社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2007－MUS－70 (13)
2007－EC－7 (13)
2007／5／11

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

糸山 克寿† 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 知能情報学専攻 ‡産業技術総合研究所

本稿では，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のモデルの作成と，楽譜情報を事前情報として与えた場合の制約付きモデルパラメータ推定手法について述べる．調波構造の有無によって楽器音の性質は大きく異なるため，従来の手法ではこれらの音を排他的に扱うことしかできなかった．本稿では，調波構造と非調波構造のそれぞれを表現する2つのモデルを統合した新たな重み付き混合モデルにより，両者の統合的手法を開発した．モデルのパラメータは最大事後確率推定に基づくEMアルゴリズムを用いて推定する．さらに，モデルの過学習を防ぎ同一楽器内のパラメータ一貫性を維持するための制約条件も同時に用いる．ポピュラー音楽のSMFを用いた評価実験で，本手法によりSNRが1.5 dB向上することを確認した．

Constrained Parameter Estimation of Harmonic and Inharmonic Models for Separating Polyphonic Musical Audio Signals

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

† Dept. of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University
‡ National Institute of Advanced Industrial Science and Technology (AIST)

This paper describes a sound source separation method for polyphonic sound mixtures of music including both harmonic and inharmonic sounds, and constrained parameter estimation using standard MIDI files as prior information. The difficulties in dealing with both types of sound together have not been addressed in most previous methods that have focused on either of the two types separately, because the properties of these sounds are quite different. We therefore developed an integrated weighted-mixture model consisting of both harmonic-structure and inharmonic tone models. On the basis of the MAP estimation using the EM algorithm, we estimated all model parameters of this integrated model under several original constraints for preventing over-training and maintaining intra-instrument consistency. We confirmed that the integrated model increased the SNR by 1.5 dB.

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どこ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．

より能動的な音楽鑑賞というユーザの要求の一つに，楽器パートの音量を自由に操作したいというものが挙げられる．この目的のために従来使用されてきたグラフィックイコライザは周波数特性を変化させ，残響コントロールなどと組み合わせて音響環境の変化を実現するための技術であるため，本質的に楽器音イコライザや特定パートイコライザを実現するには不十分である．

能動的な音楽鑑賞[1]という要求に応える技術として，吉井らはINTER:D[2]およびDrumix[3]を開発している．ユーザはDrumixを使って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可能となった．しかし，Drumix は楽曲中の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，楽器音イコライザと呼ばれるシステムを作成することである．そのためには，楽曲中に含まれる個々の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．これらの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．また，ピアノが発音時にハンマーで弦を叩く音のように，調波構造を持つ楽器音であっても，楽器の物理的な構造に由来する調波的でない成分を含むものが多い．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扱う必要がある．

しかし，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，同時の分離を行うことは想定していなかった．例えば，調波的な音を分離する研究には 4)~8) などが，非調波的な音を分離する研究には 9)~11) などがある．後藤[12] は，調波的な音を表現するモデルと非調波的な音を表現するモデルを統合する理論的な手法について言及しているが，評価はしていなかった．また，調波的な音と非調波的な音を同時に扱う他の手法として，独立成分分析 (Independent Component Analysis; ICA) などに基づくブラインド音源分離があるが，CD のような複雑な音響信号を扱うには至っていない．

我々は後藤[12] の示唆を受け，調波構造モデルと非調波構造モデルを統合した混合モデルを用いた音源分離手法を開発した．調波構造モデルは，音高を持つ楽器の単音の調波構造を表現するパラメトリックモデルに基づいており，音量，F0 の時間変化，オンセット，音長，各倍音成分の相対強度，パワーエンベロープの時間変化といったパラメータで表現される．非調波構造モデルは，ノンパラメトリックモデルに基づいており，調波構造では表現が難しいドラム音などのパワースペクトルをそのまま表現する．また前述のように，ピアノやギターなどの調波構造をもつ楽器音であっても，発音時には弦をハンマーで叩く音や弦を弾く音など，非調波成分を含んでいるので，このような音も非調波構造モデルが表現する．

モデルのパラメータの推定には，最大事後確率 (Maximum *A Posteriori*) 推定に基づく EM アルゴリズムを用いる．非調波構造モデルは大きな自由度を持っておりあらゆるパワースペクトルを表現できるため，推定の結果，非調波構造モデルが全ての混合音を表現してしまう問題がある．この問題を解決するため，事前情報として音響信号に同期した MIDI ファイル (Standard MIDI File; SMF) ☆を用い，さらに同一楽器に対する制約や非調波成分に対する制約を用いる．このようにして得られた調波・非調波併用モデルを用いることで，パワースペクトルの分離が可能となる．

2. 問題の所在と解決へのアプローチ

多重奏音楽音響信号と同期が取られている SMF が与えられたとき，我々の目標は音響信号を SMF の各トラックに対応づけられた楽器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 を 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．この音をテンプレート音と呼ぶ．SMF とテンプレート音が与えられたとき，我々が解くべき課題は以下の 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，これをそのまま分離に使うことはできず，何らかの方法でテンプレート音と入力信号の違いを吸収する必要がある．

