(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6859499号
(P6859499)
(45)発行日　令和3年4月14日(2021.4.14)　(24)登録日　令和3年3月30日(2021.3.30)

(51)Int.Cl.　F I
G10L 25/84 (2013.01)　G10L 25/84

請求項の数 17　(全 19 頁)

(21)出願番号　特願2019-520035(P2019-520035)
(86)(22)出願日　平成29年9月26日(2017.9.26)
(65)公表番号　特表2019-535039(P2019-535039A)
(43)公表日　令和1年12月5日(2019.12.5)
(86)国際出願番号　PCT/CN2017/103489
(87)国際公開番号　WO2018/068636
(87)国際公開日　平成30年4月19日(2018.4.19)
審査請求日　令和1年6月12日(2019.6.12)
(31)優先権主張番号　201610890946.9
(32)優先日　平成28年10月12日(2016.10.12)
(33)優先権主張国・地域又は機関
中国(CN)

早期審査対象出願

(73)特許権者　520015461
アドバンスド　ニュー　テクノロジーズ　カンパニー　リミテッド
英国領ケイマン諸島　グランド　ケイマン　ケーワイ１－９００８　ジョージ　タウン　ホスピタル　ロード　２７　ケイマン　コーポレート　センター
(74)代理人　100188558
弁理士　飯田　雅人
(74)代理人　100205785
弁理士　▲高▼橋　史生

最終頁に続く

(54)【発明の名称】音声信号検出方法及び装置

(57)【特許請求の範囲】
【請求項１】
コンピュータにより実施される方法であって、
ユーザ端末により、オーディオ信号を取得するステップと；
所定の音声信号のサンプリングレートと前記所定の音声信号の周波数との比率を特定するステップと；
前記ユーザ端末により、前記オーディオ信号を、前記比率で示される数のサンプルを含む、最大数量の短時間エネルギーフレームに分割するステップと；
前記ユーザ端末により、各短時間エネルギーフレームのエネルギーを特定するステップと；
前記ユーザ端末により、各短時間エネルギーフレームの前記エネルギーに基づいて、前記オーディオ信号が音声信号を含むかどうかを特定するステップと；を備える、
コンピュータにより実施される方法。
【請求項２】
前記オーディオ信号は、前記サンプリングレートで収集され、パルス符号変調（ＰＣＭ）方式である、
請求項１に記載の方法。
【請求項３】
前記取得したオーディオ信号が、非ＰＣＭ方式であり、
前記オーディオ信号を分割する前に、

前記オーディオ信号をパルス符号変調（ＰＣＭ）方式に変換するステップと；
前記オーディオ信号の前記サンプリングレートを識別するステップと；を更に備える、
請求項１に記載の方法。
【請求項４】
前記各短時間エネルギーフレームのエネルギーは、各短時間エネルギーフレームの各サンプリングポイントに関連付けられたエネルギーの合計であり、前記各サンプリングポイントに関連付けられたエネルギーは、前記短時間エネルギーフレームのサンプリングポイントに対応する前記オーディオ信号の振幅に基づいて特定される、
請求項１に記載の方法。
【請求項５】
前記オーディオ信号が音声信号を含むかどうかを特定するステップは、
複数の高エネルギーフレームを特定するステップであって、前記複数の高エネルギーフレームの各高エネルギーフレームは、エネルギーが所定の閾値よりも大きい短時間エネルギーフレームである、ステップと；
前記オーディオ信号に含まれる前記短時間エネルギーフレームの量に対する前記複数の高エネルギーフレームの量の比によって表される高エネルギーフレーム比率を特定するステップと；
前記高エネルギーフレーム比率が所定の値より大きいかどうかを特定するステップと；
前記高エネルギーフレーム比率が前記所定の値より大きいと特定された場合に、
前記オーディオ信号には音声信号が含まれていると特定するステップ；又は、
前記高エネルギーフレーム比率が前記所定の値より大きくないと特定された場合に、
前記オーディオ信号には音声信号が含まれていないと特定するステップ；を備える、
請求項１に記載の方法。
【請求項６】
前記高エネルギーフレーム比率が所定の値よりも大きいと特定され、更に、
前記オーディオ信号に含まれる前記短時間エネルギーフレームから、所定数の連続する短時間エネルギーフレームがあるかどうかを特定するステップであって、前記所定数の連続する短時間エネルギーフレームのそれぞれは、前記所定の閾値よりも大きいエネルギーを有する、ステップと；
肯定の場合に、前記オーディオ信号には音声信号が含まれていると特定するステップ；又は、
肯定でない場合に、前記オーディオ信号には音声信号が含まれていないと特定するステップ；を備える、
請求項５に記載の方法。
【請求項７】
所定の操作を実行するためにコンピュータシステムによって実行可能な１又は複数の命令を格納する、非一時的なコンピュータ読取可能媒体であって、前記操作は、
ユーザ端末により、オーディオ信号を取得するステップと；
所定の音声信号のサンプリングレートと前記所定の音声信号の周波数との比率を特定するステップと；
前記ユーザ端末により、前記オーディオ信号を、前記比率で示される数のサンプルを含む、最大数量の短時間エネルギーフレームに分割するステップと；
前記ユーザ端末により、各短時間エネルギーフレームのエネルギーを特定するステップと；
前記ユーザ端末により、各短時間エネルギーフレームのエネルギーに基づいて、前記オーディオ信号が音声信号を含むかどうかを特定するステップと；を備える、
非一時的なコンピュータ読取可能媒体。
【請求項８】
前記オーディオ信号は、前記サンプリングレートで収集され、パルス符号変調（ＰＣＭ

）方式である、
請求項７に記載の非一時的なコンピュータ読取可能媒体。
【請求項９】
前記取得したオーディオ信号が、非ＰＣＭ方式であり、
前記オーディオ信号を分割する前に、
前記オーディオ信号をパルス符号変調（ＰＣＭ）方式に変換するステップと；
前記オーディオ信号の前記サンプリングレートを識別するステップと；を更に備える
、
請求項７に記載の非一時的なコンピュータ読取可能媒体。
【請求項１０】
前記各短時間エネルギーフレームのエネルギーは、各短時間エネルギーフレームの各サンプリングポイントに関連付けられたエネルギーの合計であり、前記各サンプリングポイントに関連付けられたエネルギーは、前記短時間エネルギーフレームのサンプリングポイントに対応する前記オーディオ信号の振幅に基づいて特定される、
請求項７に記載の非一時的なコンピュータ読取可能媒体。
【請求項１１】
前記オーディオ信号が音声信号を含むかどうかを特定するステップは、
複数の高エネルギーフレームを特定するステップであって、前記複数の高エネルギーフレームの各高エネルギーフレームは、エネルギーが所定の閾値よりも大きい短時間エネルギーフレームである、ステップと；
前記オーディオ信号に含まれる前記短時間エネルギーフレームの量に対する前記複数の高エネルギーフレームの量の比によって表される高エネルギーフレーム比率を特定するステップと；
前記高エネルギーフレーム比率が所定の値より大きいかどうかを特定するステップと；
前記高エネルギーフレーム比率が前記所定の値より大きいと特定された場合に、
前記オーディオ信号には音声信号が含まれていると特定するステップ；又は、
前記高エネルギーフレーム比率が前記所定の値より大きくないと特定された場合に、
前記オーディオ信号には音声信号が含まれていないと特定するステップ；を備える、
請求項７に記載の非一時的なコンピュータ読取可能媒体。
【請求項１２】
前記高エネルギーフレーム比率が所定の値よりも大きいと特定され、更に、
前記オーディオ信号に含まれる前記短時間エネルギーフレームから、所定数の連続する短時間エネルギーフレームがあるかどうかを特定するステップであって、前記所定数の連続する短時間エネルギーフレームのそれぞれは、前記所定の閾値よりも大きいエネルギーを有する、ステップと；
肯定の場合に、前記オーディオ信号には音声信号が含まれていると特定するステップ；又は、
肯定でない場合に、前記オーディオ信号には音声信号が含まれていないと特定するステップ；を備える、
請求項１１に記載の非一時的なコンピュータ読取可能媒体。
【請求項１３】
コンピュータにより実施されるシステムであって、
１又は複数のコンピュータと；
前記１又は複数のコンピュータと相互運用可能に接続され、前記１又は複数のコンピュータによって実行されると１又は複数の操作を実行する１又は複数の命令を格納する有形の非一時的な機械読取可能媒体を備えた、１又は複数のコンピューターメモリデバイスであって、前記１又は複数の操作は、
ユーザ端末により、オーディオ信号を取得するステップと；
所定の音声信号のサンプリングレートと前記所定の音声信号の周波数との比率を特定するステップと；

