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

吉村　貴克[†]　　徳田　恵一[†]　　益子　貴史[††]　　小林　隆夫[††]
北村　正[†]

Simultaneous Modeling of Spectrum, Pitch and Duration in HMM-Based Speech Synthesis

Takayoshi YOSHIMURA[†], Keiichi TOKUDA[†], Takashi MASUKO[††], Takao KOBAYASHI[††], and Tadashi KITAMURA[†]

あらまし　本論文では，HMMに基づく音声合成において，スペクトル，ピッチ，継続長をHMMの枠組みで統一的にモデル化する手法について述べる．本システムでは，スペクトル，ピッチ，継続長モデルとして，それぞれ連続分布HMM，多空間確率分布HMM（MSD-HMM），多次元ガウス分布を用い，音素環境，アクセント，品詞などのコンテクストを考慮したコンテクスト依存モデルを構築する．コンテクスト依存モデルは，決定木に基づくコンテクストクラスタリング手法によりクラスタリングされる．決定木構築の際，節分割はMDL基準により行う．このため，新たにMSD-HMMに対するMDL基準によるコンテクストクラスタリング手法を導出している．音声合成実験において，自然性の高い合成音声が得られること，更に自動学習によりシステムを構築可能であることを確認した．

キーワード　テキスト音声合成，隠れマルコフモデル，コンテクストクラスタリング，MDL基準，自動学習

1. まえがき

現在，テキストから音声を合成する音声合成方式が数多くあるが，自由に話者の声質を変えたり，怒り，悲しみ，喜びといった発話スタイルを表現できるものは少ない．その理由として，現在提案されているほとんどの音声合成システムが音声の素片を選び，つなぎ合わせる方式をとっていることが挙げられる．この方式において多様な話者性，発話スタイルを実現するには，様々な話者，発話スタイルの音声が含まれる音声素片データベースを用意しなければならない．しかし，このような音声素片データベースの構築にはトランスクリプションだけでなく，ラベル境界情報の付与，ピッチマークの付与など，人手による確認，修正の必要な処理に莫(ばく)大な作業を要し，その精度は合成音声の品質に大きくかかわる．

このような背景から我々は，合成時に音声素片データベースを必要とせず，かつ多様な話者性，発話スタイルを容易に実現することができる音声合成システムの構築を目的とし，HMMに基づく音声合成システムを提案してきた[1]．ほかにもHMMを用いた音声合成システムがいくつか提案されているが[2], [3]，それらがHMMを音素等の音声素片を選択する際に用いているのに対し，我々のシステムはHMM自身から音声パラメータを出力し音声を合成する点を特徴としている．そのため本システムには，

（a）　静的及び動的特徴の統計量に基づいて音声パラメータを生成することにより，音声単位の接続ひずみの問題が生じにくい．

（b）　HMMのパラメータを何らかの手法を用いて変換することにより，様々な声質の音声を合成することができる．

（c）　学習データにラベル境界情報がなくとも適切な初期モデルが用意できれば，トランスクリプション

[†] 名古屋工業大学知能情報システム学科，名古屋市
Department of Computer Science, Nagoya Inst. of Tech., Gokiso-cho, Showa-ku, Nagoya-shi, 466–8555 Japan
[††] 東京工業大学大学院総合理工学研究科，横浜市
Interdisciplinary Graduate School of Science and Engineering, Tokyo Inst. of Tech., 4259 Nagatsuta, Midori-ku, Yokohama-shi, 226–8502 Japan

を用いた自動学習により，システムを構築することができる．

などの利点がある．（a）に関しては，スペクトルパラメータ生成において，滑らかで自然性の高い音声スペクトル系列が得られることを確認している [1]．また，（b）に関しては，話者適応 [4]，話者補間 [5] の手法を適用することにより，声質を自由に変換できることを示した．

ただし，上記の文献 [1], [4], [5] のシステムは韻律情報（ピッチ，継続長など）を扱っていなかった．そこで本論文では，話者適応などの手法を用いて音韻モデルのパラメータとともに韻律モデルのパラメータも変換し，多様な音声を合成することを目的とし，スペクトル，ピッチを同時にモデル化する手法 [6] に HMM による継続長のモデル化手法 [7] を新たに定式化し統合することにより，HMM の枠組みでスペクトル，ピッチ，継続長を同時に扱う音声合成システムを構築する．

