手話復習ツールのための識別対象区間の抽出方法の検討

Investigation of Method for Extracting Classification Target Segments for Sign Language Review Tool

亀田 凱聖† 川口 大夢† 笹島 和哉†
川喜田 佑介† 西村 広光† 田中 博†

Kaisei KAMEDA† Hiromu KAWAGUCHI† Kazuya SASAJIMA†
Yuusuke KAWAKITA† Hiromitsu NISHIMURA† and Hiroshi TANAKA†

†神奈川工科大学 情報学部
†Kanagawa Institute of Technology, Faculty of Information and Computer Sciences

E-mail: {s2121104, s2121022, 2123026}@cco.kanagawa-it.ac.jp,
{kwkt, nisimura, h_tanaka}@ic.kanagawa-it.ac.jp

1. はじめに

ニュース映像の中にも手話映像が含まれ，手話はより身近なものになっている．実際に手話学習のための映像コンテンツやホームページは多数存在する[1]．しかし，これらは基本的に映像やイラストを視聴することが中心になっており，覚えた手話の妥当性を評価することは身近に手話利用者がいない場合難しい．筆者らはこの問題解決の一手段として，ユーザがカメラの前で覚えた手話動作を再現し，その結果を AI によって判定，フィードバックする手話復習ツールを開発している[2]．

現ツールでは識別モデルを作成するために効率性を重視して取得した手話動作のデータ[3]と，手話復習ツールにおいてユーザの手話動作の開始終了判定を自動で行うことを重視して取得した動画データでは手話の開始・終了ポーズと時間軸上の判定区間が異なり，復習ツールとしての識別評価に悪影響を与えることが明らかになった．

本報告では識別モデル作成のための動画データの手話動作区間と，手話復習ツール利用時のユーザによる手話動作の識別対象区間を一致させる手法について検討し，その効果を確認した結果を述べる．

2. 復習ツールにおける課題

2.1 復習ツールの概要

本検討で扱う手話復習ツールは主に 2 つの機能部から構成している(図 1)．モデル構築部では，手話動画データに対して Google 社が提供する MediaPipe を適用し，得られた骨格情報から特徴抽出を行い，SVM (Support Vector Machine)アルゴリズムを用いて識別モデルを作成する．そして，動作識別部でモデル構築部において作成した識別モデルを使い，カメラの前で行った手話動作やあらかじめ撮影した手話動作動画に対して識別を行う．識別結果は SVM の出力結果から得られる尤度情報をもとにスコア上位 5 位までの単語の識別結果を表示する．ユーザは，この識別結果から自分の行った手話動作の確認ができる仕組みである．

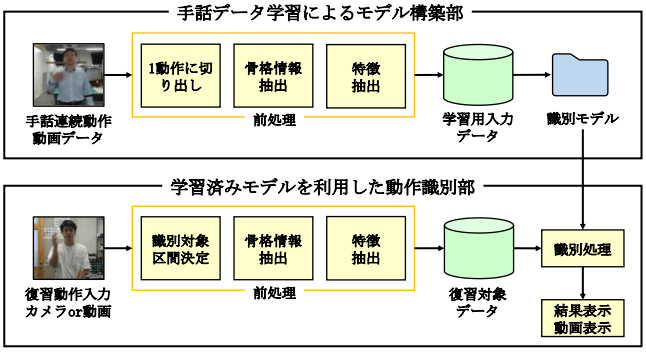

図 1. 手話復習ツールの構成

2.2 識別モデルのための手話動作動画データ

識別モデル作成のための手話動作は，株式会社ケイ・シー・シー様の手話学習者向けの動画辞典アプリ Smart Deaf[1]から，参照数が多い健康医療関連の単語を中心に選定した．新たに記憶できる手話動作数を考慮して，10 単語を 1 グループとした．復習ツールとして，この 1 グループ 10 単語をグループ単位で追加できるようにしている．学習対象として選定した単語の一例を表 1 に示す．

なお，手話動画データは手話指導資格を有する指導者のもと，指導者を含めた 4 名分の手話単語動作の動画データを取得している．データ取得の際，効率を重視し，1 単語につき 12 動作連続で行った．この時，各手話動作の開始と終了時に「気を付けの姿勢」をしている．動画

の撮影にはロジクール社のカメラ(型番：C920)を使用し，撮影条件はフレームレート 30fps，解像度は 960×720 である．

表 1．対象手話動作の一例(1 グループ)

1．イライラ	2．顔色	3．近視
4．健康	5．たくましい	6．ダイエット
7．涙	8．発達障害	9．パニック
10．肥満		

2.3 復習動画データの取得

復習動画データの取得方法を図 2 に示す．手話復習ツールのユーザの手話の復習動作の取得区間は，ツールがユーザの手話動作開始と終了を認識しやすいように，「腹部で手を組む」という開始ポーズを 3 秒カウントしたときを開始とし，「気を付けの姿勢」である終了ポーズを 1 秒カウントしたとき終了とする．この開始から終了までの動作を復習動画データとして取得している．

