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

2008－MUS－74 (7)
2008－SLP－70 (7)
2008／2／8

# 声質と歌唱スタイルを自動学習可能な歌声合成システム

酒向慎司† 才野慶二郎† 南角吉彦†
徳田恵一† 北村正†

声質や歌唱法など歌い手の特徴を歌声データと楽譜から自動学習し，それらを再現するような歌声合成システムについて述べる．本システムでは，歌い手の声質とピッチに関する特徴を確率モデルによる統一的な枠組みでモデル化している．特に，リズムやメロディといった音楽特有の表現要素が，音声信号のスペクトルや基本周波数パターンの変動に大きく関係していることから，楽譜から得られる音階や音長などを考慮したモデル化を行い，楽譜と歌詞を入力として，個人性を備えた歌声を合成するシステムを構築してきた．本手法の特徴は，このような歌声合成モデルを楽譜と歌声データから自動学習できることにある．本報告では，音楽固有のコンテキストの導入，実際の歌声データと楽譜の音符列の間のずれに着目した時間構造モデルについて検討する．実験では，童謡60曲の男性1名の歌声データを用いた歌声合成システムを構成し，ずれモデルの導入による自然性の向上が確認できた．

# A trainable singing voice synthesis system capable of representing personal characteristics and singing styles

Shinji Sako,† Keijiro Saino,† Yoshihiko Nankaku,† Keiichi Tokuda†
and Tadashi Kitamura†

We describe a trainable singing voice synthesis system, that can automatically learns the model parameters from singing voice waveform and musical scores by applying HMM-based speech synthesis technique. In this system, a sequence of spectrum and fundamental freqency ($F0$) are modeled simultaneously in a unified framework of HMM, and context dependent HMMs are constructed by taking account of contextual factors that affects singing voice. In addition, the distributions for spectral and $F0$ parameter are clustered independently by using a decision-tree based context clustering technique. Synthetic singing voice is generated from HMMs themselves by using parameter generation algorithm. We introduced an additional "time-lag" model to control start timing of each musical note. In the experiments, we confirmed that smooth and natural-sounding singing voice is synthesized. It is also maintains the characteristics and personality of the donor of the singing voice data for HMM training.

## 1. はじめに

我々はこれまでに，データ提供者の声質や歌唱スタイルの再現が可能な歌声合成手法を提案してきた[1),2)]．特定の歌い手や歌唱スタイルなどの表現を可能とする歌声合成技術は，音楽制作の支援の他，エンタテイメントやアミューズメント分野への応用を考えることができ，同様の試みは，これまでにも幾つか提案されている[3)-5)]．

このような歌声合成システムの基本要素として，文字列から人間の声を合成するテキスト音声合成がある．我々の研究グループで開発してきた隠れマルコフモデル (Hidden Markov model; HMM) に基づいた音声合成手法[6)]を利用することで，多様な表現力を備えた歌声合成が可能となると考えている．

現在主流となっている音声合成システムの多くは単位選択という方式に分類される．これは音素や音韻といった音声単位ごとに分類した波形データを，合成したいテキストに従って接続する手法である．実際に発声された音声波形を利用するため，クリアな合成音を得やすいというメリットがある一方，接続部分の歪みが生じやすく，また，多様な声質や発声のスタイルなどを制御する場合，時間領域での波形操作は難しいといえる．

これに対して，HMM に基づいた音声合成手法は，音声認識の分野で広く使われている HMM を用い，実音声から音声の生成モデルを統計的に学習するアプローチである．この手法では，動的特徴量を考慮した

† 名古屋工業大学 大学院工学研究科
Department of Computer Science and Engineering, Nagoya Institute of Technology

パラメータ生成アルゴリズム[7]によって，接続歪みのない滑らかに変化する音声パラメータが得られるほか，モデルパラメータの変換により，別の話者への適応や，多様な声質や感情を表現した音声の合成に柔軟に対応できるなどの発展性がある[8)-10)]．

このような枠組みを利用した歌声合成システムの特徴は，合成システムにおけるすべてのモデルパラメータを学習データから自働学習できる点にある．つまり，歌声の波形データとその曲の楽譜情報を元に，その歌い手の特徴をモデルとして獲得し，合成時にそれらの特徴を再現するような歌声合成システムを自動的に構成することができる．

