

(19)日本国特許庁(JP)

**(12)特許公報(B2)**

(11)特許番号  
**特許第7130637号**  
(P7130637)

(45)発行日 令和4年9月5日(2022.9.5)

(24)登録日 令和4年8月26日(2022.8.26)

(51)国際特許分類

G 10 L	15/28 (2013.01)	F I	G 10 L	15/28	230 J
G 10 L	15/22 (2006.01)		G 10 L	15/22	300 Z
G 10 L	15/10 (2006.01)		G 10 L	15/10	200 W
G 10 L	15/00 (2013.01)		G 10 L	15/00	200 A
G 06 F	3/16 (2006.01)		G 06 F	3/16	620

請求項の数 16 (全38頁) 最終頁に続く

(21)出願番号	特願2019-522753(P2019-522753)
(86)(22)出願日	平成29年11月3日(2017.11.3)
(65)公表番号	特表2020-500330(P2020-500330)
	A)
(43)公表日	令和2年1月9日(2020.1.9)
(86)国際出願番号	PCT/US2017/059955
(87)国際公開番号	WO2018/085671
(87)国際公開日	平成30年5月11日(2018.5.11)
審査請求日	令和2年10月21日(2020.10.21)
(31)優先権主張番号	62/417,281
(32)優先日	平成28年11月3日(2016.11.3)
(33)優先権主張国・地域又は機関	米国(US)
(31)優先権主張番号	15/801,307
(32)優先日	平成29年11月1日(2017.11.1)
	最終頁に続く

(73)特許権者	502208397 グーグル エルエルシー Google LLC アメリカ合衆国 カリフォルニア州 94043 マウンテン ビュー アンフィシアター パークウェイ 1600 Amphitheatre Parkway 94043 Mountain View, CA U.S.A.
(74)代理人	110001195弁理士法人深見特許事務所
(72)発明者	ミクスター,ケネス アメリカ合衆国、94043 カリフォルニア州、マウンテン・ビュー、anford シアター・パークウェイ、1600 シェケル,トマー
(72)発明者	最終頁に続く

(54)【発明の名称】 音声インターフェース装置におけるフォーカスセッション

**(57)【特許請求の範囲】****【請求項 1】**

1つ以上のマイクロフォン、スピーカ、1つ以上のプロセッサ、及び前記1つ以上のプロセッサによる実行のための1つ以上のプログラムを格納するメモリを含む第1の電子装置であって、共通のネットワークサービスに通信可能に結合される、コネクテッド電子装置のローカルグループのメンバである第1の電子装置が、

第1の操作の要求を含む第1の音声命令を受信することと、

コネクテッド電子装置の前記ローカルグループの中から前記第1の操作のための第1のターゲット装置を決定することと、

前記第1のターゲット装置に関してフォーカスセッションを確立することと、

前記共通のネットワークサービスの操作を介して、前記第1のターゲット装置によって前記第1の操作が実行されるようにすることと、

第2の操作の要求を含む第2の音声命令を受信することと、

前記第2の音声命令が第2のターゲット装置の明示的な指定を含まないと決定することと、

前記第2の操作が前記第1のターゲット装置によって実行することができると決定することと、

前記第2の音声命令が、前記第1の音声命令の受信に対して第1の所定の時間間隔内にまたは前記第1の音声命令の受信に対して第2の所定の時間間隔内に前記第1の電子装置によって受信されるかどうかを決定することによって、前記第2の音声命令が1つ以上のフ

オーカスセッション維持基準を満たすかどうかを判定することとを備え、前記第2の所定の時間間隔は前記第1の所定の時間間隔の後に続き、さらに、前記第1の電子装置が、

前記第2の音声命令が前記第2の所定の時間間隔内に受信されたという決定に従って、前記第2の音声命令のためのターゲット装置として前記第1のターゲット装置を確認する要求を出力すること、および確認する要求に応答する前記第1のターゲット装置の肯定的な確認に従って、前記第1のターゲット装置を前記第2の音声命令のためにターゲット装置として選択することと、

前記第2の音声命令が前記フォーカスセッション維持基準を満たすという決定に従つて、前記第2の操作を前記共通のネットワークサービスの操作を介して前記第1のターゲット装置によって実行させることとを備える、方法。 10

#### 【請求項2】

コネクテッド電子装置の前記ローカルグループの中から前記第1の操作のための第1のターゲット装置を決定することは、

前記第1の音声命令から前記第1のターゲット装置の明示的な指定を取得することを備える、請求項1に記載の方法。

#### 【請求項3】

コネクテッド電子装置の前記ローカルグループの中から前記第1の操作のための第1のターゲット装置を決定することは、

前記第1の音声命令が前記第1のターゲット装置の明示的な指定を含まないと決定することと、 20

コネクテッド電子装置の前記ローカルグループの中の第2の電子装置により前記第1の操作を実行することができると決定することと、

前記第2の電子装置を前記第1のターゲット装置として選択することとを備える、請求項1に記載の方法。

#### 【請求項4】

前記第2の音声命令は前記フォーカスセッション維持基準を満たすという前記決定に従つて、前記第1のターゲット装置に関して前記フォーカスセッションを延長することをさらに備える、請求項1から3のいずれか1項に記載の方法。

#### 【請求項5】

前記第1のターゲット装置に関して前記フォーカスセッションを確立することは、 30

前記第1の音声命令のタイムスタンプを記憶することと、

前記第1のターゲット装置の識別子を記憶することとを備える、請求項1から4のいずれか1項に記載の方法。

#### 【請求項6】

前記第2の音声命令が前記第1の所定の時間間隔内に受信されたという決定に従つて、前記第2の音声命令のためのターゲット装置として前記第1のターゲット装置を選択することをさらに備える、請求項1～5のいずれか1項に記載の方法。

#### 【請求項7】

第3の操作の要求と、コネクテッド電子装置の前記ローカルグループの中での第3のターゲット装置の明示的な指定とを含む、第3の音声命令を受信することと、 40

前記第1のターゲット装置に関してフォーカスセッションを終了することと、

前記第3のターゲット装置に関してフォーカスセッションを確立することと、

前記共通のネットワークサービスの操作を介して、前記第3のターゲット装置によって前記第3の操作を実行させることとをさらに備える、請求項1から6のいずれか1項に記載の方法。

#### 【請求項8】

第1のターゲット装置は前記第1の電子装置であり、

前記方法は、

第4の操作の要求と、コネクテッド電子装置の前記ローカルグループの中での第4の 50

ターゲット装置の明示的な指定とを含む、第4の音声命令を受信することをさらに備え、前記第4のターゲット装置は、コネクテッド電子装置の前記ローカルグループの第3の電子装置メンバであり、前記第3の電子装置は前記第1の電子装置とは異なり、前記方法は、さらに、

前記第1のターゲット装置に関して前記フォーカスセッションを維持することと、

前記共通のネットワークサービスの操作を介して、前記第4のターゲット装置によって前記第4の操作を実行させることとを備える、請求項1から7のいずれか1項に記載の方法。

#### 【請求項9】

前記第2の音声命令は、前記第4の操作を前記第4のターゲット装置によって実行させた後に受信され、10

前記第1の操作はメディア再生動作であり、

前記第2の操作はメディア停止操作であり、

前記方法は、

第5の操作の要求と、及びコネクテッド電子装置の前記ローカルグループの中での第5のターゲット装置の明示的な指定とを含む、第5の音声命令を受信することをさらに備え、前記第5のターゲット装置は前記第3の電子装置であり、前記方法は、

前記第1のターゲット装置に関して前記フォーカスセッションを終了することと、

前記第5のターゲット装置に関してフォーカスセッションを確立することと、

前記共通のネットワークサービスの操作を介して、前記第5のターゲット装置によって前記第5の操作を実行させることとをさらに備える、請求項8に記載の方法。20

#### 【請求項10】

所定の操作終了要求を含む第5の音声命令を受信することと、

前記第5の音声命令を受信することに従って、

前記第1のターゲット装置によって前記第1の操作を実行させることを止めることと、

前記第1のターゲット装置に関して前記フォーカスセッションを終了することとをさらに備える、請求項1から9のいずれか1項に記載の方法。

#### 【請求項11】

前記第1の操作はメディア再生操作であり、

前記第2の操作は、メディア停止操作、メディア巻き戻し操作、メディア早送り操作、音量を上げる操作、及び音量を下げる操作のうちの1つである、請求項1から10のいずれか1項に記載の方法。30

#### 【請求項12】

前記第1の操作は、複数の装置状態のうちの第1の状態への装置状態変化操作であり、

前記第2の操作は、複数の装置状態のうちの第2の状態への装置状態変更操作である、請求項1から11のいずれか1項に記載の方法。

#### 【請求項13】

前記第1の操作は、マグニチュードスケールにおいて、第1の方向へのマグニチュード変更操作であり、

前記第2の操作は、前記マグニチュードスケールにおいて、前記第1の方向と反対の第2の方向へのマグニチュード変更操作である、請求項1から12のいずれか1項に記載の方法。40

#### 【請求項14】

前記第1の電子装置は1つ以上のLEDのアレイを備え、

前記方法は、

LEDの前記アレイにおける前記LEDの1つ以上を照明することによって前記フォーカスセッションの状態を示すことをさらに備える、請求項1から13のいずれか1項に記載の方法。

#### 【請求項15】

1つ以上のマイクロフォン、50

スピーカ、

1つ以上のプロセッサ、及び

前記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格納するメモリを備え、前記1つ以上のプログラムは請求項1から1-4のいずれか1項に記載の方法を実行するための指示を備える、電子装置。

**【請求項16】**

1つ以上のマイクロフォン、スピーカ、及び1つ以上のプロセッサを備える電子装置によって実行されると、前記電子装置に請求項1から1-4のいずれか1項に記載の方法を実行させる命令を備える、コンピュータプログラム。

**【発明の詳細な説明】**

**【技術分野】**

**【0001】**

技術分野

開示された実施形態は、一般に、ターゲット装置が未知又は音声命令自体から曖昧であるときに音声命令をターゲット装置に送るための方法及びシステムを含むがこれらに限定されない、音声インターフェース及び関連装置に関する。

**【背景技術】**

**【0002】**

背景

音声インターフェースを有する電子装置は、ユーザから音声入力を収集し、音声入力に従って異なる音声起動機能を実行するために広く使用してきた。これらの音声起動機能は、操作を実行するようにターゲット装置に指示又は命令することを含み得る。例えば、ユーザは音声インターフェース装置へ音声入力を発して、ターゲット装置をオン又はオフにしたり、又はターゲット装置でのメディア再生を制御するように指示したりすることができる。

**【0003】**

典型的には、ユーザがターゲット装置に操作を実行させるように指示する音声入を行ないたい場合、ユーザは音声入力においてターゲット装置を指定するであろう。しかしながら、そのようなすべての音声入力に対してターゲット装置を明示的に指定しなければならないことは、ユーザにとって面倒であり、かつ煩わしいものである。音声入力がターゲットを指定しない場合、又は曖昧なターゲットを指定する場合でも、音声インターフェース装置は音声入力のためのターゲット装置を有することが望ましい。

**【発明の概要】**

**【課題を解決するための手段】**

**【0004】**

概要

したがって、音声入力におけるターゲット装置の指定が存在しないか又は曖昧な際でも、音声入力のためのターゲット装置を決定又は割り当てるための方法及びシステムを組み込んだ音声アシスタントシステム及び/又は音声アシスタントサーバシステムを有する、電子装置が必要である。本出願に記載されている様々な実施形態では、操作環境は、音声アシスタントサービスへのインターフェースを提供する音声起動式電子装置、及び音声アシスタントサービスを介した音声入力によって制御され得る複数の装置（例えばキャストデバイス、スマートホームデバイス）を含む。音声起動式電子装置は、音声入力を記録するように構成され、その音声入力から音声アシスタンスサービス（例えば音声アシスタンスサーバシステム）がユーザの音声要求（例えばメディア再生要求、電源状態変更要求）を決定する。次いで、音声アシスタンスサーバシステムは、ユーザの音声要求を、音声入力によって示されるターゲット装置に伝える。音声起動式電子装置は、後続の音声入力を記録するように構成されており、この音声入力においてはターゲット装置の指示が存在しないか又は曖昧である。電子装置又は音声アシスタンスサーバシステムは、この音声入力のためにターゲット装置を割り当て、この音声入力に含まれるユーザの音声要求を決定し、

10

20

30

40

50

割り当てられたターゲット装置にユーザの音声要求を送る。

**【0005】**

いくつかの実施形態に従って、1つ以上のマイクロフォン、スピーカ、1つ以上のプロセッサ、及び1つ以上のプロセッサによる実行のための1つ以上のプログラムを格納するメモリを有する第1の電子装置において、方法が実行される。第1の電子装置は、共通のネットワークサービスに通信可能に結合される、コネクテッド電子装置のローカルグループのメンバである。当該方法は、第1の操作の要求を含む第1の音声命令を受信すること；コネクテッド電子装置のローカルグループの中から、第1の操作のための第1のターゲット装置を決定することと、第1のターゲット装置に関してフォーカスセッションを確立することと、共通のネットワークサービスの操作を介して、第1のターゲット装置に第1の操作を実行させることと、第2の操作の要求を含む第2の音声命令を受信することと、第2の音声命令が第2のターゲット装置の明示的な指定を含まないと決定することと、第2の操作が第1のターゲット装置によって実行されると決定することと、第2の音声命令が1つ以上のフォーカスセッション維持基準を満たすかどうかを決定することと、及び第2の音声命令がフォーカスセッション維持基準を満たすとの決定に従って、共通のネットワークサービスの操作を介して、第2の操作を第1のターゲット装置に実行させることとを含む。

**【0006】**

いくつかの実施形態に従って、電子装置は、1つ以上のマイクロフォン、スピーカ、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格納するメモリを含む。1つ以上のプログラムは、上記の方法を実行するための命令を含む。

**【0007】**

いくつかの実施形態に従って、非一時的なコンピュータ読み取り可能な記憶媒体は1つ以上のプログラムを格納する。その1つ以上のプログラムは命令を含み、その命令は一つ以上のマイクロフォン、スピーカ、及び一つ以上のプロセッサを有する電子装置によって実行される時、電子装置に上述の方法の操作を実行させる。

**【0008】**

上記に記載された様々な実施形態のより良い理解のために、添付の図面と併せて、以下の実施についての説明を参照する必要がある。図面全体を通して、同様の参照番号は対応する部分を指す。

**【図面の簡単な説明】**

**【0009】**

【図1】いくつかの実施形態に従う例示的操縦環境を示す。

【図2】いくつかの実施形態に従う例示的音声起動式電子装置を示す。

【図3A】いくつかの実施形態に従う例示的音声アシスタンスサーバシステムを示す。

【図3B】いくつかの実施形態に従う例示的音声アシstantサーバシステムを示す。

【図4A】いくつかの実施形態に従うフォーカスセッションの一例を示す。

【図4B】いくつかの実施形態に従うフォーカスセッションの一例を示す。

【図4C】いくつかの実施形態に従うフォーカスセッションの一例を示す。

【図4D】いくつかの実施形態に従うフォーカスセッションの一例を示す。

【図5】いくつかの実施形態に従う、フォーカスセッションを確立し、フォーカスセッションに従って音声入力に応答する例示的プロセスのフロー図を示す。

【図6A】いくつかの実施形態に従う、音声起動式電子装置の正面図である。

【図6B】いくつかの実施形態に従う、音声起動式電子装置の背面図である。

【図6C】いくつかの実施形態に従う、開いた構成における電子装置190の基部に含まれるスピーカを示す、音声起動式電子装置190の斜視図である。

【図6D】いくつかの実施形態に従う、音声起動式電子装置の側面図であり、その中に含まれる電子部品を示す図である。

【図6E】図6E(1)～(4)は、いくつかの実施形態に従う、音声起動式電子装置の

タッチセンスアレイ上で検出された4つのタッチイベントのうちの一つを示し、図6E(5)は、いくつかの実施形態に従う、音声起動式電子装置の背面にあるボタンのユーザの押しを示す図である。

【図6F】いくつかの実施形態に従う、音声起動式電子装置の上面図である。

【図6G】いくつかの実施形態に従う、音声処理状態を示すためのフルカラーレディオLEDのアレイによって表示される、例示的視覚的パターンを示す図である。

**【発明を実施するための形態】**

**【0010】**

図面全体を通して、同じ参照番号は、その対応する部分を指すものとする。

**実施形態の説明**

デジタル革命は今まで、情報のオープンな共有化からグローバルな一体感まで、多くの恩恵をもたらしてきた一方で、新しい技術は、消費者の間で混乱、疑い、恐れをしばしば生じさせ、その結果、消費者がその技術から恩恵を受けることを妨げる。電子装置は、ユーザからの音声入力を受信して音声動作を開始する機能を持つ音声インターフェースとして便利に使用される。それにより、電子装置は、アイズフリー且つハンズフリーの解決策を提供して、既存技術と新技術の両方に近づく。具体的には、電子装置で受信された音声入力は、ユーザの視線が不明瞭で、手が塞がっていたとしても、指示や情報を伝えることができる。ハンズフリー且つアイズフリーの体験を可能にするために、音声起動式電子装置は、常時又はトリガされたときにのみ、周囲を聴取する（すなわち、周囲から収集された音声信号を常時処理する）。一方、ユーザのアイデンティティは、ユーザの声と使用する言語とに関連付けられている。ユーザのアイデンティティを保護するために、これらの音声起動式電子装置は、通常、非公開の場所で使用され、その場所は保護、管理され、且つ親密な空間（例えば、家庭及び車）である。

