社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

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計的音声スペクトルモデル

亀岡弘和†

† 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
〒 243-0198 神奈川県厚木市森の里若宮 3-1
E-mail: †kameoka@cs.brl.ntt.co.jp

あらまし　音声情報処理研究の歴史において，音声生成過程に関する重要な数理モデルとして，声帯振動の制御機構を模した藤崎モデルと，縦続連結した等長音響管で声道断面積関数をモデル化したことと等価な全極型声道モデルが有名である。本発表では，ある特定条件のウェーブレット変換のスペクトル領域において，これら両生成過程モデルを同時に内包する確率モデルが立てられることを示し，そのパラメータの推論法について述べる。提案法は，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同時に直接推定することが可能な初めての音声分析手法であり，様々な興味深い応用展開が期待される。

キーワード　全極型声道モデル，F_0 パターン生成過程モデル (藤崎モデル)，ウェーブレット変換，EM アルゴリズム

Statistical speech spectrum model incorporating all-pole vocal tract model and F_0 contour generating process model

Hirokazu KAMEOKA†

† NTT Communication Science Laboratories
3-1 Morinosato Wakamiya, Atsugi, Kanagawa, 243-0198 Japan
E-mail: †kameoka@cs.brl.ntt.co.jp

Abstract　In this paper, we propose to introduce a statistical speech spectrum model, simultaneously incorporating the all-pole vocal tract model and the F_0 contour generating process model.

Key words　All-pole vocal tract model, F_0 contour generating process model, wavelet transform, EM algorithm

1. ま え が き

音声は，肺から供給された空気が声帯振動によって周期的な疎密波となり，さらにその音波が声道で共振されることによって「声」となり，口から発せられる。人は音声対話において，声帯の固有振動数を巧みに制御しながら音声に適切な抑揚を与え，同時に，声道の形状を巧みに変化させながら異なる種類の母音を生成することで，聴き手に意図した情報を伝達している。音声情報処理研究の歴史において，音声生成過程に関する重要な数理モデルとして，声帯振動の制御機構を模した藤崎モデルと，縦続連結した等長音響管で声道断面積関数をモデル化したことと等価な全極型声道モデルが有名である。本稿では，全極型声道モデルと藤崎モデルを同時に内包する統計的音声スペクトルモデルを提案する。

声道スペクトルと基本周波数 (F_0) は，分析，合成，認識，強調，符号化などの幅広い音声アプリケーションにおいて根幹となる音声の重要な特徴量であり，音声信号からこれらを自動的に推定する技術は長年にわたって研究され，発展を遂げてきた。声道スペクトルを推定する問題と F_0 を推定する問題は，一見すると別の問題のようで実は密接に関係している。例えば，以下，音声信号が声道のインパルス応答と声帯による周期的なパルス波との畳み込みで表現できる場合を考えると，声道スペクトルは調波成分のピークをつないだ包絡線に対応するので，F_0 がもし既知であればどの成分が調波成分であるかが特定でき，声道スペクトルを精度良く推定することができる。一方で，F_0 推定では，実際の F_0 の整数分の一倍と推定し誤ってしまうことがしばしば問題となるが，この半ピッチ誤りの問題は，声道スペクトルが滑らかであるという先験的な知識を使えば効果的に回避することができる (半ピッチ誤りの状況では滑らかでないスペクトル包絡を仮定してしまっていることに相当するため)。また，音声の F_0 の時間軌跡 (F_0 パターンと呼ぶ。) も通常は滑らかであり，これを制約とすれば同様の推定誤りを回避できることが多い。以上のように，声道スペクトルと F_0 パターンは各々の推定において互いに有用な情報であるため，別個に求めるより，同時最適化問題として推定できるようにした方が多分に合理的である。以上の動機に基づき，スペクトル包絡構造と F_0 パターンを同時モデル化したことが提案モデルにおける第一のポイントである。

提案モデルの第二の重要なポイントは，音声生成過程の確率モデル化を通して，与えられた任意の信号がどれだけ音声のものらしいかを定量的に評価することが可能になる点である。音声生成過程から生成しえない信号には低いスコアを，音声生成過程から生成しうる信号には高いスコアを与えるような規準がもしあれば，音声分離，雑音除去，残響除去などの音声を対象としたあらゆる逆問題において音声を抽出する手掛かりとして大いに活用できることが期待されるからである。

以上より，スペクトル包絡構造と F_0 パターンをいかにして同時モデル化するかという点と，いかに音声の生成過程を正確に統計モデルの形で表現できるかという点に主眼を置き，以下で音声スペクトルのモデル化を行う。

2. 音声パワースペクトルモデル

本章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 $n\theta(u)+\rho_n$ の擬似周期信号の解析信号表現

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

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。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

$$\psi_{x,t}(u):=\frac{1}{\sqrt{2\pi e^{-x}}}\psi\left(\frac{u-t}{e^{-x}}\right) \tag{2}$$

と定義する。e^{-x} と t はスケールとシフトパラメータに該当し，$\psi_{x,t}(u)$ は時刻 t 周辺に局在する周期 e^{-x} の成分を測るためのウェーブレットである。x は対数周波数に対応する。これを用いて，s の連続ウェーブレット変換を

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

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，これらを時刻 t 周辺で

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

のように 0 次近似および 1 次近似する。$\dot{\theta}$ 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

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

を得る。ただし，$\xi_n(t):=n\theta(t)+\rho_n$ 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 $\psi_{x,t}$ 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

$$S_{\mathrm{w}}(x,t)\simeq\langle s_t,\psi_{x,t}\rangle \tag{6}$$

$$=\sum_{n=1}^{N} w_n(t)e^{\mathrm{j}\xi_n(t)}\Psi^*(ne^{-x+\Omega(t)}) \tag{7}$$

により近似できる。ただし，Ψ はマザーウェーブレット ψ のフーリエ変換である。また，$\Omega(t):=\log\dot{\theta}(t)$ であり，$\Omega(t)$ は対数瞬時基本周波数を表す。以上より，ウェーブレット変換のスペクトルはマザーウェーブレット $\psi(t)$ を具体的にどう定めるかによって決まる。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

$$\Psi(\omega):=\begin{cases} e^{-\frac{1}{4\sigma^2}(\log\omega)^2} & (\omega>0)\\ 0 & (\omega\leqq 0)\end{cases} \tag{8}$$

