

# 競技かるたの決まり字に関する統計解析

松田 孟留

理化学研究所・脳神経科学研究センター

2021年9月8日

# はじめに

- 競技かるた…百人一首の札を取る速さを競うスポーツ



- マンガ等の影響で競技人口が急増加 → コロナでピンチ



# 競技かるたのルール

- 100枚の札それぞれに和歌の下の句が書いてある
- 読手は上の句を詠み上げる
- 選手は対応する下の句の札を取る

く め か  
る に ら  
と み く  
は つ れ  
く な

ふ て わ  
り に か  
つ ゆ こ  
つき ろ  
は も

つ に く  
き し も  
か よ か  
な は く  
の れ



# 決まり字

- 決まり字…どの札が詠まれたか確定する文字
- 例：「い」で始まる札は次の3枚
  - ▶ いにしへの～ → 「いに」決まり
  - ▶ いまこんと～ → 「いまこ」決まり
  - ▶ いまはただ～ → 「いまわ」決まり

ぬ	へ	け
る	に	ふ
か	に	こ
な	ほ	こ
	ひ	の

い	つ	あ
て	き	り
つ	を	あ
る	か	ま
か	な	け
な	ち	の

し	ら	ひ
も	て	と
か	い	つ
な	ふ	て
	よ	な

- ひらがなレベルで表しているが…
- 実際にはもっと早いタイミングで認識できていそう？

## 選手の声

- とある慶應OBの方が書いた文章（抜粋）

「一字決まりの札を半音で取れたら」「三字決まりの札を二字で取れたら」と考えるのは競技かるたを競技する多くの人の願望であろう。…

世の中には、「す」「さ」「せ」のSの半音の違いが分かるという人もいるのである。聞き分けようと神経を集中すればわかるようになるのかもしれない。また、「あきの」と「あきか」は「き」を詠むときのトーンの違いで二音目で区別することが可能であるという人もいるし、「おおえ」の「オー」は、「おほけ」と「あふこ」の「オー」とは違うので見当がつくという人もいる。…

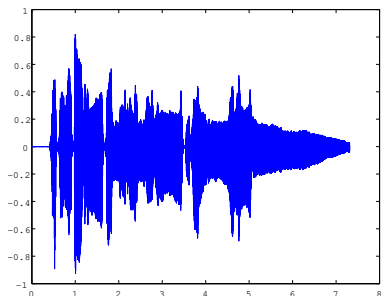
「もしも、百枚すべてについて、こうした特徴をつかむことができたなら…。」まさに夢である。夢を実現する選手はあらわれないものだろうか。

- 「夢を実現する選手」になりたい

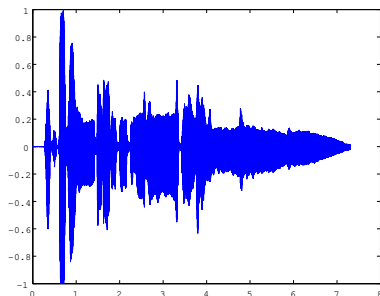
- 第58期クイーン戦1回戦
  - ▶ <https://www.youtube.com/watch?v=RhLb7AklHWQ>

# 音声の波形

- 音声 = 空気の振動
- 音によってスペクトルなどが異なる



: 「あきの」の波形

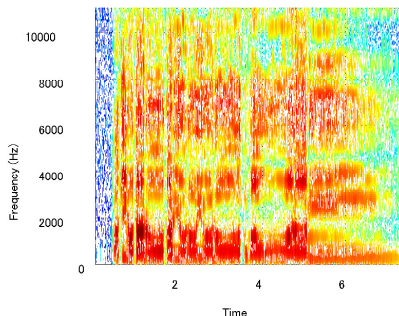


: 「あきか」の波形

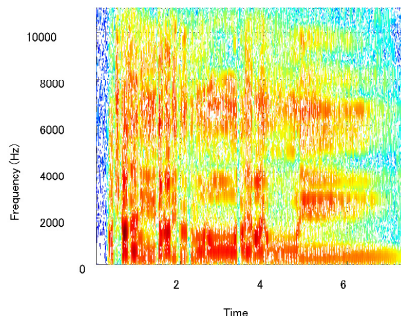


# 音声の波形

- スペクトログラム（各周波数のパワーの時間変化）
  - ▶ 横軸は時間，縦軸は周波数



: 「あきの」



: 「あきか」

- どの時点で差が生じている??

