

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5201602号
(P5201602)

(45) 発行日 平成25年6月5日(2013.6.5)

(24) 登録日 平成25年2月22日(2013.2.22)

(51) Int.Cl.	F I
G 1 O L 21/0308 (2013.01)	G 1 O L 21/02 2 O 3 Z
G 1 O H 1/00 (2006.01)	G 1 O H 1/00 1 O 2 Z
G 1 O L 21/028 (2013.01)	G 1 O H 1/00 B
	G 1 O L 21/02 2 O 1 Z

請求項の数 12 (全 34 頁)

(21) 出願番号 特願2009-511801 (P2009-511801)	(73) 特許権者 504132272 国立大学法人京都大学 京都府京都市左京区吉田本町36番地1
(86) (22) 出願日 平成20年4月14日(2008.4.14)	(73) 特許権者 301021533 独立行政法人産業技術総合研究所 東京都千代田区霞が関1-3-1
(86) 国際出願番号 PCT/JP2008/057310	(74) 代理人 100091443 弁理士 西浦 ▲嗣▼晴
(87) 国際公開番号 W02008/133097	(72) 発明者 糸山 克寿 京都府京都市左京区吉田本町 国立大学法人京都大学大学院情報学研究科内
(87) 国際公開日 平成20年11月6日(2008.11.6)	(72) 発明者 奥乃 博 京都府京都市左京区吉田本町 国立大学法人京都大学大学院情報学研究科内
審査請求日 平成23年3月28日(2011.3.28)	
(31) 優先権主張番号 特願2007-106576 (P2007-106576)	
(32) 優先日 平成19年4月13日(2007.4.13)	
(33) 優先権主張国 日本国(JP)	

最終頁に続く

(54) 【発明の名称】 音源分離システム、音源分離方法及び音源分離用コンピュータプログラム

(57) 【特許請求の範囲】

【請求項1】

複数種類の楽器から発生した複数種類の楽器音に対応する複数の楽器音信号が含まれる入力音響信号と時間的に同期し且つ前記複数の入力音響信号に対応した前記複数種類の楽器によりそれぞれ演奏されるべき複数種類の楽譜についての楽譜情報データを記憶する楽譜情報データ記憶部と、

前記複数種類の楽器に関して、それぞれの前記楽器から生じる複数種類の単音を、それぞれ調波モデルと非調波モデルとからなる複数の調波・非調波混合モデルで表現するために予め作成された、前記調波・非調波混合モデルをそれぞれ構成するための複数のパラメータを含む複数のモデルパラメータを用いて、前記複数種類の楽譜に含まれる複数の単音をそれぞれ前記複数のモデルパラメータに置き換えて、前記複数種類の楽譜に対応した、複数の前記モデルパラメータが集まって構成される複数種類のモデルパラメータ集合データを作成して記憶手段に記憶するモデルパラメータ集合データ作成記憶部と、

前記複数種類のモデルパラメータ集合データから、各時刻における前記モデルパラメータを読み出し、読み出した複数の前記モデルパラメータにそれぞれ含まれる前記複数のパラメータと予め定めた第1のモデルパラメータ変換式とを用いて、読み出した前記複数のモデルパラメータに対応した複数の初期パワースペクトルを生成して記憶手段に記憶する第1のパワースペクトル生成記憶部と、

各時刻において、前記第1のパワースペクトル生成記憶部に記憶された前記複数の初期パワースペクトルを合成して各時刻毎の合成パワースペクトルを作り、各時刻毎の前記合

成パワースペクトルに対する各時刻毎の前記複数の初期パワースペクトルの割合を示す複数の初期分配関数を計算して記憶手段に記憶する初期分配関数計算記憶部と、

第1回目の分離処理においては、各時刻の前記入力音響信号のパワースペクトルから、各時刻における前記複数の初期分配関数を用いて、各時刻における前記複数種類の楽器に対応する複数のパワースペクトルを分離して記憶手段に記憶し、第2回目以降の分離処理においては更新分配関数を用いて、各時刻における前記複数種類の楽器に対応する複数のパワースペクトルを分離して当該記憶手段に記憶するパワースペクトル分離記憶部と、

各時刻において分離された前記複数のパワースペクトルから、それぞれ前記調波・非調波混合モデルで表現するために必要な複数のパラメータを含む複数の更新モデルパラメータを推定し、複数の前記更新モデルパラメータが集まって構成される複数種類の更新モデルパラメータ集合データを作成して記憶手段に記憶する更新モデルパラメータ推定記憶部と、

前記更新モデルパラメータ推定記憶部に記憶されている前記複数種類の更新モデルパラメータ集合データから、各時刻における前記更新モデルパラメータを読み出して、読み出した複数の前記更新モデルパラメータにそれぞれ含まれる前記複数のパラメータと予め定めた第2のモデルパラメータ変換式とを用いて、読み出した前記複数の更新モデルパラメータに対応した複数の更新パワースペクトルを生成して記憶手段に記憶する第2のパワースペクトル生成記憶部と、

各時刻において、前記第2のパワースペクトル生成記憶部に記憶された前記複数の更新パワースペクトルを合成して各時刻毎の合成パワースペクトルを作り、各時刻毎の前記合成パワースペクトルに対する各時刻毎の前記複数の更新パワースペクトルの割合を示す複数の前記更新分配関数を計算して記憶手段に記憶する更新分配関数計算記憶部とを備え、

前記パワースペクトル分離記憶部が第2回目以降の前記分離処理を行うたびに、前記更新モデルパラメータ推定記憶部は、前記更新パワースペクトルが前記初期パワースペクトルに近い状態から前記パワースペクトル分離記憶部に記憶された最新の前記複数のパワースペクトルに近い状態に徐々に近づくように、前記複数の更新モデルパラメータにそれぞれ含まれる前記複数のパラメータを推定するように構成され、

前記更新パワースペクトルが前記初期パワースペクトルに近い状態から前記パワースペクトル分離記憶部に記憶された最新の前記複数のパワースペクトルに近い状態になるまで、前記パワースペクトル分離記憶部と、前記更新モデルパラメータ推定記憶部と、前記第2のパワースペクトル生成記憶部と、前記更新分配関数計算記憶部とが処理動作を繰り返すことを特徴とする音源分離システム。

【請求項2】

前記パワースペクトル分離記憶部が前記分離処理を行うたびに、前記パワースペクトル分離記憶部に記憶された各時刻における前記複数のパワースペクトルと前記第2のパワースペクトル生成記憶部に記憶された各時刻における前記複数の更新パワースペクトルとの間のKLダイバージェンス $J_1 \times$ (前記は0から1を満たす実数)と、前記第2のパワースペクトル生成記憶部に記憶された各時刻における前記複数の更新パワースペクトルと前記第1のパワースペクトル生成記憶部に記憶された各時刻における前記複数の初期パワースペクトルとの間のKLダイバージェンス $J_2 \times$ (1から)のすべての総和 J_0 に基づいてコスト関数 J を定め、該コスト関数を最小化するように、前記複数の更新モデルパラメータにそれぞれ含まれる前記複数のパラメータを推定するべく更新モデルパラメータ推定記憶部が構成され、前記は前記分離処理が行われるたびに増加し、1になるまで、前記パワースペクトル分離記憶部と、前記更新モデルパラメータ推定記憶部と、第2のパワースペクトル生成記憶部と、前記更新分配関数計算記憶部とが処理動作を繰り返すことを特徴とする請求項1に記載の音源分離システム。

【請求項3】

前記第1及び第2のモデルパラメータ変換式では、それぞれ下記の調波・非調波混合モデルを用い、

$$h_{kl} = r_{kl} c (H_{kl}(t, f) + I_{kl}(t, f))$$

10

20

30

40

50

前記 h_{kl} は単音のパワースペクトルであり、

前記 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、

$H_{kl}(t, f)$ は音量、基本周波数 F_0 の時間変化、パワーエンベロープの概形を表現する y 番目のガウシアン重み係数、 n 次倍音成分の相対強度、オンセット時刻、音長及び周波数方向の分散という特徴量を表す複数のパラメータにより構成される調波モデルであり、

前記 $I_{kl}(t, f)$ はノンパラメトリックな関数によって表される非調波モデルである請求項 2 に記載の音源分離システム。

【請求項 4】

前記更新モデルパラメータ推定記憶部で使用する前記コスト関数は、前記非調波モデルが調波構造を表現しないための制約条件を含んでいる請求項 3 に記載の音源分離システム。

10

【請求項 5】

前記調波モデルは、音高の時間変化を扱うための関数 $\mu_{kl}(t)$ を含んでおり、

前記更新モデルパラメータ推定記憶部で使用する前記コスト関数は、前記基本周波数 F_0 が時間的に不連続値をとらないための制約条件を含んでいる請求項 4 に記載の音源分離システム。

【請求項 6】

前記更新モデルパラメータ推定記憶部で使用する前記コスト関数は、前記調波モデルについて同一楽器の単音に対して倍音成分の相対強度比を一定にするための制約条件を含んでいる請求項 5 に記載の音源分離システム。

20

【請求項 7】

前記更新モデルパラメータ推定記憶部で使用する前記コスト関数は、前記非調波モデルについて同一楽器の単音に対して非調波成分比を一定にするための制約条件を含んでいる請求項 6 に記載の音源分離システム。

【請求項 8】

前記複数種類の楽器に関して、それぞれの前記楽器から生じる複数種類の単音に対応した複数の標準的なパワースペクトルにより表現された複数のテンプレートに基づいて、前記複数種類の単音を、それぞれ調波モデルと非調波モデルとからなる複数の調波・非調波混合モデルで表現するために、前記調波・非調波混合モデルをそれぞれ構成するための複数のパラメータを含む複数のモデルパラメータを作成して記憶手段に予め記憶する音モデル構成用モデルパラメータ作成記憶部を更に備えており、

30

前記モデルパラメータ集合データ作成記憶部は、前記音モデル構成用モデルパラメータ作成記憶部に記憶された前記複数のモデルパラメータを用いて、前記モデルパラメータ集合データを作成する請求項 1 に記載の音源分離システム。

【請求項 9】

前記楽譜情報データに含まれる前記複数の楽器ごとの複数の単音の情報を複数のパラメータ音に変換する音響変換手段と、前記複数のパラメータ音の複数のパワースペクトルを、それぞれ調波モデルと非調波モデルとからなる複数の調波・非調波混合モデルで表現するために、前記調波・非調波混合モデルをそれぞれ構成するための複数のパラメータを含む複数のモデルパラメータを作成する音モデル構成用モデルパラメータ作成部とを更に備えており、

40

前記モデルパラメータ集合データ作成記憶部は、前記音モデル構成用モデルパラメータ作成部が作成した前記複数のモデルパラメータを用いて、前記モデルパラメータ集合データを作成する請求項 1 に記載の音源分離システム。

【請求項 10】

複数種類の楽器から発生した複数種類の楽器音に対応する複数の楽器音信号が含まれる入力音響信号と時間的に同期し且つ前記複数の入力音響信号に対応した前記複数種類の楽器によりそれぞれ演奏されるべき複数種類の楽譜についての楽譜情報データを用意するステップと、

50

前記複数種類の楽器に関して、それぞれの前記楽器から生じる複数種類の単音を、それぞれ調波モデルと非調波モデルとからなる複数の調波・非調波混合モデルで表現するために予め作成された、前記調波・非調波混合モデルをそれぞれ構成するための、複数のパラメータを含む複数のモデルパラメータを用いて、前記複数種類の楽譜に含まれる複数の単音をそれぞれ前記複数のモデルパラメータに置き換えて、前記複数種類の楽譜に対応した、複数の前記モデルパラメータが集まって構成される複数種類のモデルパラメータ集合データを作成するステップと、

前記複数種類のモデルパラメータ集合データから、各時刻における前記モデルパラメータを読み出し、読み出した複数の前記モデルパラメータにそれぞれ含まれる前記複数のパラメータと予め定めた第1のモデルパラメータ変換式とを用いて、読み出した前記複数のモデルパラメータに対応した複数の初期パワースペクトルを生成するステップと、

10

各時刻において、前記複数の初期パワースペクトルを合成して各時刻毎の合成パワースペクトルを作り、各時刻毎の前記合成パワースペクトルに対する各時刻毎の前記複数の初期パワースペクトルの割合を示す複数の初期分配関数を計算するステップと、

第1回目の分離処理においては、各時刻の前記入力音響信号のパワースペクトルから、各時刻における前記複数の初期分配関数を用いて、各時刻における前記複数種類の楽器に対応する複数のパワースペクトルを分離し、第2回目以降の分離処理においては更新分配関数を用いて、各時刻における前記複数種類の楽器に対応する複数のパワースペクトルを分離するステップと、

各時刻において分離された前記複数のパワースペクトルから、それぞれ前記調波・非調波混合モデルで表現するために必要な複数のパラメータを含む複数の更新モデルパラメータを推定し、複数の前記更新モデルパラメータが集まって構成される複数種類の更新モデルパラメータ集合データを作成するステップと、

20

前記複数種類の更新モデルパラメータ集合データから、各時刻における前記更新モデルパラメータを読み出して、読み出した複数の前記更新モデルパラメータにそれぞれ含まれる前記複数のパラメータと予め定めた第2のモデルパラメータ変換式とを用いて、前記読み出した複数の更新モデルパラメータに対応した複数の更新パワースペクトルを生成するステップと、

各時刻において、前記複数の更新パワースペクトルを合成して各時刻毎の合成パワースペクトルを作り、各時刻毎の前記合成パワースペクトルに対する各時刻毎の前記複数の更新パワースペクトルの割合を示す複数の前記更新分配関数を計算するステップと、

