

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7230205号
(P7230205)

(45)発行日 令和5年2月28日(2023.2.28)

(24)登録日 令和5年2月17日(2023.2.17)

(51)国際特許分類 F I
G 1 0 L 15/32 (2013.01) G 1 0 L 15/32 2 0 0 Z
G 1 0 L 15/16 (2006.01) G 1 0 L 15/16

請求項の数 20 (全24頁)

(21)出願番号	特願2021-531511(P2021-531511)	(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 P arkway 94043 Mounta in View, CA U.S.A.
(86)(22)出願日	令和1年11月27日(2019.11.27)	(74)代理人	100142907 弁理士 本田 淳
(65)公表番号	特表2022-509880(P2022-509880 A)	(72)発明者	アレクシッチ、ペーター 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
(43)公表日	令和4年1月24日(2022.1.24)		最終頁に続く
(86)国際出願番号	PCT/US2019/063555		
(87)国際公開番号	WO2020/117586		
(87)国際公開日	令和2年6月11日(2020.6.11)		
審査請求日	令和3年7月15日(2021.7.15)		
(31)優先権主張番号	62/774,507		
(32)優先日	平成30年12月3日(2018.12.3)		
(33)優先権主張国・地域又は機関	米国(US)		

(54)【発明の名称】 音声入力処理

(57)【特許請求の範囲】

【請求項1】

コンピューティングデバイスによって、発話のオーディオデータを受信することと、
前記コンピューティングデバイスによって、(i)ニューラルネットワークまたは(ii)音響モデルおよび言語モデルを使用して、前記発話の複数の候補転写を含むとともに、複数の転写信頼スコアを含み、該複数の転写信頼スコアのそれぞれが、それぞれの候補転写が前記発話と一致する可能性を反映する単語ラティスを生成することと、

前記コンピューティングデバイスによって、前記コンピューティングデバイスの文脈を判定することと、

前記コンピューティングデバイスの前記文脈に基づいて、前記コンピューティングデバイスによって、前記複数の候補転写に対応する複数の文法を特定することと、

現在の前記文脈に基づいて、前記コンピューティングデバイスによって、前記複数の候補転写のそれぞれについて、それぞれの文法がそれぞれの候補転写と一致する可能性を反映する複数の文法信頼スコアを判定することと、

前記複数の文法信頼スコアのうちの最高の文法信頼スコアを、前記複数の文法のいずれも前記複数の候補転写の1つについてパースできない場合に適用されるデフォルト意図文法の文法信頼スコアに一致させるように増加させるための正規化係数を前記コンピューティングデバイスによって判定することと、

前記複数の候補転写のそれぞれについて判定された前記複数の文法信頼スコアを、前記正規化係数を用いて前記コンピューティングデバイスによって正規化することと、

10

20

前記複数の転写信頼スコアおよび正規化された前記複数の文法信頼スコアに基づいて、前記コンピューティングデバイスによって、前記複数の候補転写の中から候補転写を選択することと、

前記コンピューティングデバイスによる出力のために、選択された前記候補転写を前記発話の転写として提供することと、

を含む、コンピュータにより実装される方法。

【請求項 2】

前記複数の文法信頼スコアを正規化することは、

前記複数の文法のうちの 2 つ以上が前記複数の候補転写のうちの 1 つに対応すると判定することと、

前記複数の文法のうちの 2 つ以上が前記複数の候補転写のうちの 1 つに対応すると判定することに基づいて、前記 2 つ以上の文法について前記複数の文法信頼スコアを正規化することと、を含む、

前記コンピューティングデバイスは、前記複数の候補転写の中から、前記転写信頼スコアおよび正規化された前記文法信頼スコアに基づいて、前記候補転写を選択する、

請求項 1 に記載の方法。

【請求項 3】

前記 2 つ以上の文法について前記複数の文法信頼スコアを、前記正規化係数を用いて正規化することは、

前記 2 つ以上の文法のそれぞれについて、前記複数の文法信頼スコアのそれぞれを係数で増加させることを含む、

請求項 2 に記載の方法。

【請求項 4】

前記複数の候補転写のそれぞれについて、前記それぞれの転写信頼スコアと正規化された前記それぞれの文法信頼スコアとの積を判定することを含む、

前記コンピューティングデバイスは、前記複数の候補転写の中から、前記転写信頼スコアと正規化された前記それぞれの文法信頼スコアとの前記積に基づいて、前記候補転写を選択する、

請求項 1 に記載の方法。

【請求項 5】

前記コンピューティングデバイスによって、前記コンピューティングデバイスの前記文脈を判定することは、前記コンピューティングデバイスの位置と、前記コンピューティングデバイスのフォアグラウンドで実行しているアプリケーションと、時刻とに基づいている、

請求項 1 に記載の方法。

【請求項 6】

前記言語モデルは、前記単語ラティスに含まれる語のシーケンスについて、確率を特定するように構成され、

前記音響モデルは、前記オーディオデータの一部に一致する音素を特定するように構成されている、

請求項 1 に記載の方法。

【請求項 7】

前記コンピューティングデバイスによって、選択された前記候補転写と、選択された前記候補転写に一致する文法とに基づくアクションを実行することを含む、

請求項 1 に記載の方法。

【請求項 8】

1 つまたは複数のコンピュータと、

動作可能な命令を記憶する 1 つまたは複数の記憶装置と、を備え、前記命令は、前記 1 つまたは複数のコンピュータによって実行されると、前記 1 つまたは複数のコンピュータに、

10

20

30

40

50

コンピューティングデバイスによって、発話のオーディオデータを受信することと、
前記コンピューティングデバイスによって、(i)ニューラルネットワークまたは(i
i)音響モデルおよび言語モデルを使用して、前記発話の複数の候補転写を含むとともに、
複数の転写信頼スコアを含み、該複数の転写信頼スコアのそれぞれが、それぞれの候補
転写が前記発話と一致する可能性を反映する単語ラティスを生成することと、

前記コンピューティングデバイスによって、前記コンピューティングデバイスの文脈を
判定することと、

前記コンピューティングデバイスの前記文脈に基づいて、前記コンピューティングデバ
イスによって、前記複数の候補転写に対応する複数の文法を特定することと、

現在の前記文脈に基づいて、前記コンピューティングデバイスによって、前記複数の候
補転写のそれぞれについて、それぞれの文法がそれぞれの候補転写と一致する可能性を反
映する複数の文法信頼スコアを判定することと、

前記複数の文法信頼スコアのうちの最高の文法信頼スコアを、前記複数の文法のいづれ
も前記複数の候補転写の1つについてパースできない場合に適用されるデフォルト意図文
法の文法信頼スコアに一致させるように増加させるための正規化係数を前記コンピュー
ティングデバイスによって判定することと、

前記複数の候補転写のそれぞれについて判定された前記複数の文法信頼スコアを、前記
正規化係数を用いて前記コンピューティングデバイスによって正規化することと、

前記複数の転写信頼スコアおよび正規化された前記複数の文法信頼スコアに基づいて、
前記コンピューティングデバイスによって、前記複数の候補転写の中から候補転写を選択
することと、

前記コンピューティングデバイスによる出力のために、選択された前記候補転写を前記
発話の転写として提供することと、を含む動作を実行させる、

システム。

【請求項9】

前記複数の文法信頼スコアを正規化することは、

前記複数の文法のうちの2つ以上が前記複数の候補転写のうちの1つに対応すると判定
することと、

前記複数の文法のうちの2つ以上が前記複数の候補転写のうちの1つに対応すると判定
することに基づいて、前記2つ以上の文法について前記複数の文法信頼スコアを正規化す
ることと、を含み、

前記コンピューティングデバイスは、前記複数の候補転写の中から、前記転写信頼スコ
アおよび正規化された前記文法信頼スコアに基づいて、前記候補転写を選択する、

請求項8に記載のシステム。

【請求項10】

前記2つ以上の文法について前記複数の文法信頼スコアを、前記正規化係数を用いて正
規化することは、

前記2つ以上の文法のそれぞれについて、前記複数の文法信頼スコアのそれぞれを係数
で増加させることを含む、

請求項9に記載のシステム。

【請求項11】

前記動作は、

前記複数の候補転写のそれぞれについて、前記それぞれの転写信頼スコアと正規化され
た前記それぞれの文法信頼スコアとの積を判定することを含み、

前記コンピューティングデバイスは、前記複数の候補転写の中から、前記転写信頼スコ
アと正規化された前記それぞれの文法信頼スコアとの前記積に基づいて、前記候補転写を
選択する、

請求項8に記載のシステム。

【請求項12】

前記コンピューティングデバイスによって、前記コンピューティングデバイスの前記文

10

20

30

40

50

脈を判定することは、前記コンピューティングデバイスの位置と、前記コンピューティングデバイスのフォアグラウンドで実行しているアプリケーションと、時刻とに基づいている、

請求項 8 に記載のシステム。

【請求項 1 3】

前記動作は、

前記言語モデルは、前記単語ラティスに含まれる語のシーケンスについて、確率を特定するように構成されていることと、

前記音響モデルは、前記オーディオデータの一部に一致する音素を特定するように構成されていることと、を含む、

請求項 8 に記載のシステム。

【請求項 1 4】

前記動作は、

前記コンピューティングデバイスによって、選択された前記候補転写と、選択された前記候補転写に一致する文法とに基づくアクションを実行することを含む、

請求項 8 に記載のシステム。

【請求項 1 5】

1 つまたは複数のコンピュータによって実行可能な命令を含むソフトウェアを記憶する非一時的コンピュータ可読媒体であって、前記命令は、実行時に、前記 1 つまたは複数のコンピュータに

コンピューティングデバイスによって、発話のオーディオデータを受信することと、

前記コンピューティングデバイスによって、(i) ニューラルネットワークまたは (i i) 音響モデルおよび言語モデルを使用して、前記発話の複数の候補転写を含むとともに、複数の転写信頼スコアを含み、該複数の転写信頼スコアのそれぞれが、それぞれの候補転写が前記発話と一致する可能性を反映する単語ラティスを生成することと、

