(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7171911号
(P7171911)

(45)発行日　令和4年11月15日(2022.11.15)

(24)登録日　令和4年11月7日(2022.11.7)

(51)国際特許分類		FI		
G06F 3/16 (2006.01)		G06F	3/16	650
G10L 15/10 (2006.01)		G06F	3/16	690
G10L 15/28 (2013.01)		G10L	15/10	200W
		G10L	15/28	230K

請求項の数　20　（全55頁）

(21)出願番号　特願2021-520598(P2021-520598)
(86)(22)出願日　令和2年6月9日(2020.6.9)
(65)公表番号　特表2022-540263(P2022-540263A)
(43)公表日　令和4年9月15日(2022.9.15)
(86)国際出願番号　PCT/US2020/036749
(87)国際公開番号　WO2021/251953
(87)国際公開日　令和3年12月16日(2021.12.16)
審査請求日　令和3年6月14日(2021.6.14)

(73)特許権者　502208397
グーグル エルエルシー
Ｇｏｏｇｌｅ ＬＬＣ
アメリカ合衆国 カリフォルニア州 ９４０４３ マウンテン ビュー アンフィシアター パークウェイ １６００
１６００ Ａｍｐｈｉｔｈｅａｔｒｅ Ｐａｒｋｗａｙ ９４０４３ Ｍｏｕｎｔａｉｎ Ｖｉｅｗ， ＣＡ Ｕ．Ｓ．Ａ．
(74)代理人　100108453
弁理士　村山 靖彦
(74)代理人　100110364
弁理士　実広 信哉
(74)代理人　100133400
弁理士　阿部 達彦

最終頁に続く

(54)【発明の名称】　ビジュアルコンテンツからのインタラクティブなオーディオトラックの生成

(57)【特許請求の範囲】

【請求項１】

異なる様式の間を遷移するためのシステムであって、

データ処理システムであって、

ネットワークを介して、データ処理システムの遠隔のコンピューティングデバイスのマイクロフォンによって検出された入力オーディオ信号を含むデータパケットを受信すること、

要求を特定するために前記入力オーディオ信号を解析すること、

前記要求に基づいて、ビジュアル出力フォーマットを有するデジタルコンポーネントオブジェクトを選択することであって、前記デジタルコンポーネントオブジェクトが、メタデータに関連付けられる、選択すること、

前記コンピューティングデバイスの種類に基づいて、前記デジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換すると決定すること、

前記デジタルコンポーネントオブジェクトを前記オーディオ出力フォーマットに変換する決定に応じて、前記デジタルコンポーネントオブジェクトに関するテキストを生成すること、

前記デジタルコンポーネントオブジェクトのコンテキストに基づいて、前記テキストをレンダリングするためのデジタル音声を選択すること、

前記デジタル音声によってレンダリングされた前記テキストを用いて前記デジタルコンポーネントオブジェクトの基礎となるオーディオトラックを構築すること、

前記デジタルコンポーネントオブジェクトに基づいて、口頭でないオーディオの合図を生成すること、
前記デジタルコンポーネントオブジェクトのオーディオトラックを生成するために前記口頭でないオーディオの合図を前記デジタルコンポーネントオブジェクトの前記基礎となるオーディオの形態と組み合わせること、および
前記コンピューティングデバイスからの前記要求に応じて、前記コンピューティングデバイスのスピーカを介して出力するために前記コンピューティングデバイスに前記デジタルコンポーネントオブジェクトの前記オーディオトラックを提供することを行うための1つまたは複数のプロセッサを含む、データ処理システムを含む、システム。

【請求項２】
スマートスピーカを含む前記コンピューティングデバイスの前記種類に基づいて前記デジタルコンポーネントオブジェクトを前記オーディオ出力フォーマットに変換すると決定するための前記データ処理システムを含む請求項1に記載のシステム。

【請求項３】
デジタルアシスタントを含む前記コンピューティングデバイスの前記種類に基づいて前記デジタルコンポーネントオブジェクトを前記オーディオ出力フォーマットに変換すると決定するための前記データ処理システムを含む請求項1に記載のシステム。

【請求項４】
前記要求に応じて、リアルタイムコンテンツ選択プロセスに入力されたコンテンツ選択基準に基づいて前記デジタルコンポーネントオブジェクトを選択するための前記データ処理システムを含み、前記デジタルコンポーネントオブジェクトが、複数のサードパーティのコンテンツプロバイダによって提供された複数のデジタルコンポーネントオブジェクトから選択される請求項1に記載のシステム。

【請求項５】
前記要求の前に前記コンピューティングデバイスによってレンダリングされたコンテンツに関連するキーワードに基づいて前記デジタルコンポーネントオブジェクトを選択するための前記データ処理システムを含み、前記デジタルコンポーネントオブジェクトが、複数のサードパーティのコンテンツプロバイダによって提供された複数のデジタルコンポーネントオブジェクトから選択される請求項1に記載のシステム。

【請求項６】
自然言語生成モデルによって、前記デジタルコンポーネントオブジェクトの前記メタデータに基づいて前記デジタルコンポーネントオブジェクトに関する前記テキストを生成するための前記データ処理システムを含む請求項1に記載のシステム。

【請求項７】
音声モデルによって、前記デジタルコンポーネントオブジェクトのコンテキストに基づいて前記デジタル音声を選択するための前記データ処理システムを含み、前記音声モデルが、オーディオおよびビジュアルメディアコンテンツを含む履歴的なデータセットを用いて機械学習技術によって訓練される請求項1に記載のシステム。

【請求項８】
音声の特性のベクトルを生成するために、音声モデルに前記デジタルコンポーネントオブジェクトの前記コンテキストを入力することであって、前記音声モデルが、オーディオおよびビジュアルメディアコンテンツを含む履歴的なデータセットを用いて機械学習エンジンによって訓練される、入力すること、ならびに
前記音声の特性のベクトルに基づいて複数のデジタル音声から前記デジタル音声を選択することを行うための前記データ処理システムを含む請求項1に記載のシステム。

【請求項９】
前記メタデータに基づいて、前記オーディオトラックにトリガワードを追加すると決定するための前記データ処理システムを含み、第2の入力オーディオ信号内の前記トリガワードの検出が、前記データ処理システムまたは前記コンピューティングデバイスに前記トリガワードに対応するデジタルアクションを実行させる請求項1に記載のシステム。

【請求項１０】
前記デジタルコンポーネントオブジェクトのカテゴリを決定し、
データベースから、前記カテゴリに関連する複数のデジタルアクションに対応する複数のトリガワードを取り出し、
トリガキーワードの履歴的な実行に基づいて訓練されたデジタルアクションモデルを使用して、前記デジタルコンポーネントオブジェクトの前記コンテキストおよび前記コンピューティングデバイスの前記種類に基づいて前記複数のトリガワードをランク付けし、
前記オーディオトラックに追加するために最も高いランク付けのトリガキーワードを選択するための前記データ処理システムを含む請求項1に記載のシステム。
【請求項１１】
前記デジタルコンポーネントオブジェクト内のビジュアルオブジェクトを特定するために前記デジタルコンポーネントオブジェクトに対して画像認識を実行し、
データベースに記憶された複数の口頭でないオーディオの合図から前記ビジュアルオブジェクトに対応する前記口頭でないオーディオの合図を選択するための前記データ処理システムを含む請求項1に記載のシステム。
【請求項１２】
画像認識技術によって前記デジタルコンポーネントオブジェクト内の複数のビジュアルオブジェクトを特定し、
前記メタデータおよび前記複数のビジュアルオブジェクトに基づいて、複数の口頭でないオーディオの合図を選択し、
前記ビジュアルオブジェクトの各々と前記メタデータとの間の一致のレベルを示す前記ビジュアルオブジェクトの各々に関する一致スコアを決定し、
前記一致スコアに基づいて前記複数の口頭でないオーディオの合図をランク付けし、
前記複数の口頭でないオーディオの合図の各々と前記テキストをレンダリングするために前記コンテキストに基づいて選択された前記デジタル音声との間のオーディオの干渉のレベルを判定し、
最も高いランクに基づいて、閾値未満のオーディオの干渉の前記レベルに関連する前記複数の口頭でないオーディオの合図から前記口頭でないオーディオの合図を選択するための前記データ処理システムを含む請求項1に記載のシステム。
【請求項１３】
履歴的な実行データを使用して訓練された挿入モデルに基づいて、前記コンピューティングデバイスによって出力されるデジタルメディアストリーム内の前記オーディオトラックに関する挿入点を特定し、
前記コンピューティングデバイスに前記デジタルメディアストリーム内の前記挿入点において前記オーディオトラックをレンダリングさせるために前記コンピューティングデバイスに命令を与えるための前記データ処理システムを含む請求項1に記載のシステム。
【請求項１４】
異なる様式の間を遷移するための方法であって、
ネットワークを介してデータ処理システムの1つまたは複数のプロセッサによって、前記データ処理システムの遠隔のコンピューティングデバイスのマイクロフォンによって検出された入力オーディオ信号を含むデータパケットを受信するステップと、
前記データ処理システムによって、要求を特定するために前記入力オーディオ信号を解析するステップと、
前記要求に基づいて前記データ処理システムによって、ビジュアル出力フォーマットを有するデジタルコンポーネントオブジェクトを選択するステップであって、前記デジタルコンポーネントオブジェクトが、メタデータに関連付けられる、ステップと、
前記コンピューティングデバイスの種類に基づいて前記データ処理システムによって、前記デジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換すると決定するステップと、
前記デジタルコンポーネントオブジェクトを前記オーディオ出力フォーマットに変換す

る決定に応じて前記データ処理システムによって、前記デジタルコンポーネントオブジェクトに関するテキストを生成するステップと、
前記デジタルコンポーネントオブジェクトのコンテキストに基づいて前記データ処理システムによって、前記テキストをレンダリングするためのデジタル音声を選択するステップと、
前記データ処理システムによって、前記デジタル音声によってレンダリングされた前記テキストを用いて前記デジタルコンポーネントオブジェクトの基礎となるオーディオトラックを構築するステップと、
前記デジタルコンポーネントオブジェクトに基づいて前記データ処理システムによって、口頭でないオーディオの合図を生成するステップと、
前記データ処理システムによって、前記デジタルコンポーネントオブジェクトのオーディオトラックを生成するために前記口頭でないオーディオの合図を前記デジタルコンポーネントオブジェクトの前記基礎となるオーディオの形態と組み合わせるステップと、
前記コンピューティングデバイスからの前記要求に応じて前記データ処理システムによって、前記コンピューティングデバイスのスピーカを介して出力するために前記コンピューティングデバイスに前記デジタルコンポーネントオブジェクトの前記オーディオトラックを提供するステップとを含む、方法。
【請求項１５】
前記データ処理システムによって、スマートスピーカを含む前記コンピューティングデバイスの前記種類に基づいて前記デジタルコンポーネントオブジェクトを前記オーディオ出力フォーマットに変換すると決定するステップを含む請求項14に記載の方法。
【請求項１６】
前記要求に応じて前記データ処理システムによって、リアルタイムコンテンツ選択プロセスに入力されたコンテンツ選択基準に基づいて前記デジタルコンポーネントオブジェクトを選択するステップであって、前記デジタルコンポーネントオブジェクトが、複数のサードパーティのコンテンツプロバイダによって提供された複数のデジタルコンポーネントオブジェクトから選択される、ステップを含む請求項14に記載の方法。
【請求項１７】
前記データ処理システムによって、前記要求の前に前記コンピューティングデバイスによってレンダリングされたコンテンツに関連するキーワードに基づいて前記デジタルコンポーネントオブジェクトを選択するステップであって、前記デジタルコンポーネントオブジェクトが、複数のサードパーティのコンテンツプロバイダによって提供された複数のデジタルコンポーネントオブジェクトから選択される、ステップを含む請求項14に記載の方法。
【請求項１８】
異なる様式の間を遷移するためのシステムであって、
データ処理システムであって、
コンピューティングデバイスによってレンダリングされるデジタルストリーミングコンテンツに関連するキーワードを特定すること、
前記キーワードに基づいて、ビジュアル出力フォーマットを有するデジタルコンポーネントオブジェクトを選択することであって、前記デジタルコンポーネントオブジェクトが、メタデータに関連付けられる、選択すること、
前記コンピューティングデバイスの種類に基づいて、前記デジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換すると決定すること、
前記デジタルコンポーネントオブジェクトを前記オーディオ出力フォーマットに変換する決定に応じて、前記デジタルコンポーネントオブジェクトに関するテキストを生成すること、
前記デジタルコンポーネントオブジェクトのコンテキストに基づいて、前記テキストをレンダリングするためのデジタル音声を選択すること、
前記デジタル音声によってレンダリングされた前記テキストを用いて前記デジタルコン

ポーネントオブジェクトの基礎となるオーディオトラックを構築すること、
前記デジタルコンポーネントオブジェクトの前記メタデータに基づいて、口頭でないオーディオの合図を生成すること、
前記デジタルコンポーネントオブジェクトのオーディオトラックを生成するために前記口頭でないオーディオの合図を前記デジタルコンポーネントオブジェクトの前記基礎となるオーディオの形態と組み合わせること、および
前記コンピューティングデバイスのスピーカを介して出力するために前記コンピューティングデバイスに前記デジタルコンポーネントオブジェクトの前記オーディオトラックを提供することを行うための1つまたは複数のプロセッサを含む、データ処理システムを含む、システム。

【請求項１９】
スマートスピーカを含む前記コンピューティングデバイスの前記種類に基づいて前記デジタルコンポーネントオブジェクトを前記オーディオ出力フォーマットに変換すると決定するための前記データ処理システムを含む請求項18に記載のシステム。

【請求項２０】
リアルタイムコンテンツ選択プロセスに入力された前記キーワードに基づいて前記デジタルコンポーネントオブジェクトを選択するための前記データ処理システムを含み、前記デジタルコンポーネントオブジェクトが、複数のサードパーティのコンテンツプロバイダによって提供された複数のデジタルコンポーネントオブジェクトから選択される請求項19に記載のシステム。

【発明の詳細な説明】
【背景技術】
【０００１】
データ処理システムは、コンピューティングデバイスにデジタルコンテンツを提示させるためにコンピューティングデバイスにデジタルコンテンツを提供し得る。デジタルコンテンツは、コンピューティングデバイスがディスプレイを介して提示することができるビジュアルコンテンツを含み得る。デジタルコンテンツは、コンピューティングデバイスがスピーカを介して出力することができるオーディオコンテンツを含み得る。

【発明の概要】
【課題を解決するための手段】
【０００２】
この技術的な解決策の少なくとも1つの態様は、オーディオトラックを生成するためのシステムを対象とする。システムは、データ処理システムを含み得る。データ処理システムは、1つまたは複数のプロセッサを含み得る。データ処理システムは、ネットワークを介して、データ処理システムの遠隔のコンピューティングデバイスのマイクロフォンによって検出された入力オーディオ信号を含むデータパケットを受信し得る。データ処理システムは、要求を特定するために入力オーディオ信号を解析し得る。データ処理システムは、要求に基づいて、ビジュアル出力フォーマットを有するデジタルコンポーネントオブジェクトを選択することが可能であり、デジタルコンポーネントオブジェクトは、メタデータに関連付けられる。データ処理システムは、コンピューティングデバイスの種類に基づいて、デジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換すると決定し得る。データ処理システムは、デジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換する決定に応じて、デジタルコンポーネントオブジェクトに関するテキストを生成し得る。データ処理システムは、デジタルコンポーネントオブジェクトのコンテキストに基づいて、テキストをレンダリングするためのデジタル音声を選択し得る。データ処理システムは、デジタル音声によってレンダリングされたテキストを用いてデジタルコンポーネントオブジェクトの基礎となるオーディオトラックを構築し得る。データ処理システムは、デジタルコンポーネントオブジェクトのメタデータに基づいて、口頭でないオーディオの合図(audio cue)を生成し得る。データ処理システムは、デジタルコンポーネントオブジェクトのオーディオトラックを生成するために口頭でないオーディ

オの合図をデジタルコンポーネントオブジェクトの基礎となるオーディオの形態と組み合わせ得る。データ処理システムは、コンピューティングデバイスからの要求に応じて、コンピューティングデバイスのスピーカを介して出力するためにコンピューティングデバイスにデジタルコンポーネントオブジェクトのオーディオトラックを提供し得る。

【０００３】

この技術的な解決策の少なくとも1つの態様は、オーディオトラックを生成する方法を対象とする。方法は、データ処理システムの1つまたは複数のプロセッサによって実行され得る。方法は、データ処理システムがデータ処理システムの遠隔のコンピューティングデバイスのマイクロフォンによって検出された入力オーディオ信号を含むデータパケットを受信するステップを含み得る。方法は、データ処理システムが要求を特定するために入力オーディオ信号を解析するステップを含み得る。方法は、データ処理システムが要求に基づいてビジュアル出力フォーマットを有するデジタルコンポーネントオブジェクトを選択するステップであって、デジタルコンポーネントオブジェクトがメタデータに関連付けられる、ステップを含み得る。方法は、データ処理システムがコンピューティングデバイスの種類に基づいてデジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換すると決定するステップを含み得る。方法は、データ処理システムが、デジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換する決定に応じて、デジタルコンポーネントオブジェクトに関するテキストを生成するステップを含み得る。方法は、データ処理システムが、デジタルコンポーネントオブジェクトのコンテキストに基づいて、テキストをレンダリングするためのデジタル音声を選択するステップを含み得る。方法は、データ処理システムがデジタル音声によってレンダリングされたテキストを用いてデジタルコンポーネントオブジェクトの基礎となるオーディオトラックを構築するステップを含み得る。方法は、データ処理システムがデジタルコンポーネントオブジェクトに基づいて口頭でないオーディオの合図を生成するステップを含み得る。方法は、データ処理システムが、デジタルコンポーネントオブジェクトのオーディオトラックを生成するために口頭でないオーディオの合図をデジタルコンポーネントオブジェクトの基礎となるオーディオの形態と組み合わせるステップを含み得る。方法は、データ処理システムが、コンピューティングデバイスからの要求に応じて、コンピューティングデバイスのスピーカを介して出力するためにコンピューティングデバイスにデジタルコンポーネントオブジェクトのオーディオトラックを提供するステップを含み得る。

【０００４】

この技術的な解決策の少なくとも1つの態様は、オーディオトラックを生成するためのシステムを対象とする。システムは、1つまたは複数のプロセッサを有するデータ処理システムを含み得る。データ処理システムは、コンピューティングデバイスによってレンダリングされるデジタルストリーミングコンテンツに関連するキーワードを特定し得る。データ処理システムは、キーワードに基づいて、ビジュアル出力フォーマットを有するデジタルコンポーネントオブジェクトを選択することが可能であり、デジタルコンポーネントオブジェクトは、メタデータに関連付けられる。データ処理システムは、コンピューティングデバイスの種類に基づいて、デジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換すると決定し得る。データ処理システムは、デジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換する決定に応じて、デジタルコンポーネントオブジェクトに関するテキストを生成し得る。データ処理システムは、デジタルコンポーネントオブジェクトのコンテキストに基づいて、テキストをレンダリングするためのデジタル音声を選択し得る。データ処理システムは、デジタル音声によってレンダリングされたテキストを用いてデジタルコンポーネントオブジェクトの基礎となるオーディオトラックを構築し得る。データ処理システムは、デジタルコンポーネントオブジェクトに基づいて、口頭でないオーディオの合図を生成し得る。データ処理システムは、デジタルコンポーネントオブジェクトのオーディオトラックを生成するために口頭でないオーディオの合図をデジタルコンポーネントオブジェクトの基礎となるオーディオの形態と組み合わせ得る。データ処理システムは、コンピューティングデバイスのスピーカを介して出力

するためにコンピューティングデバイスにデジタルコンポーネントオブジェクトのオーディオトラックを提供し得る。

【０００５】

これらのおよびその他の態様および実装が、下で詳細に検討される。上述の情報および下の詳細な説明は、様々な態様および実装の説明のための例を含み、主張される態様および実装の本質および特徴を理解するための概要または枠組みを提供する。図面は、様々な態様および実装を例示し、さらに理解させ、本明細書の一部に組み込まれ、本明細書の一部を構成する。

【０００６】

添付の図面は、正しい縮尺で描かれるように意図されていない。様々な図面における同様の参照番号および参照指示は、同様の要素を示す。明瞭にする目的で、あらゆる図面においてあらゆるコンポーネントがラベル付けされるとは限らない可能性がある。

【図面の簡単な説明】

【０００７】

【図１】実装によるオーディオトラックを生成するためのシステムの図である。

【図２】実装によるオーディオトラックを生成する方法の図である。

【図３】図1に示されるシステムおよび図2に示される方法の要素を実装するために使用され得るコンピュータシステムのための大まかなアーキテクチャを示すブロック図である。

【発明を実施するための形態】

【０００８】

以下は、オーディオトラックを生成する方法、装置、およびシステムに関連する様々な概念ならびにそれらの方法、装置、およびシステムの実装のより詳細な説明である。たとえば、方法、装置、およびシステムは、ビジュアルコンテンツからオーディオトラックを生成することができる。上で導入され、下でより詳細に検討される様々な概念は、多数の方法のいずれかで実装される可能性がある。

【０００９】

この技術的な解決策は、概して、オーディオトラックを生成することを対象とする。技術的な解決策のシステムおよび方法は、ビジュアルコンテンツを処理して、口頭のおよび口頭でないオーディオの合図をともなうオーディオトラックを生成することができる。たとえば、特定の種類のコンピューティングデバイスは、オーディオのみのインターフェースを提供する(たとえば、ユーザから音声入力を受け取り、入力を処理し、デジタル音声によるオーディオ出力または口頭の出力を提供する)可能性がある。特定のコンピューティングデバイスは、オーディオユーザインターフェースを主に使用する可能性があり、または特定の状況においてオーディオインターフェースを主に使用する可能性がある。たとえば、モバイルコンピューティングデバイスのユーザは、乗り物を運転している間、走っている間、または音楽ストリーミングサービスを聞いている間、オーディオのみのインターフェースを主に使用する可能性がある。主なインターフェースがオーディオに基づくとき、データ処理システムは、オーディオデジタルコンポーネントオブジェクト(たとえば、オーディオコンテンツアイテム)を提供し得る。たとえば、データ処理システムは、サードパーティのオーディオコンテンツプロバイダによって確立されたまたは提供されたオーディオコンテンツアイテムを選択し得る。データ処理システムは、ユーザからのコンテンツ要求に応じてまたは別のトリガイベントに基づいてオーディオコンテンツアイテムを提供し得る。しかし、サードパーティのコンテンツプロバイダによって確立されたコンテンツアイテムは、オーディオコンテンツアイテムではない可能性がある。データ処理システムは、キーワード、関連性、またはその他の要因などのマッチング基準に基づいてそのようなコンテンツアイテムを選択することを決定する可能性がある。しかし、データ処理システムは、コンピューティングデバイスがオーディオに基づくインターフェースしか持っていないためにコンピューティングデバイスにコンテンツアイテムを提供することができない可能性がある。あるいは、場合によっては、コンピューティングデバイスがオーディオインターネットを主に利用するか、またはオーディオインターフェースが最も効率的なインタ

ーフェースである場合、データ処理システムは、ビジュアルコンテンツアイテムを提供し、コンピューティングデバイスにコンピューティングデバイスのディスプレイを使用してビジュアルコンテンツアイテムをレンダリングさせることによって非効率的なもしくは無駄なコンピュータの使用または否定的なユーザエクスペリエンスを引き起こし得る。ディスプレイを使用することは、モバイルコンピューティングデバイス(たとえば、スマートフォン、スマートウォッチ、またはその他のウェアラブルデバイス)のバッテリ電力を無駄にする可能性がある。したがって、データ処理システムは、オーディオコンテンツが好ましい場合にビジュアルコンテンツを提供すること、または最も関連性の高いコンテンツがビジュアルフォーマットでのみ利用可能であるためにそのコンテンツを提供することができないことにより、モバイルコンピューティングデバイスによる無駄なコンピューティングリソースの利用または低下したユーザエクスペリエンスをもたらす可能性がある。

【００１０】

さらに、たとえば、コンテンツアイテムを生成すべきフォーマットの決定、何らかのテキストを含む可能性がありまたはいかなるテキストも含まない可能性があるビジュアルコンテンツアイテムに関して口頭で伝えられるテキストをどのようにして正確に生成するのか、生成された口頭で伝えられるテキストのための適切な音声の選択、および口頭でないオーディオの合図の追加を含む様々な技術的な問題が原因で、コンテンツアイテムを異なるフォーマットで生成することは技術的に難しくなり得る。この技術的な解決策のシステムおよび方法は、フォーマット(たとえば、オーディオのみ、オーディオビジュアルフォーマット、ならびにコンピューティングデバイスの種類およびコンピューティングデバイスの現在のコンテキストに基づくインタラクションのモード)を選択し、ビジュアルコンテンツアイテムおよび関連するメタデータに基づいてテキストを自動的に生成し、生成された口頭で伝えられるテキストのための適切なデジタル声紋(voice print)を選択し、口頭でないオーディオの合図を選択し、口頭で伝えられるテキストと一緒に提供するために、機械学習技術および履歴的なデータを使用して訓練された自然言語処理およびモデルを使用することができる。

