Vol.2013-MUS-99 No.8
2013/5/11

スペクトル包絡と基本周波数の同時推定のための無限カーネル線形予測分析法

吉井 和佳[1,a)]　後藤 真孝[1,b)]

概要：本稿では，音声信号のスペクトル包絡と基本周波数とを同時に推定するための新しい線形予測分析法について述べる．従来，ソース・フィルタ理論に基づく線形予測分析法では，所与の観測信号はガウス性白色雑音を入力信号とする自己回帰系からの出力信号であると仮定して，全極型フィルタの係数を推定することが行われていた．しかし，声帯振動に起因する周期パルス（周波数領域では調波構造）が入力信号である場合には，推定されたスペクトル包絡（全極型フィルタの周波数応答）は，調波構造の倍音周波数で不必要に大きなピークをもつ問題があった．この問題を解決するため，入力信号と出力信号との関係をノンパラメトリックベイズガウス過程回帰モデルで表現する無限カーネル線形予測分析法を提案する．本手法では，異なる基本周波数に対応する可算無限個のカーネルを考え，それらの凸結合で入力信号の周期性を表現する．ここで，無限個の非負の重みに対してガンマ過程事前分布を導入すると，スパースなマルチカーネル学習を行うことができる．すなわち，全極型フィルタの係数を推定すると同時に，基本周波数に対応する優勢なカーネルを同定できる．本手法を話声・歌声信号に対して適用し，基本周波数をもつ有声区間を同定しながら，基本周波数の影響を考慮したスペクトル包絡を推定できることを確かめた．

1. はじめに

音響信号のスペクトル包絡推定は，話声・歌声信号分析の基礎をなす重要な技術である．音声分野におけるこれまでの研究により，人間の発声機構はソース・フィルタ理論でよく説明できることが知られている（図 **1**）．具体的には，声帯振動に起因する音源信号が声道を通ることでその音響特性が変化する過程を考える．時間領域では，音源信号に声道フィルタのインパルス応答が畳みこまれた出力信号が観測信号であると解釈できる．一方，周波数領域では，観測スペクトルの微細構造と包絡構造がそれぞれ，音源信号とフィルタの周波数特性に対応していると仮定することが一般的である．本研究ではこのような仮定のもと，観測信号が与えられたときに，フィルタの周波数応答，すなわちスペクトル包絡を推定する問題に取り組む．

線形予測分析 (Linear Prediction: LP) [1] は，スペクトル包絡推定のための数学的に確立された方法のひとつである．線形予測分析では，観測信号は自己回帰過程に従う，すなわち，フィルタは全極型伝達関数で記述できると仮定する．声道は単純な音響管の連結であるとみなすと鼻子音以外には反共振は存在せず，人間の聴覚はスペクトルのピーク（フォルマント）に敏感であることから，この仮定は妥当であると考えられている．古典的な線形予測分析では，音源信号がガウス性白色雑音であれば，全極型フィルタの係数を最尤推定の枠組みで精度良く求めることができる．しかし，観測信号が明確な音高をもっている，すなわち，音源信号が周期的であると，推定されたフィルタの周波数応答（スペクトル包絡）は調波構造のピークの位置で不必要に鋭いピークをもつようなバイアスがかかる．

この問題を解決するため，これまで多くの研究が行われてきた．El-Jaroudi ら [2] は，全極型フィルタの周波数応答を調波構造の離散的なピークに対してフィットさせる手法 (Discrete All-pole Filtering: DAP) を提案している．この手法は，Badeau ら [3] によって自己回帰・移動平均モデルにも適用可能なように拡張されている．一方，Oudot ら [4] は，全極型フィルタがなめらかな周波数応答をもつような制約をかける手法を提案している．Villavicencio ら [5] は，ケプストラム平滑化を反復的に適用する手法を提案している．線形予測分析以外のスペクトル包絡推定法として，河原ら [6] は，音声スペクトルを周期・非周期成分とスペクトル包絡に精度よく分離できる分析合成系 STRAIGHT を開発している．中野ら [7] は基本周波数の影響を回避するため，隣接するフレームにわたってスペクトル包絡を平均化する手法を提案している．これらの手法はスペクトル包絡を精度良く求めることができるが，基本周波数の値が既知であることが前提であった．

近年，基本周波数とスペクトル包絡をモデル化するうえで，確率モデルに基づくアプローチが有望視されている．佐宗ら [8] は自己回帰隠れマルコフモデル (AR-HMM) と

[1] 産業技術総合研究所
Umezono 1-1-1, Tsukuba, Ibaraki 305–8568, Japan
[a)] k.yoshii(at)aist.go.jp
[b)] m.goto(at)aist.go.jp

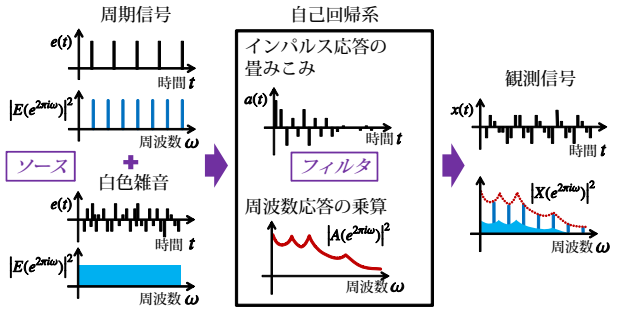

図 1　発声機構に対するソース・フィルタモデル

呼ばれる，状態遷移が循環するように拘束をかけた HMM を用いて音源信号の周期性を表現する手法を提案している．戸田ら [9] は，あらかじめ基本周波数の値を与える必要があるものの，音源信号に起因する調波構造の時間的なダイナミクスをトラジェクトリ HMM を用いて表現する手法を提案している．亀岡ら [10,11] は，基本周波数とスペクトル包絡を同時に推定するための先駆的な研究を行っている．例えば，異なる基本周波数に対応する多数の周期カーネルを内包するガウス過程に基づくマルチカーネル線形回帰分析法 (Multiple Kernel Linear Prediction: MKLP) を提案している [10]．優勢なカーネルを決定（基本周波数を推定）するには，スペクトル包絡を推定すると同時にカーネルの重みに対する事後分布最大化推定が行われる．我々は，この研究をさらに発展させることを試みる．