となるようなマザーウェーブレットを選ぶと，S_{w} は

$$S_{\mathrm{w}}(x,t)=\sum_{n=1}^{N} w_n(t)e^{\mathrm{j}\xi_n(t)}e^{-\frac{1}{4\sigma^2}(x-\Omega(t)-\log n)^2} \tag{9}$$

と具体的に書ける。このマザーウェーブレットによって構成される定 Q フィルタバンクを定 Q 対数正規分布型フィルタバンクと呼ぶことにする。ここで，$|S_{\mathrm{w}}(x,t)|^2$ は

$$|S_{\mathrm{w}}(x,t)|^2=\lambda(x,t)+o(x,t) \tag{10}$$

$$\lambda(x,t):=\sum_{n=1}^{N} w_n(t)^2e^{-\frac{1}{2\sigma^2}(x-\Omega(t)-\log n)^2} \tag{11}$$

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

$$Y[k,l]\simeq\lambda[k,l] \tag{12}$$

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，

$$Y[k,l]\sim\mathrm{Pois}(\lambda[k,l]) \tag{13}$$

と具体的に仮定する。ただし，Pois はポアソン分布を表す。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。

以上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，それぞれ音声生成過程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] に基づいて設計し，スペクトルの生成過程を階層モデルにより表現する点にある。以下，各事前分布の設計方法を順に述べる。

3. 全極モデルに基づく瞬時振幅の確率分布

本章では，全極型声道モデルに基づき $w_n[l]$ の事前分布を導く。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

$$s_l[i] = \sum_{m=0}^{M} a_l[m]s_l[i-m] + \epsilon_l[i] \tag{14}$$

と考えよう。$a_l[0], \cdots, a_l[M]$ 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，$s_l[i]$ の基本周波数は $e^{\Omega[l]}$ であるため，声帯音源に対応する入力 $\epsilon_l[i]$ も基本周波数が $e^{\Omega[l]}$ の周期信号

$$\epsilon_l[i] = \sum_{n=1}^{N} v_{l,n} e^{\mathrm{j}ne^{\Omega[l]}iu_0} \tag{15}$$

と仮定しておく必要がある (線形フィルタには入力の周波数成分の周波数値を変える効力はないため)。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

$$v_{l,n} \sim \mathcal{N}_{\mathbb{C}}(0,1), \quad n = 1, \cdots, N \tag{16}$$

を仮定する。ただし，$\mathcal{N}_{\mathbb{C}}$ は複素正規分布を表す。よって，この仮定は，駆動音源に関してすべての調波成分が等しいパワースペクトル密度をもち，位相を一様にとりうることを仮定していることに相当する。式 (14) の両辺に離散時間フーリエ変換 (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，

$$S_l(\omega) = \frac{\sqrt{2\pi}}{A_l(e^{\mathrm{j}\omega})} \sum_{n=1}^{N} v_{l,n}\delta(\omega - ne^{\Omega[l]}u_0) \tag{17}$$

となる。ただし，S_l は $s_l[1], \cdots, s_l[T]$ の DTFT，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

$$\begin{aligned} s_l[i] &= \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \frac{\sqrt{2\pi}}{A_l(e^{\mathrm{j}\omega})} \sum_{n=1}^{N} v_{l,n}\delta(\omega - ne^{\Omega[l]}u_0)e^{\mathrm{j}i\omega}\mathrm{d}\omega \\ &= \sum_{n=1}^{N} \frac{v_{l,n}}{A_l(e^{\mathrm{j}ne^{\Omega[l]}u_0})} e^{\mathrm{j}ne^{\Omega[l]}iu_0} \end{aligned} \tag{18}$$

を得る。式 (5) を離散時間表現したもの

$$s_l[i] = \sum_{n=1}^{N} w_n[l]e^{\mathrm{j}(ne^{\Omega[l]}(iu_0 - t_l) + c_n(t_l))} \tag{19}$$

と式 (18) を比較すると，

$$w_n[l] = \left| \frac{v_{l,n}}{A_l(e^{\mathrm{j}ne^{\Omega[l]}u_0})} \right|, \quad n = 1, \cdots, N \tag{20}$$

なる関係式を導ける。$e^{\Omega[l]}u_0$ は基本周波数 $e^{\Omega[l]}$ を規格化角周波数の単位で表わしたものなので，$1/|A_l(e^{\mathrm{j}e^{\Omega[l]}u_0})|$, $\cdots$, $1/|A_l(e^{\mathrm{j}Ne^{\Omega[l]}u_0})|$ 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，全極モデルパラメータおよび $\Omega[l]$ が与えられた下での $w_n[l]$ が従う分布を具体的に導くことができる。式 (16) より

$$\frac{v_{l,n}}{A_l(e^{\mathrm{j}ne^{\Omega[l]}u_0})} \sim \mathcal{N}_{\mathbb{C}}\big(0, 1/|A_l(e^{\mathrm{j}ne^{\Omega[l]}u_0})|^2\big) \tag{21}$$

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レイリー分布に従うため，式 (20), (21) より

$$w_n[l] \sim \mathrm{Rayleigh}\big(1/|A_l(e^{\mathrm{j}ne^{\Omega[l]}u_0})|\big) \tag{22}$$

が得られる。この事前分布のパラメータ (ハイパーパラメータ) は，$\pi_w = \{a_l[m] \mid 1 \leqq l \leqq L, 0 \leqq m \leqq M\}$ である。この事前分布は，$\Omega[l]$ に依存する点に注意が必要である。

4. 藤崎モデルに基づく F_0 パターンの確率分布

本章では，藤崎モデルに基づいて $\boldsymbol{\Omega} := (\Omega[1], \cdots, \Omega[L])^{\mathrm{T}}$ の事前分布を導く。藤崎モデルは喉頭の声帯振動制御機構の物理モデルであるが，これを離散時間表現の確率過程として記述した確率モデルが最近筆者らによって提案された [4]。本章の内容はこの新モデルがベースになっている。

