98-01031

指手運動軌跡による手話認識・学習支援システムの研究(継続)

代表研究者　堀口　進　北陸先端科学技術大学院大学情報科学研究科教授
共同研究者　阿部　亨　北陸先端科学技術大学院大学情報科学研究科教授
共同研究者　山森　一人　北陸先端科学技術大学院大学情報科学研究科教授

はじめに

聴覚障害を持つ人々と聴覚障害を持たない人々のコミュニケーション手段として、手話は一般的かつ重要な手段である。しかし、現状では手話通訳者が不足しており、聴覚障害者のコミュニケーションの手段が十分確保されているとは言えない。手話習得には時間がかかり、さらに手話通訳者となるには何年もの経験が必要となる。そこで、手話通訳をコンピュータに代行させる手話認識や指文字認識の研究の重要性が増している[1]。

手話の一部である指文字認識に関しては、ニューラルネットを用いた手法[2]や、手の関節角をコード化して認識を行う手法[3]、ベイズ識別を用いた手法[4]などが行われてきた。手話認識に関する研究には、手および腕の状態を画像を用いて検出／認識をする研究[5]や、センサーを装着してデータを採取し、圧縮連続DP照合を用いて認識する手法[6]などがある。

平成10年度の本研究では、マッチングを用いて位置姿勢および手形状入力装置を使用した指手運動軌跡を用いたマッチング手話単語認識システムを提案した。マッチングを用いた手話単語認識手法では、特定話者については比較的高い認識率が得られたが不特定話者については有効でないことが分った。そこで、本年度は、FFTを用いて手話単語の指手運動軌跡を分析し複数話者を対象とした高調波分析による手話単語認識手法を提案し、手話単語認識システムの評価を行い、その有効性と問題点を明かにする。

2　手話単語の入力システム

2.1　システムの構成

ハードウェアの構成を図1に示す。ホストコンピュータとしてSilicon Graphics 社製 OnyxVTX(150MHz、64Mbytes)を用いる。 ハードウェア的なシステム構成はホストコンピュータと手形状および位置姿勢からなる入力装置から成る。入力装置は、操作者が装着するセンサ部分、センサ部分からのデータ採取とホストコンピュータとの通信を行うセンサ制御装置部分から構成される。

これら手話単語入力システムの構築、ならびに編集を行うソフトウェア・システムを作成した。作成したシステムは機能的に3つのプロセス・モジュールに分けた。手話単語入力システムの機能的な概略を図2に示す。 手話単語入力システムは、センサ制御装置と通信を行うセンサ制御モジュール、入力手話データを可視化するモジュール、データを総合的に制御し情報を交換しながら全体の処理を制御をするモジュールによって構成される。

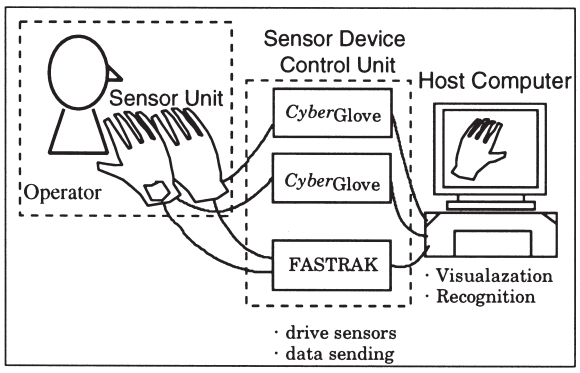

図1　単語データ入力システム

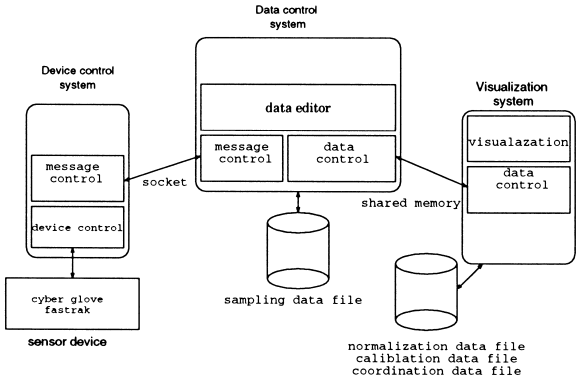

図2　手話単語入力システムの概略

2.2　入力装置

本システムでは手形状入力装置として Virtual Technologies 社製の CyberGlove[8]を用いる。 CyberGlove は手袋状のセンサ装置と制御装置から構成される。 片手あたり手指の主要18関節のデータを採取可能となっている。装置を手に装着した外観を図3に示す。

関節角センサはグローブ内部に縫い込まれており、各関節の外転角を測定する。CyberGloveは測定素子として光ファイバが使用されている。関節を曲げた場合、動作に従って光ファイバに曲げが生じ、光ファイバの屈折率が変化する。 光ファイバの場合、屈折率は曲げ角に関して線形に変化する性質があり、この屈折率を測定することで関節の曲げ角を検出する。関節角センサーは図4に示される部分に装着されている。ここでTRは親指と掌との開き角を示す。各指の第1、第2関節に対応した屈曲角を測定するセンサ(Ordering ID:(1、2)、 (4、5)、 (7、8)、 (10、11)、 (13、14))、指と指の開き角を測定するセンサ(Ordering ID:3、6、9、12)が取り付けられている。また、親指には手の掌上を小指方向にどれだけ回転したかを測定するセンサ(Ordering ID:0)がある。 指に関するセンサは全部で14個となる。これらに加え手首の曲げに関して2個、手の甲の曲げに1個あり、図4に示される手の主要な18関節の屈折角が測定可能である。

