全極型声道モデルと F_0 パターン生成過程モデルを内部にもつ統一的音声生成モデル*

○亀岡弘和

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 はじめに

人は，音声対話において，声帯の振動周期と声道の共振特性を巧みに時間変化させながらそれぞれに抑揚や意図に関係する情報と音韻に関係する情報を時間に載せて人に伝達している。音声情報処理研究の歴史において，声道および声帯振動に関する重要な数理モデルとして，全極型モデルによる声道スペクトルモデル [1] と藤崎の F_0 パターン生成過程モデル [2] が双璧をなしている。本稿では，これらのモデルを同時に内包する初めての音声モデルを提案し，これを用いて観測信号から声道スペクトルモデルのパラメータと F_0 パターンモデルのパラメータを一挙に推定できることを示す。

2 音声パワースペクトルの統計モデル

本稿では，音声信号モデルとして，n 次調波成分の瞬時振幅が $w_n(u)$，瞬時位相が $n\theta(u)+\varphi_n$ の擬似周期信号の解析信号表現

$$s(u)=\sum_{n=1}^{N} w_n(u)e^{\mathrm{j}(n\theta(u)+\varphi_n)},\quad w_n(u)\geq 0 \tag{1}$$

を考え (u は時刻，N は調波成分の最大次数)，この信号のウェーブレット変換スペクトルを導く。まず，周波数が 1 の，アドミッシブル条件を満たす適当なマザーウェーブレットを $\psi(u)$ とし，ウェーブレット基底関数を

$$\psi_{\alpha,t}(u):=\frac{1}{\sqrt{2\pi}\alpha}\psi\left(\frac{u-t}{\alpha}\right),\quad \alpha>0 \tag{2}$$

と定義する。α はスケールパラメータ，t はシフトパラメータであり，$\psi_{\alpha,t}(u)$ は時刻 t 周辺に局在する周期 α の成分を測るためのウェーブレットである。これを用いて，s の連続ウェーブレット変換を

$$S_{\mathrm{w}}(\log\tfrac{1}{\alpha},t)=\langle s,\psi_{\alpha,t}\rangle=\int_{-\infty}^{\infty} s(u)\psi^*_{\alpha,t}(u)\mathrm{d}u \tag{3}$$

と定義する。ただし，$\langle\cdot,\cdot\rangle$ は内積を表す。

ここで，時刻 t におけるウェーブレット変換スペクトルについて着目する。音声信号 s において，θ と w_n の時間変化はいずれも十分緩やかであると仮定し，まずこれらを時刻 t 周辺でそれぞれ

$$w_n(u)\simeq w_n(t),\quad \theta(u)\simeq\theta(t)+\mu(t)(u-t) \tag{4}$$

のように 0 次近似および 1 次近似する。μ は θ の導関数 (時間微分) であり，s の瞬時基本周波数を表す。式 (4) を式 (1) に代入すると，時刻 t 周辺の s を近似した信号

$$s_t(u)=\sum_{n=1}^{N} w_n(t)e^{\mathrm{j}(n\mu(t)(u-t)+c_n(t))} \tag{5}$$

を得る。ただし，$c_n(t):=n\theta(t)+\varphi_n$ である。s と s_t は大域的にはまったく異なる信号であるが，時刻 t 周辺においてのみ似通っている点がポイントである。通常ウェーブレット基底関数 $\psi_{\alpha,t}$ は時刻 t 周辺でのみ優勢であるため，時刻 t におけるウェーブレット変換スペクトルは，式 (5) を式 (3) に代入して得られる s_t のウェーブレット変換

$$S_{\mathrm{w}}(\log\tfrac{1}{\alpha},t)\simeq\langle s_t,\psi_{\alpha,t}\rangle \tag{6}$$

$$=\sum_{n=1}^{N} w_n(t)e^{\mathrm{j}c_n(t)}\Psi^*(\alpha n\mu(t)) \tag{7}$$

により近似できる。ただし，Ψ はマザーウェーブレット ψ の Fourier 変換である。α は周期を表すので対数周波数 $x=\log\frac{1}{\alpha}$ の関数として S_{w} を表すと，

$$S_{\mathrm{w}}(x,t)=\sum_{n=1}^{N} w_n(t)e^{\mathrm{j}c_n(t)}\Psi^*(ne^{-x+\Omega(t)}) \tag{8}$$

