

手話 CG アニメーション生成における空間位置制御手法の評価 Evaluation of Placement Control Method for the Sign Language Avatar Animation

内田 翼[†] 箱崎 浩平[†] 金子 浩之[†]
Tsubasa Uchida Kohei Hakozaiki Hiroyuki Kaneko



図1 提案法による空間位置制御の例：「東京と神奈川に大雨特別警報」の手話表現について，同一の手話単語モーションを使用して従来法（上）と提案法（下）を用いて生成した手話 CG を示す．従来法は全ての手話単語を話者正面で表現するが，提案法は異なる地名の固有名詞を左右の手話空間に分けて表現し，固有名詞の境界を明確化できる．

1. はじめに

手話を母語とする聴覚障害者の情報アクセシビリティ改善に向けて，手話による情報保障サービスが増えている．手話は音声言語とは異なる独自の文法を持つ言語である．手話母語者の中には，母語ではない日本語のテキストで提供される情報の内容理解が困難な人がいる．そのため，放送において日本語テキストによる字幕のみでなく，手話による情報保障の拡充について要望がある．

放送事業者が手話による情報保障サービスを拡充するにあたって，ニュースの専門的な内容を適切な手話に翻訳して伝えることができる手話キャスターや手話通訳者の数は限られる．そのため，緊急時に備えてニュースの手話表現に精通する専門家を常に放送局に確保することも難しい．そこで，ニュースの手話コンテンツ拡充に向けて，Computer Graphics (CG) アニメーションで手話を再現する手話 CG の技術を用いて，日本語文から手話 CG を自動生成する任意文手話 CG 翻訳システムを提案している [1]．任意文手話 CG 翻訳システムにおける CG アバターの制御には，辞書形と呼ばれる手話の辞書に掲載されている動作の基本形でモーションキャプチャした手話単語モーションを補間接続して使用している．しかし，手話母語者への評価実験の結果では，自動生成した手話 CG において，動作速度や間（ま），空間位置の変化や表情といった手話の言語的な特徴（以降，本論文では手話の文法要素と呼ぶ）が再現できておらず，実写の手話と比べて不自然な動作となり正確に内容を理解できないことが示されている [2][3]．これは，手話文モーションを生成するにあたって辞書形の手話単語モーションを補間接続した場合に，文意に応じた手話の文法要素が再現できないことに起因している．特に，辞書形の手話は全て話者正面の同じ位置で表現されており，文意に応じて変化する手話の空間位置が再現できず，手話文中の複数の固有名詞や文節の境界が伝わりにくいことが

課題となる．手話 CG の利用用途であるニュースのような話題の場合，内容を理解するうえで文中に登場する様々な固有名詞が重要となるが，空間位置が再現できなければ固有名詞の読み取りに影響が出るため文全体の内容理解の低下につながる．以降，本論文では文意に応じた空間位置の変化を空間位置の活用と呼ぶ．

日本手話の空間位置の活用に関して具体例を挙げる．

「東京と神奈川に大雨特別警報」というニュースの日本語文を手話に翻訳する場合，手話話者の左側で {東京} の手話単語を表現し，右側で {神奈川} の手話単語を表現する¹．{東京} と {神奈川} の2つの手話単語をそれぞれ左右の異なる位置で表現することで，異なる固有名詞であることが明確になる．辞書形の手話単語モーションを接続して全て同じ位置で表現した場合，固有名詞の境界が不明確になる．例えば，日本語の「東京」と「東」は，手話では同じ動作で表現する場合があるため，「東神奈川に大雨特別警報」といった意味に誤認識される可能性がある．しかし，手話 CG で空間位置の活用を再現するために，手話単語モーションごとに全ての空間位置のパターンを事前にモーションキャプチャしておくことは困難である．

そこでニュースにおいて重要となる固有名詞を手話で正確に表現するため，本研究では，手話 CG の自動生成において課題となる手話文法要素の中から，空間位置の活用に着目して空間位置制御手法を提案する．モーションデータの水平方向の空間位置を文意に応じて左・中央・右に制御することで，複数の固有名詞を異なる空間で表現する手話の空間活用を再現する．先行研究[5]において提案した手話記述言語 Hamburg Sign Language Notation System (HamNoSys) [6]で定義される手話の構成要素に基づいて，日本語文の文意に応じた空間活用を再現するためのモーション制御手法を開発した．「東京と神奈川に大雨特別警報」の空間活用の例を，同一の手話単語モーションを使用して，従来法と提案法それぞれを用いて生成した手話 CG を図1

