7M-07

複合ウェーブレットモデルを用いたテキスト音声合成の検討*

小口　純矢†　　濱田　康弘†　　嵯峨山　茂樹†

明治大学総合数理学部‡

1. はじめに

テキストから音声へ変換する音声合成技術(Text-to-Speech)の手法として HMM 音声合成が一定の成果を収めている. 動的特徴量を考慮しているため, 接続部分にひずみの少ない滑らかな合成が可能であり, パラメータの変換による多様な声質や発声のスタイルを表現することができる.

従来の HMM 音声合成では, 合成時にケプストラム特徴量と基本周波数から MLSA フィルタ[1]による合成を行っている. しかし, 巡回型フィルタを用いた合成音声は基本周波数成分とスペクトル包絡のピークが重なる場合にはスペクトルのピークが鋭くなり, 一部が不自然に大きく聞こえてしまうといった利得特性の問題が音質の劣化を引き起こしていた.

ここで, 巡回型フィルタを用いないでパワースペクトルから信号波形を得る方法がこの問題を解決する手段として有効であることが先行研究[2][4]では示されている.

本研究では, 複合ウェーブレットモデル(Composite Wavelet Model: CWM)を用いることで, 従来の HMM 音声合成で用いられてきた巡回型フィルタの持つ利得特性が悪化する問題の改善を試みた結果を報告する.

2. フィルタを用いない CWM 音声合成

CWM はスペクトル包絡に対し, 混合ガウスモデル(Gaussian Mixture Model: GMM)による近似を行うことで得られたパラメータをもとに Gabor ウェーブレットの基本波形を生成し, これをピッチ周期ごとに重ね合わせることによって合成を行うため, 合成音の品質はフィルタ方式に比べて多少劣るものの, 巡回型フィルタを用いたことに起因する利得特性の問題は生じないと考えられる.

先行研究[3]ではスペクトル包絡の生成系列と合成の両方に CWM を使用し合成を行っていたが, 本研究ではスペクトル包絡の生成系列には HMM から生成されたケプストラム特徴量を使用し, そこから得られたスペクトル包絡に対して CWM を適用することで音声合成を行う方法を検討した.

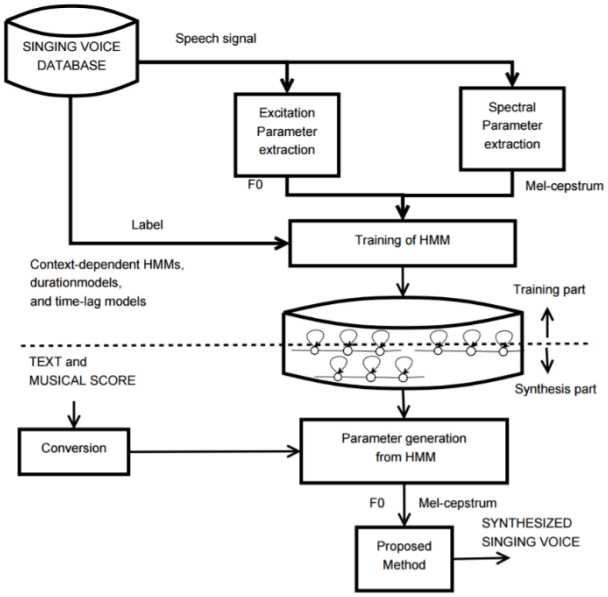

Figure1. Overview of proposed HMM-based speech synthesis system [4]

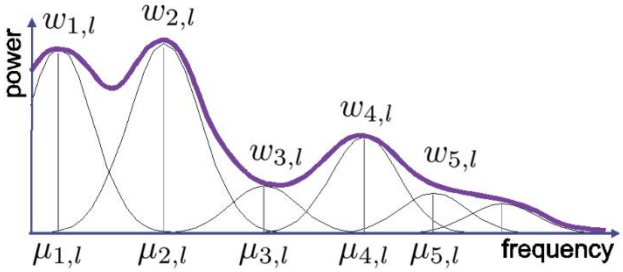

Figure2. Spectrum envelope by CWM [3]

3. HMM と CWM を用いた音声合成の方法

HMM 音声合成システムとして, HTS[2]で生成された一般化メルケプストラムを用いて算出されるスペクトル[1]に対して, CWM を適用[2]することで合成音声を得られる. この原理を用いて以下の方法によって合成を行った.

(1) HTS から得られた一般化メルケプストラム係数をスペクトルに変換する.

