論 文 の 要 旨

深層学習を用いたセンサグローブによる

連続指文字認識システムの開発

令和 3 年度

筑波技術大学大学院技術科学研究科

産業技術学専攻

土屋 智彦

主指導教員　　白石 優旗　准教授

副指導教員　　加藤 伸子　教授

1. はじめに

近年，音声認識，及びその関連技術についての研究が進み，音声による入力機能を備えた情報機器が広く普及してきている．それに伴い，音声認識による聴覚障害者のための情報保障システムとして，「こえとら」，「UD トーク」，「Cloud Speech-to-Text」などのアプリケーションやサービスが公開されている．これにより，聴者の音声を聴覚障害者が読み取ることが可能になってきている．

一方，聴覚障害者同士の日常会話では，主要なコミュニケーション手段として手話が使われることが多い．ここで，手話言語は音声言語とは異なる特徴を持つ言語であり，手指の屈伸，手の方向，手の動き，顔の表情などによって表現する言語である．よって，聴者が手話言語を習得し，読み取れるようになるためには外国語習得と同程度の学習が必要である．したがって，手話を音声情報，あるいは文字情報に変換し，聴者に対する情報保障を行う手話認識システムが求められている．手話認識システムについての研究がいくつか報告されているものの，音声認識と比較して，実用レベルに達しているとは言い難い．

そこで，本研究では，聴者と聴覚障害者のコミュニケーションを円滑にするための手話認識への第一歩として，日本手話の指文字（Japanese Finger Spelling，以下，JFS）入力インターフェースの実現を目指す．JFS とは，平仮名に 1 対 1 対応した手指の表現のことである．入力インターフェイスには，センサグローブを採用する．カメラを採用する場合は，カメラに手が映るように設置する必要がある．また，カメラは環境の影響を受けやすいため，外出先などでの利用が困難である．さらに，講演などで人の前に立った際にカメラに向けて話す必要があり，発言者に負担がかかってしまう．センサグローブであれば手に装着するだけで自由に動くことが可能なため，利用が容易な局面が多くある．

我々は，センサグローブとして，重量とコストを抑え，かつ着用者が手を動かしやすい手法である導電繊維編み込み手法を採用する．まずは，濁音，半濁音，拗音，長音を加えた JFS の全 76 文字を対象として，単一指文字認識の評価実験を行う．その際，学習モデルとしては，CNN（Convolutional Neural Network）を用いる．次に，JFS の全 76 文字を 2 文字以上で連続して提示した指文字を対象として，連続指文字認識実験の評価実験を行う．学習モデルとしては，単一指文字認識の評価実験の時に構成したモデルに，さらに LSTM（Long Short-Time Memory）を入れ加えたニューラルネットワークを構築し，比較検討する．データセットについては，日本手話の指文字の特徴を考慮し，連続指文字認識実験に最適と考えられるものを新たに提案し，その妥当性についての解析と考察を行う．

本論文では，本研究の目的を達成するために，以下の 3 つの検討および開発に取り組む．

１．日本手話の全指文字の識別を可能

動的指文字の識別に対応できるように，加速度と角速度の情報を加え，効率的な情報を採取できるようにする．

２．データスクリーニングと特徴量の改良

加速度と角速度を用いて，Madgwick フィルタを通して，角度に算出した情報も加え，さらにジャイロドリフトを抑えことにより，識別率の向上を目指す．また，移動平均や動作範囲抽出によって効率的なデータを生成する．

