Instrogram

発音時刻検出とF0推定の不要な楽器音認識手法

北原 鉄朗*　後藤 真孝**

駒谷 和範*　尾形 哲也*　奥乃 博*

***京都大学大学院情報学研究科**

****産業技術総合研究所**

7-8 Aug. 2006

研究の目的

楽器構成に基づいた音楽情報検索

i. **楽器を直接指定**

e.g.**ピアノ曲が聴きたい，弦楽四重奏が聴きたい**

ii. **楽器構成の観点からの類似楽曲検索**

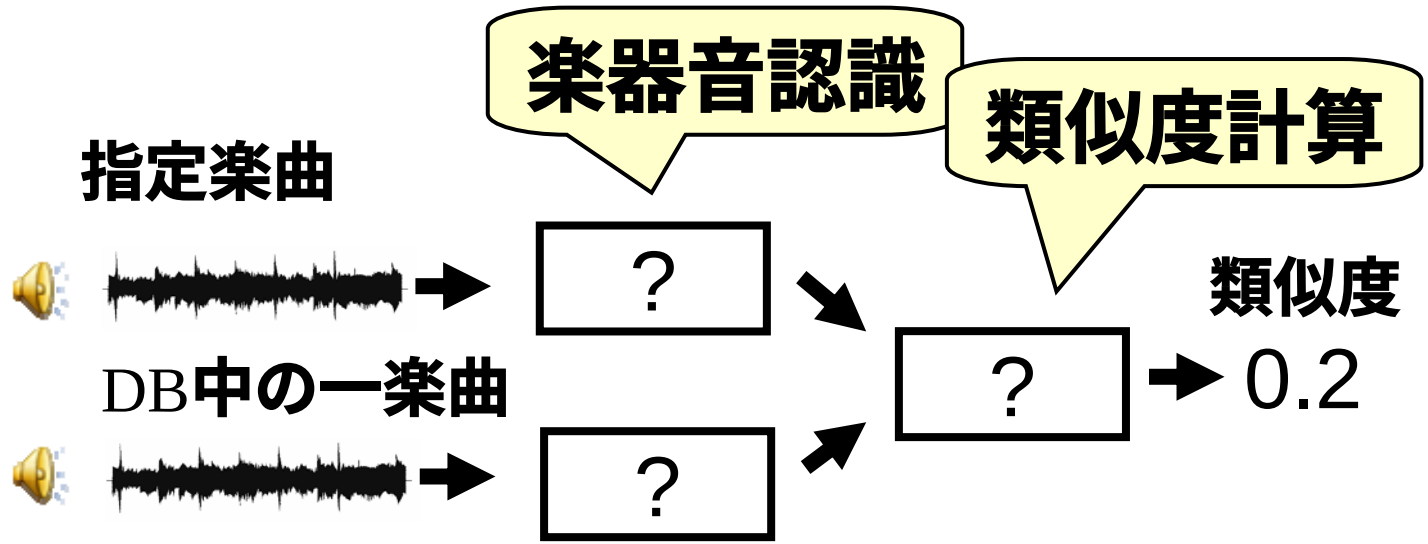
- **内容に基づく音楽情報検索のニーズ拡大**

  「この曲**が聴きたい**」**から**「こんな曲**が聴きたい**」**へ**

- **楽器構成は楽曲の雰囲気に多大な影響**

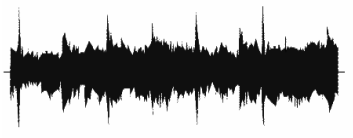
要求される要素技術

- **音響信号からの楽器の認識**
- **楽器認識結果の類似度計算**

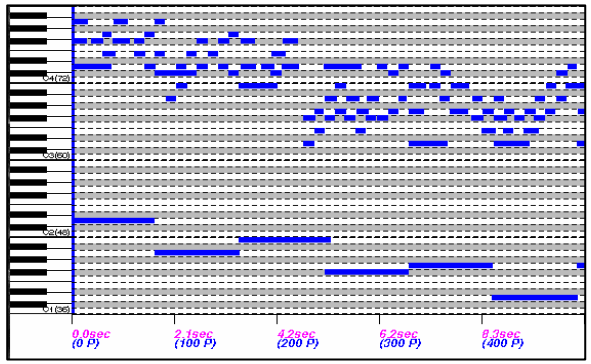


楽器音認識―従来手法―

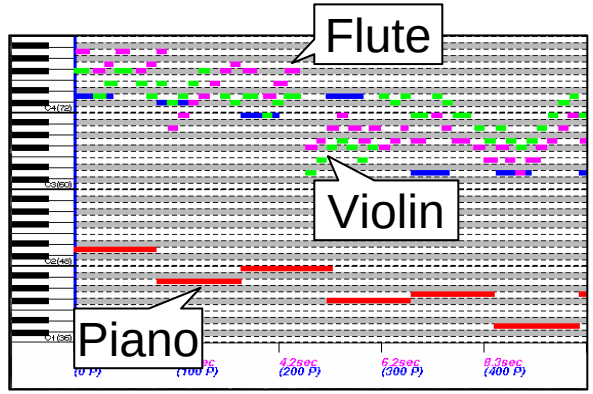
一般的な処理の流れ＝**Notewise sequential framework**



