Vol.2017-MUS-115 No.56
2017/6/18

複合ウェーブレットモデルを用いたケプストラム特徴量からの音声合成

小口純矢[†1]　濱田康弘[†1]　嵯峨山茂樹[†1]

概要：本研究では，安定した音声合成を目指し，複合ウェーブレットモデルを用いたケプストラム特徴量からの音声合成法を検討する．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 フィルタを用いた合成を行っていたが，巡回型フィルタを用いた方法ではフィルタの時間特性，ゲイン特性が音質の劣化を招いていた。そこで，本発表では，フィルタを用いない方法として，HMM 音声合成により得られたケプストラム特徴量からパワースペクトルを生成し，複合 Gabor ウェーブレットを基本波形として，これをピッチ周期ごとに重ね合わせることにより音声合成を行った結果を報告する．

キーワード：音声合成，ウェーブレット，ケプストラム，HMM

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[1]が主要な音声合成技術の一つとなっている．本研究では，その波形生成手法について検討した．

HMM 音声合成は，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という利点がある．

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メル対数スペクトル近似(MLSA: Mel-log Spectrum Approximate)フィルタ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引き起こす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[5]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(CWM: Composite Wavelet Model)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引き起こす問題の改善を試みた結果を報告する．

†明治大学 総合数理学部
Meiji University, 4-21-1, Nakano, Tokyo, Japan
1. ev50552@meiji.ac.jp
2. hamada@meiji.ac.jp
3. sagayama@meiji.ac.jp

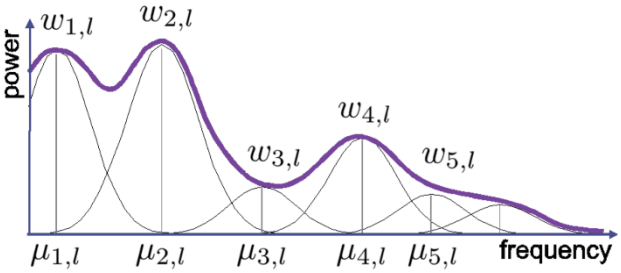

Figure 1. Spectrum envelope by CWM

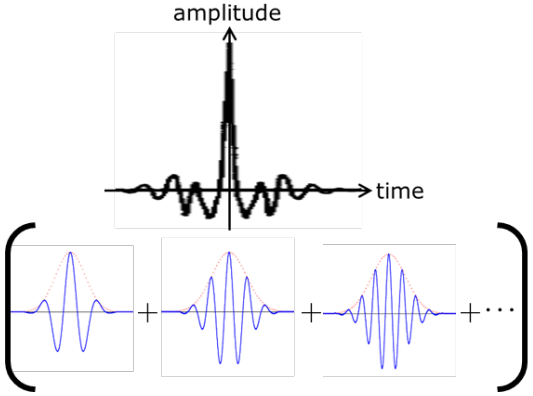

Figure 2. Basic Waveform by CWM

2. CWM 音声波形合成

まず，有声音の合成には，以下に示されるように，ガウス関数の逆フーリエ変換が Gabor 関数，つまりガウス関数と三角関数の積であることに注目する．

$$\mathcal{F}\left[\frac{a}{2\sqrt{b\pi}}\exp\left(-\frac{t^2}{4b}+jct\right)\right]=\exp(-b(\omega-c)^2)$$

ただし，ω を周波数，t を時間，a, b, c は任意の実数である．

周波数領域でのガウス関数の平均は，時間領域においては，掛け合わされる三角関数の周波数に相当する．スペクトル包絡に対し，混合ガウスモデル(GMM: Gaussian Mixture Model)を適

用することでガウス関数の和としてスペクトル包絡を近似(Fig. 1)することができるので, GMM の逆フーリエ変換を Gabor 関数の重ね合わせとして複合ウェーブレットの基本波形(Fig. 2)が得られる.

時間領域において, 得られた基本波形を基本周波数に対応する時間間隔で並べることにより音声波形を合成できる. これは, FIR フィルタを基本周波数に対応したデルタ列で駆動することに相当する. 無声音部分は基本波形を並べる間隔をランダムにし, 周期性を失わせることで実現できる. これはフィルタを雑音源で駆動することに相当し, 無声音部分では波形の非周期性が高くなることを考慮している.

CWM ではスペクトル包絡から直接音声波形を合成するため, 巡回型フィルタの利得特性が引き起こす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 を使用していたが, 本研究では CWM の音声合成手法としての側面に注目し, 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(Fig.3)を, そこから得られたスペクトル包絡に対しては CWM を適用することで音声合成を行う方法を検討した.

3. HMM と CWM を用いたテキスト音声合成

3.1 CWM 音声合成手順

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波形を得られる. この原理を用いて以下の手順によって音声合成を行った.

【CWM 音声合成手順】

(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.
(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) 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.

3.2 ケプストラム特徴量と F_0 の生成

特徴パラメータ学習・生成は典型的な HMM 音声合成システムである HTS に従って行った. 合成部において, HTS では MLSA フィルタを用いているのに対して, 本研究では CWM を用いている. HMM 音声合成システムではメル周波数ケプストラム係数系列(MFCCs: Mel-Frequency Cepstrum Coefficients)が用いられ, 励起パラメータとして F_0 が用いられる. 合成部では与えられたテキスト情報からこれらのパラメータが推定される.

