

Kinect を用いた
フィードバック機能を有する
指文字練習システムの開発と評価

平成 25 年度

筑波技術大学大学院修士課程技術科学研究科

産業技術学専攻

田中 陽土

目次

第1章	はじめに	1
1.1	研究背景	1
1.2	先行研究例	1
1.3	研究目的	3
1.4	本論文の構成	4
第2章	Kinect を用いた指文字練習システムの概要	5
2.1	指文字練習システムの開発の難しさ	5
2.2	指文字練習システムの開発における基本的な考え方	6
2.3	指文字練習システムの各部の概要	9
2.3.1	画像入力部	9
2.3.2	辞書	10
2.3.3	類似度計算部	11
2.3.4	フィードバック表示部	12
第3章	Kinect を用いた指文字練習システムの詳細	14
3.1	Kinect の配置	15
3.2	手領域の切り出しと正規化	17
3.3	特徴抽出	19
3.4	辞書の構成	21
3.5	照合パラメータの決定方法	23
3.6	フィードバック画面の構成と UI の開発	25
第4章	試作システムの実験的評価	30
4.1	リアルタイム照合実験	30
4.1.1	閾値 $Th1$ と閾値 $Th2$ の決定	31
4.1.2	辞書構成に協力した手話通訳者の場合	35
4.1.3	日常的に指文字を使用している聴覚障害者の場合	36
4.1.4	考察	38
4.2	指文字の練習効果に関する評価実験	40
4.2.1	実験 I:フィードバックがある場合	41
4.2.2	実験 II:フィードバックがない場合	44
4.2.3	実験 I と実験 II の比較	45
4.2.4	考察	47
第5章	今後の課題	49
第6章	まとめ	50
	謝辞	52

参考文献	53
研究業績	55
付録 A 指文字の標準形の図一覧	56
付録 B 距離値と実際の距離の関係	60
付録 C 指文字ごとの閾値 $Th1$ と閾値 $Th2$ の一覧	62
付録 D 最適なパラメータを決定するための実験	63
D.1 手領域画像の大きさの決定	63
D.2 カーネルパラメータの決定	65
D.3 K-mean 法による削減枚数の決定	67
D.4 辞書部分空間の次元数の決定	68
D.5 入力部分空間の次元数の決定	69

目次

図 2.1	経時変動による指文字表現の変化の例	6
図 2.2	指文字練習システムのイメージ図	8
図 2.3	練習システムの処理の流れ	9
図 2.4	Kinect の外観	10
図 2.5	相互部分空間法のイメージ図	12
図 3.1	指文字練習システムの全体図	14
図 3.2	Kinect を用いた手形状の撮影方法	15
図 3.3	Kinect の傾き方による手領域の写り方の違い	16
図 3.4	手領域の切り出し処理の流れ	18
図 3.5	ノイズがある図とノイズがない図	18
図 3.6	特徴抽出の流れ	20
図 3.7	中級者が表出した場合の曖昧な表現	20
図 3.8	データを収集するときの手の動かし方	21
図 3.9	手を動かし過ぎた例（「よ」が「み」の文字に見える）	22
図 3.10	判定処理の流れ	24
図 3.11	フィードバック画面の例	26
図 3.12	誘導メッセージの種類と実際の例	27
図 3.13	類似度の棒グラフと判定結果	28
図 3.14	判定結果の色別表示	29
図 3.15	判定結果のリスト表示	29
図 4.1	指文字以外の手形状のパターン 36 種類	32
図 4.2	入力部分空間の構成方法	32
図 4.3	「ら」の閾値 $Th1$ を決定するときの FAR と FRR の図	34
図 4.4	聴覚障害者が照合成功したパターンと失敗したパターンの例	39
図 4.5	手領域の切り出しが成功したパターンと失敗したパターンの例	39
図 4.6	指文字練習システムに関するアンケート	42
図 4.7	被験者が指文字表現の誤りに気づき、修正する様子	43
図 4.8	実験 I と実験 II の照合成功率の推移	46
図 4.9	実験 I と実験 II の平均フレーム数の推移	46
図 4.10	アンケートの集計結果	47
図 A.1	「あ」～「こ」の指文字の標準形一覧	56
図 A.2	「さ」～「と」の指文字の標準形一覧	57
図 A.3	「な」～「ほ」の指文字の標準形一覧	58

図 A.4 「ま」～「わ」の指文字の標準形一覧.....	59
図 B.1 距離値と実際の距離の関係を調べる時の撮影方法.....	60
図 B.2 距離値と実際の距離の関係.....	61
図 D.1 手領域画像の幅を変更した場合における照合成功率の結果.....	64
図 D.2 カーネルパラメータを変更した場合における照合成功率の結果.....	66
図 D.3 削減枚数を変更した場合における照合成功率の結果.....	67
図 D.4 辞書部分空間の次元数を変更した場合における照合成功率の結果.....	68

表目次

表 4.1	手話通訳者の照合結果.....	35
表 4.2	聴覚障害者の照合結果.....	37
表 4.3	実験 I の照合成功率と平均フレーム数の結果.....	43
表 4.4	実験 II の照合成功率と平均フレーム数の結果	44
表 4.5	実験 I と実験 II における練習回数ごとの差の検定結果	45
表 C.1	指文字ごとの閾値 $Th1$ と閾値 $Th2$ の一覧	62
表 D.1	入力部分空間の次元数を変更した場合における照合成功率の結果.....	69

筑波技術大学

修士（工学）学位論文

第1章 はじめに

1.1 研究背景

指文字は聴覚障害者にとって重要なコミュニケーション手段のひとつである。指文字は日本語のかなをひとつの手形状や手の動きで表現したものであり、指文字だけで会話を行うことが可能である。一方、手話も聴覚障害者にとって重要なコミュニケーション手段のひとつであり、単語ごとにひとつの手の動きで表現することができるが、人名や専門用語などはその単語を表す手話の表現が定義されていないことが多く、手話を用いて表現することができないときは指文字が用いられる。したがって、手話を学習するときは最初に指文字を身に付けることが多く、手話の講習会では指文字を覚えることから始める。

指文字を学習するときは、本や DVD などのメディアを参考とした一方向の練習、実際に指文字を用いて会話する双方向の練習がある。学習において重要であるのは双方向の練習であるが、これは本や DVD を用いた一方向の練習では、指文字の表現を誤って覚えてしまうと、その誤りを指摘する相手がおらず、自分で気づくことができない恐れがあるからである。したがって、その誤りを指摘することで、指文字の表現を正しく覚えてもらうために、双方向の練習が必要となる。文献 [1]においても、本や DVD を用いた一方向の練習では指文字を文字または言語として定着させることが難しいと報告されていることから、双方向の情報のやりとりをしながらの指文字の習得がいかに重要であるかがわかる。

しかし、双方向の練習には会話の相手が必要であるが、一般的に音声言語である日本語と比べて指文字や手話を使う人が少ない。したがって、周囲に指文字を覚えている人が少ない場合は指文字や手話を学習する機会が少なくなる。他にも指文字を覚えるために手話の講習会などに参加することができるが、講習会に参加することは労力が必要であり、手軽に指文字を学習することが難しくなってしまう背景がある。

そこで、筆者は双方向の練習が可能となる指文字練習システムを開発することで、手軽に指文字を学習する機会を増やすことを考える。

1.2 先行研究例

ここで、指文字練習システムの先行研究例について述べる。鈴木らは振動子付き手形状入力装置を用いることで指文字学習者への間違いを指摘するフィードバックを可能にした指文字教育システムを提案している [2]。まず、ユーザはあらかじめ振動子付き手形状入力装置を装着し、メニュー画面から自分の学習したい指文字を選択する。指文字が選択されたあとはディスプレイに指文字が表示され、その表示に合わせて手形状を呈示する。そして、各関節に対応した関節角センサを通して屈曲角を測定することで、5本の指のポーズ(回

内、屈曲、伸展など) から構成される手形状を認識する。また、レシーバを用いて手形状入力装置から発生される磁気を観測することで、手形状入力装置の連続的な 3 次元座標のデータを計測し、手形状の向きや動きを認識する。認識した手形状・向き・動きの情報をもとに、事前に登録した手本の手形状・向き・動きと比較を行うことで、ユーザが呈示した手形状に誤りがあるかどうか判定する。もし、誤りが見つかった場合、その誤った箇所に対して振動を与え、同時にコンピュータグラフィックス (CG) とアニメーションを用いてその誤り箇所と訂正方法をユーザに表示することで、フィードバックを行っている。

その他にも、田畑らによって提案されているデータグローブを用いた指文字教育システム [3] など、特殊なデバイスを用いた先行研究例はいくつか見られるが、特殊なデバイスを用いた場合は、ユーザの自然な動きが制限されてしまい、上手く指文字の学習ができない可能性がある。

一方、筆者の共同研究者である瀬戸山らは特殊なデバイスを用いず、画像認識を用いた指文字練習システムを開発中であり、形状相関マップを用いたフィードバックのある指文字練習システムを提案することで、指文字練習に有効であることを示唆している [4]。形状相関マップとはユーザの手形状と手本の指文字がどの程度類似しているか類似関係を直感的に教えるマップであり、ユーザはそのマップを参考に、手本の指文字がどのような手形状か推測しながら練習を行うことができる。指文字を認識するときはカメラ脇に LED 照明を設置し、撮影時には LED を点滅させながら手形状の撮影を行う。ユーザの手は LED 照明の近くにあるため、撮影した画像には手領域に対し、点灯時と消灯時で明度に差が生じる。この明度の差分を用いることで画像に写る手領域と背景領域を切り出すことができ、特殊なデバイスを用いずに手領域を対象として指文字の認識を行うことができる。形状相関マップを構成するときは相互部分空間法をベースとした類似度の計算方法を用いる。まず、動きのある指文字「の」、「も」、「り」、「を」、「ん」を除く 41 文字の手本となるデータを集め、相互部分空間法を用いて、それぞれの指文字データ同士の類似度を計算する。1 つの指文字から、40 文字に対する指文字の類似度が得られ、その類似度をもとに距離を定義する。距離は指文字同士の類似関係を表し、指文字同士の類似度が高いほど距離は 0、類似度が低いほど距離は 1 に近づく。例えば、手形状が似ている「あ」と「さ」の場合、距離は 0.2、手形状が似ていない「い」、「お」の場合、距離は 0.8 といったように指文字同士の類似関係を定量的に表すことができる。指文字同士の距離を計算した後は、距離をもとに形状相関マップの表示位置を決定し、ユーザにそのマップを表示する。ユーザが手形状を呈示すると、ユーザの手形状データと手本のデータ 41 文字との類似関係を計算し、マップ上にはリアルタイムでその手形状データと 41 文字の類似関係を反映した位置に表示する。マップ上での位置が「あ」と「さ」の中間に位置する場合、ユーザが呈示した指文字は「あ」と「さ」の指文字を呈示していないが、その中間の手形状を出していると感じることができるため、指文字練習において形状相関マップを用いたフィードバックが有効であることを示唆している。

しかし、このシステムには撮影時に LED 照明が点滅することでユーザの目に負荷がかかる問題がある。また、画像入力デバイスに高価なカメラを使用しており、安価でユーザへの負担が少ない画像入力デバイスを用いることが課題となっている。さらに、練習システムとしてのユーザインタフェースが充分でないことから練習効果の有効性を確認することが難しいという問題がある。

1.3 研究目的

以上より、指文字練習システムを開発するにあたって、安価でユーザへの負担の少ない画像入力デバイスを用いて効率的に手領域を抽出し、手軽に指文字を練習することができること、またユーザの呈示した指文字が誤っていた場合はその誤りを指摘することができることが重要であることがわかる。

そこで、本論文では以下の 3 点を目的として指文字練習システムの開発と評価を行う。ここで、標準形とは文献 [5] に記載している指文字の表現方法を参考にした表現であり、本論文ではこの指文字の表現を標準の表現として考え、標準形として定義する。なお、標準形の詳細については付録 A に記載する。

- (1) 安価でユーザへの負担が少ない画像入力デバイスとしてゲームなどのエンターテインメント用として一般的に普及している「Kinect」を用いて、安定して指文字を表す手領域を切り出し、手領域を効率的に抽出する。
- (2) 手領域特徴の照合方法と「ユーザ独立」の辞書における構成方法の検討を行い、ユーザが呈示する手形状と標準形との 3 次元形状との照合を高精度に行う。
- (3) 照合結果をリアルタイムにフィードバック表示するための方法とユーザインタフェースを検討し、双方向の情報のやり取りを可能とする。

(2) の「ユーザ独立」とは、辞書に本システムの対象となるユーザのデータが含まれない場合のことである。本システムの対象となるユーザは本やイラストを参考に覚え始めた初級レベルのユーザや、指文字を覚えているが指文字の表現に一部誤りがある中級レベルのユーザである。対象となるユーザのデータを用いずに構成したユーザ独立の辞書は汎用性が高いが、ユーザ依存の辞書と比べて高精度に照合を行うことが難しくなる傾向にある。本システムにユーザ独立の辞書を用いる理由は、ユーザ依存の辞書を用いる場合は対象となるユーザのデータを収集する必要があり、そのときに標準形ではない誤ったデータを収集してしまうことがあるため、標準形ではない表現を呈示するときに誤って照合が成功してしまう可能性があるからである。そこで、標準形を常に呈示することができる上級レベルのユーザを対象にデータを収集し、ユーザ独立の辞書を構成することで標準形ではない表現に対して誤って照合が成功しないようにする。

上記の 3 つの目的を達成するために以下の 6 つの方法の検討および開発に取り組む。ここで方法 1 と方法 2 は目的(1)、方法 3～方法 5 は目的(2)、方法 6 は目的(3)に対応している。

1. Kinect の配置方法
2. Kinect からの距離画像を入力とした手領域の切り出し方法
3. 特徴抽出の方法
4. 標準形とその辞書の構成方法
5. 相互部分空間法をベースとした類似度計算方法
6. 照合結果のフィードバック方法

これらの方法の検討および開発を行うことで、実際に目的(1)～目的(3)を達成できるかどうか評価実験を行う。評価実験を行い、習得に要する時間および習得レベルが従来に比べて向上することを確認することで、試作した本システムが指文字の練習に有効であることを示し、本システムを使用することで初級レベルのユーザや中級レベルのユーザを常に標準形を呈示することができる上級レベルのユーザへと上達させることができるようにする。

1.4 本論文の構成

以下、本論文では、第 2 章で指文字練習システムを開発する際の難しさとその難しさに対処する練習システムの基本的な考え方および指文字練習システムの概要について述べる。第 3 章では Kinect を用いた指文字練習システムの詳細について述べる。第 4 章では実際に試作した本システムを用いて確認実験を行い、指文字の練習効果について評価を行う。第 5 章では、今後の課題について述べる。第 6 章では本論文のまとめについて述べ、本論文を締め括る。

第2章 Kinect を用いた指文字練習システムの概要

第2章では、Kinect を用いた指文字練習システムにおける開発の難しさとその難しさに対処するための基本的な考え方について述べる。基本的な考え方を述べた後は指文字練習システムを構成する各部の概要について述べる。

2.1 指文字練習システムの開発の難しさ

指文字練習システムを開発するためには、まず手形状を認識することが重要であるが、手形状を認識することは容易ではない。なぜならば、手形状を認識するときは、以下のよう
に3種類の変動があり、それらの変動に対応する必要があるからである。

- ・ 身体的特徴による個人差
- ・ 表現の個人差
- ・ 表現の経時変動

身体的特徴による個人差とは指の長さ、指の太さ、掌の大きさなどであり、これらの個人差に影響されず、手形状の認識を行わなければならない。

表現の個人差とは指文字を呈示するときの指の開き方、指の曲げ方および手の向きなどであり、様々な種類がある。さらに初級レベルのユーザが指文字を呈示する場合、指文字を一部誤って表現するだけでなく、指文字以外の手形状を呈示する可能性もあり、その種類は膨大な数となる。

表現の経時変動とは、図 2.1 のように時間が経過すると指文字の表現が少しずつ変化することである。これはユーザが無意識に指文字の表現を変えているため、時間が経過すると、以前は標準形を呈示できていたが、現在は標準形と一部異なる表現を呈示することがあると考えられる。したがって、その微妙な変動もシステムが認識できるようにする必要がある。

このように指文字練習システムを開発するにあたって、それらの様々な変動に対応できるように手形状の認識を行う必要がある。ただし、様々な変動に対応しながら手形状を認識できたとしても、次はユーザに標準形を呈示してもらうためのフィードバックを行う必要がある。標準形である、標準形ではないまたは他の指文字を呈示しているという照合結果をユーザにフィードバックするだけでは、ユーザは標準形とどの程度似ているか知ることができない。そのため、ユーザはどのように修正すべきか推測することができない。そこで、指文字練習システムの開発には、ユーザが標準形を呈示しているか認識するだけでなく、ユーザの手形状の表現が標準形とどの程度似ているか類似関係を定量的に測り、その情報をフィードバックすることも必要である。



図 2.1 経時変動による指文字表現の変化の例

さらに、ユーザが呈示した手形状を標準形であると評価するためには、比較の対象となる辞書のデータが必要である。辞書のデータを収集するとき、指文字を認識することが目的ならば、同じ指文字である場合に限り、表現が多様であるほうが指文字を認識しやすくなる。しかし、指文字を練習することが目的ならば、同じ指文字であっても、標準形と必ず一致したものでなければならない。なぜならば、ユーザにフィードバックを行うときに、類似関係を定量的に測る必要があるが、標準形ではない表現が辞書のデータに含まれる場合、その影響によってユーザの手形状の表現と標準形との類似関係を正しく評価することができない可能性があるからである。そのため、指文字練習システムを開発するための辞書のデータを収集する際には、標準形ではないデータを省くことが必要であり、データを収集することが難しくなる。

以上より、指文字練習システムを開発するには、手形状の様々な変動への対応、標準形とどの程度一致するか類似関係を定量的に測ること、辞書のデータの収集といった難しさがあり、これらの問題に対処しながら開発する必要がある。

2.2 指文字練習システムの開発における基本的な考え方

そこで、筆者は指文字練習システムを開発するために、画像認識技術を用いる。画像認識とはパターン認識のひとつであり、データを収集するときは特殊なデバイスを用いずに対象となるデータを撮影するだけで認識することができる技術である。画像認識は撮影した画像データとあらかじめ正解と設定した画像データのセットとの差異を検出することで正解データとの照合を行う、または撮影した画像データに対して様々な処理を行い、あらかじめ設定されたクラス（本研究では指文字）のどれに該当するのかという答えを出力する一連の処理を行う。ここで、正解と設定した画像データのセットはパターン認識の分野では辞書と呼び、辞書に登録しているデータのパターンが豊富であればあるほど、より幅広く認識することができる。また、辞書に登録しているデータのパターンが一様である場

合は、そのパターンと一致するものだけ認識することが考えられる。指文字練習システムを開発する際には、様々な人を対象に、かつ標準形のデータだけ収集し、辞書を構成することができれば、どのような人でも標準形を呈示していれば認識し、標準形を呈示していなければ認識しないことが可能になることが考えられる。このように、画像認識技術は特殊なデバイスを用いずに手形状の様々な変動に対応し、標準形とどの程度一致しているか検出することができる技術であり、指文字練習システムの開発に適していると考えられる。

ただし、画像認識技術を用いて手形状を認識する場合、前述の 2.1 で挙げた 3 種類の変動のほかに照明条件や背景の変動が追加される。例えば、手形状を認識するとき RGB 画像に写る手領域の輝度を用いる場合、朝と夜で照明条件が変動することで手領域の輝度も変動し、手形状の認識に影響することがあると考えられる。また、手形状を撮影するときは背景も同時に写るため、あらかじめ肌色などを対象に手領域を検出する必要がある。しかし、背景が肌色に近い色であるときは背景によるノイズが入り、手領域のみ検出することは難しくなる。さらに、辞書を構成するときは第 1 章の 1.3 で述べたようにユーザ独立の辞書を構成する必要があるため、対象となるユーザのデータを含めずに可能な限り上級レベルのユーザのデータを収集しなければならない。

そこで、赤外線を用いて距離情報を取得することができる Kinect を画像入力デバイスとして用いることで画像認識技術を用いる際の照明条件の変動や手領域のみを対象とした切り出しの問題に対応する。具体的には Kinect を用いて手形状を撮影し、Kinect からの距離が最短の部分からある幅を持った領域を手領域とみなして切り出しを行うことで、背景の変動に対応し、切り出した手領域の距離情報を認識に用いることで、照明条件の変動に対応する。

また、辞書データを構成するときは手話通訳者 12 名を対象にデータを収集する。手話通訳者は聴覚障害者にわかりやすく通訳するために指文字を常に丁寧に呈示する必要があるため、本システムが想定する上級レベルのユーザであると考えられ、手話通訳者の呈示する指文字は標準形の模範例として辞書のデータに用いることに適していると考えられる。

このように画像認識技術と Kinect を併用し、さらに辞書のデータ構成方法を工夫することにより、指文字練習システムを開発することが容易になると考えられる。

上記の問題に対応した場合の指文字練習システムを図 2.2 のようにイメージした。指文字練習システムは、ユーザ（指文字学習者）が画像入力部（Kinect）を用いて手形状を入力し、照合器（コンピュータ）がその手形状を照合し、その結果をユーザにフィードバックするという構成で考える。

文献 [4] の練習システムでは、ユーザは任意の指文字の手形状を呈示することができるようになっており、正しい指文字にガイドすることが必ずしも容易ではなかった。本システムでは対象とするユーザは指文字を完全に習得していない初級レベルから中級レベルのユーザに的を絞って、ユーザが呈示すべき指文字をシステム側から指定するようにして、フィードバック機能を設計する。

さらに、ユーザが実際に指文字練習システムを用いて練習する場合、練習の流れは以下のようにになると考える。

1. コンピュータが指文字を指定し、ユーザはコンピュータの指示を受けて一定時間手形状を呈示する。
2. Kinect を用いて手形状を撮影し、コンピュータが撮影した画像に写っている手領域に対して標準形と一致するかどうか照合を行うことで、ユーザが呈示している手形状が標準形であるかどうか判定する。
3. ユーザにその判定結果を出力し、標準形から外れていると判定した場合は、標準形からどの程度外れているかユーザが標準形を推測させることができるようにビジュアルにフィードバックする、あるいはシステム側でユーザの誤りを推測して、ユーザにその誤りを伝える（ガイダンスを行う）。
4. 1～3の動作を一定時間繰り返す。