音響信号



各単音の発音時刻・音高を推定



各単音の楽器名を同定

問題点：各単音の発音時刻・音高の推定必要

検索目的なら全単音の認識必要なし

本研究のアプローチ

Instrogram

- **楽器存在確率** $p(\omega_i; t, f)$

  時刻 t **において周波数** f **をF0とする楽器** ω_i **の音が存在する確率**

  を全時刻，全周波数にわたって網羅的に計算

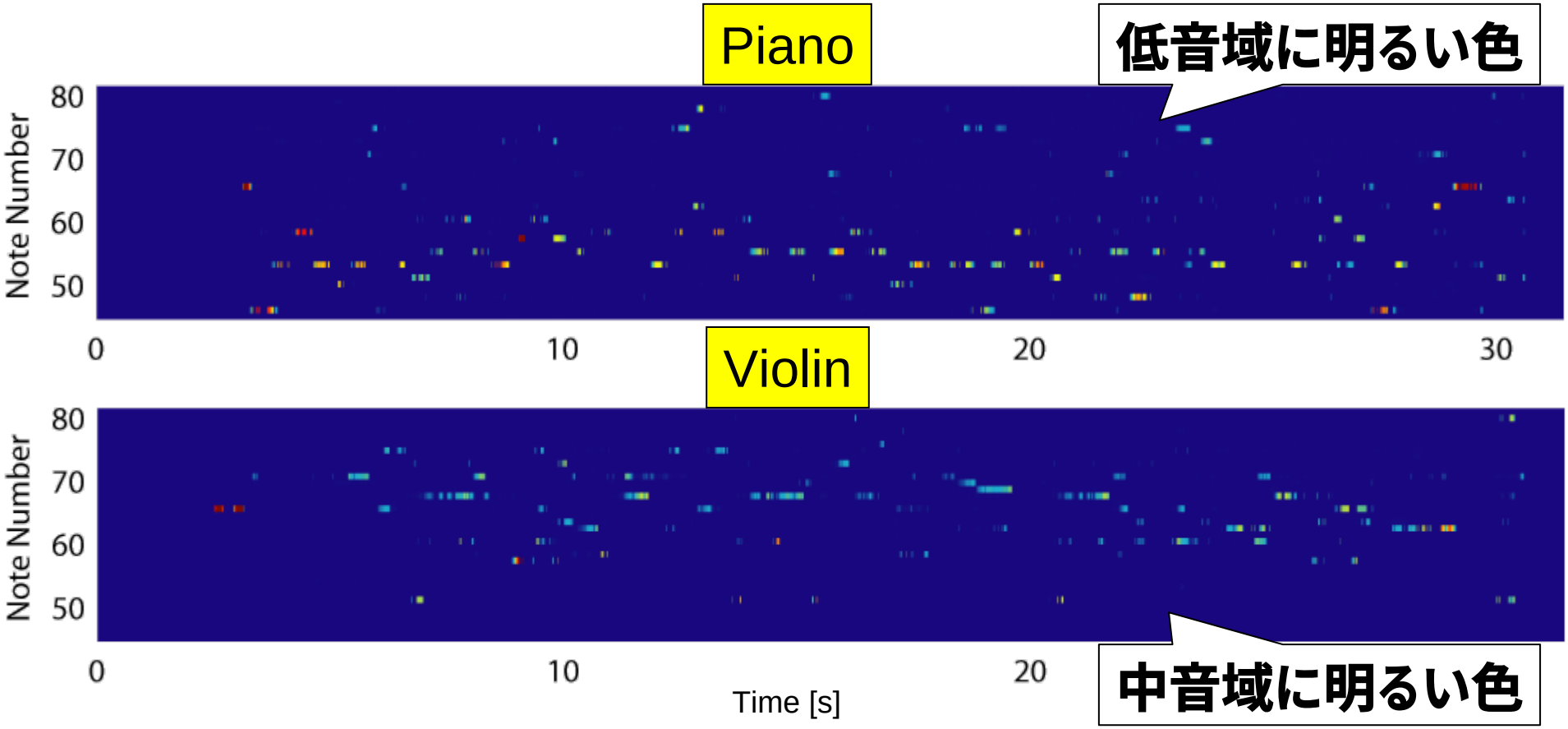
  ⇨ **Instrogram: 楽器存在確率のスペクトログラムライクな視覚表現**

- **発音時刻・音高推定相当処理を確率計算に包含し，明示的な前処理不要**

Instrogram (1/2)

