

補助事業成果報告書

補助事業の名称	AIを用いた手話動画認識による手話学習支援アプリの研究開発
補助事業の概要	人工知能(AI)による手話動作認識のためのモデル研究、および予め取得した見本となる手話との差異を検出し、手話学習者に差異を提示するモデルを研究する。 さらに5Gなどの高速通信技術を利用したモビリティが高く、簡単に手話を学習できる操作性の良いシステムを提供することにより、手話を話すことのできる介護福祉士などの医療福祉専門職を育成に貢献し、結果的に支援・介護が必要な障害者とのコミュニケーション問題を解消する。

【研究開発の実施内容と成果】

今年度は申請書の研究計画に記載した通り、以下の4点について研究開発を行った。

1. シナリオ作成と動画収集
実証実験対象の業務の選定、シナリオの特定、シナリオに従い手話動画を収集する。
2. システム開発
ろう者の手話を職員向けに翻訳するシステムと、職員の発話をろう者向けに文字変換する機能があり、窓口で双方向のコミュニケーションが可能なシステムを開発する。
3. 実行環境構築
区役所窓口の設定場所も鑑み、高速ネットワークとクラウドシステムを利用したシステム環境を構築する。
4. 実証実験
実際の区役所の窓口をお借りして実証実験を行い、手話翻訳機能を使って窓口業務が可能かどうか検証する。

今年度の研究開発の実施内容について、上記4点に分けて説明する。

1. シナリオ作成と動画収集
札幌市北区、東区の保健福祉課に協力いただき、ろう者が窓口に来ることが多いシーンから4つのシナリオを作成することを決定した。
保健福祉課にて、窓口での会話から、やり取りをまとめていただいた。これをもとに、実証実験の評価のためにシナリオごとにレベル差を設定した。例えば、身体障害者に関する申請ではろう者は「はい/いいえ」の回答で済むように難易度を低くし、障がい福祉サービスの相談申請は会話のやり取り回数や情報量など難易度を高くした。
レベル差を設定したのは、実証実験で想定した業務をすべて実行できなかった場合、どこまでのレベルであれば実施できたのかを評価できるようにするためである。

シナリオ	概要
------	----

身体障害者手帳に関する申請	身体障害者手帳の住所変更(市外転入)、再交付、再交付後の受け取り
補装具費に関する申請	補装具費(補聴器)の新規・再交付申請の手続き
日常生活用具給付事業の申請	日常生活用具(FAX、携帯用会話補助装置など)の購入補助金給付に関する新規・再支給の受付
障がい福祉サービスの相談・申請	障がい福祉サービスによる居宅介護、就労移行支援、就労継続支援などの相談・申請の受付

作成したシナリオをもとに、札幌聴覚障害者協会に協力いただき、ろう者視点でのシナリオ修正と手話翻訳を実施した。翻訳したシナリオおよび手話単語から、深層学習のための手話動画を収集した。収集したデータは以下の通りである。

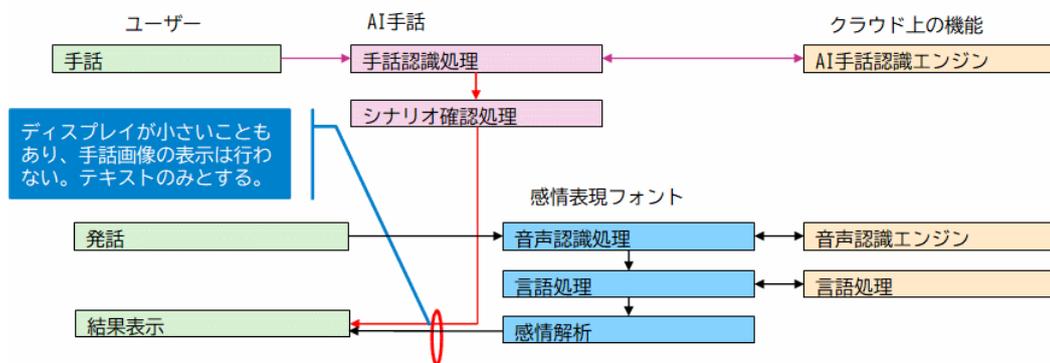
- ・ 文書データ：24種類、2179動画
- ・ 単語データ：52種類、3007動画

2. システム開発

今回のシステム開発では、音声変換(健聴者の発話した音声をテキスト変換する)は大日本印刷株式会社の「感情表現フォントシステム」(音声認識エンジンはアイシン)を使用した。手話翻訳(ろう者の手話を認識しテキスト変換する)は、後に述べるクラウド上の手話認識エンジンを使用して翻訳し、結果をクライアントに戻している。

上記の通り音声変換と手話翻訳を別々のシステムで行うが、2つの出力を合わせて1つのディスプレイに表示している。

表示はジャパンディスプレイ社製の透明ディスプレイを使用する。これは、12.3型(291.6mm×109.35mm)と小型で、透過型のため窓口に置いても圧迫感がないという特徴がある。

