匿名コミュニケーションのための
手話映像生成に関する研究

平成２６年度

筑波技術大学大学院技術科学研究科

産業技術学専攻

松岡　通浩

目次

図目次

表目次

筑波技術大学

修士（工学）学位論文

第1章 序論

1.1 研究背景

音声チャットやビデオチャットなどのアプリケーションには，ボイスチェンジャーやアバター機能が搭載されており，匿名性を確保したまま他者とのコミュニケーションを行うことができる．活用場面として，音声チャットが搭載されているオンラインゲームや会社内の会議などが挙げられる．音声には個人を特定する情報が含まれているため，不特定多数のユーザーを相手にするオンラインゲームではボイスチェンジャーを利用するユーザーもいる．また，社内でも上司，部下など立場関係なく発言することにより，会議を活性化して生産性の向上を狙うことも可能である．しかし，主に聴覚障害者が使用する手話による匿名コミュニケーションについては十分に検討されていない．

現在普及している音声チャットやビデオチャットにおいて，聴覚障害者が匿名性を確保しながらコミュニケーションを行うことが困難な理由は 2 つある．一つ目は，聴覚障害者のなかには，音声の正確な聞き取りおよび明瞭な発音が困難な人がおり，音声のみによるコミュニケーションは非常に難しいためである．したがって，ボイスチェンジャーのような音声ベースで匿名性を確保する方法は利用できない（図 1.1）．二つ目は，図 1.2 のように顔や体をアバター機能で単に隠蔽した状態で，手話による円滑なコミュニケーションを行うことは非常に困難ということである．手話は手指動作だけでなく，手指動作以外の表情や頭の動き，眉の動き，まばたき，視線，顎の動き，口型などの非手指動作（NMS : Non Manuals Signals）も重要であり，それらが欠けていると内容が伝わりにくくなる可能性がある．例えば，相手がマスクを着用した状態で手話を表現しても顔半分が覆い隠されて口型が全く見えず，目から下の表情も乏しくなるため，相手の表現内容を判断することが困難である．しかし，顔や体が表出すると，性別や容姿，服装，障害などから匿名性の確保ができない可能性があるだけでなく自分のコンプレックスが他人の目に晒されることになり，コミュニケーションに悪影響を及ぼすことも考えられる [1].

[1] 近年では，マスクを着用して顔を覆うことによって安心感を得たり，顔のコンプレックスを隠したりする者がいる．このような本来の衛生上の理由とは異なる目的でのマスクの利用は「だてマスク」と呼ばれている．
「だてマスク」，Wikipedia，http://ja.wikipedia.org/wiki/だてマスク，（2015/02/15 参照）

一方で，オンラインゲームや SNS においては，文字チャットを利用して聴覚障害者でも匿名コミュニケーションを行うことは可能である．しかし，手話を母語とする聴覚障害者の中には日本語の理解が十分でない人もいるため，文章の把握に苦労する場合がある．また，文章入力の煩わしさや伝達速度から，手話でコミュニケーションしたいという人もいる．

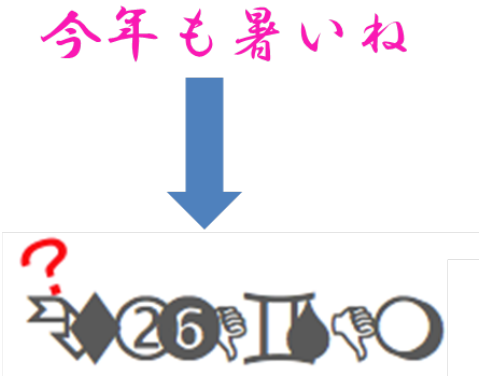

図 1.1 聴覚障害者の音声利用のイメージ

図 1.2 顔の隠蔽

手話者の匿名性を確保しながら手話表現する方法として，リアルタイムに CG モデルを動かし，手話を表現する方法が考えられる．手話者が手話を表現する際にモーションキャプチャやフェイストラッキングで得られたデータを CG モデルに入力し，リアルタイムに手話，表情などを再現する．つまり，手話者が表出した手話をリアルタイムに CG で表現して匿名手話コミュニケーションに利用する．図 1.3 のように，生成した匿名手話映像を従来のビデオチャットアプリケーションの入力映像として利用することで手話による匿名コミュニケーションを実現できる．また，聴覚障害者が SNS や動画共有サービスなどで気軽に手話映像を投稿することで，匿名性を確保したまま音声の代わりに手話で情報を発信できるようになり，社会参加の機会をより拡大することが可能になると考えられる．

また，手話を CG モデルで表現する方法では，モデルを動作させるための骨格，表情や手指情報のみで匿名手話映像を表現できるようになるため，実写映像よりもデータ量を少なくすることが可能になる．遠隔からネットワークを介して手話通訳を提供する遠隔手話通訳に応用できれば，手話通訳者の骨格，手指情報のみに限定することで送信するデータ量を減らすことができ，より安定で遅延の少ない遠隔手話通訳を実現できると考えられる．

図 1.3 匿名手話映像のオンラインコミュニケーションへの活用

1.2 研究目的

本研究は，手話による匿名コミュニケーションが可能な匿名手話映像の生成法を提案，実装し生成した手話映像を評価することを目的とする．本生成法では，モーションキャプチャとフェイストラッキングが可能なデバイスを利用する．手話者の動きの骨格情報をモーションキャプチャで取得して CG モデルに適用させる．また，細かい動きの計測が困難な手指と顔の部位はモーションキャプチャとフェイストラッキングで対象となる領域を取得して，実写画像から領域を切り出して CG モデルに合成して表現する．実写画像をそのまま表出すると匿名性を確保できない可能性があるため，手指と顔の部位を実写画像のまま合成する方法（以後，実写表現）と，色と形を変換して合成する方法（以後，変換表現）で 2 種類の匿名手話映像を生成する．実写表現と変換表現のイメージを図 1.4 に示す．

本生成法で生成した匿名手話映像について，コミュニケーション時の匿名性と可読性を評価するために実写表現，変換表現それぞれの方法について比較評価実験を行う．2 種類の方法で生成した匿名手話映像の手話の読み取りやすさ，匿名性の確保，違和感の有無を被験者に評価してもらい，本生成法の有効性を明らかにする．

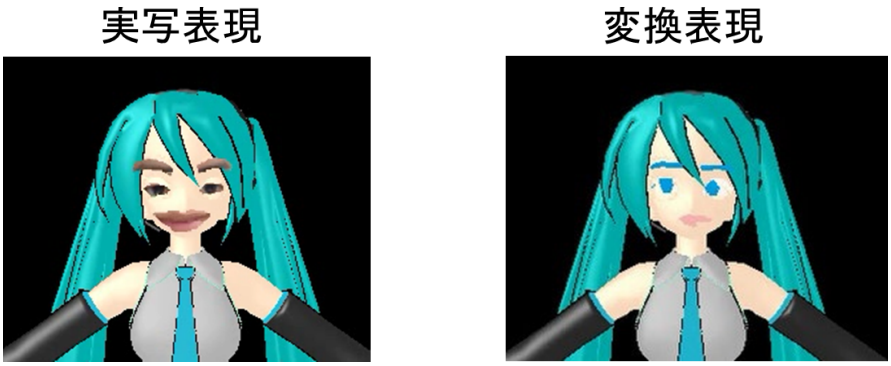

図 1.4 実写表現と変換表現

1.3 本論文の構成

本論文の構成は以下の通りである．

第 1 章では，本研究の研究背景と目的について述べた．

第 2 章では，手話を構成する要素について説明し，手話の匿名化の課題を挙げる．また，手話 の CG 表現と匿名性の関連研究を挙げ，本研究の位置付けについて述べる．

第 3 章では，手話による匿名コミュニケーションのための手話映像の生成法を提案する．手話者が手話を表現し，表現した内容を匿名手話映像に変換する方法について述べる．また，生成した匿名手話映像の動作についても述べる．

第 4 章では，第 3 章で述べた方法で生成した匿名手話映像に関して匿名性の確保，手話の読み取りやすさ，違和感の有無について評価実験を行った結果について述べる．

最後に，第 5 章でまとめ，今後の課題について述べ，本論文を締めくくる．

第2章 関連研究

2.1 手話の匿名化

聴覚障害者がコミュニケーションに用いている手話は，図 2.1 のように，手指動作と非手指動作から構成されている[1]．手指動作は，手型，手のひら方向，位置，運動の 4 要素から構成されている．非手指動作は，表情，頭の動き，眉の動き，まばたき，視線，顎の動き，口型などの手指動作以外の要素からなり，文法機能と感情の表現の役割を果たしている．例えば，平叙文の場合，「私の名前+（うなずき，場合によっては顎上げ）+佐々木です．」，疑問文の場合，「あなたの名前+（眉上下+顎出し+視線）+何？」となる．

手話でのコミュニケーションに重要なのは，非手指動作の活用であると言われている[2]．手型を拳形に固定したドラえもん手話と，顔を隠した状態で手話を表現する能面手話では，ドラえもん手話の方が，伝達力があることが実験的に示されている．ドラえもん手話は，手型が固定化されるために同音異義語のような「同型手話語」が多く識別が困難である．しかし，非手指動作が文法機能と感情の表現の役割を果たし，それによって文脈の推理が容易になるため，内容の理解がしやすくなる．つまり，手話でのコミュニケーションにおける非手指動作の重要性が指摘されている．

一方で，非手指動作に使われる顔の部位が表出されていると，匿名性を確保できない可能性がある．そのため，手話者の匿名性を確保し表現した内容を相手に正確に伝えるためには，顔や体を隠蔽しながら手指動作，非手指動作を何らかの方法で表現する必要がある．

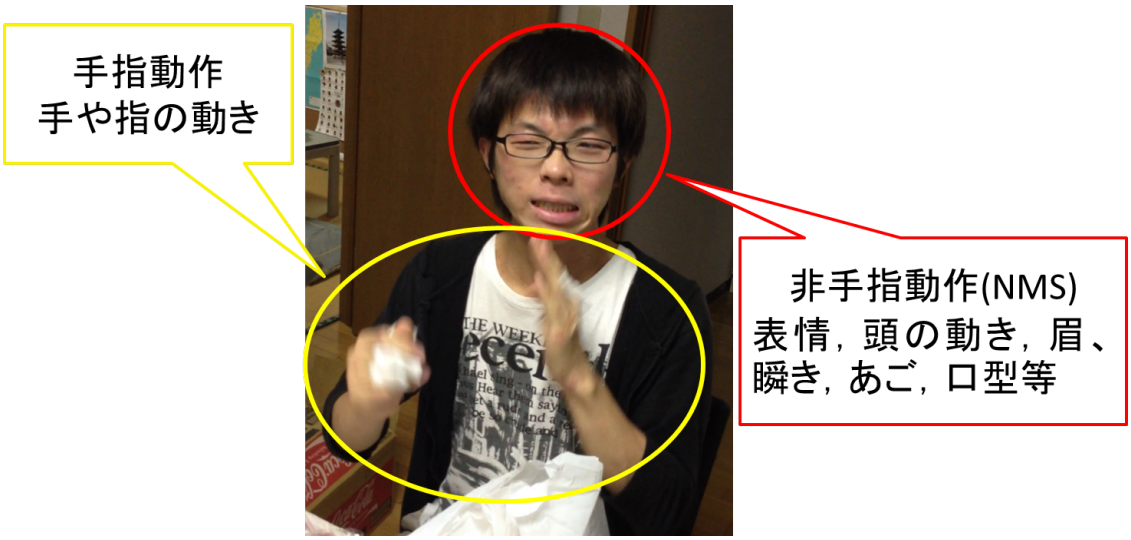

図 2.1 手話を構成する手指動作と非手指動作

2.2 手話 CG 表現

手話を CG で表現する関連研究としては，図 2.2 のように日本語のテキストを入力し，テキストの中の単語に相当する手話のアニメーションを CG により自動生成する方法が報告されている [3][4]. 手話 CG 生成システムは日本語と手話という言語同士を変換する自動翻訳技術の面と，手話のアニメーション映像を生成する CG 技術の面から成り立っている．

まず，言語翻訳について述べる．手話は視覚言語であるため，最終的には手話 CG への変換が必要となる．しかし，日本語文から手話映像へ直接変換することは難しいため，日本語文から手話単語列に変換し，変換された手話単語列から手話 CG に変換する．

次に，手話アニメーション生成について述べる．手話アニメーションを生成するためには，日本語の単語に対応した手話 CG を用意する必要がある．手話者の手や指，頭，足だけでなく顔の特徴的な部位にもマーカーを装着し，手話単語ごとにモーションキャプチャを行う．そして，TVML（TV program Making Language）を用いて日本語に合わせて手話 CG アニメーションを生成する．しかし，現状では，任意の話題の文章を日本語から手話に変換することは極めて技術的なハードルが高く現実的ではないため，気象情報のニュースを対象としている．

また，人体と CG モデルを同期させて動かす方法として，サイバーグローブやマーカーによるモーションキャプチャ，フェイストラッキングを用いて体の動きや顔の表情を再現する方法がある．しかし，上記のデバイスは環境整備にかかる負担が大きく，一般的なユーザーへの普及には至っていない．手話コミュニケーションに利用するにはユーザーに負担の少ない非接触のデバイスを選択する必要がある．

本研究では，日本語のテキストを入力するのではなく，実際に手話者が手話を表現し，手話者の匿名性を確保しながらリアルタイムにその内容を手話 CG に変換する．さらに，安価でユーザーへの負担が少ないデバイスを用いて手軽に匿名で手話コミュニケーションを行う. このような匿名コミュニケーションを目的としたリアルタイム性のある手話 CG 表現に関連する研究は調査した範囲では例を見ない.

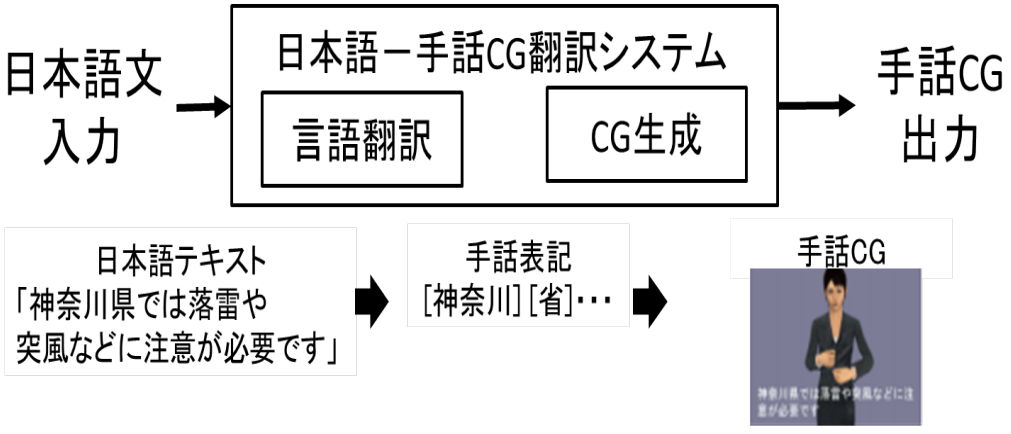

図 2.2 日本語から手話 CG への翻訳システム

2.3 匿名性

匿名性について調査した関連研究では,アメリカ手話を使用する手話者は非手話者より顔の部位の違いの判断が優れているという報告がある[5]．同報告の実験では，まず，図 2.3 のように被験者にあらかじめ目標の顔画像を被験者に提示し覚えさせる．次に，2 つの顔画像を提示する．一つは目標の顔画像と同じ顔画像で，もう一つは目標の顔画像の眉，目，鼻を別人のものと入れ替えた顔画像である.どちらかが目標の顔画像と同一人物かを当てさせる実験である．結果として手話者の方が，正答率が高いことがわかり，非手話者よりも顔の部位の違いで顔の識別ができるということが明らかになった．これは，読唇術や手話経験の影響があると言われている.

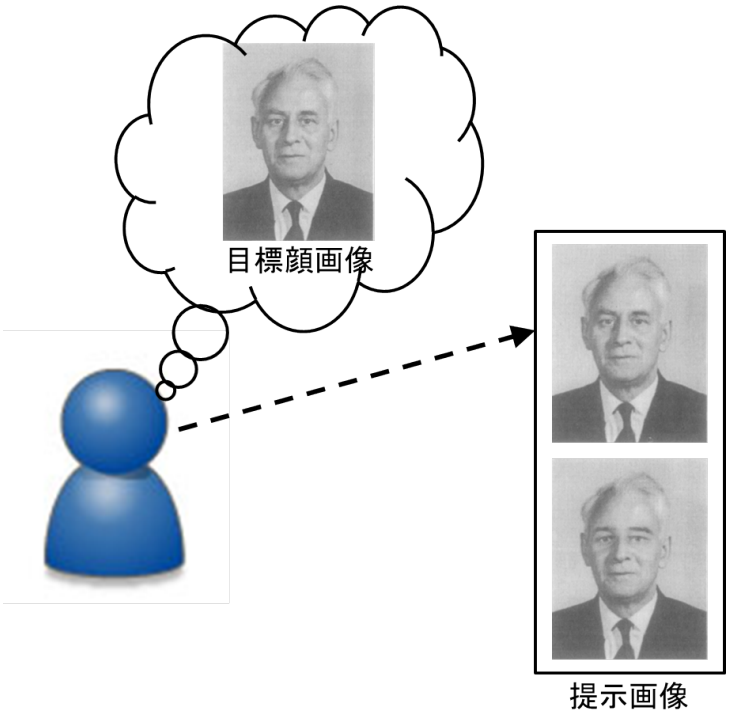

図 2.3 顔識別の実験概要

1.2 節で述べた本研究で提案する実写表現では顔の部位に実写画像を用いるため，日常から手話を使用している手話者は顔の部位を見ることで顔の特定ができる可能性があると考えられる．お互いに聴覚障害者であり手話でコミュニケーションを行うことを想定すると，実写表現では眉，目，口が表出されているので手話者本人を特定できる可能性がある．つまり，実写表現では匿名性を確保できない可能性がある．そこで，実写画像の色と形を CG モデルに合わせて変更して合成する変換表現も実装する．実験では，実写表現と変換表現による匿名手話映像の比較実験を行い，匿名性が確保されているかについて確認する.