楽器ごとに時間･周波数平面上に楽器存在確率を可視化

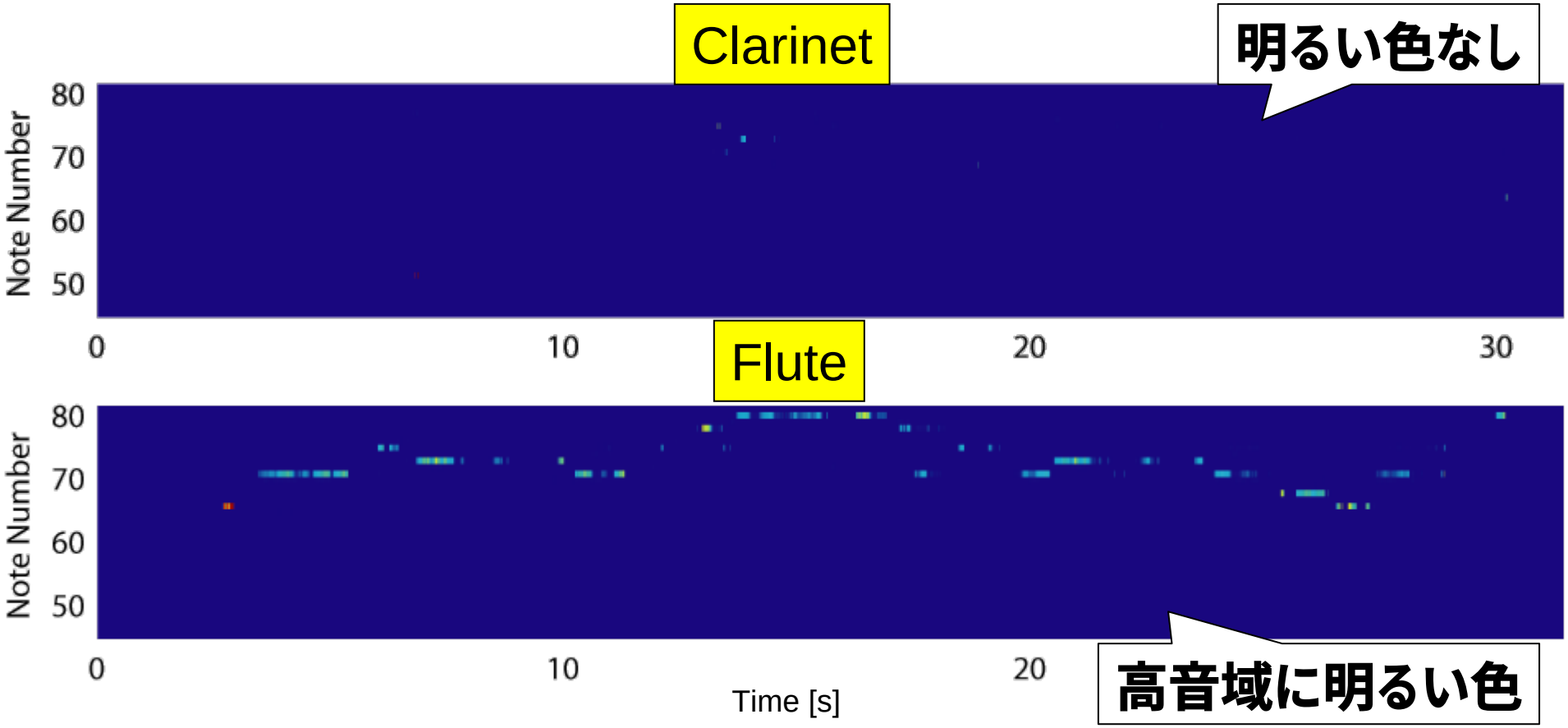
例：蛍の光（Flute-Violin-Piano）



Instrogram (2/2)

楽器ごとに時間･周波数平面上に楽器存在確率を可視化

例：蛍の光（Flute-Violin-Piano）



本研究のアプローチ

- **音響信号からの楽器の認識**

  音響信号→Instrogram（各楽器の楽器存在確率）

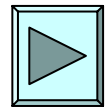
  ⇨ 前処理としての発音時刻・音高推定不要

- **楽器認識結果の類似度計算**

  Instrogram間の距離（非類似度）を計算

  ⇨ 楽器構成を連続量として表すので，
  連続的な距離尺度を自然に定義可能

単音を処理単位とせず，楽曲全体を見て，
どんな楽器があるかをおおまかにとらえる



楽器存在確率の定式化

楽器存在確率 $p(\omega_i; t, f)$ **を以下のように分解できる**

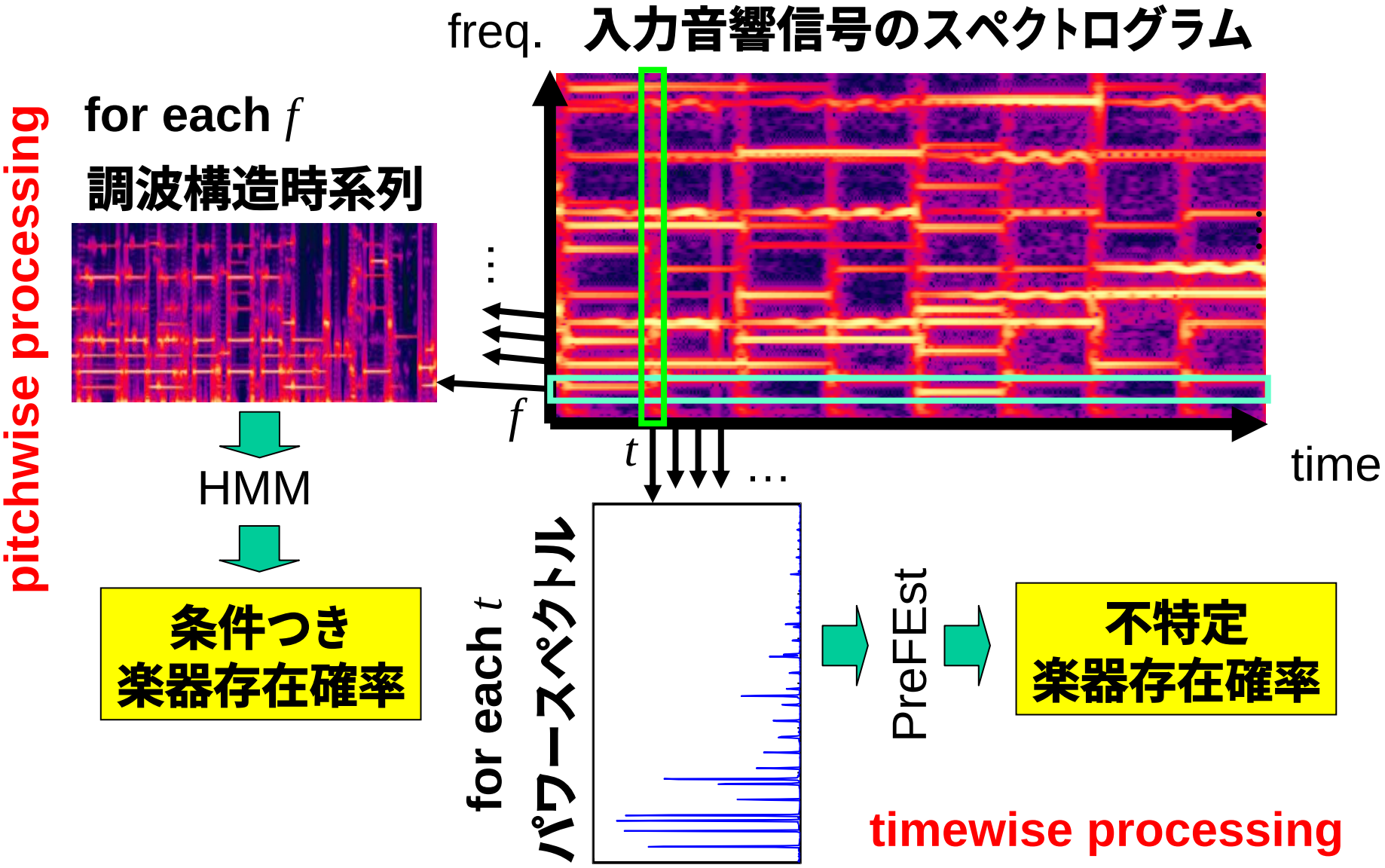
$$p(\omega_i; t, f) = p(\mathrm{X}; t, f)\ p(\omega_i | \mathrm{X}; t, f)$$

