

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第4870921号
(P4870921)

(45) 発行日 平成24年2月8日 (2012. 2. 8)

(24) 登録日 平成23年11月25日 (2011. 11. 25)

(51) Int. Cl.	F I
G 1 0 L 19/00 (2006. 01)	G 1 0 L 19/00 3 3 0 Z
G 1 0 L 11/00 (2006. 01)	G 1 0 L 11/00 4 0 2 Z
G 0 6 F 12/00 (2006. 01)	G 0 6 F 12/00 5 0 1 B
G 1 1 B 27/00 (2006. 01)	G 1 1 B 27/00 D

請求項の数 10 (全 23 頁)

(21) 出願番号	特願2004-296842 (P2004-296842)	(73) 特許権者	500046438
(22) 出願日	平成16年10月8日 (2004. 10. 8)		マイクロソフト コーポレーション
(65) 公開番号	特開2005-202357 (P2005-202357A)		アメリカ合衆国 ワシントン州 9805
(43) 公開日	平成17年7月28日 (2005. 7. 28)		2-6399 レッドモンド ワン マイ
審査請求日	平成19年10月4日 (2007. 10. 4)		クロソフト ウェイ
(31) 優先権主張番号	60/514, 490	(74) 代理人	100077481
(32) 優先日	平成15年10月24日 (2003. 10. 24)		弁理士 谷 義一
(33) 優先権主張国	米国 (US)	(74) 代理人	100088915
(31) 優先権主張番号	10/785, 561		弁理士 阿部 和夫
(32) 優先日	平成16年2月24日 (2004. 2. 24)	(72) 発明者	クリストファー ジェイ. シー. バーゲス
(33) 優先権主張国	米国 (US)		アメリカ合衆国 98052 ワシントン
			州 レッドモンド ワン マイクロソフト
			ウェイ マイクロソフト コーポレーシ
			ョン内
			最終頁に続く

(54) 【発明の名称】 オーディオ重複検出器

(57) 【特許請求の範囲】

【請求項 1】

オーディオ情報を管理するためのシステムであって、
複数のオーディオファイルの一部を対応するフィンガプリントにマッピングするフィン
ガプリンティングコンポーネントと、
前記フィンガプリント間の距離に部分的に基づいて、データ記憶デバイスから削除され
る可能性のあるものとして前記オーディオファイルの1つまたは複数にタグ付けする検出
器と、
重複したまたは欠陥のあるオーディオファイルのリストをユーザインターフェースに出
力するのに利用されるデータベースと、
雑音のあるオーディオファイルを識別するのに利用される拒否権フィンガプリントを記
憶するためのデータベースと
を備え、
前記検出器は、前記データベースに基づいて、前記オーディオファイル进行处理する間に
エラー状況を記録し、前記エラー状況に関連するファイルのリストを前記ユーザインター
フェースに出力することを特徴とするシステム。

【請求項 2】

前記検出器は、フィンガプリント間の前記距離が所定のしきい値未満であることに基
づいて前記オーディオファイルにタグ付けすることを特徴とする請求項 1 に記載のシステム
。

【請求項 3】

前記フィンガプリンティングコンポーネントはさらに、ファイルについて複数のフィンガプリントを生成し、前記複数のフィンガプリントは前記ファイル中のオーディオの時間ウィンドウに対応し、前記検出器は、各ファイルごとに前記複数のフィンガプリントと1つまたは複数の記憶済みフィンガプリントとの間の最低距離に基づいて、前記オーディオファイルにタグ付けすることを特徴とする請求項 1 に記載のシステム。

【請求項 4】

前記フィンガプリンティングコンポーネントは、前記オーディオファイル開始からの時間オフセットと、前記ファイル中の時間ウィンドウの継続時間とを受け入れるようになされたことを特徴とする請求項 3 に記載のシステム。

10

【請求項 5】

前記フィンガプリンティングコンポーネントは、1秒よりも長いオーディオから生成され約 64 ビットの浮動小数点数からなるフィンガプリントを計算することを特徴とする請求項 1 に記載のシステム。

【請求項 6】

前記検出器は、DB1 および DB2 と呼ばれる少なくとも 2 つの内部データベースを利用し、DB1 中では、レコードはフィンガプリントと、正規化係数を含めた関連する数値量とを含み、DB2 中では、レコードはファイル名と、ID インデックスと呼ばれる関連するインデックスと、「オフセット」パラメータと、「距離」パラメータとの、少なくとも 4 つのオブジェクトを含むことを特徴とする請求項 1 に記載のシステム。

20

【請求項 7】

前記検出器は、すでに処理されたファイル中で最もよくマッチする位置を見つけるために、時間ウィンドウ中のすべてのフィンガプリントを計算して比較することを特徴とする請求項 4 に記載のシステム。

【請求項 8】

前記検出器は、オーディオファイルの識別を決定するのに利用されることを特徴とする請求項 7 に記載のシステム。

【請求項 9】

前記オーディオファイルの前記識別はオーディオファイルに関連するメタデータにより実施されることを特徴とする請求項 8 に記載のシステム。

30

【請求項 10】

複数のオーディオファイルの一部を対応するフィンガプリントにマッピングするフィンガプリンティングコンポーネントと、

前記フィンガプリント間の距離に部分的に基づいて、データ記憶デバイスから削除される可能性のあるものとして前記オーディオファイルの1つまたは複数にタグ付けする検出コンポーネントと、

重複したまたは欠陥のあるオーディオファイルのリストをユーザインターフェースに出力するのに利用されるデータベースコンポーネントと、

雑音のあるオーディオファイルを識別するのに利用される拒否権フィンガプリントを記憶するためのデータベースコンポーネントと

40

を、コンピュータが実行されたときに、コンピュータに実現させるためのコンピュータ実施可能命令が記憶され、

前記検出コンポーネントは、前記データベースコンポーネントに基づいて、前記オーディオファイルを処理する間にエラー状況を記録し、前記エラー状況に関連するファイルのリストを前記ユーザインターフェースに出力することを特徴とするコンピュータ可読記憶媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は一般にコンピュータシステムに関し、より詳細には、オーディオフィンガプリ

50

ントを利用して、余分なまたは破損したオーディオファイルを自動的に管理するシステムおよび方法に関する。

【背景技術】

【0002】

近年の多くのソフトウェアシステムによってサポートされている現行機能の1つは、オーディオファイルを記憶および再生する能力である。これらのシステムの多くでは、ユーザがオーディオファイルの様々な集まりを記憶および管理することができる。しかし時間が経つにつれて、多くのユーザは、より多くの記憶空間を塞ぐことが避けられない多数のファイルがあることに不満を感じるようになる。また、集まりが大きくなるにつれて、希望のオーディオ情報を取り出して再生することは、より困難で時間のかかることになる。多くのシステムは、増加し続けるこれらの多量のオーディオ情報をユーザが管理するのを補助するためのソフトウェアを提供している。例えばこれらのシステムには、MP3、Ogg Vorbis (OGG)、Windows (登録商標) Media Audio (WMA)、MPCおよびMP+ファイルなどを含めた一般的なオーディオファイルフォーマットをサポートするオーディオマネージャを含めることができる。これによりユーザは、ユーザのオーディオファイルの集まり全体をカタログ化して、お気に入りの曲を即座に検索したり、アルバムカバーをサムネイルとして使用してアルバムを閲覧したり、報告を作成したり、その他の有用な機能を実施したりすることができる。

10

【0003】

オーディオファイルを構成することに加えて、これらのシステムは、例えばタグ編集、名前変更、歌詞編集、CD作成、アーティスト情報検索によってファイルを管理するための、ツールのセットも提供する。ユーザは、ハードディスク、CD-ROM、ネットワークドライブ、ZIPドライブ、またはその他の任意の取外し可能な媒体に記憶されたオーディオファイルを扱うことができる。ツールセットには、ユーザが複数の再生リストを再生することを可能にするツール、および各タイトルに関連する画像を表示することを可能にするツールが含まれる。他の機能には、自動生成されるデータベース統計、個人評価、ジャンル別やムード別や年別のソート、カスタムデータベース照会が含まれる。

20

【0004】

最近、ストリーム中またはファイル中のオーディオを識別するための強力な方法として、オーディオフィンガプリンティング (AFP) が出現してきた。現在、複数の会社がオーディオフィンガプリンティングに基づく音楽サービスを提供している。これらのサービスでは、識別すべきオーディオから1つまたは複数のフィンガプリントを抽出し、これらのフィンガプリントを、あらかじめ計算されたフィンガプリントの大規模なデータベースに対してチェックする必要がある。

30

【発明の開示】

【発明が解決しようとする課題】

【0005】

しかし、オーディオファイルを素早く解析するのは(サムネイルを使用できる画像とは異なり)現在是不可能なので、オーディオの大きな集まりを管理するのは困難である。ユーザは一般にラベリングに依拠しなければならないが、これさえも限定的な助けでしかない。すなわち、ラベリングはしばしば不正確だが、正確なラベリングを使用しても、通常のPCの集まりの中で重複オーディオファイルを識別するのは非常に時間がかかる。したがって、重複したオーディオファイル、および/または雑音を含むかその他の問題を有するオーディオファイルを、ユーザのPCまたはその他の媒体デバイス上で自動的に識別するシステムおよび方法であって、ユーザが自分のオーディオファイルをより効率的に管理できるようにするシステムおよび方法が必要とされている。

