2022 年度

早稲田大学大学院基幹理工学研究科情報理工・情報通信専攻 修士論文

機械学習を用いたデータグローブによる手話のリアルタイム認識に関する研究

吉　莉
（5121F038-6）

提出日：2023.1.23
指導教員：嶋本薫教授
研究指導名：ワイヤレスアクセス研究

目次

第1章 序論

1.1 研究背景

言語にはさまざまな形態がありますが、体の不自由がない健常者の多くは、音声を媒介とした言語を使って他人とコミュニケーションを取っている。しかし、音声が聞こえない、または話すことができない人、あるいはその両方に障害を持った聴覚・言語障碍であるろうあ者にとって、他人と正常にコミュニケーションをするには容易なことではない。

ろうあ者にとって、手話という特殊な言語を通して外部の世界としかコミュニケーションをとることができない。そのため、ろうあ者は健常者のように容易に社会に溶け込むことができず、健常者よりも生活しづらいのである。

平成 30 年厚生労働省「平成 28 年生活のしづらさなどに関する調査結果」[1]により、日本は約 34 万人の聴覚・言語障害者がいる。彼らがよく利用されているコミュニケーション手段として、補聴器のほか手話・手話通訳も挙げられている。

手話は、手や身振りの形・動き・位置によって意味を表し、顔の表情や唇の動きで補完する視覚に基づく言語である。しかし、手話の学習コストが高く、地域や国によって表現方法が異なるため、ニッチな言語になっている。手話は社会的に普及していないため、ろうあ者たちが教育・医療・労働などにおいて様々な困難に直面している。

日進月歩の科学進歩に伴い、ろうあ者たちが直面している困難を現代の科学技術で解決し、ろうあ者と健常者の間のコミュニケーションギャップを埋めようとして、先進的な IT 技術を用いて手話の意味を解析し、健常者に理解しやすいテキストや音声に変換する方法 — 手話認識（Sign Language Recognition, SLR）が研究されてきた。

手話認識（Sign Language Recognition, SLR）は、コンピュータが人間の手話を理解することを目的として、手話の習得や機械学習、パターン認識など様々な学科の知識を有することが求められ、非常に難しい課題である。手話認識の研究は、前述の科学分野の発展に寄与するだけでなく、ろうあ者向けのテレビ番組や手話教育など、生活の様々な場面において応用することができ、人類社会の調和ある発展を促す上で非常に重要なポジションである[2]。

IT 技術を駆使して手話認識システムを構築することができたら、手話の学習経験の

ない健常者が理解できない手話を翻訳することが可能となり、ろうあ者たちが心の奥底にある考えを障壁なく表現し、健常者がその意味を読み取ったうえで彼らをより深く理解することができるのであろう。ろうあ者と健常者のコミュニケーションの問題を解決されると期待でき、これにより、聴覚障害者がより社会に溶け込み、健常者と同じように教育を受け、仕事をし、病院に通い、生活の品質を向上させることができると考える。

そのため、手話認識に関する研究は非常に有意義である。

1.2 日本手話

言語は、形式と意味との対応関係に関する知識の総体である[3]。ここでいう形式とは、語句、節、また聴覚上の形式（音声）、視覚上の形式（文字、手話）であって、ある表現形式が常にある特定の意味と結びつく、その対応関係の知識の総体を言語という。

日本手話は言語のひとつであり、ろうあ者の間で自然発生した、日本語とは異なる体系を持つ言語の表現形式である。日本手話は、日本語の単語を手話単号に置き換えて、そのまま並べて意味を表す。その一例を示す。

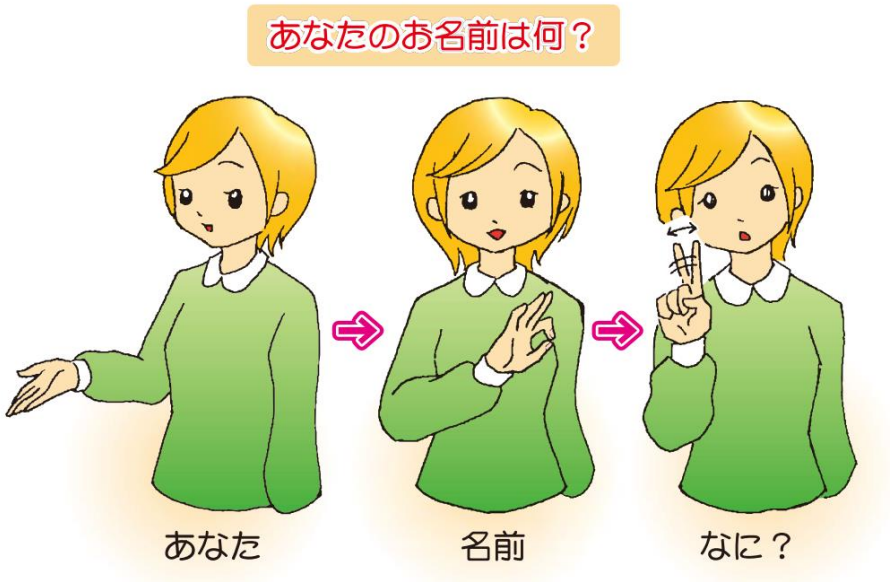


図 1.1　手話の例(兵庫県手話ハンドブック[4]より引用)

日本手話の文章の組み立てる際に、基本的なルールは二つがある[5]。一つ目は、文章を単語にして表現すること。二つ目は、助詞(「～は」や「～が」、「～を」など)は省略するして表現すること。以下の例を用いて、日本手話と日本語の違いを説明する。

日本語	昨日、家までタクシーで帰った
日本手話	昨日＋家（まで）＋タクシー（で）＋帰る

「昨日、家までタクシーで帰った」という文を日本手話で表現する際に、「まで」「で」という助詞を表現しない。このように日本手話は単語を並べることによって文章の意味を表現することができる。

さらに詳しく分類すると、手話には、図 1.2 に示す言葉を表現する純粋な「手話」と、図 1.3 に示す 1 文字 1 文字を表現する「指文字」の 2 つがある。

図 1.2　あいさつの手話(久喜市 手話、指文字の紹介[6]より引用)

図 1.3　指文字(久喜市 手話、指文字の紹介[6]より引用)

第2章 関連研究

現在、手話認識（Sign Language Recognition, SLR）の研究方法は主に 2 つある。コンピュータービジョンを利用した手話認識方式とウェアラブルデバイスを利用した手話認識方式の 2 種類である。コンピュータービジョンによる手話認識方式は、一般的に、カメラで撮影した画像データを使用してジェスチャのセグメンテーションと特徴抽出処理を実行した後、機械学習または深層学習アルゴリズムを適用して処理された画像を認識して、手話認識の結果を取得する。ウェアラブルデバイスによる手話認識方法は、加速度センサー、曲げセンサ、または筋電位センサ等を使用して手話実行中にセンサデータを取得し、データ処理と特徴抽出を経て、機械学習または深層学習アルゴリズムによってセンサデータを分類し、手話認識を行う。

2.1 コンピュータービジョンによる認識

コンピュータービジョンによる手話認識方式が現在の主流であるが、最初は一般の RGB カメラを使って画像を撮影したりビデオを録画して手話データを取得していた[7]。1988 年、田村ら[8]は 10 種類の手話動作を認識できる肌色閾値に基づく日本手話認識システムを確立し、視覚に基づく手話認識の研究が始まった。 Kinect[9]などの深度カメラの誕生により、手話認識研究に新しい方向性がもたらされた。とは言って、手話データを収集するためのデバイスは RGB カメラを使うにせよ深度カメラを使うにせよ、カメラをデータ収集用のデバイスとして、ウェアラブルデバイスを使うよりは比較的安い価格で済むことができる。現在、コンピュータービジョンによる手話認識手法は、従来の機械学習手法と深層学習手法の 2 種類に分けられる。

従来の機械学習に基づく手話認識は、通常、データの前処理、特徴抽出、および認識モデルの構築で構成される。研究者は通常、特徴抽出および認識モデルの構築により多くの努力を注いでいる。 従来の機械学習方法では、通常、Scale-Invariant Feature Transform（SIFT）や Gradient Oriented Gradient（Histogra of Oriented Gradient、HOG）などのアルゴリズムを使用して手動で特徴を設計し、モデリングと認識に従来の機械学習分類器（SVM や HMM など）を使う。Vogler と Metaxas ら[10]はカメラを

使用して手話入力デバイスを構築し、53 個の手話単語を含む手話データベースを作成した。HMM を使用して手話データを識別し、89.9%の精度を得た。 2010 年、Yang ら[11]は時空間特徴に基づく新しいモデリング手法を提案し、その特徴を SVM に入力して中国語手話認識システムを構築しました。中国語のアルファベット画像に対して検証した結果、アルファベット「F」の正解率は 99.7%を達成した。

HMM と SVM の手法以外、SIFT という手法も手話データ抽出のフェーズにおいて良いパフォーマンスが得られることで広く使用されていた。 2011 年、Dardas ら[12]は SIFT を使用して手話画像の手の形の特徴を抽出し、ベクトル量子化を使用してこれらの特徴を K-means でクラスタリングした後に統合された特徴を 3 次元ヒストグラムにマッピングし、SVM を使用してさらにを分類し、高い認識精度を実現した。 2013 年、インドの研究者 Hartanto ら[13]は SIFT を使用して手話データの特徴を抽出し、24 の英語アルファベットの手話ジェスチャーを 62.6% の精度で認識した。 上記の方法はすべて、2 次元カメラによって抽出されたデータの特徴に基づいて特徴量を作成したが、認識精度は画像データのクオリティと深くかかわっており、特定条件で画像データが求められる。 Kinect などの深度カメラが誕生した後、研究者は新しい研究の道筋を示し、マルチモーダル手話データは手話認識研究にも使用できます。

マルチモーダル AI を利用した手話認識も研究されている。2013 年に Chai ら[14]は、Kinect 深度カメラを介して手話の手の動きの軌跡を収集し、それらを正規化し、テンプレートマッチング法を使用してそれらを識別して、96.32%と高い認識率を得られた。2015 年に Sun ら[15]は、Kinect 深度カメラを使用して、2000 個の手話単語を含むアメリカ手話（ASL）データセットを収集した. データセットには、手話データの色彩情報、深度情報、スケルトン情報が含まれており、SVM を使用して手話を分類し、この方法の実行可能性を実証した。

手話単語に対する識別する研究が盛んで中、連続的手話認識も始まった。1998 年、アメリカの研究者 Starner は、ランダムの 40 個の英語単語からなる短い文章を、複数の角度から手話を撮影し手話の画像データを収集した。手の位置と角度などの情報を特徴として手動で抽出し、HMM によって手話認識システムが構築され、98% という高い精度を実現した[16]。2007 年、Morency ら[17]は潜在的動的条件付き確率場（LDCRF）に基づく手話認識方法を提案した。2016 年、中国科学技術大学の研究者である Yang ら[18]は、Kinect 深度カメラをデータ取得デバイスとして使用し、階層構造と HMM を組み合わせて時系列セグメンテーションと連続手話文の認識を実現した。

人工知能の発展に伴い、画像処理や自然言語処理における畳み込みニューラルネット

ワーク（CNN）などのモデルの優れた成果は、手話認識の研究者に新しいアイデアを提供した。 ディープラーニングに基づく手話認識は、従来の手話認識における特徴抽出とモデリングのステップを組み合わせたもので、ニューラルネットワークを構築することで、手話データの特徴を抽出するだけでなく、手話の時系列モデリングを直接行うことができる。 CNN の強力な特徴学習能力により、従来の手動による特徴抽出から CNN の自己学習方法によって画像の特徴を取得し、特徴抽出フェーズの手動方法への依存を取り除くことができた。2014 年、Pigou ら[19]は、Kinect 深度カメラを通じて手話データを収集し、CNN を使用して手の形や動き等の情報を自動的に学習し、20 個の手話ジェスチャを 91.7%の高精度で認識することができた。CNN が手話認識に適用された後、研究者たちがは、強力な時系列モデリング機能を持つ機械学習手法である HMM を CNN と組み合わせることを提案した。2016 年、Koller ら[20]はハイブリッド CNN-HMM モデルを提案した。CNN の強力な識別能力と HMM のシーケンスモデリング機能をエンドツーエンドの埋め込みによって構成したシステムは、3 つの困難な手話ジェスチャに対して認識を行い、他のモデルと比較して、15% から 38%に精度を向上させることができた。2018 年、Masood ら[21]は、従来の畳み込みニューラルネットワーク VGG-16 を使用して手話の特徴を抽出し、2524 個のアメリカ手話を含む画像データセットを 96% の精度で認識した。