(2) **奏法に独立な楽器音一貫性の達成．** ある楽器が楽譜上では同じ F0 や音長で演奏されていても，奏法やビブラートなどの違いにより，音響信号上には違いがあるため，単音ごとに何らかのモデル化をする必要がある．しかし，これらの音を他の楽器音と比べると，同じ楽器の音には何らかの一貫性があるため，完全に単音ごとのモデル化をするとこのような性質を表現できない．

これらの詳題に対して，以下のアプローチを取る．

(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．これは，モデル適応ともとらえることができる．

(2) **同一楽器内パラメーター貫性に対する制約．** 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 (KL) ダイバージェンスを最小化するよう

☆ 本稿では，SMF と音響信号とは何らかの従来手法[13)~17)] を用いて同期がとられていると仮定する．

な制約を加えることで達成できる．

3. 定 式 化

問題は，分離すべき音響信号のパワースペクトル $g^{(O)}(c,t,f)$（以下単に $g^{(O)}$ と記す）を，単音ごと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，t は時刻，f は周波数を表す．本手法では，入力信号のチャンネル数や，同時刻に発音されている単音数に一切の制限を定めない．$g^{(O)}$ 中では K 個の楽器が演奏されており，各楽器は L_k 個の単音を持つとする．ここで，k 番目の楽器，l 番目の単音のテンプレート音のパワースペクトルを $g_{kl}^{(T)}(t,f)$ とし，対応する単音を表すモデルを $h_{kl}(c,t,f)$ とする（同じくそれぞれを $g_{kl}^{(T)}, h_{kl}$ と記す）．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，$g_{kl}^{(T)}$ は 1 チャンネルとなっている．

3.1 分 離 処 理

観測パワースペクトル $g^{(O)}$ を，各モデル h_{kl} に基づいてそれぞれのモデルが表す単音に分解するために，パワースペクトル分配関数 $m_{kl}(c,t,f)$（以下単に m_{kl} と記す）を導入する．この関数は，あらゆる (c,t,f) において，k 番目の楽器，l 番目の単音が $g^{(O)}$ に対して占める割合を表す．すなわち，$g^{(O)}m_{kl}$ は k 番目の楽器，l 番目の単音の分離されたパワースペクトルを表す．パワースペクトルの加算性が成り立つことを仮定すると，m_{kl} はあらゆる (c,t,f) において

$$0 \le m_{kl} \le 1, \quad \sum_{k,l} m_{kl} = 1 \tag{1}$$

を満たせばよい．

ここで，この分離の良し悪し $J_1(k,l)$ は，次式で示した $g^{(O)}m_{kl}$ とモデル h_{kl} との間の KL ダイバージェンスで定義する．

$$J_1(k,l) = \sum_c \iint g^{(O)} m_{kl} \log \frac{g^{(O)} m_{kl}}{h_{kl}} \, dt \, df \tag{2}$$

また，推定されたモデルの良し悪し $J_2(k,l)$ は，次式で示したテンプレート $g_{kl}^{(T)}$ とモデル h_{kl} との間の KL ダイバージェンスで定義する．

$$J_2(k,l) = \sum_c \iint g_{kl}^{(T)} \log \frac{g_{kl}^{(T)}}{h_{kl}} \, dt \, df \tag{3}$$

さらに，全ての楽器，全ての単音についての分離とモデル推定を統合した全体での良し悪しは，これらの KL ダイバージェンスをあらゆる k,l について足し合わせた次式で表す．

$$\sum_{k,l} (\alpha J_1(k,l) + (1-\alpha) J_2(k,l)) \tag{4}$$

ここで，$\alpha\ (0 \le \alpha \le 1)$ は，分離とモデル推定のどちらを重視するかを表す重みパラメータである．α を最初は 0 に設定し（最初はモデルをテンプレートに近くなるように推定する），徐々に 1 に近づけていく（徐々に分離されたパワースペクトルに近づける）ことで，モデル適応と分離の繰り返しにおいてモデルの過学習を防ぐことができると考えられる．

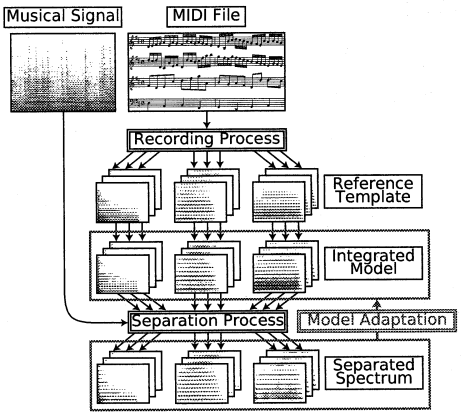

図 1 分離とモデル適応の処理の流れ

図 1 に全体の処理の流れを示す．分離とモデル適応の繰り返しは，m_{kl} の推定と h_{kl} の更新を，交互に一方を固定して他方を計算することでなされる．ここで，$\lambda(c,t,f)$ をラグランジュの未定乗数項として式 (4) に式 (1) の制約を加えると，最小化すべきコスト関数 J_0 は

$$\begin{aligned} J_0 = {} & \alpha \sum_{k,l,c} \iint g^{(O)} m_{kl} \log \frac{g^{(O)} m_{kl}}{h_{kl}} \, dt \, df \\ & + (1-\alpha) \sum_{k,l,c} \iint g_{kl}^{(T)} \log \frac{g_{kl}^{(T)}}{h_{kl}} \, dt \, df \\ & - \sum_c \iint \lambda(c,t,f) \left(\sum_{k,l} m_{kl} - 1 \right) dt \, df \end{aligned} \tag{5}$$

