

Title	手指運動軌跡のコード化による手話単語認識に関する研究
Author(s)	鈴木, 信勝
Citation	
Issue Date	1999-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1259
Rights	
Description	Supervisor:堀口 進, 情報科学研究科, 修士

修士論文

手指運動軌跡のコード化による 手話単語認識に関する研究

指導教官 堀口 進 教授

北陸先端科学技術大学院大学
情報科学研究科 情報システム学専攻
マルチメディア統合システム講座

610060 鈴木 信勝

1999年2月15日

目次

1	序論	1
1.1	研究の背景と目的	1
1.2	本論文の構成	2
2	聴覚障害と手話	3
2.1	はじめに	3
2.2	聴覚障害	3
2.2.1	聴覚障害と聾啞	3
2.2.2	聴覚障害者のコミュニケーション手段	4
2.3	手話言語学および手話音韻論から見た運動軌跡	5
2.3.1	手話の動作を記述するための音韻表記法	5
2.3.2	音韻構造の部門とカテゴリー	6
2.3.3	手話音韻構造の表現	7
2.4	まとめ	9
3	手話単語データの入力および編集システムの構築	10
3.1	はじめに	10
3.2	システムの構成	10
3.3	入力装置	12
3.3.1	手形状入力装置	12
3.3.2	位置・姿勢入力装置	12
3.4	入力装置の制御	13
3.5	手話単語データの可視化とデータ受送信	16
3.5.1	手話単語データの可視化	17
3.6	装置の初期調整	19

3.6.1	手形状入力装置の調整	19
3.6.2	手形状入力装置による関節角の情報収集	19
3.6.3	位置・姿勢入力装置による身体部位の情報収集	21
3.6.4	データの編集加工	23
3.6.5	手話単語記録時のインターフェイス	25
3.6.6	手話単語の手動切り出し	25
3.7	まとめ	29
4	手話単語および運動軌跡の認識手法	30
4.1	はじめに	30
4.2	DP マッチング	30
4.2.1	中心近接尺度と辞書データの作成	33
4.3	FFT による認識手法	35
4.3.1	離散フーリエ変換と FFT	35
4.3.2	辞書パターンの作成	37
4.3.3	誤り率	37
4.4	運動軌跡の認識手法	38
4.4.1	既得音韻構造情報からの情報抽出	38
4.4.2	音韻抽出による辞書作成	39
4.5	まとめ	43
5	認識実験と評価	44
5.1	はじめに	44
5.2	対象単語	44
5.2.1	対象とする手話単語	44
5.2.2	単語パターンの採取	44
5.3	認識実験手順	46
5.4	DP マッチングを用いた手話単語実験	46
5.5	FFT を用いた手話単語実験	52
5.5.1	実験データの前処理と調波分析による辞書作成	52
5.5.2	認識実験	52
5.6	運動軌跡の認識実験	57
5.6.1	音韻情報分析による運動軌跡グループ	57
5.6.2	運動軌跡認識実験結果	58

5.7	比較と検討	59
5.8	まとめ	60
6	結論	61
6.1	今後の課題	62
	謝辞	63
	研究業績	64
	研究業績	66

目 次

2.1	音韻構造のツリー表現	6
3.1	単語データ入力システム	11
3.2	システムモデル	11
3.3	CyberGlove の外観	13
3.4	手形状入力装置の関節角測定点	14
3.5	FASTRAK の概観	15
3.6	手話単語データの流れ	17
3.7	手話データ可視化画面	18
3.8	観測値が不完全である場合の警告の正規化画面	20
3.9	観測値が要件を満たした場合の正規化画面	20
3.10	オフセットとゲインの関節角調整画面	21
3.11	関節角調整画面全体	22
3.12	身体位置収集画面	24
3.13	身体位置情報可視化画面	24
3.14	データ制御インターフェイス	26
3.15	手話単語採取インターフェイス	27
3.16	単語切り出し編集用インターフェイス	28
4.1	D P マッチング	31
4.2	D P マッチングの傾斜制限	33
4.3	パターンの距離関係例	34
4.4	単語対単語認識時の辞書作成	38
4.5	音韻構造のツリー表現	40
4.6	単語から音韻抽出認識のための辞書作成	42
5.1	認識実験手順	47

5.2	DP マッチングを用いた特定話者認識結果	48
5.3	DP マッチングを用いた複数話者認識結果	50
5.4	DP マッチングの間引き間隔による演算時間	51
5.5	整合窓と認識率の関係	51
5.6	周波数分析による特定話者認識結果	53
5.7	3名を対象とした有効高調波階数と認識率の関係	54
5.8	4人を対象にした複数話者認識の結果	55
5.9	5人を対象にした複数話者認識の結果	55
5.10	有効高調波階数と認識時間の関係	56
5.11	34単語に対する特定話者の運動軌跡認識結果	58
5.12	34単語に対する複数話者の運動軌跡認識結果	59
5.13	認識時間の比較	60

表 目 次

3.1	身体位置情報収集の採取位置	23
4.1	音韻解析による音韻の種類および出現回数	41
5.1	対象単語	45
5.2	辞書に使用された各被験者の発話パターン数	49
5.3	DP マッチングによる認識率の平均	50
5.4	周波数分析による 3 名を対象とした複数話者認識結果	54
5.5	周波数分析による 4 名を対象とした複数話者認識結果	56
5.6	周波数分析による 5 名を対象とした複数話者認識結果	56
5.7	使用単語およびグループ	57

第 1 章

序論

1.1 研究の背景と目的

近年, 聴覚障害者の社会進出が進むにつれ, 聴覚障害者と健聴者がコミュニケーションを必要とする機会が増えている. 聴覚障害者同士の場合, コミュニケーション手段として手話は一般的な方法である. また, 口頭による意志交換手段としての即時性や明瞭性などの観点から, 健聴者と聴覚障害者のコミュニケーション手段としてもその地位が高まっている. しかし, 手話を知らない人間と手話を用いる人間の間では必然的に手話通訳者が必要になる. しかし, 現状ではその手話通訳者少なく, コミュニケーションの機会が十分保証されているとは言い難い. そこで, 計算機による手話通訳の代行への要望が高まっており, 現在, コンピュータを使用した手話通訳に関して, 手話認識や指文字認識の研究の重要性が盛んに行われている.

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 [1] や, 高橋らの手の関節角をコード化して認識を行う手法 [2], ベイズ識別を用いた手法 [3][4] が行われてきた. 手話認識に関する研究 [5] には, 画像を用いた手および腕の状態を検出し認識をする研究 [6] や, 手および腕の状態をセンサーを装着してデータを採取して認識を行う研究がある. 手や腕の状態を検知する装置として手形状・位置姿勢入力装置と呼ばれる装置があり, 佐川ら [7][8] は, これらの装着型装置を用いて入力した手話パターンを採取し, 標準パターンと照合する圧縮連続 DP 照合を提案している. この方式は, 620 語を用いた大語彙の手話単語認識実験を行った結果, 98.7% の高い認識率が得られている. しかし, 認識に用いた実験データは個人データであり, 認識システムを利用できる話者が限定されるという問題がある.

より多くの話者による手話を認識するために, 複数話者を対象とした手話単語認識手法

の確立が必要となる。本論文では、FFT 分析を用いた複数話者を対象とした手話単語認識手法を提案する。入力デバイスとして位置姿勢および手形状入力装置を使用する。さらに、FFT 分析を応用して、手話単語における、運動軌跡の認識について検討を行う。

1.2 本論文の構成

本論文の構成は次のとおりである。第 2 章では、聴覚障害とコミュニケーション手段としての手話について述べる。

第 3 章では、まず本研究で用いる、位置姿勢および手形状入力装置について触れる。さらにそれら入力装置を制御や入力したデータの加工、および手話単語データの収集のための作成したデータ編集システムについて述べる。

第 4 章では、従来特定話者用に用いられてきた DP マッチングについて述べる。次に、本研究で提案する時系列入力データを FFT を用いて周波数領域に変換し誤り率を用いて行う単語対単語の認識手法について述べる。その後、提案手法を用いて運動軌跡の認識を行う手法について述べる。

第 5 章では、第 4 章で述べたそれぞれの手法に基づいて構築した認識システムを用いて、不特定話者を対象とした認識手法とその認識結果について述べる。

第 6 章で結論を述べる。

第 2 章

聴覚障害と手話

2.1 はじめに

本章では聴覚障害について述べ、次にコミュニケーション手段としての手話について述べる。さらに手話言語学と手話の音韻についての概略について触れ、運動軌跡についての検討を行う。

2.2 聴覚障害

2.2.1 聴覚障害と聾啞

聴覚障害とは、聴覚器官である耳からその情報を処理する脳までの経路のうち一部あるいは全体に損傷があるため、音や言葉を聞き取れない、もしくは聞き取り難い障害を指す。それらの機能障害を持つ者を聴覚障害者とし、副次的に発生する音声による言語習得の困難さから、広義では発話を含む音声によるコミュニケーションが著しく困難な者をその範疇とする場合もある。

聴覚障害者が用いるコミュニケーション手段として一般的、かつ重要な方法として手話がある。また、手話以外の手段として、障害の程度や相手によって幾つかの方法がある。相手が手話を理解しないなどの状況では相手の口の形でその発声内容を知覚する口話や筆談などがある。

口話や筆談は、表現形態が多少異なるが、どれも日本語そのものである。一方の手話は、文化的な側面から日本と語の類似点も多く見い出されるが、日本語に単語レベルで一对一の対応関係にはなく、日本語とは異なる言語である。一方で、手話で用いられる手話単語を日

本語文に一对一で対応させた 日本語対应手話 (Sim-Com: simultaneous communication) と呼ばれる表現もあるが、これは一般的な手話としては用いられる場面は全体からすると少ない。

手話は単純な手真似のようなものではなく、音声言語学の音韻と同様に幾つかの構成要素の組み合わせの上に成り立っている。日本で用いられる手話は日本手話と呼ばれ、健聴者が用いる日本語と関係する部分も多いが、昨今の言語学的見地からの研究により、日本語とは異なる言語体系を有していることが明らかになっている [9]。

日本以外にも手話が存在する国がある。たとえばアメリカには A S L (American Sign Language) と呼ばれる手話があり、これは日本手話とは単語 / 文法とも別言語である。

すべての聴覚障害者が手話を用いるわけではなく、聴覚障害者個人の失聴の過程により手話の重要性が異なる。幼児期の言語習得期間までに失聴し音声言語が習得できない場合、その大部分が手話を母国語として習得する。逆に、音声言語習得後に失聴した場合は、母国語としての音声言語があるため、そうでない聴覚障害者と比較すると、習得の要件や失聴の度合いにより手話の重要性が低くなる傾向がある。

手話を母国語とするような聴覚障害者を聾啞者あるいは聾者とよび、一方、音声言語を母国語とする聴覚障害者を途中失聴者と区別する。

2.2.2 聴覚障害者のコミュニケーション手段

本来、音声言語を伴わないコミュニケーションにはさまざまなものがある。例えば筆談もその範疇になり、相手の唇の形を読む口話もある。これらはいわば伝統的な非音声言語コミュニケーションである。また、近年の計算機の普及に伴いパーソナルコンピュータなどのキーボードを介した手段も注目されている。

しかし、聾啞者にとっては手話こそが母国語である。これは言語習得期に手話を母国語として習得したため、音声言語を母国語とする人間がその母国語で思考するように、聾啞者は手話を用いて思考することを意味する。そのためにも、日本語を介してのコミュニケーションではなく手話そのものの認識手法の確立が必要とされているのである。

音声言語においても顔の表情がそのコミュニケーションの質を変化させることがあるが、手話はそれ以上に顔の表情が重要であることは見逃せない。さらに手話においては、少数ではあるが手指運動と同時に口の形を変化させることで全く別の意味を表す場合もあり、相手の顔を見ながら話すということが、単なるコミュニケーションの形式以上に重要だといえる。

また自然言語であるため、地域による差異、すなわち方言などもある。しかし、日本語が

そうであったように NHK による手話放送により、今後、標準手話が確立されると考える。

2.3 手話言語学および手話音韻論から見た運動軌跡

手話は、音韻的な構成や形態的な文の構成など、言語としての性格を持ち合わせている。本節では特に音韻論と音韻の記号表記方法の概略を中心として、手話単語の取り扱いについて述べる。その上で、音韻と音韻運動軌跡について検討する。

2.3.1 手話の動作を記述するための音韻表記法

我々が手話単語を扱う際の要望として、手話単語を一意に表せる道具が必要となる。音声言語であれば通常、それに一対一に対応する文字表現が存在する。また IPA(国際音声記号) とよばれる表記方法があり、これを用いれば文字を持たない言語であっても、音韻レベルでの一意な表記が可能である。

一方手話に関しては唯一の表現という方法は未だになく、またそれぞれの方法についても発展途上である。

古くはストーキー (Stokoe) によるケロロジー (Cherology) と呼ばれる仕組みがある。また、サットン (V.Sutton 1978) 式表記法とよばれる、元々はバレエの動きを記述するところから始まった人間の動きの記述法があり、計算機上でサットン式で記述するためのプログラム (SignBank I) が開発されている。その他にもハンブルグ大学でプリルウィッツ (S. Prillwitz) らが開発したハンブルグ手話表記システム (Hamburg Notation System for Sign Languages), 略称ハムノーシス (HamNoSys) とよばれるシステムも存在する。

通常、日本手話であればそれぞれの手話単語は日本語の訳語と一緒に紹介される。しかし、それぞれが自然言語としての性格を有し、すべての単語が1対1することはない。通常は訳語として日本語ラベルが存在するに留まる。手話単語を日本語ラベルだけで扱う場合、しばしば重複や揺らぎ発生する。例えば、日本語の異なる動詞と名詞が、手話では1単語で表されるなどは、手話単語を厳密に扱おうとする場合、非常に厄介な問題となる。また、番号による管理も恒久的な方法とはなりえない。その意味でも手話単語を音韻の組み合わせとしてとられその組合せの記述法の確立は、手話認識の研究にとっても非常に興味があるところである。

日本手話に特化した表記法としては神田らが提案した音韻表記法がある。これに関してはのべ 660 単語に関して記述されたものが入手なのであり、本論文でも音韻に関してはこれを拠り所とする。

2.3.2 音韻構造の部門とカテゴリー

音韻レベルにおいて手話言語が音声言語と比較して言えることとして、時間的空間的複雑さがある。音声言語の場合、ある音韻が提示されている時間にはほかの音韻が存在することはなく、時間軸上で単調である。一方、手話に関しては腕が2本あることに由来する時間的空間的共立が生じること以上に、静的情報である手の形や手が位置する場所、動的情報である動きが共立関係にある単語が多数あるという複雑さを持つ。

伝統的音声音韻論を考えた場合、音声言語は音素に対して記号を与え取り扱う。同様に手話言語でも音韻というレベルの存在の示唆より、それぞれの音韻としてどのようなものが存在するのか、さらにそれぞれの音韻がどのように結合して手話として表現されるのか議論されている。

日本手話の表記法に関して唯一の表記法というものはまだ存在しない。ただし、現在までの日本手話や日本手話の表記法をまとめた研究として、「日本手話の音韻表記法 [10]」が提案されている。

「日本手話の音韻表記法」に基づく手話の音韻表記法では、音韻構造はツリー構造とみなす。音韻構造は3部門(手の構え、身体位置、運動)から構成される。さらにそれぞれの部門にはカテゴリーが存在し、それぞれのカテゴリーが音韻を持ち得る。図に 2.1 その部門およびカテゴリーの構成を示す。

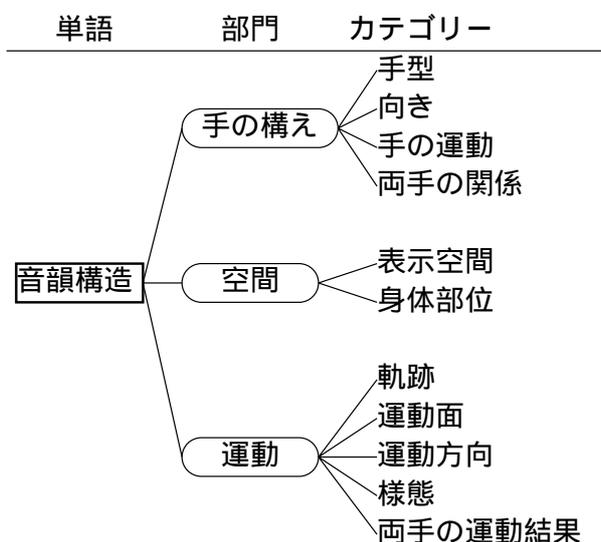


図 2.1: 音韻構造のツリー表現

2.3.3 手話音韻構造の表現

手話は、手の構え、空間、運動の3つの部門から構成されている。手の構えは、手形、向き、手の運動のカテゴリに分けられる。以下にそれら3部門とそれぞれのカテゴリの概要を示す。