# 本研究の概要

- 統計解析によって「**真の決まり字**」を明らかにしたい

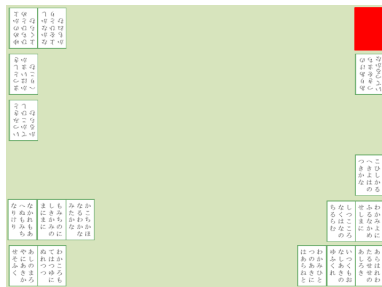
↓どうやって？

- 札ごとに詠み音声の発生過程を統計モデルとして学習する
  - ▶ 隠れマルコフモデル
- 学習した統計モデルをもとに、いま詠まれているのがどの札なのか推理していく
  - ▶ 逐次ベイズ推定
- 札が特定される時点を調べる
- 実際に選手が札を認識する過程に沿ったつもり



# 札を取る過程のモデリング

- 選手は詠みを聞きながら逐次的に「事後分布」 $\pi(k | X_1, \dots, X_t)$ を更新
  - ▶ ベイズの定理を利用
  - ▶ この計算に $p(X_1, \dots, X_t | k)$ が必要（詠み音声の生成モデル）
- 事後分布が特定の札に集中したら、選手は札を取る
  - ▶  $\pi(k | X_1, \dots, X_{t^*}) \approx 1$
  - ▶ この時点 $t = t^*$ が「真の決まり字」

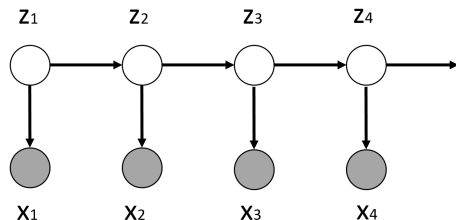


## 補足

- 選手によって札認識の早さが異なる（「感じ」の早さ）
- 各選手内の詠み音声の生成モデルの違いに起因するかも
  - ▶ 読手の生成モデルと選手の生成モデルが近いほど早く札が認識できる
  
- 空札( $k \notin K$ )のときは事後分布が1枚の札に集中しないはず
- 事後分布が集中する前に札に触るとお手つきになる
  - ▶ 例：「おお」で「おおえ」を触る
  
- 多くの選手は札を特定する前から手を動かしている
- 事後分布が高い場所へ手を出している？

# 隠れマルコフモデル (HMM)

- 系列データの統計モデル
  - ▶ 音声認識でよく用いられる
- 観測系列 $X = (X_t)$ の背後に「隠れ状態」系列 $Z = (Z_t)$ を仮定
  - ▶  $Z_t$ は有限個の値のみととする



- 状態 $Z_t$ はマルコフ連鎖で遷移
- 観測 $X_t$ は状態 $Z_t$ に対応した分布から発生

# 隠れマルコフモデル (HMM)

- パラメータ :  $\theta = (\pi, A, \phi)$ 
  - ▶  $\pi$ は初期状態の分布
  - ▶  $A$ はマルコフ連鎖の確率遷移行列
  - ▶  $\phi$ は各状態に対する出力分布のパラメータ

$$p(X, Z | \theta) = p(Z_1 | \pi) \left( \prod_{t=2}^T p(Z_t | Z_{t-1}, A) \right) \left( \prod_{t=1}^T p(X_t | Z_t, \phi) \right)$$

- $\theta$ の推定 : EMアルゴリズム

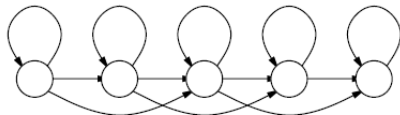
$$\hat{\theta} = \hat{\theta}(X_1, \dots, X_T)$$

- $Z_t$ の逐次推定 : フィルタリング

$$\begin{aligned} p(Z_t | X_1, \dots, X_t, \theta) &\rightarrow p(Z_{t+1} | X_1, \dots, X_t, \theta) \\ &\rightarrow p(Z_{t+1} | X_1, \dots, X_t, X_{t+1}, \theta) \end{aligned}$$

