(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5362178号
(P5362178)

(45)発行日　平成25年12月11日(2013.12.11)　(24)登録日　平成25年9月13日(2013.9.13)

(51)Int.Cl.		F I		
G10L 25/18	(2013.01)	G10L	11/00	101G
G10L 25/54	(2013.01)	G10L	11/00	402B
G10L 15/02	(2006.01)	G10L	15/02	300G

請求項の数 48　(全 19 頁)

(21)出願番号	特願2006-230500 (P2006-230500)
(22)出願日	平成18年8月28日 (2006.8.28)
(65)公開番号	特開2007-65659 (P2007-65659A)
(43)公開日	平成19年3月15日 (2007.3.15)
審査請求日	平成21年8月25日 (2009.8.25)
(31)優先権主張番号	11/219385
(32)優先日	平成17年9月1日 (2005.9.1)
(33)優先権主張国	米国 (US)

(73)特許権者　501438485
ヤフー！　インコーポレイテッド
アメリカ合衆国　カリフォルニア州　９４０８９　サニーヴェイル　ファースト　アヴェニュー　７０１
(74)代理人　100082005
弁理士　熊倉　禎男
(74)代理人　100067013
弁理士　大塚　文昭
(74)代理人　100086771
弁理士　西島　孝喜
(74)代理人　100109070
弁理士　須田　洋之
(74)代理人　100109335
弁理士　上杉　浩

最終頁に続く

(54)【発明の名称】オーディオ信号からの特徴的な指紋の抽出とマッチング

(57)【特許請求の範囲】

【請求項１】

オーディオフレームからオーディオ指紋を抽出するための方法であって、

前記オーディオフレームを複数の周波数バンドにフィルタ処理して、対応する複数のフィルタ処理したオーディオ信号を生成するステップと、

前記フィルタ処理したオーディオ信号を非直線時間スケールで再サンプリングするステップと、

前記周波数バンド毎に前記再サンプリングしたオーディオ信号を変換して、前記周波数バンドに対する特徴ベクトルを生成するステップと、

前記特徴ベクトルのセットに基づいて前記オーディオ指紋を計算するステップと、

を具備する方法。

【請求項２】

前記オーディオフレームを複数の周波数バンドにフィルタ処理するステップは、前記複数の周波数バンド毎の前記オーディオフレームをバンドパスフィルタ処理するステップを含む請求項１の方法。

【請求項３】

前記オーディオフレームを複数の周波数バンドにフィルタ処理するステップは、前記オーディオサンプルに高速フーリエ変換（ＦＦＴ）を実行するステップを含む請求項１の方法。

【請求項４】

前記オーディオフレームは、ＭＰ３フォーマットで格納されるオーディオファイルの一部である請求項１の方法。
【請求項５】
前記フィルタ処理したオーディオ信号は、対数時間スケールで再サンプリングされる請求項１の方法。
【請求項６】
前記周波数バンドは、周波数軸で直線的に離間している請求項１の方法。
【請求項７】
前記周波数バンドはオーバーラップしている請求項１の方法。
【請求項８】
特定周波数バンドの、前記再サンプリングしてフィルタ処理したオーディオ信号を変換するステップは、前記再サンプリングしたオーディオ信号に高速フーリエ変換（ＦＦＴ）を実行するステップで構成される請求項１の方法。
【請求項９】
前記オーディオ指紋を計算するステップは、前記周波数バンドに対する前記特徴ベクトルを差分エンコードするステップを含む請求項１の方法。
【請求項１０】
前記オーディオ指紋に対するインデックス値を計算するステップを更に具備し、前記インデックス値は前記オーディオ指紋の一部を含む請求項１の方法。
【請求項１１】
前記インデックス値は、前記変換したオーディオ信号の低周波数成分のセットに対応する前記オーディオ指紋の一部を含む請求項１０の方法。
【請求項１２】
前記オーディオ指紋の一部を無関連化するステップを更に具備し、前記オーディオ指紋の前記無関連化した部分は、音響モデルにより重要でないと決定された周波数範囲に対応する請求項１の方法。
【請求項１３】
前記音響モデルは心理音響モデルである請求項１２の方法。
【請求項１４】
前記音響モデルはオーディオエンコード処理の特性を模す請求項１２の方法。
【請求項１５】
前記音響モデルは環境特性を模す請求項１２の方法。
【請求項１６】
前記音響モデルはオーディオ信号の特性を模す請求項１２の方法。
【請求項１７】
オーディオ指紋のデータベースを作成するための方法であって、
請求項１に記載された方法に従ってオーディオフレームからオーディオ指紋を抽出することに基づき、複数のオーディオ指紋を生成するステップと、
前記生成された各オーディオ指紋に対するインデックス値を計算するステップであって、前記インデックス値は、前記オーディオ指紋の一部を含むものと、
それらのインデックス値に従って前記データベース内の前記オーディオ指紋をグループ化するステップと
を具備する方法。
【請求項１８】
オーディオ断片のエッジを検出するための方法であって、
前記オーディオ断片を表す入力信号を受信するステップと、
前記受信入力信号の最も最近の部分を有限長バッファに格納するステップと、
前記入力信号を受信している間の複数の時間毎に、請求項１の方法に従って、前記バッファに格納された前記受信した入力信号の一部についての指紋を計算するステップと、
前記計算された指紋それぞれと、前記オーディオ断片と関係付けられる参照指紋との間

のビットエラー率（ＢＥＲ）を計算するステップと、
時間を通じて前記計算されたＢＥＲに基づいて前記オーディオ断片のエッジを識別するステップと、
を具備する方法。
【請求項１９】
前記オーディオ断片のエッジを識別するステップは、
時間を通じて低下していくＢＥＲの期間中に、前記計算されたＢＥＲの一部を一致させる最適ラインを見出すステップと、
前記最適ラインと、ＢＥＲが５０％に等しくなる水平ラインとの交点として前記オーディオ断片の開始エッジを識別するステップと、を含む請求項１８の方法。
【請求項２０】
前記オーディオ断片のエッジを識別するステップは、
時間を通じて増加していくＢＥＲの期間中に、前記計算されたＢＥＲの一部を一致させる最適ラインを見出すステップと、
前記最適ラインと、ＢＥＲが５０％に等しくなる水平ラインとの交点から、前記バッファ長を減算した点として、前記オーディオ断片の終了エッジを識別するステップと、を含む請求項１８の方法。
【請求項２１】
特定の周波数バンドの前記再サンプリングしたオーディオ信号を変換する前記ステップは、前記再サンプリングしたオーディオ信号に離散コサイン変換（ＤＣＴ）を施すことを含む請求項１の方法。
【請求項２２】
前記周波数バンドは、周波数軸に沿って対数的な中間周波数分布を持つことを特徴とする請求項１の方法。
【請求項２３】
前記フィルタ処理したオーディオ信号は指数的時間スケールで再サンプリングされることを特徴とする請求項１の方法。
【請求項２４】
前記オーディオ指紋を計算する前記ステップは、コードブック表に従ってビット値を割り当てることにより前記周波数バンドにつき前記特徴ベクトルをエンコードすることを含む請求項１の方法。
【請求項２５】
前記コードブック表に一組のコードブック値が決定されるようにトレーニングするステップと、
決定された前記一組のコードブック値に基づいて前記コードブック表をチューニングすることと
を更に具備する請求項２４の方法。
【請求項２６】
オーディオフレームからオーディオ指紋を抽出するための方法であって、
前記オーディオフレームに対するスペクトログラムを計算するステップと、
前記スペクトログラム内の複数の周波数バンドに対して非直線時間スケールで前記スペクトログラムをサンプリングするステップと、
前記サンプリングされた周波数バンドそれぞれからの前記サンプルを用いて、長時間特徴ベクトルを抽出するステップと、
前記特徴ベクトルに基づいて前記オーディオ指紋を生成するステップと、
を具備する方法。
【請求項２７】
前記スペクトログラムを対数時間スケールでサンプリングする請求項２６の方法。
【請求項２８】
オーディオフレームからオーディオ指紋を抽出するためのコンピュータプログラム製品

