スペシャルセッション
「音楽情報処理と機械学習」

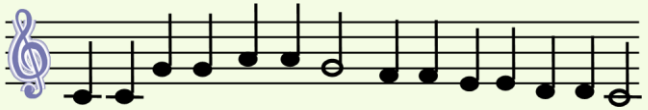
歌声情報処理：歌唱スタイルモデル

NTT コミュニケーション科学基礎研究所
メディア情報研究部　大石康智

研究の根底にある興味

「きらきら星」

★音符の高さと長さ



★歌詞

きらきら ひかる
おそらの ほしよ

★うまさ

響きのある声質，音量変化，装飾物，音符からの逸脱

⇒ **意識的に制御する**

★らしさ

身体的・心理的な状態

⇒ **意識的に制御しない**

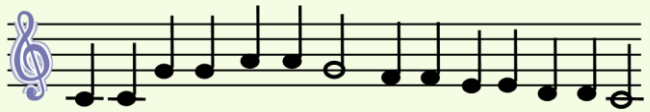
研究の根底にある興味

歌い手の「うまさ」や「らしさ」のモデリング

（記号や文字として，書き起こせない情報）

「きらきら星」

★音符の高さと長さ



★歌詞

きらきら ひかる

おそらの ほしよ

これまでの音楽情報処理

★うまさ

響きのある声質，音量変化，装飾物，音符からの逸脱

⇒ **意識的に制御する**

★らしさ

身体的・心理的な状態

⇒ **意識的に制御しない**

雑音とみなされがち

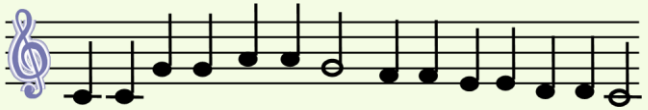
研究の根底にある興味

歌い手の「うまさ」や「らしさ」のモデリング

（記号や文字として，書き起こせない情報）

「きらきら星」

★音符の高さと長さ



★歌詞

きらきら ひかる

おそらの ほしよ

これまでの音楽情報処理

★うまさ

響きのある声質，音量変化，装飾物，音符からの逸脱

⇒ **意識的に制御する**

★らしさ

身体的・心理的な状態

⇒ **意識的に制御しない**

雑音とみなされがち

★楽器演奏や話声でも同じことが言える

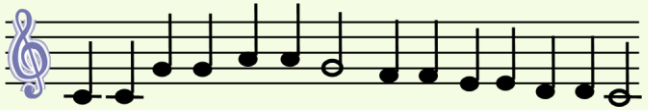
研究の根底にある興味

歌い手の「うまさ」や「らしさ」のモデリング

（記号や文字として，書き起こせない情報）

「きらきら星」

★音符の高さと長さ



★歌詞

きらきら ひかる
おそらの ほしよ

これまでの音楽情報処理

★うまさ

響きのある声質，音量変化，装飾物，音符からの逸脱

⇒ **意識的に制御する**

