

トピックモデルを用いた 歌声特徴量の分析

中野 倫靖, 吉井 和佳, 後藤 真孝
(産業技術総合研究所)

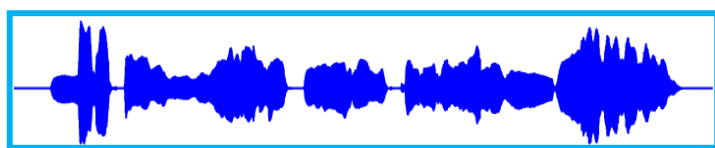
2013年9月1日

情報処理学会 音楽情報科学研究会 第100回記念シンポジウム

研究の背景

歌の特性を定量的に説明(モデル化)したい

- 歌手毎の歌い方の違いや類似性とは何か
 - 例) 違う楽曲でも同じ歌手なら歌い方が似ている
同じ楽曲でも違う歌手だと歌い方が違う



歌声



処理

歌い方モデル
(定量的な説明)

□ 歌い方の違い

着目

[鈴木 他, 2011] [齋藤, 榊原, 2011]

- 他人の歌い方を真似ると音高(F0)変化、音量変化、

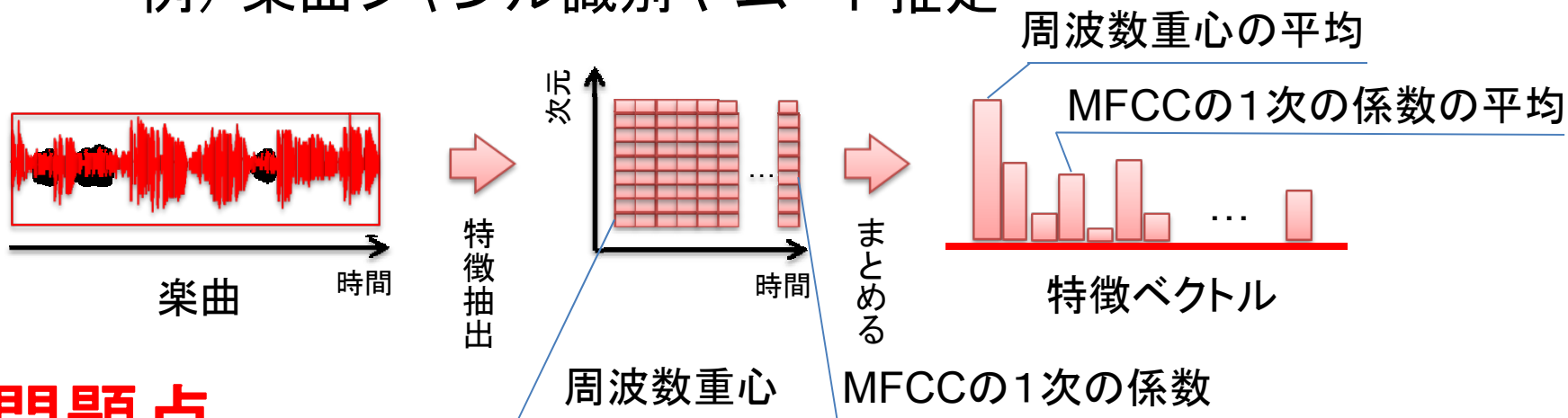
着目

声質(スペクトル包絡)、音韻継続長などが変化

楽曲を定量的に説明する方法

□ 平均的な特徴ベクトルを抽出

- 周波数重心の**平均**や**分散**、MFCCの**平均**、等
 - 例) 楽曲ジャンル識別やムード推定

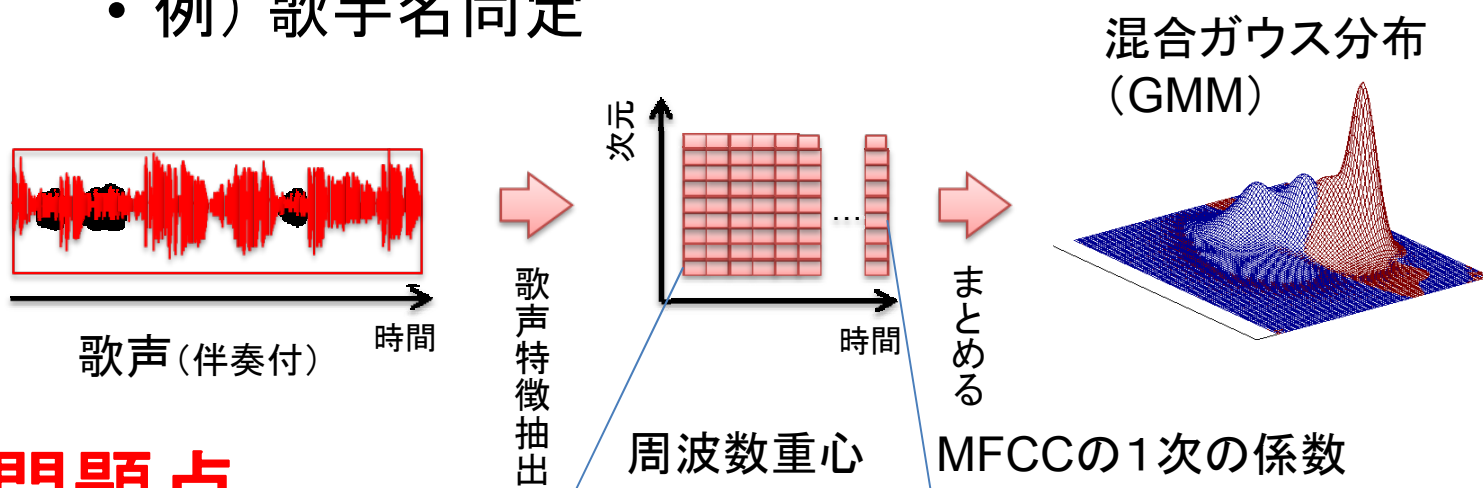


□ 問題点

- 時系列特徴のまとめ方は十分か？ 例) 平均？分散？
- 特徴量(次元)間の関係が使われない 有益な可能性

