別紙

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	**高速情報通信と人工知能技術を組み合わせることで、質の高い双方向的な学習をいつでもどこでも誰にでも提供することを目的とする。** 上記目的のために、手話に特化した深層学習のモデルを研究開発し、段階的に扱う手話単語数を増やすことで深層学習モデルを成長させる。さらに、単に手話を認識するだけでなく、学習者の行った動作を人工知能が認識しないときには、深層学習の内部構造を解析することで、学習者の動作のどの部分が誤っているのかを明らかにする技術を開発する。この解析結果を学習者にフィードバックすることによって、学習の改善へとつなげる。

【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. 手話動作動画データ
   全国手話検定4級相当の手話をAIに学習させるために、手話データを収集する。
2. 動作修正教示法の開発
   深層学習で誤認識の原因となった差異を、学習者に判りやすくフィードバックする研究を行う。
3. システムの軽量化とクラウドシステムの構築
   能力の低い端末でも学習できるように、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を構築する。
4. UIリニューアル
   学習システムが判りやすく使いやすいシステムとなるようなUIを開発する。

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. 手話動作動画データ

札幌聴覚障害者協会に協力していただき、全国手話検定4級相当の手話単語(500～600単語)の手話動画データを5回に分けて撮影した。4級の出題範囲として公開されている501単語を撮影対象とした。

同じ言葉でもいくつかの言い回し(手話)があるが、手話動画を深層学習するためには、同じ言い回しに統一する必要がある。言い回し全国手話検定が扱う標準手話に統一し、かつスムーズに手話動画を収集するために、まずサンプル手話動画を撮影し、そのサンプル動画を組み込んだ撮影用アプリケーションを作成した。

撮影用アプリケーションを使用し、札幌聴覚者協会所属のろう者10名に協力いただき、各単語を5回ずつ繰返し撮影した。撮影した手話動画は、深層学習の学習データとして使用した。

2.　動作修正教示法の開発

　全国手話検定試験4級で必要とされる手話動作（501単語）を取得した動画から、自動で認識するシステム開発を昨年度より継続して研究している。

・昨年度の手話動作認識モデルの改良：

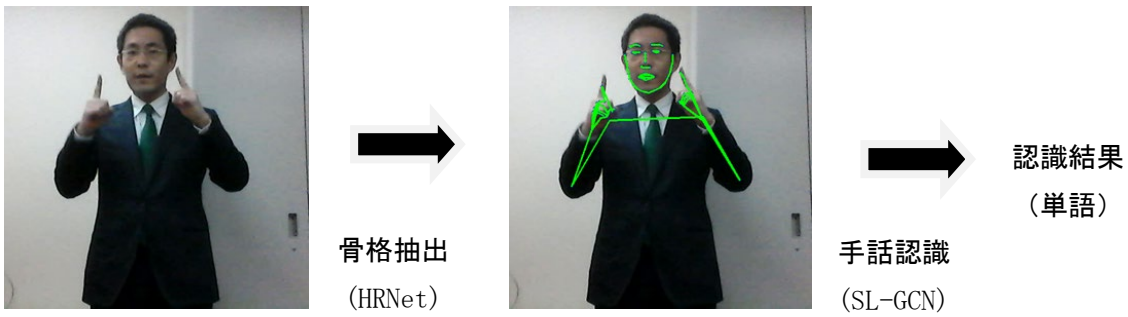
　昨年度の認識モデルでは画像そのものを使用したが、訓練データ数が限られたため、人の外見や背景といった手話とは関係のない特徴への依存があった。そのため、手話動作認識にも骨格抽出を使用し、骨格の位置変化から手話動作を認識している。成果としては、外見・背景に惑わされることなく、手話動作を高い精度で認識できた。さらに、骨格の上半身のみを抽出するため、姿勢に対する頑健性が向上した。

・今年度収集した検定４級レベルのデータ導入：

　昨年度の認識モデルは369動作の17802データを学習した。今年度は、815動作の38551データを使用して学習している。動作の種類が増えたため、より困難な学習状況であるにも係わらず、大きく精度を落とすことなく、高い学習精度を維持することができた。認識率については後述する。

・手話認識モデル：

　前述の通り、骨格を抽出した後に手話単語の認識を行う。骨格座標は27か所を選別して使用している。現状では、日本手話で重要とされる顔の座標を使用しておらず、改善が必要である。



[Sun et al., 2019]　　[Jiang et al., 2021]

