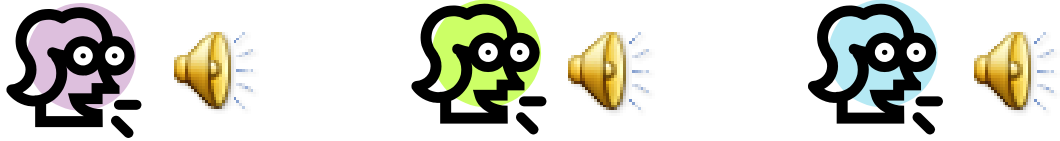
歌声の旋律と動的変動を特徴付けるための確率的な表現手法に関する検討

大石 康智[1], 後藤 真孝[2], 伊藤 克亘[3], 武田 一哉[1]

[1]名古屋大学大学院情報科学研究科
[2]産業技術総合研究所
[3]法政大学情報科学部

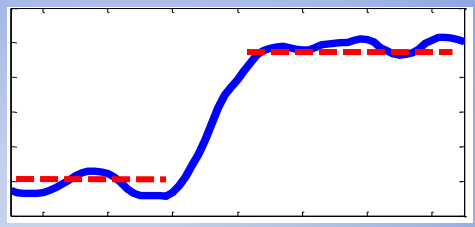
はじめに

- ある楽曲を，複数の歌唱者が歌う

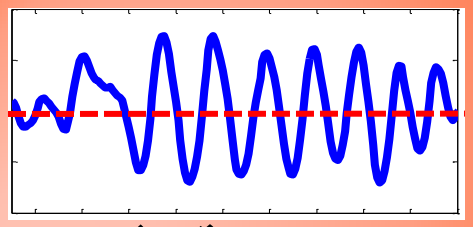


旋律情報だけでなく，歌声の動的変動
（演奏表現や癖）を特徴付けた信号モデル

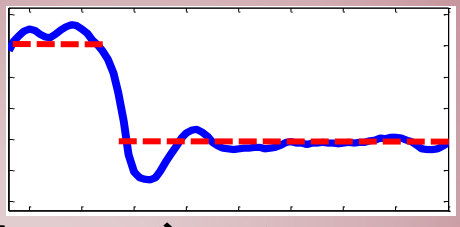
旋律を表す基本周波数（F0）の軌跡



音高の連続的な遷移



ビブラート

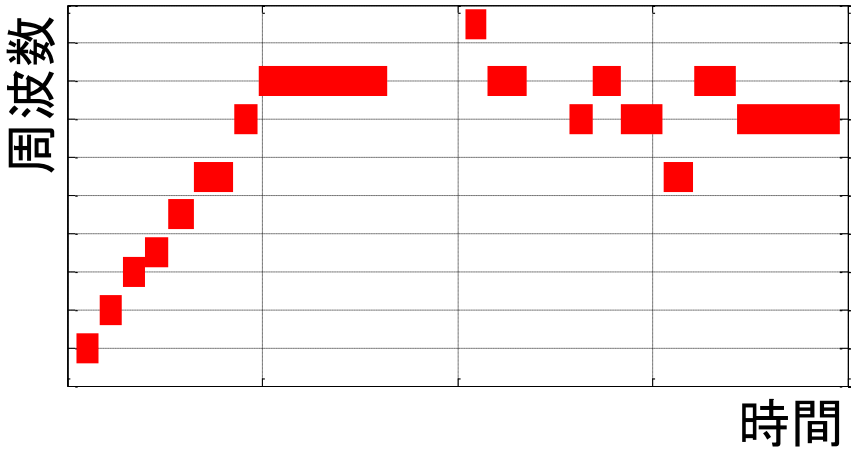


オーバーシュート

連続的な動きを表現でき，かつ，
個人の演奏表現や癖を説明することのできるモデル

歌声の動的変動を特徴付ける必要性

- 歌声合成
  - HMMや制動2次系のインパルス応答の利用
  - 合成音声の多様性についてはさらなる検討課題
- 歌唱力評価，歌唱支援
  - 歌唱者の癖の改善，歌い方の提案
- ハミング検索，歌声による採譜