歌声をモデル化する一つの方法

- 歌声毎に特徴ベクトルからGMMを学習
 - 周波数重心やMFCC等の特徴ベクトルをモデル化
 - 例) 歌手名同定



□ 問題点

- 得られた結果の意味が分かりにくい
- 他の歌声を考慮していない

提案:トピックモデルで歌声を説明する

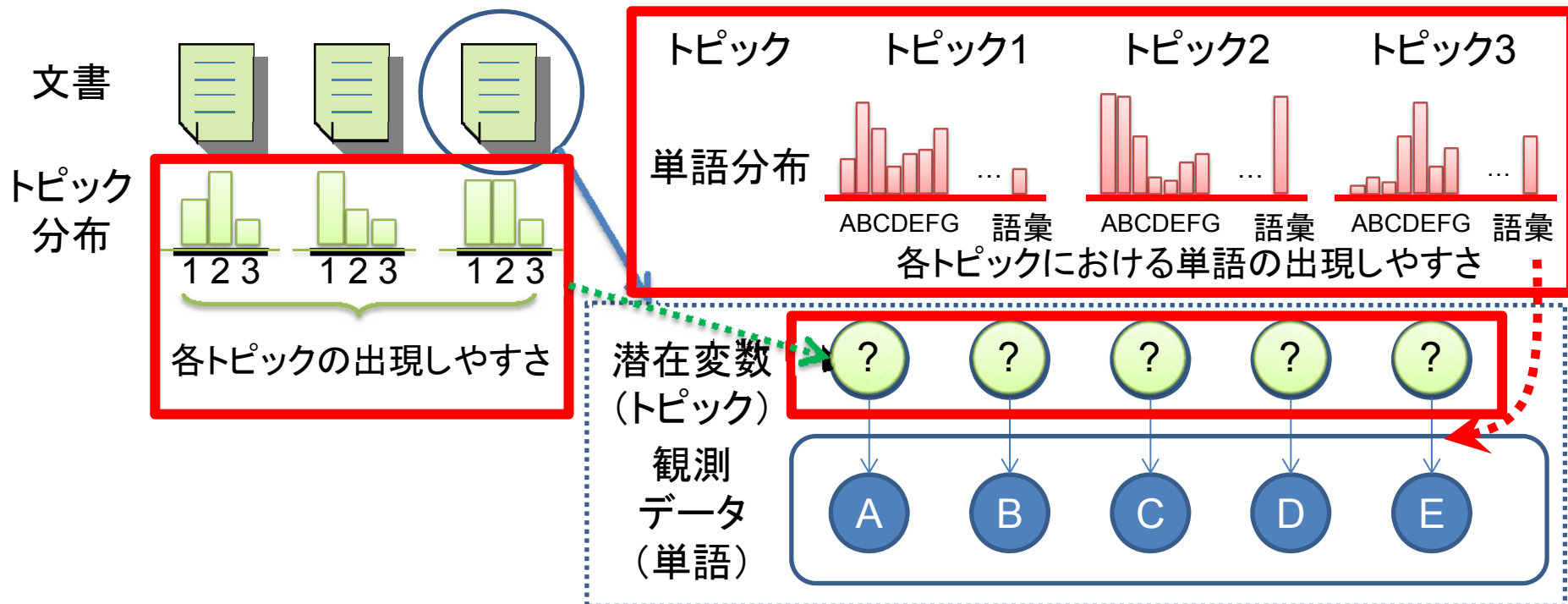
- 複数の歌声から**潜在意味(トピック)**を学習する
≡ **典型的な歌声**
 - 各歌声に**内在する隠れた構造を抽出**できる
 - トピック分布(複数トピックにおける割合)から
歌声間の類似度も算出できる
 - 複数の歌声の音響特徴量から
それらの潜在的な意味を解析する研究はなかった
- 潜在的ディリクレ配分法を利用
 - Latent Dirichlet Allocation (LDA)

LDA概要：トピックモデル=「マルチトピック」「グループデータ」に対する混合モデル

**単語系列からトピック分布、
単語分布、潜在変数系列を推定**

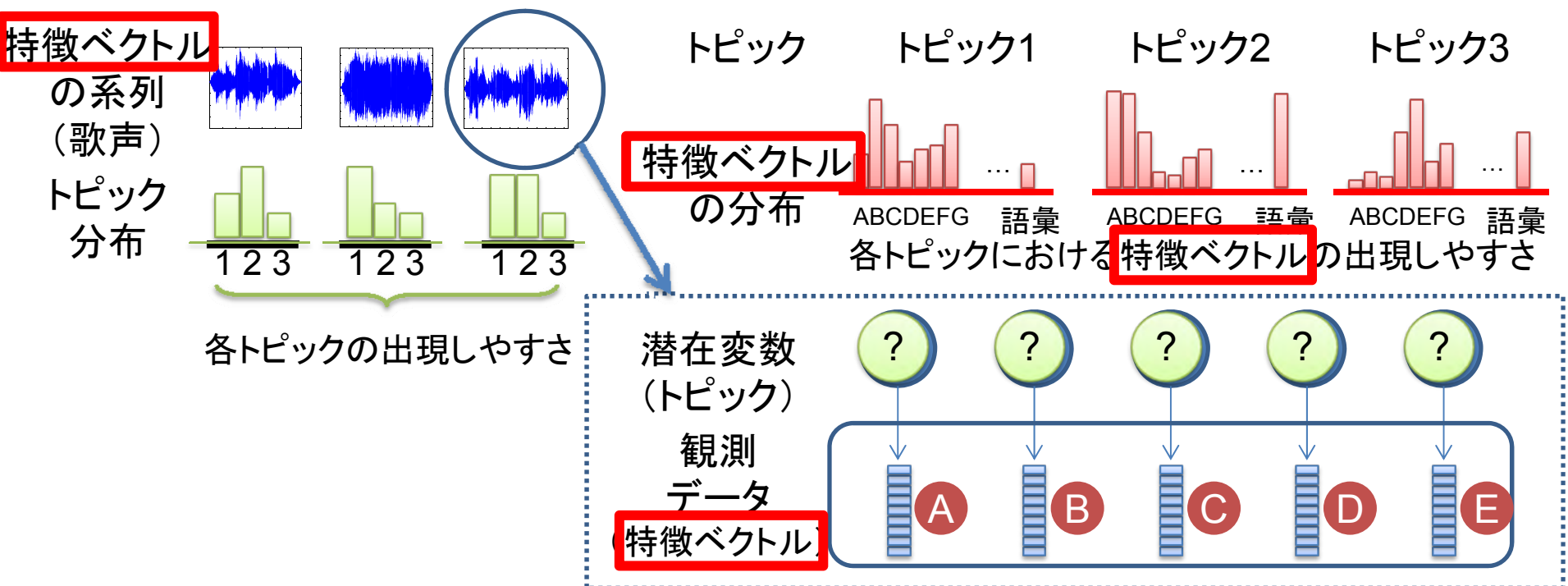
□ 文書の場合

- 各文書はトピック分布に従って生成
- 各単語はあるトピックの単語分布に従って生成



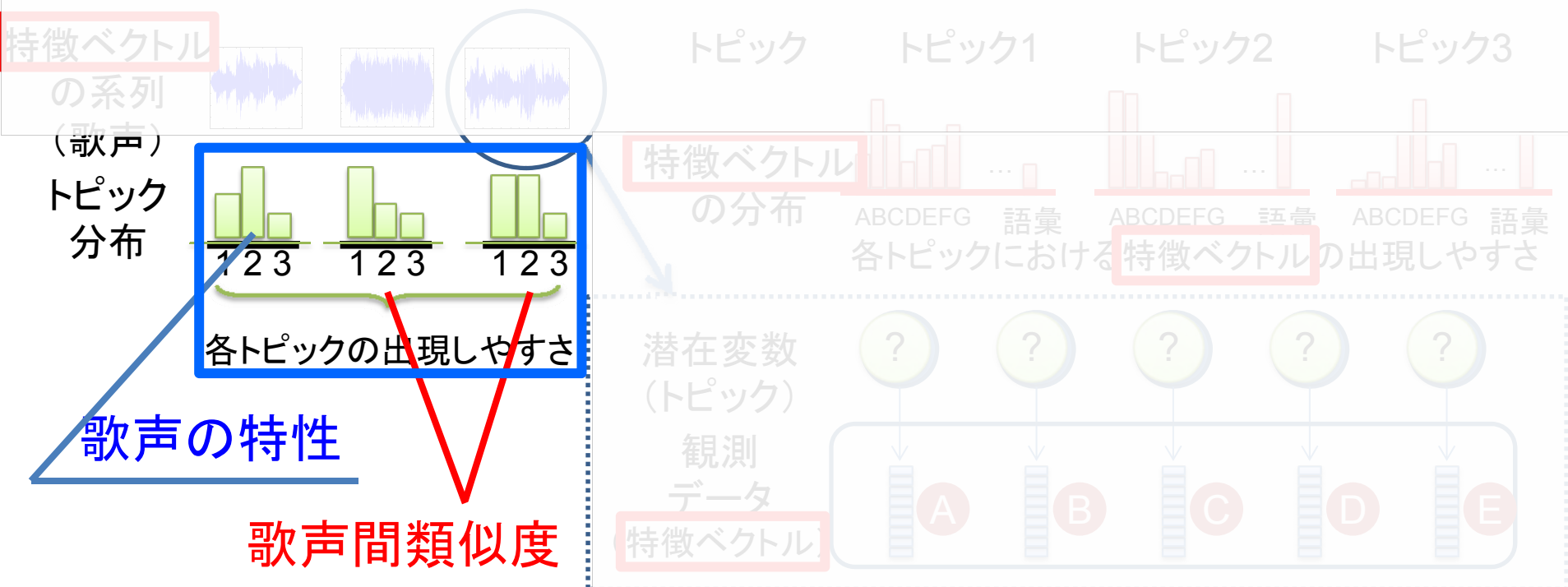
歌声トピックモデル(LDA)

