

# 調波構造・時間包絡・音色の 統合的クラスタリングによる楽音分析

宮本 賢一<sup>†</sup> 亀岡 弘和<sup>†</sup> 西本 卓也<sup>†</sup>  
小野 順貴<sup>†</sup> 嵯峨山 茂樹<sup>†</sup>

概要：本稿では、調波構造・時間包絡の連続性・音色の類似性に基づいた音響エネルギーのクラスタリングによって、複数楽器音楽信号から、単音のクラスタリングと単音の教師なし音色クラスタリングを同時に実現する統合的な楽音分析手法 HTTC (Harmonic-Temporal-Timbral Clustering) を提案し、問題の数理的な解法を論じる。また実際の楽曲を用いての実装実験を行ない、その性能について議論・考察する。

## Harmonic, Temporal and Timbral Unified Clustering for Music Signal Analysis

KENICHI MIYAMOTO,<sup>†</sup> HIROKAZU KAMEOKA,<sup>†</sup>  
TAKUYA NISHIMOTO,<sup>†</sup> NOBUTAKA ONO<sup>†</sup>  
and SHIGEKI SAGAYAMA<sup>†</sup>

In this article, we propose an algorithm to analyze multi-instrument polyphonic music signal named Harmonic-Temporal-Timbral Clustering (HTTC) via unified clustering of acoustic spectral energies based on similarities in harmonic, temporal and timbral characteristics. We also discuss the performance of the present algorithm tested in some experiments using a multi-instrument music consisting of two instruments.

### 1. はじめに

本稿では、単一チャンネルの複数楽器音楽信号から単音と音色のクラスタリングを同時に実現する楽音分析手法を議論する。この問題は自動楽譜作成やパート追跡、音楽検索など様々な応用が挙げられる関心の高い研究課題であるが、多重音からの音高推定や楽器・音色の分類など多くの問題が内在しており、それらを同時に分析することは極めて困難な問題とされてきた。

その要素問題の一つである多声音楽信号からの音高推定手法として、我々は HTC (Harmonic-Temporal Clustering)<sup>1),2)</sup> を開発した。これは、聴覚情景分析における Bregman の分凝要件を基に設計された調波構造・時間包絡を持った単音モデルに基づき、時間数

周波数平面に分布するエネルギー（時間周波数成分）を単音ごとにクラスタリングする方法論であり、多重音信号から各単音の音高や発音時刻などの演奏情報を推定することができる。この手法は、従来のフレームワイズな手法<sup>3),4)</sup> とは異なり、スペクトルの時間・周波数双方向の構造を同時に推定する単音のクラスタリング手法として、高い性能を得ている。本稿では、この HTC の発展形として、時間周波数成分を単音ごとにクラスタリングしながら共通音色を有する単音を同一カテゴリに分類（これもいわゆるクラスタリング）していくことで、各単音の音高、発音時刻の推定だけでなく音色識別を同時に行える楽音分析の手法を検討する。

本手法を検討する上で関連する研究は、混合音の音高推定だけでなく、楽器音認識や音色特徴量推定が挙げられる。楽器認識は従来より難しい問題とされ、単旋律が主に対象とされてきたが、近年では多重音を対象とした楽器音認識の研究も行われてきている<sup>5)</sup>。主流なのは、事前処理として多重音信号の音高推定を行った結果から単音の音色の特徴量を抽出し、特徴量

<sup>†</sup> 東京大学大学院情報理工学系研究科  
Graduate School of Information Science and Technology,  
The University of Tokyo  
現在, NTT コミュニケーション科学基礎研究所に勤務  
Presently with NTT Communication Science Laboratories

空間で識別を行ったり<sup>6),7)</sup>、あらかじめ音色ラベルづけがされた楽器音色の学習を行ったり<sup>10)</sup>、教師音色特徴量とのマッチングによる楽器同定を行なう手法<sup>8),9)</sup>であった。また、最近では、音高推定を介さずに楽器音存在確率を時系列として可視化できるシステムも開発されている<sup>11)</sup>。

