

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2017-67903

(P2017-67903A)

(43) 公開日 平成29年4月6日(2017.4.6)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 0 L 25/87 (2013.01)	G 1 0 L 25/87	5 D 1 8 2
G 1 0 G 1/00 (2006.01)	G 1 0 G 1/00	

審査請求 未請求 請求項の数 5 O L (全 18 頁)

(21) 出願番号	特願2015-191028 (P2015-191028)	(71) 出願人	000004075
(22) 出願日	平成27年9月29日 (2015.9.29)		ヤマハ株式会社
			静岡県浜松市中区中沢町10番1号
		(74) 代理人	100125689
			弁理士 大林 章
		(74) 代理人	100128598
			弁理士 高田 聖一
		(74) 代理人	100121108
			弁理士 高橋 太朗
		(72) 発明者	有元 慶太
			静岡県浜松市中区中沢町10番1号 ヤマハ株式会社内
		Fターム(参考)	5D182 AC01 AD10

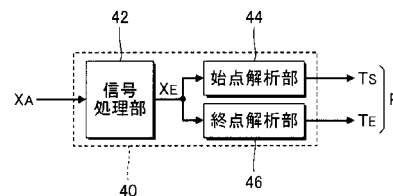
(54) 【発明の名称】 音響解析装置

(57) 【要約】

【課題】音響信号のうち発音源の発音の直後の発音区間を高精度に特定する。

【解決手段】発音区間検出部40は、音響信号XAのうち音響の発音が開始される発音始点TSと当該音響の発音が終了する発音終点TEとを解析する要素であり、音響信号XEの強度の極大点を発音始点TSとして特定する始点解析部44と、発音始点TSの経過後に音響信号XEの強度が経時的に減少する過程で強度が増加に反転する極小点を、増加後の極大点での強度と当該極大点までの強度の最小値との差分に応じた変動指標が終点閾値を上回る場合に発音終点として特定する終点解析部46とを具備する。

【選択図】 図4



【特許請求の範囲】**【請求項 1】**

音響信号のうち音響の発音が始まる発音始点と当該音響の発音が終了する発音終点とを解析する音響解析装置であって、

前記音響信号の強度の極大点を前記発音始点として特定する始点解析部と、

前記発音始点の経過後に前記音響信号の強度が経時的に減少する過程で強度が増加に反転する極小点を、前記増加後の極大点での強度と当該極大点までの強度の最小値との差分に応じた変動指標が終点閾値を上回る場合に前記発音終点として特定する終点解析部とを具備する音響解析装置。

【請求項 2】

前記終点解析部は、前記極小点を前記発音終点として特定する以前に、前記発音始点での強度に応じた減衰閾値を下回るまで前記音響信号の強度が当該発音始点から減少した場合に、前記減衰閾値を強度が下回る時点を前記発音終点として特定する

請求項 1 の音響解析装置。

【請求項 3】

前記始点解析部は、前記音響信号の強度の極大点を順次に検出する一方、前記極大点での強度と当該極大点までの強度の最小値との差分に応じた変動指標が、前記終点閾値よりも大きい始点閾値を上回る場合に、当該極大点を前記発音始点として特定する

請求項 1 または請求項 2 の音響解析装置。

【請求項 4】

前記変動指標は、前記極大点での強度と当該極大点までの強度の最小値との差分を当該極大点での強度により除算した数値である

請求項 1 から請求項 3 の何れかの音響解析装置。

【請求項 5】

前記始点解析部は、前記音響信号の強度の第 1 極大点以降の待機区間内に、前記第 1 極大点を上回る強度の第 2 極大点を検出した場合に、前記第 1 極大点を前記発音始点の候補から除外する

請求項 1 から請求項 4 の何れかの音響解析装置。

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、音響を解析する技術に関する。

【背景技術】**【0002】**

音響信号のうち音響が発音される発音区間を解析するための各種の技術が従来から提案されている。例えば特許文献 1 には、音響信号の S/N (Signal to Noise) 比が所定の条件を充足する期間を発音区間として特定する構成が開示されている。

【先行技術文献】**【特許文献】****【0003】**

【特許文献 1】特開 2008 - 170806 号公報

【発明の概要】**【発明が解決しようとする課題】****【0004】**

ところで、打楽器等の各種の楽器が発音する音響の解析には当該楽器の発音区間の特定が重要である。しかし、例えば打楽器が素早く連打された場合のように発音源が短い間隔で複数回にわたり発音した場合には、最初の発音による音響が十分に減衰する以前に直後の発音が始まる。したがって、特許文献 1 の技術のもとでは、複数回にわたる発音が 1 個の発音区間に包含される可能性がある。しかし、実際の演奏音の解析の場面では、発音

10

20

30

40

50

の開始直後の特性の解析が重要である場合が想定されるから、発音源が短い間隔で複数回にわたり発音した場合でも、最初の発音による音響のみを含む発音区間を高精度に特定することが重要である。以上の事情を考慮して、本発明は、音響信号のうち発音源の発音の直後の発音区間を高精度に特定することを目的とする。

【課題を解決するための手段】

【0005】

以上の課題を解決するために、本発明の好適な態様に係る音響解析装置は、音響信号のうち音響の発音が開始される発音始点と当該音響の発音が終了する発音終点とを解析する音響解析装置であって、音響信号の強度の極大点を発音始点として特定する始点解析部と、発音始点の経過後に音響信号の強度が経時的に減少する過程で強度が増加に反転する極小点を、増加後の極大点での強度と当該極大点までの強度の最小値との差分に応じた変動指標が終点閾値を上回る場合に発音終点として特定する終点解析部とを具備する。以上の態様では、発音始点の経過後に音響信号の強度が経時的に減少する過程で強度が増加に反転する極小点を、変動指標が終点閾値を上回る場合に発音終点として特定する。すなわち、発音源が短い間隔で複数回にわたり発音した場合（最初の発音による音響が十分に減衰する以前に直後の発音が開始する場合）には、発音始点に対応する最初の発音のみを発音区間が包含するように発音終点が特定される。したがって、音響信号のうち発音源の発音の直後の発音区間を高精度に特定することが可能である。

10

【0006】

本発明の好適な態様において、終点解析部は、極小点を発音終点として特定する以前に、発音始点での強度に応じた減衰閾値を下回るまで音響信号の強度が当該発音始点から減少した場合に、減衰閾値を強度が下回る時点を発音終点として特定する。以上の態様では、極小点が発音終点として特定される以前に、発音始点での強度に応じた減衰閾値を下回るまで音響信号の強度が発音始点から減少した場合に、強度が減衰閾値を下回る時点が発音終点として特定される。したがって、発音始点の経過後に発音源が発音することなく音響信号が減衰する場合に、発音始点からの減衰の度合いに応じた適切な発音終点を設定できるという利点がある。

20

【0007】

本発明の好適な態様において、始点解析部は、音響信号の強度の極大点を順次に検出する一方、極大点での強度と当該極大点までの強度の最小値との差分に応じた変動指標が、終点閾値よりも大きい始点閾値を上回る場合に、当該極大点を発音始点として特定する。以上の態様では、音響信号の強度の極大点が順次に検出される一方、極大点での強度と当該極大点までの強度の最小値である基準値との差分に応じた変動指標が始点閾値を上回る場合に、当該極大点が発音始点として特定される。したがって、音響信号から検出される複数の極大点のうち発音源の明瞭な発音の開始を発音始点として高精度に特定できるという利点がある。

30

【0008】

本発明の好適な態様において、変動指標は、極大点での強度と当該極大点までの強度の最小値との差分を当該極大点での強度により除算した数値である。以上の態様では、極大点での強度と基準値との差分を極大点での強度により除算することで変動指標が算定される。すなわち、差分が音響信号の音量の大小に依存しない数値に正規化される。したがって、音響信号の音量に関わらず発音始点および発音終点を適切に特定することが可能である。

40

【0009】

本発明の好適な態様において、始点解析部は、音響信号の強度の第1極大点以降の待機区間内に、第1極大点を上回る強度の第2極大点を検出した場合に、第1極大点を発音始点の候補から除外する。以上の態様では、音響信号の強度の極大点以降の待機区間内に、当該極大点を上回る強度の極大点が発検出された場合に、極大点が発音始点の候補から除外される。したがって、発音源による1回の発音の開始から音響信号の強度が増加する過程で複数の極大点が発検出される場合でも、当該発音に対応した1個の極大点を含む発音区間

