# 複数のパラメータ共有構造を用いた音響モデリング

塩田　さやか　モデリング研究系　特任助教

## 【研究背景：統計的機械学習に基づく音声認識】

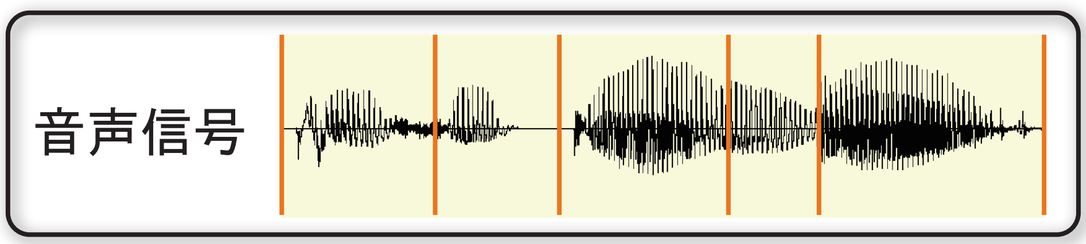
- 音声を媒体としたインタフェースの普及
  （携帯電話やカーナビゲーションなど）
- 音声の要素技術への統計的アプローチ

統計的アプローチ

音声合成　音声認識　話者認識　話者適応　音声翻訳

- ウェブ上等の膨大な音声データの存在
  ⇒ データの膨大化，複雑化

- 音響モデル
  音声信号から音の特徴をモデル化したもの




様々な要素技術の核となるモデル
⇒ 音声に関する要素技術全てに影響

- 音響モデルの性能はまだ不十分であるため
  更なる性能向上を目指した研究が重要となります

## 【研究内容：ベイズ基準に基づくモデル構造選択】

- 概要
  - 音響モデルとして広く使われている隠れマルコフモデル（HMM）
    - 学習アルゴリズム：　EMアルゴリズム
    - 学習基準：　尤度最大化(ML)基準
      ⇒ 初期値の影響が大きいため改善の必要あり
  - 音声認識に必要なモデル構造
    - モデルの複雑化や複数のモデル構造の組み合わせた研究の活発化（e.g., ROVER, random forest）

本研究では，複数のモデル構造を用いベイズ基準と確定的アニーリングEMアルゴリズムを適用することで初期値に依存しにくい高性能な音響モデルの推定を行う。

- 目的関数の導出
  - 複数のモデル構造を隠れ変数として含む周辺尤度関数

$$P(\boldsymbol{O}) = \sum_{m}\sum_{\boldsymbol{Z}} \int P(\boldsymbol{O}, \boldsymbol{Z}, m, \boldsymbol{\Lambda}_m) d\boldsymbol{\Lambda}_m$$
$$P(\boldsymbol{O}, \boldsymbol{Z}, m, \boldsymbol{\Lambda}_m) = P(\boldsymbol{O}, \boldsymbol{Z} \mid m, \boldsymbol{\Lambda}_m) P(\boldsymbol{\Lambda}_m \mid m) P(m)$$

$\boldsymbol{Z}$ ：状態系列
$\boldsymbol{O}$ ：観測
$m \in \{1, ..., M\}$
：モデル構造
$\boldsymbol{\Lambda} \in \{\boldsymbol{\Lambda}_1, ..., \boldsymbol{\Lambda}_M\}$
：モデルパラメータセット

- 確定的アニーリングEMアルゴリズムの適用
  - 自由エネルギー関数の再定義

$$\bar{\mathcal{F}}_\beta = -\frac{1}{\beta} \log \sum_{m}\sum_{\boldsymbol{Z}} \int P^\beta(\boldsymbol{O}, \boldsymbol{Z}, m, \boldsymbol{\Lambda}_m) d\boldsymbol{\Lambda}_m$$

$\beta$：Temperature parameter

  - 変分近似を用いることでモデルパラメータ・状態系列・モデル構造の変分事後確率を導出

$$\tilde{Q}(\boldsymbol{\Lambda}_m \mid m) = C_{\boldsymbol{\Lambda}_m} P^\beta(\boldsymbol{\Lambda}_m \mid m) \exp\left\langle \log P^\beta(\boldsymbol{O}, \boldsymbol{Z} \mid m, \boldsymbol{\Lambda}_m) \right\rangle_{\tilde{Q}(\boldsymbol{Z})}$$
$$\tilde{Q}(\boldsymbol{Z}) = C_{\boldsymbol{Z}} \exp\left\langle \left\langle \log P^\beta(\boldsymbol{O}, \boldsymbol{Z} \mid m, \boldsymbol{\Lambda}_m) \right\rangle_{\tilde{Q}(\boldsymbol{\Lambda}_m \mid m)} \right\rangle_{\tilde{Q}(m)}$$
$$\tilde{Q}(m) = C_m P^\beta(m) \exp\left\langle \left\langle \log P^\beta(\boldsymbol{O}, \boldsymbol{Z} \mid m, \boldsymbol{\Lambda}_m) \right\rangle_{\tilde{Q}(\boldsymbol{Z})} + \log \frac{P^\beta(\boldsymbol{\Lambda}_m \mid m)}{\tilde{Q}(\boldsymbol{\Lambda}_m \mid m)} \right\rangle_{\tilde{Q}(\boldsymbol{\Lambda}_m \mid m)}$$