本システムでは，スペクトル（パワー情報を含む），ピッチ，継続長はそれぞれ，連続分布 HMM，多空間確率分布 HMM（MSD-HMM）[8]，多次元ガウス分布でモデル化する．HMM の特徴ベクトルはスペクトル部，ピッチ部の二つのストリームからなっており，それぞれの音素 HMM は継続長モデルをもつ．これらのモデルは，スペクトル，ピッチ，継続長に影響を与える変動要因（ここではコンテクストと呼ぶ）の考えられるすべての組合せを考慮して作成する．決定木に基づくコンテクストクラスタリング手法 [9] により，コンテクスト依存のスペクトル，ピッチ，継続長モデルをそれぞれ独立にクラスタリングし，ガウス分布を共有化して信頼性を向上させるとともに，学習データには存在しないコンテクストの組合せに対しても対応するモデルを用意できるようにする．本論文では，特にピッチをモデル化する MSD-HMM のためのクラスタリング手法を，ゆう度基準によるもの [6] から，MDL 基準 [10] によるものに拡張し用いている．

以下，**2.** ではスペクトル，ピッチ，継続長の同時モデル化について述べ，**3.** でコンテクストを考慮したモデルの作成について述べる．**4.** では提案する音声合成システムについて説明する．**5.** において音声合成システムの構築とその合成の評価，更に自動学習によるシステムの構築を行う．最後に **6.** で結論を述べる．

2. スペクトル，ピッチ，継続長の同時モデル化

2.1 スペクトル，ピッチのモデル化

本論文ではスペクトルパラメータとして，メルケプストラム係数 [11] を用い，連続分布 HMM によりモデル化する．合成音声は，MLSA フィルタ [12] により，メルケプストラム係数から直接生成することができる．

ピッチパターンは有声区間では連続的な値をとり，無声区間では定数をとるため，連続分布 HMM や離散分布 HMM では直接モデル化することができない．そこで本論文では，ピッチパターンを多空間分布に基づく HMM（MSD-HMM）[13] によりモデル化する．MSD-HMM は，一つの状態に離散分布と連続分布を混在させることができ，0 次元のベクトルと多次元のベクトルが混在する観測ベクトル系列をモデル化することが可能である．本論文では，ピッチを有声区間に対応する 1 次元空間と，無声区間に対応する 0 次元空間の二つの空間から出力される観測事象と考え MSD-HMM によりモデル化する．

スペクトルモデル，ピッチモデルは連結学習により構築される．連結学習は，適切な初期モデルを用いれば，ラベル境界の情報を必要としないため，自動学習を行うのに適している．ただし，スペクトルモデル，ピッチモデルを別々にトレーニングした場合，両者のモデル間で境界のずれが生じる．そこで本論文では，図 1 のようにスペクトル，ピッチの静的特徴量 $\mathbf{c}$, p 及びそれぞれの動的特徴量を結合したものを特徴ベクトルとし，HMM をトレーニングする．

2.2 継続長のモデル化

本論文では，継続長の制御は HMM の状態継続長の制御により行う．状態継続長は多次元ガウス分布でモ

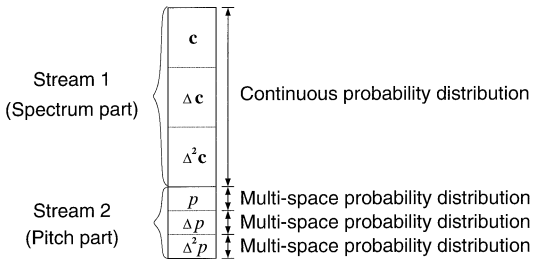


図 1 観測事象
Fig. 1 Observation.

デル化する [7]．ガウス分布の次元は音素 HMM の状態数に等しく，ガウス分布の n 次元目は，HMM の第 n 状態の状態継続長分布に対応している．継続長モデルの学習には HMM に状態継続長分布を含めて行う方法があるが，計算時間が非常にかかり効率が悪い．そこで本論文では，HMM の連結学習の際に各状態の状態滞在確率を使い，状態継続長モデルを構築する．これにより，自動学習による状態継続長モデルの構築が可能となる．

各状態の状態継続長分布の平均 ξ，分散 σ^2 は，学習データを用いた HMM の連結学習の際に作られるトレリス上において，以下のように計算される．

$$\xi(i) = \frac{\sum_{t_0=1}^{T}\sum_{t_1=t_0}^{T}\chi_{t_0,t_1}(i)(t_1-t_0+1)}{\sum_{t_0=1}^{T}\sum_{t_1=t_0}^{T}\chi_{t_0,t_1}(i)} \tag{1}$$

$$\sigma^2(i) = \frac{\sum_{t_0=1}^{T}\sum_{t_1=t_0}^{T}\chi_{t_0,t_1}(i)(t_1-t_0+1)^2}{\sum_{t_0=1}^{T}\sum_{t_1=t_0}^{T}\chi_{t_0,t_1}(i)} - \xi^2(i) \tag{2}$$