- 手の構え（部門）指を主とする手に関する記述がある。静的な手の形を示すカテゴリや、両手に関する運動に関する記述がある。
 - － 手形
手形には48種類の分類が割り当てられている。手形は両手に独立して出現する。それぞれに、指文字や手話を参考に考案された記号が与えられている。
 - － 向き
手の向きは、掌と手首の部位に着目しそれぞれを「掌の向き」と「手首の向き」として記述される。向きは両手に独立して出現する。音韻を表す記号は6種類あり、場合によりそれらを組み合わせて記述される。また、それぞれの部位に関して同じ方向を向いているかを表す記号、逆を向いていることを示す記号がある。
 - － 手の運動
手の運動は、指、手首、肘の部位に着目する。手の運動は両手に独立して出現する。動きには基礎的な音韻として6種類が割り当てられており、特殊な場合として3種類が追加されている。さらに、運動の種類が変化する場合についても記述がある。
 - － 両手の関係（同じ手形の場合）
両手手話の単語の場合、5種類の関係が定義されている。音韻表記としては両手に関して修飾をすると見なせる。
 - － 異左手形
左手の手形と右手の手形が異なる場合を表す。直接異左手形を表す記号があるわけではなく、左手と右手の音韻表記を並列に表すための区切り記号がありそれによって右手、左手の順番に表記するものである。
- 空間（部門）主に手の位置についての記述を行う部門である。大まかな提示場所を示す表示空間や、しばしば単語の意味に密接に関係する身体部位とがある。

- 表示空間

体を基準に手話がどこで提示されるかについて表す。部門「手の構え」のなかの 카테고리「向き」と同じ 6 種類の記号が定義されている。
- 身体部位

表示空間が比較的抽象度が高いのに対し、手話が提示される場所をより具体的な身体の 25 の部位によって記述する。また、その部位からの距離を 4 段階に分類している。
- 運動（部門）

手の運動のに関する記述を行う。運動の軌跡として定義されるものや、その運動がある平面上を移動する場合についての定義がされている。それぞれの 카테고리は 3 ないし 5 程度の音韻をとり得る。

 - 軌道
 - * 線形状

基本となる線形状は直線と曲線である。線形状の記述がない場合（無印）を直線とし、それ以外に曲線がある場合に記号を与えている。その他に特殊な形状について 3 種類について定め、合わせて 5 種類がある。
 - * 運動平面

運動がある平面上を移動しているとみなされる場合に定義される。それぞれの水平面、垂直面に関して 4 種類が定義され、斜面を表す記号 1 つを合わせ計 5 種類によって表される。
 - * 運動方向

運動方向は表示空間と同じ記号を用いて、基本となる「どちらへ移動するか」を表す。特別な場合としてさらに 2 種類の音韻が追加されている。
 - 様態

物理的な「速い」や「遅い」を表すものに加え、運動の程度に関して「大きい」「小さい」、(例えば移動距離が大きいなど)、を表す記号がある。さらに繰り返しがある手話に対する属性も様態に分類されている。
 - 両手の運動結果

通常の手の関係については部門「手の構え」の 카테고리「両手の関係」で記述されているが、さらに運動の結果が変化する場合に関して 3 種類の音韻を定義している。

動作を伴うという観点からは、運動軌跡は運動部門に一番近いことが推察できる。しかし、手話の提示位置に関する空間部門についても、手が提示位置まで移動することを考えれば、運動軌跡に準じた取り扱いが可能である。また手話は、単語がすべての部門とカテゴリを持つのではなく、必要に応じてカテゴリが音韻をとり得る。すなわち、あるカテゴリに着目した場合、すべての単語がそのカテゴリに関して音韻を持つわけではない。複数の手話単語があった場合、カテゴリによってその使用頻度にはバラツキがあることになる。

音韻情報から運動軌跡を取り出す場合、運動軌跡認識に関してはカテゴリ単体では利用は難しい。運動軌跡を行う場合、複数のカテゴリを組み合わせる単語のグループを作る必要がある。

2.4 まとめ

聴覚障害と非音声コミュニケーション手段としての手話について述べた。次に、手話単語に関する取り扱いと音韻表記法を述べ、最後に音韻表記法を構成する要素と運動軌跡との関係について検討した。

第 3 章

手話単語データの入力および編集システムの構築

3.1 はじめに

本章では, 本論文で用いた装置, 並びに作成したシステムについて述べる. 手話話者 (被験者) の動作を記録するために, 本システムでは身体に装着する手形状および位置・姿勢入力装置を用いる. まず, これらの入力装置について述べる. 次に, それら入力装置を用いて手話単語データを採取するために作成したシステムについて述べる. さらに, 入力されたデータを可視化し編集を行うためのシステムについて述べる.

3.2 システムの構成

ハードウェアの構成を, 図 3.1 に示す. ホストコンピュータとして Silicon Graphics 社製 Onyx VTX(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, 手形状入力装置および位置姿勢からなる入力装置とから成る. また, 入力装置は2つの部分から成る. 1つは操作者が装着するセンサ部分であり, もう1つはセンサ部分からのデータの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である.

これらハードウェアを利用しデータの採取および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のモジュールに分けられており, それぞれが独立したプロセスとして実行される. 機能的な概要を図 3.2 に示す. それぞれは, センサ制御装置との通信・制御をおこなうセンサ制御モジュール, データの可視化モジュール, 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である.

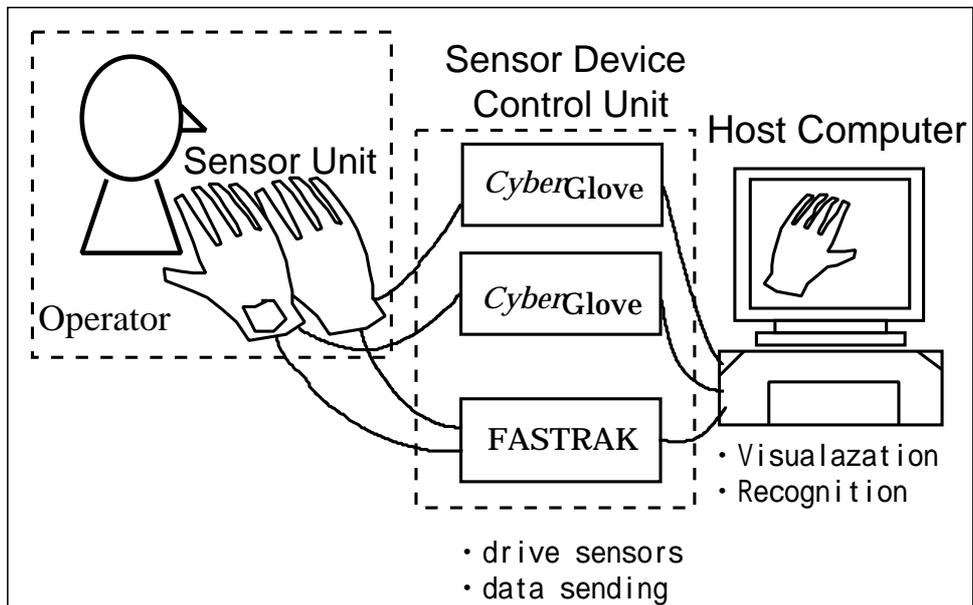


図 3.1: 単語データ入力システム

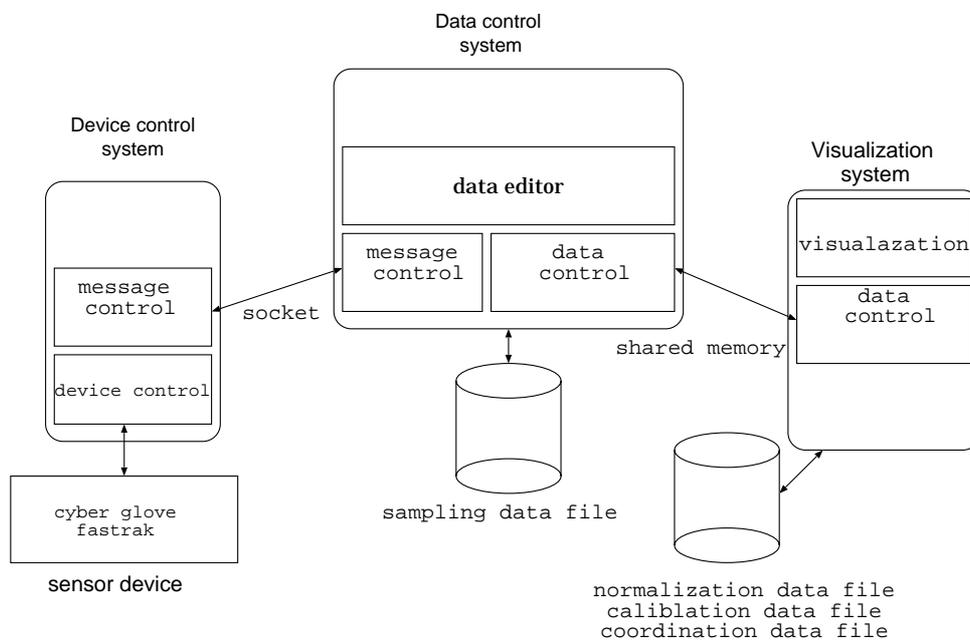


図 3.2: システムモデル

3.3 入力装置

3.3.1 手形状入力装置

本システムでは、手形状入力装置として Virtual Technologies 社製の CyberGlove[11] 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、片手あたり手指の主要 18 関節の屈曲角のデータを取得できる。装置を手に装着した外観を図 3.3 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 では関節角センサに光ファイバが使用されている。関節を曲げる動作に従って光ファイバに曲げが生じる。光ファイバの曲げによって、光ファイバの屈折率が変化しその値を測定することで、関節の曲げ角を検出する。屈折率は曲げ角に関して線形である性質を利用し、センサの出力は指の曲げ角度に対して線形な値である。

関節角センサーは図 3.4 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 1,2 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): (第 1 関節, 第 2 関節)), 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外転したかを測定するセンサがあり (Ordering ID:0), 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 3.4 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。

3.3.2 位置・姿勢入力装置

指文字を認識する際には、手の位置と向いている方向の情報が必要となる。これらの 3 次元の位置・姿勢を測定する装置として、図 3.5 に示す POLHEMUS 社製の FASTRAK[12] 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。得られるデータはトランスミッタに対するレシーバの 3 次元座標 (x,y,z) と回転座標 $(yaw,pitch,roll)$ 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。

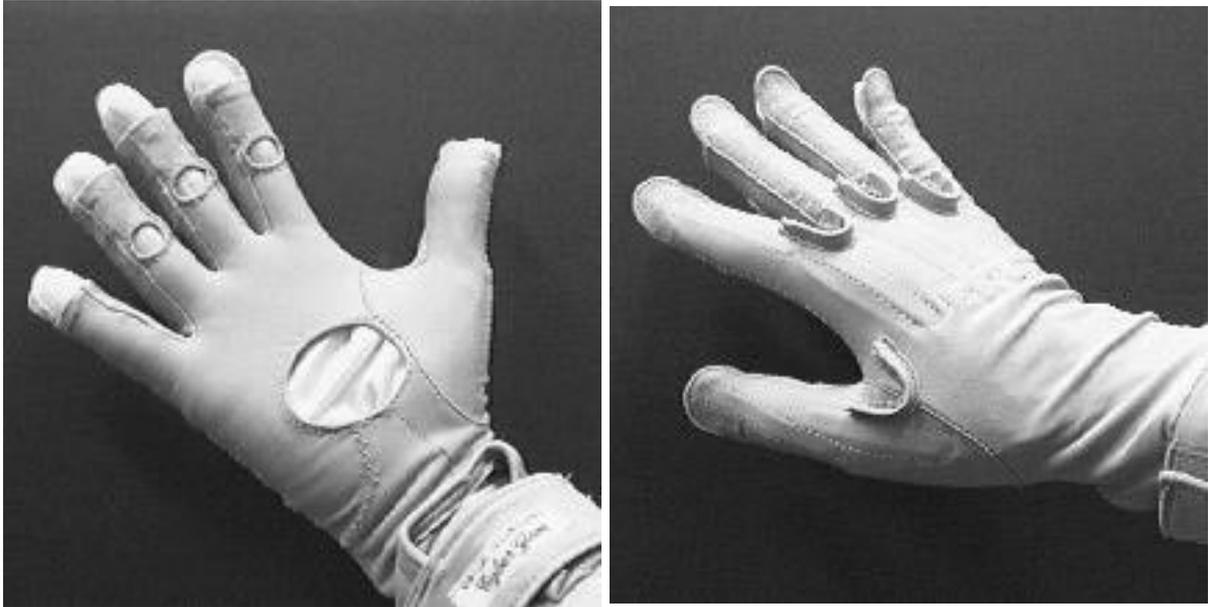


図 3.3: CyberGlove の外観

本システムでは、レシーバを CyberGlove の手首の位置に装着することで、連続的に 3 次元座標のデータを計測を行う。

3.4 入力装置の制御

センサ装置のデータ送信モード

手形状入力装置および位置姿勢入力装置はデータの送信に関して「Request & Send」と「Stream」との 2 種類のモードを持っている。「Request & Send」はセンサ制御装置がホストコンピュータの 1 回のデータリクエストメッセージに対して 1 回のセンサデータの送信を行うモードである。「Stream」は一定の間隔でセンサ制御装置が自律的にセンサデータの送信を行う「Stream」モードである。

「Request & Send」モードの特徴としては、センサへの 1 回の送信要求に対して 1 回のセンサデータ送信であるため、制御が比較的簡単となることである。一定の間隔でデータを採取したい場合には、ホストコンピュータが送信要求を一定の間隔で送出することが必須となる。しかし、ホストコンピュータがシングルプロセスではない場合、すなわちマルチプロセスとして稼働している場合、「Request & Send」モードはホストコンピュータの負荷の具合や送信経路上のデバイスドライバ等のバッファの影響により等間隔の連続データ

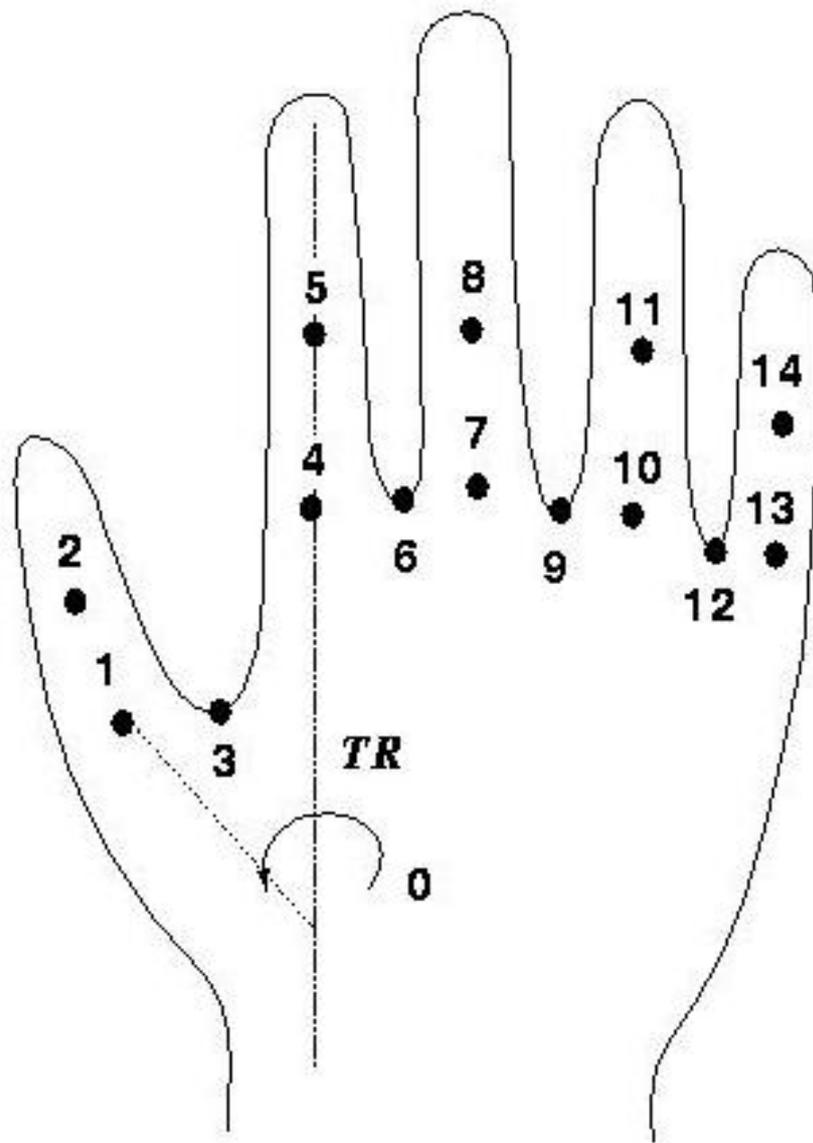


図 3.4: 手形状入力装置の関節角測定点



図 3.5: FASTRAK の概観

サンプリングには適さない。

一方「Stream」モードの特徴は、センサ制御装置内部のクロックに同期してセンサデータの送信が可能なことである。また、通信速度の範囲内でサンプリングレートは可変である。本システムで用いるホストコンピュータはマルチプロセスのシステムであり、手話データの採取においてサンプリングレートを一定にするためには「Stream」モードを使用する必要がある。