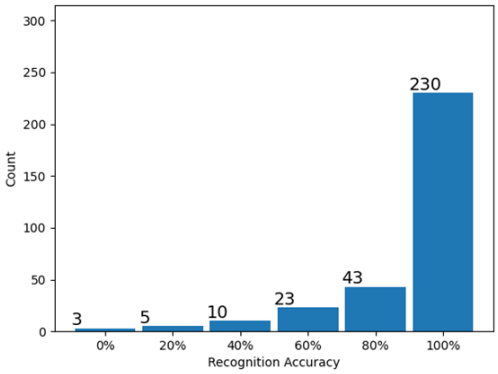
・データの内訳とモデルの説明：

	対象単語数	取得した手話動画数
全国手話検討5級相当	314	13,363
全国手話検討4級相当	501	25,188

・5級のみのモデル：

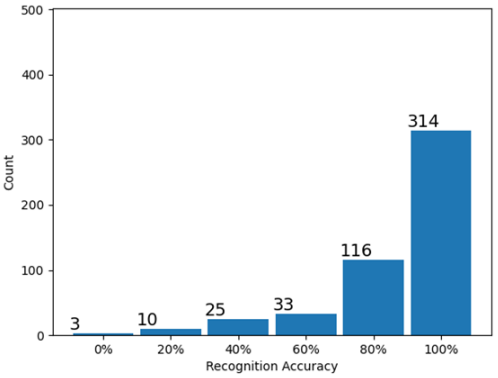
単語ごとのテスト正答率の分布を右に示す。図は230個の単語について、テスト正答率が100%であったことを示す。一部の単語で正答率が低い結果となっており、動作が近い単語の分離ができていない。

深層学習後のテスト正答率は、全体で91.1%となった。昨年度のモデル、昨年のモデル（(2+1)D ResNet）では最大テスト正答率85.8%であり、今年度のモデルでは、外見や姿勢に対する頑健性も獲得できている。



・4級のみのモデル：

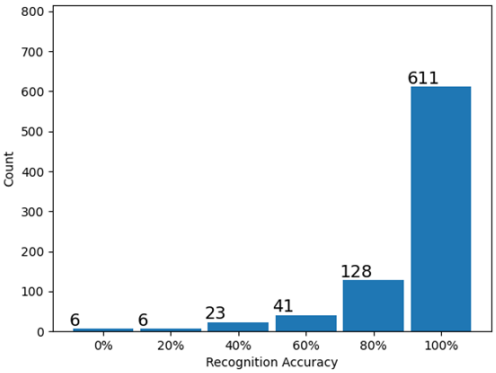
5級のみのモデルと同じ傾向である。深層学習後のテスト正答率は87.5%となっている。



・5級+4級のモデル：

5級のみ4級のみのモデルと同じく一部単語で正答率が低いが、5級と4級を個別に学習するより良い結果となっている。この正答率の向上については、偶然であるのか、もしくはデータ数が増えることによる好影響であるかは未検証である。

深層学習後のテスト正答率は91.8%であり、5級と4級の両方のデータを使用しても高い精度を維持している。



・間違い指摘モデル：

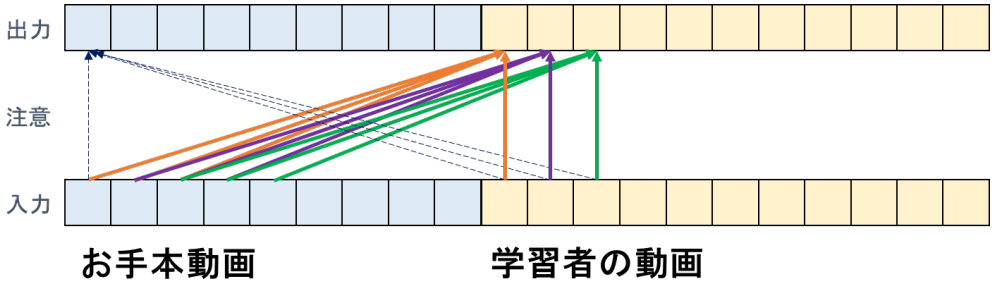
学習者が行う単語のお手本動作と学習者の動作を入力し、同じ動作を意図しているかどうかを判定する。そして、異なる場合にはその異なる部分動作を抽出する。Spatial Temporal Transformer (ST-TR)を使用してモデルを開発した。

