

隠れマルコフモデルに基づく指文字動画像生成

Motion Generation for Japanese Finger Language Based on Hidden Markov Model

森 健史[†] 南角 吉彦[†] 宮島 千代美[‡] 徳田 恵一[†] 北村 正[†]
Kenji Mori Yoshihiko Nankaku Chiyomi Miyajima Keiichi Tokuda Tadashi Kitamura

1. まえがき

日本では手話の習得率は非常に低く、聴覚障害者のコミュニケーションの幅は狭められている。この状況を打開するため、近年、コンピュータを利用した聴覚障害者との対話支援システムが注目されている。

本研究では、手話の中で固有名詞等の表現に用いられる指文字を対象とした動画像生成システムを構築する。HMM(隠れマルコフモデル)音声合成の枠組みを指文字データに適用することにより、任意の指文字列を滑らかに生成することができる。HMM音声合成技術を適用する際に重要となる、指文字データのコンテキストの設計、コンテキストクラスタリングに必要な質問の定義を新たに行った。また、HMMから滑らかなパラメータ系列を生成するための動的特徴量の効果について評価実験を行った。

2. HMMに基づく指文字動画像生成システム

HMMに基づく指文字動画像生成システムは、学習部と生成部から構成される。学習部では、学習データから特徴ベクトルを抽出し、各指文字に対応したHMM(指文字HMM)を学習する。単語毎でなく、指文字毎にモデルを学習することで、学習データには存在しない任意の指文字列を生成することができる。生成部では、生成する単語に基づいて指文字HMMを連結した単語HMMを作成し、特徴ベクトル系列を生成する。合成システムは、学習部・生成部ともに尤度最大化(ML)基準によって導出されており、数学的に整備された枠組となっている。

2.1 動的特徴量に基づくパラメータ生成

HMMの状態出力確率をガウス分布と仮定した場合、各状態から生成されるパラメータは、ガウス分布の平均ベクトルと一致する。このため、ある状態系列が与えられたときに生成されるパラメータ系列は、状態が継続している区間では平均ベクトルが続き、状態の切り替わる部分で不連続な系列となってしまう。この問題を解決するために、動的特徴量に基づくパラメータ生成アルゴリズム[2]が提案されている。この手法を用いることで、特徴ベクトルに、データの速度と加速度に対応する、1次と2次の回帰係数を連結したベクトル列によりHMMを学習し、パラメータ生成時には、速度と加速度を考慮することにより、滑らかなパラメータ系列を生成することができる。本研究では、動的特徴量を用いたパラメータ生成アルゴリズムを指文字動画像生成に適用し、その有効性を実験的に評価する。

2.2 指文字データのコンテキストクラスタリング

指文字HMMの学習を行う際に、各指文字モデルを前後の指文字、手の向き、手の動き、指の屈伸などの指の動作に関する情報を考慮したコンテキスト依存モデルへ拡張する。これにより、より多くの情報を考慮した精度の高いモデルを構築することができる。しかし、詳細なモデル化が可能となる一方、モデル数は指数的に増加し、学習外データの出現頻度の増加や各指文字HMMに対する学習データ量の不足といった問題が起こる。この問題を解決するために、決定木に基づくコンテキストクラスタリング[2]が用いられる。この手法では、コンテキ

表1: 設計したコンテキスト

コンテキスト	種類
手の向き	正面, 縦, 上, 下, 横, 裏, 動
手の動き	普通音, 濁音, 拗音, 発音, 長音
親指の屈伸	普通, 上げる, 握る, 半立ち, 曲げる
親指以外の指の屈伸	上げる, 握る, 半立ち, 曲げる
単語中の当該モデルの位置	位置情報
単語長	単語長情報

表2: クラスタ分割用の質問

対応コンテキスト	質問内容
モデル	指文字の表す文字は何か?
手の向き	手の向きは何か?
手の動き	手の動きは何か?
親指の屈伸	親指の屈伸状態は何か?
親指以外の4本の指の屈伸	親指以外の指は同時に屈伸しているか?
人差し, 中指の屈伸	人差し, 中指は同時に屈伸しているか?
中, 薬指の屈伸	中, 薬指は同時に屈伸しているか?
薬指, 小指の屈伸	薬指, 小指は同時に屈伸しているか?
人差し, 中, 薬指の屈伸	人差し, 中, 薬指は同時に屈伸しているか?
中, 薬, 小指の屈伸	中, 薬, 小指は同時に屈伸しているか?

