

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2017-146672  
(P2017-146672A)

(43) 公開日 平成29年8月24日(2017.8.24)

(51) Int.Cl.			F I			テーマコード (参考)
<b>G06F</b>	<b>3/16</b>	<b>(2006.01)</b>	G06F	3/16	610	5C058
<b>G10L</b>	<b>15/00</b>	<b>(2013.01)</b>	G10L	15/00	200U	5E555
<b>H04N</b>	<b>5/74</b>	<b>(2006.01)</b>	H04N	5/74	Z	
<b>G06F</b>	<b>3/0481</b>	<b>(2013.01)</b>	G06F	3/16	620	
			G06F	3/16	650	

審査請求 未請求 請求項の数 7 O L (全 28 頁) 最終頁に続く

(21) 出願番号 特願2016-26191 (P2016-26191)  
(22) 出願日 平成28年2月15日 (2016.2.15)

(71) 出願人 00005223  
富士通株式会社  
神奈川県川崎市中原区上小田中4丁目1番1号  
(74) 代理人 110002147  
特許業務法人酒井国際特許事務所  
(72) 発明者 高橋 潤  
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内  
(72) 発明者 田中 正清  
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内  
(72) 発明者 村瀬 健太郎  
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

最終頁に続く

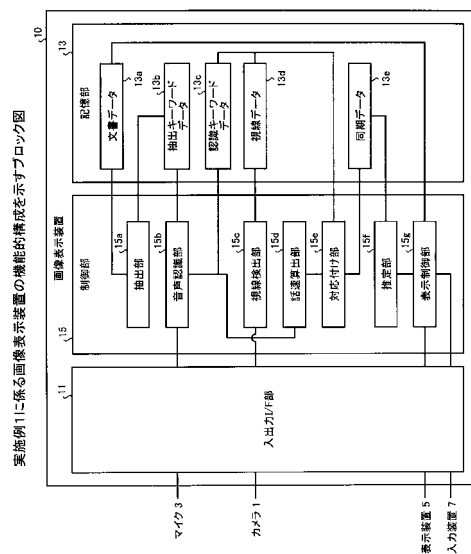
(54) 【発明の名称】 画像表示装置、画像表示方法、画像表示プログラム及び画像表示システム

(57) 【要約】

【課題】説明箇所の推定精度を向上させること。

【解決手段】画像表示装置10は、話速を算出する算出部と、音声認識により得られたキーワードと、話速とからキーワードの発話開始時刻を推定する第1推定部と、視線検出により検出された視線位置のうちキーワードの発話開始時刻に検出された視線位置とキーワードとに基づいて、表示装置に表示される文書データのページが分割された領域のうち説明箇所に対応する領域を推定する第2推定部と、説明箇所に対応する領域のハイライト表示を実行する表示制御部とを有する。

【選択図】 図1



## 【特許請求の範囲】

## 【請求項 1】

話速を算出する算出部と、

音声認識により得られたキーワードと、前記話速とから前記キーワードの発話開始時刻を推定する第 1 推定部と、

視線検出により検出された視線位置のうち前記キーワードの発話開始時刻に検出された視線位置と前記キーワードとに基づいて、表示装置に表示される文書データのページが分割された領域のうち説明箇所に対応する領域を推定する第 2 推定部と、

前記説明箇所に対応する領域のハイライト表示を実行する表示制御部とを有することを特徴とする画像表示装置。

10

## 【請求項 2】

前記第 2 推定部は、所定期間の視線位置から求まる視線の移動量のうち前記キーワードの発話開始時刻に対応する視線の移動量が所定の閾値以上である場合に、前記視線の移動量が前記閾値未満となるまで遡った時点の視線位置と前記キーワードとに基づいて、前記説明箇所に対応する領域を推定することを特徴とする請求項 1 に記載の画像表示装置。

## 【請求項 3】

前記算出部は、前記音声認識によりキーワードが認識される度に、前記話速を算出することを特徴とする請求項 1 または 2 に記載の画像表示装置。

## 【請求項 4】

前記算出部は、前記音声認識によりキーワードが認識される度に、前記キーワードが認識された時点から過去の所定期間の音声データを参照して、単位時間あたりに発話が行われたモーラ数を算出することを特徴とする請求項 3 に記載の画像表示装置。

20

## 【請求項 5】

話速を算出する処理と、

音声認識により得られたキーワードと、前記話速とから前記キーワードの発話開始時刻を推定する処理と、

視線検出により検出された視線位置のうち前記キーワードの発話開始時刻に検出された視線位置と前記キーワードとに基づいて、表示装置に表示される文書データのページが分割された領域のうち説明箇所に対応する領域を推定する処理と、

前記説明箇所に対応する領域のハイライト表示を実行する処理とがコンピュータにより実行されることを特徴とする画像表示方法。

30

## 【請求項 6】

話速を算出する処理と、

音声認識により得られたキーワードと、前記話速とから前記キーワードの発話開始時刻を推定する処理と、

視線検出により検出された視線位置のうち前記キーワードの発話開始時刻に検出された視線位置と前記キーワードとに基づいて、表示装置に表示される文書データのページが分割された領域のうち説明箇所に対応する領域を推定する処理と、

前記説明箇所に対応する領域のハイライト表示を実行する処理とをコンピュータに実行させることを特徴とする画像表示プログラム。

40

## 【請求項 7】

第 1 装置と第 2 装置とを有する画像表示システムであって、

前記第 1 装置は、

表示を行う表示装置と、

音声を入力するマイクと、

前記表示装置に対する視線を検出する視線検出部と、

前記マイクにより入力される音声データと、前記視線検出部により検出される視線位置とを前記第 2 装置へ送信する送信部とを有し、

前記第 2 装置は、

前記音声データに音声認識を行う音声認識部と、

50

話速を算出する算出部と、

前記音声認識部により認識されたキーワードと、前記話速とから、前記キーワードの発話開始時刻を推定する第1推定部と、

前記送信部により送信される視線位置のうち前記キーワードの発話開始時刻に検出された視線位置と前記キーワードとに基づいて、前記表示装置に表示される文書データのページが分割された領域のうち説明箇所に対応する領域を推定する第2推定部と、

前記表示装置に表示される前記ページのうち前記説明箇所に対応する領域をハイライト表示させる表示制御部と

を有することを特徴とする画像表示システム。

【発明の詳細な説明】

10

【技術分野】

【0001】

本発明は、画像表示装置、画像表示方法、画像表示プログラム及び画像表示システムに関する。

【背景技術】

【0002】

会議、講演やパンフレット紹介などの各種の場面では、複数人によって同一の内容の文書、例えば進捗アジェンダやスライド資料などに関する文書が共有された状態で会話によるコミュニケーションが行われる場合がある。

【0003】

20

例えば、会話によるコミュニケーションを支援する技術の一例として、音声認識を活用することにより、共有文書に含まれる内容のうち会話の内容に対応する部分を提示するシステムが実現されている。この他、視線検出の結果を利用して、共有文書に含まれる内容のうち会話の内容に対応する部分を提示することも提案されている。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開2010-256391号公報

【特許文献2】特開2009-206924号公報

【特許文献3】特開2014-222290号公報

30

【特許文献4】特開2010-262413号公報

【特許文献5】特開2014-085998号公報

【特許文献6】特開2004-007358号公報

【特許文献7】特開2005-338173号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

このように、音声認識または視線検出のいずれかが個別に利用することが提案されている。これら音声認識及び視線検出を併用することにより、話者が発話により説明を行う箇所を推定する精度を向上させることも考えられる。例えば、音声認識の結果として得られるキーワードに対応する箇所のうち視線検出により得られた視線の位置と最も近い箇所を説明箇所として抽出することが考えられる。

40

【0006】

しかしながら、音声認識及び視線検出を併用する場合、説明箇所の推定精度が低下する場合がある。

【0007】

すなわち、音声認識では、キーワードに対応する表音文字列の発話が終了した時点で始めて話者により発話されたキーワードが認識される。このように、キーワードの発話が終了した時点では、必ずしもキーワードに対応する箇所が話者によって注視されているとは限らない。なぜなら、一例として、話者が次に説明する箇所を目視により先行して確認し

50

ながら、それよりも前の箇所を読み上げて説明する状況も起こり得るからである。

【0008】

このような状況の下、発話中の箇所および次に発話予定の箇所の間で同一のキーワードが出現する場合、視線検出では次に発話予定の箇所から視線が検出される。この結果、発話中の箇所が説明箇所として抽出されず、誤って次に発話予定の箇所が説明箇所として抽出されることがある。

【0009】

1つの側面では、本発明は、説明箇所の推定精度を向上させることができる画像表示装置、画像表示方法、画像表示プログラム及び画像表示システムを提供することを目的とする。

【課題を解決するための手段】

【0010】

一態様では、画像表示装置は、話速を算出する算出部と、音声認識により得られたキーワードと、前記話速とから前記キーワードの発話開始時刻を推定する第1推定部と、視線検出により検出された視線位置のうち前記キーワードの発話開始時刻に検出された視線位置と前記キーワードとに基づいて、表示装置に表示される文書データのページが分割された領域のうち説明箇所に対応する領域を推定する第2推定部と、前記説明箇所に対応する領域のハイライト表示を実行する表示制御部とを有する。