歌声のピッチや長さは楽譜のメロディやテンポに従うものであり，ピッチの時間変化やリズムの時間構造を楽譜から一意に定めることもできるが，そこから合成される歌声は単調で機械的なものになり，歌声としての魅力に欠けるものである．実際の歌声には，楽譜通りの画一化されたものだけではなく，声質のほかに声の高さやそれらの時間的な構造の変化により，それぞれの歌い手独自のスタイルが存在している．そこで，自然の歌声にあるような音の高さの変化を再現するため，スペクトル情報とピッチ情報の時系列変化をモデル化し，さらに楽譜情報を考慮することで，モデルの高精度化を図る．

歌声は，通常の会話やテキストの読み上げなどの場合と比較して，発声する音の高さや時間的な長さ，または声の強弱などの変動の様子が大きく異なる．本手法では，楽譜から得られる音高や音長といった情報を「コンテキスト」として考え，それらのコンテキストの組み合わせを考慮したモデルの分類を行う．これらのモデルに決定木によるコンテキストクラスタリング[11)]を適用することにより，限られた学習データにおいても，汎用性の高い合成モデルを得ることが可能である．このようにして学習データに基づいて構成される歌声モデルは，楽譜上では表現できないスペクトルやピッチの変動などの歌い手の持つ様々な特徴を備え，合成時には入力として与えられた歌唱曲に応じて，それらの特徴を再現することが可能となる．

本稿では，**図1**のように楽譜が定める音符の境界と実際に歌われた歌声データの時間境界の揺らぎに着目した．このような揺らぎにも歌声の個人性が含まれていると考え，学習データにおける発声のタイミングと楽譜とのずれを確率モデルとして定式化し，歌声合成システムの時間構造モデルに導入する．実験では，60曲の童謡からなる男性1名の歌声データベースを用いて歌声合成モデルを学習し，主観評価試験によって自然性の向上を確認する．

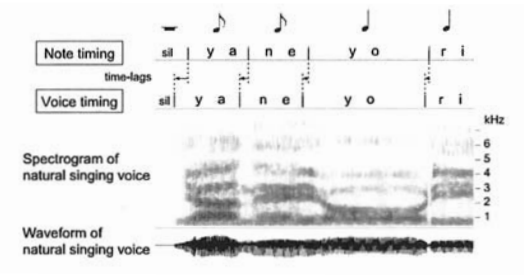


**図1** 歌声信号と音符列とのずれ
Fig. 1 An example of "time-lag"

以下，2節ではHMMに基づいた歌声合成システム，3節で実験及び合成された歌声の評価と考察，最後に4節でまとめと今後の展望について述べる．

## 2. HMMに基づいた歌声合成システム

本研究で提案する歌声合成システムの概略を**図2**に示す．このシステムは，学習部と合成部の2つのパートから構成される．学習部では，楽譜情報と歌声の波形データからなる歌声データベースを用いて歌声合成モデルを学習する．合成部では，合成したい歌唱曲の情報として歌詞とメロディに相当する情報を与えることで，歌声を合成する．なお，各部ではMIDIを楽譜情報として利用している．

### 2.1 個人性を備えた歌声のモデル化

歌声を含む人の声の基本構成は，声帯の振動による音源の生成と声道形状の変化による調音，口唇からの放射として考えることができる．このようなモデルにおいて，音声信号から音源と声道特性を分析抽出し，そこから元の信号に再合成する分析合成技術は，音声信号処理の基礎技術になっている．

本システムの歌声合成モデルは，あるメロディーにそった歌詞を発声する際，声帯の振動や声道特性の歌声の特徴パラメータがどのような時間変化をしながら発声されるか，という生成モデルに相当する．本節では，実際の歌声データから得られる特徴パラメータを確率モデルであるHMMによって学習する手法について述べる．

音声合成モデルは，歌声データから求めたスペクトル，基本周波数パラメータ，およびその時間構造を音素単位でモデル化したものである．歌声音声のスペクトルパラメータは，連続HMMによってモデル化することができるが，基本周波数は有声区間では連続値を