[†] 日本放送協会 Japan Broadcasting Corporation

¹ 本稿における表記について，日本語を「」，手話単語を表す日本語のラベル（手話グロス）を { } とする．

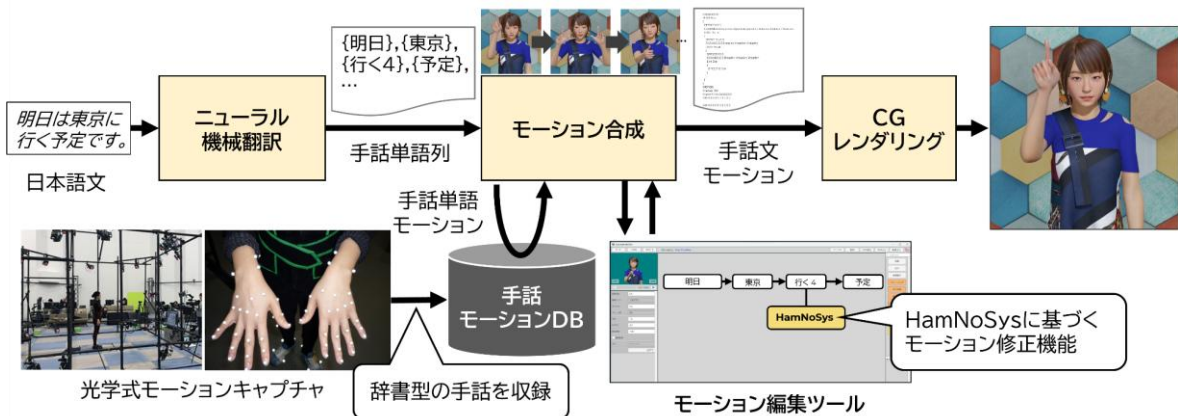


図2 任意文手話 CG 翻訳システムの概要

に示す。従来法は全ての手話単語を話者正面で表現しているのに対し、提案法は2つの異なる地名の固有名詞に対して左右の手話空間の使い分けが再現できている。

手話 CG アニメーション生成において、本研究で明らかにした項目を以下に示す。

(1) 提案法によって手話単語モーションを制御することで、追加のモーションキャプチャなしで文意に応じた複数の固有名詞の手話単語モーションの空間位置変更が可能となることを確認した。

(2) 評価実験を通じて、提案法により空間位置を制御することで手話 CG の内容理解に関する質問の正答率が向上し、日本語文の内容をより正確に手話 CG で表現できることを確認した。

2. 関連研究

2.1 モーションデータを用いた手話 CG 生成

CG アバターを制御して手話のCGアニメーションを生成する様々な手法が提案されており[7]、主に3つの代表的な手法がある。音声言語の音素に相当する手話を構成する最小単位である手話素のレベルで身体の動きを合成する手法、アニメーターによって制作した動作をつなぎ合わせるキーフレームアニメーションの手法、モーションキャプチャデータで CG アバターを制御する手法である。その中でも、実写の人間に近い自然な手話動作を再現できる手法が、手話話者の全身動作をモーションキャプチャしたモーションデータを使って CG アバターを制御する手法である[8][9][10][11]。

我々が提案する日本語文から手話 CG を自動生成する任意文手話 CG 翻訳システム[1]においても、モーションキャプチャデータを使用して CG アバターを制御している。システムの概要を図2に示す。日本語から手話への翻訳結果は、手話単語が表現順に時系列に並んだ手話単語列となる。手話単語列を基にモーションデータベースから各手話単語のモーションデータを読み込み、複数のデータ間を補間接続して手話文のモーションを生成する。手話文モーションを使って CG アバターを制御することで、手話 CG アニメーションを生成する。しかし、モーションデータによる生成手法はあらかじめキャプチャしたデータを CG アバター制御のために編集せずに使用する用途で用いられ、文意に応じて変形する手話動作を再現するための事後のデータ編

集処理が難しい。そのため、単語単位の手話動作の再現度は高いが、手話文を生成する場合に、文意に応じた手話表現を再現するための柔軟性の低さが課題として指摘されている[7][12]。1章で先述したように、視覚言語である手話には、傾きや指差し、動きの間や表情といった音声言語とは異なる様々な手話特有の文法要素がある。任意文手話 CG 翻訳システム[1]によって生成した手話 CG アニメーションにおいて、仮に手話文を構成する各手話単語と文全体の語順に誤りのない翻訳精度で手話単語列を出力しても、内容が理解できないという評価結果が示されている[2][3]。これは、手話単語モーションは基本的に辞書形でモーションキャプチャしたデータを使用しているため、文意に応じた手話の文法要素が再現できず内容が適切に伝わらないためである。

そこで、手話 CG 生成において様々な文法要素を効率的に再現するための研究が行われている。我々は、手話単語モーションを補間接続する際に、翻訳結果に応じて単語モーション同士の時間間隔や傾きの挿入タイミングを自動制御する手法[13]や、口型に特化したモーションデータを接続することで固有名詞の手話を読み取る際に補助的な情報となる読み仮名の口型を再現する手法[25]、手話記述言語 HamNoSys で定義される手話の構成要素に基づいてモーションデータを組み合わせることで手話の語形変化である Classifier (CL) 述語を再現する手法を開発している[5]。CL 述語は音声言語には存在しない概念であり、話者が手話文中の個々の物体の位置や動きを示すために、体の周りの空間と特有の手の形・動きを用いる言語現象である[14]。先行研究[5]において、HamNoSys の構成要素である手型と手首の回転に基づいて CL 述語を再現することで、手話 CG の内容理解に関する品質が改善することが示されている。しかし、その他の構成要素である手話の空間位置については、モーション編集手法の効果を検証できていない。

2.2 手話 CG における空間活用の再現

辞書形のモーションデータを組み合わせて生成した手話 CG において、手話文法要素である空間位置の活用を再現できないことが課題のひとつとして指摘されている[1][2]。特に空間活用に関しては、連続する2つの固有名詞がある場合に、それぞれの提示位置を話者の左側と右側の空間に変更するなど手指動作の変化量が大きい場合、要素が欠落していることが視覚的に分かり易い。そのため、手話 CG

の課題として指摘頻度の高い文法要素となっている [2][15]. しかし, 1 章で先述したように, 手話 CG で空間活用を再現するために, 空間位置を変えた全てのパターンの手話単語モーションを事前にモーションキャプチャし, 文意に応じて使い分けることは困難である. そのため, 様々な手話言語において空間活用を CG アバターで再現するための手法が検討されている[9][16][17][18].

先行研究[17]は, アメリカ手話の CG アニメーション生成において多用される空間活用である CL 述語に特化したシステムを提案している. 空間位置や手型, 動きなどの要素を組み合わせたテンプレートから手話アニメーションを生成する手法である. しかし, 人物や建物などの位置関係を状況説明するケースに特化し, 手話文中に登場するオブジェクトを手話空間に写像的に再現するアニメーションに限って評価検証されている.

