

# 複合ウェーブレットトラジェクトリモデルに基づく音声合成\*

北条伸克<sup>1</sup>, 亀岡弘和<sup>1,2</sup>, 嵯峨山茂樹<sup>1</sup>  
(<sup>1</sup>東大院 情報理工, <sup>2</sup>NTT CS 研)

## 1 はじめに

本稿では、高品質なテキスト音声合成方式へ将来的に組み込むことを見据え、音声スペクトル包絡時系列の生成過程の確率モデルを提案する。統計的モデルに基づくテキスト音声合成では、音声における様々な性質や挙動をいかに適切に生成モデルの形で記述できるかが合成音声の品質を左右する。特に、音声のダイナミクスを適切にモデル化することは、高品質なテキスト音声合成のために重要である。例えば、スペクトル包絡特徴量の動的特徴には話者の個性が現れることが知られている [1]。

トラジェクトリ HMM [2] では、音素やコンテキストラベルなどの離散的なシンボルと連続的に時間変化する音声スペクトルとの関係を、メルケプストラムの静的、動的特徴を出力する隠れマルコフモデル (Hidden Markov Model; HMM) を用いてモデル化することにより、高品質な音声合成が可能となった。我々の研究室では、スペクトル包絡系列を生成モデル化する際、その時間変化を、トラジェクトリ HMM と同様にスペクトル特徴量のダイナミクスのモデルにより表現することを目指す一方で、スペクトル包絡系列の情報において特に重要であるフォルマント周波数軌跡に着目し、その生成過程を直接モデル化する手法を検討している。例えば、我々の研究室では、複合ウェーブレットモデル (Composite Wavelet Model; CWM) [4, 5] のフォルマント周波数に相当するパラメータの時間軌跡をモデル化することにより、スペクトル包絡系列の生成過程をモデル化した [3]。

フォルマント周波数軌跡のモデルを考える際、無声音のスペクトル包絡をいかに表現するかが問題となる。声帯を振動源、声道をフィルタとして有声音の生成過程を捉えたとき、フォルマント周波数は声道における共振周波数として捉えられ、発話する音素系列に対応した声道の形状変化に基づき時間方向になめらかに変化すると考えられる。[3] の手法は、このようなフォルマントの性質に着目し、提案されたものであった。一方で、呼気の摩擦や破裂などにより生成される無声音のスペクトル包絡では、その生成過程が上記の有声音のものとは異なり、一般にフォルマントに相当するスペクトル包絡ピークは観測されない。したがって、フォルマントによりスペクトル包絡系列全体を表現するような生成モデルにおいては、特に無声音区間において表現に無理があり、フォルマント周波数軌跡も不自然なものとなる可能性がある。

本稿では、フォルマントに相当するスペクトル包絡ピークを持つ有声音スペクトルモデルと、平坦な形状を持つ無声音スペクトルモデルとの重畳によりモデル化されるスペクトル包絡の時系列としてスペクトル包絡系列全体が表現可能であると仮定し、有声音、無声音のスペクトルモデルを、それぞれ異なる生成過程を経る CWM パラメータセットにより表現するモデルを提案する。このようなモデル化により、フォルマント周波数の時間軌跡と、無声音のスペクトル包絡の双方が自然に表現可能であると期待される。本稿は、提案モデルの定式化及びパラメータ推定アルゴリズムについて述べ、実音声のスペクトル包絡からフォルマント周波数の時間軌跡を推定する実験

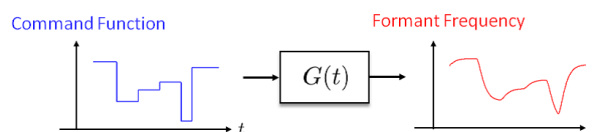


Fig. 1 線形系によるフォルマント周波数軌跡の生成過程を通して、提案手法の有効性を確認する。

## 2 音声スペクトル生成過程の確率モデル化

### 2.1 複合ウェーブレットモデル

スペクトル包絡における各ピークを Gauss 分布関数で近似的に表現できるとすると、スペクトル包絡全体を Gauss 分布関数の重ね合わせ、すなわち混合 Gauss 分布関数モデル (Gaussian Mixture Model; GMM) で表現することができる。このようなスペクトル包絡の表現を、複合ウェーブレットモデル (Composite Wavelet Model; CWM) [4, 5] と呼ぶ。CWM におけるスペクトル包絡モデル  $\phi_{\omega, l}$  は

$$\phi_{\omega, l} = \sum_k^K \frac{w_{k, l}}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(\omega - \mu_{k, l})^2}{2\sigma_{k, l}^2}\right) \quad (1)$$