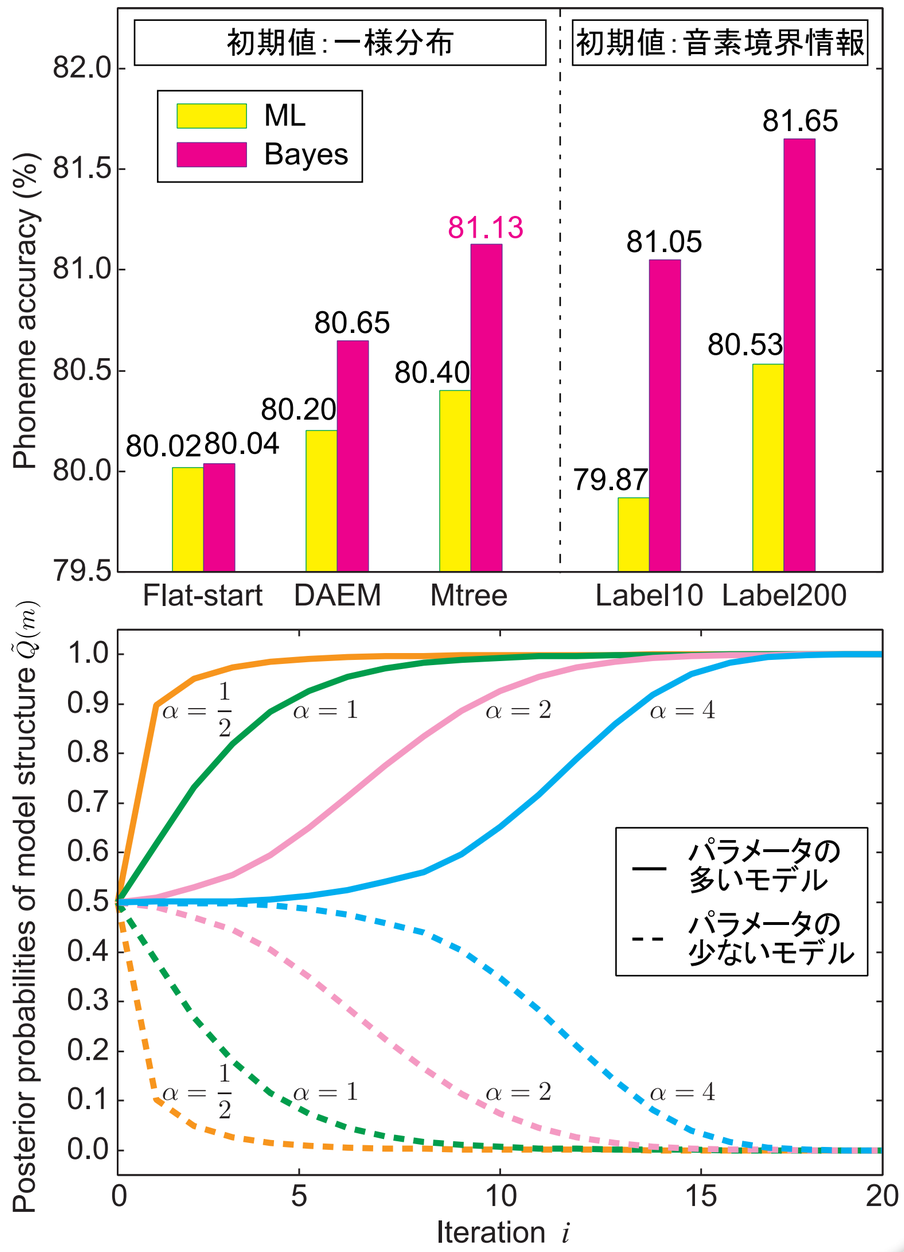
  - 温度パラメータを0（高温）から1（低温）に変えていきながら
    各温度で $\tilde{Q}(\boldsymbol{Z})$, $\tilde{Q}(m)$, $\tilde{Q}(\boldsymbol{\Lambda}_m \mid m)$ を繰り返し推定していくことで
    初期値に依存しにくい高精度なモデルを推定することが可能

アニーリングの枠組みを用いたモデル学習


※アニーリング：高温から低温に温度を少しずつ下げながら形を整える手法

温度スケジュール　$\beta(i) = \left(\frac{i}{I}\right)^\alpha, \quad i = 0, ..., I$
$\alpha = 1/8, 1/4, ..., 4, 8$

- 実験結果




音声を用いた統計的機械学習と応用研究に
興味のある大学院生及び共同研究者を歓迎致します