(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7618811号
(P7618811)

(45)発行日　令和7年1月21日(2025.1.21)

(24)登録日　令和7年1月10日(2025.1.10)

(51)国際特許分類

Ｇ１０Ｌ　15/28　(2013.01)	Ｇ１０Ｌ	15/28	２３０Ｋ
Ｇ１０Ｌ　15/10　(2006.01)	Ｇ１０Ｌ	15/10	２００Ｗ
Ｇ１０Ｌ　15/32　(2013.01)	Ｇ１０Ｌ	15/32	２２０Ｚ

請求項の数　20　（全28頁）

(21)出願番号　特願2023-535767(P2023-535767)
(86)(22)出願日　令和3年11月17日(2021.11.17)
(65)公表番号　特表2023-553995(P2023-553995A)
(43)公表日　令和5年12月26日(2023.12.26)
(86)国際出願番号　PCT/US2021/059785
(87)国際公開番号　WO2022/125284
(87)国際公開日　令和4年6月16日(2022.6.16)
審査請求日　令和5年8月10日(2023.8.10)
(31)優先権主張番号　17/118,783
(32)優先日　令和2年12月11日(2020.12.11)
(33)優先権主張国・地域又は機関
米国(US)

(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】　単一の発話におけるデバイスまたはアシスタント固有ホットワードの組合せ

(57)【特許請求の範囲】

【請求項１】

第1のアシスタント対応デバイス(AED)のデータ処理ハードウェアにおいて、ユーザによって話され、前記ユーザに関連付けられた2以上のAEDのうちの第1のAEDと第2のAEDとに向けられた発話に対応するオーディオデータを受け取るステップであって、前記オーディオデータは、行うべき動作を特定するクエリを含む、ステップと、

前記データ処理ハードウェアにより、ホットワード検出モデルを用い、前記オーディオデータの中の第1のホットワードを検出するステップであって、前記第1のホットワードは、前記第1のAEDに割り当てられ、前記第2のAEDに割り当てられた第2のホットワードとは異なる、ステップと、

前記オーディオデータの中の前記第1のAEDに割り当てられた前記第1のホットワードを検出したことに応答して、前記データ処理ハードウェアにより前記オーディオデータに対する処理を開始し、前記第2のAEDに割り当てられた前記第2のホットワードと少なくとも部分的に一致し前記クエリに先行する1または複数の用語を前記オーディオデータが含むことを決定するステップと、

前記第2のホットワードと少なくとも部分的に一致し前記クエリに先行する前記1または複数の用語を前記オーディオデータが含むという決定に基づき、前記データ処理ハードウェアにより、前記第1のAEDと前記第2のAEDとを互いに協働させて前記クエリによって特定された前記動作の遂行を果たす協働ルーチンを実行するステップと、

を含む、方法であって、

前記協働ルーチンの実行の間に、前記第1のAEDと前記第2のAEDとは、前記第1のAEDまたは前記第2のAEDの一方を指定することにより、
前記オーディオデータに関する音声認識結果を生成し、
前記音声認識結果に対してクエリ解釈を行って、行うべき前記動作を特定する前記クエリを前記音声認識結果が識別することを決定し、
前記音声認識結果に対して行われた前記クエリ解釈を、前記第1のAEDまたは前記第2のAEDの他方と共有することによって、
互いに協働する、方法。

【請求項２】

前記オーディオデータが前記第1のホットワードを含むという決定に応答した前記オーディオデータに対する処理の開始は、
前記オーディオデータに対して音声認識を行うように音声認識器に命令して、前記オーディオデータに関する音声認識結果を生成するステップと、
前記オーディオデータに関する前記音声認識結果を用いて、前記第2のホットワードと少なくとも部分的に一致する前記1または複数の用語が前記オーディオデータにおいて認識されると決定するステップと、
を含む、請求項1に記載の方法。

【請求項３】

前記オーディオデータに対して音声認識を行うように前記音声認識器に命令するステップは、
サーバ側の音声認識器に、前記オーディオデータに対して音声認識を行うように命令するステップ、または
前記オーディオデータに対して音声認識を行うために、前記第1のAEDの前記データ処理ハードウェア上で動作するように前記音声認識器に命令するステップ、
の一方を含む、請求項2に記載の方法。

【請求項４】

前記第2のAEDに割り当てられた前記第2のホットワードと少なくとも部分的に一致し前記クエリに先行する前記1または複数の用語を前記オーディオデータが含むことを決定するステップは、
前記ユーザに関連付けられる前記２以上のAEDの各々に割り当てられた1または複数のホットワードのそれぞれのリストを含むホットワードレジストリにアクセスするステップと、
前記第2のAEDに割り当てられた1もしくは複数のホットワードの前記それぞれのリストの中の前記第2のホットワードと一致するもしくは少なくとも部分的に一致する前記1または複数の用語を、前記オーディオデータにおいて認識するステップと、
を含む、請求項1に記載の方法。

【請求項５】

前記ホットワードレジストリにおいて前記２以上のAEDの各々に割り当てられた1または複数のホットワードの前記それぞれのリストは、各ホットワードに関連付けられる1もしくは複数のバリアントをさらに含んでおり、
前記第2のホットワードと少なくとも部分的に一致し前記クエリに先行する前記1または複数の用語を前記オーディオデータが含むことを決定するステップは、前記オーディオデータにおいて認識された前記1または複数の用語が前記第2のホットワードに関連付けられる前記1または複数のバリアントのうちの1つと一致すると決定するステップを含む、請求項4に記載の方法。

【請求項６】

前記ホットワードレジストリは、
前記第1のAED、
前記第2のAED、
前記ユーザに関連付けられる前記２以上のAEDのうちの第3のAED、または

前記ユーザに関連付けられる前記２以上のAEDと通信するサーバ
のうちの少なくとも1つに記憶されている、請求項4に記載の方法。
【請求項７】
前記第2のホットワードと少なくとも部分的に一致し前記クエリに先行する前記1または複数の用語を前記オーディオデータが含むことを決定するステップは、ユーザデバイスに割り当てられた前記第2のホットワードをユーザが話すことを意図したかどうかの蓋然性を決定するように訓練された機械学習モデルへの入力として、前記オーディオデータを提供するステップを含む、請求項1に記載の方法。
【請求項８】
前記オーディオデータにおいて前記クエリに先行する前記1または複数の用語が前記第2のホットワードと単に部分的に一致するときに、前記協働ルーチンを実行することにより、前記クエリによって特定された前記動作の遂行を果たすため、前記第1のAEDに、目を覚まし前記第1のAEDと協働するように前記第2のAEDを呼び出させる、請求項1に記載の方法。
【請求項９】
前記クエリによって特定される前記動作は、前記第1のAEDと前記第2のAEDとの各々に対して行うべきデバイスレベルの動作を含み、
前記協働ルーチンの実行の間に、前記第1のAEDと前記第2のAEDとは、前記デバイスレベルの動作の遂行を独立に果たすことによって互いに協働する、
請求項1に記載の方法。
【請求項１０】
行うべき前記動作を特定する前記クエリは、前記第1のAEDと前記第2のAEDとが時間を要する動作を行うためのクエリを含み、
前記協働ルーチンの実行の間に、前記第1のAEDと前記第2のAEDとは、
前記時間を要する動作の継続時間の間、相互との対を形成し、
前記第1のAEDと前記第2のAEDとの間の行うべき前記時間を要する動作と関係するサブアクションの遂行を調整することによって、
互いに協働する、請求項1に記載の方法。
【請求項１１】
第1のアシスタント対応デバイス(AED)であって、
データ処理ハードウェアと、
前記データ処理ハードウェアと通信するメモリハードウェアと、
を備えており、前記メモリハードウェアは、前記データ処理ハードウェア上で実行されると前記データ処理ハードウェアに動作を行わせる命令を記憶し、前記動作は、
ユーザによって話され、前記ユーザに関連付けられる２以上のAEDのうちの前記第1のAEDと第2のAEDとに向けられた発話に対応するオーディオデータを受け取ることであって、前記オーディオデータは、行うべき動作を特定するクエリを含む、オーディオデータを受け取ることと、
ホットワード検出モデルを用い、前記オーディオデータの中の第1のホットワードを検出することであって、前記第1のホットワードは、前記第1のAEDに割り当てられ、前記第2のAEDに割り当てられた第2のホットワードとは異なる、第1のホットワードを検出することと、
前記第1のAEDに割り当てられた前記第1のホットワードを前記オーディオデータにおいて検出したことに応答して、前記オーディオデータに対する処理を開始し、前記第2のAEDに割り当てられた前記第2のホットワードと少なくとも部分的に一致し前記クエリに先行する1または複数の用語を前記オーディオデータが含むことを決定することと、
前記第2のホットワードと少なくとも部分的に一致し前記クエリに先行する前記1または複数の用語を前記オーディオデータが含むという決定に基づき、前記第1のAEDと前記第2のAEDとを互いに協働させて前記クエリによって特定された前記動作の遂行を果たす協働ルーチンを実行することと、

を含む、第1のアシスタント対応デバイスであって、
前記協働ルーチンの実行の間に、前記第1のAEDと前記第2のAEDとは、前記第1のAEDまたは前記第2のAEDの一方を指定することにより、
前記オーディオデータに関する音声認識結果を生成し、
前記音声認識結果に対してクエリ解釈を行って、行うべき前記動作を特定する前記クエリを前記音声認識結果が識別することを決定し、
前記音声認識結果に対して行われた前記クエリ解釈を、前記第1のAEDまたは前記第2のAEDの他方と共有することによって、
互いに協働する、第1のアシスタント対応デバイス。
【請求項１２】
前記オーディオデータが前記第1のホットワードを含むという決定に応答した前記オーディオデータに対する処理の開始は、
前記オーディオデータに対して音声認識を行うように音声認識器に命令して、前記オーディオデータに関する音声認識結果を生成することと、
前記オーディオデータに関する前記音声認識結果を用いて、前記第2のホットワードと少なくとも部分的に一致する前記1または複数の用語が前記オーディオデータにおいて認識されると決定することと、
を含む、請求項11に記載のデバイス。
【請求項１３】
前記オーディオデータに対して音声認識を行うように前記音声認識器に命令することは、
サーバ側の音声認識器に、前記オーディオデータに対して音声認識を行うように命令すること、または
前記オーディオデータに対して音声認識を行うために、前記第1のAEDの前記データ処理ハードウェア上で実行するように前記音声認識器に命令すること、
の一方を含む、請求項12に記載のデバイス。
【請求項１４】
前記第2のAEDに割り当てられた前記第2のホットワードと少なくとも部分的に一致し前記クエリに先行する前記1または複数の用語を前記オーディオデータが含むことを決定することは、
前記ユーザに関連付けられる２以上のAEDの各々に割り当てられた1または複数のホットワードのそれぞれのリストを含むホットワードレジストリにアクセスすることと、
前記第2のAEDに割り当てられた1もしくは複数のホットワードの前記それぞれのリストの中の前記第2のホットワードと一致するもしくは少なくとも部分的に一致する前記1または複数の用語を、前記オーディオデータにおいて認識することと、
を含む、請求項11に記載のデバイス。
【請求項１５】
前記ホットワードレジストリにおいて前記２以上のAEDの各々に割り当てられた1または複数のホットワードの前記それぞれのリストは、各ホットワードに関連付けられる1もしくは複数のバリアントをさらに含んでおり、
前記第2のホットワードと少なくとも部分的に一致し前記クエリに先行する前記1または複数の用語を前記オーディオデータが含むことを決定することは、前記オーディオデータにおいて認識された前記1または複数の用語が前記第2のホットワードに関連付けられる前記1または複数のバリアントのうちの1つと一致すると決定することを含む、請求項14に記載のデバイス。
【請求項１６】
前記ホットワードレジストリは、
前記第1のAED、
前記第2のAED、
前記ユーザに関連付けられる２以上のAEDのうちの第3のAED、または
前記ユーザに関連付けられる２以上のAEDと通信するサーバ

