

# 多変量関数主成分分析を用いた 手話学習支援システムの構築

高須優<sup>1</sup> 櫻田京之介<sup>1</sup> 荒木由布子<sup>1</sup> 西田昌史<sup>1</sup> 綱川隆司<sup>1</sup> 西村雅史<sup>1</sup>

**概要:** これまで我々は手話人口を増やせるように、独学で容易に手話を学習できることを目指して、Kinect と 3D モデルを用いた手話学習支援システムを提案した。しかし、学習者が手話のどこがどれくらいお手本とずれているかがわからず、動作の特徴が捉えられないという問題があった。そこで、本研究では複数の観測部位の動きを関数化して同時に扱い、動作ごとの動きを捉えることができる多変量関数主成分分析を用いた手話学習支援システムを提案する。評価実験を行った結果、従来の手話学習支援システムに比べて提案システムのほうが学習者の手話をよりお手本に近い動きにさせることができた。

**キーワード:** 聴覚障がい者、手話、学習、多変量関数主成分分析、Kinect、3D モデル

## Development of sign language learning support system using multivariate functional principal component analysis

YU TAKASU<sup>†1</sup> KYONOSUKE SAKURADA<sup>†1</sup> YUKO ARAKI<sup>†1</sup> MASAFUMI  
NISHIDA<sup>†1</sup> TAKASHI TSUNAKAWA<sup>†1</sup> MASAFUMI NISHIMURA<sup>†1</sup>

**Abstract:** In order to increase the sign language population, we proposed a sign language learning support system using Kinect and 3D model. However, there was a problem that the learner could not grasp the features of the action, because he did not know where the sign language deviated from the example. In this paper, we propose a sign language learning support system using multivariate function principal component analysis. As a result of the evaluation experiment, the proposed system can move the sign language of the learner closer to the model compared to the conventional sign language learning support system.

**Keywords:** hearing impaired, sign language, learning, multivariate function principal component analysis, Kinect, 3D model

### 1. はじめに

聴覚障がい者がコミュニケーションを行うための言語として手話がある。手話は手の動きや手の形、手の位置、表情などを使って相手とコミュニケーションを行う視覚言語である。平成 29 年 4 月に公示された「特別支援学校小学部・中学部学習指導要領」の中で手話を適切に活用してコミュニケーションを行うようにと明記されており[1]、耳が聴こえない生徒が手話を習得する機会が増加している。聴覚障がい者の手話使用者の増加に伴い、健常者も手話を学習する機会が増えている。

現在、手話を学ぶ方法として手話サークルや手話講習会といったグループ学習と本や DVD を使用した自宅学習がある。グループ学習は学習者の好きな時間に学習ができないという問題がある。自宅で学習するためによく使われる本は手話の手の動きが把握しにくいという問題があり、映像コンテンツでの学習が有効であることが既に示されている[2]。映像コンテンツとしてよく使われている DVD は手

話が正面から撮影されていることが多く前後の方向の動きが把握しにくい、学習者自身が行っている手話が本当に正しいのか確認する方法がないといった問題があった。そのため映像コンテンツとともに学習者の手話がどのくらいお手本の手話と似ているかを評価できる手話学習支援システムの構築が必要である。

伊藤ら[3]は、Web カメラを用いることで学習者の手話とお手本手話の映像を比較可能な手話学習支援システムを提案し評価を行った。実際にシステムを使用してもらい、使用者にアンケートを行った結果、手話初学者に対して有効なシステムであるという結果が得られた。

鎌田ら[2]は、手話通訳において重要である手話に対応する適切な日本語訳の選択に関する学習に焦点を当てた手話学習システムを検討した。

高林ら[4]は、深度センサを用いて手領域を抽出し、正しい指文字を判定し、正しくなければ何が間違っているかフィードバックを与えるシステムを試作している。これは事

<sup>1</sup> 静岡大学  
Shizuoka University

前によくある間違いを学習させることで実現している。

永井ら[5]はモーションセンサを用いて指文字の3次元情報を取得し、指文字学習支援アプリケーションの検討を行った。習熟度を測るテストモードを実装することで、繰り返し学習を行うことへの動機づけも行った。