センサデバイス制御とデータ欠損

CyberGlove では 1 関節あたりの 256 階調 (8bit) の情報量で表される。これが 18 関節に関してあり、先頭の識別子 2Byte を含めて 20 ワードのデータとなる。一方、Fastrak は 1 次元あたり float 型の精度であり 4Byte のデータ長を必要とするため、6 次元に関して 24Byte のデータ長となる。開始識別子および終了識別子を含めて 29 バイトとなる。

ホストコンピュータとセンサデバイスとは RS-232C によって接続されている。位置姿勢入力装置と手形状入力装置はそれぞれ転送速度が変更可能であるが、用いたホストコンピュータ側の転送速度の上限が 38400bps であるため、その転送速度内でとり得る最高のサンプリングレートとして 60Hz を採用した。

センサデバイスとホストコンピュータ間の通信において、より長いバイト長を必要とする Fastrak の場合、60frame(1 秒) が経過した時点で、1740Byte となる。すなわち、短期間でもバッファからのデータ取り込みが滞った場合、OS 内部のバッファ溢れが生じ、セン

サデバイスから送られてきたデータに関して情報欠損が生じることとなる。定常状態では、ホストコンピュータは十分な速さでデータをスキャンし、バッファ溢れは生じない。しかし、他のプロセスが入出力制御や CPU を占有した場合などは情報欠損が発生することになる。これらのはバッファ溢れの外的な要因である。また、アプリケーションプログラム自体が他のシーケンス（ファイル入出力、ユーザインタフェース処理）をすることを考えた場合、これらの処理によってもバッファ溢れが生じる。これらは内因的な情報欠損の原因といえる。

そこで、内因的なデータ欠損を回避するために、ファイル入出力などのデータ制御をおこなう部分を分離した。手話単語の記録のためにはソフトウェア的に、システムをセンサ装置の制御を行う部分と、データを記録する部分の2つが必要となる。

以下の特徴を持つ位置・姿勢入力装置の制御モジュールとして実装した。

- デバイスの初期化および終了処理
- バッファ溢れ時の復旧
- 3 台のセンサデバイス装置の同期
- socket 経由の独自プロトコルコマンドの受送信と処理

3.5 手話単語データの可視化とデータ受送信

データ制御部におけるデータの受送信

データ制御モジュールが行うデータの受送信は目的に応じて2つの経路がある。1つはセンサ制御モジュールからのデータ送信、もう1つはデータ可視化モジュールへのデータ送信である。図 3.6 にその流れを示す。

2 種類 3 台のセンサ装置から送信されたセンサデータは、センサ制御モジュールによってタイミングを 60Hz のデータとして調整され、1 回のサンプリングデータとして両手で 48 次元のデータとして扱う。これを 1 フレームのデータとする。

センサ制御モジュールからデータ制御モジュールへのデータ転送においてはデータを 1 フレームのセンサデータの集まりとする。1 フレームのデータはヘッダ情報を付加した 98Byte のデータとして送信する。センサ制御モジュールからデータ制御モジュールの間では、制御のための通信は双方向で行われるが、センサデータはデータ制御モジュールへの一方通行である。データ制御モジュールへ搬出されるセンサデータは、時系列に沿って送られる必要があり、十分なバッファ能力が必要であるため socket を用いて受送信を行う。

データ制御モジュールと可視化モジュールとのデータの受渡しは、データ制御モジュールとセンサ制御部とのデータ受渡しの場合とは異なったアプローチが必要になる。センサ制御部からのデータはすべてを取りこぼすことなく、しかも送信された順番通りに受信しなければならないという条件がある。一方、可視化部では描画速度以上のデータは必要ではなく、たとえデータの送信が滞ってもセンサ制御部からのデータ受信と比較してその影響は少ない。そこで、可視化部との情報交換は共有メモリを使用する。

データ制御部はセンサ制御部からのデータのうち、最近の1フレームだけを共有メモリに書き込む。可視化部では描画速度に準じた速度で共有メモリを読みとることで、無駄なデータの送信が減り、しかも複雑な同期を不要とした。

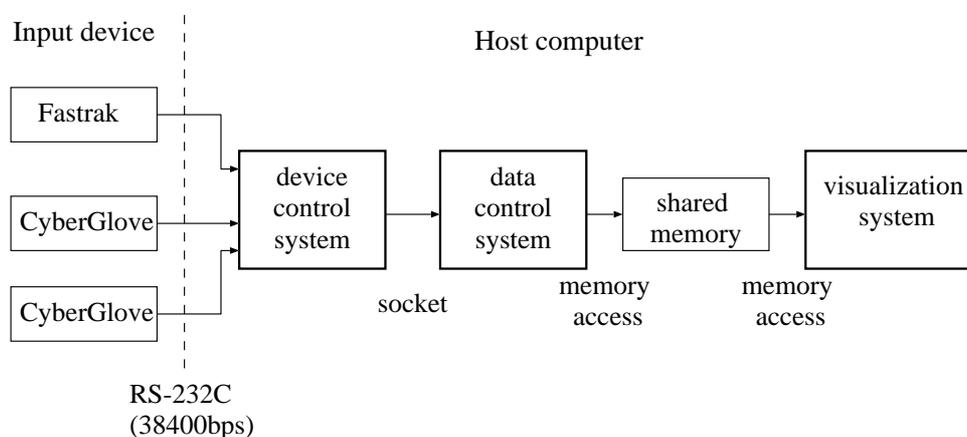


図 3.6: 手話単語データの流れ

3.5.1 手話単語データの可視化

「Request & Send」モードを用い GUI 処理を含めた描画処理を単一プロセスで行った場合、最速でも1秒間に15回の描画に留まる。描画とセンサ制御のプロセスを分割することで、描画速度が1秒間40回前後とその反応性が高まった。手話単語データの可視化によって単語データを採取しながらのデータ確認が可能になった。

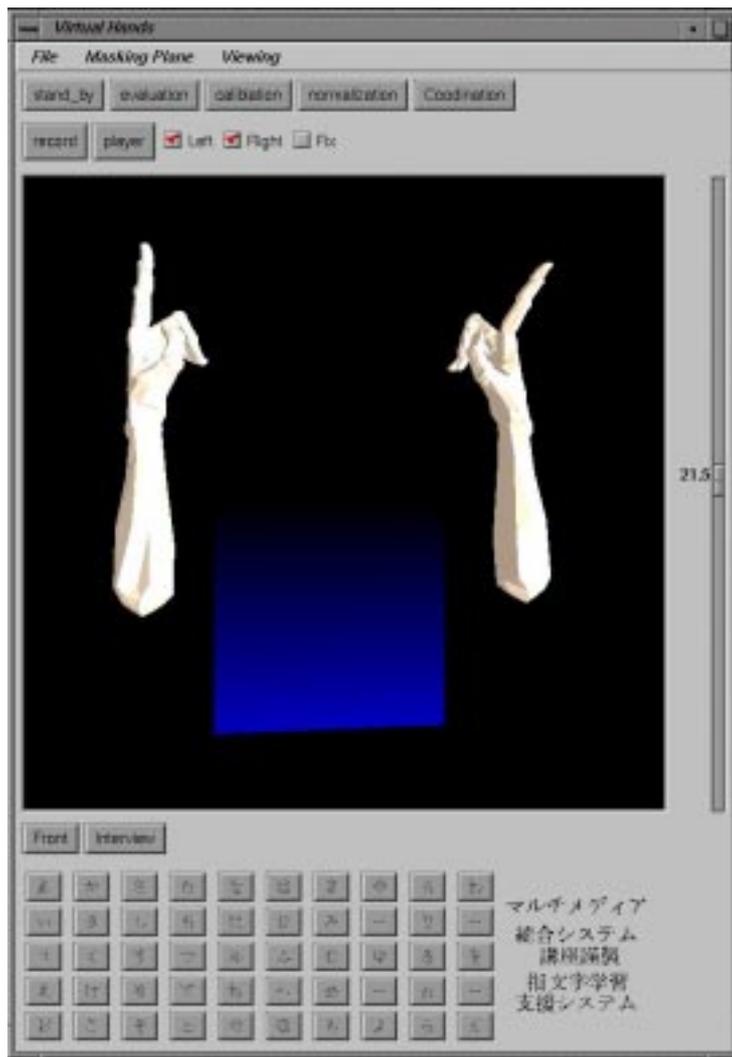


図 3.7: 手話データ可視化画面

3.6 装置の初期調整

3.6.1 手形状入力装置の調整

人間の手は、その大きさや関節の曲がり具合などに関して個人差が存在する。手形状入力装置を用いる場合、操作者ごとの手の動作特性の違いや、装着時におけるセンサーのずれなどをに対処するため、被験者は計測前にキャリブレーションを行う必要がある。手形状入力装置の間接角の最大値、最小値を記録する。具体的なキャリブレーションは以下の手順で行う。

1. 5指全部を反らす。
2. 親指以外の指を曲げる。
3. 親指だけを曲げる。
4. 5指全部を横で密着させる。
5. 5指全部を外側に開く。
6. 親指と小指を接触させる。
7. 手首を前後左右に屈折させる触させる。

これによって、被検者個人の5指全て関節角、および各指間の外転センサー出力の最大値、最小値を得る。センサーから出力される値を $Cmax_i, Cmin_i (i = 0, 1, 2, \dots, 14)$, と表す。

屈伸率

各関節の測定値を正規化し、関節角を指が曲ってる割合を屈伸率を用いて表す。

入力値を $input$ とした時、各関節の屈伸率 $bend_i$ は次式で求めることができる。

$$bend_i = \frac{input_i - Cmin_i}{Cmax_i - Cmin_i} \quad (i = 0, 1, 2, \dots, 14) \quad (3.1)$$

認識の際には手形状入力装置から得られる測定データは屈伸率で表現し比較を行う。実際のシステム上では図 3.8 に示すようなコントロールパネルを生成する。最大値と最小値を観測しその差 (レンジ) がある閾値より小さい場合は正規化が不十分であるとして、警告を行う。センサーデータとして記録するのに十分なレンジが得られれば図 3.9 に示すような画面となり、データ採取における正規化が終了する。

3.6.2 手形状入力装置による関節角の情報収集

関節角に関しては正規化情報の採取の他に、装置の調整 (calibration) の作業が必要となる。センサ出力の値は線形性が保証されているのみで、物理単位量としての意味は持たな

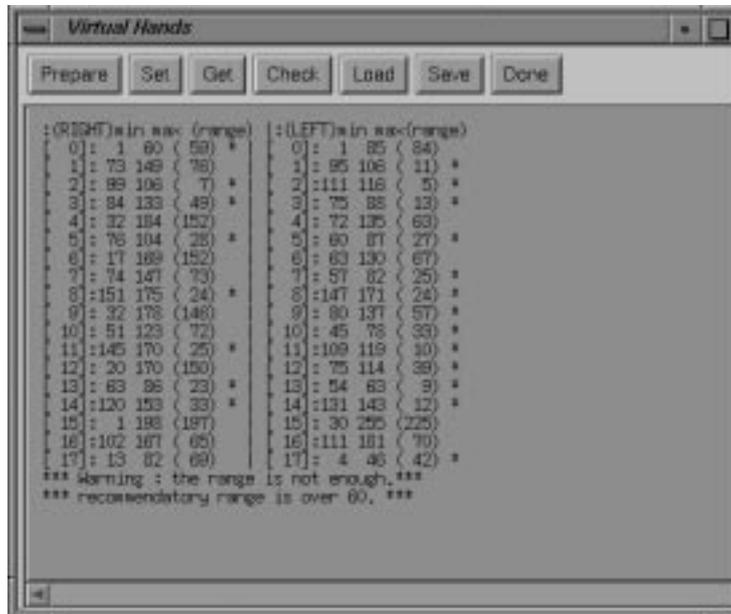


図 3.8: 観測値が不完全である場合の警告の正規化画面

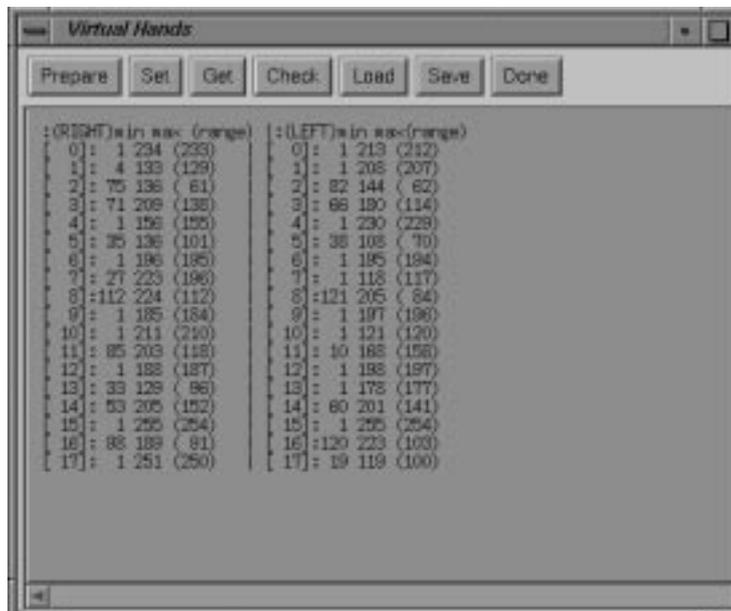


図 3.9: 観測値が要件を満たした場合の正規化画面

い. そこで, センサ出力から角度の物理的な単位への変換するパラメータを決定することが必要になる. 正規化作業では各関節角のセンサ出力の最大最小値のみに注目していたが, 手形状入力装置の調整は比較可能な物理単位 (ここでは Degree 表現) に対応変換するパラメータの決定を行う. 調整するパラメータはゲインとオフセットの 2 種類で, 実際の手の関節をある角度に固定し, 次にパラメータを固定した角度と同じになるように調整する. 図 3.10 に 1 関節に対するインタフェース要素を示す. またこの際に, CG による可視化された画面によって作業の目視が可能である.

調整作業のためのコントロールパネルを図 3.11 に示す. 18 関節に対して, 任意にオフセットとゲインが可能である. また, 現在のセンサ出力値とパラメータによる変換値を表示している. 左右の手に関しては任意に切替え可能である.

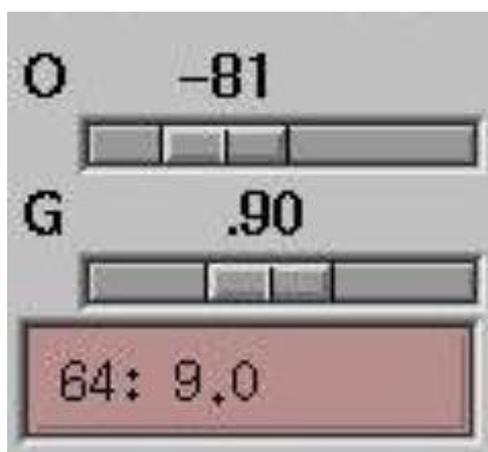


図 3.10: オフセットとゲインの関節角調整画面

3.6.3 位置・姿勢入力装置による身体部位の情報収集

位置・姿勢入力装置はホストコンピュータへのデータ形式が, 位置に関してはセンチメートルを単位に, 姿勢に関しては Degree 単位であり, 手形状入力装置のような値変換の必要性はない. ここでは, 被験者の身体的特徴や, 単語開始位置の記録や, 基準装置と座る位置とのバラつきを補正するための情報の取得について述べる.