のうちの少なくとも1つに記憶されている、請求項14に記載のデバイス。
【請求項１７】
前記第2のホットワードと少なくとも部分的に一致し前記クエリに先行する前記1または複数の用語を前記オーディオデータが含むことを決定することは、ユーザデバイスに割り当てられた前記第2のホットワードをユーザが話すことを意図したかどうかの蓋然性を決定するように訓練された機械学習モデルへの入力として、前記オーディオデータを提供することを含む、請求項11に記載のデバイス。
【請求項１８】
前記オーディオデータにおいて前記クエリに先行する前記1または複数の用語が前記第2のホットワードと単に部分的に一致するときに、前記協働ルーチンを実行することにより、前記クエリによって特定された前記動作の遂行を果たすため、前記第1のAEDに、目を覚まし前記第1のAEDと協働するように前記第2のAEDを呼び出させる、請求項11に記載のデバイス。
【請求項１９】
前記クエリによって特定される前記動作は、前記第1のAEDと前記第2のAEDとの各々に対して行うべきデバイスレベルの動作を含み、
前記協働ルーチンの実行の間に、前記第1のAEDと前記第2のAEDとは、前記デバイスレベルの動作の遂行を独立に果たすことによって互いに協働する、
請求項11に記載のデバイス。
【請求項２０】
行うべき前記動作を特定する前記クエリは、前記第1のAEDと前記第2のAEDとが時間を要する動作を行うためのクエリを含み、
前記協働ルーチンの実行の間に、前記第1のAEDと前記第2のAEDとは、
前記時間を要する動作の継続時間の間、相互との対を形成し、
前記第1のAEDと前記第2のAEDとの間の行うべき前記時間を要する動作と関係するサブアクションの遂行を調整することによって、
互いに協働する、請求項11に記載のデバイス。
【発明の詳細な説明】
【技術分野】
【０００１】
本開示は、単一の発話におけるデバイスまたはアシスタント固有ホットワードを組み合わせることに関する。
【背景技術】
【０００２】
音声対応環境(speech-enabled environment)(たとえば、自宅、職場、学校、車両、など)は、ユーザが、声を出してクエリまたはコマンドをコンピュータベースのシステムに向かって話し、そのシステムが、クエリに対処して回答を与えること、および/または、コマンドに基づいて機能を行うことを可能とする。音声対応環境は、その環境の様々な部屋または領域を通じて分散され接続されたマイクロフォンデバイスのネットワークを用いて、実装されることが可能である。これらのデバイスは、その環境に存在する他の個人に向けられた発話とは異なって、与えられた発話がシステムに向けられているときには、見分けることを助けるために、ホットワードを用い得る。したがって、これらのデバイスは、休眠状態または冬眠状態で動作し、検出された発話がホットワードを含むときにのみ、覚醒し得る。いったん覚醒すると、これらのデバイスは、完全なオンデバイスでの自動音声認識(ASR)またはサーバベースのASRなど、より高価な処理を行うように進むことができる。
【発明の概要】
【課題を解決するための手段】
【０００３】
本開示のある態様は、単一の発話におけるホットワードを組み合わせるための方法を提

供する。この方法は、第1のアシスタント対応デバイス(AED)のデータ処理ハードウェアにおいて、ユーザに関連付けられる2以上のAEDのうちの第1のAEDと第2のAEDとに向けられた、ユーザによって話された発話に対応するオーディオデータを受け取るステップを含んでおり、このオーディオデータは、行うべき動作を特定するクエリを含む。この方法は、また、データ処理ハードウェアにより、ホットワード検出モデルを用いて、オーディオデータの中の第1のホットワードを検出するステップを含み、ここで、第1のホットワードは、第1のAEDに割り当てられ、第2のAEDに割り当てられた第2のホットワードとは異なる。オーディオデータ中の第1のAEDに割り当てられた第1のホットワードを検出したことに応答して、この方法は、さらに、データ処理ハードウェアによりオーディオデータに対する処理を開始し、第2のAEDに割り当てられた第2のホットワードと少なくとも部分的に一致しクエリに先行する1または複数の用語をオーディオデータが含むことを決定するステップを含む。第2のホットワードと少なくとも部分的に一致しクエリに先行する1または複数の用語をオーディオデータが含むという決定に基づき、この方法は、追加的に、データ処理ハードウェアにより、第1のAEDと第2のAEDとを互いに協働させてクエリによって特定された動作の遂行を果たす協働ルーチンを実行するステップを含む。

【０００４】

本開示の他の態様は、単一の発話において組み合わされたホットワードを解釈するアシスタント対応デバイスを提供する。このデバイスは、データ処理ハードウェアと、データ処理ハードウェアと通信するメモリハードウェアとを含む。メモリハードウェアは、データ処理ハードウェア上で実行されるとデータ処理ハードウェアに動作を行わせる命令を、記憶する。これらの動作は、ユーザに関連付けられる2以上のAEDのうちの第1のAEDと第2のAEDとに向けられた、ユーザによって話された発話に対応するオーディオデータを受け取ることを含んでおり、このオーディオデータは、行うべき動作を特定するクエリを含む。これらの動作は、また、ホットワード検出モデルを用いて、オーディオデータの中の第1のホットワードを検出することを含み、ここで、第1のホットワードは、第1のAEDに割り当てられ、第2のAEDに割り当てられた第2のホットワードとは異なる。オーディオデータ中の第1のAEDに割り当てられた第1のホットワードを検出したことに応答して、これらの動作は、さらに、オーディオデータに対する処理を開始し、第2のAEDに割り当てられた第2のホットワードと少なくとも部分的に一致しクエリに先行する1または複数の用語をオーディオデータが含むことを決定することを含む。第2のホットワードと少なくとも部分的に一致しクエリに先行する1または複数の用語をオーディオデータが含むという決定に基づき、これらの動作は、追加的に、第1のAEDと第2のAEDとを互いに協働させてクエリによって特定された動作の遂行を果たす協働ルーチンを実行することを含む。

【０００５】

本開示のどの態様の実装形態も、以下のオプショナルな特徴のうちの1または複数を含み得る。いくつかの実装形態では、オーディオデータが第1のホットワードを含むという決定に応答したオーディオデータに対する処理の開始は、オーディオデータに対して音声認識を行うように音声認識器に命令して、オーディオデータに関する音声認識結果を生成することと、オーディオデータに関する音声認識結果を用いて、第2のホットワードと少なくとも部分的に一致する1または複数の用語がオーディオデータにおいて認識されると決定することと、を含む。これらの実装形態では、オーディオデータに対して音声認識を行うように音声認識器に命令することは、サーバ側の音声認識器に、オーディオデータに対して音声認識を行うように命令すること、またはオーディオデータに対して音声認識を行うために、第1のAEDのデータ処理ハードウェア上で動作するように音声認識器に命令することの一方を含む。いくつかの例では、第2のAEDに割り当てられた第2のホットワードと少なくとも部分的に一致しクエリに先行する1または複数の用語をオーディオデータが含むことを決定することは、ユーザに関連付けられる2以上のAEDの各々に割り当てられた1または複数のホットワードのそれぞれのリストを含むホットワードレジストリにアクセスすることと、第2のAEDに割り当てられた1もしくは複数のホットワードのそれぞれのリストの中の第2のホットワードと一致するもしくは少なくとも部分的に一致する1また

は複数の用語を、オーディオデータにおいて認識することと、を含む。これらの例では、ホットワードレジストリにおいて2以上のAEDの各々に割り当てられた1または複数のホットワードのそれぞれのリストは、各ホットワードに関連付けられる1もしくは複数のバリアントをさらに含んでおり、第2のホットワードと少なくとも部分的に一致しクエリに先行する1または複数の用語をオーディオデータが含むことを決定することは、オーディオデータにおいて認識された1または複数の用語が第2のホットワードに関連付けられる1または複数のバリアントのうちの1つと一致すると決定することを含む。また、これらの例では、ホットワードレジストリは、第1のAED、第2のAED、ユーザに関連付けられる2以上のAEDのうちの第3のAED、またはユーザに関連付けられる2以上のAEDと通信するサーバのうちの少なくとも1つに記憶されている。

【０００６】

いくつかの構成では、第2のホットワードと少なくとも部分的に一致しクエリに先行する1または複数の用語をオーディオデータが含むことを決定することは、ユーザデバイスに割り当てられた第2のホットワードをユーザが話すことを意図したかどうかの蓋然性を決定するように訓練された機械学習モデルへの入力として、オーディオデータを提供するステップを含む。いくつかの例では、オーディオデータにおいてクエリに先行する1または複数の用語が第2のホットワードと単に部分的に一致するときに、協働ルーチンを実行することにより、第1のAEDに、目を覚まし第1のAEDと協働するように第2のAEDを呼び出させてクエリによって特定された動作の遂行を果たす。

【０００７】

いくつかの実装形態では、協働ルーチンの実行の間に、第1のAEDと第2のAEDとは、第1のAEDまたは第2のAEDの一方を指定することにより、オーディオデータに関する音声認識結果を生成し、音声認識結果に対してクエリ解釈を行って、行うべき動作を特定するクエリを音声認識結果が識別することを決定し、音声認識結果に対して行われたクエリ解釈を、第1のAEDまたは第2のAEDの他方と共有することによって、互いに協働する。他の実装形態では、協働ルーチンの実行の間に、第1のAEDと第2のAEDとは、各々が独立に、オーディオデータに関する音声認識結果を生成し、音声認識結果に対してクエリ解釈を行って、行うべき動作を特定するクエリを音声認識結果が識別することを決定することによって、互いに協働する。いくつかの例では、クエリによって特定されるアクションは、第1のAEDと第2のAEDとの各々に対して行うべきデバイスレベルのアクションを含み、協働ルーチンの実行の間に、第1のAEDと第2のAEDとは、デバイスレベルのアクションの遂行を独立に果たすことによって互いに協働する。いくつかの構成では、行うべきアクションを特定するクエリは、第1のAEDと第2のAEDとが時間を要する動作を行うためのクエリを含み、協働ルーチンの実行の間に、第1のAEDと第2のAEDとは、時間を要する動作の継続時間の間、相互との対を形成し、第1のAEDと第2のAEDとの間で行うべき時間を要する動作と関係するサブアクションの遂行を調整することによって、互いに協働する。

【０００８】

本開示の追加的な態様は、単一の発話においてホットワードを組み合わせるための他の方法を提供する。この方法は、アシスタント対応デバイス(AED)のデータ処理ハードウェアにおいて、ユーザによって話されAEDによって捕捉された発話に対応するオーディオデータを受け取るステップを含んでおり、この発話は、第1のデジタルアシスタントと第2のデジタルアシスタントとが動作を行うためのクエリを含む。この方法は、また、データ処理ハードウェアにより、第1のホットワード検出モデルを用いて、オーディオデータの中の第1のホットワードを検出するステップを含み、ここで、第1のホットワードは、第1のデジタルアシスタントに割り当てら、第2のデジタルアシスタントに割り当てられた第2のホットワードとは異なる。この方法は、さらに、データ処理ハードウェアにより、第2のデジタルアシスタントに割り当てられた第2のホットワードと少なくとも部分的に一致しクエリに先行する1または複数の用語をオーディオデータが含むことを決定するステップを含む。第2のホットワードと少なくとも部分的に一致しクエリに先行する1または複数の用語をオーディオデータが含むという決定に基づき、この方法は、追加的に、データ処理

ハードウェアにより、第1のデジタルアシスタントと第2のデジタルアシスタントとを互いに協働させて動作の遂行を果たす協働ルーチンを実行するステップを含む。

【０００９】

本開示の実装形態は、以下のオプショナルな特徴のうちの1または複数を含み得る。いくつかの実装形態では、第2のホットワードと少なくとも部分的に一致しクエリに先行する1または複数の用語をオーディオデータが含むことを決定することは、第2のホットワード検出モデルを用いて、オーディオデータ中の第2のホットワードと完全に一致する1または複数の用語を検出することを含む。いくつかの例では、この方法は、さらに、オーディオデータの中の第1のホットワードを検出したことに応答して、データ処理ハードウェアによってオーディオデータに対する処理を開始し、音声認識器に、オーディオデータに対して音声認識を行うように命令して、オーディオデータに関する音声認識結果を生成し、音声認識結果に対してクエリ解釈を行って、音声認識結果がクエリを識別することを決定することにより、第1のデジタルアシスタントと第2のデジタルアシスタントとが動作を行うためのクエリをオーディオデータが含むことを決定することを含む。第2のホットワードと少なくとも部分的に一致しクエリに先行する1または複数の用語をオーディオデータが含むことを決定することは、オーディオデータに関する音声認識結果を用いて、第2のホットワードと少なくとも部分的に一致する1または複数の用語がオーディオデータにおいて認識されることを決定することを含み得る。第1のデジタルアシスタントは、第1のボイスサービスに関連付けられ得るし、第2のデジタルアシスタントは、第2のボイスサービスに関連付けられるが、第1のボイスサービスと第2のボイスサービスとは、異なるエンティティによって提供される。第1のデジタルアシスタントと第2のデジタルアシスタントとは、互いに協働して動作の遂行を果たしながら、ユーザに関連付けられる異なる組のリソースにアクセスし得る。

【００１０】

本開示の1または複数の実装形態の詳細は、添付の図面および以下の説明に記載されている。他の態様、特徴、および利点は、この説明および図面から、ならびに特許請求の範囲から、明らかになるだろう。