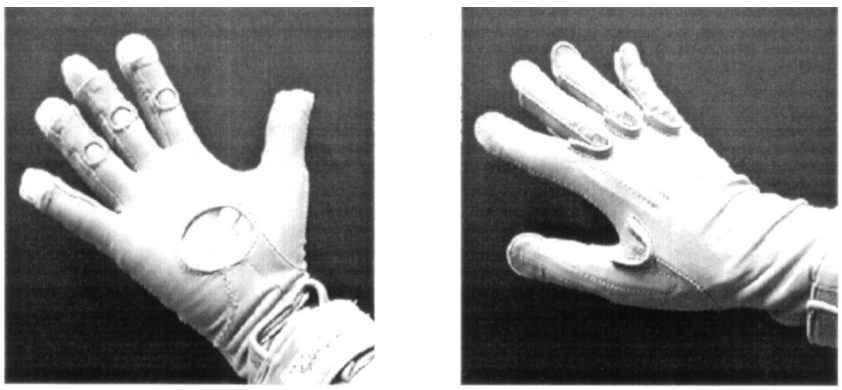

図3　CyberGloveの外観

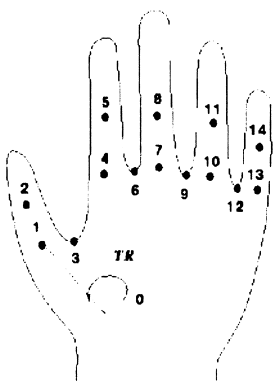

図4　手形状入力装置の関節角側定点

手話を認識するためには、手の位置および姿勢情報が必要となる。3次元の位置・姿勢を測定する装置として、図5に示すPOLHEMUS社製のFASTRAK[9]を用いる。

FASTRAKは磁気によって位置・姿勢を計測する。トランスミッタと観測基準点となるレシーバを組み合わせ、３次元位置・姿勢情報を測定する。得られるデータは３次元座標(x、y、z) と回転座標(yaw、pich、roll)の6自由度となる。 制御装置からホストコンピュータへ送られてくるデータはANSI/IEEE のFloating-Point Arithmetic 754-1985 に準じたfloat型の精度である。 サンプリング周波数はレシーバ１個の場合に最大120Hzであり、レシーバを複数個用いた場合はレシーバの個数で120Hzを分割した周波数となる。レシーバは最大で４個同時に観測可能であり、その際のサンプリング周波数は30Hzとなる。 本システムでは両手に関する情報を得るために2個のレシーバを用いる。よって60Hzのサンプリングレートとなる。

本システムでは、レシーバをCyberGloveの手首の位置に装着し、３次元位置および姿勢の計測を行う。

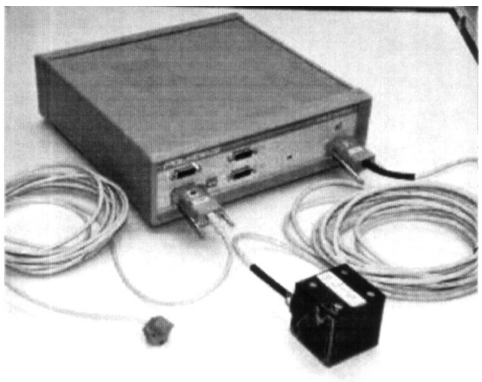

図5　FASTRAKの概念

2.3　手話単語データの可視化

手話単語データの採取中および採取後において、手話データの可視化はデータの確認のために必要である。 本システムでは、可視化のためのグラフィクス・ライブラリとしてOpenGLを用い、隠線隠面およびシェーディング機能を導入した。デバイス制御処理と描画処理を１つのプロセス内で直列に処理した場合、データサンプリング速度、描画速度のそれぞれの処理回数が約15[回／秒]に留まった。 これに対し、デバイス制御プロセスと描画プロセスを並列に実行することで、データサンプリング速度を60[回／秒]、描画速度を約40[回／秒]として実行可能となった。

3　認識対象の手話単語

対象単語は電子辞書「ムサシ」[10]を参考に任意の330単語を選んだ。 対象とした単語を表1に示す、

手話単語は5名の被験者から採取した。１単語に関して複数の同じ動作を採取する。 １単語に関して複数回提示されたそれぞれのパターンを、１サンプルとする。辞書作成用に5サンプル、さらに辞書作成用サンプルに5サンプルを採取した。手話単語１語の区間は以下のように定め、採取を行った。

1　両手をひざの上に置く。
2　１語を提示する。
3　再度、両手をひざの上に置く。

ここで、「両手をひざの上におく」は身体部位の情報収集において採取した「膝」の位置に相当する。また、片手手話の場合、使わない左手は膝の上で指をそろえて（指文字の「テ」に相当）置くものとする。被験者の手話に対する習熟度は、5名の被験者の内、3名が手話経験があり、各人は10年、4年、2年の者である。 残り２名は手話経験がなく電子辞書を見ながらそれを真似して提示した被験者である。

それぞれの被験者が提示した単語パターンは手話音韻学的には必ずしも完全一致してはいない。 例えば、身体に対する手腕の提示空間が異なったり、あるいは繰り返しの回数がことなる場合もある。 また、手話音韻学的には一致していても、より主観的な運動の区分として異なる場合もある。それは例えば、交互運動における順番(右手からかあるいは左手からか、など)であったり、提示速度の違いなどである。これらは手話単語のもつ表現の曖昧さである。本実験ではこれらの曖昧さを持った複数話者による手話単語認識を行う。