★らしさ

身体的・心理的な状態

⇒ **意識的に制御しない**

雑音とみなされがち

★楽器演奏や話声でも同じことが言える

アプローチ

①大量データの利用，② 現象の物理的特性の考慮

機械学習を用いた研究の進展

(1)「歌声と話声の違い」を探る

(2)「音高(F_0)の動き」を探る

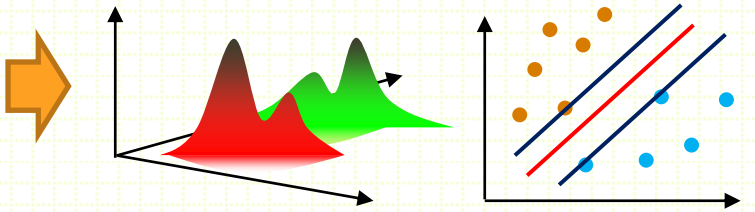
(3)「F_0軌跡の生成過程」を探る

機械学習を用いた研究の進展

(1)「歌声と話声の違い」を探る

通常の話声と大きく違う，歌声特有の音響的特徴にこそ，歌い手の「らしさ」や「うまさ」が隠される！

ガウス混合モデル（GMM）
サポートベクターマシン（SVM）



(2)「音高（F_0）の動き」を探る

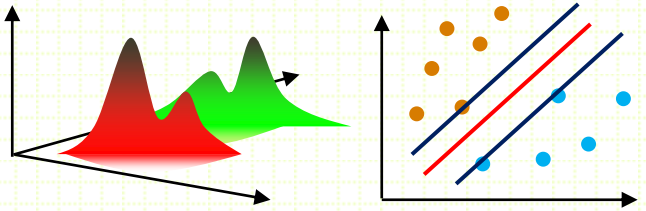
(3)「F_0軌跡の生成過程」を探る

機械学習を用いた研究の進展

(1)「歌声と話声の違い」を探る

通常の話声と大きく違う，歌声特有の音響的特徴にこそ，歌い手の「らしさ」や「うまさ」が隠される！

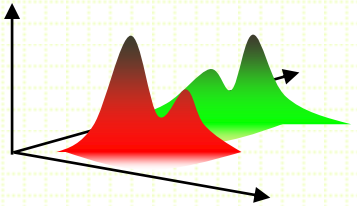
ガウス混合モデル（GMM）
サポートベクターマシン（SVM）



(2)「音高（F_0）の動き」を探る

ビブラートをはじめ，様々な動的変動成分が含まれるF_0の動きを可視化して分析したい！

ガウス混合モデル（GMM）



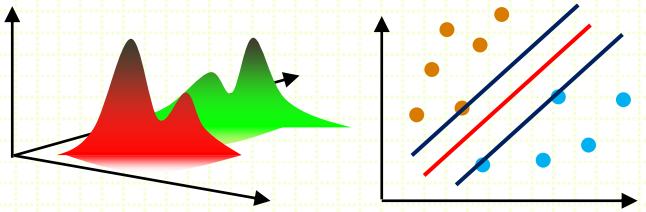
(3)「F_0軌跡の生成過程」を探る

機械学習を用いた研究の進展

(1)「歌声と話声の違い」を探る

通常の話声と大きく違う，歌声特有の音響的特徴にこそ，歌い手の「らしさ」や「うまさ」が隠される！

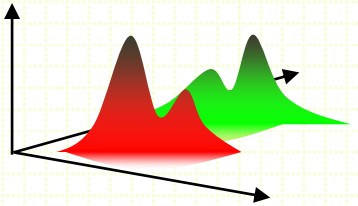
ガウス混合モデル（GMM）
サポートベクターマシン（SVM）



(2)「音高（F_0）の動き」を探る

ビブラートをはじめ，様々な動的変動成分が含まれるF_0の動きを可視化して分析したい！

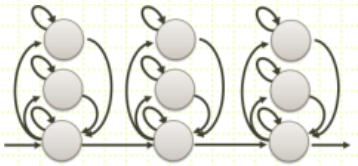
ガウス混合モデル（GMM）

