日本音響学会2010年秋季研究発表会

全極型声道モデルと F0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

日本電信電話株式会社
NTTコミュニケーション科学基礎研究所

本研究の目的・意義

目的：音声信号から声道スペクトルと基本周波数（F0）を同時推定

→そのための音声生成モデルを提案

提案モデル：

全極型声道モデルとF0パターン生成過程モデル（藤崎モデル）を同時内包する音声スペクトルモデル

同時モデル化によるメリット：

- 声道スペクトル推定
  - F0推定が正確なほど推定精度が上がる [El-Jaroudi & Makhoul, 1991]
- 音声F0推定
  - スペクトル包絡が滑らかという拘束があれば倍ピッチ・半ピッチ誤りを減らせる [Kameoka et al., 2006]
  - F0パターンが滑らかという拘束があればGross errorを減らせる
- 藤崎モデルパラメータ推定
  - 観測スペクトルから直接的に推定可能になる

提案法の導出アウトライン

有声音の信号モデル

↓

ウェーブレット変換
スペクトル表現
（定Q log-normal
フィルタバンクと命名）

未知パラメータ：w, Ω

↓

統計モデル化

$P(Y|w,\Omega)$

w の事前分布の設計
（全極型声道モデルに基づく）

ハイパーパラメータ：π_w

$P(w|\Omega,\pi_w)$

Ω の事前分布の設計
（藤崎モデル（前発表）に基づく）

ハイパーパラメータ：π_Ω

$P(\Omega|\pi_\Omega)$

パラメータ推定

$$P(w,\Omega,\pi_w,\pi_\Omega|Y) \propto P(Y|w,\Omega)P(w|\Omega,\pi_w)P(\Omega|\pi_\Omega)$$

を局所最大化する反復アルゴリズム

提案法の導出アウトライン

有声音の信号モデル

↓

ウェーブレット変換
スペクトル表現
（定Q log-normal
フィルタバンクと命名）

未知パラメータ：w, Ω

↓

統計モデル化

$P(Y|w,\Omega)$

w の事前分布の設計
（全極型声道モデルに基づく）

ハイパーパラメータ：π_w

$P(w|\Omega,\pi_w)$

Ω の事前分布の設計
（藤崎モデル（前発表）に基づく）

ハイパーパラメータ：π_Ω

$P(\Omega|\pi_\Omega)$

パラメータ推定

$P(w,\Omega,\pi_w,\pi_\Omega|Y) \propto P(Y|w,\Omega)P(w|\Omega,\pi_w)P(\Omega|\pi_\Omega)$
を局所最大化する反復アルゴリズム

信号モデル → ウェーブレット変換スペクトル

◆時刻 t 周辺の音声信号モデル

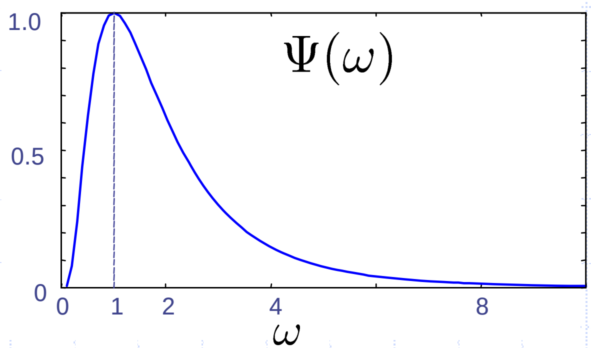
$$s_t(u) = \sum_{n=1}^{N} w_n(t) e^{\mathrm{j}(n\mu(t)(u-t)+c_n(t))}$$

u: 時刻
N: 調波成分数
$w_n(u) \geq 0$: 第 n 調波成分の瞬時振幅
$\mu(t)$: 瞬時基本周波数

(導出省略)

ウェーブレット変換

$$S_{\mathrm{w}}(x,t) = \langle s_t, \psi_{x,t} \rangle \overset{\because\ \text{一般化Parsevalの定理}}{=} \langle \underbrace{S_t(\omega)}_{s_t(u)\text{のFourier変換}}, \Psi(e^{-x}\omega) \rangle_{\omega \in \mathbb{R}}$$