$$\because \omega_i \cap \mathrm{X} = \omega_i$$

$\mathrm{X} = \omega_1 \cup \cdots \cup \omega_m$ **(何らかの楽器が発音している事象)**

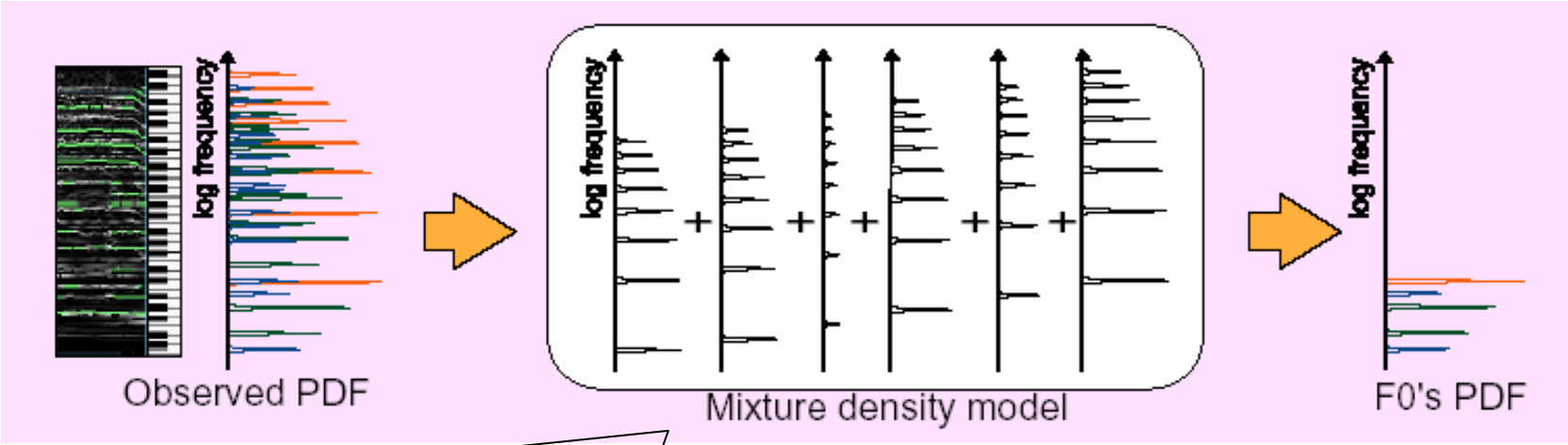
- $p(\mathrm{X}; t, f)$ **不特定楽器存在確率**
  (t, f)**に何らかの楽器音が存在する確率**
- $p(\omega_i | \mathrm{X}; t, f)$ **条件つき楽器存在確率**
  (t, f)**に楽器音が存在するとすると，それが楽器** ω_i **である確率**

確率計算アルゴリズムの概要



不特定楽器存在確率の計算

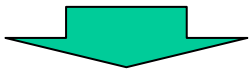
- **フレームごとに，観測されたパワースペクトルから各F0に音が存在する確率を計算**
  - ⇨ PreFEst(-core) [**後藤** ’99] **で求める**



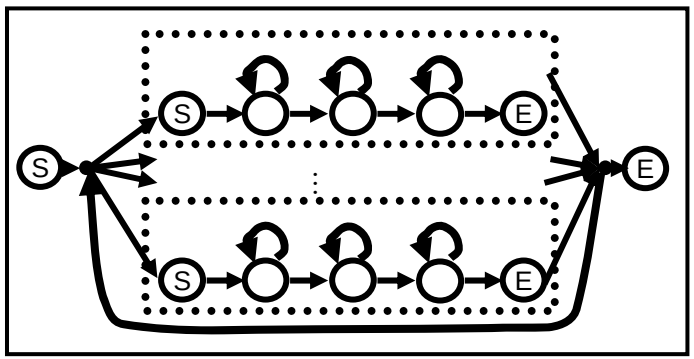
観測スペクトルをF0の異なる音モデルの加重混合とみなし，各モデルの重みをEMアルゴリズムで推定