- 文書 ⇒ 歌声特徴量の時系列(楽曲、フレーズ)
- 単語 ⇒ 各時刻の**特徴ベクトル**
 - **ベクトル量子化(VQ)して適切な語彙数にまとめる**



歌声トピックモデル(LDA)

- 文書 ⇒ 歌声特徴量の時系列(楽曲、フレーズ)
- 単語 ⇒ 各時刻の**特徴ベクトル**
 - ベクトル量子化(VQ)して適切な語彙数にまとめる



歌声トピックモデル(LDA)

□ 文書 → 歌声特徴量の時系列(楽曲、フレーズ)

□ 単語 → 特徴ベクトル

各トピックの意味を
もっと分かりやすくしたい

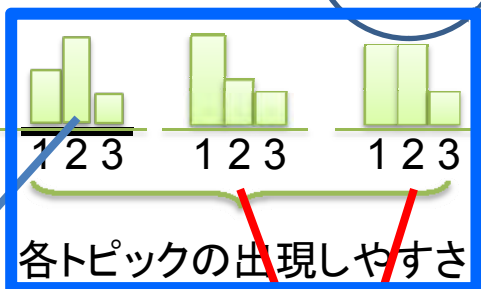
■ ベクトル → 単語彙数にまとめる

特徴ベクトル

の系列

(歌声)

トピック
分布



特徴ベクトルの分布

ABCDEF

各トピックにおける

特徴ベクトルの出現しやすさ

特徴ベクトル

ABCDEF

単語彙

潜在変数
(トピック)
観測
データ



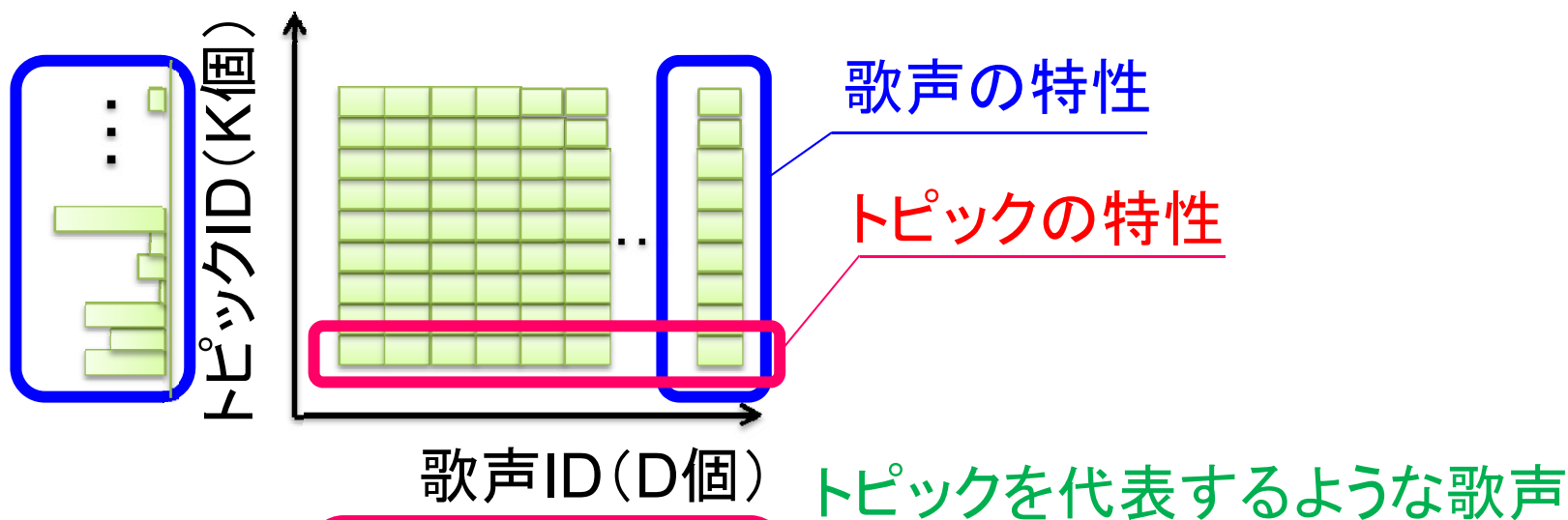
特徴ベクトル

歌声の特性

歌声間類似度

トピック分布の意味を可視化する

トピック分布はトピック数 $K \times$ 歌声数 D の行列



比率の強さに応じて
歌手名を一覧表示する

歌の特性を定量的に説明(モデル化)したい

□ トピックモデルの導入による解決

- 問題1: 時系列特徴をどうまとめるか

複数の歌声を用いて自動的に決定する

- 問題2: 特徴量トピック分布の意味が分かりにくい

歌手名を活用して可視化する

拡張 音楽情報検索の 可能性を拡げる

検索対象を拡張する音楽情報検索

□ 従来の音楽情報検索

- 「既にあるもの」から近いものを検索

□ 検索対象を変形して探索空間を拡張する

- 音高シフト: ボーカル声質の変形 (周波数軸の伸縮)

- 性別を超えた歌声情報検索

- よく知られた事例

– 「一青窈 (もらい泣き) を3半音下げると平井堅に似ている」

を自動的に推定する

※敬称略



一青窈 (もらい泣き)