**【0011】**

いくつかの実施形態に従って、音声命令においてターゲット装置の指示が存在しないか、又は曖昧である場合、音声起動式電子装置は、ターゲット装置を決定するか、又はターゲット装置に音声命令において行われた要求を割り当てる。音声起動式電子装置は、音声命令において明示的に指定又は指示されたターゲット装置に関して、フォーカスセッションを確立する。音声起動式電子装置が、ターゲット装置の指定又は指示が、存在しないか又は曖昧である後続の音声命令を受信した場合、音声起動式電子装置は、音声命令が一つ以上の基準を満たす場合、フォーカスセッションのターゲット装置に、音声命令を割り当てる。

**【0012】**

いくつかの実施形態では、ユーザが別の装置を制御するために音声インターフェース装置と対話するときに、音声インターフェース装置は、どの装置がユーザによって（例えば、フォーカスセッションにおいて）ターゲットにされていたかを記憶する。それ以降の期間、制御のためのデフォルトのターゲット装置は、その記憶された装置である。例えば、例を挙げると、ユーザが最初に「キッチンのライトを点けて」という音声命令を発し、そして次に「ライトを消して」と発した場合、第1の命令の直後に第2の命令が受信されれば、第2の音声命令のためのターゲット装置はデフォルトで「キッチンのライト」になる。別の例として、第1の命令が「リビングルームスピーカで音楽を再生して」であり、後続の命令が「音楽を止めて」である場合、第1の命令の直後に第2の命令が受信されると、第2の音声命令のためのターゲット装置はデフォルトで「リビングルームスピーカ」になる。

**【0013】**

警告9 さらに、いくつかの実施形態では、音声入力間でより長い時間間隔が存在する場合、ユーザは、最後に使用されたターゲット装置が意図されたターゲット装置であることを、確認又は検証するように求められ得る。例えば、第1の音声命令が「リビングルームスピーカで音楽を再生して」であり、第1の音声命令からより長い時間間隔を経て受信された後続の命令が「音楽を停止する」である場合、音声インターフェース装置はユーザに「リビングルームスピーカの音楽を止めたいですか？」と尋ねて、ターゲット装置が「リビ

10

20

30

40

50

「シングルームスピーカ」であることを確認してもよい。

**【0014】**

このようにして、ユーザはありとあらゆる音声入力において、自分の要求の完全なコンテキストを明細に言う必要があるという負担を、免れることができる（例えば、実行する操作を要求するありとあらゆる音声入力に、ターゲット装置の指定を含める必要があるという負担から、免れることができる）。

**【0015】**

**音声アシスタントの操作環境**

図1は、いくつかの実施形態に従う例示的操作環境である。操作環境100は、一つ以上の音声起動式電子装置104（例えば、音声起動式電子装置104\_1から104\_N、以下「音声起動式装置（複数可）」と称する）を含む。一つ以上の音声起動式装置104は、一つ以上の場所（例えば、構造物内の複数の空間全体、又は複数の構造物全体（例えば、一つは家の中に、一つはユーザの車の中に）に広がる構造物の部屋又は空間内のすべて）に配置してもよい。

10

**【0016】**

環境100はまた、1つ以上の制御可能電子装置106を含む。（例えば、電子装置106\_1から106\_N、以下「制御可能装置（複数可）」と称する）。制御可能装置106の例は、メディア装置（スマートテレビ、スピーカシステム、ワイヤレススピーカ、セットトップボックス、メディアストリーミング装置、キャスト装置）、及びスマートホーム装置（例えば、スマートカメラ、スマートサーモスタット、スマートライト、スマート危険探知機、スマートドアロック）を含む。

20

**【0017】**

音声起動式装置104及び制御可能装置106は、通信ネットワーク110を通して音声アシスタントサービス140に（例えば、音声アシスタントサービス140の音声アシスタンスサーバシステム112に）通信可能に結合される。いくつかの実施形態では、音声起動式装置104及び制御可能装置106のうちの一つ以上は、ローカルネットワーク108に通信可能に結合され、そのローカルネットワーク108は、通信ネットワーク110に通信可能に結合される；音声起動式装置（複数可）104及び／又は制御可能装置（複数可）106は、ローカルネットワーク108を介して、通信ネットワーク（複数可）110に（及び通信ネットワーク110を介して、音声アシスタンスサーバシステム112に）通信可能に結合される。いくつかの実施形態では、ローカルネットワーク108は、ネットワークインターフェース（例えばルータ）にて実装されたローカルエリアネットワークである。ローカルネットワーク108に通信可能に結合される、音声起動式装置104及び制御可能装置106もまた、ローカルネットワーク108を通して、互いに通信する場合がある。

30

**【0018】**

任意に、音声起動式装置104のうちの一つ以上は、通信ネットワーク110と通信可能に結合され、ローカルネットワーク108上にはない。例えば、これらの音声起動式装置は、ローカルネットワーク108に対応するWi-Fi（登録商標）ネットワーク上にはないが、セルラー接続を介して通信ネットワーク110に接続されている。いくつかの実施形態では、ローカルネットワーク108上にある音声起動式装置104と、ローカルネットワーク108上にない音声起動式装置104との間の通信は、音声アシスタンスサーバシステム112を通して行われる。音声起動式装置104（ローカルネットワーク108上にあろうと、ネットワーク110上にあろうと）は、音声アシスタントサービス140の装置レジストリ118に登録されているので、音声アシスタンスサーバシステム112に知られている。同様に、ローカルネットワーク108上にない音声起動式装置104は、音声アシスタントサーバシステム112を通して制御可能装置106と通信することができる。制御可能装置106もまた、（ローカルネットワーク108上にあろうと、ネットワーク110上にあろうと）装置レジストリ118に登録されている。いくつかの実施形態では、音声起動式装置104と制御可能装置106との間の通信は、音声アシス

40

50

タンスサーバシステム 112 を経由する。

【0019】

いくつかの実施形態では、環境 100 は一つ以上のコンテンツホスト 114 も含む。コンテンツホスト 114 は、そこからコンテンツが、ユーザの音声入力又は命令に含まれる要求に従って、ストリーミングされる、又はそうでなければ取得されるリモートコンテンツソースとなることがある。コンテンツホスト 114 は、音声アシスタンスサーバシステム 112 がユーザの音声要求に従って情報を検索する情報源であり得る。

【0020】

いくつかの実施形態では、制御可能装置 106 は、（例えば、音声起動式装置 104 及び / 又は音声アシスタンスサーバシステム 112 から）指定された操作を実行する、又は指定された状態に遷移する、命令又は要求を受信することができ、受信した命令又は要求に従って、操作を実行する又は状態を遷移することができる。

10

【0021】

いくつかの実施形態では、一つ以上の制御可能装置 106 は、メディアコンテンツ、ニュース、及び / 又は他の情報を一つ以上のユーザに提供するために操作環境 100 において配備されるメディア装置である。いくつかの実施形態では、メディア装置によって提供されるコンテンツは、ローカルコンテンツソースに記憶されるか、リモートコンテンツソース（例えば、コンテンツホスト（複数可）114）からストリーミングされるか、又はローカルに（例えば、ローカルテキストから、カスタマイズされたニュースの簡潔な説明、電子メール、文章、地域の天気予報などを、操作環境 100 を利用している人の 1 人以上に読み上げる音声プロセッサまで）生成される。いくつかの実施形態では、メディア装置は、メディアコンテンツを視聴者（例えば 1 人以上のユーザ）に直接出力するメディア出力装置と、メディアコンテンツをメディア出力装置にストリーミングするようにネットワーク接続されたキャスト装置とを含む。メディア出力装置の例は、テレビ（TV）表示装置及び音楽プレーヤを含むが、これらに限定されない。キャスト装置の例は、セットトップボックス（STB）、DVD プレーヤ、TV ボックス、及びグーグルの Chrome cast（登録商標）メディアストリーミング装置などのメディアストリーミング装置を含むが、これらに限定されない。

20

【0022】

いくつかの実施形態では、制御可能装置 106 は音声起動式装置 104 である。いくつかの実施形態では、音声起動式装置 104 は制御可能装置 106 である。例えば、制御可能装置 106 は、音声アシスタンスサービス 140（例えば、ユーザの音声入力を受信、処理、応答することもできるメディア装置）への音声インターフェースを含むことがある。別の例として、音声起動式装置 104 はまた、音声入力における要求又は命令に従って、特定の操作を実行し、特定の状態に移行することがある（例えば、ストリーミング音楽を再生することもできる音声インターフェース装置）。

30

【0023】

いくつかの実施形態では、音声起動式装置 104 及び制御可能装置 106 は、それぞれのアカウントを有するユーザ、又はユーザドメインにおいてそれぞれのユーザアカウントを持つ複数のユーザ（例えば、家族又は組織におけるユーザなど、関連ユーザのグループ；より一般的にはプライマリユーザ及び 1 人以上の承認された追加のユーザなど）に関連付けられる。ユーザは、音声起動式装置 104 に音声入力又は音声命令を入力することができる。音声起動式装置 104 はユーザ（例えば、ユーザ 102）からこれらの音声入力を受信し、音声起動式装置 104 及び / 又は音声アシスタンスサーバシステム 112 は、音声入力における要求を決定するように、及び要求に対する応答を生成するように進む。

40

【0024】

いくつかの実施形態では、音声入力において含まれる要求は、制御可能装置 106 に操作を実行させる（例えば、メディアを再生させる、メディアを一時停止させる、メディアを早送り又は巻き戻しさせる、音量を変更させる、画面の明るさを変更させる、光の明るさを変更させる）、又は別の状態へ移行させる（例えば、操作モードを変更させる、オン

50

又はオフにさせる、スリープモードへ移行させる、又はスリープモードから復帰させる)、命令又は要求である。

#### 【0025】

いくつかの実施形態では、音声起動式電子装置104は、音声命令に対する音声応答を生成すること及び提供すること(例えば、「何時ですか?」という質問に応答して現在の時刻を話すこと);ユーザに(例えば、「Bach Boysの曲を再生して」と)要求されたメディアコンテンツをストリーミングすること;ユーザのために準備された、ニュース記事又は毎日のニュースの概要を読み上げること;個人的なアシスタント装置又はローカルネットワーク上に記憶されたメディアアイテムを再生すること;状態を変更すること又は操作環境100内に接続された一つ以上の他の装置を操作すること(例えば、電気、電化製品又はメディア装置のオン/オフ、錠をかけること/開けること、窓の開放など);又はネットワーク110を介してサーバに対応する要求を発行することによって、音声入力に応答する。

10

#### 【0026】

いくつかの実施形態では、一つ以上の音声起動式装置104は、様々な機能(例えば、メディア装置のメディア再生機能)を開始するための音声入力を収集するために操作環境100において配備される。いくつかの実施形態では、これらの音声起動式装置104(例えば、装置104-1から104-N)は、例えば、キャスト装置及びメディア出力装置と同じ部屋の制御可能装置104(例えば、メディア装置)の近くに配備される。あるいは、いくつかの実施形態では、音声起動式装置104は、一つ以上のスマートホーム装置を有するがどのメディア装置も持たない構造内に配備される。あるいは、いくつかの実施形態では、音声起動式装置104は、一つ以上のスマートホーム装置及び一つ以上のメディア装置を有する構造内に配備される。あるいは、いくつかの実施形態では、音声起動式装置104は、ネットワークに繋がった電子装置を持たない場所に配備される。さらには、いくつかの実施形態では、構造における部屋又は空間には、複数の音声起動式装置104があることがある。

20

#### 【0027】

いくつかの実施形態では、音声起動式装置104は、少なくとも一つ以上のマイクロフォン、スピーカ、プロセッサ、及びプロセッサによる実行のための少なくとも1つのプログラムを格納するメモリを含む。スピーカは、音声起動式装置104が音声メッセージ及び他の音声(例えば可聴トーン)を、音声起動式装置104が操作環境100において位置する場所に伝達することを可能にするように構成され、それによって音楽を放送したり、音声入力処理の状態を報告したり、音声入力装置104のユーザと会話したり、又は音声入力装置104のユーザに指示を与えるたりする。音声メッセージに代わるものとして、音声入力処理の状態について音声起動式装置104のユーザにフィードバックを提供するために、視覚的信号を使用することもできる。音声起動式装置104がモバイル装置(例えば、携帯電話又はタブレットコンピュータ)であるとき、その表示画面は、音声入力処理の状態に関する通知を表示するように構成される。

30

#### 【0028】

いくつかの実施形態では、音声起動式装置104は、音声アシスタンスサーバシステム112を用いて音声認識機能を提供するようにネットワーク接続されている音声インタフェース装置である。例えば、音声起動式装置104は、音楽をユーザに提供し、音声アシスタンツサービス(例えば、グーグルアシstant)へのアイズフリー及びハンズフリーのアクセスを可能にするスマートスピーカを含む。任意に、音声起動式装置104は、デスクトップ又はラップトップコンピュータ、タブレット、マイクロフォンを含む携帯電話、マイクロフォン及び任意にスピーカを含むキャスト装置、マイクロフォン及びスピーカを含むオーディオシステム(例えば、ステレオシステム、スピーカシステム、持ち運び可能なスピーカなど)、マイクロフォン及びスピーカを含むテレビジョン、及びマイクロフォン、及びスピーカ、及び任意にディスプレイを含む自動車のユーザインタフェースシステムのうちの1つである。任意に、音声起動式装置104は単純で低コストの音声インタ

40

50

フェース装置である。一般に、音声起動式装置 104 は、ネットワーク接続が可能であり、マイクロフォン、スピーカ、並びに音声アシスタントサービスと対話するための、プログラム、モジュール、及びデータを含む任意の装置であり得る。音声起動式装置 104 の単純性と低コスト性を考えると、音声起動式装置 104 は全表示画面というよりも発光ダイオード (LEDs) のアレイを含み、音声入力処理の状態を示すために LED 上に視覚的パターンを表示する。いくつかの実施形態では、LED はフルカラー LED であり、LED の色は LED 上に表示される視覚的パターンの一部として採用されてもよい。例えば、情報又は装置の状態（例えば、フォーカスセッションが開始されたかどうかがアクティブであることを示していることに関連付けられている状態が延長されている、及び / 又は複数のユーザのうちのどの個々のユーザが特定のフォーカスセッションに関連しているか）を伝えるために視覚的パターンを表示するための LED を使用することの複数の例は、図 6 を参照して以下に説明される。いくつかの実施形態では、音声処理操作の状態を示す視覚的パターンは、音声処理操作を実行している音声起動式装置に関連する従来のディスプレイ上に示される特徴的な画像を使用して表示される。

#### 【0029】

いくつかの実施形態では、LED 又は他の視覚的なディスプレイは、参加している複数の電子装置の集合的な音声処理状態を伝えるのに使用される。例えば、複数の音声処理又は音声インターフェース装置（例えば、図 6 A に示されるような複数の電子装置 104；図 1 の複数の音声起動式装置 104）がある操作環境では、それぞれの電子装置に関連するカラー LED（例えば、図 6 に示されるような LED 604）のグループは、どの電子装置がユーザを聞いているのか、そして聞いている装置のうちどれがリーダ（"リーダ" の装置は、ユーザによって発行された音声要求に応えることにおいて一般的に先導的な役割を担う。）であるのかを伝えるために、使用することができる。

#### 【0030】

より一般的には、図 6 を参照した以下の議論では、ホットワード検出状態、聞き取り状態、思考モード、作業モード、応答モード、及び / 又は話し中モードなどの、電子装置の様々な音声処理状態を、LED の集まりを使用して視覚的に示すための「LED 設計言語」について説明する。いくつかの実施形態では、本明細書で説明されている音声処理操作の一意の状態は、「LED 設計言語」の一つ以上の態様に従って、LED のグループを使用して表される。これらの視覚的インジケータはまた、音声処理操作を実行している電子装置によって生成された一つ以上の可聴インジケータと組み合わせることができる。結果として生じる音声及び / 又は視覚的インジケータは、音声対話環境内のユーザが環境内の様々な音声処理電子装置の状態を理解し、自然で直感的な方法でそれらの装置と効果的に対話することを可能にする。

#### 【0031】

いくつかの実施形態では、音声起動式装置 104 への音声入力がキャスト装置を介してメディア出力装置を制御するために使用されるとき、音声起動式装置 104 はキャスト対応メディア装置の新しいレベルの制御を効果的に可能にする。具体的な例では、音声起動式装置 104 は、遠距離の音声アクセス機能を有するカジュアルエンジョイメントスピーカ（casual enjoyment speaker）を含み、音声アシスタントサービスのための音声インターフェース装置として機能する。音声起動式装置 104 は、操作環境 100 における任意の領域に配備することができる。複数の音声起動式装置 104 が複数の部屋に分散されている場合、それらは同期されてこれらの部屋からの音声入力を与えるキャスト音声受信機になる。

