音声 F_0 パターン生成過程の確率モデル*

○亀岡弘和, ルルージョナトン, 大石康智
日本電信電話 (株) NTT コミュニケーション科学基礎研究所

1 序 論

音声の韻律的特徴を表現する物理量の一つである基本周波数 (F_0) の時間変化パターン (以後，F_0 パターン) には，構文や意図の伝達に関連する情報が豊富に含まれており，言語情報に関連する量であるフォルマントの時間変化パターンと並んで音声の情報伝達における重要な役割を担っている。フォルマントは声道の形状の変化に付随して変化する声道共振特性を特徴づける量であるのに対し，F_0 パターンは肺からの空気供給量と甲状軟骨の運動に付随して変化する声帯の伸びに応じて決まる量である。それゆえ，音声からは音声生成メカニズムの物理的制約の範囲内の F_0 パターンしか生成され得ない。このように考えると，ある F_0 パターンがどれだけ音声のものであるらしいかを客観的に評価するためには，それが音声生成メカニズムの物理的制約をどれだけ満たしているかを規準に評価すれば良い，ということになる。従って，F_0 パターンの生成過程のモデル化は，音声の自然性を考慮に入れることで性能向上が望めるいかなる音声アプリケーションにおいて大変有益となる可能性がある。

F_0 パターン生成過程のモデルの代表格といえば藤崎のモデル [1] (以後，藤崎モデル) であろう。藤崎モデルは，発話の言語学的情報と密接な関係にあるパラメータを与えることによって，実測の音声 F_0 パターンを極めて良く近似するパターンを生成することが可能であり，その有効性は音声合成に広く利用されていることからも実証されている。また，実測の F_0 パターンから藤崎モデルのパラメータを推定する逆問題を扱った研究例も多く [2]，藤崎モデルのパラメータを音声認識や感情認識などの精度向上に役立てる試みも取り組まれている。しかし，逆問題の解析的な複雑さゆえにいまだ有効なパラメータ推定法は確立されていないのが現状である。

F_0 パターンのモデル化の重要性と藤崎モデルの (潜在的な) 有効性を踏まえ，本稿では，藤崎モデルを確率過程に基づいて統計モデル化する。その目的は，(1) 統計的手法を駆使した強力なパラメータ推定法の枠組を確立すること，(2) 統計モデルに基づく多くの音声処理問題 (分析，合成，分離，強調) に，F_0 パターンの統計モデルを新たな音声らしさの規準として組み込めるような下地を作ること，にある。

2 藤崎の F_0 パターン生成過程モデル [1]

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，声帯の伸びと対数 F_0 の変化が比例関係にあるという仮定に基づき F_0 パターンがモデル化される。甲状軟骨の平行移動運動に関係する F_0 パターンをフレーズ成分，回転運動に関係する F_0 パターンをアクセント成分と呼び，それぞれ $y_{\rm p}(t), y_{\rm a}(t)$ とする。ただし，t は時刻である。$y_{\rm p}(t)$ の生成過程 (フレーズ制御機構) はフレーズ指令と呼ぶパルス波を入力とした臨界制動の二次線形系，$y_{\rm a}(t)$ の生成過程 (アクセント制御機構) はアクセント指令と呼ぶ矩形波を入力とした臨界制動の二次線形系により表現される。以上の二つの成分と，声帯の物理的性質によって決まるベースライン成分 y_b の和 $y_{\rm p}(t) + y_{\rm a}(t) + y_b$ で F_0 パターン $y(t)$ を表したものが藤崎モデルである。フレーズ成分は短区間の上昇のあとに緩やかな下降をなす成分で，アクセント成分は急激な上昇下降をなす成分であるため，多くの言語に共通して前者が句単位の比較的緩やかな音調を，後者が語または音節単位の比較的急激で局所的な音調を表現する役割を担っていると考えられている。フレーズ制御機構およびアクセント制御機構は

$$G_{\rm p}(t) = \begin{cases} \alpha^2 t e^{-\alpha t} & (t \geq 0) \\ 0 & (t < 0) \end{cases} \tag{1}$$

$$S_{\rm a}(t) = \begin{cases} 1-(1+\beta t)e^{-\beta t} & (t \geq 0) \\ 0 & (t < 0) \end{cases} \tag{2}$$

で与えられるインパルス応答 $G_{\rm p}(t)$ とステップ応答 $S_{\rm a}(t)$ によって特徴づけられる。α と β はそれぞれの制御機構の固有角周波数であるが，これらは話者ごとにほぼ一定値をとること，話者の個人差も言語による差も比較的小さいことが確かめられており，おおよそ $\alpha = 3$rad/s, $\beta = 20$rad/s 程度であることが経験的に知られている。これらを用いて F_0 パターン $y(t)$ は具体的に

$$\begin{aligned} y(t) = y_{\rm b} &+ \sum_i A_{{\rm p},i} G_{\rm p}(t - T_{0,i}) \\ &+ \sum_j A_{{\rm a},j} \big\{ S_{\rm a}(t - T_{1,j}) - S_{\rm a}(t - T_{2,j}) \big\} \end{aligned} \tag{3}$$

