

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

宮崎 優子^{1,a)} 馬場 哲晃^{1,b)}

概要: 手話や指文字は聴覚障がい者にとってのコミュニケーションツールである一方で、認識処理には高度な画像処理技術やデバイスが必須であったために健聴者が気軽に手話や指文字をインタラクティブに学習することが困難であった。本研究では、ブラウザで手軽に動作する javascript をベースとした深層学習による画像認識技術に着目した。とりわけ MediaPipe Hands による手指骨格のトラッキングと機械学習を用いた指文字認識システムの開発を行い、健聴者が気軽に指文字を学習するためのブラウザゲーム制作を目指す。本稿では、前者に当たる指文字認識システムの設計を手指骨格トラッキング情報を用いて実装し、その報告を行う。

1. はじめに

厚生労働省が平成 28 年に実施した調査によれば、日本国内で障がい者手帳が交付されている聴覚・言語障がい者数は約 34 万 1000 人であり、このうち聴覚障がい者数は約 29 万 7000 人だと報告されている。また同調査によると、65 歳未満の聴覚障がい者のうち 25.0%が手話や手話通訳を日常のコミュニケーションツールとして利用している。手話は手の形や動き、位置によって単語を表す言語である。一方で、日本語の五十音を手の形や動きに対応させた指文字も同様に聴覚障がい者のコミュニケーションに使われる。

聴覚障がい者数に対して、2019 年 9 月 1 日時点で聴力障害者情報文化センターに登録されている手話通訳士数が 3714 人であり、そのうち聴力障害者情報文化センターが 2019 年に行なった調査に回答した 1884 人の 34.8%が日々の業務に手話通訳業務が含まれていないと回答した。日常的に手話を使用する健聴者は少なく、また学習の機会も少ない。同調査によれば、手話通訳士になるための手話・手話通訳学習方法は手話サークルが 83.8%、手話通訳者養成講座が 59.8%、手話奉仕員養成講座が 51.0%、手話通訳専門養成校が 7.0%になっており、これらに通うには時間的・金銭的にコストがかかる。

この問題を解消するため、これまでにコンピュータを使った手話や指文字の認識研究は多くなされてきたが、

Kinect[1] や Leap Motion[2]、カラーグローブ、データグローブ [3] といったデバイスが必要であったり、カメラから自動で手領域を抽出する [4] といった専用デバイスや高い処理コストが必要とする手法が一般的であった。

そこで本研究は、ブラウザで動作する javascript ベースの深層学習による画像認識技術に着目し、PC に標準搭載されているカメラのみで体験できる指文字ブラウザタイピングゲーム制作を行う。

2. 関連研究

現在、特別なデバイスを使わずに手話を学習できる Web サービスには、日本財団が公開した手話タウン^{*1}や、アメリカの Hello Monday 社が開発した Fingerspelling<https://fingerspelling.xyz/>がある。手話タウンは香港中文大学と関西学院大学、Google 社によって開発された Web サービスで、Google 社が提供する機械学習用オープンソースライブラリ TensorFlow を活用し、人のポーズやジェスチャーを認識する PoseNet や顔の表情を認識する FaceMesh、手指骨格を検出する HandPose の 3 つの機械学習モデルを組み合わせて手話の認識を行なっている。日本手話と香港手話の 2 つを学ぶことができるが、指文字の学習は行えない。Fingerspelling は、ASL(American Sign Language) を学ぶことができるブラウザゲームである。英単語と手の 3D モデルが提示され、単語のスペルを綴っていくタイピングゲームとなっており、スペリングの時間が得点に換算される。指文字の識別には Google 社の提供する機械学習ソリューションフレームワークである

