(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7395509号
(P7395509)

(45)発行日　令和5年12月11日(2023.12.11)

(24)登録日　令和5年12月1日(2023.12.1)

(51)国際特許分類

		FI		
G10L 15/28 (2013.01)		G10L	15/28	230K
G10L 15/22 (2006.01)		G10L	15/22	200Z
G10L 19/018 (2013.01)		G10L	19/018	

請求項の数　13　（全27頁）

(21)出願番号　特願2020-565375(P2020-565375)
(86)(22)出願日　令和1年5月22日(2019.5.22)
(65)公表番号　特表2021-525385(P2021-525385A)
(43)公表日　令和3年9月24日(2021.9.24)
(86)国際出願番号　PCT/US2019/033571
(87)国際公開番号　WO2019/226802
(87)国際公開日　令和1年11月28日(2019.11.28)
審査請求日　令和4年4月20日(2022.4.20)
(31)優先権主張番号　62/674,973
(32)優先日　平成30年5月22日(2018.5.22)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号　16/418,415
(32)優先日　令和1年5月21日(2019.5.21)
最終頁に続く

(73)特許権者　502208397
グーグル エルエルシー
Ｇｏｏｇｌｅ ＬＬＣ
アメリカ合衆国 カリフォルニア州 ９４０４３ マウンテン ビュー アンフィシアター パークウェイ １６００
１６００ Ａｍｐｈｉｔｈｅａｔｒｅ Ｐａｒｋｗａｙ ９４０４３ Ｍｏｕｎｔａｉｎ Ｖｉｅｗ， ＣＡ Ｕ．Ｓ．Ａ．
(74)代理人　100108453
弁理士　村山 靖彦
(74)代理人　100110364
弁理士　実広 信哉
(74)代理人　100133400
弁理士　阿部 達彦
最終頁に続く

(54)【発明の名称】　ホットワード抑制

(57)【特許請求の範囲】
【請求項１】
コンピュータ実装方法であって、
コンピューティングデバイスによって、発声の再生に対応するオーディオデータを受信するステップと、
前記コンピューティングデバイスによって、モデルに、前記オーディオデータを入力として与えるステップであって、前記モデルが、
(i)所与のオーディオデータサンプルがオーディオウォーターマークを含むかどうかを判断するように構成されており、
(ii)オーディオウォーターマークサンプルを各々が含む、ウォーターマークつきオーディオデータサンプルと、オーディオウォーターマークサンプルを各々が含まない、ウォーターマークなしオーディオデータサンプルとを使ってトレーニングされており、前記ウォーターマークつきオーディオデータサンプルは、少なくとも2つの異なるウォーターマークつきオーディオデータサンプルを含む、ステップと、
前記コンピューティングデバイスによって、前記モデルから、前記オーディオデータが前記オーディオウォーターマークを含むかどうかを示すデータを受信するステップと、
前記オーディオデータが前記オーディオウォーターマークを含むかどうかを示す前記データに基づいて、前記コンピューティングデバイスによって、前記オーディオデータの処理を続けるか、またはやめると判断するステップとを含む方法。
【請求項２】

前記オーディオデータが前記オーディオウォーターマークを含むかどうかを示す前記データを受信することは、前記オーディオデータが前記オーディオウォーターマークを含むことを示す前記データを受信することを含み、
前記オーディオデータの処理を続けるか、またはやめると判断することは、前記オーディオデータが前記オーディオウォーターマークを含むことを示す前記データを受信したことに基づいて、前記オーディオデータの処理をやめると判断することを含み、
前記方法は、前記オーディオデータの処理をやめると判断したことに基づいて、前記コンピューティングデバイスによって、前記オーディオデータの処理をやめるステップをさらに含む、請求項1に記載の方法。
【請求項３】
前記オーディオデータが前記オーディオウォーターマークを含むかどうかを示す前記データを受信することは、前記オーディオデータが前記オーディオウォーターマークを含まないことを示す前記データを受信することを含み、
前記オーディオデータの処理を続けるか、またはやめると判断することは、前記オーディオデータが前記オーディオウォーターマークを含まないことを示す前記データを受信したことに基づいて、前記オーディオデータの処理を続けると判断することを含み、
前記方法は、前記オーディオデータの処理を続けると判断したことに基づいて、前記コンピューティングデバイスによって、前記オーディオデータの処理を続けることをさらに含む、請求項1に記載の方法。
【請求項４】
前記オーディオデータの前記処理は、
前記オーディオデータに対して音声認識を実施することによって、前記発声の転写を生成することを含む、請求項1から3のいずれか一項に記載の方法。
【請求項５】
前記オーディオデータの前記処理は、
前記オーディオデータが、特定の、あらかじめ定義されたホットワードの発声を含むかどうかを判断することを含む、請求項1から4のいずれか一項に記載の方法。
【請求項６】
前記モデルに前記オーディオデータを入力として与える前に、前記コンピューティングデバイスによって、前記オーディオデータが特定の、あらかじめ定義されたホットワードの発声を含むと判断するステップを含む、請求項1から5のいずれか一項に記載の方法。
【請求項７】
前記コンピューティングデバイスによって、前記オーディオデータが特定の、あらかじめ定義されたホットワードの発声を含むと判断するステップを含み、
前記モデルに前記オーディオデータを入力として与えることは、前記オーディオデータが特定の、あらかじめ定義されたホットワードの発声を含むとの判断に応答する、請求項1から6のいずれか一項に記載の方法。
【請求項８】
前記コンピューティングデバイスによって、各々がオーディオウォーターマークを含む、前記ウォーターマークつきオーディオデータサンプルと、各々がオーディオウォーターマークを含まない、前記ウォーターマークなしオーディオデータサンプルと、ウォーターマークつきおよびウォーターマークなしオーディオサンプルの各々がオーディオウォーターマークを含むかどうかを示すデータとを受信するステップと、
前記コンピューティングデバイスによって、および機械学習を使って、各々がオーディオウォーターマークを含む、前記ウォーターマークつきオーディオデータサンプルと、各々が前記オーディオウォーターマークを含まない、前記ウォーターマークなしオーディオデータサンプルと、ウォーターマークつきおよびウォーターマークなしオーディオサンプルの各々がオーディオウォーターマークを含むかどうかを示す前記データとを使って前記モデルをトレーニングするステップと
を含む、請求項1から7のいずれか一項に記載の方法。

【請求項９】
前記ウォーターマークつきオーディオデータサンプルの少なくとも一部分は各々、複数の周期的ロケーションにおいてオーディオウォーターマークを含む、請求項8に記載の方法。
【請求項１０】
前記ウォーターマークつきオーディオデータサンプルのうちの1つにおけるオーディオウォーターマークは、前記ウォーターマークつきオーディオデータサンプルのうちの別の1つにおけるオーディオウォーターマークとは異なる、請求項8に記載の方法。
【請求項１１】
前記コンピューティングデバイスによって、発声の再生に対応する前記オーディオデータの受信の第1の時間を判断するステップと、
前記コンピューティングデバイスによって、追加コンピューティングデバイスが発声の再生に対応する前記オーディオデータを出力用に与えた第2の時間、および前記オーディオデータがウォーターマークを含んでいたかどうかを示すデータを受信するステップと、
前記コンピューティングデバイスによって、前記第1の時間が前記第2の時間と一致すると判断するステップと、
前記第1の時間が前記第2の時間と一致すると判断したことに基づいて、前記コンピューティングデバイスによって、前記オーディオデータがウォーターマークを含んでいたかどうかを示す前記データを使って前記モデルをアップデートするステップとを含む、請求項1から10のいずれか一項に記載の方法。
【請求項１２】
1つまたは複数のコンピュータと、
命令を記憶する1つまたは複数の記憶デバイスとを備えるシステムであって、前記1つまたは複数の記憶デバイスは、前記1つまたは複数のコンピュータによって実行されると、前記1つまたは複数のコンピュータに、請求項1から11のいずれか一項に記載の方法を実施させるように動作可能である、システム。
【請求項１３】
1つまたは複数のコンピュータによって実行可能な命令を備えるソフトウェアを記憶するコンピュータ可読記録媒体であって、前記命令は、実行されると、前記1つまたは複数のコンピュータに、請求項1から11のいずれか一項に記載の方法を実行させる、コンピュータ可読記録媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
関連出願の相互参照
本出願は、2018年5月22日に出願された米国特許出願第62/674,973号の利益を主張する、2019年5月21日に出願された米国特許出願第16/418,415号の利益を主張し、両出願の内容が参照によって組み込まれている。
【０００２】
本開示は概して、自動化音声処理に関する。
【背景技術】
【０００３】
音声対応ホームまたは他の環境、すなわち、ユーザが問合せまたはコマンドを声に出して話すだけでよく、コンピュータベースのシステムが問合せを受け止め、答え、かつ/またはコマンドを実施させる環境の現実性は、我々次第である。音声対応環境(たとえば、家庭、職場、学校など)は、環境の様々な部屋またはエリア中に分散された被接続マイクロフォンデバイスのネットワークを使って実装することができる。そのような、マイクロフォンのネットワークを通して、ユーザには、コンピュータまたは他のデバイスが自分の前または付近にさえもある必要なく、本質的には環境のどこからでもシステムに口頭で問い合わせる力がある。たとえば、台所で料理している間、ユーザは、システムに「3カップは何

ミリリットル?」と尋ね、それに応答して、たとえば、合成ボイス出力の形で、システムから答えを受け取る場合がある。代替として、ユーザが、「一番近いガソリンスタンドが閉まるのはいつ」、または、家を出る支度をしたときの、「今日はコートを着るべき?」などの質問をシステムに尋ねる場合がある。

【０００４】

さらに、ユーザは、システムに問合せを尋ね、かつ/またはユーザの個人情報に関するコマンドを発行し得る。たとえば、ユーザは、システムに「ジョンと会うのはいつ?」と尋ね、またはシステムに「家に帰ってきたらジョンに電話することを思い出させて」と命じる場合がある。

【発明の概要】

【課題を解決するための手段】

【０００５】

音声対応システムのために、システムとのユーザの対話法は、排他的ではないとしても主に、ボイス入力を用いるように設計される。したがって、システムは、システムに向けられていないものを含む、周辺環境において行われたすべての発声(Utterance)を拾い出す可能性があり、任意の所与の発声が、たとえば、環境の中にいる個人に向けられるのとは反対に、システムに向けられたときを見分ける何らかのやり方を有していなければならない。これを遂行するための1つのやり方は、環境の中のユーザの間の合意によって、システムの注意を喚起するために話される所定の単語または複数の単語として予約されている「ホットワード」を使うものである。ある例示的環境では、システムの注意を喚起するのに使われるホットワードは、「OKコンピュータ」という語句である。したがって、「OKコンピュータ」という語句は、話されるたびに、マイクロフォンによって拾い出され、システムに伝えられ、システムは、音声認識技法を実施するか、またはオーディオ特徴およびニューラルネットワークを使って、ホットワードが話されたかどうかを判断してもよく、話された場合、その後のコマンドまたは問合せを待ち受ける。したがって、システムに向けられた発声は、[HOTWORD][QUERY]という一般的形をとり、この例における「HOTWORD」は「OKコンピュータ」であり、「QUERY」は、システムによって、単独で、またはネットワークを介してサーバとともに音声認識され、解析され、作用され得るいかなる質問、コマンド、宣言、または他の要求であってもよい。

【０００６】

本開示は、録音された音声、たとえば、放送された音声またはテキスト-音声オーディオを、ライブ音声とは区別するための、オーディオウォーターマーキングベースの手法について論じる。この区別は、録音された音声を含む入力の中の偽ホットワードトリガの検出を可能にし、偽ホットワードトリガを抑制させる。ただし、ユーザからのライブ音声入力はウォーターマーキングされず、ウォーターマーキングされていないと判断された、音声入力の中のホットワードは抑制されなくてよい。ウォーターマーク検出機構は、騒がしく、反響する環境に対して堅牢であり、小さいフットプリント、メモリと計算の両方、および低遅延という目標を満足するように設計されている、畳み込みニューラルネットワークベースの検出器を使い得る。この手法のスケーラビリティ利点は、視聴者の多いテレビジョンイベント中の、数百万のデバイス上での同時ホットワードトリガを防止する際に際立つ。