と書ける。ただし，$\Omega(t):=\log\mu(t)$ であり，$\Omega(t)$ は対数瞬時基本周波数を表す。

ここで，周波数特性 $\Psi(\omega)$ が，次のような $\omega=1$ で最大値をとる単峰的な実関数

$$\Psi(\omega):=\begin{cases} e^{-\frac{1}{4\sigma^2}(\log\omega)^2} & (\omega>0)\\ 0 & (\omega\leq 0)\end{cases} \tag{9}$$

となるようなマザーウェーブレットを選ぶと，S_{w} は

$$S_{\mathrm{w}}(x,t)=\sum_{n=1}^{N} w_n(t)e^{\mathrm{j}c_n(t)}e^{-\frac{1}{4\sigma^2}(x-\Omega(t)-\log n)^2} \tag{10}$$

と具体的に書ける。よってそのパワースペクトルは

$$|S_{\mathrm{w}}(x,t)|^2=\lambda(x,t)+o(x,t) \tag{11}$$

$$\lambda(x,t):=\sum_{n=1}^{N} w_n(t)^2e^{-\frac{1}{2\sigma^2}(x-\Omega(t)-\log n)^2} \tag{12}$$

---

*Unified speech production model incorporating all-pole vocal-tract model and F_0 contour generating process model. by KAMEOKA, Hirokazu (NTT Communication Science Laboratories)

と書ける。$o(x,t)$ は調波間干渉を表す交差項であり，調波成分同士の重なりが少ないほど小さくなる。

周期 u_0 でサンプリングされた実際の観測音声信号に対して上記と同条件で算出したウェーブレット変換のパワースペクトルを $Y[k,l] := Y(x_k,t_l)$(ただし k は対数周波数インデックス，l は離散時刻インデックス) とすると，$w_n(t)$ と $\Omega(t)$ が真の値と等しい場合には以上より

$$Y[k,l] \simeq \lambda[k,l] \tag{13}$$

となることが期待される。ただし，$\lambda[k,l] := \lambda(x_k,t_l)$ である。$w_n(t)$ と $\Omega(t)$ が真値と等しくともなお生じる $Y[k,l]$ と $\lambda[k,l]$ との間の誤差にはさまざまな要因があり，例えば，式 (1) の形で定義した擬似周期性の仮定からの逸脱，式 (4) の近似，交差項 o の存在，などが挙げられる。提案法の枠組では，これらの複合的な誤差要因を詳細にモデル化することはせず，まとめて一挙に確率的な現象と見なし，

$$Y[k,l] \sim \mathrm{Pois}(\lambda[k,l]) \tag{14}$$

と具体的に仮定する。ただし，$\mathrm{Pois}(c)$ は c をパラメータとする Poisson 分布を表す。この仮定の下での λ の最尤推定問題は，スペクトル間の近さを測る尺度の一つとして近年音響信号処理分野で多用される I ダイバージェンスと呼ぶ歪み尺度を規準とした Y と λ の最適フィッティング問題と等価となる。提案法の枠組におけるこの仮定の有難みはのちに明らかにする。

3 音声生成モデルに基づく事前分布設計

前述のパワースペクトルモデル $\lambda[k,l]$ において推定すべき未知変数は，瞬時振幅 $w_n[l] := w_n(t_l)$ および対数基本周波数 $\Omega[l] := \Omega(t_l)$ である。提案法のポイントは，これら両未知変数の事前分布を，それぞれ音声生成プロセスのモデルとして古典的で有名な全極型声道モデル [1] と F_0 パターン生成過程モデル [2] に基づいて設計する点にある。以下，その設計方法を順に述べる。