連続的手話認識の研究において、時間分割と冗長情報の処理が課題である。 2018 年に Huang ら[22]は、キーフレームベースの手話認識モデルを提案し、この方法は、手話ビデオからキーフレームを代表情報として抽出し、フレームを単語に変換するタスクに成功した。さらに異なる代表情報に異なる重みを付け、310 の中国語手話単語を含むデータセットで実行したところ、良好な認識結果が得られた. 中国科学院の Mao ら[23]は、CNN と Long Short-Term Memory (LSTM) を使用して手話認識モデルを構築した。CNN を使用して手話データの色彩の空間的特徴を抽出し、LSTM を使用したコーデックで時間特徴とコンテキスト情報を学習を行う。ジェスチャー特徴と手の軌跡特徴を融合し、 90 個の中国語手話単語を含んだデータセッに対して認識を行い、97.8%の精度を得られた。2017 年 Camoz ら[24]は、色彩画像フレームシーケンスとハンドジェスチャフレームシーケンスを CNN-LSTM ネットワークに入力し、得られた特徴を時間分類モデルに入力し、デコードを行う。中国の研究者 Huang ら[25]は、データのラベル付けとシーケンスのセグメンテーションの手間を回避するために、フレームワークは階層を使用して手話ビデオを単語にマッピングし、潜在空間結合セマンティックを構築した。この方法は、2 つの大規模な手話データセットでテストした結果高い精度を達成した。

2019 年に Cuir ら[26]は、ディープニューラルネットワークに基づく連続手話認識フレームワークを提案し、手話動画を直接順序付けられたラベルのシーケンスに変換し、特徴抽出モジュールとして時間融合層を使用し、シーケンス学習モジュールとして双方向リカレントニューラルネットワークを使用し、データの表現力を深層ニューラルネットワークで十分に活用するための反復最適化プロセスを提案した。この手法は、カラー画像とオプティカルフロー画像の融合の研究に貢献し、両データベースにおける認識結果を非常に大きく向上させることに成功した。同年、Zhou ら[27] は CNN-RNN-CTC フレームワークに基づく動的擬似ラベル復号法を提案し、BGRU と 1D-CNN からなる時間統合モジュールを導入して、異なる時間の特徴を統合し、二つの大規模連続手話データセットでより良い認識結果を達成した。

コンピュータービジョンに基づく手話認識方式は、手話実演者を制約から解放し、人と器具の非接触を可能にした。これにより、日常のコミュニケーション習慣に即したものとなり、設備コストも削減することができた。同時に、ディープラーニングを用いたアプローチにより、人力に依存する従来の手法のデメリットを解消し、ニューラルネットワークによってデータの特徴を自動学習することで認識を向上させることができた。コンピュータービジョンを使用した手話認識手法が急速に発展し、深度カメラの登場によってマルチモーダルデータが脚光を浴びる中、より多くの研究者がディープラーニング手法を駆使してマルチモーダルデータを処理し始めた。 しかし、マルチモーダルデータにはデータの冗長性という問題があり、異なるデータが表現する特徴も大きく異なるため、マルチモーダルデータに基づく手話認識は研究者にとって困難な課題となっている。

2.2 ウェアラブルデバイスによる認識

ウェアラブルデバイスを用いた手話認識では、一般的にセンサーやモーションキャプチャを装着したデータグローブデバイスを用いて手の動きのデータを収集し、このデータから有効な情報を抽出して手の姿勢をモデル化・分類することで手話の認識を実現するのである。

データ収集装置には、手、手首、指などの正確な空間情報や動作軌跡を得るために、指の曲率、指の伸び、手首の向きなどを測定するセンサーが搭載されていることが多い。現在、手話認識研究者が使用しているウェアラブルデバイスは、Zimmerman らが開発した VPL データグローブ[28](複数のセンサーで各指の湾曲度や超音波装置で手の位置・

姿勢を計測し、コンピュータ側で手を 3D モデルにマッピング）と Kramer らが開発したサイバーグローブデータグローブが主流であります。 サイバーグローブデータグローブは 1983 年に Kramer らによって開発され、第 3 世代のサイバーグローブデータグローブ[29]は図 2.1 に示すように、指の部分の屈曲センサーと外筋センサーによって指や手のひらの屈曲度や手首の回転角など重要な情報を測定することが可能である。

図 2.1 Cyberglove III データグローブ(参考文献[29]より引用)

1993 年、Felsand Hinton ら[30]は、手話データを学習する 5 つのニューラルネットワークを構築し、DECtalk システムに接続し、身振りを音声に変換する手話認識システムの構築に成功した。 1996 年、Kim ら[31]は、VPL データグローブを用いて、韓国語 25 単語からなる手の動き情報を含む手話データを収集し、動き分類技術でジェスチャーを分類し、ファジーニューラルネットワークと組み合わせて認識し、85%の精度率を達成した。 同年、Kadous ら[32]は、パワーグローブのデータを用いて 95 語の手話データを収集し、先行研究よりも豊富な手話情報を得たが、技術的に洗練されていなかったため、精度は 80%にとどまった。 2002 年、Hemadez ら[33]はデータ収集装置としてデータグ

ローブを用い、アルファベット 26 文字の手話ジェスチャーを 100%の精度で認識し、個々のカテゴリーでは最悪の場合でも 78%の精度で認識した。2004 年、研究チーム[34]は、データグローブと 2 本の腕の骨格を用いて手話データを収集し、ジェスチャーする際の手の形状、位置、動作軌跡など様々な情報に分類し、テンプレートマッチングを用いて 176 個の手話単語を最大 95%の精度で認識した。 2008 年、Kong ら[35]は、サイバーグローブと電磁トラッカーを組み合わせて英語の手話データを取得し、線形判別による決定木を用いて手話データを分類し、28 の手話単語について 86.8%の精度を達成した。 さらに、データグローブから得られた手話データを識別するために、ベイジアンネットワークを用いて信号をセグメント化し、2 層の条件付きランダムフィールドとサポートベクターマシン（SVM）を用いて手話データを分類するセグメント化、確率的アプローチを提案し、89%の精度を達成した[36]。

2000 年には、中国科学院の Gao らがデータグローブを使って手話データを取り込み、隠れマルコフモデル（HMM）、ニューラルネットワーク、動的計画法を用いて手話認識システムを構築し、91.4%の精度で認識した[37]。 2001 年、Zhang ら[38]は新しいタイプのデータグローブを提案し、これを通じて中国語の手話データセットを収集し、手話認識システムに用いて中国語の手話語彙を認識し、良好な認識結果を達成した。2004 年、Gao らは CyberGlove とトラッカーを使って手話データを共同で取得し、5113 語の手話単語を含む大規模な中国語データセットを構築し、自己組織化特徴マッピングと HMM を使って認識した。 同年、中国科学院自動化技術研究所（IATC）は、手話単語と連続した手話発話の認識において、それぞれ 82.9%と 86.3%の精度を達成した[39]。同年、中国科学院自動化技術研究所の Fu Yujin ら[40]は、手首上のノードに複数のセンサーを設置し、デジタル信号処理（DSP）コントローラを組み込んだ新しいタイプの手袋、CAS-Glove を開発した。この手袋を使用することで、手話表現の安定性と手話認識の有効性を向上させることができた。以上により、ウェアラブルデバイスを用いた手話認識手法は、手話認識において高い精度で手の動き・位置情報を特定することが可能である。

一方、コンピュータービジョンを利用した手法と比べ、ウェアラブルデバイスを使用した手話認識はデータ収集装置においてコストが高いのも事実である。手話実演の際にはかさばるグローブを装着しなければならず、手話実演者が窮屈に感じることも少なくない。

2.3 まとめ

コンピュータービジョンをベースとした手話認識と比較すると、ウェアラブルデバイスをベースと下手話認識は、カメラ種類や撮影現場の環境に大きく依存しないことがメリットの一つである。 また、ウェアラブルデバイスを使うことで、使用者の手の姿勢の変化に関する情報をより直感的に入手することができ、手に関する情報をより正確に収集できるため、この種の手法では認識速度が速く、精度も高くなると予想できる。特に、データグローブに基づく手話認識は、グローブに設置されたモーションセンサーを用いて、手のジェスチャー変化の空間移動軌跡をリアルタイムで取得し、手のジェスチャーと移動情報を迅速に検出し、認識アルゴリズムに従って手話動作を分類し、最後に対応する応答モジュールにマッピングして人間とコンピュータの対話を完成させることができる。

従って、本論文は高い精度でかつリアルタイムの認識を実現させるために、センサーを装着したデータグローブをベースに、機械学習手法を利用したリアルタイム認識システムを提案する。

第3章 提案システム

第 3 章では提案システムについて説明する。

節 3.1 ではシステムの概要について説明する。

節 3.2 ではデータ収集を行うためのデータグローブについて説明し、節 3.2.3〜節 3.2.2 はデータグローブを構成した部分を詳しく説明する。さらに、データの収集対象となる内容は節 3.2.4 で説明する。

節 3.3 は収集されたデータに対する処理方法について説明する。

節 1.1 では 4 つの提案システムが適用をした機械学習アルゴリズムについて説明する。

節 1.1 ではリアルタイム認識システムの考え方や構成について説明する。

3.1 概要

提案したリアルタイム認識システムの目的は、ろうあ者の手話の意味をリアルタイムに翻訳し健常者に伝えることである。そのため、ろうあ者の手話データを取り込むためのデータグローブを作り、機械学習アルゴリズムによって手話データをリアルタイムに分類し、最終的にろうあ者の手話をテキストで表現するシステム構成する。リアルタイム手話認識システムの全体的な設計フレームワークを図 3.1 に示す。

提案したシステムは三つの部分に分ける。

第一部分は、データの収集である。手話の認識を実現するうえで、データの収集は必要不可欠な部分である。節 3.2 で紹介した自作のデータグローブを使用して右手に装着し、Arduino マイコンを経由でセンサのデータをパソコンに転送される。

第二部分は、データの処理である。転送されてきたデータは未加工の生データであるため、そこから不要な情報を削除し、必要な手話情報が含まれている部分のみの切り取りを行い、手話データセットを作成する。作成したデータセットから、平均と分散を計算し、その値を特徴量とし、特徴量データセットも作成する。

第三部分は、リアルタイム認識である。第二部分で作成した特徴量データセットを 4 つの機械学習アルゴリズムモデルにそれぞれ学習と認識を行い、正解率を比較した後、パフォーマンスが良いモデルを選出し、リアルタイム認識に組み込む。

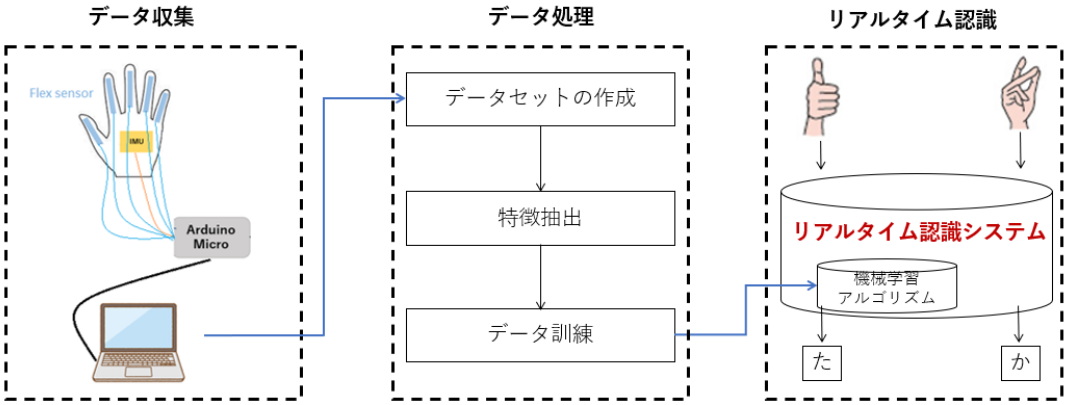


図 3.1　提案システムの構成

3.2 データグローブによるデータの収集

手話の調動[41]は，表 3.1 に示す手指信号（Manual Signals = MS)および非手指信号(Non Manual Signals = NMS)で構成される．手指信号は，指の形・掌の方向・手の位置・大局的な運動により構成され，これらの動きを時間軸上に同時あるいは連続的に提示することにより，主に語の形成に寄与する．非手指信号は，表情・口型・うなずき・視線などの要素で構成され，主に統語論的，意味論的な作用があるとされている．

表 3.1　手話を構成する要素

	手話の構成要素
手指信号(MS)	指の形、掌の方向、手の位置
非手指信号 (NMS)	表情、口型、うなずき、視線

表情、口型、うなずき、視線等の非手指信号は、ろうあ者とのコミュニケーションの

中で目視で確認するができるため、本システムは非手指信号の識別・認識・翻訳はしないことにする。したがって、本システムは手指信号のみに対してデータの収集を行う。これらの手指信号の情報データを収集するための手袋は表 3.2 に示した要素で構成される。

表 3.2　データグローブの構成

データグローブの構成要素	機能
馬術手袋	データグローブのベースとなる
曲げセンサ	指の形のデータを収集する
IMU	掌の方向、手の位置のデータを収集する

3.2.1 曲げセンサ

表 3.1 に示した手指信号である指の形のデータを収集するために、曲がり具合を測定できる曲げセンサを使用する。曲げセンサは対象物の曲げ具合を測定するもので、上層は柔軟なフィルムと感圧層が積層され、下層は柔軟なフィルムと導電線が積層されており、曲げセンサーが曲がると上層の感圧層が下層の断線部を導通し、変形の度合いに応じてセンサーの抵抗値が変化するようになっている。