ここで，$\chi_{t_0,t_1}(i)$ は，時刻 t_0 から t_1 に，状態 i にいる確率で，

$$\chi_{t_0,t_1}(i) = (1-\gamma_{t_0-1}(i))\cdot\prod_{t=t_0}^{t_1}\gamma_t(i)\cdot(1-\gamma_{t_1+1}(i)) \tag{3}$$

のように計算される．ただし，$\gamma_t(i)$ は 時刻 t に状態 i にいる確率で，$\gamma_{-1}(i)=\gamma_{T+1}(i)=0$ とする．

3. コンテクスト依存モデル

3.1 コンテクストの種類

スペクトル，ピッチ，継続長に影響を与えるコンテクストには，アクセント型，品詞，当該・先行・後続音素など，様々なコンテクストの組合せが考えられる．モデル構築の際にコンテクストを多数用意すれば，より精度の高いモデルが得られると期待できる．本論文では以下のコンテクストを考慮した．

- 文の長さ
- 当該呼気段落の位置
- { 先行，当該，後続 } 呼気段落の長さ
- 当該アクセント句の位置，前後のポーズの有無
- { 先行，当該，後続 } アクセント句の長さ，アクセント型
- { 先行，当該，後続 } の品詞（23 種類），活用形（7 種類），活用型（7 種類）
- 当該音素のアクセント句内でのモーラ位置
- { 先行，当該，後続 } 音素（42 音素）

ここで，長さ，位置の単位はモーラとする．

なお，本論文ではモデルを音素単位で用意し，文頭，文末の無音，文中のポーズも音素として扱っている．

3.2 決定木に基づくコンテクストクラスタリング

考慮するコンテクストの種類が増加するとコンテクストの組合せが指数的に増加するため，モデル当りの学習データが著しく減少し，モデルのパラメータの推定精度が低下する．また，可能なすべてのコンテクストの組合せを網羅する学習データを用意することは現実的には不可能であるため，生成時に学習データ中に存在しないコンテクストの組合せが必要となった場合に，対応するモデルを用意できずパラメータを生成することができなくなる．

この問題を解決するために，決定木を用いたコンテクストクラスタリング [9] を，コンテクスト依存のスペクトル，ピッチ，継続長モデルに適用する．決定木は 2 分木であり，それぞれの節（node）ごとにコンテクストを二つに分割する質問が用意されている．すべてのコンテクストは根（root node）からそれぞれの節の質問に従って木を下っていき，葉（leaf）のうちのどれかに達するため，いったん決定木を構築すれば，学習データに出現しないコンテクストの組合せに対しても対応するモデル（クラスタ）が一意に決定される．スペクトル，ピッチ，継続長はそれぞれ，影響を受けるコンテクストの種類が異なると考えられるため，本論文では図 2 のようにそれぞれ独立にクラスタリングを行う．また，このコンテクストクラスタリング手法は，文献 [6] において MSD-HMM に対し拡張されているが，本論文では更に，決定木構築における分割する節及び質問の選択に MDL 基準 [10] を用いる．MSD-HMM に対する MDL 基準の具体的な数式は次節で導出する．

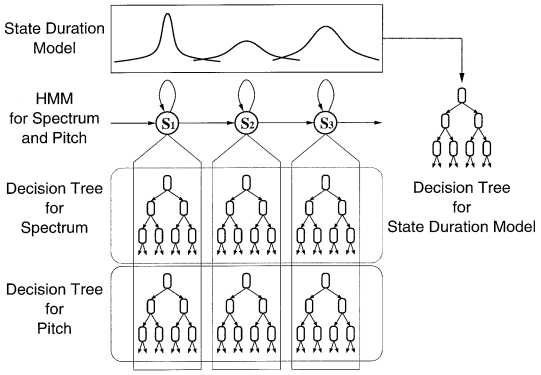


図2 決 定 木
Fig. 2 Decision trees.

3.3 MDL 基準による MSD-HMM のコンテクストクラスタリング

本節では，MSD-HMM に対する MDL 基準によるコンテクストクラスタリング手法を導出する．

クラスタリングによりクラスタのセット $\mathcal{S}$ が以下のようになったとする．

$$\mathcal{S} = \{S_1, S_2, \cdots, S_i, \cdots, S_M\} \tag{4}$$

このとき対数ゆう度 $\mathcal{L}$ は

$$\mathcal{L} = -\sum_{s\in\mathcal{S}}\sum_{g=1}^{G}\frac{1}{2}(n_g(\log(2\pi)+1)+\log|\boldsymbol{\Sigma}_{sg}| - 2\log w_{sg})\sum_{t\in T(\boldsymbol{O},g)}\gamma_t(s,g) \tag{5}$$

