〈一般研究課題〉　手話言語習得に効果的な
メタ言語的視覚フィードバックに関する研究

助 成 研 究 者　名古屋工業大学　　酒向　慎司

手話言語習得に効果的な
メタ言語的視覚フィードバックに関する研究

酒向　慎司
（名古屋工業大学）

A study on metalinguistic visual feedback for sign language learning

Shinji SAKO
(Nagoya Institute of Technology)

Abstract：

Sign language is a form of visual language and a natural language. As such, it is not as easy to learn as a foreign language and effective learning methods are not considered to be well established. On the other hand, there are several initiatives and research findings to support language acquisition, mainly in spoken languages, which show the importance of corrective feedback presented to the learner. In one of the few research cases on sign language acquisition support, Han et al. confirmed the effectiveness of capturing the body movements of a sign language learner and visually presenting the correct sign and a video of the sign performed by the learner using a CG avatar [1]. This is classified as an explicit correction type of corrective feedback and is important for learning the basic of language because it provides input that clarifies the correct use of the language. On the other hand, feedback classified as output-mandated feedback, which encourages the learner to self-correct errors, is said to further enhance learning efficiency, and a proper balance between input-mandated and output-mandated feedback is said to be important. However, in the case of sign language learning, the identification of linguistic errors and their content, and how to present them to the learner have not been sufficiently studied.

1. はじめに

手話は視覚言語の一種であり自然言語である．そのため，外国語と同じように習得は容易ではなく，効果的な学習法についても十分に確立されていないと考えられる．一方で，言語習得を支援する研究は，音声言語を中心に様々な取り組みや知見があり，学習者に提示する訂正フィードバックが重要であることが示されている．

手話言語の習得支援に関する数少ない研究事例の一つとして，Hanらは手話学習者の身体動作を取得しCGアバターによって正しい手話と学習者の行った手話の映像を視覚的に提示することの効果を確認している[1]．これは訂正フィードバックのうち，明示的訂正に該当し，言語の正しい使い方を明示するインプット提供型に分類され，基礎学習に重要である．これに対し，学習者自身に誤りの自己訂正を促すアウトプット強制型に分類されるフィードバックは学習効率をさらに高める効果があるとされ，インプット提供型とアウトプット強制型の適度なバランスが重要であるといわれている．しかし，手話の習得においていうと，言語的な誤り箇所やその内容を特定することや，それをどのように学習者に提示するかについてはまだ十分に研究されていない．

そこで本研究では手話学習者の手話映像に対する手話認識技術を応用し，手話として適切ではない表現に対してアウトプット強制型のフィードバックであるメタ言語的視覚フィードバックを自動生成して学習者に提示することにより，手話学習者の言語習得効率を高める支援システムの要素技術として，手話映像の3Dトラッキングと自己投影アバター生成と手話の口型認識について取り組んだ．

2. 研究課題

2.1 手話映像の3Dトラッキングと自己投影アバター生成

手話の映像に対してOpenPoseやmediapipeなどに代表される身体の2Dまたは3Dトラッキング技術を利用することで手話学習者の身体動作をリアルタイムで取得することができる．三次元トラッキングされた身体特徴軌跡をもとに3D-CGモデルを動かすことでリアルな手話アニメーションを生成できる．手話の学習支援という観点では，自由な視点から手話の映像を参照できることなどの優位性があるが，自分の行った手話が第三者の外観で再現されることの違和感が指摘されている．過去の研究でも，本人らしい3D-CGアバターが学習支援に効果的であることが示唆されている．

そこで，撮影された手話映像から高精度な3Dトラッキング技術により抽出した身体動作データとともに，同じ人物を撮影した画像データから3D-CGアバターを生成し，これらを組み合わせることで撮影された人物の外観を持つ手話CGアニメーションを生成する手法を検討した．

2.2 口形認識技術に基づいた手話の口形認識と分類

手話の構成要素には，手指の動きや位置を表出する手指信号と，表情・口型・頷き・視線の動きを表出する非手指信号が存在する．後者の一つである口型は手指表現とともに特定の語彙を日本手話の口型には音声言語由来のマウジングと手話独自のマウスジェスチャが存在するといわれ，その他にも手話中では表情に付随した口の動きもみられる．