【００１１】

たとえば、コンピューティングデバイスは、ビジュアルユーザインターフェース(たとえば、ユーザ入力のためのタッチスクリーンを備えたディスプレイスクリーン)とオーディオに基づくユーザインターフェース(たとえば、マイクロフォンおよびスピーカ)との両方を用いて構成され得る。コンピューティングデバイスは、コンピューティングデバイスに関連するスピーカを介して出力するための音楽を現在ストリーミングしていることがあり得る。データ処理システムは、要求、クエリ、またはストリーミング音楽に関連する情報を使用してサードパーティのコンテンツアイテムを選択し得る。選択されるサードパーティのコンテンツアイテムは、ビジュアルコンテンツアイテム(たとえば、テキストを含み得る画像)であることが可能である。データ処理システムは、要求、クエリ、またはストリーミング音楽に関連するキーワードに基づいてこのビジュアルコンテンツアイテムを選択し得る。たとえば、選択されるビジュアルコンテンツアイテムは、コンテンツアイテムが少なくとも関連性スコア(relevancy score)に基づいてランク付けされ得るリアルタイムのコンテンツ選択プロセスに基づいて決定された最も高いランク付けのコンテンツアイテムであることが可能である。データ処理システムは、コンピューティングデバイスがビジュアルユーザインターフェース(たとえば、ディスプレイスクリーンおよびタッチスクリーン入力)ならびにオーディオユーザインターフェース(たとえば、出力のためのスピーカおよび入力のためのマイクロフォン)を用いて構成されると判定し得る。しかし、データ処理システムは、コンピューティングデバイスの現在の機能に基づいて、現在主に使用されているインターフェースがオーディオに基づくインターフェースであるとさらに判定し得る。したがって、データ処理システムは、コンピューティングデバイスがビジュアルインターフェースとオーディオインターフェースとの両方を用いて構成されるが、現在使用されている主なインターフェースがオーディオインターフェースであり、コンテンツアイテムによるレンダリングのために提供するためのオーディオコンテンツアイテムをビジュアルコン

テンツアイテムに基づいて生成することがコンピューティングデバイスによるバッテリ消費または無駄なコンピューティングリソースを削減し(たとえば、コンピューティングデバイスのディスプレイを休止状態から起動する代わりにストリーミング音楽と一緒にオーディオコンテンツアイテムを提供し)、コンピューティングデバイスによって提供されるユーザエクスペリエンスを向上させる(たとえば、気を散らさない方法でオーディオコンテンツアイテムを提供する)と判定し得る。したがって、この技術的な解決策は、ユーザインターフェースの機能およびユーザエクスペリエンスを向上させながらバッテリまたはコンピューティングリソースの利用を削減するためにコンテンツアイテムを異なる様式の間で継ぎ目なく遷移させることができる。

【００１２】

オーディオコンテンツアイテムを生成すると、データ処理システムは、オーディオ音楽ストリームへの挿入時間を決定し得る。データ処理システムは、オーディオコンテンツアイテムにビジュアルインジケータを付けるべきかどうかおよびコンテンツアイテムのためにどの種類のインタラクティブ性を構成すべきかをさらに動的に決定し得る。

【００１３】

図1は、オーディオトラックを生成する例示的なシステム100を示す。システム100は、ビジュアルコンテンツからオーディオトラックを生成することができる。システム100は、コンテンツ選択インフラストラクチャを含み得る。システム100は、データ処理システム102を含み得る。データ処理システム102は、1つもしくは複数のプロセッサ(たとえば、図3に示されるプロセッサ310)を含み得るかまたは1つもしくは複数のプロセッサ(たとえば、図3に示されるプロセッサ310)上で実行され得る。データ処理システム102は、ネットワーク105を介して3Pデジタルコンテンツプロバイダデバイス160またはコンピューティングデバイス140(たとえば、クライアントデバイス)のうちの1つまたは複数と通信することができる。ネットワーク105は、インターネット、ローカルエリアネットワーク、広域ネットワーク、メトロエリアネットワーク、またはその他のエリアネットワークなどのコンピュータネットワーク、イントラネット、衛星ネットワーク、および音声またはデータモバイル電話ネットワークなどのその他の通信ネットワークを含み得る。ネットワーク105は、ラップトップ、デスクトップ、タブレット、携帯情報端末、スマートフォン、ポータブルコンピュータ、またはスピーカなどの少なくとも1つのコンピューティングデバイス140上で提示されるか、出力されるか、レンダリングされるか、または表示され得るウェブページ、ウェブサイト、ドメイン名、またはユニフォームリソースロケータなどの情報リソースにアクセスするために使用され得る。たとえば、ネットワーク105を介して、コンピューティングデバイス140のユーザは、3Pデジタルコンテンツプロバイダデバイス160によって提供される情報またはデータにアクセスすることができる。コンピューティングデバイス140は、ディスプレイデバイス146およびスピーカ(たとえば、オーディオドライバ150によって駆動されるトランスデューサ)を含み得る。コンピューティングデバイス140は、ディスプレイを含む可能性がありまたは含まない可能性があり、たとえば、コンピューティングデバイスは、マイクロフォンおよびスピーカ(たとえば、スマートスピーカ)などの限られた種類のユーザインターフェースを含む可能性がある。場合によっては、コンピューティングデバイス140の主なユーザインターフェースは、マイクロフォンおよびスピーカである可能性がある。コンピューティングデバイス140は、音声に基づくコンピューティング環境とインターフェースを取り得るかまたは音声に基づくコンピューティング環境に含まれ得る。

【００１４】

ネットワーク105は、クライアントコンピューティングデバイス140によって提示されるか、出力されるか、レンダリングされるか、または表示され得るアプリケーション、ウェブページ、ウェブサイト、ドメイン名、またはユニフォームリソースロケータなどの情報リソースにアクセスするためにデータ処理システム102によって使用され得る。たとえば、ネットワーク105を介して、クライアントコンピューティングデバイス140のユーザは、3Pデジタルコンテンツプロバイダデバイス160によって提供される情報またはデータ

にアクセスすることができる。ネットワーク105は、コンテンツ掲載または検索エンジン結果システムに関連付けられるか、またはサードパーティのデジタルコンポーネントをデジタルコンポーネント掲載キャンペーン(digital component placement campaign)の一部として含むのにふさわしいインターネット上で利用可能な情報リソースのサブネットワークを含むかまたは成すことが可能である。

【００１５】

ネットワーク105は、任意の種類または形態のネットワークである可能性があり、以下、すなわち、ポイントツーポイントネットワーク、ブロードキャストネットワーク、広域ネットワーク、ローカルエリアネットワーク、電気通信ネットワーク、データ通信ネットワーク、コンピュータネットワーク、ATM(非同期転送モード)ネットワーク、SONET(同期光ネットワーク)ネットワーク、SDH(同期デジタルハイアラーキ: Synchronous Digital Hierarchy)ネットワーク、ワイヤレスネットワーク、および有線ネットワークのいずれかを含む可能性がある。ネットワーク105は、赤外線チャネルまたは衛星帯域などのワイヤレスリンクを含む可能性がある。ネットワーク105のトポロジーは、バス型、スター型、またはリング型ネットワークトポロジーを含む可能性がある。ネットワークは、アドバンストモバイル電話プロトコル(「AMPS: advanced mobile phone protocol」)、時分割多元接続(「TDMA」)、符号分割多元接続(「CDMA」)、移動体通信用グローバルシステム(「GSM: global system for mobile communication」)、汎用パケット無線サービス(「GPRS: general packet radio services」)、またはユニバーサル移動体通信システム(「UMTS: universal mobile telecommunications system」)を含むモバイルデバイスの間で通信するために使用される任意の1つのプロトコルまたは複数のプロトコルを使用するモバイル電話ネットワークを含む可能性がある。異なる種類のデータが、異なるプロトコルによって送信される可能性があり、または同じ種類のデータが、異なるプロトコルによって送信される可能性がある。

【００１６】

データ処理システム102は、ネットワーク105を介して通信するためのプロセッサを有するコンピューティングデバイスなどの少なくとも1つの論理デバイスを含み得る。データ処理システム102は、少なくとも1つの計算リソース、サーバ、プロセッサ、またはメモリを含み得る。たとえば、データ処理システム102は、少なくとも1つのデータセンターに置かれた複数の計算リソースまたはサーバを含み得る。データ処理システム102は、複数の論理的にグループ分けされたサーバを含み、分散型コンピューティング技術を促進することができる。サーバの論理的グループは、データセンター、サーバファーム、またはマシンファームと呼ばれる可能性がある。また、サーバは、地理的に散らされ得る。データセンターまたはマシンファームは、単一のエンティティ(entity)として管理される可能性があり、またはマシンファームは、複数のマシンファームを含むことが可能である。各マシンファーム内のサーバは、異種であることができる--サーバまたはマシンのうちの1つまたは複数が、1つまたは複数の種類のオペレーティングシステムプラットフォームによって動作することができる。

【００１７】

マシンファーム内のサーバは、関連するストレージシステムと一緒に高密度ラックシステムに収容され、エンタープライズデータセンターに置かれ得る。たとえば、このようにしてサーバをまとめることは、サーバおよび高性能ストレージシステムを局所的な高性能ネットワーク上に置くことによってシステムの管理の容易性、データセキュリティ、システムの物理的セキュリティ、およびシステムの性能を改善する可能性がある。サーバおよびストレージシステムを含み、それらを高度なシステム管理ツールに結合するデータ処理システム102のコンポーネントのすべてまたは一部の集中化は、電力および処理の要件を減らし、帯域幅の使用を削減する、サーバリソースのより効率的な使用を可能にする。

【００１８】

データ処理システム102は、少なくとも、ネットワーク105を介してまたはデータ処理システム102の様々なコンポーネントの間でデータパケットまたは情報を受信および送信

することができるインターフェース104を含み得る。データ処理システム102は、音声またはオーディオ入力を受け取り、入力オーディオ信号を処理または解析することができる少なくとも1つの自然言語プロセッサコンポーネント106を含み得る。データ処理システム102は、1つまたは複数の3Pデジタルコンテンツプロバイダデバイス160によって提供されたデジタルコンポーネントアイテム(たとえば、コンテンツアイテム)を選択するために設計され、構築され、動作可能な少なくとも1つのコンテンツセレクタコンポーネント108を含み得る。データ処理システム102は、少なくとも、第1の様式またはフォーマットのコンテンツアイテムを異なる様式またはフォーマットに変換すべきかどうかを決定するためのコンテンツ変換コンポーネント110を含み得る。コンテンツアイテムを変換することは、異なるフォーマットの新しいコンテンツアイテムを生成すること(たとえば、ビジュアルコンテンツアイテムからオーディオトラックを生成すること、またはビジュアルのみのコンテンツアイテムからオーディオビジュアルコンテンツアイテムを生成すること)を指し得るまたは含み得る。新しいコンテンツアイテムは、元のコンテンツアイテムの一部を含む可能性があり、または含まない可能性がある。コンテンツ変換コンポーネント110は、フォーマットセレクタ112、テキストジェネレータ114、音声セレクタ116、アクションジェネレータ136、またはオーディオの合図ジェネレータ118を含み得る。データ処理システム102は、コンテンツアイテムをいつまたはどこに挿入すべきかを決定することができる少なくとも1つのコンテンツ挿入コンポーネント120を含み得る。データ処理システム102は、少なくとも1つの機械学習エンジン122を含み得る。データ処理システム102は、少なくとも1つのデータリポジトリ124を含み得る。データリポジトリ124は、1つまたは複数のデータ構造、データファイル、データベース、またはその他のデータを含み得るまたは記憶し得る。データリポジトリ124は、1つまたは複数のローカルまたは分散型データベースを含むことができ、データベース管理システムを含むことができる。データリポジトリ124は、コンピュータデータストレージまたはメモリを含み得る。

【００１９】

データリポジトリ124は、音声モデル126、アクションモデル128、挿入モデル130、コンテンツデータ132、またはオーディオの合図134を含み得る、記憶し得る、または保有し得る。音声モデル126は、オーディオまたはオーディオビジュアルコンテンツを含む履歴的なコンテンツアイテムおよび履歴的なコンテンツアイテムに関連するメタデータに基づいて機械学習エンジン122を使用して訓練されたモデルを含み得る。また、音声モデル126は、履歴的なコンテンツアイテムに関連する実行情報を使用して訓練され得る。

【００２０】

アクションモデル128は、コンテンツアイテムに関してアクションまたはインタラクションの種類を決定することができる機械学習エンジン122を使用して訓練されたモデルを含み得る。たとえば、ユーザは、コンテンツアイテムについてのさらなる情報を要求すること、購入すること、ハイパーリンクを選択すること、コンテンツアイテムを一次停止すること、早送りすること、巻き戻すこと、もしくはスキップすることによってコンテンツアイテムとインタラクションするか、または何らかのその他のアクションを実行し得る。データ処理システム102は、アクションモデル128を使用してコンテンツアイテムとの起こりそうなインタラクションを決定または予測し、それから、予測されたインタラクションのためにコンテンツアイテムを構成することができる。アクションモデル128は、予め決められたアクションにマッピングされるコンテンツアイテムのカテゴリも含み得る。

【００２１】

挿入モデル130は、デジタル音楽ストリームのどこに挿入するかなど、生成されたコンテンツアイテムをどこに挿入すべきかを決定するために機械学習エンジン122を使用して訓練され得る。挿入モデル130は、異なる種類のコンテンツアイテムがデジタル音楽ストリームのどこに挿入されたかなどの履歴的なデータを使用して訓練され得る。

【００２２】

コンテンツデータ132は、3Pデジタルコンテンツプロバイダデバイス160によって提供されたコンテンツアイテムまたはデジタルコンポーネントオブジェクトについてのデータ

を含み得る。コンテンツデータ132は、たとえば、ビジュアルコンテンツアイテムもしくはビジュアルコンテンツアイテムのインジケーション、コンテンツキャンペーン(content campaign)のパラメータ、キーワード、またはコンテンツの選択もしくはコンテンツの配信を容易にするその他のデータを含み得る。

【００２３】

オーディオの合図134は、基礎となるオーディオトラックに追加されることが可能である口頭でないオーディオの合図を指し得る。オーディオの合図134は、オーディオファイルおよびオーディオファイルを説明するメタデータを含み得る。例示的なオーディオの合図は、海の波音、鳥のさえずり、スポーツイベントの観客の声援、風の吹く音、または車のエンジン音であることが可能である。

【００２４】

インターフェース104、自然言語プロセッサコンポーネント106、コンテンツセレクタコンポーネント108、コンテンツ変換コンポーネント110、フォーマットセレクタコンポーネント112、テキストジェネレータコンポーネント114、音声セレクタコンポーネント116、アクションジェネレータ136、オーディオの合図ジェネレータ118、コンテンツ挿入コンポーネント120、機械学習エンジン122、またはデータ処理システム102のその他のコンポーネントは、それぞれ、互いにまたはその他のリソースもしくはデータベースと通信するように構成された少なくとも1つの処理ユニットもしくはプログラマブル論理アレーエンジンなどのその他の論理デバイスまたはモジュールを含み得るかまたは利用し得る。インターフェース104、自然言語プロセッサコンポーネント106、コンテンツセレクタコンポーネント108、コンテンツ変換コンポーネント110、フォーマットセレクタコンポーネント112、テキストジェネレータコンポーネント114、音声セレクタコンポーネント116、オーディオの合図ジェネレータ118、コンテンツ挿入コンポーネント120、機械学習エンジン122、またはデータ処理システム102のその他のコンポーネントは、別個のコンポーネント、単一のコンポーネント、またはデータ処理システム102の一部であることが可能である。システム100およびデータ処理システム102などのそのコンポーネントは、1つまたは複数のプロセッサ、論理デバイス、または回路などのハードウェア要素を含み得る。データ処理システム102のコンポーネント、システム、またはモジュールは、少なくとも部分的にデータ処理システム102によって実行され得る。

【００２５】

コンピューティングデバイス140は、少なくとも1つのセンサー148、トランスデューサ144、オーディオドライバ150、プリプロセッサ142、またはディスプレイデバイス146とのインターフェースを含み得るかまたはこれらと通信し得る。センサー148は、たとえば、環境光センサー、近接センサー、温度センサー、加速度計、ジャイロスコープ、モーションディテクタ、GPSセンサー、位置センサー、マイクロフォン、またはタッチセンサーを含み得る。トランスデューサ144は、スピーカまたはマイクロフォンを含み得る。オーディオドライバ150は、ハードウェアトランスデューサ144にソフトウェアインターフェースを提供することができる。オーディオドライバは、対応する音響波または音波を生成するようにトランスデューサ144を制御するためにデータ処理システム102によって提供されるオーディオファイルまたはその他の命令を実行することができる。ディスプレイデバイス146は、図3に示されるディスプレイ335の1つまたは複数のコンポーネントまたは関数を含み得る。プリプロセッサ142は、トリガキーワード、予め決められたホットワード、開始キーワード、またはアクティブ化キーワードを検出するように構成され得る。場合によっては、トリガキーワードは、(アクションモデル128を使用してアクションジェネレータ136によって選択されたアクションなどの)アクションを実行する要求を含み得る。場合によっては、トリガキーワードは、コンピューティングデバイス140を有効化またはアクティブ化するための予め決められたアクションキーワードを含むことが可能であり、要求キーワードは、トリガキーワードまたはホットワードの後に続くことが可能である。プリプロセッサ142は、キーワードを検出し、キーワードに基づいてアクションを実行するように構成され得る。プリプロセッサ142は、ウェークアップワードまたはその他

のキーワードもしくはホットワードを検出し、検出に応じて、コンピューティングデバイス140によって実行されるデータ処理システム102の自然言語プロセッサコンポーネント106を呼び出すことができる。場合によっては、プリプロセッサ142は、さらなる処理のためにデータ処理システム102に語を送信する前に1つまたは複数の語をフィルタリングして取り除くかまたは語を修正することができる。プリプロセッサ142は、マイクロフォンによって検出されたアナログオーディオ信号をデジタルオーディオ信号に変換し、デジタルオーディオ信号を運ぶ1つまたは複数のデータパケットをデータ処理システム102にまたはネットワーク105を介してデータ処理システム102に送信または提供することができる。場合によっては、プリプロセッサ142は、そのような送信を実行するための命令を検出することに応じて入力オーディオ信号の一部またはすべてを運ぶデータパケットを自然言語プロセッサコンポーネント106またはデータ処理システム102に提供することができる。命令は、たとえば、入力オーディオ信号を含むデータパケットをデータ処理システム102に送信するためのトリガキーワードまたはその他のキーワードまたは承認を含み得る。

【００２６】

クライアントコンピューティングデバイス140は、(センサー148を介して)音声クエリをクライアントコンピューティングデバイス140にオーディオ入力として入力し、トランスデューサ144(たとえば、スピーカ)から出力された、データ処理システム102(または3Pデジタルコンテンツプロバイダデバイス160)からクライアントコンピューティングデバイス140に提供され得るコンピュータによって生成された音声の形態のオーディオ出力を受け取るエンドユーザに関連付けられ得る。コンピュータによって生成された音声は、本物の人からの録音またはコンピュータによって生成された言葉を含み得る。

【００２７】

コンピューティングデバイス140は、アプリケーション152を実行することができる。データ処理システム102は、コンピューティングデバイス140がアプリケーション152を実行することができるオペレーティングシステムを含み得るかまたは実行し得る。アプリケーション152は、クライアントコンピューティングデバイス140が実行するか、走らせるか、起動するか、またはそれ以外の方法で提供するように構成される任意の種類のアプリケーションを含み得る。アプリケーション152は、マルチメディアアプリケーション、音楽プレーヤー、ビデオプレーヤー、ウェブブラウザ、ワードプロセッサ、モバイルアプリケーション、デスクトップアプリケーション、タブレットアプリケーション、電子ゲーム、電子商取引アプリケーション、またはその他の種類のアプリケーションを含み得る。アプリケーション152は、電子リソースに対応するデータを実行、レンダリング、ロード、解析、処理、提示、またはそうでなければ出力することができる。電子リソースは、たとえば、ウェブサイト、ウェブページ、マルチメディアWebコンテンツ、ビデオコンテンツ、オーディオコンテンツ、デジタルストリーミングコンテンツ、旅行コンテンツ、エンターテインメントコンテンツ、商品もしくはサービスの買い物に関連するコンテンツ、またはその他のコンテンツを含み得る。

【００２８】

コンピューティングデバイス140上で実行されるアプリケーション152は、サードパーティの(「3P」)電子リソースサーバ162から電子リソースに関連するデータを受信することができる。3P電子リソースサーバ162は、アプリケーションによる実行のために電子リソースを提供することができる。3P電子リソースサーバ162は、ファイルサーバ、ウェブサーバ、ゲームサーバ、マルチメディアサーバ、クラウドコンピューティング環境、またはアプリケーションにコンピューティングデバイス140を介して電子リソースを提示もしくは提供させるためのデータを提供するように構成されたその他のバックエンドコンピューティングシステムを含み得る。コンピューティングデバイス140は、ネットワーク105を介して3P電子リソースサーバ162にアクセスすることができる。

【００２９】

3P電子リソースサーバ162の管理者は、電子リソースを作る、確立する、維持する、ま

たは提供することができる。3P電子リソースサーバ162は、電子リソースの要求に応じてコンピューティングデバイス140に電子リソースを送信することができる。電子リソースは、ユニフォームリソースロケータ(「URL」)、統一資源識別子、ウェブアドレス、またはファイル名、またはファイルパスなどの識別子に関連付けられ得る。3P電子リソースサーバ162は、アプリケーション152から電子リソースの要求を受信することができる。電子リソースは、電子ドキュメント、ウェブページ、マルチメディアコンテンツ、ストリーミングコンテンツ(たとえば、音楽、ニュース、もしくはポッドキャスト)、オーディオ、ビデオ、テキスト、画像、ビデオゲーム、またはその他のデジタルもしくは電子コンテンツを含み得る。

【００３０】

データ処理システム102は、少なくとも1つの3Pデジタルコンテンツプロバイダデバイス160にアクセスし得るかまたはそうでなければ少なくとも1つの3Pデジタルコンテンツプロバイダデバイス160とインタラクションし得る。3Pデジタルコンテンツプロバイダデバイス160は、たとえば、ネットワーク105を介してコンピューティングデバイス140またはデータ処理システム102と通信するためのプロセッサを有するコンピューティングデバイスなどの少なくとも1つの論理デバイスを含み得る。3Pデジタルコンテンツプロバイダデバイス160は、少なくとも1つの計算リソース、サーバ、プロセッサ、またはメモリを含み得る。たとえば、3Pデジタルコンテンツプロバイダデバイス160は、少なくとも1つのデータセンターに置かれた複数の計算リソースまたはサーバを含み得る。3Pデジタルコンテンツプロバイダデバイス160は、広告主デバイス、サービスプロバイダデバイス、または商品プロバイダデバイスを含み得るまたは指し得る。

【００３１】

3Pデジタルコンテンツプロバイダデバイス160は、コンピューティングデバイス140による提示のためにデジタルコンポーネントを提供することができる。デジタルコンポーネントは、コンピューティングデバイス140のディスプレイデバイス146を介して提示するためのビジュアルデジタルコンポーネントであることが可能である。デジタルコンポーネントは、検索クエリまたは要求に対する応答を含み得る。デジタルコンポーネントは、データベース、検索エンジン、またはネットワーク化されたリソースからの情報を含み得る。たとえば、デジタルコンポーネントは、ニュースの情報、天気の情報、スポーツの情報、百科事典の項目、辞書の項目、またはデジタル教科書からの情報を含み得る。デジタルコンポーネントは、広告を含み得る。デジタルコンポーネントは、「スニーカーを購入したいですか。」と述べるメッセージなどの商品またはサービスの申し出を含み得る。3Pデジタルコンテンツプロバイダデバイス160は、クエリに応じて提供され得る一連のデジタルコンポーネントを記憶するためのメモリを含み得る。3Pデジタルコンテンツプロバイダデバイス160は、ビジュアルまたはオーディオに基づくデジタルコンポーネント(またはその他のデジタルコンポーネント)をデータ処理システム102に提供することもでき、データ処理システム102において、それらのビジュアルまたはオーディオに基づくデジタルコンポーネントは、コンテンツセレクタコンポーネント108による選択のために記憶され得る。データ処理システム102は、デジタルコンポーネントを選択し、デジタルコンポーネントをクライアントコンピューティングデバイス140に提供する(または提供するようにコンテンツプロバイダコンピューティングデバイス160に命令する)ことができる。デジタルコンポーネントは、ビジュアルのみのデータであるか、オーディオのみのデータであるか、またはテキスト、画像、もしくはビデオデータを用いたオーディオデータとビジュアルデータとの組合せであることが可能である。デジタルコンポーネントまたはコンテンツアイテムは、1つまたは複数のフォーマットの画像、テキスト、ビデオ、マルチメディア、またはその他の種類のコンテンツを含み得る。