【０００７】

ホットワードベースのトリガリングは、仮想アシスタントをアクティブ化するための機構であり得る。ライブ音声中のホットワードを、録音された音声、たとえば、広告中のものと区別することは、偽ホットワードトリガが、仮想アシスタントの意図的でないアクティブ化につながるので、問題となり得る。その上、ユーザが、仮想アシスタントを複数のデバイス上にインストールしてある所では、ある仮想アシスタントからの音声出力が、別の仮想アシスタントを意図せずにトリガするホットワードを含むことも可能である。仮想アシスタントの意図的でないアクティブ化は概して、望ましくない場合がある。たとえば、仮想アシスタントが、ホームオートメーションデバイスを制御するのに使われる場合、

仮想アシスタントの意図的でないアクティブ化はたとえば、照明、暖房または空調機器が意図せずにオンにされることにつながる場合があり、そうすることによって、不必要なエネルギー消費につながり、ならびにユーザにとって不便である。また、デバイスは、オンにされると、他のデバイスへメッセージを送信する(たとえば、他のデバイスから情報を取り出すため、その状況を他のデバイスにシグナリングするため、検索エンジンと通信して検索を実施するため、など)場合があり、そうすることによって、デバイスを意図せずにオンにすることは、不必要なネットワークトラフィックおよび/または処理容量の不必要な使用に、不必要な電力消費などにもつながり得る。その上、照明、暖房または空調機器などの機器の意図的でないアクティブ化は、機器の不必要な消耗を引き起こし、その信頼性を低下させる可能性がある。さらに、仮想アシスタント制御機器およびデバイスの範囲が増すと、仮想アシスタントの意図的でないアクティブ化が潜在的に危険になり得る可能性も増す。また、仮想アシスタントの意図的でないアクティブ化は、プライバシーに対する懸念を引き起こす可能性がある。

【０００８】

本出願に記載する主題のある発明的態様によると、ホットワードを抑制するための方法は、コンピューティングデバイスによって、発声の再生に対応するオーディオデータを受信するアクションと、コンピューティングデバイスによって、(i)所与のオーディオデータサンプルがオーディオウォーターマークを含むかどうかを判断するように構成されており、(ii)オーディオウォーターマークサンプルを各々が含む、ウォーターマークつきオーディオデータサンプル、およびオーディオウォーターマークサンプルを各々が含まない、ウォーターマークなしオーディオデータサンプルを使ってトレーニングされたモデルに、オーディオデータを入力として与えるアクションと、コンピューティングデバイスによって、(i)所与のオーディオデータサンプルがオーディオウォーターマークを含むかどうかを判断するように構成されており、(ii)オーディオウォーターマークを含む、ウォーターマークつきオーディオデータサンプル、およびオーディオウォーターマークを含まない、ウォーターマークなしオーディオデータサンプルを使ってトレーニングされたモデルから、オーディオデータがオーディオウォーターマークを含むかどうかを示すデータを受信するアクションと、オーディオデータがオーディオウォーターマークを含むかどうかを示すデータに基づいて、コンピューティングデバイスによって、オーディオデータの処理を続けるか、またはやめると判断するアクションとを含む。

【０００９】

これらおよび他の実装形態は各々、任意選択で、以下の特徴のうちの1つまたは複数を含み得る。オーディオデータがオーディオウォーターマークを含むかどうかを示すデータを受信するアクションは、オーディオデータがオーディオウォーターマークを含むことを示すデータを受信することを含む。オーディオデータの処理を続けるか、またはやめると判断するアクションは、オーディオデータがオーディオウォーターマークを含むことを示すデータを受信したことに基づいて、オーディオデータの処理をやめると判断することを含む。アクションは、オーディオデータの処理をやめると判断したことに基づいて、コンピューティングデバイスによって、オーディオデータの処理をやめることをさらに含む。オーディオデータがオーディオウォーターマークを含むかどうかを示すデータを受信するアクションは、オーディオデータがオーディオウォーターマークを含まないことを示すデータを受信することを含む。オーディオデータの処理を続けるか、またはやめると判断するアクションは、オーディオデータがオーディオウォーターマークを含まないことを示すデータを受信したことに基づいて、オーディオデータの処理を続けると判断することを含む。

【００１０】

アクションは、オーディオデータの処理を続けると判断したことに基づいて、コンピューティングデバイスによって、オーディオデータの処理を続けることをさらに含む。オーディオデータの処理のアクションは、オーディオデータに対して音声認識を実施することによって、発声の転写(Transcription)を生成することを含む。オーディオデータの処理

のアクションは、オーディオデータが、特定の、あらかじめ定義されたホットワードの発声を含むかどうかを判断することを含む。アクションは、(i)所与のオーディオデータサンプルがオーディオウォーターマークを含むかどうかを判断するように構成されており、(ii)オーディオウォーターマークサンプルを各々が含む、ウォーターマークつきオーディオデータサンプル、およびオーディオウォーターマークサンプルを各々が含まない、ウォーターマークなしオーディオデータサンプルを使ってトレーニングされたモデルにオーディオデータを入力として与える前に、コンピューティングデバイスによって、オーディオデータが特定の、あらかじめ定義されたホットワードの発声を含むと判断することをさらに含む。アクションは、コンピューティングデバイスによって、オーディオデータが特定の、あらかじめ定義されたホットワードの発声を含むと判断することをさらに含む。(i)所与のオーディオデータサンプルがオーディオウォーターマークを含むかどうかを判断するように構成されており、(ii)オーディオウォーターマークサンプルを各々が含む、ウォーターマークつきオーディオデータサンプル、およびオーディオウォーターマークサンプルを各々が含まない、ウォーターマークなしオーディオデータサンプルを使ってトレーニングされたモデルにオーディオデータを入力として与えるアクションは、オーディオデータが特定の、あらかじめ定義されたホットワードの発声を含むとの判断に応答する。

【００１１】

アクションは、コンピューティングデバイスによって、各々がオーディオウォーターマークを含む、ウォーターマークつきオーディオデータサンプル、各々がオーディオウォーターマークを含まない、ウォーターマークなしオーディオデータサンプル、ならびに各ウォーターマークつきおよびウォーターマークなしオーディオサンプルがオーディオウォーターマークを含むかどうかを示すデータを受信することと、コンピューティングデバイスによって、および機械学習を使って、各々がオーディオウォーターマークを含む、ウォーターマークつきオーディオデータサンプル、オーディオウォーターマークを各々が含まない、ウォーターマークなしオーディオデータサンプル、ならびに各ウォーターマークつきおよびウォーターマークなしオーディオサンプルがオーディオウォーターマークを含むかどうかを示すデータを使ってモデルをトレーニングすることとをさらに含む。ウォーターマークつきオーディオデータサンプルの少なくとも一部分は各々、複数の周期的ロケーションにおいてオーディオウォーターマークを含む。ウォーターマークつきオーディオデータサンプルのうちの1つにおけるオーディオウォーターマークは、ウォーターマークつきオーディオデータサンプルのうちの別のものにおけるオーディオウォーターマークとは異なる。アクションは、コンピューティングデバイスによって、発声の再生に対応するオーディオデータの受信の第1の時間を判断することと、コンピューティングデバイスによって、追加コンピューティングデバイスが発声の再生に対応するオーディオデータを出力用に与えた第2の時間、およびオーディオデータがウォーターマークを含んでいたかどうかを示すデータを受信することと、コンピューティングデバイスによって、第1の時間が第2の時間と一致すると判断することと、第1の時間が第2の時間と一致すると判断したことに基づいて、コンピューティングデバイスによって、オーディオデータがウォーターマークを含んでいたかどうかを示すデータを使ってモデルをアップデートすることとをさらに含む。

【００１２】

本態様の他の実装形態は、方法の動作を実施するように各々が構成された、対応するシステムと、装置と、コンピュータ記憶デバイス上に記録されたコンピュータプログラムとを含む。本態様の他の実装形態は、1つまたは複数のコンピュータによって実行可能な命令を備えるソフトウェアを記憶するコンピュータ可読媒体を含み、命令は、そのように実行されると、1つまたは複数のコンピュータに、本明細書に記載する方法のうちのいずれかを含む動作を実施させる。

【００１３】

本明細書に記載する本主題の特定の実装形態は、以下の利点のうちの1つまたは複数を実現するように実装され得る。コンピューティングデバイスは、ライブ音声に含まれるホ

ットワードに応答し得るが、記録済み媒体に含まれるホットワードには応答しない。このことは、デバイスの意図的でないアクティブ化を削減または防止し、そうすることによって、コンピューティングデバイスのバッテリー電力および処理容量を節約することができる。より少ないコンピューティングデバイスが、オーディオウォーターマークのあるホットワードを受信すると検索クエリを実施することで、ネットワーク帯域幅も保持され得る。

【００１４】

本明細書において説明される主題の1つまたは複数の実施態様の詳細は、添付の図面および以下の説明に記載される。本主題の他の特徴、態様および利点は、説明、図面および特許請求の範囲から明らかになるであろう。

【図面の簡単な説明】

【００１５】

【図１】記録済み媒体中にホットワードを検出したときにホットワードトリガを抑制するための例示的システムを示す図である。

【図２】記録済み媒体中にホットワードを検出したときにホットワードトリガを抑制するための例示的プロセスのフローチャートである。

【図３】ウォーターマーキング領域中のフレームについての例示的最小マスキング閾値、エネルギー、および聴覚の絶対閾値を示す図である。

【図４】(a)ホスト信号の例示的規模スペクトグラム、(b)ウォーターマーク信号の例示的規模スペクトグラム、(c)ウォーターマーク信号の例示的な複製された符号行列、および(d)符号行列の単一の事例をもつ、複製された符号行列パターンの例示的相関を示す図であって、垂直線は、複製の間のウォーターマークパターンの例示的境界を表す。

【図５】ウォーターマーク検出器用に使われる例示的ニューラルネットワークアーキテクチャを示す図である。

【図６】相互相関パターンの複製によって作成された例示的照合フィルタを示す図である。

【図７】ウォーターマークなし信号用の、例示的ニューラルネットワーク出力および例示的な照合フィルタリングされたニューラルネットワーク出力を示す図である。

【図８】コンピューティングデバイスおよびモバイルコンピューティングデバイスの例を示す図である。

【発明を実施するための形態】

【００１６】

図面において、類似の参照番号は、全体を通して対応する部分を表す。

【００１７】

図1は、記録済み媒体中に「ホットワード」を検出したときにホットワードトリガを抑制するための例示的システム100を示す。簡潔には、および以下でより詳しく説明するように、コンピューティングデバイス104は、オーディオウォーターマーク116を含む発声108と、あらかじめ定義されたホットワード110の発声とを出力する。コンピューティングデバイス102は、発声108を検出し、オーディオウォーターマーク識別モデル158を使うことによって、発声108がオーディオウォーターマーク134を含むと判断する。発声108がオーディオウォーターマーク134を含むことに基づいて、コンピューティングデバイス102は、あらかじめ定義されたホットワード110には応答しない。

【００１８】

より詳しくは、コンピューティングデバイス104は、Nugget Worldのコマーシャルを再生中である。コマーシャルの間、コマーシャルに出ている俳優が、発声108、すなわち「Okコンピュータ、ナゲットには何が入っている?」を言う。発声108は、ホットワード110「Okコンピュータ」と、「ナゲットには何が入っている?」という他の言葉を含む問合せ112とを含む。コンピューティングデバイス104は、拡声器を通して発声108を出力する。マイクロフォンの付近にあるどのコンピューティングデバイスも、発声108を検出することができる。