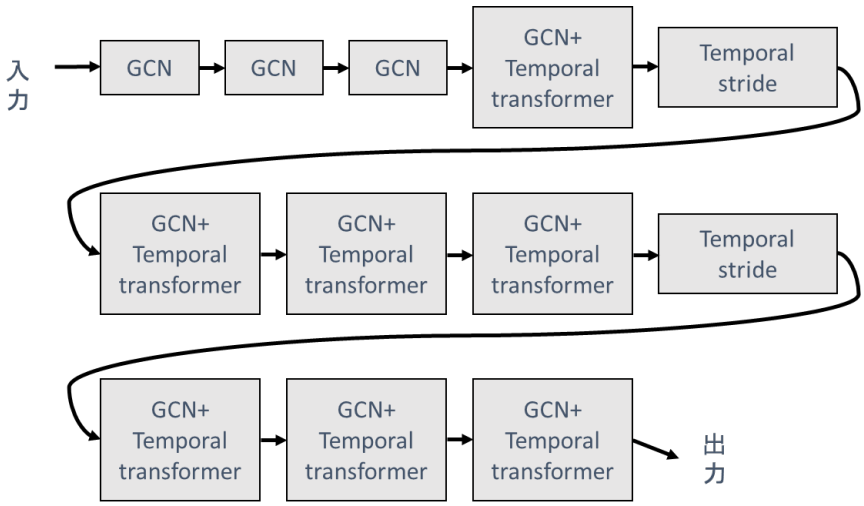
時間方向のアテンションに時間幅の制限を設定する。具体的には、時間的に幅を持ったウィンドウ内で動作を見比べ、同じ動作かどうかを認識する。

- 一様分布から最も遠い分布の注意（同じ色の矢印に対応）に対し、入力が同じ手話であれば距離を近づけ、違う手話であれば距離を遠ざける。

- 距離は平均二乗誤差を使用する。
- アテンションの幅は[31, 15, 7](単位：フレーム）に設定する。
  （時間方向畳み込みで切り替え）

動作の違いについて識別することには成功しているが、学習システムへの組み込みまではできていない。どのように学習者に提示するかも含め、継続して検証を行う。





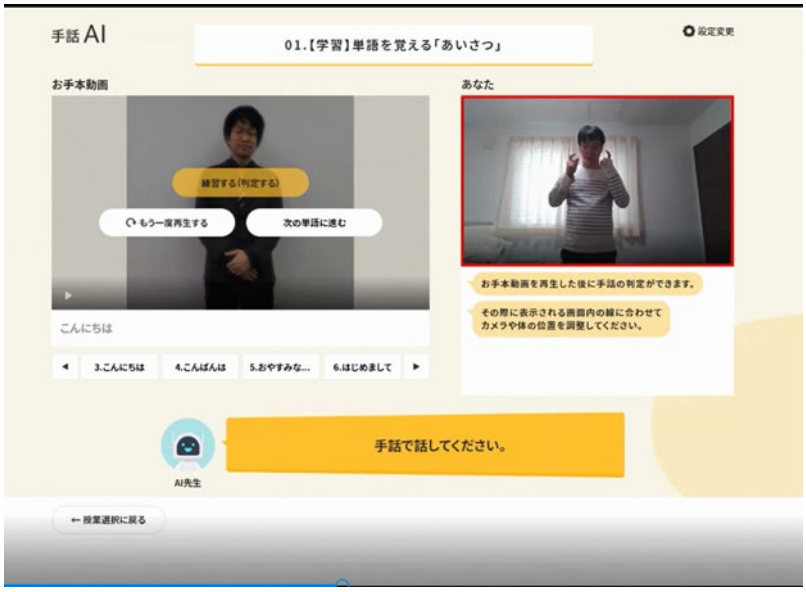
3. システムの軽量化とクラウドシステムの構築

Microsoft Azureの東京リージョンでIaaSを構築した。深層学習アルゴリズムが使用するGPUを搭載しているNCv3をベースにシステムを構築している。

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することを目的としてクラウド化を行っている。一般的なスペックのPCで、高負荷にならずに学習が行えることを目指したが、前述の間違い指摘モデルの複雑さもあり、応答に10秒以上かかるケースがあり、改善が必要な状態である。

4. UIリニューアル

使いやすく、親しみやすいシステムを目指してデザイン設計を行い、学習システムへ実装を行った。さらに、社内モニターなどによるUIの評価を実施した。



※別添資料

（様式２－３別紙 添付）（R3年度）【AI手話】状況説明資料.pptx

以上