隠れマルコフモデルに基づくハンドジェスチャー生成の検討
A Study of HMM-based Hand-Gesture Synthesis

高御堂 雄三 † **羽岡 哲郎** † **益子 貴史** † **小林 隆夫** †
Yuzo Takamido Tetsuo Haoka Takashi Masuko Takao Kobayashi

1. まえがき

近年，人間のコミュニケーションに用いられるジェスチャーに関する研究が盛んに行われている [1]〜[7]．その応用例として，コンピュータグラフィックスを用いたアニメーション製作，手話生成などが挙げられる．また，仮想対話エージェントなどを用いて，より扱いやすいユーザインターフェイスを実現するためには，音声の他にジェスチャーを交えたマルチモーダルコミュニケーションが必要となる．従って，人間のコミュニケーションにおいて重要な意味を持つしぐさ (ジェスチャー) を，自然で滑らかに生成する技術がますます重要になると考えられる．

ハンドジェスチャーや人体動作のアニメーションを生成する手法としては，モーションキャプチャによって取得した手や人体の動作を再現する方法 [1]〜[3] や，動作をコード化してコンピュータで合成する方法 [5]〜[7] などがある．前者は，自然なアニメーションを生成できる点において優れているが，取得した以外の動作を生成することは困難である．また，いずれの方法も，異なる動作の結合部分を滑らかにするため，補間処理を行って中間的な形状を生成している．しかし，補間方法によっては，生成された動作は滑らかではあるが，不自然なものになってしまうことがある．我々はこれまでに，自然で滑らかなアニメーションを実現するために，隠れマルコフモデル (HMM:Hidden Markov Model) に基づくパラメータ生成手法 [8] を用いたハンドジェスチャーアニメーション生成手法を提案している [9]．この手法では，取得した動作を用いて学習した HMM を基に，尤度最大の意味で最適なジェスチャーを生成する．その結果，動作の統計的性質を反映したジェスチャーアニメーション生成が可能になっている．この際，動的特徴量を考慮することによって，特別な補間処理や平滑化処理を行わずに，滑らかなアニメーション生成が可能となる．本論文では，前後の動作を考慮し，さらに，コンテキストクラスタリング [10] を行うことにより，学習データにないアニメーションを生成する手法を提案する．

2. ハンドジェスチャーアニメーション生成

精密な手の動きを表示するためには，手の動きを考慮した物理形状モデルを導入する必要がある．本研究では，文献 [9] と同様に，図 1 のような手の骨格を基にしたモデルを用いる．これは，透視投影によって平面に描写しており，球が関節と指先の位置を表し，円柱が関節の接続 (骨) を表す．この手形状モデルのパラメータは，各関節の角度である．

手形状モデルパラメータの生成には，図 2 に示すような HMM に基づくパラメータ生成システムを用いる．本研究では，一連の動作をいくつかの基本的なパターン

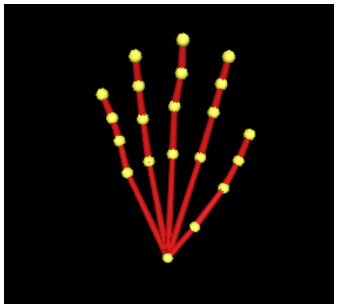

図 1: 画像による手形状の例

(プリミティブ) の列で表現する．以下では，これらの各基本パターンに付けられた名前を「ジェスチャーラベル」と呼ぶ．パラメータ生成システムの学習部では，与えられた学習データからジェスチャーラベルごとに尤度最大化基準に基づいて HMM の学習 [11] を行う．合成部では，合成したい任意の動作をジェスチャーラベル列で表現し，このラベル列に従って学習した HMM から **3.**. で述べるパラメータ生成手法を用いて手形状モデルのパラメータ列を生成する．