先行研究[19]は, JASigning システムを用いた手話アニメーション生成手法を提案している. スイス・ドイツ手話の手話アニメーションによる列車の運行情報に特化した評価実験を実施し, 時刻や列車の番号といった数字が連続する手話表現において空間位置をずらすべきという被験者からの指摘を受けて, 定型的な表現に特化して数字モーションの空間位置を左右に変更する手法を提案している. こちらも, 再現できる空間活用表現は時刻を表す定型表現に限定されている.

先行研究[20]は, 手話グロスを使用しない手話表記法である AZee 表記を手話アニメーション生成システムと連結し, AZee 表記に含まれる空間内の手の位置の情報を使用して空間位置も考慮したアニメーションを生成する手法を提案している. しかし, AZee 表記とアニメーション生成をマッピングするための検討に留まっており, 空間位置の情報を基に生成したアニメーションの品質について評価されていない. 先行研究[21]も, AZee 表記に従ってアニメーターが制作したモーションで CG アバターを制御する手法を提案しており, 手話文中に登場するオブジェクトの組み合わせに応じて空間位置を変更する手法を提案している. 再現できる手話表現の汎用性が高い一方で, 生成したアニメーションの品質評価や音声言語から手話への翻訳システムとの連結が考慮されていない. また, AZee 表記は一般的な表記法ではないため手話のアノテーションに専門的な知識を要し, 音声言語から手話への自動翻訳システムを構築する際に, 翻訳の学習データ構築のコストが高いことも課題となる.

先行研究[22]は, 手指動作に加えて表情や視線といった非手指動作を考慮して CG アバターを制御する手法を提案している. 翻訳の学習データ構築時に, 顔などの情報に加え, 空間位置をアノテーションできる仕組みを取り入れている. 複数の身体部位のタイミングなどを制御して組み合わせるアニメーションデータパッケージを提案しており, 左右の手や口・表情を独立して制御できるが, 空間位置の変更に関する制御機能のシステム実装は今後の課題となっている.

先行研究[23]は, 手話単語モーションを手話の構成要素(位置・手型・動き)に分割し, 各要素を Inverse Kinematics (IK) を適用して変形させることで, 指差しや指示動詞・空間活用を再現する汎用的なアニメーション生成手法を提案している. しかし, 機械翻訳結果を考慮した

評価は行われておらず, 音声言語を基に生成した手話アニメーションにおける提案手法の有効性は示されていない.

また, 日本手話においても人や物, 出来事などを表す際に空間を文法的に使用することが知られている[4][14]. 先行研究[4]では, 日本手話の空間位置の活用に関する具体例として, 「手話話者の右側で {コーヒー} という手話単語を作り, 左側に {紅茶} という手話単語を作り, その後, 右側で {ない} という手話をし, 左側で {余る} という手話をした場合, 「コーヒーはなくなったけれど紅茶は余っている」という意味になる。」という用例を取り上げている.

先行研究[14]は, 手話表記法のひとつである jjs 表記法 (Japanese gloss-based Japanese Sign language notation) を用いた日本語から手話への機械翻訳において, 日本手話でも多用される空間表現である CL 述語に対応した手法を提案している. 翻訳時に日本語入力文から手話空間内の位置関係・動きの動作パスを生成し, パス上に手話単語や CL 手型を配置する手法である. しかし, 翻訳結果の手話を jjs 表記法でテキスト出力するに留まっており, 最終的な CG アニメーション生成までは考慮されておらず, その有効性は示されていない.

先行研究[11]は, 深層学習を用いたデータ駆動型の日本手話翻訳・生成システムの実現に向けて, 文法的な要素を記述した日本手話のモーションキャプチャデータのコーパスを提案している. コーパスは, 手話記述言語である SignWriting[24]や手話グロスとモーションデータがペアになったものである. 手話アニメーション生成のための高度なシーケンス生成モデルに簡単に適用できるように定義されており, 多くの空間的および時間的参照と非手指動作の情報が組み込まれているが, それらを用いた翻訳やアニメーション生成は実現できていない.

基本的にこれらの先行研究は, 用途や話題を限定しており, ニュースのような汎用的な話題に関する手話アニメーションの自動生成における有効性については評価検証されていない. 一方で, 固有名詞が重要となるニュースの日本手話表現においては, 日本語文に地名や人名, 組織名といった固有名詞が 2 つ以上含まれる場合に, 先行研究[4]の例のように話者正面の手話空間を水平方向に分割して固有名詞を提示し分けることで明確に区別する空間活用が多用される. そこで我々は, 同様の空間活用を手話 CG で再現するために, HamNoSys に基づくモーション編集手法[5]を応用し, 翻訳元となる日本語文中の固有名詞に対応する複数の手話単語モーションデータの空間位置を修正する空間位置制御手法を提案する. さらに, ニュースの中でも緊急性の高いニュース速報を対象とした手話 CG の評価実験により提案法の有効性を示す.

3. 提案手法

ここでは, モーションデータの水平方向の空間位置を文意に応じて左・中央・右に制御することで, 複数の固有名詞を異なる空間で提示する手話の空間活用を再現する手法について提案する. 提案手法の詳細を 3.1 節と 3.2 節で述べる.

3.1 固有名詞表現に特化した空間位置制御手法

辞書形のモーションデータを編集して手話の空間位置を変更するために、先行研究[5]の HamNoSys の手話空間位置の制御手法を応用する。図3に HamNoSys に基づく空間位置制御のインターフェースを示す。

先行研究[5]の位置制御手法では、HamNoSys で定義される CG アバターの身体各部位との相対位置を基に、水平・垂直・奥行の3つの位置の組み合わせで指定する対象座標に対して、IKによって手首関節の位置を移動することで手話の空間位置を変更する。辞書形の手話単語モーションに対して、単純に手首関節のみを対象として IK による位置変更処理を適用した場合、手首の回転を考慮できていないことで位置変更前の手指の形が保持されず、位置変更後の手話表現の意味が変わることになる。そこで、位置変更前の手首の回転の値を位置変更後の手首関節にオフセットとして適用することで、位置変更前後で CG 空間内のカメラに相対する手指の形を保持する。図4に HamNoSys に基づく空間位置制御の例を示す。図4の「家」の例のように、位置変更後は左手の手首の屈曲が大きくなり、位置変更前後で手話「家」の意味を表すうえで重要な屋根の形が保持されていることが分かる。また、両手を同時に使う両手手話と利き手のみを使う片手手話に応じて、位置変更の対象とする腕を両手と片手から選択することが可能である。