とり，無声区間では値を持たない可変次元の時間系列信号であるため，通常の連続 HMM や離散 HMM で直接モデル化することはできない．そこで，可変次元に対応した多空間上の確率分布に基づく HMM (Multi-Space probability distribution HMM; MSD-HMM)[12] を用い，スペクトルパラメータとしてメルケプストラム[13] を多次元ガウス分布，基本周波数の有声音を 1 次元空間，無声音を 0 次元空間のガウス分布として単一の枠組みの中で同時にモデル化する．

また，音素は音韻的な分類の基本要素であるが，そのスペクトルや *F*0 の形状は様々な要因によって変動を伴う．例えば，ある程度の規則性を持って音素の繋がりによって母音が無声化することが知られている．歌声の音声信号でも，広い範囲では歌唱スタイルやテンポなど，局所的には前後の歌詞や音階・音長などによって，それらの特徴が異なってくると考えられる．テキスト音声合成においても，テキストから得られる言語的な情報が音声パラメータに影響を与えていると考え，それらの要因をコンテキストと呼び，コンテキストを考慮したモデル化が行われている．音楽では，音楽固有のコンテキストを考慮したモデル化が必要である．

また，コンテキストの導入により詳細なモデル分類がなされるが，コンテキストの種類に応じてその組み合わせの数も莫大となってしまう．すべてのコンテキストの組み合わせに対応したモデルを学習するためにはあらゆるパターンを網羅したデータベースが必要となるが，これは現実的ではない．

この問題に対する解決法として，クラスタリングによって類似したモデル間でパラメーターを共有させる手法がある[14]．これは，二分木を用いて，モデルの集合を木構造に分割することで，類似したコンテキストの組み合わせごとにモデルパラメータをクラスタリングする手法である．木の各ノードには，コンテキストを二分する質問があり，各リーフノードには，特定のモデルに相当するモデルパラメータがある．任意のコンテキストの組み合わせは，ノードにある質問に沿って木を辿ることで何らかのリーフノードに到達でき，該当するモデルを選択することができる．

### 2.2 学 習 部

まず，初期モデルとして，テキスト読み上げ文による音声データベースから音素単位のモデルを作成する．次に歌声データベースを用いて楽譜情報を考慮した学習を行う．ここでは，歌声のモデル化に効果的なコンテキストとして以下にあるものを考え，それぞれ当該

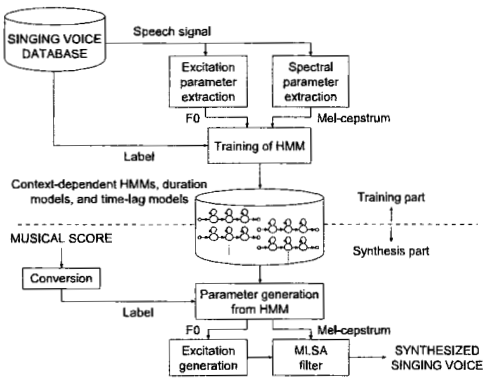


**図 2** 歌声合成システムの概要

**Fig. 2 The overview of the HMM-based singing voice synthesis system**

および前後の環境に依存した歌声モデルを学習する．さらに，これらのコンテキストに基づいて MDL 基準を用いた決定木によるコンテキストクラスタリングを行う．

- **歌詞**：音素名や母音・子音・有声音など音素の分類
- **音高**：各音符の MIDI 音階値の高低の分類
- **音長**：各音符の長さを 100ms 単位で表したもの
- **小節内位置**：小節内における音符の位置情報

なお，状態継続長モデルは，音素に対応する各歌声モデル内部の時間構造を表し，各 HMM の状態遷移回数を多次元ガウス分布でモデル化したものである．このモデルパラメータは HMM の連結学習時に作られるトレリス上で求める[15]．状態継続長モデルに関しても同様にコンテキストクラスタリングを適用する．

### 2.3 合 成 部

合成部では，楽譜情報（歌詞を付与した MIDI データ）を入力として歌声を合成する．まず，楽譜から得られる，歌詞，音高，音長情報に基づいて，歌声モデルから対応するモデルを選択する．次に，楽譜から与えられた各音符の長さを制約として，音符内の音素継続長及び音素内部の状態継続長を，各モデルの状態継続長分布に基づいた尤度最大化基準により決定する．得られた状態系列から，パラメータ生成アルゴリズムによってメルケプストラムと基本周波数パラメータの列を生成する[7]．最後に，生成されたパラメータに基づいて MLSA フィルタを励振させることで，歌声を合成する[16]．