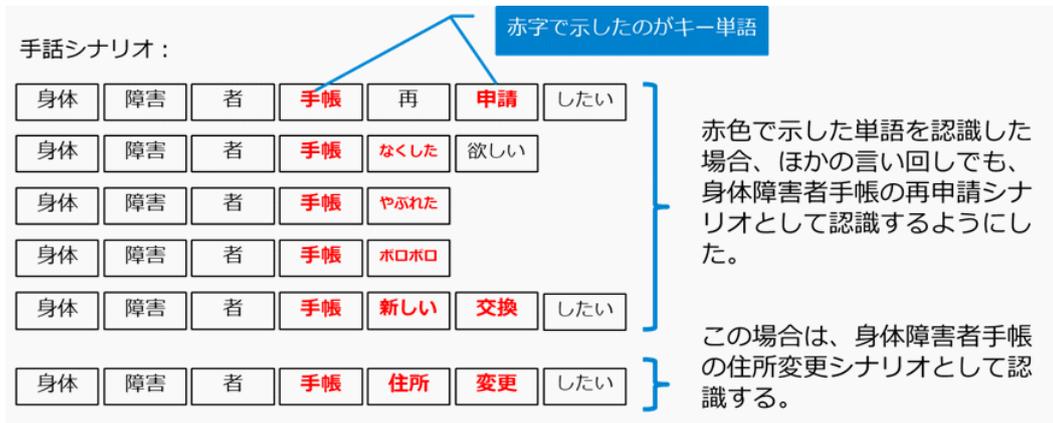


2.1 認識エンジンについて

窓口でろう者が行う手話の言い方・言い回しや表現方法などにより正しく手話を認識

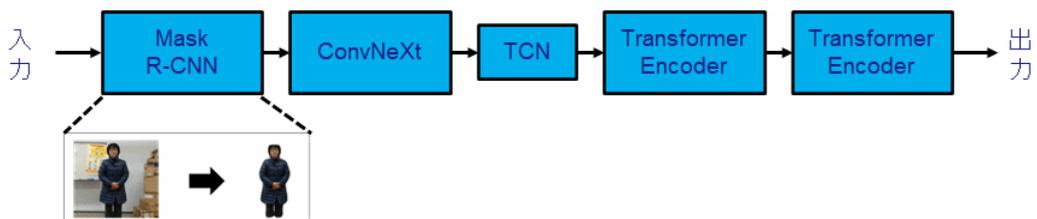
できない可能性がある。

このため、手話認識において、手話を文章ではなく単語単位で認識し、各シナリオで設定したキー単語を認識した場合はそのシナリオを発話したと認識するようなアルゴリズムとした。



以下のような階層構造で単語の認識を行っている。

Mask R-CNN	背景をピクセル単位で除去し、各フレームから人間を切り出す。 →余分な背景情報を取り除くことで、背景に惑わされない予測を行う。
ConvNeXt	各フレーム画像の特徴量を抽出する。
TCN (Temporal Convolutional Network)	系列データ向けのCNN。
Transformer Encoder	時間情報を処理し、各単語が動画内に出現したかどうかを0から1のスコアで出力する。

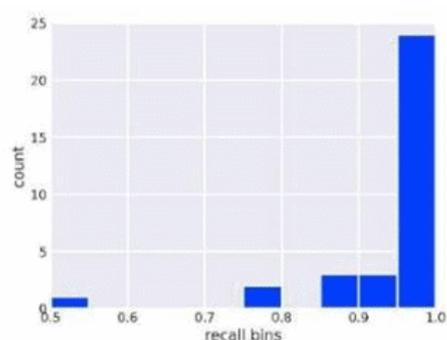
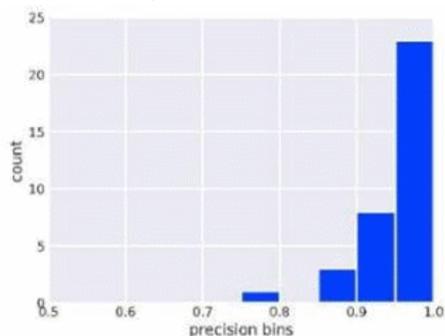


上記のアルゴリズムで認識したときの、単語ごとのPrecision、Recallの分布を以下に示す。ほとんどの単語は0.95~1.0の高い値で分類可能だが、一部の単語は性能が低くなっている。低くなっているのは、似た単語を誤って予測していることによる。

単語を誤って予測してしまう問題への対応として、Transformer Encoder処理を多段化しても改善することができなかった。

最初にMask R-CNN による人物抽出ではなく、対象の骨格抽出が有効であることが昨年の研究からわかっているが、処理時間が長くなる問題があるため、今年度のシステムで

