

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6581174号
(P6581174)

(45) 発行日 令和1年9月25日 (2019.9.25)

(24) 登録日 令和1年9月6日 (2019.9.6)

(51) Int. Cl.

F I

G 0 6 F 3/16 (2006.01)

G 1 0 L 15/28 (2013.01)

G 1 0 L 15/24 (2013.01)

G 1 0 L 15/10 (2006.01)

H 0 4 M 11/00 (2006.01)

G 0 6 F 3/16 6 5 0

G 1 0 L 15/28 2 3 0 K

G 1 0 L 15/24 Z

G 1 0 L 15/10 2 0 0 W

H 0 4 M 11/00 3 0 2

請求項の数 20 外国語出願 (全 45 頁) 最終頁に続く

(21) 出願番号 特願2017-250005 (P2017-250005)
 (22) 出願日 平成29年12月26日 (2017.12.26)
 (62) 分割の表示 特願2015-557147 (P2015-557147)
 の分割
 原出願日 平成26年2月7日 (2014.2.7)
 (65) 公開番号 特開2018-109980 (P2018-109980A)
 (43) 公開日 平成30年7月12日 (2018.7.12)
 審査請求日 平成30年1月18日 (2018.1.18)
 (31) 優先権主張番号 61/762,260
 (32) 優先日 平成25年2月7日 (2013.2.7)
 (33) 優先権主張国・地域又は機関
 米国 (US)

(73) 特許権者 503260918
 アップル インコーポレイテッド
 Apple Inc.
 アメリカ合衆国 95014 カリフォル
 ニア州 クパチーノ アップル パーク
 ウェイ ワン
 One Apple Park Way,
 Cupertino, Californ
 ia 95014, U. S. A.
 (74) 代理人 100076428
 弁理士 大塚 康徳
 (74) 代理人 100115071
 弁理士 大塚 康弘
 (74) 代理人 100112508
 弁理士 高柳 司郎

最終頁に続く

(54) 【発明の名称】 デジタルアシスタントのためのボイストリガ

(57) 【特許請求の範囲】

【請求項 1】

1つ以上のプロセッサと、前記1つ以上のプロセッサによる実行のための命令を記憶するメモリと、を含む電子機器にて実行される、ボイストリガを動作させるための方法であって、

音入力を受け取る工程と、

前記音入力の少なくとも一部が所定の種類の音に対応するか判定する工程と、

前記音入力の少なくとも一部が前記所定の種類に対応することを判定すると、前記音入力の入力レプリゼンテーションを1つ以上のリファレンスレプリゼンテーションと比較することに基づいて前記音入力が入力レプリゼンテーションと一致するか判定する工程であって、前記入力レプリゼンテーションは前記音入力のオーディオシグニチャを表現する、工程と、

前記音入力が入力レプリゼンテーションと一致することを判定すると、スピーチベースのサービスを開始する工程と、

前記電子機器の少なくとも前面で検出された光量に基づいて、前記スピーチベースのサービスを開始する前記ボイストリガの動作モードを判定する工程と、
 を含むことを特徴とする方法。

【請求項 2】

前記音入力が入力レプリゼンテーションと一致するか判定する前記工程が第1の音検出器によって行われ、前記音入力が入力レプリゼンテーションと一致するか判定する前記工程が第2の音検出器によって行われ、前記第1の音検出器が、前記第2の音検出器よりも少ない電力を動作中に消費する

10

20

ことを特徴とする、請求項 1 に記載の方法。

【請求項 3】

前記第 1 の音検出器による、前記音入力の前記所定の種類に対応することの判定に応じて前記第 2 の音検出器が起動されることを特徴とする、請求項 2 に記載の方法。

【請求項 4】

前記第 1 の音検出器による、前記音入力の前記所定の種類に対応することの判定後、前記第 2 の音検出器が少なくとも所定の時間動作されることを特徴とする、請求項 2 に記載の方法。

【請求項 5】

前記所定の種類が人の声であり、前記所定の内容が 1 つ以上の単語であることを特徴とする、請求項 1 に記載の方法。

10

【請求項 6】

前記所定の内容が、1 つ以上の所定の音素であることを特徴とする、請求項 1 に記載の方法。

【請求項 7】

前記 1 つ以上の所定の音素が、少なくとも 1 つの単語を構成することを特徴とする、請求項 6 に記載の方法。

【請求項 8】

前記音入力の前記所定の種類の音に対応するか判定する前に、前記音入力の前記条件を満たすか判定する工程を更に含むことを特徴とする、請求項 2 に記載の方法。

20

【請求項 9】

前記所定の条件が、振幅閾値であることを特徴とする、請求項 8 に記載の方法。

【請求項 10】

前記音入力の前記条件を満たすか判定する前記工程が第 3 の音検出器によって行われ、前記第 3 の音検出器が、前記第 1 の音検出器よりも少ない電力を作動中に消費することを特徴とする、請求項 8 に記載の方法。

【請求項 11】

前記音入力の少なくとも一部をメモリ内に記憶する工程と、

前記スピーチベースのサービスが開始されると、前記音入力の前記一部を前記スピーチベースのサービスに提供する工程と、
を更に含むことを特徴とする、請求項 1 に記載の方法。

30

【請求項 12】

前記音入力の前記特定ユーザの声に対応するか判定する工程を更に含むことを特徴とする、請求項 1 に記載の方法。

【請求項 13】

前記音入力の前記所定の内容を含むこと及び前記音入力の前記特定ユーザの前記声に対応することを判定すると、前記スピーチベースのサービスが開始されることを特徴とする、請求項 12 に記載の方法。

【請求項 14】

前記音入力の前記所定の内容を含むこと及び前記音入力の前記特定ユーザの前記声に対応しないことを判定すると、前記スピーチベースのサービスが限定アクセスモードにて開始されることを特徴とする、請求項 13 に記載の方法。

40

【請求項 15】

前記音入力の前記特定ユーザの前記声に対応することを判定すると、前記特定ユーザの名前を含む音声プロンプトを出力する工程を更に含むことを特徴とする、請求項 13 に記載の方法。

【請求項 16】

前記電子機器が所定の向きにあるか判定する工程と、

前記電子機器が前記所定の向きにあることを判定すると、前記ボイストリガの所定のモードを有効にする工程と、

50

を更に含むことを特徴とする、請求項 1 に記載の方法。

【請求項 17】

電子機器の 1 つ以上のプロセッサによる実行のためのプログラムであって、
音入力を受け取るための命令と、

前記音入力の少なくとも一部が所定の種類の音に対応するか判定するための命令と、

前記音入力の少なくとも一部が前記所定の種類に対応することを判定すると、前記音入力の入力レプリゼンテーションを 1 つ以上のリファレンスレプリゼンテーションと比較することに基づいて前記音入力が入力の内容を含むか判定するための命令であって、前記入力レプリゼンテーションは前記音入力のオーディオシグニチャを表現する、命令と、

前記音入力が入力の内容を含むことを判定すると、スピーチベースのサービスを開始するための命令と、

前記電子機器の少なくとも前面で検出された光量に基づいて、前記スピーチベースのサービスを開始するボイストリガの動作モードを判定するための命令と、

を含むことを特徴とする、プログラム。

【請求項 18】

1 つ以上のプロセッサと、

メモリと、

1 つ以上のプログラムと、を備える電子機器であって、前記 1 つ以上のプログラムが、前記メモリに記憶されており、前記 1 つ以上のプロセッサによって実行されるように構成されており、前記 1 つ以上のプログラムが、

音入力を受け取るための命令と、

前記音入力の少なくとも一部が所定の種類の音に対応するか判定するための命令と、

前記音入力の少なくとも一部が前記所定の種類に対応することを判定すると、前記音入力の入力レプリゼンテーションを 1 つ以上のリファレンスレプリゼンテーションと比較することに基づいて前記音入力が入力の内容を含むか判定するための命令であって、前記入力レプリゼンテーションは前記音入力のオーディオシグニチャを表現する、命令と、

前記音入力が入力の内容を含むことを判定すると、スピーチベースのサービスを開始するための命令と、

前記電子機器の少なくとも前面で検出された光量に基づいて、前記スピーチベースのサービスを開始するボイストリガの動作モードを判定するための命令と、

を含むことを特徴とする、電子機器。

【請求項 19】

前記所定の種類が人の声であり、前記所定の内容が 1 つ以上の単語であることを特徴とする、請求項 18 に記載の電子機器。

【請求項 20】

前記 1 つ以上のプログラムが、前記音入力が入力の種類に対応するか判定する前に、前記音入力が入力の内容を満たすか判定するための命令を更に含むことを特徴とする、請求項 18 に記載の電子機器。

【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、名称が「VOICE TRIGGER FOR A DIGITAL ASSISTANT」であり、2013 年 2 月 7 日に本出願の米国仮出願第 61 / 762 , 260 号の利益を主張し、すべての目的に対してそのエンティティに、参照によりこれにて組み込まれている。

【0002】

< 技術分野 >

開示された実施態様は、全般的にはデジタルアシスタントに関連し、より具体的には、デジタルアシスタントのためのボイストリガのための方法及びシステムに関連する。

10

20

30

40

50

【背景技術】

【0003】

近年、アップル社（Apple）のSIRI（登録商標）などの、ウェブの検索及びナビゲーションなどの様々なタスクを扱うための、音声ベースのデジタルアシスタントが市場に導入されている。そのような音声ベースのデジタルアシスタントの1つの利点は、ユーザが、機器を操作又は視認することなく、ハンズフリーの状態では機器と双方向作用できることである。ハンズフリー操作は特に、運転中などの、人が機器を物理的に操作できない、又はすべきでない際に有用となり得る。しかし、音声ベースのアシスタントを開始するには、ユーザは一般的にボタンを押すか、又はタッチスクリーン上のアイコンを選択する必要がある。この触覚による入力、ハンズフリー体験を阻害する。それに応じて、触覚による入力ではなく、音声入力又は信号を使用して、音声ベースのデジタルアシスタント（又は、他のスピーチベースのサービス）を有効にする方法及びシステムを提供することが有利となる。

10

【0004】

音声入力を使用して音声ベースのアシスタントを有効にすることは、音声入力を検出するためにオーディオチャネルを監視する必要がある。この監視は、そのような音声ベースのデジタルアシスタントが多くの場合に実行される、バッテリーに依存するハンドヘルド機器又はポータブル装置上の限られたリソースである電力を消費する。したがって、機器上の音声ベースのサービス及び／又はスピーチベースのサービスを開始するために使用され得る、エネルギー効率の良いボイストリガを提供することは有益となるであろう。

20

【発明の概要】

【0005】

それに応じて、限られた電力リソースを過剰に消費することなく、「常時リスニング」式のボイストリガ機能を提供できる低電力ボイストリガが必要とされている。以下に説明する実施態様は、電子機器にてボイストリガを使用して、音声ベースのアシスタントを開始するためのシステム及び方法を提供する。音声ベースのデジタルアシスタント（又は、音声テキスト化書き換えサービスなどの、他のスピーチベースのサービス）との双方向作用は多くの場合、デジタルアシスタントを有効にするために、ユーザが機器上のアフオーダンス（例えば、ボタン又はアイコン）を押すと開始し、続いて機器は、光、音（例えば、ピープ）、又は発声出力（例えば、「ご用件は何ですか？」）などの、デジタルアシスタントが有効でリスンしていることの、ユーザへのいくらかの表示を提供する。本明細書に説明するように、ボイストリガはまた、ユーザによる物理的な双方向作用を必要とすることなく、特定かつ所定の単語、フレーズ、又は音に応じて有効となるよう実装され得る。例えば、ユーザは、フレーズ「SIRIへ」と呼びかけることで、IPHONE（登録商標）のSIRIデジタルアシスタント（双方とも、本出願の譲受人であるアップル社（Apple Inc.）より提供されている）を有効にできてもよい。これに応じて、機器はピープ、音、又はスピーチ出力を発し（例えば、「ご用件は何ですか？」）、ユーザにリスニングモードが有効であることを示す。それに応じて、ユーザはデジタルアシスタント機能を提供する機器を物理的にタッチする必要なく、デジタルアシスタントとの双方向作用を開始できる。

30

40

【0006】

ボイストリガを伴うスピーチベースのサービスを開始するための1つの技術は、スピーチベースのサービスに、所定のトリガ語、フレーズ、又は音（これらのいずれかは本明細書にて「トリガ音」と呼ばれもよい）を継続的にリスンさせるものである。しかし、スピーチベースのサービス（例えば、音声ベースのデジタルアシスタント）を継続して動作させることは、相当の音声処理及びバッテリー電力を必要とする。ボイストリガ機能を提供することでの電力消費を削減するため、種々の技術が採用されてもよい。いくつかの実施態様では、電子機器のメインプロセッサ（つまり、「アプリケーションプロセッサ」）は、（例えば、アプリケーションプロセッサに依存しないために）低電力消費の1つ以上の音検出器が有効に維持されている間に、低電力状態又は無電力状態に維持される。（低電

50

力状態又は無電力状態にある場合は、アプリケーションプロセッサ若しくは任意の他のプロセッサ、プログラム、又はモジュールは、無効又はスタンバイモードにあるものとして説明されてもよい。) 例えば、アプリケーションプロセッサが無効であっても、トリガ音のためにオーディオチャネルを監視するよう、低電力音検出器が使用されている。この音検出器は時に、本明細書にてトリガ音検出器と呼ばれる。いくつかの実施態様では、特定の音、音素、及び/又は単語を検出するよう構成されている。トリガ音検出器(ハードウェア構成要素及び/又はソフトウェア構成要素を含む)は、特有の単語、音、又はフレーズを認識するよう設計されているが、そのようなタスクは大きな計算リソース及び電力リソースを必要とするため、完全な音声テキスト化機能を一般的に提供できないか、又は最適化されていない。したがって、いくつかの実施態様では、トリガ音検出器は、音声入力

10

【0007】

いくつかの実施態様では、トリガ音検出器は、一式の単語、フレーズ、音、及び/又はそれらの組み合わせなどの、種々の異なるトリガ音を検出するよう構成されている。ユーザは続いて、これらの音のいずれかを使用して、スピーチベースのサービスを開始できる。一例では、ボイストリガは、フレーズ「S I R I へ」、「S I R I 起動」、「デジタルアシスタント呼び出し」、又は「ハロー、H A L、聞こえますか、H A L?」に応答するよう予め構成されている。いくつかの実施態様では、ユーザは、予め構成されたトリガ音の1つを、単一のトリガ音として選択する必要がある。いくつかの実施態様では、ユーザが異なるトリガ音でスピーチベースのサービスを開始できるよう、ユーザは、予め構成されたトリガ音のサブセットを選択する。いくつかの実施態様では、予め構成されたトリガ音のすべては、正当なトリガ音に維持される。

20

【0008】

いくつかの実施態様では、別の音検出器が使用され、トリガ音検出器は、多くの時間に低電力モード又は無電力モードにも維持され得る。例えば、異なる種類の音検出器(例えば、トリガ音検出器よりも少ない電力を使用するもの)を使用して、オーディオチャネルを監視し、音入力特定の種類の音に対応するか判定する。音は、特定の識別可能な音の特性に基づいて、異なる「種類」に分類されている。例えば、「人の声」の種類に属する音は、特定のスペクトルコンテンツ、周期性、基本周波数、等を有する。他の種類の音(例えば、口笛、拍手、等)は、異なる特性を有する。異なる種類の音は、本明細書に説明するように、音声処理技術及び/又は信号処理技術を用いて識別されている。この音検出器は時に、本明細書にて「音種類検出器」と呼ばれる。例えば、所定のトリガフレーズが「S I R I へ」の場合、音種類検出器は、入力人が人の話し声におおよそ対応するか判定する。トリガ音が、口笛などの非有声音の場合、音種類検出器は、音入力人が口笛におおよそ対応するか判定する。適切な種類の音を検出されると、音種類検出器は、トリガ音検出器を起動し、音を更に処理及び/又は分析する。音種類検出器は、トリガ音検出器よりも少ない電力を必要とするため(例えば、必要電力が低い回路及び/又はトリガ音検出器よりも効率の良いオーディオ処理アルゴリズムを使用するため)、ボイストリガ機能は、トリガ音検出器単独よりも少ない電力を消費する。

30

40

【0009】

いくつかの実施態様では、更に別の音検出器が使用され、上記の音種類検出器及びトリガ音検出器の双方は、多くの時間に低電力モード又は無電力モードに維持され得る。例えば、音種類検出器よりも少ない電力を使用する音検出器は、オーディオチャネルの監視に使用され、音入力人が、振幅閾値(例えば、音量)などの、所定の条件を満たすか判定する。この音検出器は、本明細書にてノイズ検出器とも呼ばれてもよい。ノイズ検出器が、所定の閾値を満たす音を検出すると、ノイズ検出器は音種類検出器を起動し、音を更に処理及び/又は分析する。ノイズ検出器は、音種類検出器又はトリガ音検出器よりも少ない電

50

力を要求するため（例えば、要求電力が低い回路及び／又は効率の良いオーディオ処理アルゴリズムを使用するため）、ボイストリガ機能は、ノイズ検出器を伴わない音種類検出器及びトリガ音検出器の組み合わせよりも少ない電力を消費する。

【0010】

いくつかの実施態様では、上記の音検出器のうちの任意の1つ以上は、「オン」と「オフ」との間の状態をサイクルするデューティサイクルに従って動作されている。これは、ボイストリガの電力消費を削減することを更に助ける。例えば、いくつかの実施態様では、ノイズ検出器は10ミリ秒間「オン」となり（つまり、オーディオチャネルを積極的に監視し）、続いて90ミリ秒間「オフ」となる。このように、継続的なノイズ検出機能を依然として効果的に提供する一方で、ノイズ検出器は90%の時間「オフ」となる。いくつかの実施態様では、音検出器に対するオン及びオフの各持続時間は、トリガ音が依然として入力されている間に、検出器のすべてが有効となっているよう選択されている。例えば、「SIRIへ」のトリガフレーズについては、音検出器は、トリガフレーズがデューティサイクル（単数又は複数）のどこで始まっても、トリガ音検出器が遅れずに有効となり、十分な量の入力を分析するよう構成されてもよい。例えば、トリガ音検出器は遅れずに有効となり、音がトリガフレーズに一致することを判定するように十分な音「IRIへ」を受け取り、処理し、及び分析する。いくつかの実施態様では、音入力は受け取った状態でメモリに記憶され、音入力の大半の部分が分析され得るよう、上流の検出器（upstream detector）に渡される。それに応じて、トリガフレーズが発言された後までトリガ音検出器が起動されていなくても、記録されたトリガフレーズの全体を依然として分析できる。

【0011】

いくつかの実施態様は、ボイストリガを動作させる方法を提供する。この方法は、1つ以上のプロセッサ及び、1つ以上のプロセッサによって実行される命令を記憶するメモリを含む電子機器にて実行されている。この方法は、音入力を受け取ることを含む。この方法は、音入力の少なくとも一部が、所定の種類の音に対応するか判定することを更に含む。この方法は、音入力の少なくとも一部が、所定の種類に対応することを判定すると、音入力が所定の内容を含むか判定することを更に含む。この方法は、音入力が所定の内容を含むことを判定すると、スピーチベースのサービスを開始することを更に含む。いくつかの実施態様では、スピーチベースのサービスは、音声ベースのデジタルアシスタントである。いくつかの実施態様では、スピーチベースのサービスは、口述サービスである。

【0012】

いくつかの実施態様では、音入力が所定の種類の音に対応するか判定することが、第1の音検出器によって行われ、音入力が所定の内容を含むか判定することが、第2の音検出器によって行われる。いくつかの実施態様では、第1の音検出器は、第2の音検出器よりも少ない電力を作動中に消費する。いくつかの実施態様では、第1の音検出器は、音入力の周波数領域分析を行う。いくつかの実施態様では、音入力が所定の種類の音に対応するか判定することが、音入力が所定の条件を満たすことを判定すると行われる（例えば、以下に説明する、第3の音検出器により判定されるように）。

【0013】

いくつかの実施態様では、第1の音検出器は、デューティサイクルに従って、オーディオチャネルを定期的に監視する。いくつかの実施態様では、デューティサイクルは、約20ミリ秒のオンタイム、及び約100ミリ秒のオフタイムを含む。

【0014】

いくつかの実施態様では、所定の種類は人の声であり、所定の内容は1つ以上の単語である。いくつかの実施態様では、音入力の少なくとも一部が所定の種類の音に対応するか判定することが、音入力の少なくとも一部が人の声の周波数特性を含むか判定することを含む。

【0015】

いくつかの実施態様では、第2の音検出器は、音入力が所定の種類に対応することの、

第1の音検出器による判定に応じて起動されている。いくつかの実施態様では、第2の音検出器は、音入力が所定の種類に対応することの、第1の音検出器による判定後、少なくとも所定の時間動作されている。いくつかの実施態様では、所定の時間は、所定の内容の持続時間に対応する。

【0016】

いくつかの実施態様では、所定の内容は、1つ以上の所定の音素である。いくつかの実施態様では、1つ以上の所定の音素は、少なくとも1つの単語を構成する。

【0017】

いくつかの実施態様では、この方法は、音入力が所定の種類の音に対応するか判定する前に、音入力が所定の条件を満たすか判定することを含む。いくつかの実施態様では、所定の条件は振幅閾値である。いくつかの実施態様では、音入力が所定の条件を満たすか判定することは、第3の音検出器によって行われ、第3の音検出器は、第1の音検出器よりも少ない電力を作動中に消費する。いくつかの実施態様では、第3の音検出器は、デューティサイクルに従って、オーディオチャネルを定期的に監視する。いくつかの実施態様では、デューティサイクルは、約20ミリ秒のオンタイム、及び約500ミリ秒のオフタイムを含む。いくつかの実施態様では、第3の音検出器は、音入力の時間領域分析を行う。

【0018】

いくつかの実施態様では、この方法は、音入力の少なくとも一部をメモリに記憶すること、及び、スピーチベースのサービスが開始されると、音入力の一部を、スピーチベースのサービスに提供することを含む。いくつかの実施態様では、音入力の一部は、ダイレクトメモリアクセスを使用してメモリに記憶されている。