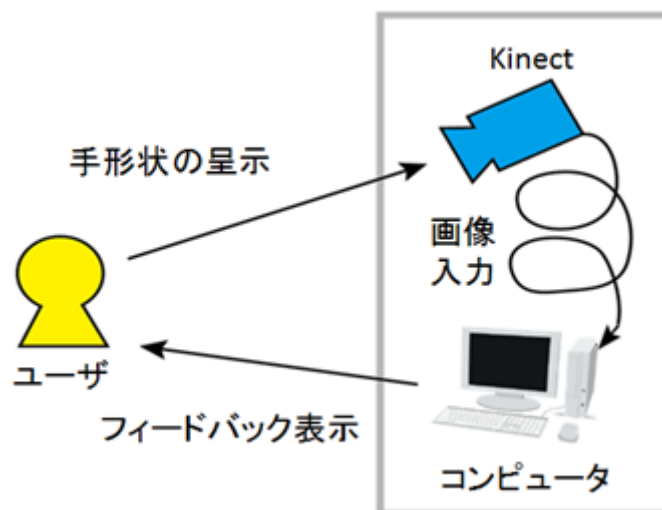


図 2.2 指文字練習システムのイメージ図

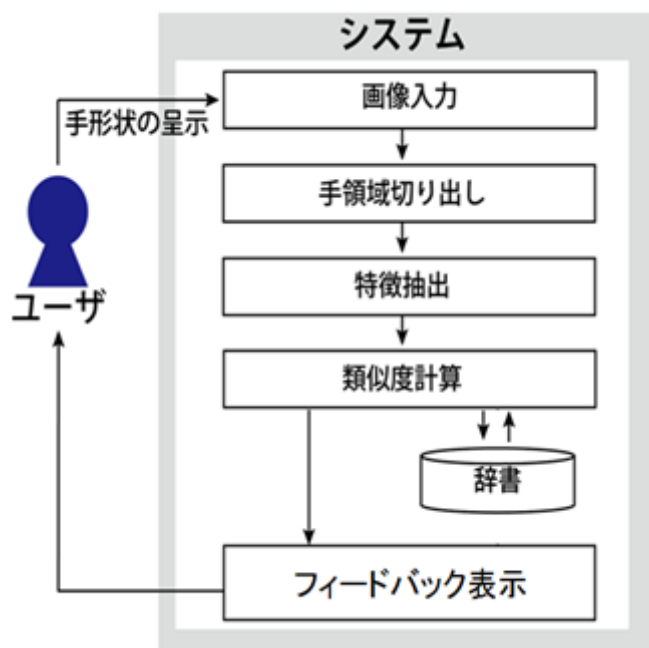


図 2.3 練習システムの処理の流れ

図 2.3 に画像認識技術と Kinect を併用した場合のシステムの処理の流れを示す。ユーザは手形状を Kinect に向かって呈示し、システムに手形状の画像を入力する。入力された画像に対して前処理として、手領域の切り出しが行われる。切り出された手領域に対して特徴抽出が行われ、入力画像の特徴量と辞書にある画像特徴量との類似度の計算が行われる。類似度の計算には辞書が必要となり、前述したように標準形のみを対象にデータを収集する必要があるが、その構成方法については第 3 章で詳述する。類似度が計算された後は、ユーザにその指文字の表現が標準形であるかどうかフィードバックを行う。このように指文字練習システムを構成することで、指文字練習システムを開発するときの難しさに対処することが可能であると考えられる。以降の第 2 章では練習システムの各部の概要を説明する。なお、手領域の切り出し部や特徴抽出部においては第 3 章で詳述する。

2.3 指文字練習システムの各部の概要

2.3.1 画像入力部

Kinect とは Microsoft 社が開発した RGB カメラと距離センサを搭載したデバイスであり (図 2.4)、このデバイスを用いることで RGB 画像と距離画像を得ることができる。Kinect から得られる距離画像の画素値は Kinect から物体までの距離を表し、40cm 以上の距離から取得可能である。本論文では Kinect から得られる距離画像の画素値を距離値と呼ぶ。なお、距離値と実際の距離の関係については巻末の付録 B に記載している。



図 2.4 Kinect の外観

2.3.2 辞書

本システムにおいて，収集した指文字データの学習と照合にはカーネル非線形直交相互部分空間法（KOMSM：Kernel Orthogonal Mutual Subspace Method） [6]を用いる．これは，文字認識や顔認識において，認識精度が良いとされている相互部分空間法（MSM：Mutual Subspace Method） [7]を拡張した手法のひとつであり，手形状の識別においても，その高い性能が大川らによって報告されている [8]．

初めに，MSM のもとになる部分空間について説明する．部分空間とは複数のデータの中から，それらのデータに共通する特徴を取り出したものである．部分空間の構成方法は主成分分析を行い，自己相関行列の対角化によって得られた主成分を部分空間の基底ベクトルとする [9]．

例えば，手領域を切り出したあとの N 枚の画像データから抽出した特徴ベクトル系列を X とすると， X は以下の式で表すことができる．

$$X = \{\vec{x}_1, \dots, \vec{x}_v, \dots, \vec{x}_n\} \quad (2.1)$$

この式を用いて自己相関行列 C を求めると，式 2.2 の計算式になる．

$$C = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \vec{x}_i \vec{x}_j^T \quad (2.2)$$

C を用いて以下の固有値問題を解くことで得られた固有ベクトル \vec{v}_j が部分空間の基底ベクトルとなる．ただし，添字 j は固有値 λ_j の大きな順に並べるものとする．

$$C\vec{v} = \lambda\vec{v} \quad (2.3)$$

式 2.3 により得られた固有ベクトルの系列で構成した部分空間はデータ間の局所的な違いを省いた大局的な特徴の集合と考えられる．この部分空間をテンプレートの集合と考えて識別を行うことで，未知のデータに対しても頑健に識別できると考えられる．

指文字練習システムにおいて、辞書を構成するときは、指文字ごとに複数のデータを集め、それぞれ部分空間をひとつずつ作る。これらの部分空間を辞書部分空間と呼び、この辞書部分空間を用いて指文字の照合を行う。

2.3.3 類似度計算部

類似度の計算部分には、前述で述べたように MSM を拡張した KOMSM を用いている。ここで、MSM について簡略的に述べる。MSM は前田らによって考案されたパターン認識の手法のひとつであり、図 2.5 に示すように辞書部分空間と入力部分空間が成す正準角を用いて 2 つの部分空間の類似度を計算する。辞書部分空間を構成するだけでなく、入力に対しても複数の画像で入力部分空間を構成することにより、1 枚のデータと照合を行うときよりも変動やノイズに強い識別を行うことができる。

いま、 M 次元の辞書部分空間 P と N 次元の入力部分空間 Q があるとき、その 2 つの部分空間の間には N ($M \geq N$) 個の正準角が存在し、このときに最も小さい角度をもつ正準角を第 1 正準角 θ_1 (最小正準角 θ_1) と呼び、以下、第 2 正準角 θ_2 , \dots , 第 N 正準角 θ_N と続く。 i 番目の正準角について式で書くと次のようになる。

$$\cos^2 \theta_i = \max_{\substack{u_i \perp u_j \\ v_i \perp v_j}} \frac{|(u_i \cdot v_i)|}{\|u_i\|^2 \cdot \|v_i\|^2} \quad (2.4)$$

このとき、 $i = 1, 2, \dots, N$ であり、 $i = 1, 2, \dots, N-1$ である。また、 u_i は P に含有されるベクトルであり、 v_i は Q に含有されるベクトルである。ただし、 $u_i \neq 0$, $v_i \neq 0$ とする。

ここで、部分空間同士の類似度 sim は以下の式 2.5 を用いる。本論文では、式 2.5 により、全ての正準角を用いて類似度を計算する。計算した類似度 sim は 0~1 の範囲となり、1 に近ければ近いほど、その 2 つの部分空間は類似しているということになる。

$$sim = \frac{1}{N} \sum_{i=1}^N \cos^2 \theta_i \quad (2.5)$$

本研究において、対象となるデータは動画画像である。様々な視点から入力した距離画像や指の長さや太さなどの変動のある距離画像を収集し、それを学習することで、視点や個人差などの変動に対して頑健な照合が可能となる。また、標準形のデータのみ収集し、それを学習することで、標準形と高精度に照合を行うことが可能となる。MSM における学習の際には、指文字ごとに入力された動画画像から辞書部分空間を生成する。照合の際には未知の手形状の動画画像を入力し、ここから入力部分空間を生成する。そして、各辞書部分空間との類似度を計算し、類似度の大きさによって照合結果を出力する。

このような手順で画像の識別を行う MSM は、線形分布に対する識別においては非常に良い性能を示すが、手形状などのような非線形分布においては高い識別精度が期待できない。その理由として、非線形に分布する場合、別のクラス同士が重なってしまい、学習時

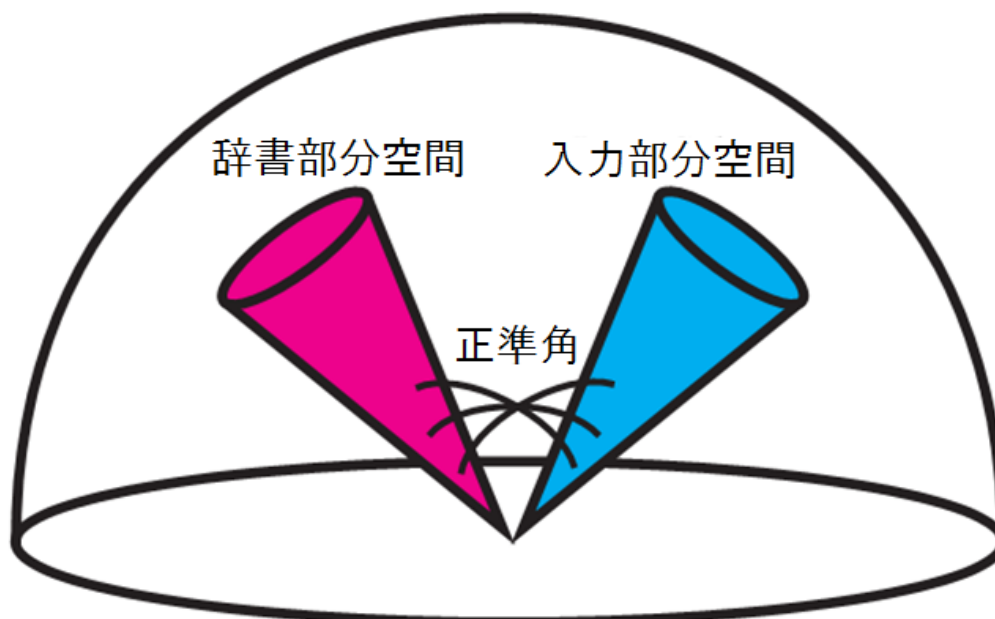


図 2.5 相互部分空間法のイメージ図

に上手くクラスごとに分割することができない。そのため、認識性能が落ちてしまう。そこで、非線形に分布する物体においても高い精度での識別を可能にするために MSM を拡張した KOMSM が考案された。KOMSM では非線形に分布するパターンをより高い次元または無限次元の空間に非線形写像を行う。その写像されたパターンに対して MSM を適用させ、高精度な識別を可能としている [8]。しかしながら、KOMSM は学習時に多くの計算量を必要とするため、学習データ数が多い場合には適応が難しい。そこで、本システムでは、文献 [8]を参考にして全学習データに対して K-means 法を用いてクラスタリングを行い、学習データ量を削減することにより、KOMSM の高速化を図っている。

なお、本システムにおいて手領域の切り出し部から類似度計算部までは文献 [10]と同じ方法を用いた。文献 [10]では距離画像から手領域を切り出し、切り出した画像を 30×30 pixels の大きさに正規化したあとに、View 特徴を用いて指文字を識別している。この識別処理を行うことにより、静止指文字 41 種類に対し、高い認識率を得ることが報告されている。

2.3.4 フィードバック表示部

フィードバック表示を行うに当たって、以下の情報をユーザに伝える必要があると考える。

- I. 呈示している指文字は正解であるか
- II. 呈示している指文字の表現は標準形であるか

- III. 呈示している指文字の表現が標準形ではない場合，標準形からどの程度誤っているか
- IV. 呈示している指文字の表現が標準形ではない場合，標準形を呈示するためにどのように修正すべきであるか

本システムでは I～III までの情報をユーザにフィードバックする処理を開発した。このフィードバックの処理については，第 3 章で詳述する。ただし，IV の情報をフィードバックする処理はまだ実装しておらず，将来的には IV の情報も提供することができるシステムを開発したいと考える。

第3章 Kinect を用いた指文字練習システムの詳細

第3章では、Kinect を用いた指文字練習システムの処理方法の詳細を述べる。指文字練習システムのハードウェアは画像入力部（Kinect）と計算機本体、ディスプレイから構成されており、図 3.1 にその外観を示す。

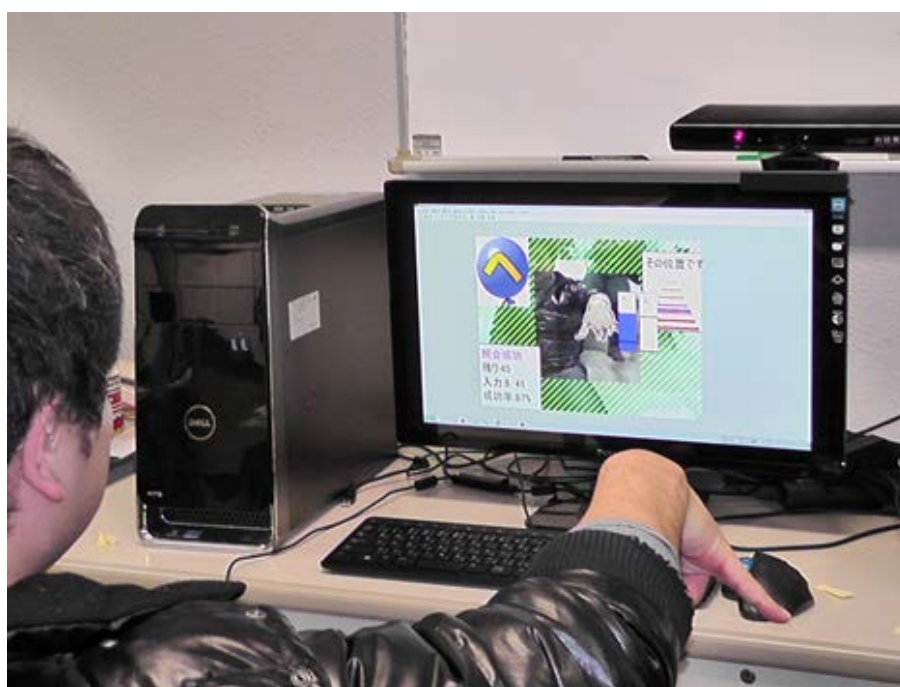


図 3.1 指文字練習システムの全体図

3.1 Kinect の配置

Kinect を用いて画像入力を行うときは、図 3.2 のように Kinect をディスプレイの上に設置し、斜め下に 20° 傾けて撮影を行う。ここで、斜め下に配置した理由は、撮影時に手首や腕の一部が手形状に隠れて画像に写らなくなるため、標準形との類似度を計算するときに手首や腕の一部が影響することが少なくなるからである。

実際に Kinect の距離画像から 80cm 以上に相当する距離値を背景領域とし、背景領域以外の領域を抽出すると、図 3.3 のように Kinect を 20° 傾けて撮影した場合と、地面と水平に向けて撮影した場合において、前者は手首や腕の一部が写らなくなるが、後者は手首や腕の一部が写っており、類似度の計算に影響を与えてしまうことが考えられる。

また、腕を伸ばして手形状を呈示することで、手首の部分を写さず手形状を照合することも可能であるが、長時間指文字の練習を行う場合は、その間に腕を伸ばし続けて手形状を呈示する必要があり、被験者の肩に負担がかかってしまい、手軽に練習することが難しくなる。そこで、Kinect の持つ手軽さを発揮し、かつ指文字の照合を高精度に行うために、Kinect を斜め下に 20° 傾けることで手軽に指文字の練習が行えるように設定する。

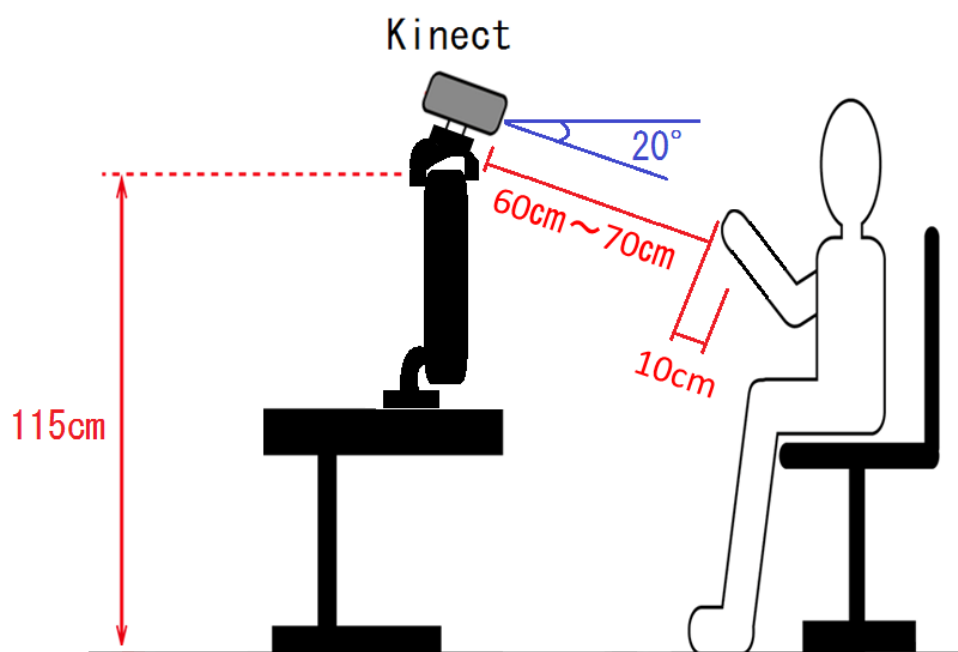


図 3.2 Kinect を用いた手形状の撮影方法

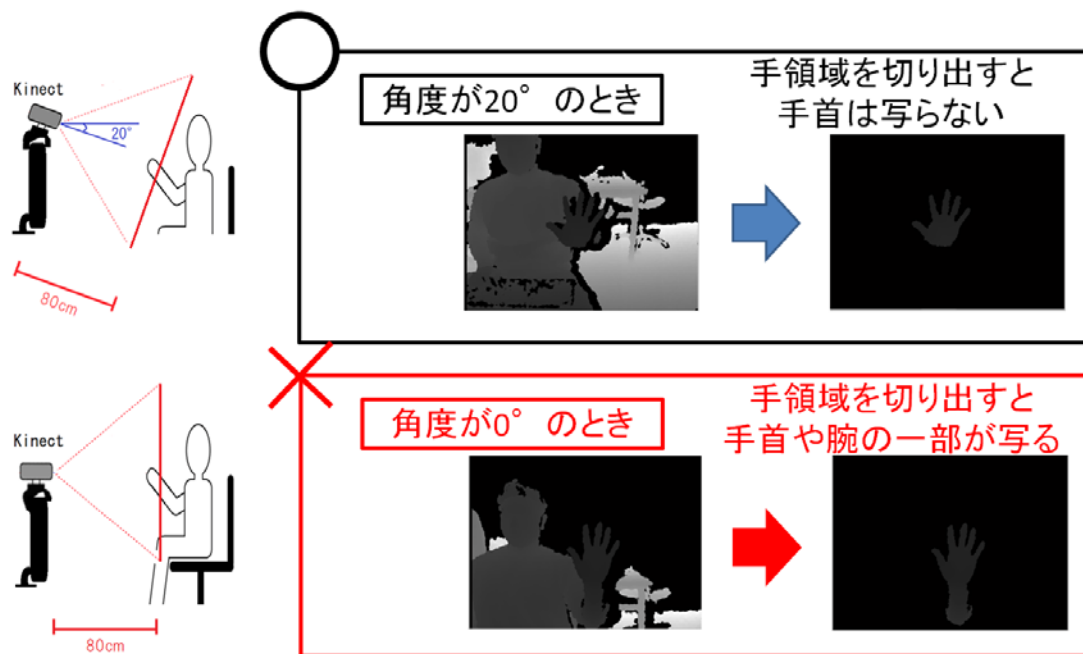


図 3.3 Kinect の傾き方による手領域の写り方の違い

3.2 手領域の切り出しと正規化

手領域の切り出しには Kinect から得られる距離画像を用いた。距離画像を照合に用いる場合、距離の情報を用いているため、RGB 画像を用いる場合と比べて照明条件の変動による影響が少ない。RGB 画像を照合に用いる場合、外光によって手領域の明度が変動することがあり、辞書のデータを収集した時間帯によっては、朝は照合することができるが、夕方は照合することができなくなることがあると考えられる。そのため、Kinect から得られる距離画像を照合に用いることにより、照明条件による変動の影響を少なくすることができるため、RGB 画像を照合に用いる場合より安定して指文字の練習を行うことができるメリットがある。

手領域の切り出しの流れは図 3.4 の通りであり、撮影時は Kinect からの距離が最短である物体を手とみなして手領域の切り出しを行う。まず、距離画像に対して距離値の最小値 +100 (距離値の 100 は約 10cm に相当する) より大きい距離値をすべて 0 (背景領域) とする。次に背景領域以外の領域を手領域とみなし、手領域において最大の縦幅または横幅のどちらか大きい方に合わせ、正方形で切り出しを行う。

ただし、切り出した画像をそのまま照合に用いると、切り出した幅の大きさによって画像の大きさが変動するため、類似度の計算に影響することが考えられる。また、Kinect からの距離によって距離値が変動するため、同様に類似度の計算に影響することが考えられる。

そこで、手領域を切り出した画像 (手領域画像) に対し、以下の 2 つの処理を行った。

- ・ 画像の大きさの正規化
- ・ 距離値の正規化

画像の大きさの正規化とは手領域を切り出した後、 50×50 pixels の大きさに縮小する処理である。この処理を行うことで、画像の大きさの変動による影響を抑えることができる。

距離値の正規化とは切り出した手領域画像に対し、以下の流れを行う処理である。

1. 手領域画像全体の距離値を合計する。
2. 手領域 (距離値が 0 ではない) の画素数を計算する。
3. 1 の合計に対し、2 で計算した画素数で割り、その数値を距離値の平均値とする。
4. 手領域の距離値から平均値を引き、100 を足す。

この正規化の処理を行うことにより、距離の変動による影響を抑えることができる。手形状の照合には 2 つの正規化を行った手領域画像を用いる。

Kinect は 40cm から距離情報を得ることができるが、40cm~50cm 付近では図 3.5 のように指の欠損といったノイズが大量に発生する。そこで、本システムでは手が Kinect から 60cm 以内にあるときは手領域の切り出しを行わないように条件を指定する。