【００１９】

発声108のオーディオは、音声部分114およびオーディオウォーターマーク116を含む

。コマーシャルの作成者は、発声108を検出するコンピューティングデバイスが、ホットワード110に応答しないようにするために、オーディオウォーターマーク116を追加してよい。いくつかの実装形態では、オーディオウォーターマーク116は、人間聴覚範囲よりも高いか、または低いオーディオ周波数を含み得る。たとえば、オーディオウォーターマーク116は、20kHzよりも大きいか、または20Hzよりも小さい周波数を含み得る。いくつかの実装形態では、オーディオウォーターマーク116は、人間聴覚範囲内であるが、音が雑音と類似しているので、人間によって検出可能でないオーディオを含み得る。たとえば、オーディオウォーターマーク116は、8と10kHzとの間の周波数パターンを含み得る。異なる周波数帯域の強度は、人には知覚不可能であり得るが、コンピューティングデバイスによって検出可能であり得る。周波数ドメイン表現115によって示されるように、発声108は、可聴部分114よりも高い周波数範囲の中にあるオーディオウォーターマーク116を含む。

【００２０】

いくつかの実装形態では、コンピューティングデバイス104は、音声データ118にウォーターマークを追加するのに、オーディオウォーターマーカー120を使えばよい。音声データ118は、「Okコンピュータ、ナゲットには何が入っている?」という、録音された発声108であってよい。オーディオウォーターマーカー120は、周期的間隔で、音声データ118の中にウォーターマークを追加することができる。たとえば、オーディオウォーターマーカー120は、ウォーターマークを200ミリ秒ごとに追加し得る。いくつかの実装形態では、コンピューティングデバイス104は、たとえば、音声認識を実施することによって、ホットワード110を含む、音声データ118の部分を識別することができる。オーディオウォーターマーカー120は、ホットワード110の前、および/またはホットワード110の後に、ホットワード110のオーディオにわたって周期的ウォーターマークを追加することができる。たとえば、オーディオウォーターマーカー120は、周期的間隔で、「okコンピュータ」というオーディオにわたって3つ(または任意の他の数)のウォーターマークを追加することができる。

【００２１】

ウォーターマーク116を追加するための技法については、図3～図7に関して後で詳しく論じる。概して、各ウォーターマーク116は、各音声データサンプル用に異なる。オーディオウォーターマーカー120は、200または300ミリ秒ごとに、発声108のオーディオにオーディオウォーターマークを追加し、「Okコンピュータ、チーズピザを注文して」という発声のオーディオに、200または300ミリ秒ごとに、異なるか、または同じオーディオウォーターマークを追加し得る。オーディオウォーターマーカー120は、ウォーターマークがオーディオサンプルの歪みを最小限にするように、各オーディオサンプル用にウォーターマークを生成してよい。このことは、オーディオウォーターマーカー120が、人間が検出することができる周波数範囲内のウォーターマークを追加することができるので、重要であり得る。コンピューティングデバイス104は、ウォーターマークつき音声112中のウォーターマークつきオーディオサンプルを、コンピューティングデバイス104によって後で出力するために記憶してもよい。

【００２２】

いくつかの実装形態では、コンピューティングデバイス104がウォーターマークつきオーディオを出力するたびに、コンピューティングデバイス104は、出力されたオーディオを示すデータを再生ログ124に記憶し得る。再生ログ124は、出力されたオーディオ108、オーディオ108を出力した日時、コンピューティングデバイス104、コンピューティングデバイス104のロケーション、オーディオ108の転写、およびウォーターマークのないオーディオ108のどの組合せも識別するデータを含み得る。

【００２３】

コンピューティングデバイス102は、マイクロフォンを通して発声108を検出する。コンピューティングデバイス102は、オーディオを受信することが可能な、どのタイプのデバイスであってもよい。たとえば、コンピューティングデバイス102は、デスクトップコ

ンピュータ、ラップトップコンピュータ、タブレットコンピュータ、ウェアラブルコンピュータ、セルラーフォン、スマートフォン、音楽プレーヤ、eブックリーダ、ナビゲーションシステム、スマートスピーカーおよびホームアシスタント、ワイヤレス(たとえば、Bluetooth)ヘッドセット、補聴器、スマートウォッチ、スマートグラス、活動追跡器、またはどの他の適切なコンピューティングデバイスであってもよい。図1に示すように、コンピューティングデバイス102はスマートフォンである。コンピューティングデバイス104は、たとえば、テレビジョン、ラジオ、音楽プレーヤ、デスクトップコンピュータ、ラップトップコンピュータ、タブレットコンピュータ、ウェアラブルコンピュータ、セルラーフォン、またはスマートフォンなど、オーディオを出力することが可能ないかなるデバイスであってもよい。図1に示すように、コンピューティングデバイス104はテレビジョンである。

【００２４】

コンピューティングデバイス102のマイクロフォンは、オーディオサブシステム150の一部であってよい。オーディオサブシステム150は、マイクロフォンを通して受信されたオーディオを最初に処理するように各々が設計されているバッファ、フィルタ、アナログデジタルコンバータを含み得る。バッファは、マイクロフォンを通して受信され、オーディオサブシステム150によって処理された現在のオーディオを記憶し得る。たとえば、バッファは、前の5秒間のオーディオデータを記憶する。

【００２５】

コンピューティングデバイス102は、オーディオウォーターマーク識別器152を含む。オーディオウォーターマーク識別器152は、マイクロフォンを通して受信され、かつ/またはバッファに記憶されたオーディオを処理し、オーディオに含まれるオーディオウォーターマークを識別するように構成される。オーディオウォーターマーク識別器152は、処理されたオーディオを、入力としてオーディオウォーターマーク識別モデル158に与えるように構成され得る。オーディオウォーターマーク識別モデル158は、オーディオデータを受信し、オーディオデータがウォーターマークを含むかどうかを示すデータを出力するように構成され得る。たとえば、オーディオウォーターマーク識別器152は、オーディオサブシステム150を通して処理されたオーディオを、オーディオウォーターマーク識別モデル158に絶えず与え得る。オーディオウォーターマーク識別器152がより多くのオーディオを与えると、オーディオウォーターマーク識別モデル158の精度が増し得る。たとえば、300ミリ秒後、オーディオウォーターマーク識別モデル158は、1つのウォーターマークを含むオーディオを受信している場合がある。500ミリ秒後、オーディオウォーターマーク識別モデル158は、2つのウォーターマークを含むオーディオを受信している場合がある。どの1つのオーディオサンプル中のウォーターマークもすべて、互いと同一である実施形態では、オーディオウォーターマーク識別モデル158は、より多くのオーディオを処理することによって、精度を向上させ得る。

【００２６】

いくつかの実装形態では、オーディオウォーターマーク識別器152は、オーディオサブシステム150から受信されたオーディオから、どの検出されたウォーターマークも削除するように構成されてよい。ウォーターマークを削除した後、オーディオウォーターマーク識別器152は、ウォーターマークのないオーディオを、ホットワーダ154および/または音声認識器162に与えてよい。いくつかの実装形態では、オーディオウォーターマーク識別器152は、オーディオサブシステム150から受信されたオーディオを、ウォーターマークを削除せずにホットワーダ154および/または音声認識器162に渡すように構成されてよい。

【００２７】

ホットワーダ154は、マイクロフォンを通して受信され、かつ/またはバッファに記憶されたオーディオ中のホットワードを識別するように構成される。いくつかの実装形態では、ホットワーダ154は、コンピューティングデバイス102が電源投入されたときはいつでもアクティブであってよい。ホットワーダ154は、バッファに記憶されたオーディオデー

タを絶えず分析することができる。ホットワーダ154は、バッファ中の現在のオーディオデータがホットワードを含む見込みを反映するホットワード信頼性スコアを計算する。ホットワード信頼性スコアを計算するために、ホットワーダ154は、オーディオデータから、フィルタバンクエネルギーまたはメル周波数ケプストラム係数などのオーディオ特徴を抽出してもよい。ホットワーダ154は、たとえばサポートベクトルマシンまたはニューラルネットワークを使うことによって、これらのオーディオ特徴を処理するのに、分類ウィンドウを使ってよい。いくつかの実装形態では、ホットワーダ154は、ホットワード信頼性スコアを判断するために(たとえば、受信されたオーディオから抽出されたオーディオ特徴を、ホットワードのうちの1つまたは複数についての対応するオーディオ特徴と比較することによって、ただしオーディオデータに対して音声認識を実施するために、抽出されたオーディオ特徴を使わずに)音声認識を実施しない。ホットワード信頼性スコアがホットワード信頼性スコア閾値を満足する場合、ホットワーダ154は、オーディオがホットワードを含むと判断する。たとえば、ホットワード信頼性スコアが0.8であり、ホットワード信頼性スコア閾値が0.7である場合、ホットワーダ154は、発声108に対応するオーディオがホットワード110を含むと判断する。いくつかの事例では、ホットワードは、起動語または注目語と呼ばれ得る。

【００２８】

音声認識器162は、入来オーディオに基づいて転写を生成する、どのタイプのプロセスを実施してもよい。たとえば、音声認識器162は、バッファ中のオーディオデータの中の音素を識別するのに、音響モデルを使ってもよい。音声認識器162は、音素に対応する転写を判断するのに、言語モデルを使えばよい。別の例として、音声認識器162は、バッファ中のオーディオデータを処理し、転写を出力する単一モデルを使ってよい。

【００２９】

オーディオウォーターマーク識別モデル158が、オーディオがウォーターマークを含むと判断する事例では、オーディオウォーターマーク識別器152は、音声認識器162および/またはホットワーダ154を非アクティブ化し得る。音声認識器162および/またはホットワーダ154を非アクティブ化することによって、オーディオウォーターマーク識別器152は、ホットワード110および/または問合せ112に応答するようにコンピューティングデバイス102をトリガし得るオーディオのさらなる処理を防止することができる。図1に示すように、オーディオウォーターマーク識別器152は、ホットワーダ154を非アクティブ状態156に、および音声認識器162を非アクティブ状態160にセットする。

【００３０】

いくつかの実装形態では、ホットワーダ154のデフォルト状態はアクティブ状態であってよく、音声認識器162のデフォルト状態はアクティブ状態であってよい。この事例では、非アクティブ状態156および非アクティブ状態160は、所定の時間量の後で満了し得る。たとえば、5秒(または別の所定の時間量)後、ホットワーダ154と音声認識器162の両方の状態はアクティブ状態に戻り得る。5秒の期間は、オーディオウォーターマーク識別器152がオーディオウォーターマークを検出するたびに更新されてよい。たとえば、発声108のオーディオ115が、オーディオの持続時間を通してウォーターマークを含む場合、ホットワーダ154および音声認識器162は、非アクティブ状態156および非アクティブ状態160にセットされてよく、コンピューティングデバイス104の端部が発声108を出力した後、追加の5秒間、その状態に留まってよい。別の例として、発声108のオーディオ115が、ホットワード110の発声を通してウォーターマークを含む場合、ホットワーダ154および音声認識器162は、非アクティブ状態156および非アクティブ状態160にセットされてよく、コンピューティングデバイス104が、問合せ112の出力と重複するホットワード110を出力した後、追加の5秒間、その状態に留まってよい。

【００３１】

いくつかの実装形態では、オーディオウォーターマーク識別器152は、オーディオウォーターマーク識別器152がウォーターマークを識別した日時を示すデータを、識別ログ164に記憶し得る。たとえば、オーディオウォーターマーク識別器152は、2019年6月10日

午後3:15における発声108のオーディオ中のウォーターマークを識別し得る。識別ログ164は、ウォーターマークの受信の時刻および日付、ウォーターマーク134を含む発声の転写、コンピューティングデバイス102、ウォーターマーク134、ウォーターマークを検出したときのコンピューティングデバイス102のロケーション、基底オーディオ132、オーディオとウォーターマークの組合せ、ならびに発声108またはウォーターマーク134の、ある時間期間前または後に検出された任意のオーディオのどの組合せも識別するデータを記憶し得る。

