社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2008－SLP－74（37）
2008 / 12 / 10

HMM音声合成における共分散パラメータの共有に関する検討

大浦圭一郎† 全 炳河† 南角 吉彦† 李 晃伸† 徳田 恵一†

† 名古屋工業大学 大学院工学研究科 情報工学専攻
〒 466–8555 愛知県 名古屋市 昭和区 御器所町
E-mail: †{uratec,zen,nankaku,ri,tokuda}@sp.nitech.ac.jp

あらまし 本報告では隠れマルコフモデル (Hidden Markov Model; HMM) に基づく音声合成システムにおける共分散パラメータの共有について述べる．近年，音声合成システムへの需要が高まっており，HMM に基づいた音声合成システムでは音声波形の断片をそのまま利用するのではなく，音声波形の特徴を HMM によりモデル化し，HMM のモデルパラメータを合成システムに保持するため，同程度の音質の波形接続法式に比べてフットプリントが小さい利点がある．中でも組み込み向けのシステムには携帯電話，PDA，カーナビ，情報家電，ゲーム機等への用途があるが，必要な CPU，メモリ等が制限されることが多く，更なるフットプリントの縮小が必要である．HMM に基づく音声合成システムにコンテキスト依存モデルを用いることで高精度な音響モデルを構築することができ，決定木に基づくコンテキストクラスタリングを用いて状態共有構造を構築する際に，組み込み用途向けに決定木のサイズを小さくすることも考えられるが，音質が劣化する．本報告では，平均に比べて共分散が音質に与える影響が小さいことに注目し，全てのパラメータの共分散を共有する手法を提案する．このパラメータ共有を仮定した上でのコンテキストクラスタリングを行い，主観評価実験により，パラメータ数を大幅に削減するのみならず，若干の品質改善を達成した．
キーワード 隠れマルコフモデル，音声合成，決定木，コンテキストクラスタリング，MDL 基準，組み込み機器

Tying covariance parameters for HMM-based speech synthesis

Keiichiro OURA†, Heiga ZEN†, Yoshihiko NANKAKU†, Akinobu LEE†, and Keiichi TOKUDA†

† Depertment of Computer Science and Enginnering, Naogya Institute of Technology
Gokiso-cho, Showak-ku, Nagoya, Aichi, 466-8555 Japan
E-mail: †{uratec,zen,nankaku,ri,tokuda}@sp.nitech.ac.jp

Abstract In this paper, we proposed a tying covariance technique in hidden Markov model (HMM) basedspeech synthesis system. In recent years, context-dependent model are used for training high quality model in hidden Markov model (HMM) based speech synthesis system. However, the use of context-dependent models results in too many free-parameters in a system, hence it is difficult to estimate the model which is statistically reliable. This is a fatal problem for development of embedded devices (mobile phone, PDA, etc...) especially. Therefore, variaous parameter clustering techniques have been proposed. The use of desicion tree based context-clustering approach is a good solution to this problem. The splitting procedure of the decision tree provides a way of keeping the balance of model complexity and robustness. Furthermore, by incorporating phonetic knowledge into questions, it can assign unseen context-dependent models to the leaf node of decision trees. In this paper, a new approach is proposed by tying all covariances. In subjective experimental results, proposed technique archived higher MOS score and smaller number of parameters than conditional technique.
Key words Hidden Markov Model, Speech Synthesis, Desicion Tree, Context-Clustering, MDL Criterion, Embedded Device

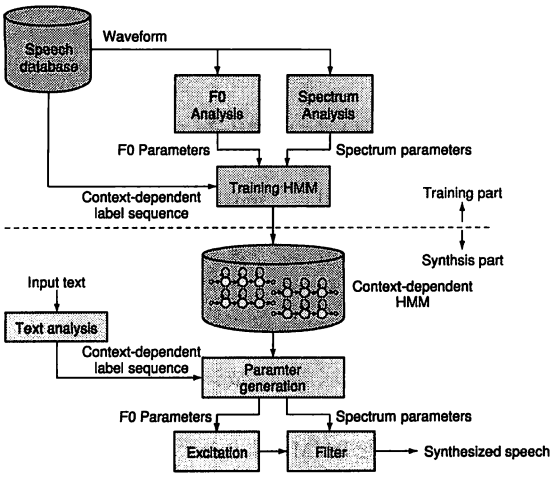

図 1　HMM 音声合成システム

Fig. 1　HMM based speech synthesis system.

