(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)公開番号 特開2022-50309 (P2022-50309A)

(43)公開日 令和4年3月30日(2022.3.30)

(51)国際特許分類		FI		
H 0 4 N	7/15 (2006.01)	H 0 4 N	7/15	1 5 0
G 1 0 L	15/30 (2013.01)	G 1 0 L	15/30	
G 1 0 L	15/00 (2013.01)	G 1 0 L	15/00	2 0 0 U
H 0 4 N	21/24 (2011.01)	H 0 4 N	21/24	

審査請求 有 請求項の数 19 OL 外国語出願 公開請求 (全17頁)

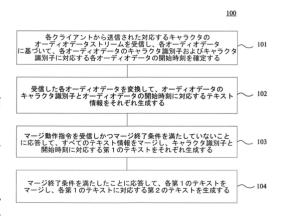
(21)出願番号	特願2021-99367(P2021-99367)	(71)出願人	514322098
(22)出願日	令和3年6月15日(2021.6.15)		ベイジン バイドゥ ネットコム サイエ
(31)優先権主張番号	202011140819.X		ンス テクノロジー カンパニー リミテ
(32)優先日	令和2年10月22日(2020.10.22)		ッド
(33)優先権主張国・地域又は機関			中華人民共和国 ペキン 100085,
	中国(CN)		ハイディアン ディストリクト , シャン
			ディ テンス ストリート, 10番, バ
(特許庁注:以下のものは登録商標)			イドゥ キャンパス 2階
1 . Z I G B E E		(74)代理人	110000578
			名古屋国際特許業務法人
		(72)発明者	リー シンビァオ
			中華人民共和国 100085 ベイジン
			ハイディエン ディストリクト シャンデ
			ィー テンス ストリート ナンバー 10
			バイドゥ キャンパス 2/エフ
			最終頁に続く

(54)【発明の名称】 情報処理方法、装置、システム、電子機器、記憶媒体およびコンピュータプログラム

(57)【要約】 (修正有)【課題】ユーザが会議情報を迅速に振り返り、会議キャラクタを識別することを容易にする、会議シーンにより適合した情報処理方法、装置、システム、電子機器及び記憶媒体を提供する。

【解決手段】情報処理方法は、各クライアントから送信された対応するキャラクタのオーディオデータストリームを受信し、各オーディオデータのキャラクタ識別プ及び各オーディオデータの開始時刻を確定するステップと、受信した各オーディオデータを変換して、各テキスト情報を生成するステップと、マージ動作指令を受信しかつマージ終了条件を満たしていないことに応答して、すべてのテキスト情報をマージし、各第1のテキストを生成するステップと、マージ終了条件を満たしたことに応答して、各第1のテキストをマージし、各第1のテキストに対応する第2のテキストを生成するステップと、を含む。

【選択図】図1



【特許請求の範囲】

【請求項1】

各クライアントから送信された対応するキャラクタのオーディオデータストリームを受信し、各前記オーディオデータに基づいて、各前記オーディオデータのキャラクタ識別子と、前記キャラクタ識別子に対応する各前記オーディオデータの開始時刻と、を確定するステップと、

受信した各オーディオデータを変換して、前記オーディオデータのキャラクタ識別子と前記オーディオデータの開始時刻とに対応する各テキスト情報を生成するステップと、

マージ動作指令を受信しかつマージ終了条件を満たしていないことに応答して、すべてのテキスト情報に対して、同じキャラクタ識別子のテキスト情報をマージすることを表すためのマージ動作を行い、前記キャラクタ識別子と前記開始時刻とに対応する各第1のテキストを生成するステップと、

マージ終了条件を満たしたことに応答して、各前記第1のテキストに対して、キャラクタ識別子および開始時刻に基づいて各前記第1のテキストを順列組合することを表すための統合動作を行い、各前記第1のテキストに対応する第2のテキストを生成するステップと、を含む情報処理方法。

【請求項2】

前記マージ動作指令は、クライアントから送信された同じキャラクタのオーディオデータストリームを受信したことに応答してトリガされるか、または現在の時点が予め設定された閾値に規定された時点を超えたことに応答してトリガされる、請求項 1 に記載の情報処理方法。

【請求項3】

前記第2のテキストを各前記クライアントまたは指定したクライアントに送信するステップをさらに含む、請求項1に記載の情報処理方法。

【請求項4】

マージ動作指令を受信しかつマージ終了条件を満たしていないことに応答して、すべてのオーディオデータに対して、各オーディオデータの開始時刻に基づいてすべてのオーディオデータをソートすることを表すための組み合わせ動作を行い、すべてのオーディオデータに対応する第1のオーディオデータを生成するステップと、

マージ終了条件を満たしたことに応答して、前記第1のオーディオデータを各前記クライアントまたは指定したクライアントに送信するステップと、をさらに含む、請求項1に記載の情報処理方法。

【請求項5】

特定の期間内に同じキャラクタ識別子の各オーディオデータの累積時間を表すための、キャラクタ識別子に対応するオーディオ総持続時間を、各前記オーディオデータの持続時間に基づいて確定するステップと、

マージ終了条件を満たしたことに応答して、前記オーディオ総持続時間を各前記クライアントまたは指定したクライアントに送信するステップと、をさらに含む、請求項1に記載の情報処理方法。

【請求項6】

情報処理要求を受信したことに応答して、各キャラクタの異なる時刻におけるオーディオデータを採集し、前記オーディオデータをサーバに送信し、サーバから送信された前記第2のテキストを受信したことに応答して、各キャラクタ識別子および対応する各開始時刻に基づいて、前記第2のテキストを表示するように構成されるクライアントと、

請求項1~5のいずれか1項に記載の情報処理方法を実行するように構成されるサーバと、を含む、情報処理システム。

【請求項7】

前記クライアントはさらに、サーバから送信された第1のオーディオデータを受信したことに応答して、前記第1のオーディオデータに対応する音声マップを表示するように構成されている、請求項6に記載の情報処理システム。

10

20

30

50

40

【請求項8】

前記クライアントはさらに、

前記音声マップに対してトリガされた再生動作を受信したことに応答して、再生動作に対応するオーディオデータを取得し、前記オーディオデータに基づいて、前記オーディオデータに対応する各キャラクタ識別子と、前記キャラクタ識別子に対応する各開始時刻と、前記キャラクタ識別子および前記開始時刻に対応する各テキスト情報と、を確定し、表示された前記第2のテキストに対してトリガされた選択動作を受信したことに応答して、選択動作に対応する前記第2のテキストにおけるテキスト情報に基づいて、前記テキスト情報に対応するオーディオデータ