50

を適切に特定することが可能である。

【図面の簡単な説明】

【0010】

【図1】本発明の第1実施形態に係る音響処理装置の構成図である。

【図2】音響解析部の構成図である。

【図3】音響信号の各発音区間の説明図である。

【図4】発音区間検出部の構成図である。

【図5】発音区間検出部の動作の説明図である。

【図6】始点解析処理のフローチャートである。

【図7】終点解析処理のフローチャートである。

【図8】音源識別部の構成図である。

【図9】調波解析処理のフローチャートである。

【図10】音源識別処理のフローチャートである。

【図11】第2実施形態における始点解析処理の説明図である。

【発明を実施するための形態】

【0011】

<第1実施形態>

図1は、本発明の第1実施形態の音響処理装置12の構成図である。図1に例示される通り、音響処理装置12には複数の收音装置14と放音装置16とが接続される。複数の收音装置14の各々は、当該收音装置14の周囲の音響を表す音響信号XAを生成する。音響信号XAは、例えば左右2チャンネルのステレオ形式の信号である。複数の收音装置14が生成した複数の音響信号XAが音響処理装置12に並列に供給される。なお、收音装置14が生成した音響信号XAをアナログからデジタルに変換するA/D変換器の図示は便宜的に省略した。

【0012】

各收音装置14は相異なる発音源の近傍に配置される。発音源は、例えば演奏により楽音を発音する楽器や歌唱音声を発音する歌唱者である。第1実施形態では、収録スタジオ等の音響空間の内部で歌唱者と複数の楽器とにより音楽を演奏する場合を想定する。各收音装置14が生成する音響信号XAには、当該收音装置14の近傍の発音源から発音された音響が優勢に含有されるが、当該音響と比較して小音量で他の発音源の音響も含有され得る。

【0013】

第1実施形態の各発音源は、調波音または非調波音を発音する。調波音は、基本周波数の基音成分と複数の倍音成分とを周波数軸上に配列した調波構造が明瞭に観測される調波性の音響である。例えば弦楽器または管楽器等の調波楽器の楽音や歌唱音声等の人間の発音が調波音の典型例である。他方、非調波音は、調波構造が明瞭に観測されない非調波性の音響である。例えばドラムやシンバル等の打楽器の楽音が非調波音の典型例である。

【0014】

なお、調波音は、調波性の音響成分を非調波性の音響成分と比較して優勢に含有する音響を意味する。したがって、調波性の音響成分のみで構成される音響のほか、調波性の音響成分と非調波性の音響成分との双方を含有するが全体としては調波性が優勢である音響も、調波音の概念に包含される。同様に、非調波音は、非調波性の音響成分を調波性の音響成分と比較して優勢に含有する音響を意味する。したがって、非調波性の音響成分のみで構成される音響のほか、調波性の音響成分と非調波性の音響成分との双方を含有するが全体としては非調波性が優勢である音響も、非調波音の概念に包含される。以下の説明では、調波音に関連する要素の符号に添字H（H：Harmonic）を付加し、非調波音に関連する要素の符号に添字P（P：Percussive）を付加する場合がある。

【0015】

音響処理装置12は、複数の音響信号XAに対する音響処理で音響信号XBを生成する。具体的には、第1実施形態の音響処理装置12は、複数の音響信号XAの混合（ミキシン

10

20

30

40

50

グ)により左右2チャンネルのステレオ形式の音響信号XBを生成する。放音装置16(例えばスピーカやヘッドホン)は、音響処理装置12が生成した音響信号XBに応じた音響を放音する。なお、音響処理装置12が生成した音響信号XBをデジタルからアナログに変換するD/A変換器の図示は便宜的に省略した。また、図1では各收音装置14と放音装置16とを音響処理装置12とは別個の要素として図示したが、複数の收音装置14と放音装置16とを音響処理装置12に搭載することも可能である。

【0016】

図1に例示される通り、音響処理装置12は、制御装置122と記憶装置124とを具備するコンピュータシステムで実現される。記憶装置124は、例えば磁気記録媒体や半導体記録媒体等の公知の記録媒体または複数種の記録媒体の組合せであり、制御装置122が実行するプログラムや制御装置122が使用する各種のデータを記憶する。制御装置122は、記憶装置124が記憶するプログラムを実行することで、複数の音響信号XAの各々を解析する音響解析部20と、音響解析部20による解析結果を利用して複数の音響信号XAから音響信号XBを生成する音響処理部30とを実現する。なお、制御装置122の機能の一部または全部を専用の電子回路で実現する構成や、制御装置122の機能を複数の装置に分散した構成も採用され得る。

10

【0017】

音響解析部20は、複数の收音装置14から供給される複数の音響信号XAの各々について、当該音響信号XAが表す音響の発音源の種類を特定する。具体的には、音響解析部20は、各音響信号XAの発音源の種類を示す情報(以下「音源識別情報」という)Dを生成する。音源識別情報Dは、例えば発音源の名称(具体的には楽器名や演奏パート名)である。

20

【0018】

図2は、音響解析部20の構成図である。図2に例示される通り、第1実施形態の音響解析部20は、発音区間検出部40と特徴量抽出部50と音源識別部60とを具備する。なお、以下の説明では、任意の1系統の音響信号XAに対する処理に便宜的に着目するが、複数の音響信号XAの各々について同様の処理が実行される。

【0019】

図2の発音区間検出部40は、音響信号XAについて複数の発音区間Pを検出する。図3には、音響信号XAの波形と発音区間Pとの関係が図示されている。図3から理解される通り、各発音区間Pは、音響信号XAが表す音響が発音される時間軸上の区間であり、音響が発音を開始する時点(以下「発音始点」という)TSから終点(以下「発音終点」という)TEまでの区間である。

30

【0020】

図2の特徴量抽出部50は、音響信号XAの特徴量Fを抽出する。第1実施形態の特徴量抽出部50は、発音区間検出部40が検出した発音区間P毎に特徴量Fを順次に抽出する。特徴量Fは、発音区間P内の音響信号XAの音響的な特徴を表す指標である。第1実施形態の特徴量Fは、相異なる複数種の特性値 $f(f_1, f_2, \dots)$ を包含するベクトルで表現される。具体的には、音響信号XAの音色を表すMFCC(Mel-frequency cepstral coefficients)、発音区間P内の音響の立上がりの急峻度、基音成分に対する倍音成分の強度比、音響信号XAの強度の符号が反転する回数または頻度である零交差数等の複数種の特性値 f が特徴量Fに包含される。

40

【0021】

各発音源が発音する音響の特徴は、発音始点TSの直後に特に顕著となる。第1実施形態では、音響信号XAの発音始点TS毎(発音区間P毎)に音響信号XAの特徴量Fが抽出されるから、発音の有無や時点とは無関係に音響信号XAを区分した区間毎に特徴量Fを抽出する構成と比較して、発音源の種類毎に固有の特徴が顕著に反映された特徴量Fを抽出できるという利点がある。もっとも、発音源による発音の有無や時点とは無関係に音響信号XAを時間軸上で区分した区間毎に特徴量Fを抽出する(したがって発音区間検出部40は省略される)ことも可能である。音源識別部60は、特徴量抽出部50が抽出した

50

特徴量 F を利用して音響信号 X A の発音源の種類を識別することで音源識別情報 D を生成する。

【 0 0 2 2 】

図 1 の音響処理部 3 0 は、音響解析部 2 0 が音響信号 X A 毎に解析した音源識別情報 D を参照して複数の音響信号 X A に音響処理を実行することで音響信号 X B を生成する。具体的には、音響信号 X A の音源識別情報 D が示す発音源の種類毎に事前に設定された音響処理が当該音響信号 X A に対して実行される。音響信号 X A に対する音響処理としては、例えば残響効果や歪効果等の各種の音響効果を付与する効果付与処理（エフェクタ）や、周波数帯域毎の音量を調整する特性調整処理（イコライザ）、音像が定位する位置を調整する定位調整処理（パン）、音量を調整する音量調整処理が例示される。効果付与処理で音響信号 X A に付与される音響効果の種類や度合、特性調整処理で音響信号 X A に付与される周波数特性、定位調整処理で調整される音像の位置、音量調整処理による調整内容（ゲイン）等の各種のパラメータが、音源識別情報 D が示す発音源の種類毎に個別に設定される。そして、音響処理部 3 0 は、以上に例示した音響処理後の複数の音響信号 X A を混合（ミキシング）することで音響信号 X B を生成する。すなわち、第 1 実施形態の音響処理部 3 0 は、調波性解析部 6 2 による発音源の識別結果を反映した自動ミキシングを実現する。以下、第 1 実施形態における発音区間検出部 4 0 および音源識別部 6 0 の各々の具体的な構成を説明する。