前記コンピューティングデバイスによって、前記コンピューティングデバイスの文脈を判定することと、

前記コンピューティングデバイスの前記文脈に基づいて、前記コンピューティングデバイスによって、前記複数の候補転写に対応する複数の文法を特定することと、

現在の前記文脈に基づいて、前記コンピューティングデバイスによって、前記複数の候補転写のそれぞれについて、それぞれの文法がそれぞれの候補転写と一致する可能性を反映する複数の文法信頼スコアを判定することと、

前記複数の文法信頼スコアのうちの最高の文法信頼スコアを、前記複数の文法のいずれも前記複数の候補転写の 1 つについてパースできない場合に適用されるデフォルト意図文法の文法信頼スコアに一致させるように増加させるための正規化係数を前記コンピューティングデバイスによって判定することと、

前記複数の候補転写のそれぞれについて判定された前記複数の文法信頼スコアを、前記正規化係数を用いて前記コンピューティングデバイスによって正規化することと、

前記複数の転写信頼スコアおよび正規化された前記複数の文法信頼スコアに基づいて、前記コンピューティングデバイスによって、前記複数の候補転写の中から候補転写を選択することと、

前記コンピューティングデバイスによる出力のために、選択された前記候補転写を前記発話の転写として提供することと、を含む動作を実行させる、

非一時的コンピュータ可読媒体。

【請求項 1 6】

前記複数の文法信頼スコアを正規化することは、

前記複数の文法のうちの 2 つ以上が前記複数の候補転写のうちの 1 つに対応すると判定することと、

前記複数の文法のうちの 2 つ以上が前記複数の候補転写のうちの 1 つに対応すると判定することに基づいて、前記 2 つ以上の文法について前記複数の文法信頼スコアを正規化す

10

20

30

40

50

ることと、を含み、

前記コンピューティングデバイスは、前記複数の候補転写の中から、前記転写信頼スコアおよび正規化された前記文法信頼スコアに基づいて、前記候補転写を選択する、

請求項 15 に記載の媒体。

【請求項 17】

前記動作は、

前記複数の候補転写のそれぞれについて、前記それぞれの転写信頼スコアと正規化された前記それぞれの文法信頼スコアとの積を判定することを含み、

前記コンピューティングデバイスは、前記複数の候補転写の中から、前記転写信頼スコアと正規化された前記それぞれの文法信頼スコアとの前記積に基づいて、前記候補転写を選択する、

10

請求項 15 に記載の媒体。

【請求項 18】

前記コンピューティングデバイスによって、前記コンピューティングデバイスの前記文脈を判定することは、前記コンピューティングデバイスの位置と、前記コンピューティングデバイスのフォアグラウンドで実行しているアプリケーションと、時刻とに基づいている、

請求項 15 に記載の媒体。

【請求項 19】

前記動作は、

前記言語モデルは、前記単語ラティスに含まれる語のシーケンスについて、確率を特定するように構成されていることと、

前記音響モデルは、前記オーディオデータの一部に一致する音素を特定するように構成されていることと、を含む、

請求項 15 に記載の媒体。

20

【請求項 20】

前記動作は、

前記コンピューティングデバイスによって、選択された前記候補転写と、選択された前記候補転写に一致する文法とに基づくアクションを実行することを含む、

請求項 15 に記載の媒体。

30

【発明の詳細な説明】

【技術分野】

【0001】

本明細書は、一般に音声認識に関する。

【背景技術】

【0002】

音声入力を使用してコンピュータとの対話を実行できるようにすることがますます求められている。これには、入力処理、特に自然言語データを処理および分析するようにコンピュータをプログラムする方法の開発が必要である。このような処理には、コンピュータによる話された言語の認識およびテキストへの変換を可能にする計算言語学の分野である音声認識が伴う場合がある。

40

【発明の概要】

【発明が解決しようとする課題】

【0003】

本明細書は、一般に音声認識に関する。

【課題を解決するための手段】

【0004】

ユーザが音声を通じてコンピューティングデバイスに入力を提供できるようにするため、音声入力処理システムは、文脈を使用して、自動音声認識部によって生成された複数の候補転写に適用する複数の文法を識別できる。各文法は、話者の異なる意図、あるいはシ

50

システムが同じ候補転写に対して実行するアクションを示している可能性がある。システムは、候補転写をパースする文法と、文法がユーザの意図と一致する可能性と、候補転写がユーザの発言と一致する可能性とに基づいて、文法および候補転写を選択し得る。次に、システムは、選択された候補転写に含まれる詳細を使用して、文法に対応するアクションを実行することができる。

【 0 0 0 5 】

より詳細には、音声処理システムは、ユーザからの発話を受け取り、単語ラティスを生成する。単語ラティスは、発話における可能性の高い単語と、各単語の信頼スコアとを反映するデータ構造である。システムは、単語ラティスから、複数の候補転写と各候補転写の転写信頼スコアとを特定する。システムは、ユーザの特性、システムの位置、システム
10
の特性、システム上で実行しているアプリケーション（例えば、現在アクティブなアプリケーションまたはフォアグラウンドで実行しているアプリケーション）、またはその他の類似の文脈データに基づいて、現在の文脈を特定する。文脈に基づいて、システムは、各候補転写をパースする複数の文法の文法信頼スコアを生成する。複数の文法が同じ候補転写に適用される可能性がある場合、システムは複数の文法信頼スコアのうちの一部を調整する場合がある。システムは、調整された複数の文法信頼スコアと複数の転写信頼スコアとの組み合わせに基づいて、文法と候補転写を選択する。

【 0 0 0 6 】

本出願に記載される主題の革新的な態様によれば、音声入力処理方法は、コンピューティングデバイスによって、発話のオーディオデータを受信することと、コンピューティング
20
デバイスによって、音響モデルおよび言語モデルを使用して、発話の複数の候補転写を含むとともに、複数の転写信頼スコアを含み、該複数の転写信頼スコアのそれぞれが、それぞれの候補転写が発話と一致する可能性を反映する単語ラティスを生成することと、コンピューティングデバイスによって、コンピューティングデバイスの文脈を判定することと、コンピューティングデバイスの文脈に基づいて、コンピューティングデバイスによって、複数の候補転写に対応する複数の文法を特定することと、現在の文脈に基づいて、コンピューティングデバイスによって、複数の候補転写のそれぞれについて、それぞれの文法がそれぞれの候補転写と一致する可能性を反映する複数の文法信頼スコアを判定することと、複数の転写信頼スコアおよび複数の文法信頼スコアに基づいて、コンピューティング
30
デバイスによって、複数の候補転写の中から候補転写を選択することと、コンピューティングデバイスによる出力のために、選択された候補転写を発話の転写として提供することと、に係るアクションを含む。

【 0 0 0 7 】

これらおよびその他の実装には、それぞれ任意選択により次のフィーチャの1つ以上を含めることができる。アクションは、複数の文法のうちの2つ以上が複数の候補転写のうちの1つに対応すると判定することと、複数の文法のうちの2つ以上が複数の候補転写のうちの1つに対応すると判定することに基づいて、2つ以上の文法について複数の文法信頼スコアを調整することと、を含む。コンピューティングデバイスは、複数の候補転写の中から、転写信頼スコアおよび調整された文法信頼スコアに基づいて、候補転写を選択する。2つ以上の文法について複数の文法信頼スコアを調整するアクションは、2つ以上の
40
文法のそれぞれについて、複数の文法信頼スコアのそれぞれを係数で増加させることを含む。アクションは、複数の候補転写のそれぞれについて、それぞれの転写信頼スコアとそれぞれの文法信頼スコアとの積を判定することを含む。コンピューティングデバイスは、複数の候補転写の中から、転写信頼スコアとそれぞれの文法信頼スコアとの積に基づいて、候補転写を選択する。コンピューティングデバイスによって、コンピューティングデバイスの文脈を判定するアクションは、コンピューティングデバイスの位置と、コンピューティングデバイスのフォアグラウンドで実行しているアプリケーションと、時刻とに基づいている。言語モデルは、単語ラティスに含まれる語のシーケンスについて、確率を特定するように構成されている。音響モデルは、オーディオデータの一部に一致する音素を特定するように構成されている。アクションは、コンピューティングデバイスによって、選
50

択された候補転写と、選択された候補転写に一致する文法とに基づくアクションを実行することを含む。

【0008】

この態様の他の実施形態は、対応するシステム、装置、およびコンピュータ記憶装置に記録されたコンピュータプログラムを含み、それぞれが方法の動作を実行するように構成されている。

【0009】

本明細書に記載されている主題の特定の実施形態は、以下の利点のうちの1つまたは複数を実現するように実施することができる。音声認識システムは、受信した音声入力と判定された文脈の両方を使用して、受信した音声入力をさらに処理してコンピューティングデバイスにアクションを実行させるために使用される文法を選択することができる。このように、音声認識システムは、限られた数の文法を複数の候補転写に適用することにより、ヒューマンマシンインタフェースの待ち時間を短縮することができる。音声認識システムは、システムが予期しない入力を受信したときに音声認識システムが転写を出力できるように、言語のすべてまたはほぼすべての単語を含む語彙を使用することができる。

10

【0010】

本明細書に記載主題の1つまたは複数の実施形態の詳細は、添付の図面および以下の説明に記載されている。主題の他の特徴、態様、および利点は、説明、図面、および特許請求の範囲から明らかとなる。

