(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7549061号
(P7549061)

(45)発行日　令和6年9月10日(2024.9.10)

(24)登録日　令和6年9月2日(2024.9.2)

(51)国際特許分類

Ｇ１０Ｌ　15/19　(2013.01)		Ｇ１０Ｌ	15/19	
Ｇ１０Ｌ　15/22　(2006.01)		Ｇ１０Ｌ	15/22	３００Ｕ

請求項の数　20　（全23頁）

(21)出願番号　特願2023-21323(P2023-21323)
(22)出願日　令和5年2月15日(2023.2.15)
(62)分割の表示　特願2021-531511(P2021-531511)の分割
原出願日　令和1年11月27日(2019.11.27)
(65)公開番号　特開2023-53331(P2023-53331A)
(43)公開日　令和5年4月12日(2023.4.12)
審査請求日　令和5年2月15日(2023.2.15)
(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)

(73)特許権者　502208397
グーグル エルエルシー
Ｇｏｏｇｌｅ ＬＬＣ
アメリカ合衆国 カリフォルニア州 ９４０４３ マウンテン ビュー アンフィシアター パークウェイ １６００
１６００ Ａｍｐｈｉｔｈｅａｔｒｅ Ｐａｒｋｗａｙ ９４０４３ Ｍｏｕｎｔａｉｎ Ｖｉｅｗ， ＣＡ Ｕ．Ｓ．Ａ．
(74)代理人　100142907
弁理士　本田 淳
(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

最終頁に続く

(54)【発明の名称】　音声入力処理

(57)【特許請求の範囲】

【請求項１】

コンピュータにより実装される方法であって、データ処理ハードウェアによって実行されるとき、

ユーザに関連付けられているコンピューティングデバイスが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ことを示す文脈を判定することと、

ユーザによって行われて前記コンピューティングデバイスによってキャプチャされた発話のオーディオデータを受信することと、

前記オーディオデータから、前記発話に対応する複数の音素を判定することと、

前記発話に対応する前記複数の音素を用いて、前記発話についての１つ以上の候補転写を生成することであって、前記１つ以上の候補転写の各々は対応する転写信頼スコアを有する、生成することと、

前記コンピューティングデバイスが前記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ことを示す前記文脈に基づいて、複数の文法から文法を選択することと、

最高の前記転写信頼スコアを有する前記１つ以上の候補転写のうちの候補転写について、選択された前記文法を用いてパースを行って、前記コンピューティングデバイスが実行すべきアクションを識別することと、

を含む動作を前記データ処理ハードウェアに実行させ、

前記複数の文法は、互いに異なった、複数の用語からなる予め指定された構造を含む、方法。

【請求項２】
前記動作は、識別された前記アクションを実行するよう前記コンピューティングデバイスに命令することをさらに含む、
請求項１に記載の方法。
【請求項３】
前記複数の文法からアラームの設定に用いられるアラーム文法が選択される可能性は、前記コンピューティングデバイスがアラームアプリケーションを実行している場合、前記コンピューティングデバイスが前記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場合と比べて高い、
請求項１に記載の方法。
【請求項４】
前記文脈が、前記コンピューティングデバイスが近くのデバイスと通信していることを含む、
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は、携帯電話またはウェアラブルデバイスを含む、
請求項１に記載の方法。
【請求項６】
前記文法は、デフォルト文法を含む、
請求項１に記載の方法。
【請求項７】
前記１つ以上の候補転写の各々は複数の用語を含む、
請求項１に記載の方法。
【請求項８】
前記複数の用語の各々は、対応する用語信頼スコアを含む、
請求項７に記載の方法。
【請求項９】
前記転写信頼スコアは、前記候補転写が前記ユーザによって行われた前記発話と一致する可能性を示している、
請求項１に記載の方法。
【請求項１０】
前記データ処理ハードウェアは、前記コンピューティングデバイス上にある、
請求項１に記載の方法。
【請求項１１】
システムであって、
データ処理ハードウェアと、
前記データ処理ハードウェアと通信するメモリハードウェアと、を備え、前記メモリハードウェアは命令を記憶しており、前記命令は前記データ処理ハードウェア上において実行されるとき、
ユーザに関連付けられているコンピューティングデバイスが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ことを示す文脈を判定することと、
ユーザによって行われて前記コンピューティングデバイスによってキャプチャされた発話のオーディオデータを受信することと、
前記オーディオデータから、前記発話に対応する複数の音素を判定することと、
前記発話に対応する前記複数の音素を用いて、前記発話についての１つ以上の候補転写を生成することであって、前記１つ以上の候補転写の各々は対応する転写信頼スコアを有する、生成することと、
前記コンピューティングデバイスが前記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ことを示す前記文脈に基づいて複数の文法から文法を選択することと、
最高の前記転写信頼スコアを有する前記１つ以上の候補転写のうちの候補転写について