前記オーディオデータが表す音声を再生し、前記音声に対応するテキスト情報が現在のページにあることに応答して、前記音声に対応する各テキスト情報を表示するように構成されている、請求項7に記載の情報処理システム。

【請求項9】

を確定し、

前記クライアントはさらに、前記音声に対応するテキスト情報が現在のページに存在していないことを検出したことに応答して、ジャンプボタンを表示するように構成されている、請求項8に記載の情報処理システム。

【請求項10】

前記クライアントはさらに、サーバから送信された、キャラクタ識別子に対応するオーディオ総持続時間を受信したことに応答して、キャラクタ識別子に基づいて前記オーディオ総持続時間を表示するように構成されている、請求項 6 に記載の情報処理システム。

【請求項11】

前記クライアントはさらに、前記第2のテキストおよび議事録テンプレートに基づいて、 前記第2のテキストに対応する議事録を生成するように構成されている、請求項6に記載 の情報処理システム。

【請求項12】

各クライアントから送信された対応するキャラクタのオーディオデータストリームを受信し、各前記オーディオデータに基づいて、各前記オーディオデータのキャラクタ識別子と、前記キャラクタ識別子に対応する各前記オーディオデータの開始時刻と、を確定するように構成される受信ユニットと、

受信した各オーディオデータを変換して、前記オーディオデータのキャラクタ識別子と前記オーディオデータの開始時刻とに対応する各テキスト情報を生成するように構成される 変換ユニットと、

マージ動作指令を受信しかつマージ終了条件を満たしていないことに応答して、すべての テキスト情報に対して、同じキャラクタ識別子のテキスト情報をマージすることを表すた めのマージ動作を行い、前記キャラクタ識別子と前記開始時刻とに対応する各第1のテキ ストを生成するように構成されるマージユニットと、

マージ終了条件を満たしたことに応答して、各前記第1のテキストに対して、キャラクタ識別子および開始時刻に基づいて各前記第1のテキストを順列組合することを表すための統合動作を行い、各前記第1のテキストに対応する第2のテキストを生成するように構成される情報処理ユニットと、を含む情報処理装置。

【請求項13】

前記マージユニットの前記マージ動作指令は、

クライアントから送信された同じキャラクタのオーディオデータストリームを受信したことに応答してトリガされるか、または現在の時点が予め設定された閾値に規定された時点を超えたことに応答してトリガされる、請求項12に記載の情報処理装置。

【請求項14】

前記第2のテキストを各前記クライアントまたは指定したクライアントに送信するように 構成される情報送信ユニットをさらに含む、請求項12に記載の情報処理装置。

【請求項15】

10

30

20

40

マージ動作指令を受信しかつマージ終了条件を満たしていないことに応答して、すべてのオーディオデータに対して、各オーディオデータの開始時刻に基づいてすべてのオーディオデータをソートすることを表すための組み合わせ動作を行い、すべてのオーディオデータに対応する第 1 のオーディオデータを生成するように構成される生成ユニットをさらに含み、

前記情報送信ユニットはさらに、マージ終了条件を満たしたことに応答して、前記第1のオーディオデータを各前記クライアントまたは指定したクライアントに送信するように構成されている、請求項14に記載の情報処理装置。

【請求項16】

特定の期間内に同じキャラクタ識別子の各オーディオデータの累積時間を表すための、キャラクタ識別子に対応するオーディオ総持続時間を、各前記オーディオデータの持続時間に基づいて確定するように構成される確定ユニットをさらに含み、

前記情報送信ユニットはさらに、マージ終了条件を満たしたことに応答して、前記オーディオ総持続時間を各前記クライアントまたは指定したクライアントに送信するように構成されている、請求項14に記載の情報処理装置。

【請求項17】

少なくとも 1 つのプロセッサと、前記少なくとも 1 つのプロセッサと通信可能に接続されるメモリとを含む電子機器であって、

前記メモリに前記少なくとも1つのプロセッサによって実行可能な指令が格納されており、前記指令が前記少なくとも1つのプロセッサによって実行されると、前記少なくとも1つのプロセッサに請求項1~5のいずれか1項に記載の情報処理方法を実行する、電子機器。

【請求項18】

コンピュータ指令が格納されている非一時的コンピュータ可読記憶媒体であって、

前記コンピュータ指令はコンピュータに請求項1~5のいずれか1項に記載の情報処理方法を実行させるためのものである、非一時的コンピュータ可読記憶媒体。

【請求項19】

プロセッサにより実行されると、請求項 1 ~ 5 のいずれか 1 項に記載の情報処理方法を実現するコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

[0001]

本出願の実施形態は、人工知能技術分野に関し、具体的には音声認識、音声マージ、および自然言語処理技術分野に関し、特に情報処理方法および装置に関する。

【背景技術】

[0002]

人工知能は、コンピュータに人間の一部の思考過程と知能行為(例えば学習、推論、思考、計画など)をシミュレーションさせることを研究する学科であり、ハードウェアレベルの技術もあれば、ソフトウェアレベルの技術もある。人工知能ハードウェア技術は一般的に、センサ、専用人工知能チップ、クラウドコンピューティング、分散記憶、ビッグデータ処理などの技術を含む。人工知能ソフトウェア技術は主にコンピュータ視覚技術、音声認識技術、自然言語処理技術および機械学習 / ディープラーニング、ビッグデータ処理技術、知識画像技術などのいくつかの方向を含む。

[0003]

人工知能の発展に伴い、オンライン音声ビデオ会議サービスを利用することが企業における日常会議の主流となっている。実際のシーンでは、ユーザが会議を見落とした場合、その後に会議内容を見たい場合、重要な会議の参加者が会議の重要な情報を振り返る必要がある場合など、会議記録機能が音声ビデオ会議ソフトウェアの重要な機能となっている。

[0004]

現段階の音声ビデオ会議の録音・録画シーンでは、録音・録画ソフトは音声保存オーディ

10

20

30

40

オのみを録音したり、音声を一区切りの文字に変換したりする形を採用しており、同じ時刻に複数人が発話していると録音効果が悪く、音声から変換された文字も乱れてしまう。 そのため、ユーザが会議情報を迅速に振り返り、会議キャラクタを識別することを容易に する、会議シーンにより適合した会議記録方式をユーザに提供する必要がある。

【発明の概要】

[0005]