トを2つのグループ(クラスタと呼ばれる)に分割する“質問”を用いて、2分木を構築する。結果として、全てのコンテキストは有限なクラスタに分類されるため、学習データに出現しなかった指文字にも対応できる。また、クラスタ内で学習データを共有することにより、モデルパラメータの推定精度を高めることができる。システムの構築にあたり、設計したコンテキストを表1に示す。また、コンテキストについて決定木を作成する際の質問を表2に示す。但し、コンテキスト依存モデルに拡張する際には、先行、後続指文字を考慮したモデルとした。

3. 指文字データベースの構築

本研究では、指文字合成システムの構築と評価のため、新たに指文字データベースを構築した。

収録単語セットは、指文字の出現エントロピーが最大となるように設計された単語セット[4]を用いた。単語数は、学習用が902単語、評価用が1000単語である。

指文字データの収録には、Polhemus社の3SPACE Fastrak, Immersion社のCyberGloveを使用した。3SPACE Fastrakからは手首の位置と回転角について6種類のデータが得られ、CyberGloveからは指の角度情報について18種類のデータが得られる。サンプリング周波数は103Hzである。収録された座標データは、描画プログラム[3]により指文字動画像として再現することができる。

4. 指文字合成実験

本研究で構築した指文字動画像生成システムを用いて合成実験を行った。

4.1 実験条件

学習データは、収録した学習用902単語を用い、24次元の座標・角度データを直接特徴ベクトルとして用いた。動的特徴量として、1次、2次の回帰係数を用いた。各指文字HMMの状態数は30とした。また、コンテキストクラスタリングには、MDL(記述長最小化)基準を用いた。

[†]名古屋工業大学 大学院工学研究科 情報工学専攻

[‡]名古屋大学 大学院情報科学研究科 メディア科学専攻

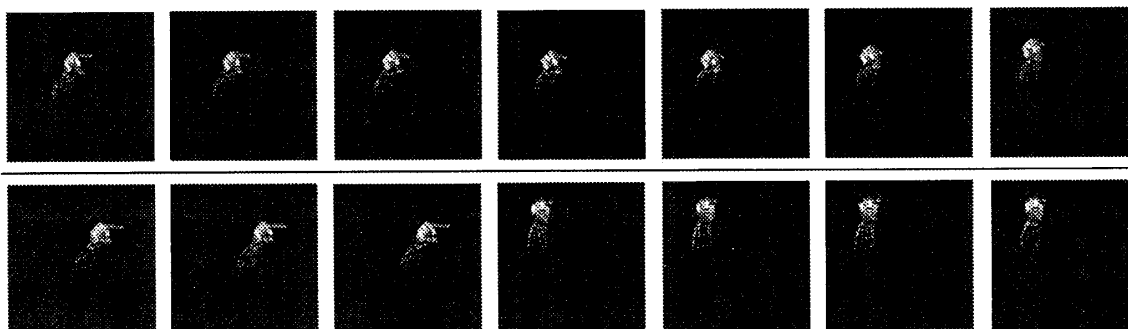


図2: 生成された動画像における動的特徴量の効果。(上) 動的特徴量有り, (下) 動的特徴量無し。

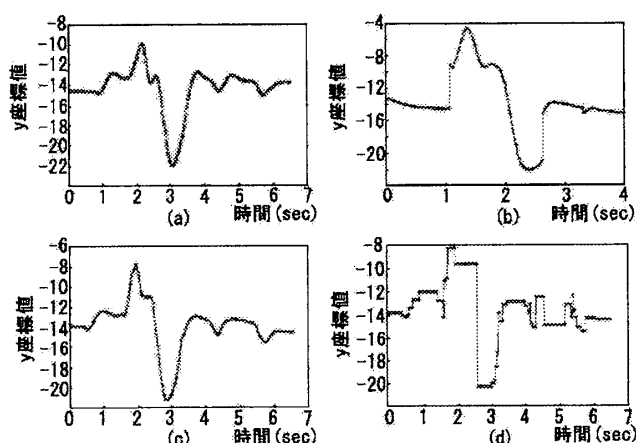


図1: 座標データの変化。(a) 収録データ (b) 従来法 (c) 提案法(動的特徴量あり) (d) 提案法(動的特徴量なし)

