Vol.2012-MUS-94 No.11
Vol.2012-SLP-90 No.11
2012/2/3

歌唱における表現意図を考慮した歌声 F_0 生成過程とその統計的モデリング

大石康智[†1]　亀岡弘和[†1,†2]
持橋大地[†3]　柏野邦夫[†1]

歌声 F_0 軌跡における楽譜逸脱成分（F_0 動的変動成分）を楽譜情報と分離して抽出することを目的として，F_0 軌跡の新しい生成過程モデルを提案する．従来，オーバーシュートのような音符の立ち上がりに関する F_0 動的変動成分は 2 次系を用いてモデル化されたが，ビブラートやポルタメントのような意図的表現による変動成分および微細変動成分はすべて雑音としてモデル化された．提案する F_0 生成過程は楽譜に記載される音符の並びを表現するノート指令信号と歌唱者の音楽的な表現意図を表す表現指令信号によって 2 次系が駆動されるモデルであり，ノート成分と表現成分を出力する．これらの成分とガウス性白色雑音に従う微細変動成分との和によって F_0 軌跡を記述する．そのモデルパラメータを推定する逆問題の解法アルゴリズムを導出し，評価実験では，推定される表現成分に歌唱者の音楽的な表現意図が含まれるかどうかを客観的かつ主観的に評価して，提案モデルの有効性について議論する．

Statistical Modeling of Singing Voice F_0 Contours by Considering Musical Expressive Intentions

Yasunori Ohishi,[†1] Hirokazu Kameoka,[†1,†2]
Daichi Mochihashi[†3] and Kunio Kashino[†1]

We present a novel statistical model of singing voice fundamental frequency (F_0) contours for characterizing both musical-note information and various dynamic components such as *overshoot* and *vibrato*. Previous studies have modeled the dynamics using a second-order linear system and estimated the model parameters, but so far with limited success due to lack of flexibility for modeling the varied dynamics. Therefore we introduce the process of generating F_0 contours based on musical note and expression command functions, which generate the temporal attack of the note such as *overshoot* and the intended expression such as *vibrato*, respectively. Then we formulate a discrete-time stochastic process version of this model and propose a powerful framework for the estimation of the model parameters. In our experiment, the proposed method successfully decomposes F_0 contours into musical note and dynamic components.

1. はじめに

楽曲のメロディを歌った歌声の音高 (F_0) 軌跡には，そのメロディを構成する楽譜の音符の並びだけでなく，楽譜に記載されない，“楽譜から逸脱した” 動的変動成分が含まれる（**図 1**）．これらは，発声器官の物理的制約に起因する成分（特に，オーバーシュートや微細変動成分[1)–3)]）と歌唱者の意図的表現による成分（特に，ビブラートやポルタメント[4)–6)]）からなると考えられ，知覚的には，前者は人間らしさ・自然性に関係し，後者は巧拙感に関係することがわかってきている[7)–9)]．さらに，後者は意図して意図通りに表現できた場合と，意図通りに制御できなかった場合とに分かれ，習熟度に関連すると考えられる．ただし，これらの逸脱成分を F_0 軌跡から特徴抽出し，歌唱者ごとにその特性を精緻に学習することまではまだ十分に検討されていない．

本研究の目的は，このような物理的制約もしくは意図的表現による楽譜逸脱成分を F_0 軌跡から楽譜情報と分離して抽出し，歌唱者ごとにどのようなパターンをもちうるのか，各パターンが文脈（楽譜の音符列）にどう依存するかを計算機に学習させることである．ここでは，うろ覚えの状態で歌った歌声ではなく，楽曲のメロディまたはその楽譜を既知として，歌唱者なりに表情付けして歌った歌声を対象とする．本研究は，歌唱者の歌い方や個性，癖を学習することを目指しており，歌唱力評価や歌唱者識別，そして現在盛んに研究される歌声合成や歌声変換[10)–18)] への応用が期待できる．例えば，ある歌声を別の歌唱者の歌い方に変換して合成することが可能となるだろう．ここでは事前に歌唱者の歌い方が学習されるため，どんなメロディにもその歌い方を転写できることを特長とする．

本研究では，物理モデルに基づいて F_0 軌跡の生成過程を記述し，そこから楽譜逸脱成分の特徴抽出に取り組む．従来，線形 2 次系を利用して歌声の F_0 に含まれる動的変動成分を制御するモデルが提案された[19)–22)]．これらの研究では，日本語の話声の F_0 パターンを表現する藤崎モデル[23)] が参考にされた．藤崎モデルは，臨界制動 2 次系のインパルス応答とステップ応答を利用して，日本語の句頭から句末に向けて緩やかに下降するフレーズ成分と，語句に対応して急激に上昇下降するアクセント成分を表現し，これらを重畳することで

†1 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation
†2 東京大学情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo
†3 大学共同利用機関法人 情報・システム研究機構 統計数理研究所
Institute of Statistical Mathematics

Vol.2012-MUS-94 No.11
Vol.2012-SLP-90 No.11
2012/2/3

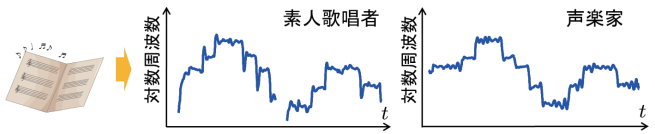

図 1 歌声の基本周波数（F_0）軌跡とその動的変動成分

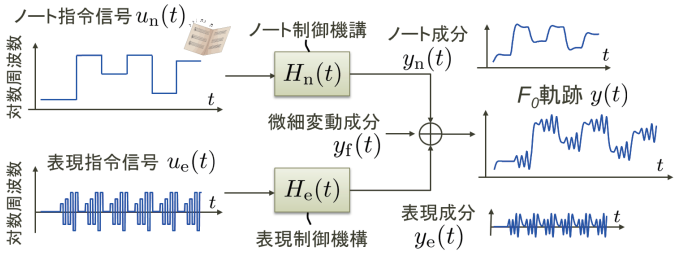

図 2 提案する歌声の F_0 生成過程の概略図

F_0 を記述する．ただし，歌声のメロディに伴った急激な F_0 の上昇・下降の制御及び，ビブラートのような準周期的な振動は臨界制動系では表現できない．そのため，歌声の F_0 制御モデルでは 2 次系の伝達関数

$$\mathcal{G}(s) = \frac{\Omega^2}{s^2 + 2\zeta\Omega s + \Omega^2} \tag{1}$$

における減衰率 ζ を調整することによって，指数減衰 ($\zeta > 1$)，減衰振動 ($0 < \zeta < 1$，オーバーシュートを表現する)，臨界制動 ($\zeta = 1$)，定常振動 ($\zeta = 0$，ビブラートを表現する) からなる様々な振動現象を表現する．文献 22) では，楽譜の音符列を表す階段状信号に式 (1) のインパルス応答を部分的に畳み込んで得られる F_0 軌跡を利用して，表情豊かな歌声合成音を実現した．しかしながら，制御パラメータ（減衰率 ζ と固有周波数 Ω）は手作業あるいは規則に基づいて決定された．

これに対し，我々は観測される F_0 軌跡から制御パラメータを推定する逆問題の解法を検討してきた．文献 24) は，楽譜の音符の並びを表す階段状信号を隠れマルコフモデル（HMM）からの出力で表現し，そこに 2 次系のインパルス応答が畳み込まれて，オーバーシュートのような動的変動成分が表現された．一方，ビブラートやプレパレーション，微細変動成分のような音符の音高が安定するときの楽譜逸脱成分はすべてガウス性白色雑音で表現された．そして，モデルパラメータの推定アルゴリズムを導出することで，楽譜逸脱成分の特徴抽出を試みた．ただ，ビブラートや微細変動成分がすべてガウス性白色雑音としてモデル化されたため，歌唱者の意図的表現による動的変動成分（ビブラートやポルタメントなど）を微細変動成分と分離して特徴付けられなかった．

本稿では，歌唱の意図的表現を特徴抽出するために，ノート指令信号と表現指令信号によって駆動される歌声 F_0 軌跡の生成過程を提案する（**図 2**）．ここで，ノート指令信号は楽譜に記載される音符の並びを表現する．一方，表現指令信号は歌唱者の音楽的な表現意図を矩形状の細かい指令として表現する．ノート成分と表現成分はこれらの指令信号によって駆動されるノート制御機講と表現制御機講の出力である．ノート制御機講と表現制御機講はフィルタに相当し，2 次系で表現される．ノート制御機講はオーバーシュートなどのノート（音符）の立ち上がり方を制御する．表現制御機講は矩形状の細かい指令信号を制御してビブラートやポルタメントを生成する．ただし，表現制御機講は臨界制動系 ($\zeta = 1$ の場合) で構成される．微細変動成分は文献 22) にならって，10Hz 以上の不規則な振動成分を想定する．最終的に，対数スケールの F_0 軌跡 $y(t)$（ここで，t は時間を表す）は，これら 3 つの成分の重ね合わせであると想定する．この F_0 生成過程を想定した理由は 2 点ある．

