日本音響学会2010年秋季研究発表会

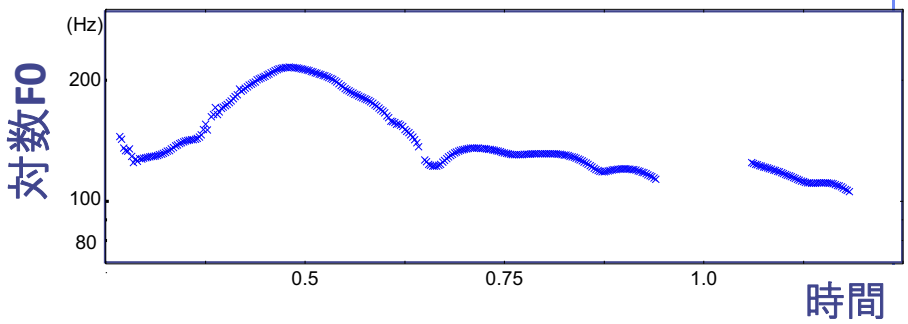
# 音声$F_0$パターン生成過程の確率モデル

○亀岡弘和
ルルージョナトン
大石康智

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

# 音声の基本周波数(F0)パターンについて

- F0パターンとは？
  - →声帯振動の周期の時間変化のパターン
    - 甲状軟骨の運動
      ⇒声帯の長さが微小変化⇒固有振動数が変化




- 音声対話において果たす役割は？
  - →音声の韻律的特徴を表現
    - 構文、意味、感情、焦点

- 音声情報処理における有用性
  - →F0パターンの自然性は音声らしさを表す有用な指標の一つ

「F0パターンが生成過程の物理的制約をどれだけ満たしているか」
→音声らしさを測るための客観的な規準に！

F0パターンのモデル化・分析は
様々な音声アプリケーションに有益なはず

→ **本研究の動機**

# 藤崎のF0パターン生成過程モデル(藤崎モデル) [藤崎1988]

[仮定1] 甲状軟骨の運動は二種類（平行移動と回転）

[仮定2] 各運動を臨界制動の二次線形系の応答として表現

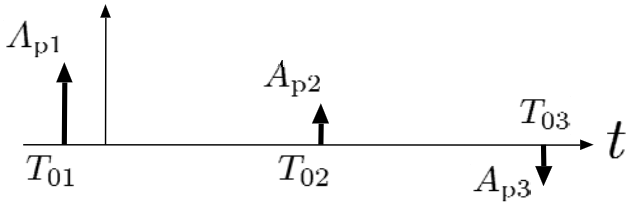
[仮定3] 各運動に伴う声帯の伸びの変化の合計が対数F0の変化に比例

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

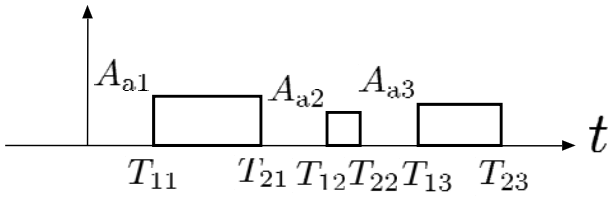
アクセント成分　$\Omega_{\mathrm{a}}(t)$ ：回転運動による変位に比例するF0パターン成分

ベースライン成分　$\Omega_{\mathrm{b}}$ ：声帯の物理的性質によって決まる定数

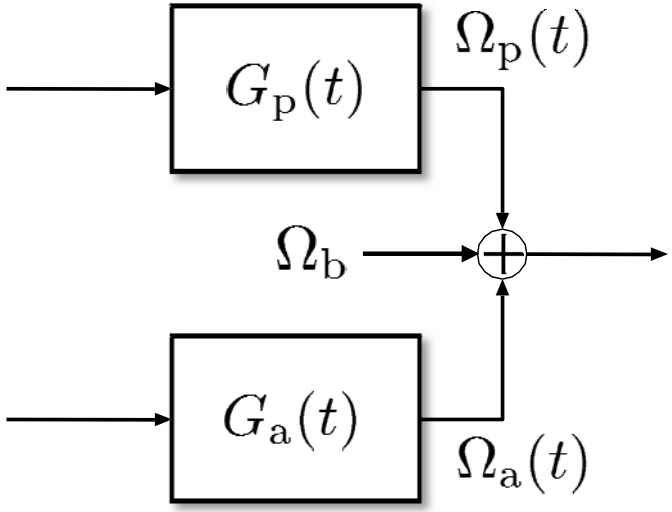
フレーズ指令 $u_{\mathrm{p}}(t)$



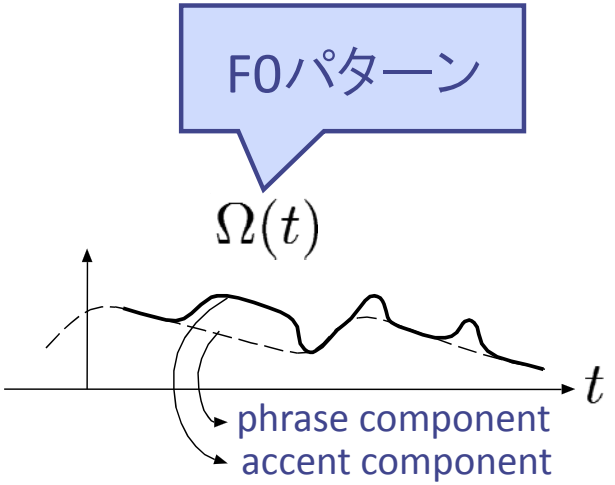

アクセント指令 $u_{\mathrm{a}}(t)$




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




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




原音声

アクセント指令
加工後音声

# 本研究の目的

藤崎モデルの離散時間表現(1)を統計モデル化(2)

- 統計的手法を駆使した強力なパラメータ推定法の枠組を確立
- 統計モデルに基づく音声処理問題(合成、分析、分離、強調)にスムーズに組み込める基盤を構築

## 発表アウトライン

- モデル化の手順
  - (1)後退差分近似による連続時間藤崎モデルの離散化
  - (2)フレーズ指令＆アクセント指令の確率モデル化
  - → (1)と(2)からF0パターンの確率密度関数を導出
- 提案モデルの使用方法
- パラメータ推定アルゴリズムの導出

# 本研究の目的

藤崎モデルの離散時間表現(1)を統計モデル化(2)

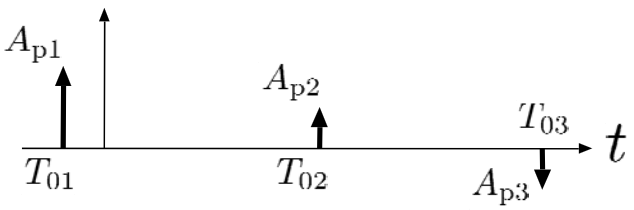
- 統計的手法を駆使した強力なパラメータ推定法の枠組を確立
- 統計モデルに基づく音声処理問題(合成、分析、分離、強調)にスムーズに組み込める基盤を構築

## 発表アウトライン

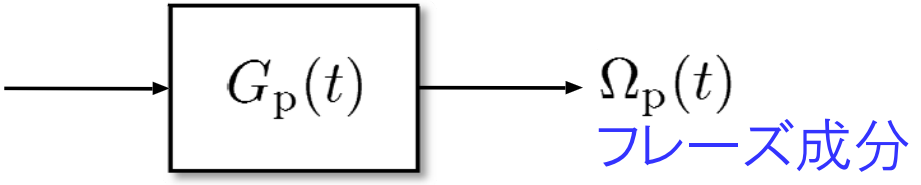
- モデル化の手順
  - (1)後退差分近似による連続時間藤崎モデルの離散化
  - (2)フレーズ指令＆アクセント指令の確率モデル化
  - → (1)と(2)からF0パターンの確率密度関数を導出
- 提案モデルの使用方法
- パラメータ推定アルゴリズムの導出

# 連続時間系から離散時間系へ（フレーズ制御系）

フレーズ指令入力 $u_{\mathrm{p}}(t)$




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




フレーズ成分

- フレーズ制御系のラプラス変換 $\alpha^2/(s+\alpha)^2$
- 逆システムのラプラス変換 → $(s+\alpha)^2/\alpha^2$
- 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
- 逆システムの伝達関数 $\dfrac{z^{-2} - 2(\alpha t_0 + 1)z^{-1} + (\alpha t_0 + 1)^2}{\alpha^2 t_0^2}$

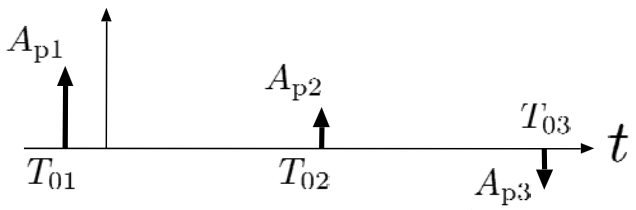
⇒ $u_{\mathrm{p}}[l] = g_0^{\mathrm{p}}\Omega_{\mathrm{p}}[l] + g_1^{\mathrm{p}}\Omega_{\mathrm{p}}[l-1] + g_2^{\mathrm{p}}\Omega_{\mathrm{p}}[l-2]$