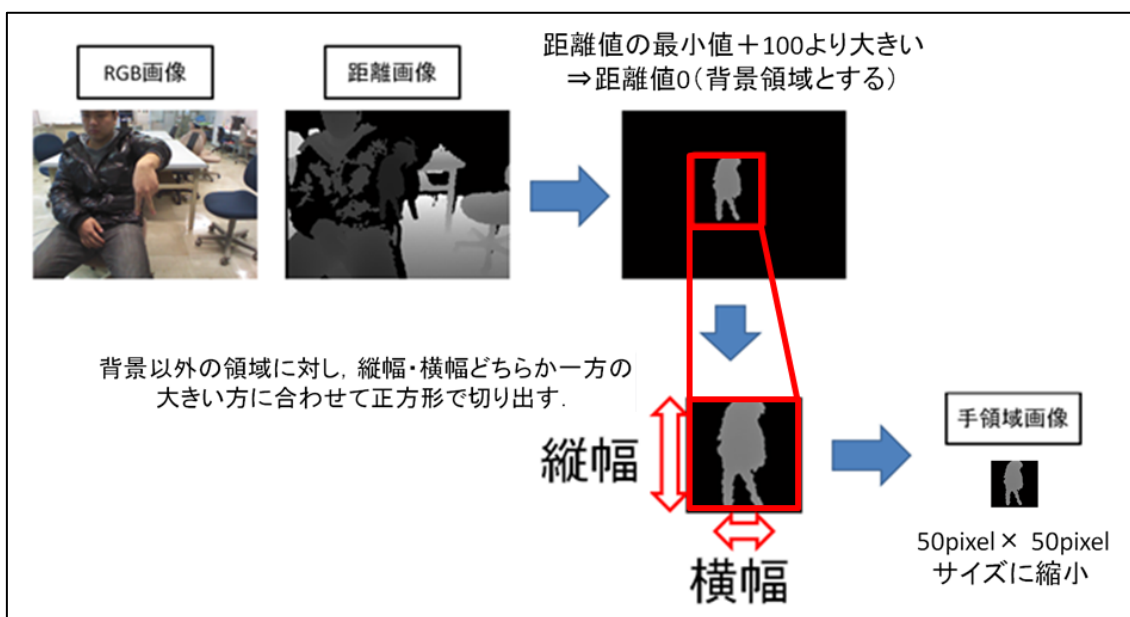


図 3.4 手領域の切り出し処理の流れ

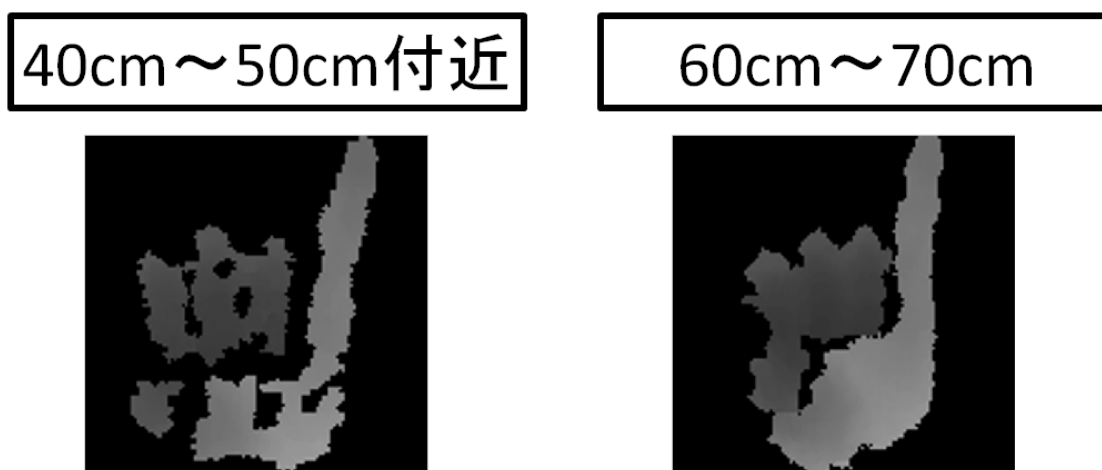


図 3.5 ノイズがある図とノイズがない図

3.3 特徴抽出

手領域の特徴抽出は、距離値の行列を列ベクトルに変換したものをを用いる。前述の 3.2 の処理により切り出した手領域画像は距離値の行列であり、図 3.6 より距離値の行列を変換してベクトル化を行う。本論文ではこの特徴量を **View 特徴** と呼ぶ。

文献 [4]では、手領域画像の特徴抽出に高次局所自己相関特徴量 (**HLAC : Higher-order Local Auto-Correlation**) [11]を用いていた。HLAC は位置不変性という特徴を有しているため、撮影時の手の位置に依存せずに特徴量を抽出することができる。しかし、ユーザへフィードバックを行うためにはユーザが呈示した手形状と標準形の指の位置関係や開き方などの差を検出し、指文字の誤りを教えることができるようにする必要がある。HLAC は位置不変性という特徴を有しているため、指の位置関係や指の開き方といった差を検出できない可能性がある。

指文字の表現において、指の位置関係や指の開き方などは重要な意味を持つ。例えば、中級者が「あ」を呈示した場合、親指の位置が図 3.7 のように「あ」、「さ」にどちらも受け取れる曖昧な表現となることがある。親指を大きく開いた場合は「あ」になるが、親指を閉じた場合は「さ」になるため、指の開き方によって意味が変わる。本システムにおいて、曖昧な表現と標準形との微妙な差を検出し、標準形ではないことをユーザに伝える必要があり、指の位置関係や開き方などがわかる特徴量を選択する必要がある。

View 特徴 は各画素の距離値を列に並べ替えただけであり、画素同士の順序情報を持つ特徴量である。そのため、**View 特徴** を標準形との照合に用いることで、画素同士の順序情報により指の位置関係や開き方などの差を検出することができるようになることが考えられ、本システムの特徴量に適していると考えられる。

View 特徴 を用いる場合、特徴量の次元数は画像の大きさに依存し、画像が $10 \times 10 \text{pixels}$ の大きさならば、次元数は $10 \times 10 = 100$ 、 $100 \times 100 \text{pixels}$ の大きさならば、次元数は $100 \times 100 = 10000$ となる。このように手領域画像の大きさが変わると照合結果に影響を与えることが考えられる。そこで、手領域画像の大きさはあらかじめ実験的に決める必要があり、その決定方法は付録 D.1 で記述する。

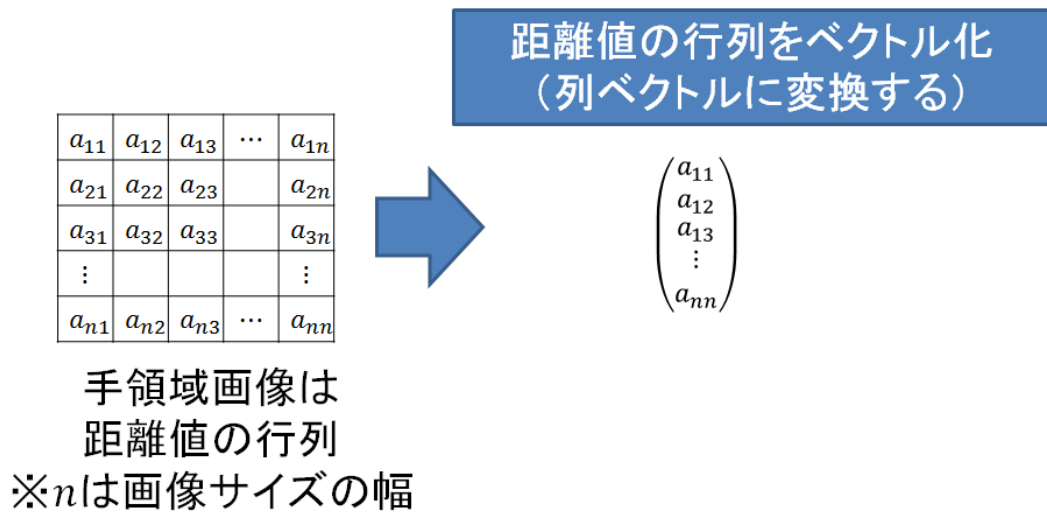


図 3.6 特徴抽出の流れ

初級者や中級者が
指文字を曖昧に
出してしまうとき

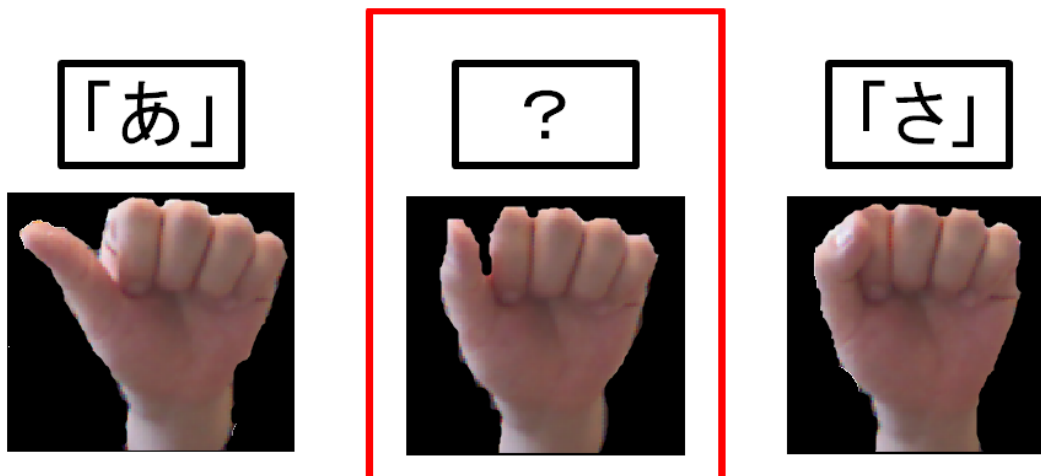


図 3.7 中級者が表出した場合の曖昧な表現

3.4 辞書の構成

辞書の構成には手話通訳者 12 名のデータを用いる。手話通訳者は第 2 章の 2.2 で述べたように聴覚障害者にわかりやすく通訳するために指文字を常に丁寧に呈示する必要があり、本システムが想定する上級レベルのユーザであると考えられる。

手話通訳者には清音指文字 46 文字のうち、動きのある指文字 5 文字（「の」、「も」、「り」、「を」、「ん」）を除いた 41 文字を対象に、右手で標準形を呈示するように依頼する。一文字あたり平均 4360 枚のデータを収集する。また、データを収集するときは図 3.8 のように手を画像の中心に呈示したあと、手をらせん状に右回りで 3 回、肩を軸に動かすように依頼する。辞書のデータを収集するときに手を動かす理由は、Kinect に対する手の向きが変動しても手形状が照合できるようにするためである。Kinect に対する手の向きによって、距離画像に写る手領域の形状が変動し、類似度の計算に影響を与えることが考えられるため、あらかじめ様々な手の向きを辞書に追加することにより、手の向きが変動しても手形状を照合できるようにする。

らせん状に手を動かすと様々な方向から撮影したデータを取得することができるが、手を動かし過ぎると角度によって他の指文字に似てしまうデータが存在するため、そのようなデータに対しては辞書から省くように確認する必要がある。例えば、図 3.9 のように「よ」を大きく動かした場合、手を大きく下に向けると小指の部分が隠れてしまい、「み」の文字に似てしまうことがある。そこで、データを収集するときは筆者が常に撮影画面に注視し、他の指文字に似てしまうまたは標準形ではない表現があった場合は、そのデータを省く。

このような方法で収集したデータを辞書データとし、この辞書データを用いて K-means 法により指文字に対応するデータごとに枚数の削減を行い、そこから 41 文字に対応する 41 種類の辞書部分空間を構成する。以降の実験はすべてこの辞書部分空間を用いて、ユーザが呈示した手形状との類似度を計算する。

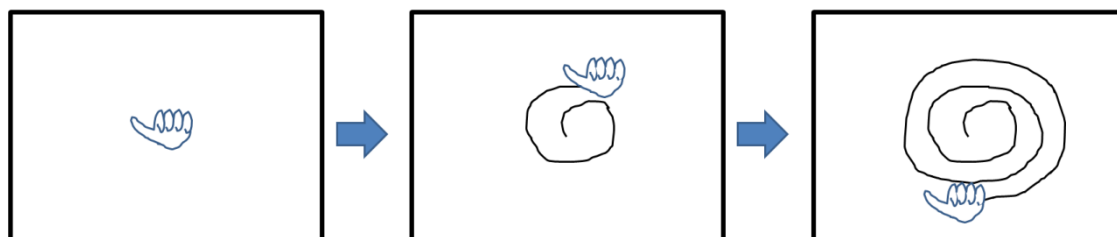


図 3.8 データを収集するときの手の動かし方

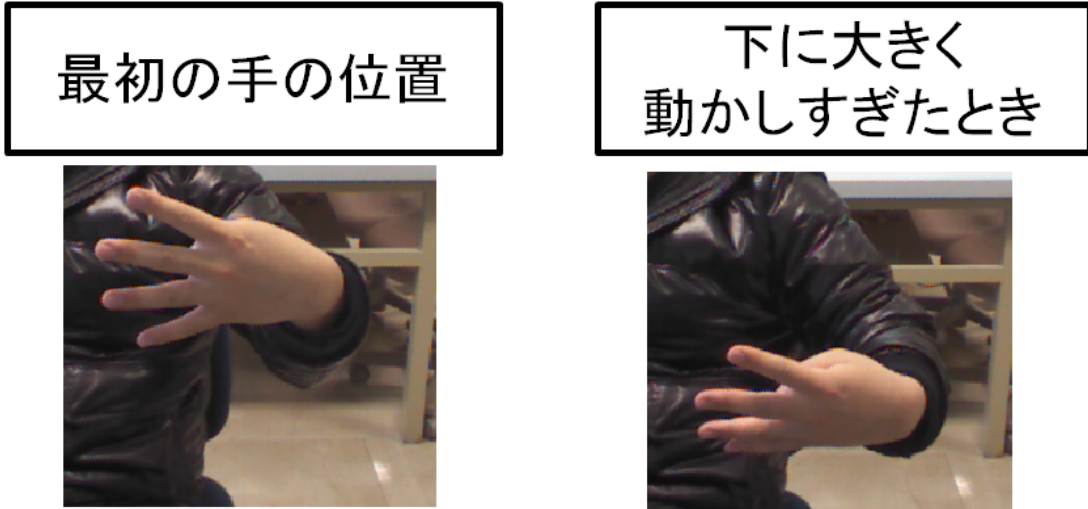


図 3.9 手を動かし過ぎた例（「よ」が「み」の文字に見える）

3.5 照合パラメータの決定方法

ここでは、指文字練習システムの照合処理において、ユーザに適切なフィードバックを行うために最適な方法を検討する。標準形との照合を行うにあたって、前述の 2.3.4 で述べたように以下の情報をユーザにフィードバックできるようにする必要がある。

- I. 呈示している指文字は正しいか
 - II. 呈示している指文字の表現は標準形であるか
 - III. 呈示している指文字の表現が標準形ではない場合、標準形からどの程度離れているか
- ただし、相互部分空間法をベースに指文字の照合を行う場合、類似度を計算することができるが、類似度は辞書データと入力がある程度似ているか表しているだけであり、システムはどの程度似ていれば標準形と判定するべきかどうか分からない。そのため、システムに標準形と標準形ではない表現の基準を教える必要がある。

そこで、類似度を用いて、照合成功の条件を以下のように指定することで、システムに標準形と標準形ではない表現の基準を教えた。 Sim_1 は指定した指文字の類似度、 Sim_2 は指定した指文字を除く 40 文字の中で最大の大きさである指文字の類似度である。また、 $Diff$ は Sim_1 と Sim_2 の差である。

- (ア) 全ての指文字の中で類似度が最大の指文字（第 1 位）は指定した指文字と一致する。
- (イ) Sim_1 が閾値 Th_1 以上である。
- (ウ) $Diff$ が閾値 Th_2 以上である。

(イ)の条件は、指定した指文字を標準形で呈示すれば、ある程度類似度が高い、つまりある程度標準形の表現と一致しているという考え方をもとに設定する。(ウ)の条件は、指定した指文字を標準形で呈示すれば指定した指文字以外の類似度と差がある、つまり他の指文字の表現に受け取ってしまう曖昧な表現ではないという考え方をもとに設定する。本論文ではこの 2 つの考え方から、ユーザが呈示した手形状の表現がこれらのすべての条件を満たすとき、ユーザの指文字は標準形を呈示していると考えられる。閾値 Th_1 と閾値 Th_2 の決定方法は第 4 章で詳述する。

図 3.10 より(ア)~(ウ)の条件をもとにした判定処理の流れを示す。4 種類の判定結果はそれぞれ標準形との近さの度合いに応じたものとなっており、判定結果をフィードバックすることで上記の I~III に関係した情報を提供することができると考えられる。このように照合時に条件を定め、条件の組み合わせごとに標準形との度合いの基準を定義することで、ユーザが呈示した指文字の表現が標準形とどの程度離れているか推測することができる。

以上より、閾値 Th_1 と閾値 Th_2 を決定し、照合成功の条件を決めることでシステムに標準形と標準形ではない表現の基準を教える。システムはこの基準を用いてユーザへのフィードバックを行う。

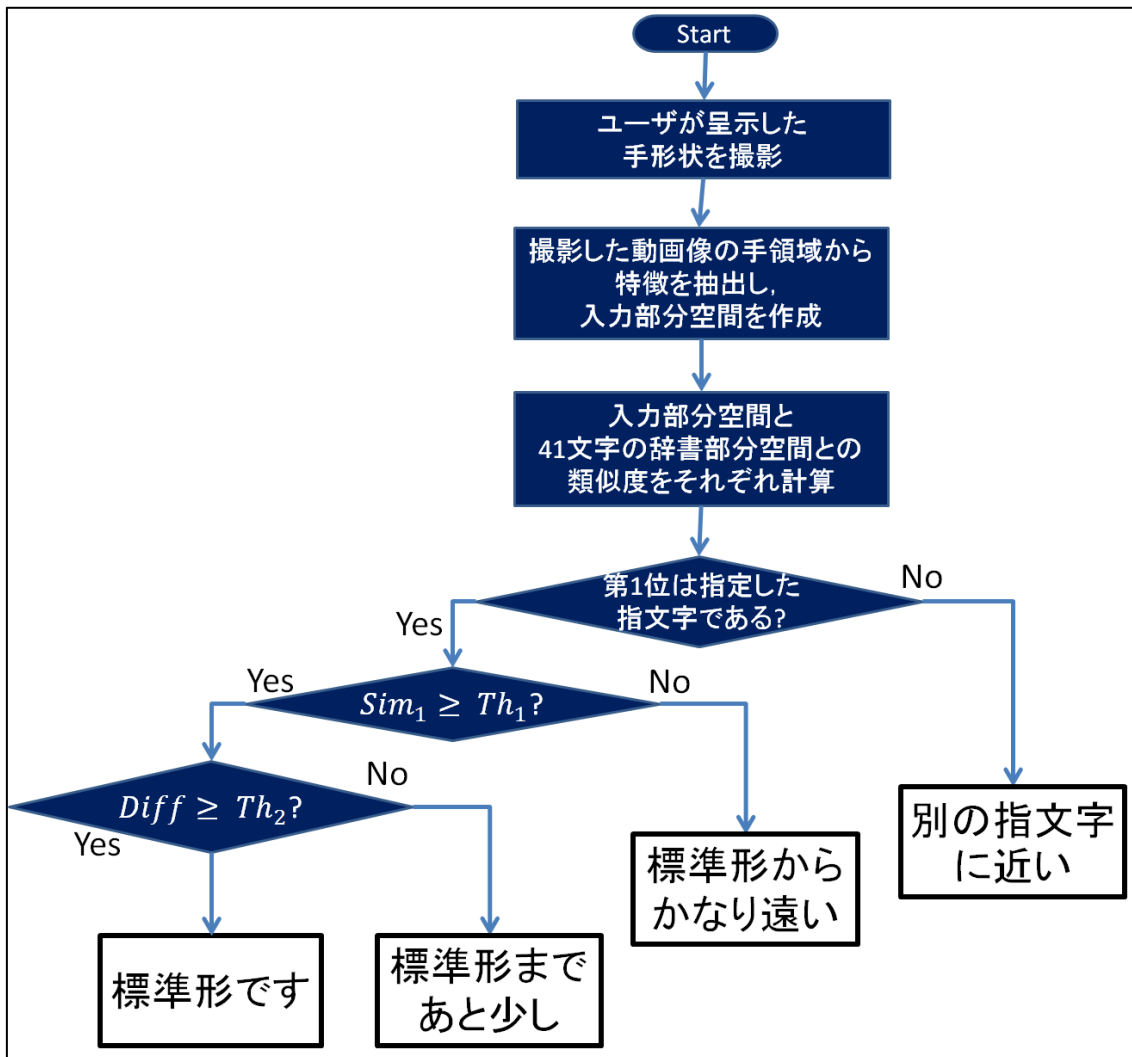


図 3.10 判定処理の流れ

3.6 フィードバック画面の構成と UI の開発

図 3.10 の判定処理の結果を用いてリアルタイムでフィードバックを行うユーザインタフェース（対話処理）について述べる。まず、本システムの練習の流れは以下の通りである。

1. システムが指文字をランダムに指定する。
2. ユーザが指定した指文字の指示に従って手形状を呈示すると、照合を開始する。
3. 指定した指文字に関して照合成功の条件をすべて満たせばその指文字はその指文字は照合成功とし、一定時間呈示しても照合成功の条件を満たさない場合は照合失敗とする。
4. 照合が終了すれば、それまでの照合結果、類似度、照合するまでの経過フレーム数などを記録する。
5. 1～4 の行程を 41 文字繰り返す。
6. 5 の行程が終了したら、41 文字の判定結果をリストで表示する。

4 に関して、照合成功率や照合が終了するまでのフレーム数を記録することにより、ユーザの練習効果を評価することができる。例えば、初級者がはじめて練習システムを使用するときは照合成功率が 60% で、照合が終了するまでの指文字ごとの平均フレーム数が 30 フレームとする。10 回目に使用するときには照合成功率が 85% で、平均フレーム数が 10 フレームならば、初級者は正確に早く指文字を標準形で呈示することができるようになったと評価することができる。なお、本システムは 1 秒間に 5.7 フレームの入力であるので、平均フレーム数が 10 フレームならば、ユーザは 2 秒以内に標準形を呈示することができると考えられる。

ここで、1～5 の行程のときのユーザへのフィードバック画面について述べる。図 3.11 にフィードバック画面を示す。以下はフィードバック画面で構成されている箇所を説明するものである。

- ① 指文字の指示
- ② Kinect からの距離メータと手の位置を誘導するメッセージの出力
- ③ 手領域の輪郭
- ④ 現在の照合結果の出力
- ⑤ 類似度を表す棒グラフ
- ⑥ 判定結果の色別表示

②は付録 B を参考に Kinect からユーザの呈示した手までの距離をメータで表したものである。図 3.11 のようにバーを下から順に 10cm 間隔で 40cm, 50cm, …, 110cm の位置に描画し、60cm と 70cm に対応するバーを青線、それ以外のバーを赤線で描画する。ユーザの手の現在位置は黒線で描画しており、Kinect から手を離せば黒線が上に移動するように見え、Kinect に手を近づけると黒線が下に移動するように見えるため、ユーザの手が Kinect からどの程度離れているかどうか直感的に確認することができる。本システムは Kinect か

