

# 音色特徴量分布の利用による 調波・非調波統合モデルのパラメータ推定

糸山 克寿<sup>†</sup> 後藤 真孝<sup>‡</sup> 駒谷 和範<sup>†</sup> 尾形 哲也<sup>†</sup> 奥乃 博<sup>†</sup>

<sup>†</sup> 京都大学大学院 情報学研究科 知能情報学専攻 <sup>‡</sup> 産業技術総合研究所

本稿では、調波・非調波統合モデルのモデルパラメータ推定における楽器の奏法と個体差に対する問題点を改善する手法について述べる。我々は、CDなどの複雑な多重奏音楽音響信号中の調波構造を持つ楽器音と持たない楽器音を同時に分離するために調波・非調波統合モデルおよびそのモデルパラメータの推定手法を設計し、調波構造モデルや非調波構造モデルをそれぞれ単独で用いるよりも統合モデルを用いることで分離性能が向上することを実験で示した。本システムで実楽曲を扱うためには、楽器の多様な奏法や個体差に対処する必要がある。そのための問題点は、MIDIは奏法を扱うことがほとんど不可能な点、単一のテンプレート音を用いると個体差に対処できない点である。そこで本稿ではこれらの問題点を解決するため、楽器音認識で用いられる音色特徴量の確率分布を用いてモデルパラメータを推定する手法を提案する。音色特徴量の確率分布は多数のデータを用いて学習されるため、特定のテンプレート音の影響を除去したパラメータ推定が可能になる。モデルから抽出した音色特徴量の確率分布に対する尤度を最大化するような制約を追加することで、モデルが表現する楽器音の特徴を多く満たすようなモデルパラメータが推定される。実験によって、音色特徴量分布の尤度を考慮することで分離性能が向上することを確認した。

## Parameter Estimation for Harmonic and Inharmonic Models by Using Timbre Feature Distributions

KATSUTOSHI ITOYAMA<sup>†</sup>, MASATAKA GOTO<sup>‡</sup>,  
KAZUNORI KOMATANI<sup>†</sup>, TETSUYA OGATA<sup>†</sup> and HIROSHI G. OKUNO

<sup>†</sup> Dept. of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University

<sup>‡</sup> National Institute of Advanced Industrial Science and Technology (AIST)

This paper describes an improved parameter estimation method for an integrated weighted-mixture model consisting of both harmonic-structure and inharmonic tone models. Although we have developed a sound source separation method by using the integrated model, this method has difficulties to deal with various performance styles and individual differences of musical instruments. To solve this problem, we propose a new parameter estimation method by using probabilistic distributions of musical timbre features. Since the probabilistic distributions are trained by using various audio signals, dependency from particular template sounds decreases. By adding a new constraint of maximizing the likelihood of the probabilistic distributions of timbre features extracted from an estimated model, the model parameters can be estimated so that they can well express musical timbre features. The experimental results showed that the performance of separation improved.

### 1. はじめに

デジタルオーディオが普及し、価値観が多様化する中で、より能動的に音楽を楽しみたいというユーザの要求が現れてきた。これまでのオーディオ再生技術は、受動的な音楽の楽しみ方をより豊かにする方向に進歩し、ユーザの要求に応えてきた。例えば、5.1次元や

7.1次元などの大掛かりなシステムで忠実な音環境の再現を目指すというものや、アクティブノイズキャンセルなどの簡便な装置で静かな音環境を作ることでも手軽に音楽鑑賞を楽しむというものがある。一方、能動的な音楽の楽しみ方には作曲や編曲、演奏などがある。一般的には能動的に音楽を楽しめるのは技術や道具を持っている人に限られており、受動的な楽しみと能動的な楽しみの間には大きなギャップがあった。

能動的な音楽鑑賞<sup>1)</sup>という要求に応える研究事例として、吉井らはドラムスを対象とした楽器音イコライザ INTER:D<sup>2)</sup> および Drumix<sup>3)</sup> を実現した。ユーザは Drumix を使って楽曲中のドラムスの音量を操作し、音色を置き換え、また、ドラムパターンを編集でき、その結果能動的な音楽鑑賞がより簡便に可能となった。しかし、これらのシステムはドラムスだけを対象としており、一般の楽器音に対して適用するまでには至っていないかった。