本稿では，近年着目されているノンパラメトリックベイズ理論とカーネル法の強みを同時に取り入れた無限カーネル線形予測分析法 (Infinite Kernel Linear Prediction: IKLP) を提案する．我々は，MKLP におけるカーネルの個数が無限に発散した極限を考え，それら無限個のカーネルの重みに対してガンマ過程事前分布を仮定する．変分ベイズ法を用いて事後分布推定を行うことで，ほとんど全てのカーネルの重みがゼロにほぼ等しくなり，基本周波数に対応するカーネルの重みのみが有意な値をもつようなスパースな学習をスペクトル包絡推定と同時に行うことができる．本研究は，線形予測分析法の背後にある観測信号の生成過程に対する仮定を明らかにすることで信号処理分野に貢献するだけではなく，マルチカーネル学習 [12] に対する効率的な収束保証付きの最適化アルゴリズムを提案することで機械学習分野にも貢献することができると考える．

2. 線形予測分析

本章では，音声・歌声などの観測信号からスペクトル包絡（全極型フィルタの係数）を推定するための確率モデルについて概観し，その最新の手法として亀岡らの手法 [10] を紹介する．まず，音源信号がガウス性白色雑音に従うと仮定する古典的な自己回帰モデルについて説明する．次に，音源信号が周期性をもつ場合でも精度のよい推定が可能なカーネル化されたモデルについて紹介する．

2.1 確率モデルの定式化

まず，観測信号が自己回帰過程に従うと仮定する基本的な定式化について説明する．あるフレームに含まれる M 個の連続するサンプルを $\boldsymbol{x} = (x_1, x_2, \cdots, x_M)^T$ とする．いま，$\boldsymbol{x}$ が P 次の自己回帰過程に従うと仮定すると，

$$x_m = \sum_{p=1}^{P} a_p x_{m-p} + \epsilon_m \tag{1}$$

と書くことができる．ここで，$\boldsymbol{a} = (a_1, \cdots, a_P)^T$ は全極型フィルタの P 個の係数であり，線形予測係数と呼ばれる．$\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_M)^T$ は誤差項である．ソース・フィルタ理論の観点からは，$\boldsymbol{\epsilon}$ が声帯振動に起因する音源信号（ソース）に対応し，$\boldsymbol{a}$ が声道（フィルタ）の反響特性を表していると解釈できる．この自己回帰モデルは，ϵ_m を入力として x_m を出力する線形系とみなすことができ，その全極型伝達関数は $A(z) = 1/(1 - a_1 z^{-1} \cdots - a_P z^{-P})$ で与えられる．すなわち，$\boldsymbol{x}$ および $\boldsymbol{\epsilon}$ の Z 変換を $X(z)$ および $E(z)$ とすると，$X(z) = E(z)A(z)$ が成り立つ．全極型フィルタの周波数応答（スペクトル包絡）は，m を周波数ビンのインデックスすると考えると $|A(e^{2\pi i m/M})|^2$ で与えられる（図 1）．

我々の目的は，観測信号 $\boldsymbol{x}$ が与えられたときに，確率的な枠組みのもとでフィルタ係数 $\boldsymbol{a}$ を求めることである．この問題はフィルタ係数 $\boldsymbol{a}$ および音源信号 $\boldsymbol{\epsilon}$ が両方とも未知のもとでは不良設定問題であるため，音源信号 $\boldsymbol{\epsilon}$ の性質に関するなんらかの仮定が必要になる．一般的には，$\boldsymbol{\epsilon}$ はガウス性白色雑音であると仮定する．

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \nu \boldsymbol{I}) \tag{2}$$

ここで，ν はノイズの分散であり，$\boldsymbol{I}$ は単位行列である．すなわち，$\boldsymbol{\epsilon}$ の M 個の成分は $\mathcal{N}(0, \nu)$ に従って独立同分布することを示している．ここで，行列 $\boldsymbol{\Psi} \in \mathbb{R}^{M \times M}$ および行列 $\boldsymbol{X} \in \mathbb{R}^{M \times P}$ を

$$\boldsymbol{\Psi} = \begin{bmatrix} 1 & & & & \\ -a_1 & \ddots & & 0 & \\ \vdots & \ddots & \ddots & & \\ -a_P & & & & \\ & \ddots & & \ddots & \ddots \\ 0 & & -a_P & \cdots & -a_1 \ 1 \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 0 & \cdots & 0 \\ x_1 & \ddots & \vdots \\ \vdots & \ddots & 0 \\ \vdots & & x_1 \\ \vdots & & \vdots \\ x_{M-1} & \cdots & x_{M-P} \end{bmatrix} \tag{3}$$

とすると，式 (1) は $\boldsymbol{\epsilon} = \boldsymbol{\Psi}\boldsymbol{x}$ と簡潔に書けて，

$$\boldsymbol{x} = \boldsymbol{\Psi}^{-1}\boldsymbol{\epsilon} \tag{4}$$

を得る．式 (2) および (4) を用いると，観測信号 $\boldsymbol{x}$ の尤度は

$$\boldsymbol{x} \sim \mathcal{N}(0, \nu \boldsymbol{\Psi}^{-1}\boldsymbol{\Psi}^{-T}) \tag{5}$$

で与えられる．式 (5) が古典的な線形予測分析の確率モデルである．この確率モデルに対して最尤推定を行う場合には，最適なフィルタ係数 $\boldsymbol{a}$ は正規方程式 $\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{a} = \boldsymbol{X}^T\boldsymbol{x}$ の解として求めることができる．