【0019】

いくつかの実施態様では、この方法は、音入力 that 特定ユーザの声に対応するか判定することを含む。いくつかの実施態様では、スピーチベースのサービスは、音入力 that 所定の内容を含むこと、及び音入力 that 特定ユーザの声に対応することを判定すると開始されている。いくつかの実施態様では、スピーチベースのサービスは、音入力 that 所定の内容を含むこと、及び音入力 that 特定ユーザの声に対応しないことを判定すると、限定アクセスモードにて開始されている。いくつかの実施態様では、この方法は、音入力 that 特定ユーザの声に対応することを判定すると、特定ユーザの名前を含む音声プロンプトを出力することを含む。

【0020】

いくつかの実施態様では、音入力 that 所定の内容を含むか判定することが、音入力のレプリゼンテーションをリファレンスレプリゼンテーションに比較すること、及び音入力のレプリゼンテーションがリファレンスレプリゼンテーションと一致する場合に、音入力 that 所定の内容を含むことを判定することを含む。いくつかの実施態様では、一致は、音入力のレプリゼンテーションが、所定の信頼値でリファレンスレプリゼンテーションに一致する場合に判定されている。いくつかの実施態様では、この方法は、この音入力を含む複数の音入力を受け取ること、及び、各音入力 that 所定の内容を含むことを判定することに応じて、複数の音入力の各1つを使用して、リファレンスレプリゼンテーションを繰り返し調整することを含む。

【0021】

いくつかの実施態様では、この方法は、電子機器が所定の向きにあるか判定すること、及び、電子機器が所定の向きにあることを判定すると、ボイストリガの所定のモードを有効にすることを含む。いくつかの実施態様では、所定の向きは、機器のディスプレイスクリーンが実質的に水平で下向きにあることに対応し、所定のモードはスタンバイモードである。いくつかの実施態様では、所定の向きは、機器のディスプレイスクリーンが実質的に水平で上向きにあることに対応し、所定のモードはリスニングモードである。

【0022】

いくつかの実施態様は、ボイストリガを動作させる方法を提供する。この方法は、1つ以上のプロセッサ及び、1つ以上のプロセッサによって実行される命令を記憶するメモリ

10

20

30

40

50

を含む電子機器にて実行されている。この方法は、ボイストリガを第１のモードにて動作させることを含む。この方法は、電子機器のマイクロフォン及びカメラのうちの１つ以上が塞がっていることを検出することで、電子機器が実質的な閉鎖空間内にあるか判定することを更に含む。この方法は、電子機器が実質的な閉鎖空間内にあることを判定すると、ボイストリガを第２のモードに切り換えることを更に含む。いくつかの実施態様では、第２のモードは、スタンバイモードである。

【００２３】

いくつかの実施態様は、ボイストリガを動作させる方法を提供する。この方法は、１つ以上のプロセッサ及び、１つ以上のプロセッサによって実行される命令を記憶するメモリを含む電子機器にて実行されている。この方法は、電子機器が所定の向きにあるか判定すること、及び、電子機器が所定の向きにあることを判定すると、ボイストリガの所定のモードを有効にすることを含む。いくつかの実施態様では、所定の向きは、機器のディスプレイスクリーンが実質的に水平で下向きにあることに対応し、所定のモードはスタンバイモードである。いくつかの実施態様では、所定の向きは、機器のディスプレイスクリーンが実質的に水平で上向きにあることに対応し、所定のモードはリスニングモードである。

【００２４】

いくつかの実施態様によると、電子機器は、音入力を受け取るよう構成された受音ユニットと、受音ユニットに連結された処理ユニットと、を含む。この処理ユニットは、音入力の少なくとも一部が所定の種類の音に対応するか判定し、音入力の少なくとも一部が所定の種類に対応することを判定すると、音入力が所定の内容を含むか判定し、音入力が所定の内容を含むことを判定すると、スピーチベースのサービスを開始するよう構成されている。いくつかの実施態様では、処理ユニットは、音入力が所定の種類の音に対応するか判定する前に、音入力が所定の条件を満たすか判定するよう更に構成されている。いくつかの実施態様では、処理ユニットは、音入力が特定ユーザの声に対応するか判定するよう更に構成されている。

【００２５】

いくつかの実施態様によると、電子機器は、ボイストリガを、複数のモードの第１のモードにて動作させるよう構成されているボイストリガユニットと、ボイストリガユニットに連結された処理ユニットと、を含む。いくつかの実施態様では、処理ユニットは、電子機器のマイクロフォン及びカメラのうちの１つ以上が塞がっていることを検出することで、電子機器が実質的な閉鎖空間内にあるか判定し、電子機器が実質的な閉鎖空間内にあることを判定すると、ボイストリガを第２のモードに切り換えるよう構成されている。いくつかの実施態様では、処理ユニットは、電子機器が所定の向きにあるか判定し、電子機器が所定の向きにあることを判定すると、ボイストリガの所定のモードを有効にするよう構成されている。

【００２６】

いくつかの実施態様によると、コンピュータ可読記憶媒体（例えば、永続的コンピュータ可読記憶媒体）が提供されており、このコンピュータ可読記憶媒体は、電子機器の１つ以上のプロセッサによって実行される１つ以上のプログラムを記憶し、この１つ以上のプログラムは、本明細書に記載の方法のいずれかを行う命令を含む。

【００２７】

いくつかの実施態様によると、本明細書に記載のいずれかの方法を行うための手段を含む電子機器（例えば、ポータブル電子機器）が提供されている。

【００２８】

いくつかの実施態様によると、本明細書に記載のいずれかの方法を行うよう構成された処理ユニットを含む電子機器（例えば、ポータブル電子機器）が提供されている。

【００２９】

いくつかの実施態様によると、１つ以上のプロセッサ及び、１つ以上のプロセッサによって実行される１つ以上のプログラムを記憶するメモリを含み、この１つ以上のプログラムは、本明細書に記載のいずれかの方法を行う命令を含む電子機器（例えば、ポータブル

10

20

30

40

50

電子機器)が提供されている。

【0030】

いくつかの実施態様によると、電子機器内にて使用するための情報処理装置が提供されており、この情報処理装置は、本明細書に記載の方法のいずれかを行うための手段を含む。

【図面の簡単な説明】

【0031】

【図1】いくつかの実施態様に係る、デジタルアシスタントが動作する環境を示すブロック図である。

【0032】

【図2】いくつかの実施態様に係る、デジタルアシスタントクライアントシステムを示すブロック図である。

【0033】

【図3A】いくつかの実施態様に係る、スタンドアロンのデジタルアシスタントシステム又はデジタルアシスタントサーバシステムを示すブロック図である。

【0034】

【図3B】いくつかの実施態様に係る、図3Aに示すデジタルアシスタントの機能を示すブロック図である。

【0035】

【図3C】いくつかの実施態様に係る、オントロジの一部を示すネットワーク図である。

【0036】

【図4】いくつかの実施態様に係る、ボイストリガシステムの構成要素を示すブロック図である。

【0037】

【図5】いくつかの実施態様に係る、ボイストリガシステムを動作させるための方法を示すフローチャートである。

【図6】いくつかの実施態様に係る、ボイストリガシステムを動作させるための方法を示すフローチャートである。

【図7】いくつかの実施態様に係る、ボイストリガシステムを動作させるための方法を示すフローチャートである。

【0038】

【図8】いくつかの実施形態に係る、電子機器の機能ブロック図である。

【図9】いくつかの実施形態に係る、電子機器の機能ブロック図である。

【0039】

類似の参照番号は、図面全体を通して対応する部分を指す。

【発明を実施するための形態】

【0040】

図1は、いくつかの実施態様に係る、デジタルアシスタントの動作環境100のブロック図である。用語「デジタルアシスタント」、「仮想アシスタント」、「インテリジェント自動アシスタント」、「音声ベースのデジタルアシスタント」、又は「自動デジタルアシスタント」は、口頭及び/又はテキスト形式の自然言語入力を解釈してユーザ意図を推定し(例えば、自然言語入力に対応するタスクの種類を特定し)、推定されたユーザ意図に基づきアクションを実行する(例えば、特定されたタスクの種類に対応するタスクを実行する)任意の情報処理システムを指す。例えば、推定されたユーザ意図に基づいてアクションするために、システムは、以下のうちの1つ以上を実行することができる。推定されたユーザ意図を果たすように設計されたステップ及びパラメータを有するタスクフローを特定すること(例えば、タスクの種類を特定すること)、推定されたユーザ意図からの具体的な要求をタスクフローに入力すること、プログラム、方法、サービス、API、又は同様のものを呼び出すことによってタスクフローを実行すること(例えば、サービスプロバイダに要求を送信すること)、並びに可聴(例えば、会話)及び/又は視覚形式のユ

10

20

30

40

50

ーザへの出力応答を生成することである。

【 0 0 4 1 】

具体的には、一度開始されると、デジタルアシスタントシステムは、自然言語命令、要求、陳述、叙述、及び／又は照会の形で少なくとも部分的にユーザ要求を受け入れることができる。全般的に、ユーザ要求は、デジタルアシスタントシステムによる、情報を提供する答え又はタスクの実行のどちらかを求めている。一般的に、ユーザ要求への満足な応答は、要求された情報回答の提供、要求されたタスクの実行のいずれか、又はその２つの組み合わせとなる。例えば、ユーザはデジタルアシスタントシステムに「私は今どこにいますか？」等の質問をする場合がある。ユーザの現在の場所に基づき、デジタルアシスタントは、「あなたはセントラルパーク内の西門の近くにいます」と回答する場合がある。ユーザはまた、例えば、「来週私のガールフレンドの誕生パーティーに私の友人を招待してほしい。」と述べることにより、タスクの実行を要求することができる。これに応じて、デジタルアシスタントは、「はい、すぐに。」との音声出力を生成することにより、要求を確認し、その後、ユーザの電子アドレス帳又は連絡先リストに記載されたユーザの友人のそれぞれに、ユーザの電子メールアドレスから、適切なカレンダー招待状を送信してもよい。情報又は様々なタスクの実行を要求するために、デジタルアシスタントと対話する方法は他にも数多く存在する。言葉による応答を提供し、プログラムされたアクションを取ることに加えて、デジタルアシスタントは、他の視覚又はオーディオ形式の応答を（例えば、テキスト、警報、音楽、ビデオ、アニメーション等として）提供することもできる。

10

20

【 0 0 4 2 】

図１に示すように、いくつかの実施態様では、デジタルアシスタントシステムはクライアント-サーバモデルにしたがって実装される。デジタルアシスタントシステムは、ユーザ機器（例えば、１０４ a 及び１０４ b ）上で実行されるクライアント側部分（例えば、１０２ a 及び１０２ b ）（以後、「デジタルアシスタント（ＤＡ）クライアント１０２」）、及びサーバシステム１０８上で実行されるサーバ側部分１０６（以後「デジタルアシスタント（ＤＡ）サーバ１０６」）を含む。ＤＡクライアント１０２は１つ以上のネットワーク１１０を通じてＤＡサーバ１０６と通信する。ＤＡクライアント１０２は、ユーザ対応入力及び出力処理、並びにＤＡサーバ１０６との通信等のクライアント側機能を提供する。ＤＡサーバ１０６は、それぞれのユーザ機器１０４（クライアント機器又は電子機器とも呼ばれる）上に各々常駐する任意の数のＤＡクライアント１０２のためにサーバ側機能を提供する。

30

【 0 0 4 3 】

いくつかの実施態様では、ＤＡサーバ１０６は、クライアント対応（facing）Ｉ／Ｏインターフェース１１２、１つ以上の処理モジュール１１４、データ及びモデル１１６、外部サービスへのＩ／Ｏインターフェース１１８、写真及びタグデータベース１３０、並びに写真タグモジュール１３２を含む。クライアント対応Ｉ／Ｏインターフェースはデジタルアシスタントサーバ１０６のためのクライアント対応入力及び出力処理を促進する。１つ以上の処理モジュール１１４は、データ及びモデル１１６を利用し、自然言語入力に基づきユーザの意図を判断し、推定されたユーザ意図に基づきタスクを実行する。写真及びタグデータベース１３０は、デジタル写真のフィンガープリント、及び、任意にデジタル写真自体、並びに、デジタル写真に関連付けられたタグを記憶する。写真タグモジュール１３２は、写真及び／又はフィンガープリントに関連してタグを作成し、記憶し、自動的に写真にタグ付けし、写真内の場所にタグをリンクする。

40

【 0 0 4 4 】

いくつかの実施態様では、ＤＡサーバ１０６は、タスク完了又は情報取得のために、ネットワーク（単数又は複数）１１０を通じて外部サービス１２０（例えば、ナビゲーションサービス（単数又は複数）１２２ - １、メッセージングサービス（単数又は複数）１２２ - ２、情報サービス（単数又は複数）１２２ - ３、カレンダーサービス１２２ - ４、電話サービス１２２ - ５、写真サービス（単数又は複数）１２２ - ６、等）と通信する。外部

50

サービスへの I / O インターフェース 118 はこのような通信を促進する。

【0045】

ユーザ機器 104 の例としては、限定するものではないが、ハンドヘルドコンピュータ、無線携帯情報端末 (personal digital assistant、PDA)、タブレットコンピュータ、ラップトップコンピュータ、デスクトップコンピュータ、セルラー電話、スマートホン、拡張型汎用パケット無線サービス (enhanced general packet radio service、EGPRS) 携帯電話、メディアプレーヤ、ナビゲーション機器、ゲームコンソール、テレビ、遠隔制御装置、又はこれらのデータ処理機器のうちの任意の 2 つ以上の組み合わせ、あるいは任意の他の好適なデータ処理機器が挙げられる。ユーザ機器 104 に関する更なる詳細が、図 2 に示される例示的なユーザ機器 104 に関して提供される。

10

【0046】

通信ネットワーク (単数又は複数) 110 の例としては、ローカルエリアネットワーク (LAN) 及び、例えば、インターネットのようなワイドエリアネットワーク (WAN) が挙げられる。通信ネットワーク (単数又は複数) 110 は、イーサネット (登録商標)、ユニバーサルシリアルバス (Universal Serial Bus、USB)、FIREWIRE (登録商標)、移動通信用のグローバルシステム (Global System for Mobile Communications、GSM (登録商標))、拡張データ GSM 環境 (Enhanced Data GSM Environment、EDGE)、符号分割多元接続 (code division multiple access、CDMA)、時分割多元接続 (time division multiple access、TDMA)、Bluetooth (登録商標)、Wi-Fi (登録商標)、ボイスオーバーインターネットプロトコル (voice over Internet Protocol、VoIP)、Wi-MAX (登録商標)、又は任意の他の好適な通信プロトコル等の、様々な有線又は無線プロトコルを含む、任意の周知のネットワークプロトコルを用いて実施することができる。

20

【0047】

サーバシステム 108 は、少なくとも 1 つのデータ処理装置、及び / 又はコンピュータの分散型ネットワーク上に実装することができる。いくつかの実施態様では、サーバシステム 108 は、また、様々な仮想機器及び / 又はサードパーティサービスプロバイダ (例えば、サードパーティクラウドサービスプロバイダ) のサービスを利用して、サーバシステム 108 の基本的なコンピューティングリソース及び / 又はインフラストラクチャリソースを提供する。

30

【0048】

図 1 に示されるデジタルアシスタントシステムは、クライアント側部分 (例えば、DA クライアント 102) 及びサーバ側部分 (例えば、DA サーバ 106) の両方を含むが、いくつかの実施態様では、デジタルアシスタントシステムはサーバ側部分 (例えば、DA サーバ 106) のみを指す。いくつかの実施態様では、デジタルアシスタントの機能は、ユーザ機器上にインストールされるスタンドアロンのアプリケーションとして実装することができる。加えて、デジタルアシスタントのクライアント部分とサーバ部分との間の機能の分配は実施態様によって異なり得る。例えば、いくつかの実施態様では、DA クライアント 102 は、ユーザ対応入力及び出力処理機能のみを提供し、デジタルアシスタントのすべての他の機能を DA サーバ 106 に委ねるシンクライアントである。例えば、いくつかの実施態様では、DA クライアント 102 は、DA サーバ 106 の 1 つ以上の機能を実行するか又は支援するように構成されている。

40

【0049】

図 2 は、いくつかの実施態様に係る、ユーザ機器 104 のブロック図である。ユーザ機器 104 は、メモリインターフェース 202、1 つ以上のプロセッサ 204、及び周辺インターフェース 206 を含む。ユーザ機器 104 内の様々な構成要素は、1 つ以上の通信バス又は信号線によって連結されている。ユーザ機器 104 は、周辺インターフェース 206 に連結されている様々なセンサ、サブシステム、及び周辺機器を含む。センサ、サブシステム、及び周辺機器は情報を収集し、及び / 又はユーザ機器 104 の様々な機能を促進する。

50

【 0 0 5 0 】

例えば、いくつかの実施態様では、動きセンサ 2 1 0（例えば、加速度計）、光センサ 2 1 2、GPS 受信機 2 1 3、温度センサ、及び近接センサ 2 1 4 が、向き、光、及び近接センシングの機能を促進するために、周辺インターフェース 2 0 6 に結合されている。いくつかの実施態様では、生体センサ、気圧計等などの他のセンサ 2 1 6 が、関連する機能を促進するために、周辺インターフェース 2 0 6 に接続されている。

【 0 0 5 1 】

いくつかの実施態様では、ユーザ機器 1 0 4 は、周辺インターフェース 2 0 6 に連結されたカメラサブシステム 2 2 0 を含む。いくつかの実施態様では、カメラサブシステム 2 2 0 の光学センサ 2 2 2 が、写真の撮影及びビデオクリップの録画等のカメラ機能を促進する。いくつかの実施態様では、ユーザ機器 1 0 4 は、通信機能を提供する 1 つ以上の有線及び／又は無線通信サブシステム 2 2 4 を含む。通信サブシステム 2 2 4 は通常、様々な通信ポート、無線周波数受信機及び送信機、並びに／又は光（例えば、赤外線）受信機及び送信機を含む。いくつかの実施態様では、ユーザ機器 1 0 4 は、音声認識、音声応答、デジタル録音、及び電話機能などの、音声使用可能機能を促進するために、1 つ以上のスピーカ 2 2 8 及び 1 つ以上のマイクロフォン 2 3 0 に連結されたオーディオサブシステム 2 2 6 を含む。いくつかの実施態様では、オーディオサブシステム 2 2 6 は、ボイストリガシステム 4 0 0 に連結されている。いくつかの実施態様では、ボイストリガシステム 4 0 0 及び／又はオーディオサブシステム 2 2 6 は、例えば、1 つ以上のアナログデジタル変換器、デジタル信号プロセッサ（DSP）、音検出器、メモリバッファ、コーデック、等を含む、音入力を受け取る及び／又は分析するための、（つまり、ハードウェア及び／又はソフトウェアを含む）低電力オーディオ回路及び／又はプログラムを含む。いくつかの実施態様では、低電力オーディオ回路は（単独又は、ユーザ機器 1 0 4 の他の構成要素に加えて）、音声ベースのデジタルアシスタント又は他のスピーチベースのサービスなどの、ユーザ機器 1 0 4 の 1 つ以上の態様に対する音声（又は音）トリガ機能を提供する。いくつかの実施態様では、低電力オーディオ回路は、プロセッサ（単数又は複数）2 0 4、I/O サブシステム 2 4 0、メモリ 2 5 0、等などの、ユーザ機器 1 0 4 の他の構成要素が停止していても及び／又はスタンバイモードにあっても、ボイストリガ機能を提供する。このボイストリガシステム 4 0 0 は、図 4 に関して更に詳細に説明されている。

【 0 0 5 2 】

いくつかの実施態様では、I/O サブシステム 2 4 0 も周辺インターフェース 2 0 6 に連結されている。いくつかの実施態様では、ユーザ機器 1 0 4 はタッチスクリーン 2 4 6 を含み、I/O サブシステム 2 4 0 は、タッチスクリーン 2 4 6 に連結されたタッチスクリーンコントローラ 2 4 2 を含む。ユーザ機器 1 0 4 がタッチスクリーン 2 4 6 及びタッチスクリーンコントローラ 2 4 2 を含む場合には、タッチスクリーン 2 4 6 及びタッチスクリーンコントローラ 2 4 2 は、通常、例えば、容量性、抵抗性、赤外線、表面超音波技術、近接センサアレイ、及び同様のもの等の複数のタッチ感知技術のいずれかを用いて、接触及び移動又はその中断を検出するように構成される。いくつかの実施態様では、ユーザ機器 1 0 4 は、タッチセンシティブ面を含まないディスプレイを含む。いくつかの実施態様では、ユーザ機器 1 0 4 は、分離したタッチセンシティブ面を含む。いくつかの実施態様では、ユーザ機器 1 0 4 はその他の入力コントローラ（単数又は複数）2 4 4 を含む。ユーザ機器 1 0 4 がその他の入力コントローラ（単数又は複数）2 4 4 を含む場合には、その他の入力コントローラ（単数又は複数）2 4 4 は通常、1 つ以上のボタン、ロックスイッチ、サムホイール、赤外線ポート、USB ポート、及び／又はスタイラス等のポイント入力機器等の、その他の入力／制御機器 2 4 8 に連結されている。

【 0 0 5 3 】

メモリインターフェース 2 0 2 は、メモリ 2 5 0 に連結される。いくつかの実施態様では、メモリ 2 5 0 は、高速ランダムアクセスメモリ及び／又は不揮発性メモリ（例えば、1 つ以上の磁気ディスク記憶機器、1 つ以上のフラッシュメモリ機器、1 つ以上の光学記憶機器、及び／又は他の不揮発性ソリッドステート記憶機器）などの、永続的コンピュー