【発明の効果】

【0011】

説明箇所の推定精度を向上させることができる。

【図面の簡単な説明】

【0012】

【図1】図1は、実施例1に係る画像表示装置の機能的構成を示すブロック図である。

【図2】図2は、視線の検出時刻とキーワードの認識時刻の関係の一例を示す図である。

【図3】図3は、実施例1に係る抽出キーワードデータの生成処理の手順を示すフローチャートである。

【図4】図4は、実施例1に係る音声認識処理の手順を示すフローチャートである。

【図5】図5は、実施例1に係る同期データの生成処理の手順を示すフローチャートである。

【図6】図6は、実施例1に係る表示制御処理の手順を示すフローチャートである。

【図7】図7は、視線の移動量の一例を示す図である。

【図8】図8は、実施例2に係る画像表示システムの構成例を示す図である。

【図9】図9は、画像表示サービスの会議システムへの適用例を示す図である。

【図10】図10は、画像表示サービスの会議システムへの適用例を示す図である。

【図11】図11は、実施例1及び実施例2に係る画像表示プログラムを実行するコンピュータのハードウェア構成例を示す図である。

【発明を実施するための形態】

【0013】

以下に添付図面を参照して本願に係る画像表示装置、画像表示方法、画像表示プログラム及び画像表示システムについて説明する。なお、この実施例は開示の技術を限定するものではない。そして、各実施例は、処理内容を矛盾させない範囲で適宜組み合わせることが可能である。

【実施例1】

【0014】

[画像表示装置が搭載する機能の一側面]

図1は、実施例1に係る画像表示装置の機能的構成を示すブロック図である。図1に示す画像表示装置10は、複数人によって同一の内容の文書、例えば進捗アジェンダやスライド資料などに関する文書に含まれるページ画面、例えばスライドのうち、話者の説明箇所に対応する部分をハイライト表示させる画像表示サービスを提供する。

10

20

30

40

50

## 【 0 0 1 5 】

かかる画像表示装置 10 では、上記の説明箇所を推定するために、音声認識および視線検出が併用される。このように音声認識および視線検出が併用される構成の下、画像表示装置 10 は、画像表示サービスの一環として、音声認識の結果としてキーワードが認識された時刻と話速とからキーワードの発話開始時刻を逆算し、当該発話開始時刻における視線検出位置とキーワードとに基づいて説明箇所を推定する。

## 【 0 0 1 6 】

このようにキーワードの発話開始時刻を逆算するのは、キーワードの発話が始まった時点では話者がキーワードに対応する説明箇所を注視している公算が高いからである。すなわち、一例として、説明を行う話者が説明箇所を先行して黙読してから説明箇所を音読する場合、キーワードの発話開始時点とキーワードの発話終了時点とでは、次のような違いがある。例えば、たとえ黙読の速度が音読の速度よりも大きいといえども、キーワードの発話開始時点では、説明箇所の黙読が終了しておらず、発話と視線が同じ説明箇所に向けられる可能性が高い。一方、キーワードの発話が終了した時点では、黙読の速度が音読の速度よりも大きいので、黙読が次の説明箇所に遷移するのに伴って視線も次の説明箇所に遷移している可能性が高い。

## 【 0 0 1 7 】

図 2 は、視線の検出時刻とキーワードの認識時刻の関係の一例を示す図である。図 2 の上段に示す時間軸には、視線検出に関する時間経過が示されており、話者の視線位置が検出された時刻のうち視線位置が説明箇所に始めて含まれた時刻が抜粋して白丸でプロットされると共に説明箇所に対応するインデックスが吹き出しで示されている。また、図 2 の下段に示す時間軸には、音声認識に関する時間経過が示されており、話者が発話することにより認識されたキーワードに対応する説明箇所のインデックスが丸印の中に記入されている。なお、図 2 の例では、説明箇所「A」、説明箇所「B」、説明箇所「C」の順にスライドの説明がなされる場合が想定されている。

## 【 0 0 1 8 】

ここで、視線検出では、視線検出が実行されるサンプリング周期ごとに視線位置が得られる。例えば、カメラ 1 により撮像される画像から話者の視線位置が検出される場合、画像のフレームレートごとに視線位置が得られる。一方、音声認識では、キーワードに対応する表音文字列の発話が終了するまではキーワードが認識されない。すなわち、音声認識が実行されるサンプリング周期ごとにキーワードの認識結果が得られる訳ではなく、キーワードに対応する表音文字列の発話が終了して始めてキーワードの認識結果が得られる。

## 【 0 0 1 9 】

このような状況の下、図 2 に示す例では、時刻  $t_1$  で視線位置が説明箇所「A」内で検出された後に時刻  $t_2$  で視線位置が説明箇所「A」内で検出される一方で、時刻  $t_3$  で説明箇所「A」に含まれるキーワードが認識される。これら視線検出および音声認識における時刻  $t_1$  及び時刻  $t_3$  の時間差は、音声認識および視線検出を単純に組み合わせた場合、説明箇所の推定精度の低下をまねく場合がある。

## 【 0 0 2 0 】

例えば、時刻  $t_3$  では、話者が次の説明箇所「B」を目視により先行して確認しながら、その前の説明箇所「A」を読み上げて説明している状態であるにもかかわらず、説明箇所「A」及び説明箇所「B」の間で同一のキーワードが出現する場合、説明箇所「A」に含まれるキーワードが認識された時刻  $t_3$  に最も近い時刻  $t_2$  に検出された視線位置が含まれる説明箇所「B」を組み合わせて推定が行われるおそれがある。

## 【 0 0 2 1 】

一方、説明箇所「A」及び説明箇所「B」の間で同一のキーワードが出現する場合であっても、説明箇所「A」に含まれるキーワードの発話が始まった時点では、説明箇所「A」の黙読が終了しておらず、説明箇所「A」に視線を向けている可能性が高いので、キーワードの発話開始時刻では、発話と視線が同じ説明箇所に向けられる可能性が高い。

## 【 0 0 2 2 】

10

20

30

40

50

このことから、発話開始時刻における視線検出位置とキーワードとに基づいて説明箇所を推定することで、発話と視線が同じ説明箇所に向けられた公算が高いタイミングの音声認識結果および視線検出結果を説明箇所の推定に用いることができる。したがって、説明箇所の推定精度を向上させることができる。

#### 【0023】

ここで、以下では、一例として、上記のハイライト表示に関する機能がプレゼンテーションソフトにアドオンされる場合を想定し、当該プレゼンテーションソフトを用いて作成された文書ファイルが含む1または複数のスライドを表示装置5に表示させることによってプレゼンテーションが進行される場合を想定する。かかるスライドには、テキストや図形を始め、他のアプリケーションプログラムによって作成されたコンテンツをインポート

10

#### 【0024】

##### [周辺機器]

図1に示すように、画像表示装置10には、カメラ1と、マイク3と、表示装置5と、入力装置7とが接続される。これらマイク3、表示装置5及び入力装置7などの周辺機器と、画像表示装置10との間は、有線または無線により接続される。

#### 【0025】

カメラ1は、CCD (Charge Coupled Device) やCMOS (Complementary Metal Oxide Semiconductor) などの撮像素子を搭載する撮像デバイスである。

20

#### 【0026】

一実施形態として、カメラ1は、視線検出用の画像を撮像するのに用いられる。ここでは、視線検出の目標とする話者の一例としてプレゼンタを想定し、表示装置5の一例としてプロジェクタがスクリーンへスライドの画像を投影することにより表示を実現する場合を想定する。この場合、カメラ1は、プレゼンタの顔の目の部分、すなわち角膜反射および瞳孔を所定の画素以上で撮像できる位置に設置される。このようなカメラ1の設置状況の下、カメラ1は、カメラ1の近傍に配置された図示しない光源を制御することによって被写体に赤外線を照射させ、被写体からの反射光を受光した上でデジタル信号に変換された被写体の画像を後述の視線検出部15cへ出力する。このとき、被写体に利用者の眼球

30

#### 【0027】

マイク3は、音声を電気信号に変換する装置である。ここで言う「マイク」は、マイクロフォンの略称である。

#### 【0028】

例えば、マイク3は、話者、例えばプレゼンテーションを実施するプレゼンタに装着させることができる。この場合、ヘッドセット型やタイピン型のマイクをプレゼンタの身体や衣服の所定位置に装着させたり、ハンド型のマイクをプレゼンタに携帯させたりすることができる。また、マイク3は、プレゼンタの発話が集音できる範囲の所定位置に設置することもできる。この場合、マイク3には、取付け型や据置き型のマイクを採用することもできる。これらいずれの場合においても、マイク3には、任意のタイプの指向性を持つマイクを採用できるが、プレゼンタの発話以外の音声、例えば聴講者等の発話や騒音などの雑音が集音されるのを抑制するために、マイクの感度をプレゼンタの発声方向に限定することもできる。なお、マイク3には、ダイナミック型、エレクトレットコンデンサ型、コンデンサ型などの任意の変換方式を採用することができる。このマイク3に音声を採取することにより得られたアナログ信号は、デジタル信号へ変換された上で画像表示装置10へ入力される。

40

#### 【0029】

表示装置5は、各種の情報を表示する装置である。

50

## 【 0 0 3 0 】

例えば、表示装置 5 には、発光により表示を実現する液晶ディスプレイや有機 E L (electroluminescence) ディスプレイなどを採用することもできるし、投影により表示を実現するプロジェクタを採用することもできる。また、表示装置 5 の設置台数は、必ずしも 1 台に限定されずともよく、複数の台数であってかまわない。以下では、一例として、プレゼンテーションの参加者であるプレゼンタ及び聴講者の両者が閲覧する共用の表示装置としてプロジェクタ及びプロジェクタが投影する画像を映すスクリーンが実装される場合を想定する。