10

【 0 0 2 3 】

< 発音区間検出部 4 0 >

20

図 4 は、発音区間検出部 4 0 の構成図である。図 4 に例示される通り、第 1 実施形態の発音区間検出部 4 0 A は、信号処理部 4 2 と始点解析部 4 4 と終点解析部 4 6 とを具備する。なお、以下の説明では、任意の 1 系統の音響信号 X A に対する処理に便宜的に着目するが、実際には複数の音響信号 X A の各々について同様の処理が実行される。

【 0 0 2 4 】

信号処理部 4 2 は、收音装置 1 4 から供給される音響信号 X A の信号処理で音響信号 X E を生成する。音響信号 X E は、音響信号 X A の時間軸上の包絡線（エンベロープ）に相当する。具体的には、信号処理部 4 2 は、音響信号 X A の各信号値を絶対値に変換したうえで高周波成分を抑圧（平滑化処理）することで音響信号 X E を生成する。音響信号 X E の波形が図 5 に例示されている。なお、外部装置で生成された音響信号 X E が音響処理装置 1 2 に供給される構成では、音響処理装置 1 2 から信号処理部 4 2 が省略され得る。

30

【 0 0 2 5 】

図 4 の始点解析部 4 4 は、音響信号 X E のうち音響の発音が始まる発音始点 T S を特定する。終点解析部 4 6 は、音響信号 X E のうち音響の発音が終了する発音終点 T E を特定する。第 1 実施形態では、始点解析部 4 4 による発音始点 T S の特定と終点解析部 4 6 による発音終点 T E の特定とが、音響信号 X E の生成に並行して実時間的に音響信号 X E の始点から時間の経過とともに順次に行われる。始点解析部 4 4 および終点解析部 4 6 の各々の動作を以下に説明する。

【 0 0 2 6 】

< 始点解析部 4 4 >

40

図 5 に例示される通り、第 1 実施形態の始点解析部 4 4 は、音響信号 X E の強度（振幅またはパワー）Q が増加から減少に反転する極大点（ピーク）x H を発音始点 T S として特定する。ただし、第 1 実施形態の始点解析部 4 4 は、音響信号 X E から検出される全部の極大点 x H を発音始点 T S とするのではなく、音響信号 X E から検出される複数の極大点 x H のうち所定の条件を充足する極大点 x H を選択的に発音始点 T S として特定する。

【 0 0 2 7 】

具体的には、始点解析部 4 4 は、図 5 に例示された極大点 x H1 のように、極大点 x H の音響信号 X E の強度 Q H と基準値 Q REF との差分 (Q H - Q REF) に応じた変動指標 が所定の閾値（以下「始点閾値」という）Z S を上回る場合 (> Z S) に当該極大点 x H を発音始点 T S として確定する。他方、図 5 に例示された極大点 x H0 のように、変動指標 が始

50

点閾値 ZSを下回る極大値 xHは発音始点 TSとされない。

【 0 0 2 8 】

基準値 QREFは、直前の発音始点 TS（処理開始の直後は音響信号 XEの始点）以降における音響信号 XEの強度 Qの最小値となるように発音始点 TSの解析処理の進行とともに随時に更新される。変動指標 は、例えば、極大点 xHでの強度 QHと基準値 QREFとの差分 (QH - QREF)を当該強度 QHで除算した数値 ($= (QH - QREF) / QH$) である。強度 QHでの除算により、変動指標 は、音響信号 XEの全体的な音量の大小に依存しない数値に正規化される。始点閾値 ZSは、事前に選定された所定の正数である。

【 0 0 2 9 】

図 6 は、始点解析部 4 4 が発音始点 TSを特定する処理（以下「始点解析処理」という）のフローチャートである。始点解析部 4 4 は、音響信号 XEの始点から順次に極大点 xHを検出し、極大点 xHの検出毎に図 6 の始点解析処理を開始する。

10

【 0 0 3 0 】

音響信号 XEの極大点 xHの検出を契機として始点解析処理を開始すると、始点解析部 4 4 は、当該極大点 xHでの強度 QHと現時点での基準値 QREFとの差分 (QH - QREF)に応じた変動指標 が始点閾値 ZSを上回るか否かを判定する (SC1)。変動指標 が始点閾値 ZSを下回る場合 (SC1:NO)、始点解析部 4 4 は、今回の極大点 xHを発音始点 TSとして特定することなく始点解析処理を終了する。他方、変動指標 が始点閾値 ZSを上回る場合 (SC1:YES)、始点解析部 4 4 は、今回の極大点 xHを発音始点 TSとして特定する (SC2)。そして、始点解析部 4 4 は、基準値 QREFを今回の極大点 xHでの強度 QHに更新する (SC3)。発音始点 TSが経過すると音響信号 XEは減衰するから、発音始点 TSの経過後は基準値 QREFは経時的に減少していく。以上が始点解析処理の好適例である。

20

【 0 0 3 1 】

< 終点解析部 4 6 >

図 4 の終点解析部 4 6 は、前述の通り、音響信号 XEのうち音響の発音が終了する発音終点 TEを特定する。図 7 は、終点解析部 4 6 が発音終点 TEを特定する処理（以下「終点解析処理」という）のフローチャートである。始点解析部 4 4 による発音始点 TSの特定 (SC2)を契機として図 7 の終点解析処理が開始される。

【 0 0 3 2 】

発音始点 TSの特定を契機として終点解析処理を開始すると、終点解析部 4 6 は、当該発音始点 TSから所定の時間 が経過したか否かを判定する (SD1)。発音始点 TSから所定の時間 が経過していない場合 (SD1:NO)、終点解析部 4 6 は、現時点の音響信号 XEの強度 Qが所定の閾値（以下「減衰閾値」という）Z0を下回るか否かを判定する (SD2)。減衰閾値 Z0は、直前の発音始点 TSでの音響信号 XEの強度 QHに応じた数値に設定される。具体的には、発音始点 TSでの強度 QHに 1 未満の正数（例えば 0.4 ~ 0.6 の任意の数値）を乗算した数値が減衰閾値 Z0として好適である。強度 Qが閾値 Z0を下回る場合 (SD2:YES)、終点解析部 4 6 は現時点を発音終点 TEとして特定する (SD3)。すなわち、発音始点 TSの経過後で音響信号 XEの強度 Qが減衰閾値 Z0を下回るまで減少した時点が発音終点 TEとして特定される。

30

【 0 0 3 3 】

ところで、例えば打楽器が素早く連打された場合のように発音源が短い間隔で複数回にわたり発音した場合には、最初の発音による音響が十分に減衰する以前に直後の発音が始まる。したがって、音響信号 XEの強度 Qが減衰閾値 Z0を下回る時点を発音終点 TEとして特定するだけでは、発音始点 TSから発音終点 TEまでの 1 個の発音区間 Pに発音源の複数回にわたる発音が包含される結果となる。しかし、例えば特徴量抽出部 5 0 による特徴量 Fの抽出や音源識別部 6 0 による発音源の種類識別等の音響信号 XAの解析の場面では、発音源の発音の開始直後の特性の解析が重要である。以上の事情を考慮して、第 1 実施形態の終点解析部 4 6 は、発音源が短い間隔で複数回にわたり発音した場合でも、発音始点 TSに対応する最初の発音のみを発音区間 Pが包含するように（すなわち第 2 回目以降の発音が発音区間 Pに包含されないように）、発音終点 TEを特定する。

40

50

【 0 0 3 4 】

具体的には、音響信号 XE の強度 Q が減衰閾値 Z0 を上回る場合 (SD2 : NO)、終点解析部 4 6 は、発音始点 TS 以降に音響信号 XE の強度 Q が減少から増加に反転する極小点 (ディップ) xL が検出されたか否かを判定する (SD4)。極小点 xL が検出されない場合 (SD4 : NO)、終点解析部 4 6 は処理をステップ SD1 に移行して、発音始点 TS から時間 が経過するか (SD1 : YES)、音響信号 XE の強度 Q が減衰閾値 Z0 を下回るまで (SD2 : YES)、極小点 xL の発生を監視する。

