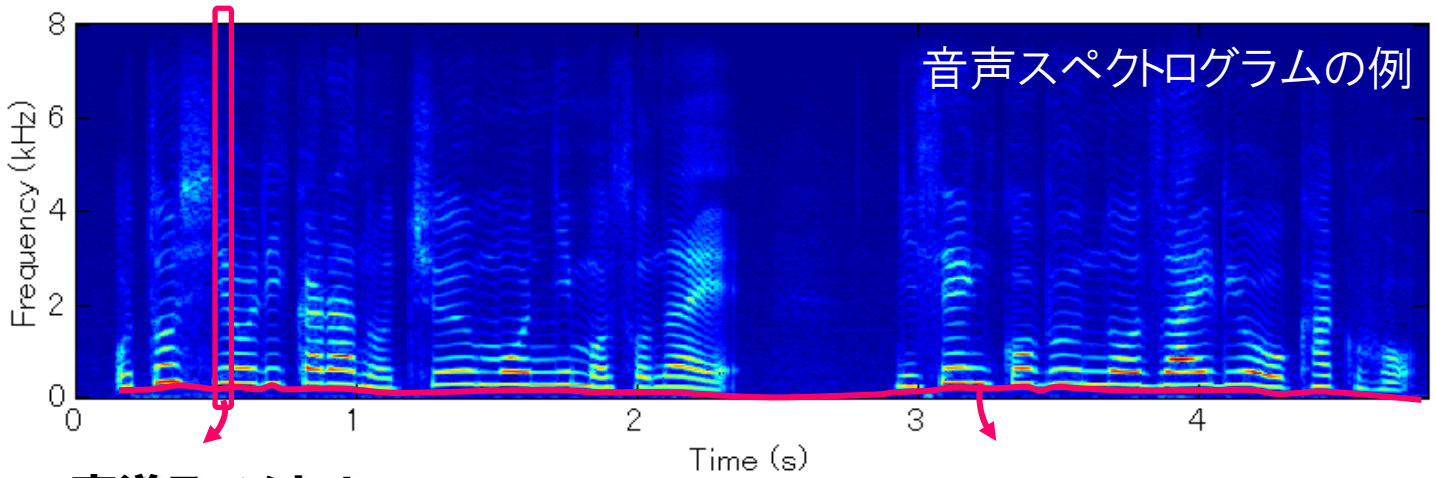
音声生成過程の確率モデル

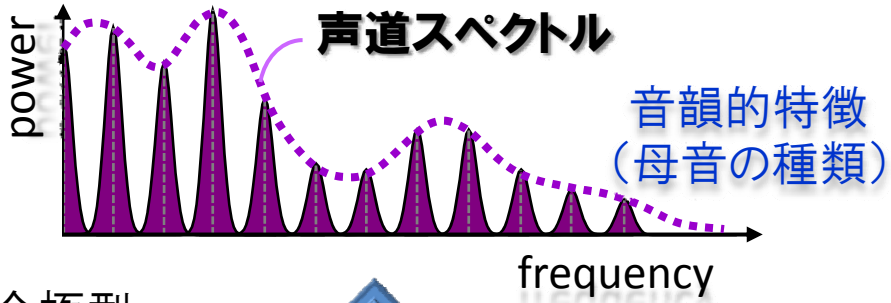
亀岡弘和

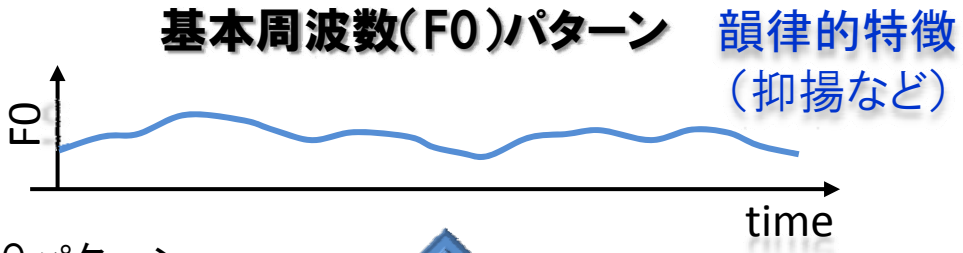
日本電信電話株式会社
NTTコミュニケーション科学基礎研究所

音声スペクトログラムについて

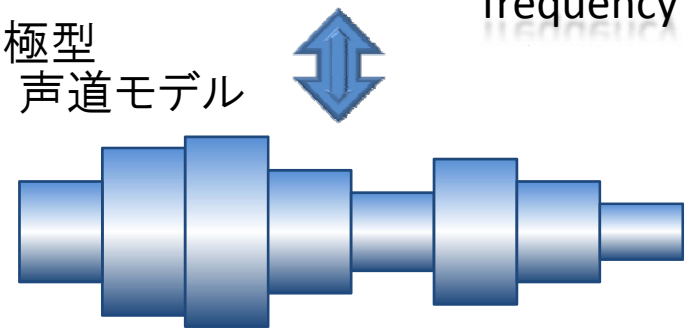
- 発声機構に基づく特有の構造がある





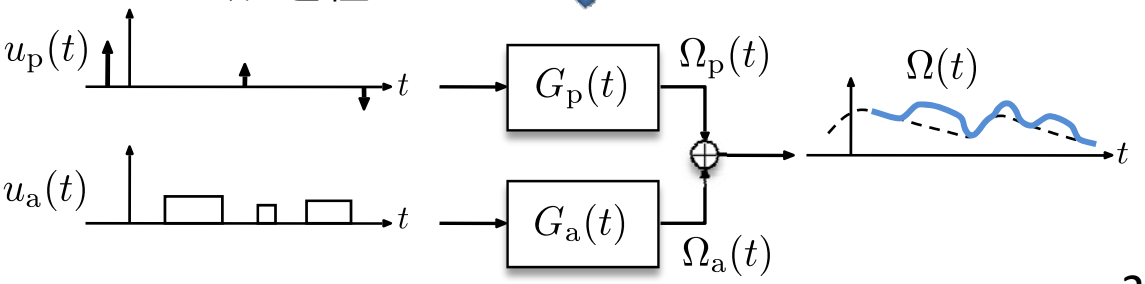


全極型
声道モデル



F0パターン
生成過程モデル

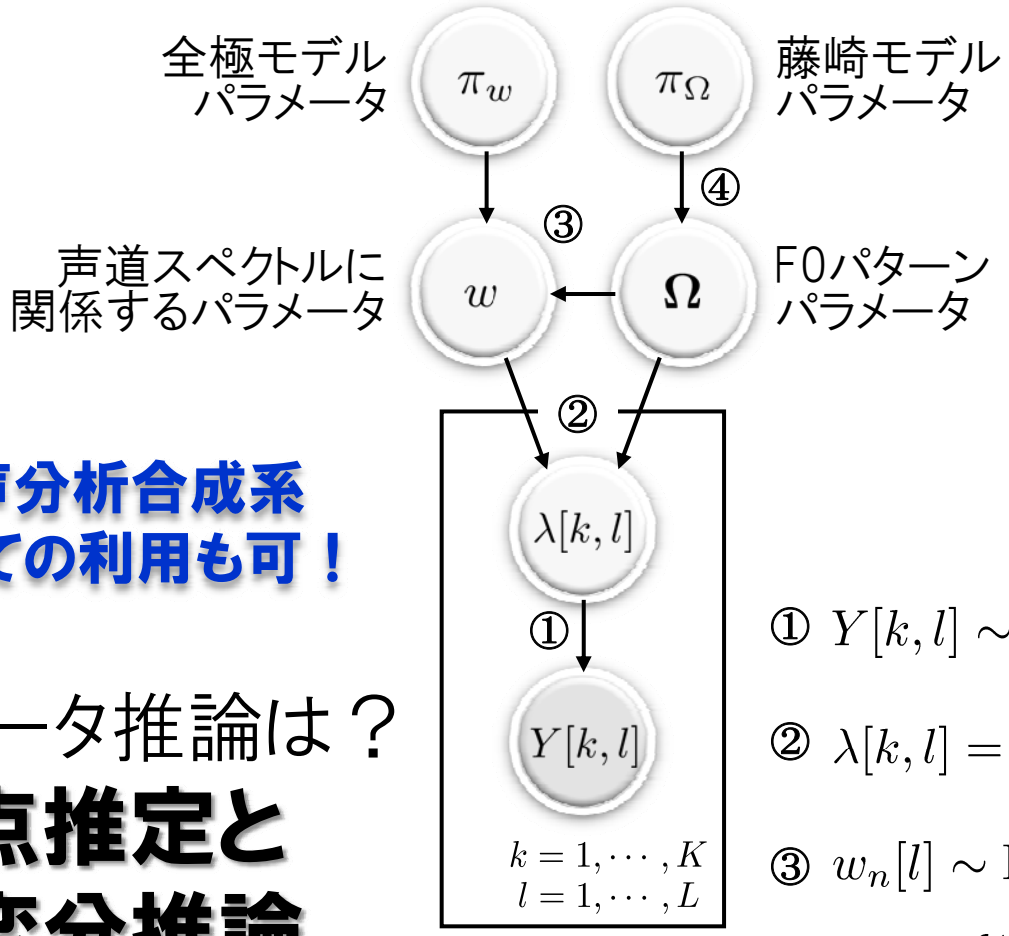
（藤崎モデル）



目的と概要

- 任意の信号が音声らしいかを測るための規準が作れるか？

➡ **音声生成過程の確率モデル化**



音声生成器が生成
しうる信号には高いスコア
しえない信号には低いスコア

音声を対象とした逆問題（雑音除去等）に活用可！

音声分析合成系としての利用も可！

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

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

- パラメータ推論は？

➡ **点推定と変分推論**

提案法の導出アウトライン

有声音の波形モデル
（基本モデル）
未知パラメータ: w, Ω

↓

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）

↓

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

$$P(\pi_w, \pi_\Omega|Y) \propto \iint P(Y|w, \Omega)P(w|\Omega, \pi_w)P(\Omega|\pi_\Omega)\mathrm{d}w\mathrm{d}\boldsymbol{\Omega}$$

の局所最大化を目的とした反復アルゴリズム

波形モデル→ウェーブレットスペクトルモデル

- 時刻 t 周辺の音声信号モデル

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

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）

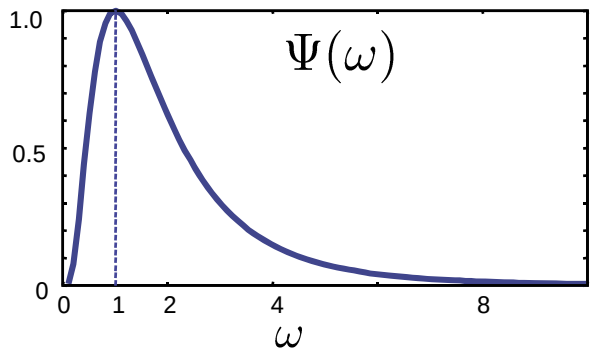
ウェーブレット変換

∵一般化Parsevalの定理

$$S_{\mathrm{w}}(x,t) = \langle s_t, \psi_{x,t} \rangle = \langle \underline{S_t(\omega)}, \Psi(e^{-x}\omega) e^{\mathrm{j}\omega t} \rangle_{\omega \in \mathbb{R}}$$