【００３２】

データ処理システム102は、少なくとも1つの計算リソースまたはサーバを有するコンテンツ掲載システムを含み得る。データ処理システム102は、少なくとも1つのコンテンツセレクタコンポーネント108を含み得るか、少なくとも1つのコンテンツセレクタコン

ポーネント108とインターフェースを取り得るか、またはそうでなければ少なくとも1つのコンテンツセレクタコンポーネント108と通信し得る。データ処理システム102は、少なくとも1つのデジタルアシスタントサーバを含み得るか、少なくとも1つのデジタルアシスタントサーバとインターフェースを取り得るか、またはそうでなければ少なくとも1つのデジタルアシスタントサーバと通信し得る。

【００３３】

データ処理システム102は、複数のコンピューティングデバイス140に関連する匿名のコンピュータネットワーク活動情報を取得することができる。コンピューティングデバイス140のユーザは、ユーザのコンピューティングデバイス140に対応するネットワーク活動情報を取得することをデータ処理システム102に肯定的に認可することが可能である。たとえば、データ処理システム102は、1つまたは複数の種類のネットワーク活動情報を取得することに同意するようにコンピューティングデバイス140のユーザに促すことができる。コンピューティングデバイス140のユーザのアイデンティティ(identity)は、匿名のままであることができ、コンピューティングデバイス140は、一意識別子(たとえば、データ処理システムまたはコンピューティングデバイスのユーザによって提供されるユーザまたはコンピューティングデバイスの一意識別子)に関連付けられ得る。データ処理システム102は、各観測値(observation)を対応する一意識別子と関連付けることができる。

【００３４】

3Pデジタルコンテンツプロバイダデバイス160は、電子コンテンツキャンペーンを確立することができる。電子コンテンツキャンペーンは、コンテンツセレクタコンポーネント108のデータリポジトリにコンテンツデータとして記憶され得る。電子コンテンツキャンペーンは、共通のテーマに対応する1つまたは複数のコンテンツグループを指し得る。コンテンツキャンペーンは、コンテンツグループ、デジタルコンポーネントデータオブジェクト、およびコンテンツ選択基準を含む階層的なデータ構造を含み得る。コンテンツキャンペーンを作成するために、3Pデジタルコンテンツプロバイダデバイス160は、コンテンツキャンペーンのキャンペーンレベルパラメータに関する値を指定することができる。キャンペーンレベルパラメータは、たとえば、キャンペーン名、デジタルコンポーネントオブジェクトを掲載するための好ましいコンテンツネットワーク、コンテンツキャンペーンのために使用されるリソースの値、コンテンツキャンペーンの開始日および終了日、コンテンツキャンペーンの継続時間、デジタルコンポーネントオブジェクトの掲載のためのスケジュール、言語、地理的場所、デジタルコンポーネントオブジェクトを提供すべきコンピューティングデバイスの種類を含み得る。場合によっては、インプレッションが、デジタルコンポーネントオブジェクトがそのソース(たとえば、データ処理システム102または3Pデジタルコンテンツプロバイダデバイス160)からいつフェッチされるかを指すことが可能であり、数えられ得る。場合によっては、クリック詐欺の可能性があるため、インプレッションとして、ロボットの活動がフィルタリングされ、除外され得る。したがって、場合によっては、インプレッションは、ロボットの活動およびエラーコードからフィルタリングされ、コンピューティングデバイス140上に表示するためにデジタルコンポーネントオブジェクトをレンダリングする機会にできるだけ近い時点で記録される、ブラウザからのページ要求に対するウェブサーバからの応答の測定値を指し得る。場合によっては、インプレッションは、可視インプレッション(viewable impression)または可聴インプレッション(audible impression)を指すことが可能であり、たとえば、デジタルコンポーネントオブジェクトは、クライアントコンピューティングデバイス140のディスプレイデバイス上で少なくとも部分的に(たとえば、20%、30%、30%、40%、50%、60%、70%、もしくはそれ以上)可視であるか、またはコンピューティングデバイス140のスピーカを介して少なくとも部分的に(たとえば、20%、30%、30%、40%、50%、60%、70%、もしくはそれ以上)可聴である。クリックまたは選択は、可聴インプレッションに対する音声応答、マウスクリック、タッチインタラクション、ジェスチャ、振り動かし、オーディオインタラクション、またはキーボードクリックなどのデジタルコンポーネントオブジェクトとのユーザインタラクションを指し得る。コンバージョンは、ユーザがデジタルコンポ

ーネントオブジェクトに関連して所望のアクションを行うこと、たとえば、製品もしくはサービスを購入すること、調査を完了すること、デジタルコンポーネントに対応する物理的な店舗を訪れること、または電子取引を完了することを指し得る。

【００３５】

3Pデジタルコンテンツプロバイダデバイス160は、コンテンツキャンペーンに関する1つまたは複数のコンテンツグループをさらに確立することができる。コンテンツグループは、1つまたは複数のデジタルコンポーネントオブジェクトと、キーワード、単語、語、語句、地理的場所、コンピューティングデバイスの種類、時刻、関心、話題、またはバーティカル(vertical)などの対応するコンテンツ選択基準とを含む。同じコンテンツキャンペーンの下のコンテンツグループは、同じキャンペーンレベルパラメータを共有することが可能であるが、キーワード、(たとえば、主コンテンツに除外キーワードが存在する場合にデジタルコンポーネントの掲載を阻止する)除外キーワード、キーワードの入札単価(bid)、または入札単価もしくはコンテンツキャンペーンに関連するパラメータなどの特定のコンテンツグループレベルパラメータに関するカスタマイズされた仕様を有する可能性がある。

【００３６】

新しいコンテンツグループを作成するために、3Pデジタルコンテンツプロバイダデバイス160は、コンテンツグループのコンテンツグループレベルパラメータに関する値を与えることができる。コンテンツグループレベルパラメータは、たとえば、コンテンツグループ名もしくはコンテンツグループのテーマ、および異なるコンテンツ掲載機会(たとえば、自動掲載もしくは管理された掲載)または結果(たとえば、クリック、インプレッション、もしくはコンバージョン)の入札単価を含む。コンテンツグループ名またはコンテンツグループのテーマは、コンテンツグループのデジタルコンポーネントオブジェクトが表示のために選択されるべきである話題または主題を捕捉するために3Pデジタルコンテンツプロバイダデバイス160が使用することができる1つまたは複数の語であることが可能である。たとえば、自動車の特約販売店は、その特約販売店が扱う車両の各ブランドのために異なるコンテンツグループを作成することができ、その特約販売店が扱う各モデルのために異なるコンテンツグループをさらに作成する可能性がある。自動車の特約販売店が使用することができるコンテンツグループのテーマの例は、たとえば、「Aスポーツカーを製造する」、「Bスポーツカーを製造する」、「Cセダンを製造する」、「Cトラックを製造する」、「Cハイブリッドを製造する」、または「Dハイブリッドを製造する」を含む。例示的なコンテンツキャンペーンのテーマは、「ハイブリッド」であり、たとえば、「Cハイブリッドを製造する」と「Dハイブリッドを製造する」との両方のためのコンテンツグループを含み得る。

【００３７】

3Pデジタルコンテンツプロバイダデバイス160は、各コンテンツグループに1つまたは複数のキーワードおよびデジタルコンポーネントオブジェクトを与えることができる。キーワードは、デジタルコンポーネントオブジェクトに関連するかまたはデジタルコンポーネントオブジェクトによって特定される製品またはサービスに関連する語を含み得る。キーワードは、1つまたは複数の語または語句を含み得る。たとえば、コンテンツグループまたはコンテンツキャンペーンに関するキーワードとして自動車の特約販売店は、「スポーツカー」、「V-6エンジン」、「四輪駆動」、「燃費」を含み得る。場合によっては、除外キーワードが、特定の語またはキーワードに対するコンテンツ掲載を避けるか、防止するか、阻止するか、または無効にするためにコンテンツプロバイダによって指定され得る。コンテンツプロバイダは、デジタルコンポーネントオブジェクトを選択するために使用される、完全一致(exact match)、フレーズ一致、または部分一致(broad match)などのマッチングの種類を指定することができる。

【００３８】

3Pデジタルコンテンツプロバイダデバイス160は、3Pデジタルコンテンツプロバイダデバイス160によって提供されるデジタルコンポーネントオブジェクトを選択するために

データ処理システム102によって使用される1つまたは複数のキーワードを提供することができる。3Pデジタルコンテンツプロバイダデバイス160は、入札する1つまたは複数のキーワードを特定し、様々なキーワードに関する入札額を与えることができる。3Pデジタルコンテンツプロバイダデバイス160は、デジタルコンポーネントオブジェクトを選択するためにデータ処理システム102によって使用される追加的なコンテンツ選択基準を提供することができる。複数の3Pデジタルコンテンツプロバイダデバイス160が、同じまたは異なるキーワードに入札することができ、データ処理システム102は、電子的メッセージのキーワードのインジケーションを受け取ることに応じてコンテンツ選択プロセスまたは広告オークションを実行することができる。

【００３９】

3Pデジタルコンテンツプロバイダデバイス160は、データ処理システム102による選択のために1つまたは複数のデジタルコンポーネントオブジェクトを提供することができる。(たとえば、コンテンツセレクタコンポーネント108を介して)データ処理システム102は、リソース割り当て、コンテンツのスケジュール、最大入札単価、キーワード、およびコンテンツグループに関して指定されたその他の選択基準に一致するコンテンツ掲載機会が利用可能になるときにデジタルコンポーネントオブジェクトを選択することができる。音声デジタルコンポーネント、オーディオデジタルコンポーネント、テキストデジタルコンポーネント、画像デジタルコンポーネント、動画デジタルコンポーネント、マルチメディアデジタルコンポーネント、またはデジタルコンポーネントリンクなどの異なる種類のデジタルコンポーネントオブジェクトが、コンテンツグループに含まれ得る。デジタルコンポーネントを選択すると、データ処理システム102は、コンピューティングデバイス140を介して提示するためにデジタルコンポーネントオブジェクトを送信することができ、コンピューティングデバイス140またはコンピューティングデバイス140のディスプレイデバイス上でレンダリングする。レンダリングは、ディスプレイデバイス上にデジタルコンポーネントを表示すること、またはコンピューティングデバイス140のスピーカによってデジタルコンポーネントを再生することを含み得る。データ処理システム102は、デジタルコンポーネントオブジェクトをレンダリングするためにコンピューティングデバイス140に命令を与えることができる。データ処理システム102は、オーディオ信号または音響波を生成するようにコンピューティングデバイス140の自然言語プロセッサコンポーネント106またはコンピューティングデバイス140のオーディオドライバ150に命令することができる。データ処理システム102は、選択されたデジタルコンポーネントオブジェクトを提示するようにコンピューティングデバイス140によって実行されるアプリケーションに命令することができる。たとえば、アプリケーション(たとえば、デジタル音楽ストリーミングアプリケーション)は、デジタルコンポーネントオブジェクトが提示され得るスロット(たとえば、コンテンツスロット)(たとえば、オーディオスロットまたはビジュアルスロット)を含むことができる。

【００４０】

データ処理システム102は、少なくとも1つのインターフェース104を含み得る。データ処理システム102は、たとえば、データパケットを使用して情報を受信および送信するように設計されたか、構成されたか、構築されたか、または動作可能であるインターフェース104を含み得る。インターフェース104は、ネットワークプロトコルなどの1つまたは複数のプロトコルを使用して情報を受信および送信することができる。インターフェース104は、ハードウェアインターフェース、ソフトウェアインターフェース、有線インターフェース、またはワイヤレスインターフェースを含み得る。インターフェース104は、あるフォーマットから別のフォーマットにデータを変換するかまたはフォーマットすることを容易にすることができる。たとえば、インターフェース104は、ソフトウェアコンポーネントなどの様々なコンポーネントの間で通信するための定義を含むアプリケーションプログラミングインターフェースを含み得る。インターフェース104は、自然言語プロセッサコンポーネント106と、コンテンツセレクタコンポーネント108と、コンテンツ変換コンポーネント110と、データリポジトリ124との間など、システム100の1つまたは複

数のコンポーネントの間の通信を容易にすることができる。

【００４１】

インターフェース104は、データ処理システム102の遠隔のコンピューティングデバイス140のマイクロフォン(たとえば、センサー148)によって検出された入力オーディオ信号を含むデータパケットをネットワーク105を介して受信することができる。コンピューティングデバイス140のユーザは、コンピューティングデバイス140にスピーチまたは音声入力を与え、入力オーディオ信号またはプリプロセッサ142によってオーディオ信号に基づいて生成されたデータパケットをデータ処理システム102に送信するようにコンピューティングデバイス140に命令するかまたはそれ以外の方法でコンピューティングデバイス140に送信させることができる。

【００４２】

データ処理システム102は、データパケットまたは入力オーディオ信号を解析するように設計され、構築された、動作可能な自然言語プロセッサコンポーネント106を含み得るか、自然言語プロセッサコンポーネント106とインターフェースを取り得るか、またはそうでなければ自然言語プロセッサコンポーネント106と通信し得る。自然言語プロセッサコンポーネント106は、データ処理システム102のハードウェア、電子回路、アプリケーション、スクリプト、またはプログラムを含み得る。自然言語プロセッサコンポーネント106は、入力信号、データパケット、またはその他の情報を受信することができる。自然言語プロセッサコンポーネント106は、スピーチを含む入力オーディオ信号を処理してスピーチをテキストに書き起こし、それから、書き起こされたテキストを理解するための自然言語処理を実行するように構成されたスピーチ認識器を含み得るかまたはスピーチ認識器と呼ばれ得る。自然言語プロセッサコンポーネント106は、インターフェース104を介してデータパケットまたはその他の入力を受信することができる。自然言語プロセッサコンポーネント106は、データ処理システム102のインターフェース104から入力オーディオ信号を受信し、出力オーディオ信号をレンダリングするようにクライアントコンピューティングデバイスのコンポーネントを駆動するためのアプリケーションを含み得る。データ処理システム102は、オーディオ入力信号を含むかまたは特定するデータパケットまたはその他の信号を受信することができる。たとえば、自然言語プロセッサコンポーネント106は、オーディオ信号を受信するかまたは取得し、オーディオ信号を解析することができるNLP技術、機能、またはコンポーネントを用いて構成され得る。自然言語プロセッサコンポーネント106は、人とコンピュータとの間のインタラクションを提供することができる。自然言語プロセッサコンポーネント106は、自然言語を理解し、データ処理システム102が人間のまたは自然言語入力から意味を導出することを可能にするための技術を用いて構成され得る。自然言語プロセッサコンポーネント106は、統計的機械学習などの機械学習に基づく技術を含み得るかまたはそのような技術を用いて構成され得る。自然言語プロセッサコンポーネント106は、入力オーディオ信号を解析するために決定木、統計モデル、または確率モデルを利用することができる。自然言語プロセッサコンポーネント106は、たとえば、固有表現(named entity)認識(たとえば、テキストのストリームが与えられたものとして、テキスト内のどのアイテムが人または場所などの適切な名前にマッピングされるか、およびそれぞれのそのような名前の種類が人、場所、または組織などのどれであるのかを決定すること)、自然言語生成(たとえば、コンピュータデータベースからの情報または意味的意図(semantic intent)を理解可能な人間の言語に変換すること)、自然言語理解(たとえば、テキストをコンピュータモジュールが操作することができる一階論理構造などのより整然とした表現に変換すること)、機械翻訳(たとえば、テキストをある人間の言語から別の人間の言語に自動的に翻訳すること)、形態素分割(たとえば、考慮されている言語の言葉の形態または構造の複雑さに基づいて困難であり得る、単語を個々の形態素に分け、形態素のクラスを特定すること)、質問応答(たとえば、特定的であるかまたは自由であることが可能である人間の言語の質問に対する答えを決定すること)、意味処理(たとえば、特定された単語を同様の意味を有するその他の単語に関連付けるために、単語を特定し、その単語の意味を符号化した後に行われる得る処理)などの機能を実行するこ

とができる。
【００４３】
自然言語プロセッサコンポーネント106は、(たとえば、NLP技術、機能、またはコンポーネントを利用して)訓練データが含むことに基づく機械学習モデルの訓練を使用してオーディオ入力信号を認識されたテキストに変換することができる。オーディオ波形の組が、データリポジトリ124、またはデータ処理システム102がアクセス可能なその他のデータベースに記憶され得る。代表的な波形が、ユーザの大きな組全体で生成されることが可能であり、それから、ユーザからのスピーチのサンプルによって増強される可能性がある。オーディオ信号が認識されたテキストに変換された後、自然言語プロセッサコンポーネント106は、たとえば、ユーザ全体にわたって訓練されたデータリポジトリ124に記憶されたモデルを使用することによってまたは手動で指定することによって、データ処理システム102が提供することができるアクションと関連付けられる単語にテキストをマッチングすることができる。
【００４４】
オーディオ入力信号は、クライアントコンピューティングデバイス140のセンサー148またはトランスデューサ144(たとえば、マイクロフォン)によって検出され得る。トランスデューサ144、オーディオドライバ150、またはその他のコンポーネントを介して、クライアントコンピューティングデバイス140は、データ処理システム102にオーディオ入力信号を提供することができ、データ処理システム102において、オーディオ入力信号は、(たとえば、インターフェース104によって)受信され、NLPコンポーネント106に提供され得るかまたはデータリポジトリ124に記憶され得る。
【００４５】
自然言語プロセッサコンポーネント106は、入力オーディオ信号を取得することができる。入力オーディオ信号から、自然言語プロセッサコンポーネント106は、少なくとも1つの要求または少なくとも1つのトリガキーワード、キーワード、もしくは要求を特定することができる。要求は、入力オーディオ信号の意図または主題を示し得る。キーワードは、行われる可能性が高いアクションの種類を示し得る。たとえば、自然言語プロセッサコンポーネント106は、入力オーディオ信号を解析して、アプリケーションを呼び出すか、コンテンツアイテムとインタラクションする少なくとも1つの要求、またはコンテンツの要求を特定することができる。自然言語プロセッサコンポーネント106は、入力オーディオ信号を解析して、夜に食事会に参加し、映画を見るために家を出る要求などの少なくとも1つの要求を特定することができる。キーワードは、行われるアクションを示す少なくとも1つの単語、語句、語根もしくは部分的な単語、または派生語を含み得る。たとえば、入力オーディオ信号からのキーワード「行く」または「～に行くために」は、輸送の必要性を示し得る。この例において、入力オーディオ信号(または特定された要求)は、輸送の意図を直接表さないが、キーワードが、輸送が要求によって示される少なくとも1つのその他のアクションの補助的なアクションであることを示す。
【００４６】
自然言語プロセッサコンポーネント106は、入力オーディオ信号を解析して、要求およびキーワードを特定するか、判定するか、取り出すか、またはそれ以外の方法で取得することができる。たとえば、自然言語プロセッサコンポーネント106は、キーワードまたは要求を特定するために入力オーディオ信号に意味処理技術を適用することができる。自然言語プロセッサコンポーネント106は、1つまたは複数のキーワードを特定するために入力オーディオ信号に意味処理技術を適用することができる。キーワードは、1つまたは複数の語または語句を含み得る。自然言語プロセッサコンポーネント106は、デジタルアクション(digital action)を実行する意図を特定するために意味処理技術を適用することができる。
【００４７】
たとえば、コンピューティングデバイス140は、クライアントコンピューティングデバイス140のセンサー148(たとえば、マイクロフォン)によって検出された入力オーディオ

信号を受信することができる。入力オーディオ信号は、「デジタルアシスタント、私は、誰かに自分の洗濯と自分のドライクリーニングをしてもらわないとならない。」であることが可能である。クライアントコンピューティングデバイス140のプリプロセッサ142は、「デジタルアシスタント」などの入力オーディオ信号内のウェークアップワード、ホットワード、またはトリガキーワードを検出することができる。プリプロセッサ142は、入力オーディオ信号内のオーディオシグネチャ(audio signature)または波形をトリガキーワードに対応するモデルオーディオシグネチャまたは波形と比較することによってウェークアップワード、ホットワード、またはトリガキーワードを検出することができる。プリプロセッサ142は、入力オーディオ信号が自然言語プロセッサコンポーネント106によって処理されるべきであることを示すウェークアップワード、ホットワード、またはトリガキーワードを入力オーディオ信号が含むと判定し得る。ホットワード、ウェークアップワード、またはトリガキーワードを検出することに応じて、プリプロセッサ142は、検出された入力オーディオ信号を決定するか、認可するか、ルーティングするか、転送するか、または自然言語プロセッサコンポーネント106による処理のためにデータ処理システム102にその他の方法で提供することができる。

【００４８】

自然言語プロセッサコンポーネント106は、入力オーディオ信号を受信し、トリガ語句「自分の洗濯をする」および「自分のドライクリーニングをする」を特定するために文を含む入力オーディオ信号に意味処理技術またはその他の自然言語処理技術を適用することができる。場合によっては、自然言語プロセッサコンポーネント106は、自然言語プロセッサコンポーネント106に入力オーディオ信号を処理させるために入力オーディオ信号に対応するデータパケットをデータ処理システム102に提供することができる。自然言語プロセッサコンポーネント106は、デジタルアシスタントサーバと連携してまたはデジタルアシスタントサーバを介して入力オーディオ信号を処理することができる。自然言語プロセッサコンポーネント106は、洗濯およびドライクリーニングなどの複数のキーワードをさらに特定し得る。

【００４９】

自然言語プロセッサコンポーネント106は、検索または情報のその他の要求を実行することに対応する検索クエリ、キーワード、意図、または語句を特定し得る。自然言語プロセッサコンポーネント106は、入力オーディオ信号が話題、イベント、現在のイベント、ニュースイベント、辞書の定義、履歴的なイベント、人物、場所、または物についての情報の要求に対応すると判定し得る。たとえば、自然言語プロセッサコンポーネント106は、入力オーディオ信号が旅行の手配をする、乗車予約をする、情報を取得する、ウェブ検索を実行する、株価をチェックする、アプリケーションを起動する、ニュースをチェックする、食べ物を注文する、またはその他の製品、商品、もしくはサービスを買うためのクエリ、要求、意図、またはアクションに対応すると判定し得る。

【００５０】

自然言語プロセッサコンポーネント106は、入力オーディオ信号を解析または処理するために1つまたは複数の技術を使用し得る。技術は、規則に基づく技術または統計的技術を含み得る。技術は、機械学習または深層学習を利用し得る。例示的な技術は、固有表現認識、感情分析、テキストの要約、アスペクトマイニング(aspect mining)、またはトピックマイニング(topic mining)を含み得る。技術は、テキスト埋め込み(たとえば、文字列の実数値ベクトル表現)、機械翻訳(たとえば、言語分析および言語生成)、または対話および会話(たとえば、人工知能によって使用されるモデル)を含み得るかまたはそれらに基づき得る。技術は、レンマ化(lemmatization)、形態素分割、単語分割(word segmentation)、品詞のタグ付け、解析、文の分割(sentence breaking)、またはステミング(stemming)などのシンタックス技術(たとえば、文法に基づく文中の単語の配列)を含み得るか、決定し得るか、または利用し得る。技術は、固有表現認識(たとえば、特定され、アプリケーション152、人、もしくは場所の名前などの現在のグループにカテゴリ分けされ得るテキストの部分を決定すること)、語意の曖昧性の解消、または自然言語生成などのセマン