と表される．ここで g は空間インデックス，w_{sg} はクラスタ s における空間 g の空間重み，$T(\boldsymbol{O},g)$ は観測事象 $\boldsymbol{o_t}$ の 空間インデックス集合が空間インデックス g を含むような 時刻 t の集合であり，$\gamma_t(s,g)$ は時刻 t においてクラスタ s の空間 g をとる確率である．また，式 (5) において 0 次元の空間の場合，$\log|\boldsymbol{\Sigma}_{sg}| = 0$ とする．この $\mathcal{L}$ を用いて記述長 l を表すと，

$$l = \sum_{s\in\mathcal{S}}\sum_{g=1}^{G}\frac{1}{2}(n_g(\log(2\pi)+1)+\log|\boldsymbol{\Sigma}_{sg}| - 2\log w_{sg})\sum_{t\in T(\boldsymbol{O},g)}\gamma_t(s,g) + \left(\sum_{s\in\mathcal{S}}\sum_{g=1}^{G}\frac{1}{2}(2n_g+1)\right)\cdot\left(\log\sum_{s\in\mathcal{S}}\sum_{g=1}^{G}\sum_{t\in T(\boldsymbol{O},g)}\gamma_t(s,g)\right) \tag{6}$$

となる．クラスタ S_i が S_{i+}, S_{i-} に分かれたときの記述長を l' とすると，記述長の変化量 δl は，

$$\begin{aligned}\delta l &= l' - l \\ &= \sum_{s\in\{S_{i+},S_{i-}\}}\sum_{g=1}^{G}\frac{1}{2}(\log|\boldsymbol{\Sigma}_{sg}| - 2\log w_{sg})\cdot\sum_{t\in T(\boldsymbol{O},g)}\gamma_t(s,g) \\ &\quad - \sum_{s\in\{S_i\}}\sum_{g=1}^{G}\frac{1}{2}(\log|\boldsymbol{\Sigma}_{sg}| - 2\log w_{sg})\cdot\sum_{t\in T(\boldsymbol{O},g)}\gamma_t(s,g) \\ &\quad + \left(\sum_{g=1}^{G}\frac{1}{2}(2n_g+1)\right)\cdot\left(\log\sum_{s\in\mathcal{S}}\sum_{g=1}^{G}\sum_{t\in T(\boldsymbol{O},g)}\gamma_t(s,g)\right)\end{aligned} \tag{7}$$

となる．記述長の減少が最も大きくなるようなクラスタと質問の組合せを選び，クラスタの分割を行う．すべてのクラスタと質問の組合せで記述長が増加する場合に，クラスタの分割を停止する．ゆう度基準の場合と異なり，学習データ量に応じた分割が行われることが期待され，また，ゆう度の変化量，節の最小学習データ数などの分割停止基準を設定あるいは調整する必要がないという利点もある．

4. 音声合成システム

音声合成システムの合成部のブロック図を図 3 に示す．合成時には，まず，合成したい任意のテキストをコンテクストに基づいたラベル列に変換する．このラベル列に従ってコンテクスト依存 HMM を結合し，一つの文 HMM を構成する．状態継続長分布に従って各状態の継続長を決定し，文献 [14] に示されるゆう度最大化基準に基づくパラメータ生成アルゴリズムにより，文 HMM からメルケプストラム列を出力し，ピッチパターンも有声/無声を考慮し同様の手法で出力する．最後に，生成したメルケプストラム列とピッチパターンを用いて，MLSA フィルタ [12] により音声を合成する．

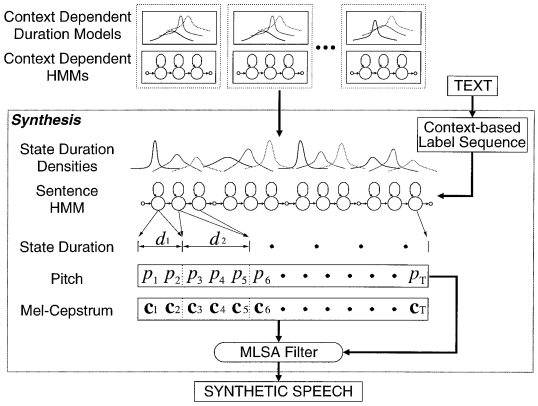


図 3　音声合成システムの合成部
Fig. 3　Synthesis part of the system.