表1　対象単語

挨拶, 会う, 赤, 明るい, 秋, 朝, 浅い, 明後日, 明日, 遊ぶ, 暖かい, 頭, 新しい, 熱い, 集まる, あなた, 兄, 姉, 危ない, ありがとう, ある, 歩く, 安心, 言う, 家, 以下, 怒る, 以外, 生きる, 行く, 幾つ, 石川, 医者, 椅子, 忙しい, 痛い, 一日, 一年, 一番, 一緒, 一般, 意味, 妹, いろいろ, 上, 嘘, 美しい, 旨い, 生まれる, 裏, 売る, 選ぶ, 多い, 大きい, 教える, 遅い, 教わる, 夫, 弟, 男, 一昨日, 大人, 同じ, 覚える, おめでとう, 重い, 思う, 面白い, 表, 女, 会社, 買う, 顔, 書く, 過去, 貸す, 家族, 固い, 悲しい, 金, 通う, 借りる, 軽い, 可愛い, 変る, 間, 考える, 関係, 簡単, 学校, 頑張る, 北, 昨日, 決める, 今日, 兄弟, 嫌い, 疑問, 臭い, 曇り, 悔しい, 暮す, 比べる, 来る, 苦しい, 車, 黒, 計算, 結婚, 決心, 健康, 現在, 恋人, 答え, 断る, 子供, 細かい, 困る, 最高, 最後, 最初, 探す, 淋しい, 寒い, さようなら, 賛成, 残念, しかし, 試験, 仕事, 自然に, 下, しっかり, 姉妹, 趣味, 手話, 障害者, 小学, 勝負, 昭和, 調べる, 白, 信じる, 自慢, 住所, 自由, 上手, 好き, 過ぎる, 少し, 捨てる, 全て, すみません, する, 座る, 生活, 相談, 卒業, 空, 大切, 高い, 立つ, 例えば, 楽しい, 食べる, 大学, 大丈夫, 騙される, 騙す, だめ, 誰, だんだん, 小さい, 近い, 違う, 父, 中学, 長, 通訳, 使う, 月, 次, 机, 作る, 都合, 続く, 妻, 強い, 適当, テレビ, 天気, 電車, 電話, 東京, 遠い, 時, 得意, 友達, 取る, どこ, どちら, 無い, 中, 長い, 泣く, なぜ, 懐かしい, 何, 名前, 苦手, 西, 日曜日, 日本, 入学, 人気, 盗む, 願う, 眠い, 眠る, 寝る, 年齢, 農業, 飲む, 入る, 始める, 恥ずかしい, 話合い, 母, 速い, 春, 晴れ, 反対, 場所, 火, 東, 引く, 飛行機, 筆談, 必要, 人, 人々, 暇, 開く, 平等, 深い, 不思議, 不満, 冬, 古い, 文化, 下手, 部屋, 返事, ほとんど, 本, 本当, 毎日, ますます, まずい, 貧しい, まだ, 町, 間違い, 待つ, まで, 短い, 水, 道, 南, 未来, 見る, 息子, 娘, 難しい, 無駄, 無理, 明治, 迷惑, 珍しい, 盲人, 目的, もし, もっと, もらう, 森, 約束, 安い, 休み, 破る, 柔らかい, 指文字, 良い, 用事, 読む, 夜, 離婚, 両親, 料理, 恋愛, 聾唖, 老人, 若い, わからない, わかる, 別れる, 分ける, 忘れる, 私, 悪い, ０, １, ２, ３, ４, ５, ６, ７, ８, ９, １０, ２０, ３０, ４０, ５０, ６０, ７０, ８０, ９０, 百, 千

認識実験手順は図7に示されるように大きく分けて２つのフェーズがある。 これら認識実験の手順は、DPマッチングを用いた手法、あるいは調波分析を用いた手法のどちらにおいても同じである。一つは辞書を作成するための手順であり、もう一つは実際の認識のための手順である。 辞書作成／認識手順のどちらの手順の場合でも、前処理としてデータの平滑化、単語区間設定が必要である。 ここで単語区間は、手動にて切り出しを行った。

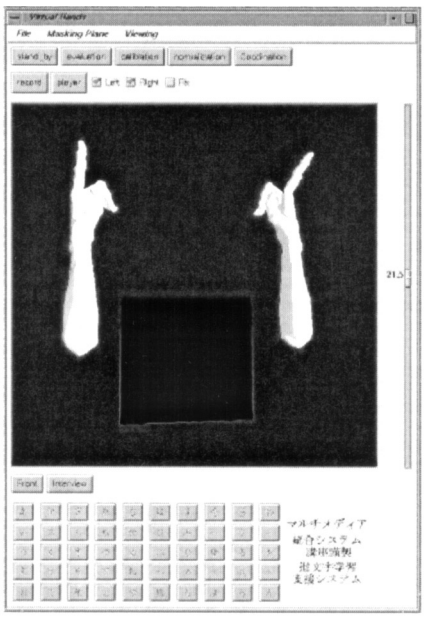

図6　手話データ可視化画面

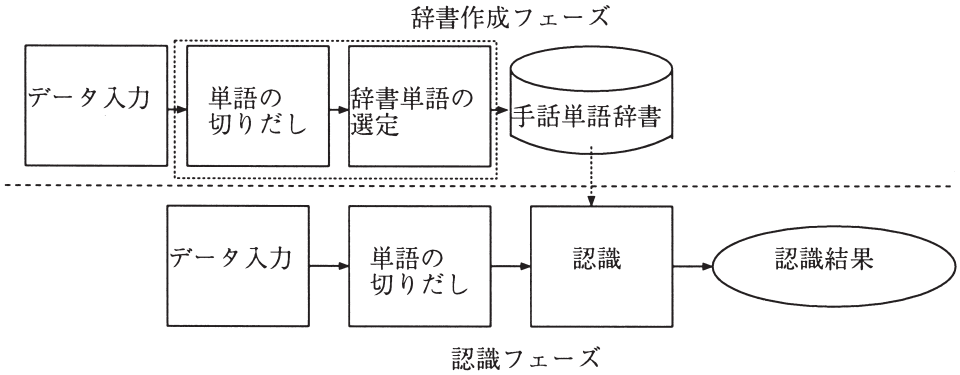

図7　認識実験手順

4　DPマッチングを用いた手話単語認識

4.1　DPマッチング

DPマッチング認識手法は、動的計画法(Dynamic programing)にもとづき、２つのパターン間距離の算出方法を定義し、入力テストパターンに対して一番距離が近い辞書パターンを認識パターンとする手法である。
図8にDPマッチングのモデル図を示す。