#### 【0032】

具体的には、いくつかの実施形態では、音声起動式装置 104 は、音声起動音声アシスタントサービス（例えば、グーグルアシスタント）に接続されているマイクロフォンを有する Wi-Fi スピーカを含む。ユーザは、音声起動式装置 104 のマイクロフォンを介してメディア再生要求を発行し、音声起動式装置 104 自体又は他の接続されたメディア出力装置上で、音声アシスタントサービスにメディアコンテンツを再生するように依頼す

ることができる。例えば、ユーザはWi-Fiスピーカに「オーケー、グーグル、私のリビングルームTVでネコの動画を再生して」と言って、メディア再生要求を発行することができる。次いで、音声アシスタントサービスは、デフォルトの又は指定されたメディアアプリケーションを使用して、要求された装置上で要求されたメディアコンテンツを再生することによって、メディア再生要求を満たす。

#### 【0033】

いくつかの実施形態では、ユーザは、音声起動式装置104のマイクロフォンを介して、すでに再生されているか又は表示装置上で再生中のメディアコンテンツに関して、音声要求を発行することができる（例えば、ユーザはメディアコンテンツに関する情報を求める、オンラインストアでメディアコンテンツを購入する、又はメディアコンテンツに関するソーシャルポストを作成及び発行することが可能となる）。

10

#### 【0034】

いくつかの実施形態では、ユーザは、それらが家中を移動するときにそれらと現在のメディアセッションを利用したいと思うことがある、音声起動式装置104の一つ以上から、そのようなサービスを要求することができる。これは、音声アシスタントサービス140が現在のメディアセッションを、第1のキャスト装置から、第1のキャスト装置に直接接続されていないか又は第1のキャスト装置の存在を知らない第2のキャスト装置に転送することを要求する。メディアコンテンツの転送に続いて、第2のキャスト装置に結合されている第2の出力装置は、第1のキャスト装置に結合されている第1の出力装置において前に再生されていたメディアコンテンツを、第1の出力装置でメディアコンテンツの再生がなされていた音楽トラック又はビデオクリップ内の正確な時点から、再生を継続する。いくつかの実施形態では、メディアセッションを転送するための要求を受信する音声起動式装置104は、その要求を満たすことができる。いくつかの実施形態では、メディアセッションを転送する要求を受信した音声起動式装置104は、その要求を処理のための、別の装置又はシステム（例えば、音声アシスタンスサーバシステム112）に中継する。

20

#### 【0035】

さらに、いくつかの実施形態では、ユーザは、音声起動式装置104のマイクロフォンを介して、情報に対する要求、又は動作若しくは操作の実行に対する要求を発行することができる。要求される情報は、個人的（例えば、ユーザの電子メール、ユーザのカレンダーアイベント、ユーザの飛行便情報など）、非個人的（例えば、スポーツのスコア、ニュース記事など）、又はそれらの間のどこか（例えば、ユーザが好むチーム又はスポーツのスコア、ユーザが好む情報源からのニュース記事など）であり得る。要求された情報、又は動作／操作は、個人情報（例えば、ユーザによって提供される支払い情報を有するデジタルメディアアイテムの購入、物理的商品の購入）へのアクセスを含み得る。音声起動式装置104は、ユーザへの音声メッセージ応答で要求に応答し、応答は、例えば、要求を満たすための追加情報の要求、要求が満たされたことの確認、要求を満たすことができないことの通知等を含み得る。

30

#### 【0036】

いくつかの実施形態では、音声起動式装置104、及び制御可能装置106の中のメディア装置に加えて、操作環境100はまた、制御可能装置106の中に一つ以上のスマートホーム装置を含み得る。一体化されたスマートホーム装置は、スマートホームネットワークにおいて互いに、及び／又は中央サーバ若しくはクラウドコンピューティングシステムとシームレスに統合して様々な有用なスマートホーム機能を提供する、インテリジェントで多センサ式でネットワーク接続された装置を含む。いくつかの実施形態では、スマートホーム装置は、キャスト装置及び／又は出力装置と同じ操作環境100の場所に配備され、したがってキャスト装置及び出力装置と近接して、又は既知の距離を置いて配置される。

40

#### 【0037】

操作環境100におけるスマートホーム装置は、インテリジェントで多センサ式でネット

50

トワーク接続された一つ以上のサーモスタッフ、インテリジェントでネットワーク接続され、多センサ式の一つ以上のハザード検出器、インテリジェントで多センサ式でネットワーク接続された一つ以上のインターフェース装置及び（以下、「スマートドアベル」と及び「スマートドアロック」と称する）、インテリジェントで多センサ式でネットワーク接続された一つ以上の警報システム、インテリジェントで多センサ式でネットワーク接続された一つ以上のカメラシステム、インテリジェントで多センサ式でネットワーク接続された一つ以上の壁スイッチ、インテリジェントで多センサ式でネットワーク接続された一つ以上の電源ソケット、及びインテリジェントで多センサ式でネットワーク接続された一つ以上のライトを含み得るが、これらに限定されない。いくつかの実施形態では、図1の操作環境100内のスマートホーム装置は、冷蔵庫、ストーブ、オーブン、テレビ、洗濯機、乾燥機、ライト、ステレオ、インターホンシステム、ガレージドア開閉装置、板の間扇風機、天井扇風機、壁掛けエアコン、プールヒーター、灌漑システム、防犯システム、暖房機器、窓用ACユニット、電動ダクト通気口などの、インテリジェントで多センサ式でネットワーク接続された複数の電化製品（以下、「スマートアプライアンス」と称する）を含み得る。いくつかの実施形態では、これらのスマートホーム装置タイプのうちのいずれか1つは、現住者又はユーザからの音声要求に全体的又は部分的に応答するように、マイクロフォン及び本明細書に記載の一つ以上の音声処理能力を備えることができる。

#### 【0038】

いくつかの実施形態では、制御可能装置104及び音声起動式装置104のそれぞれは、他の制御可能装置106、音声起動式電子装置104、中央サーバ又はクラウドコンピューティングシステム、及び／又はネットワーク接続された他の装置（例えばクライアント装置）と、データ通信及び情報共有することができる。データ通信は、様々な慣用又は標準の無線プロトコル（例えば、IEEE 802.15.4、Wi-Fi、ZigBee（登録商標）、6LoWPAN、スレッド、Z-Wave（登録商標）、ブルートゥーススマート、ISA100.11a、WirelessHART（登録商標）、MiWi（登録商標）等）及び／又は、様々な慣用又は標準の有線プロトコルのいずれか（例えば、イーサネット（登録商標）、ホームプラグなど）、又はこの文書の提出日現在でまだ開発されていない通信プロトコルを含む、任意の他の適切な通信プロトコル、のいずれかを使用して実行され得る。

#### 【0039】

通信ネットワーク（例えば、インターネット）110を介して、制御可能装置106及び音声起動式装置104は、サーバシステム（本明細書では中央サーバシステム及び／又はクラウドコンピューティングシステムとも称される）と通信することができる。任意に、サーバシステムは、制御可能装置、及びユーザに表示されるメディアコンテンツに関連付けられている、製造業者、サポートエンティティ、又はサービスプロバイダに関連付けられてもよい。したがって、サーバシステムは、音声起動式装置104によって収集された音声入力を処理する音声アシスタンスサーバ112、表示されたメディアコンテンツを提供する一つ以上のコンテンツホスト114、任意に、分散型装置端末に基づいて仮想ユーザドメインを作成するクラウドキャストサービスサーバ、及び仮想ユーザ環境における分散型装置端末の記録を保持する装置レジストリ118を含む。分散型装置端末の例は、制御可能装置106、音声作動装置104、及びメディア出力装置を含むが、これらに限定されない。いくつかの実施形態では、これらの分散型装置端末は、仮想ユーザドメインにおけるユーザアカウント（例えば、Googleユーザアカウント）にリンクされている。これらの入力に対する応答の生成を含む、音声起動式装置104によって収集された音声入力の処理は、音声起動式装置104において、音声アシスタンスサーバ112において、別のスマートホーム装置（例えば、ハブ装置又は制御可能装置106）、又は上記の全て若しくはサブセットの組み合わせにおいて、ローカルに実行できることを理解されたい。

#### 【0040】

いくつかの実施形態では、音声起動式装置（複数可）104は、スマートホーム装置が

10

20

30

40

50

無い環境においても機能することが理解されよう。例えば、音声起動式装置 104 は、スマートホーム装置が無くても、情報若しくは動作の実行に対するユーザの要求に応答、及び／又は様々なメディア再生機能を開始若しくは制御することができる。音声起動式装置 104 はまた、限定はしないが、車両、船、ビジネス、又は製造環境を含む広範囲の環境で機能することができる。

#### 【0041】

いくつかの実施形態では、音声起動式装置 104 はホットワード（「ウェイクワード」とも称される）を含む音声入力によって「起動」される（例えば、音声起動式装置 104 上の音声アシスタントサービスのためにインターフェースを起動し、音声起動式装置 104 を、音声起動式装置 104 が音声アシスタントサービスへの音声要求を受信する準備ができる状態にする）。いくつかの実施形態では、音声起動式装置 104 が音声入力の受信に関して、少なくとも所定の時間（例えば、5 分間）休止している場合、音声起動式装置 104 は起動することを必要とする；所定の時間は、音声インターフェイスセッション又は会話が時間切れになるまでに許容される休止時間の長さに対応する。ホットワードは、単語又は句であり得、所定のデフォルトであり得、及び／又はユーザによってカスタマイズされ得る（例えば、ユーザは特定の音声起動式装置 104 に対するニックネームを、装置のホットワードとして設定し得る）。いくつかの実施形態では、音声起動式装置 104 を目覚めさせることができる複数のホットワードがあり得る。ユーザはホットワードを話し、音声起動式装置 104 からの承認応答（例えば音声起動式装置 104 が挨拶を出力する）を待つことができ、それらは第 1 の音声要求を行う。あるいは、ユーザは、ホットワードと第 1 の音声要求とを 1 つの音声入力に組み合わせることができる（例えば、音声入力は、音声要求が後に続くホットワードを含む）。

#### 【0042】

いくつかの実施形態では、音声起動式装置 104 は、制御可能装置 106（例えば、メディア装置、スマートホーム装置）、クライアント装置、又は操作環境 100 のサーバシステムといいくつかの実施形態に従って対話する。音声起動式装置 104 は、音声起動式装置 104 に近くの環境から音声入力を受信するように構成される。任意に、音声起動式装置 104 は音声入力を記憶し、音声入力を少なくとも部分的にはローカルに処理する。任意に、音声起動式装置 104 は、受信した音声入力、又は部分的に処理された音声入力を、さらなる処理のために通信ネットワーク 110 を介して音声アシスタンスサーバシステム 112 に伝達する。音声起動式装置 104、又は音声アシスタンスサーバシステム 112 は、音声入力において要求があるかどうか、及びその要求は何かを決定し、その要求に対する応答を決定及び生成、及びその要求を一つ以上の制御可能装置（複数可）106 に伝達する。応答を受信する制御可能装置（複数可）106 は応答に従って、操作を実行するか又は状態を変更するように構成される。例えば、メディア装置は、音声入力における要求に対する応答に従って、メディア装置に結合される出力装置上に表示するために、1 つ以上のコンテンツホスト 114 から、メディアコンテンツ又はインターネットコンテンツを取得するように構成される。

#### 【0043】

いくつかの実施形態では、制御可能装置 106（複数可）及び音声起動式装置（複数可）104 は、ユーザドメインにおいて互いにリンクしており、そしてより具体的には、ユーザドメインにおけるユーザアカウントを介して互いに関連付けられている。制御可能装置 106（ローカルネットワーク 108 上又はネットワーク 110 上かに関わらず）及び音声起動式装置 104（ローカルネットワーク 108 上又はネットワーク 110 上かに関わらず）に関する情報は、ユーザアカウントに関連して装置レジストリ 118 に格納される。いくつかの実施形態では、制御可能装置 106 用の装置レジストリと、音声起動式装置 104 用の装置レジストリがある。制御可能装置レジストリは、ユーザドメインにおいて関連付けられている音声起動式装置レジストリにおける装置を参照することができ、逆もまた同様である。

#### 【0044】

10

20

30

40

50

いくつかの実施形態では、一つ以上の音声起動式装置 104（及び一つ以上のキャスト装置）及び一つ以上の制御可能装置 106 が、クライアント装置 103 を介して音声アシスタントサービス 140 に権限委託（commission）される。いくつかの実施形態では、音声起動式装置 104 は表示画面をまったく含まず、クライアント装置 103 に依存して権限委託過程に、ユーザインタフェースを提供する。そして、制御可能装置 106 についても同様である。具体的には、新しい音声起動式装置 104、及び／又はクライアント装置の近くに配備された制御可能装置 106 の権限委託をユーザインタフェースが容易にすることを可能にするアプリケーションが、クライアント装置 103 にインストールされる。ユーザは、クライアント装置 103 のユーザインタフェース上で要求を送信し、権限委託される必要がある新しい電子装置 104 / 106 に対して権限委託過程を開始することがある。権限委託の要求を受信した後、クライアント装置 103 は、権限委託される必要がある新しい電子装置 104 / 103 との短距離通信リンクを確立する。任意に、短距離通信リンクは、近距離無線通信（NFC）、ブルートゥース（登録商標）、ブルートゥースローエナジー（BLE）、及びその他同種のものに基づいて確立される。それから、クライアント装置 103 は、無線ローカルエリアネットワーク（WLAN）（例えば、ローカルネットワーク 108）に関連する無線設定データを、新たな装置又は電子装置 104 / 106 に伝達する。無線設定データは、少なくとも WLAN セキュリティコード（すなわち、サービスセット識別子（SSID）パスワード）を含み、任意に、SSID、インターネットプロトコル（IP）アドレス、プロキシ設定、及びゲートウェイ設定を含む。近距離通信リンクを介して無線設定データを受信した後、新しい電子装置 104 / 106 は無線設定データをデコードして回復し、無線設定データに基づいて WLAN に参加する。

#### 【0045】

いくつかの実施形態では、追加のユーザドメイン情報は、クライアント装置 103 上に表示されたユーザインタフェース上に入力され、新たな電子装置 104 / 106 をユーザドメインにおけるアカウントにリンクさせるために使用される。任意に、追加のユーザドメイン情報は、短距離通信リンクを介して無線通信データと共に新たな電子装置 104 / 106 に伝達される。任意に、新たな装置が WLAN に参加した後に、追加のユーザドメイン情報は WLAN を介して新たな電子装置 104 / 106 に伝達される。

#### 【0046】

電子装置 104 / 106 が一度ユーザドメインに権限委託されると、他の装置、及びそれらの関連する動作は、複数の制御パスを介して制御され得る。ある制御パスに従って、クライアント装置 103 にインストールされたアプリケーションは、他の装置及びその関連する動作（例えば、メディア再生動作）を制御するために使用される。あるいは、別の制御パスに従って、電子装置 104 / 106 は、他の装置、及びその関連する動作の、アイズフリー及びハンズフリーな制御を可能にするよう使用される。

#### 【0047】

いくつかの実施形態では、音声起動式装置 104 及び制御可能装置 106 は、ユーザによって（例えば、それらの装置がユーザドメインにおいて関連付けられているプライマリユーザによって）ニックネームを割り当てられてもよい。例えば、リビングルームのスピーカ装置は、「リビングルームスピーカ」というニックネームを割り当てられてもよい。このようにして、ユーザは装置ニックネームを話すことにより、音声入力で装置をより容易に参照することができる。いくつかの実施形態では、装置ニックネーム、及び対応する装置へのマッピングは、音声起動式装置 104（音声起動式装置と同じユーザに関連付けられた装置だけのニックネームを記憶する）、及び／又は音声アシスタンスサーバシステム 112 に記憶される（これは、異なるユーザに関連付けられた装置ニックネームを記憶する）。例えば、音声アシスタンスサーバシステム 112 は、異なる装置及びユーザにわたる多数の、装置ニックネーム及びマッピングを記憶し、そして特定のユーザに関連付けられた音声起動式装置 104 は、ローカルな記憶のために特定のユーザに関連付けられた装置に対して、ニックネーム及びマッピングをダウンロードする。

10

20

30

40

50

**【 0 0 4 8 】**

いくつかの実施形態では、ユーザは、音声起動式装置 104 及び / 又は制御可能装置 106 のうちの一つ以上を、ユーザによって作成された装置のグループにグループ分けすることができる。個々の装置をニックネームにより参照するのと同様に、グループには名前が付けられ、そして装置のグループをグループ名で参照されてもよい。装置ニックネームと同様に、装置グループ及びグループ名は音声起動式装置 104 及び / 又は音声アシスタンスサーバシステム 112 に記憶されてもよい。