前記ユーザ端末により、前記オーディオ信号を、前記比率で示される数のサンプルを含む、最大数量の短時間エネルギーフレームに分割するステップと；
前記ユーザ端末により、各短時間エネルギーフレームのエネルギーを特定するステップと；
前記ユーザ端末により、各短時間エネルギーフレームのエネルギーに基づいて、前記オーディオ信号が音声信号を含むかどうかを特定するステップと；を備える、前記１又は複数のコンピューターメモリデバイスと；を備える、
コンピュータにより実施されるシステム。
【請求項１４】
前記オーディオ信号は、前記サンプリングレートで収集され、パルス符号変調（ＰＣＭ）方式である、
請求項１３に記載のコンピュータにより実施されるシステム。
【請求項１５】
前記取得したオーディオ信号が、非ＰＣＭ方式であり、
前記オーディオ信号を分割する前に、
前記オーディオ信号をパルス符号変調（ＰＣＭ）方式に変換するステップと；
前記オーディオ信号の前記サンプリングレートを識別するステップと；を更に備える、
請求項１３に記載のコンピュータにより実施されるシステム。
【請求項１６】
前記各短時間エネルギーフレームのエネルギーは、各短時間エネルギーフレームの各サンプリングポイントに関連付けられたエネルギーの合計であり、前記各サンプリングポイントに関連付けられたエネルギーは、前記短時間エネルギーフレームのサンプリングポイントに対応する前記オーディオ信号の振幅に基づいて特定される、
請求項１３に記載のコンピュータにより実施されるシステム。
【請求項１７】
前記オーディオ信号が音声信号を含むかどうかを特定するステップは、
複数の高エネルギーフレームを特定するステップであって、前記複数の高エネルギーフレームの各高エネルギーフレームは、エネルギーが所定の閾値よりも大きい短時間エネルギーフレームである、ステップと；
前記オーディオ信号に含まれる前記短時間エネルギーフレームの量に対する前記複数の高エネルギーフレームの量の比によって表される高エネルギーフレーム比率を特定するステップと；
前記高エネルギーフレーム比率が所定の値より大きいかどうかを特定するステップと；
前記高エネルギーフレーム比率が前記所定の値より大きいと特定された場合に、
前記オーディオ信号には音声信号が含まれていると特定するステップ；又は、
前記高エネルギーフレーム比率が前記所定の値より大きくないと特定された場合に、
前記オーディオ信号には音声信号が含まれていないと特定するステップ；を備える、
請求項１３に記載のコンピュータにより実施されるシステム。
【発明の詳細な説明】
【技術分野】
【０００１】
本願はコンピュータ技術の分野に関し、特に、音声信号検出方法及び装置に関する。
【背景技術】
【０００２】
人々は実生活の中でスマートデバイス（例えば、スマートフォンやタブレットコンピュータ）を使って音声メッセージを送信することが多い。しかし、スマートデバイスを使って音声メッセージを送信する場合、通常は、音声メッセージを送信する前にスマートデバイスのスクリーン上の開始ボタン又は終了ボタンをタップする必要があり、これらのタップ操作はユーザにとって非常に不便である。

【０００３】
ユーザがボタンをタップすることなく音声メッセージの送信を終えるには、スマートデバイスが連続的に、又は、所定の周期に基づいて録音を実行し、取得されたオーディオ信号（ａｕｄｉｏ　ｓｉｇｎａｌ）が音声信号（ｖｏｉｃｅ　ｓｉｇｎａｌ）を含むかどうか特定する必要がある。取得されたオーディオ信号が音声信号を含む場合、スマートデバイスは音声信号を抽出してから、音声信号を処理して送信する。そのようにして、スマートデバイスは音声メッセージの送信を終える。
【０００４】
既存の技術では、取得されたオーディオ信号が音声信号を含むかどうかを検出するために、通常は、二重閾値法、自己相関最大値に基づく検出法、及びウェーブレット変換に基づく検出法などの音声信号検出法が用いられる。しかし、これらの方法では、通常、フーリエ変換のような複雑な計算を用いてオーディオ情報の周波数特性を求め、更にその周波数特性に基づいてオーディオ情報が音声信号を含むかどうか特定する。したがって、より多くのバッファデータを計算する必要があり、メモリ使用量が比較的多くなり、比較的多くの計算が必要であり、処理速度は比較的遅く、消費電力も比較的大きくなる。
【発明の概要】
【０００５】
本願の実施は音声信号検出方法及び装置を提供し、既存の技術における音声信号検出方法では処理速度が比較的低く、リソース消費が比較的高いという問題を軽減する。
【０００６】
以下の技術的解決策が本願の実施で用いられる。
【０００７】
音声信号検出方法が提供され、この方法は：オーディオ信号を取得するステップと；所定の音声信号の周波数に基づいて、前記オーディオ信号を複数の短時間エネルギーフレームに分割するステップと；各短時間エネルギーフレームのエネルギーを特定するステップと；各短時間エネルギーフレームの前記エネルギーに基づいて、前記オーディオ信号が音声信号を含んでいるかどうかを検出するステップと；を含む。
【０００８】
音声信号検出装置が提供され、この装置は：オーディオ信号を取得するよう構成された取得モジュールと；所定の音声信号の周波数に基づいて、前記オーディオ信号を複数の短時間エネルギーフレームに分割するよう構成された分割モジュールと；各短時間エネルギーフレームのエネルギーを特定するよう構成された特定モジュールと；各短時間エネルギーフレームの前記エネルギーに基づいて、前記オーディオ信号は音声信号を含んでいるかどうかを検出するよう構成された検出モジュールと；を含む。
【０００９】
本願の実施において用いられる先に述べた技術的解決策の少なくとも１つは、以下の有益な効果を奏する。
【００１０】
既存の技術では、フーリエ変換のような複雑な計算を通して、オーディオ信号が音声信号を含むかどうか特定される。対照的に、本願の実施で用いられる音声信号検出方法では、フーリエ変換のような複雑な計算を行う必要はない。取得されたオーディオ信号は、所定の音声信号の周波数に基づいて複数の短時間エネルギーフレームに分割され、各短時間エネルギーフレームのエネルギーが更に特定され、そして、各短時間エネルギーフレームのエネルギーに基づいて、取得されたオーディオ信号が音声信号を含むかどうかを検出できる。したがって、本願の実施で提供される音声信号検出方法においては、既存の技術における音声信号検出方法では処理速度が比較的低くリソース消費が比較的高い、という問題を軽減できる。
【図面の簡単な説明】
【００１１】
本明細書で述べる添付図面は本願の更なる理解を提供し、本願の一部を構成するもので