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，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 $\Omega_{\mathrm{p}}(t)$, $\Omega_{\mathrm{a}}(t)$ とする。$\Omega_{\mathrm{p}}(t)$ 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，$\Omega_{\mathrm{a}}(t)$ 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 Ω_b の和 $\Omega_{\mathrm{p}}(t) + \Omega_{\mathrm{a}}(t) + \Omega_b$ で F_0 パターン $\Omega(t)$ を表したものが藤崎モデルである。フレーズ制御機構およびアクセント制御機構は

$$G_{\mathrm{p}}(t) = \alpha^2 te^{-\alpha t} \quad (t \geqq 0) \tag{23}$$

$$G_{\mathrm{a}}(t) = \beta^2 te^{-\beta t} \quad (t \geqq 0) \tag{24}$$

で与えられるインパルス応答 $G_{\mathrm{p}}(t)$ と $G_{\mathrm{a}}(t)$ 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。フレーズ指令とアクセント指令をそれぞれ $u_{\mathrm{p}}(t)$ と $u_{\mathrm{a}}(t)$ のように時間の関数として表せば，$\Omega(t)$ は $G_{\mathrm{p}}(t)$ と $G_{\mathrm{a}}(t)$ を用いて

$$\Omega(t) = \Omega_{\mathrm{b}} + \Omega_{\mathrm{p}}(t) + \Omega_{\mathrm{a}}(t) \tag{25}$$

$$\Omega_{\mathrm{p}}(t) = (G_{\mathrm{p}} * u_{\mathrm{p}})(t) \tag{26}$$

$$\Omega_{\mathrm{a}}(t) = (G_{\mathrm{a}} * u_{\mathrm{a}})(t) \tag{27}$$

と表される。ただし，$*$ は畳み込みを表す。

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 (ラプラス変換) は $\mathcal{G}_{\mathrm{p}}(s) = \mathcal{L}\big[G_{\mathrm{p}}(t)\big] = \alpha^2/(s+\alpha)^2$ である。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 (ただし，$t_0 = t_l - t_{l-1}$)，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 $\mathcal{G}_{\mathrm{p}}^{-1}(z) = g_{\mathrm{p}2}z^{-2} + g_{\mathrm{p}1}z^{-1} + g_{\mathrm{p}0}$ と書ける。ただし，$g_{\mathrm{p}2} = (\phi-1)^2$, $g_{\mathrm{p}1} = -2\phi(\phi-1)$, $g_{\mathrm{p}0} = \phi^2$, $\phi = 1 + 1/(\alpha t_0)$ である。ここで，時刻 t_l におけるフレーズ指令関数およびフレーズ成分をそれぞれ $u_{\mathrm{p}}[l]$, $\Omega_{\mathrm{p}}[l]$ と表すと，$\Omega_{\mathrm{p}}[l]$ は単一のパラメータ ϕ(または 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

$$u_{\mathrm{p}}[l] = g_{\mathrm{p}0}\Omega_{\mathrm{p}}[l] + g_{\mathrm{p}1}\Omega_{\mathrm{p}}[l-1] + g_{\mathrm{p}2}\Omega_{\mathrm{p}}[l-2] \tag{28}$$

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

$$u_{\mathrm{a}}[l] = g_{\mathrm{a}0}\Omega_{\mathrm{a}}[k] + g_{\mathrm{a}1}\Omega_{\mathrm{a}}[l-1] + g_{\mathrm{a}2}\Omega_{\mathrm{a}}[l-2] \tag{29}$$

と書くことができる。ただし，$g_{\mathrm{a}2} = (\varphi-1)^2$, $g_{\mathrm{a}1} = -2\varphi(\varphi-1)$, $g_{\mathrm{a}0} = \varphi^2$, $\varphi = 1 + 1/(\beta t_0)$ である。ベースライン成分 $\Omega_{\mathrm{b}}(t)$ の離散時間表現を $\Omega_{\mathrm{b}}[l]$ 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 $\Omega[l] = \Omega_{\mathrm{p}}[l] + \Omega_{\mathrm{a}}[l] + \Omega_{\mathrm{b}}[l]$ で与えられる。

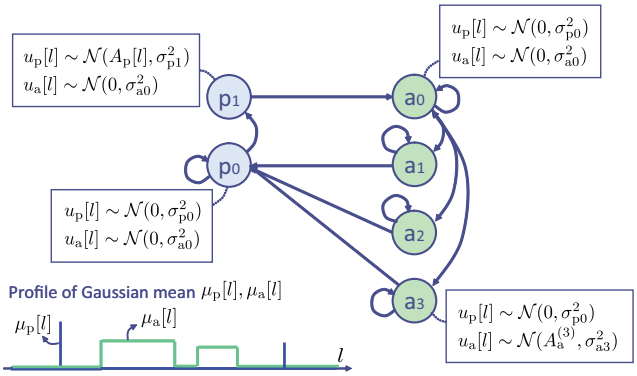

図 1　Command function modeling with HMM.

次に，$u_{\mathrm{p}}[l]$ と $u_{\mathrm{a}}[l]$ を確率モデル化する。そこで，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ことに注意する。

(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)** 複数の指令が同時に発生することはない。

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 $u_{\mathrm{p}}[l]$ と $u_{\mathrm{a}}[l]$ をモデル化する。$u_{\mathrm{p}}[l]$ と $u_{\mathrm{a}}[l]$ を組にしたベクトル $\boldsymbol{o}[l] := (u_{\mathrm{p}}[l], u_{\mathrm{a}}[l])^{\mathrm{T}}$ を考え，$\boldsymbol{o}[l]$ を

$$\boldsymbol{o}[l] = \boldsymbol{\nu}[l] + \boldsymbol{\epsilon}[l] \tag{30}$$

のようにフレーズ指令からなる時系列 $\mu_{\mathrm{p}}[l]$ とアクセント指令からなる時系列 $\mu_{\mathrm{a}}[l]$ を組にしたベクトル $\boldsymbol{\nu}[l] = (\mu_{\mathrm{p}}[l], \mu_{\mathrm{a}}[l])^{\mathrm{T}}$ に雑音 $\boldsymbol{\epsilon}[l]$ が加わったものと見なす。$\boldsymbol{\epsilon}[l]$ を共分散行列が $\boldsymbol{\Upsilon}[l] = \mathrm{diag}(\upsilon^2_{\mathrm{p}}[l], \upsilon^2_{\mathrm{a}}[l])$ のガウス性雑音とすると，以上より

$$\boldsymbol{o}[l] \sim \mathcal{N}\left(\boldsymbol{\nu}[l], \boldsymbol{\Upsilon}[l]\right) \tag{31}$$