**【 0 0 4 9 】**

ユーザからの音声入力は、音声入力における要求のための、ターゲット制御可能装置 106、又は装置のターゲットグループを明示的に指定し得る。例えば、ユーザは、「リビングルームスピーカでクラシック音楽を演奏して」という音声入力を発することができる。音声入力におけるターゲット装置は「リビングルームスピーカ」であり；音声入力における要求は「リビングルームスピーカ」にクラシック音楽を演奏させる要求である。別の例として、ユーザは、「ハウススピーカ」が装置のグループの名前である場合に、「ハウススピーカでクラシック音楽を演奏して」という音声入力を発することがある。音声入力におけるターゲット装置のグループは「ハウススピーカ」であり；音声入力における要求は、「ハウススピーカ」グループにおける装置にクラシック音楽を再生させる要求である。

10

**【 0 0 5 0 】**

ユーザからの音声入力には、ターゲット装置、又は装置グループの明示的な指定がない場合がある；名前によるターゲット装置又は装置グループへの参照は、音声入力においては存在しない。例えば、上記の例示的音声入力「リビングルームスピーカでクラシック音楽を演奏して」に続いて、ユーザは後続の音声入力「一時停止」を発することがある。音声入力は、一時停止操作に対する要求のためにターゲット装置の指定を含まない。いくつかの実施形態では、音声入力におけるターゲット装置の指定は、曖昧であり得る。例えば、ユーザが装置名を不完全に発したのかもしれない。いくつかの実施形態では、明示的なターゲット装置の指定が存在しないか、又はターゲット装置の指定が曖昧な場合は、以下で説明されるように、ターゲット装置又は装置グループは、音声入力に割り当てられ得る。

20

**【 0 0 5 1 】**

いくつかの実施形態では、音声起動式装置 104 が、ターゲット装置又は装置グループの明示的な指定を伴う音声入力を受信する際、音声起動式装置 104 は、指定されたターゲット装置又は装置グループに関してフォーカスセッションを確立する。いくつかの実施形態では、音声起動式装置 104 は、フォーカスセッションに対して、セッション開始時間（例えば、フォーカスセッションが開始されたことに基づく音声入力のタイムスタンプ）及び、フォーカスセッションに対してフォーカスされた装置として、指定されたターゲット装置又は装置グループを記憶する。いくつかの実施形態では、音声起動式装置 104 はまた、フォーカスセッションにおける後続の音声入力のログを取る。音声起動式装置 104 は、フォーカスセッションにおいて少なくとも最新の音声入力のログを取り、そして任意に、フォーカスセッション内の先行する音声入力も同様にログを取り保持する。いくつかの実施形態では、音声アシスタンスサーバシステム 112 はフォーカスセッションを確立する。いくつかの実施形態では、フォーカスセッションは、異なるターゲット装置又は装置グループを明示的に指定する音声入力により終了されることがある。

30

**【 0 0 5 2 】**

装置に関するフォーカスセッションがアクティブであり、音声起動式装置が音声入力を受信する間、音声起動式装置 104 は音声入力に関して一つ以上の決定を行う。いくつかの実施形態では、その決定は以下を含む：それらは、音声入力が明示的なターゲット装置の指定を含むかどうか、音声入力における要求がフォーカスされた装置によって満たすことができるものであるかどうか、並びにフォーカスセッションにおける最後の音声入力の時間及び / 又はセッション開始時間と比較した音声入力の時間である。音声入力が明示的なターゲット装置の指定を含まず、フォーカスされた装置によって満たされることができ、並びにフォーカスセッションにおける最後の音声入力の時間及び / 又はセッション開始

40

50

時刻に関して所定の時間基準を満たす場合、その後、フォーカスされた装置が、音声入力に対してターゲット装置として割り当てられる。フォーカスセッションに関する更なる詳細は以下に説明される。

### 【0053】

#### 操作環境における装置

図2は、いくつかの実施形態に従う、操作環境（例えば、操作環境100）においてユーザの音声命令を収集するための音声インターフェースとして適用される例示的音声起動式装置104を示すブロック図である。音声起動式装置104は、典型的には、一つ以上の処理装置（CPU）202、一つ以上のネットワークインターフェース204、メモリ206、及びこれらの構成要素（チップセットと称されることがある）を相互接続するための一つ以上の通信バス208を含む。音声起動式装置104は、ボタン212、タッチセンスアレイ214、及び一つ以上のマイクロフォン216など、ユーザ入力を容易にする一つ以上の入力装置210を含む。音声起動式装置104は、一つ以上のスピーカ220、任意にLED222のアレイ、及び任意にディスプレイ224を含む、一つ以上の出力装置218も含む。いくつかの実施形態では、LED222のアレイはフルカラーLEDのアレイである。いくつかの実施形態では、音声起動式装置104は、装置の種類に応じて、LED222のアレイ、又はディスプレイ224のどちらか一方、又は両方を含む。いくつかの実施形態では、音声起動式装置104は、位置検出装置226（例えばGPSモジュール）及び一つ以上のセンサ228（例えば加速度計、ジャイロスコープ、光センサなど）も含む。10

### 【0054】

メモリ206は、DRAM、SRAM、DDR RAM、又は他のランダムアクセス固体メモリ装置のような高速ランダムアクセスメモリを含み；そして任意に、一つ以上の磁気ディスク記憶装置、一つ以上の光ディスク記憶装置、一つ以上のフラッシュメモリ装置、又は一つ以上の他の不揮発性固体記憶装置などの不揮発性メモリを含む。メモリ206は、任意に、一つ以上の処理装置202から遠隔に配置された一つ以上の記憶装置を含む。メモリ206、あるいはメモリ206内の不揮発性メモリは、非一時的なコンピュータ読み取り可能な記憶媒体を含む。いくつかの実施形態では、メモリ206、又はメモリ206の非一時的なコンピュータ読み取り可能な記憶媒体は、以下のプログラム、モジュール、及びデータ構造、又はそれらのサブセット又はスーパーセットを格納する：20

- ・様々な基本的なシステムサービスを取り扱うための手順、及びハードウェア依存のタスクを実行するための手順を含むオペレーティングシステム232；

- ・音声起動式装置104を他の装置（例えば、音声アシスタンスサービス140、1つ以上の制御可能装置106、1つ以上のクライアント装置103、及び他の音声起動式装置（複数可）104）に、一つ以上のネットワークインターフェース204（有線又は無線）、及びインターネット、他の広域ネットワーク、ローカルエリアネットワーク（例えばローカルネットワーク108）、メトロポリタンエリアネットワークなどの一つ以上のネットワーク110、を介して接続するためのネットワーク通信モジュール234；

- ・一つ以上の入力装置を介して入力を受信し、一つ以上の出力装置218を介して音声起動式装置104にて情報の提示を可能にするための入力／出力制御モジュール236であって、以下を含むもの：

音声起動式装置104を取り囲む環境において収集される、音声入力又は音声メッセージを処理するための、又は収集された音声入力又は音声メッセージを音声アシスタンスサーバシステム112での処理に対して準備するための音声処理モジュール238；40

音声起動式装置104の装置状態に従って、LED222上に視覚的パターンを生成するためのLED制御モジュール240；及び

音声起動式装置104の上面（例えば、タッチセンサアレイ214上）でのタッチイベントを検知するためのタッチセンスモジュール242；

- ・音声起動式装置104に関連するデータを少なくとも記憶するための音声起動式装置データ244であって、以下を含むもの：

10

20

30

40

50

共通の装置設定（サービス層、装置モデル、記憶容量、処理能力、通信能力など）、ユーザドメインにおける1つ以上のユーザアカウントの情報、装置ニックネーム及び装置グループ、未登録のユーザを扱う場合の制限に関する設定、並びにLED222によって表示される1つ以上の視覚的パターンに関連する表示仕様を含む、音声起動式装置104自体に関連する情報を記憶するための音声装置設定246；並びに

音声信号、音声メッセージ、応答メッセージ、及び音声起動式装置104の音声インターフェース機能に関連する他のデータを記憶するための音声制御データ248；

・音声アシスタンスサーバシステム112によって生成された音声要求応答に含まれる命令を実行し、いくつかの実施形態では、一定の音声入力に対する応答を生成するための応答モジュール250；並びに

・装置に関してフォーカスセッションを確立、管理、及び終了するためのフォーカスセッションモジュール252。

#### 【0055】

いくつかの実施形態では、音声処理モジュール238は以下のモジュール（図示せず）を含む：

・音声入力装置104に音声入力するユーザを識別し、曖昧さをなくすためのユーザ識別モジュール；

・音声入力が音声起動式装置104を起動させるためにホットワードを含んでいるかどうかを決定し、音声入力においてそのようなものを認識するためのホットワード認識モジュール；及び

・音声入力において含まれるユーザ要求を決定するための要求認識モジュール。

#### 【0056】

いくつかの実施形態では、メモリ206は、以下を含む、未解決のフォーカスセッションのためのフォーカスセッションデータ254も格納する：

・未解決のフォーカスセッションにおいてフォーカスされた装置又は装置グループの識別子（例えば、装置ニックネーム、装置グループ名、装置のMACアドレス（複数可）を記憶するための、セッションがフォーカスされた装置（複数可）256；

・未解決のフォーカスセッションの開始のためのタイムスタンプを記憶するためのセッション開始時間258；及び

・少なくとも最新の要求／命令を含む、フォーカスセッションにおける以前の要求又は命令のログを記憶するためのセッション命令履歴260。ログは、ログに記録された以前の要求（複数可）／命令（複数可）のタイムスタンプ（複数可）を少なくとも含む。

#### 【0057】

上記の識別された要素のそれぞれは、前述のメモリ装置のうちの一つ以上において格納されてもよく、上記の機能を実行するための一組の指令に対応する。上記で識別されたモジュール又はプログラム（すなわち指令のセット）は、別々のソフトウェアプログラム、手順、モジュール又はデータ構造として実施される必要はなく、したがってこれらのモジュールの様々なサブセットは、様々な実施において結合、又は別の方法で再配列され得る。いくつかの実施形態では、メモリ206は任意に、上記で識別されたモジュール及びデータ構造のサブセットを格納する。さらに、メモリ206は任意に、上記で説明されていない追加のモジュール及びデータ構造を格納する。いくつかの実施形態では、メモリ206に格納されている、プログラム、モジュール、及び／又はデータのサブセットは、音声アシスタンスサーバシステム112上に記憶され、及び／又は音声アシスタンスサーバシステム112により実行されることができる。

#### 【0058】

いくつかの実施形態では、上述のメモリ206における一つ以上のモジュールは、モジュールの音声処理ライブラリの一部である。音声処理ライブラリは、多種多様な装置にて実施され、且つ埋め込まれることがある。

#### 【0059】

図3A～図3Bは、いくつかの実施形態による、操作環境（例えば、操作環境100）

10

20

30

40

50

の音声アシスタントサービス 140 の例示的音声アシスタンスサーバシステム 112 を示すブロック図である。サーバシステム 112 は、典型的には、一つ以上の処理装置（CPU（複数可））302、一つ以上のネットワークインターフェース 304、メモリ 306、及びこれらの構成要素（チップセットと称されることがある）を相互接続するための一つ以上の通信バス 308 を含む。サーバシステム 112 は、キーボード、マウス、音声命令入力ユニット若しくはマイクロフォン、タッチ画面表示、タッチ感知入力パッド、ジェスチャキャプチャカメラ、又は他の入力ボタン若しくは制御などのユーザ入力を容易にする、一つ以上の入力装置 310 を含むことができる。さらに、サーバシステム 112 は、キーボードを補うために又はキーボードに取って代わるために、マイクロフォン及び音声認識、又はカメラ及びジェスチャ認識を使用してもよい。いくつかの実施形態では、サーバシステム 112 は、例えば電子装置上に印刷されたグラフィックシリーズコードの画像を取り込むための、一つ以上のカメラ、スキャナ、又は光学センサユニットを含む。サーバシステム 112 はまた、一つ以上のスピーカ及び／又は一つ以上の視覚的なディスプレイを含む、ユーザインターフェース及びディスプレイコンテンツの提示を可能にする一つ以上の出力装置 312 を含み得る。

#### 【 0060 】

メモリ 306 は、DRAM、SRAM、DDR RAM、又は他のランダムアクセス固体メモリ装置のような、高速ランダムアクセスメモリを含む；そして、任意に、一つ以上の磁気ディスク記憶装置、一つ以上の光ディスク記憶装置、一つ以上のフラッシュメモリ装置、又は一つ以上の他の不揮発性固体記憶装置などの不揮発性メモリを含む。メモリ 306 は、任意に、一つ以上の処理装置 302 から遠隔に配置された一つ以上の記憶装置を含む。メモリ 306、あるいはメモリ 306 内の不揮発性メモリは、非一時的なコンピュータ読み取り可能な記憶媒体を含む。いくつかの実施形態では、メモリ 306、又はメモリ 306 の非一時的なコンピュータ読み取り可能な記憶媒体は、以下のプログラム、モジュール、及びデータ構造、又はそれらのサブセット又はスーパーセットを格納する：

- ・ 様々な基本的なシステムサービスを取り扱うための手順、及びハードウェア依存のタスクを実行するための手順を含むオペレーティングシステム 316；

- ・ サーバシステム 112 を他の装置（例えば、クライアント装置 103、制御可能装置 106、音声起動式装置 104）に、一つ以上のネットワークインターフェース 304（有線又は無線）、及びインターネット、他の広域ネットワーク、ローカルエリアネットワーク、メトロポリタンエリアネットワークなどの一つ以上のネットワーク 110 を介して接続するためのネットワーク通信モジュール 318；

- ・ 情報（例えば、アプリケーション（複数可）322～328、ウィジェット、ウェブサイト及びそのウェブページ、並びに／又は、ゲーム、オーディオ及び／若しくはビデオのコンテンツ、テキストなどを提示するためのグラフィカルユーザインターフェース）の提示をクライアント装置で可能にするためのユーザインターフェースモジュール 320；

- ・ サーバ側での実行のための命令実行モジュール 321（例えば、クライアント装置 103、制御可能装置 106、音声起動式装置 104、並びにスマートホーム装置を制御し、そのような装置によって取り込まれたデータを検討するための、ゲーム、ソーシャルネットワークアプリケーション、スマートホームアプリケーション、及び／又は他のウェブ若しくは非ウェブベースのアプリケーション）であって、以下を一つ以上含むもの：

キャスト装置（複数可）に関連する、装置提供、装置制御、及びユーザアカウント管理のための、サーバ側の機能を提供するために実行されるキャスト装置アプリケーション 322；

対応するメディアソースに関連する、メディア表示及びユーザアカウント管理のためのサーバ側の機能を提供するために実行される、一つ以上のメディアプレーヤアプリケーション 324；

対応するスマートホーム装置の、装置提供、装置制御、データ処理及びデータ検討のためのサーバ側の機能を提供するために実行される、一つ以上のスマートホーム装置アプリケーション 326；並びに

音声起動式装置 104 から受信した音声メッセージの音声処理を整理するため、又はユーザの音声命令、及びユーザの音声命令（例えば、キャスト装置、又は別の音声起動式装置 104 の指定）のための一つ以上のパラメータを抽出するために音声メッセージを直接処理する音声アシスタンスアプリケーション 328；並びに

- ・メディア表示の自動制御に関するデータ（例えば、自動メディア出力モード、及びフォローアップモードにおいて）、及び他のデータを少なくとも記憶するサーバシステムデータ 330 であって、以下のうちの一つ以上を含むもの：

共通の装置設定（例えば、サービス層、装置モデル、記憶容量、処理能力、通信能力など）、及び自動メディア表示制御のための情報を含む、一つ以上のクライアント装置と関連する情報を記憶するためのクライアント装置設定 332；

アカウントアクセス情報、装置設定のための情報（例えば、サービス層、装置モデル、記憶容量、処理能力、通信能力など）、及び自動メディア表示制御のための情報のうち一つ以上を含む、キャスト装置アプリケーション 322 のユーザアカウントに関連する情報を記憶するためのキャスト装置設定 334；

アカウントアクセス情報、メディアコンテンツタイプのユーザ好み、レビュー履歴データ、及び自動メディア表示制御のための情報のうちの一つ以上を含む、一つ以上のメディアプレーヤアプリケーション 324 のユーザアカウントに関連する情報を記憶するためのメディアプレーヤアプリケーション設定 336；

アカウントアクセス情報、1つ以上のスマートホーム装置のための情報（例えば、サービス層、装置モデル、記憶容量、処理能力、通信能力など）のうちの1つ以上を含む、スマートホームアプリケーション 326 のユーザアカウントに関連する情報を記憶するためのスマートホーム装置設定 338；

一つ以上のアカウントアクセス情報、一つ以上の音声起動式装置 104 のための情報（例えば、サービス層、装置モデル、記憶容量、処理能力、通信能力など）を含む、音声アシスタンスアプリケーション 328 のユーザアカウントに関連する情報を記憶するための音声アシスタンスデータ 340；

ユーザの定期購読（例えば、音楽ストリーミングサービス定期購読、ビデオストリーミングサービス定期購読、ニュースレター定期購読）、ユーザ装置（例えば、各ユーザ、装置ニックネーム、装置グループに関連する装置レジストリ 118 において登録された装置）、ユーザアカウント（例えばユーザの電子メールアカウント、カレンダーアカウント、金融アカウントなど）、及びその他のユーザデータを含む、ユーザドメインにおけるユーザに関する情報を記憶するためのユーザデータ 342；