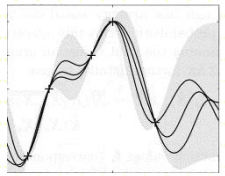


(3)「F_0軌跡の生成過程」を探る

発声器官を模擬した物理モデルを利用して，歌い手の「うまさ」や「らしさ」を分離して特徴抽出する！

隠れマルコフモデル（HMM）
EMアルゴリズム，ガウス過程





引用：Gaussian processes for Machine Learning

(1)「歌声と話声の違い」を探る

- 話声と大きく異なる，歌声特有の音響的特徴の調査

  ⇒ 歌い手の「らしさ」や「うまさ」が隠されている特徴を探す！

  - 声質：メル周波数ケプストラム係数（MFCC），ΔMFCC
  - 音高：基本周波数（F_0），ΔF_0

(1)「歌声と話声の違い」を探る

- 話声と大きく異なる，歌声特有の音響的特徴の調査

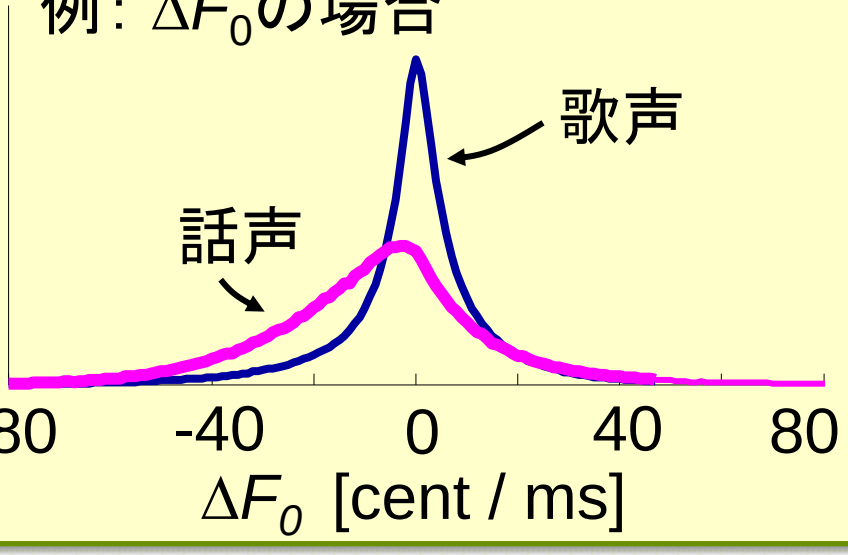
  ⇒ 歌い手の「らしさ」や「うまさ」が隠されている特徴を探す！

  - 声質：メル周波数ケプストラム係数（MFCC），ΔMFCC
  - 音高：基本周波数（F_0），ΔF_0

機械学習の導入

⇒ **GMM**による分布の学習

例：ΔF_0の場合



(1)「歌声と話声の違い」を探る

- 話声と大きく異なる，歌声特有の音響的特徴の調査

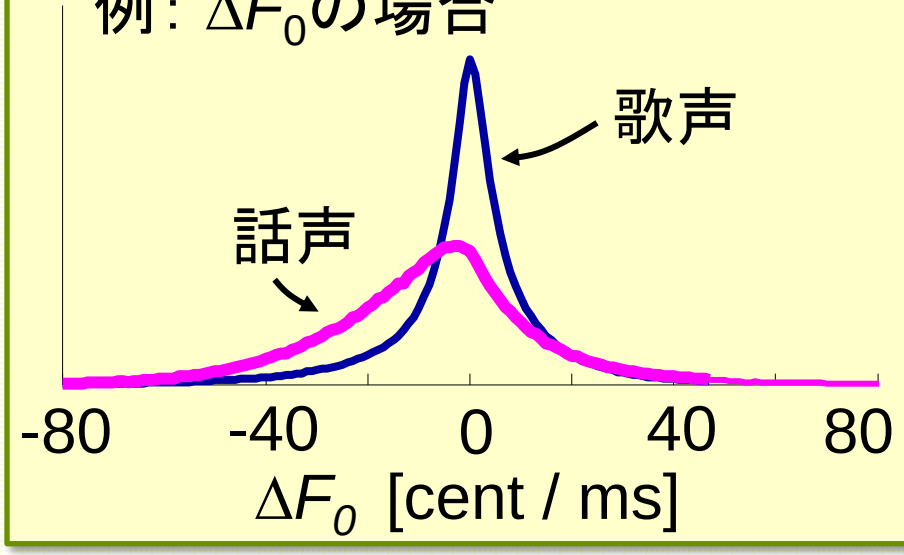
  ⇒ 歌い手の「らしさ」や「うまさ」が隠されている特徴を探す！

  - 声質：メル周波数ケプストラム係数（MFCC），ΔMFCC
  - 音高：基本周波数（F_0），ΔF_0

機械学習の導入

⇒ **GMM**による分布の学習

例：ΔF_0の場合



事後確率に基づく識別

	識別率（音声1秒）
MFCC	67.5%
ΔMFCC	78.0%
ΔF_0	**80.0%**

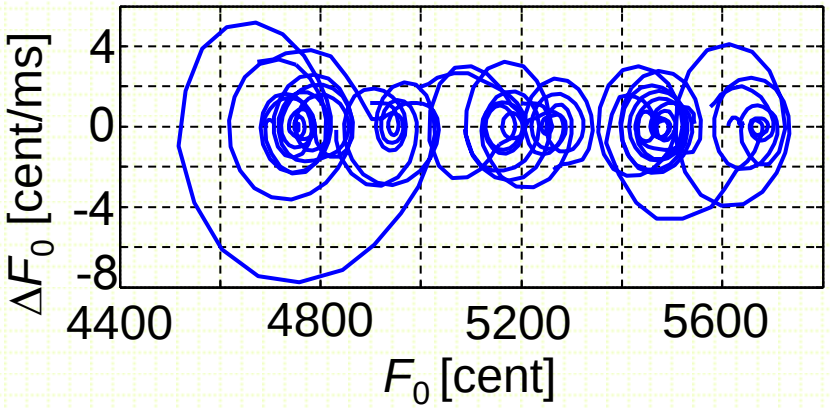
※人間の歌声／話声の識別能力と整合する結果

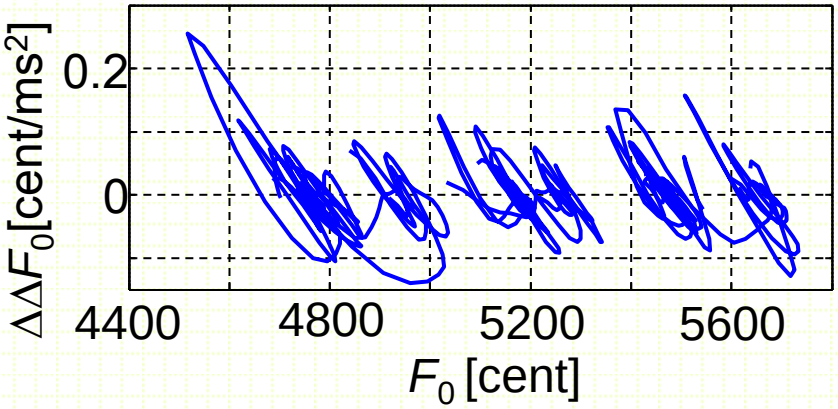
F_0の動きを分析しよう！

(2)「音高(F_0)の動き」を探る

- 相空間(F_0, ΔF_0, $\Delta\Delta F_0$, …で構成される空間)の利用