ある。本願の例示の実施とその記述は本願を説明するものであり、本願に制限を設けるものではない。添付図面について以下のとおり説明する。

【００１２】

【図１】図１は、本願の実施に係る音声信号検出方法を示すフローチャートである。

【００１３】

【図２】図２は、本願の実施に係る別の音声信号検出方法を示すフローチャートである。

【００１４】

【図３】図３は、本願の実施に係る所定の持続時間の音声信号を示す表示図である。

【００１５】

【図４】図４は、本願の実施に係る音声信号検出装置の構造を示す概略図である。

【発明を実施するための形態】

【００１６】

本願の目的、技術的解決策及び利点を明瞭にするために、以下では、本願の具体的な実施及び添付図面を参照しながら本願の技術的解決策を明確且つ包括的に記述する。記述するこれらの実施は本願の実施の全てではなく、むしろそのいくつかに過ぎないことは言うまでもない。創造的な努力なく本願の実施に基づいて当業者により得られるその他の全ての実施は、本願の保護範囲に含まれる。

【００１７】

本願の実施で提供される技術的解決策を、添付の図面を参照して、以下詳細に説明する。

【００１８】

既存の技術の音声信号検出方法における比較的低い処理速度及び比較的高いリソース消費という問題を軽減するために、本願の実施は音声信号検出方法を提供する。

【００１９】

本方法を実行する主体は、携帯電話、タブレットコンピュータ、又はパーソナルコンピュータ（Ｐｅｒｓｏｎａｌ　Ｃｏｍｐｕｔｅｒ、ＰＣ）などのユーザ端末であってもよいが、これらに限定されず、これらユーザ端末上で作動するアプリケーション（ＡＰＰ：以後「アプリ」とする）であっても、サーバなどのデバイスであってもよい。

【００２０】

説明を容易にするために、本方法を実行する主体がアプリである実施例を用いて、本方法の実施を、以下説明する。言うまでもなく本方法はアプリによって実行されるが、これは説明のための例にすぎず、本方法に対する限定として解釈されるべきではない。

【００２１】

図１は、本方法の手順の概略図である。本方法は以下のステップを含む。

【００２２】

ステップ１０１：オーディオ信号を取得する。

【００２３】

オーディオ信号は、オーディオ収集デバイスを用いてアプリにより収集されたオーディオ信号であっても、アプリにより受信されたオーディオ信号であってもよく、例えば、別のアプリ又はデバイスによって送信されたオーディオ信号であってもよい。実施については本願で限定されない。オーディオ信号を得た後、アプリはオーディオ信号をローカルに格納できる。

【００２４】

本願は、オーディオ信号に対応するサンプリングレート、持続時間、方式（フォーマット）、サウンドチャンネルなどに対して制限しない。

【００２５】

本願のこの実施において提供される音声信号検出方法では、アプリがオーディオ信号を取得することができ、取得されたオーディオ信号に対して音声信号検出を実行できるのであれば、アプリは、チャットアプリや決済アプリなどの任意のタイプのアプリであってもよい。

【００２６】
ステップ１０２：所定の音声信号の周波数に基づいて、オーディオ信号を複数の短時間エネルギーフレームに分割する。
【００２７】
短時間エネルギーフレームは、実際には、ステップ１０１で取得されたオーディオ信号の一部である。
【００２８】
具体的には、所定の音声信号の周波数に基づいて所定の音声信号の周期を特定でき、この特定された周期に基づいて、ステップ１０１で取得されたオーディオ信号が、対応する持続時間が周期である複数の短時間エネルギーフレームに分割される。例えば、ステップ１０１で取得されたオーディオ信号の持続時間に基づいて、所定の音声信号の周期が０．０１秒であると仮定すると、オーディオ信号を、持続時間が０．０１秒であるいくつかの短時間エネルギーフレームに分割できる。注記すると、ステップ１０１で取得されたオーディオ信号を分割する場合、代替として、オーディオ信号を、実際の状態と所定の音声信号の周波数とに基づいて、少なくとも２つの短時間エネルギーフレームに分割してもよい。後に続く説明を分かり易くするために、オーディオ信号が複数の短時間エネルギーフレームに分割される例を本願のこの実施で用いて、以下説明する。
【００２９】
更に、ステップ１０１でアプリがオーディオ収集デバイスを用いてオーディオ信号を収集する場合、一般に、オーディオ信号を収集することは、ある特定のサンプリングレートで、実際にはデジタル信号を形成するためのアナログ信号であるオーディオ信号、すなわちパルスコード変調（Ｐｕｌｓｅ　Ｃｏｄｅ　Ｍｏｄｕｌａｔｉｏｎ、ＰＣＭ）方式のオーディオ信号を収集することであるため、オーディオ信号は、オーディオ信号のサンプリングレートと所定の音声信号の周波数とに基づいて、更に複数の短時間エネルギーフレームに分割できる。
【００３０】
具体的には、所定の音声信号の周波数に対するオーディオ信号のサンプリングレートの比率ｍを特定でき、次いで、収集されたデジタルオーディオ信号内の各ｍ個のサンプリング点は、比率ｍに基づいて１つの短時間エネルギーフレームにグループ化される。比率ｍが正の整数である場合、オーディオ信号を、ｍに基づいて最大数量の短時間エネルギーフレームに分割でき、ｍが正の整数ではない場合、オーディオ信号を、正の整数に丸められる（端数処理する）ｍに基づいて最大数量の短時間エネルギーフレームに分割できる。注記すると、ステップ１０１で取得されたオーディオ信号に含まれるサンプリング点の数量がｍの整数倍でない場合、オーディオ信号が最大数量の短時間エネルギーフレームに分割された後に、残りのサンプリング点を破棄してもよい、又は、その代わりに、残りのサンプリング点を後続の処理のための短時間エネルギーフレームとして用いてもよい。所定の音声信号の周期における、ステップ１０１で取得されたオーディオ信号に含まれるサンプリング点の数量を表すためにＭを用いる。
【００３１】
例えば、所定の音声信号の周波数が８２Ｈｚの場合、ステップ１０１で取得されたオーディオ信号の持続時間は１秒であり、サンプリングレートは１６０００Ｈｚであり、比率ｍ＝１６０００／８２＝１９５．１である。ここで、ｍは正の整数ではないので、１９５．１は正の整数１９５に丸められる。オーディオ信号の持続時間とサンプリングレートとに基づき、オーディオ信号に含まれるサンプリング点の数量は１６０００であると特定できる。オーディオ信号に含まれるサンプリング点の数量は１９５の整数倍ではないので、オーディオ信号が８２の短時間エネルギーフレームに分割された後、残りの１０のサンプリング点は破棄してもよい。各短時間エネルギーフレームに含まれるサンプリング点の数量は１９５である。
【００３２】
ステップ１０１で取得されたオーディオ信号が別のアプリ又はデバイスによって送信さ