であって、コンピュータに、
前記オーディオフレームを複数の周波数バンドにフィルタ処理して、対応する複数のフィルタ処理したオーディオ信号を生成するステップと、
前記フィルタ処理されたオーディオ信号を非直線時間スケールで再サンプリングするステップと、
前記周波数バンド毎に前記再サンプリングされたオーディオ信号を変換して、前記周波数バンドに対する特徴ベクトルを生成するステップと、
前記特徴ベクトルのセットに基づいて前記オーディオ指紋を計算するステップと、
を実行させるためのコンピュータプログラムコードを含むコンピュータ読み取り可能媒体を含むコンピュータプログラム製品。

【請求項２９】
前記オーディオフレームを複数の周波数バンドにフィルタ処理するステップは、前記複数の周波数バンド毎の前記オーディオフレームをバンドパスフィルタ処理するステップを含む請求項２８のプログラム製品。

【請求項３０】
前記オーディオフレームを複数の周波数バンドにフィルタ処理するステップは、前記オーディオサンプルに高速フーリエ変換（ＦＦＴ）を実行するステップを含む請求項２８のプログラム製品。

【請求項３１】
前記オーディオフレームは、ＭＰ３フォーマットで格納されるオーディオファイルの一部であり、前記フィルタ処理したオーディオ信号は、前記オーディオファイルに関係付けられるＭＰ３ハイブリッドフィルタバンクから得られる請求項２８のプログラム製品。

【請求項３２】
前記フィルタ処理されたオーディオ信号は、対数時間スケールで再サンプリングされる請求項２８のプログラム製品。

【請求項３３】
前記周波数バンドは、周波数軸で直線的に離間している請求項２８のプログラム製品。

【請求項３４】
前記周波数バンドはオーバーラップしている請求項２８のプログラム製品。

【請求項３５】
特定周波数バンドの前記再サンプリングし、フィルタ処理したオーディオ信号を変換するステップは、前記再サンプリングしたオーディオ信号に高速フーリエ変換（ＦＦＴ）を実行するステップを含む請求項２８のプログラム製品。

【請求項３６】
前記オーディオ指紋を計算するステップは、前記周波数バンドに対する前記特徴ベクトルを差分エンコードするステップを含む請求項２８のプログラム製品。

【請求項３７】
前記オーディオ指紋に対するインデックス値を計算するステップを更に含み、前記インデックス値は前記オーディオ指紋の一部を含む請求項２８のプログラム製品。

【請求項３８】
前記インデックス値は、前記変換したオーディオ信号の低周波数成分のセットに対応する前記オーディオ指紋の一部を含む請求項３７のプログラム製品。

【請求項３９】
前記オーディオ指紋の一部を無関連化するステップを更に含み、前記オーディオ指紋の前記無関連化した部分は、音響モデルにより重要でないと決定した周波数範囲に対応する請求項２８のプログラム製品。

【請求項４０】
前記音響モデルは心理音響モデルである請求項３９のプログラム製品。

【請求項４１】
前記音響モデルはオーディオエンコード処理の特性を模す請求項３９のプログラム製品

。
【請求項４２】
前記周波数バンドは、周波数軸に沿って対数的な中間周波数分布を持つことを特徴とする請求項２８のプログラム製品。
【請求項４３】
前記音響モデルは環境特性を模す請求項３９のプログラム製品。
【請求項４４】
前記フィルタ処理したオーディオ信号は指数的時間スケールで再サンプリングされることを特徴とする請求項２８のプログラム。
【請求項４５】
前記音響モデルはオーディオ信号の特性を模す請求項３９のプログラム製品。
【請求項４６】
特定の周波数バンドの前記再サンプリングしたオーディオ信号を変換する前記ステップは、前記再サンプリングしたオーディオ信号に離散コサイン変換（ＤＣＴ）を施すことを含む請求項２８のプログラム製品。
【請求項４７】
前記オーディオ指紋を計算する前記ステップは、コードブック表に従ってビット値を割り当てることにより前記周波数バンドにつき前記特徴ベクトルをエンコードすることを含む請求項２８のプログラム製品。
【請求項４８】
前記コードブック表に一組のコードブック値が決定されるようにトレーニングするステップと、
決定された前記一組のコードブック値に基づいて前記コードブック表をチューニングするステップと
を、更に含むことを特徴する請求項４７のプログラム製品。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、一般にオーディオ信号処理に関し、より詳細には、オーディオ信号から特徴的な指紋を抽出し、かかる指紋のデータベースを検索することに関する。
【背景技術】
【０００２】
ファイルフォーマット、圧縮技術および他のデータ表現方法が多様性に富むので、データ信号を識別したり、それを他と比較したりすることは、技術的に著しく困難な課題となる。例えば、コンピュータ上のデジタル音楽ファイルの場合、歌をエンコードおよび圧縮するためのフォーマットが多数存在する。更に、これらの歌は異なるデータレートでデジタル形式にサンプリングされることが多いので、異なる特性（例えば、異なる波形）を有する。録音されたアナログオーディオはノイズおよび歪みも含む。これらの著しい波形の違いは、ファイルまたは信号の効率的な認識または比較の面で、かかるファイルの直接比較を拙い選択にしてしまう。同様に、ファイルを直接比較しても、異なるフォーマットでエンコードされたメディアを比較（例えば、ＭＰ３およびＷＡＶでエンコードした同一の歌の比較）することはできない。
【０００３】
これらの理由により、インターネット上に配布されるような、メディアおよび他のコンテンツの識別および追跡は、メディアに関する識別情報を含むメタデータ、透かし、または他の何らかのコードを添付することにより行われることが多い。しかし、この添付情報は、不完全、不正確、またはその両方であることが多い。例えば、メタデータは完全であったためしがなく、ファイル名に一貫性があったことはさらにない。加えて、透かし等の手法は侵入性があるので、追加データまたはコードによって元のファイルを変更してしまう。これら手法の他の欠点は、改竄に脆弱なことである。全てのメディアファイルが、メ

タデータまたは透かし等の正確な識別データを含んでいたとしても、その情報を首尾よく削除できたとしたら、ファイルは「ロックが外された」（つまり著作権を侵害される）ことになる。

【０００４】

これらの問題を回避するために、データ信号自体のコンテンツを解析するという考えに基づいて、他の方法が開発されてきた。あるクラスの方法では、オーディオセグメントに対してオーディオ指紋が生成され、その指紋は、元のオーディオを識別するために用いることができるオーディオ関係特性情報を含む。一例では、オーディオ指紋は、オーディオセグメントを識別するデジタルシーケンスを含む。オーディオ指紋を生成する処理は、生成中のその指紋のオーディオの音響的、知覚的特性に基づくことが多い。オーディオ指紋のサイズは、普通は、元のオーディオコンテンツよりずっと小さいので、簡便なツールとして用いて、オーディオコンテンツを識別、比較、および検索することができる。オーディオ指紋は、放送モニタ、オーディオコンテンツの組織化、Ｐ２Ｐネットワークのコンテンツのフィルタ処理、および歌または他のオーディオコンテンツの識別を含む広範な用途に用いることができる。これら様々な領域へ応用する場合、オーディオ指紋技法は、指紋データベース検索アルゴリズムはもとより、指紋抽出に関与するのが普通である。

【０００５】

ほとんどの既存の指紋技法は、周波数領域のオーディオサンプルからのオーディオ特徴抽出に基づいている。オーディオを、先ず複数フレームにセグメント化し、フレーム毎に一セットの特徴を計算する。オーディオ特徴のうち、高速フーリエ変換（ＦＦＴ）係数、メル周波数ケプストラム係数（ＭＦＣＣ）、スペクトル平坦化、シャープネス、線形予測コーディング（ＬＰＣ）係数および変調周波数を用いることができる。通常は、計算した特徴を、微分、平均または分散を用いて、変換される特徴ベクトルに組み立てる。特徴ベクトルを、隠れマルコフモデル、または主成分分析に続く量子化等の、アルゴリズムを用いて、より簡単な表現にマッピングし、オーディオ指紋を生成する。通常、単一オーディオフレームの処理により得られる指紋のサイズは比較的小さく、所望の信頼性程度を伴って元のオーディオシーケンスを十分識別できるほど一意であるとは言えない。指紋の一意性を強化し、よって正しい認識の確率を高める（および偽の陽性率を下げる）ために、約３～５秒のオーディオを表す、より大きなブロックに、小さなサブ指紋を組合せることができる。