これに対して我々の目的は、CD などによる音楽音響信号（混合音）中のあらゆる楽器パートに対して自由に音量を操作できる楽器音イコライザを実現することである。そのためには、楽曲中に含まれる各音を正しく推定する、すなわち全ての音を分離する必要がある。一般に音楽音響信号には、ピアノやフルートのような調波構造を持つ楽器音と、ドラムスのような調波構造を持たない楽器音の両方が含まれる。それゆえ、あらゆる楽器に対して適用可能な楽器音イコライザを実現するためには、調波的な音と非調波的な音の双方を同時に扱う必要があるが、従来の音源分離に関する研究の多くはこれらの2種類の音の一方のみに着目しており、両者が混在した音の分離を行うことは避けてきた。

我々は調波・非調波統合モデルを用いることであらゆる楽器音を統一的に扱える音源分離手法、及びそのモデルパラメータ推定手法<sup>4)</sup>を設計し、実現した。モデルパラメータ推定においては、非調波構造モデルに対する周波数方向への平滑化という制約を用いることでモデルの過学習問題を解決した。また、MIDI 音源から生成した混合音を用いて、統合モデルを用いることで調波構造モデルや非調波構造モデルを単独で用いるよりも分離性能が向上することを実験で確認した。

本システムで実楽曲を扱うためには、楽器音の多様な奏法および個体差に対処する必要がある。奏法や個体差に対処するための問題点は2つある。(1) MIDI では奏法に関するメッセージはほとんど規定されていないため、MIDI 音源のみで様々な奏法を持つ楽器音の特徴を全て表現することは困難である点。(2) 単一のテンプレート音だけを用いると個体差に対処できない点。この2つの問題点に対して、様々な楽器音から抽出した音色特徴量の分布を確率モデルによってあらかじめ学習しておき、楽器音が新たに与えられたときにその楽器音から抽出した音色特徴量の尤度を最大化するようなパラメータを推定する手法を提案する。音色特徴量とは、楽器音の音響信号の特徴を表すベクトルで、楽器音認識問題でよく用いられる。テンプレート音は特定の楽器個体や奏法から生成されるため、テンプレート音だけを用いるとその個体や奏法の影響を避けることができないが、複数の楽器個体や奏法を用いて学習した統計的な音色特徴量の分布を用いることで、奏法や個体差に頑健なパラメータ推定が可能になる。

以下、2章で我々が以前に開発した調波・非調波統

合モデルについて述べ、3章で本稿で問題とする、奏法・個体差に対する対処方法を述べる。4章で評価実験を行い、5章で結論を述べる。

## 2. 調波・非調波統合モデルによる音源分離

本章では、まず音源分離問題を定義する。その後、調波・非調波統合モデルおよび分離処理を定式化し、モデルパラメータの推定方法について述べる。

### 2.1 音源分離処理

本研究課題は、多重奏の音楽音響信号と同期が取られている標準 MIDI ファイル (Standard MIDI File; SMF) が与えられたとき、音響信号を SMF の各トラックに対応づけられた楽器パートごとの音響信号に分離することである。SMF の各トラックは、通常は各楽器パートに対応している。言い換えれば、我々の目標は各パートの全ての単音に対して、単音に対応する調波構造モデルと非調波構造モデルの全パラメータを推定することである。与えられた SMF の各単音を個別に MIDI 音源で演奏することで、音響信号中の各単音にある程度近い「音のサンプル」を作成できる。この音のサンプルをテンプレート音と呼ぶ。この問題設定において、解くべき問題は以下の2点である。

- (1) テンプレート音と実演奏とのずれの吸収。テンプレート音と入力信号との間には必ず音響的な違いがあるので、テンプレート音と入力信号との音響的差異を吸収する手法が必要となる。
- (2) 奏法に独立な楽器音一貫性の達成。ある楽器が同じ音高や音長をもつ単音を演奏していても、奏法やビブラートなどの違いにより、何らかの音響的な違いが存在する。そのため、モデル化は単音ごとに行う必要があるが、完全に単音ごとに独立したモデル化を行うと、同じ楽器音を表現するモデルが全く異なる音を表現してしまう可能性がある。これらの問題を、以下のアプローチで解決する。

- (1) モデルパラメータ適応。テンプレート音で初期化した音モデルのパラメータを、モデルと入力音響信号とのパワースペクトル上での音響的差異を最小化するように更新する。
- (2) 同一楽器内パラメータ一貫性に対する制約。楽器内での一貫性を保ちつつも各単音の微小な違いを許容するような制約の下でモデルパラメータの更新を行う。これは、同一楽器に属する各単音のモデルパラメータの平均値と現在着目している単音のモデルパラメータとの間の Kullback-Leibler Divergence (KLD) を最小化するような制約を加えることで達成できる。