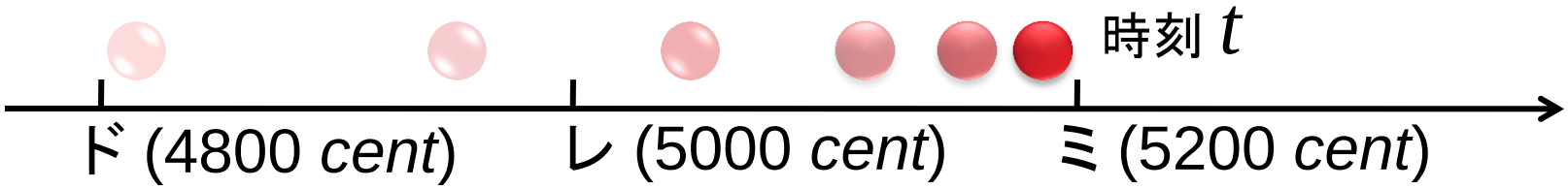
F0軌跡をシンボル列（擬似音符列）へ変換

タタタ，チャチャチャ ○

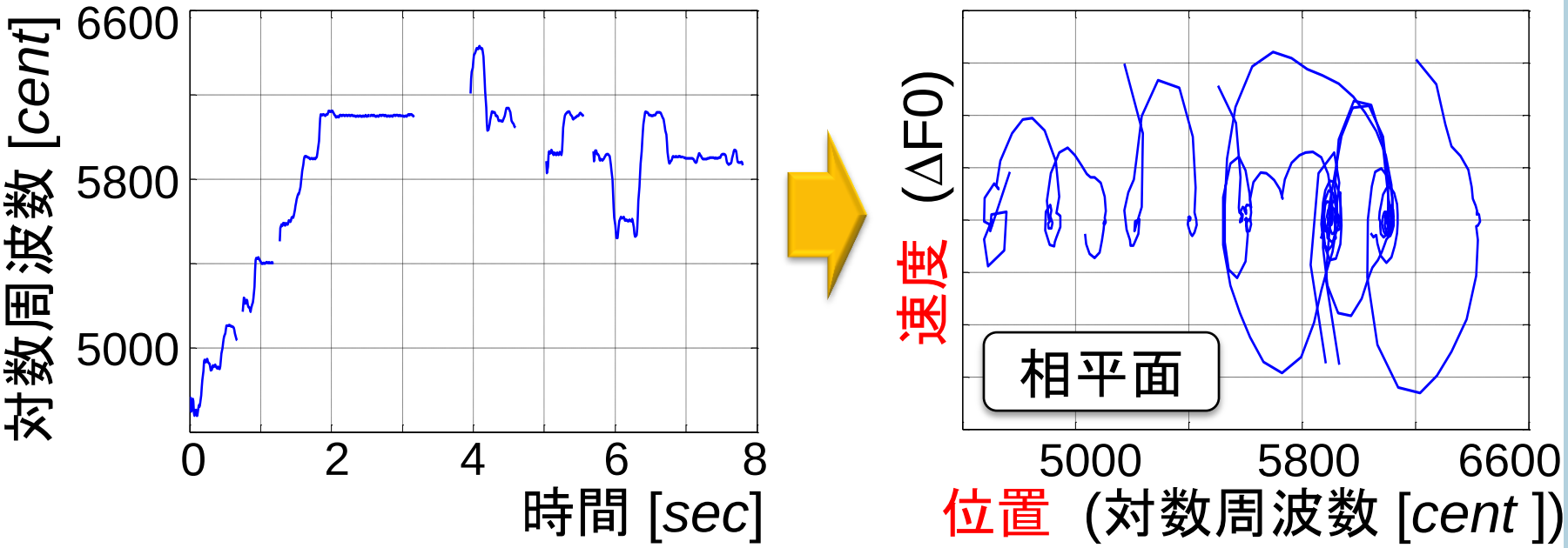
歌詞付きの歌唱 ✖

相平面における歌声のF0軌跡

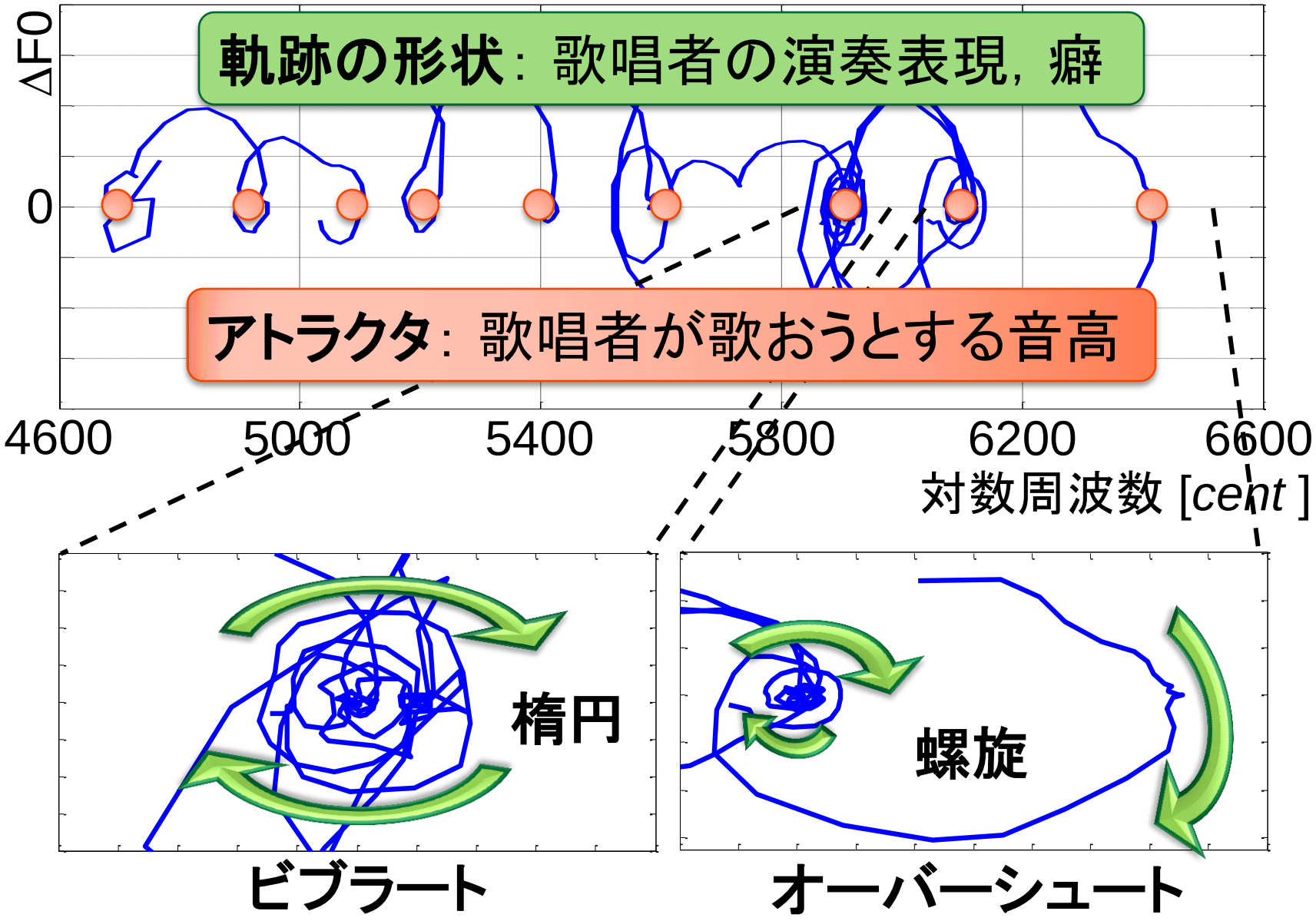
- 旋律の動き（F0軌跡）を力学的にとらえる



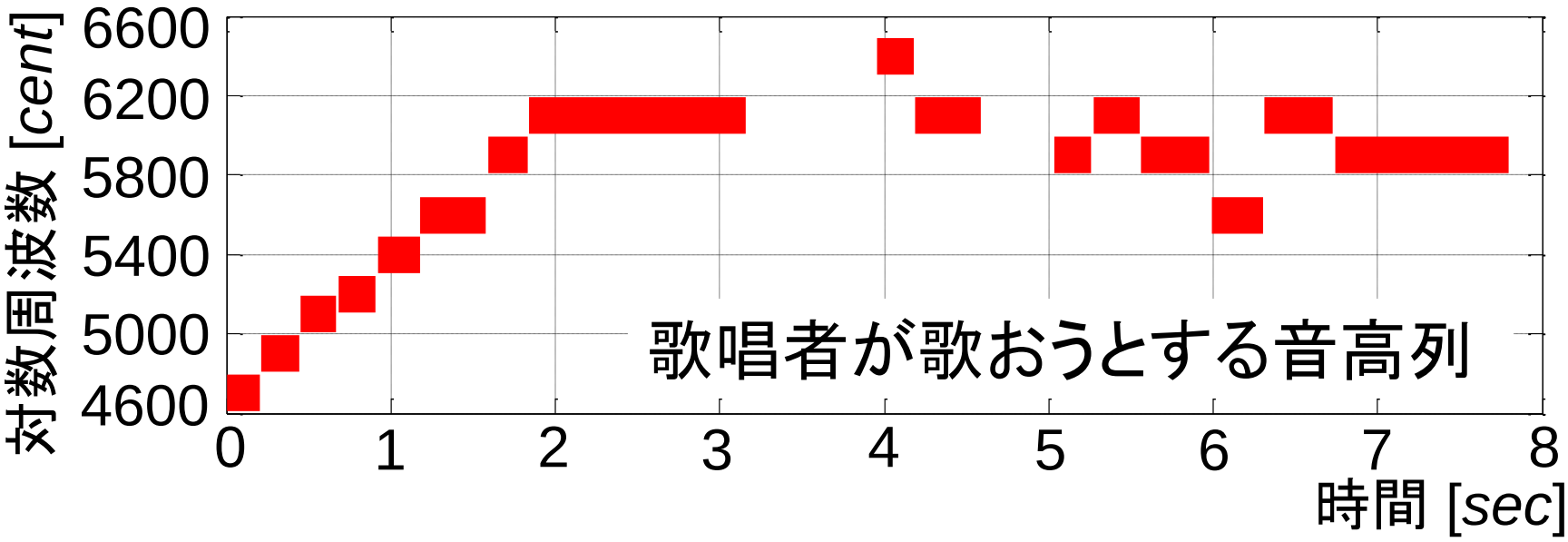
時刻 t の質点の状態は，位置と速度によって決まる

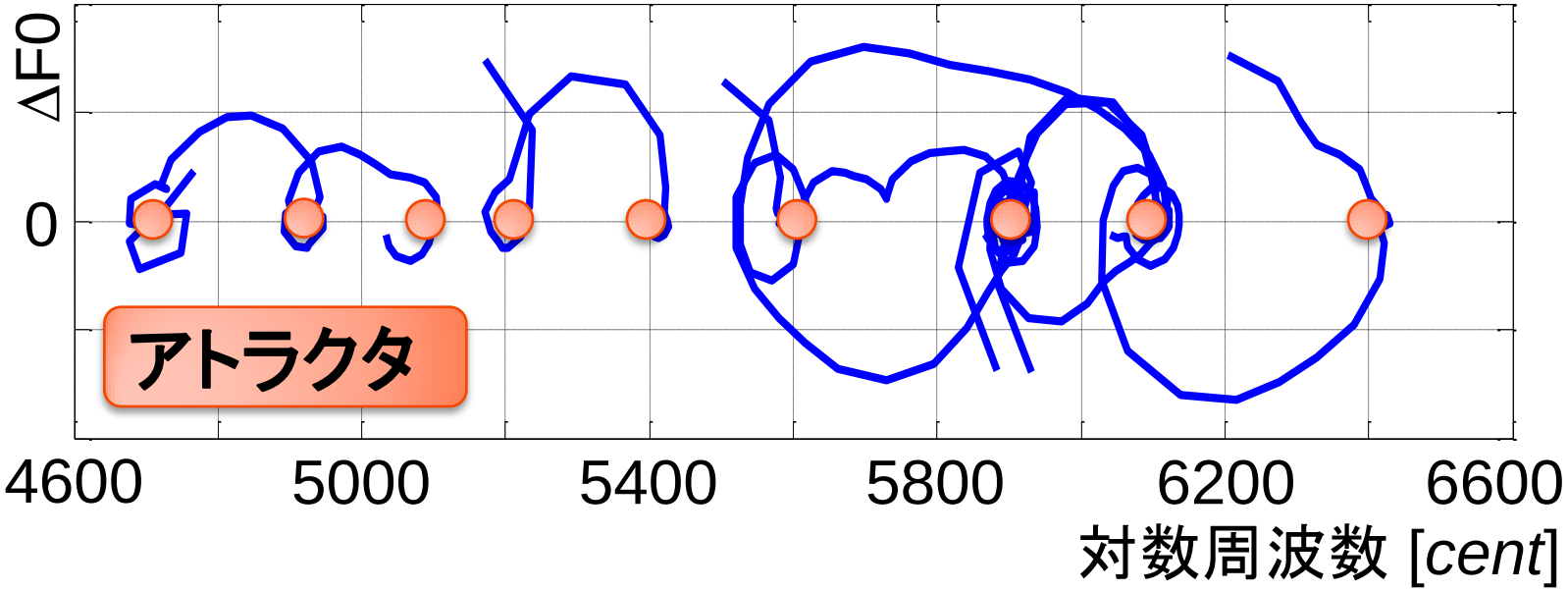


相平面における歌声のF0軌跡



本研究の目的