と表される．

まず，分離を行うために，h_{kl} を固定して J_0 が最小化されるような m_{kl} を求める．J_0 を偏微分すると，

$$\begin{cases} \dfrac{\partial J_0}{\partial m_{kl}} = \alpha g^{(O)} \log \dfrac{g^{(O)} m_{kl}}{h_{kl}} - \lambda(c,t,f) \\ \dfrac{\partial J_0}{\partial \lambda(c,t,f)} = \displaystyle\sum_{k,l} m_{kl} - 1 \end{cases} \tag{6}$$

となる．これらを用いて，連立方程式

$$\frac{\partial J_0}{\partial m_{kl}} = 0, \quad \frac{\partial J_0}{\partial \lambda(c,t,f)} = 0 \tag{7}$$

を解くと，

$$m_{kl} = \frac{h_{kl}}{\sum_{k,l} h_{kl}} \tag{8}$$

を得る．

表 1　確率密度関数とパワースペクトルとの対応関係

確率密度関数	意味	パワースペクトル
$p(c,t,f)$	観測確率密度	$g^{(O)}$
$p(k,l,t,f)$	事前確率密度	$g_{kl}^{(T)}$
$p(k,l,c,t,f\|\theta)$	完全データ	h_{kl}
$p(k,l\|c,t,f,\theta)$	不完全データ	m_{kl}

ただしパワースペクトルに関しては，個々の関数を全ての変数に対して積分した結果が 1 になるように正規化すると、上記のように対応付けられる。

次に，モデル適応を行うために，m_{kl} を固定して J_0 が最小化されるような h_{kl} を求める．これはモデルの定義と密接に関連しているため，第 4 章，第 5 章で詳述する．

3.2 EM アルゴリズムとの関連

上記の分離とモデル適応の繰り返し，すなわち m_{kl} と h_{kl} の最適化は，MAP 推定に基づく EM アルゴリズムとも解釈できる．式 (9) で定義される Q 関数を考えると，このことはより明確になる．

$$Q(\theta,\tilde{\theta}) = \alpha \sum_{k,l,c} \iint p(k,l|c,t,f,\theta)p(c,t,f) \log p(k,l,c,t,f|\tilde{\theta})dt\,df + (1-\alpha)\sum_{k,l,c}\iint p(k,l,t,f) \log p(k,l,c,t,f|\tilde{\theta})dt\,df \quad (9)$$

Q 関数がコスト関数 J_0 に対応し，さらに表 1 に示すようにそれぞれの確率密度関数は 3.1 節で述べた関数に対応する．

$$p(k,l|c,t,f,\theta) = \frac{p(k,l,c,t,f|\theta)}{\sum_{k,l} p(k,l,c,t,f|\theta)} \quad (10)$$

であることを考えると，式 (8) での分配関数の導出は確率密度関数上においても妥当であることが分かる．この式からも分かるように，$p(k,l|c,t,f,\theta)$（すなわち m_{kl}）の導出は，完全データの尤度の条件付き期待値を計算していることに相当するので，EM アルゴリズムの E (Expectation) ステップに相当する．また，θ（すなわち h_{kl}）の更新は，Q 関数を θ に関して最大化することに相当するので，M (Maximization) ステップに相当する．

4. 調波・非調波併用モデル

モデル h_{kl} 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 $I_{kl}(t,f)$ の線形和であり，(それぞれ以下単に H_{kl}, I_{kl} と記す）以下のように定義される．

$$h_{kl}(c,t,f) = r_{klc}\left(H_{kl}(t,f) + I_{kl}(t,f)\right) \quad (11)$$

r_{klc} は各チャンネルの相対的な強度を表すパラメータであり，以下の条件を満たす．

表 2　調波構造モデルのパラメータ

記号	意味
w_{kl}	調波構造モデル全体の音量
$\mu_{kl}(t)$	F0 の軌跡
u_{kly}	パワーエンベロープの概形を表現する y 番目のガウシアンの重み係数 ($\sum_y u_{kly} = 1$ 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)
τ_{kl}	オンセット時刻
$Y\phi_{kl}$	音長（Y は定数）
σ_{kl}	倍音の周波数方向の広がりを表す標準偏差

$$\sum_c r_{klc} = 1 \quad (12)$$

4.1 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 $F_{kln}(t,f)$（それぞれ以下 E_{kly}, F_{kln} と記す）を用いて以下のように定義する．ただし，Y, 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．

$$H_{kl} = \sum_{y=0}^{Y-1}\sum_{n=1}^{N} w_{kl}E_{kly}F_{kln} \quad (13)$$

$$E_{kly} = \frac{u_{kly}}{\sqrt{2\pi}\phi_{kl}} e^{-\frac{(t-\tau_{kl}-y\phi_{kl})^2}{2\phi_{kl}^2}} \quad (14)$$

$$F_{kln} = \frac{v_{kln}}{\sqrt{2\pi}\sigma_{kl}} e^{-\frac{(f-n\mu_{kl}(t))^2}{2\sigma_{kl}^2}} \quad (15)$$

