(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6727315号
(P6727315)
(45)発行日　令和2年7月22日(2020.7.22)　(24)登録日　令和2年7月2日(2020.7.2)

(51)Int.Cl.		F I		
G06F 40/56	(2020.01)	G06F	40/56	
G06F 40/20	(2020.01)	G06F	40/20	
G10L 13/08	(2013.01)	G10L	13/08	124
G10L 13/00	(2006.01)	G10L	13/00	100M

請求項の数 29　(全 29 頁)

(21)出願番号　特願2018-539396(P2018-539396)
(86)(22)出願日　平成28年12月29日(2016.12.29)
(65)公表番号　特表2019-511034(P2019-511034A)
(43)公表日　平成31年4月18日(2019.4.18)
(86)国際出願番号　PCT/US2016/069182
(87)国際公開番号　WO2017/131924
(87)国際公開日　平成29年8月3日(2017.8.3)
審査請求日　平成30年9月25日(2018.9.25)
(31)優先権主張番号　15/009,432
(32)優先日　平成28年1月28日(2016.1.28)
(33)優先権主張国・地域又は機関
米国(US)

前置審査

(73)特許権者　502208397
グーグル　エルエルシー
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
(74)代理人　100108453
弁理士　村山　靖彦
(74)代理人　100110364
弁理士　実広　信哉
(74)代理人　100133400
弁理士　阿部　達彦
(72)発明者　マシュー・シャリフィ
スイス・８００２・チューリッヒ・ブランドシェンケシュトラーセ・１１０

最終頁に続く

(54)【発明の名称】適応的テキスト－音声出力

(57)【特許請求の範囲】
【請求項１】
1つまたは複数のコンピュータによって実施される方法であって、
前記1つまたは複数のコンピュータによって、クライアントデバイスのユーザの言語習熟度を判定するステップと、
前記1つまたは複数のコンピュータによって、前記ユーザの前記判定された言語習熟度に基づいて、テキスト-音声モジュールによる出力用のテキストセグメントを決定するステップであって、テキストセグメントを決定するステップが、前記ユーザのテキスト－音声出力用の候補として複数のテキストセグメントを識別するステップと、異なるレベルの言語複雑度を示す言語複雑度スコアを各々が有する複数のテキストセグメントの中から、前記クライアントデバイスの前記ユーザの前記判定された言語習熟度を記述する基準スコアと最もよく一致する言語複雑度スコアを有するテキストセグメントを選択するステップとを含み、前記複数のテキストセグメントの中から選択するステップが、前記複数のテキストセグメントのそれぞれについての言語複雑度スコアを判定するステップを含む、ステップと、
前記1つまたは複数のコンピュータによって、前記テキストセグメントの合成された発声を含むオーディオデータを生成するステップと、
前記1つまたは複数のコンピュータによって前記クライアントデバイスに、前記テキストセグメントの前記合成された発声を含む前記オーディオデータを提供するステップと
を含む、方法。

【請求項２】
前記クライアントデバイスは、テキスト－音声インターフェースを使用するモバイルアプリケーションを表示する、請求項1に記載の方法。
【請求項３】
前記ユーザの前記言語習熟度を判定するステップは、前記ユーザによって提出された以前のクエリに少なくとも基づいて前記ユーザの言語習熟度を推論するステップを含む、請求項1または2に記載の方法。
【請求項４】
前記テキスト-音声モジュールによる出力用の前記テキストセグメントを決定するステップは、
前記ユーザへのテキスト－音声出力用のテキストセグメントを識別するステップと、
前記テキスト－音声出力用の前記テキストセグメントの複雑度スコアを計算するステップと、
前記ユーザの前記判定された言語習熟度および前記テキスト－音声出力用の前記テキストセグメントの前記複雑度スコアに少なくとも基づいて、前記ユーザへの前記テキスト－音声出力用の前記テキストセグメントを修正するステップと
を含む、請求項1から3のいずれか一項に記載の方法。
【請求項５】
前記ユーザへの前記テキスト－音声出力用の前記テキストセグメントを修正するステップは、
前記ユーザの前記判定された言語習熟度に少なくとも基づいて、前記ユーザについての全体的複雑度スコアを判定するステップと、
前記ユーザへの前記テキスト－音声出力用の前記テキストセグメント内の個々の部分についての複雑度スコアを判定するステップと、
前記ユーザについての前記全体的複雑度スコアよりも大きい複雑度スコアを有する前記テキストセグメント内の1つまたは複数の個々の部分を識別するステップと、
前記全体的複雑度スコアを下回るまで複雑度スコアを低減するように、前記テキストセグメント内の前記1つまたは複数の個々の部分を修正するステップと
を含む、請求項4に記載の方法。
【請求項６】
前記ユーザへの前記テキスト－音声出力用の前記テキストセグメントを修正するステップは、
前記ユーザに関連付けられたコンテキストを示すデータを受信するステップと、
前記ユーザに関連付けられた前記コンテキストについての全体的複雑度スコアを判定するステップと、
前記テキストセグメントの前記複雑度スコアが、前記ユーザに関連付けられた前記コンテキストについての前記全体的複雑度スコアを超えると判定するステップと、
前記ユーザに関連付けられた前記コンテキストについての前記全体的複雑度スコアを下回るまで前記複雑度スコアを低減するように、前記テキストセグメントを修正するステップと
を含む、請求項4に記載の方法。
【請求項７】
システムであって、
1つまたは複数のコンピュータと、
前記1つまたは複数のコンピュータに結合される非一時的コンピュータ可読記録媒体であって、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き、前記1つまたは複数のコンピュータに、
前記1つまたは複数のコンピュータによって、クライアントデバイスのユーザの言語習熟度を判定することと、
前記1つまたは複数のコンピュータによって、前記ユーザの前記判定された言語習熟度

に基づいて、テキスト-音声モジュールによる出力用のテキストセグメントを決定することであって、テキストセグメントを決定することが、前記ユーザのテキスト－音声出力用の候補として複数のテキストセグメントを識別することと、異なるレベルの言語複雑度を示す言語複雑度スコアを各々が有する複数のテキストセグメントの中から、前記クライアントデバイスの前記ユーザの前記判定された言語習熟度を記述する基準スコアと最もよく一致する言語複雑度スコアを有するテキストセグメントを選択することとを備え、前記複数のテキストセグメントの中から選択することが、前記複数のテキストセグメントのそれぞれについての言語複雑度スコアを判定することを含む、ことと、

前記1つまたは複数のコンピュータによって、前記テキストセグメントの合成された発声を含むオーディオデータを生成することと、

前記1つまたは複数のコンピュータによって前記クライアントデバイスに、前記テキストセグメントの前記合成された発声を含む前記オーディオデータを提供することと

を含む動作を実施させるように記憶された命令を有する、非一時コンピュータ可読媒体と

を備える、システム。

【請求項８】

前記クライアントデバイスは、テキスト－音声インターフェースを使用するモバイルアプリケーションを表示する、請求項7に記載のシステム。

【請求項９】

前記ユーザの前記言語習熟度を判定することは、前記ユーザによって提出された以前のクエリに少なくとも基づいて前記ユーザの言語習熟度を推論することを含む、請求項7または8に記載のシステム。

【請求項１０】

前記テキスト-音声モジュールによる出力用の前記テキストセグメントを決定することは、

前記ユーザへのテキスト－音声出力用のテキストセグメントを識別することと、

前記テキスト－音声出力用の前記テキストセグメントの複雑度スコアを計算することと、

前記ユーザの前記判定された言語習熟度および前記テキスト－音声出力用の前記テキストセグメントの前記複雑度スコアに少なくとも基づいて、前記ユーザへの前記テキスト－音声出力用の前記テキストセグメントを修正することと

を含む、請求項7から9のいずれか一項に記載のシステム。

【請求項１１】

前記ユーザへの前記テキスト－音声出力用の前記テキストセグメントを修正することは、

前記ユーザの前記判定された言語習熟度に少なくとも基づいて、前記ユーザについての全体的複雑度スコアを判定することと、

前記ユーザへの前記テキスト－音声出力用の前記テキストセグメント内の個々の部分についての複雑度スコアを判定することと、

前記ユーザについての前記全体的複雑度スコアよりも大きい複雑度スコアを有する前記テキストセグメント内の1つまたは複数の個々の部分を識別することと、

前記全体的複雑度スコアを下回るまで複雑度スコアを低減するように、前記テキストセグメント内の前記1つまたは複数の個々の部分を修正することと

を含む、請求項10に記載のシステム。

【請求項１２】

1つまたは複数のコンピュータによって実施される方法であって、

ユーザに関連付けられたコンテキストを示すデータを受信するステップと、

前記ユーザに関連付けられた前記コンテキストについての全体的複雑度スコアを判定するステップであって、前記ユーザが前記コンテキストにあると判定されたとき、前記ユーザによって以前に提出されたクエリ内に含まれる用語を識別するステップと、前記識別さ

れた用語に少なくとも基づいて、前記ユーザに関連付けられた前記コンテキストについての全体的複雑度スコアを判定するステップとを含む、ステップと、
前記ユーザへのテキスト－音声出力用のテキストセグメントを識別するステップと、
前記テキストセグメントの複雑度スコアが、前記ユーザに関連付けられた前記コンテキストについての前記全体的複雑度スコアを超えると判定するステップと、
前記ユーザに関連付けられた前記コンテキストについての前記全体的複雑度スコアを下回るまで前記複雑度スコアを低減するように、前記テキストセグメントを修正するステップと
を含む、方法。
【請求項１３】
前記ユーザに関連付けられた前記コンテキストを示す前記データは、前記ユーザによって以前に提出されたクエリを含む、請求項12に記載の方法。
【請求項１４】
前記ユーザに関連付けられた前記コンテキストを示す前記データは、前記ユーザに関連付けられた現在のロケーションを示すGPS信号を含む、請求項12または13に記載の方法。
【請求項１５】
前記ユーザに関連付けられた前記コンテキストを示すデータは、前記ユーザのモバイルデバイスからのセンサデータを含む、請求項12から14のいずれか一項に記載の方法。
【請求項１６】
1つまたは複数の処理デバイスによって実施される方法であって、
前記1つまたは複数の処理デバイスによって、デバイスへの音声入力の複雑度レベルを判定するステップと、
前記1つまたは複数の処理デバイスによって、前記音声入力に応答して出力用のメッセージを決定するステップであって、前記メッセージは、前記音声入力に関連付けられた前記判定された複雑度レベルに基づいて決定され、前記メッセージは、前記判定された複雑度レベルに基づいて、出力用のベースラインメッセージの複雑度レベルを変化させることによって調整される、ステップと、
前記1つまたは複数の処理デバイスによって、前記メッセージの合成された発声を含むオーディオデータを生成するステップと、
前記1つまたは複数の処理デバイスによって、前記音声入力に応答して、出力用の前記合成された発声を含む前記オーディオデータを提供するステップと
を含み、
前記出力用のメッセージを決定するステップは、
前記音声入力に応答して、出力用のベースラインメッセージを取得するステップと、
前記デバイスへの前記音声入力についての前記判定された複雑度レベルに基づいて、前記ベースラインメッセージの複雑度レベルを増大することによって、調整されたメッセージを生成するステップと
を含む、方法。
【請求項１７】
前記音声入力の前記複雑度レベルは、前記音声入力の言語複雑度を含む、請求項16に記載の方法。
【請求項１８】
前記デバイスに前記音声入力を提出したユーザの言語習熟度を判定するステップをさらに含み、
前記メッセージを決定するステップは、前記デバイスに前記音声入力を提出した前記ユーザの前記言語習熟度に基づく、請求項16または17に記載の方法。
【請求項１９】
前記出力用のメッセージを決定するステップは、
前記音声入力に応答して、出力用のベースラインメッセージを取得するステップと、
前記デバイスへの前記音声入力についての前記判定された複雑度レベルに基づいて、前