30

第2回目以降の前記分離処理を行うたびに、前記更新モデルパラメータを推定するステップでは、前記更新パワースペクトルが前記初期パワースペクトルに近い状態から前記パワースペクトルを分離するステップで分離された最新の前記複数のパワースペクトルに近い状態に徐々に近付くように、前記複数の更新モデルパラメータにそれぞれ含まれる前記複数のパラメータを推定し、

前記更新パワースペクトルが前記初期パワースペクトルに近い状態から前記パワースペクトルを分離するステップで分離された最新の前記複数のパワースペクトルに近い状態になるまで、前記パワースペクトルを分離するステップと、更新モデルパラメータを推定するステップと、前記更新パワースペクトルを生成するステップと、前記更新分配関数を計算するステップとを繰り返すことをコンピュータが実施することを特徴とする音源分離方法。

40

【請求項11】

前記パワースペクトルを分離するステップで、第2回目以降の前記分離処理を行うたびに、各時刻における前記複数のパワースペクトルと各時刻における前記複数の更新パワースペクトルとの間のKLダイバージェンス $J_1 \times$ (前記 は $0 \leq$ 1 を満たす実数) と、各時刻における前記複数の更新パワースペクトルと各時刻における前記複数の初期パワースペクトルとの間のKLダイバージェンス $J_2 \times (1 -)$ のすべての総和 J_0 に基づいてコスト関数 J を定め、該コスト関数を最小化するように、前記複数の更新モデルパラメータにそれぞれ含まれる前記複数のパラメータを推定し、前記 を前記分離処理が行

50

われるたびに増加させて、前記 が 1 になると前記分離処理を停止する請求項 10 に記載の音源分離方法。

【請求項 12】

コンピュータにインストールされて下記のステップを実行する音源分離用コンピュータプログラムであって、

複数種類の楽器から発生した複数種類の楽器音に対応する複数の楽器音信号が含まれる入力音響信号と時間的に同期し且つ前記複数の入力音響信号に対応した前記複数種類の楽器によりそれぞれ演奏されるべき複数種類の楽譜についての楽譜情報データを用意するステップと、

前記複数種類の楽器に関して、それぞれの前記楽器から生じる複数種類の単音を、それぞれ調波モデルと非調波モデルとからなる複数の調波・非調波混合モデルで表現するために予め作成された、前記調波・非調波混合モデルをそれぞれ構成するための複数のパラメータを含む複数のモデルパラメータを用いて、前記複数種類の楽譜に含まれる複数の単音をそれぞれ前記複数のモデルパラメータに置き換えて、前記複数種類の楽譜に対応した、複数の前記モデルパラメータが集まって構成される複数種類のモデルパラメータ集合データを作成するステップと、

前記複数種類のモデルパラメータ集合データから、各時刻における前記モデルパラメータを読み出し、読み出した複数の前記モデルパラメータにそれぞれ含まれる前記複数のパラメータと予め定めた第 1 のモデルパラメータ変換式とを用いて、読み出した前記複数のモデルパラメータに対応した複数の初期パワースペクトルを生成するステップと、

各時刻において、前記複数の初期パワースペクトルを合成して各時刻毎の合成パワースペクトルを作り、各時刻毎の前記合成パワースペクトルに対する各時刻毎の前記複数の初期パワースペクトルの割合を示す複数の初期分配関数を計算するステップと、

第 1 回目の分離処理においては、各時刻の前記入力音響信号のパワースペクトルから、各時刻における前記複数の初期分配関数を用いて、各時刻における前記複数種類の楽器に対応する複数のパワースペクトルを分離し、第 2 回目以降の分離処理においては更新分配関数を用いて、各時刻における前記複数種類の楽器に対応する複数のパワースペクトルを分離するステップと、

各時刻において分離された前記複数のパワースペクトルから、それぞれ前記調波・非調波混合モデルで表現するために必要な複数のパラメータを含む複数の更新モデルパラメータを推定し、複数の前記更新モデルパラメータが集まって構成される複数種類の更新モデルパラメータ集合データを作成するステップと、

前記複数種類の更新モデルパラメータ集合データから、各時刻における前記更新モデルパラメータを読み出して、読み出した複数の前記更新モデルパラメータにそれぞれ含まれる前記複数のパラメータと予め定めた第 2 のモデルパラメータ変換式とを用いて、読み出した複数の更新モデルパラメータに対応した複数の更新パワースペクトルを生成するステップと、

各時刻において、前記複数の更新パワースペクトルを合成して各時刻毎の合成パワースペクトルを作り、各時刻毎の前記合成パワースペクトルに対する各時刻毎の前記複数の更新パワースペクトルの割合を示す複数の前記更新分配関数を計算するステップと、

第 2 回目以降の前記分離処理を行うたびに、前記更新モデルパラメータを推定するステップでは、前記更新パワースペクトルが前記初期パワースペクトルに近い状態から前記パワースペクトルを分離するステップで分離された最新の前記複数のパワースペクトルに近い状態に徐々に近付くように、前記複数の更新モデルパラメータにそれぞれ含まれる前記複数のパラメータを推定し、

前記更新パワースペクトルが前記初期パワースペクトルに近い状態から前記パワースペクトルを分離するステップで分離された最新の前記複数のパワースペクトルに近い状態になるまで、前記パワースペクトルを分離するステップと、前記更新モデルパラメータを推定するステップと、前記更新パワースペクトルを生成するステップと、前記更新分配関数を計算するステップとを繰り返すことを特徴とする音源分離用コンピュータプログラム。

10

20

30

40

50

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、複数種類の楽器音信号が含まれる入力音響信号から各楽器に対応する楽器音信号を分離することを可能にする音源分離システム、方法及びプログラムに関するものである。本発明は特に、調波構造と非調波構造の両方の信号成分を持った「複数の楽器演奏による混合音の音響信号」を各楽器パートごとの音源に分離する音源分離システム、方法及びこれを実現するためのコンピュータプログラムに関するものである。

【背景技術】

【0002】

スピーカから出力される音楽音響信号（以下、単に「音響信号」という。）に含まれるドラムなどの非調波構造の信号成分を分離して、他の信号成分に影響を与えずに独立に音量を増減させることができる音響信号処理システムが知られている（例えば、特許文献1参照）。

【0003】

この従来のシステムは専ら音響信号に含まれる非調波構造の信号のみを対象とするものである。そのため、従来のシステムでは、「非調波構造と調波構造の両方の信号成分を含む混合音」を、それぞれの楽器音ごとに分離することはできなかった。

【0004】

調波構造を表すモデル（以下、本明細書では「調波モデル」という。）と非調波構造を表すモデル（以下、本明細書では「非調波モデル」という。）を同時に扱うモデル（以下、本明細書では「調波・非調波混合モデル」という。）を用いた音源分離技術については報告例が見当たらない。

【特許文献1】特開2006-5807号公報

【発明の開示】

【発明が解決しようとする課題】

【0005】

一般に、調波構造の信号波形は、基本周波数（ F_0 ）とその n 次倍音からなる調波とが重ね合わされたものである。したがって調波構造の信号波形としては、直感的には、音の高さを持つ楽器（ピアノやフルート、ギターなど）が奏でる音の信号波形が該当する。調波構造を持つ信号波形のモデルは、その音響信号のパワースペクトルの特徴量（音高、強度、オンセット時刻、音長、音色など）を推定することで音源分離が可能となることが知られている。特徴量の抽出方法には種々の方法が提案されている。多くの方法では、パラメータを含む関数を定義してそのパラメータを適応学習により推定するというものである。

【0006】

これに対し、非調波構造の信号波形は、調波構造のような基本周波数も高調波もいずれも持たないものである。例えば、音の高さを持たない楽器（ドラムなど）が発する音の波形がこれに当たる。非調波構造を持つ信号波形のモデルは、パワースペクトルでしか表現することができない。

【0007】

両構造を同時に扱うことの難しさは、モデルパラメータに対する制約条件が殆どないために、あらゆるパラメータを同時に扱わなければならない点にある。あらゆるパラメータを同時に扱うと、適応学習を行った際にモデルパラメータを望ましい形で収束させることができない。

【0008】

しかし、多重奏中のあらゆる楽器のパートの音量を自由に調節するためには、調波構造と非調波構造の両方を同時に扱うことが不可欠である。また、一般には調波構造に区分される楽器音であっても、楽器の物理的な構造に由来して厳密には調波的でない信号波形を含む場合も多い。例えば、ピアノの発音機構は、ハンマーが発音時に弦を叩くことによ

10

20

30

40

50

て発生した音をピアノの本体部分で共鳴させることによる。そのため、ピアノの音は厳密には共鳴による調波構造とハンマーの打撃音による非調波構造の両方の音響信号を含むものである。

【0009】

すなわち、楽曲中に含まれる全ての音源を分離するためには、調波的な音響信号と非調波的な音響信号の両方を同時に扱いながらモデルパラメータを望ましい形で収束させることが重要となる。

【0010】

本発明は、調波的な音響信号と非調波的な音響信号の両方の成分を含む混合音の音源を分離するための音源分離システムとこれを実現するためのコンピュータプログラム及び音源分離方法を提供することを主たる目的とする。

10

【課題を解決するための手段】

【0011】

本発明に係る音源分離システムは、楽譜情報データ記憶部と、モデルパラメータ集合データ作成記憶部と、第1のパワースペクトル生成記憶部と、初期分配関数計算記憶部と、パワースペクトル分離記憶部と、更新モデルパラメータ推定記憶部と、第2のパワースペクトル生成記憶部と、更新分配関数計算記憶部とを少なくとも備えている。

【0012】

楽譜情報データ記憶部は、複数種類の楽器から発生した複数種類の楽器音に対応する複数の楽器音信号が含まれる入力音響信号（混合音信号）と時間的に同期し且つ複数の入力音響信号に対応した複数種類の楽器によりそれぞれ演奏されるべき複数種類の楽譜についての楽譜情報データを記憶する。楽譜情報データとしては、例えば標準MIDIファイル（SMF）を利用することができる。

20

【0013】

モデルパラメータ集合データ作成記憶部では、複数種類の楽器に関して、それぞれの楽器から生じる複数種類の単音を、それぞれ調波モデルと非調波モデルとからなる複数の調波・非調波混合モデルで表現するために予め作成された、調波・非調波混合モデルをそれぞれ構成するための複数のパラメータを含む複数のモデルパラメータを用いる。まずモデルパラメータ集合データ作成記憶部では、調波・非調波混合モデルをそれぞれ構成するための複数のパラメータを含む複数のモデルパラメータを用いて、複数種類の楽譜に含まれる複数の単音をそれぞれ複数のモデルパラメータに置き換える。そして複数種類の楽譜に対応した、複数のモデルパラメータが集まって構成される複数種類のモデルパラメータ集合データを作成して記憶手段に記憶する。

30

【0014】

調波・非調波混合モデルをそれぞれ構成するための複数のパラメータを含む複数のモデルパラメータの作成方法は任意である。例えば、音モデル構成用モデルパラメータ作成記憶部を設けてもよい。この音モデル構成用モデルパラメータ作成記憶部は、複数種類の楽器に関して、それぞれの楽器から生じる複数種類の単音に対応した複数の標準的なパワースペクトルにより表現された複数のテンプレートに基づいて、複数種類の単音を、それぞれ調波モデルと非調波モデルとからなる複数の調波・非調波混合モデルで表現するために、調波・非調波混合モデルをそれぞれ構成するための複数のパラメータを含む複数のモデルパラメータを作成して記憶手段に予め記憶する。このような音モデル構成用モデルパラメータ作成記憶部を設ける場合には、モデルパラメータ集合データ作成記憶部は、音モデル構成用モデルパラメータ作成記憶部に記憶された複数のモデルパラメータを用いて、モデルパラメータ集合データを作成する。

40

【0015】

ここでテンプレートとは、例えばMIDIファイルの楽譜に基づいてMIDI音源で生成した各単音のサンプル音（テンプレート音）のパワースペクトルである。具体的には、テンプレートとは、ある種類の楽器が発することができる複数種類の単音（異なる音階の複数種類の単音）を、それぞれ標準的なパワースペクトルで表現したものである。すなわ

50

ち標準的なギターの「ド」の音を標準的なパワースペクトルで表現したものが、テンプレートである。このギターの「ド」の単音のテンプレートのパワースペクトルは、入力音響信号中に含まれているギターの楽器音信号中の「ド」の単音のパワースペクトルとは、ある程度近似しているが、同じではない。調波・非調波混合モデルは、時刻 t 、周波数 f 、 k 番目の楽器、 l 番目の単音に対して、調波構造を表現する調波モデル $H_{kl}(t, f)$ と非調波構造を表現する非調波モデル $I_{kl}(t, f)$ の線形和として定義される。この調波・非調波混合モデルは、一つのモデルで調波構造の信号成分と非調波構造の信号成分とを含む単音のパワースペクトルを表現するモデルである。したがって k 番目の楽器、 l 番目の単音に対するパワースペクトルを $J_{kl}(t, f)$ としたした場合、この調波・非調波混合モデルは、概念的には $J_{kl}(t, f) = H_{kl}(t, f) + I_{kl}(t, f)$ と表すことができる。

10

【0016】

複数種類の単音に対応した複数のテンプレートも、この調波・非調波混合モデルを満たすものである。

【0017】

また調波・非調波混合モデルをそれぞれ構成するための複数のパラメータを含む複数のモデルパラメータを作成するために、楽譜情報データに含まれる複数の楽器ごとの複数単音の情報をパラメータ音に変換する音響変換手段と、複数のパラメータ音の複数のパワースペクトルを、それぞれ調波モデルと非調波モデルとからなる複数の調波・非調波混合モデルで表現するために、調波・非調波混合モデルをそれぞれ構成するための複数のパラメータを含む複数のモデルパラメータを作成する音モデル構成用モデルパラメータ作成部とを用いることもできる。