3.1 全極型声道モデル

時刻 t_l における式 (5) の音声信号モデルの，観測信号と同じサンプリング周期 u_0 の下での離散時間表現を $s_l[i] := s_{t_l}(iu_0)$ とし，これを全極型システムからの出力

$$s_l[i] = \sum_{m=0}^{M} a_l[m]s_l[i-m] + \epsilon_l[i] \tag{15}$$

と考える。$a_l[0],\cdots,a_l[M]$ は全極モデルのパラメータであり，線形予測モデルにおいて予測係数と呼ぶものに該当する。ここで，声帯音源に対応する入力 $\epsilon_l[i]$ を

$$\epsilon_l[i] = \sum_{n=1}^{N} v_{l,n} e^{\mathrm{j}n\mu[l]iu_0} \tag{16}$$

のような基本周波数 $\mu[l] := \mu(t_l)$ の周期信号とする。また，上記声帯音源モデルにおける複素振幅 $v_{l,n}$ について

$$v_{l,n} \sim \mathcal{N}_{\mathbb{C}}(0,1), \quad n = 1,\cdots,N \tag{17}$$

を仮定する。ただし $\mathcal{N}_{\mathbb{C}}$ は複素正規分布を表す。式 (15) の両辺に離散時間 Fourier 変換 (DTFT) を行い，$A_l(z) := a_l[0] - a_l[1]z^{-1}\cdots - a_l[M]z^{-M}$ と置いて整理すると，

$$S_l(\omega) = \frac{\sqrt{2\pi}}{A_l(e^{\mathrm{j}\omega})} \sum_{n=1}^{N} v_{l,n}\delta(\omega - n\mu[l]u_0) \tag{18}$$

となる。ただし，ω は規格化角周波数である。上式に対し逆 DTFT を行うと $s_l[i]$ の陽な表現

$$\begin{aligned} s_l[i] &= \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \frac{\sqrt{2\pi}}{A_l(e^{\mathrm{j}\omega})} \sum_{n=1}^{N} v_{l,n}\delta(\omega - n\mu[l]u_0) e^{\mathrm{j}i\omega}\mathrm{d}\omega \\ &= \sum_{n=1}^{N} \frac{v_{l,n}}{A_l(e^{\mathrm{j}n\mu[l]u_0})} e^{\mathrm{j}n\mu[l]iu_0} \end{aligned} \tag{19}$$

を得る。式 (5) を離散時間表現したもの

$$s_l[i] = \sum_{n=1}^{N} w_n[l] e^{\mathrm{j}(n\mu[l](iu_0 - t_l) + c_n(t_l))} \tag{20}$$

と式 (19) を比較すると，

$$w_n[l] = \left| \frac{v_{l,n}}{A_l(e^{\mathrm{j}n\mu[l]u_0})} \right|, \quad n = 1,\cdots,N \tag{21}$$

なる関係式を導ける。$\mu[l]u_0$ は基本周波数 $\mu[l]$ を規格化角周波数の単位で表わしたものなので，$1/|A_l(e^{\mathrm{j}\mu[l]u_0})|$, $\cdots$, $1/|A_l(e^{\mathrm{j}N\mu[l]u_0})|$ は全極スペクトルを基本周波数の間隔でサンプリングした値に他ならない。以上より $w_n[l]$ が従う分布を具体的に導くことができる。式 (17) より

$$\frac{v_{l,n}}{A_l(e^{\mathrm{j}n\mu[l]u_0})} \sim \mathcal{N}_{\mathbb{C}}\big(0, 1/|A_l(e^{\mathrm{j}n\mu[l]u_0})|^2\big) \tag{22}$$

が言え，実部と虚部が独立で，かつ等しい分散の正規分布に従う複素数の絶対値は Rayleigh 分布に従うことが知られており，式 (21), (22) より最終的に

$$w_n[l] \sim \mathrm{Rayleigh}\big(1/|A_l(e^{\mathrm{j}n\mu[l]u_0})|\big) \tag{23}$$