40

【課題を解決するための手段】

【0006】

以下に、本発明のいくつかの態様に関する基本的な理解を提供するために、本発明の簡単な要約を提示する。この要約は本発明の広範な概観ではない。これは本発明の鍵となる

50

要素／クリティカルな要素を識別するものではなく、本発明の範囲を線引きするものでもない。この唯一の目的は、後で提示するより詳細な説明への前置きとして、本発明のいくつかの概念を簡略化された形で提示することである。

【 0 0 0 7 】

本発明は、重複または破損したオーディオファイルを検出して、このようなファイルの管理および削除を容易にするためのシステムおよび方法に関する。大きなオーディオファイルを素早く解析することは、例えば画像およびテキストに比べると問題を含むので、オーディオの大きな集まりを管理することは困難である。過去、ユーザはラベリングに依拠してきたが、ラベリングは不正確な場合がある。本発明は、ユーザのオーディオファイル検索を補助すること、相互に重複しているかもしれないファイルを識別すること、破損したファイル、雑音のあるファイル、またはジャンクファイルを識別すること、このようなファイルをユーザのデータベースから削除するのを容易にするためのツールを提供することによって、従来のシステムの欠点および短所の多くを解決する。一態様では、ユーザはシステムに2つのパラメータ（フィンガプリントを抽出するためのオーディオ開始からの秒数（ t ）と、スロップウィンドウ（ $slop\ window$ ）のサイズ（ s ））を供給する。次いで本発明は、ユーザのオーディオファイルを突き止め、（ t ）および（ s ）に部分的に基づいてフィンガプリントを計算する。自動的に決定された重複または破損ファイルをユーザが削除できるようにすることと共に、これらおよび他のパラメータを構成することのために、ユーザインターフェースが提供される。

【 0 0 0 8 】

本発明の一態様では、重複検出器によって処理されるオーディオファイルのためのデータベースを備えた、オーディオ重複検出器システムが提供される。重複検出器は、例えば重複オーディオファイルが存在するか否か、ファイルが破損しているかどうかなどの特徴を判定するためにオーディオファイル进行处理する分析器からの中間結果を保持するための、1つまたは複数の内部データベースを備える。フィンガプリントコンポーネントおよび正規化コンポーネントが、分析器内で動作して、オーディオファイルの一部が他のファイルと重複しているかどうかを判定し、それによりマッチ検出器が、このような部分を重複（または破損ファイルやジャンクファイルなど、他の何らかの指示）としてタグ付けする。ユーザインターフェースが、特徴の中でもとりわけ、ユーザがシステムパラメータ（例えば検出分析パラメータ）を入力できるようにするための入力、ならびに、削除するファイルを選択できるようにするための入力を備える。インターフェースからの出力には、ユーザのデータベースから削除されるオーディオファイルのリスト、およびシステムに対する構成オプションのリストを含めることができる。

【 0 0 0 9 】

一般に、ユーザは、データベースを検索する際の起点となる最上レベルディレクトリ（「ルート」と呼ばれる）を選択し、次いでユーザインターフェースを介して検出器を始動させる。通常、ユーザは2つ（またはそれ以上）のパラメータをシステムに供給する。すなわち、オーディオファイル開始から何秒でそれぞれのフィンガプリントを抽出するか（ T ）、およびどのサイズの「スロップウィンドウ」を使用するか（ S ）であり、 S は秒でも測定される。次いで検出器は、ルートディレクトリから下にサブディレクトリを検索してオーディオファイルを取り出すが、検出器はこれらをその拡張子で識別する。検出器は、オーディオファイルを見つけると、このファイルをロードし、ファイル開始後 $T - S$ 秒からファイル開始後 $T + S$ 秒までの一連のフィンガプリントを計算する。これらのどのフィンガプリントもデータベース中の既存のフィンガプリントとマッチしない場合は、ファイル開始後 T 秒で計算されたフィンガプリントがデータベースに追加される。このようにして、オーディオファイルは、単一のフィンガプリントがデータベースに追加されるようにするか（マッチが見つからない場合）、あるいはフィンガプリントがデータベースに追加されないようにする（マッチが見つかった場合）ことになる。一例では、6秒のオーディオから生成されるフィンガプリントであって、64個の浮動小数点数および1個の正規化係数（これもやはり浮動小数点数である）からなるフィンガプリントが計算される。理

解できるように、その他の時間値、およびその他の数の浮動小数変数を利用することもできる。

【 0 0 1 0 】

前述の目的および関係する目的を達成するために、本発明のいくつかの例示的な態様を、以下の記述および添付の図面に関連して本明細書に述べる。これらの態様は、本発明を実施することのできる様々な方法を示すものであり、これらはすべて本発明によってカバーされるものとする。本発明のその他の利点および新規な特徴は、以下の本発明の詳細な説明を図面と共に考えれば明らかになるであろう。

【 発明を実施するための最良の形態 】

【 0 0 1 1 】

本発明は、データベース中の重複および/または破損したオーディオファイルを自動的に検出するのを容易にするためのシステムおよび方法に関する。ファイルが検出されると、望むなら、後でデータベースから削除されるようにファイルにタグ付けすることができる。一態様では、オーディオ情報を管理するためのシステムが、オーディオファイルの一部を識別するためのフィンガプリンティングコンポーネントを備える。検出器が、決定されたオーディオファイル間のコストに基づいて、データ記憶域から削除される可能性のあるものとして1つまたは複数のオーディオファイルにタグ付けする。検出器は、コストが所定のしきい値未満であることに基づいて、または最低コスト分析に基づいて、オーディオファイルにタグ付けすることができる。別の態様では、オーディオファイルを管理するためのユーザインターフェースが提供される。インターフェースは、データベースから削除される可能性のあるオーディオファイルに関する1つまたは複数のオプションを提供する表示コンポーネントと、これらのオプションを選択して、削除される可能性のあるオーディオファイルを決定する自動オーディオ削減コンポーネントを構成するための入力コンポーネントとを備える。

【 0 0 1 2 】

本明細書で使用される用語「コンポーネント」「オブジェクト」「検出器」「システム」などは、ハードウェア、ハードウェアとソフトウェアの組合せ、ソフトウェア、実行中のソフトウェアのいずれかの、コンピュータ関連エンティティを指すものとする。例えばコンポーネントは、限定しないが、プロセッサ上で実行されるプロセス、プロセッサ、オブジェクト、実行可能ファイル、実行のスレッド、プログラム、および/またはコンピュータとすることができる。例示として、サーバ上で実行されるアプリケーションもサーバも両方とも、コンポーネントとすることができる。1つまたは複数のコンポーネントが1つのプロセスおよび/または実行スレッド内にあるものともでき、1つのコンポーネントが、1つのコンピュータ上に局所化されてもよく、複数のコンピュータ間で分散されてもよい。また、これらのコンポーネントは、様々なデータ構造が記憶された様々なコンピュータ可読媒体から実行することができる。コンポーネントは、1つまたは複数のデータパケット（例えば、ローカルシステムや分散システム中の別のコンポーネントと対話するコンポーネントからのデータ、および/または、インターネットなどのネットワークを介して他のシステムと信号によって対話するコンポーネントからのデータ）を有する信号などに従って、ローカルおよび/またはリモートプロセスを介して通信することができる。

【 0 0 1 3 】

最初に図1を参照すると、本発明の一態様によるオーディオ重複検出器システム100が示されている。システム100は、重複検出器120によって処理されるオーディオファイルのデータベース110を備える。重複検出器は、例えば重複オーディオファイルが存在するか否か、ファイルが破損しているかどうかなどの特徴を判定するためにオーディオファイルを処理する分析器140からの中間結果を保持するための、1つまたは複数の内部データベース124および130を備える。フィンガプリントコンポーネントおよび正規化コンポーネントが、分析器140内で動作して、オーディオファイルの一部が他のファイルと重複しているかどうかを判定し、マッチ検出器が、このような部分を重複（ま

10

20

30

40

50

たは破損ファイルやジャンクファイルなど、他の何らかの指示)としてタグ付けする。ユーザインターフェース150が、特徴の中でもとりわけ、ユーザがシステムパラメータ(例えば検出分析パラメータ)を入力できるようにするための入力、ならびに、削除するファイルを選択できるようにするための入力を備える。インターフェース150からの出力には、例えば、削除されるオーディオファイルのリスト、およびシステムに100に対する構成オプションのリストを含めることができる。

【0014】

一態様では、フィンガプリントコンポーネントは、ロバストオーディオ認識エンジン(RARE、Robust Audio Recognition Engine)を利用するが、これについては後でより詳細に述べる。一般に、ユーザは、データベース110を検索する際の起点となる最上レベルディレクトリ(「ルート」と呼ばれる)を選択し、次いでユーザインターフェース150を介して検出器120を始動させる。通常、ユーザは2つ(またはそれ以上)のパラメータをシステムに供給する。すなわち、オーディオファイル開始から何秒でそれぞれのフィンガプリントを抽出するか(T)、およびどのサイズの「スロップウィンドウ」を使用するか(S)である。Sは秒でも測定され、これについては後述する。次いで検出器120は、ルートディレクトリから下にサブディレクトリを検索してオーディオファイルを取り出すが、検出器はこれらをその拡張子(例えば、MP3、.WMA、.WAVファイル)で識別する。検出器120は、オーディオファイルを見つけると、このファイルをロードし、ファイル開始後T-S秒からファイル開始後T+S秒までの一連のフィンガプリントを計算する。一例では、6秒のオーディオから生成されるフィンガプリントであって、64個の浮動小数点数および1個の正規化係数(これもやはり浮動小数点数である)からなるフィンガプリントが計算される。理解できるように、その他の時間値、およびその他の数の浮動小数変数を利用することもできる。