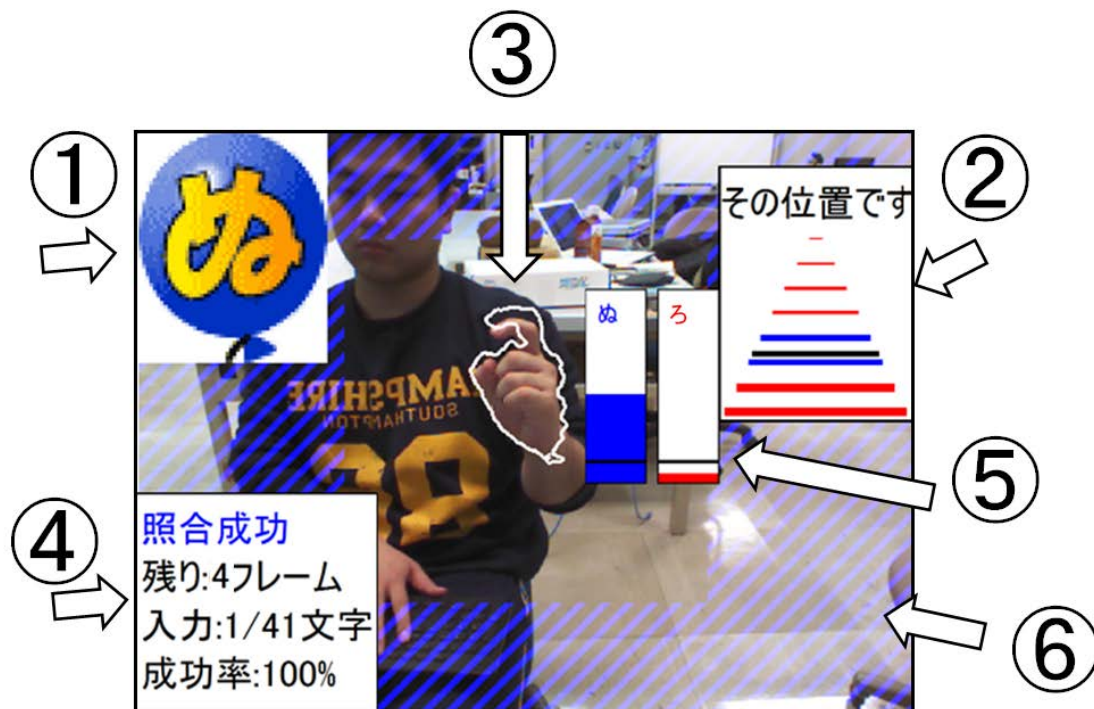


図 3.11 フィードバック画面の例

ら手の間の距離が 60cm~70cm 以内にあり、ユーザが手形状を画面中央に呈示しているときに照合を行うように条件を指定する。条件を指定することで、距離や手形状の Kinect に対する向きの変動を抑える。なお、照合を行う条件を満たすときは③を表示し、照合を行う条件を満たさないときは③を表示しないように設定する。距離メータの上の誘導メッセージは図 3.12 の通りであり、Kinect からの距離、画面の中の手の位置に応じてメッセージを出力することで、ユーザが自ら照合を行う条件を満たすように誘導する。

④は照合結果、照合が終了するまでの残りフレーム数、練習が終了した文字の数、照合成功率が表示される。特に、残りフレーム数は照合が終了するまでの制限時間を表し、手領域の検出が 1 回成功するごとに残りフレーム数が 1 つずつ減り、残りフレーム数が 0 になるまでに照合成功の条件を満たすフレームがなければその指文字は照合失敗となり、練習が終了する。なお、照合時は第 4 章の 4.1.1 で述べる入力部分空間の構成方法を用いており、4 フレーム前から現在のフレームまでの 5 枚の手領域画像セットを用いて入力部分空間を構成し、類似度を計算するようにする。そのため、本論文では照合開始を 5 フレーム、照合終了を 50 フレームとする。

⑤は標準形とユーザの手形状との類似度を表示するものであり、左の棒グラフは Sim_1 、右の棒グラフは Sim_2 を表す。今回の図 3.11 では「ぬ」を呈示したとき、「ぬ」と指文字の表現が似ている「ろ」の類似度が右の棒グラフに表示される。棒グラフには黒線が描画されているが、左の棒グラフの黒線は照合成功の条件である閾値 Th_1 を表し、右の棒グラフの

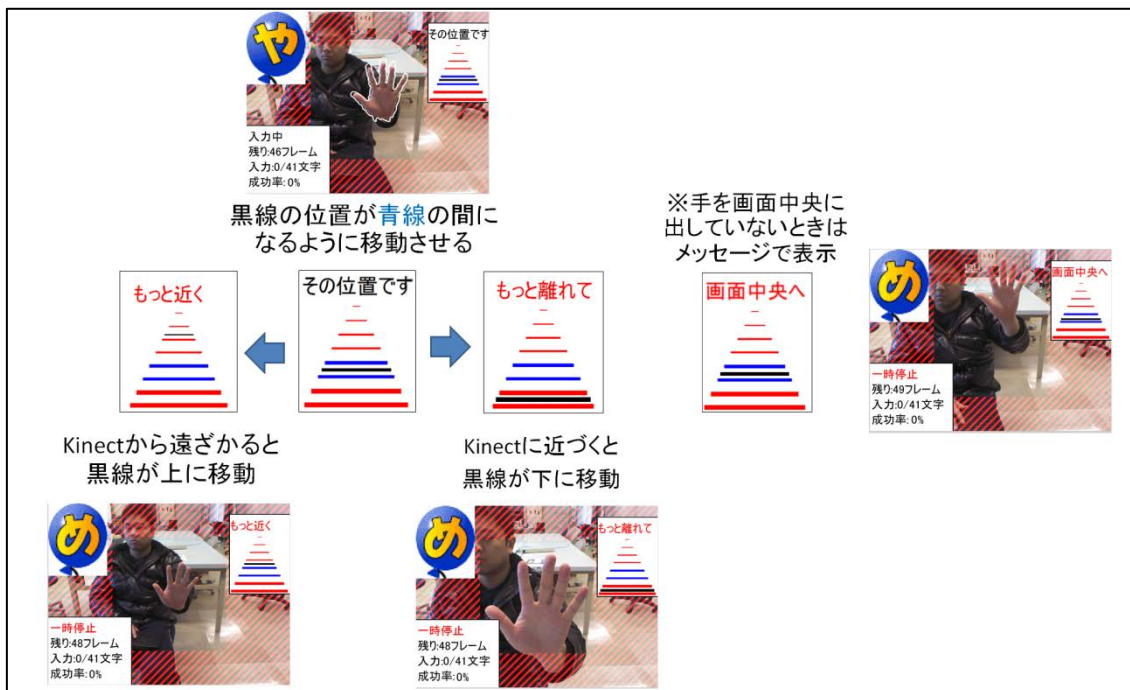


図 3.12 誘導メッセージの種類と実際の例

黒線は Sim_1 と閾値 Th_2 の差を表す。また、右の棒グラフの黒線はフレームごとに Sim_1 が計算されるため、フレームごとに Sim_1 と閾値 Th_2 の差に合わせて動く。例えば、 Sim_1 の類似度の計算結果が0.7で閾値 Th_2 が0.2の場合、右の棒グラフの黒線は類似度が0.5の位置に描画される。図3.13に類似度の棒グラフと判定結果を示す。この棒グラフを確認することで類似度が変動していることがわかるため、ユーザに指文字の表現に関する情報を提供することができる。また、棒グラフを表示するときは常に検出した手領域の隣に表示することにより、類似度の変動と指文字の表現を同時に確認することができるようにする。なお、類似度が計算されるのは5フレーム以降であるので、4フレーム目までは⑤を表示せず、5フレーム目で⑤を表示するように設定する。

ここで、文献[4]では標準形とユーザの手形状の類似関係を表すために形状相関マップを用いることを提案していた。しかし、形状相関マップは初級者にとって扱いにくいものであると考えられる。なぜならば、形状相関マップは2次元または3次元であり、類似関係をマップ上の位置で表しているため、標準形を呈示するためには大きさや方向を確認する必要があり、複雑であるからである。一方、本システムではフィードバック画面の⑤より、類似関係を棒グラフで表しているため、標準形を呈示するためには大きさのみ確認すればよく、シンプルである。本システムの対象となるユーザには初級者が含まれていることから、類似関係を表すときは、類似度の棒グラフを採用する。

⑥は判定結果を表し、図3.10の判定結果により、斜線の色が図3.14のようにリアルタイムで変わる。赤、黄色、青の3色で表現することによって、ユーザが呈示した指文字が標

準形であるかどうか、または標準形からどの程度離れているかどうか直感的に理解することができる。

6 の行程のとき、判定結果をリストで示すフィードバック画面は図 3.15 の通りである。判定結果の種類ごとに該当する指文字を 4 分割で表示することにより、標準形から外れている指文字は何か、ユーザに一目でわかるように表示する。ここで、判定結果のリストを表示するときは、4 種類の判定結果ごとに標準形に最も近い順に順位をつけ、50 フレームの中で最も順位が高いときの判定結果を対象に指文字を振り分ける。

このような方法でフィードバック表示を行うことにより、ユーザが指文字の表現の誤りに気づかせる、または修正させるように支援を行うことができる。

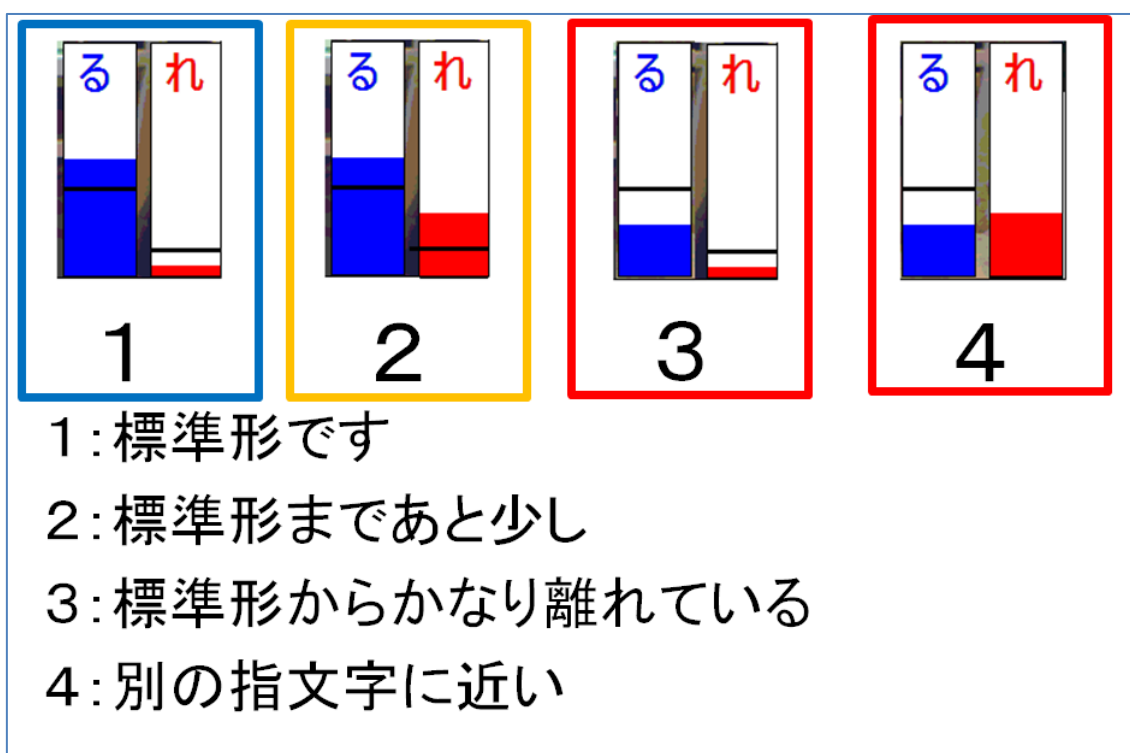


図 3.13 類似度の棒グラフと判定結果



図 3.14 判定結果の色別表示

別の指文字に近い	標準形からかなり離れている	標準形まであと少し	標準形です
いうけさと にはふほゆ ら	こ	きせぬる	あえおかく しすそたち つてなねひ へまみむめ やよれろわ

図 3.15 判定結果のリスト表示

第4章 試作システムの実験的評価

第4章では、試作した本システムを用いて評価実験を行うことにより、第1章の1.3で述べた目的が達成できたかどうかを確認する。

- (1) 安価でユーザへの負担が少ない画像入力デバイスとしてゲームなどのエンターテインメント用として一般的に普及している「Kinect」を用いて、安定して指文字を表す手領域を切り出し、手領域を効率的に抽出する。
- (2) 手領域特徴の照合方法とユーザ独立の辞書における構成方法の検討を行い、ユーザが呈示する手形状と標準形との3次元形状の照合を高精度に行う。
- (3) 照合結果をリアルタイムにフィードバック表示するための方法とユーザインタフェースを検討し、双方向の情報のやり取りを可能とする。

第3章では(1)の目的を達成するために、3.1のKinectの配置および3.2の手領域の切り出し方法を検討した。また、(2)の目的を達成するために3.3の特徴抽出、3.4の辞書の構成および3.5の照合パラメータの決定方法を検討した。さらに、(3)の目的を達成するために3.6のフィードバック画面の構成とUIの開発を行った。そこで、後述の4.1で辞書構成に協力した手話通訳者、聴覚障害者を対象に本システムを使用した場合の照合成功率を調べることにより、3.1~3.5の有効性の確認を行う。また、4.2で指文字の練習効果に関する評価実験を行うことにより、3.6の有効性を確認し、本システムは指文字の練習に有効であることを示す。

ここで、本システムで用いた計算機本体はIntel Core i7 (3.40GHz)のCPUに、メモリが12GB、OSはMicrosoft Windows 7 UltimateのデスクトップPCである。また、指文字練習システムのソフトウェアは、画像の入力処理においてはKinectによる撮影を制御するデバイスドライバであるKinect for Windows SDK v1.6を用いており、指文字の照合処理と照合結果などのフィードバック表示処理ではMath Works社のMATLABを用いて作成した。

4.1 リアルタイム照合実験

ここでは、4.2で指文字の練習効果に関する実験を行う前に、以下の3つを評価実験で確認する。

1. 指定した指文字と異なる文字、または指文字以外の手形状を呈示しても照合しない。
2. 辞書構成に協力した手話通訳者ならば、標準形を呈示して照合する。
3. 辞書構成に協力していない聴覚障害者でも標準形を呈示すれば照合する。

ここで、1を確認する理由はユーザが指定した指文字と異なる手形状を呈示し、誤って照合が成功してしまった場合は、ユーザはその誤って呈示した手形状が標準形であると解釈

してしまうため、それを未然に防ぐ必要があるからである。

4.1.1 閾値 Th_1 と閾値 Th_2 の決定

まず、第 3 章で構成した辞書を用いて、辞書のデータ収集に協力してもらった手話通訳者から任意で選んだ 2 名と日常的に指文字を使用する聴覚障害者 2 名に対し、条件(イ)の閾値 Th_1 、条件(ウ)の閾値 Th_2 を決定した。このとき、閾値 Th_1 と閾値 Th_2 を決定するときは、実際に指定した指文字と異なる文字、または指文字以外の手形状を呈示した場合に誤って照合しないようにする。

データは以下の 2 種類を収集し、閾値 Th_1 の決定にはデータセット 1 を用い、閾値 Th_2 の決定にはデータセット 2 を用いた。ここで、1 セットとは 1 文字に対し、それぞれ 50 回撮影した枚数のことである。また、データは全て右手で呈示した。

1. 手話通訳者 2 名の 41 文字のデータ×5 セット
2. 聴覚障害者 2 名の 41 文字以外の 36 種類の手形状のデータ×2 セット

データセット 2 は図 4.1 の手形状のパターンであり、初級者が指文字の一部を間違えて文字を呈示した場合を想定し、収集したものである。

初級者が指文字を間違えて呈示する場合、その誤りのパターンには膨大な数が存在し、すべての誤りを想定して実験を行うことが困難である。そこで、41 文字以外の手形状を選別するときは、文献 [5]と文献 [12]を参考に、人から見て間違いやすい手形状、手形状が似ているために誤って覚えてしまう可能性の高いものに対象を絞った。例えば、図 4.1 の手形状 P1 は「し」と「す」の間のような手形状を呈示しており、曖昧な手形状である。また、手形状 P18 は「る」と「て」の間の手形状であるが、「る」を呈示するときに薬指も立てるといった指の本数による誤りがあると考えられる。このような考え方をもとに 36 種類の手形状を選別した。

閾値 Th_1 の決定方法について説明する。まず、データセット 1 をテストデータとして入力部分空間を構成した。入力部分空間は指文字ごとに 5 枚の画像から 1 つの部分空間が構成する。例えば、指文字「あ」の画像が 50 枚入力された場合、図 4.2 のように入力部分空間 1、入力部分空間 2 といったように入力部分空間を 46 個生成する。この方法でデータセット 1 に対し、1 文字あたり合計 230 個 (=46 個×5 セット) の入力部分空間を構成した。この方法は本システムでの入力部分空間の構成方法となっており、本論文では入力部分空間の構成方法はすべてこの方法を用いる。

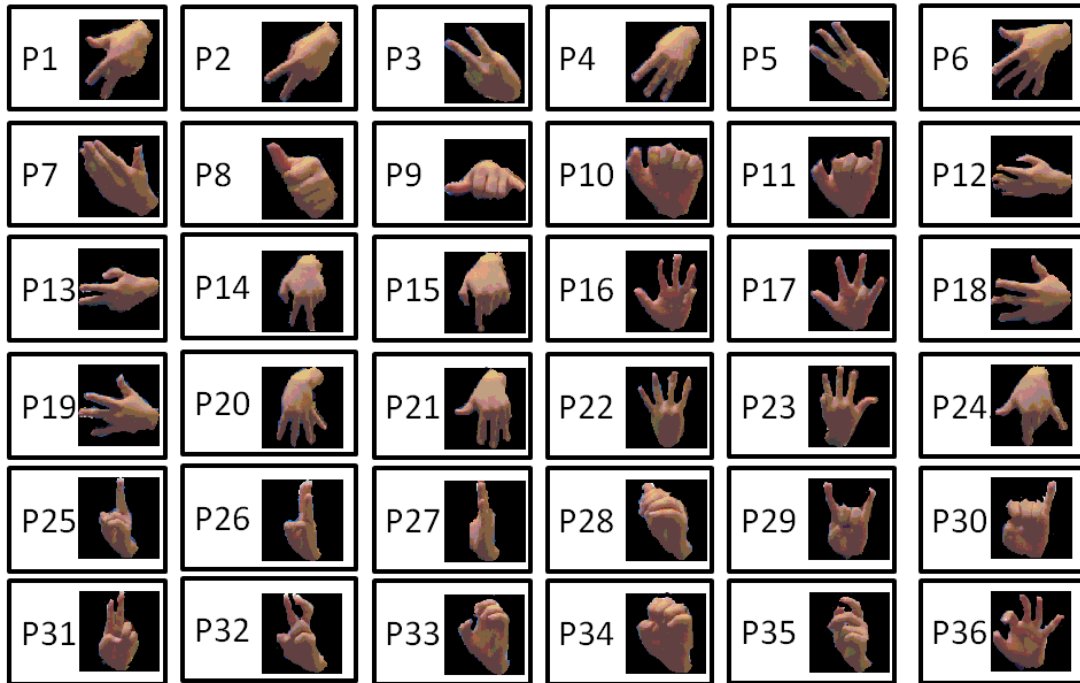


図 4.1 指文字以外の手形状のパターン 36 種類

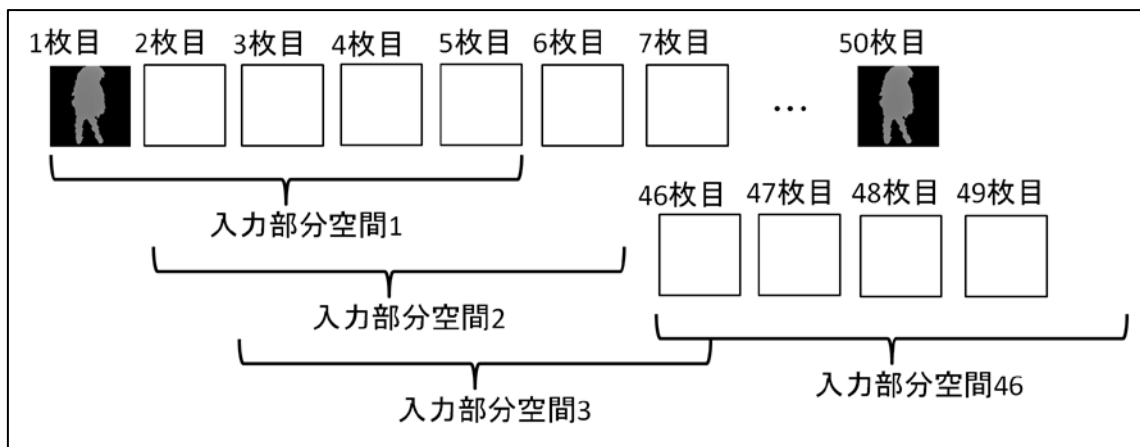


図 4.2 入力部分空間の構成方法

次に、構成した入力部分空間と 41 文字の辞書部分空間との類似度をそれぞれ計算した。類似度を計算するときは、以下の 2 種類の類似度データを求める。

- i. 入力した文字と指定した文字が同じである場合の類似度データ
- ii. 入力した文字と指定した文字が異なる場合の類似度データ

1 文字あたり、入力部分空間が 230 個 (=46 個×5 セット) あるので、辞書部分空間との類似度の計算で得られる類似度データ i は 230 個となる。類似度データ ii は入力した文字と指定した文字が異なる組み合わせは 40 通りあるので、 $40 \times 230 = 9200$ 個となる。

そして、類似度データ i と類似度データ ii を用いて、閾値 Th_1 を 0~1 の間で 0.01 ずつ変更しながら等価エラー率 (EER : Equal Error Rate) を計算した。EER とは入力した文字と指定した文字が同じであるのに不正解と認識したエラー率 (FRR : False Reject Rate) と入力した文字と指定した文字が異なるのに正解と認識したエラー率 (FAR : False Accept Rate) が等しくなる時のエラー率のことである。EER はパターン認識の分野において広く用いられている尺度であり、EER の数値が低ければ低いほど、その結果に対する信頼性があることを示す。FRR および FAR の計算方法は以下の通りである。

$$FRR[\%] = \frac{a}{\text{類似度データ i の個数}} \times 100 \quad (4.1)$$

$$FAR[\%] = \frac{b}{\text{類似度データ ii の個数}} \times 100 \quad (4.2)$$

ここで、 a は入力した文字と指定した文字が同じであるのに不正解と認識したエラー数、 b は入力した文字と指定した文字が異なるのに正解と認識したエラー数を表す。類似度データ i について閾値 Th_1 以上でないデータが存在する場合、入力した文字と指定した文字が同じであるのに不正解と認識したエラーとし、類似度データ ii について閾値 Th_1 以上のデータが存在する場合は入力した文字と指定した文字が同じであるのに正解と認識したエラーとした。図 4.3 のように閾値 Th_1 を 0~1 の間で 0.01 ずつ変更することで、FRR と FAR の差が最も少ない閾値 Th_1 を探索した。これを 41 文字繰り返すことで指文字ごとに閾値 Th_1 をそれぞれ決定した。

閾値 Th_2 の決定方法について説明する。まず、閾値 Th_1 を決定するときと同様の入力部分空間の構成方法で、データセット 1 に対して 1 文字あたり合計 230 個 (= 46 個×5 セット)、データセット 2 に対して 1 文字あたり合計 92 個 (= 46 個×2 セット) の入力部分空間を構成した。

次に、構成した入力部分空間を用いて 41 文字の辞書部分空間との類似度データ ii を計算した。閾値 Th_1 を決定するとき、類似度データ ii の個数は $40 \times 46 \times 5 = 9200$ 個であったが、閾値 Th_2 を決定するときは 36 種類の手形状のデータが追加されているので、1 文字に対し、 $40 \times 46 \times 5 + 36 \times 46 \times 2 = 12512$ 個である。