1. はじめに

近年，音声合成システムへの需要が高まっている．中でも組み込み向けのシステムには携帯電話，PDA，カーナビ，情報家電，ゲーム機等への用途があるが，必要な CPU，メモリ等が制限されることが多く，このような条件下で高音質な音声を合成することが課題となっている．本報告では，組み込み用途向けの音声合成システムの小メモリ化，高音質化に関して検討する．

音声合成システムにはあらかじめ収録した音声波形を断片化し，それを繋ぎ合わせて音声を合成する手法が多く採用されている．この手法ではデータベース中の話者の音声をほぼそのまま使用するために，肉声感は高いものの波形の接続部分で不連続感が目立つ弱点がある．大規模なデータベースを用いることで不連続感を軽減することができるが，データ量が膨大になるために組み込みは難しいという問題があった．これに対し，隠れマルコフモデル (Hidden Markov Model; HMM) に基づいた音声合成システム (図 1) では音声波形の断片をそのまま利用するのではなく，音声波形の特徴を HMM によりモデル化し，HMM モデルパラメータを合成システムに保持するため，音声合成時に必要なデータサイズが同程度音質の波形接続法式に比べて小さい利点がある．

一般に音素の音響的特徴は前後の音素や言語環境，品詞，アクセントなどにより変化することが知られている．これらのコンテキストにより分割してモデル化を行うコンテキスト依存 HMM は，音響的特徴をより精密にモデル化できると考えられ，多くの研究機関において研究・利用されている．

しかし，コンテキスト依存モデルを用いることでその総モデル数は膨大なものとなり，全てのモデルを学習データ中に用意することが困難，各モデルあたりの学習データ量が不足しパラメータ推定精度が低下，といった問題が生じる．このため，様々なモデルパラメータ共有手法が提案されており [2]〜[5]，音素決定木に基づくコンテキストクラスタリング [5] はこの問題の優れた解決法の 1 つである．本手法はコンテキストを分割条件として，コンテキスト依存の HMM 状態の集合に対してトップダウンにクラスタリングを行い，クラスタリング終了時の決定木のリーフクラスタに含まれる状態を共有することにより HMM 状態共有構造を構築する手法である．また，決定木をたどることで，学習データ中に存在しないモデルを対応するリーフクラスタに割り当てて生成できる．

コンテキストクラスタリングに関する研究は盛んに行われており，コンテキスト分割条件を選択するための基準に関する研究 [6]〜[8]，単一ガウス分布から混合ガウス分布への拡張 [9], [10] など様々な手法が提案されている．

これらの手法はいずれも HMM 状態単位での共有構造を構築しており，同数の平均と共分散のパラメータが得られる．より十分な精度の学習を行うためには共分散より平均に重みをおいたクラスタリングをする必要があると考えられる．そこで本報告では全クラスタの共分散を共有し，MDL (Minimum Description Length) [11] に基づく分割基準を用いて平均の共有構造を構築するモデルの学習手法を提案する (図 2)．スペクトルパラメータを 0 から 39 次元のベクトルとその Δ，Δ^2 とし，リーフクラスタが 1000 クラスタ，1 パラメータにつき 4byte と考えると，従来法では $(120+120)\times 1000\times 4 = 937$KB のデータサイズが必要であるのに対し，共分散を共有することで $((120\times 1000)+120)\times 4 = 469$KB のデータサイズに削減できる．なお，全共分散を用いた HMM の学習には膨大な学習データを必要とするため，ここでの共分散には対角共分散を用いた．共分散を共有する場合に適した平均の状態共有構造を構築するために，共分散を共有する仮定のもとでコンテキストクラスタリングを導出し，主観評価実験で評価した．