10

20

30

40

50

タ可読媒体を含む。いくつかの実施態様では、メモリ 250 は、オペレーティングシステム 252、通信モジュール 254、グラフィカルユーザインターフェースモジュール 256、センサ処理モジュール 258、電話モジュール 260、及びアプリケーション 262、並びにそれらのサブセット又はスーパーセットを記憶する。オペレーティングシステム 252 は、基本システムサービス処理する命令、及びハードウェア依存タスクを実行する命令を含む。通信モジュール 254 は、1 つ以上の追加機器、1 つ以上のコンピュータ及び / 又は 1 つ以上のサーバとの通信を促進する。グラフィカルユーザインターフェースモジュール 256 はグラフィックユーザインターフェース処理を促進する。センサ処理モジュール 258 は、センサ関連の処理及び機能（例えば、1 つ以上のマイクロフォン 228 を用いて受け取った音声入力の処理）を促進する。電話モジュール 260 は電話関連のプロセス及び機能を促進する。アプリケーションモジュール 262 は、電子メッセージング、ウェブブラウジング、メディア処理、ナビゲーション、イメージング及び / 又はその他のプロセス及び機能等のユーザアプリケーションの様々な機能を促進する。いくつかの実施態様では、ユーザ機器 104 は、それぞれ外部サービスプロバイダの少なくとも 1 つに関連した 1 つ以上のソフトウェアアプリケーション 270 - 1 及び 270 - 2 をメモリ 250 内に記憶する。

10

【0054】

上述したように、いくつかの実施態様では、メモリ 250 はまた、デジタルアシスタントのクライアント側機能を提供するために、クライアント側のデジタルアシスタント命令（例えば、デジタルアシスタントクライアントモジュール 264 内の）、及び様々なユーザデータ 266（例えば、ユーザ固有の語彙データ、嗜好データ、及び / 又はユーザの電子アドレス帳若しくは連絡先リスト、to-do リスト、買い物リストなどのその他のデータ）も記憶する。

20

【0055】

いくつかの実施態様では、デジタルアシスタントクライアントモジュール 264 は、ユーザ機器 104 の様々なユーザインターフェース（例えば、I/O サブシステム 244）を通じて音声入力、テキスト入力、タッチ入力、及び / 又はジェスチャ入力を受け入れることができる。デジタルアシスタントクライアントモジュール 264 はまた、オーディオ、視覚、及び / 又は触覚形態の出力を提供することもできる。例えば、出力は、音声、音響、警報、テキストメッセージ、メニュー、グラフィック、ビデオ、アニメーション、振動、及び / 又は上記のもののうちの 2 つ以上の組み合わせとして提供することができる。動作時、デジタルアシスタントクライアントモジュール 264 は、通信サブシステム 224 を用いて、デジタルアシスタントサーバ（例えば、デジタルアシスタントサーバ 106、図 1）と通信する。

30

【0056】

いくつかの実施態様では、デジタルアシスタントクライアントモジュール 264 は、ユーザ入力に関連付けられるコンテキストを確立するために、様々なセンサ、サブシステム及び周辺機器を利用してユーザ機器 104 の周囲環境から追加情報を収集する。いくつかの実施態様では、デジタルアシスタントクライアントモジュール 264 は、ユーザの意図の推定を助けるために、ユーザ入力とともにコンテキスト情報又はそのサブセットをデジタルアシスタントサーバ（例えば、デジタルアシスタントサーバ 106、図 1）に提供する。

40

【0057】

いくつかの実施態様では、ユーザ入力に伴い得るコンテキスト情報は、センサ情報、例えば、周囲環境の照明、周囲雑音、周囲温度、画像又は映像などを含む。いくつかの実施態様では、コンテキスト情報は、また、機器の物理状態、例えば、機器の向き、機器の場所、機器温度、電力レベル、速度、加速度、運動パターン、セルラー信号強度などを含む。いくつかの実施態様では、ユーザ機器 106 のソフトウェア状態に関係する情報、例えば、ユーザ機器 104 の稼働中のプロセス、インストール済みのプログラム、過去と現在のネットワーク活動、バックグラウンドサービス、エラーログ、リソース使用量なども、

50

ユーザ入力に関連したコンテキスト情報として、デジタルアシスタントサーバ（例えば、デジタルアシスタントサーバ１０６、図１）に提供される。

【００５８】

いくつかの実施態様では、ＤＡクライアントモジュール２６４は、デジタルアシスタントサーバからの要求に応じて、ユーザ機器１０４上に記憶されている情報（例えば、ユーザデータ２６６の少なくとも一部）を選択的に提供する。いくつかの実施態様では、デジタルアシスタントクライアントモジュール２６４はまた、デジタルアシスタントサーバ１０６（図１）による要求に応じて、自然言語ダイアログ又はその他のユーザインターフェースを介して、ユーザから追加入力も引き出す。デジタルアシスタントクライアントモジュール２６４は、ユーザ要求に表されるユーザ意図の推定及び／又はユーザ意図の達成に向けて、デジタルアシスタントサーバ１０６を支援するために、追加入力をデジタルアシスタントサーバ１０６に渡す。

10

【００５９】

いくつかの実施態様では、メモリ２５０は追加の命令又はより少数の命令を含んでもよい。更に、ユーザ機器１０４の様々な機能は、１つ以上の信号処理及び／又は特定用途向け集積回路の形態を含む、ハードウェアの形態及び／又はファームウェアの形態で実装されてもよく、それゆえ、ユーザ機器１０４は、図２に示されているすべてのモジュール及びアプリケーションを含む必要はない。

【００６０】

図３Ａは、いくつかの実施態様に係る、例示的なデジタルアシスタントシステム３００（デジタルアシスタントとも呼ばれる）のブロック図である。いくつかの実施態様では、デジタルアシスタントシステム３００はスタンドアロンのコンピュータシステム上に実装される。いくつかの実施態様では、デジタルアシスタントシステム３００は複数のコンピュータにわたって分散される。いくつかの実施態様では、デジタルアシスタントのモジュール及び機能のいくつかはサーバ部分及びクライアント部分に分割される。クライアント部分はユーザ機器（例えば、ユーザ機器１０４）上に常駐し、例えば、図１に示されているように、１つ以上のネットワークを通じてサーバ部分（例えば、サーバシステム１０８）と通信する。いくつかの実施態様では、デジタルアシスタントシステム３００は、図１に示されるサーバシステム１０８（及び／又はデジタルアシスタントサーバ１０６）の一実施形態である。いくつかの実施態様では、デジタルアシスタントシステム３００はユーザ機器（例えば、ユーザ機器１０４、図１）内に実装され、それにより、クライアント-サーバシステムの必要性を取り除いている。デジタルアシスタントシステム３００はデジタルアシスタントシステムの単なる一例にすぎず、デジタルアシスタントシステム３００は、図示されているよりも多数又は少数の構成要素を有してもよく、２つ以上の構成要素を組み合わせてもよく、又は構成要素の異なる構成又は配置を有してもよいことに留意されたい。図３Ａに示される様々な構成要素は、１つ以上の信号処理及び／若しくは特定用途向け集積回路を含む、ハードウェア、ソフトウェア、ファームウェア、又はそれらの組み合わせの形態で実装されてもよい。

20

30

【００６１】

デジタルアシスタントシステム３００は、メモリ３０２、１つ以上のプロセッサ３０４、入力／出力（Ｉ／Ｏ）インターフェース３０６、及びネットワーク通信インターフェース３０８を含む。これらの構成要素は１本以上の通信バス又は信号線３１０を通じて互いに通信する。

40

【００６２】

いくつかの実施態様では、メモリ３０２は、高速ランダムアクセスメモリ及び／又は不揮発性コンピュータ可読記憶媒体（例えば、１つ以上の磁気ディスク記憶装置、１つ以上のフラッシュメモリ装置、１つ以上の光記憶装置、及び／又はその他の不揮発性固体メモリ装置）等の、永続的コンピュータ可読媒体を含む。

【００６３】

Ｉ／Ｏインターフェース３０６は、ディスプレイ、キーボード、タッチスクリーン、及

50

びマイクロフォン等の、デジタルアシスタントシステム300の入出力機器316をユーザインターフェースモジュール322に連結する。I/Oインターフェース306は、ユーザインターフェースモジュール322と協働して、ユーザ入力（例えば、音声入力、キーボード入力、タッチ入力等）を受け取り、それらを適宜処理する。いくつかの実施態様では、デジタルアシスタントがスタンドアロンのユーザ機器上に実装される場合には、デジタルアシスタントシステム300は、図2におけるユーザ機器104に関して説明された構成要素並びにI/O及び通信インターフェースのいずれか（例えば、1つ以上のマイクロフォン230）を含む。いくつかの実施態様では、デジタルアシスタントシステム300はデジタルアシスタント実装のサーバ部分を表し、ユーザ機器（例えば、図2に示されるユーザ機器104）上に常駐するクライアント側部分を通じてユーザと対話する。

10

【0064】

いくつかの実施態様では、ネットワーク通信インターフェース308は、有線通信ポート（単数又は複数）312及び/又は無線送受信回路314を含む。有線通信ポート（単数又は複数）は、1つ以上の有線インターフェース、例えば、イーサネット、ユニバーサルシリアルバス（USB）、FIREWIRE（登録商標）などを介して通信信号を受信及び送信する。無線回路314は、通常、通信ネットワーク及び他の通信機器を起点/終点として、RF信号及び/又は光信号を受信及び送信する。無線通信は、GSM（登録商標）、EDGE、CDMA、TDMA、Bluetooth（登録商標）、Wi-Fi（登録商標）、VoIP、Wi-MAX（登録商標）、あるいは任意のその他の好適な通信プロトコル等の、複数の通信規格、プロトコル及び技術のいずれかを用いることができる。ネットワーク通信インターフェース308は、デジタルアシスタントシステム300と、インターネット、イントラネット、等のネットワーク、及び/又はセルラー電話ネットワーク、無線ローカルエリアネットワーク（LAN）、等の無線ネットワーク、及び/又はメトロポリタンエリアネットワーク（metropolitan area network、MAN）、及び他の機器、との間の通信を可能にする。

20

【0065】

いくつかの実施態様では、メモリ302の永続的コンピュータ可読記憶媒体は、オペレーティングシステム318、通信モジュール320、ユーザインターフェースモジュール322、1つ以上のアプリケーション324、及びデジタルアシスタントモジュール326のすべて又はサブセットを含む、プログラム、モジュール、命令、及びデータ構造を記憶する。1つ以上のプロセッサ304はこれらのプログラム、モジュール、命令を実行し、データ構造から/へと読み出し/書き込みを実行する。

30

【0066】

オペレーティングシステム318（例えば、Darwin（登録商標）、RTXC（登録商標）、LINUX（登録商標）、UNIX（登録商標）、OS X（登録商標）、iOS（登録商標）、ウィンドウズ（登録商標）、又はVxWorks等の組み込みオペレーティングシステム）は、一般システムタスク（例えば、メモリ管理、記憶装置制御、電力管理等）を制御及び管理するための様々なソフトウェア構成要素及び/又はドライバを含み、様々なハードウェア、ファームウェア、及びソフトウェア構成要素の間の通信を促進する。

40

【0067】

通信モジュール320は、ネットワーク通信インターフェース308を通じたデジタルアシスタントシステム300とその他の機器との間の通信を促進する。例えば、通信モジュール320は、図2に示される機器104の通信モジュール254と通信することができる。通信モジュール320はまた、無線回路314及び/又は有線通信ポート312によって受信されたデータを処理するための様々なソフトウェア構成要素も含む。

【0068】

いくつかの実施態様では、ユーザインターフェースモジュール322はI/Oインターフェース306を介して（例えば、キーボード、タッチスクリーン、及び/又はマイクロフォンから）ユーザからのコマンド及び/又は入力を受け取り、ディスプレイ上にユーザ

50

インターフェースオブジェクトを提供する。

【 0 0 6 9 】

アプリケーション 3 2 4 は、1 つ以上のプロセッサ 3 0 4 によって実行されるように構成されたプログラム及び / 又はモジュールを含む。例えば、デジタルアシスタントシステムがスタンドアロンのユーザ機器上に実装される場合には、アプリケーション 3 2 4 は、ゲーム、カレンダーアプリケーション、ナビゲーションアプリケーション、又は電子メールアプリケーション等の、ユーザアプリケーションを含んでもよい。デジタルアシスタントシステム 3 0 0 がサーバファーム上に実装される場合には、アプリケーション 3 2 4 は、例えば、資源管理アプリケーション、診断アプリケーション、又はスケジューリングアプリケーションを含んでもよい。

10

【 0 0 7 0 】

メモリ 3 0 2 はまた、デジタルアシスタントモジュール（即ちデジタルアシスタントのサーバ部分）3 2 6 も記憶する。いくつかの実施態様では、デジタルアシスタントモジュール 3 2 6 は、以下のサブモジュール、又はそれらのサブセット若しくはスーパーセットを含む。即ち、入出力処理モジュール 3 2 8、音声テキスト化（S T T）処理モジュール 3 3 0、自然言語処理モジュール 3 3 2、ダイアログフロー処理モジュール 3 3 4、タスクフロー処理モジュール 3 3 6、サービス処理モジュール 3 3 8、及び写真モジュール 1 3 2、である。これら処理モジュールのそれぞれは、デジタルアシスタント 3 2 6 の以下のデータ及びモデルのうちの 1 つ以上、又はそれらのサブセット又はスーパーセットへのアクセスを有する。オントロジ 3 6 0、語彙索引 3 4 4、ユーザデータ 3 4 8、分類モジュール 3 4 9、曖昧性除去モジュール 3 5 0、タスクフローモデル 3 5 4、サービスモデル 3 5 6、写真タグ化モジュール 3 5 8、検索モジュール 3 6 0、及びローカルタグ / 写真ストレージ 3 6 2。

20

【 0 0 7 1 】

いくつかの実施態様では、デジタルアシスタントモジュール 3 2 6 内に実装された処理モジュール（例えば、入出力処理モジュール 3 2 8、S T T 処理モジュール 3 3 0、自然言語処理モジュール 3 3 2、ダイアログフロー処理モジュール 3 3 4、タスクフロー処理モジュール 3 3 6、及び / 又はサービス処理モジュール 3 3 8）、データ、及びモデルを使用して、デジタルアシスタントシステム 3 0 0 は少なくとも以下のいくつかを行う。ユーザから受け取った、自然言語入力にて表されたユーザの意思を識別すること、ユーザの意思を完全に推測するために必要な情報（例えば、単語、名称、意思、等の曖昧性除去によって）を積極的に引き出し、取得すること、推測した意思を満たすためのタスクフローを判定すること、及び推測した意思を満たすためにタスクフローを実行すること。いくつかの実施態様では、デジタルアシスタントはまた、様々な理由で満足のいく応答がユーザに提供されなかったか、又は提供できなかった場合に、適切なアクションを取る。

30

【 0 0 7 2 】

いくつかの実施態様では、以下に説明するように、デジタルアシスタントシステム 3 0 0 は、自然言語入力から、ユーザの意思を識別してデジタル写真をタグ付けし、デジタル写真を適切な情報とともにタグ付けするよう、自然言語入力を処理する。いくつかの実施態様では、デジタルアシスタントシステム 3 0 0 は、自然言語入力を使用してのデジタル写真の検索、写真の自動タグ付け、等などの、写真に関連する他のタスクをも行う。図 3 B に示すように、いくつかの実施態様では、I / O 処理モジュール 3 2 8 は、図 3 A の I / O 機器 3 1 6 を通してユーザと双方向作用するか、又は図 3 A のネットワーク通信インターフェース 3 0 8 を通してユーザ機器（例えば、図 1 のユーザ機器 1 0 4）と双方向作用し、ユーザ入力（例えば、スピーチ入力）を取得し、及び、ユーザ入力に対する応答を提供する。I / O 処理モジュール 3 2 8 は、ユーザ入力の受け取りとともに、又はその受け取り直後に、ユーザ入力に関連付けられたコンテキスト情報をユーザ機器から任意に獲得する。コンテキスト情報は、ユーザ入力に関連するユーザ固有のデータ、語彙、及び / 又は嗜好を含む。いくつかの実施態様では、コンテキスト情報はまた、ユーザ要求を受け取る時点における機器（例えば、図 1 におけるユーザ機器 1 0 4）のソフトウェア及びハ

40

50

ードウェア状態、及び／又はユーザ要求を受け取った時点におけるユーザの周囲環境に関する情報も含む。いくつかの実施態様では、I / O 処理モジュール 3 2 8 はまた、ユーザ要求について、ユーザに追加質問を送り、ユーザから回答を受け取る。いくつかの実施態様では、ユーザ要求が I / O 処理モジュール 3 2 8 によって受け取られ、ユーザ要求が音声入力を包含する場合には、I / O 処理モジュール 3 2 8 は、音声テキスト化変換のために、音声入力を音声テキスト化 (S T T) 処理モジュール 3 3 0 へ転送する。

【 0 0 7 3 】

いくつかの実施態様では、音声テキスト化処理モジュール 3 3 0 は、I / O 処理モジュール 3 2 8 を通じて音声入力（例えば、音声記録にキャプチャされたユーザの発言）を受け取る。いくつかの実施態様では、音声テキスト化処理モジュール 3 3 0 は、音声入力を、音素の列、最終的には、1 つ以上の言語で書かれた単語又はトークンの列として認識するために、様々な音響及び言語モデルを用いる。音声テキスト化処理モジュール 3 3 0 は、隠れマルコフモデル、動的タイムワーピング (D T W) ベースの音声認識、並びにその他の統計的及び／又は解析的技法等の、任意の好適な音声認識技法、音響モデル、及び言語モデルを用いて実装される。いくつかの実施態様では、音声テキスト化処理は、サードパーティーサービスによって、又はユーザの機器上で少なくとも部分的に実行され得る。音声テキスト化処理モジュール 3 3 0 が音声テキスト化処理の結果（例えば、単語又はトークンの列）を獲得すると、意図推定のために、その結果を自然言語処理モジュール 3 3 2 へ渡す。デジタルアシスタント 3 2 6 の自然言語処理モジュール 3 3 2 (「自然言語プロセッサ」) は、音声テキスト化処理モジュール 3 3 0 によって生成された単語又はトークンの列 (「トークン列」) を獲得し、トークン列を、デジタルアシスタントによって認識される 1 つ以上の「実施可能な意図」に関連付けようと試みる。本明細書で使用する場合、「実施可能な意図」は、デジタルアシスタント 3 2 6 及び／又はデジタルアシスタントシステム 3 0 0 (図 3 A) によって実行することができ、タスクフローモデル 3 5 4 内に実装される、関連付けられたタスクフローを有するタスクを表す。関連付けられたタスクフローは、タスクを実行するためにデジタルアシスタントシステム 3 0 0 が取る一連のプログラムされたアクション及びステップである。デジタルアシスタントシステムの能力の範囲は、タスクフローモデル 3 5 4 内に実施され、記憶されているタスクフローの数及び種類に依存し、又は、換言すれば、デジタルアシスタントシステム 3 0 0 が認識する「実施可能な意図」の数及び種類に依存する。ただし、デジタルアシスタントシステム 3 0 0 の有効性はまた、自然言語で表現されたユーザ要求から正確な「実施可能な意図 (単数又は複数)」を推測するデジタルアシスタントシステムの能力にも依存する。

【 0 0 7 4 】

いくつかの実施態様では、音声テキスト化処理モジュール 3 3 0 から獲得される単語又はトークンの列に加えて、自然言語プロセッサ 3 3 2 はまた、ユーザ要求に関連付けられたコンテキスト情報も受け取る（例えば、I / O 処理モジュール 3 2 8 から）。自然言語プロセッサ 3 3 2 は、音声テキスト化処理モジュール 3 3 0 から受信したトークン列内に包含されている情報を明らかにし、補完し、及び／又は更に明確にするために、コンテキスト情報を任意に用いる。コンテキスト情報は、例えば、ユーザの嗜好、ユーザ機器のハードウェア及び／又はソフトウェアの状態、ユーザ要求の前、最中、又は直後に集められたセンサ情報、デジタルアシスタントとユーザとの間の以前の対話（例えば、ダイアログ）、及び同様のものを含む。

【 0 0 7 5 】

いくつかの実施態様では、自然言語処理はオントロジ 3 6 0 に基づく。オントロジ 3 6 0 は、複数のノードを包含する階層構造であり、各ノードは、「複数の実施可能な意図群」又はその他の「複数の属性」のうちの 1 つ以上に関連する「実施可能な意図」又は「属性」のいずれかを表す。上述したように、「実施可能な意図」は、デジタルアシスタントシステム 3 0 0 が実行する能力を有するタスク（例えば、「実施可能」であるか又は実施の対象とすることができるタスク）を表す。「属性」は、実施可能な意図、又は別の属性の下位態様に関連付けられるパラメータを表す。オントロジ 3 6 0 内の実施可能な意図ノ

ードと属性ノードとの間のリンクは、属性ノードによって表されるパラメータが、実施可能な意図ノードによって表されるタスクにどのように関連するのかを定義する。いくつかの実施態様では、オントロジ360は、実施可能な意図ノード及び属性ノードで構成される。オントロジ360内において、各々の実施可能な意図ノードは、1つ以上の属性ノードに直接、若しくは1つ以上の中間属性ノードを介してリンクされる。同様に、それぞれの属性ノードは、1つ以上の実施可能な意図ノードに直接、若しくは1つ以上の中間属性ノードを介してリンクされる。例えば、図3Cに示すオントロジ360は、実施可能な意図ノードである「レストラン予約」ノードを含む。属性ノードである「レストラン」、「日付/時間」(予約用)及び「関係者人数」が、「レストラン予約」ノード(即ち、実施可能な意図ノード)にそれぞれ直接に接続される。更に、属性ノードである「料理」、「価格帯」、「電話番号」及び「場所」は、属性ノード「レストラン」のサブノードであり、中間属性ノード「レストラン」を介して「レストラン予約」ノードにそれぞれ接続される。別の例について、図3Cに示すオントロジ360は、別の実施可能な意図ノードである「リマインダ設定」ノードも含む。属性ノード「日付/時間」(リマインダ設定用)及び「テーマ」(リマインダ用)は、「リマインダ設定」ノードにそれぞれ接続される。属性ノード「日付/時間」はレストラン予約をするタスク及びリマインダを設定するタスクの両方に関連していることから、属性ノード「日付/時間」はオントロジ360において「レストラン予約」ノード及び「リマインダ設定」ノードの両方に接続される。

