HMM 音声合成ツールキットの概要 *

◎全 炳河　　吉村 貴克　　田村 正統 †　　益子 貴史 †　　徳田 恵一　　(名工大, † 東工大)

1. はじめに

HMM よりスペクトル列や基本周波数パターンなどの音声パラメータ列を生成し, 音声波形を合成する HMM 音声合成システム [1] を提案した. HMM 音声合成システムを構築するためのツール群は, HTK[2] のツール群の拡張として開発されている. 本稿では, この HMM 音声合成ツールキットの概要について述べる.

2. HMM 音声合成システム

HMM 音声合成システムの概要を図 1 に示す. システムは学習部, 合成部から構成される.

学習部では, 音声データからスペクトルパラメータとしてメルケプストラム, 基本周波数パラメータとして対数基本周波数を求め, これらの 1 次及び 2 次の動的特徴量をフレーム毎に結合して特徴ベクトルとする. スペクトルは通常の連続分布 HMM, 基本周波数は多空間上の確率分布に基づいた HMM(MSD-HMM)[3][4], 継続長は HMM の各モデルの状態継続長を多次元のガウス分布 [1] でそれぞれモデル化される. スペクトル, 基本周波数, 継続長は音素, アクセント型, 形態素など様々な要因によって変動することから, モデル化の際, これらの変動要因の組み合わせ (コンテキスト) 毎に別々にモデル化し, MDL 基準に基づく決定木によるコンテキストクラスタリング [5] を適用して分布の共有化を行っている. この際, スペクトル, 基本周波数, 継続長はそれぞれ異なる変動要因に依存すると考えられるため, 別個にクラスタリングを行う. また, MDL 基準による Tree-based クラスタリングは, [1] において, MSD-HMM に対して拡張されている.

合成部では, まず, 合成する文章を変動要因を考慮したラベル列に変換し, 得られたラベル列に従ってコンテキスト依存 HMM を結合し, 文章に対応する HMM を構成する. 次に継続長分布に従って各状態の継続長を決定し, メルケプストラム列及び基本周波数パターンをパラメータ生成アルゴリズム ([6] の case 1) により生成し, MLSA フィルタ [7] を用いて波形を合成する.

この手法の利点としては, HMM のパラメータを適切に変化させることにより, 合成音の声質を容易に変化させることができる点である. 話者補間 [8], 話者適応 [9], 固有声 [10] などの手法が適用され, 声質が変化することを確認している.

3. HMM 音声合成ツールキット

HMM 音声合成ツールキットは, HTK[2] の機能追加版として提供される. 追加された主要な機能について以下に示す.

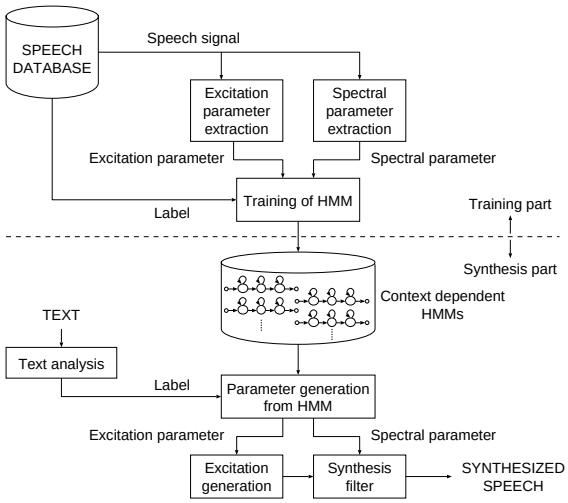

図 1．HMM-based speech synthesis system

MDL 基準による Tree-based クラスタリング

ML 基準に基づく Tree-based クラスタリングは, HMM のパラメータを共有する際に広く用いられている. しかし, 分割停止条件を内包していないため, 分割による尤度変化量の閾値をヒューリスティックに与える必要がある. これに対して, MDL 基準に基づく Tree-based クラスタリング [5] では, 学習データ量に応じて尤度変化量の閾値が自動的に設定される. 本ツールキットでは, Tree-based クラスタリングの際の基準を ML 基準, MDL 基準のどちらかを選択できる.