対数周波数 [cent]
6600
6200
5800
5400
5000
4600
0
1
2
3
4
5
6
7
8
時間 [sec]
歌唱者が歌おうとする音高列


ΔF0
0
アトラクタ
4600
5000
5400
5800
6200
6600
対数周波数 [cent]

提案手法（特徴抽出）

- F0推定
  - YIN (De Cheveigne et.al, 2002) を使用
  - 10msごとに推定
- 対数周波数への変換 （半音: 100 *cent*）

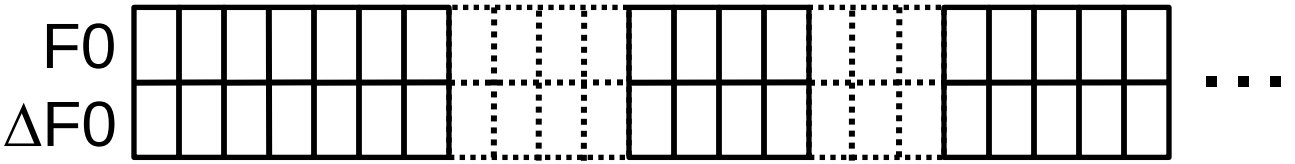
$$F0[t]_{cent} = 1200 \log_2 (F0[t]_{Hz} / 440 \times 2^{\frac{3}{12}-5})$$