本出願は情報処理方法、装置、システム、電子機器、記憶媒体およびコンピュータプログラムを提供し、人工知能技術分野に関し、具体的に音声認識、音声合成、および自然言語処理の技術分野に関する。

[0006]

第1態様において、本出願の実施形態は、各クライアントから送信された対応するキャラクタのオーディオデータは基づいて、各オーディオデータに基づいて、各オーディオデータに基づいて、各オーディオデータに基づいて、各オーディオデータのキャラクタ識別子と、サャラクタ識別子に対応する各オーディオデータを変換して、オーディオデータの開始時刻とに対応する各テキスト情報に対して、同じキャラクタ識別子とオーディスト情報に対して、同じキャラクタ識別子のテキスト情報に対して、同じキャラクタ識別子のテキスト情報に対して、キャラクタ識別子と開始時刻とに応答して、ちてのテキスト情報に対して、キャラクタ識別子とに応答したことに応答と第1のテキストを生成するステップと、マージ終了条件を満たしたことに応答と第1のテキストに対して、キャラクタ識別子および開始時刻に基づいて各第1のテキストに対応する第2のテキストを生成するステップと、を含む、情報処理方法を提供する。

[0007]

第2態様において、本出願の実施形態は、情報処理要求を受信したことに応答して、各キャラクタの異なる時刻におけるオーディオデータを採集し、オーディオデータをサーバに送信し、サーバから送信された第2のテキストを受信したことに応答して、各キャラクタ識別子および対応する各開始時刻に基づいて、第2のテキストを表示するように構成されるクライアントと、上記いずれか1項に記載の情報処理方法を実行するように構成されるサーバと、を含む、情報処理システムを提供する。

[0008]

第3態様において、本出願の実施形態は、各クライアントから送信された対応するキャラクタがイーディオデータストリームを受信し、各オーディオデータの識別子と、キャラクタ識別トと、では、受信した名ののアーディオデータが高いに構成される受信したことででは、カーディオデータが高いたが、受信したののでは、カーディオデータのおうに構成されるで換して、オーデーをでは、カージのでは

[0009]

第4態様において、本出願の実施形態は、少なくとも1つのプロセッサと、少なくとも1つのプロセッサと通信可能に接続された記憶装置とを含む電子機器であって、記憶装置に少なくとも1つのプロセッサによって実行可能な指令が記憶されており、指令が少なくとも1つのプロセッサによって実行されると、少なくとも1つのプロセッサに第1態様のいずれかの実施形態に記載の方法が実装される電子機器を提供する。

[0010]

20

10

30

20

30

40

50

第 5 態様において、本出願の実施形態は、コンピュータ指令が記憶されている非一時的コンピュータ可読記憶媒体であって、コンピュータ指令はコンピュータに第 1 態様に記載のいずれかの方法を実施させるためのものである、非一時的コンピュータ可読記憶媒体を提供する。

[0011]

第6態様において、本出願の実施形態は、プロセッサにより実行されると、本出願の第1 態様のいずれかの実施形態に記載の方法が実装される、コンピュータプログラムを提供する。

[0012]

上記説明は、本出願の実施形態の肝心または重要な特徴を認識することが意図されているわけではなく、本出願の範囲を限定するためにも使用されないことを理解されたい。本出願の他の特徴は、以下の説明によって理解しやすくなるであろう。

【図面の簡単な説明】

[0013]

図面は、本出願をよりよく理解するために使用されるものであって、本出願の限定を構成 しない。

- 【図1】本出願に係る情報処理方法の一実施形態を示す模式図である。
- 【図2】本出願の実施形態に係る情報処理方法を実行可能なシーン図である。
- 【図3】本出願に係る情報処理方法の別の実施形態を示す模式図である。
- 【図4】本出願に係る情報処理システムの一実施形態の構成模式図である。
- 【 図 5 】 本 出 願 に 係 る 情 報 処 理 方 法 を ユ ー ザ に 示 す イ ン タ ー フ ェ ー ス 模 式 図 で あ る 。
- 【図6】本出願に係る情報処理装置の一実施形態の構成模式図である。
- 【 図 7 】 本 発 明 の 実 施 形 態 に 係 る 情 報 処 理 方 法 を 実 行 す る た め の 電 子 機 器 の ブ ロ ッ ク 図 で あ る 。

【発明を実施するための形態】

[0014]

以下、図面に関連して、本出願の例示的な実施形態を説明する。理解を容易にするために、本出願の実施形態の様々な詳細を含むが、それらは例示的なものにすぎないとみなされるべきである。したがって、当業者であれば、本出願に記載された実施形態は、本出願の範囲および趣旨から逸脱することなく、様々な変更および修正を行うことができることを認識するであろう。同様に、以下の説明では、明確化かつ簡略化のために、公知の機能および構造の説明を省略する。

[0 0 1 5]

なお、矛盾を生じない限り、本出願の実施形態および実施形態における特徴は、互いに組み合わせられてもよい。以下、図面を参照し、実施形態に関連して、本出願について詳細に説明する。

[0016]

図 1 は、本出願に係る情報処理方法の一実施形態を示す模式図 1 0 0 である。この情報処理方法は、具体的には、以下のステップ(ステップ 1 0 1 ~ 1 0 4)を含む。

[0017]

ステップ101では、各クライアントから送信された対応するキャラクタのオーディオデータストリームを受信し、各オーディオデータに基づいて、各オーディオデータのキャラクタ識別子と、キャラクタ識別子に対応する各オーディオデータの開始時刻とを確定する

[0018]

本実施形態では、実行主体(例えば、サーバまたはクラウドプラットフォーム)はクライアントから送信された情報処理要求を受信すると、各クライアントから送信された対応するキャラクタのオーディオデータストリームを有線接続方式または無線接続方式で受信し、そして、各オーディオデータに対してコンテンツ解析を行うか、または各オーディオデータを用いてデータベースに問い合わせることにより、各オーディオデータのキャラクタ

