Vol.2009-MUS-81 No.4
2009/7/29

相平面に描かれる F_0 の動的変動成分を利用した歌唱様式の自動分類

加古達也[†1] 大石康智[†2] 亀岡弘和[†2]
柏野邦夫[†2] 武田一哉[†1]

本研究では，歌唱様式の確率表現手法を提案する．歌声の基本周波数 (F0) の軌跡に現れるビブラートのような動的変動成分には個人ごとに特徴があり，このような F_0 軌跡の振る舞いが歌唱様式であると考える．歌声の F_0 軌跡を相平面上に描くことで動的変動成分を可視化できる．歌唱様式が表れる相平面上の渦軌跡をモデル化するために混合ガウス分布 (GMM) による分布の学習をし，相平面を確率的に表現する．歌唱様式の確率表現手法の有効性を確認するために，歌唱カテゴリ識別実験を行った．その結果 96% の識別率を得た．

Automatic Identification for Singing Style Based on Sung Melodic Contour Characterized in Phase Plane

Tatsuya Kako,[†1] Yasunori Ohishi,[†2]
Hirokazu Kameoka,[†2] Kunio Kashino[†2]
and Kazuya Takeda [†1]

In this paper, a stochastic representation of singing styles is proposed. The dynamic property of the melodic contour, i.e. fundamental frequency (F_0) sequence, is assumed to be the main cue of singing styles because typical ornamentations such as *vibrato* has an individuality. F_0 signal trajectories in the phase plane are used as the basic representation. By fitting Gaussian mixture models to the observed F_0 trajectories in the phase plane, a parametric representation is obtained by a set of GMM parameters. The effectiveness of the proposed method is confirmed through experimental evaluation where 96% accuracy for singer-class discrimination was obtained.

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．しかし，先行研究によって，歌声の性質を左右する音響的特徴として歌唱フォルマント[1),2)] やビブラート[3)–11)]，オーバーシュート[12)]，プレパレーション[13)]，微細変動成分[14)] などの F_0 動的変動成分の存在が明らかにされた．さらに，心理実験に基づきこれらの音響的特徴が歌声の個人性知覚にどの程度寄与するかについても検討されている[15)]．その結果，スペクトル，F_0 変化パターンの順で，個人性知覚への影響があることが示された．これらの先行研究から，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 (音高列や歌詞の音韻) 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．

これまでにも 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ビブラートを検出し，歌唱者識別を行った研究[16)] 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ビ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[17)] 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．また歌声合成に関する研究では，二次系モデルに基づいてビブラートやオーバーシュートを表現する研究[15),18),19)] があるが，歌唱様式との対応づけについては議論されていなかった．

本稿では，歌唱様式を相平面を利用することで可視化し，相平面を確率表現する手法を提案する．そして，本手法を評価するために歌唱様式の識別を行う．相平面に歌唱様式の特徴を表現する利点は，ビブラートのような F_0 の動的変動成分を検出するための処理や，曲の楽譜情報が必要ないことがあげられる．

先行研究[20)] で，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し，ハミング検索へ利用する手法が提案された．この手法では，F_0 の動的変動の揺れを吸収し歌唱者の目標とする音高のみを抽出しハミング検索へ利用していた．本研究では対照的に，歌唱様式のモデル化を行うために F_0 の動的変動成分を利用する．また，歌唱様式は楽曲によらず歌唱

†1 名古屋大学大学院情報科学研究科
Graduate School of Information Science, Nagoya University
†2 NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, NTT Corporation

Vol.2009-MUS-81 No.4
2009/7/29

者ごとに表れる特徴のため．楽曲の楽譜情報は用いない．

本稿では，提案する表現方法を評価するために歌唱カテゴリ識別実験を行った．提案したモデルによって歌声から動的変動成分を取り出すことができているかを歌唱様式の似ている歌唱者のグループを作成し確認した．