音源信号 $\boldsymbol{\epsilon}$ が式 (2) で与えられる等方的なガウス分布に従う場合には，このスペクトル包絡推定法は有効に機能する．しかし，観測信号が音声信号や歌声信号である場合には，声帯の周期振動に起因する明確な周期性をもっているため，なんらかの対策が必要になる．

2.2 ガウス過程に基づくカーネル化

亀岡ら [10] は，式 (5) で与えられる確率モデルをガウス過程回帰の観点でカーネル拡張する方法を提案している．これまで，音源信号 $\boldsymbol{\epsilon}$ はガウス性白色雑音であると仮定してきた．一方，ここでは，音源信号 $\epsilon(t)$ は時間 t 上に定義された連続関数であると考え，時刻 t から関数 $\epsilon(t)$ を線形回帰する問題について考える．いま，$\{\epsilon_m\}_{m=1}^M$ は，時刻 $\{t_m\}_{m=1}^M$ における関数 $\epsilon(t)$ の出力値であるとみなす．我々の目的は，連続関数 $\epsilon(t)$ を J 個の基底関数 $\{\phi_j(t)\}_{j=1}^J$ の和で表現することである．

$$\epsilon(t) = \sum_{j=1}^J w_j \phi_j(t) + \eta(t) = \boldsymbol{\phi}(t)^T \boldsymbol{w} + \eta(t) \tag{6}$$

ここで，$\eta(t)$ は誤差関数，$\boldsymbol{\omega} \in \mathbb{R}^J$ は基底関数の重みであり，$\boldsymbol{\phi}(t) = (\phi_1(t), \cdots, \phi_J(t))^T$ とした．時刻 $\{t_m\}_{m=1}^M$ における関数 $\eta(t)$ の出力値を並べたものを $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_M)^T$ とし，計画行列を $\boldsymbol{\Phi} = (\boldsymbol{\phi}(t_1), \cdots, \boldsymbol{\phi}(t_M))^T \in \mathbb{R}^{M \times J}$ とすると，式 (6) の時刻 $\{t_m\}_{m=1}^M$ における回帰モデルは

$$\boldsymbol{\epsilon} = \boldsymbol{\Phi}\boldsymbol{w} + \boldsymbol{\eta} \tag{7}$$

で与えられる．いま，重み $\boldsymbol{\omega}$ および誤差信号 $\boldsymbol{\eta}$ が等方的なガウス分布に従うと仮定すると

$$\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \nu_w \boldsymbol{I}), \quad \boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \nu_e \boldsymbol{I}) \tag{8}$$

と書ける．ここで，ν_w および ν_e はガウス分布の分散を表す．式 (7) および (8) を用いると，最終的に

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \nu_w \boldsymbol{\Phi}\boldsymbol{\Phi}^T + \nu_e \boldsymbol{I}) \tag{9}$$

を得る．任意の時刻 $\{t_m\}_{m=1}^M$ における関数 $\epsilon(t)$ の周辺分布がガウス分布であることから，関数 $\epsilon(t)$ はガウス過程 [13] に従うことが分かる．その振る舞いはグラム行列（カーネル）$\boldsymbol{K} = \boldsymbol{\Phi}\boldsymbol{\Phi}^T$ によって規定される．カーネル $\boldsymbol{K}$ の各要素は基底関数の内積として定義されている．

$$K_{m,m'} = \boldsymbol{\phi}(t_m)^T \boldsymbol{\phi}(t_{m'}) \tag{10}$$

一方，任意の正定値行列はグラム行列として有効であることが知られており，基底関数を明示せずに $K_{m,m'} = k(t_m, t_{m'})$ として直接 $\boldsymbol{K}$ を計算することもできる（カーネルトリック）．ここで，$k(t, t')$ はカーネル関数と呼ばれる．このとき，式 (4) および (9) を用いると，観測信号 $\boldsymbol{x}$ の尤度は

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Psi}^{-1}(\nu_w \boldsymbol{K} + \nu_e \boldsymbol{I})\boldsymbol{\Psi}^{-T}) \tag{11}$$

で与えられる．これはガウス過程回帰モデル [10] であり，式 (5) をその特別な場合として含んでいる．実際，$\boldsymbol{\Phi}\boldsymbol{\Phi}^T = \boldsymbol{I}$ であれば，すなわち，J 個の基底関数が互いに独立であれば，$\nu = \nu_w + \nu_e$ とすると式 (5) が得られる．

2.3 マルチカーネル学習

次に，音源信号 $\boldsymbol{\epsilon}$ の性質を反映するようなグラム行列 $\boldsymbol{K}$ の設計法について述べる．本稿では，観測信号 $\boldsymbol{x}$（あるいは音源信号 $\boldsymbol{\epsilon}$）が基本周波数をもつ周期信号である場合を考えているので，$\boldsymbol{K}$ として周期カーネルを利用するのが自然である．例えば，$k(t, t') = \exp(-2\sin^2(\pi \frac{t-t'}{T})/l^2)$ は良く知られた周期カーネルで，その周期は T である．このとき，全ての基底関数が周期 T（基本周波数は $1/T$）の周期関数であることが暗黙的に仮定されている．

亀岡ら [10] は，各基底関数を H 個の等しいパワーを持つ正弦波の足し合わせで表現する方法を提案している．

$$\phi_j(t) = \sum_{h=1}^H \sin\left(2\pi h \frac{t-c_j}{T}\right) \quad (c_j \text{は位相}) \tag{12}$$

このとき，グラム行列 $\boldsymbol{K}$ は式 (10) から求められる．しかし，実際には音源信号 $\boldsymbol{\epsilon}$ の周期 T は未知であり，観測信号 $\boldsymbol{x}$ が与えられたときに，周期 T を推定する必要がある．