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

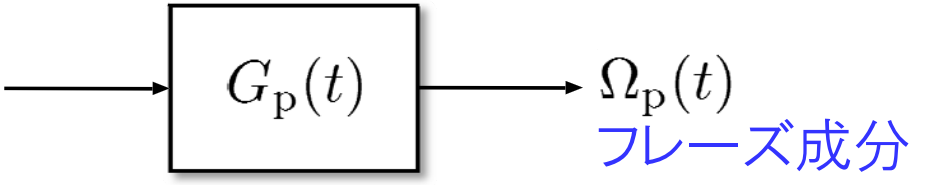
パラメータ $\psi$ で特性が決まる拘束つき全極モデル

# 連続時間系から離散時間系へ（フレーズ制御系）

フレーズ指令入力 $u_{\mathrm{p}}(t)$




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




- フレーズ制御系のラプラス変換 $\alpha^2/(s+\alpha)^2$
- 逆システムのラプラス変換 → $(s+\alpha)^2/\alpha^2$
- 後退差分近似 … 表現のサンプリング周期
- 逆システムの … $\cdots + (\alpha t_0 + 1)^2$

アクセント制御系も同形
なので以上と同様！

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

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

パラメータ $\psi$ で特性が決まる拘束つき全極モデル

# 本研究の目的

藤崎モデルの離散時間表現(1)を統計モデル化(2)

- 統計的手法を駆使した強力なパラメータ推定法の枠組を確立
- 統計モデルに基づく音声処理問題(合成、分析、分離、強調)にスムーズに組み込める基盤を構築

## 発表アウトライン

- モデル化の手順
  - (1)後退差分近似による連続時間藤崎モデルの離散化
  - (2)フレーズ指令＆アクセント指令の確率モデル化
  - → (1)と(2)からF0パターンの確率密度関数を導出
- 提案モデルの使用方法
- パラメータ推定アルゴリズムの導出

# 本研究の目的

藤崎モデルの離散時間表現(1)を統計モデル化(2)

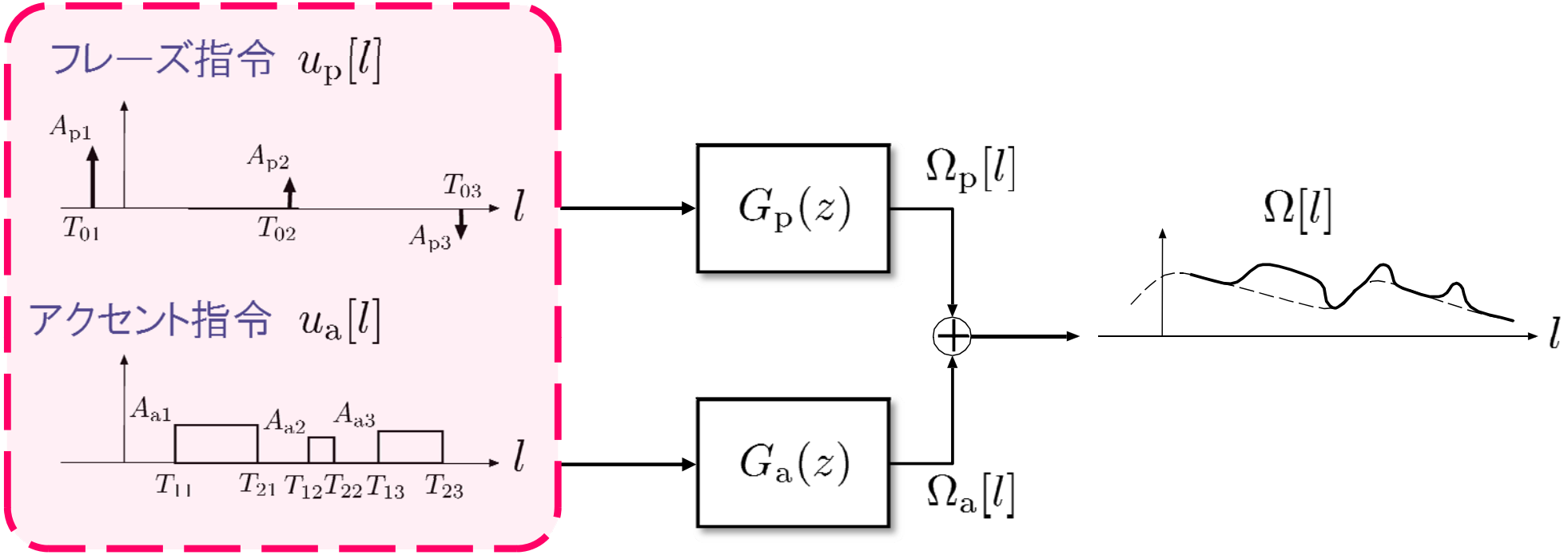
- 統計的手法を駆使した強力なパラメータ推定法の枠組を確立
- 統計モデルに基づく音声処理問題(合成、分析、分離、強調)にスムーズに組み込める基盤を構築

## 発表アウトライン

- モデル化の手順
  - (1)後退差分近似による連続時間藤崎モデルの離散化
  - (2)フレーズ指令＆アクセント指令の確率モデル化
  - → (1)と(2)からF0パターンの確率密度関数を導出
- 提案モデルの使用方法
- パラメータ推定アルゴリズムの導出