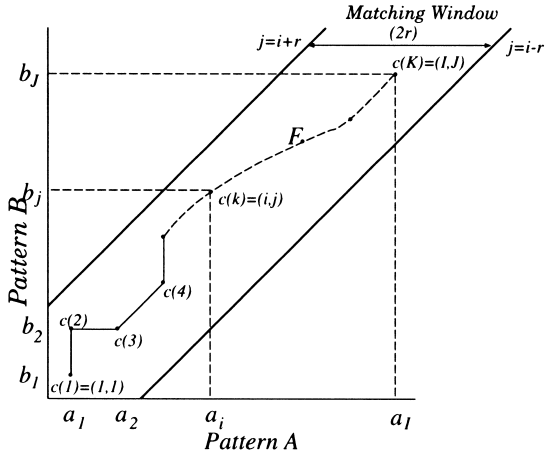

図8　DPマッチング

4.2　中心近接尺度と辞書の作成

DPマッチングによる認識手法では認識の際に辞書データを必要とする。DPマッチング自体は2パターン間の距離を算出するだけである。そこで、より標準的な辞書を与えるという目的で中心近接尺度による辞書データの作成を行う。 1単語について辞書データ作成手順を以下に示す。

1　1単語について複数の辞書候補となるパターンを採取する。
2　すべての辞書候補に対してそれぞれのパターン間の距離を定義する。
3　距離に従って、空間上に辞書候補パターンを配置する。
4　辞書候補パターンが占める領域内で、もっとも中心に近いパターンを辞書パターンとする。

中心近接尺度の算出方法と、中心近接尺度を用いた辞書パターンを作成手順を以下に示す。

1　N個のパターンを採取し、それぞれをパターンP_nとする。(n=1, . . . , N)
2　パターンP_a, P_bの距離$D(P_a, P_b)$を定義する。

3　パターンP_xに関する中心隣接尺度を以下のように定義する。

$$C_x = \frac{1}{N-1}\sum_{i \neq x}^{N} D(P_x, P_i)$$

4　中心近接尺度C_xを最初にするパターンP_xを辞書パターンとする。

距離を算出する関数$D(P_a, P_b)$には DPマッチングを用いる。その例を図9に示す。パターンP_Aは中心隣接尺度C_Aがすべてのパターンの中で最小であり、標準に一番近いと考える。また、最大のC_Cを持つパターンP_Cは標準とするのには一番適さないパターンである。

複数の被験者からそれぞれの単語に対して手話単語サンプルを採取して、それらの中から中心近接尺度に基づいて辞書データを作成するモデル図を図10に示す。 図が示すように辞書データとして保持されるデータは、特定の被験者の発話によるデータそのものである。

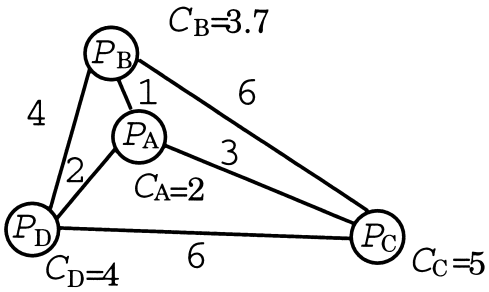

図9　パターンの距離関係例

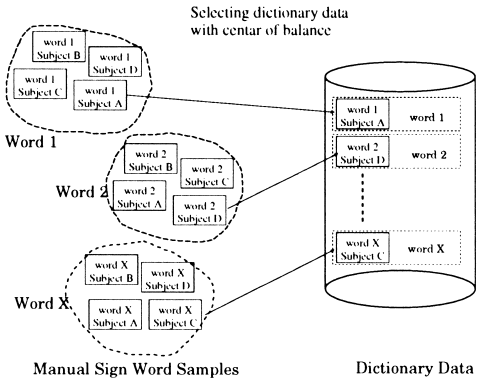

図10　中心近接尺度を用いたDPマッチングのための辞書作成

4.3　特定話者の手話単語認識

本実験で取り扱う手話単語パターンは、時系列要素として48次元の特徴ベクトルから構成される。DPマッチングを行う場合の前処理として、特徴ベクトルの各要素の分散を求める。分散はすべての辞書データの全単語区間に渡って求める。特徴ベクトル間の距離$D(i, j)$を定義する際には、 求められた分散によって正規化を行い、すべてのベクトル要素の距離の加算平均を求める。 この演算によって時系列要素の距離を算出し、DPマッチングを行う。

DPマッチングによって単語パターン間の距離を求め、辞書パターンの中からテストパターンに距離が一番近い辞書を認識パターンとする手話単語の認識実験を行う。 まず、特定話者認識におけるDPマッチングを行い、特定話者認識におけるDPマッチングの認識率の確認を行う。 その後に、調波分析を用いた手法と比較するために、DPマッチングを用いて複数話者に対して手話単語認識実験を行う。 この際、サンプリングレートを変化させた場合と同様の条件を作り、認識演算時間との比較を行う。

辞書パターンとテストパターンを採取した被験者が同じ場合(特定話者認識)について実験を行う。辞書作成においては、１単語に対して10パターン採取した内から、5パターンを辞書選定用パターンとした。中心隣接距離を用いて辞書選定用パターンから１パターンを認識用辞書に選定した。