【図面の簡単な説明】

20

【0011】

【図1】文脈に基づいて発話に適用する文法を選択する例示的システムを示す図。

【図2】文脈に基づいて発話に適用する文法を選択する例示的システムの構成要素を示す図。

【図3】コンテキストに基づいて文法を選択する例示的プロセスを示すフローチャート。

【図4】コンピューティングデバイスおよびモバイルコンピューティングデバイスの例を示す図。

【発明を実施するための形態】

【0012】

さまざまな図面での同様の参照番号と指示は、同様の要素を示す。

30

図1は、文脈に基づいて発話に適用する文法を選択する例示的システム100を示す。以下で簡潔に、および、より詳細に説明するように、ユーザ102は発話104を行う。コンピューティングデバイス106は、発話104を検出し、発話に応じてアクション108を実行する。コンピューティングデバイス106は、コンピューティングデバイス106の文脈110を使用して、ユーザ102の可能性の高い意図を判定することができる。ユーザ102の可能性の高い意図に基づいて、コンピューティングデバイス106は、適切なアクションを選択する。

【0013】

より詳細には、ユーザ102は、コンピューティングデバイス106の近くで発話104を行う。コンピューティングデバイス106は、オーディオを検出するように構成された任意のタイプのコンピューティングデバイスであり得る。例えば、コンピューティングデバイス106は、携帯電話、ラップトップコンピュータ、ウェアラブルデバイス、スマートアプライアンス、デスクトップコンピュータ、テレビ、またはオーディオデータを受信することができる任意の他のタイプのコンピューティングデバイスであってよい。コンピューティングデバイス106は、発話104のオーディオデータを処理し、単語ラティス112を生成する。単語ラティス112は、複数の単語からなる互いに異なる組み合わせを表し、それぞれについて、発話および信頼スコアに対応し得る。

40

【0014】

図1に示すように、ユーザ102は、明瞭に話しておらず、「フライツアウト (flights out)」のように聞こえる発話104を行う。コンピューティングデバイス

50

106は、ユーザの音声のオーディオを検出し、自動音声認識を使用して、ユーザ102が何を言ったかを判定する。コンピューティングデバイス106は、音響モデルをオーディオデータに適用する。音響モデルは、オーディオデータに対応する可能性が高いさまざまな音素を判定するように構成され得る。例えば、音響モデルは、オーディオデータが、他の音素に加えて「f」音素の音素が含まれていると判定する場合がある。コンピューティングデバイス106は、言語モデルを音素に適用することができる。言語モデルは、音素に一致する、可能性の高い単語および一連の単語を特定するように構成されてよい。一部の実装形態では、コンピューティングデバイス106は、オーディオデータを別のコンピューティングデバイス、例えば、サーバに送信することができる。サーバは、オーディオデータに対して音声認識を実行する場合がある。

10

【0015】

言語モデルは、発話104に一致する複数の語からなる互いに異なる組み合わせを含む単語ラティス112を生成する。単語ラティスは、発話104の最初の単語であり得る2つの単語を含む。「ライツ(lights)」という単語114が最初の単語である場合もあれば、「フライツ(flights)」という単語116が最初の単語である場合もある。言語モデルは、単語ラティス内の単語がユーザが話した単語である可能性を反映する信頼スコアを計算し得る。例えば、言語モデルは、単語114「ライツ(lights)」の信頼スコア118が0.65であり、単語116「フライツ(flights)」の信頼スコア120が0.35であると判定できる。単語ラティス112は、2番目の単語としてあり得る単語を含む。この例では、言語モデルは、2番目の単語として1つのあり得る単語122「アウト(out)」のみを特定した。

20

【0016】

一部の实装形態では、コンピューティングデバイス106は、音響モデルおよび言語モデルの代わりに、シーケンスツーシーケンスニューラルネットワーク、または他のタイプのニューラルネットワークを使用することができる。ニューラルネットワークは、1つまたは複数の隠れ層を有することができる。各サンプルの発話に対応するサンプルの発話および転写のオーディオデータを含む機械学習およびトレーニングデータを使用してトレーニングすることができる。この例では、シーケンスツーシーケンスニューラルネットワークは、単語ラティス112と信頼スコア118および120を生成することができる。シーケンスツーシーケンスニューラルネットワークは、音響モデルや言語モデルのように、音素および単語の組み合わせに対して個別の信頼スコアを生成しない場合がある。代わりに、シーケンスツーシーケンスニューラルネットワークは、音響モデルによって生成された音素の信頼スコアと、言語モデルによって生成された単語の組み合わせの信頼スコアと、の組み合わせである信頼スコアを生成する場合がある。

30

【0017】

単語ラティス112に基づいて、コンピューティングデバイス106は、発話104の2つの候補転写を特定する。第1の候補転写は「消灯(lights out)」であり、第2の候補転写は「フライツアウト(flights out)」である。第1の候補転写の信頼スコアは0.65であり、第2の候補転写の信頼スコアは0.35である。各候補転写の信頼スコアは、候補転写の各単語の信頼スコアの積であってもよい。フレーズ(phase)「ファイツアウト(fights out)」は、発話104に最も近い音響一致であり得るが、言語モデルに基づいて、「ファイト(fight)」と「アウト(out)」の組み合わせは生じる可能性が低い。したがって、「ファイツアウト(fights out)」は転写の候補とはならない。音響モデルと言語モデルの代わりにニューラルネットワークが使用される実装形態では、ニューラルネットワークは、音響モデルと言語モデルを使用して個別の信頼スコアを生成しない場合がある。

40

【0018】

コンピューティングデバイス106は、コンピューティングデバイス106の文脈110を判定することができる。文脈110は、コンピューティングデバイス106上またはその周囲に存在する要因の任意の組み合わせに基づいていてもよい。例えば、文脈110

50

は、コンピューティングデバイス 106 がユーザ 102 の家に位置していることを含み得る。文脈 110 は、現在の時刻が午後 10 時であり、曜日が火曜日であることを含み得る。文脈 110 は、コンピューティングデバイス 106 が、コンピューティングデバイス 106 のフォアグラウンドでデジタルアシスタントアプリケーションを実行している携帯電話であることを含み得る。

【0019】

一部の実装形態では、文脈 110 は、追加情報を含み得る。例えば、文脈 110 は、テーブル上またはユーザの手の上で平らであるなど、コンピューティングデバイス 106 の配向に関するデータを含み得る。文脈 110 は、バックグラウンドで実行しているアプリケーションを含み得る。文脈は、発話 104 を受信する前にコンピューティングデバイス 106 によって出力されたオーディオまたはコンピューティングデバイスの画面に表示されたデータを含み得る。例えば、「こんにちは、いかがいたしましょう？」というプロンプトを含む表示によって、コンピューティングデバイス 106 がフォアグラウンドで仮想アシスタントアプリケーションを実行していることが示され得る。文脈 110 は、天気、ユーザ 102 の識別情報、ユーザ 102 の人口統計データ、および連絡先などのコンピューティングデバイス 106 に記憶されているかまたはコンピューティングデバイス 106 によってアクセス可能なデータを含み得る。

【0020】

コンピューティングデバイス 106 は、文脈を使用して、転写に適用する文法を選択することができる。文法は、バックス・ナウア記法 (Backus - Naur form) などの一般的な表記法を使用して記述できる複数の単語からなる任意の構造とすることができる。各文法は、特定のユーザの意図に対応している場合がある。例えば、ユーザの意図は、ホームオートメーションコマンドまたはメディア再生コマンドを発行することである可能性がある。文法の一例には、アラームの文法が含まれ得る。アラーム文法では、 $\$DIGIT = (0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9)$ という表記を使用して、数字を 0、1、2、3、4、5、6、7、8、9、または 0 として定義できる。アラーム文法は、時刻に 2 桁、その次にコロン、その次に 2 桁、その次に「am」または「pm」とすることを示す $\$TIME = \$DIGIT\$DIGIT : \$DIGIT\$DIGIT (am | pm)$ という表記を使用して時刻を定義できる。アラーム文法は、アラームをアラームモードにするかタイマーモードにするかを示す $\$MODE = (alarm | timer)$ という表記を使用して、アラームのモードを定義できる。最後に、アラーム文法は、ユーザが「午前 6 時にアラームを設定 (set alarm for 6:00 am)」または「20:00 にタイマーを設定 (set timer for 20:00)」とすることができることを示す、アラームシンタックスを $\$ALARM = set \$MODE for \$TIME$ として定義できる。コンピューティングデバイス 106 は、文法を使用して、話されたコマンドまたはタイプされたコマンドをパースし、コンピューティングデバイス 106 が実行するべきアクションを特定する。したがって、文法によって、コマンドをパースするためにコンピューティングデバイス 106 を特定の態様で動作させる機能データが提供される。

【0021】

各文法は、特定の文脈でアクティブになる場合がある。例えば、コンピューティングデバイス 106 のフォアグラウンドで実行しているアプリケーションがアラームアプリケーションである場合、アラーム文法はアクティブであり得る。さらには、コンピューティングデバイスのフォアグラウンドで実行しているアプリケーションがデジタルアシスタントアプリケーションである場合、アラーム文法がアクティブになる可能性がある。ユーザがアラームアプリケーションでアラームを設定しようとする可能性は、デジタルアシスタントアプリケーションの場合よりも高い可能性があるため、コンピューティングデバイス 106 は、コンピューティングデバイス 106 がアラームアプリケーションを実行している場合、コマンドおよびユーザの意図と一致するアラーム文法の可能性に 0.95 の確率を割り当ててもよく、コンピューティングデバイス 106 がデジタルアシスタントアプリケ

10

20

30

40

50

ーションを実行している場合、0.03の確率を割り当ててもよい。

【0022】

図1に示す例では、コンピューティングデバイス106は、文法に適用され得る互いに異なる候補転写に基づいて、各文法の信頼スコアを判定する。換言すると、コンピューティングデバイス106は、候補転写のそれぞれをパースする文法を特定する。適用された文法124は、候補転写である「消灯(light s out)」をパースする文法を示し、これにより、各文法がユーザの意図に対して正しい文法であるという文法信頼スコアが示されている。より高い文法信頼スコアは、その文法は所与の転写と文脈に基づいている可能性が高いことを示す。転写が「消灯(light s out)」の場合、ホームオートメーション文法の文法信頼スコア126は0.35である。転写が「消灯(light s out)」の場合、映画コマンド文法の文法信頼スコア128は0.25である。転写が「消灯(light s out)」の場合、音楽コマンド文法の文法信頼スコア130は0.20である。各転写の文法信頼スコアの合計は1.0となるべきである。その場合、コンピューティングデバイス106は、残りの文法信頼スコアをデフォルト意図文法に割り当てることができる。この例では、転写が「消灯(light s out)」の場合、デフォルト意図文法の文法信頼スコア132は0.20である。デフォルト意図文法は、特定のアクションまたは意図に限定されない場合があり、すべてまたはほぼすべての転写をパース可能である。