- 時刻 t におけるF0の速度

$$\dot{F}0[t] \approx \Delta F0[t]_{cent} = \sum_{k=-2}^{k=2} k \cdot F0[t+k]_{cent} \Big/ \sum_{k=-2}^{k=2} k^2$$

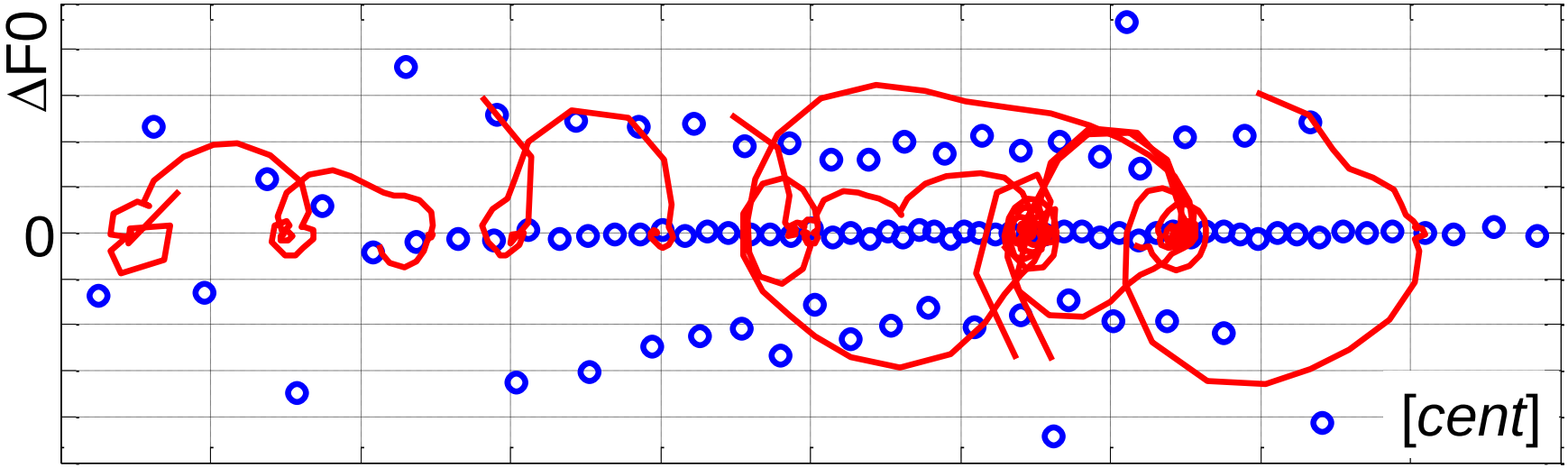
(50msにわたる回帰係数)

特徴ベクトル列



提案手法（アトラクタの位置の表現）

- 学習：相平面のクラスタリング



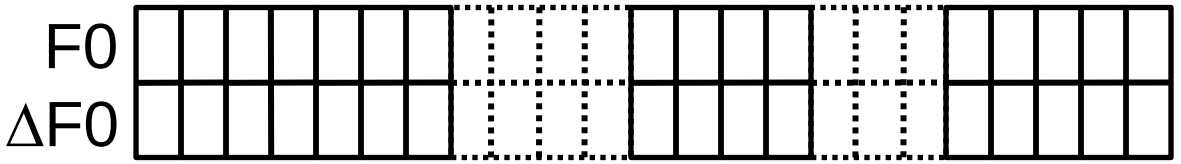
M個の領域にクラスタリング
（LBGアルゴリズム）

符号帳 $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_M\}$

符号帳の出現確率

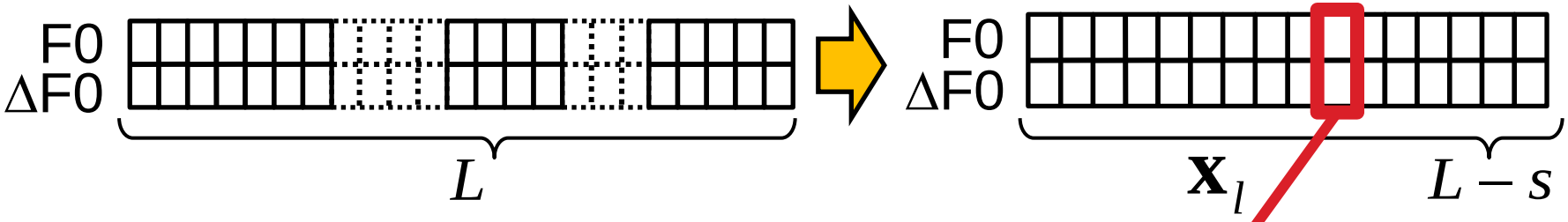
$\mathbf{P} = \{p_1, p_2, \ldots, p_M, p_{\text{無音}}\}$

特徴ベクトル列



提案手法（アトラクタの位置の表現）

- 特徴ベクトル列に対する符号帳の出現確率
  - 無声音・無音区間除去　$p_{\text{無音}} = \frac{s}{L}$

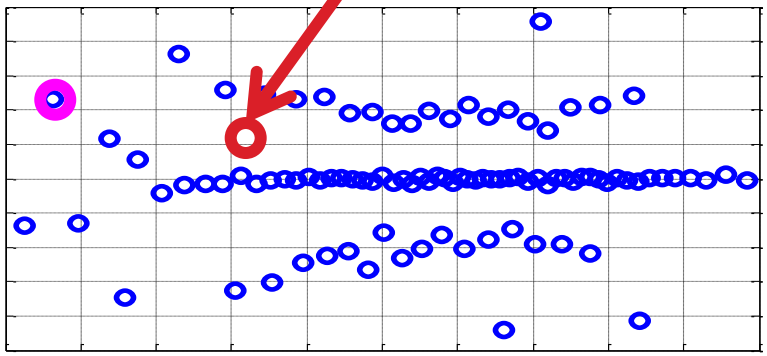


  - 特徴ベクトル $\mathbf{x}_l$ の重心ベクトル $\mathbf{v}_m$ に対する重み

$$w_{lm} = \frac{1/\| \mathbf{x}_l - \mathbf{v}_m \|_2}{\sum_{n=1}^{M} 1/\| \mathbf{x}_l - \mathbf{v}_n \|_2}$$