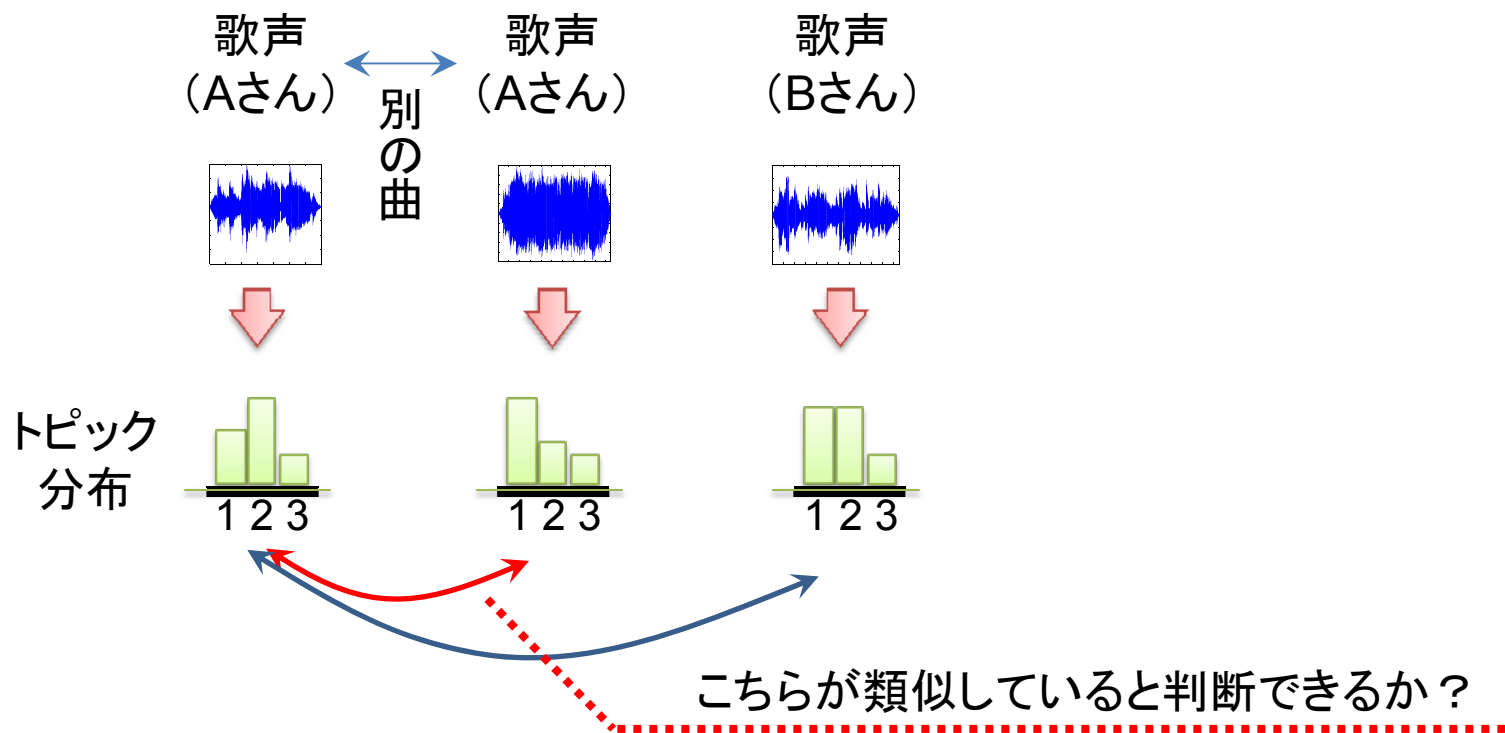
⇒ 音高シフト: -3半音

- テンポシフト: 時間変化に影響 (時間軸の伸縮)

実験

実験の目的

- トピック分布(トピックの混合比)が
歌声の類似性評価に活用できるか調査



実験条件(対象楽曲)

- 歌声を含む音楽音響信号
 - サンプリング周波数16kHz, 1ch
 - 男女6人ずつ、それぞれ3曲
 - オリコンチャート上位20曲から選曲
 - 2000年～2008年
 - ボーカルが一人である楽曲
 - 36曲: 12人 × 3曲

【男性アーティスト】

(6人、3曲ずつ)

- (M1) ASIAN KUNG-FU
GENERATION
- (M2) BUMP OF CHICKEN
- (M3) 福山 雅治
- (M4) GLAY
- (M5) 氷川きよし
- (M6) 平井堅

【女性アーティスト】

(6人、3曲ずつ)

- (F1) aiko
- (F2) JUDY AND MARY
- (F3) 一青窈
- (F4) 東京事変
- (F5) 宇多田ヒカル
- (F6) 矢井田瞳

実験条件 (LDA: モデル化)

□ 特徴ベクトル

- $\Delta F0$, LPMCC (線形予測メルケプストラム係数)
 - F0: PreFEst (混合音中で最も優勢な音高推定)
- ボーカルらしさが高いフレームのみを利用

□ 特徴ベクトルの量子化

- k-means法: クラスタ数100

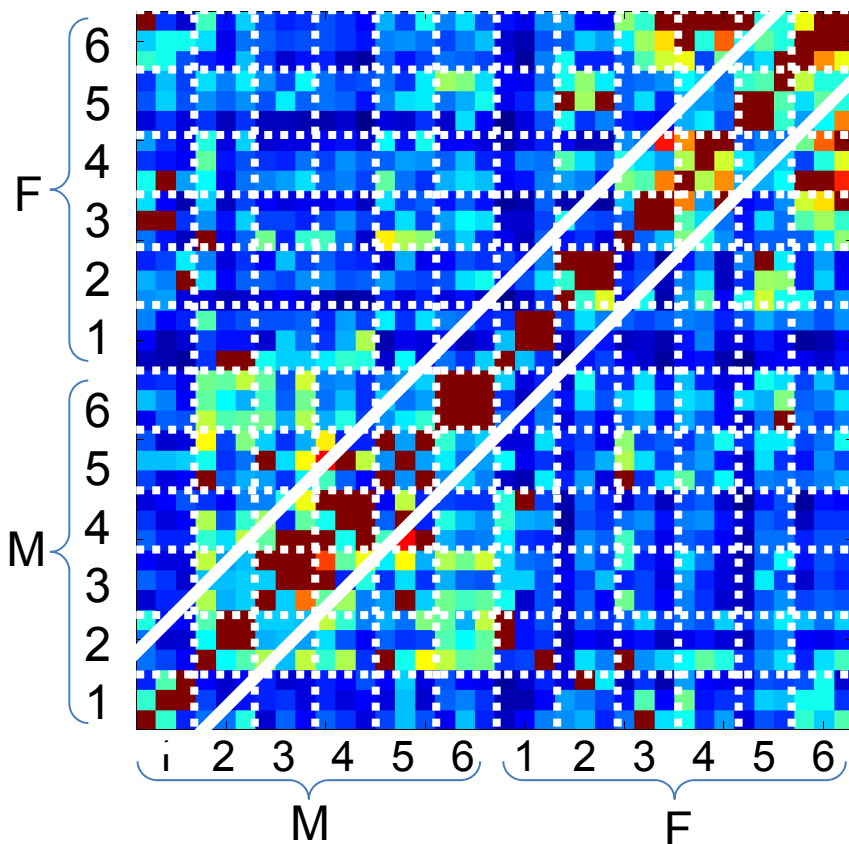
□ LDAの学習

- トピック数100
- 事前分布 (ディリクレ分布のハイパーパラメータ)
 - α (トピック混合比): 全て1
 - β (クラスタのユニグラム確率): 全て0.1とし、経験ベイズ法で最適化

実験結果：楽曲毎のトピック分布間距離

対称カルバック・ライブラ距離

音高シフト・テンポシフトなし



同一アーティストの距離が近い
= 類似性判断にトピック分布を活用できる可能性

(M5) 水戸黄門
(M6) 平井堅

(F1) aiko
(F2) JUDY AND MARY
(F3) 一青窈
(F4) 東京事変
(F5) 宇多田ヒカル
(F6) 矢井田瞳

※上位3曲を赤で塗りつぶした

実験2

実験条件：性別を超えた類似歌声検索

- 歌声を含む音楽音響信号
 - サンプリング周波数16kHz, 1ch
 - 男女6人ずつ、それぞれ3曲
 - オリコンチャート上位20曲から選曲
 - 2000年～2008年
 - ボーカルが一人である楽曲