手話単語データの採取時にはサンプリングレートを60Hzとして単語パターンを採取した。 認識実験では、単語のサンプリングレートを変化させた場合の認識率の変化を観測するために、間引き間隔を変化させて認識率を比較する。ここで間引き間隔とは60Hzで採取したデータから一定間隔でデータを抜き出す際のパラメータである。間引き間隔2、4はそれぞれ30Hz、 15Hzでデータを採取したのと同等になる。 また、DPマッチングにおける整合窓の大きさ$D_{\text{window size}}$は比較される2つのパターンの長さの平均をそれぞれL_{patternA}、 L_{patternB}としたとき、以下のように動的に決定する。

$$D_{\text{window size}} = \frac{1}{3}\frac{L_{\text{patternA}} + L_{\text{patternB}}}{2} \tag{1}$$

それぞれの単語に関して、辞書選定用パターンとして5語パターンを使用している。 テストパターンはこれら辞書選定用に使用したパターン以外の3パターンを使用する。認識結果は、3回の異なる入力テストパターンに対する平均の認識率を示す。実験結果を図11に示す。平均認識率は、間引き間隔2で99.3%となり、間引き間隔32で行った実験でも93.9%と高い認識結果を示した。

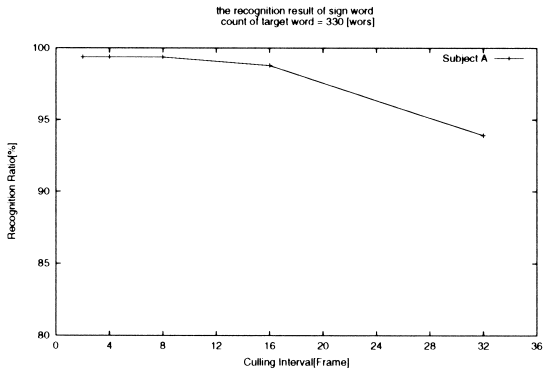

図11　DPマッチングを用いた特定話者認識結果

4.4　複数話者の手話単語認識

中心隣接距離を用いて認識用辞書を作成した後、複数話者認識の実験を行った。被験者は4名であり、各人の手話単語パターンから任意の5パターンを辞書選定用とし、計20パターンから辞書パターンを選択した。整合窓の大きさ$D_{\text{window size}}$は特定話者認識と同様、式(1)を用いる。

4人の被験者に対して間引き間隔を変化させて行った認識実験結果を図12に示す。認識率は間引き間隔を変化させてもほとんど変化がなかった。演算においてはPentiumIIプロセッサ、クロック450Hz、メモリ512Mbyteを使用して行った。

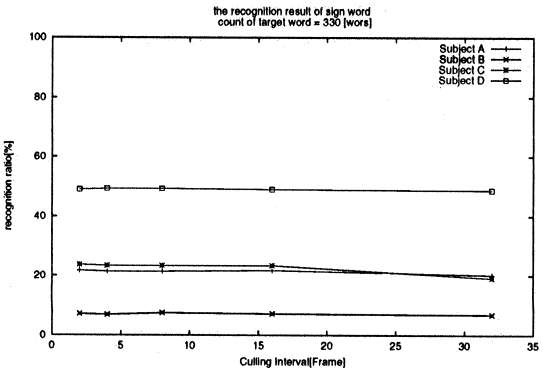

図12　DPマッチングを用いた複数話者認識結果

表2に、330単語に関して、辞書パターンとして選択されたそれぞれの被験者の単語パターン数と、平均認識率を示す。認識率は被験者Cに関して49.2%となった。被験者Bに関しては認識率は7.1%となった。 これらの認識率は、それぞれの被験者の発話データが単語データとして何パターン採用されたか(使用単語割合)との間に高い相関が見られる。また、それぞ

れの単語認識を詳細に観察した場合でも、辞書単語パターンとして選択されたパターンの発話者(被験者)以外の発話者のテストパターンが正しく認識されることは極めて少なかった。総テスト単語パターン数1320語に対する認識率は25.07%となった。 これは、総テスト単語パターン数に対して、辞書データの発話者がテストデータの発話者と同じ組合せである場合の比率に極めて近い。 これらの結果より、複数話者認識において、DPマッチングは入力話者の発話パターンが、辞書として含まれていない場合ほとんど誤認識される結果となった。

表2　辞書に使用された各被験者の発話パターン数と平均認識率

被験者名	使用単語数	使用単語割合 [%]	平均認識率 [%]
被験者 A	74	22.4	21.4
被験者 B	20	6.1	7.1
被験者 C	160	48.4	49.2
被験者 D	76	23.1	22.3

5　FFTを用いた手話単語認識

5.1　離散フーリエ変換

DPマッチングを用いた複数話者に対する手話単語認識持つ問題の解決を目的に、離散フーリェ変換を用いた調波成分によるを認識手法を提案する。

複数話者に対する手話単語認識を行うにあたって考えなければならないのが、手話単語をパターンとして捉えた場合に、その分散の度合をいかに考慮するかである。 手話単語は局所的な伸縮を伴う時系列パターンであり、時間領域で一意に時間要素の対応づけを行うのは困難である。 そこで、周波数領域における比較を検討する。 ただし、不揃いな長さの単語パターンに対してFFTを行うのでは、周波数領域でもその長さは不定になる。 そこで、時間軸上でパターンの長さ(フレーム数)を正規化し、この正規化したデータに対してFFTを行うことで、 周波数領域での取り扱いを容易にする。単に時間領域でサンプル個数を一定にして比較する場合と異なり、FFTによって位相成分を得ることで時間的な揺らぎを吸収することが可能となる。これらの処理によりパターンの分散を考慮した辞書の作成が可能となる。

入力$x(t)$があった場合、フーリエ級数展開は以下のようになる。

$$x(t) = \frac{1}{T}\sum_{n=-\infty}^{\infty} X_n \exp(j2\pi\frac{n}{T}t) \tag{2}$$