図3に示すように、HamNoSys の空間位置は細かく定義されており、胴体との相対位置については、垂直方向に Top / Shoulder / Breast / Belly / Abdominal、水平方向に Left to / Left / Center / Right / Right to、奥行方向に Contact / Close / Standard / Maximum となっている。任意文手話 CG 翻訳システムにおいて、手話文中の複数の固有名詞の境界を明確化するためには、各固有名詞を構成する手話単語モーションの空間位置を文意に応じて制御する必要がある。さらに、総務省が公開しているテレビにおけるニュースの手話通訳の事例集において、複数の企業名や政党・議員といった固有名詞の位置を左右横並びに分けて表現することが指導されている[27]。そこで、HamNoSys の定義に基づいてあらかじめ設定した左・中央・右の3種類の指定座標に対して空間位置を制御する、固有名詞表現に特化した処理を新たに開発した。具体的には、水平位置は Right / Center / Left の3つ、垂直位置は全て Top に固定し、奥行位置は辞書形の手話単語モーションの位置を編集せずに使用する。制御する空間位置を3種類に限定することで、システムで生成した手話 CG において、複数の固有名詞を区別して提示する際の動作がパターン化できる。「東京と神奈川に大雨特別警報」の例では、{東京}を左、{神奈川}を右に提示することで、2つの異なる地名を表現していることが明確になり、空間位置活用によるニュースの手話文の内容理解向上につながる。

3.2 空間位置変更による身体干渉回避オフセット処理

先行研究[5]において、辞書形のモーションデータを編集して空間位置変更した手話モーションにおいて、位置変更後に手指や腕が身体に干渉し、CG アバターが不自然な姿勢となる場合がある。位置変更前の辞書形モーションの肩の回転値が小さく脇を閉じているような姿勢の場合に、手首関節の位置を対象座標に移動することで、その姿勢からさらに脇を閉じるような処理が行われるためである。また、



図3 HamNoSys に基づく空間位置制御のインターフェース

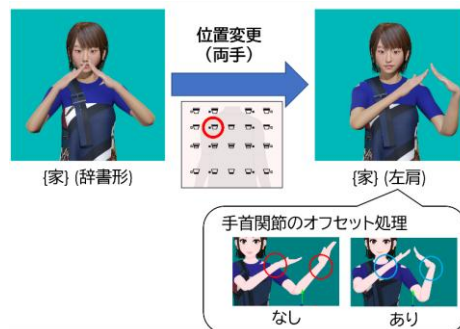


図4 HamNoSys に基づく空間位置制御例

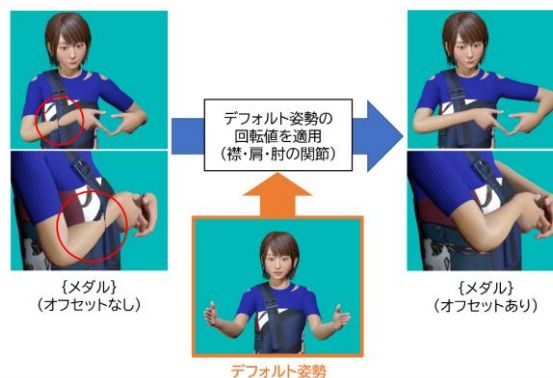


図5 身体干渉回避オフセット処理有無の比較（「メダル」の空間位置変更の例）

辞書形モーションの肘の回転値が小さく、元々の手の奥行方向の位置が身体に近い場合も、移動後に前腕部が身体に干渉し易くなる。IKによって手首関節の位置を移動する際に、各関節可動域に人体構造を考慮した解剖学的な最低限の制約を設けているが、CG アバターの体形や衣装などのボリュームは考慮できていないため身体干渉が顕著になる。この位置変更に伴う身体への干渉を解消するために、位置変更処理の前処理としてオフセット用のデフォルト姿勢をキープポーズとして挿入する処理を新たに追加する。身体干渉回避オフセット処理で使用するデフォルト姿勢と、オフセット処理有無の位置変更の比較を図5に示す。位置変更時の IK 処理適用前に、CG アバターの腕を構成する襟・肩・肘の3つの関節に対して、両肩を左右にやや広げ両手を前方に差し出したデフォルト姿勢（図5中央）における同一関節の回転値をオフセットとして適用する。両肩を広げることで位置変更によって脇を閉じた時の水平方向の干渉を軽減し、肘を曲げて両手を前方に出すことで手の位置が前方になり奥行方向の干渉を軽減できる。例えば、{メダル}を左に位置変更する場合、オフセット処理なしの場合



図6 評価した手話動画の例
左：実写，中央：提案法（位置変更あり），
右：従来法（辞書形）

合は、図5左のように右腕の前腕部分が胸の下に干渉してしまう。提案法によりデフォルト姿勢で両腕をオフセットすることで両肩が左右に広がり、図5右のように、位置変更後の腕の身体への干渉を解消することが可能となる。片手手話の空間位置変更の場合、本オフセット処理はデフォルト姿勢の片手の値のみを適用する。

4. 評価実験

提案法を用いて空間位置を変更した手話 CG について、内容理解の向上を確認するために、手話を母語とするろう者による評価実験を実施した。

4.1 実験概要

評価にあたって、実写映像（手話母語者による手話）・提案法（位置変更した手話 CG）・従来法（位置変更していない手話 CG）の3種類の手話動画を用意し、各手話動画について内容理解や自然さを確認した。被験者は手話を母語とするろう者21名で、年齢が偏らないように考慮して20代4名、30代4名、40代4名、50代5名、60代4名の年代構成とした。各被験者の手話の使用頻度については、毎日使用する人が18名、週2～3日程度使用する人が2名、ほとんど使用しない人が1名であった。

