

# 相平面確率モデルを用いた 歌唱様式の自動分類に関する研究\*

加古達也 (名大), 大石康智, 亀岡弘和, 柏野邦夫 (NTT), 武田一哉 (名大)

## 1 はじめに

音楽情報処理の分野ではまだ歌唱様式が明確に定義されていない。歌声の性質に影響を与える音響的特徴として歌唱フォルマントやピブラート、オーバーシュート、プレバレーション、微細変動成分などの  $F_0$  動的変動成分の存在が明らかにされた [1]。先行研究から、歌唱様式は歌声のスペクトルや  $F_0$  変化に含まれる、楽譜情報以外の成分に対応づけられると考えられる。特に本研究では、歌声の  $F_0$  に含まれる動的変動成分に着目し、歌唱様式のモデル化手法を提案する。

これまでも  $F_0$  の動的変動成分を利用した先行研究がいくつか提案されてきた。ケプストラム分析に基づいてピブラートを検出し、歌唱者識別を行った研究 [2] や、 $F_0$  軌跡から抽出した相対音高情報、 $F_0$  軌跡を FFT することによって検出されたピブラート情報を利用して、楽譜情報を使わない歌唱力自動評価に関する研究 [3] がある。このように局所的な  $F_0$  の変動を利用した様々な研究があるが、歌唱様式を厳密に定義しそれを自動識別しようとする研究はこれまでなかった。また、 $F_0$  軌跡の新たな表現方法として相平面上に  $F_0$  軌跡を描くことで動的変動成分を可視化する手法 [4] が提案された。本稿では、歌唱様式を相平面を利用することで可視化し、この相平面上の軌跡を確率を用いてモデル化する手法を提案する。提案モデルを評価するために歌唱カテゴリ識別実験を行う。

## 2 相平面を利用した歌唱様式の確率表現

### 2.1 $F_0$ の相平面表現

$F_0$  が歌唱者ごとに決まる微分方程式に従って生成されると仮定し、この解 ( $F_0$ ) の性質を調べるために相平面を利用する。相平面とは、 $F_0$  と  $F_0$  の 1 階微分や 2 階微分によって構成される平面である。 $F_0$  の 1 階微分は式 (1) のように微小区間の回帰係数  $\Delta F_0$  として求め、同様に 2 階微分は  $\Delta F_0$  の回帰係数  $\Delta\Delta F_0$  として求めた。

$$\Delta F_0(n) = \frac{\sum_{k=-K}^K k \cdot F_0(n+k)}{\sum_{k=-K}^K k^2}, \quad (1)$$

$F_0$ ,  $\Delta F_0$ ,  $\Delta\Delta F_0$  の 3 つの成分に着目し、相平面上に  $F_0$  軌跡を表現した。相平面上に描かれる  $F_0$  軌跡の例を図 1 に示す。図 1 の上段は  $F_0$  の時間軌跡で、中段は ( $F_0$ ,  $\Delta F_0$ ) で構成される相平面、下段は ( $F_0$ ,  $\Delta\Delta F_0$ ) で構成される相平面が描かれている。 $(F_0, \Delta F_0)$  で構成される相平面上に描かれる渦軌跡の中心は歌唱者の意図する目標音高であり、周期的な変動のピブラートであれば目標音高を中心に円を描く軌跡が観測できる。また、オーバーシュートは螺旋を描きながら目標音高へ収束する動きとして観測でき、 $(F_0, \Delta\Delta F_0)$  構成の相平面上の軌跡は負の方向に傾いた直線が描かれることが確認できた。

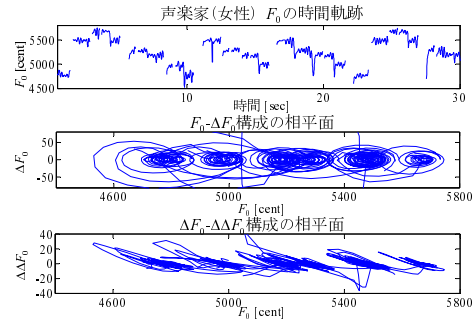


Fig. 1 声楽家女性の  $F_0$  軌跡, 上段:  $F_0$  の時間軌跡, 中段:  $F_0$ - $\Delta F_0$  構成の相平面上に描かれる  $F_0$  軌跡, 下段:  $F_0$ - $\Delta\Delta F_0$  構成の相平面上に描かれる  $F_0$  軌跡。

### 2.2 相平面確率モデル

歌唱様式は相平面上の軌跡に表れるため、この軌跡をモデル化することを考える。相平面上に描かれる  $F_0$  軌跡に確率密度関数をフィッティングし、これを相平面確率モデルと呼ぶ。これが歌唱様式をモデル化していると考えられる。相平面上に描かれる  $F_0$  軌跡は、複数の目標音高を中心に分布している。この分布を同時に確率密度関数で表現するために、混合ガウスモデル (GMM) を用いる。GMM は以下のように表される

$$\sum_{m=1}^M \lambda_m \mathcal{N}(\mathbf{f}_0(n); \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m), \quad (2)$$