本論文は、図 3.2 に示す Spectra Symbol's 社製の曲げセンサーを使用する。Spectra Symbol's 社製の曲げセンサーは、高速かつ長寿命であり、本論文では、指を曲げたときに発生する変形信号を取り込み、指の曲げ具合の指標とした。

本論文では、5 本の指を動かす際の動きを考慮して、図 3.3 に示すように、5 つの曲げセンサをそれぞれ指の位置に固定することにした。

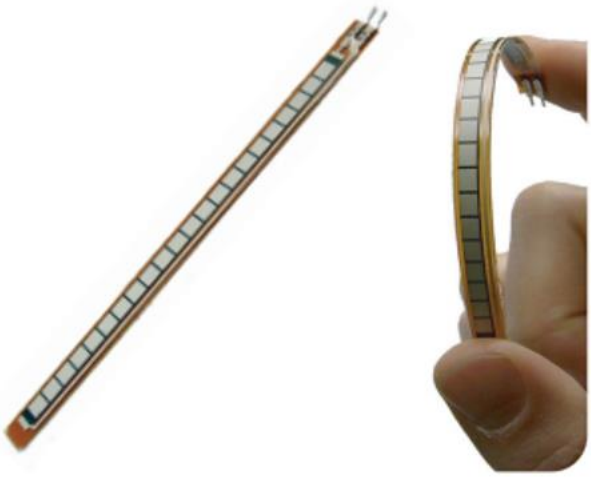


図 3.2　Spectra Symbol's 社製の曲げセンサー

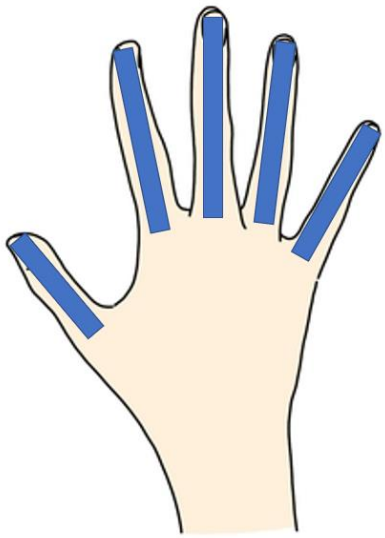


図 3.3　曲げセンサの装着位置

3.2.2 IMU

表 3.1 に示す掌の方向、手の位置に関するデータを取得するため、慣性計測ユニット(Inertial Measurement Uni, IMU)を使用する。慣性計測ユニットは、主にナビゲーションや姿勢の解像度に使用され、6 軸の慣性計測ユニットと 9 軸の慣性計測ユニットに分けられる。6 軸慣性計測ユニットはジャイロスコープと加速度センサーを内蔵し、物体の移動中に 3 軸加速度信号と 3 軸角速度信号を得ることができる。9 軸慣性計測ユニットは加速度センサー、ジャイロスコープ、地磁気センサーを内蔵し、物体の移動中に 3 軸加速度信号、3 軸角速度信号、3 軸地磁気信号の 3 つを得ることができる。 本論文では、手話する際の慣性信号を取得するために、6 軸慣性計測ユニットで十分と考え、図 3.4　6 軸慣性ユニット MPU-6050 図 3.4 に示す MPU-6050 を使用し、図 3.5 に示す位置に固定することにした。

図 3.4　6 軸慣性ユニット MPU-6050

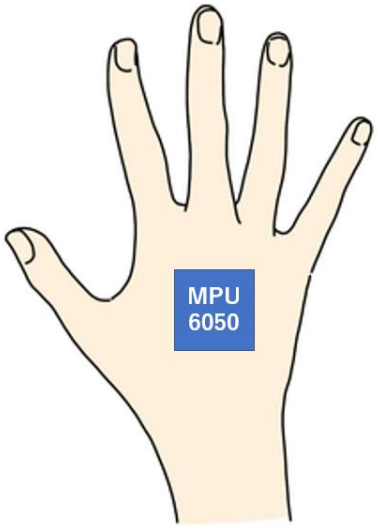


図 3.5　MPU-6050 の装着位置

3.2.3 馬術手袋

手指の変形信号を正確に捉えるためには、曲げセンサーを指に密着させる必要がある。このため、本論文ではデータ用手袋として、図 3.6 に示す LIXADA 社製の乗馬用手袋を使用する。乗馬用手袋は手の大きさが異なる人でも着用できるように伸縮性のある生地を使用している。指部分はユーザーの手の形にぴったりとフィットし，収集した変形信号は外部からの影響をほとんど受けずにユーザーの指の変形を忠実に反映させることができる。

本論文では、手袋の 5 本の指に 5 つの曲げセンサを固定し、手袋の手の甲部分に 1 つの慣性計測ユニットを固定してデータ手袋を構成し、図 3.7 に示している。データの収集は、被験者がセンサ付き手袋を装着し、手話を行った時のセンサからの連続的な時系列データを取得する。具体的な流れを以下に示す。

(1) 被験者が右手にセンサ付きの手袋を装着し、テーブルに両手を平に置く。

(2) 被験者が手話の動作を 1 回行った後に再び手がテーブルに平置きの状態に戻す。

(3) 1 個の手話につき、(2)の動作を 100 回繰り返す。なお、手話及び指文字の動作の速度は指定しない。一回の動作の長さは 1~2 秒である。

図 3.6　LIXADA 社製の乗馬用手袋

図 3.7　データ収集用データグローブ

3.2.4 認識対象

節 1.2 で説明されたように、日本手話は言葉を表現する純粋な「手話」と、1 文字 1 文字を表現する「指文字」の 2 つがある。本論文は指文字に焦点を当てて、指文字で表現された図 3.8〜図 3.11 に示す「あ」行から「た」行までのかなを認識の対象として実験を行った。

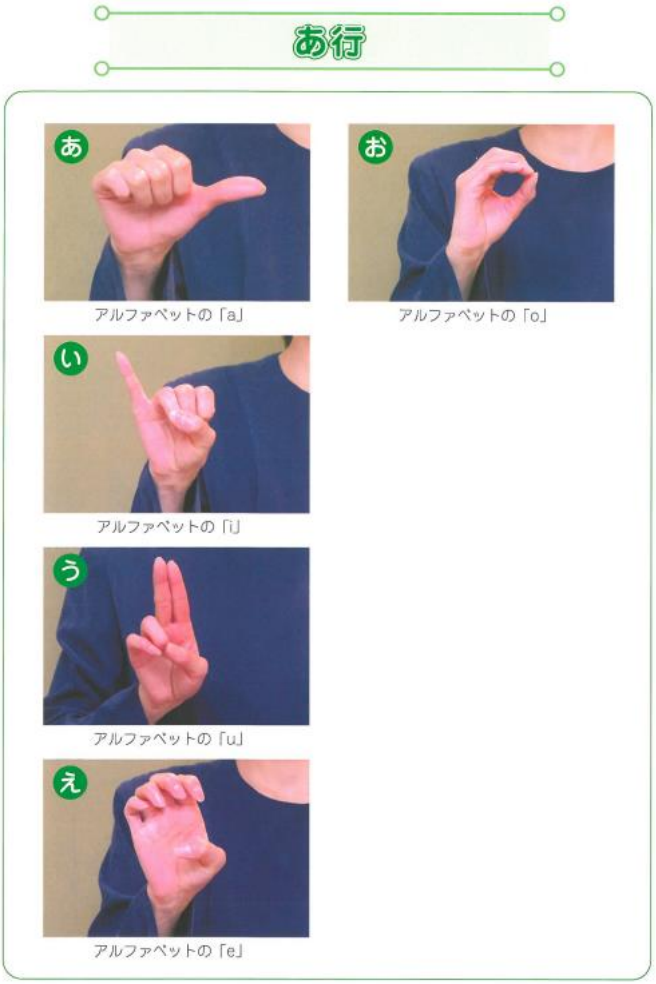


図 3.8 「あ」行の指文字(手話マニュアル[42]より引用)

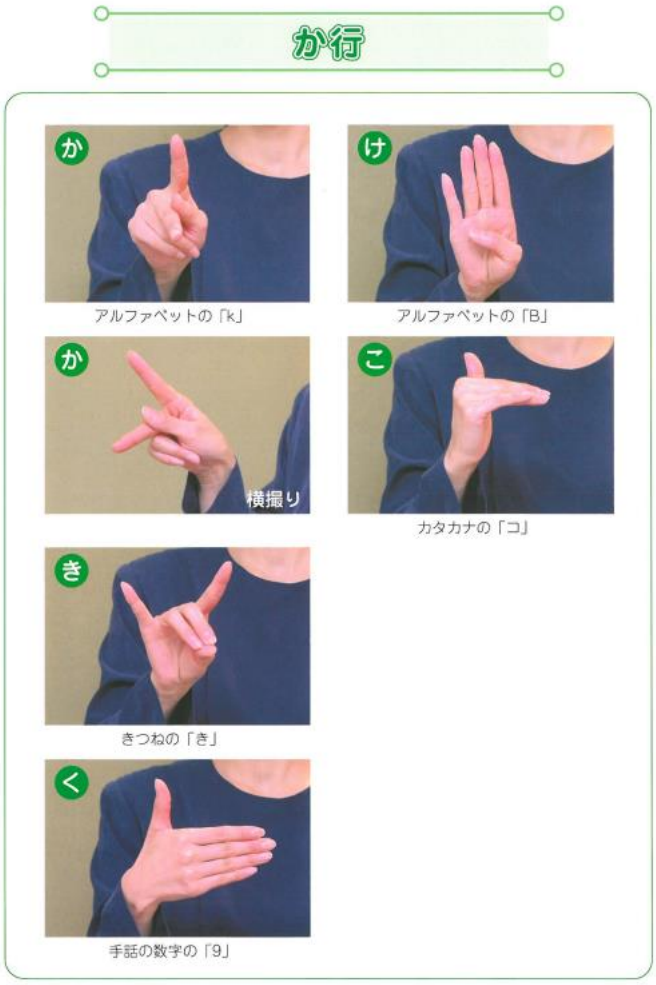


図 3.9　「か」行の指文字(手話マニュアル[42]より引用)

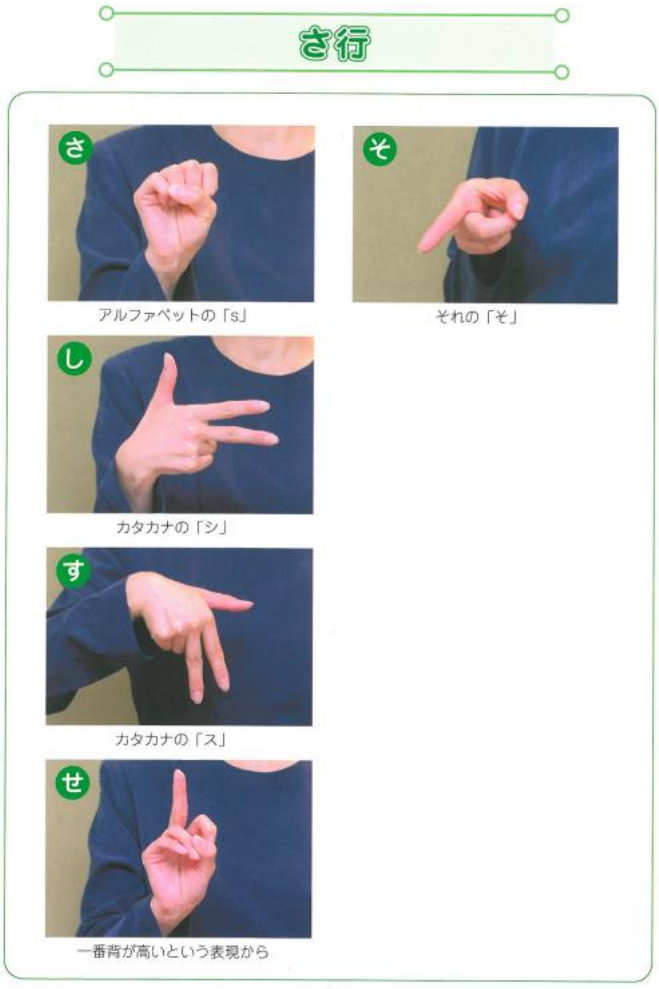


図 3.10　「さ」行の指文字(手話マニュアル[42]より引用)