【0015】

前述のように、検出器120は、少なくとも2つの内部データベース124および130を利用し、これらはDB1およびDB2とも呼ばれる。当技術分野で知られているように、これらのデータベースは、別々のデータベースである必要はない。これらは同じデータベース中の2つのテーブルであってもよく、あるいはデータベースソフトウェアを必要とせずにC++などの標準的なソフトウェアで実現することさえできる。DB1中では、レコードは概して2つのオブジェクトからなる。すなわち、フィンガプリントとそれに関連する正規化である。DB2中では、レコードは概して3つのオブジェクトからなる。すなわち、ファイル名(一般にファイル名は、ファイルの名前と共に、完全なパス名を含むものとされる)と、IDインデックスと呼ばれる関連するインデックスと、「オフセット」および「距離」と呼ばれる関連する2つの出力パラメータである。

【0016】

IDインデックスの目的の1つは、どのファイルが重複として識別されるかを常に把握しておくことである。すなわち、オーディオファイルの処理後に、2つのファイルが同じIDインデックスを有する場合は、これらは重複として識別されており、異なるIDインデックスを有する場合は、重複ではないとして識別されている。これらのデータベースは通常はシステム内部にあるが、外部データベースを採用することもできる。また、検出器120は、通常はフィンガプリントを計算するために約2MBのデータをロードするが、動作のために外部フィンガプリントデータベースを必要とはしない。

【0017】

第1のオーディオファイルがロードされると、位置Tで開始するフィンガプリントが計算されて記憶され、ファイルの名前および位置も記憶される。一般に、すべてのファイルのフィンガプリントが記憶されるわけではないが、ロードされたすべてのファイルについては、それらのフィンガプリントが計算され、それらの名前および位置が記憶される。第2のオーディオファイルがロードされると、このファイル中の位置T-Sでフィンガプリントが計算され、このフィンガプリントが、第1のファイルから計算されたフィンガプリントと比較される。マッチがある場合は、検出器120は、これらのIDインデックスを

同じ数に設定することによって（処理済みの第1のファイルのIDインデックスは0に設定されている）、2つのファイルが同じオーディオを含むことを記録してから、次のファイルに移る。マッチがない場合は、検出器は、第2のファイルのフィンガプリントを、ファイル開始後 $T - S + D$ 秒から計算する。Dはステップサイズである（例えば $D = 186$ ミリ秒）。

【0018】

以上のことから、一方または両方のファイルの最初の一部が欠けている場合があっても、オーディオファイルを重複として識別することができる。あるオーディオファイルのフィンガプリントFがデータベース中にあり、新しいファイルがFに対してテストされている場合、マッチが宣言されるためには、ウィンドウ（ファイル開始後 $T - S$ 秒で始まりファイル開始後 $T + S$ 秒で終わる）中で計算されたフィンガプリントのうちの1つがFにマッチすればよい。パラメータSはユーザ選択とすることができ、したがってユーザは、一方または両方のオーディオファイルの最初の一部が除去されることに対する頑強性に対して、重複検出速度のバランスをとることができる。速度はまた、次のように比較を順序付けることによってさらに高めることもできる。すなわち、最初に、ファイル開始後T秒で計算されたフィンガプリントをフィンガプリントデータベースに対してチェックし、次いで $T - D$ 秒でチェックし、次いで $T + D$ 秒、次いで $T - 2D$ 秒、次いで $T + 2D$ 秒のように、ウィンドウ全体（ $2S$ 秒の継続時間）がチェックされるまで同様にする。このようにして、最もマッチの可能性の高い位置が最初にチェックされ、マッチが見つかったら、システムはスロップ（slop）ウィンドウにわたるフィンガプリンティングを終了する（以下「保釈する」（bailing）と言う）ことができ、これによってプロセスを高速化することができる。

【0019】

上のプロセスは、マッチが見つかるまで、またはスロップウィンドウ中のすべてのフィンガプリント（すなわち第2のファイル中の $T - S$ 秒で開始するフィンガプリントから $T + S$ 秒で開始するフィンガプリントまで）が計算されるまで継続する。マッチが見つかった場合、検出器は、ファイル2のIDインデックスをDB2中のファイル1のIDインデックスと等しく（この場合は0に）設定することによって、ファイル2がファイル1と重複していることを示すフラグをセットし、次いで「保釈する」（ファイル2の処理を停止する）。マッチが見つからない場合は、検出器は以下のことを実施する。

（1）第2のファイル開始後T秒における（すでに計算された）フィンガプリントをDB1に追加する。

（2）対応する正規化を計算して保存し、このデータもDB1に保存する。

（3）そのIDインデックスを `id_index_max` に設定し、DB2に保存する。

（4）カウンタ `id_index_max` をインクリメントする。`id_index_max` は通常は0に初期化されることに留意されたい。このようにすれば、`id_index_max` は実際、ここまでで見つかった異なるオーディオクリップの総数になる。

【0020】

本発明の代替の一態様では、検出器120は、保釈の代わりに、スロップウィンドウ中のすべてのフィンガプリントを計算して比較し、第2のファイル中で最もよくマッチする位置を見つける。例えば、第2のファイル中の計算された第3のフィンガプリントが第1のファイル中のフィンガプリントにマッチしても、後続のフィンガプリントがよりよいマッチをもたらすかもしれず、そのため、最良のマッチを見つけるためにスロップウィンドウ中のすべてのフィンガプリントが計算される。これは、第2のファイル中のフィンガプリントが、予想された位置よりもどれだけ遠かったかを計算するのに有用である。この分析は、一方のファイルが他方のファイルに対してオーディオの最初で削られていること（clip）を示し、また、廃棄されたオーディオの継続時間も提供する。この態様では、DB2に記憶される2つの出力パラメータのうちの一方は、最もよくマッチするフィンガプリントが見つかった位置と、削られていなかったら存在していたと思われる位置との隔

たりを秒で表した継続時間である。この態様ではまた、マッチが見つかった場合に第2のパラメータ、すなわち最良のマッチに関連する「距離」もDB2に記憶される。距離は、比較された2つのフィンガプリントがどれだけ異なるかを示す尺度であり(2つのフィンガプリントが同一の場合は0)、したがって、マッチした2つのオーディオファイルがどれだけ異なるかを示す尺度として使用することができる。例えば、マッチした2つのファイルは同じ曲のリミックスである可能性があり、距離は、しきい値未満ではあるが相応に大きい可能性がある。

【0021】

保釈が採用されるか否かにかかわらず、検出器120は、記憶されたフィンガプリントと計算されたフィンガプリントとの間の距離に基づいて、マッチを宣言するか否かを決定する。距離が固定しきい値未満の場合はマッチが宣言され、代替の態様では、最低のこのような距離(スロップウィンドウ中のすべてのフィンガプリントをデータベース中のすべてのフィンガプリントと比較することによって計算されたもの)がDB2に記憶される。オフセットおよび距離の計算を常に行うのではなく別のルーチンで行う理由の1つは、これらの量を計算するのは計算コストがより高い(したがってより遅い)からである。というのは、これを行うためには、スロップウィンドウ中のすべてのフィンガプリントが計算されて比較されるからである。保釈が採用される本発明の第1の態様では、検出器はマッチが見つかったときに保釈することができる。この場合でも関連の距離およびオフセットを保存することができるが、このようにして計算されたオフセットは実際のオフセットの大まかな概算であり、このようにして計算された距離は実際の(最低の)距離の上限ではないので、この情報の有用性は劣る可能性がある。

【0022】

上記の後、次いで検出器は、ユーザによって選択されたディレクトリツリー中のすべてのオーディオファイルを同様にして反復する。すなわち、各オーディオファイルがロードされ、そのフィンガプリントがスロップウィンドウ中で計算され、このようないずれかのフィンガプリントがDB1中の既存のフィンガプリントとマッチする場合は、上記のようにマッチが宣言されて保存され、そうでない場合は、このオーディオクリップはそれまでにロードされたすべてのオーディオクリップとは異なると宣言され、そのフィンガプリント(T秒で計算されたもの)がDB1に保存され、そのIDインデックスがDB2中でid_index_maxに設定され、id_index_maxはインクリメントもされる。また、まだロードされていない他のオーディオクリップに対応するフィンガプリントとの比較のために、オーディオファイル正規化が計算されてDB1に保存される。このようにして、各ファイルは1度だけロードされ、ファイル进行处理するには1つのパスしか必要ない。ファイルが処理された後、DB2を使用して重複ファイルのリストがユーザインターフェース160に出力される。これを達成する方法の1つは、すべてのファイル名を印刷することであり、ファイル名とファイル名の間が「空白行なし」であるのは、これらが重複していることを示し、空白行は、空白行の上と下のファイルが重複していないことを示す。前述の代替態様では、印刷されたそれぞれのファイル名の後に、それに対応する距離およびオフセット(重複リスト中の最初のファイルでは両方とも0)が続く。