【0076】

実施可能な意図ノードは、その接続された概念ノードとともに、「ドメイン」として説明することができる。本説明では、各ドメインは、それぞれの実施可能な意図に関連付けられ、特定の実施可能な意図に関連付けられたノード群(及びそれらの間の関係)を指す。例えば、図3Cに示すオントロジ360には、レストラン予約ドメイン362の例及びオントロジ360内のリマインダドメイン364の例が含まれている。レストラン予約ドメインは、実施可能な意図ノードである「レストラン予約」、属性ノードである「レストラン」、「日付/時間」、及び「関係者人数」、並びに下位属性ノードである「料理」、「価格帯」、「電話番号」及び「場所」を含む。リマインダドメイン364には、実施可能な意図ノードである「リマインダ設定」、及び属性ノードである「テーマ」及び「日付/時間」が含まれている。いくつかの実施態様では、オントロジ360は、多数のドメインから構成されている。各ドメインは、1つ以上の他のドメインと、1つ以上の属性ノードを共有することができる。例えば、「日付/時間」の属性ノードは、レストラン予約ドメイン362及びリマインダドメイン364に加えて、多くの他のドメイン(例えば、スケジュールリングドメイン、旅行予約ドメイン、映画チケットドメインなど)に関連付けることができる。図3Cは、オントロジ360内に2つの例示的なドメインを示すが、オントロジ360は、「通話を開始する」、「道順を探す」、「会議を予定する」、「メッセージを送る」、及び「質問の答えを提供する」、「写真にタグを付ける」などの他のドメイン(即ち、実施可能な意図)を含んでもよい。例えば、「メッセージを送信する」のドメインは、「メッセージを送信する」の実施可能な意図ノードに関連付けられており、「受信者(単数又は複数)」、「メッセージタイプ」、及び「メッセージ本文」などの属性ノードを更に含むことができる。属性ノードの「受信者」は更に、例えば、「受信者名」と「メッセージアドレス」などの下位属性ノードによって定義してもよい。

【0077】

いくつかの実施態様では、オントロジ360は、デジタルアシスタントが理解し、作用を及ぼすことができるすべてのドメイン(したがって実施可能な意図)を含む。いくつかの実施態様では、オントロジ360は、ドメイン若しくはノードを追加又は除去すること、又はオントロジ360内のノード間の関係を変更することなどによって、修正してもよい。

【0078】

いくつかの実施態様では、複数の関連する実施可能な意図に関連付けられるノードは、オントロジ360内の「上位ドメイン」の下にクラスタ化されてもよい。例えば、「旅行

10

20

30

40

50

」上位ドメインは、旅行に関連する属性ノード及び実施可能な意図ノードのクラスタを含んでもよい。旅行に関連する実施可能な意図ノードは、「航空券予約」、「ホテル予約」、「カーレンタル」、「道順を知る」、「見所を探す」などを含んでもよい。同じ上位ドメイン（例えば、「旅行」上位ドメイン）の下の実施可能な意図ノードは、多くの属性ノードを共有し得る。例えば、「航空券予約」、「ホテル予約」、「カーレンタル」、「道順を知る」、「見所を探す」についての実施可能な意図ノードは、属性ノードである「出発場所」、「目的地」、「出発日／時間」、「到着日／時間」、及び「関係者人数」のうちの1つ以上を共有する可能性がある。

【0079】

いくつかの実施態様では、オントロジ360内の各ノードは、そのノードによって表される属性又は実施可能な意図に関連する単語及び／又はフレーズのセットに関連付けられる。各ノードに関連付けられる単語及び／又はフレーズのそれぞれのセットは、そのノードに関連付けられるいわゆる「語彙」である。各ノードに関連付けられる単語及び／又はフレーズのそれぞれのセットは、そのノードによって表される属性又は実施可能な意図に関連して、語彙索引344（図3B）内に記憶することができる。例えば、図3Bに戻ると、「レストラン」の属性についてのノードに関連付けられる語彙は、「食べ物」、「飲み物」、「料理」、「空腹」、「食べる」、「ピザ」、「ファーストフード」、「食事」などの単語を含んでもよい。別の例として、「通話を開始する（initiate a phone call）」という実施可能な意図のノードに関連付けられる語彙は、「電話する（call）」、「電話（phone）」、「ダイヤルする（dial）」、「電話をかける（ring）」、「この番号に電話する（call this number）」、「～に電話をかける（make a call to）」などの単語及びフレーズを含んでもよい。語彙索引344は異なる言語の単語及びフレーズを任意に含む。いくつかの実施態様では、図3Bに示す自然言語プロセッサ332は、音声テキスト化処理モジュール330からトークン列（例えば、テキスト文字列）を受信し、どのノードがトークン列内の単語によって暗示されているのかを判断する。いくつかの実施態様では、トークン列内の単語又はフレーズが（語彙索引344を介して）オントロジ360内の1つ以上のノードに関連付けられることが分かれば、その単語又はフレーズは、それらのノードを「トリガ」又は「アクティブにする」ことになる。複数のノードが「トリガされる」場合には、アクティブにされたノードの量及び／又は相対的重要度に基づき、自然言語プロセッサ332は、実施可能な意図のうちの1つをユーザがデジタルアシスタントに実行させようと意図したタスク（又はタスクの種類）として選択することになる。いくつかの実施態様では、最も多く「トリガされた」ノードを有するドメインが選択される。いくつかの実施態様では、最も高い信頼値を有するドメイン（例えば、その様々なトリガされたノードの相対的重要度に基づく）が選択される。いくつかの実施態様では、ドメインは、トリガされたノードの数及び重要度の組み合わせに基づき選択される。いくつかの実施態様では、ノードを選択する際には、デジタルアシスタントシステム300がユーザからの同様の要求を以前に正確に解釈したかどうか等の追加の因子も考慮される。

【0080】

いくつかの実施態様では、デジタルアシスタントシステム300はまた、特定のエンティティの名前も語彙索引344内に記憶する。そのため、これらの名前の1つがユーザ要求内に検出されると、自然言語プロセッサ332は、その名前はオントロジ内の属性又は下位属性の特定のインスタンスを指していると認識できることになる。いくつかの実施態様では、特定のエンティティの名前は、企業、レストラン、人、映画、及び同様のものの名前である。いくつかの実施態様では、デジタルアシスタントシステム300は、ユーザのアドレス帳、連絡先リスト、映画データベース、ミュージシャンデータベース、及び／又はレストランデータベース等の、他のデータ源から特定のエンティティ名を検索し、これらを識別することができる。いくつかの実施態様では、自然言語プロセッサ332が、トークン列内のある単語が特定のエンティティの名前（ユーザのアドレス帳内又は連絡先リスト内の名前等）であると識別すると、その単語は、ユーザ要求に対するオントロジ内の実施可能な意図を選択する際に追加的な重要性を与えられる。例えば、ユーザ要求から

10

20

30

40

50

単語「Mr. Santo」が認識され、語彙索引 344 内にユーザの連絡先リスト内の連絡先の 1 つとして姓「Santo」が見つかった場合、その時は、ユーザ要求は、「メッセージを送信する」又は「通話を開始する」ドメインに対応する可能性が高い。別の例としては、単語「ABC Cafe」がユーザ要求内に見つかり、語彙索引 344 内にユーザの市内の特定のレストランの名前としてターム「ABC Cafe」が見つかった場合、その時は、ユーザ要求は「レストラン予約」ドメインに対応する可能性が高い。

【0081】

ユーザデータ 348 は、ユーザ固有の語彙、ユーザの嗜好、ユーザアドレス、ユーザの既定言語及び第 2 言語、ユーザの連絡先リスト、並びに各ユーザに関するその他の短期的又は長期的情報等の、ユーザ固有の情報を含む。自然言語プロセッサ 332 は、ユーザ入力内に包含されている情報を補完してユーザ意図を更に明確にするために、ユーザ固有の情報をを用いることができる。例えば、ユーザ要求「私の誕生日パーティーに私の友人を招待してほしい」に対して、自然言語プロセッサ 332 は、誰が「友人」であり、いつ及びどこで「誕生日パーティー」が催されるのかを判断するために、ユーザの要求内のこうした情報を明示的に提供するようにユーザに要求する代わりに、ユーザデータ 348 にアクセスすることができる。

【0082】

いくつかの実施態様では、自然言語プロセッサ 332 は、分類モジュール 349 を含む。いくつかの実施態様では、分類モジュール 349 は、テキスト文字列（例えば、デジタル写真に関連付けられた音声入力に対応する）内の 1 つ以上のタームのそれぞれが、以下でより詳細に説明するような、エンティティ、動作、又は場所のいずれかであるか否かを判定する。いくつかの実施態様では、分類モジュール 349 は、1 つ以上のタームの各タームを、エンティティ、動作、又は場所の 1 つであるとして分類する。自然言語プロセッサ 332 がユーザ要求に基づき、実施可能な意図（又はドメイン）を識別すると、自然言語プロセッサ 332 は、識別された実施可能な意図を表すように、構造化されたクエリを生成する。いくつかの実施態様では、構造化されたクエリは、実施可能な意図に関するドメイン内の 1 つ以上のノードのためのパラメータを含み、パラメータの少なくともいくつかにはユーザ要求内で指定された特定の情報及び要求事項が追加される。例えば、ユーザは「私のために 7 時に寿司屋で夕食の予約を取ってください」と言う場合がある。この場合には、自然言語プロセッサ 332 は、ユーザ入力に基づき、実施可能な意図を「レストラン予約」であると正確に特定することができてよい。オントロジによれば、「レストラン予約」ドメインのための構造化されたクエリは、{料理}、{時間}、{日付}、{関係者人数}、及び同様のもの等のパラメータを含み得る。ユーザの発言内に包含されている情報に基づき、自然言語プロセッサ 332 はレストラン予約ドメインのための部分的な構造化されたクエリを生成してもよい。ここでは、部分的な構造化されたクエリは、パラメータ {料理 = 「寿司」} 及び {時間 = 「午後 7 時」} を含む。しかし、本例では、ユーザの発言は、ドメインに関連付けられる構造化されたクエリを完成するために十分な情報を包含していない。したがって、{関係者人数} 及び {日付} 等のその他の必要パラメータは、現在利用可能な情報に基づいては、構造化されたクエリ内に指定されない。いくつかの実施態様では、自然言語プロセッサ 332 は、構造化されたクエリのいくつかのパラメータに、受信したコンテキスト情報を追加する。例えば、ユーザが「私の近くの」寿司レストランを要求した場合には、自然言語プロセッサ 332 は、構造化されたクエリ内の {場所} パラメータにユーザ機器 104 から GPS 座標を追加してもよい。

【0083】

いくつかの実施態様では、自然言語プロセッサ 332 は、構造化されたクエリ（任意の完成したパラメータを含む）をタスクフロー処理モジュール 336（「タスクフロープロセッサ」）に渡す。タスクフロープロセッサ 336 は、自然言語プロセッサ 332 から、構造化されたクエリを受信すること、構造化されたクエリを完成させること、及びユーザの最終的な要求を「完了する」ために必要とされるアクションを実行すること、のうちの 1 つ以上を実行するように構成されている。いくつかの実施態様では、これらのタスクを

10

20

30

40

50

完了させるために必要な様々な手順がタスクフローモデル 354 内で提供される。いくつかの実施態様では、タスクフローモデル 354 は、ユーザから追加情報を入手するための手順、及び実施可能な意図に関連付けられるアクションを実行するためのタスクフローを含んでいる。上述したように、構造化されたクエリを完成させるために、タスクフロープロセッサ 336 は、追加情報を入手し、及び / 又は潜在的に曖昧な発言の曖昧さを除去するために、ユーザとの追加のダイアログを開始することが必要になる場合がある。このような対話が必要となる場合には、タスクフロープロセッサ 336 は、ユーザとの対話に携わるためにダイアログ処理モジュール 334 (ダイアログプロセッサ) を呼び出す。いくつかの実施態様では、ダイアログ処理モジュール 334 は、どのように (及び / 又はいつ) ユーザに追加情報を求めるべきかを決定し、ユーザ応答を受け取り、これら进行处理する。いくつかの実施態様では、I/O 処理モジュール 328 を通じてユーザに質問を提供し、ユーザから回答を受け取る。例えば、ダイアログ処理モジュール 334 はオーディオ及び / 又は視覚出力を介してユーザにダイアログ出力を提示し、口頭の又は物理的な (例えば、タッチジェスチャ) 応答を介してユーザから入力を受け取る。上述の例を続けると、タスクフロープロセッサ 336 が、ドメイン「レストラン予約」に関連付けられる構造化されたクエリについての「関係者人数」及び「日付」情報を決定するためにダイアログプロセッサ 334 を呼び出すと、ダイアログプロセッサ 334 は、「何人分ですか?」及び「どの日にしますか?」等の質問を、ユーザに渡すために生成する。ユーザから回答を受け取ると、ダイアログ処理モジュール 334 は、構造化されたクエリに欠けている情報を追加するか、又は構造化されたクエリから欠けている情報を完成させるために、その情報をタスクフロープロセッサ 336 に渡す。

【0084】

場合によっては、タスクフロープロセッサ 336 は、1つ以上の曖昧な属性を有する構造化されたクエリを受信することがある。例えば、「メッセージを送信する」ドメインのための構造化されたクエリは、意図された受信者は「Bob」であることを指示する場合があり、ユーザは、「Bob」という名前の連絡先を複数有する場合がある。タスクフロープロセッサ 336 は、ダイアログプロセッサ 334 が、構造化されたクエリのこの属性の曖昧さを除去するように要求することになる。その結果として、ダイアログプロセッサ 334 は「どのBobですか?」とユーザに尋ね、ユーザが選び得る「Bob」という名前の連絡先のリストを表示してもよい (又は読み上げてもよい)。

【0085】

いくつかの実施態様では、ダイアログプロセッサ 334 は、曖昧性除去モジュール 350 を含む。いくつかの実施態様では、曖昧性除去モジュール 350 は、1つ以上の曖昧なターム (例えば、デジタル写真に関連付けられた音声入力に対応するテキスト文字列の1つ以上の曖昧なターム) の曖昧性を除去する。いくつかの実施態様では、曖昧性除去モジュール 350 は、1つ以上のタームの第1のタームは、複数の候補となる意味を有しており、第1のタームについてユーザにプロンプトで追加情報を求め、ユーザからプロンプトに対応する追加情報を受け取って、その追加情報に応じて、第1のタームに関連付けられるエンティティ、動作、場所を識別する。

【0086】

いくつかの実施態様では、曖昧性除去モジュール 350 は、代名詞の曖昧性を除去する。このような実施態様では、曖昧性除去モジュール 350 は、1つ以上のタームの1つを代名詞として識別し、その代名詞が指す名詞を決定する。いくつかの実施態様では、曖昧性除去モジュール 350 は、電子機器のユーザに関連付けられた連絡先リストを使用して、代名詞が指す名詞を決定する。あるいは、又は更に、曖昧性除去モジュール 350 は、以前にタグ付けされたデジタル写真に関連付けられた従前の音声入力で識別されたエンティティ、動作、又は場所の名前として、その代名詞が指す名詞を決定する。あるいは、又は更に、曖昧性除去モジュール 350 は、以前にタグ付けされたデジタル写真に関連付けられた従前の音声入力に基づいて識別された人の名前として、その代名詞が指す名詞を決定する。いくつかの実施態様では、曖昧性除去モジュール 350 は、タームのうちの1つ

以上の意味を決定するために、ハンドヘルド電子機器（例えば、ユーザ機器 104）の 1 つ以上のセンサ（例えば、近接センサ 214、光センサ 212、GPS 受信機 213、温度センサ 215、動きセンサ 210）から得られた情報にアクセスする。いくつかの実施態様では、曖昧性除去モジュール 350 は、エンティティ、動作、又は場所のいずれかにそれぞれ関連付けられた 2 つのタームを識別する。例えば、2 つのタームのうち第 1 のタームは人を指し、2 つのタームのうちの第 2 のタームは場所を指す。いくつかの実施態様では、曖昧性除去モジュール 350 は、エンティティ、動作、又は場所のいずれかにそれぞれ関連付けられた 3 つのタームを識別する。

【0087】

タスクフロープロセッサ 336 が、実施可能な意図のための構造化されたクエリを完成すると、タスクフロープロセッサ 336 は、実施可能な意図に関連付けられる最終的なタスクの実行に進む。それに応じて、タスクフロープロセッサ 336 は、構造化されたクエリ内に包含されている具体的なパラメータに応じて、タスクフローモデル内のステップ及び命令を実行する。例えば、「レストラン予約」の実施可能な意図のためのタスクフローモデルは、レストランと連絡を取り、特定の時間に特定の関係者人数のための予約を実際に希望するためのステップ及び命令を含み得る。例えば、{ レストラン予約, レストラン = ABC Cafe, 日付 = 3 / 12 / 2012, 時間 = 午後 7 時, 関係者人数 = 5 } などの構造化されたクエリを使用して、タスクフロープロセッサ 336 は、(1) ABC Cafe のサーバ又は ABC Cafe などの複数のレストランへの予約を受け付けるよう構成されたレストラン予約システムへログインするステップと、(2) ウェブサイト上のフォームに日付、時間、及び関係者人数情報を入力するステップと、(3) フォームを提出するステップと、(4) ユーザのカレンダーに予約をカレンダーエントリするステップと、を行ってもよい。別の例では、以下でより詳細に説明するが、タスクフロープロセッサ 336 は、例えば、写真モジュール 132 と連携して、音声入力に応じて、デジタル写真のタグ付け又は検索に関連するステップ及び命令を実行する。いくつかの実施態様では、タスクフロープロセッサ 336 は、ユーザ入力で要求されたタスクを完了するため、又はユーザ入力で要求された情報の回答を提供するために、サービス処理モジュール 338（「サービスプロセッサ」）の支援を用いる。例えば、サービスプロセッサ 338 は、タスクフロープロセッサ 336 の代わりに、電話をかけたり、カレンダー項目を設定したり、地図検索を呼び出したり、ユーザ機器上にインストールされているその他のユーザアプリケーションを呼び出すか又はそれと双方向作用したり、及びサードパーティーサービス（例えば、レストラン予約ポータル、ソーシャルネットワークのウェブサイト又はサービス、バンキングポータル等）を呼び出すか又はそれと双方向作用したりするアクションを行うことができる。いくつかの実施態様では、各サービスによって必要とされるプロトコル及びアプリケーションプログラミングインターフェース（Application Programming Interface、API）は、サービスモデル 356 の中のそれぞれのサービスモデルによって指定され得る。サービスプロセッサ 338 は、サービスのための適当なサービスモデルにアクセスし、サービスモデルに係るサービスによって必要とされるプロトコル及び API にしたがってサービスの要求を生成する。

【0088】

例えば、レストランがオンライン予約サービスを可能にしている場合には、レストランは、予約を行うために必要なパラメータ、及び必要なパラメータの値をオンライン予約サービスへ連絡するための API を指定するサービスモデルを提示することができる。タスクフロープロセッサ 336 によって要求されると、サービスプロセッサ 338 は、サービスモデル 356 に記憶されているウェブアドレスを用いて、オンライン予約サービスとのネットワーク接続を確立し、必要な予約のパラメータ（例えば、時間、日付、関係者人数）を、オンライン予約サービスの API に従う形式でオンライン予約インターフェースへ送信することができる。

【0089】

いくつかの実施態様では、ユーザの意図を推定して明確にし、ユーザの意図を更に明ら

10

20

30

40

50

かにして絞り込むための情報を入手し、ユーザの意図を達成する応答を最終的に生成する（例えば、ユーザに出力を提供するか、又はタスクを完了する）ために、自然言語プロセッサ 332、ダイアログプロセッサ 334、及びタスクフロープロセッサ 336 は共同でかつ反復的に用いられる。

【0090】

いくつかの実施態様では、ユーザの要求を実現するために必要なタスクがすべて実行された後に、デジタルアシスタント 326 は確認応答を策定し、I/O 処理モジュール 328 を通じてユーザへ応答を送り返す。ユーザ要求が情報の回答を求めている場合には、確認応答は、要求されている情報をユーザに提示する。いくつかの実施態様では、デジタルアシスタントはまた、デジタルアシスタント 326 によって作成された応答にユーザは満足であるかどうかを指示するようにユーザに要求する。

10

【0091】

ここで、いくつかの実施態様に係る、ボイストリガシステム 400 の構成要素を示すブロック図である図 4 に注目する。（ボイストリガシステム 400 は音声に限定されておらず、本明細書に説明する実施態様は等しく非音声に適用する。）ボイストリガシステム 400 は、電子機器 104 内に、様々な構成要素、モジュール、及び/又はソフトウェアプログラムにて構成されている。いくつかの実施態様では、ボイストリガシステム 400 は、ノイズ検出器 402、音種類検出器 404、トリガ音検出器 406、スピーチベースのサービス 408、及びオーディオサブシステム 226 を含み、それぞれはオーディオバス 401 に連結されている。いくつかの実施態様では、これらより多い又は少ないモジュールが使用されている。音検出器 402、404、及び 406 はモジュールと呼んでもよく、本明細書に説明する機能を実行するための、ハードウェア（例えば、回路、メモリ、プロセッサ、等）、ソフトウェア（例えば、プログラム、チップ上のソフトウェア、ファームウェア、等）、及び/又はいずれのそれらの組み合わせを含んでもよい。いくつかの実施態様では、図 4 に破線で示すように、音検出器は、互いに通信可能に、プログラムにて、物理的に、及び/又は操作可能に、（例えば、通信バス経由で）連結されている。（説明を簡易にするため、図 4 は、隣接する音検出器にのみ連結されているそれぞれの音検出器を示す。それぞれの音検出器は、同様に他の音検出器のいずれかに連結され得ることが理解されるであろう。）

20

【0092】

いくつかの実施態様では、オーディオサブシステム 226 は、コーデック 410、オーディオデジタル信号プロセッサ（DSP）412、及びメモリバッファ 414 を含む。いくつかの実施態様では、このオーディオサブシステム 226 は、1つ以上のマイクロフォン 230（図 2）及び、1つ以上のスピーカ 228（図 2）に連結されている。オーディオサブシステム 226 は、処理及び/又は分析のために、音入力を、音検出器 402、404、406 及びスピーチベースのサービス 408（同様に、電話及び/又は電話のベースバンドサブシステムなどの、他の構成要素又はモジュール）に提供する。いくつかの実施態様では、オーディオサブシステム 226 は、少なくとも 1つのマイクロフォン 418 及び少なくとも 1つのスピーカ 420 を含む外部オーディオシステム 416 に連結されている。