- 話声の F_0 生成過程を記述する藤崎モデルでは，甲状軟骨の二つの独立な運動 (平行移動と回転) に伴う声帯の長さの変化の合計が F_0 の時間的変化をもたらすと解釈され，平行移動運動に関係する成分をフレーズ成分，回転運動に関係する成分をアクセント成分とした[23]．これらの運動が歌声にも存在すると仮定し，フレーズ成分のような大域的な変化を表すノート成分と，アクセント成分のような歌唱者が意図的に制御できる表現成分から歌声 F_0 軌跡が構成されると考えた．
- 従来，ビブラートは正弦波で表現され，そのパラメータは Vibrato rate（ビブラートの速さ）と Vibrato extent（ビブラートの深さ）と考えられた[11),22)]．これらのパラメータが指数的に変化するモデルも提案された[9)]．しかし実際に，様々な歌唱者による多様な歌声の F_0 を観察した結果，これらのパラメータは時々刻々と変化するものの，指数的な変化であるとは言えない．この複雑な振動現象をモデル化して歌い方を認識するために，意図を表す矩形状の細かい指令信号が臨界制動系によって制御されるビブラートの生成過程を想定した．

以降の節では，上記の歌声 F_0 軌跡の生成過程を離散時間表現して確率モデル化し，統計的手法に基づいてモデルパラメータの推定アルゴリズムを導出する．そして，提案モデルの有効性を客観的かつ主観的に評価する．文献 25) では，同じ枠組みで藤崎モデルを確率過程に基づいて統計モデル化する試みも行っているのでこちらも参照されたい．

Vol.2012-MUS-94 No.11
Vol.2012-SLP-90 No.11
2012/2/3

2. ノート指令と表現指令信号によって駆動される歌声 F_0 生成過程の確率モデル

歌声 F_0 軌跡の生成過程を離散時間表現し，確率過程に基づいて統計モデル化する．

2.1 F_0 生成過程の離散時間表現

連続時間領域で表現されるノート制御機講，表現制御機講の 2 次系の伝達関数の離散時間表現を得るために，従来法 24) にならって後退差分変換を利用する．後退差分変換は，時間微分演算子 s を z 領域における後退差分演算子 $s \simeq (1-z^{-1})/t_0$ に置き換える変換であり（t_0 は離散時間表現におけるサンプリング周期とする），この変換によりノート制御機構の逆システムの伝達関数 $\mathcal{H}_{\rm n}^{-1}(s)$ は z 領域で，

$$\mathcal{H}_{\rm n}^{-1}(z) = a_2 z^{-2} + a_1 z^{-1} + a_0 \tag{2}$$

と書くことができる．ただし，

$$a_2 = \varphi^2,\ a_1 = -2\varphi(\psi+\varphi),\ a_0 = 1 + 2\varphi\psi + \varphi^2 \tag{3}$$

および，φ, ψ は $\varphi = 1/(\Omega t_0)$, $\psi = \zeta$ と表現される．ここで，k を離散時刻インデックスとし，ノート指令信号およびノート成分の離散時間表現をそれぞれ $u_{\rm n}[k]$，$y_{\rm n}[k]$ とすると，$y_{\rm n}[k]$ は，ノート制御パラメータ φ, ψ によって特性が決まる拘束つき全極モデルからの出力

$$u_{\rm n}[k] = a_0 y_{\rm n}[k] + a_1 y_{\rm n}[k-1] + a_2 y_{\rm n}[k-2] \tag{4}$$

と見なすことができる．

同様に，表現指令信号 $u_{\rm e}[k]$ と表現成分 $y_{\rm e}[k]$ の関係も

$$u_{\rm e}[k] = b_0 y_{\rm e}[k] + b_1 y_{\rm e}[k-1] + b_2 y_{\rm e}[k-2] \tag{5}$$

と書くことができる．ただし，$b_2 = \xi^2$, $b_1 = -2\xi(1+\xi)$, $b_0 = 1+2\xi+\xi^2$ であり，表現制御パラメータ ξ は $\xi = 1/(\Omega t_0)$ と表現される．微細変動成分 $y_{\rm f}(t)$ の離散時間表現を $y_{\rm f}[k]$ とすると，提案モデルによる歌声 F_0 軌跡の離散時間表現は，これら 3 つの成分の和

$$y[k] = y_{\rm n}[k] + y_{\rm e}[k] + y_{\rm f}[k] \tag{6}$$

で与えられる．

2.2 F_0 生成過程の確率モデル化

ノート指令信号と表現指令信号はそれぞれ，楽譜に記載されるメロディの音符の並びと歌唱者の音楽的な表現意図を表す (図 2)．これらの指令信号を表現するために，HMM を利用して，$u_{\rm n}[k]$ と $u_{\rm e}[k]$ を確率モデル化する．$\boldsymbol{o}[k] := (u_{\rm n}[k], u_{\rm e}[k])^{\rm T}$ を

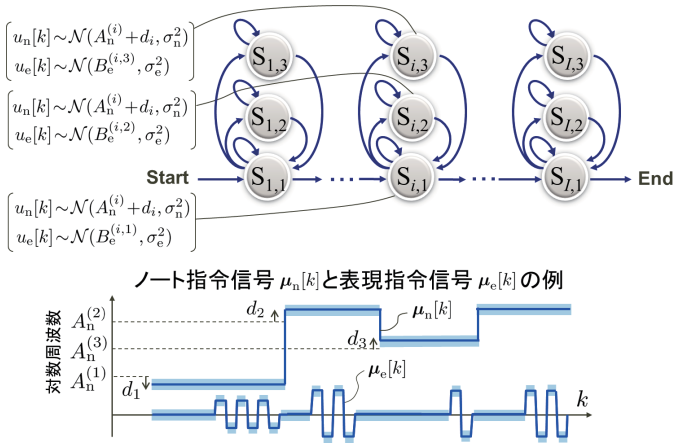

図 3 隠れマルコフモデルによる指令信号の統計モデル化

$$\boldsymbol{o}[k] \sim \mathcal{N}(\boldsymbol{\nu}[k], \boldsymbol{\Upsilon}),\quad \boldsymbol{\nu}[k] := \begin{bmatrix} \mu_{\rm n}[k] \\ \mu_{\rm e}[k] \end{bmatrix},\ \boldsymbol{\Upsilon} := \begin{bmatrix} \sigma_{\rm n}^2 & 0 \\ 0 & \sigma_{\rm e}^2 \end{bmatrix} \tag{7}$$

のように正規分布する確率変数と見なし，平均 $\boldsymbol{\nu}[k]$ が**図 3** のような状態遷移に伴って変化するモデルを考える．このように $\boldsymbol{o}[k]$ を HMM でモデル化することにより，状態遷移の経路制限（状態遷移確率の設定）を通して，$\boldsymbol{\nu}[k]$ に制約を与えることが可能となる．

具体的には，この HMM は $I \times J$ 個の状態集合 $\mathcal{S} := \{{\rm S}_{i,j}\}_{i=1,j=1}^{I,J}$ からなる．状態 ${\rm S}_{i,j}$ では，$\mu_{\rm n}[k]$ は $A_{\rm n}^{(i)} + d_i$ の値をとる．ここで，$A_{\rm n}^{(i)}$ は楽譜に記載されるメロディの i 番目の音符の音高を表し（本稿では，この値は楽譜から与えられるものとする），d_i はその音高からの推移（音符の絶対音高からのずれであり，音高シフトパラメータと呼ぶ），I は歌唱するメロディに含まれる音符の総数を表す．図 3 に示す状態遷移により，$\mu_{\rm n}[k]$ は I 個の音符区間からなる階段状信号を表現する．一方，$\mu_{\rm e}[k]$ は $B_{\rm e}^{(i,j)}$ の値をとり，これは歌唱者の表現意図を表すための，矩形の表現指令の大きさを表す．ここで，i 番目の音符では，${\rm S}_{i,1}$ を通らずして，状態 ${\rm S}_{i,j}$ から別の状態 ${\rm S}_{i,j'}$ $(j \neq j',\ 2 \leq j \leq J,\ 2 \leq j' \leq J)$ へ直接に遷移できない制約を設けることによって，$\mu_{\rm e}[k]$ は図 3 に示すような矩形信号を表現する．J は表現指令を構成するための状態数を表す（図 3 は $J = 3$ である）．指令信号を生成する HMM

