社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

ベイズ的アプローチに基づくHMM音声合成

南角　吉彦† 　全　 炳河† 　徳田　恵一† 　北村　 正† 　益子　貴史††

† 名古屋工業大学知能情報システム学科
† 東京工業大学大学院総合理工学研究科

あらまし　本論文では，HMM 音声合成に変分ベイズ法を適用し，ベイズ基準による音声のパラメータ生成アルゴリズムを導出する．HMM 音声合成では，これまで，HMM の学習と音声パラメータの生成に ML 基準が用いられてきた．しかし，ML 推定はモデルパラメータを確定的変数として点推定するため，学習データが十分に得られない場合，モデルの推定精度が低下するという問題がある．提案法は，ML 基準に比べて，音声特徴をモデル化する統計モデルの汎化能力を高めることが可能であり，合成音声の品質改善が期待できる．

キーワード　変分ベイズ法，隠れマルコフモデル，音声合成

A Bayesian Approach to HMM-Based Speech Synthesis

Yoshihiko NANKAKU†, Heiga ZEN†, Keiichi TOKUDA†, Tadashi KITAMURA†,
and Takashi MASUKO††

† Department of Computer Science, Nagoya Inst. of Tech. Gokiso-cho, Shouwa-ku, Nagoya, 466-8555 Japan
† Interdisciplinary Graduate School of Science and Engineering, Tokyo Inst. of Tech. 4259, Nagatsuta, Midori-ku, Yokohama, 226-8502 Japan

Abstract　In this paper, we applied the variational Bayes method to HMM-based speech synthesis and derived a speech parameter generation algorithm based on the Bayesian criterion. In HMM-based speech synthesis, the ML criterion has been used for training HMM and generating speech parameters. However, the ML estimator produces a point estimate of HMM parameters and the accuracy of estimation may be reduced when little training data is available. By using the proposed algorithm, it is expected that the higher generalization ability of the statistical model for speech features is achieved and leads to improving the quality of synthesized speech.

Key words　variational Bayes method, hidden Markov models, speech synthesis

1. まえがき

コーパスに基づく音声合成では，データを多く集めることにより合成音声の品質を高めることができるが，その一方で限られたデータ量で品質の良い音声合成を実現することが重要な課題である．HMM 音声合成は，メルケプストラムや基本周波数などの音声特徴量を用いた分析合成系により音声が合成されるため，波形接続型の音声合成に比べて，少ない学習データで滑らかな音声が合成可能である．また，音声特徴の動的特徴量を考慮したパラメータ生成アルゴリズムにより音声単位境界において歪みの少ない音声パラメータを得ることができる．HMM 音声合成における音声品質の上限は，ポストフィルタの使用などの方法もあるが，基本的には音声特徴量を用いた分析合成系に依存している．しかし，学習データが十分に得られない場合，統計モデルの精度が低下する可能性があり，統計モデルの汎化能力を高めることにより，合成音声の品質を改善する余地があると考えられる．また，この点において，統計的学習の様々な研究成果を応用できることが，HMM 音声合成の大きな利点であると言える．

HMM 音声合成では，これまで，HMM の学習や音声パラメータの生成に ML(Maximum Likelihood) 基準が用いられてきた．しかし，ML 推定は，モデルパラメータを確定的変数として点推定するため，学習データが十分に得られない場合，モデルの推定精度が低下するという問題がある．これに対し，ベイズ学習では，モデルパラメータを確率変数としてとらえ，学習データに対する事後確率分布を推定する．事後分布を用いて，すべての可能なモデルパラメータを考慮することにより，学習データ量が少ない場合においても，高い汎化性能が得られる．しかし，事後確率分布の推定は困難な積分計算を伴うため，何らかの近似を必要とする．従来，マルコフ連鎖モンテカルロ法などのサンプリング手法が用いられてきたが，計算量が膨大となり，大規模な問題への適用は困難であった．この問題に対し，近年，変分ベイズ法 [1], [2] が提案され，ML 基準により学習されていた様々なモデルにベイズ学習が適用可能となり，

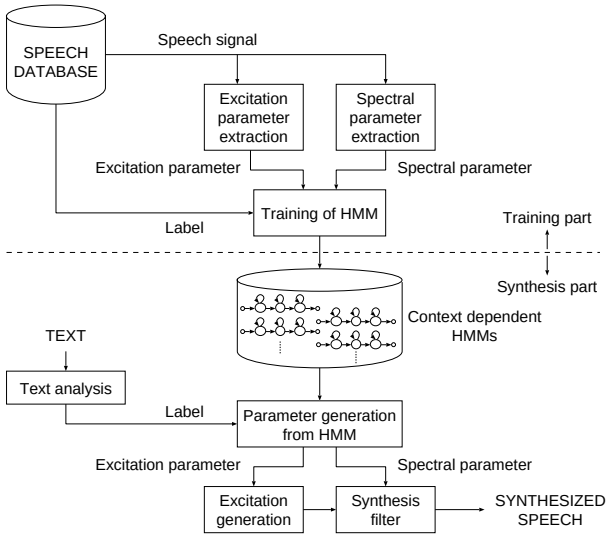

図 1　HMM に基づいた音声合成システムの構成

HMM に基づいた音声認識にも適用されている [3]．本研究では，HMM 音声合成に変分ベイズ法を適用し，ベイズ基準による音声パラメータの生成アルゴリズムを導出する．また，コンテキストクラスタリングにおいても，ベイズ基準を適用することにより，データ量に応じた最適なモデル構造を決定する．

2. ML 基準による HMM 音声合成

まず，ML 基準による音声合成システムについて説明する．図 1 にシステムの概要を示す [4]．学習部では，音声データベースからスペクトルパラメータと励振源パラメータを抽出し，音素などの音声単位で HMM を学習する．本論文では，スペクトルパラメータとしてメルケプストラム，励振源パラメータとして基本周波数を用いることとする．合成部では，合成すべきテキストに対応するラベル列に従い，音素 HMM を連結した文章 HMM を作成し，この HMM に対する尤度が最大となる音声パラメータを生成する．基本周波数に従って励振源モデルから励振源を生成し，生成したメルケプストラムから構築された合成フィルタ（MLSA フィルタ）を励振することにより合成音声が得られる．

2.1　隠れマルコフモデルの定義

音声合成に用いる隠れマルコフモデルを定義する．音声特徴量の観測ベクトル列 $\boldsymbol{O} = (\boldsymbol{o}_1, \boldsymbol{o}_2, \ldots, \boldsymbol{o}_T)$ と状態遷移を表す隠れ変数列 $\boldsymbol{Z} = (z_1, z_2, \ldots, z_T)$, $z_t \in \{1, \ldots, N\}$ から成る完全データ $\{\boldsymbol{O}, \boldsymbol{Z}\}$ の対数尤度は次式で定義される．

$$\begin{aligned}\log P(\boldsymbol{O}, \boldsymbol{Z}|\boldsymbol{\Lambda}) \;=\; & \sum_{t=1}^{T-1}\sum_{i=1}^{N}\sum_{j=1}^{N} z_t^i z_{t+1}^j \log a_{ij} \\ & + \sum_{t=1}^{T}\sum_{i=1}^{N} z_t^i \log \mathcal{N}(\boldsymbol{o}_t|\boldsymbol{\mu}_i, \boldsymbol{S}_i^{-1}) \qquad (1)\end{aligned}$$