モデルのパラメータを表 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 (Harmonic-Temporal-structured Clustering; HTC) で用いられる音源モデル[6] を参考に設計した．

亀岡らの HTC 音源モデルでは，$\mu_{kl}(t)$ は時間 t に関する多項式として定義されていたのに対して，我々はより柔軟な音高の時間変化を扱うために，ノンパラメトリックな関数として $\mu_{kl}(t)$ を定義した．しかし，この定義では各時刻でとる値に対して何も制限が加えられていないので，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新たな制約を導入する．

$$\beta_\mu \int \left(\bar{\mu}_{kl}(t)\log\frac{\bar{\mu}_{kl}(t)}{\mu_{kl}(t)} - (\bar{\mu}_{kl}(t) - \mu_{kl}(t))\right)dt \quad (16)$$

ただし $\bar{\mu}_{kl}(t)$ は，$\mu_{kl}(t)$ 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．積分中の第一項は一般的に用いられる KL ダイバージェンスで，これを最小化することで $\mu_{kl}(t)$ を $\bar{\mu}_{kl}(t)$ に近付ける作用を持つ．第二項の括弧で囲まれた部分は，$\bar{\mu}_{kl}(t)$ を積分した値と $\mu_{kl}(t)$ を積分した値を近付ける作用を持つ．つまり，この制約を用いることで，単音の F0 が急

激に変化することを防ぐことができる．

4.2 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で定義され，パワースペクトルを直接表現する．第1章で述べたように，この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新たな制約を導入する．

$$\beta_{I2}\iint\left(\bar{I}_{kl}\log\frac{\bar{I}_{kl}}{I_{kl}}-(\bar{I}_{kl}-I_{kl})\right)dt\,df \qquad (17)$$

ここで，$\bar{I}_{kl}$ は I_{kl} 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．この式の各項は式(16)と同様に設計している．この制約は I_{kl} を $\bar{I}_{kl}$ 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．

4.3 同一楽器内での制約

第2章で述べたように，調波・非調波併用モデル h_{kl} の推定されたパラメータは，同一の楽器内では類似しており，かつ各単音ごとに少しずつ異なっているという，同一楽器内での一貫性を満たしている必要がある．この性質を満たすようにパラメータ推定を行うために，新たに2つの制約を導入する．

第1の制約は，調波構造モデルに対する制約で，以下の式で与えられる．

$$\beta_v\sum_n\left(\bar{v}_{kn}\log\frac{\bar{v}_{kn}}{v_{kln}}-(\bar{v}_{kn}-v_{kln})\right) \qquad (18)$$

$\bar{v}_{kn}$ は，v_{kln} の同一楽器内での平均をとったものである．この式の各項は式(16)と同様に設計している．この式を最小化することで，v_{kln} を $\bar{v}_{kn}$ に近付けることができる．つまりこの制約は，同一楽器の単音に対して倍音成分の相対強度を類似させる作用を持っている．

第2の制約は，非調波構造モデルに対する制約で，以下の式で与えられる．

$$\beta_{I1}\iint\left(\bar{I}_{k}\log\frac{\bar{I}_{k}}{I_{kl}}-(\bar{I}_{k}-I_{kl})\right)dt\,df \qquad (19)$$

$\bar{I}_k$ は，I_{kl} の同一楽器内での平均をとったものである．この式の各項は式(16)と同様に設計している．この式を最小化することで，I_{kl} を $\bar{I}_k$ に近付けることができる．つまりこの制約は，同一楽器の単音に対して，非調波成分を類似させる作用を持っている．

5. モデル適応

第4章でモデルを定義したので，第3章で述べたように，m_{kl} を固定して h_{kl} を最適化することで，コスト関数 J を最小化することができる．ここで，J は全ての単音に対するコストである．

観測パワースペクトル全体のモデルは各単音の線形和で表現され，個々のモデルは調波構造モデルと非調波構造モデルの線形和で表現され，さらに調波構造モデルは基底関数であるガウス分布関数の線形和で表現されている．これらの事から，観測パワースペクトル全体を各単音の個々のガウス分布関数と非調波構造モデルに分解できればモデルパラメータを解析的に最適化することが可能になる．