が言える。ここで，$\boldsymbol{\nu}[l]$ および $\boldsymbol{\Upsilon}[l]$ が状態遷移に伴って変化する図 1 のような HMM を考える。HMM による $\boldsymbol{o}[l]$ のモデル化により，状態遷移の経路制限を通して両指令の組 $\boldsymbol{\nu}[l]$ に対して (A1), (A2) を満たすような制約を巧く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。

出力系列: $\{\boldsymbol{o}[l]\}_{l=1}^{L}$
状態集合: $\mathcal{S} := \{\mathrm{p}_0, \mathrm{p}_1, \mathrm{a}_0, \cdots, \mathrm{a}_N\}$
状態系列: $\{s_l\}_{l=1}^{L}$
出力分布: $P(\boldsymbol{o}[l]|s_l = j) = \mathcal{N}(\boldsymbol{c}_j[l], \boldsymbol{D}_j)$
$\boldsymbol{c}_{\mathrm{p}_0}[l] = \boldsymbol{c}_{\mathrm{a}_0}[l] = (0,0)^{\mathrm{T}}$
$\boldsymbol{c}_{\mathrm{p}_1}[l] = (A_{\mathrm{p}}[l], 0)^{\mathrm{T}}$, $\boldsymbol{c}_{\mathrm{a}_n}[l] = (0, A^{(n)}_{\mathrm{a}})^{\mathrm{T}}$
$\boldsymbol{D}_{\mathrm{p}_0} = \boldsymbol{D}_{\mathrm{a}_0} = \mathrm{diag}(\sigma^2_{\mathrm{p}_0}, \sigma^2_{\mathrm{a}_0})$
$\boldsymbol{D}_{\mathrm{p}_1} = \mathrm{diag}(\sigma^2_{\mathrm{p}_1}, \sigma^2_{\mathrm{a}_0})$, $\boldsymbol{D}_{\mathrm{a}_n} = \mathrm{diag}(\sigma^2_{\mathrm{p}_0}, \sigma^2_{\mathrm{a}_n})$
遷移確率: $\phi_{j',j} := \log P(s_l = j | s_{l-1} = j')$

簡単のため状態遷移確率および分散 $\sigma^2_{\mathrm{p}_0}$, $\sigma^2_{\mathrm{p}_1}$ $\sigma^2_{\mathrm{a}_0}, \cdots, \sigma^2_{\mathrm{a}_N}$ 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未知パラメータは，状態遷移系列 s_l，フレーズ指令の大きさ $A_{\mathrm{p}}[l]$，アクセント指令の大きさ $\{A^{(n)}_{\mathrm{a}}\}_{n=1}^{N}$ であり，これらをまとめて π_u と記す。また，平均値系列 $\{\mu_{\mathrm{p}}[l]\}_{l=1}^{L}$, $\{\mu_{\mathrm{a}}[l]\}_{l=1}^{L}$ および分散値系列 $\{\upsilon_{\mathrm{p}}[l]\}_{l=1}^{L}$, $\{\upsilon_{\mathrm{a}}[l]\}_{l=1}^{L}$ は，状態遷移系列 $\{s_l\}_{l=1}^{L}$ がひとたび決まれば $(\mu_{\mathrm{p}}[l], \mu_{\mathrm{a}}[l])^{\mathrm{T}} \leftarrow \boldsymbol{c}_{s_l}[l]$, $\mathrm{diag}(\upsilon_{\mathrm{p}}[l], \upsilon_{\mathrm{a}}[l]) \leftarrow \boldsymbol{D}_{s_l}$ $(l = 1, \cdots, L)$ により一挙に決まる。

前述の指令入力モデルに基づき $\boldsymbol{\Omega} = (\Omega[1], \cdots, \Omega[L])^{\mathrm{T}}$ の確率密度関数を導く。式 (31) より，$\boldsymbol{u}_{\mathrm{p}} := (u_{\mathrm{p}}[1], \cdots, u_{\mathrm{p}}[L])^{\mathrm{T}}$, $\boldsymbol{u}_{\mathrm{a}} := (u_{\mathrm{a}}[1], \cdots, u_{\mathrm{a}}[L])^{\mathrm{T}}$, $\boldsymbol{\mu}_{\mathrm{p}} := (\mu_{\mathrm{p}}[1], \cdots, \mu_{\mathrm{p}}[L])^{\mathrm{T}}$, $\boldsymbol{\mu}_{\mathrm{a}} := (\mu_{\mathrm{a}}[1], \cdots, \mu_{\mathrm{a}}[L])^{\mathrm{T}}$ とすると，

$$\boldsymbol{u}_{\mathrm{p}} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{p}}, \boldsymbol{\Sigma}_{\mathrm{p}}), \quad \boldsymbol{\Sigma}_{\mathrm{p}} = \mathrm{diag}(\upsilon^2_{\mathrm{p}}[1], \cdots, \upsilon^2_{\mathrm{p}}[L]) \tag{32}$$
$$\boldsymbol{u}_{\mathrm{a}} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{a}}, \boldsymbol{\Sigma}_{\mathrm{a}}), \quad \boldsymbol{\Sigma}_{\mathrm{a}} = \mathrm{diag}(\upsilon^2_{\mathrm{a}}[1], \cdots, \upsilon^2_{\mathrm{a}}[L]) \tag{33}$$