# フレーズ指令＆アクセント指令の確率モデル化

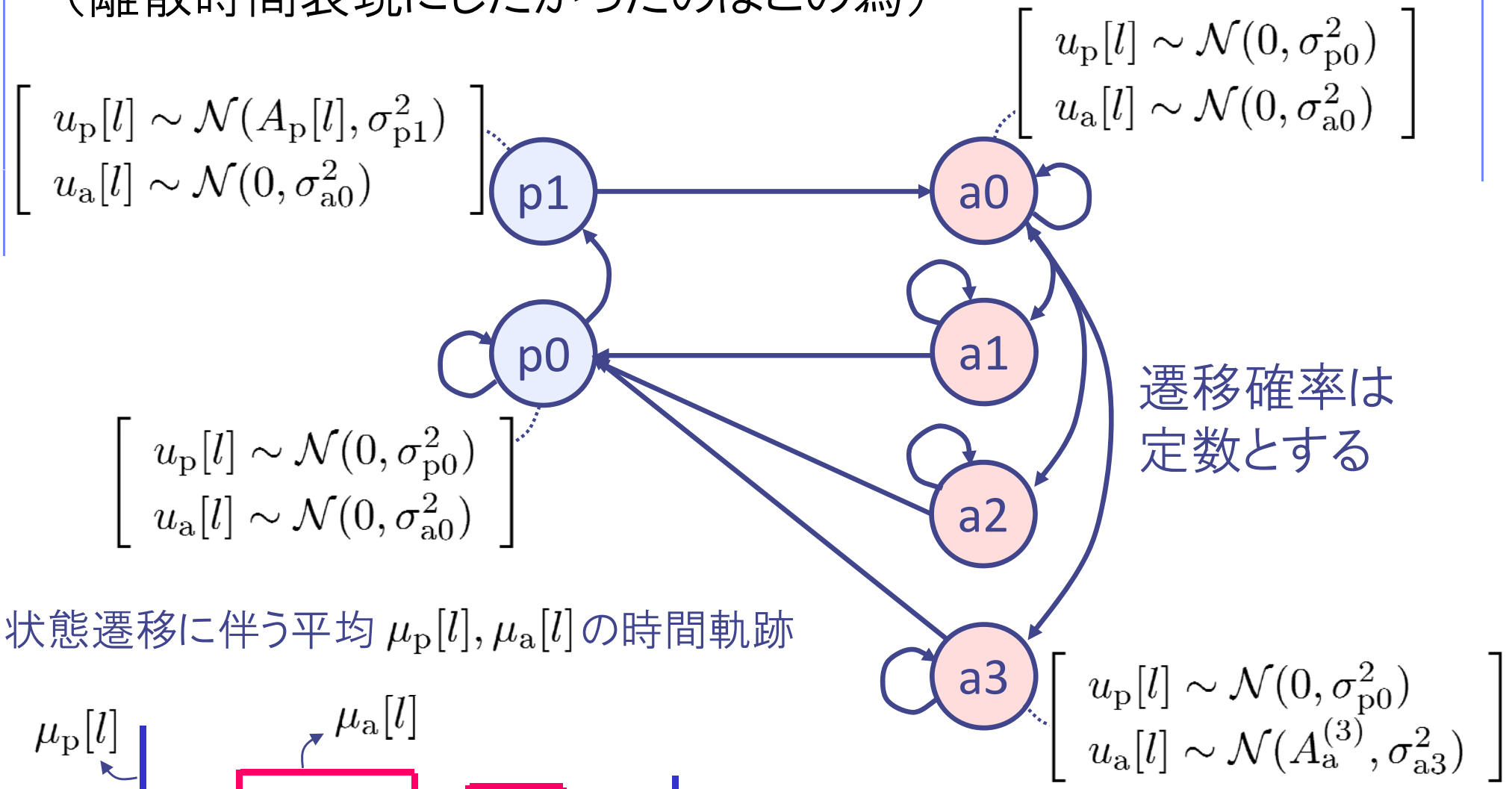
- 両指令は無制約で良い訳ではない




- 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
  - 指令は同時刻に2つ以上生起しない

# フレーズ指令＆アクセント指令の確率モデル化

- 経路制限付き隠れマルコフモデルでモデル化
  （離散時間表現にしたかったのはこの為）



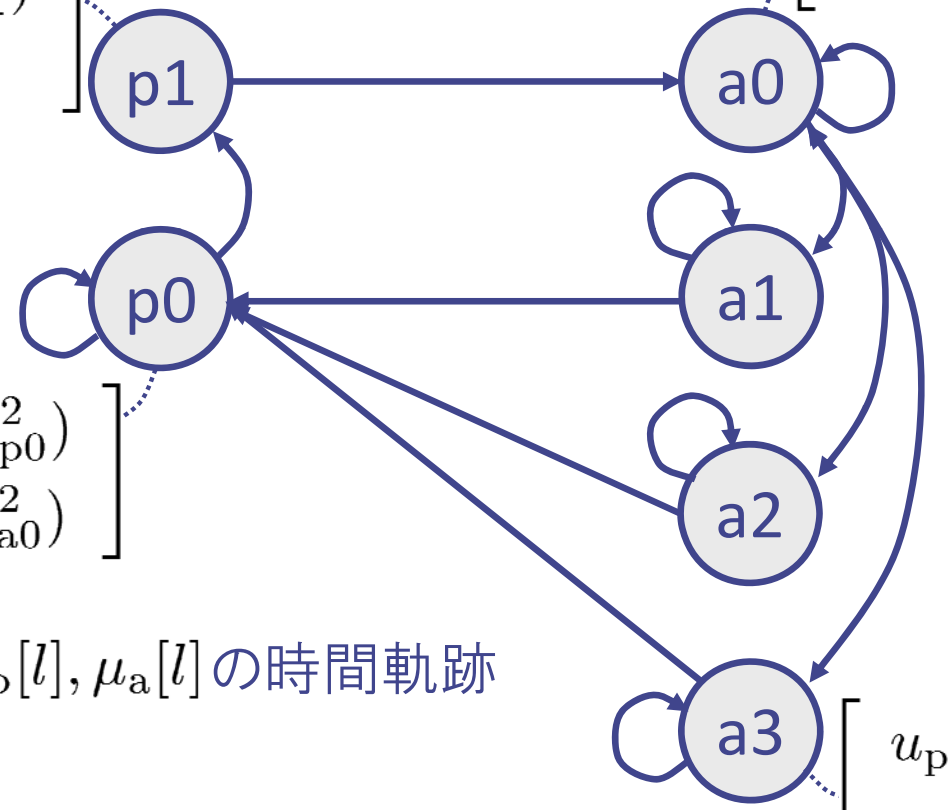

状態遷移に伴う平均 $\mu_{\mathrm{p}}[l], \mu_{\mathrm{a}}[l]$ の時間軌跡

# フレーズ指令＆アクセント指令の確率モデル化

- 経路制限付き隠れマルコフモデルでモデル化
  （離散時間表現にしたかったのはこの為）

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

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

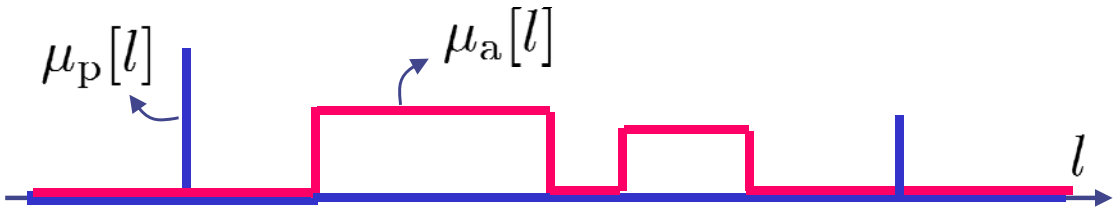



$$\begin{bmatrix} u_{\mathrm{p}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{p0}}) \\ u_{\mathrm{a}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{a0}}) \end{bmatrix}$$

遷移確率は
定数とする

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

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

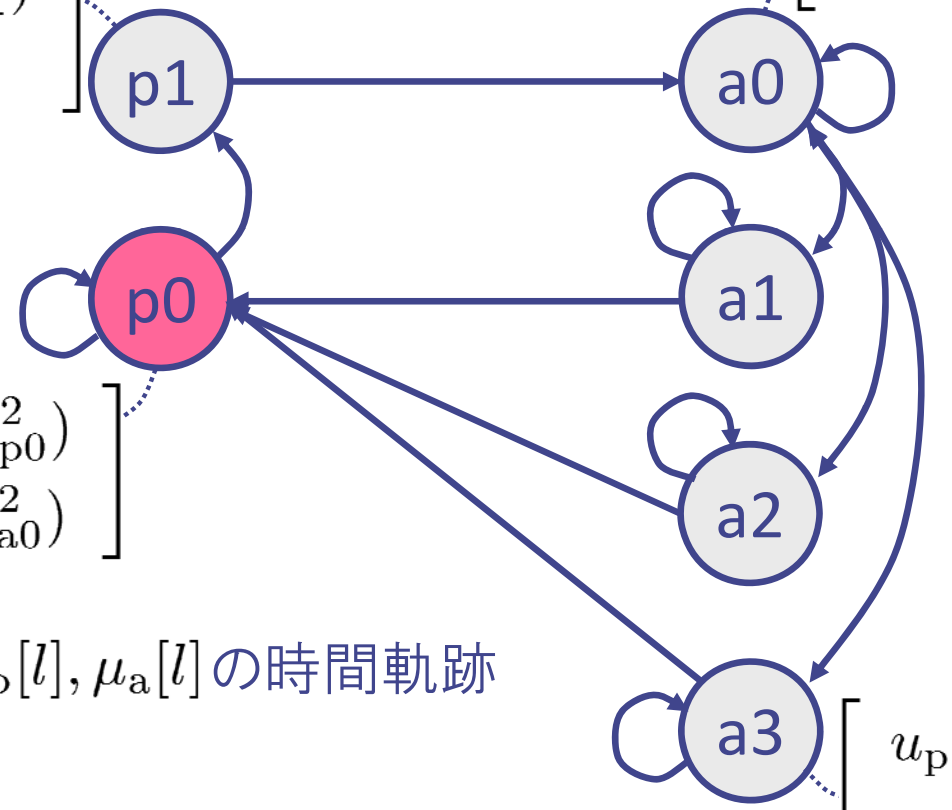


# フレーズ指令＆アクセント指令の確率モデル化

- 経路制限付き隠れマルコフモデルでモデル化
  （離散時間表現にしたかったのはこの為）

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

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

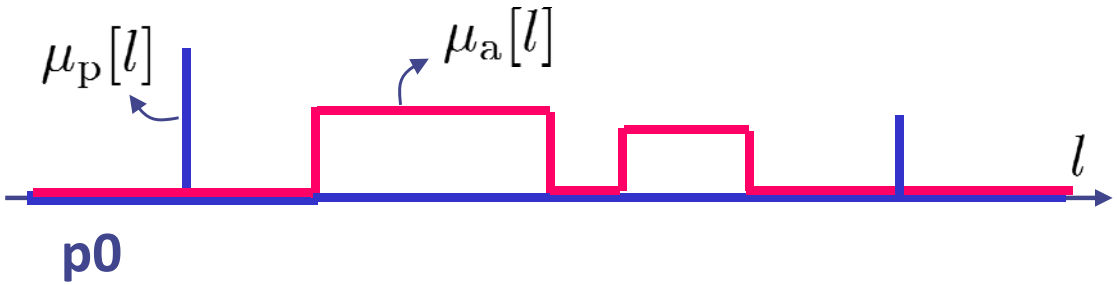



$$\begin{bmatrix} u_{\mathrm{p}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{p0}}) \\ u_{\mathrm{a}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{a0}}) \end{bmatrix}$$

遷移確率は
定数とする

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

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

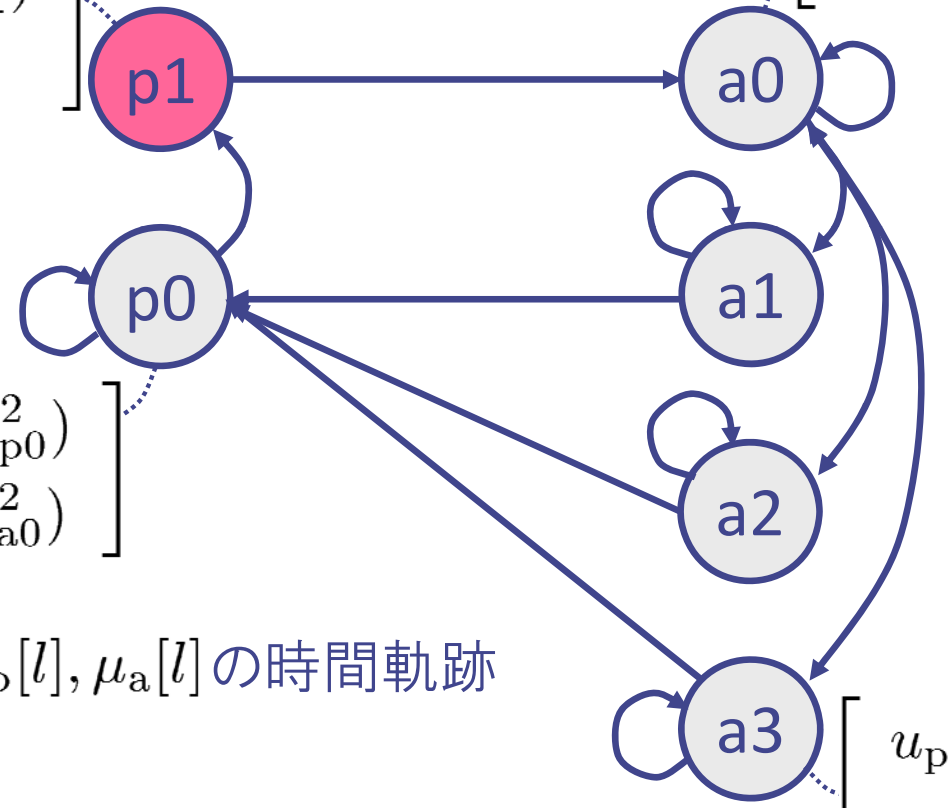


# フレーズ指令＆アクセント指令の確率モデル化

- 経路制限付き隠れマルコフモデルでモデル化
  （離散時間表現にしたかったのはこの為）

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

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

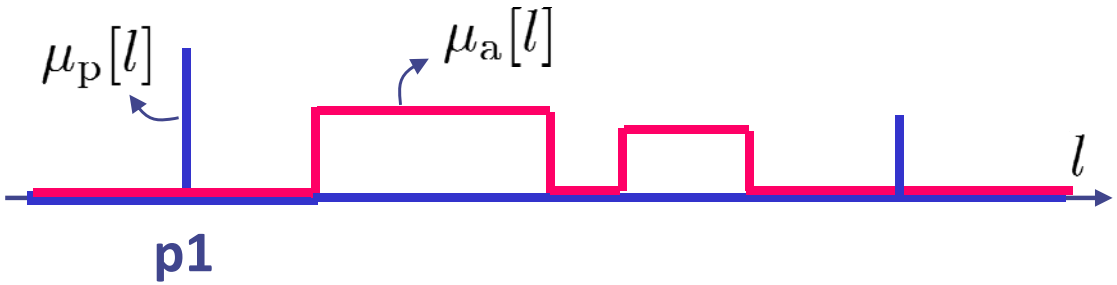



$$\begin{bmatrix} u_{\mathrm{p}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{p0}}) \\ u_{\mathrm{a}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{a0}}) \end{bmatrix}$$