【 0 0 3 5 】

他方、音響信号 XE の強度 Q が減衰閾値 Z0 を下回る以前に極小点 (以下では特に「対象極小点」という) xL が検出されると (SD4 : YES)、終点解析部 4 6 は、対象極小点 xL の直後の極大点 xH が検出されたか否かを判定する (SD5)。極大点 xH が検出されない場合 (SD5 : NO)、終点解析部 4 6 は処理をステップ SD1 に移行する。なお、対象極小点 xL が検出された場合 (SD4 : YES) に、当該対象極小点 xL での強度 Q が現時点の基準値 QREF を下回るときには ($Q < QREF$)、基準値 QREF が当該対象極小点 xL での強度 Q に更新される。すなわち、基準値 QREF は、前述の通り、発音始点 TS 以降における強度 Q の最小値 (例えば対象極小点 xL での強度 Q) となるように更新される。

【 0 0 3 6 】

図 5 には、対象極小点 xL の直後の極大点 xH2 が例示されている。極大点 xH2 が検出されると (SD5 : YES)、終点解析部 4 6 は、当該極大点 xH2 での強度 QH と現時点での基準値 QREF との差分 ($QH - QREF$) に応じた変動指標 が終点閾値 ZE を上回るか否かを判定する (SD6)。変動指標 は、前述の通り、強度 QH と基準値 QREF との差分 ($QH - QREF$) を当該強度 QH で除算した数値である。なお、現時点の基準値 QREF は、対象極小点 xL での強度 Q である可能性が高い。また、終点閾値 ZE は、発音始点 TS の特定に利用される前述の始点閾値 ZS を下回る所定の正数に設定される ($ZE < ZS$)。

【 0 0 3 7 】

変動指標 が終点閾値 ZE を下回る場合 (SD6 : NO) には、対象極小点 xL の直後に極大点 xH が観測されたものの発音始点 TS の直後の発音源の発音による強度 Q の増加 (第 2 回目以降の発音) とまでは推定できない。したがって、発音終点 TE をまだ確定せずに引続き音響信号 XE の強度 Q を監視する必要がある。そこで、終点解析部 4 6 は、処理をステップ SD1 に移行して、発音始点 TS から時間 が経過するか (SD1 : YES)、音響信号 XE の強度 Q が減衰閾値 Z0 を下回るまで (SD2 : YES)、極小点 xL の発生を監視する。

【 0 0 3 8 】

他方、変動指標 が終点閾値 ZE を上回るほど極大点 xH の強度 Q が増加した場合 (SD6 : YES) には、対象極小点 xL の直後の極大点 xH は、発音始点 TS の直後の発音源の発音 (すなわち最初の発音の直後の第 2 回目以降の発音) による強度 Q の増加と推定される。したがって、発音始点 TS から対象極小点 xL までを発音区間 P として確定し、第 2 回目以降の発音に対応する直後の極大点 xH は発音区間 P から除外する必要がある。そこで、終点解析部 4 6 は、対象極小点 xL を発音終点 TE として特定する (SD7)。すなわち、対象極小点 xL の直後の極大点 xH について変動指標 が終点閾値 ZE を上回る場合に、当該対象極小点 xL が事後的に発音終点 TE として確定される。

【 0 0 3 9 】

以上の説明から理解される通り、第 1 実施形態の終点解析部 4 6 は、発音始点 TS の経過後に音響信号 XE の強度 Q が経時的に減少する過程で検出される対象極小点 xL を、変動指標 が終点閾値 ZE を上回る場合 (SD6 : YES) に発音終点 TE として特定し (SD7)、変動指標 が終点閾値 ZE を下回る場合 (SD6 : NO) には発音終点 TE としない。なお、対象極小点 xL の直後に検出された図 5 の極大点 xH2 については、図 6 を参照して説明した通り、変動指標 が始点閾値 ZS を上回ることを条件として発音始点 TS として特定される。変動指標 が始点閾値 ZS を上回る場合には終点閾値 ZE も当然に上回るから、当該極大点 xH の直前の対象極小点 xL は発音終点 TE として確定される。

【 0 0 4 0 】

10

20

30

40

50

他方、音響信号 XE の強度 Q が減衰閾値 Z0 を下回る (SD2: YES) ことも、発音始点 TS の経過後の極小点 xL が発音終点 TE として特定される (SD7) こともなく、直前の発音始点 TS から時間 t が経過すると (SD1: YES)、終点解析部 46 は、発音始点 TS から時間 t が経過した時点 を発音終点 TE として特定する (SD8)。以上の説明から理解される通り、終点解析部 46 は、基本的には音響信号 XE の強度 Q が減衰閾値 Z0 を下回る時点 を発音終点 TE として特定する一方 (SD3)、発音始点 TS の直後の発音源の発音が推定される場合 (SD6: YES) には、発音区間 P から当該発音が除外されるように極小点 xL を発音終点 TE として確定し (SD7)、何れの条件も成立しない場合には発音始点 TS から時間 t が経過した時点 を発音終点 TE として特定する (SD8)。

【0041】

10

以上に説明した通り、第1実施形態では、発音始点 TS の経過後に音響信号 XE の強度 Q が経時的に減少する過程で強度 Q が増加に反転する極小点 xL を、変動指標 ΔQ が終点閾値 ZE を上回る場合に発音終点 TE として特定する。すなわち、発音源が短い間隔で複数回にわたり発音した場合 (最初の発音による音響が十分に減衰する以前に直後の発音が開始する場合) には、発音始点 TS に対応する最初の発音のみを発音区間 P が包含するように発音終点 TE が特定される。したがって、音響信号 XA の解析に重要な発音直後の区間を発音区間 P として高精度に特定することが可能である。音源識別部 60 による発音源の識別には、発音源の種類毎の相違が顕著となる発音直後の特性が特に重要である。したがって、発音直後の区間を発音区間 P として高精度に特定できる第1実施形態は格別に好適である。

20

【0042】

また、第1実施形態では、変動指標 ΔQ が終点閾値 ZE を上回る極小点 xL の到来前に、発音始点 TS での強度 QH に応じた減衰閾値 Z0 を下回るまで音響信号 XE の強度 Q が発音始点 TS と比較して減少した場合 (SD2: YES) に、当該強度 Q が減衰閾値 Z0 を下回る時点が発音終点 TE として特定される。したがって、発音始点 TS の経過後に発音源が発音することなく音響信号 XE が減衰する場合に、発音始点 TS からの減衰の割合に応じた適切な発音終点 TE を設定できるという利点がある。

【0043】

第1実施形態では、音響信号 XE の強度 Q の極大点 xH を順次に検出する一方、極大点 xH での強度 QH と当該極大点 xH までの強度 Q の最小値である基準値 QREF との差分 (QH - QREF) に応じた変動指標 ΔQ が始点閾値 ZS を上回る場合に、当該極大点 xH が発音始点 TS として特定される。したがって、音響信号 XE から検出される複数の極大点 xH のうち発音源の明瞭な発音の開始を発音始点 TS として高精度に特定できるという利点がある。

30

【0044】

また、極大点 xH での強度 QH と当該極大点 xH までの強度 Q の最小値である基準値 QREF との差分 (QH - QREF) を極大点 xH での強度 QH により除算することで変動指標 ΔQ が算定される。すなわち、差分 (QH - QREF) が音響信号 XE の音量の大小に依存しない数値に正規化される。したがって、音響信号 XE の音量に関わらず発音始点 TS および発音終点 TE を適切に特定することが可能である。

【0045】

40

< 音源識別部 60 >

図8は、第1実施形態の音源識別部 60 の構成図である。図8に例示される通り、第1実施形態の音源識別部 60 は、調波性解析部 62 と第1解析部 64 と第2解析部 66 と音源特定部 68 とを具備する。

【0046】

調波性解析部 62 は、音響信号 XA が表す音響 (以下「対象音」という) が調波音および非調波音の何れに該当するかを音響信号 XA の特徴量 F から解析する。第1実施形態の調波性解析部 62 は、対象音が調波音に該当する確度 WH (第1確度) と対象音が非調波音に該当する確度 WP (第2確度) とを算定する。