匿名性の評価の関連研究としては，顔画像に対するプライバシー保護処理を定量的評価する指標として ID 可到達性が提案されている[6]．匿名性を確保するために，画像や映像を公開する際に人物の顔などに対してぼかしや塗りつぶしなどの画像処理を施すことが多い．一方で，人物の未検出や処理対象の領域のずれ，画像処理の性質によって匿名化すべき対象人物の容姿の漏洩のリスクが異なるといった，匿名性の影響を定量的に評価する研究は少なかった．同提案では観察者が被写体にどの程度馴染みがあるかを表す親密度と，どの程度観察者の印象に残るかを表す特徴度が，ID 可到達性に与える影響も考慮されている．本研究では生成する匿名手話映像の匿名性について ID 可到達性を用いて評価する．ID 可到達性については 4.3.2 節で述べる．

2.4 本研究の位置付け

手話表現は手指動作と非手指動作からなり，手話者の匿名性を確保し，内容を正確に伝えるためには，顔や体を隠しつつ手指動作，非手指動作を表現する必要がある．これを実現する方法として手話者の代わりに CG モデルを用いて手話を表現することが考えられる．しかし,先行研究では日本語のテキストを手話 CG アニメーションに翻訳する研究が多く，匿名コミュニケーションを目的として手話者の手話表現をリアルタイムに CG で表現する研究は例がない．そこで，本研究では，ユーザーに負担が少なく非接触で手軽にモーションキャプチャとフェイストラッキングを行うことができるデバイスを用いて，手話者の動きと表情，手指の情報を取得し，CG で匿名手話映像を生成する方法を提案する．

本方法では非手指動作を表現するために，顔の部位の実写画像を利用する．しかし，手話者は非手話者よりも顔の部位の違いで人物を識別する能力が高いことが示されているため，実写画像を CG モデルに合成した実写表現では匿名性を確保できない可能性がある．そこで，実写画像の色と形を CG モデルに合わせて変更し合成する変換表現も実装して，実写表現と変換表現とで匿名性ならびにその違和感について評価を行う．

また，匿名性が確保されていても手話の可読性が低ければコミュニケーションに利用することはできない．このため，実写表現と変換表現で生成した匿名手話映像を読み取る実験を行い，読み取ることができた内容の定量的な評価と，主観的な評価を行う．

本研究では提案法を実装したシステムを作成し，本システムで生成した匿名手話映像を手話コミュニケーションに活用することを想定して，匿名性，違和感ならびに可読性について評価する．本方法で十分な有効性が確認できれば，ビデオチャットやオンラインゲームなどの匿名でのコミュニケーションに利用できる．また，容姿やコンプレックスにとらわれない情報発信が可能になり，SNS などのオンラインサービスにおいて，手話で気軽に投稿することが可能になる．このように，これまで不可能だった匿名での手話表現が可能になれば，聴覚障害者のコミュニケーションの機会を増やすことができ，聴覚障害者の社会環境の向上に資することができると考えられる．

第3章 匿名手話映像の生成法

3.1 匿名手話映像の生成法の概要

匿名手話映像を生成するためには，手話者の匿名性を確保することが重要である．しかし，手話者の匿名性を確保しながら手話を表現することは容易ではない．顔や体を隠蔽する場合，手指動作，非手指動作の表現が困難だからである．匿名手話映像の生成法を検討するにあたって，手話者の匿名性を確保しながら手指動作と非手指動作を正確に伝える方法を検討する必要がある．また，手話コミュニケーションを想定した場合，表現する手話がわかりやすいかどうかも調査しなければならない．

そこで筆者は，非接触で手軽にスケルトントラッキングとフェイストラッキングが可能な Kinect を用いて，手話を CG モデルでリアルタイムに表現する方法を検討した．Kinect は実写画像，深度画像を 30fps で取得することができるデバイスであるため，手話者が手話を表現したとほぼ同時に CG モデルを動かすための骨格情報（スケルトン）を得ることができる．このスケルトンを CG モデルのボーンに対応させることによって，リアルタイムで CG モデルを動かすことができる．つまり，手話者を CG モデルに置き換えることができるため，匿名性を確保しながらリアルタイムで手話を表現することが可能になる．しかし，手指の型と非手指動作についてはスケルトントラッキングによって取得することができない．このため，手指については手の領域推定を行い，実写画像より同領域を抽出して CG モデルに合成する．顔についても必要な部位のみをフェイストラッキングで取得し，手指と同様に実写画像から CG モデルへ合成する．

匿名手話映像の生成法の概要を図 3.1 に示す．本研究では，手話者を Kinect と Web カメラで撮影して以下の 3 つの処理を行い，実写表現の匿名手話映像を生成する．

- 腕の動作の CG モデル表現（図 3.1(a)）
- 実写画像を用いた手指表現（図 3.1(b)）
- 実写画像を用いた非手指動作の表現（図 3.1(c)）

また，匿名性をさらに高めるために，実写画像の変換（図 3.1(d)）の処理をすることで，変換表現の匿名手話映像も生成する．

腕の動作の CG モデル表現については，手話者の腕の動きをスケルトントラッキングで取得し，CG モデルを動かすための向きや角度を計算して CG モデルに適用させることによって手話者の腕の動きを CG モデルで再現する．本処理方法については 3.3 節で述べる．

Kinect を含めた従来のセンサーでは指先の動きの計測や表情の認識が困難である．そのため，本研究では実写画像を用いた手指表現と実写画像を用いた非手指動作の表現によって実写画像ベースでの解決を試みた．実写画像から手指と顔の部位を抽出し CG モデルに

合成することで，より手話者の動きに近い手指動作と非手指動作を表現する．本処理方法については 3.4 節と 3.5 節で述べる．

実写画像を用いた手指表現と非手指動作の表現では，より高解像度な実写画像を利用することが望ましい．しかし， Kinect の実写画像は 640×480 であり解像度が十分でなく，手指の形や動きを把握することが困難であることが明らかになった[A1]．そこで，Kinect よりも高解像度な実写画像を得ることができる Web カメラ（1920×1080）を併用することで，高解像度な実写画像を抽出するための処理も実装した（図 3.1(e)）．本処理方法については 3.6 節で述べる．

ただし，手指や顔の部位に対応した実写画像を CG に合成すると匿名性を確保できない可能性がある．また，CG の質感を考慮せずに合成すると違和感が生じる．そこで，匿名性を確保しつつ自然な表現を行うために，実写画像の変換を行った上で合成する．本処理方法については 3.7 節で述べる．

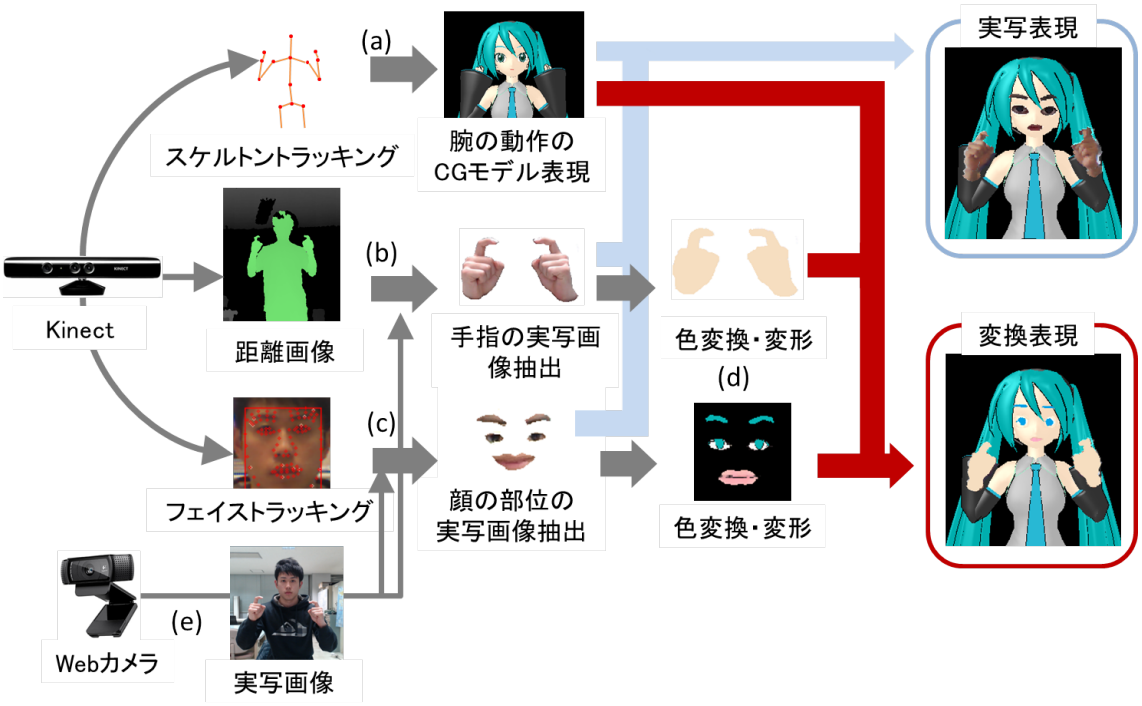

図 3.1 匿名手話映像の生成法の概要

3.2 開発環境

匿名手話映像の生成システムの開発環境の外観を図 3.2 に示す．また，ハードウェアの構成を表 3.1，ソフトウェアの構成を表 3.2 に示す．本方法の実装に使用したソフトウェアは Microsoft Visual Studio 2012 であり，開発言語は C++である．デバイスは Kinect，Web カメラを，ライブラリは Kinect for Windows SDK，Face Tracking SDK，DX ライブラリ，OpenCV を利用した．

図 3.2 匿名手話映像の生成システムの外観

表 3.1 匿名手話映像生成システムのハードウェア構成

名称	型番	仕様	用途
ノート PC	Panasonic Let's Note CF-AX2	CPU：Intel Core i7-3537U メモリ：4GB OS：Windows8.1 Pro	プログラム開発 デバッグ 実験
ディスプレイ	DELL 2407WFPb	サイズ：24 インチ 解像度：1920×1200	ノート PC の大画面サブモニター
3 次元計測センサー	Kinect for Windows センサー	RGB カメラ（解像度／フレームレート）：640×480／30fps 距離センサー（解像度／フレームレート）：640×480／30fps 有効距離：0.8m～4m（通常モード） 0.4m～3m（Near モード）	スケルトントラッキング フェイストラッキング 深度情報の取得
Web カメラ	Logicool HD Pro Webcam C920	解像度：1920×1080 フレームレート：30fps	実写画像の撮影

表 3.2 匿名手話映像生成システムのソフトウェア構成

名称	型番	用途
統合開発環境	Visual Studio Express 2012 for Windows Desktop	本方法の実装
Kinect アプリケーション開発キット	Kinect for Windows SDK v1.7	深度情報の取得 スケルトントラッキング
フェイストラッキング用ライブラリ	Face Tracking SDK	フェイストラッキング
C++言語用のゲームライブラリ	DX ライブラリ Ver3.12a	CG モデルの設定・操作・描画 テクスチャ画像の変形
画像処理ライブラリ	OpenCV2.4.9	Kinect と Web カメラ間のカメラキャリブレーション 顔の部位の輪郭線描画 実写画像の色変換

3.2.1 デバイス

モーションキャプチャとフェイストラッキングに用いるデバイスは Microsoft 社の Kinect で，撮影対象の 3 次元計測が可能なデバイスである．Kinect を図 3.3 に示す．本体内には，RGB カメラと距離センサーが搭載されている．距離センサーは PrimeSense の Light Coding という方式が採用されている[7]．Light Coding は，ある赤外線パターンをプロジェクターで照射し，赤外線カメラで投影された赤外線パターンの歪みから深度を計算する．そのため，距離センサーは赤外線プロジェクターと赤外線カメラで構成されている．その他に，マルチアレイマイクロフォンによる音声認識機能も搭載されているが，本研究ではマルチアレイマイクロフォンを用いない．

高解像度な実写画像を取得するためのデバイスは Logicool 社の Web カメラ HD Pro Webcam C920 である（図 3.4）．同カメラは 1920×1080 で Kinect よりも高解像度な実写画像を取得することができる．図 3.5 のように手話者の撮影用のカメラシステムを，Kinect の RGB カメラと縦軸が一致するように Web カメラをマウントして作成した．

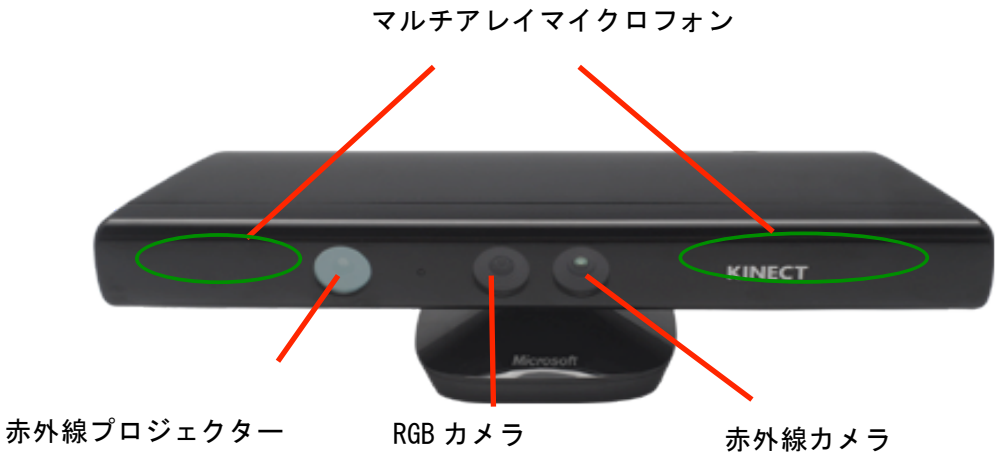

図 3.3 Kinect

図 3.4 Web カメラ

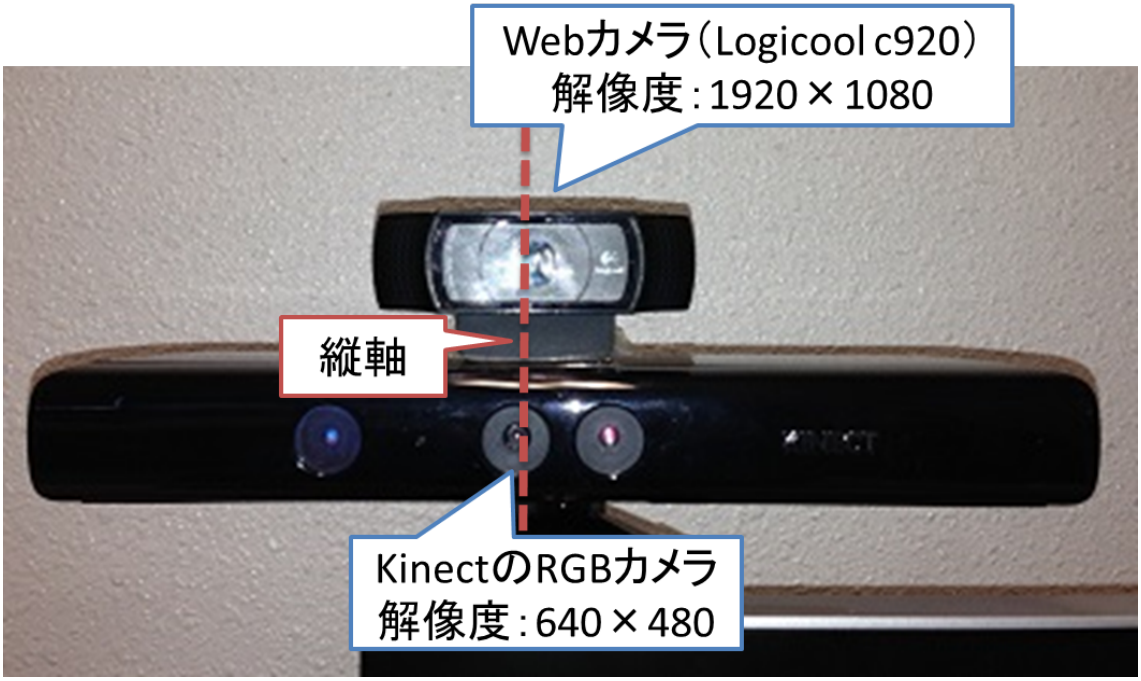

図 3.5 手話撮影用のカメラシステム

3.2.2 ライブラリ

Kinect for Windows SDK

Kinect で撮影した実写画像と深度画像の取得，スケルトントラッキングに Kinect for Windows SDK を用いた．Kinect for Windows SDK は Microsoft 社が無償で配布している Kinect アプリケーションの開発キットである[8]． Kinect の各センサーから実写画像，人物領域，深度値，人物のスケルトン，音声などの情報を取得することができる．人物領域については，人物を 6 人まで認識することができ，認識した人物には ID が振られ，深度値とともに画素ごとに識別される．深度値の検出可能範囲は，Default Mode（800[mm]～4000[mm]），Near Mode（400[mm]～3000[mm]）に限られる．本研究では，上半身を画像中に映す必要がないため，近深度値が取得可能な Near Mode をベースに説明を進める．スケルトントラッキングについては，画像中にいる人物を認識し，図 3.6 で示すようにその人物の 20 箇所の関節情報を取得することができ，それぞれ右手座標系の 3 次元座標値[m]で表される．3.3 節で述べる CG モデルの動きの表現については腕，頭を，3.4 節で述べる実写画像による手指表現については手を対象としているため，頭，首，左肩，左肘，左手首，左手，右肩，右肘，右手首，右手の 10 箇所に関してスケルトントラッキングを行う．

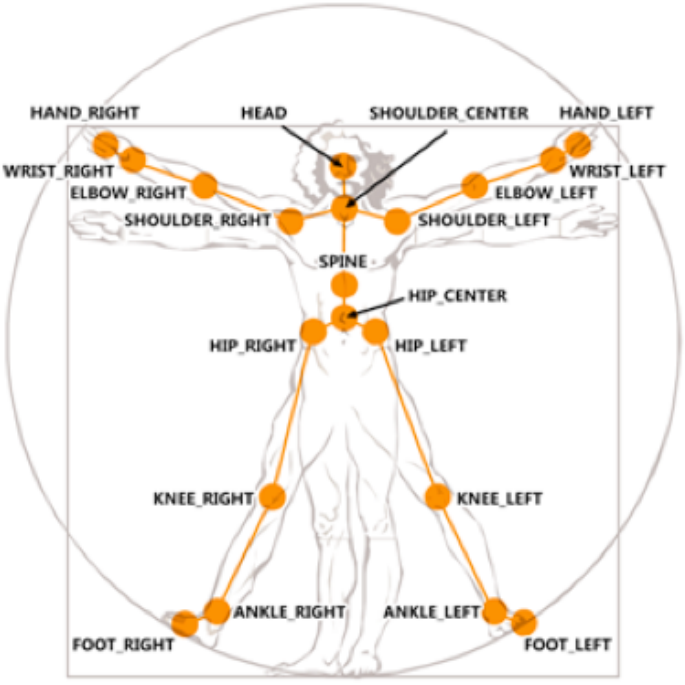

図 3.6 Kinect で取得可能な関節位置

Face Tracking SDK

顔の部位の特徴点を抽出するために用いたライブラリは Face Tracking SDK である．Face Tracking SDK は，Kinect for Windows SDK の Developer Toolkit に同梱されている．そのライブラリは，眉，目，鼻，口，および輪郭といった顔のパーツの位置，顔の向き，口の開閉などを取得することができる[8]．顔のパーツの特徴点は 87 箇所であり，図 3.7 にそれらを示す[9]．3.5 節で述べる実写画像による非手指動作の表現については眉，目，口を対象としているため，右目には 0～7 番目，左目には 8～15 番目，右眉には 16～25 番目，左眉には 26～35 番目，口には 48～59 番目の特徴点を利用する．眼鏡を着用した場合でも，光で反射しない限りフェイストラッキングを行うことが可能であり，筆者自身も確認できた．