ティクス技術を含み得るか、決定し得るか、または利用し得る。

【００５１】

場合によっては、自然言語プロセッサコンポーネント106は、アプリケーション152を起動する要求を特定し、アプリケーション152を起動するためにコンピューティングデバイス140に命令を与えることができる。場合によっては、アプリケーション152は、自然言語プロセッサコンポーネント106が入力オーディオ信号を受信する前に既に起動されている可能性がある。たとえば、入力オーディオ信号を処理または解析することに基づいて、自然言語プロセッサコンポーネント106は、呼び出す、起動する、開く、またはそれ以外の方法でアクティブ化するアプリケーション152を特定することができる。自然言語プロセッサコンポーネント106は、入力オーディオ信号を解析して語、キーワード、トリガキーワード、または語句を特定することに基づいてアプリケーション152を特定することができる。自然言語プロセッサコンポーネント106は、特定された語、キーワード、トリガキーワード、または語句を使用してデータリポジトリ124の検索を実行してアプリケーション152を特定することができる。場合によっては、キーワードは、「Application_Name_A」または「Application_Name_B」などのアプリケーション152の識別子を含み得る。場合によっては、キーワードは、相乗りアプリケーション、レストラン予約アプリケーション、映画チケットアプリケーション、ニュースアプリケーション、天気アプリケーション、ナビゲーションアプリケーション、ストリーミング音楽アプリケーション、ストリーミングビデオアプリケーション、レストラン批評アプリケーション、またはその他の種類もしくはカテゴリのアプリケーション152などのある種類またはカテゴリのアプリケーション152を示し得る。入力オーディオ信号の受信の前にアプリケーション152が既に起動され、実行されている可能性がある場合に関して、自然言語プロセッサコンポーネント106は、入力オーディオ信号を処理して、アプリケーション152において、またはアプリケーション152によってレンダリングされる電子リソースを介して提示される行動喚起に応じて実行されるアクションを決定することができる。

【００５２】

データ処理システム102は、コンピューティングデバイス140上で提示するためのコンテンツの要求をコンピュータネットワークを介して受信することができる。データ処理システム102は、クライアントコンピューティングデバイス140のマイクロフォンによって検出された入力オーディオ信号を処理することによって要求を特定することができる。要求は、要求に関連するデバイスの種類、場所、およびキーワードなどの要求の選択基準を含み得る。選択基準は、コンピューティングデバイス140のコンテキストについての情報を含み得る。コンピューティングデバイス140のコンテキストは、コンピューティングデバイス140上で実行されているアプリケーションについての情報、コンピューティングデバイス140の場所についての情報、コンピューティングデバイス140を介して(たとえば、アプリケーション152を介して)レンダリング、提示、提供、またはアクセスされているコンテンツについての情報を含み得る。たとえば、コンテンツ選択基準は、デジタルストリーミング音楽アプリケーション152によって再生されている音楽に関連するアーティスト、歌のタイトル、またはジャンルなどの情報またはキーワードを含み得る。場合によっては、コンテンツ選択基準は、アプリケーション152の閲覧履歴に関連するキーワードを含み得る。

【００５３】

データ処理システム102は、3Pデジタルコンテンツプロバイダデバイス160によって提供されるデジタルコンポーネントを選択すると決定し得る。データ処理システム102は、コンピューティングデバイス140からの要求に応じてデジタルコンポーネントを選択すると決定し得る。データ処理システム102は、アプリケーション152内のコンテンツスロットを特定することに応じてデジタルコンポーネントを選択すると決定し得る。データ処理システム102は、イベント、条件、トリガに応じて、または時間間隔に基づいてデジタルコンポーネントを選択すると決定し得る。

【００５４】

データ処理システム102は、データリポジトリ124、または1つもしくは複数の3Pデジタルコンテンツプロバイダデバイス160によって提供されたコンテンツを含み得るデータベースからデジタルコンポーネントオブジェクトを選択し、コンピューティングデバイス140によって提示するためのデジタルコンポーネントをネットワーク105を介して提供することができる。コンピューティングデバイス140は、デジタルコンポーネントオブジェクトとインタラクションすることができる。コンピューティングデバイス140は、デジタルコンポーネントに対するオーディオ応答を受信することができる。コンピューティングデバイス140は、コンピューティングデバイス140が商品もしくはサービスプロバイダを特定すること、商品もしくはサービスプロバイダの商品もしくはサービスを要求すること、サービスを実行するようにサービスプロバイダに命令すること、サービスプロバイダに情報を送信すること、またはそうでなければ商品もしくはサービスプロバイダデバイスに問い合わせることを引き起こすまたは可能にするデジタルコンポーネントオブジェクトに関連するハイパーリンクまたはその他のボタンを選択するインジケーションを受信し得る。

【００５５】

データ処理システム102は、要求、クエリ、キーワード、またはコンテンツ選択基準を受信し、受信された情報に基づいてデジタルコンポーネントを選択するためにコンテンツセレクタコンポーネント108を含み得るか、実行し得るか、またはそうでなければコンテンツセレクタコンポーネント108と通信し得る。データ処理システム102は、リアルタイムコンテンツ選択プロセスに入力されたコンテンツ選択基準に基づいてデジタルコンポーネントオブジェクトを選択することができる。データ処理システム102は、複数のサードパーティのコンテンツプロバイダ160によって提供された複数のデジタルコンポーネントオブジェクトを記憶するデータリポジトリ124からデジタルコンポーネントオブジェクトを選択することができる。

【００５６】

データ処理システム102は、データリポジトリ124内のコンテンツデータ132のデータ構造またはデータベースにデジタルコンポーネントオブジェクトを選択するために使用される情報を記憶することができる。コンテンツデータ132は、コンテンツ選択基準、デジタルコンポーネントオブジェクト、履歴的な実行情報、プリファレンス、またはデジタルコンポーネントオブジェクトを選択し、配信するために使用されるその他の情報を含み得る。

【００５７】

コンテンツセレクタコンポーネント108は、リアルタイムコンテンツ選択プロセスを介してデジタルコンポーネントを選択することができる。コンテンツ選択プロセスは、たとえば、検索エンジンによって検索を実行すること、または3Pデジタルコンテンツプロバイダデバイス160などの遠隔のサーバもしくはデバイスに記憶されたデータベースにアクセスすることを含み得る。コンテンツ選択プロセスは、サードパーティのコンテンツプロバイダ160によって提供されたスポンサー付きデジタルコンポーネントオブジェクトを選択することを指し得るかまたは含み得る。リアルタイムコンテンツ選択プロセスは、複数のコンテンツプロバイダによって提供されたデジタルコンポーネントがコンピューティングデバイス140に提供する1つまたは複数のデジタルコンポーネントを選択するために解析されるか、処理されるか、重み付けされるか、またはマッチングされるサービスを含み得る。コンテンツセレクタコンポーネント108は、コンテンツ選択プロセスをリアルタイムで実行することができる。コンテンツ選択プロセスをリアルタイムで実行することは、クライアントコンピューティングデバイス140を介して受信されたコンテンツの要求に応答してコンテンツ選択プロセスを実行することを指し得る。リアルタイムコンテンツ選択プロセスは、要求を受信する時間間隔(たとえば、1秒、2秒、5秒、10秒、20秒、30秒、1分、2分、3分、5分、10分、または20分)以内に実行される(たとえば、開始されるかまたは完了される)ことが可能である。リアルタイムコンテンツ選択プロセスは、クライアントコンピューティングデバイス140との通信セッション中に、または通信セッションが終了された後にある時間間隔以内に実行され得る。リアルタイムコンテンツ選択プロセスは、

オンラインのコンテンツアイテムのオークションを指し得るまたは含み得る。
【００５８】
音声に基づく環境において提示するためのデジタルコンポーネントを選択するために、データ処理システム102は(たとえば、自然言語プロセッサコンポーネント106のNLPコンポーネントによって)、入力オーディオ信号を解析してクエリ、キーワードを特定し、キーワードおよびその他のコンテンツ選択基準を使用して一致するデジタルコンポーネントを選択することができる。コンテンツセレクタコンポーネント108は、要求の前にコンピューティングデバイス140上で実行されるアプリケーション152によってレンダリングされたコンテンツに関連するキーワードに基づいてデジタルコンポーネントオブジェクトを選択し得る。データ処理システム102は、部分一致、完全一致、またはフレーズ一致に基づいて一致するデジタルコンポーネントを選択し得る。たとえば、コンテンツセレクタコンポーネント108は、候補デジタルコンポーネントの主題がクライアントコンピューティングデバイス140のマイクロフォンによって検出された入力オーディオ信号のキーワードまたは語句の主題に対応するかどうかを判定するために候補デジタルコンポーネントの主題を分析するか、解析するか、またはそうでなければ処理することができる。コンテンツセレクタコンポーネント108は、画像処理技術、文字認識技術、自然言語処理技術、またはデータベース検索を使用して候補デジタルコンポーネントの音声、オーディオ、語、文字、テキスト、記号、または画像を特定するか、分析するか、または認識する可能性がある。候補デジタルコンポーネントは、候補デジタルコンポーネントの主題を示すメタデータを含む可能性があり、その場合、コンテンツセレクタコンポーネント108は、候補デジタルコンポーネントの主題が入力オーディオ信号に対応するかどうかを判定するためにメタデータを処理する可能性がある。
【００５９】
3Pデジタルコンテンツプロバイダ160は、デジタルコンポーネントを含むコンテンツキャンペーンを設定するときに追加的なインジケータを提供する可能性がある。コンテンツプロバイダは、候補デジタルコンポーネントについての情報を使用して検索を実行することによってコンテンツセレクタコンポーネント108が特定する可能性があるコンテンツキャンペーンまたはコンテンツグループレベルの情報を提供する可能性がある。たとえば、候補デジタルコンポーネントは、コンテンツグループ、コンテンツキャンペーン、またはコンテンツプロバイダにマッピングされる可能性がある一意識別子を含む可能性がある。コンテンツセレクタコンポーネント108は、データリポジトリ124内のコンテンツキャンペーンデータ構造に記憶された情報に基づいて3Pデジタルコンテンツプロバイダデバイス160についての情報を決定する可能性がある。
【００６０】
コンテンツセレクタコンポーネント108によって選択されるデジタルコンポーネントオブジェクトのフォーマットまたは様式は、ビジュアル、オーディオビジュアル、またはオーディオのみであることが可能である。ビジュアルのみのフォーマットを有するデジタルコンポーネントオブジェクトは、画像またはテキスト付きの画像であることが可能である。オーディオのみのフォーマットを有するデジタルコンポーネントオブジェクトは、オーディオトラックであることが可能である。オーディオビジュアルフォーマットを有するデジタルコンポーネントオブジェクトは、ビデオクリップであることが可能である。デジタルコンポーネントオブジェクトは、デジタルコンポーネントオブジェクトのフォーマットに基づいて異なる種類のインタラクションのために構成され得る。たとえば、ビジュアルのみのデジタルコンポーネントオブジェクトは、(たとえば、デジタルコンポーネントオブジェクトに埋め込まれたハイパーリンクを選択するための)キーボード、マウス、またはタッチスクリーン入力によるインタラクションのために構成され得る。オーディオのみのデジタルコンポーネントオブジェクトは、音声入力によるインタラクションのために構成され得る(たとえば、アクションを実行するための予め決められたキーワードを検出するように構成され得る)。
【００６１】

しかし、コンテンツセレクタコンポーネント108によって選択されるデジタルコンポーネントオブジェクトのフォーマットは、コンピューティングデバイス140と互換性がない可能性があり、またはコンピューティングデバイス140による提示のために最適化されない可能性がある。場合によっては、データ処理システム102は、コンピューティングデバイス140と互換性があるかまたはコンピューティングデバイス140のために最適化されるデジタルコンポーネントオブジェクトに関するフィルタリングを行い得る。フォーマットまたは様式に基づくフィルタリングは、コンピューティングデバイス140と互換性があるかまたはコンピューティングデバイス140のために最適化されるフォーマットを有するデジタルコンポーネントオブジェクトを選択する結果となり得る。フォーマットまたは様式に基づくフィルタリングは、デジタルコンポーネントオブジェクトのフォーマットがコンピューティングデバイス140と互換性がないまたはコンピューティングデバイス140のために最適化されない可能性があるために、コンテンツ選択基準と関連性があるまたはより適合する可能性があるデジタルコンポーネントオブジェクトの選択を防止する可能性がある。たとえば、コンピューティングデバイス140がディスプレイデバイス146のないスマートスピーカである場合、ビジュアルコンテンツアイテムは、フィルタリングして取り除かれるかまたは選択を防止される可能性があり、選択をオーディオのみのコンテンツアイテムに制限する。ビジュアルコンテンツアイテムが選択されるオーディオのみのコンテンツアイテムよりもコンテンツ選択基準に適合するキーワードを含んでいたとした場合でも、データ処理システム102は、最も一致するコンテンツアイテムを提供することができない可能性がある。最も一致するコンテンツアイテムを提供しないことによって、データ処理システム102は、提示するために関連性のないコンテンツアイテムを提供することによってコンピューティングリソースの消費、ネットワーク帯域幅、またはコンピューティングデバイス140のバッテリ電力を無駄にしている可能性がある。

【００６２】

したがって、この技術的な解決策は、コンテンツ選択基準に基づいて、フォーマットに関係なく最も一致するデジタルコンポーネントオブジェクトを選択し、それから、デジタルコンポーネントオブジェクトをコンピューティングデバイス140に最適化されたまたはコンピューティングデバイス140と互換性があるフォーマットに変換することができる。フォーマットまたは様式が原因でコンテンツアイテムを取り除かないまたはコンテンツアイテムが選択されることを妨げないことによって、この技術的な解決策は、関連性およびその他のコンテンツ選択基準に基づいて最も高いランク付けのコンテンツアイテムを選択肢、それから、コンテンツアイテムを所望のフォーマットにリアルタイムで変換することができる。

【００６３】

そのようにするために、データ処理システム102は、3Pデジタルコンテンツプロバイダデバイス160によって提供されたデジタルコンポーネントオブジェクトの元のフォーマットと異なるフォーマットのデジタルコンポーネントオブジェクトを生成するように設計され、構築され、動作可能なコンテンツ変換コンポーネント110を含み得る。3Pデジタルコンテンツプロバイダデバイス160は、第1のフォーマットの元のコンテンツアイテムを提供することが可能であり、データ処理システム102は、元のコンテンツアイテムに基づく第2のフォーマットの第2のコンテンツアイテムを生成することが可能である。たとえば、コンテンツセレクタコンポーネント108は、要求またはコンテンツ選択基準に基づいて、ビジュアル出力フォーマットを有するデジタルコンポーネントオブジェクトを選択することができる。コンテンツ変換コンポーネント110は、コンピューティングデバイス140がディスプレイデバイスを持たないが、オーディオインターフェースを有すると判定し得る。それから、コンテンツ変換コンポーネント110は、オーディオのみのフォーマットを有する新しいデジタルコンポーネントオブジェクトを生成することができる。コンテンツ変換コンポーネント110は、デジタルコンポーネントオブジェクトに関連するメタデータを使用して、オーディオのみのフォーマットを有する新しいデジタルコンポーネントオブジェクトを生成することができる。コンテンツ変換コンポーネント110は、新しいコンテン

ツアイテムのためのフォーマットを選択し、元のコンテンツアイテムに基づいてテキストを生成し、新しいコンテンツアイテムのための音声を選択し、新しいコンテンツアイテムのための口頭でないオーディオの合図を生成し、新しいコンテンツアイテムとインタラクションするために使用されるアクションを生成し、それから、コンピューティングデバイス140に提供する新しいコンテンツアイテムを生成することができる。

【００６４】

コンテンツ変換コンポーネント110は、コンテンツセレクタコンポーネント108によって選択されたデジタルコンポーネントオブジェクトに基づいてデジタルコンポーネントオブジェクトを生成すべきフォーマットを選択するように設計され、構築され、動作可能なフォーマットセレクタ112を含み得る。フォーマットセレクタ112は、デジタルコンポーネントオブジェクトを変換すべきフォーマットを決定するために様々な技術または要因を使用し得る。要因は、たとえば、コンピューティングデバイス140の種類、コンピューティングデバイス140の利用可能なインターフェース、コンピューティングデバイス140の残バッテリ電力、コンピューティングデバイス140の場所、コンピューティングデバイス140に関連する運搬の仕方(たとえば、車、列車、飛行機、歩行、駆け足、自転車、もしくは静止)、コンピューティングデバイス140のフォアグラウンドで実行されるアプリケーション152の種類、コンピューティングデバイス140の状態、またはその他の要因を含み得る。場合によっては、要因は、料理をする、仕事をする、またはくつろぐなどのユーザの活動を含み得る。データ処理システム102は、時刻または最近の検索活動(たとえば、レシピの検索)に基づいてユーザが料理をしていると判定し得る。データ処理システム102は、時刻、曜日、および場所(たとえば、職場)に基づいてユーザが仕事をしているかどうかを判定し得る。データ処理システム102は、時刻、場所、およびコンピューティングデバイス140上での活動(たとえば、映画のストリーミング)に基づいてユーザがくつろいでいるかどうかを判定し得る。

【００６５】

フォーマットセレクタ112は、選択されたデジタルコンポーネントオブジェクトを変換すべきフォーマットをコンピューティングデバイス140の種類に基づいて選択することができる。フォーマットセレクタ112は、コンテンツの要求と一緒にコンピューティングデバイス140の種類についての情報を受信し得る。たとえば、コンピューティングデバイス140によって与えられたコンテンツの要求が、コンピューティングデバイス140の種類を示し得る。コンテンツの要求が受信されない場合、フォーマットセレクタ112は、コンピューティングデバイス140に関連するアカウント情報もしくはプロファイル情報、またはコンピューティングデバイス140上で実行されるアプリケーション152から受信された情報に基づいてコンピューティングデバイス140の種類を判定し得る。場合によっては、フォーマットセレクタ112は、コンピューティングデバイス140の種類についての情報に関してコンピューティングデバイス140に問い合わせることができる。コンピューティングデバイス140の例示的な種類は、ラップトップ、タブレット、スマートウォッチ、ウェアラブルデバイス、スマートフォン、スマートスピーカ、スマートテレビ、またはモノのインターネットデバイス(たとえば、スマート家電もしくはスマート照明)を含み得る。デバイスの種類は、コンピューティングデバイス140上で利用可能なインターフェースの種類(たとえば、ビジュアル出力インターフェース、オーディオ出力インターフェース、オーディオ入力インターフェース、タッチ入力インターフェース、またはキーボードおよびマウスインターフェース)を示し得る。たとえば、コンピューティングデバイス140の種類がスマートスピーカである場合、データ処理システム102は、コンピューティングデバイス140のための主なインターフェースがオーディオインターフェースであり、コンピューティングデバイス140がディスプレイデバイスを持たないと判定し得る。フォーマットセレクタ112は、デバイスの種類のための主なインターフェースがオーディオのみのインターフェースであることに応じて、元のビジュアルデジタルコンポーネントオブジェクトをオーディオのみのフォーマットのデジタルコンポーネントオブジェクトに変換すると決定し得る。別の例において、コンピューティングデバイスの種類がスマートテレビである場合、

データ処理システム102は、主なインターフェースがオーディオビジュアルインターフェースであると判定し得る。フォーマットセレクタ112は、主なインターフェースがオーディオビジュアルインターフェースであると判定することに応じて、元のビジュアルのみのデジタルコンポーネントオブジェクトをオーディオビジュアルデジタルコンポーネントオブジェクトに変換すると決定し得る。デジタルコンポーネントオブジェクトをコンピューティングデバイス140の種類のための主なフォーマットに変換することによって、データ処理システム102は、コンピューティングデバイス140におけるデジタルコンポーネントオブジェクトのレンダリングまたは提示を最適化することができる。レンダリングまたは提示を最適化することは、コンピューティングデバイス140の主なユーザインターフェースまたはユーザインターフェースの主な組合せを使用してデジタルコンポーネントオブジェクトを出力することを指し得る。

【００６６】

フォーマットセレクタ112は、コンピューティングデバイス140の利用可能なインターフェースに基づいて変換のためのフォーマットを選択することができる。コンピューティングデバイス140の種類は、コンピューティングデバイス140が含むインターフェースの種類を示し得る。しかし、インターフェースのうちの1つまたは複数が、利用不可能である可能性があり、その場合、フォーマットセレクタ112は、利用可能なインターフェースを特定し、それから、デジタルコンポーネントオブジェクトを利用可能なインターフェースに対応するフォーマットに変換することができる。たとえば、コンピューティングデバイス140は、ディスプレイデバイス146を無効化しながらまたはオフにしながらストリーミングデジタル音楽などのオーディオを出力することができ、それが、電力消費を削減することができる。この例において、フォーマットセレクタ112は、ディスプレイデバイス146がオフにされたためにビジュアルインターフェースが現在利用不可能であると判定し得るが、オーディオ出力インターフェースがオーディオを現在出力しているのでそのオーディオ出力インターフェースが利用可能であると判定し得る。別の例において、フォーマットセレクタ112は、オーディオがミュートされた場合、オーディオインターフェースが利用不可能であると判定し、ディスプレイデバイス146がビジュアル出力を活発に提供している場合、ビジュアル出力インターフェースが利用可能であると判定し得る。したがって、オーディオインターフェースが利用不可能である場合、フォーマットセレクタ112は、デジタルコンポーネントオブジェクトのためにビジュアル出力フォーマットを選択することが可能であり、ビジュアルインターフェースが利用不可能である場合、フォーマットセレクタ112は、デジタルコンポーネントオブジェクトのためにオーディオ出力フォーマットを選択することが可能である。ビジュアル出力インターフェースもオーディオ出力インターフェースも利用可能でない場合、フォーマットセレクタ112は、無駄なコンピューティングリソースの利用およびネットワーク帯域幅の利用を避けるためにコンテンツの変換を終了し、デジタルコンポーネントオブジェクトの配信を遮断し得る。

【００６７】

フォーマットセレクタ112は、コンピューティングデバイス140の残バッテリ電力に基づいて出力インターフェースを決定することができる。たとえば、残バッテリ電力が閾値(たとえば、10%、15%、20%、25%、またはその他の閾値)未満である場合、フォーマットセレクタ112は、ディスプレイデバイスに比べて少ないエネルギーを消費し得るオーディオ出力などの、レンダリングするために最も少ない量のエネルギーを利用するフォーマットを選択すると決定し得る。

【００６８】

フォーマットセレクタ112は、コンピューティングデバイス140の運搬の仕方に基づいてデジタルコンポーネントオブジェクトのためのフォーマットを選択することができる。運搬の例示的な仕方は、車、列車、飛行機、歩行、駆け足、自転車、または静止(たとえば、移動しないもしくは運搬なし)を含み得る。フォーマットセレクタ112は、運搬の仕方が車、駆け足、または自転車である場合、運搬のそれらの仕方ではユーザがビジュアル出力を知覚することができない可能性があるので、ユーザの気を散らすことを避け、無駄なエ

ネルギー消費を避けるためにオーディオのみの出力フォーマットを選択し得る。運搬の仕方が歩行、静止、公共交通機関、または飛行機である場合、フォーマットセレクタ112は、ビジュアル出力がユーザの気を散らさない可能性があり、ユーザがビジュアル出力を知覚することができる可能性が高いので、ビジュアル出力またはオーディオビジュアル出力フォーマットを選択し得る。

【００６９】

フォーマットセレクタ112は、コンピューティングデバイス140のフォアグラウンドで実行されるアプリケーション152の種類に基づいてデジタルコンポーネントオブジェクトのためのフォーマットを選択することができる。アプリケーション152の主な出力インターフェースがデジタル音楽ストリーミングアプリケーション152のようにオーディオのみである場合、フォーマットセレクタ112は、たとえば、オーディオのみのフォーマットを選択し得る。アプリケーション152の主な出力インターフェースがビジュアルのみのフォーマットである場合、フォーマットセレクタ112は、ビジュアルのみの出力を選択し得る。アプリケーション152の主な出力インターフェースがデジタルビデオストリーミングアプリケーション152のようにオーディオビジュアル出力の組合せである場合、フォーマットセレクタ112は、デジタルコンポーネントオブジェクトのためにオーディオビジュアル出力フォーマットを選択し得る。

【００７０】

フォーマットセレクタ112は、コンピューティングデバイス140の種類がデジタルアシスタントデバイスであることに基づいて、またはコンピューティングデバイス140がデジタルアシスタントアプリケーションを含むアプリケーション152を実行することに基づいてデジタルコンポーネントオブジェクトのためのフォーマットを選択し得る。デジタルアシスタントアプリケーションは、バーチャルアシスタントを指し得るかまたは含み得る。デジタルアシスタントアプリケーションは、コマンドまたは質問に基づいてタスクまたはサービスを実行することができるソフトウェアエージェントを含み得る。デジタルアシスタントアプリケーション152は、(たとえば、ユーザによって話された)自然言語入力を受け取り、処理し、それから、入力に応じてタスク、アクションを実行するかまたは提供するように構成され得る。フォーマットセレクタ112は、デジタルアシスタントアプリケーション152の主なインターフェースが音声に基づく(またはオーディオに基づく)インターフェースであることが可能であるので、アプリケーション152またはコンピューティングデバイス140の種類がデジタルアシスタントであることに応じて、デジタルコンポーネントオブジェクトのためにオーディオのみのフォーマットを選択すると決定し得る。