遷移確率は
定数とする

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

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

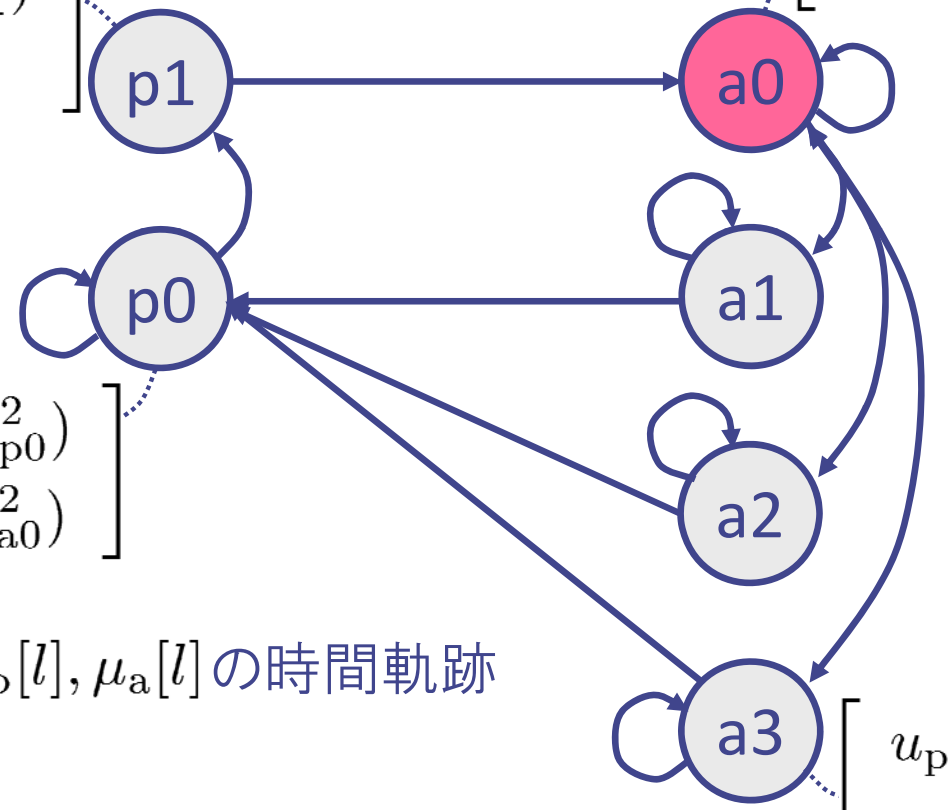


# フレーズ指令＆アクセント指令の確率モデル化

- 経路制限付き隠れマルコフモデルでモデル化
  （離散時間表現にしたかったのはこの為）

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

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

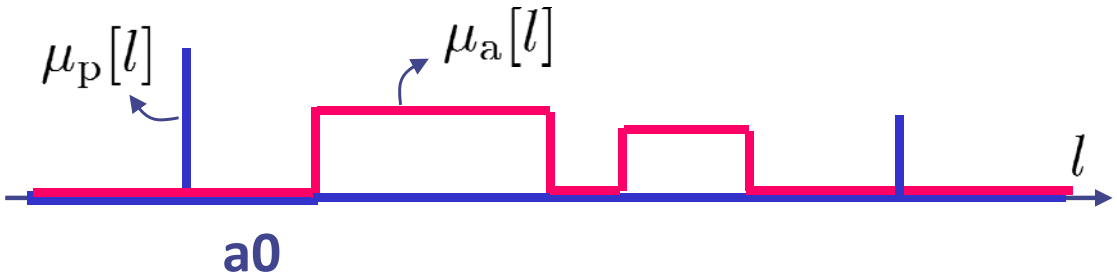



$$\begin{bmatrix} u_{\mathrm{p}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{p0}}) \\ u_{\mathrm{a}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{a0}}) \end{bmatrix}$$

遷移確率は
定数とする

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

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

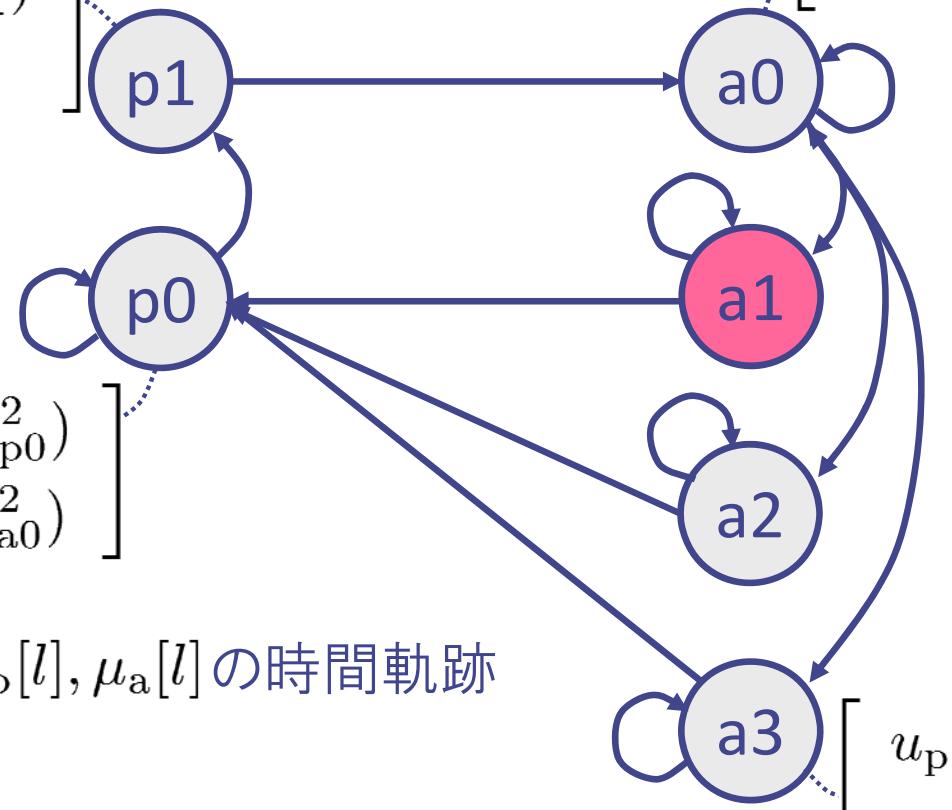


# フレーズ指令＆アクセント指令の確率モデル化

- 経路制限付き隠れマルコフモデルでモデル化
  （離散時間表現にしたかったのはこの為）

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

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

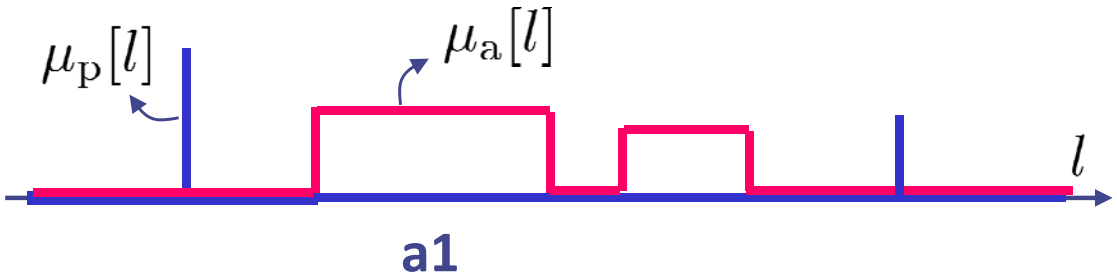



$$\begin{bmatrix} u_{\mathrm{p}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{p0}}) \\ u_{\mathrm{a}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{a0}}) \end{bmatrix}$$

