

論文 / 著書情報
Article / Book Information

論題(和文)	HMMによる規則音声合成の検討
Title(English)	
著者(和文)	立和 航, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1999年春季講演論文集, , No. 2-3-7, pp. 239-240
Citation(English)	, , No. 2-3-7, pp. 239-240
発行日 / Pub. date	1999, 3

◎立和 航 古井 貞熙(東工大)

1. はじめに

音声合成においては、音素や音節といった素片を接続して文を作るため、素片の間に生じる不連続性が品質に大きく影響を及ぼす。より自然な合成音声を得るためには、品質を評価する音響特徴の中に、前後のフレームにおける音響ベクトルとの関係を表す動的特徴量を含めることが必要である。

連続混合分布型 HMM で表現された音素モデルと文の長さから、動的特徴量を考慮しつつ文全体の尤度を最大にするパスおよび音響パラメータ列を生成する手法については、有効な手段が提案されている[1]。この方法では、文中の各音素の持続時間は音素モデルから尤度最大化規準によるパスの探索によって自動的に決定されるが、各音素の持続時間は、例えば句境界直前の音素は長く発声されるなど、文脈や発声環境によって様々に変化し、フレーム単位の遷移モデルでは特徴を的確に捉えることが困難である[2]。より自然な合成音声を得るには、韻律規則によって時間制御することが必要である。

そこで本研究では、韻律モデルから各音素の持続時間が与えられた場合に、HMM で表現された音素モデルから効果的に尤度の高い音響パラメータ列を抽出するアルゴリズムを提案する。

2. 生成アルゴリズム

音響パラメータ $o_t (t=1, 2, \dots, T)$ を、ケプストラム c_t 及びその動的特徴量であるデルタケプストラム Δc_t を用いて $o_t = (c_t, \Delta c_t)'$ と定義する。デルタケプストラムは次式

$$\Delta c_t = \frac{\sum_{l=1}^L l(c_{t+l} - c_{t-l})}{2 \sum_{l=1}^L l^2} \quad (1)$$

によって定義される回帰係数であり、それぞれの時刻における音響パラメータは、その前後 L フレームの音響パラメータの変化の影響を受ける。

我々の目標は HMM λ と各音素の継続長データからこのような動的特徴量を考慮した上で、尤もらしいケプストラムの系列を生成することである。ここでいう尤もらしい系列とは、次に示す式(2)を最大にするような o である。HMM λ からパラメータ系列 o が観測される確率は、一般に起こり得る全ての遷移パス q についての和

$$P(o|\lambda) = \sum_{\text{all } q} P(o|q, \lambda) P(q|\lambda) \quad (2)$$

で表される。このような音響パラメータを求めるには、全て、あるいは一定の範囲の遷移パスを探索しなければならない。ここで、文中の各音素の継続フレーム数が入力として与えられた場合について考える。通常の分析では、音素を HMM でモデル化する場合の状態数は 1~5 状態であり、無音を除く各音素

の継続フレーム数は長いもので 10 数フレームである。この中で音素の継続フレーム数が拘束されれば、遷移パスに関する選択の自由度は非常に小さいと考えられる。そこで、音素の継続フレーム数を遷移確率の逆数で比例配分したものを、その音素に含まれる各状態の継続長として近似する。この近似によって時刻 t における状態 s が一意に定まる。このときの遷移パスを $\hat{q}$ とする。

一つに定まった遷移パスのもとでは式(2)は単純に

$$P(o|\lambda) = P(o|\hat{q}, \lambda) \quad (3)$$

$$= \prod_{t=1}^T b_{q_t}(o_t) \quad (4)$$

と書ける。ここで $b_{q_t}(o_t)$ は、遷移パス $\hat{q}$ 上の第 t フレームにおける状態 q_t においてベクトル o_t が観測される確率である。連続混合分布型 HMM では、この確率は各特徴量の分布の平均値ベクトル μ と共分散行列 Σ を持つ多次元正規分布 N の重み付き和で定義される。

ここで、音響パラメータ列の第 t フレームのケプストラム c_t の値を変化させて $\tilde{c}_t$ としたときに、文全体の尤度がどう変化するかを考えると、先に述べたように、第 t フレームの変化は周囲 L フレームのパラメータ列 $o_{t-L}, o_{t-L+1}, \dots, o_{t+L}$ のデルタ係数の部分に影響を与える。このときの文全体の尤度変化は、式(4)より、

$$P(\tilde{o}|\hat{q}, \lambda) - P(o|\hat{q}, \lambda) \quad (5)$$

$$= \sum_{l=-L}^L b_{q_{t+l}}(\tilde{o}_{t+l}) - \sum_{l=-L}^L b_{q_{t+l}}(o_{t+l}) \quad (6)$$

となり、尤度の差分を第 t フレーム近傍の分布係数からのみによって求めることができる。 L の値としては 2 程度の値がよく用いられ、この計算量は一般にわずかである。したがって、それぞれのフレームのケプストラムについて、更新前と更新後の尤度を比較し、更新後の尤度が高ければ新しい値を採用するという方式で、勾配法により常に文全体の尤度が大きくなる方向へパラメータ列を更新し続けることが可能である。

3. 実験

音素モデルとして、ATR 日本語音声 DB の音素パランス文 503 文(話者 MHT) に対して、サンプリング周波数 20kHz、フレーム長 25.6ms、フレーム周期 12.8ms でメルケプストラム分析[3]を行い、得られた 24 次のメルケプストラム、及びそのデルタ係数を音響パラメータとして総状態数 1528、3 ループ 5 状態 4 混合 tied triphone HMM モデルを作成した。

この音素モデルと、自然音声から抽出した韻律情報によるパルスと雑音を音源として音声合成した。ATR 音声 DB の話者 MHT の会話音声に含まれる

* A study of speech synthesis using HMMs.