【００３２】

いくつかの実装形態では、オーディオウォーターマーク識別器152は、オーディオウォーターマーク識別器152がウォーターマークを識別せず、ホットワーダ154がホットワードを識別した日時を示すデータを、識別ログ164に記憶し得る。たとえば、2019年6月20日午後7:15に、オーディオウォーターマーク識別器152は、発声のオーディオ中のウォーターマークを識別しない場合があり、ホットワーダ154は、発声のオーディオ中のホットワードを識別する場合がある。識別ログ164は、ウォーターマークなしオーディオおよびホットワードの受信の時刻および日付、発声の転写、コンピューティングデバイス102、コンピューティングデバイスのロケーション、発声またはホットワードの、ある時間期間前または後に検出されたオーディオのどの組合せも識別するデータを記憶し得る。

【００３３】

いくつかの実装形態では、ホットワーダ154は、オーディオウォーターマーク識別器152の前、後、またはそれと同時に、オーディオサブシステム150から受信されたオーディオを処理することができる。たとえば、オーディオウォーターマーク識別器152は、発声108のオーディオがウォーターマークを含むと判断する場合があり、同じときに、ホットワーダ154は、発声108のオーディオがホットワードを含むと判断する場合がある。この事例では、オーディオウォーターマーク識別器152は、音声認識器162の状態を非アクティブ状態160にセットしてよい。オーディオウォーターマーク識別器152は、ホットワーダ154の状態156をアップデートすることができない場合がある。

【００３４】

いくつかの実装形態では、オーディオウォーターマーク識別器152がオーディオウォーターマーク識別モデル158を使う前に、コンピューティングデバイス106は、ウォーターマーク識別データ130を生成し、ウォーターマーク識別データ130をコンピューティングデバイス102に与える。コンピューティングデバイス106は、オーディオウォーターマーク識別モデル148を生成するのに、ウォーターマークなし音声サンプル136と、オーディオウォーターマーカー138と、機械学習を使う訓練器144とを使う。

【００３５】

ウォーターマークなし音声サンプル136は、様々な条件の下で収集された様々な音声サンプルを含み得る。ウォーターマークなし音声サンプル136は、異なる言葉を話す、同じ言葉を話す、異なるタイプのバックグラウンドノイズをもつ言葉を話す、異なる言語で言葉を話す、異なるアクセントで言葉を話す、異なるデバイスによって録音された言葉を話す、など、異なるユーザのオーディオサンプルを含み得る。いくつかの実装形態では、ウォーターマークなし音声サンプル136は各々、ホットワードの発声を含む。いくつかの実装形態では、ウォーターマークなし音声サンプル136のうちのいくつかのみが、ホットワードの発声を含む。

【００３６】

オーディオウォーターマーカー138は、各ウォーターマークなし音声サンプル用に、異なるウォーターマークを生成することができる。オーディオウォーターマーカー138は、各ウォーターマークなし音声サンプル用に、1つまたは複数のウォーターマークつき音声サンプル140を生成することができる。同じウォーターマークなし音声サンプルを使って、オーディオウォーターマーカー138は、200ミリ秒ごとにウォーターマークを含むウォーターマークつき音声サンプルと、300ミリ秒ごとにウォーターマークを含む別のウォーターマークつき音声サンプルとを生成し得る。オーディオウォーターマーカー138は、も

しあれば、ホットワードとのみ重複するウォーターマークを含む、ウォーターマークつき音声サンプルも生成し得る。オーディオウォーターマーカー138は、ホットワードと重複し、ホットワードと先行するウォーターマークを含む、ウォーターマークつき音声サンプルも生成し得る。この事例では、オーディオウォーターマーカー138は、同じウォーターマークなし音声サンプルを用いて、4つの異なるウォーターマークつき音声サンプルを作ることができる。オーディオウォーターマーカー138は、4よりも多いか、または少ないものを作ることもできる。いくつかの事例では、オーディオウォーターマーカー138は、オーディオウォーターマーカー120と同様に動作してよい。

【００３７】

訓練器144は、機械学習と、ウォーターマークなし音声サンプル136およびウォーターマークつき音声サンプル140を含むトレーニングデータとを使って、オーディオウォーターマーク識別モデル148を生成する。ウォーターマークなし音声サンプル136およびウォーターマークつき音声サンプル140は、ウォーターマークを含むか、またはウォーターマークを含まないものとして標示されるので、訓練器144は、ウォーターマークなし音声サンプル136、および各サンプルがウォーターマークを含まないことを示すラベル、ならびにウォーターマークつき音声サンプル140、および各サンプルがウォーターマークを含むことを示すラベルを含むトレーニングデータを使ってもよい。訓練器144は、機械学習を使って、オーディオサンプルを受信し、オーディオサンプルがウォーターマークを含むかどうかを出力することができるように、オーディオウォーターマーク識別モデル148を生成する。

【００３８】

コンピューティングデバイス106は、オーディオウォーターマーク識別モデル148にアクセスし、受信されたオーディオデータを処理する際に使うために、モデル128をコンピューティングデバイス102に与えることができる。コンピューティングデバイス102は、モデル128をオーディオウォーターマーク識別モデル158に記憶することができる。

【００３９】

コンピューティングデバイス106は、再生ログ142および識別ログ146に基づいて、オーディオウォーターマーク識別モデル148をアップデートしてよい。再生ログ142は、コンピューティングデバイス104から受信され、再生ログ124に記憶された再生データ126などのデータを含み得る。再生ログ142は、ウォーターマークつきオーディオを出力した複数のコンピューティングデバイスからの再生データを含み得る。識別ログ146は、コンピューティングデバイス102から受信され、識別ログ164に記憶された識別データ130などのデータを含み得る。識別ログ146は、オーディオウォーターマークを識別し、ウォーターマークつきオーディオに含まれるどのコマンドまたは問合せの実行も防止するように構成されている複数のコンピューティングデバイスからの追加識別データを含み得る。

【００４０】

訓練器144は、再生ログ142と識別ログ146を比較して、コンピューティングデバイスがウォーターマークつきオーディオを出力し、別のコンピューティングデバイスがウォーターマークつきオーディオ中のウォーターマークを識別したことを示す一致エントリを識別し得る。訓練器144は、識別ログ146および再生ログ142中のウォーターマーク識別エラーも識別し得る。第1のタイプのウォーターマーク識別エラーは、コンピューティングデバイスがウォーターマークを識別したことを識別ログ146が示すが、再生ログ142がウォーターマークつきオーディオの出力を示さないときに起こり得る。第2のタイプのウォーターマーク識別エラーは、再生ログ142はウォーターマークつきオーディオの出力を示すが、識別ログ146は、ウォーターマークつきオーディオの付近のコンピューティングデバイスがウォーターマークを識別しなかったことを示すときに起こり得る。

【００４１】

訓練器144は、エラーをアップデートし、対応するオーディオデータを、オーディオウォーターマーク識別モデル148をアップデートするための追加トレーニングデータとして使ってよい。訓練器144は、コンピューティングデバイスがウォーターマークを適切に識

別した場合、オーディオを使ってオーディオウォーターマーク識別モデル148をアップデートしてもよい。訓練器144は、コンピューティングデバイスによって出力されたオーディオと、コンピューティングデバイスによって検出されたオーディオの両方を、トレーニングデータとして使ってよい。訓練器144は、機械学習と、再生ログ142および識別ログ146に記憶されたオーディオデータとを使って、オーディオウォーターマーク識別モデル148をアップデートしてよい。訓練器144は、再生ログ142および識別ログ146中で与えられるウォーターマーキングラベルと、機械学習トレーニングプロセスの一部として上述したエラー識別技法からの訂正されたラベルとを使えばよい。

【００４２】

いくつかの実装形態では、コンピューティングデバイス102およびいくつかの他のコンピューティングデバイスは、オーディオ115を、サーバ上で稼動している、サーバベースのホットワーダおよび/またはサーバベースの音声認識器による処理のために、サーバへ送信するように構成され得る。オーディオウォーターマーク識別器152は、オーディオ115がオーディオウォーターマークを含まないことを示す場合がある。その判断に基づいて、コンピューティングデバイス102は、オーディオを、サーバベースのホットワーダおよび/またはサーバベースの音声認識器によるさらなる処理のために、サーバへ送信してよい。いくつかの他のコンピューティングデバイスのオーディオウォーターマーク識別器も、オーディオ115がオーディオウォーターマークを含まないことを示す場合がある。それらの判断に基づいて、他のコンピューティングデバイスの各々が、それらのそれぞれのオーディオを、サーバベースのホットワーダおよび/またはサーバベースの音声認識器によるさらなる処理のために、サーバへ送信してよい。サーバは、各コンピューティングデバイスからのオーディオが、ホットワードを含むかどうかを判断し、かつ/またはオーディオの転写を生成し、結果を各コンピューティングデバイスへ返送し得る。

【００４３】

いくつかの実装形態では、サーバは、ウォーターマーク決定の各々についてのウォーターマーク信頼性スコアを示すデータを受信し得る。サーバは、コンピューティングデバイス102および他のコンピューティングデバイスのロケーション、受信されたオーディオの特性、各オーディオ部分を同様の時間に受信したこと、ならびにどの他の同様のインジケータにも基づいて、コンピューティングデバイス102および他のコンピューティングデバイスによって受信されたオーディオが、同じソースからであると判断し得る。いくつかの事例では、ウォーターマーク信頼性スコアの各々は、範囲の一端におけるウォーターマーク信頼性スコア閾値と、5パーセント未満など、ウォーターマーク信頼性スコア閾値からのパーセント差であってよい別の信頼性スコアとを含む特定の範囲内であってよい。たとえば、範囲は、0.80～0.76のウォーターマーク信頼性スコア閾値であってよい。他の事例では、範囲の他端は、0.05など、ウォーターマーク信頼性スコア閾値からの固定距離であってよい。たとえば、範囲は、0.80～0.75のウォーターマーク信頼性スコア閾値であってよい。

【００４４】

サーバが、ウォーターマーク信頼性スコアの各々が、ウォーターマーク信頼性スコア閾値に近いがその閾値を満足しない範囲内であると判断した場合、サーバは、ウォーターマーク信頼性スコア閾値が調節されるべきであると判断してよい。この事例では、サーバは、ウォーターマーク信頼性スコア閾値を、範囲の低い方の端に調節してもよい。いくつかの実装形態では、サーバは、各コンピューティングデバイスから受信されたオーディオをウォーターマークつき音声サンプル140に含めることによって、ウォーターマークつき音声サンプル140をアップデートしてよい。訓練器144は、機械学習およびアップデートされたウォーターマークつき音声サンプル140を使って、オーディオウォーターマーク識別モデル148をアップデートし得る。

【００４５】

図1は、上述した異なる機能を実施する3つの異なるコンピューティングデバイスを示すが、1つまたは複数のコンピューティングデバイスのどの組合せも、機能の任意の組合せ

を実施することができる。たとえば、別個のコンピューティングデバイス106がオーディオウォーターマーク識別モデル148をトレーニングするのではなく、コンピューティングデバイス102がオーディオウォーターマーク識別モデル148をトレーニングしてよい。

【００４６】

図2は、記録済み媒体中にホットワードを検出したときにホットワードトリガを抑制するための例示的プロセス200を示す。概して、プロセス200は、受信されたオーディオを処理して、オーディオがオーディオウォーターマークを含むかどうかを判断する。オーディオがオーディオウォーターマークを含む場合、プロセス200は、オーディオのさらなる処理を抑制し得る。オーディオがオーディオウォーターマークを含まない場合、プロセス200は、オーディオを処理し、オーディオに含まれるどの問合せまたはコマンドも実行し続ける。プロセス200は、1つまたは複数のコンピュータ、たとえば、図1に示すコンピューティングデバイス102、104、および/または106を備えるコンピュータシステムによって実施されるものとして記載される。

【００４７】

システムは、発声の再生に対応するオーディオデータを受信する(210)。たとえば、テレビジョンがコマーシャルを再生中であってよく、コマーシャルに出ている俳優が、「Okコンピュータ、明かりをつけて」という場合がある。システムはマイクロフォンを含み、マイクロフォンは、俳優の発声を含む、コマーシャルのオーディオを検出する。

