

平成30年度
戦略的基盤技術高度化・連携支援事業
戦略的基盤技術高度化支援事業

「手話の自動翻訳を実現させる高精度な動作検出と
動作のパターンマッチングの技術開発」

研究開発成果等報告書

平成31年3月

担当局 中部経済産業局
補助事業者 株式会社ユニオンソフトウェアマネジメント

目次

第1章 研究開発の概要.....	3
1-1. 研究開発の背景・研究目的および目標.....	3
1-2. 研究体制.....	10
1-3. 成果概要.....	11
1-4. 当該研究開発の連絡窓口.....	13
第2章 本論.....	14
2-1. 3次元座標の取得/補正・補完/プロット.....	14
2-2. 手話翻訳デバイスの試作品製作.....	15
2-3. 手話動作認識アルゴリズムの作成.....	18
2-4. 翻訳精度評価.....	21
最終章 全体総括.....	24
3-1. 補助事業の成果.....	24
3-2. 補助事業終了後の技術的な課題.....	24
3-3. 手話翻訳デバイス「Tesign」の事業化見通し.....	24
3-4. 事業化展開.....	25
引用文献.....	26

第1章 研究開発の概要

1-1. 研究開発の背景・研究目的および目標

1-1-1. 研究開発の背景

株式会社ユニオンソフトウェアマネージメント（以下、当社）では、障害者雇用施設である非営利特定法人つくしとの協業において、手話の翻訳装置があることで、障害者の社会進出がより活発になると知り、赤外線センサを用いた手話翻訳の研究開発を進めている。

平成18年の身体障害児・者実態調査結果によると、日本の聴覚障害者数は約29万2000人にも上る [1]。また、日本補聴器工業会によれば、難聴者数は、平成29年時点で約1430万人と世界でも多く、聴覚障害者数、難聴者数合わせて約1460万人もの人が、重度は様々であるが耳に不自由を抱えて生活していることになる [2]。近年、障害者雇用に関する労働基準法の改正が行われており、大企業だけでなく中小企業でも障害者雇用の促進についての対策が取られ、障害者雇用率は増加傾向にある。しかし、その一方で障害者と健常者間のコミュニケーションの取り方について、多くの課題が残されている。当社が着目している聴覚障害者ならびに難聴者と健聴者のコミュニケーションの取り方について、以下の対応例が挙げられる。

[双方へのコミュニケーション援助]

- ① 手話通訳士の雇用
- ② 手話が分かる労働者の雇用
- ③ 筆談
- ④ メールやチャットツール

[一方へのコミュニケーション援助]

- ① 聴覚障害者ならびに難聴者による読唇
- ② 健聴者が利用する音声認識ツールを利用した文字起こし

これらの対応例の内、[双方へのコミュニケーション援助]①および②は、金銭的成本がかかること、手話通訳士、手話が分かる労働者の人数に制限があることから、いつでも、どんな企業でも、行える対策とは言えない。[双方へのコミュニケーション援助]③、④は[双方へのコミュニケーション援助]①、②と比較して、金銭的成本がかからないメリットはあるが、「文字を書く」手間や「文章を打つ」手間があると考える。また、[一方へのコミュニケーション援助]は、一つの対策では双方のコミュニケーションは難しく、別の対処との併用となる上に、[一方へのコミュニケーション援助]①については、聴覚障害者、難聴者の全員が体得しているわけ

ではない。

関連研究として、既存の手話翻訳の取組事例を図 1 に示す。接触式では、日立製作所中央研究所の開発したシステムがある。手袋型のセンサを装着して手指の動きや形状を取得し、ビデオカメラから頭部動作を認識して手話の動作を取得することができたが、装置が大掛かりで高価であるため普及には至らなかった。非接触式では、代表的な 3 次元画像センサを用いた手話認識システムが、みずほ証券と千葉大学の共同研究により提案されている [3]。これは銀行窓口の対話を想定したシステムで、プロトタイプでは 20 語彙程度の認識が可能であるが、手首と腕の動きだけを対象としており手の形は使用していない。3 次元画像センサは、複雑な手の形状を取り出すには解像度が不足しているためである。従来技術では、価格面、精度、語彙量、速度、人の癖などへの多様性について多くの課題が残されている。



図 1 既存の手話翻訳の取組事例

当社では、すでに赤外線センサを用いた手話翻訳の独自技術を開発しており、特許申請も済ませている。(出願番号 2015-045920) しかし赤外線センサのみでは、手元の動作しか検知できないため、腕や肩の動きを伴う手話については認識できない。そこで本事業では、図 2 の通り、身体全体の動作検知に 3 次元画像センサを、手指の詳細動作検知に赤外線センサを用い、加えてカメラ画像も活用した手話の自動翻訳デバイスの開発を行った。認識部位に特化したセンサを用いることで、非接触式で実用的な価格且つ、高精度な動作認識を実現できる。また、本システムでは日常会話ができる語彙量を確保し、言語出力までの高速な処理速度、人の癖などへの多様性についても対応する。本事業の手話翻訳システムの流れを図 3 に、使用イメージを図 4 に示す。図 3、図 4 にある通り、複数のセンサやカメラを用いて動作検知を行い、取得した動作の座標データをあらかじめ登録されている手話データと照合する。照合した結果を音声もしくは文字で返し、結果を出力する。また、ソフトウェアとしての研究開発だけでなく、デバイスの製作も研究として実施し、タブレット端末を目指した試作品の製作を行った。