### 2.4 ずれモデルの導入

従来手法では，歌声の時間構造を決定するモデルとして音素の状態継続長を多次元ガウス分布によって構成していた．合成時には楽譜から音符単位の長さを定

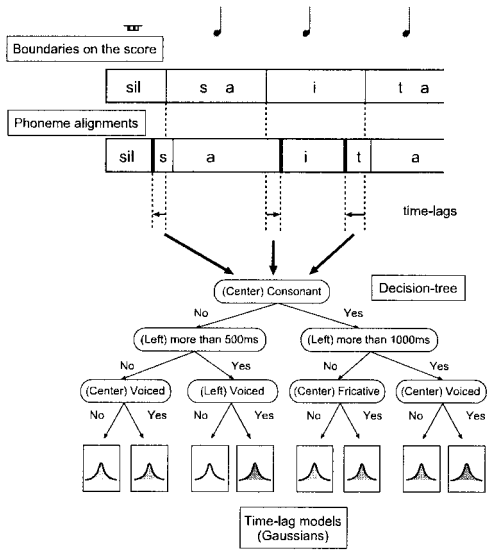


図 3　ずれモデル
Fig. 3　"time-lag" model

め，その区間内で対応する歌詞に含まれる音素の時間長を，状態継続長モデルを用いて決定していた．しかし**図 3** が示すように，歌声データの音素境界は楽譜が定める音符の境界とは一致しないことが多い．このような時間的な揺らぎは，歌声に限らず楽器演奏にも見られ，音楽表現という観点からは本質的なものといえる．そこで，歌声におけるこのような時間変動は，発声する際の音韻の違いや音高やリズムなど，様々な影響をうけた偏りがあると仮定し，学習データにおける発声のタイミングと楽譜とのずれをモデル化する枠組みを導入する．

音符単位のずれモデルとして，音符単位に見た歌声と楽譜のずれを 1 次元ガウス分布によって表し，他のスペクトルや基本周波数モデルと同様にして，コンテキスト依存モデルとして扱うことができる．

歌声合成の際には，まず楽譜の表す時間境界を定めた後，音符単位のずれモデルと音素状態継続長モデルから，以下のような両者の同時確率を最大化することで，学習データにおける音符の揺らぎを考慮した時間構造を決定することができる．

$$P(\boldsymbol{d}, \boldsymbol{g}|\boldsymbol{T}, \Lambda) = P(\boldsymbol{d}|\boldsymbol{g}, \boldsymbol{T}, \Lambda) \cdot P(\boldsymbol{g}|\Lambda) \tag{1}$$

$$= \prod_{k=1}^{N} P(\boldsymbol{d}_k|T_k, g_k, g_{k-1}, \Lambda) \cdot P(g_k|\Lambda) \tag{2}$$

$N$ は合成したい曲の音符数，$\boldsymbol{d}_k$ は $k$ 番目の音符の各状態継続長，$T_k$ は，楽譜から定められる $k$ 番目の音符長，$g_k$ は $k+1$ 番目の音符の開始時刻のずれを表わす．なお，最初と最後の音符に該当するため $g_0 = g_N = 0$ とする．式 1 を最大化する $\boldsymbol{g}$ と $\boldsymbol{d}$ は，以下の線型方程式を解くことで得られる．

$$\boldsymbol{A}\boldsymbol{g} = \boldsymbol{b} \tag{3}$$

$$\boldsymbol{d}_k = \boldsymbol{\mu}_{d_k} + \rho_k \cdot \mathrm{diag}^{-1}\left(\boldsymbol{\Sigma}_{d_k}\right) \tag{4}$$