【0047】

50

具体的には、特徴量 F の解析で調波音と非調波音とを判別する公知のパターン認識器が調波性解析部 6 2 として任意に利用される。第 1 実施形態では、教師あり学習を利用した統計モデルの代表例であるサポートベクターマシン (SVM : Support Vector Machine) を調波性解析部 6 2 として例示する。すなわち、調波性解析部 6 2 は、調波音と非調波音とを含む多数の音響の学習データを適用した機械学習で事前に決定された超平面を利用して、特徴量 F の対象音が調波音および非調波音の何れに該当するかを特徴量 F 毎 (発音区間 P 毎) に順次に判別する。そして、調波性解析部 6 2 は、例えば所定の期間内に対象音が調波音であると判別した回数の比率 (調波音と判別した回数 / 当該期間内の判別の総回数) を調波音の確度 WH として算定する一方、対象音が非調波音であると判別した回数の比率を非調波音の確度 WP として算定する (WH + WP = 1)。以上の説明から理解される通り、音響信号 XA の対象音が調波音である可能性 (尤度) が高いほど確度 WH は大きい数値となり、対象音が非調波音である可能性が高いほど確度 WP は大きい数値となる。

10

【 0 0 4 8 】

第 1 解析部 6 4 は、音響信号 XA の対象音の発音源が複数種の調波音源の何れに該当するかを音響信号 XA の特徴量 F から解析する。調波音源は、調波音を発音する発音源 (例えば調波楽器) を意味する。図 8 では、ベース (Bass) , ギター (Guitar) , 男性歌唱者 (male Vo.) , 女性歌唱者 (female Vo.) の 4 種類が、対象音の発音源の候補となる調波音源として例示されている。具体的には、第 1 実施形態の第 1 解析部 6 4 は、N 種類 (N は 2 以上の自然数) の調波音源の各々について、対象音の発音源が当該調波音源に該当する確度に応じた評価値 EH(n) (EH(1) ~ EH(N)) を設定する。

20

【 0 0 4 9 】

図 9 は、第 1 解析部 6 4 が評価値 EH(1) ~ EH(N) を設定する処理 (以下「調波解析処理」という) のフローチャートである。特徴量抽出部 5 0 による特徴量 F の抽出毎 (したがって発音区間 P 毎) に図 9 の調波解析処理が実行される。

【 0 0 5 0 】

調波解析処理を開始すると、第 1 解析部 6 4 は、事前に選定された N 種類の調波音源から任意の 2 種類の調波音源を選択する全通り ($_N C_2$ 通り) の組合せの各々について、対象音の発音源が当該組合せの 2 種類の調波音源の何れに該当するかを、特徴量 F を利用して判別する (SA1)。以上の判別には、2 種類の調波音源を判別候補とするサポートベクターマシンが好適に利用される。すなわち、調波音源の組合せに相当する $_N C_2$ 通りのサポートベクターマシンに特徴量 F を適用することで、当該組合せ毎に対象音の発音源が 2 種類の調波音源から選択される。

30

【 0 0 5 1 】

第 1 解析部 6 4 は、N 種類の調波音源の各々について、対象音の発音源が当該調波音源に該当する確度 CH(n) (CH(1) ~ CH(N)) を算定する (SA2)。任意の 1 個 (第 n 番目) の調波音源の確度 CH(n) は、例えば、合計 $_N C_2$ 回にわたる判別のうち対象音の発音源が第 n 番目の調波音源に該当すると判別された回数の比率 (調波音源に該当すると判別された回数 / $_N C_2$) である。以上の説明から理解される通り、音響信号 XA の対象音の発音源が N 種類のうち第 n 番目の調波音源に該当する可能性 (尤度) が高いほど確度 CH(n) は大きい数値となる。

40

【 0 0 5 2 】

第 1 解析部 6 4 は、調波音源毎に算定された確度 CH(n) の順位に対応した数値 (得点) を評価値 EH(n) として N 種類の調波音源の各々について設定する (SA3)。具体的には、確度 CH(n) が大きいほど評価値 EH(n) が大きい数値となるように確度 CH(n) の順位に応じた数値が各調波音源の評価値 EH(n) に付与される。例えば、確度 CH(n) の降順で最上位に位置する調波音源の評価値 EH(n) は数値 1 (例えば 1 = 1 0 0) に設定され、確度 CH(n) が第 2 位に位置する調波音源の評価値 EH(n) は数値 2 (例えば 2 = 8 0) に設定され、確度 CH(n) が第 3 位に位置する調波音源の評価値 EH(n) は数値 3 (例えば 3 = 6 0) に設定され、所定の順位を下回る残余の調波音源の評価値 EH(n) は最小値 (例えば 0) に設定される、という具合である。以上の説明から理解

50

される通り、音響信号 XA の対象音の発音源が N 種類のうち第 n 番目の調波音源に該当する可能性が高いほど評価値 EH(n) は大きい数値となる。以上が調波解析処理の好適例である。

【 0 0 5 3 】

図 8 の第 2 解析部 6 6 は、音響信号 XA の対象音の発音源が複数種 of 非調波音源の何れに該当するかを音響信号 XA の特徴量 F から解析する。非調波音源は、非調波音を発音する発音源（例えば打楽器等の非調波楽器）を意味する。図 8 では、バスクラム（Kick）、スネアドラム（Snare）、ハイハット（Hi-Hat）、フロアタム（F-Tom）、シンバル（Cymbal）の 5 種類が、対象音の発音源の候補となる非調波音源として例示されている。具体的には、第 1 実施形態の第 2 解析部 6 6 は、M 種類（M は 2 以上の自然数）の非調波音源の各々について、対象音の発音源が当該非調波音源に該当する確度に応じた評価値 EP(m)（EP(1) ~ EP(M)）を設定する。なお、調波音源の種類数 N と非調波音源の種類数 M との異同は不問である。

10

【 0 0 5 4 】

第 2 解析部 6 6 による M 個の評価値 EP(1) ~ EP(M) の設定（非調波解析処理）は、図 9 に例示した調波解析処理（第 1 解析部 6 4 による評価値 EH(n) の設定）と同様である。具体的には、第 2 解析部 6 6 は、M 種類の非調波音源から 2 種類を選択する全通り（ ${}_M C_2$ 通り）の組合せの各々について、対象音の発音源が当該組合せの 2 種類の非調波音源の何れに該当するかを判別し、対象音の発音源が第 m 番目の非調波音源に該当する確度 CP(m) を非調波音源毎に算定する。非調波音源の判別には、調波解析処理での調波音源の判別と同様にサポートベクターマシンが好適に利用される。

20

【 0 0 5 5 】

そして、第 2 解析部 6 6 は、M 種類の非調波音源の各々について、確度 CP(m) の順位に対応した数値を評価値 EP(m) として設定する。確度 CP(m) の任意の順位に位置する非調波音源の評価値 EP(m) には、確度 CH(n) の順番で同順位に位置する調波音源の評価値 EH(n) と同等の数値が付与される。具体的には、確度 CP(m) の降順で最上位に位置する非調波音源の評価値 EP(m) は数値 1 に設定され、確度 CP(m) が第 2 位に位置する非調波音源の評価値 EP(m) は数値 2 に設定され、確度 CP(m) が第 3 位に位置する非調波音源の評価値 EP(m) は数値 3 に設定され、所定の順位を下回る残余の調波音源の評価値 EP(m) は最小値（例えば 0）に設定される。したがって、音響信号 XA の対象音の発音源が M 種類のうち第 m 番目の非調波音源に該当する可能性（尤度）が高いほど評価値 EP(m) は大きい数値となる。

30

【 0 0 5 6 】

特徴量抽出部 5 0 が音響信号 XA から抽出する任意の 1 個の特徴量 F は、前述の通り、相異なる特性値 f1（第 1 特性値）および特性値 f2（第 2 特性値）を含む複数の特性値 f で構成される。第 1 実施形態の第 1 解析部 6 4 は、特徴量 F の特性値 f1 を利用して、対象音の発音源が N 種類の調波音源の各々に該当する確度 CH(n) を解析する。他方、第 2 解析部 6 6 は、特徴量 F の特性値 f2 を利用して、対象音の発音源が M 種類の非調波音源の各々に該当する確度 CP(m) を解析する。すなわち、第 1 解析部 6 4 が調波音源の確度 CH(n) の算定に利用する特徴量 F（特性値 f1）と第 2 解析部 6 6 が非調波音源の確度 CP(m) の算定に適用する特徴量 F（特性値 f2）とは相違する。