図 2 本事業で用いた新技術の特徴

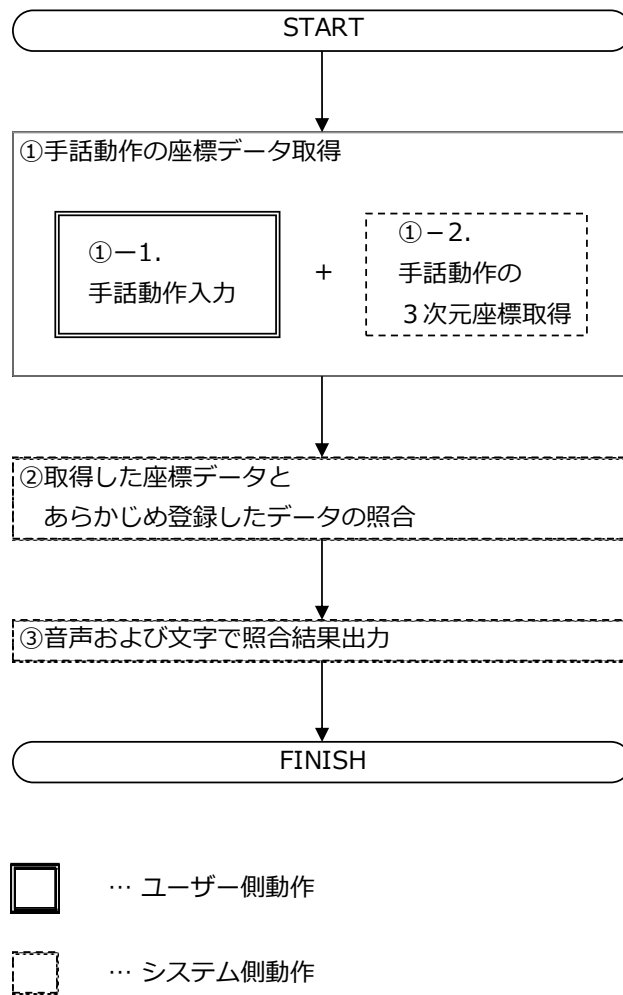


図 3 本事業における手話翻訳システムの流れ

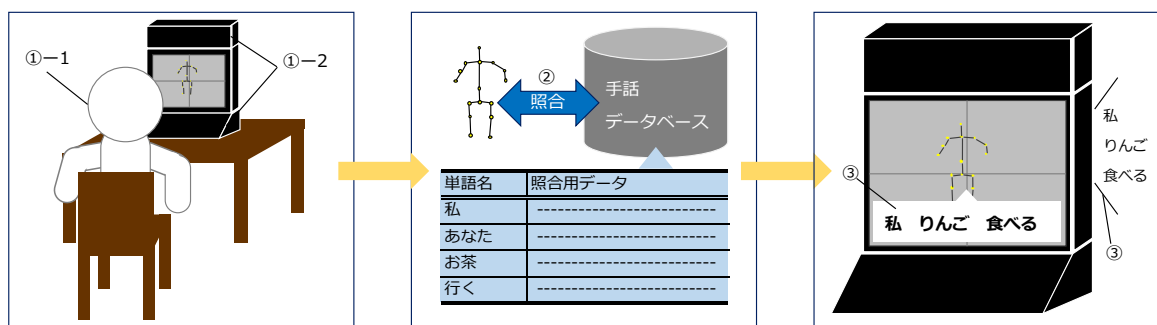


図 4 本事業における手話翻訳システムの使用イメージ

平成 28 年度の事業において、海外の動向調査を行ったところ、ポーランドにある VoicePIN 社が Kinect1 台を用いた手話翻訳のシステム開発を研究していることが分かった。VoicePIN 社では「病院の予約をとるため」の限定されたシチュエーションにおいて「一問一答形式」での動作認識を行うシステムを製作していた。センサに Kinect 1 台を用いていることから手の細部の形状を取得するにはやはり課題があり、当社が研究をしている完全な手話翻訳のシステムを実現するには、まだ課題が残っている。

また、平成 29 年度の事業において、海外の動向調査を行ったところ、アメリカにある Motion Savvy 社が Leap Motion 1 台を用いた手話翻訳のシステム開発を研究していることが分かった。Motion Savvy 社では、あらかじめジェスチャを登録させ、同様のジェスチャを行うことで翻訳を行うシステムを製作しており、フレーズと単語の組み合わせで 2 語分程度の翻訳を実現している。Motion Savvy 社が開発する「UNI」と呼ばれる手話翻訳システムは、ロチェスター空港と 60~65 のデバイス導入の契約を結んでいるが、「チケットセンター」、「セキュリティチェック」、「売店」、「搭乗ゲート」など、空港内のそれぞれの箇所に導入するため、こちらも限定されたシチュエーションでの導入であり、当社が研究をしている完全な手話翻訳のシステムを実現するには、まだ課題が残っている。

本事業では、端緒として身体全体の動作検知用に Kinect V2（図 5）を、手の細部の動作検知に Leap Motion（図 6）を用いて研究開発を行った。

1-1-2. 研究目的および目標

本事業では、既存の高性能赤外線センサと画像カメラを活用して手話の動作検知を高精度に実現する研究開発を行った。従来の手話翻訳の研究において、単一のビデオカメラを用いた画像ベースで、100 語の認識では 8 割、300 語では 7 割の正解率となっている。また、赤外線センサ Kinect®を用いた研究では 100 語の認識では 7 割の正解率となっている。本事業は、赤外線センサとカメラを組み合わせた動作検知により、1000 語で 8 割の正解率に向けて取り組んだ。また、従来の研究では、語彙量が増えると正解率が低下したり、照合に時間がかかったりし、



図 5 Kinect V2



図 6 Leap Motion

リアルタイムに出力ができなかったが、効率的なマッチングシステムを開発することで、正解率を8割に維持しつつ、リアルタイム（1秒以内）の出力の実現に向けて取り組んだ。手話の自動翻訳機器の実現には以下の技術が必要である。それぞれの目標について記載する。

[手話の自動翻訳機器の実現に必要な技術]

1. 手や指の3次元座標の、精密な位置座標の取得→精密3次元座標解析
2. 手や指の座標を、時系列での処理・判断→3次元座標の時系列処理
3. 前後の単語の関係から、効率的な単語選出→文脈依存モデル
4. 語彙増加に伴う動作と単語のパターンマッチング高速化→高速パターンマッチング
5. 多様な人や動作に対する高精度、高速な照合、多様性の許容→階層化認識モデル

1-1-2-1. 精密3次元座標解析プログラムの開発

精密3次元座標解析について、図7に示す。名古屋工業大学の事前研究により、動作の検知において、人の身体の部位は5cm、手の細部は1cmの精度であれば、動作パターンは正確に認識可能であることが分かっている。これにより、目標値として、人の身体の部位は5cm、手の細部は1cmの精度で座標位置を取得する。

また、正確な3次元座標を取得するタブレット端末の試作機を作成する。タブレット端末の完成イメージを図8に示す。人物を正確に捕捉には、赤外線センサの特性やユーザの位置に対して赤外線センサの位置が大きく関わってくる。また、日常使いしやすい製品とするため、持ち運び可能なタブレット端末が媒体として望ましい。そのため、赤外線センサを最も有効に活用できる位置であり、ユーザが使用する際の端末の位置も考慮した、正確な3次元座標が取得できる赤外線センサの位置を決定する。また、タブレットPC用のマザーボード、メモリ、SSDなどの部品を調達し、3Dプリンタを用いて、マザーボードと