  ⇒ **渦(アトラクタ)**が「うまさ」や「らしさ」を特徴づける！

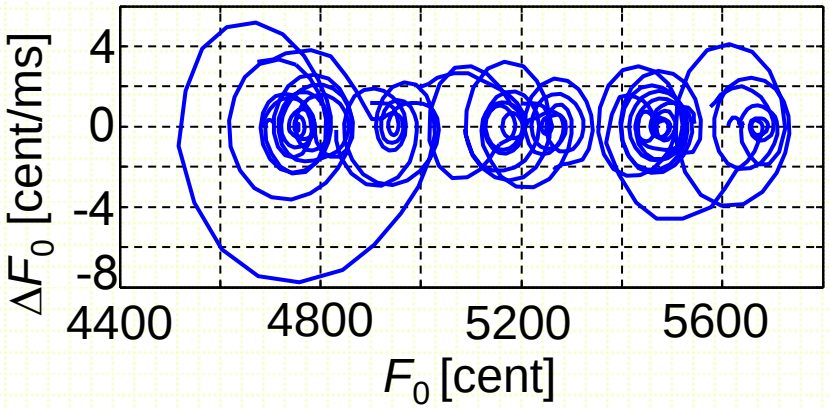


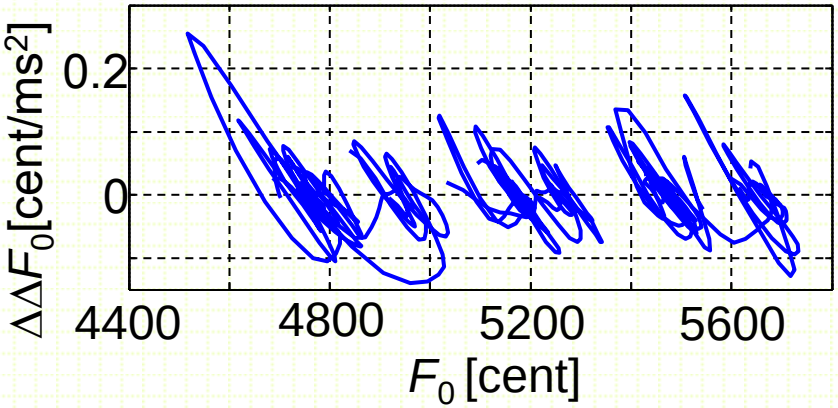


(2)「音高(F_0)の動き」を探る

- 相空間(F_0, ΔF_0, $\Delta\Delta F_0$, …で構成される空間)の利用

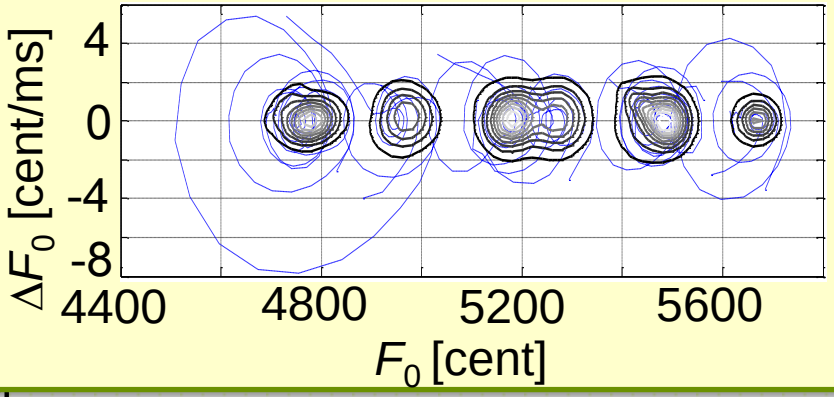
  ⇒ **渦(アトラクタ)**が「うまさ」や「らしさ」を特徴づける！





機械学習の導入

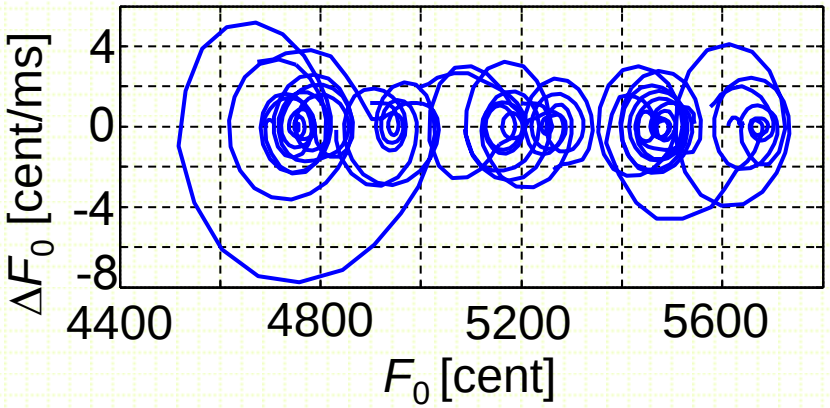
⇒ **GMM**による分布の学習

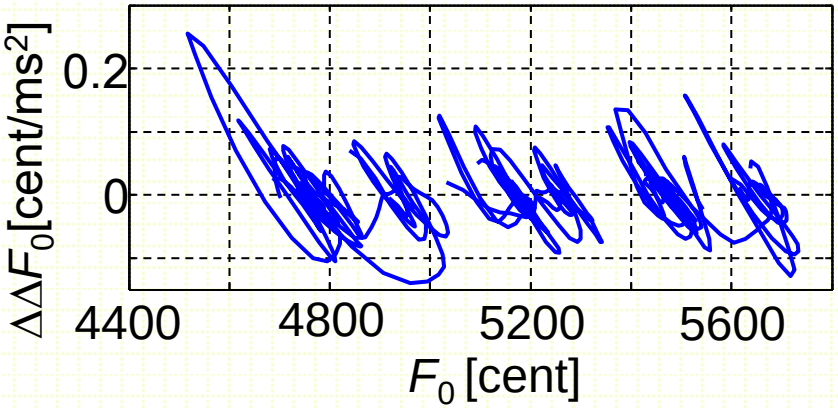


(2)「音高(F_0)の動き」を探る

- 相空間(F_0, ΔF_0, $\Delta\Delta F_0$, …で構成される空間)の利用

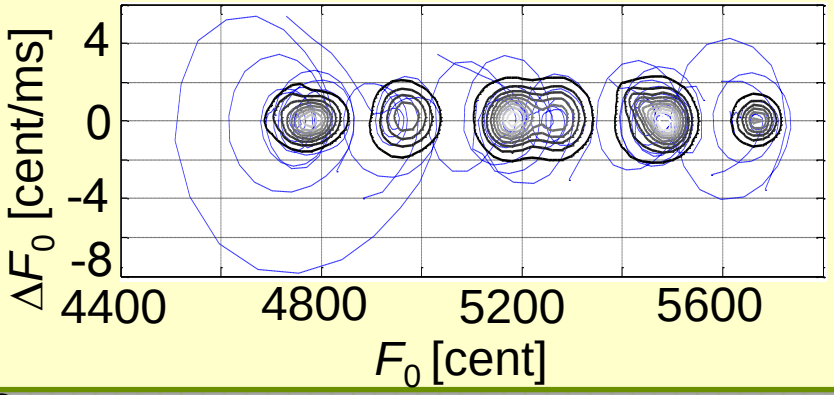
  ⇒ **渦(アトラクタ)**が「うまさ」や「らしさ」を特徴づける！





機械学習の導入

⇒ **GMM**による分布の学習



事後確率に基づく識別

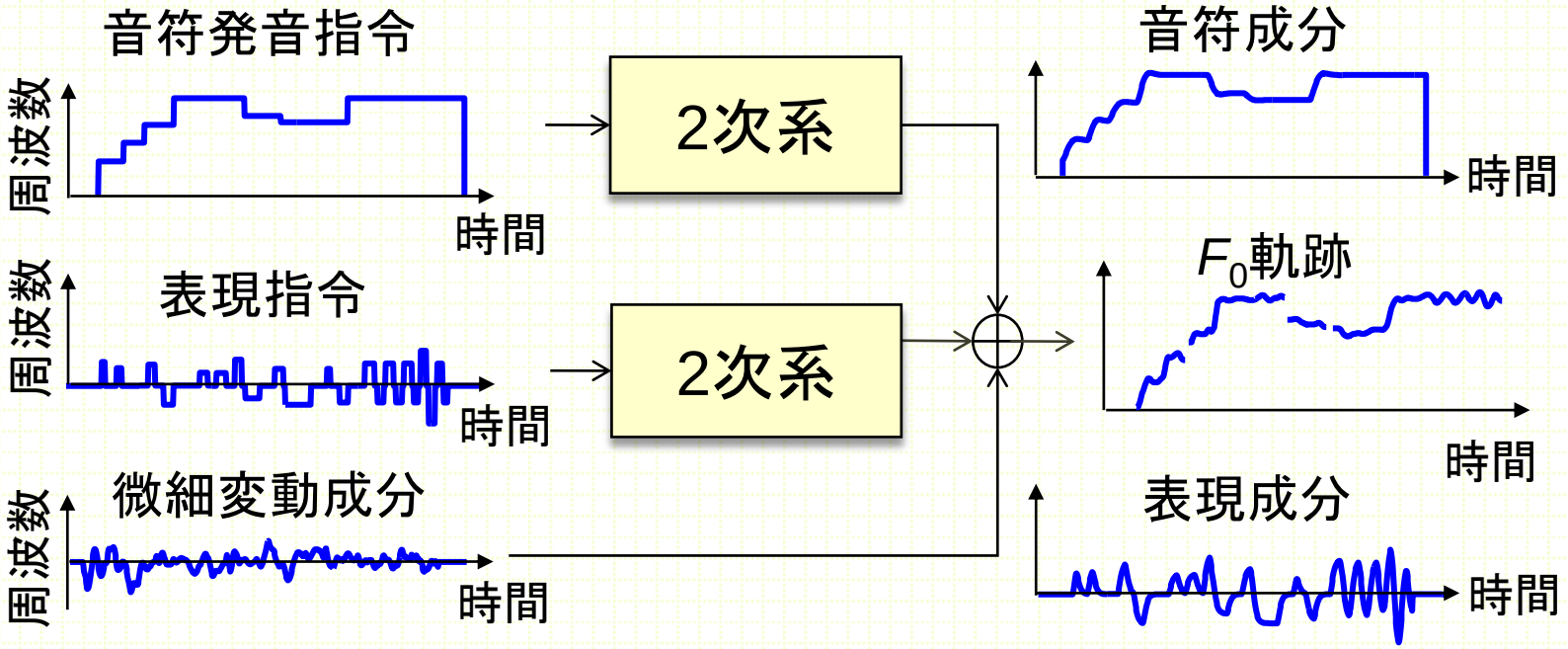
相空間	カテゴリ識別率
MFCC	78.5%
F_0	**90.0%**

軌跡をモデル化したい！

(3)「F_0軌跡の生成過程」を探る

- F_0軌跡の生成モデル(藤崎モデル，齋藤モデル)を参考

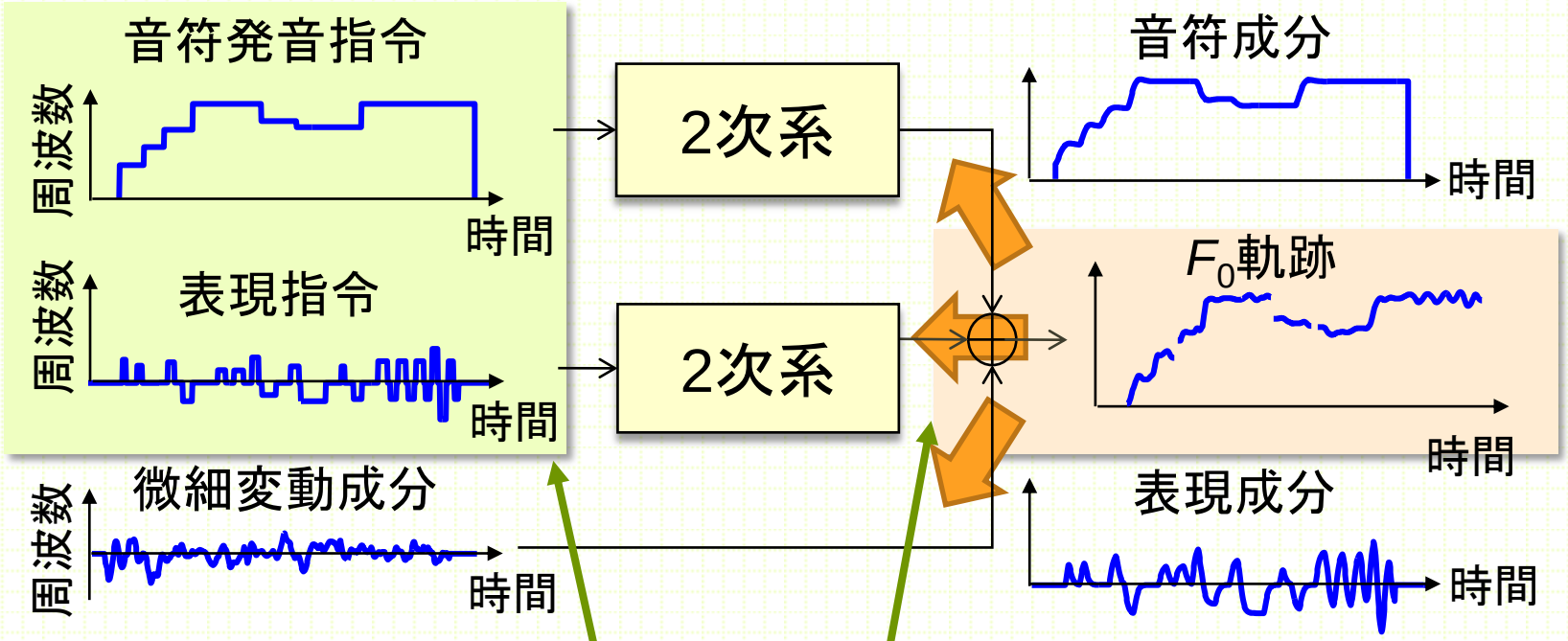
  ⇒ 発声器官の物理制約の下，「うまさ」や「らしさ」の特徴抽出



(3)「F_0軌跡の生成過程」を探る

- F_0軌跡の生成モデル(藤崎モデル，齋藤モデル)を参考

  ⇒ 発声器官の物理制約の下，「うまさ」や「らしさ」の特徴抽出



機械学習の導入

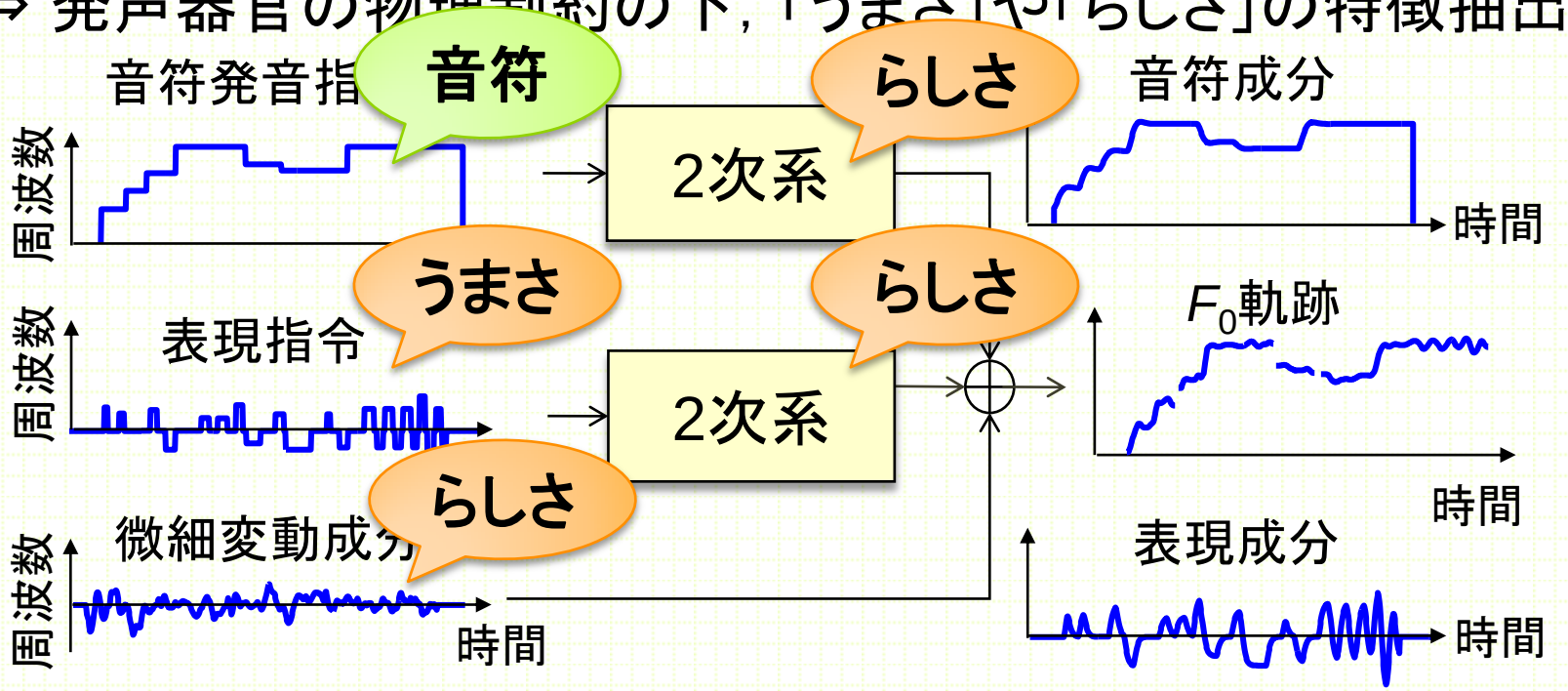
⇒ **HMM**による指令モデル