和泉ら[6]は、Webカメラの代わりにKinectを用い、学習者の手の動きを軌跡として表示し、その軌跡とお手本者の手話動作の類似性評価を行う機能、両手の各方向におけるお手本と学習者の手話の類似度をDPスコアで表示する機能、お手本と学習者の手話を3Dモデルでモデル化して様々な角度から手話を確認する機能を持った手話の学習支援システムを提案している。

しかし、これらのシステムでは各個体の動作の差や時間変化に伴う手話動作の変化にどのような特徴があるかを捉えられていなかったため、正確に手話を学習することが難しいという問題がある。また、手話の動きを的確に捉え、動作の特徴を明らかにするための手法や手話を効率的に学習する方法はいまだ確立されていない。

それに対して櫻田ら[7]は多変量関数主成分分析を用いた統計分析によって観測部位ごとの分析結果を用いて、手話動作の中で個体ごとの特徴は、どの動作に現れるのか、その特徴は時間の経過とともにどのように変化するか、大きさはどれくらいなのかを明らかにした。また、単語の種類によって個体差の違いも分析できることを示した。さらに、従来の学習システムでは学習者本人のデータのみを扱っているが、本研究ではシステム使用者以外の学習者のデータも分析に用いる点は新たな試みである。

本研究では多変量関数主成分分析による手話分析方法を従来のKinectと3Dモデルを用いた手話学習支援システムに実装し、新たな手話学習支援システムを提案する。

## 2. Kinectと3Dモデルを用いた手話学習支援システム

従来システムでは手話時のデータの収集にKinect for Windows v2(以降Kinect)を用いる。図1がKinectである。本システムではカラー画像と両手の位置データ、各関節の回転角度をKinectによって収集する。また、システムを操作するためのデバイスとして**エラー! 参照元が見つかりません**。のようなUSBフットペダルスイッチを使用している。フットペダルスイッチを押して録画することで、学習者単身でも実際の手話部分のみを記録することができる。また、手話の終了と同時に手話データの録画の停止をすることができるため、終了位置のセグメンテーションを行う必要がないといった利点もある。



図1 Kinect for Windows v2(左図)



図2 USBフットペダルスイッチ[8](右図)

次に、本システムの処理の流れを説明する。

システムを起動すると図2のような学習単語選択画面が表示される。画面左側に表示されている単語を選択すると、画面右側にその単語のお手本手話映像と手の動きや手型の説明が表示される。右下の学習開始ボタンを押すことで手話データの記録フェーズに遷移する。なお、手の動きや手型の説明については日本語・手話辞典[9]および日本語・手話辞典[10]より引用している。



図2 学習単語選択画面

記録フェーズに遷移するとKinectが起動し、Kinectのカラーカメラの映像に、Kinectの公式SDKで取得可能な手の位置のプロットを合成した映像が表示される。録画開始の入力をする手話データの記録を始める。手話データはカラーカメラの画像と両手の位置データ、3Dモデルを動かす処理で使用する各関節の回転角度の3つであり、1/30秒単位で記録する。手の位置データはKinectを原点とし、図3の矢印の方向が正となるような座標軸上の点として出力され、3次元のデータとして出力される。上下方向をx軸、左右方向をy軸、前後方向をz軸として位置データを取得した。

録画停止の入力まで手話データを記録し続け、録画停止の入力をする手話データの処理フェーズに遷移する。図4が手話データの記録フェーズにおける実際のシステムの画面である。右上の画像がKinectの映像であり、青と赤の点は推定した両手の位置をカラーカメラの座標軸に変換し、プロットしたものである。



図 3 手の位置の座標軸



図 4 手話データの記録画面

処理フェーズでは手話画像の映像化、手話データのテキスト出力、DP 照合の 3 つの処理を行う。出力したカラー画像に、そのフレームまでの手の位置をプロットし、各プロットを線で結ぶことで軌跡付きの動画を生成している。

学習者の手話とお手本の手話を比較し数値化するために DP 照合を用いる。DP 照合を行う前に、移動平均を用いて平滑化し、初期フレームの手の位置が原点となるよう全ての手の位置データを平行移動させる前処理を行う。それぞれ前処理を行った学習者の手話とお手本手話の手の位置のデータ列を DP 照合でパターンマッチングし、正規化されたパターン間距離を計算する。DP 照合でのパターン要素間の距離は対応パターンの手の位置を左右、上下、前後の 3 次元ごとの要素にし、次元ごとに減算し、絶対値を取ったものを用いる。