しかし、音高・発音時刻推定と音色識別は実はいわゆる「鶏と卵」の関係にある。すなわち、以上の楽器認識や音色特徴量推定の先行研究からも理解されるように、混合音中の各単音の音高・発音推定が精度良く行われれば、各単音の音色識別の精度は向上する。一方で、各単音の音色が正しく識別されていれば、重なり合う調波成分を分離する手がかりとなったり、オクターブ位置を同定するための手がかりとなり、音高推定の精度は飛躍的に向上する。従って、音高・発音時刻推定を行ったのちに音色識別を行う多くの従来の方針に対し、音高・発音時刻推定と音色識別は同時に行うべきであると我々は考える。

複数楽器で演奏された音楽を聴くとき、人間は、たとえ未知の楽器であっても異なる音色であればそれらを自然に区別するし、類似する音色はグルーピングして聴くことができる。本研究では、人間のこうした学習機能を計算論的に実現する手法を開発することが目的である。聴覚情景分析 (ASA) においては、人間の聴覚は、混合音をいくつかのひとまとまりの音 (単音) としてクラスタリングする機能だけでなく、各単音をいくつかの音色としてクラスタリングする機能を持つことを 1 つの仮説としている。HTC は、前者の機能を計算論的に実現したものに相当するが、後者の機能を含めることでより人間の聴覚機能に近付くと予想される。本稿で提案するこの統一的な楽音分析手法を Harmonic-Temporal-Timbral Clustering (HTTC) と呼ぶ。

## 2. 問題設定

### 2.1 音響エネルギーの観測モデル

短時間周波数分析により、入力音響信号から音響エネルギーを表すスペクトログラム  $W(x, t)$  ( $x$ : 対数周波数、 $t$ : 時刻) が得られる。この  $W(x, t)$  は、複数の音色が様々な時刻・ピッチ・音長・音量で演奏された単音の音響エネルギーの和として観測されると考えられる (図 1 参照)。

本研究の問題は、この観測モデルの逆問題として、観測された音響エネルギーを音のまとまり (以後音響オブジェクトと呼ぶ) にクラスタリングする問題、各音のピッチや音量を推定する問題、音色特徴量から音

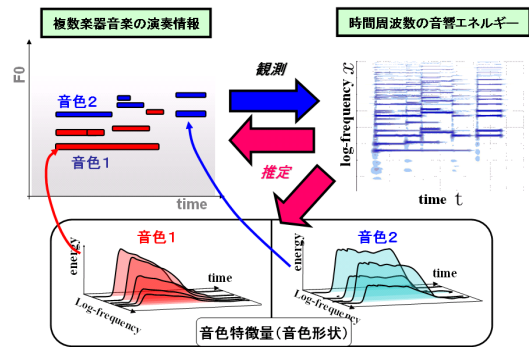


図 1 音響エネルギーの観測モデル

響オブジェクトを音色ごとに分類・各音色カテゴリの音色特徴量を推定する問題、をすべて同時に実現することといえる。

### 2.2 音色特徴量の定義

音色を決定する特徴量の定義については従来より様々な聴覚情景分析の知見や研究がある。中でも聴覚的な知見として良く知られた音の三要素 (音量・ピッチ・音色) においては、音色は音量やピッチで表せない音の性質と定義されている。また一般的に人間の聴覚では、楽器音色のクラスタリングは音長に依存しないと考えられる。以上の知見や考察から、時間-対数周波数領域の音響エネルギー分布における音色特徴量を、ピッチ・音量・発音時刻・音長に依存しない単音の音響エネルギー形状 (音色形状と呼ぶ) と定義できる。

音色の例として、ピアノとバイオリンの単音のエネルギー形状を図 2 に示す。両者は倍音成分の大きさや時間方向の減衰の仕方が大きく異なっているのが分かる。本稿では、この形状の差を音色を決定付ける要因として考え、同じ音色クラスタに属する単音は、ピッチや音量に依らず互いにエネルギーの時間・周波数形状が類似していることを仮定することにする。