そこで，新たな2つのパワースペクトル分配関数 $m^{(H)}_{klyn}(t,f), m^{(I)}_{kl}(t,f)$ を導入する．それぞれの関数は，k 番目の楽器，l 番目の単音の分離パワースペクトル $g^{(O)}m_{kl}$ を $\{y,n\}$ ラベル付きのガウス分布関数および非調波構造モデルに分配する関数であり，

$$\begin{cases}\displaystyle\sum_{y,n}m^{(H)}_{klyn}(t,f)+m^{(I)}_{kl}(t,f)=1\\ 0\le m^{(H)}_{klyn}(t,f)\le 1\\ 0\le m^{(I)}_{kl}(t,f)\le 1\end{cases} \qquad (20)$$

を満たす．モデル h_{kl} を固定した状態で，J を最小化するこれらの分配関数を導出すると，

$$\begin{cases}m^{(H)}_{klyn}=\dfrac{w_{kl}E_{kly}F_{kln}}{H_{kl}+I_{kl}}\\ m^{(I)}_{kl}=\dfrac{I_{kl}}{H_{kl}+I_{kl}}\end{cases} \qquad (21)$$

となる．詳細は省略するが，m_{kl} と同様の導出過程で求めることができる．

$\lambda^{(r)}_{kl},\lambda^{(u)}_{kl},\lambda^{(v)}_{kl}$ をそれぞれ r_{klc},u_{kly},v_{kln} に対するラグランジュの未定乗数項として，

$$\begin{cases}G_{kl}(c,t,f)=\alpha g^{(O)}m_{kl}+(1-\alpha)g^{(T)}_{kl}\\ G^{(H)}_{klyn}(c,t,f)=m^{(H)}_{klyn}G_{kl}(c,t,f)\\ G^{(I)}_{kl}(c,t,f)=m^{(I)}_{kl}G_{kl}(c,t,f)\end{cases} \qquad (22)$$

とすると，各単音のモデル h_{kl} の各パラメータに対する更新式は式(23)のコスト関数から求めることができる．ただし $G^{(H)}_{klyn}(c,t,f), G^{(I)}_{kl}(c,t,f)$ は，以降ではそれぞれ $G^{(H)}_{klyn}, G^{(I)}_{kl}$ と記す．具体的な更新式の導出は付録に記した．

6. 評価実験

提案手法の性能を確認するため，評価実験を行った．

6.1 実験の目的

本実験の目的は，本稿で構築した調波・非調波併用モデルの有効性を確認することである．具体的には，以下の3つの条件で分離を行い，ミックス前の信号とのSNRを比較した．

$$
\begin{aligned}
J = \sum_{k,l} \Bigg(& \sum_{c,y,n} \iint \left(G^{(H)}_{klyn} \log \frac{G^{(H)}_{klyn}}{r_{klc} w_{kl} E_{kly} F_{kln}} - G^{(H)}_{klyn} + r_{klc} w_{kl} E_{kly} F_{kln} \right) dt\, df \\
& + \sum_{c} \iint \left(G^{(I)}_{kl} \log \frac{G^{(I)}_{kl}}{r_{klc} I_{kl}} - G^{(I)}_{kl} + r_{klc} I_{kl} \right) dt\, df \\
& + \beta_v \sum_{n} \left(\bar{v}_{kn} \log \frac{\bar{v}_{kn}}{v_{kln}} - \bar{v}_{kn} + v_{kln} \right) + \beta_\mu \int \left(\bar{\mu}_{kl}(t) \log \frac{\bar{\mu}_{kl}(t)}{\mu_{kl}(t)} - \bar{\mu}_{kl}(t) + \mu_{kl}(t) \right) dt \\
& + \beta_{I1} \iint \left(\bar{I}_k \log \frac{\bar{I}_k}{I_{kl}} - \bar{I}_k + I_{kl} \right) dt\, df + \beta_{I2} \iint \left(\bar{I}_{kl} \log \frac{\bar{I}_{kl}}{I_{kl}} - \bar{I}_{kl} + I_{kl} \right) dt\, df \\
& - \lambda^{(r)}_{kl} \left(\sum_{c} r_{klc} - 1 \right) - \lambda^{(u)}_{kl} \left(\sum_{y} u_{kly} - 1 \right) - \lambda^{(v)}_{kl} \left(\sum_{n} v_{kln} - 1 \right) \Bigg)
\end{aligned}
\tag{23}
$$

表 3 実験条件

Frequency analysis	
sampling rate	44.1 kHz
STFT window	2048 points Gaussian
Parameters	
# of partials: N	20
# of kernels in E_{kly}: Y	10
β_v	0.1
β_μ	0.1
β_{I1}	3.5
β_{I2}	0.5
MIDI sound generator	
test data	Yamaha MU2000
template sounds	Roland SD-90

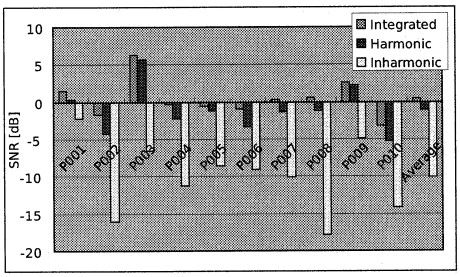

図 2 実験結果

(1) 調波・非調波併用モデルを用いた場合（本手法）
(2) 調波構造モデルのみを用いた場合
(3) 非調波構造モデルのみを用いた場合

本来，調波構造モデルだけではドラム音を表現することは非常に難しいが，本実験では用いたモデル以外の条件を同一にするために，調波構造モデルだけを用いた場合でもドラムパートを含む混合音の分離を行った．また，非調波構造モデルだけを用いた場合は，非調波構造モデルを平滑化する制約を用いるとモデルが調波構造を表現することができなくなるため，この制約は用いていない．