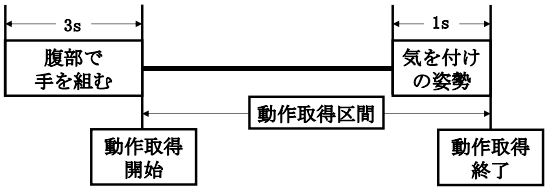

図 2．復習対象データの取得方法

2.4 動作データの開始・終了ポーズの相違

識別モデル作成のための動画データの手話動作開始ポーズと終了ポーズは 2 章 2 節で述べた通り，「気を付けの姿勢」であり，手首は画面外に外れる．一方で復習対象データでは「腹部で手を組む」開始ポーズと「気を付けの姿勢」の終了ポーズであり，手首が画面内に位置するため，識別モデル作成のための動画データと復習対象データの開始・終了ポーズに相違がある(図 3)．「たくましい」という問題が提示された画面である．

図 3．開始・終了ポーズの相違

2.5 識別モデルのための動画データ動作区間

ポーズの相違により，識別モデル作成のための手話動作区間と，手話復習ツール利用時の手話動作の識別対象区間が異なる．まず，識別モデルのための動画データの手話動作区間の抽出方法について述べる．2 章 2 節で述べたとおり，識別モデルのための動画データは 12 連続動作であるため，1 手話動作ごとに動作の開始，動作の終了の判定を行い 1 動作に切り出す必要がある．今回は以下の方法で切り出しを行った．まず，動画データ下限から 5%の位置を閾値とし、式(1)の条件を設定した(図 4).

・動作の開始
10 フレーム中 6 フレーム以上で式(1)が真であるという条件を最初に満たしたフレーム
・動作の終了
式(1)が真であるとき，10 フレーム中 6 フレーム以上で式(1)が偽になったフレーム

この動作の開始から終了までを識別モデル作成のための動画データとした．

図 4．識別モデルのための動画データと閾値の設定

$$\frac{W_{Ly}+W_{Ry}}{2} \leq Threshold \qquad (1)$$

ここで，
W_{Ly}：右手首の y 座標値
W_{Ry}：左手首の y 座標値
$Threshold$：閾値
である．

3. 識別対象区間の決定

本検討以前では手話動作の前後に手話動作ではない動作が 1 秒あると仮定し，動作取得区間から 1 秒を考慮して取り除いた区間を識別対象区間としていた(図 5)．しかし，この手法は識別モデルのための動画データと復習時の識別対象データとして使用する動作区間が異なると

いう課題がある．

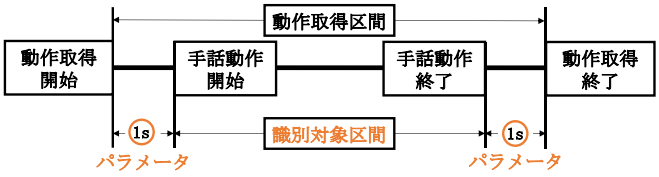

図 5. 従来の識別対象区間決定の方法

3.1 手話動作区間の一致方法

識別モデル作成時の動画データの手話動作区間と復習時の識別対象区間を一致させる目的で，識別対象区間の決定方法を提案する．その方法を図 6 に示す．

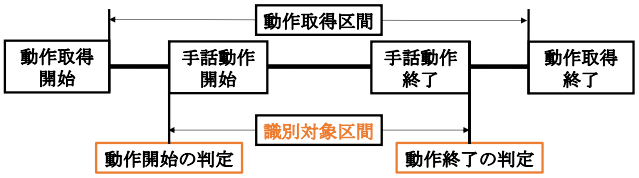

図 6. 提案する識別対象区間決定の方法

本提案手法では，2 章 5 節で示した識別モデルのための動画データの手話動作区間と同一の条件で復習対象動画データの識別対象区間を抽出し，両データを一致させる．以下に識別対象区間の抽出方法を示す．

・動画の下限の決定
識別モデルのための動画データの下限に相当する位置を復習ユーザの肘と手首の間とし，復習対象データの開始ポーズは両肘と両手首の y 座標平均値の位置を識別モデル作成時の動画データの下限と同じ位置とみなす(図 7).