遷移確率は
定数とする

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

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

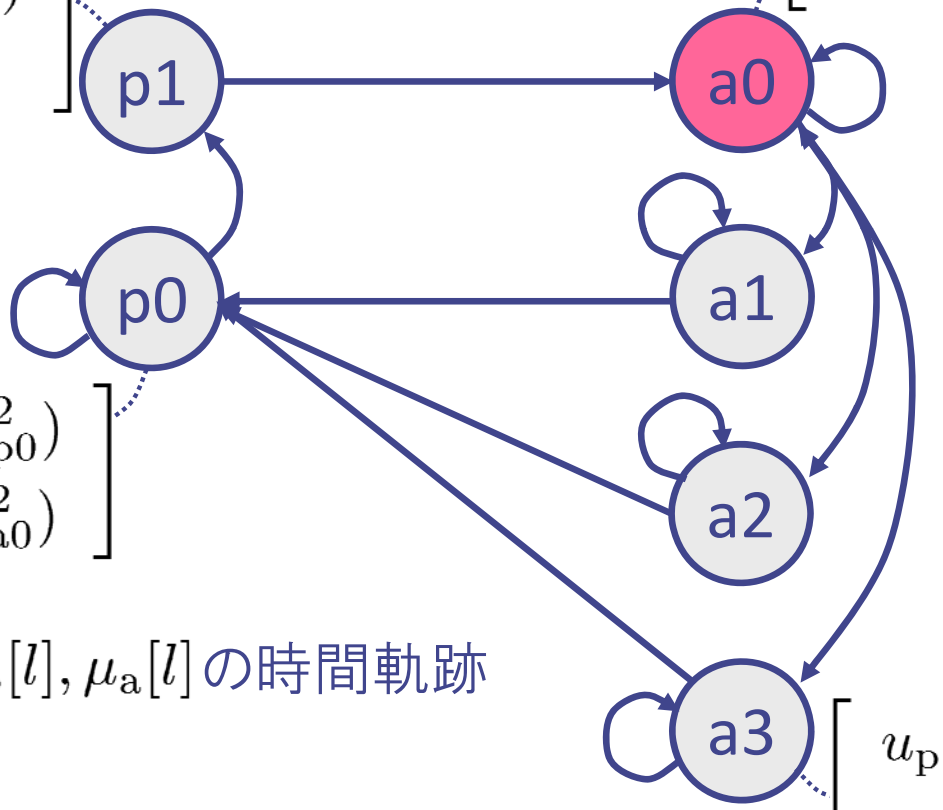


# フレーズ指令＆アクセント指令の確率モデル化

- 経路制限付き隠れマルコフモデルでモデル化
  （離散時間表現にしたかったのはこの為）

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

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

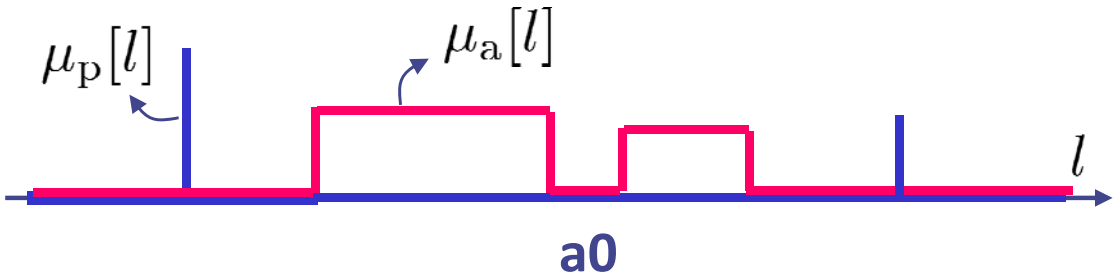



$$\begin{bmatrix} u_{\mathrm{p}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{p0}}) \\ u_{\mathrm{a}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{a0}}) \end{bmatrix}$$

遷移確率は
定数とする

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

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

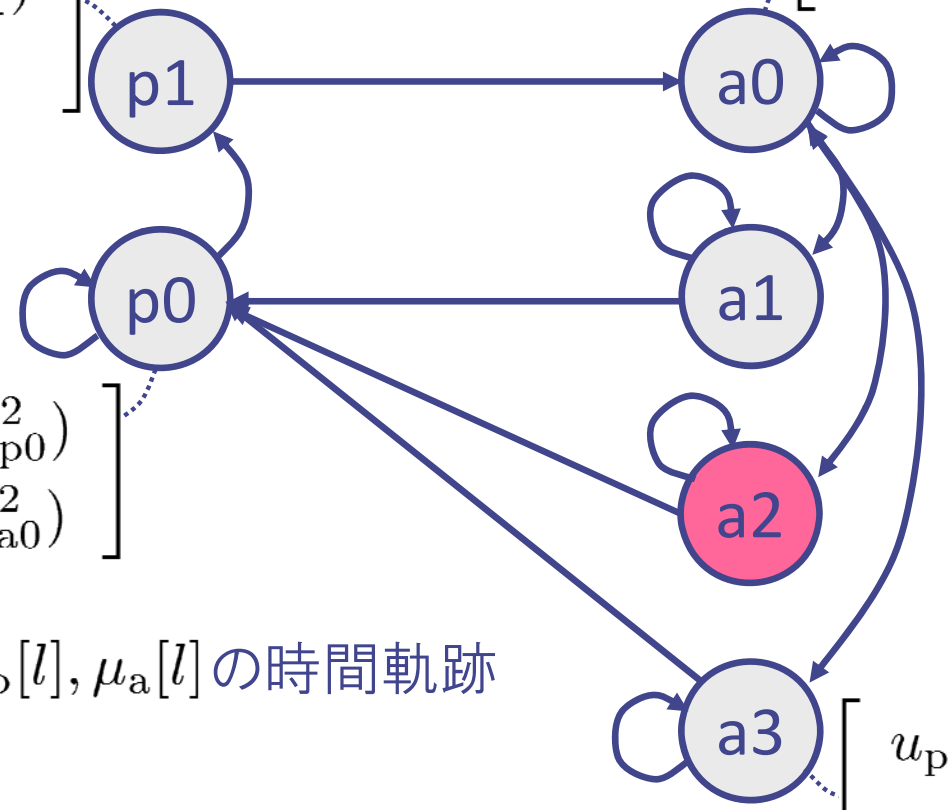


# フレーズ指令＆アクセント指令の確率モデル化

- 経路制限付き隠れマルコフモデルでモデル化
  （離散時間表現にしたかったのはこの為）

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

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

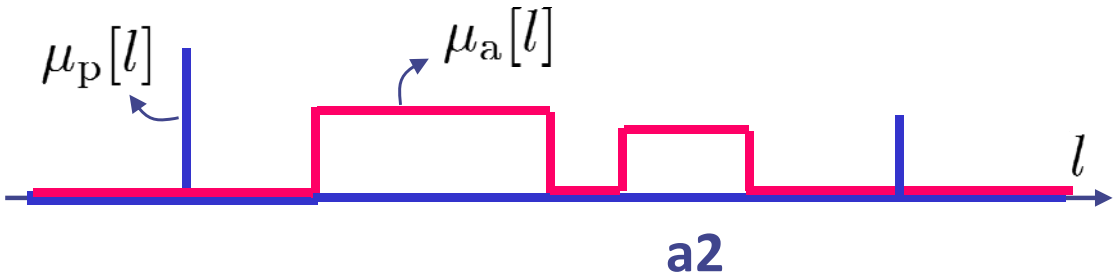



$$\begin{bmatrix} u_{\mathrm{p}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{p0}}) \\ u_{\mathrm{a}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{a0}}) \end{bmatrix}$$

遷移確率は
定数とする

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

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

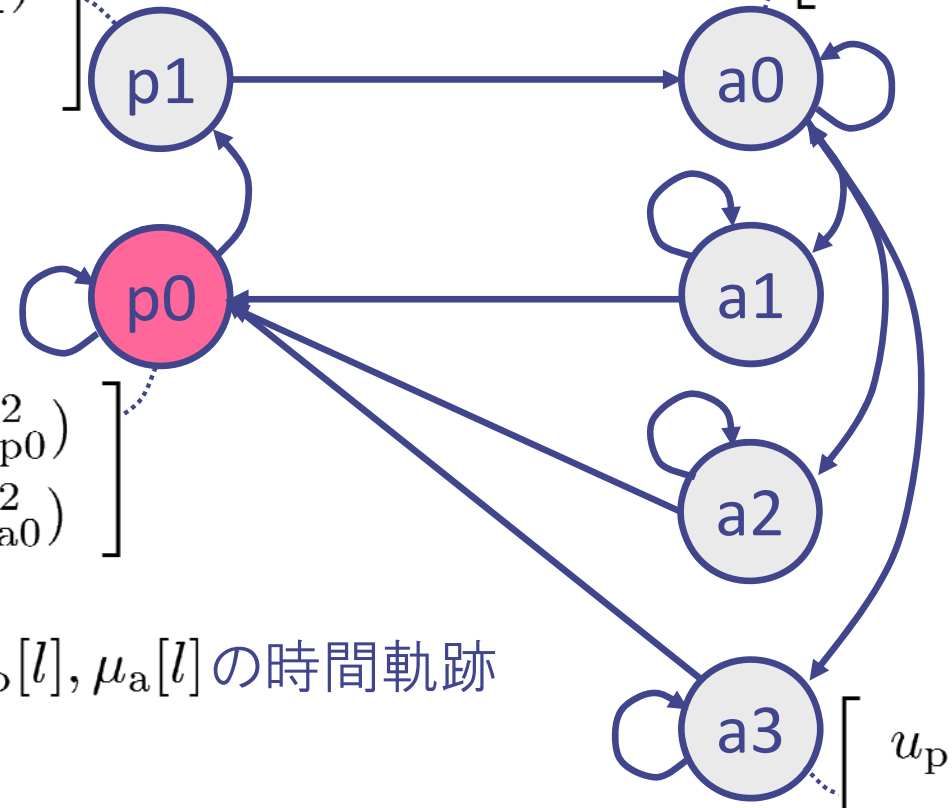


# フレーズ指令＆アクセント指令の確率モデル化

- 経路制限付き隠れマルコフモデルでモデル化
  （離散時間表現にしたかったのはこの為）

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

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

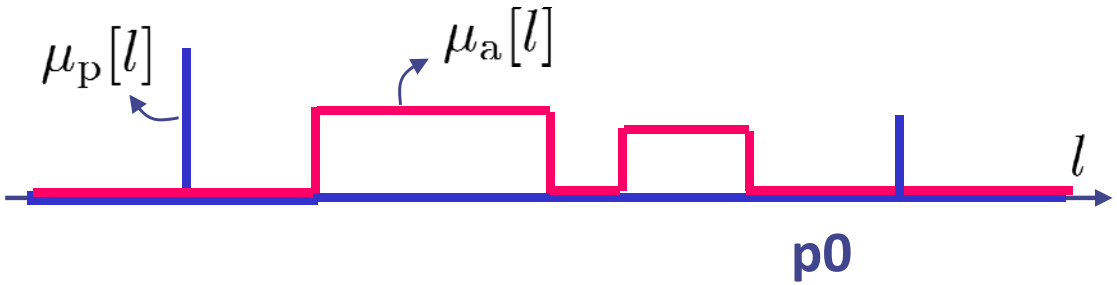



$$\begin{bmatrix} u_{\mathrm{p}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{p0}}) \\ u_{\mathrm{a}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{a0}}) \end{bmatrix}$$