この問題に対する強力な解法のひとつに，マルチカーネル学習 [12] が知られている．具体的には，グラム行列 $\boldsymbol{K}$ を I 個の異なるグラム行列の重みつけ和として表現する．

$$\boldsymbol{K} = \sum_{i=1}^I \theta_i \boldsymbol{K}_i \tag{13}$$

ここで，$\boldsymbol{\theta} = \{\theta_i\}_{i=1}^I$ はカーネルの重みであり，$\boldsymbol{K}_i$ は周期 T_i をもつ周期カーネルである．カーネルの重みは，観測信号 $\boldsymbol{x}$ において各周期カーネルがどの程度優勢であるか，すなわち，基本周波数がどのあたりに存在するかを示している．最終的に，観測信号 $\boldsymbol{x}$ の尤度は以下で与えられる．

$$\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Psi}^{-1}\left(\nu_w \textstyle\sum_{i=1}^I \theta_i \boldsymbol{K}_i + \nu_e \boldsymbol{I}\right)\boldsymbol{\Psi}^{-T}\right) \tag{14}$$

フィルタ係数 $\boldsymbol{a}$ およびカーネルの重み $\boldsymbol{\theta}$ は，EM アルゴリズムを用いて推定することができる．文献 [10] においては，異なる周期 $\{T_i\}_{i=1}^I$ をもつ数百個のカーネルを考え，カーネルの重み $\boldsymbol{\theta}$ に対する事前分布として一般化ガウス分布をおくことで，$\boldsymbol{\theta}$ を事後確率最大化 (MAP) 推定により求めている．しかし，MAP 推定の枠組みでは，$\boldsymbol{\theta}$ を完全にスパースに導くことは原理的にできない．

3. 無限カーネル線形予測分析

本章では，音声・歌声信号のスペクトル包絡と基本周波数を確率的な枠組みで同時に推定するため，ノンパラメトリックベイズモデルに基づく無限カーネル線形予測分析法 (IKLP) を提案する．まず，式 (14) で与えられるマルチカーネル線形予測分析法 (MKLP) の確率モデルにおいて，カーネルの個数を無限に発散させたときの極限（$I \to \infty$）について考える．次に，確率モデルのベイズ的な取り扱いを可能にするため，未知のパラメータに対して適切な事前分布を設計する．最後に，未知のパラメータの事後分布を計算するため，変分ベイズ法に基づく効率的で収束の保証された最適化アルゴリズムの導出を行う．

3.1 ノンパラメトリックベイズモデル

式 (14) において $I \to \infty$ とすると，観測信号 $\boldsymbol{x}$ の尤度は

$$\boldsymbol{x} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Psi}^{-1}\left(\nu_w \sum_{i=1}^{I\to\infty} \theta_i \boldsymbol{K}_i + \nu_e \boldsymbol{I}\right)\boldsymbol{\Psi}^{-T}\right) \quad (15)$$

で与えられる．まず，無限次元の非負ベクトル $\boldsymbol{\theta}$ に対する事前分布として，ガンマ過程を利用する．具体的には，

$$\theta_i \sim \mathrm{Gamma}(\alpha/I, \alpha) \quad (16)$$

として，打ち切りレベル I を無限大に発散させたときに，$\boldsymbol{\theta}$ は，集中度 α をもつガンマ過程から得られる無限次元の非負系列となる．このとき，任意の正整数 $\epsilon > 0$ に対して，$\theta_i > \epsilon$ となる要素の個数 I^+ はほとんど確実に有限であることが証明されており，無限次元の空間においてスパースな学習が可能である根拠となっている．現実的には，打ち切りレベル I を集中度 α より十分大きくしておけば，$\boldsymbol{\theta}$ 中のいくつかの要素のみがゼロより大きな有意な値を持つことが期待できる．

完全なベイズ的な取り扱いのため，ガウス分布の分散 ν_w および ν_e に対する事前分布に，ガンマ分布を利用する．

$$\nu_w \sim \mathrm{Gamma}(a_w, b_w), \quad \nu_e \sim \mathrm{Gamma}(a_e, b_e) \quad (17)$$

ここで，a_* および b_* はガンマ分布の形状パラメータおよびレートパラメータである．さらに，フィルタ係数 $\boldsymbol{a}$ に対する事前分布として，ガウス分布を利用する．

$$\boldsymbol{a} \sim \mathcal{N}(\mathbf{0}, \lambda \boldsymbol{I}) \quad (18)$$

ここで，λ は超パラメータである．

基本周波数を推定するには，重みの期待値 $\mathbb{E}[\theta_i]$ が最大となるカーネル $\boldsymbol{K}_i$ を同定すればよい．このとき，基本周波数は $1/T_i$ となる．スペクトル包絡は，フィルタ係数の期待値 $\mathbb{E}[\boldsymbol{a}]$ を用いて計算できる．このモデルの副次的効果として，MKLP [10] と同様に，分散の比率 $\mathbb{E}[\nu_w]/\mathbb{E}[\nu_w + \nu_e]$ から有声・無声の判定が可能である．

3.2 変分ベイズ法

我々の目的は，観測信号 $\boldsymbol{x}$ が与えられたときに，未知パラメータの同時的な事後分布 $p(\boldsymbol{\theta}, \boldsymbol{a}, \nu_w, \nu_e|\boldsymbol{x})$ をベイズの定理 $p(\boldsymbol{\theta}, \boldsymbol{a}, \nu_w, \nu_e|\boldsymbol{x}) = p(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{a}, \nu_w, \nu_e)/p(\boldsymbol{x})$ に従って計算することである．正規化項である周辺尤度 $p(\boldsymbol{x})$ を解析的に計算することは困難であるが，変分ベイズ法を用いれば真の事後分布の近似を効率よく求めることができる．具体的には，真の事後分布 $p(\boldsymbol{\theta}, \boldsymbol{a}, \nu_w, \nu_e|\boldsymbol{x})$ を，以下のように因子分解可能な変分事後分布

$$q(\boldsymbol{\theta}, \boldsymbol{a}, \nu_w, \nu_e) = q(\boldsymbol{a})q(\nu_w)q(\nu_e)\prod_i q(\theta_i) \quad (19)$$