【図面の簡単な説明】

【００１１】

【図１Ａ】複数のホットワードが単一の発話において組み合わされる例示的な音声環境の概略図である。

【図１Ｂ】複数のホットワードが単一の発話において組み合わされる例示的な音声環境の概略図である。

【図１Ｃ】図1Aおよび図1Bの音声環境からの例示的なアシスタント対応デバイスの概略図である。

【図１Ｄ】図1Aおよび図1Bの音声環境からの例示的なアシスタント対応デバイスの概略図である。

【図２】音声環境において動作している例示的な協働ルーチンである。

【図３】複数のデジタルアシスタントを含む例示的なアシスタント対応デバイスの概略図である。

【図４】デバイス固有の複数のホットワードを単一の発話において組み合わせる方法のための例示的な動作構成のフローチャートである。

【図５】アシスタント固有の複数のホットワードを単一の発話において組み合わせる方法のための例示的な動作構成のフローチャートである。

【図６】本明細書に記載されているシステムおよび方法を実装するために用いられ得る例示的なコンピューティングデバイスの概略図である。

【発明を実施するための形態】

【００１２】

様々な図面における同様の参照符号は、同様の要素を示す。

【００１３】

理想的には、デジタルアシスタントインターフェースと会話するときには、ユーザは、そのデジタルアシスタントインターフェースを動作させているアシスタント対応デバイスの方向に向けられて話されたリクエストを経由して、自分自身があたかも他の人に話しかけているように伝えることができるべきである。デジタルアシスタントインターフェースは、話されたリクエストを処理し認識する結果としてアクションが行われることを可能とするように、これらの話されたリクエストを自動音声認識器に提供する。しかし、実際には、スマートフォンやスマートウォッチなど、リソースに制限があるボイス対応デバイス(voice-enabled device)において音声認識を連続的に動作させることは法外に高価であるから、これらの話されたリクエストにデバイスが常に応答するのは、困難である。

【００１４】

常にオンである音声をサポートするユーザ体験を生じさせるために、アシスタント対応デバイス(assistant-enabled devices)は、典型的には、狭い組のフレーズを特徴付けるオーディオ特徴を認識するように構成されたコンパクトなホットワード検出モデルを動作させ、このオーディオ特徴は、ユーザによって話されると、ユーザによって話されたどの後続の音声に対しても完全自動音声認識(ASR)を始動させる。有利には、ホットワード検出モデルは、デジタル信号プロセッサ(DSP)チップなどの低電力ハードウェア上で動作が可能であり、「やあグーグル」または「やあリビングルームのスピーカ」など、様々な固定されたフレーズコマンドに応答し得る。

【００１５】

ユーザの環境(たとえば、自宅や職場)におけるアシスタント対応デバイスの個数が増えるにつれて、ユーザは、たとえば、一群のアシスタント対応スマートスピーカの全体で音量レベルを調整するために、または一群のアシスタント対応スマート照明の全体で照明レベルを調整するために、複数のアシスタント対応デバイスを同時にトリガすることを望むことがあり得る。同様に、複数の異なるボイスアシスタントサービスを提供する単独のアシスタント対応デバイスの場合には、ユーザは、ユーザのクエリを果たすために、これらのボイスサービスのうちの2以上を同時にトリガすることを望む場合があり得る。ユーザが複数の異なるアシスタント対応デバイスをトリガすることを望む場合でも、または複数の異なるボイスアシスタントサービスをトリガすることを望む場合でも、ユーザは、それぞれのデバイスまたはデジタルアシスタントサービスに対し、別個のクエリを独立に生じさせることが、現に要求される。たとえば、ユーザの自宅において、台所の照明と食堂の照明とを消灯するためには、ユーザは、「やあ、台所の電球、消えて下さい」および「やあ、食堂の電球、消えて下さい」というように、別個のクエリを話さなければならないことになり得る。

【００１６】

本明細書における実装形態は、すべてのデバイスまたはデジタルアシスタントサービスをトリガすることによりユーザによって話される発話の中の後続のクエリを処理するために、デバイス固有の複数のホットワードをユーザによって話される単一の発話の中でユーザが組み合わせることを可能にすることに向けられている。以下で詳細に説明されるように、ユーザ環境において同じ場所にある複数のアシスタント対応デバイス(AED)は、それぞれのAEDが、それぞれのデバイス固有ホットワードに応答し、ユーザ環境において同じ場所にある他のAEDのうちの1または複数に代わって部分的なデバイス固有ホットワードを検出/認識するようにも構成され得るように、互いに協働し得る。たとえば、あるユーザが、それぞれがそれ自体のデバイス固有ホットワード(たとえば、やあデバイス1、および、やあデバイス2)に応答する2つのスマートスピーカを有しており、このユーザが、両方のスピーカで、彼または彼女のジャズプレイリストを再生させることを希望するというシナリオでは、ユーザは、両方のスマートスピーカでリクエストされたプレイリストの再生を開始させるために、「やあ、デバイス1およびデバイス2、私のジャズプレイリストを再生して下さい」という単一のクエリを話すことができる。このシナリオでは、ユーザは、完全なデバイス固有ホットワードである「やあデバイス1」と話しているが、第2のスマートスピーカに対しては、デバイス固有ホットワードを単に部分的にしか話していない(たと

えば、「やあ」という用語が、話されたフレーズである「デバイス2」の直前に来ていなかった)。しかし、第1のスマートスピーカが「やあデバイス1」というフレーズを検出することが、目を覚まし、ユーザによって話された発話を認識するようにASRを始動させるように、デバイスをトリガしている。これら2つのスマートスピーカは対を成し互いに協働するように構成されているから、「やあデバイス1」というフレーズを検出して現にASRを動作させている第1のスマートスピーカは、第2のスマートスピーカのための部分的ホットワードの一致として「デバイス2」というフレーズを認識し、ユーザが第2のスマートスピーカを呼び出すことも意図していると決定することができる。このシナリオでは、第1のスマートスピーカは、ジャズプレイリストからの曲が両方のスピーカから同時に演奏されるようにするために、クエリを処理するためにも第2のスマートスピーカに目を覚ますように命令すること、および/または、第2のスマートスピーカに代わってクエリを果たすことがあり得る。ユーザは、複数のAEDに向けられた単一のクエリを同時に話すだけでよく、それにより、それぞれがAEDのうちの異なるものに向けられた複数のクエリを提供する必要がなかったため、ユーザの時間を節約することになったのが、有利であった。

【００１７】

図1A～図1Dを参照すると、いくつかの実装形態において、音声環境100は、複数のアシスタント対応デバイス110(デバイス110、ユーザデバイス110、またはAED110とも称される)の方向に向けられた発話20を話すユーザ10を含む。ここで、ユーザ10によって話された発話20は、ストリーミングオーディオ12として1または複数のデバイス110によって捕捉され、クエリ22に対応し得る。たとえば、クエリ22は、アクション、動作、またはタスクを行うようにとのリクエストを指し、より特定すると、デバイス110のうちの1または複数において動作しているデジタルアシスタントインターフェース120に対してアクション、動作、またはタスクを行うようにとのリクエストを指す。ユーザ10は、それぞれのデバイス110上で動作しているホットワード検出器130(図1Cおよび図1D)によってストリーミングオーディオ12において1または複数のホットワード24が検出されるときに、休眠または冬眠状態(すなわち、低電力状態)から目覚めるように、1または複数のデバイス110、110a～nをトリガする呼び出しフレーズとして、1または複数のホットワード24および/または部分ホットワード24、24pをクエリ22の前に配置し得る。この意味で、ユーザ10は、コンピューティング活動を行うまたは質問への回答を見つけるために、AEDデバイス110上で動作しているデジタルアシスタントインターフェース120との間で会話的な対話を有し得る。

【００１８】

デバイス110は、ユーザ10に関連付けらており環境100からのオーディオを捕捉することができる任意のコンピューティングデバイスと対応し得る。ユーザデバイス110のいくつかの例は、これらに限定されることはないが、モバイルデバイス(たとえば、携帯電話、タブレット、ラップトップ、電子書籍リーダなど)、コンピュータ、ウェアラブルデバイス(たとえば、スマートウォッチ)、音楽プレーヤ、キャストデバイス(casting device)、スマート機器(たとえば、スマートテレビ)およびモノのインターネット(IoT)デバイス、リモコン、スマートスピーカなどを含む。デバイス110は、データ処理ハードウェア112dと、データ処理ハードウェア112dと通信するメモリハードウェア112mとを含んでおり、メモリハードウェア112mは、データ処理ハードウェア112dによって実行されると、音声処理に関係する1または複数の動作をデータ処理ハードウェア112dに行わせる命令を記憶する。

【００１９】

デバイス110は、さらに、音声環境100におけるオーディオを捕捉してオーディオデータ14と称される電気信号(図1Cおよび図1Dのオーディオデータ14)に変換するためのオーディオ捕捉デバイス(たとえば、1または複数のマイクロフォンのアレイ)114を備えたオーディオサブシステムを含む。示されている例では、デバイス110はオーディオ捕捉デバイス114(一般的に、マイクロフォン114とも称される)を実装しているが、オーディオ捕捉デバイス114は、物理的にデバイス110上に存在せずに、オーディオサブシステム(たとえ

ば、デバイス110の周辺機器)と通信する場合もあり得る。たとえば、デバイス110は、車両の全体に配置されたマイクロフォンのアレイを活用する車両インフォテイメントシステムに対応し得る。マイクロフォンなどのオーディオ捕捉デバイス114に加えて、デバイス110のオーディオサブシステムは、スピーカなど、オーディオ再生デバイス116も含み得る。スピーカ116を用いると、デバイス110は、ユーザ10および/またはデバイス110が位置する環境100のために、オーディオを再生し得る。これにより、デバイス110(たとえば、アシスタントインターフェース120)が、デバイス110に関連付けられる1または複数のスピーカにおける合成された再生オーディオ出力を用いて、クエリ22に応答することが可能になり得る。たとえば、ユーザ10がアシスタントインターフェース120に「きょうの天候はどんな様子ですか?」と質問すると、スピーカ116は、「きょうは晴天で、摂氏約21度(華氏70度)です」と述べる合成された音声を出力し得る。

【００２０】

デバイス110は、また、グラフィカルユーザインターフェース(GUI)要素(たとえば、ウィンドウ、スクリーン、アイコン、メニュなど)および/またはグラフィカルコンテンツを表示するディスプレイ118を含み得る。たとえば、デバイス110は、GUI要素または他のグラフィカルコンテンツをディスプレイ118のために生成するアプリケーションを、ロードまたは起動し得る。ディスプレイ118において生成されたこれらの要素は、ユーザ10によって選択可能であり得るし、また、デバイス110上で生じるアクティビティ/動作を処理するための何らかの形式の視覚的フィードバック、またはクエリ22への視覚的な応答を提供するようにも機能し得る。さらにまた、デバイス110はボイス対応デバイス110であるから、ユーザ10は、様々なボイスコマンドを用いて、ディスプレイ118上に生成された要素と対話し得る。たとえば、ディスプレイ118は、特定のアプリケーションのためのオプションのメニュを示す場合があり、ユーザ10は、音声を介してオプションを選択するために、インターフェース120を用い得る。

【００２１】

例示を与えると、ユーザ10は、ユーザの自宅の居間に位置する2つのスマート電球に対応する2つのAED110、110d～eに、発話20を向け得る。ここで、ユーザ10は、居間で映画を見ていて、居間の照明を暗くすることを望むことがあり得る。このシナリオでは、ユーザ10は、「デバイス1とデバイス2、照明を暗くして下さい」というクエリを話し得る。ここで、クエリ22は、第1のスマート電球110dに関連付けられる完全なデバイス固有ホットワード(「デバイス1」)と第2のスマート電球110eに関連付けられる完全なデバイス固有ホットワード(「デバイス2」)との前に配置されるのであるが、これが、デバイス110d、110eの両方を、目を覚まし、クエリ22によって独立に特定される動作を果たす、すなわち、それぞれのスマート電球がその照度を暗い照明の特徴であるレベルまで下げることを果たすことによって互いに協働するように、トリガする。追加的にまたはその代わりに、このクエリ22に応答して、デバイス110d、110eの一方または両方が、他のデバイス110cに、電球110d、110eのそれぞれの暗さレベルを制御/調整するためのスライダグラフィカルユーザインターフェース(GUI)をユーザ10に提供するGUIをディスプレイ118上に表示するように、命じる。この例をさらに拡張するため、2つのデバイス110がこのクエリ22を受け取ると、これらのデバイスは、クエリ22を実行し、第3のデバイス110、110cと協働することがあり得るが、この第3のデバイスとは、ユーザ10の近くに位置しており、第1および/または第2のデバイス110d～eと通信するモバイルデバイス110cである。

【００２２】

音声対応インターフェース(たとえば、デジタルアシスタントインターフェース)120は、デバイス110によって捕捉された発話20で運ばれたクエリ22またはコマンドを処理し得る。音声対応インターフェース120(インターフェース120またはアシスタントインターフェース120とも称される)は、一般に、発話20に対応するオーディオデータ14を受け取ること、およびオーディオデータ14に対する音声処理または発話20から生じる他のアクティビティを調整することを容易にする。インターフェース120は、デバイス110のデー

タ処理ハードウェア112d上で動作する。インターフェース120は、発話20を含むオーディオデータ14を、音声処理またはクエリ遂行に関係する様々なシステムに伝え得る。