日本手話におけるマウジングに関する研究によると，マウジングの役割は，固有名詞・数字・同じ動きで異なる意味を複数持つ同形異義語などを表出する際に，それらを区別するために手指信号と併用して表出するものとされている[2,3]．しかし，マウジングの働きは不明確な点もあることや，本研究の予備調査により，実際の手話では音声の表記とは異なり様々なパターンで表出される

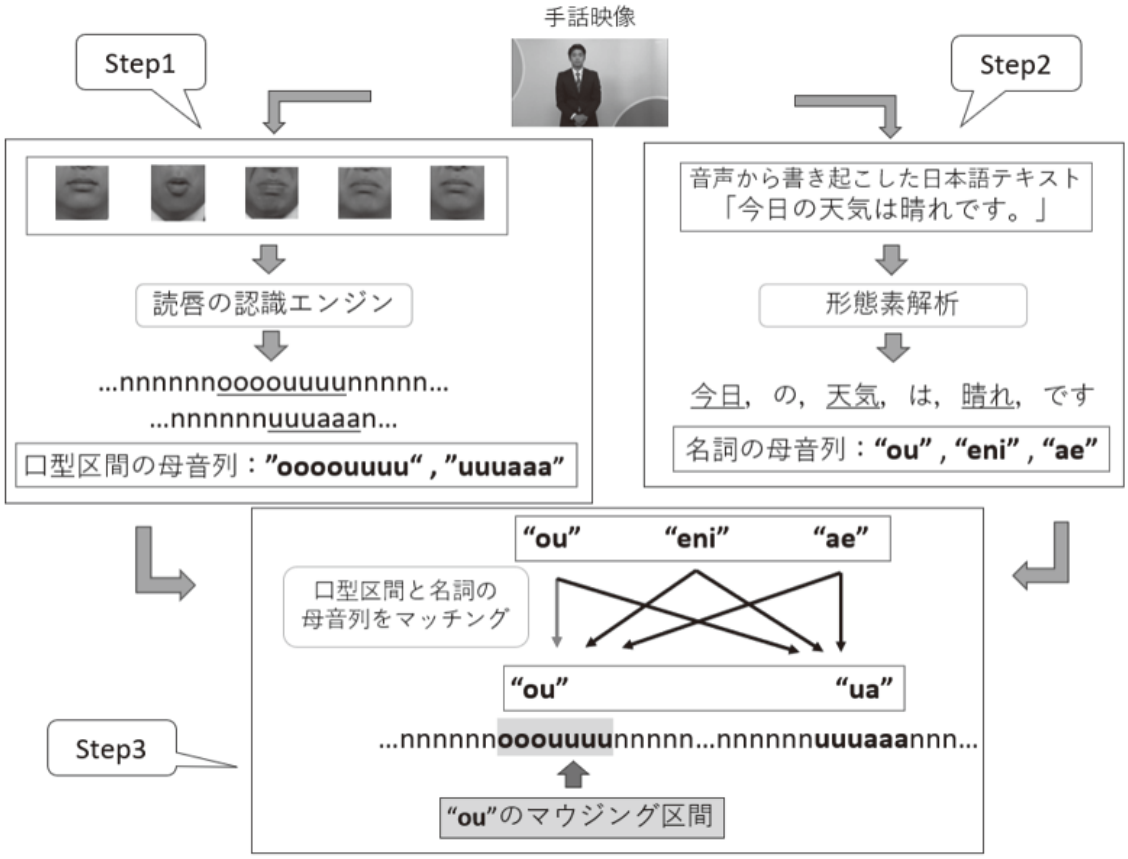

図1: マウジング検出手法の概要図

ことが分かり，その表出のされ方も不規則で一般化が困難であった．そこで2つ目の課題として手話の映像からマウジングを検出しフィードバックする手法について検討する．

本研究のマウジング検出手法の流れを図1 に示す．に変化がない区間は口型が現れていないと仮定して排除した結果，残った区間を口型が出現した区間として検出する．

ここでは，一般的にマウジングが表出されるとする名詞を対象とし，対訳文の形態素解析の結果から名詞を抽出することでマウジングの候補となる語を特定する．また，本研究では実際に表出したマウジングに該当する語を検出し，該当する語に対してマウジングとして表出される可能性のあるパターンをマウジング候補として追加する．