、選択された前記文法を用いてパースを行って、前記コンピューティングデバイスが実行すべきアクションを識別することと、
を含む動作を前記データ処理ハードウェアに実行させ、
前記複数の文法は、互いに異なった、複数の用語からなる予め指定された構造を含む、システム。
【請求項１２】
前記動作は、識別された前記アクションを実行するよう前記コンピューティングデバイスに命令することをさらに含む、
請求項１１に記載のシステム。
【請求項１３】
前記複数の文法からアラームの設定に用いられるアラーム文法が選択される可能性は、前記コンピューティングデバイスがアラームアプリケーションを実行している場合、前記コンピューティングデバイスが前記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場合と比べて高い、
請求項１１に記載のシステム。
【請求項１４】
前記文脈が、前記コンピューティングデバイスが近くのデバイスと通信していることを含む、
請求項１１に記載のシステム。
【請求項１５】
前記コンピューティングデバイスは、携帯電話またはウェアラブルデバイスを含む、
請求項１１に記載のシステム。
【請求項１６】
前記文法は、デフォルト文法を含む、
請求項１１に記載のシステム。
【請求項１７】
前記１つ以上の候補転写の各々は複数の用語を含む、
請求項１１に記載のシステム。
【請求項１８】
前記複数の用語の各々は、対応する用語信頼スコアを含む、
請求項１７に記載のシステム。
【請求項１９】
前記転写信頼スコアは、前記候補転写が前記ユーザによって行われた前記発話と一致する可能性を示している、
請求項１１に記載のシステム。
【請求項２０】
前記データ処理ハードウェアは、前記コンピューティングデバイス上にある、
請求項１１に記載のシステム。
【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。
【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】
ユーザが音声を介し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。
【０００５】
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。
【０００６】
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。
【０００７】
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。

【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】

【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。

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。

【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。

【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。

【００１６】

一部の実装形態では、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、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は、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。

【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。

【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、バッカス・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）」と言うことができることを示す、アラームシンタックスを＄ＡＬＡＲＭ＝ｓｅｔ＄ＭＯＤＥｆｏｒ＄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ので、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（ｌｉｇｈｔｓ　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（ｌｉｇｈｔｓ　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（ｌｉｇｈｔｓ　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（ｌｉｇｈｔｓ　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（ｌｉｇｈｔｓ　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。

【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（ｆｌｉｇｈｔｓ　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（ｆｌｉｇｈｔｓ　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。

【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（ｌｉｇｈｔｓ　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（ｌｉｇｈｔｓ　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（ｌｉｇｈｔｓ　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（ｌｉｇｈｔｓ　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。「フライツアウト（ｆｌｉｇｈｔｓ　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「フライツアウト（ｆｌｉｇｈｔｓ　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「フライツアウト（ｆｌｉｇｈｔｓ　ｏｕｔ）」を提供する結果となる可能性がある。

【００２５】

インターネットで「フライツアウト（ｆｌｉｇｈｔｓ　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。

【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。

【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、組み合わさ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。

【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。

【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、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＝０．１６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。

【００３１】

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「フライツアウト（ｆｌｉｇｈｔｓ　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。