【0023】

次に図2を参照すると、本発明による重複プロセッサ200および関連する処理態様が示されている。重複プロセッサ200は、オーディオファイル管理を容易にするために、図1に関して上述したことに加えてオーディオファイルをさらに処理することを含む。これには、210で、検出されたエラー状況にフラグを立てること、および220で、このようなエラーに関連するファイルをリストすることを含めることができる。ファイルをロードしようとしたときに、ファイルを読み取ることができない場合、またはヘッダがオーディオファイル中で予想されるヘッダと一致しない場合、またはその他の何らかのエラー状況に遭遇した場合は、プロセッサ200は単に、そのファイルに関するエラー状況の記録を作成して次のファイルに移行する。次いで、ファイル処理が完了したときにシステムはまず、遭遇したファイルの名前を、そのファイル进行处理することができなかった理由(

10

20

30

40

50

例えば「サポートされないサンプリングレート」や「D R M制限のせいで開くことができない」と共に出力する。このリストは、220の「破損したファイル」と題した単一のブロックで出力することができる。

【0024】

別の有用な機能は、オーディオとしてうまくロードされたものの例えば雑音や他の何らかの欠陥を含んでいるファイルに適用されるものである。各ファイルごとに、他の処理が適用される前に230で、そのファイルから抽出されたフィンガプリントが「拒否権 (veto)」フィンガプリントのリストと比較される。「拒否権」フィンガプリントは、白雑音や、入力がないときの様々なサウンドカードから録音されたオーディオなど、少数の特別なタイプのオーディオのフィンガプリントおよび正規化を計算することによって事前に計算されたものである。当該のファイルのフィンガプリントがこれらの「拒否権フィンガプリント」のいずれかとマッチする場合は、このファイルは「ジャンク」であることがわかり、そのようにフラグを立てることができる。この場合も、オーディオファイルの処理が完了したとき、240の「ジャンクファイル」と題した1つのブロックでこれらの「ジャンク」ファイルをリストすることができる。このようにしてシステムは、重複を見つけるだけでなく、ユーザがチェックすることができおそらく削除したいであろう、その他のファイルを識別することもできる。

10

【0025】

図3に移ると、本発明の一態様によるユーザインターフェース処理300が示されている。所与のファイルに何らかの重複または破損が見つかったとき、システムは、重複を比較して、どちらを保存するかについての様々なオプションをユーザインターフェース310を介してユーザに提供することを試みることができる。削除のためのユーザインターフェース310は、2つまたはそれ以上の段階に分割することができる。320の段階1で、削除される可能性のあるものとしてファイルがラインナップされる。330の段階2で、ユーザは、それぞれのファイルのどれを実際に削除するか選択する。

20

【0026】

320の段階1に関して、本発明は、(1)340で品質比較に基づいて1つまたは複数の重複を保存する能力をユーザに提供することができ、(2)350でユーザが使用する符号化に基づいてファイルの優遇を提供することができ、かつ/または(3)360でデジタル権利管理に基づいて優遇を提供することができる。例えば、品質比較340で、2つのファイルが同じ符号化タイプを有し(例えば両方ともWMA、または両方ともMP3)、異なるビットレートを有するが他の点では同様である場合は、ビットレートが高い方(したがって品質が高い方)のファイルを保存することができる(すなわち他方のファイルは削除される可能性のあるものとしてラインナップされる)。同様に、2つのファイルが同じ符号化タイプを有し、一方が他方よりも継続時間が長いことを除いてはすべての点で同様である場合は、継続時間が長い方のファイルを保存することができる(したがって、削除されるファイルは通常、削られた部分のあるファイルになる)。一般に、2つのファイルが1点を除くすべての点で同様である場合は、品質が低い方のファイルが、削除される可能性のあるものとして選択されることになる。あるいは、ユーザの目的が主にディスク空間を増やすことである場合は、ユーザは、品質は劣るが小さい方のファイルを保存することを選択することもできる。

30

40

【0027】

350での符号化に関して、一方のファイルの符号化タイプがWMAであり、他方のファイルの符号化タイプがMP3である場合、WMAが保存される可能性がある。これは、どちらの符号化を好むかをコンテンツ/サービスプロバイダが選択できることに対処する。例えば、ある会社は、様々なビジネス上の理由でWMA符号化を好む場合がある。360でのデジタル権利管理に関して、あるファイルが他のファイルよりも強力なデジタル管理権利保護を受ける場合は、他のファイルの方を、削除される可能性のあるものとしてラインナップされるファイルとすることができる(例えば、あるファイルはD R Mによりコピー可能でないが、他のファイルはコピー可能である場合は、他のファイルの方を削

50

除されるものとしてラインナップすることができる)。この態様では、重複検出を使用して、このユーザが同意した最も強力なDRM制約に対して所与の曲を「整合」させている。当然、ユーザが故意に同じ曲のコピーを様々な強度のDRMで欲するという状況がある場合もあり、この状況は330の段階2で対処することができる。理解できるように、上の例はファイル間の2方向の比較に基づいていたが、多方向の比較を適用することもできる。

【0028】

330の段階2に関して、370で、どのファイルを実際に削除するかについての様々な選択子がユーザに提示され、ユーザからの確認後、削除されるように選択されたファイルが削除される。ユーザには、ファイルが本当に重複であることを重複検出器がどれだけ確信しているかに基づいて、様々なレベルの警告を提示することもできる。さらに段階2では、ユーザインターフェース310は、2つの重複を比較するための「高速閲覧」機能をユーザに提供することもできる。例えば、ファイル(A)とファイル(B)が重複であることがわかっており、(A)の方が長いことを除いては(A)と(B)は同様であると仮定する。重複検出器は、その動作の中で、2つのコピー中のどこにマッチするフィンガプリントがあるかを決定することができるので、また、各オーディオの総継続時間を決定するので、最初で(A)が(B)よりも長いのか、最後で(A)が(B)よりも長いのか、あるいはこの両方なのかかわかる。次いでユーザインターフェース310は、ファイル(A)の中の、ファイル(B)に含まれない可能性が非常に高い部分を再生することができ、ユーザは、本当に2つのファイルのうちの長い方を保存したいかを決定することができる(例えばユーザは、違いが単に拍手または静寂でしかない場合は、2つのうちの短い方を好む場合もある)。

【0029】

さらに、2つのファイルが本当に重複であるかどうかの確信がユーザになく、各ファイルを別々に聴かなければならないのは嫌である場合、ユーザは、重複検出器が再び両方のファイルに対して実行されるよう、ただし今回は様々な位置の多くのフィンガプリントを使用して実行されるよう要求することができる。それにより、2つのファイルのどの部分が実際に重複しているかを決定することができる。例えば、一方のファイルが他方のファイルよりも3倍長い場合、これらは最初だけが同一である場合がある。このようにして、ユーザインターフェース310は、重複検出器を再び呼び出して、特定のファイルに関する情報をさらに得ることができる。これは、重複検出器が最初にユーザのオーディオファイルすべてに対して実行される間に計算するには、計算コストが高すぎるであろう。上記の検出の態様はすべて、ビデオの重複を識別するのにとも(例えばオーディオトラックの重複を検索することによって)利用できることに留意されたい。

【0030】

図4は、本発明の一態様によるオーディオ重複処理を示す流れ図である。説明を簡単にするために、方法を一連の動作として図示および記述するが、いくつかの動作は、本発明により、本明細書に図示および記述するのとは異なる順序で行うこともでき、かつ/または他の動作と同時に行うこともできるので、本発明は動作の順序によって限定されないことを理解および認識されたい。例えば、別法として、相互に関係付けられた一連の状態またはイベントとして状態図などに方法を表すこともできることは、当業者なら理解および認識するであろう。さらに、本発明による方法を実施するために、図示したすべての動作が必要であるとは限らない。

【0031】

ロバストオーディオ認識エンジン(RARE)重複検出器DupDetが、図4の基本プロセス400に従って動作することができ、例えばディレクトリツリー中のオーディオファイルを再帰的に処理する。各フィンガプリントから、他のオーディオファイルから計算されたフィンガプリントの大きな集まりへの平均ユークリッド距離が1つになるように、生み出される各フィンガプリントごとに正規化係数も生み出される。これは、フィンガプリントの対の間で計算された距離を有益に比較することができるように実施される。

【 0 0 3 2 】

プロセス 4 0 0 は、各ファイルごとにトレースのセットを生成し、これを他のオーディオファイルについて生成されたフィンガプリントのセットに対してチェックする。ここで、用語「トレース」は、データベースに対してテストされるファイルから計算されたフィンガプリントについて用い、「フィンガプリント」は、データベースに記憶されたフィンガプリントについて用いる。トレースとフィンガプリントとの間の $D(\cdot, \cdot)$ (距離) がしきい値未満である場合、関連するオーディオファイルは重複であると宣言される。各ファイルにつき、フィンガプリントはファイル中の固定位置 T で計算され、トレースは、4 1 0 で、 T を中心とした検索ウィンドウ S 中で計算される。 S および T はユーザ定義である。