例えば、ユーザの音声モデル又は音声指紋、及びユーザの快適な音量レベルのしきい値を含む、ユーザの音声プロファイルをユーザドメインにおいて記憶するためのユーザ音声プロファイル 344；並びに

複数の装置のためのフォーカスセッションデータを記憶するためのフォーカスセッションデータ 346。

#### 【0061】

- ・装置レジストリ 118 を管理するための装置登録モジュール 348；
- ・電子装置 104 を取り囲む環境において収集された音声入力又は音声メッセージを処理するための音声処理モジュール 350；並びに

- ・装置に関してフォーカスセッションを確立、管理、及び終了するためのフォーカスセッションモジュール 352。

#### 【0062】

図 3B を参照して、いくつかの実施形態では、メモリ 306 は、以下を含む、一つ以上の未解決のフォーカスセッション 3462 1 から 3462 M のためのフォーカスセッションデータ 346 も格納する：

- ・フォーカスセッションが確立されている装置の識別子を記憶するためのセッションソース装置 3464；

- ・未解決のフォーカスセッションにおいてフォーカスされた装置又は装置グループの識

10

20

30

40

50

別子（例えば、装置のニックネーム、装置グループ名、装置のMACアドレス（複数可））を記憶するための、セッションがフォーカスされた装置3466（複数可）；

・未解決のフォーカスセッションの開始に対するタイムスタンプを記憶するためのセッション開始時間3468；及び

・最新の要求／命令を少なくとも含む、フォーカスセッションにおける以前の要求又は命令のログを記憶するためのセッション命令履歴3470。

#### 【0063】

いくつかの実施形態では、音声アシスタンスサーバシステム112は音声入力の処理を主に担当し、したがって、図2を参照して上述したメモリ206における一つ以上のプログラム、モジュール、及びデータ構造は、メモリ306においてそれぞれのモジュールに含まれる（例えば、音声処理モジュール238に含まれる、プログラム、モジュール、及びデータ構造は、音声処理モジュール350に含まれる）。音声起動式装置104は、取り込まれた音声入力を処理のための音声アシスタンスサーバシステム112に伝達するか、又は最初に音声入力を前処理し処理のための音声アシスタンスサーバシステム112に前処理された音声入力を伝達する。いくつかの実施形態では、音声アシスタンスサーバシステム112及び音声起動式装置104は、音声入力の処理に関して、いくつかの共有される責任、及びいくつかの分割された責任を有し、図2に示されるプログラム、モジュール、及びデータ構造は、音声アシスタンスサーバシステム112及び音声起動式装置104の両方に含まれ得るか、又は音声アシスタンスサーバシステム112及び音声起動式装置104の中で分割される。図2に示される他のプログラム、モジュール、及びデータ構造、又はそれらの類似物もまた、音声アシスタンスサーバシステム112に含まれてもよい。

10

#### 【0064】

上記要素の各々は、前述のメモリ装置のうちの1つ以上に格納されてもよく、上述の機能を実行するための命令セットに対応する。上記モジュール又はプログラム（つまり、命令セット）は、別々のソフトウェアプログラム、プロシージャ、モジュール、又はデータ構造として実装される必要がないため、これらのモジュールの様々なサブセットが、様々な実施形態において組み合わされる又は並べ替えられてもよい。いくつかの実施形態では、メモリ306は、必要であれば、上記モジュール及びデータ構造のサブセットを格納する。さらに、メモリ306は、必要であれば、上に記載されないさらなるモジュール及びデータ構造を格納する。

20

#### 【0065】

##### フォーカスセッション例

図4A-4Dは、いくつかの実施形態に従うフォーカスセッションの一例を示す。音声起動式装置104（例えば、操作環境100）及び複数の制御可能装置106を備える操作環境では、その環境においてユーザが制御可能装置106のうちの1つをターゲット装置として指定する音声入力を行うと、ターゲット装置がフォーカスされた装置としてフォーカスセッションが確立することがある。

30

#### 【0066】

図4Aは、音声起動式装置404（例えば音声起動式装置104）、並びに操作環境（例えば、操作環境100）における3つの制御可能装置406、408、及び410（例えば、制御可能装置106）を示す。装置は、ユーザ402と同じ空間内（例えば、同じ部屋内）にあってもよく、又はユーザが位置する構造全体に広がっていてもよい。装置406は、「マスターべッドルームスピーカ」とニックネームが付けられたスピーカシステムである。装置408は、「リビングルームTV」とニックネームが付けられたメディア装置である。装置410は、「ゲームルームTV」とニックネームが付けられたメディア装置である。現時点ではフォーカスセッションはなく；フォーカスセッション418は空である。

40

#### 【0067】

ユーザ402は、「ゲームルームTVで猫のビデオを再生して」という音声入力403

50

を発し、音声起動式装置 404 はその音声入力を受信する。音声起動式装置 404 は、音声入力 403 における要求が猫のビデオを再生するための要求であると決定し、ターゲット装置は、音声入力 403 において明示的に指定された「ゲームルーム TV」装置 410 である。フォーカスされた装置が「ゲームルーム TV」装置 410 であるセッション 418 は、図 4B に示すように、音声起動式装置 404 にて確立される。猫のビデオを再生するための命令は（装置 404 又は音声アシスタンスサーバシステム 112 によって）「ゲームルーム TV」装置 410 に送信され、装置 410 は操作 416 を実行する。

#### 【0068】

図 4C を参照して、続いて、「ゲームルーム TV」410 がフォーカスされた状態のセッション 418 がアクティブであり、操作 416 が装置 410 によって実行されている間、ユーザ 402 は別の音声入力「一時停止」420 を発する。音声起動式装置 404 は、音声入力 420 がターゲット装置の指定を含むかどうか、及び音声入力 420 における要求がフォーカスされた装置 410 によって実行されることができるかどうかを決定する。特定の音声入力 420 「一時停止」の場合、音声起動式装置 404 は、音声入力 420 がターゲット装置の指定を含まないこと、及び音声入力における要求（何が再生していくようが「一時停止」）が、フォーカスされた装置によって実行されることを決定する。いくつかの実施形態では、音声入力 420 がターゲット装置の指定を含むかどうかを決定することは、音声入力における装置ニックネームに対する一致を探すこと（例えば、音声入力に関する、会話からテキストへの認識の実行、及び装置ニックネームを探すためのテキストの解析）を含む。いくつかの実施形態では、音声入力における要求がフォーカスされた装置によって実行され得るかどうかを決定することは、音声入力における要求が何であるかを決定すること、及び、セッションにおける最後の命令との整合性（例えば、「音楽の一時停止」要求が「音楽の一時停止」である最新の命令と矛盾している）のために、要求を現在のフォーカスセッション 418 の命令履歴（例えば履歴 260）と比較すること、及び整合性に対する要求とフォーカスされた装置の機能とを比較する（例えば、「音楽の一時停止」要求はスマートライトの機能と矛盾している）ことを含む。

#### 【0069】

いくつかの実施形態では、音声起動式装置 404 はまた、音声入力 420 が一つ以上のフォーカスセッション維持基準を満たすかどうかを決定する。いくつかの実施形態では、フォーカスセッション維持基準は、音声入力 420 のタイムスタンプがアクティブセッションにおける最後の音声入力 403 のタイムスタンプから一定の時間内にあることである（例えば、第 2 の音声入力は、前の第 1 の音声入力の一定時間内に受信される）。いくつかの実施形態では、この基準に対して複数の時間しきい値がある。例えば、第 1 の短い方の時間しきい値（例えば 20 分）、及び第 2 の長い方の時間しきい値（例えば 4 時間）があり得る。音声入力 420 が最後の音声入力 403 の第 1 の短い方のしきい値内で受信され、かつ上記の他の基準が満たされる場合、フォーカスされた装置が音声入力 420 のためにターゲット装置として設定される（そして、いくつかの実施形態において、音声入力 420 を処理のために音声アシスタンスサーバシステム 112 へ伝達するときは同様にこのターゲット装置設定を伝達する）。例えば、音声入力 420 は、ターゲット装置の指定を含まないこと、及び要求「一時停止」が最後の命令「猫のビデオを再生して」と整合性があることが決定される。「音声入力 420 が音声入力 403 の短い方の時間のしきい値内に受信された場合、フォーカスされた装置である「ゲームルーム TV」装置 410 が音声入力 420 のためにターゲット装置として設定され、図 4D に示すように、「ゲームルーム TV」装置 410 にて実行されている操作 416 が、音声入力 420 に従って猫のビデオを一時停止している。

#### 【0070】

音声入力 420 が、最後の音声入力 403 の第 1 の短い方のしきい値の後、且つ第 2 の長い方のしきい値以内に受信され、且つ上記の他の基準が満たされる場合、音声起動式装置 404 は音声プロンプトを出力して、フォーカスされた装置が音声入力 420 にとって所望のターゲット装置であるというユーザからの確認を要求する。音声起動式装置 404

10

20

30

40

50

は、フォーカスされた装置が所望のターゲット装置であるという確認を受け取ると、セッション 418 を維持し、フォーカスされた装置を音声入力 420 のためのターゲット装置として設定する（そして、いくつかの実施形態においては、処理のための音声アシスタンスサーバシステム 112 へ音声入力 420 を伝達するときは同様にこのターゲット装置設定を伝達する）。ユーザがターゲット装置を確認しない場合、音声起動式装置 404 は、ユーザがターゲット装置を指定すること、ユーザが音声入力を再度言うが、ターゲット装置の指定を含めること、及び／又はセッション 418 を終了することを要求してもよい。いくつかの実施形態では、音声入力 420 が、最後の音声入力 403 から第 2 の長い方のしきい値の後に受信された場合、又は上記の他の基準が満たされない場合、セッション 418 は終了される。いくつかの実施形態では、これらの時間しきい値の値はメモリ 206 及び／又はメモリ 306 において格納されている。音声入力間の経過時間が、これらのしきい値と比較される。

#### 【0071】

いくつかの実施形態では、音声入力における明示的に指定されたターゲット装置の欠如、及び音声入力における要求と、最後の音声入力及びフォーカスされた装置の機能との整合性も、フォーカスセッション維持基準と見なされる。

#### 【0072】

##### プロセス例

図 5 は、いくつかの実施形態に従う、ユーザの音声入力に応答する方法 500 を示すフロー図である。いくつかの実施形態では、方法 500 は、一つ以上のマイクロフォン、スピーカ、一つ以上のプロセッサ、及び一つ以上のプロセッサによる実行のための一つ以上のプログラムを格納するメモリを備える第 1 の電子装置（例えば音声起動式装置 104）にて実施される。この第 1 の電子装置は、共通のネットワークサービス（例えば、音声アシスタンスサービス 140）に通信可能に（ネットワーク 110 を介して）結合される、コネクテッド電子装置（例えば、ユーザアカウントに関連付けられた、音声起動式装置 104 及び制御可能装置 106；特定の音声起動式装置 104 に関連付けられた制御可能装置 106 など）のローカルグループのメンバである。

#### 【0073】

第 1 の電子装置は、第 1 の操作の要求を含む第 1 の音声命令を受信する（502）。例えば、音声起動式装置 404 は第 1 の音声入力 403 を受信する。

#### 【0074】

第 1 の電子装置は、コネクテッド電子装置のローカルグループの中から、第 1 の操作のための第 1 のターゲット装置を決定する（504）。音声起動式装置 404 は、（例えば、音声処理モジュール 238 による処理に基づいて）装置 406、408、及び 410 の中から音声入力 403 のためのターゲット装置（又は装置グループ）を決定する。音声起動式装置 404 は、音声入力 403 におけるターゲット装置の指定「ゲームルーム TV」を、「ゲームルーム TV」装置 410 として認識する。

#### 【0075】

第 1 の電子装置は、第 1 のターゲット装置（又は装置グループ）に関してフォーカスセッションを確立する（506）。音声起動式装置 404（例えば、フォーカスセッションモジュール 252）は、「ゲームルーム TV」装置 410 がフォーカスされた装置だとして、フォーカスセッション 418 を確立する。

#### 【0076】

第 1 の電子装置は、共通のネットワークサービスの操作を介して、第 1 の操作を第 1 のターゲット装置（又は装置グループ）によって実行させる（508）。音声起動式装置 404 又は音声アシスタンスサーバシステム 112 は、音声入力 403 において要求された操作を実行するために、音声アシスタンスサービス 140 を介して装置 410 に命令を伝達する。

#### 【0077】

第 1 の電子装置は、第 2 の操作の要求を含む第 2 の音声命令を受信する（510）。音

10

20

30

40

50

声起動式装置 404 は、第 2 の音声入力 420 を受信する。

**【0078】**

第 1 の電子装置は、第 2 の音声命令が第 2 のターゲット装置（又は装置グループ）の明示的な指定を含まないと決定する（512）。音声起動式装置 404 は、音声入力 420 のためのターゲット装置を（例えば、音声処理モジュール 238 による処理に基づいて）決定し、音声入力 420 がターゲット装置の指定を含まないことを認識する。

**【0079】**

第 1 の電子装置は、第 2 の操作が第 1 のターゲット装置（又は装置グループ）によって実行することができると決定する（514）。音声起動式装置 404 は、音声入力 420 において要求された操作がフォーカスされた装置 410 により実行可能であり、音声入力 403 において要求され、フォーカスされた装置 410 によって実行されている最後の操作と整合性があると決定する。10

**【0080】**

第 1 の電子装置は、第 2 の音声命令が一つ以上のフォーカスセッション維持基準を満たすかどうかを決定する（516）。音声起動式装置 404 は、音声入力 420 が音声入力 403 の一定時間内に受信されたかどうかを決定する。

**【0081】**

第 2 の音声命令がフォーカスセッション維持基準を満たすという決定に従って、第 1 の電子装置は、共通のネットワークサービスの操作を介して、第 2 の操作を第 1 のターゲット装置（又は装置グループ）によって実行させる（518）。音声起動式装置 404 は、音声入力 420 が音声入力 403 の第 1 の短い方の時間しきい値内に受信されたと決定し、その判断に従って、音声入力 420 のためのターゲット装置をフォーカスされた装置 410 に設定する。音声起動式装置 404 又は音声アシスタンスサーバシステム 112 は、音声入力 420 において要求された操作を実行するために、音声アシスタンスサービス 140 を介して装置 410 に命令を伝達する。20

**【0082】**

いくつかの実施形態では、コネクテッド電子装置のローカルグループの中から第 1 の操作のための第 1 のターゲット装置を決定することは、第 1 の音声命令から第 1 のターゲット装置の明示的な指定を取得することを含む。音声起動式装置 404 は、音声入力 403 がターゲット装置の明示的な指定を含むかどうかを決定するために音声入力 403 を前処理してもよい。あるいは、音声起動式装置 404 は、音声入力 403 を処理した音声アシスタンスサーバシステム 112 からターゲット装置の明示的な指定を受け取ってもよい。30

**【0083】**

いくつかの実施形態では、コネクテッド電子装置のローカルグループの中から第 1 の操作のための第 1 のターゲット装置を決定することは、第 1 の音声命令が第 1 のターゲット装置の明示的な指定を含まないことを決定すること、コネクテッド電子装置のローカルグループの間で第 2 の電子装置によって第 1 の操作が実行され得ることを決定すること、及び第 2 の電子装置を第 1 のターゲット装置として選択することを含む。第 1 の音声入力がターゲットの明示的な指定を含まないが、第 1 の音声入力内に含まれる要求がグループ内の単一の装置によって実行されることがあるものである（例えば、ビデオに関連した命令、そしてグループにおいてはただ 1 つのビデオ対応装置がある）場合、その単一の装置が最初の音声入力のためにターゲット装置として設定される。さらに、いくつかの実施形態では、音声起動式装置の他にただ 1 つの制御可能装置がある場合、その制御可能装置は音声入力に対してデフォルトのターゲット装置であり、その音声入力はターゲット装置を明示的に指定せず、その音声入力の要求された操作は制御可能装置により実行ができる。40

**【0084】**

いくつかの実施形態では、ユーザの音声入力履歴（例えば、音声アシスタンスサーバシステム 112 によって収集され、メモリ 306 に格納され、音声起動式装置 104 によって収集され、メモリ 206 に格納される）は、特定の音声起動式装置 104 が特定の制御

10

20

30

40

50

可能装置 106 を制御するために頻繁に使用されていることを履歴が示すかどうかを決定するために（例えば、音声アシスタンスサーバシステム 112 又は音声起動式装置 104 により）分析されることがある。履歴がそのような関係を示す場合、特定の制御可能装置は、音声起動式装置への音声入力のためのデフォルトのターゲット装置として設定されてもよい。

#### 【 0 0 8 5 】

いくつかの実施形態では、デフォルトのターゲット装置の指定（例えば、識別子）は、音声起動式装置 104 及び / 又は音声アシスタンスサーバシステム 112 に記憶されている。

#### 【 0 0 8 6 】

いくつかの実施形態では、第 2 の音声命令がフォーカスセッション維持基準を満たすという決定に従って、第 1 のターゲット装置に対してフォーカスセッションを延長する。いくつかの実施形態では、フォーカスセッションは、一定の時間後にタイムアウトする（すなわち終了する）。第 2 の音声入力 420 がフォーカスセッション維持基準を満たす場合、フォーカスセッション 418 は時間的に延長され得る（例えば、タイムアウトタイマーをリセットする）。