で与えられる。ただし、 $\omega, l$  は周波数と時刻のインデックス、 $k$  は Gauss 関数のインデックスであり、 $K$  は GMM の混合数である。また、 $\mu_{k, l}, \sigma_{k, l}^2, w_{k, l}$  は、それぞれ Gauss 分布関数を統計分布と見なした際の平均・分散・重みに対応し、各スペクトル包絡ピークの周波数・鋭さ・強度に対応するパラメータである。

### 2.2 複合ウェーブレットトラジェクトリモデル

#### 2.2.1 フォルマント軌跡の生成過程に関する仮説

声帯振動が共振することによって生じるフォルマントの周波数軌跡には声道の運動に伴う何らかの物理的な制約が付随すると考えられるが、フォルマント周波数軌跡が藤崎の  $F_0$  パターン生成過程モデル [6] と同様のメカニズムによって生じると仮定する。具体的には、フォルマント周波数の対数の時間軌跡を、Fig. 1 に示すように音素区間ごとに一定値をとる階段状の指令関数 (以後、音素指令関数) にインパルス応答

$$G(t) = \begin{cases} \alpha^2 t e^{-\alpha t} & (t \geq 0) \\ 0 & (t < 0) \end{cases} \quad (2)$$

が畳み込まれ ( $\alpha$  は固有角周波数)、二次線形系の出力として生じたものとする。なお、同様のフォルマント周波数軌跡のモデルが、規則音声合成 [7]、フォルマント周波数軌跡推定 [8] において過去に提案されている。また、二次線形系の仮定が置かれた他の音声生成過程モデルの例として、音素認識を目的とした調音運動の動的モデルが提案されている [9]。

#### 2.2.2 確率モデルの定式化

前述のフォルマント軌跡モデルに基づきスペクトル包絡系列をモデル化する手法を考える。ここで、1

\* Speech synthesis based on composite wavelet trajectory model. by HOJO, Nobukatsu, KAMEOKA Hirokazu, SAGAYAMA Shigeki (The University of Tokyo)

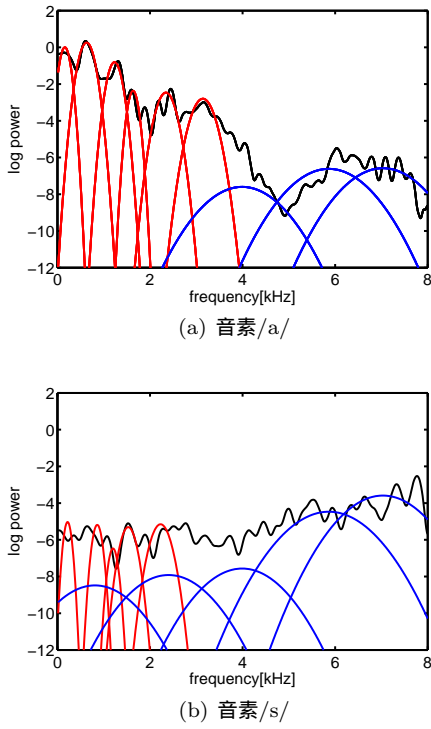


Fig. 2 有声音スペクトルモデル (赤) と無声音スペクトルモデル (青) の CWM パラメータの例

章で述べたように、有声音のものとは性質の異なる無声音のスペクトル包絡をもフォルマントの存在を前提としてモデル化することは不自然である。したがって、有声音、無声音の双方が現れるスペクトル包絡系列をモデル化するためには、例えば、無声音のスペクトル包絡を別途モデル化する方法が考えられる。また、有声子音では、呼気の摩擦や破裂等と声帯の振動が同時に生じるため、そのスペクトル包絡は有声音、無声音の双方の特徴を併せ持つ可能性がある。

