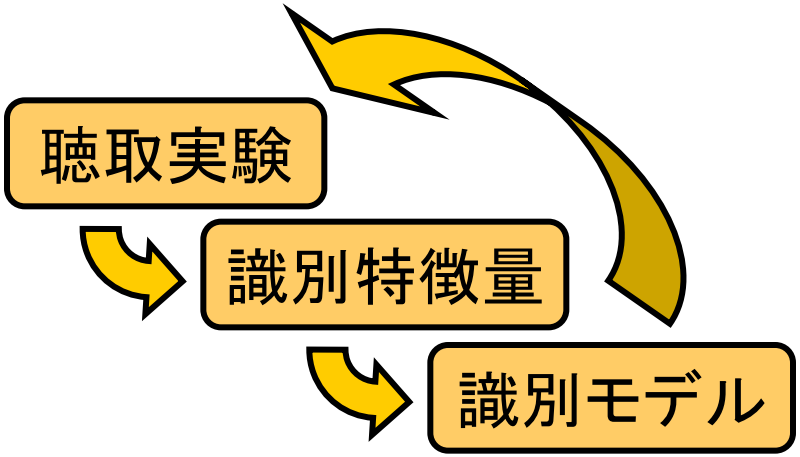
# 音声だけでシームレスにハミング検索と曲名検索が可能な楽曲検索システム

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

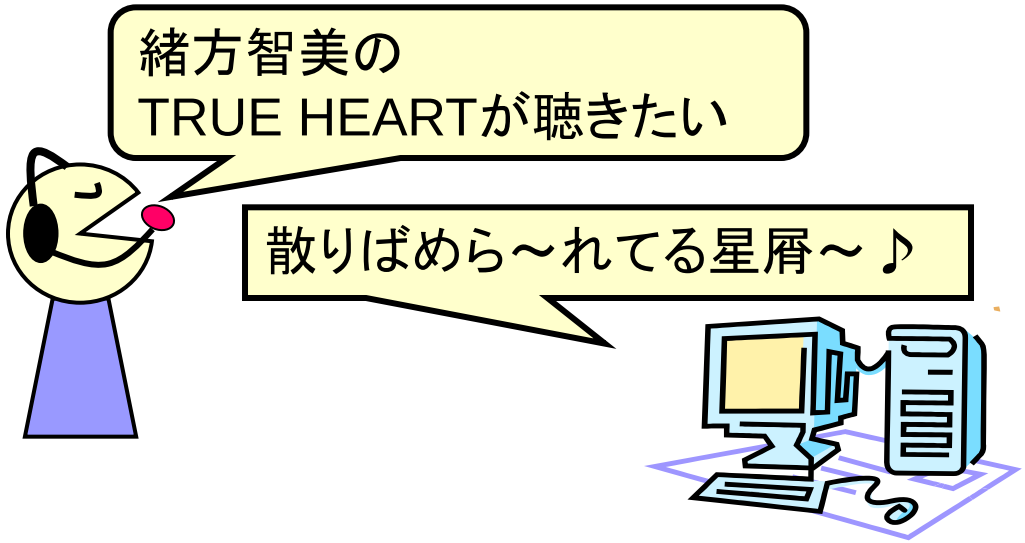
# はじめに

- 歌声と朗読音声の識別
  - 短時間スペクトル特徴
  - 基本周波数の時間変化

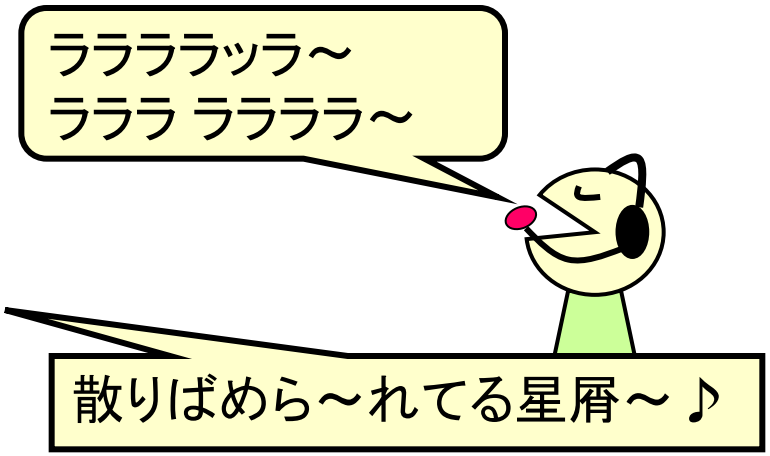



- 歌っても，曲名を読み上げても検索可能な楽曲検索システム

＜曲名検索＞




＜ハミング検索＞

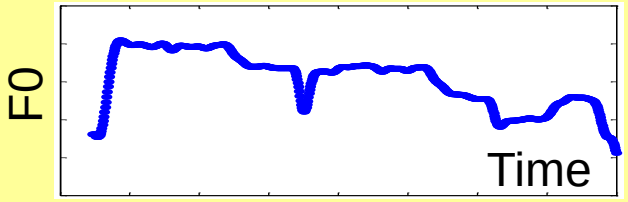


# デモをご覧ください

- RWC音楽データベース：ポピュラー音楽の No. 46 森元康介作曲の “線路はつづくよ” を歌声と曲名の朗読音声で検索します。 

# システムの概要

検索キーのメロディ(F0)
をもつ曲を検索する




ハミング検索

歌声

音声
入力

音声識別器

朗読音声

楽曲
データベース

曲名検索 (音声認識)