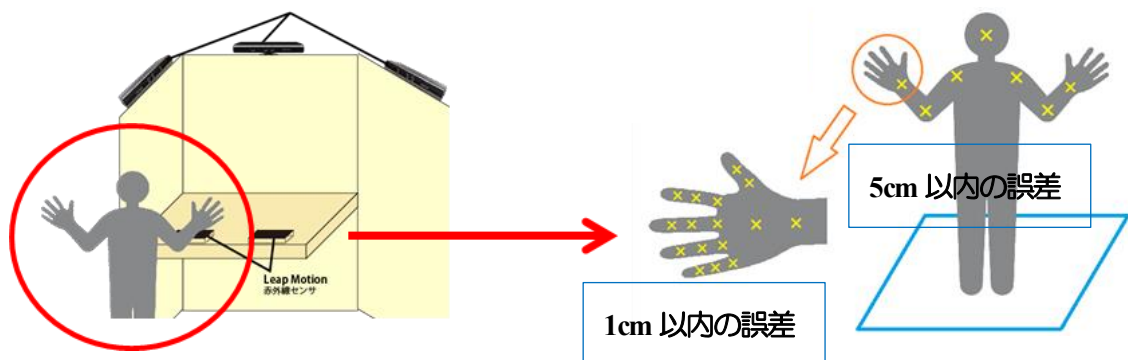


図7 精密3次元座標解析



図 8 手話翻訳デバイス（タブレット端末）の完成イメージ

センサを結線した複数種のタブレット端末を試作する。被験を通じて決定したアルゴリズム設計に最適な端末形状を決定する。

1-1-2-2. 3次元座標の時系列処理

3次元座標の時系列処理について、図 9 に示す。手話動作を一連の動きとして処理するために、1m 四方の空間をパソコンモニタ上の仮想空間に縮小再現し、取得した身体および手の座標位置を 1/10 秒単位で画面上にプロットする。アドバイザーである障害者雇用施設の非営利特定法人つくしからは、手話翻訳を実現するためには 1/10 秒単位で動作を再現できれば十分であるとの助言を頂いている。なぜならば、日常会話において手話は 1 秒間に 10 単語の表現はしないためである。よって、目標値としては、再現時間単位を 100m 秒（1/10 秒）とする。

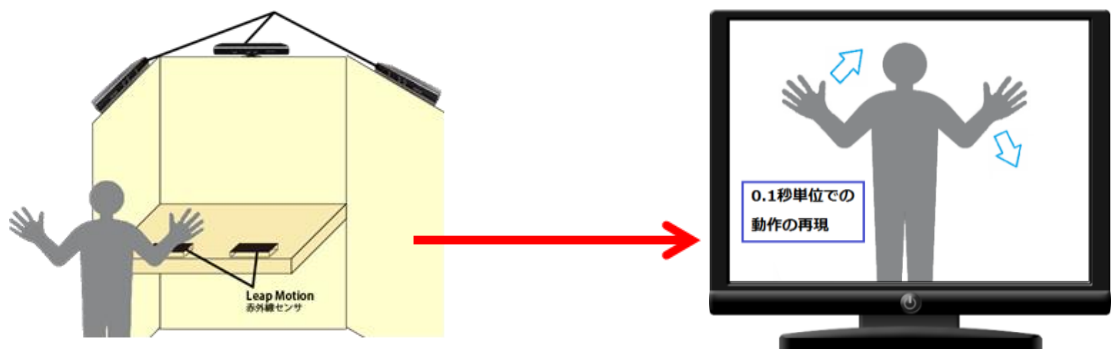


図 9 3次元座標の時系列処理

1-1-2-3. 文脈依存モデルプログラムの開発

日常会話が可能な語彙量を確保するため、収録単語の拡大（大語彙化）を行う。大語彙化に伴う認識率低下を防ぐ方法として、文脈依存モデルプログラムを開発する。文脈依存モデルプログラムを使用することで、現在 100 語に対して 70%の認識率であるが、1000 語に対して 80%の認識率に向上させる。よって目標値としては、被験者 10 名×動作パターン 1000 種類で 80%以上の認識率を実現する。

1-1-2-4. 高速パターンマッチングプログラムの開発

会話に支障のない処理時間を実現するために、高速パターンマッチングプログラムを開発する。高速パターンマッチングプログラムを使用することで、動作後 1 秒以内に応答することを目標とする。名古屋工業大学の事前研究では、一連の手話動作が終了してから 3~4 秒かかっていることから、目標値として、被験者 10 名×動作パターン 1000 種類で 80%以上の認識率を、手話動作が終了してから 1 秒以内に応答させる。

1-1-2-5. 階層化認識モデルの開発

様々な人物の違いに対応するため、階層化認識モデルデータベースを構築する。階層化認識モデルデータベースを用いた認識により、100 名で 80%の認識率を目指す。目標値としては、被験者 100 名×動作パターン 1000 種類で 80%以上の認識率を実現する。

1-2. 研究体制

研究体制について、図 10 に示す。株式会社ユニオンソフトウェアマネジメントは名古屋工業大学と、動作検知技術およびパターンマッチング技術の共同開発を行いながら、手話翻訳デバイスの開発を行った。PL は、株式会社ユニオンソフトウェアマネジメント 代表取締役 青井基行が担当し、本事業における 5 つの技術課題の解決に向けた実施体制の構築およびそれぞれの技術課題を組み込み 1 つの統合された高度な技術への発展のための研究を主に担当した。さらに、本事業において試作した製品を販売するための事業化に向けて必要な行動を行った。SL は名古屋工業大学の酒向准教授が担当した。SL は本事業における 5 つの技術課題のうち「文脈依存モデル」「高速パターンマッチング」「階層化認識モデル」を確立するために必要な AI（人工知能）要素を持つプログラム アルゴリズムについて大学の知見を活用し、技術の確立を目指した研究を実施した。PL と SL は、各自の研究状況において情報を共有し、連携を取ることで円滑な事業運営を行った。アドバイザーの一般社団法人愛知県聴覚障害者協会と非営利特定法人つくしには、よ