$$K_m = \sum_{l=1}^{L-s} w_{lm}$$

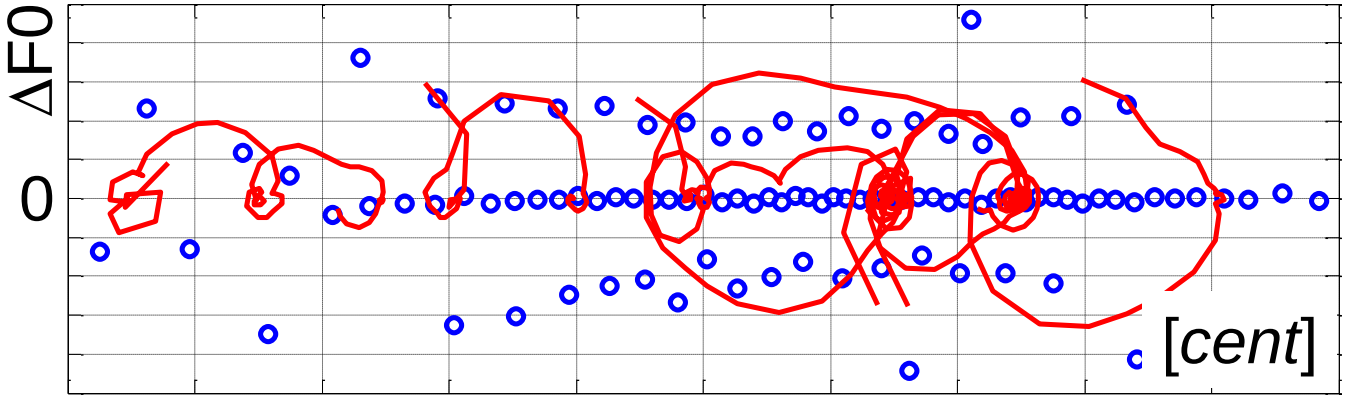
$$p_m = \frac{K_m}{L} \qquad (m = 1,2,\cdots,M)$$



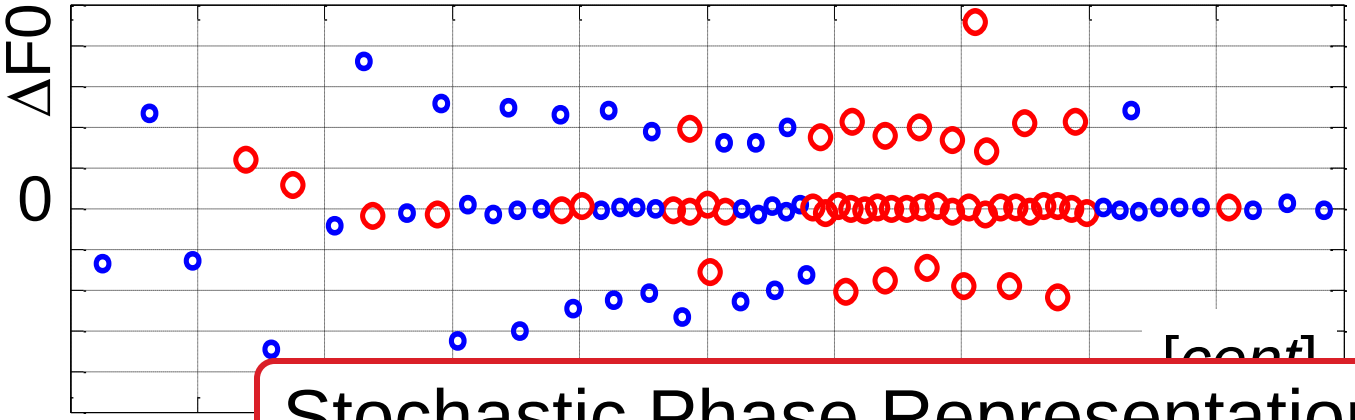
特徴ベクトルと重心ベクトルとの距離に基づいたベクトル量子化

提案手法（アトラクタの位置の表現）

- アトラクタの位置の確率的表現



特に出現確率の大きい重心ベクトルを
赤色で強調すると



Stochastic Phase Representation (SPR)

評価実験

- SPRによる旋律の表現方法の有効性
  - 楽曲データベース (参照信号)

    RWC研究用音楽データベース：ポピュラー音楽100曲

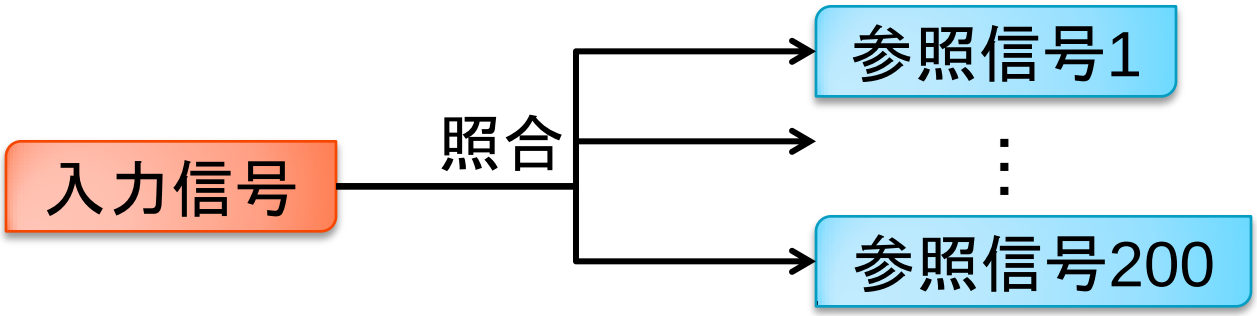
    歌唱の出だし部分
    盛り上がる主題の部分
    ｝合計200種類の旋律（平均 11.7秒）
  - 歌声データベース (入力信号)