20

【0018】

また第1のパワースペクトル生成記憶部は、複数種類のモデルパラメータ集合データから、各時刻におけるモデルパラメータを読み出し、読み出した複数のモデルパラメータにそれぞれ含まれる複数のパラメータと予め定めた第1のモデルパラメータ変換式とを用いて、読み出した複数のモデルパラメータに対応した複数の初期パワースペクトルを生成して記憶手段に記憶する。

【0019】

なお第1のモデルパラメータ変換式としては、下記の調波・非調波混合モデルを用いることができる。

30

【0020】

$$h_{kl} = r_{kl} c (H_{kl}(t, f) + I_{kl}(t, f))$$

上記式において、 h_{kl} は単音のパワースペクトルであり、 $r_{kl} c$ は各チャンネルの相対的な強度を表すパラメータである。そして $H_{kl}(t, f)$ は音量、基本周波数 F_0 の時間変化、パワーエンベロープの概形を表現する y 番目のガウシアン加重係数、 n 次倍音成分の相対強度、オンセット時刻、音長及び周波数方向の分散という特徴量を表す複数のパラメータにより構成される調波モデルである。また $I_{kl}(t, f)$ はノンパラメトリックな関数によって表される非調波モデルである。

【0021】

40

初期分配関数計算記憶部は、まず各時刻（楽譜上において1つの単音が存在する時刻に相当する）において、第1のパワースペクトル生成記憶部に記憶された複数の初期パワースペクトルを合成して各時刻毎の合成パワースペクトルを作る。次に、初期分配関数計算記憶部は、各時刻毎の合成パワースペクトルに対する各時刻毎の複数の初期パワースペクトルの割合（比率）を示す複数の初期分配関数を計算して記憶手段に記憶する。この初期分配関数は、パワースペクトルに含まれる複数の周波数成分ごとの複数の前記割合を含むものである。初期分配関数は、パワースペクトルを構成する調波モデルと非調波モデルのそれぞれについても同様に分配することができる関数である。

【0022】

パワースペクトル分離記憶部は、第1回目の分離処理においては、各時刻の入力音響信

50

号のパワースペクトルから、各時刻における複数の初期分配関数を用いて、各時刻における複数種類の楽器に対応する複数のパワースペクトルを分離して記憶手段に記憶する。そして第2回目以降の分離処理においては、更新分配関数を用いて、各時刻における複数種類の楽器に対応する複数のパワースペクトルを分離して当該記憶手段に記憶する。

【0023】

更新モデルパラメータ推定記憶部は、各時刻において分離された複数のパワースペクトルから、それぞれ調波・非調波混合モデルで表現するために必要な複数のパラメータを含む複数の更新モデルパラメータを推定する。そして更新モデルパラメータ推定記憶部は、複数の更新モデルパラメータが集まって構成される複数種類の更新モデルパラメータ集合データを作成して記憶手段に記憶する。なお更新モデルパラメータ推定記憶部における推定処理については後に説明する。

10

【0024】

第2のパワースペクトル生成記憶部は、更新モデルパラメータ推定記憶部に記憶されている複数種類の更新モデルパラメータから、各時刻における更新モデルパラメータを読み出して、読み出した複数の更新モデルパラメータに含まれる複数のパラメータと予め定めた第2のモデルパラメータ変換式とを用いて、読み出した複数の更新モデルパラメータに対応した複数の更新パワースペクトルを生成して記憶手段に記憶する。第2のモデルパラメータ変換式としては、第1のモデルパラメータ変換式と同様の変換式を用いることができる。

【0025】

20

また更新分配関数計算記憶部は、各時刻において、第2のパワースペクトル生成記憶部に記憶された複数の更新パワースペクトルを合成して各時刻毎の合成パワースペクトルを作る。次に、更新分配関数計算記憶部は、各時刻毎の合成パワースペクトルに対する各時刻毎の複数の更新パワースペクトルの割合を示す複数の更新分配関数を計算して記憶手段に記憶する。更新分配関数も、初期分配関数と同様に、パワースペクトルを構成する調波モデルと非調波モデルのそれぞれについても同様に分配することができる関数である。

【0026】

更新モデルパラメータ推定記憶部は、パワースペクトル分離記憶部が第2回目以降の分離処理を行うたびに、更新パワースペクトルが初期パワースペクトルに近い状態からパワースペクトル分離記憶部に記憶された最新の複数のパワースペクトルに近い状態に徐々に近づくように、複数の更新モデルパラメータにそれぞれ含まれる複数のパラメータを推定するべく構成されている。そして更新パワースペクトルが初期パワースペクトルに近い状態からパワースペクトル分離記憶部に記憶された最新の複数のパワースペクトルに近い状態になるまで、パワースペクトル分離記憶部と、更新モデルパラメータ推定記憶部と、第2のパワースペクトル生成記憶部と、更新分配関数計算記憶部とが処理動作を繰り返す。したがって最終的な各単音の更新モデルパラメータに基づいて作成された更新パワースペクトルは、調波モデルと非調波モデルとを含んで構成された、入力音響信号に含まれる一つの楽器の単音のパワースペクトルに近いものとなる。よって本発明によれば、調波モデルと非調波モデルを含む形で楽器音のパワースペクトルを分離することができる。すなわち本発明によれば、入力音響信号中の楽器音に近い楽器音（音源）を分離することができる。

30

40

【0027】

更新モデルパラメータ推定記憶部におけるパラメータの推定は、コスト関数を用いて行うのが好ましい。コスト関数としては、例えば、パワースペクトル分離記憶部が分離処理を行うたびに、パワースペクトル分離記憶部に記憶された各時刻における複数のパワースペクトルと第2のパワースペクトル生成記憶部に記憶された各時刻における複数の更新パワースペクトルとの間のKLダイバージェンス $J_1 \times (\quad)$ （ \quad は0から1を満たす実数）と、第2のパワースペクトル生成記憶部に記憶された各時刻における複数の更新パワースペクトルと第1のパワースペクトル生成記憶部に記憶された各時刻における複数の初期パワースペクトルとの間のKLダイバージェンス $J_2 \times (1 - \quad)$ のすべての総和 J_0 に基

50

づくコスト関数 J を用いるのが好ましい。そしてコスト関数を最小化するように、複数の更新モデルパラメータにそれぞれ含まれる複数のパラメータを推定する。なおこの場合、更新モデルパラメータ推定記憶部は、 α を分離処理が行われるたびに増加させるように構成する。そして α が 1 になるまで、パワースペクトル分離記憶部と、更新モデルパラメータ推定記憶部と、第 2 のパワースペクトル生成記憶部と、更新分配関数計算記憶部とが処理動作を繰り返すことにより、音源の分離を行うことができる。なおこの場合には、パワースペクトル分離記憶部が第 1 回目の分離処理を行う際には、 α が 0 になる。特にこのようにして更新モデルパラメータに含まれるパラメータを推定すると、確実に且つ安定した状態で、更新モデルパラメータに含まれるパラメータを収束させることができる。

【0028】

10

このようなコスト関数を用いると、種々の制約を付けることができ、しかもパラメータの推定精度を高くすることができる。例えば、このコスト関数に、非調波モデルが調波構造を表現しないための制約条件を含めることができる。この制約条件を含めると、パラメータを推定する際に、非調波モデルが調波構造になって生じることがある誤推定の発生を確実に防止することができる。

【0029】

また調波モデルが、音高の時間変化を扱うための関数 $\mu_{kl}(t)$ を含んでいる場合には、コスト関数に、基本周波数 F_0 が時間的に不連続値をとらないための制約条件を含めることができる。このようにすると分離音が瞬間的に大きな変化をすることがなくなる。

【0030】

20

さらにコスト関数には、調波モデルについて同一楽器の単音に対して倍音成分の相対強度比を一定にするための制約条件及びまたは非調波モデルについて同一楽器の単音に対して非調波成分比を一定にするための制約条件を含めることができる。これらの制約条件を含めると、同一楽器の単音間で音感が極端に異なるようになることがなくなる。

【0031】

本発明の音源分離方法は、コンピュータを用いて以下のステップを実施する。

【0032】

(S1) 複数種類の楽器から発生した複数種類の楽器音に対応する複数の楽器音信号が含まれる入力音響信号と時間的に同期し且つ前記複数の入力音響信号に対応した複数種類の楽器によりそれぞれ演奏されるべき複数種類の楽譜についての楽譜情報データを用意するステップ。

30

【0033】

(S2) 複数種類の楽器に関して、それぞれの楽器から生じる複数種類の単音を、それぞれ調波モデルと非調波モデルとからなる複数の調波・非調波混合モデルで表現するために予め作成された、調波・非調波混合モデルをそれぞれ構成するための複数のパラメータを含む複数のモデルパラメータを用いて、複数種類の楽譜に含まれる複数の単音をそれぞれ複数のモデルパラメータに置き換えて、複数種類の楽譜に対応した、複数のモデルパラメータが集まって構成される複数種類のモデルパラメータ集合データを作成するステップ。

【0034】

40

(S3) 複数種類のモデルパラメータ集合データから、各時刻におけるモデルパラメータを読み出し、読み出した複数のモデルパラメータにそれぞれ含まれる複数のパラメータと予め定めた第 1 のモデルパラメータ変換式とを用いて、読み出した複数のモデルパラメータに対応した複数の初期パワースペクトルを生成するステップ。

【0035】

(S4) 各時刻において、複数の初期パワースペクトルを合成して各時刻毎の合成パワースペクトルを作り、各時刻毎の合成パワースペクトルに対する各時刻毎の複数の初期パワースペクトルの割合を示す複数の初期分配関数を計算するステップ。

【0036】

(S5) 第 1 回目の分離処理においては、各時刻の入力音響信号のパワースペクトルが

50

ら、各時刻における複数の初期分配関数を用いて、各時刻における複数種類の楽器に対応する複数のパワースペクトルを分離し、第2回目以降の分離処理においては更新分配関数を用いて、各時刻における複数種類の楽器に対応する複数のパワースペクトルを分離するステップ。

【0037】

(S6) 各時刻において分離された複数のパワースペクトルから、それぞれ調波・非調波混合モデルで表現するために必要な複数のパラメータを含む複数の更新モデルパラメータを推定し、複数の更新モデルパラメータが集まって構成される複数種類の更新モデルパラメータ集合データを作成するステップ。

【0038】

(S7) 複数種類の更新モデルパラメータ集合データから、各時刻における更新モデルパラメータを読み出して、読み出した複数の更新モデルパラメータにそれぞれ含まれる複数のパラメータと予め定めた第2のモデルパラメータ変換式とを用いて、読み出した複数の更新モデルパラメータに対応した複数の更新パワースペクトルを生成するステップ。

【0039】

(S8) 各時刻において、複数の更新パワースペクトルを合成して各時刻毎の合成パワースペクトルを作り、各時刻毎の合成パワースペクトルに対する各時刻毎の複数の更新パワースペクトルの割合を示す複数の更新分配関数を計算するステップ。

【0040】

(S9) 更新モデルパラメータ集合データを作成するステップにおいて、第2回目以降の分離処理を行うたびに、更新モデルパラメータを推定するステップでは、更新パワースペクトルが初期パワースペクトルに近い状態からパワースペクトル分離記憶部に記憶された最新の複数のパワースペクトルに近い状態に徐々に近づくように、複数の更新モデルパラメータにそれぞれ含まれる複数のパラメータを推定するステップ。

【0041】

(S10) 更新パワースペクトルが初期パワースペクトルに近い状態からパワースペクトルを分離するステップで分離された最新の複数のパワースペクトルに近い状態になるまで、パワースペクトルを分離するステップと、更新モデルパラメータを推定するステップと、更新パワースペクトルを生成するステップと、更新分配関数を計算するステップとを繰り返すステップ。

【0042】

また本発明の音源分離用コンピュータプログラムは、上記方法の各ステップをコンピュータにおいて実行するように構成されている。

【図面の簡単な説明】

【0043】

【図1】音源分離システムをコンピュータを用いて実現する場合の構成例を示すブロック図である。

【図2】図1のコンピュータ内に本発明の音源分離用プログラムをインストールして実現する複数の機能実現手段の関係を示すブロック図である。

【図3】音源分離用プログラムのアルゴリズムの一例を示すフローチャートである。

【図4】本実施形態に係る音源分離システムの流れを視覚的に説明するための概念図である。

【図5】本実施形態に係る音源分離システムの流れを視覚的に説明するための概念図である。

【図6】分配関数の求め方を概念的に説明するために用いる図である。

【図7】分離関数を用いた分離処理を概念的に説明するために用いる図である。

【図8】本発明で採用されるモデルパラメータ反復推定処理の一例の大まかな手順をフローチャートで表したものである。

【図9】各楽器パートのSNR (Signal to Noise Ratio) を楽曲ごとに平均したものと、全楽曲、全パートについてSNRを平均した結果を示す一覧表である。