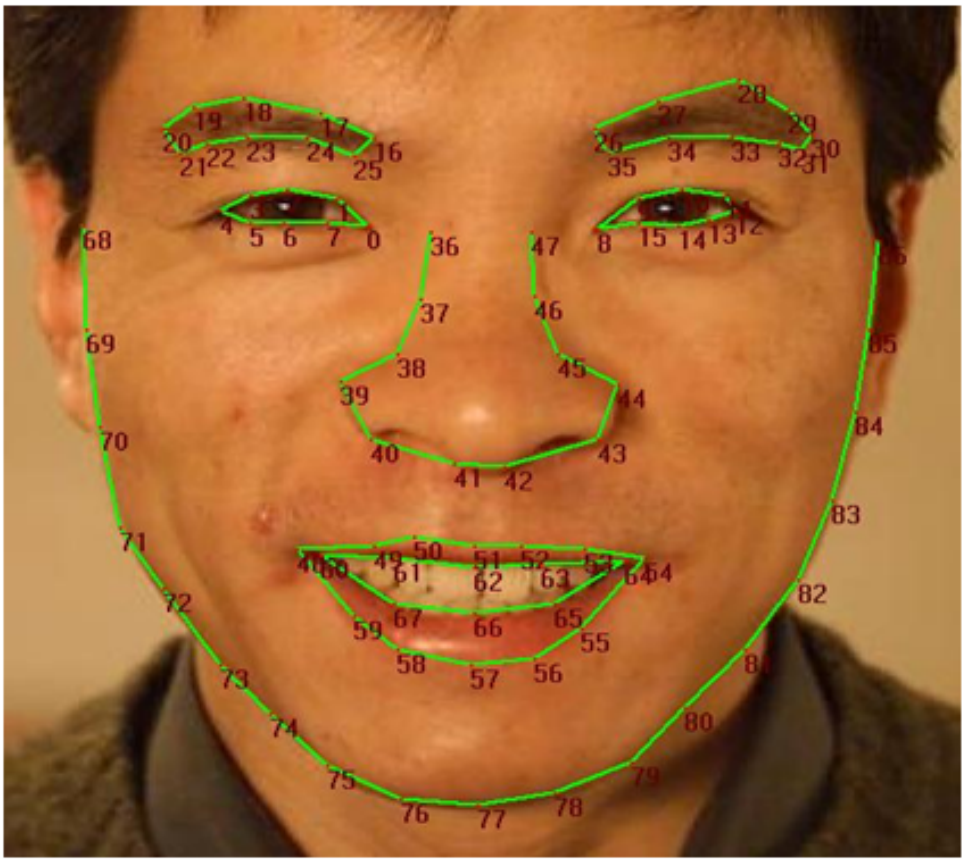

図 3.7 Kinect で取得可能な顔の部位の特徴点

DX ライブラリ

CG モデルの設定・操作・描画，テクスチャ画像の変形に用いたライブラリは DX ライブラリである．DX ライブラリは，DirectX や Windows 関連のプログラムを使いやすくまとめた形で利用できるようにした C++言語用のゲームライブラリである[10]．

OpenCV

Kinect と Web カメラ間のカメラキャリブレーション，顔の部位の輪郭線生成，実写画像の色変換に用いたライブラリは OpenCV である．OpneCV は，インテルが開発・公開したオープンソースのコンピュータビジョン向けのライブラリである[11].

3.3　腕の動作の CG モデル表現

CG モデルで手話の腕の動きを表現するために，手話者の腕の動作のモーションキャプチャを行う．2.1 節で述べたように，手指動作は手型，手のひら方向，位置，運動の 4 要素から構成されている．これらのうち，位置，運動の要素は，腕の動きによるものが大きい．手指だけでは手型の移動量，体に対する相対位置を示すことが困難だからである．したがって，それらの動きを取得し CG モデルに対応させる．

本研究では，ユーザーの各関節の 3 次元座標から位置姿勢を推定し CG モデルのボーンに適用させることで，手話者の腕の動きと同じになるように CG モデルを動かす．人間の姿勢は，各関節の回転角によって表現される．そのため，ユーザーの位置姿勢推定には，関節間のベクトルを用いてある関節に関して回転軸と回転角を計算する必要がある．関節間のベクトルは，2 つの関節の 3 次元座標で計算ができる．そのベクトルを用いて，内積で回転角を，外積で回転軸を求める．求めた回転角と回転軸を CG モデルのボーンに適用させることで，手話者と同じ動きの表現を実現できる．その処理方法の概要を図 3.8 に示す．図中の矢印は単位ベクトルとする．

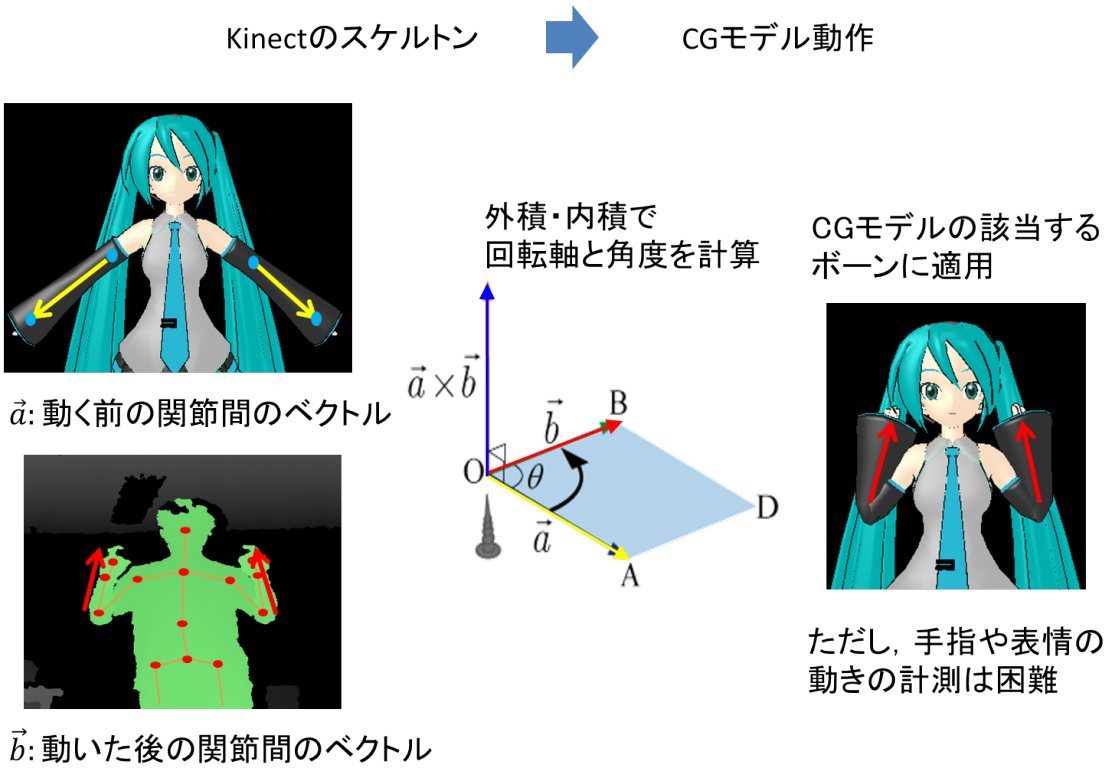

図 3.8 腕の動作の CG モデル表現の処理法

CG モデルを動かす処理方法について述べる．ユーザーの姿勢を CG モデルに反映させるためには，CG モデルを動かす前の姿勢からユーザーの姿勢へ変化させるための各関節の回転軸と回転角を計算する必要がある．CG モデルの各関節の 3 次元座標は既知であり，ユーザーの各関節の 3 次元座標は Kinect のスケルトントラッキングで取得することができる．CG モデル，ユーザーの関節間の単位ベクトルをそれぞれ$\vec{a}$，$\vec{b}$とする．そして，その 2 つのベクトルを用いて回転角 θ，回転軸$\vec{c}$を求める．回転角 θ，回転軸$\vec{c}$はそれぞれ

$$\theta = \cos^{-1}\left(\frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|}\right) \tag{3.1}$$

$$\vec{c} = \frac{\vec{a}\times\vec{b}}{|\vec{a}\times\vec{b}|} \tag{3.2}$$

で求まる．最後に，求めた回転角 θ と回転軸$\vec{c}$を CG モデルのボーンに適用する．他のボーンに関しても同様に処理を行うことで，CG モデルをスケルトントラッキングの結果に合わせて動かすことができる．

ただし，Kinect のスケルトントラッキングでは手指の細かな部分の 3 次元座標を取得することができないため，CG モデルの手指を動かすことは困難である．このため，本方法では実写画像を用いて手指を表現する．次節で詳細について述べる．

3.4　実写画像を用いた手指表現

3.4.1実写画像を用いた手指表現の概要

手話を表現するにあたって，手指動作である手型，手のひら方向の情報が不可欠である．前述したように，Kinect のスケルトントラッキングでは手指の細かな動きの計測が困難なため，実写画像を用いて手指を表現する．実写画像から手指領域を抽出して CG モデルに合成する方法について述べる．

3.4.2手指領域抽出

実写画像による手指表現を行うために，手指領域をリアルタイムに抽出する必要がある．深度情報を用いた手指領域の抽出方法では，手のひらの位置座標を中心に，手指を包含できる長さを半径とした楕円球体内の手指領域を抽出する方法がある[A1]．しかし，図 3.9 のように「こんにちは」「食べる」などのような顔と手が重なる手話の場合，その球体内に顔の領域も含まれてしまうことがある．つまり，顔と手の距離が小さい場合には顔領域と手指領域を分離することが困難になる．

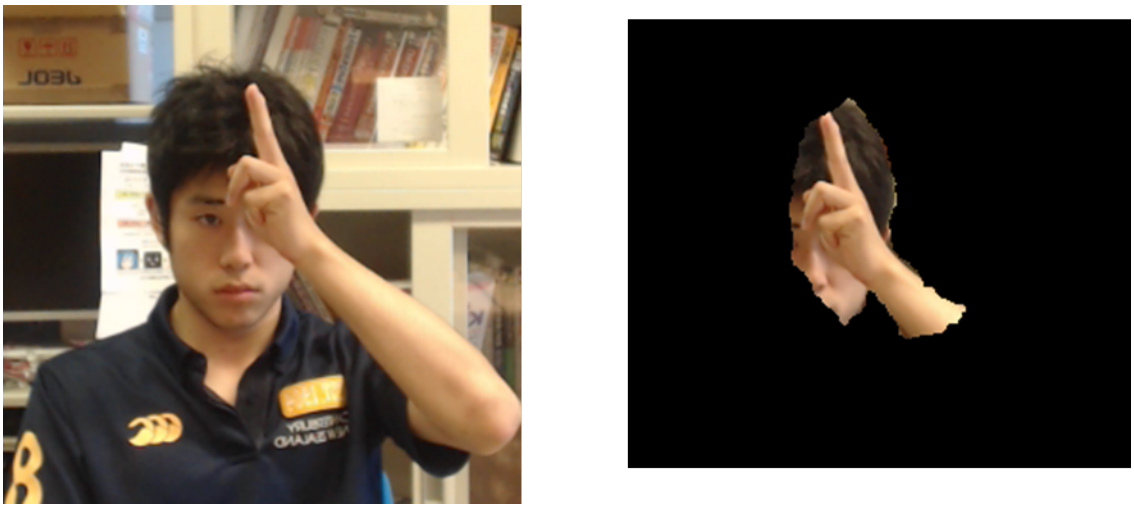

図 3.9 手指領域の誤認識

そこで本研究では，各画素の深度値から手指の体積を計算し，手指領域のみを抽出する方法を利用した[12]．手型をカメラで撮影したときに，実際に撮影される範囲は手型の表面である．手型を横から見たときに，カメラ側の半分よりも前の部分は手型の体積の半分程度である．その性質に基づいて体積を考慮することで，手指領域のみを抽出できる．

Kinect は撮影対象の 3 次元座標の取得が可能であり，手指の表面を推定できるため体積を計算することができる．手の位置を中心とした球体の範囲内で行うことで，手と顔が重なっても手指のみをある程度抽出できる．

片手の位置座標を中心とした楕円体内に範囲限定する方法について述べる．図 3.10 は Kinect で認識したユーザーと，手の位置（図 3.6 の HAND_LEFT と HAND_RIGHT）に丸をつけて表示した結果である．認識されたユーザー領域の各画素に対して，以下の 2 つの値がしきい値以内であれば手指領域候補とする．

- 深度画像上のユーザーの片手の中心座標と各画素の座標間の距離
- ユーザーの片手の中心座標の深度値と各画素の深度値の差分

図 3.10 Kinect で識別されたユーザーと手の例

深度画像上のユーザーの片手の中心座標を(x_0, y_0)，画素iの座標を(x_i, y_i)とすると，その間の距離値t_iは

$$t_i = \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2} \tag{3.3}$$

で求めることができる．また，ユーザーの片手の中心座標の深度値をz_0，画素iの深度値をz_iとすると，その間の差分d_iは

$$d_i = z_i - z_0 \tag{3.4}$$

で求めることができる．深度画像上の各画素を走査していき，式(3.3)，(3.4)を計算して下記の 1 から 3 の条件を満たせば片手領域候補画素として判断する．

1. 現在調べている画素iがユーザーと判断されていれば 2 に進む．
2. 現在調べている画素iのt_iがしきい値以内であれば 3 に進む．

3.　現在調べている画素iのd_iがしきい値以内であれば片手領域候補画素とする．

次に，手領域候補画素群から体積を考慮した片手領域の抽出方法について述べる．Kinectから物体までの深度値zと Kinect の画角から，深度画像の占める横方向の幅l_w，縦方向の幅l_hの推定が可能である．深度画像の幅w，高さhであるとき，深度画像中のzの位置にある画素のもつ縦方向の幅はl_h/h，横方向の幅はl_w/wで表せるため，その画素の占める面積$s(z)$は

$$s(z) = \frac{l_w * l_h}{w * h} \tag{3.5}$$

で求まる．この例を図 3.11 に示す．

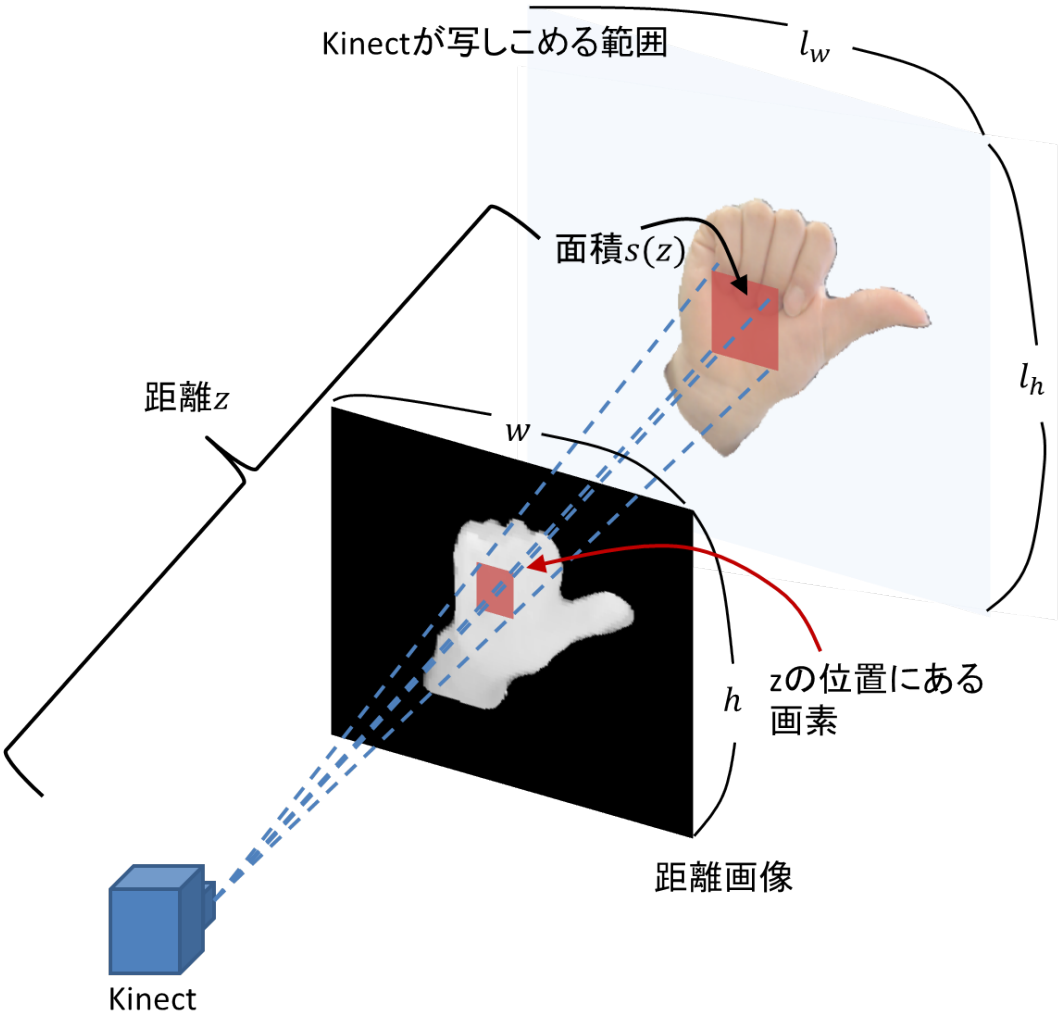

図 3.11 深度画像中のある画素が占める面積

ある画素と，それよりも Kinect から遠い位置にある画素までの深度値をそれぞれz_f，z_nとしたとき，これらの画素の間で形成される四角錐台の体積$V(z_f, z_n)$は，

$$V(z_f, z_n) = \frac{s(z_n) * z_n - s(z_f) * z_f}{3} \tag{3.6}$$

で計算できる．まず，楕円球体内に範囲限定した各画素の持つ深度値を小さい順にソートしたものを配列に格納する．この配列のインデックス番号をj，個数をk，z値をz_jと表す．次に，jを 0 から順に 1 ずつ増やしながら，z_jの位置の画素に対応する四角錐台の体積を計算する．深度値が小さい順にk個の画素が形成する四角錐台群の体積の合計$V(k)$を

$$V(k) = \sum_{j=0}^{k-1} V(z_j, z_{j+1}) \tag{3.7}$$

で計算する．

本研究では，式(3.7)で計算される手指領域の体積$V(k)$は，実際の手の体積の半分程度であり，手型によらず一定であると仮定する．つまり，図 3.12 のように，体積のしきい値V_tを設定し，kを 0 から順に 1 ずつ増やしながら$V(k) \geq V_t$となるkをnとすると，$j = 0,1, \dots, n-1$に対応する画素を手指領域として抽出する．体積のしきい値V_tについては，一般人の手の体積は平均 320[cm³]と報告されており[13]，今回はその半分の 160[cm³]をしきい値とした．

抽出した手領域の各画素の RGB の値を取得し，用意した画像バッファに描画した結果を図 3.13 に示す．顔の前に手指が重なっても，手指のみを抽出することができることがわかる．もう一方の手指でも同様に処理することで，両手の抽出を行う．

