

画像認識技術を用いた特殊音に着目した  
指文字学習支援アプリケーションの試作Prototype of a Fingerspelling Learning Support Application  
Focusing on Special Sounds Using Image Recognition Technology中下 誇天<sup>†</sup> 石井 幹大<sup>‡</sup> 伊藤 一成<sup>†</sup>  
Goten Nakashimo Mikihiko Ishii Kazunari Ito

## 1. はじめに

現代社会において、多様性の尊重がますます重要視されている。多様な背景を持つ人々が互いに理解し合い、共存する包括的な社会の実現が求められており、LGBT に対する取り組みや移民受け入れなど、世界的にその取り組みが加速され SDGs の一環としても取り組まれている。日本においても「心のバリアフリー」を掲げ、異なる心身の特性や価値観をもつ人々が、円滑なコミュニケーションを図れる社会が目指されている[1]。

しかし、依然として多様性への理解が十分に行き渡っていない部分も存在する。その一例として、聴覚に障がいを持つ方々と非障害者との間にあるコミュニケーションの壁が挙げられる。手話や指文字は、聴覚に障がいを持つ人々にとって重要なコミュニケーション手段であるが、日本において非障がい者がこれらを学ぶ環境は十分に整備されていない。その結果、双方がコミュニケーションを取ることが難しく、理解し合うことが妨げられている。この課題を解決するには、非障がい者が手話や指文字を日常生活の中で気軽に習得可能とすることが望ましいが、忙しい現代社会において、十分な学習時間を確保することは非常に難しい。そこで、アプリケーションを活用することで個人の生活スタイルに合わせた学習を可能とし、継続的な学習を促進することを目指す。

今回、画像認識技術を活用した指文字学習支援アプリケーションを開発したので、本稿で報告する。本アプリケーションでは、特殊音（促音、拗音、濁音、半濁音）の扱い方に着目する。本稿の構成は次のとおりである。2章で聴覚障がい者におけるコミュニケーション手段についてまとめ、3章で開発アプリケーションについて述べたのち、4章でまとめる。

## 2. 先行研究

聴覚障がい者におけるコミュニケーション手段には手話、指文字、読話・補聴器といった方法が存在し[2]、それぞれの場面に応じて適切な、かつ負担が少ない手段を採用することが求められる。各手法の有効性の違いについて理解することで、多様な社会実現に近づくことができる。

指文字認識に関する研究分野では、各種センサを用いた多様な手法が提案されている。例えば、永井ら[3]や森本ら[4]は、Leap Motion およびモーションセンサ（PC カメラ）を併用し、手指の動作を同時に取得するシステムを提案し、

3DCG を判定結果と同時に提示することで正確な動作習得を支援するアプリケーションを開発した。小松・白井[5]は、実際にセンサグローブを作成し、画像認識の手法として提案しており、比較的高い精度の認識が可能であると結論づけている。また、宮崎・馬場[6]は、MediaPipe を利用して指文字学習アプリケーションの開発に取り組み、高精度の認識を行うには複数人のデータを取る必要があるが、作成者のデータだけでも、限定的な指文字であればゲームとして成立すると結論づけている。さらに、Leap Motion と MediaPipe の比較を行った生野・外村[7]は、認識距離や認識精度の観点から、MediaPipe によって高精度のトラッキングが可能であると結論づけている。

このように、各種センサを用いた手指動作認識に関する研究は多く行われているが、より多くの人が気軽に利用できるようにするためには、必要な機材を最小限に抑えることが重要と我々は考えている。また、日本における指文字は、直音文字に対応した 45 の片手の手型と、特殊音節の文字に対応した手型を移動させる型から構成される[8]。先行研究の多くは、直音文字に対応した型について取り上げており、手型の移動が必要な特殊音節の文字に対応した型を取り上げた研究は少ない。そこで、本稿では、MediaPipe を使用し、直音文字に加え、手型の移動が必要な特殊音節の文字にも対応したシステムを開発・提案する。

## 3. 開発アプリケーション

本章では、2章で述べた方針に基づき開発したアプリケーションについて説明する。本アプリケーションは、MediaPipe を使用し、手型の移動が必要な音にも対応することで、直音文字に加え、促音・拗音・濁音・半濁音といった特殊音節の文字も取り扱う。また、ユニークさを追求し、ユーザの継続的な学習を促進することも目指す。

## 3.1 手法

パーソナルコンピュータ上で動作するアプリケーションの構築を前提とし、ウェブカメラから取得される映像データに対して MediaPipe を用いた手指のランドマーク検出を行い、指文字の認識を試みる。指文字の判定には、第1筆者が各種指文字の動作を行ってデータを収集し、それを基に機械学習を実施することで認識モデルを作成する。作成したモデルを用いて、実際の映像入力から各指文字を分類・判定する仕組みを構築した。手型の移動が必要な特殊音節を表す指文字については、MediaPipe によって得られる X, Y, Z 各軸の座標データから手型の動きを解析し、時系列における位置の変化量に基づいて動作の有無を判定する手法を採用した。