EM法によるパラメータ推定

(3)「F_0軌跡の生成過程」を探る

- F_0軌跡の生成モデル(藤崎モデル，齋藤モデル)を参考

  ⇒ 発声器官の物理制約の下，「うまさ」や「らしさ」の特徴抽出



機械学習の導入

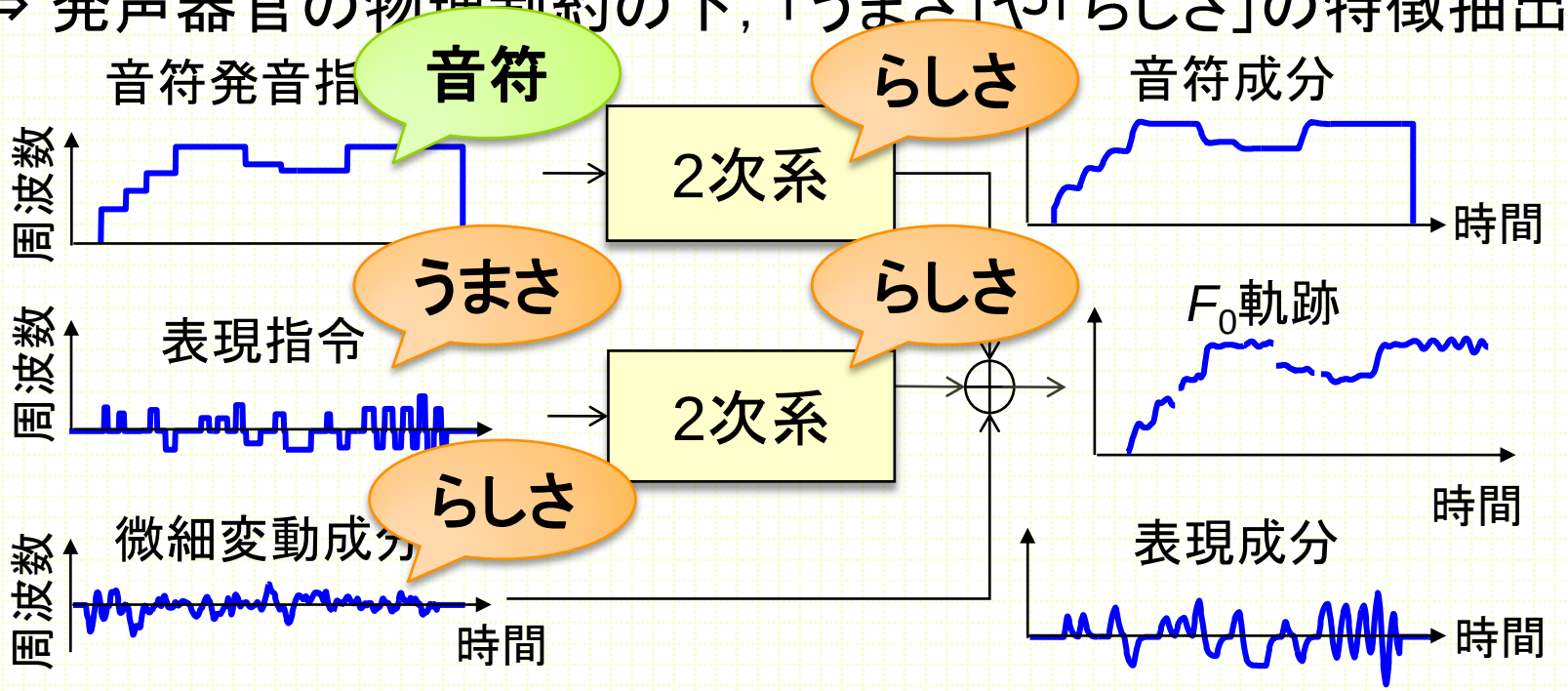
⇒ **HMM**による指令モデル

EM法によるパラメータ推定

(3)「F_0軌跡の生成過程」を探る

- F_0軌跡の生成モデル(藤崎モデル，齋藤モデル)を参考

  ⇒ 発声器官の物理制約の下，「うまさ」や「らしさ」の特徴抽出



機械学習の導入

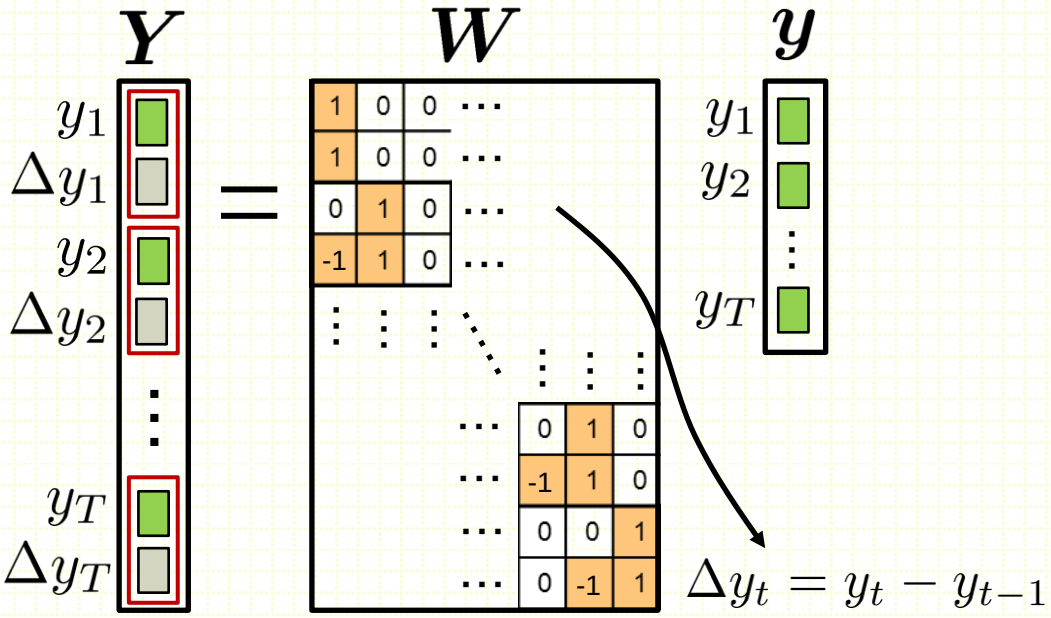
⇒ **HMM**による指令モデル

EM法によるパラメータ推定

認識・検索・合成の観点から評価中

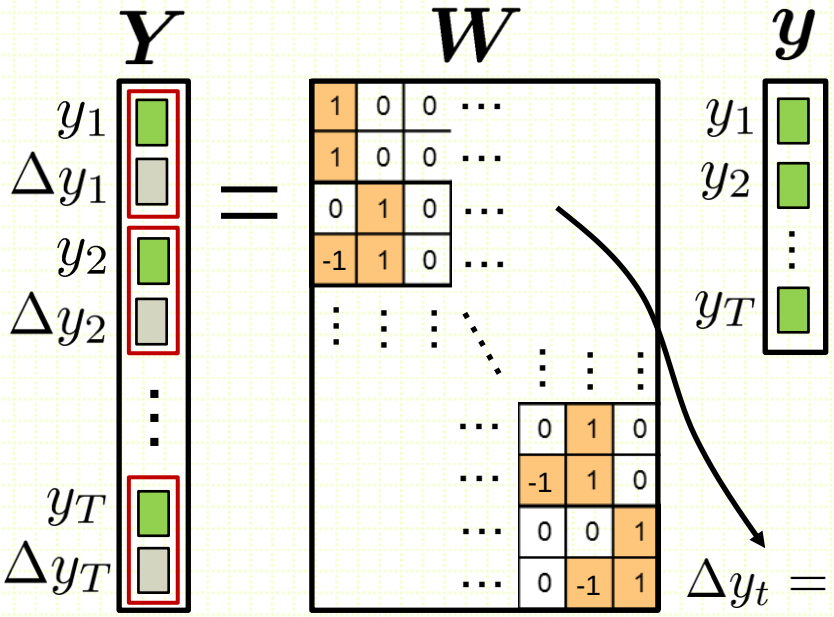
補足1:トラジェクトリモデルとの関係

- 静的特徴量系列ベクトルyを確率変数としたモデルの尤度最大化

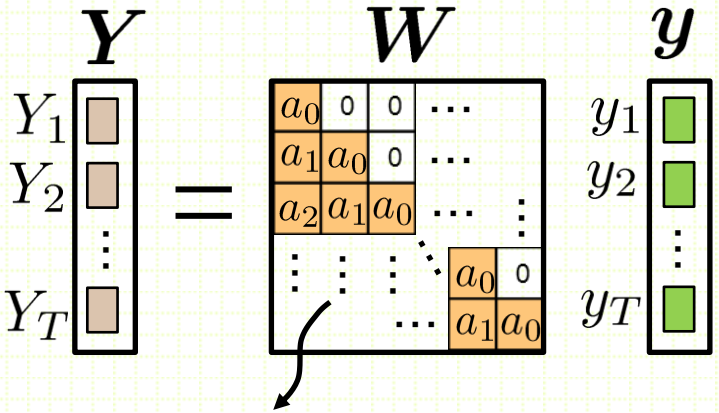


補足1:トラジェクトリモデルとの関係

- 静的特徴量系列ベクトルyを確率変数としたモデルの尤度最大化



F_0軌跡の確率モデル



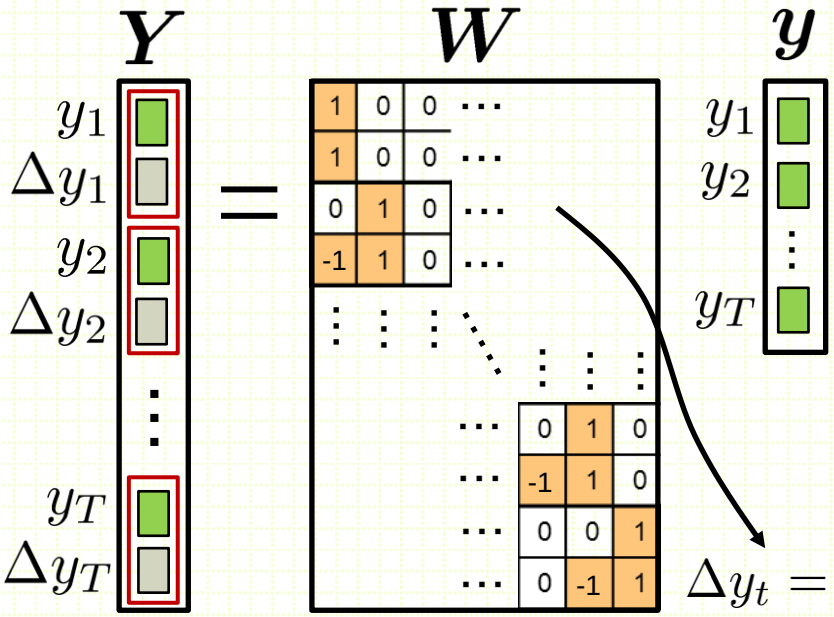
2次系の微分方程式の差分近似

$$Y_t = a_0 y_t + a_1 y_{t-1} + a_2 y_{t-2}$$

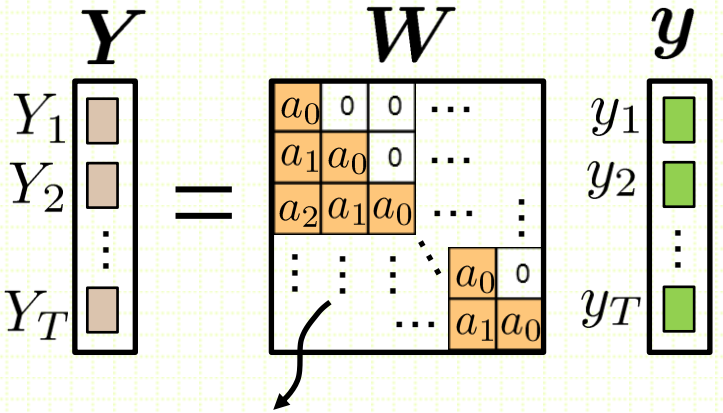
a_0, a_1, a_2 : 固定 or 可変

補足1:トラジェクトリモデルとの関係

- 静的特徴量系列ベクトル$\boldsymbol{y}$を確率変数としたモデルの尤度最大化



F_0軌跡の確率モデル



2次系の微分方程式の差分近似

$$Y_t = a_0 y_t + a_1 y_{t-1} + a_2 y_{t-2}$$

a_0, a_1, a_2 : 固定 or 可変

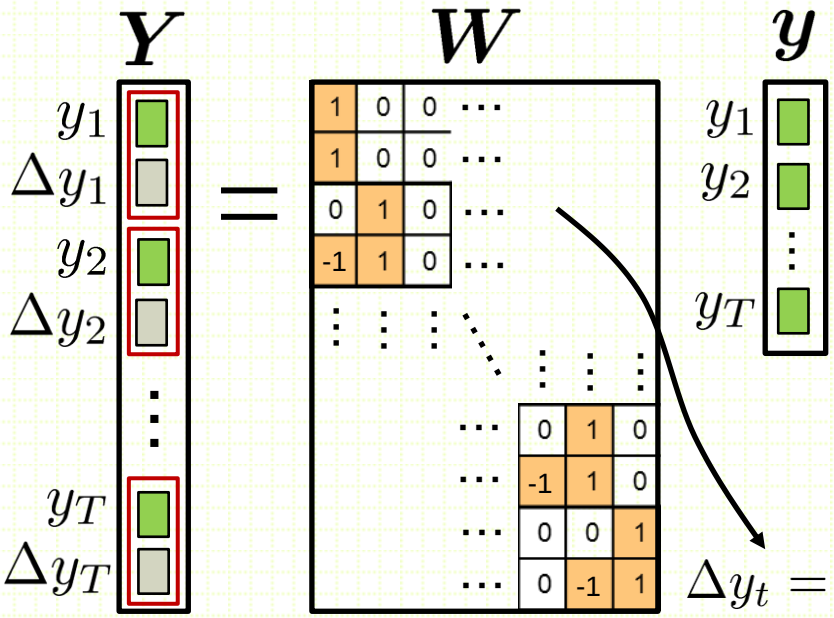
出力確率分布を単一ガウス分布とするHMMを想定すると,