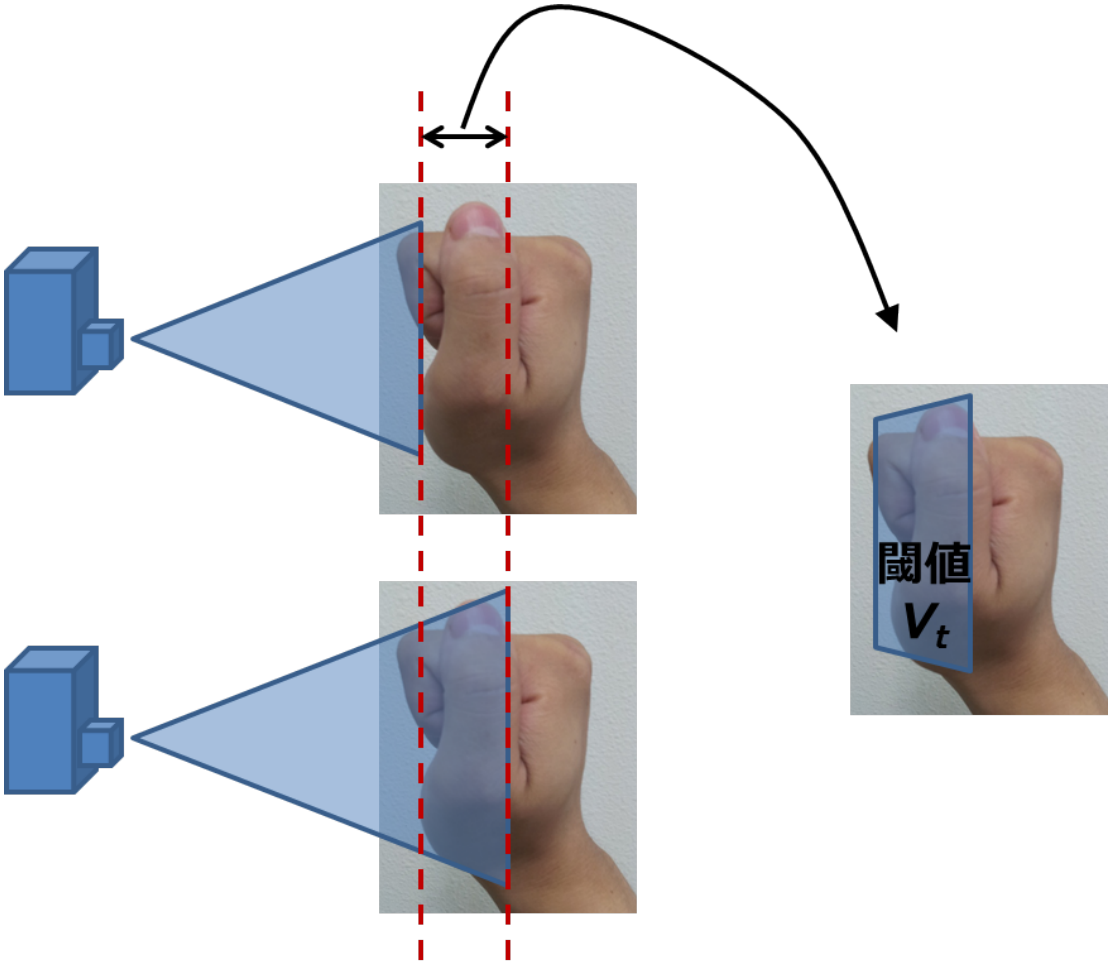

図 3.12 体積を用いた手指領域抽出のイメージ

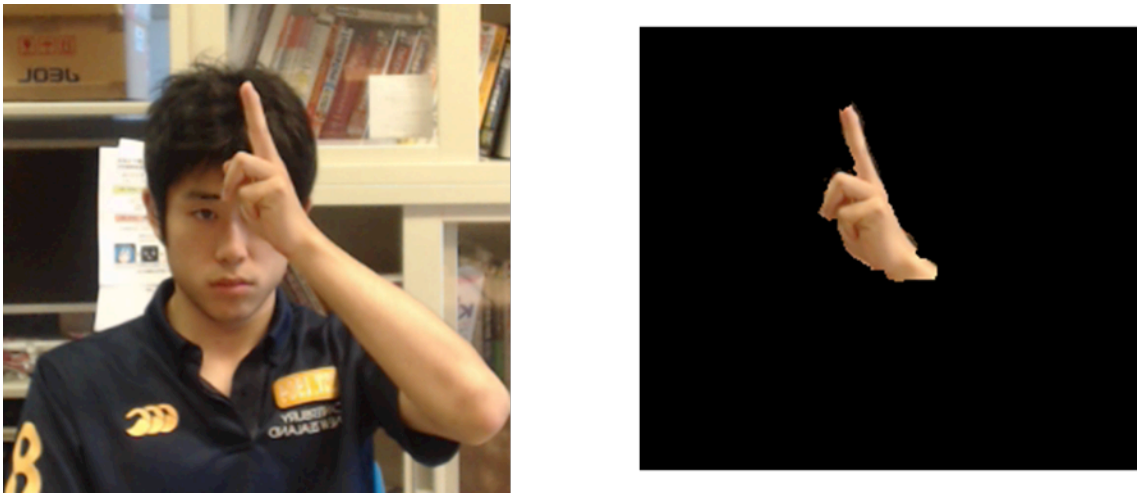

図 3.13 体積を用いた手指領域抽出の結果

3.4.3手指の実写画像の CG への合成

抽出した手指領域の実写画像を CG モデルの対応する部分に合成する方法について述べる．本研究では，手指の実写画像をテクスチャ画像として扱い，CG モデルを描画する画像に対してテクスチャマッピングする．

テクスチャの貼り付けを行うために，テクスチャ画像の座標と，CG モデルを描画する画像上の座標をそれぞれ対応づける．図 3.14 のようにテクスチャ座標系は横軸をs，縦軸をtとする座標系であり，それぞれ 0.0～1.0 の範囲をとる．

図 3.15 のように，画像上の CG モデルの手首の位置座標を(u, v)と仮定する．テクスチャ画像のサイズが$a \times a$とすると，テクスチャ画像を CG モデルの手首に合わせて貼り付けるためには，図 3.16 のように CG の座標$(u - a/2,\ v - a/2,\ 0.0)$，$(u - a/2,\ v + a/2,\ 0.0)$，$(u + a/2,\ v - a/2,\ 0.0)$，$(u + a/2,\ v + a/2,\ 0.0)$で生成される矩形面の頂点それぞれに対して，テクスチャ座標$(0.0,\ 0.0)$，$(0.0,\ 1.0)$，$(1.0,\ 0.0)$，$(1.0,\ 1.0)$を設定する．矩形面のz成分を 0.0 にすることで手指の実写画像を最前面に描画する．テクスチャ画像のサイズについては，手指がすべて画像に入るように 128×128 の大きさにした．画像上の CG モデルの手首の位置座標については，DX ライブラリの関数[2]で得ることができる．左右の手指の実写画像を CG モデルに合成した結果を図 3.17 に示す．

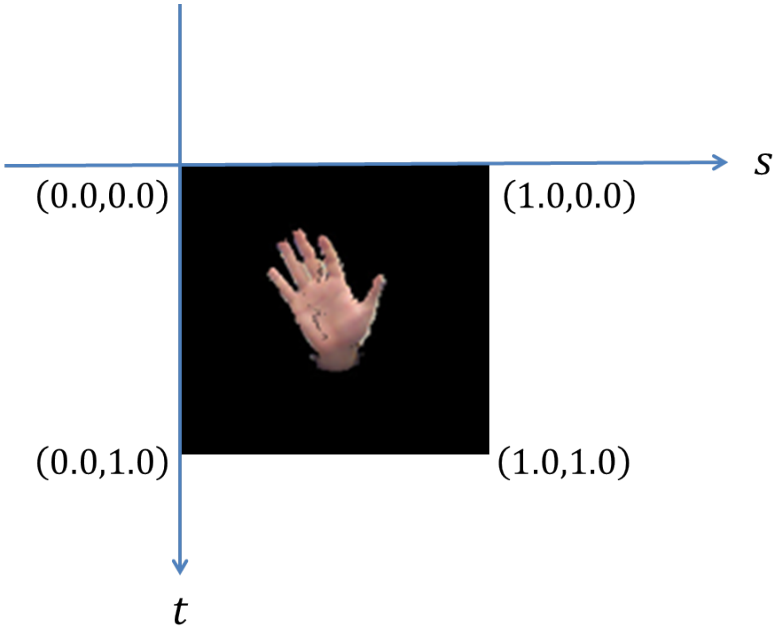

図 3.14 テクスチャ座標系

[2] VECOTR MV1GetFramePosition1：CG モデルのボーンのワールド座標を取得する．
VECTOR ConvWorldPosToScreenPos：ワールド座標を，CG モデルを描画する画像上の座標に変換する．

図 3.15 画像上の手首の位置座標

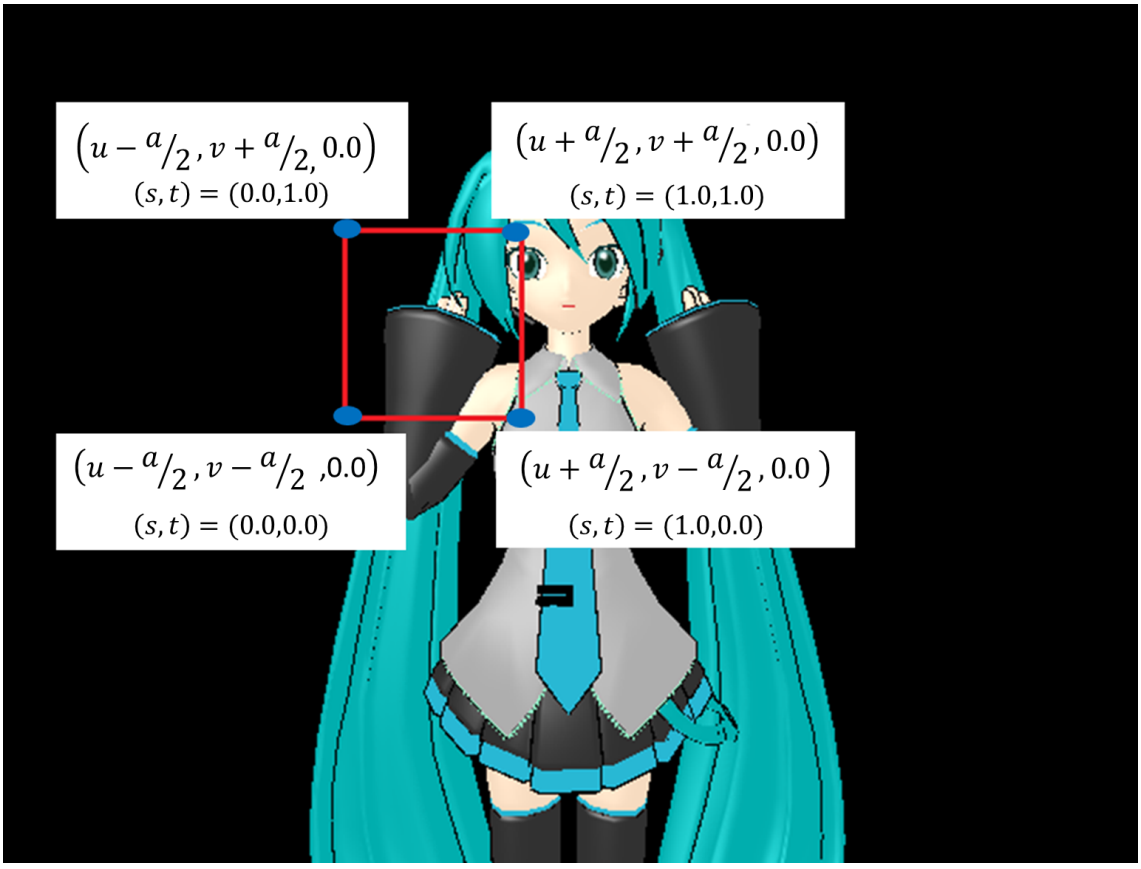

図 3.16 矩形面の頂点

図 3.17 手指の実写画像の合成

3.5　実写画像を用いた非手指動作の表現

3.5.1実写画像を用いた非手指動作の表現の概要

非手指動作の中で主に使われている眉の動き，まばたき，視線，口型も表現することで，文章の種類の区別がしやすくなり，内容を読み取りやすくなることが期待できる．ここでは,Kinect で計測した顔の部位を実写画像から抽出して CG モデルに合成する方法について述べる．

3.5.2顔の部位領域抽出

本研究では，眉，目，口を対象に実写画像から抽出して，眉の動き，まばたき，視線，口型を表現する．そのため，まず，左眉，右眉，左目，右目，口に関して輪郭線を生成するための画像バッファと実写画像から抽出するための画像バッファを用意する．この画像バッファのサイズは実写画像のサイズと同じとする．

Kinect のフェイストラッキングで得られた顔の部位の特徴点を図 3.18 左に示す．次に，図 3.18 右のように 1 つの顔の部位ごとに各画像バッファ中に特徴点間を直線で結んだ輪郭線を描く．そして，輪郭線内の領域の各画素に 1：左眉，2：右眉，3：左目，4：右目，5：口の各部位領域として番号を割り付ける．顔の部位の各領域を図 3.19 に示す．割り付けた番号の各画素の座標を参照して実写画像から RGB の値を取得し，用意した画像バッファに描画することで，左眉，右眉，左目，右目，口の実写画像を生成する．今回は実写画像の大きさは手指の実写画像と同じ 128×128 とした．

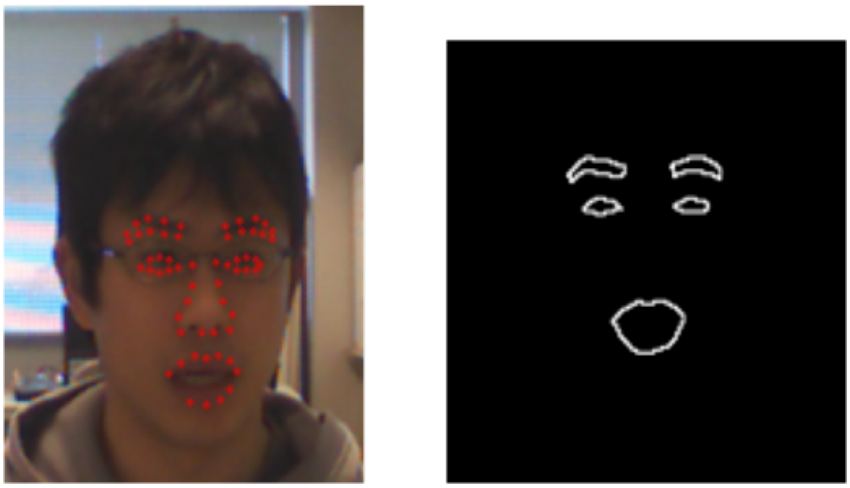

図 3.18 顔の部位の輪郭線生成

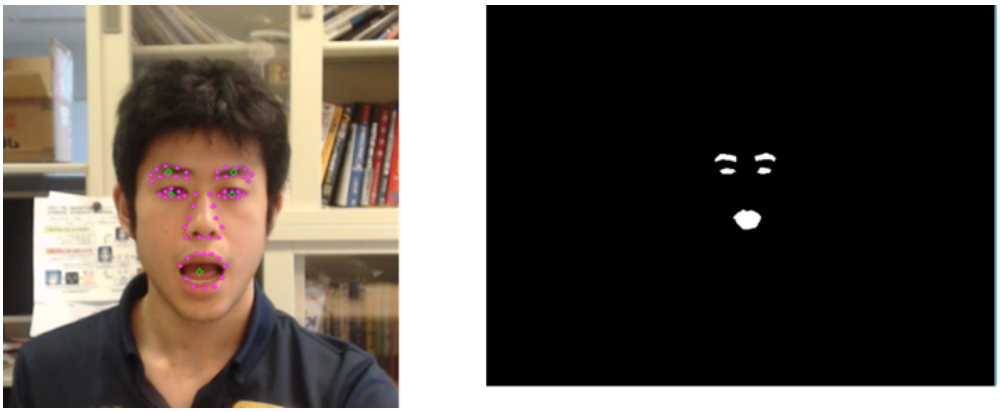

図 3.19 顔の部位の領域

3.5.3顔の部位の実写画像の CG への合成

3.5.2 節で述べた左眉，右眉，左目，右目，口の実写画像を CG モデルの対応する部分に合成する方法について述べる．顔の部位の実写画像を CG に合成するために，前準備として CG モデルの眉，目，口のテクスチャを消去した．合成方法については 3.4.3 節で述べたように，顔の各部位の実写画像をテクスチャ画像として扱い，各部位の中心点から矩形面を生成し，矩形面の各頂点にテクスチャ座標を設定することによって，顔の各部位の実写画像を最前面に描画する．

左目，右目，口に関しては，CG モデルの部位に含まれているため，左目，右目，口の 3 次元座標から画像上の 2 次元座標を推定し，それぞれの座標を中心に実写画像を貼り付けることができる．しかし，左眉，右眉に関しては CG モデルの部位に含まれていないため，左眉，右眉の部位の 3 次元座標は取得できない．そこで，左眉，右眉はそれぞれ左目，右目の上あたりに位置していることを前提に，本研究では眉の実写画像の矩形面の中心はそれぞれ目の中心のx座標，目の中心のy座標+nとする．nは対象の人物の眉，目の大きさ，CG モデルの顔の部位の位置のバランスを考慮して手動で設定した．今回は$n = 10[pixel]$でとした．左眉，右眉，左目，右目，口の実写画像を画像上の CG モデルの対応するところに合成した結果を図 3.20 に示す．

図 3.20 顔の部位の実写画像の合成

3.6　高解像度な実写画像の利用[A3]

3.6.1高解像度な実写画像利用の概要

Kinect で取得可能な実写画像から手指や顔の部位の抽出，CG への合成処理を実装したが，図 3.21 上のように解像度が十分でなく手指や顔の部位の動きの把握が困難であった．そこで，より高解像度な実写画像を用いることで，手指や顔の部位の見やすさの向上を図った（図 3.21 下）．高解像度な実写画像を取得するデバイスとして，Web カメラを併用した．図 3.5 で示したように Kinect の RGB カメラと縦軸が一致するように Web カメラをマウントしたカメラシステムを作成した．しかし，Kinect と Web カメラの位置が異なり両カメラ間で視差が生じるため，画像間の位置のずれがそのため，両カメラ間の座標変換を求めて画像間を対応づける．本研究では，次節に述べるカメラキャリブレーションによる方法で両カメラ間の座標変換を推定する．