<sup>†</sup> 青山学院大学 社会情報学部

School of Social Informatics, Aoyama Gakuin University

<sup>‡</sup> 青山学院大学 理工学部

College of Science and Engineering, Aoyama Gakuin University

### 3.2 アプリケーション説明

本アプリケーションは、「指文字学習モード」と、学習効果の向上を図る「リズムゲームモード」から構成される。

3.2.1 節で指文字学習モードについて、3.2.2 節でリズムゲームモードについて、それぞれ述べる。

#### 3.2.1 指文字学習モード

指文字学習モードの表示画面を図1に示す。ユーザは、画面右側の領域で任意の指文字を選択し、画面左側の領域に表示された見本を模倣しながら学習を進める。

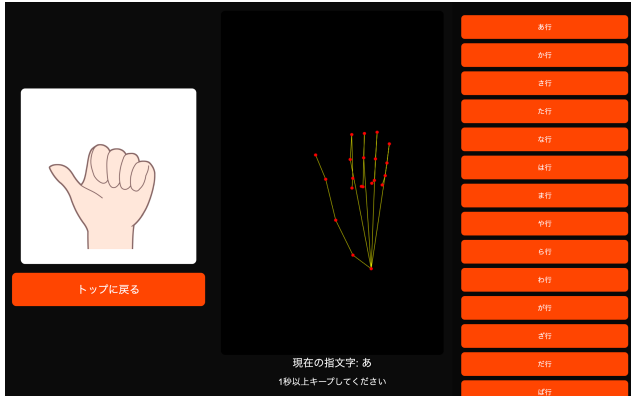


図1 指文字学習モードの画面

画面右側の領域には、直音の文字およびと特殊音節の文字が行別に分類され、プルダウン形式で配置されている。ユーザは、任意の行を選択し、その行に属する文字群から学習対象の文字を選択する。図1は、「あ」が選択された状態であり、画面左側の領域には「あ」を表す指文字の見本の図が表示される。ユーザは、見本の図を見ながら、選択した指文字を模倣する。ユーザが模倣した結果は、画面中央の領域にリアルタイムで描画される。ユーザの手型の動きは、ウェブカメラによって取得している。また、画面中央下側の領域には、「現在の指文字」を表示する欄を設けており、ユーザの手形と事前学習されたモデルを照合し、最も類似している文字を表示する。ユーザが正確な手型を1秒以上保持した場合、「正解」として認識され、図2の画面に遷移する。ただし、手型の移動が必要な特殊音節の文字をはじめ、動的動作を伴う一部の文字については、一種でも正しい動作が認識できれば正解とみなす例外処理を実装している。

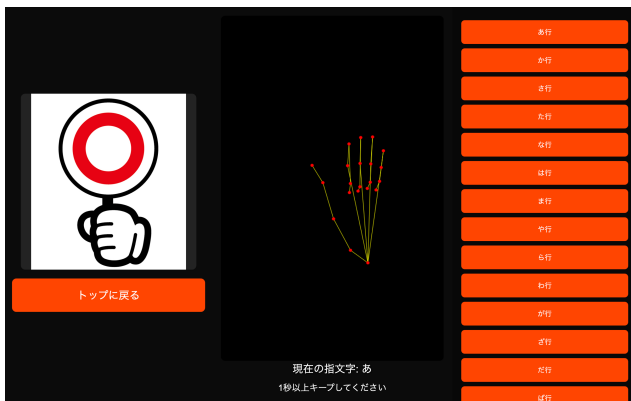


図2 指文字学習モードの画面（正解時）

#### 3.2.2 リズムゲームモード

本モードでは、あらかじめ設定された楽曲の歌詞が画面の上に流れる。画面を流れる文字が指定位置に達した段階でユーザが当該の文字を指文字で表現すると、認識アルゴリズムが正誤判定を実施する。

図3に、リズムゲームモードの初期画面を示す。初期画面は、ユーザが楽曲の選択を行う画面である。選択対象の楽曲には、一般的に認知度が高く、学習者のスキーマと関連付けやすい童謡を採用した。これにより、学習内容の定着を促し、学習効果の向上を図っている。



図3 リズムゲームモードの初期画面

リズムゲームモードでのプレイ画面の一例を図4に示す。画面中央には横長のバーが表示されており、楽曲の進行に合わせて指文字で表現すべき文字（例：「よ」「せ」）が右から左へと流れていく。画面左側に配置された赤い半透明の円は判定位置を示している。画面上を流れる文字が判定位置に到達したタイミングでユーザが指文字を作ると、認識アルゴリズムにより正誤判定が行われる。判定結果に応じて、スコアやコンボ数が加算される。