さらに音楽において興味深いのは、同一楽器音ならば音長が異なっても同一楽器音色として認知できる点である。これは、単純な時間周波数パターンのクラスタリングでは音色が扱えない面があることを示唆しており、音色のクラスタリングと個々の音響オブジェクトに依存するパラメータ (音長など) の推定は同時に行うことが必要と考えられる。

### 2.3 問題設定に関する考察

前述した問題設定は、人間の聴覚で行なっている単音と音色のクラスタリングを、音響エネルギーの時間・周波数領域でモデル化したものと考えられる。音色特徴量の定義や観測モデルをエネルギー領域で設定することは、位相よりも振幅を知覚しやすい人間の聴覚の

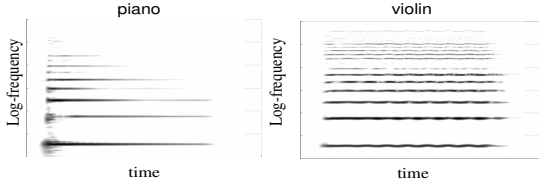


図 2 音色形状分布の例左:piano、右:violin

知見や、非負性による解の多義性解消の面で妥当であろうと考えられる。

また前節で定義した音色特徴量は、1つの楽器の全ての音色を完全に表現できる性質とはいえない。つまり音色のクラスタリングは、必ずしも楽器同定あるいは楽器分類とは一致しないと言える。例えば弦楽器の arco と pizzicato は別の音色として分類されることが自然である。またピッチによって倍音構造の相違がある楽器も多い。しかしこのような音色の例では、予備知識が無ければ人間の聴覚でも別楽器と考えることは十分あり得る。従って、HTTC における音色クラスタリングは実際の楽器同定としては原理的限界があると考えられ、人間が間違ふような音色クラスタリングの違いが生じることが予想される。

### 3. HTTC における単音・音色形状モデル

前章で述べた問題設定に対する解法として、HTC と同様、音響オブジェクトのエネルギーモデルを用いたモデルフィッティングが有効であると考えられる。そこで本章では、音色の定義や音脈の分凝条件を用いた音響オブジェクトモデルの設計について述べる。

#### 3.1 音色形状を用いた音響オブジェクトモデル

本節では音響オブジェクト  $k$  のエネルギーモデル  $q_k(x, t)$  について述べる。前述した音色の定義から、単音のエネルギー形状自体は音色カテゴリ  $c$  に依存する  $T_c(x, t)$  ( $c$ :音色カテゴリ番号、 $\iint T_c(x, t) dx dt = 1$ ) で表現されると仮定し、各音響オブジェクトは独立したパラメータとして音量  $w_k$ 、ピッチ  $\mu_k$ 、発音時刻  $\tau_k$ 、音長  $\gamma_k$  と所属音色カテゴリ  $c_k$  を持つと定義できる ( $k$ :オブジェクト番号)。ここにおいて、 $\mu_k$  と  $\tau_k$  は所属音色カテゴリの音色形状  $T_{c_k}(x, t)$  を時間-対数周波数平面上で平行移動する役割を、 $w_k$  はその音色形状エネルギーを定数倍する役割を持つ。また、 $T_c(x, t)$  が時間方向に十分長い包絡を持つと仮定すると、 $\gamma_k$  はその時間包絡を途中でミュートする役割になると考えられる (3.3 節)。