$s_t(u)$ のFourier変換

時刻 t 周辺に含まれる対数周波数 x の成分を測るための変換



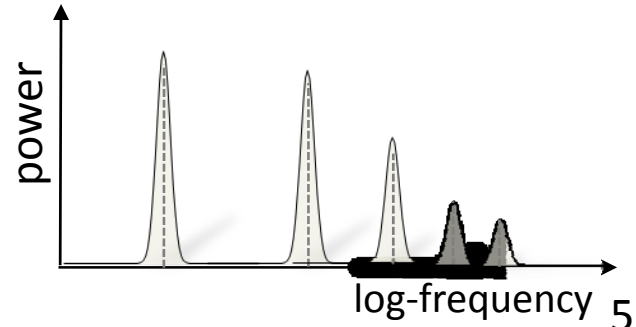
$$\Psi(\omega) = \begin{cases} e^{-\frac{1}{4\sigma^2}(\log \omega)^2} & (\omega > 0) \\ 0 & (\omega \leq 0) \end{cases}$$

「定Q log-normalフィルタバンク」

ウェーブレットスペクトルモデル

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

$\log \mu(t)$ 瞬時**対数**基本周波数



ウェーブレットスペクトルの統計モデル化

- 観測ウェーブレットスペクトル $\boxed{Y[k,l]} := Y(x_k, t_l) \geq 0$

  k: 対数周波数インデックス
  l: 時刻インデックス

- 「現象」$Y[k,l]$ をできるだけ良く説明するスペクトルモデル $\boxed{\lambda[k,l]} := \lambda(x_k, t_l)$ を求めたい

  未知パラメータ: $\underline{w_n[l]} := w_n(t_l)$ $\underline{\Omega[l]} := \Omega(t_l)$

  調波成分の振幅　　F0パターン（対数F0の時間軌跡）

- 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 [Csiszar, 1975]

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

  - $I(\lambda)$ を λ に関して最小化することは,

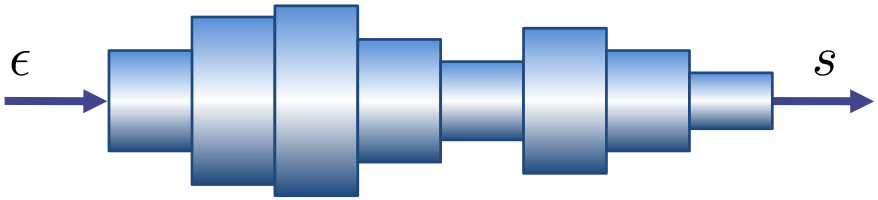
$$\boxed{Y[k,l] \sim \mathrm{Pois}(\lambda[k,l])} \longrightarrow Y[k,l] \text{ の統計モデル化}$$

を仮定した下での λ の最尤推定と等価

全極モデルに基づく w の事前分布設計

- 音声信号モデル

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



全極システムのイメージ（等長音響管モデルと等価）

$$\longrightarrow \quad s_l[i] := s_{t_l}(iu_0) = \sum_{n=1}^{N} w_n[l] e^{\mathrm{j}(ne^{\Omega[l]}(iu_0-t_l)+\xi_n(t_l))}$$

離散時間表現
（周期 u_0 でサンプリング）

- $s_l[i]$ 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 $A_l(z) = a_0 - a_l[1]z^{-1} \cdots - a_l[M]z^{-M}$