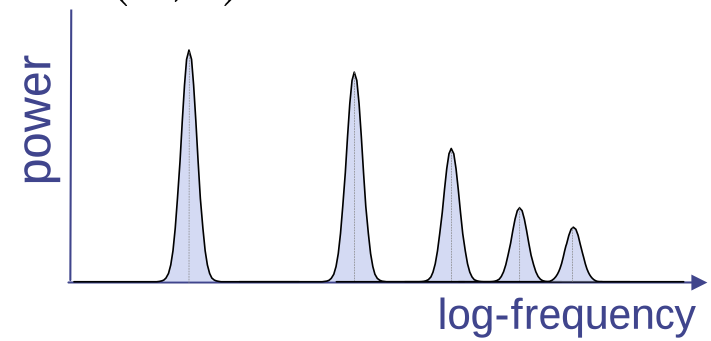


$$\Psi(\omega) = \begin{cases} e^{-\frac{1}{4\sigma^2}(\log \omega)^2} & (\omega > 0) \\ 0 & (\omega \leq 0) \end{cases}$$

「定Q log-normalフィルタバンク」と命名

$$|S_{\mathrm{w}}(x,t)|^2 \simeq \sum_{n=1}^{N} w_n(t)^2 e^{-\frac{1}{2\sigma^2}(x - \underbrace{\Omega(t)}_{\log \mu(t)} - \log n)^2} =: \lambda(x,t)$$

瞬時対数基本周波数



ウェーブレット変換スペクトルの統計モデル

- 観測ウェーブレット変換スペクトル $Y[k,l] := Y(x_k, t_l)$

k: 対数周波数インデックス
l: 時刻インデックス

- $Y[k,l]$ にできるだけ良くフィットするスペクトルモデル $\lambda[k,l] := \lambda(x_k, t_l)$ を求めたい

未知パラメータ: $\underbrace{w_n[l]}_{\text{調波成分の振幅}} := w_n(t_l)$, $\underbrace{\Omega[l]}_{\text{F0パターン（対数F0の時間軌跡）}} := \Omega(t_l)$

- スペクトル間の近さを測るための尺度の導入
  - Iダイバージェンス

$$I(\lambda) := \sum_{k,l} \left(Y[k,l] \log \frac{Y[k,l]}{\lambda[k,l]} - Y[k,l] + \lambda[k,l] \right)$$

  - $I(\lambda)$ を λ に関して最小化することは,

$$Y[k,l] \sim \mathrm{Pois}(\lambda[k,l]) \longrightarrow Y[k,l] \text{ の統計モデル化}$$

を仮定した下での λ の最尤推定と等価

提案法の導出アウトライン

有声音の信号モデル

↓

ウェーブレット変換
スペクトル表現
（定Q log-normal
フィルタバンクと命名）

未知パラメータ：w, Ω

↓

統計モデル化

$P(Y|w,\Omega)$

w の事前分布の設計
（全極型声道モデルに基づく）

ハイパーパラメータ：π_w

$P(w|\Omega,\pi_w)$

Ω の事前分布の設計
（藤崎モデル（前発表）に基づく）

ハイパーパラメータ：π_Ω

$P(\Omega|\pi_\Omega)$

パラメータ推定

$P(w,\Omega,\pi_w,\pi_\Omega|Y) \propto P(Y|w,\Omega)P(w|\Omega,\pi_w)P(\Omega|\pi_\Omega)$
を局所最大化する反復アルゴリズム

提案法の導出アウトライン

有声音の信号モデル

↓

ウェーブレット変換スペクトル表現
（定Q log-normal フィルタバンクと命名）

未知パラメータ：w, Ω

↓

統計モデル化

$P(Y|w,\Omega)$

w の事前分布の設計
（全極型声道モデルに基づく）

ハイパーパラメータ：π_w

$P(w|\Omega,\pi_w)$

Ω の事前分布の設計
（藤崎モデル（前発表）に基づく）

ハイパーパラメータ：π_Ω

$P(\Omega|\pi_\Omega)$

パラメータ推定

$P(w,\Omega,\pi_w,\pi_\Omega|Y) \propto P(Y|w,\Omega)P(w|\Omega,\pi_w)P(\Omega|\pi_\Omega)$