と表される。ただし，$A_{{\rm p},i}$ と $T_{0,i}$ はそれぞれ i 番目のフレーズ指令の大きさと生起時刻であり，$A_{{\rm a},j}$ と $T_{1,j}$，と $T_{2,j}$ はそれぞれ j 番目のアクセント指令の振幅と始端時刻と終端時刻を表す。ところで，アクセント制御機構のインパルス応答 $G_{\rm a}(t)$ は $S_{\rm a}(t)$ の時間微分ゆえ

$$G_{\rm a}(t) = \begin{cases} \beta^2 t e^{-\beta t} & (t \geq 0) \\ 0 & (t < 0) \end{cases} \tag{4}$$

となり，$G_{\rm p}(t)$ と同形であることが分かる。従って，アクセント成分 $y_{\rm a}(t)$ は複数の矩形波からなるアクセント指令関数 $u_{\rm a}(t)$ と $G_{\rm a}(t)$ の畳み込みによって表される。

---

*A probabilistic model of speech F_0 contours. by KAMEOKA Hirokazu, LE ROUX Jonathan and OHISHI Yasunori (NTT Communication Science Laboratories)

3　藤崎モデルの離散時間表現

本章では，藤崎モデルの離散時間表現を得るため，連続時間システムのフレーズ制御機構およびアクセント制御機構に対して後退差分変換により離散化を行う。まず，フレーズ制御機構の伝達関数 (Laplace 変換) は $\mathcal{G}_{\mathrm{p}}(s) = \mathcal{L}\big[G_{\mathrm{p}}(t)\big] = \alpha^2/(s+\alpha)^2$ で与えられる。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 は変換先の離散時間表現におけるサンプリング周期)，この変換により，$\mathcal{G}_{\mathrm{p}}^{-1}(s)$ の逆システムの伝達関数を z 領域で

$$\mathcal{H}_{\mathrm{p}}^{-1}(z) = a_2 z^{-2} + a_1 z^{-1} + a_0 \tag{5}$$

と書くことができる。ただし，

$$a_2 = (\psi-1)^2,\quad a_1 = -2\psi(\psi-1),\quad a_0 = \psi^2 \tag{6}$$

および，$\psi = 1 + 1/(\alpha t_0)$ である。ここで，k を離散時刻インデックスとし，フレーズ指令関数およびフレーズ成分の離散時間表現をそれぞれ $u_{\mathrm{p}}[k]$ および $y_{\mathrm{p}}[k]$ とすると，$y_{\mathrm{p}}[k]$ は，単一のパラメータ ψ(あるいは α) によって特性が決まる拘束つき全極モデルからの出力

$$u_{\mathrm{p}}[k] = a_0 y_{\mathrm{p}}[k] + a_1 y_{\mathrm{p}}[k-1] + a_2 y_{\mathrm{p}}[k-2] \tag{7}$$

と見なすことができる。同様に，アクセント指令関数 $u_{\mathrm{a}}[k]$ とアクセント成分 $y_{\mathrm{a}}[k]$ の関係も

$$u_{\mathrm{a}}[k] = b_0 y_{\mathrm{a}}[k] + b_1 y_{\mathrm{a}}[k-1] + b_2 y_{\mathrm{a}}[k-2] \tag{8}$$

と書くことができる。ただし $b_2 = (\varphi-1)^2$，$b_1 = -2\varphi(\varphi-1)$，$b_0 = \varphi^2$，$\varphi = 1 + 1/(\beta t_0)$ である。ベースライン成分 $y_{\mathrm{b}}(t)$ の離散時間表現を $y_{\mathrm{b}}[k]$ とすると，藤崎モデルの離散時間表現はこれらの三成分の和 $y[k] = y_{\mathrm{p}}[k] + y_{\mathrm{a}}[k] + y_{\mathrm{b}}[k]$ で与えられる。

4　藤崎モデルの統計モデル化

まず，$u_{\mathrm{p}}[k]$ と $u_{\mathrm{a}}[k]$ のモデル化について述べる。藤崎モデルにおいて，フレーズ指令とアクセント指令に関して以下のような規則が定められている。

(A1) フレーズ指令はパルス波，アクセント指令は矩形波である。(A2) 発話の開始または先行フレーズ内のアクセント指令終了後にフレーズ指令が発生する。また，フレーズ指令の後にアクセント開始指令が発生する。つまり，アクセント指令発生中はフレーズ指令は発生しない。(A3) アクセント指令の開始した後には必ずアクセント終了指令が発生する。つまり，隣接するアクセント指令同士は重なり合うことはない。

藤崎モデルのパラメータ推定における難しさの一つは，これらの制約の下で最適推定をいかにして行えるかという点にあろう。特に，上記の (A2), (A3) より，$u_{\mathrm{p}}[k]$ と $u_{\mathrm{a}}[k]$ は相互に依存し合う関係がある点に注意が必要である。提案法では，これを解決するため隠れマルコフモデル (HMM) を用いて指令入力信号を確率モデル化する。