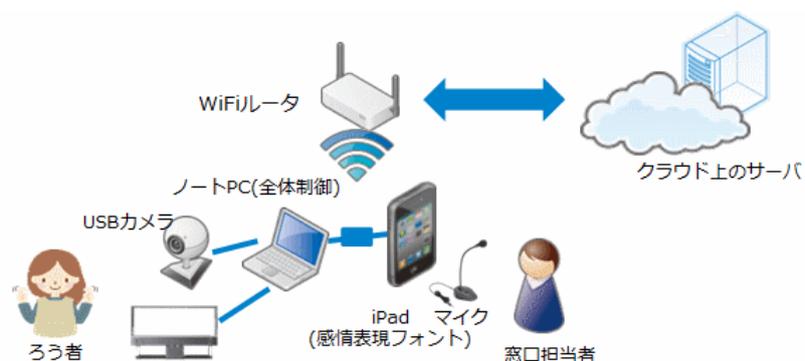
は採用していない。



3. 実行環境構築

Microsoft Azureにてサーバを構築した。深層学習アルゴリズムではGPUが必須となるため、GPUを搭載している仮想サーバを構築し、その上に手話認識エンジンを構築している。クライアントは手話動画をサーバに転送しサーバにて手話認識を行う仕組みとしている。つまり、クライアントとサーバの処理を分離し、クライアント側の処理を軽量化している。

窓口にはサーバを設置するスペースがないため、以上の仕組みにより小型PCでも使用できるようなシステム構成とした。



4. 実証実験

手話翻訳システムが実用レベルとなっているかを確認するため、札幌市の協力のもとに実証実験を行った。

ただし、実際の業務との兼ね合いで、区役所の窓口を長時間占有できず、札幌聴覚障害者協会に場所を借りて行う実験と、窓口で行う実験の2パターンで実施している。のべ、29名のろう者に協力いただき、予め設定した評価ポイントを確認している。評価ポイントでは、AIによる手話の認識率や、窓口での有効性などを明確化している。設定した評価ポイントは以下のとおりである。

観点	明らかにしたいこと	評価ポイント	手法
ニーズ充足度	窓口担当者側のニーズ充足度	窓口担当者の満足度が向上したか	窓口担当者インタビュー/アンケート
	ろう者側のニーズ充足度	ろう者の満足度が向上したか	ろう者アンケート
	共通のニーズ充足度	手話が正確に翻訳されているか（50%以上の認識率）	ログ情報からの測定/アンケート
		誤った手続き内容になっていないか	利用者アンケート
サービスの魅力度	他のコミュニケーション方法より優れているか	筆談での対応の約2倍以内でコミュニケーションできるか	シナリオ到達までの時間比較
サービスの発展性	他の業種での発展性	他の窓口で同様サービスがあった場合に利用したいか	ろう者アンケート
運用方法	サービス提供者の運用負荷	サービス提供側の運用作業は想定範囲内か	保守担当者インタビュー/アンケート
ユーザビリティ	操作方法は分かりやすいか	直感的に操作できるか	ろう者/窓口担当者アンケート

評価結果は以下の通りとなった。

評価ポイント	結果
1 窓口担当者の満足度が向上したか	 使用感について職員の100%が良いと回答しており、満足度は高い。これまではろうあ相談員がいなければ対応できなかったが、限定的でも対応可能であること、筆談等で行っていた対応が自動で文字に変換されることについて、高評価となった。
2 ろう者の満足度が向上したか	 使用感についてろう者の78%が良いと回答しており、満足度は高い。ただし、60%が翻訳できていないとの回答や、54%が筆談の方がよいと回答しており、手話認識率について更なる改善が必要である。
3 手話が正確に翻訳されているか（50%以上の認識率）	 手話認識率は69.4%であり、50%以上の認識率を達成している。全体の認識率に比べてアンケートの認識率が低い状況であり、肯定/否定を示す基本的な単語の認識率が低いためである。
4 誤った手続き内容になっていないか	 実証実験では、誤変換から誤った手続きとなる事象は発生しなかった。
5 筆談での対応の約2倍以内でコミュニケーションできるか	 筆談の所要時間はシステムを使用した場合の2.8倍程度であり、システムを使った方が迅速なコミュニケーションが取れることが判る。窓口での手続きでは職員による説明が多く、これらを筆談で行うことは担当者の負担になる。
6 他の窓口で同様サービスがあった場合に利用したいか	 ろう者の88%が利用したいと回答しており、高評価である。利用したくないと回答した12%は、誤変換を理由としており、認識率向上が課題となる。
7 サービス提供側の運用作業は想定範囲内か	 手話認識率は、動画を撮影するカメラの位置に影響を受けやすい傾向があり、サービス化に向けては課題となる。AIが認識できる手話語彙がまだ少ないため、ろう者が異なる手話表現をした場合に認識できないなど、語彙に関する課題がある。
8 直感的に操作できるか	 手話動作の開始と終了時にスイッチを押す方式としており、この部分がろう者には不評であった。タイミングの抽出方法などが今後の課題となる。

以上