が得られる。この事前分布において，$\Theta_w = \{a_l[m] \mid 1 \le l \le L, 0 \le m \le M\}$ がハイパーパラメータとなる。

3.2 F_0 パターン生成過程モデル

次に，対数基本周波数の時間軌跡 (以後，F_0 パターン) $\boldsymbol{\Omega} := (\Omega[1],\cdots,\Omega[L])^{\mathrm{T}}$ の事前分布について述べる。音声 F_0 パターンの代表的なモデルとして藤崎モデル [2] が有名である。藤崎モデルは喉頭の生理的，物理的特性に基づいて声帯振動制御機構をモデル化したものであり，その離散時間表現の確率過程のモデルが最近筆者らによって定式化された [4]。藤崎モデルの詳細およびその確率モデルの導出過程については [4] に委ねることとす

るが，藤崎モデルパラメータ Θ_Ω が与えられた下での $\boldsymbol{\Omega}$ の条件つき分布が，多次元正規分布

$$\boldsymbol{\Omega} \sim \mathcal{N}(\boldsymbol{\mu}_\Omega, \boldsymbol{\Sigma}_\Omega) \tag{24}$$

の形となることさえ踏まえておけば以後の議論においては十分である。なお，$\boldsymbol{\mu}_\Omega$, $\boldsymbol{\Sigma}_\Omega$ は藤崎モデルパラメータ Θ_Ω の値に応じて決まる変数であるため，Θ_Ω がこの事前分布のハイパーパラメータとなる。

4 パラメータ推定アルゴリズム

以上のモデルにおいて，$\boldsymbol{\Omega}$, $w := \{w_n[l] \mid 1 \le n \le N, 1 \le l \le L\}$, Θ_Ω, Θ_w が未知のパラメータであり，観測パワースペクトル $Y := \{Y[k,l] \mid 1 \le k \le K, 1 \le l \le L\}$ を最も良く説明するような $\boldsymbol{\Omega}$, w, Θ_Ω, Θ_w の値を推定したい。そこで，モデルパラメータの事後確率

$$\begin{aligned} p(\boldsymbol{\Omega}, w, \Theta_\Omega, \Theta_w | Y) &\propto p(Y|\boldsymbol{\Omega}, w) p(\boldsymbol{\Omega}, w, \Theta_\Omega, \Theta_w) \\ &= p(Y|\boldsymbol{\Omega}, w) p(\boldsymbol{\Omega}|\Theta_\Omega) p(w|\Theta_w) p(\Theta_\Omega) p(\Theta_w) \end{aligned} \tag{25}$$

を $\boldsymbol{\Omega}$, w, Θ_Ω, Θ_w に関して最大化する最大事後確率 (MAP) 推定問題を考える。2 章および 3 章より各確率は

$$p(Y|\boldsymbol{\Omega}, w) = \prod_{k,l} \frac{\lambda[k,l]^{Y[k,l]} e^{-\lambda[k,l]}}{Y[k,l]!} \tag{26}$$

$$p(\boldsymbol{\Omega}|\Theta_\Omega) = \frac{|\boldsymbol{\Sigma}_\Omega^{-1}|^{1/2}}{(2\pi)^{L/2}} e^{-\frac{1}{2}(\boldsymbol{\Omega}-\boldsymbol{\mu}_\Omega)^{\mathrm{T}} \boldsymbol{\Sigma}_\Omega^{-1} (\boldsymbol{\Omega}-\boldsymbol{\mu}_\Omega)} \tag{27}$$

$$p(w|\Theta_w) = \prod_{n,l} \frac{w_n[l]}{P_n[l]} e^{-\frac{w_n[l]^2}{2P_n[l]}} \tag{28}$$

で与えられる。ただし，$\lambda[k,l]$ は式 (12) のとおりであり，$P_n[l] := 1/|A_l(e^{\mathrm{j}n\mu[l]u_0})|^2$ である。ここで，以下を仮定する。