さらに検出された口型区間から，マウジングに対応する区間を特定する．特定したマウジング候補の母音列に対して口型区間の認識結果の繰り返しの文字を削除した母音列のDPマッチングにより算出された類似スコアを算出する．類似スコアが閾値以上となった場合に，その口型区間を候補の語のマウジング区間とする．

3. 実験結果

3.1 手話映像の3Dトラッキングと自己投影アバター生成

近年は汎用的なWebカメラで撮影された動画像から人体の2Dポーズや3Dポーズをリアルタイムで推定するソフトウエアが利用できるようになっている．また，VTuberをはじめとして3D-CGによるアバターを生成する技術も普及している．

これらの既存システムを調査し，手話の訓練に有効なリアルタイムセルフフィードバックシステムを構築するため，カメラによって得られた映像からリアルタイムモーショントラッキングを行

い，学習者本人を模したアバターを動かすシステムをUnityで実装した．3Dモーショントラッキングには Webcam Motion Capture[4]を，3Dアバター生成にはin3D[5]をそれぞれ採用した．

図2: 合成された3D-CGアバターによる手話アニメーション

図2は，訓練時の手本となる別の人物が表出した手話動作に対して，手話学習者を模して生成されたCGアバターがその動作を行うようにアニメーション合成をした例である．このように手話学習者本人を模した3D-CGアニメーションが容易に生成でき，手話学習者本人の動作を3D-CGアニメーションとして再現できるほか，本人以外のお手本となる動作を模倣したアニメーションを生成可能となった．図1ではKoSign[6,7]に含まれる高精度に記録された手話モーションキャプチャデータを利用してモーションの再合成を行っている．このシステムを基盤として，手話学習システムにおいて学習効果の高い視覚的フィードバックについて検証を行うことが今後の課題となる．

現時点では，手話の表現のうち両手を組み合わせて行った場合や，手が前後に重なってオクルージョンが発生した場合には指のトラッキングに失敗する傾向があることが確認されている．モーショントラッキングの高精度化のほか，全身のトラッキングと手指のトラッキングに適したシステムを併用によって改善を図ることを検討している．

3.2 口形認識技術に基づいた手話の口形認識と分類

日本語の対訳付きの手話映像を収集し，手話中のマウジング検出の精度を確かめる評価実験を行った．本実験では，各手話者1 分程度の計10 分の映像を用いる．データセットには1 文章あたり約10 秒前後の66 文章が含まれる．さらに対訳文を用いて文単位に分割し，画像列の区間と日本語の書き起こしテキストとの紐付けを行った．また，正解ラベルとして，名詞を表すマウジングが現れたフレーム区間の開始および終了時点と母音列のアノテーションを行った．

口型区間として検出されたものに対応するマウジングを候補の語の中から探して，マウジング区間を特定したときの結果を評価する．評価指標には再現率，適合率，F 値を用いる．再現率は正解ラベルとして付与されたマウジングの数に対して検出に成功したマウジングの数の割合，適合率は，マウジングが検出された回数に対してマウジング検出に成功した回数の割合を表し，F 値は再現率と適合率から算出する．

口型が表出された区間を決定した結果を表1に示す．全体的には8割強のマウジング区間が含まれていた．しかし，マウジングの候補の語と口型が表出された区間の母音列のマッチングによってマウジング区間を特定した結果は，表2に示す通りどの閾値においても誤検出が多くなった．

マウジング検出の精度評価から，口型が表出された区間にはマウジング区間が含まれているにもかかわらず，マウジング区間特定では十分な結果が得られないことが分かった．そこで，検出されまた，マウジング区間の分析を行った．マウジングの正解ラベルと推定ラベルおよびその区間と，区間同士の重複割合と不整合割合，それらの類似スコアを調査した．まず口型認識の誤りがみられ