以上より、スペクトル包絡系列の生成過程は、有声音、無声音の特徴をそれぞれ持つスペクトルや、その双方の特徴を併せ持つスペクトルが各時刻で生成される過程として見なすことができる。このようなスペクトル包絡系列の生成過程を自然にモデル化するためには、例えば、有声音、無声音のスペクトルを別途モデル化し、それらが重畳されたものをスペクトル包絡系列のモデルとする方法が可能であると考えられる。複合ウェーブレットトラジェクトリモデルでは、有声音、無声音のスペクトルを、それぞれ異なる生成過程を経る CWM パラメータのセットによりモデル化し、それらを重畳することによりスペクトル包絡系列のモデルとする。便宜的に、以後 GMM の分散パラメータ  $\sigma^2$  の逆数  $\rho$  をパラメータとみなし、有声音、無声音のスペクトルモデルの CWM パラメータセットをそれぞれ、 $\{\rho^{(v)}, \mathbf{w}^{(v)}, \boldsymbol{\mu}^{(v)}\}$ ,  $\{\rho^{(u)}, \mathbf{w}^{(u)}, \boldsymbol{\mu}^{(u)}\}$  とする。ここで  $m = 1, \dots, M$ ,  $M$  は有声音スペクトルモデルの GMM 混合数,  $n = 1, \dots, N$ ,  $N$  は無声音スペクトルモデルの GMM 混合数であり、 $\rho^{(v)} = \{\rho_{m,l}^{(v)}\}_{m,l}$  などとした。特に  $\boldsymbol{\mu}^{(v)}$  はフォルマント周波数軌跡に対応すると見なすことができる。

提案モデルのパラメータ例を Fig. 2 に示した。有声音 /a/ のスペクトルは赤で示された有声音スペクトルモデルのパラメータにより、無声音 /s/ のスペクトルは青で示された無声音スペクトルモデルのパラメータにより主にモデル化される。

有声音スペクトルの生成過程は、2.2.1 節のフォルマント軌跡の生成過程に関する仮説に基づき定式化される。将来的に提案モデルを HMM 音声合成の枠

組みに組み込む際には、音素指令関数の生成過程は、[3] と同様に、音素に対応する HMM 状態の出力分布として定式化することが可能である。本稿では、提案モデルが HMM 音声合成のために十分なスペクトル包絡系列の表現能力があることを確認するため、提案モデルによる分析合成のための定式化を行う。分析合成においては、指令関数  $\{u_{m,l}\}_{m,l}$  は、各時刻  $l$  で共通の分布に従い生成されると仮定する。次に、2.2.1 節で議論した様に、生成された指令関数  $\{u_{m,l}\}_{m,l}$  に二次線形系のインパルス応答が畳み込まれて Gauss 分布関数の平均値の軌跡  $\{\mu_{m,l}\}_{m,l}$  が生じるとする。提案モデルでは、それぞれの生成過程を、

$$P(u_{m,l}) = \mathcal{N}(u_{m,l}; m_m, \eta_m^2) \quad (3)$$

$$P(\mu_{m,l} | \{u_{m,l}\}_l) = \mathcal{LN}(G_{m,l} * u_{m,l}, \nu_m^2) \quad (4)$$

とする。ここで  $\mathcal{LN}(x; \mu, \sigma^2)$  は対数正規分布であり、 $\log x$  が正規分布  $\mathcal{N}(x; \mu, \sigma^2)$  に従うことと等価である。また、各フォルマントの帯域幅やパワーに対応するパラメータの生成過程を、

$$P(w_{m,l}^{(v)}) = \text{Gamma}(w_{m,l}^{(v)}; a_m^{(v)}, b_m^{(v)}) \quad (5)$$

$$P(\rho_{m,l}^{(v)}) = \text{Gamma}(\rho_{m,l}^{(v)}; c_m^{(v)}, d_m^{(v)}) \quad (6)$$

の様にモデル化する。ただし、 $\text{Gamma}(x; a, b)$  はガンマ分布

$$\text{Gamma}(x; a, b) = x^{a-1} \frac{\exp(-x/b)}{\Gamma(a) b^a} \quad (7)$$

である。

以上のモデル化により、各フォルマント周波数、帯域幅及びパワーの先験的情報を、それぞれ (3)、(5) 及び (6) のパラメータの設定により、モデルに組み込むことが可能である。