以下，第 2 章では相平面上に表現した F_0 軌跡の確率表現について述べ，歌声の動的変動成分が確率的にモデル化できているかを確認する．第 3 章では，提案手法の効果を調査するために歌唱カテゴリ識別実験を行う．第 4 章では，実験結果を示し本稿のまとめと今後の展開を述べる．

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

ビブラートのような F_0 軌跡にみられる動的変動成分には個人性が現れる．ここで，歌声から観測される F_0 は人間の発声器官によって調節されて出力される．この F_0 軌跡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 とその時間微分 $\frac{dF_0}{dt}$ からなる 2 次元平面であり，F_0 軌跡の動的変動成分を可視化するために用いる．

しかし，観測する F_0 の信号列は連続的な時間関数 $F_0(t)$ ではないため，本手法では信号列 $\{F_0(n)\}_{n=1,\ldots,N}$ から時刻 n の時間微分を回帰係数 ΔF_0 を用いて推定した．

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

同様に ΔF_0 の回帰係数 $\Delta\Delta F_0$ を計算し 2 次の微係数 $\frac{d^2F_0}{dt^2}$ の近似値として用いた．

F_0, ΔF_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 と ΔF_0 で構成される相平面，下段は F_0 と $\Delta\Delta F_0$ で構成される相平面が描かれている．F_0-ΔF_0 構成の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る．周期的な変動のビ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，F_0-$\Delta\Delta F_0$ 構成の相平面上の軌跡は $-45°$ 傾いた直線が現れている．これは，F_0 が正弦波成分の足し合わせ

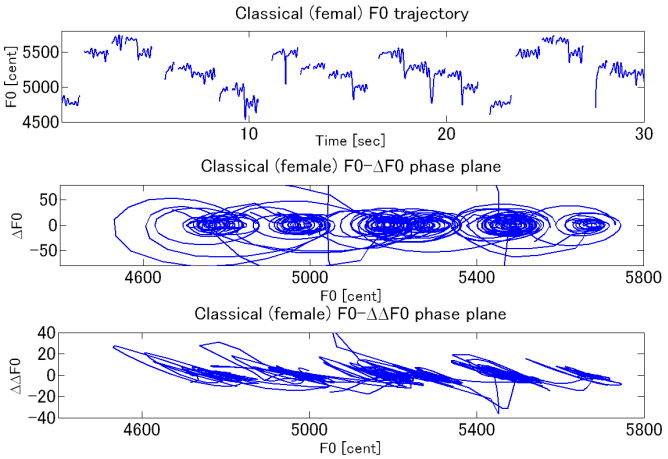

図 1 声楽家女性の F_0 軌跡，上段：F_0 の時間軌跡，中段：F_0-ΔF_0 構成からなる相平面上に描かれる F_0 軌跡，下段：F_0-$\Delta\Delta F_0$ 構成からなる相平面上に描かれる F_0 軌跡．

Fig. 1 Melodic contour (top), corresponding phase planes for F_0-ΔF_0 (middle) and F_0-$\Delta\Delta F_0$ (bottom)

と考えれば $\Delta\Delta F_0$ には式 (2) のような関係があるためである．

$$\Delta\Delta F_0 = -F_0. \tag{2}$$

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た．

2.2 相平面の確率表現

歌唱様式は相平面上の軌跡に現れる．この軌跡を工学的に取り扱うためにパラメータ化する必要がある．しかし，F_0 軌跡は歌い方によって変化するために確定的な信号としてモデル化することが難しい．そこで，確率的に F_0 軌跡を表現する．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することで相平面確率モデル (SPP) を生成し，歌唱様式をモデル化する．相平面上に描かれる F_0 軌跡は，複数の音高目標を中心に分布する．つまりデータは多峰型に分布する．この分布を確率密度関数で表現するために，混合ガウスモデル (Gaussian Mixture Model:GMM) を用いる．GMM は以下のように表される

$$\sum_{m=1}^{M} \lambda_m \mathcal{N}(\mathbf{f}_0(n); \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m), \tag{3}$$