10

【 0 0 3 3 】

4 2 0 に示すように、プロセス 4 0 0 は、フィンガプリントの生成と重複のチェックとを同時に 1 つのパスで行うことができる。4 3 0 で第 1 のオーディオファイルが読み取られると、位置 T の 6 秒フィンガプリントが計算され保存される。第 2 のオーディオファイルがロードされると、4 1 0 で、 $T - S$ から $T + S$ までのウィンドウ中で開始するトレースが時間順に計算される (各トレースは前のトレースから 6 分の 1 秒後に計算される)。4 4 0 でトレースのうちの 1 つがマッチである場合、ファイルは重複であると宣言され、4 5 0 でこのフィンガプリントについての重複リストに追加される。検索ウィンドウ中のトレースセット全体でマッチが見つからない場合は、4 6 0 でフィンガプリント (すでにオーディオ中の位置 T で計算されたもの) がデータベースに保存され、これは (今のところ) 固有のクリップを表す。システムはまた、雑音 (例えば静寂や、入力なしのサウンドカードなど) から収集されたフィンガプリントである 6 個 (またはその他の数) の「拒否権フィンガプリント」も使用する。拒否権フィンガプリントにマッチするオーディオファイルには、「ジャンクファイル」とラベル付けすることもできる。図示のように、4 7 0 でマッチに重複としてタグ付けすることができ、4 8 0 および 4 9 0 で、最良マッチスコア、オフセット、マッチするフィンガプリント ID を保存することに関係するその他の処理を行うことができる。

20

【 0 0 3 4 】

図 5 に、本発明の一態様によるオーディオフィンガプリンティングシステム 5 0 0 を示す。ストリームオーディオフィンガプリンティング (S A F) では、オーディオストリーム中でオーディオセグメントを識別することがそのタスクだが、ストリームは雑音で破損している場合がある。図 5 のシステム 5 0 0 に全体的なセットアップを示す。最初に、入来オーディオストリーム 5 1 0 の固定長セグメントが、5 2 0 の低次元トレース (期間として示されているベクトル) に変換される。次いでこの入力トレース 5 2 0 は、5 3 0 の、記憶された事前計算済みのトレース (フィンガプリント) の大きなセットと比較される。記憶された各フィンガプリントは、あらかじめ特定のオーディオセグメント (例えば曲) から抽出されたものである。

30

【 0 0 3 5 】

入力トレース 5 2 0 は、ストリーム 5 1 0 中の反復される期間で計算され、データベース 5 3 0 と比較される。次いで、データベース中で見つかった入力トレースは、2 次的なフィンガプリントを使用して、ごくわずかな追加の計算コストで確認することができる。代表的な適用例には、ロイヤルティ評価のためや、コマーシャルがスポンサーに対するサービスとして放送されたことを確認するために放送オーディオを識別すること、ソフトウェアプレーヤがユーザによって生成された CD 上のトラックを識別できるようにすること、ラベル付けされていないオーディオに関するメタデータを見つけること、大規模なオーディオデータベース中で自動的に重複を検出することが含まれる。本発明は、オーディオから雑音頑強特徴を自動的に抽出するために、歪み判別分析 (DDA 、 $Distortion Discriminant Analysis$) と呼ばれるアルゴリズムを採用することができる。 DDA 特徴は、線形かつ重畳のニューラルネットワークによって計算され、各レイヤが、指向性主成分分析 ($OPCA$ 、 $oriented Principal$

40

50

Components Analysis) 次元縮小のバージョンの1つを実施する。
【0036】

歪みに対する頑強性を構築するために、DDAは、訓練信号セットの歪んだバージョンが利用可能であると仮定する。歪んだ信号のサンプルを必要とすることは、実際の雑音モデルを知る必要があることよりも厳しくなく、より一般的である。DDAは一般に、歪みが加法的であるとは仮定せず、非線形の歪みも扱う。テスト段階で予想される特定の歪みに対して訓練できることが有用な場合もあるが、DDAは、訓練に使用されない歪みに対して頑強であるという点で、一般化することができる。事前計算済みのトレースは、オーディオセグメントを一意に識別するのに使用されるので「フィンガプリント」と呼ばれる。通常、1つのオーディオクリップにつき1つまたは2つのフィンガプリントが使用されるが、より多くのフィンガプリントを使用すれば、エラーレートをさらに抑えることができる。また、以下の記述ではいくらかの数学的背景が提供されるが、ベクトルは太字フォントで表され、それらの成分は通常のフォントで表され、ダッシュは転置を表す。

10

【0037】

ベクトル集合

【0038】

【数1】

$$\mathbf{x}_i \in \mathbb{R}^d$$

【0039】

, $i = 1, \dots, m$ がある場合に、各

20

【0040】

【数2】

$$\mathbf{x}_i$$

【0041】

は信号を表し(この箇所および以下では、歪みのないデータを「信号」データと呼ぶ)、各

【0042】

【数3】

30

$$\mathbf{x}_i$$

【0043】

ごとに、N個の歪んだバージョンの集合

【0044】

【数4】

$$\tilde{\mathbf{x}}_i^k$$

【0045】

, $k = 1, \dots, N$ があると仮定する。対応する差分ベクトル

40

【0046】

【数5】

$$\mathbf{z}_i^k \equiv \tilde{\mathbf{x}}_i^k - \mathbf{x}_i$$

【0047】

(以下、「雑音」ベクトルと呼ぶ)を定義する。一般に、すべてのkの場合の

【0048】

【数6】

$$\mathbf{z}_i^k$$

50

【 0 0 4 9 】

にできるだけ直交するが、これに沿って元の信号

【 0 0 5 0 】

【 数 7 】

$\mathbf{x}_i$

【 0 0 5 1 】

の分散が最大化される、線形射影を見つけることが望まれる。望まれる射影を定義する単位ベクトルを

【 0 0 5 2 】

10

【 数 8 】

$\mathbf{n}_i$

【 0 0 5 3 】

, $i = 1, \dots, M$ で表すが、 M はユーザによって選択されることになる。この考察は、一時的に $M = 1$ を選択することによって単純化することができる。

【 0 0 5 4 】

平均二乗再構築誤差 ($1 / m N$)

【 0 0 5 5 】

【 数 9 】

20

$$\sum_{i,k} (\mathbf{x}_i - \hat{\mathbf{x}}_i^k)^2$$

【 0 0 5 6 】

を最小化する特徴抽出子

【 0 0 5 7 】

【 数 1 0 】

$\mathbf{n}$

【 0 0 5 8 】

を構築することができ、

【 0 0 5 9 】

30

【 数 1 1 】

$$\hat{\mathbf{x}}_i^k \equiv (\tilde{\mathbf{x}}_i^k \cdot \mathbf{n}) \mathbf{n}$$

【 0 0 6 0 】

である。 R_1 および R_2 をそれぞれ

【 0 0 6 1 】

【 数 1 2 】

$\mathbf{x}_i$

【 0 0 6 2 】

40

および

【 0 0 6 3 】

【 数 1 3 】

$\mathbf{z}_i$

【 0 0 6 4 】

の相関行列とすると、この問題を解決する

【 0 0 6 5 】

【 数 1 4 】

$\mathbf{n}$

50

【 0 0 6 6 】

が、最大固有値を有する $R_1 - R_2$ の固有ベクトルであることを証明するのは単純である。しかし、この特徴抽出子は、雑音および信号のベクトルが 2 つの異なるスケール係数で大域的にスケールされた場合に方向

【 0 0 6 7 】

【 数 1 5 】

$\mathbf{n}$

【 0 0 6 8 】

が変化するという望ましくない特性を有する。そうではなく、OPCA (Oriented PCA) の方向は、以下の一般化されたレイリー商を最大化する方向

10

【 0 0 6 9 】

【 数 1 6 】

$\mathbf{n}$

【 0 0 7 0 】

として定義される。

【 0 0 7 1 】

【 数 1 7 】

$$q_0 = \frac{\mathbf{n}'\mathbf{C}_1\mathbf{n}}{\mathbf{n}'\mathbf{C}_2\mathbf{n}} \quad (1)$$

20

【 0 0 7 2 】

上式で、 $\mathbf{C}_1$ は信号の共分散行列であり、 $\mathbf{C}_2$ は雑音の共分散行列である。ただし、元の形の OPCA とは対照的に、共分散行列ではなく雑音の相関行列が使用される。というのは、平均雑音信号ならびにその分散にペナルティを課すことが望まれるからである。明示的には、以下の式を利用する。

【 0 0 7 3 】

【 数 1 8 】

$$\mathbf{C} \equiv \frac{1}{m} \sum_i (\mathbf{x}_i - E[\mathbf{x}])(\mathbf{x}_i - E[\mathbf{x}])' \quad (2)$$

$$\mathbf{R} \equiv \frac{1}{mN} \sum_{i,k} \mathbf{z}_i^k (\mathbf{z}_i^k)' \quad (3)$$

30

そして、以下の一般化されたレイリー商を最大化する。

【 0 0 7 4 】

【 数 1 9 】

$$q = \frac{\mathbf{n}'\mathbf{C}\mathbf{n}}{\mathbf{n}'\mathbf{R}\mathbf{n}} \quad (4)$$

【 0 0 7 5 】

(4) における分子は、単位ベクトル

40

【 0 0 7 6 】

【 数 2 0 】

$\mathbf{n}$

【 0 0 7 7 】

に沿った信号データの射影の分散であり、分母は、射影された平均二乗「誤差」(

【 0 0 7 8 】

【 数 2 1 】

$\mathbf{n}$

【 0 0 7 9 】

に沿って射影されたすべての雑音ベクトル

50

【 0 0 8 0 】

【 数 2 2 】

 $\mathbf{z}_i^k$

【 0 0 8 1 】

の平均二乗モジュラス)である。

【 0 0 8 2 】

方向

【 0 0 8 3 】

【 数 2 3 】

10

 $\mathbf{n}_j$

【 0 0 8 4 】

は、 $q = 0$ に設定することによって見つけることができ、これは以下の一般固有値問題をもちたらず。

【 0 0 8 5 】

【 数 2 4 】

 $q = qR\mathbf{n}$

(5)