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

w が満たすべき制約

全極モデルに基づく w の事前分布設計

- システム入力に確率的な揺らぎをもたせる

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

$$v_{l,n} \sim \mathcal{N}_{\mathbb{C}}(0,1)$$

これにより基本周波数が $e^{\Omega[l]}$ の周期的な雑音 ϵ を表現したことになる

実部と虚部が同じ分散の独立な正規分布に従うなら振幅は？

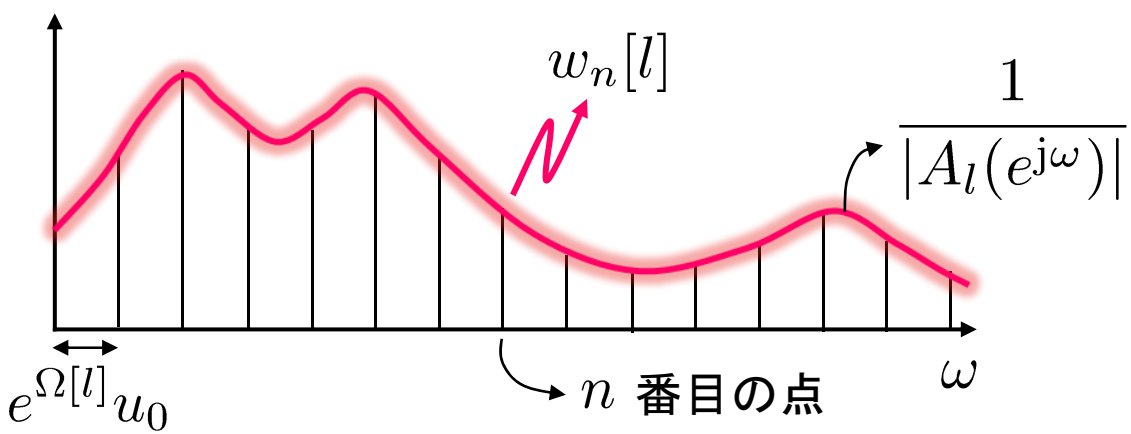
（w と v の関係式）

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

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

ハイパーパラメータ

$$\pi_w = \{a_l[m]\}_{0 \le m \le M, 1 \le l \le L}$$



藤崎モデルに基づくΩの事前分布設計

- **藤崎モデル**[Fujisaki, 1988]**とは**

  声帯振動の制御（F0パターン生成）機構の簡易な物理モデル

  A) 甲状軟骨の運動は二種類（平行移動と回転）あると仮定

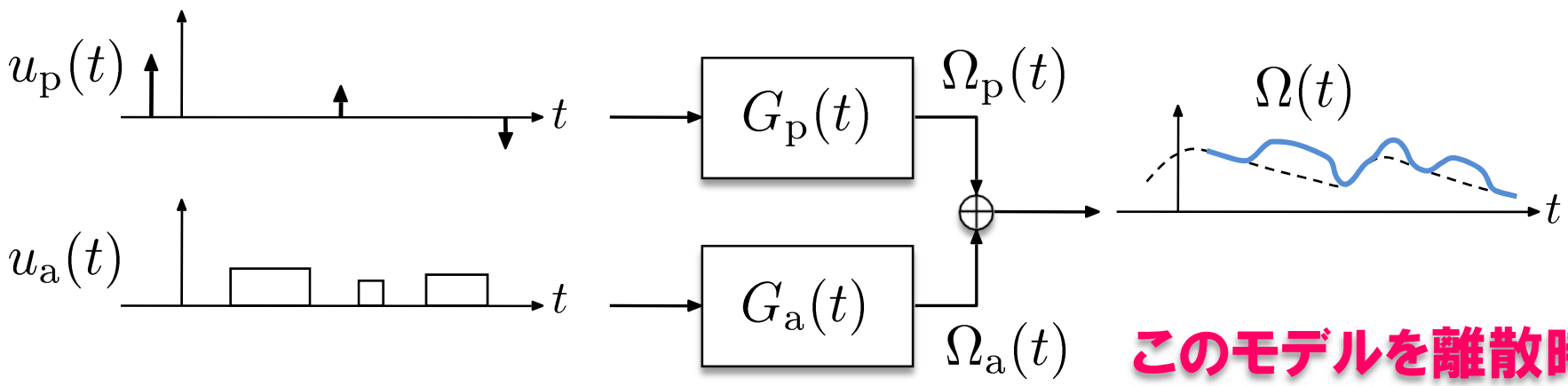
  B) 各運動を臨界制動の二次線形系の応答として表現

  C) 各運動に伴う声帯の伸びの変化の合計が対数F0の変化に比例すると仮定

フレーズ成分 $\Omega_{\mathrm{p}}(t)$：平行移動による変位に比例するF0パターン成分

アクセント成分 $\Omega_{\mathrm{a}}(t)$：回転運動による変位に比例するF0パターン成分

ベースライン成分 Ω_{b}：声帯の物理的性質によって決まる定数

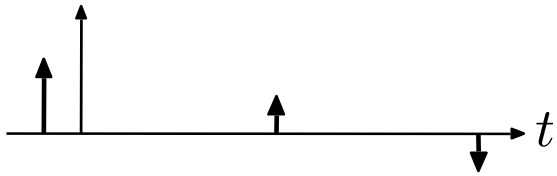


このモデルを離散時間確率過程で表現しよう！

藤崎モデルに基づくΩの事前分布設計

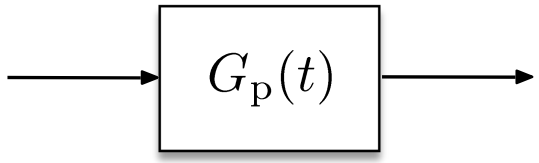
- 連続時間系から離散時間系への変換

フレーズ指令入力 $u_{\mathrm{p}}(t)$



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

→ 臨界制動系のインパルス応答



$\Omega_{\mathrm{p}}(t)$
フレーズ成分

– フレーズ制御系のラプラス変換 $G_{\mathrm{p}}(s) = \alpha^2/(s+\alpha)^2$

– 逆システムのラプラス変換 $G_{\mathrm{p}}^{-1}(s) = (s+\alpha)^2/\alpha^2$

– 後退差分近似 $s = \dfrac{1 - z^{-1}}{t_0}$ → 離散時間表現のサンプリング周期

– 逆システムのz変換 $G^{-1}(z) = \dfrac{z^{-2} - 2(\alpha t_0 + 1)z^{-1} + (\alpha t_0 + 1)^2}{\alpha^2 t_0^2}$

$$u_{\mathrm{p}}[l] = g_0^{\mathrm{p}}\Omega_{\mathrm{p}}[l] + g_1^{\mathrm{p}}\Omega_{\mathrm{p}}[l-1] + g_2^{\mathrm{p}}\Omega_{\mathrm{p}}[l-2]$$

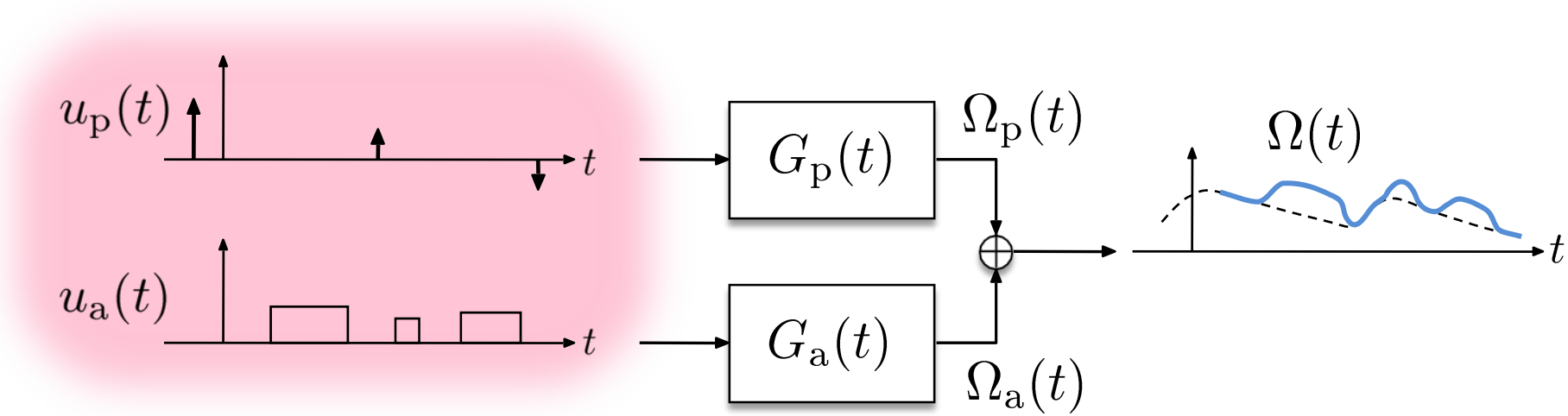
→ アクセント成分についても同様

$g_0^{\mathrm{p}} = (\psi - 1)^2, \quad g_1^{\mathrm{p}} = -2\psi(\psi - 1), \quad g_2^{\mathrm{p}} = \psi^2, \quad \psi = 1 + 1/(\alpha t_0)$