を局所最大化する反復アルゴリズム

③全極モデルに基づく w の事前分布設計 (1/2)

音声信号モデル

$$s_{t_l}(u) = \sum_{n=1}^{N} w_n[l] e^{\mathrm{j}(ne^{\Omega[l]}(u-t_l)+c_n(t_l))}$$

$$\xrightarrow[\text{（周期}u_0\text{でサンプリング）}]{\text{離散時間表現}} \quad s_l[i] := s_{t_l}(iu_0) = \sum_{n=1}^{N} w_n[l] e^{\mathrm{j}(ne^{\Omega[l]}(iu_0-t_l)+c_n(t_l))}$$

$s_l[i]$ を全極システムの出力信号と仮定

- 拘束式: $s_l[i] = \sum_{m=0}^{M} a_l[m] s_l[i-m] + \epsilon_l[i]$
- システム入力: $\epsilon_l[i] = \sum_{n=1}^{N} v_{l,n} e^{\mathrm{j}ne^{\Omega[l]}iu_0}$ ← 基本周波数 $e^{\Omega[l]}$

$$s_l[i] = \sum_{n=1}^{N} \frac{v_{l,n}}{A_l(e^{\mathrm{j}ne^{\Omega[l]}u_0})} e^{\mathrm{j}ne^{\Omega[l]}iu_0}$$

ただし $A_l(z) = 1 - a_l[1]z^{-1} \cdots - a_l[M]z^{-M}$

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

③全極モデルに基づく w の事前分布設計 (2/2)

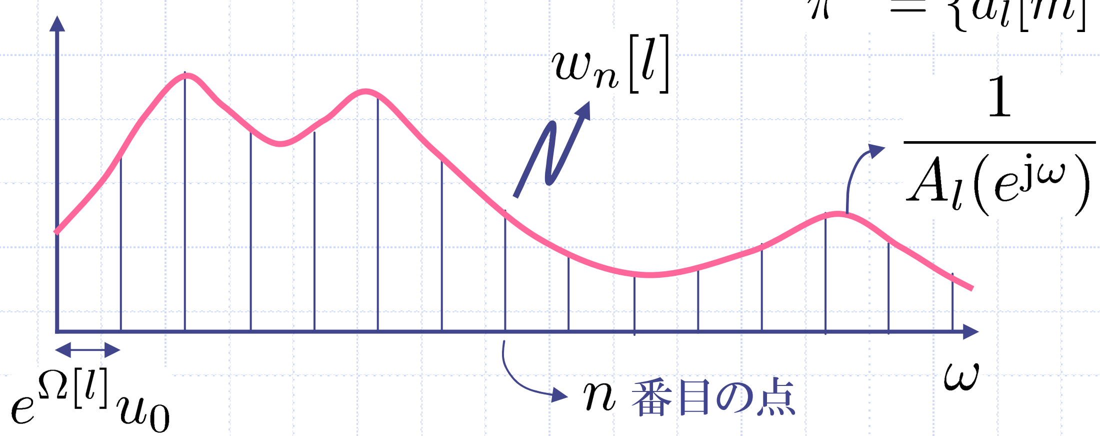
- システム入力の確率モデル化

$$\epsilon_l[i] = \sum_{n=1}^{N} \underline{v_{l,n}} e^{\mathrm{j} n e^{\Omega[l]} i u_0} \longrightarrow v_{l,n} \sim \mathcal{N}_{\mathbb{C}}(0,1)$$

$$w_n[l] = \left| \frac{v_{l,n}}{A_l(e^{\mathrm{j}ne^{\Omega[l]}u_0})} \right| \longrightarrow w_n[l] | \Omega[l] \sim \mathrm{Rayleigh}\left(\frac{1}{|A_l(e^{\mathrm{j}ne^{\Omega[l]}u_0})|} \right)$$