【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。

【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（ｌｉｇｈｔｓ　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（ｌｉｇｈｔｓ　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（Ｌｉｇｈｔｓ　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の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。一部の実装形態では、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合がある。例えば、候補転写は「時間をセット（ｓｅｔ　ｔｉｍｅ）」であり、０．７５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（ｓｅｔ　ｃｈｉｍｅ）」であり、０．２５の転写信頼スコアを有する可能性がある。

【００４０】

音声認識部２２５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２４０は、システム２００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２４０は、システムの任意のセンサからセンサデータ２４５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、ジャイロスコープ、重力センー、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２４５はまた、システム２００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２００のバッテリレベル、信号強度、システム２００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２００の他の同様の状態を含み得る。

【００４１】

文脈特定部２４０はまた、システム２００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２５０を収集することができる。プロセスデータ２５０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２００が実行しているアプリケーション、システム２００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。

【００４２】

例として、文脈特定部２４０は、システム２００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２５０およびセンサデータ２４５を受信することができる。

【００４３】

文法スコア生成部２５５は、文脈特定部２４０からシステム２００の文脈を受信し、音声認識部２２５から単語ラティスを受信する。文法スコア生成部２５５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２６０を特定する。場合によっては、文法２６０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２５５は、文法２６０のいずれによってもパース（ｐａｒａｂｌｅ）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。

【００４４】

一部の例では、単一の文法２６０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２５５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２５５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。

【００４５】

一部の例では、複数の文法２６０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２５５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２５５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。

【００４６】

文法スコア正規化部２６５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２６５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２５５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２６５は、当該文法信頼スコアを１．０に増やす。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。

【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。

【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る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。

【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。一部の実装形態では、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。

【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。

【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。

【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。

【００５５】

システムは、システムの文脈を判定する（３３０）。一部の実装形態では、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様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。

【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（ｌｉｇｈｔｓ　ｏｕｔ）」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）。一部の実装形態では、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、組み合わされた信頼スコアが最も高い文法と転写を選択する。

【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）。一部の実装形態では、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。

【００６０】

一部の実装形態では、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。一部の実装形態では、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。一部の実装形態では、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】

一部の実装形態では、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。一部の実装形態では、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。

【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。

【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量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。

【００６５】

一部の実装形態では、プロセス３００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０．３２、映画コマンド０．２５、音楽コマンド０．２０、一般検索０．１０、オンラインビジネス０．０３、社会集団０．０３、およびオンライン百科事典検索０．０２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、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１０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０．０３２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。

【００６６】

プロセス３００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３００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３００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３００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。

【００６７】

一部の実装形態では、プロセス３００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、タガー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得られる可能性がある。上位Ｎ（Ｎ＝１００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ので、前述の最良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。

【００６８】

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４００およびモバイルコンピューティング装置４５０の例を示す。コンピューティング装置４００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４５０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。

【００６９】

コンピューティング装置４００は、プロセッサ４０２、メモリ４０４、記憶装置４０６、メモリ４０４および複数の高速拡張ポート４１０に接続する高速インタフェース４０８、ならびに低速拡張ポート４１４および記憶装置４０６に接続する低速インタフェース４１２を含む。プロセッサ４０２、メモリ４０４、記憶装置４０６、高速インタフェース４０８、高速拡張ポート４１０、および低速インタフェース４１２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４０２は、高速インタフェース４０８に接続されたディスプレイ４１６等の外部入力／出力装置にＧＵＩのグラフィック情報を表示するために、メモリ４０４または記憶装置４０６に記憶された命令を含む、コンピューティング装置４００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）。

【００７０】

メモリ４０４は、コンピューティング装置４００内の情報を記憶する。いくつかの実装においては、メモリ４０４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４０４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４０４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。

【００７１】

記憶装置４０６は、コンピューティング装置４００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４０６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４０２）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４０４、記憶装置４０６、またはプロセッサ４０２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。