10

20

30

40

50

【発明を実施するための最良の形態】

【0044】

以下、本発明を実施するための最良の形態（以下「実施形態」という。）を詳細に説明する。

【0045】

図1は、本発明の実施形態の音源分離システムをコンピュータ10を用いて実現する場合の構成例を示すブロック図である。コンピュータ10は、CPU（Central Processing Unit）11と、DRAM等のRAM（Random Access Memory）12と、ハードディスクドライブ（以下、「ハードディスク」という。）その他の大容量記憶手段13と、フレキシブルディスクドライブ又はCD-ROMドライブ等の外部記憶部14と、LAN（Local Area Network）又はインターネットなどの通信ネットワーク20との通信を行う通信部18とを備える。また、コンピュータ10は、キーボード又はマウス等の入力部15と、液晶ディスプレイ等の表示部16とを備えている。さらに、コンピュータ10は、MIDI音源等の音源17を搭載している。

10

【0046】

CPU11は、後述するパースペクトルの分離処理と更新モデルパラメータのパラメータの推定（モデル適応）処理を行うための各ステップを実行する計算手段として動作する。

【0047】

音源17は、後述する入力音響信号を備えている。また音源分離を行うための入力音響信号と時間的に同期した標準MIDIファイル（Standard MIDI File、以下「SMF」という。）を楽譜情報データとして備えている。SMFはCD-ROM等や、通信ネットワーク20を介してハードディスク13に記録される。なお、「時間的に同期した」とは、SMF中の各楽器パートの単音（楽譜の音符に相当するもの）のオンセット時刻（発音時刻）と音長が実際の入力楽曲の音響信号における各楽器パートの単音と完全に同期しているということである。

20

【0048】

なお、MIDI信号の記録、編集、再生などは、シーケンサ、またはシーケンスソフトウェア（図示を省略する）で行われる。ここでMIDI信号は、MIDIファイルとして扱われる。SMFとは、MIDI音源の演奏データを記録するための基本ファイルフォーマットである。SMFは、異なるシーケンサ、あるいはシーケンスソフトウェア間でのMIDIファイルの互換性を保つ統一規格である"チャンク"と呼ばれるデータ単位によって構成されている。SMFフォーマット化されたMIDIファイルデータのイベントとしては、大きく分けて、ミディ・イベント（MIDI Event）、システム・エクスクルーシブ・イベント（SysEx Event）、そしてメタ・イベント（Meta Event）の3種類がある。ミディ・イベントには、演奏データそのものが示されている。システム・エクスクルーシブ・イベントには、主にMIDIのシステム・エクスクルーシブ・メッセージが示されている。システム・エクスクルーシブ・メッセージは、特定の楽器にしかない情報を交換したり、特別の非音楽情報、イベント情報などを伝達するために用いられる。メタ・イベントには、テンポや拍子などの演奏全体の情報や、シーケンサやシーケンスソフトが利用する歌詞や著作権情報などの付加的な情報が示されている。全てのメタ・イベントは0xFFで始まり、次にイベント・タイプを表すバイトが続き、更にデータ長及びデータ自体が続く。MIDI演奏プログラムは、自分が認識できないメタ・イベントを無視するように設計されている。また、各イベントには、そのイベントを実行する時間的タイミングに関するタイミング情報が付加されている。このタイミング情報は、直前のイベントの実行からの時間差で示されている。例えば、このタイミング情報が"0"のときは、直前のイベントと同時にこのタイミング情報が付加されたイベントが実行されることになる。

30

40

【0049】

一般に、MIDI規格を用いた音楽再生では、各種信号、楽器固有の音色をモデル化し、そのデータを格納した音源を各種パラメータで制御するシステムをとっている。そして

50

、S M Fの各トラックはそれぞれの楽器パートに対応しており、各楽器パートとこの分離信号が含まれている。また、S M Fには、音高とオンセット時刻と音長又はオフセット時刻と楽器ラベルなどの情報が含まれている。

【0050】

したがって、S M Fが与えられていれば、これをM I D I音源で演奏することによって、入力される音響信号中の各単音にある程度近い音のサンプル（これを、「テンプレート音」という。）を生成することができる。テンプレート音から、ある楽器から生じる単音に対応した標準的なパワースペクトルにより表現されたデータのテンプレートを作成することができる。

【0051】

テンプレート音またはテンプレートは、実際の入力音響信号の単音または単音のパワースペクトルとは完全には同一でなく、必ず音響的な違いがある。そのためテンプレート音またはテンプレートをそのまま分離音または分離のためのパワースペクトルとして使用することはできない。しかし後に詳しく説明するように、単音の更新パワースペクトルが後述する初期パワースペクトルに近い状態から、入力音響信号から分離した単音の最新のパワースペクトルに近い状態に徐々に近付ける学習（これを「モデル適応」という。）を行うことにより、更新モデルパラメータに含まれる複数のパラメータを最終的に望ましい形で収束させることができれば、正解又はそれに近い分離音と推定されることになる。

【0052】

また、S M Fのトラックを利用することで分離後の音響信号がどれだけ合成前の音響信号に近いかという定量的な評価が可能になる。

【0053】

図2は、図1のコンピュータ10内に本発明の音源分離用プログラムをインストールして実現する複数の機能実現手段の関係を示すブロック図である。また図3は、音源分離用プログラムのアルゴリズムの一例を示すフローチャートである。さらに図4及び図5は、本実施形態に係る音源分離システムの流れを視覚的に説明するための概念図である。まず図1乃至図5を用いて、音源分離システムの基本的な構成について説明し、後に理論的な説明をする。

【0054】

本実施形態の音源分離システムは、入力音響信号記憶部101と、入力音響信号パワースペクトル作成記憶部102と、楽譜情報データ記憶部103と、モデルパラメータ作成記憶部104と、モデルパラメータ集合データ作成記憶部106と、第1のパワースペクトル生成記憶部108と、初期分配関数計算記憶部110と、パワースペクトル分離記憶部112と、更新モデルパラメータ推定記憶部114と、第2のパワースペクトル生成記憶部116と、更新分配関数計算記憶部118とを備えている。

【0055】

入力音響信号記憶部101は、複数種類の楽器から発生した複数種類の楽器音に対応する複数の楽器音信号が含まれる入力音響信号（混合音信号）を記憶する。入力音響信号は、音楽を再生する目的と、パワースペクトルを得る目的のために用意されている。また入力音響信号パワースペクトル作成記憶部102は、入力音響信号からパワースペクトルを作成して記憶する。図4及び図5には、入力音響信号から得たパワースペクトルAの一例が示されている。このパワースペクトルは、横軸が時間で縦軸が周波数である。図4及び図5の例では、複数の時刻における複数のパワースペクトルが横に並ぶように表示されている。

【0056】

楽譜情報データ記憶部103は、入力音響信号と時間的に同期し且つ複数の入力音響信号に対応した複数種類の楽器によりそれぞれ演奏されるべき複数種類の楽譜についての楽譜情報データを記憶している。図4及び図5には、楽譜情報データBを理解しやすいように実際の楽譜として図示してある。本実施形態では、楽譜情報データBとしては、前述の標準M I D Iファイル（S M F）を利用する。

10

20

30

40

50

【 0 0 5 7 】

モデルパラメータ作成記憶部 1 0 4 は、複数種類の楽器に関して、それぞれの楽器から生じる複数種類の単音それぞれについて、調波モデルと非調波モデルとからなる複数の調波・非調波混合モデルで表現する（構成する）ための複数のパラメータを含むモデルパラメータを作成して、記憶手段 1 0 5 に記憶する。これらモデルパラメータを作成するために、本実施形態では、楽譜情報データ B に含まれている楽器パートで使用される複数種類の楽器に関して、それぞれの楽器から生じる複数種類の単音（楽器から生じる全ての単音）に対応した複数の標準的なパワースペクトルにより表された複数のテンプレートをを用いて、複数種類の単音についての複数のモデルパラメータを作成する。

【 0 0 5 8 】

モデルパラメータ集合データ作成記憶部 1 0 6 では、モデルパラメータ作成記憶部 1 0 4 の記憶手段 1 0 5 に記憶された、調波・非調波混合モデルをそれぞれ構成するための複数のパラメータを含んで構成されたモデルパラメータを用いて、複数種類の楽譜に含まれる複数の単音をそれぞれ複数のモデルパラメータに置き換える。そしてモデルパラメータ集合データ作成記憶部 1 0 6 は、複数種類の楽譜に対応した、複数のモデルパラメータが集まって構成される複数種類のモデルパラメータ集合データを作成し、これを記憶手段 1 0 7 に記憶する。

【 0 0 5 9 】

なお後に説明する他の実施の形態では、MIDIファイルの楽譜情報データを音響変換手段で音に変換したテンプレート音に基づいて、モデルパラメータを作成する。前述のように、テンプレート音とは、楽譜に基づいてMIDI音源で生成する各単音のサンプルである。そしてテンプレートとは、ある種類の楽器が発することができる複数種類の単音（異なる音階の複数種類の単音）を、それぞれ標準的なパワースペクトルで表現したものである。図 5 には、テンプレートは示されていないが、各単音についての各テンプレートは、図 5 の出力に示した「分離音」の文字の下に示した複数のパワースペクトルの図と同様に、時間軸と周波数軸とを有するパワースペクトルとして表される。具体的には、例えば、標準的なギターの「ド」の音を標準的なパワースペクトルで表現したものが、テンプレートである。このギターの「ド」の単音のテンプレートのパワースペクトルは、入力音響信号中に含まれているギターの楽器音信号中の「ド」の単音のパワースペクトルとは、ある程度近似しているが、同じではない。

【 0 0 6 0 】

ここで調波・非調波混合モデルは、時刻 t 、周波数 f 、 k 番目の楽器、 l 番目の単音に対して、調波構造を表現する調波モデル $H_{kl}(t, f)$ と非調波構造を表現する非調波モデル $I_{kl}(t, f)$ の線形和として定義される。この調波・非調波混合モデルは、一つのモデルで調波構造の信号成分と非調波構造の信号成分とを含む単音のパワースペクトルを表現するモデルである。したがって k 番目の楽器、 l 番目の単音に対するパワースペクトルを $J_{kl}(t, f)$ としたした場合、この調波・非調波混合モデルは、 $J_{kl}(t, f) = H_{kl}(t, f) + I_{kl}(t, f)$ と表すことができる。本実施形態では、複数種類の単音に対応した複数のテンプレートは、この調波・非調波混合モデルを構成するための複数のパラメータによって構成されるモデルパラメータに変換される。このモデルパラメータは、単音の音モデルとも呼ばれる。モデルパラメータを音モデルとしてイメージ的に示すと、図 5 中の「中間表現」の文字の下に示した「音モデル」の文字の下に示した複数の図のようになる。前述のモデルパラメータ作成記憶部 1 0 4 の記憶手段 1 0 5 には、複数種類の楽器に関して、それぞれ複数種類の単音に対応する複数のモデルパラメータが記憶されている。

【 0 0 6 1 】

前述のモデルパラメータ集合データ作成記憶部 1 0 6 の記憶手段 1 0 7 には、図 4 に示すように、複数種類の楽譜または楽器に対応して、複数のモデルパラメータ ($MP11 \sim MP1l$) \sim ($MPk1 \sim MPkl$) がそれぞれ集まって構成されたモデルパラメータ集合データ $MPD1 \sim MPDk$ が記憶されている。図 4 は、対応する楽譜の一つの単音が一つのモデ

10

20

30

40

50

ルパラメータ（音モデル）で表現されていることを、一つのモデルパラメータを一枚のシートに模して表示している。

【 0 0 6 2 】

第1のパワースペクトル生成記憶部108は、図4に示すように、複数種類のモデルパラメータ集合データMPD1～MPDkから、各時刻におけるモデルパラメータ（MP11～MP1l）～（MPk1～MPkl）を読み出す。そして第1のパワースペクトル生成記憶部108は、読み出した複数のモデルパラメータにそれぞれ含まれる複数のパラメータと予め定めた第1のモデルパラメータ変換式とを用いて、読み出した複数のモデルパラメータに対応した複数の初期パワースペクトル（PS11～PS1l）～（PSk1～PSkl）を生成して記憶手段109に記憶する。

10

【 0 0 6 3 】

なお第1のパワースペクトル生成記憶部108で使用する第1のモデルパラメータ変換式としては、下記の調波・非調波混合モデルを用いることができる。

【 0 0 6 4 】

$$h_{kl} = r_{klc} (H_{kl}(t, f) + I_{kl}(t, f))$$

上記式において、 h_{kl} はパワースペクトルであり、 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る。そして $H_{kl}(t, f)$ は音量、基本周波数 F_0 の時間変化、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 n の重み係数、 n 次倍音成分の相対強度、オンセット時刻、音長及び周波数方向の分散という特徴量を表す複数のパラメータにより構成される調波モデルである。また $I_{kl}(t, f)$ はノンパラメトリックな関数によって表される非調波モデルである。この調波モデルの複数のパラメータと非調波モデルの関数とが、モデルパラメータにそれぞれ含まれる複数のパラメータである。

20

【 0 0 6 5 】

初期分配関数計算記憶部110は、図6に示すように、まず各時刻において、第1のパワースペクトル生成記憶部108の記憶手段109に記憶された複数の初期パワースペクトル（例えばPS11, PS21, …, PSk1）を合成して各時刻毎の合成パワースペクトルTPS（例えばPS11+PS21+…+PSk1）を作る。次に、初期分配関数計算記憶部110は、各時刻毎の合成パワースペクトルTPSに対する各時刻毎の複数の初期パワースペクトルの割合（比率）{例えば、[PS11/TPS]}を示す複数の初期分配関数（DF11～DFkl）を計算して記憶手段111に記憶する。図4には、初期パワースペクトルと初期分配関数とを、1つのシート図で兼用して示してある。記憶手段111に記憶されている複数の初期分配関数の数は、時刻の数（単音の数 l の最大値）に楽器の数 k または楽譜の種類 n の数を掛けた数に等しい。そして初期分配関数は、図6に示すように、パワースペクトルに含まれる複数の周波数成分ごとの複数の割合 $R_1 \sim R_9$ を含むものである。