れた受信オーディオ信号である場合、オーディオ信号は、前述の方法のうちのいずれか１つを用いて複数の短時間エネルギーフレームに分割できる。注記すると、オーディオ信号の方式がＰＣＭ方式ではない場合がある。前述の方法でオーディオ信号のサンプリングレートと所定の音声信号の周波数とに基づいて分割することにより短時間エネルギーフレームが得られる場合、受信オーディオ信号をＰＣＭ方式のオーディオ信号に変換する必要がある。更に、オーディオ信号を受信したときには、オーディオ信号のサンプリングレートを特定する必要がある。オーディオ信号のサンプリングレートを識別する方法は、既存の技術における識別方法であってよい。ここでは説明を簡単にするために詳細は省略する。

【００３３】

ステップ１０３：各短時間エネルギーフレームのエネルギーを特定する。

【００３４】

本願のこの実施では、ＰＣＭ方式のオーディオ信号が、前述の方法で、同じくＰＣＭ方式のいくつかの短時間エネルギーフレームに分割されるとき、短時間エネルギーフレームのエネルギーは、短時間エネルギーフレーム内の各サンプリング点に対応するオーディオ信号の振幅に基づいて特定できる。具体的には、短時間エネルギーフレーム内の各サンプリング点に対応するオーディオ信号の振幅に基づいて各サンプリング点のエネルギーを特定し、次いで、サンプリング点のエネルギーを合計する。最終的に取得されたエネルギーの合計は、短時間エネルギーフレームのエネルギーとして用いられる。

【００３５】

例えば、短時間エネルギーフレームのエネルギーは以下の式を用いて特定できる。

【数１】

$$\text{エネルギー}=\sum_{i}^{i+n}(A_i[t])^2$$

式中、ｉはオーディオ信号のｉ番目のサンプリング点を表し、ｎは短時間エネルギーフレームに含まれるサンプリング点の数量であり、Ａ$_{i}$［ｔ］はｉ番目のサンプリング点に対応するオーディオ信号の振幅であり、短時間エネルギーフレームの振幅の値の範囲は、－３２７６８から３２７６７である。

【００３６】

更に、本願のこの実施においては、計算を簡素化し、リソースを節約するために、振幅を３２７６８で除した値を更に短時間エネルギーフレームの正規化振幅として使用できる。振幅は、オーディオ信号が収集されたときに得られる。短時間エネルギーフレームの正規化振幅の値の範囲は、－１から１である。

【００３７】

短時間エネルギーフレームがＰＣＭ方式ではない場合、振幅計算関数を各瞬間における短時間エネルギーフレームの振幅に基づいて特定でき、積分はその関数の２乗に対して実行される。そして最終的に得られる積分結果は短時間エネルギーフレームのエネルギーである。

【００３８】

ステップ１０４：各短時間エネルギーフレームのエネルギーに基づいて、オーディオ信号に音声信号が含まれているかどうかを検出する。

【００３９】

具体的には、オーディオ信号に音声信号が含まれているかどうかを特定するために、次の２つの方法を用いることができる。

【００４０】

方法１：全ての短時間エネルギーフレームの総量に対する、エネルギーが所定の閾値よりも大きい短時間エネルギーフレームの量の比率（以下、高エネルギーフレーム比率と呼ぶ）が特定され、特定された高エネルギーフレーム比率は所定の比率より大きいかどうか特定される。それが肯定であれば、オーディオ信号は音声信号を含むと特定され、そうで

なければ、オーディオ信号は音声信号を含まないと特定される。
【００４１】
所定の閾値の値及び所定の比率の値は、実際の要求に基づいて設定できる。本願のこの実施において、所定の閾値は２に設定でき、所定の比率は２０％に設定できる。高エネルギーフレーム比率が２０％より大きい場合、オーディオ信号は音声信号を含むと特定され、そうでなければ、オーディオ信号は音声信号を含まないと特定される。
【００４２】
本願のこの実施では、人が話すとき、実生活の中では外部環境にいくらかのノイズがあり、このノイズのエネルギーは、一般に、人の声よりも低いので、方法１を用いてオーディオ信号が音声信号を含むかどうか特定できる。この場合、エネルギーが所定の閾値よりも大きい短時間エネルギーフレームをオーディオ信号セグメントが含み、これらの短時間エネルギーフレームがオーディオ信号セグメントの特定の比率を構成する場合、オーディオ信号は、音声信号を含むと特定できる。
【００４３】
方法２：最終的な検出結果をより正確にするために、方法１を用いて、高エネルギーフレーム比率を特定し、特定された高エネルギーフレーム比率が所定の比率より大きいかどうかを特定できる。否定であれば、オーディオ信号は音声信号を含まないと特定される。肯定であれば、エネルギーが所定の閾値より大きい短時間エネルギーフレーム内に少なくともＮ個の連続する短時間エネルギーフレームがある場合、オーディオ信号は音声信号を含むと特定され、エネルギーが所定の閾値より大きい短時間エネルギーフレーム内に少なくともＮ個の連続する短時間エネルギーフレームがない場合、オーディオ信号は音声信号を含まないと特定される。Ｎは任意の正の整数であってよい。本願のこの実施では、Ｎを１０に設定できる。
【００４４】
具体的には、方法１に基づいて、方法２では、オーディオ信号が音声信号を含むかどうか特定するために以下の要件が追加される。すなわち、エネルギーが所定の閾値よりも大きい短時間エネルギーフレーム内に、少なくともＮ個の連続する短時間エネルギーフレームがあるかどうかが特定される。そのようにして、ノイズを効果的に減らすことができる。実生活では、ノイズは人の声よりもエネルギーが低く、オーディオ信号はランダムである。方法２では、オーディオ信号が過度のノイズを含む場合を効果的に排除でき、外部環境におけるノイズの影響が低減され、ノイズリダクション機能を果たす。
【００４５】
注記すると、本願のこの実施において提供される音声信号検出方法は、モノラルオーディオ信号、バイノーラルオーディオ信号、マルチチャンネルオーディオ信号等の検出に適用できる。１つのサウンドチャネルを用いて収集されたオーディオ信号はモノラルオーディオ信号であり、２つのサウンドチャネルを用いて収集されたオーディオ信号はバイノーラルオーディオ信号であり、複数のサウンドチャンネルを用いて収集されたオーディオ信号はマルチチャンネルオーディオ信号である。
【００４６】
図１に示す方法でバイノーラルオーディオ信号及びマルチチャンネルオーディオ信号を検出する場合、ステップ１０１乃至ステップ１０４で説明した操作を実行することにより、各チャンネルの取得されたオーディオ信号を検出でき、最後に、各チャンネルのオーディオ信号の検出結果に基づいて、取得されたオーディオ信号が音声信号を含むかどうかを特定する。
【００４７】
具体的には、ステップ１０１で取得されたオーディオ信号がモノラルオーディオ信号である場合、そのオーディオ信号に対してステップ１０１乃至ステップ１０４で説明した操作を、直接、実行でき、検出結果が最終的な検出結果として用いられる。
【００４８】
ステップ１０１で取得されたオーディオ信号がモノラルオーディオ信号ではなくバイノ

