学会誌 20120 年 7 月号解説記事より再掲

ＣＧによる手話アニメーションの自動生成システム

比留間伸行† (正会員)　清水俊宏†　梅田修一†　加藤直人†　宮﨑太郎†　井上誠喜†　金子浩之†　長嶋祐二††

†ＮＨＫ放送技術研究所，††工学院大学

Automatic Generation System of CG Sign Language Animation

Nobuyuki HIRUMA† (*Member*), Toshihiro SHIMIZU†, Shuichi UMEDA†, Naoto KATO†, Taro MIYAZAKI†, Seiki INOUE†, Hiroyuki KANEKO†, Yuji NAGASHIMA††

†NHK Science and Technology Research Laboratories, ††Kogakuin University

1.　はじめに

ＮＨＫは公共放送として，視覚や聴覚に障害がある方にも情報をお届けすることを重視している．視覚障害者に向けては，デジタル放送の電子番組表（ＥＰＧ）やデータ放送の文字データを音声で読み上げたり，点字ディスプレイなどで触覚により利用したりできる受信端末を開発している [1]．また，聴覚障害者のための字幕（クローズドキャプション）のデータを制作するための音声認識技術の開発も進めている [2]．

一方，先天的なろう者を中心に使われている「日本手話」は，音声言語の日本語とは語彙も文法体系も異なる別の言語であり，これを第一言語とする人々からは，字幕に加えて手話によるサービスの充実の要望をいただいている．これにお応えすることを目指して，我々は日本語の入力テキストから，相当する手話のアニメーションをコンピュータグラフィクス（ＣＧ）により自動生成する技術の研究に取り組んでいる．

ただし現状では，任意の話題の日本語から手話への変換は，極めて技術的なハードルが高く現実的ではない．そこで我々は，気象のニュースを対象とすることとした．これは，気象情報は緊急性を有する事があるので人間の手話通訳者が放送現場に不在となる深夜，早朝の緊急のニュースに応用できれば有効であると考えられ，また，使用される語彙や文体がある程度限られることから技術的実現性があると考えられるためである．本稿では，この研究開発の現況を概説する．

2.　手話のモーションキャプチャ

本研究には，２つの側面がある．日本語と手話という異なる言語同士を変換する自動翻訳技術としての面と，人体の自然な動作のアニメーション映像を生成するコンピュータグラフィクス技術の面である．これらの双方の基礎となる重要なデータが，手話を演じる人体の動作の電子的な記録である．

言語の研究のためのデータというと，その言語の筆記録を思い浮かべるが，手話は文字を持たない言語である．文字の無い言語は世界に例が多いが，その研究に際しては，音声言語であれば，通常，発話を発音記号などで筆記する事ができる．これとは異なり手話は視覚的な言語であり，同一時点に複数の言語的な要素が表出されるので，表記は簡単ではない．手話研究のための表記法として，sIGNDEX，HamNoSysなどが有名であるが，その記述の実践には高度な習熟を要するため我々の目的には容易とは言えず，現状では映像として記録するのが現実的であると判断した．映像記録として実写映像をそのまま収録することは，高品質な記録が容易に得られるメリットがあるが，収録後応用に用いる際の被写体のポーズの加工等は困難である．手話においても音声言語と同様に，文内での接続や文法上の役割等によって語の変形が起きる事から，収録後の加工・変形が適用しやすいモーションデータ記録に基づくＣＧアニメーションにより手話語彙を記録した．すなわち，手話単語ごとにモーションキャプチャを行い３Ｄキャラクタモデルで動作するＣＧアニメーションを作成した．

2.1.　収録単語の選定 [3]

モーションデータを収録する手話語彙の選定にあたっては，（財）全日本ろうあ連盟発行の「わたしたちの手話」[4]を中心に，いくつかの冊子型手話辞典を参考にした．その中には最小の意味単位である語(形態素)から，その組み合わせによる語（複合語）まで幅広く含んでいる．理想的には，形態素のみでＣＧ アニメーションを作成し，複合語は形態素単位のＣＧアニメーションを合成することも考えられる．しかしながら，ＣＧアニメーションの合成が必ずしも滑らかにできない場合のあることが予想されるので，よく使われるような複合語は一つの動作としてＣＧアニメーションを作成した．また，

機械翻訳などの言語処理の立場からも，複合語のほうが語義を一意に特定しやすく翻訳精度の向上が期待できるので，手話語彙には複合語も含めた．これまでに収集した手話語彙数は，指文字や数量表現も含め約7,000語である．「わたしたちの手話」の手話語彙数は約6,000語であるので，かなりの手話語彙をカバーしていると考えられる．