遷移確率は
定数とする

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

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

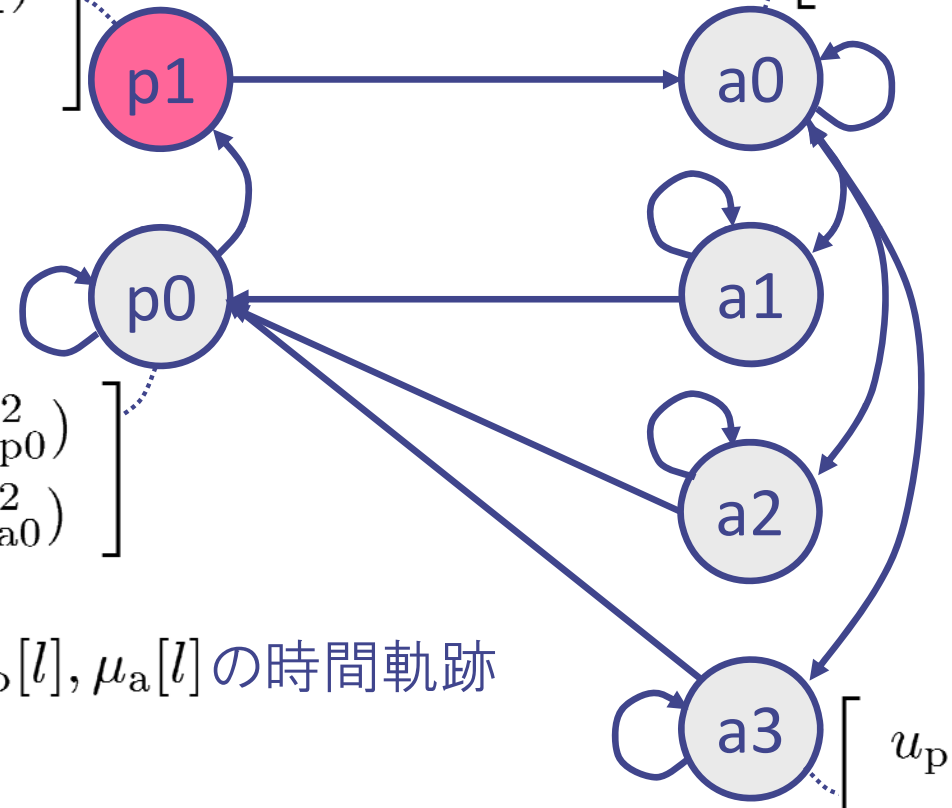


# フレーズ指令＆アクセント指令の確率モデル化

- 経路制限付き隠れマルコフモデルでモデル化
  （離散時間表現にしたかったのはこの為）

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

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

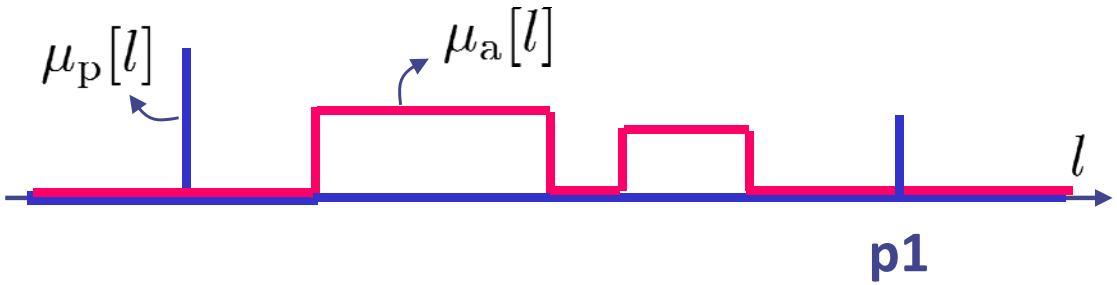



$$\begin{bmatrix} u_{\mathrm{p}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{p0}}) \\ u_{\mathrm{a}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{a0}}) \end{bmatrix}$$

遷移確率は
定数とする

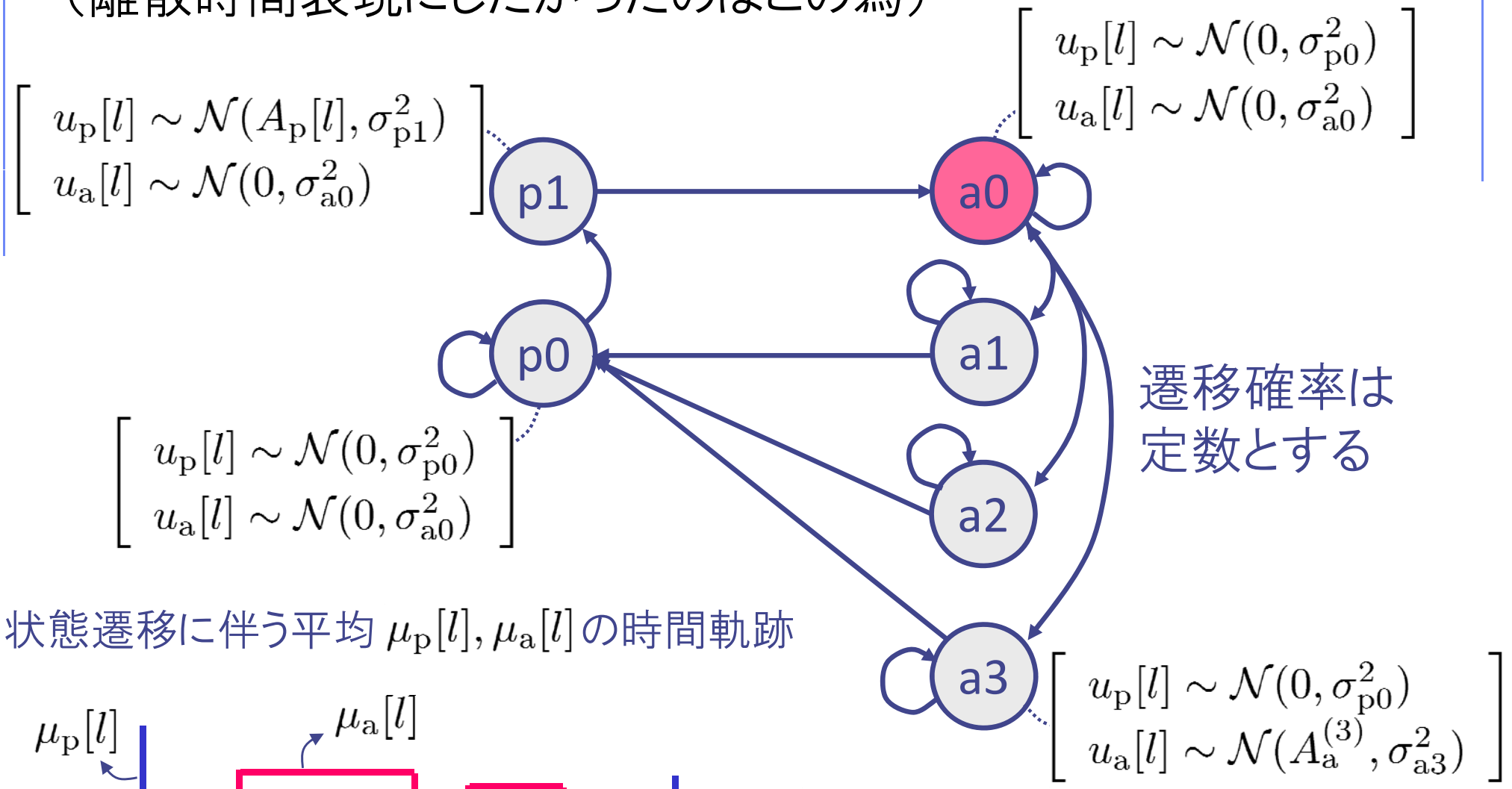
$$\begin{bmatrix} u_{\mathrm{p}}[l] \sim \mathcal{N}(0, \sigma^2_{\mathrm{p0}}) \\ u_{\mathrm{a}}[l] \sim \mathcal{N}(A^{(3)}_{\mathrm{a}}, \sigma^2_{\mathrm{a3}}) \end{bmatrix}$$

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



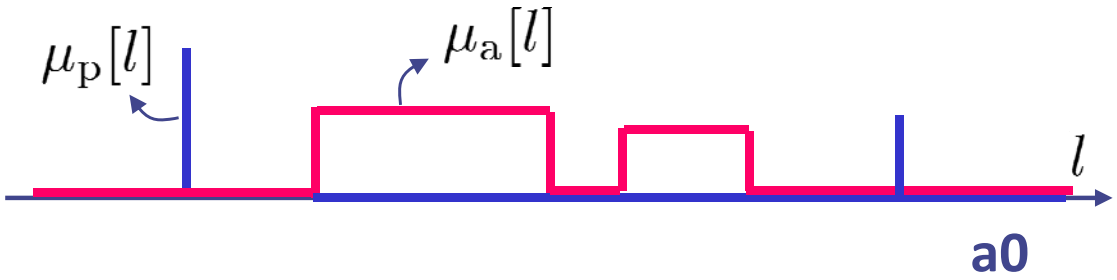
# フレーズ指令＆アクセント指令の確率モデル化

- 経路制限付き隠れマルコフモデルでモデル化
  （離散時間表現にしたかったのはこの為）




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



# 本研究の目的

藤崎モデルの離散時間表現(1)を統計モデル化(2)