ーラルオーディオ信号又はマルチチャンネルオーディオ信号である場合、ステップ１０１乃至ステップ１０４で説明した操作を実行することによって各チャンネルの音声信号を処理できる。各チャンネルのオーディオ信号が音声信号を含まないことが検出された場合、ステップ１０１で取得されたオーディオ信号は音声信号を含まないと特定される。少なくとも１つのチャンネルのオーディオ信号が音声信号を含むことが検出された場合、ステップ１０１で取得されたオーディオ信号は音声信号を含むと特定される。

【００４９】

更に、ステップ１０２で説明した所定の音声信号の周波数は、任意の音声の周波数とすることができる。実施は本願において限定されない。実際には、現実のケースに基づいて、ステップ１０１で取得された異なるオーディオ信号に対して異なる周波数の所定の音声信号を設定できる。注記すると、所定の音声信号の周波数は、分割を通して最終的に得られる短時間エネルギーフレームが以下の要求、すなわち短時間エネルギーフレームに対応する持続時間は、ステップ１０１で取得されたオーディオ信号に対応する周期以上であるとの要求、を満たすという条件で、最高音（ソプラノ）の音声周波数又は最低音（バス）の音声周波数などの任意の音声信号の周波数であってよい。より良好な検出効果を確保して、できるだけ多くのリソースを節約し、処理速度を向上させるために、本願のこの実施では、所定の音声信号の周波数を、人の最低音声周波数、すなわち８２Ｈｚ、に設定できる。周期は周波数の逆数であるので、所定の音声信号の周波数が人の最低音声周波数である場合、所定の音声信号の周期は人の最高音声周期である。したがって、ステップ１０１で取得されたオーディオ信号の周期にかかわらず、短時間エネルギーフレームに対応する持続時間は、先に取得されたオーディオ信号の周期以上である。

【００５０】

注記すると、本願のこの実施では、人の音声の特徴に基づいてオーディオ信号が音声信号を含むかどうか特定するためにここで論じた検出方法が用いられるので、短時間エネルギーフレームに対応する持続時間は、ステップ１０１で取得されたオーディオ信号の周期以上であることが要求される。ノイズと比較して、人の音声はより高いエネルギーを持ち、より安定しており、そして連続的である。短時間エネルギーフレームに対応する持続時間がステップ１０１で取得されたオーディオ信号の周期より短い場合、短時間エネルギーフレームに対応する波形は全周期（ｃｏｍｐｌｅｔｉｏｎ　ｐｅｒｉｏｄ）の波形を含まず、短時間エネルギーフレームの期間は比較的短い。この場合、高エネルギーフレーム比率が所定比率よりも大きく、エネルギーが所定の閾値よりも大きい短時間エネルギーフレーム内に少なくともＮ個の連続する短時間エネルギーフレームがある場合でも、それはオーディオ信号が音響信号（ｓｏｕｎｄ　ｓｉｇｎａｌ）を含むことを単に示すだけであり、この音響信号が音声信号であることを示すものではない。したがって、本願のこの実施では、ステップ１０１で取得されたオーディオ信号の持続時間は、人の最高音声周期よりも長くなければならない。

【００５１】

更に、本願のこの実施において提供される音声信号検出方法は、特に、ユーザのタップ操作なくチャットアプリを用いることによって音声メッセージの送信を終えることができるアプリケーションシナリオに適用可能である。シナリオに基づいて、本願のこの実施において提供される音声信号検出方法を、以下、詳細に説明する。このシナリオでは、図２は、本方法の手順の概略図である。本方法は以下のステップを含む。

【００５２】

ステップ２０１：リアルタイムでオーディオ信号を収集する。

【００５３】

ユーザは、アプリを起動した後に、タップ操作をせずにチャットアプリが音声メッセージの送信を終えることを期待する場合がある。この場合、アプリは、外部環境を連続的に録音してリアルタイムでオーディオ信号を収集し、ユーザの音声の抜けを減らす。更に、オーディオ信号を収集した後、アプリはオーディオ信号をリアルタイムでローカルに格納できる。ユーザがアプリを停止した後、アプリは録音を停止する。

【００５４】
ステップ２０２：リアルタイムで収集したオーディオ信号から所定の持続時間を持つオーディオ信号を切り取る。
【００５５】
アプリがオーディオ信号をリアルタイムで検出する代わりに録音を続けると、音声メッセージはリアルタイムで送信されない。したがって、アプリは、ステップ２０１で収集されたオーディオ信号から、所定の持続時間を持つオーディオ信号をリアルタイムに切り取り、所定の持続時間を持つオーディオ信号に対して後続の検出を実行できる。
【００５６】
所定の持続時間を持つ現在切り取られたオーディオ信号は、現在のオーディオ信号（ｃｕｒｒｅｎｔ　ａｕｄｉｏ　ｓｉｇｎａｌ）と呼ぶことができ、所定の持続時間を持つ最後に切り取られたオーディオ信号は、最後に取得されたオーディオ信号（ｌａｓｔ　ｏｂｔａｉｎｅｄ　ａｕｄｉｏ　ｓｉｇｎａｌ）と呼ぶことができる。
【００５７】
ステップ２０３：所定の音声信号の周波数に基づいて、所定の持続時間内のオーディオ信号を複数の短時間エネルギーフレームに分割する。
【００５８】
ステップ２０４：各短時間エネルギーフレームのエネルギーを特定する。
【００５９】
ステップ２０５：各短時間エネルギーフレームのエネルギーに基づいて、所定の持続時間内のオーディオ信号が音声信号を含むかどうかを検出する。
【００６０】
現在のオーディオ信号が音声信号を含むことが検出された場合、最後に取得されたオーディオ信号が音声信号を含むかどうかが特定される。最後に取得されたオーディオ信号が音声信号を含まないと特定されると、現在のオーディオ信号の開始点を音声信号の開始点として特定でき、最後に取得されたオーディオ信号が音声信号を含むと特定されると、現在のオーディオ信号の開始点は音声信号の開始点ではない。
【００６１】
現在のオーディオ信号が音声信号を含まないことが検出されると、最後に取得されたオーディオ信号が音声信号を含むかどうか特定される。最後に取得されたオーディオ信号が音声信号を含むと特定されると、最後に取得されたオーディオ信号の終了点は音声信号の終了点として特定でき、最後に取得されたオーディオ信号が音声信号を含まないと特定されると、現在のオーディオ信号の終了点も、最後に取得されたオーディオ信号の終了点も音声信号の終了点ではない。
【００６２】
例えば、図３に示すように、Ａ、Ｂ、Ｃ、Ｄは、所定の持続時間を持つ４つの隣接するオーディオ信号である。オーディオ信号ＡとＤとは音声信号を含まず、オーディオ信号ＢとＣとは音声信号を含む。この場合、オーディオ信号Ｂの開始点を音声信号の開始点と特定し、オーディオ信号Ｃの終了点を音声信号の終了点と特定できる。
【００６３】
時に、現在のオーディオ信号がユーザの文言の開始部分又は終了部分であり、そのオーディオ信号には少しの音声信号が含まれていることがある。この場合、アプリは、オーディオ信号が音声信号を含まない、と間違って特定する可能性がある。現在のオーディオ信号は音声信号を含むことが検出された後、間違った特定によるユーザの音声の抜けを減らすために、最後に取得されたオーディオ信号が音声信号を含むかどうか特定でき、最後に取得されたオーディオ信号は音声信号を含まないと特定された場合、最後に取得されたオーディオ信号の開始点を音声信号の開始点として特定できる。更に、現在のオーディオ信号が音声信号を含まないことが検出された後、最後に取得されたオーディオ信号が音声信号を含むかどうか特定でき、最後に取得されたオーディオ信号が音声信号を含むと特定されると、現在のオーディオ信号の終了点を音声信号の終了点として特定できる。前述の例