記ベースラインメッセージの複雑度レベルを低下することによって、調整されたメッセージを生成するステップと
を含む、請求項16に記載の方法。
【請求項２０】
前記デバイスは、テキスト－音声インターフェースを使用するモバイルアプリケーションを稼動する、請求項16に記載の方法。
【請求項２１】
1つまたは複数のコンピュータによって実施される方法であって、
前記1つまたは複数のコンピュータによって、(i)特定の音声入力が第1のユーザによって提供され、(ii)前記特定の音声入力が、前記第1のユーザとは異なる第2のユーザによって提供されたことを示すデータを取得するステップと、
前記1つまたは複数のコンピュータによって、(i)前記第1のユーザについての第1の言語習熟度スコアおよび(ii)前記第2のユーザについての第2の言語習熟度スコアを判定するステップであって、前記第1の言語習熟度スコアは、前記第2の言語習熟度スコアよりも高いレベルの言語習熟度を示す、ステップと、
前記1つまたは複数のコンピュータによって、(i)前記第1の言語習熟度スコアに基づく第1のメッセージの合成された発声を含む第1のオーディオデータおよび(ii)前記第2の言語習熟度スコアに基づく第2のメッセージの合成された発声を含む第2のオーディオデータを生成するステップであって、前記第1のメッセージは、前記第2のメッセージよりも高い言語複雑度を有し、前記第1のオーディオデータを生成するステップは、前記第1の言語習熟度スコアに基づいて前記第1のメッセージのテキストを決定するステップを含み、前記第1のメッセージのテキストを決定するステップが、前記第1のユーザのテキスト－音声出力用の候補として複数のテキストセグメントを識別するステップと、異なるレベルの言語複雑度を示す言語複雑度スコアを各々が有する複数のテキストセグメントの中から、前記第1のユーザの前記判定された言語習熟度を記述する基準スコアと最もよく一致する言語複雑度スコアを有するテキストセグメントを選択するステップとを備え、前記複数のテキストセグメントの中から選択するステップが、前記複数のテキストセグメントのそれぞれについての言語複雑度スコアを判定するステップを含む、ステップと、
前記1つまたは複数のコンピュータによって、(i)前記特定の音声入力に応答して前記第1のユーザのクライアントデバイスに前記第1のオーディオデータを、および(ii)前記特定の音声入力に応答して前記第2のユーザのクライアントデバイスに前記第2のオーディオデータを提供するステップと
を含む、方法。
【請求項２２】
前記第2のオーディオデータを生成するステップは、前記第2の言語習熟度スコアに基づいて前記第2のメッセージのテキストを決定するステップを含む、請求項21に記載の方法。
【請求項２３】
前記第1のユーザについての前記第1の言語習熟度スコアおよび前記第2のユーザについての前記第2の言語習熟度スコアを判定するステップは、前記第1のユーザおよび前記第2のユーザによって提出されたそれぞれの以前のクエリに少なくとも基づいて、前記第1のユーザおよび第2のユーザのそれぞれの言語習熟度を推論するステップを含む、請求項21または22に記載の方法。
【請求項２４】
前記第1のオーディオデータを生成するステップは、
前記第1のユーザへのテキスト－音声出力用のテキストセグメントを識別するステップと、
前記テキストセグメントの複雑度スコアを計算するステップと、
前記第1のユーザの前記第1の言語習熟度スコアおよび前記テキスト－音声出力用の前記テキストセグメントの前記複雑度スコアに少なくとも基づいて、前記第1のユーザへの前

記テキスト－音声出力用の前記テキストセグメントを修正するステップと
を含む、請求項21から23のいずれか一項に記載の方法。
【請求項２５】
前記第1のユーザへの前記テキスト－音声出力用の前記テキストセグメントを修正するステップは、
前記第1のユーザの第1の言語習熟度スコアに少なくとも基づいて、前記第1のユーザについての全体的複雑度スコアを判定するステップと、
前記第1のユーザへの前記テキスト－音声出力用の前記テキストセグメント内の個々の部分についての複雑度スコアを判定するステップと、
前記第1のユーザについての前記全体的複雑度スコアよりも大きい複雑度スコアを有する前記テキストセグメント内の1つまたは複数の個々の部分を識別するステップと、
前記全体的複雑度スコアを下回るまで複雑度スコアを低減するように、前記テキストセグメント内の前記1つまたは複数の個々の部分を修正するステップと
を含む、請求項24に記載の方法。
【請求項２６】
前記第1のユーザへの前記テキスト－音声出力用の前記テキストセグメントを修正するステップは、
前記第1のユーザに関連付けられたコンテキストを示すデータを受信するステップと、
前記第1のユーザに関連付けられた前記コンテキストについての全体的複雑度スコアを判定するステップと、
前記テキストセグメントの前記複雑度スコアが、前記第1のユーザに関連付けられた前記コンテキストについての前記全体的複雑度スコアを超えると判定するステップと、
前記第1のユーザに関連付けられた前記コンテキストについての前記全体的複雑度スコアを下回るまで前記複雑度スコアを低減するように、前記テキストセグメントを修正するステップと
を含む、請求項24または25に記載の方法。
【請求項２７】
前記第1のオーディオデータおよび前記第2のオーディオデータを提供するステップは、
前記1つまたは複数のコンピュータによって、(i)コンピュータネットワークを介して前記第1のユーザの前記クライアントデバイスに前記第1のオーディオデータを、および(ii)コンピュータネットワークを介して前記第2のユーザの前記クライアントデバイスに前記第2のオーディオデータを提供するステップ
を含む、請求項21から26のいずれか一項に記載の方法。
【請求項２８】
システムであって、1つまたは複数の処理デバイスと、前記1つまたは複数の処理デバイスによって実施されるとき、前記システムに請求項1～6および請求項12～27のいずれか一項に記載の方法を実施させる命令を記憶する1つまたは複数の機械可読ストレージデバイスとを備える、システム。
【請求項２９】
1つまたは複数の処理デバイスによって実施されるとき、前記1つまたは複数の処理デバイスに請求項1～6および請求項12～27のいずれか一項に記載の方法を実施される命令を記憶する、１つまたは複数の機械可読ストレージデバイス。
【発明の詳細な説明】
【技術分野】
【０００１】
関連出願の相互参照
本出願は、参照によりその全体が本明細書に組み込まれる、2016年1月28日に出願され、"ADAPTIVE TEXT-TO-SPEECH-OUTPUTS"と題する米国特許出願第15/009,432号に対する優先権を主張する。
【０００２】

本明細書は概して、電子通信について記載する。

【背景技術】

【０００３】

音声合成は、人間音声を人工的に作り出すことを指す。音声合成器は、テキストに対応する音声出力を生成するためのソフトウェアまたはハードウェア構成要素で実装することができる。たとえば、テキスト－音声(TTS)システムが一般に、データベースに記憶されている、記録された音声の断片を連結することによって通常の言語テキストを音声に変換する。

【発明の概要】

【課題を解決するための手段】

【０００４】

音声合成は、電子コンピューティングのより大部分がデスクトップからモバイル環境にシフトしたのに従い、ユーザエクスペリエンスにとってより中心的になっている。たとえば、ディスプレイなしの比較的小さいモバイルデバイスの使用の増加が、モバイルデバイス上に表示されるコンテンツにアクセスし、使用するためのテキスト－音声(TTS)システムの使用の増加につながっている。

【０００５】

本明細書は、改善されたユーザインターフェース、詳細には、改善されたTTSを通した、より良好な、コンピュータとユーザ間の通信を開示する。

【０００６】

既存のTTSシステムに伴う1つの特定の課題は、そのようなシステムがしばしば、異なるユーザの変動する言語習熟度に適応することができないことである。この柔軟性の欠如がしばしば、限られた言語習熟度を有するユーザが複雑なテキスト－音声出力を理解するのを妨げる。たとえば、TTSシステムを使用する非ネイティブ言語話者は、限られた言語精通のせいで、テキスト－音声出力を理解するのに問題を有する可能性がある。既存のTTSシステムに伴う別の課題は、テキスト－音声出力を理解するためのユーザの瞬間的能力も、特定のユーザコンテキストに基づいて変動する可能性があることである。たとえば、いくつかのユーザコンテキストは、比較的長い、または比較的複雑なテキスト－音声出力をより理解しにくくするバックグラウンドノイズを含む。

【０００７】

いくつかの実装形態において、システムが、ユーザの言語習熟度に基づいて、テキスト－音声出力に使用されるテキストをユーザがテキスト－音声出力を把握することができる見込みを増大するように調整する。たとえば、ユーザの言語習熟度は、先行ユーザアクティビティから推論し、テキスト－音声出力をユーザの言語習熟度と釣り合う適切な複雑度に調整するために使用することができる。いくつかの例では、システムが、異なるレベルの言語習熟度に対応する複数の候補テキストセグメントを取得する。システムは次いで、ユーザの言語習熟度に最も良く一致するとともに最も厳密に対応する候補テキストセグメントを選択し、選択されたテキストセグメントの合成された発声をユーザへの出力用に提供する。他の例では、システムが、テキストセグメント中のテキストをテキスト－音声出力を生成するのに先立って、ユーザの言語習熟度により良好に対応するように改変する。ボキャブラリ、文構造、長さなどを含むテキストセグメントの様々な側面が調整されることが可能である。システムは次いで、改変されたテキストセグメントの合成された発声をユーザへの出力用に提供する。

【０００８】

ここで議論されるシステムがユーザについての個人情報を収集するか、または個人情報を利用する場合がある状況のために、ユーザには、プログラムまたは特徴が個人情報、たとえば、ユーザのソーシャルネットワーク、ソーシャルアクションもしくはアクティビティ、職業、ユーザの嗜好、またはユーザの現在のロケーションについての情報を収集するかどうかを制御するための、あるいはユーザにより関連する場合があるコンテンツをコンテンツサーバから受信するかどうか、および/またはどのようにして受信するかを制御す

るための機会が提供されてもよい。さらに、特定のデータが、記憶され、または使用される前に、個人的に識別可能な情報が削除されるように1つまたは複数のやり方で匿名化されてもよい。たとえば、ユーザのアイデンティティが、ユーザについての個人を識別できる情報が判定されることが可能でないように匿名化されてもよく、またはロケーション情報が取得される場合は、ユーザの特定のロケーションを判定することができないように、ユーザの地理的ロケーションが、たとえば、市、郵便番号、または州レベルに一般化されてもよい。したがって、ユーザは、情報がどのように、自分について収集され、コンテンツサーバによって使用されるかに対する制御を有してもよい。

【０００９】

一態様では、コンピュータ実装方法は、1つまたは複数のコンピュータによって、クライアントデバイスのユーザの言語習熟度を判定するステップと、1つまたは複数のコンピュータによって、ユーザの判定された言語習熟度に基づいて、テキスト－音声モジュールによる出力用のテキストセグメントを決定するステップと、1つまたは複数のコンピュータによって、テキストセグメントの合成された発声を含むオーディオデータを生成するステップと、1つまたは複数のコンピュータによって、およびクライアントデバイスに、テキストセグメントの合成された発声を含むオーディオデータを提供するステップとを含むことができる。

【００１０】

他のバージョンは、対応するシステムと、コンピュータ記憶デバイス上で符号化された、方法のアクションを実施するように構成されたコンピュータプログラムとを含む。

【００１１】

1つまたは複数の実装形態は、以下のオプションの特徴を含むことができる。たとえば、いくつかの実装形態において、クライアントデバイスは、テキスト－音声インターフェースを使用するモバイルアプリケーションを表示する。

【００１２】

いくつかの実装形態において、ユーザの言語習熟度を判定することは、ユーザによって提出された以前のクエリに少なくとも基づいてユーザの言語習熟度を推論することを含む。

【００１３】

いくつかの実装形態において、テキスト－音声モジュールによる出力用のテキストセグメントを決定することは、ユーザのテキスト－音声出力用の候補として複数のテキストセグメントを識別すること、複数のテキストセグメントは異なるレベルの言語複雑度を有すること、およびクライアントデバイスのユーザの判定された言語習熟度に少なくとも基づいて、複数のテキストセグメントの中から選択することを含む。

【００１４】

いくつかの実装形態において、複数のテキストセグメントの中から選択することは、複数のテキストセグメントのそれぞれについての言語複雑度スコアを判定すること、およびクライアントデバイスのユーザの言語習熟度を記述する基準スコアと最も良く一致する言語複雑度スコアを有するテキストセグメントを選択することを含む。