30

40

【0093】

いくつかの実施態様では、スピーチベースのサービス 408 は音声ベースのデジタルアシスタントであり、図 1～図 3C に関連して上述のデジタルアシスタントシステムの 1つ以上の構成要素又は機能に対応する。いくつかの実施態様では、スピーチベースのサービスは、音声テキスト化サービス、口述サービス、等であり、いくつかの実施態様では、ノイズ検出器 402 はオーディオチャネルを監視し、オーディオサブシステム 226 からの音入力が、振幅閾値などの所定の条件を満たすか判定する。オーディオチャネルは、1つ以上のマイクロフォン 230 などの、1つ以上の収音機器によって受信した音声情報のストリームに対応する（図 2）。オーディオチャネルは、その処理状態に関わらずに音声情報を参照するか、又は音声情報を処理及び/又は送信している特定のハードウェアを参照

50

する。例えば、オーディオチャネルは、マイクロフォン 230 からのアナログ電気インパルス（及び／又は、それらが伝播されている回路）、並びに、（例えば、オーディオサブシステム 226 及び／又は、電子機器 104 の任意の他のオーディオ処理システムによる）アナログ電気インパルスの処理の結果として、デジタル的にエンコードされたオーディオストリームを参照してもよい。

【0094】

いくつかの実施態様では、所定の条件は、音入力、所定の時間に特定の音量を超えるか否かである。いくつかの実施態様では、ノイズ検出器は、音入力の時間領域分析を使用し、（例えば、音種類検出器 404、トリガ語検出器 406、及び／又はスピーチベースのサービス 408 によって行われるような）他の種類の分析に比較して、これは相対的に少ない計算リソース及びバッテリーリソースを必要とする。いくつかの実施態様では、例えば、周波数領域分析を含む、他の種類の信号処理及び／又は音声分析が使用される。音入力、所定の条件を満たすことをノイズ検出器 402 が判定すると、（例えば、1 つ以上の処理ルーチンを開始するよう制御信号を提供することで、及び／又は上流の音検出器に電力を提供することで）音種類検出器 404 などの、上流の音検出器を起動する。いくつかの実施態様では、上流の音検出器は、他の満たされている条件に応じて起動されている。例えば、いくつかの実施態様では、上流の音検出器は、機器が閉鎖空間内に保管されていないことを判定することに応じて起動されている（例えば、光の閾値レベルを検出する光検出器に基づいて）。

【0095】

音種類検出器 404 はオーディオチャネルを監視し、音入力、人の声、口笛、拍手、等に独特の音などの特定の種類の音に対応するか判定する。音種類検出器 404 が認識するよう構成されている音の種類は、ボイストリガが認識するよう構成されている特定のトリガ音（単数又は複数）に対応する。トリガ音が話し言葉又はフレーズである実施態様では、音種類検出器 404 は、「音声活動検出器（voice activity detector）」（VAD）を含む。いくつかの実施態様では、音種類検出器 404 は、音入力の周波数領域分析を使用する。例えば、音種類検出器 404 は、（例えば、フーリエ変換を使用して）受け取った音入力のスペクトログラムを生成して、音入力のスペクトル構成要素を分析し、音入力、特定の音の種類又は分類（例えば、人の話し声）に対応すると思われるか判定する。したがって、トリガ音が話し言葉又はフレーズである実施態様では、オーディオチャネルが、人の話し声ではなく、背景音（例えば、トラフィックノイズ）を拾う場合は、VAD はトリガ音検出器 406 を起動しない。いくつかの実施態様では、音種類検出器 404 は、いずれの下流の音検出器（例えば、ノイズ検出器 402）の所定の条件が満たされる限りは、有効に維持される。例えば、いくつかの実施態様では、音種類検出器 404 は、（ノイズ検出器 402 によって判定されるように）音入力、所定の振幅閾値を上回る音を含む限りは、有効に維持され、音が所定の閾値を下回ると無効となる。いくつかの実施態様では、一度起動されると、音種類検出器 404 は、タイマの終了（例えば、1、2、5、又は 10 秒、若しくは任意の他の適切な持続時間）、音種類検出器 404 の特定のオン／オフサイクル数の終了、又は事象の発生（例えば、ノイズ検出器 402 及び／又は音種類検出器 404 により判定されるように、音の振幅が第 2 の閾値を下回った）などの条件が満たされるまで、有効に維持される。

【0096】

上述のように、音入力、所定の種類の音に対応することを音種類検出器 404 が判定すると、トリガ音検出器 406 などの、上流の音検出器を（例えば、1 つ以上の処理ルーチンを開始するよう制御信号を提供することで、及び／又は上流の音検出器に電力を提供することで）起動する。

【0097】

トリガ音検出器 406 は、音入力、特定の所定の内容の少なくとも一部（例えば、トリガ語、フレーズ、又は音の少なくとも一部）を含むか判定するよう構成されている。いくつかの実施態様では、トリガ音検出器 406 は、音入力のレプリゼンテーション（「入力

レプリゼンテーション」)を、トリガ語の1つ以上のリファレンスレプリゼンテーションと比較する。入力レプリゼンテーションが、1つ以上のリファレンスレプリゼンテーションの少なくとも1つと、許容できる信頼値で一致する場合は、トリガ音検出器406は、スピーチベースのサービス408を(例えば、1つ以上の処理ルーチンを開始するよう制御信号を提供することで、及び/又は上流の音検出器に電力を提供することで)開始する。いくつかの実施態様では、入力レプリゼンテーション及び1つ以上のリファレンスレプリゼンテーションはスペクトログラム(又は、その数学的レプリゼンテーション)であり、これは信号のスペクトル密度が時間とともにどのように変化するかを表す。いくつかの実施態様では、レプリゼンテーションは、他の種類のオーディオシグネチャ又は声紋である。いくつかの実施態様では、スピーチベースのサービス408を開始することは、1つ以上の回路、プログラム、及び/又はプロセッサをスタンバイモードから退出させ、音ベースのサービスを呼び出すことを含む。音ベースのサービスは続いて、より包括的なスピーチ認識、音声テキスト化処理、及び/又は自然言語処理を提供する準備をする。いくつかの実施態様では、ボイストリガシステム400は、音入力、機器の所有者/ユーザなどの、特定の人の声に対応するか判定できるよう、音声認証機能を含む。例えば、いくつかの実施態様では、音種類検出器404は、声紋化技術を使用して、音入力、正規ユーザによって発言されたことを判定する。音声認証及び声紋化は、そのすべてにおいて参照により本明細書に組み込まれている、本願の譲受人に帰属する、米国特許出願第13/053,144号に更に詳細に説明されている。いくつかの実施態様では、音声認証は、本明細書に説明される音検出器のいずれかに含まれている(例えば、ノイズ検出器402、音種類検出器404、トリガ音検出器406、及び/又はスピーチベースのサービス408)。いくつかの実施態様では、音声認証は、上記の音検出器とは別のモジュールとして実装されており(例えば、音声認証モジュール428として、図4)、ノイズ検出器402の後に、音種類検出器404の後に、トリガ音検出器406の後に、又は任意の他の適切な位置に操作可能に配置されてもよい。

【0098】

いくつかの実施態様では、トリガ音検出器406は、いずれの下流の音検出器(単数又は複数)(例えば、ノイズ検出器402及び/又は音種類検出器404)の条件が満たされる限りは、有効に維持される。例えば、いくつかの実施態様では、トリガ音検出器406は、音入力、(ノイズ検出器402により検出されるように)所定の閾値を上回る音を含む限りは、有効に維持される。いくつかの実施態様では、それは、音入力、(音種類検出器404により検出されるように)特定の種類の音を含む限りは、有効に維持される。いくつかの実施態様では、それは、前述の条件の双方が満たされる限りは、有効に維持される。

【0099】

いくつかの実施態様では、一度起動されると、トリガ音検出器406は、タイマの終了(例えば、1、2、5、又は10秒、若しくはいずれの他の適切な持続時間)、トリガ音検出器406の特定のオン/オフサイクル数の終了、又は事象の発生(例えば、音の振幅が第2の閾値を下回った)などの条件が満たされるまで、有効に維持される。いくつかの実施態様では、1つの音検出器が別の検出器を起動すると、双方の音検出器が有効に維持される。しかし、音検出器は、何度も有効又は無効とされてよく、上流の音検出器を有効とするために、下流の(例えば、低電力及び/又は精巧化)音検出器のすべてを有効にする(又は、各条件が満たされる)必要はない。例えば、いくつかの実施態様では、ノイズ検出器402及び音種類検出器404が、それらの各条件が満たされていることを判定し、トリガ音検出器406が起動された後、ノイズ検出器402及び音種類検出器404の1つ又は双方は、トリガ音検出器406の作動中に無効となる、及び/又はスタンバイモードとなる。他の実施態様では、ノイズ検出器402及び音種類検出器404の双方(又は、1つ若しくは他方)は、トリガ音検出器406の作動中に有効に維持される。様々な実施態様では、異なる組み合わせの音検出器が、異なる時間に有効となり、1つが有効又は無効となるかは、他の音検出器の状態に依存してもよく、又は他の音検出器の状態とは

10

20

30

40

50

独立としてもよい。

【0100】

図4は3つの個別の音検出器を説明するが、それぞれは、異なる態様の音入力を検出するように構成されており、ボイストリガの様々な実施態様では、これより多い、又は少ない音検出器が使用されている。例えば、いくつかの実施態様では、トリガ音検出器406のみが使用されている。いくつかの実施態様では、トリガ音検出器406は、ノイズ検出器402又は音種類検出器404のいずれかと併せて使用されている。いくつかの実施態様では、検出器402～406のすべてが使用されている。いくつかの実施態様では、追加的な音検出器が同様に含まれている。

【0101】

そのうえ、異なる組み合わせの音検出器が、異なる時間に使用されてもよい。例えば、特定の組み合わせの音検出器及びそれらがどのように双方向作用するかは、コンテキスト又は機器の作動状態などの、1つ以上の条件に依存してもよい。1つの具体例として、機器が電源に接続されている（及び、したがってバッテリー電力のみに依存しない）場合は、トリガ音検出器406は、ノイズ検出器402及び音種類検出器404が無効に維持されている間に、有効である。別の一例では、機器がポケット又はバックパック内にある場合は、すべての音検出器が無効である。多くの電力を必要とする検出器が、低電力を必要とする検出器によって、必要時にのみ呼び出される、上述したような音検出器をカスケード接続することで、省電力のボイストリガ機能が提供され得る。上述したように、デューティサイクルに従って、音検出器のうちの1つ以上を動作させることで、更なる省電力が達成されている。例えば、いくつかの実施態様では、ノイズ検出器402は、ノイズ検出器が少なくとも一時的に無効となっていて、効率よく継続的なノイズ検出を行うよう、デューティサイクルに従って作動する。いくつかの実施態様では、ノイズ検出器402は、10ミリ秒間オンとなり、90ミリ秒間オフとなる。いくつかの実施態様では、ノイズ検出器402は、20ミリ秒間オンとなり、500ミリ秒間オフとなる。他のオン及びオフの各持続時間も可能である。

【0102】

いくつかの実施態様では、ノイズ検出器402が、その「オン」インターバル中にノイズを検出すると、ノイズ検出器402はオンに維持され、音入力を更に処理及び/又は分析する。例えば、ノイズ検出器402は、所定の時間（例えば、100ミリ秒）に所定の振幅を上回る音を検出すると、上流の音検出器を起動するよう構成されてもよい。したがって、ノイズ検出器402が、その10ミリ秒間の「オン」インターバル中に、所定の振幅を上回る音を検出すると、直ちに「オフ」インターバルには入らない。その代り、ノイズ検出器402は有効に維持され、音入力の処理を継続し、所定の全持続時間（例えば、100ミリ秒）に、閾値を超えるかを判定する。

【0103】

いくつかの実施態様では、音種類検出器404は、デューティサイクルに従って作動する。いくつかの実施態様では、音種類検出器404は、20ミリ秒間オンとなり、100ミリ秒間オフとなる。他のオン及びオフの各持続時間も可能である。いくつかの実施態様では、音種類検出器404は、そのデューティサイクルの「オン」インターバル内に、音入力が所定の種類の音に対応するか判定することができる。したがって、音が特定の種類であることを音種類検出器404がその「オン」インターバル中に判定すると、音種類検出器404は、トリガ音検出器406（又は、任意の他の上流の音検出器）を起動する。あるいは、いくつかの実施態様では、音種類検出器404が、「オン」インターバル中に、所定の種類に対応できる音を検出すると、検出器は直ちに「オフ」インターバルには入らない。その代り、音種類検出器404は有効に維持され、音入力の処理を継続し、それが所定の種類の音に対応するか判定する。いくつかの実施態様では、所定の種類の音が検出されたことを音検出器が判定すると、トリガ音検出器406を起動し、音入力を更に処理し、トリガ音が検出されたか判定する。ノイズ検出器402及び音種類検出器404と同様に、いくつかの実施態様では、トリガ音検出器406は、デューティサイクルに従っ

10

20

30

40

50

て作動する。いくつかの実施態様では、トリガ音検出器 406 は、50 ミリ秒間オンとなり、50 ミリ秒間オフとなる。他のオン及びオフの各持続時間も可能である。トリガ音に対応できる音があることを、トリガ音検出器 406 がその「オン」インターバル中に検出すると、検出器は直ちに「オフ」インターバルには入らない。その代り、トリガ音検出器 406 は有効に維持され、音入力の処理を継続し、それがトリガ音を含むか判定する。いくつかの実施態様では、そのような音が検出されると、トリガ音検出器 406 は有効に維持され、1、2、5、又は 10 秒間、又は任意の他の適切な持続時間などの、所定の持続時間に、音声処理する。いくつかの実施態様では、持続時間は、検出するよう構成されている特定のトリガ語又は音の長さに基づいて選択されている。例えば、トリガフレーズが「SIRIへ」の場合、トリガ語検出器は約 2 秒間作動し、音入力にフレーズを含むか判定する。

10

【0104】

いくつかの実施態様では、音検出器のいくつかは、デューティサイクルに従って動作され、他は有効時に継続して作動する。例えば、いくつかの実施態様では、第 1 の音検出器のみがデューティサイクルに従って動作され（例えば、図 4 のノイズ検出器 402）、上流の音検出器は、一度起動すると継続して動作されている。いくつかの他の実施態様では、ノイズ検出器 402 及び音種類検出器 404 はデューティサイクルに従って動作され、一方でトリガ音検出器 406 は継続して動作されている。特定の音検出器が継続して動作されるか、又はデューティサイクルに従って動作されるかは、コンテキスト又は機器の作動状態などの、1 つ以上の条件に依存する。いくつかの実施態様では、機器が電源に接続されており、バッテリー電力のみに依存しない場合は、音検出器のすべてが、一度起動すると継続して作動する。他の実施態様では、機器が（例えば、センサ及びノ又はマイクロフォン信号により判定されるように）ポケット又はバックパック内にある場合は、ノイズ検出器 402（又は、音検出器のいずれか）は、デューティサイクルに従って作動するが、機器が保管されていない可能性があるとは判定されると、継続して作動する。いくつかの実施態様では、特定の音検出器が継続して動作されるか、又はデューティサイクルに従って動作されるかは、機器のバッテリー充電レベルに依存する。例えば、ノイズ検出器 402 は、バッテリー充電が 50 % 超の場合は継続して作動し、バッテリー充電が 50 % 未満の場合はデューティサイクルに従って作動する。いくつかの実施態様では、ボイストリガは、ノイズ、エコー、及びノ又は音キャンセル機能を含む（総称してノイズキャンセルと呼ばれる）。いくつかの実施態様では、ノイズキャンセルは、オーディオサブシステム 226 によって（例えば、オーディオ DSP 412 によって）行われている。ノイズキャンセルは、音検出器によって処理される前に、音入力から不要なノイズ又は音を削減又は除去する。場合によっては、不要なノイズは、ファン又はキーボードからのクリック操作などの、ユーザの環境からの背景ノイズである。いくつかの実施態様では、不要なノイズは、所定の振幅又は周波数の、これを上回る、又は下回る任意の音である。例えば、いくつかの実施態様では、一般的な人の発声範囲（例えば、3,000 Hz）を上回る音は、信号からフィルタアウト又は除去される。いくつかの実施態様では、複数のマイクロフォン（例えば、マイクロフォン 230）を使用して、受け取った音のどの構成要素を削減及びノ又は除去すべきか判定することを助ける。例えば、いくつかの実施態様では、オーディオサブシステム 226 はビーム形成技術を使用し、空間内の単一のポイント（例えば、ユーザの口）から生じる音又は音入力の各部分を識別する。オーディオサブシステム 226 は続いて、すべてのマイクロフォンによって等しく受け取られた音（例えば、いずれの特定の方向から生じていない背景音）を音入力から除去することで、この音に集中する。

20

30

40

【0105】

いくつかの実施態様では、DSP 412 は、デジタルアシスタントが作動している機器によって出力されている音を、音入力からキャンセル又は除去するよう構成されている。例えば、オーディオサブシステム 226 が、音楽、ラジオ、ポッドキャスト、音声出力、又は任意の他のオーディオコンテンツを（例えば、スピーカ 228 を介して）出力している場合、DSP 412 は、マイクロフォンによって拾われ、音入力に含まれた、出力され

50

た音のいずれかを除去する。したがって、この音入力は、この出力された音声を含まない（又は、少なくとも出力された音声より少なく含む）。それに応じて、音検出器に提供される音入力はよりクリーンであり、より正確なトリガである。ノイズキャンセルの態様は、そのすべてにおいて参照により本明細書に組み込まれている、本願の譲受人に帰属する、米国特許第7,272,224号に更に詳細に説明されている。

【0106】

いくつかの実施態様では、異なる音検出器は、音入力異なる方法にてフィルタリング及び/又は前処理されることを必要とする。例えば、いくつかの実施態様では、ノイズ検出器402は60と20,000Hzとの間の時間領域音声信号を分析するよう構成されており、音種類検出器は、60と3,000Hzとの間の音声の周波数領域分析を行うよう構成されている。したがって、いくつかの実施態様では、オーディオDSP412（及び/又は、機器104の他のオーディオDSP）は、受け取った音声を、音検出器の各ニースにしたがって前処理する。いくつかの実施態様では、一方で、音検出器は、それら特有のニースにしたがって、オーディオサブシステム226からの音声をフィルタリング及び/又は前処理するよう構成されている。そのような場合では、オーディオDSP412は、音入力を音検出器に提供する前に、ノイズキャンセルを依然として行ってもよい。いくつかの実施態様では、電子機器のコンテキストは、ボイストリガが作動しているか、及びどのように作動しているか判定することを助けるために使用されている。例えば、機器がポケット、財布、又はバックパックに入れてある場合には、音声ベースのデジタルアシスタントなどの、スピーチベースのサービスをユーザが呼び出すことはないと思われる。また、音の大きいロックコンサート中に、スピーチベースのサービスをユーザが呼び出すことはないと思われる。ユーザによっては、特定の時刻（例えば、深夜）にスピーチベースのサービスを呼び出すことはないと思われる。一方で、ボイストリガを使用して、スピーチベースのサービスをユーザが呼び出すことがおおいにあり得るコンテキストもある。例えば、ユーザによっては、運転中、一人にいる時、勤務中、等にボイストリガを使用することはおおいにありうる。様々な技術を使用して、機器のコンテキストが判定されている。様々な実施態様では、この機器は、以下の構成要素からの情報、又は機器のコンテキストを判定するための情報源のうちの任意の1つ以上からの情報を使用する。即ち、GPS受信機、光センサ、マイクロフォン、近接センサ、向きセンサ、慣性センサ、カメラ、通信回路及び/又はアンテナ、充電回路及び/又は電源回路、スイッチ位置、温度センサ、コンパス、加速度計、カレンダー、ユーザの嗜好、等。機器のコンテキストは続いて、ボイストリガを動作させるか、及びどのように動作させるかの調整に使用され得る。例えば、特定のコンテキストでは、ボイストリガは、コンテキストが維持される限りは無効となる（又は、異なるモードにて動作される）。例えば、いくつかの実施態様では、ボイストリガは、電話が所定の向きにある（例えば、表面の上で下向きに置かれている）、所定の期間中（例えば、10:00PMと8:00AMとの間）、電話が（例えば、スイッチ位置、モード設定、又はユーザの嗜好に基づいて）「サイレント」又は「マナー（do not disturb）」モードにある、機器が実質的な閉鎖空間内（例えば、ポケット、バッグ、財布、引き出し、又はグローブボックス）にある、（例えば、近接センサ、音声通信/無線通信/赤外線通信に基づく）ボイストリガ及び/又はスピーチベースのサービスを有する他の機器の近くに機器がある場合、等、の場合に無効となる。いくつかの実施態様では、無効にされる代りに、ボイストリガシステム400は、（例えば、ノイズ検出器402を、10ミリ秒間の「オン」インターバル及び5秒間の「オフ」インターバルを伴うデューティサイクルに従って動作させることで）低電力モードにて動作される。いくつかの実施態様では、オーディオチャネルは、ボイストリガシステム400が低電力モードにて動作されている際に、より低い頻度で監視されている。いくつかの実施態様では、ボイストリガは、通常モードにある際よりも、低電力モードにある際に、異なる音検出器又は音検出器の組み合わせを使用する。（ボイストリガは、多くの異なるモード又は作動状態を可能としてもよく、それぞれは異なる量の電力を使用してもよく、異なる実施態様は、それら特有の設計にしたがってそれらを使用する。）