が言える。式 (28), (29) の関係式より，フレーズ成分 $\boldsymbol{\Omega}_{\mathrm{p}} := (\Omega_{\mathrm{p}}[1], \cdots, \Omega_{\mathrm{p}}[L])^{\mathrm{T}}$ と $\boldsymbol{u}_{\mathrm{p}}$ の関係，および，アクセント成分 $\boldsymbol{\Omega}_{\mathrm{a}} := (\Omega_{\mathrm{a}}[1], \cdots, \Omega_{\mathrm{a}}[L])^{\mathrm{T}}$ と $\boldsymbol{u}_{\mathrm{a}}$ の関係は，

$$\boldsymbol{G}_{\mathrm{p}} := \begin{bmatrix} g_{\mathrm{p}_0} & & O \\ g_{\mathrm{p}_1} & g_{\mathrm{p}_0} & \\ g_{\mathrm{p}_2} & g_{\mathrm{p}_1} & g_{\mathrm{p}_0} \\ & \ddots & \ddots & \ddots \\ O & & g_{\mathrm{p}_2} & g_{\mathrm{p}_1} & g_{\mathrm{p}_0} \end{bmatrix}, \quad \boldsymbol{G}_{\mathrm{a}} := \begin{bmatrix} g_{\mathrm{a}_0} & & O \\ g_{\mathrm{a}_1} & g_{\mathrm{a}_0} & \\ g_{\mathrm{a}_2} & g_{\mathrm{a}_1} & g_{\mathrm{a}_0} \\ & \ddots & \ddots & \ddots \\ O & & g_{\mathrm{a}_2} & g_{\mathrm{a}_1} & g_{\mathrm{a}_0} \end{bmatrix}$$

と置くと，それぞれ $\boldsymbol{u}_{\mathrm{p}} = \boldsymbol{G}_{\mathrm{p}}\boldsymbol{\Omega}_{\mathrm{p}}$, $\boldsymbol{u}_{\mathrm{a}} = \boldsymbol{G}_{\mathrm{a}}\boldsymbol{\Omega}_{\mathrm{a}}$ のように表現できることから，式 (32), (33) より

$$\boldsymbol{\Omega}_{\mathrm{p}} \sim \mathcal{N}\left(\boldsymbol{G}_{\mathrm{p}}^{-1}\boldsymbol{\mu}_{\mathrm{p}}, \boldsymbol{G}_{\mathrm{p}}^{-1}\boldsymbol{\Sigma}_{\mathrm{p}}(\boldsymbol{G}_{\mathrm{p}}^{-1})^{\mathrm{T}}\right) \tag{34}$$
$$\boldsymbol{\Omega}_{\mathrm{a}} \sim \mathcal{N}\left(\boldsymbol{G}_{\mathrm{a}}^{-1}\boldsymbol{\mu}_{\mathrm{a}}, \boldsymbol{G}_{\mathrm{a}}^{-1}\boldsymbol{\Sigma}_{\mathrm{a}}(\boldsymbol{G}_{\mathrm{a}}^{-1})^{\mathrm{T}}\right) \tag{35}$$

が導かれる。ベース成分 $\Omega_{\mathrm{b}}[l]$ についても，同様に $\boldsymbol{\Omega}_{\mathrm{b}} \sim \mathcal{N}(\mu_{\mathrm{b}}\mathbf{1}, \boldsymbol{\Sigma}_{\mathrm{b}})$ とし，$\boldsymbol{\Omega}_{\mathrm{p}}$, $\boldsymbol{\Omega}_{\mathrm{a}}$ とは独立と仮定する。ただし，$\boldsymbol{\Sigma}_{\mathrm{b}}$ は定数とする。仮定より，$\boldsymbol{\Omega}_{\mathrm{p}}$, $\boldsymbol{\Omega}_{\mathrm{a}}$, $\boldsymbol{\Omega}_{\mathrm{b}}$ は独立なので，藤崎モデルパラメータ $\pi_\Omega := \{\pi_u, \alpha, \beta, \mu_{\mathrm{b}}\}$ が与えられたもとでの F_0 パターン $\boldsymbol{\Omega} = \boldsymbol{\Omega}_{\mathrm{p}} + \boldsymbol{\Omega}_{\mathrm{a}} + \boldsymbol{\Omega}_{\mathrm{b}}$ の確率密度関数は，

$$\begin{aligned}\boldsymbol{\Omega} \sim \mathcal{N}\big(&\boldsymbol{G}_{\mathrm{p}}^{-1}\boldsymbol{\mu}_{\mathrm{p}} + \boldsymbol{G}_{\mathrm{a}}^{-1}\boldsymbol{\mu}_{\mathrm{a}} + \mu_{\mathrm{b}}\mathbf{1}, \\ &\boldsymbol{G}_{\mathrm{p}}^{-1}\boldsymbol{\Sigma}_{\mathrm{p}}(\boldsymbol{G}_{\mathrm{p}}^{-1})^{\mathrm{T}} + \boldsymbol{G}_{\mathrm{a}}^{-1}\boldsymbol{\Sigma}_{\mathrm{a}}(\boldsymbol{G}_{\mathrm{a}}^{-1})^{\mathrm{T}} + \boldsymbol{\Sigma}_{\mathrm{b}}\big)\end{aligned} \tag{36}$$

で与えられる。以上より，以下を得る。

$$\begin{aligned} P(\boldsymbol{\Omega}|\pi_\Omega) &= \frac{|\boldsymbol{\Sigma}^{-1}|^{1/2}}{(2\pi)^{T/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\Omega} - \boldsymbol{\mu}_\Omega)^{\mathrm{T}}\boldsymbol{\Sigma}_\Omega^{-1}(\boldsymbol{\Omega} - \boldsymbol{\mu}_\Omega)\right\} \\ \boldsymbol{\mu}_\Omega &= \boldsymbol{G}_{\mathrm{p}}^{-1}\boldsymbol{\mu}_{\mathrm{p}} + \boldsymbol{G}_{\mathrm{a}}^{-1}\boldsymbol{\mu}_{\mathrm{a}} + \mu_{\mathrm{b}}\mathbf{1} \\ \boldsymbol{\Sigma}_\Omega &= \boldsymbol{G}_{\mathrm{p}}^{-1}\boldsymbol{\Sigma}_{\mathrm{p}}\big(\boldsymbol{G}_{\mathrm{p}}^{\mathrm{T}}\big)^{-1} + \boldsymbol{G}_{\mathrm{a}}^{-1}\boldsymbol{\Sigma}_{\mathrm{a}}\big(\boldsymbol{G}_{\mathrm{a}}^{\mathrm{T}}\big)^{-1} + \boldsymbol{\Sigma}_{\mathrm{b}} \end{aligned} \tag{37}$$

