

補助事業成果報告書

補助事業の名称	SureTalkの高度化によるきこえない人へのアクセシビリティ向上のための研究開発
補助事業の概要	きこえない人へのアクセシビリティ向上のため、AI を活用して SureTalk を高度化するための下記研究開発を行う。 1. 手話認識精度向上のための読唇技術、正解ラベルの自動抽出機能開発 2. 利用者の利便性向上のための手話認識変換技術等の開発 3. SureTalk機能向上のための音声⇒手話出力機能、手話⇒音声出力機能開発

【研究開発の実施内容と成果】

1. 手話シーンの撮影

今年度研究開発にあたって、別紙に記載されている通り、手話シーンの撮影を行った。

2. 読唇に関する技術開発

2－1. 研究開発の実施内容

(1) 手話シーンの収集

SureTalkで収録している1,500種類の手話単語シーンは殆どが口の動きがないシーンであった。そのため、そのままでは読唇技術の有用性の評価は難しいことから、最初に本研究開発用に49単語を選定し新たに収集した。表1に例を示す。2単語は同じ手指動作で異なる意味をもつ単語を示す。手話単語の選定には、筑波技術大学・加藤伸子教授に外部指導を受けた。

表 1 収集手話単語例

春	言葉
暖かい	単語
夏	世界
暑い	外国

(2) 手話シーンに対する読唇実験

手話シーンを収集するまでの期間において、手話シーンにおいて口元が隠れる場合と隠れない場合の自動判定処理の開発に取り組んだ。

手話シーンを収集後は、最初にこれまでの読唇技術を手話シーンに適用する。認識精度が低い場合は、口元が隠れる場合と隠れない場合に分けた読唇、深層学習モデル、モデルの学習方法の改善や、発話シーンの追加により認識精度向上を図った。

読唇実験を行うにあたり、前項で選定した582シーンを4つのグループに分けて、4分割交差検証法を適用して読唇実験を実施した。読唇処理を適用する前に、撮影シーンから口唇領域（lipROI）を抽出する必要がある。これは、既存手法を利用した。図2の読唇入力前にある顔画像が抽出した顔領域画像例である。手話シーンであったが、手指の動きの影響を受けず、正しく顔領域が抽出されることを確認した。

(3) 手話認識と読唇ハイブリットシステム開発

手話認識は手指動作ベース管理番号を出力しており、またこの管理番号には多義語が含まれることがある。そこで、(2)にて研究開発した読唇により、発音された言葉の母音を抽出する。次に、手話認識結果の多義語に含まれる各候補の母音を取得し、入力データの母音と手話認識結果の母音を照合し、最も近いものを選ぶことで、より正確な翻訳結果を出力する簡易システム開発を行った。アルゴリズムの概要を図1に示す。複数の機械学習モデルを組み合わせたものをハイブリッドシステムと呼び、図2に示す。

