

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7577204号
(P7577204)

(45)発行日 令和6年11月1日(2024.11.1)

(24)登録日 令和6年10月24日(2024.10.24)

(51)国際特許分類 F I
 G 0 1 C 21/26 (2006.01) G 0 1 C 21/26 A
 G 0 1 C 21/36 (2006.01) G 0 1 C 21/36
 G 0 6 F 3/16 (2006.01) G 0 6 F 3/16 5 4 0

請求項の数 20 (全29頁)

(21)出願番号	特願2023-519747(P2023-519747)	(73)特許権者	502208397
(86)(22)出願日	令和2年10月22日(2020.10.22)		グーグル エルエルシー
(65)公表番号	特表2023-547324(P2023-547324 A)		Google LLC
(43)公表日	令和5年11月10日(2023.11.10)		アメリカ合衆国 カリフォルニア州 9 4 0 4 3 マウンテン ビュー アンフィシ
(86)国際出願番号	PCT/US2020/056756		アター パークウェイ 1 6 0 0
(87)国際公開番号	WO2022/086528		1 6 0 0 Amphitheatre P
(87)国際公開日	令和4年4月28日(2022.4.28)		arkway 9 4 0 4 3 Mounta
審査請求日	令和5年5月26日(2023.5.26)		in 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】 コンテンツ認識ナビゲーション指示

(57)【特許請求の範囲】

【請求項1】

コンテンツ認識ナビゲーション指示を生成するための方法であって、前記方法が、マッピングアプリケーションを介してクライアントデバイス内の1つまたは複数のプロセッサによって、ルートに沿って出発地から目的地まで移動するための1つまたは複数の音声ナビゲーション指示を取得するステップと、

前記1つまたは複数のプロセッサによって、前記マッピングアプリケーションとは異なるソースから再生されている電子メディアコンテンツを識別するステップであって、前記ソースが前記クライアントデバイスにおいてまたは前記クライアントデバイスに近接して実行されている、ステップと、

前記1つまたは複数のプロセッサによって、前記電子メディアコンテンツの特性を決定するステップであって、

前記電子メディアコンテンツの特性は、前記電子メディアコンテンツの発話速度、前記電子メディアコンテンツの転写、前記電子メディアコンテンツの長さ、または電子メディアコンテンツ内に含まれる関心地点(POI)もしくは地理的トピックのうち、少なくとも1つを含む、決定するステップと、

前記1つまたは複数のプロセッサによって、前記電子メディアコンテンツの前記特性に従って前記1つまたは複数の音声ナビゲーション指示のうち少なくとも1つを調整するステップと、

前記1つまたは複数のプロセッサによって、前記少なくとも1つの調整された音声ナビゲ

ーション指示をユーザに提示するステップとを含む、方法。

【請求項 2】

前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つを調整するステップが、

前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つが提示されるタイミングを調整するステップ、

前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つが提示される言語を調整するステップ、

前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つが提示される速度を調整するステップ、または

前記ルートに沿った関心地点(POI)に関する推奨事項を提供するステップのうちの少なくとも1つを含む、

請求項1に記載の方法。

【請求項 3】

前記電子メディアコンテンツの特性を決定するステップが、

前記電子メディアコンテンツが前記ソースによって提示される速度を決定するステップ、

前記電子メディアコンテンツが前記ソースによって提示される言語を決定するステップ、

前記電子メディアコンテンツにおける一時停止を識別するステップ、または

前記電子メディアコンテンツ内に含まれる関心地点(POI)もしくは地理的トピックを識別するステップ

のうちの少なくとも1つを含む、

請求項1または2に記載の方法。

【請求項 4】

前記電子メディアコンテンツの前記特性に従って前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つを調整するステップが、

前記1つまたは複数のプロセッサによって、前記少なくとも音声ナビゲーション指示における操縦の複雑さのレベルを決定するステップであって、前記操縦の複雑さのレベルは前記音声ナビゲーション指示によって示される操縦タイプに応じて決定されるレベルである、決定するステップと、

前記1つまたは複数のプロセッサによって、前記電子メディアコンテンツの関連性レベルを決定するステップであって、前記関連性レベルは前記電子メディアコンテンツの主題、前記ユーザの現在位置、および/または前記音声ナビゲーション指示の目的地に基づいて決定される、決定するステップと、

前記1つまたは複数のプロセッサによって、少なくとも1つの音声ナビゲーション指示の前記操縦の複雑さのレベルを前記電子メディアコンテンツの前記関連性レベルと比較するステップと、

前記1つまたは複数のプロセッサによって、前記比較に基づいて、前記少なくとも1つの音声ナビゲーション指示を提示するかどうかを決定するステップと

を含む、

請求項1から3のいずれか一項に記載の方法。

【請求項 5】

前記比較に基づいて、前記少なくとも1つの音声ナビゲーション指示を提示するかどうかを決定するステップが、

前記少なくとも1つの音声ナビゲーション指示の前記操縦の複雑さのレベルが前記電子メディアコンテンツの前記関連性レベルを超えると判断したことに応答して、前記1つまたは複数のプロセッサによって、前記少なくとも1つの音声ナビゲーション指示を提示するステップを含む、

請求項4に記載の方法。

10

20

30

40

50

【請求項 6】

前記比較に基づいて前記少なくとも1つの音声ナビゲーション指示を提示するかどうかを決定するステップが、

前記電子メディアコンテンツの前記関連性レベルが前記音声ナビゲーション指示の前記操縦の複雑さのレベルを超えていると判断したことに応答して、前記1つまたは複数のプロセッサによって、前記少なくとも1つの音声ナビゲーション指示を提示しないステップを含む、

請求項4に記載の方法。

【請求項 7】

前記マッピングアプリケーションとは異なるソースから再生されている電子メディアコンテンツを識別するステップが、

前記1つまたは複数のプロセッサによって、前記マッピングアプリケーションとは異なる前記クライアントデバイス上で実行されている音声アプリケーションから音声再生データを取得するステップ、

前記1つまたは複数のプロセッサによって、前記クライアントデバイスに通信可能に結合されたデバイスから音声再生データを取得するステップ、または

前記1つまたは複数のプロセッサによって、周囲音声フィンガープリントを、所定のメディアコンテンツの1つまたは複数の音声フィンガープリントと比較するステップのうちの少なくとも1つを含む、

請求項1から6のいずれか一項に記載の方法。

【請求項 8】

前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つを調整するステップが、

前記1つまたは複数のプロセッサによって、前記電子メディアコンテンツ内の文の終わりを識別するステップと、

前記1つまたは複数のプロセッサによって、前記文の終了後に前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つを提示するステップとを含む、

請求項1から7のいずれか一項に記載の方法。

【請求項 9】

前記1つまたは複数のプロセッサによって、前記電子メディアコンテンツが前記1つまたは複数の音声ナビゲーション指示に従って提示される方法を調整するステップをさらに含む、請求項1から8のいずれか一項に記載の方法。

【請求項 10】

前記電子メディアコンテンツの前記特性に従って前記少なくとも1つの音声ナビゲーション指示を調整するステップが、

前記電子メディアコンテンツの前記特性を事前に記憶された規則のセットに適用することによって、前記少なくとも1つの音声ナビゲーション指示を調整するステップ、または

特定の電子メディアコンテンツに対して音声ナビゲーション指示を調整するための機械学習モデルを生成し、前記電子メディアコンテンツの前記特性を前記機械学習モデルに適用することによって、前記少なくとも1つの音声ナビゲーション指示を調整するステップの少なくとも1つを含む、

請求項1から9のいずれか一項に記載の方法。

【請求項 11】

コンテンツ認識ナビゲーション指示を生成するためのクライアントデバイスであって、前記クライアントデバイスが、

スピーカと、

1つまたは複数のプロセッサと、

前記1つまたは複数のプロセッサおよび前記スピーカに結合され、前記1つまたは複数のプロセッサによって実行されると、前記クライアントデバイスに、

10

20

30

40

50

マッピングアプリケーションを介して、ルートに沿って出発地から目的地まで移動するための1つまたは複数の音声ナビゲーション指示を取得する動作と、

前記マッピングアプリケーションとは異なるソースから再生されている電子メディアコンテンツを識別する動作であって、前記ソースが前記クライアントデバイスにおいてまたはクライアントデバイスに近接して実行されている、動作と、

前記電子メディアコンテンツの特性を決定する動作であって、
前記電子メディアコンテンツの特性は、前記電子メディアコンテンツの発話速度、前記電子メディアコンテンツの転写、前記電子メディアコンテンツの長さ、または電子メディアコンテンツ内に含まれる関心地点(POI)もしくは地理的トピックのうち、少なくとも1つを含む、決定する動作と、

10

前記電子メディアコンテンツの前記特性に従って前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つを調整する動作と、

前記スピーカを介して、前記少なくとも1つの調整された音声ナビゲーション指示をユーザに提示する動作と

を実行させる指示を記憶するコンピュータ可読メモリと
を備える、
クライアントデバイス。

【請求項12】

前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つを調整するために、前記指示が、前記クライアントデバイスに、

20

前記1つまたは複数の音声ナビゲーション指示のうち少なくとも1つが提示されるタイミングを調整する動作、

前記1つまたは複数の音声ナビゲーション指示のうち少なくとも1つが提示される言語を調整する動作、

前記1つまたは複数の音声ナビゲーション指示のうち少なくとも1つが提示される速度を調整する動作、または

前記ルートに沿った関心地点(POI)に関する推奨事項を提供する動作
のうちの少なくとも1つを実行させる、
請求項11に記載のクライアントデバイス。

【請求項13】

30

前記電子メディアコンテンツの特性を決定するために、前記指示が、前記クライアントデバイスに、

前記電子メディアコンテンツが前記ソースによって提示される速度を決定する動作、

前記電子メディアコンテンツが前記ソースによって提示される言語を決定する動作、

前記電子メディアコンテンツにおける一時停止を識別する動作、または

前記電子メディアコンテンツ内に含まれる関心地点(POI)もしくは地理的トピックを識別する動作

のうちの少なくとも1つを実行させる、

請求項11または12に記載のクライアントデバイス。

【請求項14】

40

前記電子メディアコンテンツの前記特性に従って少なくとも1つの音声ナビゲーション指示を調整するために、前記指示が、前記クライアントデバイスに、

前記少なくとも1つの音声ナビゲーション指示における操縦の複雑さのレベルを決定する動作であって、前記操縦の複雑さのレベルは前記音声ナビゲーション指示によって示される操縦タイプに応じて決定されるレベルである、決定する動作と、

前記電子メディアコンテンツの関連性レベルを決定する動作であって、前記関連性レベルは前記電子メディアコンテンツの主題、前記ユーザの現在位置、および/または前記音声ナビゲーション指示の目的地に基づいて決定される、決定する動作と、

前記少なくとも1つの音声ナビゲーション指示の前記操縦の複雑さのレベルを前記電子メディアコンテンツの前記関連性レベルと比較する動作と、

50

前記比較に基づいて、前記少なくとも1つの音声ナビゲーション指示を提示するかどうかを決定する動作と
 を実行させる、
 請求項11から13のいずれか一項に記載のクライアントデバイス。

【請求項15】

前記比較に基づいて、前記少なくとも1つの音声ナビゲーション指示を提示するかどうかを決定するために、前記指示が、前記クライアントデバイスに、

前記少なくとも1つの音声ナビゲーション指示の前記操縦の複雑さのレベルが前記電子メディアコンテンツの前記関連性レベルを超えると判断したことに応答して、前記少なくとも1つの音声ナビゲーション指示を提示させる、

10

請求項14に記載のクライアントデバイス。

【請求項16】

前記マッピングアプリケーションとは異なるソースから再生されている電子メディアコンテンツを識別するために、前記指示が、前記クライアントデバイスに、

前記マッピングアプリケーションとは異なる前記クライアントデバイス上で実行されている音声アプリケーションから音声再生データを取得する動作、

前記クライアントデバイスに通信可能に結合されたデバイスから音声再生データを取得する動作、または

周囲音声フィンガープリントを、所定のメディアコンテンツの1つまたは複数の音声フィンガープリントと比較する動作

20

のうちの少なくとも1つを実行させる、

請求項11から15のいずれか一項に記載のクライアントデバイス。

【請求項17】

前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つを調整するために、前記指示が、前記クライアントデバイスに、