【０００６】

フィリップスが開発した一つの指紋技法は、短時間フーリエ変換（ＳＴＦＴ）を用いて、オーディオ信号の１１．８ｍｓｅｃの間隔毎に３２ビットのサブ指紋を抽出する。オーディオ信号を、先ず０．３７秒の長さのオーバーラップするフレームにセグメント化し、フレームを３１／３２のオーバーラップ因子を有するハミング窓により重み付けし、ＦＦＴを用いて周波数領域に変換する。得られた周波数領域データは、横軸を時間、縦軸を周波数とするスペクトログラム（例えば、時間－周波数ダイヤグラム）として表すことができる。フレーム毎のスペクトル（スペクトログラムの列）を、３００Ｈｚから２０００Ｈｚの範囲の対数間隔の、オーバーラップしていない３３の周波数バンドにセグメント化する。バンド毎にスペクトルエネルギーを計算し、３２ビットのサブ指紋を、時間および周波数軸に沿う連続バンドのエネルギー差の符号を用いて生成する。一つのフレーム内の二つのバンド間のエネルギー差が、前のフレームの同一バンド間のエネルギー差より大きい場合、アルゴリズムは、サブ指紋内の対応ビットに対して「１」を出力し、さもなければ対応ビットに「０」を出力する。指紋は、３秒間のオーディオに相当する２５６の後続３２ビットサブ指紋を、単一指紋ブロックに組合せることにより組み立てられる。

【０００７】

普通の種類のオーディオ処理、ノイズおよび歪みに対してはロバストであるように設計されるが、このアルゴリズムは、得られるスペクトルスケーリングが原因で、大きな速度変化に対して十分にロバストであるとは言えない。従って、スケール不変フーリエ－メリン領域内でオーディオ指紋を抽出する修正アルゴリズムが提案された。この修正アルゴリ

ズムは、オーディオフレームを周波数領域に変換した後に実行する追加ステップを含む。これらの追加ステップは、第２フーリエ変換が後に続く、スペクトル対数マッピングのステップを含む。従って、フレーム毎に第１ＦＦＴが適用され、その結果が対数マッピングされてパワースペクトルが得られ、そして第２ＦＦＴが適用される。これは、対数で再サンプリングしたフーリエ変換のフーリエ変換として説明することができ、会話認識で広く用いられる周知のＭＦＣＣ方法と類似している。主たる違いは、フーリエ－メリン変換が全スペクトルの対数マッピングを用いるのに対して、ＭＦＣＣは、メル周波数スケール（人の聴覚システム特性を模した、１ｋＨｚまでは直線で、それ以上高い周波数では対数間隔）に基づいている。

【０００８】

フィリップスアルゴリズムは、２つの連続フレームだけのスペクトル係数を用いてサブ指紋を計算するので、いわゆる短時間解析アルゴリズムのカテゴリに入る。スペクトログラム内のオーバーラップした多数のＦＦＴフレームを用いてスペクトル特徴を抽出する他のアルゴリズムもある。多数の時間フレームの評価に基づく方法のいくつかは、長時間スペクトログラム解析アルゴリズムとして既知である。

【０００９】

例えば、Ｓｕｋｉｔｔａｎｏｎの「コンテンツ識別のための変調スケール解析」、ＩＥＥＥ Ｔｒａｎｓａｃｔｉｏｎ ｏｎ Ｓｉｇｎａｌ Ｐｒｏｃｅｓｓｉｎｇ，５２巻、Ｎｏ．１０（２００４年１０月）、に記載された一つの時間解析アルゴリズムは変調周波数の推定に基づく。このアルゴリズムでは、オーディオをセグメント化し、スペクトログラムをそれについて計算する。次に、スペクトログラムバンド（例えば、スペクトログラムの周波数範囲）毎に、スペクトログラムの時間行（例えば横軸）に沿って第２変換を適用することにより、変調スペクトルを計算する。これは、スペクトログラムの周波数列（例えば縦軸）に沿って第２ＦＦＴを適用する修正フィリップス法とは異なる。この手法において、スペクトログラムをＮ個の周波数バンドにセグメント化し、バンド毎に、同一数Ｎの連続ウエーブレット変換（ＣＷＴ）を計算する。

【００１０】

このアルゴリズムの開発者は、フィリップスアルゴリズムに比較べて優位な性能を主張しているが、既存のアルゴリズムは、依存として幾つかの欠点を示している。例えば、それらアルゴリズムは、歪みのある会話および音楽、特にオーディオを、ＣＥＬＰオーディオコーデックを用いて圧縮している場合（例えば、ＧＳＭ等の、携帯電話オーディオと関係付けられる）、それらを高い信頼性をもって識別するほど十分にロバストであるとは言えない。更にこれらのアルゴリズムは、一般的に、マイクロホン録音と関係付けられるような、ノイズおよびアナログ歪みに敏感である。また、アルゴリズムが、単一種類の歪みのある状態で、オーディオを識別できるとしても、現実世界のシナリオでは、より普通な、そしてそのシナリオによりに近い、多数の歪みの組合せを扱うことができない（例えば、携帯電話と同様に、ノイズの多い多少残響感のある部屋で、マイクロホンにより録音して、ＧＳＭ圧縮を加えたオーディオ）。

【００１１】

従って、実際の用途に適用する場合、既存の指紋法のエラー率は、許容できないほど高く（例えば、偽の陽性や偽の陰性）、生成する指紋は商業化するには余りにも大きく、および／または余りにも速度が低い。従って、現在のオーディオ認識技法が解決できなかった既存の限界を克服する必要がある。

【発明の開示】

【００１２】

従って、本発明は、オーディオ信号の信号コンテンツに基づく、オーディオ信号からの特徴的な指紋の抽出を可能とする。この指紋を、一セットの参照指紋（例えば、データベース内の）に対して一致させて、２つの信号間の、信号同定、または類似性を決定することができる。この指紋抽出アルゴリズムの性質により、それは既存の解決策を悩ませる多くの問題を蒙ることはなく、そのような解決策に比べて、高速、高効率、高精度で、スケ

ール変更可能であるとともに、ロバストである。

【００１３】

オーディオ指紋を生成するための方法の実施の形態では、オーディオ信号をサンプリングし、その信号からスペクトログラム情報を計算する。スペクトログラムを複数の周波数バンドに分割する。バンド毎のシーケンスサンプルを、対数で再サンプリングすることにより、バンドサンプルの対数マッピングを得る。次に、対数マッピングしたバンドサンプルに第２ＦＦＴを適用して、バンド毎の特徴ベクトルを取得する。次いで、オーディオ指紋を特徴ベクトルに基づいて計算する。オーディオ指紋は、コンピュータ可読媒体に格納でき、または伝送可能な信号として瞬時に準備できる。

【００１４】

以前のオーディオ指紋とは異なり、本発明の実施の形態は、時間的に非直線的に（例えば、対数的に）離間した一連の周波数バンドサンプルから長時間の特徴ベクトルを抽出する。以前の方法はスペクトログラムの周波数軸に沿う対数マッピングを用いたが（例えば、フーリエ－メリン変換およびバーク（bark）スケール）、直線時間スケールが用いられている。対照的に、本発明の実施の形態は、サブバンドサンプルを処理するために非直線的（例えば対数的）な時間スケールを用いることにより、指紋抽出、および一致アルゴリズムのロバスト性を著しく改良できる。

【００１５】

例えば、サブバンドサンプルの時間対数マッピングにより、アルゴリズムはオーディオ再生速度、ならびに時間圧縮および時間伸長の変動に強くなる。この理由は、対数再サンプリングにより、再生速度の何らかのスケーリングが対数マッピングスペクトログラムでは直線シフトとなり、直線シフトはＦＦＴにより除去されるからである。このように、オーディオ信号の指紋は、再生速度の変動に関わらず、もしくは時間圧縮または時間伸長に起因する変動がほとんどないか、または全くない。また、対数時間スケールの使用は、第２の時間－周波数ＦＦＴ変換の低周波数分解能を改良する。これにより、スペクトログラム変調スペクトルの解析に用いる複雑なウエーブレット変換の代わりに、単純なＦＦＴを用いることができ、以前の方法と比較べて、実装がより効率的かつ高速になる。