- 1764曲：12人 × 3曲 × 7 × 7
 - 音高シフト：7バージョン
 - 3半音～3半音（1半音毎）
 - テンポシフト：7バージョン
 - 0.7～1.3倍速（0.1倍速毎）

【男性アーティスト】

(6人、3曲ずつ)

- (M1) ASIAN KUNG-FU GENERATION
- (M2) BUMP OF CHICKEN
- (M3) 福山 雅治
- (M4) GLAY
- (M5) 氷川きよし
- (M6) 平井堅

【女性アーティスト】

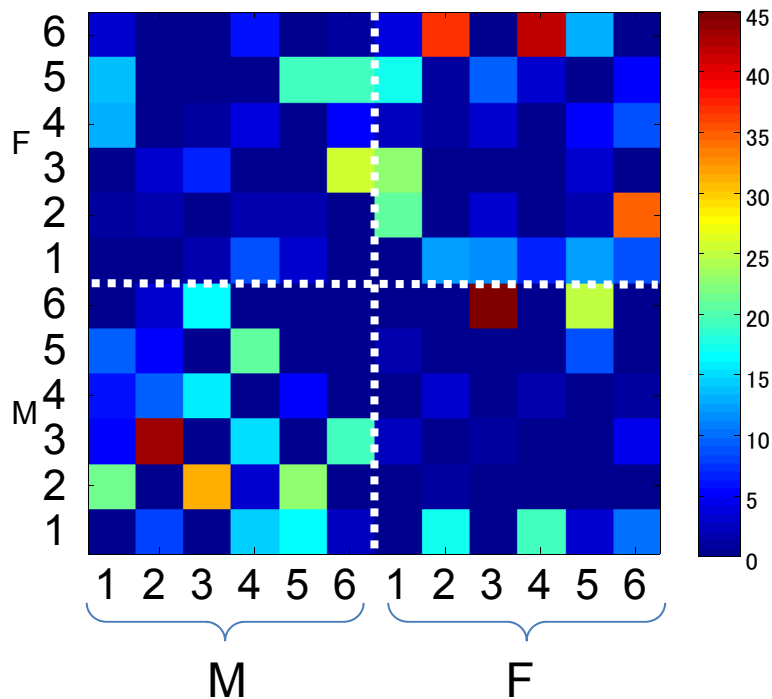
(6人、3曲ずつ)

- (F1) aiko
- (F2) JUDY AND MARY
- (F3) 一青窈
- (F4) 東京事変
- (F5) 宇多田ヒカル
- (F6) 矢井田瞳

実験結果：性別を超えた類似歌声検索

※敬称略

自分以外で
最も似ているとされた歌手の数
(赤茶色いほど多い)



【男性アーティスト】

(6人、3曲ずつ)

- (M1) ASIAN KUNG-FU GENERATION
- (M2) BUMP OF CHICKEN
- (M3) 福山 雅治
- (M4) GLAY
- (M5) 氷川きよし
- (M6) 平井堅

【女性アーティスト】

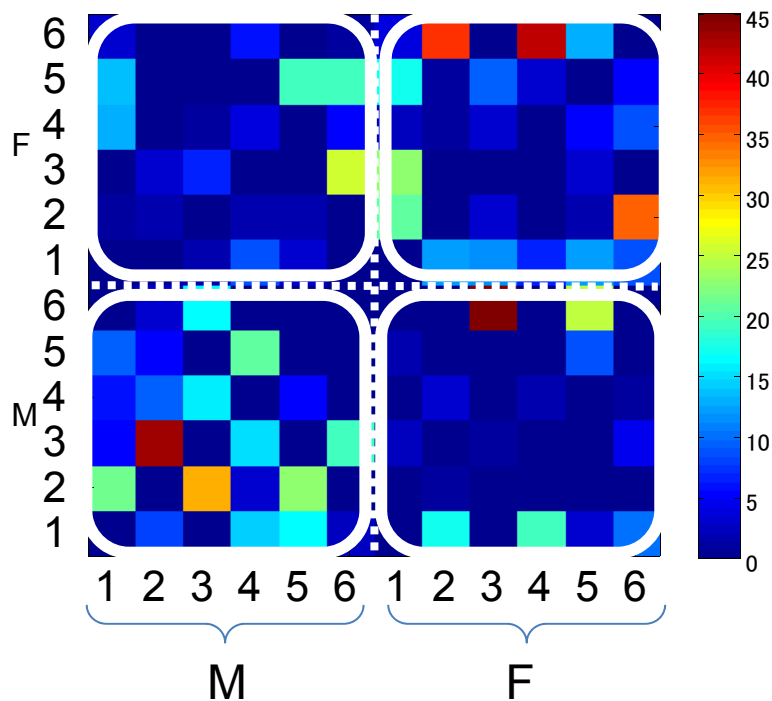
(6人、3曲ずつ)

- (F1) aiko
- (F2) JUDY AND MARY
- (F3) 一青窈
- (F4) 東京事変
- (F5) 宇多田ヒカル
- (F6) 矢井田瞳

実験結果：性別を超えた類似歌声検索

※敬称略

自分以外で
最も似ているとされた歌手の数
(赤茶色いほど多い)



【男性アーティスト】

(6人、3曲ずつ)

- (M1) ASIAN KUNG-FU GENERATION
- (M2) BUMP OF CHICKEN
- (M3) 福山 雅治
- (M4) GLAY
- (M5) 氷川きよし
- (M6) 平井堅

【女性アーティスト】

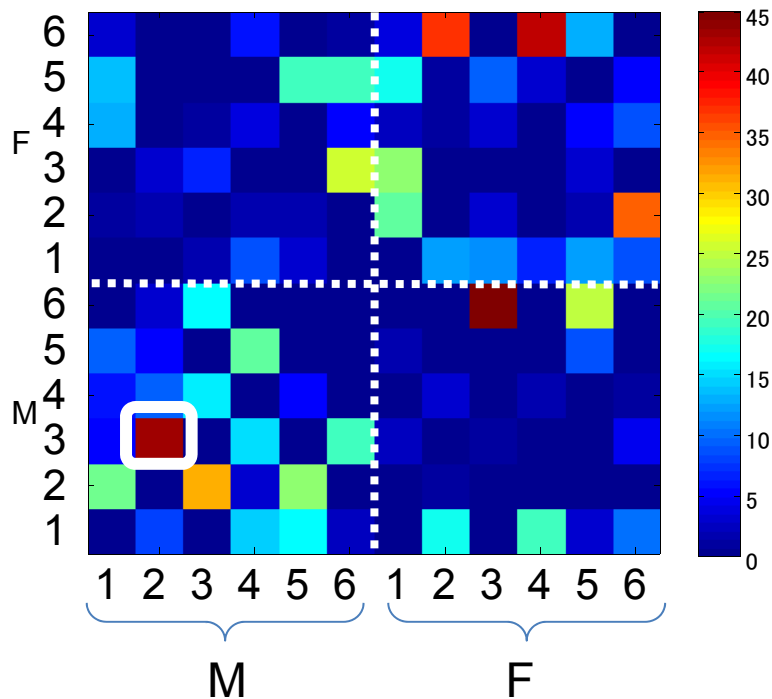
(6人、3曲ずつ)