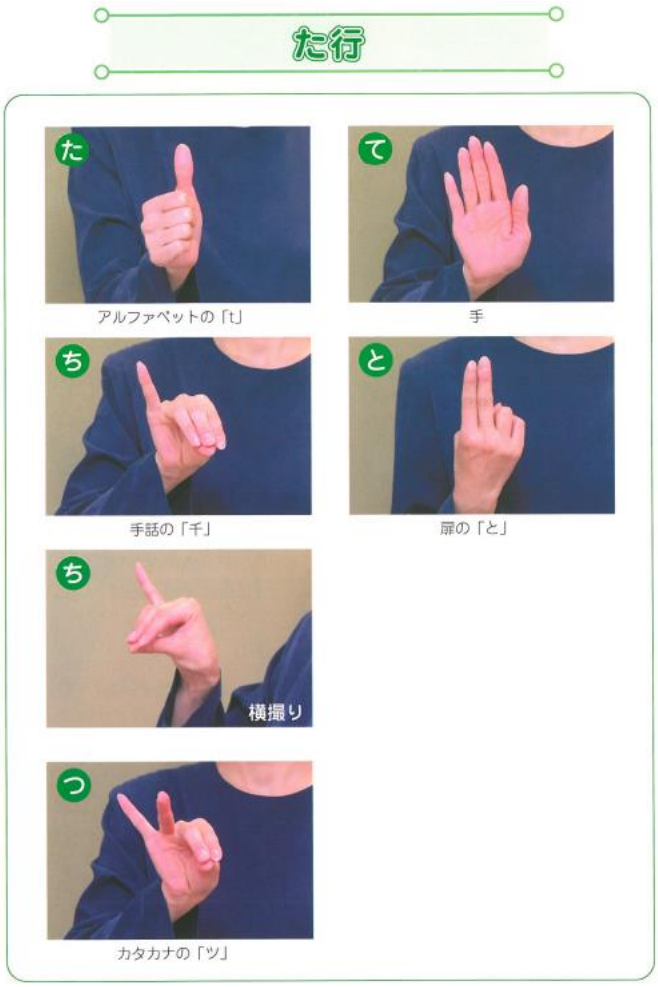


図 3.11　「た」行の指文字(手話マニュアル[42]より引用)

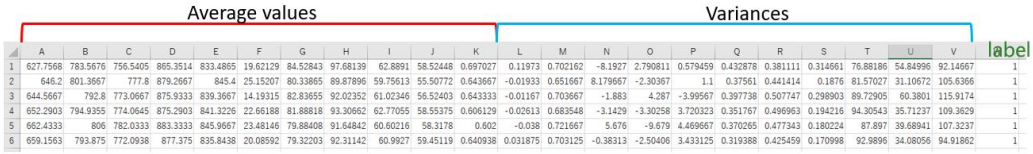
3.3 手話データの処理

節 3.2 のデータ収集プロセスを経て、「あ」行から「た」行の指文字に関する生データすべて取得できた。取得された生のデータ中に不要な情報も含まれており、図 3.13 に示したように、青枠で囲まれた部分のみ指文字情報が含まれている。生データから不要な情報を捨て、指文字情報が含まれている部分のみに対して切り取りを行う。

切り取りを行った後、「あ」～「た」までの 20 個の指文字に対して、データセットを作成する。1 個の指文字に対して 100 グループのデータがあり、1 グループに 11 チャ

ンネル（曲げセンサの 5 チャンネル＋IMU 6 チャンネル）からのデータセットが含まれており、チャンネルごとに平均と分散を計算し、計算した値をモデル学習・認識する際の特徴量とする。

図 3.12 に示したように、最終的に指文字ごとに 22 次元 x100 グループのデータを得ることができる。モデルに指文字に対する学習と認識はこのデータセットを使用する。



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	label
1	627.7568	783.5676	756.5405	865.3514	833.4865	19.62129	84.52843	97.68139	62.8891	58.52448	0.697027	0.11973	0.702162	-8.1927	2.790811	0.579459	0.432878	0.381111	0.314661	76.88186	54.84996	92.14667	1
2	646.2	801.3667	777.8	879.2667	845.4	25.15207	80.33865	89.87896	59.75613	55.50772	0.643667	-0.01933	0.651667	8.179667	-2.30367	1.1	0.37561	0.441414	0.1876	81.57027	31.10672	105.6366	1
3	644.5667	792.8	773.0667	875.9333	839.3667	14.19315	82.83655	92.02352	61.02346	56.52403	0.643333	-0.01167	0.703667	-1.883	4.287	-3.99567	0.397738	0.507747	0.298903	89.72905	60.3801	115.9174	1
4	652.2903	794.9355	774.0645	875.2903	841.3226	22.66188	81.88818	93.30662	62.77055	58.55375	0.606129	-0.02613	0.683548	-3.1429	-3.30258	3.720323	0.351767	0.496963	0.194216	94.30543	35.71237	109.3629	1
5	662.4333	806	782.0333	883.3333	845.9667	23.48146	79.88408	91.64842	60.60216	58.3178	0.602	-0.038	0.721667	5.676	-9.679	4.469667	0.370265	0.477343	0.180224	87.897	39.68941	107.3237	1
6	659.1563	793.875	772.0938	877.375	835.8438	20.08592	79.32203	92.31142	60.9927	59.45119	0.640938	0.031875	0.703125	-0.38313	-2.50406	3.433125	0.319388	0.425459	0.170998	92.9896	34.08056	94.91862	1

図 3.12　モデル学習・認識に使用するデータセットの様子

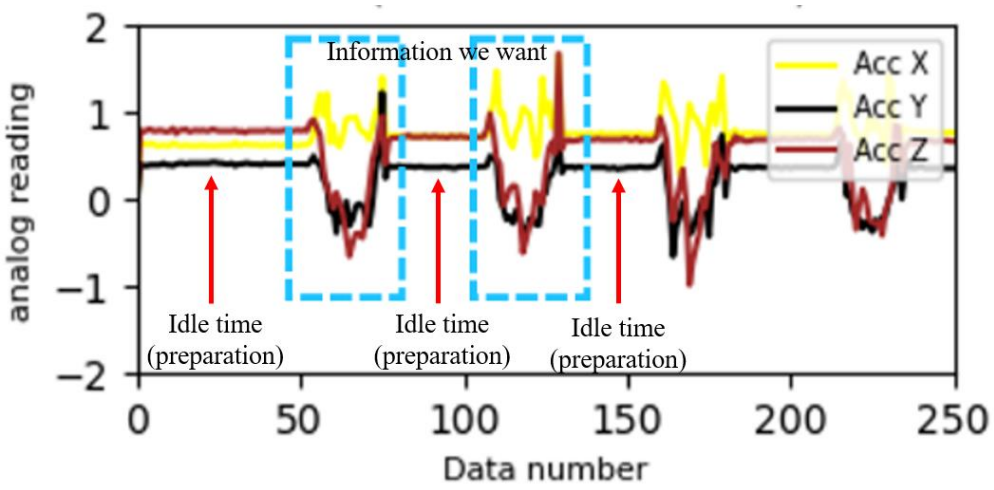


図 3.13 「そ」を表現する際の加速度情報（3 軸）

3.4 機械学習アルゴリズム

機械学習とは，コンピュータコンピューターが大量のデータを学習し、分類や予測などのタスクを遂行するアルゴリズムやモデルを自動的に構築する技術である．大量のデータからパターンやルールを発見し、それをさまざまな物事に利用することで判別や予測をすることが効能であるため，現在の AI の中核技術である．

機械学習は使用する学習データのタイプ、分析方法や状況によって「教師あり学習」、「教師なし学習」、「強化学習」の 3 つに大きく分けられる．

教師あり学習

「教師あり学習」とは，学習データに正解を与えた状態で学習させる手法である．教師あり学習と呼ばれる理由は、そのアルゴリズムがトレーニングデータから学習するプロセスが、先生が生徒に学習を指導しているように見えるからである。正解が用意されていて、アルゴリズムがトレーニングデータに対して予測をし、先生（正解)によってその予測結果が直される、といったような流れになっている。教師あり学習で解く問題で代表的なのが、「回帰」と「分類」 である。

「回帰」とは，連続する数値を予測するものである。平均気温や天候といったデータと売り上げの関係を学習し，将来の売り上げを予測するといったものが「回帰」に当たる。「分類」とは、あるデータがどのクラスに属するかを予測するものである.たとえば、迷惑メールと通常のメールとクラス分けされているデータから文章の特徴とクラスの関係を学習し，新たに来たメールが迷惑メールかどうかを予測するといったものが「分類」に当たる。

よく利用される教師あり学習アルゴリズムは以下の 4 つである。

・サポートベクトルマシン(Support-Vector Machine, SVM)

・K 近傍法(K-Nearest Neighbor algorithm, k-NN)

・決定木(Decision Tree)

・ランダムフォレスト(Random Forests)

教師なし学習

教師なし学習とは、学習データに正解を与えない状態で学習させる手法である。教師

なし学習は上記の教師あり学習と違って、与えられた学習データは正解がなく。アルゴリズム自身がそのデータを探索することで、データの構造やパターンなどを抽出したり、データを分類している。つまり、教師なし学習の最終目標は、与えられた学習データに対する理解を深めるためにデータの基本的な構造や分布をモデル化することである。

強化学習

強化学習は、システム自身が試行錯誤しながら、最適なシステム制御を実現する機械学習手法のひとつである。教師なし学習と同じく正解データは与えらていないがデータの出力を価値づけし、その価値を最大化するための行動をとるようにアルゴリズムを最適化するのである。たとえば、株式の売買でもっとも利益を出すためにはどのタイミングで売るべきか、ゲームで高いスコアを出すためにはどうするかなどの判断処理が強化学習に該当する。

本論文は、センサから収集した手話のデータを元に学習及び認識を行うため，機械学習の中の教師あり学習に該当する。リアルタイム認識システムの構築に精度の高いアルゴリズムを適用する必要があると考え、本論文は 4 つの教師あり学習アルゴリズムのそれぞれの精度の比較を行った。本論文で使用したサポートベクトルマシン，ランダムフォレスト， K 近傍法，決定木について節 3.4.1～節 3.4.4 で説明する．

3.4.1 サポートベクトルマシン

ポートベクトルマシン（Support Vector Machine, SVM）は 1995 年頃に AT&T の V.Vapnik が発表したパターン識別用の教師あり機械学習方法であり、現在知られている方法としては、最も優秀なパターン識別能力を持つとされているデータを 2 つのクラスに分類する問題において SVM の分類能力は優れているのだが，多クラスの分類では計算量が多い、カーネル関数の選択の基準がないなどの課題も指摘されている．2 クラスのパターン識別は SVM の主要任務とは 2 つのクラス間の最も距離の離れた箇所（最大マージン）を見つけ出すことである。

3.4.2 K 近傍法

K 近傍法（K-Nearest Neighbor KNN）は特徴空間における最も近い訓練例に基いた分類の手法であり、パターン認識でよく使われる。K 近傍法は、機械学習アルゴリズムの中でも簡単なアルゴリズムと言われている。理由は、インスタンスの分類を、その近傍のオブジェクト群の多数決で行うことで決定するからのである。

図 3.14 と図 3.15 は K 3 と K=5 の場合，K 近傍法の識別様子を示している。2 種類の学習データ（青色の四角と赤色の丸をベクトル空間上にプロットしておき、未知のデータ緑色の星が得られたら、その道のデータから距離が近い順に任意の K 個を取得し、多数決でデータが属するクラスを推定する。したがって，K=3 の場合，未知のデータの識別結果が青色の四角となり，K=5 の場合，未知のデータの識別結果が赤色の丸となるのである。