5.　音声合成システムの構築

HMM の学習データとして，ATR 日本語音声データベースの男性話者 MHT による音韻バランス文 503 文のうち 450 文を学習データとして用いた．サンプリング周波数は 16 kHz，分析周期は 5 ms とした．25 ms 長ブラックマン窓を用いてメルケプストラム分析を行い，パワー情報を含む 0〜24 次のメルケプストラム係数を求めた(注 1)．特徴ベクトルは，メルケプストラム係数及びそのデルタ，デルタデルタとピッチ及びそのデルタ，デルタデルタを合わせた全 78 次元のベクトルとした．使用した HMM は，対角共分散単一ガウス分布をもつ 5 状態 left-to-right モデルであり，継続長モデルは 5 次元ガウス分布である．

5.1　決　定　木

スペクトル，ピッチモデルは音素内での状態位置ごとにコンテクストクラスタリングを行って，それぞれ五つの決定木を作成し，継続長モデルは全体で一つの決定木を作成した．クラスタリングの結果，スペクトルモデル，ピッチモデル，継続長モデルの総分布数はそれぞれ 943, 2452, 579 となった．

図 4 は，コンテクストクラスタリングの際に自動生成されたスペクトル，ピッチ，継続長モデルの決定木の例である．図中の “L_*”，“C_*”，“R_*” はそれぞれ先行，当該，後続環境を意味し，“*_breath_*”，“*_accent_*”，はそれぞれ呼気段落，アクセント句に関する質問を意味している．“silence” は，文頭，文末の無音，文中のポーズ，“pit_s2_*” と “dur_s2_*” は決定木の葉を表している．また，“C_mora <= 1” は，当該音素のモーラ位置が当該アクセント句内で 1 以下であることを表している．これらの図から，スペクトルモデルに関しては，音素に関する質問がはじめに多く適用されており，音素環境の影響を強く受けているのがわかる．また，ピッチモデルに関しては，最初に有声か無声かで分かれ，状態継続長分布に関しては，最初に無音か無音でないかで分かれており，自動生成された決定木がそれぞれの特徴をよくとらえているのがわかる．

5.2　動的特徴量の効果

HMM からスペクトル系列，ピッチパターンを出力する際，(a) 静的特徴量のみ，(b) デルタパラメータを考慮したもの，(c) 更にデルタデルタパラメータを考慮したものの 3 通りの出力を行った．システムから生成された合成音声（「平均倍率を下げた形跡がある」）のスペクトルの一部とピッチの生成例をそれぞれ図 5, 図 6 に示す．文章は学習データには含まれないものを用いている．

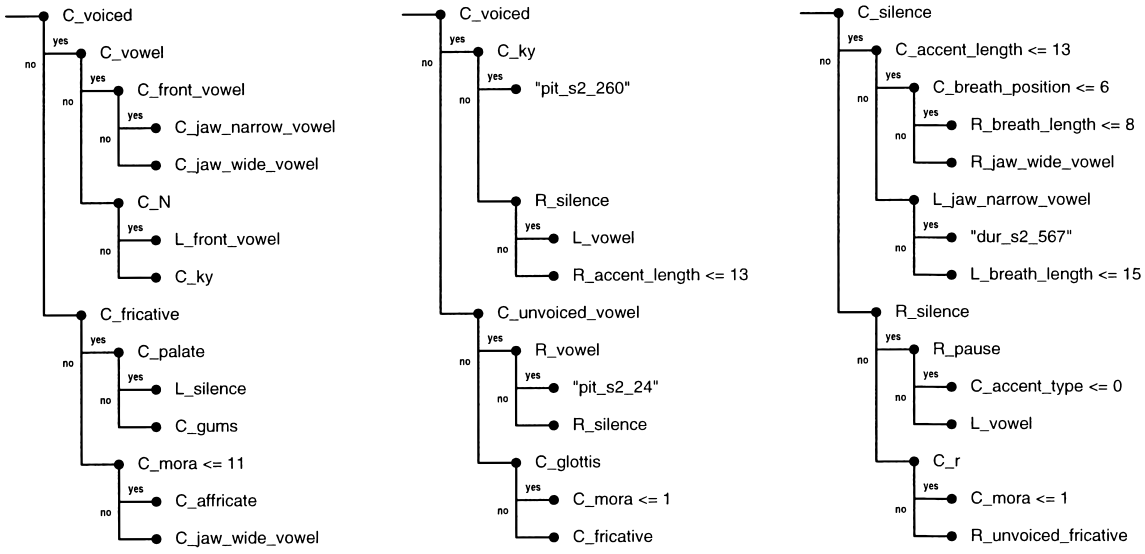
図 5 より，スペクトルに関しては動的特徴量を考慮することにより，滑らかに変化するスペクトル系列が得られていることがわかる．また，図 6 より，ピッチパターンに関しても動的特徴量を考慮することにより，滑らかなピッチパターンが得られ，動的特徴量がピッチパターンの生成においても大きな効果をもつことが確認された．