## 【 0 0 3 1 】

この表示装置 5 は、一例として、画像表示装置 1 0 からの指示にしたがってプレゼンテーション画面を表示する。例えば、表示装置 5 は、画像表示装置 1 0 上で動作するプレゼンテーションソフトが開く文書ファイルのスライドを表示する。この場合、表示装置 5 は、文書ファイルが含むスライドのうちプレゼンタが入力装置 7 を介して指定する任意のスライドを表示させることもできるし、プレゼンテーションソフトが有するスライドショーの機能が ON 状態に設定された場合、各スライドが作成されたページ順に文書ファイルが含むスライドを切り替えて表示させることもできる。

10

## 【 0 0 3 2 】

入力装置 7 は、各種の情報に対する指示入力を受け付ける装置である。

## 【 0 0 3 3 】

例えば、表示装置 5 がプロジェクタとして実装される場合、スクリーンに映し出されたスライド上の位置を指し示すレーザーポインタを入力装置 7 として実装することができる。すなわち、レーザーポインタの中には、スライドのページを進めたり、戻したりする各種のボタンなどの操作部が設けられたリモコン機能付きのレーザーポインタも存在する。このリモコン機能付きのレーザーポインタが有する操作部を入力装置 7 として援用することもできる。この他、マウスやキーボードを入力装置 7 として採用したり、レーザーポインタによって指し示されたポインタの位置のセンシング、プレゼンタの視線検出やジェスチャ認識を行うためにスクリーンまたはプレゼンタの所定の部位が撮像された画像を入力する画像センサを入力装置 7 として採用することもできる。なお、表示装置 5 が液晶ディスプレイとして実装される場合、入力装置 7 には、液晶ディスプレイ上に貼り合わせられたタッチセンサを採用することもできる。

20

30

## 【 0 0 3 4 】

この入力装置 7 は、一例として、画像表示装置 1 0 上でプレゼンテーションソフトに実行させる文書ファイルの指定、スライドのページを進める操作やスライドのページを戻す操作などを受け付ける。このように入力装置 7 を介して受け付けられる操作は、画像表示装置 1 0 へ出力されることになる。

## 【 0 0 3 5 】

画像表示装置 1 0 は、上記の画像表示サービスを実行するコンピュータである。

## 【 0 0 3 6 】

一実施形態として、画像表示装置 1 0 には、デスクトップ型またはノート型のパーソナルコンピュータなどの情報処理装置を採用することができる。この他、画像表示装置 1 0 には、上記のパーソナルコンピュータなどの据置き型の端末のみならず、各種の携帯端末装置を採用することもできる。例えば、携帯端末装置の一例として、スマートフォン、携帯電話機や P H S (Personal Handyphone System) などの移動体通信端末、さらには、P D A (Personal Digital Assistants) などのスレート端末などがその範疇に含まれる。

40

## 【 0 0 3 7 】

なお、本実施例では、あくまで一例として、画像表示装置 1 0 が上記のプレゼンテーションソフトを外部のリソースに依存せずに単独で実行するスタンドアロンで上記の画像表示サービスを提供する場合を想定する。詳細は後述するが、上記の画像表示サービスは、スタンドアロンで提供される実装に限定されない。例えば、プレゼンテーションソフト

50

トを実行するクライアントに対し、上記の画像表示サービスを提供するサーバを設けることによってクライアントサーバシステムとして構築することができる。その他、プレゼンテーションソフトをサーバに実行させ、その実行結果をサーバがクライアントへ伝送して表示させるシンクライアントシステムとして構築したりすることもできる。

【 0 0 3 8 】

[ 画像表示装置 1 0 の構成 ]

続いて、本実施例に係る画像表示装置 1 0 の機能的構成について説明する。図 1 に示すように、画像表示装置 1 0 は、入出力 I / F ( InterFace ) 部 1 1 と、記憶部 1 3 と、制御部 1 5 とを有する。なお、図 1 には、データの入出力の関係を表す実線が示されているが、図 1 には、説明の便宜上、最小限の部分について示されているに過ぎない。すなわち、各処理部に関するデータの入出力は、図示の例に限定されず、図示以外のデータの入出力、例えば処理部及び処理部の間、処理部及びデータの間、並びに、処理部及び外部装置の間のデータの入出力が行われることとしてもかまわない。

10

【 0 0 3 9 】

入出力 I / F 部 1 1 は、マイク 3、表示装置 5 及び入力装置 7 などの周辺機器との間で入出力を行うインタフェースである。

【 0 0 4 0 】

一実施形態として、入出力 I / F 部 1 1 は、入力装置 7 から入力された各種の操作を制御部 1 5 へ出力する。また、入出力 I / F 部 1 1 は、制御部 1 5 から出力されたスライドの画像データを表示装置 5 へ出力したり、スライドに含まれる領域に対するハイライト指示またはそのキャンセル指示を表示装置 5 へ出力したりする。また、入出力 I / F 部 1 1 は、マイク 3 から入力された音声データを制御部 1 5 へ出力する。

20

【 0 0 4 1 】

記憶部 1 3 は、制御部 1 5 で実行される O S ( Operating System ) やプレゼンテーションソフトを始め、アプリケーションプログラムなどの各種プログラムに用いられるデータを記憶するデバイスである。

【 0 0 4 2 】

一実施形態として、記憶部 1 3 は、画像表示装置 1 0 における主記憶装置として実装される。例えば、記憶部 1 3 には、各種の半導体メモリ素子、例えば R A M ( Random Access Memory ) やフラッシュメモリを採用できる。また、記憶部 1 3 は、補助記憶装置として実装することもできる。この場合、H D D ( Hard Disk Drive )、光ディスクや S S D ( Solid State Drive ) などを採用できる。

30

【 0 0 4 3 】

記憶部 1 3 は、制御部 1 5 で実行されるプログラムに用いられるデータの一例として、文書データ 1 3 a、抽出キーワードデータ 1 3 b、認識キーワードデータ 1 3 c、視線データ 1 3 d 及び同期データ 1 3 e を記憶する。なお、上記の文書データ 1 3 a 以外の抽出キーワードデータ 1 3 b、認識キーワードデータ 1 3 c、視線データ 1 3 d 及び同期データ 1 3 e は、各データが参照または登録される場面で併せて説明することとする。また、記憶部 1 3 には、上記のデータ以外にも、他の電子データ、例えばハイライト表示に関する定義データなども併せて記憶することもできる。

40

【 0 0 4 4 】

文書データ 1 3 a は、文書に関するデータである。

【 0 0 4 5 】

一実施形態として、文書データ 1 3 a には、プレゼンテーションソフトを用いて 1 または複数のスライドが作成された文書ファイルを採用できる。かかるスライドには、テキストや図形を始め、他のアプリケーションプログラムによって作成されたコンテンツをインポートすることができる。例えば、ワープロソフトで作成された文書、表計算ソフトで作成された表やグラフをインポートしたり、撮像デバイスで撮像された画像や動画、さらには、画像編集ソフトで編集された画像や動画などをインポートしたりすることができる。このように、テキスト以外のコンテンツには、音声認識によるキーワード検索を実現する

50



ために、プレゼンテーションの開始前までに当該コンテンツの説明語句や説明文などの文字列を含むメタ情報を付与しておくことができる。

【0046】

制御部15は、各種のプログラムや制御データを格納する内部メモリを有し、これらによって種々の処理を実行するものである。

【0047】

一実施形態として、制御部15は、中央処理装置、いわゆるCPU (Central Processing Unit) として実装される。なお、制御部15は、必ずしも中央処理装置として実装されずともよく、MPU (Micro Processing Unit) やDSP (Digital Signal Processor) として実装されることとしてもよい。また、制御部15は、ASIC (Application Specific Integrated Circuit) やFPGA (Field Programmable Gate Array) などのハードワイヤードロジックによっても実現できる。 10

【0048】

制御部15は、各種のプログラム、例えば上記の画像表示サービスに対応する機能がアドオンされたプレゼンテーションソフトを実行することによって下記の処理部を仮想的に実現する。例えば、制御部15は、図1に示すように、抽出部15aと、音声認識部15bと、視線検出部15cと、話速算出部15dと、対応付け部15eと、推定部15fと、表示制御部15gとを有する。

【0049】

抽出部15aは、文書ファイルに含まれるスライドから音声認識で用いる辞書データに登録する単語を抽出キーワードデータ13bとして抽出する処理部である。 20

【0050】

一実施形態として、抽出部15aは、上記の抽出キーワードデータ13bを抽出する処理を自動的に開始することもできるし、手動設定で開始することもできる。例えば、自動的に開始する場合、プレゼンテーションソフトが文書ファイルを記憶部13に保存した上で閉じる場合、あるいはプレゼンテーションを介する文書ファイルの編集に文書ファイルが記憶部13に保存された場合に、処理を起動させることができる。また、手動設定で開始する場合、入力装置7を介してプレゼンテーションの前処理の実行指示を受け付けた場合に、処理を起動させることができる。いずれの場合においても、記憶部13に記憶された文書データ13aが含む文書ファイルのうち、保存または実行指示に対応する文書ファイルを読み出すことによって処理が開始される。 30

【0051】

かかる抽出キーワードデータ13bの生成について説明すると、抽出部15aは、記憶部13に記憶された文書データ13aが含む文書ファイルのうち保存が実行された文書ファイルあるいはプレゼンテーションの前処理の実行指示を受け付けた文書ファイルを読み出す。ここでは、一例として、抽出部15aが記憶部13から文書ファイルを読み出す場合を例示したが、文書ファイルの入手経路はこれに限定されない。例えば、抽出部15aは、ハードディスクや光ディスクなどの補助記憶装置またはメモリカードやUSB (Universal Serial Bus) メモリなどのリムーバブルメディアから文書ファイルを取得することもできる。また、抽出部15aは、外部装置からネットワークを介して受信することによって文書ファイルを取得することもできる。 40