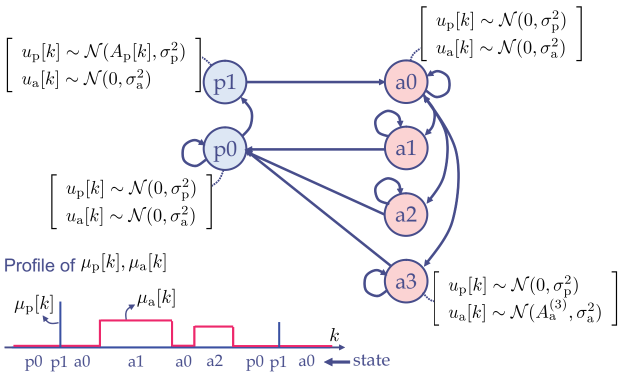

図 1　Command function modeling with HMM.

まず，$\boldsymbol{o}[k] := (u_{\mathrm{p}}[k], u_{\mathrm{a}}[k])^{\mathrm{T}}$ を，

$$\boldsymbol{o}[k] \sim \mathcal{N}\left(\boldsymbol{\nu}[k], \boldsymbol{\Upsilon}\right) \tag{9}$$

$$\boldsymbol{\nu}[k] := \begin{bmatrix} \mu_{\mathrm{p}}[k] \\ \mu_{\mathrm{a}}[k] \end{bmatrix},\quad \boldsymbol{\Upsilon} := \begin{bmatrix} \sigma_{\mathrm{p}}^2 & 0 \\ 0 & \sigma_{\mathrm{a}}^2 \end{bmatrix} \tag{10}$$

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 1 のような状態遷移に伴って変化するモデルを考える。これは HMM に他ならず，このように $\boldsymbol{o}[k]$ を HMM でモデル化したことにより，状態遷移の経路制限 (状態遷移確率の設定) を通して $\boldsymbol{\nu}[k]$ に対して上記の (A1)〜(A3) を満たすような制約を与えることが可能となる。提案する HMM の構成は以下のとおりである。

$$\begin{array}{rl}
\text{出力値系列:} & \{\boldsymbol{o}[k]\}_{k=1}^{K} \\
\text{状態集合:} & \mathcal{S} := \{\mathrm{p}_0, \mathrm{p}_1, \mathrm{a}_0, \cdots, \mathrm{a}_N\} \\
\text{状態系列:} & \{s_k\}_{k=1}^{K} \\
\text{状態出力分布:} & P(\boldsymbol{o}[k] | s_k = i) = \mathcal{N}(\boldsymbol{c}_i[k], \boldsymbol{\Upsilon}) \\
 & \boldsymbol{c}_i[k] = \begin{cases} (0, 0)^{\mathrm{T}} & (i = \mathrm{p}_0, \mathrm{a}_0) \\ \left(A_{\mathrm{p}}[k], 0\right)^{\mathrm{T}} & (i = \mathrm{p}_1) \\ \left(0, A_{\mathrm{a}}^{(n)}\right)^{\mathrm{T}} & (i = \mathrm{a}_n) \end{cases} \\
\text{状態遷移確率:} & \phi_{i',i} := \log P(s_k = i | s_{k-1} = i')
\end{array}$$

簡単のため状態遷移確率を定数とすると，以上より指令入力モデルにおいて推定すべきパラメータは，フレーズ指令の大きさ $A_{\mathrm{p}}[k]$，状態遷移系列 s_k，アクセント指令の大きさ $\{A_{\mathrm{a}}^{(n)}\}_{n=1}^{N}$，指令入力信号の分散 $\sigma_{\mathrm{p}}^2, \sigma_{\mathrm{a}}^2$ であり，これらをまとめて $\boldsymbol{\theta}_u$ と記す。また，平均値系列 $\{\mu_{\mathrm{p}}[k]\}_{k=1}^{K}$ および $\{\mu_{\mathrm{a}}[k]\}_{k=1}^{K}$ は，状態遷移系列 $\{s_k\}_{k=1}^{K}$ が与えられたもとで $(\mu_{\mathrm{p}}[k], \mu_{\mathrm{a}}[k])^{\mathrm{T}} \leftarrow \boldsymbol{c}_{s_k}[k]$ で与えられる。