各ストリーム毎の Tree-based クラスタリング

HMM 音声合成システムでは, スペクトルと励振源の情報を合わせた特徴ベクトルを利用する. スペクトルと励振源は異なる変動要因により変化すると考えられるが, HMM の状態単位での Tree-based クラスタリングでは, それぞれに異なる決定木を構築できない. 本ツールキットでは, 別のストリームでモデル化したスペクトル, 励振源情報に対してストリーム毎に Tree-based クラスタリングを適用することができる.

多空間確率分布による出力分布

基本周波数パターンは, 有声区間では 1 次元の連続値, 無声区間では無声を表す離散シンボルとして観測されるため, 通常の連続 HMM や離散 HMM ではモデル化できない. そこで, 基本周波数パターンを有声を表す N 次元空間からの出力と無声を表す 0 次元空間からの出力が時間的に混在した系列としてとらえ, 多空間上の確率分布に基づく HMM を用いてモデル化する. 本ツールキットでは, 若干の制約はあるが, HMM の出力分布を多空間確率分布を用いてモデル化できる.

∗ A toolkit for HMM-based speech synthesis

By Heiga Zen, Takayoshi Yoshimura, Masatsune Tamura†, Takashi Masuko†, and Keiichi Tokuda
(Nagoya Institute of Technology, †Tokyo Institute of Technology)

HMM からのパラメータ生成

HMM 音声合成システムでは, 音声パラメータを HMM から生成し, 音声波形を合成する. パラメータ生成アルゴリズムには幾つかの種類があるが [6], 本ツールキットでは, コレスキー分解による高速なパラメータ生成アルゴリズムが実装されている.

多次元ガウス分布による状態継続長のモデル化

HMM 音声合成では継続時間長を制御するために, 多次元ガウス分布やガンマ分布を用いて HMM の各状態の継続長をモデル化している. 次元数は HMM の状態数に等しく, n 次元目は HMM の第 n 状態継続長分布に対応している. 状態継続長モデルの学習は, HMM に状態継続長を含めて行う方法があるが, 計算時間が非常にかかり効率が悪い. 本ツールキットでは, 状態継続長モデルを含まない HMM の連結学習の際に構築されるトレリス上の状態滞在確率を用いて, 状態継続長分布を計算している.

4. 応用

HMM 音声合成ツールキットと他のツールキットやその応用の概念図を図 2 に示す. HMM 音声合成ツールキットは, HMM 音声合成システムの構築を行うだけでなく, 次にあげるような利用法が考えられる.

擬人化音声対話エージェント

HMM 音声合成は, 擬人化音声対話エージェント [11] の音声合成部 [12] における波形生成エンジンとして利用されている. 本ツールキットを用いて, 対話エージェントで利用できる音響モデルを容易に構築できる.

Festival Speech Synthesis System

音声合成における代表的なフレームワークの一つである Festival speech synthesis system[13] の波形生成エンジンとして HMM 音声合成システムを組み込むことができる. 文献 [14] において, 言語解析部に Festival より提供されているものを用い, 波形生成部に HMM 音声合成を使用した英語 TTS が開発されている.

音声認識

HMM 音声合成ツールキットで提供される機能の幾つかは, 音声合成だけでなく, 音声認識においても有用であると考えられる. ストリーム単位の Tree-based クラスタリングは, 現在広く利用されている状態単位の共有構造を持つ HMM と異なるトポロジーの HMM を構築できる. また, [15] では, パラメータ生成アルゴリズムにより得られた系列を音声認識に利用している.

単位選択型音声合成

HMM を音声合成に利用する試みは, これまで数多く行われている. 単位選択型音声合成方式では, ユニット選択の際のコスト関数をどのように与えるかが重要な問題となってくる. 文献 [16] では, HMM 音声合成システムで用いる統計量を単位選択に用いている.

5. まとめ

本稿では, HMM 音声合成システムツールキットの概要について述べた. 本ツールキットは, [17] において, 公開する予定である. 今後は, HMM 音声合成システムの開発で有益と思われた機能を順次統合し, 公開する予定である.