30

【 0 0 6 6 】

パワースペクトル分離記憶部112は、図7に示すように、第1回目の分離処理において、各時刻の入力音響信号のパワースペクトルA1から、各時刻における複数の初期分配関数（例えば、DF11～DFk1）を用いて、各時刻における複数種類の楽器に対応する複数のパワースペクトルPS11～PSk1を分離して記憶手段113に記憶する。すなわちパワースペクトル分離記憶部112は、第1回目の分離処理において、入力音響信号のパワースペクトルA1に初期分配関数（例えば、DF11～DFk1）を乗算することにより、各時刻における複数種類の楽器に対応する複数のパワースペクトル（一単音のパワースペクトル）PS11～PSk1をそれぞれ分離する。後に説明するように、パワースペクトル分離記憶部112は、2回目以降の分離処理では、更新分配関数を用いてパワースペクトルの分離処理を行う。

40

【 0 0 6 7 】

更新モデルパラメータ推定記憶部114は、図4に示すように、各時刻において分離された複数種類の楽器に対応する複数のパワースペクトルPS11～PSk1それぞれを調波・非調波混合モデルで表現するために必要な複数のパラメータを含む複数の更新モデ

50

ルパラメータ (MP11 ~ MPk1) を推定する。図4においては、分離したパワースペクトルと更新モデルパラメータとを、1つのシート図で兼用して示してある。そして更新モデルパラメータ推定記憶部114は、複数の更新モデルパラメータが集まって構成される複数種類の更新モデルパラメータ集合データMPD1 ~ MPDkを作成して記憶手段115に記憶する。なお更新モデルパラメータ推定記憶部114における推定処理については後で説明する。図5では、これら最初のモデルパラメータMP11 ~ MPk1または更新モデルパラメータMP11 ~ MPk1によって表現される音モデルを「中間表現」として示してある。そして図5では、各時刻において分離された複数種類の楽器に対応する複数のパワースペクトルデータPS11 ~ PSk1から複数のパラメータから構成される更新モデルパラメータ(MP11 ~ MPk1)を推定することを「パラメータの推定」と記載してある。

10

【0068】

図2に戻って、第2のパワースペクトル生成記憶部116は、更新モデルパラメータ推定記憶部114の記憶手段115に記憶されている複数種類の更新モデルパラメータ集合データから、各時刻における更新モデルパラメータ(MP11 ~ MPk1)を読み出して、読み出した複数の更新モデルパラメータに含まれる複数のパラメータと予め定めた第2のモデルパラメータ変換式とを用いて、読み出した複数の更新モデルパラメータ(MP11 ~ MPk1)に対応した複数の更新パワースペクトル(PS11" ~ PSk1" : 図示せず)を生成して記憶手段117に記憶する。第2のモデルパラメータ変換式としては、第1のモデルパラメータ変換式と同様の変換式を用いることができる。

20

【0069】

また更新分配関数計算記憶部118は、初期分配関数計算記憶部110における計算と同様にして更新分配関数を計算する。すなわち更新分配関数計算記憶部118は、各時刻において、第2のパワースペクトル生成記憶部116に記憶された複数の更新パワースペクトル(PS11" ~ PSk1" : 図示せず)を合成して各時刻毎の合成パワースペクトルTPSを作る。そして次に、更新分配関数計算記憶部118は、各時刻毎の合成パワースペクトルTPSに対する各時刻毎の複数の更新パワースペクトルの割合(例えばPS11"/TPS)を示す複数の更新分配関数(DF11 ~ DFk1 : 図示せず)を計算して記憶手段119に記憶する。更新分配関数(DF11 ~ DFk1)も、初期分配関数(DF11 ~ DFk1)と同様に、パワースペクトルを構成する調波モデルと非調波モデルのそれぞれについても同様に分配することができる関数である。

30

【0070】

次に、更新モデルパラメータ推定記憶部114における推定処理について説明する。更新モデルパラメータ推定記憶部114は、パワースペクトル分離記憶部112が第2回目以降の分離処理を行うたびに、更新パワースペクトル(PS11" ~ PSk1" : 図示せず)が初期パワースペクトルに近い状態からパワースペクトル分離記憶部112の記憶手段113に記憶された最新の複数のパワースペクトルに近い状態に徐々に近づくように、複数の更新モデルパラメータ(MP11 ~ MPk1)にそれぞれ含まれる複数のパラメータを推定するように構成されている。そして更新パワースペクトル(PS11" ~ PSk1")が初期パワースペクトル(PS11 ~ PSk1)に近い状態からパワースペクトル分離記憶部112の記憶手段113に記憶された最新の複数のパワースペクトル(PS11 ~ PSk1)に近い状態になるまで、パワースペクトル分離記憶部112と、更新モデルパラメータ推定記憶部114と、第2のパワースペクトル生成記憶部116と、更新分配関数計算記憶部118とが処理動作を繰り返す。したがって最終的な各単音の更新モデルパラメータ(MP11 ~ MPk1)に基づいて作成された更新パワースペクトル(PS11" ~ PSk1")は、調波モデルと非調波モデルとを含んで構成された、入力音響信号に含まれる一つの楽器の単音のパワースペクトルに近いものとなる。

40

【0071】

なお後に詳しく説明するように、更新モデルパラメータ推定記憶部114における更新モデルパラメータのパラメータの推定は、コスト関数を用いて行うのが好ましい。コスト

50

関数としては、例えば、パワースペクトル分離記憶部 112 が分離処理を行うたびに、パワースペクトル分離記憶部 112 の記憶手段 113 に記憶された各時刻における複数のパワースペクトル ($PS_{11} \sim PS_{k1}$) と第 2 のパワースペクトル生成記憶部 116 の記憶手段 117 に記憶された各時刻における複数の更新パワースペクトル ($PS_{11}'' \sim PS_{k1}''$) との間の KL ダイバージェンス $J_{1 \times}$ (α は $0 < \alpha < 1$ を満たす実数) と、第 2 のパワースペクトル生成記憶部 116 の記憶手段 117 に記憶された各時刻における複数の更新パワースペクトル ($PS_{11}'' \sim PS_{k1}''$) と第 1 のパワースペクトル生成記憶部 108 の記憶手段 119 に記憶された各時刻における複数の初期パワースペクトル ($PS_{11} \sim PS_{k1}$) との間の KL ダイバージェンス $J_{2 \times}$ ($1 - \alpha$) のすべての総和 J_0 に基づくコスト関数 J を用いるのが好ましい。そしてコスト関数 J を最小化するように、複数の更新モデルパラメータ ($MP_{11} \sim MP_{k1}$) にそれぞれ含まれる複数のパラメータを推定する。なおこの場合、更新モデルパラメータ推定記憶部 114 は、 α を分離処理が行われるたびに増加させるように構成する。そして α が 1 になるまで、パワースペクトル分離記憶部 112 と、更新モデルパラメータ推定記憶部 114 と、第 2 のパワースペクトル生成記憶部 116 と、更新分配関数計算記憶部 118 とが処理動作を繰り返すことにより、音源の分離を行うことができる。なおこの場合には、パワースペクトル分離記憶部 112 が第 1 回目の分離処理を行う際には、 α が 0 になる。特にこのようにして更新モデルパラメータ ($MP_{11} \sim MP_{k1}$) に含まれるパラメータを推定すると、確実に且つ安定した状態で、更新モデルパラメータ ($MP_{11} \sim MP_{k1}$) に含まれるパラメータを収束させることができる。

【0072】

図 3 には、上記実施の形態及び本発明の方法をコンピュータを用いて実施する場合に用いるコンピュータプログラムのアルゴリズムの一例を示してある。このアルゴリズムでは、ステップ S1 で、複数種類の楽器から発生した複数種類の楽器音に対応する複数の楽器音信号と、入力音響信号と時間的に同期し且つ複数の入力音響信号に対応した複数種類の楽器によりそれぞれ演奏されるべき複数種類の楽譜についての楽譜情報データを用意する。そしてステップ S2 では、複数種類の楽器に関して、それぞれの楽器から生じる複数種類の単音を、それぞれ調波モデルと非調波モデルとからなる複数の調波・非調波混合モデルで表現するために予め作成された、調波・非調波混合モデルをそれぞれ構成するための複数のパラメータを含む複数のモデルパラメータを用意する。そしてこれらのモデルパラメータを用いて、複数種類の楽譜に含まれる複数の単音をそれぞれ複数のモデルパラメータ ($MP_{11} \sim MP_{11}$) \sim ($MP_{k1} \sim MP_{k1}$) に置き換えて、複数種類の楽譜に対応した、複数のモデルパラメータ ($MP_{11} \sim MP_{11}$) \sim ($MP_{k1} \sim MP_{k1}$) が集まって構成される複数種類のモデルパラメータ集合データ $MPD_1 \sim MPD_k$ を作成する。そしてステップ S3 では、複数種類のモデルパラメータ集合データ $MPD_1 \sim MPD_k$ から、各時刻におけるモデルパラメータを読み出し、読み出した複数のモデルパラメータ ($MP_{11} \sim MP_{k1}$) にそれぞれ含まれる複数のパラメータと予め定めた第 1 のモデルパラメータ変換式とを用いて、読み出した複数のモデルパラメータ ($MP_{11} \sim MP_{k1}$) に対応した複数の初期パワースペクトル $PS_{11} \sim PS_{k1}$ を生成する。次にステップ S4 では、各時刻において、複数の初期パワースペクトルを合成して各時刻毎の合成パワースペクトルを作る。そして各時刻毎の合成パワースペクトルに対する各時刻毎の複数の初期パワースペクトルの割合を示す複数の初期分配関数 ($DF_{11} \sim DF_{k1}$) を計算する。次にステップ S5 では、第 1 回目の分離処理においては、各時刻の入力音響信号のパワースペクトルから、各時刻における複数の初期分配関数 ($DF_{11} \sim DF_{k1}$) を用いて、各時刻における複数種類の楽器に対応する複数のパワースペクトル $PS_{11} \sim PS_{k1}$ を分離する。そして第 2 回目以降の分離処理においては更新分配関数 ($DF_{11} \sim DF_{k1}$) を用いて、各時刻における複数種類の楽器に対応する複数のパワースペクトルを分離する。次にステップ S6 では、各時刻において分離された複数のパワースペクトル $PS_{11} \sim PS_{k1}$ から、それぞれ調波・非調波混合モデルで表現するために必要な複数のパラメータを含む複数の更新モデルパラメータ ($MP_{11} \sim MP_{k1}$) を推定するためのコスト関数 J を決定する。そしてステ

10

20

30

40

50

ステップ S 7 では、コスト関数を最小化するように、複数のモデルパラメータ (MP11 ~ MPk1) を推定する。次にステップ S 8 では、複数の更新モデルパラメータ (MP11 ~ MPk1) が集まって構成される複数種類の更新モデルパラメータ集合データ MPD1 ~ MPDk を作成する。第 1 回目の分離処理の推定では、 α は 0 となる。2 回目以降の分離処理から α が増分する。そしてステップ S 9 では、 β に α を加える。 β の値は、分離処理回数を何回行うかによって定まる。分離精度上げるためには、 β は小さいほうが好ましい。さらにステップ S 10 では、複数種類の更新モデルパラメータ集合データから、各時刻における更新モデルパラメータ (MP11 ~ MPk1) を読み出して、読み出した複数の更新モデルパラメータ (MP11 ~ MPk1) に含まれる複数のパラメータと予め定めた第 2 のモデルパラメータ変換式とを用いて、読み出した複数の更新モデルパラメータ (MP11 ~ MPk1) に対応した複数の更新パワースペクトル (PS11" ~ PSk1") を生成する。次にステップ S 11 では、各時刻において、複数の更新パワースペクトル (PS11" ~ PSk1") を合成して各時刻毎の合成パワースペクトルを作り、各時刻毎の合成パワースペクトルに対する各時刻毎の複数の更新パワースペクトル (PS11" ~ PSk1") の割合を示す複数の更新分配関数 (DF11 ~ DFk1) を計算する。ステップ S 12 で α が 1 であるか否かの判断が行われ、 α が 1 で無ければ、ステップ S 5 へとジャンプする。更新パワースペクトルが初期パワースペクトルに近い状態からパワースペクトルを分離するステップで分離された最新の複数のパワースペクトルに近い状態になるまで、パワースペクトルを分離するステップ S 5 と、更新モデルパラメータを推定するステップ S 6 からステップ S 9 と、更新パワースペクトルを生成するステップ S 10 と、更新分配関数を計算するステップ S 11 とが繰り返される。そして α が 1 になると、終了する。

【0073】

以下に示す (1) ~ (4) において、本実施の形態に係る音源分離システム及び方法を実現するために利用する事項について詳細に説明する。

【0074】

(1) 楽譜情報の利用

音源分離とは広義には混合音中に含まれる音響信号がいかなる音源 (楽器音信号) の組み合わせであるのかを推定した上で、これを分離することである。本来的には、混合音から各音源 (楽器音信号) を分離して抽出するステップと、分離された音源 (楽器音信号) がいかなる楽器による演奏かを推定する音源推定ステップとが含まれることになる。後者のステップは、楽器音認識技術と呼ばれる分野に属するものである。楽器音認識技術は、多重奏の音響信号を入力信号としたときであれば、例えば、ピアノとフルートとバイオリンの三重奏である、といった演奏楽曲中の音源を推定する技術によって実現される。