パラメータ ψ で特性が決まる**拘束つき全極モデル**

藤崎モデルに基づくΩの事前分布設計

- 両「指令」関数は無制約で良いわけではない



- 藤崎モデルにおける基本ルール
  - フレーズ指令はパルス波
  - アクセント指令は矩形波
  - 同時刻に二つ以上の指令は発生しない

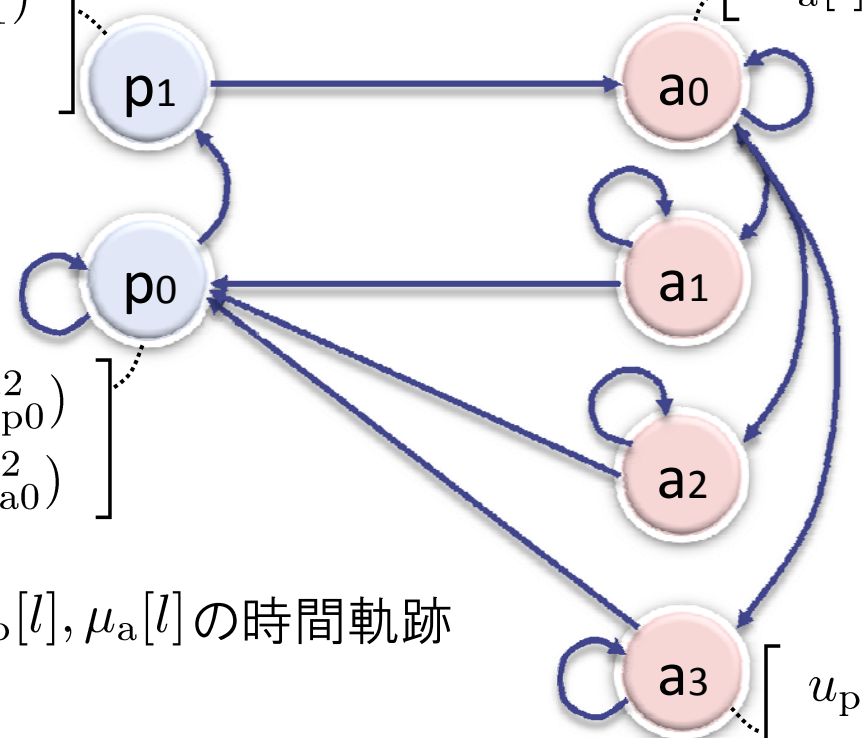
藤崎モデルに基づくΩの事前分布設計

- 隠れマルコフモデルによるフレーズ・アクセント指令の確率モデル化

自己遷移させないことにより，パルス波を生成可能に

$$\begin{bmatrix} u_{\mathrm{p}}[l] \sim \mathcal{N}(A_{\mathrm{p}}[l], \sigma_{\mathrm{p1}}^2) \\ u_{\mathrm{a}}[l] \sim \mathcal{N}(0, \sigma_{\mathrm{a0}}^2) \end{bmatrix}$$

$$\begin{bmatrix} u_{\mathrm{p}}[l] \sim \mathcal{N}(0, \sigma_{\mathrm{p0}}^2) \\ u_{\mathrm{a}}[l] \sim \mathcal{N}(0, \sigma_{\mathrm{a0}}^2) \end{bmatrix}$$



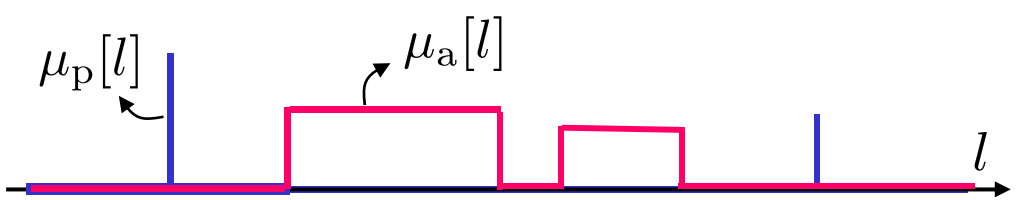
$$\begin{bmatrix} u_{\mathrm{p}}[l] \sim \mathcal{N}(0, \sigma_{\mathrm{p0}}^2) \\ u_{\mathrm{a}}[l] \sim \mathcal{N}(0, \sigma_{\mathrm{a0}}^2) \end{bmatrix}$$

a_nからa_m ($n \neq m$, $n \neq 0$, $m \neq 0$)に遷移できないようにすることで矩形波を表現可能に

$$\begin{bmatrix} u_{\mathrm{p}}[l] \sim \mathcal{N}(0, \sigma_{\mathrm{p0}}^2) \\ u_{\mathrm{a}}[l] \sim \mathcal{N}(A_{\mathrm{a}}^{(3)}, \sigma_{\mathrm{a3}}^2) \end{bmatrix}$$

簡単のため遷移確率は定数とする

状態遷移に伴う平均 $\mu_{\mathrm{p}}[l], \mu_{\mathrm{a}}[l]$ の時間軌跡



藤崎モデルに基づくΩの事前分布設計

- フレーズ成分 $\boldsymbol{\Omega}_{\mathrm{p}} = (\Omega_{\mathrm{p}}[1], \cdots, \Omega_{\mathrm{p}}[T])^{\mathrm{T}}$ の確率密度関数

  p. 10で得た拘束式 $u_{\mathrm{p}}[l] = g_0^{\mathrm{p}}\Omega_{\mathrm{p}}[l] + g_1^{\mathrm{p}}\Omega_{\mathrm{p}}[l-1] + g_2^{\mathrm{p}}\Omega_{\mathrm{p}}[l-2]$

  p. 12で得た確率モデル $u_{\mathrm{p}}[l] \sim \mathcal{N}(\mu_{\mathrm{p}}[l], \sigma_{\mathrm{p}}[l]^2)$

  状態系列によって決まる平均と分散の遷移系列

$$
\overbrace{\begin{bmatrix} g_0^{\mathrm{p}} & & & & 0 \\ g_1^{\mathrm{p}} & g_0^{\mathrm{p}} & & & \\ g_2^{\mathrm{p}} & g_1^{\mathrm{p}} & g_0^{\mathrm{p}} & & \\ & \ddots & \ddots & \ddots & \\ 0 & & g_2^{\mathrm{p}} & g_1^{\mathrm{p}} & g_0^{\mathrm{p}} \end{bmatrix}}^{\boldsymbol{G}_{\mathrm{p}}}
\overbrace{\begin{bmatrix} \Omega_{\mathrm{p}}[1] \\ \Omega_{\mathrm{p}}[2] \\ \Omega_{\mathrm{p}}[3] \\ \vdots \\ \Omega_{\mathrm{p}}[L] \end{bmatrix}}^{\boldsymbol{\Omega}_{\mathrm{p}}}
=
\overbrace{\begin{bmatrix} u_{\mathrm{p}}[1] \\ u_{\mathrm{p}}[2] \\ u_{\mathrm{p}}[3] \\ \vdots \\ u_{\mathrm{p}}[T] \end{bmatrix}}^{\boldsymbol{u}_{\mathrm{p}}}
\sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{p}}, \boldsymbol{\Sigma}_{\mathrm{p}})
$$

対角成分が$\sigma_{\mathrm{p}}[l]^2$の対角行列

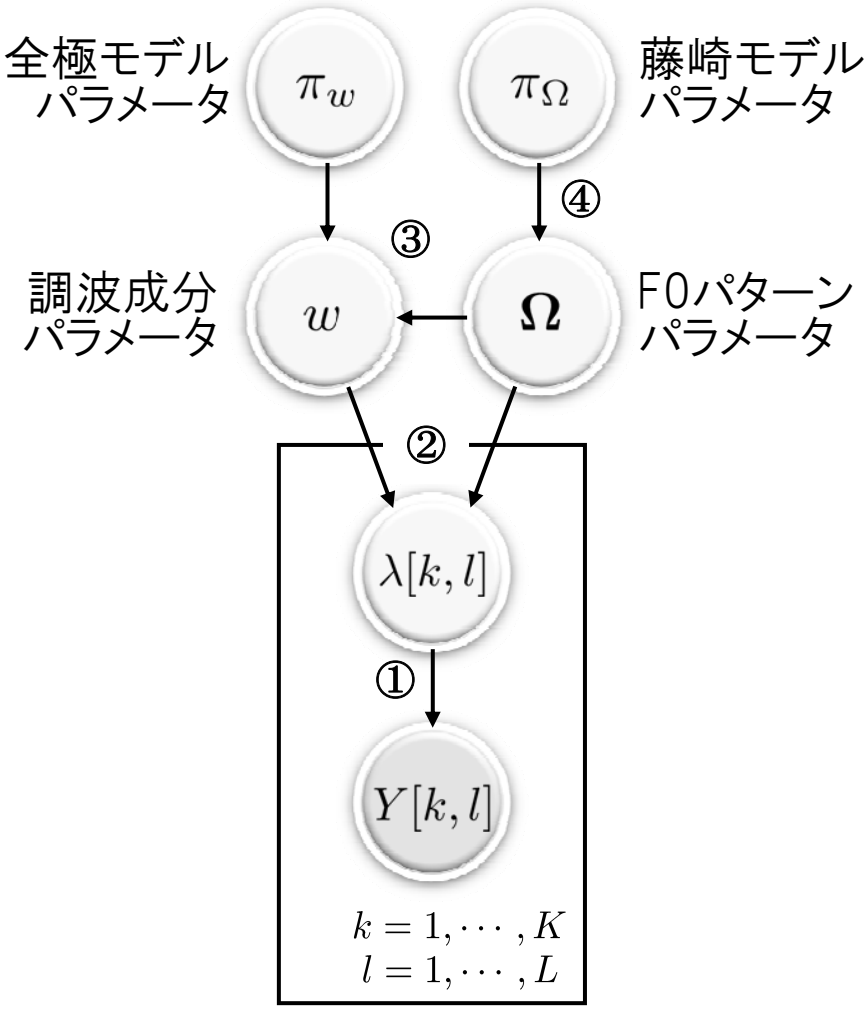
$$
\therefore \boldsymbol{\Omega}_{\mathrm{p}} \sim \mathcal{N}(\boldsymbol{G}_{\mathrm{p}}^{-1}\boldsymbol{\mu}_{\mathrm{p}}, \boldsymbol{G}_{\mathrm{p}}^{-1}\boldsymbol{\Sigma}_{\mathrm{p}}(\boldsymbol{G}_{\mathrm{p}}^{-1})^{\mathrm{T}})
$$