#### 【 0 0 8 7 】

いくつかの実施形態では、第 1 のターゲット装置に関してフォーカスセッションを確立することは、第 1 の音声命令のタイムスタンプを記憶すること、及び第 1 のターゲット装置の識別子を記憶することを含む。音声入力 403 を受信した後にフォーカスセッションが確立されると、音声起動式装置 404 は、（例えば、セッション命令履歴 260 における）音声入力 403 の時間、及び（例えば、セッションがフォーカスされた装置 256 における）フォーカスされた装置 410 の識別子を記憶する。

#### 【 0 0 8 8 】

いくつかの実施形態では、フォーカスセッション維持基準は、第 1 の音声命令の受信に対して第 1 の所定の時間間隔内に、又は第 1 の音声命令の受信に対して第 2 の所定の時間間隔で、第 2 の音声命令が第 1 の電子装置によって受信されるという基準を含み、第 2 の所定の時間間隔は第 1 の所定の時間間隔の後に続く；そして、第 2 の音声命令が一つ以上のフォーカスセッション維持基準を満たすかどうかを決定することは、第 2 の音声命令が第 1 の所定の時間間隔内、又は所定の第 2 の時間間隔内のいずれかで受信されるかどうかを決定することを含む。音声起動式装置 404 は、音声入力 420 が音声入力 403 の第 1 の時間しきい値又は第 2 の時間しきい値内に受信されたかどうかを含む、音声入力 420 が一つ以上のフォーカスセッション維持基準を満たすかどうかを決定する。

#### 【 0 0 8 9 】

いくつかの実施形態では、第 2 の音声命令が第 1 の所定の時間間隔内に受信されたという決定に従って、第 1 の電子装置は、第 2 の音声命令のためのターゲット装置として第 1 のターゲット装置を選択する。音声入力 420 が音声入力 403 からの第 1 の短い方の時間しきい値内に受信されると決定された場合、フォーカスされた装置 410 は音声入力 420 に対してターゲット装置に設定される。

#### 【 0 0 9 0 】

いくつかの実施形態では、第 2 の音声命令が第 2 の所定の時間間隔内に受信されたという決定に従って、第 1 の電子装置は、第 2 の音声命令のターゲット装置として第 1 のターゲット装置を確認する要求を出力し；そして、確認するための要求に応答する、第 1 のターゲット装置の肯定的な確認に従って、第 1 のターゲット装置を第 2 の音声命令のためのターゲット装置として選択する。音声入力 420 が、第 1 の短い方の時間しきい値外であるが第 2 の長い方の時間しきい値内で受信されると音声入力 403 から決定された場合、音声起動式装置はユーザにターゲット装置の確認するよう促す（例えば、フォーカスされた装置 410 が意図されたターゲット装置であるかどうかをユーザに尋ねる）。フォーカスされた装置 410 が意図されたターゲット装置であることをユーザが確認した場合、フォーカスされた装置 410 は、音声入力 420 のターゲット装置として設定される。

10

20

30

40

50

**【 0 0 9 1 】**

いくつかの実施形態では、第1の電子装置は、第3の操作の要求、及びコネクテッド電子装置のローカルグループの中の第3のターゲット装置の明示的な指定を含む第3の音声命令を受信し、第1のターゲット装置に関してフォーカスセッションを終了し、第3のターゲット装置に対してフォーカスセッションを確立し、共通のネットワークサービスの操作を介して第3のターゲット装置によって第3の操作を実行させる。音声起動式装置404は、音声入力420の後に、装置410以外のターゲット装置（例えば、装置406又は408）の明示的な指定を含む新しい音声入力を受信することがある。その音声入力の受信に従って、フォーカスされた装置410とのフォーカスセッション418が終了し、フォーカスされた新しいターゲット装置との新しいセッションが確立される。音声起動式装置404又は音声アシスタンスサーバシステム112は、音声アシスタンスサービス140を介して、新しい音声入力において要求された操作を実行するために新しいターゲット装置に命令を伝達する。

**【 0 0 9 2 】**

いくつかの実施形態では、第1のターゲット装置は第1の電子装置である。第1の電子装置は、第4の操作の要求、及びコネクテッド電子装置のローカルグループの中の第4のターゲット装置の明示的な指定を含む第4の音声命令を受信する。ここで、第4のターゲット装置は、コネクテッド電子装置のローカルグループの第3の電子装置のメンバであり、第3の電子装置は第1の電子装置とは異なる；第1の電子装置は、第1のターゲット装置に関してフォーカスセッションを維持し；共通のネットワークサービスの操作を介して、第4のターゲット装置によって第4の操作を実行させる。もし、音声起動式装置404でのアクティブフォーカスセッション418のためのフォーカスされた装置が音声起動式装置404それ自体であり、そして別の装置をターゲットとして指定する新しい音声入力が音声入力420の後に受信される場合、音声起動式装置404又は音声アシスタンスサーバシステム112は、新しい音声入力において要求された操作を実行するように、音声アシスタンスサービス140を介して、別のターゲット装置に命令を伝達するが、フォーカスセッションは音声起動式装置404がフォーカスされた状態で維持される。

**【 0 0 9 3 】**

いくつかの実施形態では、第2の音声命令は、第4の操作が第4のターゲット装置によって実行させられた後に受信され、第1の操作はメディアの再生操作であり、第2の操作はメディアの停止操作である。第1の電子装置は、第5の操作の要求、及びコネクテッド電子装置のローカルグループの中から第5のターゲット装置の明示的な指定を含む第5の音声命令を受信し、そのローカルグループにおいては第5のターゲット装置は第3の電子装置である；第1の電子装置は、第1のターゲット装置に関してフォーカスセッションを終了し；第5のターゲット装置に関してフォーカスセッションを確立し；共通のネットワークサービスの操作を介して、第5のターゲット装置に第5の操作を実行させる。音声起動式装置404でのアクティブフォーカスセッション418のためにフォーカスされた装置が音声起動式装置404自体であり、音声入力403がメディア再生を開始する要求を含んでおり、音声入力403が、音声入力403の結果としてメディア再生を一時停止する要求を含んでおり、異なる装置をターゲットとして指定する新しい音声入力が音声入力420の後に受信される場合、音声起動式装置404又は音声アシスタンスサーバシステム112は、新しい音声入力において要求された操作を実行するために、音声アシスタンスサービス140を介して異なるターゲット装置に命令を伝達する。そして、フォーカスされた音声起動式装置とのフォーカスセッションは終了し、フォーカスされた新しいターゲット装置との新しいフォーカスセッションが確立する。

**【 0 0 9 4 】**

いくつかの実施形態では、第1の電子装置は、所定の操作終了要求を含む第5の音声命令を受信し、第5の音声命令の受信に従って、第1の操作が第1のターゲット装置によって実行されなくし、第1のターゲット装置に関してフォーカスセッションを終了する。音声起動式装置404が所定の終了命令（例えば、「停止」）を受信した場合、音声起動式

10

20

30

40

50

装置 404 又は音声アシスタンスサーバシステム 112 は、音声アシスタンスサービス 140 を介して装置 410 に命令を伝達して、操作 416 を実行することを中止させ、フォーカスセッション 418 が終了される。

#### 【0095】

いくつかの実施形態では、第 1 の操作はメディアの再生操作であり、第 2 の操作はメディア停止操作、メディア巻き戻し操作、メディア早送り操作、音量アップ操作、及び音量ダウン操作のうちの 1 つである。音声入力 403 における要求は、メディアコンテンツ（例えば、ビデオ、音楽）の再生を開始するための要求であり得、音声入力 420 における要求は、再生を制御するための要求（例えば、一時停止、巻き戻し、早送り、音量を上げる／下げる、次のアイテム／トラック、前のアイテム／トラック）であり得る。

10

#### 【0096】

いくつかの実施形態では、第 1 の操作は、複数の装置状態のうちの第 1 の状態への装置状態変更操作であり、第 2 の操作は、複数の装置状態のうちの第 2 の状態への装置状態変更操作である。音声入力 403 における要求は、第 1 の状態へ遷移するための要求（例えば、照明又は装置の電源を入れる、スリープモードへの移行）であり得、音声入力 420 の要求は、第 2 の状態へ遷移するための要求（照明又は装置の電源を切る、スリープモードからの復帰）であり得る。

#### 【0097】

いくつかの実施形態では、第 1 の操作は、マグニチュードスケールにおける第 1 の方向へのマグニチュード変更操作であり、第 2 の操作は、マグニチュードスケールにおける第 1 の方向とは反対の第 2 の方向へのマグニチュード変更操作である。

20

#### 【0098】

音声入力 403 における要求は、一方向にマグニチュードを変える（例えば、明るくする、音量を上げる）要求であり、音声入力 420 にある要求は、反対方向にマグニチュードを変える（例えば、薄暗くする、音量を下げる）要求であり得る。

#### 【0099】

いくつかの実施形態では、第 1 の電子装置は一つ以上の LED のアレイを含む。第 1 の電子装置は、LED のアレイにおいて一つ以上の LED を照明することによってフォーカスセッションの状態を示す。音声起動式装置 404 は、LED アレイ上にパターンを表示することによって、アクティブフォーカスセッション、又は他の状態及びフォーカスセッションに関連する他の情報（例えば、どれだけの時間フォーカスセッションがアクティブであったか、又は最後の音声入力からどれだけ時間が経過したかの表示）があることを示し得る。

30

#### 【0100】

いくつかの実施形態では、識別されたユーザごとにフォーカスセッションを確立することがあり得る。例えば、ユーザがターゲット装置を指定する音声入力をする場合、ユーザは識別され、音声入力で指定されたターゲット装置がフォーカスされた状態で、識別されたユーザに対してフォーカスセッションが確立される。異なるユーザが音声入力をし、異なるターゲット装置を指定すると、異なるターゲット装置がフォーカスされた状態で、異なるユーザが識別され、識別された異なるユーザに関して別のフォーカスセッションが確立される。異なるユーザによって話され、ターゲット装置を指定しない音声入力は、それぞれの識別されたユーザに対応するアクティブフォーカスセッションに基づいて異なるターゲット装置を割り当てられ得る。

40

#### 【0101】

##### フォーカスセッションの追加の例

以下の実施形態は、メディア装置である一つ以上の制御可能装置 106 と同じ部屋にある音声起動式装置 104 のコンテキストにおける実施形態を説明する。以下に説明される実施形態は、他の種類の制御可能装置 106（例えば、スマートホーム装置）及び他の装置の配置設定に適合され得ることを理解されたい。

#### 【0102】

50

いくつかの実施形態では、音声起動式装置上で再生中のメディアがまだ無い場合、音声起動式装置以外の制御可能装置であるフォーカス装置を用いて、フォーカスセッションを開始することができる。いくつかの実施形態では、音声起動式装置で再生されているメディアが一時停止されている場合、音声起動式装置以外の制御可能装置をフォーカス装置としてフォーカスセッションを開始することができる。

#### 【0103】

いくつかの実施形態では、音声起動式装置に（そして、任意に、音声起動式装置と同じ Wi-Fi ネットワーク上で）関連付けられた装置又は装置グループに向けられた明示的なターゲット装置を有する任意の妥当な要求をユーザが発行した場合、フォーカスセッションが開始される。そのような妥当な要求の例としては、「リビングルームスピーカで音楽を再生する」、「ベッドルーム TV で音量を上げる」、「ホームグループの次の曲」、「リビングルームスピーカを一時停止する」などを含む。明示的なターゲット装置は、フォーカスセッションのためのフォーカスされた装置になる。

10

#### 【0104】

いくつかの実施形態では、要求が明らかにビデオに関連付けられた要求であり、関連付けられた制御可能装置の中に単一のビデオ対応装置がある場合、ビデオ対応装置がフォーカスされた装置として、フォーカスセッションは確立され得る。

#### 【0105】

いくつかの実施形態では、音声起動式装置が能動的にメディアを再生している間に、別の装置がターゲット装置として要求が受信されると、フォーカスは音声起動式装置に残るが、音声起動式装置は一度セッションを停止又は一時停止すると、別の装置でメディアを再生又は制御するという任意の新しい要求は、その別の装置にフォーカスを移動する。

20

#### 【0106】

例えば、ユーザは「Lady Gaga を再生して」と要求し、音声起動式装置は Lady Gaga の音楽の再生を開始し、音声起動式装置がフォーカスされた状態でフォーカスセッションを開始する。次いで、ユーザは「一時停止」を要求し、音声起動式装置はレディガガの音楽を一時停止する（そして、例えば 2 時間フォーカスセッションを維持する）。1 時間が経過した後、ユーザは「私の TV で猫の動画を再生して」と要求する。フォーカスは TV へ移動し、TV が猫のビデオの再生を開始する。

#### 【0107】

30

別の例として、ユーザは「Lady Gaga を再生して」と要求し、音声起動式装置は Lady Gaga の音楽の再生を開始し、音声起動式装置がフォーカスされた状態でフォーカスセッションを開始する。次に、ユーザが「自分の TV で猫のビデオを表示して」と要求すると、猫のビデオは TV で表示され始めるが、フォーカスは音声起動式装置に残ったままである。次に、ユーザは「次へ」を要求し、音声起動式装置はその要求に従つてレディガガ音楽における次のトラックに進む。その後、ユーザは「一時停止」を要求し、音声起動式装置での音楽は一時停止される。次に、ユーザが「私の TV での次のスライド」を要求すると、次のスライドが TV で始まり、フォーカスは TV に移される。

#### 【0108】

40

いくつかの実施形態では、妥当な要求は、音楽の開始、ビデオの開始、ニュースの読み上げの開始（ニュース記事の読み上げなど）、ポッドキャストの開始、写真の開始（写真の表示又はスライドショーなど）、及び任意のメディア制御命令（任意の現在のフォーカスセッションを終了する所定の STOP 命令以外）を含む。

#### 【0109】

いくつかの実施形態では、以下のいずれかが発生するときにフォーカスセッションが終了する：

- ・ フォーカスセッションが（音声入力、例えば、異なる装置を明示的に指定する音声入力を介して）異なる装置に転送され、この場合、フォーカスセッションが異なる装置とともに開始される；

- ・ 音声入力又は別の装置からのキャスティングを介して（例えば音声を介して：「レデ

50

イガガを＜音声インターフェース装置のニックネーム＞で再生して」、「ローカルでレディガガを再生して」など；キャスティングを介して：ユーザはクライアント装置上のアプリケーションを介して音声起動式装置にコンテンツをキャストする）、音声起動式装置上でフォーカスセッションが開始又は（一時停止状態から）再開される；

しかしながら、音声起動式装置が、メディアを再生しようとしているグループのメンバー（フォロワ又はリーダ）である場合、（再生中であっても）フォーカスを停止することはない。そのため、フォーカスはグループのリーダに残る（リーダは別の音声起動式装置である可能性がある）；

- ・要求が、フォーカスされた制御可能装置への所定の（すべての関連する文法を含む）「停止」命令であるとき；

- ・タイムアウト関連命令：

タイムアウトは、制御可能装置が、明示的に指定されているか、フォーカスセッションのフォーカスされた装置に基づいて設定されているかに関わらず、制御可能装置に与えられた、最後の要求又は所定の「停止」命令以外の命令から測定することができる；

タイムアウトはさまざまな可能な命令にわたって240分である；そして

- ・ユーザが、一時停止／再生に使用する、音声起動式装置のボタンを押したとき（さらに、一時停止した任意のコンテンツが音声起動式装置でローカルに再開される）。

#### 【0110】

いくつかの実施形態では、音声起動式装置は、ターゲット装置のユーザ確認を要求する。以下のように、ユーザは制御可能装置でメディアを再生したい場合、確認を求められる：

- ・確認要求は、メディア開始に対して（例えば、何も再生されていないところで音楽を開始する）（早送り又は次のトラックなどのメディア制御に対して）トリガされる；
- ・フォーカスセッションがアクティブになると、確認要求がトリガされる；そして
- ・確認要求は、制御可能装置が、明示的に指定されるか、フォーカスセッションのフォーカスされた装置に基づいて設定されるかどうかに関わらず、現在の音声起動式装置から制御可能装置に与えられた、所定の「停止」命令以外の最後の音声命令から一定時間（例：20分）が経過した後でトリガされる。

#### 【0111】

確認のための要求は、例えば次のとおりである：

- ・音声起動式装置は、「私に＜制御可能装置名＞で再生して欲しいですか？」と出力する。

#### 【0112】

ユーザが「はい。」と応答する。すると、要求されたメディアはフォーカスされた制御可能装置上で再生され、フォーカスはその装置上で維持される。

#### 【0113】

ユーザが「いいえ」と応答する。すると、要求されたメディアは音声起動式装置上で再生され、フォーカスセッションは終了される。

#### 【0114】

その他：例えば、ユーザの応答が不明瞭な場合、音声起動式装置が「すみません、あなたの応答を理解できませんでした」と出力することがある。

#### 【0115】

いくつかの実施形態では、フォーカスセッションが開始されるとき、メディア開始及び音声ベースの制御命令は、フォーカスされた制御可能装置に適用される。非メディア要求（例えば、検索、質問）は音声起動式装置によって回答され、非メディア要求はフォーカスセッションを終了しない。