¹ 東京都立大学
Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan

a) miyazaki-yuko@ed.tmu.ac.jp

b) baba@tmu.ac.jp

^{*1} <https://signtown.org/>

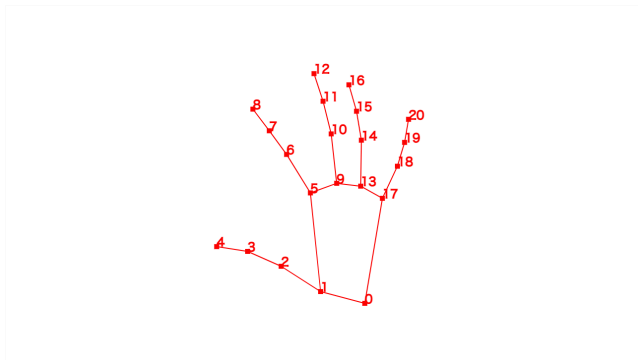


図 1 MediaPipe Hands による手の関節座標

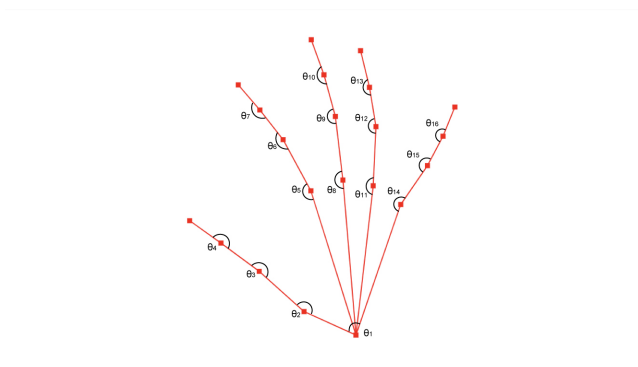


図 2 特徴量として使用した 16 角度

MediaPipe の中のハンドトラッキングモジュールである MediaPipe Hands が使用されている。

本研究でも MediaPipe Hands を使用した手指骨格のトラッキングと機械学習によって認識システムを開発し、日本語の単語を指文字で綴るタイピングゲームを制作する。

3. 実装

3.1 設計

認識システムは Shimizu ら [5] の関節の角度を特徴量とする手法を参考にした。MediaPipe Hands によって得られる 21 個の関節座標 (図 1 参照) から、各関節の角度と手のひらの付け根の角度を計算し、図 2 に示す 16 個の特徴量を使って機械学習を行なった。機械学習には、ブラウザ上で手軽に機械学習を行える ml5.js の classification を使用した。ml5.js^{*2}は TensorFlow.js をベースに作られたライブラリであり、学生やアーティストなどの幅広い層に向けて開発されているために取り扱いやすく、リファレンスも豊富であったためこれを利用した。

データは、図 3 のようにひらがなが描かれたボタンを押し、画面に向かってそのひらがなに対応した指文字を提示することで 16 個の関節の角度を収集する。ボタンを押した 1 秒後にデータ収集が始まり、10 秒間収集を続ける。MediaPipe Hands は 30fps で動作するため、一音あたり約 300 データが収集される。収集されたデータは json ファイ