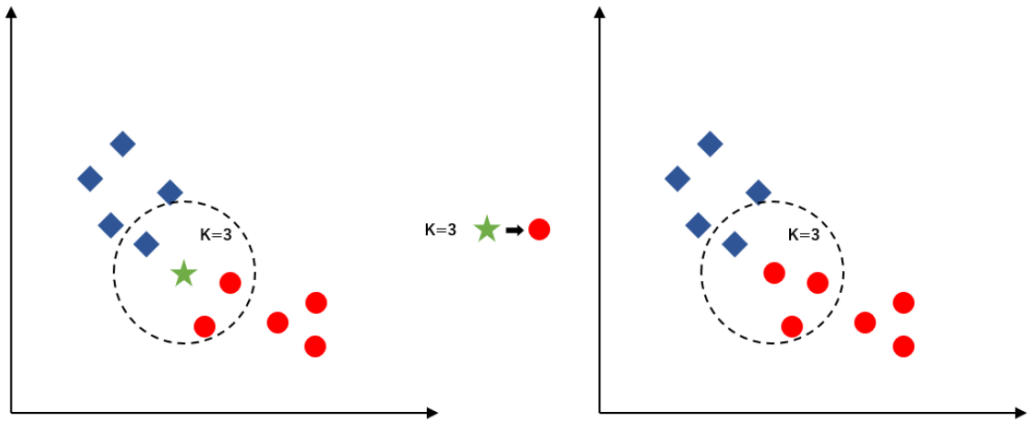


図 3.14　K=3 の場合における K 近傍法の識別結果

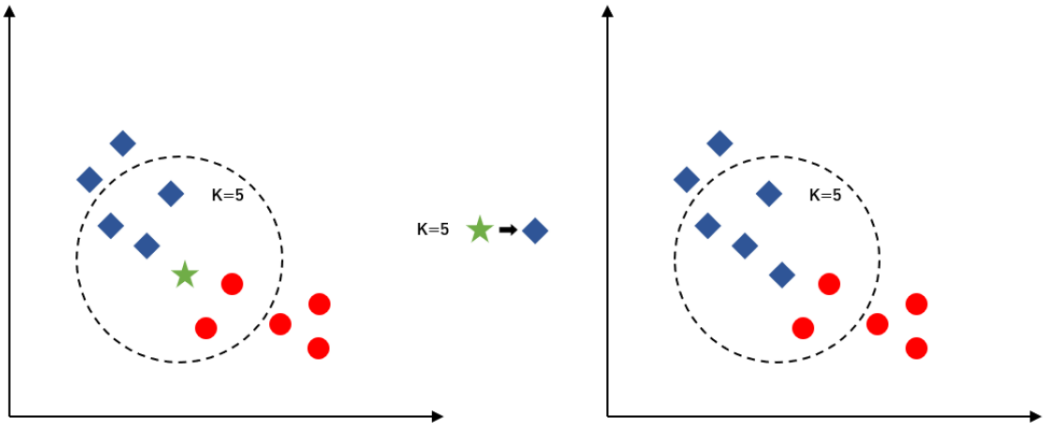


図 3.15　K=5 の場合における K 近傍法の識別結果

3.4.3 決定木

決定木とは木構造を用いて分類や回帰を行う機械学習の手法の一つである．分類木と回帰木の総称して決定木という．分類木は対象を分類する問題を解き，回帰木は対象の数値を推定する問題を解くのである。たとえば，図 3.16(a)は 「温度と湿度がどのようなときに 暑いと 感じるのか？」 といった 問題を分類木で表現したものである．図 3.16(b)は 「温度と湿度がどのようなときに水を何 L 飲むか？」 といった問題を回帰木で表現したものである。

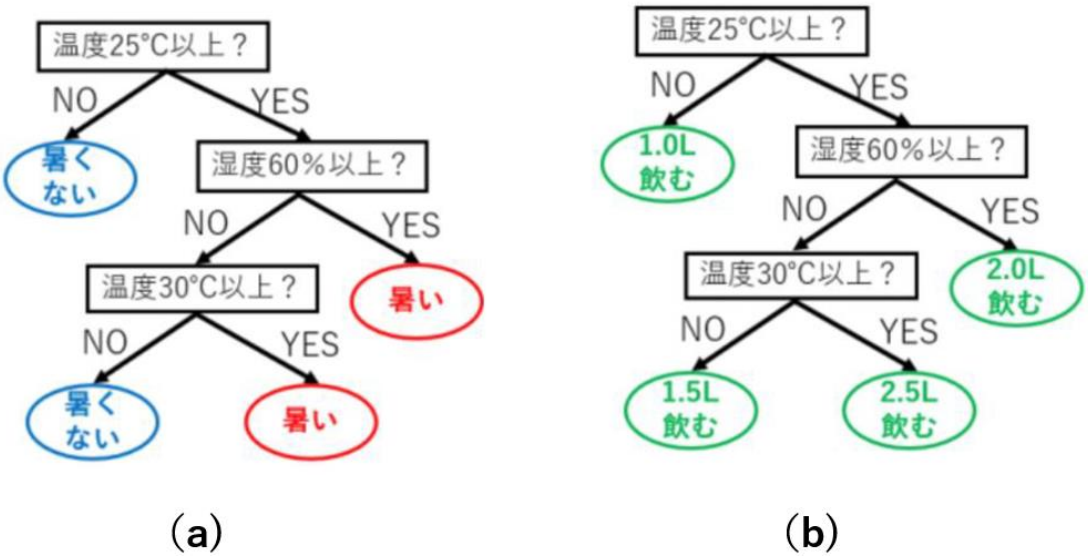


図 3.16　決定木で問題を解くイメージ

3.4.4 ランダムフォーレスト

ランダムフォレストはざっくりというと複数の決定木を集まって,「分類」または「回帰」をする機械学習の手法である。決定木の集合体なので、条件分岐をもった幾つかの決定木をランダムに構築して、それらの結果を組み合わせて、「分類」または「回帰」をすることができる.

3.5 リアルタイム手話認識システム

リアルタイム手話認識システムは図 3.17 に示したような流れで実行する。最初はユーザーの準備時間としてカウントダウンしてからデータの読み取りが開始する。読み取ったデータに対してリアルタイムで平均と分散を計算し、事前に構築した KNN に基づく学習モデルへわたし、近似データを比較する。最も類似度が高い指文字のデータを認識の結果とし、ターミナルに表示する。