【００４８】

システムは、(i)所与のオーディオデータサンプルがオーディオウォーターマークを含むかどうかを判断するように構成されており、(ii)オーディオウォーターマークサンプルを各々が含む、ウォーターマークつきオーディオデータサンプル、およびオーディオウォーターマークサンプルを各々が含まない、ウォーターマークなしオーディオデータサンプルを使ってトレーニングされたモデルに、オーディオデータを入力として与える(220)。いくつかの実装形態では、システムは、オーディオデータがホットワードを含むと判断し得る。ホットワードを検出したことに基づいて、システムは、オーディオデータを、入力としてモデルに与える。たとえば、システムは、オーディオデータが「okコンピュータ」を含むと判断し得る。「okコンピュータ」を検出したことに基づいて、システムはオーディオデータをモデルに与える。システムは、ホットワードを含んでいた、オーディオデータの部分と、ホットワードの後に受信されたオーディオとを与え得る。いくつかの事例では、システムは、ホットワードの前からのオーディオの部分を与え得る。

【００４９】

いくつかの実装形態では、システムは、オーディオデータを分析して、オーディオデータがホットワードを含むかどうかを判断し得る。分析は、オーディオデータを入力としてモデルに与える前または後に起こり得る。いくつかの実装形態では、システムは、機械学習と、各々がオーディオウォーターマークを含む、ウォーターマークつきオーディオデータサンプル、各々がオーディオウォーターマークを含まない、ウォーターマークなしオーディオデータサンプル、ならびに各ウォーターマークつきおよびウォーターマークなしオーディオサンプルがオーディオウォーターマークを含むかどうかを示すデータとを使って、モデルをトレーニングすることができる。システムは、モデルへのオーディオ入力がウォーターマークを含むか、それともウォーターマークを含まないかを示すデータを出力するように、モデルをトレーニングすることができる。

【００５０】

いくつかの実装形態では、異なるウォーターマークつきオーディオ信号は、互いとは異なるウォーターマークを含み得る(どの1つのオーディオサンプル中のウォーターマークもすべて、互いと同一であり得るが、あるオーディオ信号中のウォーターマークは、別のオーディオ信号中のウォーターマークとは異なる)。システムは、オーディオ信号中の歪みを最小限にするように、各オーディオ信号用に異なるウォーターマークを生成してよい。いくつかの実装形態では、システムは、ウォーターマークを、オーディオ信号中に周期的間隔で配置してよい。たとえば、システムは、ウォーターマークを200ミリ秒ごとに配置し

てよい。いくつかの実装形態では、システムは、ホットワードを含むオーディオおよび/またはホットワードの前の時間期間にわたって、ウォーターマークを配置してよい。

【００５１】

システムは、(i)所与のオーディオデータサンプルがオーディオウォーターマークを含むかどうかを判断するように構成されており、(ii)オーディオウォーターマークを含む、ウォーターマークつきオーディオデータサンプル、およびオーディオウォーターマークを含まない、ウォーターマークなしオーディオデータサンプルを使ってトレーニングされたモデルから、オーディオデータがオーディオウォーターマークを含むかどうかを示すデータを受信する(230)。システムは、オーディオデータがウォーターマークを含むという指示を受信するか、またはオーディオデータがウォーターマークを含まないという指示を受信し得る。

【００５２】

システムは、オーディオデータがオーディオウォーターマークを含むかどうかを示すデータに基づいて、オーディオデータの処理を続けるか、またはやめる(240)。いくつかの実装形態では、システムは、オーディオデータがオーディオウォーターマークを含む場合、オーディオデータの処理をやめてよい。いくつかの実装形態では、システムは、オーディオデータがオーディオウォーターマークを含まない場合、オーディオデータの処理を続けてよい。いくつかの実装形態では、オーディオデータの処理は、オーディオデータに対して音声認識を実施すること、および/またはオーディオデータがホットワードを含むかどうかを判断することを含み得る。いくつかの実装形態では、処理は、オーディオデータに含まれる問合せまたはコマンドを実行することを含み得る。

【００５３】

いくつかの実装形態では、システムは、システムがオーディオデータを受信した時刻および日付をロギングする。システムは、この時刻および日付を、オーディオデータを出力するコンピューティングデバイスから受信された時刻および日付と比較してもよい。システムが、オーディオデータの受信の日時が、オーディオデータを出力した日時と一致すると判断した場合、システムは、オーディオデータを追加トレーニングデータとして使ってモデルをアップデートしてよい。システムは、オーディオデータがウォーターマークを含んでいたかどうかを判断する際に、モデルが正しかったかどうかを識別し、オーディオデータが、トレーニングデータに追加されるときに正しいウォーターマークラベルを含むようにしてもよい。

【００５４】

より詳しくは、ユーザ向けのタスクを実施することができるソフトウェアエージェントは概して、「仮想アシスタント」と呼ばれる。仮想アシスタントは、たとえば、ユーザからのボイス入力によって作動されてよく、すなわち、たとえば、ユーザによって話されると、仮想アシスタントをアクティブ化させ、話されたトリガ語に関連付けられたタスクを実施させる1つまたは複数のトリガ語を認識するようにプログラムされてよい。そのようなトリガ語はしばしば、「ホットワード」と呼ばれる。仮想アシスタントは、たとえば、ユーザのコンピュータ携帯電話または他のユーザデバイス上に設けられ得る。代替として、仮想アシスタントは、いわゆる「スマートスピーカー」(1つまたは複数のホットワードの助けにより、対話型アクションおよびハンズフリーアクティブ化を提供する統合型仮想アシスタントをもつタイプのワイヤレススピーカー)など、別のデバイス中に統合されてよい。

【００５５】

スマートスピーカーの広い採用に伴い、追加問題が生じる。大勢の観衆がいるイベント、たとえば、1億人を超える視聴者を引きつけるスポーツイベント中、ホットワードのある広告は、仮想アシスタントの同時トリガリングにつながる可能性がある。多数の視聴者がいることにより、音声認識サーバへの同時問合せの大幅な増加がある場合があり、そのことがサービス拒否(DOS)につながり得る。

【００５６】

偽ホットワードをフィルタリングするための2つの可能機構は、(1)オーディオフィンガープリンティングであって、問合せオーディオからのフィンガープリントが、広告のような、既知のオーディオからのフィンガープリントのデータベースと照合されて、偽トリガをフィルタアウトする、オーディオフィンガープリンティング、および(2)オーディオウォーターマーキングであって、オーディオが公開元によってウォーターマーキングされ、仮想アシスタントによって録音された問合せが、フィルタリングするためのウォーターマークを調べられる、オーディオウォーターマーキングに基づくものである。

【００５７】

本開示は、畳み込みニューラルネットワークを使う、低遅延の、小型フットプリントウォーターマーク検出器の設計について記載する。このウォーターマーク検出器は、対象のシナリオにおいて頻繁であり得る、騒がしく、反響する環境に対して堅牢であるようにトレーニングされる。

【００５８】

オーディオウォーターマーキングは、著作権保護および第2画面アプリケーションにおいて使われ得る。著作権保護では、ウォーターマーク検出は概して、オーディオ信号全体が検出に応じられるとき、遅延に敏感である必要はない。第2画面アプリケーションのケースでは、高遅延ウォーターマーク検出によりもたらされる遅れは、許容可能であり得る。これらの2つのシナリオとは異なり、仮想アシスタントにおけるウォーターマーク検出は、遅延に非常に敏感である。

【００５９】

ウォーターマーク検出を伴う、知られているアプリケーションでは、ウォーターマークを構成する埋め込みメッセージは通常、前もっては知られておらず、ウォーターマーク検出器は、メッセージシーケンスがウォーターマークを含むかどうかを判断し得る前に、メッセージシーケンスを復号し、ウォーターマークを含む場合、ウォーターマークを判断しなければならない。ただし、本明細書に記載するいくつかのアプリケーションでは、ウォーターマーク検出器は、デコーダ/ウォーターマーク検出器によって正確に知られているウォーターマークパターンを検出中であり得る。すなわち、録音された音声コンテンツの公開元またはプロバイダは、これにウォーターマークでウォーターマーキングしてもよく、ウォーターマークの詳細を、たとえば、仮想アシスタントのプロバイダおよび/または仮想アシスタントを含むデバイスのプロバイダに対して入手可能にしてもよい。同様に、仮想アシスタントのプロバイダは、仮想アシスタントからの音声出力がウォーターマークを与えられるように手配し、ウォーターマークの詳細を入手可能にしてもよい。その結果、受信されたメッセージ中でウォーターマークが検出されると、受信されたメッセージは、ユーザからのライブ音声入力ではないことがわかり、受信されたメッセージ中のどのホットワードから生じた仮想アシスタントのアクティブ化も、メッセージ全体が受信され、処理されるまで待つ必要なく抑制され得る。これにより、遅延が削減される。

【００６０】

ホットワード抑制のためのいくつかの実装形態は、オーディオフィンガープリンティング手法を使用する。この手法は、既知のオーディオのフィンガープリントデータベースを必要とする。デバイス上でのこのデータベースの維持は非自明なので、そのようなソリューションのデバイス上展開は実現可能でない。ただし、オーディオフィンガープリンティング手法の多大な利点は、オーディオ公開プロセスに対する修正を必要としなくてよいことである。したがって、この手法は、オーディオ公開元が協力者でない敵対的シナリオにも取り組むことができる。

【００６１】

本開示は、ウォーターマークベースのホットワード抑制機構について記載する。ホットワード抑制機構は、メモリおよび計算フットプリントの設計制約をもたらすデバイス上展開を使い得る。さらに、ユーザエクスペリエンスに対する影響を避けるために、遅延に対する制約がある。

【００６２】

ウォーターマークベースの手法は、ウォーターマークを追加するための、オーディオ公開プロセスの修正を必要とし得る。したがって、これらの手法は、協力者によって公開されたオーディオを検出するのに使われ得るだけであることがある。ただし、これらの手法は、フィンガープリントデータベースの維持を必要としなくてよい。この特徴により、いくつかの利点が可能になる。

【００６３】

第1の利点は、デバイス上展開の実行可能性であり得る。これは、いくつかの仮想アシスタントが一斉にトリガされる場合がある高視聴者数イベント中に、利点となり得る。これらの偽トリガを検出するための、サーバベースのソリューションは、同時トリガのスケールにより、サービスの拒否につながり得る。第2の利点は、協力者によって公開された未知のオーディオ、たとえば、公開元が協力的であり得るが、オーディオが前もっては知られていないテキスト-音声(TTS)シンセサイザ出力の検出であり得る。第3の利点は、スケーラビリティであり得る。オンラインプラットフォーム上のオーディオ/ビデオ公開元などのエンティティは、それらのオーディオに、仮想アシスタントをトリガするのを避けるためにウォーターマーキングすることができる。いくつかの実装形態では、これらのプラットフォームは、実際にはオーディオフィンガープリンティングベースの手法を使って扱うことができない、数百万時間のコンテンツをホストする。

【００６４】

いくつかの実装形態では、本明細書に記載するウォーターマークベースの手法は、敵対的エージェントに取り組む能力を有し得るオーディオフィンガープリンティングベースの手法と組み合わせることができる。

【００６５】

以下の記述は、ウォーターマーク埋め込み器およびウォーターマーク検出器について記載する。

【００６６】

ウォーターマーク埋め込み器は、FFTドメインにおけるスペクトル拡散ベースのウォーターマーキングに基づき得る。ウォーターマーク埋め込み器は、ウォーターマーク信号の振幅を形成するのに使われる最小マスキング閾値(ＭＭＴ)を推定するのに、心理音響モデルを使えばよい。