By Wataru Tachiwa and Sadaoki Furui (Tokyo Institute of Technology)

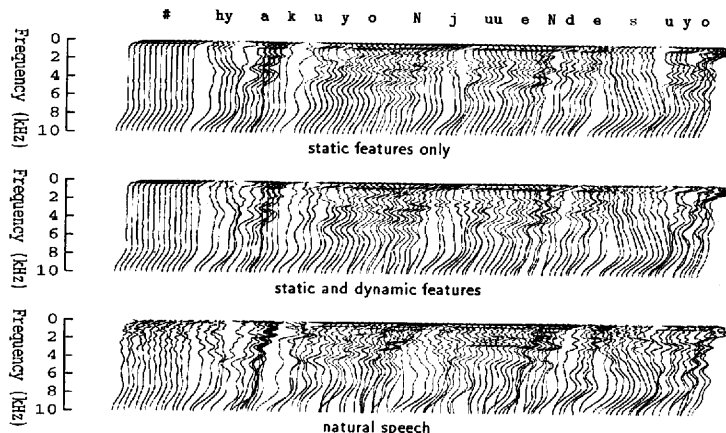


図 1. 合成音声のスペクトル変化

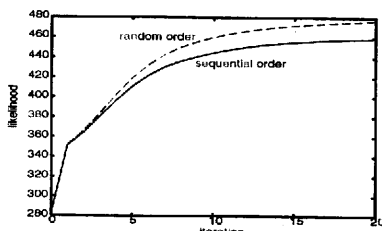


図 2. 勾配法による尤度の変化

会話文(2-7) {# hy a k u y o N j u e N d e s u y o} の音声合成したときのスペクトル変化の様子を図 1 に示す。

勾配法によるパラメータ修正によって、主に有声音の接続が滑らかになる一方で、急激に変化すべき箇所はそのまま残され明瞭性が確保されている。

図 2 に、このときの尤度の変化の模様を示す。ここでは、各フレームのケプストラムのそれぞれ全ての要素に対して、飽和するまで更新を繰り返すことを全フレームについて行う計算を 20 回繰り返した、回数を重ねる毎に、文全体の尤度が大きくなっている様子がわかる。

また、各回において先頭フレームから順番に更新して行くのではなく、ランダムな順序で更新した方がより早く収束に向かうことがわかった。図 2 において点線で示される曲線は、ランダムな順序でフレームを更新した場合の尤度の変化を表している。これは、先頭フレームから更新すると、尤度を飽和させたばかりのフレームの音響パラメータが、すぐ次のフレームの更新に影響されて変化するためであると考えられる。

また、この音声を生成する際には第 3 音素/k/の直前で文章を二つにわけて別々にパラメータを求めた上で接続した。音素/k/は無声閉鎖音に属し、音の直前に常に無音区間が出現するためである。無音区間を挟むと通常、その前後のスペクトルは接続しないため、独立した 2 つの文章として扱うことができる。同様に直前に無音区間が発生する音素の一覧を表 1 に示す。

これらの音素で文を短く分けることにより、一文

表 1. 無音区間が発生する音素

無声閉鎖音	/p/, /t/, /k/
無声破擦音	/ch/, /ts/
拗音	/py/, /ky/
促音閉鎖音	/q/

あたりのフレーム数を削減して、バスや音響パラメータ列の探索作業を高速化することができる。主観評価による受聴では、文を上記の子音で分けた場合と分けない場合で、品質には大きな差異は見られなかった。今回実験で用いた音素セット(母音 12, 子音 30)で、ATR 日本語音声 DB の音素バランス文 503 文について一文あたりの平均音素数を測定した結果では、54.89 音素/文に対し、無声区間を含む子音で分けた場合は 5.69 音素/文となった。

4. 結果

固定パスのもとで尤度が局所的に最大となる音響パラメータ列を勾配法によって探索するアルゴリズムを提案し、なめらかな合成音生成できることを確認した。

謝辞

HMM を利用した音声合成法に関して御助言を頂いた東工大精密工学研究所の益子貴史氏に感謝致します。音素モデル作成にあたって御指導を頂いた NTT ヒューマンインタフェース研究所の大附克年氏に感謝致します。

参考文献

- [1] 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, Mar.1997.
- [2] 広瀬 啓吉 “韻律情報の処理”, Journal of Signal Processing, Vol.2, No.6, pp.415-423, Nov.1998.
- [3] 徳田 恵一, 小林 隆夫, 深田 俊明, 齊藤 博徳, 今井 聖, “メルケプストラムをパラメータとする音声のスペクトル推定”, 信学論 (A), J74-A. 8, pp.1240-1248, Aug.1991.