【００７２】

高速インタフェース４０８は、コンピューティング装置４００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４１２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４０８は、メモリ４０４、ディスプレイ４１６（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４１０に接続される。実装において、低速インタフェース４１２は、記憶装置４０６および低速拡張ポート４１４に接続される。低速拡張ポート４１４は、種々の通信ポート（例えば、ＵＳＢ、Ｂｌｕｅｔｏｏｔｈ（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。

【００７３】

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されてよい。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スス、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。

【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（エンハンスト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。

【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。

【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。

【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。

【００８３】

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、ＰＬＤ（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。

【００８４】

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、ＣＲＴ（陰極線管）またはＬＣＤ（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、ユーザからの入力は、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。

【００８５】

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することが可能な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、ＬＡＮ（ローカルエリアネットワーク）、ＷＡＮ（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。一部の実装形態では、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。

【００８６】

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。

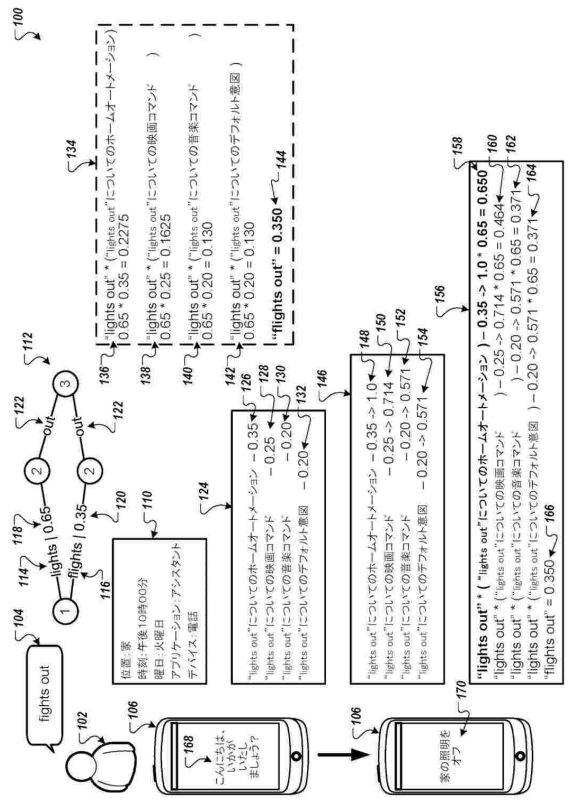
【００８７】

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

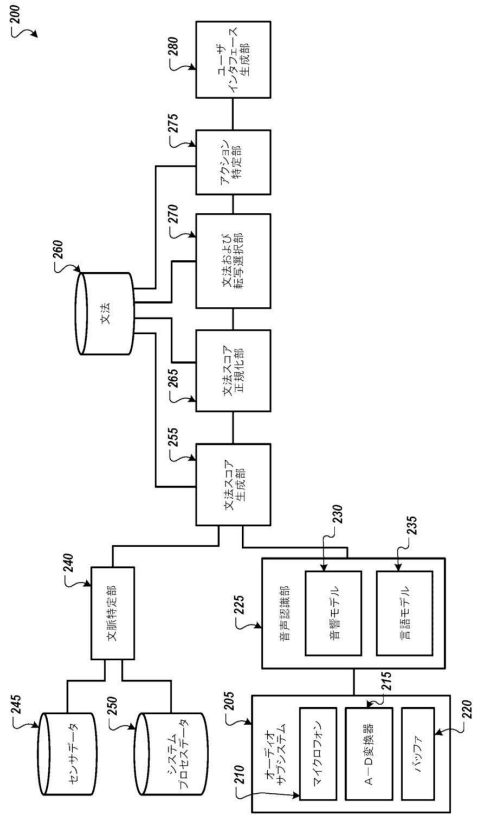
アプリケーションはデリゲートにアクセスするものとして説明されているが、他の実装では、デリゲートは、１つまたは複数のサーバで実行されるアプリケーション等、１つまたは複数のプロセッサによって実装された他のアプリケーションによって使用されてよい。さらに、図に示された論理の流れは、望ましい結果を得るために、示された特定の順序または連続した順序を必要とはしない。さらに、説明された流れに他の動作が提供されたり、または流れから除去されてよく、説明されたシステムに他の構成要素が追加されたり、またはシステムから除去されてよい。したがって、他の実装は、以下の特許請求の範囲内にある。