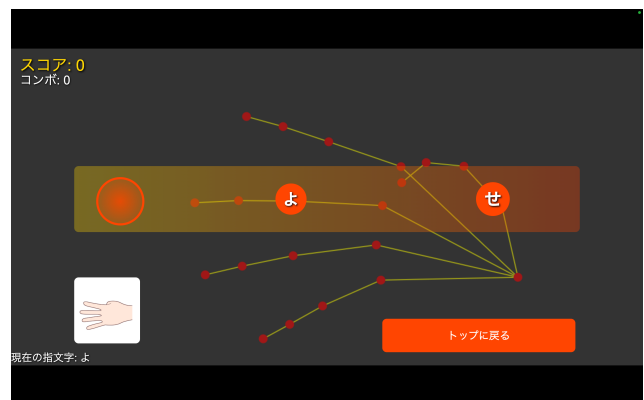


図4 リズムゲームモードのプレイ画面

ユーザの手型の動きはウェブカメラを通じてリアルタイムで取得され、その映像が画面中央部に表示される。この機能により、ユーザーは自身の動作を視覚的に確認しながらプレイすることが可能となる。また、初期画面にて「ヒントモード」をオンにすると、プレイ画面の左下には、画面上を流れる文字に対応した指文字の図が表示される。これにより、ユーザは表示された図を参照しながら模倣する

ことが可能となり、初学者に対する動作の理解と習得を効果的に促すことが期待される。

歌詞に設定された全ての文字が画面上を流れ終わると、楽曲の終了と同時に結果画面が表示される。図5に結果画面を示す。結果画面では、最終スコアとともに、ユーザが誤った指文字の一覧が表示される。



図5 リズムゲームモードの結果画面

#### 4. おわりに

本項では、MediaPipeを活用し、動的動作を含む指文字にも対応した学習アプリケーションの開発について報告した。動的動作を含む指文字に対応したことで、直音文字だけでなく、促音・拗音・濁音・半濁音といった特殊音節の文字も扱うことが可能となった。多様性が重視される現代において、聴覚に障がいのある人々とそうでない人々との間のコミュニケーションの壁を低くするためには、誰もが指文字や手話に触れ、学ぶ機会を得られる環境づくりが求められる。本稿は、これまで静的な指文字の認識に焦点が当てられてきた既存研究とは異なり、動的な表現も含めて学習できる点で、一定の意義を持つといえる。

本アプリケーションの特徴は、MediaPipeを利用することで特別なセンサを必要とせず、PCやスマートフォンのカメラを通じて、誰でも手軽に指文字を学べる環境を実現した点である。一方、MediaPipeの仕様上、Z軸方向の動きの認識精度には課題が残っている。また、複数の指が重なるような手型では、カメラ画像からの情報取得が不完全となり、認識精度が著しく低下する場面も見られた。今後、学習データの増強や認識アルゴリズムの改良を行い、さらなる精度向上を図る。

また、学習者に対する情報提示は、静止画を用いたものが中心となっているが、動的動作の理解を促すためには、実際の手の動きを記録した動画やアニメーションを導入することが有効と考えられる。加えて、学習を継続的に行うための工夫として、リズムゲーム形式にエフェクトを取り入れることで、学習者の興味やモチベーションを引き出す手法の導入も検討している。今後も本アプリケーションの改良を続けるとともに、被験者実験を通じて、本アプリケーションの有効性を検証する。

#### 参考文献

[1] 中野 泰志, “心のバリアフリーに対する取り組み”, Vol.24, No.1, pp.30-33 (2022).

- [2] 福田 友美子, 森本 行雄, 四日市 章, “聴覚障害者のコミュニケーション手段の使用に関する実態調査”, *Audiology Japan*, Vol.37, No.3, pp. 229-235 (1994).
- [3] 永井 敦, 竹内 健人, 森本 正志, “モーションセンサを用いた指文字学習支援アプリケーションに関する検討”, *情報処理学会研究報告 コンピュータビジョンとイメージメディア (CVIM)*, Vol.2016, No.5 (2016).
- [4] 森本 正志, 川除 慎吾, 加藤 雅弥, 田端 壱成, 加藤 秀康, 永井 敦, 竹内 健人, “モーションセンサを用いた指文字学習支援アプリケーション”, *情報処理学会論文誌 デジタルコンテンツ (DCON)*, Vol.8, No.2, pp.28-40 (2020).
- [5] 小松 周生, 白井 優旗, “深層学習を用いたセンサグローブによる指文字認識の検討”, *情報処理学会研究報告 アクセシビリティ (AAC)*, Vol.2018, No.4, pp.1-5 (2018).
- [6] 宮崎 優子, 馬場 哲晃, “手軽な指文字学習支援ゲームを目的とした認識システムの開発”, *情報処理学会研究報告 アクセシビリティ (AAC)*, Vol.2022, No.10, pp.1-3 (2022).
- [7] 生野 優輝, 外村 佳伸, “手指ジェスチャー認識に向けた Leap Motion と MediaPipe の比較検討”, 2020 年度情報処理学会関西支部 支部大会 講演論文集 (2020).
- [8] 井口 亜希子, 原島 恒夫, 田原 敬, “聴覚障害幼児の言語獲得における指文字の役割に関する文献的考察 -指文字獲得過程と語彙獲得の側面から-”, *障害科学研究*, Vol.43, No.1, pp.137-148 (2019).