【００１６】

更に、時間を非直線的（例えば対数的）に再スケール化するので、バンド出力フレームは、ほとんどの部分で、解析したオーディオシーケンスの開始を表すサンプルを含む。従って、結果として得られる指紋は、シーケンス開始時に主として配置されるサンプルを用いて生成される。オーディオシーケンスの比較的小さい部分が、結果的に得られる指紋に最も寄与するので、この指紋を用いて、より短いオーディオシーケンスを一致させることができる。例えば、一実施の形態では、５秒間の元のオーディオフレームから生成した指紋を、２倍短いオーディオ断片から採ったサンプルと、高い信頼性をもって一致させることができる。

【００１７】

また、本指紋技法の実施の形態は、ノイズおよび信号歪みにも耐性がある。一実施の形態は、１００％のホワイトノイズの存在下で会話様信号を検出できる（すなわち、信号対雑音比が０ｄＢ）。本技法は、フィルタ処理、圧縮、周波数イコライジングおよび位相歪みにも耐性がある。例えば、本発明の実施の形態は、±５％のピッチ変動（テンポを保存した状態で）および±２０％のタイミング変動（ピッチを保存した状態で）を伴うオーディオを高い信頼性で認識できる。

【００１８】

別の実施の形態では、生成した指紋フレームを、特定数の周波数バンドを用いて形成する場合、音響モデルを用いて重要でない周波数バンドにマークする。重要でないバンドには、オーディオサンプルを区別する際に、知覚できる値を実質的に何も加えないバンドが含まれる。関係する周波数バンドだけを処理することが、信号対雑音比を高め、指紋一致処理全体のロバスト性を改善する。更に、無関係のバンドを除外すると、帯域制限されたオーディオコンテンツの認識効率、例えば、非常に低いビットレートでエンコードした、

または低速テープ速度でアナログ録音した会話の場合、を著しく改善できる。

【００１９】

本実施の形態は、また、高速インデックスおよび大規模データベースにある指紋に対する効率的な検索を提供する。例えば、各オーディオ指紋に対するインデックスは、指紋のコンテンツの一部から計算できる。一実施の形態では、指紋からの一セットのビットを、指紋のインデックスとして用いるが、その場合、非直線的（例えば対数的）な再サンプリングにより、ビットはより安定な低周波数の係数と対応する。試験指紋をデータベース内の一セットの指紋と一致させるために、試験指紋をインデックスに対して一致させて候補の指紋のグループを得ることができる。次いで、試験指紋を候補の指紋と一致させるので、試験指紋をデータベースにある全ての指紋と一致させる必要がない。

【００２０】

別の実施の形態では、エッジ検出アルゴリズムを用いて、解析したオーディオフレームまたは断片の正確なエッジを決定する。用途によっては、特に、オーディオサンプルが、サンプル全体の短い時間の間だけ異なる場合、オーディオサンプル内の解析したオーディオフレームのエッジの配置を知ることが重要である。エッジ検出アルゴリズムは、直線回帰技法を用いて、オーディオフレームのエッジを識別できる。

【００２１】

指紋技法の実施の形態の用途は、数え切れないほど多く、オーディオストリーム、および他のオーディオコンテンツ（例えば、ストリーミングメディア、ラジオ、広告、インターネット放送、ＣＤの歌、ＭＰ３ファイル、または何らかの他の種類のオーディオコンテンツ）のリアルタイム識別が含まれる。このように、本発明の実施の形態は、効率的、かつリアルタイムのメディアコンテンツの検査および他の報告を可能にする。

【００２２】

【００２３】

【００２４】

【００２５】

【発明を実施するための最良の形態】

【００２６】

概要

本発明の実施の形態は、オーディオサンプルからの特性情報（例えばオーディオ指紋）の抽出、および抽出した特性情報を用いたオーディオの一致または識別、を可能にする。図１に示すように、オーディオサンプル１００から採取したオーディオのフレーム１０５は、指紋抽出アルゴリズム１１０に入力される。オーディオサンプル１００は、任意の幅広いソースにより提供できる。オーディオフレーム１０５のシーケンスを用いて、指紋抽出アルゴリズム１１０は、シーケンスの特性となる一つ以上のオーディオ指紋１１５を生成する。区別するための識別子としての機能を果たすために、オーディオ指紋１１５は、オーディオサンプル１００のフレーム１０５のシーケンスの識別、または他の特性に関する情報を提供する。特に、オーディオサンプル１００に対する一つ以上の指紋１１５により、オーディオサンプル１００を一意に識別できる。指紋抽出アルゴリズム１１０の実施の形態を、詳細に以下説明する。

【００２７】

一旦生成すると、抽出された指紋１１５は、次に更なるプロセスで用いるか、または後で使用するためにメディア上へ格納できる。例えば、指紋１１５は、指紋１１５を指紋データベース１２５（例えば、既知のソースからのオーディオ指紋の収集）の登録と比較して、オーディオサンプル１００の識別を決定する指紋一致アルゴリズム１２０により用いることができる。指紋を用いるための各種の方法を以下に説明する。

【００２８】

オーディオサンプル１００は、指紋システムの用途に依存して、任意の幅広いソース起源であってよい。一実施の形態では、オーディオサンプル１００は、メディア放送から受信した放送からサンプリングされ、デジタル化される。代替として、デジタル化の必要を

予め避けるために、メディア放送はデジタル形式でオーディオを伝送してもよい。メディア放送の種類には、ラジオ放送局、衛星放送局、およびケーブルオペレータが含まれるが、これらに限定されない。このように、指紋システムを用いて、これらの放送を監査して、どのオーディオがどの時間に放送されるかを決定できる。これにより、放送制限、ライセンス契約等の準拠を確実にするための自動システムが可能になる。指紋抽出アルゴリズム１１０は、放送信号の正確な開始、および終了を知らなくても動作するので、メディア放送者の協力または知識なしに動作させて、独立した、かつ先入観のない結果を確保できる。

【００２９】

別の実施の形態では、メディアサーバが、メディアライブラリからオーディオファイルを検索し、指紋抽出アルゴリズム１１０が使用するためのデジタル放送をネットワーク（例えばインターネット）上に伝送する。ストリーミングインターネットラジオ放送は、この種のアーキテクチャの一例であり、メディア、広告、および他のコンテンツが個人またはユーザグループへ配送されている。かかる実施の形態では、指紋抽出アルゴリズム１１０および一致アルゴリズム１２０は、通常、オーディオサンプル１００のストリーミングコンテンツ内に含まれる個々のメディアアイテムの開始または終了時間に関する情報を何も持たない。しかしながら、これらのアルゴリズム１１０および１２０がストリーミングコンテンツを識別するのに、この情報は不要である。

【００３０】

別の実施の形態では、指紋抽出アルゴリズム１１０は、オーディオサンプル１００またはその一連のフレーム１０５を、オーディオファイルを収納する格納装置にアクセス権を持つクライアントコンピュータから受信する。クライアントコンピュータは、格納装置から個々のオーディオファイルを検索し、ファイルから一つ以上の指紋１１５を生成するために、ファイルを指紋抽出アルゴリズム１１０に送信する。代替として、クライアントコンピュータは、格納装置１４０からファイルのバッチを検索し、ファイル毎に一セットの指紋を生成するために、指紋抽出器１１０に順次送信する。（本明細書で用いる時、言うまでもなく「セット」は、単一アイテムを含むグループ内の任意数のアイテムを含む）。指紋抽出アルゴリズム１１０は、クライアントコンピュータまたはネットワーク上でクライアントコンピュータに接続されたリモートサーバにより実行できる。

【００３１】

アルゴリズム

図１に示す指紋抽出アルゴリズム１１０を実装する指紋抽出システム２００の一実施の形態を図２に示す。指紋抽出システム２００は、複数の処理チャンネル（それぞれは、本明細書で要素２１０および２１５としてラベル付けされる一つ以上の処理モジュールを含む）に接続され、順に、オーディオ指紋１１５を生成するための差分エンコーダ２２５に接続される解析フィルタバンク２０５を備える。指紋抽出システム２００は、オーディオ指紋を生成すべきオーディオフレーム１０５を受信するよう構成される。