【0052】

続いて、抽出部15aは、先に読み出した文書ファイルに含まれるスライドを複数の領域へ分割する。例えば、抽出部15aは、一文、行、段落などの単位でスライドを分割する。この場合、抽出部15aは、スライドが含む文字列を走査して、スペース、句点または改行に対応する区切り文字を検出し、当該区切り文字を領域の境界に設定する。かかる境界を前後に、抽出部15aは、スライドが含む文字列を区切る。これによって、スライドが複数の領域へ区切り文字ごとに分割される。その上で、抽出部15aは、スライドの分割によって得られた領域に当該領域を識別するインデックスを割り当てる。なお、ここでは、スライドを自動的に分割する場合を例示したが、入力装置7等を介して領域の境界 50

を指定させることによってスライドを手動設定で分割することとしてもかまわない。

【0053】

スライドの分割後に、抽出部15aは、当該スライドに含まれる複数の領域のうち領域を1つ選択する。続いて、抽出部15aは、先に選択された領域が含む文字列に対し、自然言語処理を実行することによって単語及びその読みを抽出する。例えば、抽出部15aは、領域内の文字列に形態素解析等を実行することにより得られた形態素のうち品詞が名詞である単語及び単語の読みを抽出する。その後、抽出部15aは、スライドが含む領域が全て選択されるまで上記の単語及び単語の読みの抽出を繰り返し実行する。その上で、抽出部15aは、単語、単語の読み及び単語が含まれる領域のインデックスが対応付けられた抽出キーワードデータ13bを記憶部13へ登録する。

10

【0054】

スライドの分割後に、抽出部15aは、当該スライドに含まれる複数の領域のうち領域を1つ選択する。続いて、抽出部15aは、先に選択された領域が含む文字列に対し、自然言語処理を実行することによって単語を抽出する。例えば、抽出部15aは、領域内の文字列に形態素解析等を実行することにより得られた形態素のうち品詞が名詞である単語や、文節を形成する単語などを抽出する。そして、抽出部15aは、先に抽出された各単語に当該単語が含まれる領域に割り当てられたインデックスを付与する。その後、抽出部15aは、スライドが含む領域が全て選択されるまで上記の単語の抽出及び上記のインデックスの付与を繰り返し実行する。

20

【0055】

このようにして全ての領域から単語が抽出された後に、抽出部15aは、スライドに含まれる単語ごとに当該単語kの文書中の出現回数を算出する。かかる文書中の出現回数は、一例として、単語kが文書中に出現する回数を集計することによって算出される。その上で、抽出部15aは、単語k、インデックスid<sub>x</sub>及び単語kの文書中の出現回数が対応付けられた抽出キーワードデータ13bを記憶部13へ登録する。

【0056】

音声認識部15bは、音声認識を実行する処理部である。

【0057】

一実施形態として、音声認識部15bは、プレゼンテーションソフトが文書ファイルを開いた状態でプレゼンテーションの開始指示を受け付けた場合に起動し、マイク3から所定時間長の音声信号が入力されるまで待機する。例えば、少なくとも1フレーム分の時間長、例えば10msの音声信号が入力されるのを待機する。そして、音声認識部15bは、マイク3から所定時間長の音声信号が入力される度に、当該音声信号にワードスポッティングなどの音声認識を実行する。このとき、音声認識部15bは、記憶部13に記憶された抽出キーワードデータ13bのうちプレゼンテーションソフトが実行中である文書ファイルが含むスライドであり、かつ表示装置5に表示中であるスライドに関する抽出キーワードデータ13bをワードスポッティングに適用する。これによって、音声認識部15bは、プレゼンタの発話の中に表示中のスライドに含まれる各領域から抽出された単語が存在するか否かを認識する。そして、音声認識部15bは、音声信号から単語の読みが認識された場合、当該単語及びその単語が認識された時間が対応付けられた認識キーワードデータ13cを記憶部13へ登録する。なお、同一の単語が時間経過に伴って複数回にわたって認識される場合には、最後、すなわち最新に認識された時刻が記憶部13へ登録される。

30

40

【0058】

これら一連の動作により、音声認識部15bは、表示中であるスライドの中でプレゼンタが説明中である可能性が高い単語を認識する。以下では、抽出キーワードデータ13bに含まれる単語のことを「抽出キーワード」と記載すると共に、認識キーワードデータ13cに含まれる単語のことを「認識キーワード」と記載し、互いのラベルを区別する場合がある。

【0059】

50

視線検出部 15c は、視線検出を実行する処理部である。

【0060】

一実施形態として、視線検出部 15c は、カメラ 1 から画像が取得される度に、当該画像に角膜反射法などのアルゴリズムを適用する。具体的には、視線検出部 15c は、カメラ 1 から取得された画像から楕円状の眼球の瞳孔の中心位置及び角膜の曲率中心を結ぶ視線方向、いわゆる視線ベクトルを検出する。その上で、視線検出部 15c は、先に検出された視線ベクトルから表示装置 5 のスクリーン上の注視点、例えばスクリーン上の座標を算出する。ここでは、あくまで一例として、角膜反射法を用いて視線検出を実現する場合を例示したが、これ以外の方法を用いて視線検出を実現することもできる。例えば、視線検出部 15c は、表示装置 5 のスクリーンを分割し、分割した領域を見る眼の形状を学習しておき、カメラ 1 から入力される被写体の画像から検出される眼の形状とテンプレートマッチングを実行することによって視線方向を検出することもできる。また、視線検出部 15c は、プレゼンタが装着するヘッドセットによって検出された視線方向を取得することとしてもかまわない。

10

【0061】

このように注視点が算出された後、視線検出部 15c は、視線検出が実行された時刻および視線検出の結果として得られた注視点が対応付けられたエントリを視線データ 13d の 1レコードとして記憶部 13へ追加登録する。以下では、音声認識部 15b によりキーワードが認識される時刻と、視線検出部 15c により注視点が検出された時刻とを区別する観点から、前者のことを「認識時刻」と記載すると共に、後者のことを「検出時刻」と記載する場合がある。

20

【0062】

なお、図 1 には、視線検出部 15c がソフトウェアにより実装される場合を例示したが、制御部 15 等のプロセッサとは別のハードウェアとして実装することもできる。例えば、視線検出部 15c に対応する機能を有するマイコン等とカメラ 1 とが一体で形成された小型視線センサとして実装することもできる。

【0063】

話速算出部 15d は、話速を算出する処理部である。

【0064】

一実施形態として、話速算出部 15d は、記憶部 13 に記憶された認識キーワードデータ 13c に対する更新の有無を監視する。そして、話速算出部 15d は、認識キーワードデータ 13c に新たな認識キーワードが追加される度に、話者の話速を算出する。例えば、話速算出部 15d は、当該認識キーワードが記憶部 13 に登録された時点から過去の所定期間、例えば 3 秒間の音声データに含まれる母音数をもとに、単位時間、例えば 1 秒間あたりに発話されたモーラ数を「話速」として算出する。このように、話速算出部 15d は、認識キーワードが追加される度に話速を逐次的に算出する。これによって、一側面として、話者による説明の速度が変化してもそれに追従して上記の発話開始時刻を推定することが可能になる。

30

【0065】

なお、ここでは、一例として、単位時間あたりに発話されたモーラ数を話速として算出する場合を例示したが、話速の算出方法はこれに限定されない。例えば、音声認識の対象とする言語が英語等である場合、単位時間あたりに発話された音節の数を話速として算出することもできる。また、上記の話速は、必ずしも認識キーワードが追加される度に話速を逐次的に算出せずともかまわない。例えば、プレゼンタの話速を固定値で設定しておき、かかる話速を新たな認識キーワードが追加される度に参照することもできる。

40

【0066】

対応付け部 15e は、音声認識結果及び視線検出結果の対応付けを行う処理部である。

【0067】

一実施形態として、対応付け部 15e は、話速算出部 15d により話速が算出された場合、認識キーワードデータ 13c に追加された認識キーワードと、話速算出部 15d によ

50

り算出された話速とから、当該認識キーワードの発話が始まった時刻、すなわち上記の発話開始時刻を推定する。例えば、対応付け部 15 e は、下記の式 (1) にしたがって上記の「発話開始時刻」を推定する。

【0068】

発話開始時刻 = 認識時刻 - 反応時間 - 発話所要時間 - 認識処理時間・・・(1)

【0069】

ここで、上記の式 (1) に記載の「認識時刻」とは、認識キーワードが音声認識部 15 b により認識された時刻を指す。また、「反応時間」とは、話者がスライド上のキーワードに視線を向けてから当該キーワードの内容を視認して次の動作、例えばキーワードの表音文字列の読み上げなどを行うまでの反応に費やす時間を指す。例えば、反応時間には、200 msec 程度の値を設定することができる。さらに、「発話所要時間」とは、認識キーワードに対応する表音文字列の発話を開始してから当該表音文字列の発話が終了するまでの時間を指す。例えば、対応付け部 15 e は、認識キーワードに対応する表音文字列のモーラ数を話速算出部 15 d により算出された話速で乗算することにより、上記の「発話所要時間」を算出することができる。また、「認識処理時間」とは、話者によるキーワードの発話というイベントが発生してから当該キーワードが音声認識部 15 b により認識されるまでの音声認識処理の所要時間を指す。これは、一側面として、音声認識エンジンが実行されるハードウェア、例えば CPU やメモリの性能に依存する。よって、CPU やメモリの性能が事前測定等により既知である場合、CPU やメモリの性能にしたがって上記の「認識処理時間」を設定しておくことができる。