ここで，

$$a_{i,i} = -1 - \frac{\sum_{t=1}^{K \cdot n_{(i+1)}} \sigma^2_{d_{(i+1),t}}}{\sum_{t=1}^{K \cdot n_i} \sigma^2_{d_{i,t}}} - \frac{\sum_{t=1}^{K \cdot n_{(i+1)}} \sigma^2_{d_{(i+1),t}}}{\sigma^2_{g_i}} \tag{5}$$

$$a_{(i+1),i} = \frac{\sum_{t=1}^{K \cdot n_{(i+1)}} \sigma^2_{d_{(i+1),t}}}{\sum_{t=1}^{K \cdot n_i} \sigma^2_{d_{i,t}}} \tag{6}$$

$$a_{i,(i+1)} = 1 \tag{7}$$

$$a_{i,j} = 0 \quad (j \neq i \pm 1, j \neq i) \tag{8}$$

$$b_i = \frac{\sum_{t=1}^{K \cdot n_{(i+1)}} \sigma^2_{d_{(i+1),t}}}{\sum_{t=1}^{K \cdot n_i} \sigma^2_{d_{i,t}}} \left(T_i - \sum_{t=1}^{K \cdot n_i} \mu_{d_{i,t}}\right) - \left(T_{i+i} - \sum_{t=1}^{K \cdot n_{i+1}} \mu_{d_{(i+1),t}}\right) - \frac{\sum_{t=1}^{K \cdot n_{(i+1)}} \sigma^2_{d_{(i+1),t}}}{\sigma^2_{g_i}} \mu_{g_i} \tag{9}$$

$$(1 \leq i \leq N-1, 1 \leq j \leq N-1)$$

$$\rho_k = \frac{(T_k - g_{k-1} + g_k) - \sum_{t=1}^{K \cdot n_k} \mu^2_{d_{k,t}}}{\sum_{t=1}^{K \cdot n_k} \sigma^2_{d_{k,t}}} \tag{10}$$

$n_k$ は $k$ 番目の音符に対応する音素列，$K$ は音素 HMM の状態数，$a_{i,j}$ は $A$ の $(i, j)$ 要素，$b_i$ は $\boldsymbol{b}$ の $i$ 番目の要素，, $\mu_{d_{k,t}}$ と $\sigma^2_{d_{k,t}}$ は $k$ 番目の音符の $t$ 番目の音素に対応する状態継続長分布の平均と分散を表わしている．以上のように，$A$ は三重対角非対称行列となり式 3 は容易に解くことができる．

## 3. 実　　験

歌声データベースを用いて歌声モデルを学習し，学習データに含まれない未知の曲を入力として歌声を合成した．さらに主観評価試験を実施し，ずれモデルの導入による有効性を確認した．

### 3.1 学習データの作成

歌声音声の学習には，**表 1** に示すような，これまでに構築した歌声データベースを利用する．このデータベースは，童謡など 60 曲を男性 1 名が歌った音声を録音し整理したものである．

歌声データのメルケプストラム分析と基本周波数抽出を行い，HMM の学習データを作成した．基本周波数の抽出には TEMPO[17] を用いた．メルケプストラム分析の条件を**表 2** に示す．

得られた分析データから，0〜24 次のメルケプスト

表 1　歌声データベース
Table 1　Singing voice database

| | |
|---|---|
| 楽曲 | 唱歌・童謡など 60 曲 (約 72 分) |
| 歌唱者 | 男性 1 名 |
| サンプリング周波数 | 44.1kHz |
| サンプルサイズ | 16bit |

表 2　歌声データの分析条件
Table 2　Experimental condition for Mel-cepstral analysis

| | |
|---|---|
| サンプリング周波数 | 16kHz (44.1kHz からダウンサンプリング) |
| フレーム周期 | 5ms |
| 分析窓長 | 25ms |
| 窓関数 | Blackman 窓 |
| 分析次数 | 24 次 |

ラム係数ベクトルと対数基本周波数の時系列を静的特徴量とし，これに前後のフレームから計算される動的特徴量を加えたものを歌声モデルの学習データとした．

$t$ 番目のフレームのメルケプストラムの静的特徴をそれぞれ $\boldsymbol{c}_t$ としたとき，その動的特徴量 $\Delta\boldsymbol{c}_t$ および 2 次動的特徴量 $\Delta^2\boldsymbol{c}_t$ は以下の式 11,12 から計算した．

$$\Delta\boldsymbol{c}_t = \frac{1}{2}(-\boldsymbol{c}_{t-1} + \boldsymbol{c}_{t+1}) \tag{11}$$

$$\Delta^2\boldsymbol{c}_t = \frac{1}{4}(\boldsymbol{c}_{t-1} - 2\boldsymbol{c}_t + \boldsymbol{c}_{t+1}) \tag{12}$$

