

自律移動支援プロジェクトでの 手話映像合成技術の応用

鈴木 雄介 宮本 一郎
竹内 晃一

厚生労働省の統計によれば日本における身体障害者手帳保有数は30万人余りであり、そのうち日本語と言語体系の異なる「日本手話」を母語とする人は約6万人いるとされている。

日常生活において、音声アナウンスなど音声のみで提供されている情報は少なからずあり、聴覚障害者にとっては情報にアクセスできないという問題となる。こうした情報を手話として提供することは、聴覚障害者の利便となる可能性が高い。

手話映像を表現するためのシステムは多数研究されており、中には商品化されているものもある^{1) 2) 3)}。これらのシステムでは、手話の表示にCG（コンピュータグラフィクス、以下CG）によって合成された映像を使っているものが多い。CGには任意の文章を表す手話を合成できる可能性があるという利点があるが、手話にとって重要な、非手指信号（Non Manual Signals 以下NMS）と呼ばれる表情や体の動き、口の形などの表現が難しいなどの課題も残されている^{4) 5) 6)}。

筆者らは、実写の映像を利用した手話映像表示システムの開発を行っている。実写映像はCGに比して上記NMSなどの表現力が高く、実際の手話使用者に好まれるという利点がある⁷⁾。しかし、CGに比べ自由な文章の合成が困難であるという問題も残されている。本稿では映像合成の技法を利用して上記の問題の解決をめざすシステムについて説明する。

手話映像合成システムについて

手話映像合成システムは実際に手話の動作を行っている人物を撮影した映像を用いている。これは手の動きだけでなく、NMSが表現上重要な役割を持つという手話の特性を重視したためである。NMSの分析、研究はまだまだ途上にあり、CGで工学的に合成することが可能なレベルまで理解が進んでいない。またCGでは手話動作データをシステムに追加する際、人物の動作を取り込むためにモーションキャプチャと呼ばれる技術を用いることが多いが、

表情などの動きは正確なキャプチャやデータの扱いが困難であり、またデータの校正に多くの手作業を必要とする。

実際に手話を行っている人物を撮影した映像はCGに比してNMSの表現力が高いため手話が読み取りやすく、また日常見慣れているため、実際の手話使用者に好まれるという利点がある。そのため、実際に手話を行っている人物を撮影した映像を使用することが、手話で情報提示を行う場合には理想的であろう。しかし、撮影された映像は柔軟性を欠き、わずかな内容変更の場合にも新たな撮影作業が必要になるなどCGに比べ自由な文章の合成が困難であるという問題もある。

筆者らはこれらの問題を考慮し、二つの方式の利点をうまく結びつけた手話映像合成を実現することを目標としている。すなわち、CGの柔軟性と実写映像の表現力の高さを併せ持つ手話映像合成技術の開発である。上記映像の作成方法と利点を表1にまとめて示した。○がついている項目がその方法の優れた点をあらわす。

表1 手話映像システムの比較

	方法		
	撮影した 動画	CG	本 システム
状況の変化に対する 動的な対応		○	○
拡張性		○	○
NMSの表現力	○		○
ろう者にとっての 親しみやすさ	○		○

本システムは手話を行っている人物を撮影した実写の映像を、単語を単位として分節化し、それらを必要に応じて選択、連結して一連の手話文章をあらわす映像を合成する。また本システムは基本的には撮影と映像の切り出しのみでシステムへのデータの追加が容易に行えるという点で、データの拡張性が高い。加えて、本システムは連結する映像間の動作の空白枠を映像合成の手法で補

完するという特徴を持つ。

試作中であるシステムの構成の概念図を図1に示す。

このシステムは入力された文章を形態素解析して、分割された各日本語の単語に相当する手話映像ファイルをデータベースから選択し、連結して一連の文章として出力するという構成になっている。以下、図1に基づき、各要素とその実際の動作について説明する。