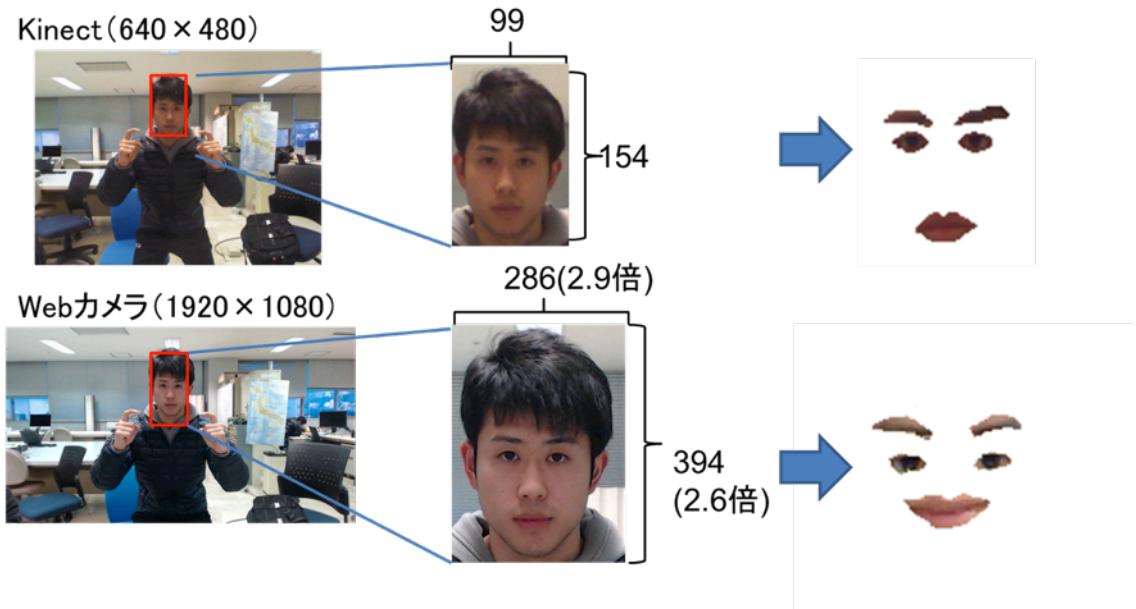

図 3.21 解像度の違いの例（上：Kinect　下：高解像度な Web カメラ）

3.6.2特徴点の座標変換

図 3.22 のように，Kinect 座標系と Web カメラ座標系を一致させるために，カメラキャリブレーションにより両カメラ間の変換を推定する．Kinect と Web カメラを同時にカメラキャリブレーションすることで，ステレオカメラとしての内部パラメータM_p，外部パラメータM_tを求める．カメラキャリブレーションは，位置が既知である多数の 3 次元座標と，その位置をカメラで観察した画像座標の組から，カメラの内部パラメータと外部パラメータを求める計算である．OpenCV に Zhang の手法によるカメラキャリブレーションが実装されているため，今回はこれを利用した．

Kinect は 3 次元座標$\boldsymbol{P_k}$を計測することができる．Kinect 座標系から Web カメラ座標系に座標変換するための外部パラメータ$\boldsymbol{M_t}$と，Web カメラ座標系から画像平面上に射影変換するための内部パラメータ$\boldsymbol{M_p}$が既知である場合，特徴点$\boldsymbol{P_k}$に対応する Web カメラの画像平面上の座標$\boldsymbol{P_w}$は

$$\boldsymbol{P_w = M_p\, M_t\, P_k} \tag{3.6}$$

で計算できる． 実例として Kinect で得られた特徴点を Web カメラ画像上に変換して描画した結果を図 3.23 に示す. このように Kinect の特徴点で構成される手指と顔の部位の領域に対応する Web カメラのより高解像度な実写画像を抽出できるため，3.4 節と 3.5 節で述べた実写画像として Web カメラの実写画像を利用する．

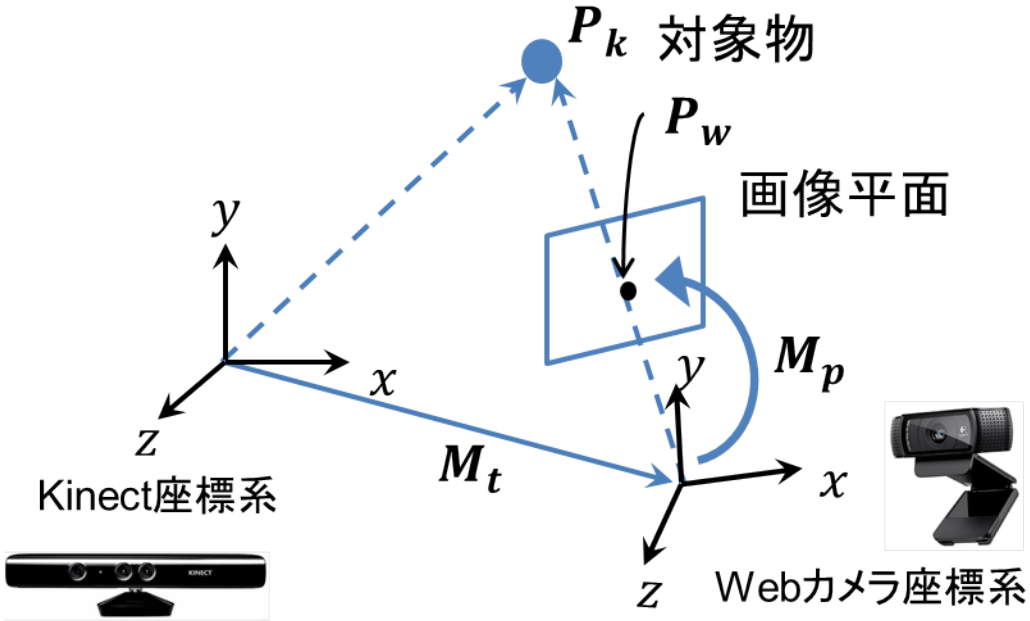

図 3.22 Kinect から Web カメラへの変換

図 3.23 手指・顔の部位の特徴点の座標変換結果の例（左：Kinect　右：Web カメラ）

3.7　実写画像の変換

3.7.1匿名性の向上

2.3 節で述べたように，手指や顔の部位に対応した実写画像を CG モデルに合成すると匿名性を確保できない可能性がある．また，CG の質感を考慮せずに合成すると，違和感が生じる可能性もある．そこで，匿名性を確保しつつ自然な表現を行うために，実写画像の色変換と変形を行った上で CG に合成する．

3.7.2実写画像の色変換

CG モデルに合成する際の違和感を小さくするために，実写画像をイラスト風に変換する．まず，図 3.24 のように顔の部位，手指の実写画像に対してバイラテラルフィルタを適用して輪郭をぼかさずノイズを除去する．次に，ノイズを除去した顔の部位，手指の実写画像を CG モデルに合わせるための色変換を行う．使われている顔の部位は眉，目，口であり，顔の各部位と手指の色を考慮して，実写画像の各画素の RGB の値に応じて黒色（眉，黒目），白色（白目），赤色（口），肌色（肌）の 4 色に階調化する．各色のしきい値は対象の人物に合うように手動で設定し，CG モデルに合わせて各色を変更した（図 3.25 例では黒色を水色，赤色を桃色に変更）．今回は，実写画像にバイラテラルフィルタをカーネルサイズ 5，$\sigma_1 = 50$，$\sigma_2 = 100$のパラメータで適用し，眉，黒目，白目，口，肌の部分の色変換については，表 3.3 に従って変換を行った．

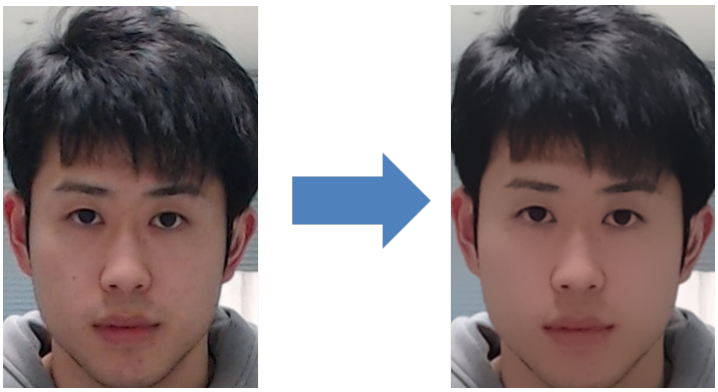

図 3.24 実写画像のノイズ除去

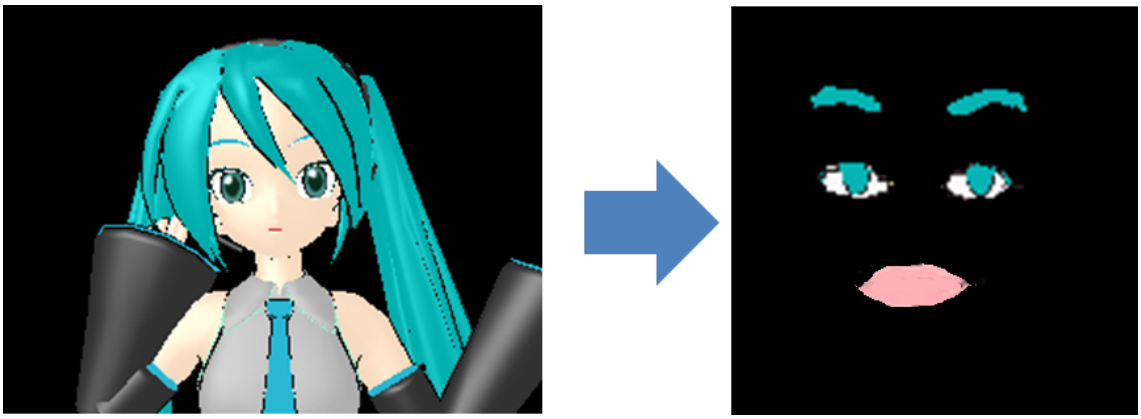

図 3.25 CG モデルに合わせた色変換

表 3.3 色変換の際のパラメータ

	RGB のしきい値	変換後の RGB 値
眉, 黒目	0<R<90，30<G<90，0<B<90	R=7，G=150，B=212
白目	130<R<160，100<G<140，100<B<140	R=255，G=255，B=255
口	140<R<210，80<G<130，60<B<130	R=255，G=180，B=184
肌	170<R<230，110<G<170，110<B<130	R=246，G=217，B=188

3.7.3顔の部位の実写画像の変形

色を変換しただけでは本人が特定される可能性がある．また，違和感を小さくするために，顔の部位を CG モデルに合わせて変形することが望ましい．本研究では，抽出した顔の部位の実写画像をz軸まわりの回転，およびxy方向の拡大縮小を行い，CG モデルに貼り付ける．

3.5.3 節で述べたように顔の部位の実写画像はテクスチャ画像として合成しているため，それぞれの矩形面の頂点の座標にz軸まわりの回転，xy方向の拡大縮小の変換行列を適用することで実写画像を変形することができる．ただし，テクスチャ画像を回転や拡大縮小するためには，画面の左上を原点としてテクスチャ画像の中心を原点に合わせる必要がある．テクスチャ画像のサイズは$a \times a$とすると，矩形面の頂点をそれぞれ$(-a/2, -a/2, 0.0)$，$(a/2, -a/2, 0.0)$，$(-a/2, a/2, 0.0)$，$(a/2, a/2, 0.0)$に設定することでテクスチャ画像の中心を原点に合わせる．次に，変換前の頂点を点$\boldsymbol{P}$，回転行列を$\boldsymbol{R}$，拡大縮小行列を$\boldsymbol{S}$とすると，変換後の頂点$\boldsymbol{P'}$は

$$\boldsymbol{P'} = \boldsymbol{RSP} \tag{3.6}$$

で求めることができる．このようにして，図 3.26 のように実写画像を変形する．最後に，3.5.3 節で述べた方法で画面上の CG モデルの対応するところに貼り付ける．回転，拡大縮小の際のパラメータは，対象者の顔の部位の大きさ，カメラからの距離に合わせて手動で設定した．色変換と変形を行った結果を図 3.27 に示す．

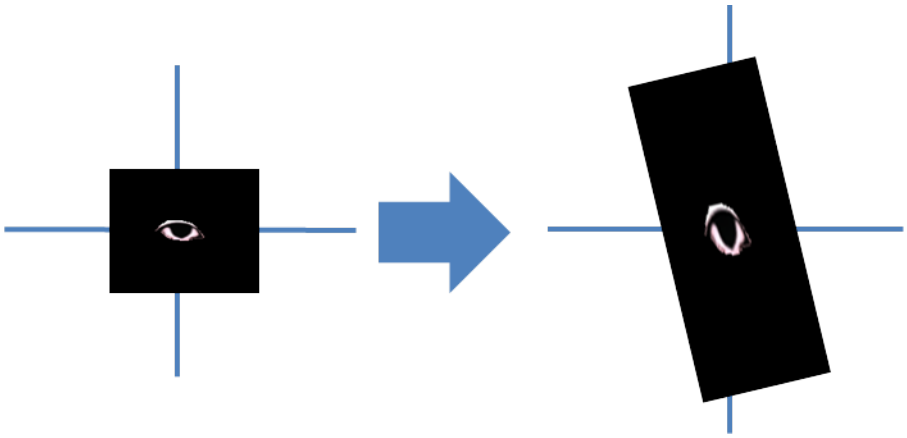

図 3.26 顔の部位の実写画像の変形

図 3.27 色変換と変形後の結果

第4章 匿名手話映像の評価実験[A6]

4.1 匿名手話映像生成の処理速度とデモンストレーション

第 3 章で述べた処理方法を実装し，表 3.1 で示したハードウェア構成で匿名手話映像生成を行った．手話者を撮影してから匿名手話映像が表示されるまでに要する処理時間は，実写表現で約 36ms，変換表現で約 40ms であった．フレームレート（匿名手話映像の更新速度）に換算するとそれぞれ約 28fps，25fps である [3].

実際に PEPNet-Japan シンポジウムの機器展示コーナーで，実写表現，変換表現で生成した匿名手話映像のデモンストレーションを行い，聴覚障害者やその支援に携わる方々に試してもらった．体験者からは「面白い」，「実用化してほしい」，「顔の部位が CG モデルと合わないから違和感がある」などの意見をいただくことができた[A2][A4]．体験者によっては，動きの誤認識や手指，顔の部位の色変換が上手くいかないなどの問題も見られたが，撮影環境の調整や利用者に合わせてしきい値設定でこれらの問題は解決できる．動きの誤認識については，認識範囲の妨げとなる障害物を近くに置かずに Kinect の正面に立つことで防ぐことができる．色変換については，ユーザーの肌や唇などに合わせてしきい値パラメータ設定することによって解決できる．実写表現で生成した匿名手話映像の様子を図 4.1，変換表現で生成した匿名手話映像の様子を図 4.2 に示す．

本デモンストレーションを通して，筆者以外の人物でも匿名手話映像を生成可能であることを確認できた．

図 4.1 実写表現で生成した匿名手話映像の様子

[3] Kinect と Web カメラのフレームレートはフレームバッファを更新する速度であり，その速度は 30fps である．本方法では，1 フレームの匿名手話映像を描画する際に最新のフレームバッファを利用して非同期で処理を行うため，カメラのフレームレートの制約は受けない．

図 4.2 変換表現で生成した匿名手話映像の様子

4.2 評価実験の概要

3 章で述べた実写表現と変換表現で生成した匿名手話映像について，匿名性の確保，手話の読み取りやすさについて評価実験を行った．また，生成した映像の違和感についても合わせて評価を行った．

匿名性を確保するためには非手指動作で活用される顔をうまく隠すことが重要であると考えられる．目，眉および口のみと抽出してそのまま CG モデルに合成する実写表現では，2.3 節で述べたように顔の識別能力が高い手話者に対して十分な匿名性を確保できない可能性がある．そこで，著名な人物の顔画像を元に生成した実写表現と変換表現の画像を生成し，匿名性が確保されているか，違和感がどの程度あるかについて実験を行う．

一方，匿名性が確保されていたとしても，手話表現を読み取ることが困難であればコミュニケーションには利用できない．実写表現と変換表現で文を表現した匿名手話映像を生成し．読み取りやすさについての定量的な評価と主観的な評価を行った結果について述べる．

4.3 節で匿名性の確保に関する実験と結果について述べ，4.4 節で匿名手話映像の読み取りに関する実験と結果について述べる．両実験で得られた結果をもとに本方法で生成した匿名手話映像の特徴や有効性について考察する．

4.3 匿名性の確保に関する実験

4.3.1実験目的

実写表現と変換表現で生成した画像について匿名性の確保に関する実験を行った.手話単語の表現方法や,手指や腕の動かし方の癖などで手話者が特定されてしまう可能性があるが,本研究では特に本人が特定されやすい非手指動作で用いられる顔の部位を対象に実写表現と変換表現で処理した顔画像について匿名性の評価を行った.

匿名性の評価方法は，2.3 節で述べた顔画像に対するプライバシー保護処理に対して，親密度，特徴度別に ID 可到達性を算出して定量的評価する方法がある．対象の人物をどのくらい知っているかの指標となる親密度と,人物の顔が印象に残るかどうか指標の特徴度について匿名性の確保の結果が変わると考えられるため,実験に使用する対象人物の顔画像に対して親密度，特徴度を調査し ID 可到達性を求めることによって匿名性の評価を行う.

4.3.2 節で匿名性の評価に用いる ID 可到達性について説明し,4.3.3 節で実験方法,4.3.4 から 4.3.6 節で実験結果について述べる.