ただし，$z_t^i = \delta(z_t, i)$ であり，$z_t = i$ のとき 1, それ以外は 0 とする．N は HMM の状態数，HMM のパラメータ $\boldsymbol{\Lambda} = \{\boldsymbol{a}, \boldsymbol{b}\}$ は，状態遷移確率 $\boldsymbol{a} = \{a_{ij}\}_{i,j=1}^N$, $a_{ij} = P(z_{t+1} = j|z_t = i)$ と出力確率 $\boldsymbol{b} = \{b_i(\boldsymbol{o}_t)\}_{i=1}^N$, $b_i(\boldsymbol{o}_t) = P(\boldsymbol{o}_t|z_t = i) = \mathcal{N}(\boldsymbol{o}_t|\boldsymbol{\mu}_i, \boldsymbol{S}_i^{-1})$ により与えられる．ただし，$\mathcal{N}(\cdot|\boldsymbol{\mu}_i, \boldsymbol{S}_i^{-1})$ は，$\boldsymbol{\mu}_i$ と $\boldsymbol{S}_i$ を，それぞれ，平均ベクトルと共分散行列の逆行列として持つガウス分布である．出力確率分布としては，多次元ガウス分布の重み付き和で表される多次元ガウス混合分布が用いられることが多いが，ここでは，簡単のため単一の多次元ガウス分布を仮定する．

2.2　ML 基準によるパラメータ生成

ML 基準による音声パラメータの生成では，動的特徴量を考慮することにより，状態遷移やモデル接続部分で歪みの少ない滑らかな音声パラメータを生成することができる [5], [6]．以下にそのアルゴリズムを説明する．まず，学習部では与えられた学習データ $\boldsymbol{O}$ の尤度を最大にするモデルパラメータ $\boldsymbol{\Lambda}^{(ML)}$ を推定する．学習データ $\boldsymbol{O}$ は，音声認識などで良く用いられるように，音声パラメータの静的特徴量に動的特徴量を連結したベクトルを特徴ベクトルとして用いる．

$$\boldsymbol{\Lambda}^{(ML)} \;=\; \arg\max_{\boldsymbol{\Lambda}} P(\boldsymbol{O}|S, \boldsymbol{\Lambda}) \qquad (2)$$

ここで，S は学習データのラベル列である．合成部では，推定されたモデルから尤度最大となる音声パラメータ $\boldsymbol{x} = \left[\boldsymbol{x}_1^\top, \boldsymbol{x}_2^\top, \ldots, \boldsymbol{x}_{T_x}^\top\right]^\top$ を生成する．

$$\boldsymbol{x}^{(ML)} \;=\; \arg\max_{\boldsymbol{x}} P(\boldsymbol{W}\boldsymbol{x}|s, \boldsymbol{\Lambda}^{(ML)}) \qquad (3)$$

ここで，s は合成音声のラベル列である．また，$\boldsymbol{W}$ は学習データと同様に，音声の静的特徴量 $\boldsymbol{x}$ に動的特徴量を加える行列であり，$\boldsymbol{W}\boldsymbol{x}$ の時刻 t における要素は，次式で表される．

$$(\boldsymbol{W}\boldsymbol{x})_t \;=\; [\boldsymbol{x}_t^\top, \Delta\boldsymbol{x}_t^\top, \Delta^2\boldsymbol{x}_t^\top]^\top \qquad (4)$$

また，動的な特徴ベクトル $\Delta\boldsymbol{x}_t$，$\Delta^2\boldsymbol{x}_t$ は，静的特徴ベクトル $\boldsymbol{x}_t$ から

$$\Delta\boldsymbol{x}_t \;=\; \sum_{\tau=-L^{(1)}}^{L^{(1)}} w_1(\tau)\boldsymbol{x}_{t+\tau} \qquad (5)$$

$$\Delta^2\boldsymbol{x}_t \;=\; \sum_{\tau=-L^{(2)}}^{L^{(2)}} w_2(\tau)\boldsymbol{x}_{t+\tau} \qquad (6)$$

により計算されるものとする．ただし，$w_1(\tau)$，$w_2(\tau)$ は動的特徴量を計算するための重み係数である．$\boldsymbol{x}_t$ が D 次元とすれば，$\boldsymbol{x}$, $\boldsymbol{W}\boldsymbol{x}$ は，それぞれ，TD 次元，$3TD$ 次元である．$\boldsymbol{W}$ は，$3TD \times TD$ の行列であり，1 部の要素に係数 $1, w_1(\tau), w_2(\tau)$ をもち，他の多くの要素は 0 となる．

式 (3) において，$P(\boldsymbol{W}\boldsymbol{x}|s, \boldsymbol{\Lambda}^{(ML)})$ は隠れ変数を含むため，すべての隠れ変数列を考慮する必要があるが，ここでは簡単のため，あらかじめ最適な状態系列が与えられると仮定する(注 1)．

$$\begin{aligned}\boldsymbol{x}^{(ML)} \;&=\; \arg\max_{\boldsymbol{x}} \sum_{\boldsymbol{q}} P(\boldsymbol{W}\boldsymbol{x}|\boldsymbol{q}, \boldsymbol{\Lambda}^{(ML)}) P(\boldsymbol{q}|s, \boldsymbol{\Lambda}^{(ML)}) \\ &\simeq\; \arg\max_{\boldsymbol{x}} P(\boldsymbol{W}\boldsymbol{x}|\boldsymbol{q}', \boldsymbol{\Lambda}^{(ML)}) \qquad (7)\end{aligned}$$

(注 1)：式 (3) を EM アルゴリズムのような繰返しアルゴリズムで解く手法 [7] が提案されている．また，隠れ変数である状態系列を確率モデルで最適化する場合，ここで定義した状態遷移確率では不十分であり，状態継続長モデルなどの学習が必要となる．

ここで，$\boldsymbol{q} = (q_1, q_2, \ldots, q_{T_x})$ は $\boldsymbol{x}$ の状態系列，$\boldsymbol{q}'$ は与えられた最適な状態系列を表す．状態系列 $\boldsymbol{q}'$ が与えられたときの $\boldsymbol{x}$ の尤度 $P(\boldsymbol{W}\boldsymbol{x}|\boldsymbol{q}', \boldsymbol{\Lambda}^{(ML)})$ は，高次元のガウス分布と見なすことができる．

$$\begin{aligned}\log P(\boldsymbol{W}\boldsymbol{x}|\boldsymbol{q}', \boldsymbol{\Lambda}) &= -\frac{DT}{2}\log(2\pi) \\ &+\frac{1}{2}\log|\boldsymbol{S}| - \frac{1}{2}(\boldsymbol{W}\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{S}(\boldsymbol{W}\boldsymbol{x}-\boldsymbol{\mu}) \quad (8)\end{aligned}$$