データの処理終了後、結果の提示を行う。図 5 が実際の結果表示画面である。画面右上に学習者の、画面右下にお手本の軌跡を描画した手話動画を並べ、再生することで学習者がお手本と比較しながら自身の手話を確認できる。

左下には学習者の手話とお手本手話がどれだけ近いかを表す DP スコアが表示される。これは学習者の手話とお手本手話の手の位置データを DP 照合で比較したスコアを両手の左右、上下、前後の 3 次元ごとの要素ごとに出力した結果である。

左上は人型 3D モデルに手話を再現させ、さまざまな方向から手の動きの確認ができる機能である。Kinect の公式 SDK と Unity[11]を用いることで、各関節の回転角度を取得することができる。この関節の回転角度を 3D モデルに対応させることで、Kinect で認識された人物の動きを 3D モ

デルに反映させることができる[12]。この技術を応用し、手話時の関節データを記録しておくことで、3D モデルに手話の動きをさせることができる。これにより 3D モデルをさまざまな方向から見て手話の動きを確認することができる。



図 5 結果表示画面

### 3. 多変量関数主成分分析に基づく手話学習支援システムの構築

#### 3.1 多変量関数主成分分析

多変量関数主成分分析とは関数データ解析法の一つである。関数データ解析法とは、各個体や対象に対して、離散点で経時的・空間的に観測・測定された一組のデータを滑らかな関数として捉えて関数化して次元縮小(平滑化)を行うことで、その関数化データの集合から有効に情報を抽出する統計モデルを用いた手法である。

関数データ解析法は、情報損失量を最小限にするデータの次元縮小によりノイズに頑健であり、個体間で観測点数や位置が異なる場合でも、精度よく安定した推定量を得ることが可能となるといった利点を有する。

今回の目的である手話動作の特徴を捉えるには、手指と複数の部位における動作を同時に考慮する必要があるため、関数データ解析の中でも複数の種類の関数データのばらつきを同時に捉えてその特徴を捉えることのできる多変量関数主成分分析を用いる。これにより、手のひら、肘、肩の動作のように、観測部位が複数にわたる場合の動作を同時に扱って、その共分散を計算し、動作ごとの大きく異なる動きの違いを捉えることができる。

#### 3.2 多変量関数主成分分析に基づく手話分析

櫻田ら[7]の提案手法を用いることで、ばらつきと主成分という指標で手話の分析を行うことができる。ばらつきとは、ある部位がある方向にどのように時間変動したかを特徴抽出した際に、個体(学習者)の動きにどれほどの差をもつかを表した指標である。図 6 は「富士山」という単語において、左手指先の y 方向(左右)に対するばらつきの時間変動であり、黒線が学習者 10 人の部位の平均、赤線が被験者全体の平均との差を表している。この図から動作の約 50%時点からこの動作の振幅やタイミング、傾きが個体に

よって異なっていることが分かる。

また、各個体の各部位から得られた関数を多変量関数主成分分析することで得られる主成分スコアを用いることで、図7のようなお手本との類似度を視覚的にとらえることができる。図7は「富士山」という単語において、A~Jの個体が第1主成分、第2主成分のスコアがお手本にどれほど近いかを二次元グラフで表したものである。第1主成分は動作の平面方向において、第2主成分は動作の奥行においての特徴を示している。

また多変量関数主成分分析は個体ではなく単語を量ととらえることで、単語ごとの動きの複雑さ、つまり難易度を分析することも可能である。

以上の多変量関数主成分分析に基づく手話分析結果を提示する機能を現行の手話学習支援システムに実装した。図8は実際の多変量関数主成分分析の分析結果表示画面である。画面左のグラフは図7と同様に選択した単語において、A~Jの個体の各主成分スコアがお手本にどれほど近いかを二次元グラフで表したものである。真ん中の6つのグラフは両手のx(上下), y(左右), z(前後)各方向それぞれの平均動作に対するばらつきの時間変動を表したものである。右の二つの表は、システムで学習可能な単語の難易度を表にしたものである。