両手の位置センサに注目し, それぞれを以下の対応する被験者の身体部位に置き, その位置情報を保持する. 表 3.1 に採取身体位置を示す. 初めの 6 項目に関しては体の各部位にセンサを接触させることで行う. 膝の位置情報は 5.2.2 章で述べる単語の開始および終了位置と同等の姿勢を採取する. 最後の 3 項目は肘を伸ばし肩を基準にその位置までの線分が

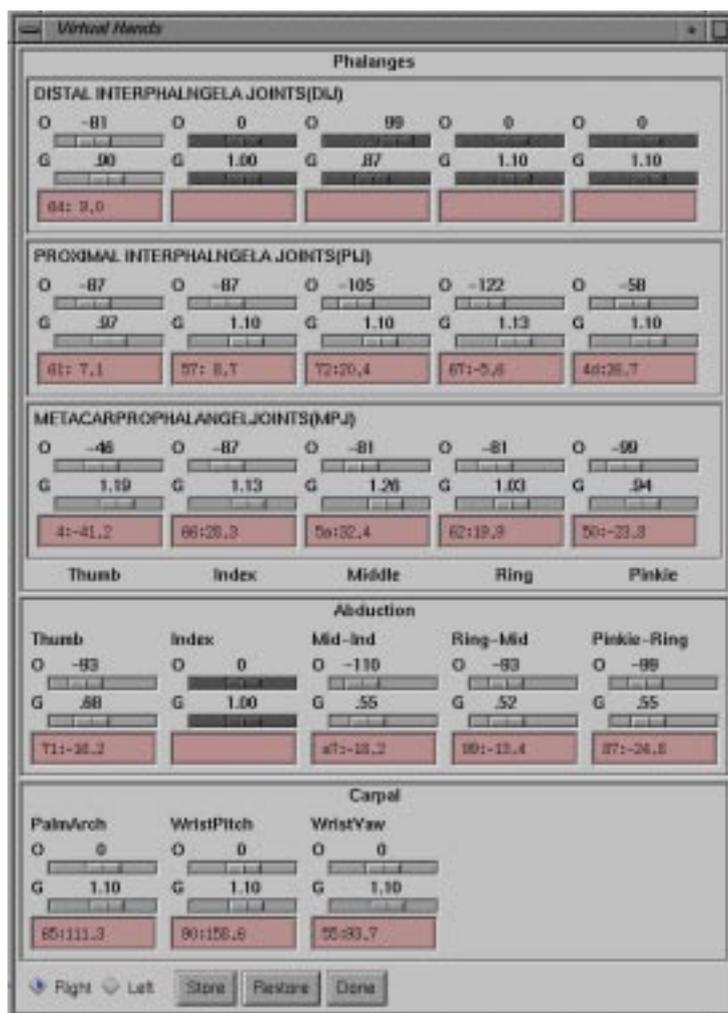


図 3.11: 関節角調整画面全体

それぞれ直行するように位置を採取する。身体部位の情報収集に関しては図 3.13 に示すコントロールパネルで制御し、その結果は現在の腕位置と併せて画面上に可視化される。可視化の実例を図 3.13 に示す。これにより、データ採取中の体軸のずれなどを目視により確認可能となる。

表 3.1: 身体位置情報収集の採取位置

1. 頭部頂点
2. 額中央部
3. 鼻頂点
4. 顎先端 (掌を前方に向ける)
5. 喉 (喉ぼとけ付近)
6. 肩
7. 臍
8. 膝 (単語開始終了位置姿勢と同等)
9. 肩から水平側方
10. 肩から水平前方
11. 肩から垂直上方

3.6.4 データの編集加工

第 3.4 節では、センサ制御装置とセンサ制御装置をソフトウェアから制御するモジュールについて述べた。本章ではそれらセンサ制御モジュールを、ユーザの制御下に置くためのデータ制御モジュールについて述べる。データ制御モジュールとセンサ制御モジュールとを分割した理由の一つとして、センサ制御モジュールが直接ユーザインタフェースを行った場合、その端末を経由した操作者の行った入出力自体がセンサデータの制御処理を圧迫することがある。

データ制御モジュールはグラフィカルユーザインタフェース (GUI) を持ったシステムとして実装した。センサ制御モジュールを含めた他のモジュールとのデータの交換、データ

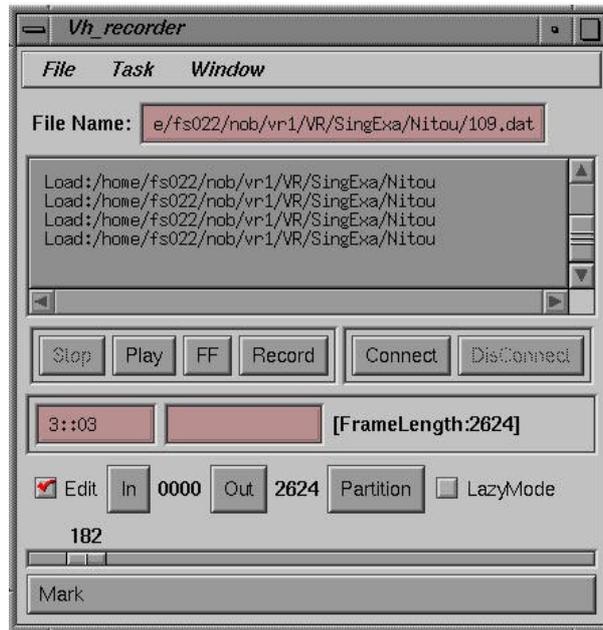


图 3.12: 身体位置收集画面

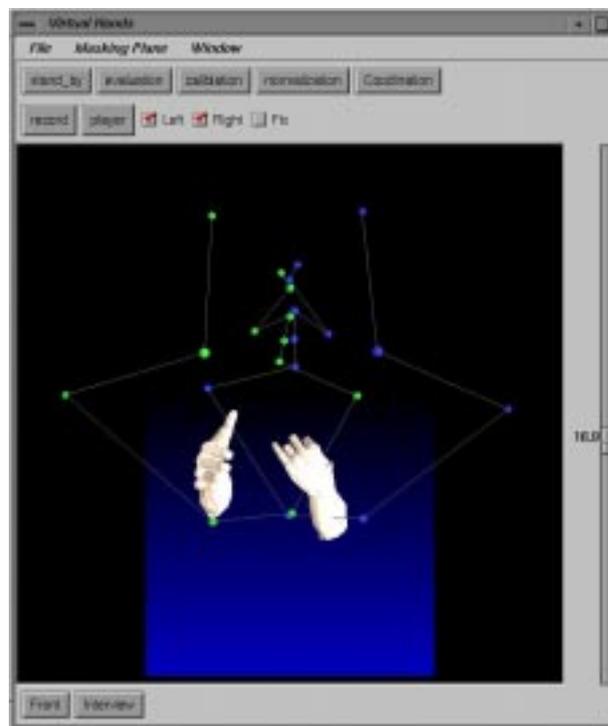


图 3.13: 身体位置情報可視化画面

のファイルとの入出力, その他ユーザが行う編集指示の対応と作業内容の実行を行い, 円滑なデータサンプリング作業の環境を提供する. 図 3.14 その表示画面を示す. データ制御モジュールは以下のような機能を持つよう実装した.

- ファイルへのデータの入出力操作
- センサ制御モジュールへの送信要求・送信停止要求
- センサ制御モジュールからのデータの受信
- 被験者を用いたサンプリングの場合の, 各種インターフェイス
- データの手動分割のためのインターフェイス

3.6.5 手話単語記録時のインターフェイス

手話単語の採取の際に記録単語の日本語ラベル表示などを行うインターフェイスを作成した. このユーザインターフェイスは, 現在対象としている単語を設定し, 単語に対応する, 日本語ラベルの表示, データの保存時のファイル名の決定, およびデータ保存の後に自動的に対象単語を次の単語へ変更する. これにより, 単語データ入力時の便宜を図る. 図 3.15にその画面を示す.

3.6.6 手話単語の手動切り出し

採取した手話単語のデータは必要に応じて使用部分の切り出しが必要になる. これは手動で行っている. データの切り出し作業を可能にする編集機能を含んだ単語切り出しのための機能およびインターフェイスを作成した. 作成したインターフェイスを 図 3.16に示す.

手話データ採取においては同一単語に関して複数回の提示を行っており, この時点のデータファイルには同一単語が複数回記録されている. 認識実験では1ファイルに対して1語のデータである必要があり, 現在は目視確認による手動切り出しを行っている.

単語切り出し編集では, 操作者が参照位置, 参照位置からの距離と速度の閾値の3種類の条件を与えることで, 複数の単語パターンから1語の切り出しを行う機能を持つ. いくつかのデータでは正しい切り出しが行えることを確認しており, 今後の自動切り出しへの応用が考えられる.



図 3.14: データ制御インターフェイス

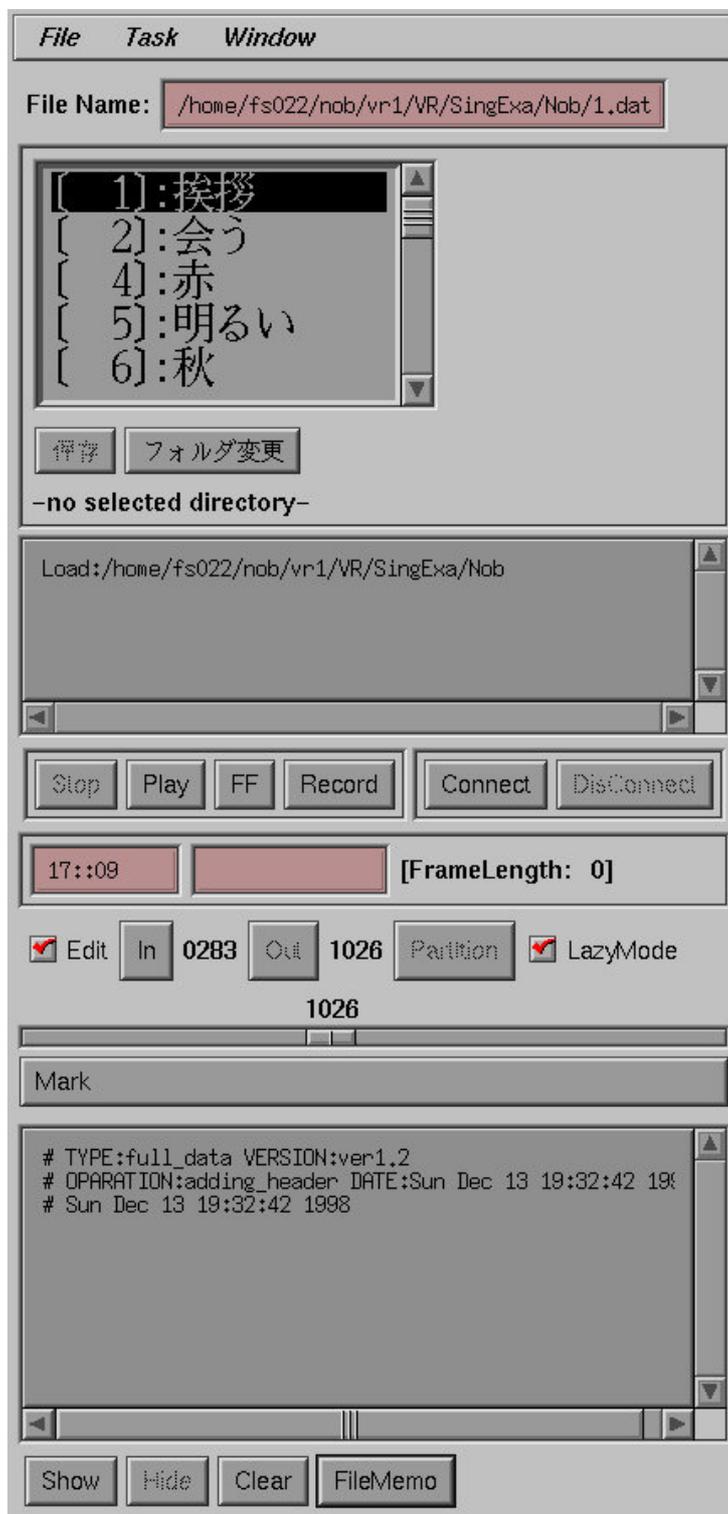


図 3.15: 手話単語採取インターフェイス

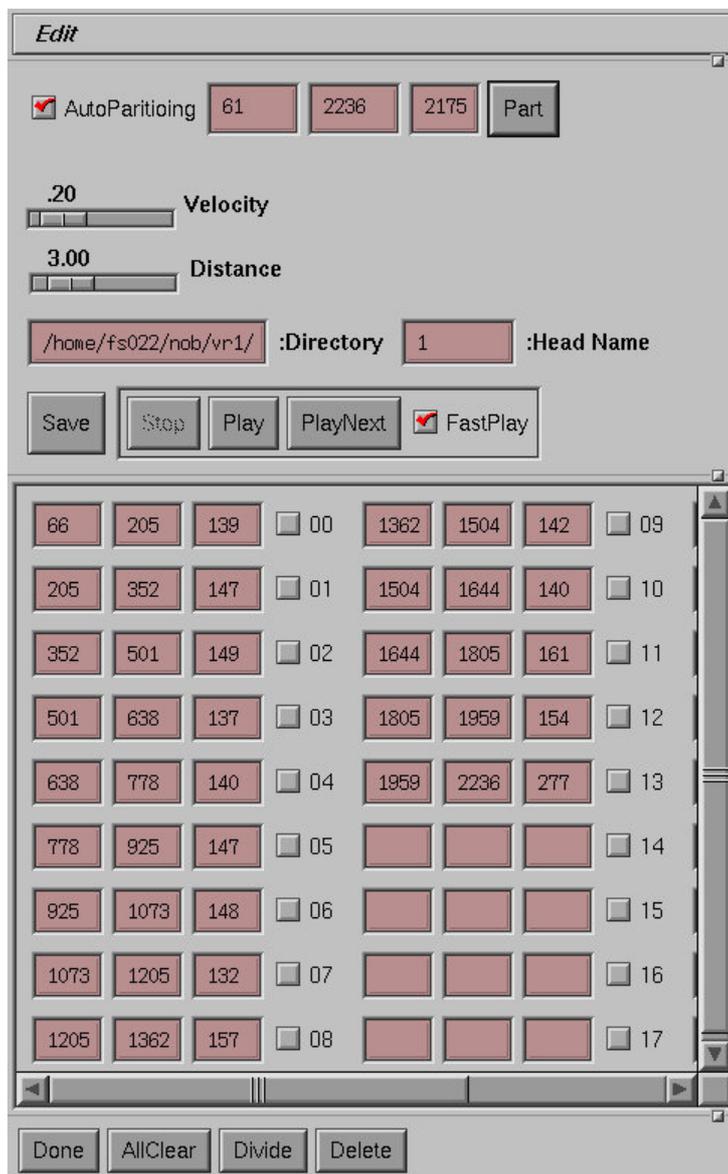


図 3.16: 単語切り出し編集用インターフェイス

3.7 まとめ

手話単語のデータ入力のためのセンサ装置および構成システムについて述べた。さらに、それら装置を手話単語データの入力装置として使用するために作成したシステムについて述べた。これらにおいては装置のキャリブレーションを行うためにも使用される。次に、センサデータの取得方法とデータの可視化と編集を行うためのシステムについて述べた。

第 4 章

手話単語および運動軌跡の認識手法

4.1 はじめに

手話単語認識は特定話者認識に関しては、従来研究で DP マッチングを用いた非常に認識率の高い手法が提案されている。手話単語認識に関して、次に必要となるのは複数話者への対応である。本章では、DP マッチングを用いた認識手法と、調波解析を用いた手話単語認識手法について述べる。それぞれの手法において、複数話者の手話単語認識に用いるということは、単語パターンを複数話者から採取し、それらより単語に関する辞書を作成することを意味する。

手話単語を認識する手法として、DP マッチングを用いた手法の説明と、調波成分を用いた手法の提案を行った後、調波成分を用いて手話の運動軌跡を認識する手法について述べる。

4.2 DP マッチング

音声認識などの時間の概念があるパターンについて、認識で用いられる手法として DP マッチングがある。個人を対象とした手話単語認識に関して非常に高い認識率が得られている。

DP マッチングは、まず動的計画法 (Dynamic programming) に基づき 2 つのパターン間の距離を定義する。この定義を用いてテストパターンに対して、複数の辞書パターンから一番距離が近いパターンを認識パターンとする方法である。

長さがそれぞれ I, J のパターン A, B を特徴ベクトルの時系列として次のように表現する。図 4.1 に DP マッチングのモデル図を示す。

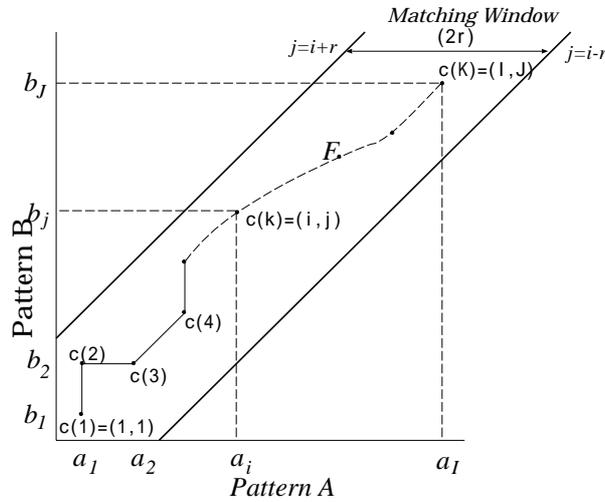


図 4.1: DP マッチング

$$\left. \begin{aligned} A &= a_0, a_1, a_2, \dots, a_i \\ B &= b_0, b_1, b_2, \dots, b_j \end{aligned} \right\} \quad (4.1)$$

パターン A の i 番目の特徴ベクトル a_i と、パターン B の j 番目の特徴ベクトル b_j との対応づけを $c = (i, j)$ と表す. $i-j$ 平面上で $c = (i, j)$ が格子点となって表れる. ここで, それぞれのパターンの始点から終点までを K 個の対応付けを行った場合, $c(1) = (1, 1), c(K) = (I, J)$ となり, この系列を時間変換関数 F とするとき, 以下のように表すことができる.

$$F = c(1), c(2), c(3), \dots, c(K) \quad (4.2)$$

時間変換関数 F は $i-j$ 平面上で $(1, 1)$ から (I, J) の各格子点を結ぶ 1 本の線である. 2 つの特徴ベクトル a_i と b_j の距離を $d(i, j)$ と定義する.