10

20

30

40

50

【 0 1 0 7 】

一方で、この機器がいくつかの他のコンテキストにある場合は、コンテキストが維持されている限り、ボイストリガは有効となる（又は、異なるモードにて動作される）。例えば、いくつかの実施態様では、ボイストリガは、電源に接続されている、電話が所定の向きにある（例えば、表面の上で上向きに置かれている）、所定の期間中（例えば、8:00 AMと10:00 PMとの間）、機器が移動中である及び／又は車内にある（例えば、GPS信号、BLUETOOTH（登録商標）接続、又は車両と接続中、等に基づく）、等、の間は有効に維持される。機器が車内にある際の裏付けを検出する態様は、そのすべてにおいて参照により本明細書に組み込まれている、本願の譲受人に帰属する、米国仮特許出願第61/657,744号に更に詳細に説明されている。特定のコンテキストを判定する方法の種々の具体例を以下に提供する。様々な実施形態では、異なる技術及び／又は情報源を使用して、これら及び他のコンテキストが検出される。

10

【 0 1 0 8 】

上述したように、ボイストリガシステム400が有効であるか否か（例えば、リスニング中）は、機器の物理的向きに依存し得る。いくつかの実施態様では、ボイストリガは、機器が表面の上で「上向き」に置かれている場合に有効となっており（例えば、ディスプレイ及び／又はタッチスクリーンの表面が見える状態）、及び／又は「下向き」の場合に無効となっている。これは、設定メニュー、スイッチ、又はボタンの操作を必要とすることなく、ボイストリガを有効及び／又は無効にする容易な方法をユーザに提供する。いくつかの実施態様では、この機器は、表面の上で上向きに置かれているか又は下向きに置かれているかを、光センサ（例えば、機器104の前後面への入射光の差異に基づいて）、近接センサ、磁気センサ、加速度計、ジャイロスコープ、チルトセンサ、カメラ、等を使用して検出する。いくつかの実施態様では、他の作動モード、設定、パラメータ、又は嗜好は、機器の向き及び／又は位置に影響される。いくつかの実施態様では、ボイストリガがリスンしている特定のトリガ音、単語、又はフレーズは、機器の向き及び／又は位置に依存する。例えば、いくつかの実施態様では、ボイストリガは、機器が1つの向きにある（例えば、表面の上で上向きに置かれている）場合は、第1のトリガ語、フレーズ、又は音をリスンし、機器が別の向きにある（例えば、下向きに置かれている）場合は、異なるトリガ語、フレーズ、又は音をリスンする。いくつかの実施態様では、下向きに対するトリガフレーズは、上向きに対するものよりも長く、及び／又はより複雑なものである。したがって、ユーザは、他の人が周りにいる場合、又は雑音の多い環境にいる場合は、機器を下向きに置くことができ、より短い、又はより簡易なトリガ語に対してより頻繁に起こるであろう不正の承認も削減しながら、ボイストリガを依然として作動できるようにする。1つの具体例として、上向きトリガフレーズは「SIRIへ」としてもよく、一方で下向きトリガフレーズは「SIRIへ、アンドリューです、起動してください」としてもよい。より長いトリガフレーズはまた、音検出器及び／又は音声認証器に対して、処理及び／又は分析のためにより長い音声サンプルを提供し、したがってボイストリガの精度を上げ、不正の承認を減らす。

20

30

【 0 1 0 9 】

いくつかの実施態様では、この機器104は、この機器が車内にあるか検出する（例えば、自動車）。ボイストリガは、機器及び／又はスピーチベースのサービスの操作に必要な物理的双方向作用を減らすことを助けるため、ユーザが車両内にいる場合にスピーチベースのサービスを呼び出すことに、特に有用である。実際に、音声ベースのデジタルアシスタントの利点の1つは、機器を見てタッチ操作することが不可能又は危険である場合に、タスクを行うよう使用できることである。したがって、デジタルアシスタントを呼び出すためにユーザが機器をタッチする必要がないよう、機器が車内にある場合に、ボイストリガを使用してもよい。いくつかの実施態様では、この機器は、BLUETOOTH（登録商標）通信（又は、他のワイヤレス通信）又は、ドッキングコネクタ若しくはケーブルなどのものを通して、車両に接続されている及び／又は車両とペアリングされていることを検出することで、車内にあることを判定する。いくつかの実施態様では、この機器は、

40

50

(例えば、GPS受信機、加速度計、及び/又はジャイロスコープを使用して)機器の位置及び/又は速度を判定することで、車内にあることを判定する。例えば、毎時20マイル超で移動しており、道に沿って移動することを判定していることから、機器が車内にあると思われると判定し、ボイストリガは続いて有効に維持される、及び/又は、高電力状態若しくは高感度状態に維持される。

【0110】

いくつかの実施態様では、この機器は、実質的な閉鎖空間内にあるか判定することで、機器が(例えば、ポケット、財布、バッグ、引き出し、等内に)保管されているか検出する。いくつかの実施態様では、この機器は、光センサ(例えば、専用周囲光センサ及び/又はカメラ)を使用して、保管されていることを判定する。例えば、いくつかの実施態様では、この機器は、光センサが微弱な光を検出する又は光を検出しない場合に、おそらくは保管されている。いくつかの実施態様では、時刻及び/又は機器の位置も考慮される。例えば、高い光レベルが予想される場合(例えば、日中)に、光センサが低い光レベルを検出すると、機器は保管されており、ボイストリガシステム400は不要としてもよい。したがって、このボイストリガシステム400は、低電力状態又はスタンバイ状態となる。いくつかの実施態様では、機器の対向する面に位置するセンサによって検出された光の差異は、その位置、したがって保管されているか否かの判定に使用され得る。具体的には、この機器が、ポケット又はバッグ内に保管されておらず、テーブル又は表面に置いてある場合に、ユーザはボイストリガを有効にするよう試みようとする。この機器がテーブル又はデスクなどの表面の上に下向き(又は、上向き)に置かれている場合、機器の片面は塞がれ、他の表面が周囲光に露出する一方で、その片面には微弱な光しかあたらない又は光がまったくあたらない。したがって、機器の前面及び背面の光センサが、著しく異なる光レベルを検出する場合、機器は保管されていないと判定する。一方で、対向する面にある光センサが、同じ又は類似の光レベルを検出する場合、機器は実質的な閉鎖空間内に保管されていると判定する。また、双方の光センサが、日中に低い光レベルを検出する場合(又は、電話が明るい環境にあると機器が予想する場合)、機器は保管されていると大きな信頼値を伴って判定する。

【0111】

いくつかの実施態様では、他の技術を使用して(光センサの代り又はこれに加えて)、機器が保管されているか判定する。例えば、いくつかの実施態様では、この機器は、スピーカ又はトランスデューサ(例えば、スピーカ228)から、1つ以上の音(例えば、トーン、クリック、ピン、等)を発し、1つ以上のマイクロフォン又はトランスデューサ(例えば、マイクロフォン230)を監視し、省略音(単数又は複数)のエコーを検出する。(いくつかの実施態様では、この機器は、人の可聴範囲外の音などの、不可聴信号を発する。)エコーから、この機器は、周辺環境の特性を判定する。例えば、相対的に広い環境(例えば、室内又は車内)は、相対的に狭い、閉鎖環境(例えば、ポケット、財布、バッグ、引き出し、等)とは異なって音を反射する。

【0112】

いくつかの実施態様では、ボイストリガシステム400は、(ボイストリガ及び/又はスピーチベースのサービスを有する他の機器などの)他の機器に近い場合は、他の機器から離れている場合とは異なって動作されている。これは、例えば、多くの機器が互いに近い場合に、一人がトリガ語を発言すると、他の周辺の機器が同様にトリガされないよう、ボイストリガシステム400を止める又は感度を下げることに有用とすることができる。いくつかの実施態様では、機器は、RFID、近接通信、赤外線信号/音響信号、等を使用して、他の機器への近接を判定する。上述したように、ボイストリガは、ユーザが運転中などの、機器がハンズフリーモードで動作されている際に、特に有用である。そのような場合では、ユーザはしばしば、有線ヘッドセット又は無線ヘッドセット、スピーカ及び/又はマイクロフォン付き腕時計、車両内蔵マイクロフォン及びスピーカ、等、などの、外部オーディオシステムを使用し、機器を顔の近くに持つ必要なく、電話を掛ける又はテキスト入力を口述する。例えば、無線ヘッドセット及び車両オーディオシステムは、BL

U E T O O T H (登録商標) 通信、又は任意の他の適切なワイヤレス通信を使用して、電子機器に接続してもよい。しかしそれは、無線アクセサリとオープンなオーディオチャネルを維持するために必要な電力を理由として、無線オーディオアクセサリを介して受信した音声を監視するボイストリガにとって非効率的となり得る。特に、無線ヘッドセットは、数時間の継続的な通話時間を提供するために、十分な電力をそのバッテリーに保持することができ、したがって、周囲の音声を単に監視し、潜在的なトリガ音を待つために使用する代りに、ヘッドセットが実際の通信に必要な際のために、バッテリーを蓄えておくことに好適である。そのうえ、有線の外部ヘッドセットアクセサリは、オンボードマイクロフォン単独よりも過大な電力を必要とする場合があり、ヘッドセットのマイクロフォンを有効に維持することは、機器のバッテリー充電電力を消耗する。これは、無線ヘッドセット又は有線ヘッドセットにより受け取った周囲音声は通常、大部分が無音又は無関係の音で構成されることを考慮すると、特にあてはまる。したがって、いくつかの実施態様では、ボイストリガシステム 400 は、機器が外部マイクロフォン (有線又は無線) に連結されていても、機器上のマイクロフォン 230 からの音声を監視する。続いて、ボイストリガがトリガ語を検出すると、この機器は、外部マイクロフォンとの有効なオーディオリンクを開始し、続いての (音声ベースのデジタルアシスタントへのコマンドなどの) 音入力を、機器上のマイクロフォン 230 ではなく、外部のマイクロフォンを介して受け取る。特定の条件が満たされるとしたら、有効な通信リンクが、(機器 104 に有線又は無線を介して通信可能に連結されてもよい) 外部オーディオシステム 416 と機器との間に維持され得、ボイストリガシステム 400 は、機器上のマイクロフォン 230 の代りに (又は、これに加えて)、外部オーディオシステム 416 を介して、トリガ音をリッスンし得る。例えば、いくつかの実施態様では、電子機器及び / 又は外部オーディオシステム 416 の (例えば、各機器上の加速度計、ジャイロスコープ、等により判定される) 動きの特性は、ボイストリガシステム 400 が、機器上のマイクロフォン 230 又は外部マイクロフォン 418 を使用して背景音を監視すべきかを判定することに使用されている。具体的には、機器と外部オーディオシステム 416 との間の動きの差異は、外部オーディオシステム 416 が実際に使用中かについての情報を提供する。例えば、機器及び無線ヘッドセットの双方が、実質的に等しく動いている場合 (又は、動いていない場合)、ヘッドセットは使用されていない又は着用されていないと判定されてもよい。これは、例えば、双方の機器は互いに近く、アイドル状態であるために起こり得る (例えば、テーブル上に置いてある、又はポケット、バッグ、財布、引き出し、等に入れてある)。それに応じて、これらの条件下では、ボイストリガシステム 400 は、ヘッドセットが実際に使用されていることがありそうもないため、機器上のマイクロフォンを監視する。無線ヘッドセットと機器との間の動きに差異がある場合、ユーザがヘッドセットを着用していることを判定する。これらの条件は、例えば、機器が置かれているため (例えば、表面の上又はバッグ内)、(装着者が相対的に静止していても、少なくとも少量の動きが起こり得るであろう) ユーザの頭にヘッドセットが着用されている間に発生する場合がある。これらの条件下では、ヘッドセットが着用されていると考えられるため、ボイストリガシステム 400 は、有効な通信リンクを維持し、ヘッドセットのマイクロフォン 418 を、機器上のマイクロフォン 230 の代りに、(又は、これに加えて) 監視する。この技術は機器及びヘッドセットの動きにおける差異に注目するため、双方の機器に共通の動きは相殺される。これは、例えば、機器 (例えば、携帯電話) が、カップホルダ内、空席のシート上、又はユーザのポケット内にあり、ヘッドセットがユーザの頭に着用されている、移動中の車両内でユーザがヘッドセットを使用している場合に有用となり得る。双方の機器に共通の動きが相殺されると (例えば、車両の動き)、機器に比較されるようなヘッドセットの相対運動が (もし存在すれば) 判定され得、ヘッドセットがおそらくは使用中であるか (又は、ヘッドセットが着用されていないか) 判定する。上記説明は無線ヘッドセットについて言及するが、類似の技術が同様に、有線ヘッドセットに適用される。

【0113】

人の声は大きく異なるため、ボイストリガを調律して特定ユーザの声を認識することに

10

20

30

40

50

おけるその精度を改善することは、必要又は有用となり得る。また、人の声は、例えば、病気、加齢又はホルモンの変化による自然な声の変化、等により、時間とともに変化し得る。したがって、いくつかの実施態様では、ボイストリガシステム 400 は、特定のユーザ又はユーザグループに対する、その音声及び / 又は音認識プロファイルに順応することができる。上述したように、音検出器（例えば、音種類検出器 404 及び / 又はトリガ音検出器 406）は、音入力のリプレゼンテーション（例えば、ユーザにより提供された音又は発言）を、1つ以上のリファレンスレプリゼンテーションと比較するよう構成されてもよい。例えば、入力レプリゼンテーションが、リファレンスレプリゼンテーションに、所定の信頼値レベルにて一致する場合、音検出器は、音入力が所定の種類の音に対応する（例えば、音種類検出器 404）、又は音入力が所定の内容を含む（例えば、トリガ音検出器 406）ことを判定する。ボイストリガシステム 400 を調律するため、いくつかの実施態様では、この機器は、入力レプリゼンテーションが比較される、リファレンスレプリゼンテーションを調整する。いくつかの実施態様では、リファレンスレプリゼンテーションは、音声登録手順又は「トレーニング」手順の一部として調整（又は、作成）されており、ここでユーザは、機器がリファレンスレプリゼンテーションを調整（又は、作成）できるよう、トリガ音を数回出力する。この機器は続いて、その人の実際の声を使用して、リファレンスレプリゼンテーションを作成する。

10

【0114】

いくつかの実施態様では、この機器は、通常の使用条件下にて受け取られるトリガ音を使用し、リファレンスレプリゼンテーションを調整する。（例えば、トリガリング基準のすべてを満たす音入力が見つかった）正常なボイストリガリング事象の後、例えば、この機器は、音入力からの情報を使用し、リファレンスレプリゼンテーションを調整及び / 又は調律する。いくつかの実施態様では、トリガリング基準のすべて又は一部を、特定の信頼値レベルで満たすことが判定された音入力のみが、リファレンスレプリゼンテーションの調整に使用されている。したがって、音入力トリガ音に対応する又はこれを含むことについて、ボイストリガの信頼性が低い場合、その音声入力は、リファレンスレプリゼンテーションを調整する目的に対して無視されてもよい。一方で、いくつかの実施態様では、ボイストリガシステム 400 を低い信頼値にて満たした音入力は、リファレンスレプリゼンテーションの調整に使用されている。

20

【0115】

いくつかの実施態様では、機器 104 は、ますます多くの音入力を受け取られると、ユーザの音声の経年によるわずかな変化に順応できるよう、（これら又は他の技術を使用して）リファレンスレプリゼンテーションを繰り返し調整する。例えば、いくつかの実施態様では、機器 104（及び / 又は関連機器若しくはサービス）は、それぞれの正常なトリガリング事象の後で、リファレンスレプリゼンテーションを調整する。いくつかの実施態様では、機器 104 は、それぞれの正常なトリガリング事象に関連する音入力を分析し、リファレンスレプリゼンテーションが、その入力（例えば、特定の条件が満たされている場合）に基づいて調整されるべきか判定し、そうすることが適切な場合のみ、リファレンスレプリゼンテーションを調整行う。いくつかの実施態様では、機器 104 は、リファレンスレプリゼンテーションの移動平均を長期にわたって維持する。いくつかの実施態様では、ボイストリガシステム 400 は、（例えば、音検出器のうちの1つ以上により判定されるように）トリガリング基準のうちの1つ以上を満たさない音を検出するが、これは正規ユーザによって、そうするように実際に試みられてもよい。例えば、ボイストリガシステム 400 は、「S I R I へ」などのトリガフレーズに応答するよう構成されてもよいが、ユーザの音声（例えば、病気、加齢、アクセント / 音調の変化、等により）変化すると、ボイストリガシステム 400 は、機器を有効にするユーザの試みを認識しなくともよい。（これはまた、ボイストリガシステム 400 がデフォルト条件に設定されている、及び / 又は、ユーザが初期化していない、又は、ユーザの音声に対して、ボイストリガシステム 400 をカスタマイズするトレーニング手順を行っていない、などの、そのユーザの特定の音声に対して、ボイストリガシステム 400 が適切に調律されていなかった場合に

30

40

50

生じてもよい。)ボイストリガシステム400が、ボイストリガを有効にするユーザの第1の試みに応答しない場合、ユーザはおそらくトリガフレーズを繰り返す。この機器は、これらの繰り返された音入力互いに似ており、及び/又はこれらがトリガフレーズに似ていることを、(ボイストリガシステム400にスピーチベースのサービスを有効にさせるように充分似ていなくとも)検出する。そのような条件が満たされると、この機器は、音入力が、ボイストリガシステム400を有効にする正当な試みに対応することを判定する。それに応じて、いくつかの実施態様では、ボイストリガシステム400は、それらの受け取った音入力を使用し、ユーザによる類似の発言が、正当なトリガとして将来に承認されるよう、1つ以上の態様のボイストリガシステム400を調整する。いくつかの実施態様では、これらの音入力は、特定の条件又は条件の組み合わせが満たされる場合にのみ、ボイストリガシステム400を順応させるよう使用されている。例えば、いくつかの実施態様では、音入力は、所定の数の音入力が連続して受け取られた場合に(例えば、2、3、4、5、又は任意の他の適切な数)、音入力がリファレンスレプリゼンテーションに十分類似する場合に、音入力が互いに充分類似する場合に、音入力が互いに近い場合に(例えば、所定の期間内、及び/又は、所定のインターバルにて、又は、その付近で受け取られた場合に)、及び/又は、これら又は他の条件の任意の組み合わせにて、ボイストリガシステム400を順応させるために使用されている。場合によっては、ボイストリガシステム400は、スピーチベースのサービスの(例えば、ボタン又はアイコンを押すことでの)手動での開始が続く、トリガリング基準のうちの1つ以上を満たさない、1つ以上の音入力を検出してもよい。いくつかの実施態様では、ボイストリガシステム400は、音入力を受け取ったすぐ後にスピーチベースのサービスが開始されたことにより、音入力が実際に、失敗したボイストリガリングの試みに対応することを判定する。それに応じて、ボイストリガシステム400は、上述したように、それらの受け取った音入力を使用し、ユーザによる発言が、正当なトリガとして将来に承認されるよう、1つ以上の態様のボイストリガシステム400を調整する。

【0116】

上記の順応技術が、リファレンスレプリゼンテーションを調整することに言及する一方で、トリガ音検出技術の他の態様は、リファレンスレプリゼンテーションを調整することに加えて、又はその代りに、同じ又は類似の方法で調整されてもよい。例えば、いくつかの実施態様では、この機器は、音入力の特定の周波数又は周波数範囲に集中及び/又はこれらを削減するなどの、音入力がどのようにしてフィルタリングされるか、及び/又はどのようなフィルタが音入力に適用されるかを調整する。いくつかの実施態様では、この機器は、入力レプリゼンテーションとリファレンスレプリゼンテーションとの比較に使用されているアルゴリズムを調整する。例えば、いくつかの実施態様では、入力レプリゼンテーションとリファレンスレプリゼンテーションとの間の差異の判定に使用される数学関数の1つ以上の項が、変更、追加、又は除去されるか、又は異なる数学関数に置き換えられている。いくつかの実施態様では、上記のそれらなどの順応技術は、ボイストリガシステム400が提供できる、又は提供するよう構成されているものよりも多くのリソースを必要とする。特に、音検出器は、リファレンスレプリゼンテーション及び/又は音検出アルゴリズム(又は、任意の他の適切な態様のボイストリガシステム400)の反復順応を行うために必要な、プロセッサ、データ、又はメモリの量又は種類、又はそれらへのアクセスを有しなくともよい。したがって、いくつかの実施態様では、上記の順応技術のうちの一つ以上は、アプリケーションプロセッサ(例えば、プロセッサ(単数又は複数)204)などの、より強力なプロセッサによって、又は、異なる機器(例えば、サーバシステム108)によって実行されている。しかし、ボイストリガシステム400は、アプリケーションプロセッサがスタンバイモードにあっても作動するよう設計されている。したがって、ボイストリガシステム400の順応に使用される音入力は、アプリケーションプロセッサが有効でなく、音入力を処理できない場合に受け取られている。それに応じて、いくつかの実施態様では、音入力は、受け取り後に更に処理及び/又は分析され得るよう、機器によって記憶されている。いくつかの実施態様では、音入力は、オーディオサブシステ

10

20

30

40

50

ム 2 2 6 のメモリバッファ 4 1 4 内に記憶されている。いくつかの実施態様では、音入力
は、ダイレクトメモリアクセス (DMA) 技術 (例えば、アプリケーションプロセッサを
起動する必要なく、データをコピー又は移動できるよう、DMA エンジンを使用すること
を含む) を使用して、システムメモリ (例えば、メモリ 2 5 0、図 2) 内に記憶されてい
る。記憶された音入力は続いて、アプリケーションプロセッサ (又は、サーバシステム 1
0 8、若しくは別の適切な機器) に、起動後、アプリケーションプロセッサが、上記の順
応技術のうちの 1 つ以上を実行できるよう、提供されている又はこれによってアクセスさ
れている。いくつかの実施態様では。