【００６７】

この技法を要約すると、ウォーターマーク追加を受けるホスト信号の領域が、最小エネルギー基準に基づいて選択される。離散フーリエ変換(DFT)係数が、これらの領域中のあらゆるホスト信号フレーム(25msウィンドウ-12.5msホップ)について推定される。これらのDFT係数は、心理音響モデルを使って最小マスキング閾値(ＭＭＴ)を推定するのに使われる。ＭＭＴは、ウォーターマーク信号のフレームについての規模スペクトルを形成するのに使われる。図3は、推定されたＭＭＴを、ホスト信号エネルギーおよび聴覚の絶対閾値とともに提示する。ホスト信号の位相は、ウォーターマーク信号用に使われてよく、DFT係数の符号は、メッセージペイロードから判断される。メッセージビットペイロードは、多重スクランブル化を使って、フレームのチャンクに広げられ得る。いくつかの実装形態では、システムは、問合せがウォーターマーキングされているかどうかを検出中である場合があり、どのペイロードも送信する必要はない場合がある。したがって、システムは、フレームのチャンク(たとえば、16フレームまたは200ms)にわたって、符号行列をランダムに選び、この符号行列をウォーターマーキング領域にわたって繰り返せばよい。符号行列のこの繰返しは、ウォーターマーク検出器出力を後処理し、検出性能を向上するのに活用され得る。個々のウォーターマークフレームの重複追加により、ウォーターマーク信号が生成し得る。図2のサブプロット(a)および(b)は、ホスト信号およびウォーターマーク信号の規模スペクトルを表し、サブプロット(c)は符号行列を表す。垂直線は、行列の2つの複製の間の境界を表す。

【００６８】

ウォーターマーク信号は、ウォーターマークの不可聴性をさらに確実にするために、時

間ドメインにおけるホスト信号を一定の因子(たとえば、α∈[0,1])だけスケーリングした後、ホスト信号に追加され得る。いくつかの実装形態では、αは、音質の知覚評価(PEAQ)のような客観評価メトリックを使って反復的に判断される。いくつかの実装形態では、システムは、伝統的スケーリング因子(たとえば、α∈{0.1、0.2、0.3、0.4、0.5})を使い、これらのスケーリング因子の各々での検出性能を評価し得る。

【００６９】

いくつかの実装形態では、ウォーターマーク検出器向けの設計要件は、モデルのメモリフットプリントと、その計算の複雑さの両方に対して大幅な制約を課す、デバイス上展開であり得る。以下の記述は、デバイス上キーワード検出のための畳み込みニューラルネットワークベースのモデルアーキテクチャについて記載する。いくつかの実装形態では、システムは、時間畳み込みニューラルネットワークを使い得る。

【００７０】

いくつかの実装形態では、ニューラルネットワークは、200msパターンの1つのインスタンスと同じ200msパターンの複製であり得る、埋め込まれたウォーターマーク符号行列(図4、サブプロット(c))の相互相関を推定するようにトレーニングされる。図4のサブプロット(d)は、相互相関を示す。相互相関は、各符号行列ブロックの開始についての情報を符号化することができ、ホスト信号内のウォーターマーク信号の持続時間全体に対して非ゼロであってよい。

【００７１】

システムは、マルチタスク損失関数を使ってニューラルネットワークをトレーニングし得る。一次タスクは、グランドトゥルース相互相関の推定であってよく、補助タスクは、エネルギー摂動パターンおよび/またはウォーターマーク規模スペクトルの推定であってよい。グランドトゥルースとネットワーク出力との間の平均2乗誤差が計算され得る。補助損失を正規化定数でスケーリングした後、損失の一部または全部が補間され得る。いくつかの実装形態では、各ネットワーク出力を、対応するグランドトゥルースのダイナミックレンジをちょうどカバーするように境界すると、性能が向上し得る。

【００７２】

いくつかの実装形態では、システムはネットワーク出力を後処理し得る。いくつかの実装形態では、ウォーターマークはペイロードメッセージを有していない場合があり、単一の符号行列が、ウォーターマーキング領域全体を通して複製される。これにより、周期的である相互相関パターンが生じ得る(図4、サブプロット(d))。この態様は、ネットワーク出力における擬似ピークをなくすのに活用することができる。いくつかの実装形態では、および性能を向上するために、システムは、対象の周波数を分離する帯域通過フィルタにわたって相互相関パターン(図6参照)を複製することによって作成された照合フィルタを使い得る。図7は、照合フィルタリングの前および後の、ウォーターマークなし信号用に生成されたネットワーク出力を比較する。いくつかの実装形態では、周期性を有していない擬似ピークは大幅に抑制することができる。グランドトゥルース705は、約0.0(たとえば、-0.01と0.01との間)であってよく、ネットワーク出力710および照合フィルタリングされたネットワーク出力720よりも入念にx軸を追跡し得る。ネットワーク出力710は、グランドトゥルース705および照合フィルタリングされたネットワーク出力720よりも、x軸に関して変化し得る。照合フィルタリングされたネットワーク出力720は、ネットワーク出力710よりも入念に、x軸を追跡してよく、グランドトゥルース705ほど入念にx軸を追跡しなくてよい。照合フィルタリングされたネットワーク出力720は、ネットワーク出力710よりも平滑であり得る。照合フィルタリングされたネットワーク出力720は、ネットワーク出力710よりも狭い範囲内に留まり得る。たとえば、照合フィルタリングされたネットワーク出力720は、-0.15と0.15との間に留まり得る。ネットワーク出力710は、-0.30と0.60との間に留まり得る。

【００７３】

ニューラルネットワークは、トレーニングされると、ニューラルネットワークを具現化するモデルをオーディオデータサンプルに適用することによって、所与のオーディオデー

タサンプルがオーディオウォーターマークを含むかどうかを判断する方法において使われ得る。この方法は、オーディオデータがオーディオウォーターマークを含む見込みを反映する信頼性スコアを判断するステップと、オーディオデータがオーディオウォーターマークを含む見込みを反映する信頼性スコアを信頼性スコア閾値と比較するステップと、オーディオデータがオーディオウォーターマークを含む見込みを反映する信頼性スコアを信頼性スコア閾値と比較したことに基づいて、オーディオデータに対して追加処理を実施するかどうかを判断するステップとを含み得る。

【００７４】

ある実施形態では、方法は、オーディオデータがオーディオウォーターマークを含む見込みを反映する信頼性スコアを信頼性スコア閾値と比較したことに基づいて、信頼性スコアが信頼性スコア閾値を満足すると判断するステップを含み、オーディオデータに対して追加処理を実施するかどうかを判断することは、オーディオデータに対する追加処理の実施を抑制すると判断することを含む。ある実施形態では、方法は、発声がオーディオウォーターマークを含む見込みを反映する信頼性スコアを信頼性スコア閾値と比較したことに基づいて、信頼性スコアが信頼性スコア閾値を満足しないと判断するステップを含み、オーディオデータに対して追加処理を実施するかどうかを判断することは、オーディオデータに対して追加処理を実施すると判断することを含む。ある実施形態では、方法は、ユーザから、オーディオデータに対する追加処理の実施を確認するデータを受信するステップと、オーディオデータに対する追加処理の実施を確認するデータを受信したことに基づいて、モデルをアップデートするステップとを含む。ある実施形態では、オーディオデータに対する追加処理は、オーディオデータの転写に基づいてアクションを実施すること、またはオーディオデータが、特定の、あらかじめ定義されたホットワードを含むかどうかを判断することを含む。ある実施形態では、方法は、オーディオデータに、(i)所与のオーディオデータサンプルがオーディオウォーターマークを含むかどうかを判断するように構成されており、(ii)オーディオウォーターマークを含む、ウォーターマークつきオーディオデータサンプル、およびオーディオウォーターマークを含まない、ウォーターマークなしオーディオデータサンプルを使ってトレーニングされたモデルを適用する前に、オーディオデータが、特定の、あらかじめ定義されたホットワードを含むと判断するステップを含む。ある実施形態では、方法は、オーディオデータが、特定の、あらかじめ定義されたホットワードを含むと判断するステップを含み、オーディオデータに、(i)所与のオーディオデータサンプルがオーディオウォーターマークを含むかどうかを判断するように構成されており、(ii)オーディオウォーターマークを含む、ウォーターマークつきオーディオデータサンプル、およびオーディオウォーターマークを含まない、ウォーターマークなしオーディオデータサンプルを使ってトレーニングされたモデルを適用することは、オーディオデータが、特定の、あらかじめ定義されたホットワードを含むとの判断に応答する。ある実施形態では、方法は、オーディオウォーターマークを含む、ウォーターマークつきオーディオデータサンプル、およびオーディオウォーターマークを含まない、ウォーターマークなしオーディオデータサンプルを受信するステップと、機械学習を使って、オーディオウォーターマークを含む、ウォーターマークつきオーディオデータサンプル、およびオーディオウォーターマークを含まない、ウォーターマークなしオーディオデータサンプルを使うモデルをトレーニングするステップとを含む。ある実施形態では、ウォーターマークつきオーディオデータサンプルの少なくとも一部分が、複数の周期的ロケーションにおけるオーディオウォーターマークを含む。

【００７５】

図8は、本明細書に記載する技法を実装するのに使われ得るコンピューティングデバイス800およびモバイルコンピューティングデバイス850の例を示す。コンピューティングデバイス800は、ラップトップ、デスクトップ、ワークステーション、携帯情報端末、サーバ、ブレードサーバ、メインフレーム、および他の適切なコンピュータなど、様々な形のデジタルコンピュータを表すことを意図している。モバイルコンピューティングデバイス850は、携帯情報端末、セルラー電話、スマートフォン、および他の同様のコンピュー

ティングデバイスなど、様々な形のモバイルデバイスを表すことを意図している。ここに示される構成要素、それらの接続および関係、ならびにそれらの機能は、例であることのみを目的としており、限定的であることは目的としていない。

【００７６】

コンピューティングデバイス800は、プロセッサ802と、メモリ804と、記憶デバイス806と、メモリ804および複数の高速拡張ポート810に接続する高速インターフェース808と、低速拡張ポート814および記憶デバイス806に接続する低速インターフェース812とを含む。プロセッサ802、メモリ804、記憶デバイス806、高速インターフェース808、高速拡張ポート810、および低速インターフェース812の各々は、様々なバスを使って相互接続され、共通マザーボード上に、または必要に応じて他の様式で搭載され得る。プロセッサ802は、GUIについてのグラフィカル情報を、高速インターフェース808に結合されたディスプレイ816などの外部入力/出力デバイス上に表示するための、メモリ804中または記憶デバイス806上に記憶された命令を含む、コンピューティングデバイス800内での実行のための命令を処理することができる。他の実装形態では、複数のプロセッサおよび/または複数のバスが、必要に応じて、複数のメモリおよびタイプのメモリとともに使われてよい。また、複数のコンピューティングデバイスが接続されてよく、各デバイスは、必要な動作の部分を(たとえば、サーババンク、ブレードサーバのグループ、またはマルチプロセッサシステムとして)提供する。

【００７７】

メモリ804は、コンピューティングデバイス800内に情報を記憶する。いくつかの実装形態において、メモリ804は、1つまたは複数の揮発性メモリユニットである。いくつかの実装形態において、メモリ804は、1つまたは複数の不揮発性メモリユニットである。メモリ804は、磁気または光ディスクなど、別の形のコンピュータ可読媒体であってもよい。

【００７８】

記憶デバイス806は、コンピューティングデバイス800に大容量記憶を提供することが可能である。いくつかの実装形態において、記憶デバイス806は、フロッピーディスクデバイス、ハードディスクデバイス、光ディスクデバイス、もしくはテープデバイス、フラッシュメモリもしくは他の同様の固体メモリデバイス、または記憶エリアネットワークもしくは他の構成におけるデバイスを含むデバイスのアレイなどのコンピュータ可読媒体であるか、またはそれらを含み得る。命令は、情報キャリアに記憶することができる。命令は、1つまたは複数の処理デバイス(たとえば、プロセッサ802)によって実行されると、上述したような1つまたは複数の方法を実施する。命令は、コンピュータまたは機械可読媒体など、1つまたは複数の記憶デバイス(たとえば、メモリ804、記憶デバイス806、またはプロセッサ802上のメモリ)によって記憶することもできる。