ただし，$\boldsymbol{\mu}$，$\boldsymbol{S}$ は，状態系列 $\boldsymbol{q}'$ に従って並べたガウス分布の平均ベクトルと共分散行列を一つにまとめたものである．

$$\boldsymbol{\mu} = [\boldsymbol{\mu}_{z_1}^\top, \boldsymbol{\mu}_{z_2}^\top, \ldots, \boldsymbol{\mu}_{z_T}^\top]^\top \quad (9)$$

$$\boldsymbol{S} = \mathrm{diag}[\boldsymbol{S}_{z_1}, \boldsymbol{S}_{z_2}, \ldots, \boldsymbol{S}_{z_T}] \quad (10)$$

もし，動的特徴量を用いない場合，尤度を最大にする音声パラメータは平均ベクトル列と等しくなり，滑らかな音声パラメータは得られない．しかし，動的特徴量を考慮することにより，$P(\boldsymbol{W}\boldsymbol{x}|\boldsymbol{q}', \boldsymbol{\Lambda})$ を最大にする $\boldsymbol{x}$ は，

$$\frac{\partial \log P(\boldsymbol{W}\boldsymbol{x}|\boldsymbol{q}', \boldsymbol{\Lambda})}{\partial \boldsymbol{x}} = \boldsymbol{0} \quad (11)$$

とおくことによって得られる線形方程式

$$\boldsymbol{W}^\top \boldsymbol{S}\boldsymbol{W}\boldsymbol{x}^{(ML)} = \boldsymbol{W}^\top \boldsymbol{S}\boldsymbol{\mu} \quad (12)$$

により定められる．$\boldsymbol{W}^\top\boldsymbol{S}\boldsymbol{W}$ は，$TD \times TD$ の巨大な行列であるが，コレスキー分解あるいは QR 分解を用いて少ない演算量で解くことができる [8], [9]．また，時間方向に再帰的な形で計算を行うアルゴリズム [10] も提案されている．

3. ベイズ基準による HMM 音声合成

3.1 変分法による予測分布の近似

ML 基準による音声合成では，モデルパラメータを点推定するのに対し，ベイズ音声合成では，学習データの事後確率分布を計算し，すべてのモデルパラメータを重み付けした予測分布から音声パラメータが生成される．ベイズ基準による合成音声は次式で得られる．

$$\begin{aligned}\boldsymbol{x}^{(Bayes)} &= \arg\max_{\boldsymbol{x}} P(\boldsymbol{W}\boldsymbol{x}|s, \boldsymbol{O}, S) \\ &= \arg\max_{\boldsymbol{x}} P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}|s, S) \quad (13)\end{aligned}$$

ここで，$P(\boldsymbol{W}\boldsymbol{x}|s, \boldsymbol{O}, S)$ は合成音声の予測分布である．また，動的特徴量を含む合成パラメータ $\boldsymbol{W}\boldsymbol{x}$ と学習データ $\boldsymbol{O}$ の周辺尤度は，

$$\begin{aligned}&P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}|s, S) \\ &= \sum_{\boldsymbol{q}}\sum_{\boldsymbol{Z}}\int P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{q}, \boldsymbol{O}, \boldsymbol{Z}, \boldsymbol{\Lambda}|s, S)d\boldsymbol{\Lambda} \\ &= \sum_{\boldsymbol{q}}\sum_{\boldsymbol{Z}}\int P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{q}|s, \boldsymbol{\Lambda})P(\boldsymbol{O}, \boldsymbol{Z}|S, \boldsymbol{\Lambda})P(\boldsymbol{\Lambda})d\boldsymbol{\Lambda} \quad (14)\end{aligned}$$

と書くことができる．ここで，$P(\boldsymbol{\Lambda})$ は事前確率であり，あらかじめ得られる学習対象に関する知識を与えることができる．また，$P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{q}|s, \boldsymbol{\Lambda})$ と $P(\boldsymbol{O}, \boldsymbol{Z}|S, \boldsymbol{\Lambda})$ は，それぞれ，合成データと学習データの完全な尤度を表す．対数周辺尤度 $\log P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}|s, S)$ に対し，任意の分布 $Q(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda})$ を導入し，Jensen の不等式を用いて，下限 $\mathcal{F}$ を定義する．

$$\begin{aligned}&\log P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}|s, S) \\ &= \log\sum_{\boldsymbol{q}}\sum_{\boldsymbol{Z}}\int P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}, \boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda}|s, S)d\boldsymbol{\Lambda} \\ &= \sum_{\boldsymbol{q}}\sum_{\boldsymbol{Z}}\int \log Q(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda})\frac{P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}, \boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda}|s, S)}{Q(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda})}d\boldsymbol{\Lambda} \\ &\geqq \sum_{\boldsymbol{q}}\sum_{\boldsymbol{Z}}\int Q(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda})\log\frac{P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}, \boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda}|s, S)}{Q(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda})}d\boldsymbol{\Lambda} \\ &= \left\langle \log\frac{P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}, \boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda}|s, S)}{Q(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda})} \right\rangle_{Q(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda})} \\ &= \mathcal{F} \quad (15)\end{aligned}$$

ここで，下限 $\mathcal{F}$ は，$Q(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda})$ を変関数とする汎関数であり，不等式における $\log P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}|s, S)$ と $\mathcal{F}$ の差は，$Q(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda})$ と $P(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda}|\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}, s, S)$ の KL-divergence で表される．

$$\begin{aligned}\mathcal{F} &= \left\langle \log\frac{P(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda}|\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}, s, S)P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}|s, S)}{Q(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda})} \right\rangle \\ &= \log P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}|s, S) \\ &\quad -KL(Q(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda})||P(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda}|\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}, s, S)) \quad (16)\end{aligned}$$

ここで，$P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}|s, S)$ は $\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda}$ に関して定数であるため，$Q(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda})$ に関して $\mathcal{F}$ を最大化することは KL-divergence を最小にすることと等価である．また，$Q(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda}) = P(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda}|\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}, s, S)$ のとき，KL-divergence は 0 となり，$\log P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}|s, S) = \mathcal{F}$ となる．よって，変関数 $Q(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda})$ に，積分計算が可能となるような拘束条件を与えた上で，下限 $\mathcal{F}$ を最大化することにより，最適な $P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}|s, S)$ の近似分布を求めることができる．ここでは，以下の拘束条件を与える．

$$Q(\boldsymbol{q}, \boldsymbol{Z}, \boldsymbol{\Lambda}) = Q(\boldsymbol{q})Q(\boldsymbol{Z})Q(\boldsymbol{\Lambda}) \quad (17)$$