ここで、 $\mathbf{f}_0(n) = [F_0(n), \Delta F_0(n), \Delta\Delta F_0(n)]^T$  であり、 $\mathcal{N}(\cdot)$  はガウス分布、 $\Theta = \{\lambda_m, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m\}$  はモデルパラメータである。各パラメータの  $\lambda_m$ ,  $\boldsymbol{\mu}_m$ ,  $\boldsymbol{\Sigma}_m$  はそれぞれ相対頻度、平均ベクトル、全共分散行列である。

プロの声楽家女性、ポップス歌手女性、音楽経験のない素人女性の 3 名の歌唱者の ( $F_0$ ,  $\Delta F_0$ ) 構成の相平面確率モデルを図 2 に示す。プロの声楽家を見ると、深いピブラートをつかった歌唱法のためガウス分布の垂直方向に大きな分散が確認できる。一方、素人では水平方向に大きな分散が確認できる。また、ポップス歌手では垂直方向の分散が小さく、深いピブラートが確認できないが、水平方向の分散も小さいため音高を正確に保った歌唱法であることがわかる。このように、歌唱様式は相平面確率モデルのパラメータセット  $\Theta$  によってモデル化することができる。

## 3 評価実験

歌唱様式をモデル化した、相平面確率モデルを利用して、3 つの歌唱カテゴリの識別実験を行った。

### 3.1 実験条件

プロの声楽家、ポップス歌手、音楽経験のない素人 (男女各 1 名ずつ) の計 6 名からなる歌声データを利用した。歌唱曲は「きらきら星」(1 番, 2 番, ハミング)、「喜びの歌」(1 番, 2 番, ハミング)、「練習曲」

\* Automatic identification for singing style based on stochastic representation of phase plane by KAKO Tatsuya(Nagoya Univ.), OHISHI Yasunori, KAMEOKA Hirokazu, KASHINO Kunio(NTT), and TAKEDA Kazuya(Nagoya Univ.)

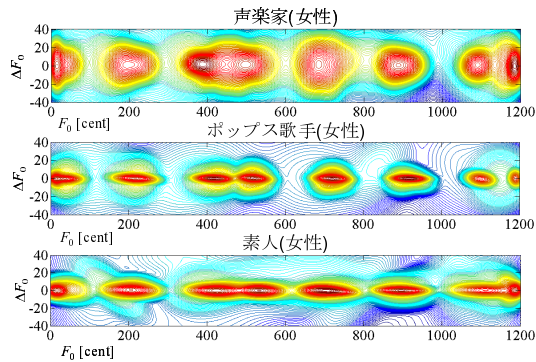


Fig. 2  $F_0$ - $\Delta F_0$  構成の相平面確率モデル, 上段: プロの声楽家女性, 中段: ポップス歌手女性, 下段: 素人女性

Table 1  $F_0$  推定の実験条件

サンプリング周波数	16kHz
$F_0$ 算出のフレーム長	64ms
窓関数	ハニング窓
窓シフト長	10ms
$F_0$ 平滑化 (移動平均窓長)	50ms

(5パターン)であり, 収録方法は, 各歌唱者がヘッドホンで伴奏・ガイドトーンを聴きながら歌ったものと, 何も聴かずに歌ったものの歌唱者ごと28種類, 合計168種類の歌声データを収録した。 $F_0$ は, 後藤ら[5]の提案した $F_0$ 推定手法を利用した。 $F_0$ 推定の実験条件を表1に示す。また式(1)の回帰係数は $K=2$ として $\Delta F_0, \Delta\Delta F_0$ を算出した。

実験に用いた歌声データは歌唱者ごとに異なる音高で歌われているが, 本手法では歌唱者ごとの歌唱様式に着目するため歌声データの音高を正規化する。正規化の方法を以下に示す。まず, 周波数 [Hz] の $F_0$ を対数スケールの周波数 [cent] に次式のように変換し, 音高を平均律に変換する。

$$1200 \times \log_2 \frac{F_0}{440 \times 2^{3/12-5}} \quad [\text{cent}]. \quad (3)$$

次に $F_0$ の音域を半音単位の(0,100) [cent]に制限した。

$$\text{mod}(F_0 + 50, 100). \quad (4)$$

また $\Delta, \Delta\Delta$ 成分は平均0, 分散1に正規化した。GMMの学習用データとして「きらきら星」と「練習曲」を用い, $F_0, \Delta F_0, \Delta\Delta F_0$ の3次元ベクトルの分布をEMアルゴリズムを用いて分布をGMMで学習した。評価用データには「喜びの歌」を用い歌声データから5~15[s]を切り出し, 3カテゴリ識別実験を行った。