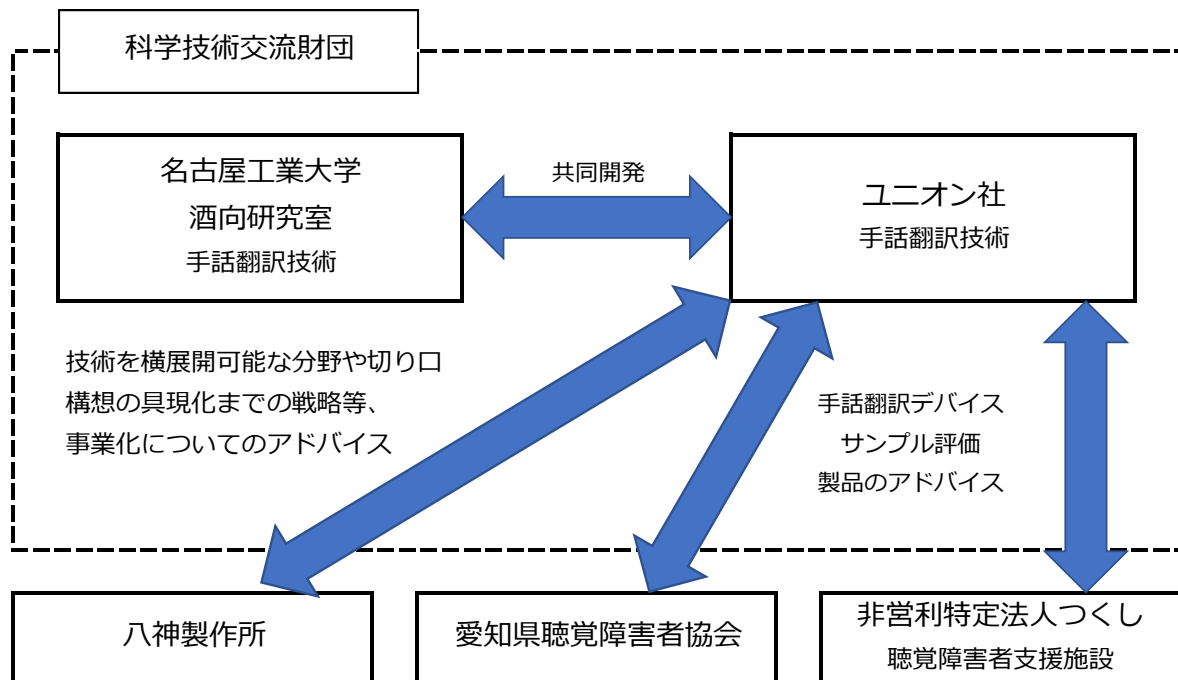


図 10 研究体制図

りユーザ目線に近い観点から手話翻訳デバイスのサンプルを評価して頂き、川下企業として製品改良のアドバイスを行って頂いた。また製品化が実現した後は、愛知県聴覚障害者協会や非営利特定法人つくしが参画する障害者向けのイベントなどにおいて製品のプロモーション活動に協力頂き、製品を世の中に認知させるための協力を頂くことになっている。また、製品化後も引き続き利用者として、製品の改良点や問題点を指摘頂き、製品の利便性向上や製品の完成度向上に協力を頂く。八神製作所には、技術の横展開が可能な分野やその切り口、それらの構想を具現化するまでの戦略等、事業化についてのアドバイスを頂く。

1-3. 成果概要

【1】 3次元座標の取得/補正・補完/プロット

【1-1】 取得した身体および手の3次元座標の画面プロット

Kinect V2、Leap Motion から取得した身体および手の3次元座標の画面プロットを行った。取得した3次元座標をリアルタイムに画面にプロットするだけでなく、取得した3次元座標データを、時間と紐づけて撮り溜めることで、時系列処理を行えるようにプログラムを構築した。3次元座標データの取得は 1/30 秒単位で行い、目標を達成できている。

【1-2】 3次元座標データの補正・欠損補完

Kinect V2、Leap Motion から取得した身体および手の3次元座標は、何らかの原因で曖昧な3

次元座標が取得された場合や、座標データが取得できなかった（欠損した）場合、前後フレームの座標データからデータの有るべき3次元座標を割り出す処理を構築した。

【2】 手話翻訳デバイスの試作品製作

【2-1】 3D プリンタ、3DCAD マシンの導入

手話翻訳デバイスの試作品を製作するため、3D プリンタ、3DCAD 専用 PC を導入した。

【2-2】 筐体用パーツの導入

手話翻訳デバイスの試作品を製作するため、筐体用パーツとして、マザーボード、CPU、メモリ、AC アダプタ、タッチパネルディスプレイ、SSD、無線 LAN 子機、OS を導入し、3D プリンタで造形した筐体と結線することで、筐体セットとして計 10 台分製作した。

【2-3】 試作品の筐体製作

3D プリンタと 3DCAD マシンを使用して、手話翻訳デバイスの試作品製作を行った。身体および手の3次元座標の取得が十分に行えるように赤外線センサの位置を決定し、マザーボードやディスプレイなどのパーツを結線することで筐体の製作を行った。Kinect V2 と Leap Motion を 1 台ずつ用いる筐体だけでなく、RealSense 単体、Xtion2 単体、RealSense と Leap Motion を 1 台ずつ用いる筐体など、各赤外線センサの特徴に合わせて複数種類の筐体製作を行った。最終的に縦 30cm×横 25cm×奥行 15cm の筐体サイズになり、小型化が課題となった。

【3】 手話動作認識アルゴリズムの作成

【3-1】 単語抽出

入力された 1 文章の手話動作に含まれる単語データを 5 要素認識にかけるため、1 文章分の手話動作データから 1 単語ずつデータを抽出するアルゴリズムを作成した。研究開始当初は、通常手話で会話する時同様、連続して単語を入力させた 1 文章データから、単語である箇所を自動で抽出する自動単語抽出アルゴリズムの作成を目指していた。しかし、手話を行う速度や大きさなどの人の癖や、単語の表し方の複雑さから、単語を表している箇所と、単語と単語の渡りである箇所の区別は難しいと分かった。そこで、ユーザが単語表現終了時に定位置に手を戻すことで単語間の境目を決定し、単語抽出を実現した。

【3-2】 手話単語の認識アルゴリズム

手話動作と単語名を紐付けるための手話ディクショナリを作成した。動作を解析し、単語に変換するため、手話動作における「手の形」、「手の位置」、「手の動き」、「掌の向き」、「指の向き」の各要素認識アルゴリズムを作成した。各アルゴリズム処理後、手話ディクショナリと照合し、単語翻訳を実現した。「手の形」、「手の動き」は、他要素と比較して認識精度が低かった。

【3-3】文脈依存モデル

認識精度向上のため、単語間のつながりを考慮した文脈依存モデルプログラムを開発した。

【3-4】高速パターンマッチング

動作入力後 1 秒以内に結果を出力する高速パターンマッチングプログラムを開発した。

【3-5】階層化認識モデル

様々な人物の違いに対応するため、階層化認識モデルを開発した。認識モデルとパラメータから階層構造を作成するプログラムと最適モデル探索用のプログラムを作成した。

【4】 翻訳精度評価

【4-1】 翻訳精度評価用データ収集

翻訳精度を評価するため、また、認識精度向上に向けた取り組みに用いるため、データ収集を行った。データ収集は複数回に渡って実施し、非営利特定法人つくしのスタッフ（健聴者、聴覚障害者共）を中心にデータ収集にご協力頂き、3次元座標データの収録とアンケート回答を収集した。収集したデータは 41 名分のデータであり、目標の 100 名には届かなかった。