【００１５】

いくつかの実装形態において、テキスト－音声モジュールによる出力用のテキストセグメントを決定することは、ユーザへのテキスト－音声出力用のテキストセグメントを識別すること、テキスト－音声出力用のテキストセグメントの複雑度スコアを計算すること、ならびにユーザの判定された言語習熟度およびテキスト－音声出力用のテキストセグメントの複雑度スコアに少なくとも基づいて、ユーザへのテキスト－音声出力用のテキストセグメントを修正することを含む。

【００１６】

いくつかの実装形態において、ユーザへのテキスト－音声出力用のテキストセグメントを修正することは、ユーザの言語習熟度の判定に少なくとも基づいて、ユーザについての全体的複雑度スコアを判定すること、ユーザへのテキスト－音声出力用のテキストセグメ

ント内の個々の部分についての複雑度スコアを判定すること、ユーザについての全体的複雑度スコアよりも大きい複雑度スコアを有する、テキストセグメント内の1つまたは複数の個々の部分を識別すること、および複雑度スコアを全体的複雑度スコアを下回るまで低減するように、テキストセグメント内の1つまたは複数の個々の部分を修正することを含む。

【００１７】

いくつかの実装形態において、ユーザへのテキスト－音声出力用のテキストセグメントを修正することは、ユーザに関連付けられたコンテキストを示すデータを受信すること、ユーザに関連付けられたコンテキストについての全体的複雑度スコアを判定すること、テキストセグメントの複雑度スコアが、ユーザに関連付けられたコンテキストについての全体的複雑度スコアを超えると判定すること、および複雑度スコアをユーザに関連付けられたコンテキストについての全体的複雑度スコアを下回るまで低減するように、テキストセグメントを修正することを含む。

【００１８】

別の態様では、コンピュータプログラムが、コンピューティング装置によって実行されると、上記方法のいずれかを実施させる機械可読命令を含む。

【００１９】

別の全般的態様では、コンピュータ実装方法が、ユーザに関連付けられたコンテキストを示すデータを受信するステップと、ユーザに関連付けられたコンテキストについての全体的複雑度スコアを判定するステップと、ユーザへのテキスト－音声出力用のテキストセグメントを識別するステップと、テキストセグメントの複雑度スコアが、ユーザに関連付けられたコンテキストについての全体的複雑度スコアを超えると判定するステップと、複雑度スコアをユーザに関連付けられたコンテキストについての全体的複雑度スコアを下回るまで低減するように、テキストセグメントを修正するステップとを含む。

【００２０】

いくつかの実装形態において、ユーザに関連付けられたコンテキストについての全体的複雑度スコアを判定するステップは、ユーザがコンテキストにあると判定されたとき、ユーザによる、以前に提出されたクエリ内に含まれる用語を識別するステップと、識別された用語に少なくとも基づいて、ユーザに関連付けられたコンテキストについての全体的複雑度スコアを判定するステップとを含む。

【００２１】

いくつかの実装形態において、ユーザに関連付けられたコンテキストを示すデータは、ユーザによって以前に提出されたクエリを含む。

【００２２】

いくつかの実装形態において、ユーザに関連付けられたコンテキストを示すデータは、ユーザに関連付けられた現在のロケーションを示すGPS信号を含む。

【００２３】

1つまたは複数の実装形態の詳細が、添付の図面および以下の説明において説明される。他の可能な特徴および利点は、説明、図面および特許請求の範囲から明らかになるであろう。

【００２４】

これらの態様の他の実装形態は、コンピュータ記憶デバイス上で符号化された方法のアクションを実施するように構成された、対応するシステム、装置およびコンピュータプログラムを含む。

【図面の簡単な説明】

【００２５】

【図１】言語習熟度に基づいてテキスト－音声出力を生成するためのプロセスの例を示す図である。

【図２】ユーザコンテキストに基づいて適応的テキスト－音声出力を生成するためのシステムの例を示す図である。

【図３】テキスト－音声出力内の文構造を修正するためのシステムの例を示す図である。
【図４】クラスタ化技法の使用に基づいて適応的テキスト－音声出力を生成するためのシステムの例を示すブロック図である。
【図５】適応的テキスト－音声出力を生成するためのプロセスの例を示すフロー図である。
【図６】本明細書に記載されるプロセス、またはその一部が、その上で実装されることが可能であるコンピューティングデバイスのブロック図である。
【発明を実施するための形態】
【００２６】
図面において、類似の参照番号は、全体を通して対応する部分を表す。
【００２７】
図1は、言語習熟度に基づいてテキスト－音声出力を生成するためのプロセス100Aおよび100Bの例を示す図である。プロセス100Aおよび100Bは、テキストクエリ104用に、それぞれ、言語習熟度が高いユーザ102aおよび言語習熟度が低いユーザ102b向けに異なるテキスト－音声出力を生成するために使用される。図示のように、ユーザデバイス106aおよび106b上でクエリ104を受信した後、プロセス100Aは、ユーザ102a向けに高複雑度テキスト－音声出力108aを生成し、プロセス100Bは、ユーザ102b向けに低複雑度出力108bを生成する。さらに、プロセス100Aおよび100Bを実行するTTSシステムは、言語習熟度推定器110、テキスト－音声エンジン120を含むことができる。加えて、テキスト－音声エンジン120は、テキストアナライザ122、言語学アナライザ124、および波形生成器126をさらに含むことができる。
【００２８】
概して、テキスト－音声出力を生成するために使用されるテキストの内容は、ユーザの言語習熟度に従って決定することができる。さらに、または代替として、テキスト－音声出力を生成するために使用されるべきテキストは、ユーザのコンテキスト、たとえば、ユーザのロケーションまたはアクティビティ、存在するバックグラウンドノイズ、ユーザの現在のタスクなどに基づいて決定することができる。さらに、可聴形に変換されるべきテキストは、ユーザがタスクを完了するのに失敗したか、またはアクションを繰り返しているという指示など、他の情報を使用して調整または決定されてもよい。
【００２９】
本例では、2人のユーザ、すなわちユーザ102aおよびユーザ102bが、それぞれ、ユーザデバイス106aおよび106b上で同じクエリ104をアプリケーション、ウェブページ、または他の検索機能への入力として提供する。たとえば、クエリ104は、その日の天気予報を判定するために、ユーザデバイス106aおよび106bに送信される音声クエリであることが可能である。クエリ104は次いで、クエリ104に応答してテキスト－音声出力を生成するためにテキスト－音声エンジン120に送信される。
【００３０】
言語習熟度推定器110は、ユーザデータ108aに基づいて、特定のユーザ(たとえば、ユーザ102aまたはユーザ102b)に関連付けられた言語習熟度スコアを判定する、TTSシステム内のソフトウェアモジュールであることが可能である。言語習熟度スコアは、特定の言語での通信を理解するための、具体的には、特定の言語で音声を理解するための、ユーザの能力の推定値であることが可能である。言語習熟度の1つの測度は、音声制御タスクの完了に成功するためのユーザの能力である。カレンダ予約の設定、方角の参照などのような多くのタイプのタスクは、ユーザとデバイスが言語通信を交換する一連の対話に従う。ユーザが、音声インターフェースを通してこれらのタスクワークフローの完了に成功する割合は、ユーザの言語習熟度の強力なインジケータである。たとえば、ユーザが開始する10個の音声タスクのうち9つを完了するユーザは、高い言語習熟度を有する見込みがある。一方、ユーザが開始する音声タスクのうちの大多数を完了するのに失敗するユーザは、低い言語習熟度を有すると推論することができ、それはそのユーザは、デバイスからの通信を十分に理解していない場合があるか、または適切な言葉による応答を提供することができ

なかった場合があるからである。さらに以下で議論するように、ユーザが、標準TTS出力を含むワークフローを完了しないと、低い言語習熟度スコアを生じ、TTSは、様々なタスクを理解し、完了するためのユーザの能力を増大する場合がある、適応された、単純化された出力を使用してもよい。

【００３１】

図示のように、ユーザデータ108aは、ユーザによって提出された先行テキストクエリ内で使用される単語と、TTSシステムによって使用される英語、それとも他のいずれかの言語が、ユーザのネイティブ言語であるかという指示と、ユーザの言語把握技能を反映するアクティビティおよび/または挙動のセットとを含むことができる。たとえば、図1に示されるように、ユーザのタイピング速度は、ある言語でのユーザの言語の流暢さを判定するために使用することができる。さらに、言語ボキャブラリ複雑度スコアまたは言語習熟度スコアが、以前のテキストクエリ中でユーザによって使用された単語に、所定の複雑度を関連付けることに基づいて、ユーザに割り当てられることが可能である。別の例では、先行クエリ中の誤認識された単語の数もまた、言語習熟度スコアを判定するために使用することができる。たとえば、誤認識された単語の高い数字は、低い言語習熟度を示すために使用することができる。いくつかの実装形態において、言語習熟度スコアは、クエリ104の提出に先立ってユーザについて判定された、ユーザに関連付けられた記憶されたスコアを参照することによって判定される。

【００３２】

図1は、言語習熟度推定器110をTTSエンジン120とは別個の構成要素として示すが、いくつかの実装形態では、図2に示されるように、言語習熟度推定器110は、TTSエンジン120内の一体型ソフトウェアモジュールであることが可能である。そのような事例では、言語習熟度推定を伴う動作は、TTSエンジン120によって直接調節することができる。

【００３３】

いくつかの実装形態において、ユーザに割り当てられる言語習熟度スコアは、ユーザについて推定される特定のユーザコンテキストに基づいてもよい。たとえば、図2に関してより詳細に記載されるように、ユーザコンテキスト判定は、ユーザに、限られた言語把握能力を一時的に持たせることができるコンテキスト特有言語習熟度を判定するために使用することができる。たとえば、ユーザコンテキストが大幅なバックグラウンドノイズを示す場合、またはユーザが運転などのタスクに携わっている場合、言語習熟度スコアは、ユーザの現時点での言語把握能力が、他のユーザコンテキストに相対して一時的に低下されることを示すために使用することができる。

【００３４】

いくつかの実装形態では、以前のユーザアクティビティに基づいて言語習熟度を推論するのではなく、言語習熟度スコアは代わりに言語習熟度推定器110を使用せずにTTSエンジン120に直接提供することができる。たとえば、言語習熟度スコアは、ユーザの言語習熟度レベルを指定する登録プロセス中のユーザ入力に基づいて、ユーザに対して明示することができる。たとえば、登録中ユーザはユーザの技能レベルを指定する選択を提供することができ、それは次いでユーザについての適切な言語習熟度を算出するために使用することができる。他の例では、ユーザは、ユーザの言語習熟度レベルを指定するために使用することができる、人口統計情報、教育レベル、居住地などのような、他のタイプの情報を提供することができる。

【００３５】

上述した例では、言語習熟度スコアは、最近生成されたユーザアクティビティデータに基づいて定期的に調整される離散値のセット、または登録プロセス中に最初に明示される継続スコアのいずれかであることが可能である。第1の事例では、言語習熟度スコアの値は、ユーザの現時点での言語把握および習熟度が減衰される場合があることを示す1つまたは複数の要因(たとえば、大幅なバックグラウンドノイズを示すユーザコンテキスト)に基づいてバイアスすることができる。第2の事例では、言語習熟度スコアの値は、初回算出の後にプリセットし、ユーザの言語習熟度が増大したこと(たとえば、所与の言語につ

いてのタイピングレートの増加または訂正レートの減少)を示す特定のマイルストーンイベントの後でのみ調整することができる。他の事例では、これらの2つの技法の組合せを特定のテキスト入力に基づいてテキスト－音声出力を可変的に調整するために使用することができる。そのような事例では、ユーザの言語技能の特定の側面をそれぞれが表す複数の言語習熟度スコアをユーザ向けにテキスト－音声出力をどのようにして最も良く調整するかを決定するために使用することができる。たとえば、ある言語習熟度スコアはユーザのボキャブラリの複雑度を表すことができ、別の言語習熟度スコアは、ユーザの文法技能を表すために使用することができる。