本稿で扱う音源分離問題とは、入力された混合音のパワースペクトル  $X(c, t, f)$  を、 $k$  番目の楽器、 $l$  番目の単音のパワースペクトルに分解することである。ここで、 $c$  は左右などのチャンネル（総チャンネル数  $C$ ）、 $t$  は時刻、 $f$  は周波数を表す。本手法では、入力信号のチャン

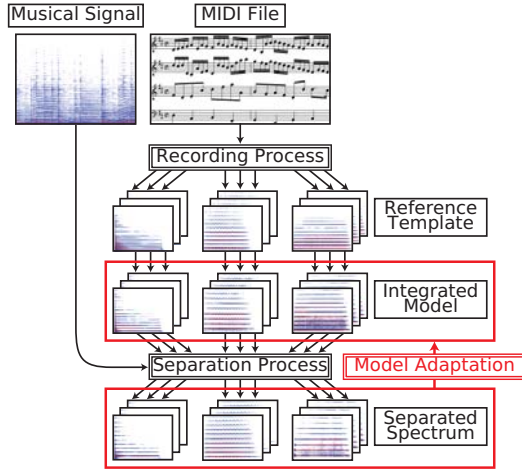


図1 分離とモデル適応の処理の流れ

ネル数や、同時に発音されている単音数に一切の制限を設けない．事前情報として与えられる SMF から、 $X(c, t, f)$  中では  $K$  種類の楽器が演奏されており、各楽器は  $L_k$  個の単音を持つものとする．ここで、 $k$  番目の楽器、 $l$  番目の単音の単音を表すモデルを  $J_{kl}(c, t, f)$  とする．SMF での定位情報は必ずしも音響信号での定位とは一致しない場合があるので、テンプレート音  $Y_{kl}(t, f)$  は  $c$  を持たず、1 チャンネルとして扱う．分離処理の流れを図1に示す．

## 2.2 調波・非調波統合モデル

調波・非調波統合モデル  $J_{kl}(c, t, f)$  は、調波構造を表現するモデル  $H_{kl}(t, f)$  と非調波構造を表現するモデル  $I_{kl}(t, f)$  に各モデルの重み  $w_{kl}^{(H)}, w_{kl}^{(I)}$  による重み付き和にモデルの音量  $w_{kl}$  およびチャンネルごとの重み  $r_{kl}(c)$  を乗じたもので、以下の式で定義する．

$$J_{kl}(c, t, f) = w_{kl} r_{kl}(c) \left( w_{kl}^{(H)} H_{kl}(t, f) + w_{kl}^{(I)} I_{kl}(t, f) \right) \quad (1)$$

$w_{kl}^{(H)}$  と  $w_{kl}^{(I)}$ 、 $r_{kl}(c)$ 、および  $w_{kl}$  は任意の  $k, l$  に対して以下の各条件を満たす．

$$w_{kl}^{(H)} + w_{kl}^{(I)} = 1, 0 \leq w_{kl}^{(H)}, w_{kl}^{(I)} \leq 1 \quad (2)$$

$$\sum_c r_{kl}(c) = C, 0 \leq r_{kl}(c) \leq C \quad (3)$$

$$\sum_{k,l} w_{kl} = \sum_c \iint X(c, t, f) dt df \quad (4)$$

### 2.2.1 調波構造モデル

調波構造モデルは、パラメトリックな基底関数であるガウス分布関数の線形和として、パワーエンベロープを表現する  $E_{kly}(t)$  と各時刻の調波構造を表現する  $F_{kln}(t, f)$  を用いて以下の式で定義する．ただし、 $M, N$  は定数で、それぞれパワーエンベロープを表現するガウシアンの数と、調波構造の倍音成分の数を表す．

表1 調波構造モデルのパラメータ

記号	意味
$w_{kl}^{(H)}$	調波構造モデル全体の音量
$\omega_{kl}(t)$	F0 の軌跡
$u_{klm}$	パワーエンベロープの概形を表現する $m$ 番目のガウシアン の重み係数 ( $\sum_m u_{klm} = 1$ を満たす)
$v_{kln}$	$n$ 次倍音成分の相対強度 ( $\sum_n v_{kln} = 1$ を満たす)
$\tau_{kl}$	発音時刻
$M\phi_{kl}$	音長 ( $M$ は定数)
$\sigma_{kl}$	倍音の周波数方向の広がりを表す標準偏差