以下，次章では共分散の共有手法について紹介し，第 3 章では主観評価実験について述べる．

2. 共分散の共有

2.1 決定木に基づくコンテキストクラスタリング

決定木に基づくコンテキストクラスタリングはコンテキストを分割条件として，コンテキスト依存の HMM 状態の集合に対してトップダウンにクラスタリングを行い，クラスタリング終了時の決定木のリーフクラスタに含まれる状態を共有することにより HMM 状態共有構造を構築する手法である．

ある状態 m に対する観測ベクトル $\boldsymbol{o}$ の平均 $\boldsymbol{\mu}_m$ と共分散 $\boldsymbol{\Sigma}_m$ は次のように定義される．

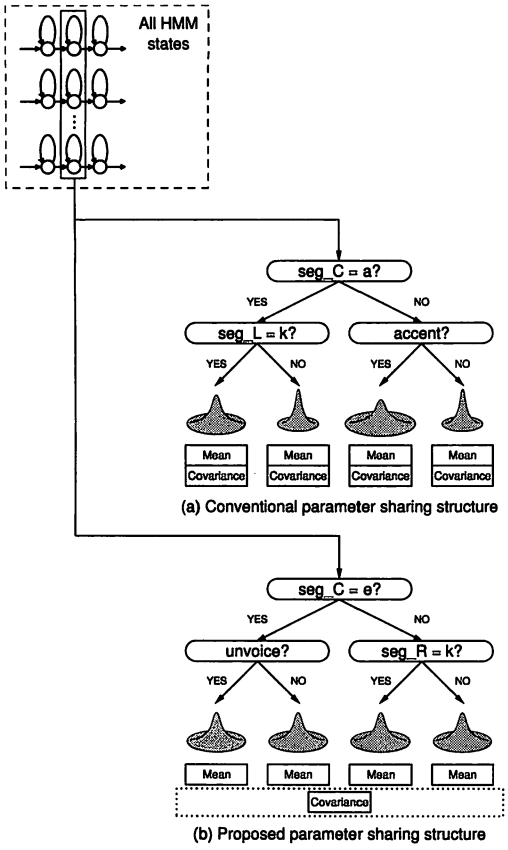

図 2　パラメータ共有構造

Fig. 2　Parameter sharing structures

$$\boldsymbol{\mu}_m = \frac{\sum_{t=1}^{T} \gamma_t \boldsymbol{o}_t}{\sum_{t=1}^{T} \gamma_t} \tag{1}$$

$$\begin{aligned}\boldsymbol{\Sigma}_m &= \frac{\sum_{t=1}^{T} \gamma_t (\boldsymbol{o}_t - \boldsymbol{\mu}_m)(\boldsymbol{o}_t - \boldsymbol{\mu}_m)^\top}{\sum_{t=1}^{T} \gamma_t} \\ &= \frac{\sum_{t=1}^{T} \gamma_t \boldsymbol{o}_t \boldsymbol{o}_t^\top}{\sum_{t=1}^{T} \gamma_t} - \boldsymbol{\mu}_m \boldsymbol{\mu}_m^\top \end{aligned} \tag{2}$$

ここで γ は各フレームにおける学習データ量である．さらに観測ベクトル $\boldsymbol{o}$ の対数尤度は次の式で定義される．

$$\begin{aligned}&\sum_{t=1}^{T} \gamma_t(m) \log P(\boldsymbol{o}_t \mid \lambda_m) \\ &= \sum_{t=1}^{T} \gamma_t(m) \Big(-\frac{1}{2} \log(2\pi |\boldsymbol{\Sigma}_m|) \\ &\quad -\frac{1}{2} (\boldsymbol{o}_t - \boldsymbol{\mu}_m)^\top \boldsymbol{\Sigma}_m^{-1} (\boldsymbol{o}_t - \boldsymbol{\mu}_m) \Big) \\ &= -\frac{1}{2} \sum_{t=1}^{T} \gamma_t(m) (\log(2\pi |\boldsymbol{\Sigma}_m|) + n) \end{aligned} \tag{3}$$