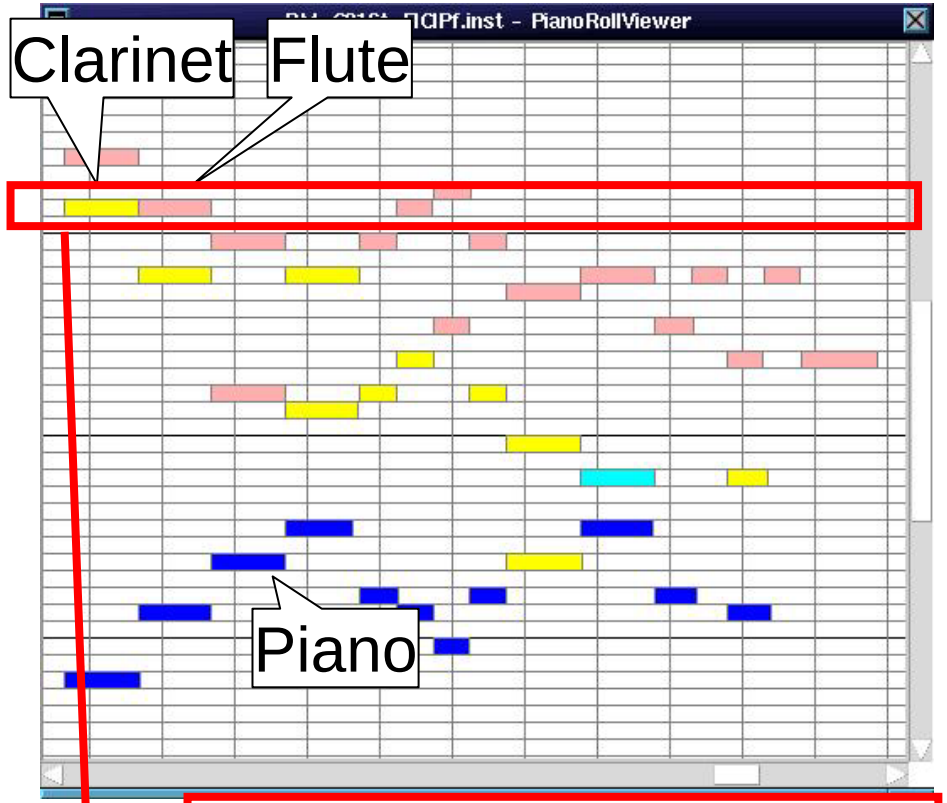
条件つき楽器存在確率の計算

- 基本的な考え方
  あるF0に着目すると，
  演奏楽器は
  (PF | … | FL | silence)+



F0ごとに分けて考えると，
音声認識と同じ枠組み
(HMM) 適用可能





この周波数に着目すると，
「Silence→Clarinet
→Flute→Silence
→Flute→Silence」

条件つき楽器存在確率の計算

- 具体的な処理手順

  F_l [Hz]**から**F_h [Hz]**まで** Δf [cent]**ごとに以下を行う**

① 調波構造抽出

当該周波数 f **を**F0**とする調波構造** $H(t,\ f)$ **を抽出**

② 特徴抽出

$H(t,\ f)$ **から特徴ベクトルの時系列**$\{x(t,\ f)\}$**を抽出**

③ 確率計算

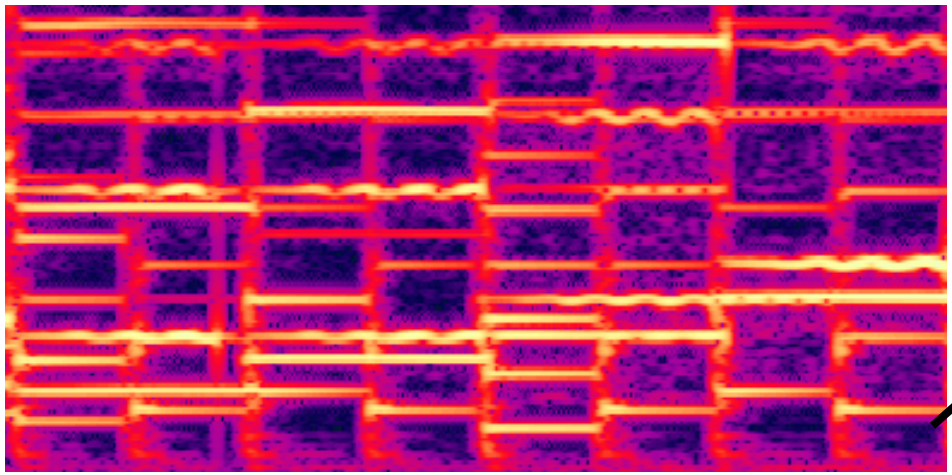
$x(t, f)$ **に対する各楽器**HMM**の尤度を計算**

条件つき楽器存在確率の計算