ハイパーパラメータは全極モデルパラメータ

$$\pi^w = \{a_l[m] \mid 0 \le m \le M, 1 \le l \le L\}$$



提案法の導出アウトライン

有声音の信号モデル

↓

ウェーブレット変換スペクトル表現
（定Q log-normal フィルタバンクと命名）

未知パラメータ：w, Ω

↓

統計モデル化

$P(Y|w,\Omega)$

w の事前分布の設計
（全極型声道モデルに基づく）

ハイパーパラメータ：π_w

$P(w|\Omega,\pi_w)$

Ω の事前分布の設計
（藤崎モデル（前発表）に基づく）

ハイパーパラメータ：π_Ω

$P(\Omega|\pi_\Omega)$

パラメータ推定

$P(w,\Omega,\pi_w,\pi_\Omega|Y) \propto P(Y|w,\Omega)P(w|\Omega,\pi_w)P(\Omega|\pi_\Omega)$

を局所最大化する反復アルゴリズム

提案法の導出アウトライン

有声音の信号モデル

↓

ウェーブレット変換
スペクトル表現
（定Q log-normal
フィルタバンクと命名）

未知パラメータ：w, Ω

↓

統計モデル化

$P(Y|w, \Omega)$

w の事前分布の設計
（全極型声道モデルに基づく）

ハイパーパラメータ：π_w

$P(w|\Omega, \pi_w)$

Ω の事前分布の設計
（藤崎モデル（前発表）に基づく）

ハイパーパラメータ：π_Ω

$P(\Omega|\pi_\Omega)$

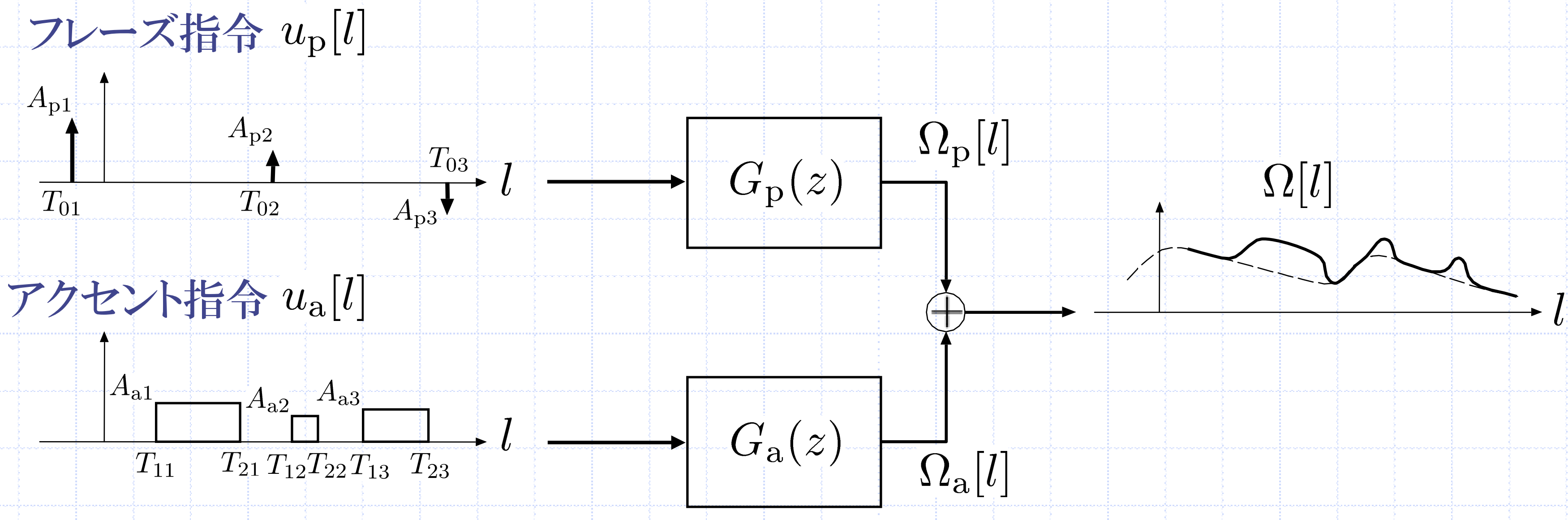
パラメータ推定

$P(w, \Omega, \pi_w, \pi_\Omega|Y) \propto P(Y|w, \Omega)P(w|\Omega, \pi_w)P(\Omega|\pi_\Omega)$