【００３６】

TTSエンジン120は、ユーザの言語習熟度スコアによって示される言語習熟度に適応されたテキスト－音声出力を生成するために言語習熟度スコアを使用することができる。いくつかの事例では、TTSエンジン120は、テキストクエリ104用の候補TTS文字列のセットから特定のTTS文字列を選択したことに基づいて、テキスト－音声出力を適応させる。そのような事例では、TTSエンジン120は、候補TTS文字列のそれぞれがユーザによって正確に解釈される見込みを予測するためのユーザの言語習熟度スコアの使用に基づいて、特定のTTS文字列を選択する。これらの技法に関連した、より具体的な記述が、図2に関して提供される。代替として、他の事例では、TTSエンジン120は、ベースラインTTS文字列を選択し、言語習熟度スコアによって示されるユーザの言語習熟度レベルに基づいて、TTS文字列の構造を調整することができる。そのような事例では、TTSエンジン120は、ユーザによって理解される見込みがより高い適応されたTTS文字列を生成するために、ベースラインTTS文字列の文法を調整し、単語置換を行い、かつ/または文複雑度を低減することができる。これらの技法に関連した、より具体的な記述が、図3に関して提供される。

【００３７】

依然として図1を参照すると、TTSエンジン120は、ユーザについての言語習熟度スコアが異なるので、ユーザ102aおよび102b向けに異なるテキスト－音声出力を生成してもよい。たとえば、プロセス100Aにおいて、言語習熟度スコア106aは、ユーザ102aが複雑ボキャブラリを有し、英語を第1言語として有し、先行ユーザクエリ中に比較的高い毎分単語数を有することを示すユーザデータ108aから推論される高い英語習熟度を示す。言語習熟度スコア106aの値に基づいて、TTSエンジン120は、複雑な文法構造を含む高複雑度テキスト－音声出力108aを生成する。図示のように、テキスト－音声出力108aは、今日の最高気温および最低気温についての付加情報を含む従属節に加え、今日の予報が晴天であることを記述する独立節を含む。

【００３８】

プロセス100Bの例では、言語習熟度スコア106bは、ユーザ102bが単純ボキャブラリを有し、英語を第2言語として有し、以前に10個の不適当なクエリを与えたことを示すユーザアクティビティデータ108bから推論される低い英語習熟度を示す。この例では、TTSエンジン120は、テキスト－音声出力108aに相対して、より単純な文法構造を含む低複雑度テキスト－音声出力108bを生成する。たとえば、単一の文に複数の節を含むのではなく、テキスト－音声出力108bは、テキスト－音声出力108aと同じ主情報(たとえば、今日の予報が晴天である)を伝える単一の独立節を含むが、今日の最高および最低気温に関連した付加情報は含まない。

【００３９】

TTS出力用のテキストの適応は、異なる様々なデバイスおよびソフトウェアモジュールによって実施することができる。たとえば、サーバシステムのTTSエンジンは、言語習熟度スコアに基づいてテキストを調整し、次いで、調整されたテキストの合成された発声を含むオーディオを出力するための機能を含んでもよい。別の例として、サーバシステムの前処理モジュールは、テキストを調整し、調整されたテキストを音声合成のためにTTSエンジンに渡してもよい。別の例として、ユーザデバイスは、適切なTTS出力を生成することができるように、TTSエンジン、またはTTSエンジンおよびテキストプリプロセッサを含んでもよい。

【００４０】
いくつかの実装形態において、TTSシステムは、クライアントデバイスのサードパーティモバイルアプリケーションまたはウェブページと通信を交換するように構成されたソフトウェアモジュールを含むことができる。たとえば、システムのTTS機能は、アプリケーションパッケージインターフェース(API)を通して、サードパーティモバイルアプリケーションに対して利用可能にされてもよい。APIは、TTSエンジン120を稼動するサーバシステムに対してTTSオーディオを要求するためにアプリケーションまたはウェブサイトが使用することができるプロトコルの定義済みセットを含むことができる。いくつかの実装形態において、APIは、ユーザのデバイス上でローカルに稼動するTTS機能を利用可能にすることができる。たとえば、APIは、プロセス間通信(IPC)、リモートプロシージャコール(RPC)、または他のシステムコールもしくは関数を通して、アプリケーションまたはウェブページに対して利用可能であってもよい。TTSエンジン、および関連言語習熟度分析またはテキスト前処理は、ユーザの言語習熟度向けの適切なテキストを決定し、また合成された音声用のオーディオを生成するためにユーザのデバイス上でローカルに稼動されることが可能である。
【００４１】
たとえば、サードパーティアプリケーションまたはウェブページは、サードパーティアプリケーションまたはウェブページの音声インターフェースのタスクフローに基づいて、ユーザに提供される音声命令のセットを生成するためにAPIを使用することができる。APIは、アプリケーションまたはウェブページが、音声に変換されるべきテキストを提供するべきだと指定することができる。いくつかの事例では、ユーザ識別子または言語習熟度スコアなど、他の情報が提供されることが可能である。
【００４２】
TTSエンジン120がAPIを使用してサードパーティアプリケーションと通信を交換する実装形態では、TTSエンジン120は、サードパーティアプリケーションからのテキストセグメントが、テキスト用のテキスト－音声出力を生成するのに先立って調整されるべきかどうかを判定するために使用することができる。たとえば、APIは、適応的テキスト－音声出力の生成を開始するサードパーティアプリケーション内での条件を指定するコンピュータ実装プロトコルを含むことができる。
【００４３】
一例として、あるAPIが、TTS出力用の候補として複数の異なるテキストセグメントをアプリケーションが提出することを許可する場合があり、異なるテキストセグメントは、異なるレベルの言語習熟度に対応する。たとえば、候補は、等価な意味を有するが異なる複雑度レベル(たとえば、高複雑度応答、中複雑度応答、および低複雑度応答)を有するテキストセグメントであることが可能である。TTSエンジン120は次いで、各候補を理解するために必要とされる言語習熟度のレベルを判定し、ユーザに対する適切な言語習熟度スコアを判定し、言語習熟度スコアに最もよく対応する候補テキストを選択してもよい。TTSエンジン120は次いで、選択されたテキスト用の合成されたオーディオをたとえばAPIを使用してネットワークを介してアプリケーションに返す。いくつかの事例では、APIは、ユーザデバイス106aおよび106b上でローカルに利用可能であることが可能である。そのような事例では、APIは、様々なタイプのプロセス間通信(IPC)を介して、またはシステムコールによりアクセス可能であてもよい。たとえば、ユーザデバイス106aおよび106b上でのAPIの出力は、APIがユーザデバイス106aおよび106b上でローカルに動作するので、TTSエンジン120のテキスト－音声出力であることが可能である。
【００４４】
別の例では、APIにより、サードパーティアプリケーションは、単一のテキストセグメントと、TTSエンジン120が、異なる複雑度を有するテキストセグメントを生成するようにテキストセグメントを修正することを許可されるかどうかを示す値とを提供することができる。アプリまたはウェブページが改変を許可されることを示す場合、TTSシステム120は、元のテキストが、ユーザが口頭応答で理解することができるよりも複雑であることを言

語習熟度スコアが示唆するとき、たとえば、テキストの複雑度を低減するためにテキストに対する様々な変更を行ってもよい。さらに他の例では、APIにより、サードパーティアプリケーションは、TTSエンジン120が、ユーザに関連付けられたユーザコンテキストを判定し、判定されたユーザコンテキストに基づいて、特定のテキスト－音声出力を調整および生成することができるように、テキストセグメントとともにユーザデータ(たとえば、サードパーティアプリケーションにおいて提出された先行ユーザクエリ)もまた提供する。同様に、APIにより、アプリケーションは、ユーザデバイスからのコンテキストデータ(たとえば、全地球測位信号、加速度計データ、環境ノイズレベルなど)またはTTSエンジン120に、サードパーティアプリケーションを通してユーザに最終的に提供されるテキスト－音声出力を調整させるためのユーザコンテキストの指示を提供することができる。いくつかの事例では、サードパーティアプリケーションはまた、ユーザの言語習熟度を判定するために使用することができるデータをAPIに提供することができる。

【００４５】

いくつかの実装形態において、TTSエンジン120は、ユーザの言語習熟度を使用することも、ユーザに関連付けられたコンテキストを判定することもなく、ユーザクエリ用のテキスト－音声出力を調整することができる。そのような実装形態では、TTSエンジン120は、ユーザが出力を誤認しているという信号を受信したこと(たとえば、同じクエリまたはタスクに対する複数の再試行)に基づいて、初回テキスト－音声出力がユーザにとって複雑すぎると判定することができる。それに応答して、TTSエンジン120は、再試行されたクエリまたは関連クエリについて、後続テキスト－音声応答の複雑度を低減することができる。したがって、ユーザが、アクションの完了に成功することができないとき、TTSエンジン120は、TTS出力を理解するために求められる詳細の量または言語習熟度をユーザが理解するレベルに達するまで漸進的に低減してもよい。

【００４６】

図2は、ユーザコンテキストに基づいてテキスト－音声出力を適応的に生成するシステム200の例を示す図である。簡潔には、システム200は、クエリアナライザ211と、言語習熟度推定器212と、補間器213と、言語学アナライザ214と、リランカ(reranker)215と、波形生成器216とを含むTTSエンジン210を含むことができる。システム200はまた、コンテキストプロファイル232のセットを記憶するコンテキストリポジトリ220、およびユーザ履歴データ234を記憶するユーザ履歴マネージャ230を含む。いくつかの事例では、TTSエンジン210は、図1に関して記載されたTTSエンジン120に対応する。

【００４７】

本例では、ユーザ202は最初に、その日のユーザの最初の会議に関連した情報に対する要求を含むクエリ204をユーザデバイス208上で提出する。ユーザデバイス208は次いで、ユーザ202に関連付けられたクエリ204およびコンテキストデータ206をそれぞれクエリアナライザ211および言語習熟度推定器212に送信することができる。クエリへの応答ではない他のタイプのTTS出力、たとえば、カレンダリマインダ、通知、タスクワークフローなどが、同じ技法を使用して適応されてもよい。

【００４８】

コンテキストデータ206は、繰り返されるテキストクエリの間の時間間隔、ユーザ202に関連付けられた、ロケーション、速度、または動きパターンを示す全地球測位信号(GPS)データ、特定の時間期間内にTTSエンジン210に提出された先行テキストクエリ、またはTTSエンジン210に関連したユーザアクティビティを示すことができる他のタイプのバックグラウンド情報など、ユーザ202に関連付けられた特定のコンテキストに関する情報を含むことができる。いくつかの事例では、コンテキストデータ206は、クエリ204が、ユーザアクションに関連付けられたテキストセグメント、それともテキスト－音声出力を生成するためにTTSエンジン210に送信される命令であるかなど、TTSエンジン210に提出されたクエリ204のタイプを示すことができる。

【００４９】

クエリ204を受信した後、クエリアナライザ211は、クエリ204に応答する情報を識別す

るためにクエリ204を解析する。たとえば、クエリ204が音声クエリであるいくつかの事例では、クエリアナライザ211は最初に、音声クエリのトランスクリプションを生成し、次いで、たとえば、検索エンジンにクエリを提供することによってクエリ204に応答する情報を決定し、検索結果を受信するためにクエリ204内の個々の単語またはセグメントを処理する。クエリのトランスクリプションおよび識別された情報は次いで、言語学アナライザ214に送信することができる。

【００５０】

ここで言語習熟度推定器212を参照すると、コンテキストデータ206を受信した後、言語習熟度推定器212は、図1に関して記載された技法を使用して受信されたコンテキストデータ206に基づいて、ユーザ202についての言語習熟度を計算する。具体的には、言語習熟度推定器212は、リポジトリ220上に記憶された様々なコンテキストプロファイル232を通して解析する。コンテキストプロファイル232は、特定のユーザコンテキストに関連付けられた関連タイプの情報を含むアーカイブされたライブラリであることが可能であり、テキスト－音声出力に含めることができる。コンテキストプロファイル232は、ユーザ202が現時点で、コンテキストプロファイル232に関連付けられたコンテキスト内にあるとき、各情報タイプがユーザ202によって理解される見込みがある限度を表す各情報タイプに関連付けられた値をさらに指定する。