【0075】

しかしながら、現在の楽器音認識技術は未だあまり成熟していない。最新の研究でもせいぜい調波構造のみからなる 4 和音程度の混合音を認識対象としたものに留まっている。音源数が増えると楽器音の認識はさらに困難となる。

【0076】

そこで、本発明では、音源の分離精度を高めるため、事前に楽器ラベルと各楽器パートごとの音符情報を含んだ楽譜情報 (これを、本明細書では「楽譜情報データ」という。) が与えられるとの前提条件を設けることとした。楽譜情報を事前知識として用いることで、後述するように様々な制約を考慮した音源分離が可能となる。

【0077】

(2) 調波・非調波混合モデルの定式化

時刻 t 、周波数 f 、 k 番目の楽器、 l 番目の単音における調波モデルと非調波モデルを統合した「調波・非調波混合モデル hkl 」 (パワースペクトル) を、調波構造を表現するモデル $Hkl(t, f)$ と非調波構造を表現するモデル $Ikl(t, f)$ との線形和で、次式 (1) により定義する。

【数 1】

$$h_{kl} = r_{klc} (H_{kl}(t, f) + I_{kl}(t, f)) \quad (1)$$

【0078】

上記式(1)において、 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。

【数 2】

$$\sum_c r_{klc} = 1 \quad 10$$

【0079】

上記(1)式において、調波モデル $H_{kl}(t, f)$ は、音高を持つ楽器音の調波構造を表現するパラメトリックモデル(パラメータによって表現されるモデル)に基づいて定義される。すなわち調波モデル $H_{kl}(t, f)$ は、音量及び基本周波数(F_0)の時間変化、オンセット時刻、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等といった特徴量を表すパラメータにより表現される。

【0080】

本実施形態では、調波モデルとして調波時間構造化クラスタリング(Harmonic-Temporal-structured Clustering: H T C)で用いられる音源モデル(以下、「H T C音源モデル」という。)で用いる複数のパラメータをベースとして調波モデルを構成している。しかしこの音源モデルは、基本周波数 F_0 の軌跡 $\mu_{kl}(t)$ が時刻 t に関する多項式として定義されているために、柔軟な音高の時間変化を扱うことができない。そこで、本実施形態では、さらに柔軟に音高の時間変化を扱うために、H T C音源モデルを修正して、以下の式(2)乃至(4)を満たすように、軌跡 $\mu_{kl}(t)$ をノンパラメトリック関数として定義して自由度を高めている。

【数 3】

$$H_{kl} = \sum_{y=0}^{Y-1} \sum_{n=1}^N w_{kly} E_{kly} F_{kln} \quad (2) \quad 30$$

$$E_{kly} = \frac{u_{kly}}{\sqrt{2\pi\phi_{kl}}} e^{-\frac{(t-\tau_{kl}-y\phi_{kl})^2}{2\phi_{kl}^2}} \quad (3)$$

$$F_{kln} = \frac{v_{kln}}{\sqrt{2\pi\sigma_{kl}}} e^{-\frac{(f-n\mu_{kl}(t))^2}{2\sigma_{kl}^2}} \quad (4) \quad 40$$

【0081】

上記式(2)において、 w_{kly} は調波成分の重みを表すパラメータであり、 E_{kly} はパワーエンベロープの時間変化を、 F_{kln} は各時刻または各時刻の調波構造をそれぞれ表している。また、 E_{kly} 、 F_{kln} は上記式(3)及び(4)のように表される。なお E_{kly} は、実際には、 $E_{kly}(t)$ と、 F_{kln} は $F_{kln}(t)$ と表示されるものがあるが、ここでは (t) を便宜上省略してある。

【 0 0 8 2 】

また、上記調波モデルのパラメータを表 1 に示す。なお表 1 に示された、複数のパラメータは、後述するモデルパラメータ、更新モデルパラメータを構成する複数のパラメータの主たるものである。

【表 1】

記号	意味	
w_{kl}	調波構造モデル全体の音量	10
$\mu_{kl}(t)$	F0 の軌跡	
u_{kly}	パワーエンベロープの概形を表現する y 番目のガウシアン加重係数 $\sum_y u_{kly} = 1$ 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)	20
σ_{kl}	周波数方向の分散	

【 0 0 8 3 】

一方、非調波モデルは、ノンパラメトリック関数として定義される。そのため非調波モデルは、パワースペクトルで直接表現される。非調波モデルは、バスドラムやスネアドラムなどの非調波的な音（パワースペクトルにおいて個別の周波数成分を明確に特定できない音）を表現する。また、上述のとおりピアノやギターなどの調波構造を持つ楽器音であっても、発音時の音には弦をハンマーで叩く音や弦を引く音などのような非調波成分を含んでいる。そこで本実施形態では、このような非調波成分も非調波モデルによって表現する。

【 0 0 8 4 】

そして、本実施形態では、以上のようにして定式化した調波・非調波混合モデルを構成する複数のパラメータを含むモデルパラメータを、望ましい形で収束させる必要がある。換言すると、各楽器パートの全ての単音に対して単音に対応する調波・非調波混合モデルを構成する複数のパラメータを含むモデルパラメータを推定するために、本実施の形態では、以下のようにモデルパラメータに含まれる複数のパラメータを推定するために用いる後述するコスト関数 [後に説明する式 (2 1) に示される関数] に以下の制約条件を付加している。

【 0 0 8 5 】

(3) 調波・非調波混合モデルのモデルパラメータに対する種々の制約条件の設定

本実施形態におけるモデルパラメータに対する制約条件は、大きく分けて 3 つからなる。なお、以下に示す制約条件は、後述するコスト関数 J [式 (2 1)] に加算されて全体のコストを増大させる要素となるものである。そしてこれらはコスト関数 J を最小化することに対する制約条件として作用するものである。

【 0 0 8 6 】

[第 1 の条件] : 基本周波数 F_0 の連続性に関する条件

既に述べたように、上述した (2) の調波・非調波混合モデルに含まれる調波モデルは、柔軟に音高の時間変化を扱うために、ノンパラメトリック関数として $\mu_{kl}(t)$ を含むものとして定義される。その結果、基本周波数 F_0 が時間的に不連続に変化することが可能となるという問題が生じる。

10

20

30

40

50

【 0 0 8 7 】

この問題を解決するため、基本周波数 F_0 の不連続な変化を一定条件の下で禁止する制約、具体的には、以下の式 (5) で与えられる制約を後に説明するコスト関数 J [式 (21)] に示すように付加するのが好ましい。

【数 4】

$$\beta_{\mu} \int \left(\bar{\mu}_{kl}(t) \log \frac{\bar{\mu}_{kl}(t)}{\mu_{kl}(t)} - (\bar{\mu}_{kl}(t) - \mu_{kl}(t)) \right) dt \quad (5)$$

10

【 0 0 8 8 】

上記 (5) 式中の μ は係数である。また上記式中の μ の上に $-$ が付いた関数 (以下 μ - kl (t) とする) は、基本周波数 F_0 を更新する際に $\mu_{kl}(t)$ をガウシアンフィルタで時間方向に平滑化したもので、現在の F_0 を周波数方向に平滑化する作用を持つ。この制約は、 $\mu_{kl}(t)$ を μ - kl (t) に近づける作用を持っている。基本周波数が不連続な変化するとは、基本周波数 F_0 が変わる際に、大きな変化をすることを意味する。

【 0 0 8 9 】

[第2の条件]: 非調波モデルに対する条件

上述した (2) の調波・非調波混合モデルに含まれる非調波モデルは入力パワースペクトルで直接表現されたものである。そのため非常に大きな自由度を持つ。その結果、調波・非調波混合モデルを用いた場合には、入力パワースペクトルから分離した複数のパワースペクトルの多くが非調波モデルだけで表現される可能性がある。すなわち、後述の (4) において説明する更新モデルパラメータの反復推定処理の結果、調波モデルを含む混合音に含まれる複数の楽器音を示す楽器音信号を非調波モデルで表現してしまうという問題が生じることがある。

20

【 0 0 9 0 】

そこで、この問題を解決するため、以下の式 (6) で与えられる制約を後に説明するコスト関数 J [式 (21)] に示すように付加するのが好ましい。

【数 5】

$$\beta_{I2} \iint \left(\bar{I}_{kl} \log \frac{\bar{I}_{kl}}{I_{kl}} - (\bar{I}_{kl} - I_{kl}) \right) dt df \quad (6)$$

30

【 0 0 9 1 】

上記式で I_2 は、係数である。そして上記式において、 I の上に $-$ を付けた関数を以下 I - kl とする。この I - kl は、 I_{kl} をガウシアンフィルタで周波数方向に平滑化したものである。この制約は、 I_{kl} を、 I - kl に近づける作用を持っている。このような制約を付けることにより、調波モデルが非調波モデルだけで表現される可能性がなくなる。

【 0 0 9 2 】

[第3の条件]: 調波・非調波混合モデルに関する一般的制約 (同一楽器間での音色の一貫性に関する制約)

ある楽器が楽譜上では同じ基本周波数 F_0 や音長で表現されていても、奏法やピブラートなどの違いにより、音響信号上には違いが生じる。そのため、単音ごとに調波・非調波混合モデルを用いてモデル化 (複数のパラメータからなるモデルパラメータで表現) する必要がある。しかし、ある楽器のある音をその楽器の他の音 (楽器音) と比べた場合、同じ楽器の複数の音には何らかの一貫性 (すなわち、同じ楽器の複数の音は同じような性質を持っているということ) がある。しかし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。換言すると、分離処理を行って得たパワースペクトルから更新モデルパラメータを推定した際に、推定される更新モデルパラメータを構成する複数のパラメータ

40

50

は、同じの楽器の複数の音では類似しており、かつ各単音ごとに少しずつ異なっているという、同じ楽器の複数の音が有する一貫性を満たしている必要がある。

【0093】

そこで、同一の楽器による演奏においては複数の楽器音の一貫性を保持しつつ微小な違いを許容するための制約条件を調波モデルと非調波モデルの両方に対して加えるために、下記の式を後に説明するコスト関数 J [式(21)] に示すように付加するのが好ましい。

【0094】

(3-1: 調波モデルに対する同一楽器における複数の音モデル間での条件)

同一楽器間における調波モデルに対する制約は、具体的には以下の式(7)で与えられる。 10

【数6】

$$\beta_v \sum_n \left(\bar{v}_{kn} \log \frac{\bar{v}_{kn}}{v_{kln}} - (\bar{v}_{kn} - v_{kln}) \right) \quad (7)$$

【0095】

上記式で v は係数である。また v の上に $-$ が付いた関数を以下 $v-kn$ と言う。ここで、 $v-k,n$ は、 n 次倍音成分の相対強度 v_{kln} を同一楽器の複数の音モデルに関して平均をとったものである。この制約は、一つの楽器の複数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。 20

【0096】

(3-2: 非調波モデルに対する同一楽器における複数の音モデル間での条件)

同一楽器における複数の音モデルにおける非調波モデルに対する制約は、具体的には以下の式(8)で与えられる。

【数7】

$$\beta_{I1} \iint \left(\bar{I}_k \log \frac{\bar{I}_k}{I_{kl}} - (\bar{I}_k - I_{kl}) \right) dt df \quad (8) \quad 30$$

【0097】

上記式 I1 は、係数である。また I の上に $-$ が付いた関数を以下 $I-k$ と言う。ここで、 $I-k$ は、同一楽器の複数の音モデルの I_{kl} の平均をとったものである。この制約は、同一楽器の複数の単音(または複数の単音についての複数の音モデル)に対して非調波成分を類似させる作用を持っている。

【0098】

(4) モデルパラメータの反復推定処理

以上の第1乃至第3の制約条件の下で、観測対象となるパワースペクトル(入力音響信号のパワースペクトル) $g(0)(c, t, f)$ を、(2)に示す調波・非調波混合モデルを構成するモデルパラメータに変換するために、観測対象となるパワースペクトル(入力音響信号のパワースペクトル)を複数の単音に対応する複数のパワースペクトルに分解する処理(これを、「分離処理」という。)を行う。この処理を行うために、パワースペクトルの分配関数 $m_{kl}(c, t, f)$ を導入する。なお、以下単に、パワースペクトル $g(0)(c, t, f)$ を $g(0)$ 、分配関数 $m_{kl}(c, t, f)$ を m_{kl} と表記する場合がある。本発明では、第1回目に使用する分配関数を初期分配関数と呼び、第2回目以降の分離処理に用いる分配関数を更新分配関数と呼んでいる。 40

【0099】

なお、 c は左右などのチャンネル、 t は時刻、 f は周波数である。また各記号の添字 k 50

は k 番目の楽器 ($1 \leq k \leq K$)、 l は l 番目の単音 ($1 \leq l \leq L$) を表すものとする。本実施形態において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。すなわち、観測対象となるパワースペクトル $g(O)$ には K 個の楽器で演奏されたパワースペクトルの全てが包含されており、各楽器は L_k 個の単音を持つとする。また、 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 (テンプレート) を $g_{kl}(T)(t, f)$ と表し、対応する単音のパワースペクトルを $h_{kl}(c, t, f)$ とする [以下テンプレート音のパワースペクトル $g_{kl}(T)$ を $g_{kl}(T)$ 、音モデル $h_{kl}(c, t, f)$ を h_{kl} と書き表す]。事前の楽譜情報データ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。

【0100】

10

図 8 は、本発明で採用されるモデルパラメータ反復推定処理の一例の大まかな手順をフローチャートで表したものである。この例では、先の実施の形態と異なって、パワースペクトルによって表現される各楽器の複数の単音の複数のテンプレートを、複数のテンプレート音から作成している。