基本周波数 $p_t$ についても同様に $\Delta p_t, \Delta^2 p_t$ を求め，メルケプストラムと基本周波数の 2 つのストリームからなる学習ベクトルの次元数は合計 78 次元となる．

### 3.2　歌声モデルの学習

歌声データから抽出されたメルケプストラムと基本周波数を MSD-HMM によってモデル化する．HMM は単混合 5 状態の left-to-right モデルとし，音素はポーズと無音を含んだ 36 種類とした．

まず，音素バランスの考慮されたテキスト読み上げ文データベースから，音素単位の HMM を学習した．これを初期モデルとして，歌声データと対応する楽譜情報を利用し，歌声合成モデルを学習した．前節で述べたとおり，歌詞から得られる音素の他，MIDI データの音階表現値を利用した音高と，当該音素が属するモーラの時間長を 100ms 単位で分類した音長のコンテキスト，小節内の音符位置について，その前後環境（先行，当該，後続）を考慮してモデルを分類し，さらに各モデルの状態パラメータを共有化するため，MDL 基準に基づいたコンテキストクラスタリングを行った．

なお，コンテキストクラスタリングは，メルケプストラム，基本周波数，音素状態継続長，ずれモデルのそれぞれに対して個別に行った．クラスタリングによって得られた木の総分布数は，スペクトル，$F0$，状態継続長，ずれモデルのそれぞれで，4265, 2477, 197, 317 となった．

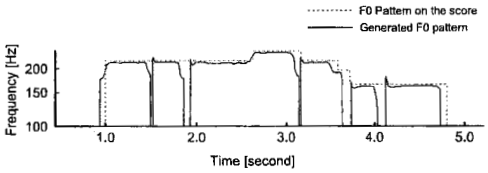


図 4　生成された基本周波数パターン
Fig. 4　Examples of generated $F0$ pattern

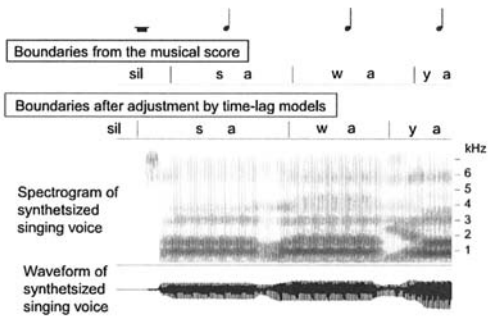


図 5　生成されたスペクトル
Fig. 5　Examples of generated spectra

### 3.3　歌声合成

学習データに含まれていない未知の曲を入力とし，歌声合成システムから各パラメータを生成，歌声を合成した．

生成された $F0$ 系列とスペクトルの一部を**図 4, 図 5** にそれぞれ示す．何れの図からも，楽譜上の音符が示す時間境界からのずれを伴った $F0$ やスペクトルが生成されている．そのずれの傾向も一様ではなく，周辺のコンテキストによって制御されるようなモデルが学習されていることがわかる．また，$F0$ 系列は，楽譜の示す音階と比較すると，全体的にやや低くなっていることや，発声のタイミングも全体的に見ると早い傾向であることから，この学習データに含まれる歌い手個人の特徴が，合成モデルに反映されていることを示唆している．

### 3.4　主観評価試験

合成された歌声の品質を評価するため，聞き取りによる主観評価実験を行った．ここでは，ずれモデルの導入によって自然性が向上したかを確認する．評価データとして，ずれモデルの有無による 2 つの条件で学習に含まれていない曲を合成し，6 小節単位に分割した

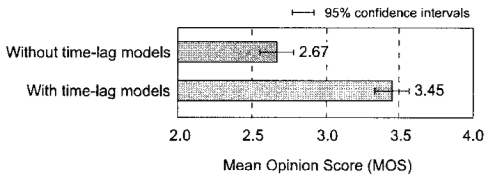


**図 6** ずれモデルの有無による自然性の評価
**Fig. 6** Evaluation of naturalness for each methods

合計 27 サンプルを作成した．各サンプルの自然性について 5 段階評価を行い，合計 14 名の被験者から得られた MOS 値を**図 6** に示す．

評価結果から，ずれモデルの導入による自然性の向上が確認でき，音符間のずれを考慮した時間構造モデルを利用することで，より歌声らしい歌声合成が可能であることが示された．