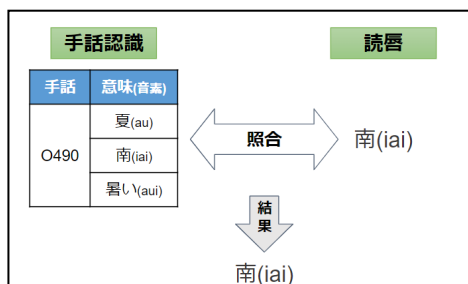


図1 アルゴリズム概要

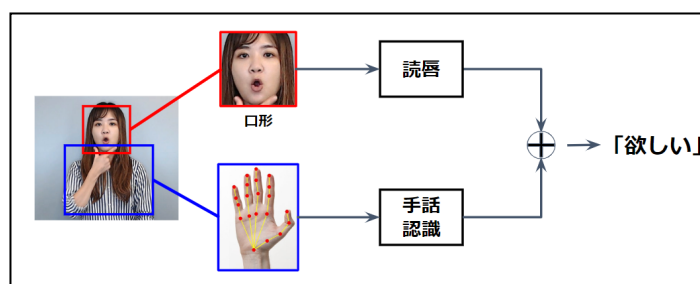
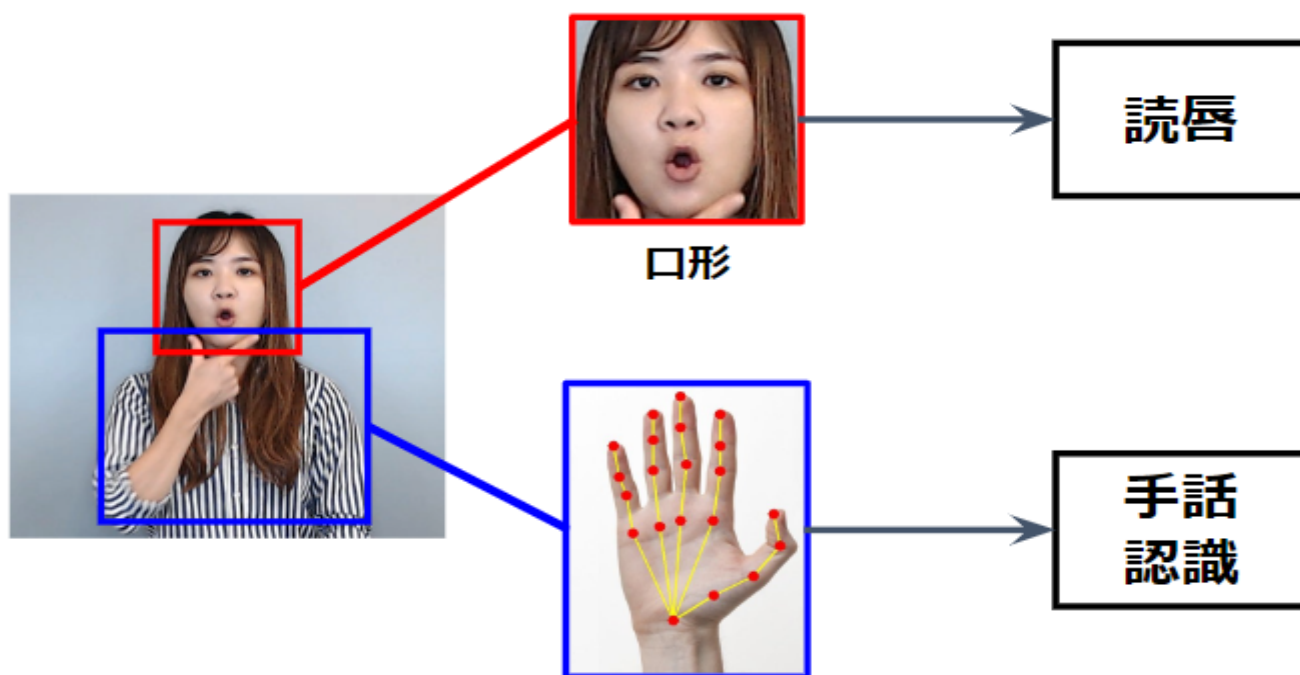


図2 ハイブリットシステム

2-2 研究開発の成果

(1) 手話シーンの収集



(3) 簡易ハイブリットシステム開発と認識実験

実験を行うにあたり、読唇実験と同様582シーンを利用した。認識実験を実施したところ、手話認識(日本語ベース認識率に換算)と簡易ハイブリットシステムの平均認識率はそれぞれ20.3%、36.3%であった。単語毎の認識精度をグラフ化した結果を図3に示す。

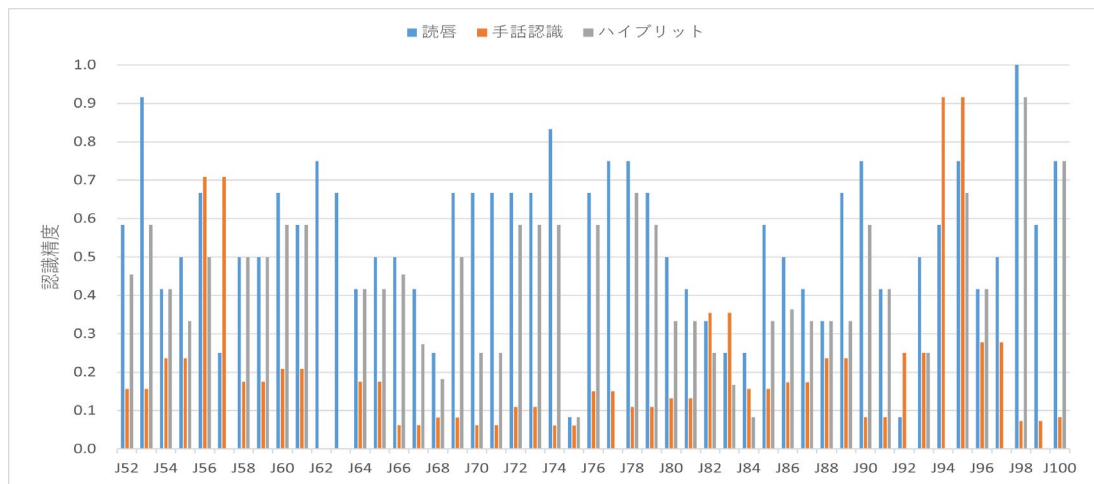


図 1 認識精度

手話認識と読唇は独立しているため、片方が誤認識する傾向があると、統合後も誤認識する傾向があるのは予想通りであった。しかし、手話認識と比較して簡易ハイブリットシステムの平均認識率が高く、手指動作が同じ手話単語に対して読唇の結果を考慮した手話表現の識別が出来たため、読唇技術および統合システムの有効性を確認した。