実験は全て対面で行い、進行役を担当するろう者が被験力者の目の前に設置したモニタに評価対象となる手話動画を提示する形で実施した。実験中の進行役と被験者の間の全てのコミュニケーションは手話で実施した。

4.1.1 評価した手話動画

評価にはニュース速報に関する手話10文を使用した。手話文の選定において、実写映像の手話表現を確認し、必ず固有名詞に対する空間位置の活用が使用されている文を選定した。ニュース速報には、災害／政治／芸能／交通の4つの話題が含まれる。各手話文に対して先述した3種類の手話動画（実写・提案法・従来法）を生成し、計30本の動画を評価に用いた。事前に手話母語者によって日本語文から手話へ翻訳した表現で実写映像を収録し、2種類の手話 CG 動画（提案法・従来法）についても、実写映像の手

表1 読み取れた内容に関する質問項目の例

ニュース速報文	「英保守党の新党首にトラス外相 ジョンソン氏の後任の首相に就任へ」
質問1	誰に関する話題ですか？
質問2	何人登場しましたか？
質問3	各登場人物の役職・立場は？

話表現に従って手話単語モーションを接続して生成した。本評価実験においては、3種類全ての手話動画において手話文を構成する各手話単語と文全体の語順が共通であるため、翻訳処理によって生じる各単語の選択と文全体の語順の誤りによる影響がなく、CG生成処理の違いに限定した評価結果を得ることができる。

評価した3種類の手話動画の例を図6に示す。ここで示した「村」と「関」という手話は、「新潟県の村上市と関川村に大雨特別警報 最大級の警戒が必要」というニュース速報を手話で表現した文の一部である。実写の手話母語者による翻訳では、「村上市」と「関川村」という異なる2つの地名が、それぞれ話者の左右の異なる位置で提示されている。従来法では全ての手話単語が辞書形のまま話者の正面に提示されるのに対し、提案法は実写のように位置を左右に変更して提示可能となる。

手話動画の提示に関しては、3つの提示パターン（X/Y/Z）を用意した。各提示パターンにおける手話動画は、3種類の手話動画全30本の中から手話文の重複のない10本ずつの組み合わせで構成した。例えば、提示パターンXは、実写4本、提案法3本、従来法3本の手話動画の組み合わせで手話文10文を構成した。被験者21名を7名ずつ3つのグループに分けて、各グループにX/Y/Zそれぞれの提示パターンを割り当てて評価した。

4.1.2 実験手順

被験者は手話動画を1本視聴する度に、3つの項目に対して回答を実施する。ひとつ目は、手話動画から読み取れたニュース速報文の内容に関する質問である。例えば、「新潟県の村上市と関川村に大雨特別警報 最大級の警戒が必要」という内容を表現した手話動画に対して、進行役が「警報が出た場所はどこですか？」という形で各ニュース速報の内容に合わせて質問した（質問1）。質問1は全ての手話文において、複数の地名や人名、組織名などの空間位置の活用が発生する固有名詞部分の読み取りに関する質問とした。その後、「場所の数はいくつでしたか？」という形で手話文中に表現された固有名詞の数について質問し（質問2）、さらに手話文10文中2文は内容を詳細に確認するための補足質問（質問3）を実施した。表1に質問項目の例を示す。

2つ目は、手話の理解しやすさに関する主観評価で、0～100のVisual Analog Scale（VAS）について、0から100の間でスライドバーを操作して回答する（0：理解しづらい～100：理解しやすい）。

3つ目は、手話の自然さに関する主観評価で、理解しやすさ同様、0～100のVASで回答する（0：不自然～100：自然）。

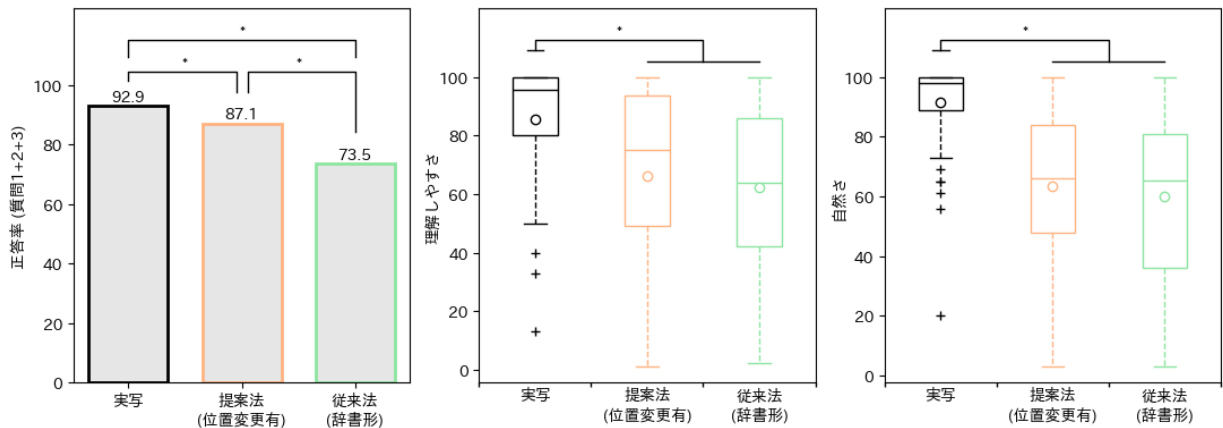


図7 評価実験の結果：各グラフは、左側が全質問の正答率、中央が理解しやすさ、右側が自然さの評価結果を示す。各グラフとも黒/橙/緑色が実写/提案法（位置変更有）/従来法（辞書形）のスコアを示す。また統計検定の結果から有意差があると判定された箇所に“*”のマークがつく。箱ひげ図の○は平均値、-は中央値を示す。