    AISTハミングデータベース

    日本人歌唱者75名（男性37名，女性38名）

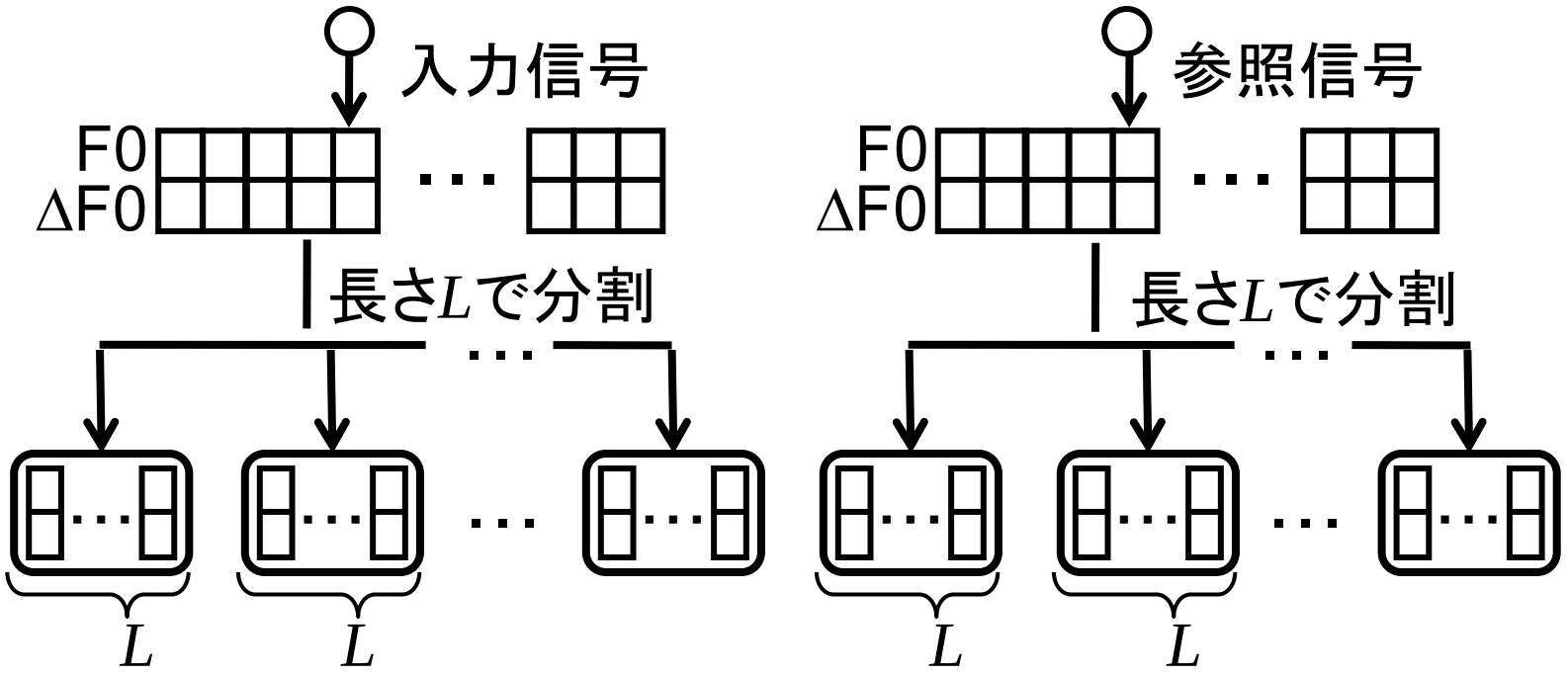
    楽曲データベースの50種類を歌詞付きで歌唱 (平均12.0秒)