【4-2】 翻訳精度検証

翻訳精度の評価を行うため、精度検証を行った。精度検証は段階に分けて実施し、まずは単語精度検証を 1 名の被験者データを用いて実施した。検証の結果、5 文章で 90%、10 文章で 89%、20 文章で 83%、40 文章で 73%、80 文章で 60%、175 文章で 47%の単語認識精度が確認された。1 名分のデータでも、1000 語に満たない単語数で 80%の認識率を達成できず、当初の目標であった、100 名の 1000 語に対して 80%の認識精度は、実現には至らなかったが、ポーランドにある VoicePIN 社の「病院の予約システム」が限られたシチュエーション内の 20 単語程度であまり精度が出ていないことと比較すれば、実用に近づいていると言える。

1-4. 当該研究開発の連絡窓口

株式会社ユニオンソフトウェアマネジメント

技術グループ 尾山隆二

〒460-0003

愛知県名古屋市中区錦一丁目 13 番 26 号

名古屋伏見スクエアビル 8 階

TEL : (052) 219-5020 / FAX : (052) 219-5021

E-mail info@union-soft.co.jp

第2章 本論

2-1. 3次元座標の取得/補正・補完/プロット

(1). はじめに

身体および手の3次元座標を取得し、画面にプロットするアルゴリズムについて記載する。本事業では、Kinect V2、Leap Motion から取得した身体および手の3次元座標の画面プロットを行った。

(2). 方法および結果/考察

(2) - 1. 取得した身体および手の3次元座標の画面プロット

Kinect V2 および Leap Motion を用いて身体および手の3次元座標を取得し、画面にプロットした結果を図 11 に示す。図 11 のように身体の3次元座標、手の3次元座標がそれぞれ画面に縮小されてプロットされた。画面にプロットした座標による各関節間の距離と実寸値の関節間の距離を計測したところ、身体は 5cm、手の細部は 1cm の精度で座標位置の取得を実現した。画面へのプロットは 30FPS（1/30 秒単位の座標取得）を実現した。この時、取得した3次元座標をリアルタイムに画面へプロットするだけでなく、取得した3次元座標データを時間と紐づけて撮り溜めることで、時系列処理ができると確認された。

また、本事業では、Kinect V2 と Leap Motion を用いた手話翻訳デバイスの研究開発を進めてきたが、当社では Kinect V2 だけでなく、Intel 社の RealSense、ASUS 社の Xtion2 などの赤外線センサへの対応も視野に入れ、赤外線センサからデータ取得する部分のみカスタ



図 11 身体および手の3次元座標の画面プロット

マイズを行うことで、Kinect 以外の赤外線センサを用いた場合も動作検知を行える、センサフリーなアルゴリズム設計を行った。

(2) - 2. 3次元座標データの補正・欠損補完

Kinect V2 および Leap Motion から取得した身体や手の3次元座標が何らかの原因で欠損したり、曖昧なデータを出力したりすることがあると分かった。これらに対応するために、当社では、3次元座標データの補正および欠損補完アルゴリズムを作成した。このアルゴリズムは、取得したデータの一部が曖昧である、あるいは、欠損しているデータであった場合、前後フレームの情報から有るべき座標値を算出し、データの補正、補完を行う。

3次元座標データの補正・欠損補完の翻訳精度への影響度を評価した。評価方法は、同じ被験者の同じ文章データを用いて、補完を実施しないデータと補完を実施したデータの「手の形」、「手の位置」、「手の動き」、「掌の向き」、「指の向き」の5要素の認識結果を比較することで精度比較を行った。評価の結果、「手の位置」、「掌の向き」、「指の向き」はほとんど差が見られなかったが、「手の形」は3%程度、「手の動き」は0.5%程度、補完を行ったデータの方が補完を行わなかったデータに比べて、認識率が高かった。

また、研究を進める中で、手の座標データの欠損の定義には、以下の複数パターンがあることが分かった。

[手の座標データにおける欠損の定義]

- ① ユーザが手話動作を行っていないため、センサから意図的に手を外している
- ② ユーザが手話動作を行っている最中に、センサの認識範囲外に出てしまった
- ③ ユーザが手話動作を行っている最中に、手は認識範囲内にあるが、センサのトラッキングが失敗した

これらのパターンについて、手の認識用センサから取得されるデータのみで判断することは難しい。そのため、欠損の定義と補完対象について、引き続き検討し、必要に応じた補正および補完を行う必要がある。

2-2. 手話翻訳デバイスの試作品製作

(1). はじめに

手話翻訳デバイスであるタブレット端末の試作品製作について記載する。

(2). 方法および結果/考察

まず、タブレット端末の試作品の製作を行った。CAD ソフトを用いて筐体の設計を行い、3D プリンタを用いて筐体の造形した筐体を用意したマザーボードや赤外線センサ、ディスプレイなどを結線し、組み上げて製作した。製作した筐体を図 12 に示す。筐体サイズは、縦 25cm×横 25cm×奥行 15cm となり、当社が目指している「タブレット端末」とするには、小型化の課題が残っている。筐体が大型である原因には以下が挙げられる。

[筐体が大型である原因]

- ① Kinect V2 のセンサ自体が大型（縦 6.5cm×横 25cm×奥行 8cm）であること
- ② Kinect V2 に付属のアダプタおよびケーブル類が多い（延 8m25cm 程度）こと
- ③ Kinect V2 と Leap Motion を 50cm 以上離れた配置となる形状必要であること

また、研究開発中であった 2017 年に Kinect V2 の生産中止が発表され、在庫のみの販売となることが発表された。筐体小型化の課題および Kinect V2 生産中止を踏まえて、当社では Kinect V2 を用いた研究開発と並行して、代替センサの検討、代替センサを用いた筐体の試作品製作を行った。代替センサとして、Intel 社の RealSense、ASUS 社の Xtion2 を検討し、筐体の製作を行ったが、最終的に RealSense D435 を採用することに決定した。RealSense D435 と Leap Motion を 1 台ずつ用いる筐体を図 13 に示す。筐体サイズは、縦 30cm×横 25cm×奥行



図 12 Kinect V2 を用いた筐体

15cm となり、小型化の課題は残されたままだった。RealSense D435 にセンサを変えたにもかかわらず、筐体小型化の課題が残っている原因には以下が挙げられる。

[筐体小型化の課題が残っている原因]

- ① ユーザが製品使用時に、カメラから頭部が見切れてしまうことを防ぐため、筐体の土台部分に高さを追加する必要があったこと
- ② 筐体製作の際、3D プリンタで造形するため、造形物の反り返りおよび強度を考慮すると分厚い形状にする必要があったこと
- ③ 今回用いたマザーボードが Kinect V2 の動作に必要なコンピュータのスペックであるため、CPU の Intel Core i7 の動作に大型のファンが付随していること