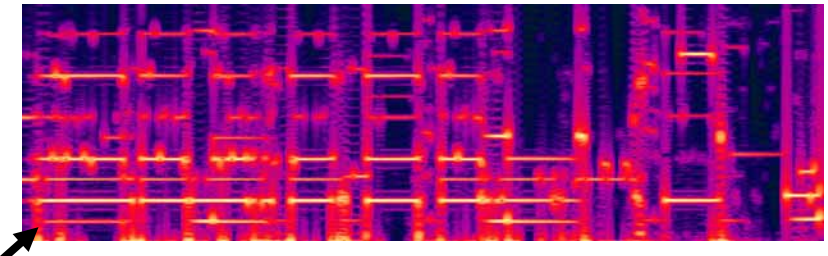
① 調波構造抽出

当該周波数 f をF0とする調波構造 $H(t,\ f\ )$ を抽出

入力音響信号のスペクトログラム



F0 = f **の調波構造** $H(t, f)$



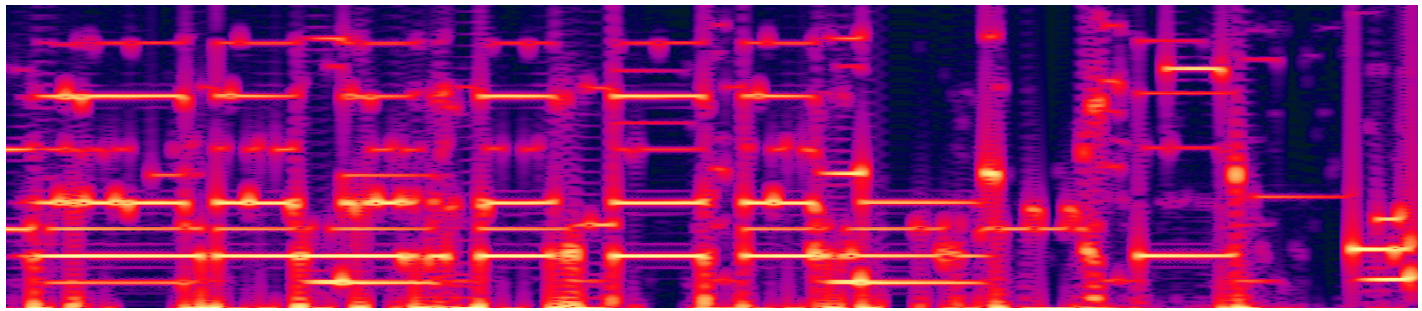
条件つき楽器存在確率の計算

② 特徴抽出

$H(t,\ f)$ **から特徴ベクトルの時系列**$\{x(t,\ f)\}$**を抽出**

F0 = f **の調波構造**

$H(t, f)$



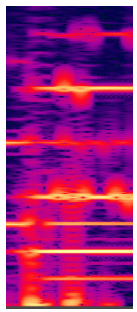
T秒分抽出

… **t秒ずらして繰り返す**

特徴ベクトル $x(t, f)$

$H_t(\tau, f)$

$(t \le \tau < t + T)$



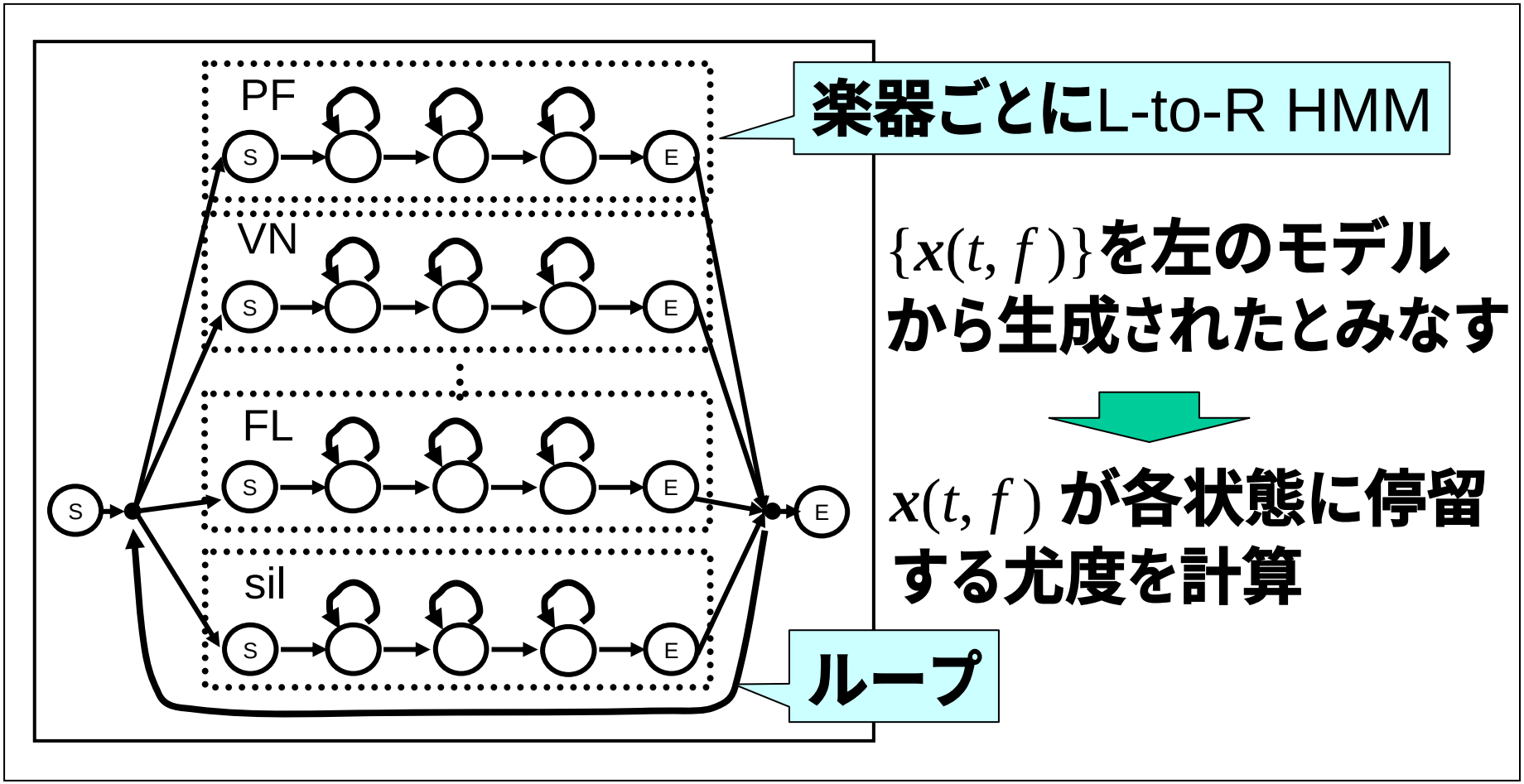
特徴抽出

- **周波数重心**
- **パワー包絡の近似直線の傾き**
- AM, FM**の振幅と振動数,** etc. (28**個**)