Vol.2012-MUS-94 No.11
Vol.2012-SLP-90 No.11
2012/2/3

（指令信号モデル）の構成をまとめると，以下のとおりである．

出力系列 : $\{\boldsymbol{o}[k]\}_{k=1}^{K}$
状態集合 : $\mathcal{S} := \{\mathrm{S}_{i,j}\}_{i=1,j=1}^{I,J}$
状態系列 : $\{s_k\}_{k=1}^{K}$
出力確率分布 : $P(\boldsymbol{o}[k]|s_k = n) = \mathcal{N}(\boldsymbol{c}_n[k], \boldsymbol{\Upsilon}),\ \boldsymbol{c}_n[k] = (A_{\mathrm{n}}^{(i)} + d_i,\ B_{\mathrm{e}}^{(i,j)})^{\mathrm{T}} \quad (n = \mathrm{S}_{i,j})$
状態遷移確率 : $\phi_{n',n} := \log P(s_k = n|s_{k-1} = n')$

簡単のため，状態遷移確率 $\phi_{n',n}$ は定数とすると，指令信号モデルにおいて推定すべきパラメータは，HMM の経路を表す状態系列 $\{s_k\}_{k=1}^{K}$，音高シフトパラメータ $\{d_i\}_{i=1}^{I}$，表現指令の大きさパラメータ $\{B_{\mathrm{e}}^{(i,j)}\}_{i=1,j=1}^{I,J}$，指令信号の分散パラメータ $\sigma_{\mathrm{n}}^2,\ \sigma_{\mathrm{e}}^2$ であり，

$$\theta_{\mathrm{u}} := \left\{\{s_k\}_{k=1}^{K},\ \{d_i\}_{i=1}^{I},\ \{B_{\mathrm{e}}^{(i,j)}\}_{i=1,j=1}^{I,J},\ \sigma_{\mathrm{n}}^2,\ \sigma_{\mathrm{e}}^2\right\} \tag{8}$$

と記述する．また，平均値系列 $\{\mu_{\mathrm{n}}[k]\}_{k=1}^{K}$ および $\{\mu_{\mathrm{e}}[k]\}_{k=1}^{K}$ は，状態系列 $\{s_k\}_{k=1}^{K}$ が与えられたもとで，$(\mu_{\mathrm{n}}[k],\ \mu_{\mathrm{e}}[k])^{\mathrm{T}} \leftarrow \boldsymbol{c}_{s_k}[k]$ で与えられる．

指令信号モデルに基づいて，$\boldsymbol{y} = (y[1], \ldots, y[K])^{\mathrm{T}}$ の確率密度関数を導く．式 (7) より，$\boldsymbol{u}_{\mathrm{n}} := (u_{\mathrm{n}}[1], \ldots, u_{\mathrm{n}}[K])^{\mathrm{T}},\ \boldsymbol{u}_{\mathrm{e}} := (u_{\mathrm{e}}[1], \ldots, u_{\mathrm{e}}[K])^{\mathrm{T}},\ \boldsymbol{\mu}_{\mathrm{n}} := (\mu_{\mathrm{n}}[1], \ldots, \mu_{\mathrm{n}}[K])^{\mathrm{T}},\ \boldsymbol{\mu}_{\mathrm{e}} := (\mu_{\mathrm{e}}[1], \ldots, \mu_{\mathrm{e}}[K])^{\mathrm{T}}$ とすると

$$\boldsymbol{u}_{\mathrm{n}}|\theta_u \sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{n}}, \boldsymbol{\Sigma}_{\mathrm{n}}), \quad \boldsymbol{\Sigma}_{\mathrm{n}} = \sigma_{\mathrm{n}}^2 \boldsymbol{I}_K \tag{9}$$

$$\boldsymbol{u}_{\mathrm{e}}|\theta_u \sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{e}}, \boldsymbol{\Sigma}_{\mathrm{e}}), \quad \boldsymbol{\Sigma}_{\mathrm{e}} = \sigma_{\mathrm{e}}^2 \boldsymbol{I}_K \tag{10}$$

が言える．ここで，$\boldsymbol{I}_K$ は $K \times K$ の単位行列を表す．ノート成分 $\boldsymbol{y}_{\mathrm{n}} := (y_{\mathrm{n}}[1], \ldots, y_{\mathrm{n}}[K])^{\mathrm{T}}$ とノート指令信号 $\boldsymbol{u}_{\mathrm{n}}$ の関係，および表現成分 $\boldsymbol{y}_{\mathrm{e}} := (y_{\mathrm{e}}[1], \ldots, y_{\mathrm{e}}[K])^{\mathrm{T}}$ と表現指令信号 $\boldsymbol{u}_{\mathrm{e}}$ の関係は，

$$\boldsymbol{A} := \begin{bmatrix} \boldsymbol{A}^{(1)} \\ \vdots \\ \boldsymbol{A}^{(i)} \\ \vdots \\ \boldsymbol{A}^{(I)} \end{bmatrix} = \begin{bmatrix} a_0^{(1)} & & & & O \\ a_1^{(1)} & a_0^{(1)} & & & \\ \ddots & \ddots & \ddots & & \\ & a_2^{(i)} & a_1^{(i)} & a_0^{(i)} & \\ & & \ddots & \ddots & \ddots \\ O & & & a_2^{(I)} & a_1^{(I)} \quad a_0^{(I)} \end{bmatrix},\ \boldsymbol{B} := \begin{bmatrix} b_0 & & & & O \\ b_1 & b_0 & & & \\ b_2 & b_1 & b_0 & & \\ & \ddots & \ddots & \ddots & \\ O & & b_2 & b_1 & b_0 \end{bmatrix} \tag{11}$$

と置くと，それぞれ，

$$\boldsymbol{u}_{\mathrm{n}} = \boldsymbol{A}\boldsymbol{y}_{\mathrm{n}}, \quad \boldsymbol{u}_{\mathrm{e}} = \boldsymbol{B}\boldsymbol{y}_{\mathrm{e}} \tag{12}$$

と表現できる．ここでは，先行研究で得られた知見 22) に基づいて，ノート制御パラメータ $\varphi,\ \psi$ がノートごとに異なるものと想定し，式 (11) の $a_0^{(i)},\ a_1^{(i)},\ a_2^{(i)}$ は，式 (3) より

$$a_2^{(i)} = (\varphi^{(i)})^2,\ a_1^{(i)} = -2\varphi^{(i)}(\psi^{(i)} + \varphi^{(i)}),\ a_0^{(i)} = 1 + 2\varphi^{(i)}\psi^{(i)} + (\varphi^{(i)})^2 \tag{13}$$

と表現され，$\{\varphi^{(i)},\ \psi^{(i)}\}_{i=1}^{I}$ を制御パラメータとする．ノート成分と表現成分の確率密度関数は，

$$\boldsymbol{y}_{\mathrm{n}}|\theta_u, \{\varphi^{(i)}, \psi^{(i)}\}_{i=1}^{I} \sim \mathcal{N}(\boldsymbol{A}^{-1}\boldsymbol{\mu}_{\mathrm{n}}, \boldsymbol{A}^{-1}\boldsymbol{\Sigma}_{\mathrm{n}}(\boldsymbol{A}^{-1})^{\mathrm{T}}) \tag{14}$$

$$\boldsymbol{y}_{\mathrm{e}}|\theta_u, \xi \sim \mathcal{N}(\boldsymbol{B}^{-1}\boldsymbol{\mu}_{\mathrm{e}}, \boldsymbol{B}^{-1}\boldsymbol{\Sigma}_{\mathrm{e}}(\boldsymbol{B}^{-1})^{\mathrm{T}}) \tag{15}$$

が導出される．微細変動成分 $\boldsymbol{y}_{\mathrm{f}} := (y_{\mathrm{f}}[1], \ldots, y_{\mathrm{f}}[K])^{\mathrm{T}}$ はガウス性白色雑音を想定する．

$$\boldsymbol{y}_{\mathrm{f}}|\sigma_{\mathrm{f}}^2 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\mathrm{f}}), \quad \boldsymbol{\Sigma}_{\mathrm{f}} = \sigma_{\mathrm{f}}^2 \boldsymbol{I}_K \tag{16}$$

仮定より，$\boldsymbol{y}_{\mathrm{n}},\ \boldsymbol{y}_{\mathrm{e}},\ \boldsymbol{y}_{\mathrm{f}}$ は独立なので，$\Theta := \{\theta_{\mathrm{u}},\ \{\varphi^{(i)},\ \psi^{(i)}\}_{i=1}^{I},\ \xi,\ \sigma_{\mathrm{f}}^2\}$ が与えられた下での歌声 F_0 軌跡 $\boldsymbol{y} = \boldsymbol{y}_{\mathrm{n}} + \boldsymbol{y}_{\mathrm{e}} + \boldsymbol{y}_{\mathrm{f}}$ の確率密度関数は式 (14)，(15) と式 (16) より，

$$\boldsymbol{y}|\Theta \sim \mathcal{N}\left(\boldsymbol{A}^{-1}\boldsymbol{\mu}_{\mathrm{n}} + \boldsymbol{B}^{-1}\boldsymbol{\mu}_{\mathrm{e}},\ \boldsymbol{A}^{-1}\boldsymbol{\Sigma}_{\mathrm{n}}(\boldsymbol{A}^{-1})^{\mathrm{T}} + \boldsymbol{B}^{-1}\boldsymbol{\Sigma}_{\mathrm{e}}(\boldsymbol{B}^{-1})^{\mathrm{T}} + \boldsymbol{\Sigma}_{\mathrm{f}}\right) \tag{17}$$