前記電子メディアコンテンツ内の文の終わりを識別する動作と、

前記文の終了後に前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つを提示する動作と

を実行させる、

請求項11から16のいずれか一項に記載のクライアントデバイス。

30

【請求項18】

前記指示がさらに、前記クライアントデバイスに、

前記電子メディアコンテンツが前記1つまたは複数の音声ナビゲーション指示に従って提示される方法を調整させる、

請求項11から17のいずれか一項に記載のクライアントデバイス。

【請求項19】

1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサに、

マッピングアプリケーションを介して、ルートに沿って出発地から目的地まで移動するための1つまたは複数の音声ナビゲーション指示を取得する動作と、

40

前記マッピングアプリケーションとは異なるソースから再生されている電子メディアコンテンツを識別する動作であって、前記ソースがクライアントデバイスにおいてまたは前記クライアントデバイスに近接して実行されている、動作と、

前記電子メディアコンテンツの特性を決定する動作であって、前記電子メディアコンテンツの特性は、前記電子メディアコンテンツの発話速度、前記電子メディアコンテンツの転写、前記電子メディアコンテンツの長さ、または電子メディアコンテンツ内に含まれる関心地点(POI)もしくは地理的トピックのうち、少なくとも1つを含む、決定する動作と、

前記電子メディアコンテンツの前記特性に従って前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つを調整する動作と、

スピーカを介して、前記少なくとも1つの調整された音声ナビゲーション指示をユーザ

50

に提示する動作と
 を実行させる指示を記憶するコンピュータ可読メモリ。

【請求項 20】

前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つを調整するために、
 前記指示が、前記1つまたは複数のプロセッサに、

前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つが提示されるタイ
 ミングを調整する動作、

前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つが提示される言語
 を調整する動作、

前記1つまたは複数の音声ナビゲーション指示のうちの少なくとも1つが提示される速度
 を調整する動作、または

前記ルートに沿った関心地点(POI)に関する推奨事項を提供する動作

のうちの少なくとも1つを実行させる、

請求項19に記載のコンピュータ可読メモリ。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、コンテンツ認識(content-aware)ナビゲーション指示に関し、より具体的に
 は、エリア内または同じデバイス上で再生される他のメディア/音声コンテンツに基づいて
 音声ナビゲーション指示を調整することに関する。

【背景技術】

【0002】

本明細書で提供する背景説明は、本開示の文脈を一般的に提示する目的のためのもの
 である。この背景セクションにおいて説明されている範囲までの現在名前が挙げられている
 発明者の業績、ならびに出願時に場合によっては従来技術として適格ではない可能性があ
 る本明細書本文の態様は、本開示に対する従来技術として明示的にも暗黙的にも認めら
 れない。

【0003】

今日、コンピュータ、スマートフォンなど、または組み込みデバイス内で実行されるソ
 フトウェアアプリケーションは、段階的なナビゲーション指示を生成する。典型的には、
 ユーザが、出発地と目的地とを指定し、ソフトウェアアプリケーションは、すぐに、およ
 び/またはユーザが出発地から目的地に移動するときに、指示を表示、または音声フォー
 マットにおいて提示する。

【0004】

これらのソフトウェアアプリケーションは、ルートに基づいてナビゲーション指示を生
 成するために、一般に、距離、通りの名前、建物の番号の指標を利用する。例えば、これ
 らのシステムは、「4分の1マイル進み、次いでMaple Streetを右折してください」のよ
 うな指示を運転者に提供することができる。

【発明の概要】

【発明が解決しようとする課題】

【0005】

車載ナビゲーションシステムによって提供される音声指示は、一般に有益なものでなけ
 ればならず、聞いている人、特に運転者の気を不必要に散らすことなく、従うべきル
 ートについて十分かつ適切なタイミングで指示を提供しなければならない。それに加えて、音
 声指示と同時に他の音声車両内で再生されるので、車載ナビゲーションシステムは、一
 般に運転者の注意をひくために競争しなければならない。複数の音声出力の組合せは、ユ
 ーザがナビゲーション指示を聞くことができない、またはナビゲーション指示を理解でき
 ない、最適ではない体験になる可能性がある。車両内の他の音声は、運転者の好みの言語
 、運転者の好みの音声再生速度、運転者の好みの目的地などの運転者の好みを示す情報
 を提供する場合もある。したがって、ナビゲーション中に再生される他のメディアコンテン

10

20

30

40

50

ツと深く統合され、車両内で再生される他の音声に従って適応される音声指示を提供する改善されたシステム-ユーザ音声インターフェースが有益である場合がある。

【課題を解決するための手段】

【0006】

いくつかの実装形態において、音声ナビゲーション指示を提供するマッピングアプリケーションは、エリア内または同じデバイス上(例えば、同じ車両内、同じ室内、アプリケーションのしきい値距離内の屋外エリア内など)で再生される他のメディア/音声コンテンツを識別し得る。次いで、マッピングアプリケーションは、メディア/音声コンテンツに従って音声ナビゲーション指示を調整し得る。例えば、メディア/音声コンテンツが音声ブック、ポッドキャスト、音楽などである場合、マッピングアプリケーションは、メディア/音声コンテンツが途切れるまで、次のナビゲーション指示を提供するのを待ち得る。より具体的には、マッピングアプリケーションは、セグメントの終わり、歌の終わり、または文の終わりの後に、次のナビゲーション指示を再生し得る。別の例において、メディア/音声コンテンツが特定の言語(例えば、スペイン語)において提示される場合、マッピングアプリケーションは、その言語において音声ナビゲーション指示を提示し得る。さらに別の例において、マッピングアプリケーションは、メディア/音声コンテンツにおいて言及されているルートに沿った関心地点(POI)に関する提案または推奨事項を提供し得る。

10

【0007】

エリア内で再生されている他のメディア/音声コンテンツを識別するために、マッピングアプリケーションは、例えば、アプリケーションプログラミングインターフェース(API)を介して、クライアントデバイス上で実行されている他のアプリケーションと通信し得る。クライアントデバイス上で実行されている他のアプリケーションは、他のアプリケーションが再生されている速度(例えば、発話速度)(例えば、1.5x、2xなど)、メディア/音声コンテンツの言語、メディア/音声コンテンツにおける話者の声の特性、メディア/音声コンテンツの転写、メディア/音声コンテンツの長さなどの、他のアプリケーションによって再生されているメディア/音声コンテンツの特性を含む音声再生データを提供し得る。それに加えて、マッピングアプリケーションは、短距離通信リンクを介するなどして、メディア/音声コンテンツを再生している他のデバイス(例えば、車両ヘッドユニット)と通信し得る。他のデバイスはまた、メディア/音声コンテンツの特性を含む音声再生データをマッピングアプリケーションに提供し得る。さらに、マッピングアプリケーションは、所定のメディア/音声コンテンツの音声フィンガープリントを周囲のエリア内の周囲音声と比較することによって、他のメディア/音声コンテンツを識別し得る。例えば、マッピングアプリケーションは、ポピュラー音楽、音声ブック、ラジオ番組、ポッドキャストなどからのメディア/音声に対応する音声フィンガープリントのライブラリを取得し得る。マッピングアプリケーションが、音声フィンガープリントのうちの1つとの一致を識別した場合、マッピングアプリケーションは、メディア/音声コンテンツがエリア内で提示されていると判断し得る。

20

30

【0008】

次いで、マッピングアプリケーションは、コンテンツに関連付けられたメタデータを受信するために、コンテンツを提示するアプリケーションもしくはデバイスと通信することによって、または音声認識技法を使用してメディア/音声コンテンツを分析することによって、メディア/音声コンテンツの特性を識別し得る。

40

【0009】

本開示の技法の1つの例示的な実施形態は、コンテンツ認識ナビゲーション指示を生成するための方法である。方法は、マッピングアプリケーションを介してクライアントデバイスにおいて、ルートに沿って出発地から目的地まで移動するための1つまたは複数の音声ナビゲーション指示を取得するステップと、マッピングアプリケーションとは異なるソースから再生されている電子メディアコンテンツを識別するステップであって、ソースがクライアントデバイスにおいてまたはクライアントデバイスに近接して実行されている、ステップと、電子メディアコンテンツの特性を決定するステップとを含む。方法は、電子

50

メディアコンテンツの特性に従って1つまたは複数の音声ナビゲーション指示のうち少なくとも1つを調整するステップと、少なくとも1つの調整された音声ナビゲーション指示をユーザに提示するステップとをさらに含む。

【0010】

本開示の技法の別の例示的な実施形態は、コンテンツ認識ナビゲーション指示を生成するためのクライアントデバイスである。クライアントデバイスは、スピーカと、1つまたは複数のプロセッサと、1つまたは複数のプロセッサおよびスピーカに結合され、指示を記憶する非一時的コンピュータ可読メモリとを含む。指示は、1つまたは複数のプロセッサによって実行されると、クライアントデバイスに、マッピングアプリケーションを介して、ルートに沿って出発地から目的地まで移動するための1つまたは複数の音声ナビゲーション指示を取得する動作と、マッピングアプリケーションとは異なるソースから再生されている電子メディアコンテンツを識別する動作であって、ソースがデバイスにおいてまたはデバイスに近接して実行されている、動作と、電子メディアコンテンツの特性を決定する動作とを実行させる。指示は、さらに、クライアントデバイスに、電子メディアコンテンツの特性に従って1つまたは複数の音声ナビゲーション指示のうち少なくとも1つを調整させ、スピーカを介して、少なくとも1つの調整された音声ナビゲーション指示をユーザに提示させる。

10

【0011】

本開示の技法のさらに別の例示的な実施形態は、指示を記憶する非一時的コンピュータ可読メモリである。指示は、1つまたは複数のプロセッサによって実行されると、1つまたは複数のプロセッサに、マッピングアプリケーションを介して、ルートに沿って出発地から目的地まで移動するための1つまたは複数の音声ナビゲーション指示を取得する動作と、マッピングアプリケーションとは異なるソースから再生されている電子メディアコンテンツを識別する動作であって、ソースがクライアントデバイスにおいてまたはクライアントデバイスに近接して実行されている、動作と、電子メディアコンテンツの特性を決定する動作とを実行させる。指示は、さらに、1つまたは複数のプロセッサに、電子メディアコンテンツの特性に従って1つまたは複数の音声ナビゲーション指示のうち少なくとも1つを調整させ、スピーカを介して、少なくとも1つの調整された音声ナビゲーション指示をユーザに提示させる。

20

【0012】

本明細書で説明する方法は、他のメディア(音声)コンテンツの音声特性を考慮に入れることによって、音声ナビゲーション指示の出力を改善する。方法は、音声ナビゲーション指示に従うリスナーによって容易に理解されるように、音声ナビゲーション指示の出力を調整する様々な方法について説明する。この改善された明瞭さおよび理解は、音声ナビゲーション指示がより安全であることを意味する。例えば、音声ナビゲーション指示は、言語などのメディアコンテンツの特性に一致するように調整され得、またはその出力は、文の終わりなどのメディアコンテンツにおける自然な中断に一致する時間において発生するように遅延/調整され得る。別の例において、メディアコンテンツは、音声ナビゲーション指示の出力を可能にするために一時停止され得、したがって、音声の重複を回避する。これらおよび他の例について、以下に説明する。

30

【図面の簡単な説明】

【0013】

【図1】本開示の技法がコンテンツ認識音声ナビゲーション指示を生成するために使用されることが可能な例示的な車両を示す図である。

【図2】コンテンツ認識音声ナビゲーション指示を生成するための技法が実装されることが可能な例示的なシステムのブロック図である。

【図3】図2のコンテンツ認識音声ナビゲーション生成システムが特定のメディアコンテンツに従ってナビゲーション指示を調整するために利用することができる例示的なナビゲーション指示データテーブルを示す図である。