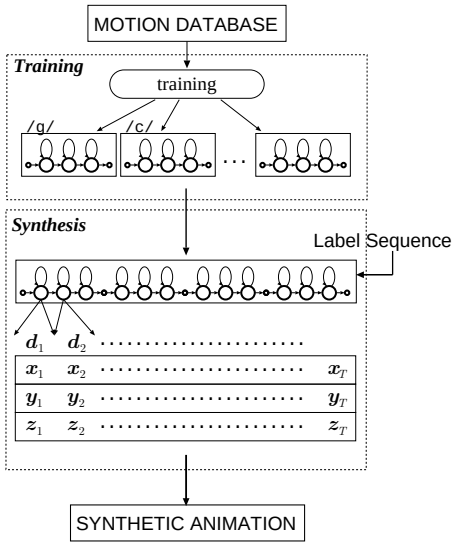

図 2: 合成システムブロック図

3. HMM に基づくパラメータ生成

3.1 動的特徴量

特徴ベクトル，すなわち，HMM が出力するベクトルとして，静的パラメータ (手形状モデルパラメータ)，お

†東京工業大学 大学院総合理工学研究科

よび，動的パラメータ (デルタパラメータ) を一つに結合したベクトルを用いる．

手形状モデルパラメータの次数を M として，時刻 (フレーム) t の動作パラメータを $\boldsymbol{x}_t$ (M 次の実ベクトル) とおく．長さ T の動作パラメータ列 $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T)$ が与えられたとき，$\Delta^{(0)}\boldsymbol{x} = \boldsymbol{x}_t$，$\Delta^{(1)}\boldsymbol{x}_t = \Delta\boldsymbol{x}_t$，$\Delta^{(2)}\boldsymbol{x}_t = \Delta^2\boldsymbol{x}_t$ とおけば，式 (1) によって，$\boldsymbol{x}_t$ の n 次デルタパラメータ $\Delta^{(n)}\boldsymbol{x}_t$ が定義される．

$$\Delta^n \boldsymbol{x}_t = \sum_{i=-L_n}^{L_n} w^{(n)}(i)\boldsymbol{x}_{t+i}, \quad 0 \le n \le 2 \tag{1}$$

ただし，$w^{(n)}(i)$ は $\Delta^n\boldsymbol{x}_t$ を計算するための係数を表し，$L_0 = 0$，$w^{(0)}(0) = 0$，$i \neq 0$ のとき $w^{(0)}(i) = 0$ である．

3.2 生成

連続出力分布型 HMM のパラメータセットを λ で表し，λ からある状態遷移系列 $\boldsymbol{q}$ に沿って長さ T の出力ベクトル系列 $(\boldsymbol{o}_1, \boldsymbol{o}_2, \ldots, \boldsymbol{o}_T)$ を生成することを考える．ここでは，HMM のそれぞれの状態 q は，平均 m_q，分散 σ_q^2 のガウス分布でモデル化した状態継続長分布 $p_q(d_q) = \mathcal{N}(d_q; m_q, \sigma_q^2)$ をもつとする．ただし，$p_q(d_q)$ は，状態 q が d_q フレーム継続する確率を表す．また，HMM は単一出力分布型 left-to-right モデルである．HMM の状態遷移系列を $\boldsymbol{q} = (\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_T)$ とし，$\boldsymbol{q}$ に沿って出力されるパラメータ系列からなるベクトルを $\boldsymbol{O} = [\boldsymbol{o}_1', \boldsymbol{o}_2', \ldots, \boldsymbol{o}_T']'$ とする．与えられた HMM λ に対し，出力ベクトル $\boldsymbol{O}$ は，$P(\boldsymbol{q}, \boldsymbol{O}|\lambda)$ を $\boldsymbol{O}$ と $\boldsymbol{q}$ に関して最大化することにより得られる．状態遷移確率 $P(\boldsymbol{q}|\lambda)$ にかける重みを α とすると，出力ベクトル $\boldsymbol{O}$ と状態遷移系列 $\boldsymbol{q}$ の同時生起確率 $P(\boldsymbol{q}, \boldsymbol{O}|\lambda)$ の対数は，

$$\log P(\boldsymbol{q}, \boldsymbol{O}|\lambda) = \alpha \log P(\boldsymbol{q}|\lambda) + \log P(\boldsymbol{O}|\boldsymbol{q}, \lambda) \tag{2}$$

と表すことができる．ここで，α を十分に大きくした場合，状態遷移系列 $\boldsymbol{q}$ は状態遷移確率 $P(\boldsymbol{q}|\lambda)$ のみによって決定される．HMM は left-to-right モデルであるため，状態遷移確率 $P(\boldsymbol{q}|\lambda)$ は状態継続長分布 $p_q(d_q)$ のみにより表されるとすると，

$$\log P(\boldsymbol{q}|\lambda) = \sum_{i=1}^{N} \log \mathcal{N}(p_i(d_i); m_i, \sigma_i^2) \tag{3}$$