識別子と、キャラクタ識別子に対応する各オーディオデータの開始時刻とを確定すること ができる。本実施形態では、クライアントは、ユーザから発した議事録イベント(例えば 、 所 定 の ボ タン / メ ニ ュ ー ま た は 設 定 さ れ た シ ョ ー ト カ ッ ト キ ー が ク リ ッ ク / ト リ ガ さ れ たイベント)を聞くことにより、ユーザの会議記録指示を受け、会議記録指示に基づいて 情報処理要求を生成して送信することができる。各キャラクタは、会議ソフトウェアに登 録された各発話者アカウントを表すことができる。各キャラクタのオーディオデータは、 会議シーンにおいて各キャラクタが異なる時刻に発話したオーディオデータを表すことが できる。異なる時刻のオーディオデータは、予め設定された時間に基づいて、オーディオ データの時間間隔が予め設定された閾値より大きいか否かを判断することで得られる。例 えば、同じキャラクタの長時間発話におけるポーズ時間が5秒を超えるとセグメント化し すなわち、次の時刻のオーディオデータとする。あるいは同じキャラクタの発話時間が 6 0 秒 継 続 す る と 、 そ の キ ャ ラ ク タ 発 話 に お け る ポ ー ズ 時 間 が 2 秒 を 超 え る と セ グ メン ト 化する。なお、上記の無線接続方式は、3G、4G、5G接続、Wi-Fi接続、ブルー トゥース(登録商標)接続、WiMAX接続、Zigbee接続、UWB(Ultra Wideband)接続、および他の現在に知られているまたは将来に開発される無線接 続方式を含むことができるが、これらに限定されない。

[0019]

ステップ 1 0 2 では、受信した各オーディオデータを変換して、オーディオデータのキャラクタ識別子とオーディオデータの開始時刻とに対応する各テキスト情報を生成する。

[0020]

本実施形態では、実行主体は、音声変換方法に基づいて、ステップ101で受信した異なるキャラクタの各オーディオデータを変換し、オーディオデータのキャラクタ識別子とオーディオデータの開始時刻とに対応する各テキスト情報を生成することができる。

[0021]

ステップ103では、マージ動作指令を受信しかつマージ終了条件を満たしていないこと に応答して、すべてのテキスト情報をマージし、キャラクタ識別子と開始時刻とに対応す る各第1のテキストを生成する。

[0022]

本実施形態では、実行主体はマージ動作指令を受信しかつマージがマージ終了条件を満たしていないと判断した後、すべてのテキスト情報に対して、同じキャラクタタグのテキスト情報をマージすることを表すためのマージ動作を行い、キャラクタ識別子と開始時では、対応する各第1のテキストを生成する。マージ終了条件は予め設定されることができる。例えば、マージ時間が60分間に達した場合には、マージ終了条件を満たしたと判断され、および/または、マージテキストのサイズ(すなわち、各第1のテキストのサイズの合計)が予め設定された閾値を超えた場合には、マージ終了条件を満たしたと判断される。マージ終了条件はさらに、ユーザが動作ボタンをクリックしてマージ動作指令を終了した場合には、マージ終了条件を満たしたと判断される。

[0 0 2 3]

ステップ104では、マージ終了条件を満たしたことに応答して、各第1のテキストをマージし、各第1のテキストに対応する第2のテキストを生成する。

[0024]

本実施形態では、実行主体はマージがマージ終了条件を満たしたと判断した場合、各第 1 のテキストに対して、キャラクタ識別子および開始時刻に応じて各第 1 のテキストを順列組合することを表すための統合動作を行い、各第 1 のテキストに対応する第 2 のテキストを生成する。

[0 0 2 5]

図 2 によると、本実施形態に係る情報処理方法 2 0 0 は、電子機器 2 0 1 において実行される。電子機器 2 0 1 は、情報処理要求を受信すると、まず、各クライアントから送信さ

10

20

30

40

れた対応するキャラクタのオーディオデータストリームを受信し、各オーディオデータに基づいて、各オーディオデータのキャラクタ識別子と、キャラクタ識別子に対応する各オーディオデータの開始時刻とを確定する(202)。そして、電子機器201は、受信した各オーディオデータを変換し、オーディオデータのキャラクタ識別子とオーディオデータの開始時刻とに対応する各テキスト情報を生成する(203)。電子機器201は、マージ動作指令を受信しかつマージ終了条件を満たしていない場合には、すべてのテキスト情報をマージし、キャラクタ識別子と開始時刻に対応する各第1のテキストを生成する(204)。電子機器201は、マージ終了条件を満たした場合、各第1のテキストをマージし、各第1のテキストに対応する第2のテキストを生成し(205)、第2のテキストを各クライアントまたは指定したクライアントに送信する。

[0026]

本出願の上述した実施形態に係る情報処理方法は、各クライアントから送信された対応す るキャラクタのオーディオデータストリームを受信し、各オーディオデータに基づいて、 各 オ ー デ ィ オ デ ー タ の キ ャ ラ ク タ 識 別 子 と 、 キ ャ ラ ク タ 識 別 子 に 対 応 す る 各 オ ー デ ィ オ デ ータの開始時刻とを確定し、受信した各オーディオデータを変換し、オーディオデータの キャラクタ識別子とオーディオデータの開始時刻とに対応する各テキスト情報を生成し、 マージ動作指令を受信しかつマージ終了条件を満たしていないことに応答して、すべての テ キ ス ト 情 報 を マ ー ジ し 、 キ ャ ラ ク タ 識 別 子 と 開 始 時 刻 と に 対 応 す る 各 第 1 の テ キ ス ト を 生成し、マージ終了条件を満たしたことに応答して、各第1のテキストをマージし、各第 1のテキストに対応する第2のテキストを生成することにより、複数のクライアントにお ける異なるキャラクタのオーディオデータを受信・変換した後、テキスト統合を経て送信 情報を確定する方法を実現した。各クライアントにおける異なるキャラクタのオーディオ データをそれぞれ受信し、異なるキャラクタに応じて音声の採集と処理を行うことにより 、従来技術に見られる同じ時刻に複数人がオンラインで話すと録音効果が悪くなるという 問題を解決した。キャラクタ別にオーディオを採集・変換し、オーディオデータのキャラ ク タ 識 別 子 と オ ー デ ィ オ デ ー タ の 開 始 時 刻 と に 対 応 す る 各 テ キ ス ト 情 報 を 生 成 す る こ と に より、従来技術に見られる音声から変換した文字が乱れるという問題を解決した。この方 法により、会議シーンにおいては、会議内容の可視化が可能となる。会議のオーディオを テキスト形式に変換してクライアントに送信し、テキスト情報をユーザに提示することに より、ユーザは筋道がよく立って迅速に会議内容を振り返ることができる。ユーザの問合 せ時間を節約し、ユーザの読解体験を向上させることができる。

[0027]

図3は、情報処理方法の別の実施形態の模式図300を示す。当該方法のフローは、以下のステップ(ステップ301~305)を含む。

[0028]