#### 3.2 音色形状分布モデル

次に、音色形状分布  $T_c(x, t)$  のパラメトリックモデ

ルの設計を述べる。音程を持つ楽器音は調波構造を持つことから、対数周波数方向に関しては図 3 のように調波構造拘束付の GMM (Gaussian Mixture Model) で表現できる。また聴覚的な知見から、音色形状は連続的な時間包絡を持つといえる。この表現には様々なモデルが考えられるが、少数パラメータでの表現性・目的関数最小化のパラメータの解の求めやすさから、本研究では、時間包絡形状に関しても図 3 のような GMM で表現する。よって、音色形状  $T_c(x, t)$  は調波構造拘束付 2 次元 GMM で表現できる。音響オブジェクト依存のピッチ、発音時刻を原点に配置すると、

$$T_c(x, t) = \sum_{n,y} T_{c,n,y}(x, t) \quad (1)$$

$$T_{c,n,y}(x, t) = \frac{v_{c,n} u_{c,n,y}}{2\pi\sigma\phi} e^{-\frac{(x-\log n)^2}{2\sigma^2} - \frac{(t-y\phi)^2}{2\phi^2}} \quad (2)$$

$$\forall c, \sum_n v_{c,n} = 1, \quad \forall c, \forall n, \sum_y u_{c,n,y} = 1 \quad (3)$$

と定式化される。ここにおいて、 $v_{c,n}$  は調波構造 GMM の Gaussian の重み、 $u_{c,n,y}$  は倍音ごとの時間包絡 GMM の Gaussian の重みであり、形状のスケールの規格化条件として式 (3) を満たす。なお、音色クラスタリングにおいて音色形状の比較をやすくするために、Gaussian の分散は全音色カテゴリで共通とした。

#### 3.3 音長を考慮した音響オブジェクトモデル

音長の自由度を持ったオブジェクトのエネルギー形状の表現は、各音響オブジェクトにおいて音色形状の時間包絡 GMM の重み  $u_{c,n,y}$  に、音長  $\gamma_k$  までは 1、以降は 0 となるような連続消音関数  $R(t - \gamma_k)$  をかけることで実現できる。この消音関数についても様々な形状が考えられるが、最も簡単な関数として sigmoid 関数  $R(t - \gamma_k) = \frac{1}{1 + e^{p(t - \gamma_k)}}$  が考えられる。

以上の設計方針をまとめると、 $q_k(x, t)$  は

$$\begin{aligned} q_k(x, t) &= \sum_{n,y} S_{k,n,y}(x, t) \\ S_{k,n,y}(x, t) &= w_k T_{c_k,n,y}(x - \mu_k, t - \tau_k; \gamma_k) \\ &= w_k \sum_{n,y} \frac{1}{1 + e^{p(y\phi - \gamma_k)}} \frac{v_{c_k,n} u_{c_k,y}}{2\pi\sigma\phi} \\ &\quad e^{-\frac{(x - \mu_k - \log n)^2}{2\sigma^2} - \frac{(t - \tau_k - y\phi)^2}{2\phi^2}} \quad (4) \end{aligned}$$

と定式化できる。このモデルの推定パラメータをまとめると表 1 のようになる。

## 4. 反復推定アルゴリズム

4.1 解法の方針:音響エネルギーのクラスタリング  
先行研究の HTC<sup>1)</sup> においては、観測された音響エネルギーを各音響オブジェクトごとに分配することで、

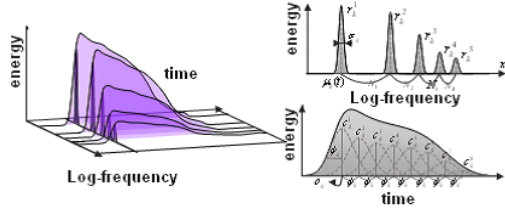


図 3 音色形状分布左:概形、右上:対数周波数方向の GMM、右下:時間方向の GMM

表 1 音響オブジェクトモデルの推定パラメータ

表記	物理的意味
$w_k$	オブジェクト $k$ の音量
$\mu_k$	オブジェクト $k$ のピッチ周波数
$\tau_k$	オブジェクト $k$ の発音時刻
$\gamma_k$	オブジェクト $k$ の音長
$c_k$	オブジェクト $k$ の所属音色カテゴリ番号
$v_{c,n}$	音色 $c$ の第 $n$ 倍音のエネルギー比
$u_{c,n,y}$	音色 $c$ の第 $n$ 倍音における時間方向エネルギー比