また，合成音声の品質に及ぼす動的特徴の効果を主観評価実験により確かめた．被験者 8 人にそれぞれ，学習データに含まれない 53 文章の中からランダムに選んだ 6 文章を聞かせた．プリファレンススコアを図 7 に示す．この図より，スペクトルのみならず，ピッチについても，静的特徴量のみでなく，動的特徴量をも考慮した方が合成音声の品質が向上し，更にスペクトル，ピッチ両方について動的特徴量を考慮することにより，合成音声の品質が大きく向上していることがわかる．

5.3　自動学習によるシステムの構築

初期モデルとして不特定話者性別依存モデルを用いた自動学習により男性話者 MHT の音声合成システムの構築を行った．ATR 日本語音声データベースの話者 MHT を除く男性話者 5 人による音韻バランス文 450 文を用いて不特定話者性別依存 triphone モデルを学

(注 1)：http://kt−lab.ics.nitech.ac.jp/~tokuda/SPTK/ にメルケプストラム分析・合成のソースコードがある．



(a) Tree for Spectrum Model (1st state)

(b) Tree for Pitch Model (1st state)

(c) Tree for State Duration Model

図 4　決定木の例
Fig. 4　Examples of decision trees.

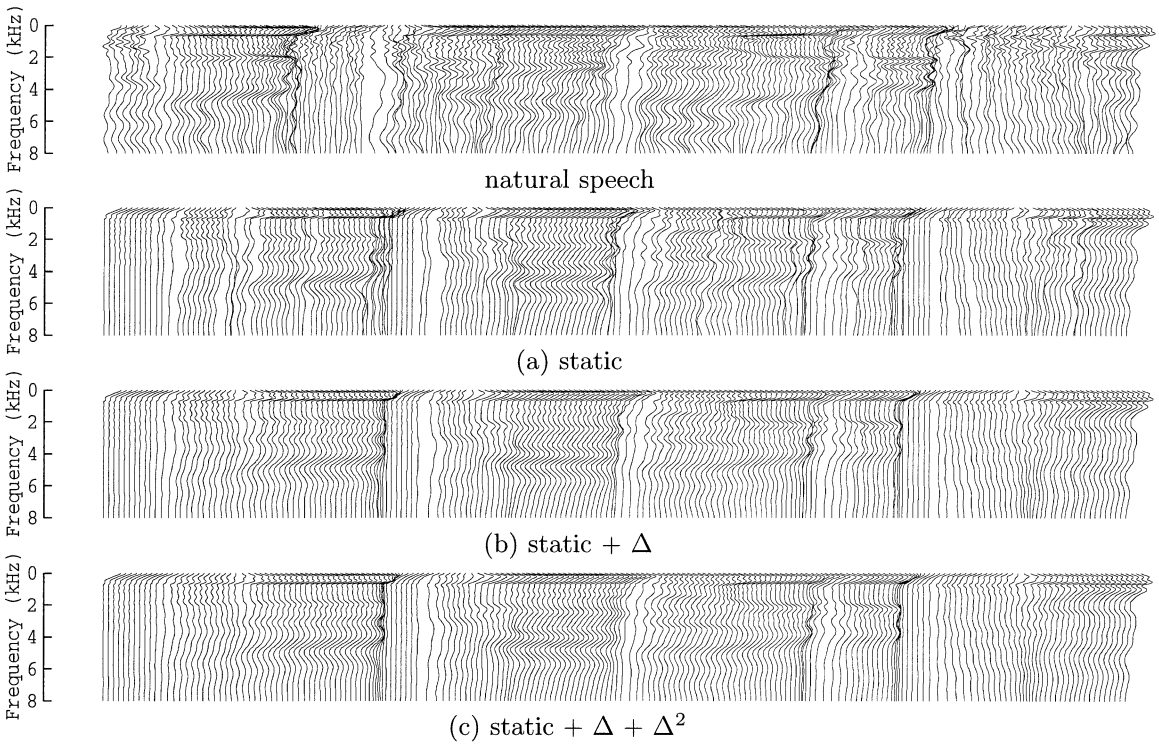


図 5　生成されたスペクトル「平均倍率」
Fig. 5　Generated spectra for a phrase “heikiNbairitsu.”