【図4】クライアントコンピューティングデバイスにおいて実装されることが可能な、コ

40

50

コンテンツ認識音声ナビゲーション指示を生成するための例示的な方法のフロー図である。

【発明を実施するための形態】

【0014】

一般的に言えば、コンテンツ認識ナビゲーション指示を提供するための技法は、1つもしくはいくつかのクライアントデバイス、車両ヘッドユニット、1つもしくはいくつかのネットワークサーバ、またはこれらのデバイスの組合せを含むシステムにおいて実装されることが可能である。しかしながら、明確にするために、以下の例は、マッピングアプリケーションを実行するクライアントデバイスが、ユーザをルートに沿って出発地から目的地までナビゲートするための音声ナビゲーション指示のセットを取得する実施形態に主に焦点を当てる。例えば、ユーザは、マッピングアプリケーションを介してユーザの現在位置から目的地までのナビゲーション指示を要求し得、クライアントデバイスは、要求をナビゲーションサーバに転送し得る。次いで、ナビゲーションサーバは、音声ナビゲーション指示のセットを生成し、音声ナビゲーション指示のセットをユーザへの再生のためにクライアントデバイスに送信し得る。

10

【0015】

それに加えて、マッピングアプリケーションは、クライアントデバイス上で実行される他のアプリケーションと(例えば、APIを介して)、または車両ヘッドユニットもしくは他のクライアントデバイスなどの、クライアントデバイスの近傍内の他のデバイス(例えば、短距離通信リンク)と通信し得る。通信は、他のアプリケーション/デバイスによって再生されているメディアコンテンツの指標を含み得、メディアコンテンツの言語、メディアコンテンツの発話速度、メディアコンテンツの転写、メディアコンテンツの話者の声の特性などの、メディアコンテンツの特性を含み得る。メディアコンテンツを識別するために他のアプリケーション/デバイスと通信することに加えて、マッピングアプリケーションは、メディアコンテンツを識別するために、エリア内で再生されている周囲音声を、ポピュラーソング、ラジオ番組、ポッドキャストなどのポピュラーなメディアコンテンツに関する音声フィンガープリントのセットと比較し得る。

20

【0016】

いずれにしても、マッピングアプリケーションは、メディアコンテンツの特性を分析し、メディアコンテンツに従って音声ナビゲーション指示をどのように調整するかを決定するために、メディアコンテンツの特性をサーバデバイスに送信し得る。次いで、サーバデバイスは、調整された音声ナビゲーション指示、または音声ナビゲーション指示の再生をどのように調整するかをクライアントデバイスに送信し得る。他の実施形態において、マッピングアプリケーションは、メディアコンテンツに従って音声ナビゲーション指示をどのように調整するかを決定し得る。

30

【0017】

本明細書で使用されるメディアコンテンツは、ラジオ番組、ポッドキャスト、音声ブック、音楽、広告、テレビ番組、映画、ビデオ、または音声成分を含む任意の他のタイプのメディアを含み得る。本明細書で説明する音声ナビゲーション指示は、車両を介した目的地への運転指示を含むが、音声ナビゲーション指示は、徒歩、自転車、公共交通機関など、任意の適切な交通手段の様式ためのものであり得る。

40

【0018】

例示的なハードウェアおよびソフトウェア構成要素

図1を参照すると、上記で概説した技法が実装されることが可能な例示的な環境1は、携帯デバイス10と、ヘッドユニット14を有する車両12とを含む。携帯デバイス10は、例えば、スマートフォン、タブレットコンピュータ、または車載ナビゲーションシステムであり得る。携帯デバイス10は、有線(例えば、ユニバーサルシリアルバス(USB))またはワイヤレス(例えば、Bluetooth、Wi-Fi Direct)であり得る通信リンク16を介して、車両12のヘッドユニット14と通信する。携帯デバイス10は、第4世代または第3世代セルラーネットワーク(それぞれ、4Gまたは3G)などのワイヤレス通信ネットワークを介して、様々なコンテンツプロバイダ、サーバなどと通信することもできる。

50

【 0 0 1 9 】

ヘッドユニット14は、デジタル地図などのナビゲーション情報を提示するためのディスプレイ18を含むことができる。ディスプレイ18は、タッチスクリーンであり、目的地、出発地などの名前または住所を含み得るテキスト入力を入力するためのソフトウェアキーボードを含む。それぞれヘッドユニット14およびステアリングホイール上のハードウェア入力制御20および22は、英数字を入力するため、またはナビゲーション指示を要求するための他の機能を実行するために使用されることが可能である。ヘッドユニット14は、例えば、マイクロフォン24およびスピーカ26などの、音声入力および出力構成要素を含むこともできる。スピーカ26は、携帯デバイス10から送信された音声指示を再生するために使用されることが可能である。

10

【 0 0 2 0 】

コンテンツ認識音声ナビゲーション生成システムが実装されることが可能な例示的な通信システム100が図2に示されている。通信システム100は、「マッピングアプリケーション122」と呼ばれる場合もある地理アプリケーション122を実行するように構成されたクライアントデバイス10を含む。実装形態に応じて、アプリケーション122は、対話型デジタル地図を表示し、運転、徒歩、または音声ナビゲーション指示を含む他のナビゲーション指示を提供するためにルーティングデータを要求および受信し、様々な地理位置コンテンツなどを提供することができる。クライアントデバイス10は、様々な場所にナビゲートしながらデジタル地図を表示するユーザ(本明細書では「運転者」とも呼ばれる)によって操作され得る。通信システム100は、Bluetooth、Wi-Fi Directなどの短距離通信リンクを介してクライアントデバイス10と通信し得る車両ヘッドユニット14も含む。さらに、通信システム100は、Bluetooth、Wi-Fi Directなどの短距離通信リンクを介してクライアントデバイス10と通信し得る、クライアントデバイス10の近傍内の他のコンピューティングデバイス92を含み得る。例えば、クライアントデバイス10が運転者のスマートフォンである場合、他のコンピューティングデバイス92は、車両内の同乗者のスマートフォン、または運転者のタブレットもしくはウェアラブルデバイスを含み得る。

20

【 0 0 2 1 】

クライアントデバイス10に加えて、通信システム100は、コンテンツ認識音声ナビゲーション指示をクライアントデバイス10に提供するように構成されたサーバデバイス60を含む。サーバデバイス60は、例示的な実装形態において、特定のメディア/音声コンテンツに対して音声ナビゲーション指示を調整するための機械学習モデルを記憶するデータベース80に通信可能に結合されることが可能である。トレーニングデータは、ユーザに以前に提供された音声ナビゲーション指示のセット、音声ナビゲーション指示が提供されたときに再生されたメディアコンテンツの特性、音声ナビゲーション指示に対してユーザによって行われた調整の指標、および/または音声ナビゲーション指示に対するユーザの満足度に関する自己申告の指標を含み得る。トレーニングデータについて、図3を参照して以下でさらに詳細に説明する。それに加えて、データベース80は、音声ナビゲーションパラメータを調整するための規則のセットを記憶し得る。

30

【 0 0 2 2 】

より一般的に、サーバデバイス60は、任意のタイプの適切な地理空間情報、または地理的コンテキストにリンクされることが可能な情報を記憶する1つまたはいくつかのデータベースと通信することができる。通信システム100は、例えば、運転、徒歩、自転車、または公共交通機関の指示を提供するナビゲーションデータサーバ34を含むこともできる。さらに、通信システム100は、地図表示を生成するために地図データをサーバデバイス60に提供する地図データサーバ50を含むことができる。通信システム100内で動作するデバイスは、通信ネットワーク30を介して相互接続されることが可能である。

40

【 0 0 2 3 】

様々な実装形態において、クライアントデバイス10は、スマートフォンまたはタブレットコンピュータであり得る。クライアントデバイス10は、メモリ120と、1つまたは複数のプロセッサ(CPU)116と、グラフィック処理ユニット(GPU)112と、マイクロフォンと

50

スピーカとを含むI/Oモジュール14と、ユーザインターフェース(UI)32と、全地球測位サービス(GPU)モジュールを含む1つまたはいくつかのセンサ19とを含み得る。メモリ120は、非一時的メモリとすることができ、ランダムアクセスメモリ(RAM)、読み取り専用メモリ(ROM)、フラッシュメモリ、他のタイプの永続メモリなどの、1つまたはいくつかの適切なメモリモジュールを含むことができる。I/Oモジュール114は、例えば、タッチスクリーンであり得る。様々な実装形態において、クライアントデバイス10は、図2に示すよりも少ない構成要素を含むことができ、または逆に、追加の構成要素を含むことができる。他の実装形態において、クライアントデバイス10は、任意の適切な携帯型または非携帯型のコンピューティングデバイスであり得る。例えば、クライアントデバイス10は、ラップトップコンピュータ、デスクトップコンピュータ、スマートウォッチまたはスマート眼鏡などのウェアラブルデバイスなどであり得る。

10

【0024】

メモリ120は、任意のタイプの適切なモバイルまたは汎用オペレーティングシステムであることが可能なオペレーティングシステム(OS)126を記憶する。OS126は、アプリケーションがセンサ読み取り値を取得することを可能にするアプリケーションプログラミングインターフェース(API)機能を含むことができる。例えば、コンピューティングデバイス10上で実行するように構成されたソフトウェアアプリケーションは、その時点におけるクライアントデバイス10の現在位置を取得するためにOS126のAPIを呼び出す指示を含むことができる。APIは、APIが推定値をどの程度確信しているかの定量的な指標(例えば、パーセンテージ)を返すこともできる。

20

【0025】

メモリ120は、上記で示したように、対話型デジタル地図を生成するおよび/または他の地理的機能を実行するように構成されたマッピングアプリケーション122も記憶する。マッピングアプリケーション122は、ナビゲーション指示、音声ナビゲーション指示、および/または音声ナビゲーション指示パラメータを受信し、音声ナビゲーション指示パラメータに従って音声ナビゲーション指示を提示することができる。マッピングアプリケーション122は、運転、徒歩、自転車、または公共交通機関の指示を表示することもでき、一般に、地理、ジオロケーション、ナビゲーションなどに関連する機能を提供することもできる。

【0026】

図2は、マッピングアプリケーション122をスタンドアロンアプリケーションとして示しているが、マッピングアプリケーション122の機能はまた、クライアントデバイス10上で実行されるウェブブラウザを介してアクセス可能なオンラインサービスの形態において、クライアントデバイス10上で実行される別のソフトウェアアプリケーションのためのプラグインまたは拡張などとして提供されることが可能であることに留意されたい。マッピングアプリケーション122は、異なるそれぞれのオペレーティングシステムに対して異なるバージョンにおいて提供されることが可能である。例えば、クライアントデバイス10のメーカーは、Android(商標)プラットフォームのためのマッピングアプリケーション122を含むソフトウェア開発キット(SDK)、iOS(商標)プラットフォームのための別のSDKなどを提供することができる。

30

【0027】

マッピングアプリケーション122に加えて、メモリ120は、メディア/音声コンテンツを再生する、音楽アプリケーション、ビデオアプリケーション、ストリーミングアプリケーション、ラジオアプリケーション、ソーシャルメディアアプリケーションなどの他のクライアントアプリケーション132を記憶する。これらのアプリケーション132は、マッピングアプリケーション122と通信するためのAPIを公開し得る。

【0028】

いくつかの実装形態において、サーバデバイス60は、1つまたは複数のプロセッサ62と、メモリ64とを含む。メモリ64は、有形の非一時的メモリであり得、ランダムアクセスメモリ(RAM)、読み取り専用メモリ(ROM)、フラッシュメモリ、他のタイプの永続メモリ

50

などを含む、任意のタイプの適切なメモリモジュールを含み得る。メモリ64は、クライアントデバイス10の近傍内で提示されるメディア/音声コンテンツの特性を取得し、メディア/音声コンテンツの特性に従って音声ナビゲーション指示のセット内の音声ナビゲーション指示を調整することができるコンテンツ認識音声ナビゲーション生成器68を構成する、プロセッサ62上で実行可能な指示を記憶する。いくつかの実装形態において、コンテンツ認識音声ナビゲーション生成器68は、特定のメディア/音声コンテンツに対して音声ナビゲーション指示を調整するための機械学習モデルを生成し得る。コンテンツ認識音声ナビゲーション生成器68はまた、出発地から目的地までのユーザのためのナビゲーション指示の要求を受信し得る。次いで、コンテンツ認識音声ナビゲーション生成器68は、次いでスピーカ26を介してマッピングアプリケーション122によって提示される音声ナビゲーション指示のセットを取得し、クライアントデバイス10に提供し得る。