そして、1 セットごとに 41 文字の誤照合率を計算した。誤照合率とは標準形ではないものを一定時間呈示し、その間で照合成功の条件を満たしてしまった時のエラー率を表し

たものである。これは、実際に本システムを用いて標準形ではないものを呈示したときを想定したものである。誤照合率の計算方法は式 4.3 の通りである。

$$\text{誤照合率}[\%] = \frac{\text{誤照合の数}}{41 \text{ 文字} \times 5 \text{ セット} + 36 \text{ 種類} \times 2 \text{ セット}} \times 100 \quad (4.3)$$

ここで、照合成功の条件(イ)の閾値 Th_1 は前述の閾値 Th_1 の決定方法により指文字ごとに設定したものである。閾値 Th_2 を決定するときは指文字ごとに 0~1 の間で 0.01 ずつ変更しながら誤照合率が 0.5%以下となる閾値を閾値 Th_2 とした。

このように誤照合率が 0.5%以下まで抑えたときの閾値 Th_1 と閾値 Th_2 を用いることにより、ユーザが誤って手形状を呈示した場合でも、ある程度誤照合を未然に防ぐことができる。

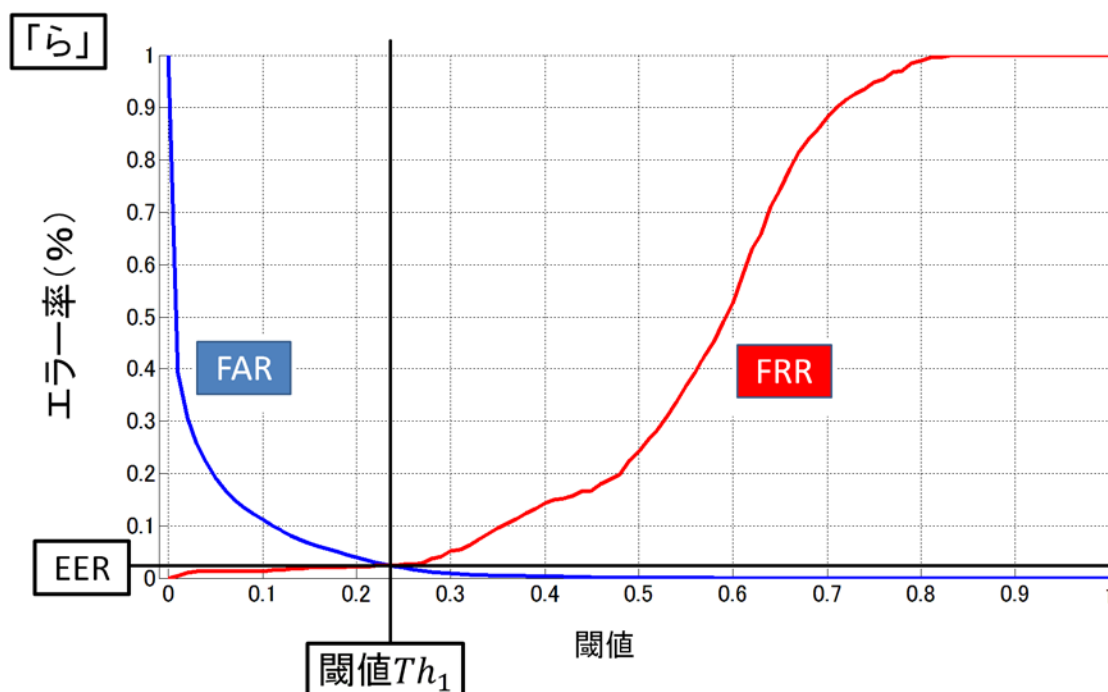


図 4.3 「ら」の閾値 Th_1 を決定するときの FAR と FRR の図

4.1.2 辞書構成に協力した手話通訳者の場合

ここでは、実際に第3章で構成した辞書を用いて、辞書作成に協力した任意の2名の手話通訳者を対象に照合実験を行うことで、ユーザ依存の辞書ならば標準形を照合できるかどうか確認を行う。テストデータは前述の閾値 Th_1 と閾値 Th_2 を決定するときに用いた手話通訳者2名の5セットと同じデータを用い、そのときの照合成功率を調べた。表4.1に手話通訳者の照合成功率の結果を示す。

表 4.1 手話通訳者の照合結果

文字	L1	L2	平均	文字	L1	L2	平均
あ	100%	100%	100%	さ	100%	100%	100%
い	100%	100%	100%	し	100%	100%	100%
う	80%	100%	90%	す	80%	80%	80%
え	100%	100%	100%	せ	100%	100%	100%
お	100%	100%	100%	そ	100%	100%	100%
か	100%	100%	100%	た	100%	80%	90%
き	100%	100%	100%	ち	100%	80%	90%
く	100%	100%	100%	つ	80%	100%	90%
け	100%	100%	100%	て	100%	80%	90%
こ	100%	100%	100%	と	100%	100%	100%
文字	L1	L2	平均	文字	L1	L2	平均
な	80%	100%	90%	ま	100%	100%	100%
に	100%	100%	100%	み	100%	100%	100%
ぬ	100%	80%	90%	む	100%	100%	100%
ね	100%	100%	100%	め	100%	100%	100%
の				も			
は	100%	100%	100%	や	100%	100%	100%
ひ	100%	100%	100%				
ふ	100%	100%	100%	ゆ	100%	100%	100%
へ	100%	100%	100%				
ほ	100%	100%	100%	よ	100%	100%	100%
文字	L1	L2	平均	L1:手話通訳者1 L2:手話通訳者2 全体の照合成功率の平均 97.3%			
ら	100%	100%	100%				
り							
る	100%	60%	80%				
れ	100%	100%	100%				
ろ	100%	100%	100%				
わ	100%	100%	100%				
を							
ん							

ここで、照合成功率について述べる。照合が成功したとは、本システムにおいて 50 フレーム指文字を呈示し、50 フレームの中でひとつでも照合成功の条件を満たした場合である。照合成功率は 41 文字の中で照合が成功した指文字の数の割合を表し、式 4.4 で計算する。

$$\text{照合成功率}[\%] = \frac{\text{照合成功数}}{41 \text{ 文字}} \times 100 \quad (4.4)$$

表 4.1 を見ると、照合成功率の平均は 97.3%であり、さらにすべての指文字の照合成功率が 80%以上であった。

4.1.3 日常的に指文字を使用している聴覚障害者の場合

ここで、日常的に指文字を使用している聴覚障害者 2 名に対して実際に指文字練習システムを 5 回使用してもらい、ユーザ独立の辞書でも標準形を呈示すれば照合できるかどうか確認を行う。被験者は閾値 Th_1 と閾値 Th_2 を決定したときの聴覚障害者とは異なる 2 名である。

日常的に指文字を使用している聴覚障害者の場合、指文字の表現にくせが見られ、標準形から外れた表現となることが多く、中級レベルのユーザであると考えられる。指文字の練習効果に関する実験を行うときは、被験者のデータを含まないユーザ独立の辞書で本システムを使用してもらう必要があり、事前に標準形を呈示すれば高精度に照合を行うことができることを確認しなければならない。そこで、実験時は付録 A の標準形一覧を渡しておき、常に確認しながら標準形を呈示してもらい、そのときの照合成功率を調べることで、ユーザ独立の辞書でも標準形を呈示すれば照合できることを確認した。

表 4.2 に聴覚障害者の照合成功率の結果を示す。これを見ると、照合成功率の平均は 92.2%であり、辞書に対象ユーザのデータを含まないユーザ独立の辞書でも標準形を呈示すれば、一部を除いて照合できることがわかった。41 文字において、80%以上照合できた指文字は 37 文字であり、それらの指文字に関しては標準形を呈示すれば、照合できることがわかった。一方、照合成功率が 80%未満の文字は「つ」、「ぬ」、「ら」、「ろ」の 4 文字であった。

表 4.2 聴覚障害者の照合結果

文字	D1	D2	平均	文字	D1	D2	平均
あ	100%	100%	100%	さ	100%	100%	100%
い	100%	80%	90%	し	100%	100%	100%
う	80%	100%	90%	す	100%	100%	100%
え	100%	100%	100%	せ	100%	100%	100%
お	80%	80%	80%	そ	80%	100%	90%
か	100%	60%	80%	た	100%	100%	100%
き	100%	100%	100%	ち	80%	80%	80%
く	100%	100%	100%	つ	40%	60%	50%
け	100%	100%	100%	て	100%	80%	90%
こ	80%	100%	90%	と	80%	100%	90%

文字	D1	D2	平均	文字	D1	D2	平均
な	100%	100%	100%	ま	100%	100%	100%
に	100%	100%	100%	み	100%	100%	100%
ぬ	40%	80%	60%	む	100%	100%	100%
ね	100%	100%	100%	め	100%	100%	100%
の				も			
は	100%	100%	100%	や	100%	100%	100%
ひ	100%	100%	100%				
ふ	100%	100%	100%	ゆ	100%	100%	100%
へ	100%	100%	100%				
ほ	100%	100%	100%	よ	100%	100%	100%

文字	D1	D2	平均
ら	40%	60%	50%
り			
る	100%	100%	100%
れ	100%	100%	100%
ろ	20%	60%	40%
わ	100%	100%	100%
を			
ん			

D1:聴覚障害者1
D2:聴覚障害者2
全体の照合成功率の平均
92.2%

4.1.4 考察

まず、手話通訳者の場合の照合成功率が 97.3%、聴覚障害者の場合の照合成功率が 92.2% 得られたことについて考察する。これは、本システムで検討した特徴抽出方法、辞書の構成方法および照合パラメータの決定方法が有効であり、ユーザ依存の辞書だけではなく、ユーザ独立の辞書でも標準形と高精度に照合を行うことができたということがいえる。

次に、聴覚障害者の照合成功率が低い指文字について考察する。照合が失敗した画像データを調べると、それらのデータには標準形を呈示しても、手の角度によって成功するパターンと失敗するパターンがあることがわかった。例えば、図 4.4 の「つ」の場合、照合が成功するときは指を相手側にまっすぐ突き出しているが、照合が失敗するときは指を少し下に突き出している傾向があった。他の指文字も同様に手の角度によって成功するパターンと失敗するパターンがあった。これは辞書データのパターン不足が考えられ、失敗したパターンを対象にデータを追加するべきであると考えられる。ただし、「ぬ」と「ろ」に関しては図 4.5 のように、辞書データのパターン不足だけではなく、手領域切り出しが失敗し、手領域の中で手首や服の一部が入ることで照合が失敗するパターンもあった。これは、「ぬ」と「ろ」の指文字は、掌を Kinect に対して横に向けており、手形状と腕が地面に対して垂直になりやすく、Kinect からの距離が手形状と腕において差が少なくなるため、他の指文字より手首や服の一部が入りやすくなることが原因と考えられる。また、「お」、「か」、「こ」など「ぬ」、「ろ」と同様に掌を Kinect に対して横に向けている文字では、ある程度手首や服の一部が入っているデータがみられた。しかし、「お」、「か」、「こ」の場合は手首や服の一部が入っても標準形を呈示すれば 80%以上照合成功することができた。このことから、手首や服の一部が入っても辞書にそのパターンのデータを含めれば、手首や服の一部が入っても標準形を呈示すれば照合を高精度に行うことが可能になることが考えられる。したがって「ぬ」と「ろ」に関しては手領域の切り出し方法を改良する、または辞書データを追加することで、より高精度に照合することができると考えられる。

以上より、手領域の切り出しに関しては、「お」、「か」、「こ」、「ぬ」、「ろ」など掌が Kinect に対して横を向いている 5 文字以外は安定して手領域の切り出しができていたことから、本システムの Kinect の配置や手領域の切り出し方法は一部を除いて成功していることがいえる。

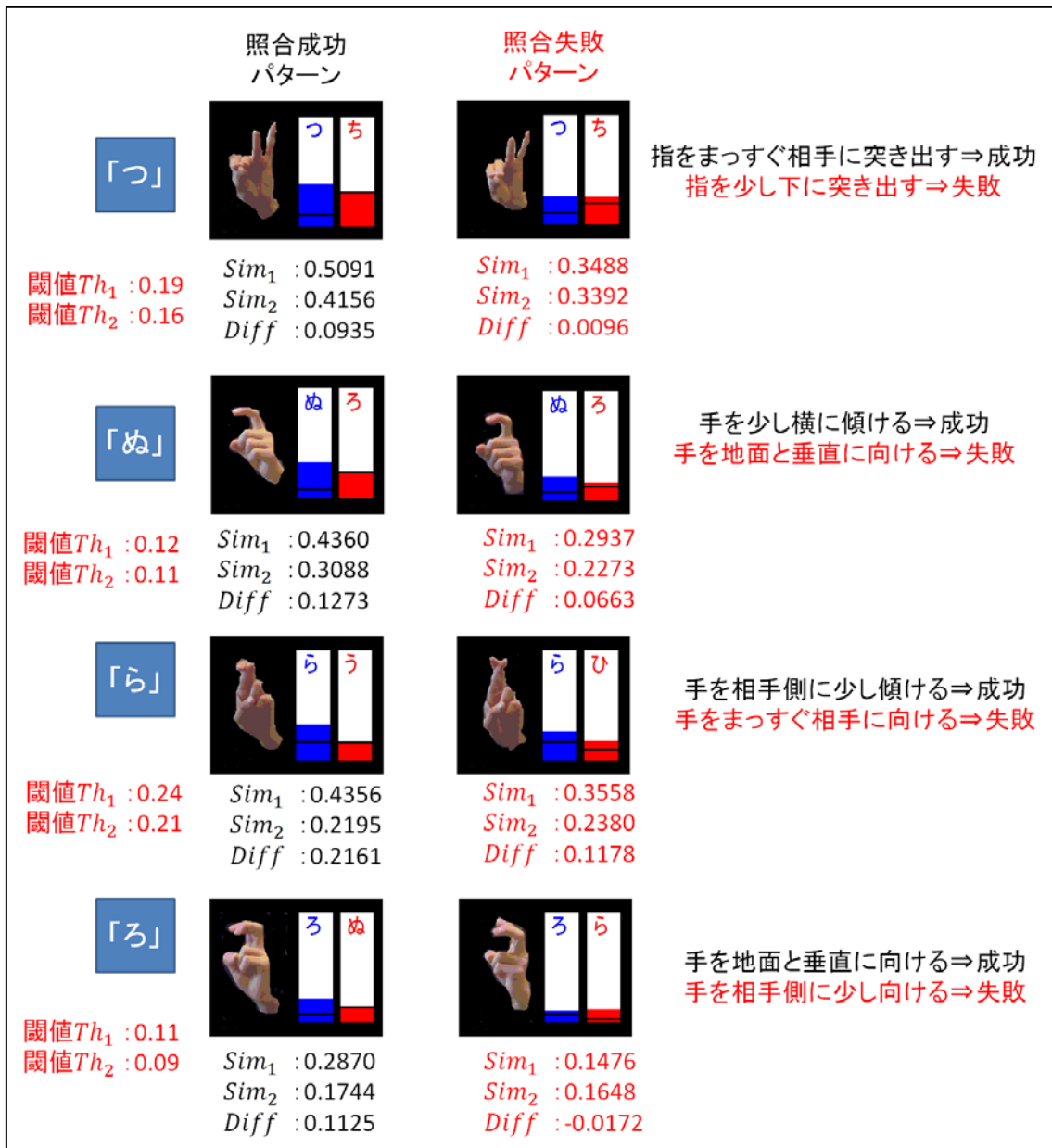


図 4.4 聴覚障害者が照合成功したパターンと失敗したパターンの例



図 4.5 手領域の切り出しが成功したパターンと失敗したパターンの例

4.2 指文字の練習効果に関する評価実験

ここでは、本システムにおける第3章のフィードバック画面とUIの有効性の確認のために、フィードバックがある場合の練習システムとフィードバックがない場合の練習システムに分けて比較実験を行う。フィードバックがある場合の練習システムとは、本システムのことであり、図3.11のフィードバック画面において①～⑥の要素で構成し、練習終了後に判定結果のリストを表示するシステムのことであり、本システムは図3.11の④の照合結果、⑤の類似度の棒グラフ、⑥の判定結果の色別表示というフィードバック表示を行うことでユーザに呈示している指文字に関する情報をリアルタイムで提供し、練習終了後に判定結果のリストを表示することでユーザにどの指文字が標準形と異なっているかどうか伝えることができるため、双方向の練習ができると考えられる。一方、フィードバックがない場合の練習システムとは図3.11のフィードバック画面において①～③の要素のみで構成し、練習終了後に判定結果のリストを表示せずに終了するシステムのことであり、このシステムを用いて指文字を練習した場合、ユーザは呈示している指文字に関して標準形であるかどうか確認することができないため、本やDVDを用いた一方向の練習を行った場合と同じ効果になることが考えられる。フィードバックがある場合の練習システムとフィードバックがない場合の練習システムをそれぞれ使用してもらい、そのときの照合成功率と照合が成功するまでに経過したフレーム数を比較することで、本やイラストを用いるよりも双方向の練習ができる本システムを用いるほうが有効であることを確認する。

実験の流れは以下のように行う。まず、日常的に指文字を使用している聴覚障害者男性10名、女性10名を対象に、それぞれ男性5名、女性5名の割合でフィードバックがある場合の練習システムを使用したグループ、フィードバックがない場合の練習システムを使用したグループをそれぞれ作る。次にそれぞれのグループで練習システムを10回ずつ使用してもらい、そのときの照合成功率と照合が終了するまでに経過した平均フレーム数を求める。最後に、フィードバックがある場合の練習システムを使用したグループとフィードバックがない場合の練習システムを使用したグループの照合成功率と平均フレーム数を比較する。ここで、照合成功率とは41文字の中でユーザが標準形を出すことができる割合を表しており、照合成功率が高ければ高いほど、ユーザは標準形をより多く呈示することができるようになったということがいえる。また、照合が終了するまでに経過した平均フレーム数とはユーザが標準形を早く呈示することができるフレーム数を表し、平均フレーム数が低ければ低いほど、ユーザは標準形を早く呈示することができるようになったということがいえる。

4.2.1 実験 I: フィードバックがある場合

実験 I の対象となる被験者は聴覚障害者 10 名であり、男性 5 名と女性 5 名の構成である。また、10 名の内 8 名はそれぞれ日常的に指文字を使用している中級レベルのユーザ、2 名は指文字を覚え始めて 1 年未満の初級レベルのユーザである。練習システムは合計で 10 回使用してもらい、最初の 3 回は標準形の資料を見せず、練習システムの使い方に慣れてもらった。4 回目から 10 回目までは練習システムを使用する前に 5 分間標準形の資料を確認してもらい、確認した後に練習システムを使用することを繰り返した。練習が 10 回終わった後は被験者に図 4.6 の指文字練習システムに関するアンケートについて回答してもらい、システムに対する評価や感想を書いてもらった。最後に、照合成功率と平均フレーム数において 1 回目の分散との検定および平均値との検定を行うことにより、繰り返し使用することによる練習効果がみられるかどうか確認した。分散の差の検定を行うときは F 検定で分散に有意差があるか確認し、平均値の差の検定を行うときは t 検定で有意差があるか確認した。なお、F 検定において分散に 5% 水準で有意差がみられたときは不等分散による t 検定を用い、分散に有意差がないときは等分散による t 検定を用いた。

表 4.3 に実験 I の照合成功率と平均フレーム数の結果を示す。表 4.3 の照合成功率と平均フレーム数は実験 I の被験者 10 人の平均値を表す。照合成功率における 1 回目と 10 回目の平均値を比較すると、10 回目で 28% 上昇しており、平均フレーム数における 1 回目と 10 回目の平均値を比較すると、10 回目で 14.1 フレーム減少していた。さらに、表 4.3 の平均値の検定結果から、照合成功率と平均フレーム数ともに 1% 水準で 4 回目以降に有意差がみられた。

ここで、筆者が被験者が練習システムを使用しているときの様子を観察すると、指文字の表現が誤っていた場合は、図 4.7 のように被験者が類似度の変動を棒グラフで確認することにより、自身が呈示している指文字が標準形とは異なることに気づき、標準形へ修正を行う様子がみられた。

指文字練習システムに関するアンケート

	非常に そう思わない	1・2・3・4・5・6・7	どちらとも 言えない	非常に そう思う
(1) 練習システムを使ってみて、面白いと思いますか?	1	2	3	4
(2) 練習システムをもっと使ってみたいですか?	1	2	3	4
(3) 練習システムは使いやすいですか?	1	2	3	4
(4) 表示画面のレイアウトは見やすいですか?	1	2	3	4
(5) 練習システムを使う前より指文字を 素早く出すことができるようになりましたか?	1	2	3	4
(6) 練習システムを使うと、指文字を早く覚えることができそうですか?	1	2	3	4
(7) 自分のどの指文字が標準形と異なるか把握できましたか?	1	2	3	4
(8) 表示画面のどの情報が役に立ちましたか?	()			
(9) 標準形と異なっていた指文字を具体的に挙げて下さい。	()			
(10) 練習システムを使ってみて、ここを直したほうが いいなと思う点があれば記入してください。 ※自由回答	()			
(11) 他に何か感想があれば、記入してください。 ※自由回答	()			
(12) 最後にあなた自身のことについてお伺いします。 性別 (男 ・ 女) 出身地 () 指文字の経験年数 () 出身校 (ろう学校・一般校・両方とも通学経験あり) 指文字を使用している頻度				
1. 指文字をあまり覚えていないので、使わない	※以下の項目に関しては、あてはまるものに全て○をお願いします。			
2. できるだけ手話で表し、指文字を使わないようにしている	友達と話すとき (口話・声のみ・手話・指文字・筆談)			
3. 手話でわからない単語があれば、時々使う程度	家族と話すとき (口話・声のみ・手話・指文字・筆談)			
4. 指文字に頼って話すことが多い	その他 (学校の先生など) (口話・声のみ・手話・指文字・筆談)			

アンケートは以上です。ご記入ありがとうございました。

図 4.6 指文字練習システムに関するアンケート

表 4.3 実験 I の照合成功率と平均フレーム数の結果

練習回数	照合成功率	標準偏差	p値 (F検定)	p値 (t検定)	1回目の平均値との 有意差
1	57%	0.0984			
2	59%	0.1513	0.21600	0.70494	
3	54%	0.1090	0.76574	0.53628	
4	76%	0.0754	0.44122	0.00011	1%水準で有意差あり
5	79%	0.1068	0.81093	0.00013	1%水準で有意差あり
6	83%	0.0512	0.06470	0.00000	1%水準で有意差あり
7	83%	0.0770	0.47591	0.00000	1%水準で有意差あり
8	85%	0.0700	0.32493	0.00000	1%水準で有意差あり
9	88%	0.0784	0.50872	0.00000	1%水準で有意差あり
10	85%	0.0879	0.74293	0.00000	1%水準で有意差あり