ステップ 3 0 1 では、各クライアントから送信された対応するキャラクタのオーディオデータストリームを受信し、各オーディオデータに基づいて、各オーディオデータのキャラクタ識別子と、キャラクタ識別子に対応する各オーディオデータの開始時刻とを確定する

[0029]

ステップ302では、受信した各オーディオデータを変換して、オーディオデータのキャラクタ識別子とオーディオデータの開始時刻とに対応する各テキスト情報を生成する。

[0030]

ステップ303では、マージ動作指令を受信しかつマージ終了条件を満たしていないことに応答して、すべてのテキスト情報をマージし、キャラクタ識別子と開始時刻とに対応する各第1のテキストを生成し、すべてのオーディオデータに対して組み合せ動作を行い、すべてのオーディオデータに対応する。

[0 0 3 1]

本実施形態では、実行主体はマージ動作指令を受信しかつマージがマージ終了条件を満たしていないと判断した後、すべてのテキスト情報をマージし、キャラクタ識別子と開始時

10

20

30

40

20

30

40

50

刻とに対応する各第1のテキストを生成し、すべてのオーディオデータに対して、各オーディオデータの開始時刻に応じてすべてのオーディオデータをソートすることを表すための組み合せ動作を行い、すべてのオーディオデータに対応する第1のオーディオデータを生成することができる。

[0032]

本実施形態のいくつかのオプション的な実施形態では、マージ動作指令は、クライアントから送信された同じキャラクタのオーディオデータストリームを受信したことに応答してトリガされるか、またはマージ動作指令は、現在の時点が予め設定された閾値に規定された時点を超えたことに応答してトリガされる。様々なマージ動作指令のトリガ条件を設けることにより、柔軟で多様なテキストマージを実現し、システムおよびユーザの様々なニーズを満たすとともに、システム効率を向上させることができる。

[0 0 3 3]

ステップ304では、各オーディオデータの持続時間に基づいて、キャラクタ識別子に対応するオーディオ総持続時間を確定する。

[0034]

本実施形態では、実行主体は、各オーディオデータの持続時間に基づいて、キャラクタ識別子に対応するオーディオ総持続時間を確定することができ、オーディオ総持続時間は、 ある期間内に同じキャラクタ識別子の各オーディオデータの累積時間を表す。

[0035]

ステップ305では、マージ終了条件を満たしたことに応答して、各第1のテキストをマージし、各第1のテキストに対応する第2のテキストを生成し、第2のテキスト、第1のオーディオデータ、およびオーディオ総持続時間を各クライアントまたは指定したクライアントに送信する。

[0036]

本実施形態では、実行主体は、マージがマージ終了条件を満たしたと判断した後、各第1のテキストをマージし、各第1のテキストに対応する第2のテキストを生成し、そして第2のテキスト、ステップ303で生成された第1のオーディオデータ、およびステップ304で確定されたオーディオ総持続時間を各クライアントまたは指定したクライアントに送信してユーザに表示することができる。

[0037]

なお、上述した条件判断およびデータ統合方法は、現在広く研究および適用されている公 知技術であり、ここではこれ以上説明しない。

[0038]

本実施形態では、ステップ 3 0 1 および 3 0 2 の具体的な動作は、図 1 に示す実施形態におけるステップ 1 0 1 および 1 0 2 の動作と実質的に同じであり、ここではこれ以上説明しない。

[0039]

図3から分かるように、図1に対応すると比較して、本実施形態における情報処理方法の模式図300は、マージ動作指令を受信しかつマージ終済と時間におけるに応答して、すべてのテキスト情報をマージを識別して組み合せ動作を受信した。すべてのオーディオデータに対して組み合き割1のオーディオデータに対応する第1のオーディオがロータでは成が持続時間に基満たしたことに応第1のテキストを生成し、キャラクタに対応するオーディオ総持続時間である第1のテキストを生成し、第2のテキストに対応する第1のオーディを第1のテキストを生成のテキストを生成し、第2のテキストに対応する第1でを第1のテキストに対応する第1でで表がローガル機器のテキストである第1でを生がしたがよりにはいて、第2のデキストを生成がローカル機器の音声できずのに他の端末からにという問題を解決し、より豊富な情報に送を実現した。これによりの発生したすいという問題を解決し、より豊富な情報をユーザに提示することがで

きる。

[0040]

図4は、本出願に係る情報処理システムの一実施形態の構成模式図400を示す。図4に示すように、この情報処理システムは、情報処理要求を受信したことに応答して、各キャラクタの異なる時刻のオーディオデータを採集し、オーディオデータをサーバに送信し、サーバから送信された第2のテキストを受信したことに応答して、各キャラクタ識別子および対応する各開始時刻に基づいて、図5に示すように第2のテキストを表示するように構成されるサーバ402と、を含む。具体的には、テキストを表示する過程で、同じ時間帯に複数人が発話する場合、発話時間の早いから遅い順にソートし、複数人の発話の開始時間が同じである場合、発話者の名前の頭文字、数字順にソートする。

[0041]

システムにおいて、クライアントはさらに、サーバから送信された第1のオーディオデータを受信したことに応答して、第1のオーディオデータに対応する音声マップを表示するように構成されている。クライアントはサーバから送信された第1のオーディオデータを受信した場合、図5に示すように第1のオーディオデータに対応する音声マップをユーザに提示することができる。

[0042]

システムにおいて、クライアントはさらに、音声マップにトリガされた再生動作を受信したことに応答して、再生動作に対応するオーディオデータを取得し、オーディオデータに対応する各キャラクタ識別子、キャラクタ識別子に対応する各界が表した第2のテキストにおけるテキスト情報を取得したことに応答しておけるテキスト情報を取得し、第2のテキストにおけるテキスト情報を取得し、第2のテキストにおけるテキスト情報を取得し、第2のテキストにおけるテキスト情報に基づいて、テキスト情報に対応するオーディオデータを確定にあることができるように構成されている。これによりで、音声に対応する各テキスト情報を表示するように構成されている。これによりに対対が発話した場合には、複数人の文字の位置特定により、合きにハイライト表示することができる。テキストと音声の相互位置特定により、と中にテキストは、音声再生に伴って自動的にスクロールするとともに文字ごとにハイライト表示される。

[0043]

システムにおいて、クライアントはさらに、音声に対応するテキスト情報が現在のページに存在しないことを検出したことに応答して、ジャンプボタンを表示するように構成されている。これにより、文字の段落全体が積み重ねられて表示されており、ユーザはキー情報を迅速に位置特定、認識することができず、音声再生中にユーザは対応するテキスト情報に高度に注意を払う必要があり、労力がかかり、読みにくいという問題を解決し、キー情報を位置特定、認識する効率を向上させることができる。