♪
♪♪

楽曲
出力

楽曲再生

発話を音声認識し，
曲名から曲を検索する

検索キーに対して
最も類似度の高い曲を再生する

<s> se N ro wa tu du ku yo </s>
線路は続くよ

# 音声識別器

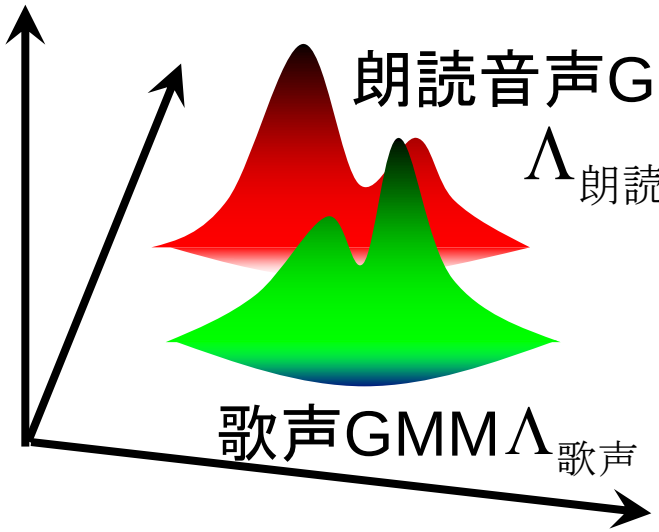
- 識別特徴量

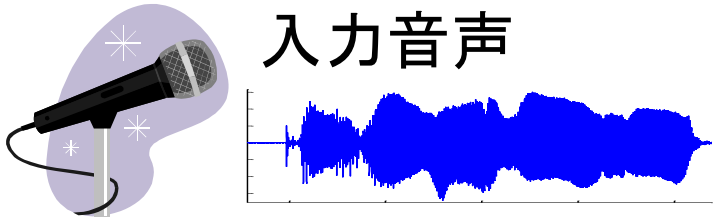
短時間スペクトルの特徴抽出
・MFCC, ΔMFCC

韻律の特徴抽出
・F0の時間変化 ΔF0

- 識別モデルの学習

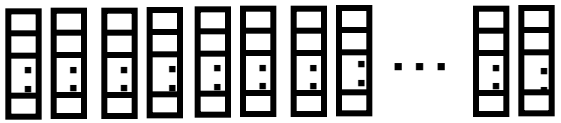
混合ガウス分布(GMM)の利用




特徴抽出

特徴ベクトル系列 $\mathbf{x}_t (t = 1, \ldots, T)$



識別関数

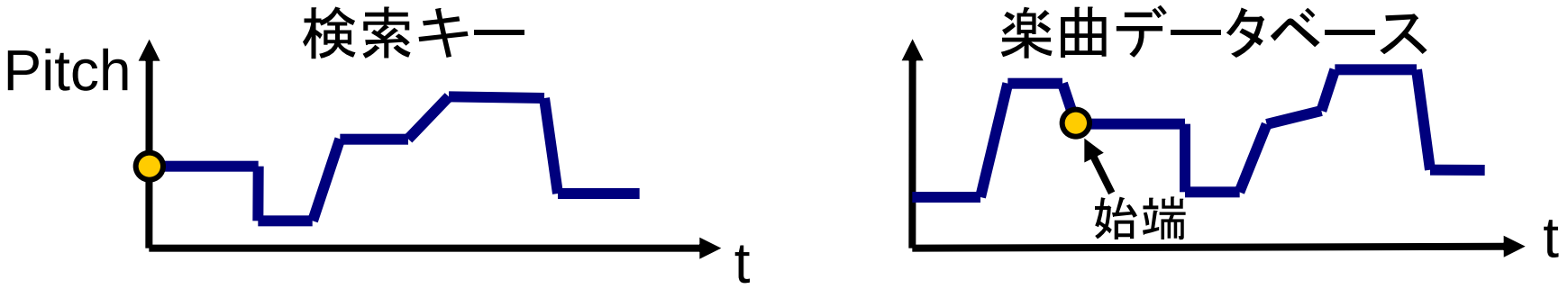
$$\hat{d} = \underset{d=\text{歌声,朗読音声}}{\arg\max} \frac{1}{T} \sum_{t=1}^{T} \log p(\mathbf{x}_t; \Lambda_d)$$

# 従来のハミング検索器

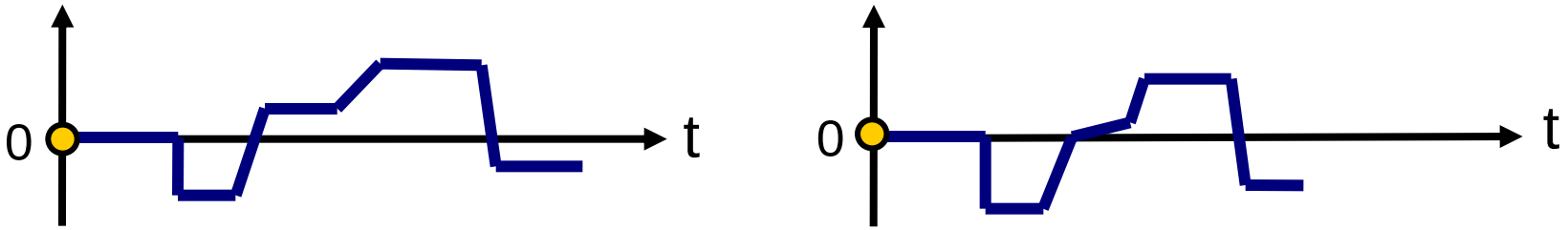
- 記号・パターンベースの検出手法
  - 始端特徴依存連続DPを用いた鼻歌検索手法