2.2. モーションキャプチャの実際

モーションキャプチャの手法は各種提案されているが，我々は高い精度が期待できる光学式モーションキャプチャを用いた．手話においては，指で作る輪が閉じているか否か，隣接する指同士が接触しているか否か，などのディテールが読み取りの上で大切になるからである．一方で光学式モーションキャプチャの弱点であるマーカの遮蔽によるデータの脱落については，編集作業で注意深く補正することとした．

光学式モーションキャプチャでは，再帰性反射材でコーティングされた複数個のマーカを被写体に装着し，レンズ周辺に赤外線LEDを搭載した複数台の赤外線カメラを用いて反射光を収録する．収録したカメラ映像を元に各マーカの三次元位置を計測することによって，被写体の動作を算出する．

図１にモーションキャプチャの撮影風景を示す．手話の演者（手話者）は，CODA（ Child of Deaf Adults：当人は健聴者だが，親（養育者）がろう者で幼少時から手話で育てられた方．いわば，日本語と手話のバイリンガリストである）で手話通訳士の資格を持つ女性１名である．この方に，今回選定した手話語彙すべてを演じてもらった．マーカは手話で重要な手や指を中心に頭から足まで装着した．また，手話では顔の表情など非手指の動作も重要であることが指摘されており，手話者の顔面の特徴的な部位にもマーカを装着してモーションデータを記録した．

図１ 手話語彙のモーションキャプチャ

Fig.1 Motion Capture of Sign Language Words

2.3. ＣＧキャラクタのモデル

モーションキャプチャにより得られたデータは骨格構造からなるＣＧキャラクタ（図２）の動作データに変換され，BVH (BioVision Hierarchy)ファイル形式で保存されている．BVHファイル形式はBiovision 社[5]が開発したモーションキャプチャのファイルフォーマットで，アスキー形式で記述されているので，ファイル内容を計算機で操作することが容易である．

骨格構造を定義するにあたっては，手指は各指の関節に加えて親指の付け根や掌の細やかな動きを可能とし，身体は鎖骨に関節を増やすことで肩周辺の動作を可能とした[6]．

図２ 手話ＣＧキャラクタのための骨格モデル

Fig.2 Bone Model of Sign Language CG Character

さらに，分かりやすい手話ＣＧアニメーションを生成する上で課題となる，非手指動作が表出できるモデルの開発に取り組んでいる [7]．表情をはじめとする非手指動作には，口型の他に眉の上下・寄せ，頬や顎の動き，うなずきなどがある．

顔の表情をＣＧにより表現する技術としては，ＣＧモデルの顔において皮膚下の表情筋と，これに連動する顔の皮膚表面を物理モデルとして定義する方法があるが，表情筋モデルの制作は複雑であり，導入が容易ではない．本研究において，我々は，光学式のモーションキャプチャによって顔の皮膚形状の動きをデジタル的に記録したデータによりＣＧモデルの皮膚形状を制御する手法を選択することとした．表情筋モデルと比較して制御点（関節）が多くなるものの， 物理モデル

によるシミュレーションが不要でありＣＧ生成時の計算負荷が少なく，ＣＧモデルの制作が容易である（図３）．

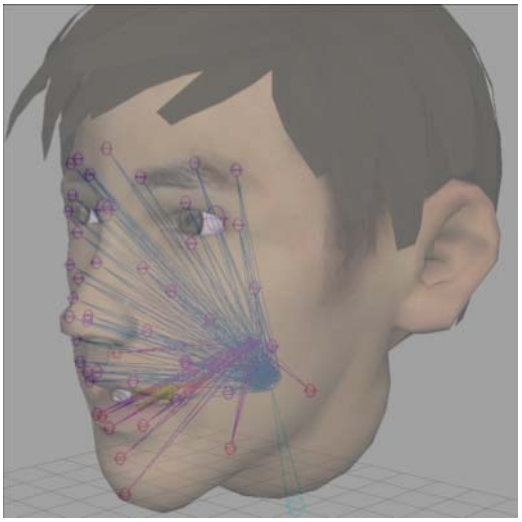

図３　表情のＣＧモデル

Fig.3　CG Model of Facial Expression

3.　手話コーパスの構築と日本語-手話翻訳技術

言語翻訳の手法は，規則翻訳，用例翻訳，統計翻訳に大別される．日本語から手話への翻訳を開発するにあたっては，以下のように考えて，用例翻訳と統計翻訳を用いることとした[8]．なお通例，言語翻訳とは文字列から文字列への変換をいうが，手話は視覚言語であるので，最終的には映像への変換が必要となる．しかし，日本語単文字列から手話映像へ直接変換することは難しいので，まずは手話単語列（手話動作を単語列に書き起こしたもの）に変換し，変換された手話単語列から手話映像ＣＧに変換するという手順を採用した．