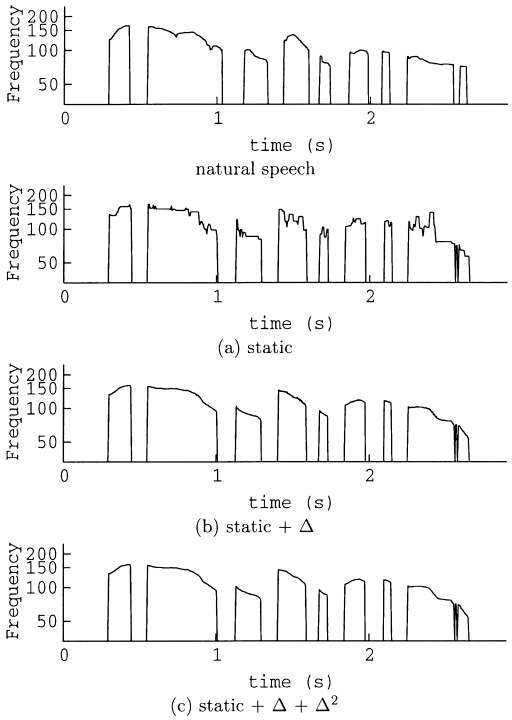


図 6　生成されたピッチパターン
「平均倍率を下げた形跡がある」
Fig. 6　Generated pitch pattern for a sentence
“heikiNbairitsuwo sageta keisekiga aru.”

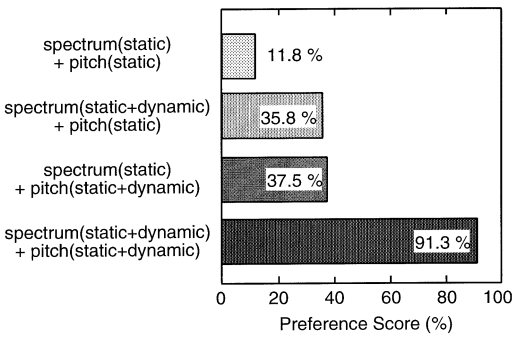


図 7　動的特徴量の効果
Fig. 7　Effect of dynamic feature.

習した．ただしピッチは，音素環境より，構文情報の影響を強く受けると考え，初期モデルにおけるピッチパラメータの学習は行っていない．

システムの自動学習は以下の手順で行った．

（1）　不特定話者性別依存 triphoneHMM を，コンテクスト依存 HMM にコピーする．

（2）　ターゲット話者 MHT の学習データにより，コンテクスト依存 HMM の連結学習を行う．

（3）　コンテクストクラスタリングを行う．

（4）　再び連結学習を行う．同時に継続長モデルを構築する．

（5）　継続長モデルのコンテクストクラスタリングを行う．

構築されたシステムを用いて音声を合成し，非公式な受聴試験を行った結果，ラベル境界情報を用いてシステムを構築した場合とほぼ同品質の合成音声が得られることを確認した(注 2)．

6. むすび

本論文では，スペクトル，ピッチ，継続長を HMM の枠組みで同時に扱う音声合成システムの構築を行った．動的特徴量を考慮することにより，滑らかで自然性の高い音声スペクトル系列，ピッチパターンが得られ，学習データ発話者の個人性をよく再現した自然な音声が得られることを確認した．その際，スペクトルと同様，ピッチについても動的特徴量を考慮することにより，主観品質が大きく改善されることが明らかになった．更に，初期モデルとして不特定話者性別依存 triphone モデルを用い，自動学習により音声合成システムの構築が可能であることを確認した．

なお，提案した音声合成システムにはまだ改善の余地がある．例えば，本システムにおいて音源は単純に，有声区間に関してはパルス列，無声区間に関しては白色雑音を用いているが，音源情報も HMM の枠組みでモデル化し，音声符号化の技術である MELP [15] などで用いられている手法と同様の音源生成手法を導入することにより，音質は更に向上すると考えられる（連続値，離散値を含む音源情報は MSD-HMM によりモデル化可能であることに注意する）．また本論文では，パワーはスペクトルパラメータに含めて学習しているが，これを分離し，パワーも独立にモデル化すれば，合成音声の品質の改善が見込まれる．今後の課題としては，これらの改善を行うことと，その合成音声の評価，更に話者適応，話者補間の手法を適用して，各モデルのパラメータを変換することにより，様々な話者の声質の合成音声を生成することが挙げられる．また，本システムでは話者ごとに HMM を学習しているが，これをある発話スタイル（例えば，喜んでいる

(注 2)：http://kt−lab.ics.nitech.ac.jp/˜yossie/TTS/ に最新の合成音声のサンプルがある．

ような話し方）で発声された音声で HMM を学習し，この HMM を話者適応の手法を用いて別の発話スタイル（例えば，怒っているような話し方）に対応した HMM に変換できれば，その HMM から音声パラメータを生成し，様々な発話スタイルの音声を合成できると期待している．

謝辞 本研究の一部は，文部省科学研究費補助金（基盤研究 B(2) 課題番号 10555125），(財) 中部電力基礎技術研究所研究助成金によった．

文　献

[1] 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.

