

## 類似手話語彙の平面データによる光学的識別法 (2)

## —平面データと立面データの比較—

## The Recognition of Plane Data of Similar Signs (2): Comparison of Plane and Cubic Data

神田 和幸<sup>†</sup> 木村勉<sup>†</sup>  
Kazuyuki Kanda Tsutomu Kimura

## 1. はじめに

神田(2019b)<sup>[1]</sup>では、考察において、伝統的な手話学の枠組みである「手話の音素」による類似手話の成分分析と OpenPose<sup>[2]</sup>による光学的なデータ分析が対比された。これまでの手話の光学的分析は手話の音素を先験的情報として画像処理をしており、モーションキャプチャ(mocap)を用いた分析も同様であった。このため手型データ取得に拘り、大がかりな仕組みと設備を用意し、人的資源と高額資金を要してきた。それはそれなりの研究価値をもつと考えられるが、反面研究者数が限定されるという弱点があった。

OpenPose による分析は比較的安価で誰にでも利用できるため、研究者数の増加が見込まれる。その際、手型に拘る必要がなく、さらに画像の奥行きもある程度無視できるのであれば、利便性は高くなる。そのためには旧来の手話の音素のような枠組みに拘ることなく、新たな枠組みを設定することで、より簡便な分析方法を確立できるのが望ましい。

枠組みは全く新規に創造するのは困難で、ある程度、経験から導き出されるものである。経験が枠組みの弱点を見つけ出すのである。

本論では、同じ手話語彙について、過去に mocap によるデータも取得しているの、上記論文と全く同じ手法による分析を試みた。手話モデルは同一人物だが、時期が異なるので、違いがあることは確かだが、熟練の手話者であるため、誤差は少ないと予想される。

## 2. 類似手話語彙

高橋他(2019)<sup>[3]</sup>で識別率が低かったとして指摘しているのは、<兄><姉>と<妹><娘>である。神田(2019b)では<兄><姉><弟><妹><息子><娘>を比較した。従来の手話学の枠組みでは、この6手話語彙の成分は次のようになっている。(表1)

しかし光学的分析では手型の測定が難しいことがわかっている。OpenPose による分析では関節の動きのみに着目するので、位置についても上記のような分類はできない。本論の mocap による測定でも、位置の空間指定はできるが、身体部位や中立といった意味をもつ位置の指定はできない。動きについても物理的な軌跡と移動時間が測定できる点である点は OpenPose と変わらないが、z 軸方向の測定ができる点が異なる。動きのみに着目する手法については、神田(2010)<sup>[4]</sup>、神田・木村(2019a)が説明している通り、動きから検索することで速く目標に達することがわかっているからである。

神田(2019b)との比較のため、まったく同じ手順で mocap データを分析するのが本論である。

## 2.1 &lt;姉&gt;の全データによる外観

OpenPose と mocap では取得データの関節数が異なる。また mocap ではカメラ数によりロストすることも多い。本論

表 1 手話語彙の成分

語	手型	位置	動き
兄	男	中立	上
姉	女	中立	上
弟	男	中立	下
妹	女	中立	下
息子	男	腹	前
娘	女	腹	前

で用いる mocap は6台のカメラにより 21 個の光点を観測するシステムなので、ロストデータもそれなりにある。光点は次の 21 箇所になる。

光点：頭上、頭右、頭左、右肩、左肩、右肘、左肘、右小指、右親指、右中指、右手首、左小指、左親指、左中指、左手首、右腰、左腰、右膝、左膝、右爪先、左爪先

全観測データをも csv データに変換したので、その時間的変位グラフを作成したが、本論では紙幅の都合で省略する。

ここで手話のホームポジションについて説明しておく。手話モデルの録画では、手話単語の最初と最後で、両手をだらんと下げるタイプ(A)と両手を腹の前で組むタイプ(B)がある。プロの手話録画や通訳場面では主としてBタイプが用いられるが、実験として多くの被験者に真似してもらった場合はAタイプの方が切れ目が認識しやすいこともあり、Aタイプが用いられる。本論ではBタイプのモデルを使用した mocap データを解析するが、OpenPose 解析の神田(2019b)ではAタイプが用いられた。

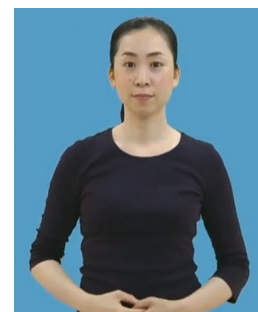
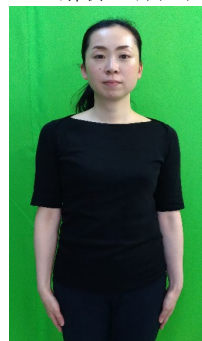