- (F1) aiko
- (F2) JUDY AND MARY
- (F3) 一青窈
- (F4) 東京事変
- (F5) 宇多田ヒカル
- (F6) 矢井田瞳

実験結果：性別を超えた類似歌声検索

※敬称略

自分以外で
最も似ているとされた歌手の数
(赤茶色いほど多い)



【男性アーティスト】

(6人、3曲ずつ)

(M1) ASIAN KUNG-FU
GENERATION

(M2) BUMP OF CHICKEN

(M3) 福山 雅治

(M4) GLAY

(M5) 氷川きよし

(M6) 平井堅

【女性アーティスト】

(6人、3曲ずつ)

(F1) aiko

(F2) JUDY AND MARY

(F3) 一青窈

(F4) 東京事変

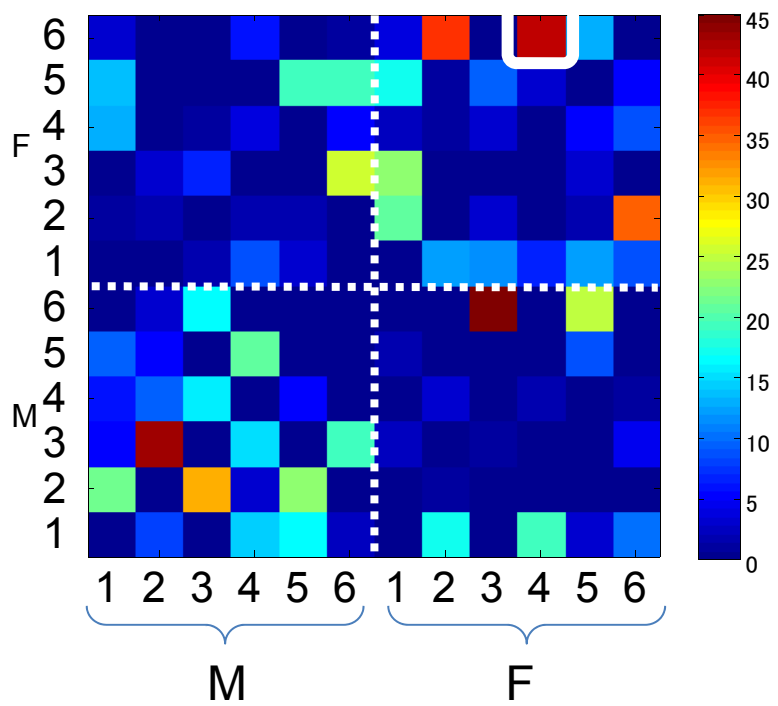
(F5) 宇多田ヒカル

(F6) 矢井田瞳

実験結果：性別を超えた類似歌声検索

※敬称略

自分以外で
最も似ているとされた歌手の数
(赤茶色いほど多い)



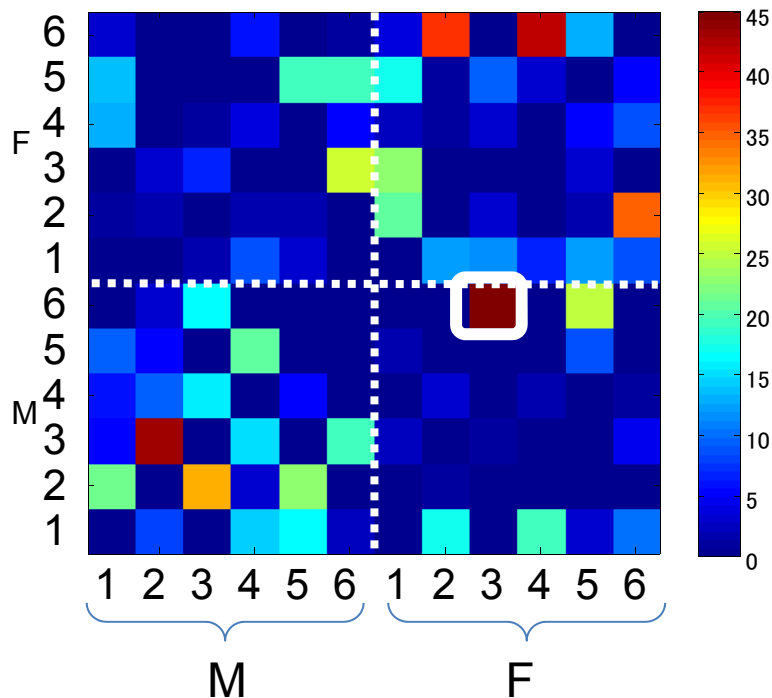
- 【男性アーティスト】**
(6人、3曲ずつ)
- (M1) ASIAN KUNG-FU GENERATION
 - (M2) BUMP OF CHICKEN
 - (M3) 福山 雅治
 - (M4) GLAY
 - (M5) 氷川きよし
 - (M6) 平井堅

- 【女性アーティスト】**
(6人、3曲ずつ)
- (F1) aiko
 - (F2) JUDY AND MARY**
 - (F3) 一青窈
 - (F4) 東京事変
 - (F5) 宇多田ヒカル
 - (F6) 矢井田瞳**

実験結果：性別を超えた類似歌声検索

※敬称略

自分以外で
最も似ているとされた歌手の数
(赤茶色いほど多い)



【男性アーティスト】

(6人、3曲ずつ)

- (M1) ASIAN KUNG-FU GENERATION
- (M2) BUMP OF CHICKEN
- (M3) 福山 雅治
- (M4) GLAY
- (M5) 氷川きよし
- (M6) 平井堅

【女性アーティスト】

(6人、3曲ずつ)

- (F1) aiko
- (F2) JUDY AND MARY
- (F3) 一青窈
- (F4) 東京事変
- (F5) 宇多田ヒカル
- (F6) 矢井田瞳

この12人の中では最も類似

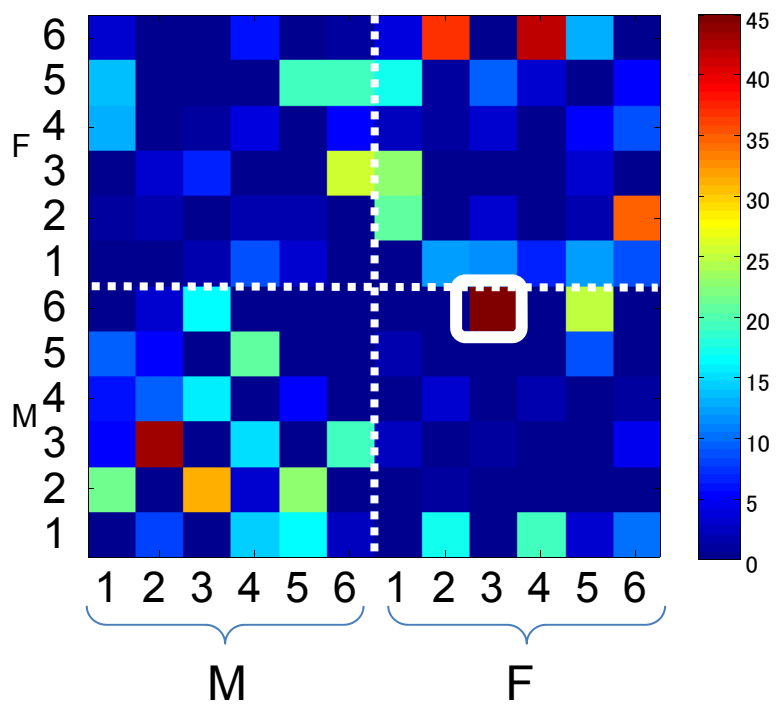
一青窈 (もらい泣き) → 平井堅 (思いがかさなるその前に・・・)
音高シフト+3, テンポシフト×1.1