Vol.2009-MUS-81 No.4
2009/7/29

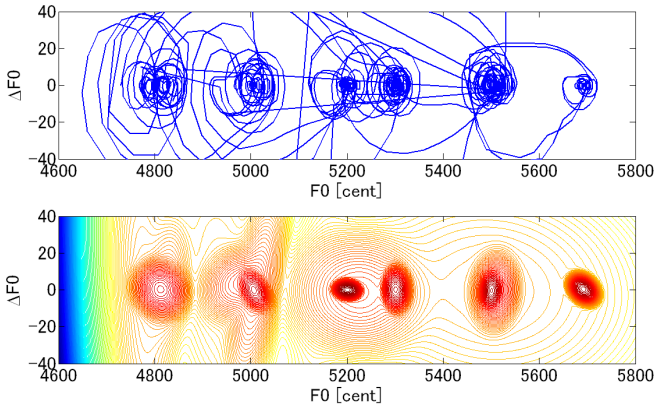

図 2 相平面上に描かれる F_0 軌跡の分布を学習した GMM.
Fig. 2 Gaussian Mixture model fitted to the F_0 contour in the phase plane.

ここで，$\mathbf{f}_0(n)$ は

$$\mathbf{f}_0(n) = [F_0(n), \Delta F_0(n), \Delta\Delta F_0(n)]^{\mathrm{T}}, \tag{4}$$

であり，$\mathcal{N}(\cdot)$ はガウス分布，

$$\boldsymbol{\Theta} = \{\lambda_m, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m\}_{m=1,\cdots,M} \tag{5}$$

はモデルパラメータであり，λ_m, $\boldsymbol{\mu}_m$, $\boldsymbol{\Sigma}_m$ はそれぞれ相対頻度，平均ベクトル，共分散行列である．

F_0 軌跡を GMM で表現したものを図 2 に示す．F_0 軌跡の形状がモデルに学習できていることが確認できる．それぞれのガウス分布の水平方向の分散は目標音高周りの持続的な F_0 軌跡を表しており，垂直方向の分散はビブラートの深さを表している．このように，歌唱様式は確率表現された相平面のパラメータセット $\boldsymbol{\Theta}$ によってモデル化することができる．

2.3 相平面確率モデル例

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-ΔF_0 構成の相平面確率モデルを図 3 に示す．プロの声楽家をみると，深いビ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の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ある特徴が確認できる．ポップス歌手には深いビブラートが確認できないが，水平方向の分散が小さいため正確な音高を保った歌唱法であることがわかる．

また，F_0-$\Delta\Delta F_0$ 構成の相平面確率モデルを図 4 に示す．プロの声楽家にのみ，F_0 と

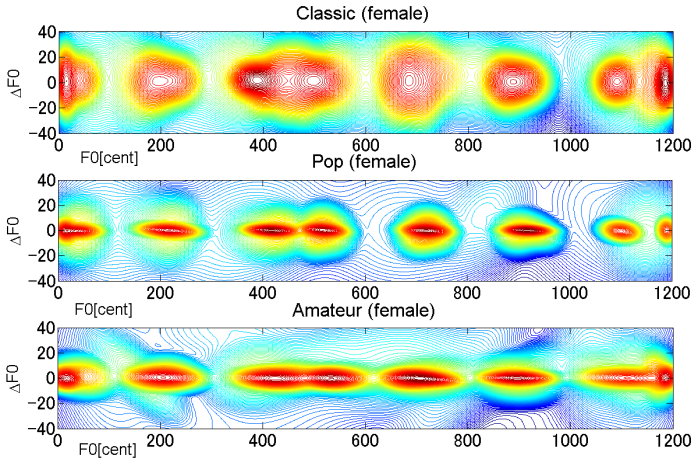

図 3 F_0-ΔF_0 構成の相平面確率モデル，上段：プロの声楽家女性，中段：ポップス歌手女性，下段：素人女性
Fig. 3 Stochastic phase plane models for a professional classic (top), a professional pops (middle) and an amateur (bottom) singer.