３．連続指文字認識実験

単一指文字認識の発展と，手話認識への発展に貢献するために，連続指文字認識実験を行う．

以上より，考察をまとめ，手話認識システム開発に向けての提案を示す．

2. 単一指文字認識実験

はじめにで述べた１.については，サンプリング周波数と，手の方向と手の動きの次元数を見直すことで，手指動作時の手指の屈伸，手の方向，手の動きの各情報を漏れなく採取できていることを確認した．また，JFS データ採取実験として，実験協力者 20 人に対し，JFS76 文字毎に，11 次元（手指：5 次元，加速度：3 次元，角速度：3 次元）入力で 1 秒間（200 サンプル）のデータ採取を 5 回繰り返し実施した．次に，はじめにで述べた．２．について，Madgwick フィルタを用いて，加速度と角速度から角度（3 次元）を算出し，Sin と Cos に変換した．これにより，合計 6 次元を増やすことで，手指（5 次元），手の方向と手の動き（6 次元）と合わせて 17 次元となった．さらに，サンプリング周波数の見直しを行った．200 サンプル/秒は情報を漏れなく採取できるが，ノイズの混入や学習に時間がかかるという課題がある．よって，移動平均により，4 サンプル/秒にデータ圧縮した．識別実験では，最初に手指と加速度と角速度と角度をそれぞれ分岐し，後の層で結合するニューラルネットワークを構築した．20-fold CV で評価実験を行った結果，約 70.0%の識別率を得られた．しかし，手指の密着問題，キャリブレーション問題，一部採取データへのノイズ混入，採取データ不足などの課題が残された．

3. 連続指文字認識実験

はじめにで述べた３．については，まず安定したデータ採取の実現ため，センサグローブの改良を行った．データ採取では，JFS の特徴を生かしたデータセットを提案した上，64 語の単語を選定した．その際，データ不足を解決するために，データを 12 人分増やし，32 人とした．64 単語毎に，11 次元（手指：5 次元，加速度：3 次元，角速度：3 次元）入力で 8 秒間（120 サンプル/秒×8 秒＝960 サンプル）のデータ採取と動画採取を 5 回繰り返し実施した．その際，毎回最初に特定の動作を行うことで,キャリブレーションをすることとした.その後,加速度と角速度を用いて，Madgwick フィルタにより角度（3 次元）を算出し，Sin と Cos に変換した．合計 6 次元を増やすことで，手指（5 次元）と動作方向（6 次元）と合わせて 17 次元となった．次に，移動平均により 32 サンプル（4 サンプル/秒×8 秒＝32 サンプル）にデータ圧縮した．識別実験では，LSTM だけのニューラルネットワークと，CNN と LSTM の両方を用いたニューラルネットワークとで比較実験を行った．なお，CNN と LSTM の両方を用いたニューラルネットワークについては，最初に手指と加速度と角速度と角度をそれぞれ分岐し，後の層で結合した．5-fold CV で評価した結果，CNN と LSTM を組み合わせたモデルで約 92.1%の識別率が得られた．今回，キャリブレーション問題やノイズ混入については改善されたものの，手指の密着問題，手指の動作状態において静的指文字と動的指文字の区別が困難になってしまう課題，採取データの不足などの課題は残された．

4. まとめと今後の課題

本論文では，センサグローブの重量とコストを抑え，着用者が手を動かしやすい手法である導電繊維編み込み手法を採用し，単一指文字認識実験と連続指文字認識実験をそれぞれ行った．結果，単一指文字認識実験では約 70.0%，連続指文字認識実験では約 92.1%の識別率が得られた．

今後の重要な課題として，以下の 3 つがあげられる．1 つ目は，φ（手が机から離れたとき，机に置くとき，文字間のわたりの 3 つを示すシンボル）のデータ数が JFS76 文字よりも非常に多いことである．2 つ目は，静的指文字と動的指文字を明確に区別することである．3 つ目は，手指の密着問題を解決することである．

手話認識システム開発に発展するために取り組むべきこととして，以下の 3 つが考えられる．1 つ目は，現在，音声データが多いのに対し，手話データが少ないことである．2 つ目は，日本手話のデータを用意する必要があることである．3 つ目は，日本手話に適した学習モデルを開発する必要があることである．

これらの課題に取り組むことで，手話認識システムへの発展を期待できると考える．