と表される．ここで N は HMM の状態数である．このとき，$P(\boldsymbol{q}|\lambda)$ を最大にする $\boldsymbol{q}$ は

$$\boldsymbol{q} = (\underbrace{1, 1, \ldots, 1}_{\lceil m_1 + 1/2 \rceil}, \underbrace{2, 2, \ldots, 2}_{\lceil m_2 + 1/2 \rceil}, \ldots, \underbrace{N, N, \ldots, N}_{\lceil m_N + 1/2 \rceil}) \tag{4}$$

$$T = \sum_{i=1}^{N} \lceil m_i + 1/2 \rceil, \tag{5}$$

ただし，$\lceil m \rceil$ は，m を超えない最大の整数，と求まる．

次に，このようにして得られた $\boldsymbol{q}$ に対し，出力確率 $P(\boldsymbol{O}|\boldsymbol{q}, \lambda)$ を最大にする $\boldsymbol{O}$ を求めることを考える．出力確率の対数は

$$\begin{aligned} \log P(\boldsymbol{O}|\boldsymbol{q}, \lambda) &= -\frac{1}{2}\log|\boldsymbol{U}| - \frac{3MT}{2} log 2\pi \\ &\quad -\frac{1}{2}(\boldsymbol{W}\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{U}^{-1}(\boldsymbol{W}\boldsymbol{x} - \boldsymbol{\mu}) \end{aligned} \tag{6}$$

と表すことができる．ここで，

$$\boldsymbol{\mu} = [\boldsymbol{\mu}_{q_1}', \boldsymbol{\mu}_{q_2}', \ldots, \boldsymbol{\mu}_{q_T}']' \tag{7}$$

$$\boldsymbol{U} = \mathrm{diag}[\boldsymbol{U}_{q_1}, \boldsymbol{U}_{q_2}, \ldots, \boldsymbol{U}_{q_T}] \tag{8}$$

であり，$\boldsymbol{\mu}_{q_t}$ および $\boldsymbol{U}_{q_t}$ はそれぞれ状態 q_t の平均および共分散，$\mathrm{diag}[\boldsymbol{U}_{q_1}, \boldsymbol{U}_{q_2}, \ldots, \boldsymbol{U}_{q_T}]$ は行列 $\boldsymbol{U}_{q_1}, \boldsymbol{U}_{q_2}, \ldots, \boldsymbol{U}_{q_T}$ を対角に並べた $3MT \times 3MT$ 行列である．$\boldsymbol{O}$ は，静的なパラメータ系列からなるベクトル $\boldsymbol{x} = [\boldsymbol{x}_1', \boldsymbol{x}_2', \ldots, \boldsymbol{x}_T']'$ と

$$\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_T]' \tag{9}$$

$$\boldsymbol{w}_t = \begin{bmatrix} (\boldsymbol{w}_t)_{11} & (\boldsymbol{w}_t)_{12} & \ldots & (\boldsymbol{w}_t)_{1T} \\ (\boldsymbol{w}_t)_{21} & (\boldsymbol{w}_t)_{22} & \ldots & (\boldsymbol{w}_t)_{2T} \\ (\boldsymbol{w}_t)_{31} & (\boldsymbol{w}_t)_{32} & \ldots & (\boldsymbol{w}_t)_{3T} \end{bmatrix}' \tag{10}$$

$$(\boldsymbol{w}_t)_{ij} = w^{(i)}(j - t)\boldsymbol{I}_M \tag{11}$$

から

$$\boldsymbol{O} = \boldsymbol{W}\boldsymbol{x} \tag{12}$$

と計算される．ただし，$\boldsymbol{I}_M$ は $M \times M$ の単位行列である．従って，$\boldsymbol{O}$に関する $P(\boldsymbol{O}|\boldsymbol{q}, \lambda)$ の最大化は，静的なパラメータ系列からなるベクトル $\boldsymbol{x} = [\boldsymbol{x}_1', \boldsymbol{x}_2', \ldots, \boldsymbol{x}_T',]'$ に関する最大化となる．式 (6),(12) より，$P(\boldsymbol{O}|\boldsymbol{q}, \lambda)$ を最大化する $\boldsymbol{x}$ は，次式の連立線形方程式

$$\boldsymbol{W}'\boldsymbol{U}^{-1}\boldsymbol{W}\boldsymbol{x} = \boldsymbol{W}'\boldsymbol{U}^{-1}\boldsymbol{\mu} \tag{13}$$