#### 【0116】

いくつかの実施形態では、フォーカスセッションが開始されたときでも、物理的対話が依然として音声起動式装置を制御するので、音量を変更、及び一時停止／再生するための音声起動式装置との物理的対話（例えばボタンを押すこと、タッチ感知領域に触れること）は音声起動式装置に影響し、必ずしも制御可能装置には影響しない。

10

20

30

40

50

**【0117】**

いくつかの実施形態では、音声起動式装置上で再生されているタイマー／アラーム／テキスト読み上げに対して発行された要求又は命令は、フォーカスされた制御可能装置に対する同様の要求又は命令よりも高い優先度を有する。例えば、音声起動式装置がタイマー又はアラームを鳴らしており、ユーザが「停止」と発したとき、音声起動式装置はタイマー又はアラームの鳴動を停止する。その後、ユーザが「音量を<上げる／下げる>」と発した場合、タイマー又はアラーム鳴動は依然として停止され、制御可能装置の音量は変更され、上げられる又は下げられる。

**【0118】**

別の例として、音声起動式装置がテキスト読み上げを再生していて（例えばユーザの電子メールを読み上げていて）、ユーザが「停止」と発した場合、音声起動式装置はテキスト読み上げを停止する。その後、ユーザが「音量を<上げる／下げる>」と発すると、音声起動式装置上の音量は変更され、上げられる又は下げられる。

10

**【0119】**

さらに別の例として、音声起動式装置が休止、一時停止、又はアプリケーションがロードされ、ユーザが「停止」と発した場合、制御可能装置でのメディア再生は停止され、フォーカスセッションは終了される。その後、ユーザが「音量を<上げる／下げる>」と発すると、制御可能装置上の音量は変更され、上げられる又は下げられる。

**【0120】****音声起動式電子装置の物理的特徴**

20

図6A及び図6Bは、いくつかの実施形態に従う、音声起動式電子装置104（図1）の正面図600及び背面図620である。電子装置104は、一つ以上のマイクロフォン602とフルカラーLED604のアレイとを含む。フルカラーLED604は、電子装置104の上面の下に隠され、それらが点灯していないときにはユーザに見えないようにすることができる。いくつかの実施形態では、フルカラーLED604のアレイは、物理的にリング状に配列される。さらに、電子装置104の背面は、電源に結合するように構成された電源コネクタ608を任意に含む。

**【0121】**

いくつかの実施形態では、電子装置104は目に見えるボタンが無いきれいな外観を提示し、電子装置104との対話は、音声及びタッチジェスチャに基づく。あるいは、いくつかの実施形態では、電子装置104は、限られた数の物理的なボタン（例えば、その背面上のボタン606）を含み、電子装置104との対話は、音声及びタッチジェスチャに加えて、さらにボタンを押すことに基づく。

30

**【0122】**

電子装置104においては、1つ以上のスピーカが配備されている。図6Cは、いくつかの実施形態に従う、開いた構成での電子装置104の基部610に収容されたスピーカ622を示す、音声起動式電子装置104の斜視図660である。電子装置104は、フルカラーLED604のアレイ、1つ以上のマイクロフォン602、スピーカ622、デュアルバンドWi-Fi802.11a/c無線、ブルートゥースLE無線、環境光センサ、USBポート、プロセッサ、及びプロセッサによる実行のためのプログラムを少なくとも1つ格納するメモリを含む。

40

**【0123】**

図6Dを参照すると、電子装置104は、電子装置104の上面のタッチイベントを検出するように構成されたタッチセンスアレイ624をさらに含む。タッチセンスアレイ624は、電子装置104の上面の下に配備され隠されてもよい。いくつかの実施形態では、タッチセンスアレイは、ビアホールのアレイを含む回路基板の上面に配列され、フルカラーLED604は、回路基板のビアホール内に配備される。回路基板が電子装置104の上面のすぐ下に配置されると、フルカラーLED604とタッチセンスアレイ624の両方が同様に電子装置104の上面のすぐ下に配備される。

**【0124】**

50

図 6 E ( 1 ) ~ 図 6 E ( 4 ) は、いくつかの実施形態に従う、音声起動式電子装置 104 のタッチセンスアレイ 624 上で検出された 4 つのタッチイベントを示す。図 6 E ( 1 ) 及び 6 E ( 2 ) を参照すると、タッチセンスアレイ 624 は音声起動式電子装置 104 の上面上の回転スワイプを検出する。時計回りのスワイプの検出に応答して、音声起動式電子装置 104 はその音声出力の音量を増加させ、反時計回りのスワイプの検出に応答して音声起動式電子装置 104 はその音声出力の音量を減少させる。図 6 E ( 3 ) を参照すると、タッチセンスアレイ 624 は音声起動式電子装置 104 の上面上のシングルタップタッチを検出する。第 1 のタップタッチの検出に応答して、音声起動式電子装置 104 は、第 1 のメディア制御操作を実行し（例えば、特定のメディアコンテンツを再生し）、第 2 のタップタッチの検出に応答して、音声起動式電子装置 104 は、第 2 のメディア制御操作を実行する（例えば、現在再生されている特定のメディアコンテンツを一時停止する）。図 6 E ( 4 ) を参照すると、タッチセンスアレイ 624 は、音声起動式電子装置 104 の上面上のダブルタップタッチ（例えば、2 回の連続タッチ）を検出する。2 つの連続するタッチは、所定の長さより短い時間で隔てられる。しかしながら、それらが所定の長さより長い時間で隔てられるとき、2 つの連続したタッチは 2 つのシングルタップタッチと見なされる。いくつかの実施形態では、ダブルタップタッチの検出に応答して、音声起動式電子装置 104 は、ホットワード検出状態を開始し、ホットワード検出状態において電子装置 104 は、一つ以上のホットワード（例えば、所定のキーワード）を聞いて認識する。電子装置 104 がホットワードを認識するまで、電子装置 104 は音声入力を音声アシスタンスサーバ 112 又はクラウドキャストサービスサーバ 118 に一切送信しない。いくつかの実施形態では、一つ以上のホットワードの検出に応答してフォーカスセッションが開始される。

#### 【 0125 】

いくつかの実施形態では、フルカラー LED 604 のアレイは、LED 設計言語に従って 1 組の視覚パターンを表示し、音声起動式電子装置 104 の上面上の、時計回りのスワイプ、反時計回りのスワイプ、シングルタップ、又はダブルタップの検出を示すように構成される。例えば、フルカラー LED 604 のアレイは、それぞれ図 6 E ( 1 ) 及び 6 E ( 2 ) に示すように、時計回り又は反時計回りのスワイプを追跡するために順次点灯することができる。電子装置 104 の音声処理状態に関連する視覚パターンに関するさらなる詳細は、図 6 F 及び 6 G ( 1 ) ~ 6 G ( 8 ) を参照して以下に説明される。

#### 【 0126 】

図 6 E ( 5 ) は、いくつかの実施形態に従う、音声起動式電子装置 104 の裏側のボタン 606 へのユーザの例示的なタッチ又は押しを示す。ボタン 606 へのユーザの第 1 のタッチ又は押しに応答して、電子装置 104 のマイクロフォンはミュートされ、ボタン 606 へのユーザの第 2 のタッチ又は押しに応答して、電子装置 104 のマイクロフォンが起動される。

#### 【 0127 】

##### 音声ユーザインターフェースの視覚的快適さのための LED 設計言語

いくつかの実施形態では、電子装置 104 は、全表示画面ではなくフルカラー発光ダイオード（LED）のアレイを含む。フルカラー LED のアレイの照明を構成し、電子装置 104 の異なる音声処理状態を示す異なる視覚的パターンを可能にするために LED 設計言語が採用される。LED 設計言語は、色、パターン、及びフルカラー LED の固定されたセットに適用される特定の動作の文法から構成されている。その言語における要素は、電子装置 104 の使用中に特定の装置状態を視覚的に示すために組み合わされる。いくつかの実施形態では、フルカラー LED の照明は、他の重要な状態の中で、電子装置 104 の受動的な聞き取り状態及び能動的な聞き取り状態を明確に描写することを目的としている。類似の LED 設計言語要素を用いて LED（例えば LED 604）によって視覚的に示すことができる状態は、一つ以上のフォーカスセッションの状態、一つ以上の特定のフォーカスセッションに関連付けられた 1 人以上のユーザのアイデンティティ、及び / 又は一つ以上のアクティブフォーカスセッションの持続期間を含む。例えば、いくつかの実施

10

20

30

40

50

形態では、フォーカスセッションがアクティブであること、第2の音声入力の検出のために延長されたこと、及び／又は電子装置104とのユーザの音声対話の欠如のために最近失効したことを示すために、LED604の異なる光のパターン、色の組み合わせ、及び／又は特定の動きを使用することができる。特定のフォーカスセッションに関連する1人以上のユーザの1つ以上のアイデンティティもまた、視覚的に特定のユーザを識別するLED604の異なる光のパターン、色の組み合わせ、及び／又は特定の動きとともに示すことができる。フルカラーLEDの配置は、電子装置104の物理的制約に準拠し、フルカラーLEDのアレイは、特定の技術（例えば、グーグルアシスタント）に基づいて第三者の相手先ブランド名製造会社（OEM）によって製造されたスピーカにおいて使用することができる。

10

#### 【0128】

音声起動式電子装置104においては、電子装置104がその周囲の環境から収集された音声入力を処理するが、音声入力を記憶したり音声入力を任意のリモートサーバに伝達したりしないときに、受動的な聞き取りが起こる。対照的に、能動的な聞き取りは、電子装置104がその周囲の環境から収集された音声入力を記憶するとき、及び／又は音声入力をリモートサーバと共有するときに起こる。本出願のいくつかの実施形態によれば、電子装置104は、電子装置104のユーザのプライバシーを侵害することなく、その周囲の環境において音声入力を受動的に聞き取るだけである。

#### 【0129】

図6Gは、いくつかの実施形態に従う、音声起動式電子装置104の上面図であり、図6Hは、いくつかの実施形態に従う、音声処理状態を示すためにフルカラーLEDのアレイによって表示される6つの例示的な視覚的パターンを示す。いくつかの実施形態では、電子装置104はいかなる表示画面も含まず、フルカラーLED604は、全表示画面と比較して単純で低成本の視覚的ユーザインターフェースを提供する。フルカラーLEDは、電子装置の上面の下に隠れ、点灯していないときにはユーザに見えなくてもよい。図6G及び図6Hを参照すると、いくつかの実施形態では、フルカラーLED604のアレイは物理的にリング状に配列されている。例えば、図6H(6)に示されるように、フルカラーLED604のアレイは、それぞれ図6F(1)及び6F(2)に示されるように時計回り又は反時計回りのスワイプを追跡するために順次点灯してもよい。

20

#### 【0130】

音声処理状態を視覚的に示すための方法が電子装置104にて実施される。電子装置104は、一つ以上のマイクロフォン602を介して電子装置に近接した環境から音声入力を収集し、音声入力を処理する。処理は、環境内のユーザからの音声入力を識別し、音声入力に応答することのうちの一つ以上を含む。電子装置104は、複数の所定の音声処理状態の中から処理の状態を決定する。フルカラーLED604のそれぞれについて、電子装置104は、決定された音声処理状態に関連するそれぞれの所定のLED照明仕様を識別する。照明仕様は、LED照明持続時間、パルス繰り返し数、デューティサイクル、色順序及び輝度のうちの一つ以上を含む。いくつかの実施形態では、電子装置104は、複数のユーザのうちの1人のアイデンティティに従ってフルカラーLED604の所定のLED照明仕様（例えば色順序）のうちの少なくとも1つをカスタマイズすることにより、音声処理状態（いくつかの実施形態ではフォーカスセッションの状態を含む）が複数のユーザのうちの1人に関連付けられると決定し、フルカラーLED604の所定のLED照明仕様を識別する。

30

#### 【0131】

さらに、いくつかの実施形態では、決定された音声処理状態に従って、フルカラーLEDの色は所定の色の組を含む。例えば、図6G(2)、6G(4)及び6G(7)～(10)を参照すると、所定の色の組は、青、緑、黄及び赤を含むグーグルブランドの色を含み、フルカラーLEDのアレイは4つの象限に分けられ、それぞれがグーグルブランドの色の1つに関連付けられている。

40

#### 【0132】

50

フルカラーＬＥＤの識別されたＬＥＤ照明仕様に従って、電子装置104は、フルカラー－ＬＥＤのアレイの照明を同期させて、決定された音声処理状態（いくつかの実施形態ではフォーカスセッションの状態を含む）を示す視覚的パターンを提供する。いくつかの実施形態では、音声処理状態を示す視覚的パターンは、複数の個別的なＬＥＤ照明画素を含む。いくつかの実施形態では、視覚的パターンは開始セグメント、ループセグメント、及び終了セグメントを含む。ループセグメントは、フルカラーＬＥＤのＬＥＤ照明持続期間に関連する期間にわたって続き、音声処理状態の長さ（例えば、アクティブフォーカスセッションの持続期間）と一致するように構成される。

#### 【0133】

いくつかの実施形態では、電子装置104は、ＬＥＤ設計言語によって表される20を超える異なる装置状態（複数の所定の音声処理状態を含む）を有する。任意に、複数の所定の音声処理状態は、ホットワード検出状態、聞き取り状態、思考状態、及び応答状態のうちの一つ以上を含む。いくつかの実施形態では、上述のように、複数の所定の音声処理状態は一つ以上のフォーカスセッション状態を含む。

10

#### 【0134】

これまで実施形態について詳細に言及されたが、その例は添付の図面に示されている。先述の詳細な説明では、記載されている様々な実施形態の徹底した理解を提供するために、多数の具体的な詳細が説明されてきた。しかしながら、記載された様々な実施形態が、これらの具体的な詳細が無しに実施されてよいことは当業者にとっては明らかであろう。他の例では、その実施形態の局面を必要に曖昧にしないように、よく知られている方法、手順、構成要素、回路、及びネットワークは詳細には説明されていない。

20

#### 【0135】

様々な要素を説明するために、いくつかの例において、第1の、第2の用語などが本明細書において使用され得るが、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とも理解されるだろう。これらの用語は、1つの要素を別の要素と区別するために使用されるにすぎない。例えば、様々な記載された実施形態の範囲から逸脱することなく、第1の装置を第2の装置と称することができ、同様に、第2の装置を第1の装置と称することができる。第1の装置及び第2の装置は、両方のタイプの装置であるが、同じ装置ではない。

#### 【0136】

本明細書中に記載されるさまざまな実施形態の説明で用いる術語は、特定の実現例のみを説明する目的のものであり、限定を意図するものではない。記載されるさまざまな実現例の説明及び添付の請求項で用いるような「a」、「an」、及び「the」という単数形は、文脈が明確にそうでないと示していなければ、複数形も含むことが意図される。本明細書中で用いるような「及び／又は」という用語は、列挙される関連の項目のうち1つ以上のありとあらゆる可能な組合せを参照しつつこれを包含することも理解される。「含む」、「含んでいる」、「備える」、及び／又は「備えている」という用語は、この明細書で用いる場合、述べる特徴、整数、ステップ、操作、要素、及び／又は構成要素の存在を特定するが、1つ以上の他の特徴、整数、ステップ、操作、要素、構成要素、及び／又はその群の存在又は追加を排除するわけではないことがさらに理解される。

30

#### 【0137】

本明細書中で用いるような「ならば（if）」という用語は、任意に、文脈に依存して、「したとき」又は「すると」又は「判定することに応答して」又は「検出することに応答して」又は「という判定に従って」を意味すると解釈される。同様に、「判定されれば」又は「[述べた条件又はイベントが]検出されれば」という句は、任意に、文脈に依存して、「判定すると」又は「判定に応答して」又は「[述べた条件又はイベントを]検出すると」又は「[述べた条件又はイベントを]検出するのに応答して」又は「[述べた条件又はイベントが]検出される」という判定に従って」を意味すると解釈される。

40

#### 【0138】

上述のシステムがユーザに関する情報を収集する状況では、ユーザは、個人的な情報（

50

例えば、ユーザの好み又はスマート装置の使用に関する情報)を収集し得るプログラム又は機能をオプトイン又はオプトアウトする機会を与えられる。さらに、いくつかの実施形態では、あるデータは、それが記憶又は使用される前に一つ以上的方法で匿名化され、その結果、個人的に識別可能な情報が取り除かれる。例えば、個人的に識別可能な情報を決定することができない、又はユーザに関連付けることができないように、且つユーザの好み又はユーザインタラクションが特定のユーザに関連付けられるのではなく一般化される(例えばユーザの統計に基づいて一般化される)ように、ユーザのアイデンティティは匿名化されてもよい。

#### 【0139】

さまざまな図面のうちのいくつかは特定の順序で複数の論理的段階を示すが、順序に依存しない段階を再度順序付けてもよく、他の段階を組合せたり切離したりしてもよい。何らかの再順序付け又は他のグループ分けに具体的に言及がされるが、他のものが当業者には自明であり、そのため、本明細書中に提示される順序付け及びグループ分けは代替策の網羅的な列挙ではない。さらに、ハードウェア、ファームウェア、ソフトウェア、又はその任意の組合せで段階を実現可能であろうことを認識すべきである。