5.　パラメータ推定アルゴリズム

以上のモデルをグラフィカルモデルで表わすと図 2 のようになる。$\boldsymbol{\Omega}$, $w := \{w_n[l] \mid 1 \leqq n \leqq N, 1 \leqq l \leqq L\}$, π_Ω, π_w が未知のパラメータであり，観測スペクトル $Y := \{Y[k,l]\}_{1 \leqq k \leqq K, 1 \leqq l \leqq L}$ を最も良く説明するような $\boldsymbol{\Omega}$, w, π_Ω, π_w の値を推定することが目的である。そこで，モデルパラメータの事後確率

$$\begin{aligned} p(\boldsymbol{\Omega}, w, \pi_\Omega, \pi_w|Y) &\propto p(Y|\boldsymbol{\Omega}, w)p(\boldsymbol{\Omega}, w, \pi_\Omega, \pi_w) \\ &= p(Y|\boldsymbol{\Omega}, w)p(\boldsymbol{\Omega}|\pi_\Omega)p(w|\boldsymbol{\Omega}, \pi_w)p(\pi_\Omega) \end{aligned} \tag{38}$$

を $\boldsymbol{\Omega}$, w, π_Ω, π_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.〜4. より各確率は

$$p(Y|\boldsymbol{\Omega}, w) = \prod_{k,l} \frac{\lambda[k,l]^{Y[k,l]} e^{-\lambda[k,l]}}{Y[k,l]!} \tag{39}$$

$$p(\boldsymbol{\Omega}|\pi_\Omega) = \frac{|\boldsymbol{\Sigma}_\Omega^{-1}|^{1/2}}{(2\pi)^{L/2}} e^{-\frac{1}{2}(\boldsymbol{\Omega} - \boldsymbol{\mu}_\Omega)^{\mathrm{T}}\boldsymbol{\Sigma}_\Omega^{-1}(\boldsymbol{\Omega} - \boldsymbol{\mu}_\Omega)} \tag{40}$$

$$p(w|\boldsymbol{\Omega}, \pi_w) = \prod_{n,l} \frac{w_n[l]}{P_n[l]} e^{-\frac{w_n[l]^2}{P_n[l]}} \tag{41}$$

で与えられる。ただし，$\lambda[k,l]$ は式 (11) のとおりであり，$P_n[l] := 1/|A_l(e^{jne^{\Omega[l]}u_0})|^2$ である。以上の MAP 推定問題の大域最適解は解析的に解くことはできないが，以下に示すように Expectation-Maximization(EM) アルゴリズムにより局所解を得ることができる。

EM アルゴリズムの導出では，潜在データと呼ぶ，観測不可な架空のデータをどのように定義するかがポイントとなる。ここでは，観測スペクトル $Y[k,l]$ のうち n 番目の高調波に対応しているであろう成分 $C_n[k,l]$ を潜在データと見なす。従って，

$$Y[k,l] = \sum_n C_n[k,l] \tag{42}$$

を満たす。今，我々は $Y[k,l] \sim \mathrm{Pois}(\lambda[k,l])$ を仮定しているため，これと矛盾しないように $C_n[k,l]$ が従う分布を仮定する必要がある。スペクトルモデル $\lambda[k,l]$ が n 個の調波成分

$$\lambda_n[k,l] = w_n[l]^2 e^{-\frac{1}{2\sigma}^2(x_k - \Omega[l] - \log n)^2} \tag{43}$$

の和で表されていることを思い出せば，

$$C_n[k,l] \sim \mathrm{Pois}(\lambda_n[k,l]), \quad (n = 1, \cdots, N) \tag{44}$$

を仮定すれば，ポアソン分布の再生性より，矛盾することなく $Y[k,l] \sim \mathrm{Pois}(\lambda[k,l])$ が成り立つことが分かる。以上より，観測データ $Y := \{Y[k,l]\}_{1\leqq k\leqq K, 1\leqq l\leqq L}$ と潜在データ $C := \{C_n[k,l]\}_{1\leqq n\leqq N, 1\leqq k\leqq K, 1\leqq l\leqq L}$ を組にした完全データ $\{Y, C\}$ の確率密度関数が

$$p(Y, C|\boldsymbol{\Omega}, w) = p(Y|C)p(C|\boldsymbol{\Omega}, w) \tag{45}$$

$$p(Y|C) = \prod_{k,l} \delta\Big(Y[k,l] - \sum_n C_n[k,l]\Big) \tag{46}$$

$$p(C|\boldsymbol{\Omega}, w) = \prod_n \prod_{k,l} \frac{\lambda_n[k,l]^{C_n[k,l]} e^{-\lambda_n[k,l]}}{C_n[k,l]!} \tag{47}$$

で与えられる。完全データの対数確率の，観測データとモデルパラメータが与えられた下での条件つき期待値を Q 関数と呼び，Q 関数をもとに EM アルゴリズムの各ステップの更新則が導かれる。以下で本問題における Q 関数を導く。$\log p(Y, C|\boldsymbol{\Omega}, w)$ の中で $\boldsymbol{\Omega}, w$ に関係する項に対し，$C_n[k,l]$ の期待値をとったものが Q 関数となるので，Q 関数は

$$Q(\boldsymbol{\Omega}, w) = \sum_n \sum_{k,l} \big(\langle C_n[k,l]\rangle_{p(C_n[k,l]|Y[k,l],\boldsymbol{\Omega}',w')} \log\lambda_n[k,l] - \lambda_n[k,l]\big) \tag{48}$$

で与えられる。ただし $\langle f(x)\rangle_{p(x|y)}$ は条件つき期待値 $\mathbb{E}_x[f(x)|y]$ を表す。ここで，ポアソン分布の性質より，

$$C_n[k,l]|Y[k,l] \sim \mathrm{Binom}\Big(Y[k,l], \frac{\lambda'_n[k,l]}{\lambda'[k,l]}\Big) \tag{49}$$