さらに系列 F に沿って定義された関数によって特徴ベクトル間の距離を求め, その加重平均をとる. その値 $E(F)$ を, 格子点の系列 F を用いた場合のパターン A, B 間の距離とする.

$$E(F) = \frac{\sum_{k=1}^K d(c(k))}{\sum_{k=1}^K w(k)} \quad (4.3)$$

$w(k)$ は重み係数である。系列 F は $c(1) = (1, 1)$ から $c(K) = (I, J)$ までの様々な経路をとりうるが、その中で $E(F)$ を最小にするような対応づけ F を、パターン A, B 間の距離 D とする。

$$D(A, B) = \min_F \{E(F)\} \quad (4.4)$$

この D を最小にするような対応づけを動的計画法によって求める。変換関数 F には次のような制限を設ける。

$$\left. \begin{array}{ll} (1) \text{ 単調性} & i(k-1) \leq i(k), j(k-1) \leq j(k) \\ (2) \text{ 連続性} & i(k) - i(k-1) \leq 1, j(k) - j(k-1) \leq 1 \\ (3) \text{ 境界条件} & i(1) = 1, j(1) = 1, i(K) = I, j(K) = J \\ (4) \text{ 整合窓の条件} & |i(k) - j(k)| \leq r \end{array} \right\} \quad (4.5)$$

ここで整合窓の条件とは、時間軸において極端な対応づけをしないための制限である。この条件は幅 $2r$ で F の対応づけを制限する。

上記の制限のもとで、 $E(F)$ を最小にするための $w(k)$ を定める必要がある。これを定義することにより、最小化の対象となる目的関数が加法的となり動的計画法に基づき演算可能な問題となる。 $w(k)$ には幾つかのものが提案されているが以下で示されるように i, j について対称な以下の重みづけを用いる。これを傾斜制限として図示したものを図 4.2 に示す。傾斜制限を当てはめることにより動的計画法に基づく漸化式の作成が可能となる。

$$w(k) = (i(k) - i(k-1)) + (j(k) - j(k-1)) \quad (4.6)$$

$c(1)$ から $c(k)$ でのうち、ある部分系列 $c(1), \dots, c(k)$ について以下の式を定義する。

$$\begin{aligned} q(c(k)) &= g(i, j) \\ &= \min_{c(1), \dots, c(k-1)} \{ \sum_{l=1}^{k-1} d(c(l)) \cdot w(l) \} \end{aligned} \quad (4.7)$$

さらに、この式を変形することで以下の式が得られる。

$$\begin{aligned} q(c(k)) &= \min_{c(1), \dots, c(k-1)} \{ \sum_{l=1}^{k-1} d(c(l)) \cdot w(l) \} \\ &= \min_{(c(k-1))} \{ g(c(k-1)) + d(c(k)) \cdot w(k), \dots, c(k-1) \} \end{aligned} \quad (4.8)$$

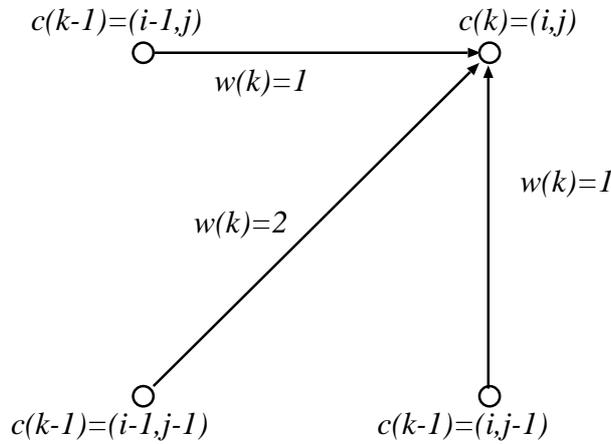


図 4.2: DP マッチングの傾斜制限

$c(k)$ は式 (4.6) から, 3 通りが候補となる. 上式では 3 通りの $c(k-1)$ のうち $g(c(k-1))$ と $d(c(k-))$ の和が最小となるものを選び, その値を次の部分問題 $g(c(k))$ とすればよいことになる.

これらをまとめると以下の漸化式となる

$$g(i, j) = \min \begin{cases} g(i, j-1) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i-1, j) + d(i, j) \end{cases} \quad (4.9)$$

初期条件を $g(1, 1) = 2d(1, 1)$ とし, $i = 1 \sim I, j = 1 \sim J$ について式 (4.9) を計算すれば以下の式よりパターン A, B の距離が定義できる.

$$D(A, B) = \frac{I+J}{1} g(I, J) \quad (4.10)$$

4.2.1 中心近接尺度と辞書データの作成

DP マッチングにおいては認識の際に, 辞書データを必要とする. 本来, DP マッチングによる認識手法は, あらかじめ採取したデータとの距離を算出する以上の意味を持たない. 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を導入し, 中心近接尺度より辞書データの決定を行う. 中心近接尺度の着目点は次のようになる.

1. 複数の辞書候補となるパターンを得る.

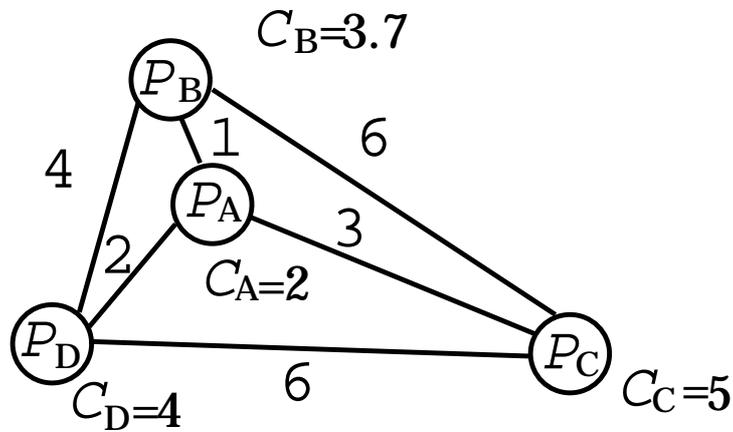


図 4.3: パターンの距離関係例

2. 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より中心に近いパターンを辞書パターンとする.

距離の算出関数としては以下の条件を満たす必要がある.

- $D(P_n, P_i)|_{i=n} = 0$
- $D(P_i, P_j) = D(P_j, P_i)$

中心近接尺度の算出とその利用の手順を以下に示す.

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.
2. パターン P_a, P_b の距離を算出する $D(P_a, P_b)$ 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に $D(P_a, P_b)$ DP マッチングを用いる. ここで, 最小の C を持つパターンは標準に一番近いこといえる. また最大の C を持つパターンは標準から一番遠いパターンである. その例を図 4.3 に示す.

DP マッチングの問題点として、標本と標本の距離を算出する手法ではあるが標本としての分散の考慮には至っていないことがある。また、DP マッチングには整合窓の大きさというパラメータについて一般化が難しい問題がある。整合窓の目的は標本の時系列要素の対応づけで時系列上で極端な対応づけを制限するものであるが、一方では DP マッチングの柔軟性を制限する。これらはトレードオフの関係にあり、しかもそれは比較されるパターンに依存する。とくに不特定話者認識ではその傾向が顕著である。その他にも、ベクトル演算という性質上サンプル数や整合窓の増加は演算時間に関して指数関数的な増加となって現れる。

4.3 FFT による認識手法

複数話者に対する手話単語認識に関する DP マッチングの持つ諸問題の解決を目的に、離散フーリエ変換を用いた調波成分による認識手法を提案する。

DP マッチングは特定話者のような同一パターンと認識されるべき単語のバラツキが少ない場合はその効果を発揮するが、対応させるべき要素が一意に決めることが難しく、その単語がどの程度のバラツキがあるかの考慮は困難であった。

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、時間領域でこれらを行うのは困難である。そこで、周波数領域での比較を検討する。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を正規化し、それに対して FFT を行うことで、周波数領域での取り扱いを容易にすることとする。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 によって位相成分が得ることで時間的な揺れが吸収可能となる。これらの処理によりパターンの分散を考慮可能となる。

4.3.1 離散フーリエ変換と FFT

時間関数と時間関数をフーリエ変換を用いて変換を行ったのちの関数表現、および離散時間関数から周波成分を求める方法である高速フーリエ変換 (FFT) について述べる。

$x(t)$ があった場合、フーリエ級数展開の複素数関数による表現は以下のようになる。

$$x(t) = \frac{1}{T} \sum_{n=-\infty}^{\infty} X_n \exp(j2\pi \frac{n}{T}t) \quad (4.11)$$

時間 t の変数 $x(t)$ が領域 $-T/2 \leq Y/2$ 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.

$$X_n = \int_{-T/2}^{T/2} x(t) \exp(-j2\pi \frac{N}{T} t) dt \quad (4.12)$$

入力されるデータを離散データとした場合, 有限長フレームデータのそれぞれの要素は, 有限長の離散データと見なすことができる. スペクトルのサンプル値系列が $2f_x$ に相当するサンプル数 N を周期とする周期関数であることに注目すれば, 積和の k の変域を $k = 0$ から $N - 1$ にしてもよい. よって,

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 $2f_x$ の周期で同じ波形を繰り返す時間関数とそのスペクトルとの間の変換対として以下の式が得られる.

$$X_k = \text{DFT}(x_n) = \sum_{n=0}^{N-1} x_n \exp(-j2\pi \frac{N}{kn}) \quad (4.13)$$

$$x_n = \text{IDFT}(X_k) = \frac{1}{N} \sum_{k=0}^{N-1} X_k \exp(j2\pi \frac{N}{kn}) \quad (4.14)$$

DFT に用いる複素数指数関数を以下のようにする.

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

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 DFT および IDFT の式を書き直すと次のようになる.

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

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

高速フーリエ変換とは 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. とくに M が 2 の冪乗である場合に最も効率がよいアルゴリズムが知られており, 直接計算すれば M^2 回の複素計算を必要とするのに対して, $M \log_2 M$ ですむことが知られている. FFT のアルゴリズムの詳細は他に譲ものとする.

時系列データを変換し周波数領域で比較するためには通常, 時間領域で同じ長さの区間を区切り FFT を行うこととなる. 一方, 手話単語データは非線形な伸縮を伴うがゆえに,

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 の冪乗の値が望ましい。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。

特徴ベクトルの要素数が 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。

4.3.2 辞書パターンの作成

DP マッチングの場合と異なり、周波数領域でパターンの比較を行うことになる。DP マッチングの場合、比較する要素を漸化式を用いて値を最小にするような試行演算が必要であった。線形補間による時系列要素の正規化と、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とが可能になる。これによって対応させる要素を一意に決めることが可能となる。要素 n, m の値を $s_{n,m}$ とするとき、の平均値 $MD_{n,m}$ と標準偏差 $SD_{n,m}$ はそれぞれ以下のように求めることができる。辞書パターンの作成は図 4.4 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 (語パターン) から生成される。認識率の評価時には、認識されるべきテストパターンの単語ラベルと認識された辞書パターンの単語ラベルが等しいかで評価する。

$$MD_{n,m} = \bar{x}_{n,m} = \frac{1}{I} \sum_{i=1}^I s_{n,m} \quad (4.15)$$

$$SD_{n,m} = \hat{x}_{n,m} = \frac{1}{I} \sum_{i=1}^I \sqrt{(s_{n,m} - \bar{x}_{n,m})^2} \quad (4.16)$$

4.3.3 誤り率

辞書パターンとテストパターンの距離の算出には以下に示すような、重み付き距離を用いる。

単語 l に関して、辞書として持っている要素の平均値 $\bar{x}_{n,m}^l$ とし、入力された $x_{n,m}$ との距離をそれぞれの標準偏差 $s_{n,m}^l$ によって正規化することで、 D^l を求める。 D を最小にするような l を認識単語とする。ここで、 N は特徴ベクトルの次元数を示し、 M は 0 次高調波からの帯域幅数を表す。

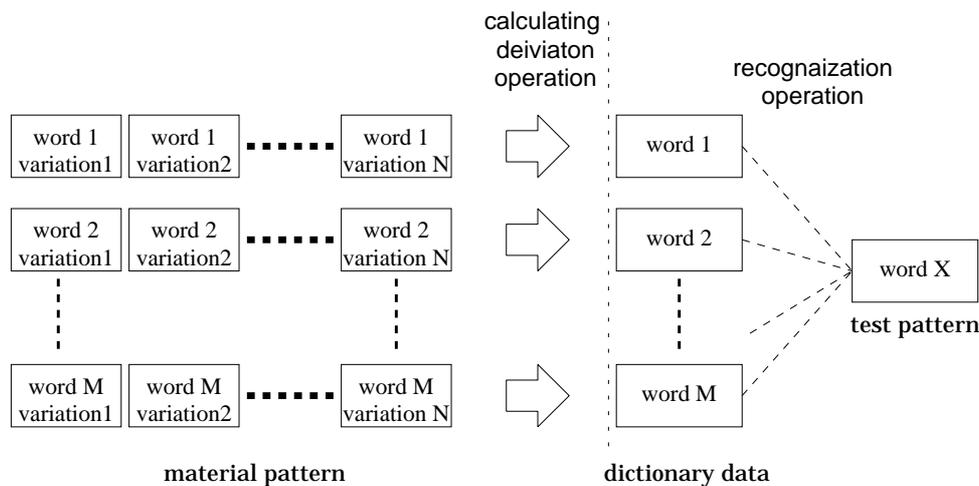


図 4.4: 単語対単語認識時の辞書作成

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

$$D^i = \min_l D^l (i \text{ を認識単語とする})$$

4.4 運動軌跡の認識手法

前節の第 4.3.2 節では、単語と単語のクラスで行う認識に述べた。本節では、同じ音韻要素を持つ入力単語を同じ音韻要素のカテゴリーへ認識を行う手法について述べる。

4.4.1 既得音韻構造情報からの情報抽出

まず、手話には言語学的に音韻論という形で体系付けられた音韻が存在する。音韻要素の機械可読な記号への対応付けと、計算機可読な形式の音韻構造を表現する手段が提案されており、これに基づいて手話単語に対して音韻を表現した音韻構造表現が公開されている。これを既得音韻情報とする。

本論文では、その既得音韻情報の分析を行うため、音韻構造表現を計算機内部表現に変換するシステムを作成した。内部表現は音韻構造表現に沿ったツリー構造となる。図 4.5 に実際の音韻構造表現を構文解析し内部表現に変換した後、その内部表現を表示したもの

をツリー構造として示す。これらを、単語同士の距離定義や単語のグループ分けに際して用いる。

既得音韻構造情報から単語間の距離を定義する場合は、次のような手順で行う。

1. 部門の数を DN とする。
2. カテゴリーの総数を CN とする。
3. カテゴリー c が幾つの音韻要素を採り得るかを $en(c)$ として求める。
4. すべてのカテゴリーにおいて取り得る音韻要素数の総数を $EN = \sum_{c=1}^{CN} en(c)$ とする。
5. すべての音韻要素の $e(c)$ の有無を表すことのできる状態変数、すなわち桁数 EN の 2 進表現値を、音韻要素表現変数 s と定める。
6. すべてのカテゴリーの音韻 ce の音韻要素表現変数 s における位置を $P(ce)$ とする。
7. 音韻要素表現変数 s に対して、手続き $P(ce)$ によって音韻が含まれれば真、含まれなければ偽とし、それぞれを 1,0 として置き換える。

以上の手続きにより、既得音韻構造情報から単語 W の音韻要素表現変数 $s(W)$ を求めることができる。 $P(ce)$ を参考に 2 単語間で該当する音韻要素に関して、排他論理和を求め、それらを積算することで音韻情報による単語間の距離が定義できる。同じ構造であればその距離は 0 となり、異なる要素が多ければそれに応じて距離が長くなる。また、マスクを用いることで、特定の部門やカテゴリーに関しての距離が定義できるようになる。

これらの方法によって求められた既得音韻構造情報の分析結果を、図 4.1 に音韻要素の種類と出現回数を 660 単語に関して計数したものを示す。ここで該当音韻数はその音韻カテゴリーが取り得る音韻の種類を示す。出現回数は該当音韻カテゴリー音韻が出現した個数を表す。出現単語数は該当音韻カテゴリー音韻を 1 つでも含むような単語の数を示す。

対象とした単語数が 660 単語であること出現単語数から、手形はすべての単語に関して存在する。また手形は出現回数は総単語数より多い。これは 1 つの単語が、両手、あるいは変化することで複数の手型をもつことを示す。また、音韻表記で定義されている「運動の軌跡」を含む単語は 660 単語中 56 語であることなどがわかる。