6.2 実験データ

実験には，RWC 研究用音楽データベース：ポピュラー音楽 (RWC-MDB-P-2001)[18] から選んだ 10 曲 (Nos. 1–10) を用いた．各楽曲は開始から 30 秒の区間を利用した．我々の音源分離手法は CD のような複雑な楽曲を扱うことを想定しているが，本実験では以下の理由により，MIDI 音源から録音した音響信号を分離対象とした．

- 本実験の目的に適した音楽データベースがない．目的に適したデータベースとは，楽曲の音響信号とそれに同期がとられた SMF，さらにミックス前の各楽器パートごとの分離信号であるマスタートラックの全てが利用できるものである．マスタートラックは，分離後の音響信号がどれだけミックス前の音響信号に近いかという定量的な評価のために必要である．
- 本手法は，歌声を扱うことを想定していない．これは，歌声が楽器音に比べて音響的特徴の変化が複雑であるためモデルで表現可能な範囲を超えていること，歌声テンプレートの作成が困難なことがその理由である．

テンプレート音と分離対象となるテスト用楽曲は，異なる MIDI 音源で生成した．その他の詳細な実験条件を表 3 に示す．これらのパラメータは，実験的に最適なものを求めたものである．

6.3 実験結果

図 2 に，各楽器パートの SNR を楽曲ごとに平均した結果 (p001～p010) と，それらの SNR をさらに全 10 曲で平均した結果 (Average) を示す．全 10 曲の SNR を平均した結果を見ると，併用モデルを用いた場合 (Integrated) に SNR は最も高くなっており，提案手法の有用性が示されている．全体的な傾向として，併用モデルを用いた場合 (Integrated) と調波構造モデルだけを用いた場合 (Harmonic) では SNR にそれほど大きな差は現れていないのに対して，非調波構造モデルだけを用いた場合 (Inharmonic) はおよそ 10 dB の悪くなっている。非調波構造モデルに対する制約は調波構造と非調

波構造が分離されていることが要求されるために，非調波構造モデルのみを用いた場合，モデル更新の際にこれらの制約を用いることができずモデルが大幅に過学習を起こしてしまったためと考えられる．

また，ドラムパートを除くほぼ全ての楽器パートに共通していた特徴として，ドラム音が他のパートに混ざるということがあった．これは，ドラム音の非調波成分の一部が他の楽器音の非調波構造モデルで表現されてしまっていることを意味している．これによってドラムパートが減衰し，楽器音イコライザでドラムパートの音量を下げても他の楽器パートに混ざったドラム音が残るため，楽器音イコライザとして不十分な性能になってしまうなどの悪影響が考えられる．

7. おわりに

本稿では，調波構造モデルと非調波構造モデルを統合したモデルを用いた多重奏の音源分離手法と，そのためのモデル適応手法について述べた．また，本手法の性能を示すために評価実験を行い，併用モデルを用いることの有効性を確認した．我々は，本手法で分離した音楽音響信号を用いた楽器音イコライザを開発している．今後は，実演奏楽曲での評価や分離における前提条件の緩和を行っていく予定である．

謝辞 本研究の一部は、科学研究費補助金（基盤研究 (A), 特定領域「情報学」），21 世紀 COE プログラム「知識社会基盤構築のための情報学拠点形成」，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