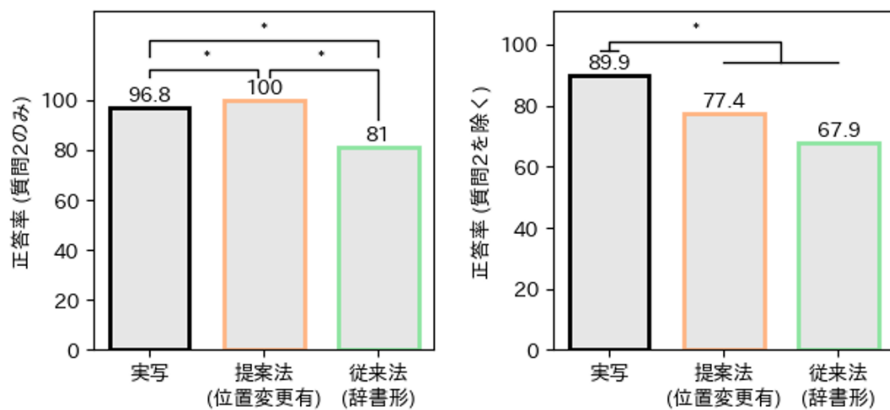


図8 質問2のみの正答率（左）と質問1と3の正答率（右）：各図とも黒/橙/緑色が実写/提案法（位置変更有）/従来法（辞書形）のスコアを示す。また統計検定の結果から有意差があると判定された箇所に“*”のマークがつく。

実験開始前に、各被験者に対してサンプルの手話動画1本を提示し、実験手順を説明する時間を設けた。また、評価時の各手話動画の視聴回数は2回とした。

4.2 実験結果

実験の結果を図7に示す。図7の左側の棒グラフは、空間位置変更が発生する固有名詞部分の読み取りに関する質問について、被験者全21名分の正答率の平均値である。質問1から3までを含んだ全ての質問に対する結果である。正答率は、各質問に対して得られた回答を“正解”、“部分正解”、“不正解”に分類し、それぞれに対して100、50、0というスコアを与えて算出した。スコア付けについては、複数の固有名詞を全て正確に認識できた場合は100、固有名詞の認識に一部でも誤りがあった場合は50、単語を固有名詞として認識できなかった場合や全く読み取れなかった場合は0という基準を定義した。

図7に示した通り、正答率が高い順から実写が92.9%、提案法（位置変更有）が87.1%、従来法（辞書形）が73.5%となった。提案法は実写と比べると正答率が5.8%低い結果となったが、提案法による空間位置変更によって手話CGに関する正答率が従来法と比べ13.6%改善することが示された。100、50、0に分類されたスコアを手話動画の

種別ごとに計上し、3種類の手話動画間の正答率についてカイ二乗検定（P値0.01、両側検定）を実施したところ、実写、提案法、従来法の全ての手話動画の間に有意な差があることが示された。

次に、理解しやすさと自然さに関する主観評価の結果を図7の中央と右側のグラフに示す。各グラフはスコアの平均値を示している。全体の傾向を見ると、理解しやすさと自然さどちらも図7左側の正答率の結果と同様、実写、提案法、従来法の順でスコアの平均値が高い結果となった。有意差については、マン・ホイットニーのU検定（P値0.01、両側検定）を実施し、実写と2つの手話CGの間では有意差があったが、手話CGの提案法と従来法の間には有意な差がないという結果となった。

5. 考察

5.1 手話動画ごとの比較

4.2節の評価結果を基に分析を実施した。質問1から3まで全ての質問に対する正答率について、手話CGの提案法と従来法を比較すると、提案法によって手話CGの正答率が向上したものは10文中6文であった。3文は位置変更の有無で正答率に変化がなく、1文のみ提案法が従来法より低い正答率となった。提案法の位置変更による正答率の

表2 手話文ごとの正答率（全21名の平均）

番号	話題	正答率		
		実写	提案法	従来法
1	政治	71.4	85.7	57.1
2	政治	100	78.6	85.7
3	交通	78.6	50.0	50.0
4	災害	100	78.6	78.6
5	災害	85.7	85.7	57.1
6	災害	100	57.1	50.0
7	芸能	85.7	92.9	92.9
8	政治	85.7	92.9	64.3
9	交通	100	85.7	78.6
10	政治	71.4	50.0	28.6

改善例として、「新潟県の村上市と関川村に大雨特別警報最大級の警戒が必要」という手話文で説明する。従来法の手話動画において、「村上市」と「関川村」という2つの地域を辞書形のまま正面の同じ位置で提示した場合には、複数の地域に関する話題ではなく、「村上市の関川村」というひとつの地域に関する情報であると誤って認識する被験者が多かった。一方で、提案法により空間位置を左右に位置変更した手話動画では、「村上市」と「関川村」を2つの異なる地名として正確に判別することができた。

また、提案法が従来法より低い正答率になった1文は、政治関連のニュース速報で、手話文中に登場する2つの政党名に関して質問した。ヒアリング時のコメントから、回答を誤った被験者全員が、該当箇所の手話表現の速度が速すぎたため読み取れなかったことが分かった。2つの政党名の手話表現は、どちらも5つの手話単語を連続して組み合わせることで表現される複合語であった。提案法と従来法では手話モーションの再生速度は同一であったが、複数の手話単語を連続して提示する速い動作に左右への位置変更を追加したことで、動作速度がさらに増したように認識された可能性がある。手話文中の固有名詞などの重要な単語は再生速度を下げた方が読み取り易いという指摘もあり[3]、空間位置と再生速度の相関については、今後の検討課題である。