モデルフィッティングによりピッチや発音時刻を得ることを可能にしていた。この分配エネルギーは単音のエネルギー形状を表現していることから、本研究における観測エネルギーの分配は、ピッチなどを推定するだけでなく、同時に音色の分類や形状推定にも有益であると考えられる。

以上の考察に基づき、本研究の解法方針を述べる。各  $(x, t)$  において観測エネルギーから音響オブジェクト  $k$  への分配関数  $m_k(x, t)$  ( $0 \leq m_k(x, t) \leq 1$ ,  $\sum_k m_k(x, t) = 1$ ) を導入し、前章で設計した音響オブジェクトモデル  $q_k(x, t)$  とオブジェクト  $k$  への分配エネルギー  $m_k(x, t)W(x, t)$  とのフィッティングを行なうことで、ピッチや音量、音色カテゴリや各音色形状の推定が実現できる。さらに推定された演奏情報や音色形状から  $m_k(x, t)$  を適切に再更新することで、さらによりパラメータの推定が可能になると考えられる。

#### 4.2 分布間距離最小化の 4 ステップ反復更新アルゴリズム

オブジェクト  $k$  の分配エネルギー  $m_k(x, t)W(x, t)$  と音響オブジェクトモデル  $q_k(x, t)$  のフィッティングの「良さ」を表す分布間距離として I ダイバージェンスを用いると、目的関数

$$J = \sum_k J_k \quad (5)$$

$$J_k = \iint_D m_k(x, t)W(x, t) \log \frac{m_k(x, t)W(x, t)}{q_k(x, t)} - (m_k(x, t)W(x, t) - q_k(x, t)) dx dt \quad (6)$$

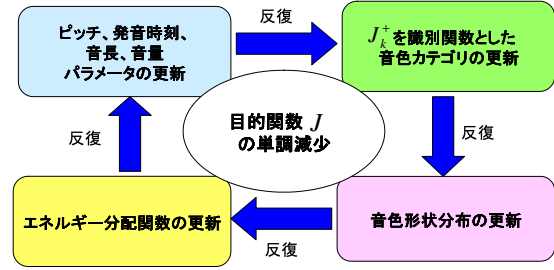


図 4 目的関数最小化を実現する 4 ステップ反復アルゴリズム

を制約条件

$$\forall x, \forall t, \sum_k m_k(x, t) = 1 \quad (7)$$

の元で最小化する問題として定式化できる。

ここで各オブジェクトに分配されたエネルギー  $m_k(x, t)W(x, t)$  をさらに各 2 次元 Gaussian  $S_{k,n,y}(x, t)$  に分配する関数  $m_{k,n,y}(x, t)$  を導入すると、Jensen の不等式より

$$J_k \leq \sum_{n,y} \iint_D m_k(x, t) m_{k,n,y}(x, t) W(x, t) \log \left( \frac{m_k(x, t) m_{k,n,y}(x, t) W(x, t)}{q_k(x, t)} \right) = J_k^+ \quad (8)$$

を満たす。このとき等式は

$$m_{k,n,y}(x, t) = \frac{S_{k,n,y}(x, t)}{\sum_{n,y} S_{k,n,y}(x, t)} \quad (9)$$

のときのみ成立する。以上より、 $J^+ = \sum_k J_k^+$  として

- (1) 分配関数  $m_k(x, t) m_{k,n,y}(x, t)$  の更新
- (2)  $J^+$  を最小化するオブジェクト依存パラメータ  $w_k, \mu_k, \tau_k, \gamma_k$  の更新
- (3)  $J^+$  を最小化する各オブジェクトの音色カテゴリ  $c_k$  更新
- (4)  $J^+$  を最小化する各音色カテゴリの音色形状  $T_c(x, t)$  の更新