で与えられる．以上より，

$$P(\boldsymbol{y}|\Theta) = \frac{|\boldsymbol{\Sigma}^{-1}|^{1/2}}{(2\pi)^{K/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\right\} \tag{18}$$

$$\boldsymbol{\mu} = \boldsymbol{A}^{-1}\boldsymbol{\mu}_{\mathrm{n}} + \boldsymbol{B}^{-1}\boldsymbol{\mu}_{\mathrm{e}},\ \boldsymbol{\Sigma} = \boldsymbol{A}^{-1}\boldsymbol{\Sigma}_{\mathrm{n}}(\boldsymbol{A}^{-1})^{\mathrm{T}} + \boldsymbol{B}^{-1}\boldsymbol{\Sigma}_{\mathrm{e}}(\boldsymbol{B}^{-1})^{\mathrm{T}} + \boldsymbol{\Sigma}_{\mathrm{f}}$$

が歌声 F_0 軌跡 $\boldsymbol{y}$ が与えられたときのモデルパラメータ Θ の尤度関数である．

Θ の事前確率については，各パラメータは互いに独立で，パラメータ $\sigma_{\mathrm{n}}^2,\ \sigma_{\mathrm{e}}^2,\ \sigma_{\mathrm{f}}^2$ は一様に分布すると仮定し，以下のように記述する．

$$P(\Theta) \propto P(\varphi^{(1)}) \ldots P(\varphi^{(I)}) P(\psi^{(1)}) \ldots P(\psi^{(I)}) P(\xi)$$
$$P(d_1) \ldots P(d_I) P(B_{\mathrm{e}}^{(1,1)}) \ldots P(B_{\mathrm{e}}^{(I,J)}) P(s_1) \prod_{k=2}^{K} P(s_k|s_{k-1}) \tag{19}$$

3. パラメータ推定アルゴリズム

$\boldsymbol{y}$ が与えられた下で，Θ の事後確率 $P(\Theta|\boldsymbol{y})$ を最大化する問題は解析的に解くことはできないが，$\boldsymbol{x} := (\boldsymbol{y}_{\mathrm{n}}^{\mathrm{T}},\ \boldsymbol{y}_{\mathrm{e}}^{\mathrm{T}},\ \boldsymbol{y}_{\mathrm{f}}^{\mathrm{T}})^{\mathrm{T}}$ を完全データと見なすことで，EM アルゴリズムによる不完全データ問題に帰着できる．不完全データと完全データの関係は $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}$ となり，ここ

Vol.2012-MUS-94 No.11
Vol.2012-SLP-90 No.11
2012/2/3

で $\boldsymbol{H} := [\boldsymbol{I}_K\ \boldsymbol{I}_K\ \boldsymbol{I}_K]$ とする．この場合，完全データの対数尤度は，

$$\log P(\boldsymbol{x}|\Theta) \stackrel{\mathrm{c}}{=} \frac{1}{2}\log|\boldsymbol{\Lambda}^{-1}| - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}-\boldsymbol{m}) \tag{20}$$

$$\boldsymbol{m} := \begin{bmatrix} \boldsymbol{A}^{-1}\boldsymbol{\mu}_{\mathrm{n}} \\ \boldsymbol{B}^{-1}\boldsymbol{\mu}_{\mathrm{e}} \\ \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{\Lambda}^{-1} := \begin{bmatrix} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{n}}^{-1}\boldsymbol{A} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{e}}^{-1}\boldsymbol{B} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{\Sigma}_{\mathrm{f}}^{-1} \end{bmatrix}$$

で与えられる．このとき，Q 関数 $Q(\Theta,\Theta')$ は，

$$\begin{aligned} Q(\Theta,\Theta') \stackrel{\mathrm{c}}{=} \frac{1}{2}\Big[&\log|\boldsymbol{\Lambda}^{-1}| - \mathrm{tr}(\boldsymbol{\Lambda}^{-1}\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{y};\Theta']) \\ &+ 2\boldsymbol{m}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta'] - \boldsymbol{m}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\boldsymbol{m}\Big] + \log P(\Theta) \end{aligned} \tag{21}$$

となる．ここで，$\mathrm{tr}(\cdot)$ は行列のトレースを表し，$\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta']$ と $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{y};\Theta']$ は条件付きガウス分布の性質より，

$$\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta'] = \boldsymbol{m} + \boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}})^{-1}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{m}) \tag{22}$$

$$\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{y};\Theta'] = \boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}})^{-1}\boldsymbol{H}\boldsymbol{\Lambda} + \mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta']\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta']^{\mathrm{T}} \tag{23}$$

と書ける．E ステップでは，直前のステップで更新されたパラメータを Θ' に代入し，上記に基づいて $\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta']$ と $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{y};\Theta']$ が算出される．$\boldsymbol{y}_{\mathrm{n}}$，$\boldsymbol{y}_{\mathrm{e}}$，$\boldsymbol{y}_{\mathrm{f}}$ に対応するように，$\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta']$ および $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{y};\Theta']$ を

$$\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta'] = \begin{bmatrix} \bar{\boldsymbol{x}}_{\mathrm{n}} \\ \bar{\boldsymbol{x}}_{\mathrm{e}} \\ \bar{\boldsymbol{x}}_{\mathrm{f}} \end{bmatrix}, \quad \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{y};\Theta'] = \begin{bmatrix} \boldsymbol{R}_{\mathrm{n}} & * & * \\ * & \boldsymbol{R}_{\mathrm{e}} & * \\ * & * & \boldsymbol{R}_{\mathrm{f}} \end{bmatrix} \tag{24}$$

のように区分表現すると，Q 関数は

$$\begin{aligned} Q(\Theta,\Theta') \stackrel{\mathrm{c}}{=} \frac{1}{2}\Big[&\log|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{n}}^{-1}\boldsymbol{A}| + \log|\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{e}}^{-1}\boldsymbol{B}| + \log|\boldsymbol{\Sigma}_{\mathrm{f}}^{-1}| \\ &- \mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{n}}^{-1}\boldsymbol{A}\boldsymbol{R}_{\mathrm{n}}) + 2\boldsymbol{\mu}_{\mathrm{n}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{n}}^{-1}\boldsymbol{A}\bar{\boldsymbol{x}}_{\mathrm{n}} - \boldsymbol{\mu}_{\mathrm{n}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{n}}^{-1}\boldsymbol{\mu}_{\mathrm{n}} \\ &- \mathrm{tr}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{e}}^{-1}\boldsymbol{B}\boldsymbol{R}_{\mathrm{e}}) + 2\boldsymbol{\mu}_{\mathrm{e}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{e}}^{-1}\boldsymbol{B}\bar{\boldsymbol{x}}_{\mathrm{e}} - \boldsymbol{\mu}_{\mathrm{e}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{e}}^{-1}\boldsymbol{\mu}_{\mathrm{e}} \\ &- \mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{f}}^{-1}\boldsymbol{R}_{\mathrm{f}})\Big] + \log P(\Theta) \end{aligned} \tag{25}$$

と書き直せて，各パラメータについて M ステップの更新式を求めることができる．

1) 状態系列: Q 関数の中で $s := \{s_k\}_{k=1}^K$ に関係する項は

$$\mathcal{I}_1(s) := -\frac{1}{2}\sum_{k=1}^{K}(\boldsymbol{o}[k]-\boldsymbol{c}_{s_k}[k])^{\mathrm{T}}\boldsymbol{\Upsilon}^{-1}(\boldsymbol{o}[k]-\boldsymbol{c}_{s_k}[k]) + \log P(s_1) + \sum_{k=2}^{K}\log P(s_k|s_{k-1})$$

となる．ここで，$\boldsymbol{o}[k] := ([\boldsymbol{A}\bar{\boldsymbol{x}}_{\mathrm{n}}]_k, [\boldsymbol{B}\bar{\boldsymbol{x}}_{\mathrm{e}}]_k)^{\mathrm{T}}$ であり，$[\cdot]_k$ はベクトルの k 番目の要素を表す．これを最大化する状態系列 $\{s_k\}_{k=1}^K$ は動的計画法により効率的に解くことができる．まず，状態 $S_{1,1}$ について

$$\delta_1(\mathrm{S}_{1,1}) = -\frac{1}{2}(\boldsymbol{o}[1]-\boldsymbol{c}_{\mathrm{S}_{1,1}}[1])^{\mathrm{T}}\boldsymbol{\Upsilon}^{-1}(\boldsymbol{o}[1]-\boldsymbol{c}_{\mathrm{S}_{1,1}}[1]) + \log P(\mathrm{S}_{1,1}) \tag{26}$$