【００７１】

コンテンツ変換コンポーネント110は、デジタルコンポーネントオブジェクトに基づいてテキストを生成するように設計され、構築され、動作可能なテキストジェネレータ114を含み得る。たとえば、フォーマットセレクタ112がビジュアルデジタルコンポーネントオブジェクトをオーディオのみのデジタルコンポーネントオブジェクトに変換すると決定することに応じて、テキストジェネレータ114は、デジタルコンポーネントオブジェクトを処理して、オーディオによって出力され得るテキストを生成することができる。ビジュアルコンポーネントオブジェクトに基づいてテキストを生成するために、テキストジェネレータ114は、ビジュアルコンポーネント内のテキストを解析し得るか、ビジュアルデジタルコンポーネントオブジェクトを処理するために画像処理技術を適用し得るか、または光学式文字認識技術を適用し得る。テキストジェネレータ114は、ビジュアルコンポーネントオブジェクトに関連するメタデータを取得し、メタデータを解析または処理してテキストを生成し得る。メタデータは、たとえば、製品の仕様または製品の説明を含み得る。したがって、テキストジェネレータ114は、ビジュアルデジタルコンポーネント内のテキスト、デジタルコンポーネントオブジェクトに埋め込まれたハイパーリンクもしくはユニフォームリソースロケータ、製品へのリンク、または製品の説明のタプルを使用し得る。

【００７２】

テキストジェネレータ114は、ビジュアルデジタルコンポーネントから取得されたテキ

スト、メタデータ、または対応するリンクのタプルを、テキストを生成するための自然言語生成モデルに入力し得る。テキストジェネレータ114は、自然言語生成エンジンもしくはコンポーネントを含み得るか、自然言語生成エンジンもしくはコンポーネントを用いて構成され得るか、または自然言語生成エンジンもしくはコンポーネントにアクセスし得る。自然言語生成は、構造化されたデータを自然言語に変換するプロセスを指し得る。テキストジェネレータ114は、自然言語生成を使用して、テキストトゥスピーチシステムによって読み出され得るテキストを生成することができる。

【００７３】

自然言語生成技術を用いて構築されたテキストジェネレータ114は、コンテンツの決定(たとえば、テキスト内でどんな情報を述べるべきかを決定すること)、ドキュメントの構造化(たとえば、伝える情報の全体的な編成)、集約(たとえば、読みやすさおよび自然さを高めるための類似した文の合併)、語彙の選択(たとえば、概念を言葉で表現すること)、指示表現(referring expression)の生成(たとえば、オブジェクトおよび領域を特定する指示表現を作成すること)、ならびに具現化(たとえば、シンタックス、形態論、および正書法の規則に従って適正であり得る実際のテキストを作成すること)などの複数の段階でテキストを生成することができる。

【００７４】

テキストジェネレータ114は、人間によって書かれたテキストの大きなコーパス上などで機械学習を使用して統計モデルを訓練することによって自然言語生成を実行することができる。機械学習は、たとえば、モデルを訓練するために3Pデジタルコンテンツプロバイダデバイス160によって提供されたデジタルコンポーネントオブジェクトに対応する人間によって書かれたテキストを処理することができる。

【００７５】

テキストジェネレータ114は、シーケンストゥシーケンス(sequence-to-sequence)モデルを使用してテキストを生成することができる。シーケンストゥシーケンスモデルは、2つの部分、エンコーダおよびデコーダを含み得る。エンコーダおよびデコーダは、1つのネットワークへと組み合わされる2つの異なるニューラルネットワークモデルであることが可能である。ニューラルネットワークは、長期短期記憶(「LSTM: long short-term memory」)ブロックなどの再帰型ニューラルネットワーク(「RNN」)であることが可能である。ネットワークのエンコーダ部分は、入力シーケンス(たとえば、ビジュアルデジタルコンポーネント内のテキスト、デジタルコンポーネントオブジェクト内に埋め込まれたハイパーリンクもしくはユニフォームリソースロケータ、製品へのリンク、または製品の説明に対応するタプル)を理解し、それから、入力のより低い次元の表現を作成するように構成され得る。エンコーダは、この表現をデコーダネットワークに転送することができ、デコーダネットワークは、出力を表すシーケンスを生成するように構成され得る。デコーダは、デコーダの反復の各時間ステップにおいて単語を1つずつ生成することができる。

【００７６】

テキストジェネレータ114は、テキストを生成するために敵対的生成ネットワーク(「GAN: generative adversarial network」)を使用することができる。GANは、生成されたテキストが「本物」かまたは「偽物」かを検出するように構成される敵対者(adversary)(たとえば、判別器(discriminator)ネットワーク)を導入することによって本物らしいサンプルを生成するように訓練される生成器(generator)ネットワークを指し得る。たとえば、判別器は、生成器を調整するために使用される動的に更新される評価測定基準であり得る。GANの生成器および判別器は、平衡点が達せられるまで継続的に改善し得る。

【００７７】

したがって、テキストジェネレータ114は、自然言語生成技術を使用してビジュアルデジタルコンポーネントオブジェクトに基づいてテキストを生成することができる。コンテンツ変換コンポーネント110は、テキストをスピーチとして出力するために使用するデジタル声紋を選択することができる。コンテンツ変換コンポーネント110は、テキストをレンダリングするためのデジタル音声を選択するように設計され、構築され、動作可能な音

声セレクタ116を含み得る。コンテンツ変換コンポーネント110は、デジタルコンポーネントオブジェクトのコンテキストに基づいて、または生成されたテキストに基づいてデジタル音声を選択することができる。音声セレクタ116は、デジタルコンポーネントオブジェクトの種類またはテキストのコンテキストに合致するデジタル音声を選択することができる。たとえば、音声セレクタ116は、枕の広告のために、アクション映画の広告に比べて異なるデジタル音声を選択し得る。

【００７８】

テキストジェネレータ114によって生成されたテキストのオーディオトラックを生成するためのデジタル声紋を選択するために、音声セレクタ116は、履歴的なデータを使用して機械学習エンジン122によって訓練された音声モデル126を使用することができる。音声モデル126を訓練するために使用される履歴的なデータは、たとえば、コンピューティングデバイス140またはその他の媒体を介して提示するために3Pデジタルコンテンツプロバイダによって作成されたオーディオデジタルコンポーネントオブジェクトを含み得る。履歴的なデータは、3Pデジタルコンテンツプロバイダによって作成されたオーディオデジタルコンポーネントオブジェクトの各々に関連するメタデータまたはコンテキスト情報を含み得る。メタデータまたはコンテキスト情報は、たとえば、話題、概念、キーワード、地理的場所、ブランド名、バーティカルカテゴリ、製品カテゴリ、サービスカテゴリ、またはオーディオデジタルコンポーネントオブジェクトの様相を説明するその他の情報を含み得る。履歴的なデータは、オーディオデジタルコンポーネントオブジェクトに関連する実行情報を含み得る。実行情報は、オーディオデジタルコンポーネントオブジェクト上の選択またはコンバージョンなど、エンドユーザがオーディオデジタルコンポーネントとインタラクションしたかどうかを示し得る。

【００７９】

たとえば、履歴的なデジタルコンポーネントは、テレビ、ラジオ、またはコンピューティングデバイス140で放送するために3Pコンテンツプロバイダによって作成されたラジオ広告(たとえば、放送ラジオまたはデジタルストリーミングラジオ局)、テレビ広告(たとえば、放送もしくはケーブルテレビ、またはデジタルストリーミングテレビチャンネル)を含み得る。これらの履歴的なデジタルコンポーネントは、それらがテレビ上で提示されている場合、オーディオおよびビジュアルコンポーネントを含み得る。テレビ広告に関連するメタデータまたはコンテキスト情報は、製品の種類(たとえば、自動車、旅行、家庭用電化製品、もしくは食品)、サービスの種類(たとえば、税務サービス、電話サービス、インターネットサービス、レストラン、配送サービス、もしくは家事サービス)、製品もしくはサービスについての説明情報、製品もしくはサービスを提供する会社もしくは主体についての情報、広告が提供されるべき地理的場所(たとえば、州、地理的領域、都市、もしくは郵便番号)、またはその他のキーワードを含み得る。したがって、履歴的なデータは、オーディオ(またはオーディオビデオ)3Pデジタルコンポーネントオブジェクトに対応するオーディオトラックおよびオーディオトラックに関連するメタデータを含み得る。履歴的な3Pデジタルコンポーネントオブジェクトを記憶する例示的なデータ構造が、Table 1(表1)に示される。

【００８０】

【表１】

一意ID	オーディオファイル	製品/サービス	バーティカル	場所	ブランド	説明
1	Audio_1.mp3	製品	自動車	USA	Company_A	高級スポーツカー
2	Audio_2.mp3	サービス	銀行	ニューイングランド	Company_B	低金利のクレジットカードの申し出

Table 1: 履歴的なデータの説明のための例

【００８１】

Table 1(表1)は、テキストジェネレータ114によって生成されたテキストレンダリングするために使用するデジタル音声を選択するために音声セレクタ116によって使用される音声モデル126を訓練するために機械学習エンジン122によって使用される履歴的なデータの説明のための例を提供する。Table 1(表1)に示されるように、それぞれの履歴的な3Pデジタルコンポーネントオブジェクトは、オーディオトラック(たとえば、Audio_1.mp3およびAudio_2.mp3)、広告が製品のためのものであるのかまたはサービスのためのものであるのかのインジケーション、バーティカル市場のインジケーション(たとえば、自動車または銀行)、広告が提供される場所のインジケーション(たとえば、米国全土、またはニューイングランドなどの地理的領域)、広告のブランドまたはプロバイダ(たとえば、Company_AまたはCompany_B)、ならびにデジタルコンポーネントオブジェクトに関連する追加的な説明またはキーワード(たとえば、高級スポーツカー、または低金利のクレジットカードの申し出)を含み得る。オーディオファイルは、たとえば、.wav、.mp3、.aac、または任意のその他のオーディオフォーマットを含む任意のフォーマットであることが可能である。場合によっては、履歴的なデジタルコンポーネントオブジェクトは、オーディオとビデオとの両方を含むことが可能であり、その場合、オーディオファイルフォーマットは.mp4、.mov、.wmv、.flv、またはその他のファイルフォーマットなどのオーディオおよびビジュアルファイルフォーマットを指し得る。

【００８２】

データ処理システム102は、履歴的なデジタルコンポーネントデータを前処理して、データを、機械学習エンジン122が音声モデル126を訓練するためにデータを処理するのに好適なフォーマットにすることができる。たとえば、音声セレクタ116または機械学習エンジン122は、履歴的なデジタルコンポーネントデータを処理してデータ内の特徴を特定するためのオーディオ処理技術または解析技術を用いて構成され得る。特徴は、たとえば、オーディオファイル内のオーディオの特性、製品/サービス、バーティカルカテゴリ、説明からのキーワード、またはその他の情報を含み得る。例示的なオーディオの特性は、音声の性別、音声の年齢層、高さ、周波数、振幅もしくは音量、イントネーション、方言、言語、アクセント、言葉が話される速さ、またはその他の特性を含み得る。

【００８３】

機械学習エンジン122は、履歴的なデータを分析し、音声モデル126を訓練するために任意の機械学習または統計技術を使用することができる。機械学習エンジン122は、予測または判断を行うためにサンプルデータまたは訓練データ(たとえば、履歴的なデジタルコンポーネントオブジェクト)に基づいてモデルを構築することができる学習技術または機能を用いて構成され得る。機械学習エンジン122は、教師ありもしくは教師なし学習技術、半教師あり学習、強化学習、自己学習、特徴学習、スパース辞書学習(sparse dictionary learning)、または相関ルール(association rule)を用いて構成され得る。機械学習を実行するために、機械学習エンジン122は、訓練データで訓練された音声モデル126を作成

することができる。モデルは、たとえば、人工ニューラルネットワーク、決定木、サポートベクターマシン、回帰分析、ベイジアンネットワーク、または遺伝的アルゴリズムに基づき得る。

【００８４】

音声セレクタ116は、ビジュアルデジタルコンポーネントに基づいてテキストジェネレータ114によって生成されたテキストを受信すると、機械学習エンジン122によって訓練された音声モデル126を使用してデジタル声紋を選択するために、テキストをビジュアルデジタルコンポーネントに関連するメタデータと一緒に使用することができる。たとえば、データ処理システム102は、音声モデル126によって、デジタルコンポーネントオブジェクトのコンテキストに基づいてデジタル音声を選択することができる。コンテキストは、デジタルコンポーネントオブジェクトに関連するテキスト、メタデータ、またはその他の情報を含み得るか、指し得る。コンテキストは、コンピューティングデバイス140に関連する情報を指し得るかまたは含み得る。場合によっては、音声セレクタ116は、運搬の仕方、場所、プリファレンス、実行情報、またはコンピューティングデバイス140に関連するその他の情報などのコンピューティングデバイス140のコンテキストに基づいてデジタル音声を選択し得る。

【００８５】

音声セレクタ116は、音声の特性のベクトルを生成し、それから、テキストをレンダリングするためのデジタル音声を選択するために、音声モデル126にデジタルコンポーネントオブジェクトのコンテキストを入力することができる。テキストおよびメタデータは、製品、サービス、バーティカルカテゴリ、キーワードについての情報、または音声モデル126に入力され得るその他の情報を示すことができる。音声モデル126への入力は、ビジュアルデジタルコンポーネントに基づいて生成されたテキスト、またはテキストとビジュアルデジタルコンポーネントのメタデータとの組合せであることが可能である。音声モデル126の出力は、テキストをレンダリングするために使用するデジタル声紋の特徴を予測する音声の特性のベクトルであることが可能である。出力は、デジタル声紋の性別(たとえば、男性もしくは女性)、イントネーション(たとえば、イントネーションはテキストの各音節に関して変わり得る)、アクセント、発声、高さ、音の大きさ、スピーチの速さ、トーン、テクスチャ(texture)、音の大きさ、またはその他の情報を示し得る。音声モデル126の出力は、バス、バリトン、テノール、アルト、メゾソプラノ、およびソプラノなどのその他の音声の種類を含み得る。

【００８６】

音声セレクタ116は、一致するデジタル声紋または最も一致するデジタル声紋を特定するために、音声モデル126によって出力された音声の特性のベクトルをデータリポジトリ124に記憶された利用可能なデジタル声紋と比較することができる。デジタル声紋は、性別、アクセント、発声、高さ、音の大きさ、スピーチの速さ、またはその他の情報に基づいてカテゴリ分けされ得る。音声セレクタ116は、テキストをレンダリングするために使用する最も一致するデジタル声紋を選択するために、音声モデル126の出力を記憶されたまたは利用可能なデジタル声紋と比較することができる。音声セレクタ116は、一致する声紋を選択するための特性を重み付けし得る。たとえば、性別などの特性が、アクセントなどの特性よりも重く重み付けされ得る。スピーチの速さなどの特性が、アクセントなどの特性よりも重く重み付けされ得る。場合によっては、音声セレクタ116は、最も多くの特性と一致するデジタル音声を選択し得る。音声セレクタ116は、音声モデル126の出力に基づいてデジタル声紋を選択するために任意のマッチング技術を使用することができる。

【００８７】

選択されたデジタル声紋は、デジタル声紋を特定する一意識別子を含み得る。デジタル声紋は、コンテンツ変換コンポーネント110がテキストトゥスピーチを実行するために使用することができる情報を含み得る。デジタル声紋は、テキストトゥスピーチエンジンのための命令を含み得る。コンテンツ変換コンポーネント110は、デジタル声紋によって示される音声の特性を使用してテキストをレンダリングするために任意の種類のテキストト

ゥスピーチテクノロジーを使用することができる。たとえば、コンテンツ変換コンポーネント110は、デジタル声紋によって定義される人間のような音声を使用してテキストをレンダリングするためにニューラルネットワーク技術を使用することができる。

【００８８】

コンテンツ変換コンポーネント110は、テキストトゥスピーチ技術を使用して、選択されたデジタル声紋に基づいてレンダリングされる基礎となるオーディオトラックを生成することができる。たとえば、コンテンツ変換コンポーネント110は、デジタル音声によってレンダリングされるテキストを用いてデジタルコンポーネントオブジェクトの基礎となるオーディオトラックを構築するように設計され、構築され、動作可能なオーディオの合図ジェネレータ118を含み得る。オーディオの合図ジェネレータ118は、テキストトゥスピーチエンジンを使用して、デジタル音声によってテキストをレンダリングまたは合成することができる。

【００８９】

コンテンツ変換コンポーネント110は、基礎となるオーディオトラックに口頭でないオーディオの合図を追加すると決定し得る。オーディオの合図ジェネレータ118は、基礎となるオーディオトラックに追加する口頭でないオーディオの合図を生成するように設計され、構成され、動作可能であり得る。口頭でない合図は、効果音を含み得るかまたは指し得る。口頭でない合図は、たとえば、海の波、風、木の葉のそよぎ、自動車のエンジン、車の運転、飛行機の離陸、群衆の声援、スポーツ、アクション映画の効果(たとえば、高速のカーチェイス、ヘリコプターなど)、駆け足、自転車の運転、またはその他の効果音の音を含み得る。したがって、口頭でないオーディオの合図は、単語または数を用いたスピーチ(たとえば、口頭で伝えられる言葉)を含まない音または効果音を指し得る。

【００９０】

オーディオの合図ジェネレータ118は、デジタルコンポーネントオブジェクトのテキストまたはメタデータに基づいて1つまたは複数の口頭でないオーディオの合図を生成し得る。オーディオの合図ジェネレータ118は、テキストのために選択されたデジタル音声に基づいて1つまたは複数の口頭でないオーディオの合図を生成し得る。オーディオの合図ジェネレータ118は、コンピューティングデバイス140のコンテキスト(たとえば、運搬の仕方、コンピューティングデバイスの種類、コンピューティングデバイス140のフォアグラウンドで実行されるアプリケーション152の種類、アプリケーション152において提示されているコンテンツ、またはコンピューティングデバイス140から受信された要求)に基づいて口頭でないオーディオの合図を選択し得る。

【００９１】

オーディオの合図ジェネレータ118は、オーディオの合図モデル134またはオーディオの合図データストアを使用して基礎となるオーディオトラックに追加する1つまたは複数の口頭でないオーディオの合図を選択し得る。オーディオの合図134のデータストアは、効果音があることのインジケータなどのメタデータによってタグ付けされる効果音を含み得る。たとえば、海の波の効果音は、「海の波」などの効果音の説明によってタグ付けされ得る。オーディオの合図134のデータストアは、複数の種類の海の波の効果音を含むことが可能であり、それぞれの海の波を区別する対応するタグまたは説明を含むことが可能である。

【００９２】

場合によっては、オーディオの合図は、デジタル音声内の特性のために構成され得るかまたは最適化され得る。特定のオーディオの合図は、特定の音声の特性のベクトルを有するデジタル音声のための背景効果音として提示するために最適化される可能性がある。最適化された効果音は、テキストの目的がユーザによって理解可能であるようにして、テキストを妨害または邪魔することなくデジタル音声によるテキストのレンダリングと一緒にレンダリングされ、それによって、改善されたユーザインターフェースを提供することができる効果音を指し得る。たとえば、デジタル音声と同じ周波数および振幅を有する効果音は、効果音のある中でデジタル音声を知覚、識別、または区別することを難しくする可

能性があり、これは、効果のない出力を提供することによって低下したユーザエクスペリエンスおよび無駄なコンピューティングリソースをもたらし得る。

【００９３】

オーディオの合図ジェネレータ118は、ビジュアルデジタルコンポーネントに対して画像認識を実行してデジタルコンポーネントオブジェクト内のビジュアルオブジェクトを特定することができる。オーディオの合図ジェネレータ118は、ビジュアルデジタルコンポーネントに関連する任意のテキストを無視し、画像認識を実行してオブジェクトを検出し得る。オーディオの合図ジェネレータ118は、任意の画像処理技術またはオブジェクト検出技術を使用することができる。オーディオの合図ジェネレータ118は、機械学習エンジン122によって、オブジェクトの説明によってタグ付けされるオブジェクトの画像を含む訓練データセットに基づいて訓練されたモデルを使用することができる。機械学習エンジン122は、オーディオの合図ジェネレータ118がモデルに入力される新しい画像内のオブジェクトを検出するためにモデルを使用し得るように、訓練データを用いてモデルを訓練することができる。オーディオの合図ジェネレータ118は、訓練されたモデルを使用してビジュアルデジタルコンポーネントオブジェクト内の画像を検出することができる。したがって、オーディオの合図ジェネレータ118は、デジタルコンポーネントオブジェクト内のビジュアルオブジェクトを特定するためにビジュアルデジタルコンポーネントオブジェクトに対して画像認識を実行することができる。オーディオの合図ジェネレータ118は、データリポジトリ124に記憶された口頭でないオーディオの合図134から、ビジュアルオブジェクトに対応する口頭でないオーディオの合図を選択することができる。

【００９４】

また、オーディオの合図ジェネレータ118は、デジタルコンポーネントオブジェクトに対応するランディングウェブページへのリンクなどのビジュアルデジタルコンポーネントに埋め込まれたリンクにアクセスすることによってオブジェクトを特定し得る。オーディオの合図ジェネレータ118は、ウェブページを解析してビジュアルオブジェクトおよび追加的なコンテキスト情報、キーワード、またはメタデータを特定することができる。オーディオの合図ジェネレータ118は、オブジェクト、コンテキスト情報、キーワード、またはメタデータに基づいてオーディオの合図を選択することができる。たとえば、ランディングウェブページのテキストは、キーワード「ビーチ、休暇、クルーズ」を含み得る。オーディオの合図ジェネレータ118は、キーワードのうちの1つまたは複数に対応するオーディオの合図を選択することができる。

【００９５】

オーディオの合図ジェネレータ118がビジュアルオブジェクト内の画像またはビジュアルデジタルコンポーネントもしくはデジタルコンポーネントにリンクされたウェブページのメタデータに関連するその他のキーワードもしくはコンテキスト情報に基づいて複数の候補のオーディオの合図を特定する場合、オーディオの合図ジェネレータ118は、1つまたは複数の口頭でないオーディオの合図を選択し得る。オーディオの合図ジェネレータ118は、いくつの口頭でないオーディオの合図を選択すべきかを決定するためのポリシーを使用し得る。ポリシーは、すべての特定されたオーディオの合図を選択すること、予め決められた数のオーディオの合図をランダムに選択すること、オーディオトラック全体を通じて異なるオーディオの合図を交替で選択すること、1つもしくは複数のオーディオの合図を重ね合わせるかもしくはミックスすること、または予め決められた数の最も高いランク付けのオーディオの合図を選択することであり得る。

【００９６】

たとえば、オーディオの合図ジェネレータ118は、ビジュアルデジタルコンポーネント内の最も目立つオブジェクトを特定し、最も目立つオブジェクトに対応するオーディオの合図を選択することができる。オブジェクトが目立つことは、ビジュアルデジタルコンポーネント内のオブジェクトのサイズを指す可能性がある(たとえば、ビジュアルデジタルコンポーネント内の最も大きなオブジェクトが、最も目立つオブジェクトであり得る)。目立つことは、オブジェクトが背景にあるのとは対照的に画像の前景にあることに基づき得る

。オーディオの合図ジェネレータ118は、テキストジェネレータ114によって生成されたテキストに最も関連性があるビジュアルデジタルコンポーネント内のオブジェクトを特定することができる。関連性は、オブジェクトの説明およびテキストに基づいて決定され得る。たとえば、生成されたテキストがオブジェクトの名前およびオブジェクトの説明内のキーワードを含む場合、オブジェクトは、テキストに関連性があると判定される可能性がある。オーディオの合図ジェネレータ118は、テキストに最も関連性があるキーワードまたは概念を決定し、オーディオの合図に関してそれらのキーワードを選択し得る。

【００９７】

オーディオの合図ジェネレータ118は、目立つこと、関連性、または目立つことと関連性との両方に基づいてオブジェクトをランク付けし得る。オーディオの合図ジェネレータ118は、ランクに基づいて1つまたは複数のオーディオの合図を選択すると決定し得る。たとえば、オーディオの合図ジェネレータ118は、最も高いランク付けのオーディオの合図、上位2つのランク付けの合図、上位3つのランク付けの合図、または何らかのその他の数のオーディオの合図を選択し得る。

【００９８】

場合によっては、オーディオの合図ジェネレータ118は、オーディオの合図がデジタル音声によってレンダリングされるテキストを邪魔することに基づいてオーディオの合図をフィルタリングして取り除くか、削除するか、またはオーディオの合図が基礎となるオーディオトラックに追加されることを防止することができる。たとえば、オーディオの合図ジェネレータ118は、画像認識技術によってデジタルコンポーネントオブジェクト内の複数のビジュアルオブジェクトを特定し得る。オーディオの合図ジェネレータ118は、メタデータ(たとえば、3Pデジタルコンテンツプロバイダデバイス160によって提供されたメタデータまたはデジタルコンポーネントオブジェクトにおけるリンクに対応するランディングページに関連するキーワード)およびテキストに基づいて、複数の口頭でないオーディオの合図を特定することができる。オーディオの合図ジェネレータ118は、ビジュアルオブジェクトの各々とメタデータとの間の一致のレベルを示すビジュアルオブジェクトの各々に関する一致スコアを決定することができる。オーディオの合図ジェネレータ118は、関連性、部分一致、フレーズ一致、または完全一致などの一致スコアを決定するための任意のマッチング技術を使用することができる。オーディオの合図ジェネレータ118は、一致スコアを決定するためにコンテンツセレクタコンポーネント108と同様の技術を使用することができる。オーディオの合図ジェネレータ118は、一致スコアに基づいて口頭でないオーディオの合図をランク付けすることができる。場合によっては、オーディオの合図ジェネレータ118は、1つまたは複数の最も高いランク付けのオーディオの合図を選択し得る。