2-3 研究成果の考察と令和6年度計画の方針策定について

本研究開発では、手話認識と読唇を組み合わせることで、顔の動き（表情や口の形）を考慮した手話の全体的な表現を識別できることが明らかになり、手話翻訳システムの性能向上において読唇との併用する事の有効性が確認できた。

しかし一方で、複数の機械学習モデルを組み合わせたシステムは計算量およびメモリー量が増加し、リアルタイム性が低下、運用コスト高、全体最適化が困難といった様々な問題点が確認された。これらを解決するために、令和6年度の研究開発テーマとしては骨格情報から直接テキストへのマッピングを行うEnd to Endモデルを採用したいと考える。この構造によりシステムがより簡素化され、計算量とメモリー量の増加を抑えながら、手話認識の性能向上が可能と考えられる。

3. 学習モデル生成に関する技術開発

3-1 研究開発の実施内容

(1) 単語列情報を管理しないDBの検討と実装

・SureTalk データベースにおける単語列の管理方法とその問題点

SureTalkシステム開発の初期におけるデータベース（以下、DBと略す）設計では、手話で表現する際の単語、及び文章について次の2つのテーブルで管理し、現在までこの方式を踏襲している。（図4

1. 手話単語、及び手話文章を手話動作毎に管理するテーブル（単語・文章マスターテーブル）

2. 手話文章がどのような単語列（手話動作）で構成させているかを管理するテーブル

(単語列管理テーブル)

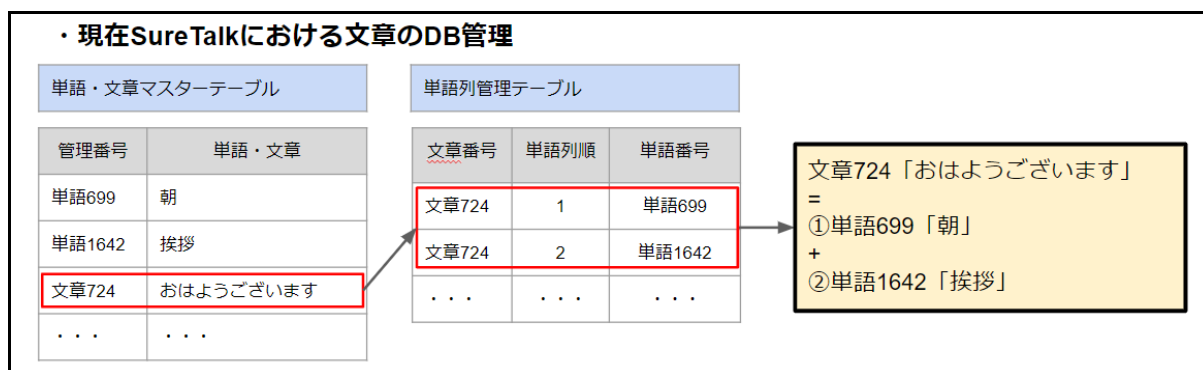


図 4

限りある人的リソースの中、一つずつ目視で動きを確認しながら各単語番号を付与しなければならない点では十分なデータ量に溜まるには膨大な時間を要してしまう。そのため、単語列の管理そのものを省く仕組み作りを検討した。

・単語列管理の廃止

本研究では、これら単語列の管理そのものを廃止する方法を検討し、手話認識に至るまでの全工程を再精査した上で実装に至った。但し、この方法は単に人的な作業を省くだけでは完結せず、AI学習、アルゴリズムを単語列管理から脱却した形に作り変える必要がある。このシステム対応方法については別項「(3) 新しいAIアルゴリズムの検討及び検証」にて詳細を説明する。また、今回の新たなDBに外部データを投入するための仕組みを令和6年度事業で実現するためのAPIなどを検討して実装する。