ここで n は分析次数である．

本報告ではコンテキストクラスタリングの停止規準に MDL 基準を用いており，記述長が最小になるように構築される．質問 q を用いてクラスタ S を S_{q+} と S_{q-} に分割する際，クラスタ分割前後の記述長の変化量は次のように定義される．

$$\begin{aligned}\Delta_q &= \frac{1}{2}\Gamma(S_{q+}) \left(\log\left(2\pi \left|\Sigma_{S_{q+}}\right|\right) + n\right) \\ &\quad + \frac{1}{2}\Gamma(S_{q-}) \left(\log\left(2\pi \left|\Sigma_{S_{q-}}\right|\right) + n\right) \\ &\quad - \frac{1}{2}\Gamma(S_q) \left(\log\left(2\pi \left|\Sigma_{S_q}\right|\right) + n\right) + K \log \Gamma(S_0) \\ &= \frac{1}{2}\Big\{\Gamma(S_{q+}) \log \left|\boldsymbol{\Sigma}_{S_{q+}}\right| + \Gamma(S_{q-}) \log \left|\boldsymbol{\Sigma}_{S_{q-}}\right| \\ &\quad - \Gamma(S) \log |\boldsymbol{\Sigma}_S| \Big\} + K \log \Gamma(S_0) \end{aligned} \tag{4}$$

ここで，$\Gamma(\cdot)$ は各クラスタの学習データ量，S_0 は決定木のルートクラスタである．K は分割によって増えるパラメータ数であり，$\boldsymbol{\Sigma}$ が対角共分散と定義される場合は $K = n + n$ になり，全共分散の場合は $K = n + \frac{n(n+1)}{2}$ になる．これは ML 基準に基づくコンテキストクラスタリングにおいて，分割前後の尤度の変化量の閾値が $K \log \Gamma(S_0)$ となっていると考えることができる．

2.2　共分散を共有したコンテキストクラスタリング

HMM 音声合成システムに用いられる決定木に基づくコンテキストクラスタリングは HMM 状態単位での共有構造を構築しており，同数の平均と共分散のパラメータが得られる．より十分な精度の学習を行うためには共分散より平均に重みをおいたクラスタリングをする必要があると考えられる．そこで本報告では全クラスタの共分散を共有し，MDL に基づく分割基準を用いて平均の共有構造を構築する

共有した共分散を $\boldsymbol{\Sigma}_g$ とすると，観測ベクトル $\boldsymbol{o}$ の対数尤度は次のように定義される．

$$
\begin{aligned}
&\sum_{t=1}^{T}\gamma_t(m)\log P\left(\boldsymbol{o}_t \mid \lambda'_m\right)\\
&= \sum_{t=1}^{T}\gamma_t(m)\left(-\frac{1}{2}\log\left(2\pi\left|\boldsymbol{\Sigma}_g\right|\right)\right.\\
&\quad \left.-\frac{1}{2}\left(\boldsymbol{o}_t-\boldsymbol{\mu}_m\right)^{\top}\boldsymbol{\Sigma}_g^{-1}\left(\boldsymbol{o}_t-\boldsymbol{\mu}_m\right)\right)\\
&= -\frac{1}{2}\sum_{t=1}^{T}\gamma_t(m)\log\left(2\pi\left|\boldsymbol{\Sigma}_g\right|\right)\\
&\quad -\frac{1}{2}\sum_{t=1}^{T}\gamma_t(m)\left(\boldsymbol{o}_t-\boldsymbol{\mu}_m\right)^{\top}\boldsymbol{\Sigma}_g^{-1}\left(\boldsymbol{o}_t-\boldsymbol{\mu}_m\right)\\
&= -\frac{1}{2}\sum_{t=1}^{T}\gamma_t(m)\left(\log\left(2\pi\left|\boldsymbol{\Sigma}_g\right|\right)+tr\left(\boldsymbol{\Sigma}_m\boldsymbol{\Sigma}_g^{-1}\right)\right) \qquad (5)
\end{aligned}
$$

同様に分割前後の記述長の変化量を示す.