- アクセント成分も同様 $\boldsymbol{\Omega}_{\mathrm{a}} \sim \mathcal{N}(\boldsymbol{G}_{\mathrm{a}}^{-1}\boldsymbol{\mu}_{\mathrm{a}}, \boldsymbol{G}_{\mathrm{a}}^{-1}\boldsymbol{\Sigma}_{\mathrm{a}}(\boldsymbol{G}_{\mathrm{a}}^{-1})^{\mathrm{T}})$
- F0パターンベクトル $\boldsymbol{\Omega} = \boldsymbol{\Omega}_{\mathrm{p}} + \boldsymbol{\Omega}_{\mathrm{a}} + \boldsymbol{\Omega}_{\mathrm{b}}$

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

パラメータ推論その1

- 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 $w_n[l] \sim \mathrm{Rayleigh}(\hat{P}_n[l])$

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

$$\underset{w,\Omega,\pi_w,\pi_\Omega}{\mathrm{argmax}} \ \log P(w, \Omega, \pi_w, \pi_\Omega | Y)$$

$$= \underset{w,\Omega,\pi_w,\pi_\Omega}{\mathrm{argmax}} \ \underline{\log P(Y|w,\Omega)} P(w|\Omega,\pi_w) P(\Omega|\pi_\Omega)$$

log sumの形を含む

点推定アルゴリズム (1/3)

- EMアルゴリズム
  - 完全データ(観測データ＆潜在データ)の対数確率
  →観測データとパラメータが与えられた下での期待値＝Q関数
  - Q関数からEステップとMステップを導出

- 潜在データ:観測スペクトルを構成する各調波成分 $c_n[k,l]$

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

- Poisson分布の再生性より

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

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

$$P(Y, c|w, \Omega) = P(c|w, \Omega)P(Y|c, w, \Omega)$$ ・・・完全データの確率密度関数

点推定アルゴリズム (2/3)

- 完全データ(観測データ＆潜在データ)の対数確率：

$$P(c|w,\Omega) = \prod_n \prod_{k,l} \frac{\lambda_n[k,l]^{c_n[k,l]} e^{-\lambda_n[k,l]}}{c_n[k,l]!}$$

$$P(Y|c,\cancel{w,\Omega}) = \prod_{k,l} \delta\left(Y[k,l] - \sum_n c_n[k,l]\right)$$

$$\longrightarrow \quad \log P(Y,c|w,\Omega) = \log P(c|w,\Omega) + \log P(Y|c)$$

$$\stackrel{c}{=} \sum_n \sum_{k,l} c_n[k,l] \log \lambda_n[k,l] - \lambda_n[k,l]$$

- 観測データとモデルパラメータが与えられた下での期待値：

$$Q(w,\Omega|w',\Omega') = \sum_n \sum_{k,l} \langle c_n[k,l] \rangle_{P(c_n[k,l]|Y[k,l],w',\Omega')} \log \lambda_n[k,l] - \lambda_n[k,l]$$

Poisson分布の性質より

$$c_n[k,l] \big| Y[k,l] \sim \mathrm{Binom}\left(Y[k,l], \frac{\lambda'_n[k,l]}{\sum_m \lambda'_m[k,l]}\right)$$

$$\frac{Y[k,l]\lambda'_n[k,l]}{\sum_m \lambda'_m[k,l]}$$

Q関数の完成！

点推定アルゴリズム (3/3)

- Q関数からEステップとMステップを導出

$$R := \sum_n \sum_{k,l} \frac{Y[k,l]\lambda'_n[k,l]}{\sum_m \lambda'_m[k,l]} \log \lambda_n[k,l] - \lambda_n[k,l]$$
$$+ \log P(w|\Omega, \pi_w) + \log P(\Omega|\pi_\Omega)$$