【００３２】

以下に、より詳細に説明するが、入力されるオーディオフレーム１０５毎に、解析フィルタバンク２０５は、概して、周波数範囲全体に亘って受信信号に対するパワースペクトル情報を計算する。図示の実施の形態では、それぞれの処理チャンネルは、バンドがオーバーラップすることもある周波数範囲内の周波数バンドと対応する。従って、チャンネル毎に対応するバンドについて処理を実行できるように、指紋抽出システム２００が実行する処理をチャンネルが分割する。他の実施の形態では、複数のバンドに対する処理を、単一モジュールにより単一チャンネル内で実行でき、または用途に対して、およびシステムの技術的制約に対して、しかるべく任意の他の構成に処理を分割できる。

【００３３】

解析フィルタバンク２０５は、オーディオフレーム１０５（図１に示すオーディオサンプル１００からのフレーム１０５等）を受信する。解析フィルタバンク２０５は、オーディオフレーム１０５を時間領域から周波数領域に変換して、周波数範囲全体のフレーム１

０５に対するパワースペクトル情報を計算する。一実施の形態では、約２５０～２２５０Ｈｚの範囲の信号に対するパワースペクトルは、幾つかの周波数バンドに分割される（例えば、Ｍ＝１３のＭバンド）。バンドは、直線または対数の中間周波数分布（または、任意の他のスケール）を有し、オーバーラップしていてもよい。フィルタバンクの出力は、複数のバンドそれぞれについて信号エネルギーの大きさを含む。一実施の形態では、平均エネルギーの大きさは、バンド内の平均スペクトルエネルギーの立方根を用いて取得する。

【００３４】

解析フィルタバンク２０５の各種の実装が、ソフトウエア、およびハードウエア要件、およびシステムの制約に応じて可能である。一実施の形態では、解析フィルタバンク２０５は、オーディオフレーム１０５の信号を周波数バンドそれぞれに分離する幾つかのバンドパスフィルタを備え、次いで、エネルギー評価、およびダウンサンプリングが続く。別の実施の形態では、解析フィルタバンク２０５は、短時間高速フーリエ変換（ＦＦＴ）を用いて実装する。例えば、８ｋＨｚでサンプリングされたオーディオ１００は、６４ｍｓのフレーム１０５（すなわち５１２サンプル）にセグメント化する。次いで、２つのオーディオフレーム１０５（すなわち１０２４サンプル）から成るそれぞれ５０％オーバーラップしたセグメントのパワースペクトルを、ハン窓をかけ、ＦＦＴを実行して計算し、次いで、Ｍ個の等しく、または対数的に離間したオーバーラップ三角窓を用いてバンドフィルタ処理する。

【００３５】

一実施の形態では、パワースペクトルは周波数バンド内で平均化し、フレームシーケンスのエネルギー変化のみを、幾つかの実施の形態（以下に説明する）に対する特徴ベクトルの計算のために取得する。絶対振幅の代わりにエネルギー変化を使用し、かつフィルタバンク２０５のスペクトル特性の要件が低いために、各種の時間－周波数エネルギー変換を、上記ＦＦＴの代わりに用いることができる。例えば、修正離散コサイン変換（ＭＤＣＴ）を用いることができる。ＭＤＣＴの一つの利点は、ｎ／４ポイントＦＦＴを一つと、幾つかのサンプルの前回転および後回転だけを用いて計算できるので、複雑でないことである。従って、ＭＤＣＴを実装するフィルタバンク２０５は、ＦＦＴを実装するものより良好な動作をすると予想され。例えば、２倍の速度で変換計算ができる。

【００３６】

別の実施の形態では、解析フィルタバンク２０５は、カスケード接続の多相フィルタおよびエイリアスキャンセルが後に続くＭＤＣＴを含むＭＰ３ハイブリッドフィルタバンクを用いて実装される。ＭＰ３フィルタバンクは、５７６サンプルからなるオーディオのフレーム１０５毎に、５７６のスペクトル係数を生成する。８ｋＨｚでサンプリングされるオーディオでは、得られるフレームレートは、上記１０２４ポイントＦＦＴフィルタバンクの１５．６２６ｆｐｓと比較して１３．８ｆｐｓである。フレームレート差は、データが再サンプリングされた場合、以下に説明するように、時間－周波数解析中に補う。解析フィルタバンク２０５は、直交ミラーフィルタ（ＱＭＦ）を用いても実装できる。ＭＰ３ハイブリッドフィルタバンクの第１段は、３２等分幅バンドをもつＱＭＦを利用する。従って、１１，０２５Ｈｚのオーディオ信号の２５０～２２５０Ｈｚの周波数範囲は１３バンドに分割される。

【００３７】

ＭＰ３フィルタバンクの一利点は移植性にある。異なるＣＰＵのためのＭＰ３フィルタバンクの高度に最適化された実装が存在する。従って、指紋生成ルーチンは、ＭＰ３エンコーダと容易に一体化でき、追加処理せずにＭＰ３フィルタバンクからスペクトル係数を得ることができる。従って、指紋生成ルーチンは、ＭＰ３デコーダーと容易に一体化でき、完全なデコーディングをせずにＭＰ３ビットストリームから直接スペクトルデータを得ることができる。他のオーディオコードとの一体化も可能である。

【００３８】

一旦決定すると、サブバンドサンプルをバッファに入れ、一つ以上の非直線再サンプラ

２１０に提供する。一実施の形態では、非直線再サンプラ２１０はそれぞれ、Ｍ個の周波数バンドの内の一つと対応する。こうして、非直線再サンプラ２１０はそれぞれ、直線的な時間間隔をもつ特定周波数バンドに対するＳサンプルのシーケンスを受信する（例えば、フィルタバンクの実装に依存してＳを６４～８０に選択する）。一実施の形態では、各再サンプラ２１０は、サブバンドサンプルをＭ個のバンドの内の一つにマッピングし、対数的な時間間隔をもつ一連のＴサンプル（例えば、Ｔ＝６４）を生成する。これをＭバンドのそれぞれについて実行する場合、データを、対数スケールの時間軸（横軸）を有するサンプリングしたスペクトログラムと対応する［Ｍ×Ｔ］のマトリックスに格納できる。しかしながら、対数サンプリングは、一つの可能性に過ぎず、他の実施の形態では、指数的再サンプリング等の、他の種類の非直線サンプリングを実行できる。

【００３９】

次いで、バンド毎に非直線マッピングしたサンプルに変換を実行する一つ以上のＦＦＴモジュール２１５に、サブバンドサンプルを提供する。一実施の形態では、バンド毎に対数マッピングしたバンドサンプルに（例えば、［Ｍ×Ｔ］マトリックスの各行に）、ＴポイントＦＦＴを実行する。ＦＦＴから得られた一連の係数を特徴ベクトルと呼ぶ。一実施の形態では、各バンドに対する特徴ベクトルは、上行する周波数の順にそのバンドに対して計算したＦＦＴの他の係数全てを備える。従って、各特徴ベクトルは、Ｎ個の係数（例えば、Ｎ＝Ｔ／２＝３２）を含むことになる。

【００４０】

サブバンドサンプルにＦＦＴを実行するとしてＦＦＴモジュール２１５を説明するが、他の実施の形態では、ＦＦＴモジュール２１５は、他の時間－周波数変換を実行する処理モジュールにより置き換える。例えば、ＦＦＴの代わりに、離散コサイン変換（ＤＣＴ）または離散ハートレー変換（ＤＨＴ）を用いて、サブバンドサンプルを変換する。特に、ＤＨＴを用いると、偽の陽性率および非相関インデックス値が低くなる傾向があり、検索アルゴリズムを高速化するのに役立つ。別の実施の形態では、ＦＦＴモジュール２１５の代わりに線形予測コーディングを第２変換として用いる。