【００２３】

図1Cおよび図1Dなど、いくつかの例では、インターフェース120は、ホットワード検出器130、音声認識器140、および/またはインタプリタ170と通信する。音声認識器140は、インタプリタ170を実装ことがあり得るし、または、インタプリタ170が別個のコンポーネントであり得る。ここでは、インターフェース120が、発話20に対応するオーディオデータ14を受け取り、オーディオデータ14をホットワード検出器130に提供する。ホットワード検出器130は、1または複数のホットワード検出段を含み得る。たとえば、ホットワード検出器130は、「常にオン」であってホットワードの存在を最初に検出するように構成されている第1段のホットワード検出器を含み得るが、いったんホットワード候補が検出されると、第1段のホットワード検出器は、そのホットワード候補24を特徴付けるオーディオデータ14を、オーディオデータ14がホットワード候補24を含むかどうかを確認する第2段のホットワード検出器に送り得る。第2段のホットワード検出器が、第1段のホットワード検出器によって検出されたホットワード候補を拒絶することにより、デバイス110が休眠または冬眠状態から目覚めることを回避することがあり得る。第1段のホットワード検出器は、ホットワードの存在を粗く待機するためにデジタル信号プロセッサ(DSP)を実行させるホットワード検出モデルを含み得るし、第2段のホットワード検出器は、第1段のホットワード検出器によって検出されたホットワード候補を受諾するまたは拒絶するために、第1段よりも計算量の多い(computationally-intensive)ホットワード検出モデルを含み得る。第2段のホットワード検出器は、第1段のホットワード検出器がストリーミングオーディオにおいてホットワード候補24を検出するとトリガするアプリケーション・プロセッサ(CPU)上で動作し得る。いくつかの例では、第2段のホットワード検出器は、オーディオデータ14に対して音声認識を行って、ホットワード24がオーディオデータ14において認識されるかどうかを決定する音声認識器140を含む。

【００２４】

ホットワード検出器130に関連付けられるホットワード検出モデルが、発話20に対応するオーディオデータ14がデバイス110に割り当てられたホットワード24を含むことを検出すると、インターフェース120(またはホットワード検出器130それ自体)は、オーディオデータ14に対する音声処理を開始するために、オーディオデータ14を音声認識器140に送り得る。たとえば、インターフェース120は、オーディオデータ14に対する処理を開始するために、オーディオデータ14を音声認識器140に中継するのであるが、これは、オーディオデータ14が、他のデバイス110に割り当てられたホットワード24と少なくとも部分的に一致し発話20のクエリ22に先行する1または複数の用語を含むかどうかを、決定するためである。オーディオデータ14が、他のデバイス110に割り当てられた他のホットワード24と少なくとも部分的に一致しクエリ22に先行する1または複数の用語を含む、という決定に基づき、インターフェース120は、2つのデバイス110を互いに協働させてクエリ22によって特定された動作の遂行を果たす協働ルーチン200を、実行し得る。

【００２５】

図1Cおよび図1Dを参照しながら図1Aの例を参照すると、ユーザ10によって話された発話20は、「やあデバイス1そしてデバイス2、私のジャズプレイリストを再生して下さい」を含む。ここで、発話20は、第1のデバイス110、110aに割り当てられた第1のホットワード24、24a「やあ、デバイス1」を含んでおり、オーディオデータ14中のこの第1のホットワードは、検出されると、「そしてデバイス2、私のジャズプレイリストを再生して下さい」という用語に対応する以後に捕捉されたオーディオデータ14を処理のために音声認識システム140に中継するように、第1のデバイス110a上で動作しているインターフェース120をトリガする。すなわち、第1のデバイス110aは、休眠または冬眠状態にあり、オーディオストリーム12におけるホットワード24または部分ホットワード24、24pの存在を検出するために、ホットワード検出器130を動作させ得る。たとえば、「デバイス2」は、第2のデバイス110bに対する部分ホットワード24pであると考えられ得るが、そ

の理由は、第2のデバイス110bに割り当てられた全体/完全なホットワード24はフレーズ「やあデバイス2」を含むからである。よって、発話20は、全体のホットワードフレーズである「やあデバイス2」の用語「やあ」を欠いており、用語「デバイス2」は全体的な/完全なホットワード24の部分ホットワード24pに関連付けられている。本明細書で用いられているホットワード24は、一般に、全体ホットワード24または部分ホットワード24pのいずれかを指し得る。呼び出しフレーズとして機能するホットワード24は、ホットワード検出器130によって検出されると、目を覚ましホットワード24および/またはホットワード23に続く1もしくは複数の用語(たとえば、「そしてデバイス2、私のジャズプレイリストを再生して下さい」という用語)に対する音声認識を開始するように、デバイス110をトリガする。たとえば、発話20は、第1のデバイス110aに割り当てられた第1のホットワード24aと第2のデバイス110bに割り当てられた第2のホットワード24b(たとえば、部分ホットワード24p)を含むから、図1Aは、ユーザ10のジャズプレイリストからの音楽を演奏するために、目を覚まし互いに協働している第1および第2のデバイス110a～bを示しているが、他方で、第3のデバイス110cは、発話20を捕捉するためにユーザ10の十分に近い範囲内に存在するが、発話20が第3のデバイス110cに割り当てられたいかなるホットワード24も含まないため、目を覚まさない。この例では、発話20が第2のホットワード24b$_p$と部分一致するだけの1または複数の用語を含むため、第2のデバイス110b上で動作しているホットワード検出器130は、ホットワードの存在を検出して、目を覚ますように第2のデバイス110bをトリガすることはない。代わりに、第1のデバイス110aは、第2のホットワード「やあデバイス2」と部分的に一致する1または複数の用語「デバイス2」を識別するため、音声認識を開始し、オーディオデータに対してASR結果142に対して意味論的な解釈を行い、次に、クエリ22によって特定されるジャズプレイリストの再生動作を果たすために、目を覚まし、第1のデバイス110aと協働するように、第2のデバイス110bを呼び出す。

【００２６】

図1Cおよび図1Dを参照すると、ホットワード検出を行うために、ホットワード検出器130は、音声認識または意味論的解析を行うことなくホットワード24を示す音響的特徴を検出するように構成された、ニューラルネットワークベースのモデルなどの、ホットワード検出モデルを含む。ホットワード検出器130を用いることにより、ホットワード24の検出は、(たとえば、データ処理ハードウェア112dに関連付けられる)デバイスの計算処理装置(CPU)の消費を回避するDSPチップなどの低電力ハードウェアにおいて、生じ得る。上述されたように、第1段のホットワード検出器は、ホットワード候補の存在を最初に検出するために、DSPチップ上で動作し、次に、ホットワードの存在を確認するために、目を覚まして第2段のホットワード検出器(ホットワード検出モデルまたは音声認識器)を実行させるように、CPUを呼び出し得る。この検出器がホットワード24を検出すると、ホットワード24は、目を覚ましてより高価な処理(たとえば、ASRおよび自然言語理解(NLU))を要求する音声認識を開始させるように、デバイスをトリガし得る。ここで、デバイスは、データ処理ハードウェア112d(たとえば、CPU)上で音声認識器140を動作させることによって、オンデバイスASRを行い得る。オプションであるが、デバイス110は、サーバ(たとえば、図1Bのリモートシステム160)とのネットワーク接続を確立して、オーディオデータ14に対してサーバ側のASRおよび/またはNLUを行うために、オーディオデータ14をサーバに提供することがあり得る。いくつかの実装形態では、環境100における各デバイス110は、それ自体のホットワード検出器130を動作させる。

【００２７】

ホットワード検出器130がオーディオデータ14においてホットワード24を検出したことに応答して、インターフェース120は、この発話20に対応するオーディオデータ14を音声認識器140に中継し、音声認識器140は、オーディオデータ14に対する音声認識を行って、発話20に対する自動音声認識(ASR)結果(たとえば、文字起こし)142を生成する。音声認識器140および/またはインターフェース120は、ASR結果142を、インタプリタ170(たとえば、NLUモジュール)に提供して、ASR結果142に対して意味論的解釈を行い、

第2のデバイス110bに割り当てられた「やあデバイス2」という第2のホットワードと部分的に一致する1または複数の用語「デバイス2」をオーディオデータ14が含むことを決定することができる。したがって、第2のホットワード24、$24b_p$と部分的に一致する1または複数の用語をオーディオデータ14が含むという決定に基づき、インタプリタ170は、発話20もまた第2のデバイス110bの方向に向けられていたと決定することにより、目を覚まし第1のデバイス110aと協働するように第2のデバイス110bを呼び出すために、協働ルーチン200の実行を開始する命令172を提供する。特に、もし発話20が代わりに第2のホットワード24に対するすべての用語「やあデバイス2」を含んでいた場合には、第2のデバイス110b上で動作しているホットワード検出器130は、第2のホットワードを検出し、やはり目を覚まして音声認識と意味論的解釈とを独立に行うように第2のデバイス110bをトリガすることにより、後続のクエリ22によって特定される動作を果たすために第2のデバイス110aとの協働するために協働ルーチン200を実行していた場合があり得る。

【００２８】

この例では、クエリ22は、可聴の再生のためにジャズ音楽のプレイリストをストリーミングするという長時間にわたる動作を行うための、第1および第2のデバイス110a、110bに対するクエリを含む。したがって、協働ルーチン200の実行の間、第1および第2のデバイス110a、110bは、長時間にわたる動作の継続時間の間には相互との対になり、第1のデバイス110aと第2のデバイス110bとの間での長時間にわたる動作に関係するサブアクションの遂行を調整することによって、互いに協働し得る。言い換えると、クエリ22は音楽演奏コマンドに対応するため、協働ルーチン200は、第1のデバイス110aと第2のデバイス110bとを相互との対にさせ、ユーザのジャズプレイリストからの曲をステレオ構成で可聴的に再生させ得ることにより、第1のデバイス110aは、左側のオーディオチャネルの役割をサブアクションとして想定し、第2のデバイス110bは、右側のオーディオチャネルの役割をサブアクションとして想定することになる。

【００２９】

いくつかの実装形態では、図1Bに示されているように、デバイス110は、ネットワーク150を経由して、リモートシステム160と通信する。リモートシステム160は、リモートデータ処理ハードウェア164(たとえば、リモートサーバもしくはCPU)および/またはリモートメモリハードウェア166(たとえば、リモートデータベースもしくは他のストレージハードウェア)など、リモートリソース162を含み得る。デバイス110は、音声処理および/またはクエリ遂行と関係する様々な機能を行うために、リモートリソース162を利用し得る。たとえば、音声認識器140の機能の一部または全部は、リモートシステム160上(すなわち、サーバ側)に存在し得る。ある例では、音声認識器140は、オンデバイス自動音声認識(ASR)を行うためのデバイス110上に存在する。他の例では、音声認識器140は、サーバ側のASRを提供するために、リモートシステム160上に存在する。さらに他の例では、音声認識器140の機能が、デバイス110とサーバ160とにわたり分離されている。

【００３０】

いくつかの構成では、音声認識器140は、音声認識の間に用いられる音声認識モデルのタイプに応じて、異なる位置(たとえば、オンデバイスまたはリモート)に存在し得る。エンドツーエンドまたはストリーミングベースの音声認識モデルは、その空間効率的なサイズのためにデバイス110上に存在し得るが、他方、複数のモデル(たとえば、音響モデル(AM)、発音モデル(PM)、および言語モデル(LM))から構築される、より大型で、より従来型の音声認識モデルは、オンデバイスであるよりもむしろ、リモートシステム140に存在するサーバベースのモデルであり得る。言い換えると、音声認識の所望レベルおよび/または音声認識を行う所望速度に応じて、インターフェース120は、音声認識器140による音声認識がオンデバイスで(すなわち、ユーザ側で)で生じるか、またはリモートで(すなわち、サーバ側で)生じるのかを命令し得る。

【００３１】

図1Bなどのいくつかの例では、環境100は、第1のネットワーク150、150aと、第2のネットワーク150、150bとを含む。ここでは、第1のネットワーク150aは、ユーザの自

宅に関連付けられる個人ネットワークなど、ローカルエリアネットワーク(LAN)と対応し得る。LANとして、第1のネットワーク150aは、ユーザ10に関連付けられる複数のデバイス110、110a～nが互いに接続可能である、および/または互いに通信可能であるように構成されているローカルネットワークレイヤを参照し得る。たとえば、デバイス110は、WiFi、Bluetooth、Zigbee、Ethernet、または他の無線ベースのプロトコルなど、有線および/または無線通信プロトコルを用いて、互いに接続する。第1のネットワーク150aでは、あるデバイス110が、クエリ22を果たすために、情報(たとえば、協働ルーチン200に関連付けられる命令)を、1または複数の他のデバイス110にブロードキャストし得る。デバイス110は、ネットワーク150aの中への加入の際に通信手段を確立するためのデイスカバリプロセスにおいて、互いに通信するように設定され得るか、または特定の組のデバイス110を呼び出すクエリ22に応答して、相互とのペアリングプロセスを経験し得る。第1のネットワーク150aまたはローカルネットワークも、第2のネットワーク150bまたはリモートネットワークと通信するように構成され得る。ここで、リモートネットワークとは、広い地理的領域にわたって拡大するワイドエリアネットワーク(WAN)を指し得る。第2のネットワーク150bと通信可能であることにより、第1のネットワーク150aは、リモートシステム160と通信し得る、またはリモートシステム160へのアクセスを有し得ることになり、それによって、1または複数のデバイス110が、サーバ側の音声認識、サーバ側のホットワード検出、もしくはいくつかの他のタイプのサーバ側の音声処理もしくはクエリ遂行などのサービスを、行うことが可能になる。いくつかの構成では、スーパーバイザ(たとえば、コンピュータベースのソフトウェア)は、第1のネットワーク150a上でまたはユーザ10に関連付けられるローカルネットワーク上で動作するデバイス110と協働するように構成され得ることにより、スーパーバイザは、ユーザ10からの発話20が複数のデバイス110を目覚めさせ、目覚めさせられたデバイス110の間での協働ルーチン200をスーパーバイザが容易にするまたは開始することを認識し得る。