以上を踏まえて、当社では引き続き筐体の小型化の課題に取り組む。製品リリース前に行う無償貸与でのユーザからの声やデータを集めた上で、製品のハードウェア製作を外部に業務委託し、筐体およびマザーボードの設計を再度行うことで、サイズ、強度、放熱効率などの対応が可能となる。また、③の通り、今回用いたマザーボードに搭載されている CPU は RealSense D435 には過度なスペックのため、適した CPU を用いることで、大型なファンが必要なくなり、より小型なデバイスの製作が可能と考える。



図 13 RealSense D435 を用いた筐体

2-3. 手話動作認識アルゴリズムの作成

(1). はじめに

手話動作の認識を行うためのアルゴリズムについて記載する。本項では、単語抽出、手話単語の認識アルゴリズム、文脈依存モデル、高速パターンマッチング、階層化認識モデルについて記載する。

(2). 方法および結果/考察

(2)-1. 単語抽出

入力された1文章の手話動作に含まれる単語データを5要素認識にかけるため、1文章分の手話動作データから1単語ずつデータを抽出するアルゴリズムを作成した。研究開始当初はユーザが自然に近い入力を実現させるために、通常手話で会話する時同様、連続して単語を入力させた1文章データから、単語である箇所を自動で抽出する自動単語抽出アルゴリズムの作成を目指していた。例として、「あなた、季節、好き、何（日本語訳: あなたの好きな季節は何ですか）」の文章データにおける、左右それぞれの手の移動速度に着目した単語切り出し箇所のグラフを図14に示す。手話を行う速度や大きさなどの人の癖や、単語の表し方の複雑さから、単語を表している箇所と単語と単語の渡りである箇所の区別をつけることが難しいと分かり、自動単語抽出アルゴリズムの作成には、多くの手話ユーザデータが必要であると判断した。そこで、補助事業終了後にも引き続きデータを集めることで、自動単語抽出アルゴリズムの完成を目指す。

現在の単語抽出はユーザ側の動作で決定している。ユーザが、単語表現が終了した際に定位置に手を戻すことで、単語と単語の境目を決定でき、単語抽出を実現した。ユーザに

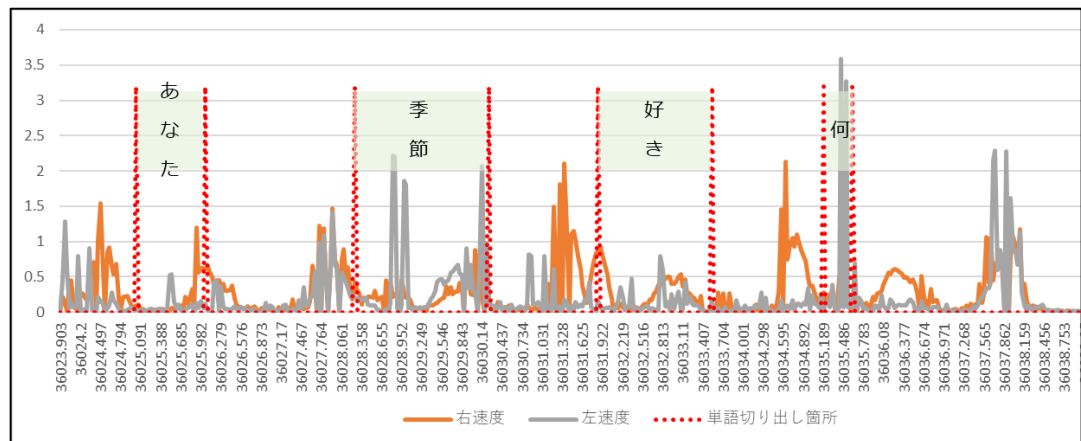


図 14 手の移動速度と単語切り出し箇所

としては、普段通りの連続した手話できないことや、抑揚をつけにくいなどの課題は残っているが、ユーザが単語と単語の境目を明示することで、抽出単語数が95%以上の正解率を実現できており、残りの5%以下も使い方を誤ったことでの誤認識のため、使い方を理解したユーザが行えば、限りなく100%に近い精度で単語の抽出を行えている。

(2) - 2. 手話単語の認識アルゴリズム

まず、手話動作と単語名を紐付けるための手話ディクショナリを作成した。手話ディクショナリには、左右の手それぞれについて「手の形」、「手の位置」、「手の動き」、「掌の向き」、「指の向き」を意味するデータを持たせた。次に、動作を解析し、単語に変換するため、手話動作における「手の形」、「手の位置」、「手の動き」、「掌の向き」、「指の向き」の各要素認識アルゴリズムを作成した。各アルゴリズムは、対象となる要素の認識モデルを用いて形状や位置、動き、向きを判定する。各アルゴリズム処理後の結果を手話ディクショナリと照合することで、単語翻訳を実現した。

5要素それぞれの認識アルゴリズムを用いて、各要素の認識精度を評価した結果、「手の形」は、71%、「手の位置」は90%、「手の動き」は50%、「掌の向き」は97%、「指の向き」は97%の認識率であった。「手の形」、「手の動き」の認識率について課題があり、他の要素と比較して認識精度が低かった。これは他の要素と比較して、「手の形」、「手の動き」は候補数が多く、誤認識しやすいためだと考えられる。「手の形」、「手の動き」の認識は、アルゴリズムの設計を見直すなどの対応を行い、認識精度を向上させる。

(2) - 3. 文脈依存モデル

認識精度を上げるため、単語翻訳に加えて、単語間のつながりを考慮した、文脈依存モデルプログラムの開発を行った。単語翻訳結果は1単語目の1位単語、1単語目の2位単語…、2単語目の1位単語、2単語目の2位単語…n単語目のm位単語と、それぞれの単語について候補を持たせ、単語翻訳結果として1位の単語であっても、前後の単語とのつながりが薄い場合や、逆に単語翻訳結果として5位の単語であっても、前後の単語とのつながりが濃い場合に、それらが考慮されるアルゴリズムを作成した。文脈依存モデルの有用性について評価を行った結果、文脈依存モデルをかける前後の翻訳結果を比較すると、文脈依存モデルをかけた後の方が8%高い結果が得られた。しかし、文脈依存モデルをかける前の翻訳精度が14%、文脈依存モデルをかけた後の翻訳精度が22%と、両者共に目指す精度を大きく下回った。この原因として、単語の翻訳精度が高くないことと文脈依存モ

