G-101

HMM を用いた手話単語の画像合成
Sign Language Synthesis by HMM

田口弘貴*　山本博史*
Taguchi Hiroki　Yamamoto Hirohumi

1　はじめに

健常者が使用する日本語と，それとは全く体系の異なる日本手話との間で，より話者の意図を正確に伝えられる通訳システムを目指す．本稿では，特に手話単語の画像合成についての手法を提案する．

2　背景

日本手話は健常者には手の動きで日本語の代替を図っている言語だと考えられがちだが，実際には腕の細かな動かし方や顔の表情なども含めてコミュニケーションを行う，言葉とは異なる語彙体系と文法的構造を持った自然言語である．日本語の語順通りに手話単語を当てはめた日本語対応手話とは異なり，習得には外国語を学ぶ時と同等の労力を要する．これはろう者が日本語を理解する際にも言えることで，日本語よりも先に日本手話を母語とした者にとって，健常者が使う言葉は外国語と差異はない．日本語の理解ができないと，文字を使ったコミュニケーションも難しくなり，より一層健常者とろう者の溝を深める原因となる [1].

二つの異なる言語文化を結びつけるには，両方に造詣の深い人間が必須だが，手話通訳者の数は慢性的に不足している．そこで本稿では日本手話・日本語間通訳を支援するシステムとして，特に手話単語の画像合成に関する手法について記す．

日本手話はろう者のコミュニティの中で作り上げられた自然言語で，根本的に日本語とは異なる言語文化である．例えば「今朝は朝ご飯食べた？」を日本手話の語順で記すと「朝ご飯，食べた，今朝」となり，最後に首を傾げるなどのジェスチャーで質問の意図を表す．他にも表情で感情を表したり（「暑いです」「暑いですね」等），腕の動かし方で細かな意味合いの表現を行う．このように単純な手指の動作だけでなく，その時のシチュエーションに合わせた自然な動作や，顔の表情などのボディランゲージが文法的に非常に重要な役割を担っている．これは手話の画像を合成すると考えた時，日本語対応手話との大きな違いになる．日本語対応手話は日本語と手話単語が一対一の対応をしており，ある単語を表す手話単語は，常に 1 パターンしかない．対して日本手話は意味合いに応じた細かい動作が何パターンも存在し，通訳先であるろう者に意味が確実に伝わるように自然な動きでなくてはならない．このような特徴のある日本手話の画像合成実現のため，本手法では隠れマルコフモデル (以下 HMM と表記) を採用した．

3　提案手法

3.1　HMM

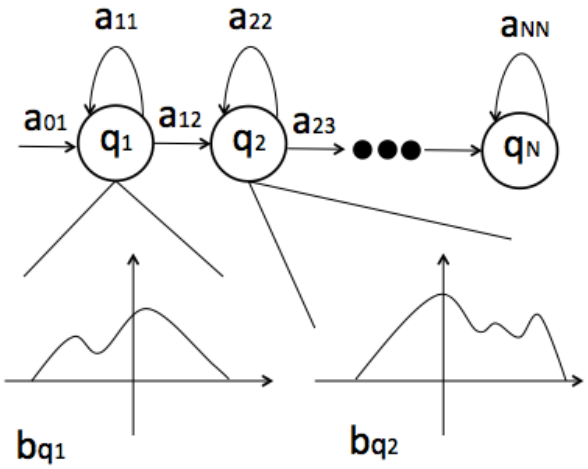

図 1　HMM 概要図

HMM とは連続的な信号パターンから，パラメータのわからないマルコフ性を持って遷移する信号の出現確率分布を構成する確率モデルである．マルコフ性とはその過程が今後持つ確率分布は，現在どのような状態かという

* 近畿大学，Kinki University

ことのみに依存し，それまでがいかなる過去であっても影響を受けない特性である．図 1 のように一方向に状態遷移を繰り返し，その状態が持つ確率分布から信号を出力する．状態の数を N，出力された信号の数を M とした時，HMM は状態の集合 $Q = \{q_1, q_2, ..., q_N\}$，状態 n の初期状態確率 π_{q_n}，状態の遷移確率 $a_{q_n,q_{n+1}}$，出力された信号集合 $O = \{o_1, o_2, ..., o_M\}$，信号パターンの出力確率分布 $b_{q_n}(o_n)$ を持つ．ある時刻 $t = 1$〜end において，時系列的に連続した状態遷移が $Qt = \{qt_1, qt_2, ..., qt_{end}\}$ となる場合に，出力信号が $Ot = \{ot_1, ot_2, ..., ot_{end}\}$ となる確率は式 (1) となる．

$$P(Ot|Qt) = \pi_{q_0} \prod_{t=1}^{end} a_{q_t,q_{t+1}} b_{q_t}(o_t) \tag{1}$$

この確率は Viterbi アルゴリズムを用いて，尤度が最も高くなるように高速に計算される．その特性から音声認識・合成によく用いられ，動作認識 [2][3] や音声に合わせた画像合成 [4][5] にも採用されている．日本手話の画像合成のためには 2 節に記したように，いくつものパターンから自然なものを，機械的ではなくより人間に近い動作として抽出しなくてはならない．よって手話の学習・動作抽出に HMM を採用する事で，ろう者が使う日本手話に近い画像合成が行えると考えた．