の 4 ステップ更新の反復によって (図 4)、 $J$  の単調減少が保証され、局所最適解に収束するアルゴリズムが実現できる。

#### 4.3 パラメータ更新式導出

本節では表 1 で示した推定パラメータの更新式を導出する。

##### 4.3.1 分配関数の更新

式 (7) の制約下で式 (5)(6) の目的関数を  $m_k(x, t)$  に関して最小化すると、 $m_k(x, t)$  の最適解は

$$\hat{m}_k(x, t) = \frac{q_k(x, t)}{\sum_k q_k(x, t)} \quad (10)$$

と求まる。これと式 (9) より、目的関数を減少させる

分配関数の更新を

$$\hat{m}_k(x, t) \hat{m}_{k,n,y}(x, t) = \frac{S_{k,n,y}(x, t)}{\sum_k S_{k,n,y}(x, t)} \quad (11)$$

と実現できる。なおこの更新によって  $J$  と  $J^+$  の値は一致する。

#### 4.3.2 オブジェクトパラメータの更新

式 (8) を音響オブジェクトの各パラメータで偏微分することで、目的関数最小化を実現する更新式を導出できる。ここでは導出は省略し更新式のみを示す。

$$\hat{w}_k = \frac{\sum_{n,y} \Phi_{k,n,y}}{\sum_{n,y} v_{c_k,n} u_{c_k,n,y} R(\gamma_k - y\phi)} \quad (12)$$

$$\hat{\mu}_k = \frac{\sum_{n,y} \iint (x - \log n) l_{k,n,y}(x, t) dx dt}{\sum_{n,y} \Phi_{k,n,y}} \quad (13)$$

$$\hat{\tau}_k = \frac{\sum_{n,y} \iint (t - y\phi) l_{k,n,y}(x, t) dx dt}{\sum_{n,y} \Phi_{k,n,y}} \quad (14)$$

$$\Phi_{k,n,y} = \iint l_{k,n,y}(x, t) dx dt \quad (15)$$

$$l_{k,n,y}(x, t) = m_k(x, t) m_{k,n,y}(x, t) W(x, t) \quad (16)$$

音長  $\gamma_k$  に関しては解析解が求まらないが、ステップサイズ探索付きの勾配法による更新が可能である。しかし更新の収束速度の問題から、本研究では離散的な候補からの最適解探索で更新する。

$$\begin{aligned} \hat{\gamma}_k &= \hat{\xi}_k \phi \\ \hat{\xi}_k &= \underset{\xi}{\operatorname{argmax}} \left( - \sum_{n,y} \iint l_{k,n,y}(x, t) \right. \\ &\quad \left. \log R((y - \xi)\phi) dx dt \right. \\ &\quad \left. + w_k \sum_{n,y} v_{c_k,n} u_{c_k,n,y} R((y - \xi)\phi) \right) \quad (17) \end{aligned}$$

#### 4.3.3 音色形状分布の更新

音色形状についても、各音色カテゴリごとに式 (3) の制約下で式 (8) を最小化する  $v_{c,n}$ 、 $u_{c,n,y}$  の更新式を導出でき、

$$\hat{v}_{c,n} = \frac{\sum_{c_k=c} \sum_y \Phi_{k,n,y}}{\sum_{c_k=c} \sum_{n,y} \Phi_{k,n,y}} \quad (18)$$

$$\hat{u}_{c,n,y} = \frac{\sum_{c_k=c} \Phi_{k,n,y}}{\lambda_{u,c,n} + \sum_{c_k=c} w_k v_{c,n} R(y\phi - \gamma_k)} \quad (19)$$

と求まる。ただし、ラグランジュ未定乗数  $\lambda_{u,c,n}$  は

$$\sum_y \frac{\sum_{c_k=c} \Phi_{k,n,y}}{\lambda_{u,c,n} + \sum_{c_k=c} w_k v_{c,n} R(\gamma_k - y\phi)} = 1 \quad (20)$$