続いて、無声音スペクトルの生成過程をモデル化する。提案モデルでは、

$$P(w_{n,l}^{(u)}) = \text{Gamma}(w_{n,l}^{(u)}; a_n^{(u)}, b_n^{(u)}) \quad (8)$$

$$P(\rho_{n,l}^{(u)}) = \text{Gamma}(\rho_{n,l}^{(u)}; c_n^{(u)}, d_n^{(u)}) \quad (9)$$

とモデル化し、各 Gauss 関数の平均パラメータ  $\mu_{n,l}^{(u)}$  は定数とする。提案モデルでは特に、スペクトル包絡の各帯域における緩やかな形状を表現するため、 $\rho_{n,l}^{(u)}$  は、 $\rho_{n,l}^{(v)}$  に比べより帯域幅が広がるように生成確率を与える。

最後に、各時刻の CWM パラメータが与えられたときに観測スペクトル包絡  $y_{\omega,l}$  が生じる確率を

$$\phi_{\omega,l} = \sum_{m=1}^M \psi_{m,\omega,l}^{(v)} + \sum_{n=1}^N \psi_{n,\omega,l}^{(u)} \quad (10)$$

$$\psi_{m,\omega,l}^{(v)} = \sum_{m=1}^M w_{m,l}^{(v)} \sqrt{\frac{\rho_{m,l}^{(v)}}{2\pi}} \exp\left(-\frac{\rho_{m,l}^{(v)}(\omega - \mu_{m,l}^{(v)})^2}{2}\right) \quad (11)$$

$$\psi_{n,\omega,l}^{(u)} = \sum_{n=1}^N w_{n,l}^{(u)} \sqrt{\frac{\rho_{n,l}^{(u)}}{2\pi}} \exp\left(-\frac{\rho_{n,l}^{(u)}(\omega - \mu_{n,l}^{(u)})^2}{2}\right) \quad (12)$$

$$P(y_{\omega,l} | \rho^{(v)}, \mathbf{w}^{(v)}, \boldsymbol{\mu}^{(v)}, \rho^{(u)}, \mathbf{w}^{(u)}) = \text{Poisson}(y_{\omega,l}; \phi_{\omega,l}) \quad (13)$$

とする．ここで  $\text{Poisson}(x; \lambda)$  は Poisson 分布である．なお，この仮定の下での  $\lambda$  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  $x$  と  $\lambda$  の最適フィッティング問題と等価となることが知られている [10]．

以上の提案モデルの推定すべきパラメータをまとめて  $\Theta = \{\rho^{(v)}, \mathbf{w}^{(v)}, \mu^{(v)}, \mathbf{u}, \rho^{(u)}, \mathbf{w}^{(u)}\}$  とする．ここで新たに  $\mathbf{u} = \{u_{m,l}\}_{m,l}$  とした．次章では，観測スペクトル包絡系列  $Y = \{y_{\omega,l}\}_{\omega,l}$  が与えられた下での事後確率  $P(\Theta|Y)$  を最大化するパラメータ推定アルゴリズムについて述べる．

### 3 パラメータ推定アルゴリズム

$P(\Theta|Y)$  を最大化する  $\Theta$  を解析的に求めることは難しいが，各変数について局所最適化を繰り返すことは可能である．この時  $\log P(\Theta|Y)$  は，

$$\log P(\Theta|Y) \stackrel{c}{=} \log P(Y|\Theta) + \log P(\Theta) \quad (14)$$

$$\begin{aligned} \log P(\Theta) &\stackrel{c}{=} \log P(\rho^{(v)}) + \log P(\mathbf{w}^{(v)}) \\ &\quad + \log P(\mathbf{u}) + \log P(\mu^{(v)}|\mathbf{u}) \\ &\quad + \log P(\rho^{(u)}) + \log P(\mathbf{w}^{(u)}) \end{aligned} \quad (15)$$

と書ける．ここで， $\stackrel{c}{=}$  は定数部分を除いた場合の等号を意味する．先述のように， $-\log P(Y|\Theta)$  は定数項を除けば観測スペクトル包絡  $y_{\omega,l}$  とスペクトル包絡モデル  $\phi_{\omega,l}$  との間の  $I$  ダイバージェンスと等しく [10]，さらに，

$$\begin{aligned} \sum_{\omega} w \sqrt{\frac{\rho}{2\pi}} \exp\left(-\frac{\rho(\omega - \mu)^2}{2}\right) \\ \simeq \int_{-\infty}^{\infty} \sqrt{\frac{\rho}{2\pi}} w \exp\left(-\frac{\rho(\omega - \mu)^2}{2}\right) d\omega \\ = w \end{aligned} \quad (16)$$