を局所最大化する反復アルゴリズム

④藤崎モデルに基づく$\boldsymbol{\Omega}$の事前分布設計

- 前発表のとおりなので詳しいことは省略



$$\boldsymbol{\Omega} \sim \mathcal{N}(\underbrace{\boldsymbol{G}_{\mathrm{p}}^{-1}\boldsymbol{\mu}_{\mathrm{p}} + \boldsymbol{G}_{\mathrm{a}}^{-1}\boldsymbol{\mu}_{\mathrm{a}} + \mu_{\mathrm{b}}\mathbf{1}}_{\boldsymbol{\mu}_{\Omega}},\ \underbrace{\boldsymbol{G}_{\mathrm{p}}^{-1}\boldsymbol{\Sigma}_{\mathrm{p}}(\boldsymbol{G}_{\mathrm{p}}^{-1})^{\mathrm{T}} + \boldsymbol{G}_{\mathrm{a}}^{-1}\boldsymbol{\Sigma}_{\mathrm{a}}(\boldsymbol{G}_{\mathrm{a}}^{-1})^{\mathrm{T}} + \boldsymbol{\Sigma}_{\mathrm{b}}}_{\boldsymbol{\Sigma}_{\Omega}})$$

提案法の導出アウトライン

有声音の信号モデル

↓

ウェーブレット変換
スペクトル表現
（定Q log-normal
フィルタバンクと命名）

未知パラメータ：w, Ω

↓

統計モデル化

$P(Y|w, \Omega)$

w の事前分布の設計
（全極型声道モデルに基づく）

ハイパーパラメータ：π_w

$P(w|\Omega, \pi_w)$

Ω の事前分布の設計
（藤崎モデル（前発表）に基づく）

ハイパーパラメータ：π_Ω

$P(\Omega|\pi_\Omega)$

パラメータ推定

$P(w, \Omega, \pi_w, \pi_\Omega|Y) \propto P(Y|w, \Omega)P(w|\Omega, \pi_w)P(\Omega|\pi_\Omega)$