さらに、理解しやすさや自然さの主観評価の結果において、提案法と従来法の間でCGの位置変更処理の有無に関わらず有意差がなかった。つまり、提案法によってモーションデータを修正した手話CGの手話動作が、辞書形モーションを加工せずに接続した従来法と同等の自然さであることが示された。提案法において身体干渉回避オフセット処理を適用することで、位置変更後の身体への干渉が軽減されたことも不自然さの軽減に寄与したと考えられる。

5.2 手話文ごとの比較

手話文を読み取るにあたって、ニュース速報の話題の違いによって理解しやすさに影響が出た可能性があるため、質問1に関する手話文ごとの正答率の傾向を分析した。手話文ごとの全21名の平均正答率を表2に示す。

表2より、正答率が高い順から実写、提案法、従来法という全ての話題を含めた場合の傾向と同様の結果になった

手話文が10文中3文あった（手話文6, 9, 10）。また、最も高い正答率については、10文中6文が実写となった（手話文2, 3, 4, 6, 9, 10）。しかし、提案法が実写より高い正答率となった手話文が10文中3文あった（手話文1, 7, 8）。この原因として、手話の空間領域の広さと位置変更の変化量の違いが考えられる。辞書形のモーションデータは基本的に手話単語を大きく明確に表現した動作で収録しており、辞書形モーションを接続して生成する手話CGは必然的に手話空間が大きくなる。一方で、実写の手話母語者による手話は、手話CGと比べると手話を表現する空間領域が狭く、空間位置変更の変化量も小さい。そのため、実写の空間位置変更と比べ変化量が大きい提案法による空間位置変更では、異なる複数の固有名詞を提示し分けているという空間活用本来の意図が明確に読み取れたことが正答率に影響したと考えられる。

5.3 提案法による固有名詞の数の認識の向上

質問2のみの正答率と、質問1と3を合わせた正答率を分けて分析したそれぞれの結果を図8に示す。図8左側の質問2のみの正答率について、提案法は被験者全員が全ての手話文に関して100%の正答率となり、実写の96.8%よりも高い結果となった。5.2節で先述したように、提案法による空間位置変更は、実写の空間位置変更と比べ空間を明確に区別して提示したためと考えられる。つまり、提案法を用いた手話CGにおいて、手話文中の複数の固有名詞や文節の切れ目を明確化するための空間の文法的な活用が再現できていると考えられる。

しかし、質問2の固有名詞の数を正確に回答できても、質問1や3で確認した固有名詞の内容を正確に回答できなかったケースがあった。図8右側の質問1と3の正答率を見ると、提案法の正答率が77.4%となり、実写の89.9%と比べ12.5%低い結果となっている。原因のひとつとして、固有名詞の読み取りに影響を与える口型の再現に課題があることが考えられる[2][25][26]。本実験における手話CGの口型制御は、固有名詞を構成する手話単語モーションに対して、事前に収録した仮名文字（音節）ごとの口型モーションを固有名詞の読み仮名に合わせて接続する制御手法を利用した[1][25]。しかし、固有名詞について置換した口型モーションと手指動作が非同期であることや、口型アニメーションの品質自体が低いことが原因で、アバターの口型が不自然となっていた。そのため、提案法によって手話文中に登場する複数の固有名詞の総数が把握できても、口型の情報が欠落し、各手話単語が示す具体的な固有名詞を理解できなかったと考えられる。そこで、本提案法による位置変更手法を、先行研究の口型修正手法[26]と併用することで、固有名詞の読み取り精度向上に関して相乗効果が見込まれる。結果的に、手話文中の固有名詞を構成する手話単語や文節の境界がより明確になり、手話CGによる手話文の内容理解が向上すると考えられる。

6. おわりに

本研究では、手話CGアニメーション生成において、ニュースで重要となる固有名詞を手話で正確に表現するため、手話の文法要素のひとつである空間活用を再現するための

空間位置制御手法を提案した。辞書形のモーションデータの水平方向の空間位置を、文意に応じて左・中央・右に制御することで、複数の固有名詞を異なる位置で提示する手話の空間活用を再現する。さらに、位置変更に伴う腕の身体への干渉を解消するために、前処理としてオフセット用のデフォルト姿勢をキーポーズとして挿入する身体干渉回避オフセット処理を追加した。評価実験を通じて、提案法を用いた空間位置変更により、生成した手話 CG の内容理解に関する質問の正答率が従来法と比べて 13.6% 改善することが示された。本手法を用いて手話の空間位置を変更することで、手話 CG において手話文中の固有名詞や文節の境界が明確になり、元の日本語文の内容を適切に伝えることが可能となる。

今後、日本語文から手話 CG を自動生成するにあたって、空間位置制御処理の自動化が課題となる。自然言語処理を活用することで日本語文中の固有名詞の数など位置制御用の情報を抽出し、手話単語列中の各単語に左・中・右といった補助的なラベル情報を付与することで位置制御処理が自動化できると考えられる。また、位置変更時にオフセット処理を適用しても身体干渉が解消しない手話モーションが存在するため、IK による腕の制御と同時に腰関節を左右に回転するような制御を適用することで制御処理の品質を改善できると考えられる。さらに、本手法を口型修正手法 [26] など他の文法要素を再現する手法と併用することで、手話 CG の内容理解に関して更なる品質向上が実現できると考えられる。