$$P(\boldsymbol{y}|\ \boldsymbol{q},\ \boldsymbol{\Lambda}) = \mathcal{N}\left((\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{q}}^{-1}\boldsymbol{W})^{-1}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{q}}^{-1}\boldsymbol{\mu}_{\boldsymbol{q}},\ (\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{q}}^{-1}\boldsymbol{W})^{-1}\right)$$

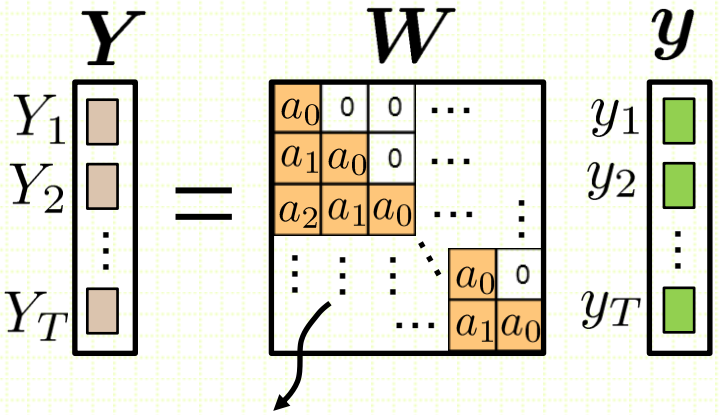
状態系列　HMMパラメータ　状態系列から決まる平均・分散

補足1:トラジェクトリモデルとの関係

- 静的特徴量系列ベクトル$\boldsymbol{y}$を確率変数としたモデルの尤度最大化



F_0軌跡の確率モデル



2次系の微分方程式の差分近似

$$Y_t = a_0 y_t + a_1 y_{t-1} + a_2 y_{t-2}$$

a_0, a_1, a_2 : 固定 or 可変

出力確率分布を単一ガウス分布とするHMMを想定

ガウス過程

$$P(\boldsymbol{y}|\ \boldsymbol{q},\ \boldsymbol{\Lambda}) = \mathcal{N}\left((\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{q}}^{-1}\boldsymbol{W})^{-1}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{q}}^{-1}\boldsymbol{\mu}_{\boldsymbol{q}},\ (\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{q}}^{-1}\boldsymbol{W})^{-1}\right)$$