多変量関数主成分分析による手話分析方法を用いることで、どの部位がどれくらいお手本とずれているかがわかり、個体の特徴から個体の動作を評価することができ、どの単語の動きが複雑か学習の難易度を知ることができる。

したがって、多変量関数主成分分析を用いることで学習者の手話能力を定量的に比べることができ、より効率的に手話を学習することができると考えられる。

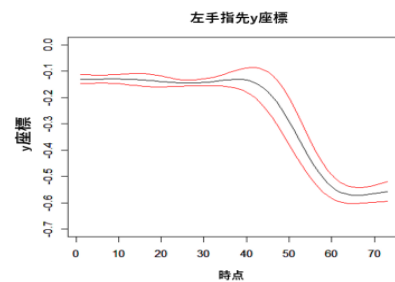


図6 「富士山」の左手指先y座標方向(左右)のばらつきの時間変動

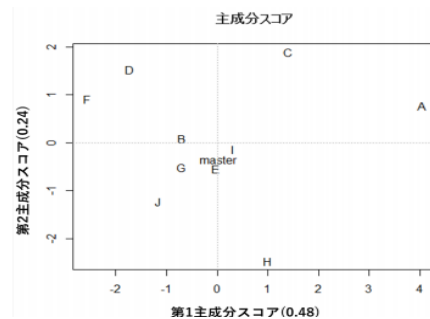


図7 学習者10名とお手本1名の「富士山」の主成分スコア類似度グラフ

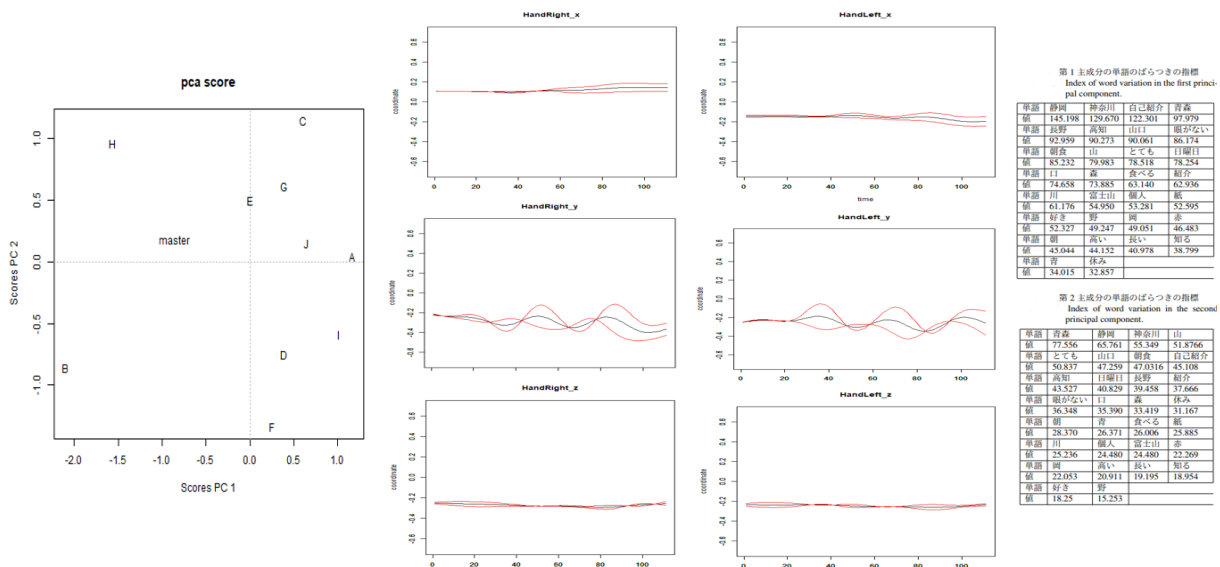


図8 多変量関数主成分分析の結果表示

## 4. 評価実験

### 4.1 実験条件

2章にて述べた従来研究の手話学習支援システム(以降

従来システム)と従来システムに多変量関数主成分分析の分析結果を実装した提案システムの比較実験を行った。

手話学習経験のない20代の学生14名に「自己紹介」という単語の手話の学習を行うために、まず両システムを5回使用してもらい、評価結果を踏まえて再度5回使用して

もらった(実験 1)。学習した単語への慣れを無くすためそれぞれの計測間に十分に時間を空けて、手法の順番を入れ替えて計測を行った。また手話学習経験のない 20 代の学生 4 名に単語数を 10 単語に増やした同様の実験を行った(実験 2)。単語は多変量関数主成分分析を用いて分析した難易度がばらつくように選出した。図 9 が選出した単語である。