【００４１】

次いで、特徴ベクトルを、オーディオサンプルについて指紋１１５を生成する差分エンコーダ２２５に提供する。一実施の形態では、差分エンコーダ２２５は、各隣接バンド対と対応する特徴ベクトルを減算する。Ｍ個のバンドがある場合、Ｍ－１個の隣接バンド対がある。２つの特徴ベクトルを減算すると、Ｎ個の差分値のベクトルが得られる。これらの差分値それぞれに対して、差分が０以上の場合、差分エンコーダ２２５は１を選択し、差分が０未満の場合、差分エンコーダ２２５は０を選択する。シーケンスの４ビットの各グループに対して、エンコーダは、コードブック表に従ってビット値を割り当てる。最良のコードブック値は、指紋アルゴリズムのチューニングおよびトレーニング中に計算する。連続するバンド対それぞれの特徴ベクトルに対してこの処理を繰り返すと、［（Ｍ－１）×Ｎ／４］マトリックスのビットを生じる。直線ビットシーケンスとして表すことができるこのマトリックスを、オーディオ指紋１１５として用いる。Ｍ＝１３およびＮ＝８の例では、指紋１１５は１２バイトの情報を有する。

【００４２】

一実施の形態では、主成分分析（ＰＣＡ）を用いて、得られた特徴ベクトルを無相関化し、サイズを小さくしてから量子化する。デジタルコサイン変換等の、他の無相関化技法を追加して、または代替として用いて、冗長性を除去し、特徴ベクトルを小型化する。

【００４３】

一実施の形態では、指紋抽出システム２００は、特定オーディオ信号の大きくオーバーラップした一連のオーディオフレームに対して、複数の指紋を生成する。一実施例では、システム２００により処理される一連のフレーム１０５それぞれは、３秒間のオーディオ信号を含み、前のシリーズが開始した後６４ｍｓに開始する。このようにして、６４ｍｓ毎に開始するオーディオ信号の３秒間の幾つかの部分に指紋を生成する。かかる仕組みを実装するために、指紋抽出システム２００は、解析フィルタバンク２０５の前後にメモリ

ーバッファを含み、バッファは、次のオーディオフレーム１０５を受信すると、次の６４ｍｓのオーディオ信号により更新される。

【００４４】

音響モデル

指紋システムの各種の用途では、特定周波数バンドは、知覚できない、オーディオサンプルに対するエンコード処理がバンドを除去する、または何らかの他の理由により、重要でないことがある。従って、一実施の形態では、音響モデル２３５を用いて、特定指紋に対する重要でない周波数バンドを識別し、マークする。心理音響モデル等の、音響モデルは、各種のオーディオ処理分野で周知である。音響モデル２３５に対する一セットのモデルパラメータは、指紋１１５の作成中に高品位の基準サンプルに対して計算でき、データベース１２５に格納できる。指紋１１５の重要でないバンドは、その対応値（すなわち、ビット）をゼロに消去することによりマークできる。指紋をデータベース記録と一致させる処理では、非ゼロの値をもつ対応するバンド対のみを用いて指紋１１５を区別するので、これにより、後続のどの一致処理でもバンドを効率的に無視できる。マスクされたバンド（すなわち、ゼロ値を有するバンド）を相互比較から除外することもできる。

【００４５】

一実施の形態では、音響モデルは、人の聴覚系のための心理音響モデルである。これは、指紋システムの目的を人の聴覚系を目標としたオーディオ識別とすると有用なことがある。かかるオーディオは、関連がないオーディオ情報を除去する一つ以上の感覚的エンコーダにより圧縮することができる。人の心理音響モデルを用いることにより、かかる関連がないバンドを指紋から識別し、除外できる。

【００４６】

しかし、心理音響モデルは、人の感覚に従ってエンコードしたオーディオに適する一種の音響モデルにすぎない。別の音響モデルは、特定の録音装置の特性を模したモデルである。かかる録音装置の音響モデルに対する各バンドは、その重要性に応じて割り当てられる重み付け因子を有する。更に別の音響モデルは、車両内または室内で見られる背景ノイズ等の、特定環境の特性を模する。かかる実施の形態では、音響モデルに対する各バンドは、システムを設計した環境での重要性に応じて割り当てられた重み付け因子を有する。

【００４７】

一実施の形態では、音響モデル２３５およびフィルタバンク２０５のパラメータは、解析するオーディオ信号１００の種類および特性に依存する。一セットのサブバンド重み付け因子、ならびに幾つかのフィルタバンクバンドおよびその周波数分布を備える異なったプロファイルを用いて、目標とするオーディオ信号の特性の良好な一致を得る。例えば、会話様オーディオでは、信号のパワーは主として低い周波数バンドに集中する一方、音楽はジャンルに応じてより高い周波数に関連する成分を含む。一実施の形態では、音響モデルのパラメータは、基準オーディオ信号により計算し、生成した指紋とともに、コンテンツデータベースに格納する。別の実施の形態では、音響モデルのパラメータは、一致処理中に、解析するオーディオ信号の特性に基づいて動的に計算する。

【００４８】

従って、音響モデル２３５の可能性がある用途には、特定環境、および／または録音装置、およびエンコードアルゴリズム特性に対するオーディオ認識パラメータのチューニングが含まれる。例えば、携帯電話オーディオパスの既知の音響特性（マイクロホン特性、オーディオ処理および圧縮アルゴリズム等）により、これらの特性を模擬した音響モデルの開発が可能になる。指紋比較中にこのモデルを用いることにより、生成した指紋の一致処理のロバスト性を著しく増大させることができる。

【００４９】

指紋のインデックス化およびマッチング

一実施の形態では、指紋インデックス器２３０が、各指紋１１５に対するインデックスを生成する。次いで、指紋１１５を指紋データベース１２５に格納するので、指紋データベース１２５のコンテンツの効率的な検索、およびマッチングが可能になる。ある実施の

形態では、指紋１１５に対するインデックスが、指紋１１５の一部を含む。従って、指紋データベース１２５内の指紋１１５は、それらについての有用な識別情報によりインデックスされる。

【００５０】

指紋１１５それぞれが［（Ｍ－１）×Ｎ／４］マトリックスのビットを含む上記実施の形態では、インデックス器２３０は、インデックスとして最も左の列からのビットを用いる。指紋１１５それぞれが１２×８マトリックスのビットである実施例では、指紋１１５に対するインデックスは、最も左の２列のビット（合計２４ビット）とすることができる。このように、指紋１１５それぞれに対するインデックスとして用いられるビットは、指紋１１５を計算するために用いる特徴ベクトルの低周波数スペクトル係数に基づく指紋１１５のサブセットである。従って、これらのビットは、対数マッピングしたスペクトログラムバンドのスペクトルの低周波数成分と対応し、ノイズおよび歪みを抑えるよう安定かつ低感度になる。従って、高いレベルの確率で、類似指紋は、同一のインデックス数値を有することになる。こうして、インデックスを用いてデータベース中の類似したかつそれらしく一致する指紋にラベル付けし、グループ化できる。

【００５１】

図３は、本発明の一実施の形態による、上記説明のインデックスを用いて、試験指紋を指紋データベース１２５とマッチングする方法を示す。試験指紋について指紋データベース１２５内に一致を見出すために、マッチングアルゴリズムは、上記のように試験指紋に対するインデックス値を計算することにより開始される３１０。このインデックス値を用いて、例えば、データベース１２５内の同一インデックス値を有する指紋全てが含まれる候補の指紋グループが得られる３２０。上記説明のように、インデックス値を計算する方法により、データベース１２５内のどの一致も、候補指紋のこのグループ内にある可能性が非常に高い。

【００５２】

候補の指紋のグループ内のどれかの一致を試験するために、試験指紋と、それぞれの候補指紋との間のビットエラー率（ＢＥＲ）を計算する３３０。２つの指紋の間のＢＥＲは、一致しない対応ビットの百分率である。関連のない完全にランダムな指紋では、ＢＥＲは５０％となると予測される。一実施の形態では、ＢＥＲが約３５％未満の場合、２つの指紋は一致している；しかしながら、偽の陽性、および／または偽の陰性に対する所望の許容値に応じて他の限界値を用いてもよい。更に、ＢＥＲ以外の計算または基準を用いて２つの指紋を比較することができる。例えば、ＢＥＲの逆数、一致率を用いることもできる。更に、２つの指紋の比較に際して、特定ビットに他より高い重み付けをしてもよい。