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

– **Eステップ**: λ' にMステップで更新した w と Ω を代入

– **Mステップ**: R を w と Ω に関して最大化

全極モデルパラメータ

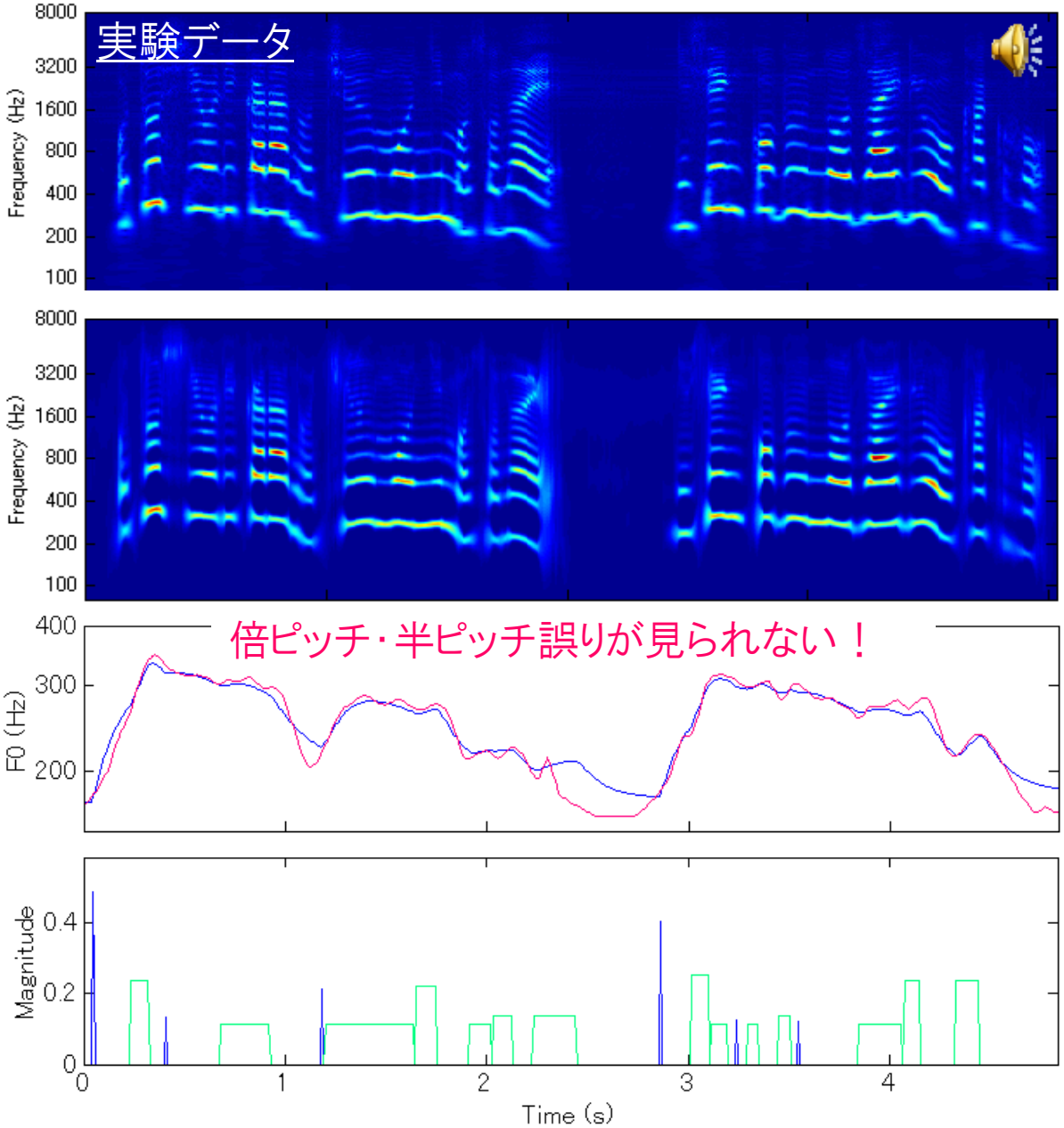
R を π_w と π_Ω に関して最大化

藤崎モデルパラメータ

1ステップ前の更新値

[近似] $P(w|\Omega, \pi_w) \simeq P(w|\Omega', \pi_w)$

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

$$Y[k,l]$$

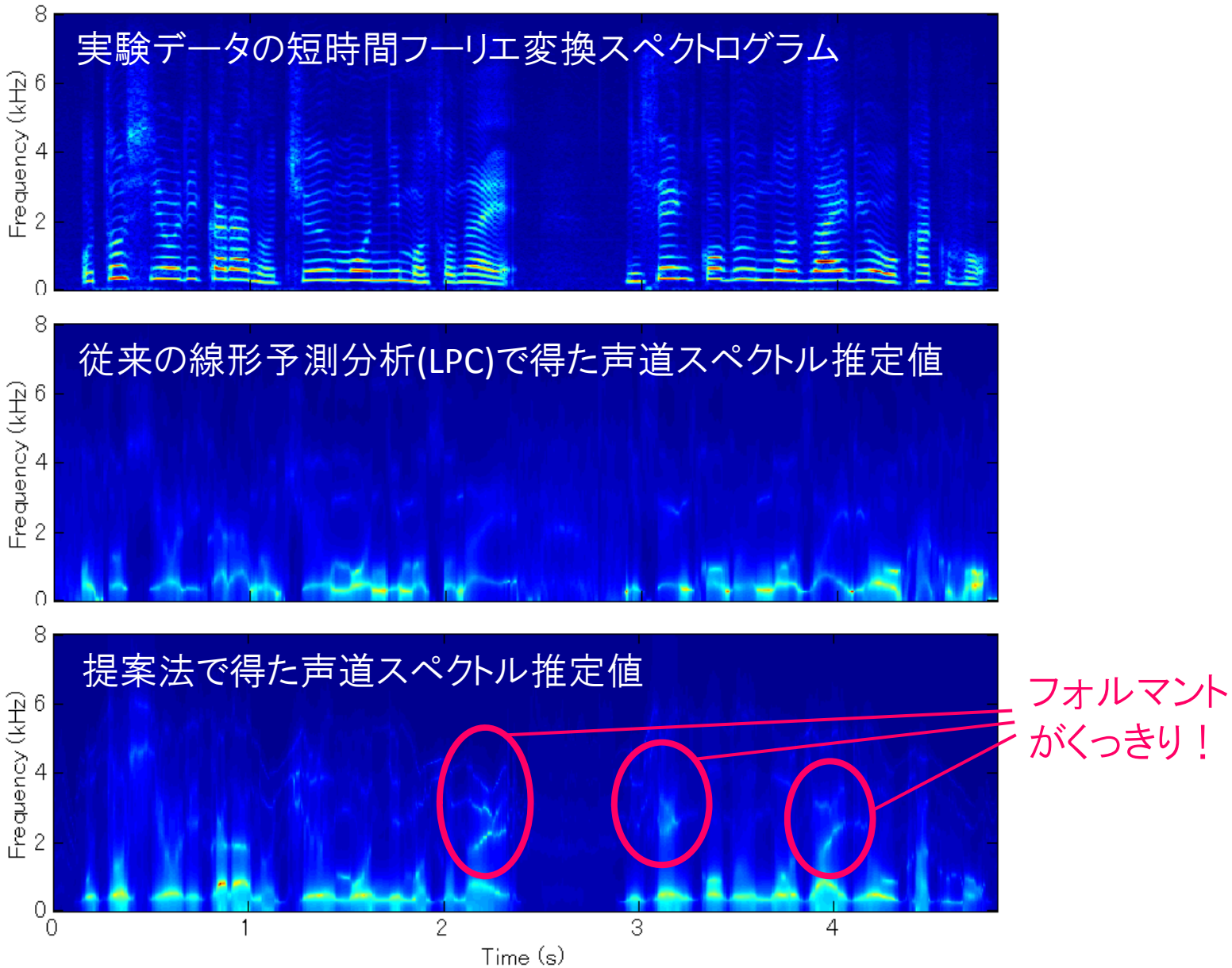
パラメータ反復推定後の局所最適モデル

$$\lambda[k,l]$$

赤：F0パターン
青：藤崎モデル

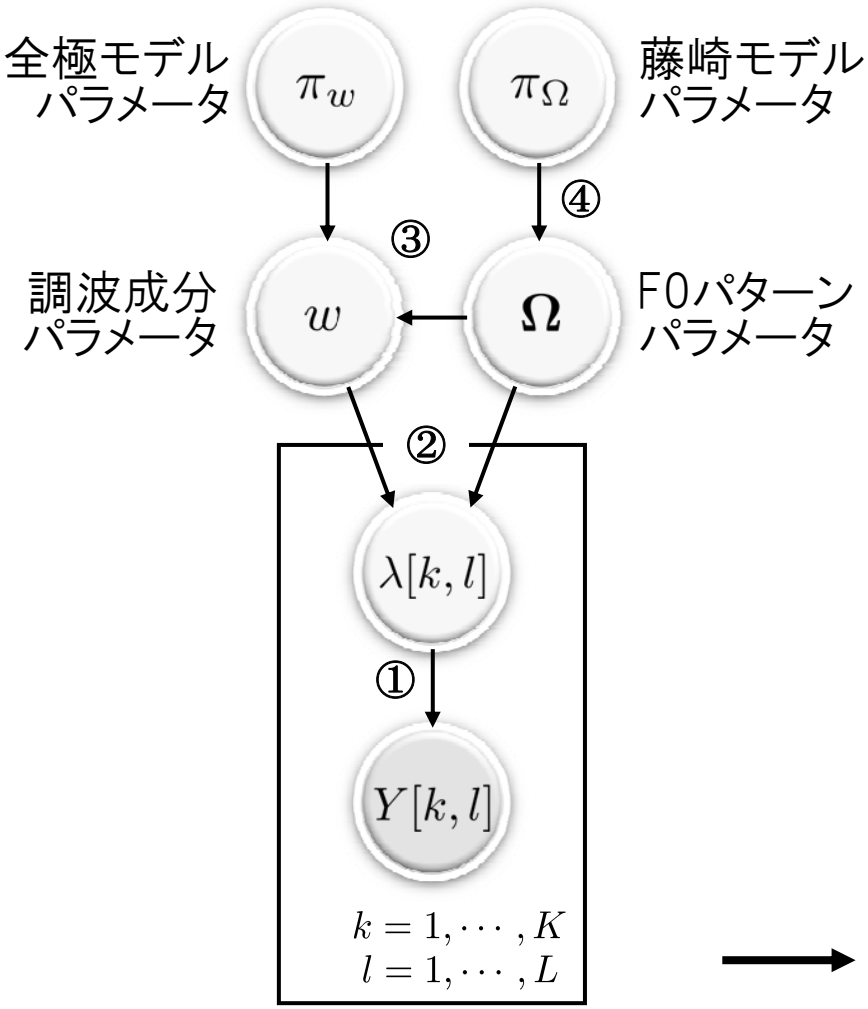
青：フレーズ指令
緑：アクセント指令

声道スペクトルの推定結果



パラメータ推論その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 $w_n[l] \sim \mathrm{Rayleigh}(\hat{P}_n[l])$