で近似することを考える．これは，事後分布において各パラメータの独立性を仮定していることを意味しており，真の事後分布との間にはいくらかの乖離が存在する．

変分ベイズ法では，変分事後分布の真の事後分布に対するカルバック・ライブラー (KL) ダイバージェンスを単調減少させるように，各因子を反復的に最適化する．これは，対数周辺尤度 $\log p(\boldsymbol{x})$ の変分下限 $\mathcal{L}$ を単調増加させることと等価である．ここで，$\mathcal{L}$ は以下の通り与えられる．

$$\begin{aligned}\log p(\boldsymbol{x}) &\geq \mathbb{E}[\log p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{a}, \nu_w, \nu_e)] \quad (20)\\ &+ \mathbb{E}[\log p(\boldsymbol{\theta})] + \mathbb{E}[\log p(\boldsymbol{a})] + \mathbb{E}[\log p(\nu_w)] + \mathbb{E}[\log p(\nu_e)]\\ &- \mathbb{E}[\log q(\boldsymbol{\theta})] - \mathbb{E}[\log q(\boldsymbol{a})] - \mathbb{E}[\log q(\nu_w)] - \mathbb{E}[\log q(\nu_e)] \equiv \mathcal{L}\end{aligned}$$

しかし，右辺の第一項（対数尤度関数の期待値）は依然として解析的に計算ができないため，$\mathcal{L} \geq \mathcal{L}'$ となるようなさらなる変分下限 $\mathcal{L}'$ を構成し，$\mathcal{L}'$ を逐次最大化することを考える．このとき，各因子の更新則は

$$\begin{aligned}q(\boldsymbol{\theta}) &\propto p(\boldsymbol{\theta}) \exp(\mathbb{E}_{q(\boldsymbol{a},\nu_w,\nu_e)}[\log q(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{a}, \nu_w, \nu_e)])\\ q(\nu_w) &\propto p(\nu_w) \exp(\mathbb{E}_{q(\boldsymbol{\theta},\boldsymbol{a},\nu_e)}[\log q(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{a}, \nu_w, \nu_e)]) \quad (21)\\ q(\nu_e) &\propto p(\nu_e) \exp(\mathbb{E}_{q(\boldsymbol{\theta},\boldsymbol{a},\nu_w)}[\log q(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{a}, \nu_w, \nu_e)])\end{aligned}$$

で与えられる．ここで，$q(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{a}, \nu_w, \nu_e)$ は $p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{a}, \nu_w, \nu_e)$ の変分下限であり，式 (24) で与えられる．ただし，共役性の問題から $\boldsymbol{a}$ の完全なベイズ的な取り扱いは困難であるため，$q(\boldsymbol{a}) = \delta_{\boldsymbol{a}^*}(\boldsymbol{a})$ であると仮定する．ここで，$\delta_{\boldsymbol{a}^*}$ は，ある $\boldsymbol{a}^*$ において関数値は無限大となり，それ以外はゼロとなるようなディラックのデルタ関数である．

3.2.1 行列に関する不等式

解析的に計算可能な変分下限 $\mathcal{L}'$ を導出するためには，行列に関する 2 つの不等式を用いる必要がある．まず，行列や行列変数関数に関する重要な概念を整理しておく．

定義 1 (行列の半正定値性) ある実対称行列 $\boldsymbol{A}$ が半正定値性を満たすとは，任意の実ベクトル $\boldsymbol{z}$ に対して $\boldsymbol{z}^T\boldsymbol{A}\boldsymbol{z} \geq 0$ が成立する，あるいは $\boldsymbol{A}$ の全ての固有値がゼロ以上である，あるいは $\boldsymbol{A} = \boldsymbol{Z}^T\boldsymbol{Z}$ となるような実行列 $\boldsymbol{Z}$ が存在することを言う．これらの条件はすべて等価である．

定義 2 (関数の凸性・凹性) 行列変数スカラー値関数 $f(\cdot)$ が凸であるとは，任意の実数 $0 \leq \lambda \leq 1$ に対して $\lambda f(\boldsymbol{A}) + (1-\lambda) f(\boldsymbol{B}) \geq f(\lambda\boldsymbol{A} + (1-\lambda)\boldsymbol{B})$ が成り立つことを言う．関数 $f(\cdot)$ が凹であるとは，$\lambda f(\boldsymbol{A}) + (1-\lambda) f(\boldsymbol{B}) \leq f(\lambda\boldsymbol{A} + (1-\lambda)\boldsymbol{B})$ が成り立つことを言う．

次に，$\boldsymbol{V}$ を任意の半正定値行列，$\boldsymbol{z}$ を任意の実ベクトルであるとすると，以下の 2 つの補題が得られる．

補題 1 関数 $f(\boldsymbol{V}) = \log|\boldsymbol{V}|$ は凹関数である．

補題 2 関数 $g(\boldsymbol{V}) = \boldsymbol{z}^T\boldsymbol{V}^{-1}\boldsymbol{z}$ は凸関数である．

紙面の都合上これらの証明は省略するが，定義 1 および定義 2 に従えば，簡単に確認することができる．

いま，各補題に関して不等式を導くことを考える．まず，凹関数 $f(\boldsymbol{V})$ に関して，任意の半正定値行列 $\boldsymbol{\Omega}$ を展開点とした 1 次のテイラー展開を考えると以下を得る．

$$\log|\boldsymbol{V}| \leq \log|\boldsymbol{\Omega}| + \mathrm{tr}(\boldsymbol{\Omega}^{-1}\boldsymbol{V}) - M \quad (22)$$