練習回数	平均フレーム数	標準偏差	p値 (F検定)	p値 (t検定)	1回目の平均値との 有意差
1	31.2	5.1			
2	29.5	5.8	0.42633	0.30977	
3	31.6	4.4	0.94667	0.87724	
4	23.7	3.8	0.70705	0.00017	1%水準で有意差あり
5	21.2	4.0	0.94562	0.00006	1%水準で有意差あり
6	17.7	5.1	0.94655	0.00001	1%水準で有意差あり
7	18.0	6.2	0.66163	0.00000	1%水準で有意差あり
8	16.5	5.8	0.51953	0.00000	1%水準で有意差あり
9	15.3	4.1	0.96046	0.00000	1%水準で有意差あり
10	17.1	4.5	0.83680	0.00001	1%水準で有意差あり

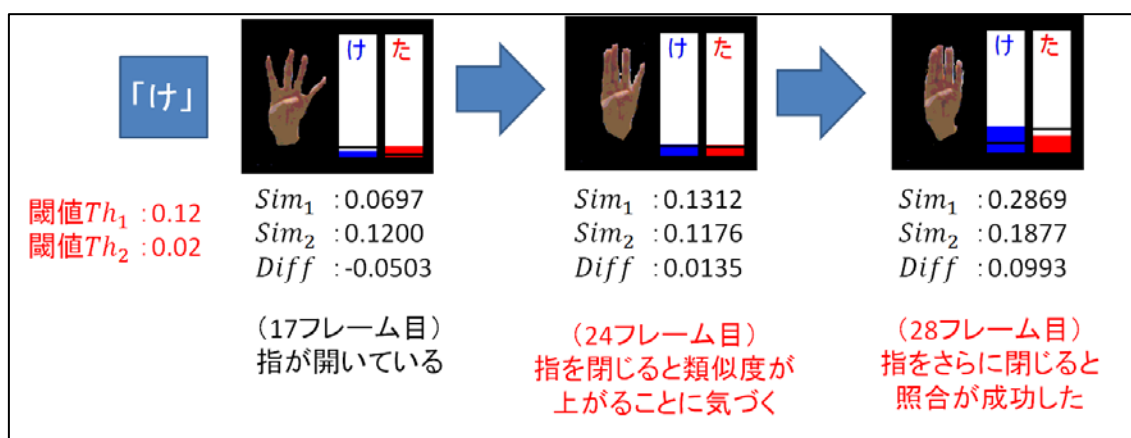


図 4.7 被験者が指文字表現の誤りに気づき、修正する様子

4.2.2 実験 II: フィードバックがない場合

実験 II の対象となる被験者は聴覚障害者 10 名であり、男性 5 名と女性 5 名の構成である。また、10 名はすべて日常的に指文字を使用している中級レベルのユーザである。実験 II の実験条件は実験 I とほぼ同じであるが、実験 II の場合は 4 回目以降も 5 分間標準形の資料を確認した後にフィードバックがない場合の練習システムを使用してもらった。

表 4.4 に実験 II の照合成功率と平均フレーム数の結果を示す。表 4.4 の照合成功率と平均フレーム数は実験 II の被験者 10 人の平均値を表す。照合成功率における 1 回目と 10 回目の平均値を比較すると 10 回目で 12% しか上昇せず、平均フレーム数における 1 回目と 10 回目の平均値を比較すると 10 回目で 2.5 フレームしか減少していないことがわかった。さらに、表 4.4 の平均値の検定結果から、照合成功率と平均フレーム数はともに検定結果をみると、照合成功率と平均フレーム数両方とも有意差がないことがわかった。

表 4.4 実験 II の照合成功率と平均フレーム数の結果

練習回数	照合成功率	標準偏差	p値 (F検定)	p値 (t検定)	1回目の平均値との 有意差
1	55%	0.2058			
2	55%	0.1586	0.44961	0.97664	
3	57%	0.1718	0.59955	0.77683	
4	67%	0.1543	0.40391	0.16715	
5	66%	0.1736	0.62092	0.21364	
6	65%	0.1710	0.59053	0.26395	
7	64%	0.2117	0.93367	0.33398	
8	63%	0.1934	0.85708	0.39365	
9	60%	0.2077	0.97857	0.58635	
10	67%	0.1464	0.32538	0.15185	

練習回数	平均フレーム数	標準偏差	p値 (F検定)	p値 (t検定)	1回目の平均値との 有意差
1	29.5	10.0			
2	31.4	7.8	0.44501	0.84996	
3	28.8	6.8	0.53385	0.59718	
4	26.5	7.6	0.58278	0.08894	
5	27.4	9.1	0.89208	0.14860	
6	26.9	8.1	0.68388	0.19878	
7	27.1	9.4	0.94853	0.27870	
8	26.3	8.4	0.86123	0.26978	
9	27.5	8.3	0.85928	0.52623	
10	27.0	6.4	0.32468	0.12784	

4.2.3 実験 I と実験 II の比較

実験 I と実験 II の結果から、それぞれのグループで照合成功率と平均フレーム数の平均値を比較することで、本システムは、双方向の情報のやり取りを可能としたフィードバックのある練習システムであり、本や DVD を用いた場合の一方方向の練習より有効であることを示す。それぞれの練習回において、まず F 検定で分散に有意差があるか確認したあとに、t 検定を用いて平均値に有意差があるか確認した。

表 4.5 に実験 I と実験 II における練習回数ごとの差の検定結果、図 4.8 に実験 I と実験 II の照合成功率の推移、図 4.9 に実験 I と実験 II の平均フレーム数を示す。表 4.5 より、照合成功率、平均フレーム数ともに 5%水準で 6 回目以降から有意差がみられた、また図 4.8 と図 4.9 の結果から、実験 I のグループのほうが実験 II のグループより標準形を多く、早く呈示することができるようになったということがわかる。

表 4.5 実験 I と実験 II における練習回数ごとの差の検定結果

照合成功率					
練習回数	実験Iの 平均値	実験IIの 平均値	p値 (F検定)	p値 (t検定)	実験Iと実験IIの 平均値の有意差
1	57%	55%	0.03859	0.76434	
2	59%	55%	0.89036	0.55706	
3	54%	57%	0.19108	0.62805	
4	76%	67%	0.04444	0.09556	
5	79%	66%	0.16377	0.05196	
6	83%	65%	0.00136	0.00709	1%水準で有意差あり
7	83%	64%	0.00592	0.01999	5%水準で有意差あり
8	85%	63%	0.00572	0.00597	1%水準で有意差あり
9	88%	60%	0.00776	0.00192	1%水準で有意差あり
10	85%	67%	0.14444	0.00328	1%水準で有意差あり

平均フレーム数					
練習回数	実験Iの 平均値	実験IIの 平均値	p値 (F検定)	p値 (t検定)	実験Iと実験IIの 平均値の有意差
1	31.2	29.5	0.05810	0.78701	
2	29.5	31.4	0.69565	0.74860	
3	31.6	28.8	0.16833	0.36220	
4	23.7	26.5	0.08265	0.54846	
5	21.2	27.4	0.06671	0.23429	
6	17.7	26.9	0.11376	0.03324	5%水準で有意差あり
7	18.0	27.1	0.02589	0.02485	5%水準で有意差あり
8	16.5	26.3	0.02072	0.01137	5%水準で有意差あり
9	15.3	27.5	0.07532	0.00152	1%水準で有意差あり
10	17.1	27.0	0.44972	0.01590	5%水準で有意差あり

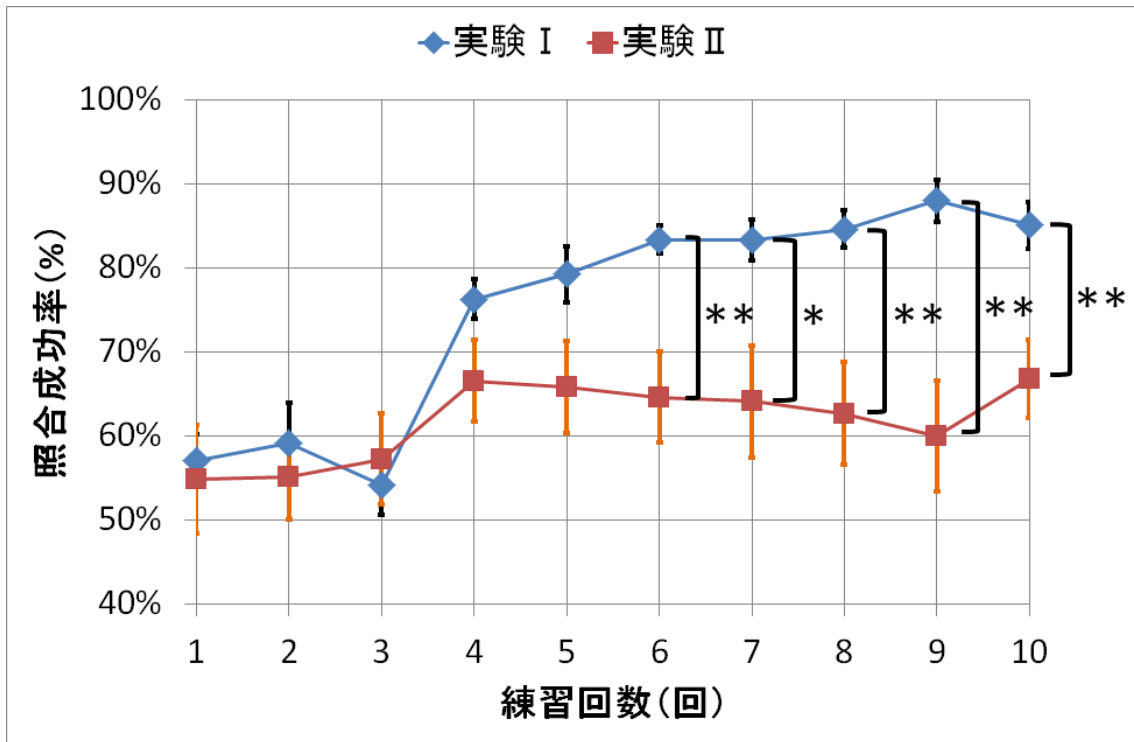


図 4.8 実験 I と実験 II の照合成功率の推移

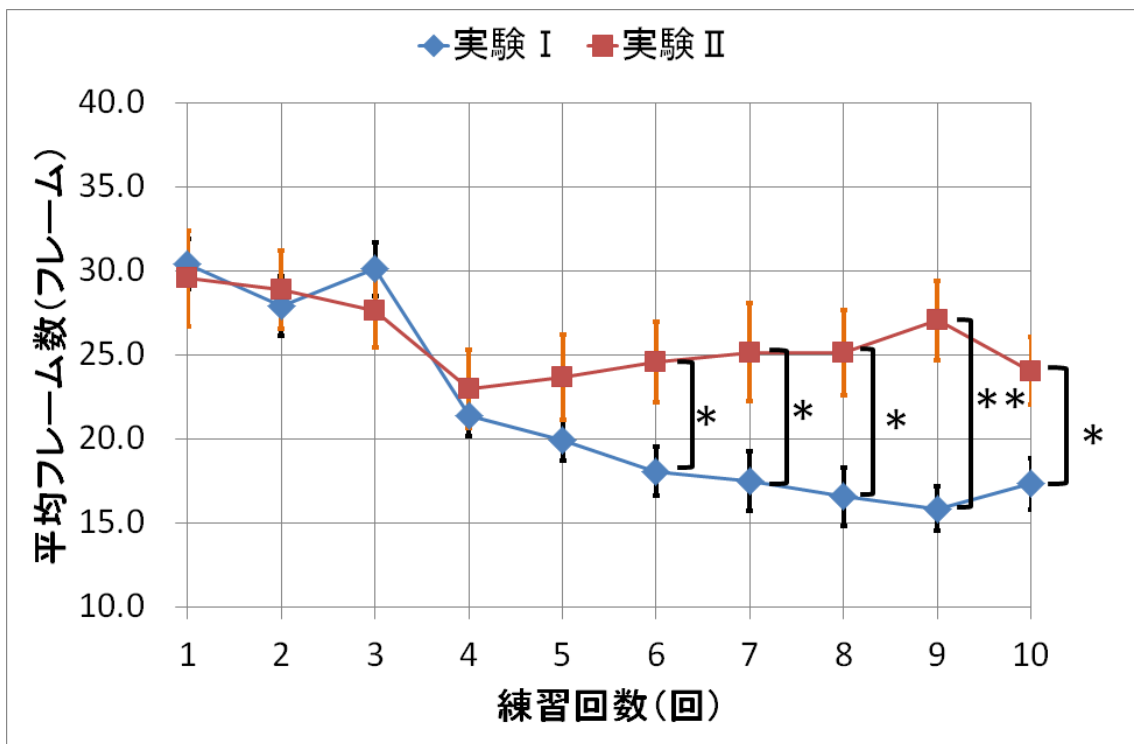


図 4.9 実験 I と実験 II の平均フレーム数の推移

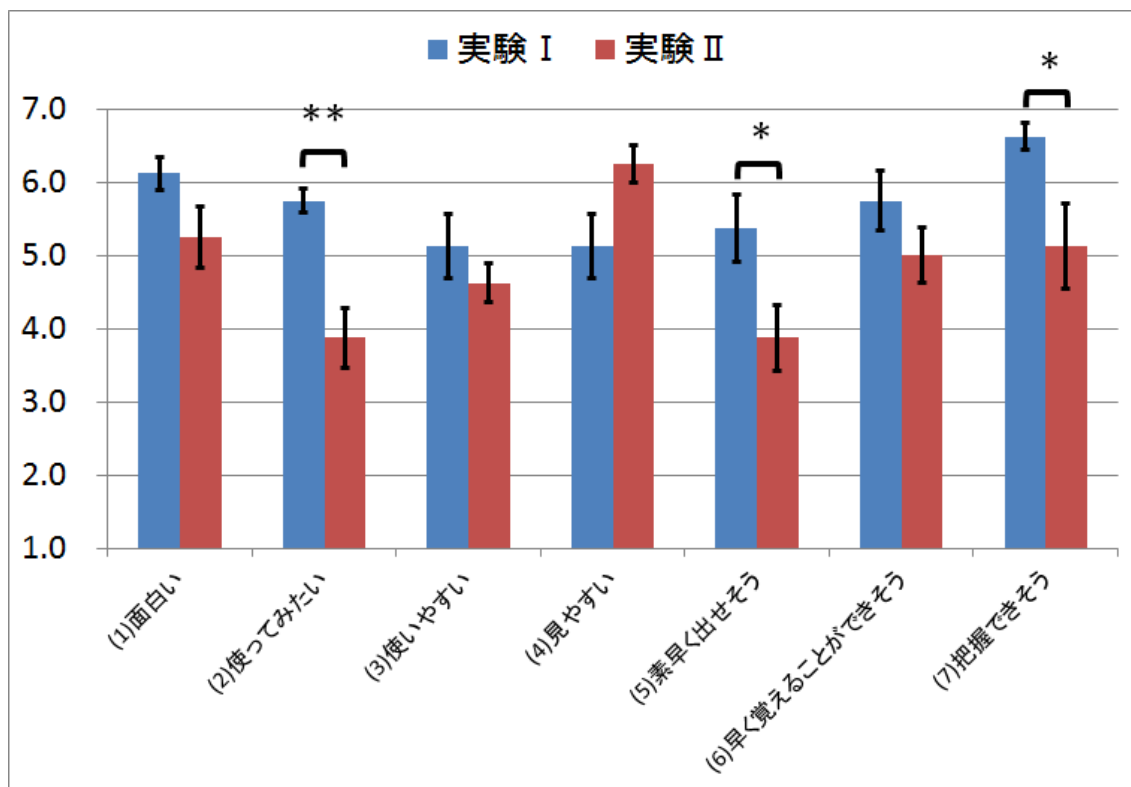


図 4.10 アンケートの集計結果

また、練習システムを使用した後のアンケートから、実験 I と実験 II において、項目(1)～(7)の回答の平均値に有意差があるかどうか評価を行った。図 4.10 に実験 I と実験 II のアンケートの結果を示す。検定の結果、(2)の項目において、1%水準で有意差がみられ、(5)と(7)の項目において 5%水準で有意差があった。

4.2.4 考察

実験 I において、照合成功率と平均フレーム数ともに 4 回目以降の平均値に 1%水準で有意差がみられたことについて考察する。標準形の資料を見せ始めたのは 4 回目ということ を考慮すると、4 回目以降で有意差がみられたということは、フィードバックがあれば練習システムを 1 回使用するだけでも十分に練習効果があるということがいえる。また、図 4.7 から被験者が呈示した指文字の表現の誤りに気づき、修正した様子がみられたことおよび実験 I のグループにおけるアンケートの回答結果から類似度の棒グラフや判定結果の色別表示が役に立ったという回答が多数得られたことから、本システムのフィードバック表示方法によりユーザに手形状の表現の誤りに気づかせることができ、双方向の練習を行うことができたということがいえる。

実験 II において、照合成功率と平均フレーム数ともに有意差がみられなかったことにつ

いて考察する。これは、フィードバックがない場合は 10 回繰り返し練習しても効果が少ないということがいえ、本や DVD を用いて一方向の練習をした場合と同じ効果となったということが考えられる。実験 II のグループにおけるアンケートの回答結果をみると、標準形を呈示しているかどうか不安である、指文字表現で修正すべき箇所がわからないという回答が一部得られたことからフィードバックがない場合の練習システムを使用した場合、呈示した指文字の表現が標準形であるかどうかわからないために標準形を呈示することが難しくなったということが考えられる。以上より、フィードバックがない場合の練習システムを使用した場合は本や DVD を用いた一方向の練習をした場合と同じ効果であり、標準形を身につけることが難しいということがいえる。

実験 I と実験 II において、照合成功率と平均フレーム数ともに 6 回目以降から 5%水準で有意差があったことについて考察する。これは、少なくとも 6 回以上フィードバックがある場合の練習システムを使用すれば、フィードバックがない場合の練習システムより練習効果があるということがいえる。つまり、本システムはユーザの指文字表現の誤りを伝えることができる、双方向の情報のやり取りができるシステムであり、本やイラストを用いた一方向の練習より有効であるということがいえる。

第5章 今後の課題

第5章では、本システムにおける今後の課題を述べる。まず、第4章の日常的に指文字を使用している聴覚障害者の場合、手領域の切り出しの問題および辞書のデータ不足により、照合成功率が92.2%しか得られなかった。手領域の切り出しについて、距離による切り出しだけではなく、手の体積による切り出し方法 [13] もあり、今後の検討が必要である。辞書のデータ不足について、今回収集した手話通訳者は12人であったが、指文字認識において40人集めれば高い認識率が得られることが報告されている [8] ことから辞書データをさらに集めることが必要であり、標準形の照合成功率を増やすことは今後の課題である。

また、本システムは清音指文字41文字だけを対象にしていたが、今後は動きのある指文字を含めた指文字練習システムが必要であり、動きのある場合の認識方法も検討する必要がある。

さらに、第4章の本システムのフィードバック画面において、現在は Sim_1 と Sim_2 を棒グラフで同時に表示しているが、フレームごとに類似度を計算しているため、常に2つの棒グラフが動き、どちらを見るべきかわかりにくいものとなっている。これを解決するためには、類似度の棒グラフを1つに統一して表示することが必要であると考えられる。 Sim_1 と Sim_2 の類似度を用いて1つの尺度で表すことができるように計算し、1つの棒グラフで表示することで、ユーザにわかりやすい情報を提供できるようにすることができる。このようにフィードバック表示方法を改良することも重要である。また、類似度の棒グラフを用いるだけではなく、形状相関マップによるガイド方法を用いて、より直感的なフィードバック表示ができるように本システムに適用する [14] ことも考えられる。

最後に、第2章の2.3.4のフィードバック表示部において、呈示している指文字の表現が標準形ではない場合、標準形を呈示するためにどのように修正すべきであるか具体的なガイド方法を開発することも今後の課題である。現在は、アニメーションやCGを用いてユーザに手形状の修正箇所を表示する方法 [2] や、事前に誤り形状を定義した辞書を作り、誤り形状に対する類似度が高ければその誤りをメッセージなどで指摘するガイド方法 [10] が提案されている。これらの手法が本システムに適用することができるかどうか検討することも今後の課題である。

第6章 まとめ

本論文では以下の3点を目的として指文字練習システムの開発と評価を行うことで、試作したシステムはユーザに指文字の表現の誤りを気づかせることができるような双方向の練習ができるシステムであり、本やイラストを用いた一方向の練習より有効であることが示した。

- (1) 安価でユーザへの負担が少ない画像入力デバイスとしてゲームなどのエンターテインメント用として一般的に普及している「Kinect」を用いて、安定して指文字を表す手領域を切り出し、手領域を効率的に抽出する。
- (2) 手領域特徴の照合方法とユーザ独立の辞書における構成方法の検討を行い、ユーザが呈示する手形状と標準形との3次元形状の照合を高精度に行う。
- (3) 照合結果をリアルタイムにフィードバック表示するための方法とユーザインタフェースを検討し、双方向の情報のやり取りを可能とする。

第2章では指文字練習システムを開発することの難しさとその対処方法を述べ、本研究で目指しているKinectを用いた指文字練習システムの概要について述べた。

第3章ではKinectを用いた場合の指文字練習システムにおいて、上記の3点を達成するために以下の6つの検討および開発方法についてその詳細を述べた。

1. Kinectの配置
2. 手領域の切り出し
3. 特徴抽出
4. 辞書の構成
5. 照合パラメータの決定
6. フィードバック画面の構成とユーザインタフェースの開発

第4章では上記の3点の目的が達成できたかどうか確認するために評価実験を行った。まず、手話通訳者と聴覚障害者が標準形を呈示して、照合成功するかどうか確認することで、上記の(1)、(2)の目的が達成したかどうか確認した。評価実験の結果、辞書構成に協力した手話通訳者が標準形を呈示した場合は97.3%、辞書構成に協力していない聴覚障害者でも標準形を呈示すれば92.2%照合成功できたことから、本システムで検討および開発した手領域切り出しの方法や辞書の構成方法などは有効であり、(1)、(2)は達成できていることを確認した。

また、実験Iと実験IIにおいて、それぞれフィードバックがある場合の練習システムとフィードバックがない場合の練習システムを使用してもらい、照合成功率や照合が終了するまでに経過した平均フレーム数を求めた。フィードバックがある場合の練習システムを使用した場合、実験Iの結果から本システムは双方向の練習ができ、練習効果があるということがわかった。一方、フィードバックがない場合の練習システムを使用した場合、実験

II の結果から、本や DVD を用いた場合の一方向の練習と同じ効果がみられ、繰り返し使用しても練習効果がみられなかった。

さらに、照合成功率や平均フレーム数についてフィードバックがある場合の練習システムを使用したグループとフィードバックがない場合の練習システムを使用したグループの平均値の差の検定を行うことにより、練習効果に差があるかどうか確認した。検定の結果、本システムを少なくとも 6 回以上使用すればフィードバックがない場合の練習システムより練習効果に差があることを示した。以上より、本システムはユーザに指文字の表現の誤りを気づかせることができるような双方向の練習ができるシステムであり、本やイラストを用いた一方向の練習より有効であることが示された。

第 5 章では、本システムにおいて、手領域の切り出し、辞書のデータ不足および動きのある指文字認識への拡張など指文字の認識に関する課題を述べた。また直感的なフィードバック方法の検討、誤りの修正への具体的なガイド方法の開発など指文字練習システムの練習方法に関する課題を述べた。