4.4.2 音韻抽出による辞書作成

既得音韻情報から得られた情報から同じ音韻を持つグループにわけ、同一グループの集合から単語認識と同様に標準辞書を作成する。

```

{TYPE:Departments
  (TYPE:=Hand_Pose_Term
    (TYPE:=Finger_Shape_and_Direction
      (TYPE:=Finger_shape: ITEM:=1)
      (TYPE:=HandDirection
        (PALM:=
          (TYPE:=Palm: ITEM:=c)
        )
      )
    )
  (TYPE:=Hand_Relation: ITEM:=idn)
) ;End of Hand_Pose
(TYPE:=Motion
  (TYPE:=Right_Left
    (RIGHT
      (TYPE:=MotionPlaneJoint
        (DIRECTION:=
          (TYPE:=MotionDirection (ITEM:=L) )
        )
      )
    )
  )# End of RIGHT
  (LEFT
    (TYPE:=MotionPlaneJoint
      (DIRECTION:=
        (TYPE:=MotionDirection (ITEM:=R) )
      )
    )
  )# End of LEFT
) #End of (R-L)
)
} ; End of Departments

```

日本語ラベル「会う」: 音韻構造表記 {1Pc,idn:L/R}

図 4.5: 音韻構造のツリー表現

表 4.1: 音韻解析による音韻の種類および出現回数

音韻カテゴリー	該当音韻数	出現回数	出現単語数	備考
Word Context	1	75	75	造語辞
Finger shape	47	819	660	手型
Palm	8	698	582	掌の向き
Wrist	8	151	138	手首の向き
Hand Motion	11	235	222	手の動き
Hand Relation	5	311	279	両手関係
Body part	26	375	350	身体位置
Body Eexpression	6	48	36	表示空間
Boyd Distance:	4	355	348	身体距離
Motion Trajectory	4	56	56	運動の軌道
Motion Plane	5	173	170	運動平面
Motion Direction	8	639	464	運動方向
Motion Condition	6	242	212	運動条件
Motion Result	3	28	28	運動結

この場合のクラス分けに際して、手話単語に対してすでに得られている音韻情報（既得音韻情報）を用いることとなる。図 4.6に同じ音韻を持つ単語データからの辞書パターン作成と認識の過程を示す。

1. すべての採取データを FFT を用いて高調波成分へ変換する。
2. 同じ音韻を持つものに対して同一の音韻ラベルを付ける。
3. 同一の音韻ラベルを持つ高調波データから平均値と分散を求め辞書データとする。
4. パターン認識の結果として音韻要素へのクラス分けを行う。

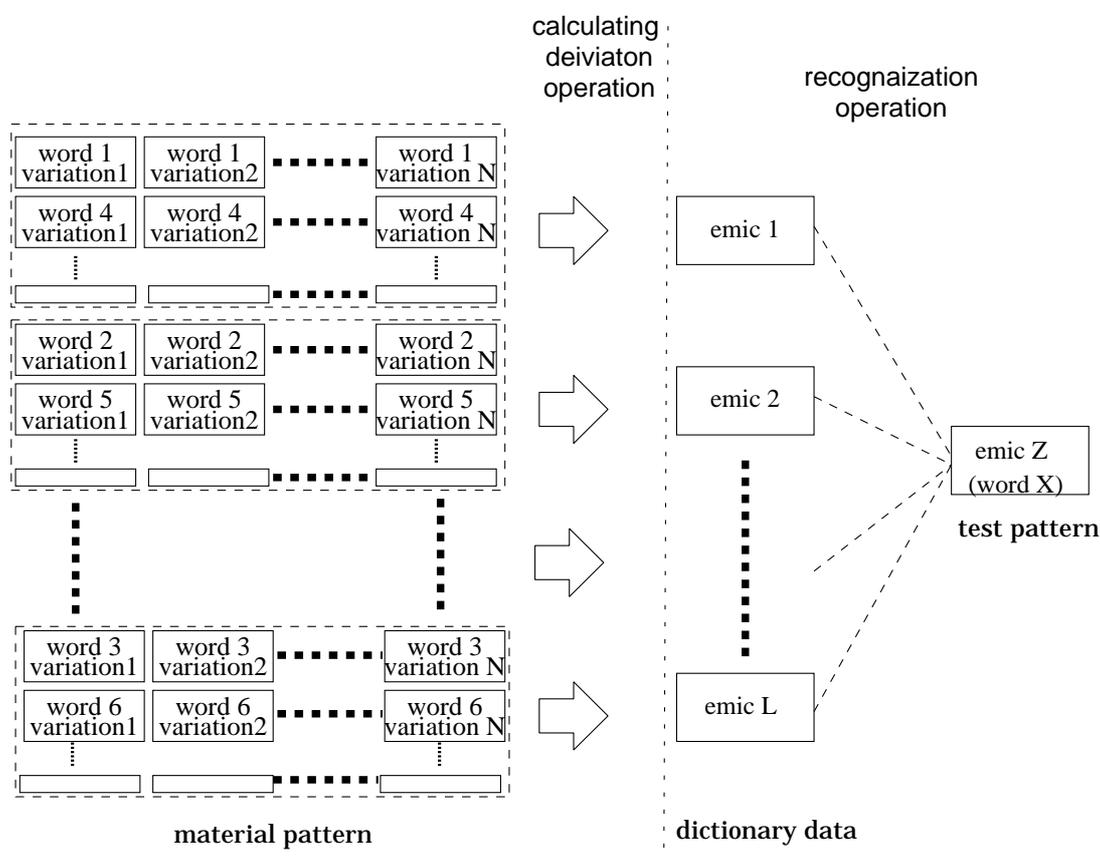


図 4.6: 単語から音韻抽出認識のための辞書作成

4.5 まとめ

DP マッチングによる認識手法を複数話者に対して行うための方法として, 中心近接尺度を定義し, 中心近接尺度による辞書パターンの選定方法について述べた.

複数話者に対して考慮した提案手法として, 調波成分を用いて手話単語認識を行う手法を述べた. 手話の運動軌跡の情報を既に得られている音韻表記から得る方法について述べた. 同じ運動軌跡を持つ手話単語をグループとし手話単語の運動軌跡を認識するための方法について述べた.

第 5 章

認識実験と評価

5.1 はじめに

本章では、従来法の一つである DP マッチングと、提案手法である調波分析を用いた認識手法とに関する実験とその比較を行う。それぞれの手法に関して特定話者の場合と複数話者を対象にした場合についての比較と検討を行う。

また、調波分析を用いて手話単語の単語の運動軌跡既得音韻情報を用いてあらかじめ単語のグループ化を行い、それらに基づいた辞書作成および運動軌跡認識の実験とその結果について述べる。

5.2 対象単語

5.2.1 対象とする手話単語

対象単語は電子辞書「ムサシ」[13]を参考に任意の 330 単語を選んだ。対象とした単語を表 5.1 に示す。

5.2.2 単語パターンの採取

手話単語は 5 人の被験者から 1 単語あたり 10 回提示してもらったパターンを使用する。1 単語に関して複数回同じ動作を提示することになる。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 単語に関する 1 語とする。また、同じ複数単語に対してことなる被験者による提示パターンがあり、被験者毎の提示された単語パターンの集まりを単語パターンセットとする。認識実験において 1 語の区間を以下のように定め、サンプリ

表 5.1: 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, 聾啞, 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

ングを行った。ここで、「両手をひざの上におく」と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。

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。

片手手話の場合は, 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, 4名の被験者の内, 3名が手話経験がそれぞれ、10年, 4年, 2年の者である。残り1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。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かかる場合もある。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, 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。

5.3 認識実験手順

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認識実験手順は図 5.1に示されるように大きく分けて2つのフェーズがある。一つは辞書を作成するためのフェーズである。もう一つは実際の認識のためのフェーズである。どちらの手順においても, 必要に応じて単語区間を設定する必要がある, 本実験では単語区間は手動にて切り出しを行った。

5.4 DP マッチングを用いた手話単語実験

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をから構成される。

DP マッチング際の前処理として, 特徴ベクトルの要素についてあらかじめ単語の最初から最後まで値の分散を求める。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。

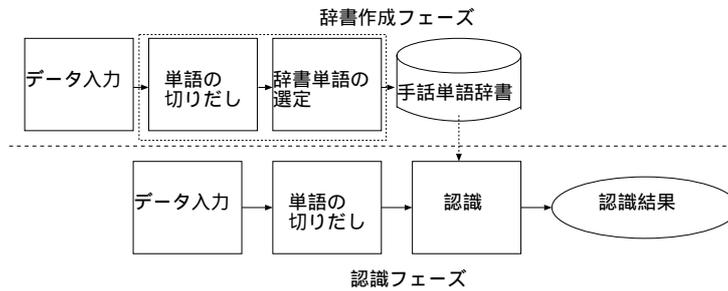


図 5.1: 認識実験手順

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。本実験ではまず、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、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する実験を行う。また、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。

演算においては PentiumII プロセッサ、クロック 450Hz、メモリ 512Mbyte を使用した。

特定話者認識

辞書パターンとテストパターンを採取した被験者同じ場合 (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 語パターン採取した内から、5 パターンを任意に取り出し、これを辞書選定用パターンとする。第 4.2.1 章で示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。

辞書作成の後、認識実験を行った。認識実験の結果を図 5.13 に示す。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り、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 $L_{\text{patternA}}, L_{\text{patternB}}$ としたとき、以下のように動的に決定する。

$$D_{\text{window size}} = \frac{1}{3} \frac{L_{\text{patternA}} + L_{\text{patternB}}}{2} \quad (5.1)$$

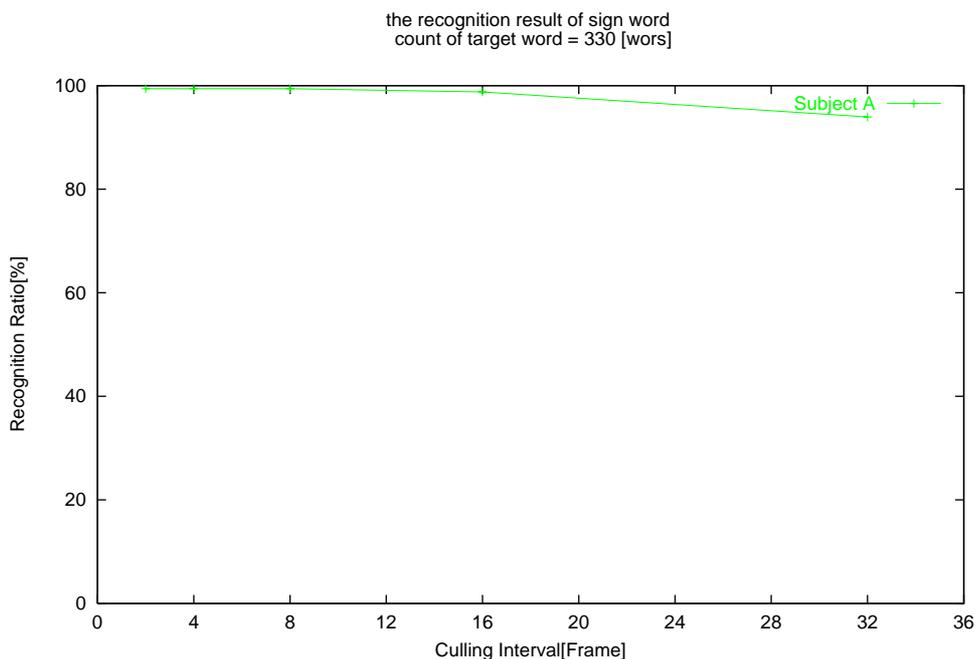


図 5.2: DP マッチングを用いた特定話者認識結果

それぞれの単語に関して, 辞書選定用パターンとして 10 語パターン中, 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 パターンのテストパターンについての認識率の平均で示す. 辞書を 330 単語とし, 330 語のテストパターンに行った認識実験結果を図 5.13 に示す. 結果は間引き間隔 2 で 99.3% となった. 間引き間隔 32 で行った実験でも 93.9% と非常に高い認識結果を示した.

複数話者認識

第 4.2.1 章で示した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, それぞれの手話単語パターンにおいて任意の 5 パターンを辞書選定用とすることで, 計 20 パターンから辞書パターンを選択した. 表 5.2 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の数を示す. 整合窓の大きさ $D_{\text{window size}}$ は特定話者認識と同様, 式 (5.1) を用いる. 4 人の被験者に対して行った認識実験結果を図 5.3 に示す. その際の演算時間を図 5.4 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. 認識率は被験者 C に関

表 5.2: 辞書に使用された各被験者の発話パターン数

被験者名	使用単語数	割合 %
被験者 A	74	22.4
被験者 B	20	6.1
被験者 C	160	48.4
被験者 D	76	23.1

して最高で 48.8%となった。被験者 B に関しては認識率は最低で 7.0%となった。それぞれの被験者の平均認識率を表 5.3 に示す。これらの認識率は 5.2 に示した比率にほぼ一致している。単語に関してテストパターンと辞書パターンを観測しても、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ほとんどなかった。総テスト単語パターン数 1320 語に対する認識率は 25.07% となり、4 名の被験者を対象とした複数話者認識においても、特定話者認識と同等の条件下以外ではほとんど誤認識されることを裏付ける結果となった。

DP マッチングの演算時間

演算時間と間引き間隔との関係を図 5.4 に示す。間引き間隔を大きくすることはパターンの長さを減少させることになる。すなわち、この結果から DP マッチングにおいてパターンの長さを増加させることは指数関数的な演算時間の増加を招くことが確認できる。

DP マッチングの演算時間に関係するパラメータには、サンプリングレート (間引き間隔) 以外に、整合窓の大きさが存在する。図 5.5 に複数話者に対する、間引き間隔 2 の場合の整合窓の大きさを変化と認識率の関係を示す。ここで間引き間隔 2 の場合、対象とした辞書パターンおよびテストパターンの長さの平均は 63.2 フレームであった。この図からは整合窓の大きさが 20, 30 で認識率が若干上がるが、それ以降 60 までではむしろ下がっている。整合窓としてはパターン長さの $1/2$ 、あるいは $1/3$ が妥当と言える。