前述の指令入力モデルに基づき $\boldsymbol{y} = (y[1], \cdots, y[K])^{\mathrm{T}}$ の確率密度関数を導く。式 (9), (10) より，$\boldsymbol{u}_{\mathrm{p}} := (u_{\mathrm{p}}[1], \cdots, u_{\mathrm{p}}[K])^{\mathrm{T}}$, $\boldsymbol{u}_{\mathrm{a}} := (u_{\mathrm{a}}[1], \cdots, u_{\mathrm{a}}[K])^{\mathrm{T}}$, $\boldsymbol{\mu}_{\mathrm{p}} := (\mu_{\mathrm{p}}[1], \cdots, \mu_{\mathrm{p}}[K])^{\mathrm{T}}$, $\boldsymbol{\mu}_{\mathrm{a}} := (\mu_{\mathrm{a}}[1], \cdots, \mu_{\mathrm{a}}[K])^{\mathrm{T}}$ とすると，

$$\boldsymbol{u}_{\mathrm{p}} | \theta_u \sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{p}}, \boldsymbol{\Sigma}_{\mathrm{p}}),\quad \boldsymbol{\Sigma}_{\mathrm{p}} = \sigma_{\mathrm{p}}^2 \boldsymbol{I} \tag{11}$$

$$\boldsymbol{u}_{\mathrm{a}} | \theta_u \sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{a}}, \boldsymbol{\Sigma}_{\mathrm{a}}),\quad \boldsymbol{\Sigma}_{\mathrm{a}} = \sigma_{\mathrm{a}}^2 \boldsymbol{I} \tag{12}$$

が言える。3 章で得た関係式より，フレーズ成分 $\boldsymbol{y}_{\mathrm{p}} := (y_{\mathrm{p}}[1], \cdots, y_{\mathrm{p}}[K])^{\mathrm{T}}$ と $\boldsymbol{u}_{\mathrm{p}}$ の関係，および，アクセント

成分 $\boldsymbol{y}_{\mathrm{a}} := (y_{\mathrm{a}}[1], \cdots, y_{\mathrm{a}}[K])^{\mathrm{T}}$ と $\boldsymbol{u}_{\mathrm{a}}$ の関係は,

$$\boldsymbol{A} := \begin{bmatrix} a_0 & & & O \\ a_1 & a_0 & & \\ a_2 & a_1 & a_0 & \\ & \ddots & \ddots & \ddots \\ O & & a_2 \; a_1 & a_0 \end{bmatrix}, \quad \boldsymbol{B} := \begin{bmatrix} b_0 & & & O \\ b_1 & b_0 & & \\ b_2 & b_1 & b_0 & \\ & \ddots & \ddots & \ddots \\ O & & b_2 \; b_1 & b_0 \end{bmatrix} \tag{13}$$

と置くと，それぞれ $\boldsymbol{u}_{\mathrm{p}} = \boldsymbol{A}\boldsymbol{y}_{\mathrm{p}}$, $\boldsymbol{u}_{\mathrm{a}} = \boldsymbol{B}\boldsymbol{y}_{\mathrm{a}}$ のように表現できることから，式 (11), (12) より

$$\boldsymbol{y}_{\mathrm{p}}|\theta_u, \alpha \sim \mathcal{N}\big(\boldsymbol{A}^{-1}\boldsymbol{\mu}_{\mathrm{p}}, \boldsymbol{A}^{-1}\boldsymbol{\Sigma}_{\mathrm{p}}(\boldsymbol{A}^{-1})^{\mathrm{T}}\big) \tag{14}$$

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

が導かれる。ベース成分 $y_{\mathrm{b}}[k]$ についても，同様に白色 Gauss 性雑音 $\epsilon_{\mathrm{b}}[k]$ に起因する確率変数 $y_{\mathrm{b}}[k] = \mu_{\mathrm{b}} + \epsilon_{\mathrm{b}}[k]$ と仮定し，$\epsilon_{\mathrm{b}}[k] \sim \mathcal{N}(0, \sigma_{\mathrm{b}}^2)$ とし，同様に，$\epsilon_\xi[j]$ と $\epsilon_{\xi'}[j']$ は $(\xi, j) \neq (\xi', j')$ のとき独立とすると,

$$\boldsymbol{y}_{\mathrm{b}}|\mu_{\mathrm{b}} \sim \mathcal{N}(\mu_{\mathrm{b}}\boldsymbol{1}, \boldsymbol{\Sigma}_{\mathrm{b}}) \tag{16}$$

が言える。ただし，$\boldsymbol{\Sigma}_{\mathrm{b}} = \sigma_{\mathrm{b}}^2\boldsymbol{I}$ であり，$\theta_{\mathrm{b}} := \{\mu_{\mathrm{b}}, \sigma_{\mathrm{b}}^2\}$ と置く。仮定より，$\boldsymbol{y}_{\mathrm{p}}, \boldsymbol{y}_{\mathrm{a}}, \boldsymbol{y}_{\mathrm{b}}$ は独立なので，$\Theta := \{\theta_u, \alpha, \beta, \theta_{\mathrm{b}}\}$ が与えられたもとでの F_0 パターン $\boldsymbol{y} = \boldsymbol{y}_{\mathrm{p}} + \boldsymbol{y}_{\mathrm{a}} + \boldsymbol{y}_{\mathrm{b}}$ の確率密度関数は，式 (14), (15) と式 (16) より,

$$\begin{aligned} \boldsymbol{y}|\Theta \sim \mathcal{N}\big(&\boldsymbol{A}^{-1}\boldsymbol{\mu}_{\mathrm{p}} + \boldsymbol{B}^{-1}\boldsymbol{\mu}_{\mathrm{a}} + \mu_{\mathrm{b}}\boldsymbol{1}, \\ &\boldsymbol{A}^{-1}\boldsymbol{\Sigma}_{\mathrm{p}}(\boldsymbol{A}^{-1})^{\mathrm{T}} + \boldsymbol{B}^{-1}\boldsymbol{\Sigma}_{\mathrm{a}}(\boldsymbol{B}^{-1})^{\mathrm{T}} + \boldsymbol{\Sigma}_{\mathrm{b}}\big) \end{aligned} \tag{17}$$

