CrestMuse Symposium project
独立行政法人科学技術振興機構 戦略的創造研究推進事業 CREST 研究領域「デジタルメディア作品の制作を支援する基盤技術」
2005年度採択テーマ「時系列メディアのデザイン転写技術の開発」研究代表者 片寄 晴弘（関西学院大学理工学部 教授）
関西学院大学理工学部／産業技術総合研究所／和歌山大学システム工学部／東京大学情報理工学系研究科／京都大学情報科学研究科　http://www.crestmuse.jp/sympo2006/

混合音中の歌声と歌詞との時間的対応付け

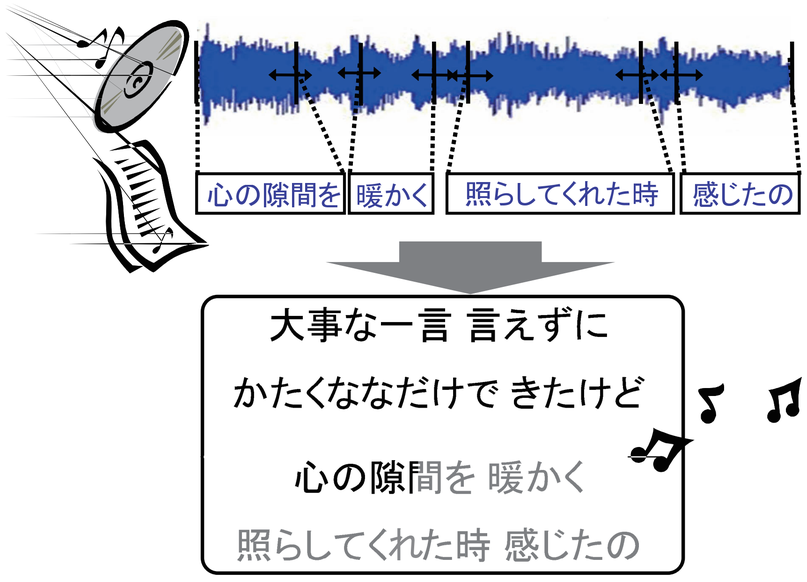
藤原 弘将　後藤 真孝†　奥乃 博　（京都大学）†産業技術総合研究所

1.背景

問題設定

混合音中の歌声と歌詞の時間的対応付け

入力：市販CD等の伴奏を含む音響信号, 対応する歌詞
出力：フレーズ毎に時間情報の付与された歌詞



従来研究

音素の時間長を用いてアラインメント[Wang2004]

➔ 各音素の発声時間は登場位置によって大きく異なる
➔ 歌声と各音素の音韻的な特徴を全く考慮していない

という問題点があった

2. 本研究の課題とアプローチ

歌声の音韻的な特徴を考慮する

➔ 音声認識で用いられるViterbiアラインメントを導入

そのためには**混在する伴奏や, 間奏部が問題となる**

➔ 以下の三つの手法を開発

伴奏音抑制	歌声区間推定	Viterbiアラインメント
混合音中から, メロディ（歌声）のみを分離する	歌声を含む領域のみを検出する	音響モデルを特定歌手に適応し, アラインメントをする

3. 歌声と歌詞の時間的対応付け手法

伴奏音抑制

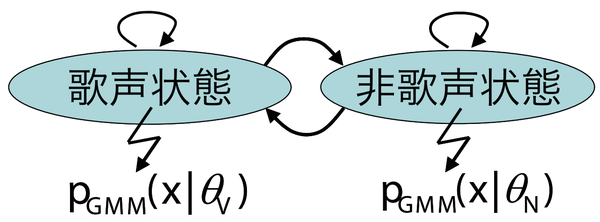
メロディのみの音響信号を得る

①メロディのF0を推定
②F0の調波構造を抽出
③正弦波重量モデルで再合成

歌声区間推定

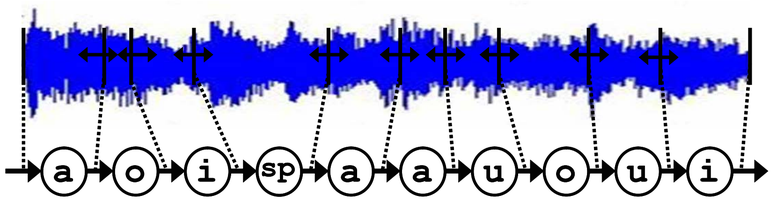
歌声を含まない領域を除去

歌声状態, 非歌声状態を行き来するエルゴディックHMM
出力確率はGMMで近似



Viterbiアラインメント

歌詞の母音のみを用いてアラインメント



MFCC、⊿MFCC、⊿パワーを使用

音響モデル以下の三段階で適応する

(i)単独歌唱の歌声に適応
(ii)分離された歌声に適応
(iii)入力楽曲の特定歌手に適応

4. 評価実験と再生インタフェース

評価実験

RWC音楽データベースから選ばれた10曲使用

10曲中8曲に対して90%以上の精度

ただし, 評価はフレーズ単位で, 正しくラベル付けされた区間の割合

再生インタフェース

再生と同期した歌詞のリアルタイム表示機能
歌詞を用いた楽曲の頭出し機能

➔ 実演デモ