表1: マウジング区間検出の再現率・適合率

手話者	正解ラベル内のマウジング	検出された区間	検出されたマウジング区間	再現率	適合率
1	17	38	16	0.94	0.42
2	26	37	25	0.96	0.68
3	25	45	21	0.95	0.47
4	16	40	12	0.75	0.30
5	15	36	13	0.87	0.36
6	22	43	13	0.59	0.30
7	19	52	16	0.84	0.31
8	18	38	18	1.0	0.47
9	22	28	20	0.91	0.71
10	21	44	18	0.86	0.41
全体	198	401	172	0.87	0.43

表2: マウジング区間検出の再現率・適合率

閾値	正解ラベル内のマウジング数	検出数	検出成功数	再現率	適合率	F 値
0.3	198	904	70	0.354	0.077	0.127
0.4	198	503	39	0.197	0.078	0.111
0.5	198	345	18	0.091	0.052	0.066

た原因として，マウジングが小さい口の動きで表出した場合や，マウジングと同時に下を向く動作の影響が考えられる．次に区間が正確に検出できなかった原因として，マウジングの前後のマウスジェスチャや頷きによってマウジング以外の動きが認識結果に影響したと考えられる．また，対象外のフレームを削除したことにより異なるマウジングが一つの区間に含まれた場合や，同じ母音が続く場合やマウジングの後に口の形を維持する場合にマウジングの1 文字が20 フレーム以上にわたるものがあり，口型区間の検出方法を改善する必要がある．

まとめ

本研究では手話の効率的な習得を支援することを目的として，学習者自身に誤りの自己訂正を促すフィードバックは学習効率をさらに高める効果があることに注目し，そのようなシステムの実現に必要な要素技術について研究を行った．手話学習者が自身の手話を振り返るほか本来の正しい動作を確認する際に，本人を模した手話映像でフィードバックするために，手話映像の3Dトラッキングと自己投影アバター生成システムを実装し，Webカメラの映像からリアルタイムで3D-CGアニメーションを生成するシステムを作成した．

一方で，本研究では，日本手話のマウジング検出の手法の提案に加え，マウジングの表出の仕方の調査やマウジング検出を行う上での技術的な課題を明らかにした．マウジングの表出には規則性

がはっきりとしない変形や省略がみられることが分かった．提案手法によるマウジング検出ではマウジング以外のマウスジェスチャや頷きの動きの影響により誤検出が多く含まれていたため，十分な結果が得られなかったが，他の手指信号などの認識と統合することで信頼性を高めることができると考えられる．

今後の展開として，これらの技術を基盤として，手話学習システムにおいて学習効果の高い視覚的フィードバックについて検討を行い，学習効果について評価実験を行うことが考えられる．

参考文献

[1] Han Duy Phan, Kirsten Ellis, Alan Dorin, and Patrick Olivie, “Feedback Strategies for Embodied Agents to Enhance Sign Language Vocabulary Learning”, IVA ’20: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, No. 47, pp. 1–8, 2020. DOI:10.1145/3383652.3423871

[2] Mayumi Bono, Kouhei Kikuchi, Paul Cibulka, and Yutaka Osugi. “A colloquial corpus of Japanese Sign Language: Linguistic resources for observing sign language conversations.” pp. 1898–1904, 2014.

[3] 加藤 直人, 宮﨑 太郎, “手話ニュースコーパスの拡張: ニュースに出現する口型の分析”, 電子情報通信学会技術研究報告, Vol. 115, No. 193, pp. 47–52, 2015.

[4] in 3D: Create Photorealistic Avatars For Metaverse, https://in3d.io/ (accessed on 2024/5/31)

[5] Webcam Motion Capture, https://webcammotioncapture.info/ja/index.php (accessed on 2024/5/31)

[6] 工学院大学. 工学院大学多用途型日本手話言語データベース(KoSign), 2021. https://www.nii.ac.jp/dsc/idr/rdata/KoSign/ (accessed on 2024/5/31)

[7] Yuji Nagashima, Keiko Watanabe, Daisuke Hara, Yasuo Horiuchi, Shinji Sako, Akira Ichikawa, “Constructing a Highly Accurate Japanese Sign Language Motion Database Including Dialogue”, Communications in Computer and Information Science, pp.76–81, 2020.

[8] 梅田 唯花, 酒向 慎司, “読唇を用いた日本手話の映像データにおける口型認識”, 情報処理学会第86回全国大会, 4ZJ-01, 2024.