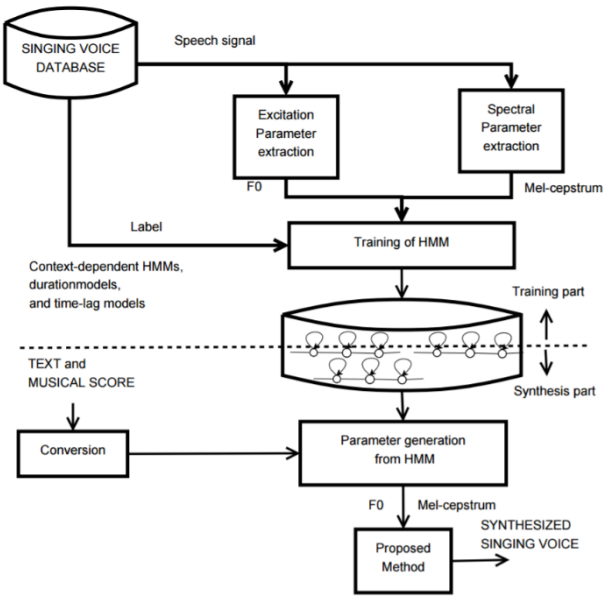

Figure 3. Overview of proposed HMM-based speech synthesis system

3.1 MFCC からスペクトルへの変換

生成された MFCCs は次のようにスペクトルへと変形される.

$$H(\omega) = s_\gamma^{-1}\left(\sum_{m=0}^{M} \tilde{c}_\gamma(m) e^{-j\tilde{\omega}t}\right)$$

$$s_\gamma^{-1}(\omega) = \begin{cases} (1+\gamma\omega)^{1/\gamma}, & 0 < |\gamma| \le 1 \\ \exp\omega, & \gamma = 0 \end{cases}$$

$$\tilde{\omega} = \tan^{-1}\frac{(1-\alpha^2)\sin\omega}{(1+\alpha^2)\cos\omega - 2\alpha}$$

ここで, $H(\omega)$はスペクトル, $s^{-1}{}_\gamma$は一般化対数関数の逆関数, $\tilde{c}_\gamma$はメル周波数ケプストラム係数を表す.

3.2 スペクトルの GMM による近似

Zolfaghari ら[5]は，音声スペクトルのフォルマント分析のために包絡を GMM で近似する手法を提案した．その手法に倣い，EM アルゴリズムに基づくスペクトルの GMM 近似によりスペクトルパラメータを抽出した．

3.2 CWM による音声波形合成

CWM 特徴量と基本周波数情報を用いて音声波形を合成する手法については，[4] の手法と同様である．

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 刻みで変化させたものを用いた．また，GMM による近似は EM アルゴリズムを用い，混合数は 10 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，基本周波数の変化により生じる利得の変化を各フレームの利得の最大値と最小値の差を用いて調べた．

利得特性を調べた結果を Fig. 4, 5 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ことが示唆される．

5. 終わりに

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

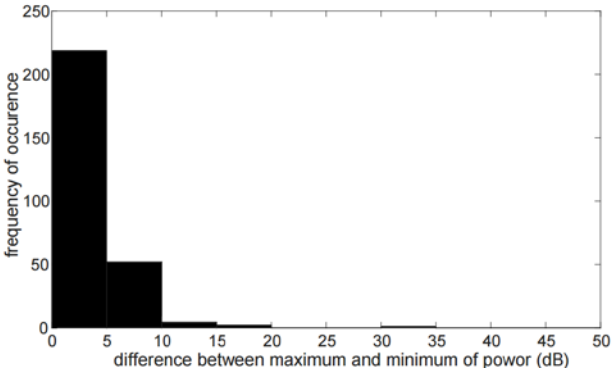

Figure 4. Gain characteristics of MLSA filter.

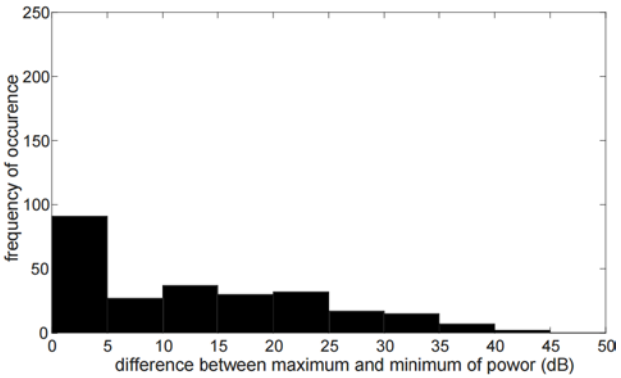

Figure 5. Gain characteristics of CWM.

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複合 Gabor ウェーブレットをピッチ周期ごとに並べて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すなどを検討し，品質の向上を図りたい．

謝辞

本研究は，日本学術振興会科研費基盤研究 (A) 課題番号 26240025 および 17H00749 の部分的支援を受けて行われた。

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.
[6] Zolfaghari 他, “Formant Analysis Using Mixture of Gaussians,” ICSLP 96, vol. 2, pp.1229−1232, 1996.