図 1 ホームポジション A 図 2 ホームポジション B

ホームポジションの違いは片手手話の左手の動きに顕著に表れる。Aタイプでは左手はほとんど動かないが、Bタイプでは左手も個人によって無意識に動く場合もある。

まず<姉>の利き手(右手)の手首と肘の動きを解析する。<姉>は図3に示すように利き手の小指を上げる動作である。この動きを図4から図6に示す。縦軸はホームポジションからの相対座標位置、横軸はフレーム数(120 FPS)である。



図 3 <姉>の動作

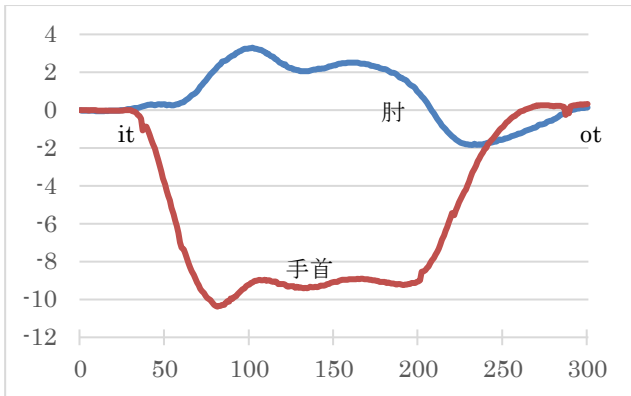


図 4 姉 X 軸

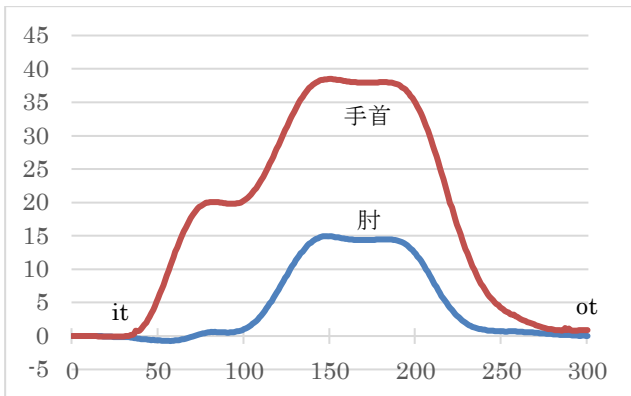


図 4 姉 Y 軸

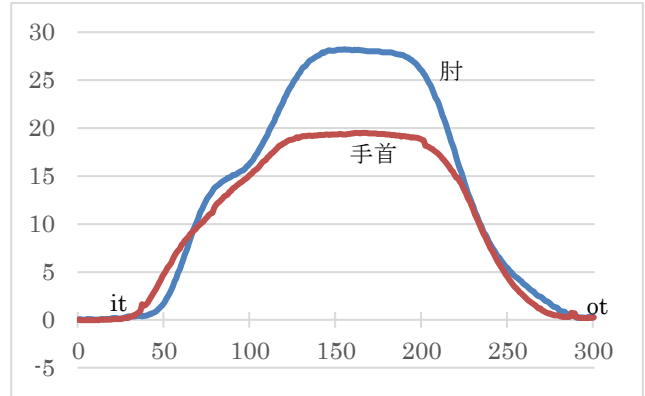


図 5 姉 Z 軸

図中の it, ot は手話動作の開始 (入り渡り: it=in-transition)と終了 (出渡り: ot=out-transition)を示している。

特徴的なのは図 4 の右手首 X 軸はマイナス方向 (右, 外方向) に動いていることである。右肘はそれに連動してプラス方向 (右, 内側) 動いている。Y 軸は手首も肘もプラス方向なので, 右腕が上方方向に動いていることが読み取れる。ホームポジションは体の中心に手を置いているので, いわゆる中立空間という胸前の右手を自然に上げる位置に手首が移動したことを示す。運動学的にいえば, 右肩を中心とした外旋運動をしつつ, 前挙していることになる<sup>10)</sup>。右手首は 2 段階の動きをしており, 最初上方に移動し (肩の外旋運動), 一旦静止した後, 再上昇 (前挙) し, 最後に元に戻っていることが読み取れる。これは一度小指を立てて前に置き, その後, 持ち上げるという 2 段階動作という動きを示している。