このとき，下限 $\mathcal{F}$ は次式で表される．

$$\begin{aligned}\mathcal{F} &= \langle \log P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{q}|s, \boldsymbol{\Lambda})\rangle_{Q(\boldsymbol{q})Q(\boldsymbol{\Lambda})} \\ &\quad + \langle \log P(\boldsymbol{O}, \boldsymbol{Z}|S, \boldsymbol{\Lambda})\rangle_{Q(\boldsymbol{Z})Q(\boldsymbol{\Lambda})} + \log P(\boldsymbol{\Lambda}) \\ &\quad + E[Q(\boldsymbol{q})] + E[Q(\boldsymbol{Z})] + E[Q(\boldsymbol{\Lambda})] \quad (18)\end{aligned}$$

ただし，$E[\cdot]$ はエントロピーを表す．まず，$Q(\boldsymbol{Z})$ について，$\mathcal{F}$ を最大化する．変関数 $Q(\boldsymbol{Z})$ は，式 (15) の不等式が成り立つため，$\sum_{\boldsymbol{Z}} Q(\boldsymbol{Z}) = 1$ を満たす必要があり，ラグランジュ係数 $\lambda_{\boldsymbol{Z}}$ を用いて，次式を最大化する．

$$\begin{aligned}\mathcal{F}_{\boldsymbol{Z}} &= \mathcal{F} - \lambda_{\boldsymbol{Z}}\left(\sum_{\boldsymbol{Z}} Q(\boldsymbol{Z}) - 1\right) \\ &= \sum_{\boldsymbol{Z}}\left(f_{\boldsymbol{Z}}(\boldsymbol{Z}, Q(\boldsymbol{Z})) - \lambda_{\boldsymbol{Z}}Q(\boldsymbol{Z})\right) + Const \quad (19)\end{aligned}$$

ただし，

$$\begin{aligned}&f_{\boldsymbol{Z}}(\boldsymbol{Z}, Q(\boldsymbol{Z})) \\ &= Q(\boldsymbol{Z})\langle \log P(\boldsymbol{O}, \boldsymbol{Z}|S, \boldsymbol{\Lambda})\rangle_{Q(\boldsymbol{\Lambda})} - Q(\boldsymbol{Z})\log Q(\boldsymbol{Z}) \quad (20)\end{aligned}$$

ここで，$\mathcal{F}_{\boldsymbol{Z}}$ を最大にする関数 $Q(\boldsymbol{Z})$ を得るため，変分法を適用する．変分法では，次式で表される変分導関数 $\delta\mathcal{F}_{\boldsymbol{Z}}$ を 0 と置くことにより，関数の最適化問題を解く．

$$\begin{aligned}\delta\mathcal{F}_{\boldsymbol{Z}} &= \sum_{\boldsymbol{Z}} \frac{\partial}{\partial Q(\boldsymbol{Z})}\Big\{ f_{\boldsymbol{Z}}(\boldsymbol{Z}, Q(\boldsymbol{Z})) - \lambda_{\boldsymbol{Z}} Q(\boldsymbol{Z}) \Big\}\delta Q(\boldsymbol{Z}) \\ &= 0 \end{aligned} \tag{21}$$

ただし，$\delta Q(\boldsymbol{Z})$ は $Q(\boldsymbol{Z})$ の変分であり，$Q(\boldsymbol{Z})$ の微細なずれを表す任意の関数である．ここで，$\mathcal{F}_{\boldsymbol{Z}}$ は，$Q(\boldsymbol{Z})$ の微分項を含まない単純な形であるため，すべての $\boldsymbol{Z}$ において，次式が成り立てば良い．

$$\frac{\partial}{\partial Q(\boldsymbol{Z})}\Big\{ f_{\boldsymbol{Z}}(\boldsymbol{Z}, Q(\boldsymbol{Z})) - \lambda_{\boldsymbol{Z}} Q(\boldsymbol{Z}) \Big\} = 0 \tag{22}$$

よって，最適な近似分布は，

$$Q(\boldsymbol{Z}) = C_{\boldsymbol{Z}} \exp \langle \log P(\boldsymbol{O}, \boldsymbol{Z}|S, \boldsymbol{\Lambda}) \rangle_{Q(\boldsymbol{\Lambda})} \tag{23}$$

となる．ただし，$C_{\boldsymbol{Z}}$ は $\sum_{\boldsymbol{Z}} Q(\boldsymbol{Z}) = 1$ を満たすための正規化定数である．同様の導出方法により，

$$Q(\boldsymbol{q}) = C_{\boldsymbol{q}} \exp \langle \log P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{q}|s, \boldsymbol{\Lambda}) \rangle_{Q(\boldsymbol{\Lambda})} \tag{24}$$

$$\begin{aligned} Q(\boldsymbol{\Lambda}) &= C_{\boldsymbol{\Lambda}} P(\boldsymbol{\Lambda}) \exp \langle \log P(\boldsymbol{O}, \boldsymbol{Z}|S, \boldsymbol{\Lambda}) \rangle_{Q(\boldsymbol{Z})} \\ &\quad \times \exp \langle \log P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{q}|s, \boldsymbol{\Lambda}) \rangle_{Q(\boldsymbol{q})} \end{aligned} \tag{25}$$

が得られる．これらの近似分布は相互に関係しているため，個別に更新を繰り返すことにより，結果的に $\mathcal{F}$ を極大に導くことができる．また，各更新において，$\mathcal{F}$ は必ず増加するため収束性が保証されている．

さらに，$Q(\boldsymbol{\Lambda})$ に関して，$\boldsymbol{a}_i = \{a_{ij}\}_{j=1}^N$，$\boldsymbol{b}_i = \{\boldsymbol{\mu}_i, \boldsymbol{S}_i\}$ とし，事前確率を $P(\boldsymbol{\Lambda}) = \prod_{i=1}^N P(\boldsymbol{a}_i) \prod_{i=1}^N P(\boldsymbol{b}_i)$ と設定すると，直ちに，

$$Q(\boldsymbol{\Lambda}) = \prod_{i=1}^N Q(\boldsymbol{a}_i) \prod_{i=1}^N Q(\boldsymbol{b}_i) \tag{26}$$

$$\begin{aligned} Q(\boldsymbol{a}_i) &= C_{\boldsymbol{a}_i} P(\boldsymbol{a}_i) \exp\Big\{ \sum_{j=1}^N \sum_{t=1}^{T-1} \langle z_t^i z_{t+1}^j \rangle \log a_{ij} \Big\} \\ &\quad \times \exp\Big\{ \sum_{j=1}^N \sum_{t=1}^{T_x-1} \langle q_t^i q_{t+1}^j \rangle \log a_{ij} \Big\} \end{aligned} \tag{27}$$

$$\begin{aligned} Q(\boldsymbol{b}_i) &= C_{\boldsymbol{b}_i} P(\boldsymbol{b}_i) \exp\Big\{ \sum_{t=1}^T \langle z_t^i \rangle \log \mathcal{N}(\boldsymbol{o}_t|\boldsymbol{\mu}_i, \boldsymbol{S}_i^{-1}) \Big\} \\ &\quad \times \exp\Big\{ \sum_{t=1}^{T_x} \langle q_t^i \rangle \log \mathcal{N}((\boldsymbol{W}\boldsymbol{x})_t|\boldsymbol{\mu}_i, \boldsymbol{S}_i^{-1}) \Big\} \end{aligned} \tag{28}$$