3.2 画像合成

手話単語の画像合成には，複数のパーツに分けた 3D モデルを用いた．指の中腹から先端，指の付け根から中腹のパーツがそれぞれ 5 つ，手のひらのパーツ，肘から手のひらまでの腕のパーツに分割されたものを一本の腕と考える．これが左右に 2 本存在することで手話を表現する．各パーツは独立に回転軸を持っており，腕から指の先端に向け従属関係が存在する．また腕のパーツのみ独自の移動軸を持っている．これらの 3 次元的な回転・移動は，各フレームの x，y，z 座標からなる 3 列のベクトル f として表現できる．これが時系列に沿って各フレームに存在しており，$t = 1$〜end において f を内包する行列を F とすると

$$F = \{f_1, f_2, .., f_t, f_{t+1}, ..., f_{end}\} \tag{2}$$

これが片手のパーツ × 2+ 腕の移動ベクトル × 2 だけ存在している．よって最終的に HMM から得られるベクトルの集合 V は式 (3) となる．

$$V = \{F_1, F_2, ..., F_{26}\} \tag{3}$$

4 結果

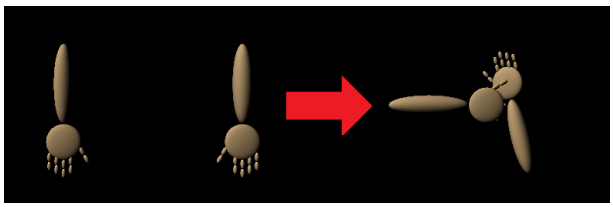

図 2　初期状態と「名前」の途中動作

図 2 は実際に手話単語の画像合成を行ったものである．左が待機状態にある初期位置で，右が「名前」を表す手話となっており，画像合成終了後は再び左の初期位置に戻る．画像合成を行うたびに状態遷移系列 $Qt = \{qt_1, qt_2, ..., qt_{end}\}$ が変化し，単一パターンではない，変化にとんだ手話単語の画像合成が行うことができた．また同一の人間でも全く同じ機械的な動作を行わないように，確率分布により細かなノイズが入ることで，自然に近い画像合成が行えた．学習する際のパラメータの調整をさらに詳細に行う事で，より手話の正確なニュアンスを伝える画像合成を実現する事ができると考えている．

5 結論

HMM を用いることで，固定された出力ではなく複数のパターンをもった，より人間の動作に近い手話単語の画像合成を行った．

今後の課題としては，まず顔モデルの対応が挙げられる．先述した通り，日本手話は顔の表情や頭の動きも文法に含まれており，同じ手話動作でも表現できる意味合いが大幅に増える．顔の画像合成を考えた際，顔のパーツを分割する事で手指と同じく各パーツごとの x，y，z 座標で表現できる．

次に学習データの収集が大きな課題である．2014 年現在，日本手話のコーパスは構築が進んでいない．これは手話の表記法が確立されておらず，日本語との対応が難しい事が一番の原因である [6]．また世間にあふれている言葉とは違い，日本手話は根本的に話者が少なく，手話を記録した映像データも数が少ない．たとえ日本手話を巧みに使用する話者に協力してもらいデータ収集を行うにしても，単純な手指の動作だけでなく表情一つで意味合いが変わる日本手話は，膨大なデータの記録と詳細な解析・クラスタリングが必要になる．近年のバリアフリーへの関心の高まりとともに日本手話の研究や記録化が進みつつあり，より効率的なデータ収集手法の確立が一番の課題であ

ると考えられる．

参考文献

[1] 池田尚志, 松本忠博, 点字と手話と自然言語処理, 電子情報通信学会 基礎・境界ソサイエティ Fundamentals Review，Vol.4，No.4，pp.282-292，2011.

[2] 稲邑哲也，HMM と人間の動作認識・教示・生成，日本ロボット学会誌, Vol.29, No.5, pp.419-422, 2011.

[3] 松尾直志，山田寛，白井良明，島田伸敬，HMM を利用した画像処理による手話単語の認識のための特徴抽出および状態分割，ヒューマインターフェース学会論文誌，Vol.15，No.1，pp.85-94，2011.

[4] 山本英里，中村哲，鹿野清宏，HMM を用いた音声からの唇動画像合成法，情報処理学会論文誌，Vol.39，No.5，pp.23-32，1998.

[5] 垣原清次，中村哲，鹿野清宏，HMM を用いた自然な発話動画像合成，電子情報通信学会論文誌 D-II，Vol.J83-D-II，No.11，pp.2498-2506，2000.

[6] 加藤直人，手話における言語資源の研究動向，NHK 技研 R&D，No.139，pp.10-19，2013.