状態系列　HMMパラメータ　状態系列から決まる平均・分散

$\boldsymbol{K}$: カーネル関数を用いて計算する

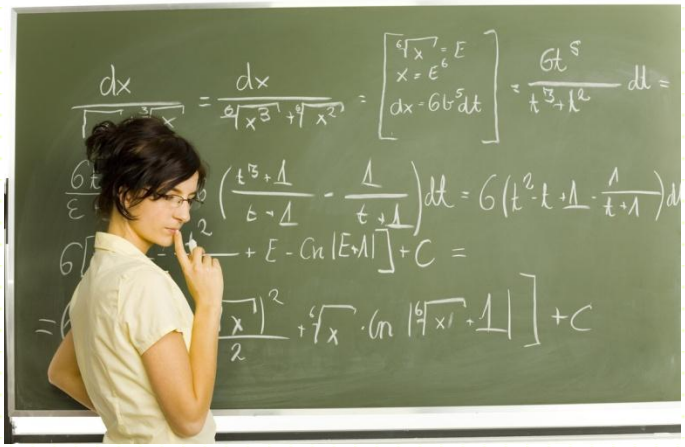
補足2:デモ

- 「ビブラート」や「こぶし」を単に検出するだけでなく,
  表現の意図として詳細に取り出せる
- どんなメロディにも「うまさ」や「らしさ」を転写できる

機械学習の効果と限界

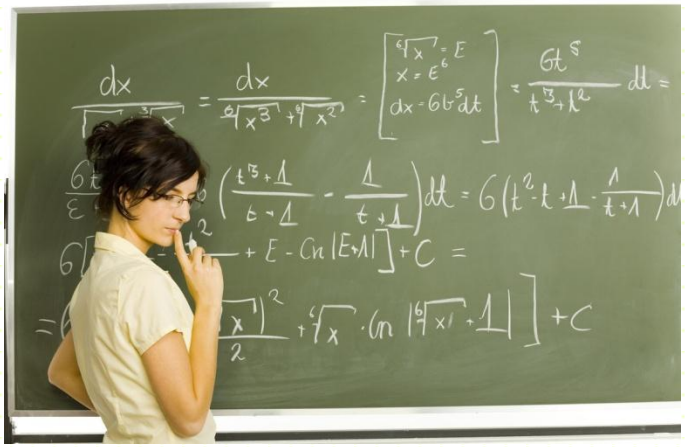
機械学習の効果と限界

- 効果
  - 大量のデータを用いさえすれば，**潜在する知識や特性をデータに語らせる**ことができる
  - 予測できる



機械学習の効果と限界

- 効果
  - 大量のデータを用いさえすれば，**潜在する知識や特性をデータに語らせる**ことができる
  - 予測できる



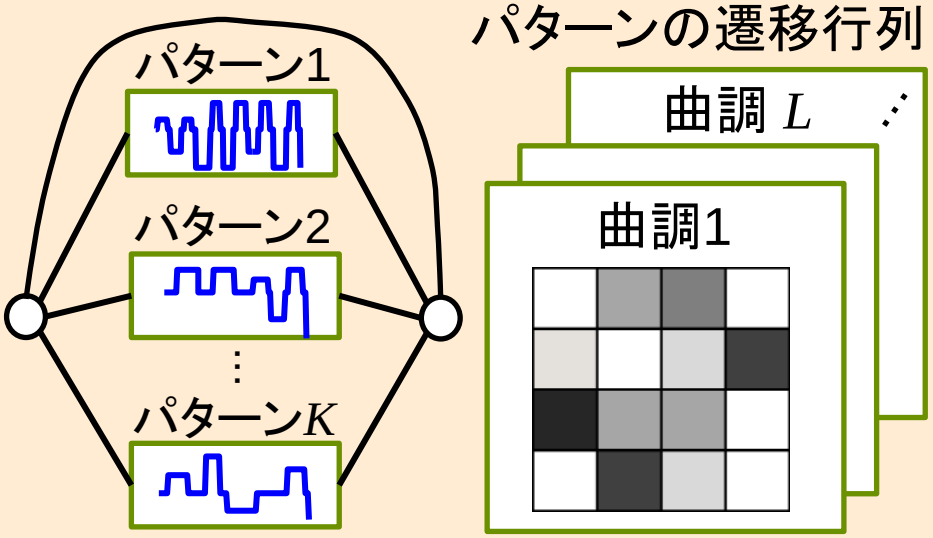
- 重要なこと
  - 音響学や生理学などの知見に基づいて，**信号の特徴やその背後にある生成過程を考慮すること**
  - **人間の知覚**と整合する結果が得られるか確認すること
  - 数理的に記述できたからといって，モデルパラメータの学習がうまくいくとは限らない，**前処理・初期状態決定**も重要
  - 大規模データを扱うため，**実装方法にも工夫**が必要