規則翻訳では人手によって構築された翻訳知識が必要となる．翻訳知識の構築には，翻訳対象の言語的知見が必要となるが，日本手話に関しては機械翻訳を開発するまでにはその解明が進んでおらず，規則翻訳で早期に翻訳システムを開発することは現実的ではない．一方，統計翻訳では対象言語の知見が必要ないという利点があるが，統計翻訳には，実際に言語が用いられた実例を集積した「対訳コーパス」と呼ばれる大規模なデータベースが必要である．このため我々は，ＮＨＫの手話ニュースを対象にコーパスの構築を進めている[9]．ＮＨＫが毎日放送する手話ニュース番組のアナウンス原稿と，それに対応する手話映像を手話単語列（頷き，指差しなどを含む）に書き起こしたデータとを対応づけて記録する作業により，現在までに，約３万対のコーパスが構築されている．しかしながら，現在報告されている他の言語の統計翻訳の研究例では100万文～1,000万文規模のコーパスが使われており，手話コーパスとしては比較的大きな我々の対訳コーパスの３万文でも，これに比べ遥かに小さい．今後も対訳コーパスを増やしていく予定であるので，将来的には統計翻訳の適用も考えられるが，現時点では難しい．その点，同じコーパスベースの翻訳手法である用例翻訳は対訳文を直接利用しているため，統計翻訳ほど大きなサイズでなくとも，ある程度の翻訳精度を得ることが可能である．さらに，我々は当面の翻訳対象を気象情報に絞っているので，そこに出現する言語現象はある程度限られたものとなる．特に，節や句単位でみると気象情報は定型的な表現が多い．もちろん用例翻訳では，統計翻訳と異なり，翻訳対象に対する言語的知見が必要となるが，規則翻訳ほど詳細なものでなくともよい．

以上の考察により，我々は用例翻訳を主とし，統計翻訳を併用する方法をとることにした．本システムでは，まず，節（句）単位で完全一致による用例翻訳を行う．その上で，用例との完全一致が得られない節（句）が現れた場合には，統計翻訳で翻訳する手法を開発した．

4.　TVML

一方，コンピュータグラフィクス技術としての本研究の特徴は，ＮＨＫ放送技術研究所が開発した番組制作記述言語TVML (TV program Making Language)[11]に基づいている点である．TVMLは，言語体系にカメラや照明などテレビ番組制作の概念やノウハウを含んだスクリプト言語であり，ＣＧキャラクタの複数表示やマルチＣＧカメラによる多彩な演出を付加した映像コンテンツが制作できる．パソコン上でTVMLによる番組台本(TVML スクリプト)を記述することで，ＣＧキャラクタにセリフや身体動作などを与えることが可能である．記述されたTVMLスクリプトはTVMLプレーヤと呼ぶソフトウェアによって順次解釈され，リアルタイムＣＧによる映像コンテンツが生成される．

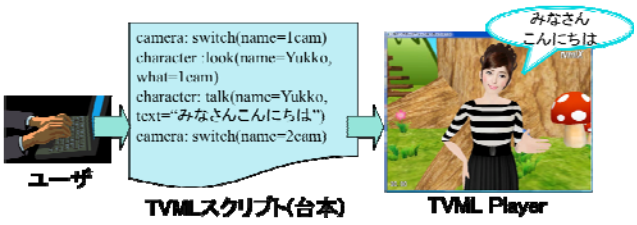

図４　TMVL のしくみ

Fig.4　Outline of TVML

TVML を用いて手話の手や指の動き(手指動作)を表現するＣＧ映像を生成するため，前述の多関節骨格構造をもつＣＧモデルに対応したTVML プレーヤを開発した[6]．

TVMLを用いる事により，手話キャラクタの入れ替え，背景の変更，視点やカメラパラメータ（ズームなど）の変更が容易に行え，番組の企画や特性に即したＣＧアニメーションを極めて柔軟かつ効率的に生成できる．

5. 手話文のアニメーション生成技術

前述のモーションキャプチャによる手話単語のモーションデータは，「きをつけ」の姿勢から１つの単語を演じ，その後「きをつけ」の姿勢に戻るまでを１シーケンスとして記録されている．これらを元に手話の文章のアニメーションを生成するためには，複数の単語の動作を接続する動作（「わたり」という）を生成する必要がある．そこで，ある手話単語の本体部分（以下，わかりやすく「実」の部分，と記す）が終了して「きをつけ」に戻るまでの動作の途中から，次の単語の「きをつけ」から「実」の部分に至る動作の途中に接続する動作を補間して生成する技術を開発した [11].