平井堅 (瞳をとじて) → 一青窈 (「ただいま」)
音高シフト-3, テンポシフト×1.2

実験結果：性別を超えた類似歌声検索

※敬称略

自分以外で
最も似ているとされた歌手の数
(赤茶色いほど多い)



- 【男性アーティスト】**
(6人、3曲ずつ)
- (M1) ASIAN KUNG-FU GENERATION
 - (M2) BUMP OF CHICKEN
 - (M3) 福山 雅治
 - (M4) GLAY
 - (M5) 氷川きよし
 - (M6) 平井堅**

- 【女性アーティスト】**
(6人、3曲ずつ)
- (F1) aiko
 - (F2) JUDY AND MARY
 - (F3) 一青窈**
 - (F4) 東京事変
 - (F5) 宇多田ヒカル
 - (F6) 矢井田瞳

この12人の中では最も類似

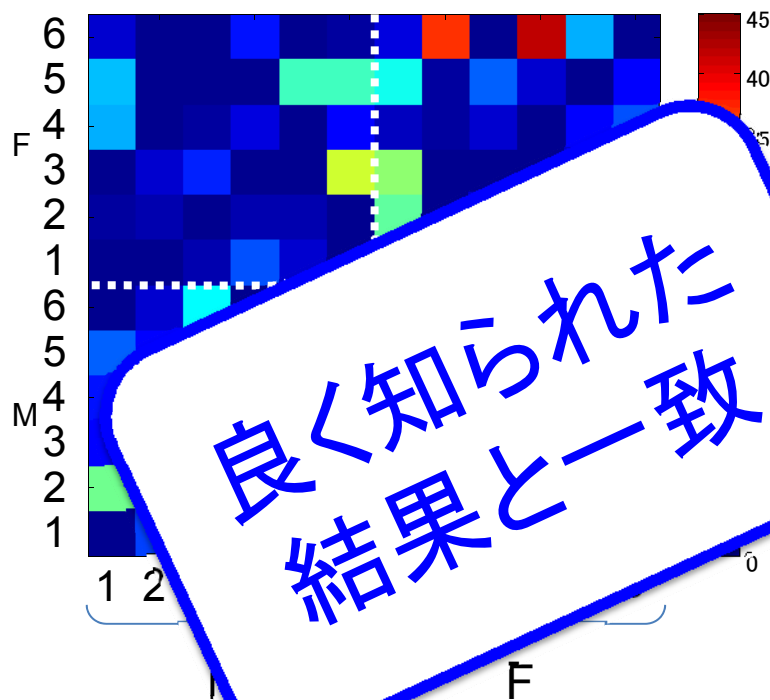
一青窈 (もらい泣き) → 平井堅 (思いがかさなるその前に...
音高シフト+3, テンポシフト×1.1

平井堅 (瞳をとじて) → 一青窈 (「ただいま」)
音高シフト-3, テンポシフト×1.2

実験結果：性別を超えた類似歌声検索

※敬称略

自分以外で
最も似ているとされた歌手の数
(赤茶色いほど多い)



【男性アーティスト】

(6人、3曲ずつ)

- (M1) ASIAN KUNG-FU GENERATION
- (M2) BUMP OF CHICKEN
- (M3) 福山 雅治
- (M4) GLAY
- (M5) 氷川きよし
- (M6) 平井堅

【女性アーティスト】

(6人、3曲ずつ)

- (F1) aiko
- (F2) JUDY AND MARY
- (F3) 一青窈
- (F4) 東京事変
- (F5) 宇多田ヒカル
- (F6) 矢井田瞳

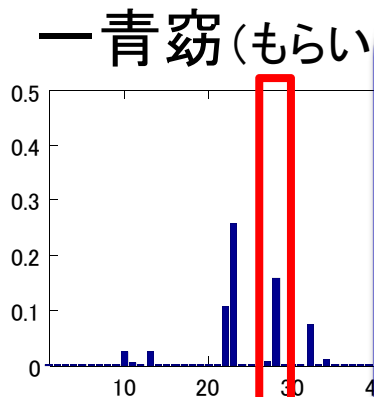
良く知られた
結果と一致

「一青窈(もらい泣き)を
3半音下げると平井堅に似ている」
を自動推定できた

一青窈(もらい泣き) ➡ 平井堅(思いがかさなるその前に・・・)
音高シフト+3, テンポシフト×1.1

平井堅(瞳をとして) ➡ 一青窈(「ただいま」)
音高シフト-3, テンポシフト×1.2

実験結果：歌手クラウドによるトピック可視化

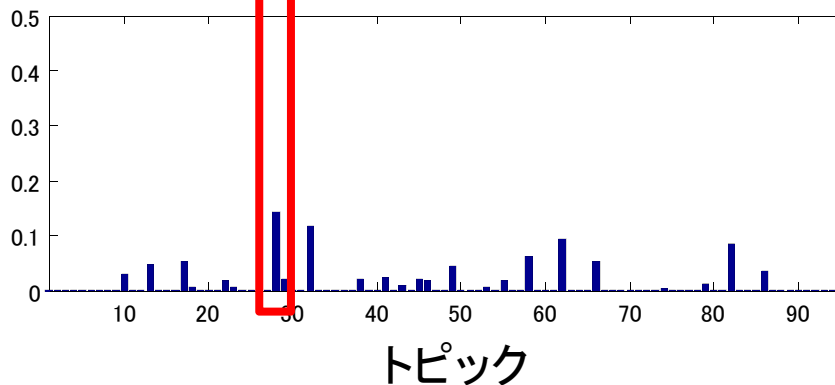


平井堅
一青窈
宇多田ヒカル
の比率が強い

トピック28の歌手クラウド
※楽曲毎にアーティスト名を表示

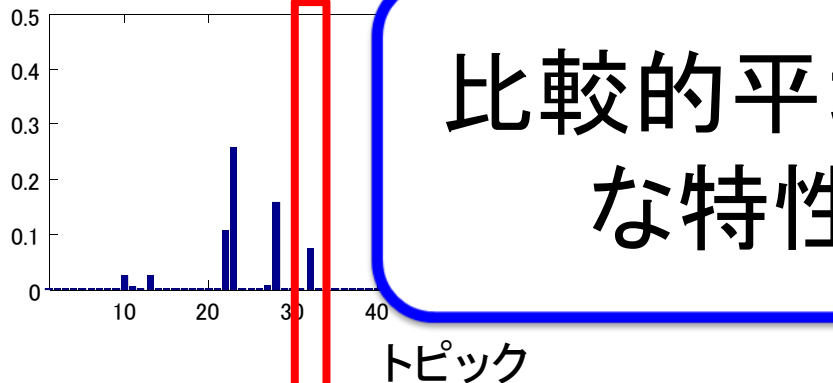


平井堅 (思いがかさなるその前に...)
音高シフト-3, テンポシフト×1.1



実験結果：歌手クラウドによるトピック可視化

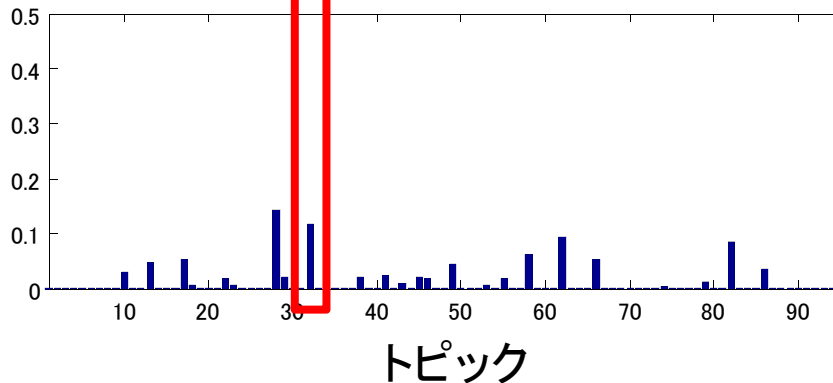
一青窈 (もらい泣き)