$\Delta\Delta F_0$ に強い負の相関がみられた．このことから，深いビブラートを用いた歌唱様式では，負の強い相関が表れることが言える．

3. 評価実験

相平面確率モデルの性能を確認する評価実験を行った．評価実験では，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),「発声練習」(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[21] 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 にしめす．また式 (1) の回帰係数は $K = 2$ として ΔF_0, $\Delta\Delta F_0$ を算出した．

Vol.2009-MUS-81 No.4
2009/7/29

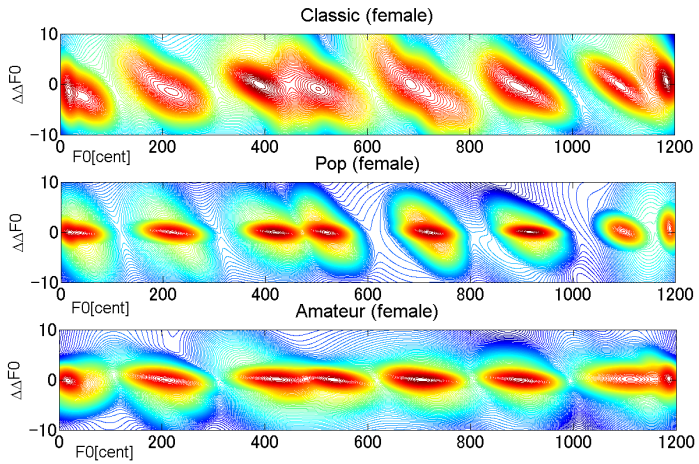

図 4 F_0-$\Delta\Delta F_0$ 構成の相平面確率モデル，上段：プロの声楽家女性，中段：ポップス歌手女性，下段：素人女性
Fig. 4 2nd order Stochastic phase plane models for a professional classic (top), a professional pops (middle) and an amateur (bottom) singer.

表 1 調波構造のパターンマッチングを用いた F_0 推定[21] の実験条件.
Table 1 Signal analysis conditions for F_0 estimation. Harmonical PSD pattern matching[21] is used with these parameters.

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

データは歌唱者ごとに異なる音高で歌われている．本実験では，歌唱者ごとの歌唱様式のみに着目するために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，音高を等間隔に並べる．

$$1200 \times \log_2 \frac{F_0}{440 \times 2^{3/12-5}} \quad \texttt{[cent]}. \tag{6}$$

次に F_0 の音域を半音単位の (0,100)[cent] に制限した．

$$\mathrm{mod}\ (F_0 + 50, 100). \tag{7}$$

GMM の学習用データとして，各歌唱者ごとに歌声データの「きらきら星」と「発声練習」から推定した F_0, ΔF_0, $\Delta\Delta F_0$ を用いて，EM アルゴリズムによって分布の学習を行い歌唱カテゴリのモデルを作成した．

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 (8) を用いて識別する．

$$\begin{aligned}\hat{s} &= \arg\max_s \left[p(s|\{F_0, \Delta F_0, \Delta\Delta F_0\})\right] \\ &= \arg\max_s \left[\frac{1}{N}\sum_{n=1}^{N} \log p(\mathbf{f}_0(n)|\Theta_s) + \log p(s)\right]\end{aligned} \tag{8}$$