どちらの実験も実験後にアンケートによる評価を行った。質問内容をまとめたものが図 10 である。アンケートの内容はシステムまたは機能についての 5 段階評価と、役にたった機能を選択してもらうもの、改善点や要望、気づいたことに対する自由記述である。項目(3)の分析結果のグラフ 1~3 は 3.2 節で示した図 8 の左のグラフ、真ん中の 6 つのグラフ、右の表である。

自己紹介	高知	森	口	食べる
富士山	好き	赤	長い	休み

図 9 実験 2 使用単語一覧

- |  |
|--|
| (1)統計分析によって動きはわかりやすくなったか<br>(2)各機能は手話の学習の参考になったか<br>(3)学習に役立った機能に○をつけてください<br>(4)各システムで今後も学習を継続したいか<br>(5)改善してほしい点や追加してほしい機能等があれば記述してください<br>(6)その他気づいた点があれば記述してください |
|--|

図 10 手話学習支援システムアンケート内容

#### 4.2 実験 1 の実験結果と考察

実験 1 の結果として、14 名の両手各方向における従来システムと提案システムの DP スコアの差を表 1 に示す。これは評価結果を踏まえたうえで、システムを 5 回使用してもらった結果の平均の値である。DP 照合によるスコアはパターン間距離であるため、値が小さいほどお手本と学習者の手話が似ているという結果になる。表の値は従来システムのスコアから、提案システムのスコアを引いたものである。よってスコア差の値が正であれば提案システムの方がお手本に近い手話が出来ているということになる。

スコアの差が負になる箇所が 0 である学習者が一人、1 である学習者が 7 人、2 である学習者が 6 人でありこのことから、提案システムの方が従来システムよりもお手本に近い手話ができていることがわかった。

学習者 G の右手上下、学習者 J の右手上下、左手上下の値がほかの部位と大きく離れた数値を示したため、評価結果を踏まえる前後の各記録の DP スコアとその軌跡付き動画を確認した。結果、学習者 G は従来システムにおいて評価の確認後 1 回目と 2 回目のみ DP スコアが大きかったが軌跡付き動画にてお手本と大きく異なった動きはしていなかった。これは Kinect の座標取得精度の問題だと考えられる。学習者 J は従来システムと提案システム間の差は大きい、それぞれ平均する前の各記録の値に大きな変動は見

られなかった。本実験では Kinect と学習者の距離等、実験環境が変わらないように実験を行ったが、学習者 J の実験の際に、細かな差異によって DP スコアの数値の基準値がずれてしまったのではないかと考える。

表 1 14 人の両システムの DP スコアの差(実験 1)

従来-提案	右手左右	右手上下	右手前後	左手左右	左手上下	左手前後
A	3.90	1.06	-0.52	2.26	5.70	2.90
B	0.32	1.56	1.22	-0.72	0.44	1.26
C	0.60	0.94	1.78	1.92	2.34	0.42
D	0.12	0.26	0.32	0.22	-0.54	1.52
E	0.56	-0.44	0.62	0.00	0.98	0.28
F	4.32	0.34	0.12	-1.36	-1.96	0.54
G	-1.02	14.04	2.00	1.38	7.08	0.00
H	1.66	1.40	2.32	-3.24	8.46	-0.20
I	1.74	4.74	-0.10	2.72	1.90	-0.50
J	4.20	15.46	-0.56	0.68	11.52	-1.86
K	0.12	0.28	0.52	0.64	0.42	-0.60
L	0.56	0.84	0.30	1.26	-3.78	0.62
M	1.06	2.84	0.36	-0.12	1.28	0.12
N	0.22	-0.64	0.12	-0.96	0.58	1.04