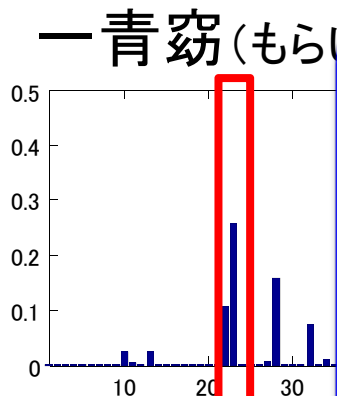
比較的平均的な特性

トピック32の歌手クラウド
※楽曲毎にアーティスト名を表示

平井堅 (思いがかさなるその前に...)
音高シフト+3, テンポシフト×1.1

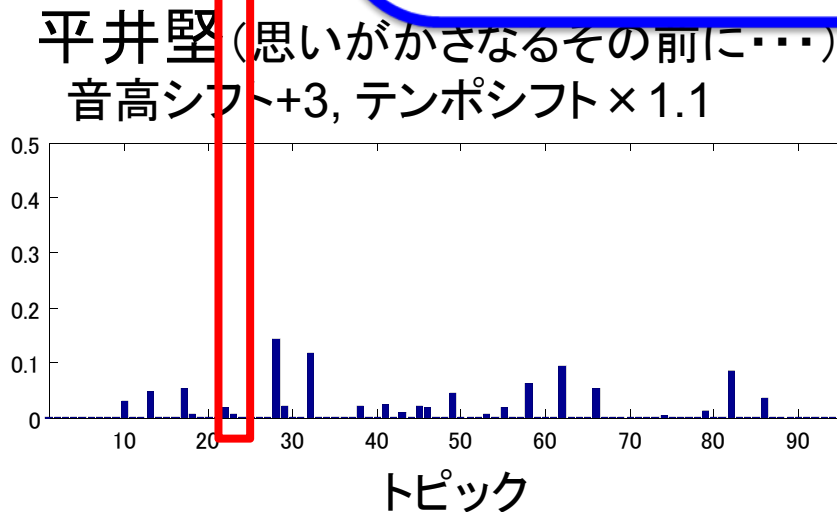


実験結果：歌手クラウドによるトピック可視化



平井堅には
ない特性
一青窈
JUDY AND MARY

トピック23の歌手クラウド
※楽曲毎にアーティスト名を表示



本研究の位置付け: 歌声の特性分析

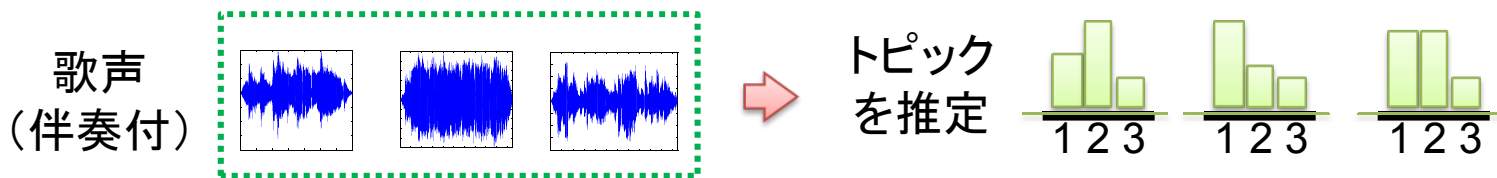
- 歌手の特性
 - 性別(男女等)
 - 声種(ソプラノやアルト等)
- 歌い方の特性
 - 声区(裏声等)
 - 音高(F0)軌跡のモデル化
- 楽曲の特性
 - 楽曲ジャンルや歌詞
- 聴取印象
 - 感情、印象評価語

本研究の範囲

複数の歌声から
個々の歌声の特性を
モデル化する

まとめ

- 歌声の音響特徴量をトピックモデルで分析
 - 複数の歌声から分かる潜在的な特性を分析



- 性別の違いを超えた類似歌手検索の提案
 - 「一青窈を3半音下げると平井堅」を実証

- 歌手クラウドによるトピックの可視化を提案



追加実験

□ アーティスト数を増やして実験

■ 120曲

- ・ オリコン2000年～2008年で上位20曲に入った楽曲
 - 男性:59アーティスト
 - 女性:61アーティスト

■ 音高シフトのみ

□ 類似事例

access(EDGE)



B'z(juice)
音高シフト+1

一青窈(もらい泣き)



森山 直太郎(さくら)
音高シフト+2

矢井田 瞳(B'coz I love you)



スピッツ(ホタル)
音高シフト+3

宇多田 ヒカル
(Wait&See～リスク～)
音高シフト+1



mihimaruGT
(ツヨクツヨク)



追加実験

□ アーティスト数を増やして実験

■ 120曲

- ・ オリコン2000年～2008年で上位20曲に入った楽曲
 - 男性:59アーティスト
 - 女性:61アーティスト

■ 音高シフトのみ

□ 類似事例

access(EDGE)



一青窈(もらい泣き)



天井田 瞳(B'coz I love you)



宇多田 ヒカル
(Wait&See～リスク～)
音高シフト+1



B'z(juice)
音高シフト+1



森山 直太郎(さくら)
音高シフト+2



スピッツ(ホタル)
音高シフト+3

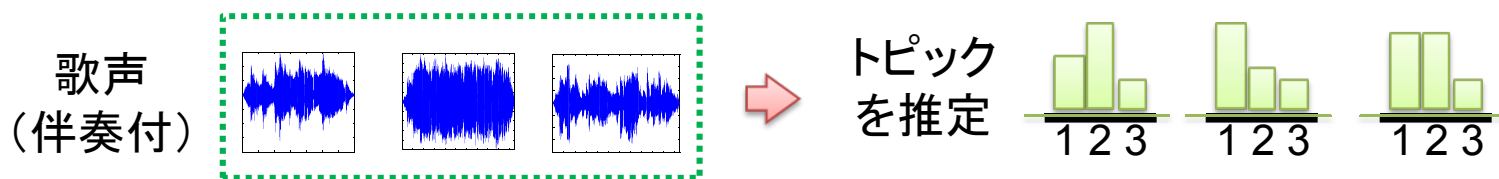


mihimaruGT
(ツヨクツヨク)



まとめ

- 歌声の音響特徴量をトピックモデルで分析
 - 複数の歌声から分かる潜在的な特性を分析



- 性別の違いを超えた類似歌手検索の提案
 - 「一青窈を3半音下げると平井堅」を実証

- 歌手クラウドによるトピックの可視化を提案