$$
\begin{aligned}
\Delta_q = &\ \frac{1}{2}\Gamma\left(S_{q+}\right)\left(\log\left(2\pi\left|\boldsymbol{\Sigma}_{S_g}\right|\right)+tr\left(\boldsymbol{\Sigma}_{S_{q+}}\boldsymbol{\Sigma}_g^{-1}\right)\right)\\
&+\frac{1}{2}\Gamma\left(S_{q-}\right)\left(\log\left(2\pi\left|\boldsymbol{\Sigma}_{S_g}\right|\right)+tr\left(\boldsymbol{\Sigma}_{S_{q-}}\boldsymbol{\Sigma}_g^{-1}\right)\right)\\
&-\frac{1}{2}\Gamma\left(S\right)\left(\log\left(2\pi\left|\boldsymbol{\Sigma}_{S_g}\right|\right)+tr\left(\boldsymbol{\Sigma}_S\boldsymbol{\Sigma}_g^{-1}\right)\right)\\
&+n\log\Gamma\left(S_0\right)\\
= &\ \frac{1}{2}\left\{\Gamma\left(S_{q+}\right)tr\left(\boldsymbol{\Sigma}_{S_{q+}}\boldsymbol{\Sigma}_g^{-1}\right)+\Gamma\left(S_{q-}\right)tr\left(\boldsymbol{\Sigma}_{S_{q-}}\boldsymbol{\Sigma}_g^{-1}\right)\right.\\
&\left.-\Gamma\left(S_q\right)tr\left(\boldsymbol{\Sigma}_S\boldsymbol{\Sigma}_g^{-1}\right)\right\}+n\log\Gamma\left(S_0\right) \qquad (6)
\end{aligned}
$$

提案法では共分散を共有しているため，増加パラメータ数は平均の分析次数である n のみである．より音質に影響を及ぼすと思われる平均を重視したクラスタリングになるので音質の向上が期待できる．

3. 主観評価実験

3.1 実験条件

提案法の有効性を示すため，MOS 値に基づく主観評価実験を行った．学習データとして，ATR 日本語音声データベース B セットの男性話者 MHT による音韻バランス文 503 文章中の 450 文章を用いた．サンプリング周波数は 16kHz，分析周期は 5ms とした．

学習に用いる特徴ベクトルはスペクトルパラメータ，基本周波数パラメータから成る．スペクトルパラメータとしては，STRAIGHT [12] によって抽出されたスペクトルに，分析パラメータを $\alpha = 0.42$ としたメルケプストラム分析 [1] を適用することにより得られた 39 次元のメルケプストラムパラメータとその Δ，Δ^2 を用いた．基本周波数パラメータとしては対数基本周波数とその Δ，Δ^2 を用いた．

HMM は 5 状態のスキップなし left-to-right の隠れセミマルコフモデル (HSMM) [13] とし，音素をモデルの単位とした．メルケプストラムパラメータは連続分布 HMM，対数基本周波数は多空間分布 HMM (MSD-HMM)，継続長は多次元ガウス分布でモデル化する．メルケプストラムパラメータ，対数基本周波数，継続長に関する出力確率分布はそれぞれ独立にコンテキストクラスタリングされる．

HMM からのパラメータ生成には発話内変動を考慮したパラメータ生成法 [14] を用いた．

MDL 基準におけるペナルティ項 (文献 [11] の式 (9)) の重みを変化させることにより，様々なパラメータ数のモデルを学習した．重みは 8.0，4.0，2.0，1.0，0.5，0.25 の 6 種類を用いた．メルケプストラムパラメータと対数基本周波数のコンテキストクラスタリングは独立に行ったが，MDL 基準におけるペナルティ項の重みは共通のものを用いている．

被験者は 10 名であり，各被験者は被験者毎に 53 文章の中からランダムに選ばれた 10 文章を 1 点から 5 点の 5 段階で評価した．

3.2 実験結果

まず，共分散を共有することによるモデルの劣化，パラメータ数の削減を調べるため，以下の 2 手法を評価した．

NORMAL 従来法．平均，共分散を共にコンテキストクラスタリング.

TIED DIAGC1 平均は NORMAL と同じ共有構造をもち，全状態の対角共分散は全体で共有.