10

【0029】

コンテンツ認識音声ナビゲーション生成器68およびマッピングアプリケーション122は、コンテンツ認識音声ナビゲーション生成システムの構成要素として動作することができる。代替的に、コンテンツ認識音声ナビゲーション生成システムは、サーバ側構成要素のみを含み、音声ナビゲーション指示を提示するための指示をマッピングアプリケーション122に単に提供することができる。言い換えれば、これらの実施形態におけるコンテンツ認識音声ナビゲーション生成技法は、マッピングアプリケーション122に透過的に実装されることが可能である。別の代替案として、コンテンツ認識音声ナビゲーション生成器68の全機能は、マッピングアプリケーション122において実装されることが可能である。

20

【0030】

簡単にするために、図2は、サーバデバイス60をサーバの唯一のインスタンスとして示す。しかしながら、サーバデバイス60は、各々が1つまたは複数のプロセッサを備え、他のサーバデバイスと独立して動作することができる1つまたは複数のサーバデバイスのグループを含む。そのようなグループ内で動作するサーバデバイスは、クライアントデバイス10からの要求を個別に(例えば、可用性に基づいて)、要求を処理することに関連する1つの動作が1つのサーバデバイス上で実行され、同じ要求を処理することに関連する別の動作が別のサーバデバイス上で実行される分散方式において、または任意の他の適切な技法に従って処理することができる。この議論の目的のために、「サーバデバイス」という用語は、個々のサーバデバイス、または2つ以上のサーバデバイスのグループを指す場合がある。

30

【0031】

動作中、クライアントデバイス10内で動作するマッピングアプリケーション122は、データを受信し、サーバデバイス60に送信する。したがって、一例において、クライアントデバイス10は、出発地から目的地までのナビゲーション指示を要求する通信を、(サーバデバイス60内に実装された)コンテンツ認識音声ナビゲーション生成器68に送信し得る。したがって、コンテンツ認識音声ナビゲーション生成器68は、音声ナビゲーション指示のセットを生成し得る。次いで、次の操縦のための音声ナビゲーション指示を再生する前に、クライアントデバイス10は、クライアントデバイス10の近傍内で再生されているメディア/音声コンテンツの特性をコンテンツ認識音声ナビゲーション生成器68に送信し得、および/または音声ナビゲーション指示の指標を送信し得る。コンテンツ認識音声ナビゲーション生成器68は、メディア/音声コンテンツに基づいて音声ナビゲーション指示をどのように調整するかを決定し得、調整された音声ナビゲーション指示、または音声ナビゲーション指示を調整するためのデータをクライアントデバイス10に送信し得る。他の実装形態において、クライアントデバイス10は、コンテンツ認識音声ナビゲーション生成器68が音声ナビゲーション指示のセットを生成する前に、クライアントデバイス10の近傍内で再生されているメディア/音声コンテンツの特性をコンテンツ認識音声ナビゲーション生成器68に送信し得る。次いで、コンテンツ認識音声ナビゲーション生成器68は、メディア/音声コンテンツの特性を考慮して、音声ナビゲーション指示のセットを生成し得る。

40

【0032】

50

いくつかの実施形態において、コンテンツ認識音声ナビゲーション生成器68は、各々の調整された音声ナビゲーション指示を生成し、クライアントデバイス10に個別に提供する。他の実施形態において、コンテンツ認識音声ナビゲーション生成器68は、テキストベースのナビゲーション指示のセットを生成し、クライアントデバイス10に提供する。次いで、各テキストベースのナビゲーション指示について、コンテンツ認識音声ナビゲーション生成器68は、クライアントデバイス10の近傍内で再生されているメディア/音声コンテンツの特性を受信し、テキストベースのナビゲーション指示に関する音声ナビゲーション指示パラメータを生成し、クライアントデバイス10に提供する。結果として、クライアントデバイス10は、音声ナビゲーション指示パラメータに従って、音声ナビゲーション指示を生成して提示する。音声ナビゲーション指示パラメータは、音声ナビゲーション指示の言語、音声ナビゲーション指示の発話速度、音声ナビゲーション指示を提示するタイミングなどを含み得る。

10

【0033】

また、いくつかの実施形態において、コンテンツ認識音声ナビゲーション生成器68は、同じナビゲーション指示の複数のバージョン(例えば、英語におけるナビゲーション指示の第1のバージョン、スペイン語におけるナビゲーション指示の第2のバージョン、ユーザに1000メートル先で左折するように命じるナビゲーション指示の第3のバージョン、メディアコンテンツのセグメントまたは文の終わりまで指示が遅れた場合に、ユーザに500メートル先で左折するように命じるナビゲーション指示の第4のバージョンなど)を含むテキストベースのナビゲーション指示のセットを生成し、クライアントデバイス10に提供する。次いで、各ナビゲーション指示について、コンテンツ認識音声ナビゲーション生成器68は、ナビゲーション指示に関する音声ナビゲーション指示パラメータを生成し、クライアントデバイス10に提供する。結果として、クライアントデバイス10は、音声ナビゲーション指示パラメータに対応するナビゲーション指示のバージョンを提示する。

20

【0034】

コンテンツ認識音声ナビゲーション生成器68は、メディアコンテンツの言語に基づいて残りのナビゲーション指示の各々について言語を調整するなど、メディアコンテンツに基づいて複数のナビゲーション指示に関する音声ナビゲーションパラメータを調整し得る。それに加えて、コンテンツ認識音声ナビゲーション生成器68は、メディアコンテンツ内の特定の文またはセグメント後に個々のナビゲーション指示を再生するように個々のナビゲーション指示のタイミングを調整するなど、メディアコンテンツに基づいて個々のナビゲーション指示に関する音声ナビゲーションパラメータを調整し得る。

30

【0035】

コンテンツ認識音声ナビゲーション生成器68は、事前に記憶された規則のセットに基づいて音声ナビゲーションパラメータを調整し得る。例えば、事前に記憶された規則は、メディアコンテンツの少なくとも1つのセグメントの言語、メディアコンテンツの少なくともしきい値数のセグメント、またはメディアコンテンツの少なくともしきい値量の時間に基づいて、音声ナビゲーション指示の言語を調整する第1の規則を含み得る。音声ナビゲーション指示の言語をメディアコンテンツの言語と一致するように調整することによって、音声ナビゲーション指示のユーザの理解が改善され、それによって、ユーザの安全性を改善する。事前に記憶された規則は、メディアコンテンツの少なくとも1つのセグメントの発話速度、メディアコンテンツの少なくともしきい値数のセグメント、またはメディアコンテンツの少なくともしきい値量の時間に基づいて、音声ナビゲーション指示の発話速度を調整する第2の規則も含み得る。再び、音声ナビゲーション指示の発話速度をメディアコンテンツの発話速度とより密接に整合するように調整することによって、ユーザの理解および安全性が改善される。

40

【0036】

さらに、事前に記憶された規則は、音声ナビゲーション指示がメディアコンテンツの文またはセグメントの終了後に再生されるように、音声ナビゲーション指示のタイミングを調整する第3の規則を含み得る。メディアコンテンツの文またはセグメントの終了後に音

50

声ナビゲーション指示を再生することによって、音声ナビゲーション指示は、コンテンツの流れが突然途切れることがないので、より容易に理解される。したがって、これも、ユーザの理解と安全性とを改善する。いくつかの実装形態において、事前に記憶された規則は、音声ナビゲーション指示の複雑さのレベル、操縦のための場所に到着する前に音声ナビゲーション指示が再生されなければならない緊急度レベル、およびメディアコンテンツの関連性レベルを決定し得る。

【0037】

例えば、メディアコンテンツの関連性レベルは、メディアコンテンツが音楽である場合は低くなり得、メディアコンテンツが、ユーザが運転している都市の音声ツアーである場合、またはメディアコンテンツが、ユーザが目的地に到着したときに実行するように計画しているタスクを実行する方法に関する説明ビデオである場合、高くなり得る。事前に記憶された規則は、音声ナビゲーション指示を再生するためにメディアコンテンツをいつ、どのように中断するかを決定するために、音声ナビゲーション指示の複雑さのレベル、操縦のための場所に到着する前に音声ナビゲーション指示が再生されなければならない緊急度レベル、および/またはメディアコンテンツの関連性レベルを比較し得る。

10

【0038】

例えば、音声ナビゲーション指示が再生されなければならない緊急度レベルが高い場合、事前に記憶された規則は、メディアコンテンツの関連性レベルが高い場合でも、メディアコンテンツを中断することを示し得る。これは、必要に応じて緊急のナビゲーション指示が依然として再生されることを保証する。一方、音声ナビゲーション指示が再生されなければならない緊急度レベルが中程度または低く、メディアコンテンツの関連性レベルが高い場合、事前に記憶された規則は、メディアコンテンツ内の文またはセグメントの終わりまで、音声ナビゲーション指示を再生するのを待つことを示し得、または音声ナビゲーション指示をメディアコンテンツと同時に再生しながら、メディアコンテンツの音量を下げることを示し得、もしくは音声ナビゲーション指示の音量を上げることを示し得る。別の例において、複雑さのレベルが高く、関連性レベルが高い場合、事前に記憶された規則は、1回目は第1の文またはセグメントの終了後、2回目は第2の文またはセグメントの終了後に、音声ナビゲーション指示を複数回再生することを示し得る。メディアコンテンツを不自然に中断することを回避するために特定のインスタンスにおいて複雑な指示を繰り返し再生することは、ユーザが明確なポイントで複数回指示を提供されるので、ユーザの安全性を改善する。さらに別の例において、複雑さのレベルが関連性レベルを超える場合、事前に記憶された規則は、音声ナビゲーション指示をすぐに再生することを示し得る。一方、関連性レベルが複雑さのレベルを超える場合、事前に記憶された規則は、音声ナビゲーション指示を再生しないこと、または文もしくはセグメントの終わりまで音声ナビゲーション指示を再生するのを待つことを示し得る。

20

30

【0039】

メディアコンテンツの文またはセグメントの終了後に音声ナビゲーション指示を再生するために、クライアントデバイス10は、メディアコンテンツ内の特定のマーカ(例えば、文の終わり、セグメントの終わりなど)を識別するために、以下でより詳細に説明するように、音声特性分類および/または自然言語処理モデルを使用して、メディアコンテンツの現在の音声ストリームを分析する。クライアントデバイス10は、音声特性分類および/または自然言語処理モデルを使用して文またはセグメントの終わりが識別されるまで、リアルタイムまたはほぼリアルタイムで現在の音声ストリームを分析し続け得る。次いで、文またはセグメントの終わりが識別されると、クライアントデバイス10は、音声ナビゲーション指示を再生し得る。

40

【0040】

音声ナビゲーション指示はまた、音声ナビゲーション指示を再生するために文またはセグメントの終わりまで待つ際の時間遅延を考慮して調整され得る。例えば、最初に音声ナビゲーション指示は、「100m先、左折してください」であり得る。文の終了後、ユーザは、操縦のための場所から50mである場合があり、したがって、音声ナビゲーション指示

50

は、「50m先、左折してください」に調整され得る。時間遅延を考慮して音声ナビゲーション指示を調整するために、音声ナビゲーション指示は、最初に、「Xメートル先、左折してください」などの可変距離で生成され得る。マッピングアプリケーション122は、音声ナビゲーション指示が提示された時点における距離を、ユーザの現在位置と操縦のための場所との間の距離として決定し得、「X」を決定された距離に置き換え得、または決定された距離を切り上げるか、もしくは最も近い50メートル、100メートル、500メートルなどに丸め得る。これを行うことによって、音声ナビゲーション指示の出力の遅延により、音声ナビゲーション指示の精度は、損なわれない。このように、本明細書で説明する方法は、音声ナビゲーション指示の精度を損なうことなく、音声ナビゲーション指示でメディアコンテンツを不自然に中断することを回避する。これは、音声ナビゲーション指示の理解を改善し、ユーザの安全性を改善する。