が言えるため，期待値は具体的に

$$\langle C_n[k,l]\rangle_{p(C_n[k,l]|Y[k,l],\boldsymbol{\Omega}',w')} = Y[k,l]\frac{\lambda'_n[k,l]}{\lambda'[k,l]} \tag{50}$$

と求まる。ただし，Binom は二項分布を表し，$\lambda'_n[k,l]$, $\lambda'[k,l]$ は $\boldsymbol{\Omega}', w'$ を代入した $\lambda_n[k,l]$, $\lambda[k,l]$ を表す。以上より Q 関数の具体形を得た。あとは，Q 関数に対数事前分布を加えた

$$I(\boldsymbol{\Omega}, w, \pi_\Omega, \pi_w) := \sum_{n,k,l} \big(Y[k,l]\frac{\lambda'_n[k,l]}{\lambda'[k,l]}\log\lambda_n[k,l] - \lambda_n[k,l]\big) + \log p(\boldsymbol{\Omega}|\pi_\Omega) + \log p(w|\boldsymbol{\Omega}, \pi_w) \tag{51}$$

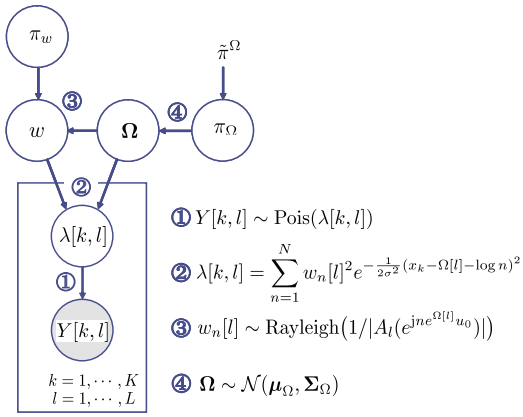

図 2　Graphical model representation of the proposed model.

を最大化する $\boldsymbol{\Omega}, w, \pi_\Omega, \pi_w$ の更新則を導出できれば，以下のとおり EM アルゴリズムが実現できる。

(E ステップ)　更新された $\boldsymbol{\Omega}, w$ を $\boldsymbol{\Omega}', w'$ に代入する。
(M ステップ)　$I(\boldsymbol{\Omega}, w, \pi_\Omega, \pi_w)$ を最大化 (または増加) するように $\boldsymbol{\Omega}, w, \pi_\Omega, \pi_w$ を更新し，E ステップに戻る。

M ステップの更新則を導出するにあたり，以下を仮定する。
(B1)　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。

$$\int_{-\infty}^{\infty} e^{-x^2/2\sigma^2}\mathrm{d}x \simeq x_0 \sum_k e^{-x_k^2/2\sigma^2} \tag{52}$$

よって，$\sum_k \lambda_n[k,l] \simeq \sqrt{2\pi}\sigma w_n[l]^2/x_0$ である。
(B2)　事前分布の信頼度を調節できるようにする目的で，式 (41) 右辺を 2γ 乗したものを $p(w|\boldsymbol{\Omega}, \Theta_w)$ とする。
(B3)　$p(w|\boldsymbol{\Omega}, \pi_w)$ において $\boldsymbol{\Omega}$ を 1 ステップ前の更新値 $\boldsymbol{\Omega}'$ に置き換える。すなわち，$p(w|\boldsymbol{\Omega}, \pi_w) \simeq p(w|\boldsymbol{\Omega}', \pi_w)$ と近似する。

以上の準備により，M ステップの更新式が導ける。

1) $\boldsymbol{F}_0$ パターン $\boldsymbol{\Omega}$: $I(\boldsymbol{\Omega}, w, \pi_\Omega, \pi_w)$ で $\boldsymbol{\Omega}$ に関係する項は

$$I(\boldsymbol{\Omega}, w, \pi_\Omega, \pi_w) \overset{c}{=} -\frac{1}{2}(\boldsymbol{\Omega} - \hat{\boldsymbol{\Omega}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega} - \hat{\boldsymbol{\Omega}}) - \frac{1}{2}(\boldsymbol{\Omega} - \boldsymbol{\mu}_\Omega)^{\mathrm{T}}\boldsymbol{\Sigma}_\Omega^{-1}(\boldsymbol{\Omega} - \boldsymbol{\mu}_\Omega) \tag{53}$$

であるので，I を最大化する $\boldsymbol{\Omega}$ は

$$\boldsymbol{\Omega} = (\boldsymbol{\Sigma} + \boldsymbol{\Sigma}_\Omega)^{-1}(\boldsymbol{\Sigma}\hat{\boldsymbol{\Omega}} + \boldsymbol{\Sigma}_\Omega\boldsymbol{\mu}_\Omega) \tag{54}$$

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

$$\hat{\Omega}_l = \sum_k \frac{Y[k,l]}{\lambda'[k,l]} \sum_n \lambda'_n[k,l](x_k - \log n) \Big/ \sum_k Y[k,l] \tag{55}$$

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

$$\Sigma_l = \sigma^2 \Big/ \sum_k Y[k,l] \tag{56}$$

の対角行列である。

2) n 次調波成分の瞬時パワー $w_n[l]$: $\partial I/\partial w_n[l] = 0$ を解くと，I を最大化する $w_n[l]$ は

$$w_n[l]^2 = \frac{\sum_k Y[k,l]\frac{\lambda'_n[k,l]}{\lambda'[k,l]} + \gamma}{\sqrt{2\pi}\sigma/x_0 + \gamma/P_n[l]} \tag{57}$$

で与えられる。この更新式は $\gamma \to \infty$, $\gamma \to 0$ の極限をとると

$$\lim_{\gamma\to\infty} \frac{\sum_k Y[k,l]\frac{\lambda'_n[k,l]}{\lambda'[k,l]} + \gamma}{\sqrt{2\pi}\sigma/x_0 + \gamma/P_n[l]} = P_n[l]$$

$$\lim_{\gamma\to 0} \frac{\sum_k Y[k,l]\frac{\lambda'_n[k,l]}{\lambda'[k,l]} + \gamma}{\sqrt{2\pi}\sigma/x_0 + \gamma/P_n[l]} = \frac{x_0}{\sqrt{2\pi}\sigma}\sum_k Y[k,l]\frac{\lambda'_n[k,l]}{\lambda'[k,l]}$$