【0101】

(S1) 初めに、事前の楽譜情報データから、少なくとも各単音の音高と、オンセット時刻と、音長又はオフセット時刻と、楽器ラベルとを含む情報を抽出し、音響信号変換手段によって事前の音楽情報を音響信号に変換して全ての単音をテンプレート音として記録する (すなわち「テンプレート音を録音」する)。

【0102】

20

(S2) 次に、テンプレート音からパワースペクトルで表現した全ての単音についての複数のテンプレートを作成する。複数のテンプレートを調波・非調波混合モデルを構成するモデルパラメータに置き換えて複数のモデルパラメータが集合して構成されるモデルパラメータ集合データを作成する。このことを「モデルパラメータをテンプレート音で初期化する」と言う。なおこのモデルパラメータ集合データから読み出した各時刻における複数のモデルパラメータに基づいて、各時刻毎の複数の初期分配関数を計算する。

【0103】

(S3) 次に、入力音響信号のパワースペクトルから各時刻における複数の単音に対応する複数のパワースペクトルを、各時刻毎の複数の初期分配関数を用いて分離する。分離処理は、入力音響信号のパワースペクトルに初期分配関数を乗算することにより実行される。そして、次に分離した各時刻における複数のパワースペクトルから更新モデルパラメータを推定する。そして分離音のパワースペクトルから生成した複数の更新モデルパラメータから作成した複数の更新パワースペクトルと入力音響信号のパワースペクトルから分離された複数のパワースペクトルとの近さ即ち KL ダイバージェンス J_1 と、前記テンプレート音に基づいて最初に作成したモデルパラメータ集合データから作成した複数の初期パワースペクトルと更新パワースペクトルとの近さすなわち KL ダイバージェンス J_2 とを、 $\alpha : (1 - \alpha)$ (但し α は $0 \leq \alpha \leq 1$ を満たす実数である。) の比率で重み付けして、それらを足し合わせて現在のコスト関数と定める。そこで最初は、 α の初期値を 0 に設定する。

30

【0104】

40

(S4) 次に、更新パワースペクトルから各時刻における複数の更新分配関数を計算する。

【0105】

(S5) 次に、更新分配関数を用いて分離処理を実行する。

【0106】

(S6) 次に、 $J_1 = 1$ か否かを判断し、 $J_1 = 1$ のときは終了する。

【0107】

(S7) もし、S6 において $J_1 \neq 1$ のときは分離したパワースペクトルからコスト関数を用いて更新モデルパラメータを推定する (モデルパラメータを更新する) と共に α を増加させる。

50

【 0 1 0 8 】

(S 8) ステップ S 4 にジャンプする。

【 0 1 0 9 】

この例では、モデルパラメータの初期値として、テンプレート音を利用し、そこで得られたモデルパラメータから生成した初期パワースペクトルに基づいて初期分配関数を作成する。そして、この初期分配関数から最初の分離音を生成する。そして、さらに分離音を分離精度の高いものにするために（または分離音の良し悪しを評価するために）、分離音のパワースペクトルと分離音を更新モデルパラメータに変換した後に得た更新パワースペクトルとの近さ J 1 と、初期モデルパラメータから得た初期パワースペクトルと更新パワースペクトルとの近さ J 2 とを、 で重み付けして を初期値 0 から徐々に 1 に近づけていくことにより、最初は更新パワースペクトルをテンプレートに近くなるように推定し、分離処理とモデル適応とを繰り返しながら徐々に分離されたパワースペクトルに近づけることで、モデルパラメータの過学習を防止している。

10

【 0 1 1 0 】

本実施形態では、更新モデルパラメータが望ましい形で収束するように (3) に示すようなモデルパラメータへの適切な制約条件を設定し、かつその条件の下で (4) に示すようなモデル適応（モデルパラメータの反復推定処理）を行っている。

【 0 1 1 1 】

なお、上述した分離とモデル適応を繰り返すという一連のステップ（ (S 1) ~ (S 8) のステップ）は、分配関数 m_{kl} と調波・非調波混合モデルで表されるパワースペクトル h_{kl} のパラメータとを最適化することに他ならず、見方を変えると最大事後確率（Maximum A Posteriori）推定に基づく EM アルゴリズムとみなすこともできる。すなわち、分配関数 m_{kl} を導出することは EM アルゴリズム中の E（Expectation）ステップに相当し、調波・非調波混合モデル h_{kl} を構成する更新モデルパラメータを更新することは M（Maximization）ステップに相当する。

20

【 0 1 1 2 】

このことは、次式 (9) で定義される Q 関数を考えると明確になる。

【 数 8 】

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

30

40

【 0 1 1 3 】

Q 関数がコスト関数 J 0 に相当し、さらに確率密度関数は表 2 に示すように、それぞれが先に述べた関数 $g(0)$ 、 $g_{kl}(T)$ 、 h_{kl} 、 m_{kl} に対応する。

【表 2】

表2 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}

10

【0114】

但し、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必要がある。

【0115】

さらに、次式(10)を考えると、後述する式(17)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(10)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。すなわちこの導出は、EMアルゴリズムのE (Expectation) ステップに相当する。また、 $p(k, l|c, t, f, \theta)$ (すなわち h_{kl})の更新は、Q関数を m_{kl} 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のである。

20

【数9】

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

【0116】

以下、モデルパラメータの推定処理における計算手法について数式を用いて具体的に説明する。

30

【0117】

観測対象である入力音響信号のパワースペクトル $g^{(O)}$ から、各調波・非調波混合モデル h_{kl} をそれぞれ構成するモデルパラメータのパラメータを推定するために、それぞれのモデルパラメータが表す単音に相当するパワースペクトルを分離するために利用するパワースペクトルの分配関数 $m_{kl}(c, t, f)$ は、 k 番目の楽器、 l 番目の単音が $g^{(O)}$ のパワースペクトルうちの程度の割合を占めるかを表している。従って、入力音響信号のパワースペクトルと分配関数との積 $g^{(O)} \cdot m_{kl}$ を計算すれば、 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が求められる。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、分配関数 m_{kl} は、下記の関係を満たせばよい。

40

【数10】

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

【0118】

を満たせばよい。

【0119】

ここで、分配関数による分離の良し悪しを評価するために、積 $g^{(O)} \cdot m_{kl}$ により求めた全ての分離した単音のパワースペクトルと全ての更新パワースペクトル h_{kl} との間

50

のKLダイバージェンス(相対エントロピー) $J_1(k, l)$ を用いる [式(11) 参照]

【数11】

$$J_1(k, l) = \sum_c \iint g^{(O)} m_{kl} \log \frac{g^{(O)} m_{kl}}{h_{kl}} dt df \quad (11)$$

10

【0120】

また、推定された更新モデルパラメータの良し悪しを評価するために、テンプレート音 $g_{kl}(T)$ から求めた初期モデルパラメータから作成した初期パワースペクトルと更新モデルパラメータから作成した更新パワースペクトル (h_{kl}) との間のKLダイバージェンス $J_2(k, l)$ を用いる [式(12) 参照]。

【数12】

$$J_2(k, l) = \sum_c \iint g_{kl}^{(T)} \log \frac{g_{kl}^{(T)}}{h_{kl}} dt df \quad (12)$$

20

【0121】

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を評価するため、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KLダイバージェンスの総和 J_0 を用いる [式(13) 参照]。更新モデルパラメータを構成する複数のパラメータの推定には、この総和 J_0 に基づくコスト関数 J [式(21)] を用いる。

【数13】

$$J_0 = \sum_{k,l} (\alpha J_1(k, l) + (1 - \alpha) J_2(k, l)) \quad (13)$$

30

【0122】

ここで、 $(0 \leq \alpha \leq 1)$ は、分離とモデル推定のどちらを重視するかを表すパラメータである。 α の値は、最初は0に設定し(すなわち、これはモデルパラメータから作成するパワースペクトルの初期値をテンプレート音に基づく初期パワースペクトルとすることを意味する)、徐々に1に近づく(すなわち、更新パワースペクトルを入力音響信号から分離したパワースペクトルに近づく)。

【0123】

40

分離とモデル適応の繰り返しは、分配関数 m_{kl} の推定とパワースペクトル (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、 J をラグランジュの未定乗数項とし、最小化すべきコスト関数として改めて J_0 を定義すると、次式(14)のように記述される。

【数 1 4】

$$\begin{aligned}
 J_0 = & \alpha \sum_{k,l,c} \iint g^{(O)} m_{kl} \log \frac{g^{(O)} m_{kl}}{h_{kl}} dt df \\
 & + (1 - \alpha) \sum_{k,l,c} \iint g_{kl}^{(T)} \log \frac{g_{kl}^{(T)}}{h_{kl}} dt df \\
 & - \lambda \left(\sum_{k,l} m_{kl} - 1 \right)
 \end{aligned} \tag{14}$$

10

【0 1 2 4】

まず、分離を行うために、パワースペクトル (h_{kl}) を固定して総和 J_0 が最小化されるような分配関数 m_{kl} を求める。 J_0 を偏微分すると、下記 (15) 式のようになる。

【数 1 5】

$$\begin{cases} \frac{\partial J_0}{\partial m_{kl}} = \alpha g^{(O)} \log \frac{g^{(O)} m_{kl}}{h_{kl}} - \lambda \\ \frac{\partial J_0}{\partial \lambda} = \sum_{k,l} m_{kl} - 1 \end{cases} \tag{15}$$

30

【0 1 2 5】

これらを用いて連立方程式

【数 1 6】

$$\frac{\partial J_0}{\partial m_{kl}} = 0, \quad \frac{\partial J_0}{\partial \lambda} = 0 \tag{16}$$

【0 1 2 6】

を解くと、

【数 1 7】

$$m_{kl} = \frac{h_{kl}}{\sum_{k,l} h_{kl}} \tag{17}$$

40

【0 1 2 7】

を得る。

【0 1 2 8】

次に、モデル適応を行うために、分配関数 m_{kl} を固定してコスト関数 J が最小化される

50

ような調波・非調波混合モデル (hkl) を求めることで、コスト関数 J を最小化することができる。

【 0 1 2 9 】

ここで、コスト関数 J は全ての単音に対するコストとみなすことができるが、式 (1) と前述の [数 2] に示した条件とより明らかなように、観測対象の入力音響信号のパワースペクトル全体のモデルは各単音の線形和である。そして一つ一つの音モデルは調波モデルと非調波モデルの線形和である。また調波モデルは基底関数の線形和で表現されていることから、観測対象の入力音響信号のパワースペクトルの全体を、各単音の個々のガウス分布関数 (調波モデルに相当) と非調波モデルとに分解することで、モデルパラメータを解析的に最適化することが可能である。

10

【 0 1 3 0 】

そこで、新たな二つのパワースペクトルの分配関数 $m_{klyn}^{(H)}(t, f)$ と $m_{kl}^{(I)}(t, f)$ とを導入する。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を $\{y, n\}$ ラベル付のガウス分布関数 (調波モデルに相当) 及び非調波モデルに分配する関数である。

【 数 1 8 】

$$\begin{cases} \sum_{y,n} m_{klyn}^{(H)}(t, f) + m_{kl}^{(I)}(t, f) = 1 \\ 0 \leq m_{klyn}^{(H)}(t, f) \leq 1 \\ 0 \leq m_{kl}^{(I)}(t, f) \leq 1 \end{cases} \quad (18)$$

20

【 0 1 3 1 】

を満たす。

【 0 1 3 2 】

調波・非調波混合モデルのパワースペクトル (hkl) を固定した状態で、コスト関数 J を最小化するこれらの分配関数を導出すると、

30

【 数 1 9 】

$$\begin{cases} m_{klyn}^{(H)} = \frac{w_{kl} E_{kly} F_{kln}}{H_{kl} + I_{kl}} \\ m_{kl}^{(I)} = \frac{I_{kl}}{H_{kl} + I_{kl}} \end{cases} \quad (19)$$

40

【 0 1 3 3 】

となる。詳細は省略するが、前述の分配関数 mkl と同様の導出過程で示すことができる。

【 0 1 3 4 】

r, u, v をそれぞれ rklc, rkly, kln に対するラグランジュの未定乗数項として、

【数 2 0】

$$\begin{cases} G_{kl}(c, t, f) = \alpha g^{(O)} m_{kl} + (1 - \alpha) g_{kl}^{(T)} \\ G_{klyn}^{(H)}(c, t, f) = m_{klyn}^{(H)} G_{kl}(c, t, f) \\ G_{kl}^{(I)}(c, t, f) = m_{kl}^{(I)} G_{kl}(c, t, f) \end{cases} \quad (20)$$

【0 1 3 5】

10

とすると、各単音の調波・非調波混合モデル (hkl) の各パラメータに対する更新式は下記の式 (21) のコスト関数 J から求めることができる。

【数 2 1】

$$\begin{aligned} J = & \sum_{k,l} \left(\sum_{c,y,n} \iint \left(G_{klyn}^{(H)} \log \frac{G_{klyn}^{(H)}}{r_{klc} w_{kl} E_{kly} F_{kln}} - G_{klyn}^{(H)} + r_{klc} w_{kl} E_{kly} F_{kln} \right) dt df \right. \\ & + \sum_c \iint \left(G_{kl}^{(I)} \log \frac{G_{kl}^{(I)}}{r_{klc} I_{kl}} - G_{kl}^{(I)} + r_{klc} I_{kl} \right) dt df \\ & + \beta_v \sum_n \left(\bar{v}_{kn} \log \frac{\bar{v}_{kn}}{v_{kln}} - \bar{v}_{kn} + v_{kln} \right) + \beta_\mu \int \left(\bar{\mu}_{kl}(t) \log \frac{\bar{\mu}_{kl}(t)}{\mu_{kl}(t)} - \bar{\mu}_{kl}(t) + \mu_{kl}(t) \right) dt \\ & + \beta_{I1} \iint \left(\bar{I}_k \log \frac{\bar{I}_k}{I_k} - \bar{I}_k + I_k \right) dt df + \beta_{I2} \iint \left(\bar{I}_{kl} \log \frac{\bar{I}_{kl}}{I_{kl}} - \bar{I}_{kl} + I_{kl} \right) dt df \\ & \left. - \lambda_r \left(\sum_c r_{klc} - 1 \right) - \lambda_u \left(\sum_y u_{kly} - 1 \right) - \lambda_v \left(\sum_n v_{kln} - 1 \right) \right) \quad (21) \end{aligned}$$