・閾値の決定
識別モデルのための動画データは H1 の 5%の高さを閾値とした．復習対象データも同様に H2 の 5%の高さを閾値とした(図 7)．2 章 5 節で示した動作の開始・終了と同様の方法で動画データから識別対象区間として抽出した．

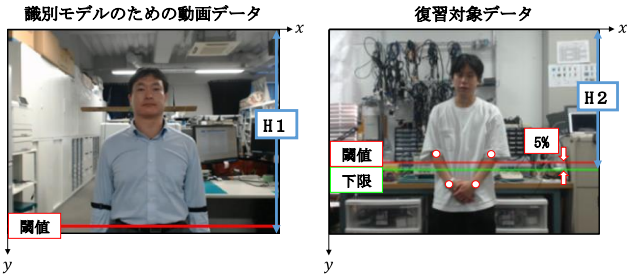

図 7. 手話動作の開始・終了の判定

3.2 実験による確認

識別モデルのための動画データの手話動作区間と提案手法により抽出した復習対象データの識別対象区間の開始・終了ポーズが一致していることを確認した(図 8)．他の手話動作も同様に開始・終了ポーズが一致していることを確認した．

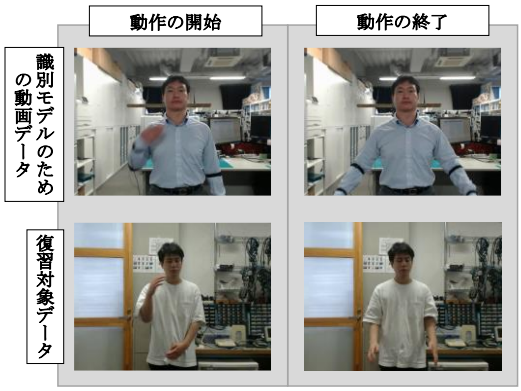

図 8. データ区間抽出結果（手話：健康）

4. 識別精度の比較

提案手法の効果を確認するために，従来手法と提案手法による本ツールの出力である識別精度を比較した．比較を行うにあたり，手話動作の位置と動作時間を考慮して表 2 に示す手話単語を表 1 の中から選定した．この手話動作を 3 人の実験者が 1 単語 5 回ずつ行った手話動作動画を収録し，従来手法と提案手法による手話動作動画の識別結果を比較した．

表 2. 比較検証で扱う手話単語

手話単語	特徴
イライラ	画面上部で行う手話
肥満	画面下部で行う手話
たくましい	手話動作時間が短い
顔色	手話動作時間が長い
健康	

各単語を 5 回行った尤度の平均値を表 3 示す．ここで、識別結果である単語とともに尤度情報をユーザが分かりやすいように 100 倍に換算して提示している．10 ポイント以上の上昇は「肥満」「顔色」「健康」で見られた．一方で，「イライラ」の尤度は 10 ポイント以上低下が見られた．また，「たくましい」は実験者によっては尤度が大きく減少した．

多くのケースで提案手法による識別精度の向上が確認できた．これは識別モデルのための動画データと復習対象データの識別対象区間の一致によるものと考えられる．

表 3. 識別結果である尤度の比較結果

		イライラ	肥満	たくましい	顔色	健康
実験者１	従来	85.4	83.1	0.3	44.2	18.3
	提案	64.9	93.1	10.5	69.2	76.4
	変化量	**-20.4**	**10.0**	**10.2**	**25.0**	**58.1**
実験者２	従来	64.1	67.0	33.6	89.1	81.6
	提案	74.9	80.9	3.4	94.7	77.6
	変化量	**10.8**	**13.8**	**-30.2**	5.6	-4.0
実験者３	従来	80.7	37.6	9.8	81.6	68.4
	提案	52.8	35.2	52.6	97.5	90.2
	変化量	**-27.9**	-2.4	**42.8**	**15.9**	**21.8**

（赤字は10ポイント以上上昇，青字は10ポイント以上低下）

5. まとめと今後の課題

識別モデル作成のための動画データの手話動作区間と，手話復習ツール利用時のユーザの手話動作の動画データの識別対象区間を一致させることを目的に，動画の抽出方法を検討，提案した．その結果，手話動作の識別結果である尤度の値を向上させることができた．今後識別対象の単語を増やすために新たに手話動作データ取得する際に，今回の検討で得られた知見を反映することを今後の課題とする．

謝　辞

本研究開発は，公益財団法人電子通信普及財団の助成を受けたものです．また，手話動作をご指導いただいた株式会社ケイ・シー・シーの関係各位に感謝いたします．

文　献

[1] 株式会社ケイ・シー・シー，手話学習者向けスマートデバイス動画辞典，SmartDeaf，https://www.smartdeaf.com/.

[2] 亀田凱聖，川口大夢，西村広光，田中博，“手話学習支援を目的とした手話復習ツールの拡張機能，”電子情報通信学会情報・システムソサイエティ特別企画予稿集，ジュニア＆学生ポスターセッション，p.60, 2024.

[3] T. Wakao, T. Sato, W. Odagiri, Y. Kawakita, H. Nishimura, H. Tanaka, and J. Mitsugi, “Accurate sign language motion classification using synchronized backscatter sensors, ” Nonlinear Circuits, Communications and Signal Processing, 2PM3-1-4, pp.385-388, 2022.