となることを用いれば，

$$\begin{aligned} -\log P(Y|\Theta) &\stackrel{c}{=} \sum_{\omega,l} \left( y_{\omega,l} \log \frac{y_{\omega,l}}{\phi_{\omega,l}} - y_{\omega,l} + \phi_{\omega,l} \right) \\ &\stackrel{c}{=} \sum_{\omega,l} (\phi_{\omega,l} - y_{\omega,l} \log \phi_{\omega,l}) \\ &\simeq \sum_{m,l} w_{m,l}^{(v)} + \sum_{n,l} w_{n,l}^{(u)} - \sum_{\omega,l} y_{\omega,l} \log \phi_{\omega,l} \end{aligned} \quad (17)$$

が言える．この式の  $-y_{\omega,l} \log \phi_{\omega,l}$  の項に対し，負の対数関数の凸性を利用し，Jensen の不等式を用いることで

$$\begin{aligned} -y_{\omega,l} \log \phi_{\omega,l} &\leq -y_{\omega,l} \sum_m \gamma_{m,\omega,l}^{(v)} \log \frac{\psi_{m,\omega,l}^{(v)}}{\gamma_{m,\omega,l}^{(v)}} \\ &\quad - y_{\omega,l} \sum_n \gamma_{n,\omega,l}^{(u)} \log \frac{\psi_{n,\omega,l}^{(u)}}{\gamma_{n,\omega,l}^{(u)}} \end{aligned} \quad (18)$$

のように上界関数を設計することができる．また， $-\log P(\mu^{(v)}|\mathbf{u})$  の  $(\sum_{\tau} G_{m,l-\tau} u_{m,\tau})^2$  の項に対し，二次関数の凸性を利用し，同様に Jensen の不等式を用いることで

$$\left( \sum_{\tau} G_{m,l-\tau} u_{m,\tau} \right)^2 \leq \sum_{\tau} \frac{(G_{m,l-\tau} u_{m,\tau})^2}{\lambda_{\tau,m,l}} \quad (19)$$

のように上界関数を設計することができる [11]．さらに  $-\log P(\mu^{(v)}|\mathbf{u})$  の中の  $(\log \mu_{m,l}^{(v)})^2$  の項に関しては，

$$\begin{aligned} (\log \mu_{m,l}^{(v)})^2 &\leq \frac{1}{\mu_{m,l}^{(v)}} + \left( \frac{2 \log \xi_{m,l}}{\xi_{m,l}} + \frac{1}{\xi_{m,l}^2} \right) \mu_{m,l}^{(v)} \\ &\quad + |\log \xi_{m,l}|^2 - 2 \log \xi_{m,l} - \frac{2}{\xi_{m,l}} \end{aligned} \quad (20)$$

のように上界関数を設計することができる [12]．ここで  $\gamma_{m,\omega,l}^{(v)}, \gamma_{n,\omega,l}^{(u)}, \lambda_{\tau,m,l}, \xi_{m,l}$  は補助変数である．式 (18) ~ (20) より， $-\log p(\Theta|Y)$  の上界関数を設計することができ，これを補助関数として補助関数法を適用することができる．まず，補助変数の更新式は，上述の不等式の等号成立条件，すなわち，

$$\gamma_{m,\omega,l}^{(v)} = \frac{\psi_{m,\omega,l}^{(v)}}{\phi_{\omega,l}} \quad (21)$$

$$\gamma_{n,\omega,l}^{(u)} = \frac{\psi_{n,\omega,l}^{(u)}}{\phi_{\omega,l}} \quad (22)$$

$$\lambda_{\tau,m,l} = \frac{G_{m,l-\tau} u_{m,\tau}}{\sum_{\tau'} G_{m,l-\tau'} u_{m,\tau'}} \quad (23)$$