4.3.2匿名性の評価方法[6]

2.3 節で述べたように，本研究の匿名手話映像を生成する処理は一種のプライバシー保護処理であると考えられる.匿名性の確保を定量的評価する指標として ID 可到達性を用いる.あるプライバシー保護処理を施した異なる顔画像を観察者にn回提示し，正しく同定できた回数をn'とした場合，ID 可到達性Nは

$$N = \frac{n'}{n} \tag{4.1}$$

で計算する．プライバシー保護能力が高い処理ほどNが小さくなる．この計算を親密度f，特徴度cごとに行った結果を評価に用いる.

$$U_f = \frac{u'_f}{u_f} \tag{4.2}$$

$$V_c = \frac{v'_c}{v_c} \tag{4.3}$$

式(4.2)については，u_fがある親密度における処理後の画像を観察者に提示した回数，u'_fがそのうち同定できた回数，U_fはその ID 可到達性である．同様に，特徴度cの ID 可到達性V_cについても式(4.3)で計算する．被写体と観察者の親密度と，被写体の特徴度を考慮し，プライバシー保護処理の有効性を定量的に評価することができる．

本研究では，プライバシー保護処理を実写表現と変換表現とし，ID 可到達性を評価する．ID 可到達性が低ければ匿名性が確保されていると判定できる．

4.3.3匿名性の確保に関する実験の方法

まず，芸能人やアスリートを中心に 10 名の著名な人物を選び，各人物について 5 枚ずつ合計 50 枚の画像をウェブから収集した．次に，図 4.3 のように 6 名の人物の参照画像をランダムに選び，この中の 1 名の画像（参照画像とは異なる画像）を処理した結果を対象画像としたシートを作成した[4]．著名な人物の顔画像から実写表現と変換表現の顔画像を生成する方法については付録 A1 で詳述する．被験者は同シートを見て実写表現あるいは変換表現を施した対象画像の元となった人物を回答する．被験者は実写表現と変換表現について各 5 回ずつ計 10 回の回答をしてもらった．また，回答の自信について 5 段階（1：全く自信がない～3：どちらともいえない～5：とても自信がある），対象画像の違和感について 5 段階（1：そう思わない～3：どちらともいえない～5：そう思う）で評価してもらった．

上記の実験を終えた後に，実験に用いた 10 名の人物に対する親密度fと特徴度cを調査した．そのシートを図 4.4 に示す．親密度fは 3 段階（1：全く知らない，2：見たことはある，3：よく知っている），特徴度cは 5 段階（1：全く特徴的だと思わない，2：あまり特徴的だと思わない，3：どちらともいえない，4：やや特徴的だと思う，5：とても特徴的だと思う）で評価してもらった．なお，この実験に協力してもらった被験者は，筑波技術大学に在籍する聴覚障害をもつ大学生 20 名である．

[4] 図 4.1 のシートは，撮影許可をいただいた一般の方の顔画像を差し替えたものだが，実験では著名な人物の顔画像を使用した．

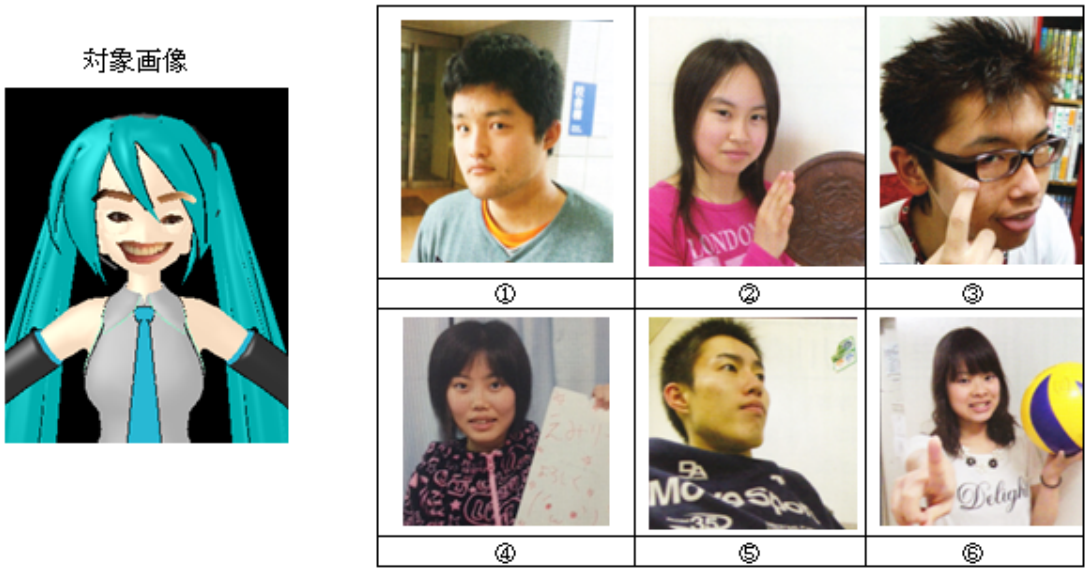

図 4.3 匿名性に関する調査シートの例 1

手話映像の匿名性の確保に関する調査

引き続き，この用紙の質問にご回答お願いします．
左側の画像で示す人物について，それぞれ 2 つの質問についてあてはまるもに○をつけてください．

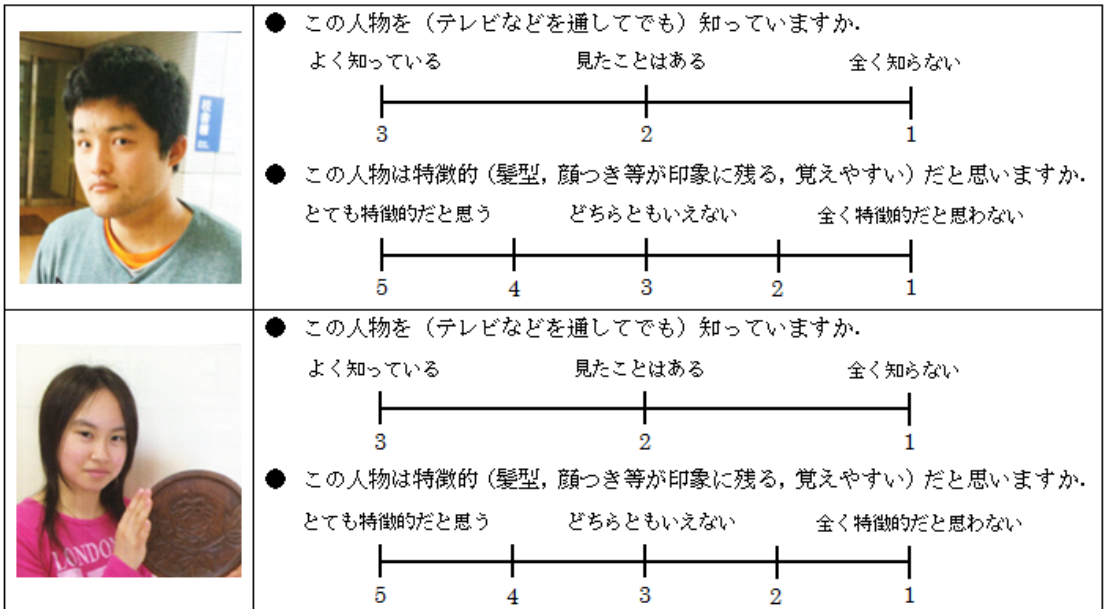

図 4.4 匿名性に関する調査シートの例 2

4.3.4匿名性に関する結果

実写表現と変換表現について，親密度fに対する ID 可到達性を図 4.5，特徴度cに対する ID 可到達性を図 4.6 に示す．図 4.5 の縦軸は親密度f，横軸は ID 可到達性，図 4.6 の縦軸は特徴度c，縦軸は ID 可到達性である．棒グラフは平均値，エラーバーは標準偏差を表す．

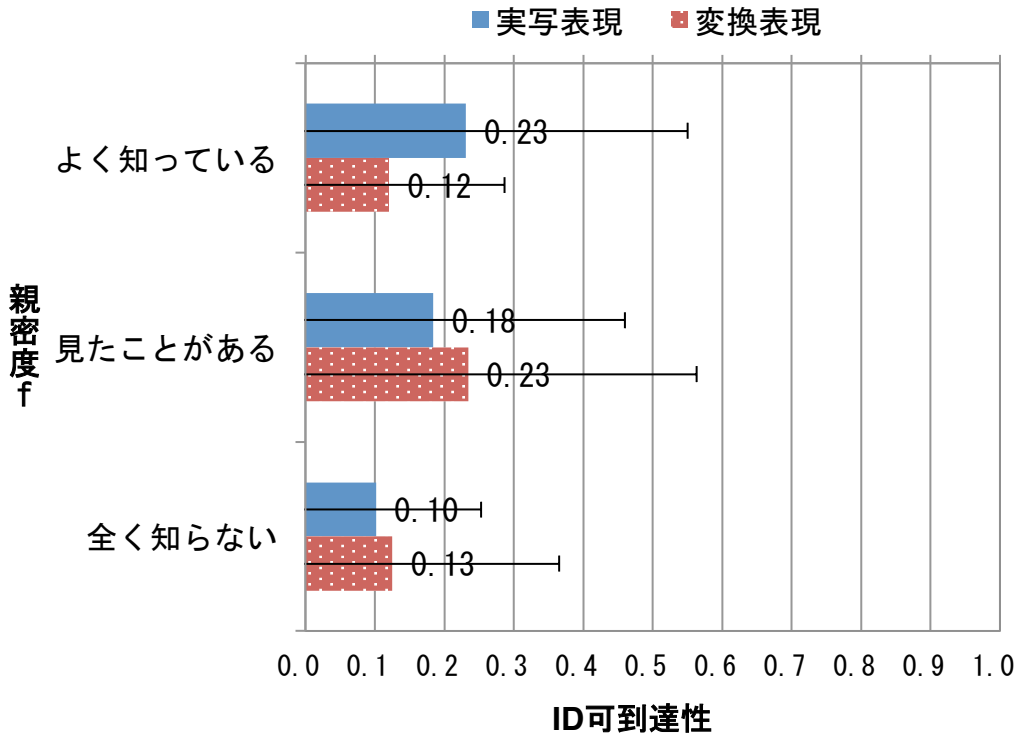

図 4.5 親密度fに対する ID 可到達性の結果

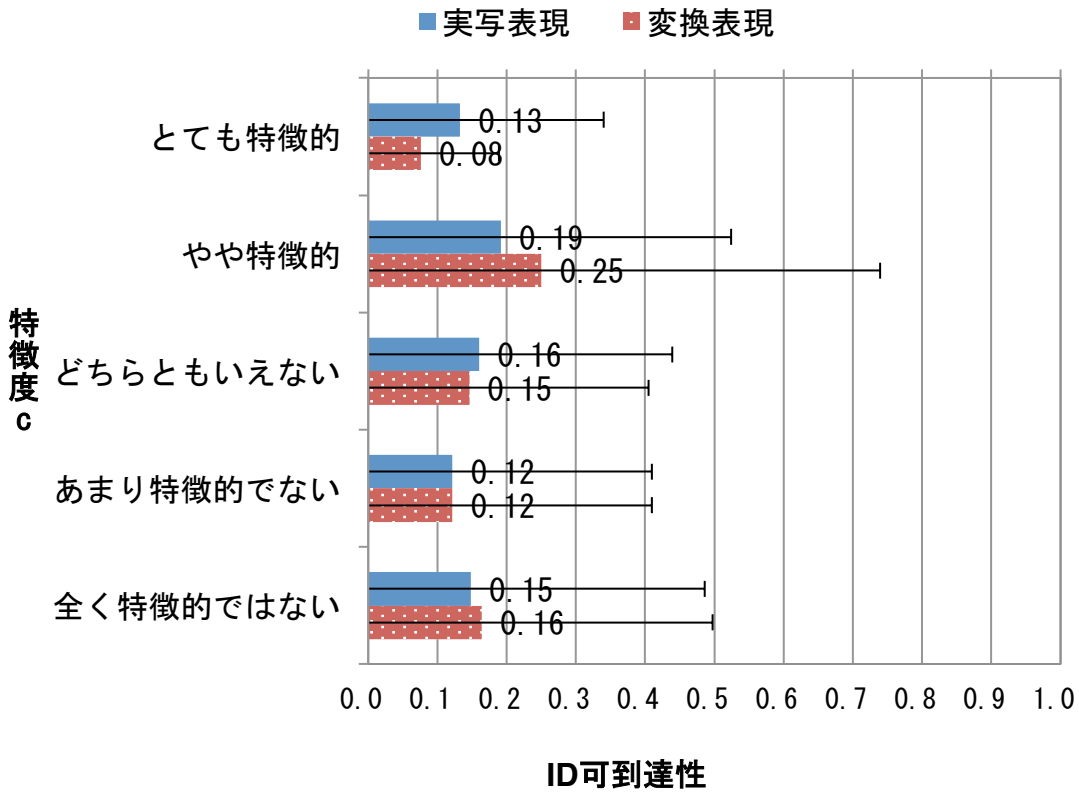

図 4.6 特徴度cに対する ID 可到達性の結果

親密度f，表現方法の要因によって ID 可到達性に影響があるかどうかを調べるために，2 要因の分散分析を行った結果を表 4.1 に示す.いずれの要因も有意な差は見られなかった.
特徴度c，表現方法の要因によって ID 可到達性に影響があるかどうかを調べるために，2 要因の分散分析を行った結果を表 4.2 に示す．いずれの要因も有意な差は見られなかった.

表 4.1 親密度fの ID 可到達性の分散分析表

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
表現方法	2	0.003	0.00153	0.036	0.965
親密度	2	0.118	0.05900	1.378	0.255
表現方法：親密度	4	0.197	0.04934	1.153	0.334
Residuals	163	6.977	0.04280		

表 4.2 特徴度cの ID 可到達性の分散分析表

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
表現方法	2	0.013	0.00670	0.122	0.886
特徴度	4	0.165	0.04119	0.748	0.560
表現方法：特徴度	8	0.158	0.01980	0.360	0.941
Residuals	235	12.937	0.05505		

全体の親密度fの ID 可到達性の平均は約 0.165，全体の特徴度cの ID 可到達性は約 0.151 であった．一方，本実験で用いたシートのように 6 枚の参照画像から 1 枚の正解を選び出す確率は約 0.16 である．つまり，今回の結果は無作為に選んだ画像が正解する確率とほぼ同程度であり，親密度，特徴度に関係なく人物の匿名性は十分に確保されていると考える．

4.3.5回答の自信に関する結果

同定できた場合，同定できなかった場合のそれぞれの自信の結果を図 4.7 に示す．縦軸は同定の可否，横軸は自信の 5 段階評価を表し，棒グラフは平均値，エラーバーは標準偏差を表す．表現方法，同定の可否の要因によって自信に影響があるかどうかを調べるために 2 要因の分散分析を行った結果，5%水準で表現方法間に有意な差が見られた（表 4.3）．変換表現は，処理した顔画像の元となった人物を回答する際に同定できたかどうかに関わらず，実写表現より自信が低いということがわかった．

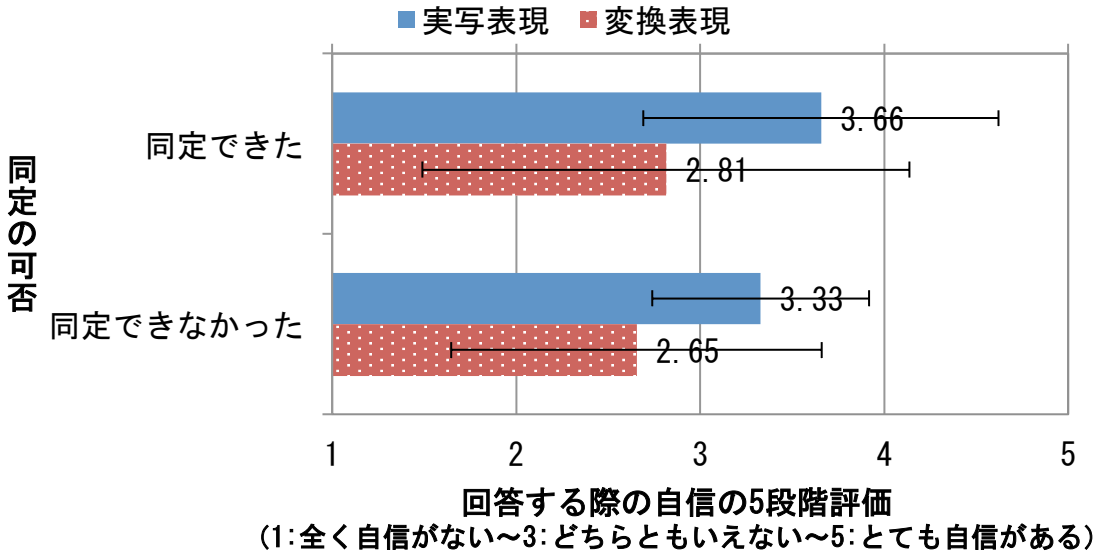

図 4.7 回答する際の自信の結果

表 4.3　自信の分散分析表

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
表現方法	1	6.59	6.589	6.715	0.0118	*
同定の可否	1	1.86	1.861	1.897	0.1732	
表現方法:同定の可否	1	0.78	0.784	0.799	0.3747	
Residuals	64	62.79	0.981			

*：p<0.05

4.3.6違和感に関する結果

違和感の評価結果を図 4.8 に示す．縦軸は表現方法，横軸は違和感に対する主観評価を表し，棒グラフは平均値，エラーバーは標準準偏差を表す．

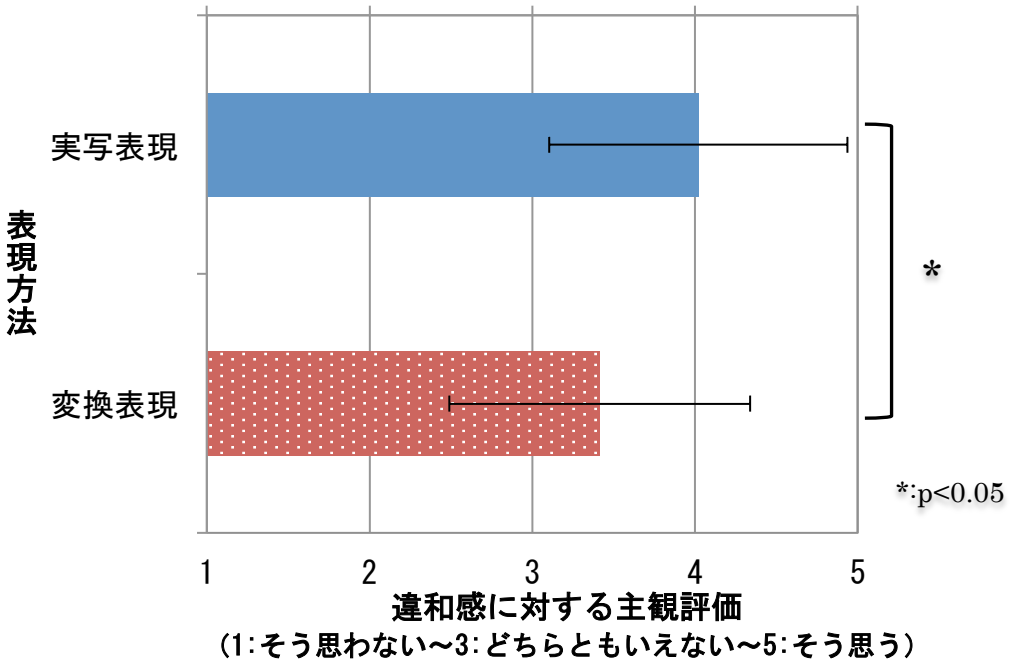

図 4.8 顔画像における違和感の主観評価の結果

実写表現と変換表現のそれぞれの違和感の5段階評価の平均値は約4.02,3.41であった．表現方法の要因によって違和感に影響があるかどうかを調べるために，t検定を行った結果，5%水準で有意な差がみられた(p=0.043)．両処理に関しても違和感が少ないとは言えないが，顔画像において変換表現は実写表現と比べて違和感を減らすのに有効であることがわかった．

4.4　手話の読み取りに関する実験

4.4.1実験目的

実写表現と変換表現で生成した匿名手話映像について，読み取りができるかどうかを調査するための実験を行う．本実験では単文と長文に分けて評価実験を行った．単文については肯定文や否定文などの文の種類を用意し，読み取れた内容の成績による定量的評価と，手話のわかりやすさ，手指動作，非手指動作の見やすさについて主観的評価を行った結果について述べる．

4.4.2匿名手話映像の内容

実写表現と変換表現で生成した匿名手話映像の読み取りについて評価を行うために，単文と長文に分けて実験を行った．単文については肯定文，否定文，疑問文および命令文の 4 種類について，表 4.4 で示すように種類ごとに 2 個ずつ合計 8 個の単文を使用した．長文については表 4.5 で示すように手話表現で約 30 秒間になる内容の異なる 2 個の文を使用した．こちらも単文と同様に実写表現と変換表現で合計 4 種類の匿名手話映像を生成した．これらの匿名手話映像を用いて映像を被験者に読み取ってもらう実験を行った．

表 4.4 実験に使用した単文の内容

文の種類	内容
肯定文	私は今朝 5 時に起きた.
	昨日は暑かった.
否定文	私はコーヒーを飲まない.
	彼は図書館へ行かなかった.
疑問文	あなたは野球が好き？
	彼女は昨日学校に来た？
命令文	携帯電話を使ってはいけない.
	掃除をしなさい.

