付けを行っていなかった。

そこで本報告では、各単語ごとに最も尤度の高くなるモデルを効率的に探索する方法について検討を行った。さらに、調動の特徴をより明確に反映させるため、コードブックに加速度ベクトルを考慮する方法についても述べる。

2 調動の量子化

掌の軌跡をトラッキングするために POLHEMUS 社製の FASTRAK を用いた。この計測装置は、磁気センサーから 1/30sec ごとに 3 次元座標値を計測可能である。これを手首に固定し、得られたデータから時間 t の掌位置と $t + \Delta t$ の掌位置の差分をとり、掌の移動ベクトルを求める。得られた移動ベクトルは、あらかじめ用意した図 1 のようなコードブック比較し方向ベクトルを量子化する。ここで、 X 軸は前額軸、 Y 軸は矢状軸、および Z 軸は垂直軸方向とする。

さらに、生成された移動ベクトル列から加速度を求め、移動ベクトルと組合せ手話の調動の量子化を行う。また、加速度も、通常の手話の調動で計測される加速度の最大値を求め、あらかじめ量子化しておく。

A Study of Optimization HMM in Recognition of Sign Language

Masao Nishida, Yuji Nagashima

Kogakuin University

2665-1 Nakano-cho, Hattouji-shi, Tokyo, 192, Japan

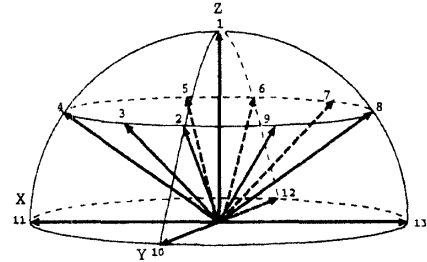


図 1: コードブックベクトル (上半球)

このような処理をすることにより

- 停止や停留状態
- 調動速度の時間的な変化

を観測シンボル系列に反映させることが可能となる。

また、手話の調動は、観測者には停止しているように見えても微振動が計測される。そこで、予備実験により得た経験値的な値を閾値とし、ある程度細かな動きに関しては停止しているもののみならずことにした。

3 最適なモデル形状探索

逐次状態分割法 (SSS) [4] と呼ばれるこの手法は、尤度を最大にするように、モデルの形状を順次変化させる方法である。これは、状態を並列方向、または直列方向へ分割するといった操作を繰り返すことにより、モデルの形状を目標となる調動に最適化する。

以下に、本手法のモデル最適化の流れを示す。

- 初期モデルとして 1 つの状態と終端と始点からなるモデルを作成する。この初期モデルに調動を学習させる。
- ある時点での状態の分割はモデルの全ての状態の中で最も情報量の多い部分を選択し、その状態を並列方向もしくは直列方向に分割する。
- 並列方向と直列方向の分割で対象となるシンボル系列の尤度が大きくなる分割方向を選択する。

このような状態分割を繰り返すことによって、少ない状態数で目標とする手話単語に最適なモデルを構築することができる。

4 評価実験

ここでは、本手法を用いて実際に単語認識の評価を行った。

モデルのパラメータ推定は、あらかじめ3の手法で各単語ごとに最適化処理を行ったモデルを用意し、2の方法で得られるシンボル系列を *Baum-Welch* アルゴリズムを用い繰り返し行った。パラメータ推定したモデルに対して未知のシンボル系列の出力確率を求め、確率が最大になるモデルを認識結果とする。なお、認識の計算は *Viterbi* アルゴリズムを用いた。

モデルの推定に用いたシンボル系列は、男性手話初心者2名の調動より単語ごとに200サンプルづつ用意し学習させ、それ以外の手話初心者4名の調動を未知シンボル系列として用いた。全ての被験者には *SIGINDEX* [5] の映像をあらかじめ観察させ、立位で行った。実験に用いた単語は片手のみの特徴的な動きの手話単語14種類を選び実験を行った。ただし、掌の軌跡だけでは同じような軌跡を描く調動（{兄},{賛成}等）があるので、そのような単語に関しては同一の単語として扱った。

また、比較のため状態分割を施さないHMMの評価も併せて行った。状態分割を施さないHMMは全ての単語に対して状態数5の *bakis* モデルを用いた。実験の結果を表1に示す。

表1: 認識結果

単語	非分割	状態数	分割	状態数
兄	47.5%	5	92.5%	9
言う	30%	5	92.5%	7
意外	35%	5	92.5%	8
色々	72.5%	5	75%	9
教える	92.5%	5	82.5%	7
教わる	72.5%	5	97.5%	6
がっかりする	82.5%	5	62.5%	4
通う	100%	5	90%	6
借りる	62.5%	5	90%	8
疑問	52.5%	5	72.5%	6
砂糖	85%	5	95%	8
賛成	32.5%	5	82.5%	6
首相	17.5%	5	80%	6
大丈夫	35%	5	75%	8
総合	57%		84%	

本方法では、非分割時と比べ総合平均で27%認識率が向上した。非分割の結果は各単語毎に非常にばらつきがあり、全体としての認識率も良いものではなかった。これは、状態数を単語の種類にかかわらず全て同じモデルで実験を行ったため、モデルの形が比較的最適化されたモデルの形状に近

い単語に関しては認識率が向上しそれ以外の単語は学習の効果が上がらなかったことを示している。

ここで提案した方法では、良好な認識率を示した。これは、あらかじめモデルを各単語ごとに適切な形状に変形しておくことで他の単語との誤認識を少なくしたためだと考えられる。

また、{教える}-{がっかりする}等、認識率が低下している単語もある。原因として状態分割評価に用いた調動の過学習がある。これは今回、状態分割の評価に用いた調動はパラメータ推定用シンボル系列を用いたが、そのために偏った形状になったのではないかと考えられる。

全体として、単語ごとに最適なモデルの形状を探索し、モデルのパラメータを推定することにより調動に対して頑健なモデルを構築できたことを示していると考えられる。

5 あとがき

ここでは磁気センサーを用いて掌の軌跡をトラッキングし、調動の特徴をより明確にラベル化するため、加速度ベクトルに対して量子化を行い、得られたシンボル系列に対してHMMを適用して学習および認識を行った。その際、HMMによる認識をよりロバストなものにするために、モデルの形状を最適化して認識評価を行った。

その結果、最適化を行っていないモデルと比較して良好な認識率を得ることができた。

今後の課題として連続な調動の認識、両手の調動認識への拡張、語形変化への対応等が挙げられる。

謝辞 本研究の一部は文部省科学研究費基盤研究(C)07808040の助成によるものである。

参考文献

- [1] 長嶋, 亀井, 杉山:形能情報駆動型手話アニメーションシステム, 信学会, ET96-86, pp.73-78(1996)
- [2] 藤井, 長嶋:2次元空間画像による調動認識, HI部会, N&R Vol10, pp17-52(1994)
- [3] 中川:確率モデルによる音声認識, 電子情報通信学会編 (1988)
- [4] 鷹見, 嵯峨山:逐次状態分割法 (SSS) による隠れマルコフネットワークの自動生成, 音響学会講演論文集, 2-5-13(1991)
- [5] 神田, 長嶋, 市川:サインデックス試案, 信学会, ET-96-83, pp.53-58(1996)