においては、オーディオ信号Ａの開始点を音声信号の開始点と特定し、オーディオ信号Ｄの終了点を音声信号の終了点として特定できる。

【００６４】

現在のオーディオ信号が音声信号を含むことを検出した後、アプリはオーディオ信号を音声識別装置へ送信でき、その結果、音声識別装置はオーディオ信号に対して音声処理を実行して音声結果を取得することができる。その後、音声識別装置はオーディオ信号を後続の処理装置へ送信し、最後に音声メッセージの形式でオーディオ信号を送信する。送信された音声メッセージ内のユーザの音声が完全な文章であることを保証するために、音声信号の特定された開始点と特定された終了点との間の全てのオーディオ信号を音声識別装置へ送信した後、アプリはオーディオ停止信号を音声識別装置へ送信してユーザが現在述べているこの文章が完了した旨を音声識別装置に通知でき、それにより、音声識別装置は全てのオーディオ信号を後続の処理装置へ送信する。最終的に、オーディオ信号は音声メッセージの形式で送信される。

【００６５】

更に、正確な特定を確実にするために、現在のオーディオ信号を得た後、所定の時間周期を持つ副信号を、最後に取得されたオーディオ信号から更に切り取ることが可能である。現在のオーディオ信号と切り取られた副信号とが連結されて、取得されたオーディオ信号（以下、連結オーディオ信号（ｃｏｎｃａｔｅｎａｔｅｄ　ａｕｄｉｏ　ｓｉｇｎａｌ）と呼ぶ）として機能する。更に、後続の音声信号検出は、連結オーディオ信号に対して実行される。

【００６６】

副信号は現在のオーディオ信号の前に連結できる。所定の時間周期は、最後に取得されたオーディオ信号のテール時間周期であってよく、時間周期に対応する持続時間は任意の持続時間であってよい。最終的な検出結果がより正確であることを保証するために、本願のこの実施では、所定の時間周期に対応する持続時間は、所定の比率と連結オーディオ信号に対応する持続時間との積以下である値に設定できる。

【００６７】

連結オーディオ信号が音声信号を含むことが検出されると、最後に取得された連結オーディオ信号が音声信号を含むかどうかを特定できる。最後に取得された連結オーディオ信号が音声信号を含まないと特定されると、連結オーディオ信号の開始点を音声信号の開始点として用いることができる。連結オーディオ信号が音声信号を含まないことが検出されると、最後に取得された連結オーディオ信号が音声信号を含むかどうかを特定できる。最後に取得された連結オーディオ信号が音声信号を含むと特定されると、連結オーディオ信号の終了点を音声信号の終了点として用いることができる。

【００６８】

本願のこの実施において、連続的な録音に加えて、アプリは周期的に録音を実行できる。実施は本願のこの実施において限定されない。

【００６９】

本願のこの実施で提供される音声信号検出方法は、音声信号検出装置を用いて更に実施できる。図４に、この装置の概略構造図を示す。音声信号検出装置は、主に以下のモジュール、すなわち、オーディオ信号を取得するよう構成された取得モジュール４１と；所定の音声信号の周波数に基づいてオーディオ信号を複数の短時間エネルギーフレームに分割するよう構成された分割モジュール４２と；各短時間エネルギーフレームのエネルギーを特定するよう構成された特定モジュール４３と；各短時間エネルギーフレームのエネルギーに基づいて、オーディオ信号が音声信号を含むかどうかを検出するよう構成された検出モジュール４４と；を含む。

【００７０】

実施において、取得モジュール４１は：現在のオーディオ信号を取得し；最後に取得されたオーディオ信号から所定の周期を持つ副信号を切り取り；そして、取得されたオーディオ信号として機能するように、現在のオーディオ信号と切り取られた副信号とを連結す

るよう構成される。
【００７１】
実施において、分割モジュール４２は：所定の音声信号の周波数に基づいて所定の音声信号の周期を特定し；そして、特定された周期に基づいて、オーディオ信号を、対応する持続時間がその周期である複数の短時間エネルギーフレームに分割するよう構成される。
【００７２】
実施において、検出モジュール４４は：エネルギーが所定の閾値よりも大きい短時間エネルギーフレームの量の、全ての短時間エネルギーフレームの総量に対する比率を特定し；比率が所定の比率より大きいかどうか特定し；肯定であればオーディオ信号は音声信号を含むと特定し；否定であればオーディオ信号は音声信号を含まないと特定する；よう構成される。
【００７３】
実施において、検出モジュール４４は、エネルギーが所定の閾値よりも大きい短時間エネルギーフレームの量の、全ての短時間エネルギーフレームの総量に対する比率を特定し；比率が所定の比率より大きいかどうか特定し；否定であればオーディオ信号は音声信号を含まない、と特定し；肯定であればエネルギーが所定の閾値より大きい短時間エネルギーフレーム内に少なくともＮ個の連続した短時間エネルギーフレームがあるとき、オーディオ信号は音声信号を含む、と特定し；エネルギーが所定の閾値よりも大きい短時間エネルギーフレーム内に少なくともＮ個の連続する短時間エネルギーフレームがないとき、オーディオ信号は音声信号を含まない、と特定するよう構成される。
【００７４】
既存の技術では、フーリエ変換のような複雑な計算を通して、オーディオ信号が音声信号を含むかどうかが特定される。対照的に、本願の実施で用いられる音声信号検出方法では、フーリエ変換のような複雑な計算を実行する必要はない。取得されたオーディオ信号は、所定の音声信号の周波数に基づいて複数の短時間エネルギーフレームに分割され、各短時間エネルギーフレームのエネルギーが更に特定され、そして、各短時間エネルギーフレームのエネルギーに基づいて、取得されたオーディオ信号が音声信号を含むかどうかを検出できる。したがって、本願の実施において提供される音声信号検出方法では、既存の技術における音声信号検出方法における、処理速度が比較的低く、リソース消費が比較的高いという問題を軽減できる。
【００７５】
本開示は、本願の実施に係る方法、デバイス（システム）、コンピュータプログラム製品のフローチャート及び／又はブロック図を参照して説明されている。フローチャート及び／又はブロック図内の各プロセス及び／又は各ブロック、並びにフローチャート及び／又はブロック図内のプロセス及び／又はブロックの組み合わせを実施するために、コンピュータプログラム命令を使用することができることを理解されたい。これらのコンピュータプログラム命令は、汎用コンピュータ、専用コンピュータ、組み込みプロセッサ、又はあらゆるその他のプログラマブルデータ処理デバイスに、マシンを生成するために提供されることができ、これにより、コンピュータ、又はあらゆるその他のプログラマブルデータ処理デバイスのプロセッサが、フローチャートの１つ以上のプロセスにおける、及び／又は、ブロック図の１つ以上のブロックにおける、特定の機能を実施するデバイスを生成できるようになる。
【００７６】
このコンピュータプログラム命令を、コンピュータ又はあらゆるその他のプログラマブルデータ処理デバイスにある方法で機能するように命令することができるコンピュータ読取可能なメモリに記憶して、これらのコンピュータ読取可能なメモリに記憶された命令が、命令装置を含むアーチファクトを作り出すようにすることができる。この命令装置は、フローチャート内の１つ以上のプロセスにおける、及び／又はブロック図内の１つ以上のブロックにおける特定の機能を実施する。
【００７７】