40

【 0 0 5 7 】

具体的には、第 1 解析部 6 4 による確度 CH(n) の算定には、調波音源の種類毎に相違が顕著となる特性値 f1 が利用される。例えば、音色を表す MFCC や、基音成分に対する倍音成分の強度比等の特性値 f1 が、調波音の確度 CH(n) の算定に好適に利用される。他方、第 2 解析部 6 6 による確度 CP(m) の算定には、非調波音源の種類毎に相違が顕著となる特性値 f2 が利用される。例えば、音響の立上がりの急峻度や零交差数等の特性値 f2 が、非調波音の確度 CP(m) の算定に好適に利用される。なお、第 1 解析部 6 4 が利用する特性値 f1 と第 2 解析部 6 6 が利用する特性値 f2 とを部分的に共通させることも可能である。

50

【 0 0 5 8 】

図 8 の音源特定部 6 8 は、調波性解析部 6 2 と第 1 解析部 6 4 と第 2 解析部 6 6 とによる以上の解析の結果に応じて音響信号 X A の発音源の種類を特定する。発音源の種類の特
定は発音区間 P 毎に実行される。図 8 に例示される通り、第 1 実施形態の音源特定部 6 8
は、乗算部 6 8 2 と乗算部 6 8 4 と選択処理部 6 8 6 とを包含する。

【 0 0 5 9 】

乗算部 6 8 2 は、第 1 解析部 6 4 が N 種類の調波音源について設定した N 個の評価値 E
H(1) ~ E H(N) の各々に、調波性解析部 6 2 が解析した調波音の確度 W H を乗算することで
N 個の識別指標 R ($R = E H(n) \times W H$) を算定する。他方、乗算部 6 8 4 は、第 2 解析部
6 6 が M 種類の非調波音源について設定した M 個の評価値 E P(1) ~ E P(M) の各々に、調波
性解析部 6 2 が解析した非調波音の確度 W P を乗算することで M 個の識別指標 R ($R = E P$
(m) \times W P) を算定する。乗算部 6 8 2 および乗算部 6 8 4 の処理により、N 種類の調波音
源と M 種類の非調波音源とを含む K 種類 ($K = N + M$) の候補音源の各々について識別指
標 R が算定される。以上の説明から理解される通り、確度 W H は、調波音の各評価値 E H(n
) に対する加重値に相当し、確度 W P は、非調波音の各評価値 E P(m) に対する加重値に相当
する。対象音が調波音に該当する確度 W H が大きいほど調波音源の識別指標 R が相対的に
優勢となり、対象音が非調波音に該当する確度 W P が大きいほど非調波音源の識別指標 R
が相対的に優勢となる。

【 0 0 6 0 】

選択処理部 6 8 6 は、乗算部 6 8 2 および乗算部 6 8 4 が算定した K 個の識別指標 R に
応じて音響信号 X A の対象音の発音源の種類を特定し、当該発音源の種類を示す音源識別
情報 D (例えば楽器名) を生成する。具体的には、選択処理部 6 8 6 は、K 種類の候補音
源のうち識別指標 R が最大となる 1 種類の候補音源を対象音の発音源として選択し、当該
候補音源を指定する音源識別情報 D を生成する。すなわち、音響信号 X A の対象音の発音
源の種類が識別される。以上に例示した処理が複数の音響信号 X A の各々について実行さ
れることで、対象音の発音源の種類を示す音源識別情報 D が音響信号 X A 毎に生成される
。音響解析部 2 0 の具体例は以上の通りである。

【 0 0 6 1 】

図 1 の音響処理部 3 0 は、調波性解析部 6 2 が音響信号 X A 毎に解析した音源識別情報
D を参照して複数の音響信号 X A に音響処理を実行することで音響信号 X B を生成する。具
体的には、音響信号 X A の音源識別情報 D が示す発音源の種類毎に事前に設定された音響
処理が当該音響信号 X A に対して実行される。音響信号 X A に対する音響処理としては、例
えば残響効果や歪効果等の各種の音響効果を付与する効果付与処理 (エフェクタ) や、周
波数帯域毎の音量を調整する特性調整処理 (イコライザ) , 音像が定位する位置を調整す
る定位調整処理 (パン) , 音量を調整する音量調整処理が例示される。効果付与処理で音
響信号 X A に付与される音響効果の種類や度合, 特性調整処理で音響信号 X A に付与される
周波数特性, 定位調整処理で調整される音像の位置, 音量調整処理による調整内容 (ゲイ
ン) 等の各種のパラメータが、音源識別情報 D が示す発音源の種類毎に個別に設定される
。そして、音響処理部 3 0 は、以上に例示した音響処理後の複数の音響信号 X A を混合 (
ミキシング) することで音響信号 X B を生成する。すなわち、第 1 実施形態の音響処理部
3 0 は、調波性解析部 6 2 による発音源の識別結果を反映した自動ミキシングを実現する
。

【 0 0 6 2 】

図 1 0 は、第 1 実施形態の音源識別部 6 0 が任意の 1 系統の音響信号 X A について対象
音の発音源の種類を特定する処理 (以下「音源識別処理」という) のフローチャートであ
る。複数の音響信号 X A の各々について、特徴量抽出部 5 0 による特徴量 F の抽出毎 (発
音区間 P 毎) に図 1 0 の音源識別処理が実行される。

【 0 0 6 3 】

音源識別処理を開始すると、調波性解析部 6 2 は、音響信号 X A が表す対象音が調波音
および非調波音の何れに該当するかを音響信号 X A の特徴量 F から解析する (S B 1) 。他

方、第1解析部64は、図9を参照して説明した調波解析処理によりN種類の調波音源の各々について評価値EH(n) (EH(1) ~ EH(N))を算定し(SB2)、第2解析部66は、調波解析処理と同様の非調波解析処理によりM種類の非調波音源の各々について評価値EP(m) (EP(1) ~ EP(M))を算定する(SB3)。そして、音源特定部68は、調波性解析部62と第1解析部64と第2解析部66とによる以上の解析の結果に応じて音響信号XAの発音源の種類を特定する(SB4)。なお、調波性解析部62による調波性の解析と、第1解析部64による調波解析処理と、第2解析部66による非調波解析処理との順序は任意である。例えば調波解析処理(SB2)および非調波解析処理(SB3)の実行後に調波性解析部62が調波性を解析することも可能である。

【0064】

以上に説明した通り、第1実施形態では、調波音と非調波音とを相互に区別して対象音の発音源の種類が特定される。具体的には、対象音が調波音および非調波音の各々に該当する確度(WH, WP)を調波性解析部62が解析した結果と、対象音の発音源がN種類の調波音源の各々に該当する確度CH(n)を第1解析部64が解析した結果と、対象音の発音源がM種類の非調波音源の各々に該当する確度CP(m)を第2解析部66が解析した結果とを利用して、対象音の発音源の種類が特定される。したがって、調波音と非調波音とを区別せずに発音源の種類を特定する構成と比較して対象音の発音源の種類を高精度に特定することが可能である。第1解析部64や第2解析部66の未学習の発音源についても音響処理部30による調波音/非調波音の識別は可能であるという利点もある。

【0065】

また、第1実施形態では、対象音が調波音に該当する確度WHと各調波音源の評価値EH(n)との乗算、および、対象音が非調波音に該当する確度WPと各非調波音源の評価値EP(m)との乗算により、K種類の候補楽器(N種類の調波音源およびM種類の非調波音源)の各々について識別指標Rが算定され、各識別指標Rに応じて対象音の発音源の種類が特定される。すなわち、対象音が調波音に該当する確度WHが大きいほど調波音源の識別指標Rが相対的に優勢となり、対象音が非調波音に該当する確度WPが大きいほど非調波音源の識別指標Rが相対的に優勢となる。したがって、K個の識別指標Rの比較により対象音の発音源の種類を簡便かつ高精度に特定できるという利点がある。