* Text-to-Speech Synthesis Using Composite Wavelet Model
† Junya Koguchi, Yasuhiro Hamada, Shigeki Sagayama
‡ Meiji University

(2) GMM によるスペクトルの近似[3]を行い, 各ガウス分布の重み, 平均, 分散を得る.
(3) ここでスペクトル領域における単一ガウス関数のフーリエ変換は Gabor ウェーブレットであるという性質を利用し, スペクトルの GMM 近似から得られた重み, 平均, 分散から Gabor ウェーブレットの重ね合わせることにより基本周期波形を生成する.
(4) これを HTS から得られたピッチ周期ごとに繰り返す.

4. 実験

[5]の実験に倣い, 検討した CWM による音声合成法が有効であるか調べるために合成音声の利得特性を調べた. 比較として MLSA フィルタによる合成音声の特性を調べた.

3.1 実験条件

実験に用いる音声は, ATR データベースより 3-5 秒程度の 5 文章を選択し, HTS により生成された γ =1.0, α =0.55, サンプリング周波数を 16000 *Hz* としたケプストラム特徴量と基本周波数を 0.8 倍から 1.2 倍まで 0.05 刻みで変更したものを用いた. また, GMM による近似は EM アルゴリズムを用い, 混合数は 5 とした.

3.2 利得特性の評価

テキスト音声合成は新たに音声を生成するため, 分析合成[3]のように元となる音声と比較することはできない. そこで, 有声区間において各フレームのパワーの最大値と最小値の差を周波数ごとに調べた.

利得特性を調べた結果を Fig. 3, 4 にヒストグラムで示す. 図は分布が右に偏るほど利得の変化が大きくなることを示している. 図より利得特性は MLSA フィルタを用いた合成手法に比べて CWM を用いた方法がより安定していると考えられる.

5. おわりに

本研究では, 安定感のある音声合成を目指し, ケプストラム特徴量から CWM モデルによる音声合成を行った.

HMM により生成されたケプストラム特徴量をスペクトルに変換し, GMM を適用することでガウス関数の重ね合わせとして近似, その結果から得られたガウス関数の重み, 平均, 分散をもとに求めた Gabor ウェーブレットの重ね合わせをピッチ周期ごとに重ね合わせることで合成音声を生成した. 得られた合成音声の利得特性を調べた結果, 改善を確認した.

今後の課題として, ガウス関数の混合数を増やして品質の向上を図りたい.

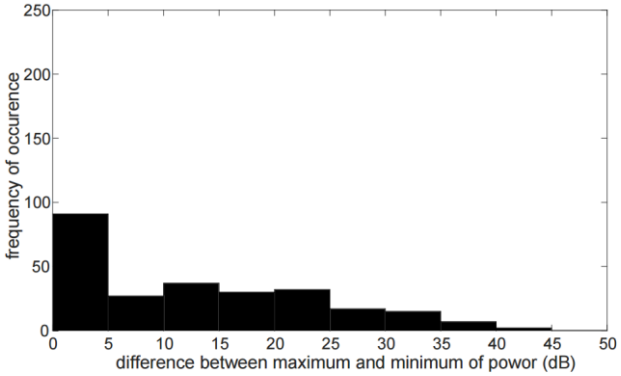

Figure3 Gain characteristics of MLSA filter.

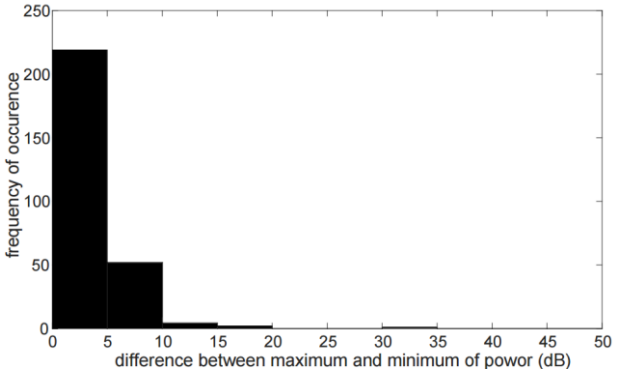

Figure 4. Gain characteristics of CWM.

文献

[1] 徳田, “HMM による音声合成の基礎,”信学論, 74, 2000.
[2] 全 他, “HMM 音声合成システム (HTS) の開発,” 情報処理学会研究報告, SLP-069(129), pp. 301-306, Dec. 2007.
[3] 槐 他, “複合ウェーブレットモデルに基づく音声の分析合成,” 電子情報通信学会技術研究報告, SP-105(370), pp. 1-6, Oct. 2005.
[4] 北条 他, “複合ウェーブレットと HMM の統合モデルに基づくテキスト音声合成,” 情報処理学会研究報告, 2013-MUS-99(21), pp. 1-5, May 2013.
[5] 濱田 他, “無矛盾位相復元を用いたケプストラム特徴量からの音声合成,” 第 78 回情報処理学会全国大会講演論文集, pp. 2-15-16, 2016.