【００３２】

図1Dを参照すると、いくつかの実装形態では、インタプリタ170が、第2のAEDに割り当てられた第2のホットワード24「やあデバイス2」と少なくとも部分的に一致しクエリ22に先行する1または複数の用語をオーディオデータ14が含むことを、ユーザ10に関連付けられる各デバイス110、110a～nに割り当てられた1または複数のホットワードのそれぞれのリスト182、182a～nを含むホットワードレジストリ180にアクセスし、オーディオデータ14に対するASR結果142が、第2のデバイス110bに割り当てられた1もしくは複数のホットワードのそれぞれのリスト182aにおける第2のホットワード「やあデバイス2」と一致もしくは部分一致する1または複数の用語を含むことを識別することによって、決定する。インタプリタ170によってアクセスされたホットワードレジストリ180は、デバイス110のうちの1つ、デバイスのうちの複数、および/またはネットワーク150(図1B)を経由してすべてのデバイス110と通信する中央サーバ(たとえば、リモートシステム160(図1B))上に、記憶され得る。よって、各デバイス110は、(1)その特定のデバイス110に対するホットワード24だけを含むデバイス固有のホットワードレジストリ180を記憶する、(2)ユーザ10に関連付けられるすべてのデバイス110に対するホットワード24を有するグローバルホットワードレジストリ180を記憶する、または(3)どのホットワードレジストリ180も記憶しない、であり得る。特定のデバイス110がグローバルホットワードレジストリ180を含むときには、そのデバイス110は、ユーザ10のデバイス110に割り当てられたホットワード24に対するローカルな中央ストレージノードとして機能し得る。グローバルホットワードレジストリ180を記憶しないデバイスは、ローカルネットワーク150aを経由して1または複数の他のデバイス110上に記憶されているグローバルホットワードレジストリ180にアクセスするか、またはリモートシステム160上に存在するグローバルホットワードレジストリにアクセスし得る。デバイス110は、新たなホットワード24がアクティブ/利用可能であるとき、および/またはホットワード24が非アクティブ/利用不可能になるときに、ホットワードレジストリ180をアクティブに更新し得る。

【００３３】

いくつかの例では、ホットワードレジストリ180において各デバイス110に割り当てられたホットワード24のそれぞれのリスト182は、各ホットワード24に関連付けられる1または複数のバリアントを含む。ここで、ある特定のデバイス10に割り当てられたホットワード24のそれぞれのバリアントは、そのデバイスに対する部分ホットワード24pに対応し得る。引き続きこの例を考察すると、図1Dは、ホットワード「やあデバイス2」と、バリアント「デバイス2」と、部分ホットワード24pに対応するバリアント「やあデバイス〈...〉2」と、を含む、第2のデバイス110bに割り当てられたホットワード24のそれぞれのリストを示している。よって、第1のデバイス110a上で動作しているインタプリタ170は、ホットワードレジストリ180にアクセスし、ASR結果142における1または複数の用語と部分ホットワード24pが一致するため、第2のデバイス110bに関連付けられるそれぞれのリスト182aがバリアント「デバイス2」をリストに含めていることを識別する。特に、リスト182aは、バリアント「やあデバイス〈...〉2」を、ユーザがクエリを「やあデバイス1と2」または「やあデバイス1とデバイス2」より先行させるときに、第2のホットワード「やあデバイス2」との部分一致を可能にする複合表現(complex expression)としてリストに含めている。

【００３４】

上の注記で述べたように、ユーザ10があるホットワードを部分的に話すだけのときには、特定のデバイス110上で動作しているホットワード検出器130は、そのホットワード24の存在を検出することはなく、つまり、ユーザ10によって部分ホットワード24pだけが話されるときには、目覚めさせるようにデバイス110をトリガすることはない。さらに例示すると、ホットワード検出器130がホットワード検出を行っているときには、ホットワード検出器130は、特定のホットワード24がストリーミングオーディオの中に存在する信頼レベルを示すホットワードスコアを生成する。ホットワードスコアが閾値を満足させる(たとえば、特定の閾値を超える)ときには、ホットワード検出器130は、ストリーミングオーディオの中に完全なホットワード24が存在することを識別する。しかし、ストリーミングオーディオの中に部分ホットワード24pだけが存在するときには、ホットワード検出器130は、閾値を満足させない対応するホットワードスコアを生成し得る。結果的に、ホットワード検出器130は、ホットワード24を検出せず、デバイス110は引き続き休眠または冬眠状態に留まることになる。この結果を回避するためには、インタプリタ170は、ホットワードレジストリ180にアクセスして、オーディオデータ14において認識された1または複数の用語(たとえば、ASR結果142における1または複数の用語)があるホットワード24に関連付けられるバリアントと一致することを決定することができる。この一致は、目を覚まさせ、クエリ22によって特定された動作を果たすためにクエリが向けられた1または複数の他のデバイス110と協働させるように、デバイス110をトリガするために、信頼スコアを効果的に上昇させることが可能である。

【００３５】

いくつかの例では、オーディオデータにおいてそのホットワードを検出するAED110は、機械学習モデル175を実行させて、発話20に対応するオーディオデータ14が他のAEDに割り当てられたホットワードを指すかどうかを決定することができる。したがって、機械学習モデル175は、オーディオデータにおいて部分ホットワードを検出するように訓練される。機械学習モデル175は、オーディオデータ14を入力として受け取り、ユーザ10が他のAEDに割り当てられたホットワードを話すことを意図していたかどうかの蓋然性を決定し得る。機械学習モデルは、1または複数のホットワードとそのバリアントとに対して、予測されるホットワード発話に関して訓練され得る。機械学習モデルは、ニューラルネットワークまたは埋め込みベースの比較モデル(embedding-based comparison model)を含み得るが、後者では、オーディオデータ14の埋め込みが、予測されるホットワード発話の埋め込みと比較される。

【００３６】

図2を参照すると、協働ルーチン200は、他のデバイス110に割り当てられたホットワードと少なくとも部分的に一致する1または複数の用語を音声認識結果142が含むことを

示す命令172をアシスタントインターフェース120が提供したことに応答して、動作する。命令172は、発話20が向けられた2以上のデバイス110、110a～nの各々に関連付けられる識別子を含み得る。命令172は、さらに、音声認識結果142を含み得る。オーディオデータの中の1または複数の用語が、デバイス110、110bに割り当てられたホットワードと部分的に一致するだけのときには、協働ルーチン200を実行させることで、トリガされたデバイス110、110aに、目を覚まし第1のデバイス110aと協働してクエリ22によって特定された動作の遂行を果たすように他方のデバイス110bを呼び出させ得る。たとえば、協働ルーチン200が、インタプリタ170から、第2のデバイス110bに割り当てられたホットワードと部分的に一致するだけの用語をASR結果142が含むことを示す命令172を受け取ると、協働ルーチン200は、目を覚ますように第2のデバイス110bを呼び出すことが可能である。

【００３７】

協働ルーチン200は、委任ステージ(delegation stage)210と遂行ステージ(fulfillment stage)220とを含み得る。委任ステージ210の間、協働するデバイス110、100a～bは、協働するデバイスの得賃少なくとも1つに処理命令を指定することによって、互いに協働する。単純にするため、第1のデバイス110aと第2のデバイス110bとに対応する2つの協働するデバイス110が存在するが、しかし、発話が2より多くのデバイス110に向けられたとインタプリタが決定するときには、他の例は、2より多くの協働するデバイス110を含み得る。処理命令212は、第1の協働するデバイス110aを、オーディオデータ14に対してASR結果142を生成し、ASR結果142に対してクエリ解釈を行って、行うべき動作を特定するクエリ22をASR結果142が識別すると決定し、ASR結果142に対して実行されたクエリ解釈を他方の協働するデバイス110bと共有するように、指定し得る。この例では、オーディオデータ14は、第2のデバイス110bに割り当てられたホットワードと部分的に一致する1または複数の用語を含み得るだけであり、したがって、委任ステージ210は、目を覚まさせ第1のデバイス110aと協働するように第2のデバイス110bを同時に呼び出しながら、オーディオデータ14の処理を継続している第1のデバイス110aに、行うべき動作を特定するクエリ22を識別させることを決定し得る。他の例では、処理命令212は、代わりに、協働するデバイスが、オーディオデータ14に対してASR結果142を各々が独立に生成させ、クエリ22を識別するためにASR結果142に対してクエリ解釈を行うことにより、協働するデバイスが互いに協働することを許容させ得る。

【００３８】

協働するデバイス110が、他のデバイスがその態様を行わない間に、音声処理および/またはクエリ解釈のうちのいくつかの態様を行うときには、ルーチン202は、ルーチン202の実行を調整するために、どの協働するデバイス110が他の協働するデバイス110と情報を共有する必要があるのかを指定し得る。たとえば、第1のデバイス110aが「私のジャズプレイリストを再生して下さい」というクエリ22に対するクエリ解釈を行う場合には、第2のデバイス110bは、解釈が第2のデバイス110bと共有されるまで、このクエリ解釈について知らないことになる。さらに、ルーチン202が、第1のデバイス110aが音声処理を行い第2のデバイス110bがクエリ解釈を行うと指定する場合には、第2のデバイスのアクションは、第1のデバイスのアクションに依存するので、第1のデバイス110aは、第2のデバイス110bがクエリ解釈を行うことを可能にするために、音声認識結果142を第2のデバイス110bと共有することを必要とする。

【００３９】

処理命令212を生じさせるときには、委任ステージ210は、処理能力、電力使用、バッテリレベル、デバイス110において利用可能なAEDモデル、それぞれのデバイスがローカルにまたはリモートでASRを遂行できるかどうか、またはデバイス110に関連付けられるいずれかの他の能力/パラメータなど、各々の協働するデバイス110の能力を評価することがあり得る。たとえば、特定の協働するデバイス110が、リソース集約的な動作を行うために、本質的に、より大きな処理リソースを有することがあり得る。言い換えると、第1のデバイス110aがスマートウォッチなどのような限定された処理リソースを有するデバイ

ス110であり、第2のデバイス110bがタブレットであるときには、スマートウォッチは、処理リソースに関して、タブレットよりも、はるかに制約され得る。したがって、協働するデバイス110の一方がスマートウォッチであるときには、委任ステージ210は、可能な場合には常に、協働する他方のデバイス110上での音声処理とクエリ解釈との遂行を、指定することがあり得る。

【００４０】

遂行ステージ220は、協働するデバイス110のうちの少なくとも1つによってオーディオデータ14から解釈されたクエリ22を、受け取る。いくつかの例では、クエリ22は、協働するデバイス110の各々に対して行うデバイスレベルのアクションを特定する。たとえば、照明を暗くする動作を特定する図1のスマートライト110d、110eの方向に向けられたクエリ22は、デバイスレベルのクエリに対応し、この場合、遂行ステージ220は、スマートライト110d、110eに対して、各々がそれらの照射を暗い照明の特徴であるレベルまで独立に低下させることによって、互いに協働するように命令し得る。

【００４１】

他の例では、クエリ22は、協働するデバイス110によって共に行われる、時間を要する動作を特定する。時間を要する動作を行うために、これらのデバイス110は、その時間を要する動作と関係する複数のサブアクション222、222a～nを行う際に、協働することが要求される。それゆえ、協働するデバイス110は、時間を要する動作の継続時間の間は相互との対を形成し、協働するデバイス110のそれぞれの間のその時間を要する動作と関係するサブアクション222の遂行を調整することによって、互いに協働し得る。したがって、遂行ステージ220は、時間を要する動作と関係するサブアクション222を識別し、協働するデバイス110の間のそれらのサブアクションの遂行を調整する。