ここで s は歌唱カテゴリ，Θ_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 $p(s)$ は一様分布と想定し，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, ΔF_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 5 に示す．識別率は評価デー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 秒のとき歌唱カテゴリ識別率が 96% で最大となった．しかし，評価データ長が 13 秒以上は識別率に変化が見られなくなった．また GMM の混合数は評価データ長 13 秒では混合数 8 が最も高い識別率となった．GMM の混合数が多いと歌唱者ごとの個人性を学習してしまうため，異なる歌唱者の歌唱カテゴリ識別においては，混合数 8 が最適であるといえる．図 6 に歌唱カテゴリ識別の次元数の識別結果の比較を示す．特徴量に F_0, (F_0, ΔF_0), (F_0, ΔF_0, $\Delta\Delta F_0$) の 3 セットを用いて，GMM の混合数 8, 評価データ長 13 秒で歌唱カテゴリ識別を行った．F_0 と (F_0, ΔF_0) の識別結果を比べると F_0 のみを特徴量に用いた場合よりも ΔF_0 も併用した方が誤り率が半減している．また，$\Delta\Delta F_0$ も使うことにより ΔF_0 の結果に比べ誤り率が減少していることがわかり，(F_0, ΔF_0, $\Delta\Delta F_0$) を特徴量に用いることの有効性が確認できる．これらの結果から相平面の確率表現が 3 種類の歌唱カテゴリの歌唱様式の表現に効果があることがわかる．

4. まとめ

本稿では，F_0 軌跡の動的変動成分の確率表現に基づく歌唱様式のモデル化手法を提案した．歌声が微分方程式に従って生成されると考え，相平面を用いることで歌声に含まれる歌

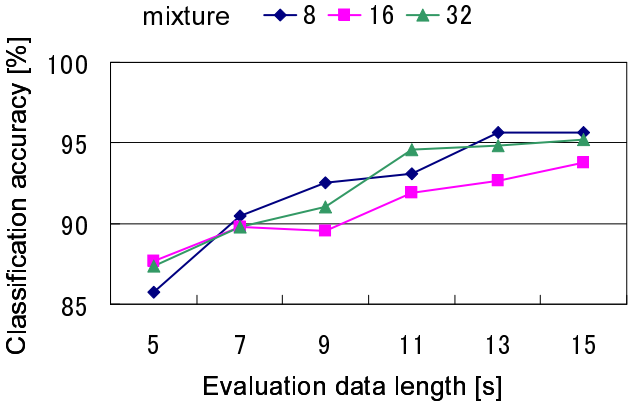

図 5 3 歌唱カテゴリの識別結果
Fig. 5 Accuracy in discriminating three singer classes.

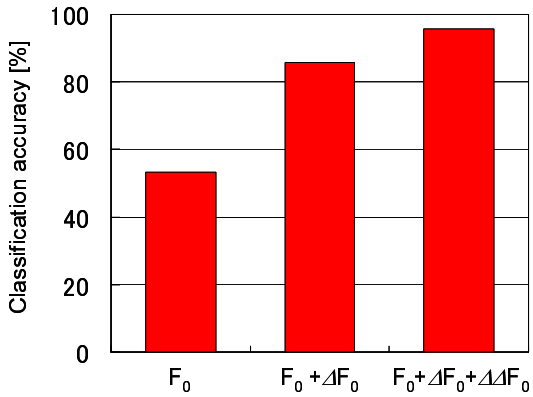

図 6 次元数を変化させたときの 3 歌唱カテゴリ識別結果
Fig. 6 Comparing the accuracy in discriminating singer classes.

唱様式の特徴を抽出した．そして，相平面上に描かれた渦軌跡の形状を GMM によって分布の学習を行い歌唱様式を数値化した．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．また実環境下で利用できるアプリケーション開発のために，カラオケのような雑音が存在する歌唱条件下で推定された F_0 を利用しても，提案手法を用いた頑健な識別を目指す．