【００９９】

場合によっては、オーディオの合図ジェネレータ118は、デジタル音声を使用して合成されるテキストを妨げない、まとめない、妨害しない、またはそのようなテキストにその他の方法で悪影響を与えない1つまたは複数の最も高いランク付けのオーディオの合図を選択し得る。たとえば、オーディオの合図ジェネレータ118は、口頭でないオーディオの合図の各々とテキストをレンダリングするために選択されたデジタル音声との間のオーディオの干渉のレベルを判定し得る。干渉のレベルは、たとえば、振幅、周波数、高さ、またはタイミングなどの1つまたは複数の要因を使用して判定され得る。説明のための例において、合成されたテキストと同じ周波数および振幅を有する効果音は、エンドユーザがレンダリングされたテキストを正しく知覚するのを妨げる高レベルの干渉を引き起こす可能性がある。別の例においては、大きな破壊の音が、テキストを邪魔し得る。しかし、口頭で伝えられるオーディオトラック全体を通じてより低い振幅のそよ風の音は、テキストを邪魔しない可能性がある。

【０１００】

干渉のレベルを決定するために、オーディオの合図ジェネレータ118は、合成されたテキストに対して口頭でないオーディオの合図によって引き起こされる干渉の量を決定し得

る。量は、テキストの割合、連続的な継続時間、またはテキストに対するデシベルレベルであることが可能である。場合によっては、干渉は、信号対雑音比または口頭でないオーディオの合図の信号に対するテキストの信号の比に基づくことが可能である。干渉のレベルは、等級(たとえば、低、中、もしくは高)または数値(たとえば、1から10までの尺度、もしくは尺度の一端が干渉がないことを表し、尺度の他端が完全な干渉を表す任意のその他の尺度による)を使用して示され得る。完全な干渉は、合成されたテキストを完全に打ち消す可能性がある相殺的干渉を指し得る。

【０１０１】

場合によっては、オーディオの合図ジェネレータ118は、合成されたテキストおよび口頭でないオーディオの合図に対応するオーディオ波形を組み合わせ、それから、組み合わされた信号を処理して、エンドユーザが合成されたテキストを知覚することができるかどうかを判定することによって干渉のレベルを決定し得る。オーディオの合図ジェネレータ118は、インターフェース104または自然言語プロセッサコンポーネント106と同様のオーディオ処理技術を使用して、合成されたテキストがデータ処理システム102自体によって正確に知覚され得るかどうかを確認、検証、または判定することができ、それは、エンドユーザがオーディオトラックを積極的にやはり知覚することができるかどうかを示し得る。

【０１０２】

予め決められた閾値未満である干渉のレベル(たとえば、低い干渉、5、6、7未満の干渉スコア、またはその他の測定基準)を有するオーディオの合図を特定すると、オーディオの合図ジェネレータ118は、閾値未満のオーディオの干渉のレベルを有する最も高いランク付けの口頭でないオーディオの合図を選択し得る。

【０１０３】

オーディオの合図ジェネレータ118は、選択された口頭でないオーディオの合図を基礎となるオーディオトラックと組み合わせてデジタルコンポーネントオブジェクトのオーディオトラックを生成することができる。オーディオトラックは、ビジュアルデジタルコンポーネントオブジェクトに基づくオーディオのみのデジタルコンポーネントオブジェクトに対応し得る。オーディオの合図ジェネレータ118は、口頭でない合図を基礎となるオーディオトラックと組み合わせるために任意のオーディオミキシング技術を使用することができる。たとえば、オーディオの合図ジェネレータ118は、基礎となるオーディオトラックに口頭でないオーディオの合図を重ねることができ、口頭でないオーディオの合図を背景オーディオとして追加することができ、口頭でないオーディオの合図を合成されたテキストの前もしくは後または合成されたテキストの間にちりばめることができる。データ処理システム102は、2つ以上の入力オーディオ信号の特性を組み合わせるか、ダイナミクスを変更するか、イコライズするか、またはそれ以外の方法で変更するように構成されたデジタルミキシングコンポーネントを含み得る。データ処理システム102は、口頭でないオーディオの合図および基礎となるオーディオトラックを受信し、2つの信号を合計して組み合わされたオーディオトラックを生成することができる。データ処理システム102は、デジタルミキシングプロセスを使用して入力オーディオ信号を組み合わせることができ、それによって、望ましくない雑音または歪みの混入を防止する。

【０１０４】

したがって、基礎となるオーディオフォーマットが合成されると、データ処理システム102は、口頭でないオーディオの合図またはメタデータから決定され得る伴奏トラックを挿入する第2の生成ステップを実行することができる。たとえば、ビジュアルデジタルコンポーネントが椰子の木のあるビーチリゾートのように見える場合、データ処理システム102は、波および木の葉を揺らす風のオーディオを合成し、合成されたオーディオをテキストトゥスピーチの基礎となるオーディオトラックに追加してオーディオトラックを生成することができる。

【０１０５】

場合によっては、アクションジェネレータ136が、基礎となるオーディオトラックまた

は口頭でない合図を有する生成されたオーディオトラックに予め決められたまたは固定のオーディオを追加し得る。たとえば、データ処理システム102は、ヒューリスティックなまたは規則に基づく技術を使用して、「私達のウェブサイトでこれについてもっと知ってください。」などの語句を追加すると決定し得る。これは、デジタルコンポーネントオブジェクトから独立したアクションをそのとき独立して実行するようにユーザに促すことができる。データ処理システム102は、履歴的な実行情報、オーディトラックの時間の長さに基づいて、または構成もしくは設定(たとえば、データ処理システム102の管理者によって設定されたデフォルト設定、もしくは3Pデジタルコンテンツプロバイダデバイス160によって提供され得る設定)に基づいて固定のオーディオを自動的に追加すると決定することができる。場合によっては、固定のオーディオは、「私達のウェブサイトでこれについてさらに知りたいですか。」などの音声入力の促しを含み得る。データ処理システム102は、促しに対応するデジタルアクションをそれから自動的に実行するために、この場合は「はい」などの、促しに対する返答の中のトリガワードを検出するようにオーディオデジタルコンポーネントオブジェクトを構成し得る。

【０１０６】

データ処理システム102は、コンピューティングデバイス140にスピーカ(たとえば、トランスデューサ144)を介してオーディオトラックを出力または提示させるためにコンピューティングデバイス140に生成されたオーディオトラックを提供することができる。場合によっては、データ処理システム102は、オーディオトラックに実行可能なコマンドを追加することができる。コンテンツ変換コンポーネント110は、オーディオトラックにトリガワードを追加するように設計され、構成され、動作可能なアクションジェネレータ136を含み得る。トリガワードは、オーディオトラックとのインタラクションを容易にし得る。プリプロセッサ142は、トリガワードをリスニングし、それから、トリガワードに応じてアクションを実行することができる。トリガワードは、オーディオトラックの再生中およびオーディオトラックの後の予め決められた量の時間(たとえば、1秒、2秒、5秒、10秒、15秒、またはその他の適切な時間の区間)などの所定の時間の区間の間アクティブなままである新しいウェークアップまたはホットワードになり得る。コンピューティングデバイス140のマイクロフォン(たとえば、センサー148)によって検出された入力オーディオ信号内のトリガワードの検出に応じて、データ処理システム102またはコンピューティングデバイス140は、トリガワードに対応するデジタルアクションを実行することができる。

【０１０７】

説明のための例において、オーディオトラックは、クルーズチケットを購入し、ビーチのある島で休暇を過ごす広告を含み得る。オーディオトラックは、「クルーズチケットの価格が知りたいですか。」などの促しを含み得る。トリガワードは、「はい」、または「価格はいくら。」、「いくらかかる。」、またはチケットの価格を要求するユーザによる意図を伝える何らかのその他の変化形であることが可能である。データ処理システム102は、プリプロセッサ142がユーザによって与えられる後続の音声入力内のトリガキーワードを検出し得るように、トリガワードをコンピューティングデバイス140またはコンピューティングデバイス140のプリプロセッサ142に提供することができる。音声入力内のトリガワードを検出することに応じて、プリプロセッサ142は、音声入力をデータ処理システム102に転送し得る。NLPコンポーネント106は、音声入力を解析し、ユーザをユーザに提示されたデジタルコンポーネントに関連するランディングページに導くこと、またはそうでなければ、要求された情報にアクセスし、提供することなどの、音声入力に対応するアクションを実行することができる。

【０１０８】

トリガキーワードは、様々なデジタルアクションにリンクされ得る。例示的なデジタルアクションは、情報を提供すること、アプリケーションを起動すること、ナビゲーションアプリケーションを起動すること、音楽もしくはビデオを再生すること、製品もしくはサービスを注文すること、電化製品を制御すること、照明デバイスを制御すること、モノの

インターネット対応デバイスを制御すること、レストランの食事を注文すること、予約をすること、相乗りを注文すること、映画のチケットを予約すること、航空券を予約すること、スマートテレビを制御すること、またはその他のデジタルアクションを含み得る。

【０１０９】

アクションジェネレータ136は、1つまたは複数の技術を使用してデジタルコンポーネントオブジェクトに関するアクションを選択することができる。アクションジェネレータ136は、ヒューリスティックな技術を使用して、予め決められたアクションの組からアクションを選択することができる。アクションジェネレータ136は、生成されたテキストを受信し、推測されたアクションを出力するように構成されたアクションモデル128を使用し得る。

【０１１０】

たとえば、アクションジェネレータ136は、デジタルコンポーネントオブジェクトのカテゴリを決定し得る。カテゴリは、自動車、銀行、スポーツ、衣料品などのバーティカルカテゴリを指し得る。アクションジェネレータ136は、カテゴリのために確立された1つまたは複数のトリガワードおよびデジタルアクションを取り出すためにカテゴリを用いて検索を実行するかまたはデータベースに問い合わせることができる。データベースは、トリガキーワードおよびデジタルアクションへのカテゴリのマッピングを含み得る。Table 2(表2)は、トリガワードおよびアクションへのカテゴリの例示的なマッピングを示す。

【０１１１】

【表２】

カテゴリ	トリガワード	デジタルアクション
相乗り	はい、乗車、乗車を注文する、～に行く	コンピューティングデバイス140上で相乗りアプリケーションを起動する、ユーザを乗せるための乗車を注文する
旅行	フライトを予約する、フライトの価格をチェックする、[都市]に行くのにいくらかかるか、[都市]への次のフライトはいつか	フライトの予約の選択肢を提供する、フライト予約アプリケーションを起動する、フライトを検索し、結果を提供する
乗り物の買い物	車はいくらか、車にどんなオプションが利用可能であるか、最も近い特約販売店はどこか	コンテンツアイテムのプロバイダのランディングページにアクセスする、要求された情報を提供する、最も近い自動車の特約販売店に案内するようにナビゲーションアプリケーションを起動する

Table 2: トリガワードおよびデジタルアクションへのカテゴリのマッピングの説明のための例

【０１１２】

Table 2(表2)は、トリガワードおよびデジタルアクションへのカテゴリの例示的なマッピングを示す。Table 2(表2)に示されるように、カテゴリは、相乗り、旅行、および乗り物の買い物を含み得る。アクションジェネレータ136は、デジタルコンポーネントオブジェクトのために生成されたテキスト、デジタルコンポーネントオブジェクトに関連するメタデータ、またはデジタルコンポーネントオブジェクトに埋め込まれたリンクに関連する解析データに基づいて、選択されたデジタルコンポーネントオブジェクトのカテゴリを決

定し得る。場合によっては、3Pデジタルコンテンツプロバイダデバイス160が、メタデータと一緒にカテゴリ情報を提供し得る。場合によっては、アクションジェネレータ136が、意味処理技術を使用して、デジタルコンポーネントオブジェクトに関連する情報に基づいてカテゴリを決定し得る。

【０１１３】

アクションジェネレータ136は、デジタルコンポーネントオブジェクトに関するカテゴリを特定または決定すると、検索を実行するかまたはマッピング(たとえば、Table 2(表2))に問い合わせることができる。アクションジェネレータ136は、マッピングから、カテゴリに関連する1つまたは複数のデジタルアクションに対応する1つまたは複数のトリガワードを取り出すことができる。たとえば、アクションジェネレータ136がカテゴリを「相乗り」と特定する場合、アクションジェネレータ136は、問い合わせまたは検索に応じて、トリガキーワード「はい」、「乗車」、「乗車を注文する」、または「～に行く」を取り出し得る。アクションジェネレータ136は、デジタルアクション、すなわち、コンピューティングデバイス140上で相乗りアプリケーションを起動すること、またはユーザを乗せるための乗車を注文することをさらに特定し得る。アクションジェネレータ136は、トリガキーワードのすべてを検出し、トリガキーワードの検出に応じて対応するデジタルアクションを実行する命令を用いてオーディオデジタルコンポーネントオブジェクトを構成し得る。

【０１１４】

場合によっては、アクションジェネレータ136は、取り出されたトリガキーワードおよびデジタルアクションのうちのすべてではない1つまたは複数をデジタルコンポーネントオブジェクトに追加すると決定し得る。たとえば、アクションジェネレータ136は、トリガキーワードの履歴的な実行に基づいて訓練されたデジタルアクションモデル128を使用して、デジタルコンポーネントオブジェクトのコンテキストおよびクライアントデバイスの種類に基づいてトリガワードをランク付けし得る。アクションジェネレータ136は、オーディオトラックに追加するために最も高いランク付けのトリガキーワードを選択し得る。

【０１１５】

アクションモデル128は、訓練データを使用して機械学習エンジン122によって訓練され得る。訓練データは、トリガキーワードに関連する履歴的な実行情報を含み得る。履歴的な実行情報は、各トリガキーワードに関して、トリガキーワードがインタラクションをもたらしたかどうか(たとえば、コンピューティングデバイス140がオーディオトラックの提示の後に受け取られた音声入力内にトリガキーワードを検出したか)、デジタルコンポーネントオブジェクトに関連するコンテキスト情報(たとえば、カテゴリ、キーワード、概念、またはインタラクションのフローにおける状態)、およびインタラクションのコンテキスト情報を含み得る。コンピューティングデバイス140のコンテキストは、たとえば、コンピューティングデバイス140の種類(たとえば、モバイルデバイス、ラップトップデバイス、スマートフォン、もしくはスマートスピーカ)、コンピューティングデバイス140の利用可能なインターフェース、運搬の仕方(たとえば、歩行、車、静止、自転車など)、またはコンピューティングデバイス140の場所を含み得る。たとえば、運搬の仕方が駆け足、自転車、または車である場合、データ処理システム102は、ビジュアルまたはタッチ入力を必要としない可能性があり、ユーザエクスペリエンスを向上させるためにオーディオ出力をもたらし得るインタラクションの種類を選択することができる。

【０１１６】

機械学習エンジン122は、アクションモデル128がデジタルコンポーネントオブジェクトおよびコンピューティングデバイス140のコンテキストに基づいてインタラクションをもたらす可能性が最も高いトリガキーワードを予測することができるように、この訓練データに基づいてアクションモデル128を訓練し得る。したがって、アクションジェネレータ136は、インタラクションをもたらす可能性が最も高いトリガキーワードを提供するためにリアルタイムのプロセスでデジタルコンポーネントオブジェクトに追加するアクションをカスタマイズし得るかまたは仕立て得る。トリガキーワードの数をアクションモデル

128によって決定されたようにインタラクションをもたらす可能性が最も高いトリガキーワードに制限することによって、アクションジェネレータ136は、コンピューティングデバイス140のユーザが望ましくないアクションの実行をうっかり引き起こす見込みを小さくしながら、プリプロセッサ142またはNLPコンポーネント106がトリガキーワードを正確におよび確実に検出する見込みを大きくすることができる。さらに、トリガワードの数を制限することによって、アクションジェネレータ136は、ネットワーク105を介して送信されるコマンドまたはデータパケットの数を減らし、プリプロセッサ142が処理するトリガワードの数を減らすことによってネットワーク帯域幅の通信およびコンピューティングリソースの利用を削減することができる。

【０１１７】

データ処理システム102は、コンピューティングデバイス140のスピーカを介して出力するためにコンピューティングデバイス140にデジタルコンポーネントオブジェクトのオーディオトラックを提供することができる。場合によっては、データ処理システム102は、デジタルコンポーネントオブジェクトのオーディオトラックに関する挿入点を決定することができる。挿入点は、コンピューティングデバイス140のオーディオ出力に関連する時点を指し得る。オーディオ出力は、デジタルストリーミング音楽、またはコンピューティングデバイス140上で実行されるアプリケーション152を介して提供されるその他のオーディオ(もしくはオーディオビジュアル)出力に対応しうる。データ処理システム102は、ユーザエクスペリエンスを改善し、生成されたオーディオトラックがエンドユーザによって知覚され、最終的にインタラクションを受け取る見込みを大きくしながら、コンピューティングデバイス140によって出力されている主オーディオコンテンツを不明瞭にすることまたは歪ませることを防止するために挿入地点を決定し得る。

【０１１８】

データ処理システム102は、オーディオトラックに関する挿入点を特定するように設計され、構築され、動作可能なコンテンツ挿入コンポーネント120を含み得る。コンテンツ挿入コンポーネント120は、コンピューティングデバイス140によって出力されるデジタルメディアストリーム内のオーディオトラックに関する挿入点を特定することができる。コンテンツ挿入コンポーネント120は、挿入モデル130を使用して挿入点を特定することができる。機械学習エンジン122は、履歴的な実行データを使用して挿入モデル130を訓練することができる。履歴的な実行データは、訓練データを含み得るかまたは訓練データと呼ばれ得る。挿入モデル130を訓練するために使用される履歴的な実行データは、デジタルメディアストリームに挿入されたオーディオトラックに関する履歴的な挿入点についてのデータを含み得る。データは、オーディオトラックがいつ挿入されたか、オーディオトラックについてのコンテキスト情報、デジタルメディアストリームについてのコンテキスト情報、ユーザがオーディオトラックとインタラクションしたかどうか、ユーザがオーディオトラックとどのようにインタラクションしたか(たとえば、ユーザがどのアクションを行ったか、またはインタラクションが肯定的であったのかもしくは否定的であったのか)、あるいはコンピューティングデバイス140についてのコンテキスト情報(たとえば、コンピューティングデバイスの種類、コンピューティングデバイスの利用可能なインターフェース、またはコンピューティングデバイスの場所)を示し得る。

【０１１９】

機械学習エンジン122は、この訓練データを使用して挿入モデル130を訓練することができる。機械学習エンジン122は、デジタルコンテンツストリーム(たとえば、ストリーミング音楽、ニュース、ポッドキャスト、ビデオ、またはその他のメディア)にビジュアルデジタルコンポーネントオブジェクトに基づいて生成されたオーディオトラックをいつ挿入すべきかを予測するために挿入モデル130が使用され得るように挿入モデル130を訓練するために任意の技術を使用することができる。

【０１２０】

コンテンツ挿入コンポーネント120は、履歴的な実行データを使用して訓練された挿入モデル130に基づいて挿入点を特定することができる。挿入点は、たとえば、現在のスト

リーミングメディアのセグメントの後および次のセグメントの始まりの前であることが可能である。各セグメントは、別の歌に対応し得る。ポッドキャストなどの別の例において、コンテンツ挿入コンポーネント120は、挿入モデル130を使用して、セグメント中にオーディオトラックを挿入すると決定し得る。コンテンツ挿入コンポーネント120は、セグメントが開始した後およびセグメントが終了する前にオーディオトラックを挿入し得る。たとえば、セグメントは、30分の継続時間を持つことが可能であり、コンテンツ挿入コンポーネント120は、挿入モデル130を使用して、セグメントを15分再生した後にオーディオトラックを挿入すると決定し得る。

【０１２１】

コンテンツ挿入コンポーネント120は、現在のコンテキスト(たとえば、生成されたオーディオトラック、ストリーミングメディア、およびコンピューティングデバイス140のコンテキスト)に基づいてカスタムの挿入点を決定し得る。コンテンツ挿入コンポーネント120は、カスタムの挿入点をリアルタイムで決定し得る。コンテンツ挿入コンポーネント120は、第1のコンピューティングデバイスおよび第2のコンピューティングデバイスが異なる種類のコンピューティングデバイスである場合(たとえば、ラップトップ対スマートスピーカ)、第2のコンピューティングデバイス140に比べて第1のコンピューティングデバイス140に関して異なる挿入点を決定し得る。コンテンツ挿入コンポーネント120は、異なるコンピューティングデバイスに関連する運搬の仕方(たとえば、歩行対車対静止)に基づいて第1のコンピューティングデバイス140および第2のコンピューティングデバイス140に関して異なる挿入点を決定し得る。

【０１２２】

コンテンツ挿入コンポーネント120は、デジタルメディアストリーム内のキーワード、語、または概念に近接してオーディオトラックを挿入すると決定し得る。コンテンツ挿入コンポーネント120は、デジタルメディアストリームを監視してオーディオトラックに関連性があるデジタルメディアストリーム内のトリガワードを検出し、それから、デジタルメディアストリーム内で検出されたトリガワードの後にまたはそのようなトリガワードに応じてオーディオのみのデジタルコンポーネントオブジェクトを挿入すると決定し得る。

【０１２３】

コンテンツ挿入コンポーネント120は、3P電子リソースサーバ162からデジタルメディアストリームのセグメントのコピーを取得することができる。コンテンツ挿入コンポーネント120は、デジタルメディアストリームのセグメントを解析してセグメント内のすべてのトークン(たとえば、キーワード、話題、または概念)および文を特定することができる。コンテンツ挿入コンポーネント120は、トークンまたは文がデジタルコンポーネントオブジェクトにどれだけ関連性があるかを決定するために各トークンおよび文に関する関連性スコアを決定し得る。コンテンツ挿入コンポーネント120は、最も高い関連性スコアを有するトークンを選択し、それから、選択されたトークンの近くに(たとえば、トークンが提示される前または提示された後に)挿入するためにオーディオのみのデジタルコンポーネントオブジェクトを提供することができる。

【０１２４】

場合によっては、コンテンツ挿入コンポーネント120は、デジタルメディアのセグメント内のすべてのトークンを特定し、オーディオトラックが各トークンの近くに挿入されるモンテカルロシミュレーションを実行することができる。コンテンツ挿入コンポーネント120は、様々な挿入点をニューラルネットワークエンジンに入力してどの挿入点が最良でありそうかを決定することができる。ニューラルネットワークは、機械学習技術を使用して、デジタルメディアストリームに挿入された人間によって格付けされたオーディオトラックを含む訓練データに基づいて訓練され得る。たとえば、コンテンツ挿入コンポーネント120は、挿入モデル130を使用して挿入点を決定することができる。訓練データは、挿入点にオーディオトラックを有するデジタルメディアストリームを格付けする人間の格付け人を含み得る。格付けは、良いもしくは悪いなどの2値であることが可能であり、またはある尺度のスコアであることが可能である(たとえば、たとえば、10が最良の音を鳴ら

すトラックを示し、0が最悪の音を鳴らすトラックを示すようにして0から10まで)。

【０１２５】

場合によっては、コンテンツ挿入コンポーネント120は、ヒューリスティックな技術を使用して、生成されたオーディオトラックに関する挿入点を決定することができる。ヒューリスティックな技術は、デジタルメディアストリームの種類に基づいて異なり得る。デジタルメディアストリームのコンテンツが歌である場合、ヒューリスティックな規則は、生成されたオーディオトラックを歌が再生を終えた後に挿入することであることが可能である。デジタルメディアストリームのコンテンツがポッドキャストである場合、ヒューリスティックな規則は、関連性のあるトークンを含む文の後にオーディオトラックを挿入することであることが可能である。

【０１２６】

挿入点を選択すると、データ処理システム102は、コンピューティングデバイス140にデジタルメディアストリーム内の挿入点においてオーディオトラックをレンダリングさせるためにコンピューティングデバイス140に命令を与えることができる。