の解として与えられる．

3.3 決定木に基づくコンテキストクラスタリング

基本パターンを HMM でモデル化する際，一つの基本パターンであっても前後のパターンによって形状や動作が異なると考えられる．ここでは前後のパターンを考慮した基本パターンをトライモーションと呼ぶこととし，トライモーション毎に HMM を学習することとする．しかし基本パターンの種類が増加すると，全ての可能なトライモーションを含む学習データを用意することは難しいため，トライモーション HMM に対して決定木に基づくコンテキストクラスタリング [10] を適用する．ここでコンテキストとは，当該および前後の基本パターンなど，手形状に影響を与えると考えられる要因の組み合わせ，つまり，本研究の場合はトライモーションを指す．

決定木は 2 分木であり，それぞれの節ごとにコンテキストを 2 つに分割する質問が用意されている．すべてのモデルは根からそれぞれの節の質問に従って木を下って行き，葉のうちのどれかに達するため，いったん決定木を構築すれば，学習データに出現しないコンテキストについても，対応するモデルが一意に決定される．また，質問の作り方によっては，学習データに含まれない基本パターンを合成することも可能となる．

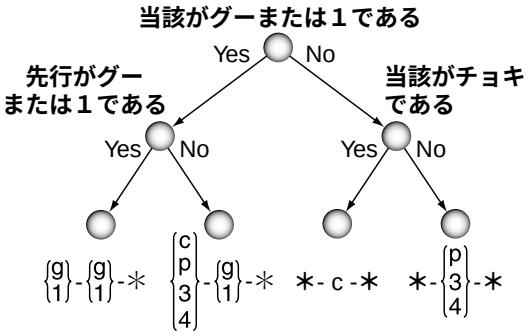

図 3: クラスタリング例

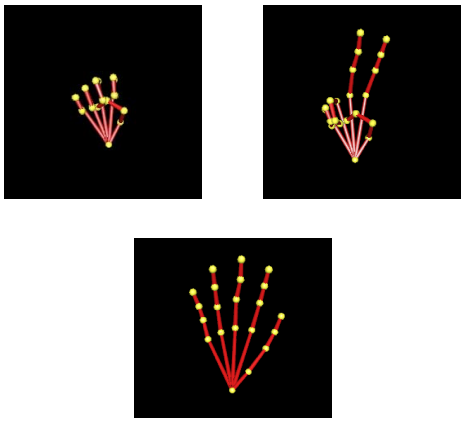

図 4: 実験により取得したジェスチャー

4. 実験

3. で述べた HMM に基づくパラメータ生成手法を用いて，ハンドジェスチャーアニメーションの生成実験を行なった．

4.1 実験条件

手の物理形状モデルとしては，図 1 に示すような，関節が 15 個 (各指 3 個) のものを用いた．このモデルは，各関節の Z，X，Y それぞれの軸に関する回転角度をパラメータとしている．各関節の自由度は 3 であるので，手形状モデル全体の自由度は 45 である．すなわち，手形状パラメータの次数は 45 次となる．

学習データとして，CyberGlove を用いて図 4 に示すようなグー (g)，チョキ (c)，パー (p) の 3 種類をランダムに 100 回続けて行ったジェスチャーを収録した．収録データのフレームレートは 30 フレーム/秒，総フレーム数は 6214 フレーム，データ長は約 207 秒であった．取得したデータの各フレームに，コンピュータグラフィックス表示された形状を見ながら，手動でラベル付けを行った．これらに人差し指のみを立てた状態 (1)，人差し指から薬指までの 3 本の指を立てた状態 (3)，人差し指から小指までの 4 本の指を立てた状態 (4) の 3 つのパターンを加えた合計 6 種類のパターンを基本パターンのセットとした．

HMM はスキップなしの 5 状態 left-to-right モデルを用いた．また，各状態の出力ガウス分布の共分散行列は対角とした．各トライモーション毎に HMM の学習を行った後，決定木に基づくコンテキストクラスタリングを行った．この際，コンテキストと手形状の対応は各指によって異なるため，指毎に異なる質問のセットを用意し，別々にクラスタリングを行った．例えば，中指のクラスタリングには以下に示す 9 つの質問のセットを用いている．