#### 4.3 実験 2 の実験結果と考察

実験 2 の結果として、単語ごとの DP スコアの差について 4 人の平均を求めたものを表 2 に示す。単語の並びは上から難易度の高いものである。実験 2 では難易度が高いもの(自己紹介、高知など)は提案システムの方がスコアが小さかったが、難易度の低いもの(好き、赤、長いなど)は人によって差の正負がばらついた。大きな原因は二つ考えられた。難易度の低い単語は意味がシンプルで動きが単純なことが多く、従来システムの学習方法で十分な学習成果が得られること、統計分析の結果として出力される各時点ごとのばらつきのグラフは動作が途中で変化する二つ以上の手話を組み合わせて出来た複合語のような複雑な手話で大きな効果を発揮することである。これらの結果から、学習する単語の難易度に合わせた結果表示などの改善が必要ではないかと考えられる。

表 2 単語ごとの DP スコアの差の平均(実験 2)

4人平均	右手左右	右手上下	右手前後	左手左右	左手上下	左手前後
自己紹介	1.25	0.59	0.36	0.63	1.93	1.37
高知	1.32	-0.05	0.17	0.27	0.82	0.19
森	0.69	-0.12	0.11	-0.22	-0.30	0.52
口	1.13	1.59	0.41	0.16	1.27	0.40
食べる	0.44	2.10	0.50	0.47	0.71	1.03
富士山	1.41	1.86	0.29	1.39	2.07	0.09
好き	1.18	0.72	0.22	-0.13	-0.02	0.16
赤	0.72	-0.18	0.23	-0.36	-0.84	-0.19
長い	-0.50	0.35	0.11	0.13	0.10	-0.17
休み	-0.20	-0.46	0.39	0.34	0.85	0.18

#### 4.4 アンケートの結果と考察

アンケートの 5 段階評価の結果を図 11 に、項目(3)の結果を図 12 に示す。項目(1)の 5 段階評価の平均は 3.8 であった。(2)の 5 段階評価で「軌跡付きの動画」、「DP スコア」、

「統計分析のグラフ」の3機能は3.6以上の評価がついた。項目(3)においても上記の3機能は6割以上の人が参考になると答えた。以上から統計分析を含めたシステムが学習に有効だと考えられる。しかし、(4)の5段階評価はどちらも3.5以上の評価がついたが、従来システムの方が評価が高かった。これは(5)、(6)の自由記述にて、特に提案システムのUIや結果の表示方法等に複数の改善要望が記述されておりこれが原因と考えられる。アンケートの項目(5)、(6)の自由記述の意見をまとめたものを図13に示す。これらの結果から、必要性・有効性が高いものを見やすく配置するなどといった改善が必要であると考えられる。