10

【0023】

一部の実装形態では、いずれの文法も特定の転写をパースできない場合がある。この場合、デフォルト意図文法が適用される。例えば、「フライトアウト(flight s out)」をパースする文法がなくともよい。このため、転写が「フライトアウト(flight s out)」の場合、デフォルト意図文法の文法信頼スコア132は1.0である。

20

【0024】

一部の実装形態では、コンピューティングデバイス106は、転写信頼スコアと文法信頼スコアとの積を計算することに基づいて、文法と転写を選択する。ボックス134は、転写信頼スコアと文法信頼スコアとの積に基づいて、転写および文法を選択した結果を示している。例えば、「消灯(light s out)」とホームオートメーション文法とによる組み合わせ信頼スコア136は0.2275である。「消灯(light s out)」と映画コマンド文法とによる組み合わせ信頼スコア138は0.1625である。「消灯(light s out)」と音楽コマンド文法とによる組み合わせ信頼スコア140は0.130である。「消灯(light s out)」とデフォルト意図文法とによる組み合わせ信頼スコア142は0.130である。「フライトアウト(flight s out)」とデフォルト意図文法とによる組み合わせ信頼スコア144は0.350である。この実装形態では、組み合わせ信頼スコア144が最高スコアである。したがって、この手法を使用して、コンピューティングデバイス106は、コマンド「フライトアウト(flight s out)」に対応するアクションを実行することができる。これは、ユーザ102にとっては妥当でない結果である可能性が高く、検索エンジンに転写「フライトアウト(flight s out)」を提供する結果となる可能性がある。

30

40

【0025】

インターネットで「フライトアウト(flight s out)」を検索することがユーザ102の意図ではない場合、ユーザは発話104を繰り返す必要がある。そうすることによって、コンピューティングデバイス106が、追加の発話を処理するために追加のコンピューティングおよび電力リソースを消費することが必要となる。ユーザ102は、コンピューティングデバイス102の画面を追加の期間においてアクティブにすることによって、追加の処理および電力リソースを使用し得るコンピューティングデバイス102に手動で所望のコマンドを入力することになる。

【0026】

コンピューティングデバイス106がユーザの意図と一致しないアクションを実行する

50

可能性を減らすために、コンピューティングデバイス 106 は、適用された文法 124 の文法信頼スコアを正規化することができる。適用された文法 124 の文法信頼スコアを正規化するために、コンピューティングデバイス 106 は、最高の文法信頼スコアを 1.0 に増加させるために必要な係数を計算することができる。換言すると、係数と最高の文法信頼スコアとの積は 1.0 となるべきである。次に、コンピューティングデバイス 106 は、他の文法信頼スコアを係数で乗算して、他の文法信頼スコアを正規化することができる。コンピューティングデバイス 106 によって実行される正規化は、確率の合計が 1 になるように調整される従来の確率正規化とは異なる場合がある。最高の文法信頼スコアを 1.0 に増加させるので、正規化された確率の合計は 1 にならない。増加させられた信頼スコアは、従来の意味での確率ではなく、疑似確率を表す場合がある。本願の他の場所で説明されている正規化処理でも、同様の疑似確率が生成される場合がある。

10

【0027】

正規化された文法信頼スコア 146 に示されているように、コンピューティングデバイス 106 は、ホームオートメーション文法の文法信頼スコアを最高文法信頼スコアとして特定し得る。文法信頼スコア 126 を 1.0 に増加させるために、コンピューティングデバイスは、文法信頼スコア 126 に $1.0 / 0.35 = 2.857$ を乗算する。コンピューティングデバイス 106 は、文法信頼スコア 128 に 2.857 を乗算することによって、正規化された文法信頼スコア 150 を計算する。コンピューティングデバイス 106 は、文法信頼スコア 130 に 2.857 を乗算することによって、正規化された文法信頼スコア 152 を計算する。コンピューティングデバイス 106 は、文法信頼スコア 132 に 2.857 を乗算することによって、正規化された文法信頼スコア 154 を計算する。転写が「フライツアウト (flights out)」の場合のデフォルト意図文法の文法信頼スコアは 1.0 であり、これは、転写「フライツアウト (flights out)」をパースする文法が他にないためである。したがって、転写が「フライツアウト (flights out)」である場合、スコアがすでに 1.0 であるので、デフォルト意図文法の文法信頼スコアを正規化する必要はない。

20

【0028】

ボックス 134 に示されているように、ボックス 124 の文法信頼スコアに単語ラティス 112 の転写信頼スコアを乗算する代わりに、コンピューティングデバイス 106 は、単語ラティス 112 からの転写信頼スコアと正規化された文法信頼スコアとを使用して、組み合わせられた信頼スコアを計算する。特に、コンピューティングデバイス 106 は、正規化された文法信頼スコアのそれぞれを、それぞれの転写の転写信頼スコアで乗算する。デフォルト意図文法のみが文法に適用される場合は、転写信頼スコアに 1.0 を乗算するので、対応する転写信頼スコアは変化しない。

30

【0029】

ボックス 156 に示すように、コンピューティングデバイス 106 は、「消灯 (lights out)」の転写信頼スコアに正規化された文法信頼スコア 148 を乗算して 0.650 の結果を得ることによって、組み合わせ信頼スコア 158 を計算する。コンピューティングデバイス 106 は、「消灯 (lights out)」の転写信頼スコアに正規化された文法信頼スコア 150 を乗算して 0.464 の結果を得ることによって、組み合わせ信頼スコア 160 を計算する。コンピューティングデバイス 106 は、「消灯 (lights out)」の転写信頼スコアに正規化された文法信頼スコア 152 を乗算して 0.371 の結果を得ることによって、組み合わせ信頼スコア 162 を計算する。コンピューティングデバイス 106 は、「消灯 (lights out)」の転写信頼スコアに正規化された文法信頼スコア 154 を乗算して 0.371 の結果を得ることによって、組み合わせ信頼スコア 164 を計算する。コンピューティングデバイス 106 は、転写を「フライツアウト (flights out)」をパースする文法ないので、「フライツアウト (flights out)」の転写信頼スコアを 0.350 に維持する。

40

【0030】

一部の実装形態では、コンピューティングデバイス 106 は、コンピューティングデバ

50

イス106の現在の文脈の現在のユーザ文脈を考慮して、可能性が高い文法に対応できるように、ボックス156内の信頼スコアを調整できる。例えば、ユーザ102は、コンピューティングデバイス106を使用して音楽を聴いている場合がある。コンピューティングデバイス106は、曲を再生するメディアデバイスであり得る。この場合、コンピューティングデバイス106は、信頼スコア162を調整し得る。コンピューティングデバイス106は、信頼スコアを係数で乗算することによって、プリセット値を信頼スコアに割り当てることによって、または別の技術によって、信頼スコアを増加させることができる。例えば、音楽コマンドの現在の確率は、メディアデバイスであるコンピューティングデバイス102を介して音楽を聴いているユーザ102の文脈に基づいて、0.9である。ボックス124内の音楽コマンドの確率は、0.20である信頼スコア130であり得る。この場合、コンピューティングデバイス106は、信頼スコア162に $0.9 / 0.2 = 4.5$ の比率を乗算してよい。結果として得られる信頼スコア162は、 $0.371 * 4.5 = 1.67$ となる。ボックス156の他の信頼スコアは、それぞれの信頼スコアについて同様の比率で調整することができる。例えば、ホームオートメーションコマンドの現在の確率は0.04である可能性がある。この場合、比率は $0.04 / 0.35 = 0.116$ となり、これは、0.04を信頼スコア126で割ったものである。コンピューティングデバイス106は、調整された、またはバイアスされた、0.10の信頼スコアを計算するために、信頼スコア158に0.16を乗算することができる。この場合、最高の信頼スコアは、音楽コマンドに対応するものとなる。

10

【0031】

20

一部の実装形態では、この追加の調整ステップは、コンピューティングデバイス106が選択する候補転写に影響を与える可能性がある。例えば、候補転写「フライトアウト(flight out)」がビデオゲームの名前であり、ユーザがビデオゲームを起動することが期待される場合、コンピューティングデバイス106は、上記で計算されたものと同様の比率に基づいて、そしてユーザがビデオゲームを起動する確率を使用して、信頼スコア166を0.8と調整することができ、これは、コンピューティングデバイス106および/またはユーザ102の現在のコンテキストに基づいている。

【0032】

一部の実装形態では、複数の文法が同じ候補転写をパースする場合など、コンピューティングデバイス106は、この再スコアリングステップを使用せずに、ユーザの意図および音声認識の検出を改善することができる。この再スコアリングステップは、コンピューティングデバイス106が、音声認識信頼スコアに基づいて可能性高いとはされない、異なる候補転写を選択することを可能にし得る。

30

【0033】

一部の実装形態では、コンピューティングデバイス106は、ボックス124の信頼スコアを考慮せずに、この再スコアリングステップを使用することができる。例えば、コンピューティングデバイスは、単語ラティス112から特定された転写信頼スコアに再スコアリングステップを適用することができる。この場合、コンピューティングデバイス106は、ボックス146およびボックス156の両方に示されている調整を実行しないと考えられる。