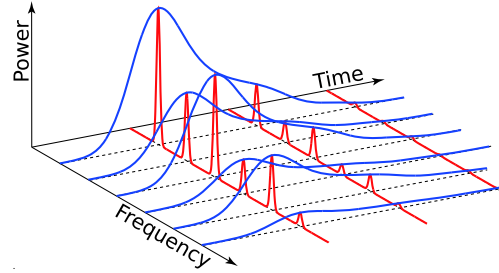


図2 調波構造モデルの概形

$$H_{kl}(t, f) = \sum_{m=0}^{M-1} \sum_{n=1}^N E_{klm}(t) F_{kln}(t, f) \quad (5)$$

$$E_{klm}(t) = \frac{u_{klm}}{\sqrt{2\pi}\phi_{kl}} e^{-\frac{(t-\tau_{kl}-m\phi_{kl})^2}{2\phi_{kl}^2}} \quad (6)$$

$$F_{kln}(t, f) = \frac{v_{kln}}{\sqrt{2\pi}\sigma_{kl}} e^{-\frac{(f-n\omega_{kl}(t))^2}{2\sigma_{kl}^2}} \quad (7)$$

モデルのパラメータを表1に、モデルの概形を図2に示す．このモデルは、亀岡らの調波時間構造化クラスタリングで用いられる音源モデル<sup>5)</sup>を参考に設計した．

### 2.2.2 非調波構造モデル

非調波構造モデルはノンパラメトリックな関数として定義され、パワースペクトルを直接表現する．このため、非調波構造モデルは任意のパワースペクトルを表現することができる．また、任意の  $k, l$  に対して以下の条件を満たす．

$$\iint I_{k,l}(t, f) dt df = 1 \quad (8)$$

理論的には、調波構造モデルのパワーエンベロープが式(6)のようにガウシアンで表現されており、時間方向に無限の長さを持つことと同様に、非調波構造モデルの時間長も無限の長さであるべきである．しかし、全ての楽器音が楽曲と同じ長さを持つことは通常の楽曲ではまず起こらないため、非調波構造モデルの時間長および発音時刻はテンプレート音と一致するように設定する．

### 2.2.3 制約条件

非調波構造モデルは任意のパワースペクトルを表現できるので、入力パワースペクトルがこのモデルだけで表現されてしまう可能性がある．しかし、調波構造モデルが表現すべき調波構造までも非調波構造モデルが表現してしまうことは望ましくない．この問題を解

決するため、以下の式で与えられる制約を導入する。

$$\gamma_{kl}^{(I1)} = \beta_{I1} \iint \bar{I}_{kl}(t, f) \log \frac{\bar{I}_{kl}(t, f)}{I_{kl}(t, f)} dt df \quad (9)$$

$\bar{I}_{kl}(t, f)$  は  $I_{kl}(t, f)$  にガウシアンフィルタを畳み込んで周波数方向に平滑化したものである。この制約は  $I_{kl}(t, f)$  を  $\bar{I}_{kl}(t, f)$  に近付ける作用を持っている。つまり、単音の非調波成分を周波数方向にピークを持たない滑らかな形状にさせ、非調波構造モデルが調波的になることを防ぐことができる。

調波構造モデルの  $\omega_{kl}(t)$  はノンパラメトリックな関数として定義しているため、各時刻でとる値に対して何も制限が加えられておらずパラメータ推定によって時間的な不連続性が生じる可能性がある。この問題を解決するため、以下の式で与えられる制約を導入する。

$$\gamma_{kl}^{(\omega)} = \beta_{\omega} \int \bar{\omega}_{kl}(t) \frac{\bar{\omega}_{kl}(t)}{\omega_{kl}(t)} dt \quad (10)$$

$\bar{\omega}_{kl}(t)$  は、 $\omega_{kl}(t)$  にガウシアンフィルタを畳み込んで時間方向に平滑化したものである。この制約を最小化することで、単音の F0 が急激に変化することを防ぐことができる。

さらに、楽器音の音色一貫性を達成するため、以下の2つの制約を追加する。

$$\gamma_{kl}^{(v)} = \beta_v \sum_n \bar{v}_{kln} \log \frac{\bar{v}_{kln}}{v_{kln}} \quad (11)$$

$$\gamma_{kl}^{(I2)} = \beta_{I2} \iint \bar{I}_k(t, f) \log \frac{\bar{I}_k(t, f)}{I_{kl}(t, f)} dt df \quad (12)$$