で与えられる。以上より,

$$\begin{aligned} P(\boldsymbol{y}|\Theta) &= \frac{|\boldsymbol{\Sigma}^{-1}|^{1/2}}{(2\pi)^{T/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right\} \\ \boldsymbol{\mu} &= \boldsymbol{A}^{-1}\boldsymbol{\mu}_{\mathrm{p}} + \boldsymbol{B}^{-1}\boldsymbol{\mu}_{\mathrm{a}} + \mu_{\mathrm{b}}\boldsymbol{1} \qquad (18) \\ \boldsymbol{\Sigma} &= \boldsymbol{A}^{-1}\boldsymbol{\Sigma}_{\mathrm{p}}\big(\boldsymbol{A}^{\mathrm{T}}\big)^{-1} + \boldsymbol{B}^{-1}\boldsymbol{\Sigma}_{\mathrm{a}}\big(\boldsymbol{B}^{\mathrm{T}}\big)^{-1} + \boldsymbol{\Sigma}_{\mathrm{b}} \end{aligned}$$

が，F_0 パターン $\boldsymbol{y}$ が与えられたときの藤崎モデルパラメータ Θ の尤度関数である。

Θ の事前確率については，各要素は独立で，状態遷移系列 $\{s[k]\}_{k=1}^K$ と制御パラメータの ψ と φ 以外のパラメータは一様に分布すると仮定し，$P(\Theta) \propto P(\psi)P(\varphi)P(s_1)\prod_{k=2}^K P(s_k|s_{k-1})$ とする。

5 提案モデルの応用場面

以上の F_0 パターンの統計モデルの具体的な応用例について述べる。まず，提案モデルを用いて実測 F_0 パターン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 データ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 データは通常観測できない。従って，実測 F_0 パターン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。統計的枠組では，この類の問題は不完全データ問題に他ならず，Expectation-Maximization(EM) アルゴリズムにより扱うことができる。$\boldsymbol{y} \in \mathbb{R}^K$ を観測 F_0 データと欠損 F_0 データからなる完全データとし，実際に観測された F_0 データを並べたベクトルを $\boldsymbol{y}_{\mathrm{obs}} \in \mathbb{R}^{K'} (K' \leq K)$ としよう。$\boldsymbol{y}$ 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。紙面の都合上導出は省略せざるを得なかったが，E ステップでは $\boldsymbol{y}$ を

$$\boldsymbol{y} \leftarrow \boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{M}^{\mathrm{T}}(\boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}^{\mathrm{T}})^{-1}(\boldsymbol{y}_{\mathrm{obs}} - \boldsymbol{M}\boldsymbol{\mu}) \tag{19}$$

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

$$\Theta \leftarrow \underset{\Theta}{\operatorname{argmax}} \log P(\boldsymbol{y}|\Theta)P(\Theta) \tag{20}$$

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 $\boldsymbol{M} = \boldsymbol{I}$ においては式 (19) 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たしかになっていることが分かる。

第二の応用方法としては，F_0 をパラメータにもつ何らかの音声モデルの事前分布として利用するやり方である。具体的な応用方法については [3] で詳しく論じているのでそちらを参照されたい。

6 パラメータ推定アルゴリズム

$P(\Theta|\boldsymbol{y})$ を最大化する問題 (式 (20) に相当) は解析的に解くことはできないが，$\boldsymbol{x} := (\boldsymbol{y}_{\mathrm{p}}^{\mathrm{T}}, \boldsymbol{y}_{\mathrm{a}}^{\mathrm{T}}, \boldsymbol{y}_{\mathrm{b}}^{\mathrm{T}})^{\mathrm{T}}$ 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。この場合，完全データの対数尤度は，先に見たとおり,

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

$$\boldsymbol{x} := \begin{bmatrix} \boldsymbol{y}_{\mathrm{p}} \\ \boldsymbol{y}_{\mathrm{a}} \\ \boldsymbol{y}_{\mathrm{b}} \end{bmatrix}, \ \boldsymbol{m} := \begin{bmatrix} \boldsymbol{A}^{-1}\boldsymbol{\mu}_{\mathrm{p}} \\ \boldsymbol{B}^{-1}\boldsymbol{\mu}_{\mathrm{a}} \\ \mu_{\mathrm{b}}\boldsymbol{1} \end{bmatrix} \tag{21}$$

$$\boldsymbol{\Lambda}^{-1} := \begin{bmatrix} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{p}}^{-1}\boldsymbol{A} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{a}}^{-1}\boldsymbol{B} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{\Sigma}_{\mathrm{b}}^{-1} \end{bmatrix} \tag{22}$$