とおくと，$k=2,3,\ldots,K$ について逐次的に $\delta_k(S_{i,j})$ を

$$\delta_k(\mathrm{S}_{i,j}) = \max_{\mathrm{S}_{i',j'}}\Big[\delta_{k-1}(\mathrm{S}_{i',j'}) - \frac{1}{2}(\boldsymbol{o}[k]-\boldsymbol{c}_{\mathrm{S}_{i,j}}[k])^{\mathrm{T}}\boldsymbol{\Upsilon}^{-1}(\boldsymbol{o}[k]-\boldsymbol{c}_{\mathrm{S}_{i,j}}[k]) + \phi_{\mathrm{S}_{i',j'},\mathrm{S}_{i,j}}\Big] \tag{27}$$

により計算できる．各ステップで選択される状態番号

$$\psi_k(\mathrm{S}_{i,j}) = \operatorname*{argmax}_{\mathrm{S}_{i',j'}}\Big[\delta_{k-1}(\mathrm{S}_{i',j'}) + \phi_{\mathrm{S}_{i',j'},\mathrm{S}_{i,j}}\Big] \tag{28}$$

を記憶しておくことで，$k=K$ まで到着後に $s_{k-1} = \psi_k(s_k)$ $(k = K, K-1, \ldots, 2)$ により選択された状態番号を辿っていくと最適経路 $s_1,\ldots,s_K$ を得る．

2) ノート制御パラメータ: $\varphi^{(i)}$ と $\psi^{(i)}$ に関する事前分布を $\varphi^{(i)} \sim \mathcal{N}(\mu_\varphi, \sigma_\varphi^2)$，$\psi^{(i)} \sim \mathcal{N}(\mu_\psi, \sigma_\psi^2)$ とする．Q 関数の中で $\varphi^{(i)}$ と $\psi^{(i)}$ に関係する項は，

$$\begin{aligned} \mathcal{I}_2(\varphi^{(i)},\psi^{(i)}) = &|\mathcal{T}_{S_{i,\cdot}}|\log(1+2\varphi^{(i)}\psi^{(i)}+(\varphi^{(i)})^2) - \frac{1}{2\sigma_{\mathrm{n}}^2}\mathrm{tr}((\boldsymbol{A}^{(i)})^{\mathrm{T}}\boldsymbol{A}^{(i)}\boldsymbol{R}_{\mathrm{n}}) \\ &+ \frac{1}{\sigma_{\mathrm{n}}^2}([\boldsymbol{\mu}_{\mathrm{n}}]_{\mathcal{T}_{S_{i,\cdot}}})^{\mathrm{T}}\boldsymbol{A}^{(i)}\bar{\boldsymbol{x}}_{\mathrm{n}} - \frac{1}{2\sigma_\varphi^2}(\varphi^{(i)}-\mu_\varphi)^2 - \frac{1}{2\sigma_\psi^2}(\psi^{(i)}-\mu_\psi)^2 \\ &\mathcal{T}_{\mathrm{S}_{i,\cdot}} = \{k|s_k \in \{\mathrm{S}_{i,1},\ldots,\mathrm{S}_{i,J}\}\} \end{aligned} \tag{29}$$

となる．ここで，$|\mathcal{T}|$ は集合 $\mathcal{T}$ の要素数を表す．また，$[\boldsymbol{\mu}]_{\mathcal{T}}$ は，集合 $\mathcal{T}$ の要素を添え字として，その添え字に相当する $\boldsymbol{\mu}$ の要素を取り出した部分ベクトルを表す．今，

$$\boldsymbol{U}_2 := \begin{bmatrix} 1 & & & & O \\ -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ O & & 1 & -2 & 1 \end{bmatrix}, \quad \boldsymbol{U}_1 := \begin{bmatrix} 2 & & & & O \\ -2 & 2 & & & \\ 0 & -2 & 2 & & \\ & \ddots & \ddots & \ddots & \\ O & & 0 & -2 & 2 \end{bmatrix}, \quad \boldsymbol{U}_0 := \begin{bmatrix} 1 & & & & O \\ 0 & 1 & & & \\ 0 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ O & & 0 & 0 & 1 \end{bmatrix}$$

Vol.2012-MUS-94 No.11
Vol.2012-SLP-90 No.11
2012/2/3

として，式 (4) から，$\boldsymbol{A}^{(i)}$ は，

$$\boldsymbol{A}^{(i)} = (\varphi^{(i)})^2 [\boldsymbol{U}_2]_{\mathcal{T}_{\mathrm{S}_{i,\cdot}}} + \varphi^{(i)}\psi^{(i)} [\boldsymbol{U}_1]_{\mathcal{T}_{\mathrm{S}_{i,\cdot}}} + [\boldsymbol{U}_0]_{\mathcal{T}_{\mathrm{S}_{i,\cdot}}} \tag{30}$$

と表現される．ここで，$[\boldsymbol{U}]_{\mathcal{T}}$ は集合 $\mathcal{T}$ の要素を添え字として，行列 $\boldsymbol{U}$ からその添え字に相当する行ベクトルを取り出して構成される部分行列を意味する．ニュートン・ラフソン法を利用して，$\mathcal{I}_2(\varphi^{(i)}, \psi^{(i)})$ を最大化する $\varphi^{(i)}$ と $\psi^{(i)}$ が数値的に導出される．

3) 表現制御パラメータ: ξ に関する事前分布を $\xi \sim \mathcal{N}(\mu_\xi, \sigma_\xi^2)$ とする．Q 関数の中で ξ に関係する項は，

$$\mathcal{I}_3(\xi) = K \log(1 + 2\xi + \xi^2) - \frac{1}{2\sigma_{\mathrm{e}}^2} \mathrm{tr}(\boldsymbol{B}^{\mathrm{T}} \boldsymbol{B} \boldsymbol{R}_{\mathrm{e}}) + \frac{1}{\sigma_{\mathrm{e}}^2} \boldsymbol{\mu}_{\mathrm{e}}^{\mathrm{T}} \boldsymbol{B} \bar{\boldsymbol{x}}_{\mathrm{e}} - \frac{1}{2\sigma_\xi^2} (\xi - \mu_\xi)^2 \tag{31}$$

となる．ニュートン・ラフソン法を利用して，$\mathcal{I}_3(\xi)$ を最大化する ξ が数値的に導出される．

4) その他のパラメータ: d_i と $B_{\mathrm{e}}^{(i,j)}$ に関して，それぞれ事前分布を $d_i \sim \mathcal{N}(0, \sigma_{\mathrm{d}}^2)$ と $B_{\mathrm{e}}^{(i,j)} \sim \mathcal{N}(\mu_{B^{(i,j)}}, \sigma_B^2)$ とする．残されたパラメータの更新式は

$$d_i = \frac{1}{|\mathcal{T}_{\mathrm{S}_{i,\cdot}}| + \sigma_{\mathrm{n}}^2/\sigma_{\mathrm{d}}^2} \sum_{k \in \mathcal{T}_{\mathrm{S}_{i,\cdot}}} ([\boldsymbol{A}\bar{\boldsymbol{x}}_{\mathrm{n}}]_k - A_{\mathrm{n}}^{(i)}) \tag{32}$$

$$B_{\mathrm{e}}^{(i,j)} = \frac{1}{|\mathcal{T}_{\mathrm{S}_{i,j}}|/\sigma_{\mathrm{e}}^2 + 1/\sigma_B^2} \Big(\sum_{k \in \mathcal{T}_{\mathrm{S}_{i,j}}} \frac{[\boldsymbol{B}\bar{\boldsymbol{x}}_{\mathrm{e}}]_k}{\sigma_{\mathrm{e}}^2} + \frac{\mu_{B^{(i,j)}}}{\sigma_B^2} \Big), \quad \mathcal{T}_{\mathrm{S}_{i,j}} = \{k | s_k = \mathrm{S}_{i,j}\} \tag{33}$$

$$\sigma_{\mathrm{n}}^2 = \left(\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{R}_{\mathrm{n}}) - 2\boldsymbol{\mu}_{\mathrm{n}}^{\mathrm{T}} \boldsymbol{A} \bar{\boldsymbol{x}}_{\mathrm{n}} + \boldsymbol{\mu}_{\mathrm{n}}^{\mathrm{T}} \boldsymbol{\mu}_{\mathrm{n}} \right) / K \tag{34}$$

$$\sigma_{\mathrm{e}}^2 = \left(\mathrm{tr}(\boldsymbol{B}^{\mathrm{T}} \boldsymbol{B} \boldsymbol{R}_{\mathrm{e}}) - 2\boldsymbol{\mu}_{\mathrm{e}}^{\mathrm{T}} \boldsymbol{B} \bar{\boldsymbol{x}}_{\mathrm{e}} + \boldsymbol{\mu}_{\mathrm{e}}^{\mathrm{T}} \boldsymbol{\mu}_{\mathrm{e}} \right) / K \tag{35}$$

$$\sigma_{\mathrm{f}}^2 = \mathrm{tr}(\boldsymbol{R}_{\mathrm{f}}) / K \tag{36}$$