【 0 1 1 7 】

図 5 ~ 図 7 は、特定の実施態様に係る、ボイストリガを動作させるための方法を表すフ
ローダイアグラムである。この方法は、任意で、コンピュータメモリ又は永続的コンピ
ュータ可読記憶媒体 (例えば、クライアント機器 1 0 4 のメモリ 2 5 0、デジタルアシスタ
ントシステム 3 0 0 に関連するメモリ 3 0 2) 内に記憶されている命令によって管理され
ており、サーバシステム 1 0 8、及び / 又はユーザ機器 1 0 4 a を含むがこれらに限定し
ない、デジタルアシスタントシステムの 1 つ以上のコンピュータシステムの 1 つ以上のプ
ロセッサによって実行されている。コンピュータ可読記憶媒体は、磁気又は光ディスク記
憶装置、フラッシュメモリなどの固体記憶装置、又は他の不揮発性メモリデバイス (単数
又は複数) を含む得る。コンピュータ可読記憶媒体上に記憶されたコンピュータ可読命令
は、ソースコード、アセンブリ言語コード、オブジェクトコード、又は、1 つ以上のプロ
セッサによって解釈実行される他の命令フォーマットのうちの 1 つ以上を含んでよい。様
々な実施態様では、各図に示されている、それぞれの方法のいくつかの動作を組み合わ
せてもよいし、及び / 又はいくつかの動作の順序を、順序から変更してもよい。また、いく
つかの実施態様では、個別の図に示す及び / 又は個別の方法に関連して説明された動作は
、他の方法を形成するよう組み合わせられてもよく、同じ図及び / 又は同じ方法に関連して
説明された動作は、異なる方法に分けられてもよい。そのうえ、いくつかの実施態様では
、方法における 1 つ以上の動作は、例えば、自然言語処理モジュール 3 3 2、ダイアログ
フロー処理モジュール 3 3 4、オーディオサブシステム 2 2 6、ノイズ検出器 4 0 2、音
種類検出器 4 0 4、トリガ音検出器 4 0 6、スピーチベースのサービス 4 0 8、及び / 又
は任意のそれらのサブモジュールを含む、デジタルアシスタントシステム 3 0 0 及び / 又
は電子機器 (例えば、ユーザ機器 1 0 4) のモジュールによって実行されている。図 5 は
、いくつかの実施態様に係る、ボイストリガシステムを動作させる方法 5 0 0 を示す (例
えば、図 4 のボイストリガシステム 4 0 0、図 4)。いくつかの実施態様では、方法 5 0
0 は、1 つ以上のプロセッサ及び、1 つ以上のプロセッサ (例えば、電子機器 1 0 4) に
よって実行される命令を記憶するメモリを含む電子機器にて実行されている。この電子機
器は、音入力を受け取る (5 0 2)。この音入力は、発言 (例えば、単語、フレーズ、又
は文章)、人による発音 (例えば、口笛、舌打ち、指鳴らし、拍手、等)、又は任意の他
の音 (例えば、電子的に生じるチャープ音、メカニカルノイズメーカー、等) に対応して
もよい。いくつかの実施態様では、電子機器は、オーディオサブシステム 2 2 6 (例えば
、コーデック 4 1 0、オーディオ DSP 4 1 2、及びバッファ 4 1 4、同様に、図 4 に関
連して説明されるマイクロフォン 2 3 0 及び 4 1 8、を含む) を介して音入力を受信する
。

【 0 1 1 8 】

いくつかの実施態様では、電子機器は、音入力 that 所定の条件を満たすか判定する (5 0
4)。いくつかの実施態様では、電子機器は、時間領域分析を音入力に適用し、音入力
が所定の条件を満たすか判定する。例えば、電子機器は、音入力のある期間にわたって分析
し、音振幅が所定のレベルに達するか判定する。いくつかの実施態様では、閾値は、音入
力の振幅 (例えば、音量) が、所定の閾値を満たす及び / 又は超える場合に満たされる。
いくつかの実施態様では、それは、音入力 that 所定の閾値を所定の時間に満たす及び / 又
は超える場合に満たされる。以下に更に詳細に説明するように、いくつかの実施態様では、
音入力 that 所定の条件を満たすか判定すること (5 0 4) は、第 3 の音検出器 (例えば、ノ

10

20

30

40

50

イズ検出器 402) によって行われている。(第3の音検出器がこの場合に使用され、この音検出器を他の音検出器(例えば、以下に説明する第1の音検出器及び第2の音検出器)から区別し、いずれの作動位置又は音検出器の順序を必ずしも示さない。)

【0119】

電子機器は、音入力が所定の種類の音に対応するか判定する(506)。上述したように、音は、特定の識別可能な音の特性に基づいて、様々な「種類」に分類されている。音入力が所定の種類に対応するか判定することは、音入力が特定の種類の特性を含む又は示すか判定することを含む。いくつかの実施態様では、所定の種類の音は人の声である。そのような実施態様では、音入力が人の声に対応するか判定することは、音入力が人の声の周波数特性を含むか判定することを含む(508)。以下に更に詳細に説明するように、いくつかの実施態様では、音入力が所定の種類の音に対応するか判定すること(506)は、第1の音検出器(例えば、音種類検出器404)によって行われている。音入力が所定の種類の音に対応することを判定すると、電子機器は、音入力が所定の内容を含むか判定する(510)。いくつかの実施態様では、所定の内容は、1つ以上の所定の音素に対応する(512)。いくつかの実施態様では、1つ以上の所定の音素は、少なくとも1つの単語を構成する。いくつかの実施態様では、所定の内容は、音である(例えば、口笛、クリック、又は拍手)。いくつかの実施態様では、以下に説明するように、音入力が所定の内容を含むか判定すること(510)は、第2の音検出器(例えば、トリガ音検出器406)によって行われている。

【0120】

音入力が所定の内容を含むことを判定すると、電子機器は、スピーチベースのサービスを開始する(514)。いくつかの実施態様では、スピーチベースのサービスは、上に詳述するように、音声ベースのデジタルアシスタントである。いくつかの実施態様では、スピーチベースのサービスは口述サービスであり、ここでスピーチ入力はテキストに変化され、(例えば、電子メール、テキストメッセージ、ワードプロセッシング、又はノート取りアプリケーション、等の)テキスト入力フィールドに含まれる及び/又は表示される。スピーチベースのサービスが音声ベースのデジタルアシスタントである実施態様では、音声ベースのデジタルアシスタントが開始されると、ユーザにプロンプト(例えば、音又はスピーチプロンプト)が発行され、ユーザが、音声入力及び/又はコマンドをデジタルアシスタントに提供できることが示される。いくつかの実施態様では、音声ベースのデジタルアシスタントを開始することは、アプリケーションプロセッサ(例えば、プロセッサ(単数又は複数)204、図2)を有効にすること、1つ以上のプログラム又はモジュール(例えば、デジタルアシスタントクライアントモジュール264、図2)を開始すること、及び/又は、リモートサーバ若しくは機器(例えば、デジタルアシスタントサーバ106、図1)への接続を創設すること、を含む。

【0121】

いくつかの実施態様では、電子機器は、音入力が特定ユーザの声に対応するか判定する(516)。例えば、機器の正規ユーザの音声に対応するか判定するために、1つ以上の音声認証技術が音入力に適用されている。音声認証技術は、上に詳述されている。いくつかの実施態様では、音声認証は、音検出器のうちの1つ(例えば、トリガ音検出器406)によって行われている。いくつかの実施態様では、音声認証は、専用音声認証モジュール(いずれの適切なハードウェア及び/又はソフトウェアを含む)によって行われている。いくつかの実施態様では、音声ベースのサービスは、音入力が所定の内容を含み、音入力 that 特定ユーザの声に対応することの判定に応じて開始されている。したがって、例えば、音声ベースのサービス(例えば、音声ベースのデジタルアシスタント)は、トリガ語又はフレーズが正規ユーザによって話された場合にのみ開始される。これは、サービスが非正規ユーザによって呼び出され得る可能性を減らし、複数の電子機器が近接している場合に、一人のユーザのトリガ音の発言が、別のユーザのボイストリガを起動しないため、特に有用とできる。

【0122】

スピーチベースのサービスが音声ベースのデジタルアシスタントであるいくつかの実施態様では、音入力が入力された内容を含むが、特定ユーザの声に対応しないことを判定することに応じて、音声ベースのデジタルアシスタントは、限定アクセスモードで開始されている。いくつかの実施態様では、限定アクセスモードは、デジタルアシスタントが、デジタルアシスタントがさもなければ提供し得るデータ、サービス、及び／又は機能のサブセットのみにアクセスできるようにする。いくつかの実施態様では、限定アクセスモードは、（例えば、デジタルアシスタントの非正規ユーザが、カレンダー、タスクリスト、連絡先、写真、電子メール、テキストメッセージ、等からのデータにアクセスできないよう）、書き込みのみモードに対応する。いくつかの実施態様では、限定アクセスモードは、スピーチベースのサービスのサンドボックス化されたインスタンスに対応し、スピーチベースのサービスが、機器 104（図 2）、又は任意の他の機器（例えば、図 3 A のユーザデータ 348、これは、図 1 のサーバシステム 108 などの、リモートサーバ上に記憶されてもよい）上のユーザデータ 266 などの、ユーザのデータから読まない又はこれらに書き込まないようにする。

【0123】

いくつかの実施態様では、音入力が入力された内容を含み、音入力が入力された特定ユーザの声に対応することの判定に応じて、音声ベースのデジタルアシスタントは、特定ユーザの名前を含むプロンプトを出力する。例えば、特定のユーザが音声認証を介して識別されると、音声ベースのデジタルアシスタントは、トーン、ビープ、又は非専有音声プロンプトなどの、より一般的なプロンプトの代りに、「Peter さん、ご用は何ですか？」などのプロンプトを出力してもよい。上述したように、いくつかの実施態様では、第 1 の音検出器は、音入力が入力された種類の音に対応するか判定し（ステップ 506 にて）、第 2 の音検出器は、音検出器が入力された内容を含むか判定する（ステップ 510 にて）。いくつかの実施態様では、第 1 の音検出器は、例えば、第 1 の音検出器が、第 2 の音検出器よりも、プロセッサ負荷の少ない技術を使用するため、第 2 の音検出器よりも少ない電力を動作中に消費する。いくつかの実施態様では、第 1 の音検出器は音種類検出器 404 であり、第 2 の音検出器はトリガ音検出器 406 であり、双方とも図 4 に関連して上述されている。いくつかの実施態様では、これらの動作中は、第 1 の音検出器及び／又は第 2 の音検出器は、図 4 に関連して上述されるように、デューティサイクルに従って、オーディオチャネルを定期的に監視する。

【0124】

いくつかの実施態様では、第 1 の音検出器及び／又は音検出器は、音入力の周波数領域分析を行う。例えば、これらの音検出器は、ラプラス変換、Z 変換、又はフーリエ変換を行い、周波数スペクトルを生成するか、又は音入力若しくはその一部のスペクトル密度を判定する。いくつかの実施態様では、第 1 の音検出器は、音入力が入力された人の声の特性である周波数（又は、人の声の特性である音入力の他の特徴、態様、又は特質）を含むか判定するよう構成されている、音声活動検出器である。

【0125】

いくつかの実施態様では、第 2 の音検出器は、第 1 の音検出器が入力された種類の音入力を検出するまでオフ又は無効である。それに応じて、いくつかの実施態様では、方法 500 は、音入力が入力された種類に対応することを判定することに応じて、第 2 の音検出器を起動することを含む。（他の実施態様では、第 2 の音検出器は、他の条件に応じて起動されているか、又は第 1 の音検出器からの判定に無関係に継続して動作されている。）いくつかの実施態様では、第 2 の音検出器を起動することは、ハードウェア及び／又はソフトウェア（例えば、回路、プロセッサ、プログラム、メモリ、等を含む）を有効にすることを含む。いくつかの実施態様では、第 2 の音検出器は、起動後の少なくとも所定の時間動作されている（例えば、有効にされ、オーディオチャネルを監視している）。例えば、音入力が入力された種類に対応する（例えば、人の声を含む）ことを、第 1 の音検出器が判定すると、第 2 の音検出器が動作し、音入力が入力された内容（例えば、トリガ語）も含むか判定する。いくつかの実施態様では、所定の時間は、所定の内容の持続時間に対応する。したがっ

10

20

30

40

50

て、所定の内容が、フレーズ「S I R Iへ」の場合、所定の時間は、フレーズが発言されたか判定するように十分に長くなる（例えば、1又は2秒、又はいずれの別の適切な持続時間）。所定の内容が、フレーズ「S I R Iへ、起動して助けてください」などの、より長いものの場合、所定の時間はより長いものとなる（例えば、5秒、又は別の適切な持続時間）。いくつかの実施態様では、第2の音検出器は、第1の音検出器が所定の種類に対応する音を検出する限りは作動する。そのような実施態様では、例えば、第1の音検出器が人の話し声を音入力内に検出する限りは、第2の音検出器は音入力処理し、所定の内容を含むか判定する。

【0126】

上述したように、いくつかの実施態様では、第3の音検出器（例えば、ノイズ検出器402）は、音入力処理が所定の条件を満たすか判定する（ステップ504にて）。いくつかの実施態様では、第3の音検出器は、第1の音検出器よりも少ない電力を作動中に消費する。いくつかの実施態様では、第3の音検出器は、図4に関して上述するように、デューティサイクルに従って、オーディオチャネルを定期的に監視する。また、いくつかの実施態様では、第3の音検出器は、音入力の時間領域分析を行う。いくつかの実施態様では、時間領域分析は、第2の音検出器によって適用される周波数領域分析よりも、プロセッサ負荷が少ないため、第3の音検出器は、第1の音検出器よりも少ない電力を消費する。

【0127】

第1の音検出器（例えば、音種類検出器404）による判定に応じて第2の音検出器（例えば、トリガ音検出器406）を起動することに関する上記説明と同様に、いくつかの実施態様では、第1の音検出器は、第3の音検出器（例えば、ノイズ検出器402）による判定に応じて起動されている。例えば、いくつかの実施態様では、音種類検出器404は、音入力処理が所定の条件（例えば、十分な持続時間に特定の音量を超える）を満たすことの、ノイズ検出器402による判定に応じて起動されている。いくつかの実施態様では、第1の音検出器を起動することは、ハードウェア及び/又はソフトウェア（例えば、回路、プロセッサ、プログラム、メモリ、等を含む）を有効にすることを含む。他の実施態様では、第1の音検出器は、他の条件に応じて起動されているか、又は継続して動作されている。いくつかの実施態様では、この機器は、音入力の少なくとも一部をメモリに記憶する（518）。いくつかの実施態様では、メモリは、オーディオサブシステム226のバッファ414である（図4）。記憶された音入力は、機器による、音入力の非リアルタイムの処理を可能とする。例えば、いくつかの実施態様では、音検出器のうちの1つ以上は、記憶された音入力を読み出し及び/又は受け取り、この記憶された音入力処理する。これは、上流の音検出器（例えば、トリガ音検出器406）が、オーディオサブシステム226による音入力の受け取り途中まで起動されない場合に、特に有用とされ得る。いくつかの実施態様では、音入力の記憶された一部は、スピーチベースのサービスが開始されると、このスピーチベースのサービスに提供される（520）。したがって、音入力の一部が受け取られるまで、スピーチベースのサービスが十分に作動しなくとも、このスピーチベースのサービスは、音入力の記憶された一部を複製、処理、又は、さもなければ作動できる。いくつかの実施態様では、音入力の記憶された一部は、電子機器の順応モジュールに提供されている。

【0128】

様々な実施態様では、ステップ（516）～（520）が、方法500内の異なる位置にて行われている。例えば、いくつかの実施態様では、ステップ（516）～（520）のうち1つ以上は、ステップ（502）と（504）との間、ステップ（510）と（514）との間、又は任意の他の適切な位置にて行われている。

【0129】

図6は、いくつかの実施態様に係る、ボイストリガシステムを動作させる方法600を示す（例えば、図4のボイストリガシステム400、図4）。いくつかの実施態様では、方法600は、1つ以上のプロセッサ及び、1つ以上のプロセッサ（例えば、電子機器104）によって実行される命令を記憶するメモリを含む電子機器にて実行されている。電

10

20

30

40

50

子機器は、所定の向きにあるか判定する（602）。いくつかの実施態様では、電子機器は、光センサ（カメラを含む）、マイクロフォン、近接センサ、磁気センサ、加速度計、ジャイロ스코プ、チルトセンサ、等を使用して、その向きを検出する。例えば、電子機器は、前面カメラのセンサに入射する光の量又は輝度を、背面カメラのセンサに入射する光の量又は輝度と比較することで、表面の上で下向き又は上向きに置かれているか判定する。前面カメラによって検出された量及び／又は輝度が、背面カメラによって検出されたそれよりも十分に大きい場合は、電子機器は、上向きにあると判定する。一方で、背面カメラによって検出された量及び／又は輝度が、前面カメラのそれよりも十分に大きい場合は、機器は、下向きにあると判定する。電子機器が所定の向きにあることを判定すると、この電子機器は、ボイストリガの所定のモードを有効にする（604）。いくつかの実施態様では、所定の向きは、機器のディスプレイスクリーンが実質的に水平で下向きにあることに対応し、所定のモードはスタンバイモードである。（606）。例えば、いくつかの実施態様では、スクリーンが下向きとなるように、スマートホン又はタブレットがテーブル又はデスクに置かれている場合、ボイストリガはスタンバイモード（例えば、電源切）となり、ボイストリガの意図しない起動を防止する。

10

【0130】

一方で、いくつかの実施態様では、所定の向きは、機器のディスプレイスクリーンが実質的に水平で上向きにあることに対応し、所定のモードはリスニングモードである（608）。したがって、例えば、スクリーンが上向きとなるように、スマートホン又はタブレットがテーブル又はデスクに置かれている場合、ボイストリガはリスニングモードとなり、トリガを検出すると、ユーザに応答できる。

20

【0131】

図7は、いくつかの実施態様に係る、ボイストリガを動作させる方法700を示す（例えば、ボイストリガシステム400、図4）。いくつかの実施態様では、方法700は、1つ以上のプロセッサ及び、1つ以上のプロセッサ（例えば、電子機器104）によって実行される命令を記憶するメモリを含む電子機器にて実行されている。電子機器は、ボイストリガ（例えば、ボイストリガシステム400）を第1のモードにて動作させる（702）。いくつかの実施態様では、第1のモードは、通常のリスニングモードである。

【0132】

電子機器は、電子機器のマイクロフォン及びカメラのうちの1つ以上が塞がっていることを検出することで、実質的な閉鎖空間内にあるか判定する（704）。いくつかの実施態様では、実質的な閉鎖空間は、ポケット、財布、バッグ、引き出し、グローブボックス、ブリーフケース、等を含む。

30

【0133】

上述したように、いくつかの実施態様では、この機器は、スピーカ又はトランスデューサから、1つ以上の音（例えば、トーン、クリック、ピン、等）を発し、1つ以上のマイクロフォン又はトランスデューサを監視し、省略音（単数又は複数）のエコーを検出することで、マイクロフォンが塞がっていることを検出する。例えば、相対的に広い環境（例えば、室内又は車内）は、相対的に狭い、実質的な閉鎖環境（例えば、財布又はポケット）とは異なって音を反射する。したがって、マイクロフォン（又は、音を発したスピーカ）が塞がっていることを、エコー（又は無エコー）に基づいて機器が検出すると、この機器は、実質的な閉鎖空間内にあると判定する。いくつかの実施態様では、この機器は、マイクロフォンが閉鎖空間特有の音を拾うことを検出することで、マイクロフォンが塞がっていることを検出する。例えば、機器がポケット内にある場合、マイクロフォンがポケットの繊維に接触又は近接することによる、特徴的な柔らかいノイズを、マイクロフォンは検出することができる。いくつかの実施態様では、この機器は、センサによる受光レベルに基づいて、又は焦点の合った画像を得ることができると判定することで、カメラが塞がっていることを検出する。例えば、高レベルの光が予想される時間（例えば、日中）に、カメラセンサが低レベルの光を検出する場合、この機器は、カメラが塞がっており、機器が実質的な閉鎖空間内にあることを判定する。別の例として、カメラは、そのセンサ上の

40

50

焦点の合った画像を取得するよう試みてもよい。通常、これは、カメラが非常に暗い場所（例えば、ポケット又はバックパック）にある場合、又は焦点を合わせようと試みる被写体に近すぎる（例えば、財布又はバックパック内）にある場合に難しくなる。したがって、カメラが焦点の合った画像を取得できない場合、機器が実質的な閉鎖空間内にあると判定する。

【 0 1 3 4 】

電子機器が実質的な閉鎖空間内にあることを判定すると、電子機器は、ボイストリガを第2のモードに切り換える（706）。いくつかの実施態様では、第2のモードは、スタンバイモードである（708）。いくつかの実施態様では、スタンバイモードにある場合、ボイストリガシステム400は、周囲の音声の監視を続けるが、ボイストリガシステム400をさもなければ起動するに関わらず、受け取った音に応答しない。いくつかの実施態様では、スタンバイモードでは、ボイストリガシステム400は無効となっており、音声を処理してトリガ音を検出することをしない。いくつかの実施態様では、第2のモードは、第1のモードとは異なるデューティサイクルに従って、ボイストリガシステム400の1つ以上の音検出器を動作させることを含む。いくつかの実施態様では、第2のモードは、第1のモードとは異なる組み合わせの音検出器を動作させることを含む。

10

【 0 1 3 5 】

いくつかの実施態様では、第2のモードは、より感度の高い監視モードに対応し、実質的な閉鎖空間内にあっても、ボイストリガシステム400がトリガ音を検出し、応答できるようにする。いくつかの実施態様では、ボイストリガが第2のモードに切り換わると、この機器は、電子機器のマイクロフォン及びカメラのうちの1つ以上が塞がっているか（例えば、ステップ（704）に関して上述の技術のいずれかを使用して）検出することで、電子機器が依然として実質的な閉鎖空間内にあるか定期的に判定する。この機器が実質的な閉鎖空間内にまだある場合、ボイストリガシステム400は第2のモードに維持される。いくつかの実施態様では、この機器が実質的な閉鎖空間内から動かされると、電子機器は、ボイストリガを第1のモードに戻す。

20

【 0 1 3 6 】

いくつかの実施によれば、図8は、前述のように本発明の原理にしたがって構成された電子機器800の機能ブロック図を示す。この機器の機能ブロックは、本発明の原理を実行するためのハードウェア、ソフトウェア、又はハードウェア及びソフトウェアの組み合わせにより実施することができる。図8に述べられる機能ブロックは、上記に述べたような本発明の原理を実施するために、組み合わせるか、又はサブブロックに分割することができるが当業者によって理解される。したがって、本明細書における説明は、あらゆる可能な組み合わせ若しくは分割、又は本明細書に述べられる更なる機能ブロックの定義を支持し得るものである。