表 5.3: DP マッチングによる認識率の平均

被験者名	認識率 %
被験者 A	21.39
被験者 B	7.21
被験者 C	49.15
被験者 D	22.54
平均	25.07

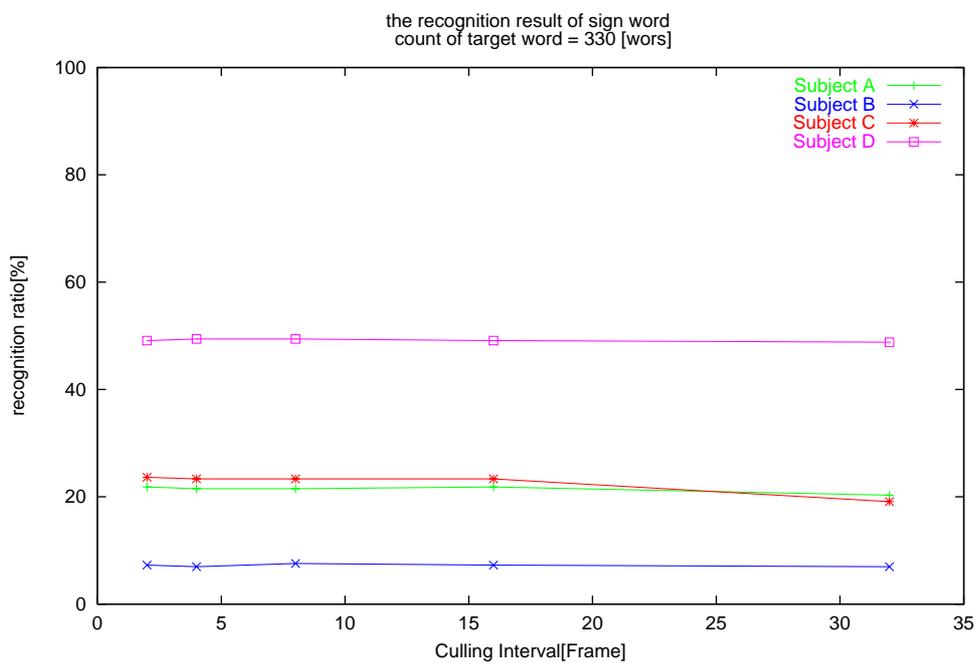


図 5.3: DP マッチングを用いた複数話者認識結果

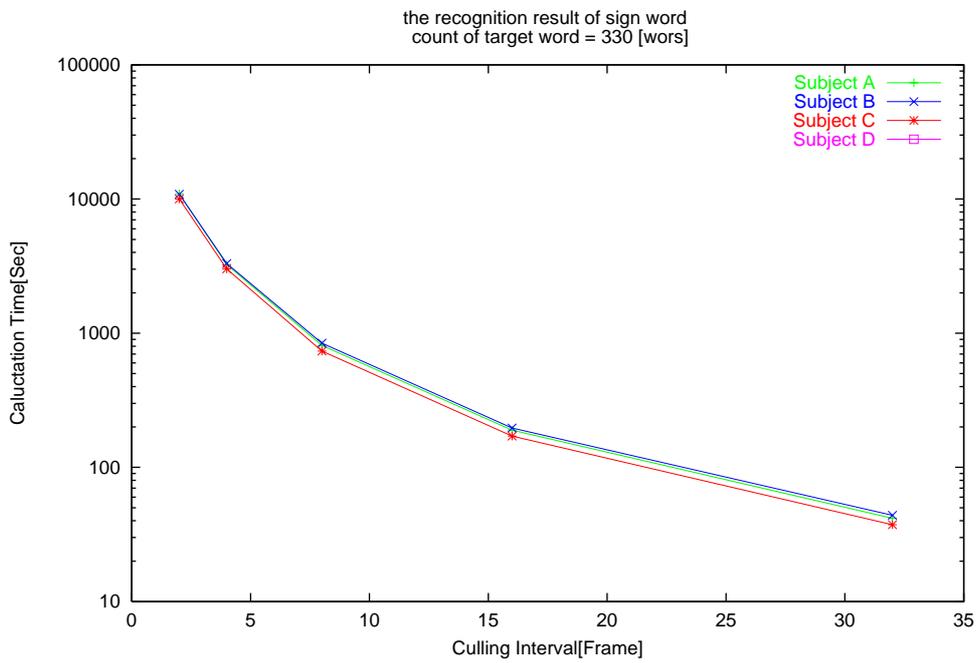


図 5.4: DP マッチングの間引き間隔による演算時間

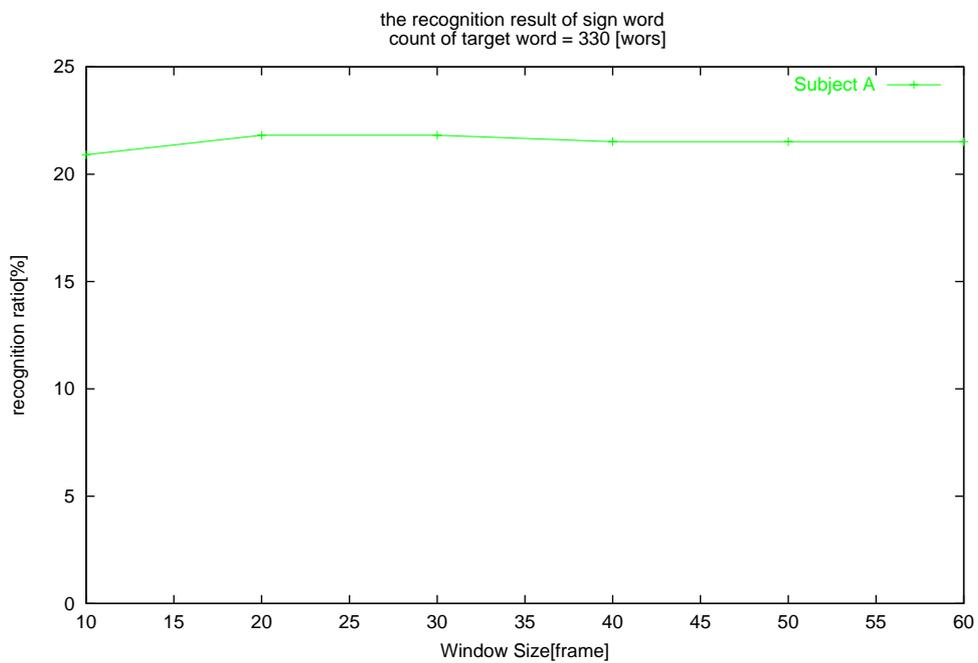


図 5.5: 整合窓と認識率の関係

5.5 FFT を用いた手話単語実験

DP マッチングによる認識手法が複数話者に適応していないことを受けて、複数話者に適応した認識手法の実験を行う。

本節では第 4.3 節で述べた 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、調波成分によって辞書の作成および認識を行う手法についての認識実験を行う。これらの実験を通じ、複数話者認識における DP マッチングを用いた手法に対して、調波解析を用いた手法の有効性を示す。調波解析による認識手法を、特定話者と複数話者 (3 名、および 4 名) を対象とした場合について実験を行う。次に、調波解析による認識手法を用いての手話単語の運動軌跡の認識実験を行う。

5.5.1 実験データの前処理と調波解析による辞書作成

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に関して正規化を行う。正規化として、線形補間を用い 32 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 の冪乗において効率が良いことから、最終的にフレーム数を 32 フレームとして実験を行うことを決定した。この際には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、なおかつ、辞書選定用に用いた単語あたり 5 パターンを使用して調波解析を行った後、辞書を作成した。

5.5.2 認識実験

調波成分を用いる際のパラメータとしては、基本波に対する高調波成分の高さがある。これを有効高調波階数とし、第 n 次高調波成分まで使用した場合を n 階有効高調波による認識とする。なお、比較は複素成分表現を用いて行う。

調波成分分析による特定話者認識

特定話者に対して、使用調波幅を変化させた場合の認識結果を、図 5.6 に示す。特定話者の認識では、著者は分析を用いた手法でも 89.0% から 94.5% と、DP マッチングの特定話者認識に準じた認識率を得た。

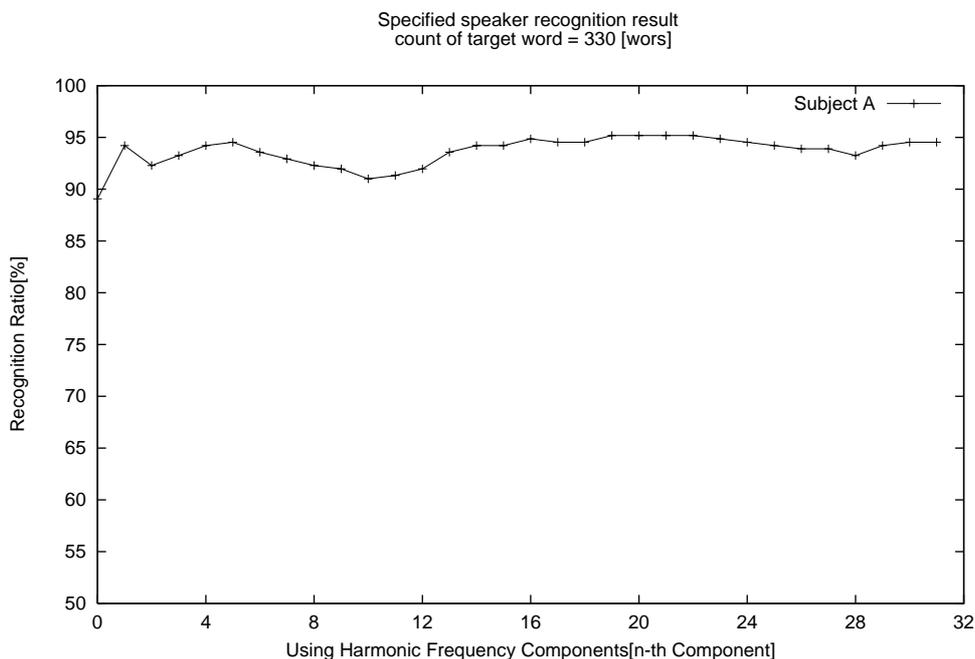


図 5.6: 周波数分析による特定話者認識結果

調波成分分析による複数話者認識

使用高調波幅を変化させた場合の、第 1 位認識率を求める。3 人の複数話者の場合の手話単語実験の結果を図 5.7 に、4 人の場合を図 5.8 に、5 人の場合を図 5.9 示す。

3 人の複数話者を対象とした実験では、すべての被験者において 12 階有効高調波を使用した場合に、一番良い認識率が得られた。表 5.4 に 12 階有効高調波使用時の第 1 位認識率および第 3 位迄認識率を示す。被験者 A から採取した手話データのテストパターンの組が最もよい認識率を示し、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パターンの組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% の認識率となった。8 階有効高調波近辺から認識率が横ばいとなることから、7 階有効高調波以降を対象として考える。その場合、すべての被験者に関する平均認識率は 83.4 % となった。

本実験では、12 階有効高調波時に一番良い認識率となった。それ以上の有効高調波の条件では認識率の若干の下降が見られた。一定有効高調波以上での認識率の下降は、測定ノイズや手話単語としての有効な情報が少なくなることが原因と考えられる。

3,4 名それぞれの複数話者認識においてすべての被験者の認識率の平均が高い高調波幅の場合の、認識率を表 5.5 および表 5.6 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。また、4階有効高調波時に高い一番高い認識率となりその場合の平均は62.3%となった。4階有効高調波以上では68.2%を示した被験者の単語パターンセットもある。しかし一方では40%程度までに認識率が低下した被験者のデータもあった。5名を対象とした実験ではさらに認識率の低下がみられ、最高の認識率でも、被験者A(N=2)で57.2%であった。

調波成分の内、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すことを考慮した場合、一定程度以上の調波成分は認識に十分な効果を果たさない結果となった。

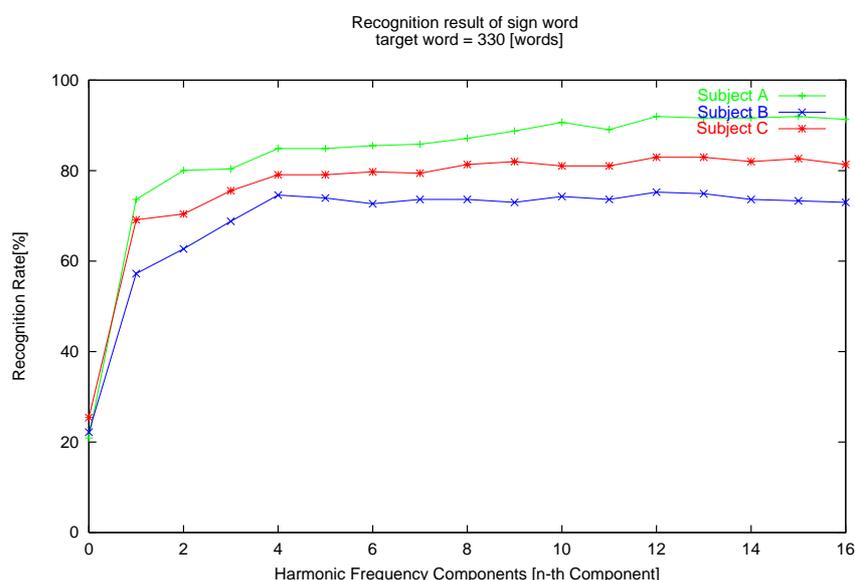


図 5.7: 3名を対象とした有効高調波階数と認識率の関係

表 5.4: 周波数分析による3名を対象とした複数話者認識結果

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

12 高調波幅時

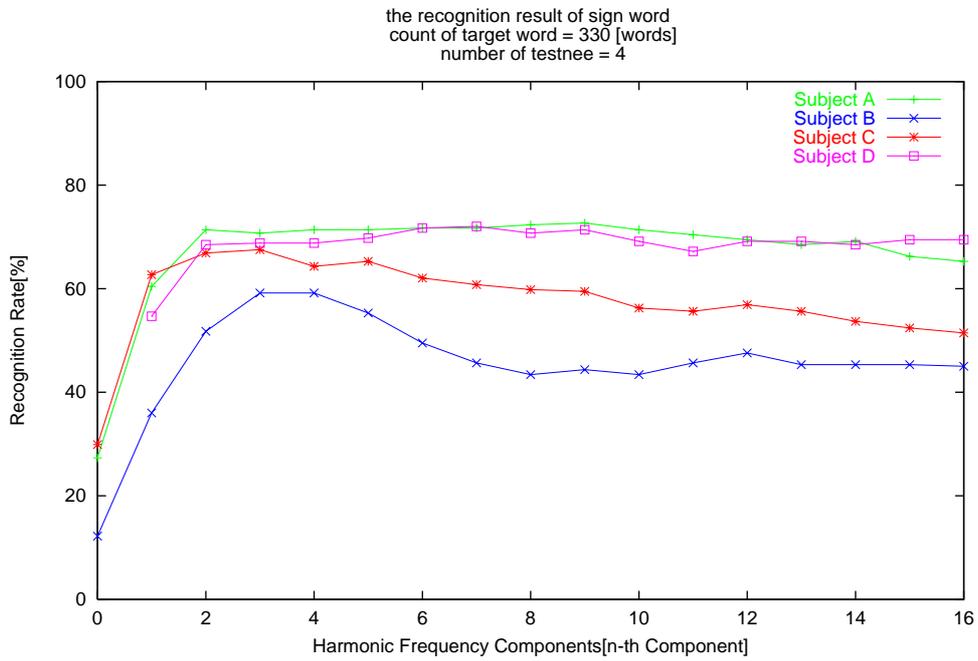


図 5.8: 4 人を対象にした複数話者認識の結果

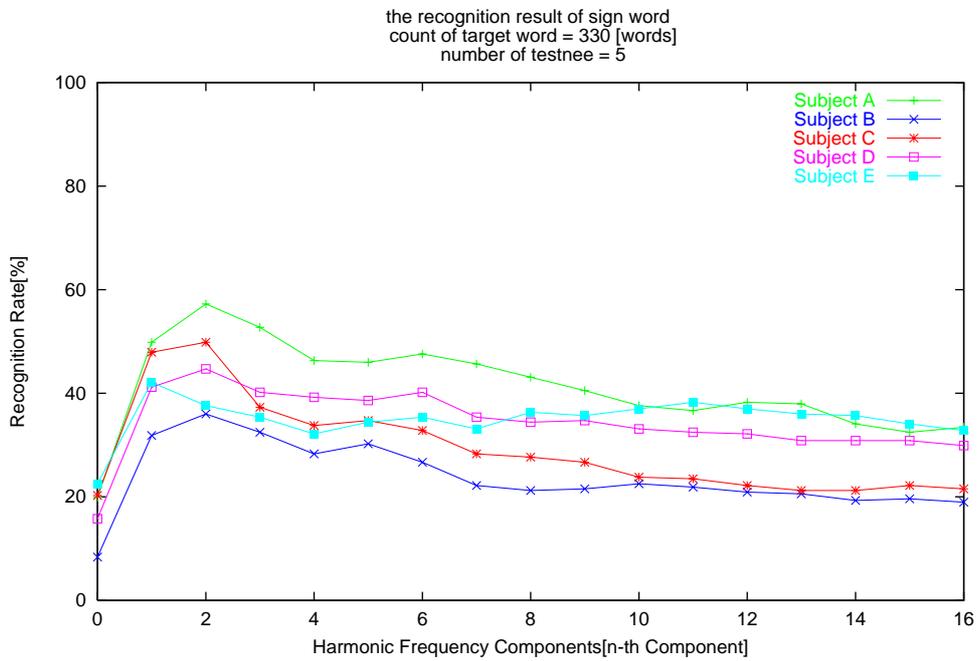


図 5.9: 5 人を対象にした複数話者認識の結果

表 5.5: 4 名を対象とした複数話者認識結果 表 5.6: 5 名を対象とした複数話者認識結果

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

$N = 2$ 高調波幅時

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

$N = 2$ 高調波幅時

調波成分分析による演算時間

有効高調波階数と認識時間の関係を図 5.10 に示す。有効高調波階数は、周波数分析を用いた手話単語認識の計算精度に関するパラメータといえる。その演算性質上、認識時間は有効高調波階数の増加に対して線形性を示すことが確認できた。

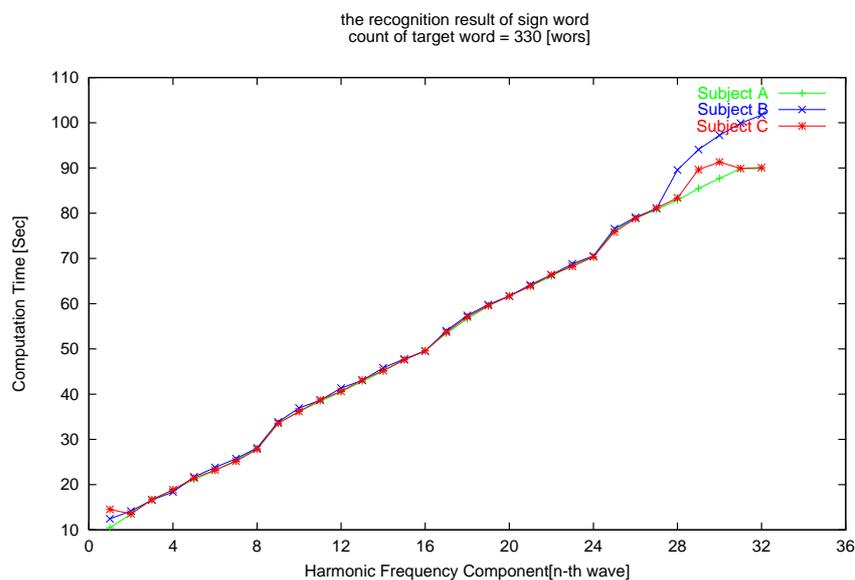


図 5.10: 有効高調波階数と認識時間の関係

5.6 運動軌跡の認識実験

複数話者の手話単語認識対しての有効性を確認した調波分析を用いた認識手法を、運動軌跡の認識に応用した場合の認識実験を行う。辞書作成および認識において、単語認識では同じ単語を辞書としていたのに対して、運動軌跡の認識は同じ運動軌跡を持つ語パターンを辞書として使用し、認識の際には認識されたパターンとテストパターンがおなじ運動軌跡を持つかどうかで認識率を算出することになる。