【 0 0 8 6 】

20

以下のことを証明するのは単純である。

【 0 0 8 7 】

1) 半正定値 (positive semidefinite) の C 、 R (この場合のように) では、一般固有値は正である。しかし、 R が全階数でない場合は、この問題が適切 (well-posed) になるように正則化される。

【 0 0 8 8 】

2) 信号と雑音のいずれかをスケールすると、固有値は変化することになるが、OPCA 方向は変化せずそのままである。

【 0 0 8 9 】

3)

30

【 0 0 9 0 】

【 数 2 5 】

 $\mathbf{n}_i$

【 0 0 9 1 】

は、一次独立であるか、またはそうなるように選択することができる。

【 0 0 9 2 】

4)

【 0 0 9 3 】

【 数 2 6 】

40

 $\mathbf{n}_i$

【 0 0 9 4 】

は必ずしも直交とは限らないが、行列 C と R の両方に対して共役である。

【 0 0 9 5 】

5)

【 0 0 9 6 】

【 数 2 7 】

 $\mathbf{n}$

【 0 0 9 7 】

50

を最高の重みの一般固有ベクトルになるように選択することによって、 q が最大化される。

【0098】

オーディオなどの高次元データの場合、OPCAはレイヤで適用することができる。例えば、6秒のオーディオから64次元のフィンガプリントを抽出することを考えてみる。オーディオ信号はモノに変換されて11025Hzにダウンサンプリングされ、後続の特徴抽出が、次元66150のベクトルを次元64のベクトルにマッピングする。この場合に一般固有値問題を直接に解決するのは不可能である。その代わり、OPCAを2つのレイヤで適用することができ、第1のレイヤは、小さいウィンドウにわたって計算された対数スペクトルに作用し、第2のレイヤは、第1のレイヤによって生成されたベクトルを集約することによって計算されたベクトルに作用する。この手法は、「歪み判別分析（DDA）」と呼ばれる。DDAは線形方法であり、所与のレイヤで生じる射影は重畳と見なすことができる。したがって、DDAは線形かつ重畳のニューラルネットワークと見なすことができ、重みはOPCAを使用して選択される。

10

【0099】

DDAでは、後続の各レイヤは、その前のレイヤよりも広い時間ウィンドウを見る。そのレイヤで見つかった固有方向は、その特定の時間スケールに適合する。これはDDAの特徴である。例えば、これを使用して整合雑音を補償することができ、整合雑音は、記憶済みのフィンガプリントが入力トレースと時間的に位相がずれている可能性があることから生じる雑音として定義される。最悪の場合、フィンガプリントは、隣接する2つの入力トレースを計算するのに使用された2つのフレームの間にあるフレームから計算されたものということになる。DDAシステムにおけるこのような時間的歪みの補償は、最後のレイヤ上で適用されるべきである。というのは、最後のレイヤは最も広い時間ウィンドウを見るからである。

20

【0100】

DDAは、テスト段階を計算効率の高いものにして、様々な時間スケールでの歪みの補償を可能にするだけでなく、訓練段階でも効率的である。必要とされる共分散行列および相関行列は、望むなら1度に1ベクトルずつ計算することができる。したがって、これらの行列は、自由裁量の多くの量のデータを使用して推定することができる。行列を推定した後、標準的な数値線形代数パッケージを使用して、一般固有値を計算することができる。

30

【0101】

図6に、本発明の一態様による歪み判別分析システム600を示す。オーディオ処理の技法、例えば音声から特徴を抽出する技法は、しばしば約20ミリ秒のフレーム持続時間を使用する。しかし、フィンガプリンティング適用例で計算オーバーヘッドを低減するためには、毎秒およそ数回にわたってストリームからトレースを生成することが望ましい。20ミリ秒の入力フレームの場合、最後のDDAレイヤで使用されるステップサイズは、100Hzの初期サンプリングレート未満でサンプリングしなければならないことになり、これはエイリアシング（aliasing）を引き起こす可能性があり、エイリアシングはさらに歪みの原因として作用することになる。図6に示すシステム600は、この問題を回避する。サンプリングレートが低下した中間レイヤがないので、概してエイリアシングはない。この要件のため、またトレースを約2分の1秒の時間スケールで生成するという要件のため、第1レイヤのフレームの可能な持続時間がかなり制約される。また、第1レイヤが時間的に広いことにより、DDAには、周波数空間で重要な方向を選択する際により多くのフレキシビリティがもたらされる。

40

【0102】

図7に、本発明の一態様による一般固有値700を示す。前述のシステム600の第1レイヤでの64出力次元の選択は、図7に示す、訓練データに対する測定された一般固有スペクトルによってガイドされる。第1レイヤからの有用な情報のほとんどは、最初の100個の射影で取り込まれる。第2レイヤ上のスペクトルは、それほど急速には降下しな

50

い。しかし、データベース検索を高速化するために、第2レイヤ上の上位64個の射影だけが考慮された。データベース検索の速度はさらに、出力を186ミリ秒ごとではなく372ミリ秒ごとにのみサンプリングすることによって、2倍にすることもできる。

【0103】

前述のストリームオーディオフィンガプリンティングシステムは、最初にステレオオーディオ信号をモノに変換し、次いで11025Hzにダウンサンプリングする。信号は、半分重なった固定長の372ミリ秒フレームに分割される。次いで、MC LT（重なったウィンドウによるフーリエ変換）が各フレームに適用される。各MC LT係数の対数モジュラスをとることによって、対数スペクトルが生成される。このストリームオーディオフィンガプリンティングシステムは、識別容易な特定の歪みを抑制する、フレームごとの2つの前処理ステップを実施する。

10

【0104】

第1の前処理ステップは、周波数等化およびボリューム調整によって引き起こされた歪みを除去する。この「等化解除しきい値処理」(de-equalization thresholding)ステップは、対数スペクトルのDCTを取り、第1成分に対する1から第6成分以上に対する0まで線形に傾斜する重みを各DCT係数に掛け、次いで逆DCTを実施することによって、対数スペクトルにローパスフィルタを適用する。この結果、対数スペクトルに対するスムーズな近似Aが得られる。次いでAは、一様に6dBだけ低減され、-70dBでクリッピングされる。この場合、第1の前処理ステップの出力ベクトルは、成分に関する対数スペクトルとAとの差だが、これはこの差が正である場合であり、そうでない場合は0である。

20

【0105】

第2の前処理ステップは、人間の聴取者に聴こえない信号の中の歪みを除去する。このステップは、第1のステップからの対数スペクトルを累乗し、次いで、アルゴリズムによって周波数依存の知覚しきい値を生成する。この場合、最終的な前処理済み信号は、対数スペクトルと対数知覚しきい値とのdBの差だが、これはこの差が正である場合であり、そうでない場合は0である。最終的な前処理済みデータは、1フレームあたり2048個の実係数(したがって2048個の帯域)からなる。

【0106】

図8を参照すると、本発明の様々な態様を実現するための例示的な環境910は、コンピュータ912を含む。コンピュータ912は、処理ユニット914、システムメモリ916、システムバス918を備える。システムバス918は、限定しないがシステムメモリ916を含めたシステムコンポーネントを、処理ユニット914に結合する。処理ユニット914は、利用可能な様々なプロセッサのいずれかとすることができる。デュアルマイクロプロセッサおよびその他のマルチプロセッサアーキテクチャを処理ユニット914として採用することもできる。

30

【0107】

システムバス918は、利用可能な様々なバスアーキテクチャのいずれかを用いた、メモリバスまたはメモリコントローラ、周辺バスまたは外部バス、および/あるいはローカルバスを含めて、いくつかのタイプのバス構造のいずれかとすることができる。バスアーキテクチャには、限定しないが16ビットバス、ISA(Industrial Standard Architecture)、MCA(Micro-Channel Architecture)、EISA(Extended ISA)、IDE(Intelligent Drive Electronics)、VESAローカルバス(VLB)、PCI(Peripheral Component Interconnect)、USB(Universal Serial Bus)、AGP(Advanced Graphics Port)、PCMCIA(Personal Computer Memory Card International Association)バス、SCSI(Small Computer Systems Interface)が含まれる。

40

50

【0108】

システムメモリ916は、揮発性メモリ920および不揮発性メモリ922を含む。不揮発性メモリ922には、起動中などにコンピュータ912内の要素間で情報を転送するための基本ルーチンを含むBIOS(basic input/output system)が記憶されている。限定ではなく例示として、不揮発性メモリ922には、読取り専用メモリ(ROM)、プログラム可能ROM(PROM)、電氣的にプログラム可能なROM(EPROM)、電氣的に消去可能なROM(EEPROM)、またはフラッシュメモリを含めることができる。揮発性メモリ920にはランダムアクセスメモリ(RAM)が含まれ、これは外部キャッシュメモリとして働く。限定ではなく例示として、RAMは、同期RAM(SRAM)、ダイナミックRAM(DRAM)、同期DRAM(SDRAM)、ダブルデータレートSDRAM(DDR SDRAM)、エンハンスドSDRAM(ESDRAM)、Synchlink DRAM(SLDRAM)、ダイレクト Rambus RAM(DRRAM)など、多くの形で入手可能である。