$$\xi_{m,l} = \mu_{m,l}^{(v)} \quad (24)$$

で与えられる．

$\mu^{(v)}$  以外のパラメータの更新式は，

$$u_{m,l} = \frac{\frac{m_m}{\eta_m^2} + \sum_{\tau \geq l} \frac{G_{k,\tau-l} \log \mu_{m,\tau}}{\nu_m^2}}{\frac{1}{\eta_m^2} + \sum_{\tau \geq l} \frac{G_{m,\tau-l}^2}{\nu_m^2 \lambda_{l,m,\tau}}} \quad (25)$$

$$\rho_{m,l}^{(v)} = \frac{2(c_m^{(v)} - 1) + \sum_{\omega} y_{\omega,l} \gamma_{m,\omega,l}^{(v)}}{\frac{2}{d_m^{(v)}} + \sum_{\omega} y_{\omega,l} \gamma_{m,\omega,l}^{(v)} (\omega - \mu_{m,l}^{(v)})^2} \quad (26)$$

$$\rho_{n,l}^{(u)} = \frac{2(c_n^{(u)} - 1) + \sum_{\omega} y_{\omega,l} \gamma_{n,\omega,l}^{(u)}}{\frac{2}{d_n^{(u)}} + \sum_{\omega} y_{\omega,l} \gamma_{n,\omega,l}^{(u)} (\omega - \mu_{n,l}^{(u)})^2} \quad (27)$$

$$w_{m,l}^{(v)} = \frac{a_m^{(v)} - 1 + \sum_{\omega} y_{\omega,l} \gamma_{m,\omega,l}^{(v)}}{\frac{1}{b_m^{(v)}} + 1} \quad (28)$$

$$w_{n,l}^{(u)} = \frac{a_n^{(u)} - 1 + \sum_{\omega} y_{\omega,l} \gamma_{n,\omega,l}^{(u)}}{\frac{1}{b_n^{(u)}} + 1} \quad (29)$$

となる． $\mu^{(v)}$  については，

$$p_3 \mu_{m,l}^{(v)3} + p_2 \mu_{m,l}^{(v)2} + p_1 \mu_{m,l}^{(v)} + p_0 = 0 \quad (30)$$

$$p_3 = \sum_{\omega} y_{\omega,l} \gamma_{m,\omega,l}^{(v)} \rho_{m,l}^{(v)} \quad (31)$$

$$p_2 = \frac{2\xi_{m,l} \log \xi_{m,l} + 1}{2\nu_m^2 \xi_{m,l}^2} - \sum_{\omega} y_{\omega,l} \gamma_{m,\omega,l}^{(v)} \rho_{m,l}^{(v)} \omega \quad (32)$$

$$p_1 = 1 - \frac{1}{\nu_m^2} \sum_{\tau \leq l} G_{m,l-\tau} u_{m,\tau} \quad (33)$$

$$p_0 = -\frac{1}{2\nu_m^2} \quad (34)$$

とにおいて，式 (30) の正の解のうち  $-\log P(\Theta|Y)$  を最も小さくする  $\mu_{k,l}^{(v)}$  を選ばよ．



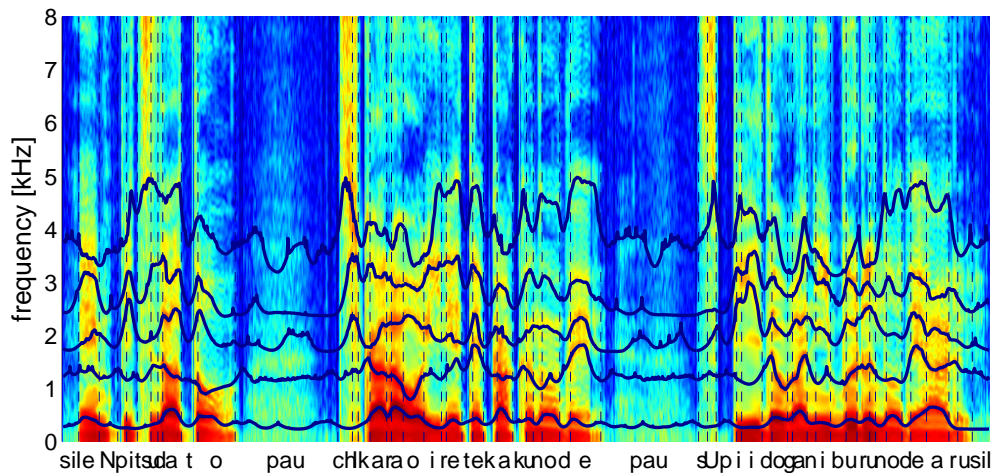


Fig. 3 No.436 のスペクトル包絡系列 (カラーマップ表示) と提案法による推定フォルマント周波数軌跡 (実線) (点線は音素境界)