(西村ら,2001)

①メロディ(音高時系列)を求める




②始端の音高を基準とした音高を求める

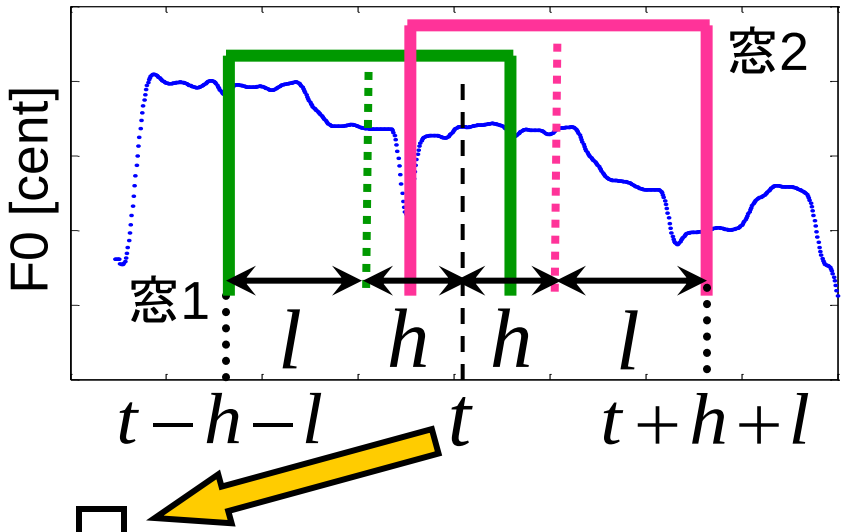


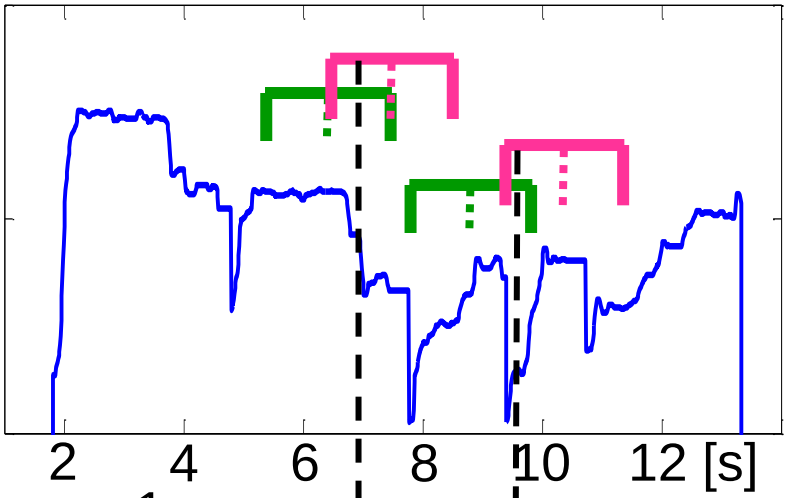

③DPマッチングにより類似度の算出

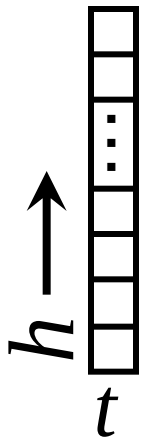
移調に対応したメロディのマッチング

# 提案するハミング検索器(特徴抽出)

- メロディの時系列を多次元に眺めることはできないか?