【００７９】

高速インターフェース808は、コンピューティングデバイス800向けの帯域消費型動作を管理し、低速インターフェース812は、より帯域低消費型動作を管理する。機能のそのような割振りは、例にすぎない。いくつかの実装形態において、高速インターフェース808は、メモリ804、ディスプレイ816に(たとえば、グラフィックスプロセッサまたはアクセラレータを通して)、および様々な拡張カードを受け得る高速拡張ポート810に結合される。本実装形態において、低速インターフェース812は、記憶デバイス806および低速拡張ポート814に結合される。低速拡張ポート814は、様々な通信ポート(たとえば、USB、Bluetooth、イーサネット、ワイヤレスイーサネット)を含み得るが、キーボード、ポインティングデバイス、スキャナなど、1つもしくは複数の入力/出力デバイス、またはスイッチもしくはルータなどのネットワークデバイスに、たとえば、ネットワークアダプタを通して結合され得る。

【００８０】

コンピューティングデバイス800は、図に示すように、いくつかの異なる形で実装されてよい。たとえば、標準サーバ820として、またはそのようなサーバのグループで複数回

、実装され得る。さらに、ラップトップコンピュータ822などのパーソナルコンピュータで実装され得る。また、ラックサーバシステム824の一部として実装され得る。代替として、コンピューティングデバイス800からの構成要素は、モバイルコンピューティングデバイス850などのモバイルデバイス中の他の構成要素と組み合わせることができる。そのようなデバイスの各々は、コンピューティングデバイス800およびモバイルコンピューティングデバイス850のうちの1つまたは複数を含んでよく、システム全体が、互いと通信する複数のコンピューティングデバイスから作られてよい。

【００８１】

モバイルコンピューティングデバイス850は、他の構成要素の中でも、プロセッサ852、メモリ864、ディスプレイ854などの入力/出力デバイス、通信インターフェース866、およびトランシーバ868を含む。モバイルコンピューティングデバイス850には、追加記憶を提供するために、マイクロドライブまたは他のデバイスなどの記憶デバイスが設けられてもよい。プロセッサ852、メモリ864、ディスプレイ854、通信インターフェース866、およびトランシーバ868の各々は、様々なバスを使って相互接続され、構成要素のうちのいくつかは、共通マザーボード上に、または必要に応じて他の方式で搭載されてよい。

【００８２】

プロセッサ852は、メモリ864中に記憶された命令を含む、モバイルコンピューティングデバイス850内の命令を実行することができる。プロセッサ852は、別個および複数のアナログおよびデジタルプロセッサを含むチップのチップセットとして実装され得る。プロセッサ852は、たとえば、ユーザインターフェース、モバイルコンピューティングデバイス850によって稼働されるアプリケーション、およびモバイルコンピューティングデバイス850によるワイヤレス通信の制御など、モバイルコンピューティングデバイス850の他の構成要素の協調を可能にし得る。

【００８３】

プロセッサ852は、ディスプレイ854に結合された制御インターフェース858およびディスプレイインターフェース856を通してユーザと通信し得る。ディスプレイ854は、たとえば、TFT(薄膜トランジスタ液晶ディスプレイ)ディスプレイもしくはOLED(有機発光ダイオード)ディスプレイ、または他の適切なディスプレイ技術であってよい。ディスプレイインターフェース856は、グラフィカルおよび他の情報をユーザに対して提示するようにディスプレイ854を駆動するための適切な回路機構を備え得る。制御インターフェース858は、コマンドを、ユーザから受信し、プロセッサ852への提出のために変換し得る。さらに、外部インターフェース862が、モバイルコンピューティングデバイス850と他のデバイスの近距離通信を可能にするように、プロセッサ852との通信を提供し得る。外部インターフェース862は、たとえば、いくつかの実装形態ではワイヤード通信を、または他の実装形態ではワイヤレス通信を提供することができ、複数のインターフェースが使われてもよい。

【００８４】

メモリ864は、モバイルコンピューティングデバイス850内に情報を記憶する。メモリ864は、1つもしくは複数のコンピュータ可読媒体、1つもしくは複数の揮発性メモリユニット、または1つもしくは複数の不揮発性メモリユニットのうちの1つまたは複数として実装され得る。拡張メモリ874が設けられ、拡張インターフェース872を通してモバイルコンピューティングデバイス850に接続されてもよく、インターフェース872は、たとえば、SIMM(シングルインラインメモリモジュール)カードインターフェースを含み得る。拡張メモリ874は、モバイルコンピューティングデバイス850に余剰記憶空間を提供することができ、またはモバイルコンピューティングデバイス850向けのアプリケーションもしくは他の情報を記憶することもできる。特に、拡張メモリ874は、上述したプロセスを実践し、または補うための命令を含むことができ、セキュアな情報も含むことができる。したがって、たとえば、拡張メモリ874は、モバイルコンピューティングデバイス850用のセキュリティモジュールとして設けられてよく、モバイルコンピューティングデバイス850のセキュアな使用を許可する命令でプログラムされ得る。さらに、ハッキングできない

ようにSIMMカード上に識別情報を置くなど、付加情報とともに、SIMMカードを介して、セキュアなアプリケーションが提供されてよい。

【００８５】

メモリは、たとえば、以下で論じるように、フラッシュメモリおよび/またはNVRAMメモリ(不揮発性ランダムアクセスメモリ)を含み得る。いくつかの実装形態では、命令は、1つまたは複数の処理デバイス(たとえば、プロセッサ852)によって命令が実行されると、上述したような1つまたは複数の方法を実施するように、情報キャリアに記憶される。命令は、1つまたは複数のコンピュータまたは機械可読媒体など、1つまたは複数の記憶デバイス(たとえば、メモリ864、拡張メモリ874、またはプロセッサ852上のメモリ)によって記憶することもできる。いくつかの実装形態において、命令は、たとえば、トランシーバ868または外部インターフェース862を介して、伝搬される信号中で受信されることが可能である。

【００８６】

モバイルコンピューティングデバイス850は、必要な場合はデジタル信号処理回路機構を含み得る通信インターフェース866を通してワイヤレスに通信することができる。通信インターフェース866は、中でも特に、GSMボイスコール(広域移動通信システム)、SMS(ショートメッセージサービス)、EMS(拡張メッセーングサービス)、もしくはＭＭＳメッセージ通信(マルチメディアメッセーングサービス)、CDMA(符号分割多元接続)、TDMA(時分割多元接続)、PDC(パーソナルデジタルセルラー)、WCDMA(登録商標)(広帯域符号分割多元接続)、CDMA2000、またはGPRS(汎用パケット無線サービス)など、様々なモードまたはプロトコルの下で通信を提供する場合がある。そのような通信は、たとえば、無線周波数を使うトランシーバ868を通して起こり得る。さらに、たとえばBluetooth、WiFi、または他のそのようなトランシーバを使って、短距離通信が起こり得る。さらに、GPS(全地球測位システム)受信機モジュール870が、追加ナビゲーションおよびロケーション関連ワイヤレスデータをモバイルコンピューティングデバイス850に提供してよく、このデータは、必要に応じて、モバイルコンピューティングデバイス850上で稼動するアプリケーションによって使われ得る。

【００８７】

モバイルコンピューティングデバイス850は、オーディオコーデック860を使って可聴的に通信することもでき、コーデック860は、発話情報を、ユーザから受信し、使用可能なデジタル情報に変換し得る。オーディオコーデック860は同様に、たとえば、モバイルコンピューティングデバイス850のハンドセット中のスピーカーを通すなどして、ユーザ向けの可聴音を生成し得る。そのような音は、ボイス通話からの音を含んでよく、記録された音(たとえば、ボイスメッセージ、音楽ファイルなど)を含んでよく、モバイルコンピューティングデバイス850上で動作するアプリケーションによって生成された音も含んでよい。

【００８８】

モバイルコンピューティングデバイス850は、図に示すように、いくつかの異なる形で実装されてよい。たとえば、セルラー電話880として実装され得る。また、スマートフォン882、携帯情報端末、または他の同様のモバイルデバイスの一部として実装され得る。

【００８９】

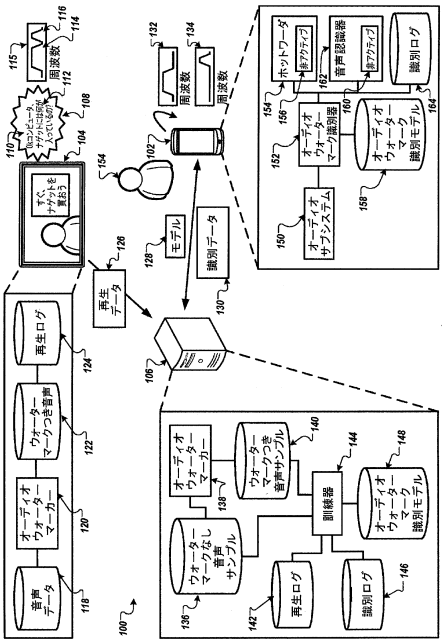
ここに記載するシステムおよび技法の様々な実装形態は、デジタル電子回路機構、集積回路機構、特別に設計されたASIC(特定用途向け集積回路)、コンピュータハードウェア、ファームウェア、ソフトウェア、および/またはそれらの組合せで実現され得る。これらの様々な実装形態は、少なくとも1つのプログラム可能プロセッサを含むプログラム可能システム上で実行可能および/または翻訳可能な1つまたは複数のコンピュータプログラムでの実装を含んでよく、プログラム可能プロセッサは、記憶システム、少なくとも1つの入力デバイス、および少なくとも1つの出力デバイスからデータおよび命令を受信するように、ならびにそれらにデータおよび命令を送信するように結合された、特殊または一般的目的であってよい。

【００９０】
これらのコンピュータプログラム(プログラム、ソフトウェア、ソフトウェアアプリケーションまたはコードとしても知られる)は、プログラム可能プロセッサ用の機械命令を含み、高度手続型および/もしくはオブジェクト指向プログラミング言語で、ならびに/またはアセンブリ/機械言語で実装され得る。本明細書で使用する機械可読媒体およびコンピュータ可読媒体という用語は、機械命令を機械可読信号として受信する機械可読媒体を含むプログラム可能プロセッサに、機械命令および/またはデータを提供するのに使われる、どのコンピュータプログラム製品、装置および/またはデバイス(たとえば、磁気ディスク、光ディスク、メモリ、プログラム可能論理デバイス(PLD))も指す。機械可読信号という用語は、プログラム可能プロセッサに機械命令および/またはデータを提供するのに使われるどの信号も指す。
【００９１】
ユーザとの対話を可能にするために、ここで記載するシステムおよび技法は、ユーザに情報を表示するための表示デバイス(たとえば、CRT(陰極線管)やLCD(液晶ディスプレイ)モニタ)と、ユーザがコンピュータに入力を与え得るためのキーボードおよびポインティングデバイス(たとえば、マウスやトラックボール)とを有するコンピュータ上で実装することができる。他の種類のデバイスも、ユーザとの対話を可能にするのに使うことができ、たとえば、ユーザに与えられるフィードバックは、どの形の感覚フィードバック(たとえば、視覚フィードバック、聴覚フィードバック、または触覚フィードバック)でもよく、ユーザからの入力は、音響、発話、または触覚入力を含む、どの形でも受信することができる。
【００９２】
ここで記載するシステムおよび技法は、バックエンド構成要素を(たとえば、データサーバとして)含む、もしくはミドルウェア構成要素(たとえば、アプリケーションサーバ)を含む、もしくはフロントエンド構成要素(たとえば、ここで記載するシステムおよび技法の実装形態とユーザが対話し得るためのグラフィカルユーザインターフェースもしくはウェブブラウザを有するクライアントコンピュータ)、またはそのようなバックエンド、ミドルウェア、もしくはフロントエンド構成要素のどの組合せも含むコンピューティングシステムで実装することができる。システムの構成要素は、どの形または媒体のデジタルデータ通信(たとえば、通信ネットワーク)によっても相互接続することができる。通信ネットワークの例には、ローカルエリアネットワーク(LAN)、ワイドエリアネットワーク(WAN)、およびインターネットがある。
【００９３】
コンピューティングシステムは、クライアントおよびサーバを含み得る。クライアントとサーバは概して、互いから離れており、通常、通信ネットワークを通して対話する。クライアントとサーバの関係は、それぞれのコンピュータ上で稼動するとともに互いとのクライアント-サーバ関係を有するコンピュータプログラムにより発生する。
【００９４】
以上、いくつかの実装形態を詳しく記載したが、他の修正が可能である。たとえば、本出願に記載される論理フローは、望ましい結果を達成するのに、図示される特定の順序、または順番を求めない。さらに、他のアクションが提供されてよく、またはアクションが、記載したフローからなくされてよく、他の構成要素が、記載したシステムに追加され、もしくはそこから削除されてよい。したがって、他の実装形態は、以下の特許請求の範囲内である。また、一態様または実装形態において記載した特徴が、どの他の態様または実装形態において適用されてもよい。
【符号の説明】
【００９５】
100 システム
102 コンピューティングデバイス
104 コンピューティングデバイス
106 コンピューティングデバイス