表 4.5 実験に使用した長文の内容

文章番号	内容
長文 1	今日は雨ですね．でも週末はとてもいい天気になるようです．私は遊園地に行くのですが，あなたはどこにでかけるつもりですか.
長文 2	私は小学校の時，学校が終わった後すぐに家に帰っておやつをたべていました．おやつを食べ終わった後は，公園で友達と遊びました.

4.4.3内容の読み取りに関する実験の方法

本実験では，筑波技術大学に在籍する聴覚障害をもつ大学生 18 名を対象に，ノートパソコン上に単文，長文の匿名手話映像を提示して読み取ってもらう実験を行った．15 インチのディスプレイに匿名手話映像をフルスクリーンで表示し，被験者はノートパソコンから 50cm ほど離れて見てもらった．まず，単文に関して，各被験者で文の種類（肯定部，否定文，疑問文および命令文）×表現方法（実写表現および変換表現）の条件，順序の組み合わせに被験者間で偏りが出ないようにランダムに匿名手話映像を提示する．被験者には，同じ条件の映像を 2 回見てもらい,読み取った内容の文を日本語で用紙に記入してもらう．その後，以下の項目に関して 5 段階（1：そう思わない，2：ややそう思わない，3：どちらともいえない，4：ややそう思う，5：そう思う）で回答してもらう.

- 手話はわかりやすい
- 手・指の動きは見やすい
- 腕の動きは見やすい
- 頭の動きは見やすい
- 目の動き（瞬き，視線等）は見やすい
- 口の形は見やすい
- 眉の動きは見やすい
- 話者の気持ちが伝わってきた
- 違和感がある

この読み取り，内容の記入，項目への回答の流れを 8 個の匿名手話映像について行ってもらった.

次に，長文の匿名手話映像も単文と同様に各被験者で文章の種類×表現方法の条件，順序の組み合わせに被験者間で偏りが出ないようにランダムに匿名手話映像を提示する．被験者には，同じ条件の映像を 2 回見てもらい，各長文に対して表 4.6 で示した内容に関する質問項目に回答してもらう.

表 4.6 長文の内容に対する質問項目と正答

文章番号	質問項目	正答
長文 1	今日はどんな天気でしたか.	雨
	私はどこへ行くのでしょうか.	遊園地
長文 2	家に帰った後何をしていたのでしょうか.	おやつを食べた
	どこで遊んでいましたか.	公園

4.4.4読み取りの評価方法

まず，単文の読み取りに関する評価方法について述べる．匿名手話映像で読み取った文章の内容の成績を 10 点満点とし，単語と助詞の観点で採点を行う．単語，助詞の区別については，次の通りであり，単語を一重線，助詞を二重線とする．

1. 私 は 今朝 5 時 に 起き た.
2. 昨日 は 暑 かった.
3. 私 は コーヒー を 飲ま ない.
4. 彼 は 図書館 へ 行か な かった.
5. あなた は 野球 が 好き ？
6. 彼女 は 昨日 学校 に 来 た ？
7. 携帯電話 を 使っては いけない.
8. 掃除 を しな さい.

それぞれ単語と助詞の数が異なるので，正答率を求めて 10 点満点に計算し直す．例として 1.では，単語が 3 個，助詞が 3 個当たった場合の正答率の計算方法は，6/7*10=8.57[点]となる．

次に，長文の読み取りに関する評価方法について述べる．1 つの匿名手話映像の内容に関する 2 つの質問項目を設けており，それらの質問に対する回答の正答率を求める．質問に対する正答は表 4.5 で示した通りである．

4.4.5単文の読み取りの成績

単文の読み取りの成績を図 4.9 に示す．縦軸は文の種類，横軸は単文の読み取り成績を表し，棒グラフは平均値，エラーバーは標準偏差を表す．

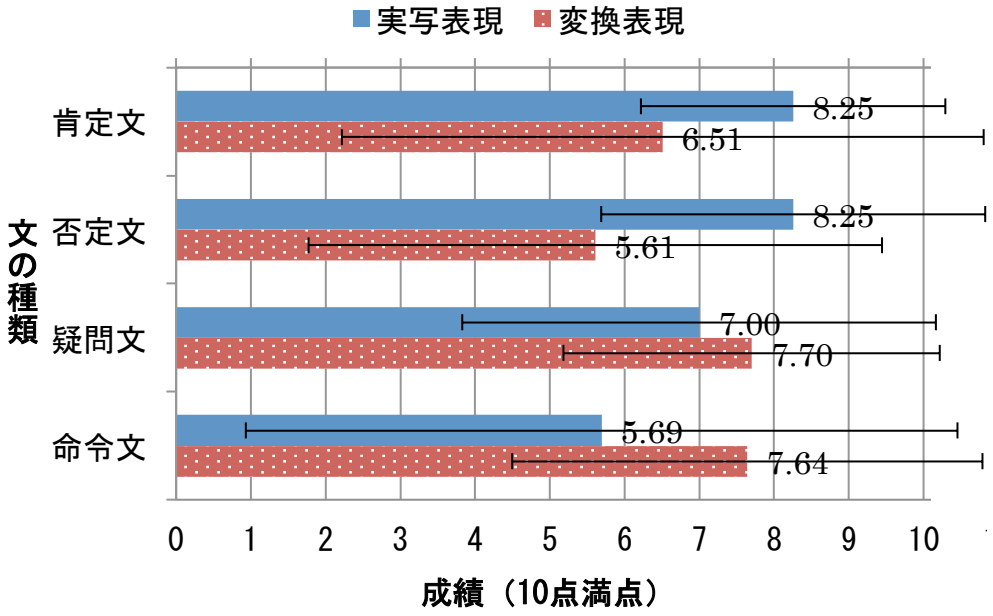

図 4.9 単文の読み取りの成績

表現方法，文の種類の要因によって単文の読み取りの成績に影響があるかどうかを調べるために，2 要因の分散分析を行った結果を表 4.7 に示す．5%水準で表現方法と文の種類の交互作用が有意であった．続いて多重比較を行ったが，いずれの要因も有意な差は見られなかった．全体の読み取りの成績は約 7.08[点]であった．

表 4.7 単文の読み取りの成績の分散分析表

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
表現方法	1	6.9	6.86	0.591	0.4434	
文の種類	3	12.8	4.26	0.367	0.7768	
表現方法:文の種類	3	122.0	40.67	3.503	0.0173	*
Residuals	136	1578.9	11.61			

*：$p<0.05$

4.4.6単文の読み取りの主観評価結果

匿名手話映像に対する評価を図 4.10 に示す．縦軸は評価項目，横軸は評価項目に対する主観評価である．棒グラフは平均値，エラーバーは標準偏差を表す．

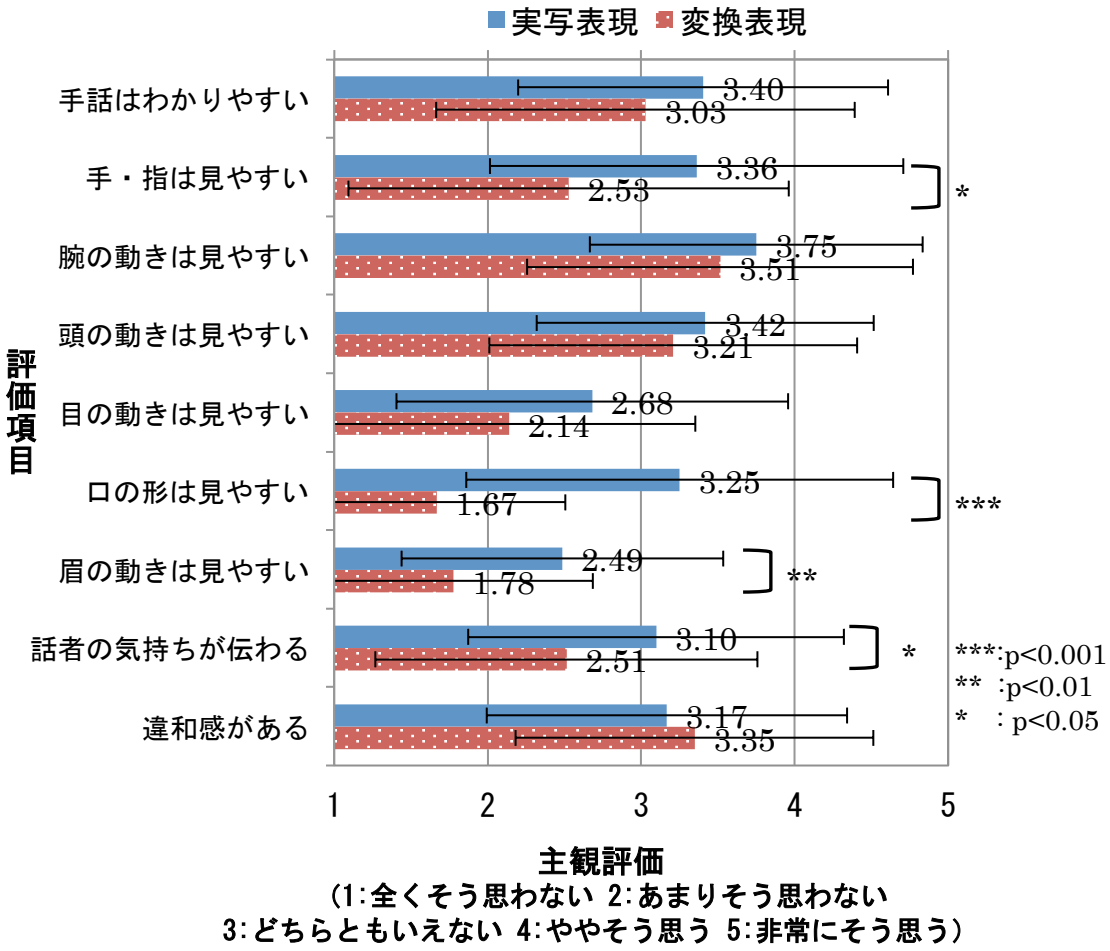

図 4.10 匿名手話映像に対する主観評価の結果

「手話はわかりやすい」「手・指は見やすい」「腕の動きは見やすい」「頭の動きは見やすい」「目の動きは見やすい」「口の形は見やすい」「眉の動きは見やすい」「話者の気持ちが伝わってきた」「違和感がある」の項目に関して，いずれも高い評価を得られたとは言えない．表現方法の要因によってそれぞれの質問に対する評価に影響があるかどうかを調べるために，t検定を行った結果，「話者の気持ちが伝わってきた」では 5%水準（$p = 0.036$），「口の形は見やすい」では 0.1%水準（$p = 6.78\times10^{-7}$），「手・指は見やすい」「眉の動きは見やすい」では 1%水準（それぞれ$p = 0.009$，$P = 0.003$）で有意な差が見られた．

結果として表現方法によって手話のわかりやすさに差はないということがわかった．また，手話表現に関わる非手指動作に関しては，部分的に実写表現の方がよいという結果になった．違和感に関しては．4.3.6 節で述べたように顔画像を対象にした実験では表現方法によって差があったが，映像を読み取る実験ではいずれの表現方法も同程度であった．

被験者に自由記述してもらった要望や意見を表 4.8 に示す．

表 4.8 匿名手話映像に対する要望や意見

①	指の動き（細かい動き）は，変換後同じ色なので見にくい．大きな動きならわかるが，細かい動きや指の動きを見やすくしてほしい．
②	変換前は口の形，手話，手指の形がはっきりしているので何を言っているかがわかる．しかし，変換後は口の形，手指の形が同一色になってしまっているため，きちんと読み取れない．ただ，腕を使う手話はわかる．
③	疑問を感じたときのはもう少し眉をひそめた方がいい．
④	口と指がもう少しはっきりしていれば読み取れる確率は上がると思う．
⑤	手の形はイラストの方がわかりやすい時もあれば写真の方がわかりやすい時もある．口の形は写真（というより映像？）の方がかなりわかりやすかったです．
⑥	指，口の形がきれいにみれない．途中で手がとぎれているのか気になった．手は CG の方が好き．
⑦	顔の動き，腕の動きは大体把握できるが，手や指の形がわかりにくかった．手や指の形がはっきりするとさらに伝わると思います．あと，表情を読み取るのも難しかった．

4.4.7長文の読み取り結果

長文の内容に対する質問の正答率を図 4.11 に示す．縦軸は表現方法，横軸は正答率を表す．棒グラフは平均値，エラーバーは標準偏差を表す．

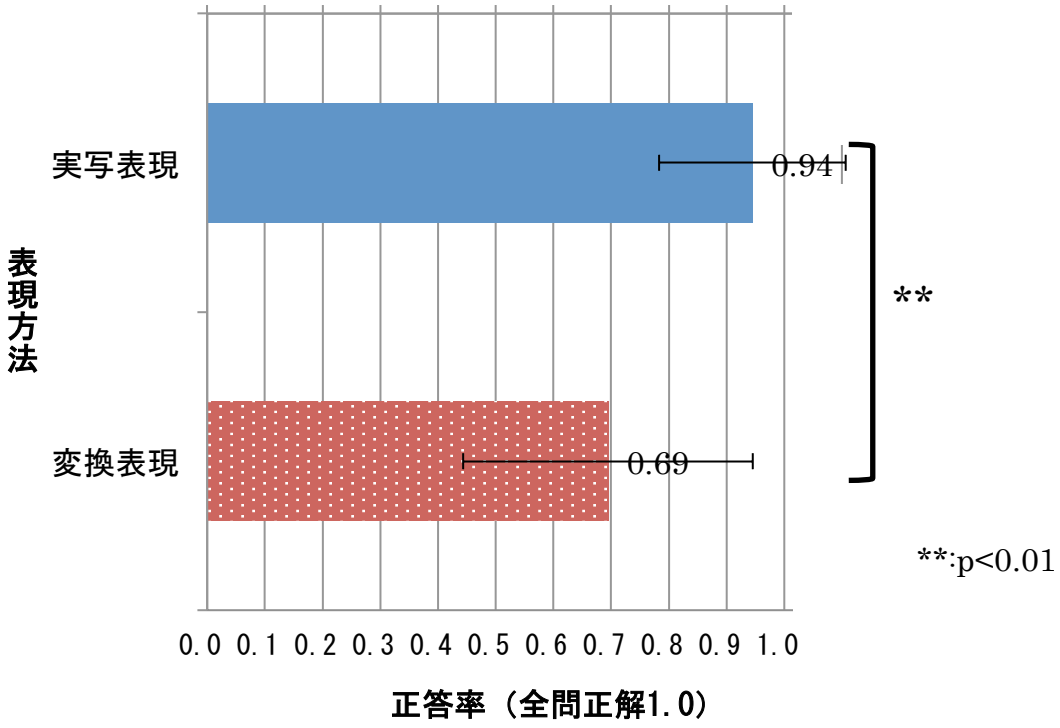

図 4.11 長文の内容に対する質問の正答率の結果

表現方法の要因によって長文の内容に対する質問の正答率に影響があるかどうかを調べるために，*t*検定を行った結果，1%水準で有意な差がみられた(P=0.003)．つまり，実写表現の方が変換表現よりも長文を読み取りやすいことがわかった．また，実写表現の場合の正答率が 9 割以上であり，内容をほぼ理解できていることがわかった．

4.5 考察

4.5.1処理速度に関する考察

手話者を撮影してから匿名手話映像を表示するまでに，実写表現では約 36ms，変換表現では約 40ms の処理時間を要することがわかった．手話通信において 660ms 以上の遅延時間が生じると自由対話に支障が生じると報告されており[14]，本研究で提案した匿名手話映像の生成法では通信遅延が 620ms 未満であれば対話コミュニケーションに支障はないと考えられる．しかし，同報告では手話の実写映像による評価であり，本方法で生成した匿名手話映像で双方向コミュニケーションを行った場合も同様かどうかは自明ではない．手話による実際のコミュニケーション場面に匿名手話映像を適用して，遅延や切断などのネットワークに起因する特性，ならびに対話しやすさや伝達内容の正確性などについて評価を行い，実用化にあたっての課題と解決方法を明らかにする必要がある．

4.5.2匿名性に関する考察

実写表現と変換表現によって匿名化された顔画像については，いずれも匿名性が確保されていることがわかった．また，変換表現は実写表現よりも回答する際の自信が低かったことから，顔の部位の形と色を変換する変換表現はさらに匿名性を向上させることができる可能性が高い．今回の匿名性に関する実験では顔画像のみを対象にしたが，手指に障害を持った人や肌の色，性別などの情報を隠すことができるか明らかではない．また，手話は本来動きをともなうものであり，手話者が顔見知りである場合は単語表現，手指や腕の動きによって本人が特定される可能性がある．これらの課題に対して匿名性を確保できるかどうかについて検討を重ねる余地が残されている．

4.5.3内容の読み取りに関する考察

図 4.10 で示した単文の内容の読み取り実験の主観評価結果では，手指動作に関係する手・指の見やすさ，非手指動作に関係する口の形と眉の動きの見やすさ，話者の気持ちが伝わるかの各設問について実写表現の方が高評価であった．一方，単文の読み取り成績を分析した結果，表現方法と文の種類の間で交互作用が見られた．多重比較を行った結果からはいずれの組み合わせでも有意な差は見られなかった．成績の分析結果からこれらの差について論じることはできないが，標準偏差は大きいものの平均値を見ると肯定文，否定文では実写表現の方が，命令文では変換表現の方が高い可能性がある．今回は文の種類ご

とに 2 つの単文を用意して実験を行ったが，より多くの単文を対象とし被験者を増やして実験を重ねることで明らかにできるものと考える．これらの実験結果から単文の読み取りについて主観的には実写表現の方が高評価であることがわかった．しかし，実際に読み取った成績については実写表現と変換表現の間に差は見られなかった．ただし，この差については表現方法と文の種類の間の有意な交互作用によって限定されるため，実験を重ねる必要があると考えられる．

長文の内容の読み取り実験では,内容に関する質問の正答率は実写表現の場合で約 9 割,変換表現の場合で約 7 割であり有意な差が見られた．単文読み取り実験の非手指動作の表現の見やすさおよび手指の見やすさの主観評価について実写表現の方が高評価であったことを考慮すると，長文の内容の理解に非手指動作の表現と手指の表現の見やすさが影響している可能性がある．本実験では単文と長文についての実写表現と変換表現で表現された匿名手話映像の読み取りやすさの差について明らかにできなかったが，非手指動作と手指の見やすさに着目した実験を行うことで明らかにできると考える．