40

【0034】

コンピューティングデバイス106は、最高の組み合わせ信頼スコアを有する文法および転写を選択する。例えば、ホームオートメーション転写および転写「消灯(light out)」は、0.650の最高の組み合わせ信頼スコアを有し得る。この場合、コンピューティングデバイス106は、ホームオートメーション文法に基づいて「消灯(light out)」を実行する。コンピューティングデバイス106は、コンピューティングデバイス106が配置されている家またはユーザ102の家の照明を消すことができる。コンピューティングデバイス106が映画コマンド文法を使用した場合、コンピューティングデバイス106は、映画「ライトアウト(Light Out)」を再生することができる。コンピューティングデバイス106が音楽コマンド文法を使用した場

50

合、コンピューティングデバイス106は、曲「ライトアウト(Lights Out)」を再生することができる。

【0035】

図1に示すように、コンピューティングデバイス106のディスプレイは、コンピューティングデバイス106によって実行されるアクションを示し得る。最初に、コンピューティングデバイス106は、デジタルアシスタントのためのプロンプト168を表示することができる。コンピューティングデバイス106は、発話104を受信し、プロンプト170を表示することによって、コンピューティングデバイス106が照明をオフにしていることを示す。

【0036】

図2は、文脈に基づいて発話に適用する文法を選択する例示的システム200の構成要素を示す。システム200は、音声オーディオを受信および処理するように構成された任意の種類コンピューティングデバイスであり得る。例えば、システム200は、図1のコンピューティングデバイス106と同様であり得る。システム200の構成要素は、単一のコンピューティングデバイスに実装することも、複数のコンピューティングデバイスに分散させることもできる。単一のコンピューティングデバイスに実装されているシステム200は、プライバシー上の理由から有益である可能性がある。

【0037】

システム200は、オーディオサブシステム205を含む。オーディオサブシステム205は、マイクロフォン210、アナログ-デジタル変換器215、バッファ220、および他の様々なオーディオフィルタを含み得る。マイクロフォン210は、音声などの周囲領域の音を検出するように構成することができる。アナログ-デジタル変換器215は、マイクロフォン210によって検出されたオーディオデータをサンプリングするように構成され得る。バッファ220は、システム200による処理のために、サンプリングされたオーディオデータを記憶することができる。一部の实装形態では、オーディオサブシステム205は、継続的にアクティブであり得る。この場合、マイクロフォン210は常に音を検出している可能性がある。アナログ-デジタル変換器215は、検出されたオーディオデータを常にサンプリングし得る。バッファ220は、音の最後の10秒などの最新のサンプリングされたオーディオデータを記憶することができる。システム200の他の構成要素がバッファ220におけるオーディオデータを処理しない場合、バッファ220は前のオーディオデータを上書きすることができる。

【0038】

オーディオサブシステム205は、処理されたオーディオデータを音声認識部225に提供する。音声認識部は、音響モデル230への入力として音声データを提供する。音響モデル230は、オーディオデータにおける音に対応する可能性のある音素を特定するように訓練されてよい。例えば、ユーザが「セット(set)」と言った場合、音響モデル230は、「s」音、「e」母音、および「t」音に対応する音素を特定し得る。音声認識部225は、特定された音素を言語モデル235への入力として提供する。言語モデル235は、単語ラティスを生成する。単語ラティスは、言語モデル235によって特定された候補用語のそれぞれについての用語信頼スコアを含む。例えば、単語ラティスは、最初の用語が「セット(set)」の可能性が高いことを示し得る。言語モデル235は、最初の用語について他の可能性のある用語を特定しない場合がある。言語モデル235は、第2の用語について2つのあり得る用語を特定し得る。例えば、単語ラティスは、あり得る第2の用語として「時間(time)」および「チャイム(chime)」という用語を含み得る。言語モデル235は、各用語に用語信頼スコアを割り当てることができる。「時間」の用語信頼スコアは0.75であり、「チャイム」の用語信頼スコアは0.25である可能性がある。

【0039】

音声認識部225は、単語ラティスに基づいて候補転写を生成することができる。各候補転写は、話者が転写における用語を話した可能性を反映する転写信頼スコアを有する場

10

20

30

40

50

合がある。例えば、候補転写は「時間をセット (s e t t i m e) 」であり、0 . 7 5 の転写信頼スコアを有する可能性がある。別の候補転写は「チャイムをセット (s e t c h i m e) 」であり、0 . 2 5 の転写信頼スコアを有する可能性がある。

【 0 0 4 0 】

音声認識部 2 2 5 が単語ラティス、候補転写、および転写信頼スコアを生成している間、文脈特定部 2 4 0 は、システム 2 0 0 の現在の文脈を示す文脈データを収集し得る。文脈特定部 2 4 0 は、システムの任意のセンサからセンサデータ 2 4 5 を収集することができる。センサは、位置センサ、温度計、加速度計、ジャイロ스코ープ、重力センサー、時間と曜日、および他の同様のセンサを含み得る。センサデータ 2 4 5 はまた、システム 2 0 0 のステータスに関連するデータを含み得る。例えば、ステータスは、システム 2 0 0 のバッテリーレベル、信号強度、システム 2 0 0 が通信している、または認識している可能性のある近くのデバイス、およびシステム 2 0 0 の他の同様の状態を含み得る。

10

【 0 0 4 1 】

文脈特定部 2 4 0 はまた、システム 2 0 0 が実行しているプロセスを示すシステムプロセスデータ 2 5 0 を収集することができる。プロセスデータ 2 5 0 は、各プロセスに割り当てられたメモリ、各プロセスに割り当てられたプロセスリソース、システム 2 0 0 が実行しているアプリケーション、システム 2 0 0 のフォアグラウンドまたはバックグラウンドで実行しているアプリケーション、システムのディスプレイ上のインタフェースの内容、および同様のプロセスデータを示し得る。

【 0 0 4 2 】

例として、文脈特定部 2 4 0 は、システム 2 0 0 がユーザの自宅であり、時刻が午後 6 時であり、曜日が月曜日であり、フォアグラウンドアプリケーションがデジタルアシスタントアプリケーション、デバイスがタブレットであることを示すシステムプロセスデータ 2 5 0 およびセンサデータ 2 4 5 を受信することができる。

20

【 0 0 4 3 】

文法スコア生成部 2 5 5 は、文脈特定部 2 4 0 からシステム 2 0 0 の文脈を受信し、音声認識部 2 2 5 から単語ラティスを受信する。文法スコア生成部 2 5 5 は、単語ラティスの候補転写のそれぞれをパースする文法 2 6 0 を特定する。場合によっては、文法 2 6 0 のいずれも候補転写をパースしない。この場合、文法スコア生成部 2 5 5 は、文法 2 6 0 のいずれによってもパース (p a r a b l e) できない候補転写にデフォルト意図文法の 1 . 0 の文法信頼スコアを与える。

30

【 0 0 4 4 】

一部の例では、単一の文法 2 6 0 が、候補転写をパースすることができる。この場合、文法スコア生成部 2 5 5 は、候補転写が実際の転写であると仮定して、単一の文法の文法信頼スコアを判定してもよい。文法信頼スコアは確率を表すので、文法信頼スコアは 1 . 0 未満である可能性がある。文法スコア生成部 2 5 5 は、1 . 0 と文法信頼スコアとの間の差を、候補転写のデフォルトの意図文法の文法信頼スコアに割り当てる。

【 0 0 4 5 】

一部の例では、複数の文法 2 6 0 が、候補転写をパースすることができる。この場合、文法スコア生成部 2 5 5 は、候補転写が実際の転写であると仮定して、複数の文法の各々の文法信頼スコアを判定してもよい。文法信頼スコアは確率の集合を表すので、文法信頼スコアの合計は 1 . 0 未満である可能性がある。文法スコア生成部 2 5 5 は、1 . 0 と文法信頼スコアの合計との間の差を、候補転写のデフォルトの意図文法の文法信頼スコアに割り当てる。

40

【 0 0 4 6 】

文法スコア正規化部 2 6 5 は、候補転写のそれぞれについての文法信頼スコアを受信し、それらの文法信頼スコアを正規化する。一部の実装形態では、文法スコア正規化部 2 6 5 は、デフォルト意図文法以外の文法の文法信頼スコアのみを正規化する。文法スコア生成部 2 5 5 が、特定の候補転写について、1 つの文法に対して 1 つの文法信頼スコアを生成する場合、文法スコア正規化部 2 6 5 は、当該文法信頼スコアを 1 . 0 に増やす。文法

50

スコア生成部 255 が、特定の候補転写について、複数の文法に対して文法信頼スコアを生成しない場合、文法スコア正規化部 265 は、デフォルト意図文法の文法信頼スコアを 1.0 に維持する。

【0047】

文法スコア生成部 255 が、特定の候補転写について、複数の文法の各々に対して複数の文法信頼スコアを生成する場合、文法スコア正規化部 265 は、特定の候補転写について最高の文法信頼スコアを特定する。文法スコア正規化部 265 は、係数と最高の文法信頼スコアとの積が 1.0 になるように、最高の文法信頼スコアを 1.0 に増加させるための係数を計算する。文法スコア正規化部 265 は、係数を他の文法信頼スコアのそれぞれに乗算することによって、同じ特定の転写についての他の文法信頼スコアを増加させる。

10

【0048】

文法および転写選択部 270 は、正規化された文法信頼スコアおよび転写信頼スコアを受信する。文法信頼スコアと転写信頼スコアの両方を使用して、文法および転写選択部 270 は、組み合わせ信頼スコアを計算することによって、受信された発話および話者の意図にほぼ一致する可能性が高い文法および転写を特定する。文法および転写選択部 270 は、各正規化された文法信頼スコアおよび対応する候補転写の転写信頼スコアの積を計算することによって、それぞれの組み合わせ信頼スコアを判定する。デフォルト意図文法が特定の候補転写の唯一の文法である場合、文法および転写選択部 270 は、組み合わせ信頼スコアとして転写信頼スコアを維持する。文法および転写選択部 270 は、最高の組み合わせ信頼スコアを有する文法および候補転写を選択する。