# 音声の隠れマルコフモデル

- 単語や音素 (/a/,/t/など) をHMMで表現
  - ▶ さらに前後の音素によって区別することもある (triphone)
  - ▶ 例: a-k+a, a-k+e, a-k+i, a-k+o, a-k+u
- 観測系列 $X$  = 音声波形から得られる特徴量の時系列
- 各状態は発声の1ステップを表現
  - ▶ 例: k-a-r-u-t-a (6状態)
  - ▶ もっと状態数を増やした方がよいかもしれない
- 時間の前後関係からマルコフ連鎖は下の形
  - ▶ left-to-right型 (状態は逆行しない)





## 札の事後分布の更新

- それぞれの札の周辺尤度を逐次計算可能
  - ▶  $p(Z_t | X_1, \dots, X_{t-1}, \theta)$  : 一期先予測分布

$$p(X_1, \dots, X_T | \theta) = p(X_1 | \theta) \prod_{t=2}^T p(X_t | X_1, \dots, X_{t-1}, \theta)$$

$$p(X_1 | \theta) = \sum_{Z_1} p(X_1 | Z_1, \theta) p(Z_1 | \theta)$$

$$p(X_t | X_1, \dots, X_{t-1}, \theta) = \sum_{Z_t} p(X_t | Z_t, \theta) p(Z_t | X_1, \dots, X_{t-1}, \theta)$$

- 周辺尤度を用いて札の事後分布を逐次更新

$$\pi(k | X_1, \dots, X_t) \propto \pi(k | X_1, \dots, X_{t-1}) p(X_t | X_1, \dots, X_{t-1}, \theta_k)$$

# データ取得

- 東京大学駒場キャンパス内の録音施設を使用
- 選手Aの「おおえ」、「おおけ」、「おおこ」の初句（最初5文字）の詠み音声を10回ずつ録音（44100Hzサンプリング）
  - ▶ （松田が）お手つきしやすい札
  - ▶ 正確には「おほえ」、「おほけ」、「あふこ」  
→ 3つ目だけ母音が違う？
- 詠む順番はランダムとした

はみま  
しすた  
たあふ  
てまみ  
のも

めまわ  
のにか  
そすた  
てみつ  
そそ

さをひ  
らもと  
まうを  
しらも  
みみ

# 参考

- 第58期クイーン戦1回戦

- ▶ <https://www.youtube.com/watch?v=RhLb7Ak1HWQ>

- 01:35 おおけ

- 07:14 おおこ

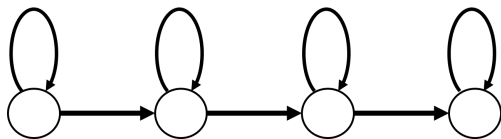
## 解析の手順

- 3首それぞれ10回ずつの音声データを使用
- 27個（3首×9回）の音声データから3つのHMMを推定
- 残り3つ（3首×1回）の音声データでテスト
  - ▶ 事前分布は $\pi(\text{ooe}) = \pi(\text{ooke}) = \pi(\text{ooko}) = \frac{1}{3}$
- 計算機に3枚の札の詠みを学習させた後，新たな詠みを「聞かせて」どの時点で識別できるか調べている



## 今回使ったモデル

- 「おおえやま」、「おおけなく」、「おおことの」のHMM
- 観測データ $X$ …音声特徴量の39次元時系列
  - ▶ L. M. LeeのMATLABプログラムを使用
  - ▶ メルケプストラムや対数パワースペクトルなど
- 出力分布…39次元の正規分布（共分散は対角行列）
- 状態のマルコフ遷移グラフは下の形