【００５１】

図2に示す例において、コンテキストプロファイル232は、ユーザ202が現在ユーザ202が自身の毎日の通勤中であることを示すコンテキストにあることを指定する。さらに、コンテキストプロファイル232は、ユーザ202によって把握される見込みがある個々の単語および句についての値も指定する。たとえば、データまたは時間情報は、"SINCE"については"0.9"という値に関連付けられ、ユーザ202が、会議に関連付けられた詳細な情報(たとえば、会議に出席する関係者、または会議のロケーション)よりもむしろ、会議に関連付けられた、一般化された情報(たとえば、次に行われる会議の時間)204を理解する見込みがより高いことを示す。この例では、値の差は、複雑または詳細な情報を理解するためのユーザの能力が低下されるので、特定のタイプの情報を理解するためのユーザの能力における差を示す。

【００５２】

個々の単語および句に関連付けられた値は、ユーザ202が以前コンテキストデータ206によって示されるコンテキスト内の以前のユーザセッションからのユーザアクティビティデータに基づいて判定することができる。たとえば、履歴ユーザデータはユーザ履歴マネージャ230から送信することができ、マネージャ230は、クエリログ234内に記憶されたデータを取り出す。本例では、日時情報についての値は、ユーザが通常、会議のロケーションよりも頻繁に会議に関連付けられた日時情報にアクセスすると判定したことに基づいて増大することができる。

【００５３】

言語習熟度推定器212が、受信されたコンテキストデータ206に対応する特定のコンテキストプロファイル232を選択した後、言語習熟度推定器212は、選択されたコンテキストプロファイル232を補間器213に送信する。補間器213は、選択されたコンテキストプロファイル232を解析し、含まれる個々の単語および句と、それらに関連付けられた値とを抽出する。いくつかの事例では、補間器213は、異なるタイプの情報および関連付けられた値をテキスト－音声出力候補240aのリストを生成するための言語学アナライザ214に直接送信する。そのような事例では、補間器213は、特定のタイプの情報と、関連付けられた値とを選択されたコンテキストプロファイル232から抽出し、言語学アナライザ214に送信する。他の事例では、補間器213はまた、選択されたコンテキストプロファイル232をリランカ215に送信することができる。

【００５４】

いくつかの事例では、TTSエンジン210には、構造化データのセット(たとえば、カレンダイベントのフィールド)を提供することができる。そのような事例では、補間器213は、

構造化データをコンテキストプロファイル232によって示されるユーザの習熟度に合致するレベルでのテキストに変換することができる。たとえば、TTSエンジン210は、構造化データ中の情報を表すための異なるレベルの詳細または複雑度を示す1つまたは複数の文法を示すデータにアクセスし、ユーザの言語習熟度スコアに基づいて適切な文法を選択してもよい。同様に、TTSエンジン210は、言語習熟度スコアを提供されると、適切な単語を選択するために辞書を使用することができる。

【００５５】

言語学アナライザ214は、クエリ204内に含まれる情報に対して、正規化などの処理動作を実施する。たとえば、クエリアナライザ211は、クエリ204内に含まれる各単語または小片に音素的トランスクリプションを割り当て、クエリ204をテキスト音素変換を使用して句、節、および文などの韻律単位に分割することができる。言語学アナライザ214はまた、クエリ204に応答するものとして識別される複数のテキスト－音声出力候補を含むリスト240aを生成する。この例では、リスト240aは、異なるレベルの複雑度を有する複数のテキスト－音声出力候補を含む。たとえば、"At 12:00PM with Mr. John near Dupont Circle"という応答は、会議の時間、会議のロケーション、会議が行われる相手を識別するので、最も複雑な応答である。相対的に、"In three hours"という応答は、会議の時間を識別するだけなので、最も複雑でない。

【００５６】

リスト240aはまた、各テキスト－音声出力候補がクエリ204に応答する見込みがある見込みに基づく、テキスト－音声候補についてのベースラインランクを含む。本例では、リスト240aは、最も複雑なテキスト－音声出力候補が、クエリ204の内容に関連付けられた最も大量の情報を含むので、クエリ204に応答する見込みが最もあることを示す。

【００５７】

言語学アナライザがテキスト－音声出力候補のリスト240aを生成した後、リランカ215は、受信されたコンテキストデータ206に基づくテキスト－音声出力候補についての調整されたランクを含むリスト240bを生成する。たとえば、リランカ215は、選択されたコンテキストプロファイル232内に含まれる特定のタイプの情報に関連付けられたスコアに基づいてランクを調整することができる。

【００５８】

本例では、リランカ215は、ユーザ202が、ユーザが通勤中であることを示すユーザの現在のコンテキストを提供されると、テキスト－音声応答内の日時情報を把握する見込みはあるが、テキスト－音声応答内の当事者名もロケーション情報も理解する見込みはないことを示すコンテキストプロファイル232に基づいて、最も単純なテキスト－音声出力を最も高くランク付けする。この点において、受信されたコンテキストデータ206は、ユーザ202がTTSエンジン210のテキスト－音声出力204cの内容を理解する見込みを増大するように、特定のテキスト－音声出力候補の選択を調整するために使用することができる。

【００５９】

図3は、テキスト－音声出力内の文構造を修正するためのシステム300の例を示す図である。簡潔には、TTSエンジン310が、クエリ302およびユーザ(たとえば、ユーザ202)についての言語習熟度プロファイル304を受信する。TTSエンジン310は次いで、クエリ302に応答する調整されたテキスト－音声出力306cを生成するために動作312、314、および316を実施する。いくつかの事例では、TTSエンジン310は、図1に関して記載されたTTSエンジン120、または図2に関して記載されたTTSエンジン210に対応する。

【００６０】

概して、TTSエンジン310は、異なるタイプの調整技法を使用してクエリ302用のベースラインテキスト－音声出力306aの文構造を修正することができる。一例として、TTSエンジン310は、個々の単語または句に関連付けられた複雑度スコアが、ユーザの言語複雑度プロファイル304によって示される閾スコアよりも大きいと判定したことに基づいて、ベースラインテキスト－音声出力306a内の単語または句を置換することができる。別の例として、TTSエンジン310は、ベースラインテキスト－音声出力306aの全体的複雑度が、言語

複雑度プロファイル304に基づいて満足できるレベルまで低減されるように、個々の文の節(sentence clause)を並べ直すことができる。TTSエンジン310はまた、テキストの複雑度を調整するために単語を並べ替え、文を分裂させ、または組み合わせ、他の変更を行うことができる。

【００６１】

より詳細には、動作312の間、TTSエンジン310は最初に、クエリ302に応答するベースラインテキスト－音声出力306aを生成する。TTSエンジン310は次いで、ベースラインテキスト－音声出力306aをセグメント312a～312cに解析する。TTSエンジン310はまた、個々のセグメントの間の区切り点を示す句読点(たとえば、コンマ、ピリオド、セミコロンなど)を検出する。TTSエンジン310はまた、セグメント312a～312cのそれぞれについての複雑度スコアを計算する。いくつかの事例では、複雑度スコアは、特定の言語における特定の単語の頻度に基づいて計算することができる。代替技法は、ユーザによる使用頻度、またはユーザによってアクセスされる履歴コンテンツ(たとえば、ニュース記事、ウェブページなど)中での出現の頻度に基づいて、複雑度スコアを計算することを含むことができる。これらの例のそれぞれにおいて、複雑度スコアは、ユーザによって把握される見込みがある単語およびユーザによって把握される見込みがない他の単語を示すために使用することができる。

【００６２】

本例では、セグメント312aおよび312bは、それぞれ、"FORECAST"および"CONSISTENT"などの高複雑用語を含むことに基づいて、比較的複雑であると判定される。ただし、セグメント312cは、含まれる用語が比較的単純なので、比較的単純であると判定される。この判定は、セグメント312cについての複雑度スコア(たとえば、0.41)と比較して、より高い複雑度スコア(たとえば、0.83、0.75)を有するセグメント312aおよび312bによって表される。

【００６３】

上述したように、言語習熟度プロファイル304は、ユーザによって把握可能な最大複雑度を示す閾複雑度スコアを計算するために使用することができる。本例では、閾複雑度スコアは、セグメント312aおよび312bがユーザによって把握される見込みがないとTTS310が判定するように、"0.7"となるように計算することができる。

【００６４】

言語習熟度プロファイル304によって示される閾複雑度スコアよりも大きい関連複雑度スコアを有する個々のセグメントを識別した後、動作314の間、TTSエンジン310は、識別された単語をユーザによって理解される見込みがより高いと予測される代替物で置換する。図3に示されるように、"FORECAST"は"WEATHER"で置換することができ、"CONSISTENT"は"CHANGE"で置換することができる。これらの例では、セグメント314aおよび314bは、言語習熟度プロファイル304によって示される閾複雑度スコアを下回る、より低複雑度のスコアを有するより単純な代替を表す。

【００６５】

いくつかの実装形態において、TTSエンジン310は、高度に複雑な単語と置き換わるべき適度に複雑な単語を決定するために教師なし技法を使用するトレーニングされたスキップグラムモデルを使用して高複雑度単語についての単語置換を処理することができる。いくつかの事例ではまた、TTSエンジン310は、高複雑単語のための単語置換を処理するためにシソーラスまたは同義語データを使用することができる。

【００６６】

ここで動作316を参照すると、クエリの文の節は、特定の文構造に関連付けられた複雑度の計算、および言語習熟度プロファイル304によって示される言語習熟度に基づいて、ユーザが文構造を理解することができるかどうかの判定に基づいて調整することができる。

【００６７】

本例では、TTSエンジン310は、ベースラインテキスト－音声応答306aが3つの文の節(た

とえば、"today's forecast is sunny"、"but not consistent"、および"and warm")を含むと判定したことに基づいて、ベースラインテキスト－音声応答306aが高い文複雑度を有すると判定する。それに応答して、TTSエンジン310は、調整された文部分316aおよび316bを生成することができ、これらは、従属節と独立節を組み合わせて、セグメント化用句読点を含まない単一の節にする。その結果、調整されたテキスト－音声応答306bは、より単純なボキャブラリ(たとえば、"WEATHER"、"CHANGE")ならびにより単純な文構造(たとえば、節分離なし)の両方を含み、調整されたテキスト－音声出力306bをユーザが理解する見込みが増大する。調整されたテキスト－音声出力306bは次いで、TTSエンジン310による出力用に出力306cとして生成される。

【００６８】

いくつかの実装形態において、TTSエンジン310は、ユーザにとって問題があると識別される特定の文構造を避けるための重み付け係数を使用してベースラインクエリ302を調整することを含むユーザ特有再構造化アルゴリズムの使用に基づいて文構造調整を実施することができる。たとえば、ユーザ特有再構造化アルゴリズムは、従属節の包含を低く重み付け、または単純な主語動詞目的語シーケンスを有する文の節を高く重み付けるための選択肢を指定することができる。

【００６９】

図4は、クラスタ化技法の使用に基づいて、テキスト－音声出力を適応的に生成するシステム400の例を示すブロック図である。システム400は、言語習熟度推定器410、ユーザ類似度判定器420、複雑度オプティマイザ430、および機械学習システム440を含む。

【００７０】

簡潔には、言語習熟度推定器410は、複数のユーザ402からデータを受信する。言語習熟度推定器410は次いで、複数のユーザ402のそれぞれについて言語複雑度プロファイル412のセットを推定し、セットは次いで、ユーザ類似度判定器420に送信される。ユーザ類似度判定器420は、類似ユーザのユーザクラスタ424を識別する。複雑度オプティマイザ430および機械学習システム440は次いで、複雑度マッピング442を生成するためにユーザクラスタ424内の各ユーザの言語複雑度プロファイル412と、複数のユーザ402から受信されたコンテキストデータとを分析する。

【００７１】

概して、システム400は、ユーザの集団についてのアクティブ言語複雑度とパッシブ言語複雑度との間の関係を分析するために使用することができる。アクティブ言語複雑度は、ユーザによって提供される検出された言語入力(たとえば、テキストクエリ、音声入力など)を指す。パッシブ言語複雑度は、ユーザに提供される音声信号を理解または把握するためのユーザの能力を指す。この点において、システム400は、各個々のユーザについての適切なパッシブ言語複雑度を判定するために複数のユーザについてのアクティブ言語複雑度とパッシブ言語複雑度との間の判定された関係を使用することができ、特定のユーザが、テキスト－音声出力を理解する最も高い見込みを有する。