となる．ただし，

$$\langle z_t^i \rangle = \sum_{\boldsymbol{Z}} Q(\boldsymbol{Z}) z_t^i, \quad \langle z_t^i z_{t+1}^j \rangle = \sum_{\boldsymbol{Z}} Q(\boldsymbol{Z}) z_t^i z_{t+1}^j \tag{29}$$

$$\langle q_t^i \rangle = \sum_{\boldsymbol{q}} Q(\boldsymbol{q}) q_t^i, \quad \langle q_t^i q_{t+1}^j \rangle = \sum_{\boldsymbol{q}} Q(\boldsymbol{q}) q_t^i q_{t+1}^j \tag{30}$$

である．また，$C_{\boldsymbol{a}_i}$ と $C_{\boldsymbol{b}_i}$ は，それぞれ $Q(\boldsymbol{a}_i)$ と $Q(\boldsymbol{b}_i)$ を正規化する定数であり，$C_{\boldsymbol{\Lambda}} = \prod_{i=1}^N C_{\boldsymbol{a}_i} \prod_{i=1}^N C_{\boldsymbol{b}_i}$ である．$Q(\boldsymbol{Z})$，$Q(\boldsymbol{q})$ についても，これらの分布を用いて，

$$\begin{aligned} Q(\boldsymbol{Z}) &= C_{\boldsymbol{Z}} \prod_{t=1}^{T-1} \exp\langle \log a_{z_t z_{t+1}} \rangle_{Q(\boldsymbol{a}_{z_t})} \\ &\quad \times \prod_{t=1}^{T} \exp\langle \log \mathcal{N}(\boldsymbol{o}_t|\boldsymbol{\mu}_{z_t}, \boldsymbol{S}_{z_t}^{-1}) \rangle_{Q(\boldsymbol{b}_{z_t})} \end{aligned} \tag{31}$$

$$\begin{aligned} Q(\boldsymbol{q}) &= C_{\boldsymbol{q}} \prod_{t=1}^{T_x-1} \exp\langle \log a_{q_t q_{t+1}} \rangle_{Q(\boldsymbol{a}_{q_t})} \\ &\quad \times \prod_{t=1}^{T_x} \exp\langle \log \mathcal{N}((\boldsymbol{W}\boldsymbol{x})_t|\boldsymbol{\mu}_{q_t}, \boldsymbol{S}_{q_t}^{-1}) \rangle_{Q(\boldsymbol{b}_{q_t})} \end{aligned} \tag{32}$$

と書ける．ここで，$Q(\boldsymbol{Z})$，$Q(\boldsymbol{q})$ は HMM の尤度関数と同じ形となっており，式 (29)，(30) の期待値は，Forward-Backward アルゴリズムを用いて計算することができる．

3.2 事前分布

モデルパラメータの事前確率 $P(\boldsymbol{\Lambda})$ として，共役事前分布を用いる．共役事前分布とは，事前分布と事後分布が同じ分布族となる事前分布のことである．尤度関数が式 (1) の HMM では，遷移確率は Dirichlet 分布，出力確率は Gauss-Wishart 分布となる．つまり，事後分布の近似分布である $Q(\boldsymbol{a}_i)$ と $Q(\boldsymbol{b}_i)$ がそれぞれ Dirichlet 分布と Gauss-Wishart 分布となる．以下に事前分布を定義する．

$$P(\boldsymbol{a}_i) = \mathcal{D}(\boldsymbol{a}_i|\boldsymbol{\phi}_i) = \frac{\Gamma\left(\sum_{j=1}^N \phi_{ij}\right)}{\prod_{j=1}^N \Gamma(\phi_{ij})} \prod_{j=1}^N a_{ij}^{\phi_{ij}-1} \tag{33}$$

$$P(\boldsymbol{b}_i) = \mathcal{N}(\boldsymbol{\mu}_i|\boldsymbol{\nu}_i, (\xi_i \boldsymbol{S}_i)^{-1}) \mathcal{W}(\boldsymbol{S}_i|\eta_i, \boldsymbol{B}_i) \tag{34}$$

$$\begin{aligned} &\mathcal{N}(\boldsymbol{\mu}_i|\boldsymbol{\nu}_i, (\xi_i \boldsymbol{S}_i)^{-1}) = \\ &\quad C_{\mathcal{N}_i} |\boldsymbol{S}_i|^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} Tr(\xi_i \boldsymbol{S}_i (\boldsymbol{\mu}_i - \boldsymbol{\nu}_i)(\boldsymbol{\mu}_i - \boldsymbol{\nu}_i)^\top \right\} \end{aligned} \tag{35}$$

$$\begin{aligned} &\mathcal{W}(\boldsymbol{S}_i|\eta_i, \boldsymbol{B}_i) \\ &\quad = C_{\mathcal{W}_i} |\boldsymbol{S}_i|^{\frac{\eta_i - D - 1}{2}} \exp\left\{ -\frac{1}{2} Tr(\boldsymbol{S}_i \boldsymbol{B}_i) \right\} \end{aligned} \tag{36}$$

$$C_{\mathcal{N}_i} = (2\pi)^{-\frac{D}{2}} \xi_i^{\frac{D}{2}} \tag{37}$$

$$C_{\mathcal{W}_i} = \frac{|\boldsymbol{B}_i|^{\frac{\eta_i}{2}}}{2^{\frac{\eta_i D}{2}} \pi^{\frac{D(D-1)}{4}} \prod_{d=1}^D \Gamma\left(\frac{\eta_i + 1 - d}{2}\right)} \tag{38}$$

とする．ここで，$\Gamma(\cdot)$ はガンマ関数，D は特徴ベクトルの次元数である．事前分布を表すパラメータをまとめると，$\{\phi_{ij}, \xi_i, \eta_i, \boldsymbol{\nu}_i, \boldsymbol{B}_i\}_{i,j=1}^N$ となる．また，共役事前分布を用いているため，事後分布も同様のパラメータセットで表すことができ，$\{\bar{\phi}_{ij}, \bar{\xi}_i, \bar{\eta}_i, \bar{\boldsymbol{\nu}}_i, \bar{\boldsymbol{B}}_i\}_{i,j=1}^N$ で表すこととする．