主観評価実験の結果を図 3 に示す．横軸が総パラメータ数，縦軸が MOS 値，図中の数値は MDL 基準におけるペナルティ項の重みである．“NORMAL” と “TIED DIAGC1” を比較したところ，平均の共有構造をそのままに共分散を共有することで，約半分のパラメータ数で従来法とほぼ同等の MOS 値が得られた．

次に以下の手法を用いて実験した．

NORMAL 従来法．平均，共分散を共にコンテキストクラスタリング.

TIED DIAGC2 全状態の対角共分散を全体で共有．平均の共有構造は，共分散を共有することを仮定して導出したコンテキストクラスタリングを用いて構築.

主観評価実験の結果を図 4 に示す．横軸が総パラメータ数，縦軸が MOS 値，図中の数値は MDL 基準におけるペナルティ項の重みである．“NORMAL” と “TIED DIAGC2” を比較したところ，提案法ではより少ないパラメータで従来法と同等かそれ以上の MOS 値が得られた．図 3 と比較すると，平均の共有構造の構築に従来のコンテキストクラスタリングを用いるより，共分散を共有することを仮定して導出したコンテキストクラスタリングを用いたほうが適していると考えられる．一つのパラメータを 4byte と考えた時，“NORMAL” と “TIED DIAGC2” で MDL 基準におけるペナルティ項の重みが 1.0 の点を比較すると，提案法はより高い MOS 値で 813KB から

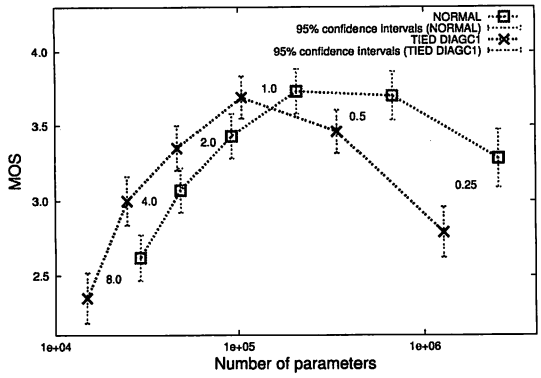

図 3 主観評価実験結果 1

Fig. 3 Experimental results 1

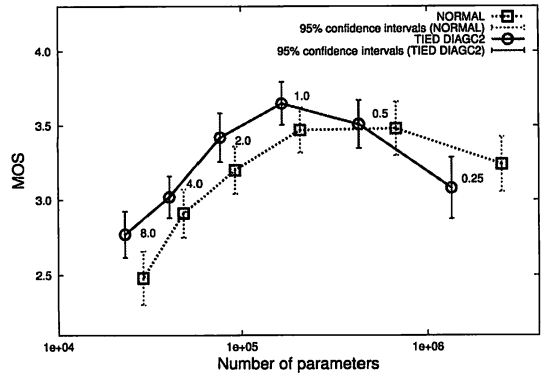

図 4 主観評価実験結果 2

Fig. 4 Experimental results 2

649KB へパラメータ数を削減できていることがわかる．さらに，従来法における重みが 1.0 の点と提案法における重みが 2.0 の点を比較すると，ほぼ同等の MOS 値で 813KB から 300KB への大幅なパラメータ数の削減が確認できた．

各手法におけるメルケプストラムの MDL 基準のペナルティ項の重みに対応するリーフクラスタ数と総パラメータ数をそれぞれ図 5，6 に示す．横軸が MDL 基準のペナルティ項の重み，縦軸がそれぞれリーフクラスタ数と総パラメータ数である．図 5 より，従来法と比較して提案法では同じ重みを用いた場合でもより多くの分割が行われることがわかる．これは従来のコンテキストクラスタリング時，平均と共分散の分割に必要だった学習データ量を，提案法では平均の分割にだけ使えるからと考えられる．だだし，図 6 の通り，共分散の共有によって総パラメータ数では提案法が従来法を下回ることが確認できる．

従来法で学習された全クラスタの共分散の平均と，提案法で共有した共分散を図 7，8，9 に示す．図 7，8，9 はそれぞれ，メルケプストラムと対数基本周波数の static パラメータ，およびそれらの Δ，Δ^2 である．図 7 より，メルケプストラムの static パラメータの共分散は従来法と比較してあまり変化がない．これは共分散の共有によって大きくなるはずの共分散を，リーフクラスタ数を多くすることで打ち消していると考えられる．図 8，9 から，メルケプストラムの static パラメータと比べて Δ，Δ^2 パラメータは共分散が少し大きくなっていることがわかるが，音質に最も影響が高いと思われる static パラメータの共分散の値に大きな変化が無いので，音質への影響は小さいと思われる．