$x(t)$が領域 $\frac{-\mathrm{T}}{2} \leq \frac{\mathrm{T}}{2}$ で定義される場合、フーリエ係数X_nは次式で求められる。

$$X_n = \int_{-T/2}^{T/2} x(t)\exp(-j2\pi\frac{N}{T}t)dt \tag{3}$$

手話データの採取によって得られたデータは、有限長の離散データと見なすことができる。入力されたデータを、離散データ数 Nを周期として扱った場合、積和のkの変域を0 〜N−1とできる。

時間区間を 0 から T、 周波数範囲を $-f_x$ から $+f_x$ とし、その外では T および $2f_x$ 周期で同じ波形を繰り返す時間関数とした場合、離散フーリエ変換DFTと離散フーリエ逆変換IDFTとして、以下の式(4)および、式(5)が得られる。

$$X_k = \mathrm{DFT}(x_n) = \sum_{n=0}^{N-1} x_n exp(-j2\pi\frac{N}{kn}) \tag{4}$$

$$x_n = \mathrm{IDFT}(X_k) = \frac{1}{N}\sum_{k=0}^{N-1} X_k exp(j2\pi\frac{N}{kn}) \tag{5}$$

離散フーリエ変換および離散フーリエ逆変換に用いる複素指数関数を式(6)に定義する。

$$W_N = exp(-j2\pi\frac{1}{N}) \tag{6}$$

W_N は複素平面上で単位円の全周を N 等分した点を表し、W_{N^p} は p の増加とともに円周上を負の方向に $1/N$ 円周刻で動く点を表すため、W_N および W_{N^p} は回転因子とよばれる。
回転因子W_N を用いて式(4)および式(5)を書き直すと、それぞれ式(7)、式(8)となる。

$$X_k = \sum_{p=0}^{N-1} x_n \exp(-j2\pi) = \sum_{p=0}^{N-1} x_n W_{N^{kn}} \tag{7}$$

$$x_n = \frac{N}{1} \sum_{k=0}^{N-1} X_k \exp(j2\pi) = \sum_{p=0}^{N-1} X_k W_{N^{-kn}} \tag{8}$$

高速フーリエ変換FFTはMがある約数の積に分解できるとき、上式を高速に計算するためのアルゴリズムである。 離散フーリエ変換の定義式によって計算する場合M^2回の複素計算を必要とするが、FFTを用いた場合、特にMが2のべき乗である場合において、$Mlog_2M$回の計算ですむことが知られている。

時系列データを変換し周波数領域で比較するために、時間領域で同じ長さの区間を区切りFFTを行うことが必要となる。一方、手話単語データは非線形な伸縮を伴うために、その長さがある範囲内で不定である。そこで、線形補間を用いて固定の長さMに時間軸方向に正規化を行う。MはFFTの効率から考えて２のべき乗の値が望ましい。この処理により、不定長の手話単語データをFFTを行い周波数領域で扱えるようになる。 各々のパターンを比較する際にはそれらの周波数領域での要素は周波数成分の意味はなくなり、基本波に対する第n高調波として取り扱う。

特徴ベクトル要素数Nの時系列データを、長さMの時間正規化データを$N \times M$の行列として扱う。 FFT処理は時系列でデータに対して複素数値をとるため 周波数領域では $N \times 2M$ の行列データとなる。これを改めて$N \times M$ の行列とする。 それぞれの要素の平均値と分散を求めることでパターンマッチングが可能となる。

5.2 FFTを用いた辞書パターン作成

FFTを用いた手法では周波数領域でパターンの比較を行う。線形補間による時系列のフレーム数の正規化とFFTを用いることで、すべてのパターンを同じ$N \times M$ の行列として扱う。これにより、対応させる要素を一意に決めることが可能となる。ここで N は特徴ベクトルの要素数であり、 M はフレーム長である。要素n, mの値を$s_{n,\ m}$とするとき、平均値 $MD_{n,\ m}$と標準偏差 $SD_{n,\ m}$はそれぞれ以下の式(9)、式(10)によって求めることができる。辞書パターンの作成は図13に示されるように、被験者が提示した同じ単語の複数パターン(サンプル)から生成する。 ここでIは辞書作成用サンプル数である。

$$MD_{n,m} = \bar{x}_{n,m} = \frac{1}{I} \sum_{i=1}^{I} s_{n,m} \tag{9}$$

$$SD_{n,m} = \dot{x}_{n,m} = \frac{1}{I} \sum_{i=1}^{I} \sqrt{(s_{n,m} - \bar{x}_{n,m})^2} \tag{10}$$

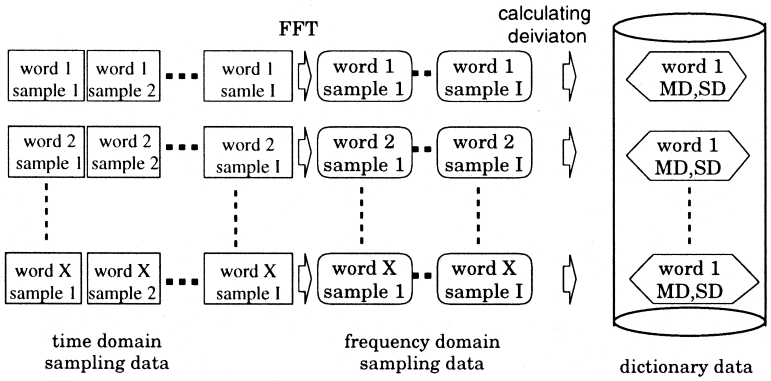

図13　FFTを用いた単語認識の辞書作成

5.3 重み付き距離による認識法

辞書パターンとテストパターンの距離算出に重み付き距離を用いる。重み付き距離D^lは次式によって求められる。iを認識単語とする。

$$D^l = \sqrt{\frac{1}{N(M+1)} \sum_{n=1}^{N} \sum_{m=0}^{M} \frac{(x_{n,m} - \bar{x}^l_{n,m})^2}{s^l_{n,m}}}$$

$$D^i = \min_l D^l$$

単語lの辞書データは各要素の平均値$\overline{x}^{l}_{n,m}$と、標準偏差$s^{l}_{n,m}$である。入力パターンを$x_{n,m}$とすると、単語lの辞書の$\overline{x}^{l}_{n,m}$との差を求め、$s^{l}_{n,m}$によって正規化することで、D^{l}を求める。D^{l}を最小にするよう単語lを認識単語とする。