20

【0049】

アクション特定部 275 は、選択された文法および選択された転写を文法および転写選択部 270 から受信し、システム 200 が実行するアクションを特定する。選択された文法は、アラームの設定、メッセージの送信、曲の再生、人への電話、またはその他の同様のアクションなどのアクションのタイプを示している場合がある。選択された転写は、アラームを設定する時間、メッセージの送信先、再生する曲、呼び出し先、またはアクションのその他の同様の詳細など、アクションのタイプの詳細を示す場合がある。文法 260 は、特定の文法のアクションのタイプに関する情報を含み得る。文法 260 は、アクション特定部 275 が候補転写をどのようにパースしてアクションのタイプの詳細を判定すべきかについての情報を含み得る。例えば、文法は \$ アラーム (\$ ALARM) 文法である可能性があり、アクション特定部 275 は選択された転写をパースして、タイマーを 20 分に設定することを判定する。アクション特定部 275 は、アクションを実行するか、またはシステム 200 の別の部分、例えば、プロセッサに命令を提供することができる。

30

【0050】

一部の実装形態では、ユーザインタフェース生成部 280 は、アクションの標示またはアクションの実行の標示、あるいはその両方を表示する。例えば、ユーザインタフェース生成部 280 は、20 分からカウントダウンするタイマー、または家の中で照明をオフとすることの確認を表示することができる。一部の例では、ユーザインタフェース生成部 280 は、実行されたアクションの標示を提供しない場合がある。例えば、アクションはサーモスタットを調整している可能性がある。システム 200 は、システム 200 上に表示するためのユーザインタフェースを生成することなく、サーモスタットを調整してもよい。一部の实装形態では、ユーザインタフェース生成部 280 は、ユーザがアクションと対話するか、またはアクションを確認するためのインタフェースを生成することができる。例えば、アクションは母親に電話することであってもよい。ユーザインタフェース生成部 280 は、システム 200 がアクションを実行する前に、母親を呼び出すアクションをユーザが確認するためのユーザインタフェースを生成することができる。

40

【0051】

図 3 は、文脈に基づいて文法を選択する例示的プロセス 300 を示すフローチャートである。一般に、プロセス 300 は、オーディオに対して音声認識を実行し、オーディオの転写およびオーディオをパースする文法に基づいて実行するアクションを特定する。プロ

50

セス300は、文法信頼スコアを正規化して、話者が意図した可能性が高いアクションを特定する。プロセス300は、1つまたは複数のコンピュータを含むコンピュータシステム、例えば、図1のコンピューティングデバイス106または図2のシステム200によって実行されるものとして説明される。

【0052】

システムは、発話のオーディオデータを受信する(310)。例えば、ユーザは、「消灯(lights out)」または「フライツアウト(flights out)」のように聞こえる発話を話すことがある。システムは、マイクを介して発話を検出するか、発話のオーディオデータを受信する。システムは、オーディオサブシステムを使用してオーディオデータを処理することができる。

10

【0053】

システムは、音響モデルおよび言語モデルを使用して、発話の複数の候補転写を含むとともに、複数の転写信頼スコアを含み、該複数の転写信頼スコアのそれぞれが、それぞれの候補転写が発話と一致する可能性を反映する単語ラティスを生成する(320)。システムは自動音声認識を使用して単語ラティスを生成する。自動音声認識プロセスは、オーディオデータの各部分に一致する様々な音素を特定する音響モデルへの入力としてオーディオデータを提供することを含み得る。自動音声認識プロセスは、発話中の各候補単語の信頼スコアを含む単語ラティスを生成する言語モデルへの入力として音素を提供することを含み得る。言語モデルは、語彙から単語ラティスの単語を選択する。語彙は、システムが認識するように構成されている言語の単語を含む場合がある。例えば、システムは英語のために構成されてよく、語彙には英語の単語を含んでいてよい。一部の実装形態では、プロセス300のアクションは、システムが認識できる語彙内の単語を制限することを含まない。換言すると、プロセス300は、システムが認識するように構成された言語の任意の単語を用いて転写を生成することができる。システムは言語における各単語を認識できるので、システムの音声認識プロセスは、ユーザが予期しないことを言ったとき、システムが認識するように構成されている言語内である限り機能できる。

20

【0054】

システムは、単語ラティスを使用して、様々な候補転写を生成できる。各候補転写は、ユーザが転写を話した可能性を反映する異なる転写信頼スコアを含む場合がある。例えば、候補転写「消灯(lights out)」の信頼スコアは0.65であり、候補転写「フライツアウト(flights out)」の信頼スコアは0.35である可能性がある。

30

【0055】

システムは、システムの文脈を判定する(330)。一部の实装形態では、文脈は、システムの場所、コンピューティングデバイスのフォアグラウンドで実行しているアプリケーション、ユーザの人口統計情報、連絡先データなど、システムに記憶されているかシステムからアクセスできるデータ、以前のユーザのクエリまたはコマンド、時刻、日付、曜日、天気、システムの向き、およびその他の同様のタイプの情報に基づいている。

【0056】

システムは、システムの文脈に基づいて、複数の候補転写に対応する複数の文法を特定する(340)。文法は、システムが実行できる様々なコマンドに対して様々な構造を含む場合がある。例えば、文法は、アラームの設定、映画の再生、インターネット検索の実行、ユーザのカレンダーのチェック、またはその他の同様のアクションのためのコマンド構造を含む場合がある。

40

【0057】

システムは、現在の文脈に基づいて、複数の候補転写のそれぞれについて、それぞれの文法がそれぞれの候補転写と一致する可能性を反映する複数の文法信頼スコアを判定する(350)。文法信頼スコアは、候補転写の1つが発話の転写であるとして、システムの文脈に基づいて文法が話者の意図と一致する可能性の条件付き確率と考えられ得る。例えば、転写が「消灯(lights out)」の場合のホームオートメーション文法の文

50

法信頼スコアは0.35である可能性がある。換言すると、転写が「消灯 (light s o u t)」であるとして、ホームオートメーション文法の条件付き確率は0.35である。システムは、候補転写をパースする文法ごとに文法信頼スコアを生成する場合がある。一部の実装形態では、文脈に起因して、システムは、候補転写をパースする一部の文法についてのみ、文法信頼スコアを生成する場合がある。各候補転写の文法信頼スコアの合計が1.0になるよう、システムは、残りの確率をデフォルト意図文法に割り当てる場合がある。

【0058】

システムは、転写信頼スコアおよび文法信頼スコアに基づいて、複数の候補転写の中から1つの候補転写を選択する(360)。一部の实装形態では、システムは、複数の文法に一致する候補転写を増やすことにより、文法の信頼スコアを正規化する。候補転写ごとに、システムは、最高の文法信頼スコアに、文法信頼スコアを1.0に正規化するために必要な係数を乗算することができる。システムは、同じ候補転写に対する他の文法信頼スコアに対して、同じ係数で乗算することができる。システムは、正規化された文法信頼スコアに転写信頼スコアを掛け合わせて、組み合わせ信頼スコアを生成することができる。システムは、組み合わせられた信頼スコアが最も高い文法と転写を選択する。

10

【0059】

システムは、出力のために、選択された候補転写を発話の転写として提供する(370)。一部の实装形態では、システムは、文法と候補転写に基づいてアクションを実行する。文法は、取るべき行動を示している場合がある。アクションには、映画の再生、連絡先への電話、照明の点灯、メッセージの送信、またはその他の同様のタイプのアクションが含まれる場合がある。選択された転写には、電話をかける連絡先、送信するメッセージ、メッセージの受信者、映画の名前、またはその他の同様の詳細が含まれる場合がある。

20

【0060】

一部の实装形態では、システムは、有限状態トランスデューサー(FST)を使用して、所与の文法セットの単語ラティスにタグを付けるプロセスを使用する。一部の实装形態では、システムは、現在の文脈に一致するすべての文法の和集合として文法を制約する場合がある。一部の实装形態では、システムは、事前にオフラインで、または実行時に動的に照合する場合がある。

【0061】

一部の实装形態では、プロセス300は、単語ラティスのエッジに重みを付けることができる。システムは、文法を制約する重み付き有限状態トランスデューサへと文法をコンパイルすることができる。アークの重み、すなわち、この有限状態トランスデューサのエッジの重みは、所与の文法の単語の確率の量をエンコードする場合があり、これは、負の対数の重みである場合がある。一部の实装形態では、単語ラティスに関連するほぼすべての文法のすべてが、共に統合される場合がある。

30

【0062】

プロセス300は、各文法の文脈依存確率を判定するシステムを続行することができ、これは、文脈を所与として、文法の確率であり得る。システムは、提供された構成要素と対応する重みによって文脈依存の確率を判定することができ、これは、オープニングデコレータアークなどの文法制約部において、アークまたはエッジにエンコードされてもよい。

40

【0063】

プロセス300は、単語ラティスの重みをクリアするシステムを続行することができる。システムは、文法制約部を使用して単語ラティスを構成する場合があり、これにより、文法に一致するスパンがマークされる場合がある。例えば、文法に一致するスパンは、スパンについては<media_commands>曲を再生(play song)</media_commands>などのように開始および終了デコレータタグで囲まれてもよい(または「曲を再生(play song)」の転写)。

【0064】

プロセス300は、確率を正規化し続けてよい。システムは、デコレータタグが削除さ

50

れたタグが付けられるラティスのコピーを生成する。システムは、トロピカルセミリング (tropical semiring) のアークコストを判定および最小化し、その結果、それぞれの一意の単語経路が得られ、単語経路は、単語経路を通じた最も高い確率をエンコードする。確率は、文法を所与として単語経路の確率に文法の確率を乗算した確率であってもよい。確率は、負の対数の重みを用いてアークの重みの符号を反転することにより、反転させることができ、この反転したラティスは、タグ付きラティスで構成されてよい。反転格子で構成することは、その格子の重みで除算を実行することと実質的に同じである可能性がある (この場合、各単語シーケンスまたは経路ごとに、文法を所与として単語経路の確率の最大値に文法の確率を乗算したもの)。したがって、システムは最適なタグ付き経路の確率をそれ自体で分割し、それぞれが 1 . 0 になる可能性がある。一部の
10
実装形態では、最適でない経路は、より低い疑似確率を受け取り、所与の単語経路の文法の疑似確率である所望の量を含むラティスを生成する。

