

手話認識における単語境界検出法の検討

7L-3

大平 栄二, 佐川 浩彦, 崎山 朝子, 大木 優

(株)日立製作所中央研究所

1. はじめに

聴覚障害者と健聴者との円滑なコミュニケーションの実現を目的として、手話通訳システムの研究を進めている。先に連続DP照合を用いた、聴覚障害者から健聴者への手話通訳システム[1]についての検討を行い報告した。本報告では、この手話通訳システムにおいて重要な手話認識の性能向上法について述べる。ここでは、手話の動作の特徴に基づいて(1)入力される手話の単語境界を検出し、(2)検出された単語区間の動作パタンの特徴づけを行う方式を提案する。本方式により、DP照合で検出される手話単語の候補の数を削減し、認識性能を向上させることが期待できる。

2. 手話認識方式の概要

手話認識システムの構成を図1に示す。手話はデータグループ(米国VPL Research社製)より1/30秒毎(この単位を以下フレームと呼ぶ)に記号化され入力される。具体的には、データグループは、手の位置や向き、ならびに各5本の指の第一、第二関節の曲げ角を特徴パラメータとして求める。本システムはパターンマッチングに基づく認識を行なう。まず、あらかじめ認識したい手話単語の特徴パラメータを標準パターンとして標準パターンメモリに格納する。そして、認識時には、入力される未知の手話の特徴パラメータ(これを入力パターンと呼ぶ)と標準パターンメモリに格納された各単語の標準パターンを比較(マッチング)することにより、入力パターンに含まれる手話の単語を認識する。

このマッチング結果としては、単語候補の集合である単語ラティスが得られる。単語ラティスには、正解の単語以外の単語も候補として検出される。手話単語列検出部は、この単語ラティスから正解の単語列を求め、手話-日本語変換部が助詞などを補足して自然な日本語に変換する。

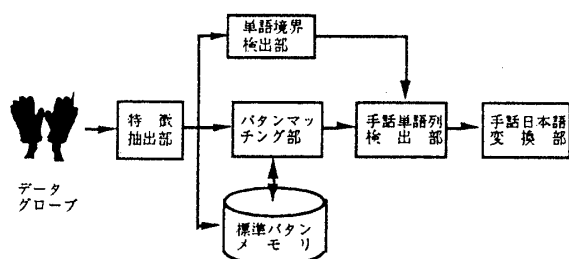


図1 手話認識システムの構成

3. 手話認識の問題点とアプローチ法

手話に限らず認識における重要な課題の1つに、パターンマッチングで検出される単語候補から、どのようにして正解の単語列を求めるかという問題がある。従来、自然言語や音声の認識では、

この処理において、文法、意味、常識などの知識を利用している。しかし、例えば、候補単語があまりに多いと、正解でない単語列の組み合わせが尤もらしい解釈になる場合がある。手話ではさらにつきのような問題がある。

(1) 手話は、日本語とは異なる独自の体系を持った言語である。しかし、まだ研究段階の言語であるため、体系化された文法などが無い。このため、日本語等に比べて文法に関する情報を制約として十分に用いることができない。

(2) 手話は、格助詞が表現されない。このため、格に基づく意味制約を十分に用いることができない。

以上のような理由により、認識性能を向上させるためには、パターンマッチング部において単語候補の十分な絞り込みが必要である。

音声では、文字に相当する音韻の他に、音韻にはあまり依存しない、声の強弱や抑揚といった韻律情報が音声を理解する上で大きな役割を果たしていることが知られている。例えば、韻律情報は、音声を単語や文節など単位に分割する[2]ほか、分割された単位の係り受け関係をも表現する場合もある。一方、手話でも音声同様この韻律情報に相当する特徴が存在すると考えられる。本研究では、このような単語境界を示す特徴を見だし、利用することにより、パターンマッチング部における単語候補を削減し、認識率を向上させることを目的とした。

4. 単語境界検出

4.1 特徴抽出

手話表現の基本単位[3]としては、(1)手形、(2)位置、(3)動きがあげられる。このうち単語にあまり依存しない、手形と動きに関する特徴量として、(1)手の形状の変化、(2)手の速度、(3)動きの方向を次のように求める。