10

【0109】

コンピュータ912は、取外し可能/固定式、揮発性/不揮発性コンピュータ記憶媒体も備える。図8には、例としてディスク記憶装置924を示す。ディスク記憶装置924には、限定しないが磁気ディスクドライブ、フロッピー(登録商標)ディスクドライブ、テープドライブ、Jazドライブ、Zipドライブ、LS-100ドライブ、フラッシュメモリカード、メモリスティックなどのデバイスが含まれる。さらに、ディスク記憶装置924には、限定しないが、コンパクトディスクROMドライブ(CD-ROM)、CD記録可能ドライブ(CD-Rドライブ)、CD書換え可能ドライブ(CD-RWドライブ)、デジタル多用途ディスクROMドライブ(DVD-ROM)などの光ディスクドライブを含めて、別個の記憶媒体、あるいは他の記憶媒体と組み合わせた記憶媒体を含めることもできる。ディスク記憶デバイス924からシステムバス918への接続を容易にするために、通常、インターフェース926などの取外し可能または固定式インターフェースが使用される。

20

【0110】

図8には、ユーザと、適した動作環境910で述べた基本的なコンピュートリソースとの間の媒介として働くソフトウェアも記述してあることを理解されたい。このようなソフトウェアには、オペレーティングシステム928が含まれる。オペレーティングシステム928は、ディスク記憶装置924に記憶されているものとしてすることができ、コンピュータシステム912のリソースを制御および割振りする働きをする。システムアプリケーション930が、システムメモリ916とディスク記憶装置924のいずれかに記憶されたプログラムモジュール932およびプログラムデータ934を介して、オペレーティングシステム928によるリソース管理を利用する。本発明は、様々なオペレーティングシステム、またはオペレーティングシステムの組合せを使用して実施できることを理解されたい。

30

【0111】

ユーザは、入力デバイス936を介してコンピュータ912にコマンドおよび情報を入力する。入力デバイス936には、限定しないが、マウスなどのポインティングデバイス、トラックボール、スタイラス、タッチパッド、キーボード、マイクロホン、ジョイスティック、ゲームパッド、衛星受信アンテナ、スキャナ、TVチューナカード、デジタルカメラ、デジタルビデオカメラ、ウェブカメラなどが含まれる。これらおよび他の入力デバイスは、インターフェースポート938を介してシステムバス918を経由し処理ユニット914に接続する。インターフェースポート938には、例えばシリアルポート、パラレルポート、ゲームポート、ユニバーサルシリアルバス(「USB」)が含まれる。出力デバイス940は、入力デバイス936と同じタイプのポートのいくつかを使用する。したがって例えば、USBポートは、コンピュータ912への入力を提供するのに使用することができ、またコンピュータ912から出力デバイス940に情報を出力するのにも使用することができる。出力デバイス940の中でもとりわけ、モニタ、スピーカ、ブ

40

50

リントなど、特別なアダプタを必要とするいくつかの出力デバイス 940 があることを例示するために、出力アダプタ 942 が提供されている。出力アダプタ 942 には、限定ではなく例示として、出力デバイス 940 とシステムバス 918 との間の接続手段を提供するビデオカードおよびサウンドカードが含まれる。リモートコンピュータ 944 など、その他のデバイス、および/またはデバイスのシステムも、入力と出力の両方の機能を提供することに留意されたい。

【0112】

コンピュータ 912 は、リモートコンピュータ 944 など 1 つまたは複数のリモートコンピュータへの論理接続を用いて、ネットワーク化された環境で動作することができる。リモートコンピュータ 944 は、パーソナルコンピュータ、サーバ、ルータ、ネットワーク PC、ワークステーション、マイクロプロセッサベースの機器、ピアデバイス、またはその他の一般的なネットワークノードなどとしてすることができ、通常はコンピュータ 912 に関して述べた要素の多くまたはすべてを備える。簡単にするために、メモリ記憶デバイス 946 だけがリモートコンピュータ 944 と共に示されている。リモートコンピュータ 944 は、ネットワークインターフェース 948 を介してコンピュータ 912 に論理接続され、次いで通信接続 950 を介して物理接続される。ネットワークインターフェース 948 は、ローカルエリアネットワーク (LAN) やワイドエリアネットワーク (WAN) などの通信ネットワークを含む。LAN 技術には、ファイバ分散データインターフェース (FDDI)、銅分散データインターフェース (CDDI)、イーサネット (登録商標) / IEEE 1102.3、トークンリング / IEEE 1102.5 などが含まれる。WAN 技術には、限定しないがポイントツーポイントリンク、統合デジタルサービス通信網 (ISDN) やその変形などの回路交換ネットワーク、パケット交換ネットワーク、デジタル加入者線 (DSL) が含まれる。

【0113】

通信接続 950 は、ネットワークインターフェース 948 とバス 918 を接続するのに利用されるハードウェア/ソフトウェアを指す。通信接続 950 は、図示をわかりやすくするためにコンピュータ 912 の内部に示してあるが、コンピュータ 912 の外部にあってもよい。ネットワークインターフェース 948 への接続に必要なハードウェア/ソフトウェアには、例示に過ぎないが、通常の電話グレードモデムやケーブルモデムや DSL モデムを含めたモデム、ISDN アダプタ、イーサネット (登録商標) カードなどの、内部技術および外部技術が含まれる。

【0114】

図 9 は、本発明が対話することのできるサンプルコンピューティング環境 1000 の概略ブロック図である。システム 1000 は、1 つまたは複数のクライアント 1010 を含む。クライアント 1010 は、ハードウェアおよび/またはソフトウェア (例えばスレッド、プロセス、コンピューティングデバイス) とすることができる。システム 1000 は、1 つまたは複数のサーバ 1030 も含む。サーバ 1030 もまた、ハードウェアおよび/またはソフトウェア (例えばスレッド、プロセス、コンピューティングデバイス) とすることができる。サーバ 1030 は、例えば、本発明を利用することによって変形を実施するためのスレッドを収容することができる。クライアント 1010 とサーバ 1030 との間の可能な対話の 1 つは、複数のコンピュータプロセス間で伝送されるように適合されたデータパケットの形とすることができる。システム 1000 は、クライアント 1010 とサーバ 1030 との間の通信を容易にするのに利用できる通信フレームワーク 1050 を含む。クライアント 1010 は、クライアント 1010 のローカルな情報を記憶するのに利用できる 1 つまたは複数のクライアントデータ記憶装置 1060 に動作可能に接続されている。同様に、サーバ 1030 は、サーバ 1030 のローカルな情報を記憶するのに利用できる 1 つまたは複数のサーバデータ記憶装置 1040 に動作可能に接続されている。

【0115】

以上に述べたことは、本発明の例を含む。当然、本発明を記述するために、考えられる

あらゆる組合せのコンポーネントおよび方法について述べることは不可能である。しかし、本発明に関するその他多くの組合せおよび置換が可能であることは、当業者なら理解するであろう。したがって本発明は、添付の特許請求の趣旨および範囲に含まれるすべての改変、修正、変形を含むものとする。