[0044]

システムにおいて、クライアントはさらに、サーバから送信されたキャラクタ識別子に対応するオーディオ総持続時間を受信したことに応答して、キャラクタ識別子に基づいてオーディオ総持続時間を表示するように構成されている。これにより、より豊富で多様な情報提示を実現し、ユーザの異なるニーズに応えることができる。

[0045]

システムにおいて、クライアントはさらに、第2のテキストおよび議事録テンプレートに基づいて、図5に示すように第2のテキストに対応する議事録を生成するように構成されている。テキストマージとテキスト統合を通じて、会議ごとのすべての発話者のテキストをマージして完全な会議内容に復元することにより、ユーザは必要に応じてテキストをスクロール閲覧、検索し、発話者、発話時間、発話内容の提示形式を採用して、より明確かつ系統的に会議内容を読むことができるため、ユーザのエネルギーを節約し、読解体験を

10

20

30

40

向上させることができる。

[0046]

図4から分かるように、この情報処理システムにより、文字段落全体が積み重ねられて表示されており、ユーザはキー情報を迅速に位置特定、認識することができないという問題を解決した。ユーザはテキスト内容を直接コピーしたり、会議のキー情報を抜粋したりして関連文書(例えば、議事録)を作成することができるため、文書作成の効率が向上している。テキストと音声マップを同時に表示することにより、より豊富な情報提示を実現し、ユーザの多様なニーズをさらに満たすことができる。

[0047]

さらに図6を参照すると、本出願では、上述した図1~3に示す方法の実装として、様々な電子機器に具体的に適用可能な情報処理装置の一実施形態を提供する。この装置の実施 形態は、図1に示す方法の実施形態に対応する。

[0048]

図6に示すように、本実施形態に係る情報処理装置600は、各クライアントから送信をれた対応するキャラクタのオーディオデータのキャラクタ識別子と、キャラクタ識別子と、キャラクタ識別子と、キャラクタ識別子と、キャラクタ識別子と、キャラクタ識別子と、カーディオデータのキャラクタ識別子と、キャラクタ識別子とでは、カーディオデータの開始時刻とを確定するように構成するように構成ないまででは、カーダーの開始に対応する各テキスト情報を生成するように構成ないまででは、カージーが表すといい、キャラクタ識別子とに対立して、とれてのテキスト情報に対して、キャラクタ識別子とに対が了条件を満たにでいますが、キャラクタ識別子とにができまれるマージを表すためのマキストを生成するように構成される情報処理ユニット604と、を含む。

[0049]

本実施形態では、情報処理装置600の受信ユニット601、変換ユニット602、マージユニット603および情報処理ユニット604の具体的な処理およびそれによる技術的効果については、それぞれ図1に対応する実施形態におけるステップ101~104に対する説明を参照することができる。ここではこれ以上説明しない。

[0050]

本実施形態のいくつかのオプション的な実施形態では、マージユニットのマージ動作指令はクライアントから送信された同じキャラクタのオーディオデータストリームを受信したことに応答してトリガされるか、またはマージユニットのマージ動作指令は現在の時点が予め設定された閾値に規定された時点を超えたことに応答してトリガされる。

[0051]

本実施形態のいくつかのオプション的な実施形態では、装置は、第2のテキストを各クライアントまたは指定したクライアントに送信するように構成される情報送信ユニットをさらに含む。

[0 0 5 2]

本実施形態のいくつかのオプション的な実施形態では、装置は、マージ動作指令を受信しかつマージ終了条件を満たしていないことに応答して、すべてのオーディオデータに対して、各オーディオデータの開始時刻に応じてすべてのオーディオデータをソートすることを表すための組み合わせ動作を行い、すべてのオーディオデータに対応する第1のオーディオデータを生成するように構成される生成ユニットをさらに含み、情報送信ユニットはさらに、マージ終了条件を満たしたことに応答して、第1のオーディオデータを各クライアントまたは指定したクライアントに送信するように構成されている。

[0053]

本実施形態のいくつかのオプション的な実施形態では、装置は、各オーディオデータの持

10

20

30

40

20

30

40

50

続時間に基づいて、キャラクタ識別子に対応する、ある期間内に同じキャラクタ識別子の各オーディオデータの累積時間を表すためのオーディオ総持続時間を確定するように構成される確定ユニットをさらに含み、情報送信ユニットはさらに、マージ終了条件を満たしたことに応答して、オーディオ総持続時間を各クライアントまたは指定したクライアントに送信するように構成されている。

[0054]

図7は、本発明の実施形態に係る情報処理方法を実行するための電子機器のブロック図である。電子機器は、ラップトップ、デスクトップコンピュータ、ワークステーション、パーソナルデジタルアシスタント、サーバ、ブレードサーバ、メインフレームコンピュータ、および他の適切なコンピュータのような様々な形態のデジタルコンピュータを表すことが意図されている。電子機器はまた、個人デジタル処理、携帯電話、スマートフォン、ウェアラブル装置、および他の同様のコンピューティング装置のような様々な形態のモバイル装置を表すことができる。本出願に示すコンポーネント、それらの接続および関係、ならびにそれらの機能は、一例に過ぎず、本出願に記載されたおよび/または要求される本出願の実装を限定することは意図されていない。

[0055]

図7に示すように、この電子機器は、1つまたは複数のプロセッサ701と、メモリ702と、高速インターフェースおよび低速インターフェースを含む様々なコンポーネントを接続するためのインターフェースと、を含む。各コンポーネントは、異なるバスで互いに接続されており、共通マザーボードに実装されていてもよく、必要に応じて他の方法で実装されていてもよい。プロセッサは、インターフェースに結合された表示装置などの外部入出力装置上にGUIのグラフィック情報を表示するために、メモリまたはメモリ上に格納された指令を含む電子機器内で実行される指令を処理することができる。他の実施形態では、複数のプロセッサおよび/または複数のバスは、必要に応じて、複数のメモリおよび複数のメモリとともに使用されてもよい。同様に、複数の電子機器を接続することがで複数のメモリとともに使用されてもよい。同様に、複数の電子機器を接続することができる、各機器は、部分的に必要な動作(例えば、サーバアレイ、ブレードサーバのセット、またはマルチプロセッサシステムとして)を提供することができる。図7では、1つのプロセッサ701を例に挙げている。

[0056]