ここで，M は行列 $\boldsymbol{V}$ のサイズであり，等号成立条件は $\boldsymbol{\Omega}=\boldsymbol{V}$ である．次に，凸関数 $g(\boldsymbol{V})$ に関して，澤田ら [14] によって提案された行列不等式を適用すると以下を得る．

$$\boldsymbol{z}^T\left(\sum_{i=1}^I \boldsymbol{V}_i\right)^{-1}\boldsymbol{z} \le \sum_{i=1}^I \boldsymbol{z}^T\boldsymbol{\Upsilon}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{\Upsilon}_i\boldsymbol{z} \qquad (23)$$

ここで，$\{\boldsymbol{V}_i\}_{i=1}^I$ は任意の半正定値行列の集合であり，$\{\boldsymbol{\Upsilon}_i\}_{i=1}^I$ は足し合わせると単位行列になるような任意の行列の集合（補助変数）である．等号成立条件は，$\boldsymbol{\Upsilon}_i = \boldsymbol{V}_i(\sum_{i'=1}^I \boldsymbol{V}_{i'})^{-1}$ である．この不等式を証明するには，ラグランジュの未定乗数法を用いて，右辺の最小値が左辺に等しくなることを確かめればよい．

3.2.2 変分下限と反復最適化

まず，解析的に計算可能な変分下限 $\mathcal{L}'$ の導出を行う．半正定値行列 $\boldsymbol{K}$ を $\boldsymbol{K}=\nu_w\sum_i\theta_i\boldsymbol{K}_i+\nu_e\boldsymbol{I}$ とし，式 (22) および (23) を用いると，$\mathbb{E}[\log p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{a},\nu_w,\nu_e)]$ (式 (20) で与えられる $\mathcal{L}$ の第一項) の変分下限は以下で与えられる．

$$\begin{aligned}\mathbb{E}[\log p(\boldsymbol{x}|\cdot)] &= -\frac{M}{2}\log(2\pi)-\frac{1}{2}\mathbb{E}[\log|\boldsymbol{K}|]-\frac{1}{2}\mathbb{E}[\boldsymbol{x}^T\boldsymbol{\Psi}^T\boldsymbol{K}^{-1}\boldsymbol{\Psi}\boldsymbol{x}]\\ &\ge -\frac{1}{2}\log|\boldsymbol{\Omega}|-\frac{1}{2}\textstyle\sum_i\mathbb{E}[\nu_w\theta_i]\,\mathrm{tr}(\boldsymbol{\Omega}^{-1}\boldsymbol{K}_i)-\frac{1}{2}\mathbb{E}[\nu_e]\,\mathrm{tr}(\boldsymbol{\Omega}^{-1})+\mathrm{const.}\\ &\quad -\frac{1}{2}\textstyle\sum_i\mathbb{E}\left[\frac{1}{\nu_w\theta_i}\right]\boldsymbol{x}^T\boldsymbol{\Psi}^T\boldsymbol{\Upsilon}_i^T\boldsymbol{K}_i^{-1}\boldsymbol{\Upsilon}_i\boldsymbol{\Psi}\boldsymbol{x}-\frac{1}{2}\mathbb{E}\left[\frac{1}{\nu_e}\right]\boldsymbol{x}^T\boldsymbol{\Psi}^T\boldsymbol{\Upsilon}_0^T\boldsymbol{\Upsilon}_0\boldsymbol{\Psi}\boldsymbol{x}\end{aligned} \qquad (24)$$

ここで，$\boldsymbol{\Omega}$ は任意の半正定値行列であり，$\boldsymbol{\Upsilon}=\{\boldsymbol{\Upsilon}_i\}_{i=0}^{I\to\infty}$ は足すと単位行列になるような補助変数である．右辺を最大化する，すなわち，等号が成立するときの条件は

$$\boldsymbol{\Omega}=\mathbb{E}[\nu_w]\textstyle\sum_i\mathbb{E}[\theta_i]\boldsymbol{K}_i+\mathbb{E}[\nu_e]\boldsymbol{I} \qquad (25)$$

$$\boldsymbol{\Upsilon}_i=\mathbb{E}\left[\frac{1}{\nu_w\theta_i}\right]^{-1}\boldsymbol{K}_i\boldsymbol{S}^{-1},\qquad \boldsymbol{\Upsilon}_0=\mathbb{E}\left[\frac{1}{\nu_e}\right]^{-1}\boldsymbol{S}^{-1} \qquad (26)$$

で与えられる．ここで，$\boldsymbol{S}=\sum_i\mathbb{E}\left[\frac{1}{\nu_w\theta_i}\right]^{-1}\boldsymbol{K}_i+\mathbb{E}[\frac{1}{\nu_e}]^{-1}\boldsymbol{I}$ とした．いま，式 (24) はパラメータ自身とその逆数に関する期待値計算を含んでいることに注意する．すなわち，十分統計量は x および $1/x$ である．一方，ガンマ分布の十分統計量は $\log(x)$ および x であるので，変分事後分布は一般化逆正規分布 (GIG) 分布で与えられる [15]．

$$q(\theta_i)=\mathrm{GIG}(\theta_i|\gamma_i,\rho_i,\tau_i) \qquad (27)$$

$$q(\nu_w)=\mathrm{GIG}(\nu_w|\gamma_w,\rho_w,\tau_w),\ q(\nu_e)=\mathrm{GIG}(\nu_e|\gamma_e,\rho_e,\tau_e)$$