【0070】

これら「認識時刻」、「反応時間」、「発話所要時間」及び「認識処理時間」を上記の式 (1) に代入することにより、対応付け部 15 e は、認識キーワードの「発話開始時刻」を求めることができる。その後、対応付け部 15 e は、記憶部 13 に記憶された視線データ 13 d に含まれる注視点のうち先に求められた「発話開始時刻」に対応する「検出時刻」と紐付けられた注視点を抽出する。その上で、対応付け部 15 d は、「認識時刻」、「認識キーワード」及び「注視点」などの項目が対応付けられたエントリを同期データ 13 e の 1 レコードとして記憶部 13 へ追加登録する。

【0071】

推定部 15 f は、同期データ 13 e を参照して、説明箇所を推定する処理部である。

【0072】

一実施形態として、推定部 15 f は、記憶部 13 に記憶された同期データ 13 e のうち推定部 15 f が記憶部 13 を参照する時刻から過去に遡って所定期間、例えば 3 秒間までに生成されたレコードを読み出す。続いて、推定部 15 f は、先に読み出されたレコードのうちレコードを 1 つ選択する。そして、推定部 15 f は、先に選択されたレコードに含まれる注視点が表示中のスライド上で属する領域と、当該レコードに含まれる認識キーワードが表示中のスライド上で属する領域との 2 つの領域が一致するか否かを判定する。

【0073】

例えば、推定部 15 f は、表示中のスライドに含まれる領域ごとに当該領域の内容が表示装置 5 のスクリーン上で表示される表示範囲が定義された表示範囲データを参照することにより、注視点が表示中のスライドに含まれる領域のうちいずれの領域に属するのかを識別することができる。かかる表示範囲データは、一例として、抽出部 15 a により設定された領域の形状及び大きさを定義する情報と、表示装置 5 のスクリーンのサイズとから生成することができる。例えば、表示中のスライドに含まれる領域の形状が矩形であるとしたとき、領域ごとに当該領域に関する矩形の対角を形成する 2 つの頂点の座標、あるいは矩形を形成する 1 つの頂点と矩形の幅及び高さなどが対応付けられた表示範囲データが生成される。これを参照することにより、推定部 15 f は、注視点の座標が表示中のスライドに含まれる領域のうちいずれの領域に属するのかを識別できる。例えば、注視点がいずれかの矩形に含まれる場合、当該矩形の領域に注視点が属すると判断できる。また、注視点がいずれの領域の外である場合でも、注視点から所定の距離以内に存在する矩形が存

10

20

30

40

50

在する場合、当該矩形の領域に注視点が属すると判断することもできる。さらに、注視点から所定の距離以内に複数の矩形が存在する場合、注視点から近い方または注視点から所定の距離以内のエリアと多く重複する方の矩形の領域に属すると判断することもできる。

【0074】

一方、推定部15fは、抽出キーワードデータ13bを参照することにより、認識キーワードが表示中のスライド上で属する領域を識別する。上記の抽出部15aの説明で述べたとおり、抽出キーワードデータ13bには、抽出キーワードと領域のインデックスが対応付けられているので、認識キーワードと同一である抽出キーワードに紐付けられた領域のインデックスを検索することにより、認識キーワードが表示中のスライド上で属する領域も識別することができる。

10

【0075】

ここで、2つの領域が一致する場合、音声認識および視線検出の両面で説明箇所の推定結果が一致していると判断できる。この場合、2つの領域が一致しない場合よりも、当該同期データ13eのレコードから推定された説明箇所が尤もらしいとみなすことができる。よって、推定部15fは、表示中のスライドに含まれる領域ごとに設けられたカウンタのうち音声認識および視線検出の両方の推定結果が一致する領域のカウンタを対象に、当該カウンタを1つインクリメントする。これによって、表示中のスライドに含まれる領域ごとに音声認識および視線検出の両方の推定結果が一致する回数を計数することができる。一方、2つの領域が一致しない場合、音声認識による説明箇所の推定結果と視線検出による説明箇所の推定結果が一致していないと判断できる。この場合、いずれの領域に関するカウンタもインクリメントされない。その後、推定部15fは、記憶部13から読み出された全てのレコードが選択されると、表示中のスライドに含まれる領域のうちカウンタの値が最大である領域を説明箇所として推定する。

20

【0076】

表示制御部15gは、表示装置5に対する表示制御を実行する処理部である。なお、ここでは、表示制御部15gが実行する表示制御のうち、スライドに関する表示制御と、ハイライトに関する表示制御との一側面について説明する。

【0077】

[スライドの表示制御]

一側面として、表示制御部15gは、プレゼンテーションソフトにより文書ファイルが開かれた場合、当該文書ファイルが含むスライドを表示装置5に表示させる。このとき、表示制御部15gは、文書ファイルが含むスライドのうち最初のページのスライドを表示させることとしてもよいし、最後に編集が行われたページのスライドを表示させることとしてもよい。その後、表示制御部15gは、プレゼンテーションの開始指示を受け付けた後、推定部15fにより推定された説明箇所に対応する領域に関するスライドを表示装置5に表示させる。また、表示制御部15gは、入力装置7を介してページの切替え指示を受け付けた場合、表示装置5に表示させるスライドを変更する。例えば、ページを進める操作を受け付けた場合、表示制御部15gは、表示中のスライドの次ページのスライドを表示装置5に表示させる。また、ページを戻す操作を受け付けた場合、表示制御部15gは、表示中のスライドの前ページのスライドを表示装置5に表示させる。

30

40

【0078】

[ハイライトの表示制御]

他の一側面として、表示制御部15gは、プレゼンテーションの開始指示を受け付けてからプレゼンテーションの終了指示を受け付けるまで下記の処理を繰り返し実行する。すなわち、表示制御部15gは、推定部15fにより推定された説明箇所の領域のハイライト表示を実行する。ここで言う「ハイライト表示」は、狭義のハイライト表示、すなわち背景色を明るくしたり、反転したりする表示制御に留まらず、広義のハイライト表示を意味する。例えば、説明箇所の囲み表示、説明箇所の塗りつぶしの強調、フォント（フォントサイズ、下線や斜体）の強調などのように、強調表示の全般を任意に実行することができる。なお、ハイライト表示は、入力装置7を介してキャンセル操作を受け付けた場合に

50

通常表示へ戻すこととしてもかまわない。また、当然のことながら、推定部 15 f によりいずれの領域も説明箇所として出力されない場合には、表示中のスライド上でハイライト表示は実行されない。

【0079】

[ 処理の流れ ]

次に、本実施例に係る画像表示装置 10 の処理の流れについて説明する。なお、ここでは、画像表示装置 10 が実行する ( 1 ) 抽出キーワードデータの生成処理、( 2 ) 音声認識処理、( 3 ) 同期データの生成処理、( 4 ) 表示制御処理の順に説明することとする。

【0080】

( 1 ) 抽出キーワードデータの生成処理

図 3 は、実施例 1 に係る抽出キーワードデータの生成処理の手順を示すフローチャートである。この処理は、自動的に開始することもできるし、手動設定で開始することもできる。例えば、自動的に開始する場合、プレゼンテーションソフトが文書ファイルを記憶部 13 に保存した上で閉じる場合、あるいはプレゼンテーションを介する文書ファイルの編集集中に文書ファイルが記憶部 13 に保存された場合に、処理を起動させることができる。また、手動設定で開始する場合、入力装置 7 を介してプレゼンテーションの前処理の実行指示を受け付けた場合に、処理を起動させることができる。いずれの場合においても、記憶部 13 に記憶された文書データ 13 a が含む文書ファイルのうち、保存または実行指示に対応する文書ファイルを読み出すことによって処理が開始される。

【0081】

図 3 に示すように、抽出部 15 a は、文書ファイルに含まれるスライドを一文、行または段落などの単位で複数の領域へ分割する ( ステップ S 101 )。続いて、抽出部 15 a は、ステップ S 101 で得られた領域に各領域を識別するインデックスを割り当てる ( ステップ S 102 )。

【0082】

そして、抽出部 15 a は、ステップ S 102 で割り当てられたインデックスのうちインデックスを 1 つ選択する ( ステップ S 103 )。続いて、抽出部 15 a は、ステップ S 103 で選択されたインデックスの領域内の文字列に形態素解析等を実行することにより得られた形態素のうち品詞が名詞である単語及びその読みを抽出する ( ステップ S 104 )。

【0083】

そして、抽出部 15 a は、ステップ S 102 で割り当てられたインデックスが全て選択されるまで ( ステップ S 105 No )、上記のステップ S 103 及びステップ S 104 の処理を繰り返し実行する。

【0084】

その後、ステップ S 102 で割り当てられたインデックスが全て選択された場合 ( ステップ S 105 Yes )、抽出部 15 a は、単語、単語の読み及び単語が含まれる領域のインデックスが対応付けられた抽出キーワードデータ 13 a を記憶部 13 へ登録し ( ステップ S 106 )、処理を終了する。