(A1) $p(\Theta_w)$ を一様分布とする。
(A2) 対数周波数のサンプリング間隔 $x_0 := x_{k+1} - x_k$ が十分小さく，以下の近似が成り立つ。

$$\int_{-\infty}^{\infty} e^{x^2/2\sigma^2} \mathrm{d}x \simeq x_0 \sum_k e^{x_k^2/2\sigma^2} \tag{29}$$

(A3) 事前分布の信頼度を調節できるようにする目的で，式 (28) 右辺を 2β 乗したものを $p(w|\Theta_w)$ とする。

以上より，対数事後確率 $\log p(\boldsymbol{\Omega}, w, \Theta_\Omega, \Theta_w | Y)$ は

$$\begin{aligned} L :=& \sum_{k,l} Y[k,l] \log \sum_n w_n[l]^2 e^{-\frac{1}{2\sigma^2}(x_k - \Omega[l] - \log n)^2} \\ &- \frac{\sqrt{2\pi}\sigma}{x_0} \sum_{n,l} w_n[l]^2 - \frac{1}{2} \log |\boldsymbol{\Sigma}_\Omega| \\ &- \frac{1}{2} (\boldsymbol{\Omega} - \boldsymbol{\mu}_\Omega)^{\mathrm{T}} \boldsymbol{\Sigma}_\Omega^{-1} (\boldsymbol{\Omega} - \boldsymbol{\mu}_\Omega) \\ &+ 2\beta \sum_{n,l} \left(\log \frac{w_n[l]}{P_n[l]} - \frac{w_n[l]^2}{2P_n[l]} \right) \end{aligned} \tag{30}$$

と定数項を除いて等しくなる。L を最大化する $\boldsymbol{\Omega}$, w, Θ_Ω, Θ_w の解を解析的に解くことはできないが，補助関数法に基づく反復計算により局所最適解を探索することができる。L の補助関数は，$\sum_n \gamma_n[k,l] = 1$ を満たす補助変数 $\gamma_n[k,l]$ を用いて表される不等式

$$\begin{aligned} &\log \sum_n w_n[l]^2 e^{-\frac{1}{2\sigma^2}(x_k - \Omega[l] - \log n)^2} \ge \\ &\sum_n \gamma_n[k,l] \left\{ \log \frac{w_n[l]^2}{\gamma_n[k,l]} - \frac{1}{2\sigma^2} (x_k - \Omega[l] - \log n)^2 \right\} \end{aligned} \tag{31}$$

に基づいて設計される下限関数

$$\begin{aligned} \tilde{L}(\gamma) :=& \sum_{n,k,l} Y[k,l] \gamma_n[k,l] \log \frac{w_n[l]^2}{\gamma_n[k,l]} \\ &- \frac{1}{2} (\boldsymbol{\Omega} - \hat{\boldsymbol{\Omega}})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\Omega} - \hat{\boldsymbol{\Omega}}) \\ &- \frac{\sqrt{2\pi}\sigma}{x_0} \sum_{n,l} w_n[l]^2 - \frac{1}{2} \log |\boldsymbol{\Sigma}_\Omega| \\ &- \frac{1}{2} (\boldsymbol{\Omega} - \boldsymbol{\mu}_\Omega)^{\mathrm{T}} \boldsymbol{\Sigma}_\Omega^{-1} (\boldsymbol{\Omega} - \boldsymbol{\mu}_\Omega) \\ &+ 2\beta \sum_{n,l} \left(\log \frac{w_n[l]}{P_n[l]} - \frac{w_n[l]^2}{2P_n[l]} \right) \end{aligned} \tag{32}$$

で与えられる。ただし，$\hat{\boldsymbol{\Omega}} := (\hat{\Omega}[1], \cdots, \hat{\Omega}[L])^{\mathrm{T}}$ は

$$\hat{\Omega}_l = \frac{\sum_{k,n} Y[k,l] \gamma_n[k,l] (x_k - \log n)}{\sum_{k,n} Y[k,l] \gamma_n[k,l]} \tag{33}$$