と導出される．

4. 実装方法

前節で導出したパラメータ推定アルゴリズムにおける実際の実装方法について述べる．

4.1 実測 F_0 軌跡からパラメータ推定方法

提案アルゴリズムを用いて，実測 F_0 軌跡からパラメータを推定する問題を考える．式 (18) では，全区間で F_0 が観測されていることが暗に想定されているが，実際には F_0 は有声音が発せられるときのみ観測可能であり，無声音の区間では F_0 は通常観測できない．したがって，実測 F_0 軌跡からのパラメータ推定を行うためには一般に欠損データの問題を扱う必要がある．この類の問題は不完全データ問題に他ならず，EM アルゴリズムにより扱うことができる．文献 25) を参照して，$\boldsymbol{y} \in \mathbb{R}^K$ を実測 F_0 データと欠損 F_0 データからなる完全データとし，実測 F_0 データを並べたベクトルを $\boldsymbol{y}_{\mathrm{obs}} \in \mathbb{R}^{K'}$ $(K' \leq K)$ とする．$\boldsymbol{y}$ と $\boldsymbol{y}_{\mathrm{obs}}$ との関係は，各行に 1 が一個あり残りはすべて 0 であるような $K' \times K$ のバイナリ行列 $\boldsymbol{M}$ を用いて $\boldsymbol{y}_{\mathrm{obs}} = \boldsymbol{M}\boldsymbol{y}$ と表される．これを利用して EM の各ステップを導くことができる．導出は割愛するが，E ステップでは $\boldsymbol{y}$ を

$$\boldsymbol{y} \leftarrow \boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{M}^{\mathrm{T}} (\boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}^{\mathrm{T}})^{-1} (\boldsymbol{y}_{\mathrm{obs}} - \boldsymbol{M}\boldsymbol{\mu}) \tag{37}$$

により更新し，M ステップではその $\boldsymbol{y}$ を用いて

$$\Theta \leftarrow \operatorname*{argmax}_{\Theta} \log P(\boldsymbol{y}|\Theta) P(\Theta) \tag{38}$$

を実行すれば良い．なお自明ではあるが，欠損区間がない場合を表す $\boldsymbol{M} = \boldsymbol{I}_K$ においては式 (37) は $\boldsymbol{y} \leftarrow \boldsymbol{y}_{\mathrm{obs}}$ となり，$\boldsymbol{y}_{\mathrm{obs}}$ をそのまま完全データと見なして良いという意味になる．

4.2 パラメータの更新方法

EM 法に基づいてパラメータ推定アルゴリズムを導出したものの，実際は推定すべきパラメータ数が多いため，各パラメータを反復更新しても，局所解に収束してしまう．これを抑制するために，本稿では表現制御パラメータ ξ を固定する．したがって，意図的表現に関する情報は表現指令信号のみを推定する．また，以下の順序で分散パラメータを小さな値に固定する制約の下で，各パラメータを反復して更新する．

(1) σ_{e}^2, σ_{f}^2 を固定して（例えば，どちらも 1），E ステップと M ステップを反復する．
(2) σ_{n}^2, σ_{f}^2 を固定して，E ステップと M ステップを反復する．
(3) σ_{n}^2, σ_{e}^2 を固定して，E ステップと M ステップを反復する．

5. 評価実験

提案モデルの有効性を，人工的に作成した F_0 軌跡，および実測の F_0 軌跡を用いて客観的かつ主観的に評価する．

5.1 人工的に作成した F_0 軌跡を用いた評価

まず，パラメータ推定アルゴリズムの挙動を調べるために，文献 22) の歌声 F_0 制御モデルを用いて，人工的に作成した F_0 軌跡をテストデータとして動作実験を行った．事前に童謡や歌謡曲から抜粋したメロディの MIDI（合計 16 種類，抜粋部分の長さは平均 11.4 秒）に基づいて，対数周波数 (単位は cent，半音は 100cent に相当する) 軸上で階段状に変化する音符列を作成する（サンプリング周期は 5ms，休符区間はその前の音符の音高

Vol.2012-MUS-94 No.11
Vol.2012-SLP-90 No.11
2012/2/3

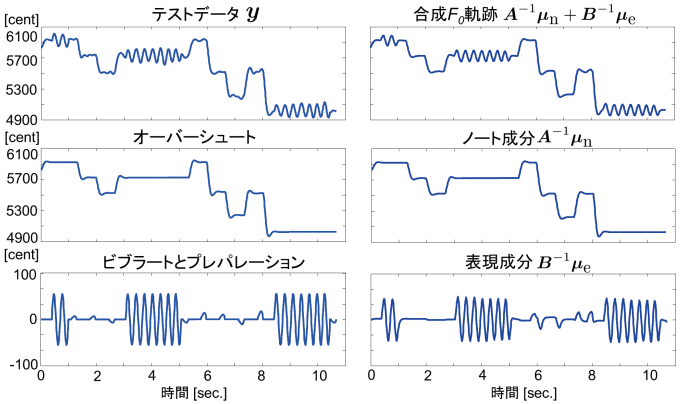

図 4 人工的に作成した F_0 軌跡に対するパラメータ推定結果

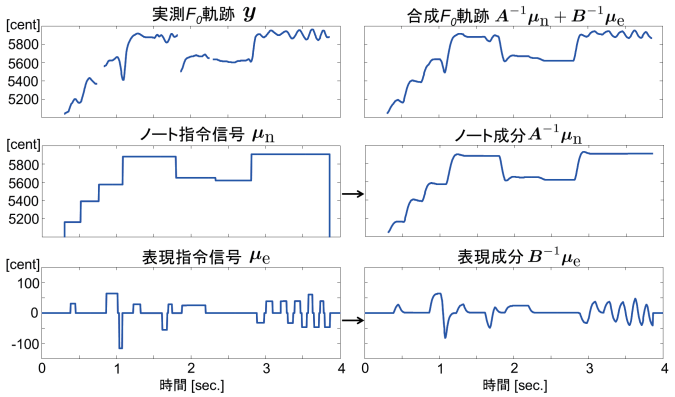

図 5 実測 F_0 軌跡に対するパラメータ推定結果

を伸ばす）．これに，文献 22) で調査されたプロ歌唱者の F_0 軌跡を生成するためのパラメータを用いて，楽譜逸脱成分を付加する (**図 4** の左側)．具体的には，オーバーシュートは $\zeta = 0.570,\ \Omega = 0.0363$ [rad/ms] とし，プレパレーションは $\zeta = 0.675,\ \Omega = 0.0308$ [rad/ms] とし，ビブラートは Vibrato extent = 83.0 cent, Vibrato rate = 6.25 Hz とし，微細変動成分は白色雑音をカットオフ周波数 10 Hz のローパスフィルタに通し，振幅の平均値を 20 cent に調整したものとし，キーシフトは+25cent，0cent，−25cent の 3 パターンとした．以上より，合計 48 個のテストデータを得る．

一方，パラメータ推定アルゴリズムの設定値を以下に示す．前節の (1)〜(3) における各反復回数は 1000 回とした．I と $\{A_{\mathrm{n}}^{(i)}\}_{i=1}^{I}$ は抜粋したメロディの MIDI 情報から与えられる．表現指令信号を構成するための状態数 J は 5 とした．HMM の状態遷移確率は，$\phi_{\mathrm{S}_{i,1},\mathrm{S}_{i,1}} = \log(0.9999 \times (J-1)/J)$，$\phi_{\mathrm{S}_{i,1},\mathrm{S}_{i,j}} = \log(0.9999/J)$，$\phi_{\mathrm{S}_{i,1},\mathrm{S}_{i+1,1}} = \log(0.0001)$，$\phi_{\mathrm{S}_{i,j},\mathrm{S}_{i,j}} = \log(0.9999)$，$\phi_{\mathrm{S}_{i,j},\mathrm{S}_{i,1}} = \log(0.0001)$，$(1 \le i \le I,\ 2 \le j \le J)$ とした．表現制御パラメータは $\xi = 3$ に固定した．パラメータの事前分布における固定パラメータは，$\mu_{\varphi} = 6$，$\sigma_{\varphi}^2 = 0.1$，$\mu_{\psi} = 0.6$，$\sigma_{\psi}^2 = 0.02$，$\sigma_{\mathrm{d}}^2 = 2500$，$\sigma_{B}^2 = 100$，$\mu_{B^{(i,1)}} = 0$，$\mu_{B^{(i,2)}} = 30$，$\mu_{B^{(i,3)}} = -30$，$\mu_{B^{(i,4)}} = 60$，$\mu_{B^{(i,5)}} = -60$，$(1 \le i \le I)$ とした．これらは文献 22) と予備実験に基づいて決定した結果である．