参考文献

- [1] 中谷真規, 内田翼, 宮崎太郎, 金子浩之, 佐野雅規, “句や節のまとまりを考慮した任意文手話 CG 生成システムの開発”, 2023 映情学年大, 31B-5 (2023).
- [2] 内田翼, 佐野雅規, “手話 CG アニメーション生成における手話特有の文法要素の影響に関する調査”, 映情学創立 70 周年記念大, 31D-2 (2020).
- [3] 内田翼, 中谷真規, 梶山岳士, 金子浩之, “ニュース速報を対象とした手話 CG アニメーションの評価”, 2024 映情学年大, 31B-7 (2024).
- [4] 武居渡, “〈コミュニケーション障害と発達心理学の接点〉言語の写像性は言語獲得を促進させるか: 手話獲得研究からの知見”, コミュニケーション障害学, Vol.23, No.2, pp.143-151 (2006).
- [5] T. Uchida, T. Miyazaki, H. Kaneko, “HamNoSys-based Motion Editing Method for Sign Language”, Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources, pp.376-385 (2024).
- [6] T. Hanke, “HamNoSys — Representing Sign Language Data in Language Resources and Language Processing Contexts”, In Proceedings of the LREC2004 Workshop on the Representation and Processing of Sign Languages: From Sign-Writing to Image Processing. Information techniques and their implications for teaching, documentation and communication, Vol.4, pp.1-6 (2004).
- [7] R. Wolfe, J. C. McDonald, T. Hanke, S. Ebling, D. V. Landuyt, F. Picron, V. Krausneker, E. Efthimiou, E. Fotinea, A. Braffort, “Sign Language Avatars: A Question of Representation”, Information, Vol.13, No.4:206, WP-B-5 (2022).
- [8] S. Gibet, N. Courty, K. Duarte, T. L. Naour, “The Signcom System for Data-driven Animation of Interactive Virtual Signers: Methodology and Evaluation”, ACM Transactions on Interactive Intelligent Systems (TiiS), Vol.1, No.1, pp.1-23 (2011).
- [9] L. Naert, C. Larboulette, S. Gibet, “LSF-ANIMAL: A Motion Capture Corpus in French Sign Language Designed for the Animation of Signing Avatars”, In 12th International Conference on Language Resources and Evaluation (LREC 2020), pp.6008-6017 (2020).
- [10] S. Gibet, P.F. Marteau, “A Text-To-SL Synthesis System Using 3D Avatar Technology”, 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), pp.1-5 (2023).
- [11] H. Brock, K. Nakadai, “Deep JSLC: A Multimodal Corpus Collection for Data-driven Generation of Japanese Sign Language Expressions”, In 11th International Conference on Language Resources and Evaluation (LREC 2018), pp.4247-4252 (2018).
- [12] 市川薫, “理工学的手法による手話研究の事例”, 手話学研究, Vol.19, pp.43-52 (2010).
- [13] T. Uchida, N. Nakatani, T. Miyazaki, H. Kaneko, M. Sano, “Motion Editing Tool for Reproducing Grammatical Elements of Japanese Sign Language Avatar Animation”, Eighth International Workshop on Sign Language Translation and Avatar Technology (SLTAT2023), WP-B-10 (2023).
- [14] 藤垣俊也, 杉山真也, 松本忠博, 加藤三保子, “日本語から手話への機械翻訳における空間表現 CL 述語翻訳の試み”, 言語処理学会第 21 回年次大会 (NLP2015) 発表論文集, pp.952-955 (2015).
- [15] M. Kipp, Q. Nguyen, A. Heloir, S. Matthes, “Assessing the Deaf User Perspective on Sign Language Avatars”, The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility (ASSETS 2011), pp.107-114 (2011).
- [16] R. Elliott, J. R. W. Glauert, J. R. Kennaway, I. Marshall, E. Safar, “Linguistic Modelling and Language-processing Technologies for Avatar-based Sign Language Presentation”, Universal Access in the Information Society, Vol.6, Issue.4, pp.375-391 (2008).
- [17] M. Huenerfauth, L. Zhao, E. Gu, J. Allbeck, “Design and Evaluation of an American Sign Language Generator”, Proceedings of the Workshop on Embodied Language Processing (Embodied NLP 2007), pp.51-58 (2007).
- [18] M. Huenerfauth, P. Lu, H. Kacorri, “Synthesizing and Evaluating Animations of American Sign Language Verbs Modeled from Motioncapture Data”, Proceedings of the 6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT 2015), pp.22-28 (2015).
- [19] S. Ebling, “Automatic Translation from German to Synthesized Swiss German Sign Language”, PhD Thesis (University of Zurich) (2016).
- [20] M. Filhol, J. McDonald, R. Wolfe, “Synthesizing Sign Language by Connecting Linguistically Structured Descriptions to a Multi-track Animation System”, 11th International Conference on Universal Access in Human-Computer Interaction (UAHCI 2017), pp.27-40 (2017).
- [21] J. C. McDonald, M. Filhol, “Natural Synthesis of Productive Forms from Structured Descriptions of Sign Language”, Machine Translation, Vol.35, Issue.4, pp.363-386 (2021).
- [22] J. H. Kim, E. J. Hwang, S. Cho, D. H. Lee, J. Park, “Sign Language Production with Avatar Layering: A critical Use Case over Rare Words”, In 13th International Conference on Language Resources and Evaluation (LREC 2022), pp.1519-1528 (2022).
- [23] L. Naert, C. Larboulette, S. Gibet, “Motion Synthesis and Editing for the Generation of New Sign Language Content: Building New Signs with Phonological Recombination”, Machine Translation, Vol.35, Issue. 3, pp.405-430 (2021).
- [24] V. Sutton, “Lessons in Sign Writing”, Textbook and Workbook (Second Edition), The Center for Sutton Movement Writing, Inc., La Jolla, CA (1995).
- [25] 森田祐介, 内田翼, 佐野雅規, “ニューラル手話翻訳における固有名詞訳出手法の検討”, 2022 映情学年大, 11B-1 (2022).
- [26] 箱崎浩平, 村上智哉, 内田翼, 金子浩之, “口型動作を修正した手話 CG ニュース文の評価と分析”, 電子情報通信学会技術研究報告, WIT2024-33 (2025).
- [27] 「テレビジョン放送における手話通訳実践に役立つ参考例集」作成委員会, “テレビジョン放送における手話通訳実践に役立つ参考例集”, (2025)