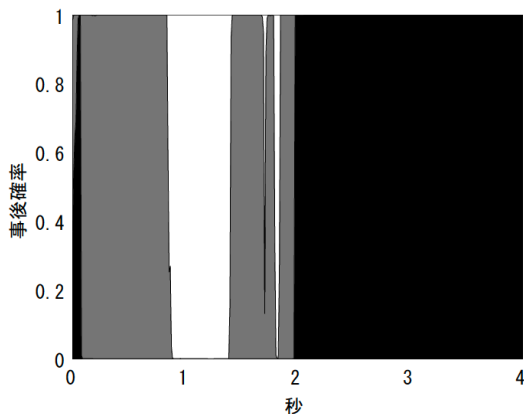




# 解析結果（おおえ）

- 事後分布の時間推移

- ▶ 黒：おおえ、灰色：おおけ、白：おおこ
- ▶ 縦軸：事後確率、横軸：秒

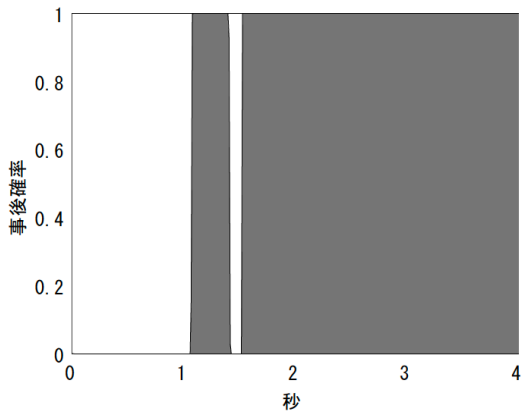


- 聞いてみると？

## 解析結果（おおけ）

- 事後分布の時間推移

- ▶ 黒：おおえ、灰色：おおけ、白：おおこ
- ▶ 縦軸：事後確率、横軸：秒



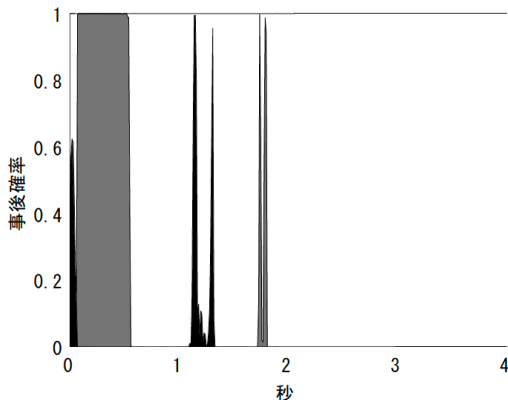
- 聞いてみると？



# 解析結果（おおこ）

## ● 事後分布の時間推移

- ▶ 黒：おおえ、灰色：おおけ、白：おおこ
- ▶ 縦軸：事後確率、横軸：秒



## ● 聞いてみると？

# 考察

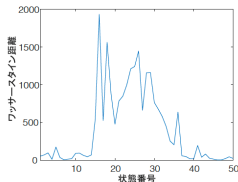
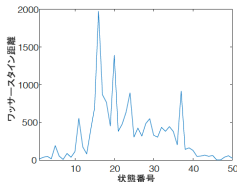
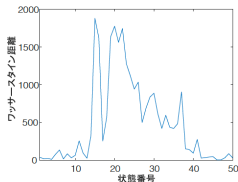
- ほとんどの時点で1枚の札に事後分布が偏る
- ある時点以降では正解の札に事後分布が集中する
  - ▶ 「ある時点」=決まり字!
- 聞いてみると「おお」の段階で正しく識別できている
  
- 決まり字直前で確信が揺らぐ
  - ▶ 実際に選手が体験していそうな現象

# 聞き分けてみると…

- おおえ
- おおけ
- おおこ

# 出力分布の比較

- 各状態の出力分布間の $L^2$ ワッサーズタイン距離
  - ▶ 左: 「おおえ」「おおけ」
  - ▶ 中: 「おおえ」「おおこ」
  - ▶ 右: 「おおけ」「おおこ」



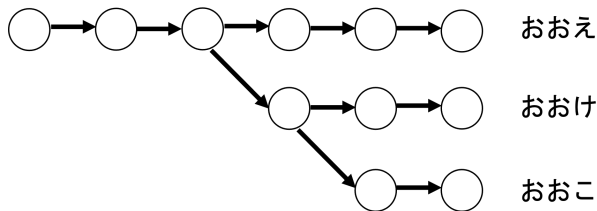
- どのペアでも15番目の状態付近でワッサーズタイン距離が増加  
→ 「おお」の終端に対応
- この時点での音声の違いを捉えることで札の識別を達成