【００７２】

複数のユーザ402は、TTSエンジン(たとえば、TTSエンジン120)に関連付けられたアプリケーションを使用する複数のユーザであることが可能である。たとえば、複数のユーザ402は、モバイルアプリケーションのユーザインターフェースを介してテキスト－音声特徴をユーザに提供するためにTTSエンジンを使用するモバイルアプリケーションを使用するユーザのセットであることが可能である。そのような事例において、複数のユーザ402からのデータ(たとえば、先行ユーザクエリ、ユーザ選択など)は、モバイルアプリケーションによって追跡し、言語習熟度推定器410による分析のために集約することができる。

【００７３】

言語習熟度推定器410は最初に、図1に関して以前に記載されたものとほぼ同様の技法を使用して複数のユーザ402についてのパッシブ言語複雑度を測定することができる。言語習熟度推定器410は次いで、複数のユーザ402のそれぞれについての個々の言語複雑度プロファイルを含む言語複雑度プロファイル412を生成することができる。各個々の言語複雑

度プロファイルは、複数のユーザ402のそれぞれについてのパッシブ言語複雑度およびアクティブ言語複雑度を示すデータを含む。

【００７４】

ユーザ類似度判定器420は、複数のユーザ402内の類似ユーザを識別するために言語習熟度プロファイル412のセット内に含まれる言語複雑度データを使用する。いくつかの事例では、ユーザ類似度判定器420は、類似アクティブ言語複雑度(たとえば、類似言語入力、提供された音声クエリなど)を有するユーザをグループ化することができる。他の事例では、ユーザ類似度判定器420は、先行のユーザ提出クエリ中に含まれる単語、モバイルアプリケーションにおける特定のユーザ挙動、またはユーザロケーションを比較することによって類似ユーザを判定することができる。ユーザ類似度判定器420は次いで、ユーザクラスタ424を生成するように類似ユーザをクラスタ化する。

【００７５】

いくつかの実装形態において、ユーザ類似度判定器420は、記憶されたクラスタデータ422に基づいて、指定されたクラスタ中のユーザについての集約データを含むユーザクラスタ424を生成する。たとえば、クラスタデータ422は、複数のユーザ402に関連付けられたパッシブ言語複雑度を示す特定のパラメータ(たとえば、不適当なクエリ応答の数など)によってグループ化することができる。

【００７６】

ユーザクラスタ424を生成した後、複雑度オプティマイザ430は、TTSシステムによる言語出力の複雑度を変え、ユーザパフォーマンスを示すためにTTSシステムによる言語出力を理解するためのユーザの能力を示すパラメータ(たとえば、理解率、音声アクションフロー完了率、または回答成功率)のセットを使用してユーザのパッシブ言語複雑度を測定する。たとえば、パラメータは、各クラスタ424内のユーザが所与のテキスト－音声出力をどの程度理解するかを特徴づけるために使用することができる。そのような事例では、複雑度オプティマイザ430は最初に、ユーザに低複雑度音声信号を与え、様々な複雑度において追加音声信号を再帰的に提供することができる。

【００７７】

いくつかの実装形態において、複雑度オプティマイザ430はまた、各ユーザクラスタ424に関連付けられた様々なユーザコンテキストについての最適パッシブ言語複雑度を判定することができる。たとえば、パラメータのセットを使用してユーザの言語習熟度を測定した後、複雑度オプティマイザ430は次いで、測定されたデータを各ユーザコンテキスト用に最適パッシブ言語複雑度を判定することができるように、複数のユーザ402から受信されたコンテキストデータによって分類することができる。

【００７８】

様々なパッシブ言語複雑度についてのパフォーマンスデータを集めた後、機械学習システム440は次いで、ユーザの言語把握が最強であることをパフォーマンスパラメータが示す特定のパッシブ言語複雑度を判定する。たとえば、機械学習システム440は、アクティブ言語複雑度と、パッシブ言語複雑度と、ユーザコンテキストとの間の関係を判定するために特定のユーザクラスタ424内のすべてのユーザについてのパフォーマンスデータを集約する。

【００７９】

ユーザクラスタ424についての集約データは次いで、ユーザクラスタ424内の各ユーザについての実際の言語複雑度スコアを判定するためにユーザクラスタ424内の各ユーザについての個々のデータと比較することができる。たとえば、図4に示されるように、複雑度マッピング442は、最適パッシブ言語複雑度にマップされたアクティブ言語複雑度に対応する実際の言語複雑度を推論するためのアクティブ言語複雑度とパッシブ言語複雑度との間の関係を表すことができる。

【００８０】

複雑度マッピング442は、複数のユーザ402内のすべてのユーザクラスタについてのアクティブ言語複雑度と、TTS複雑度と、パッシブ言語複雑度との間の関係を表し、この関係

は次いで、個々のユーザによる後続クエリについての適切なTTS複雑度を予測するために使用することができる。たとえば、上述したように、ユーザ入力(たとえば、クエリ、テキストメッセージ、Eメールなど)は、類似ユーザをユーザクラスタ424にグループ化するために使用することができる。各クラスタについて、システムは、理解するための変動するレベルの言語習熟度を求めるTTS出力を提供する。システムは次いで、各クラスタ中のユーザに適した言語複雑度のレベルを判定するためにユーザから受信された応答、および変えられたTTS出力についてのタスク完了率を評価する。システムは、クラスタ識別子と、識別されたクラスタに対応するTTS複雑度スコアとの間のマッピング442を記憶する。システムは次いで、複雑度マッピング442をユーザ向けのTTS出力についての適切なレベルの複雑度を判定するために使用する。たとえば、システムは、ユーザのアクティブ言語習熟度を表すクラスタを識別し、マッピング442中のクラスタについての対応するTTS複雑度スコア(たとえば、パッシブ言語理解のレベルを示す)を参照し、取り出されたTTS複雑度スコアによって示される複雑度レベルを有するTTS出力を生成する。

【００８１】

ユーザについて判定された実際の言語複雑度は次いで、図1～図3に関して記載された技法を使用してTTSシステムを調整するために使用することができる。この点において、類似ユーザのグループ(たとえば、ユーザクラスタ424)からの集約言語複雑度データは、単一のユーザに関して、TTSシステムの性能をインテリジェントに調整するために使用することができる。

【００８２】

図5は、テキスト－音声出力を適応的に生成するためのプロセス500の例を示すフロー図である。簡潔には、プロセス500は、クライアントデバイスのユーザの言語習熟度を判定するステップ(510)、テキスト－音声モジュールによる出力用のテキストセグメントを決定するステップ(520)、テキストセグメントの合成された発声を含むオーディオデータを生成するステップ(530)、およびオーディオデータをクライアントデバイスに提供するステップ(540)を含むことができる。

【００８３】

より詳細には、プロセス500は、クライアントデバイスのユーザの言語習熟度を判定するステップ(510)を含むことができる。たとえば、図1に関して記載されたように、言語習熟度推定器110は、様々な技法を使用してユーザについての言語習熟度を判定することができる。いくつかの事例では、言語習熟度は、言語習熟度のレベルを示す割り当てられたスコアを表すことができる。他の事例では、言語習熟度は、言語習熟度の複数のカテゴリからの割り当てられたカテゴリを表すことができる。他の事例では、言語習熟度は、ユーザの習熟度レベルを示すユーザ入力および/または挙動に基づいて判定することができる。

【００８４】

いくつかの実装形態において、言語習熟度は、異なるユーザ信号から推論することができる。たとえば、図1に関して記載されたように、言語習熟度は、ユーザ入力のボキャブラリ複雑度、ユーザのデータエントリレート、音声入力からの誤認識された単語の数、異なるレベルのTTS複雑度についての完了された音声アクションの数、またはユーザによって閲覧されるテキスト(たとえば、本、記事、ウェブページ上のテキストなど)の複雑度のレベルから推論することができる。

【００８５】

プロセス500は、テキスト－音声モジュールによる出力用のテキストセグメントを決定するステップ(520)を含むことができる。たとえば、TTSエンジンは、ユーザの判定された言語習熟度に基づいて、ベースラインテキストセグメントを調整することができる。いくつかの事例では、図2に関して記載されたように、出力用のテキストセグメントは、ユーザに関連付けられたユーザコンテキストに基づいて調整することができる。他の事例では、図3に関して記載されたように、出力用のテキストセグメントはまた、テキストセグメントの複雑度を低減するために単語置換または文再構造化によって調整することができる

。たとえば、調整は、個々の単語がテキストセグメント中に含まれるのがどれだけ珍しいか、使用される動詞のタイプ(たとえば、複合動詞、または動詞時制)、テキストセグメントの言語学的構造(たとえば、従属節の数、関連単語の間の分離の量、句が入れ子にされる程度)などに基づくことができる。他の例では、調整は、言語学的特性についての参照測定値を用いる上記の言語学的測度(たとえば、主語と動詞との間の平均的分離、形容詞と名詞との間の分離など)にも基づくことができる。そのような例では、基準測定値はまた、平均を表すことができるか、または異なる複雑度レベルについての範囲もしくは例を含むことができる。

【００８６】

いくつかの実装形態において、出力用のテキストセグメントを決定するステップは、ユーザの言語習熟度レベルを記述する基準スコアと最も良く一致するスコアを有するテキストセグメントを選択するステップを含むことができる。他の実装形態では、個々の単語または句は、複雑度についてスコア付けすることができ、次いで、最も複雑な単語を全体的複雑度がユーザについての適切なレベルを満たすように、置換し、消去し、または再構造化することができる。

【００８７】

プロセス500は、テキストセグメントの合成された発声を含むオーディオデータを生成するステップ(530)を含むことができる。

【００８８】

プロセス500は、オーディオデータをクライアントデバイスに提供するステップ(540)を含むことができる。

【００８９】

図6は、クライアントとして、または1つのサーバもしくは複数のサーバとしてのいずれかで、本文書において記載されるシステムおよび方法を実装するために使用することができるコンピューティングデバイス600、650のブロック図である。コンピューティングデバイス600は、ラップトップ、デスクトップ、ワークステーション、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など、様々な形のデジタルコンピュータを表すことを意図している。コンピューティングデバイス650は、携帯情報端末、セルラー電話、スマートフォン、および他の同様のコンピューティングデバイスなど、様々な形のモバイルデバイスを表すことを意図している。さらに、コンピューティングデバイス600または650は、ユニバーサルシリアルバス(USB)フラッシュドライブを含むことができる。USBフラッシュドライブは、オペレーティングシステムおよび他のアプリケーションを格納することができる。USBフラッシュドライブは、別のコンピューティングデバイスのUSBポートに挿入することができるワイヤレストランスミッタまたはUSBコネクタなどの入力/出力構成要素を含むことができる。ここに示される構成要素、それらの接続および関係、ならびにそれらの機能は、例示であることのみを目的としており、本文書において記載および/または請求される本発明の実装形態を限定することは目的としていない。

【００９０】

コンピューティングデバイス600は、プロセッサ602、メモリ604、記憶デバイス606、メモリ604および高速拡張ポート610に接続する高速インターフェース608、ならびに低速バス614および記憶デバイス606に接続する低速インターフェース612を含む。構成要素602、604、606、608、610、および612のそれぞれは、様々なバスを使用して相互接続され、共通マザーボード上にまたは必要に応じて他の方式で搭載されることが可能である。プロセッサ602は、GUIについてのグラフィカル情報を高速インターフェース608に結合されるディスプレイ616などの外部入力/出力デバイス上に表示するためのメモリ604中または記憶デバイス606上に記憶された命令を含むコンピューティングデバイス600内での実行のための命令を処理することができる。他の実装形態では、複数のプロセッサおよび/または複数のバスが、必要に応じて複数のメモリおよび複数のタイプのメモリとともに使用されることが可能である。また、複数のコンピューティングデバイス600が接続されることが可

能であり、各デバイスは、必要な動作の部分をたとえばサーババンク、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提供する。

【００９１】

メモリ604は、コンピューティングデバイス600内に情報を記憶する。一実装形態において、メモリ604は、1つまたは複数の揮発性メモリユニットである。別の実装形態において、メモリ604は、1つまたは複数の不揮発性メモリユニットである。メモリ604はまた、磁気または光ディスクなど、別の形のコンピュータ可読媒体であることが可能である。