デルや重み付けなどの調整が未だ甘いことが挙げられるため、引き続き対応する。

(2)－4. 高速パターンマッチング

動作入力後、1 秒以内の速さで結果を出力する高速パターンマッチングプログラムを開発した。既存研究では画像処理を駆使していた「手の形」の認識について、座標データを用いて行うことで、処理の高速化を実現した。高速パターンマッチングを行ったことで、文章の単語数に左右されるが、平均して1 単語あたり1 秒程度の応答速度を実現した。

(2)－5. 階層化認識モデル

様々な人物の違いに対応するため、階層化認識モデルの開発を行った。階層化認識モデルでは、性別、年齢、体型などのユーザ情報から階層構造を辿り、ユーザ本人用の認識モデルがなくても、既存モデルの中から最適なモデルを探すアルゴリズムである。階層構造を作成するプログラムと最適モデル探索用のプログラムを作成した。

当初、どのような単語翻訳においても、本人のデータから作成したモデルを用いることが最も認識精度が高くなることが想定されており、一方でシステムの全ユーザ分のモデルを個別に作成することは、事前データを大量に収集する点で非常に困難であることが予想されていた。そのため、階層化認識モデルでは、あらかじめユーザの性別、身長、体型、年齢、手の大きさ等のパラメータにより分類した複数の平均的なモデルを用意することで、ユーザ情報を基に適したモデルを自動で選定することで、本人モデルがない場合でも、比較的高精度な認識率を実現する仕組みを設計、構築する予定だった。しかし、階層化モデルを作成する際に材料となる個々の本人モデルの単語の認識精度が不十分であることが分かった。本来であれば複数のユーザデータから作成したモデルから、ユーザ情報を基に最適なモデルを選定し、翻訳をかけた際の認識精度の評価を行う予定であったが、まず本人モデルの認識精度を一定の水準に引き上げる必要があったため、本人モデルに限定した単語翻訳精度検証を優先的に行った。また当初の見解では、性別や年齢などのユーザ情報を用いてグルーピングを行うことで、意外な性別および年齢の組み合わせが同一の認識モデルが最適モデルとなるような階層化構造を想定しており、同年代且つ同性、体型も似ている場合は複数の認識モデルを用いた認識結果にそれぞれ差異は見られないと想定していた。しかし、同年代且つ同性、体型も似ている場合の被験者2 名について、複数の認識モデルを用いた認識結果に差異が見られた。このことから、当初の見解が覆り、性別、年齢、体型のみでは、最適にモデルを分けることや選ぶことができないことが分かった。こ

れらにより、指定したパラメータからモデルを選定するアルゴリズムのソフトウェア製作は完了しているが、100名分のデータ収集の内、完了できたのは41名分であるため、現在の手話翻訳システムへの入れ込み組み込みは控えているが、階層化モデルに含まれる個々のモデルデータは本人モデルと等価の構造であるため、組み込みのために新たなプログラムの開発は必要ない見込みである。本件については、補助事業終了後も引き続きデータを集め、データ解析により階層化認識モデルを充実させた後、システムに組み込む。

2-4. 翻訳精度評価

(1). はじめに

本事業における翻訳精度に関する評価結果を記載する。

(2). 方法および結果/考察

(2)-1. 翻訳精度評価用データ収集

翻訳精度を評価するため、また、認識精度向上に向けた取り組みに用いるため、データ収集を行った。データ収集は複数回に渡って実施し、非営利特定法人つくしのスタッフ（健聴者、聴覚障害者共）を中心にデータ収集にご協力頂き実施した。収集したデータは、日常会話に用いられる単語1000語から作成した例文の手話データ、「手の形」、「手の位置」、「手の動き」の認識モデルを作るためのアダプテーションデータであり、データ収集のためにお越し頂いた際、年齢や性別、普段手話を使っているかどうかなどのアンケートに回答頂いた。データ収集参加者について、男女比、健常者と聴覚障害者の比率、日常的に手話を使っているかのアンケート結果を図15に示す。収集したデータは41名分のデータであり、目標の100名には届かなかった。

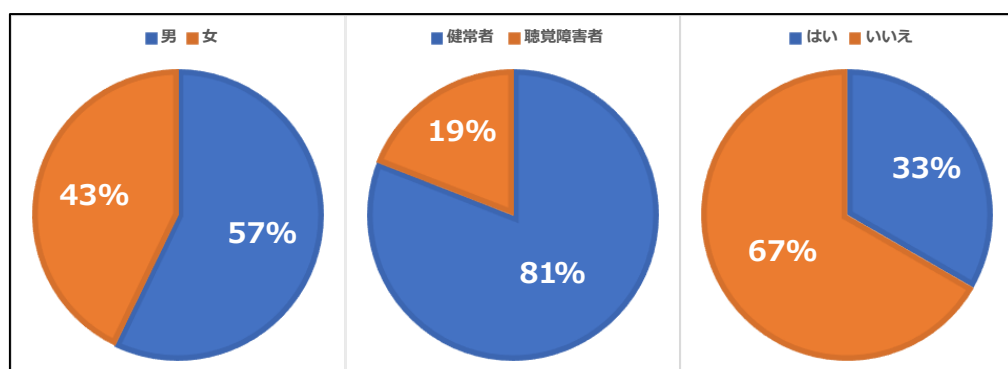


図 15 被験者データ

(左から男女比、健常者と聴覚障害者の比率、日常的に手話を使っているか)

(2) - 2. 翻訳精度検証

手話翻訳における翻訳精度について記載する。当社では、精度評価の観点を以下に設定し、それぞれについて、評価を行った。

[精度評価の観点]

① 単語翻訳

文章翻訳にかける前段階の単語候補に含まれている正解単語数の評価

② 意思疎通度

文脈依存モデル後結果を被験者に提示し、意思疎通が可能性の評価

① 単語翻訳

対象となる単語は段階的に増やすため、5 文章(単語数: 20)、10 文章(単語数: 38)、20 文章(単語数: 76)、40 文章(単語数: 185)、80 文章(単語数: 370)、175 文章(単語数: 965)と文章数を増やして行った。対象単語が 1000 語でない理由は、日常会話として使う単語を 1000 語ピックアップした内、指文字と数字を除く単語のみを対象単語としたためである。評価の結果、5 文章で 90%、10 文章で 89%、20 文章で 83%、40 文章で 73%、80 文章で 60%、175 文章で 47%であった。既存研究でも分かっていた通り、対象の単語数が増えれば増えるほど、認識精度が下がった。175 文章分の対象単語に対し、80%の認識率を達成できていない理由には、以下が考えられる。

[認識率 80%を達成できていない理由]