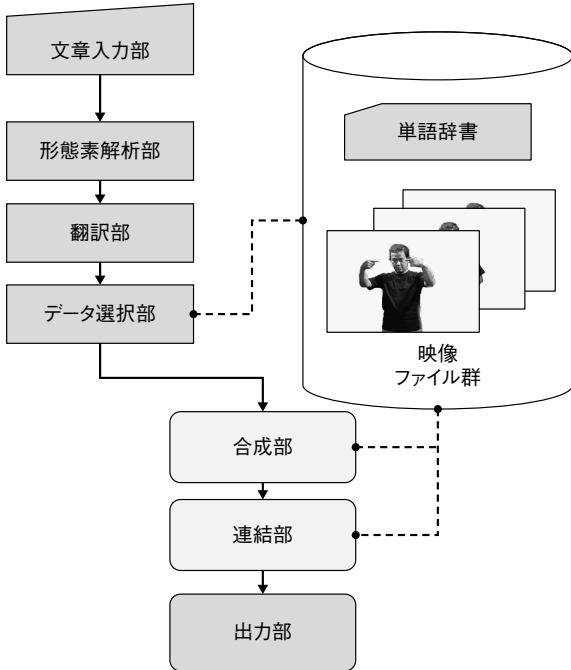


図1 システムの構成

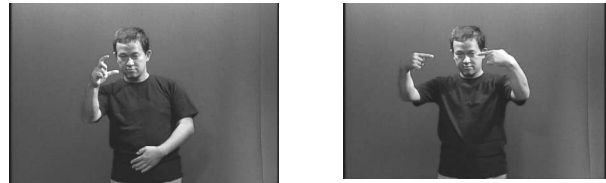
システムはデータ、文章入力部、形態素解析部、翻訳部、データ選択部、合成部、連結部、出力部から構成されている。使用するデータは主に、単語辞書、映像ファイルなどからなる。単語辞書とは日本語と、手話映像ファイルとの対応関係を示したものである。映像ファイル群とは、文章を表現するのに必要な手話の要素を、ほぼ単語を単位として分割したものである。形態素解析部は入力された日本語文章を形態素解析する。

翻訳部は単語の語順を解析の後、手話の文法に合わせて前後関係を調整するなどの翻訳処理の後、手話単語ラベルを割り当てる。手話単語ラベルとは主に、手話単語を識別するために割り当てられた日本語のラベルである。

手話単語は、日本語と手話は異なる言語であるため、日本語では一つの単語で表現される内容が複数種類の手話表現となることがある。

たとえば、気象概況に用いる単語を用いて説明すると、図2に示すように「台風が北上する」の「北上する」と

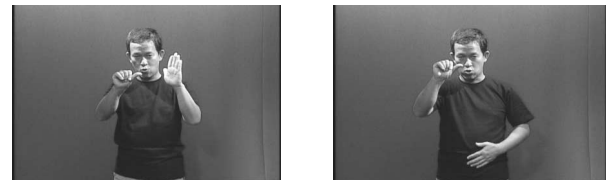
「前線が北上する」の「北上する」は形態が異なる。また図3に示すように、「数字」表現は係る単語が「台風」であるか「気温」であるかで左手の位置が大きく異なる。係りを受けて変化するそれぞれの場合ごとに一つの映像ファイルとして保持し、翻訳の後にデータ選択部によって、適宜ふさわしい映像ファイルが選択されるようにする。



「北上する」(台風)

「北上する」(前線)

図2 係り受けによる形態の違い



5(気温)

5(台風)

図3 係り受けによる数字の形態の違い

問題点

映像ファイルはほぼ単語ごとに分節化されたものであるため、これを単純に連結して表示すると各単語の終了時から次の単語の開始時までの人物の動作が最終的な連結映像に現れず、人物動作の空白域（ギャップ）が生じ、最終的に表示される映像ファイルに違和感が生ずるという問題が発生する。

この問題を解決するため映像合成技法によって、前後の映像ファイルから動作の空白部分を補完するための合成映像を作成する。最終的に合成された映像ファイルも含めた一連の手話映像を出力する。

映像合成

図4（次ページ）に従って、以下中間映像合成のフローについて説明する。映像ファイル群から2つのファイルを選び出し、先に表示されるファイルを先行ファイル、次のファイルを後続ファイルとする。

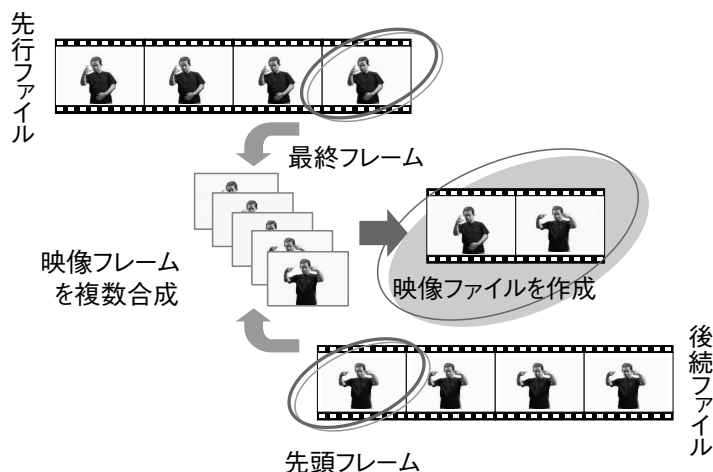


図4 映像合成のフロー

- ① 2つの映像ファイルを準備する。
- ② 先行ファイルから動作の終了時点のフレーム（最終フレーム）を取り出し、後続ファイルから動作の開始時点のフレーム（先頭フレーム）を取り出す。
- ③ 取り出した2枚のフレームに対して画像合成の技法であるモーフィング処理を行うことによって、最終フレームから先頭フレームへと変化する中間の画像フレームを複数枚生成する。この処理の詳細は次節で説明する。
- ④ 生成した画像フレームを連結して映像ファイルを作成する。
- ⑤ 作成した映像ファイルを先行ファイルと後続ファイルの中間部分に挿入する。