【図面の簡単な説明】

【0116】

【図1】本発明の一態様によるオーディオ重複検出器システムの概略ブロック図である。

【図2】本発明による重複処理態様の図である。

【図3】本発明によるユーザインターフェース態様を示す図である。

【図4】本発明の一態様によるオーディオ重複処理を示す流れ図である。

【図5】本発明の一態様によるオーディオフィンガプリンティングシステムを示す図である。

【図6】本発明の一態様による歪み判別分析を示す図である。

【図7】本発明の一態様による一般固有値を示す図である。

【図8】本発明の一態様による適した動作環境を示す概略ブロック図である。

【図9】本発明が対話することのできるサンプルコンピューティング環境の概略ブロック図である。

【符号の説明】

【0117】

110 データベース、オーディオファイル

120 重複検出器

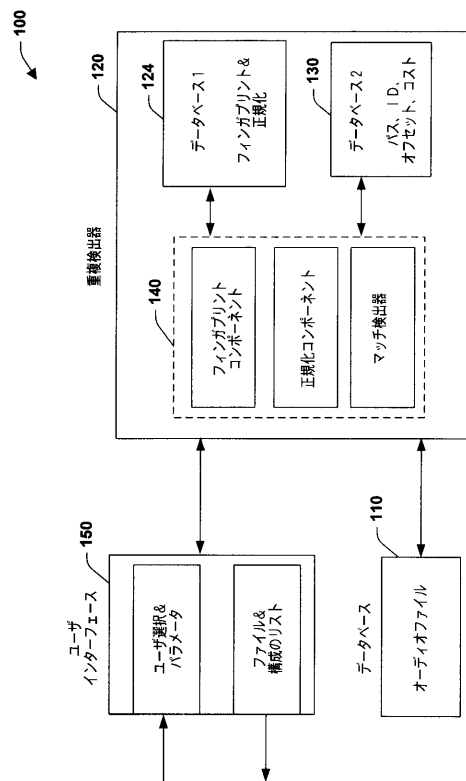
124 データベース1、フィンガプリント&正規化

130 データベース2、パス、ID、オフセット、コスト

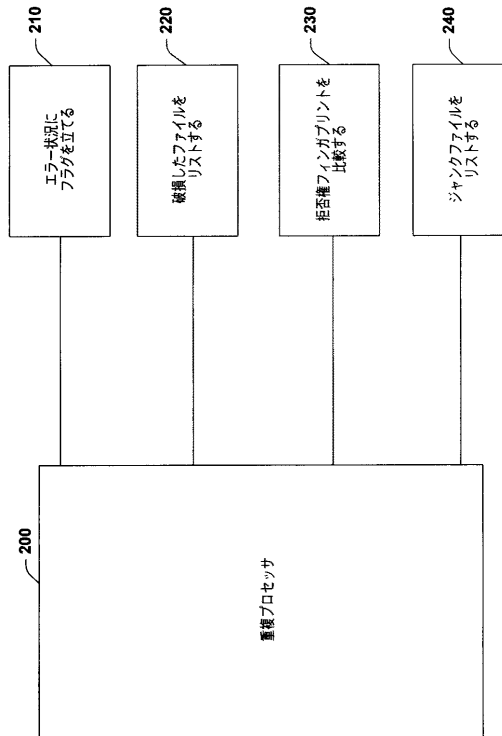
140 フィンガプリントコンポーネント、正規化コンポーネント、マッチ検出器

150 ユーザインターフェース、ユーザ選択&パラメータ、ファイル&構成のリスト

【図1】



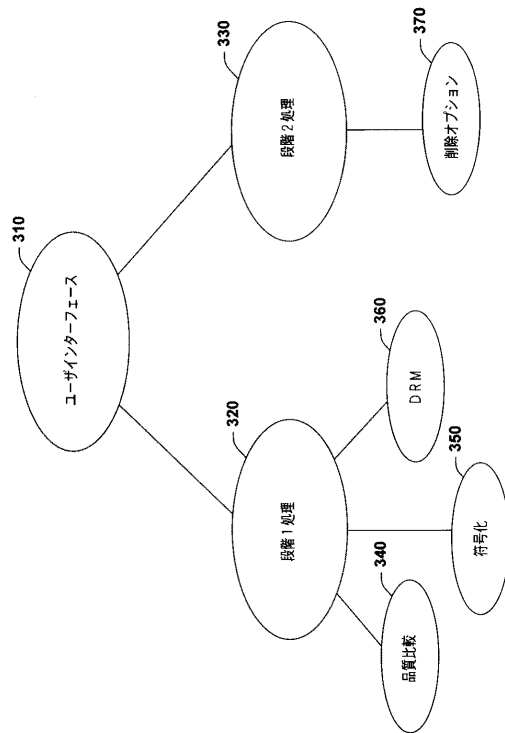
【図2】



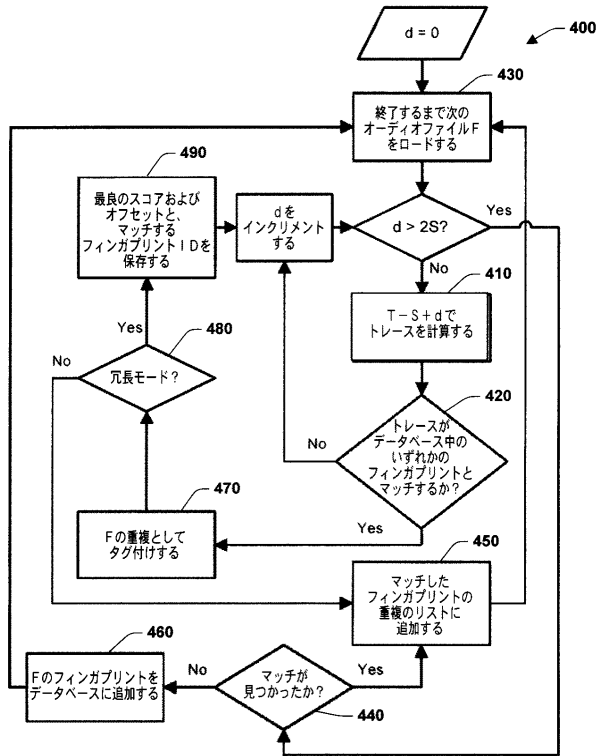
10

20

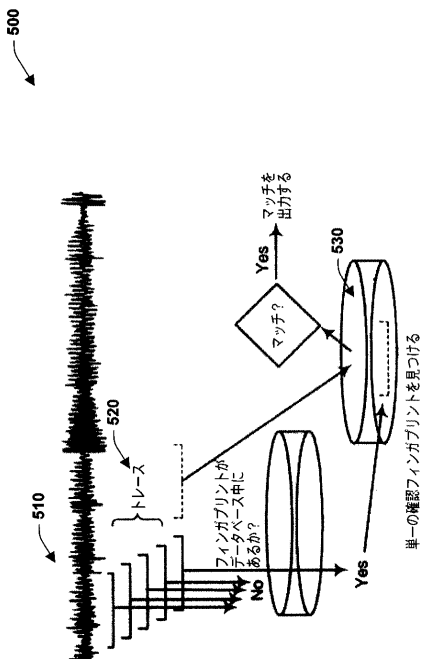
【図 3】



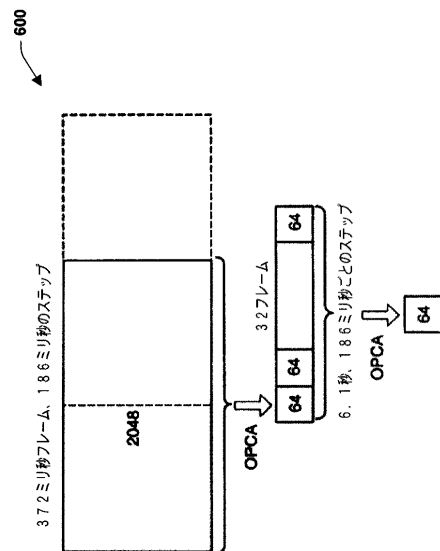
【図 4】



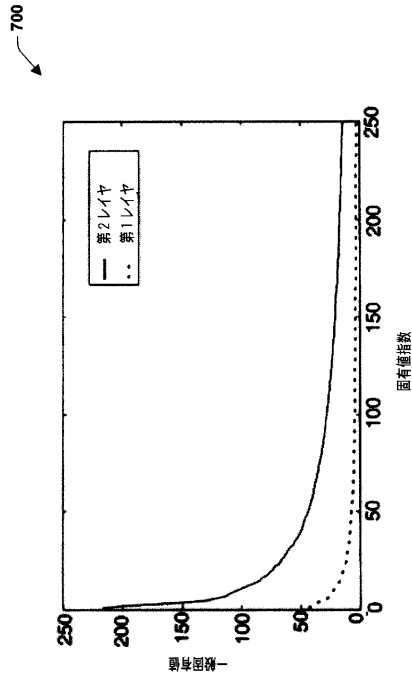
【図 5】



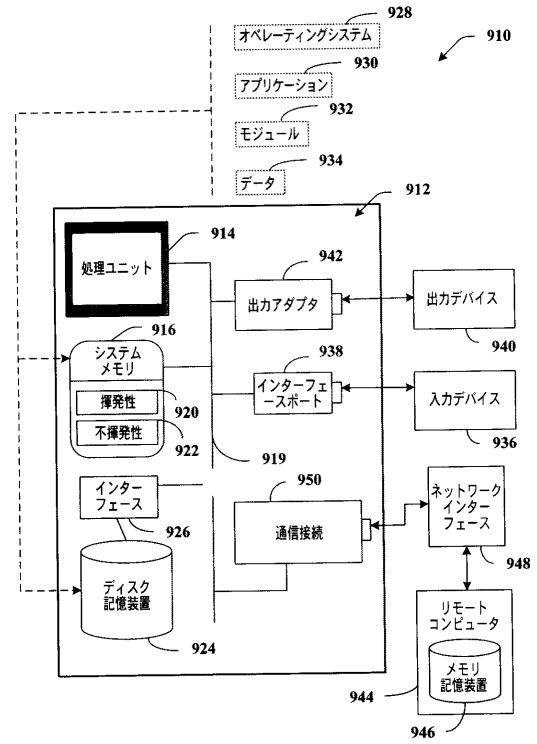
【図 6】



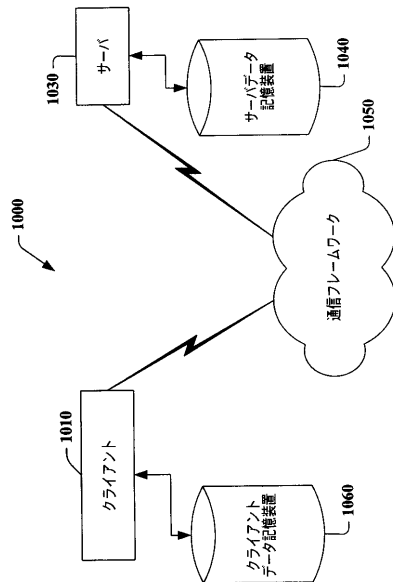
【図 7】



【図 8】



【図 9】



フロントページの続き

- (72)発明者 ダニエル プラスティナ
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション内
- (72)発明者 エリン エル・レンショウ
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション内
- (72)発明者 ジョン シー・プラット
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション内

審査官 柴垣 俊男

- (56)参考文献 特開2003-203467(JP,A)
特開2000-155759(JP,A)
特開平10-097269(JP,A)
特開2000-114692(JP,A)
特開平06-274193(JP,A)
特開2002-236496(JP,A)
特開2003-248494(JP,A)

- (58)調査した分野(Int.Cl., DB名)
- | | |
|------|-------|
| G10L | 19/00 |
| G10L | 11/00 |
| G06F | 12/00 |
| G11B | 27/00 |