図 4 はテストデータに対するパラメータの推定結果例を示す．左側がテストデータであり，右側が推定結果である．テストデータに近い軌跡が推定されたことを定性的に確認できる．二乗平均平方根（Root Mean Square，RMS）を計算したところ，ノート成分とオーバーシュートを含む信号との RMS は 7.47cent，表現成分とビブラートおよびプレパレーションを含む信号との RMS は 9.51cent であった．対数周波数上の 100cent は半音に相当し，RMS がその 10% 以下に収まっていることから，人工的な F_0 軌跡を用いた場合，パラメータ推定アルゴリズムの局所解問題を抑制できていることを確認できた．

5.2 実測 F_0 軌跡を用いた評価

次に，実測の F_0 軌跡に対する推定結果を**図 5** に示す．実測 F_0 軌跡として，「RWC 研究用音楽データベース：ポピュラー音楽」(RWC-MDB-P-2001)[26] における，歌手名：緒方智美，曲番号：No.07，曲名：PROLOGUE のメロディの F_0 を手作業でラベル付した結果[27] を用いた．本来ならばこの楽曲の音響信号から F_0 を推定すべきであるが，今回は提案手法の性能の上限を調べるためにこのようなデータを利用した．F_0 は 10ms ごとにラベル付されているので，アップサンプリングによって，5ms ごとの実測 F_0 軌跡を得た．同楽曲の楽譜情報は「RWC 研究用音楽データベース：ポピュラー音楽」の MIDI データを利用した．図 5 は推定結果の一部であり，実測 F_0 軌跡からノート指令信号と表現指令信号が推定され，それらの制御機構のインパルス応答が畳み込まれて，ノート成分と表現成分が得られる．これらを足し合わせたものを合成 F_0 軌跡とし，定性的には実測 F_0 軌跡に近い軌跡が得られたことがわかる．また，微細変動成分パラメータである σ_{f}^2 の値は 187.7 であるため，振幅が 13cent 程度の微細な変動成分が推定されており，先行研究の知見[3] と整合する．特徴的

Vol.2012-MUS-94 No.11
Vol.2012-SLP-90 No.11
2012/2/3

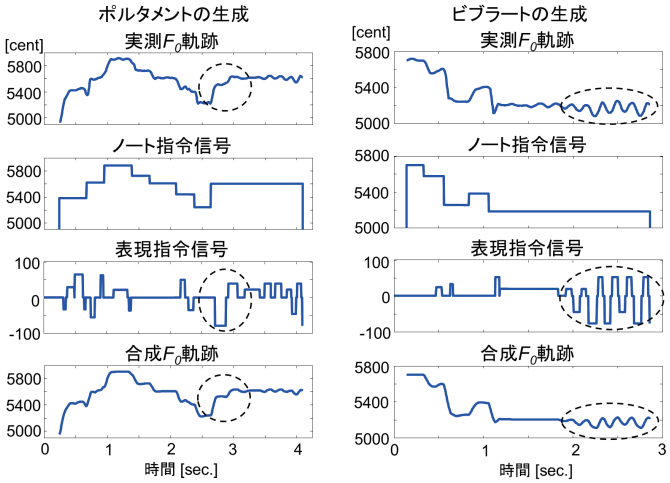

図 6 ポルタメントとビブラートが生成される様子

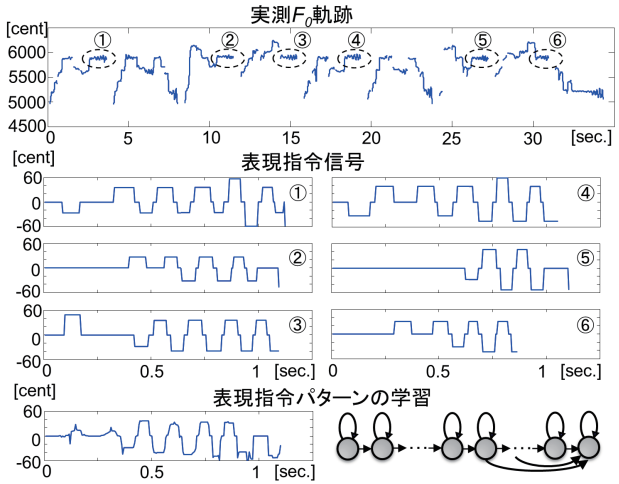

図 7 推定された表現指令信号を用いた表現指令パターンの学習

なことは，4 節で述べたように，無声音の区間を欠損データとして扱っているが，合成 F_0 軌跡ではその部分が周囲の F_0 から補間されていることである．

具体的に，推定されたノート指令信号と表現指令信号を見てみよう．**図 6** の左図はポルタメントが生成される様子を示す．楽譜（MIDI）では一つの音を伸ばし続けるよう記述されているものの，歌手の表現意図によって，半音だけ滑らかに変化させる様子が実測 F_0 軌跡に観測される．このとき提案法では，ノート指令信号は同一音符区間として推定され，表現指令信号は点線枠内に示されるような音高を変化させる指令が推定される．一方，図 6 の左図はビブラートが生成される様子を示す．実測 F_0 軌跡の最後の音符区間にビブラートが観測される．ビブラートが必ずしも正弦波に従う挙動ではないことがわかる．このとき，表現指令信号として，部分的に矩形変化する指令が推定された．ポルタメントとビブラートの生成のどちらにおいても合成 F_0 軌跡は実測 F_0 軌跡に近い軌跡が得られている．

5.3 表現指令パターンの学習

前節より，音符区間におけるポルタメントやビブラートのような表現意図を表現指令信号として推定できることがわかった．そうすると，この歌唱者の表現指令信号に頻出する標準パターンを学習したくなる．楽曲「PROLOGUE」には同一音高，同一音長の音符が複数存在するため，ある同一音符区間における表現指令入力信号を描画した結果を**図 7** に示す．ここでは 6 つの表現指令信号を示した（楽曲全体で 18 個ある）．この音符区間ではビブラートをかけるため，似たような表現指令が得られる．これらの表現指令信号の標準パターンを得るために，図 7 の下部に示す Left-to-right 型の HMM を利用して，指令信号の軌跡の学習を試みた．この音符区間の表現指令信号の平均長に基づいて，状態数は 220 とした．ただ，平均長よりも短い表現指令信号もあるため，最後の 50 状態は最終状態への遷移確率を持つように HMM のトポロジを構成した．出力確率分布は単一ガウス分布として HMM のパラメータを推定し，ガウス分布の平均値を描画した結果を図 7 の下部に示す．0.5 秒頃からビブラートの振動が開始し，その振幅が時間とともに右下がりに変動するパターンが学習された．今後の課題は，このように推定された表現指令信号から歌唱者の表現意図を音符区間ごとに大規模に学習することである．

5.4 聴取実験に基づく提案モデルの評価

最後に，提案モデルとそのパラメータ推定アルゴリズムによる歌唱の意図的表現の抽出について，主観的な聴取実験に基づいて評価する．5.2 節で示した実測 F_0 軌跡から推定される，4 種類の信号軌跡を準備する．図 5 と対応付けて見てほしい．

(1) ノート指令：$\boldsymbol{\mu}_n$
(2) ノート成分：$\boldsymbol{A}^{-1}\boldsymbol{\mu}_n$

Vol.2012-MUS-94 No.11
Vol.2012-SLP-90 No.11
2012/2/3

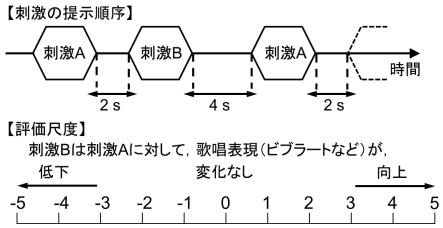

図 8 刺激の提示順序と評価尺度

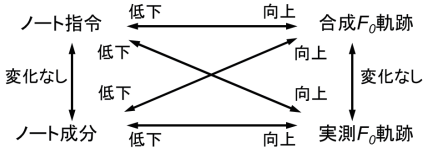

図 9 聴取実験に対する仮説

(3) 合成 F_0 軌跡：$\boldsymbol{A}^{-1}\boldsymbol{\mu}_n + \boldsymbol{B}^{-1}\boldsymbol{\mu}_e$

(4) 実測 F_0 軌跡：$\boldsymbol{y}$

これらの信号軌跡と YAMAHA の歌声合成ソフトウェア Vocaloid3[28]（歌手は Vocaloid2 用の VY1）⋆1を利用して，歌声を合成する．具体的には，推定されたノート指令信号に基づいて，発音時刻ごとにノートと歌詞を入力する（今回はすべての音符を「な，/na/」で発音する）．次に，ピッチベンドセンシティビティを 12（1 オクターブ）に設定して，(1)〜(4) の信号軌跡それぞれに対してピッチベンドの値を計算し，入力する．音量はすべて一定とする．以上の手順によって，4 種類の歌声が合成される．最後にこれらの音響信号を 4 小節ごとに切り出した全 96 サンプル（歌声ごとに 24 サンプルが切り出される）を刺激として用いる．**図 8** に示すように，ランダムに刺激を選び出して，刺激 A，刺激 B と提示し，これを 50 セット準備する．被験者数は成人男性 5 名とし，各被験者は，刺激 B は刺激 A に対して歌唱表現が向上したか否か（ビブラートやポルタメントを付加し，滑らかに歌っているか否か）を 11 段階で評価する．この聴取実験の結果に対する仮説を**図 9** に示す．着目すべきは，