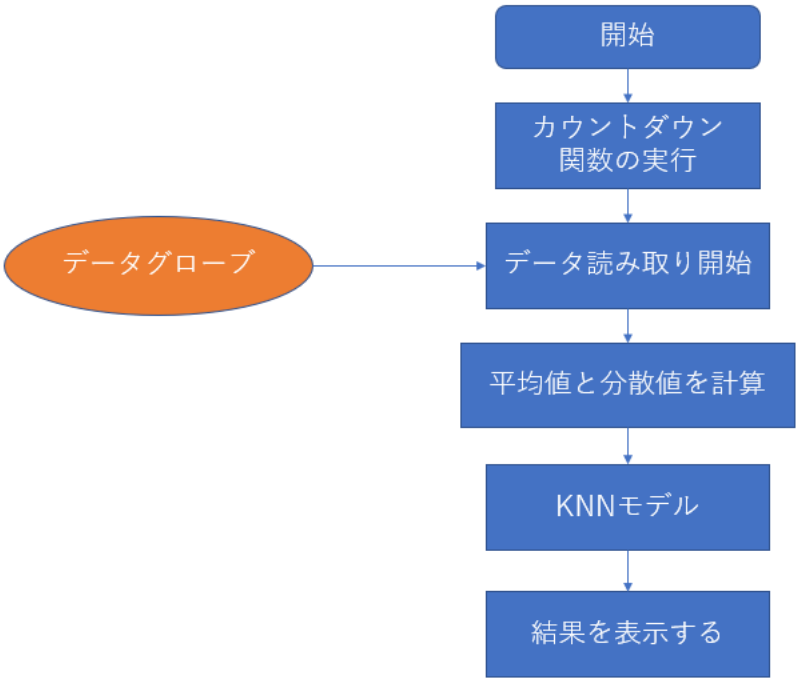


図 3.17　リアルタイム手話認識システムの概要

第4章 結果

4.1 データ収集・処理の結果

4.1.1 データ収集の結果

節 3.2 で説明した方法でデータの収集を行い、「あ」から「た」までの 20 指文字の生データの様子を図 4.1～図 4.20 に示す。

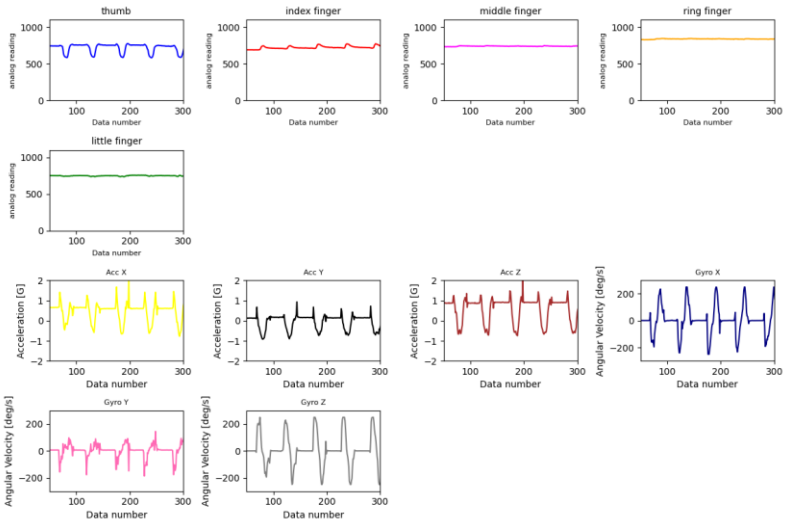


図 4.1 「あ」の生データ

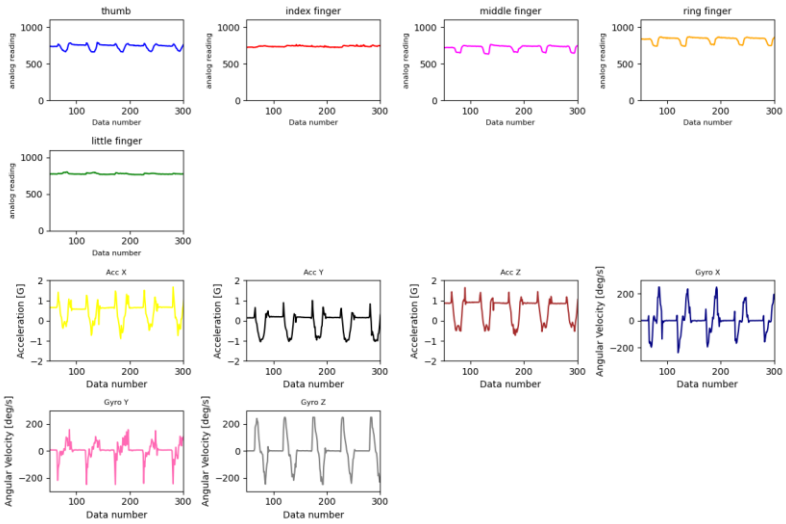


図 4.2 「い」の生データ

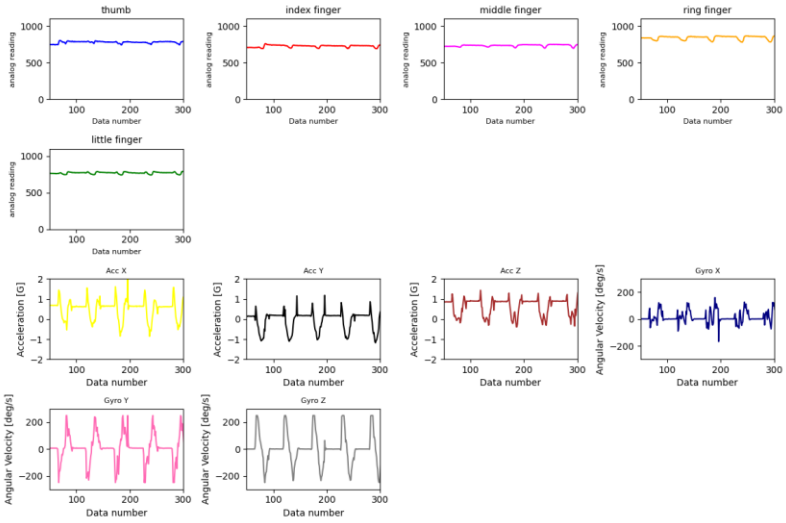


図 4.3 「う」の生データ

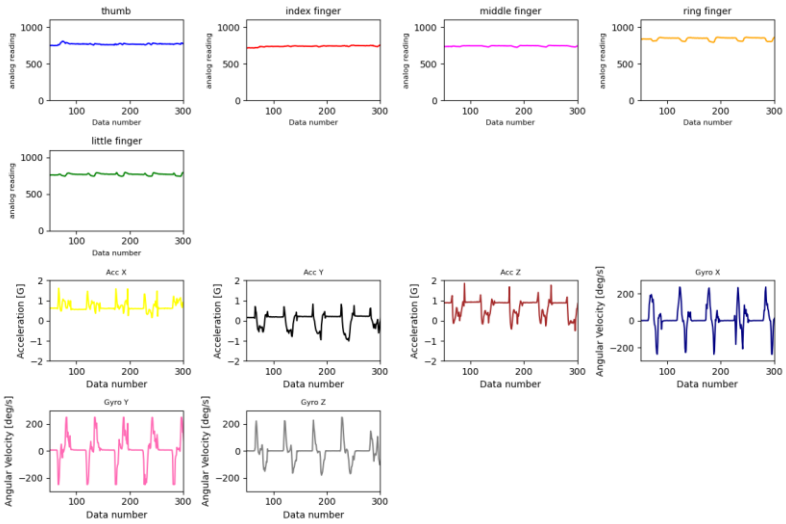


図 4.4 「え」の生データ

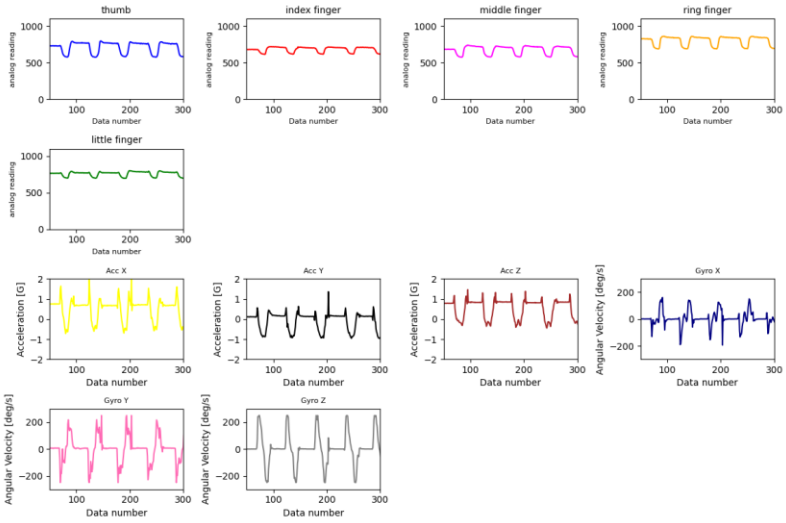


図 4.5 「お」の生データ

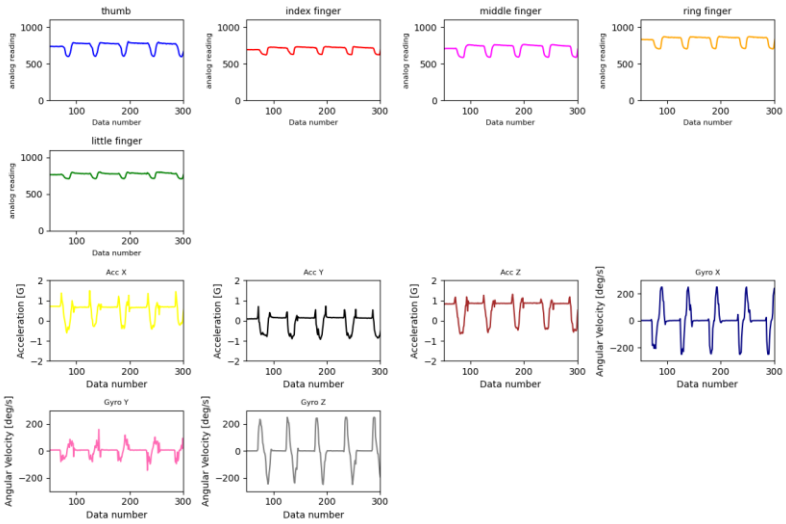


図 4.6 「か」の生データ

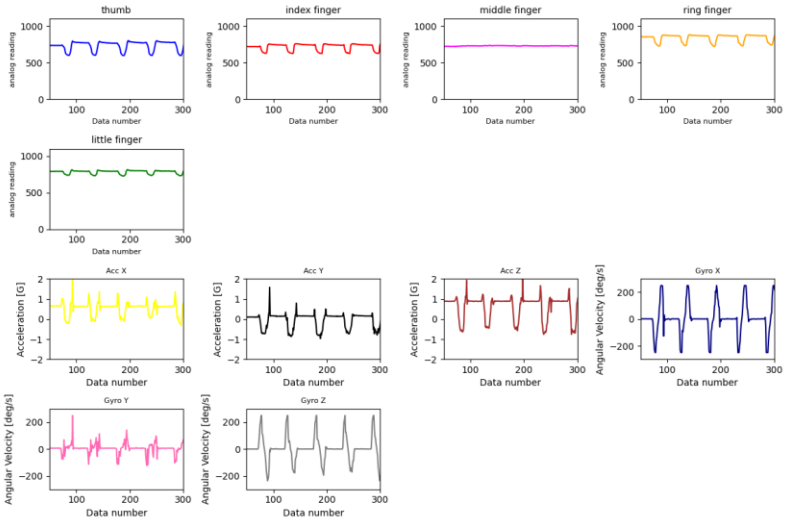


図 4.7 「き」の生データ

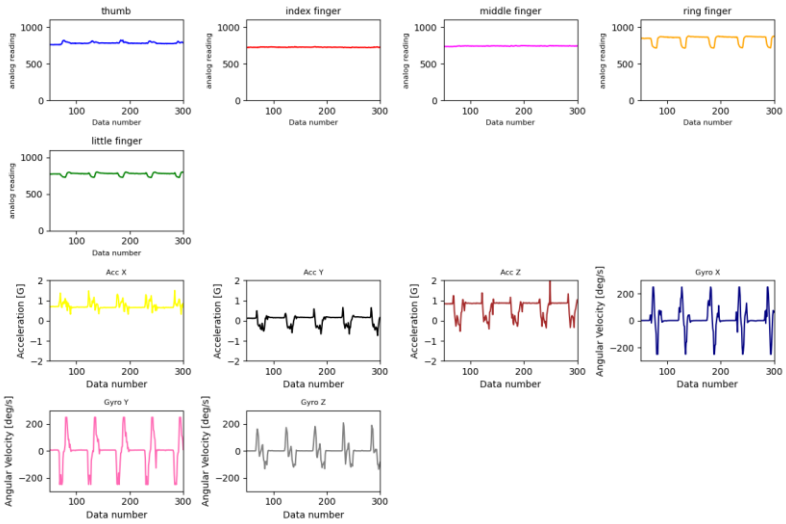


図 4.8 「く」の生データ

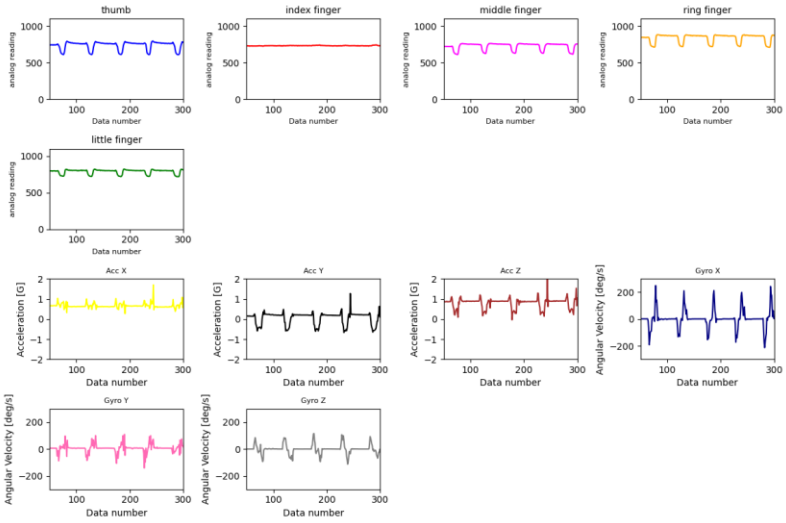


図 4.9 「け」の生データ

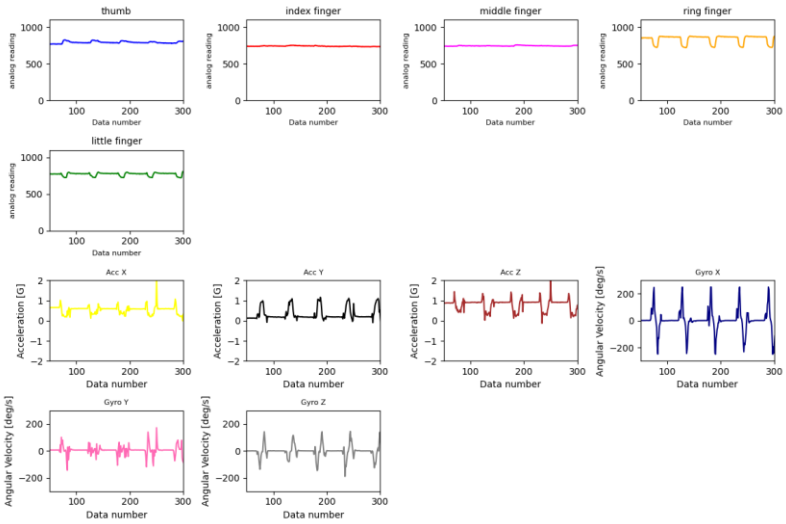


図 4.10　「こ」の生データ

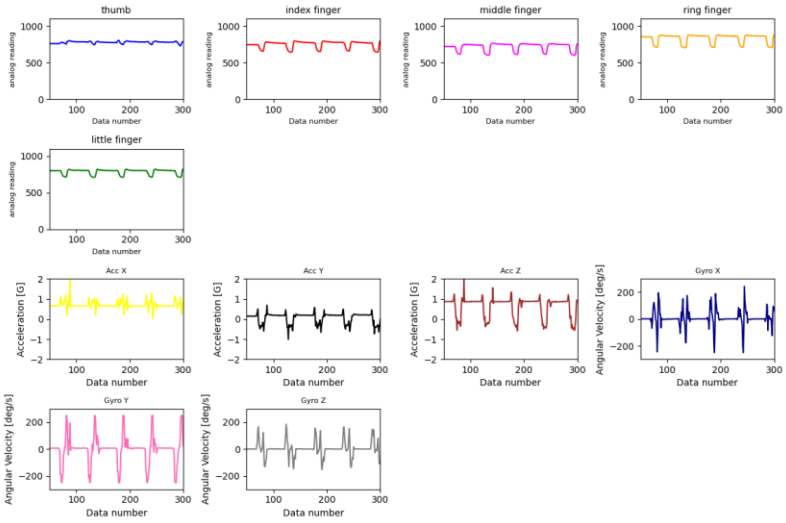


図 4.11　「さ」の生データ

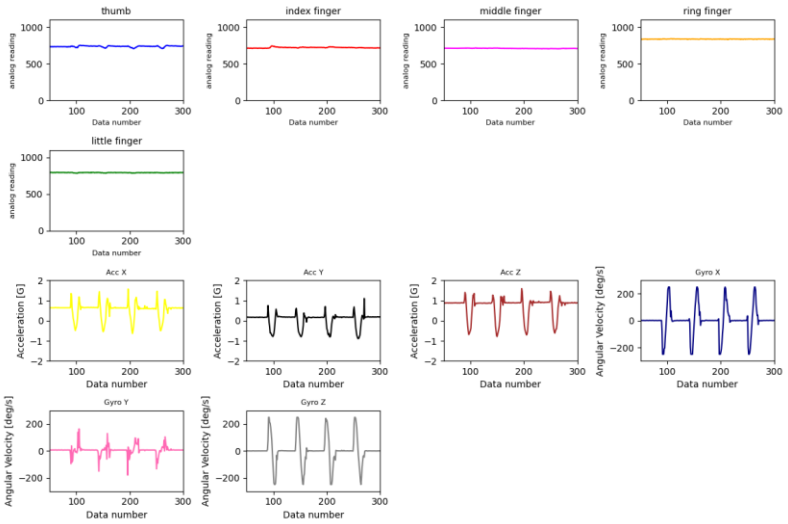


図 4.12　「し」の生データ

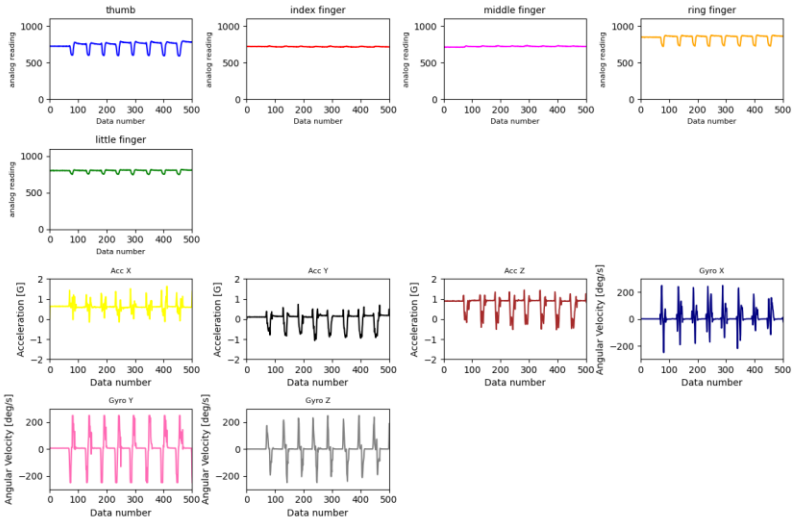


図 4.13　「す」の生データ

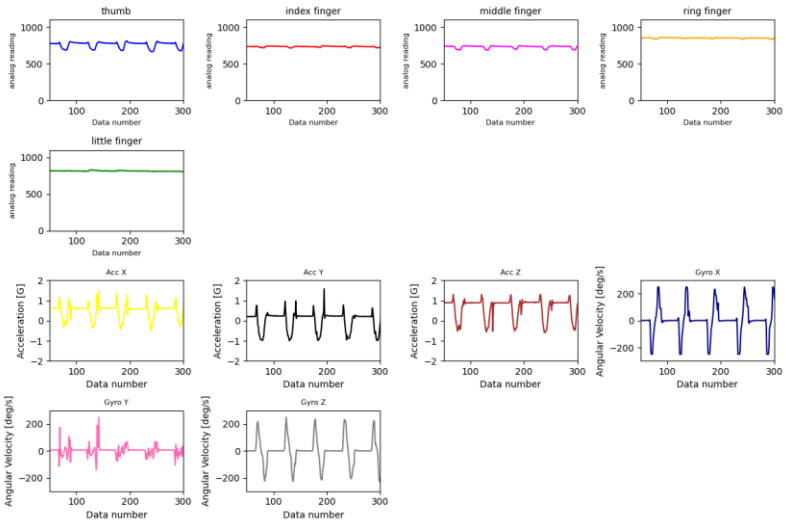


図 4.14　「せ」の生データ

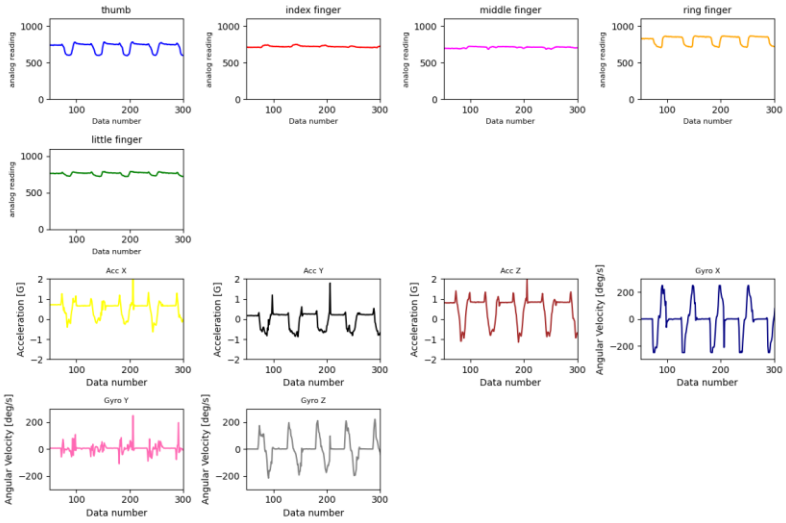


図 4.15　「そ」の生データ

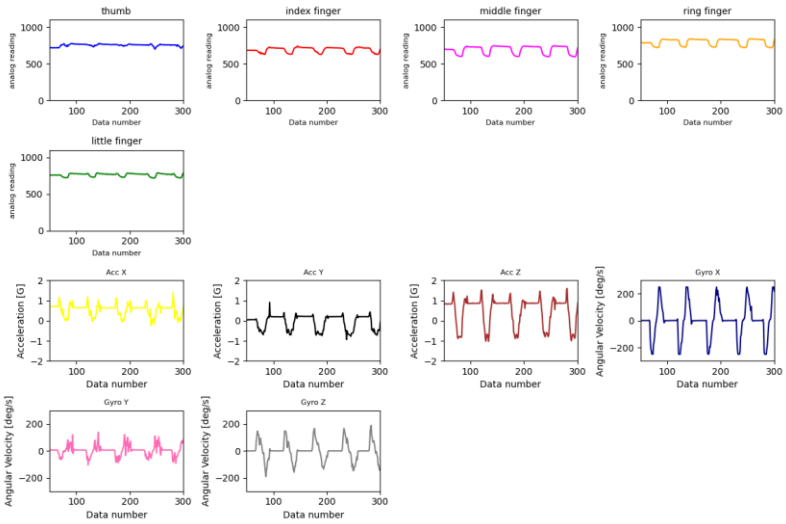


図 4.16　「た」の生データ

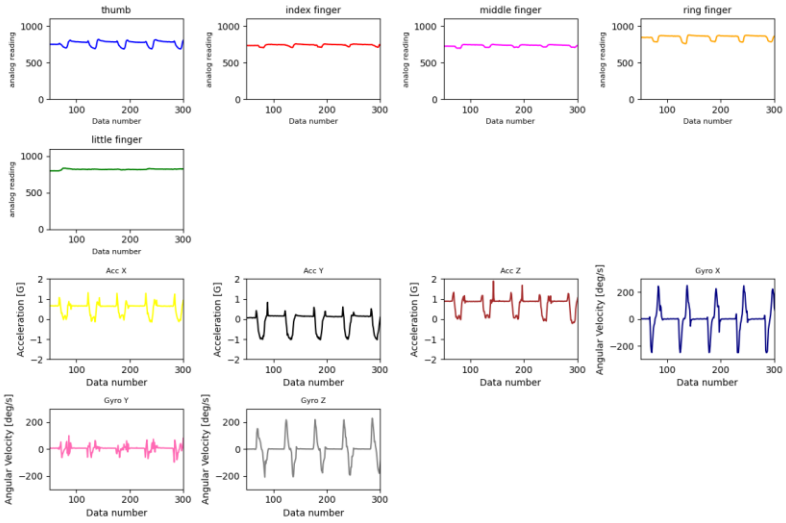


図 4.17　「ち」の生データ

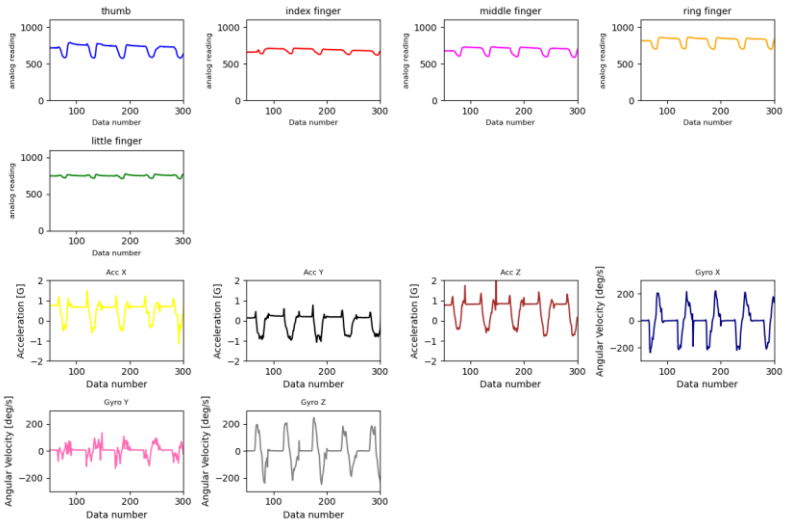


図 4.18　「つ」の生データ