- 統計的手法を駆使した強力なパラメータ推定法の枠組を確立
- 統計モデルに基づく音声処理問題(合成、分析、分離、強調)にスムーズに組み込める基盤を構築

## 発表アウトライン

- モデル化の手順
  - (1)後退差分近似による連続時間藤崎モデルの離散化
  - (2)フレーズ指令＆アクセント指令の確率モデル化
  - → (1)と(2)からF0パターンの確率密度関数を導出
- 提案モデルの使用方法
- パラメータ推定アルゴリズムの導出

# 本研究の目的

藤崎モデルの離散時間表現(1)を統計モデル化(2)

- 統計的手法を駆使した強力なパラメータ推定法の枠組を確立
- 統計モデルに基づく音声処理問題(合成、分析、分離、強調)にスムーズに組み込める基盤を構築

## 発表アウトライン

- モデル化の手順
  - (1)後退差分近似による連続時間藤崎モデルの離散化
  - (2)フレーズ指令＆アクセント指令の確率モデル化
  - → (1)と(2)からF0パターンの確率密度関数を導出
- 提案モデルの使用方法
- パラメータ推定アルゴリズムの導出

# F0パターンの確率密度関数の導出

- 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

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

（2）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

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

- 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　$\boldsymbol{\Omega} = \boldsymbol{\Omega}_{\mathrm{p}} + \boldsymbol{\Omega}_{\mathrm{a}} + \boldsymbol{\Omega}_{\mathrm{b}}$

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

# F0パターンの確率密度関数の導出

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

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

  （2）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

$G_{\mathrm{p}}$

$$\begin{bmatrix} g_0^{\mathrm{p}} & & \\ g_1^{\mathrm{p}} & g_0^{\mathrm{p}} & \\ g_2^{\mathrm{p}} & g_1^{\mathrm{p}} & g_0^{\mathrm{p}} \\ & \ddots & \ddots \\ 0 & & g_2^{\mathrm{p}} \end{bmatrix}$$

パラメータ $\Theta$：

- 状態遷移系列　$\{s_l\}_{l=1}^{L}$
- 状態出力分布の平均値　$\{A_{\mathrm{p}}[l]\}_{l=1}^{L}, \{A_{\mathrm{a}}^{(n)}\}_{n=1}^{N}$
- フレーズ制御パラメータ　$\psi$
- アクセント制御パラメータ　$\varphi$
- ベースライン値　$\mu_{\mathrm{b}}$

- アクセント成分
- F0パターンベクトル　$\Omega = \Omega_{\mathrm{p}} + \Omega_{\mathrm{a}} + \Omega_{\mathrm{b}}$

$$\Omega \sim \mathcal{N}(G_{\mathrm{p}}^{-1}\mu_{\mathrm{p}} + G_{\mathrm{a}}^{-1}\mu_{\mathrm{a}} + \mu_{\mathrm{b}}\mathbf{1},$$

$$G_{\mathrm{p}}^{-1}\Sigma_{\mathrm{p}}(G_{\mathrm{p}}^{-1})^{\mathrm{T}} + G_{\mathrm{a}}^{-1}\Sigma_{\mathrm{a}}(G_{\mathrm{a}}^{-1})^{\mathrm{T}} + \Sigma_{\mathrm{b}})$$

# 本研究の目的

藤崎モデルの離散時間表現(1)を統計モデル化(2)

- 統計的手法を駆使した強力なパラメータ推定法の枠組を確立
- 統計モデルに基づく音声処理問題(合成、分析、分離、強調)にスムーズに組み込める基盤を構築

## 発表アウトライン

- モデル化の手順
  - (1)後退差分近似による連続時間藤崎モデルの離散化
  - (2)フレーズ指令＆アクセント指令の確率モデル化
  - → (1)と(2)からF0パターンの確率密度関数を導出
- 提案モデルの使用方法
- パラメータ推定アルゴリズムの導出

# 本研究の目的

藤崎モデルの離散時間表現(1)を統計モデル化(2)

- 統計的手法を駆使した強力なパラメータ推定法の枠組を確立
- 統計モデルに基づく音声処理問題(合成、分析、分離、強調)にスムーズに組み込める基盤を構築

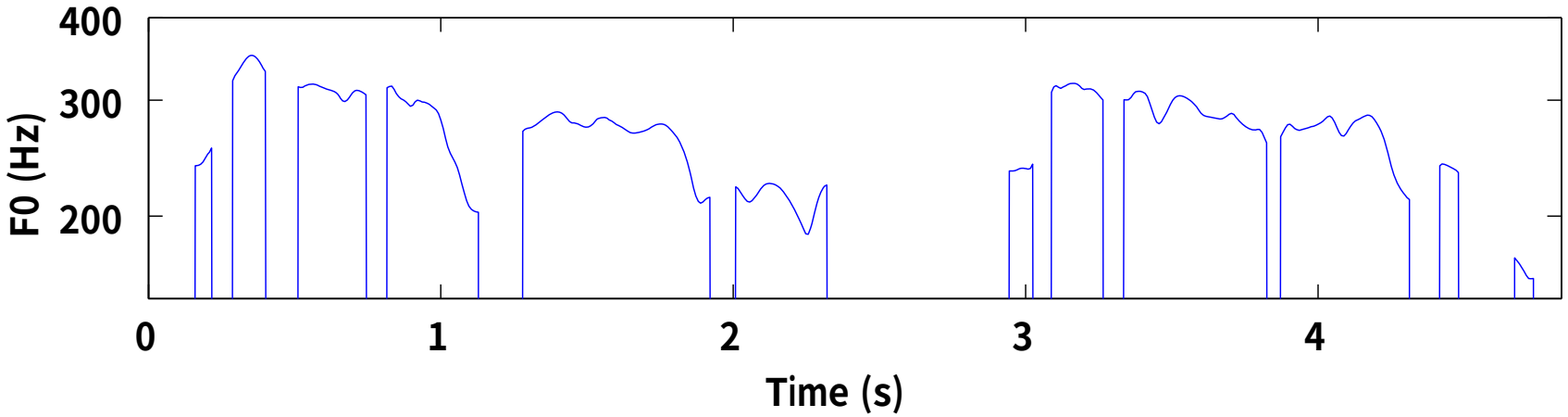
## 発表アウトライン

- モデル化の手順
  - (1)後退差分近似による連続時間藤崎モデルの離散化
  - (2)フレーズ指令＆アクセント指令の確率モデル化
  - → (1)と(2)からF0パターンの確率密度関数を導出
- 提案モデルの使用方法
- パラメータ推定アルゴリズムの導出

# 提案モデルの使用方法

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

- 実測F0パターンからの藤崎モデルの推定
  - 実測F0パターンは”欠損データ”




  - 上記統計モデルは全区間のデータが観測されていることが想定

→欠損データの下での最尤/最大事後確率推定は
EMアルゴリズムで解ける！

- F0をパラメータにもつ音声信号/スペクトルの統計モデルにおけるF0パラメータの事前分布として利用可能（詳細は直後の発表にて）

# 本研究の目的

藤崎モデルの離散時間表現(1)を統計モデル化(2)

- 統計的手法を駆使した強力なパラメータ推定法の枠組を確立
- 統計モデルに基づく音声処理問題(合成、分析、分離、強調)にスムーズに組み込める基盤を構築

## 発表アウトライン

- モデル化の手順
  - (1)後退差分近似による連続時間藤崎モデルの離散化
  - (2)フレーズ指令＆アクセント指令の確率モデル化
  - → (1)と(2)からF0パターンの確率密度関数を導出
- 提案モデルの使用方法
- パラメータ推定アルゴリズムの導出

# 本研究の目的

藤崎モデルの離散時間表現(1)を統計モデル化(2)

- 統計的手法を駆使した強力なパラメータ推定法の枠組を確立
- 統計モデルに基づく音声処理問題(合成、分析、分離、強調)にスムーズに組み込める基盤を構築

## 発表アウトライン

- モデル化の手順
  - (1)後退差分近似による連続時間藤崎モデルの離散化
  - (2)フレーズ指令＆アクセント指令の確率モデル化
  - → (1)と(2)からF0パターンの確率密度関数を導出
- 提案モデルの使用方法
- パラメータ推定アルゴリズムの導出

# パラメータの最大事後確率推定

- $\boldsymbol{\Omega}$が与えられた下で $P(\Theta|\boldsymbol{\Omega}) \propto P(\boldsymbol{\Omega}|\Theta)P(\Theta)$を最大化

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

- フレーズ成分 $\boldsymbol{\Omega}_{\mathrm{p}}$，アクセント成分 $\boldsymbol{\Omega}_{\mathrm{a}}$，ベースライン成分 $\boldsymbol{\Omega}_{\mathrm{b}}$ を完全データと見なすとEM法が適用できる！