謝辞

本論文は、筆者が筑波技術大学大学院修士課程技術科学研究科産業技術学専攻における研究の成果をまとめたものである。筑波技術大学産業技術学部産業情報学科 岡崎彰夫教授と加藤伸子教授には、指導教員として日夜を問わない叱咤激励と多くの御指導、御助言を頂いた。ここに記して深謝の意を表す。また、河野純大准教授には、本学位論文における主査として、貴重な時間を割いて研究並びに論文全体についての御助言と御指導を頂いた。ここに記して深謝の意を表す。

また、筑波大学システム情報系 福井和広教授と筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻博士課程修了生 大川泰弘氏、博士前期課程 高林大輔氏には共同研究者として、本研究に対して多くの御指導と御助言を頂いた。ここに記して深謝の意を表す。

筑波技術大学大学院技術科学研究科産業技術学専攻修士課程修了生 瀬戸山浩平氏、修士課程 近藤真暉氏には、学内における共同研究者として筆者の研究に快く協力して頂き、的確な指摘と様々な御助言、御提案を頂いた。ここに記して深謝の意を表す。

本研究における標準形の呈示方法については、筑波技術大学障害者高等教育研究支援センター 白澤麻弓准教授に多くの貴重な御意見と有意義な御助言を頂いた。ここに記して深謝の意を表す。

最後に、本研究に関する実験には、手話通訳者や筑波技術大学産業技術学部の在学生など、多くの方々に被験者として御協力を頂いた。ここに記して深謝の意を表す。

参考文献

- [1] 渡辺有香,松本秀彦,諸富隆,“指文字学習における認知的変化についての3刺激 oddball 課題を用いた検討”,作新学院大学人間文化学部紀要, 6,pp.21-32,2008.
- [2] 鈴木信勝,大橋巧治,堀口進,“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レポート, IS-RR-99-0017,pp.1-21,1999.
- [3] 田畑慶人,黒田知宏,眞鍋佳嗣,千原國宏,“手型認識を用いた指文字練習教育システム”,教育システム情報学会誌, Vol.18,No.2,pp.172-178,2001.
- [4] 瀬戸山浩平,高林大輔,田中陽土,大川泰弘,加藤伸子,福井和広,岡崎彰夫,“三次元手形状の類似性を考慮した指文字練習システムのための正準角に基づく形状相関マップの検討”,ヒューマンインタフェースシンポジウム 2012 論文集, pp.1053-1058,2012.
- [5] 神田和幸, 指文字の研究, 光生館, 1986.
- [6] K.Fukui, O.Yamaguchi, “The Kernel Orthogonal Mutual Subspace Method and its Application to 3D Object Recognition”,ACCV07 Lecture Notes in Computer Science(LNCS), Vol.4844,pp.467-476,2007.
- [7] 前田賢一,渡辺貞一,“局所的構造を導入したパターン・マッチング法”,電子情報通信学会論文誌,Vol.J68-D,No.3,pp.345-352, 1985.
- [8] Y.Ohkawa,K.Fukui, “Hand-Shape Recognition Using The Distributions of Multi-Viewpoint Image Sets”,IEICE Transaction on Information and Systems, Vol.E95-D,No.6,pp.1619-1627,2012.
- [9] 坂野鋭,“パターン認識における主成分分析・顔画像認識を例として”,統計数理,第 49 卷,第 1 号,pp.23-42, 2001.
- [10] 高林大輔,大川泰弘,瀬戸山浩平,田中陽土,加藤伸子,岡崎彰夫,福井和広,“フィードバック機能を備えた指文字練習システムの試作”,電子情報通信学会技術研究報告,Vol.112,No.483,pp.79-84, 2013.
- [11] N.Otsu, “A new scheme for practical, flexible and intelligent vision systems”, Proceedings of IAPR Workshop on Computer Vision, pp.431-435, 1988.
- [12] 平美穂子,“指文字の Intelligibility に影響を及ぼす諸要因の研究”,特殊教育学研究,24(1),pp.29-36, 1986.
- [13] 三宅太一,若月大輔,内藤一郎,“距離画像を用いた動きのある指文字の非接触認識手法の検討”,電子情報通信学会ヒューマンコミュニケーショングループ(HCG)シンポジウ

ム 2012 論文集,pp.270-275, 2012.

- [14] 瀬戸山浩平,高林大輔,田中陽土,大川泰弘,加藤伸子,福井和広,岡崎彰夫, “Kinect を用いた指文字練習システムにおけるリアルタイムフィードバック表示の検討”, ヒューマンインタフェース学会研究会報告集,Vol.15,No.2,pp.35-40, 2013.

研究業績

- [1] 田中陽土,高林大輔,瀬戸山浩平,大川泰弘,加藤伸子,福井和広,岡崎彰夫 “Kinect を用いた指文字練習システムにおける識別処理の検討”,ヒューマンインタフェース学会研究会報告集,Vol.15,No.2,pp.29-34,2013.
- [2] 田中陽土,高林大輔,近藤真暉,加藤伸子,福井和広,岡崎彰夫 “Kinect を用いた指文字練習システムにおける識別処理とユーザインタフェースの改良”,電子情報通信学会ヒューマンコミュニケーショングループ (HCG) シンポジウム 2013 論文集, pp.291-296,2013.

付録 A 指文字の標準形の図一覧

本システムのデータ収集, 確認実験において指文字の標準形を定義した. この付録では, その標準形の図一覧を示す. 標準形の定義には文献 [5]を参考にした.

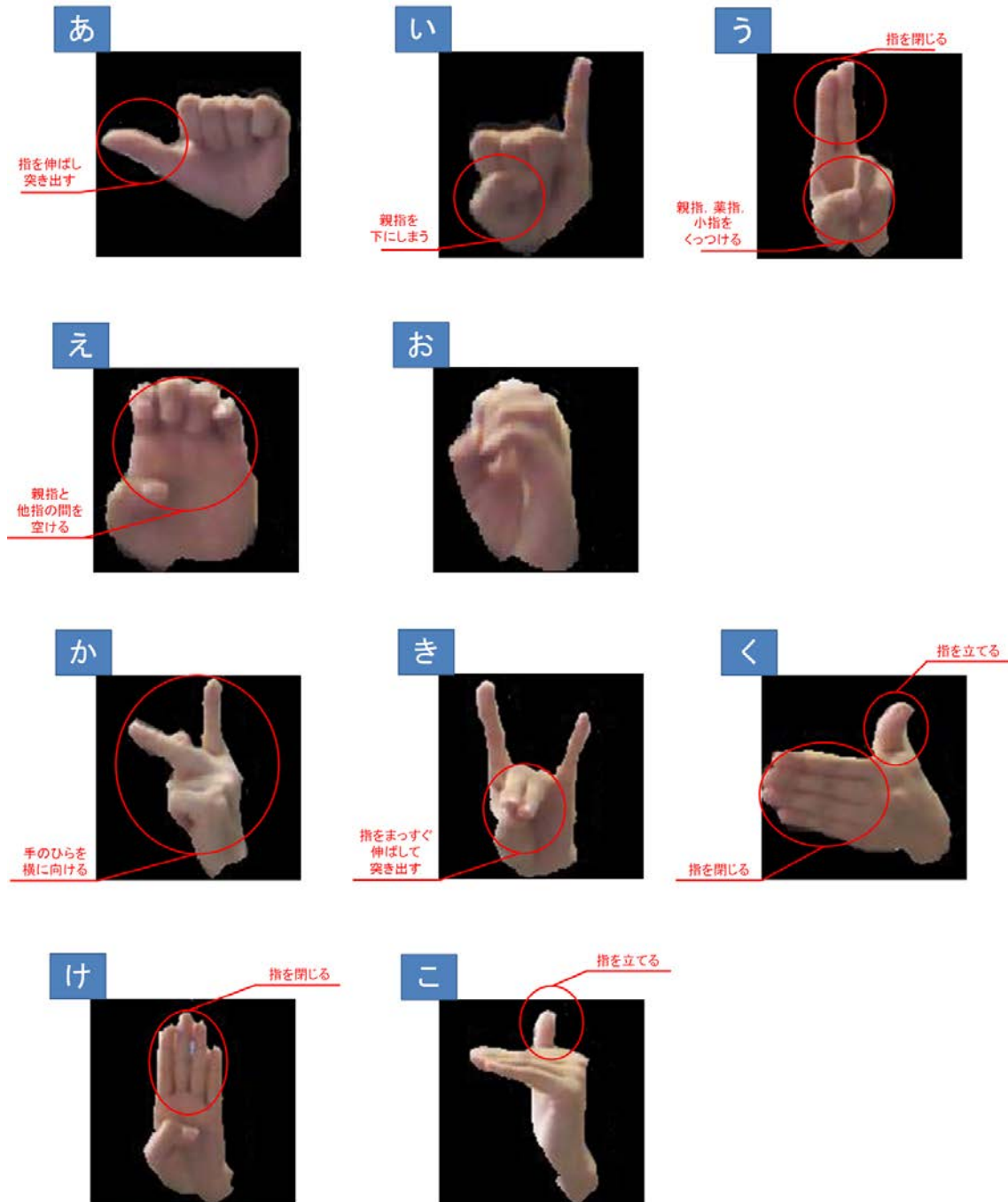


図 A.1 「あ」～「こ」の指文字の標準形一覧

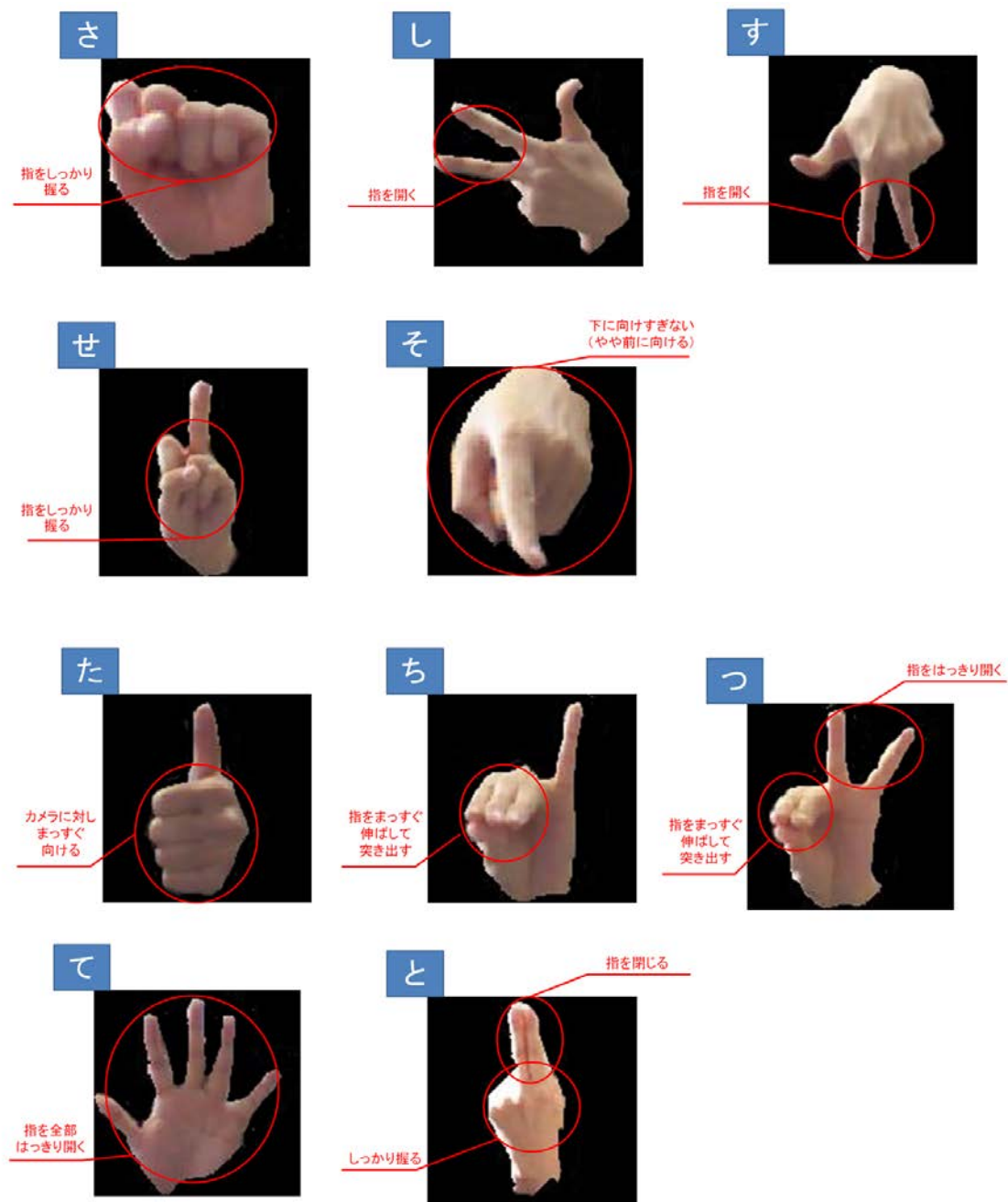


図 A.2 「さ」～「と」の指文字の標準形一覧

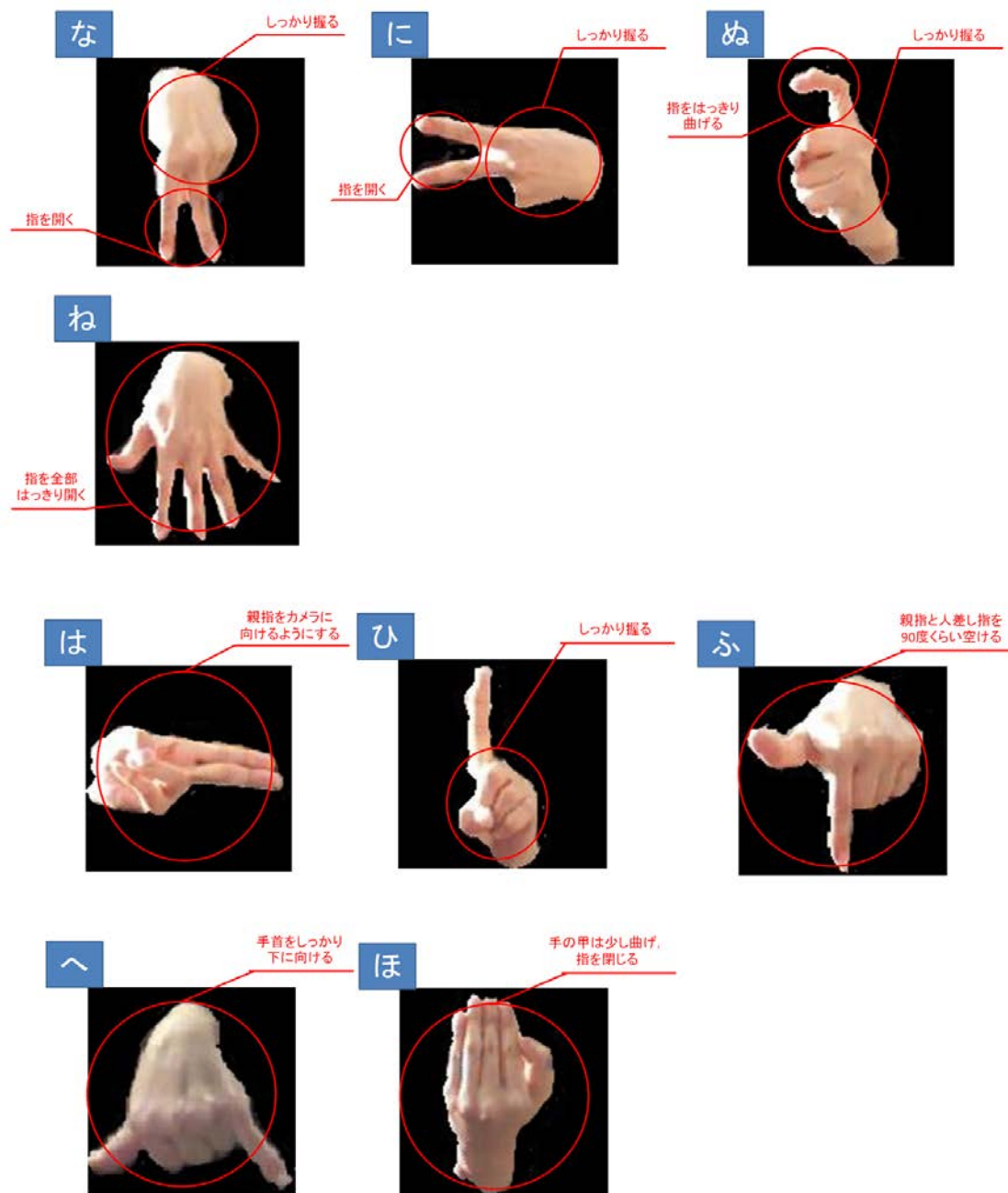


図 A.3 「な」～「ほ」の指文字の標準形一覧

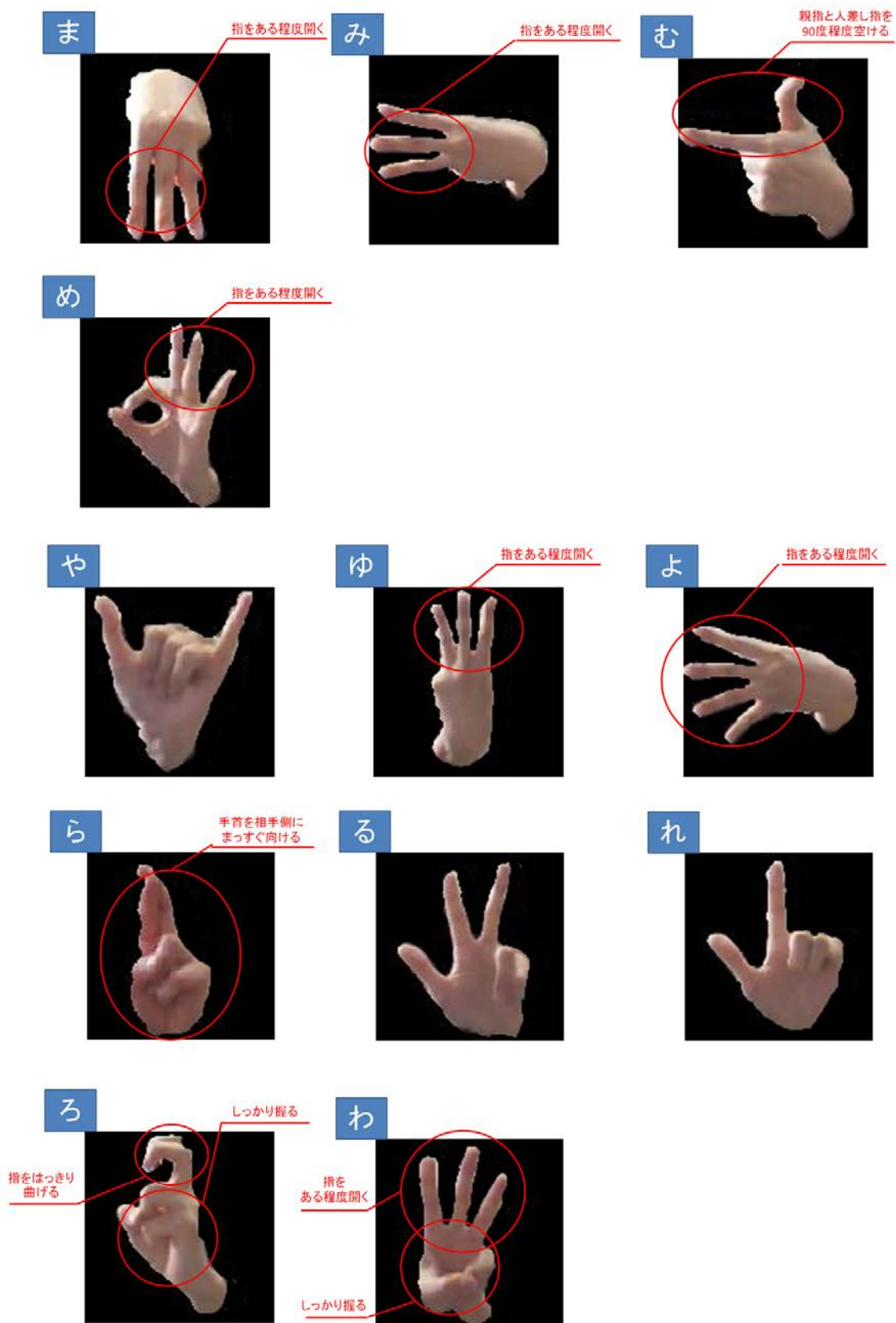


図 A.4 「ま」～「わ」の指文字の標準形一覧

付録 B 距離値と実際の距離の関係

この付録では, Kinect から得られる距離値をもとに実際の距離との関係を調べた. 図 B.1 のように適当な大きさの物体を用意し, Kinect と物体が同じ高さになるように配置して, 10cm 間隔で Kinect から物体までの距離について, カメラに映る物体の領域に対する距離値の平均を計測すると, 図 B.2 のグラフから 40cm の距離では約 450, 10cm 離れた地点では約 1050 の距離値が得られ, Kinect から近いほど距離値は小さくなり, 10cm 離れるほど距離値は 100 程度大きくなるのがわかる.

手領域の検出や距離メータの作成にはこの図を参考にした.