10

【0041】

それに加えて、クライアントデバイス10は、電子メディアコンテンツが1つまたは複数の音声ナビゲーション指示に従って提示される方法を調整し得る。より具体的には、クライアントデバイス10は、例えば、メディアコンテンツがクライアントデバイス10上で実行されているアプリケーションから再生されている場合、(例えば、APIを介してアプリケーションと通信することによって)メディアコンテンツを一時停止するか、またはメディアコンテンツの音量を調整し得る。別の例において、メディアコンテンツが(例えば、短距離通信リンクを介して)クライアントデバイス10に通信可能に結合された別のデバイスから再生されている場合、クライアントデバイス10は、他のデバイス上で再生されているメディアコンテンツを一時停止するか、またはメディアコンテンツの音量を調整する要求を他のデバイスに送信し得る。これは、他のデバイスにメディアコンテンツを一時停止させるか、またはメディアコンテンツの音量を調整させ得る。メディアコンテンツの一時停止は、メディアコンテンツと音声ナビゲーション指示の両方が同時に再生されるのを有利に回避し、それによって、音声ナビゲーション指示の明瞭さがメディアコンテンツによって影響されないことを保証する。同様に、例えば、メディアコンテンツの音量を下げることは、ユーザが音声ナビゲーション指示を聞き、理解するのを助ける。したがって、これらの概念の両方が、ユーザの安全性を改善する。

20

【0042】

依然としてさらに、事前に記憶された規則は、メディアコンテンツ内に含まれるPOIまたはメディアコンテンツにおいて議論される他の地理的トピック(例えば、古代ローマ)を識別し、POIがルートに沿ったウェイポイントのしきい値半径内にあるかどうかを判定する第4の規則を含み得る。POIがルートに沿ったウェイポイントのしきい値半径内にある場合、事前に記憶された規則は、「興味がある場合、左側にローマの遺跡があります(If you're interested, on the left is a Roman archaeological site)」などのPOIに関する情報を再生することを示し得る。それに加えてまたは代替的に、事前に記憶された規則は、「Joe's Coffeeは、約1マイル先の左側にあります。そこへのナビゲーション指示はいかがですか(Joe's Coffee is about a mile down to the left. Would you like navigation directions there?)」などの、POIにナビゲートする提案をユーザに提供することを示し得る。次いで、ユーザがJoe's Coffeeへのナビゲーション指示を希望することを示す場合、コンテンツ認識音声ナビゲーション生成器68は、ユーザの現在位置からJoe's Coffeeまでの音声ナビゲーション指示のセットを生成し得る。

30

40

【0043】

さらに、事前に記憶された規則は、例えば、メディアコンテンツが音声ブックまたはポッドキャストである場合、メディアコンテンツの話者の声の特性を決定する第5の規則を含み得る。次いで、事前に記憶された規則は、ユーザが、音声ナビゲーション指示が提示されたときに注意を促されるように、メディアコンテンツの話者と区別可能な音声ナビゲーション指示を再生するための声を識別し得る。事前に記憶された規則は、話者の声の特性(例えば、声のピッチ、トーン、および周波数)を音声記録の各々の特性と比較することによって、所定の音声記録のセットから音声記録を選択し、話者の声と最も似ていない音

50

声記録を識別し得る。メディアコンテンツの話者の声と区別可能な音声ナビゲーション指示を再生するための声を識別することによって、音声ナビゲーション指示は、ユーザによってより容易に理解される。

【0044】

事前に記憶された規則のセットに基づいて音声ナビゲーションパラメータを調整することに加えて、またはその代替として、コンテンツ認識音声ナビゲーション生成器68は、特定のメディア/音声コンテンツに対して音声ナビゲーション指示を調整するための機械学習モデルを生成し得る。機械学習モデルを生成するために、コンテンツ認識音声ナビゲーション生成器68は、ユーザに以前に提供された音声ナビゲーション指示のセット、音声ナビゲーション指示が提供されたときに再生されたメディアコンテンツの特性、音声ナビゲーション指示に対してユーザによって行われた調整の指標、および/または音声ナビゲーション指示に対するユーザの満足度に関する自己申告の指標を含むトレーニングデータを取得する。

10

【0045】

例えば、位置データおよび/または他のユーザデータを共有するオプションを選択するユーザは、音声ナビゲーション指示が提示されたときに再生されたメディアコンテンツの特性とともに、それぞれのクライアントデバイス10によって提示された音声ナビゲーション指示のセットを送信し得る。

【0046】

メディアコンテンツの特性は、メディアコンテンツの言語、メディアコンテンツの発話速度、メディアコンテンツの話者の声の特性、メディアコンテンツの転写、メディアコンテンツの長さ、メディアコンテンツのタイトル、メディアコンテンツのタイプ(例えば、ポッドキャスト、音声ブック、歌など)、メディアコンテンツ内で言及されるPOIまたは他の地理的トピックなどを含み得る。いくつかの実装形態において、クライアントデバイス10は、メディアコンテンツのソースと(例えば、短距離通信リンクまたはAPIを介して)通信することによって、メディアコンテンツの特性を決定する。他の実装形態において、クライアントデバイス10は、エリア内の周囲の音声フィンガープリントを分析することによって、メディアコンテンツの特性を決定する。

20

【0047】

例えば、ユーザは、周囲音声電子メディアコンテンツとして識別される場合、マッピングアプリケーション122がエリア内の周囲音声进行分析することを可能にするオプションを選択し得る。クライアントデバイス10、およびより具体的にマッピングアプリケーション122は、クライアントデバイス10またはクライアントデバイス10の通信範囲内にある他のコンピューティングデバイス92上で実行されている他のアプリケーション132と通信することによって、メディアコンテンツがマッピングアプリケーションとは別のソースから再生されていることを最初に識別し得る。マッピングアプリケーション122はまた、所定のメディア/音声コンテンツの音声フィンガープリントを周囲エリア内の周囲音声フィンガープリントと比較することによって、メディアコンテンツが再生されていることを識別し得る。

30

【0048】

例えば、マッピングアプリケーション122は、ポピュラー音楽、音声ブック、ラジオ番組、ポッドキャストなどから、メディア/音声コンテンツに対応する音声フィンガープリントのライブラリを取得し得る。マッピングアプリケーション122は、周囲音声からフィンガープリントを抽出し、周囲音声フィンガープリントの特徴を識別し、例えば、機械学習技法を使用して、周囲音声フィンガープリントの特徴を所定のメディア/音声コンテンツからの音声フィンガープリントの特徴と比較し得る。機械学習技法は、線形回帰、多項式回帰、ロジスティック回帰、ランダムフォレスト、ブースティング、最近傍、ベイジアンネットワーク、ニューラルネットワーク、サポートベクターマシン、または任意の他の適切な機械学習技法を含み得る。例えば、周波数、ピッチ、トーン、振幅などが、音声フィンガープリント特徴として記憶され得る。特徴は、歌全体、ポッドキャスト、音声ブックな

40

50

ど、またはメディアコンテンツの個々のセグメント(例えば、8秒ごとのセグメント)について識別され得る。次いで、所定のメディア/音声コンテンツに関するこれらの音声フィンガープリントの各々が、周囲音声フィンガープリントの特徴と比較され得る。

【0049】

いくつかの実施形態において、所定のメディア/音声コンテンツの音声フィンガープリント特徴は、最近傍アルゴリズムを使用して周囲音声フィンガープリントの特徴と比較され得る。最近傍アルゴリズムは、周囲音声フィンガープリントの特徴に最も近い所定のメディア/音声コンテンツの音声フィンガープリント特徴を識別し得る。次いで、マッピングアプリケーション122は、周囲音声フィンガープリント特徴が、所定のメディア/音声コンテンツ内の歌、音声ブック、ラジオ番組、ポッドキャストなどのうちの1つの音声フィンガープリント特徴と一致するか、またはしきい値量よりも多くの類似性を有する場合、周囲音声電子メディアコンテンツを含むと判断し得る。マッピングアプリケーション122はまた、周囲音声フィンガープリントと一致するか、またはしきい値量よりも多くの類似性を有する、所定のメディア/音声コンテンツ内の特定の歌、音声ブック、ラジオ番組、ポッドキャストなどであると判断し得る。

10

【0050】

次いで、マッピングアプリケーション122は、所定のメディア/音声コンテンツからの一致する歌、音声ブック、ラジオ番組、ポッドキャストなどの特性に基づいて、周囲音声の特性を識別し得る。マッピングアプリケーション122が、音声フィンガープリント技法を使用して、所定のメディア/音声コンテンツからの一致する歌、音声ブック、ラジオ番組、ポッドキャストなどを識別しない場合などの他の実装形態において、マッピングアプリケーション122は、音声認識、話者識別、音声特性分類、および自然言語処理モデルを含むコンテンツ分析モデルを、メディアコンテンツの音声ストリームに対して実行することができる。これらのモデルは、話者、POIまたは他の地理的トピックなどの音声ストリームにおいて言及されている用語、メディアコンテンツ内の特定のマーカ(例えば、文の終わり、セグメントの終わりなど)、メディアコンテンツの言語、メディアコンテンツの発話速度などを識別するために、音声ストリームの特性を分析し、特性をトレーニングデータと比較するニューラルネットワークまたは任意の他の適切なタイプの機械学習モデルを含み得る。

20

【0051】

例えば、モデルは、いくつかの話者からの音声ストリームを含むトレーニングデータを使用してトレーニングされ得、音声ストリームの特性(例えば、ピッチ、トーン、周波数、振幅など)は、それぞれの話者に従って分類される。マッピングアプリケーション122は、音声ストリームの特性を識別するためにメディアコンテンツの音声ストリームを分析し、同様の特性を有するいくつかの話者から話者を識別するために特性をモデルに適用し得る。

30

【0052】

別の例において、モデルは、いくつかの言語における音声ストリームを含むトレーニングデータを使用してトレーニングされ得る。マッピングアプリケーション122は、音声ストリームの特性を識別するためにメディアコンテンツの音声ストリームを分析し、同様の特性を有するいくつかの言語から言語を識別するために特性をモデルに適用し得る。

40

【0053】

さらに別の例において、モデルは、特定の用語、句、文などが話され、音声ストリームの特性がそれぞれの用語、句、文などに従って分類される、音声ストリームを含むトレーニングデータを使用してトレーニングされ得る。マッピングアプリケーション122は、音声ストリームの特性を識別するためにメディアコンテンツの音声ストリームを分析し、同様の特性を有するいくつかの用語、句、文などから用語、句、文などを識別するために特性をモデルに適用し得る。

【0054】

いずれにしても、提示された各音声ナビゲーション指示または音声ナビゲーション指示のセットについて、コンテンツ認識音声ナビゲーション生成器68は、音声ナビゲーション

50

指示が提示されたときに再生されたメディアコンテンツの特性を取得し得る。それに加えて、コンテンツ認識音声ナビゲーション生成器68は、言語を変更すること、発話速度を変更すること、音量を変更すること、音声ナビゲーション指示をミュートすること、音声ナビゲーション指示が繰り返されることを要求することなどの、音声ナビゲーション指示に対してユーザによって行われた調整の指標を取得し得る。また、いくつかの実施形態において、コンテンツ認識音声ナビゲーション生成器68は、追加のトレーニングデータとして、音声ナビゲーション指示に対するユーザの満足度に関する自己申告の指標を取得し得る。例えば、マッピングアプリケーション122は、音声ナビゲーション指示に対する不満を示すためのユーザ制御、および/またはユーザが自分の不満を説明するためのテキストフィールドを含み得る。このようにして、機械学習モデルは、ユーザが従う可能性がより高く、ユーザを悩ませるまたは邪魔する可能性がより低い音声ナビゲーション指示を生成することができる。

10

【0055】

音声ナビゲーション指示のセット、音声ナビゲーション指示パラメータ、メディアコンテンツ特性、音声ナビゲーション指示パラメータに対するユーザ調整、および/またはユーザが音声ナビゲーション指示に満足したかどうかの指標は、機械学習技法を使用して機械学習モデルを生成するためのトレーニングデータとして提供され得る。いくつかの実施形態において、各音声ナビゲーション指示パラメータに対して、別個の機械学習モデルが生成され得る。例えば、1つの機械学習モデルが、音声ナビゲーション指示の言語を決定するために生成され得る。別の機械学習モデルが、音声ナビゲーション指示を提供するタイミングを決定するために生成され得る。さらに別の機械学習モデルが、音声ナビゲーション指示の発話速度を決定するために生成され得る。