を局所最大化する反復アルゴリズム

提案法の導出アウトライン

有声音の信号モデル

↓

ウェーブレット変換
スペクトル表現
（定Q log-normal
フィルタバンクと命名）

未知パラメータ：w, Ω

↓

統計モデル化

$P(Y|w,\Omega)$

w の事前分布の設計
（全極型声道モデルに基づく）

ハイパーパラメータ：π_w

$P(w|\Omega,\pi_w)$

Ω の事前分布の設計
（藤崎モデル（前発表）に基づく）

ハイパーパラメータ：π_Ω

$P(\Omega|\pi_\Omega)$

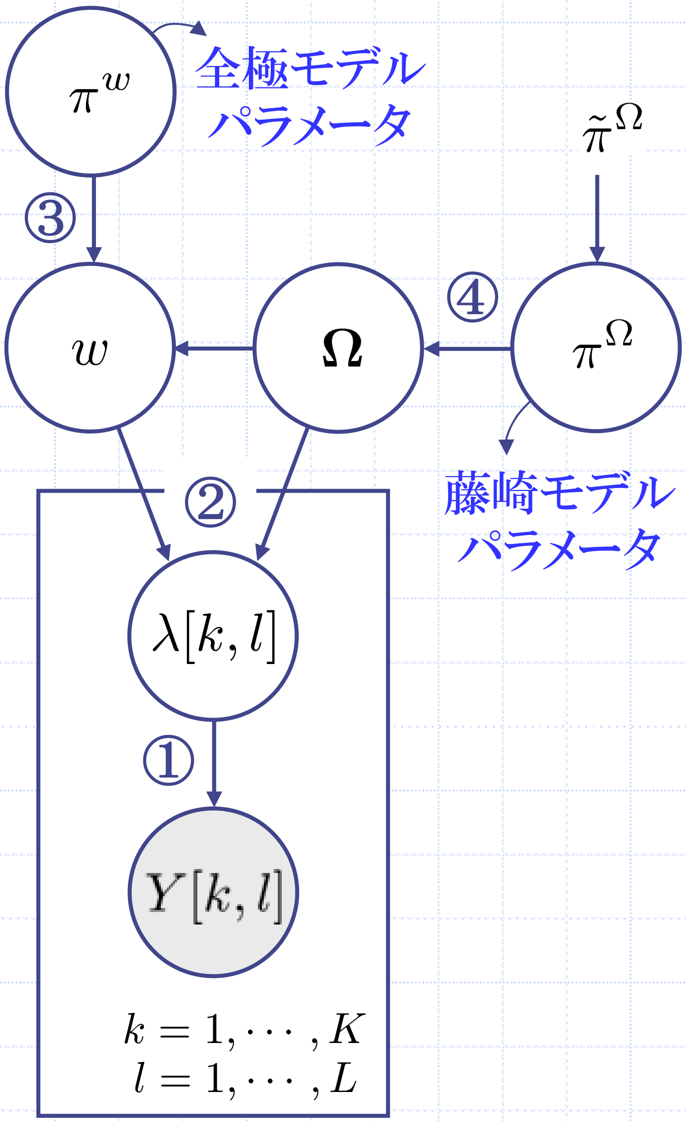
パラメータ推定

$$P(w,\Omega,\pi_w,\pi_\Omega|Y) \propto P(Y|w,\Omega)P(w|\Omega,\pi_w)P(\Omega|\pi_\Omega)$$

を局所最大化する反復アルゴリズム

パラメータ推定(1/2)

- グラフィカルモデル



① $Y[k,l] \sim \mathrm{Pois}(\lambda[k,l])$

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

③ $w_n[l] | \Omega[l] \sim \mathrm{Rayleigh}\left(\frac{1}{|A_l(e^{\mathrm{j}ne^{\Omega[l]}u_0})|}\right)$

④ $\mathbf{\Omega} \sim \mathcal{N}(\boldsymbol{\mu}_\Omega, \mathbf{\Sigma}_\Omega)$

⇩

$\underset{w,\Omega,\pi^w,\pi^\Omega}{\mathrm{argmax}}\ P(w, \Omega, \pi^w, \pi^\Omega | Y)$ を解きたい！

パラメータ推定(2/2)

- 潜在変数：観測スペクトルを構成する各調波成分 $c_n[k,l]$

$$\longrightarrow Y[k,l] = \sum_{n=1}^{N} c_n[k,l]$$

- Poisson分布の性質より

$$c_n[k,l] \sim \mathrm{Pois}(\lambda_n[k,l]) \text{ のとき } Y[k,l] \sim \mathrm{Pois}\Bigl(\underbrace{\sum_{n=1}^{N} \lambda_n[k,l]}^{\lambda[k,l]}\Bigr)$$