【０１２７】

図2は、実装によるオーディオトラックを生成する方法の図である。方法200は、たとえば、データ処理システム、インターフェース、コンテンツセレクタコンポーネント、自然言語プロセッサコンポーネント、コンテンツ変換コンポーネント、またはコンピューティングデバイスを含む、図1または図3に示される1つまたは複数のシステム、コンポーネント、またはモジュールによって実行され得る。判断ブロック202において、データ処理システムは、入力信号が受信されたかどうかを判定し得る。入力信号は、データ処理システムの遠隔にあるコンピューティングデバイスによって検出された音声入力に対応し得る。入力信号は、コンピューティングデバイスのマイクロフォンによって検出された音声入力などのオーディオ入力信号を運ぶデータパケットを含み得る。データ処理システムは、データ処理システムのインターフェースを介して入力信号を受信し得る。データ処理システムは、ネットワークを介してコンピューティングデバイスから入力信号を受信し得る。

【０１２８】

判断ブロック202において、入力信号が受信されたとデータ処理システムが判定する場合、データ処理システムは、入力信号を解析し、要求を検出するためにACT 204に進むことができる。データ処理システムは、自然言語処理技術を使用して入力信号を解析し、入力信号内の1つまたは複数のキーワード、語、概念、語句、またはその他の情報を検出することができる。

【０１２９】

データ処理システムは、コンテンツを選択すべきかどうかを決定するために判断ブロック206に進むことができる。コンテンツを選択することは、リアルタイムコンテンツ選択プロセスを実行してサードパーティのデジタルコンポーネントプロバイダによって提供されたデジタルコンポーネントオブジェクトを選択することを指し得る。コンテンツを選択することは、サードパーティのデジタルコンポーネントプロバイダによって与えられたコンテンツ選択基準を使用してリアルタイムオンラインオークションを実行することを指し得る。

【０１３０】

データ処理システムがACT 204において入力信号内にコンテンツの要求を検出する場合、データ処理システムは、判断ブロック206において、コンテンツを選択すると決定し得る。データ処理システムが、判断ブロック202において、入力信号が受信されなかったと判定する場合、データ処理システムは、判断ブロック206においてコンテンツを選択するとやはり判断し得る。たとえば、データ処理システムは、コンピューティングデバイスからデジタルコンポーネントオブジェクトの明示的な要求を受信することなしに、先を見越して、オンラインコンテンツ選択プロセスを実行し、デジタルコンポーネントオブジェクトをコンピューティングデバイスにプッシュし得る。データ処理システムは、コンピューティングデバイスによって出力されるデジタル音楽ストリーム内の(たとえば、メディアセ

グメントまたは歌の間の)提示機会を特定し、この機会にデジタルコンポーネントオブジェクトを提供すると自動的に決定し得る。したがって、場合によっては、データ処理システムは、コンテンツの要求を受信し、それから、コンテンツ選択プロセスを実行することが可能であり、一方、その他の場合、データ処理システムは、コンテンツを受信しないが、先を見越して、コンテンツ選択プロセスを実行すると決定する可能性がある。データ処理システムがコンテンツの要求を受信する場合、要求は、主コンテンツ(たとえば、入力オーディオ信号内のクエリに応じる検索結果)を求めるものであることが可能であり、データ処理システムは、要求に応じるが、入力クエリに直接応じるオーガニック検索結果とは異なり得る補足コンテンツ(たとえば、広告に対応するデジタルコンポーネントオブジェクト)を選択するためにオンラインオークションを実行することが可能である。

【０１３１】

判断ブロック206において、データ処理システムが3Pデジタルコンポーネントプロバイダのデジタルコンポーネントオブジェクトを選択するためにコンテンツ選択プロセスを実行しないと決定する場合、データ処理システムは、入力信号が受信されたかどうかを判定するために判断ブロック202に戻り得る。しかし、データ処理システムが、判断ブロック206において、コンテンツを選択すると決定する場合、データ処理システムは、コンテンツ選択プロセスを実行してコンテンツを選択するためにACT 208に進むことができる。データ処理システムは(たとえば、コンテンツセレクタコンポーネントによって)、入力された要求、コンピューティングデバイス、またはデジタルストリーミングコンテンツに関連するコンテンツ選択基準またはその他のコンテキスト情報を使用してデジタルコンポーネントオブジェクトを選択することができる。

【０１３２】

データ処理システムは、コンピューティングデバイスのディスプレイデバイスを介して表示するために構成されるビジュアルのみのデジタルコンポーネントオブジェクト、コンピューティングデバイスのスピーカを介して再生するために構成されるオーディオのみのデジタルコンポーネントオブジェクト、またはコンピューティングデバイスのディスプレイとスピーカとの両方を介して出力するために構成されるオーディオビジュアルデジタルコンポーネントなどの、フォーマットを有するデジタルコンポーネントオブジェクトを選択することができる。

【０１３３】

判断ブロック210において、データ処理システムは、選択されたデジタルコンポーネントを異なるフォーマットに変換すべきかどうかを決定し得る。たとえば、選択されたデジタルコンポーネントオブジェクトがビジュアルのみのフォーマットである場合、データ処理システムは、コンピューティングデバイスのディスプレイデバイスを介して提示するためにデジタルコンポーネントオブジェクトをコンピューティングデバイスにビジュアルフォーマットで提供すべきか、またはデジタルコンポーネントオブジェクトをスピーカなどのコンピューティングデバイスの異なる出力インターフェースを介して提示するために異なるフォーマットに変換すべきかを決定し得る。

【０１３４】

データ処理システム(たとえば、フォーマットセレクタ)は、デジタルコンポーネントオブジェクトを変換すべきかどうかを決定し得る。データ処理システムは、コンピューティングデバイスの利用可能なインターフェース、コンピューティングデバイスの主なインターフェース、コンピューティングデバイスのコンテキスト(たとえば、運搬の仕方)、コンピューティングデバイスの種類、またはその他の要因に基づいて決定を行い得る。判断ブロック210において、データ処理システムが選択されたデジタルコンポーネントオブジェクトを異なるフォーマットに変換しないと決定する場合、データ処理システムは、ACT 212に進み、選択されたデジタルコンポーネントオブジェクトをその元のフォーマットでコンピューティングデバイスに送信することができる。

【０１３５】

しかし、データ処理システムが、判断ブロック210において、デジタルコンポーネント

オブジェクトを異なるフォーマットに変換すると決定する場合、データ処理システムは、テキストを生成するためにACT 214に進むことができる。たとえば、元のフォーマットがビジュアルのみのフォーマットであり、データ処理システムがデジタルコンポーネントをオーディオのみのフォーマットに変換すると決定する場合、データ処理システムは、ビジュアルデジタルコンポーネントオブジェクトに関するテキストを生成するためにACT 214に進むことができる。(たとえば、テキストジェネレータを介して)データ処理システムは、自然言語生成技術を使用して、ビジュアルデジタルコンポーネントオブジェクトに基づいてテキストを生成することができる。データ処理システムは、デジタルコンポーネントオブジェクトのビジュアルコンテンツのみに基づいてテキストを生成し得る。ビジュアルコンテンツは、画像を指し得る。場合によっては、データ処理システムは、ビジュアルデジタルコンポーネントオブジェクトに関連するメタデータに基づいてテキストを生成し得る。

【０１３６】

ACT 216において、データ処理システムは、テキストを合成またはレンダリングするためのデジタル音声を選択し得る。データ処理システムは、選択されたデジタル音声を使用して、生成されたテキストのテキストトゥスピーチ変換を実行することができる。データ処理システムは、生成されたテキスト、デジタルコンポーネントオブジェクトに関連するコンテキスト情報(たとえば、キーワード、話題、概念、バーティカルカテゴリ)、メタデータ、コンピューティングデバイスに関連するコンテキスト情報に基づいてデジタル音声を選択することができる。データ処理システムは、機械学習技術および履歴的なデータに基づいて訓練されたモデルを使用して、生成されたテキストを合成するために使用するデジタル音声を選択することができる。たとえば、データ処理システムは、デジタルコンポーネントオブジェクトに関連するコンテキスト情報(たとえば、メタデータ)をモデルに入力することができ、モデルは、音声の特性のベクトルを出力することができる。音声の特性のベクトルは、性別、スピーチの速さ、イントネーション、音の大きさ、またはその他の特性を示し得る。

【０１３７】

データ処理システムは、音声の特性のベクトルに一致するデジタル音声を選択することができる。データ処理システムは、選択されたデジタル音声を使用して基礎となるオーディオトラックを構築することができる。データ処理システムは、音声の特性のベクトルによって示されるように基礎となるオーディオトラックを構築することができる。たとえば、デジタル音声は、性別などの固定の特性と、スピーチの速さ、イントネーション、または音の大きさなどの動的な特性とを含み得る。動的な特性は、音節毎に変わり得る。データ処理システムは、音節毎に音声の特性のベクトルに対応する固定のおよび動的な特性を使用してテキストを合成するように構成されたテキストトゥスピーチエンジンを使用し得る。

【０１３８】

ACT 218において、データ処理システムは、口頭でないオーディオの合図を生成し得る。データ処理システムは、口頭でないオーディオの合図をACT 216において生成された基礎となるオーディオトラックと組み合わせることができる。口頭でないオーディオの合図を生成するために、データ処理システムは(たとえば、オーディオの合図ジェネレータによって)、ビジュアルデジタルコンポーネント内のオブジェクトを特定することができる。データ処理システムは、ビジュアルコンポーネントのみを特定し得る。データ処理システムは、ビジュアルコンポーネントとテキストコンポーネント(たとえば、デジタルコンポーネントオブジェクトに関連するメタデータ)との両方を特定し得る。オブジェクトを特定すると、データ処理システムは、オブジェクトを特定するかまたは示すオーディオの合図を特定することができる。たとえば、データ処理システムが海の波および椰子の木を特定する場合、データ処理システムは、波の音および葉を吹き抜けるそよ風の音を選択し得る。

【０１３９】

データ処理システムは、任意のオーディオミキシング技術を使用して、選択されたオー

ディオの合図を基礎となるオーディオトラックと組み合わせることができる。データ処理システムは、基礎となるオーディオトラックの一部にまたはオーディオトラック全体に口頭でないオーディオの合図を追加し得る。データ処理システムは、基礎となるオーディオトラック内の口頭で伝えられるテキストを歪ませないまたは不明瞭にしないようにして基礎となるトラックに口頭でないオーディオの合図を追加し、それによってユーザエクスペリエンスを向上させ得る。場合によっては、データ処理システムは、組み合わされたオーディオトラックをシミュレーションし、品質をテストし得る。たとえば、データ処理システムは、組み合わされたオーディオトラックを受信することと、組み合わされたオーディオトラックに対して自然言語処理を実行することとをシミュレーションし得る。データ処理システムは、解析されたテキストをデータ処理システムのテキストジェネレータによって生成されたテキストと比較することによって、データ処理システムのNLPコンポーネントが口頭で伝えられるテキストを正確に検出することができたかどうかを確認することができる。データ処理システムが組み合わされたオーディオトラック内のテキストを正確に解読することができない場合、データ処理システムは、口頭でないオーディオの合図が口頭で伝えられるテキストに悪影響を与え、エンドユーザが口頭で伝えられるテキストを正確に特定することを妨げると判定し得る。したがって、ユーザエクスペリエンスを向上させようとして、データ処理システムは、口頭でないオーディオの合図のうちの1つまたは複数を取り除き、それから、組み合わされたオーディオトラックを再生成し、再テストすることが可能である。データ処理システムは、データ処理システムが口頭で伝えられるテキストを正確に解釈することができるまで口頭でないオーディオの合図のこの削除ならびに再生成および再テストを実行し得る。組み合わされたオーディオトラック内の口頭で伝えられるテキストが知覚可能であると判定することに応じて、データ処理システムは、提示するために、組み合わされたオーディオトラックを承認し得る。

【０１４０】

場合によっては、データ処理システムは、組み合わされたオーディオトラックを提示のためにコンピューティングデバイスに送信し得る。場合によっては、データ処理システムは、オーディオトラックに関する挿入点を決定するためにACT 220に進むことができる。データ処理システムは、機械学習技術および履歴的なデータによって訓練された挿入モデルを使用して、コンピューティングデバイスによって出力されているデジタルメディアストリームにオーディオトラックを挿入すべきかどうかを決定することができる。データ処理システムは、コンピューティングリソースの利用、ネットワーク帯域幅の消費を削減するか、デジタルメディアストリームのレイテンシもしくは遅延を防止するか、またはユーザエクスペリエンスを向上させる挿入点を決定することができる。たとえば、データ処理システムは、デジタルメディアストリームのセグメントの始めに、デジタルメディアストリームのセグメントの間に、またはデジタルメディアストリームのセグメントの後にオーディオトラックを挿入すると決定し得る。

【０１４１】

挿入点を決定すると、データ処理システムは、ACT 222に進み、コンピューティングデバイスに変換されたデジタルコンポーネントをレンダリングさせるか、再生されるか、またはそれ以外の方法で提示させるためにコンピューティングデバイスに変換されたコンテンツ(またはオーディオのみのデジタルコンポーネントオブジェクト)を提供することができる。場合によっては、データ処理システムは、デジタルアクションを呼び出すか、開始するか、または実行するように変換されたデジタルコンポーネントを構成し得る。たとえば、データ処理システムは、ユーザから入力された後続の音声内のトリガワードを検出し、それから、トリガワードに応じてデジタルアクションを実行するようにコンピューティングデバイスまたはデータ処理システムを構成するための命令を与えることができる。

【０１４２】

場合によっては、データ処理システムは、コンテンツの要求を受信しない可能性がある。たとえば、データ処理システムは、コンピューティングデバイスによってレンダリングされるデジタルストリーミングコンテンツに関連するキーワードを先を見越して特定する

ことができる。データ処理システムは、判断ブロック206において、キーワードに応じてコンテンツを選択すると決定し得る。そのとき、データ処理システムは、キーワードに基づいて、ビジュアル出力フォーマットを有するデジタルコンポーネントオブジェクトを選択し得る。データ処理システムは、コンピューティングデバイスの種類に基づいて、デジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換すると決定し得る。データ処理システムは、デジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換する決定に応じて、デジタルコンポーネントオブジェクトに関するテキストを生成し得る。データ処理システムは、デジタルコンポーネントオブジェクトのコンテキストに基づいて、テキストをレンダリングするためのデジタル音声を選択し得る。データ処理システムは、デジタル音声によってレンダリングされたテキストを用いてデジタルコンポーネントオブジェクトの基礎となるオーディオトラックを構築し得る。データ処理システムは、デジタルコンポーネントオブジェクトに基づいて口頭でないオーディオの合図を生成し得る。データ処理システムは、デジタルコンポーネントオブジェクトのオーディオトラックを生成するために口頭でないオーディオの合図をデジタルコンポーネントオブジェクトの基礎となるオーディオの形態と組み合わせ得る。データ処理システムは、コンピューティングデバイスのスピーカを介して出力するためにコンピューティングデバイスにデジタルコンポーネントオブジェクトのオーディオトラックを提供し得る。

【０１４３】

図3は、例示的なコンピュータシステム300のブロック図である。コンピュータシステムまたはコンピューティングデバイス300は、システム100、またはデータ処理システム102などのそのシステム100のコンポーネントを含み得るかまたはそれらを実装するために使用され得る。コンピューティングシステム300は、情報を伝達するためのバス305またはその他の通信コンポーネントと、情報を処理するためのバス305に結合されたプロセッサ310または処理回路とを含む。また、コンピューティングシステム300は、情報を処理するためのバスに結合された1つまたは複数のプロセッサ310または処理回路を含み得る。コンピューティングシステム300は、情報およびプロセッサ310によって実行される命令を記憶するためのバス305に結合されたランダムアクセスメモリ(RAM)またはその他のダイナミックストレージデバイスなどのメインメモリ315も含む。メインメモリ315は、データリポジトリであることが可能であるかまたはデータリポジトリを含むことが可能である。メインメモリ315は、位置情報、一時的な変数、またはプロセッサ310による命令の実行中のその他の中間情報を記憶するためにも使用され得る。コンピューティングシステム300は、静的な情報およびプロセッサ310のための命令を記憶するためのバス305に結合された読み出し専用メモリ(ROM)320またはその他のスタティックストレージデバイスをさらに含む可能性がある。ソリッドステートデバイス、磁気ディスク、または光ディスクなどのストレージデバイス325が、情報および命令を永続的に記憶するためにバス305に結合され得る。ストレージデバイス325は、データリポジトリを含み得るかまたはデータリポジトリの一部であり得る。

【０１４４】

コンピューティングシステム300は、ユーザに対して情報を表示するための液晶ディスプレイまたはアクティブマトリックスディスプレイなどのディスプレイ335にバス305を介して結合される可能性がある。英数字およびその他のキーを含むキーボードなどの入力デバイス330が、プロセッサ310に情報およびコマンド選択を伝達するためにバス305に結合される可能性がある。入力デバイス330は、タッチスクリーンディスプレイ335を含み得る。入力デバイス330は、プロセッサ310に方向情報およびコマンド選択を伝達するためおよびディスプレイ335上でカーソルの動きを制御するためのマウス、トラックボール、またはカーソル方向キーなどのカーソルコントロールも含み得る。ディスプレイ335は、たとえば、図1のデータ処理システム102、クライアントコンピューティングデバイス140、またはその他のコンポーネントの一部であることが可能である。

【０１４５】

本明細書において説明されるプロセス、システム、および方法は、メインメモリ315に

含まれる命令の配列をプロセッサ310が実行することに応じてコンピューティングシステム300によって実施され得る。そのような命令は、ストレージデバイス325などの別のコンピュータ可読媒体からメインメモリ315に読み込まれ得る。メインメモリ315に含まれる命令の配列の実行は、コンピューティングシステム300に本明細書において説明される例示的なプロセスを実行させる。マルチプロセッシング配列の1つまたは複数のプロセッサも、メインメモリ315に含まれる命令を実行するために使用される可能性がある。配線による回路が、本明細書において説明されるシステムおよび方法と一緒にソフトウェア命令の代わりにまたはソフトウェア命令と組み合わせて使用され得る。本明細書において説明されるシステムおよび方法は、ハードウェア回路とソフトウェアとのいかなる特定の組合せにも限定されない。

【０１４６】

例示的なコンピューティングシステムが図3に示されたが、本明細書に記載の動作を含む対象は、本明細書において開示される構造およびそれらの構造的均等物を含む、その他の種類のデジタル電子回路、またはコンピュータソフトウェア、ファームウェア、もしくはハードウェア、またはそれらのうちの1つもしくは複数の組合せで実装され得る。

【０１４７】

本明細書において説明されるシステムがユーザもしくはユーザデバイスにインストールされたアプリケーション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状況において、ユーザは、プログラムまたは特徴がユーザ情報(たとえば、ユーザのソーシャルネットワーク、ソーシャルな行為もしくは活動、職業、ユーザの好み、またはユーザの現在位置についての情報)を収集するかどうかを制御する機会を与えられる。加えて、または別法において、特定のデータが、個人情報が削除されるように、そのデータが記憶されるかまたは使用される前に1つまたは複数の方法で処理され得る。

【０１４８】

本明細書に記載の対象および動作は、本明細書において開示される構造およびそれらの構造的均等物を含むデジタル電子回路、またはコンピュータソフトウェア、ファームウェア、もしくはハードウェア、またはそれらのうちの1つもしくは複数の組合せで実装され得る。本明細書に記載の対象は、1つまたは複数のコンピュータプログラム、たとえば、データ処理装置による実行のために、またはデータ処理装置の動作を制御するために1つまたは複数のコンピュータストレージ媒体上に符号化されたコンピュータプログラム命令の1つまたは複数の回路として実装され得る。代替的にまたは追加的に、プログラム命令は、データ処理装置による実行のために好適な受信機装置に送信するために情報を符号化するように生成される人為的に生成される伝播信号、たとえば、機械によって生成される電気的信号、光学的信号、または電磁的信号上に符号化され得る。コンピュータストレージ媒体は、コンピュータ可読ストレージデバイス、コンピュータ可読ストレージ基板、ランダムもしくはシリアルアクセスメモリアレーもしくはデバイス、またはそれらのうちの1つもしくは複数の組合せであることが可能であり、あるいはそれらに含まれることが可能である。コンピュータストレージ媒体は、伝播信号ではないが、人為的に生成された伝播信号に符号化されたコンピュータプログラム命令の送信元または送信先であることが可能である。また、コンピュータストレージ媒体は、1つもしくは複数の別個のコンポーネントもしくは媒体(たとえば、複数のCD、ディスク、もしくはその他のストレージデバイス)であることが可能であり、またはそれらに含まれることが可能である。本明細書に記載の動作は、1つもしくは複数のコンピュータ可読ストレージデバイスに記憶された、またはその他のソースから受信されたデータに対してデータ処理装置によって実行される動作として実装され得る。

【０１４９】

用語「データ処理システム」、「コンピューティングデバイス」、「コンポーネント」、または「データ処理装置」は、例として、1つのプログラミング可能なプロセッサ、1台のコンピュータ、1つのシステムオンチップ、またはそれらの複数もしくは組合せを含む、データを処理するための様々な装置、デバイス、および機械を包含する。装置は、専用

の論理回路、たとえば、FPGA(フィールドプログラマブルゲートアレー)またはASIC(特定用途向け集積回路)を含み得る。装置は、ハードウェアに加えて、問題にしているコンピュータプログラムのための実行環境を作成するコード、たとえば、プロセッサのファームウェア、プロトコルスタック、データベース管理システム、オペレーティングシステム、クロスプラットフォームランタイム環境、仮想マシン、またはそれらのうちの1つもしくは複数の組合せを構成するコードも含み得る。装置および実行環境は、ウェブサービスインフラストラクチャ、分散コンピューティングインフラストラクチャ、およびグリッドコンピューティングインフラストラクチャなどの様々な異なるコンピューティングモデルインフラストラクチャを実現することができる。自然言語プロセッサコンポーネント106およびその他のデータ処理システム102またはデータ処理システム102のコンポーネントは、1つまたは複数のデータ処理装置、システム、コンピューティングデバイス、またはプロセッサを含み得るかまたは共有し得る。たとえば、コンテンツ変換コンポーネント110およびコンテンツセレクタコンポーネント108は、1つまたは複数のデータ処理装置、システム、コンピューティングデバイス、またはプロセッサを含み得るかまたは共有し得る。

【０１５０】

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、アプリ、スクリプト、またはコードとしても知られる)は、コンパイラ型言語もしくはインタープリタ型言語、宣言型言語もしくは手続き型言語を含む任意の形態のプログラミング言語で記述可能であり、独立型プログラムとしての形態、またはモジュール、コンポーネント、サブルーチン、オブジェクト、もしくはコンピューティング環境での使用に好適なその他の単位としての形態を含む任意の形態で展開され得る。コンピュータプログラムは、ファイルシステム内のファイルに対応し得る。コンピュータプログラムは、その他のプログラムもしくはデータを保持するファイルの一部(たとえば、マークアップ言語のドキュメントに記憶された1つもしくは複数のスクリプト)、問題にしているプログラムに専用の単一のファイル、または複数の連携されたファイル(たとえば、1つもしくは複数のモジュール、サブプログラム、もしくはコードの一部を記憶するファイル)に記憶され得る。コンピュータプログラムは、1つのコンピュータ上で、または1つの場所に置かれるか、もしくは複数の場所に分散され、通信ネットワークによって相互に接続される複数のコンピュータ上で実行されるように展開され得る。

【０１５１】

本明細書に記載のプロセスおよび論理フローは、入力データに対して演算を行い、出力を生成することによってアクションを行うために1つまたは複数のコンピュータプログラム(たとえば、データ処理システム102のコンポーネント)を1つまたは複数のプログラミング可能なプロセッサが実行することによって実行され得る。また、プロセスおよび論理フローは、専用の論理回路、たとえば、FPGA(フィールドプログラマブルゲートアレー)またはASIC(特定用途向け集積回路)によって実行されることが可能であり、さらに、装置は、それらの専用の論理回路として実装されることが可能である。コンピュータプログラム命令およびデータを記憶するのに適したデバイスは、例として、半導体メモリデバイス、たとえば、EPROM、EEPROM、およびフラッシュメモリデバイス、磁気ディスク、たとえば、内蔵ハードディスクまたはリムーバブルディスク、光磁気ディスク、ならびにCD ROMディスクおよびDVD-ROMディスクを含むすべての形態の不揮発性メモリ、媒体、およびメモリデバイスを含む。プロセッサおよびメモリは、専用論理回路によって補完され得るか、または専用論理回路に組み込まれ得る。

【０１５２】

本明細書に記載の対象は、バックエンドコンポーネントを、たとえば、データサーバとして含むか、またはミドルウェアコンポーネント、たとえば、アプリケーションサーバを含むか、またはフロントエンドコンポーネント、たとえば、ユーザが本明細書に記載の対象の実装とインタラクションすることができる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か、または1つもしくは複数のそのようなバックエンドコンポーネント、ミドルウェアコンポーネント、もしくはフ