1) Goto, M.: Active Music Listening Interfaces Based on Signal Processing, *Proc. ICASSP* (2007).
2) Yoshii, K., Goto, M. and Okuno, H.G.: INTER:D: A Drum Sound Equalizer for Controlling Volume and Timbre of Drums, *Proc. EWIMT*, pp.205–212 (2005).
3) 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, *IPSJ Journal*, Vol.48, No.3, pp.134–144 (2007).
4) Virtanen, T. and Klapuri, A.: Separation of Harmonic Sounds Using Linear Models for the Overtone Series, *Proc. ICASSP*, Vol.II, pp.1757–1760 (2002).
5) Every, M. and Szymanski, J.: A Spectral-filtering Approach to Music Signal Separation, *Proc. DAFx*, pp.197–200 (2004).
6) Kameoka, H., Nishimoto, T. and Sagayama, S.: Harmonic-temporal Structured Clustering via Deterministic Annealing EM Algorithm for Audio Feature Extraction, *Proc. ISMIR*, pp.115–122 (2005).
7) Viste, H. and Evangelista, G.: A Method for Separation of Overlapping Partials Based on Similarity of Temporal Envelopes in Multichannel Mixtures, *IEEE Transactions on Speech and Audio Processing*, Vol.14, No.3, pp.1051–1061 (2006).
8) Woodruff, J., Pardo, B. and Dannenberg, R.: Remixing Stereo Music with Score-informed Source Separation, *Proc.ISMIR*, pp.314–319 (2006).
9) Helen, M. and Virtanen, T.: Separation of Drums from Polyphonic Music Using Non-negative Matrix Factorization and Support Vector Machine, *Proc. EUSIPCO* (2005).
10) Barry, D., Fitzgerald, D., Coyle, E. and Lawlor, B.: Drum Source Separation Using Percussive Feature Detection and Spectral Modulation, *Proc. ISSC*, pp.13–17 (2005).
11) Yoshii, K., Goto, M. and Okuno, H. G.: Drum Sound Recognition for Polyphonic Audio Signals by Adaptation and Matching of Spectrogram Templates with Harmonic Structure Suppression, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol.15, No.1, pp.333–345 (2007).
12) Goto, M.: A Real-time Music-scene-description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-world Audio Signals, *Speech Communication (ISCA Journal)*, Vol.43, No.4, pp.311–329 (2004).
13) Cano, P., Loscos, A. and Bonada, J.: Score-performance Matching Using HMMs, *Proc. ICMC*, pp.441–444 (1999).
14) Dannenberg, R.B. and Hu, N.: Polyphonic Audio Matching for Score Following and Intelligent Audio Editors, *Proc. ICMC*, pp.27–33 (2003).
15) Adams, N., Marquez, D. and Wakefield, G.: Iterative Deepening for melody Alignment and Retrieval, *Proc. ISMIR*, pp.199–206 (2005).
16) Dixon, S. and Widmer, G.: MATCH: A Music Alignment Tool Chest, *Proc. ISMIR*, pp.492–497 (2005).
17) Cont, A.: Realtime Audio to Score Alignment for Polyphonic Music Instruments Using Sparce Non-negative Constraints and Hierarchical HMMs, *Proc. ICASSP*, Vol.II, pp. 641–644 (2006).
18) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC Music Database: Popular, Classical, and Jazz Music Databases, *Proc. ISMIR*, pp.287–288 (2002).

付　　録

A.1 パラメータ更新式の導出

式 23 のコスト関数を各パラメータで偏微分したものの零点を求めることで，J が極小になるようにパラメータを更新する式を導出する．

A.1.1 r_{klc}: 各チャンネルの相対強度

$$\frac{\partial J}{\partial r_{klc}} = \sum_{y,n} \iint \left(-\frac{G^{(H)}_{klyn}}{r_{klc}} + w_{kl} E_{kly} F_{kln} \right) dt\, df + \iint -\frac{G^{(I)}_{kl}}{r_{klc}}\, dt\, df - \lambda^{(r)}_{kl} = 0 \tag{24}$$

$$\frac{\partial J}{\partial \lambda^{(r)}_{kl}} = \sum_{c} r_{klc} - 1 = 0 \tag{25}$$

この連立方程式を解き，以下を得る．

$$r_{klc} = \frac{\iint \left(\sum_{y,n} G^{(H)}_{klyn} + G^{(I)}_{kl} \right) dt\, df}{\sum_{c} \iint \left(\sum_{y,n} G^{(H)}_{klyn} + G^{(I)}_{kl} \right) dt\, df} \tag{26}$$

A.1.2 w_{kl}: 調波構造モデルの重み

$$\frac{\partial J}{\partial w_{kl}} = \sum_{c,y,n} \iint \left(-\frac{G^{(H)}_{klyn}}{w_{kl}} + r_{klc} E_{kly} F_{kln} \right) dt\, df = 0 \tag{27}$$

この方程式を解き，以下を得る．

$$w_{kl} = \frac{\sum_{c,y,n} \iint G^{(H)}_{klyn}\, dt\, df}{\sum_{c,y,n} \iint E_{kly} F_{kln}\, dt\, df} \tag{28}$$

A.1.3 $\mu_{kl}(t)$: F0 の軌跡

$$\frac{\partial J}{\partial \mu_{kl}(t)} = \sum_{c,y,n} \int -\frac{n\,(f - n\mu_{kl}(t))\, G^{(H)}_{klyn}}{\sigma^2_{kl}}\, df - \beta_\mu \left(\frac{\bar{\mu}_{kl}(t)}{\mu_{kl}(t)} - 1 \right) = 0 \tag{29}$$

$$\Rightarrow \quad a_\mu \mu_{kl}(t)^2 + b_\mu \mu_{kl}(t) + c_\mu = 0 \tag{30}$$

$$\begin{cases} a_\mu = \sum_{c,y,n} \int n^2 G^{(H)}_{klyn}\, df \\ b_\mu = \sigma^2_{kl} \beta_\mu - \sum_{c,y,n} \int n f G^{(H)}_{klyn}\, df \\ c_\mu = -\sigma^2_{kl} \beta_\mu \bar{\mu}_{kl}(t) \end{cases} \tag{31}$$