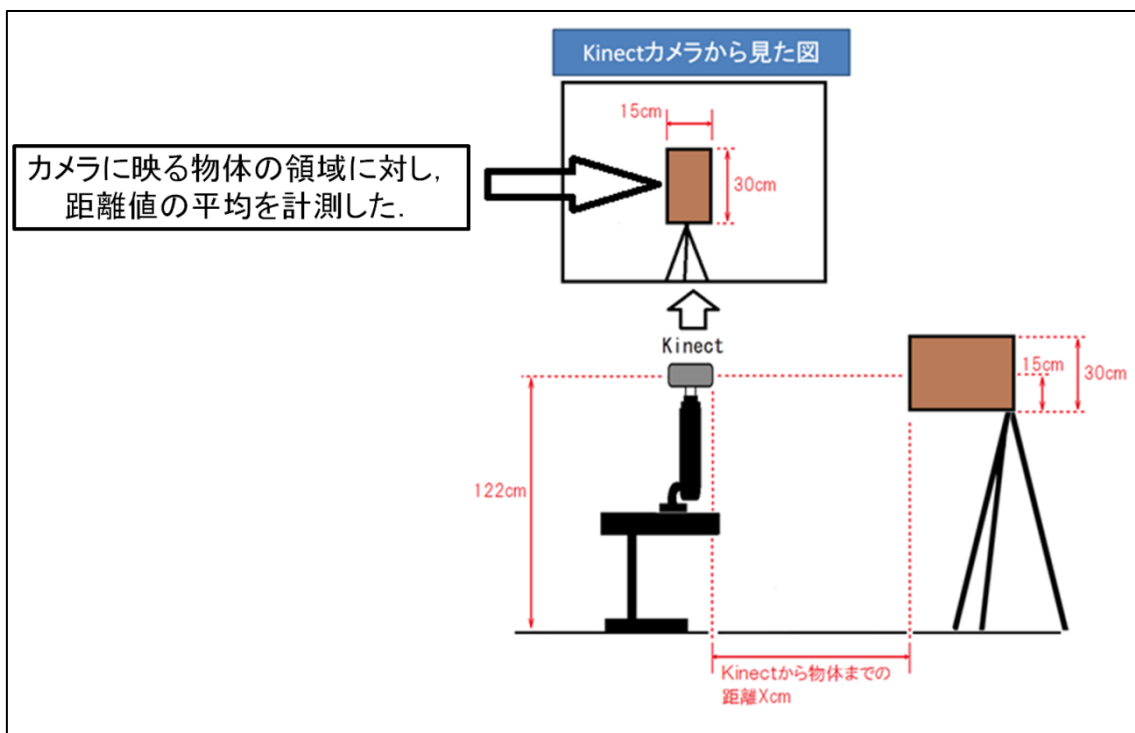


図 B.1 距離値と実際の距離の関係を調べるときの撮影方法

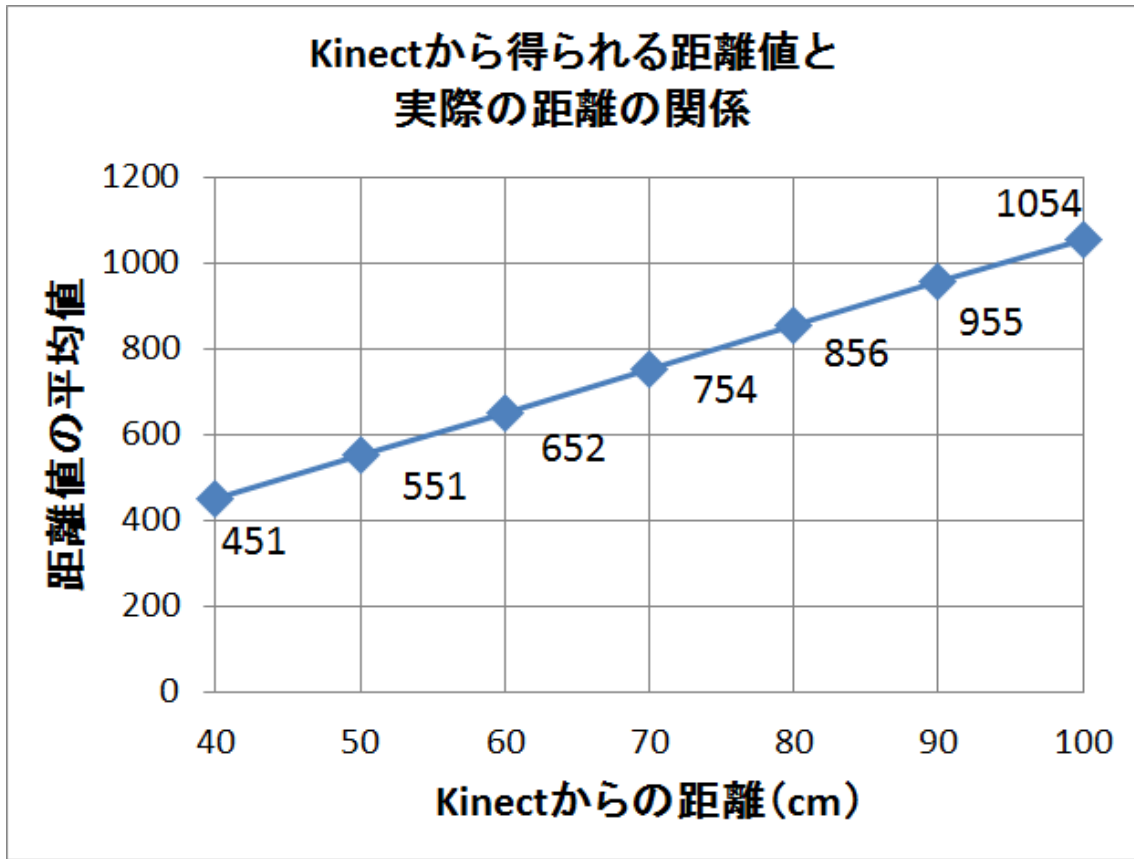


図 B.2 距離値と実際の距離の関係

付録 C 指文字ごとの閾値 Th_1 と閾値 Th_2 の一覧

第3章で, 本システムの照合成功の条件を決める閾値 Th_1 と閾値 Th_2 を決定した. そこで, この付録では, その指文字ごとの閾値 Th_1 と閾値 Th_2 の一覧を示す.

表 C.1 指文字ごとの閾値 Th_1 と閾値 Th_2 の一覧

指文字	閾値 Th_1	閾値 Th_2	指文字	閾値 Th_1	閾値 Th_2
あ	0.07	0.04	な	0.08	0.05
い	0.16	0.06	に	0.14	0.07
う	0.15	0.08	ぬ	0.12	0.11
え	0.09	0.06	ね	0.11	0.14
お	0.11	0.14	は	0.14	0.00
か	0.16	0.11	ひ	0.19	0.08
き	0.18	0.20	ふ	0.09	0.11
く	0.12	0.15	へ	0.04	0.00
け	0.12	0.02	ほ	0.10	0.03
こ	0.13	0.11	ま	0.15	0.20
さ	0.10	0.14	み	0.20	0.14
し	0.12	0.27	む	0.07	0.05
す	0.10	0.14	め	0.11	0.05
せ	0.17	0.11	や	0.10	0.01
そ	0.20	0.14	ゆ	0.08	0.03
た	0.18	0.08	よ	0.10	0.01
ち	0.19	0.16	ら	0.24	0.21
つ	0.15	0.08	る	0.07	0.12
て	0.06	0.09	れ	0.08	0.05
と	0.10	0.02	ろ	0.11	0.09
			わ	0.11	0.02

付録 D 最適なパラメータを決定するための実験

本論文では指文字を照合するために必要なパラメータについて、記述していない個所があった。付録 D ではそのパラメータについて最適な値を探索し、実験した結果を記述する。実験は指文字を照合するために必要なパラメータ以外、第 4 章の辞書構成に協力した手話通訳者の場合の実験と同じ条件で行い、そのときの照合成功率を計算した。実験したパラメータは以下の通りである。最も照合成功率の高いときのパラメータを①～⑤の順に選択していくことで、可能な限り高精度に標準形と照合を行うことができるようにした。

- ① 手領域画像の大きさ
- ② カーネルパラメータ
- ③ K-mean 法による削減枚数
- ④ 辞書部分空間の次元数
- ⑤ 入力部分空間の次元数

ここで、②のカーネルパラメータとは KOMSM において、非線形部分空間を構成するとき用いる式 D.1 の $2\sigma^2$ のことである。式 D.1 は後で述べる。③の削減枚数とは、ある枚数に設定した場合、ある枚数までデータを削減するという意味である。例えば、削減枚数を 50 枚に設定したならば、50 枚までデータを削減することになる。

D.1 手領域画像の大きさの決定

ここでは、①の手領域画像の大きさを決定する実験を行った。本システムで View 特徴を用いて高精度に標準形との照合を行う場合、2.3.3 の特徴抽出部で記述したように、次元数は画像の大きさに依存する。標準形との照合には手領域画像を用いるが、手領域画像の大きさによって View 特徴の次元数が増減し、照合成功率に影響を与える可能性がある。View 特徴の次元数が大きすぎる場合は計算量も大きくなり、リアルタイムで指文字の照合を行うことが難しくなる。そこで、3.2 の手領域の切り出しと正規化において、画像の大きさの正規化を行うときに、画像の縦横比を 1:1 に固定したまま幅を 10pixels～100pixels の間で 5pixels ずつ大きさの異なる手領域画像をそれぞれ構成した。それぞれ大きさの異なる手領域画像のデータセットを用意することで、それぞれの照合成功率を計算した。実験結果は図 D.1 の通りである。ここでは、②のカーネルパラメータを 0.1、③の削減枚数を 50 枚、④の次元数を 15 次元、⑤の入力部分空間の次元数を 3 次元とした。図 D.1 を見ると手領域画像の幅が 50pixels のときに照合成功率は 94.6% となり、最大となることがわかった。そこで、②～⑤のパラメータを決定する実験を行うときは手領域の画像の大きさを 50×50 pixels とした。

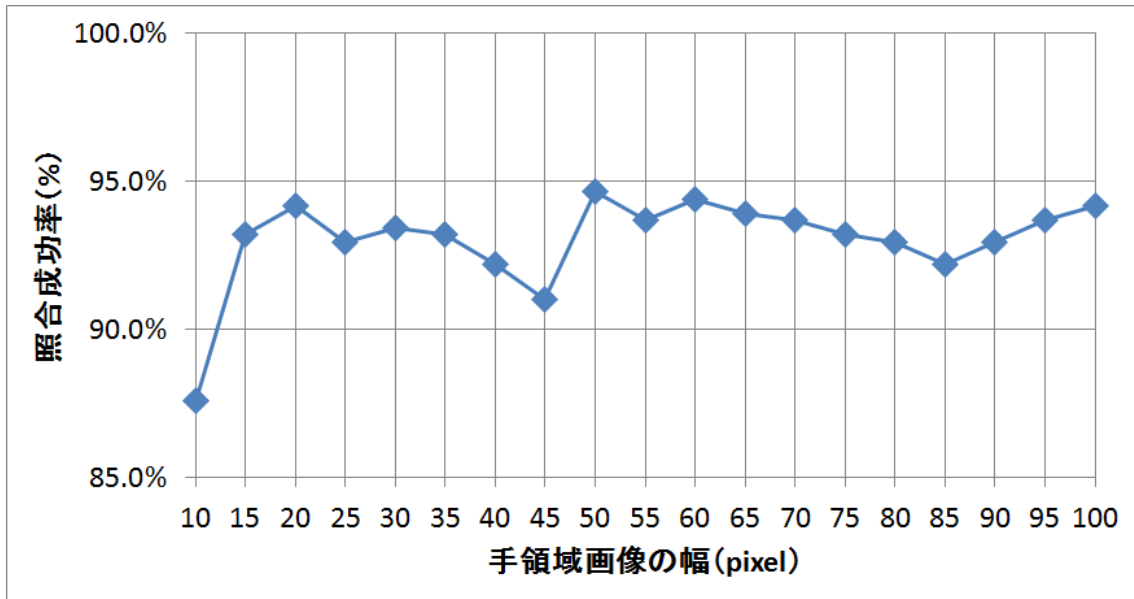


図 D.1 手領域画像の幅を変更した場合における照合成功率の結果

D.2 カーネルパラメータの決定

ここでは、②のカーネルパラメータを決定する実験を行った。本システムでの辞書と識別には KOMSM を用いており、その際にはガウシアン関数を用いた。KOMSM を用いるときは、まず、非線形に分布する特徴ベクトル x に対して非線形写像 φ により極めて高次元もしくは無限次元の特徴空間に写像する。次に、この特徴空間の上で主成分分析を行い、非線形部分空間を構成する。最後に非線形部分空間同士の正準角から類似度を計算し、照合を行う。

ここで、写像したあとの特徴空間で主成分分析を行うためには、第 2 章の式 2.2 より異なる特徴ベクトル同士 x と y のそれぞれの写像 $\varphi(x)$ と $\varphi(y)$ の内積 $(\varphi(x) \cdot \varphi(y))$ を計算する必要がある。しかし、この特徴ベクトル同士の内積を計算するためには、ベクトルの次元が極めて高次元もしくは無限次元であるため、長時間もしくは無限の時間を必要とする。そこで、カーネル関数 $k(x, y) = (\varphi(x) \cdot \varphi(y))$ の関係を利用することで、この特徴ベクトル同士の内積を短時間で計算できるようにする。この計算技法は一般的にカーネルトリックと呼ばれ、本システムでは、カーネル関数として一般的なガウシアン関数(式 D.1)を用いる。

$$k(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \quad (\text{D.1})$$

カーネルパラメータは実験的に決定する必要がある。カーネルパラメータによって照合結果が大きく変動することがある。そこで、カーネルパラメータを 0.1~0.9 で 0.2 ずつ変えながら、そのときの照合成功率を計算することにより、カーネルパラメータを決定した。実験結果は図 D.2 の通りである。図 D.2 を見るとカーネルパラメータが 0.5 のときに照合成功率は 97.0%となり、最大となることがわかった。そこで、③~⑤のパラメータを決定する実験を行うときはカーネルパラメータを 0.5 とした。

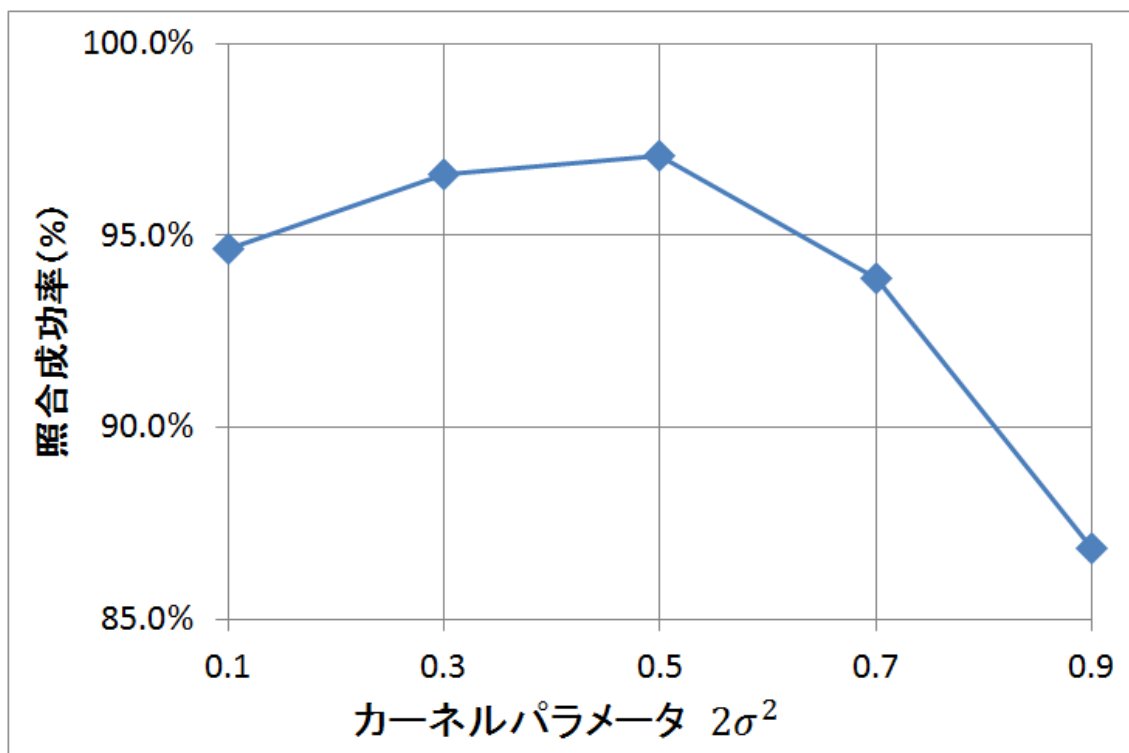


図 D.2 カーネルパラメータを変更した場合における照合成功率の結果

D.3 K-mean 法による削減枚数の決定

ここでは、③の K-mean 法による削減枚数を決定する実験を行った。第 2 章の 2.3.5 で記述したように KOMSM は学習時に多くの計算量を必要とするため、学習データ数のある程度削減する必要がある。そこで、全学習データに対して K-means 法を用いてクラスタリングを行い、学習データ量を削減することにより、KOMSM の高速化を図っている。K-mean 法は以下の流れでデータの削減を行う。

1. 各特徴ベクトルに対してランダムにクラスタを割り当てる。
2. 割り当てたクラスタごとにそれぞれの特徴ベクトルの中心を計算する。中心の計算には特徴ベクトルの要素ごとの相加重平均を用いる。
3. クラスタごとの中心と各特徴ベクトルとのユークリッド距離を求め、各特徴ベクトルに対して、クラスタの中心と最も近い距離のクラスタになるように割り当て直す。
4. 3 の処理において、全ての特徴ベクトルのクラスタの割り当てが変化しない場合は、そのクラスタの中心を新たに特徴ベクトルとし、終了する。そうでない場合は 2 に戻る。

このようにデータを削減することができるが、このとき、割り当てるクラスタの数によって、データを削減する枚数が決まる。例えば、割り当てるクラスタの数を 30 と設定すれば、データの削減枚数は 30 となり、データの枚数によって、照合成功率に影響することが考えられる。そこで、データの削減枚数を 20~80 の間で 10 ずつ変更しながら照合成功率を計算した。実験結果は図 D.3 の通りである。図 D.3 を見ると削減枚数が 50 のときに照合成功率は 97.0% となり、最大となることがわかった。そこで、④、⑤のパラメータを決定する実験を行うときは削減枚数を 50 とした。

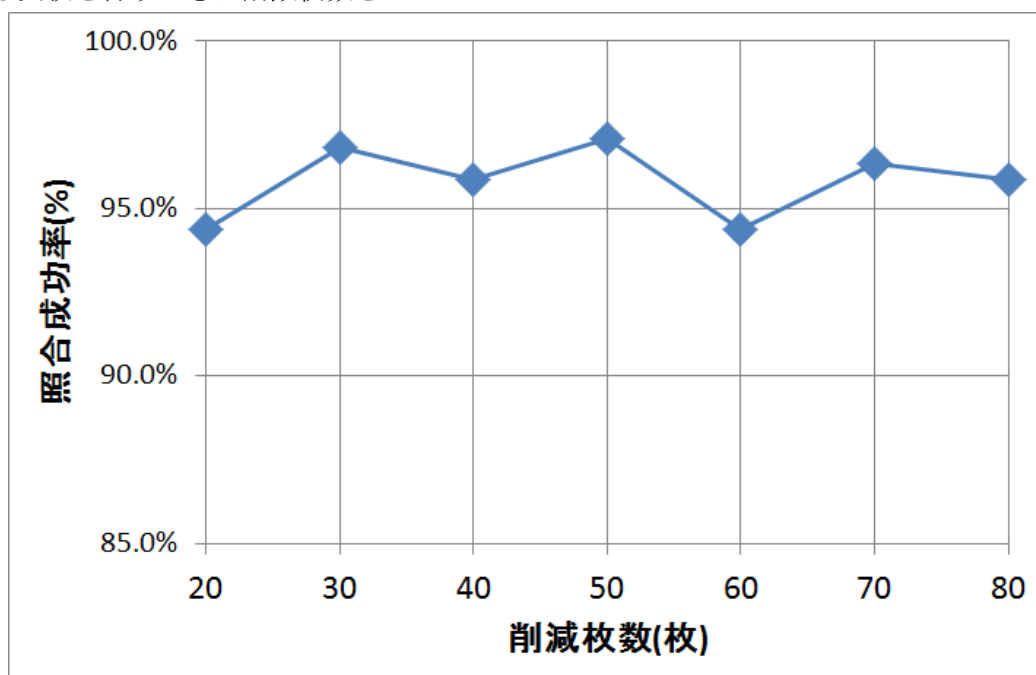


図 D.3 削減枚数を変更した場合における照合成功率の結果

D.4 辞書部分空間の次元数の決定

ここでは、④の辞書部分空間の次元数を決定する実験を行った。辞書部分空間の次元数の上限は学習データ数によって決まる。D.3の実験では削減枚数を50としたことから、学習データ数は50枚となり、辞書部分空間の次元数は最大で50まで用いることができる。そこで、辞書部分空間の次元数を5~50まで5ずつ変更しながら、そのときの照合成功率を計算した。実験結果は図D.4の通りである。図D.4を見ると、辞書部分空間の次元数が10次元のときに最大となることがわかり、その結果は97.3%であった。そこで、最後の⑤のパラメータを決定する実験を行うときは辞書部分空間の次元数を10とした。

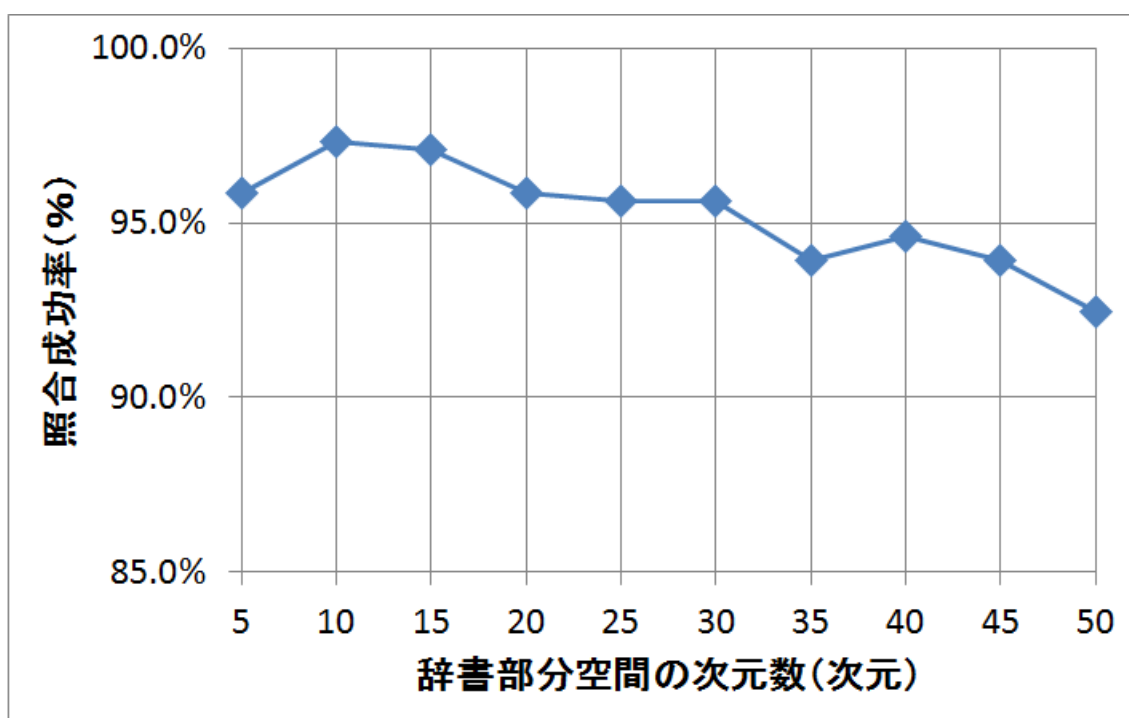


図 D.4 辞書部分空間の次元数を変更した場合における照合成功率の結果

D.5 入力部分空間の次元数の決定

ここでは、⑤の入力部分空間の次元数を決定する実験を行った。入力部分空間の次元数の上限は入力データの枚数によって決まる。第4章の4.1.1では入力部分空間の構成には5枚使用していたことから、入力部分空間の次元数の上限は5である。そこで、そこで、入力部分空間の次元数を2から5まで変更しながら、そのときの照合成功率を計算した。実験結果は表D.1の通りである。表D.1を見ると、入力部分空間の次元数が3のときに最大となることがわかり、その結果は97.3%であった。以上より、本システムを構築するにあたって、最適なパラメータは①の手領域画像の大きさでは50×50pixels、②のカーネルパラメータは0.5、③の削減枚数は50、④の辞書部分空間の次元数は10、入力部分空間の次元数は3であるということがわかった。

表 D.1 入力部分空間の次元数を変更した場合における照合成功率の結果

次元数	2	3	4	5
照合成功率	96.6%	97.3%	96.8%	95.9%