

位置と動きの動作要素に基づく手話認識に関する検討

A Study on Sign Language Recognition based on Gesture Components of Position and Movement

前畠 大 西田 昌史 堀内 靖雄 市川 薫*

概要. 従来の手話認識での研究では、単語単位で手話認識が行われてきた。しかし、単語単位での認識では、モデル数が膨大になってしまう上に、手話速度の速いものには適応に限界が来てしまう。その欠点を補うために動作要素認識が用いられているが、本研究では中でも重要な役割を果たす動きと位置の動作要素単位での認識を行う。認識方法は、各動作要素の逐次性や同時性を考慮するために隠れマルコフモデル (HMM) を用いて認識を行う。そして、各動作要素の時間情報を考慮して統合を行うことで単語数の増加によるモデル数の増大にも対応できる手話認識が可能であると考え。本研究では位置と動きを動作要素としてモデル化し、それらを統合したものを利用した、文単位、すなわち、連続単語での手話認識を目指した。実験での認識結果は位置動作要素認識率が 56.44%、動き動作要素認識率が 75.98%、単語認識率が 50.27%となった。

1 はじめに

手話認識に関する従来研究は、単語単位で手話認識するものであった [1]。しかし、孤立単語の認識では単語の増加に伴ってモデル数も増加させる必要があり、単語が増えるほど認識率が低くなる欠点があった。単語数の増加という欠点を補うものとして近年では動作要素レベルでの認識が行われている。この方式としては本研究室の滝沢らが研究を行っている [2][3]。しかし、滝沢らの研究では 20 個という単語数の少なさに問題があると思われる。そこで本研究では単語数を増やし、動作要素の中でも重要な位置と動きの動作要素を統合させ、文単位での認識を目指す。

2 手話の認識手法

2.1 認識システムの概要

入力装置は磁気センサを用いて手話文の左右の手の位置をフレームごとに 3 次元位置座標で取得したものである。次にそのデータのフレーム間の差分を取ることで運動に対するデータとして取得した。これらに対して各々独立にモデルを作り上げ、動作要素を定義した。それに対して、隠れマルコフモデル (HMM) を用いることで認識を行い、位置と動きを独立に認識した結果を統合させ単語を認識する。

2.2 位置動作要素

手話認識において、位置は重要な動作要素の一つである。手の動く領域を図 1 のように単純に 18 分

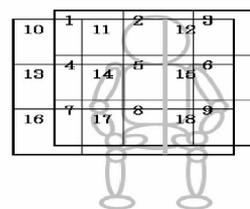


図 1. 位置の 18 分割

割し、位置のクラスタ化を行った。取得した座標データから位置動作要素モデルを作り上げた。動作要素は手の速度をグラフ化し、その極小点毎に区切って決めたものである。このモデルと HMM を用いて位置動作要素の認識を行った。本研究で用いた HMM は状態数 11、混合分布数 1 の left-to-right モデルである。例えば「食べる」という単語で言えば、まず胸の前に両手をもってきて、右手を口の前に持っていった後、右手を胸の前に戻し、また右手を口の前に持っていった後、胸の前に戻す動作である。単語に対して図 1 で表記されるクラスタにおける番号付けを行った。「食べる」という単語の例で言うとこの単語は 5→2→5→2→5 といったようなクラスタの番号で表記することが出来る。このように単語に対してラベルをつけた。学習に関しては文中の単語ごとに取得したデータにラベルを付けることで学習をさせ、それと辞書との比較によって認識を行った。

2.3 動き動作要素の認識

動きも手話認識で重要な動作要素の一つである。座標データの差分を取ることで動き動作要素のデータとした。動きは動く方向と運動の仕方 (直線・曲線など) の組み合わせによるラベル付けを行った。例

Copyright is held by the author(s).