[2] R.E. Donovan and E.M. Eide, “The IBM trainable speech synthesis system,” Proc. ICSLP, vol.5, pp.1703–1706, Nov. 1998.

[3] M. Plumpe, A. Acero, H. Hon, and X. Huang, “HMM-based smoothing for concatenative speech synthesis,” Proc. ICSLP, vol.6, pp.2751–2754, Nov. 1998.

[4] M. Tamura, T. Masuko, K. Tokuda, and T. Kobayashi, “Speaker adaptation for HMM-based speech synthesis system using MLLR,” Proc. Third ESCA/COCOSDA workshop on Speech Synthesis, pp.273–276, Dec. 1998.

[5] T. Yoshimura, K. Tokuda, T, Masuko, T. Kobayashi, and T. Kitamura, “Speaker interpolation in HMM-based speech synthesis system,”’ Proc. EUROSPEECH, vol.5, pp.2523–2526, Sept. 1997.

[6] 宮崎　昇，徳田恵一，益子貴史，小林隆夫，“多空間上の確率分布を用いた HMM によるピッチパタン生成の検討，” 信学技報，SP98-12, April 1998.

[7] 吉村貴克，徳田恵一，益子貴史，小林隆夫，北村　正，“HMM に基づく音声合成のための状態継続長モデルの構築，” 信学技報，DSP98-85, Sept. 1998.

[8] 宮崎　昇，徳田恵一，益子貴史，小林隆夫，“多空間上の確率分布に基づいた HMM とピッチパタンモデリングへの応用，” 信学技報，SP98-12, April 1998.

[9] J.J. Odell, The Use of Context in Large Vocabulary Speech Recognition, Ph.D. dissertation, Cambridge University, 1995.

[10] K. Shinoda and T. Watanabe, “Speaker adaptation with autonomous model complexity control by MDL principle,” Proc. ICASSP, pp.717–720, May 1996.

[11] 徳田恵一，小林隆夫，深田俊明，斎藤博徳，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），vol.J74-A, no.8, pp.1240–1248, Aug. 1991.

[12] 今井　聖，住田一男，古市千枝子，“音声合成のためのメル対数スペクトル近似（MLSA）フィルタ，” 信学論（A），vol.J66-A, no.2, pp.122–129, Feb. 1983.

[13] K. Tokuda, T. Masuko, N. Miyazaki, and T. Kobayashi, “Hidden Markov models based on multi-space probability distribution for pitch pattern modeling,” Proc. ICASSP, pp.229–232, May 1999.

[14] 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，” 音響誌，vol.53, no.3 pp.192–200, March 1997.

[15] A.V. McCree and T.P. Barnwell III, “A mixed excitation LPC vocoder model for low bit rate speech coding,” IEEE Trans. Speech and Audio Processing, vol.3, no.4, pp.242–250, July 1995.

（平成 12 年 2 月 29 日受付，5 月 23 日再受付）

吉村　貴克

平 11 名工大大学院博士前期課程了．現在，同大大学院博士後期課在学中．音声合成の研究に従事．日本音響学会会員．

徳田　恵一　（正員）

昭 59 名工大・工・電子卒．平 1 東工大大学院博士課程了．同年東工大電気電子工学科助手．平 8 名工大知能情報システム学科助教授．工博．音声分析，音声合成・符号化，音声認識，ディジタル信号処理の研究に従事．日本音響学会，人工知能学会，情報処理学会，IEEE 各会員．

益子　貴史　（正員）

平 5 東工大・工・情報卒．平 7 同大大学院博士前期課程了（知能科学専攻）．同年同大精密工学研究所助手．現在，同大大学院総合理工学研究科物理情報システム創造専攻助手．音声の分析・合成，音声認識の研究に従事．日本音響学会，IEEE, ISCA 各会員．

小林　隆夫　（正員）

昭 52 東工大・工・電気卒．昭 57 同大大学院博士課程了．同年同大精密工学研究所助手．工博．現在，同大大学院総合理工学研究科物理情報システム創造専攻教授．ディジタルフィルタ，音声の分析・合成，音声認識の研究に従事．日本音響学会，IEEE, ISCA 各会員．

北村　　正　（正員）

昭 48 名工大・工・電子卒．昭 53 東工大大学院博士課程了．同年東工大精密工学研究所助手．昭 58 名工大・工・電子工学科講師．昭 59 同助教授．平 7 名工大知能情報システム学科教授．工博．音声情報処理，マルチメディア情報処理の研究に従事．日本音響学会，IEEE，ISCA 各会員．