20

【0056】

図3は、特定のメディア/音声コンテンツに対して音声ナビゲーション指示を調整するために使用され得る例示的なデータ300を示す。いくつかの実施形態において、コンテンツ認識音声ナビゲーション生成器68は、音声ナビゲーション指示を調整するために、データ300を事前に記憶された規則のセットに適用し得る。データ300は、メディア/音声コンテンツ特性310と、音声ナビゲーション指示パラメータ320とを含み得る。メディア/音声コンテンツ特性310は、メディアコンテンツの転写、メディアコンテンツの発話速度、メディアコンテンツの言語、メディアコンテンツ内に含まれるPOI、メディアコンテンツの長さ、メディアコンテンツの関連性レベル、メディアコンテンツのタイトル、メディアコンテンツのタイプ、メディアコンテンツの話者の声の特性などを含み得る。

30

【0057】

メディアコンテンツの転写は、メディアコンテンツのテキスト、ならびに文の終わり、セグメントの終わり、次の文の始まり、次のセグメントの始まりなどの、メディアコンテンツ内の一時停止がある時点の指標を含み得る。

【0058】

メディアコンテンツの関連性レベルは、1から100などの関連性スコアであり得、「非常に低い」、「低い」、「中位」、「高い」、「非常に高い」などのカテゴリであり得、または任意の他の適切な方法において示され得る。メディアコンテンツの関連性レベルは、メディアコンテンツの主題、ユーザの現在位置、および/または音声ナビゲーション指示の目的地に基づいて決定され得る。例えば、関連性レベルは、メディアコンテンツが、ユーザが運転している都市の音声ツアーである場合、またはメディアコンテンツが、ユーザが目的地において実行し得るタスクを実行する方法に関する説明ビデオである場合、高くなり得る。

40

【0059】

メディアコンテンツのタイプは、ラジオ番組、ポッドキャスト、音声ブック、音楽、広告、テレビ番組、映画、ビデオなどを含み得る。話者の声の特性は、話者の声のピッチ、トーン、周波数、または任意の他の適切な特性を含み得る。

【0060】

50

例示的なメディア/音声コンテンツ特性310は、メディアコンテンツの転写、メディアコンテンツの発話速度、メディアコンテンツの言語、メディアコンテンツ内に含まれるPOI、メディアコンテンツの長さ、メディアコンテンツの関連性レベル、メディアコンテンツのタイトル、メディアコンテンツのタイプ、メディアコンテンツの話者の声の特性などを含み得、これらは、説明を容易にするだけのためのメディア/音声コンテンツ310のほんの数例に過ぎない。任意の適切なメディア/音声コンテンツ特性310が使用され得、説明した特性310のサブセットが使用され得る。

【0061】

音声ナビゲーション指示パラメータ320は、操縦のタイプ、操縦の場所、操縦の複雑さのレベル、音声ナビゲーション指示を再生するための緊急度レベル、音声ナビゲーション指示の言語、音声ナビゲーション指示の発話速度、音声ナビゲーション指示のための声、音声ナビゲーション指示のタイミング、および/または音声ナビゲーション指示の音量を含み得る。

10

【0062】

操縦の複雑さのレベルは、1から100などの複雑さスコアであり得、「非常に低い」、「低い」、「中位」、「高い」、「非常に高い」などのカテゴリであり得、または任意の他の適切な方法において示され得る。操縦の複雑さのレベルは、四差路交差点の曲がり角、六差路交差点の曲がり角、ラウンドアバウト、Uターン、高速道路の合流、高速道路の出口などの操縦タイプに基づいて決定され得る。複雑さのレベルはまた、次の操縦と前の操縦との間の時間または距離の量に基づいて決定され得る。前の操縦の直後に発生する操縦は、より高い複雑さのレベルを有し得る。さらに、複雑さのレベルは、ユーザが操縦を実行するために変更する必要がある車線の数に基づいて決定され得る。例えば、マッピングアプリケーション122は、前の操縦を実行した後のユーザの初期車線を、ユーザが次の操縦を実行するための最終車線と比較し得る。複雑さのレベルは、操縦を実行するための車線変更の数が増加すると上昇する場合がある。

20

【0063】

音声ナビゲーション指示を再生するための緊急度レベルは、1から100などの緊急度スコアであり得、「非常に低い」、「低い」、「中位」、「高い」、「非常に高い」などのカテゴリであり得、または任意の他の適切な方法において示され得る。操縦の緊急度レベルは、ユーザが操縦を実行しなければならないまでの距離および/または時間の量に基づいて決定され得る。例えば、緊急度スコアは、ユーザが操作を実行しなければならないまでの距離および/または時間の量に反比例し得る。例えば、緊急度スコアは、操縦がユーザの現在位置から1マイル離れている場合よりも、操縦がユーザの現在位置から200メートル離れている場合に高くなり得る。

30

【0064】

音声ナビゲーション指示のタイミングは、ユーザが前の操縦の場所と後続の操縦の場所との間の中間にいるときなど、音声ナビゲーション指示を提示するタイミングを含み得る。音量は、デシベル(dB)において示されるか、または低音量(例えば、第1のしきい値デシベル量未満)、中音量(例えば、第1のしきい値デシベル量と、第1のしきい値デシベル量よりも高い第2のしきい値デシベル量との間)、高音量(例えば、第2のしきい値デシベル量よりも上)などとしてカテゴリ化され得る。

40

【0065】

例示的な音声ナビゲーション指示パラメータ320は、操縦のタイプ、操縦の場所、操縦の複雑さのレベル、音声ナビゲーション指示を再生するための緊急度レベル、音声ナビゲーション指示の言語、音声ナビゲーション指示の発話速度、音声ナビゲーション指示のための声、音声ナビゲーション指示のタイミング、音声ナビゲーション指示の音量などを含み得るが、これらは、説明を容易にするだけのための音声ナビゲーション指示パラメータ320のほんの数例に過ぎない。任意の適切な音声ナビゲーション指示パラメータ320が使用され得、説明したパラメータ320のサブセットが使用され得る。

【0066】

50

データ300は、メディアコンテンツ特性および/または音声ナビゲーション指示パラメータに基づく音声ナビゲーション指示に対する調整の指標も含む。コンテンツ認識音声ナビゲーション生成器68は、事前に記憶された規則をメディアコンテンツ特性および/または音声ナビゲーション指示パラメータに適用することによって、音声ナビゲーション指示への調整を決定し得る。他のシナリオにおいて、調整は、ユーザ制御を選択するユーザから手動調整であり得る。

【0067】

例えば、Joe's Coffeeが広告内に含まれていることに応答して、コンテンツ認識音声ナビゲーション生成器68は、Joe's Coffeeの店舗の位置を決定し、その位置をナビゲーション指示のルートに沿ったウェイポイントと比較するために、地図データサーバ50と通信し得る。ルートに沿ったウェイポイントのうちの1つのしきい値半径内にJoe's Coffeeが存在する場合、コンテンツ認識音声ナビゲーション生成器68は、Joe's Coffeeに回り道するようにユーザへの推奨を音声ナビゲーション指示のセットに追加し得る。別の例において、ポッドキャストの発話速度が1.5倍であることに応答して、コンテンツ認識音声ナビゲーション生成器68は、音声ナビゲーション指示の発話速度を1.5倍に上昇させ得、次の音声ナビゲーション指示のタイミングを、ポッドキャスト内の文の終了後に再生するように調整し得る。さらに別の例において、メディアコンテンツがスペイン語の音楽を含むことに応答して、コンテンツ認識音声ナビゲーション生成器68は、音声ナビゲーション指示の言語をスペイン語に変更し得る。別の例において、ポッドキャストの発話速度がヘブライ語で2倍であることに応答して、コンテンツ認識音声ナビゲーション生成器68は、音声ナビゲーション指示の発話速度を2倍に上昇させ得、音声ナビゲーション指示の言語をヘブライ語に変更し得る。

【0068】

他の実施形態において、データ300は、機械学習モデルを生成するために使用されるトレーニングデータであり得る。いくつかの実施形態において、トレーニングデータ300は、データベース80内に記憶され得る。メディア/音声コンテンツ特性310および音声ナビゲーション指示パラメータ320に加えて、トレーニングデータ300は、音声ナビゲーション指示に対する運転者の応答を示すデータを含み得る。音声ナビゲーション指示に対する運転者の応答を示すデータは、言語を変更すること、発話速度を変更すること、音量を変更すること、音声ナビゲーション指示をミュートすること、音声ナビゲーション指示が繰り返されることを要求することなどの、音声ナビゲーション指示に対してユーザによって行われた調整を含み得る。音声ナビゲーション指示に対する運転者の応答を示すデータは、追加のトレーニングデータとして、音声ナビゲーション指示に対するユーザの満足度に関する自己申告の指標を含み得る。

【0069】

機械学習モデルを生成するために、コンテンツ認識音声ナビゲーション生成器68は、トレーニングデータ300のサブセットを、音声ナビゲーション指示に対する様々な調整に対応するものとして分類し得る。いくつかのサブセットは、音声ナビゲーション指示の発話速度および言語に対する調整を含むトレーニングデータセット300の最後のエントリなど、複数の調整カテゴリにおいて分類され得る。

【0070】

次いで、コンテンツ認識音声ナビゲーション生成器68は、機械学習モデルを生成するためにサブセットを分析し得る。機械学習モデルは、回帰分析(例えば、ロジスティック回帰、線形回帰、または多項式回帰)、k最近傍、決定木、ランダムフォレスト、ブースティング、ニューラルネットワーク、サポートベクターマシン、深層学習、強化学習、ベイジアンネットワークなどの様々な機械学習技法を使用して生成され得る。いくつかの実施形態において、コンテンツ認識音声ナビゲーション生成器68は、音声ナビゲーション指示の言語を決定するための第1の機械学習モデルを生成し得る。コンテンツ認識音声ナビゲーション生成器68は、音声ナビゲーション指示の発話速度を決定するための第2の機械学習モデルを生成し得る。コンテンツ認識音声ナビゲーション生成器68は、音声ナビゲーション

10

20

30

40

50

指示を提供するタイミングを決定するための第3の機械学習モデルと、音声ナビゲーション指示に関する音声記録を識別するための第4の機械学習モデルと、音声ナビゲーション指示において参照するPOIを決定するための第5の機械学習モデルとを生成し得る。

【0071】

例えば、音声ナビゲーション指示を提供するタイミングを決定するための機械学習モデルは、ブランチによって接続されたいくつかのノードを有する決定木であり得、各ノードは、メディア/音声コンテンツ特性および/または音声ナビゲーション指示パラメータに対するテスト(例えば、メディアコンテンツの関連性レベルが高いか)を表し、各ブランチは、テストの結果(例えば、イエス)を表し、各リーフは、音声ナビゲーション指示のタイミングに対する調整(例えば、すぐに再生する、メディアコンテンツ内の文の終了後に再生する、メディアコンテンツのセグメントの終了後に再生するなど)を表す。

10

【0072】

より具体的には、コンテンツ認識音声ナビゲーション生成器68は、第1のノードが音声指示の緊急度レベルが高いかどうかに対応する決定木を生成し得る。緊急度レベルが高くない場合、第1のブランチは、メディアコンテンツが10秒を超えて残っているかどうかに対応する第2のノードに接続し得る。第2のコンテンツが10秒を超えて残っている場合、第2のブランチは、関連性レベルが高いかどうかに対応する第3のノードに接続し得る。関連性レベルが高い場合、第3のブランチは、音声ナビゲーション指示がメディアコンテンツ内の文の終了後に提示されるべきであることを示し得るリーフノードに接続し得る。決定木は、1つのリーフノードと3つのブランチとを含むが、これは、説明を容易にするだけのための例に過ぎない。各決定木は、メディア/音声コンテンツ特性および/または音声ナビゲーション指示パラメータに対する任意の適切な数および/またはタイプのテストを有する、任意の数のノード、ブランチ、およびリーフを含み得る。

20

【0073】

いずれにしても、コンテンツ認識音声ナビゲーション生成器68は、上記で説明したように、事前に記憶された規則または機械学習技法を使用して、特定のメディア/音声コンテンツに対する音声ナビゲーション指示の調整を識別し得る。場合によっては、次いで、コンテンツ認識音声ナビゲーション生成器68は、調整された音声ナビゲーション指示を生成し得る。説明したように、そのような変更は、競合を回避するか、または他の方法で他のメディアコンテンツと区別することによって、音声ナビゲーション指示の明瞭さを改善する。例えば、コンテンツ認識音声ナビゲーション生成器68が、言語を英語からスペイン語に変更すると決定した場合、コンテンツ認識音声ナビゲーション生成器68は、出発地から目的地までのスペイン語の音声ナビゲーション指示のセットを取得し、スペイン語の音声ナビゲーション指示をクライアントデバイス10に提供し得る。別の例において、コンテンツ認識音声ナビゲーション生成器68が、メディアコンテンツにおいて言及された特定のモジュールを見るために右を見るという追加の推奨を追加した場合、コンテンツ認識音声ナビゲーション生成器68は、追加の推奨をクライアントデバイス10に提供し得る。