ロントエンドコンポーネントの組合せを含むコンピューティングシステムに実装され得る。システムのコンポーネントは、任意の形態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相互に接続されることが可能である。通信ネットワークの例は、ローカルエリアネットワーク(「LAN」)および広域ネットワーク(「WAN」)、インターネットワーク(たとえば、インターネット)、ならびにピアツーピアネットワーク(たとえば、アドホックピアツーピアネットワーク)を含む。

【０１５３】

システム100またはシステム300などの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み得る。クライアントおよびサーバは、概して互いに離れており、通常は通信ネットワーク(たとえば、ネットワーク105)を通じてインタラクションする。クライアントとサーバとの関係は、それぞれのコンピュータ上で実行されており、互いにクライアント-サーバの関係にあるコンピュータプログラムによって生じる。一部の実装において、サーバは、(たとえば、クライアントデバイスとインタラクションするユーザに対してデータを表示し、そのようなユーザからユーザ入力を受け取る目的で)クライアントデバイスにデータ(たとえば、デジタルコンポーネントを表すデータパケット)を送信する。クライアントデバイスにおいて生成されたデータ(たとえば、ユーザインタラクションの結果)は、サーバにおいてクライアントデバイスから受信され得る(たとえば、データ処理システム102のインターフェース104によって受信され得る)。

【０１５４】

動作が特定の順序で図面に示されているが、そのような動作は、示された特定の順序でまたは逐次的順序で実行される必要があるわけではなく、すべての示された動作が、実行される必要があるわけではない。本明細書に記載のアクションは、異なる順序で実行され得る。

【０１５５】

様々なシステムコンポーネントの分割は、すべての実装において分割を必要とするわけではなく、説明されたプログラムコンポーネントは、単一のハードウェアまたはソフトウェア製品に含まれ得る。たとえば、コンテンツ変換コンポーネント110およびコンテンツ挿入コンポーネント120は、単一のコンポーネント、アプリ、またはプログラム、または1つもしくは複数の処理回路を有する論理デバイスであることが可能であり、あるいはデータ処理システム102の1つまたは複数のプロセッサによって実行されることが可能である。

【０１５６】

この技術的な解決策の少なくとも1つの態様は、オーディオトラックを生成するためのシステムを対象とする。システムは、データ処理システムを含み得る。データ処理システムは、1つまたは複数のプロセッサを含み得る。データ処理システムは、ネットワークを介して、データ処理システムの遠隔のコンピューティングデバイスのマイクロフォンによって検出された入力オーディオ信号を含むデータパケットを受信し得る。データ処理システムは、要求を特定するために入力オーディオ信号を解析し得る。データ処理システムは、要求に基づいて、ビジュアル出力フォーマットを有するデジタルコンポーネントオブジェクトを選択することが可能であり、デジタルコンポーネントオブジェクトは、メタデータに関連付けられる。データ処理システムは、コンピューティングデバイスの種類に基づいて、デジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換すると決定し得る。データ処理システムは、デジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換する決定に応じて、デジタルコンポーネントオブジェクトに関するテキストを生成し得る。データ処理システムは、デジタルコンポーネントオブジェクトのコンテキストに基づいて、テキストをレンダリングするためのデジタル音声を選択し得る。データ処理システムは、デジタル音声によってレンダリングされたテキストを用いてデジタルコンポーネントオブジェクトの基礎となるオーディオトラックを構築し得る。データ処理システムは、デジタルコンポーネントオブジェクトのメタデータに基づいて、口頭でないオーディオの合図を生成し得る。データ処理システムは、デジタルコンポーネン

トオブジェクトのオーディオトラックを生成するために口頭でないオーディオの合図をデジタルコンポーネントオブジェクトの基礎となるオーディオの形態と組み合わせ得る。データ処理システムは、コンピューティングデバイスからの要求に応じて、コンピューティングデバイスのスピーカを介して出力するためにコンピューティングデバイスにデジタルコンポーネントオブジェクトのオーディオトラックを提供し得る。

【０１５７】

データ処理システムは、スマートスピーカを含むコンピューティングデバイスの種類に基づいてデジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換すると決定し得る。データ処理システムは、デジタルアシスタントを含むコンピューティングデバイスの種類に基づいてデジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換すると決定し得る。

【０１５８】

データ処理システムは、要求に応じて、リアルタイムコンテンツ選択プロセスに入力されたコンテンツ選択基準に基づいてデジタルコンポーネントオブジェクトを選択することが可能であり、デジタルコンポーネントオブジェクトは、複数のサードパーティのコンテンツプロバイダによって提供された複数のデジタルコンポーネントオブジェクトから選択される。データ処理システムは、要求の前にコンピューティングデバイスによってレンダリングされたコンテンツに関連するキーワードに基づいてデジタルコンポーネントオブジェクトを選択し得る。デジタルコンポーネントオブジェクトは、複数のサードパーティのコンテンツプロバイダによって提供された複数のデジタルコンポーネントオブジェクトから選択され得る。

【０１５９】

データ処理システムは、自然言語生成モデルによって、デジタルコンポーネントオブジェクトのメタデータに基づいてデジタルコンポーネントオブジェクトに関するテキストを生成し得る。データ処理システムは、音声モデルによって、デジタルコンポーネントオブジェクトのコンテキストに基づいてデジタル音声を選択し得る。音声モデルは、オーディオおよびビジュアルメディアコンテンツを含む履歴的なデータセットを用いて機械学習技術によって訓練され得る。

【０１６０】

データ処理システムは、音声の特性のベクトルを生成するために、音声モデルにデジタルコンポーネントオブジェクトのコンテキストを入力し得る。音声モデルは、オーディオおよびビジュアルメディアコンテンツを含む履歴的なデータセットを用いて機械学習エンジンによって訓練され得る。データ処理システムは、音声の特性のベクトルに基づいて複数のデジタル音声からデジタル音声を選択し得る。

【０１６１】

データ処理システムは、メタデータに基づいて、オーディオトラックにトリガワードを追加すると決定し得る。第2の入力オーディオ信号内のトリガワードの検出は、データ処理システムまたはコンピューティングデバイスにトリガワードに対応するデジタルアクションを実行させる。

【０１６２】

データ処理システムは、デジタルコンポーネントオブジェクトのカテゴリを決定し得る。データ処理システムは、データベースから、カテゴリに関連する複数のデジタルアクションに対応する複数のトリガワードを取り出し得る。データ処理システムは、トリガキーワードの履歴的な実行に基づいて訓練されたデジタルアクションモデルを使用して、デジタルコンポーネントオブジェクトのコンテキストおよびコンピューティングデバイスの種類に基づいて複数のトリガワードをランク付けし得る。データ処理システムは、オーディオトラックに追加するために最も高いランク付けのトリガキーワードを選択し得る。

【０１６３】

データ処理システムは、デジタルコンポーネントオブジェクト内のビジュアルオブジェクトを特定するためにデジタルコンポーネントオブジェクトに対して画像認識を実行し得

る。データ処理システムは、データベースに記憶された複数の口頭でないオーディオの合図からビジュアルオブジェクトに対応する口頭でないオーディオの合図を選択し得る。

【０１６４】

データ処理システムは、画像認識技術によってデジタルコンポーネントオブジェクト内の複数のビジュアルオブジェクトを特定し得る。データ処理システムは、複数のビジュアルオブジェクトに基づいて、複数の口頭でないオーディオの合図を選択し得る。データ処理システムは、ビジュアルオブジェクトの各々とメタデータとの間の一致のレベルを示すビジュアルオブジェクトの各々に関する一致スコアを決定し得る。データ処理システムは、一致スコアに基づいて複数の口頭でないオーディオの合図をランク付けし得る。データ処理システムは、複数の口頭でないオーディオの合図の各々とテキストをレンダリングするためにコンテキストに基づいて選択されたデジタル音声との間のオーディオの干渉のレベルを判定し得る。データ処理システムは、最も高いランクに基づいて、閾値未満のオーディオの干渉のレベルに関連する複数の口頭でないオーディオの合図から口頭でないオーディオの合図を選択し得る。

【０１６５】

データ処理システムは、履歴的な実行データを使用して訓練された挿入モデルに基づいて、コンピューティングデバイスによって出力されるデジタルメディアストリーム内のオーディオトラックに関する挿入点を特定し得る。データ処理システムは、コンピューティングデバイスにデジタルメディアストリーム内の挿入点においてオーディオトラックをレンダリングさせるためにコンピューティングデバイスに命令を与えることができる。

【０１６６】

この技術的な解決策の少なくとも1つの態様は、オーディオトラックを生成する方法を対象とする。方法は、データ処理システムの1つまたは複数のプロセッサによって実行され得る。方法は、データ処理システムがデータ処理システムの遠隔のコンピューティングデバイスのマイクロフォンによって検出された入力オーディオ信号を含むデータパケットを受信するステップを含み得る。方法は、データ処理システムが要求を特定するために入力オーディオ信号を解析するステップを含み得る。方法は、データ処理システムが要求に基づいてビジュアル出力フォーマットを有するデジタルコンポーネントオブジェクトを選択するステップであって、デジタルコンポーネントオブジェクトがメタデータに関連付けられる、ステップを含み得る。方法は、データ処理システムがコンピューティングデバイスの種類に基づいてデジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換すると決定するステップを含み得る。方法は、データ処理システムが、デジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換する決定に応じて、デジタルコンポーネントオブジェクトに関するテキストを生成するステップを含み得る。方法は、データ処理システムが、デジタルコンポーネントオブジェクトのコンテキストに基づいて、テキストをレンダリングするためのデジタル音声を選択するステップを含み得る。方法は、データ処理システムがデジタル音声によってレンダリングされたテキストを用いてデジタルコンポーネントオブジェクトの基礎となるオーディオトラックを構築するステップを含み得る。方法は、データ処理システムがデジタルコンポーネントオブジェクトに基づいて口頭でないオーディオの合図を生成するステップを含み得る。方法は、データ処理システムが、デジタルコンポーネントオブジェクトのオーディオトラックを生成するために口頭でないオーディオの合図をデジタルコンポーネントオブジェクトの基礎となるオーディオの形態と組み合わせるステップを含み得る。方法は、データ処理システムが、コンピューティングデバイスからの要求に応じて、コンピューティングデバイスのスピーカを介して出力するためにコンピューティングデバイスにデジタルコンポーネントオブジェクトのオーディオトラックを提供するステップを含み得る。

【０１６７】

方法は、データ処理システムがスマートスピーカを含むコンピューティングデバイスの種類に基づいてデジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換すると決定するステップを含み得る。方法は、データ処理システムが、要求に応じて、

リアルタイムコンテンツ選択プロセスに入力されたコンテンツ選択基準に基づいてデジタルコンポーネントオブジェクトを選択するステップであって、デジタルコンポーネントオブジェクトが、複数のサードパーティのコンテンツプロバイダによって提供された複数のデジタルコンポーネントオブジェクトから選択される、ステップを含み得る。

【０１６８】

方法は、データ処理システムが要求の前にコンピューティングデバイスによってレンダリングされたコンテンツに関連するキーワードに基づいてデジタルコンポーネントオブジェクトを選択するステップを含み得る。デジタルコンポーネントオブジェクトは、複数のサードパーティのコンテンツプロバイダによって提供された複数のデジタルコンポーネントオブジェクトから選択され得る。

【０１６９】

この技術的な解決策の少なくとも1つの態様は、オーディオトラックを生成するためのシステムを対象とする。システムは、1つまたは複数のプロセッサを有するデータ処理システムを含み得る。データ処理システムは、コンピューティングデバイスによってレンダリングされるデジタルストリーミングコンテンツに関連するキーワードを特定し得る。データ処理システムは、キーワードに基づいて、ビジュアル出力フォーマットを有するデジタルコンポーネントオブジェクトを選択することが可能であり、デジタルコンポーネントオブジェクトは、メタデータに関連付けられる。データ処理システムは、コンピューティングデバイスの種類に基づいて、デジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換すると決定し得る。データ処理システムは、デジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換する決定に応じて、デジタルコンポーネントオブジェクトに関するテキストを生成し得る。データ処理システムは、デジタルコンポーネントオブジェクトのコンテキストに基づいて、テキストをレンダリングするためのデジタル音声を選択し得る。データ処理システムは、デジタル音声によってレンダリングされたテキストを用いてデジタルコンポーネントオブジェクトの基礎となるオーディオトラックを構築し得る。データ処理システムは、デジタルコンポーネントオブジェクトに基づいて、口頭でないオーディオの合図を生成し得る。データ処理システムは、デジタルコンポーネントオブジェクトのオーディオトラックを生成するために口頭でないオーディオの合図をデジタルコンポーネントオブジェクトの基礎となるオーディオの形態と組み合わせ得る。データ処理システムは、コンピューティングデバイスのスピーカを介して出力するためにコンピューティングデバイスにデジタルコンポーネントオブジェクトのオーディオトラックを提供し得る。

【０１７０】

データ処理システムは、スマートスピーカを含むコンピューティングデバイスの種類に基づいてデジタルコンポーネントオブジェクトをオーディオ出力フォーマットに変換すると決定し得る。データ処理システムは、リアルタイムコンテンツ選択プロセスに入力されたキーワードに基づいてデジタルコンポーネントオブジェクトを選択することが可能であり、デジタルコンポーネントオブジェクトは、複数のサードパーティのコンテンツプロバイダによって提供された複数のデジタルコンポーネントオブジェクトから選択される。

【０１７１】

今やいくつかの例示的な実装を説明したが、以上は例示的であり、限定的でなく、例として提示されたことは明らかである。特に、本明細書において提示された例の多くは方法の行為またはシステムの要素の特定の組合せを含むが、それらの行為およびそれらの要素は、同じ目的を達成するためにその他の方法で組み合わされる可能性がある。1つの実装に関連して検討された行為、要素、および特徴は、その他の実装または実装の同様の役割から除外されるように意図されていない。

【０１７２】

本明細書において使用された言葉遣いおよび用語は、説明を目的としており、限定と見なされるべきでない。本明細書における「～を含む(including)」、「～を含む(comprising)」、「～を有する」、「～を含む(containing)」、「～を含む(involving)」、「～

によって特徴付けられる(characterized by)」、「～ことを特徴とする(characterized in that)」、およびこれらの変化形の使用は、その後に列挙された項目、それらの項目の均等物、および追加的な項目、ならびにその後に列挙された項目だけからなる代替的な実装を包含するように意図される。1つの実装において、本明細書に記載のシステムおよび方法は、説明された要素、行為、またはコンポーネントのうちの1つ、2つ以上のそれぞれの組合せ、またはすべてからなる。

【０１７３】

本明細書において単数形で言及されたシステムおよび方法の実装または要素または行為へのすべての言及は、複数のこれらの要素を含む実装も包含する可能性があり、本明細書における任意の実装または要素または行為への複数形のすべての言及は、単一の要素のみを含む実装も包含する可能性がある。単数形または複数形の言及は、今開示されたシステムまたは方法、それらのコンポーネント、行為、または要素を単一のまたは複数の構成に限定するように意図されていない。任意の情報、行為、または要素に基づいている任意の行為または要素への言及は、行為または要素が任意の情報、行為、または要素に少なくとも部分的に基づく実装を含む可能性がある。

【０１７４】

本明細書において開示された任意の実装は、任意のその他の実装または実施形態と組み合わされる可能性があり、「実装」、「いくつかの実装」、「1つの実装」などの言及は、必ずしも相互排他的ではなく、実装に関連して説明された特定の特徴、構造、または特色が少なくとも1つの実装または実施形態に含まれる可能性があることを示すように意図される。本明細書において使用されるそのような用語は、必ずしもすべてが同じ実装に言及しているとは限らない。任意の実装は、本明細書において開示された態様および実装に合致する任意の方法で包括的または排他的に任意のその他の実装と組み合わされる可能性がある。

【０１７５】

「または(or)」との言及は、「または(or)」を使用して記載された任意の項が記載された項のうちの1つ、2つ以上、およびすべてのいずれかを示す可能性があるように包含的であると見なされる可能性がある。項の連言的リストのうちの少なくとも1つのへの言及は、記載された項のうちの1つ、2つ以上、およびすべてのいずれかを示す包含的なまたは(OR)と見なされる可能性がある。たとえば、「『A』および『B』のうちの少なくとも一方」との言及は、「A」のみ、「B」のみ、および「A」と「B」との両方を含み得る。「～を含む(comprising)」またはその他の非限定的用語と関連して使用されるそのような言及は、追加的な項を含み得る。

【０１７６】

図面、詳細な説明、または任意の請求項の技術的な特徴が後に参照符号を付されている場合、参照符号は、図面、詳細な説明、および請求項を理解し易くするために含められた。したがって、参照符号があることもないことも、いかなる請求項の要素の範囲に対するいかなる限定的な効果も持たない。

【０１７７】

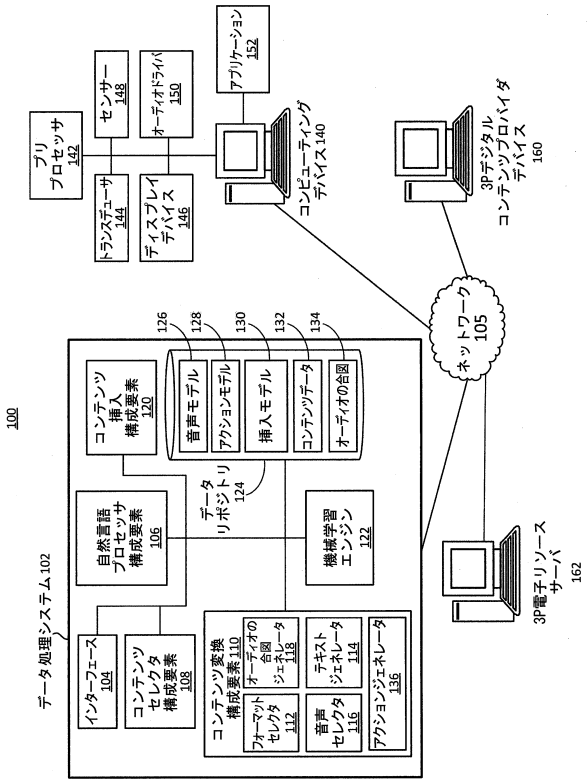
本明細書に記載のシステムおよび方法は、それらの特徴を逸脱することなくその他の特定の形態で具現化される可能性がある。たとえば、3Pデジタルコンテンツプロバイダデバイス160などの3Pまたはサードパーティと記載されたデバイス、製品、またはサービスは、部分的にもしくは完全にファーストパーティの(first party)デバイス、製品、もしくはサービスであることが可能であり、または部分的にもしくは完全にファーストパーティのデバイス、製品、もしくはサービスを含むことが可能であり、データ処理システム102またはその他のコンポーネントに関連するエンティティによって共有され得る。上述の実装は、説明されたシステムおよび方法の限定ではなく、例示的である。したがって、本明細書に記載のシステムおよび方法の範囲は、上述の説明ではなく添付の請求項によって示され、請求項の均等の意味および範囲内に入る変更は、それに包含される。

【符号の説明】

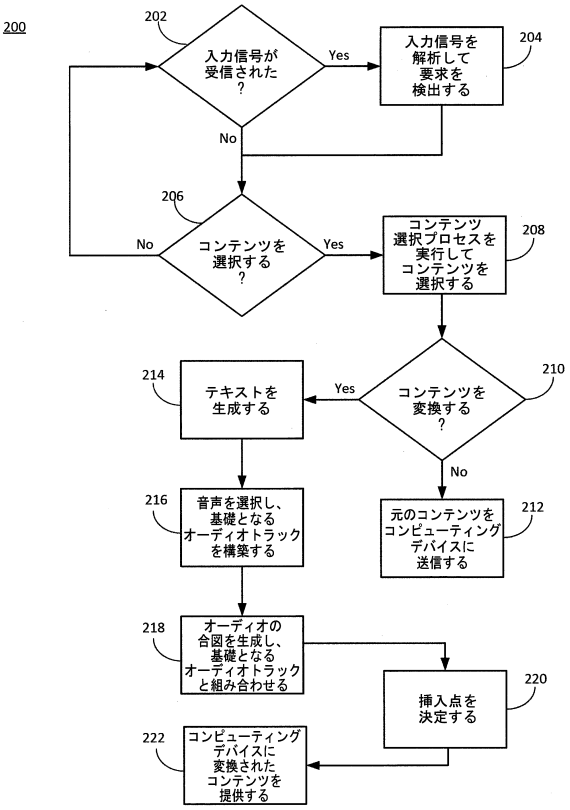
【０１７８】
100 システム
102 データ処理システム
104 インターフェース
105 ネットワーク
106 自然言語プロセッサ構成要素
108 コンテンツセレクタ構成要素
110 コンテンツ変換構成要素
112 フォーマットセレクタ
114 テキストジェネレータ
116 音声セレクタ
118 オーディオの合図ジェネレータ
120 コンテンツ挿入構成要素
122 機械学習エンジン
124 データリポジトリ
126 音声モデル
128 アクションモデル
130 挿入モデル
132 コンテンツデータ
134 オーディオの合図
136 アクションジェネレータ
140 コンピューティングデバイス
142 プリプロセッサ
144 トランスデューサ
146 ディスプレイデバイス
148 センサー
150 オーディオドライバ
152 アプリケーション
160 3Pデジタルコンテンツプロバイダデバイス
162 3P電子リソースサーバ
200 方法
300 コンピュータシステム、コンピューティングデバイス
305 バス
310 プロセッサ
315 メインメモリ
320 ROM
325 ストレージデバイス
330 入力デバイス
335 ディスプレイ

【図面】

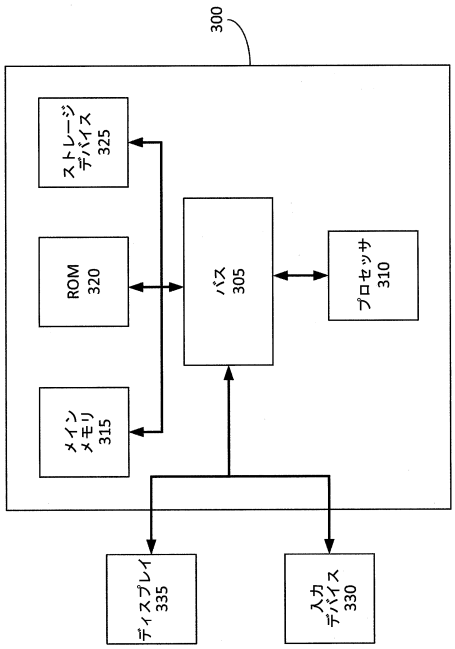
【図１】


100
データ処理システム102
プリプロセッサ 142
センサー 148
オーディオドライバ 150
アプリケーション 152
トランスデューサ 144
ディスプレイデバイス 146
コンピューティングデバイス 140
3Pデジタルコンテンツプロバイダデバイス 160
ネットワーク 105
3P電子リソースサーバ 162
コンテンツ挿入構成要素 120
音声モデル
アクションモデル
挿入モデル
コンテンツデータ
オーディオの合図
126
128
130
132
134
データリポジトリ 124
自然言語プロセッサ構成要素 106
機械学習エンジン 122
インターフェース 104
コンテンツセレクタ構成要素 108
コンテンツ変換構成要素 110
オーディオの合図ジェネレータ 118
テキストジェネレータ 114
フォーマットセレクタ 112
音声セレクタ 116
アクションジェネレータ 136

【図２】


200
202
入力信号が受信された？
Yes
No
204
入力信号を解析して要求を検出する
206
コンテンツを選択する？
No
Yes
208
コンテンツ選択プロセスを実行してコンテンツを選択する
210
コンテンツを変換する？
Yes
No
214
テキストを生成する
216
音声を選択し、基礎となるオーディオトラックを構築する
212
元のコンテンツをコンピューティングデバイスに送信する
218
オーディオの合図を生成し、基礎となるオーディオトラックと組み合わせる
220
挿入点を決定する
222
コンピューティングデバイスに変換されたコンテンツを提供する

【図３】


300
ストレージデバイス 325
ROM 320
バス 305
プロセッサ 310
メインメモリ 315
ディスプレイ 335
入力デバイス 330

フロントページの続き

(72)発明者　マシュー・シャリフィ
アメリカ合衆国・カリフォルニア・９４０４３・マウンテン・ビュー・アンフィシアター・パークウェイ・１６００
(72)発明者　ヴィクター・カルブネ
アメリカ合衆国・カリフォルニア・９４０４３・マウンテン・ビュー・アンフィシアター・パークウェイ・１６００

審査官　円子 英紀
(56)参考文献　国際公開第２０１９／２１６９６９（WO，Ａ１）
特開２００２－３２８９４９（ＪＰ，Ａ）
特表２０１３－５１７７３９（ＪＰ，Ａ）
(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　３／１６
Ｇ１０Ｌ　１５／１０
Ｇ１０Ｌ　１５／２８