- フレーズ構造，繰り返しの特徴抽出



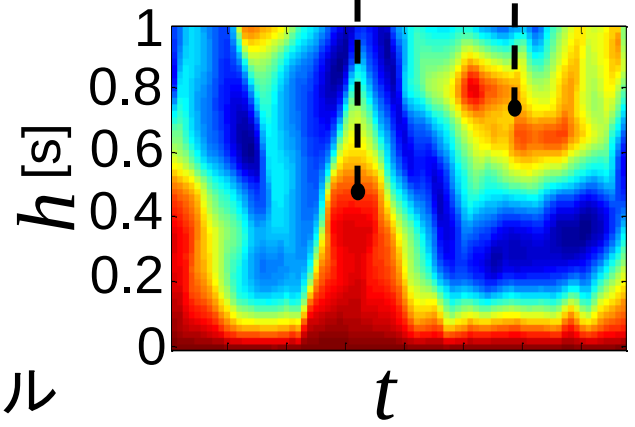




窓1と窓2に含まれる
メロディ間の相関係数を求める

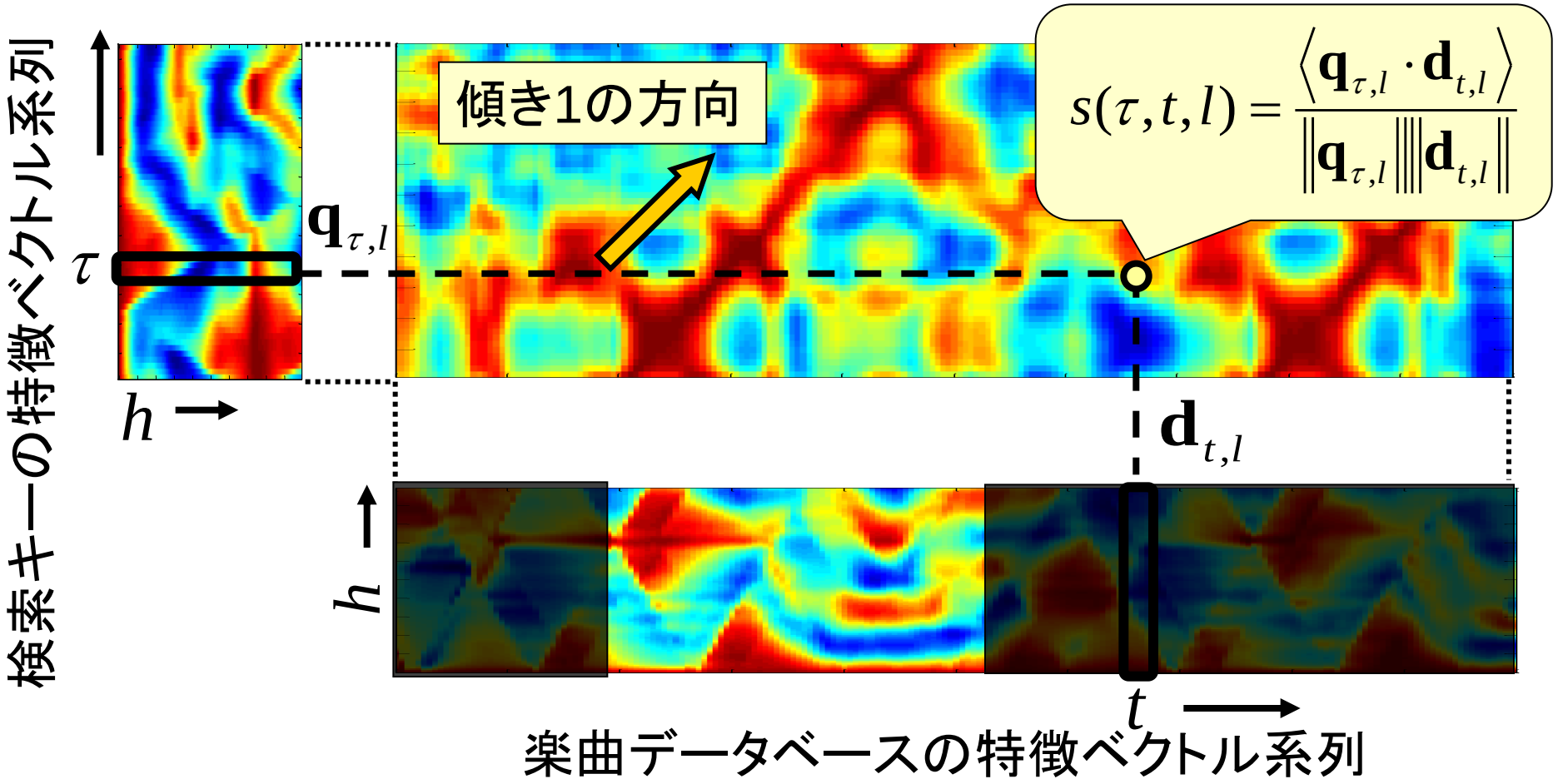
$\mathbf{q}_{t,l}$ : 検索キーの特徴ベクトル

$\mathbf{d}_{t,l}$ : 楽曲データベースの特徴ベクトル



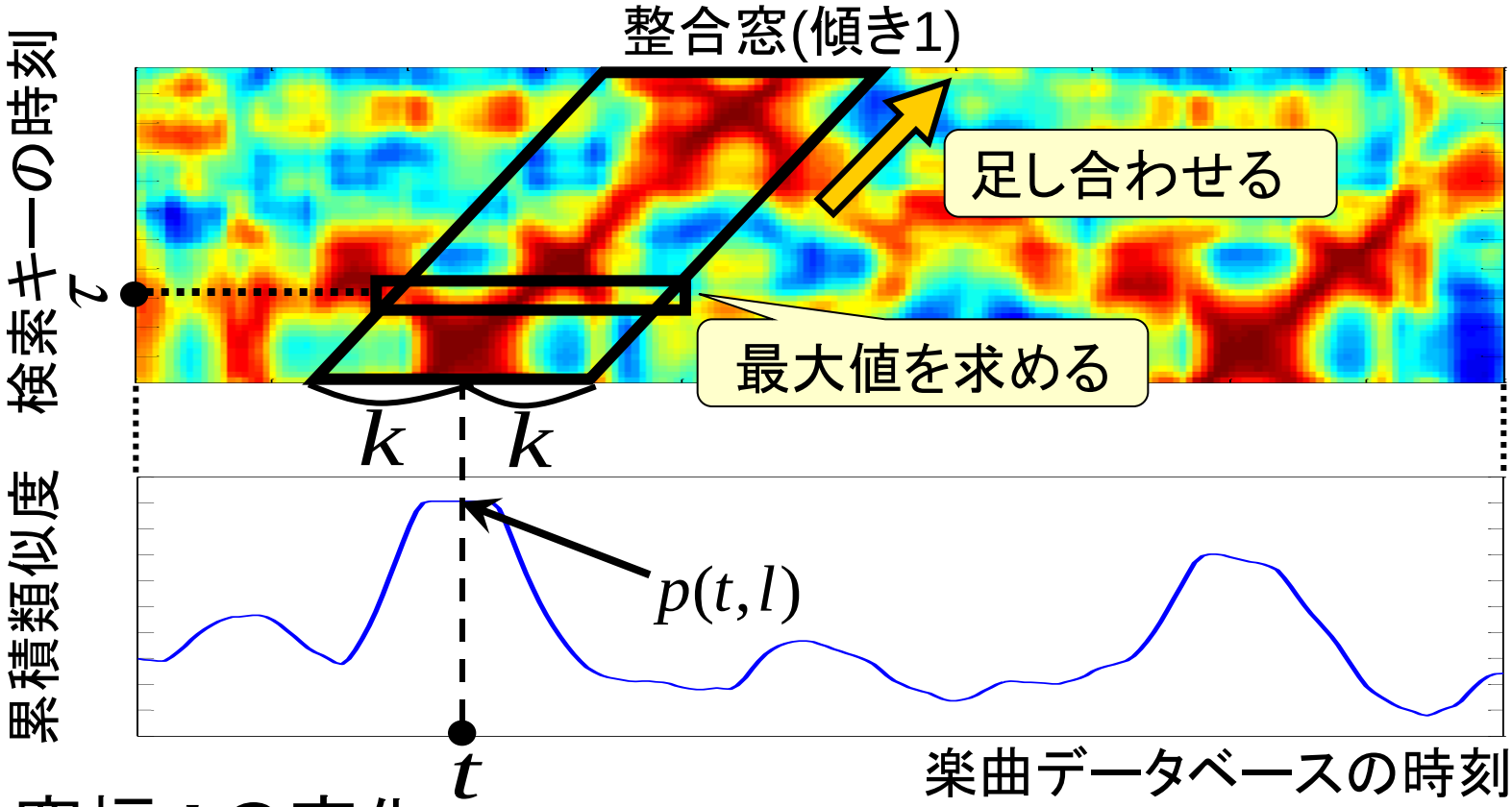
# 提案するハミング検索器(検索方法1)