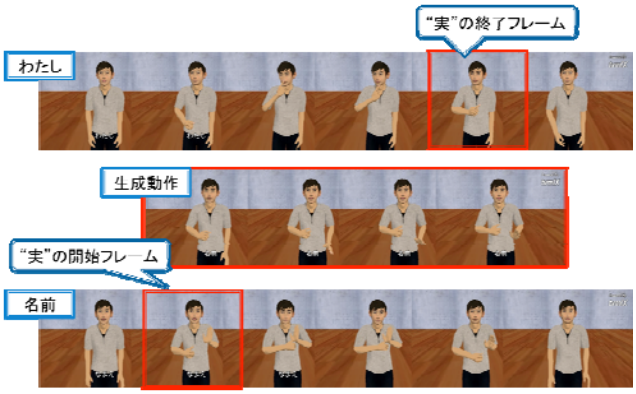

図５　手話単語の動作の接続

Fig.5　Connecting Sign Language Words

図５は，「私」という語と「名前」という語のモーションデータを接続して「私の名前」という句の動作を生成する模様を模式的に示したものである．現在のところ，前の語の「実」の終了フレームと次の語の「実」の開始フレームとの間の動作を生成するアルゴリズムは線形補間を用いている．今後，より良い補間法の検討などを進めていきたい．

6. 試作自動翻訳システム

以上の技術を集成して，入力日本語テキストから自動的にＣＧによる手話のアニメーションを生成するシステムを試作した（図６）．本システムは，話題のドメインは気象情報に限定されるが，その範囲では日本語の入力文に対し手話のアニメーションを生成できる．現在のところ，生成したアニメーションの手話としての品質に関する系統的な評価は実施しておらず今後の課題であるが，本システムをＮＨＫ放送技術研究所の一般公開などのイベントで展示し，ご覧いただいたろう者の方々からは以下のような感想を得ている．

- 理解可能な手話が表現できているところもあり，今後が期待される．
- しかしながら全体としては．単語の選択，動作の自然さなどの点で改善の余地がある．

引き続き，システムの評価と改善に取り組む予定である．

図６　試作した日本語−手話翻訳システム

Fig.6　Prototype Japanese-JSL Translation System

7. おわりに

日本語の入力に対し，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システムを試作した．今後の開発を進めるにあたっては，本稿で紹介した各要素技術の改善を進めるとともに，生成した手話アニメーションの言語としての正確性，流暢性の評価法を開発することが必要である．その際には，ろう者のコミュニティーのご理解とご協力をいただけるようにすることが重要であると考える．

参考文献

1) 坂井　忠裕，半田　拓也，大河内　直之，伊福部　達:” 視覚障害者向けデジタル放送受信機とＵＩの開発評価”，電子情報通信学会 HCG（Human Communication Group）シンポジウム 2011, HCG2011-B3-4, pp.154-159,(2011).
2) 今井　亨，奥　貴裕，小林　彰夫:”音声認識によるリアルタイム字幕放送の進展”，情報処理学会研究報告　SLP　音声言語情報処理, vol.2011-SLP88, no.4,(2011).
3) 加藤　直人，金子　浩之，井上　誠喜，清水　俊宏，長嶋　祐二: “日本語－手話対訳辞書の構築－日本語語彙の拡張－”，電子情報通信学会 HCG（Human Communication Group）シンポジウム，HCG2009-I-3, (2009).
4) （財）全日本ろうあ連盟：“わたしたちの手話(1)～ (9) ”，（財）全日本ろうあ連盟出版局.
5) http://www.biovision.com/
6) 金子　浩之，浜口　斉周，道家　守，井上　誠喜: “TVML による手話アニメーションの一検討”，電子情報通信学会技術報告，WIT2008-82, pp.79-83, (2009).
7) 金子　浩之，加藤　直人，清水　俊宏，井上　誠喜，長嶋　祐二: “非手指動作を付加した手話映像生成に関する一検討”，電子情報通信学会 HCG（Human Communication Group）シンポジウム HCG2010-A5-2, (2010)

8) 加藤　直人, 宮﨑　太郎, 金子　浩之, 井上　誠喜, 梅田　修一, 比留間　伸行, 長嶋　祐二: “気象情報の日本語－手話 CG 翻訳”, 第 18 回言語処理学会年次大会発表論文集, P1-21, pp.275-278, (2012).
9) 加藤　直人: ”手話ニュースコーパスの構築”, 第 16 回言語処理学会年次大会発表論文集, PA2-5, pp.494-497, (2010).
10) http://www.nhk.or.jp/strl/TVML/index.html
11) 金子　浩之, 加藤　直人, 清水　俊宏, 井上　誠喜, 長嶋　祐二: “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 2010, A-19-6, (2010).