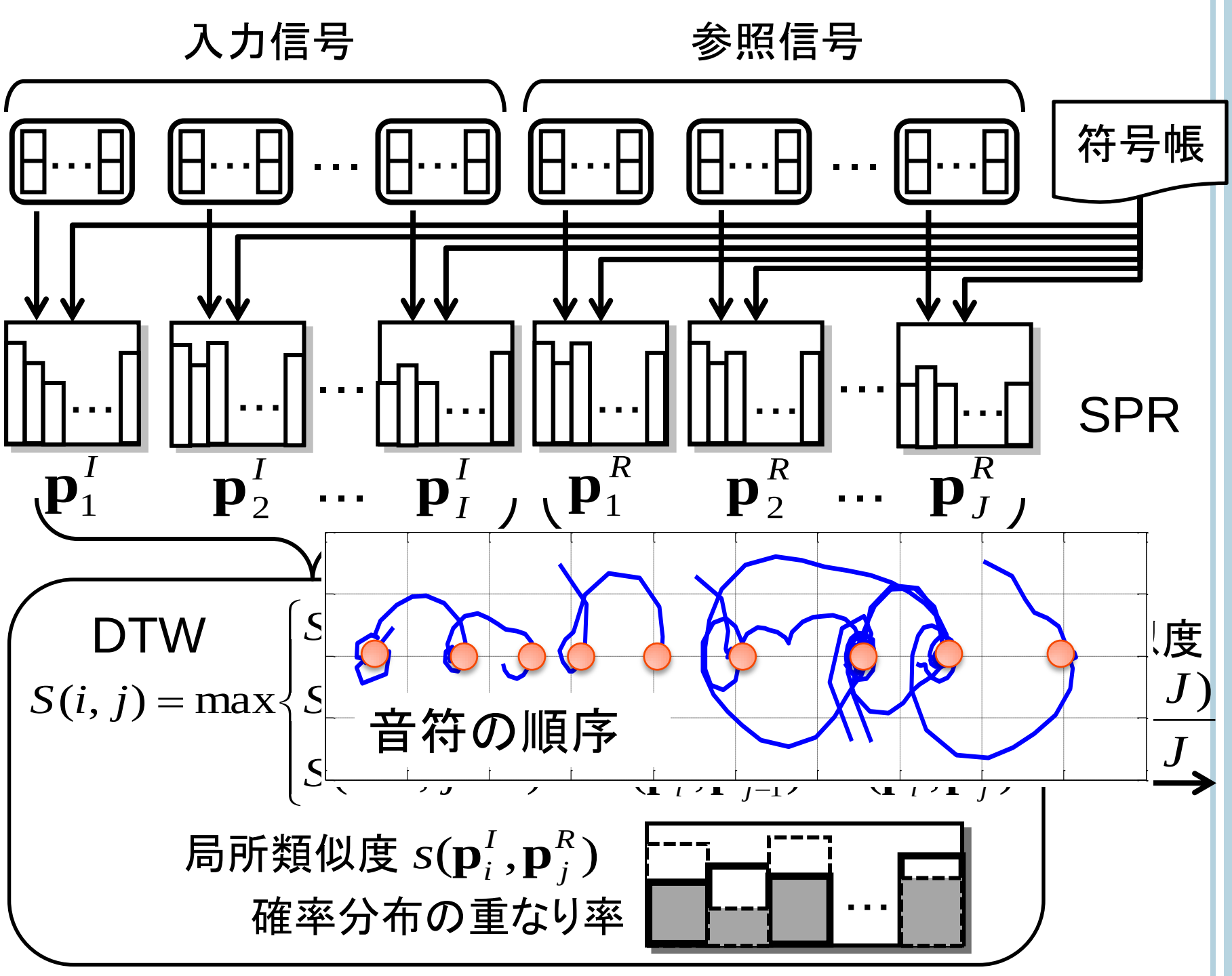
    伴奏なし，自由なテンポ，うろ覚えであるため揺れを含む



照合方法

- 移調した歌唱への対処
  - 各信号のF0の平均値を，そのF0軌跡から減算
- 符号帳の作成
  - 学習データ：200種類の旋律からなる参照信号

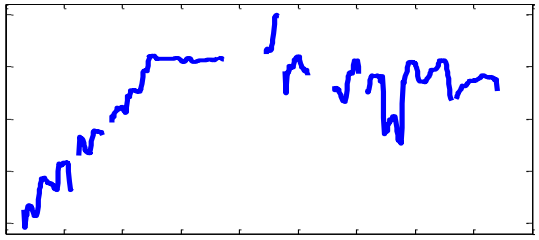



入力信号
参照信号
符号帳
SPR
$\mathbf{p}_1^I$ $\mathbf{p}_2^I$ … $\mathbf{p}_I^I$ $\mathbf{p}_1^R$ $\mathbf{p}_2^R$ … $\mathbf{p}_J^R$
DTW
$S(i,j) = \max\{$
音符の順序
$J)$
J
局所類似度 $s(\mathbf{p}_i^I, \mathbf{p}_j^R)$
確率分布の重なり率