### 2.3 パラメータ推定

入力パワースペクトルと全てのモデルを加算合計したモデルパワースペクトル間の KLD をコスト関数  $Q$  として設定し、さらに、各モデルが対応する楽器単音を表現するための規範として、テンプレート音のパワースペクトルとモデルとの間の KLD、および式 (9) の制約をコスト関数に加えた上で、このコスト関数を最小化するパラメータを求めることでモデルパラメータ推定を行う。しかし、このままでは最適なパラメータの解析的導出ができないため、入力パワースペクトルを  $k$  番目の楽器、 $l$  番目の単音に分配する関数  $S_{kl}(c, t, f)$ 、さらに単音ごとに分配されたパワースペクトルおよびテンプレート音のパワースペクトルを調波構造を表現する  $\{y, n\}$  ラベル付きのガウス分布関数および非調波構造関数へと分配する関数  $S_{klmn}^{(H)}(t, f)$ ,  $S_{kl}^{(I)}(t, f)$  を導入する。これらの分配関数は、以下の各条件を満たす。

$$\forall c, t, f, \sum_{k,l} S_{kl}(c, t, f) = 1 \quad (13)$$

$$\forall k, l, t, f, \sum_{m,n} S_{klmn}^{(H)}(t, f) + S_{kl}^{(I)}(t, f) = 1 \quad (14)$$

$$\forall k, l, c, t, f, 0 \leq S_{kl}(c, t, f) \leq 1 \quad (15)$$

$$\forall k, l, m, n, t, f, 0 \leq S_{klmn}^{(H)}(t, f) \leq 1 \quad (16)$$

$$\forall k, l, t, f, 0 \leq S_{kl}^{(I)}(t, f) \leq 1 \quad (17)$$

これらの分配関数は、入力パワースペクトルからは観測することができないため、分配関数とモデルパラメータを同時に推定することは不可能であるが、分配関数とモデルパラメータの一方を固定し他方を推定することを交互に繰り返すことによって、コスト関数を最小化するモデルパラメータの推定および分配関数の導出が可能となる。このパラメータ推定手法は、パワースペクトルやモデルを全ての変数に関して積分した結果が 1 になるように正規化したものを観測確率密度関数および完全データ確率密度関数、テンプレート音のパワースペクトルを事前確率密度関数と見なした上での EM アルゴリズムを用いたモデルパラメータの最大事後確率 (Maximum A Posteriori; MAP) 推定と等価である。

### 3. 音色特徴量分布の利用

本章では、統合モデルのパラメータ推定において音色特徴量の分布を用いる手法を新たに提案する。

#### 3.1 問題の所在と解決へのアプローチ

前章までで述べたように、我々は調波・非調波統合モデルを用いた音源分離を実現した。これを基に、実楽曲を扱うことのできる楽器音イコライザを実現するためには楽器の奏法や個体差に対処する必要がある。つまり、コスト関数において、テンプレート音に依存せずにモデルがどの程度対象としている楽器音を表現できているかを測定する尺度が必要となる。

このような尺度として、楽器音認識などに用いられる音色特徴量の分布に対する尤度がある。楽器音認識で用いられる音色特徴量の例に、高調波成分の相対強度、パワーエンベロープの立ち上がり・減衰の速度、MFCC などがある。一般的には、様々な楽器音から抽出した音色特徴量の分布を正規分布などの確率モデルを用いてあらかじめ学習しておき、楽器音が新たに与えられたときにその楽器音から抽出した音色特徴量に対する尤度が最大となる楽器を求めることで楽器音の識別がなされる。つまり、ある楽器  $k$  から学習した音色特徴量の分布に対する新たに与えられた楽器音  $l$  の音色特徴量  $\xi_l$  に対する尤度  $p(\xi_l|k)$  は、「楽器音  $l$  がどの程度楽器  $k$  らしいか」を表している。音色特徴量の負の対数尤度をコスト関数に加えることで、パラメータを最尤なものに近付ける制約として作用する。

さらに、音色特徴量の分布を学習する際に、複数の奏法・個体によって生成された音響信号から抽出した特徴量を用いることで、特徴量分布に奏法や楽器個体に対する柔軟性を持たせることができるため、特定の奏法・個体から生成されたテンプレート音を用いる場合よりも奏法や楽器個体の変化に頑健なパラメータ推定が可能になる。