【００９２】

記憶デバイス606は、コンピューティングデバイス600に大容量記憶を提供することが可能である。一実装形態において、記憶デバイス606は、フロッピーディスクデバイス、ハードディスクデバイス、光ディスクデバイス、もしくはテープデバイス、フラッシュメモリもしくは他の同様の固体メモリデバイス、または記憶エリアネットワークもしくは他の構成におけるデバイスを含むデバイスのアレイなどのコンピュータ可読媒体であるか、またはそれらを含むことができる。コンピュータプログラム製品を情報キャリア中で有形に実施することができる。コンピュータプログラム製品はまた、実行されると、上述したような1つまたは複数の方法を実施する命令を含むことができる。情報キャリアは、メモリ604、記憶デバイス606、またはプロセッサ602上のメモリなどのコンピュータまたは機械可読媒体である。

【００９３】

高速コントローラ608は、コンピューティングデバイス600向けの帯域消費型動作を管理し、低速コントローラ612は、より帯域低消費型の動作を管理する。機能のそのような割振りは例示にすぎない。一実装形態において、高速コントローラ608は、たとえば、グラフィックスプロセッサまたはアクセラレータを通して、メモリ604、ディスプレイ616に、および様々な拡張カード(図示せず)を収容することができる高速拡張ポート610に結合される。本実装形態において、低速コントローラ612は、記憶デバイス606および低速拡張ポート614に結合される。低速拡張ポートは、様々な通信ポート、たとえば、USB、Bluetooth(登録商標)、イーサネット(登録商標)、ワイヤレスイーサネットを含むことができるが、キーボード、ポインティングデバイス、マイクロフォン/スピーカーペア、スキャナなど、1つもしくは複数の入力/出力デバイス、またはスイッチもしくはルータなどのネットワークデバイスに、たとえば、ネットワークアダプタを通して結合されることが可能である。コンピューティングデバイス600は、図に示すように、いくつかの異なる形で実装されることが可能である。たとえば、標準サーバ620として、またはそのようなサーバのグループで複数回、実装されることが可能である。また、ラックサーバシステム624の一部として実装されることが可能である。さらに、ラップトップコンピュータ622などのパーソナルコンピュータで実装されることが可能である。代替として、コンピューティングデバイス600からの構成要素は、デバイス650などのモバイルデバイス(図示せず)中の他の構成要素と組み合わせることができる。そのようなデバイスのそれぞれは、コンピューティングデバイス600、650のうちの1つまたは複数を含むことができ、システム全体が、互いと通信する複数のコンピューティングデバイス600、650から構成されることが可能である。

【００９４】

コンピューティングデバイス600は、図に示すように、いくつかの異なる形で実装されることが可能である。たとえば、標準サーバ620として、またはそのようなサーバのグループで複数回実装されることが可能である。また、ラックサーバシステム624の一部として実装されることが可能である。さらに、ラップトップコンピュータ622などのパーソナルコンピュータで実装されることが可能である。代替として、コンピューティングデバイス600からの構成要素は、デバイス650などのモバイルデバイス(図示せず)中の他の構成要素と組み合わせることができる。そのようなデバイスのそれぞれは、コンピューティングデバイス600、650のうちの1つまたは複数を含むことができ、システム全体が、互いと通信する複数のコンピューティングデバイス600、650から構成されることが可能である。

【００９５】
コンピューティングデバイス650は、他の構成要素の中でも、プロセッサ652、メモリ664、およびディスプレイ654などの入力/出力デバイス、通信インターフェース666、ならびにトランシーバ668を含む。デバイス650には、追加記憶を提供するためにマイクロドライブまたは他のデバイスなどの記憶デバイスが設けられることが可能である。構成要素650、652、664、654、666、および668のそれぞれは、様々なバスを使用して相互接続され、構成要素のうちのいくつかは、共通マザーボード上に、または必要に応じて他の方式で搭載されることが可能である。
【００９６】
プロセッサ652は、メモリ664中に記憶された命令を含むコンピューティングデバイス650内の命令を実行することができる。プロセッサは、別個および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ることが可能である。さらに、プロセッサは、いくつかのアーキテクチャのうちのどれを使用しても実装することができる。たとえば、プロセッサ652は、CISC(複合命令セットコンピュータ)プロセッサ、RISC(縮小命令セットコンピュータ)プロセッサ、またはMISC(最少命令セットコンピュータ)プロセッサであることが可能である。プロセッサは、たとえば、ユーザインターフェース、デバイス650によって稼動されるアプリケーション、およびデバイス650によるワイヤレス通信の制御など、デバイス650の他の構成要素の調整を可能にすることができる。
【００９７】
プロセッサ652は、制御インターフェース658およびディスプレイ654に結合されるディスプレイインターフェース656を通してユーザと通信することができる。ディスプレイ654は、たとえば、TFT(薄膜トランジスタ液晶ディスプレイ)ディスプレイもしくは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ることが可能である。ディスプレイインターフェース656は、グラフィカルおよび他の情報をユーザに対して提示するようにディスプレイ654を駆動するための適切な回路機構を含むことができる。制御インターフェース658は、コマンドをユーザから受信し、プロセッサ652への提出のために変換することができる。さらに、デバイス650と他のデバイスの近距離通信を可能にするように、プロセッサ652と通信する外部インターフェース662が提供されることが可能である。外部インターフェース662は、たとえば、いくつかの実装形態ではワイヤード通信をまたは他の実装形態ではワイヤレス通信を提供することができ、複数のインターフェースが使用されることも可能である。
【００９８】
メモリ664は、コンピューティングデバイス650内に情報を記憶する。メモリ664は、1つもしくは複数のコンピュータ可読媒体、1つもしくは複数の揮発性メモリユニット、または1つもしくは複数の不揮発性メモリユニットのうちの1つまたは複数として実装されることが可能である。拡張メモリ674が設けられ、拡張インターフェース672を通してデバイス650に接続されることも可能であり、インターフェース672は、たとえば、SIMM(シングルインラインメモリモジュール)カードインターフェースを含むことができる。そのような拡張メモリ674はまた、デバイス650に余剰記憶空間を提供することができ、またはデバイス650向けのアプリケーションもしくは他の情報を記憶することができる。特に、拡張メモリ674は、上述したプロセスを実践し、または補うための命令を含むことができ、セキュアな情報もまた含むことができる。したがって、たとえば、拡張メモリ674は、デバイス650用のセキュリティモジュールとして設けられることが可能であり、デバイス650のセキュアな使用を許可する命令でプログラムされることが可能である。さらに、ハッキングできないようにSIMMカード上に識別情報を置くなど、付加情報とともに、SIMMカードを介して、セキュアなアプリケーションが提供されることが可能である。
【００９９】
メモリは、たとえば、以下で議論するように、フラッシュメモリおよび/またはNVRAMメモリを含むことができる。一実装形態では、コンピュータプログラム製品が、情報キャリア中で有形に実施される。コンピュータプログラム製品は、実行されると、上述したよう

な1つまたは複数の方法を実施する命令を含む。情報キャリアは、たとえば、トランシーバ668または外部インターフェース662を介して受信することができる、メモリ664、拡張メモリ674、またはプロセッサ652上のメモリなどのコンピュータまたは機械可読媒体である。

【０１００】

デバイス650は、必要な場合はデジタル信号処理回路機構を含むことができる通信インターフェース666を通してワイヤレスに通信することができる。通信インターフェース666は、それ以外のものの中でも特に、GSM(登録商標)音声コール、SMS、EMS、またはMMSメッセージ通信、CDMA、TDMA、PDC、WCDMA(登録商標)、CDMA2000、またはGPRSなど、様々なモードまたはプロトコルでの通信を提供することができる。そのような通信は、たとえば、無線周波数トランシーバ668を通して発生させることができる。さらに、たとえばBluetooth(登録商標)、Wi-Fi、または他のそのようなトランシーバ(図示せず)を使用して短距離通信を発生させることができる。さらに、GPS(全地球測位システム)レシーバモジュール670が、追加ナビゲーションおよびロケーション関連ワイヤレスデータをデバイス650に提供することができ、このデータは、必要に応じてデバイス650上で稼動するアプリケーションによって使用されることが可能である。

【０１０１】

デバイス650はまた、オーディオコーデック660を使用して可聴的に通信することができ、コーデック660は、発話情報をユーザから受信し、使用可能なデジタル情報に変換することができる。オーディオコーデック660は同様に、たとえば、デバイス650のハンドセット中のスピーカーを通すなどして、ユーザ向けの可聴音を生成することができる。そのような音は、音声通話からの音を含むことができ、記録された音、たとえば、音声メッセージ、音楽ファイルなどを含むことができ、デバイス650上で動作するアプリケーションによって生成された音もまた含むことができる。

【０１０２】

コンピューティングデバイス650は、図に示すように、いくつかの異なる形で実装されることが可能である。たとえば、セルラー電話680として実装されることが可能である。また、スマートフォン682、携帯情報端末、または他の同様のモバイルデバイスの一部として実装されることが可能である。

【０１０３】

ここに記載するシステムおよび方法の様々な実装形態は、デジタル電子回路機構、集積回路機構、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/またはそのような実装形態の組合せで実現されることが可能である。これらの様々な実装形態は、少なくとも1つのプログラム可能プロセッサを含むプログラム可能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むことができ、プログラム可能プロセッサは、記憶システム、少なくとも1つの入力デバイス、および少なくとも1つの出力デバイスからデータおよび命令を受信するように、ならびにそれらにデータおよび命令を送信するように結合される、専用または汎用であることができる。

【０１０４】

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、プログラム可能プロセッサ用の機械命令を含み、高度手続型および/もしくはオブジェクト指向プログラミング言語で、ならびに/またはアセンブリ/機械言語で実装されることが可能である。本明細書で使用する「機械可読媒体」「コンピュータ可読媒体」という用語は、機械命令を機械可読信号として受信する機械可読媒体を含むプログラム可能プロセッサに機械命令および/またはデータを提供するために使用されるどのコンピュータプログラム製品、装置および/またはデバイス、たとえば、磁気ディスク、光ディスク、メモリ、プログラム可能論理デバイス(PLD)も指す。「機械可読信号」という用語は、プログラム可能プロセッサに機械命令および/またはデータを提供するために使用されるどの信号も指す。

【０１０５】
ユーザとの対話を可能にするために、ここで記載するシステムおよび技法は、ユーザに情報を表示するためのディスプレイデバイス、たとえば、CRT(陰極線管)やLCD(液晶ディスプレイ)モニタと、ユーザがコンピュータに入力を提供することができるためのキーボードおよびポインティングデバイス、たとえば、マウスやトラックボールとを有するコンピュータ上で実装することができる。他の種類のデバイスも、ユーザとの対話を可能にするために使用されることが可能であり、たとえば、ユーザに提供されるフィードバックは、どの形の感覚フィードバックでも、たとえば、視覚フィードバック、聴覚フィードバック、または触覚フィードバックであることができ、ユーザからのインプットは、音響、発話、または触覚インプットを含むどの形でも受信されることが可能である。

【０１０６】
ここで記載するシステムおよび技法は、バックエンド構成要素をたとえばデータサーバとして含む、もしくはミドルウェア構成要素、たとえばアプリケーションサーバを含む、もしくはフロントエンド構成要素、たとえば、ここで記載されるシステムおよび技法の実装形態とユーザが対話することができるためのグラフィカルユーザインターフェースもしく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含むコンピューティングシステム、またはそのようなバックエンド、ミドルウェア、もしくはフロントエンド構成要素のどの組合せでも実装されることが可能である。システムの構成要素は、どの形または媒体のデジタルデータ通信(たとえば、通信ネットワーク)によっても相互接続されることが可能である。通信ネットワークの例には、ローカルエリアネットワーク("LAN")、ワイドエリアネットワーク("WAN")、およびインターネットがある。

【０１０７】
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むことができる。クライアントとサーバは概して、互いから離れており、通常、通信ネットワークを通して対話する。クライアントとサーバの関係は、それぞれのコンピュータ上で稼動するとともに互いとのクライアント-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。