3.3 事後分布パラメータの更新

具体的な更新式を以下に示す．まず，事前分布のパラメータと $Q(\boldsymbol{Z})$，$Q(\boldsymbol{q})$ から計算される期待値を以下に定義する．

$$\bar{N}_i = \sum_{t=1}^T \langle z_t^i \rangle + \sum_{t=1}^{T_x} \langle q_t^i \rangle \tag{39}$$

$$\bar{N}_{ij} = \sum_{t=1}^{T-1} \langle z_t^i z_{t+1}^j \rangle + \sum_{t=1}^{T_x-1} \langle q_t^i q_{t+1}^j \rangle \tag{40}$$

$$\bar{\boldsymbol{o}}_i = \frac{1}{\bar{N}_i} \left(\sum_{t=1}^T \langle z_t^i \rangle \boldsymbol{o}_t + \sum_{t=1}^{T_x} \langle q_t^i \rangle (\boldsymbol{W}\boldsymbol{x})_t \right) \tag{41}$$

$$\bar{\boldsymbol{C}}_i = \frac{1}{\bar{N}_i}\Bigg(\sum_{t=1}^{T}\langle z_t^i\rangle(\boldsymbol{o}_t-\bar{\boldsymbol{o}}_i)(\boldsymbol{o}_t-\bar{\boldsymbol{o}}_i)^\top + \sum_{t=1}^{T_x}\langle q_t^i\rangle((\boldsymbol{W}\boldsymbol{x})_t-\bar{\boldsymbol{o}}_i)((\boldsymbol{W}\boldsymbol{x})_t-\bar{\boldsymbol{o}}_i)^\top\Bigg) \quad (42)$$

これらの期待値を用いて，モデルパラメータの事後確率分布 $Q(\boldsymbol{\Lambda})$ は，次式で更新される．

$$\bar{\phi}_{ij} = \phi_{ij}+\bar{N}_{ij},\quad \bar{\xi}_i = \xi_i+\bar{N}_i,\quad \bar{\eta}_i = \eta_i+\bar{N}_i \quad (43)$$

$$\bar{\boldsymbol{\nu}}_i = \frac{\bar{N}_i\bar{\boldsymbol{o}}_i+\xi_i\boldsymbol{\nu}_i}{\bar{N}_i+\xi_i} \quad (44)$$

$$\bar{\boldsymbol{B}}_i = \bar{N}_i\bar{\boldsymbol{C}}_i+\boldsymbol{B}_i+\frac{\bar{N}_i\xi_i}{\bar{N}_i+\xi_i}(\bar{\boldsymbol{o}}_i-\boldsymbol{\nu}_i)(\bar{\boldsymbol{o}}_i-\boldsymbol{\nu}_i)^\top \quad (45)$$

また，$Q(\boldsymbol{Z})$，$Q(\boldsymbol{q})$ の更新に用いられる期待値は，次式で表される．

$$\langle\log a_{ij}\rangle_{Q(\boldsymbol{a}_i)} = \Psi(\bar{\phi}_{ij})-\Psi\left(\sum_{k=1}^{N}\bar{\phi}_{kj}\right) \quad (46)$$

$$\langle\log\mathcal{N}(\boldsymbol{o}_t|\boldsymbol{\mu}_i,\boldsymbol{S}_i^{-1})\rangle_{Q(\boldsymbol{b}_i)} = -\frac{1}{2}\Bigg\{D\log\pi+\frac{D}{\bar{\xi}_i}-\sum_{d=1}^{D}\Psi\left(\frac{\bar{\eta}_i+1-d}{2}\right)+\log|\bar{\boldsymbol{B}}_i| + Tr\left\{\bar{\eta}_i\bar{\boldsymbol{B}}_i^{-1}(\boldsymbol{o}_t-\bar{\boldsymbol{\nu}}_i)(\boldsymbol{o}_t-\bar{\boldsymbol{\nu}}_i)^\top\right\}\Bigg\} \quad (47)$$

$$\langle\log\mathcal{N}((\boldsymbol{W}\boldsymbol{x})_t|\boldsymbol{\mu}_i,\boldsymbol{S}_i^{-1})\rangle_{Q(\boldsymbol{b}_i)} = -\frac{1}{2}\Bigg\{D\log\pi+\frac{D}{\bar{\xi}_i}-\sum_{d=1}^{D}\Psi\left(\frac{\bar{\eta}_i+1-d}{2}\right)+\log|\bar{\boldsymbol{B}}_i| + Tr\left\{\bar{\eta}_i\bar{\boldsymbol{B}}_i^{-1}((\boldsymbol{W}\boldsymbol{x})_t-\bar{\boldsymbol{\nu}}_i)((\boldsymbol{W}\boldsymbol{x})_t-\bar{\boldsymbol{\nu}}_i)^\top\right\}\Bigg\} \quad (48)$$

となる．ここで，$\Psi(\cdot)$ は digamma 関数である．これらの期待値を用いて，Forward-Backward アルゴリズムを実行することにより，$\bar{N}_i,\bar{N}_{ij},\bar{\boldsymbol{o}}_i,\bar{\boldsymbol{C}}_i$ を更新する．

3.4 ベイズ基準による音声パラメータの生成

変分法による近似事後分布の最適化を行うことにより，下限 $\mathcal{F}$ は周辺尤度 $P(\boldsymbol{W}\boldsymbol{x},\boldsymbol{O}|s,S)$ を良く近似することができる．よって，最適な音声パラメータは，$\mathcal{F}$ を最大化することで得られる．下限 $\mathcal{F}$ を $\boldsymbol{x}$ について，

$$\frac{\partial\mathcal{F}}{\partial\boldsymbol{x}} = \frac{\partial}{\partial\boldsymbol{x}}\langle\log P(\boldsymbol{W}\boldsymbol{x}|\boldsymbol{q},\boldsymbol{\Lambda})P(\boldsymbol{q}|s,\boldsymbol{\Lambda})\rangle_{Q(\boldsymbol{q})Q(\boldsymbol{\Lambda})} = \left\langle\frac{\partial}{\partial\boldsymbol{x}}\log P(\boldsymbol{W}\boldsymbol{x}|\boldsymbol{q},\boldsymbol{\Lambda})\right\rangle_{Q(\boldsymbol{q})Q(\boldsymbol{\Lambda})} = \boldsymbol{0} \quad (49)$$

とおくことにより，

$$\boldsymbol{W}^\top\langle\boldsymbol{S}\rangle\boldsymbol{W}\boldsymbol{x}^{(Bayes)} = \boldsymbol{W}^\top\langle\boldsymbol{S}\boldsymbol{\mu}\rangle \quad (50)$$

が得られる．ただし，

$$\langle\boldsymbol{S}\rangle = \sum_{\boldsymbol{q}}\int Q(\boldsymbol{q})Q(\boldsymbol{\Lambda})\boldsymbol{S}d\boldsymbol{\Lambda} = \mathrm{diag}[\langle\boldsymbol{S}_{q_1}\rangle,\langle\boldsymbol{S}_{q_2}\rangle,\ldots,\langle\boldsymbol{S}_{q_T}\rangle] \quad (51)$$