#### 3.2 音色特徴量の尤度最大化

統合モデルのパラメータ  $\theta$  ( $w, u_m, v_n$  など)、モデルパラメータから音色特徴量への写像  $\Xi: \theta \mapsto \xi$ 、およびモデルが表現する楽器の音色特徴量の分布  $p(\xi)$  が与え

$$\begin{aligned}
Q = \sum_{k,l} \left( \sum_{m,n} D_{\text{KL}} \left( S_{klmn}^{(H)}(t, f) S_{kl}(c, t, f) X(c, t, f) \parallel r_{kl}(c) w_{kl}^{(H)} E_{klm}(t) F_{kln}(t, f) \right) \right. \\
+ D_{\text{KL}} \left( S_{kl}^{(I)}(t, f) S_{kl}(c, t, f) X(c, t, f) \parallel r_{kl}(c) w_{kl}^{(I)} I_{kl}(t, f) \right) \\
+ \beta_T \left( D_{\text{KL}} \left( S_{klmn}^{(H)}(t, f) Y_{kl}(t, f) \parallel w_{kl}^{(H)} E_{klm}(t) F_{kln}(t, f) \right) + D_{\text{KL}} \left( S_{kl}^{(I)}(t, f) Y_{kl}(t, f) \parallel w_{kl}^{(I)} I_{kl}(t, f) \right) \right) \\
+ \gamma_{kl}^{(\omega)} + \gamma_{kl}^{(v)} + \gamma_{kl}^{(I1)} + \gamma_{kl}^{(I2)} + \frac{1}{2} \beta_p (\boldsymbol{\xi}_{kl} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\boldsymbol{\xi}_{kl} - \boldsymbol{\mu}_k) \\
+ \left( \text{Lagrange multipliers of } w_{kl}, w_{kl}^{(H)}, w_{kl}^{(I)}, r_{kl}(c), u_{klm}, v_{kln}, I_{kl}(t, f), S_{kl}(c, t, f), S_{klmn}^{(H)}(t, f), S_{kl}^{(I)}(t, f) \right)
\end{aligned} \quad (18)$$

られたとき、その対数尤度  $\log p(\boldsymbol{\xi})$  を最大化する  $\theta$  を求める問題を考える。ここでは特徴量の分布は平均ベクトル  $\boldsymbol{\mu}$ 、共分散行列  $\Sigma$  で定められる正規分布であるとする。ただし、特徴量の性質によってはディリクレ分布なども利用可能である。正規分布の確率密度関数は単峰的であるため、それぞれの特徴量の最尤推定値を求めることは、特徴量の対数尤度の微分係数が0になる点を求めることと等しい。

$k$  番目の楽器、 $l$  番目の単音のパラメータ  $\theta_{kl} = (\theta_{kl1}, \dots, \theta_{klN_\theta})$  が与えられたとき、特徴量分布に対する尤度を最大化する  $i$  番目のパラメータ  $\theta_{kli}$  を求めるためには、次式の  $\theta_{kli}$  に関する零点を求めればよい。

$$\frac{\partial \log p(\boldsymbol{\xi}_k)}{\partial \theta_{kli}} = -\frac{1}{2} \sum_j (\xi_{klj} - \mu_{kj}) (\Sigma_k^{-1})_{ij} \frac{\partial \xi_{klj}}{\partial \theta_{kli}} \quad (19)$$

ここで問題となるのが、パラメータから音色特徴量への写像  $\xi$  がどのように構成されているかである。方程式の中に導関数  $\partial \xi_j / \partial \theta_i$  が現れており、さらに方程式が必ず解を持ち、解が解析的に求められるためには、パラメータは特徴量の線形変換で求められることが強く求められる。しかし一般には特徴量を線形変換のみで求めることは困難であるため、そのような場合には特徴量とパラメータの関係を線形回帰で学習しておく等の方法をとることで、近似的に尤度を最大化することが可能になる。

前節で挙げた特徴量の例で考えると、高調波成分の相対強度は、調波構造モデルの  $v_n$  そのものであるので、 $\partial \xi_j / \partial \theta_i = \partial v_i / \partial v_i = 1$  となる。立ち上がり・減衰を包含したパワーエンベロープの変化は、同様に調波構造モデルの  $u_m$  で表現されているので、 $\partial \xi_l / \partial \theta_i = \partial u_i / \partial u_i = 1$  となる。MFCC は、三角窓によるメル周波数フィルタバンクをかけ、対数を取り、離散コサイン変換を施して得られるため、非線形変換が含まれる。そのため、入力（パワースペクトル）と出力（MFCC）との関係を線形回帰し、パラメータ推定にはその回帰直線を用いることになる。