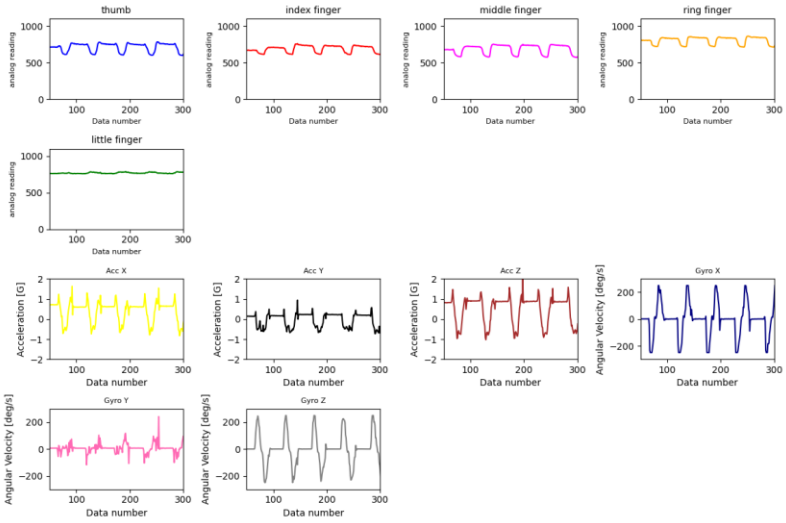


図 4.19　「て」の生データ

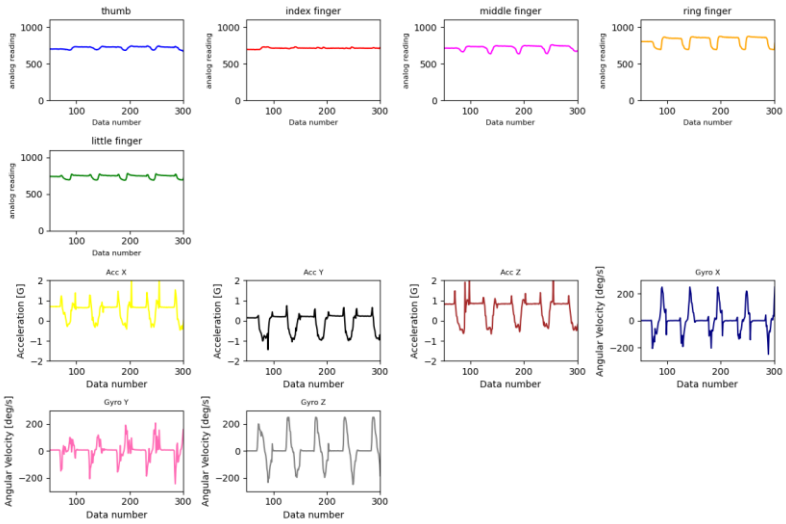


図 4.20　「と」の生データ

4.1.2 データ処理の結果

4.2 機械学習アルゴリズムの比較

表 4.1 は節 1.1 で説明した 4 つの機械学習アルゴリズムを基に構築した学習・認識モデルによって「あ」～「た」行のデータに対して認識を行った後、それぞれの正確率と平均正確率の結果になる。表 4.1 から KNN とランダムフォーレストの平均正解率が 99.75%で、4 つモデルの中で最もパフォーマンスが良いと分かった。この 2 つのアルゴリズムに共通しているのは、「あ」、「か」、「さ」行に対して 100%の認識率を達成できたことである。「た」行に対して 100%の認識を達成できなかったが、それでも高い 99%の精度で認識できた。SVM は、二番目に平均正解率が高い手法となった。「か」行と「さ」行に対して全て認識できたが、「あ」行と「た」行は誤認識が発生した。決定木はこの 4 つの中で平均正解率が最も低い結果となった。とは言って、94.25%という高い平均正解率であった。100%正しく認識できたのは「か」行のみで、「あ」行の誤認識率が最も高く、80%の正解率にしかない結果となった。提案する日本語手話認識システムにおいて、どの機械学習アルゴリズムが最も有効かを調べた結果、KNN とランダムフォーレストが最も有効であることがわかった。

節 3.4.2 と節 3.4.4 で説明した各アルゴリズムの特徴を考慮し、最終的に KNN をリアルタイム認識システムを構築する際のベースとなるモデルにするとした。

表 4.1 各アルゴリズムの正確率の結果

	正確率			
	SVM	K-NN	決定木	ランダムフォーレスト
「あ」行	97.00	100.00	80.00	100.00
「か」行	100.00	100.00	100.00	100.00
「さ」行	100.00	100.00	98.00	100.00
「た」行	99.00	99.00	99.00	99.00
平均正確率	99.00	99.75	94.25	99.75

4.3 リアルタイム認識システム

図 4.21 に、構築したリアルタイム手話認識システムの動作流れになる。ユーザーの準備時間として、まず、3 秒間のカウントダウンからプログラムがスタートし、カウントダウンが終わると「読み込み中」の文字がターミナル側で確認できる。文字が出ると同時に、データグローブがデータの読み込みが始まるため、このとき、手話の動作を行う。手話の動作が終了したら、データがバックエンドで自動的解析され、認識した結果をターミナルに表示し、次の読み込みが始まる。このように、構築したリアルタイム手話認識システムが連続的認識を実現する。