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

$$\begin{aligned} Q(\Theta, \Theta') \stackrel{c}{=} \frac{1}{2}\Big[&\log|\boldsymbol{\Lambda}^{-1}| - \mathrm{tr}(\boldsymbol{\Lambda}^{-1}\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{y};\Theta']) \\ &+ 2\boldsymbol{m}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta'] - \boldsymbol{m}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\boldsymbol{m}\Big] + \log P(\Theta) \end{aligned} \tag{23}$$

となる。ここで，$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}$(ただし，$\boldsymbol{H} = [\boldsymbol{I} \ \boldsymbol{I} \ \boldsymbol{I}]$) であるから，$\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta]$ と $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{y};\Theta]$ は，具体的に

$$\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta] = \boldsymbol{m} + \boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}})^{-1}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{m}) \tag{24}$$

$$\begin{aligned} \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{y};\Theta] = \boldsymbol{\Lambda} - &\boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}})^{-1}\boldsymbol{H}\boldsymbol{\Lambda} \\ &+ \mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta]\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta]^{\mathrm{T}} \end{aligned} \tag{25}$$

と書ける。E ステップでは，直前のステップで更新されたモデルパラメータを 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。$\boldsymbol{y}_{\mathrm{p}}, \boldsymbol{y}_{\mathrm{a}}, \boldsymbol{y}_{\mathrm{b}}$ 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

$$\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta'] = \begin{bmatrix} \bar{\boldsymbol{x}}_{\mathrm{p}} \\ \bar{\boldsymbol{x}}_{\mathrm{a}} \\ \bar{\boldsymbol{x}}_{\mathrm{b}} \end{bmatrix}, \ \ \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{y};\Theta'] = \begin{bmatrix} \boldsymbol{R}_{\mathrm{p}} & * & * \\ * & \boldsymbol{R}_{\mathrm{a}} & * \\ * & * & \boldsymbol{R}_{\mathrm{b}} \end{bmatrix} \tag{26}$$

のように区分表現すると，Q 関数は

$$
\begin{aligned}
Q(\Theta,\Theta') \stackrel{c}{=} \frac{1}{2}\Big[& \log|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{p}}^{-1}\boldsymbol{A}| + \log|\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{a}}^{-1}\boldsymbol{B}| + \log|\boldsymbol{\Sigma}_{\mathrm{b}}^{-1}| \\
& - \mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{p}}^{-1}\boldsymbol{A}\boldsymbol{R}_{\mathrm{p}}) + 2\boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{p}}^{-1}\boldsymbol{A}\bar{\boldsymbol{x}}_{\mathrm{p}} - \boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{p}}^{-1}\boldsymbol{\mu}_{\mathrm{p}} \\
& - \mathrm{tr}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{a}}^{-1}\boldsymbol{B}\boldsymbol{R}_{\mathrm{a}}) + 2\boldsymbol{\mu}_{\mathrm{a}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{a}}^{-1}\boldsymbol{B}\bar{\boldsymbol{x}}_{\mathrm{a}} - \boldsymbol{\mu}_{\mathrm{a}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{a}}^{-1}\boldsymbol{\mu}_{\mathrm{a}} \\
& - \mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{b}}^{-1}\boldsymbol{R}_{\mathrm{b}}) + 2\boldsymbol{\mu}_{\mathrm{b}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{b}}^{-1}\bar{\boldsymbol{x}}_{\mathrm{b}} - \boldsymbol{\mu}_{\mathrm{b}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{b}}^{-1}\boldsymbol{\mu}_{\mathrm{b}}\Big] \\
& + \log P(\Theta) \qquad (27)
\end{aligned}
$$

と書き直せて，これを用い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

$$
\begin{aligned}
\mathcal{I}_1(s) := & -\frac{1}{2}\sum_{k=1}^{K}(\boldsymbol{o}[k]-\boldsymbol{c}_{s_k}[k])^{\mathrm{T}}\boldsymbol{\Upsilon}^{-1}(\boldsymbol{o}[k]-\boldsymbol{c}_{s_k}[k]) \\
& + \log P(s_1) + \sum_{k=2}^{K}\log P(s_k|s_{k-1}) \qquad (28)
\end{aligned}
$$