### 3.3 コスト関数

式 (18) に、特徴量分布とテンプレートとの距離の両方を考慮した場合のコスト関数を示す。特徴量分布を用いない場合は  $\beta_p = 0$  と、テンプレートをを用いない場合は  $\beta_T = 0$  とすることで、それぞれの目的に応じたコスト関数となる。ここで、 $D_{\text{KL}}(A(x) \parallel B(x))$  は  $A(x)$  と  $B(x)$  との KLD であり、

$$D_{\text{KL}}(A(x) \parallel B(x)) = \int A(x) \log \frac{A(x)}{B(x)} dx \quad (20)$$

で定義される。また、式 (18) 中の Lagrange multipliers とは、それぞれのパラメータや分配関数に対する拘束条件を表現するためのラグランジュの未定乗数項を表している。各パラメータに対する拘束条件は、式 (2), (3), (4), (8), (13), (14), および表 1 に記されている。紙面の都合上、詳細な式は省略する。このコスト関数をモデルパラメータや分配関数で微分し、(導関数) = 0 という方程式を解くことで、対応するパラメータなどの更新式を導出できる。こちらも紙面の都合上、詳細な式は省略する。

## 4. 評価実験

本手法の性能を確認するため、評価実験を行った。

### 4.1 実験の目的

本実験の目的は、音色特徴量の分布をパラメータ推定に用いることで、分離性能がどのように変化するかを確認することである。以下の 2 条件で分離処理を行い、結果を比較した。

- 特徴量分布上での尤度を考慮した場合 (PDF)。
- テンプレート音のみを考慮した場合 (Template)。

### 4.2 実験方法・実験データ

基本的な実験方法および実験データは、我々の以前の研究報告<sup>4)</sup> に準ずる。すなわち、

- テンプレート音でモデルを初期化する。
- モデルを入力信号に適應させる。この際、実験条件 (a) では  $\beta_T = 0$  として音色特徴量分布に対するモデルパラメータの負の対数尤度を加えたコスト関数を、実験条件 (b) では  $\beta_p = 0$  としてテンプレートスペクトルとモデルとの KLD を加えたコスト関数を用いる。
- 分離信号とミックス前の信号との SNR を求める。のように実験を行った。実験データには、RWC 研究用音楽データベース：ポピュラー音楽 (RWC-MDB-P-2001)<sup>6)</sup> から選んだ 10 曲 (No. 1–10) の開始から 30 秒の区間を利用した。入力音響信号は YAMAHA MU2000、テンプレート音は Roland SD-90 で生成し、入力音響信号とテンプレート音の間に確実に音響的な差異が含まれるようにした。

特徴量には最も単純なものであるモデルパラメータの部分集合の 40 次元ベクトル  $(u_{kl1}, \dots, u_{klM}, v_{kl1},$

表 2 実験条件

項目	値
# of kernels in $E_{klm}$ : M	10
# of partials: N	30
$\beta_v$	0.1
$\beta_\omega$	0.1
$\beta_{I1}$	0.5
$\beta_{I2}$	3.5
$\beta_T$	0 (PDF), 0.5 (Template)
$\beta_p$	$1.0 \times 10^6$ (PDF), 0 (Template)

PDF: 音色特徴量分布上での尤度を用いた場合  
 Template: テンプレート音との距離を用いた場合

表 3 実験結果

楽器パート	PDF	Template
ピアノ	-5.97 dB	-5.76 dB
ベースギター	-2.70 dB	-2.82 dB
フルート	-3.25 dB	-3.38 dB
アコースティックギター	-4.34 dB	-4.53 dB
エレキギター	-4.40 dB	-4.72 dB
ストリングス	-3.05 dB	-3.16 dB

PDF: 音色特徴量分布上での尤度を用いた場合  
 Template: テンプレート音との距離を用いた場合

...,  $v_{klN}$ ) を用いた。特徴量とパラメータが同一なので、 $\partial \xi_j / \partial \theta_i = 1$  とすることで特徴量分布を用いたパラメータ推定が可能になる。 $u_m$  は調波構造モデルにおけるパワーエンベロープ、 $v_n$  は同モデルにおける高調波成分の相対強度を表しているため、各楽器の調波構造の形状を学習している。