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

- EMアルゴリズム

1ステップ前の更新値

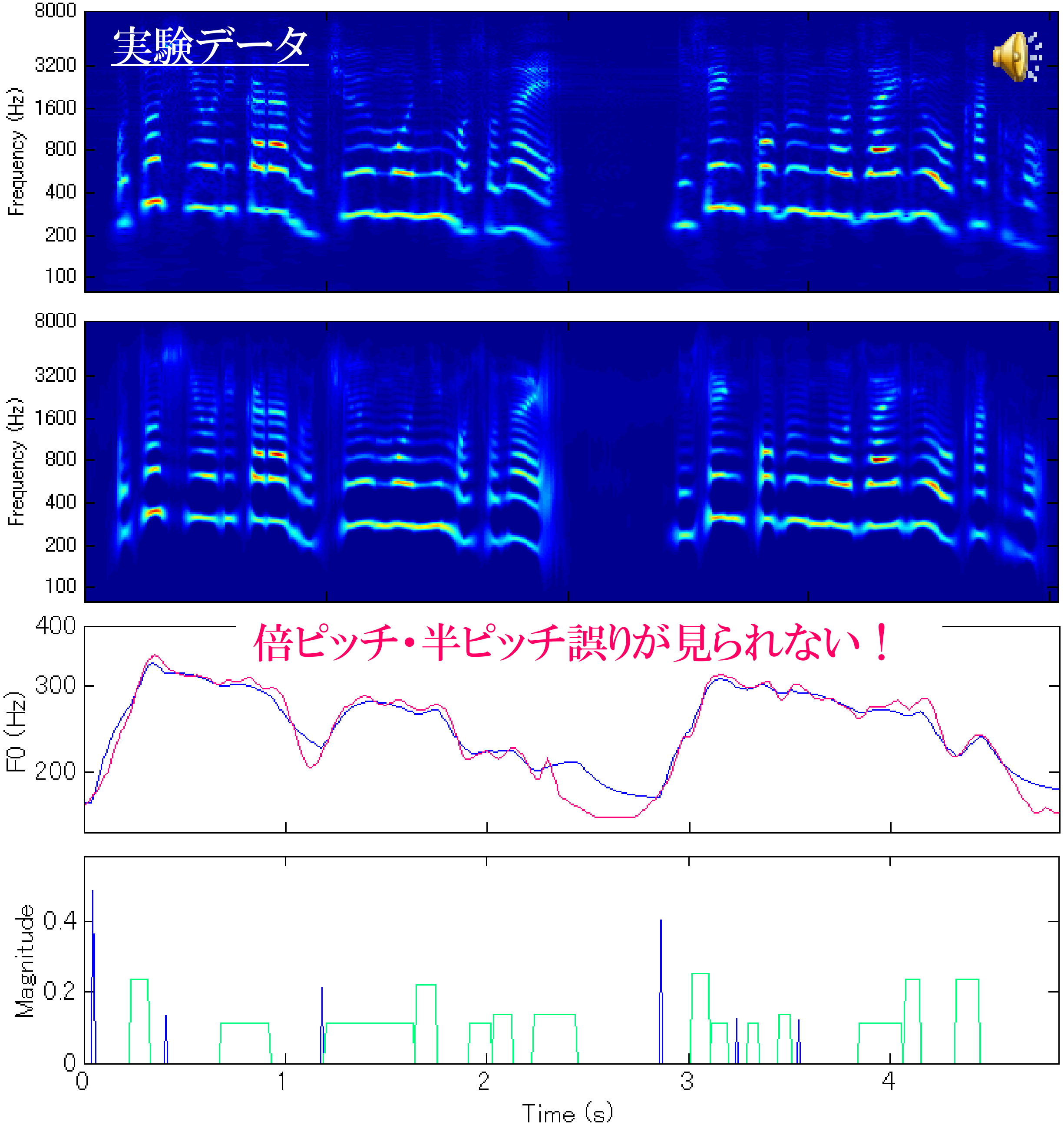
[近似] $P(w|\Omega, \pi_w) \simeq P(w|\Omega^{(\ell-1)}, \pi_w)$

更新式が綺麗に解ける！

$$\longrightarrow \begin{cases} c_n[l]^{(\ell)} = (\text{閉形式}) \\ \mathbf{\Omega}^{(\ell)} = (\text{閉形式}) \\ w_n[l]^{(\ell)} = (\text{閉形式}) \\ \pi_w^{(\ell)} = (\text{反復計算}) \\ \pi_\Omega^{(\ell)} = (\text{反復計算}) \end{cases}$$