(1) 手の形状の変化: フレーム t と隣接する $t-1$ フレームおよび $t-2$, $t-3$ フレーム間の手形状の変化 $DH1$, $DH2$, $DH3$ を求める。そして、 $DH1 \sim DH3$ が全て一定値以下の場合のみ手の形状が一定であると判断する。2フレーム間の手の形状の変化 DH は、5本の指の第一、第二関節の曲げ角のフレーム間での変化(差の絶対値)の総和として求める。

(2) 手の速度: 手の速度は次のように求める。

$$v(t)^2 = (x(t) - x(t-1))^2 + (y(t) - y(t-1))^2 + (z(t) - z(t-1))^2$$

$$v_{2d}(t)^2 = (y(t) - y(t-1))^2 + (z(t) - z(t-1))^2$$

ここで、 $x(t)$, $y(t)$, $z(t)$ は、入力される手話の t フレーム目の x , y , z 座標、 $x(t-1)$, $y(t-1)$, $z(t-1)$ は、 $t-1$ フレーム目の x , y , z 座標である。

(3) 動作の方向:

$$\text{平面動作MF} = (y(t) - y(t-m))^2 + (z(t) - z(t-m))^2$$

$$\text{上下動作率RV} = (z(t) - z(t-m))^2 / \text{MF}$$

$$\text{左右動作率RH} = (y(t) - y(t-m))^2 / \text{MF}$$

ここで、 t と $t-m$ は、隣接する分割点(次節で定義)位置のフレームである。

4.2 単語分割方式

単語分割方法を以下に示す。まず、分割候補位置として、手の

速度が一定値以下の「徐行区間」と速度の極小位置を検出する。そして、各々について、以下の規則により境界の判定を行なう。

(1) 「徐行区間」の始末端は、基本的には単語境界とする。

東京都心身障害者福祉センターの関氏や文献[3]における分析結果によると、「休止」や「静止単語」では、約200msec(6フレーム)以上の動作の停止が生じるとの報告がある。このため、連続した5フレーム(約167msec)以上の徐行区間のみ「徐行区間」とし、それ以下の場合には徐行区間とせず単に極小位置とする。ただし、3次元の速度 $v(t)^2$ では5フレーム未満の「徐行区間」であっても、2次元の速度 $v_{xy}(t)^2$ で5フレーム以上の「徐行区間」であれば、後者を「徐行区間」とする。

(2) 極小位置を単語境界候補とし、反復動作などで単語内で生じる候補位置を削除する。

単語境界は、速度が0近くになる時のみでなく、速度が極小を示す位置でも生じる。この原因の1つは、手の座標のセンサは手の甲の中心にあるため、指先の座標を表していないこと。さらに手は直線的に動くのではなく楕円を描くように滑らかに動くためであると考えられる。このため、極小位置は全て単語境界候補とし、以下に示す規則に基づいて単語境界を決定する。

- (a) 一般に、単語内では手の形状を一定に保つ手話が多い。このため、連続する3つの極小位置で手形状が一定(図2-1)であれば、中間の極小位置は単語内位置とする。
- (b) 反復動作では、単語内で手の形状が一定に保たれないことも多い。ただし、極小位置前後の区間では一定に保たれる傾向にある。このため、連続する3つの極小位置の中間の極小位置で手形状の一定区間が存在し(図2-2)、先頭と最後の極小位置の空間上の距離 $D2$ が中間と最後の極小位置の空間上の距離 $D1$ より小さく($D2 < D1$)、かつ距離 $D2$ が一定値以下なら、中間の極小位置は反復動作の単語内位置である。
- (c) 極小位置の手形状が一定で(図2-2)、速度が一定値以上であり、かつ前後の区間の動作の方向が同じであれば、中間の極小位置は単語内位置とする。
- (d) 「徐行区間」が、後続単語の準備のためのポーズ区間であれば、「徐行区間」終端が後続単語の開始点となる。この場合、「徐行区間」の終端で、手の形状の移動が完了し、後続単語の動作に移る。すなわち、「徐行区間」の終端から手形状一定となる。このため、「徐行区間」に後続する極小位置のうち、「徐行区間」から極小位置までの間の手形状が一定でないならば単語境界とする。
- (e) 途切れた「徐行区間」でも、連続しかつ両方の「徐行区間」で手形状が一定であれば、1つの「徐行区間」とする。