- 先行パターンがグーまたは 1 である
- 先行パターンがチョキである
- 先行パターンがパー，3，または 4 である
- 当該パターンがグーまたは 1 である
- 当該パターンがチョキである
- 当該パターンがパー，3，または 4 である
- 後続パターンがグーまたは 1 である
- 後続パターンがチョキである
- 後続パターンがパー，3，または 4 である

この質問のセットを用いて HMM の第 4 状態の中指の部分に対してクラスタリングした例を図 3 に示す．図 3 はクラスタリングにより作成された決定木の 2 層目までと，各ノードの質問で分類されて得られたトライモーションのセットを表す．図中 a-b-c は先行パターンが a，当該パターンが b，後続パターンが c であることを表しており，{·} は括弧に含まれるパターンからの選択を，* は全てのパターンからの選択を表している．このようにして決定木を作成することにより，学習データに含まれない 1，3，4 のパターンも生成することが可能となる．

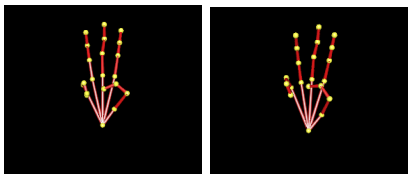

図 6: 3 つ指のジェスチャー (左：合成　右：オリジナル)

4.2 結果

学習データに含まれないジェスチャーラベル列

$$g \rightarrow 1 \rightarrow c \rightarrow 3 \rightarrow 4 \rightarrow p$$

に対して生成を行った．生成されたアニメーションの長さは 366 フレーム (約 12 秒) であった．図 5 に，生成されたアニメーションの一連の流れ図を示す．順番に指が立って行く変化過程，形状が滑らかに変化している様子が確認できる．しかし，指同士が交錯している部分もあり，これは，生成アルゴリズムが，指同士の衝突回避などの物理的拘束条件を考慮していないためや，実際の手と手形状モデルが一致していないためであると考えられる．

次に，生成された 3 つ指のジェスチャーと，あらかじめ学習データとは別に収録した 3 のジェスチャーを図 6 に示す．図から学習データにないジェスチャーも実際と似た手形状が生成されていることが確認できる．