Z 軸は肘が一旦前に出て, 静止し, さらに大きく前に出ていることが読み取れる。この前後の静止運動は X, Y には反映されないので, z 軸方向の動きが微妙な動きを測定するのに必要だといえそうだが, 意識運動というよりは手首の移動に伴う肘の付随運動と見ることができそうであり, 運動学的知見を活用すれば, 肘の動きは推定できそうである。

## 2.2 兄姉の比較

<兄>と<姉>の動きを比較する。<兄>は手型が中指を立てている以外は, <姉>と同じ動作である。

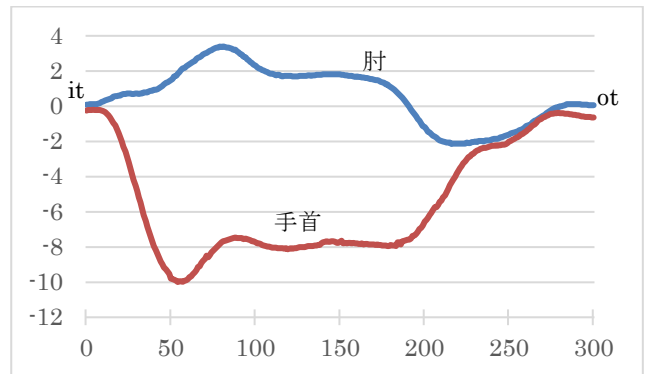


図 6 兄 X 軸

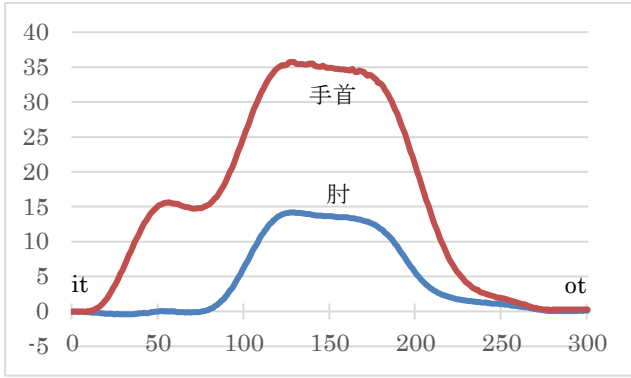


図7 兄 Y軸

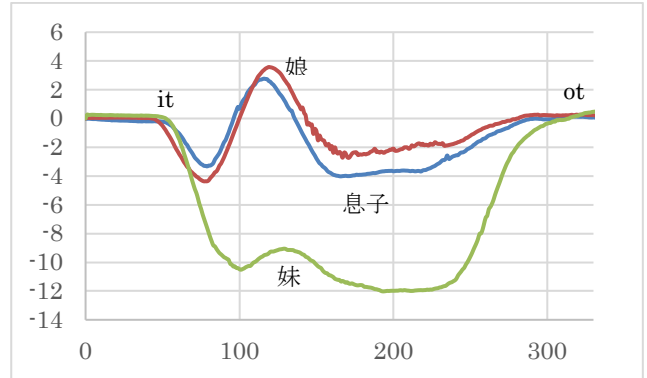


図9 息子・娘・妹 手首 X軸

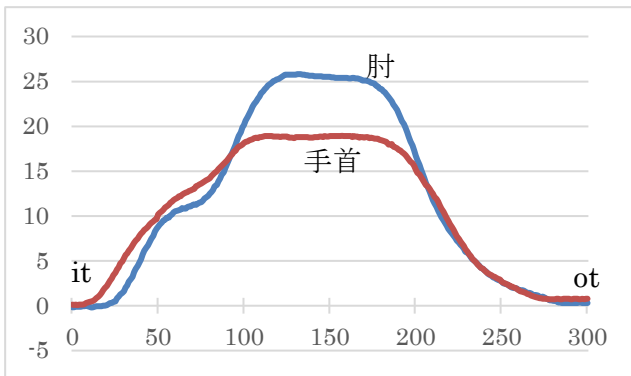


図8 兄 Z軸

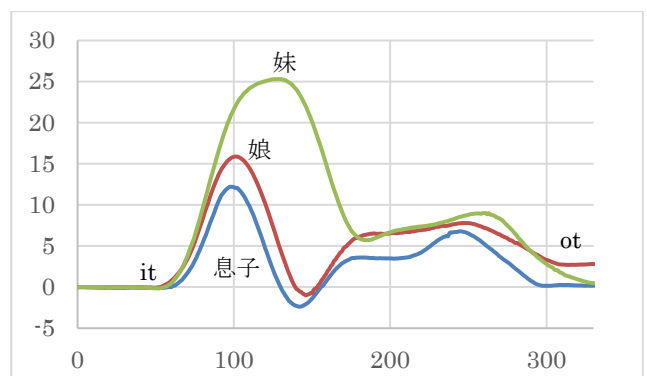


図10 息子・娘・妹 手首 Y軸

＜姉＞と＜兄＞を比較すると、予想通り差はほとんど見られない。予想されたことだが、＜姉＞と＜兄＞の動きは全体から見ると、ほとんど違いはない。動きの大きさにわずかに差が見られる程度である。