4.3 動作タイプ付与方式

前節で分割された区間の動作の特徴付けを行なう。ここでは、次の4つの動作タイプに特徴付ける。どれにも該当しない場合は、顕著な特徴がない区間である。

(1) 反復動作区間

反復動作を伴う手話単語区間である。単語分割における規則(b)で検出可能である。

(2) 上下方向、左右方向動作区間

上下方向、左右方向動作の顕著な動作である。

上下方向の動作区間:

MF \geq 約5cm/フレーム, RV \geq 0.8 (角度が約65度以上)

左右方向の動作区間:

MF \geq 約5cm/フレーム, RH \geq 0.8 (角度が約25度以下)

(3) ポーズ区間と動きの少ない手話単語区間

「徐行区間」は、「ポーズ」の場合と「動きの少ない手話単語区間」のどちらかである。

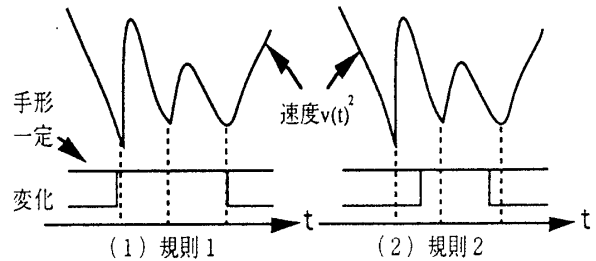


図2 手の速度と手形状

5. 評価実験

本方式を評価するため、実験を行なった。実験は文献5, 文献6に掲載されている話者1名の、11文、延べ30文(198単語)を対象に行なった。手話は連続した動作であるため、単語境界位置を一意に決定するのは難しい。このため、正解の単語境界位置は、対象の文の手話をコンピュータグラフィックスによる手話合成[4]により表示し、目視により境界位置と思われる5フレーム前後の区間を検出し、その区間に含まれていたら正解とした。実験結果を表1に示す。

表1 実験結果

対象文(単語数)	正解	誤分割	脱落
30文(198単語)	177単語	10単語	11単語

対象の198単語中、単語内を誤って単語境界としたのが10単語、単語境界を単語内と間違ったのが11単語であり、89%の分割性能が得られた。脱落のうち7例は、反復動作単語の前の準備動作をも単語内としたものであった。一方、誤分割は、ノイズにより生じた極小位置が原因となった3例の他、回転運動を伴う手話単語内(具合、口)で4例生じた。また、87単語(44%)で動作の特徴付けができた。

6. おわりに

手話の動作の特徴に基づく単語境界検出ならびに分割された区間の手話の動作パターンの特徴付ける方法を提案した。この結果は、認識性能の向上のみならず、手話の分析における基準としても有効であると考えられる。今後は、実際の認識における有効性を実証していく予定である。

謝辞

本研究に関して、貴重な助言を頂いた中京大学の神田和幸教授、東京都心身障害者福祉センター関宜正先生に感謝致します。なお、本研究の一部は、RWC(Real-World Computing)プロジェクトの委託研究として行ったものである。

参考文献

- [1]佐川他: 連続DP照合を用いた手話通訳システム, ヒューマンインタフェース44-12 (1992)
- [2]大平他: 韻律情報を用いた音声会話文の文構造推定方式, 信学論Vol.J72-A, No.1, 23-31 (1989)
- [3]神田他: 日本手話の構動素の表記, 音声研究会資料S84-102 (1985)
- [4]森地他: 手話翻訳のための身振りコーディング法, 第8回ヒューマンインタフェースシンポジウム
- [5]貞広編集: わたしたちの手話(会話編1), (財)全日本ろうあ連盟(平3-9)
- [6]飯塚; ともに歩むためにおぼえようみんなの手話 会話編, (財)全日本ろうあ連盟