これらのコンピュータプログラム命令をコンピュータ又はその他のプログラマブルデータ処理デバイスにロードして、コンピュータ又はその他のプログラマブルデバイス上で一連の操作及びステップが実行されるようにし、コンピュータで実施される処理を生成することができる。これにより、コンピュータ又はその他のプログラマブルデバイス上で実行される命令が、フローチャート内の１つ以上のプロセス及び／又はブロック図内の１つ以上のブロックにおける特定の機能を実施するデバイスを提供することを可能とする。

【００７８】

典型的な構成では、計算デバイスは１つ以上の中央処理演算装置（ＣＰＵｓ）、１つ以上の入出力インターフェース、１つ以上のネットワークインターフェース、及び１つ以上のメモリを含む。

【００７９】

メモリは、揮発性メモリ、ランダムアクセスメモリ（ＲＡＭ）、不揮発性メモリ、及び／又はリードオンリーメモリ（ＲＯＭ）やフラッシュメモリ（ｆｌａｓｈ　ＲＡＭ）のようなコンピュータ読取可能な媒体を含んでよい。メモリはコンピュータ読取可能な媒体の一例である。

【００８０】

コンピュータ読取可能な媒体には、任意の方法又は技術を用いて情報を記憶できる、永続的、非永続的、移動可能な、及び移動不能な媒体が含まれる。この情報はコンピュータ読取可能な命令、データ構造、プログラムモジュール、又はその他のデータであってよい。コンピュータの記憶媒体の例として、相変化ランダムアクセスメモリ（ＰＲＡＭ）、スタティックランダムアクセスメモリ（ＳＲＡＭ）、ダイナミックランダムアクセスメモリ（ＤＲＡＭ）、別タイプのランダムアクセスメモリ、リードオンリーメモリ（ＲＯＭ）、電気的に消去可能でプログラム可能なＲＯＭ（ＥＥＰＲＯＭ）、フラッシュメモリ、又は別のメモリ技術、コンパクトディスクＲＯＭ（ＣＤ－ＲＯＭ）、デジタル多用途ディスク（ＤＶＤ）、又は別の光学記憶装置、カセット磁気テープ、磁気テープ／磁気ディスクストレージ、他の磁気的記憶装置、又は他の任意の非伝送媒体があるが、これに限定されない。このコンピュータの記憶媒体は、計算デバイスによってアクセスできる情報を記憶するよう構成することができる。本願の定義に基づき、コンピュータ読取可能な媒体は、変調されたデータ信号及び搬送波のような一時的な媒体（ｔｒａｎｓｉｔｏｒｙ　ｍｅｄｉａ）を含まない。

【００８１】

さらに、用語「含む」、「備える」、又はこれらのその他任意の応用形は、非限定的な包含を網羅するものであるため、一連の要素を含んだ工程、方法、商品、デバイスはこれらの要素を含むだけでなく、ここで明確に挙げていないその他の要素をも含む、あるいは、このような工程、方法、商品、デバイスに固有の要素をさらに含むことができる点に留意することが重要である。「（１つの）～を含む」との用語を付けて示された要素は、それ以上の制約がなければ、その要素を含んだ工程、方法、商品、デバイス内に別の同一の要素をさらに含むことを排除しない。

【００８２】

当業者は、本願の実施が方法、システム、又はコンピュータプログラム製品として提供できることを理解するはずである。そのため、本発明は、ハードウェアのみの実施、ソフトウェアのみの実施、又は、ソフトウェアとハードウェアとの組み合わせによる実施を用いることができる。さらに、本発明は、コンピュータで使用可能なプログラムコードを含んだ１台以上のコンピュータで使用可能な記憶媒体（ディスクメモリ、ＣＤ－ＲＯＭ、光学メモリ等を含むがこれに限定されない）上で実施されるコンピュータプログラム製品を使用できる。

【００８３】

上述のものは本願の一実施の形態であり、本願を限定することを意図するものではない。当業者は、本願に対して様々な修正及び変更を加えることができる。本願の主旨及び原理から逸脱せずに為されるあらゆる修正、均等物による置換、改善は、本願の特許請求の

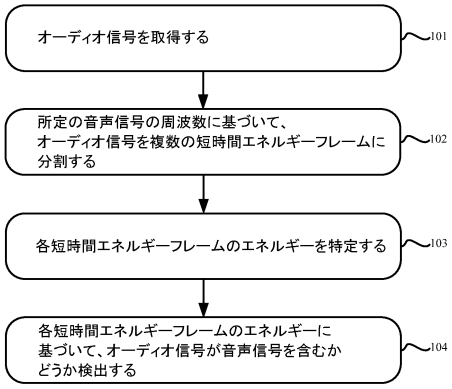
範囲に含まれるものである。
以下、本発明の実施の態様の例を列挙する。
［第１の局面］
音声信号検出方法であって：
オーディオ信号を取得するステップと；
所定の音声信号の周波数に基づいて、前記オーディオ信号を複数の短時間エネルギーフレームに分割するステップと；
各短時間エネルギーフレームのエネルギーを特定するステップと；
各短時間エネルギーフレームの前記エネルギーに基づいて、前記オーディオ信号が音声信号を備えているかどうかを検出するステップと；を備える、
音声信号検出方法。
［第２の局面］
オーディオ信号を取得する前記ステップは：
現在のオーディオ信号を取得するステップと；
最後に取得されたオーディオ信号から、所定の時間周期を持つ副信号を切り取るステップと；
前記取得されたオーディオ信号として機能するよう、前記現在のオーディオ信号と前記切り取られた副信号とを連結するステップと；を備える、
第１の局面に記載の方法。
［第３の局面］
所定の音声信号の周波数に基づいて、前記オーディオ信号を複数の短時間エネルギーフレームに分割する前記ステップは：
前記所定の音声信号の周波数に基づいて前記所定の音声信号の周期を特定するステップと；
前記特定された周期に基づいて、前記オーディオ信号を、対応する持続時間が前記周期である複数の短時間エネルギーフレームに分割するステップと；を備える、
第１の局面に記載の方法。
［第４の局面］
各短時間エネルギーフレームの前記エネルギーに基づいて、前記オーディオ信号が音声信号を備えているかどうかを検出する前記ステップは：
エネルギーが所定の閾値よりも大きい短時間エネルギーフレームの量の、全短時間エネルギーフレームの総量に対する比率を特定するステップと；
前記比率が所定の比率より大きいかどうかを特定するステップと；
肯定であれば、前記オーディオ信号は音声信号を備える、と特定し、否定であれば、前記オーディオ信号は音声信号を備えないと特定するステップと；を備える、
第１の局面に記載の方法。
［第５の局面］
各短時間エネルギーフレームの前記エネルギーに基づいて、前記オーディオ信号が音声信号を備えているかどうかを検出する前記ステップは：
エネルギーが所定の閾値よりも大きい短時間エネルギーフレームの量の、全短時間エネルギーフレームの総量に対する比率を特定するステップと；
前記比率が所定の比率より大きいかどうかを特定するステップと；
否定であれば、前記オーディオ信号は音声信号を備えない、と特定し、
肯定であれば、
エネルギーが前記所定の閾値より大きい前記短時間エネルギーフレーム内に少なくともＮ個の連続する短時間エネルギーフレームがあるときには前記オーディオ信号は音声信号を備える、と特定し、
エネルギーが前記所定の閾値より大きい前記短時間エネルギーフレーム内に少なくともＮ個の連続する短時間エネルギーフレームがないときには前記オーディオ信号は音声信号を備えない、と特定するステップと；を備える、