【００４２】

上述の例を引き続き用いると、クエリ22は、ユーザの居間に位置するスマートスピーカに対応する第1および第2のデバイス110a、110b上でユーザのジャズプレイリストを可聴的に再生する、時間を要する動作を特定する。この時間を要する動作を行うに、遂行ステージ220は、この時間を要する動作と関係するサブアクション222を識別し、第1のデバイス110aと第2のデバイス110bとに、相互との対を形成させ、第1のデバイス110aと第2のデバイス110bとの間のこの時間を要する動作と関係するサブアクション222の遂行を調整させる遂行命令225を、生成する。たとえば、ユーザのジャズプレイリストを再生するには、ジャズ音楽のプレイリストがローカルにアクセスされるか(たとえば、プレイリストが、デバイス110a～bのうちの1つに記憶されている)、ローカルネットワーク150a(図1B)上のネットワークストレージデバイス(図示せず)からアクセスされるか、または何らかのリモートサーバに存在する音楽ストリーミングサービスからストリーミングされるかのいずれかが、あり得る。この例では、ユーザのジャズプレイリストは、音楽ストリーミングサービスに関連付けられるストリーミング音楽アプリケーションにおけるプレイリストである。ここで、遂行命令225は、第2のデバイス110bに、現在の曲をジャズ音楽プレイリストからリモートネットワークを経由してストリーミングするためにリモート音楽ストリーミングサービスと接続している音楽ストリーミングアプリケーションを起動するサブアクションを行い、その曲を第1のデバイス110aに送信/ストリーミングさせ、現在の曲を演奏するオーディオ再生の責任を左側のオーディオチャネルとして想定させるように、命令し得る。他方で、遂行命令225は、第1のデバイス110aには、第2のデバイス110bからストリーミングされる現在の曲を演奏するオーディオ再生の責任を、右側のオーディオチャネルとして想定するというサブアクション222を行うことだけを、命令する。したがって、遂行命令225は、2つのデバイス110a、110bが音楽をステレオ構成で再生するように時間を要する動作を果たすために、第1のデバイス110aと第2のデバイス110bとの間のサブアクションの遂行を調整する。すると、プレイリストからの曲のストリーミングに対応するサブアクション222は、この時間を要する動作が終了する(たとえば、プレイリストが終わるか、またはユーザ10がデバイス110a～bにおいて音楽再生を停止するとき)まで反復する。時間を要する動作が終了すると、デバイス110は、分離して(たとえば、それ

らのペアリングされた接続が終了する)、低電力の状態(たとえば、休眠または冬眠状態)に戻る。

【００４３】

図3は、複数のホットワード24を、図1A～図1Dの例に類似する単一の発話20に組み合わせる例である。図3は、複数のホットワード24の各々が異なるデバイス110と対応する代わりに、複数のホットワード24が異なるアシスタントインターフェース120と対応するという点で、図1A～図1Dの例と異なる。すなわち、単一の発話20において組み合わされている複数のホットワード24は、デバイス固有ではなく、インターフェース固有である。アシスタントインターフェース120とは、デバイス110のデータ処理ハードウェア112d上で動作する1または複数のアプリケーションを指し得る。たとえば、インターフェース120は、メディアアプリケーション(たとえば、ビデオストリーミングアプリケーション、オーディオストリーミングアプリケーション、メディアプレーヤアプリケーション、メディアギャラリアプリケーションなど)、ワードプロセシングアプリケーション、ナビゲーションアプリケーション、ソーシャルメディアアプリケーション、通信アプリケーション(たとえば、メッセージングアプリケーション、電子メールアプリケーションなど)、ファイナンシャルアプリケーション、組織アプリケーション(たとえば、アドレス帳アプリケーション)、リテールアプリケーション、エンターテイメントアプリケーション(たとえば、ニュースアプリケーション、天気アプリケーション、スポーツアプリケーション)、キャストアプリケーションなどの、異なるアプリケーションとのインターフェースとなるアプリケーションプログラミングアプリケーション(API)である。いくつかのインターフェース120は、それらのアプリケーションとのインターフェースとなるために、または、おそらくは、その特定の企業のビジネス上の提供に特有のある程度の機能を含むために、企業によって開発されたプロプライエタリなソフトウェアである。図3に示されているように、2つのより一般的なアシスタントインターフェース120が、グーグル(たとえば、グーグルアシスタント(Google Assistant)と称される)とアマゾン(たとえば、アレクサ(Alexa)と称される)と、によって提供されている。それぞれのインターフェース120は、ユーザ10によってアシスタントインターフェース120に向かって話された発話20において受け取られたクエリ22に関連付けられる動作、タスク、またはアクションを行うようにインターフェース120をトリガするための、それ自体に特有の組のホットワード24を有し得る。

【００４４】

それぞれのインターフェース120は、デバイス110との通信中である他のアプリケーションとの間で異なる互換性を有し得るか、またはそれ自体の組の特有の利点を有し得るので、デバイス110のユーザ10は、多くの場合に、特定のデバイス110上で複数のインターフェース120を用い得る。また、ユーザ10は、特定のクエリ22に対して、結果/応答を比較するために、または複数の観点を得るために、同一のアクションを行うのに2つの異なるインターフェース120を用いることさえあり得る。たとえば、ユーザ10は、暴風雨または降雨を生じさせる天候に関しては、第1のインターフェース120、120aの天気予報機能の方が、第2のインターフェース120、120bの天気予報機能よりも正確であるが、他方で、湿気および暖かい天候に関しては、第2のインターフェース120、120bの天気予報機能の方が、第1のインターフェース120、120aの天気予報機能よりも正確であると考えるかもしれない。このような視点により、ユーザ10は、通常であれば2つの別個の発話20「やあグーグル、きょうの天気はどんな様子になるだろうか?」および「アレクサ、きょうの天気はどんな様子になるだろうか?」となるだろうものを、「やあグーグルとアレクサ、きょうの天気はどんな様子になるだろうか?」という単一の発話20に組み合わせることがあり得る。図3では、例が、買い物の質問を、参照している。この場合、ユーザ10は、価格設定を比較するために、または市場における価格設定に関してより多くのデータを収集するために、「やあグーグルとアレクサ、レイザークレスト(Razor Crest)のレゴセットはいくらだろうか?」と言うことによって、あるレゴセットの価格を求めて、アマゾンとグーグルとの両方にクエリ22を発する場合があり得る。

【００４５】

ホットワード24はデバイス固有ではなくインターフェース固有であるが、デバイス110の他の特徴は、同様に機能する。たとえば、インターフェース固有のホットワード24の場合には、デバイス110は、図3で見ることが可能なように、ホットワード検出器130と、音声認識器140と、コラボレータ200と、を含む。言い換えると、デバイス110が、ユーザ10によって話された発話20に対応するオーディオデータ14を受け取り、ホットワード検出器130が、オーディオデータ14において、第1のホットワード24、24aである「やあグーグル」を検出するのであるが、この場合に、第1のホットワード24、24aは、第1のデジタルアシスタント120aに割り当てられている。音声認識器140は、オーディオデータ14に対するASR結果142を生成し、インタプリタ170は、オーディオデータ14に対するASR結果142が、(たとえば、アレクサに割り当てられた)第2のデジタルアシスタント120bに割り当てられた第2のホットワード24と少なくとも部分的に一致しクエリ22に先行する1または複数の用語を含むかどうかを決定する。インタプリタ170は、既に論じられたように、ホットワードレジストリ180にアクセスして、用語「アレクサ」が第2のデジタルアシスタント120bに割り当てられたホットワードと一致することを決定することができる。

【００４６】

第2のデジタルアシスタント120bに割り当てられた1または複数の第2のホットワード24と少なくとも部分的に一致しクエリ22に先行する1または複数の用語をオーディオデータ14が含むという決定に基づき、インタプリタ170は、第1のデジタルアシスタント120aと第2のデジタルアシスタント120bとを互いに協働させて動作の遂行を果たす協働ルーチン200を開始させる命令172を送る。図1A～図1Dの例とは対照的に、複数のデジタルアシスタントインターフェース120(たとえば、第1および第2のデジタルアシスタントインターフェース120a～b)は、協働してデバイス110ではなくクエリ22に関連付けられる動作の遂行を果たす。これは、クエリ22のアクションまたはサブアクション222が、複数のインターフェース120によって並列的に(たとえば、同時に)行われ得ることを意味する。

【００４７】

クエリ22に対応する動作の遂行を複数のインターフェース120が果たしているときには、異なるインターフェース120が、異なる方法で、クエリ22を果たし得る。たとえば、あるインターフェース120が、他のインターフェース120とは異なるサービスに関連付けられることがあり得るし、またはあるインターフェース120が、異なる遂行結果を生成させることがあり得るが、その理由は、インターフェース120は、他のインターフェース120とは異なるリソースへのアクセスを有するからである。いくつかの実装形態では、異なるインターフェース120が、デバイス110のために、異なる種類のアクションを行う、または制御する。たとえば、あるインターフェース120が、ある様態で、デバイスレベルのアクションを行い、他のインターフェース120が、異なる様態で、同じデバイスレベルのアクションを行うことがあり得る。例示すると、ユーザ10が、「やあグーグルとアレクサ、データロギングをオフにして下さい」という発話20を、話したとする。この発話におけるクエリ22は、先に述べた、グーグルに関連付けられる第1のインターフェース120aが、第1のインターフェース120aのデータロギング機能を非アクティブ化するという照明の例と類似しているが、しかし、アマゾンに対応する第2のインターフェース120bでは、データロギング機能を非アクティブ化しない。代わりに、第2のインターフェース120bが、第1のインターフェース120aと同様に、そのデータロギング機能を、独立に非アクティブ化する。

【００４８】

独立に動作することに加えて、複数のインターフェース120は、応答を同期させるように、協働し得る。たとえば、第1のインターフェース120aが、「きょうの天気はどんな様子になるだろうか?」というサーチクエリ22に対して、「きょうの予報は晴れです」と応答するときには、第2のインターフェース120bは、肯定する(たとえば、「同意します」)ことによって、または第1のインターフェース120aの応答に異議を唱えることによって、第1のインターフェース120aと協働するように構成され得る。また、より詳細な応答をユ

ーザに提供するために、応答の一部が、一方のインターフェースから提供され、応答の他の一部が、他方のインターフェースから得られるという場合もあり得る。

【００４９】

図4は、デバイス固有のホットワード24を単一の発話20において組み合わせる方法400のための、例示的な動作構成のフローチャートである。動作402では、方法400は、第1のAEDデバイス110、110aのデータ処理ハードウェア112dにおいて、ユーザ10によって話されユーザ10に関連付けられる2以上のAED110、110a～nの中の第1のAED110aと第2のAED110、110bとの方向に向けられた発話20に対応するオーディオデータ14を受け取るのであるが、ここで、オーディオデータ14は、行うべき動作を特定するクエリ22を含む。動作404では、方法400は、ホットワード検出モデルを用いて、オーディオデータ14において第1のホットワード24、24aを検出し、第1のホットワード24aは、第1のAED110aに割り当てられ、第2のAED110bに割り当てられた第2のホットワード24、24bとは異なる。オーディオデータ14中の第1のAED110aに割り当てられた第1のホットワード24aを検出したことに応答して、動作406では、方法400が、オーディオデータ14に対する処理を開始し、第2のAED110bに割り当てられた第2のホットワード24bと少なくとも部分的に一致しクエリ22に先行する1または複数の用語をオーディオデータ14が含むことを決定する。第2のホットワード24bと少なくとも部分的に一致しクエリ22に先行する1または複数の用語をオーディオデータ14が含むという決定に基づいて、動作408において、方法400は、第1のAED110aと第2のAED110bとを互いに協働させてクエリ22によって特定された動作の遂行を果たす協働ルーチン202を、実行する。

【００５０】

図5は、アシスタント固有のホットワード24を単一の発話20において組み合わせる方法500のための例示的な動作構成のフローチャートである。動作502では、方法500は、アシスタント対応デバイス(AED)110、110aのデータ処理ハードウェア112dにおいて、ユーザ10によって話されAED110aによって捕捉された発話20に対応するオーディオデータ14を受け取り、ここで、発話20は、動作を行うための第1のデジタルアシスタント120、120aと第2のデジタルアシスタント120、120bとに対するクエリ22を含む。動作504では、方法500は、データ処理ハードウェア112dによって、第1のホットワード検出モデルを用いて、オーディオデータの中の第1のホットワード24、24aを検出し、ここで、第1のホットワード24aは、第1のデジタルアシスタント120aに割り当てられ、第2のデジタルアシスタント120bに割り当てられた第2のホットワード24、24bとは異なる。動作506では、方法500は、データ処理ハードウェア112dによって、第2のデジタルアシスタント120bに割り当てられた第2のホットワード24bと少なくとも部分的に一致しクエリ22に先行する1または複数の用語をオーディオデータ14が含むことを決定する。第2のホットワード24bと少なくとも部分的に一致しクエリ22に先行する1または複数の用語をオーディオデータ14が含むという決定に基づき、動作508において、方法500は、データ処理ハードウェア112dによって、第1のデジタルアシスタント120aと第2のデジタルアシスタント120bとを互いに協働させて動作の遂行を果たす協働ルーチン202を実行する。

【００５１】

図6は、本文書に記載されているシステムおよび方法を実装するのに用いられ得る例示的なコンピューティングデバイス600の概略図である。コンピューティングデバイス600は、ラップトップ、デスクトップ、ワークステーション、パーソナルデジタルアシスタント、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、様々な形式のデジタルコンピュータを表すことが意図されている。本明細書に示されるコンポーネント、それらの接続および関係、ならびにそれらの機能は、例示であることのみが意味されており、本文書において記載および/または請求される本発明の実装形態を限定することは、意味されていない。