である．また，行列の各成分は独立に期待値計算ができるため，

$$\langle\boldsymbol{S}_{q_t}\rangle = \sum_{i=1}^{N}\langle q_t^i\rangle\int Q(\boldsymbol{b}_i)\boldsymbol{S}_i d\boldsymbol{\Lambda} = \sum_{i=1}^{N}\langle q_t^i\rangle\bar{\eta}_i\bar{\boldsymbol{B}}_i^{-1} \quad (52)$$

となる．同様に，

$$\langle\boldsymbol{S}\boldsymbol{\mu}\rangle = \sum_{\boldsymbol{q}}\int Q(\boldsymbol{q})Q(\boldsymbol{\Lambda})\boldsymbol{S}\boldsymbol{\mu}d\boldsymbol{\Lambda} = [\langle\boldsymbol{S}_{q_1}\boldsymbol{\mu}_{q_1}\rangle^\top,\langle\boldsymbol{S}_{q_2}\boldsymbol{\mu}_{q_2}\rangle^\top,\ldots,\langle\boldsymbol{S}_{q_T}\boldsymbol{\mu}_{q_T}\rangle^\top]^\top \quad (53)$$

$$\langle\boldsymbol{S}_{q_t}\boldsymbol{\mu}_{q_t}\rangle = \sum_{i=1}^{N}\langle q_t^i\rangle\int Q(\boldsymbol{b}_i)\boldsymbol{S}_i\boldsymbol{\mu}_i d\boldsymbol{\mu}_i d\boldsymbol{S}_i = \sum_{i=1}^{N}\langle q_t^i\rangle\bar{\eta}_i\bar{\boldsymbol{B}}_i^{-1}\bar{\boldsymbol{\nu}}_i \quad (54)$$

である．式 (50) は，ML 基準によるパラメータ生成 (式 (12)) と同じ形をしており，$\boldsymbol{S}$，$\boldsymbol{S}\boldsymbol{\mu}$ を，それぞれ，$\langle\boldsymbol{S}\rangle$，$\langle\boldsymbol{S}\boldsymbol{\mu}\rangle$ に置き換えることにより容易に実装可能である．また，事後分布 $Q(\boldsymbol{\Lambda})$，$Q(\boldsymbol{q})$，$Q(\boldsymbol{Z})$ の更新には，$\boldsymbol{x}$ を用いるため，$\boldsymbol{x}$ の更新と事後分布パラメータの更新を繰り返すことにより，より最適な音声パラメータを求めることができる．この繰返し手順には，様々な組合せが考えられるが，例えば以下のような手順となる．

（1） 事前分布パラメータと $\boldsymbol{x}$ の初期値を与える．
（2） $Q(\boldsymbol{Z})$ の更新 (式 (46)，(47))
（3） $Q(\boldsymbol{q})$ の更新 (式 (46)，(48))．
（4） Forward-Backward アルゴリズム (式 (39)–(42))
（5） $Q(\boldsymbol{\Lambda})$ の更新 (式 (43)–(45))
（6） $\boldsymbol{x}$ の更新 (式 (50))．
（7） $\mathcal{F}$ が増加しなければ終了．それ以外は (2) へ．

3.5 コンテキストクラスタリング

HMM 音声合成では，HMM による音声認識に比べ，韻律やアクセントなどの様々な情報をコンテキストとして HMM を学習するため，コンテキストクラスタリングが重要な役割を果たす．これまで，コンテキストクラスタリングにおいても，HMM の学習と同様，ML 基準が用いられてきた．また，クラスタリングの分割停止条件として，MDL 基準が良く用いられる．しかし，MDL 基準はその導出に，学習データが十分に得られるという仮定を置いており，データ量が少ない場合，十分な精度が得られない可能性が指摘されている．この問題に対し，音声認識において，ベイズ基準によるコンテキストクラスタリング [11] が提案されており，有効性が確認されている．本研究でも同様に，決定木によるコンテキストクラスタリングにベイズ基準を用いる．

ベイズ基準のコンテキストクラスタリングでは，$Q(\boldsymbol{Z})$，$Q(\boldsymbol{q})$，$Q(\boldsymbol{a})$ を固定した状態で，$\mathcal{F}$ を最大にする $Q(\boldsymbol{b})$ の共有構造を決定する．$\mathcal{F}$ に近似事後確率を代入し，クラスタリングに関係のある項だけを考慮すると，

$$\mathcal{F} = -\log C_{\boldsymbol{\Lambda}}+E\left[Q(\boldsymbol{q})\right]+E\left[Q(\boldsymbol{Z})\right] = -\sum_{i=1}^{N}\log C_{\boldsymbol{b}_i}+Const \quad (55)$$

となる．ただし，

$$\log C_{\boldsymbol{b}_i} = \log\frac{\bar{C}_{\mathcal{N}_i}\bar{C}_{\mathcal{W}_i}}{C_{\mathcal{N}_i}C_{\mathcal{W}_i}}(2\pi)^{\frac{\bar{N}_i D}{2}} \quad (56)$$

である．簡単のため事前確率はすべてのガウス分布で等しいと仮定し，その正規化定数を $C_{\mathcal{N}_0}$，$C_{\mathcal{W}_0}$ とすると，あるクラスタが分割された時の $\mathcal{F}$ の変化量は，

$$\begin{aligned}\delta\mathcal{F} &= -\log C_{\boldsymbol{b}_y} - \log C_{\boldsymbol{b}_n} + \log C_{\boldsymbol{b}_p} \\ &= -\log \bar{C}_{\mathcal{N}_y}\bar{C}_{\mathcal{W}_y} - \log \bar{C}_{\mathcal{N}_n}\bar{C}_{\mathcal{W}_n} \\ &\quad + \log \bar{C}_{\mathcal{N}_p}\bar{C}_{\mathcal{W}_p} + \log C_{\mathcal{N}_0}C_{\mathcal{W}_0} \\ &= \delta f_y + \delta f_n - \delta f_p - \delta f_0 \end{aligned} \tag{57}$$

と書ける．ただし，

$$\delta f_i = -\frac{D}{2}\log\xi_i - \frac{\eta_i}{2}\log|\boldsymbol{B}_i| + \sum_{d=1}^{D}\log\Gamma\left(\frac{\eta_i+1-d}{2}\right) \tag{58}$$

である．ここで，$C_{\boldsymbol{b}_p}$ は分割前，$C_{\boldsymbol{b}_y}$，$C_{\boldsymbol{b}_n}$ は分割後の事後確率分布の正規化定数であり，各クラスタの統計量 $\bar{N}_i, \bar{\boldsymbol{o}}_i, \bar{\boldsymbol{C}}_i$ から計算される．$\delta\mathcal{F}$ を最大にする質問を選択しながら分割を繰返し，$\delta\mathcal{F} \leqq 0$ で分割を停止することにより，ベイズ基準によるクラスタリングを実現できる．