4.2 従来型システムとの比較実験

従来法として、指文字毎に辞書ファイルを作成し、それからデータを参照し、連結することで生成を行うシステム(波形連結型)を取り上げ、提案手法との比較を行った。波形連結型の手法では、辞書ファイルに収録された動作区間では、実際の指文字データが保持されているため、学習データに忠実な動作が再現できる。しかし、波形連結型の問題点は、コンテキストを考慮した詳細なモデル化には辞書ファイルが巨大化してしまう点、文字間の適切な補間アルゴリズムの設計が必要となる点が挙げられる。今回は提案手法と同じ条件とするため、接続規則については考慮していない。図1に、指文字単語‘チンタオ’における手首のy座標の変化を示す。波形連結型(図1(b))では接続規則を考えていないため、文字と文字の間の不連続な変化を起こしている。一方、提案手法(図1(c))では、特別な接続規則を考慮していないにも関わらず、当該文字だけでなく先行、後続文字を考慮したモデル化することに加え、動的特徴量を考慮することにより、文字間のデータを統計的に合成していることがわかる。以上より、提案手法の優位性を示すことができた。

4.3 動的特徴量の有効性に関する実験

4.3.1 客観評価実験

学習データを作成する際に、付加した動的特徴量の有効性について、比較実験を行った。図1(c)が動的特徴量有りの場合で、図1(d)が動的特徴量無しの場合である。図1(d)の動的特徴量を考慮しない場合は、HMMの

1つの状態が継続する間、平均ベクトルが出力されるため、変化は階段状になり、不自然な動作となる。これに対し、図1(c)の動的特徴量を考慮した場合、すべての区間において滑らかに変化している。実際に合成した動画像を、コマ送りにした画像を図2に示す。上段が動的特徴考慮をした場合、下段が考慮しない場合となっている。動的特徴を考慮しない場合、同じ状態が続くコマと一度大きく変化するコマになっているのが確認できる。

4.3.2 主観評価実験

動的特徴の考慮による改善が、実際に動画像で視認できるか確認するための実験を行った。一対比較試験(XAB型)に従い、実験を行った。被験者は手話の知識のない6名の健常者とし、評価点は動作の自然性に着目した。5つのデータを用意し、再生順、単語の登場順はランダムとした。結果は、動的特徴を考慮した方が自然であるという回答が全体の97%となり、動的特徴を考慮した合成データは人間が評価しても自然な動作になっていると言える。

5. むすび

本研究では、任意の指文字列を3Dアニメーションとして再生することのできる、指文字動画像生成システムを構築した。また、提案手法の利点、動的特徴導入の効果について評価実験を行い、提案手法の優位性を示すことができた。

謝辞

本研究を行うにあたり、指文字動画像の描画プログラムを提供して頂いた名古屋工業大学、舟橋健司 助教授、指文字単語セットを提供して頂いた名古屋大学 武田研究室、江本祐太 氏に深謝致します。

参考文献

- [1] 吉村貴克, 徳田恵一, 益子貴史, 小林隆夫, 北村正, “HMMに基づく音声合成におけるスペクトル・ピッチ・状態継続長の同時モデル化,” 電子情報通信学会, Vol.J83-D-2 No.11, pp.2099-2107, 2000.
- [2] 徳田恵一, 益子貴史, 小林隆夫, 今井聖, “動的特徴を用いた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, 1997.
- [3] 舟橋健司, 安田考美, 横井茂樹, 鳥脇純一郎, “3次元仮想空間における仮想手による物体操作モデルと一実現法,” 電子情報通信学会論文誌 D-II, Vol.J81-D-II, No.5, pp.822-831, 1998.
- [4] 江本祐太, 宮島千代美, 伊藤克亘, 武田一哉, “日本語指文字認識・合成用コーパスの構築,” 情報科学技術フォーラム, 7H-3, Sept.2005 (発表予定).