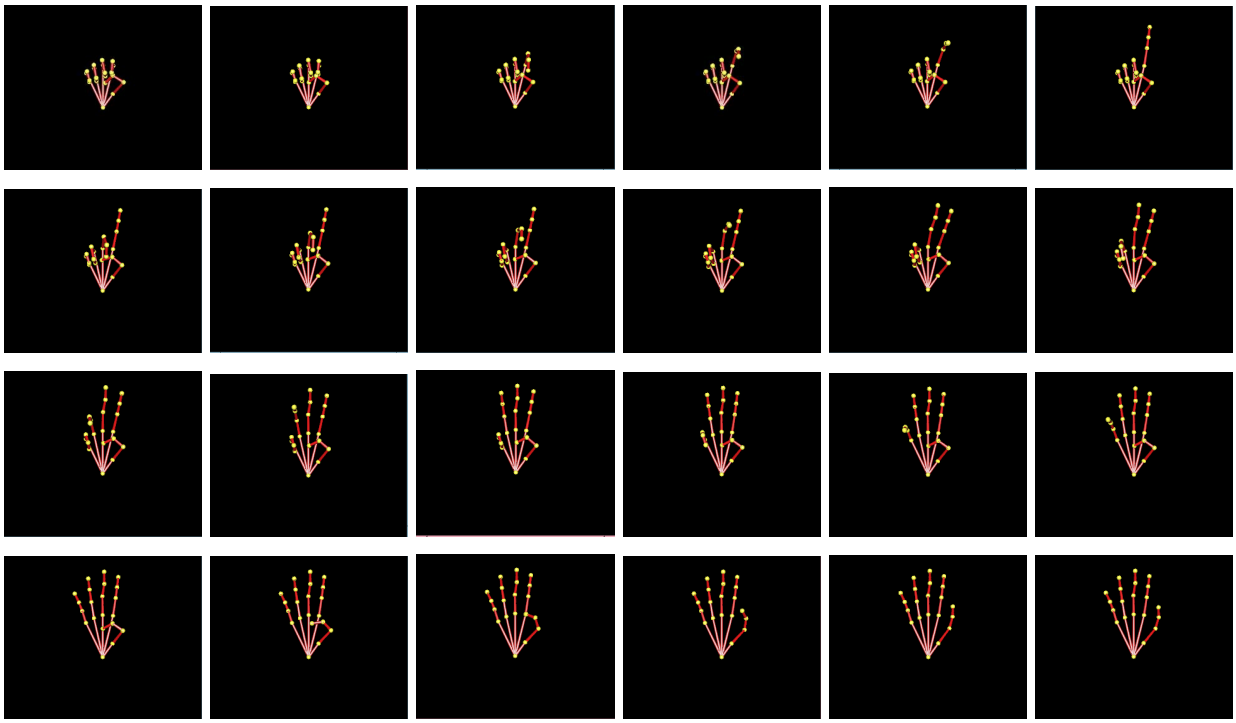

図 5: 合成結果

5. おわりに

隠れマルコフモデルに基づくパラメータ生成手法を用いた，ハンドジェスチャーアニメーション生成手法について述べた．学習データにないジェスチャーを生成する実験を行った結果，自然で滑らかなジェスチャーを生成することができた．生成された学習データにないジェスチャーと，CyberGlove によってあらかじめ収録したジェスチャーとを比較した結果，同等のジェスチャーが生成されたいることが確認できた．

しかしながら，衝突回避を行っていないため，一部で，指同士が交錯するようなジェスチャーが生成されてしまった．また，アニメーション生成部分は，骨格を表示する程度の簡単なものであり，自然な手の表現とは言えない．そこで，今後は，こうした問題点を解決してゆくとともに，複雑なジェスチャーの生成や全身の生成，機械ロボットを動かすことによるジェスチャーの表現，ジェスチャーの「自然さ」を評価する方法の確立などについても検討を行って行く予定である．

参考文献

[1] 猪木 誠二，渡辺 錬士， 呂 山， “手話アニメーション作成・編集ツール，“信学論 (D-I) vol.J84-D-I, no.6, pp.987–995, Jun. 2001.

[2] 崎山 朝子，太平 英二，佐川 浩彦，大木 優，池田 尚司，“リアルタイム手話アニメーションの合成方法，“信学論 (D-II),vol,J79-D-II, no.2, pp.182–190, Feb. 1996.

[3] S.Lu, S.Igi, H.Matsuo, and Y.Nagashima, ”Towards a dialogue system based on recognition and systhesis of Japanese sign language,” Proc. of Intl. Gesture Workshop, GW’97, pp.259–271, Bielefeld, Germany, Sep. 1997.

[4] 長嶋 裕二，“手話情報学の現状と課題，“ 信学技報，PRMU99-141，Nov，1999.

[5] S.Gibet, T.Lebourque, and P.Marteau, ”High-level specification and animation of communicative gesture,” Journal of Visual Languages and Computing,vol.12, no.6, pp.657-687, 2001.

[6] T.Lebourgue, and S.Gibet, ”A complete system for the speci fication and the generation of sign language gesture,” Proc. of Intl. Gesture Work shop, GW’99, LNAI 1739, pp.227–238, Gif-sur-Yvette, France, Mar. 1999.

[7] I.Wachsmuth, and S.Kopp, ”Lifelike gesture synthesis and timing for conversational agents,” Proc. of Intl. Gesture Workshop, GW 2001, LNAI 2298,pp.120–133, London, UK, Apr. 2001.

[8] 徳田恵一，益子貴史，小林隆夫，今井聖，“動的特徴量を用いた HMM からの音声パラメータ生成アルゴリズム，” 日本音響学会誌，vol.53, no.3, pp.192–200, Mar. 1997.

[9] 羽岡哲郎，益子貴史，小林隆夫，“隠れマルコフモデルに基づくハンドジェスチャーアニメーション生成，” 信学技報，vol.102, no.519, pp.43–48, Dec. 2002.

[10] j.j Odell, The Use of Context in Large Vocabulary Speech Recognition, Ph.D. disser tation, Cambridge University, 1995.

[11] A.P.Dempster, N.M.Laird, and D.B.Rubin, “Maximum likelihood from incomplete dat a via the EM algorithm,” Journal of Royal Statistical Society, Series B, vol39, pp.1–38, 1977.