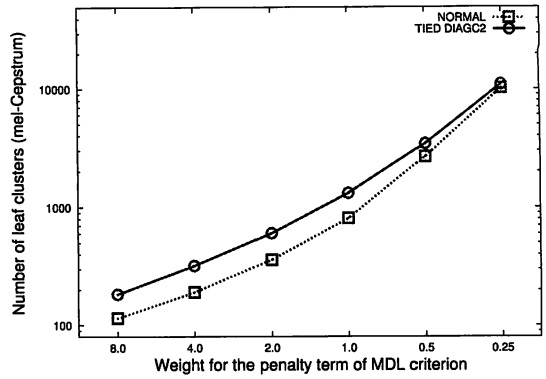

図 5 MDL 基準のペナルティ項の重みとリーフクラスタ数 (メルケプストラム)

Fig. 5 Number of leaf clusters (mel-Cepstrum) versus the weight of penalty term of MDL criterion.

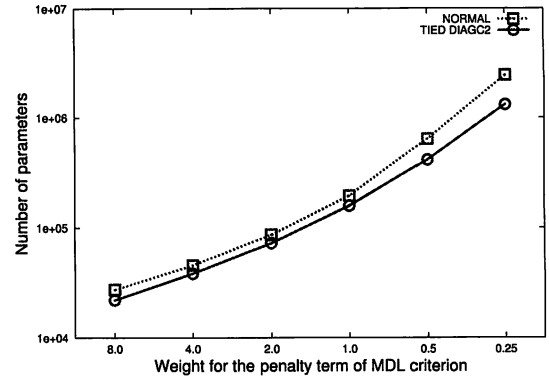

図 6 MDL 基準のペナルティ項の重みと総パラメータ数 (メルケプストラム)

Fig. 6 Number of parameters (mel-Cepstrum) versus the weight of penalty term of MDL criterion.

4. むすび

本稿では，HMM に基づく音声合成システムにおける共分散の共有に関する検討をすることにより，主観評価実験の結果，パラメータ数を大幅に削減するのみならず，若干の品質改善を達成した．今後の課題として，全共分散や STC(Semi-Tied Covariance) [15] 等の共分散の共有に関する評価があげられる．

References

[1] 徳田恵一, 小林隆夫, 千葉健司, 今井聖, "メル一般化ケプストラ

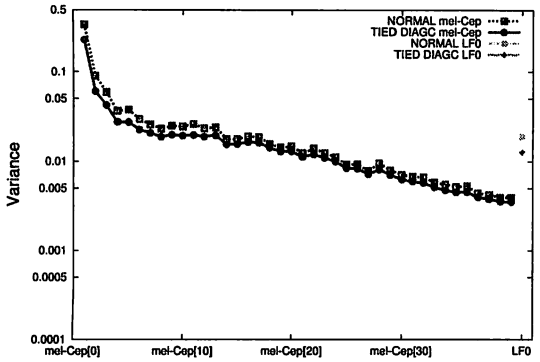

図 7 メルケプストラムパラメータと対数基本周波数の対角共分散

Fig. 7 Diagonal covariance of mel-Cepstrum parameters and log F0

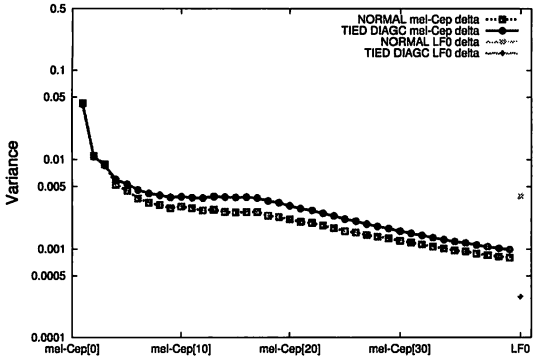

図 8 メルケプストラムパラメータと対数基本周波数の対角共分散 (Δ)