以上の結果を考察すると、＜姉＞と＜兄＞の動きはほぼ同一であると結論づけられる。

### 2.3 息子娘妹の比較

次に＜息子＞＜娘＞＜妹＞の比較をする。＜妹＞は＜姉＞が、腕が上に動くことに対して、腕を下に動かす動作である。また＜娘＞は、＜妹＞の動作と似ており、下に動かしたあと、前に出す動作が付け加わる。＜息子＞は＜娘＞と同じ動作で手型が異なる（親指だけを立てる）。

高橋他(2019)では以下のように述べている。

右手首の違いを表したのが図5（原著図：筆者注）である。（中略）ここで見られたのはyに一番大きな差があることとxのずれとyのずれが不一致であることである。（中略）次に＜息子＞＜娘＞＜妹＞の3者の右手首の動きを比較したのが図6（原著図：筆者注）である。（中略）＜息子＞＜娘＞と＜妹＞ではyもxもパターンが異なる。＜息子＞と＜娘＞は同じグループに属し、＜妹＞は別のグループに属することがわかる。（下線部筆者）

この先験情報を元に、右手首の動きに特化して検証する。

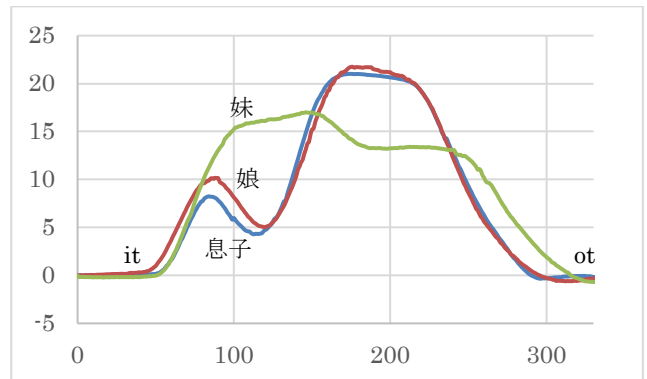


図11 息子・娘・妹 手首 Z軸

図9、図10、図11を比較すると、Y軸の動きは3者ともに似ている。＜妹＞の動きがやや大きい。X軸とZ軸の動きは＜息子＞と＜娘＞はほぼ同じだが、＜妹＞の動きは異なる。

＜息子＞＜娘＞では全軸の動きの違いはほとんどない。＜妹＞ではx軸とz軸の動きは同期しており、x軸はマイナス方向にやや大きく、z軸はプラス方向に大きく動く。つまり＜妹＞では手首を前に出していることが読み取れる。

ここで冒頭の表1をみてみると人間の認識では、＜妹＞の手の動きは中立位置から下に動くであるから、y軸が大きく動く意識があり、x,z軸の動きは認識されていない。ところが機械計測はx,z軸の動きに特徴が表れ、とくにz軸方向の変化が特徴的である。表2に改訂版を示す。

表 2 手話語彙の成分 (改訂)

語	手型	位置	動き
妹	女	中立	下
息子	男	腹	前
娘	女	腹	前

<息子><娘>では中立位置から腹位置までの移動は *it* として認識されず、前に出すという *z* 軸方向が強く意識され、*y* 軸の方向は付随的に意識される。機械認識では *y* 軸の動きは腹までの移動の無意識変遷が示され、*x, z* 軸が同期しており、*z* 軸が前方向の動きとして計測されるのは人間の意識と同じであるが、人間は *x* 軸方向の動きは認識していない。

<妹>では人間は *z* 軸前方向の動きをまったく意識されないが、機械計測ではその動きが計測された。

ここでは人間の意識と機械計測の差が明確に示された。*z* 軸方向の動きは *mocap* でないと厳密な計測は困難かもしれない。しかし *OpenPose* による 2 次元座標データから、運動学的知見を活用すれば、*z* 軸方向の推定はある程度可能なのである。