となる。ただし，$\boldsymbol{o}[k] := ([\boldsymbol{A}\bar{\boldsymbol{x}}_{\mathrm{p}}]_k, [\boldsymbol{B}\bar{\boldsymbol{x}}_{\mathrm{a}}]_k)^{\mathrm{T}}$ 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遷移系列 $\{s_k\}_{k=1}^K$ は動的計画法により効率的に解くことができる。まず，すべての状態 i について $\delta_1(i)$ を $\delta_1(i) = -\frac{1}{2}(\boldsymbol{o}[1]-\boldsymbol{c}_i[1])^{\mathrm{T}}\boldsymbol{\Upsilon}^{-1}(\boldsymbol{o}[1]-\boldsymbol{c}_i[1])$ と置くと，$k=2,\cdots,K$ について逐次的に $\delta_k(i)$ を $\delta_k(i) = \max_{i'}\left[\delta_{k-1}(i')-\frac{1}{2}(\boldsymbol{o}[k]-\boldsymbol{c}_i[k])^{\mathrm{T}}\boldsymbol{\Upsilon}^{-1}(\boldsymbol{o}[k]-\boldsymbol{c}_i[k])+\phi_{i',i}\right]$ により計算していくことができる。各ステップで選択される状態番号 $\Psi_k(i) = \mathrm{argmax}_{i'}[\delta_{k-1}(i')+\phi_{i',i}]$ を記憶しておくことで，$k=K$ まで到達後に $s_{k-1} = \Psi_k(s_k)$ $(k=K,\cdots,2)$ により選択された状態番号を辿っていくことができ，最適経路 $s_1,\cdots,s_K$ を得ることができる。

2) フレーズ制御パラメータ: ψ の事前分布を $\psi \sim \mathcal{N}(\mu_\psi, 1/\nu_\psi^2)$ とする。式 (6) より，$\boldsymbol{A}$ は定数行列 $\boldsymbol{U}_2$, $\boldsymbol{U}_1$, $\boldsymbol{U}_0$ を用いて $\boldsymbol{A} = \boldsymbol{U}_2\psi^2+\boldsymbol{U}_1\psi+\boldsymbol{U}_0$ と表せるので，Q 関数の ψ に関する偏導関数は 4 次式となり，

$$
\begin{aligned}
& 2\mathrm{tr}(\boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{U}_2\boldsymbol{R}_{\mathrm{p}})\psi^4 + 3\mathrm{tr}(\boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{U}_1\boldsymbol{R}_{\mathrm{p}})\psi^3 \\
& + \{\mathrm{tr}((2\boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{U}_0+\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{U}_1)\boldsymbol{R}_{\mathrm{p}}) - 2\boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{T}}\boldsymbol{U}_2\bar{\boldsymbol{x}}_{\mathrm{p}} + \sigma_{\mathrm{p}}^2\nu_\psi^2\}\psi^2 \\
& + \{\mathrm{tr}(\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{U}_0\boldsymbol{R}_{\mathrm{p}}) - \boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{T}}\boldsymbol{U}_1\bar{\boldsymbol{x}}_{\mathrm{p}} - 2\sigma_{\mathrm{p}}^2\nu_\psi^2\mu_\psi\}\psi - 2K\sigma_{\mathrm{p}}^2
\end{aligned}
$$

の根を解くことで極値を求めることができる。求まった 4 つの極値の中で $\mathcal{I}_2(\psi)$ を最大にする ψ が更新値となる。

3) アクセント制御パラメータ: 同様に φ の事前分布を $\varphi \sim \mathcal{N}(\mu_\varphi, 1/\nu_\varphi^2)$ とする。更新値の導出過程は 4) と同様なので省略する。

4) その他:

$$A_{\mathrm{p}}[k] = \hat{u}_{\mathrm{p}}[k],\ (k \in \mathcal{T}_{\mathrm{p}_1}) \qquad (29)$$

$$A_{\mathrm{a}}^{(n)} = \frac{1}{|\mathcal{T}_{\mathrm{a}_n}|}\sum_{k\in\mathcal{T}_{\mathrm{a}_n}}[\boldsymbol{B}\bar{\boldsymbol{x}}_{\mathrm{a}}]_k,\ \mathcal{T}_{\mathrm{a}_n} = \{k|s_k = \mathrm{a}_n\} \qquad (30)$$

$$\mu_{\mathrm{b}} = \boldsymbol{1}^{\mathrm{T}}\bar{\boldsymbol{x}}_{\mathrm{b}}/T \qquad (31)$$

$$\sigma_{\mathrm{p}}^2 = \big(\mathrm{tr}\big(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{R}_{\mathrm{p}}\big) - 2\boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{T}}\boldsymbol{A}\bar{\boldsymbol{x}}_{\mathrm{p}} + \boldsymbol{\mu}_{\mathrm{p}}^{\mathrm{T}}\boldsymbol{\mu}_{\mathrm{p}}\big)/K \qquad (32)$$

$$\sigma_{\mathrm{a}}^2 = \big(\mathrm{tr}\big(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{R}_{\mathrm{a}}\big) - 2\boldsymbol{\mu}_{\mathrm{a}}^{\mathrm{T}}\boldsymbol{B}\bar{\boldsymbol{x}}_{\mathrm{a}} + \boldsymbol{\mu}_{\mathrm{a}}^{\mathrm{T}}\boldsymbol{\mu}_{\mathrm{a}}\big)/K \qquad (33)$$

$$\sigma_{\mathrm{b}}^2 = \big(\mathrm{tr}(\boldsymbol{R}_{\mathrm{b}}) - 2\mu_{\mathrm{b}}\boldsymbol{1}^{\mathrm{T}}\bar{\boldsymbol{x}}_{\mathrm{b}}\big)/K + \mu_{\mathrm{b}}^2 \qquad (34)$$