の解であり、ニュートン法により高速に求まる。

#### 4.3.4 音色カテゴリの更新

音色カテゴリ番号  $c_k$  の更新も、 $J_k^+$  を識別関数としたクラスタリングで実現でき、更新式は

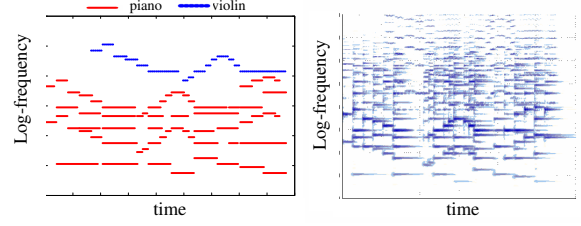


図 5 左:入力楽曲の演奏情報 赤:piano、青:violin、右: 観測音響エネルギー

$$\begin{aligned} \hat{c}_k &= \underset{c}{\operatorname{argmax}} \sum_{n,y} \Phi_{k,n,y} \log(v_{c,n} u_{c,n,y}) \\ &\quad - w_k \sum_{n,y} v_{c,n} u_{c,n,y} R(\gamma_k - y\phi) \quad (21) \end{aligned}$$

と求まる。

## 5. 実装システムの適用例

本章では、提案アルゴリズムを実装して実際の楽曲に適用した例を示し、本手法の性能について考察する。

### 5.1 実験条件

入力用の楽曲は、RWC 研究用音楽データベースより RWC-MDB-C-2001 No.39 から、ピアノとバイオリンで演奏される冒頭部 10 秒程度を抜粋して利用した。本稿では特に音色の分類と形状学習を考察するため、MIDI 形式のデータから生成した WAV 形式の信号を入力とし、音響オブジェクト数とピッチ・発音時刻の初期値は MIDI の正解を与えて実験を行なった。入力の演奏情報と観測音響エネルギーを図 5 に示す。

### 5.2 音色カテゴリ数 2 での推定

まず、音色カテゴリ数を 2 として実験を行った。推定された演奏情報と 2 つの音色形状分布を図 6,7 に示す。この結果より、ピアノのメロディ部分の音色は同一音色としてグルーピングされ、第一倍音が小さいなど形状もピアノの特徴を妥当に推定できたが、ピアノの低音とバイオリンが同一音色カテゴリに分類される結果となった。これは 2.3 節で考察したように、ピアノの音高が広い領域に分布するため、ピッチに依存して倍音比が変動することが原因であると考えられる。

### 5.3 音色カテゴリ数を増加した実験

前節の結果を踏まえて、音色カテゴリ数を 3 にした実験を行なった。結果の演奏情報と音色形状を図 8,9 に示す。音色カテゴリ数を増やしたことにより、ピアノの音色が音色 1 と 2 の二つに区別されたものの、バイオリンのオブジェクトはほとんどが音色 3 にクラスタリングされており、ピアノの低音との音色混同も大きく減少している様子が観察できる。

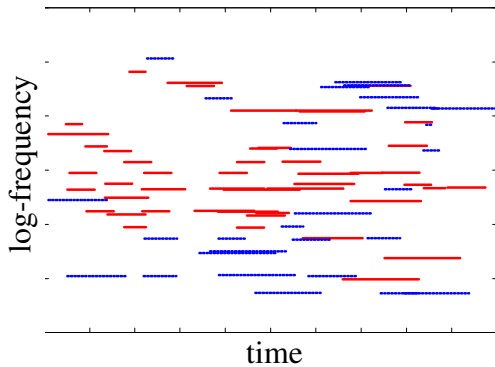


図 6 推定された演奏情報 (音色カテゴリ数 2)

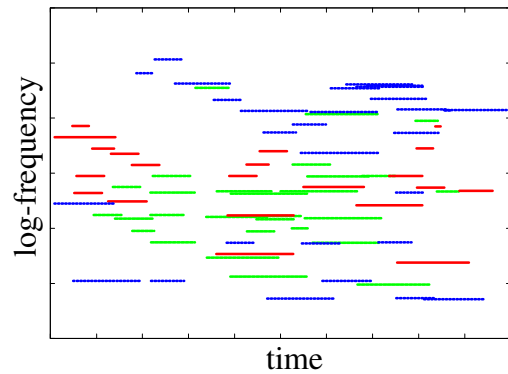


図 8 推定された演奏情報 (音色カテゴリ数 3)