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

$\underset{\pi_w, \pi_\Omega}{\mathrm{argmax}} \log P(\pi_w, \pi_\Omega | Y)$ を解きたい！

w と $\boldsymbol{\Omega}$ を周辺化したい
けど今のままでは難しい

→ w と $\boldsymbol{\Omega}$ の変分事後分布推定

$P(w|\boldsymbol{\Omega}, \pi_w) \simeq P(w|\langle \boldsymbol{\Omega} \rangle_{q(\boldsymbol{\Omega})}, \pi_w)$ と近似仮定

変分事後分布推定 (1/6)

- 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分布の再生性より

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

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

- 目標：$P(c, w, \Omega | Y, \pi^w, \pi^\Omega)$ を近似すること！

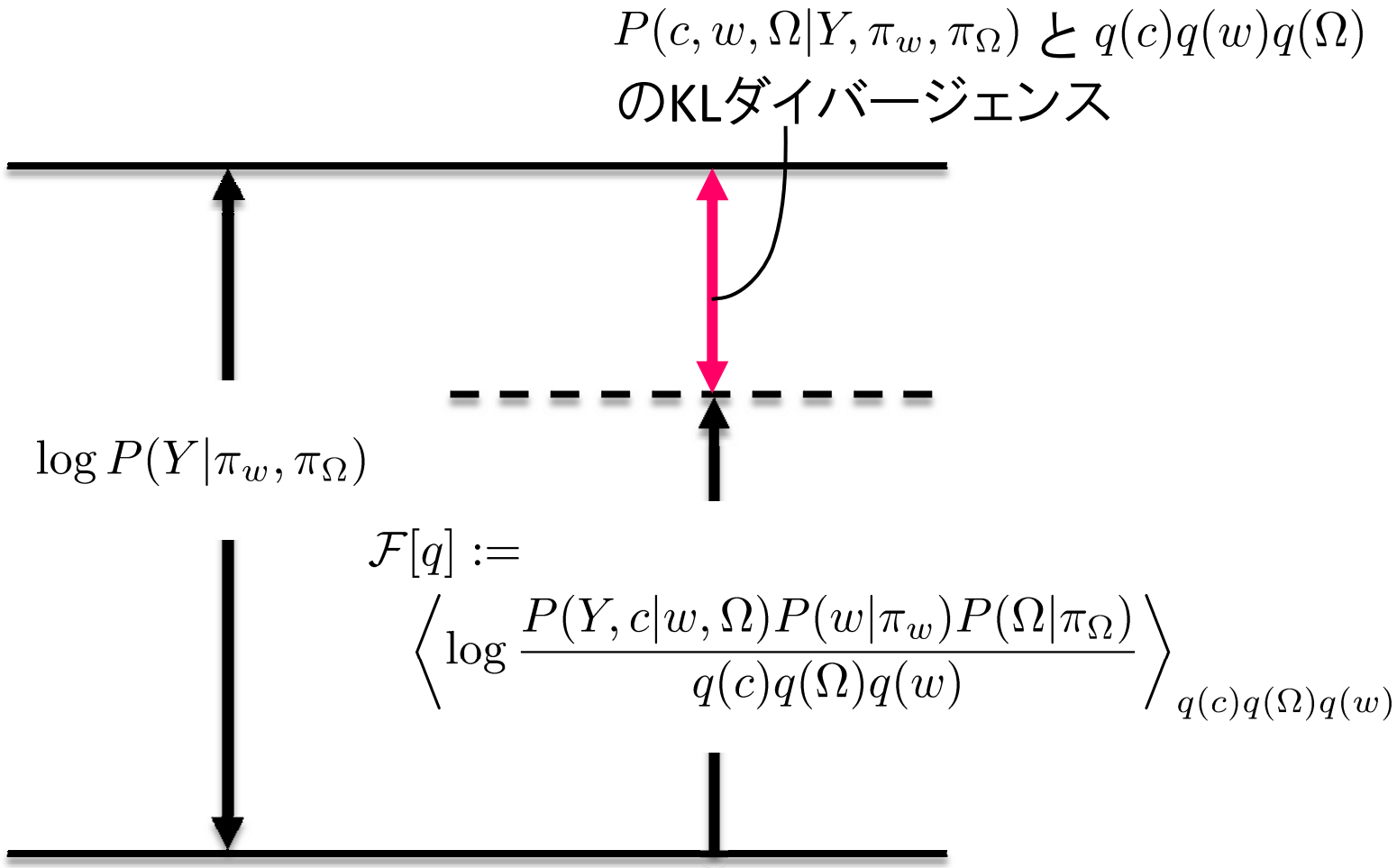
$$\log P(Y|\pi_w, \pi_\Omega) = \log \int_c \int_w \int_\Omega P(Y, c|w, \Omega) P(w|\pi_w) P(\Omega|\pi_\Omega) \mathrm{d}c\mathrm{d}w\mathrm{d}\Omega$$

$$\geq \int_c \int_w \int_\Omega q(c)q(\Omega)q(w) \log \frac{P(Y, c|w, \Omega) P(w|\pi_w) P(\Omega|\pi_\Omega)}{q(c)q(\Omega)q(w)} \mathrm{d}c\mathrm{d}w\mathrm{d}\Omega$$

$$=: \mathcal{F}[q]$$

$\log P(Y|\pi_w, \pi_\Omega) - \mathcal{F}[q]$ は $P(c, w, \Omega | Y, \pi_w, \pi_\Omega)$ と $q(c)q(w)q(\Omega)$ のKLダイバージェンスになっている！

変分事後分布推定 (2/6)


$P(c, w, \Omega | Y, \pi_w, \pi_\Omega)$ と $q(c)q(w)q(\Omega)$
のKLダイバージェンス
$\log P(Y|\pi_w, \pi_\Omega)$
$\mathcal{F}[q] :=$
$\left\langle \log \frac{P(Y, c|w, \Omega)P(w|\pi_w)P(\Omega|\pi_\Omega)}{q(c)q(\Omega)q(w)} \right\rangle_{q(c)q(\Omega)q(w)}$

変分事後分布推定 (3/6)

- $\mathcal{F}[q]$の反復最大化

$$q(c) \propto \exp\left(\left\langle \log P(Y,c|w,\Omega)P(w|\pi^w)P(\Omega|\pi^\Omega)\right\rangle_{q(w)q(\Omega)}\right) \quad \cdots(1)$$

$$q(w) \propto \exp\left(\left\langle \log P(Y,c|w,\Omega)P(w|\pi^w)P(\Omega|\pi^\Omega)\right\rangle_{q(c)q(\Omega)}\right) \quad \cdots(2)$$

$$q(\Omega) \propto \exp\left(\left\langle \log P(Y,c|w,\Omega)P(w|\pi^w)P(\Omega|\pi^\Omega)\right\rangle_{q(c)q(w)}\right) \quad \cdots(3)$$

$$P(Y,c|w,\Omega) = \underline{P(Y|c)}\,\underline{P(c|w,\Omega)}$$

$$\prod_{k,l}\delta\left(Y[k,l] - \sum_n c_n[k,l]\right) \qquad \prod_n\prod_{k,l}\frac{\lambda_n[k,l]^{c_n[k,l]}e^{-\lambda_n[k,l]}}{c_n[k,l]!}$$

変分事後分布推定 (4/6)

（1）潜在変数 $c_n[k,l]$ の変分事後分布更新式

$$q(c_1[k,l],\cdots,c_N[k,l]) \propto \delta\left(Y[k,l]-\sum_n c_n[k,l]\right)$$

$$\exp\left(\sum_n \left\{c_n[k,l]\langle \log\lambda_n[k,l]\rangle_{q(w)q(\Omega)} - \log(c_n[k,l]!)\right\}\right)$$

$$= \mathrm{Multinomial}(c_1[k,l],\cdots,c_N[k,l];Y[k,l],\underline{p_1[k,l],\cdots,p_N[k,l]})$$

$$\underline{p_n[k,l]} = \frac{\exp\underline{\langle\log\lambda_n[k,l]\rangle_{q(w)q(\Omega)}}}{\sum_{n'}\exp\underline{\langle\log\lambda_{n'}[k,l]\rangle_{q(w)q(\Omega)}}}$$