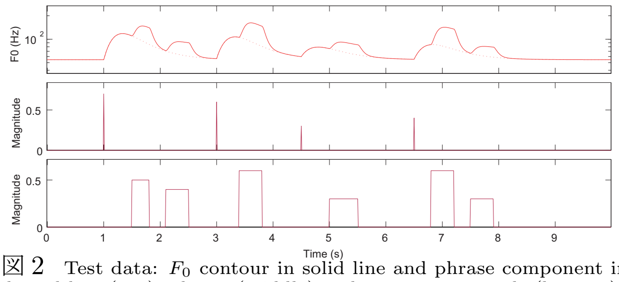

図 2　Test data: F_0 contour in solid line and phrase component in dotted line (top), phrase (middle) and accent commands (bottom).

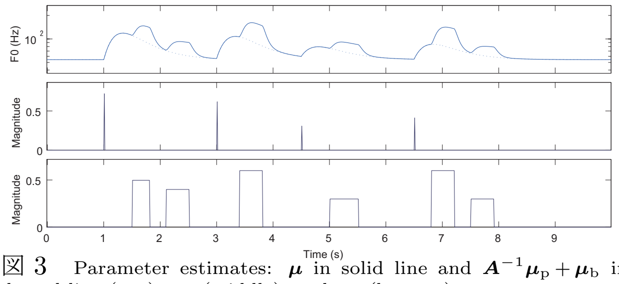

図 3　Parameter estimates: $\boldsymbol{\mu}$ in solid line and $\boldsymbol{A}^{-1}\boldsymbol{\mu}_{\mathrm{p}}+\boldsymbol{\mu}_{\mathrm{b}}$ in dotted line (top), $\boldsymbol{\mu}_{\mathrm{p}}$ (middle), and $\boldsymbol{\mu}_{\mathrm{a}}$ (bottom).

7　実験

提案モデルを 6 章で提案した最適化法のアルゴリズムとしての純粋なふるまいを調べる目的で，元の藤崎モデルを用いて合成した人工の F_0 パターン (図 2) を対象とした動作実験を行った。F_0 パターンの長さは 10s，サンプリング周期は 10ms，各パラメータは図 2 の中段と下段のとおりとした。アルゴリズムの反復回数は 10 とし，N は 10，状態遷移確率はそれぞれ $\phi_{\mathrm{p}_0,\mathrm{p}_0}=\log(0.999)$, $\phi_{\mathrm{p}_0,\mathrm{p}_1}=\log(0.001)$，$\phi_{\mathrm{p}_1,\mathrm{a}_0}=\log(1.0)$，$\phi_{\mathrm{a}_0,\mathrm{a}_0}=\log(0.999)$，$\phi_{\mathrm{a}_n,\mathrm{a}_0}=\log(0.001)$，$\phi_{\mathrm{a}_0,\mathrm{a}_n}=\log(0.0001)$，$\phi_{\mathrm{a}_n,\mathrm{a}_n}=\log(0.899)$，$\phi_{\mathrm{a}_n,\mathrm{p}_0}=\log(0.1)$，$1 \leq n \leq 10$ とした。以上の条件で得た $\boldsymbol{\mu}$, $\boldsymbol{\mu}_{\mathrm{p}}$，$\boldsymbol{\mu}_{\mathrm{a}}$ の推定結果を図 3 に示す。図 3 と図 2 を比較すると，真の値に極めて近い値が得られていることが分かる。このことから式 (20) に相当する最適化問題は提案アルゴリズムにより効果的に解けることが分かった。また，離散時間表現への変換による近似の影響はさほど大きくないことも確認できた。

8　まとめ

本稿では，藤崎モデルの離散時間表現への変換とその統計モデル化を通して F_0 パターン生成過程の統計モデルを構築した。提案モデルの応用例について簡単に触れ，パラメータ推定を EM アルゴリズムによって実現できることを示した。人工の F_0 パターンに対しパラメータ推定アルゴリズムを実行し，その基本動作を確認した。歌声の F_0 パターンのモデル化についても [4] で検討しているので興味のある読者は是非参照されたい。

参考文献

[1] H. Fujisaki, In *Vocal Physiology: Voice Production, Mechanisms and Functions*, (O. Fujimura, ed.) Raven Press, pp. 347–355, 1988.

[2] S. Narusawa et al., In *Proc. ICASSP'02*, Vol. 1, pp. 509–512, 2002.

[3] 亀岡, 音講論 (秋)'10, 1-1-4, 2010.

[4] 大石ら, 音講論 (秋)'10, 3-P-31, 2010.