【図面】

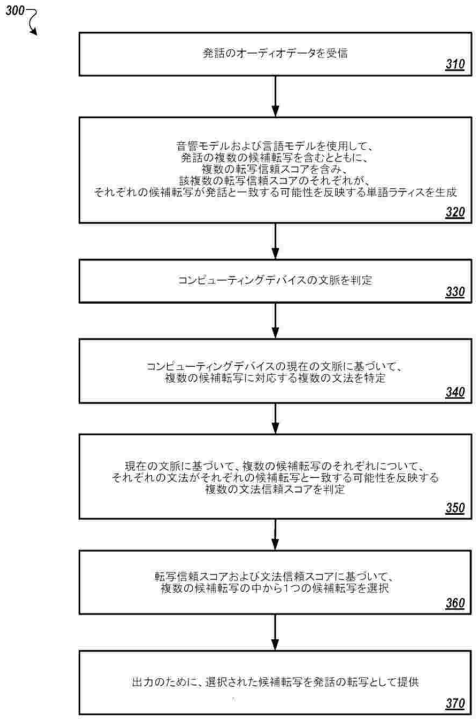
【図１】



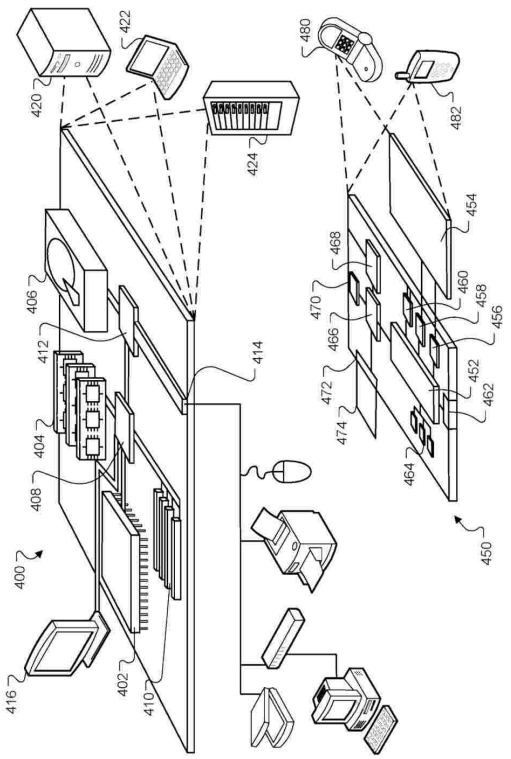
【図２】



【図３】


300
発話のオーディオデータを受信
310
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
320
コンピューティングデバイスの文脈を判定
330
コンピューティングデバイスの現在の文脈に基づいて、
複数の候補転写に対応する複数の文法を特定
340
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
350
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
360
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
370

【図４】


420
422
424
480
482
454
468
460
458
456
470
466
452
462
472
474
464
450
406
412
414
404
408
400
416
402
410

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 パークウェイ 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 パークウェイ １６００
審査官　山下 剛史
(56)参考文献　特表２０１３－５１０３４１（ＪＰ，Ａ）
特開２０００－２９３１９６（ＪＰ，Ａ）
特開２００１－５４８９（ＪＰ，Ａ）
米国特許出願公開第２０１８／００５３５０２（ＵＳ，Ａ１）
特開２００１－１５７１３７（ＪＰ，Ａ）
特開２００３－９１２９８（ＪＰ，Ａ）
特開２０１８－１８２６９２（ＪＰ，Ａ）
特表２０１８－５３３０３６（ＪＰ，Ａ）
(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ １５／００－１５／３４
ＩＥＥＥ Ｘｐｌｏｒｅ