今後の展開および実用化に向けて

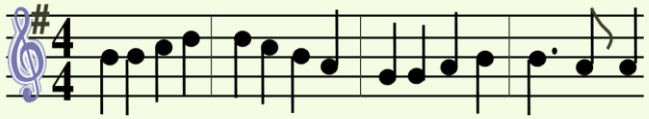
今後の展開および実用化に向けて

- 音韻・音量も含め，「うまさ」のバリエーションの学習

歌い手Aの「うまさ」モデル



楽譜と歌詞

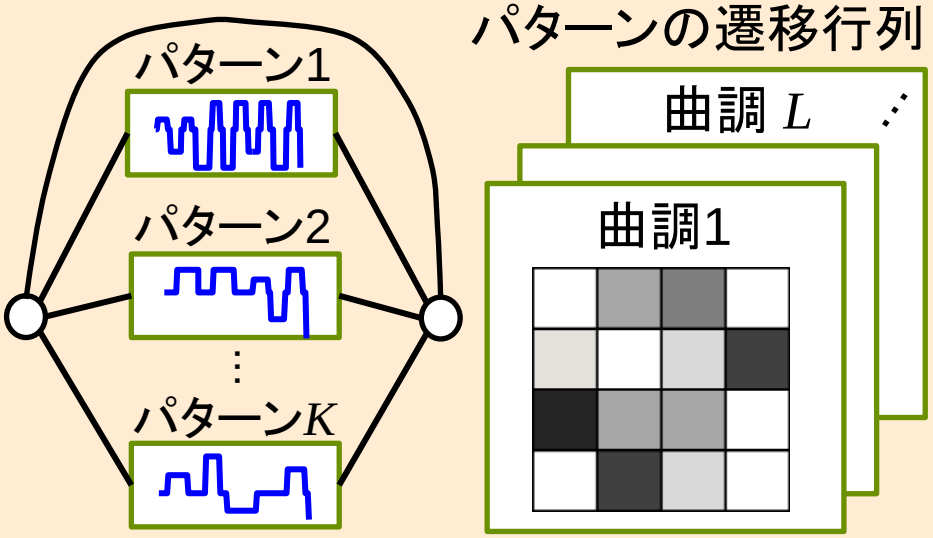


はれたる あおぞら ただよう くもよ

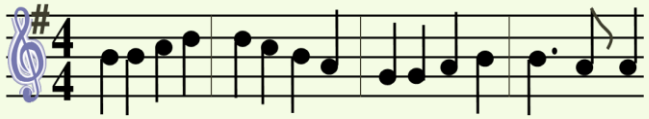
今後の展開および実用化に向けて

- 音韻・音量も含め，「うまさ」のバリエーションの学習

歌い手Aの「うまさ」モデル

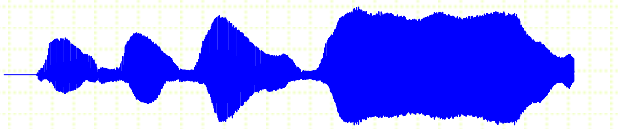


楽譜と歌詞



はれたる あおぞら ただよう くもよ

歌声音響信号

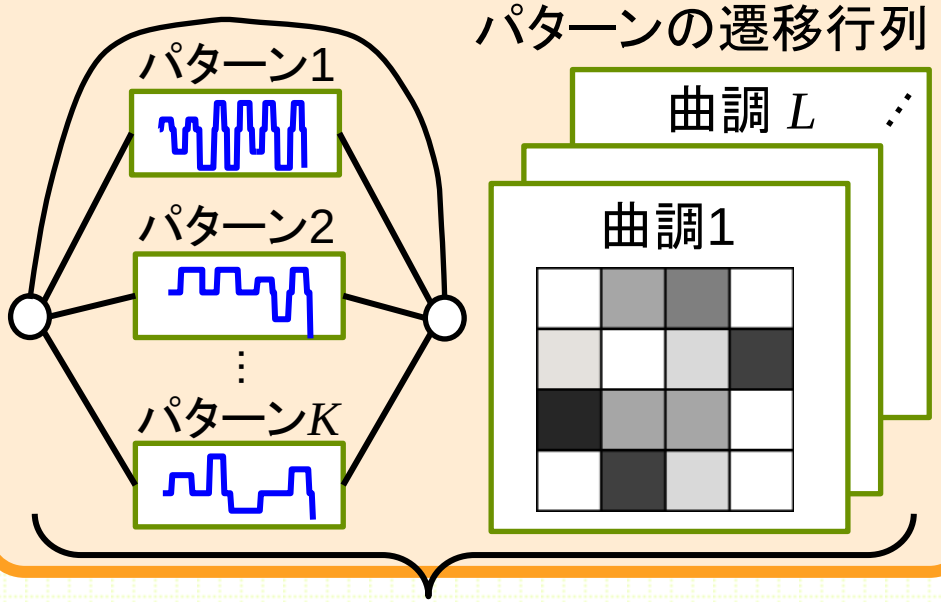


今後の展開および実用化に向けて

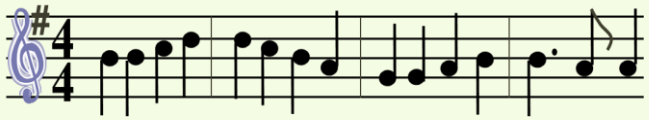
- 音韻・音量も含め，「うまさ」のバリエーションの学習

歌い手Aの「うまさ」モデル

パターンの遷移行列

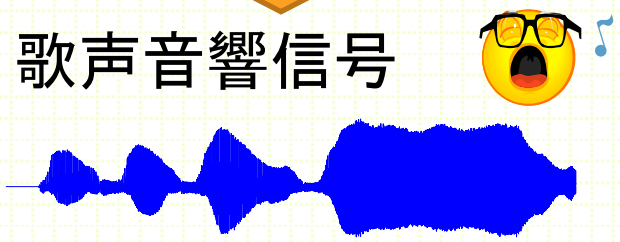


楽譜と歌詞



はれたる あおぞら ただよう くもよ

歌声音響信号



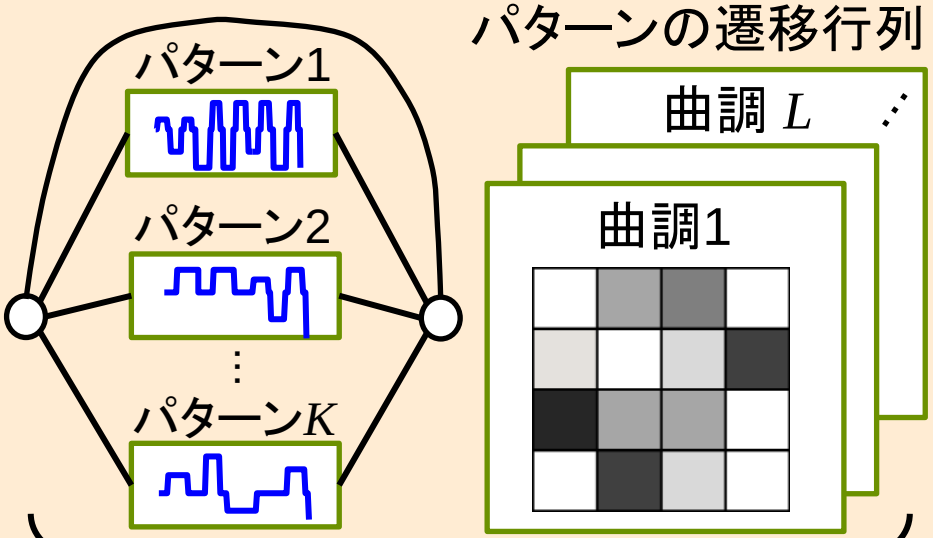
パターンの数が未知 ⇒ ノンパラメトリックベイズ法？
（モデルの複雑度も大規模データから推定する）

今後の展開および実用化に向けて

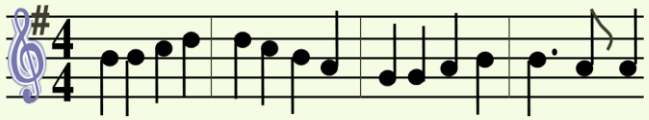
- 音韻・音量も含め，「うまさ」のバリエーションの学習

歌い手Aの「うまさ」モデル

パターンの遷移行列

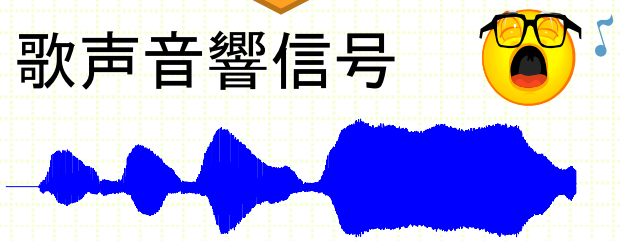


楽譜と歌詞



はれたる あおぞら ただよう くもよ

歌声音響信号

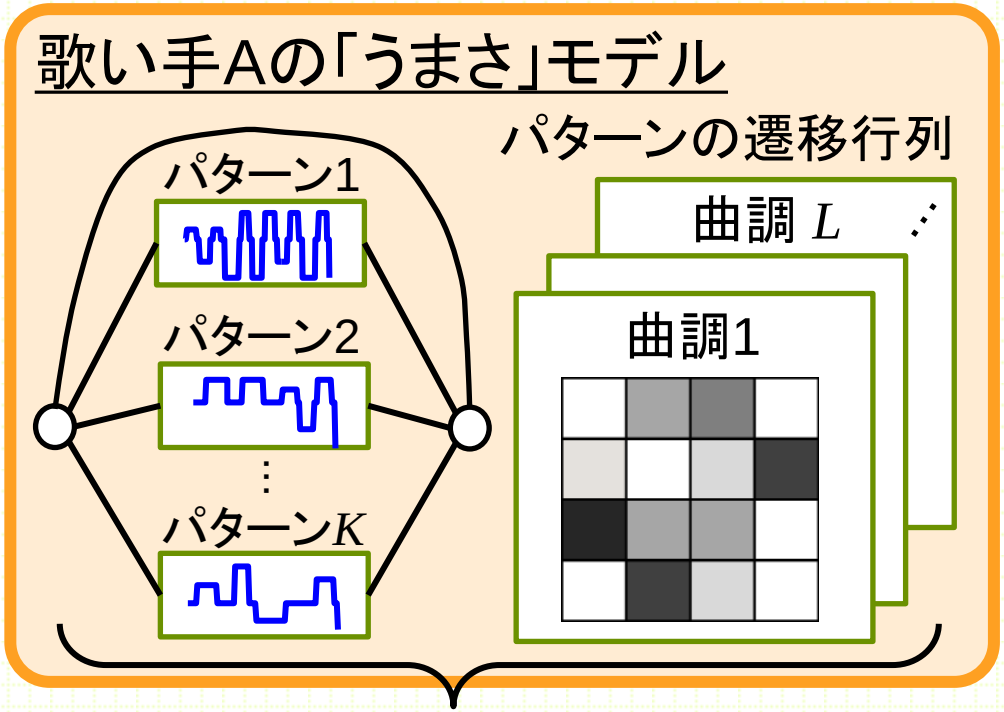


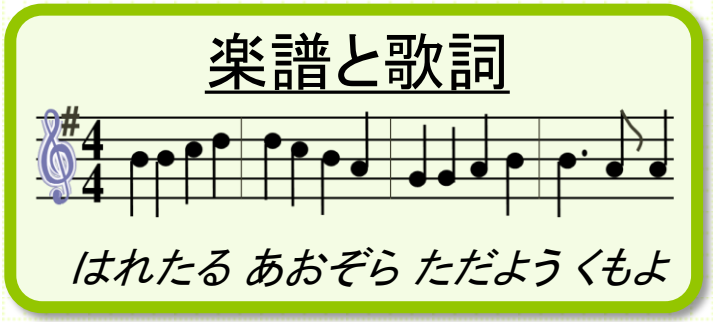
パターンの数が未知 ⇒ ノンパラメトリックベイズ法？
（モデルの複雑度も大規模データから推定する）

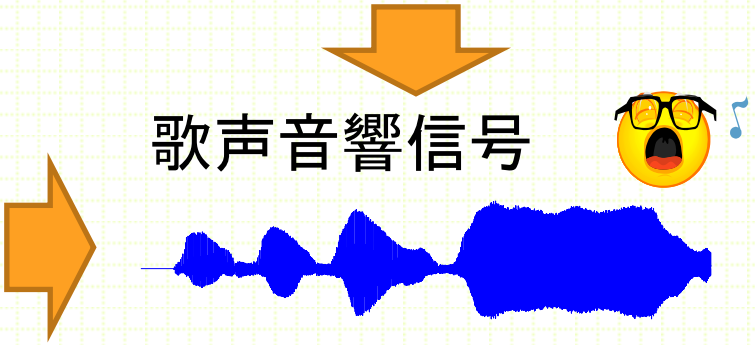
有名歌手※ の「新曲」の
歌声を自動的に作り出す!

今後の展開および実用化に向けて

- 音韻・音量も含め，「うまさ」のバリエーションの学習







パターンの数が未知 ⇒ ノンパラメトリックベイズ法？
（モデルの複雑度も大規模データから推定する）

有名歌手※ の「新曲」の
歌声を自動的に作り出す!

人間を知りたい
「知」を流通させたい