【００５２】

コンピューティングデバイス600は、プロセッサ610と、メモリ620と、ストレージデバイス630と、メモリ620および高速拡張ポート650に接続する高速インターフェース/コ

ントローラ640と、低速バス670およびストレージデバイス630に接続する低速インターフェース/コントローラ660と、を含む。コンポーネント610、620、630、640、650、および660のそれぞれは、様々なバスを用いて相互接続され、共通のマザーボード上に、または必要に応じて他の様態で搭載され得る。プロセッサ610は、グラフィカルユーザインターフェース(GUI)のためのグラフィカル情報を、高速インターフェース640に結合されたディスプレイ680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620の中にまたはストレージデバイス630上に記憶された命令を含めて、コンピューティングデバイス600内部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複数のタイプのメモリと共に、用いられることがある。また、複数のコンピューティングデバイス600が接続されることがあるが、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、一群のブレードサーバ、またはマルチプロセッサシステムとして)提供する。

【００５３】

メモリ620は、コンピューティングデバイス600の内部に、情報を非一時的に記憶する。メモリ620は、コンピュータ可読媒体、揮発性メモリユニット、または不揮発性メモリユニットであり得る。非一時的メモリ620は、プログラム(たとえば、命令のシーケンス)またはデータ(たとえば、プログラム状態情報)を、コンピューティングデバイス600による使用のために、一時的または恒久的に記憶するのに用いられる物理デバイスであり得る。不揮発性メモリの例は、これらに限定されることはないが、フラッシュメモリおよびリードオンリメモリ(ROM)/プログラマブルなリードオンリメモリ(PROM)/消去可能でプログラマブルなリードオンリ専用メモリ(EPROM)/電子的に消去可能なプログラマブルなリードオンリメモリ(EEPROM)(たとえば、典型的には、ブートプログラムなどのファームウェア用に用いられる)を含む。揮発性メモリの例は、これらに限定されないが、ランダムアクセスメモリ(RAM)、ダイナミックランダムアクセスメモリ(DRAM)、スタティックランダムアクセスメモリ(SRAM)、相変化メモリ(PCM)、およびディスクまたはテープを含む。

【００５４】

ストレージデバイス630は、コンピューティングデバイス600のための大容量ストレージを提供することが可能である。いくつかの実装形態では、ストレージデバイス630は、コンピュータ可読媒体である。様々な異なる実装形態において、ストレージデバイス630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ソリッドステートメモリデバイス、またはストレージ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であり得る。追加的な実装形態では、コンピュータプログラム製品が、情報キャリアとして、有体的に具体化される。コンピュータプログラム製品は、実行されると、上述したような1または複数の方法を行う命令を含む。情報キャリアは、メモリ620、ストレージデバイス630、またはプロセッサ610上のメモリなど、コンピュータまたは機械可読な媒体である。

【００５５】

高速コントローラ640は、コンピューティングデバイス600のための帯域集約的な動作を管理し、他方で、低速コントローラ660は、より帯域集約的でない動作を管理する。義務のそのような配分は、例示的なものにすぎない。いくつかの実装形態において、高速コントローラ640は、メモリ620、ディスプレイ680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介して)、および様々な拡張カード(図示せず)を受け入れる場合がある高速拡張ポート650に結合される。いくつかの実装形態において、低速コントローラ660は、ストレージデバイス630および低速拡張ポート690に結合される。低速拡張ポート690は、様々な通信ポート(たとえば、USB、Bluetooth、イーサネット、無線イーサネット)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介して結合され得る。

【００５６】

コンピューティングデバイス600は、図に示されるように、いくつかの異なる形式で、実装され得る。たとえば、それは、標準サーバ600aまたは複数回にわたる一群のサーバ600aとして、ラップトップコンピュータ600bとして、またはラックサーバシステム600cの一部として、実装され得る。

【００５７】

本明細書に記載されているシステムおよび技法の様々な実装形態は、デジタル電子および/もしくは光学回路構成、集積回路構成、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、ならびに/またはそれらの組合せとして、実現され得る。これらの様々な実装形態は、少なくとも1つのプログラマブルなプロセッサを含むプログラマブルなシステム上で実行可能および/または解釈可能な1または複数のコンピュータプログラムとしての実装を含み得るが、ここで、プログラマブルなプロセッサは、ストレージシステム、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、専用または汎用のものであり得る。

【００５８】

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、プログラマブルなプロセッサのための機械命令を含んでおり、高水準手続型および/もしくはオブジェクト指向プログラミング言語として、ならびに/またはアセンブリ/機械言語として、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、いずれかのコンピュータプログラム製品、非一時的コンピュータ可読媒体、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、プログラマブルな論理デバイス(PLD))を指しており、これらは、機械命令を機械可読信号として受け取る機械可読媒体を含むプログラマブルなプロセッサに、機械命令および/またはデータを提供するのに用いられる。「機械可読信号」という用語は、プログラマブルな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用いられるいずれかの信号を指している。

【００５９】

本明細書に記載されているプロセスおよび論理フローは、入力データに対して動作し、出力を生成することによって機能を行うための1または複数のコンピュータプログラムを実行する、データ処理ハードウェアとも称される1または複数のプログラマブルなプロセッサによって、遂行可能である。プロセスおよび論理フローは、また、たとえばFPGA(フィールドプログラマブルゲートアレイ)やASIC(特定用途向け集積回路)など、専用の論理回路構成によっても遂行可能である。コンピュータプログラムの実行に適したプロセッサは、例として、汎用および専用のマイクロプロセッサの両方、ならびにいずれかの種類のデジタルコンピュータのいずれか1または複数のプロセッサを含む。一般に、プロセッサは、リードオンリメモリもしくはランダムアクセスメモリまたはそれらの両方から、命令とデータとを受け取ることになる。コンピュータの本質的要素は、命令を行うためのプロセッサと、命令とデータとを記憶するための1または複数のメモリデバイスと、である。一般に、コンピュータは、また、たとえば磁気、光磁気ディスク、もしくは光ディスクなど、データを記憶するための1または複数の大容量ストレージデバイスを含んでいるか、または大容量ストレージデバイスからデータを受け取り、もしくはデータを転送し、もしくはそれら両方を行うように、大容量ストレージデバイスに動作可能に結合される。ただし、コンピュータが、そのようなデバイスを有していることは、必要ない。コンピュータプログラム命令およびデータを記憶するのに適したコンピュータ可読媒体は、例として、たとえばEPROM、EEPROM、およびフラッシュメモリデバイスなどの半導体メモリデバイスと、たとえば内部ハードディスクまたは取り外し可能ディスクなどの磁気ディスクと、光磁気ディスクと、CD-ROMおよびDVD-ROMディスクとを含む、あらゆる形式の不揮発性メモリ、媒体およびメモリデバイスを含む。プロセッサとメモリとは、専用の論理回路構成によって補完されることが、またはその中に組み込まれることが、可能である。

【００６０】

ユーザとの対話を提供するために、本開示の1または複数の態様は、コンピュータ上で実装されることが可能であり、コンピュータは、たとえば、CRT(陰極線管)、LCD(液晶ディスプレイ)モニタ、またはタッチスクリーンなど、ユーザのために情報を表示するためのディスプレイデバイスと、オプションであるが、たとえばマウスやトラックボールなど、ユーザがコンピュータに入力を与えることを可能にするためのキーボードおよびポインティングデバイスと、を有する。他の種類のデバイスをユーザとの対話を提供するのに用いられることも、同様に可能であって、たとえば、ユーザに提供されるフィードバックは、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など、いずれかの形式の感覚フィードバックであることが可能であり、ユーザからの入力は、音響、音声、または触覚入力を含む、いずれかの形式として受け取られることが可能である。さらに、コンピュータは、たとえば、ユーザのクライアントデバイス上のウェブブラウザへ、そのウェブブラウザから受け取られたリクエストに応答してウェブページを送るなど、ユーザによって用いられるデバイスへドキュメントを送り、ユーザによって用いられるデバイスからドキュメントを受け取ることにより、ユーザと対話することができる。

【００６１】

いくつかの実装形態が、以上で、説明された。しかし、本開示の精神および範囲から逸脱することなく様々な修正が行われ得る、ということが理解されるであろう。したがって、他の実装形態も、以下の特許請求の範囲に属する。

【符号の説明】

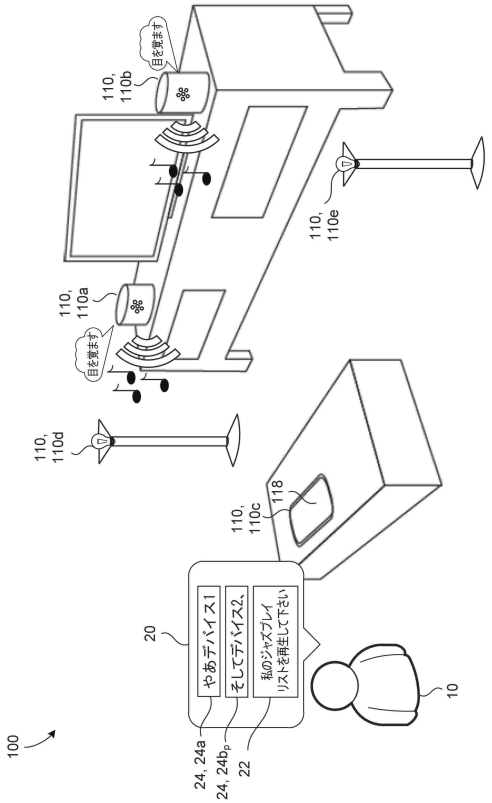
【００６２】

10 ユーザ
12 ストリーミングオーディオ
20 発話
22 クエリ
24 ホットワード、ホットワード候補
24a 第1のホットワード
24p 部分ホットワード
100 音声環境
110 アシスタント対応デバイス(AED)、デバイス、ユーザデバイス
110a 第1のデバイス
110b 第2のデバイス
110c デバイス、第3のデバイス、モバイルデバイス
110d AED、デバイス、第1のスマート電球、スマートライト
110e AED、デバイス、第2のスマート電球、スマートライト
112d データ処理ハードウェア
112m メモリハードウェア
114 オーディオ捕捉デバイス(マイクロフォン)
116 オーディオ再生デバイス、スピーカ
118 ディスプレイ
120 音声対応インターフェース、アシスタントインターフェース
130 ホットワード検出器
140 音声認識器、音声認識システム
142 自動音声認識(ASR)結果
150 ネットワーク
160 リモートシステム
162 リモートリソース
164 リモートデータ処理ハードウェア
166 リモートメモリハードウェア
170 インタプリタ
172 命令

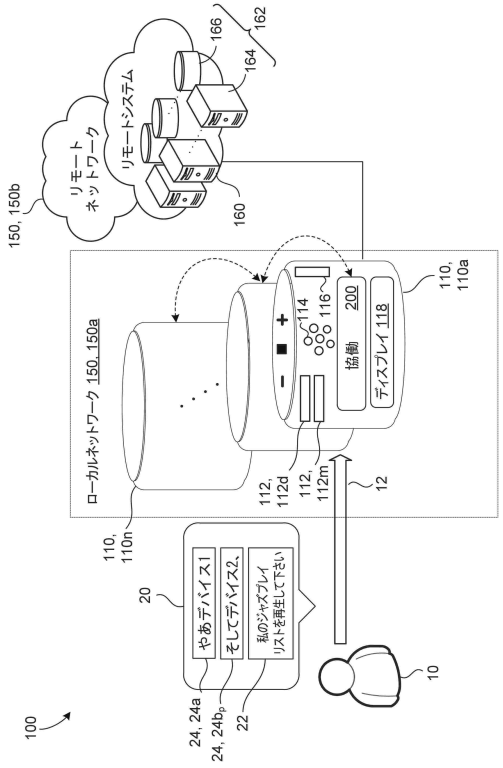
175 機械学習モデル
180 ホットワードレジストリ
182、182a～n リスト
200 協働ルーチン
210 委任ステージ
212 処理命令
220 遂行ステージ
222 サブアクション
225 遂行命令
600 コンピューティングデバイス
600a 標準サーバ
600b ラップトップコンピュータ
600c ラックサーバシステム
610 プロセッサ
620 メモリ
630 ストレージデバイス
640 高速インターフェース/コントローラ
650 高速拡張ポート
660 低速インターフェース/コントローラ
670 低速バス
680 ディスプレイ
690 低速拡張ポート