- 検索キーと楽曲データベースの特徴ベクトル間の局所類似度をコサイン距離により求める



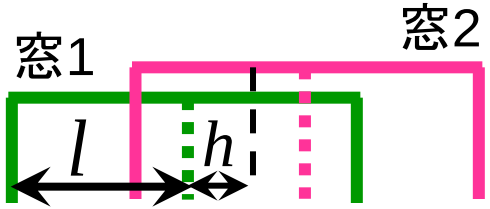
# 提案するハミング検索器(検索方法2)

- 累積類似度の算出




- 窓幅 $l$ の変化

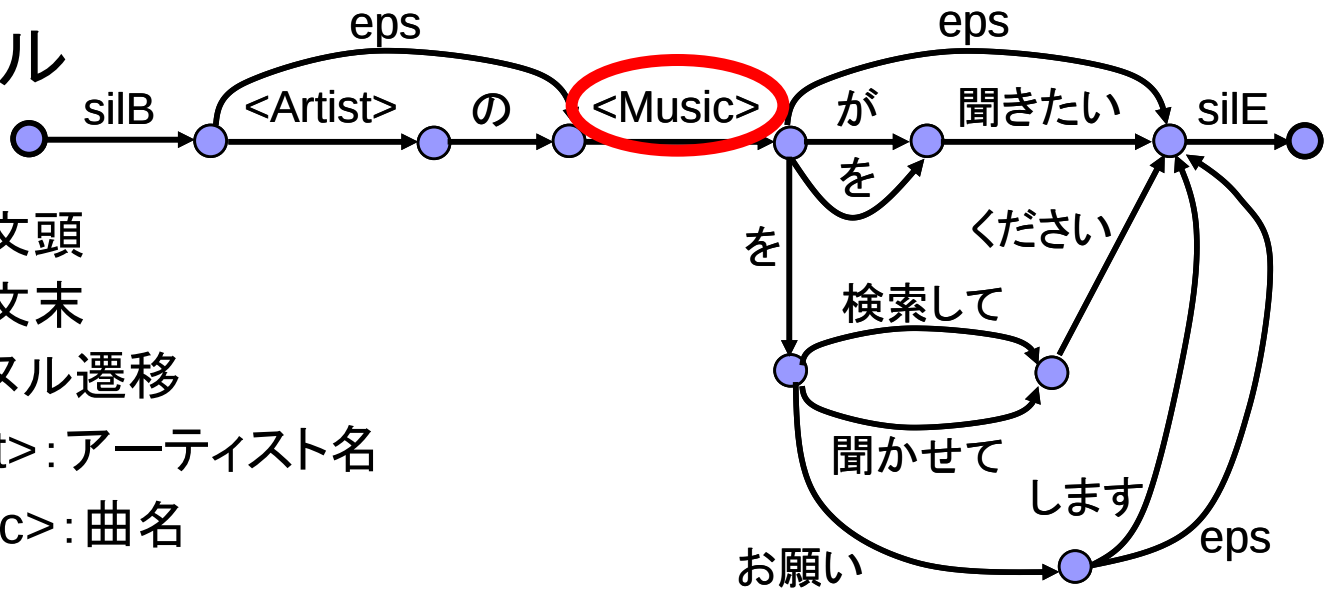
$$P(t)=\sum_{l} p(t,l)$$




$P(t)$の
最大値を求める

# 音声認識器

- 朗読音声を音声認識し，曲名から曲を検索する
- 音声認識エンジン
  - 記述文法音声認識実行キット Julian-3.4.2
- 音響モデル
  - CSRC標準日本語音響モデル
    (状態数3000/129, 性別非依存, 64混合, PTM triphone)
- 言語モデル