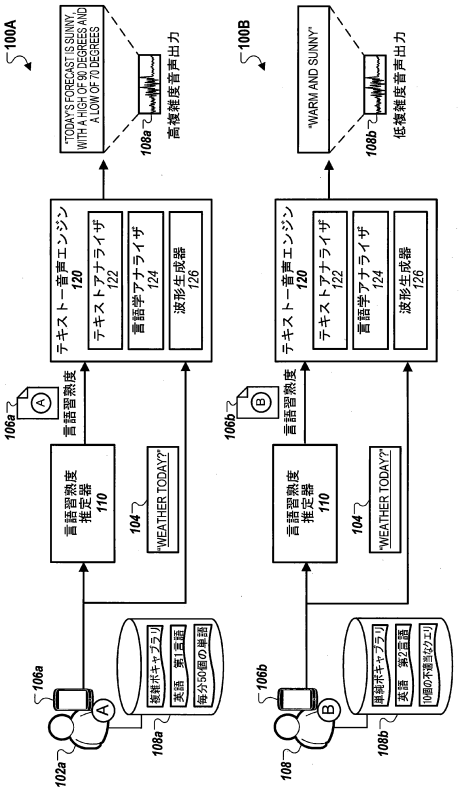
【０１０８】
いくつかの実施形態が記載された。それにも関わらず、本発明の趣旨および範囲から逸脱することなく、様々な修正を行うことができることが理解されよう。さらに、図面に示される論理フローは、望ましい結果を達成するために図示される特定の順序、または順番を求めない。さらに、他のステップが提供されることが可能であり、またはステップが、記載したフローからなくされることが可能であり、他の構成要素が、記載したシステムに追加され、もしくはそこから削除されることが可能である。したがって、他の実施形態は、以下の特許請求の範囲内である。

【符号の説明】

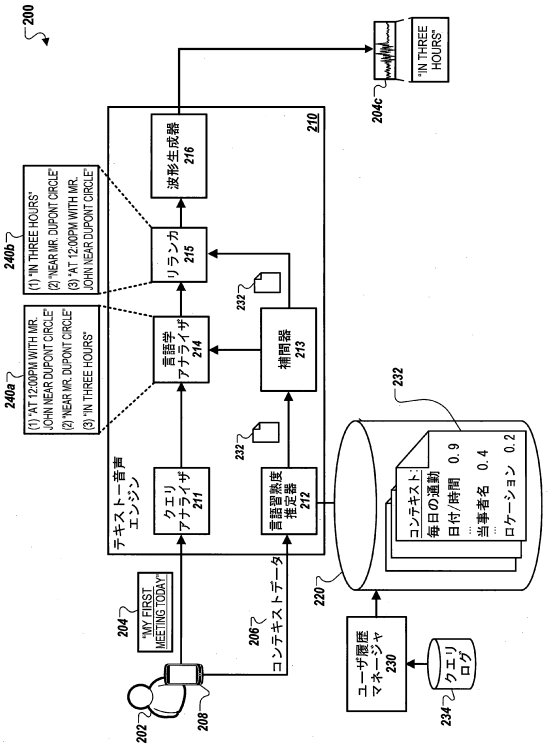
【０１０９】
100A　プロセス
100B　プロセス
104　テキストクエリ、クエリ
106a　ユーザデバイス
106b　ユーザデバイス
110　言語習熟度推定器
120　テキスト－音声エンジン、TTSエンジン
122　テキストアナライザ
124　言語学アナライザ
126　波形生成器
200　システム
208　ユーザデバイス
210　TTSエンジン
211　クエリアナライザ

212 言語習熟度推定器
213 補間器
214 言語学アナライザ
215 リランカ
216 波形生成器
220 コンテキストリポジトリ、リポジトリ
230 ユーザ履歴マネージャ
232 コンテキストプロファイル
234 ユーザ履歴データ、クエリログ
300 システム
302 クエリ
304 言語習熟度プロファイル
310 TTSエンジン
400 システム
410 言語習熟度推定器
412 言語複雑度プロファイル
420 ユーザ類似度判定器
422 クラスタデータ
430 複雑度オプティマイザ
440 機械学習システム
600 コンピューティングデバイス
602 プロセッサ、構成要素
604 メモリ、構成要素
606 記憶デバイス、構成要素
608 高速インターフェース、構成要素、高速コントローラ
610 高速拡張ポート、構成要素
612 低速インターフェース、構成要素、低速コントローラ
614 低速バス、低速拡張ポート
616 ディスプレイ
620 標準サーバ
622 ラップトップコンピュータ
624 ラックサーバシステム
650 コンピューティングデバイス、デバイス、構成要素
652 プロセッサ、構成要素
654 ディスプレイ、構成要素
656 ディスプレイインターフェース
658 制御インターフェース
660 オーディオコーデック
662 外部インターフェース
664 メモリ、構成要素
666 通信インターフェース、構成要素
668 トランシーバ、構成要素、無線周波数トランシーバ
670 GPS(全地球測位システム)レシーバモジュール
672 拡張インターフェース
674 拡張メモリ
680 セルラー電話
682 スマートフォン

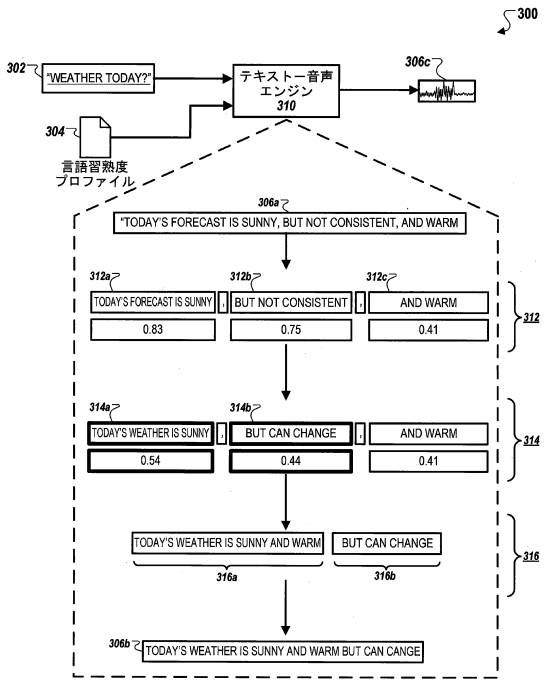
【図１】


100A
"TODAY'S FORECAST IS SUNNY, WITH A HIGH OF 90 DEGREES AND A LOW OF 70 DEGREES"
108a
高複雑度音声出力
100B
"WARM AND SUNNY"
108b
低複雑度音声出力
テキスト－音声エンジン 120
テキストアナライザ 122
言語学アナライザ 124
波形生成器 126
106a
言語習熟度
言語習熟度推定器 110
104
"WEATHER TODAY?"
106b
102a
108
108a
108b

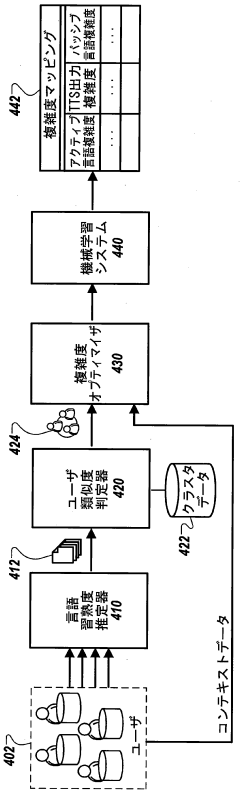
【図２】


200
"IN THREE HOURS"
204c
210
波形生成器 216
リランカ 215
言語学アナライザ 214
補間器 213
言語習熟度推定器 212
クエリアナライザ 211
テキスト－音声エンジン
240b
(1) "IN THREE HOURS"
(2) "NEAR MR. DUPONT CIRCLE"
(3) "AT 12:00PM WITH MR. JOHN NEAR DUPONT CIRCLE"
240a
(1) "AT 12:00PM WITH MR. JOHN NEAR DUPONT CIRCLE"
(2) "NEAR MR. DUPONT CIRCLE"
(3) "IN THREE HOURS"
232
コンテキスト:
毎日の通勤
日付/時間 0.9
当事者名 0.4
ロケーション 0.2
204
"MY FIRST MEETING TODAY"
206
コンテキストデータ
220
ユーザ履歴マネージャ 230
クエリログ
234
202
208

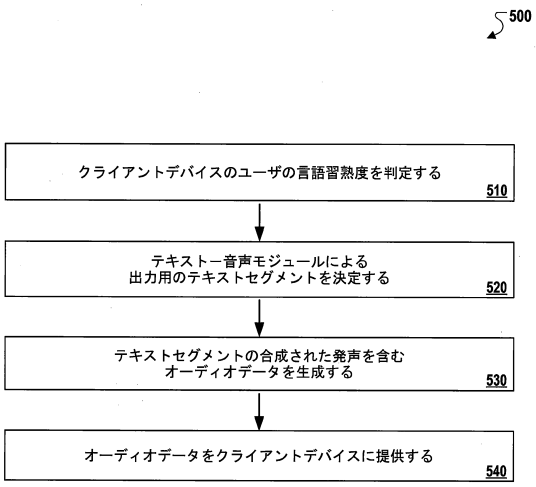
【図３】


300
302
"WEATHER TODAY?"
テキスト－音声エンジン 310
306c
304
言語習熟度プロファイル
306a
"TODAY'S FORECAST IS SUNNY, BUT NOT CONSISTENT, AND WARM
312a
TODAY'S FORECAST IS SUNNY
312b
BUT NOT CONSISTENT
312c
AND WARM
0.83
0.75
0.41
312
314a
TODAY'S WEATHER IS SUNNY
314b
BUT CAN CHANGE
AND WARM
0.54
0.44
0.41
314
TODAY'S WEATHER IS SUNNY AND WARM
BUT CAN CHANGE
316a
316b
316
306b
TODAY'S WEATHER IS SUNNY AND WARM BUT CAN CANGE

【図４】


400
442
複雑度マッピング
アクティブ言語複雑度
TTS出力複雑度
パッシブ言語複雑度
機械学習システム 440
複雑度オプティマイザ 430
424
ユーザ類似度判定器 420
クラスタデータ
422
412
言語習熟度推定器 410
コンテキストデータ
402
ユーザ

【図５】


500
クライアントデバイスのユーザの言語習熟度を判定する
510
テキスト－音声モジュールによる
出力用のテキストセグメントを決定する
520
テキストセグメントの合成された発声を含む
オーディオデータを生成する
530
オーディオデータをクライアントデバイスに提供する
540

【図６】

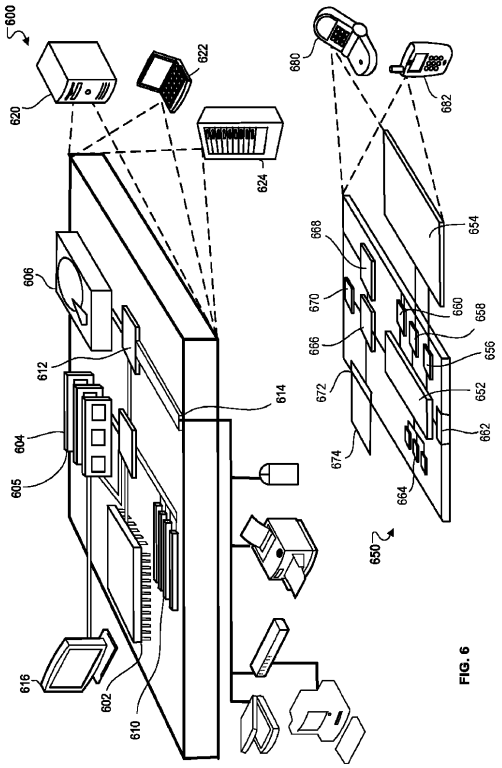

600
620
622
624
680
682
606
612
614
604
605
616
602
610
668
654
670
660
658
666
656
672
652
674
662
664
650

FIG. 6

フロントページの続き

(72)発明者　ジェイコブ・ニコラウス・フォースター
アメリカ合衆国・カリフォルニア・９４０４３・マウンテン・ヴュー・アンフィシアター・パークウェイ・１６００

審査官　早川　学

(56)参考文献　特開平０３－２２５３８９（ＪＰ，Ａ）
米国特許出願公開第２００４／０１９３４２１（ＵＳ，Ａ１）
特開２０１０－１４５８７３（ＪＰ，Ａ）
特開平０３－０３５２９６（ＪＰ，Ａ）
特開２０１４－１９９３２３（ＪＰ，Ａ）
特開２０１４－０９３０５０（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　４０／００－４０／５８
Ｇ１０Ｌ　１３／００－９９／００