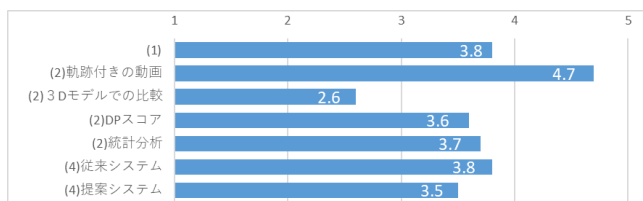


図 11 (1), (2), (4)のアンケート結果

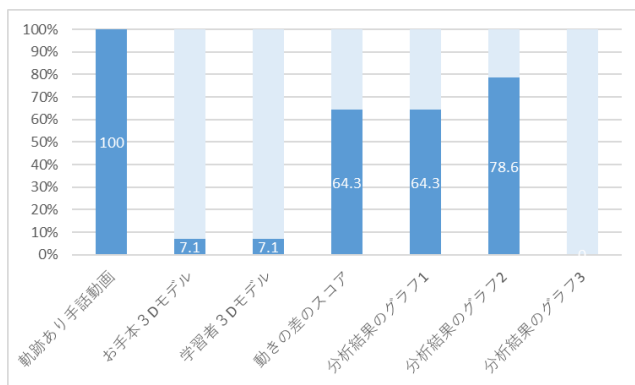


図 12 (3)のアンケート結果

3Dモデルがあまり参考にならなかった  
3Dモデルは改善するか無くしたほうが良い  
x, y, zの軸をもっとわかりやすくしてほしい  
手の位置が分かっても指の動きまで評価されないのが気になる  
録画している最中に自分の動きを確認できる機能  
動画中の軌跡の差分を表示する機能  
3Dモデルの使い方が分かりにくかった  
数値やグラフだけでなく直感的に分かるように言語化や画像で表示する機能  
動画が重なって軌跡等が見えなくなったほうが良い  
グラフ2がばらつくのはわかるが「いつ」「どこ」がわかりにくい  
統計分析を動きに反映しにくい  
3Dモデルにリプレイ機能が欲しい。もともと注意したほうが良いことなど動作のコツ等は説明があると便利  
統計分析結果の説明が分かりやすかったためそれも分析結果から読み取りやすくなると嬉しい  
もう少し図が大きければ見やすい  
masterとの比較が分かりやすかった  
動画とずれのプロットを同時に見たい  
手が隠れないことがあったり手がどこかへ飛んでいくことが何度あったため結果がぶれた

図 13 項目(5), (6)の意見まとめ

## 5. おわりに

本研究では従来研究である Kinect と 3D モデルを用いた手話学習支援システムに多変量関数主成分分析を用いた手

話分析機能を実装することで、どの部位がどれくらいお手本とずれているかがわかり、個体の特徴から個体の動作を評価でき、どの単語の動きが複雑か学習の難易度がわかり、学習者の手話能力を定量的に比べることができる手話学習システムを提案し評価実験を行った。評価実験を行った結果、単語の難易度が高いものに対して手話分析機能は有効であるという結果が得られた。アンケートの結果、提案システムが学習に有効だという評価が得られ、特に「軌跡付きの動画」、「DPスコア」、「統計分析のグラフ」の3機能が高い評価を得たが、改善点や課題があることも示唆された。

本システムの課題として 3D モデルの改善、Kinect の誤認識時の対応、分析結果のより理解しやすい提示方法の検討、および UI の改善などが挙げられる。また、聴覚障がい者によるシステム評価も今後の課題である。

上記の問題を解決し、より学習効率の高い手話学習支援システムを構築していく必要がある。また、現システムは語彙の習得のためのシステムにとどまっているため、実際に手話を使った会話を行うといったより実践的な手話の学習を行えるようなシステムへの拡張も検討していく必要がある。

## 参考文献

- [1]特別支援学校小学部・中学部学習指導要領：<[https://www.mext.go.jp/a\\_menu/shotou/tokubetu/main/1386427.htm](https://www.mext.go.jp/a_menu/shotou/tokubetu/main/1386427.htm)>(2021年1月25日取得)。
- [2]鎌田一雄，西堀義仁，西本哲也“聴者のための手話学習支援システムに関する一考察”，映像情報メディア学会技術報告，MIP99-76，1999。
- [3]伊藤奈美，上山輝“Webカメラを用いた手話学習支援システムの評価と改良”，富山大学人間発達科学部紀要，第11巻，第3号，pp. 51-58，2017。
- [4]高林大輔，大川泰弘，瀬戸山浩平，田中陽士，加藤伸子，岡崎彰夫，福井和広“フィードバック機能を備えた指文字練習システムの試作”，電子情報通信学会技術研究報告，Vol.112，No.483，pp.79-84，2013。
- [5]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，vol.2016，No.5，2016。
- [6]和泉勇希，西田昌史，綱川隆司，西村雅史“Kinectと3Dモデルを用いた手話学習支援システムの構築”，情報処理学会，第81回全国大会2019。
- [7]櫻田京之介，荒木由布子，和泉勇希，西田昌史“手話データの多変量関数主成分分析，ヒューマンインタフェース学会論文誌”，Vol. 22，No. 4，pp. 1-10，2020。
- [8]USB フットペダルスイッチ <<http://route-r.co.jp/?p=3102>>，(参照2021年1月27日取得)。
- [9]米川明彦“日本語・手話辞典”，全日本聾唖連盟。
- [10]米川明彦“新日本語・手話辞典”全日本聾唖連盟。
- [11]Unity：<<https://unity.com/>>，(参照2021年1月28日取得)。
- [12]中村薫，杉浦司，高田智広，上田智章“KINECT for Windows SDK プログラミング Kinect for Windows v2 センサー対応版”，秀和システム。