silB：文頭
silE：文末
eps：ヌル遷移
<Artist>：アーティスト名
<Music>：曲名

# 評価実験

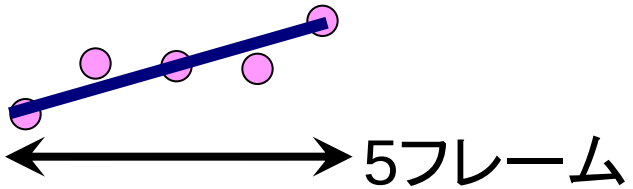
- 楽曲データベース
  - RWC音楽データベース：ポピュラー音楽100曲
- 楽曲のメロディデータ (ハミング検索器で使用)
  - メロディのF0を手作業でラベル付けした結果
- 評価データ（検索キー）
  - AISTハミングデータベース (収録被験者75名)
    - 楽曲データベースから合計25曲を選択

| | 歌声 | ハミング | 朗読音声 |
|---|---|---|---|
| 曲の出だしの部分 | 25サンプル | 25サンプル | 25サンプル |
| 曲の主題の部分 | 25サンプル | 25サンプル | 25サンプル |

  - 曲名の読み上げ音声(被験者6名による60サンプル)
    - 楽曲データベースから10曲選択，認識文法に基づいて発話

# 音声識別性能の評価

- 歌声(ハミングも含む)と朗読音声の2クラス識別
  - GMMの学習データ：曲の出だしの部分の歌声・朗読音声
    Step1. 10msごとにMFCC (12次)，F0を算出
    Step2. ΔMFCC，ΔF0の算出




5フレーム(50ms)の値から回帰係数の計算

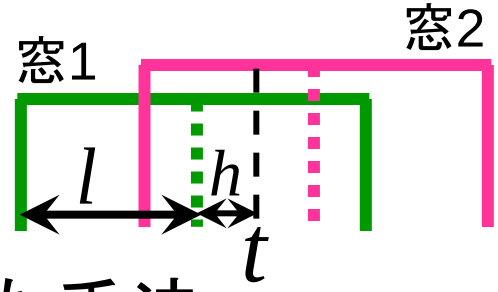
⇒ 25次元の特徴ベクトル (MFCC+ΔMFCC+ΔF0)

    Step3. GMM(16混合)による特徴ベクトルの頻度分布の学習
  - 評価(検索キー)：主題の部分の歌声・ハミング・朗読音声

| | 歌声 | ハミング | 朗読音声 | 全体 |
|---|---|---|---|---|
| 識別率 | 96.2% | 98.0% | 94.2% | 96.1% |

# ハミング検索性能の評価

- 提案手法：メロディ間の相関関係を検索に利用
  - 検索キー $\mathbf{q}_{t,l}$ ，楽曲データベース $\mathbf{d}_{t,l}$ (50次元ベクトル)
    - $h$：20msから1sまで20msずつ変化
    - $l$ ：25msから150msまで25msずつ変化
  - 整合窓 $k$：300ms