5.6.1 音韻情報分析による運動軌跡グループ

運動軌跡の認識のため、既得音韻情報を基に同じ運動軌跡を持つ運動軌跡を有するグループを作成し、辞書を作成する。手の運動軌跡に関しては、FFTを施されたデータの位置および姿勢のみを考慮し、関節角に関するデータは用いない。

表 5.7: 使用単語およびグループ

グループ	単語	動き
1	挨拶, 会う, 集まる, 結婚,	胸前で両手を近付けて, 止める
2	赤, 白, 通訳	口元に手を近付ける
3	明日, 未来, 明後日	肩から前に手を出す
4	遊ぶ 会社	両手を頭の脇で交互に動かす
5	兄 姉	片手を挙げる
6	安心, 重い, 暇	両手を下ろす
7	あなた, 言う, 行く	顔前から手を前に出す
8	妹, 弟	片手を下げる
9	一昨日, 過去, 昨日	肩越しに手を後ろにする
10	夫, 妻	胸前に両手を近付けた後, 片手を更に前に出し止める
11	間, 道	両手を肩幅で前に出す
12	兄弟, 比べる, 姉妹	両手を上下させる
13	別れる, わける	胸前で両手を寄せた後に左右に離す

34 単語に対して 13 グループに分類した後、テスト単語パターンの入力に対して正しいグループへ認識されるかで評価を行う。表 4.1に示された音韻構造情報をもとに、空間お

よび運動の部門に着目しグループを作成した。対象とした単語およびグループを表 5.7に示す。これらの単語グループを基に、特定話者と複数話者に関する運動軌跡の認識実験を行う。

5.6.2 運動軌跡認識実験結果

特定話者を対象とした実験, すなわち, 辞書の作成に用いたパターンとテストパターンに用いたパターンを同一被験者から採取した場合の認識結果を図 5.11に示す。最低有効高調波階数を 8 とするとき最低認識率は 85.3%, 最高が 94.2% という結果になった。

3 人の被験者から採取したデータから辞書を作成し, それぞれの被験者から採取したデータをテストパターンとした実験に付いて図 5.11に示す。高調波幅 $N = 8$ の場合, すべての被験者によるテストパターンにおいて認識率の最高は 88.2 % , 最低が 67.6 % となった。

これらの実験によって, 調波分析による運動軌跡の認識の有効性が確かめられた。

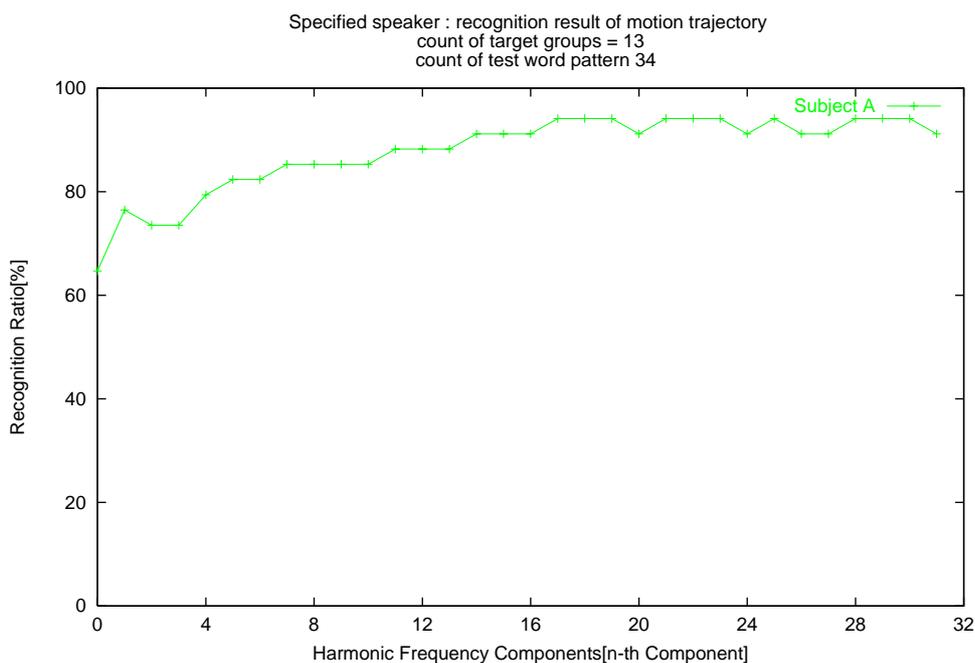


図 5.11: 34 単語に対する特定話者の運動軌跡認識結果

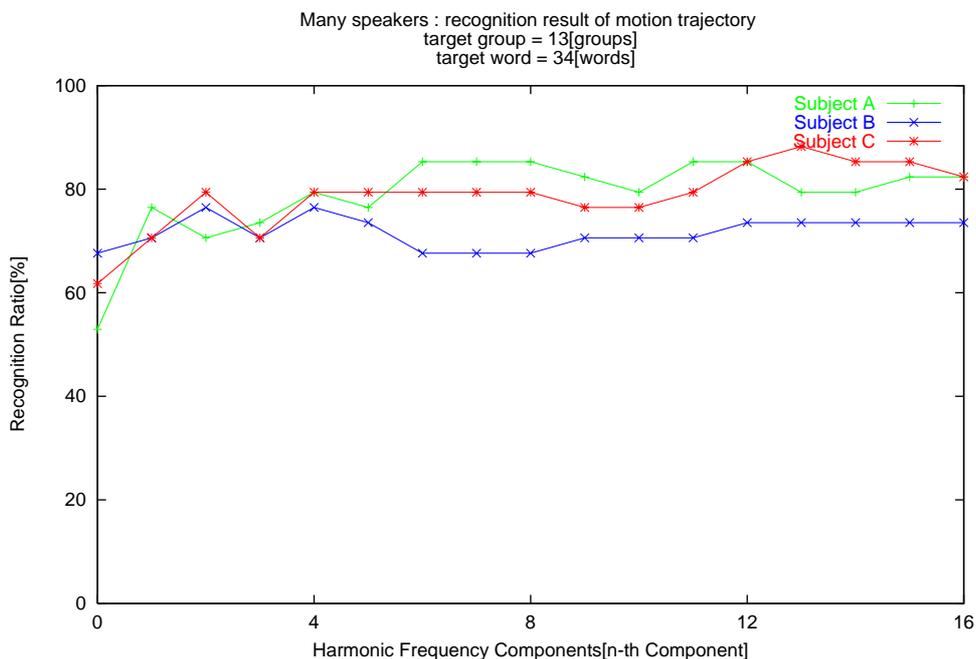


図 5.12: 34 単語に対する複数話者の運動軌跡認識結果

5.7 比較と検討

DP マッチングを用いた手法と FFT を用いた手法との比較

手話単語認識に関して、DP マッチングを用いた認識手法と FFT を用いた認識手法との認識率の比較を行う。特定話者認識では DP マッチングは最高で 99.4% となり非常に高い認識率であること確認した。FFT を用いた認識に関しての特定話者認識では DP マッチングに次ぐ認識率であった。

複数話者に対する認識実験では、4 人の被験者に対して 25.07% の認識率であったのに対し、FFT を用いた手法では 83.4 % であった。DP マッチングを用いた手法は被験者数に依存した値で、複数話者に適用できる手法ではないことが確認できた。

次に演算時間の比較を行う。2 つの認識手法ではそれぞれに固有の可変要素があるが、それら可変要素に関する演算時間の比較はできない。そこで、演算時間と認識率の関係によって比較を行う。

同一採取データを基にした特定話者認識における、演算時間の比較を図 5.13 に示す。特定話者認識において、両方の手法が重なる演算時間帯では FFT を用いた場合が若干低い結果となった。認識率が収束するまでの時間は FFT を用いた手法が速い結果が得られた。

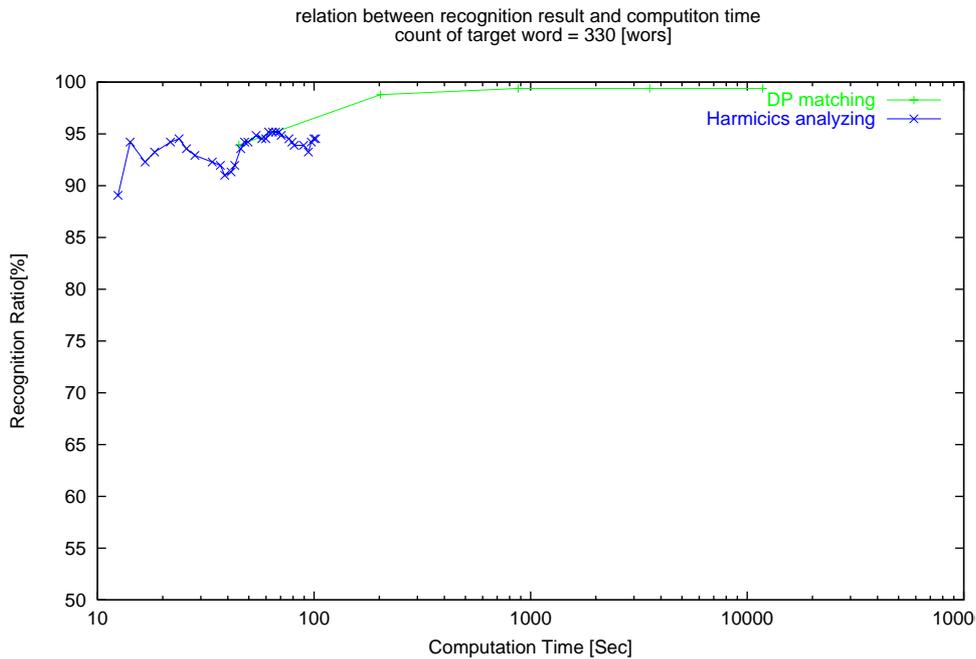


図 5.13: 認識時間の比較

5.8 まとめ

本章では、従来方法である DP マッチングを使用した認識手法と、それに対して提案をした FFT を用いた手話単語認識に関する認識実験を行った。その後、FFT を用いた運動軌跡の認識実験を行った。

複数話者認識に関しては 330 単語に対して 4 人の被験者から採取したデータを用いた場合、中心近接距離を用いて辞書を作成した DP マッチングでは 25.07% の認識率に留まった。

FFT を用いた調波解析を基に 3 人の被験者に対して辞書を作成した場合、第 1 位認識率が 83.4 %、第 3 位認識率 94.0 % という結果が得られた。次に、4 人の被験者に対して辞書を作成した場合、4 人に対する平均認識率が最高で 68.2% へと低下するという結果が得られた。

次に提案した手法によって、手話単語の運動軌跡の認識実験を行った。34 単語に対して 13 グループに分類した後、平均認識率は 85.3% という結果を示した。

これらの実験より調波分析による複数話者認識および、運動軌跡認識の有効性が確かめられた。

第 6 章

結論

昨今の福祉社会への関心の高まりとともに、聴覚障害者を含む心身に障害を持つ人々が健聴者と変わらなく生活が可能な環境への取り組みが各方面で進められている。

聴覚障害を持つ者の場合、音声言語を直接用いないでのコミュニケーション手段が必要になる。非音声のコミュニケーション手段としていくつかの方法があるが、そのなかでも手話は重要な方法であり、特に幼児期の言語取得年齢以前に失調した聾啞者に関しては手話が母国語であるといえ、欠かすことのできないものである。

一方で、健聴者がこれら聾啞者とのコミュニケーションを行う場合、それぞれの母国語である音声言語および手話ではその用を足さず、しばしば手話通訳者が必要となる。しかし現状では、手話通訳者はその需要に対して供給が少ないのが現状であり、手話通訳の代行となる手話認識の研究が盛んに行われている。また、手話認識を健聴者の手話学習に応用することを考えた場合、単に手話単語の合否を判別するだけではなく、提示された手話のどの部分が不適當であるかを示せることは、有用だと考える。

そこで本研究では、従来行われていなかった複数話者に対する手話単語を対象として、調波解析を用いた認識手法の提案ならびに認識システムの構築を行った。また特定話者認識の従来法である DP マッチングに関して、特定話者認識ならびに複数話者認識の認識に関して比較を行った。運動軌跡においては音韻表記法による手話単語の音韻情報を拠り所とした。

手話データは装着型の位置・姿勢および手形状入力装置を用いて採取した。これに伴い、データを採取するためや採取データからの単語区間の切り出しを行うための、手話データの可視化機能を含むデバイス制御およびデータ加工・編集システムを構築した。

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% の高い認識率を確認し、FFT を用いた手法では 95.2% の結果を得た。次にそれぞれの手法

を複数話者に対する認識実験を行い,DP マッチングを使用した場合には 25.1%の第 1 位認識率に留まった. FFT 処理による調波分析を用いた手法では 3 名の複数話者を対象とした場合, 83.4 % の第 1 位認識率となり, 調波分析を用いた手法の有用性を確認した. ただし,4 名を対象とした実験では平均で 62.3%と認識率の低下が見られた.

次に,手話単語の運動軌跡を認識する実験を行った. 既得音韻情報を基にして 34 単語に対して 13 グループに分類した後, 認識実験を行った. 特定話者の運動軌跡に関しては平均で 90.2%の認識率を得た. 3 人の複数話者に関する平均認識率は 85.3%となり,3 人の話者に関しては 68.2%と認識率の低下がみられた.

認識実験に関しては,PentiumII プロセッサ, クロック 450Hz, メモリ 512Mbyte を使用した. FFT を用いた認識手法の場合, 今回行った実験では 330 単語からの認識と云う条件下では, 1 単語の認識は 0.3 秒以下で終了しており, 0.15 秒となるような可変要素の実験条件でも特に認識率の低下はないことを確認している. これにより実時間処理が可能な複数の話者に対する認識手法が示唆される.

6.1 今後の課題

本研究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,単語区間の切り出しが必要である. 運動軌跡の認識に関しても,対象とする単語および運動軌跡の種類を増やした実験を行うことが肝要である. またより実用的な認識手法として,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.

謝辞

本研究を進めるにあたり、終始御指導を御鞭撻を賜りました、北陸科学技術大学院大学 情報科学研究科 堀口 進 教授、阿部 亨 助教授にここで深く感謝の気持ちを表します。

副テーマで御指導を頂きまして主テーマへの示唆を頂きました渡部 卓雄助教授 に感謝致します。

日頃より有意義な御教示と議論を頂きました、山森一人助手、井口 寧 助手、林 亮子 助手に感謝致します。

公私ともにさまざまな場面で御助力を頂きました星芝 貴行 さん、加藤 聡さん、堀井 洋くん、奥野弘之くん、に感謝致します。手話単語データの採取と手話に関するアドバイスを頂きました徳田昌晃さんに感謝致します。手話単語データの採取への御協力を頂きました、越田正美さん、伊藤徳広くん、新座良和くんにお礼申し上げます。

日頃より、お世話になった堀口・阿部研究室の皆様、に厚くお礼申し上げます。

本研究の一部は、文部省特定研究ならびに電気普及財団の研究助成を用いて行われました。関係各位に深く感謝致します。

参考文献

- [1] 内田 雅文, 石川 潔, 井出 英人 : “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.
- [2] 高橋 友一, 岸野 文郎 : “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.
- [3] 萩原 芳彦, 堀口 進 : “指手骨格モデルを用いた指文字認識”, 日本手話学会 第 21 回 全国大会予稿集, pp.32-35, Sep.1995.
- [4] 後藤 岳志, 堀口 進 : “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] 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
- [6] 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.
- [7] 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.
- [8] E.Ohira, H.Sagawa, T.Sakiyama, M.Ohki : “A Segmentation Method for Sign Language Recognition”, IEICE TRANS.INF. & SYST., VOL E78-D, NO.1 Jan.1995.
- [9] 神田和幸 : “手話学講義”, 日本福村出版,1994.
- [10] 神田和幸, 中博一 : “日本手話の音韻表記法”, 日本手話学会, 手話学研究 12(1991),31-39.
- [11] “CyberGloveTM User’s manual”, Virtual Technologies, 1993.
- [12] “3SPACE USER’S MANUAL”, POLHEMUS, 1993.

- [13] 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.
- [14] “手話法辞典”，栃木県聾学校，栃木県聴覚障害者協会，1978.
- [15] 入江 督 編：“手話辞典”，大阪ダイヤモンド工業株式会社，1993.
- [16] 伊藤，竹村：“手話入門”，廣済堂，1982.
- [17] 中野，伊藤，松本：“手話への招待”，福村出版，1977.

研究業績

- [1] 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
鈴木信勝, 後藤岳志, 阿部亨, 堀口進
電気関係学会北陸支部連合大会 F-52, pp361.1997.
- [2] 曖昧な運動を含んだ手話単語の認識
鈴木信勝, 阿部亨, 堀口進
電気関係学会北陸支部連合大会 F-49, pp353.1998.
- [3] 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
鈴木 信勝 , 堀口 進
北陸先端科学技術大学院大学リサーチレポート, IS-RR-98-0029P
- [4] 手形状入力装置を用いた指文字認識システム
鈴木 信勝 , 後藤 岳志, 堀口 進
北陸先端科学技術大学院大学リサーチレポート, IS-RR-98-0033P