### 3.2 識別方法

声楽家, ポップス歌手, 素人の3カテゴリを最大事後確率に基づいて式(5)を用いて識別する。

$$\begin{aligned} \hat{s} &= \arg \max_s [p(s|\{F_0, \Delta F_0, \Delta\Delta F_0\})] \\ &= \arg \max_s \left[ \frac{1}{N} \sum_{n=1}^N \log p(\mathbf{f}_0(n)|\Theta_s) + \log p(s) \right] \end{aligned} \quad (5)$$

ここで $s$ は歌唱カテゴリ,  $\Theta_s$ は歌唱カテゴリ $s$ のモデルパラメータ,  $N$ は評価データ長を表す。歌唱カテゴリの事前確率 $p(s)$ は等確率と想定し, 事後確率は尤度を計算することで求め, 尤度が最大となるモデルを識別結果とした。

### 3.3 実験結果

特徴量に( $F_0, \Delta F_0, \Delta\Delta F_0$ )を用いた歌唱カテゴリ識別実験の識別結果を図3に示す。識別率は評価デー

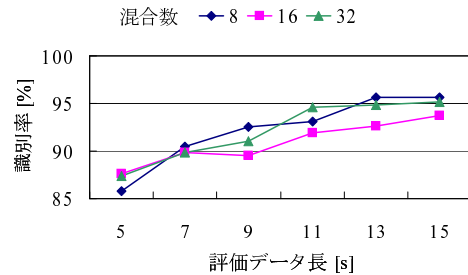


Fig. 3 3歌唱カテゴリの識別結果

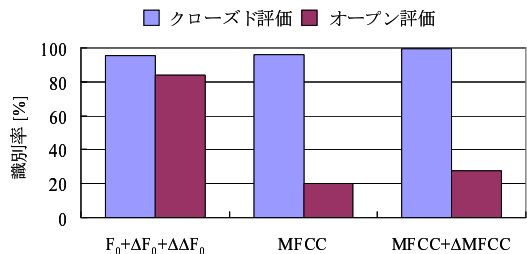


Fig. 4 歌唱者の実験条件を変えたときの識別率比較

タ長が長くなると良くなり, GMMの混合数8で評価データ長13[s]のとき歌唱カテゴリ識別率が96%で最大となった。図4に同じ歌唱者が学習データにも評価データにも含まれる実験条件をクローズド評価とし, 学習データに男性を評価データに女性を利用した実験条件をオープン評価とし, 特徴量に( $F_0, \Delta F_0, \Delta\Delta F_0$ ), 12次のMFCC, (MFCC,  $\Delta$ MFCC)の3セットを用いて, GMMの混合数32, 評価データ長15[s]で歌唱カテゴリ識別を行った結果を示す。

## 4 考察

GMMの混合数が多いと過学習となってしまう識別率が低下してしまう。歌唱者の歌唱カテゴリ識別においては, 実験結果から混合数8が最適であるといえる。また, 12次のMFCCを利用して識別したところ95%の識別率。(MFCC,  $\Delta$ MFCC)を利用した場合99%の識別率でありクローズド評価の実験条件では高い識別率を得た。しかし, オープン評価の実験条件ではMFCC,  $\Delta$ MFCCを利用した場合の識別率は低くなり, $F_0$ を利用した識別では84%の識別率が得られた。MFCCは歌唱者ごとの個人識別には適しているが, 歌唱者が異なった場合は共通する歌唱様式は分類することができない。しかし, 歌唱様式をモデル化した相平面確率モデルを利用することで歌唱者のピブラートの深さ, 音程の安定性などの歌唱方法が共通する歌唱者を分類することができる。

## 5 まとめ

本稿では, 相平面確率モデルを用いた歌唱様式の分類手法を提案した。相平面を用いることで, 歌声に含まれる歌唱様式の特徴を抽出した。相平面上に描かれた渦軌跡の形状をGMMによって確率表現した。その結果, 3種類の歌唱カテゴリ識別で96%以上の識別率を得た。本稿の実験では規模が小さいため, 歌唱者と歌唱カテゴリの数を増やして実験を行うことが今後の課題である。

## 参考文献

- [1] 齋藤 他, 音講論, 1-R-21, 457-458, 2008.
- [2] Tin *et al.*, IEEE Transactions on Audio, Speech, and Language Processing, 15 (2), 519-530, 2007.
- [3] 中野 他, 情処論, 48 (1), 227-236, 2007.
- [4] 大石 他, 情処論, 49 (11), 3789-3797, 2008.
- [5] 後藤 他, 信学論, 83 (11), 2330-2340, 2000.