手話単語認識実験で用いるN、Mそれぞれの値は以下のようになる。

N：特徴ベクトルの次元数 (96 = 48 ×2)
M：0次高調波からの帯域幅数(0 ～ 16)

5.4 前処理および辞書作成

DPマッチングによる認識手法を複数話者に適応させることは難しい。そこで、FFTを用いて手話単語を時間領域から周波数領域へ変換し高調波成分によって辞書作成および認識を行う手法の評価を行う。調波分析による認識手法を、特定話者と複数話者(3名、4名および5名)を対象とした場合について実験を行う。次に、高調波分析による認識手法を用いた手話単語の運動軌跡の認識実験を行う。

被験者より採取された手話単語データは、長さが不定である時系列データである。 時系列データの長さを一定にする目的で、手話単語データを時間軸のフレーム数に関して正規化を行う。 フレーム数正規化として、線形補間を用い64フレームに正規化する。手話の関節角や位置情報などは、周波数領域で考えた場合、音声言語と比較し、必要な帯域が狭い。また、FFTにおいては離散データの個数Nは２のべき乗において効率が良いことから、フレーム数を64フレームとして実験を行った。この際に外乱ノイズを除去する目的で移動平均を用いた平滑化を行う。

DPマッチングを用いた認識実験同様、対象単語数は330単語であり、また、辞書選定用に単語あたり5パターンを使用して高調波分析を行った後、辞書を作成した。

5.5 特定話者の認識実験

高調波成分を比較する場合、基本波に対してn番目の高調波まで使用する場合を高調波幅nとする。比較は複素成分を用いて行う。

特定話者に対して、高調波幅nを変化させた場合の認識結果を、図14に示す。特定話者認識の場合、高調波分析を用いた手法では89.0%から94.5%の認識率が得られ、DPマッチングの特定話者認識に準じた結果を得た。

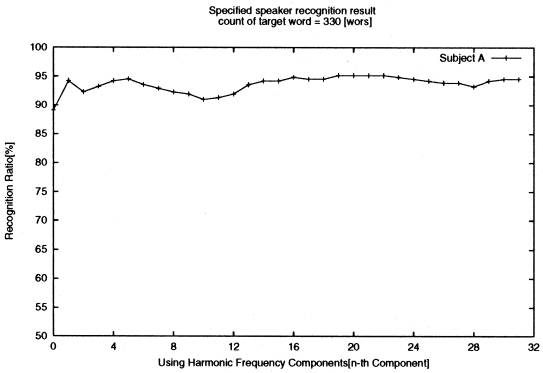

図１４　周波数分析による特定話者認識結果

5.6 複数話者の認識実験

使用高調波成分を変化させた場合の、3名の複数話者の場合の手話単語実験の結果を図15に、4名の場合を図16に、5名の場合を図17に示す。

3名の話者を対象とした実験ではすべての被験者において、高調波幅n=12とした場合に一番良い認識率が得られた。 表3に高調波幅n=12の場合の第1位認識率および第3位認識率を示す。ここで、第1位認識とは入力パターンの単語とに対して認識されるべき辞書単語が一番距離が近く、正しく認識されたことを意味する。すなわち、すべての辞書データをテスト入力パターンとの距離によって順位づけした場合に第1位となった場合である。入力パターンに対して、１番目、あるいは2、3番目までに現れた場合を第3位認識とし、この結果に基づき第3位認識率を算出する。被験者Aから採取したテスト入力パターンが最もよい認識率となり、最高で92.0%の認識率を示した。被験者Cから採取したのテスト入力パターンが最も認識率が低くなった。被験者Cの場合で最高で75.4%の認識率であった。