また，公式な評価結果ではないものの，この試験における被験者の多くが，合成された歌声について学習データ提供者の「本人らしさ」を感じ取っていたことを付け加えておく．

## 4. ま と め

HMM に基づいた歌声合成システムの歌声の自然性の向上を目的として，楽譜情報に基づいたコンテキストの導入，ずれモデルによる時間構造モデルの高精度化を行った．

従来手法では，歌声の時間構造に関するモデルは音符内の音素継続時間を扱うものであったが，提案法では楽譜の音符の位置と実際の歌声における発声との揺らぎをガウス分布によってモデル化し，そのゆらぎを再現した歌声合成が可能となった．実験では，童謡 60 曲からなる歌声データベースを用いて歌声合成モデルを学習・合成し，主観評価試験によってずれモデルの導入による自然性の向上が確認できた．

今後の展望として，歌い手の声質や歌唱スタイルを変化させるモデル適応の検討や，本手法の枠組みを利用した楽器演奏生成手法への適用がある．

## 参 考 文 献


1) 酒向慎司, 宮島千代美, 徳田恵一, 北村正: 隠れマルコフモデルに基づいた歌声合成システム, 情報処理学会論文誌, Vol. 45, No. 3, pp. 719–727 (2004).
2) Saino, K., Zen, H., Nankaku, Y., Lee, A. and Tokuda, K.: An HMM-based Singing Voice Synthesis System, *Proc. of INTERSPEECH*, Vol. 1, pp. 1–4 (2006).
3) 吉田由紀, 中嶌信弥: 歌声合成システム CyberSingers, 情報処理学会研究報告 音声言語情報処理, Vol. 25, No. 8 (1995).
4) Macon, N. W., Jensen-Link, L. J., Oliverio, J. and Clements, M. A.: A Singing voice synthesis system based on sinusoidal modeling, *Proc. of ICASSP*, Vol. 1, pp. 434–438 (1997).
5) ヤマハ 歌声合成ソフトウエア "VOCALID". `http://www.vocalid.com/`.
6) 吉村貴克, 徳田恵一, 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, pp. 2099–2107 (2000).
7) Tokuda, K., Yoshimura, T., Masuko, T., Kobayashi, T. and Kitamura, T.: Speech Parameter Generation Algorithm for HMM-based Speech, *Proc. of ICASSP*, Vol. 3, pp. 1315–1518 (2000).
8) Tamura, M., Masuko, T., Tokuda, K. and Kobayashi, T.: Speaker Adaptation for HMM-based Speech Synthesis System Using MLLR, *Proc. of The Third ESCA/COCOSDA workshop on Speech Synthesis*, pp. 273–276 (1998).
9) Yoshimura, T., Tokuda, K., Masuko, T. and Kobayashi, T.: Speaker Interpolation in HMM-based speech synthesis system, *Proc. of EUROSPEECH*, Vol. 5, pp. 2523–2526 (1997).
10) Shichiri, K., Sawabe, A., Yoshimura, T., Tokuda, K. and Kitamura, T.: Eigenvoices for HMM-based speech synthesis, *Proc. of ICSLP*, pp. 1269–1272 (2002).
11) 篠田浩一, 渡辺隆夫: 情報量基準を用いた状態クラスタリングによる音響モデルの作成, 信学技報, Vol. SP96-79, pp. 9–16 (1996).
12) 徳田恵一, 益子貴史, 宮崎昇, 小林隆夫: 多空間上の確率分布に基づいた HMM, 信学論 (D-II), Vol. J79-D-II, No. 7, pp. 1579–1589 (2000).
13) 徳田恵一, 小林隆夫, 斉藤博徳, 深田俊明, 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 (1991).
14) Odell, J. J.: *The use of context in large vocabulary speech recognition*, PhD Thesis, Cambridge University (1995).
15) 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, Vol. DSP98-85, No. 262, pp. 45–50 (1998).
16) 今井聖, 住田一男, 古市千恵子: 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 (1983).
17) Kawahara, H., Masuda, I. and Cheveigne, A.: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds, *Speech Communication*, Vol. 27, No. 3-4, pp. 187–207 (1999).