- 合成 F_0 軌跡による歌声が，実測 F_0 軌跡による歌声と比べて，歌唱表現に変化がなく評価されること
- 合成 F_0 軌跡による歌声が，ノート指令による歌声およびノート成分による歌声と比べて，歌唱表現が向上したと評価されること

と仮説を立てている点である．

評価結果を**表 1** に示す．被験者ごとに評価値の分散が異なるので，各被験者で値を正規化し，それを 5 人の被験者で平均した結果が表 1 である．仮説のとおり，刺激 A{ ノート指令，ノート成分 } → 刺激 B{ 合成 F_0 軌跡，実測 F_0 軌跡 } のとき評価値は正となり，刺激 A{ 合成 F_0 軌跡，実測 F_0 軌跡 } → 刺激 B{ ノート指令，ノート成分 } のとき評価値は負となり，合成音における歌唱表現の有無を評価できていることがわかる．なおかつ，刺激 A{ 合成 F_0 軌跡 } → 刺激 B{ 実測 F_0 軌跡 } および刺激 A{ 実測 F_0 軌跡 } → 刺激 B{ 合成 F_0 軌跡 } のとき，評価値が他に比べて 0 に近い値となった．このことから，実測 F_0 軌跡に含まれる歌唱の意図的表現が，提案法の表現成分として推定されることを主観的な観点から確認できた．ただし，合成 F_0 軌跡による歌声と実測 F_0 軌跡による歌声を比較したときの評価値が 0 でないことから，残された微細変動成分について今後詳細に分析する必要がある．また本実験では対象曲が 1 曲なので，今後は様々な歌唱者が多様な楽曲を歌った歌声を対象として，提案モデルの有効性を大規模に評価する必要がある．

表 1 聴取実験結果：評価値を被験者ごとに正規化した後の平均値

刺激 A ＼ 刺激 B	ノート指令	ノート成分	合成 F_0 軌跡	実測 F_0 軌跡
ノート指令		0.151	0.479	0.434
ノート成分	-0.106		0.560	0.632
合成 F_0 軌跡	-0.621	-0.524		0.123
実測 F_0 軌跡	-0.679	-0.540	0.091	

6. おわりに

本稿では，歌声 F_0 軌跡に含まれる物理的制約もしくは意図的表現による楽譜逸脱成分（F_0 動的変動成分）を楽譜情報と分離して抽出することを目的として，ノート指令信号と表現指令信号によって駆動される F_0 軌跡の生成過程を記述し，そのモデルパラメータを推定する逆問題の解法を提案した．提案する F_0 生成過程は楽譜に記載される音符の並びを表現するノート指令信号と歌唱者の音楽的な表現意図を表す表現指令信号が 2 次系によって制御され，それぞれノート成分と表現成分を出力する．これらの成分とガウス性白色雑音に従う微細変動成分との和によって F_0 軌跡が記述される．評価実験では，導出したパラメータ推定アルゴリズムによって推定される表現成分に，歌唱者の意図的表現が含まれることを客観的かつ主観的に確認し，提案モデルの有効性を示した．今後の課題は，抽出された楽譜

⋆1 http://www.vocaloid.com/lineup/

逸脱成分が歌唱者ごとにどのようなパターンをもちうるのか，各パターンが文脈（楽譜の音符列）にどう依存するかを学習することである．また，大規模な聴取実験を行うことで，常に人間の歌声知覚も考慮しながら研究を進める必要がある．さらには，F_0（音高）だけでなく，音韻や音量における楽譜逸脱成分についても検討を進めたい．

謝辞 本研究に対し，有益なご助言を頂いた中野允裕氏 (NTT CS 研) に感謝致します．

参 考 文 献

1) Sundberg, J.: *The Science of the Singing*, Northern Illinois University Press (1987).
2) de Krom, G. and Bloothooft, G.: Timing and Accuracy of Fundamental Frequency Changes in Singing, *In Proc. ICPhS95*, pp.206–209 (1995).
3) Akagi, M. and Kitakaze, H.: Perception of Synthesized Singing Voices with Fine Fluctuations in Their Fundamental Frequency Contours, *in Proc. ICSLP 2000*, pp. 458–461 (2000).
4) Seashore, C. E.: A Musical Ornament, the Vibrato, *In Psychology of Music*, McGraw-Hill Book Company, pp.33–52 (1938).
5) Kenji Kojima, M. Y. and Nakayama, I.: Variability of Vibrato -A Comparative Study between Japanese Traditional Singing and Bel Canto-, *In Proc. Speech Prosody 2004*, pp.151–154 (2004).
6) Nakayama, I.: Comparative Studies on Vocal Expressions in Japanese Traditional and Western Classical- Style Singing, Using a Common Verse, *In Proc. ICA 2004*, pp.1295–1296 (2004).
7) Saitou, T. et al.: Speech-To-Singing Synthesis: Converting Speaking Voices to Singing Voices by Controlling Acoustic Features Unique to Singing Voices, *in Proc. WASSPA 2007*, pp.215–218 (2007).
8) Saitou, T. et al.: Acoustic and Perceptual Effects of Vocal training in Amateur Male Singing, *in Proc. EUROSPEECH 2009*, pp.832–835 (2009).
9) 右田尚人ほか：ヴィブラート歌唱における基本周波数制御に有効な特徴量の検討，音講論集, 3-P-28，pp.395–396 (2010).
10) 河原英紀，片寄晴弘：高品質音声分析変換合成システム STRAIGHT を用いたスキャット生成研究の提案，情報処理学会論文誌， Vol.43, No.2, pp.208–218 (2002).
11) Nakano, T. et al.: An Automatic Singing Skill Evaluation Method for Unknown Melodies Using Pitch Interval Accuracy and Vibrato Features, *Proc. ICSLP 2006*, pp.1706–1709 (2006).
12) Mayor, O. et al.: The Singing Tutor: Expression Categorization and Segmentation of the Singing Voice, *in Proc. AES 121st Convention* (2006).
13) 中野倫靖，後藤真孝：VocaListener：ユーザ歌唱の音高および音量を真似る歌声合成システム，情報処理学会論文誌， Vol.52, No.12, pp.3853–3867 (2011).
14) Bonada, J. et al.: Synthesis of the Singing Voice by Performance Sampling and Spectral Models, *IEEE Signal Processing Magazine*, Vol.24, pp.67–79 (2007).
15) Kako, T. et al.: Automatic Identification for Singing Style Based on Sung Melodic Contour Characterized in Phase Plane, *in Proc. ISMIR 2009*, pp.393–397 (2009).
16) Fukayama, S. et al.: Orpheus: Automatic Composition System Considering Prosody of Japanese Lyrics, *in Proc. ICEC 2009*, pp.309–310 (2009).
17) Nakano, T. et al.: VocaListener2: A Singing Synthesis System Able to Mimic a User's Singing in terms of Voice Timbre Chages as Well as Pitch and Dynamics, *in Proc. ICASSP 2011*, pp.453–456.
18) Mase, A. et al.: HMM-based singing voice synthesis system using pitch-shifted pseudo training data, *in Proc. INTERSPEECH 2010*, pp.845–848.
19) 柏野邦夫ほか：パート譜を用いたボーカル音分離システム，音講論集, 2-9-1，pp.625–626 (1998).
20) 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).
21) Minematsu, N. et al.: Prosodic Modeling of Nagauta Singing and Its Evaluation, *Proc. SpeechProsody 2004*, pp.487–490 (2004).
22) Saitou, T., Unoki, M. and Akagi, M.: Development of an F0 control model based on F0 dynamic characteristics for singing-voice synthesis, *Speech Communication*, Vol.46, pp.405–417 (2005).
23) Fujisaki, H.: A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour, *Vocal Physiology: Voice Production, Mechanisms and Functions, (O. Fujimura, ed.)*, Raven Press, pp. 347–355 (1988).
24) 大石康智ほか：歌声 F0 系列からの楽譜逸脱成分の抽出 －動特性モデルに基づく楽譜との時間的対応付け－，音講論集, 1-8-19，pp.279–282 (2011).
25) Kameoka, H. et al.: A statistical model of speech F0 contours, *in Proc. SAPA 2010*, pp.43–48.
26) 後藤真孝，橋口博樹，西村拓一，岡 隆一：RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース，情報処理学会論文誌， Vol.45, No.3, pp.728–738 (2004).
27) Goto, M.: AIST Annotation for the RWC Music Database, *inn Proc. ISMIR 2006* (2006).
28) Kenmochi, H. and Ohshita, H.: VOCALOID - Commercial Singing Synthesizer based on Sample Concatenation, *in Proc. INTERSPEECH 2007*, pp. 4010–4011 (2007).