ただし，$\mathrm{GIG}(x|\gamma,\rho,\tau)=\frac{(\rho/\tau)^{\gamma/2}}{2K_\gamma(\sqrt{\rho\tau})}x^{\gamma-1}e^{-(\rho x+\tau/x)/2}$ である．このとき，変分パラメータの更新則は以下となる．

$$\begin{aligned}\gamma_i&=\alpha/I,\ \rho_i=2\alpha+\mathbb{E}[\nu_w]\,\mathrm{tr}(\boldsymbol{\Omega}^{-1}\boldsymbol{K}_i)\\ \tau_i&=\mathbb{E}\left[\frac{1}{\nu_w}\right]\boldsymbol{x}^T\boldsymbol{\Psi}^T\boldsymbol{\Upsilon}_i^T\boldsymbol{K}_i^{-1}\boldsymbol{\Upsilon}_i\boldsymbol{\Psi}\boldsymbol{x}\\ \gamma_w&=a_w,\ \rho_w=2b_w+\textstyle\sum_i\mathbb{E}[\theta_i]\,\mathrm{tr}(\boldsymbol{\Omega}^{-1}\boldsymbol{K}_i)\\ \tau_w&=\textstyle\sum_i\mathbb{E}\left[\frac{1}{\theta_i}\right]\boldsymbol{x}^T\boldsymbol{\Psi}^T\boldsymbol{\Upsilon}_i^T\boldsymbol{K}_i^{-1}\boldsymbol{\Upsilon}_i\boldsymbol{\Psi}\boldsymbol{x}\\ \gamma_e&=a_e,\ \rho_e=2b_e+\mathrm{tr}(\boldsymbol{\Omega}^{-1}),\ \tau_e=\boldsymbol{x}^T\boldsymbol{\Psi}^T\boldsymbol{\Upsilon}_0^T\boldsymbol{\Upsilon}_0\boldsymbol{\Psi}\boldsymbol{x}\end{aligned} \qquad (28)$$

フィルタ係数 $\boldsymbol{a}$ の MAP 推定値は，$\mathcal{L}'$ の偏微分をゼロとおくことで求まる．具体的には，正則化付きの正規方程式 $(\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}+\lambda\boldsymbol{I})\boldsymbol{a}=\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}$ の解で与えられる．ただし，$\boldsymbol{\Sigma}^{-1}=\sum_i\mathbb{E}\left[\frac{1}{\nu_w\theta_i}\right]\boldsymbol{\Upsilon}_i^T\boldsymbol{K}_i^{-1}\boldsymbol{\Upsilon}_i+\mathbb{E}\left[\frac{1}{\nu_e}\right]\boldsymbol{\Upsilon}_0^T\boldsymbol{\Upsilon}_0$ とした．

4. 実験

本章では，無限カーネル線形予測分析法 (IKLP) の基本的な振る舞いを確認するために行った実験について述べる．

4.1 実験条件

実験には，16kHz でサンプリングされた 2 種類の音響信号を用いた．ひとつは，RWC 研究用音楽データベース：音楽ジャンル RWC-MDB-G-2001 [16] の楽曲 No.91 の男性の無伴奏歌唱のうち冒頭部分 6.62s である．基本周波数軌跡の正確なアノテーションを行い，STRAIGHT [6] を用いて分析・再合成を行った信号を観測信号として用いた．もうひとつは，ATR 音声データベース [17] に収録されている女性話者の 4s の音声信号 FSUSA101 である．これらの信号は基本周波数が 300Hz 以上の帯域に存在し，従来の線形予測分析法への悪影響が大きいと考えられる．

IKLP は各短時間フレームごとに独立して行う．フレーム長は 2048 点 ($M=2048$)・シフト長は 160 点とし，窓関数にはガウス窓を用いた．超パラメータは $\alpha=1.0$, $a_w=b_w=a_e=b_e=1.0$, $P=30$, $\lambda=0.1$ とした．また，100Hz から 400Hz まで 6 セント間隔で 400 個の基本周波数に対応するカーネル $\{\boldsymbol{K}_i\}_{i=1}^{I=400}$ を準備した．

4.2 実験結果

図 **2** および図 **3** に示した実験結果から，IKLP はスペクトル包絡と基本周波数を同時に推定することができるだけではなく，MKLP と同様に有声・無声区間の検出を行えることが確認できた．図 2 において，4.2s から 4.6s 付近に存在する有声区間の検出が不安定になっている原因は，調波構造の高次倍音成分が比較的弱く，音源信号の性質を周期カーネルよりも単位行列カーネル（白色雑音）を用いて表現する方が適切であると判断されたからである．6.0s 付近に存在するビブラートにおいても，基本周波数を精度よく推定することができた．図 2 および図 3 の 2 列目および 3 列目を比較すると分かる通り，推定されたスペクトル包絡は，調波構造の倍音ピークの影響をほとんど受けていない．

IKLP は理論的に妥当ではあるものの，基本周波数推定においてしばしば半ピッチ誤りを起こすことがあった．この理由は，式 (15) で与えられる尤度関数は，モデルの共分散 $\boldsymbol{\Psi}^{-1}\boldsymbol{K}\boldsymbol{\Psi}^{-T}$ が観測した共分散 $\boldsymbol{x}\boldsymbol{x}^T$ を過小評価する場合には大きなペナルティをかけるが，過大評価しても小さなペナルティしか与えないからである．このことは，式 (15) が板倉・齋藤 (IS) ダイバージェンス [1] と密接に関係していることを示唆する．実際，$\boldsymbol{K}$ が単位行列であれば，IKLP は IS ダイバージェンス基準の自己回帰モデルの最適化問題に帰着する．IS ダイバージェンスは KL ダイバージェンスのような凸性を持たないため，初期値依存性が高く，局所解に陥りやすいと考えられている [18]．この問題を解決