メモリ702は、本出願によって提供される非一時的コンピュータ可読記憶媒体である。ここで、メモリには、少なくとも1つのプロセッサに、本出願によって提供される情報処理方法を実行させるために、少なくとも1つのプロセッサによって実行可能な指令が格納されている。本出願の非一時的コンピュータ可読記憶媒体は、本出願によって提供される情報処理方法をコンピュータに実行させるためのコンピュータ指令を記憶する。

[0 0 5 7]

メモリ702は、非一時的コンピュータ可読記憶媒体として、非一時的ソフトウェアプログラム、非一時的コンピュータ実行可能プログラム、並びに本実施形態における情報処理方法に対応するプログラム指令/モジュール(例えば、図6に示す受信ユニット601、変換ユニット602、マージユニット603、情報処理ユニット604)のようなモジュールを記憶するために使用されることができる。プロセッサ701は、メモリ702に記憶された非一時的ソフトウェアプログラム、指令、およびモジュールを実行することにより、サーバの様々な機能アプリケーションおよびデータ処理を実行する。すなわち、上述した方法の実施形態における情報処理方法を実現する。

[0058]

メモリ702は、オペレーティングシステム、少なくとも1つの機能に必要なアプリケーションを記憶することができるプログラム記憶領域と、情報処理方法を実行する電子機器の使用に応じて作成されたデータなどを記憶することができるデータ記憶領域と、を含むことができる。さらに、メモリ702は、高速ランダムアクセスメモリを含むことができ、少なくとも1つのディスク記憶装置、フラッシュメモリ装置、または他の非一時的ソリッドステート記憶装置のような非一時的メモリを含むこともできる。いくつかの実施形態

では、メモリ702は、任意に、プロセッサ701に対して遠隔設定されたメモリを含み、これらの遠隔メモリは、ネットワークを介して情報処理方法を実行する電子機器に接続されることができる。上述したネットワークの例は、インターネット、企業内ネットワーク、ローカルエリアネットワーク、移動通信網、およびそれらの組み合わせを含むが、これらに限定されない。

[0059]

情報処理方法を実行するための電子機器は、入力装置703と、出力装置704と、をさらに含むことができる。プロセッサ701、メモリ702、入力装置703、および出力装置704は、バスまたは他の方法で接続されることができる。図7は、バスを介して接続されている例を示す。

[0060]

入力装置703は、入力された数字または文字情報を受信し、タッチパネル、キーパッド、マウス、トラックパッド、タッチパッド、ポインティングレバー、1つまたは複数のマウスボタン、トラックボール、ジョイスティックなどの情報処理方法を実行する電子機器のユーザ設定および機能制御に関するキー信号入力を生成することができる。出力装置704は、表示装置、補助照明装置(例えば、LED)、および触覚フィードバック装置(例えば、振動モータ)などを含むことができる。表示装置は、液晶ディスプレイ(LCD)、発光ダイオード(LED)ディスプレイ、およびプラズマディスプレイを含むことができるが、これらに限定されない。いくつかの実施形態では、表示装置はタッチパネルであってもよい。

[0061]

本出願に記載されたシステムおよび技術の様々な実施形態は、デジタル電子回路システム、集積回路システム、特定用途向け集積回路(ASIC)、コンピュータハードウェア、ファームウェア、ソフトウェア、および/またはそれらの組み合わせで実装されてもよい。これらの様々な実施形態は、1つまたは複数のコンピュータプログラムに実装されることを含むことができる。この1つまたは複数のコンピュータプログラムは、少なくとも1つのプログラマブルプロセッサを含むプログラマブルシステム上で実行および/または解釈することができる。このプログラマブルプロセッサは、記憶システム、少なくとも1つの入力装置、および少なくとも1つの出力装置が命令を記憶システム、少なくとも1つの入力装置、および少なくとも1つの出力装置に送信することができる専用または汎用プログラマブルプロセッサであってもよい。

[0062]

これらのコンピュータプログラム(プログラム、ソフトウェア、ソフトウェアアプリケーション、またはコードとも呼ばれる)は、プログラマブルプロセッサの機械命令を含む。これらのコンピュータプログラムは、高度なプロセスおよび / またはオブジェクト指向プログラミング言語、および / またはアセンブリ / 機械言語を使用して実施されることができる。本出願で使用されるように、「機械可読媒体」および「コンピュータ可読媒体」という用語は、機械命令および / またはデータをプログラマブルプロセッサに提供するための任意のコンピュータプログラム製品、機器、および / または装置(例えば、磁気ディスク、光ディスク、メモリ、プログラマブル論理装置(PLD))を意味する。それには、機械可読信号として機械命令を受信する機械可読媒体が含まれる。「機械可読信号」という用語は、機械命令および / またはデータをプログラマブルプロセッサに提供するための任意の信号を意味する。

[0063]

ユーザとのやりとりを提供するために、本出願に記載されたシステムおよび技術は、ユーザに情報を表示するための表示装置(例えば、CRT(陰極線管)またはLCD(液晶ディスプレイ)モニタ)と、キーボードおよびポインティング装置(例えば、マウスまたはトラックボール)と、を有するコンピュータ上で実施されてもよく、ユーザは、キーボードおよびポインティング装置を介して入力をコンピュータに提供することができる。他の種類の装置はまた、ユーザとのやりとりを提供するために使用されてもよく、例えば、ユ

10

20

30

40

20

30

40

50

ーザに提供されるフィードバックは、任意の形態のセンサフィードバック(例えば、視覚フィードバック、聴覚フィードバック、または触覚フィードバック)であってもよく、ユーザからの入力は、任意の形態(音響入力、音声入力、または触覚入力を含む)で受信されてもよい。

[0064]

本出願に記載されたシステムおよび技術は、バックグラウンドコンポーネントを含むコンピュータシステム(例えば、アプリケーションサーバ)、または前側コンポーネントを含むコンピュータシステム(例えば、グラフィカルユーザインターフェースまたはウェブブラウザを有するユーザコンピュータ。ユーザは、グラフィカルユーザインターフェースまたはウェブブラウザを介して、本出願に記載されたシステムが表して、本出願に記載されたシステムがよび技術の実施形態と相互作用することができる。)、またはそのようなバックグラウンドコンポーネント、ミドルウェアコンポーネント、または前側コンポーネントの任意の組み合わせを含む、ミドルウェアコンポーネント、または前側コンポーネントの任意の組み合わせを含むにおいて実装されることができる。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(例えば、通信ネットワーク)によって相互に接続されることができる。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)、広域ネットワーク(WAN)、およびインターネットを含む。

[0065]