30

【0074】

他の例において、コンテンツ認識音声ナビゲーション生成器68は、音声ナビゲーション指示のパラメータを調整するための指示をクライアントデバイス10に提供し得る。例えば、コンテンツ認識音声ナビゲーション生成器68が、発話速度を1.5倍に変更すると決定した場合、コンテンツ認識音声ナビゲーション生成器68は、発話速度を1.5倍に変更するための指示をクライアントデバイス10に送信し得、クライアントデバイス10は、音声ナビゲーション指示を1.5倍において再生し得る。別の例において、コンテンツ認識音声ナビゲーション生成器68が、メディアコンテンツ内の文またはセグメントの終わりなどの特定の時点において音声ナビゲーション指示を再生することを決定した場合、コンテンツ認識音声ナビゲーション生成器68は、メディアコンテンツ内の文またはセグメントの終わりまで音声ナビゲーション指示を再生するのを待つための指示を、クライアントデバイス10に送信し得る。次いで、クライアントデバイス10は、クライアントデバイス10は、文またはセグメントがいつ終わったかを判断し、次いで音声ナビゲーション指示を再生するため

40

50

に、メディアコンテンツの音声ストリームを分析し得る。前述のように、コンテンツ認識音声ナビゲーション生成器68は、サーバデバイス60またはクライアントデバイス10内に実装されることが可能である。コンテンツ認識音声ナビゲーション生成器68がクライアントデバイス10上に実装されるシナリオにおいて、コンテンツ認識音声ナビゲーション生成器68は、音声ナビゲーション指示のパラメータを調整するために、クライアントデバイス10上で実行されるテキスト読み上げ(TTS)エンジンに指示を提供し得る。

【0075】

コンテンツ認識音声ナビゲーション指示を提示するための例示的な方法

図4は、コンテンツ認識音声ナビゲーション指示を生成するための例示的な方法400のフロー図を示す。方法は、コンピュータ可読メモリ上に記憶され、クライアントデバイス10の1つまたは複数のプロセッサにおいて実行可能な指示のセットにおいて実装されることが可能である。例えば、方法のステップのうちの少なくともいくつかは、マッピングアプリケーション122によって実装されることが可能である。

10

【0076】

ブロック402において、ルートに沿って出発地から目的地までユーザをナビゲートするための音声ナビゲーション指示のセットが、マッピングアプリケーション122を介して取得される。例えば、ユーザは、マッピングアプリケーション122を介して、ユーザの現在位置から目的地までのナビゲーション指示を要求し得、クライアントデバイス10は、要求をナビゲーションサーバ34に転送し得る。次いで、ナビゲーションサーバ34は、音声ナビゲーション指示のセットを生成し、ユーザへの再生のために音声ナビゲーション指示のセットをクライアントデバイス10に送信し得る。

20

【0077】

次いで、ブロック404において、マッピングアプリケーションとは異なるソースから再生されている電子メディアコンテンツが識別される。ソースは、クライアントデバイス10上で実行される別のアプリケーション132であり得、またはクライアントデバイス10の近傍内の別のコンピューティングデバイス92であり得る。電子メディアコンテンツを識別するために、マッピングアプリケーション122は、クライアントデバイス上で実行される他のアプリケーション132と(例えば、APIを介して)、または車両ヘッドユニット14または他のクライアントデバイスなどの、クライアントデバイス10の近傍内の他のコンピューティングデバイス92と(例えば、短距離通信リンクを介して)通信し得る。通信は、他のアプリケーション/デバイス92、132によって再生されている電子メディアコンテンツの指標を含み得、電子メディアコンテンツの言語、電子メディアコンテンツの発話速度、電子メディアコンテンツの転写、電子メディアコンテンツの話者の声の特性などの、電子メディアコンテンツの特性を含み得る(ブロック406)。

30

【0078】

メディアコンテンツを識別するために他のアプリケーション/デバイスと通信することに加えて、マッピングアプリケーション122は、電子メディアコンテンツを識別するために、クライアントデバイス10のマイクロフォンによってキャプチャされている場合があるエリア内で再生されている周囲音声を、ポピュラーソング、ラジオ番組、ポッドキャストなどのポピュラーメディアコンテンツの音声フィンガープリントのセットと比較し得る。電子メディアコンテンツが識別されると、マッピングアプリケーション122は、音声認識、話者識別、音声特性分類、および自然言語処理モデルを含むコンテンツ分析モデルを、電子メディアコンテンツの音声ストリームに対して実行することによって、電子メディアコンテンツの特性を識別し得る(ブロック406)。これらのモデルは、話者、POIなどの音声ストリームにおいて言及されている用語、メディアコンテンツ内の特定のマーカ(例えば、文の終わり、セグメントの終わりなど)、メディアコンテンツの言語、メディアコンテンツの発話速度などを識別するために、音声ストリームの特性を分析し、特性をトレーニングデータと比較するニューラルネットワークまたは任意の他の適切なタイプの機械学習モデルを含み得る。

40

【0079】

50

いずれにしても、次いで、音声ナビゲーション指示のセットは、電子メディアコンテンツの特性に従って調整され得る(ブロック408)。より具体的には、いくつかの実装形態において、音声ナビゲーション指示のセットは、事前に記憶された規則を使用して調整され得る。例えば、例えば、事前に記憶された規則は、メディアコンテンツの少なくとも1つのセグメントの言語、メディアコンテンツの少なくともしきい値数のセグメント、またはメディアコンテンツの少なくともしきい値量の時間に基づいて、音声ナビゲーション指示の言語を調整する第1の規則を含み得る。事前に記憶された規則は、メディアコンテンツの少なくとも1つのセグメントの発話速度、メディアコンテンツの少なくともしきい値数のセグメント、またはメディアコンテンツの少なくともしきい値量の時間に基づいて、音声ナビゲーション指示の発話速度を調整する第2の規則も含み得る。

10

【0080】

さらに、事前に記憶された規則は、音声ナビゲーション指示がメディアコンテンツの文またはセグメントの終了後に再生されるように、音声ナビゲーション指示のタイミングを調整する第3の規則を含み得る。依然としてさらに、事前に記憶された規則は、メディアコンテンツ内に含まれるPOIを識別し、POIがルートに沿ったウェイポイントのしきい値半径内にあるかどうかを判定する第4の規則を含み得る。POIがルートに沿ったウェイポイントのしきい値半径内にある場合、事前に記憶された規則は、「興味がある場合、左側にローマの遺跡があります(If you're interested, on the left is a Roman archaeological site)」などのPOIに関する情報を再生することを示し得る。それに加えてまたは代替的に、事前に記憶された規則は、「Joe's Coffeeは、約1マイル先の左側にあります。そこへのナビゲーション指示はいかがですか(Joe's Coffee is about a mile down to the left. Would you like navigation directions there?)」などの、POIにナビゲートする提案をユーザに提供することを示し得る。

20

【0081】

さらに、事前に記憶された規則は、例えば、メディアコンテンツが音声ブックまたはポッドキャストである場合、メディアコンテンツの話者の声の特性を決定する第5の規則を含み得る。次いで、事前に記憶された規則は、ユーザが音声ナビゲーション指示が提示されたときに注意を促されるように、メディアコンテンツの話者と区別可能な音声ナビゲーション指示を再生するための声を識別し得る。

【0082】

事前に記憶された規則のセットに基づいて音声ナビゲーションパラメータを調整することに加えてまたはその代替として、マッピングアプリケーション122は、特定のメディア/音声コンテンツに対して音声ナビゲーション指示を調整するための機械学習モデルを生成し得る。次いで、マッピングアプリケーション122は、音声ナビゲーション指示に対する調整を識別するために、電子メディアコンテンツの特性および/または音声ナビゲーション指示のパラメータを機械学習モデルに適用し得る。

30

【0083】

次いで、マッピングアプリケーション122は、調整された音声ナビゲーション指示を生成し得る。例えば、マッピングアプリケーション122が言語を英語からスペイン語に変更すると決定した場合、マッピングアプリケーション122は、出発地から目的地までのスペイン語の音声ナビゲーション指示のセットを取得し得る。他の例において、マッピングアプリケーション122は、音声ナビゲーション指示のパラメータを調整するために、TTSエンジンに指示を提供し得る。

40

【0084】

他の実装形態において、電子メディアコンテンツの特性に従って音声ナビゲーション指示のセットを調整するために、クライアントデバイス10は、音声ナビゲーション指示のセット、電子メディアコンテンツの特性、および/または音声ナビゲーション指示パラメータを、サーバデバイス60に送信し得る。サーバデバイス60、およびより具体的には、コンテンツ認識音声ナビゲーション生成器68は、音声ナビゲーション指示のセットを調整するための事前に記憶された規則のセットを取得し得、または特定のメディア/音声コンテンツ

50

に対して音声ナビゲーション指示を調整するための機械学習モデルを生成し得る。コンテンツ認識音声ナビゲーション生成器68は、音声ナビゲーション指示に対する調整を識別するために、電子メディアコンテンツの特性および/または音声ナビゲーション指示パラメータを、事前に記憶された規則のセットまたは機械学習モデルに適用し得る。次いで、サーバデバイス60は、調整された音声ナビゲーション指示、または音声ナビゲーション指示の再生をどのように調整するかをクライアントデバイス10に送信し得る。有益なことに、音声ナビゲーション指示に対する調整を識別するために、クライアントデバイス10の代わりにサーバデバイス60を使用することによって、このタスクを実行するための処理、時間、および消費電力がサーバデバイス60のものになるので、クライアントデバイス10のリソースが節約される。それに加えて、サーバデバイス60の処理能力がクライアントデバイス60の処理能力を上回る場合があるので、このステップをサーバデバイス60に移すことは、特にリアルタイムシナリオにおいて、処理がより効率的に実行されることを可能にし、それによって、調整がより迅速に識別されることを可能にする。

【0085】

次いで、ブロック410において、調整された音声ナビゲーション指示は、例えば、スピーカを介してユーザに提示される。

【0086】

追加の考慮事項

以下の追加の考慮事項は、前述の議論に適用される。本明細書全体を通じて、複数のインスタンスが、単一のインスタンスとして説明した構成要素、動作、または構造を実装する場合がある。1つまたは複数の方法の個々の動作は、別個の動作として図示および説明されているが、個々の動作のうちの一つまたは複数は、同時に実行され得、動作が図示された順序において実行される必要はない。例示的な構成において別個の構成要素として提示されている構造および機能は、組み合わせられた構造または構成要素として実装され得る。同様に、単一の構成要素として提示されている構造および機能は、別個の構成要素として実装され得る。これらおよび他の変形、修正、追加、および改善は、本開示の主題の範囲内に入る。

【0087】

それに加えて、特定の実施形態は、ロジック、またはいくつかの構成要素、モジュール、もしくはメカニズムを含むものとして本明細書において説明されている。モジュールは、ソフトウェアモジュール(例えば、機械可読媒体上に記憶されたコード)またはハードウェアモジュールのいずれかを構成し得る。ハードウェアモジュールは、特定の動作を実行することができる有形のユニットであり、特定の方法において構成または配置され得る。例示的な実施形態において、一つもしくは複数のコンピュータシステム(例えば、スタンドアロン、クライアント、またはサーバコンピュータシステム)、またはコンピュータシステムの一つもしくは複数のハードウェアモジュール(例えば、プロセッサまたはプロセッサのグループ)は、ソフトウェア(例えば、アプリケーションまたはアプリケーション部分)によって、本明細書で説明する特定の動作を実行するように動作するハードウェアモジュールとして構成され得る。