【図面】

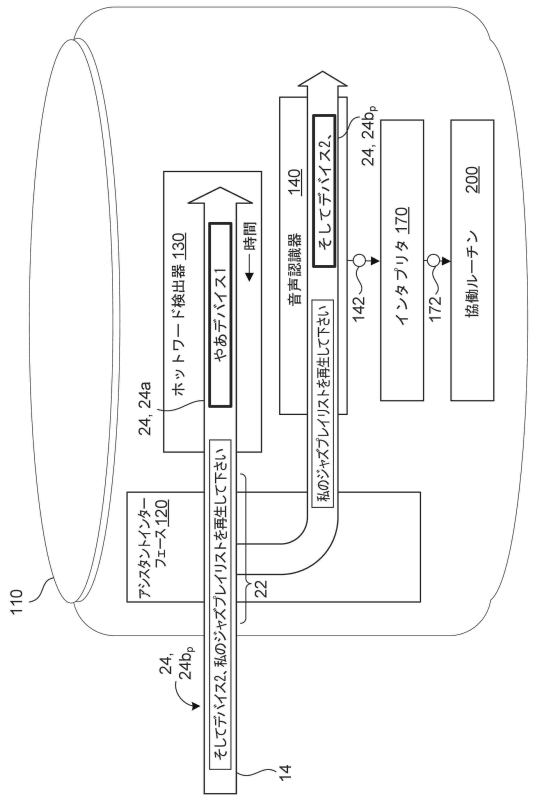
【図１Ａ】



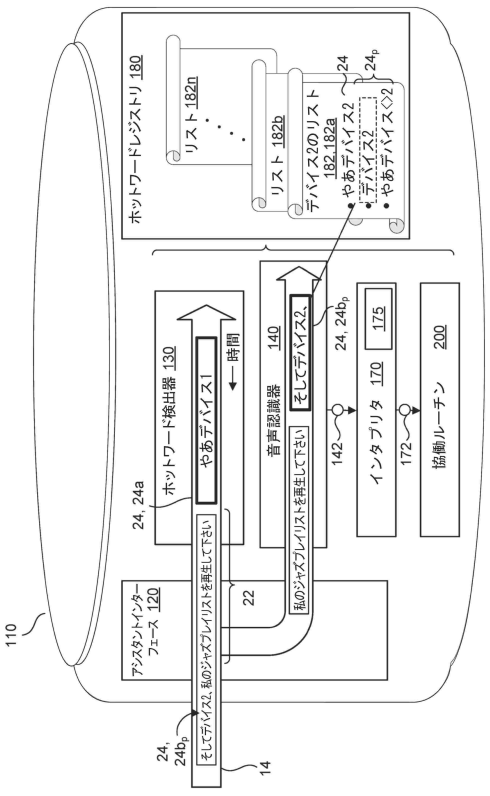
【図１Ｂ】



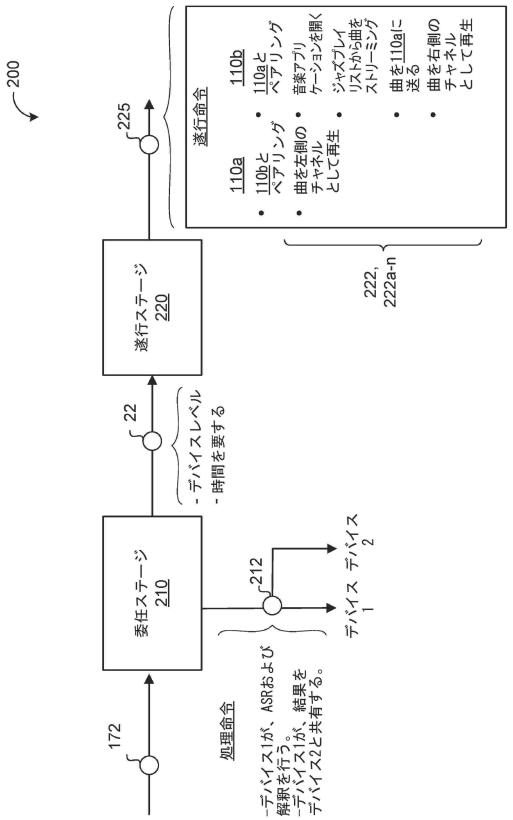
【図１Ｃ】


110
アシスタントインターフェース 120
24, 24a
ホットワード検出器 130
やあデバイス1
時間
私のジャズプレイリストを再生して下さい
音声認識器 140
そしてデバイス2、
142
インタプリタ 170
172
協働ルーチン 200
22
24, 24bp
そしてデバイス2、私のジャズプレイリストを再生して下さい
14

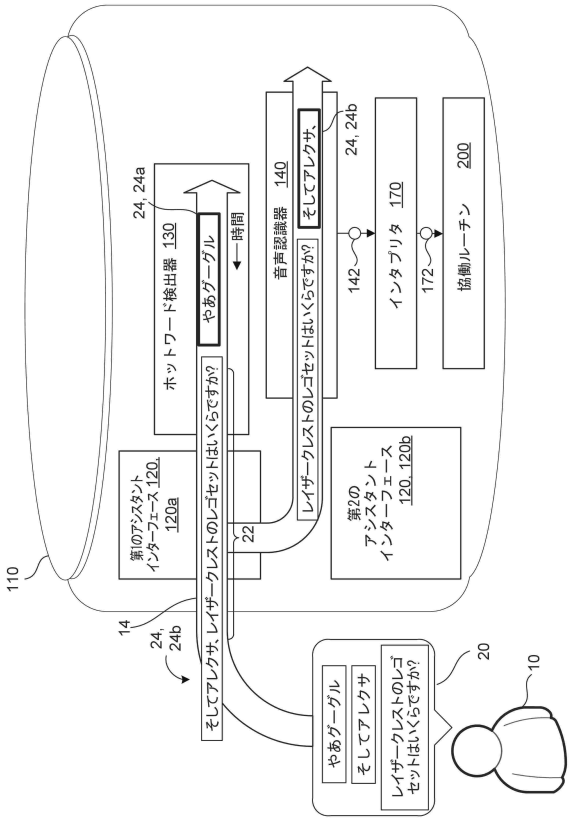
【図１Ｄ】


110
ホットワードレジストリ 180
リスト 182n
リスト 182b
デバイス2のリスト 182, 182a
24
24p
やあデバイス2
デバイス2
やあデバイス◇2
アシスタントインターフェース 120
24, 24a
ホットワード検出器 130
やあデバイス1
時間
私のジャズプレイリストを再生して下さい
音声認識器 140
そしてデバイス2、
142
24, 24bp
インタプリタ 170
175
172
協働ルーチン 200
22
24, 24bp
そしてデバイス2、私のジャズプレイリストを再生して下さい
14

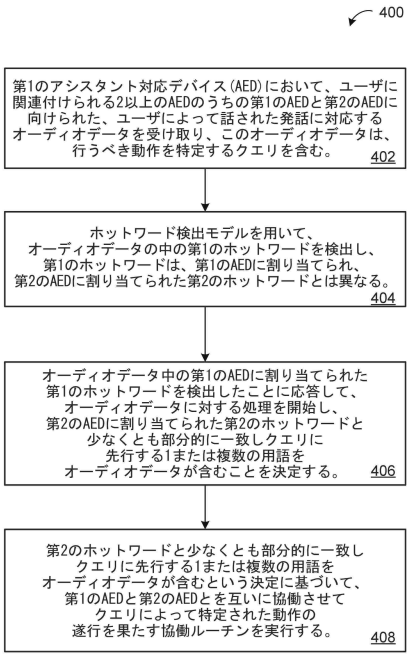
【図２】


200
225
遂行命令
110a
• 110bとペアリング
• 曲を左側のチャネルとして再生
110b
• 110aとペアリング
• 音楽アプリケーションを開く
• ジャズプレイリストから曲をストリーミング
• 曲を110aに送る
• 曲を右側のチャネルとして再生
222, 222a-n
遂行ステージ 220
22
- デバイスレベル
- 時間を要する
委任ステージ 210
212
デバイス 1
デバイス 2
172
処理命令
-デバイス1が、ASRおよび解釈を行う。
-デバイス1が、結果をデバイス2と共有する。

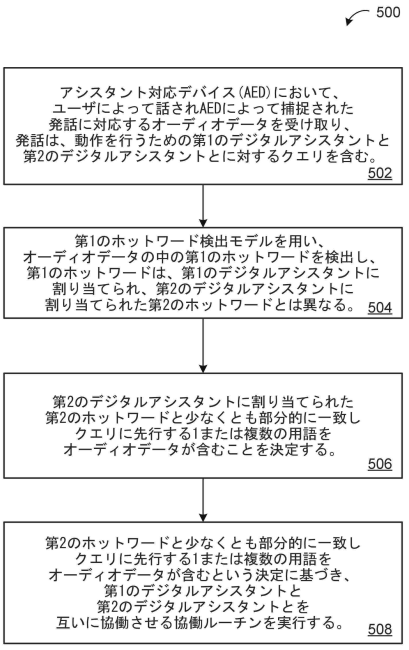
【図３】


110
第1のアシスタントインターフェース 120, 120a
第2のアシスタントインターフェース 120, 120b
24, 24a
ホットワード検出器 130
やあグーグル
時間
そしてアレクサ、レイザークレストのレコセットはいくらですか?
音声認識器 140
そしてアレクサ、
レイザークレストのレコセットはいくらですか?
142
24, 24b
インタプリタ 170
172
協働ルーチン 200
22
14
24, 24b
やあグーグル
そしてアレクサ
レイザークレストのレコセットはいくらですか?
20
10

【図４】


400
第1のアシスタント対応デバイス(AED)において、ユーザに関連付けられる2以上のAEDのうちの第1のAEDと第2のAEDに向けられた、ユーザによって話された発話に対応するオーディオデータを受け取り、このオーディオデータは、行うべき動作を特定するクエリを含む。 402
ホットワード検出モデルを用いて、オーディオデータの中の第1のホットワードを検出し、第1のホットワードは、第1のAEDに割り当てられ、第2のAEDに割り当てられた第2のホットワードとは異なる。 404
オーディオデータ中の第1のAEDに割り当てられた第1のホットワードを検出したことに応答して、オーディオデータに対する処理を開始し、第2のAEDに割り当てられた第2のホットワードと少なくとも部分的に一致しクエリに先行する1または複数の用語をオーディオデータが含むことを決定する。 406
第2のホットワードと少なくとも部分的に一致しクエリに先行する1または複数の用語をオーディオデータが含むという決定に基づいて、第1のAEDと第2のAEDとを互いに協働させてクエリによって特定された動作の遂行を果たす協働ルーチンを実行する。 408

【図５】


500
アシスタント対応デバイス(AED)において、ユーザによって話されAEDによって捕捉された発話に対応するオーディオデータを受け取り、発話は、動作を行うための第1のデジタルアシスタントと第2のデジタルアシスタントとに対するクエリを含む。 502
第1のホットワード検出モデルを用い、オーディオデータの中の第1のホットワードを検出し、第1のホットワードは、第1のデジタルアシスタントに割り当てられ、第2のデジタルアシスタントに割り当てられた第2のホットワードとは異なる。 504
第2のデジタルアシスタントに割り当てられた第2のホットワードと少なくとも部分的に一致しクエリに先行する1または複数の用語をオーディオデータが含むことを決定する。 506
第2のホットワードと少なくとも部分的に一致しクエリに先行する1または複数の用語をオーディオデータが含むという決定に基づき、第1のデジタルアシスタントと第2のデジタルアシスタントとを互いに協働させる協働ルーチンを実行する。 508

【図６】

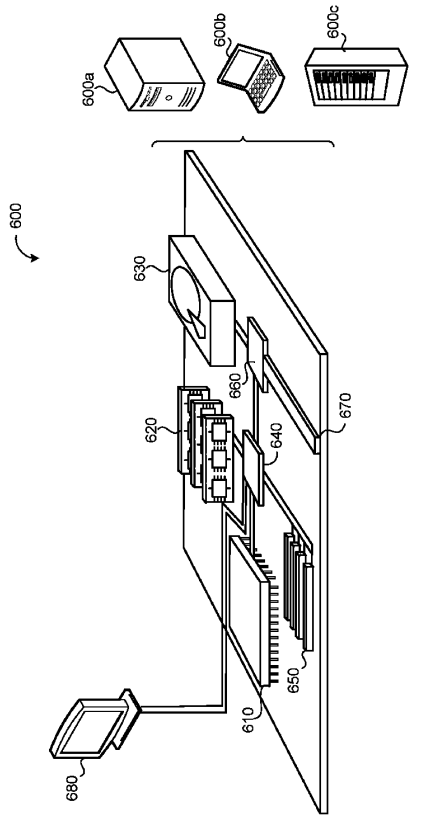

600
600a
600b
600c
630
660
620
640
670
650
610
680

FIG. 6

フロントページの続き

(72)発明者　マシュー・シャリフィ
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　ヴィクター・カルブネ
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　山下 剛史

(56)参考文献　米国特許出願公開第２０２０／０００７９８７（ＵＳ，Ａ１）
米国特許出願公開第２０１８／０２０４５６９（ＵＳ，Ａ１）
特表２０２０－５２８５６６（ＪＰ，Ａ）
米国特許第１０３６６６９２（ＵＳ，Ｂ１）
米国特許出願公開第２０２０／０３７２９０７（ＵＳ，Ａ１）
米国特許出願公開第２０２０／０２５８５１２（ＵＳ，Ａ１）

(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１３／００－９９／００