条件つき楽器存在確率の計算

③ 確率計算

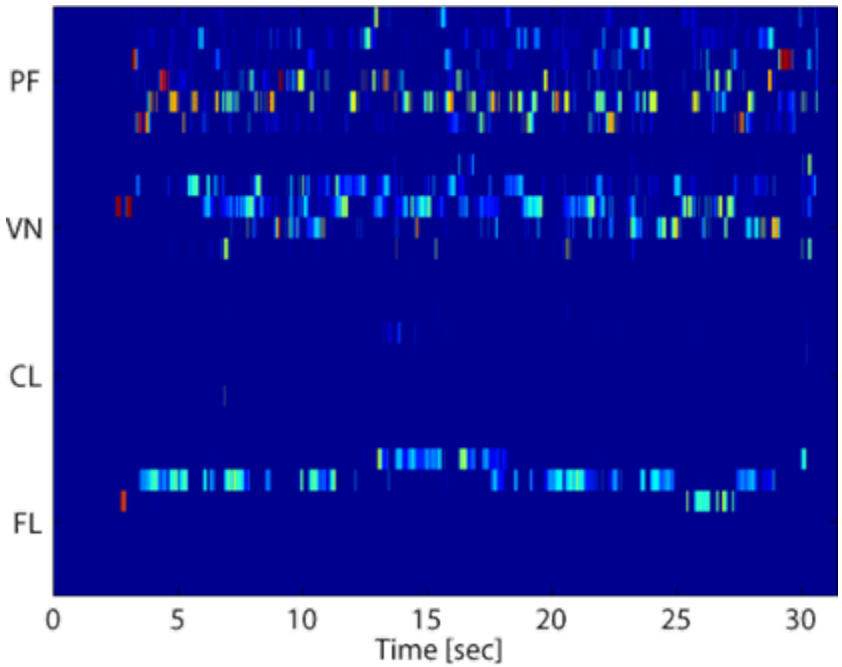
$x(t, f)$ **に対する各楽器**HMM**尤度を計算**



Instrogramの簡略化

- 周波数方向を粗く表示する
  - ⇨ 全周波数区間をN個の区間に区切り，
    k番目の区間 I_k の楽器存在確率：

$$p(\omega_i; t, I_k) = p(\omega_i; t, \sum_{f \in I_k} f)$$



Instrogram間の類似度計算

- **類似度計算のキーアイディア：**
  各時刻の楽器存在確率を特徴ベクトルとみなして，
  この時系列同士に対してDTW (**DPマッチング**)

① 時刻 t におけるベクトル $\mathbf{p}_t$：全楽器存在確率の結合

$$\mathrm{p}_t = (p(\omega_1; t, I_1), p(\omega_1; t, I_2), \cdots, p(\omega_m; t, I_N))'$$

② 2ベクトルp, q間の距離：コサイン距離

$$\mathrm{dist}(\mathrm{p}, \mathrm{q}) = 1 - (\mathrm{p}, \mathrm{q})/||\mathrm{p}|| \cdot ||\mathrm{q}||$$

$$(\mathrm{p}, \mathrm{q}) = \mathrm{p}' R \mathrm{q}$$

$$||\mathrm{p}|| = \sqrt{(\mathrm{p}, \mathrm{p})}$$

③ 上記距離尺度を用いてDTW

R：**正定値行列**
要素間の関連性考慮可能
単位行列なら通常の内積

Instrogram作成実験 (実演奏)

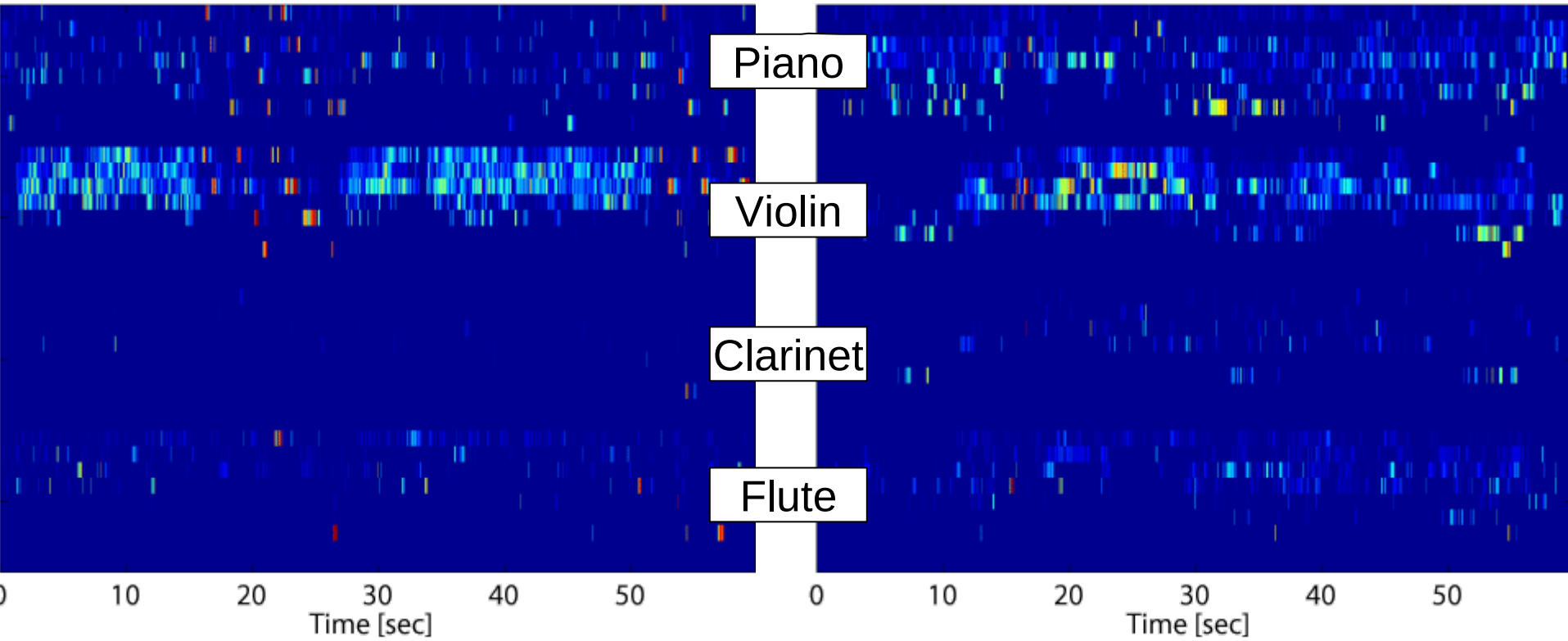
- 1〜4重奏程度の実演奏 (クラシック，ジャズ)
  (RWC音楽データベースより)
- 学習データ：RWC-MDB-I-2001，NTTMSA-P1
  で作成した切り貼り3重奏
- 対象楽器：Piano, Violin, Clarinet, Flute

Classical	(i) No. 12, 14, 21, 28	Strings
	(ii) No. 19, 40	Piano+Strings
	(iii) No. 43	Piano+Flute
Jazz	(iv) No. 1, 2, 3	Piano solo

実験結果 (1/2)

RM-C No.14 (Str.)