【００５３】

所定の一致基準内には一致がない場合（３４０）、または修正するインデックスがそれ以上ない場合は（３５０）、一致アルゴリズムがデータベース１２５内に試験指紋の一致を何も発見できなかった場合である。次いで、システムは検索を継続するか（例えば、候補指紋を得るために制限が緩い基準を用いる）、または停止してもよい。一つ以上の一致する指紋がある場合（３４０）、一致指紋のリストが返される３６０。

【００５４】

一実施の形態では、システムは、一致を検索するための別の候補指紋のセットを得るために、計算した指紋インデックスを修正した後（３７０）、上記の検索を繰り返す。計算した指紋インデックスを修正するために（３７０）、計算した指紋インデックスの一つ、または多数を入れ替えることができる。指紋インデックスが２４ビットを有する一実施例では、元の指紋インデックスを用いて一致を発見できなかった後で、２４ビットの内の異なる単一ビットをその都度入れ替えて、検索ステップを２４回繰り返す。様々な他の技法を用いて検索空間を拡大できる。

【００５５】

一実施の形態では、指紋インデックス器２３０は、音響モデル２３５により計算し、前にデータベース１２５に格納しておいた一セットの周波数バンド重み付け因子に基づいて

、インデックスビットを一つ以上の指紋から選択することにより、一つ以上のインデックスを生成する。ビット入れ替えにより得られたインデックスを含む多数のインデックスを用いる場合、候補指紋のグループには、計算したインデックス毎について得られる全ての候補が含まれる。

【００５６】

別の実施の形態では、計算したインデックス毎に得られる大部分の、または全ての候補グループ内で発見できる指紋候補だけを、事前スクリーニングし、選択することにより、検索領域を狭くすることができる。ビット入れ替えにより得られるインデックスを含む多数のインデックスを用いることによる多数の指紋候補グループの事前スクリーニングは、データベース検索の性能を著しく改良できる。一実施の形態では、可能性のある指紋候補へのインデックスおよび参照をコンピューターメモリに格納することにより、指紋候補の高速選択、および事前スクリーニングが可能になる。第２ステップでは（ステップ３２０）、所与の指紋と一致する確率が最も高い指紋候補だけが、コンピューターメモリにロードされ、比較される。この手法により、小さなインデックスだけをコンピューターメモリに保持し、一方、低速の装置（例えば、ハードディスクドライブ、またはネットワーク上）には大きな指紋を格納することにより、高速検索が可能になる。

【００５７】

オーディオフレームのエッジ検出

用途によっては、一致オーディオ断片のエッジ検出が所望されることがある。エッジ検出により、システムは、特定のオーディオ断片が一致する時間を正確に知ることができる。解析するオーディオの品位に応じて、エッジ検出アルゴリズムの実施の形態は、約０．１～０．５秒の正確さで、一致するオーディオ断片のエッジを検出することができる。

【００５８】

上記のように、指紋技法の実施の形態は、サブバンド処理バッファにオーディオサンプルを蓄積する。このバッファ化により、指紋アルゴリズムの出力は遅延され、オーディオ断片のエッジに塗りつけられる。この効果を図４に示す。図は、オーディオ断片に対する参照指紋と、入力されてくるサンプルオーディオストリームに対して生成される時間を通じての一連の指紋との間の、時間を通じてのビットエラー率（ＢＥＲ）のグラフである。図示の実施の形態では、サブバンドバッファは、３秒間のオーディオを保持し、２つの指紋が３５％以下のビットエラー率（ＢＥＲ）を有する場合、一致が宣言される。

【００５９】

最初に、時間Ｔ０では、サブバンド処理バッファは空であり、次いで、生成した指紋は、元のオーディオとのゼロ一致を生成する（すなわち、ＢＥＲがほぼ５０％に等しいと予測される）。オーディオサンプルがサブバンドバッファに加えられた時にＢＥＲが低下すると、より良好な一致を示す。時間が十分経過した後、ＢＥＲが時間Ｔ１でスレッショールド３５％以下に低下して、一致を示す。最終的に、時間Ｔ２で、バッファがサンプルで満たされるので、ＢＥＲは平坦域に入る。指紋アルゴリズムが対応するオーディオ断片の最後を通過する時間Ｔ３で、一致が少なくなる指紋を生成し始め、従って、ＢＥＲが増加し、時間Ｔ４で認識スレッショールド３５％に達する。得られた一致曲線の期間（Ｔ１～Ｔ４）、および平坦域の期間（Ｔ２～Ｔ３）は、一致したオーディオ断片の期間（Ｔ０～Ｔ３）より、それぞれ短い。

【００６０】

一実施の形態では、エッジ検出アルゴリズムを用いて、一致するオーディオフレーム、または断片の正確なエッジを決定する。図４に示すようなＢＥＲ曲線が得られる。ＢＥＲ曲線は、低下するＢＥＲ（例えば、Ｔ１～Ｔ２）を伴う一致の開始、ほぼ一定のＢＥＲ（例えば、Ｔ２～Ｔ３）を伴う平坦域、および増加するＢＥＲ（例えば、Ｔ３～Ｔ４）を伴う一致の終了、と対応する領域にセグメント化される。実際のＢＥＲ曲線は一般にノイズが多いので、回帰分析等の、しかるべき技法を用いてセグメント化する。一実施の形態では、３５％以上のＢＥＲを生成する全てのサンプルは、信頼できないので無視する。次いで、一致するオーディオ断片の開始を（すなわち、時間Ｔ１）、低下するＢＥＲ領域（例

えば、Ｔ１～Ｔ２）を最良の方法で適合させるラインと、５０％ＢＥＲと対応する水平ラインとの交点として、線形回帰を用いて計算する。類似の手法は、時間Ｔ５を推定するのにも適用でき、増加するＢＥＲ領域（例えば、Ｔ３～Ｔ４）を最良の方法で適合させるラインと、５０％ＢＥＲと対応する水平ラインとの交点を取る。しかしながら、この場合、時間Ｔ５は、サブバンドバッファの期間Ｂだけ遅延した断片の終了と対応し、一致するオーディオ断片の実際の終了とは対応しない。断片の終了（例えば、時間Ｔ３）の場所は、得られた推定時間Ｔ５から、サブバンドバッファ期間Ｂを減算することにより計算できる。

【００６１】

別の実施の形態では、一致するオーディオ断片の終了は領域Ｔ２～Ｔ３の終了として推定され、オーディオ断片の開始は、領域Ｔ２～Ｔ３の開始と対応する時間Ｔ２から、サブバンドバッファＢの期間を減算することにより計算する。

【００６２】

概略

ベクトルおよびマトリックスの点から説明したが、どの指紋、またはサブ指紋についても計算した情報を、ベクトルまたはマトリックスの値としてだけではなく任意の形式で格納し、所有できる。従って、用語ベクトルおよびマトリックスは、オーディオサンプルから抽出したデータを表す簡便なメカニズムとしてのみ用いられ、何らかの他の方法を制限することを意味しない。更に、パワースペクトルは、スペクトログラムの点から説明しているが、言うまでもなく、オーディオ信号のパワースペクトル、またはスペクトル解析を表すデータは、スペクトログラムだけではなく、他の任意の適切な形式で表され、用いることができる。

【００６３】

一実施の形態では、本明細書で説明した任意の、または全てのステップ、操作、または処理を実行するためにコンピュータープロセッサにより実行されるコンピュータープログラムコードを収納するコンピュータ可読媒体を含むコンピュータープログラム製品により、ソフトウエアモジュールが実装される。従って、本明細書で説明した任意のステップ、操作、または処理は、単独で、または他の装置と組合せて、一つ以上のソフトウエアモジュール、またはハードウエアモジュールにより実行、または実装できる。更に、ハードウエア要素の点から説明したシステムの任意の部分は、ソフトウエアで実装してもよく、ソフトウエア要素の点から説明したシステムの任意の部分は、専用回路内のハードコードとして、ハードウエアで実装してもよい。例えば、説明した方法を実行するためのコードを、ハードウエア装置、例えば、ＡＳＩＣ、または他のカスタム回路、に組み込むことができる。これにより、本発明の恩恵を多くの異なる装置の能力と組合せることができる。