第１の局面に記載の方法。
［第６の局面］
音声信号検出装置であって：
オーディオ信号を取得するよう構成された取得モジュールと；
所定の音声信号の周波数に基づいて、前記オーディオ信号を複数の短時間エネルギーフレームに分割するよう構成された分割モジュールと；
各短時間エネルギーフレームのエネルギーを特定するよう構成された特定モジュールと；
各短時間エネルギーフレームの前記エネルギーに基づいて、前記オーディオ信号は音声信号を備えているかどうかを検出するよう構成された検出モジュールと；を備える、
音声信号検出装置。
［第７の局面］
前記取得モジュールは、
現在のオーディオ信号を取得し、
最後に取得されたオーディオ信号から、所定の時間周期を持つ副信号を切り取り、
前記取得されたオーディオ信号として機能するよう、前記現在のオーディオ信号と前記切り取られた副信号とを連結するよう構成される、
第６の局面に記載の装置。
［第８の局面］
前記分割モジュールは、
前記所定の音声信号の周波数に基づいて前記所定の音声信号の周期を特定し、
前記特定された周期に基づいて、前記オーディオ信号を、対応する持続時間が前記周期である複数の短時間エネルギーフレームに分割するよう構成される、
第６の局面に記載の装置。
［第９の局面］
前記検出モジュールは、
エネルギーが所定の閾値よりも大きい短時間エネルギーフレームの量の、全短時間エネルギーフレームの総量に対する比率を特定し、
前記比率が所定の比率より大きいかどうかを特定し、
肯定であれば、前記オーディオ信号は音声信号を備える、と特定し、
否定であれば、前記オーディオ信号は音声信号を備えない、と特定するよう構成される、
第６の局面に記載の装置。
［第１０の局面］
前記検出モジュールは、
エネルギーが所定の閾値よりも大きい短時間エネルギーフレームの量の、全短時間エネルギーフレームの総量に対する比率を特定し、
前記比率が所定の比率より大きいかどうかを特定し、
否定であれば、前記オーディオ信号は音声信号を備えない、と特定し、
肯定であれば、
エネルギーが前記所定の閾値より大きい前記短時間エネルギーフレーム内に少なくともＮ個の連続する短時間エネルギーフレームがあるときには前記オーディオ信号は音声信号を備える、と特定し、
エネルギーが前記所定の閾値より大きい前記短時間エネルギーフレーム内に少なくともＮ個の連続する短時間エネルギーフレームがないときには前記オーディオ信号は音声信号を備えない、と特定するよう構成される、
第６の局面に記載の装置。

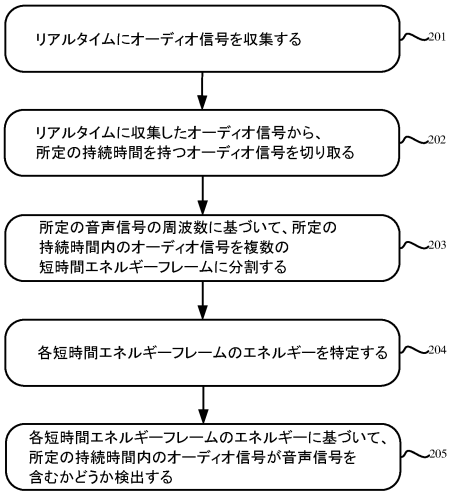
【符号の説明】
【００８４】

４１　取得モジュール
４２　分割モジュール
４３　特定モジュール
４４　検出モジュール

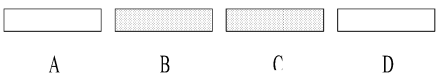
【図１】



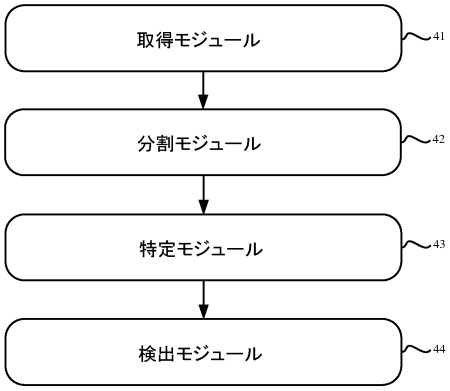
【図２】



【図３】



【図４】



フロントページの続き

(72)発明者 ジャオ, レイ
中華人民共和国, ヂィァンシァン ３１１１２１, ハンヂョウ, ユ ハンディストリクト, ウェスト ウェン イー ロード ナンバー ９６９, ビルディング ３, ５／エフ, アリババ グループ リーガル デパートメント
(72)発明者 グァン, イェンチュ
中華人民共和国, ヂィァンシァン ３１１１２１, ハンヂョウ, ユ ハンディストリクト, ウェスト ウェン イー ロード ナンバー ９６９, ビルディング ３, ５／エフ, アリババ グループ リーガル デパートメント
(72)発明者 ツァン, シャオドン
中華人民共和国, ヂィァンシァン ３１１１２１, ハンヂョウ, ユ ハンディストリクト, ウェスト ウェン イー ロード ナンバー ９６９, ビルディング ３, ５／エフ, アリババ グループ リーガル デパートメント
(72)発明者 リン, ファン
中華人民共和国, ヂィァンシァン ３１１１２１, ハンヂョウ, ユ ハンディストリクト, ウェスト ウェン イー ロード ナンバー ９６９, ビルディング ３, ５／エフ, アリババ グループ リーガル デパートメント

審査官 菊池 智紀

(56)参考文献 特開平１０－３０１６００（ＪＰ, Ａ）
特開２０００－２００１００（ＪＰ, Ａ）
特表２０１３－５０８７４４（ＪＰ, Ａ）

(58)調査した分野(Int.Cl., ＤＢ名)
Ｇ１０Ｌ ２５／７８－２５／８７, １５／０４