コンピュータシステムは、クライアントおよびサーバを含むことができる。クライアントおよびサーバは、通常互いに離れており、一般に通信ネットワークを介して相互作用する。クライアントとサーバの関係は、対応するコンピュータ上で実行され、互いにクライアント / サーバ関係にあるコンピュータプログラムによって生成される。

[0066]

本出願の実施形態の技術的解決手段によれば、各クライアントから送信された対応するキ ャラクタのオーディオデータストリームを受信し、各オーディオデータに基づいて、各オ ー ディ オ デ ー タ の キ ャ ラ ク タ 識 別 子 と 、 キ ャ ラ ク タ 識 別 子 に 対 応 す る 各 オ ー デ ィ オ デ ー タ の開始時刻とを確定し、受信した各オーディオデータを変換し、オーディオデータのキャ ラ ク タ 識 別 子 と オ ー デ ィ オ デ ー タ の 開 始 時 刻 と に 対 応 す る 各 テ キ ス ト 情 報 を 生 成 し 、 マ ー ジ動作指令を受信しかつマージ終了条件を満たしていないことに応答して、すべてのテキ スト情報をマージし、キャラクタ識別子と開始時刻に対応する各第1のテキストを生成し マージ終了条件を満たしたことに応答して、各第1のテキストをマージし、各第1のテ キストに対応する第2のテキストを生成することにより、複数のクライアントにおける異 なるキャラクタのオーディオデータを受信・変換したうえ、テキスト統合を経て送信情報 を確定する方法を実現した。各クライアントにおける異なるキャラクタのオーディオデー タをそれぞれ受信し、異なるキャラクタに応じて音声の採集と処理を行うことにより、従 来 技 術 に 見 ら れ る 同 じ 時 刻 に 複 数 人 が オン ラ イン で 話 す と 録 音 効 果 が 悪 く な る と い う 問 題 を解決した。キャラクタ別にオーディオを採集・変換し、オーディオデータのキャラクタ 識 別 子 と オ ー デ ィ オ デ ー タ の 開 始 時 刻 に 対 応 す る 各 テ キ ス ト 情 報 を 生 成 す る こ と に よ り 、 従来技術に見られる音声から変換した文字が乱れるという問題を解決した。この方法によ り、会議シーンにおいては、会議内容の可視化が可能となる。会議のオーディオをテキス ト 形 式 に 変 換 し て ク ラ イ ア ン ト に 送 信 し 、 テ キ ス ト 情 報 を ユ ー ザ に 提 示 す る こ と に よ り 、 ユーザは筋道がよく立って迅速に会議内容を振り返ることができる。ユーザの問合せ時間 を節約し、ユーザの読解体験を向上させることができる。

[0067]

なお、上述した様々な形態のフローを用いて、ステップを改めて並び替え、追加または削除を行うことができる。例えば、本出願に記載された各ステップは、本出願に開示された技術案の所望の結果が達成できる限り、並行して実行されてもよいし、順番で実行されてもよいし、異なる順番で実行されてもよい。本明細書はここで制限しない。

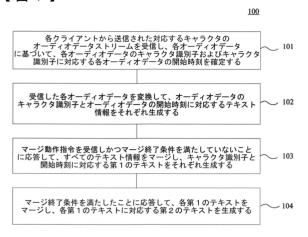
[0068]

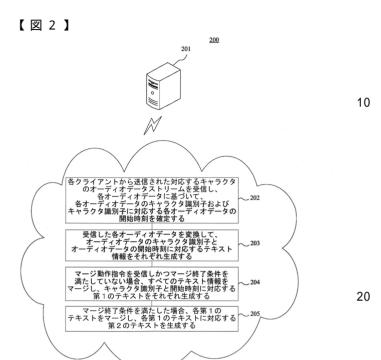
上記具体的な実施形態は、本出願の保護範囲を限定するものではない。設計要件および他

の要因に従って、様々な修正、組み合わせ、副次的な組み合わせ、および置換を行うことができることを当業者は理解すべきである。本出願の趣旨および原理を逸脱せずに行われたあらゆる修正、均等な置換および改善などは、いずれも本出願の保護範囲内に含まれるべきである。

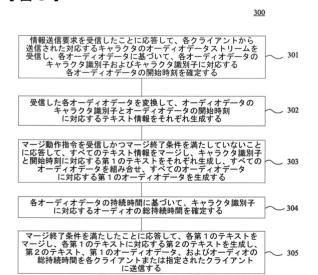
【図面】

【図1】

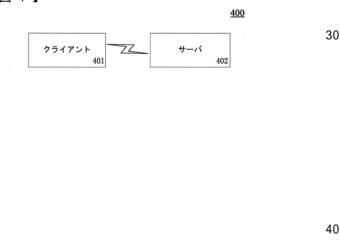




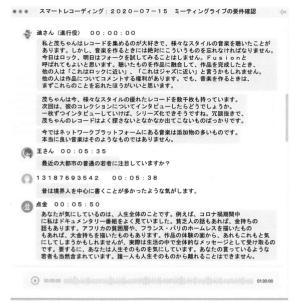
【図3】



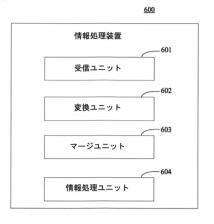
【図4】



【図5】

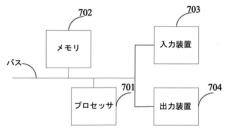


【図6】



10

【図7】



20

【外国語明細書】

2022050309000009.pdf

30

フロントページの続き

(72)発明者 シエ ハンメイ

中華人民共和国 100085 ベイジン ハイディエン ディストリクト シャンディー テンス ストリート ナンバー 10 バイドゥ キャンパス 2/エフ

(72)発明者 ファン フェイミン

中華人民共和国 100085 ベイジン ハイディエン ディストリクト シャンディー テンス ストリート ナンバー 10 バイドゥ キャンパス 2/エフ

(72)発明者 ジャオ フイビン

中華人民共和国 100085 ベイジン ハイディエン ディストリクト シャンディー テンス ストリート ナンバー 10 バイドゥ キャンパス 2/エフ

(72)発明者 ディン メイユェン

中華人民共和国 100085 ベイジン ハイディエン ディストリクト シャンディー テンス ストリート ナンバー 10 バイドゥ キャンパス 2/エフ

(72)発明者 フリナ

中華人民共和国 100085 ベイジン ハイディエン ディストリクト シャンディー テンス ストリート ナンバー 10 バイドゥ キャンパス 2/エフ