【００６４】

別の実施の形態では、指紋アルゴリズムを、携帯電話、携帯情報端末（ＰＤＡ）、ＭＰ３プレーヤおよび／またはレコーダ、セットトップボックス、またはオーディオコンテンツを格納、または演奏する任意の他の装置等の、任意の様々なオーディオ装置に組み込み、実行させる。かかる装置に指紋アルゴリズムを組み込むのには、幾つかの利点がある。例えば、携帯電話でオーディオ指紋を直接生成するのは、その電話から携帯電話ネットワーク上の指紋サーバに、圧縮したオーディオを送信するのと比較してより良好な結果をもたらすであろう。携帯電話上でアルゴリズムを実行するのは、会話を圧縮するために設計され、音楽には貧弱な性能しか発揮できないＧＳＭ圧縮により起きる歪みを除去する。従って、本手法は、携帯電話により録音されたオーディオの認識を著しく改良できる。また、ネットワークトラフィックはもとより、サーバに加わる負荷も軽減する。

【００６５】

かかる組み込み手法の別の恩恵は、プライバシー、およびユーザの権利を侵害せずに、聴取している経験を監視できることである。例えば、録音装置は、オーディオを録音し、指紋を生成し、次いで、指紋だけをサーバに解析のために送信する。録音したオーディオは装置から出ていくことがない。次いで、サーバは、指紋から元のオーディオを復元でき

なくても、送信した指紋を用いて目標とする音楽、または広告を識別する。

【００６６】

本発明の実施の形態のこれまでの説明は、説明を目的として提示したものであり、開示した形だけに本発明を徹底させたり、制限する意図はない。当該技術に習熟する者には言うまでもなく、多くの修正および改変が上記教示の観点から可能である。従って、意図していることは、本発明の範囲が、この詳細な説明によってではなく、本明細書に付帯のクレームによって制限される、ということである。

【図面の簡単な説明】

【００６７】

【図１】本発明の実施の形態によるオーディオサンプルから指紋を抽出し、使用するためのプロセスの略図である。

【図２】本発明の実施の形態による指紋抽出システムの略図である。

【図３】本発明の実施の形態による一致アルゴリズムのフロー図である。

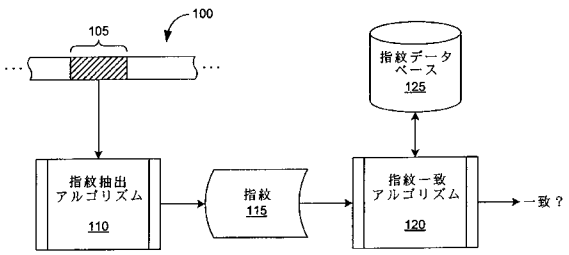
【図４】本発明の実施の形態によるエッジ検出アルゴリズムを示す。

【符号の説明】

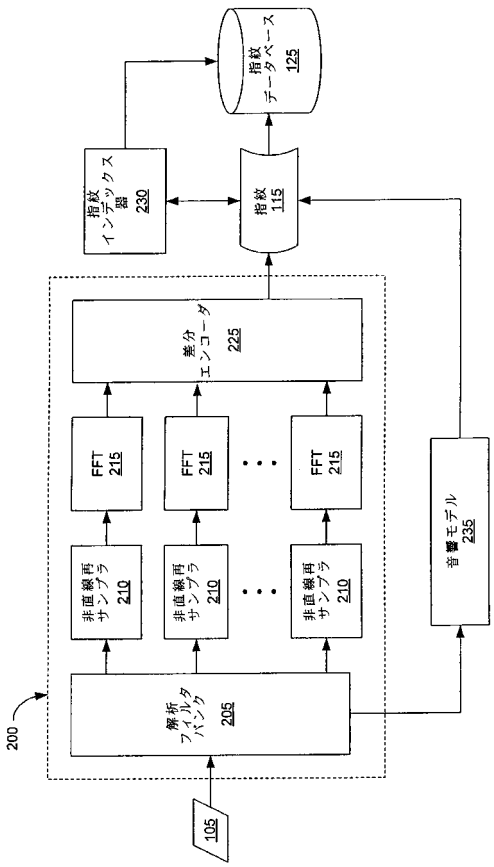
【００６８】

１００　オーディオサンプル
１１０　指紋抽出アルゴリズム
１１５　指紋
１２５　指紋データベース
２００　指紋抽出システム

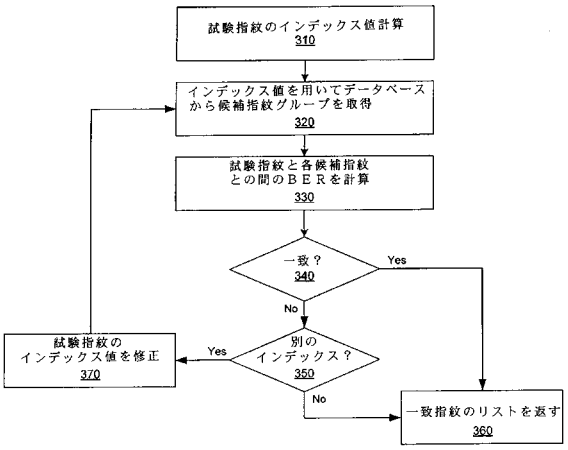
【図１】



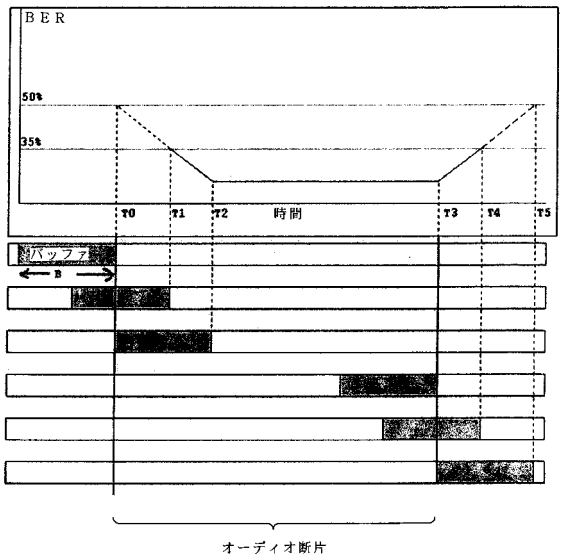
【図２】



【図３】


試験指紋のインデックス値計算
310
インデックス値を用いてデータベースから候補指紋グループを取得
320
試験指紋と各候補指紋との間のＢＥＲを計算
330
一致？
340
Yes
No
別のインデックス？
350
Yes
No
試験指紋のインデックス値を修正
370
一致指紋のリストを返す
360

【図４】


ＢＥＲ
50%
35%
T0
T1
T2
時間
T3
T4
T5
B
オーディオ断片

フロントページの続き

(74)代理人 100120525
弁理士 近藤 直樹
(72)発明者 サージー ビロブロフ
カナダ国 ブリティシュコロンビア Ｖ３Ｅ２Ｒ９, コクィトラム, マウンテンビュー コート １４１０

審査官 田部井 和彦

(56)参考文献 特開２０００－２８５１０４（ＪＰ, Ａ）
国際公開第２００４／０４４８２０（WO, Ａ１）
国際公開第０２／０１１１２３（WO, Ａ２）
特開平０５－０７３０９３（ＪＰ, Ａ）
特開２００５－２６６０１３（ＪＰ, Ａ）
特表２００６－５０５８２１（ＪＰ, Ａ）
S.Sukittanon, L.E.Atlas, J.W.Pitton, Modulation-scale analysis for content identification, Signal Processing, IEEE Transactions on, IEEE, ２００４年１０月, Volume 52, Issue 10 , p.3023 - 3035
P.Cano, E.Batle, T.Kalker, J.Haitsma, A review of algorithms for audio fingerprinting , Multimedia Signal Processing, 2002 IEEE Workshop on, IEEE, ２００２年 ９月, p.169 - 173

(58)調査した分野(Int.Cl., ＤＢ名)
Ｇ１０Ｌ １３／００－２５／９３
ＩＥＥＥ Ｘｐｌｏｒｅ