(2) インターネット上の手話動画収集の検討

手話認識の学習モデル性能向上のためにインターネット上の大量の手話映像を取り込む検討を行ってきた。検討に際して、取り込む手話映像には個人情報が含まれるため、データの取り込みならびに動画収集という行為が著作権法上問題ないか、またデータ活用の観点で「著作物」と「個人情報」に関するセキュリティ面における対応方法について社内確認を行った。AI開発・学習段階での著作物の利用に関しては、文化庁で定められている著作権法の30条の4における「著作物に表現された思想又は感情の享受を目的としない行為」に該当するかを社内の法務に確認した。その結果、今回のAI性能向上を目的として機械学習のみに手話映像を利用する行為は、著作権法30条の4の権利制限規定の対象となる可能性が高いため、手話動画の権利者の利益を不当に害しない範囲で対応を進めていくこととした。しかし、具体的なデータの種類や用法についての明記がないため社内セキュリティ部門に確認を行ったところ、映像データから骨格情報をAIに取り込んで開発することについては個人を識別することを目的としないため、SBのプライバシーポリシーで規定している利用範囲内であれば、ツールの開発を進めて問題ないという判断結果となった。

上記の結果に加え、インターネット上のYouTubeやFacebook, instagram, X(旧Twitter)などのSNSサービスに投稿されている動画の利用に関する各サービスの利用規約には、AI学習を目的とした手話映像の利用、ならびに骨格情報の取得、およびこれを活用することに対する制約が明記されていないことを確認した。令和6年度以降はツールの開発検討で本格的な

動画収集を実施する予定である。

3-2 研究開発の成果

・新しいAIアルゴリズムの検討及び検証

現在利用している手話認識アルゴリズムは、以下のように多くの課題を抱えている。

1. 学習データは「手話の骨格情報」「手話の単語列情報」「日本語テキスト」の3つが必要である

2. 特に「手話の単語列情報」の自動生成が技術的に非常に難しく、手動でのDB管理の工数が非常に大きく、継続的なDB収集が難しい

3. 「手話の骨格情報→手話単語列」と「手話単語列→日本語」のそれぞれの学習モデル生成の工数が大きい

4. 日本語に沿った手話表現に限った学習のため、日本手話に対してのAI学習が難しい

上記の課題を解決しない限り、公開動画から学習データとして取り込む際に莫大な工数が必要となるため、今回は手話の骨格情報からダイレクトに日本語テキストに翻訳できる学習モデルの開発を行った。ここではfairseqという自然言語処理向けのツールキットを活用して手話認識向けの学習モデルを生成した。概要を図5に示す。まずfairseqによる「手話単語列→日本語」の学習モデルを生成して

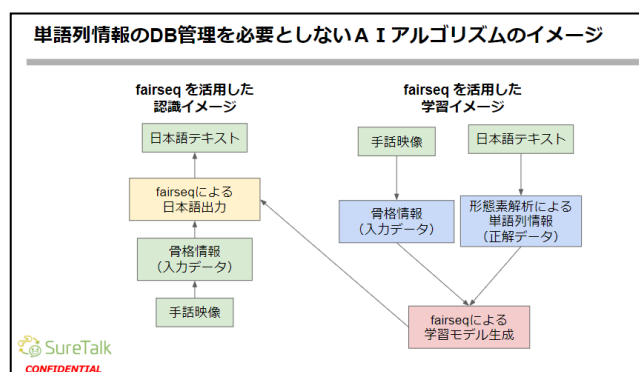


図 5

検証した結果、有効性が確認できたため、次に「手話の骨格情報→日本語」の学習モデルを生成して検証を行った。その結果を以下に示す。

【テストデータに対するBLEU評価結果（日本語ベース）】

平均BLEUスコア： 75.44

平均BLEUスコア（単語）： 73.12（商用で利用している学習モデルでは 83.86）

平均BLEUスコア（文章）： 79.14（商用で利用している学習モデルでは 53.83）

※特に単語は同じ日本語に対して違う動きを学習しているため若干落ちているが、実際に利用するには十分な結果だと考えている。

3-3 研究成果の考察と令和6年度計画の方針策定について

このAIアルゴリズムを実現したことによって、従来の手話単語列の管理が不要になり、手話データの収集スピードの加速化が可能になった。年度内にはSureTalkの会話機能に学習モデルを反映させ、手話ユーザーに認識体験していただく予定である。また、単語列を廃止したDBに手話データを蓄積するための仕様が現在では限定的なところでしか使えないため、令和6年度事業では公開動画なども含めて広範囲で使えるようにし、手話認識の学習モデルの性能向上に繋げていく予定である。