を要素とするベクトル，$\boldsymbol{\Sigma}$ は対角要素 Σ_l が

$$\Sigma_l = \sigma^2 \Big/ \sum_{k,n} Y[k,l] \gamma_n[k,l] \tag{34}$$

の対角行列である。

以上の準備により，補助関数法の各ステップにおける更新式が導ける。

1) 補助変数 $\gamma_n[k,l]$:

$$\gamma_n[k,l] \leftarrow \frac{w_n[l]^2 e^{-(x_k - \Omega[l] - \log n)^2/2\sigma^2}}{\sum_n w_n[l]^2 e^{-(x_k - \Omega[l] - \log n)^2/2\sigma^2}} \tag{35}$$

2) $\boldsymbol{F}_0$ パターン $\boldsymbol{\Omega}$:

$$\boldsymbol{\Omega} \leftarrow (\boldsymbol{\Sigma} + \boldsymbol{\Sigma}_\Omega)^{-1} (\boldsymbol{\Sigma}\hat{\boldsymbol{\Omega}} + \boldsymbol{\Sigma}_\Omega \boldsymbol{\mu}_\Omega) \tag{36}$$

3) n 次調波成分の瞬時パワー $w_n[l]$:

$$w_n[l]^2 \leftarrow \frac{\sum_k Y[k,l] \gamma_n[k,l] + \beta}{1 + \beta/P_n[l]} \tag{37}$$

この更新式は $\beta \to \infty$ および $\beta \to 0$ の極限をとると

$$\lim_{\beta\to\infty} \frac{\sum_k Y[k,l]\gamma_n[k,l] + \beta}{1 + \beta/P_n[l]} = P_n[l] \quad (38)$$

$$\lim_{\beta\to 0} \frac{\sum_k Y[k,l]\gamma_n[k,l] + \beta}{1 + \beta/P_n[l]} = \sum_k Y[k,l]\gamma_n[k,l] \quad (39)$$

となる。式 (38) は前段で推定された全極スペクトルを基本周波数の間隔でサンプリングした値，式 (39) は観測スペクトルに最も良くフィットする値をそれぞれ表しており，式 (37) は両者の中間的な値に相当する。

4) 藤崎モデルパラメータ Θ_Ω: [4] の反復アルゴリズムにより $\tilde{L}$ を増加させる Θ_Ω を得ることができる。

5) 全極モデルパラメータ Θ_w: $\tilde{L}$ において Θ_w に関係する項は [3] で設定されている目的関数と同形となるので，Θ_w の更新方法には当該文献と同様のアイディアが適用できる。詳細は [3] に委ねるが，$\boldsymbol{a}_l := (a_l[0], a_l[1], \cdots, a_l[M])^{\mathrm{T}}$ とすると，

$$\boldsymbol{a}_l \leftarrow \boldsymbol{R}^{-1}\boldsymbol{h} \quad (40)$$

$$\boldsymbol{h} \leftarrow \hat{\boldsymbol{R}}\boldsymbol{a}_l \quad (41)$$

を繰り返すことで $\tilde{L}$ を増加させる $\boldsymbol{a}_l$ を得ることができる。ただし，$\boldsymbol{R}_l$ および $\hat{\boldsymbol{R}}_l$ はそれぞれ

$$R_l(m'-m) = \frac{1}{N}\sum_{n=1}^{N} w_n[l]^2 \cos n\mu[l] u_0(m'-m) \quad (42)$$

$$\hat{R}_l(m'-m) = \frac{1}{N}\sum_{n=1}^{N} P_n[l] \cos n\mu[l] u_0(m'-m) \quad (43)$$