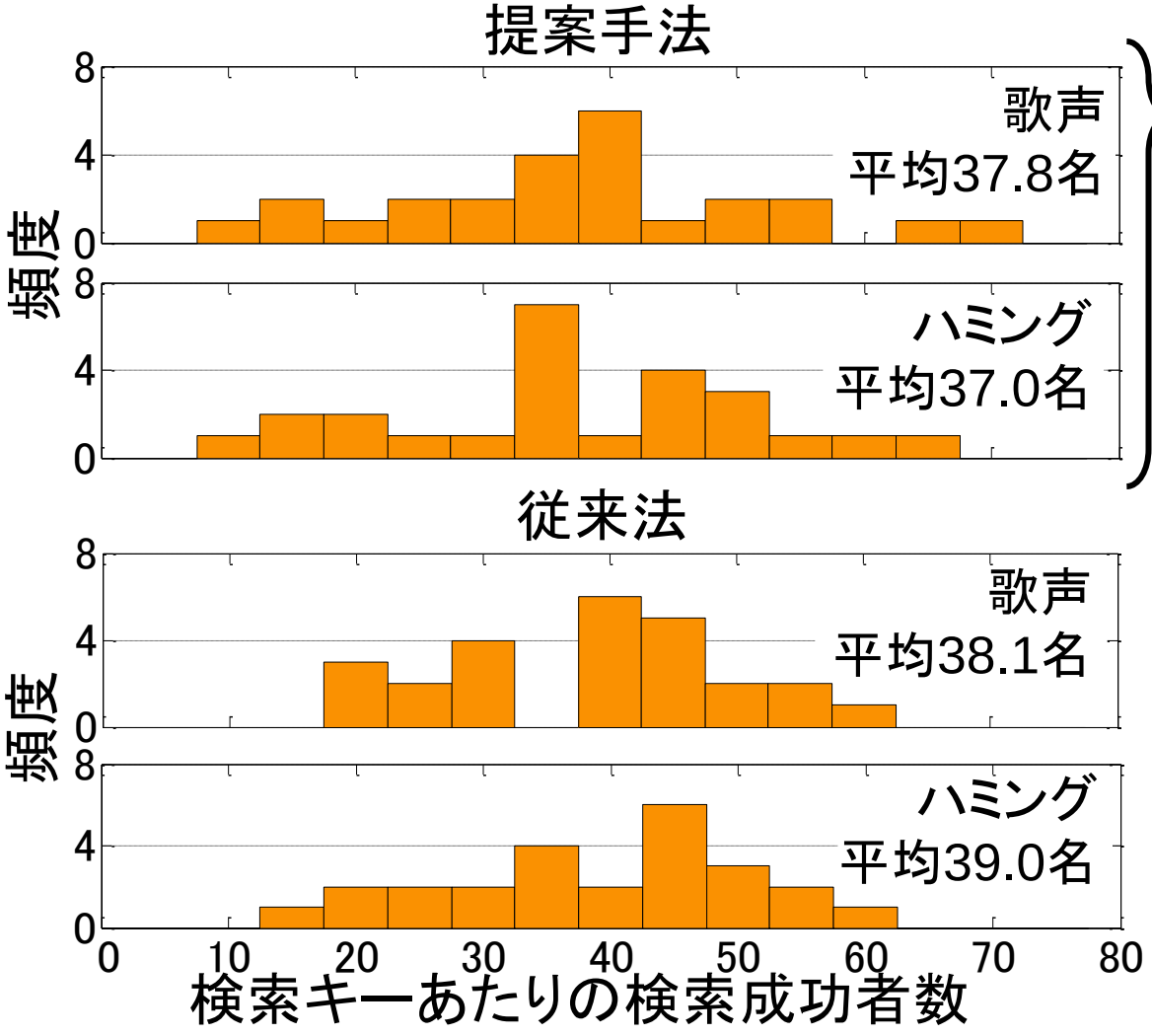

- 従来法：始端特徴依存連続DPを用いた手法
- 検索キー：25曲の主題の部分の歌声・ハミング(75名分)
- 正しい曲が検索されたとき正解として検索率を求める

| | 従来法 | | 提案手法 | |
|---|---|---|---|---|
| 検索率 | 歌声 | ハミング | 歌声 | ハミング |
| 1位 | 29.8% | 29.9% | 29.3% | 28.5% |
| 10位以内 | 50.8% | 52.1% | 50.5% | 49.3% |

検索失敗の例

# ハミング検索性能の評価

- 同一曲，同一箇所の検索キーあたりの検索成功者数
  (検索結果10位以内に正しい曲が含まれれば正解)