30

【 0 1 3 7 】

図8に示すように、電子機器800は、音入力を受け取るよう構成された受音ユニット802を含む。電子機器800はまた、スピーチ受け取りユニット802に連結された処理ユニット806を含む。いくつかの実施態様では、処理ユニット806は、ノイズ検出部808、音種類検出部810、トリガ音検出部812、サービス開始部814、及び音声認証部816を含む。いくつかの実施態様では、ノイズ検出部808は、上記のノイズ検出器402に対応し、ノイズ検出器402に関する上述のいずれの作動を行うよう構成されている。いくつかの実施態様では、音種類検出部810は、上記の音種類検出器404に対応し、音種類検出器404に関する上述のいずれの作動を行うよう構成されている。いくつかの実施態様では、トリガ音検出部812は、上記のトリガ音検出器406に対応し、トリガ音検出器406に関する上述のいずれの作動を行うよう構成されている。いくつかの実施態様では、音声認証部816は、上記の音声認証モジュール428に対応し、音声認証モジュール428に関する上述のいずれの作動を行うよう構成されている。処理ユニット806は、音入力の少なくとも一部が所定の種類の音に対応するか（例えば、音種類検出部810で）判定し、音入力の少なくとも一部が所定の種類に対応することを

40

50

判定すると、音入力が所定の内容を含むか（例えば、トリガ音検出部 8 1 2 で）判定し、音入力が所定の内容を含むことを判定すると、スピーチベースのサービスを（例えば、サービス開始部 8 1 4 で）開始するよう構成されている。

【 0 1 3 8 】

いくつかの実施態様では、処理ユニット 8 0 6 はまた、音入力 that 所定の種類の音に対応するか判定する前に、音入力 that 所定の条件を満たす（例えば、ノイズ検出部 8 0 8 で）判定するよう構成されている。いくつかの実施態様では、処理ユニット 8 0 6 はまた、音入力 that 特定ユーザの声に対応するか（例えば、音声認証部 8 1 6 で）判定するよう構成されている。

【 0 1 3 9 】

いくつかの実施によれば、図 9 は、前述のように本発明の原理にしたがって構成された電子機器 9 0 0 の機能ブロック図を示す。この機器の機能ブロックは、本発明の原理を実行するためのハードウェア、ソフトウェア、又はハードウェア及びソフトウェアの組み合わせにより実施することができる。図 9 に述べられる機能ブロックは、上記に述べたような本発明の原理を実施するために、組み合わせるか、又はサブブロックに分割することができることが当業者によって理解される。したがって、本明細書における説明は、あらゆる可能な組み合わせ若しくは分割、又は本明細書に述べられる更なる機能ブロックの定義を支持しうるものである。

【 0 1 4 0 】

図 9 に示すように、電子機器 9 0 0 は、ボイストリガユニット 9 0 2 を含む。ボイストリガユニット 9 0 2 は、様々な異なるモードにて動作され得る。第 1 のモードでは、ボイストリガユニットは、音入力を受け取り、特定の基準を満たすか判定する（例えば、リスニングモード）。第 2 のモードでは、ボイストリガユニット 9 0 2 は、音入力を受け取らず、及び / 又は処理しない（例えば、スタンバイモード）。電子機器 9 0 0 はまた、ボイストリガユニット 9 0 2 に連結された処理ユニット 9 0 6 を含む。いくつかの実施態様では、処理ユニット 9 0 6 は、1 つ以上のセンサ（例えば、マイクロフォン、カメラ、加速度計、ジャイロスコープ、等を含む）及びモード切り換えユニット 9 1 0 を含んでもよい及び / 又は接点となってもよい環境検出部 9 0 8 を含む。いくつかの実施態様では、処理ユニット 9 0 6 は、電子機器のマイクロフォン及びカメラのうちの 1 つ以上が塞がっていることを検出することで、電子機器が実質的な閉鎖空間内にあるか（例えば、環境検出部 9 0 8 で）判定し、電子機器が実質的な閉鎖空間内にあることを判定すると、ボイストリガを第 2 のモードに（例えば、モード切り換えユニット 9 1 0 で）切り換えるよう構成されている。

【 0 1 4 1 】

いくつかの実施態様では、処理ユニットは、電子機器が所定の向きにあるか（例えば、環境検出部 9 0 8 で）判定し、電子機器が所定の向きにあることを判定すると、ボイストリガの所定のモードを（例えば、モード切り換えユニット 9 1 0 で）有効にするよう構成されている。

【 0 1 4 2 】

いくつかの実施によれば、図 1 0 は、前述のように本発明の原理にしたがって構成された電子機器 1 0 0 0 の機能ブロック図を示す。この機器の機能ブロックは、本発明の原理を実行するためのハードウェア、ソフトウェア、又はハードウェア及びソフトウェアの組み合わせにより実施することができる。図 1 0 に述べられる機能ブロックは、上記に述べたような本発明の原理を実施するために、組み合わせるか、又はサブブロックに分割することができることが当業者によって理解される。したがって、本明細書における説明は、あらゆる可能な組み合わせ若しくは分割、又は本明細書に述べられる更なる機能ブロックの定義を支持しうるものである。

【 0 1 4 3 】

図 1 0 に示すように、電子機器 1 0 0 0 は、ボイストリガユニット 1 0 0 2 を含む。ボイストリガユニット 1 0 0 2 は、様々な異なるモードにて動作され得る。第 1 のモードで

10

20

30

40

50

は、ボイストリガユニットは、音入力を受け取り、特定の基準を満たすか判定する（例えば、リスニングモード）。第2のモードでは、ボイストリガユニット1002は、音入力を受け取らず、及び/又は処理しない（例えば、スタンバイモード）。電子機器1000はまた、ボイストリガユニット1002に連結された処理ユニット1006を含む。いくつかの実施態様では、処理ユニット1006は、マイクロフォン及び/又はカメラ、及びモード切り換えユニット1010を含んでもよく、及び/又は接点となってもよい、環境検出部1008を含む。

【0144】

処理ユニット1006は、電子機器のマイクロフォン及びカメラのうちの1つ以上が塞がっていることを検出することで、電子機器が実質的な閉鎖空間内にあるか（例えば、環境検出部1008で）判定し、電子機器が実質的な閉鎖空間内にあることを判定すると、ボイストリガを第2のモードに（例えば、モード切り換えユニット1010で）切り換えるよう構成されている。上述の説明は、説明目的のために、特定の実施態様を参照して説明した。しかし、上述の例示的な説明は、網羅的であること、又は、開示される実施態様を正確な形態に限定することを意図していない。上述の教示を考慮すれば、多くの修正及び変形が可能である。実施態様は、開示された思想の理念と実際の応用を最も良く説明するために、選択して説明し、それによって、当業者が、考えられる特定の用途に適した様々な変更を加えて、それを最も良く利用することができるようにしたものである。

【0145】

当然のことながら、「第1」、「第2」などの用語は、様々な要素を説明するために本明細書で使用するができるが、これらの要素はこれらの用語によって限定されるべきではないことを理解されたい。これらの用語は、ある要素を別の要素と区別するためにのみ使用される。例えば、「第1の音検出器」のすべての出現について一貫して名前が変更され、「第2の音検出器」のすべての出現について一貫して名前が変更される限り、説明の意味を変えることなく、第1の音検出器は第2の音検出器と呼ぶことができ、同様に、第2の音検出器は第1の音検出器と呼ぶことができる。第1の音検出器及び第2の音検出器は双方とも音検出器であるが、同じ音検出器ではない。

【0146】

本明細書で使用される用語は、特定の実施態様を説明する目的のためであり、特許請求の範囲を限定することを意図するものではない。説明される実施態様の説明及び添付の特許請求の範囲で使用されるように、単数形の「a」、「an」、及び「the」は、文脈がそうではないことを明確に示さない限り、複数形もまた含むことを意図している。本明細書で使用される時に、用語「及び/又は」が、関連する列挙された項目のうちの1つ以上の任意のすべての可能な組合せを指し、かつこれを含むことをもまた理解されたい。用語「comprises（含む）」及び/又は「comprising（含む）」が、本明細書で使用される場合、述べられた特徴、整数、ステップ、動作、要素、及び/又は構成要素の存在を指定するが、1つ以上の他の特徴、整数、ステップ、動作、要素、構成要素、及び/又はそれらのグループの存在又は追加を除外しないことを更に理解されたい。本明細書で使用される時には、用語「if」は、文脈に応じて、「時」又は「際に」又は、先に述べた条件が正しいとの「判定に応じて」又は「判定にしたがって」、「検出に応じて」を意味するものと解釈できる。同様に、「[先に述べた条件が正しい]と判定される場合」又は「[先に述べた条件が正しい]場合」、又は「[先に述べた条件が正しい]時」というフレーズは、先に述べた条件が正しいとの「判定に際して」、「の判定に際して」又は「判定に応じて」又は「判定にしたがって」又は「検出した際に」又は「検出に応じて」ということを意味するものと解釈できる。

10

20

30

40

【図 1】

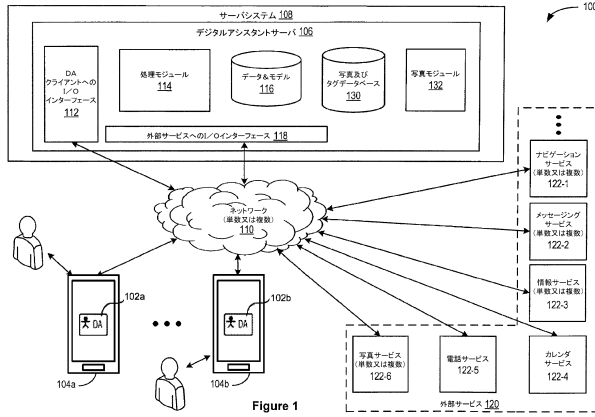


Figure 1

【図 2】

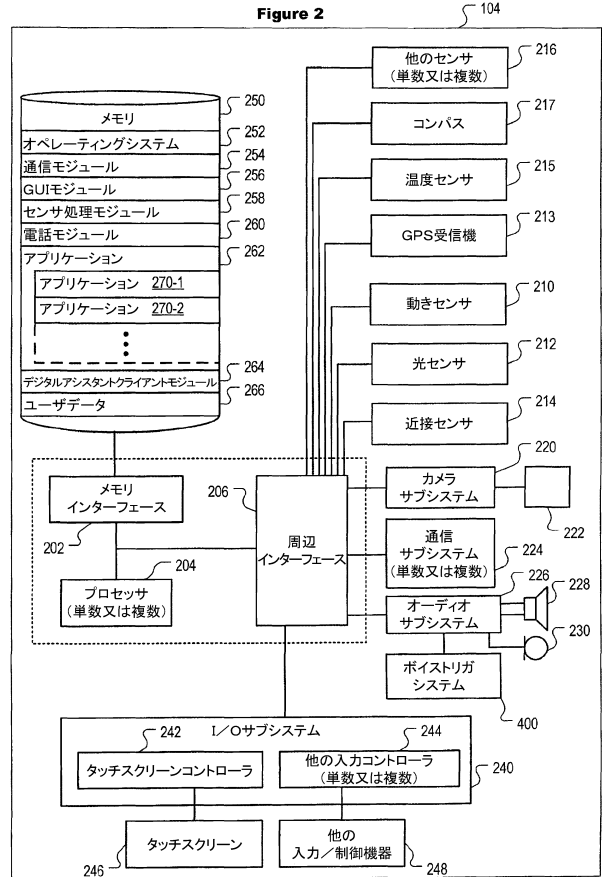
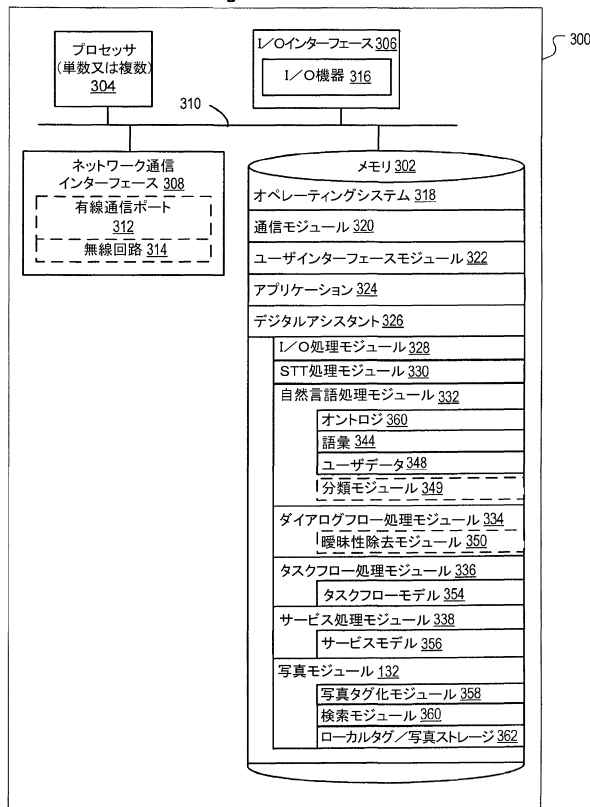


Figure 2

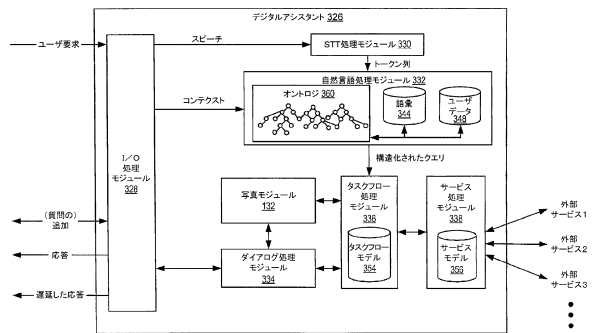
【図 3 A】

Figure 3A



【図 3 B】

Figure 3B



【図 3 C】

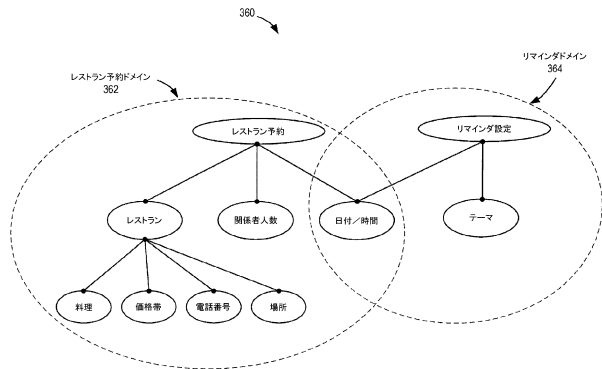
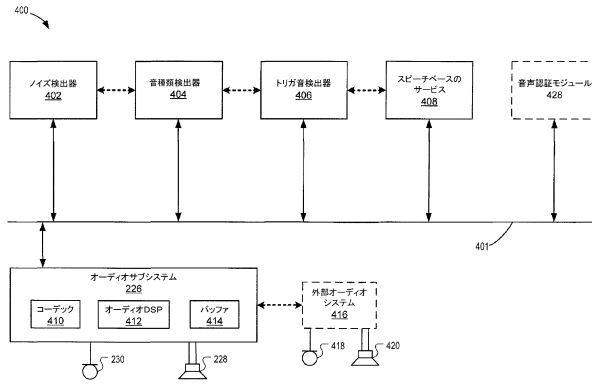
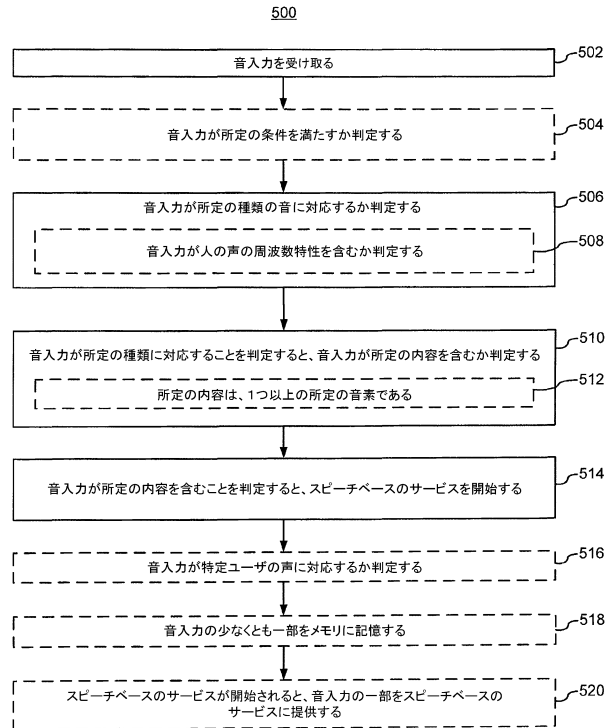


Figure 3C

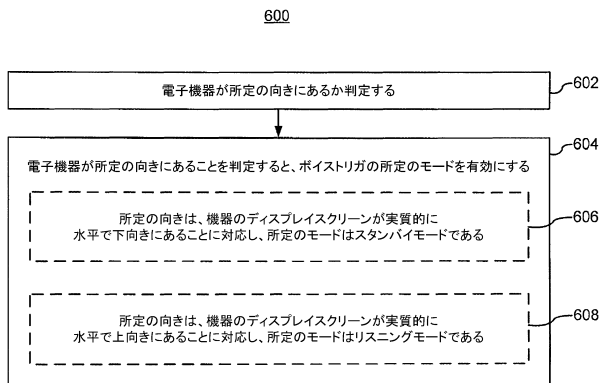
【図4】



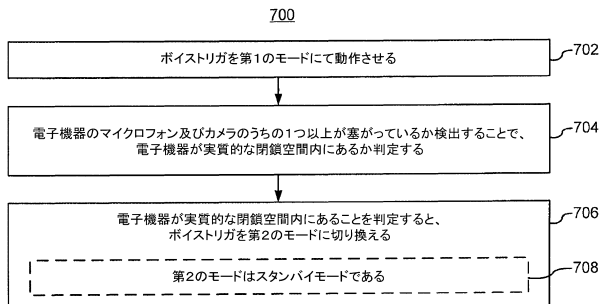
【図5】



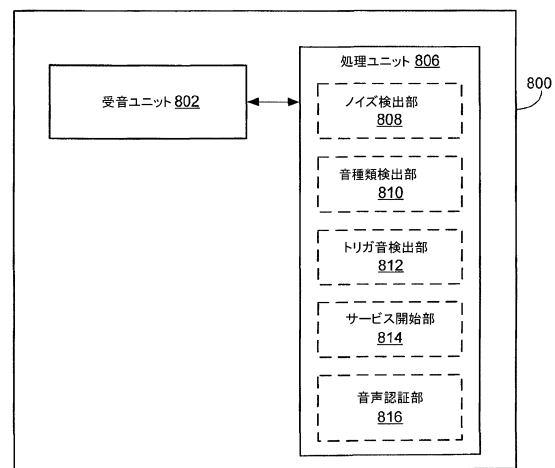
【図6】



【図7】



【図8】



【図 9】

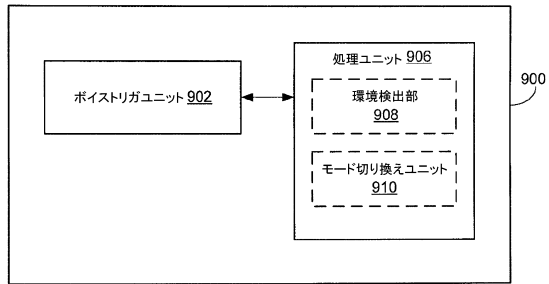


Figure 9

フロントページの続き

(51)Int.Cl.		F I	
H 0 4 M 1/00 (2006.01)		H 0 4 M 1/00	R
H 0 4 M 1/73 (2006.01)		H 0 4 M 1/73	
G 1 0 L 15/18 (2013.01)		G 0 6 F 3/16	6 4 0
		G 1 0 L 15/18	3 0 0

(74)代理人 100116894

弁理士 木村 秀二

(74)代理人 100130409

弁理士 下山 治

(74)代理人 100134175

弁理士 永川 行光

(72)発明者 ビンダー, ジャスティン, ジー.

アメリカ合衆国 カリフォルニア州 9 5 0 1 4, クパチーノ, インフィニット ループ 1
, アップル インコーポレイテッド内

(72)発明者 タッキン, オヌル

アメリカ合衆国 カリフォルニア州 9 5 0 1 4, クパチーノ, インフィニット ループ 1
, アップル インコーポレイテッド内

(72)発明者 ポスト, サミュエル, ディー.

アメリカ合衆国 カリフォルニア州 9 5 0 1 4, クパチーノ, インフィニット ループ 1
, アップル インコーポレイテッド内

(72)発明者 グルーバー, トーマス, アール.

アメリカ合衆国 カリフォルニア州 9 5 0 1 4, クパチーノ, インフィニット ループ 1
, アップル インコーポレイテッド内

審査官 萩島 豪

(56)参考文献 特開平 1 0 - 3 1 2 1 9 4 (J P , A)

特開 2 0 0 0 - 0 2 9 6 6 1 (J P , A)

特開平 0 2 - 1 4 6 0 9 9 (J P , A)

国際公開第 2 0 1 2 / 1 2 9 2 3 1 (W O , A 1)

特開 2 0 0 3 - 2 5 5 9 9 1 (J P , A)

特開 2 0 0 4 - 3 1 0 0 3 4 (J P , A)

国際公開第 2 0 1 0 / 0 5 4 3 7 3 (W O , A 2)

特開 2 0 0 3 - 0 4 4 0 9 0 (J P , A)

特開 2 0 1 2 - 0 3 3 9 9 7 (J P , A)

(58)調査した分野(Int.Cl., D B名)

G 0 6 F 1 / 2 6 - 1 / 3 2

G 0 6 F 3 / 0 1 - 3 / 0 4 8 9

G 0 6 F 3 / 1 6

G 1 0 L 1 5 / 0 0 - 1 7 / 2 6

H 0 4 M 1 / 0 0

H 0 4 M 1 / 7 3

H 0 4 M 1 1 / 0 0