（2）と（3）で計算される

$$\underline{\langle\log\lambda_n[k,l]\rangle_{q(w)q(\Omega)}} = \underline{\langle\log w_n[k,l]^2\rangle_{q(w)}} - \frac{1}{2}\log 2\pi\sigma^2$$

$$-\frac{1}{2\sigma^2}\left((x_k-\log n)^2 - 2(x_k-\log n)\underline{\langle\Omega[l]\rangle_{q(\Omega)}} + \underline{\langle\Omega[l]^2\rangle_{q(\Omega)}}\right)$$

→ $\langle c_n[k,l]\rangle_{q(c)} = Y[k,l]p_n[k,l]$ （他の更新式で使う）

変分事後分布推定 (5/6)

（2）調波成分パワー $w_n[l]$ の変分事後分布更新式

$$q(w_n[l]) \propto \exp\left(\sum_k (\langle c_n[k,l]\rangle_{q(c)} 2\log w_n[l]) - w_n[l]^2\right) \underbrace{\frac{w_n[l]}{\hat{P}_n[l]} e^{-\frac{w_n[l]^2}{2\hat{P}_n[l]}}}_{\text{事前分布}}$$

$$= \mathrm{Nakagami}(w_n[l]; \xi_n[l], \eta_n[l])$$

（1）で計算される

$$\xi_n[l] = 1 + \sum_k \langle c_n[k,l]\rangle_{q(c)}$$

$$\eta_n[l] = \frac{2\hat{P}_n[l]}{2\hat{P}_n[l]+1}\left(1 + \sum_k \langle c_n[k,l]\rangle_{q(c)}\right)$$

$$\langle \log w_n[l]^2\rangle_{q(w)} = \Psi(\xi_n[l]) - \log(\xi_n[l]/\eta_n[l])$$ （他の更新式で使う）

Digamma関数

変分事後分布推定 (6/6)

（3）F0パターン $\boldsymbol{\Omega}$ の変分事後分布更新式

$$q(\boldsymbol{\Omega}) \propto \exp\left(-\frac{1}{2}\sum_n\sum_{k,l}\frac{\langle c_n[k,l]\rangle_{q(c)}}{\sigma^2}(\Omega[l]-(x_k-\log n))^2\right)$$

$$\underline{\exp\left(-\frac{1}{2}(\boldsymbol{\Omega}-\boldsymbol{\mu}_\Omega)^{\mathrm{T}}\boldsymbol{\Sigma}_\Omega^{-1}(\boldsymbol{\Omega}-\boldsymbol{\mu}_\Omega)\right)}$$

事前分布

$$= \mathcal{N}(\boldsymbol{\Omega}; (\underline{\boldsymbol{\Sigma}}+\boldsymbol{\Sigma}_\Omega)^{-1}(\underline{\boldsymbol{\Sigma}}\underline{\hat{\boldsymbol{\Omega}}}+\boldsymbol{\Sigma}_\Omega\boldsymbol{\mu}_\Omega), \underline{\boldsymbol{\Sigma}}+\boldsymbol{\Sigma}_\Omega)$$

$$\underline{\hat{\boldsymbol{\Omega}}} = (\hat{\Omega}_1, \cdots, \hat{\Omega}_L)^{\mathrm{T}}$$

（1）で計算される

$$\hat{\Omega}_l = \frac{\sum_n\sum_k \underline{\langle c_n[k,l]\rangle_{q(c)}}(x_k-\log n)}{\sum_n\sum_k \underline{\langle c_n[k,l]\rangle_{q(c)}}}$$

$$\underline{\boldsymbol{\Sigma}} = \mathrm{diag}(\Sigma_1, \cdots, \Sigma_l) \qquad \Sigma_l = \sigma^2 \Big/ \sum_n\sum_k \underline{\langle c_n[k,l]\rangle_{q(c)}}$$

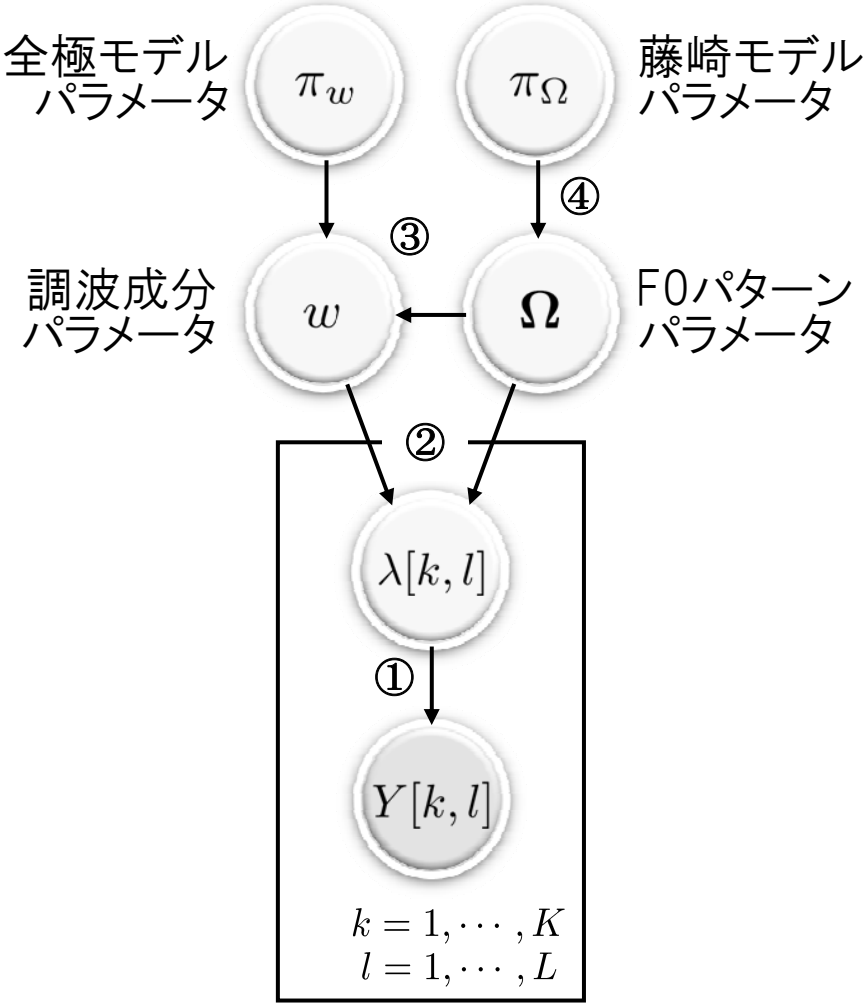
$$\langle\Omega[l]\rangle_{q(\Omega)} = \left[(\boldsymbol{\Sigma}+\boldsymbol{\Sigma}_\Omega)^{-1}(\boldsymbol{\Sigma}\hat{\boldsymbol{\Omega}}+\boldsymbol{\Sigma}_\Omega\boldsymbol{\mu}_\Omega)\right]_l$$

$$\langle\Omega[l]^2\rangle_{q(\Omega)} = \left[\boldsymbol{\Sigma}+\boldsymbol{\Sigma}_\Omega\right]_{l,l} + (\langle\Omega[l]\rangle_{q(\Omega)})^2$$

（他の更新式で使う）

ハイパーパラメータの推定

- グラフィカルモデル



q の反復計算後 $\log P(Y|\pi_w, \pi_\Omega) \simeq \mathcal{F}[q]$
となっていることを利用

$$\underset{\pi_w, \pi_\Omega}{\operatorname{argmax}} \log P(Y|\pi_w, \pi_\Omega) P(\pi_w, \pi_\Omega)$$
$$\simeq \underset{\pi_w, \pi_\Omega}{\operatorname{argmax}} \underline{\mathcal{F}[q]} + \log P(\pi_w, \pi_\Omega)$$

これが達成されたら $\log P(Y|\pi_w, \pi_\Omega)$
の値が変化するため $q(c)q(w)q(\Omega)$ を
再推定する必要あり

q と π の反復推定
（ただし収束性は保証されない）

まとめ

モデル化の話

- 有声音の波形モデル
- ウェーブレットスペクトル
  - 定Q log-normalフィルタバンク

尤度関数設計

- 全極モデル
- 藤崎モデル

音声情報処理研究の歴史を代表する，音声生成過程に関する二大数理モデル → 事前分布設計

推論の話

- 統計的推論アルゴリズム
  - EMアルゴリズム
  - ベイズと点推定の組み合わせ（収束性は保証されない）