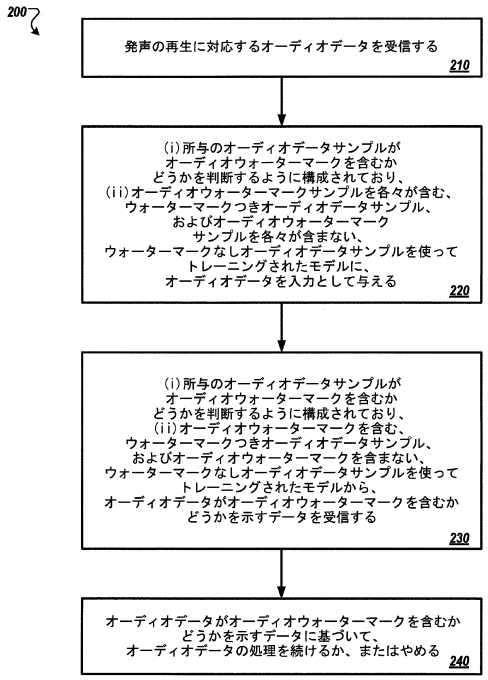
120 オーディオウォーターマーカー
124 再生ログ
138 オーディオウォーターマーカー
142 再生ログ
144 訓練器
146 識別ログ
148 オーディオウォーターマーク識別モデル
150 オーディオサブシステム
152 オーディオウォーターマーク識別器
154 ホットワーダ
158 オーディオウォーターマーク識別モデル
162 音声認識器
164 識別ログ
800 コンピューティングデバイス
802 プロセッサ
804 メモリ
806 記憶デバイス
808 高速インターフェース
810 高速拡張ポート
812 低速インターフェース
814 低速拡張ポート
816 ディスプレイ
820 標準サーバ
822 ラップトップコンピュータ
824 ラックサーバシステム
850 モバイルコンピューティングデバイス
852 プロセッサ
854 ディスプレイ
856 ディスプレイインターフェース
858 制御インターフェース
860 オーディオコーデック
862 外部インターフェース
864 メモリ
866 通信インターフェース
868 トランシーバ
870 GPS(全地球測位システム)受信機モジュール
872 拡張インターフェース
874 拡張メモリ
880 セルラー電話
882 スマートフォン

【図面】

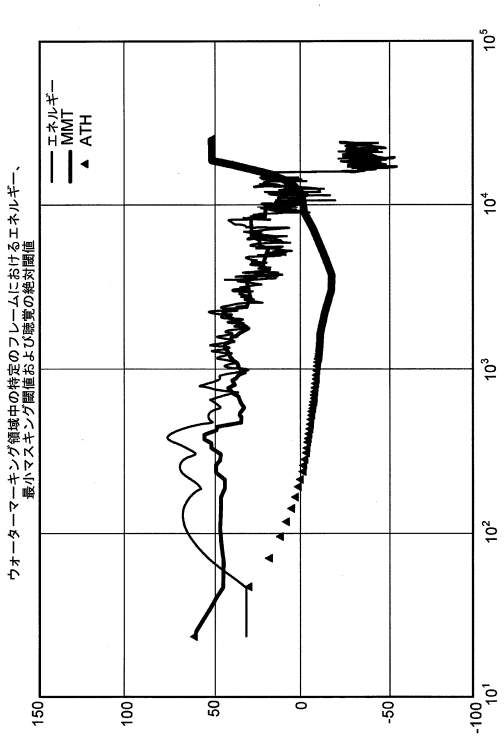
【図１】



【図２】



【図３】



【図４】

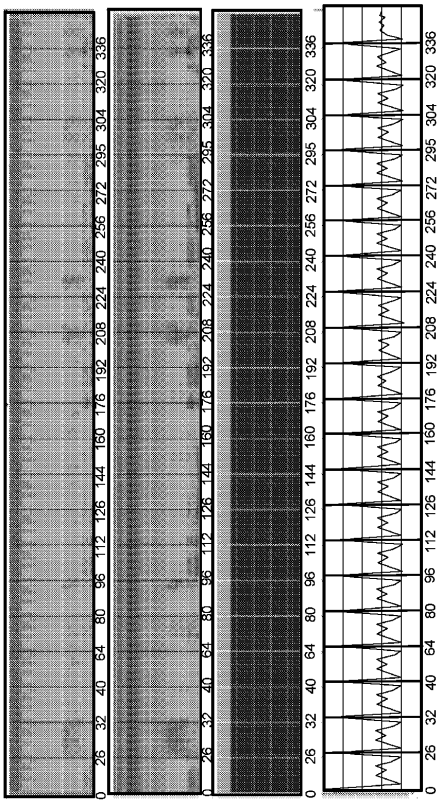
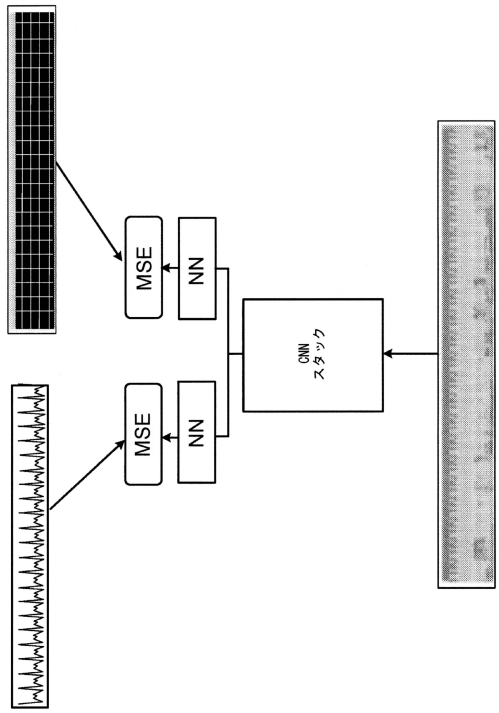


FIG. 4

【図５】



【図６】

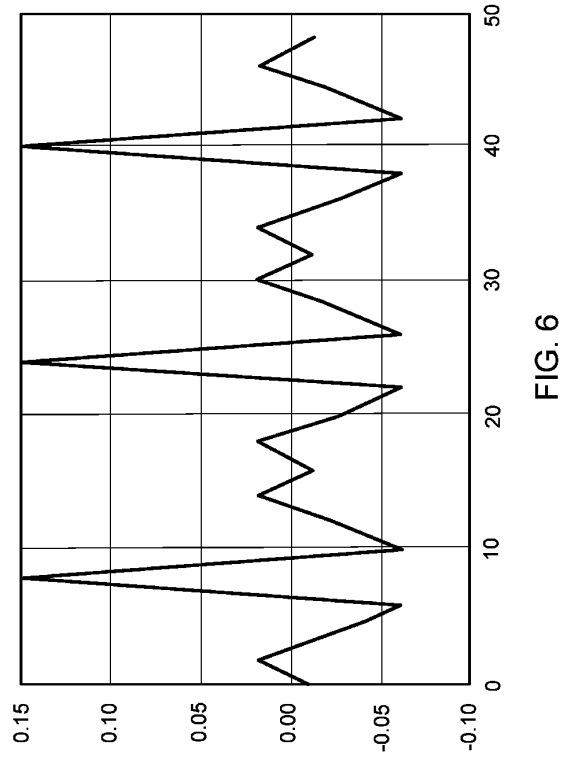
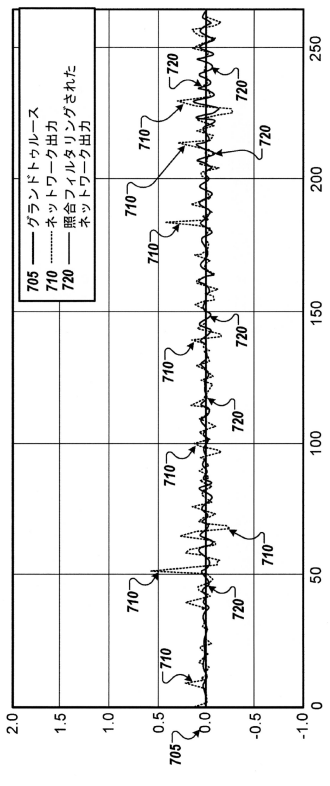


FIG. 6

【図７】



【図８】

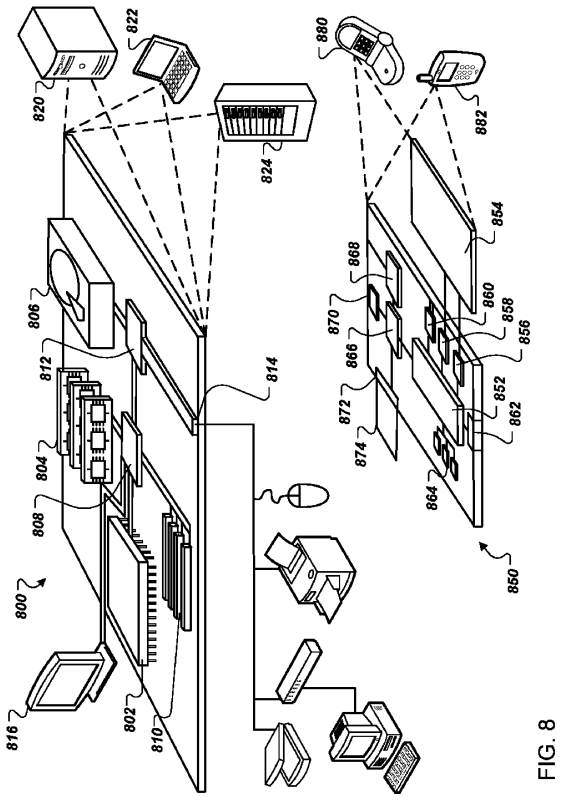


FIG. 8

フロントページの続き

(33)優先権主張国・地域又は機関
米国(US)
(72)発明者 アレクサンダー・エイチ・グルエンシュタイン
アメリカ合衆国・カリフォルニア・９４０４３・マウンテン・ビュー・アンフィシアター・パークウェイ・１６００
(72)発明者 タラル・プラディープ・ジョグレッカー
アメリカ合衆国・カリフォルニア・９４０４３・マウンテン・ビュー・アンフィシアター・パークウェイ・１６００
(72)発明者 ビジャヤディティヤ・ペディンチ
アメリカ合衆国・カリフォルニア・９４０４３・マウンテン・ビュー・アンフィシアター・パークウェイ・１６００
(72)発明者 ミヒール・エー・ユー・バッキアニ
アメリカ合衆国・カリフォルニア・９４０４３・マウンテン・ビュー・アンフィシアター・パークウェイ・１６００

審査官 菊池 智紀
(56)参考文献 米国特許第０９５４８０５３（ＵＳ，Ｂ１）
特開２０１０－１６４９９２（ＪＰ，Ａ）
特表２０２０－５２６７８１（ＪＰ，Ａ）
米国特許出願公開第２０１８／０１３０４６９（ＵＳ，Ａ１）
国際公開第２０１４／１１２１１０（WO，Ａ１）
(58)調査した分野 (Int.Cl.，ＤＢ名)
Ｇ１０Ｌ １５／００－１５／３４，１９／０１８
ＩＥＥＥ Ｘｐｌｏｒｅ