【 0 0 6 5 】

一部の实装形態では、プロセス 3 0 0 は、ビームプルーニング (beam pruning) を使用することによって、低い条件付き確率で文法を破棄することを含み得る。例えば、発話「消灯 (lights out)」に一致する文法が多すぎる場合である (ここで、ホームオートメーションの確率は 0 . 3 2、映画コマンド 0 . 2 5、音楽コマンド 0 . 2 0、一般検索 0 . 1 0、オンラインビジネス 0 . 0 3、社会集団 0 . 0 3、およびオンライン百科事典検索 0 . 0 2 である)。システムは、閾値を下回る、または最も可能性の高い解釈すなわち文法の閾値パーセンテージを下回る文法を取り除くことによって、
20
処理の負担を軽減することができる。例えば、システムは、最も可能性の高い文法の 1 0 分の 1 未満の可能性の文法を取り除いてよい。次に、システムは、所与の転写に対して 0 . 0 3 2 未満の文法を除去する場合がある。プルーニングの背後にある理論的根拠は、スコアの低いタグ付けの確率が非常に低いため、バイアシングがあっても、より可能性の高い解釈よりもその解釈が選択されるといった合理性がないことによる。プルーニングにおいては、システムが、所望のプルーニング重みで得られたラティスに動作する有限状態トランスデューサを追加する必要がある場合があり、これは、取り除くためにタグ付けが有する最良の仮定がどの遅れているかを指定する。

【 0 0 6 6 】

プロセス 3 0 0 の利点によれば、システムが文法をラティス上で直接バイアスすることを可能となる。プロセス 3 0 0 は、他の単語ラティスタグ付けアルゴリズムよりも複雑さが低い。プロセス 3 0 0 は、最大正規化を実装することによって、タグ付けの曖昧さによって引き起こされる確率の断片化を解決する。プロセス 3 0 0 は、ビームプルーニングを使用することによって、比較的可能性の低い提案されたタグ付けを破棄することを可能にする。
30

【 0 0 6 7 】

一部の实装形態では、プロセス 3 0 0 は、ラティスから n 個の最良の仮説を抽出し、すべての仮説および可能性に個別にタグを付け、オプションでそれらを最後にラティスに再結合することができる。システムは、タグが他の点では同一であると想定してよい。これによって、レイテンシが長くなり、および / または、リコールが小さくなる同じ結果が
40
得られる可能性がある。上位 N (N = 1 0 0 など) の後の n 個の最良の仮説が削除された場合、リコールは小さくなる可能性がある。これはレイテンシの管理に役立つ場合がある。ラティス内の個別の文の数は、その中の単語の数に対して指数関数的になる可能性がある。前記の最良 N 制限 (これにより、調べる選択肢の数が制限される場合がある) を使用しない限り、一部の回避策は最悪の場合に指数関数的に遅くなる可能性がある。

【 0 0 6 8 】

図 4 は、本明細書に記載した技術を実装するために使用され得るコンピューティング装置 4 0 0 およびモバイルコンピューティング装置 4 5 0 の例を示す。コンピューティング装置 4 0 0 は、ラップトップ、デスクトップ、ワークステーション、携帯情報端末、サーバ、ブレードサーバ、メインフレーム、および他の適切なコンピュータ等、様々な形態の
50

デジタルコンピュータを表すよう意図されている。モバイルコンピューティング装置 4 5 0 は、携帯情報端末、携帯電話、スマートフォン、および他の同様のコンピューティング装置等、様々な形態のモバイル装置を表すよう意図されている。本明細書に示された構成要素、構成要素の接続および関係、ならびに構成要素の機能は、例示的であることのみを意図されており、限定的であることは意図されていない。

【 0 0 6 9 】

コンピューティング装置 4 0 0 は、プロセッサ 4 0 2、メモリ 4 0 4、記憶装置 4 0 6、メモリ 4 0 4 および複数の高速拡張ポート 4 1 0 に接続する高速インタフェース 4 0 8、ならびに低速拡張ポート 4 1 4 および記憶装置 4 0 6 に接続する低速インタフェース 4 1 2 を含む。プロセッサ 4 0 2、メモリ 4 0 4、記憶装置 4 0 6、高速インタフェース 4 0 8、高速拡張ポート 4 1 0、および低速インタフェース 4 1 2 の各々は、様々なバスを介して相互接続され、共通のマザーボード上に、または必要に応じて他の方法で実装されてよい。プロセッサ 4 0 2 は、高速インタフェース 4 0 8 に接続されたディスプレイ 4 1 6 等の外部入力 / 出力装置に G U I のグラフィック情報を表示するために、メモリ 4 0 4 または記憶装置 4 0 6 に記憶された命令を含む、コンピューティング装置 4 0 0 内で実行するための命令を処理することが可能である。他の実装においては、複数のメモリおよび種類のメモリとともに、必要に応じて複数のプロセッサおよび / または複数のバスが使用されてよい。また、複数のコンピューティング装置が接続され、各装置が必要な動作の一部を提供してよい（例えば、サーババンク、ブレードサーバのグループ、またはマルチプロセッサシステムとして）。

【 0 0 7 0 】

メモリ 4 0 4 は、コンピューティング装置 4 0 0 内の情報を記憶する。いくつかの実装においては、メモリ 4 0 4 は、単数または複数の揮発性メモリユニットである。いくつかの実装においては、メモリ 4 0 4 は、単数または複数の不揮発性メモリユニットである。メモリ 4 0 4 は、磁気ディスクまたは光学ディスク等の別の形態のコンピュータ可読媒体であってもよい。

【 0 0 7 1 】

記憶装置 4 0 6 は、コンピューティング装置 4 0 0 に大容量記憶を提供することができる。いくつかの実装においては、記憶装置 4 0 6 は、フロッピー（登録商標）ディスク装置、ハードディスク装置、光学ディスク装置、またはテープ装置、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装置等のコンピュータ可読媒体、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装置を含む装置のアレイであるか、もしくはそれを含んでよい。命令は情報担体に記憶されてよい。命令は、1 つまたは複数の処理装置（例えば、プロセッサ 4 0 2）によって実行されると、上記のような 1 つまたは複数の方法を実行する。命令は、コンピュータ可読媒体または機械可読媒体（例えば、メモリ 4 0 4、記憶装置 4 0 6、またはプロセッサ 4 0 2 上のメモリ）等の 1 つまたは複数の記憶装置に記憶されてもよい。

【 0 0 7 2 】

高速インタフェース 4 0 8 は、コンピューティング装置 4 0 0 の帯域幅集中型の動作を管理する一方、低速インタフェース 4 1 2 は、より低帯域幅集中型の動作を管理する。このような機能の割り当ては一例に過ぎない。いくつかの実装形態では、高速インタフェース 4 0 8 は、メモリ 4 0 4、ディスプレイ 4 1 6（例えば、グラフィックプロセッサまたはアクセラレータを通じて）、および様々な拡張カード（図示せず）を受け得る高速拡張ポート 4 1 0 に接続される。実装において、低速インタフェース 4 1 2 は、記憶装置 4 0 6 および低速拡張ポート 4 1 4 に接続される。低速拡張ポート 4 1 4 は、種々の通信ポート（例えば、U S B、B l u e t o o t h（登録商標）、イーサネット（登録商標）、無線イーサネット）を含んでよく、例えばネットワークアダプタを通じて、キーボード、ポインティングデバイス、スキャナー、または、スイッチまたはルータ等のネットワーク装置等の、1 つまたは複数の入力 / 出力装置に接続されてもよい。

【 0 0 7 3 】

コンピューティング装置 400 は、図に示すように、多くの異なる形態で実装されてよい。例えば、標準のサーバ 420 として、またはそのようなサーバのグループに複数回実装されてよい。また、ラップトップコンピュータ 422 等のパーソナルコンピュータに実装されてよい。また、ラックサーバシステム 424 の一部として実装されてもよい。あるいは、コンピューティング装置 400 からの構成要素は、モバイルコンピューティング装置 450 等のモバイル装置（図示せず）内の他の構成要素と組み合わせられてよい。そのような装置のそれぞれは、コンピューティング装置 400 およびモバイルコンピューティング装置 450 のうちの 1 つまたは複数を含んでよく、システム全体は、互いに通信する複数のコンピューティング装置で構成されてよい。

【0074】

モバイルコンピューティング装置 450 は、構成要素の中でも特に、プロセッサ 452、メモリ 464、ディスプレイ 454 等の入力/出力装置、通信インタフェース 46、送受信機 468 を含む。モバイルコンピューティング装置 450 は、追加のストレージを提供するために、マイクロドライブまたは他の装置等の記憶装置を備えてもよい。プロセッサ 452、メモリ 464、ディスプレイ 454、通信インタフェース 466、および送受信機 468 は、様々なバスを介して各々に相互接続され、複数の構成要素は、共通のマザーボード上に、または必要に応じて他の方法で実装されてよい。

【0075】

プロセッサ 452 は、メモリ 464 に記憶された命令を含む、モバイルコンピューティング装置 450 内の命令を実行することができる。プロセッサ 452 は、別個で複数のアナログおよびデジタルプロセッサを含むチップのチップセットとして実装されてよい。プロセッサ 452 は、例えば、ユーザインタフェース、モバイルコンピューティング装置 450 によって実行されるアプリケーション、およびモバイルコンピューティング装置 450 による無線通信の制御等、モバイルコンピューティング装置 450 の他の構成要素の調整を提供してよい。