【0088】

様々な実施形態において、ハードウェアモジュールは、機械的または電子的に実装され得る。例えば、ハードウェアモジュールは、特定の動作を実行するように(例えば、フィールドプログラマブルゲートアレイ(FPGA)または特定用途向け集積回路(ASIC)などの専用プロセッサとして)永続的に構成された専用回路またはロジックを備え得る。ハードウェアモジュールは、特定の動作を実行するためにソフトウェアによって一時的に構成されるプログラム可能なロジックまたは回路も(例えば、汎用プロセッサまたは他のプログラム可能なプロセッサ内に包含されるように)備え得る。ハードウェアモジュールを機械的に、専用の永続的に構成された回路において実装するか、または一時的に構成された(例えば、ソフトウェアによって構成された)回路において実装するかは、コストおよび時間の考慮によって決定され得ることが理解されるであろう。

10

20

30

40

50

【0089】

したがって、ハードウェアという用語は、特定の方法において動作するように、または本明細書で説明されている特定の動作を実行するように物理的に構成された、永続的に構成された(例えば、配線された)、または一時的に構成された(例えば、プログラムされた)エンティティである、有形のエンティティを包含すると理解されるべきである。本明細書で使用される「ハードウェア実装モジュール」は、ハードウェアモジュールを指す。ハードウェアモジュールが一時的に構成される(例えば、プログラムされる)実施形態を考慮すると、ハードウェアモジュールの各々は、いかなる一時点においても構成またはインスタンス化される必要はない。例えば、ハードウェアモジュールがソフトウェアを使用して構成された汎用プロセッサを備える場合、汎用プロセッサは、異なる時点において、それぞれ異なるハードウェアモジュールとして構成され得る。したがって、ソフトウェアは、例えば、ある時点において特定のハードウェアモジュールを構成し、異なる時点において異なるハードウェアモジュールを構成するようにプロセッサを構成し得る。

10

【0090】

ハードウェアモジュールは、他のハードウェアモジュールに情報を提供することができ、それらから情報を受信することができる。したがって、説明されているハードウェアモジュールは、通信可能に結合されていると見なされ得る。複数のそのようなハードウェアモジュールが同時に存在する場合、通信は、ハードウェアモジュールを接続する(例えば、適切な回路およびバスを介する)信号伝送を介して達成され得る。複数のハードウェアモジュールが異なる時間において構成またはインスタンス化される実施形態において、そのようなハードウェアモジュール間の通信は、例えば、複数のハードウェアモジュールがアクセスすることができるメモリ構造における情報の記憶および取得を通じて達成され得る。例えば、1つのハードウェアモジュールは、動作を実行し、その動作の出力を、通信可能に結合されたメモリデバイス内に記憶し得る。次いで、さらなるハードウェアモジュールは、後に、記憶された出力を取得および処理するために、メモリデバイスにアクセスし得る。ハードウェアモジュールはまた、入力デバイスまたは出力デバイスとの通信を開始し得、リソース(例えば、情報の集合体)に対して動作することができる。

20

【0091】

方法400は、非一時的コンピュータ可読記憶媒体内に記憶され、コンピューティングデバイス(例えば、本明細書で説明したように、サーバデバイス、パーソナルコンピュータ、スマートフォン、タブレットコンピュータ、スマートウォッチ、モバイルコンピューティングデバイス、または他のクライアントコンピューティングデバイス)のプロセッサを使用して実行される有形のコンピュータ実行可能命令の形態における1つまたは複数の機能ブロック、モジュール、個々の関数、またはルーチンを含み得る。方法400は、任意のバックエンドサーバ(例えば、本明細書で説明するように、地図データサーバ、ナビゲーションサーバ、または任意の他のタイプのサーバコンピューティングデバイス)の一部、例えば、例示的な環境のクライアントコンピューティングデバイスモジュール、またはそのような環境の外部にあるモジュールの一部として含まれ得る。図は、説明を容易にするために他の図面を参照して説明されている場合があるが、方法400は、他のオブジェクトおよびユーザインターフェースとともに利用されることが可能である。さらに、上記の説明は、特定のデバイス(サーバデバイス60またはクライアントデバイス10など)によって実行される方法400のステップについて説明しているが、これは、説明の目的のみのために行われている。方法400のブロックは、環境の1つまたは複数のデバイスまたは他の部分によって実行され得る。

30

40

【0092】

本明細書で説明する例示的な方法の様々な動作は、関連する動作を実行するように(例えば、ソフトウェアによって)一時的に構成された、または永続的に構成された1つまたは複数のプロセッサによって少なくとも部分的に実行され得る。一時的に構成されているかまたは永続的に構成されているかにかかわらず、そのようなプロセッサは、1つまたは複数の動作または機能を実行するように動作するプロセッサ実装モジュールを構成し得る。本

50

明細書で言及するモジュールは、いくつかの例示的な実施形態において、プロセッサ実装モジュールを備え得る。

【0093】

同様に、本明細書で説明する方法およびルーチンは、少なくとも部分的にプロセッサで実施され得る。例えば、方法の動作のうち少なくともいくつかは、1つまたは複数のプロセッサまたはプロセッサ実装ハードウェアモジュールによって実行され得る。特定の動作の実行は、1つのマシン内に存在するだけでなく、いくつかのマシンにわたって展開される1つまたは複数のプロセッサ間で分散され得る。いくつかの例示的な実施形態において、プロセッサは、単一の場所内(例えば、住居環境内、オフィス環境内、またはサーバームとして)に配置され得るが、他の実施形態において、プロセッサは、いくつかの場所にわたって分散され得る。

10

【0094】

1つまたは複数のプロセッサはまた、「クラウドコンピューティング」環境内の、またはSaaSとしての関連する動作の実行をサポートするように動作し得る。例えば、上記で示したように、動作のうち少なくともいくつかは、(プロセッサを含むマシンの例として)コンピュータのグループによって実行され得、これらの動作は、ネットワーク(例えば、インターネット)と1つまたは複数の適切なインターフェース(例えば、API)とを介してアクセス可能である。

【0095】

さらに、図は、説明の目的のみのための例示的な環境のいくつかの実施形態を示している。当業者は、本明細書で説明する原理から逸脱することなく、本明細書で例示する構造および方法の代替実施形態が用いられ得ることを、以下の議論から容易に認識するであろう。

20

【0096】

本開示を読むと、当業者は、本明細書に開示された原理を通してコンテンツ認識音声ナビゲーション指示を提供するためのさらに追加の代替的な構造的および機能的な設計を理解するであろう。したがって、特定の実施形態および用途について図示および説明してきたが、開示した実施形態は、本明細書で開示した正確な構造および構成要素に限定されないことが理解されるべきである。添付の特許請求の範囲において定義されている要旨および範囲から逸脱することなく、当業者に明らかな様々な修正、変更、および変形が、本明細書で開示された方法および装置の配置、動作、および詳細においてなされ得る。

30

【符号の説明】

【0097】

1 環境

10 携帯デバイス、クライアントデバイス

12 車両

14 ヘッドユニット、車両ヘッドユニット、I/Oモジュール

16 通信リンク

18 ディスプレイ

19 センサ

20 ハードウェア入力制御

22 ハードウェア入力制御

24 マイクロフォン

26 スピーカ

30 通信ネットワーク

32 ユーザインターフェース(UI)

34 ナビゲーションデータサーバ、ナビゲーションサーバ

50 地図データサーバ

60 サーバデバイス

62 プロセッサ

40

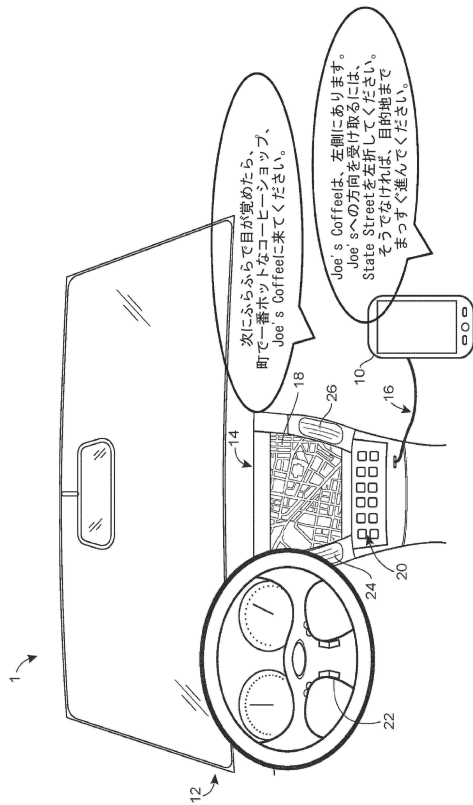
50

- 64 メモリ
- 68 コンテンツ認識音声ナビゲーション生成器
- 80 データベース
- 92 他のコンピューティングデバイス
- 100 通信システム
- 114 I/Oモジュール
- 116 プロセッサ(CPU)
- 120 メモリ
- 122 マッピングアプリケーション、地理アプリケーション、アプリケーション
- 126 オペレーティングシステム(OS)、OS
- 132 クライアントアプリケーション、アプリケーション
- 300 データ、トレーニングデータ、トレーニングデータセット
- 310 メディア/音声コンテンツ特性、特性
- 320 音声ナビゲーション指示パラメータ、パラメータ

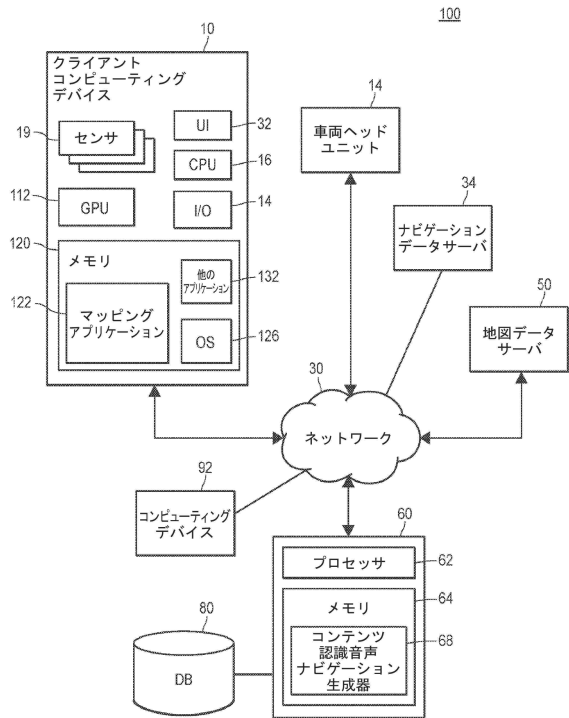
10

【図面】

【図 1】



【図 2】



20

30

40

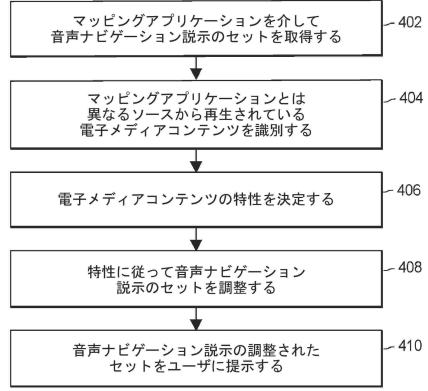
50

【 図 3 】

300
音声ナビゲーション指示パラメータ
320
音声コンテンツ特性
310

音声コンテンツ	速度	言語	POI	長さ	関連性レベル	操縦タイプ	場所	複雑さのレベル	調整
Joe's Coffeeの広告	1倍	英語	Joe's Coffee	30秒	低い	左折	State および Marinaの曲がり角	低い	広告の終了後に Joe's Coffeeへの 回り道を推奨する
スポーツのポッドキャスト	1.5倍	英語	なし	2分	低い	右折	Lake および Elmの曲がり角	低い	音声指示の速度を1.5倍に変更し、 文の終了後に 指示を唱唱する
音楽	1倍	スペイン語		10分	低い	高速道路の出口	Oak および Mapleの曲がり角	高い	音声指示の言語を スペイン語に変更する
ニュースのポッドキャスト	2倍	ヘブライ語		5分	高い	合流	Main および Oakの曲がり角	低い	音声指示の速度を2倍に変更し、言語をヘブライ語に変更する

【 図 4 】



10

20

30

40

50

フロントページの続き

- (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
- (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
- 審査官 宮本 礼子
- (56)参考文献 特開2019-117324(JP,A)
特開2010-117476(JP,A)
特開2011-095142(JP,A)
- (58)調査した分野 (Int.Cl., DB名)
G08G 1/00-99/00
G01C 21/00-21/36
G01C 23/00-25/00
G06F 3/16