4.5.4違和感に関する考察

匿名手話映像の違和感に関して，顔画像および単文の映像ともに違和感が少ないとは言えなかった．顔画像では表現方法間で有意な差が見られ変換表現の方が違和感が小さかったが，単文を表現した動画映像では有意な差が見られなかった．顔画像では時間をかけて処理後の顔画像を見て本人を特定しようとするため，実写表現で違和感を強く感じる人が多かったと考えられる．表 4.7 で示したように，違和感に関する意見は全く寄せられておらず読み取りの改善に関する意見や要望がほとんどであった．これらの結果の理由として，単文映像になると内容の把握が優先され表現方法による違和感を覚えにくかった，あるいは動きをともなうと違和感が低減される可能性があると推測する．今後はさらに違和感とその発生原因について検討を重ねる必要がある．

4.5.5全体の考察

本研究で提案した匿名手話映像の生成法では，実写表現，変換表現ともに匿名性が十分に確保されていることが明らかになった．また，実写表現による長文の内容の読み取りでは，内容をほぼ読み取れることがわかった．つまり，コミュニケーション時に実写表現による匿名手話映像で相手が提示された場合に，一定の匿名性が確保された状態で内容を読み取れることが示唆された．しかし，違和感についてはさらに改善する必要があると考えている．

変換表現では表 4.7 の①，②で被験者から指摘されているように，手指と顔の部位に影がなく色がそれぞれ同じ色で塗りつぶされているため，手指と顔の部位の動きがわかりにくいとのことであった（図 4.12）．変換表現の色と形の変形方法を改善してさらに違和感の少ない匿名手話映像を作ることが今後の課題である．

図 4.12 変換表現の問題点

第5章 結論

5.1 まとめ

聴覚障害者は声色を変えたり，体を隠蔽したりしながら匿名でコミュニケーションを行うことが困難である．そこで筆者は聴覚障害者が日常的なコミュニケーションで活用している手話で匿名コミュニケーションが可能なシステムを考える必要があると考えた．本研究では手話による匿名コミュニケーションを実現するために，Kinect を用いて CG モデルを動かし，顔の部位と手指の表現に実写画像を利用することによって，匿名手話映像を生成する方法を提案した．また，顔の部位に実写画像を利用すると匿名性を確保できなくなる可能性があったため，CG モデルに合わせて色と形の変換を行った上で CG モデルに合成する方法も提案した．

提案法を実装した匿名手話映像生成システムを構築したところ，処理速度に関して手話者を撮影してから匿名手話映像が表示されるまでに要する処理時間は，実写表現で約 36ms，変換表現で約 40ms であった．手話通信において 660ms 以上の遅延時間が生じると自由対話に支障が生じるが，本生成法の処理時間を考慮して 620ms 未満であれば対話コミュニケーションに支障はないと考えられる．

実写表現，変換表現で生成した匿名手話映像について匿名性の確保に関する評価実験を行った．その結果，実写表現と変換表現によって匿名化された顔画像についてはいずれも匿名性が十分に確保されていることが明らかになった．特に，変換表現は実写表現よりも回答する際の自信が低く，匿名性を向上することができる可能性があることがわかった．

匿名手話映像の読み取り実験については，単文では実写表現，変換表現合わせて全体の読み取り成績が約 7 割であった．非手指動作の表現の見やすさの主観評価，話者の気持ちの伝わりやすさともに実写表現の方が変換表現よりも評価が高かった．長文の内容に関する質問の正答率では有意な差が見られ実写表現の方が，正答率が高かった．単文読み取り実験の非手指動作の表現の見やすさおよび手指の見やすさの主観評価について実写表現の方が高評価であったことから，長文の内容の理解に非手指動作の表現と変換表現の見やすさが影響していると考えられる．

匿名手話映像の違和感について，顔画像では変換表現の方が違和感が小さかったが，単文の映像では表現方法間で有意差が見られなかった．これらの違和感の発生原因については明らかにできなかった．

以上の実験結果から，本研究で提案する匿名手話映像生成法の実写表現と変換表現で匿名性を確保可能であり，特に実写表現については表出された内容の把握も 9 割以上把握可

能であることがわかった．また，本方法で生成した匿名手話映像の違和感については改善の余地があり，より自然な映像生成方法について検討する必要がある．

5.2 今後の課題

匿名性の確保の課題として，手指に障害を持った人や肌の色，性別などを隠すことができるか明らかになっていないことや，単語表現，手指や腕の動きによって本人が特定される問題点が挙げられる．これらの課題に対して匿名性を確保できるかどうかについて検討する必要がある．

長文の読み取りでは，変換表現での成績が実写表現よりも低かった原因として手指と顔の部位に影がなく色がそれぞれ同じ色で塗りつぶされているため手指と顔の部位の動きがわかりにくいことが挙げられる．単文読み取り実験の非手指動作の表現の見やすさおよび手指の見やすさの主観評価について実写表現の方が高評価であったことを考慮すると，手指と顔の部位の着色を改善し動きをわかりやすくすることで読み取り成績が向上すると考えられる．単文の読み取り成績については，被験者によって成績のばらつきが大きいため，文章の内容の把握しやすさをそれぞれ同等にした上で比較実験を行い，表現方法，文の種類の間で傾向があるかどうかさらなる検討が必要である．

顔画像および単文の映像ともにCGモデルと実写画像の合成にともなう違和感については小さくすることができなかった．先述した長文の読み取りでの課題と同様に，顔の部位と手指の実写画像の着色を改善することにより違和感の少ない読み取りやすい匿名手話映像の生成についても検討する必要がある．

本方法で生成した匿名手話映像を実際の対話場面で利用することを想定して，遅延や切断などのネットワークに起因する特性，ならびに対話しやすさや伝達内容の正確性などについて評価を行い，実用化にあたっての課題と解決方法を明らかにすることが今後の課題である．

謝辞

本研究を行うにあたり，ご指導ご鞭撻をいただきました若月大輔准教授に心より感謝いたします．また，貴重な時間を割いて本論文をご精読頂き有用なコメントを頂きました修士論文主査の皆川洋喜教授，副査の井上正之准教授に深く感謝いたします．同研究室の近藤真暉氏，稲川直樹氏には，研究についての様々な助言や研究室での生活に関することなど，他にも様々な点でお世話になりました．本当にありがとうございました．

参考文献

[1] 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”,福村出版,東京,2012

[2] 神田和幸,“ドラえもん手話の実例と NMS の情報伝達”,可視化情報学会誌.suppl.,24(1),pp.277-278,2008

[3] 金子,加藤,清水,井上,“動作合成による手話文 CG アニメーション生成”,電子情報通信学会総合大会講演論文集 2010 年_基礎・境界,p.274,2010

[4] 加藤,宮崎,金子,井上,梅田,清水,比留間,長嶋,“気象情報を対象とした日本語-手話 CG 翻訳の主観評価実験”,言語処理学会第 19 回年次大会発表論文集,pp.130-133,2013

[5] McCullough S, Emmoey K,“Face Processing by Deaf ASL Signers:Evidence for Expertise in Distinguishing Local Features”, J Deaf Stud Deaf Educ, 2(4), pp.212-222, 1997

[6] 中島悠太,池野知顕,馬場口登,“顔画像に対するプライバシー保護処理の有効性の定量的評価”,電子情報通信学会技術研究報告会.ICSS,情報通信システムセキュリティ 112(128),pp.59-66,2012

[7] “上野家のホームページ-Game/KinectKinect-資料室”,http://yueno.homeip.net/xoops/modules/xpwiki/?Game%2FKinect (2015/02/23 参照)

[8] 杉浦司,中村薫,“Kinect for Windows SDK 実践プログラミング”,工学社出版,2013

[9] Microsoft,“Face Tracking SDK-MSDN-Microsoft”,http://msdn.microsoft.com/en-us/library/jj130970.aspx (2015/02/23 参照)

[10] “DX ライブラリ置き場 HOME”, http://homepage2.nifty.com/natupaji/DxLib/ (2015/02/23 参照)

[11] OpenCV 2 プログラミングブック制作チーム，“OpenCV 2 プログラミングブック”，マイナビ，2011

[12] 三宅太一，“距離画像を用いた動きのある指文字認識に関する研究”,筑波技術大学 修士（工学）学位論文

[13] 射手矢味先，浅見高明，小俣幸嗣，石島繁，梅垣浩二，“柔道選手における手部の形態的機能的左右差について”，武道学研究 22-2，pp.63-64，1989

[14] 長嶋祐二，住田英之，中園薫，“手話対話の映像遅延に及ぼす影響の基礎的な検討”，電子情報通信学会技術研究報告.WIT，福祉情報工学 105(67)，pp.25-30，2005

本研究に関する成果・発表等

[A1] 松岡通浩，若月大輔，河野純大，“匿名コミュニケーションのための手話映像生成に関する基礎的検討”，平成 25 年度電子情報通信学会信越支部大会講演論文集，p.28，2013

[A2] 松岡通浩，若月大輔，“匿名コミュニケーションのための手話映像表現”，第 9 回日本聴覚障害学生高等教育支援シンポジウム・ランチセッション「聴覚障害学生支援に関する機器展示」，p.77，2013

[A3] 松岡通浩，若月大輔，河野純大，“匿名コミュニケーションのための手話映像生成に関する検討”，2014 年電子情報通信学会情報・システムソサイエティ特別企画 ポスターセッション予稿集，ISS-P-138，2014

[A4] 松岡通浩，若月大輔，“匿名コミュニケーションのための手話映像表現”，第 10 回日本聴覚障害学生高等教育支援シンポジウム・ランチセッション「聴覚障害学生支援に関する機器展示」，p.20，2014

[A5] 松岡通浩，若月大輔，河野純大，“匿名性のある読み取りやすい手話映像の生成方法”，平成 26 年度電子情報通信学会信越支部大会講演論文集，p.38，2014

[A6] 松岡通浩，若月大輔，河野純大，“読みやすさを考慮した匿名手話映像の生成法”，電子情報通信学会 HCG シンポジウム 2014，pp.507-512，2014

付録

A1　著名な人物の顔画像に対するフェイストラッキング

4.3.3 節で述べた匿名性の確保に関する実験では，著名のような人物の顔画像に対して実写表現または変換表現の処理を行っている．そのためには有名人に対してフェイストラッキングを行う必要がある．

Kinect のフェイストラッキングはカラーバッファと深度バッファを引数として渡して，2 種類の動作を行う．一つ目はスケルトントラッキングで得られた首と頭の位置をもとに顔を追跡する方法であり，二つ目はカラーフレーム上で一から顔を検出する方法である．

実際に有名人に Kinect で撮影してもらうことは難しいので，本研究ではインターネットで収集した著名のような人物の顔画像そのものをカラーバッファとして読み込み， Kinect から撮影した深度画像を深度バッファとして読み込んだ．著名のような人物の顔画像中に顔があると思われる位置に対応する深度画像中の位置に筆者の顔を合わせて移動させると，有名人の顔の特徴点を取得することができる．

そのことから，Kinect のフェイストラッキングでは深度情報を使わずに顔の特徴点を取得していると考えられる．フェイストラッキングの詳細な処理については，Face Tracking SDK の MSDN を参考にされたい．

A2　実験に用いたアンケート用紙

4.3 節で述べた匿名性の確保に関する実験で用いたアンケートは 2 種類である．一つ目は対象の人物を同定するためのアンケート，二つ目は対象の人物に関する親密度，特徴度を調査するためのアンケートであり，アンケートの例はそれぞれ図 4.3，4.4 で示した通りである．

4.4 節で述べた手話の読み取りに関するに実験で用いたアンケート用紙のサンプルを提示する．

手話映像の文章の読み取りに関する調査

あなたの性別をお選びください。(○はひとつ)

1) 男性　　2) 女性

あなたの年齢を教えてください。

＿＿＿＿＿＿歳

あなたの手話歴を教えてください。

＿＿＿＿＿＿年

今回の実験で提示された映像の内容についてお聞きします。

Part1

問 1-1

映像 1 の内容をご記入ください。

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

問 1-2

映像 1 について，以下の項目に対して 5 段階評価でお答えください。(○は項目ごとにひとつずつ) また，気持ちについてもお尋ねします.

	5:そう思う 4:ややそう思う 3:どちらともいえない 2:ややそう思わない 1:そう思わない				
手話はわかりやすい	5	4	3	2	1
手・指の動きは見やすい	5	4	3	2	1
腕の動きは見やすい	5	4	3	2	1
頭の動きは見やすい	5	4	3	2	1
目の動き（瞬き，視線等）は見やすい	5	4	3	2	1
口の形は見やすい	5	4	3	2	1
眉の動きは見やすい	5	4	3	2	1
話者の気持ちが伝わってきた	5	4	3	2	1
違和感がある	5	4	3	2	1

問 1-3

映像 1 ではどんな気持ちが伝わってきましたか.

＿＿＿＿＿＿＿＿

問 2-1

映像 2 の内容をご記入ください。

_____

問 2-2

映像 2 について，以下の項目に対して 5 段階評価でお答えください。(○は項目ごとにひとつずつ) また，気持ちについてもお尋ねします.

	5:そう思う 4:ややそう思う 3:どちらともいえない 2:ややそう思わない 1:そう思わない				
手話はわかりやすい	5	4	3	2	1
手・指の動きは見やすい	5	4	3	2	1
腕の動きは見やすい	5	4	3	2	1
頭の動きは見やすい	5	4	3	2	1
目の動き（瞬き，視線等）は見やすい	5	4	3	2	1
口の形は見やすい	5	4	3	2	1
眉の動きは見やすい	5	4	3	2	1
話者の気持ちが伝わってきた	5	4	3	2	1
違和感がある	5	4	3	2	1

問 2-3

映像 2 ではどんな気持ちが伝わってきましたか.

_____

問 3-1

映像 3 の内容をご記入ください。

_____

問 3-2

映像 3 について，以下の項目に対して 5 段階評価でお答えください。(○は項目ごとにひとつずつ) また，気持ちについてもお尋ねします.

	5:そう思う 4:ややそう思う 3:どちらともいえない 2:ややそう思わない 1:そう思わない				
手話はわかりやすい	5	4	3	2	1
手・指の動きは見やすい	5	4	3	2	1
腕の動きは見やすい	5	4	3	2	1
頭の動きは見やすい	5	4	3	2	1
目の動き（瞬き，視線等）は見やすい	5	4	3	2	1
口の形は見やすい	5	4	3	2	1
眉の動きは見やすい	5	4	3	2	1
話者の気持ちが伝わってきた	5	4	3	2	1
違和感がある	5	4	3	2	1

問 3-3

映像 2 ではどんな気持ちが伝わってきましたか.

問 4-1

映像 4 の内容をご記入ください。

問 4-2

映像 4 について，以下の項目に対して 5 段階評価でお答えください。(○は項目ごとにひとつずつ) また，気持ちについてもお尋ねします.

	5:そう思う 4:ややそう思う 3:どちらともいえない 2:ややそう思わない 1:そう思わない				
手話はわかりやすい	5	4	3	2	1
手・指の動きは見やすい	5	4	3	2	1
腕の動きは見やすい	5	4	3	2	1
頭の動きは見やすい	5	4	3	2	1
目の動き（瞬き，視線等）は見やすい	5	4	3	2	1
口の形は見やすい	5	4	3	2	1
眉の動きは見やすい	5	4	3	2	1
話者の気持ちが伝わってきた	5	4	3	2	1
違和感がある	5	4	3	2	1

問 4-3

映像 4 ではどんな気持ちが伝わってきましたか.

問 5-1

映像 5 の内容をご記入ください。

問 5-2

映像 5 について，以下の項目に対して 5 段階評価でお答えください。(○は項目ごとにひとつずつ) また，気持ちについてもお尋ねします.

	5:そう思う 4:ややそう思う 3:どちらともいえない 2:ややそう思わない 1:そう思わない				
手話はわかりやすい	5	4	3	2	1
手・指の動きは見やすい	5	4	3	2	1
腕の動きは見やすい	5	4	3	2	1
頭の動きは見やすい	5	4	3	2	1
目の動き（瞬き，視線等）は見やすい	5	4	3	2	1
口の形は見やすい	5	4	3	2	1
眉の動きは見やすい	5	4	3	2	1
話者の気持ちが伝わってきた	5	4	3	2	1
違和感がある	5	4	3	2	1

問 5-3

映像 5 ではどんな気持ちが伝わってきましたか.

問 6-1

映像 6 の内容をご記入ください。

問 6-2

映像 6 について，以下の項目に対して 5 段階評価でお答えください。(○は項目ごとにひとつずつ) また，気持ちについてもお尋ねします.

	5:そう思う 4:ややそう思う 3:どちらともいえない 2:ややそう思わない 1:そう思わない				
手話はわかりやすい	5	4	3	2	1
手・指の動きは見やすい	5	4	3	2	1
腕の動きは見やすい	5	4	3	2	1
頭の動きは見やすい	5	4	3	2	1
目の動き（瞬き，視線等）は見やすい	5	4	3	2	1
口の形は見やすい	5	4	3	2	1
眉の動きは見やすい	5	4	3	2	1
話者の気持ちが伝わってきた	5	4	3	2	1
違和感がある	5	4	3	2	1

問 6-3

映像 6 ではどんな気持ちが伝わってきましたか.

_____

問 7-1

映像 7 の内容をご記入ください。

_____

問 7-2

映像 7 について，以下の項目に対して 5 段階評価でお答えください。(○は項目ごとにひとつずつ) また，気持ちについてもお尋ねします.

	5:そう思う 4:ややそう思う 3:どちらともいえない 2:ややそう思わない 1:そう思わない				
手話はわかりやすい	5	4	3	2	1
手・指の動きは見やすい	5	4	3	2	1
腕の動きは見やすい	5	4	3	2	1
頭の動きは見やすい	5	4	3	2	1

目の動き（瞬き，視線等）は見やすい	5	4	3	2	1
口の形は見やすい	5	4	3	2	1
眉の動きは見やすい	5	4	3	2	1
話者の気持ちが伝わってきた	5	4	3	2	1
違和感がある	5	4	3	2	1

問 7-3

映像 7 ではどんな気持ちが伝わってきましたか.

問 8-1

映像 8 の内容をご記入ください。

問 8-2

映像 8 について，以下の項目に対して 5 段階評価でお答えください。(○は項目ごとにひとつずつ) また，気持ちについてもお尋ねします.

	5:そう思う 4:ややそう思う 3:どちらともいえない 2:ややそう思わない 1:そう思わない				
手話はわかりやすい	5	4	3	2	1
手・指の動きは見やすい	5	4	3	2	1
腕の動きは見やすい	5	4	3	2	1
頭の動きは見やすい	5	4	3	2	1
目の動き（瞬き，視線等）は見やすい	5	4	3	2	1
口の形は見やすい	5	4	3	2	1
眉の動きは見やすい	5	4	3	2	1
話者の気持ちが伝わってきた	5	4	3	2	1
違和感がある	5	4	3	2	1

問 8-3

映像 8 ではどんな気持ちが伝わってきましたか.

Part2

問 1-1（長文 1 について）
家で何をしていたのでしょうか。

_____

問 1-2（長文 1 について）
どこで遊んでいましたか。

_____

問 2-1（長文 2 について）
今日はどんな天気でしたか。

_____

問 2-2（長文 2 について）
私はどこへ行くのでしょうか。

_____

その他に改善などの意見や要望があれば、ご記入ください。