* Masaru Maebatake, Masafumi Nishida, Yasuo Horiuchi, 千葉大学大学院 融合科学研究科 情報科学専攻, and Akira Ichikawa, 千葉大学



図 2. 単語認識の例

例えば、「食べる」という単語の例では、直線運動をしながら上、下、上、下というように動作を行っている。そこで、直線+上と直線+下というようなラベルをつけることで認識を行った。動き動作要素に関して位置動作要素と同様の HMM を用いており、学習・認識に対しても同様の方法を取っている。

2.4 単語認識

単語認識は先ほど用いた 2 つの動作要素の統合によって単語の認識を行う。HMM によって動作要素の検出時間が測定できるので、位置動作要素と動き動作要素を時系列ごとに追って行く図 2 のように作る。時系列ごとに追うことによって位置動作要素と動き動作要素に重なる部分が出る。この部分を単語としてみなすことで予め作っておいた辞書と比較して認識を行った。この辞書は、実際の手話文から位置・動きを見ることで生成したものである。

3 評価実験

3.1 実験条件

2 章で述べた動作要素認識とその統合の手法を用いて動作要素と手話単語の認識を行った。今回用意した例文は 25 文で単語数は 183 個で、異なり単語数は 81 個である。位置動作要素数は 36 個、動き動作要素数は 64 個であった。被験者は 4 人で 25 文を 5 回ずつ行ってもらった。この被験者はいずれも健常者である。また、今回の実験はクローズドで行った。

3.2 実験結果

上記のように行った実験結果のうち位置動作要素を表 1 に、動き動作要素については表 2、単語については表 3 のように示す。動き動作要素の認識に関

表 1. 位置動作要素の認識結果.

被験者	再現率 (%)	適合率 (%)
A	53.88(493/915)	55.37(510/921)
B	55.96(512/915)	58.40(539/923)
C	59.23(542/915)	60.20(552/917)
D	56.72(519/915)	60.27(590/979)
平均	56.44(2066/3660)	58.58(2191/3740)

表 2. 動き動作要素の認識結果.

被験者	再現率 (%)	適合率 (%)
A	75.52(691/915)	73.26(696/950)
B	79.45(727/915)	76.80(735/957)
C	68.09(623/915)	66.63(641/962)
D	80.87(740/915)	72.81(763/1048)
平均	75.98(2781/3660)	70.89(2835/3999)

表 3. 単語の認識結果.

被験者	再現率 (%)	適合率 (%)
A	53.03(480/905)	51.83(495/955)
B	46.41(420/905)	46.19(455/985)
C	49.72(450/905)	49.74(480/965)
D	51.93(470/905)	46.98(700/1490)
平均	50.27(1820/3620)	48.46(2130/4395)

してはまずまずの結果が出たと考えられるが、位置動作要素と単語認識に関しては改善の余地がある。位置動作要素認識率が低い理由として、単語のラベルを「食べる」という単語のラベル 5→2→5→2→5 のように長くなり、HMM が認識できないことが主な原因であると考えられる。単語認識率に関しては、位置動作要素認識率が低かったための結果であると推測する。改善策として位置動作要素のラベルを短くすることが考えられる。例えば「食べる」という単語例で言えば、2→5, 5→2 のように 2 つのクラスターで単語を表現することで長さの短縮化、動作要素数の減少が計られると考えられる。しかし、位置動作要素と動き動作要素に基づく手話認識が有用であることは示せたかと思える。

4 まとめ

文レベルでの認識を目指して位置と動きの動作要素による認識について検討した。今後は位置動作要素の認識率を高めることで手話認識の精度改善していく予定である。

参考文献

- [1] 金山, 白井, 島田:HMM を用いた手話単語の認識:電子情報通信学会技術研究報告, Vol.104, No.93, pp21-28(2004)
- [2] 滝沢, 堀内, 市川:動作要素の時間的関係を考慮した手話認識方式の検討:ヒューマンインタフェース学会研究報告集, Vol.3, No.5, pp47-52(2001)
- [3] 岡沢, 西田, 堀内, 市川:わり区間を含む単位を用いた手話認識手法の検討:電子情報通信学会技術研究報告, Vol.103, No.747, pp13-18(2004)