### 3. 考察

本論では次の仮説を検証した。

#### 仮説 片手手話では手首に変化が大きく表れる

本論での検証の結果、手話の動きには特徴的な成分があり、その成分をコアにすることで、認識率を高める可能性があることが示された。とくに本論が対象とした片手手話においては右手首の動きに特徴が出ることを示された。

本論での検証の利点は、*mocap* のような壮大な装置を必要としない方法を示唆する点である。また被験者がとくに特別なセンサを装着する必要がない、いわばウェアレスである点に大きな意味がある。さらに類似のしくみを利用した他の研究<sup>[6]</sup>において、50%と認識率が低いのは手型の認識を採用していると推測される。その原因は伝統的手話学の枠組みである手話音素が手型、位置、動きであることを援用しているからであろう。

本論での検証において、人間の認識と機械計測にギャップがあることが示された。これは音韻 (音素) と音声のギャップとの類似性があるのかもしれない。本論の研究者は新たな枠組みとして新手話学を提案しているが、そこでは音素に代わる像素という運動を核とした要素を提案している。本論で引用した高橋(2019)では運動に焦点を絞ったことで認識率が高くなった原因と考えている。音素は人間の認識を基礎として、対立と相補分布を核として設定されている構成素であるから、物理的計測と差が出るのは必然である。音声言語研究では、調音音声学と音響工学という認識と物理現象をつなぐ分野の発達により、示差的特徴などのより精密な構成素が考案された。手話言語研究に当てはめれば、運動生理学と運動工学がそれに該当するが、示差的特徴に相当する構成素は確定していない。分節音韻論の応用がそれに近いといえるかもしれないが、工学的研究とのリンケージ的研究は未見である。

本論の次の段階としては、運動の特徴点に焦点を当てることで、解析速度を上げることと、大量のデータによる分析である。

本論では特徴点として右手首に注目し、ここに重みをつ

けることで効率のよいデータ解析ができるのではないかと、ということを示した。

本論の着想は高橋(2019)の弱点を補強するという視点であるから、検証対象の手話を認識率の低かった語彙に限定している。そのため手話全般に敷衍できるかどうかは、他の語彙についても検証を要するであろう。

本論では *z* 軸 (前後) 方向の重要性も示された。本論では *OpenPose* による *xy* 軸データと *mocap* データの比較であったが、*z* 軸による分析も必要性も暗示している。この方法は実験室のような環境で録画と同時に分析する場合には有効だが、社会実装する段階や、既にある大量のビデオ画像を活用するには適用できない。新たに 2 カメラによる大量の手話データをとることは容易ではない。2 次元座標データから 3 次元データを推定する手法の開発が望まれる。それには運動生理学的知見を取り込む必要がある。

本論の結果は伝統的手話学の枠組みを援用した手話の工学的分析は非効率であることも副次的に証明したともいえる。本論で活用した *mocap* や *OpenPose* といった最近の技術を活用するには、伝統的手話学の枠組みである手話の音素に拘らず、新手話学の像素の概念利用が有効であることも示されたと考える。

像素については既に発表済みなので詳細は避けたが、本論が示したように右手首を動体、動きを軌跡として考えることの利点が活用できるであろう。像素の構成素の 1 つである速度についての議論は本論では避けたが、本論で用いたデータから、速度を導き出すことは容易である。その分析は今後の課題としたい。

#### 謝辞

本論は文部科学省科学研究費補助金課題番号 18K18518 (代表者神田和幸) 及び課題番号 18K18517 (代表者木村勉) による研究成果の一部である。

#### 参考文献

- [1] 神田 和幸, 木村 勉, “描素の構造—新手話学の演繹的アプローチ”, 電子情報通信学会技術研究報告, Vol.118, No.440 (2019b).
- [2] *OpenPose*, <https://github.com/CMU-Perceptual-Computing-Lab/openpose> (2019年6月8日閲覧)
- [3] 高橋 佑太, 木村 勉, 神田 和幸, “機械学習を用いた手話認識に関する研究”, 電子情報通信学会技術研究報告, Vol.118, No.440 (2019).
- [4] 神田和幸, 手話の言語的特性に関する研究—手話電子化辞書のアーキテクチャー—, 福村出版, 2010
- [5] 動作を支える骨と筋肉、関節 | 動作のしくみから理解する (1), <https://www.kango-roo.com/sn/k/view/1903> (2019年6月19日閲覧)
- [6] グラフニューラルネットで手話通訳も AI にお任せか, <https://ai-scholar.tech/image-recognition/graf-ai-75/> (2019年6月19日閲覧)