^{*2} <https://ml5js.org/>

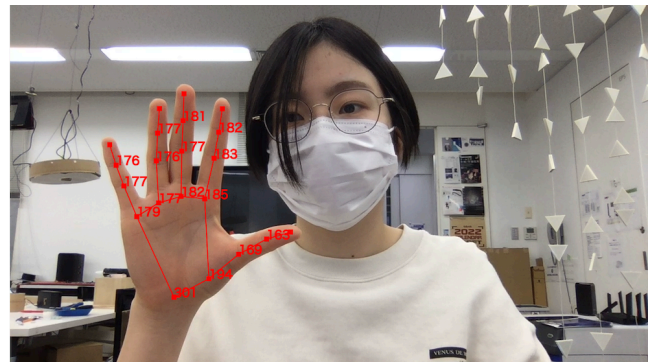


図 3 データ収集画面

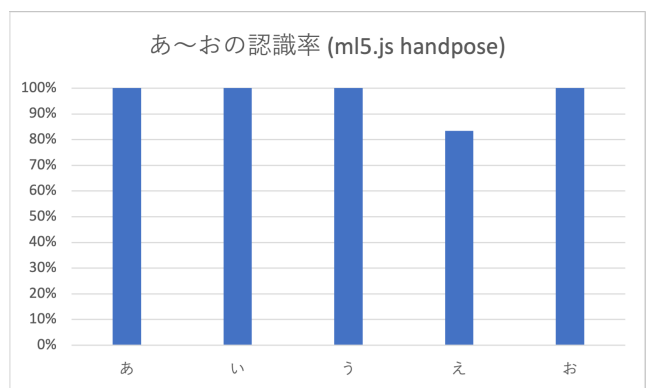


図 4 ml5.js handpose を使った「あ」～「お」の指文字の認識率



図 5 ml5.js handpose を使った「あ」～「こ」の指文字の認識率

ルとして書き出した。

3.2 予備実験

MediaPipe Hands を実装する前に、機械学習で使用している ml5.js のハンドトラッキングである handpose で指文

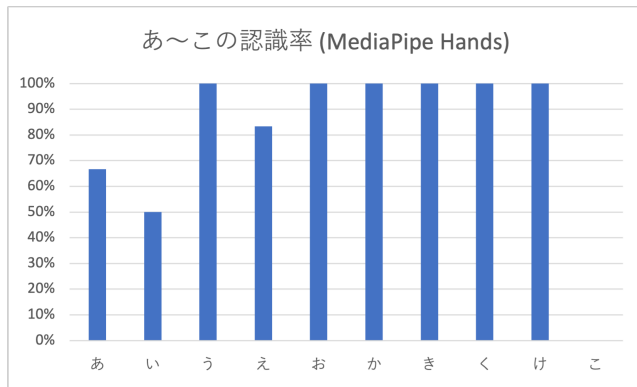


図 6 MediaPipe Hands を使った「あ」～「こ」の指文字の認識率

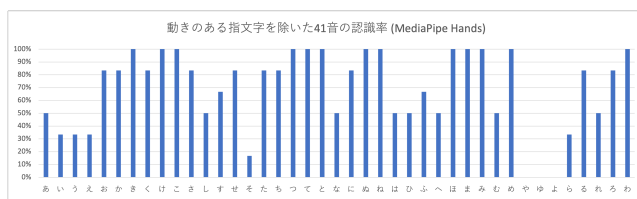


図 7 MediaPipe Hands を使った動きのある指文字を除いた 41 音の認識率

字の認識を行った。まず「あ」から「お」までの 5 種類の指文字を、手指関節座標データを収集し学習させ、識別テストを行ったところ認識が可能であったが、座標データでは画面内の手の大きさや位置が変わってしまうため認識結果が変わってしまうため、システムとしてロバスト性が低い。そのため、同じく「あ」から「お」までの指文字を、関節の角度を計算してデータ収集、学習、認識テストを行ったところ、認識率は図 4 のようになった。次に指文字の種類を増やし、「あ」から「こ」の 10 種類の指文字の関節角度を計算し認識テストを行った。結果として図 5 のようになり、「い」「え」の認識精度が低く、最後にデータを収集した「こ」が認識不可能であった。原因は不明だが、データ収集の順番を変えると「こ」は認識できたため、最後に収集した一音が認識不可能であることが判明した。

3.3 開発

ml5.js の handpose は TensorFlow.js のハンドトラッキングを古いモデルで行なっていたため、MediaPipe Hands に比べてトラッキングの精度が明らかに低かった。そのため、MediaPipe Hands に切り替えて実装を行なった。予備実験と同じく「あ」から「こ」の 10 種類を、関節の角度を特徴量として収集し、学習、認識テストを行ったところ、図 6 のようになった。「あ」の認識率は若干下がったものの、認識できていなかった「い」、「え」の認識率が向上し、認識率の低かった「う」、「か」の認識率も上昇した。ただし、モデルを切り替えても最後に収集した「こ」は識別不可能だった。

次に、46 音のうち動きを伴う指文字である「の」「も」

「り」「を」「ん」の 5 音を除いた 41 音の識別をテストした。予備実験により、最後の 1 音が識別不可能であることがわかってきたため、ダミーを 1 音入れた 42 音をデータ収集し、識別テストを行なった。結果として、ダミーと最後に収集した「や」「ゆ」「よ」の 4 音が識別不可能であることがわかったが、収集の順番を入れ替えれば、46 音全て識別可能であった。「や」「ゆ」「よ」を認識していないシステムの認識率は図 7 のようになり、文字によって認識しやすいものとしにくいものがあることがわかった。また、「え」や「そ」などは人によって手首や指の角度、開き具合が異なるため、筆者のみが教えたデータでは不十分であることがわかった。しかし、指文字を初めて行ったユーザもいる中で、41 音の内 13 音は 100% の識別率になり、限定的な文字であればゲームとして遊ぶには十分に可能であると考えられる。

4. 今後の展望

今後は全てのデータがきちんと学習できない原因を解明することが第一の課題である。また、今回除いた「の」「も」「り」「を」「ん」の動きを伴う指文字の識別方法を実装していきたい。動きを伴う指文字が識別可能になれば、濁音や半濁音も識別可能になる。認識システムが開発できれば、当事者を含めた様々な人からデータを収集し、認識の精度を上げていく必要がある。また、Web タイピングゲームのシステムや UI を制作し、ユーザー評価を行いたい。

参考文献

- [1] 裕介織茂, 祐司玉國, 大介高橋, 教佳岡本: Kinect を用いた指文字認識の検討 (学生研究発表会), 映像情報メディア学会技術報告, Vol. 38.9, No. 0, pp. 31–32 (オンライン), DOI: 10.11485/itetr.38.9.031(2014).
- [2] 真生子船阪, 由羽石川, 雅美田, 和貴 城: Leap Motion Controller を用いた指文字認識, 情報処理学会研究報告. MPS, 数理モデル化と問題解決研究報告, Vol. 2015-MPS-102, No. 8, pp. 1–6 (オンライン), 入手先 (<https://ci.nii.ac.jp/naid/110009877760/>) (2015).
- [3] 優旗白石, 智彦土屋, 伸子加藤, 文雄米山, 明寿設楽: 深層学習による多次元時系列データを用いた指文字認識, 筑波技術大学テクノレポート, Vol. 28, No. 1, pp. 58–59 (オンライン), 入手先 (<https://ci.nii.ac.jp/naid/120006949183/>) (2020).
- [4] 勝文井上: 日本語の指文字を用いたタイピングシステムのプロトタイプ構築, 日本画像学会誌, Vol. 58, No. 3, pp. 340–344 (オンライン), DOI: 10.11370/isj.58.340 (2019).
- [5] Arata, S. and Tetsuaki, B.: Real-time Finger Spelling “Yubimoji” Recognition by Skeletal Hands Features Tracking, Tokyo, Japan (2014).