なお、音色特徴量学習の際にテンプレート音が影響を及ぼすことを確実に避けるために、テンプレート音として録音した音を、テンプレートとしてそのまま使うものと特徴量学習に使うものに分割すべきである。そこで、音色特徴量分布を学習するためのデータには、分離対象となる楽曲を除く 9 曲のテンプレート音で初期化したモデルから抽出した特徴量を用いた。これはクロスバリデーションの考え方に似ているが、評価対象のデータセット 10 曲分は、これら音色特徴量学習用のテンプレート音とはさらに別に用意してある(データセット全体を変えてより厳しく評価している)ため、一般的によく行われる評価のためのクロスバリデーションとは異なる。

コスト関数における各パラメータは表 2 のように設定した。 $\beta_p$  が他のパラメータと比べて非常に大きい値をとっているが、これは特徴量分布上の尤度とテンプレートとの距離がコスト関数に与える影響を同程度にするためである。

#### 4.3 実験結果

表 3 に、各楽曲の楽器パートごとの SNR を求め、楽器パートごとに平均した結果を示す。ピアノ以外の楽器パートでは、特徴量分布を用いた方が総じて SNR が大きくなっていることが分かる。つまり多くの楽器においては、特徴量分布を用いることでテンプレート音を使うよりも分離性能が向上することを示している。

一方、ピアノパートでは特徴量分布を用いることで分離性能が低下している。ピアノは音高の変化によ

て音色が大きく変化する<sup>7)</sup>ことが知られているが、本実験で用いた特徴量分布では、全ての音高で共通の分布を用いている。このため、音高ごとに異なる音色を表現できるテンプレート音を用いた場合に SNR が大きくなったと考えられる。このような楽器音に対してパラメータ推定を行う場合には、F0 依存多次元正規分布<sup>8)</sup>などの F0 の変化を考慮した分布を用いることで、適切なパラメータが推定されると考えられる。

本実験では、学習データに複数の奏法や個体差が含まれていないため、奏法や個体差の対する頑健性を真に確かめるには至っていない。しかし、入力信号と学習データ・テンプレート音に個体差がある状況では、テンプレート音を用いるよりもよい分離性能を示しているため、本手法の潜在的な能力を確認したといえる。

#### 5. おわりに

本稿では、音色特徴量分布を利用した調波・非調波統合モデルのパラメータ推定手法について述べた。実験によって、多くの楽器パートに対しては比較的簡単な特徴量分布を用いるだけでも分離性能が向上することから、本手法の潜在的な能力を示した。今後は、複数の奏法や楽器個体を含む学習データを用いて、奏法や個体差に対する頑健性を評価する。

また今回は、特徴量として高調波成分の相対強度と、パワーエンベロープの立ち上がり・減衰の速度を扱ったが、今後は本研究の結果を応用して MFCC 等のより多くの特徴量を扱い、分布を GMM などで表すことで複数の奏法を持つ楽器音の分離に取り組んでいきたい。

謝辞 本研究の一部は、科学研究費補助金(基盤研究(S), 特定領域「情報爆発 IT 基盤」)、科学技術振興機構 CrestMuse プロジェクトによる支援を受けた。

#### 参考文献

- Goto, M.: Active Music Listening Interfaces Based on Signal Processing, *ICASSP*, Vol.IV, pp.1441-1444 (2007).
- Yoshii, K., Goto, M. and Okuno, H.G.: INTER:D: A Drum Sound Equalizer for Controlling Volume and Timbre of Drums, *EWIMT*, pp.205-212 (2005).
- Yoshii, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H.G.: Drumix: An Audio Player with Real-time Drum-part Rearrangement Functions for Active Music Listening, *IPSS Journal*, Vol.48, No.3, pp.134-144 (2007).
- 糸山克寿, 後藤真孝, 駒谷和範, 尾形哲也, 奥乃博: 多重奏音楽音響信号の音源分離のための 調波・非調波モデルの制約付きパラメータ推定, 情処研報, 2007-MUS-70, pp.81-88 (2007).
- Kameoka, H., Nishimoto, T. and Sagayama, S.: Harmonic-temporal Structured Clustering via Deterministic Annealing EM Algorithm for Audio Feature Extraction, *ISMIR*, pp.115-122 (2005).
- Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC Music Database: Popular, Classical, and Jazz Music Databases, *ISMIR*, pp.287-288 (2002).
- 安藤由典: 楽器の音響学, 音楽之友社 (1996).
- 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, 情処論, Vol.44, No.10, pp.2448-2548 (2003).