を m' 行 m 列成分とする Toeplitz 行列である。

5 動作実験

ATR 音声データベースの女性話者音声データ (サンプリング周波数 16kHz のディジタル信号) を用いて提案法の基本動作の確認を行った。図 1 にその動作実験例を示す。図 1 上段に上記の音声信号に対して $x_0 = 12$cent, $\sigma = 0.02$, $t_0 = t_{k+1} - t_k = 8$ms の条件下で算出したウェーブレット変換スペクトログラム，中段に 4 章で述べた最適化アルゴリズムにより得た $\lambda[k,l]$ の推定値を示す。また下段に F_0 パラメータ $\boldsymbol{\Omega}$(赤線) および推定した藤崎モデルパラメータにより表現される F_0 パターン (青線) を示す。現段階では提案法の具体的な性能を定量的に評価したわけではないが，動作実験の過程で既にいくつかの課題が明らかになった。具体的には，$\boldsymbol{\Omega}$ がまだ明らかに正しい F_0 の解に至っていない段階にもかかわらず収束してしまうケースがしばしば見られた。これは，$\boldsymbol{\Omega}$ と $\boldsymbol{\mu}_\Omega$ は互いに近くなるようにいずれも更新されるため，$\boldsymbol{\Omega}$ の初期値に強い拘束を受けてしまうのが原因であると考えられる。この問題は w と Θ_w についても同様

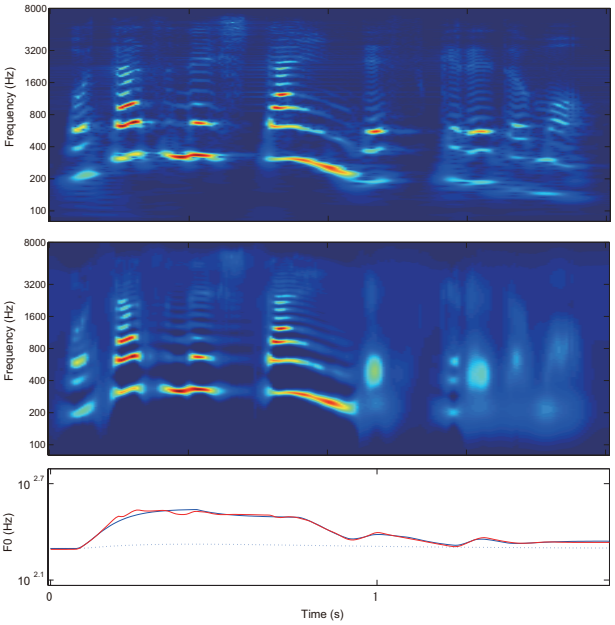

図 1 提案法の動作実験例。ウェーブレット変換による観測スペクトログラム $Y[k,l]$ (上段) と $\lambda[k,l]$ の推定値 (中段) と $\boldsymbol{\Omega}$ の推定値 (下段・赤線) および $\boldsymbol{\mu}_\Omega$ の推定値 (下段・青線)

であった。また，反復計算の初期段階において藤崎モデルのパラメータ推定が不安定な動作をすることが多かった。これは，藤崎モデルのパラメータ更新ステップにおいては，有音区間の $\Omega[l]$ の推定値と本来は無視するべきはずの無音区間における $\Omega[l]$ の推定値も同等の価値を置いて推定を行っていることに原因があると考えられる。以上のことから，$\boldsymbol{\Omega}$ と w を周辺化して Y から直接 Θ_Ω と Θ_w を推定できる方法をいずれ検討していく必要がありそうだという感触を得た。

6 まとめ

本稿では，全極型声道スペクトルモデルと藤崎の F_0 パターン生成過程モデルを事前分布の形として内包する音声ウェーブレットスペクトルの統計モデルを提案し，パラメータの効率的な最適化方法を補助関数法に基づいて導出した。この枠組により，全極型声道モデルと藤崎モデルのパラメータをスペクトルからダイレクトに同時推定することが可能となるが，動作実験の結果，解決すべきいくつかの課題が明らかになった。

参考文献

[1] F. Itakura, S. Saito, In *Proc. 6th Int' l Cong. Acoust. (ICA' 68)*, C-5-5, C17–20, 1968.

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

[3] A. El-Jaroudi, J. Makhoul, *IEEE trans. Signal Process.*, Vol. 39, No. 2, pp. 411–423, 1991.

[4] 亀岡ら, 音講論 (秋), 1-1-3, 2010.