【0085】

( 2 ) 音声認識処理

図 4 は、実施例 1 に係る音声認識処理の手順を示すフローチャートである。この音声認識処理は、プレゼンテーションソフトが文書ファイルを開いた状態でプレゼンテーションの開始指示を受け付けた場合に起動し、プレゼンテーションの終了指示を受け付けるまで繰り返し実行される。なお、音声認識処理は、視線検出部 15 c による図示しない視線検出処理と並列して実行することができ、視線検出処理と同一のサンプリング周波数で実行することもできるし、視線検出処理と異なるサンプリング周波数で実行することもできる。

【0086】

図 4 に示すように、音声認識部 15 b は、マイク 3 から所定時間長の音声信号が入力されるまで、例えば少なくとも 1 フレーム分の時間長、例えば 10 msec の音声信号が入

10

20

30

40

50

力されるまで待機する（ステップ S 3 0 1）。

【 0 0 8 7 】

そして、マイク 3 から所定時間長の音声信号が入力されると（ステップ S 3 0 1 Y e s）、音声認識部 1 5 b は、当該音声信号にワードスポッティングなどの音声認識を実行する（ステップ S 3 0 2）。かかるステップ S 3 0 2 でワードスポッティングが実行される場合には、記憶部 1 3 に記憶された抽出キーワードデータ 1 3 b のうちプレゼンテーションソフトが実行中である文書ファイルが含むスライドであり、かつ表示装置 5 に表示中であるスライドに関する抽出キーワードデータが辞書データとして適用される。

【 0 0 8 8 】

このとき、音声信号から単語が認識された場合（ステップ S 3 0 3 Y e s）、音声認識部 1 5 b は、ステップ S 3 0 2 で認識された単語及びその単語が認識された時間が対応付けられた認識キーワードデータ 1 3 c を記憶部 1 3 へ登録し（ステップ S 3 0 4）、ステップ S 3 0 1 の処理へ戻る。

10

【 0 0 8 9 】

一方、マイク 3 から所定時間長の音声信号が入力されていない場合、あるいは音声信号から単語が認識されなかった場合（ステップ S 3 0 1 N o またはステップ S 3 0 3 N o）、認識キーワードデータ 1 3 c の登録を実行せずに、ステップ S 3 0 1 の処理へ戻る。

【 0 0 9 0 】

（ 3 ）同期データの生成処理

図 5 は、実施例 1 に係る同期データの生成処理の手順を示すフローチャートである。この処理は、一例として、認識キーワードデータ 1 3 c に新たな認識キーワードが追加される度に実行される。

20

【 0 0 9 1 】

図 5 に示すように、認識キーワードデータ 1 3 c に新たな認識キーワードが追加されると（ステップ S 5 0 1 Y e s）、話速算出部 1 5 d は、ステップ S 5 0 1 で追加された認識キーワードが記憶部 1 3 に登録された時点から過去の所定期間の音声データに含まれる母音数をもとに、単位時間あたりに発話されたモーラ数を「話速」として算出する（ステップ S 5 0 2）。

【 0 0 9 2 】

続いて、対応付け部 1 5 e は、ステップ S 5 0 2 で算出された話速と、認識キーワードデータ 1 3 c に追加された認識キーワードとから、当該認識キーワードの発話開始時刻を推定する（ステップ S 5 0 3）。例えば、ステップ S 5 0 2 で算出された話速と、認識キーワードデータ 1 3 c に追加された認識キーワードとを始め、予め与えられた他の情報をもとに上記の式（ 1 ）に関する「認識時刻」、「反応時間」、「発話所要時間」及び「認識処理時間」を求めた後、これらを上記の式（ 1 ）に代入することにより、認識キーワードの「発話開始時刻」を求めることができる。

30

【 0 0 9 3 】

その後、対応付け部 1 5 e は、記憶部 1 3 に記憶された視線データ 1 3 d に含まれる注視点のうち先に求められた「発話開始時刻」に対応する「検出時刻」と紐付けられた注視点を抽出する（ステップ S 5 0 4）。例えば、対応付け部 1 5 e は、「発話開始時刻」と最も時間差が少ない「検出時刻」と紐付けられた注視点を抽出することもできるし、「発話開始時刻」の前または後のいずれかの「検出時刻」と紐付けられた注視点だけを抽出することもできる。

40

【 0 0 9 4 】

その上で、対応付け部 1 5 d は、「認識時刻」、「認識キーワード」及び「注視点」などの項目が対応付けられたエントリを同期データ 1 3 e の 1 レコードとして記憶部 1 3 へ追加登録し（ステップ S 5 0 5）、ステップ S 5 0 1 の処理へ戻る。

【 0 0 9 5 】

（ 4 ）表示制御処理

図 6 は、実施例 1 に係る表示制御処理の手順を示すフローチャートである。この表示制

50

御処理は、一例として、表示装置 5 にスライドが表示される限り、繰り返し実行される。図 6 に示すように、推定部 15 f は、記憶部 13 に記憶された同期データ 13 e のうち推定部 15 f が記憶部 13 を参照する時刻から過去に遡って所定期間、例えば 3 秒間までに生成されたレコードを読み出す（ステップ S 701）。

【0096】

続いて、推定部 15 f は、ステップ S 701 で読み出されたレコードのうちレコードを 1 つ選択する（ステップ S 702）。そして、推定部 15 f は、ステップ S 702 で選択されたレコードに含まれる注視点が表示中のスライド上で属する第 1 の領域と、当該レコードに含まれる認識キーワードが表示中のスライド上で属する第 2 の領域との 2 つの領域が一致するか否かを判定する（ステップ S 703）。

10

【0097】

これら第 1 の領域および第 2 の領域が一致する場合（ステップ S 703 Yes）、音声認識および視線検出の両面で説明箇所の推定結果が一致していると判断できる。この場合、2 つの領域が一致しない場合よりも、当該同期データ 13 e のレコードから推定された説明箇所が尤もらしいとみなすことができる。よって、推定部 15 f は、表示中のスライドに含まれる領域のうち音声認識および視線検出の両方の推定結果が一致する領域のカウントを対象に、音声認識および視線検出の両方の推定結果が一致する回数を計数するカウントを 1 つインクリメントする（ステップ S 704）。

【0098】

一方、第 1 の領域および第 2 の領域が一致しない場合（ステップ S 703 No）、音声認識による説明箇所の推定結果と視線検出による説明箇所の推定結果が一致していないと判断できる。この場合、いずれの領域に関するカウントもインクリメントされず、ステップ S 705 の処理へ移行する。

20

【0099】

その後、ステップ S 701 で読み出された全てのレコードが選択されるまで（ステップ S 705 No）、上記のステップ S 702 から上記のステップ S 704 までの処理が繰り返し実行される。

【0100】

そして、ステップ S 701 で読み出された全てのレコードが選択された場合（ステップ S 705 Yes）、推定部 15 f は、表示中のスライドに含まれる領域のうちカウントの値が最大である領域を説明箇所として推定する（ステップ S 706）。その上で、表示制御部 15 g は、推定部 15 f により推定された説明箇所の領域をハイライト表示し（ステップ S 707）、ステップ S 701 の処理へ移行する。

30

【0101】

なお、図 6 に示すフローチャートでは、カウントの値が最大である領域を説明箇所として推定する場合を例示したが、カウントの値が所定の閾値以上である領域を説明箇所として推定することもできる。この場合、説明箇所として推定される領域は必ずしも 1 つであるとは限らず、複数の領域が説明箇所として推定される場合もある。

【0102】

[効果の一側面]

40

上述してきたように、本実施例に係る画像表示装置 10 は、音声認識の結果としてキーワードが認識された時刻と話速とからキーワードの発話開始時刻を逆算し、当該発話開始時刻における視線検出位置及びキーワードに基づいて説明箇所を推定する。それ故、発話と視線が同じ説明箇所に向けられた公算が高いタイミングの音声認識結果および視線検出結果を説明箇所の推定に用いることができる。したがって、本実施例に係る画像表示装置 10 によれば、説明箇所の推定精度を向上させることができる。

【実施例 2】

【0103】

さて、これまで開示の装置に関する実施例について説明したが、本発明は上述した実施例以外にも、種々の異なる形態にて実施されてよいものである。そこで、以下では、本発

50



明に含まれる他の実施例を説明する。

【 0 1 0 4 】

[ 文書ファイルの応用例 ]

上記の実施例 1 では、プレゼンテーションソフトによって作成された文書を用いる場合を例示したが、他のアプリケーションプログラムによって作成された文書を用いることもできる。すなわち、表示時に画面単位で表示されるページを含む文書ファイルであれば、ワープロソフトの文書ファイルが有するページをスライドに読み替えたり、表計算ソフトの文書ファイルが有するシートをスライドに読み替えることによって図 3 ~ 図 6 に示した処理を同様に適用できる。

【 0 1 0 5 】

10

[ 対応付けの応用例 1 ]

上記の画像表示装置 1 0 は、ステップ S 5 0 3 等で推定されるキーワードの発話開始時刻に対応する視線データの動きが大きい場合、視線の動きが大きくなる前の視線データを利用するように補正処理を行うこともできる。例えば、画像表示装置 1 0 は、所定の期間、例えば 1 0 0 m s e c が経過する度に、所定の期間の開始時点でカメラ 1 から取得された画像から検出された第 1 の注視点と、所定の期間の終了時点でカメラ 1 から取得された画像から検出された第 2 の注視点との距離を算出する。その上で、画像表示装置 1 0 は、第 1 の注視点および第 2 の注視点の間の距離が所定の閾値、例えばスライドの幅または高さの半分以上である場合、所定期間に含まれる検出時刻に視線の移動量「大」のラベルが付与される一方で、第 1 の注視点および第 2 の注視点の間の距離が所定の閾値未満である場合、所定期間に含まれる検出時刻に視線の移動量「小」のラベルが付与される。このように各注視点が検出される検出時刻に視線の移動量「小」または「大」のラベルがさらに対応付けられたエントリが視線データ 1 3 d のレコードとして生成される。