以上の更新則を十分な回数反復することで、 $P(\Theta|Y)$  を局所最大化するパラメータ  $\Theta$  を推定することができる。

#### 4 実験

提案モデルが実音声のフォルマント周波数の時間軌跡をよく表現できていることを確認するために実験を行った。フォルマント CWM パラメータセットの混合数は  $M=5$ 、CWM パラメータセットの混合数は  $N=7$ 、 $\alpha_m = 30$  とした。スペクトル包絡の解析は、STRAIGHT[13] により行った。

音声サンプルとして、HTS2.1 のデモスクリプト [14] に同梱された男性話者の音声 (サンプリング周波数 16kHz、サンプルサイズ 16bit) のうち、No.436 の音声を用いた実験結果を Fig. 3 に示した。図は、解析されたスペクトル包絡系列に提案した推定フォルマント周波数の時間軌跡を重ねて描いた図である。推定したスペクトルピークが実音声のフォルマント周波数軌跡をうまく表現できていることが確認できる。

#### 5 おわりに

本稿では、フォルマント周波数軌跡モデルに基づくスペクトル包絡系列の生成モデルとして、有声音、無声音のスペクトルをそれぞれ異なる生成過程を経た CWM パラメータセットによりモデル化し、それらの重畳によりスペクトル包絡系列をモデル化する、複合ウェーブレットトラジェクトリモデルを提案した。実験を通して、提案モデルが十分に観測スペクトルを表現し、フォルマント周波数軌跡が推定可能であることを確認した。今後は、提案モデルを HMM に組み込むことにより、テキスト音声合成の高品質化の実現を目指す。

#### 参考文献

[1] 嵯峨山, 板倉, “音声の動的尺度に含まれる個人性情報,” 音講論 (春), No. 3-2-7, pp. 589-590, 1979.  
 [2] K. Tokuda *et al.*, “Trajectory modeling based on HMMs with the explicit relationship between static and dynamic features,” in *Proc. of Eurospeech*, pp. 837-840, 2004.

[3] 吉里他, “フォルマント周波数軌跡を潜在パラメータとした音声スペクトル生成過程の確率モデル,” 音講論 (春), No. 1-7-7, pp. 277-280, 2013.  
 [4] 槐他, “複合ウェーブレットモデルに基づく音声の分析合成,” 信学技報, Vol. 105, No. 372, pp. 1-6, 2005.  
 [5] P. Zolfaghari, T. Robinson, “Formant Analysis using Mixtures of Gaussians,” in *Proc. IC-SLP'96*, Vol. 2, pp. 1229-1232, 1996.  
 [6] H. Fujisaki, In *Vocal Physiology: Voice Production, Mechanisms and Functions*, (O. Fujimura, ed.) Raven Press, pp. 347-355, 1988.  
 [7] L. R. Rabiner, “Speech synthesis by rule : An acoustic domain approach,” Ph. D. thesis, M. I. T., 1967.  
 [8] 板橋他, “線形 2 次系モデルによるホルマント軌跡の記述とセグメンテーション,” 電子技術総合研究所彙報, Vol. 6, No.40, pp. 530-541, 1976.  
 [9] 誉田, “調音モデルにもとづく音声の特徴抽出に関する研究,” 博士論文, 早稲田大学, 1977.  
 [10] H. Kameoka, “Statistical Approach to Multipitch Analysis,” Ph.D. Thesis, The University of Tokyo, 2007.  
 [11] 亀岡他, “音声のスパース性と非負制約つき畳み込みモデルに基づくパワースペクトル領域残響除去,” 音講論 (秋), No. 3-8-10, pp. 705-708, 2008.  
 [12] H. Kameoka *et al.*, “Constrained and regularized variants of non-negative matrix factorization incorporating music-specific constraints,” in *Proc. ICASSP'12*, pp. 5365-5368, 2012.  
 [13] H. Kawahara *et al.*, “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction,” *Speech Communication*, Vol. 27, No. 3-4, pp. 187-207, 1999.  
 [14] <http://hts.sp.nitech.ac.jp/>