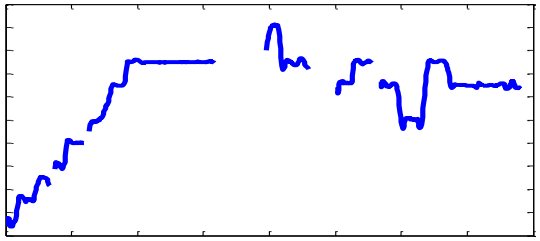
比較手法と評価方法

- DTWを利用したF0軌跡の照合方法（従来法）

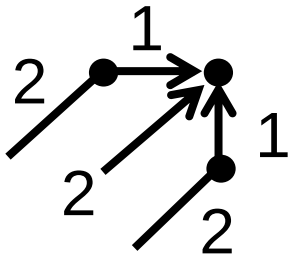
入力信号



参照信号



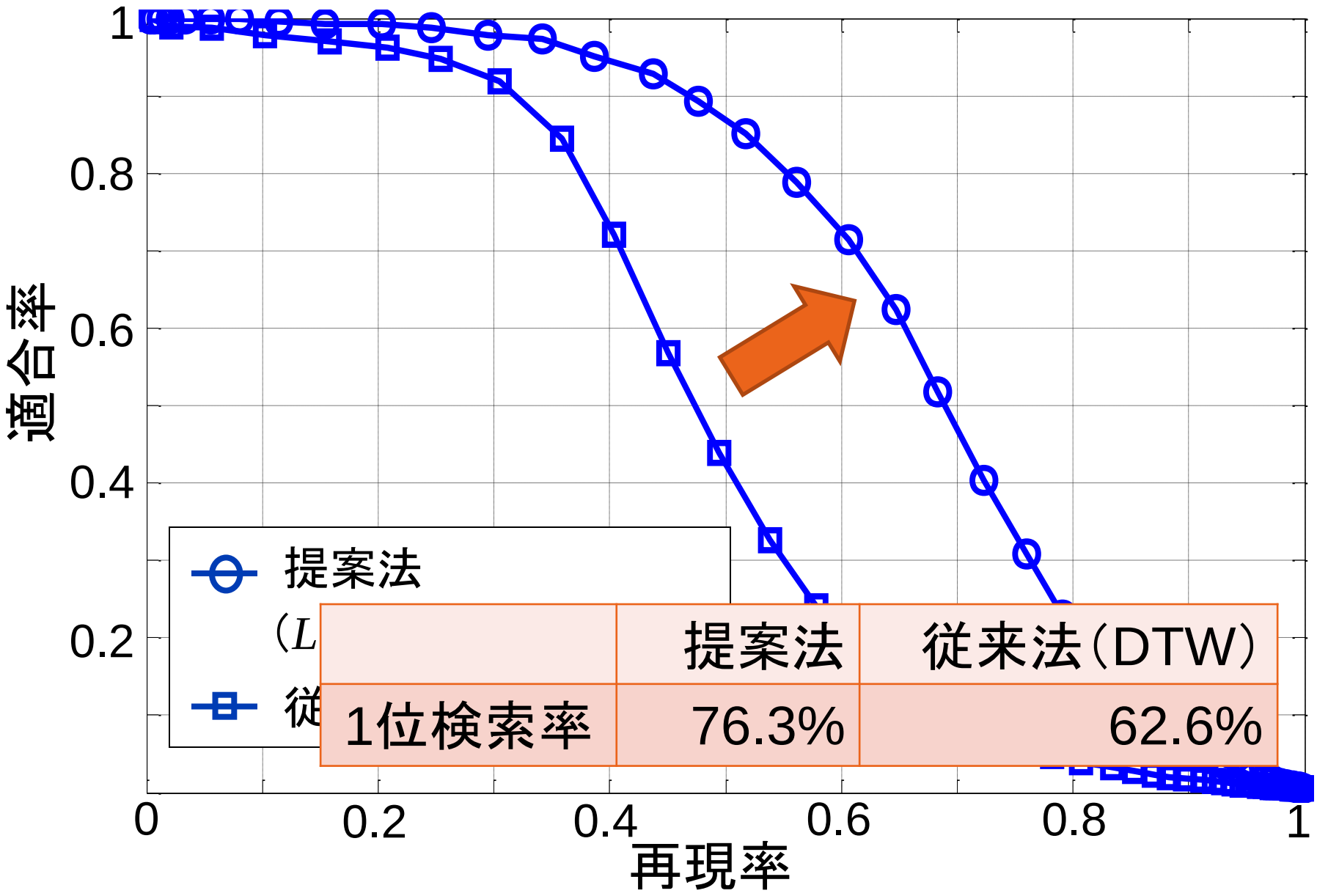
DTWの傾斜制限



F0が推定されない区間は取り除く

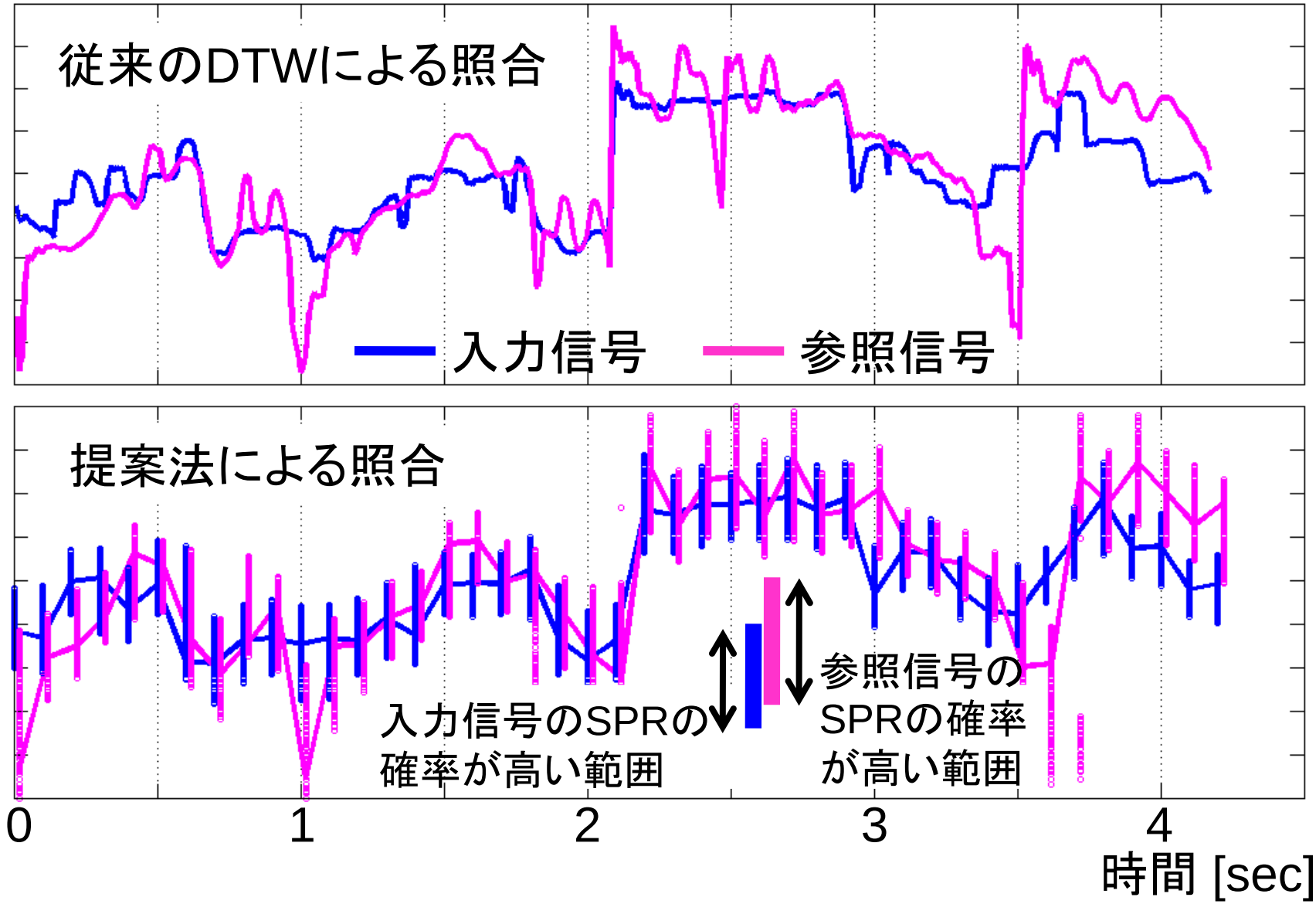
- 評価方法
  - 適合率と再現率の関係
  - 1位検索率
    - 入力信号に対して正解の参照信号が1位に検索される

実験結果



	提案法	従来法（DTW）
1位検索率	76.3%	62.6%

考察（提案法の利点）


従来のDTWによる照合
入力信号
参照信号
提案法による照合
入力信号のSPRの
確率が高い範囲
参照信号の
SPRの確率
が高い範囲
0
1
2
3
4
時間 [sec]

考察（旋律のテンポと検索性能）

- 歌声データベース　50種類の旋律
  - テンポが遅い10種類の旋律（平均72.7bpm）
    - 75名の被験者が歌唱　750サンプル
  - テンポが速い10種類の旋律（平均169.3bpm）
    - 75名の被験者が歌唱　750サンプル

	テンポの速い旋律	テンポの遅い旋律
従来法(DTW)	64.3%	70.6%
提案法	80.0%	80.8%
誤り改善率	44.0%	34.7%

旋律のテンポが速い ⇒ 歌唱者はうろ覚えのため，正しく旋律を歌唱できない，揺れを含んでしまう

まとめと今後の展開

- 相平面を利用したF0 軌跡の新しい表現手法
  - アトラクタの位置：音高が安定する箇所
  - 軌跡の形状： 歌唱者の演奏表現や癖などの動的変動
- アトラクタの位置を確率的に表現する手法
  - ハミング検索のための旋律の類似尺度に適用
  - 動的変動や部分的な音符の挿入や削除による揺れを含む歌声に対しても，効果的に検索結果を絞り込むことが可能であった

- 歌唱者の演奏表現や癖の特徴付け，歌声生成
- 相平面におけるF0軌跡の”動き”をどのようにモデル化したらよいか？