となる。上段の式は直前に更新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，下段の式はスペクトルモデルが観測スペクトルに最も良くフィットする $w_n[l]^2$ の値をそれぞれ表しており，式 (57) は両者の間を補間する値に相当する。γ はその補間係数と見ることができる。

3) 藤崎モデルパラメータ π_Ω: I において π_Ω に関係する項は

$$I(\boldsymbol{\Omega}, w, \pi_\Omega, \pi_w) \stackrel{c}{=} -\frac{1}{2}\log|\boldsymbol{\Sigma}_\Omega| - \frac{1}{2}(\boldsymbol{\Omega}-\boldsymbol{\mu}_\Omega)^{\mathrm{T}}\boldsymbol{\Sigma}_\Omega^{-1}(\boldsymbol{\Omega}-\boldsymbol{\mu}_\Omega) \quad (58)$$

であり，[4] の目的関数と同形となるので，π_Ω の更新方法には当該文献と同様の反復アルゴリズムが適用できる。詳細は [4] を参照されたい。

4) 全極モデルパラメータ π_w: I において π_w に関係する項は，

$$I(\boldsymbol{\Omega}, w, \pi_\Omega, \pi_w) \stackrel{c}{=} \gamma\sum_l\sum_n\left(\log\frac{w_n[l]^2}{P_n[l]^2} - \frac{w_n[l]^2}{P_n[l]}\right) \quad (59)$$

であり，[3] で設定されている目的関数と類似した形となるので，π_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。詳細は省略するが，$\boldsymbol{a}_l := (a_l[0], a_l[1], \cdots, a_l[M])^{\mathrm{T}}$ とすると，

$$\boldsymbol{a}_l \leftarrow \boldsymbol{R}_l^{-1}\boldsymbol{h}_l \quad (60)$$

$$\boldsymbol{h}_l \leftarrow \hat{\boldsymbol{R}}_l\boldsymbol{a}_l \quad (61)$$

を繰り返すことで R 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

$$R_l(m'-m) = \frac{1}{N}\sum_{n=1}^{N} w_n[l]^2\cos ne^{\Omega[l]}u_0(m'-m) \quad (62)$$

$$\hat{R}_l(m'-m) = \frac{2}{N}\sum_{n=1}^{N} P_n[l]\cos ne^{\Omega[l]}u_0(m'-m) \quad (63)$$

を m' 行 m 列成分とする Toeplitz 行列である。

6. 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 3 にその動作実験例を示す。図 3 上段に上記の音声信号に対して $x_0 = 12$cent, $\sigma = 0.02$, $t_0 = t_l - t_{l-1} = 8$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 5. 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。F_0 パラメータの推定結果を見ると，基本周波数推定においてしばしば問題となる倍ピッチ・半ピッチ誤りがほとんどないことが分かる。これはスペクトル包絡の滑らかさと，F_0 パターンの滑らかさの制約が効果的に働いていることを示している。また，図 4 に示すように藤崎モデルパラメータがスペクトルから直接得られる点も提案法の大きな特徴の一つである。

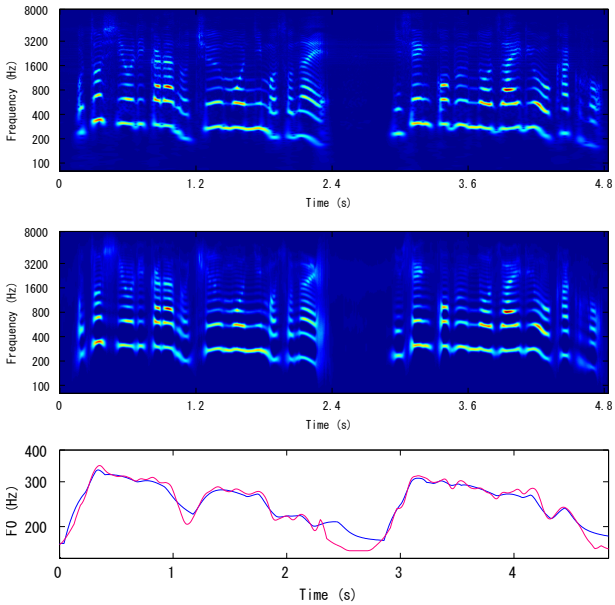

図 3　観測スペクトログラム $Y[k,l]$ (上段) と $\lambda[k,l]$ の推定値 (中段) と $\boldsymbol{\Omega}$ の推定値 (下段・赤線) および $\boldsymbol{\mu}_\Omega$ の推定値 (下段・青線)

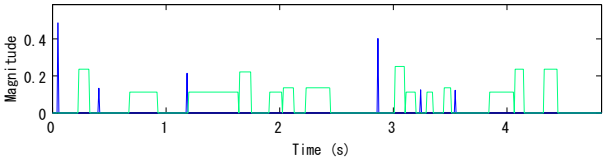

図 4　フレーズ指令 (青線) とアクセント指令 (緑線) の推定結果

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 π_w についても同様であった。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 π_Ω と π_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。筆者は既に，変分ベイズ法に基づき $\boldsymbol{\Omega}$ と w の事後分布を近似計算するアルゴリズムを導出しており，これをもとに，今後上記の課題解決に向けて検討を行っていく予定である。

7. 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 EM アルゴリズム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直接スペクトルから同時推定することが可能となった。

文　　献

[1] F. Itakura, S. Saito, “Analysis synthesis telephony based upon the maximum likelihood method,” In *Proc. 6th Int' l Cong. Acoust. (ICA ' 68)*, C-5-5, C17–20, 1968.

[2] H. Fujisaki, “A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour,” In *Vocal Physiology: Voice Production, Mechanisms and Functions*, (O. Fujimura, ed.) Raven Press, pp. 347–355, 1988.

[3] A. El-Jaroudi, J. Makhoul, “Discrete all-pole modeling,” *IEEE trans. Signal Process.*, **39**(2), pp. 411–423, 1991.

[4] 亀岡, ルルー, 大石, “音声 F_0 パターン生成過程の確率モデル,” 音講論 (秋), 1-1-3, pp.207-210, Sep. 2010.