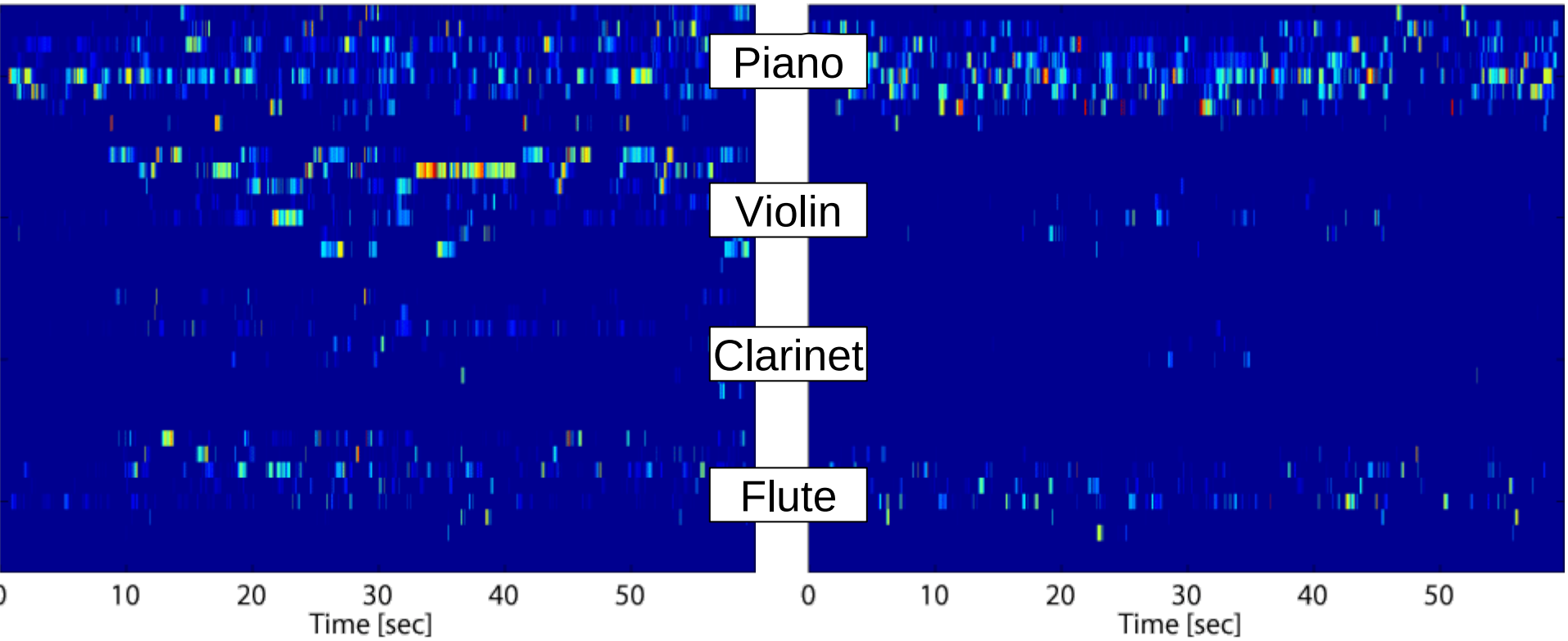
RM-C No.19 (Pf.+Str.)



- Violinの**存在確率が高い (明るい色)**
- **左に比べて右の**Pianoの**存在確率が高い**

実験結果 (2/2)

RM-C No.40 (Pf.+Str.)　　RM-J No.1 (Piano)



- **左：最初Pfだけで途中からStrが入ってくる**
- **右：Pianoだけ存在確率が高い**

Instrogram間類似度計算結果

- **グループ内非類似度は多くが7000以下**
- **弦楽曲-ピアノ曲間の非類似度はおよそ9000以上**

  ⇨実際の楽器構成を適切に反映
- **類似楽曲ベスト3**
  - Instrogram**の場合**

    弦楽曲の類似楽曲ベスト3はすべて弦楽曲

    非弦楽曲の類似楽曲ベスト3はすべて非弦楽曲
  - MFCC**の場合**

    弦楽曲と非弦楽曲の混同しばしば

考察―貢献・意義―

- **音楽可視化の立場から**
  古典的可視化・・・楽譜（自動化困難）
  音響信号可視化・・・Spectrogram, Specmurt, etc.
    ⇨**「楽器構成」の可視化は初**

- **音楽情報検索（**MIR**）の立場から**
  従来・・・Polyphonic timbre, etc.
    ⇨**楽器音認識結果を用いた**MIR**は初**

- **「しろうとの音楽理解」の立場から**
  楽譜的表現に立脚しないアプローチ

まとめ

- Instrogram: **楽器存在確率の視覚表現**
  - **不特定楽器存在確率×条件つき楽器存在確率**
  - **発音時刻検出・F0推定が不要**
  - **「どんな楽器の演奏か」を記号化せずに表現**
- Instrogram**間類似度に基づく音楽情報検索**
  - MFCC**よりも楽器構成の類似性を適切に反映**