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{\Omega}_{\mathrm{p}} \\ \boldsymbol{\Omega}_{\mathrm{a}} \\ \boldsymbol{\Omega}_{\mathrm{b}} \end{bmatrix} \sim \mathcal{N}\left( \underbrace{\begin{bmatrix} \boldsymbol{G}_{\mathrm{p}}^{-1}\boldsymbol{\mu}_{\mathrm{p}} \\ \boldsymbol{G}_{\mathrm{a}}^{-1}\boldsymbol{\mu}_{\mathrm{a}} \\ \mu_{\mathrm{b}}\mathbf{1} \end{bmatrix}}_{\boldsymbol{m}}, \underbrace{\begin{bmatrix} \boldsymbol{G}_{\mathrm{p}}^{-1}\boldsymbol{\Sigma}_{\mathrm{p}}(\boldsymbol{G}_{\mathrm{p}}^{-1})^{\mathrm{T}} & O & O \\ O & \boldsymbol{G}_{\mathrm{a}}^{-1}\boldsymbol{\Sigma}_{\mathrm{a}}(\boldsymbol{G}_{\mathrm{a}}^{-1})^{\mathrm{T}} & O \\ O & O & \boldsymbol{\Sigma}_{\mathrm{b}} \end{bmatrix}}_{\boldsymbol{\Lambda}} \right)$$

  - “不完全データ” $\boldsymbol{\Omega}$ と完全データの関係： $\boldsymbol{\Omega} = \underbrace{[\boldsymbol{I}\ \boldsymbol{I}\ \boldsymbol{I}]}_{\boldsymbol{H}} \begin{bmatrix} \boldsymbol{\Omega}_{\mathrm{p}} \\ \boldsymbol{\Omega}_{\mathrm{a}} \\ \boldsymbol{\Omega}_{\mathrm{b}} \end{bmatrix}$
  - Q関数

$$Q(\Theta, \Theta') \stackrel{\mathrm{c}}{=} -\frac{1}{2}\left[\mathrm{tr}(\boldsymbol{\Lambda}^{-1}\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{\Omega};\Theta]) - 2\boldsymbol{m}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\mathbb{E}[\boldsymbol{x}|\boldsymbol{\Omega};\Theta] + \boldsymbol{m}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\boldsymbol{m}\right]$$

# EMアルゴリズム

## E-step

F0パターンモデル

$$\mathbb{E}[\boldsymbol{x}|\boldsymbol{\Omega};\boldsymbol{\Theta}] = \boldsymbol{m} + \boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}})^{-1}(\boldsymbol{\Omega} - \boldsymbol{H}\boldsymbol{m})$$

$$\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{\Omega};\boldsymbol{\Theta}] = \boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}})^{-1}\boldsymbol{H}\boldsymbol{\Lambda} + \mathbb{E}[\boldsymbol{x}|\boldsymbol{\Omega};\boldsymbol{\Theta}]\mathbb{E}[\boldsymbol{x}|\boldsymbol{\Omega};\boldsymbol{\Theta}]^{\mathrm{T}}$$

## M-step

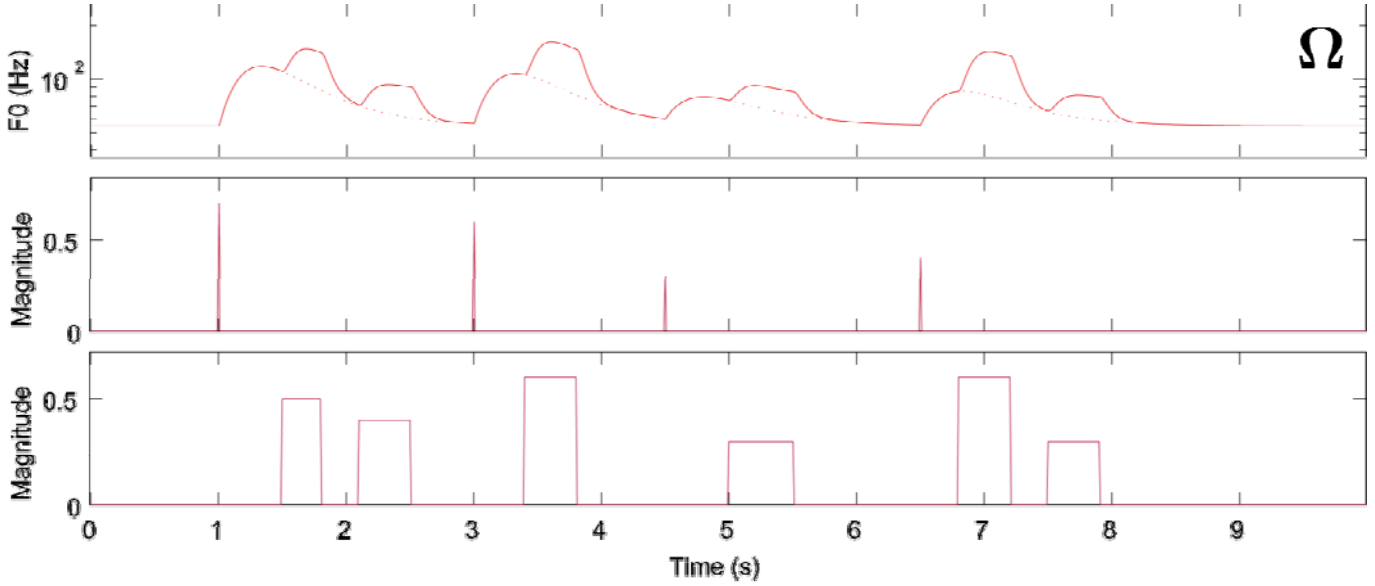
$$Q(\Theta,\Theta') \stackrel{\mathrm{c}}{=} -\frac{1}{2}\left[\mathrm{tr}(\boldsymbol{\Lambda}^{-1}\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{\Omega};\Theta]) - 2\boldsymbol{m}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\mathbb{E}[\boldsymbol{x}|\boldsymbol{\Omega};\Theta] + \boldsymbol{m}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\boldsymbol{m}\right]$$

を $\Theta$ に関して最大化

- $\psi, \varphi$ の更新は4次方程式の求解に帰着
- 状態遷移系列と状態出力分布の平均値の更新はViterbi学習により実現可能

# シミュレーション実験

- 実験データ：元の藤崎モデルから合成した人工データ

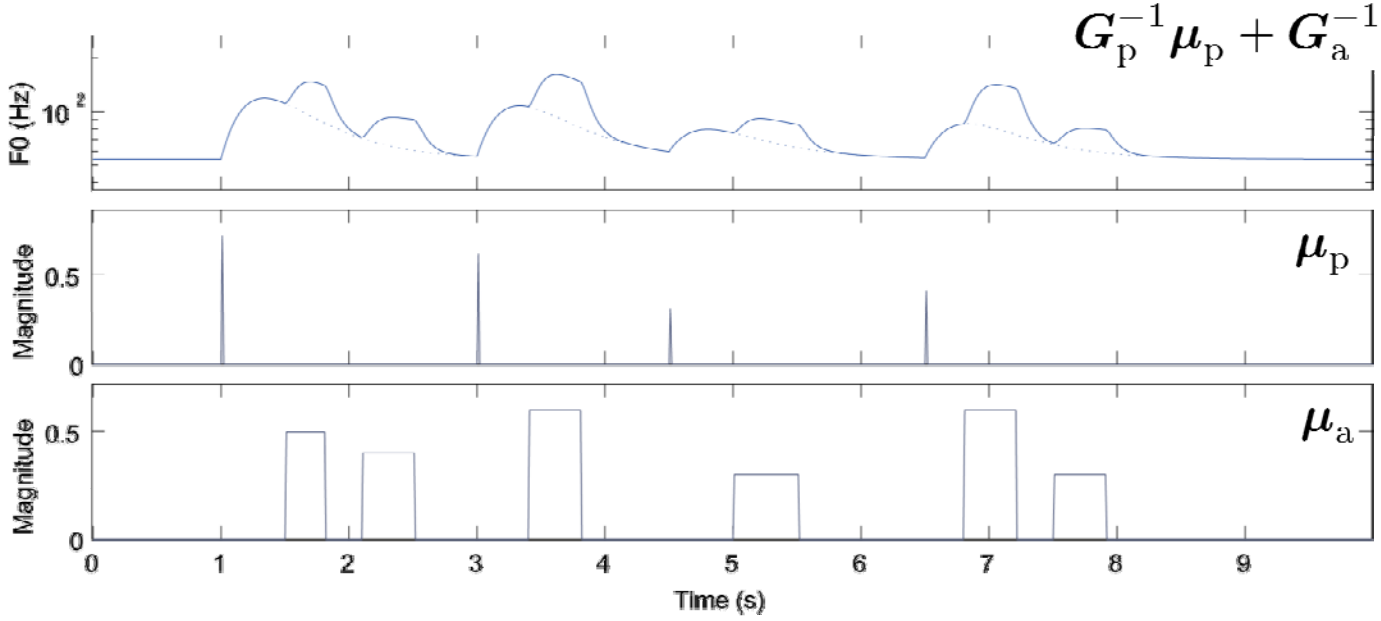



サンプリング
周期 $t_0$ =10ms

フレーズ指令

アクセント指令

- 推定結果： 状態数13，反復回数10




初期値設定：

$$\psi = 1 + \frac{1}{3t_0}$$

$$\varphi = 1 + \frac{1}{20t_0}$$

# まとめ

- 藤崎モデルの離散時間表現の確率モデルを提案
  - 統計的手法を駆使した強力なパラメータ推定法の枠組を構築
  - 音声の統計モデルに基づく各種音声処理問題にスムーズに組み込める基盤を構築
- 提案モデルのポイント
  - もともとは連続時間系の藤崎モデルを離散時間系で記述
  - フレーズ指令とアクセント指令を隠れマルコフモデルにより確率モデル化
  - EMアルゴリズムによるパラメータ推定法を導出
- 効果
  - 今のところ人工データに対する一有効性だけ確認済
  - 実測F0パターンに対してはどうか？欠損データの影響は？
  - 統計的音声モデルに組み込んだ際の効果は？
- 参考までに
  - 3-P-31 複数振動基底に基づく歌声のF0 動特性の統計的モデリング
    ◎大石 康智，亀岡 弘和，持橋 大地，永野 秀尚，柏野 邦夫