① 文章の入力データそのものに誤りがある

② 「手の形」、「手の動き」の認識精度に未だ伸び代がある

③ 5 要素の認識結果と手話ディクショナリの照合時のチューニングが甘い

以上の理由に対して、引き続きアルゴリズムの見直しや、認識モデルの調整を行い、認識精度向上を目指す。

② 意思疎通度

評価方法としては、本研究の文章データについて詳細を周知していない被験者に文章の翻訳結果を提示し、被験者の解釈を記述してもらう。被験者の解釈と文章の正解を比較して、本来手話話者が伝えたかった意味を読み取ることができるかを評価した。比較的、正解に近い翻訳が行えた文章を 10 文章、20 文章、30 文章と増やし、評価した結果、10 文章で 70%、20 文章で 40%、30 文章で 28%の正解率だった。評価結果の一部を表 1

表 1 意思疎通度評価結果（一部）

正解の日本語文	例文の正解単語列	例文の翻訳結果	日本語文書き換え
雷が鳴ると、慌てて座る。	雷 慌てる 座る 私	雷 慌てる 座る 私	私は雷に驚いて 座り込んでしまいました。
算数の本を夕方まで 貸しますよ。	算数 本 夕方 まで 貸す	暇 本 夕方 まで 貸す	暇なら本を夕方まで 貸しましょうか。
会議を欠席したから、 情報が混乱して困っている。	会議 欠席 情報 混乱 困る 私	会議 欠席 情報 ない 有る したい	欠席者用に議事録を 残して。

に示す。表 1 の 1 行目のように、例文の正解単語列と例文の翻訳結果に差異がなければ、日本語文書き換えも正解の日本語文と大きな差異がない結果が得られた。表 1 の 2 行目、3 行目は、どちらも例文の正解単語列と例文の翻訳単語列が一致していないが、2 行目は日本語文書き換えが誤った解釈に、3 行目は日本語文書き換えとしては誤っているが、正しい情報を要求していることは一致している。このように、文章の翻訳精度は意思疎通に影響しているが、文章翻訳が正解でないと意思疎通が取れないとまでは言い切れないことが分かった。本件は、言語学的知見からのアプローチが必要と考え、引き続き研究を行う。

最終章 全体総括

3-1. 補助事業の成果

本事業では、手話の自動翻訳デバイスの開発を行った。認識部位に特化したセンサを用いることで、単語の認識精度にして、20 単語（5 文章程度）で 90%、38 単語（10 文章程度）で 89%、76 単語（20 文章程度）で 83%、183 単語（40 文章程度）で 73%、370 単語（80 文章程度）で 60%、965 単語（175 文章程度）で 47%の認識率を実現した。また、動作完了から結果出力までの応答速度は 1 単語あたり 1 秒程度と、目標値に近い結果を実現した。以上から、研究および評価の結果、既存研究と比較して高精度且つ、高速な動作認識を実現できた。しかし、当初の目標値について、100 名分のデータ×1000 語の語彙数×80%の認識率の達成には至らなかった。また、単語抽出の自動化が新たな課題となった。手話翻訳デバイスとしては、タブレット端末の筐体試作を行ったが、筐体の小型化の課題が残った。

3-2. 補助事業終了後の技術的な課題

本事業を経て、残された技術的な課題を以下に記載する。残された課題の期限としては、大きく分けて、①筐体の無償貸与前までの課題、②筐体の無償貸与終了までの課題の二つに分類される。当社は、改めて以下課題について解決可能なスケジュールを組み、対応していく。

[残された技術的な課題]

①筐体の無償貸与前までの課題

- ・ユーザインタフェースの整備
- ・RealSense D435 用プログラム作成

②筐体の無償貸与までの課題

- ・筐体のダウンサイジング
- ・ユーザインタフェースの改良
- ・データ収集
- ・翻訳精度向上

3-3. 手話翻訳デバイス「Tesign」の事業化見通し

日本国内の聴覚障害者が約 36 万人であり、将来的には聴覚障害者の 5%（18,000 人）が当社の手話の自動翻訳デバイスを利用するように事業展開をする計画である。開発した手話翻訳デバイスは、障害者差別解消法の施行による官公庁への導入、企業の障害者雇用の義務化の進展による企業への導入、それらに伴う聴覚障害者の社会進出の増加による個人購入の増加、などにより聴覚障害者全体の 5%の普及は達成可能と考えている。聴覚障害者同士でのコミュニケーションは、メールなどによる文字よりも、ビデオ通話などの画像通信により手話で行っている。手話の自動翻訳デバイス

が開発されれば、聴覚障害者と健常者の会話における障害者負担が軽減される。障害者差別解消法の施行により公共の場での需要や、企業の障害者雇用の義務化の進展により健常者の中で働く聴覚障害者にとってコミュニケーションツールである翻訳デバイスの需要は高まるため、利用者は十分に見込める。

3-4. 事業化展開

事業終了後1年目に当社が、サンプル品(10台程度)を製作し、障害者支援施設であるつくしや、愛知県聴覚障害者協会の評価、助言を頂き改良を行う。これまで同社は、ジェスチャ(手振り動作)を音声に翻訳するシステムの試作を行っているが、つくしから手振り動作から発展させ複雑で動作の速い手話翻訳の要請を受けている。アドバイザー2者の要件を満たし、市場化に向けた製品を完成させる。2年目に、量産を開始し医療機器や福祉用具の販売事業を展開する八神製作所を代理店として販売する。

本製品は特定福祉製品の認定を受けることで、難聴度が高い聴覚障害者の使用において公的な購入補助が受けられる。それにより、障害者の購入の経済的負担を軽減し、普及を図る。また、平行して、ランディングページの作成、YouTube インストリーム広告、国際福祉機器展への出展等、販売促進活動を行う。

名古屋工業大学とは継続的に手話翻訳に関する追加研究を進める。障害者雇用施設につくしと、愛知県聴覚障害者協会とは、連携して製品の改良を進める。開発後には、福祉機器の製造販売において実績のある八神製作所と連携し、一般ユーザや福祉施設、公共施設への販売展開を進める。

[事業化スケジュール]

事業年度	平成31年度	平成32年度	平成33年度	平成34年度	平成35年度
サンプルの出荷	→				
追加研究	→	→			
設備投資					
製品等の生産		→	→	→	→
製品等の販売		→	→	→	→
特許出願	→				
出願公開	→				
特許権設定		→			
ライセンス付与			→	→	→

引用文献

- [1] 厚生労働省, “平成 18 年身体障害児・者実態調査結果,” 2008.
- [2] 一般社団法人 日本補聴器工業会, “JapanTrak 2018 調査報告 p. 17,” 2018.
- [3] みずほ情報総研株式会社, “手話者とのコミュニケーションを支援する手話認識システム,” 2014. [オンライン]. Available: https://www.mizuho-ir.co.jp/publication/report/2014/mhir07_shuwa_01.html. [アクセス日: 13 2019].