20

【 0 1 0 6 】

図 7 は、視線の移動量の一例を示す図である。図 7 に示すように、キーワードの発話開始時刻  $t_1$  に対応する検出時刻に視線の移動量「大」のラベルが付与されている場合、話者の視線がある領域から次に説明する領域へ移っている可能性が高い。この場合、画像表示装置 1 0 は、キーワードの発話開始時刻  $t_1$  よりも前の検出時刻であって始めて視線の移動量「小」のラベルが付与されている検出時刻  $t_0$  に紐付けられた注視点の座標がキーワード及び発話開始時刻と対応付けられたエントリを同期データ 1 3 e の 1 レコードとして記憶部 1 3 へ追加登録する。なお、図 7 には示されていないが、キーワードの発話開始時刻に対応する検出時刻に視線の移動量「小」のラベルが付与されている場合、画像表示装置 1 0 は、キーワードの発話開始時刻  $t_1$  に対応する検出時刻に紐付けられた注視点の座標がキーワード及び発話開始時刻と対応付けられたエントリを同期データ 1 3 e の 1 レコードとして記憶部 1 3 へ追加登録する。

30

【 0 1 0 7 】

[ 対応付けの応用例 2 ]

さらに、上記の画像表示装置 1 0 は、キーワードの発話開始時刻  $t_1$  に対応する検出時刻に視線の移動量「大」のラベルが付与されていることを条件に、視線の移動量「小」のラベルが付与されている検出時刻まで遡って注視点に対応付ける場合を例示したが、さらなる加重要件を設けることとしてもよい。例えば、画像表示装置 1 0 は、音声認識部 1 5 b により認識された認識キーワードが当該認識キーワードを含む領域内の末尾、例えば領域の文字全体の後ろ 1 0 % 以内に含まれる場合、かつキーワードの発話開始時刻  $t_1$  に対応する検出時刻に視線の移動量「大」のラベルが付与されている場合に始めて、視線の移動量「小」のラベルが付与されている検出時刻まで遡って注視点に対応付けることとしてもよい。このような加重要件を設定することにより、話者の視線がある領域から次に説明する領域へ移っている可能性がより高い状況下でキーワードと対応付ける注視点の検出時刻を補正できる。

40

【 0 1 0 8 】

[ 分散および統合 ]

50

また、図示した各装置の各構成要素は、必ずしも物理的に図示の如く構成されておらずともよい。すなわち、各装置の分散・統合の具体的形態は図示のものに限られず、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。例えば、抽出部 15 a、音声認識部 15 b、視線検出部 15 c、話速算出部 15 d、対応付け部 15 e、推定部 15 f または表示制御部 15 g を画像表示装置 10 の外部装置としてネットワーク経由で接続するようにしてもよい。また、抽出部 15 a、音声認識部 15 b、視線検出部 15 c、話速算出部 15 d、対応付け部 15 e、推定部 15 f または表示制御部 15 g を別の装置がそれぞれ有し、ネットワーク接続されて協働することで、上記の画像表示装置 10 の機能を実現するようにしてもよい。

10

#### 【0109】

##### [他の実装例]

上記の実施例 1 では、画像表示装置 10 が上記のプレゼンテーションソフトを外部のリソースに依存せずに単独で実行するスタンドアロンで図 3 ~ 図 6 に関する処理を実行する場合を例示したが、他の実装形態を採用することもできる。例えば、プレゼンテーションソフトを実行するクライアントに対し、図 3 ~ 図 6 に関する処理のうち一部または全部の処理を実行するサーバを設けることによってクライアントサーバシステムとして構築することもできる。この場合、パッケージソフトウェアやオンラインソフトウェアとして上記の画像表示サービスを実現する画像表示プログラムをインストールさせることによってサーバ装置を実装できる。例えば、サーバ装置は、上記の画像表示サービスを提供する W e bサーバとして実装することとしてもよいし、アウトソーシングによって上記の画像表示サービスを提供するクラウドとして実装することとしてもかまわない。この場合、クライアントは、サーバ装置に対し、ハイライト表示の開始指示、例えば少なくともプレゼンテーションに用いる文書ファイルを指定する情報をアップロードした後に、プレゼンテーションが開始される。プレゼンテーションが開始されると、クライアントは、マイク 3 から採取された音声信号または音声認識処理の結果とカメラ 1 により撮像された画像データまたは視線検出処理の結果とをアップロードし、表示装置 5 に表示中のスライドのページが切り替わる度にスライドのページ情報をアップロードする。すなわち、抽出キーワードデータの生成処理、音声認識処理や視線検出処理は、クライアント側で実行させることもできるし、サーバ側で実行させることとしてもかまわない。これによって、サーバ装置は、少なくとも図 5 や図 6 に示した処理が実行可能となる。さらに、クライアントは、図示しない入力デバイスに関する操作情報をサーバへ伝送し、サーバから伝送される処理結果だけを表示装置 5 に表示させることにより、シンクライアントシステムとして構築することもできる。この場合には、各種のリソース、例えば文書データもサーバにより保持されると共に、プレゼンテーションソフトもサーバで仮想マシンとして実装されることになる。例えば、プレゼンテーションソフトがクライアント側で実行される場合、サーバからクライアントへハイライト表示を実施する領域の識別情報、例えば上記の領域のインデックスを伝送すればよく、また、シンクライアントシステムとして実装される場合、説明箇所のハイライト表示が実施されたスライドの表示データまたはハイライト表示が行われる前の画面との差分データをサーバからクライアントへ伝送すればよい。なお、上記の実施例 1 では、画像表示プログラムがプレゼンテーションソフトにアドオンされる場合を想定したが、ライセンス権限を有するクライアントから画像表示プログラムをライブラリとして参照する要求を受け付けた場合に、画像表示プログラムをプラグインさせることもできる。

20

30

40

#### 【0110】

##### [シンクライアントシステムへの適用例]

図 8 は、実施例 2 に係る画像表示システムの構成例を示す図である。図 8 には、画像表示システム 2 の一例として、クライアント端末 20 に最低限の機能しか持たせず、サーバ装置 200 でアプリケーションやファイルなどのリソースを管理するシンクライアントシステムが示されている。なお、ここでは、画像表示システム 2 の一形態としてシンクライ

50

アントシステムを例示するが、後述のように、汎用のクライアントサーバシステムにも上記の画像表示サービスを適用できることをここで付言しておく。

【0111】

図8に示すように、画像表示システム2には、クライアント端末20と、サーバ装置200とが含まれる。

【0112】

クライアント端末20には、デスクトップ型またはノート型のパーソナルコンピュータなどの情報処理装置を採用することができる。この他、クライアント端末20には、上記のパーソナルコンピュータなどの据置き型の端末のみならず、各種の携帯端末装置を採用することもできる。例えば、携帯端末装置の一例として、スマートフォン、携帯電話機やPHSなどの移動体通信端末、さらには、PDAなどのスレート端末などがその範疇に含まれる。

10

【0113】

サーバ装置200は、上記の画像表示サービスを提供するコンピュータである。

【0114】

一実施形態として、サーバ装置200は、パッケージソフトウェアやオンラインソフトウェアとして上記の画像表示サービスを実現する画像表示プログラムをインストールさせることによってサーバ装置を実装できる。例えば、サーバ装置は、上記の画像表示サービスを提供するWebサーバとして実装することとしてもよいし、アウトソーシングによって上記の画像表示サービスを提供するクラウドとして実装することとしてもかまわない。

20

【0115】

これらクライアント端末20及びサーバ装置200は、ネットワークNWを介して、互いが通信可能な状態で接続される。かかるネットワークNWの一例として、有線または無線を問わず、インターネットを始め、LANやVPN(Virtual Private Network)などの任意の種類通信網を採用できる。

【0116】

図8に示す通り、クライアント端末20は、カメラ1と、マイク3と、表示装置5と、入力装置7と、視線検出部22と、データ授受部24とを有する。なお、図8には、図1に示した機能部と同様の機能を発揮する機能部、例えばマイク、カメラ、表示装置及び入力装置に同一の符号を付し、その説明を省略する。

30

【0117】

図8に示す視線検出部22は、図1に示した視線検出部15cと同様に、カメラ1から画像が取得される度に、当該画像に対する視線検出を実行する。ここで、図8に示す視線検出部22は、一例として、視線検出部22がプロセッサ等とは別のハードウェアとして実装される場合を想定する。この場合、カメラ1及び視線検出部22は、一例として、表示装置5の周辺に設置される小型視線センサとして実装することができる。

【0118】

データ授受部24は、サーバ装置200との間で各種のデータの授受を制御する処理部である。

【0119】

一実施形態として、データ授受部24は、一例として、クライアント端末20が有するCPUなどのプロセッサにより、シンクライアントシステムのクライアント用のプログラムが実行されることで、仮想的に実現される。

40

【0120】

例えば、データ授受部24は、カメラ1が撮像する画像に対する視線検出部22の視線検出の結果、マイク3により入力される音声データ、さらには、入力装置7が受け付けた操作情報などをサーバ装置200へ送信する。また、データ授受部24は、サーバ装置200で実行されるプレゼンテーションソフトの実行結果を含むデスクトップ画面、すなわち表示装置5のスクリーンに表示させる表示データを受信する。例えば、プレゼンテーションソフトにより文書ファイルがスライドショーで表示される場合、プレゼンテーション