## まとめ

- 統計モデルを用いて「真の決まり字」について解析した
  - ▶ 詠み音声の生成過程…隠れマルコフモデル
  - ▶ 選手が札を認識する過程…逐次ベイズ推定
- 「おおえ」「おおけ」「おおこ」を「おお」で識別できた
- 100首全部を取り扱いたい
  - ▶ 「さ」「しら」「しの」「す」「せ」のs音の違い？
  - ▶ 「たか」「たき」「たご」「たち」「たま」「たれ」の「た」の違い？
- 公認読手や専任読手の詠み音声に適用してみたい
  - ▶ 読手によらず正確に札が識別できるか？

## 今後の課題：百人一首の系統樹の推定

- 今回は札ごとに個別に隠れマルコフモデルを学習
- 共通の始状態から状態が分岐していくモデルの方が適切
  - ▶ 例：「おお」の途中までは共通

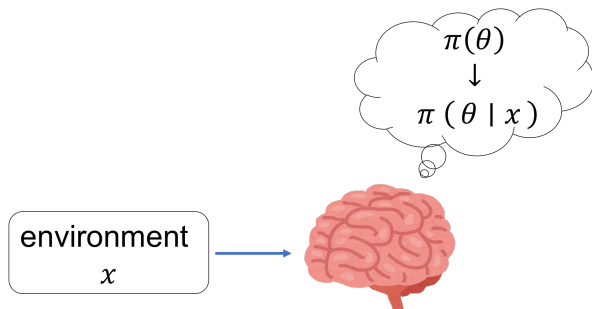


→ 百人一首の系統樹を明らかにできるか？

- 隠れマルコフモデルの状態遷移グラフ構造の学習
  - ▶ かるた以外にも応用できそう

## 今後の課題：かるたと脳

- 予測符号化(predictive coding)：脳内情報処理＝ベイズ推論



- 選手の情報処理過程を脳計測によって明らかにできるか？
  - ▶ 例：「おおえ」の「おお」と「おおこ」の「こ」を合成して聞かせるとミスマッチ陰性電位が発生するか？

# HMMのフィルタリング

- 観測をもとに状態を逐次推定
- 一期先予測分布 $p(Z_{t+1} | X_1, \dots, X_t, \theta)$ とフィルタリング分布 $p(Z_t | X_1, \dots, X_t, \theta)$ を交互に計算

$$p(Z_{t+1} | X_1, \dots, X_t, \theta) = \sum_{Z_t} p(Z_{t+1} | Z_t, \theta) p(Z_t | X_1, \dots, X_t, \theta)$$

$$p(Z_t | X_1, \dots, X_t, \theta) = \frac{p(Z_t | X_1, \dots, X_{t-1}, \theta) p(X_t | Z_t, \theta)}{\sum_{Z_t} p(Z_t | X_1, \dots, X_{t-1}, \theta) p(X_t | Z_t, \theta)}$$



# HMMの周辺尤度の計算

- 状態数選択のためにHMMの周辺尤度の計算が必要
- 一期先予測分布 $p(Z_t | X_1, \dots, X_{t-1}, \theta)$ を用いて計算可能

$$p(X_1, \dots, X_T | \theta) = p(X_1 | \theta) \prod_{t=2}^T p(X_t | X_1, \dots, X_{t-1}, \theta)$$

$$p(X_1 | \theta) = \sum_{Z_1} p(X_1 | Z_1, \theta) p(Z_1 | \theta)$$

$$p(X_t | X_1, \dots, X_{t-1}, \theta) = \sum_{Z_t} p(X_t | Z_t, \theta) p(Z_t | X_1, \dots, X_{t-1}, \theta)$$

# HMMのパラメータ推定

- EMアルゴリズムによってパラメータ $\theta$ をデータから推定

$$Q(\theta, \theta^{(k)}) = \sum_Z p(Z | X, \theta^{(k)}) \log p(X, Z | \theta)$$

$$\theta^{(k+1)} = \arg \max_{\theta} Q(\theta, \theta^{(k)})$$

- $\theta^{(k)}$ の更新式は解析的に解ける