参 考 文 献

1) Sundberg, J.: *The Science of the Singing*, Northern Illinois University Press (1987).
2) Sundberg, J.: Singing and timbre, *Music room acoustics*, Vol.17, pp.57–81 (1977).
3) Seashore, C.E.: A Musical Ornament, the Vibrato, *Psychology of Music*, McGraw-Hill Book Company, pp.33–52 (1938).
4) Large, J. and Iwata, S.: Aerodynamic Study of Vibrato and Voluntary "Straight Tone" Pairs in Singing, *J. Acoust. Soc. Am.*, Vol.49, No.1A, p.137 (1971).
5) Rothman, H.B. and Arroyo, A.A.: Acoustic variability in vibrato and its perceptual significance, *J. Voice*, Vol.1, No.2, pp.123–141 (1987).
6) Myers, D. and Michel, J.: Vibrato and pitch transitions, *J. Voice*, Vol.1, No.2, pp. 157–161 (1987).
7) Hakes, J., Shipp, T. and Doherty, E. T.: Acoustic characteristics of vocal oscillations: Vibrato, exaggerated vibrato, trill, and trillo, *J. Voice*, Vol. 1, No. 4, pp. 326–331 (1988).
8) D'Alessandro, C. and Castellengo, M.: The pitch of short-duration vibrato tones, *J. Acoust. Soc. Am.*, Vol.95, No.3, pp.1617–1630 (1994).
9) Gerhard, D.: Pitch track target deviation in natural singing, *ISMIR*, pp.514–519 (2005).
10) Kojima, K., Yanagida, M. and Nakayama, I.: Variability of Vibrato -A Comparative Study between Japanese Traditional Singing and Bel Canto-, *Speech Prosody*, pp.151–154 (2004).
11) Nakayama, I.: Comparative Studies on Vocal Expressions in Japanese Traditional and Western Classical- Style Singing, Using a Common Verse, *ICA*, pp.1295–1296 (2004).
12) de Krom, G. and Bloothooft, G.: Timing and Accuracy of Fundamental Frequency Changes in Singing, *ICPhS95*, pp.206–209 (1995).
13) 齋藤毅：歌声知覚・生成機構の解明に向けた歌声合成システム構築に関する研究，博士論文 (2006).
14) Akagi, M. and Kitakaze, H.: Perception of Synthesized Singing Voices with Fine Fluctuations in Their Fundamental Frequency Contours, *ICSLP*, pp.458–461 (2000).
15) Saitou, T., Goto, M., Unoki, M. and Akagi, M.: Speech-To-Singing Synthesis: Con-

Vol.2009-MUS-81 No.4
2009/7/29

verting Speaking Voices to Singing Voices by Controlling Acoustic Features Unique to Singing Voices, *WASPAA*, pp.215–218 (2007).
16) Nwe, T.L. and Li, H.: Exploring Vibrato-Motivated Acoustic Features for Singer Identification, *IEEE Transactions on Audio, Speech, and Language processing*, pp. 519–530 (2007).
17) 中野倫靖，後藤真孝，　平賀譲：楽譜情報を用いない歌唱力自動評価手法，情報処理学会論文誌，Vol.48, No.1 (2007).
18) Mori, H., Odagiri, W. and Kasuya, H.: F0 dynamics in singing: Evidence from the data of a baritone singer, *IEICE Trans. Inf. and Syst.*, Vol.E87-D, No.5, pp. 1086–1092 (2004).
19) Minematsu, N., Matsuoka, B. and Hirose, K.: Prosodic Modeling of Nagauta Singing and Its Evaluation, *SpeechProsody*, pp.487–490 (2004).
20) 大石康智，後藤真孝，伊藤克亘，武田一哉：相平面に描かれる歌声の基本周波数軌跡：歌唱者の意図する音高目標値系列の推定とハミング検索への応用，情報処理学会論文誌，Vol.49, No.11, pp.1234–1242 (2008).
21) 後藤真孝，伊藤克亘，　速水悟：自然発話中の有声休止箇所のリアルタイム検出システム，信学論 (D-II)，Vol.83, No.11, pp.2330–2340.