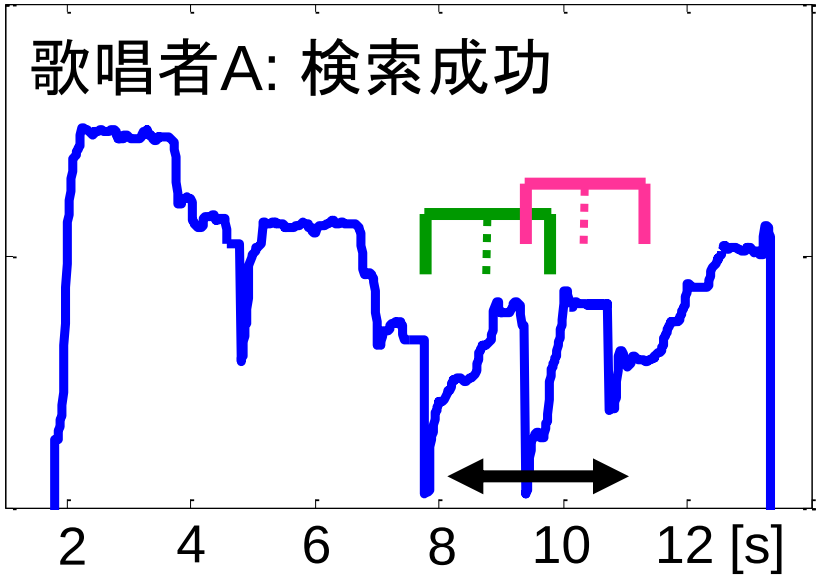

従来法に比べて，
頻度分布の分散が大きい

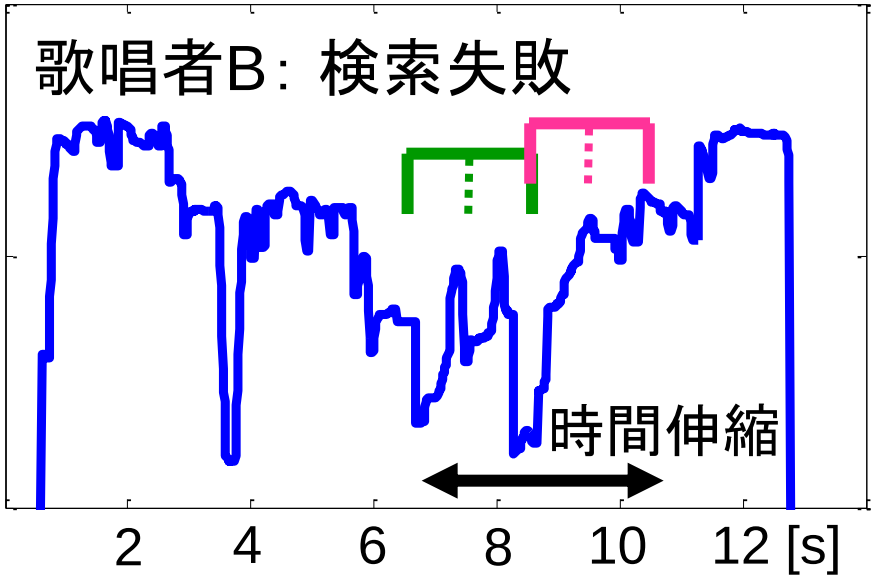
提案手法では，
検索しやすい曲もあれば
検索できない曲もある

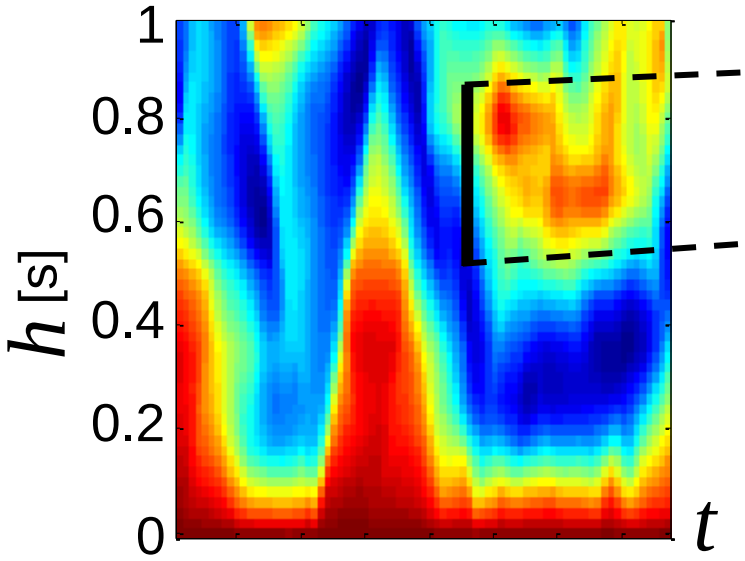
多様な曲に対応できない

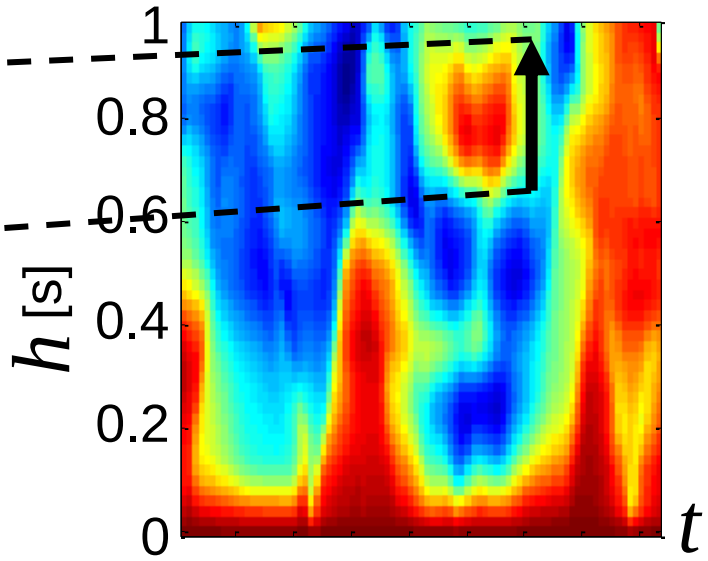
同様に提案手法では，
検索しやすい歌唱者もいれば
検索しにくい歌唱者もいる

# 提案手法の欠点→時間伸縮に対応できない



# 音声認識性能の評価

- 検索キー：曲名読み上げ音声(計60サンプル)
- 単語辞書：142単語
  (うちアーティスト33単語，曲名100単語)

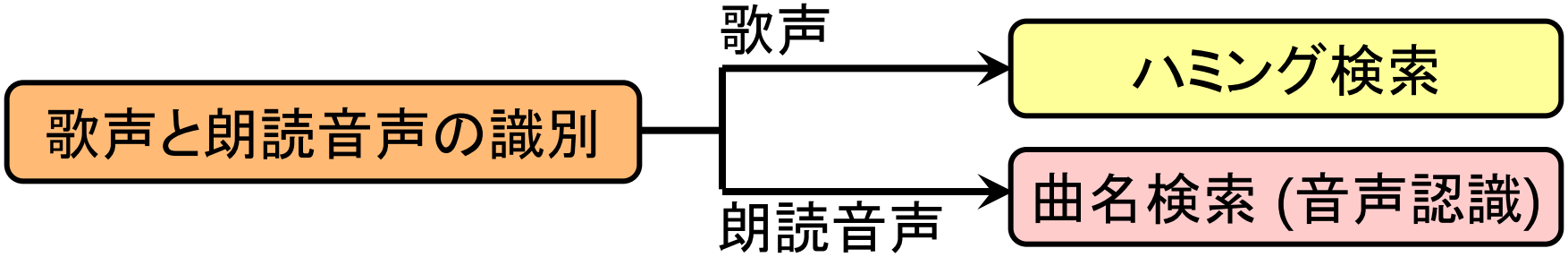
| 音声識別性能 (識別率) | 楽曲検索性能 (検索率) |
|---|---|
| 100% | 96.7% |

音声識別性能：読み上げ音声を”朗読音声”と識別した割合
楽曲検索性能：読み上げ音声により正しく曲が検索できた割合

- 認識誤り
  - 小動物(曲名)→So Long(曲名)
  - Cool motion(曲名)→Game of Love(曲名)

# まとめと今後の展開

- 歌声と朗読音声で検索可能な楽曲検索システム




- メロディ間の相関関係を利用したハミング検索手法
- 大規模な歌声データベースを利用した評価実験
  - 提案したハミング検索手法の性能は従来法に比べて低い
  - 曲や歌唱者によって検索率のばらつきが大きい
- 観測されるF0から原曲のメロディの推定