一連の処理によって、手話を行っている人物の動作が滑らかになっているように見せる効果を得ることができる。



図5 画像合成の図

映像合成の詳細

映像合成の方法の詳細を図5で説明する。Warping³⁾を基に、基礎実験を反復して改良した方法である。

まず、動画から切り出した静止画フレーム上、人物の掌からひじまでの腕領域を囲む輪郭線を複数点の集合として取得しておく。次に輪郭線を構成する複数点から上記の囲んだ領域を覆う三角形メッシュを作成する。三角形メッシュは先行ファイルの最終フレーム、後続ファイルの先頭フレームで同様に作成される。

先行する最終フレームのメッシュ領域と後続する先頭フレームメッシュ領域の位置の差から移動速度を決定し、最終フレームのメッシュを先頭フレームのメッシュへと決定した速度で移動させつつメッシュ内対応領域の画素の平均値を画面に出力する。これによってメッシュ、つま

り腕領域が形を変えながら動作している映像が得られる。

この映像に予め作成しておいた人物の腕領域を隠した体領域の映像テンプレートと、各フレームから腕領域内の映像を塗りつぶして背景が前面に現れないような映像とを重ね合わせたものを出力すると、腕が自然に動作しているような映像を得ることができる。

実験的応用事例

前節で説明した手話映像合成エンジンを用いた情報提供システムを開発し、神戸空港で実験を行った。実験は国土交通省が推進する自律移動支援プロジェクトの一環として行われた。自律移動支援プロジェクトは高齢者、障害者または怪我などによって一時的にであっても移動や、移動に関する情報の取得に障害がある人を情報技術に

よって支援し、対象者の自律的な移動を助けるという趣旨を持つプロジェクトである。

神戸空港で行われたプロジェクトの実証実験中に、聴覚障害者を対象として手話映像合成システムの有用性の評価を行った。被験者の持つ携帯端末上に作成したシステムで合成した手話映像を表示した。図6に携帯端末上に表示された手話映像のイメージを示す。評価実験後のアンケートの結果、被験者の多くが手話による情報提示を支持し、手話による情報が必要であると答えた。実験の詳細報告は本誌掲載の「ユビキタスサービスの実現に向けて～神戸空港ユビキタス実証実験への参画～」[pp.4-9]に譲る。



図6 携帯端末上に配置された手話映像のイメージ

終わりに

開発中の手話映像合成システムについて説明した。手話については未知な部分も多く、大学や他企業研究機関、また実際にシステムのユーザーとなる手話使用者との協力関係をとりながら研究をすすめる必要がある。 ◆◆

参考文献

- 1) 黒川隆夫：“手話と日本語の相互翻訳の試み”，第5回関西情報関連学会連合大会論文集，pp.25-34，2000年
- 2) 黒田知宏他：“手話伝送システムS-TEL”，電子情報通信学会技術研究報告，ET96-85，pp.65-71，1996年
- 3) 佐川浩彦：“手話アニメーションソフトMimehandとその応用”，医療とコンピュータ，Vol.13 No.8，2002年
- 4) 木村晴美，市田泰博：“はじめての手話”，日本文芸社，1995年
- 5) 米原裕貴，長嶋祐二：“手話の習熟度による注視点の変化に関する実験的検討”，ヒューマンインターフェースシンポジウム2002論文集，pp.233-236，2002年
- 6) 児玉哲彦，安村通晃：“表情の表現を含む手話アニメーションの試作”，情報処理学会研究報告，2003-HI-103，pp.23-29，2003年

7) 市川貴士他：“携帯端末画面での手話映像の見易さに関する検討”，ヒューマンインターフェースシンポジウム2003論文集，pp.309-312，2003年

8) Thaddeus Beier, Shawn Neely: “Feature-based image metamorphosis”, Proceedings of the 19th annual conference on Computer graphics and interactive techniques, pp35-42, 1992

● 筆者紹介

鈴木雄介：Yushuke Suzuki. 研究開発本部 HIラボラトリ インタラクティブコミュニケーションチーム

宮本一郎：Ichiro Miyamoto. 研究開発本部 HIラボラトリ インタラクティブコミュニケーションチーム

竹内晃一：Koichi Takeuchi. 研究開発本部 HIラボラトリ インタラクティブコミュニケーションチーム