この方程式を解き，以下を得る．

$$\mu_{kl}(t) = \frac{-b_\mu + \sqrt{b^2_\mu - 4a_\mu c_\mu}}{2a_\mu} \tag{32}$$

A.1.4 u_{kly}: パワーエンベロープの概形

$$\frac{\partial J}{\partial u_{kly}} = \sum_{c,n} \iint -\frac{G^{(H)}_{klyn}}{u_{kly}}\, dt\, df - \lambda^{(u)}_{kl} = 0 \tag{33}$$

$$\frac{\partial J}{\partial \lambda^{(u)}_{kl}} = \sum_{y} u_{kly} - 1 = 0 \tag{34}$$

この連立方程式を解き，以下を得る．

$$u_{kly} = \frac{\sum_{c,n} \iint G^{(H)}_{klyn}\, dt\, df}{\sum_{c,y,n} \iint G^{(H)}_{klyn}\, dt\, df} \tag{35}$$

A.1.5 v_{kln}: n 次倍音成分の相対強度

$$\frac{\partial J}{\partial v_{kln}} = \sum_{c,y} \iint -\frac{G^{(H)}_{klyn}}{v_{kln}}\, dt\, df - \beta_v \frac{\bar{v}_{kn}}{v_{kln}} - \lambda^{(v)}_{kl} = 0 \tag{36}$$

$$\frac{\partial J}{\partial \lambda^{(v)}_{kl}} = \sum_{n} v_{kln} - 1 = 0 \tag{37}$$

この連立方程式を解き，以下を得る．

$$v_{kln} = \frac{\beta_v \bar{v}_{kn} + \sum_{c,y} \iint G^{(H)}_{klyn}\, dt\, df}{\beta_v + \sum_{c,y,n} \iint G^{(H)}_{klyn}\, dt\, df} \tag{38}$$

A.1.6 τ_{kl}: オンセット時刻

$$\frac{\partial J}{\partial \tau_{kl}} = \sum_{c,y,n} \iint -G^{(H)}_{klyn} \frac{t - \tau_{kl} - y\phi_{kl}}{\phi^2_{kl}}\, dt df = 0 \tag{39}$$

この方程式を解き，以下を得る．

$$\tau_{kl} = \frac{\sum_{c,y,n} \iint (t - y\phi_{kl})\, G^{(H)}_{klyn}\, dt\, df}{\sum_{c,y,n} \iint G^{(H)}_{klyn}\, dt\, df} \tag{40}$$

A.1.7 $Y\phi_{kl}$: 音長

$$\frac{\partial J}{\partial \phi_{kl}} = \sum_{c,y,n} \iint G^{(H)}_{klyn} \frac{(t - \tau_{kl})\,(t - \tau_{kl} - y\phi_{kl}) - \phi^2_{kl}}{\phi^3_{kl}}\, dt\, df = 0 \tag{41}$$

$$\Rightarrow \quad a_\phi \phi^2_{kl} + b_\phi \phi_{kl} + c_\phi = 0 \tag{42}$$

$$\begin{cases} a_\phi = \sum_{c,y,n} \iint G^{(H)}_{klyn}\, dt\, df \\ b_\phi = \sum_{c,y,n} \iint y\,(t - \tau_{kl})\, G^{(H)}_{klyn}\, dt\, df \\ c_\phi = -\sum_{c,y,n} \iint (t - \tau_{kl})^2\, G^{(H)}_{klyn}\, dt\, df \end{cases} \tag{43}$$

この方程式を解き，以下を得る．

$$\phi_{kl} = \frac{-b_\phi + \sqrt{b^2_\phi - 4a_\phi c_\phi}}{2a_\phi} \tag{44}$$

A.1.8 σ_{kl}: 周波数方向の分散

$$\frac{\partial J}{\partial \sigma_{kl}} = \sum_{c,y,n} \iint -G^{(H)}_{klyn} \frac{-n^2\sigma^2_{kl} + (f - n\mu_{kl}(t))^2}{n^2\sigma^3_{kl}}\, dt\, df = 0 \tag{45}$$

この方程式を解き，以下を得る．

$$\sigma_{kl} = \sqrt{\frac{\sum_{c,y,n} \iint (f/n - \mu_{kl}(t))^2\, G^{(H)}_{klyn}\, dt\, df}{\sum_{c,y,n} \iint G^{(H)}_{klyn}\, dt\, df}} \tag{46}$$

A.1.9 $I_{kl}(t,f)$: 非調波成分

$$\frac{\partial J}{\partial I_{kl}} = \sum_{c} \left(-\frac{G^{(I)}_{kl}}{I_{kl}} + r_{klc} \right) + \beta_{I1} \left(-\frac{\bar{I}_k}{I_{kl}} + 1 \right) + \beta_{I2} \left(-\frac{\bar{I}_{kl}}{I_{kl}} + 1 \right) = 0 \tag{47}$$

この方程式を解き，以下を得る．

$$I_{kl} = \frac{\sum_c \left(G^{(I)}_{kl} + \beta_{I1} \bar{I}_{kl} + \beta_{I2} \bar{I}_k \right)}{1 + \beta_{I1} + \beta_{I2}} \tag{48}$$