3.6 実装における近似

これまでに説明したアルゴリズムでは，$Q(\boldsymbol{Z})$ は，$Q(\boldsymbol{\Lambda})$ を介して $\boldsymbol{x}$ に依存する．よって，合成音声のラベル列 s が決定した後で，すべての学習データに対する Forward-Backward アルゴリズムを計算する必要がある．しかし，実際に，この計算を行うことは計算量の観点から現実的ではない．そこで，実装においては，$Q(\boldsymbol{Z})$ は $\boldsymbol{x}$ に依存しないと仮定し，あらかじめ学習データから計算した $Q(\boldsymbol{Z})$ で固定することを考える．

合成音声のラベル列 s が与えられる前は，式 (14) における $P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{q}|s, \boldsymbol{\Lambda})$ を計算することができないため，この項を考慮しないで $Q(\boldsymbol{Z})$，$Q(\boldsymbol{\Lambda})$ を最適化する．すなわち，$P(\boldsymbol{W}\boldsymbol{x}, \boldsymbol{O}|, s, S)$ を $P(\boldsymbol{O}|S)$ に置き換える．このとき，最適な事後分布は，

$$Q(\boldsymbol{Z}) = C_{\boldsymbol{Z}} \exp\langle\log P(\boldsymbol{O}, \boldsymbol{Z}|S, \boldsymbol{\Lambda})\rangle_{Q(\boldsymbol{\Lambda})} \tag{59}$$

$$Q(\boldsymbol{\Lambda}) = C_{\boldsymbol{\Lambda}} P(\boldsymbol{\Lambda}) \exp\langle\log P(\boldsymbol{O}, \boldsymbol{Z}|S, \boldsymbol{\Lambda})\rangle_{Q(\boldsymbol{Z})} \tag{60}$$

で得られる．また，この分布を用いた期待値計算は，式 (39)–(42) を次式に置き換えることにより，計算可能である．

$$\bar{N}_i = \sum_{t=1}^{T}\langle z_t^i\rangle, \quad \bar{N}_{ij} = \sum_{t=1}^{T-1}\langle z_t^i z_{t+1}^j\rangle \tag{61}$$

$$\bar{\boldsymbol{o}}_i = \frac{1}{\bar{N}_i}\sum_{t=1}^{T}\langle z_t^i\rangle\boldsymbol{o}_t \tag{62}$$

$$\bar{\boldsymbol{C}}_i = \frac{1}{\bar{N}_i}\sum_{t=1}^{T}\langle z_t^i\rangle(\boldsymbol{o}_t - \bar{\boldsymbol{o}}_i)(\boldsymbol{o}_t - \bar{\boldsymbol{o}}_i)^\top \tag{63}$$

合成システムを実現するための手順は，以下のようになる．

- 学習部

（1） 事前分布パラメータを与える．
（2） $Q(\boldsymbol{Z})$ の更新 (式 (46)，(47))．
（3） Forward-Backward アルゴリズム (式 (61)–(63))．
（4） $Q(\boldsymbol{\Lambda})$ の更新 (式 (43)–(45))．
（5） $\mathcal{F}$ が増加しなければ終了．それ以外は (2) へ．

- 合成部

（1） ラベル列 s に基づく $\boldsymbol{x}$ の初期値を与える．
（2） $Q(\boldsymbol{q})$ の更新 (式 (46)，(48))
（3） Forward-Backward アルゴリズム (式 (39)–(42))
（4） $Q(\boldsymbol{\Lambda})$ の更新 (式 (43)–(45))
（5） $\boldsymbol{x}$ の更新 (式 (50))．
（6） $\mathcal{F}$ が増加しなければ終了．それ以外は (2) へ．

合成部では，$Q(\boldsymbol{\Lambda})$ の更新を行わないなどの様々な組合せが考えられるが，どのような手順が適当であるかは，今後，評価実験により調査する必要がある．現在，実験として，学習部の実装を終えており，合成部において状態系列 $\boldsymbol{q}$ を外部から与え，$Q(\boldsymbol{q})$ と $Q(\boldsymbol{\Lambda})$ を更新しない簡単な場合について，合成音声の生成を確認している．また，2.2 で述べたように，ここで定義した HMM は状態系列の最適化には不十分であり，状態継続長モデルが必要となるが，本論文では，その点は考慮していない．そのため，実装上は，$Q(\boldsymbol{q})$ を状態継続長モデルを用いて計算するなど，何らかの対処が必要である．HMM の定義として，遷移系列を最適化できるモデルを定義し，学習も含めた最適化を行うのが理想的であるが，計算量の問題があり，今後の課題である．

4. ま と め

本論文では，ベイズ基準による HMM 音声合成を導出した．今後の課題は，提案手法の評価と状態継続長モデルを用いた状態系列と音声パラメータの同時最適化である．

文　献

[1] H. Attias, “Inferring parameters and structure of latent variable models by variational Bayes,” Proc. UAI, 1999.
[2] 上田修功, “ベイズ学習,”（全 4 回）電子情報通信学会誌，Vol. 85, No. 4,6,7,8, 2002.
[3] S. Watanabe, Y. Minami, A. Nakamura, and N. Ueda, “An application of variational Bayesian approach to speech recognition,” NIPS15, 2002.
[4] 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・継続長の同時モデル化,” 信学論 (D-II), vol.J83-D-II, no.11, Nov. 2000.
[5] 益子貴史, 徳田恵一, 小林隆夫, 今井 聖, “動的特徴を用いた HMM に基づく音声合成,” 信学論 (D-II), vol.J79-D-II, no.12, pp.2184–2190, Dec. 1996.
[6] 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.
[7] K. Tokuda, Takayoshi Yoshimura, T. Masuko, T. Kobayashi and T. Kitamura, “Speech parameter generation algorithms for HMM-based speech synthesis,” Proc. ICASSP, pp.1315–1318, June 2000.
[8] K. Tokuda, T. Kobayashi and S. Imai, “Speech parameter generation from HMM using dynamic features,” Proc. ICASSP, pp.660–663, 1995.
[9] K. Tokuda, T. Masuko, T. Yamada, T. Kobayashi and Satoshi Imai, “An algorithm for speech parameter generation from continuous mixture HMMs with dynamic features,” Proc. EUROSPEECH, pp.757–760, 1995.
[10] K. Koishida, K. Tokuda, T. Masuko and T. Kobayashi, “Vector quantization of speech spectral parameters using statistics of dynamic features,” Proc. ICSP, vol.1, pp.247–252, Aug. 1997.
[11] 渡辺 晋治, 南 康浩, 中村 篤, 上田修功, “ベイズ的基準を用いた状態共有型 HMM 構造の選択,” 信学論 (D-II), vol.J86-D-II, no.6, pp.776–786, 2003.