20

【0 1 3 6】

30

すなわちコスト関数 J 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、コスト関数が極小となるように、更新モデルパラメータを構成するパラメータを更新する (推定する) 式をそれぞれ導出することができる。なおこのような式を導出する手法は公知であるため詳しい説明は省略する。なお式 (21) のコスト関数 J において、式中の最初の 2 つの式が (1 -) の重みを付けた前述の総和 J 0 に相当するものであり、3 番目から 7 番目までの式が前述の式 (5) 乃至 (8) の制約条件である。なおこれらの制約条件は付加することが好ましいが、必要に応じて適宜に付加すればよいものである。しかし式 (6) の制約条件は他の制約条件に優先する。また式 (6) の制約条件の次には、式 (5) の制約条件が残りの制約条件に優先する。

【0 1 3 7】

40

- 評価結果 -

本発明に係る上記音源分離方法の各ステップを実行するプログラムを作成し、公的な研究用音楽データベースの一つである RWC 研究用音楽データベースに登録されているポピュラー音楽データベース (RWC - MDB - P - 2001) から選んだ 10 曲 (Nos. 1 - 10) を用いて音源分離を行った。各楽曲は開始から 30 秒の区間を利用した。なお、詳細な実験条件を表 3 に示す。

【表 3】

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

10

20

【0138】

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。なお、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。

【0139】

本発明は調波・非調波混合モデルを利用する点の一つの特徴であるが、比較のため、調波モデルを利用した場合と非調波モデルを利用した場合についても同一条件で実験を行った。

【0140】

図9は、各楽器パートのSNR (Signal to Noise Ratio) を楽曲ごとに平均したものと、全楽曲、全パートについてSNRを平均した結果を示す一覧表である。この表は、10曲のSNRを平均すると、混合モデルを用いた場合にSNRは他の単一モデルと比較して最も高くなっていることを示している。

30

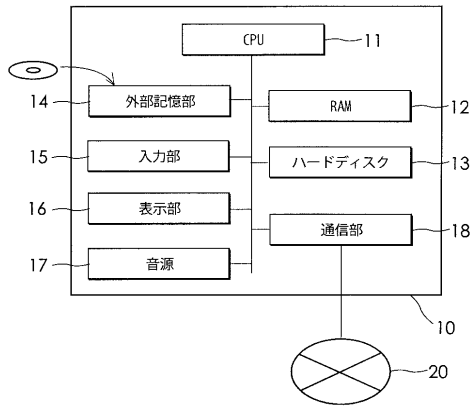
【産業上の利用可能性】

【0141】

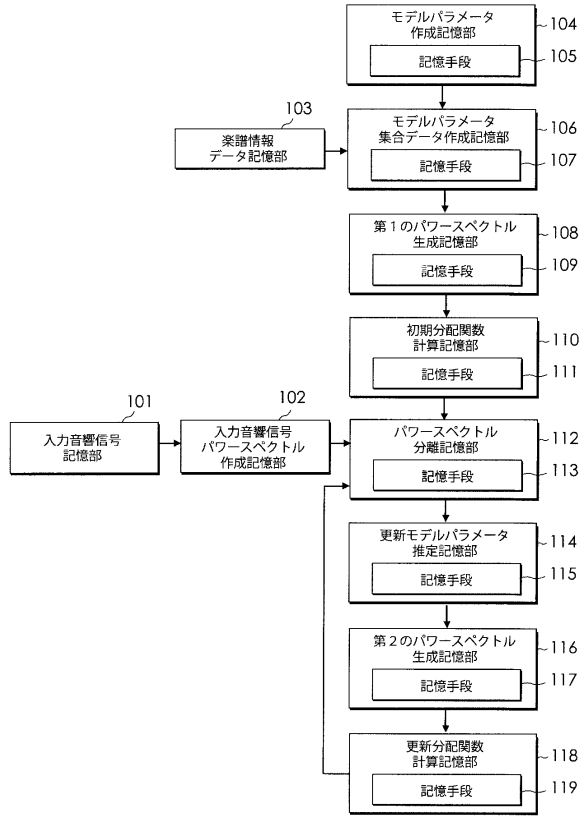
本発明によれば、調波モデルと非調波モデルを含む形で楽器音のパワースペクトルを分離することができるので、入力音響信号中の楽器音に近い楽器音(音源)を分離することができる。また本発明によれば、楽器パートごとの音量の増減やエフェクトを自由に操作することが可能になる。また本発明に係る音源分離システム及び方法は、従来は一部の専門家のみが利用していた高度な操作技術を必要とする高価な音響機器を用いることなく、個人がパーソナルコンピュータ上で楽器音ごとに音量を増減することが可能な「楽器音イコライザ」を実現することを可能にするコンピュータプログラムのキーテクノロジーとして位置づけられ、産業上の利用可能性は極めて大きい。

40

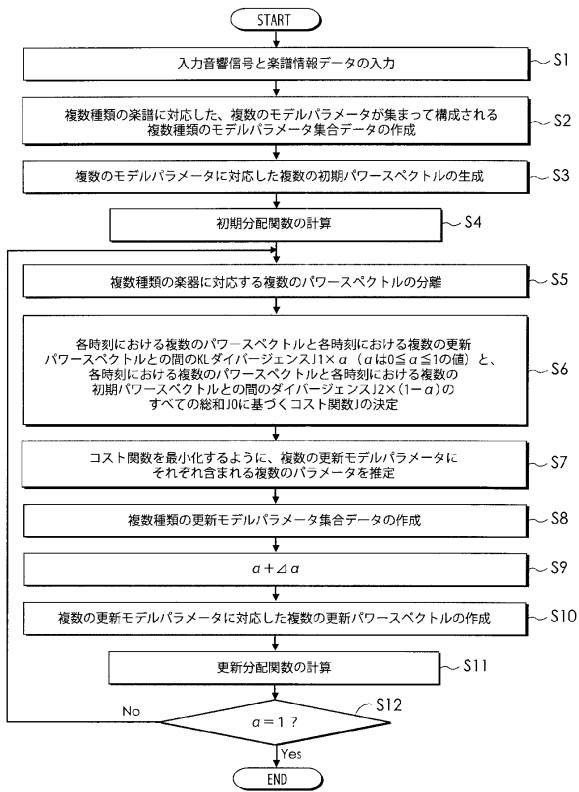
【図1】



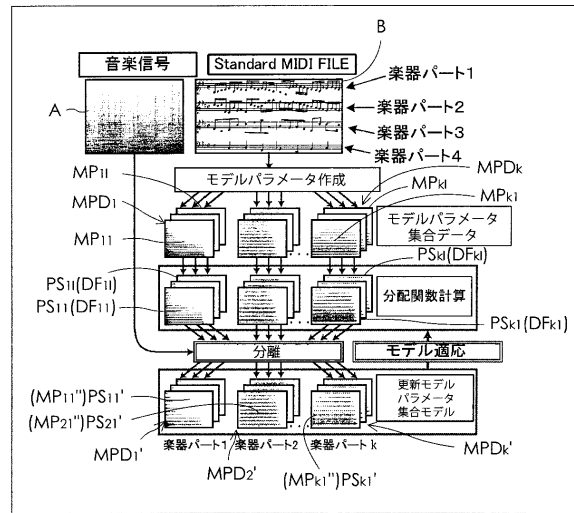
【図2】



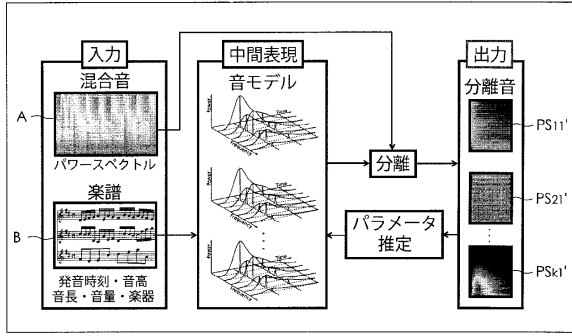
【図3】



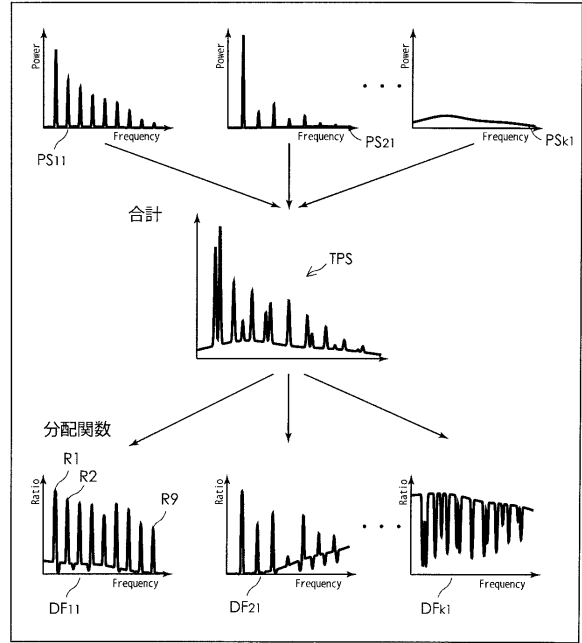
【図4】



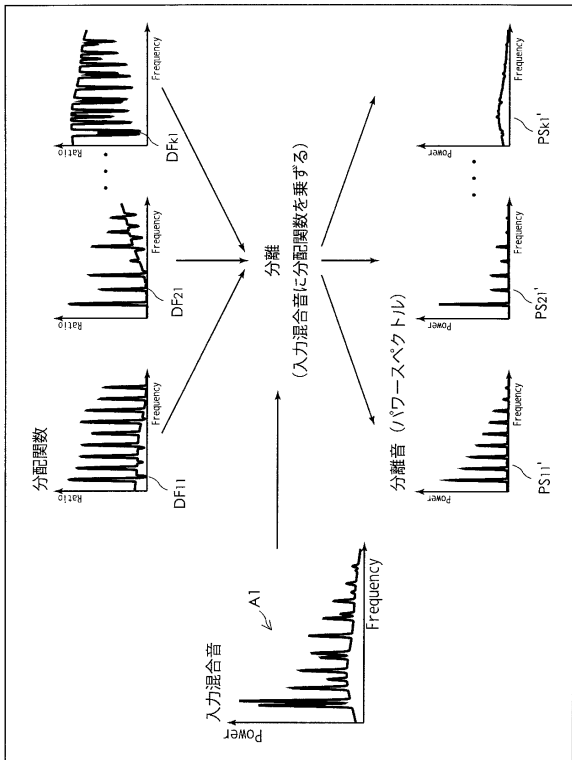
【図5】



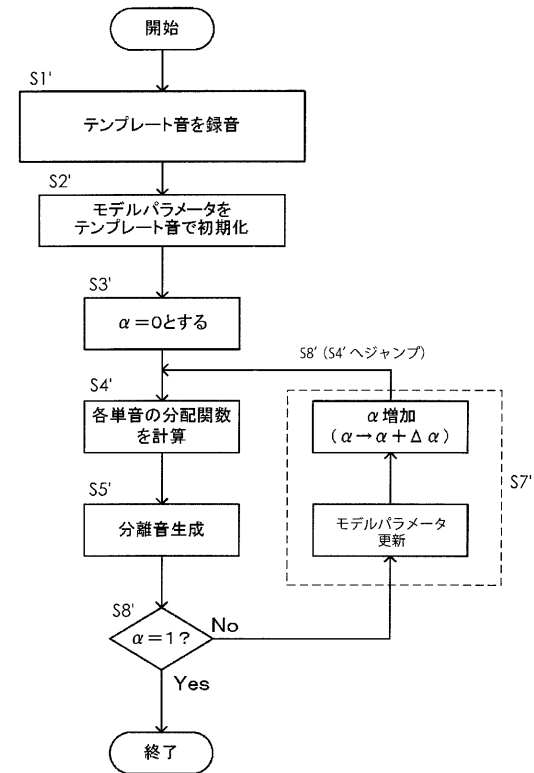
【図6】



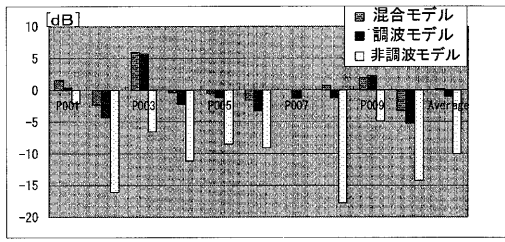
【図7】



【図8】



【 図 9 】



フロントページの続き

(72)発明者 後藤 真孝

茨城県つくば市東1-1-1 独立行政法人産業技術総合研究所 つくばセンター内

審査官 菊地 陽一

(56)参考文献 特開2002-244691(JP,A)

特開平11-095753(JP,A)

特許第3413634(JP,B2)

(58)調査した分野(Int.Cl., DB名)

G10H 1/00

G10L 21/0308

G10L 21/028