10

#### 【0140】

具体的な実現例を参照して、説明の目的のための以上の説明を記載した。しかしながら、例示的な上記の議論は、網羅的であること又は開示される正確な形態に請求項の範囲を限定することを意図しない。上記教示に鑑みて多数の修正例及び変更例が可能である。実現例は、請求項の基礎をなす原則及びそれらの実践的な適用例を最良に説明し、それにより、当業者が、企図される特定の用途に適するようなさまざまな修正例とともに実現例を最良に用いることができるようするために選択された。

20

30

40

50

## 【図面】

## 【図 1】

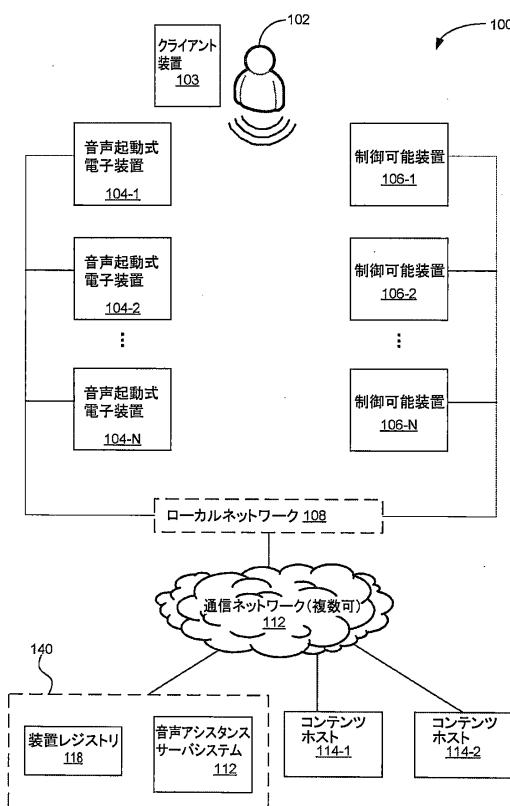


Figure 1

## 【図 2】

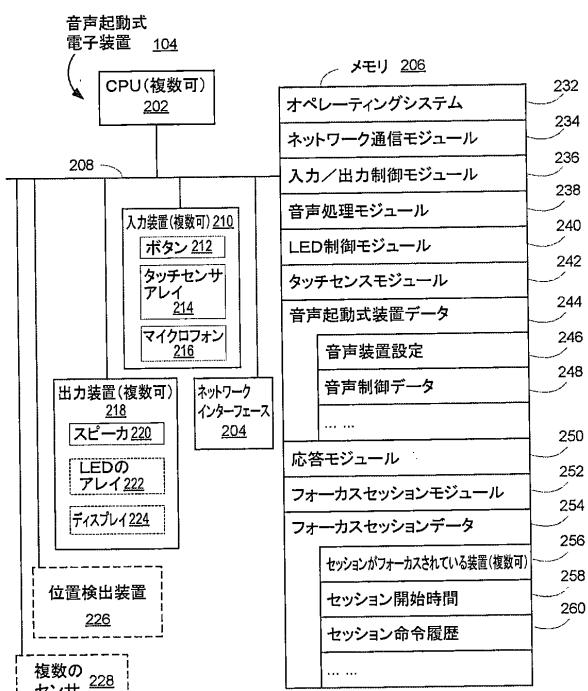


Figure 2

10

20

30

40

## 【図 3 A】

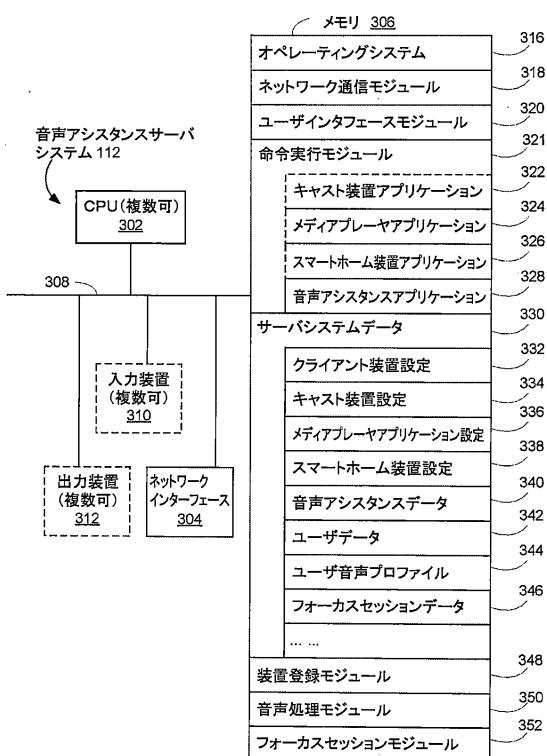


Figure 3A

## 【図 3 B】

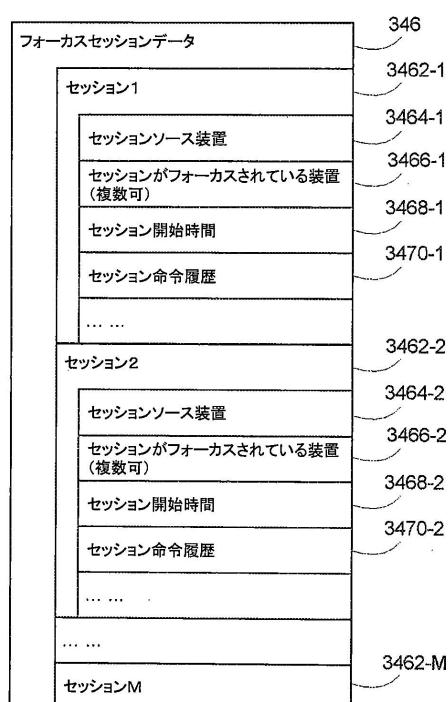


Figure 3B

50

【図 4 A】

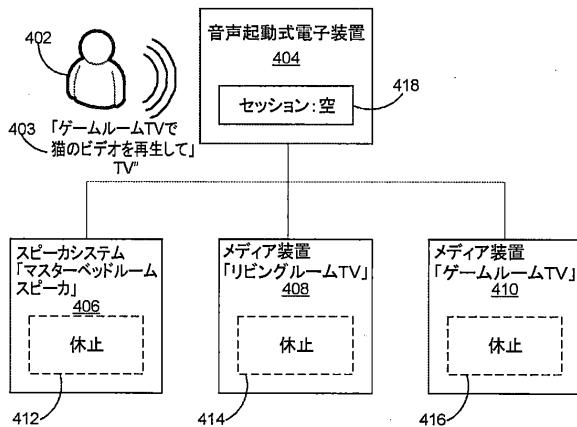


Figure 4A

【図 4 B】

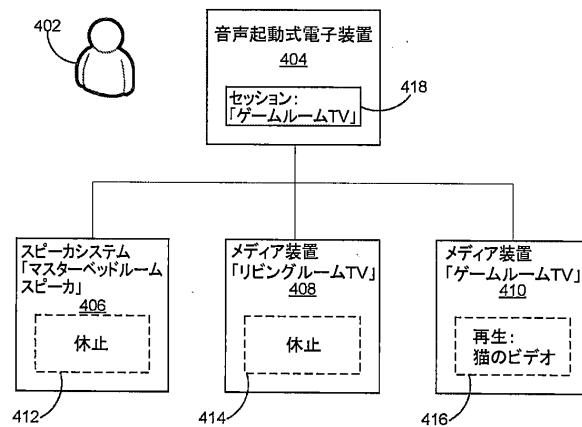


Figure 4B

10

【図 4 C】

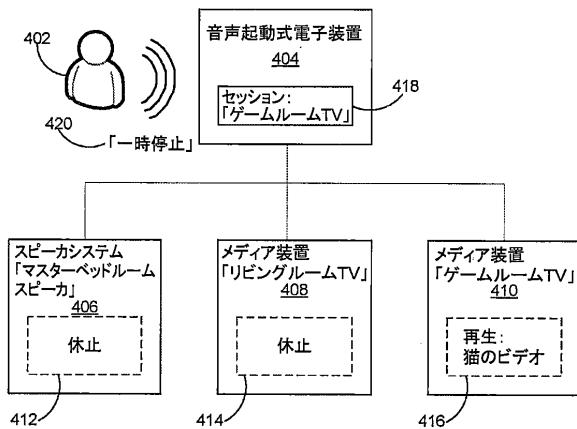


Figure 4C

【図 4 D】

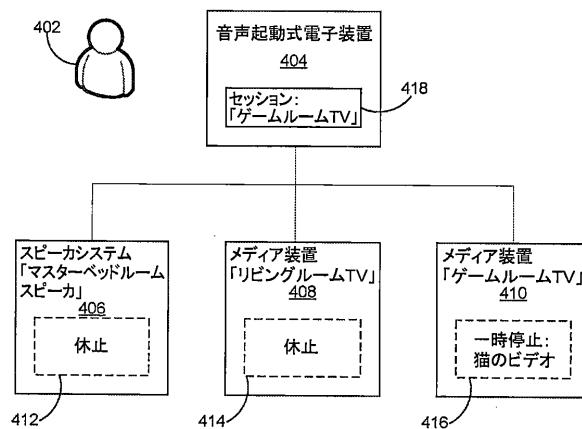


Figure 4D

20

30

40

50

【図 5】

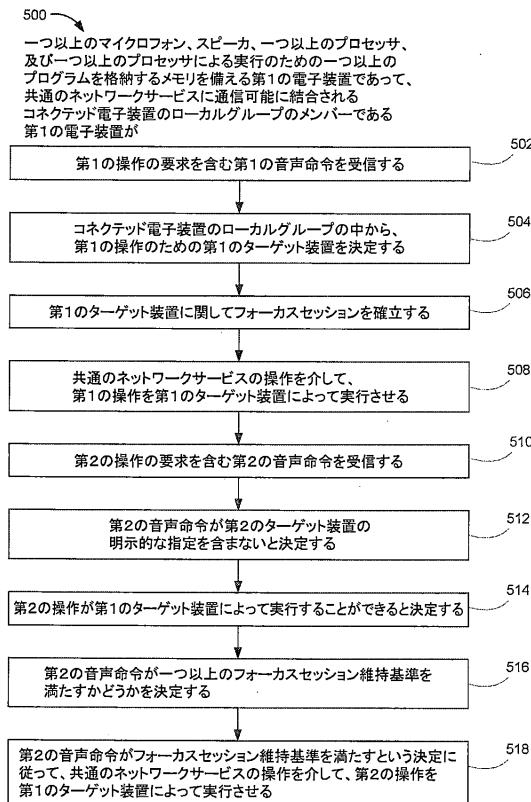


Figure 5

【図 6 A】

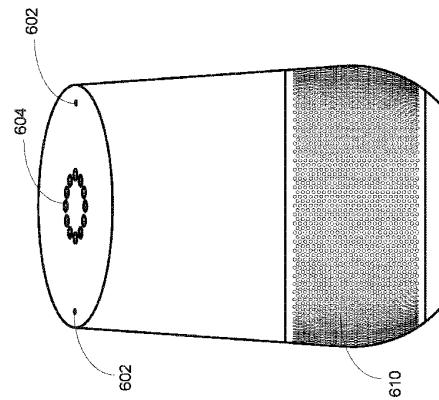


FIGURE 6A

10

【図 6 B】

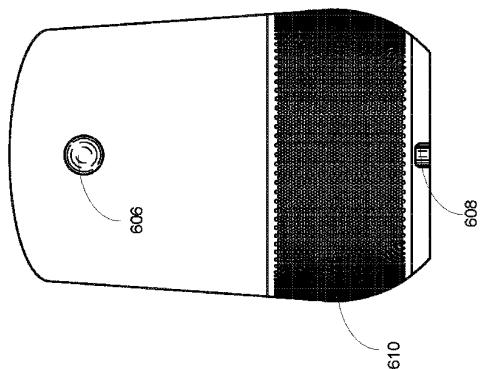
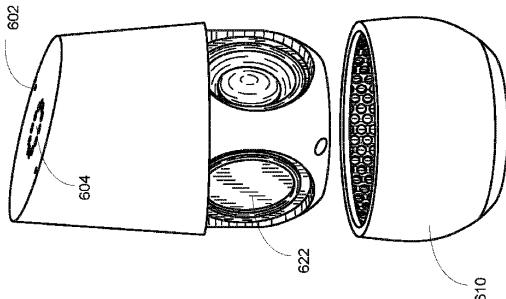


FIGURE 6B

FIGURE 6C

30

【図 6 C】



40

50

【図 6 D】

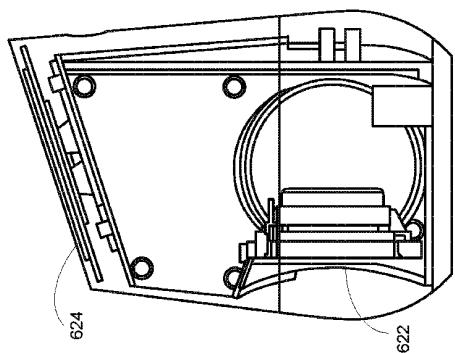
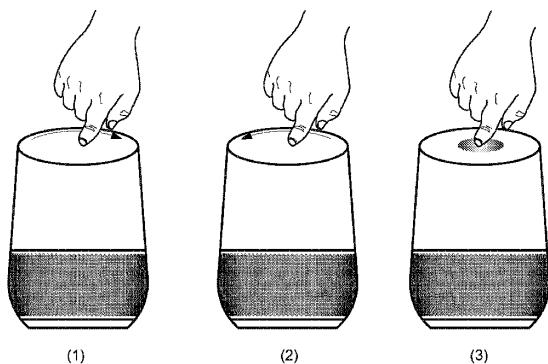
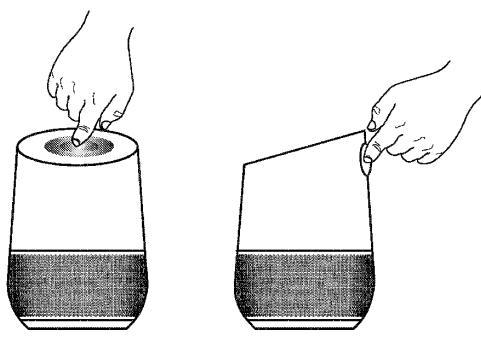


FIGURE 6D

【図 6 E】



10



20

FIGURE 6E

【図 6 F】

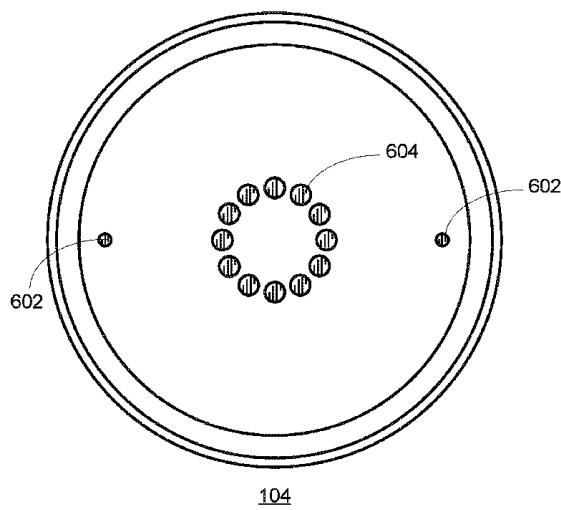


FIGURE 6F

30

【図 6 G】

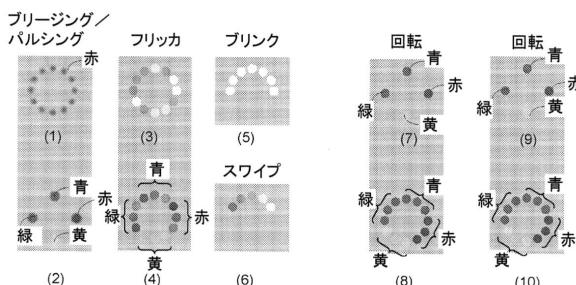


FIGURE 6G

40

50

---

フロントページの続き

(51)国際特許分類

F I		
G 0 6 F	3/16	6 3 0
G 0 6 F	3/16	6 5 0
G 1 0 L	15/00	2 0 0 N

(33)優先権主張国・地域又は機関

米国(US)

アメリカ合衆国、94043 カリフォルニア州、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 ゲエン，トゥアン・アン

アメリカ合衆国、94043 カリフォルニア州、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 大野 弘

(56)参考文献 特開2015-201739(JP,A)

特開2009-044609(JP,A)

世界のビジネスモデル予報 BUSINESS FORECAST THE INTERNET OF THING  
ブラックストーンが注目するヴィヴィントの「未来の家」，フォーブス ジャパン  
，株式会社アトミックスメディア，2016年07月25日，第3巻 第9号，pp.118-119

(58)調査した分野 (Int.Cl., DB名)

G 1 0 L 1 5 / 2 8

G 1 0 L 1 5 / 2 2

G 1 0 L 1 5 / 1 0

G 1 0 L 1 5 / 0 0

G 0 6 F 3 / 1 6