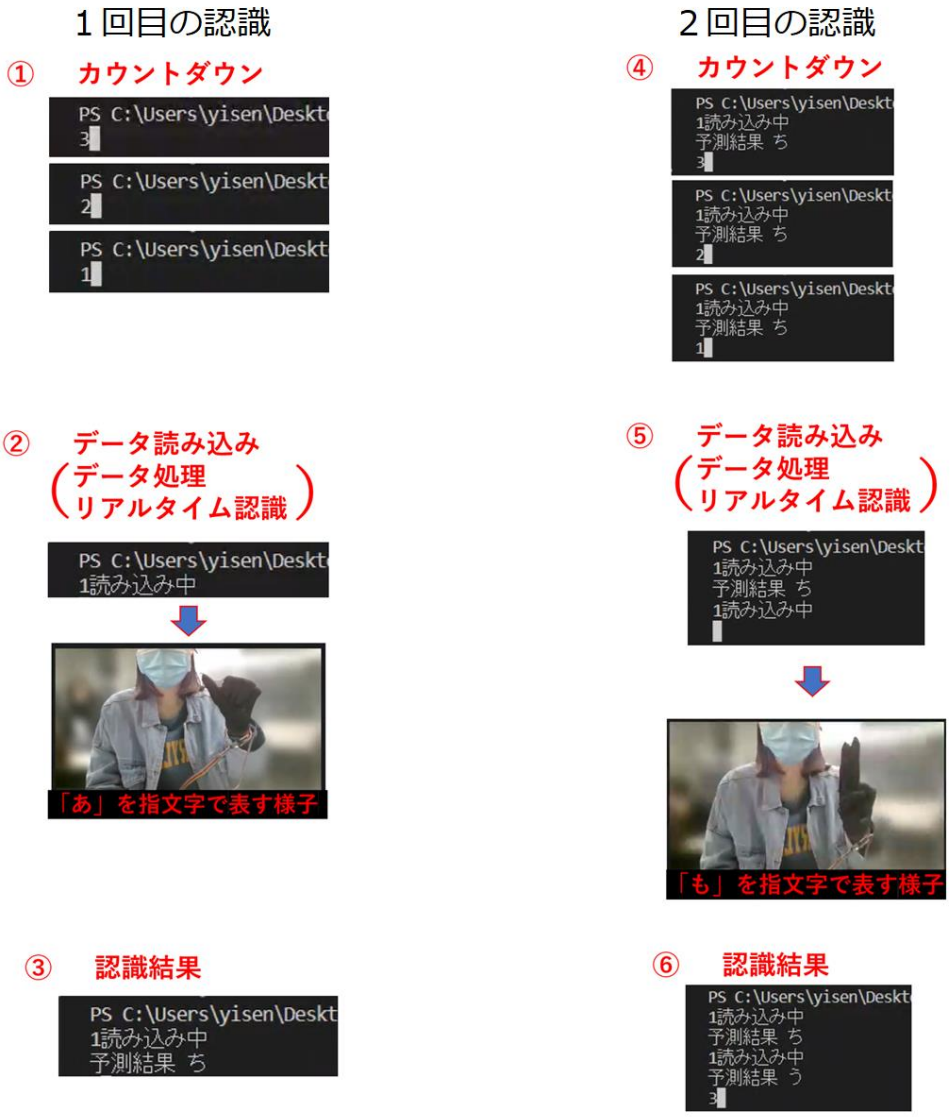


図 4.21　リアルタイム認識システムの流れ

第5章 まとめ

ろうあ者が社会に溶け込めるようにするため、また、手話通訳者の不足と支援機器の高コスト化に対応するため、本論文では、ろうあ者と健常者のコミュニケーションを可能にするリアルタイム手話認識システムを設計・構築した。ろうあ者の手話を通訳する際の利便性を考慮し、軽量で持ち運びしやすいように曲げセンサと慣性ユニットからなるデータグローブを作成した。データグローブから読み取った手話のデータをシリアルポート経由で Arduino マイコンに送信し、「あ」行から「た」行までの指文字に対して時系列データの取得を行い、データの前処理を経て特徴量に関するデータセットを作成した。さらに、本論文が構築するシステムに適合する精度が最も高いアルゴリズムを選別するために、4 つのアルゴリズムのモデルを作成し、精度の比較を行った。4 つの中で精度が高い K-NN を手話の分類と認識に適用するとしてリアルタイム手話認識システムを構築し、連続で手話の認識が可能であることが分かった。本システムは、健常者がろうあ者のコミュニケーションにおいて有用であることを示した。

第6章 今後の展望

これからの手話認識システムは、マルチモーダルフュージョンを組み合わせたり、手話認識と顔認識、指紋認識を組み合わせてアルゴリズムの対話プロセスを多様化したり、深層学習の様々な手法を組み合わせて手話認識プロセスをより解釈しやすくしてエンドツーエンド操作を可能にするなど、現在主流のいくつかの開発と組み合わせることが考えられる。ウェアラブルデバイスによる手話認識は、今後ますます注目され、その応用範囲はますます広がり、さまざまな応用分野でブレークスルーをもたらすであろう。

本論文が提案したシステムをより広く応用するためには、まだ多くの作業と多くの解決すべき問題が残されている。 この論文での改善の可能性を考慮すると、今後の研究は以下の分野に焦点を当てる。

① あらかじめ設定された手話データのスケール。

本論文では、「あ」行から「た」行のみに対してデータの収集を行った。しかし、この量は不十分であり、完全なる手話の通訳を実現するために、多数の手話を認識をさせるために、手話データのスケールをできる限り拡張する必要がある。

② 各種シナリオでの応用に向けた課題。

本論文は、データ収集のプロセスにおいて理想的な実験室の環境で実装しています。冬で室外の風雪環境や夏での高温環境などの制約のある場合では、提案したシステムの安定性はどれぐらい環境に影響されるか等の潜在的な問題点を念頭に置く必要がある。

第7章 謝辞

本論文の作成にあたり、多くの方々にご指導ご鞭撻を賜りました。終始あたたかいご指導と激励を賜りました嶋本薫教授と劉江教授に心から深謝の意を表します。

本研究の遂行にあたり、快く実験に参加頂いた皆様に、感謝いたします。嶋本研究室の皆様には、本研究の遂行にあたり多大なご助言、ご協力頂きました。ここに誠意の意を表します。

最後に，これまで私をあたたかく応援してくれた家族，私を明るく励まし続けてくれた友人に心から感謝します.

第8章 研究業績

[1] Li Ji, Jiang Liu, Shigeru Shimamoto, “Recognition of Japanese Sign Language by Sensor-Based Data Glove Employing Machine Learning”, 2022 IEEE 4rd Global Conference on Life Sciences and Technologies (Life Tech 2022), Osaka, Japan, March 2022.

[2] Li Ji, Jiang Liu, Shigeru Shimamoto, “Smart Glove based Real-time Recognition of Japanese Sign Language employing Machine Learning”, A3 Foresight Program AI-based Future IoT Technologies and Services 2022 Workshop, Tokyo, Japan, December 2022.

第9章 参考文献

[1] 厚生労働者. 平成 28 年生活のしづらさなどに関する調査 [J]. 2018

[2] ONG S R S. Automatic sign language analysis: a survey and the future beyond lexical meaning 27(6): 873-891. [M]. IEEE Transactions on Pattern Analysis and Machine Intelligence 2005.

[3] 佐々木 仁子, 正人 久. 日本手話と日本語 [J]. 言語文化論叢, 2002, 第 10 号

[4] 兵庫県手話ハンドブック. https://web.pref.hyogo.lg.jp/kf10/universal/documents/shuwahandbooka4.pdf

[5] 高橋 亘, 仲内 直子, 宮地 絵美, et al. 日本手話と日本語の構造比較と聾者にわかりやすい日本語の表現 [J]. 関西福祉学大学紀要, 2006, 第 10 号

[6] 久喜市手話（指文字）の紹介. https://www.city.kuki.lg.jp/kenko/shakai_shogai/sho_fukushi/shuwa/shuwa_shokai.html

[7] Kishore P V V, Prasad M V D, Prasad C R, et al. 4-Camera model for sign language recognition using elliptical fourier descriptors and ANN[C]. 2015 International Conference on Signal Processing and Communication Engineering Systems. Guntur, India: IEEE, 2015: 34-38.

[8] Tamura S, Kawasaki S. Recognition of sign language motion images[J]. Pattern Recognition, 1988, 21(4): 343-353.

[9] Zhang Z. Microsoft kinect sensor and its effect[J]. IEEE MultiMedia, 2012, 19(2): 4-10.

[10] Vogler C, Metaxas D. ASL Recognition based on a coupling between HMMs and 3D motion analysis[C]. 6th International Conference on Computer Vision. Bombay, India: IEEE, 1998: 363-369.

[11] Yang, Quan. Chinese sign language recognition based on video sequence appearance modelin-g[C]. IEEE Conference on Industrial Electronics and Applications, 2010, pp. 1537-1542.

[12] Dardas N H, Georganas N D. Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques[J]. IEEE Transactions on Instrumentation & Measurement, 2011, 60(11): 3592-3607.

[13] Hartanto R, Susanto A, Santosa PI. Preliminary design of static indonesian sign language recog-nition system[C]. 2013 International Conference on Information Technology and Electrical Engineering (ICITEE). Yogyakarta, Indonesia: IEEE, 2013: 187-192.

[14] Chai X, Li G, Lin Y, et al. Sign language recognition and translation with kinect [C]. 10th IEEE International Conference on Automatic Face and Gesture Recognition. Shanghai, China: IEEE, 2013.

[15] Sun C, Zhang T, et al. Latent support vector machine modeling for sign language recognition with kinect[J]. ACM Transactions on Intelligent Systems and Technology, 2015, 6(2): 1-20.

[16] Starner T, Weaver J, Pentland A. Real-time American sign language recognition using desk and wearable computer based video[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1998, 20(12): 1371-1375.

[17] Morency L P, Quattoni A, Darrell T. Latent-dynamic discriminative models for continuous gesture recognition[C]. 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, MN, USA: IEEE, 2007:1-8.

[18] Yang W, Tao J, Ye Z. Continuous sign language recognition using level building based on fast hidden Markov model[J]. Pattern Recognition Letters, 2016, 78: 28-35.

[19] Pigou L, Dieleman S, Kindermans P-J. Sign language recognition using convolutional neural networks[C]. European Conference on Computer Vision. Springer, 2014: 572-578.

[20] Koller O, Zargaran S, Ney H, et al. Deep Sign: Enabling Robust Statistical Continuous Sign Language Recognition via Hybrid CNN-HMMs [J]. International Journal of Computer Vision, 2018, 126: 1311-1325.

[21] Masood S, Thuwal H C, Srivastava A. American Sign Language Character Recognition Using Convolution Neural Network[J]. Smart Innovation, Systems and Technologies, 2018, 78:403-412.

[22] Huang S, Mao C, Tao J, et al. A Novel Chinese Sign Language Recognition Method Based on Keyframe-Centered Clips[J]. IEEE Signal Processing Letters, 2018, 25(3): 442-446.

[23] Mao C, Huang S, Li X, et al. Chinese Sign Language Recognition with Sequence to Sequence Learning[C]. China Conference on Computer Vision(ccev). Tianjin, China: Elsevier, 2017, 771:180-191.

[24] Camgoz N C, Hadfield S, Koller O, et al. SubUNets: end-to-end hand shape and continuous sign language recognition[C]. IEEE International Conference on Computer Vision. IEEE, 2017:3075-3084.

[25] Huang J, Zhou W, Zhang Q, et al. Video-based sign language recognition without temporal segmentation[C]. 32nd AAAI Conference on Artificial Intelligence. New Orleans, Louisiana, USA: AAAI 2018:2257-2264.

[26] Cui R, Liu H, Zhang C. A deep neural framework for continuous sign language recognition by iterative training[J]. IEEE Transactions on Multimedia, 2019, 21(7):1880-1891.

[27] Zhou H, Zhou W, Li H. Dynamic pseudo label decoding for continuous sign language recognition[C]. IEEE International Conference on Multimedia and Expo, Shanghai,

China: IEEE, 2019: 1282-1287.
[28]Zimmerman, Thomas, G., et al. A hand gesture interface device[C]. Human Factors in Computing Systems. Toronto Ontario, Canada: ACM SIGCHI Bulletin, 1986, 17(4): 189-192.
[29]Kramer J, Leifer L. The talking glove[J]. ACM SIGCAPH Computers and the Physically Handi-capped, 1988, 39(39): 12-16.
[30]Fels S S, Hinton G E. Glove-Talk: a neural network interface between a data-glove and a speech synthesizer[J]. IEEE Transactions on Neural Networks, 1993, 4(1): 2-8.
[31]Kim J S, Jang W. A dynamic gesture recognition system for the Korean sign language (KSL)[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 1996, 26(2): 354-359.
[32]Kadou M W. Machine recognition of auslan signs using PowerGloves: towards large-lexicon recognition of sign language[C]. Proceedings of the Workshop on the Integration of Gesture in Language & Speech, 1996: 165-174.
[33]Hernandez-Rebollar J L, Lindeman R W, Kyriakopoul N. A multi-class pattern recognition system for practical finger spelling translation[C]. 4th IEEE International Conference on Multimodal Interfaces. Pittsburgh, USA: IEEE, 2002: 185-190.
[34]Hernandez-Rebollar J L, Kyriakopoul N, Lindeman R W. A new instrumented approach for translating American sign language into sound and text[C]. 6th IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea (South), 2004: 547-552.
[35]Kong W W, Ranganath S. Signing Exact English (SEE): Modeling and recognition[J]. Pattern Recognition, 2008, 41(5): 1638-1652.
[36]Kong W W, Ranganath S. Towards subject independent continuous sign language recognition: A segment and merge approach[J]. Pattern Recognition, 2014, 47(3): 1294-1308.
[37]Gao W, Ma J Y, Shan S G, et al. HandTalker: A multimodal dialog system using sign langua-ge and 3-D virtual human[J]. 3th International Conference on Multimodal Interfaces (ICMI 2000). Beijing, China, 2000: 564-571.
[38]张亚新，原魁，杨学良. 一种用于手语识别的新型数据手套[J]. 北京科技大学学报, 2001, 23(4): 379-381.
[39]Gao W, Fang G, Zhao D, et al. A Chinese sign language recognition system based on SOFM/S-R/HMM[J]. Pattern Recognition, 2004, 37(12): 2389-2402.
[40]付玉锦，原魁，朱海兵，等. CAS-Glove 型数据手套运动建模与软件系统开发[J]. 系统仿真学报, 2004, 16(4): 660-662.
[41]寺内 美奈，渡辺 桂子，渡辺 久子, et al. NVSG 形態表記のための日本手話語彙分類法

[J]. IPSJ SIG Technical Report, 2013, Vol.2013-NL-213 No.9
[42]手話マニュアル. https://www.city.yao.osaka.jp/cmsfiles/contents/0000031/31539/sonota.pdf