詳細は予稿参照

動作実験



定Q log-normal フィルタバンクを用いて分析したスペクトログラム

$Y[k, l]$

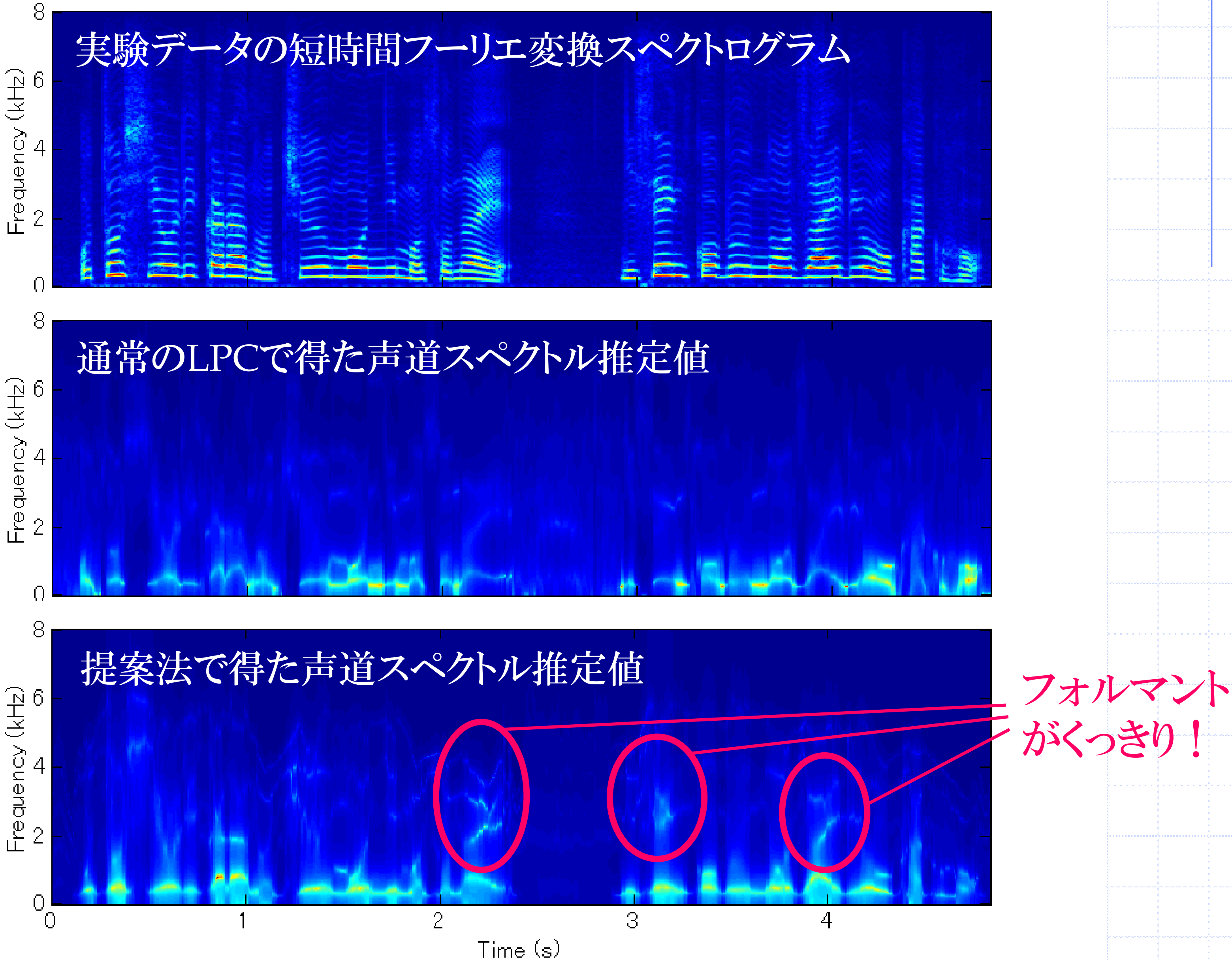
パラメータ反復推定後の局所最適モデル

$\lambda[k, l]$

赤：F0パターン
青：藤崎モデル

青：フレーズ指令
緑：アクセント指令

声道スペクトルの推定結果



まとめ： 全極型声道モデルと藤崎モデルを同時に内包する音声スペクトルモデルを提案

有声音の信号モデル

↓

ウェーブレット変換
スペクトル表現
（定Q log-normal
フィルタバンクと命名）
未知パラメータ：w, Ω

↓

統計モデル化
$P(Y|w,\Omega)$

w の事前分布の設計
（全極型声道モデルに基づく）
ハイパーパラメータ：π_w
$P(w|\Omega,\pi_w)$

Ω の事前分布の設計
（藤崎モデル（前発表）に基づく）
ハイパーパラメータ：π_Ω
$P(\Omega|\pi_\Omega)$

パラメータ推定
$P(w,\Omega,\pi_w,\pi_\Omega|Y) \propto P(Y|w,\Omega)P(w|\Omega,\pi_w)P(\Omega|\pi_\Omega)$
を局所最大化する反復アルゴリズム

- 同時モデル化のメリットを動作実験で確認
- 今後課題：きちんとした性能評価，分析合成系/音声強調系への応用