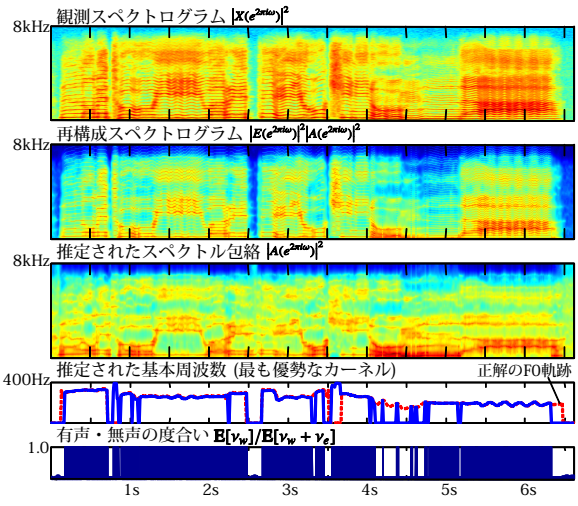

図 2　男性歌唱に対する推定結果

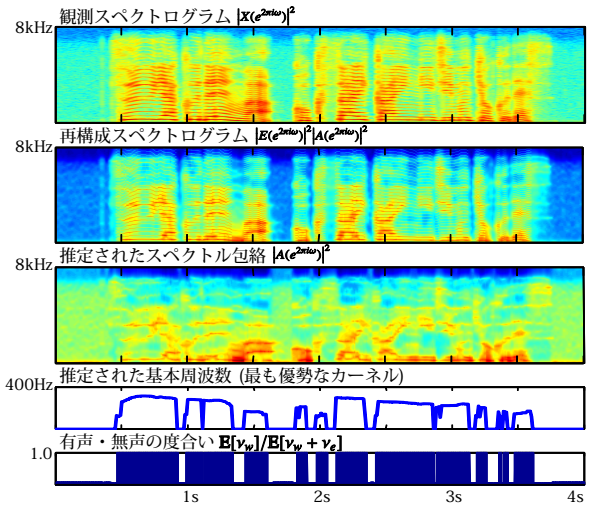

図 3　女性話声に対する推定結果

するには，基本周波数の時間的なダイナミクスを考慮したり，基本周波数の初期化を工夫する必要がある．

5. おわりに

本稿では，マルチカーネル学習の枠組みで，スペクトル包絡と基本周波数を同時に推定できるノンパラメトリックベイズモデルについて述べた．実験の結果，調波構造を考慮したなめらかなスペクトル包絡が推定できることが確認できた．従来法との詳細な比較実験は課題である．

今後の展開には，いくつかの興味深い方向性が考えられる．まず，推定精度を向上させるためには，音源信号 $\boldsymbol{\epsilon}$ に対する精緻なモデル [19,20] に基づいてカーネル関数 $\boldsymbol{\Phi}$ を設計し，カーネルの周期パラメータ $\boldsymbol{T}$ 自体を経験ベイズ法の枠組みで最尤推定する方向性が考えられる．また，提案手法は IS ダイバージェンス基準の線形予測分析 [1] の自然な拡張になっていることから，同様のマルチカーネル学習の枠組みで，IS ダイバージェンス基準の非負値行列因子分解 (NMF) [15,21] の本質的な拡張が可能になるはずである．我々はすでに，従来の NMF における基底ベクトルの要素間の相関構造を考慮した新しい音響信号分解法について研究を進めており，優れた音源分離結果を得ている．詳細については稿を改めて報告したい．

謝辞: 本研究の一部は，JSPS 科研費 23700184 および JST OngaCREST プロジェクトの支援を受けた．

参考文献

[1] F. Itakura and S. Saito. Analysis synthesis telephony based on the maximum likelihood method. *ICA*, 1968.
[2] A. El-Jaroudi and J. Makhoul. Discrete all-pole modeling. *IEEE Trans. on SP*, 39(2):411–423, 1991.
[3] R. Badeau and B. David. Weighted maximum likelihood autoregressive and moving average spectrum modeling. *ICASSP*, pp.3761–3764, 2008.
[4] M. Oudot *et al.* Estimation of the spectral envelope of voiced sounds using a penalized likelihood approach. *IEEE Trans. on SAP*, 9(5):469–481, 2001.
[5] F. Villavicencio *et al.* Improving LPC spectral envelope extraction of voiced speech by true-envelope estimation. *ICASSP*, pp.869–872, 2006.
[6] H. Kawahara *et al.* Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication*, 27(3–4):187–207, 1999.
[7] T. Nakano and M. Goto. A spectral envelope estimation method based on F0. *SAPA-SCALE*, pp.11–16, 2012.
[8] A. Sasou and K. Tanaka. Robust LP analysis using glottal source HMM with application to high-pitched and noise corrupted speech. *Eurospeech*, 2001.
[9] T. Toda and K. Tokuda. Statistical approach to vocal tract transfer function estimation based on factor analyzed trajectory HMM. *ICASSP*, pp.3925–3928, 2008.
[10] 亀岡弘和 *et al.* マルチカーネル線形予測モデルによる音声分析. 日本音響学会春季研究発表会, 2010.
[11] H. Kameoka *et al.* Speech spectrum modeling for joint estimation of spectral envelope and fundamental frequency. *IEEE Trans. on ASLP*, 18(6):1507–1516, 2010.
[12] G. Lanckriet *et al.* Learning the kernel matrix with semidefinite programming. *JMLR*, 5:27–72, 2004.
[13] C. E. Rasmussen and C. K. I. Williams, editors. *Gaussian Processes for Machine Learning*. MIT Press, 2006.
[14] H. Sawada *et al.* Efficient algorithms for multichannel extensions of Itakura-Saito nonnegative matrix factorization. *ICASSP*, pp.261–264, 2012.
[15] M. Hoffman *et al.* Bayesian nonparametric matrix factorization for recorded music. *ICML*, pp.439–446, 2010.
[16] M. Goto *et al.* RWC music database: Popular, classical, and jazz music database. *ISMIR*, pp.287–288, 2002.
[17] A. Kuramatsu *et al.* ATR Japanese speech database as a tool of speech recognition and synthesis. *Speech Communication*, 9(4):357–363, 1990.
[18] N. Bertin *et al.* A tempering approach for Itakura-Saito non-negative matrix factorization. with application to music transcription. *ICASSP*, pp.1545–1548, 2009.
[19] D. H. Klatt and L. C. Klatt. Analysis, synthesis and perception of voice quality variations among female and male talkers. *JASA*, 87(2):820–857, 1990.
[20] G. Fant *et al.* A four-parameter model of glottal flow. *STL-QPSR*, 26(4):1–13, 1985.
[21] C. Févotte *et al.* Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. *Neural Computation*, 21(3):793–830, 2009.