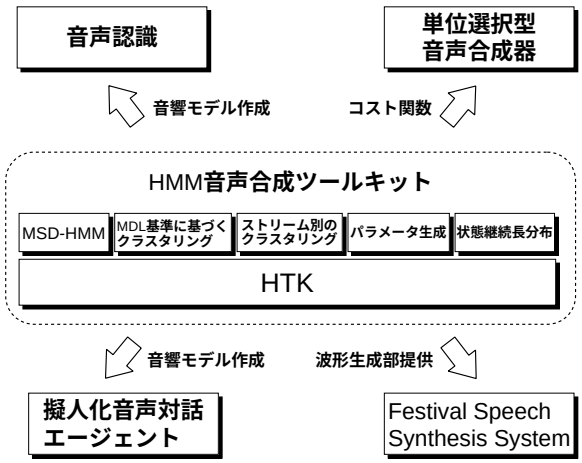

図 2．Relation for other toolkits

謝辞　HMM 音声合成システムは, 東京工業大学小林隆夫教授及び名古屋工業大学北村正教授の研究グループにより開発された. システムの開発に携わった, 全ての方々に感謝いたします.

参考文献

[1] 吉村, 徳田, 益子, 小林, 北村, “HMM に基づく音声合成におけるスペクトル・ピッチ・継続長の同時モデル化”, 信学論, vol.J83-D-II, no.12, pp.2099–2107, 2000.
[2] http://htk.eng.cam.ac.uk/.
[3] 徳田, 益子, 宮崎, 小林 “多空間上の確率分布に基づいた HMM”, 信学論, vol.J83-D-II, pp.1579–1589, 2000.
[4] 益子, 徳田, 宮崎, 小林, “多空間確率分布 HMM によるピッチパタン生成”, 信学論, vol.J83-D-II, no. 7, pp.1600–1609, 2000.
[5] K. Shinoda and T. Watanabe, “MDL-based context-dependent subword modeling for speech recognition,” *J. Acoust. Soc. Jpn.(E)*, vol. 21, no. 2, pp.79–86, 2000.
[6] K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi, T. Kitamura, “Speech parameter generation algorithms for HMM-based speech synthesis,” *Proc. of ICASSP 2000*.
[7] 今井, 住田, 古市, “音声合成のためのメル対数スペクトル近似 (MLSA) フィルタ”, 信学論, J66-A, no.2, pp.122–129, 1983.
[8] T. Yoshimura, K. Tokuda, T. Masuko, T. Kobayashi and T. Kitamura, “Speaker Interpolation in HMM-Based Speech Synthesis System,” Proc. of EUROSPEECH, vol.5, pp.2523–2526, 1997.
[9] 田村, 益子, 徳田, 小林, “HMM に基づく音声合成におけるピッチ・スペクトルの話者適応”, 信学論, vol.J85-D-II, no. 4, pp.545–553, 2002.
[10] 沢辺, 七里, 吉村, 徳田, 益子, 小林, 北村, “固有声に基づく音声合成におけるピッチのモデル化”, 音講演集, 3-2-6, 2001-10.
[11] 嵯峨山, 伊藤, 宇津呂, 甲斐, 小林, 下平, 伝, 徳田, 中村, 西本, 新田, 広瀬, 森島, 峯松, 山下, 山田, 李, “擬人化音声対話エージェント開発プロジェクト”, 音講論集, 1-5-14, 2002-3.
[12] 山下, 峯松, 徳田, 小林, 広瀬, “擬人化音声対話エージェントにおける音声合成”, 音講論集, 1-5-17, 2002-3.
[13] A. W. Black, P. Taylor and R. Caley, “The Festival Speech Synthesis System,” http://www.festvox.org/festival/.
[14] K. Tokuda, H. Zen and A. W. Black, “An HMM-based speech synthesis system applied to English,” *IEEE Speech Synthesis Workshop*, 2002.
[15] Y. Minami, E. McDermott, A. Nakamura and S. Katagiri, “A recognition method with parametric trajectory synthesized using direct relations between static and dynamic feature vector time series,” *Proc. of ICASSP 2002*.
[16] 水谷, 吉村, 徳田, 北村, “HMM に基づいた波形接続型音声合成方式の検討”, 音講論集, 2-10-5, 2002-3.
[17] http://hts.ics.nitech.ac.jp/.