【0066】

ところで、例えば対象音の発音源が調波音源に該当する確度CH(n)を評価値EH(n)として利用するとともに対象音の発音源が非調波音源に該当する確度CP(m)を評価値EP(m)として利用する構成(以下「比較例」という)では、評価値EH(n)の数値が調波音源の種類数Nに依存するとともに評価値EP(m)の数値が非調波音源の種類数Mに依存する。例えば、調波音源の種類数Nが多いほど確度CH(n)は小さい数値となる。したがって、調波音源の種類数Nと非調波音源の種類数Mとが相違する場合には、評価値EH(n)と評価値EP(m)とを適切に比較できないという問題がある。第1実施形態では、対象音の発音源が調波音源に該当する確度CH(n)の順位に応じた数値が評価値EH(n)として調波音源毎に設定され、対象音の発音源が非調波音源に該当する確度CP(m)の順位に応じた数値が評価値EP(m)として非調波音源毎に設定される。すなわち、評価値EH(n)は調波音源の種類数Nに依存しない数値に設定され、評価値EP(m)は非調波音源の種類数Mに依存しない数値に設定される。したがって、第1実施形態によれば、例えば調波音源の種類数Nと非調波音源の種類数Mとが相違する場合でも評価値EH(n)と評価値EP(m)とを適切に比較できるという利点がある。調波音源の種類数Nおよび非調波音源の種類数Mの制約が緩和されると換言することも可能である。ただし、前述の比較例も本発明の範囲には包含される。

【0067】

また、第1実施形態では、第1解析部64が調波音源の確度CH(n)の算定に利用する特徴量F(特性値f1)と第2解析部66が非調波音源の確度CP(m)の算定に適用する特徴量F(特性値f2)とが相違する。具体的には、例えば第1解析部64による確度CH(n)の算定には調波音の識別に好適な特性値f1が利用され、第2解析部66による確度CP(m)の算定には非調波音の識別に好適な特性値f2が利用される。したがって、調波音源の確

10

20

30

40

50

度 $CH(n)$ の算定と非調波音源の確度 $CP(m)$ の算定とに同種の特徴量を利用する構成と比較して、対象音の発音源を高精度に特定できるという利点がある。ただし、第1解析部64と第2解析部66とが共通の特徴量 F を利用することも可能である。

【0068】

<第2実施形態>

本発明の第2実施形態を説明する。なお、以下に例示する各形態において作用や機能が第1実施形態と同様である要素については、第1実施形態の説明で使用した符号を流用して各々の詳細な説明を適宜に省略する。

【0069】

図11に例示される通り、発音源による1回の発音（例えば打楽器の1回の打撃による発音）の開始直後に強度 Q が増加する過程において複数回の極大点 xH （ $xH1$ 、 $xH2$ ）が観測される場合がある。図11の極大点 $xH1$ の変動指標は始点閾値 ZS を上回るから、変動指標が始点閾値 ZS を上回る全部の極大点 xH を発音始点 TS として確定する第1実施形態では、極大点 $xH1$ および極大点 $xH2$ の双方が発音始点 TS として特定される。しかし、音響信号 XE の強度は、極大点 $xH1$ の直後の極大点 $xH2$ まで増加する。すなわち、極大点 $xH1$ および極大点 $xH2$ は、実際には発音源の1回の発音に対応すると推定される。したがって、極大点 $xH1$ を発音始点 TS として特定することなく直後の極大点 $xH2$ のみを発音始点 TS として特定し、極大点 $xH1$ および極大点 $xH2$ の双方を1個の発音区間 P に包含させるべきである。以上の事情を考慮して、第2実施形態では、音響信号 XE の強度 Q の1個の極大点 $xH1$ の直後に、当該極大点 $xH1$ を上回る強度 Q の極大点 $xH2$ を検出した場合に、先行の極大点 $xH1$ を発音始点 TS の候補から除外する。

【0070】

具体的には、変動指標が始点閾値 ZS を上回る任意の1個の極大点 $xH1$ （第1極大点）を第1実施形態と同様の方法で検出すると、始点解析部44は、図11に例示される通り、当該極大点 $xH1$ に対応する時間軸上の位置に待機区間 V を設定する。待機区間 V は、極大点 $xH1$ を発音始点 TS として確定することを留保する区間であり、極大点 $xH1$ 以降に設定される。第2実施形態の始点解析部44は、極大点 $xH1$ を始点とする所定長の待機区間 V を設定する。

【0071】

待機区間 V を設定すると、始点解析部44は、極大点 $xH1$ 以降の音響信号 XE について極大点 xH の探索を継続する。前述の通り、音響信号 XE の強度 Q は、極大点 $xH1$ 以降に増加する可能性がある。極大点 $xH1$ を上回る強度の極大点 $xH2$ （第2極大点）を待機区間 V 内に検出した場合、始点解析部44は、先行の極大点 $xH1$ を発音始点 TS の候補から除外する。以上の処理を順次に行い、検出済の極大点 xH を上回る強度の極大点 xH を検出することなく待機区間 V が経過すると、始点解析部44は、待機区間 V の満了前に最後に検出した極大点 xH を発音始点 TS として確定する。

【0072】

以上の説明から理解される通り、第2実施形態では、音響信号 XE の強度 Q の極大点 $xH1$ 以降の待機区間 V 内に、当該極大点 $xH1$ を上回る強度 Q の極大点 $xH2$ が検出された場合に、極大点 $xH1$ が発音始点 TS の候補から除外される。したがって、発音源による1回の発音の開始から音響信号 XE の強度 Q が増加する過程で複数の極大点 xH が検出される場合でも、当該発音に対応した1個の極大点 xH を含む発音区間 P を適切に特定することが可能である。

【0073】

なお、第2実施形態では、1個の極大点 $xH1$ を始点とする待機区間 V を設定したが、極大点 $xH1$ を上回る強度 Q の極大点 $xH2$ を検出した場合に、当該極大点 $xH2$ を始点とする待機区間 V を新規に設定する（すなわち極大点 xH の検出毎に待機区間 V を更新する）ことも可能である。

【0074】

<変形例>

10

20

30

40

50

以上に例示した各態様は多様に変形され得る。具体的な変形の態様を以下に例示する。以下の例示から任意に選択された2個以上の態様は、相互に矛盾しない範囲で適宜に併合され得る。

【0075】

(1) 前述の各形態では、調波性解析部62がサポートベクターマシンにより調波音と非調波音とを判別したが、調波性解析部62による調波音/非調波音の判別方法は以上の例示に限定されない。例えば、調波音および非調波音の各々の特徴量Fの分布傾向を表現する混合正規分布を利用して対象音を調波音と非調波音とに判別する方法や、K-meansアルゴリズムを利用したクラスタリングで対象音を調波音と非調波音とに判別する方法も採用され得る。第1解析部64および第2解析部66の各々が対象音の発音源の種類を推定する

10

【0076】

(2) 前述の各形態では、調波性解析部62が解析した調波音の確度WHをN個の評価値EH(1)~EH(N)に乘算するとともに非調波音の確度WPをM個の評価値EP(1)~EP(M)に乘算したが、調波音の確度WHおよび非調波音の確度WPを音響信号XAの発音源の種類に反映させる方法は以上の例示に限定されない。例えば、音響信号XAの対象音が調波音および非調波音の何れに該当するかを確度WHおよび確度WPに応じて判別し、N個の評価値EH(1)~EH(N)およびM個の評価値EP(1)~EP(M)の何れかを調波性の判別結果に応じて選択的に利用して、音源特定部68が発音源の種類を特定することも可能である。

20

【0077】

具体的には、調波性解析部62は、確度WHが確度WPを上回る場合には対象音を調波音と判別し、確度WPが確度WHを上回る場合には対象音を非調波音と判別する。音源特定部68は、対象音が調波音であると判別された場合には、第1解析部64が算定したN個の評価値EH(1)~EH(N)のなかの最大値に対応する調波音源を発音源の種類として特定する一方、対象音が非調波音であると判別された場合には、第2解析部66が算定したM個の評価値EP(1)~EP(M)のなかの最大値に対応する非調波音源を発音源の種類として特定する。以上に例示した構成は、前述の各形態において、確度WHおよび確度WPの一方を1に設定するとともに他方を0に設定した構成とも換言される。なお、対象音が調波音であると調波性解析部62が判別した場合に第2解析部66による非調波解析処理(M個の評価値EP(1)~EP(M)の算定)を省略する構成や、対象音が非調波音であると調波性解析部62が解析した場合に第1解析部64による調波解析処理(N個の評価値EH(1)~EH(N)の算定)を省略する構成も採用され得る。