Fig. 8 Diagonal covariance of mel-Cepstrum parameters and log F0 (Δ)

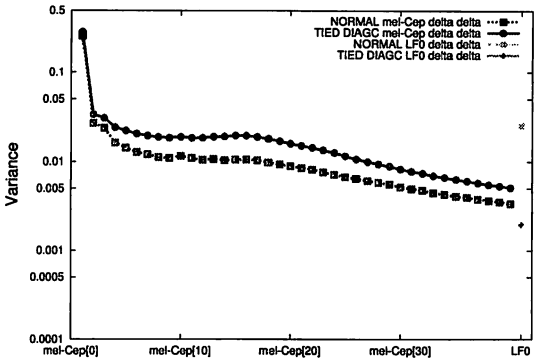

図 9 メルケプストラムパラメータと対数基本周波数の対角共分散 (Δ^2)

Fig. 9 Diagonal covariance of mel-Cepstrum parameters and log F0 (Δ^2)

ム分析による音声のスペクトル推定," 電子情報通信学会論文誌, 75-A, 7, pp. 1124–1134, 1992.

[2] K. F. Lee, "Context-Dependent Phonetic Hidden Markov models for Speaker-Independent Continuous Speech Recognition," IEEE Trans. Acoustic, Speech and Signal Processing, vol. 38, no. 4, pp. 599–609, 1990.

[3] J. Takami and S. Sagayama, "A Successive State Splitting Algorithm for Efficient Allophone Modeling," Proc. ICASSP'92, pp. 573–576, 1992.

[4] M. Y. Hwang, X. Huang, and F. Alleva, "Predicting Unseen Triphones with Senones," Proc. ICASSP'93, pp .311–314, 1993.

[5] J. J. Odell, "The Use of Context in Large Vocabulary Speech Recognition, PhD dissertation," Cambridge University, 1995.

[6] K. Shinoda and T. Watanabe, "MDL-based Context-Dependent Subword Modeling for Speech Recognition," J. Acoust. Soc. Jpn.(E), vol .21, no. 2, pp. 79-8-6, 2000.

[7] W. Chou and W. Reichl, "Decision Tree State Tying based on Panalized Bayesain Information Criterion," Proc. ICASSP'99, pp. 345–348, 1999.

[8] 渡部晋治, 南泰浩, 中村篤, 上田修功, "ベイズ的アプローチに基づく状態共有型 HMM 構造の学習," 信学技報, SP2002-14, pp. 43–48, 2002.

[9] 加藤恒夫, 黒岩眞吾, 清水徹, 樋口宜男, "混合分布 HMM における Tree-based クラスタリング," 信学論 (D-II), vol. J83-D-II, no. 11, pp. 2128–2136, 2000.

[10] H. j. Nock, "Context Clustering for Triphone-based Speech Recognition," Master Thesis, Cambridge University, 1996.

[11] K. Shinoda, T. Watanabe, "MDL-based contextdependent subword modeling for speech recognition," J. Acoust. Soc. Jpn. (E), 21 (2), pp. 79–86, 2000.

[12] H. Kawahara, M. K. Ikuyo, A. Cheneigne, "Restructuring speech representations using a pitch-adaptive timefrequency smoothing and an instantaneousfrequency-based F0 extraction: Possible role of a repetitive structure in sounds," Speech Communication, 27, pp. 187–207, 1999.

[13] H. Zen, T. Masuko, K. Tokuda, T. Kobayashi, T. Kitamura, "A Hidden Semi-Markov Model-Based Speech Synthesis System," IEICE Trans. Inf. & Sys., vol. 90D, no. 5, pp. 825–834, 2007.

[14] T. Toda, K. Tokuda, "Speech parameter generation algorithm considering global variance for HMM-based speech synthesis," Interspeech 2005, pp. 2801–2804, 2005.

[15] J. Yamagishi, M. Tamura, T. Masuko, K. Tokuda, and T. Kobayashi, "A training method of average voice model for HMM-based speech synthesis," IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences, Vol. E86-A, No. 8, pp. 1956–1963, 2003.