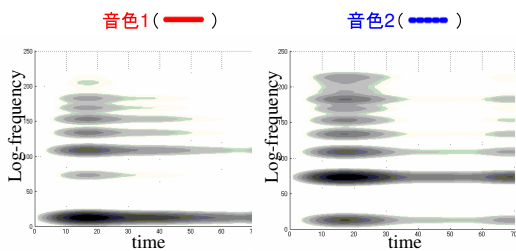


図 7 推定された音色形状 (音色カテゴリ数 2)

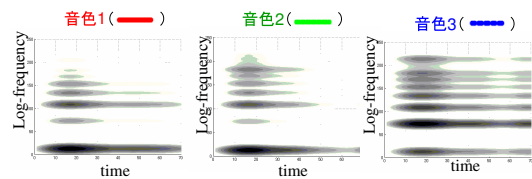


図 9 推定された音色形状 (音色カテゴリ数 3)

## 6. 結 論

本研究では、調波構造・時間包絡の連続性・音色の類似性に基づいた音響エネルギーのクラスタリングにより、単音と音色のクラスタリングを同時に実現する楽音分析手法を提案し、また実際の楽曲を用いた適用例を示した。今後の課題として、ピッチに依存した倍音比の相違を考慮した音色モデルの設計、音響オブジェクト数の妥当性を考慮したアルゴリズムの提案、非調波構造の楽器も含めた統合的な複数音楽分析手法の提案などを検討したい。

## 謝 辞

本研究の一部は科学研究費補助金・基盤研究B(課題番号 17300054) および科学技術振興機構 CREST プロジェクトの補助を受けて行なわれた。

## 参 考 文 献

- 1) H.Kameoka *et al.*, "A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering," *IEEE Trans. on Audio, Speech and Language Processing*, in Press.
- 2) 亀岡弘和, ルルー ジョナトン, 小野順貴, 嵯峨山茂樹, "調波時間構造化クラスタリングによる CASA へのアプローチ," 日本音響学会聴覚研究会, Vol. 36, No. 7, H-2006-103, pp. 575-580, 2006.

- 3) A. Klapuri, "Multiple Fundamental Frequency Estimation Based on Harmonicity and Spectral Smoothness," *IEEE Trans, Speech and Audio Proc*, 2003.
- 4) M.Goto, "A Robust Predominant-F0 Estimation Method for Real-time Detection of Melody and bass lines in CD recordings," in *Proc. ICASSP*, 2000.
- 5) T. Kitahara, "Computational Musical Instrument Recognition and Its Application to Content-based Music Information Retrieval," Ph.D. Thesis, Kyoto University, 2007.
- 6) 北原 鉄朗, 後藤 真孝, 奥乃 博, "音高による音色変化に着目した楽器音の音源同定: F0 依存多次元正規分布に基づく識別手法," 情報処理学会論文誌, Vol.44, No.10, pp.2448-2458, 2003.
- 7) 北原 鉄朗, 後藤 真孝, 奥乃 博, "音響的類似性を反映した楽器の階層表現の獲得とそれに基づく未知楽器のカテゴリレベルの音源同定," 情報処理学会論文誌, Vol.45, No.3, pp.680-689, 2004.
- 8) J. Eggink, G. J. Brown, "A Missing Feature Approach to Instrument Identification in Polyphonic Music," *Proc, ICASSP*, 2003.
- 9) K.Kashino, Hiroshi Murase, "A sound source identification system for ensemble music based on template adaptation and music stream extraction," *Speech Communication*, vol.27, 1999.
- 10) A.Klapuri, "Analysis of Musical Instrument Sounds by Source-Filter-Decay Model," *Proc, ICASSP*, 2007.
- 11) T. Kitahara, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno, "Instrogram: Probabilistic Representation of Instrument Existence for Polyphonic Music", *IPSSJ Journal*, Vol.48, No.1, pp.214-226, 2007.