30

【0078】

以上の例示から理解される通り、音源特定部68は、調波性解析部62と第1解析部64と第2解析部66とによる解析結果に応じて対象音の発音源の種類を特定する要素として包括的に表現され、第1解析部64および第2解析部66の双方の解析結果を利用するか一方の解析結果のみを利用するかは、本発明において不問である。

【0079】

(3) 前述の各形態では始点閾値ZSを固定値としたが、始点閾値ZSを可変値とすることも可能である。例えば、極大点xHでの音響信号XEの強度QHに応じた数値(例えば強度QHを所定値に乘算した数値)を始点閾値ZSとして利用し、図6のステップSC1では、極大点xHでの強度QHと基準値QREFとの差分(QH-QREF)を変動指標として始点閾値ZSと比較することも可能である。終点閾値ZEについても同様に可変値とすることが可能である。また、始点閾値ZSまたは終点閾値ZEを利用者からの指示に応じて可変に設定することも可能である。

40

【0080】

(4) 移動体通信網やインターネット等の通信網を介して端末装置(例えば携帯電話機やスマートフォン)と通信するサーバ装置で音響処理装置12を実現することも可能である。具体的には、音響処理装置12は、端末装置から通信網を介して受信した複数の音響信

50

号XAから前述の各形態と同様の処理で音響信号XBを生成して端末装置に送信する。なお、端末装置から受信した複数の音響信号XAの各々の発音源の種類(音源識別情報D)を音響解析部20が識別して端末装置に通知し、端末装置に搭載された音響処理部30が識別結果に応じて複数の音響信号XAから音響信号XBを生成することも可能である。すなわち、音響処理部30は音響処理装置12から省略され得る。また、音響信号XAの発音区
間Pの発音始点TSおよび発音終点TEを端末装置に通知する構成(例えば端末装置が特徴
量抽出部50および音源解析部20を具備する構成)では、音響処理装置12の音響解析
部20から特徴量抽出部50と音源識別部60とが省略される。

【0081】

以上の説明から理解される通り、本発明の好適な態様は、音響信号XAのうち音響の発
音を開始される発音始点TSと当該音響の発音が終了する発音終点TEとを解析する装置(音響解析装置)として包括的に表現される。音響解析装置における音響処理部30の有無は不問である。

10

【0082】

(5) 前述の各形態で例示した音響処理装置12は、前述の通り制御装置122とプログラ
ムとの協働で実現される。プログラムは、コンピュータが読取可能な記録媒体に格納さ
れた形態で提供されてコンピュータにインストールされ得る。記録媒体は、例えば非一過
性(non-transitory)の記録媒体であり、CD-ROM等の光学式記録媒体(光ディスク
)が好例であるが、半導体記録媒体や磁気記録媒体等の公知の任意の形式の記録媒体を包
含し得る。また、以上に例示したプログラムは、通信網を介した配信の形態で提供されて
コンピュータにインストールされ得る。

20

【0083】

(6) 本発明は、前述の各形態に係る音響処理装置12の動作方法としても特定される。
例えば、発音区間検出部40が音響信号XAのうち音響の発音を開始される発音始点TSと
当該音響の発音が終了する発音終点TEとを解析する方法(音響解析方法)においては、
コンピュータ(単体の装置のほか、相互に別体の複数の装置で構成されたコンピュータシ
ステムも含む)が、音響信号XAの強度Qの極大点xHを発音始点TSとして特定し(図6
の始点解析処理)、発音始点TSの経過後に音響信号XAの強度Qが経時的に減少する過程
で強度Qが増加に反転する極小点xLを、増加後の極大点xHでの強度QHと当該極大点xH
までの強度Qの最小値QREFとの差分に応じた変動指標が終点閾値ZEを上回る場合に発
音終点TEとして特定する(図7の終点解析処理)。

30

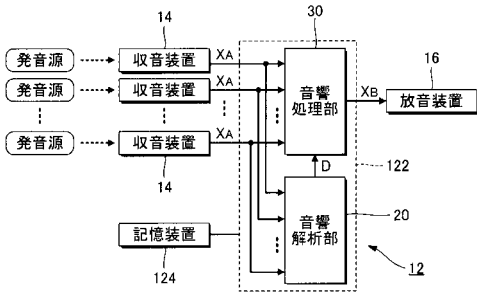
【符号の説明】

【0084】

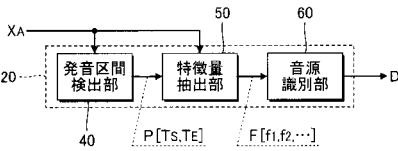
12 …… 音響処理装置、14 …… 收音装置、16 …… 放音装置、122 …… 制御装置、1
24 …… 記憶装置、20 …… 音響解析部、30 …… 音響処理部、40 …… 発音区間検出部
、42 …… 信号処理部、44 …… 始点解析部、46 …… 終点解析部、50 …… 特徴量抽出
部、60 …… 音源識別部、62 …… 調波性解析部、64 …… 第1解析部64、66 …… 第
2解析部、68 …… 音源特定部、682 …… 乗算部、684 …… 乗算部、686 …… 選択
処理部。

40

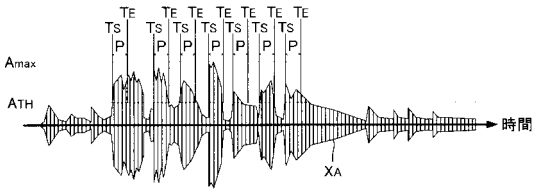
【図1】



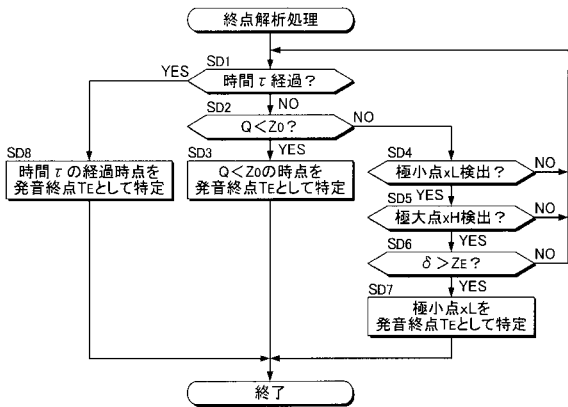
【図2】



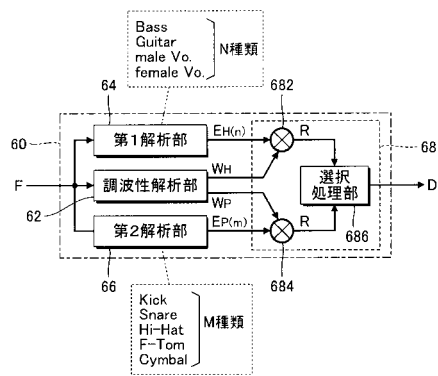
【図3】



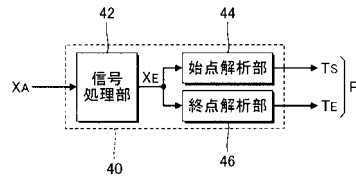
【図7】



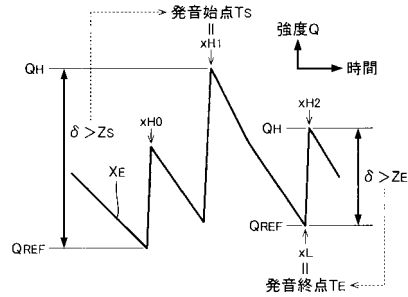
【図8】



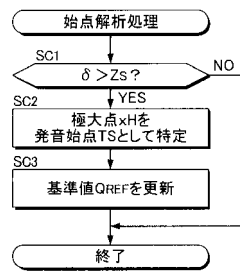
【図4】



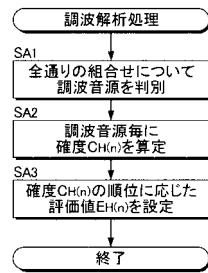
【図5】



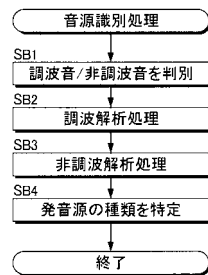
【図6】



【図9】



【図10】



【 図 1 1 】