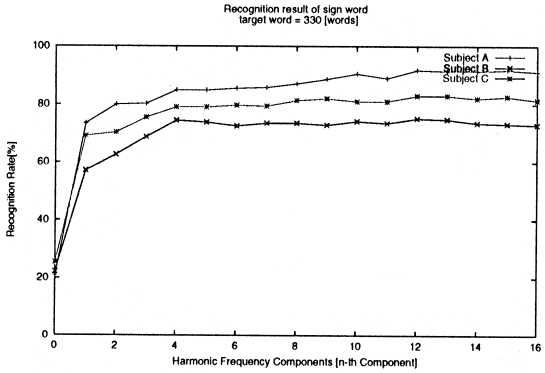

図15　3名を対象にした高調波幅と認識率の関係

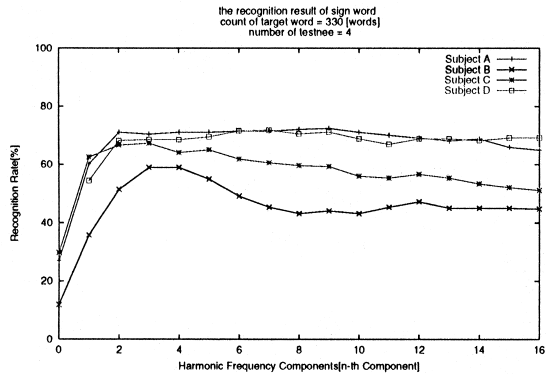

図16　5人を対象にしたFFTを用いた認識の結果

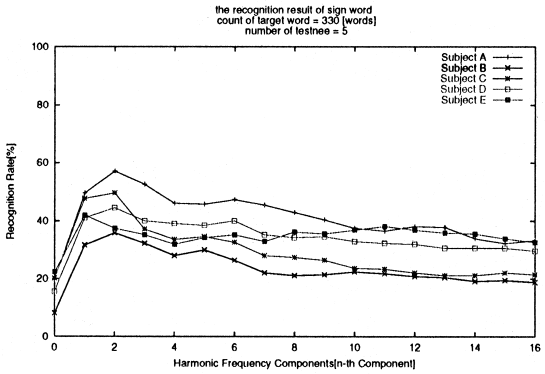

図17　5人を対象にしたFFTを用いた認識の結果

表3　3名を対象とした周波数分析による認識結果（高調波幅N=12）

被験者名	第1位認識率 %	第3位認識率 %
被験者A	92.0	97.7
被験者B	75.2	90.7
被験者C	83.0	93.6
平均	83.4	94.0

本実験では、高調波幅n=12の時に一番良い認識率となった。それ以上の高調波幅では逆に認識率が若干悪くなった。一定の高調波幅以上で認識率が下降したことは、高域における測定ノイズや、高域にでは手話単語として有効な情報が少ないことが原因と考えられる。

被験者数4名および5名の場合の、それぞれに対する認識率を、表4および表5に示す。4名を対象とした実験は、3名を対象とした場合と比較して認識率が全体的に低下した。高調波幅n=3の場合に高い一番高い認識率となり、平均認識率は64.7%となった。 被験者Aの場合、高調波幅n=3では71.4%を示した。一方で被験者Cの場合、50%程度まで認識率が低下した。被験者数5名の場合でもさらに全体的な認識率の低下みられ、平均認識率は45.0%となった。

手話単語データの高調波成分において、低調波成分は認識に関して特に有効な成分であることがわかる。反対に、複数話者認識において対象話者を増やした場合にも、使用する高調波幅に関して認識率のピークがあることが分かる。

表4　4名を対象とした認識結果（高調波幅N =2)

被験者名	認識率 %
被験者 A	71.4
被験者 B	51.8
被験者 C	66.9
被験者 D	68.5
平均	64.7

表5　5名を対象とした複数話者認識結果（高調波幅N =2)

被験者名	認識率 %
被験者 A	57.2
被験者 B	36.0
被験者 C	49.8
被験者 D	44.7
被験者 E	37.6
平均	45.0

まとめ

本報告では、指手運動軌跡による手話単語認識手法の提案ならびに認識システムの構築を行った。330単語を対象とした、特定話者認識においてはDPマッチングを用いた手法で99.3%の認識率が得られた。しかし、複数話者に対する認識実験を行った結果、 4名の複数話者の手話単語認識では、DPマッチングを使用した場合25.1%の第１位認識率に留まった。高調波成分を用いた手法では3名および4名の複数話者を対象とした場合、それぞれの認識率は83.4%、64.7%となり調波分析を用いた手法の有用性を確認した。さらに、5名を対象とした場合では45.0%となった。

本報告では単語区間は手動切り出しを行っていたが、手話単語学習支援システムを念頭に置いた場合、単語区間の自動切り出しが必要である。また、不特定話者に対する認識手法の提案が今後に期待される。

参考文献

[1] 加藤雄士, 神田和幸, 長島祐二, 市川薫, 黒川敏夫：.“ 手話工学の現状と将来の研究課題 ” ヒューマンインターフェイス, pp. 37-44, 1997.

[2] 井出英人,内田雅文.“ 手形状認識と手話への応用 ”. 電気学会論文誌, Vol. 114-C, No. 10, pp. 995-999, Oct. 1994.

[3] 岸野文郎, 高橋友一.“ 手振り認識方法とその応用 ”.電子情報通信学会論文誌, Vol. J73-D-II, No. 12, pp1985-1992, Dec. 1990.

[4] 堀口 進, 後藤岳志.“ 動作を伴う指文字を含む連続指文字認識 ”. 電気関係学会北陸支部連合大会 F-54, pp396, 1996.

[5] 福本, 間瀬, 末永.“ 画像処理を用いた指示動作検出の実験システム ”. 電子情報通信学会春期全国大会論文集, 1-251, 1991.

[6] 佐川, 酒匂, 大平, 崎山, 阿部.“ 圧縮連続dp照合を用いた手話認識方式 ”. 電子情報通信学会論文誌, Vol. J77-D-II, No. 4, pp. 753-763, Apr. 1994.

[7] 神田和幸.“ 手話学講義 ”. 日本福村出版, 1994.

[8] “ *Cyber Glove*™ *User 's manual* ”.Virtual Technologies, 1993.

[9] “ *3SPACE USER ' S MANUAL* ”. POLHEMUS, 1993.

[10] 神田和幸.“ ムサシα日本手話電子辞書 ”.アルファメディア, 1995.

＜発 表 資 料＞

題　　名	掲載誌・学会名等	発表年月
“振動子付き手形状入力装置を用いた指文字学習支援システム”	北陸先端科学技術大学院大学リサーチリポートIS-RR-98-00029P, pp. 1-28	1998年11月
“手形状入力装置を用いた指文字認識システム”	北陸先端科学技術大学院大学リサーチリポートIS-RR-98-00033P, pp. 1-29	1998年12月
“曖昧な運動を含んだ手話単語の認識”	平成10年度電気関係学会北陸支部連合大会論文集	1998年10月
“DP照合による未知操者の手話単語認識率”	平成10年度電気関係学会北陸支部連合大会論文集 p. 350	1999年10月
“振動子付き手形状入力装置を用いた指文字学習支援システムの性能評価”	北陸先端科学技術大学院大学リサーチリポートIS-RR-99-00017P, pp. 1-21	1999年4月