【0076】

プロセッサ 452 は、ディスプレイ 454 に接続された制御インタフェース 458 およびディスプレイインタフェース 456 を通じてユーザと通信してよい。ディスプレイ 454 は、例えば、TFT（薄膜トランジスタ液晶ディスプレイ）ディスプレイ、OLED（有機発光ダイオード）ディスプレイ、または他の適切なディスプレイ技術であってよい。ディスプレイインタフェース 456 は、ディスプレイ 454 を駆動してグラフィックおよび他の情報をユーザに提示するための適切な回路を含んでよい。制御インタフェース 458 は、ユーザからコマンドを受け取り、プロセッサ 452 に供給するために、コマンドを変換してよい。さらに、外部インタフェース 462 は、プロセッサ 452 との通信を提供して、モバイルコンピューティング装置 450 と他の装置との近領域通信を可能にしてよい。外部インタフェース 462 は、例えば、いくつかの実装では有線通信を提供し、他の実装では無線通信を提供してよく、複数のインタフェースを使用してもよい。

【0077】

メモリ 464 は、コンピューティング装置 450 内の情報を記憶する。メモリ 464 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の 1 つまたは複数として実装することができる。拡張メモリ 474 が提供され、拡張インタフェース 472 を介してモバイルコンピューティング装置 450 に接続されてよい。拡張インタフェース 472 は、例えば、SIMM（シングルインラインメモリモジュール）カードインタフェースを含んでよい。拡張メモリ 474 は、モバイルコンピューティング装置 450 に余分な記憶領域を提供してよく、またはモバイルコンピューティング装置 450 のアプリケーションまたは他の情報を記憶してもよい。具体的には、拡張メモリ 474 は、上述のプロセスを実行または補完する命令を含んでよく、保安情報を含んでもよい。したがって、例えば、拡張メモリ 474 は、モバイルコンピューティング装置 450 のセキュリティモジュールとして提供されてよく、モバイルコンピューティング装置 450 の安全な使用を可能にする命令でプログラムされてよい。さらに、SIMMカードを介して、ハ

10

20

30

40

50

ッキング不能な方法でSIMMカードに識別情報を配置する等の追加情報とともに、保安アプリケーションが提供されてよい。

【0078】

メモリは、以下で検討されるように、例えば、フラッシュメモリおよび/またはNVRAMメモリ(不揮発性ランダムアクセスメモリ)を含んでよい。いくつかの実装においては、命令は、情報担体に記憶される。命令は、1つまたは複数の処理装置(例えば、プロセッサ452)によって実行されると、上記のような1つまたは複数の方法を実行する。命令は、1つまたは複数のコンピュータ可読媒体または1つまたは複数の機械可読媒体(例えば、メモリ464、拡張メモリ474、またはプロセッサ452上のメモリ)等の1つまたは複数の記憶装置に記憶されてもよい。いくつかの実装においては、命令は、例えば送受信機468または外部インタフェース462を通じて、伝搬信号で受信することができる。

10

【0079】

モバイルコンピューティング装置450は、必要に応じてデジタル信号処理回路を含み得る通信インタフェース466を通じて無線で通信してよい。通信インタフェース466は、特に、GSM(登録商標)音声通話(モバイル通信用グローバルシステム)、SMS(ショートメッセージサービス)、EMS(エンハンスドメッセージングサービス)、MMSメッセージング(マルチメディアメッセージングサービス)、CDMA(符号分割多元接続)、TDMA(時分割多元接続)、PDC(パーソナルデジタルセルラ)、WCDMA(登録商標)(広帯域符号分割多元接続)、CDMA2000、またはGPRS(汎用パケット無線サービス)等、様々な態様またはプロトコルの下で通信を提供してよい。このような通信は、例えば、無線周波数を使用する送受信機468を通じて行われてよい。加えて、Bluetooth、WiFi(登録商標)、または他のそのような送受信機(図示せず)の使用等により、近距離通信が発生してよい。加えて、GPS(全地球測位システム)受信機モジュール470は、モバイルコンピューティング装置450上で実行されるアプリケーションにより適宜使用され得る、追加的なナビゲーションおよび位置関連の無線データをモバイルコンピューティング装置450に提供してよい。

20

【0080】

モバイルコンピューティング装置450は、オーディオコーデック460を使用して可聴的に通信してよく、オーディオコーデック460は、ユーザから口頭の情報を受信し、それを使用可能なデジタル情報に変換してよい。オーディオコーデック460は、同様に、例えば、モバイルコンピューティング装置450のハンドセット内のスピーカ等を通じて、ユーザに可聴音を生成してよい。このような音は、音声通話の音を含んでよく、録音された音(例えば、音声メッセージ、音楽ファイル等)を含んでよく、モバイルコンピューティング装置450上で動作するアプリケーションによって生成される音を含んでもよい。

30

【0081】

モバイルコンピューティング装置450は、図に示すように、多くの異なる形態で実装されてよい。例えば、モバイルコンピューティング装置450は、携帯電話480として実装されてよい。モバイルコンピューティング装置450は、スマートフォン482、携帯情報端末、または他の同様のモバイル装置の一部として実装されてもよい。

40

【0082】

本明細書で説明するシステムおよび技法のさまざまな実装は、デジタル電子回路、集積回路、特別に設計されたASIC(特定用途向け集積回路)、コンピュータハードウェア、ファームウェア、ソフトウェア、および/またはそれらの組み合わせで実現することができる。これらの様々な実装は、記憶装置、1つ以上の入力装置、および1つ以上の出力装置に対してデータおよび命令を送信すると共にこれらからデータおよび命令を受信するよう接続された、特定目的または汎用目的の1つ以上のプログラマブルプロセッサを備えたプログラマブルシステム上で実行可能および/または翻訳可能な1つまたは複数のコンピュータプログラムでの実装を含んでよい。

50

【 0 0 8 3 】

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、プログラマブルプロセッサのための機械語命令を含み、高水準手続き型および/またはオブジェクト指向プログラミング言語、および/またはアセンブリ/機械語で実装することができる。本明細書で使用されるように、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、プログラマブルプロセッサに機械語命令及び/又はデータを供給するために使用される、機械可読信号として機械語命令を受け取る機械可読媒体を含む、任意のコンピュータプログラムプロダクト、装置及び/又はデバイス（例えば、磁気ディスク、光学ディスク、メモリ、P L D（プログラマブルロジックデバイス））を指す。機械可読信号という用語は、機械語命令および/またはデータをプログラマブルプロセッサに提供するために使用される信号を指す。

10

【 0 0 8 4 】

ユーザとの対話を提供するため、本明細書で説明したシステムおよび技術は、ユーザに情報を表示するための表示装置（例えば、C R T（陰極線管）またはL C D（液晶ディスプレイ）モニター）と、ユーザが入力をコンピュータに提供可能なキーボードおよびポインティング装置（例えば、マウスまたはトラックボール）と、を含むコンピュータ上で実装されてよい。他の種類の装置を使用して、ユーザとの対話を提供してもよい。例えば、ユーザに提供されるフィードバックは、任意の形式の感覚的なフィードバック（例えば、視覚的フィードバック、聴覚的フィードバック、または触覚的フィードバック）であってよく、ユーザからの入力、音響的入力、音声的入力、または触覚的入力を含む任意の形式で取り込まれてよい。

20

【 0 0 8 5 】

本明細書で説明したシステムおよび技術は、バックエンドコンポーネント（例えば、データサーバ）を備えたコンピュータシステム、ミドルウェアコンポーネント（例えば、アプリケーションサーバ）を備えたコンピュータシステム、フロントエンドコンポーネント（例えば、ユーザが本明細書で説明されたシステムおよび技術の実装と対話を行うことができるグラフィカルインタフェースまたはウェブブラウザを有するクライアントコンピュータ）を備えたコンピュータシステム、または、このようなバックエンド、ミドルウェア、またはフロントエンドコンポーネントの任意の組合せを備えたコンピュータシステムで実施されてよい。システムの構成要素は、デジタルデータ通信（例えば、通信ネットワーク）の任意の形式または媒体によって相互接続されてよい。通信ネットワークの例は、L A N（ローカルエリアネットワーク）、W A N（ワイドエリアネットワーク）、およびインターネットを含む。一部の实装形態では、本明細書において説明するシステムおよび技術は、音声認識および他の処理がデバイス上で直接実行される組み込みシステム上に実装することができる。

30

【 0 0 8 6 】

コンピューティングシステムは、クライアントおよびサーバを含んでよい。クライアントとサーバは、一般には相互に離れており、典型的には通信ネットワークを通じて対話する。クライアントとサーバの関係は、各コンピュータ上で実行され、相互にクライアント・サーバ関係を有するコンピュータプログラムにより発生する。

40

【 0 0 8 7 】

いくつかの実装が詳細に説明されたが、他の変更も可能である。例えば、クライアントアプリケーションはデリゲートにアクセスするものとして説明されているが、他の実装では、デリゲートは、1つまたは複数のサーバで実行されるアプリケーション等、1つまたは複数のプロセッサによって実装された他のアプリケーションによって使用されてよい。さらに、図に示された論理の流れは、望ましい結果を得るために、示された特定の順序または連続した順序を必要とはしない。さらに、説明された流れに他の動作が提供されたり、または流れから除去されてよく、説明されたシステムに他の構成要素が追加されたり、またはシステムから除去されてよい。したがって、他の実装は、以下の特許請求の範囲内にある。

50

フロントページの続き

- (72)発明者 モレノ メンヒバル、ペドロ ジェイ .
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
エイ 1600
- (72)発明者 ベリコビッチ、レオニード
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
エイ 1600
- 審査官 山下 剛史
- (56)参考文献 米国特許出願公開第2018/0053502 (US, A1)
国際公開第2007/108500 (WO, A1)
特表2017-511899 (JP, A)
特開2013-125144 (JP, A)
特開2017-139004 (JP, A)
特表2011-513795 (JP, A)
- (58)調査した分野 (Int.Cl., DB名)
G10L 15/00 - 15/34
IEEE Xplore