50

ソフトにより生成されるウィンドウは全画面表示されるので、デスクトップ画面とウィンドウ画面とが同じ表示内容となる。ここで、データ授受部 24 は、サーバ装置 200 が伝送するデスクトップ画面の表示データを任意のフレームレートで受信することができる他、デスクトップ画面の表示データに差分がある場合に絞ってデスクトップ画面の表示データを受信することもできる。このとき、サーバ装置 200 から伝送されるデスクトップ画面の表示データは、デスクトップ画面の全体であってもよいし、デスクトップ画面の一部、例えばフレーム間の差分の表示データであってもかまわない。

【0121】

このように、クライアント端末 20 及びサーバ装置 200 の間で授受される各種のデータには、トラフィックを抑制する観点から、圧縮符号化を行うこととしてもよいし、また、セキュリティの観点から、各種の暗号化を行うこととしてもよい。

10

【0122】

図 8 に示すように、サーバ装置 200 は、記憶部 220 と、制御部 240 とを有する。なお、サーバ装置 200 は、図 8 に示す機能部以外にも既知のコンピュータが有する各種の機能部、例えば他の装置との間で通信制御を行う通信 I/F 部などの機能部を有することとしてもかまわない。

【0123】

記憶部 220 は、制御部 240 で実行される OS やプレゼンテーションソフトを始め、アプリケーションプログラムなどの各種プログラムに用いられるデータを記憶するデバイスである。

20

【0124】

一実施形態として、記憶部 220 は、サーバ装置 200 における主記憶装置として実装される。例えば、記憶部 220 には、各種の半導体メモリ素子、例えば RAM やフラッシュメモリを採用できる。また、記憶部 220 は、補助記憶装置として実装することもできる。この場合、HDD、光ディスクや SSD などを採用できる。

【0125】

例えば、記憶部 220 は、制御部 240 で実行されるプログラムに用いられるデータの一例として、図 8 に示す文書データ 221、抽出キーワードデータ 222、認識キーワードデータ 223、視線データ 224 及び同期データ 225 を記憶する。これら文書データ 221、抽出キーワードデータ 222、認識キーワードデータ 223、視線データ 224 及び同期データ 225 は、サーバ装置 200 に接続されるクライアント端末 20 のうちいずれのクライアント端末 20 に関するデータであるのかがサーバ装置 200 で識別できるように、文書データ 221、抽出キーワードデータ 222、認識キーワードデータ 223、視線データ 224 及び同期データ 225 が格納される記憶領域がクライアント端末 20 の識別情報ごとに区別されたり、あるいは文書データ 221、抽出キーワードデータ 222、認識キーワードデータ 223、視線データ 224 及び同期データ 225 がクライアント端末 20 の識別情報とさらに対応付けられたりする他は、図 1 に示した文書データ 13a、抽出キーワードデータ 13b、認識キーワードデータ 13c、視線データ 13d 及び同期データ 13e と同様のデータである。これらのデータ以外にも、他の電子データ、例えばハイライト表示に関する定義データなども併せて記憶することもできる。

30

40

【0126】

制御部 240 は、各種のプログラムや制御データを格納する内部メモリを有し、これらによって種々の処理を実行するものである。

【0127】

一実施形態として、制御部 240 は、中央処理装置、いわゆる CPU として実装される。なお、制御部 240 は、必ずしも中央処理装置として実装されずともよく、MPU や DSP として実装されることとしてもよい。また、制御部 15 は、ASIC や FPGA などのハードワイヤードロジックによっても実現できる。

【0128】

制御部 240 は、各種のプログラムを実行することによって下記の処理部を仮想的に実

50

現する。例えば、制御部 240 は、図 8 に示すように、音声認識部 241 と、話速算出部 242 と、対応付け部 243 と、推定部 244 と、表示制御部 245 とを有する。なお、ここでは、音声認識を適用する前処理を実行する抽出部 15a に対応する処理部の図示は省略したが、抽出部 15a と同様の機能を発揮する処理部をサーバ装置 200 上に設けることにより、図 3 に示す処理をサーバ装置 200 に実行させることができるのは言うまでもない。

#### 【0129】

ここで、図 8 に示す制御部 240 では、図 1 に示した制御部 15 と異なり、視線検出がクライアント端末 20 で実行された上で視線検出の結果として検出時刻及び注視点がサーバ装置 200 へ伝送されるので、サーバ装置 200 では視線検出処理を実行せずともよい点が異なる。その代わりに、制御部 240 は、クライアント端末 20 から検出時刻及び注視点が伝送される度に、検出時刻及び注視点を含むエントリを視線データ 224 の 1 レコードとして記憶部 220 へ追加登録する。

10

#### 【0130】

図 8 に示す音声認識部 241 は、図 1 に示した音声認識部 15b と同様に、音声認識を実行する処理部である。この音声認識部 241 へ入力される音声データがクライアント端末 20 からサーバ装置 200 へ伝送された音声データであるという入力の取得経路以外は、図 1 に示した音声認識部 15b との間で相違はなく、音声認識部 241 は、図 1 に示した音声認識部 15b と同様の処理を実行する。

#### 【0131】

また、話速算出部 242、対応付け部 243 及び推定部 244 も、図 1 に示した話速算出部 15d、対応付け部 15e 及び推定部 15f と同様の処理を実行する処理部である。

20

#### 【0132】

表示制御部 245 は、クライアント端末 20 の表示装置 5 に対する表示制御を実行する処理部である。

#### 【0133】

ここで、表示制御部 245 は、クライアント端末 20 のデスクトップ画面、すなわち表示装置 5 のスクリーンに表示させる表示データを所定のフレームレート、あるいはデスクトップ画面の更新を契機に送信する。このとき、表示制御部 245 は、デスクトップ画面に更新がない場合、必ずしもデスクトップ画面の表示データをクライアント端末 20 へ伝送せずともかまわない。さらに、表示制御部 245 は、デスクトップ画面の全体の表示データを送信することとしてもよいし、デスクトップ画面の一部、例えばフレーム間の差分の表示データを送信することとしてもかまわない。このようなデスクトップ画面の伝送と並行して、表示制御部 245 は、図 1 に示した表示制御部 15g と同様に、クライアント端末 20 から伝送される入力装置 7 の操作情報にしたがって上記のスライドの表示制御を実行したり、さらには、上記のハイライトの表示制御などを実行することにより、プレゼンテーションソフトにより生成されるウィンドウ画面の表示データを更新する。このようにしてデスクトップ画面の伝送時にウィンドウ画面の更新内容がサーバ装置 200 からクライアント端末 20 へ伝送されることになる。

30

#### 【0134】

以上のように、本実施例に係る画像表示システム 2 がシンクライアントシステムとして実装された場合、サーバ装置 200 の音声認識部 241 が図 4 に示した音声認識処理を実行することができる。この音声認識処理では、ステップ S301 でマイク 3 から音声データが直接取得される代わりに、クライアント端末 20 からサーバ装置 200 へ伝送される音声データが取得される以外に処理内容の差はない。さらに、サーバ装置 200 の話速算出部 242 及び対応付け部 243 が図 5 に示した同期データの生成処理を実行することができる。さらに、サーバ装置 200 の推定部 244 及び表示制御部 245 が図 6 に示した表示制御処理を実行することができる。この表示制御処理では、ステップ S706 で説明箇所の領域が推定された段階で直ちにハイライト表示が実行されるとは限らない点異なる。すなわち、ステップ S707 では、ハイライト表示が行われたウィンドウ画面を含む

40

50

デスクトップ画面の表示データが生成される。このデスクトップ画面の表示データがクライアント端末 20 へ伝送されるタイミングは、必ずしもデスクトップ画面の表示データが生成されたタイミングと同時でなくともかまわない。

【0135】

[汎用のクライアントサーバシステムへの適用例]

図 8 には、画像表示システム 2 がシンクライアントシステムとして実装される場合を例示したが、必ずしもシンクライアントシステムとして実装されずともかまわず、汎用のクライアントサーバシステムとして実装することもできる。

【0136】

例えば、図 1 に示した画像表示装置 10 をクライアント端末とし、このクライアント端末を収容する図示しないサーバ装置に、画像表示装置 10 が有する処理部のうち、話速算出部 15 d、対応付け部 15 e 及び推定部 15 f などの処理部を実装することとすればよい。この場合、クライアント端末である画像表示装置 10 が図 4 に示した音声認識処理を実行し、認識キーワードが得られる度に追加の認識キーワードもしくは認識キーワードデータの全体を図示しないサーバ装置へ伝送することにより、図示しないサーバ装置上でクライアント端末ごとに認識キーワードデータが記憶されることになる。これによって、クライアント及びサーバ間で音声データが伝送されずともよくなる。

【0137】

さらに、図示しないサーバ装置で図 5 及び図 6 に示した各処理が実行される。この場合、図示しないサーバ装置は、ステップ S 706 で説明箇所に対応する領域が推定された場合、当該領域に付与されたインデックスをクライアント端末である画像表示装置 10 に伝送する。このように説明箇所に対応する領域のインデックスを受け取った画像表示装置 10 の表示制御部 15 g は、説明箇所に対応する領域のハイライト表示を実行することができる。

【0138】

以上のように、汎用のクライアントサーバシステムにも上記の画像表示サービスを適用できる。

【0139】

[会議システムへの適用例]

例えば、上記の実施例 1 では、話者と聴講者が 1 つの表示装置 5 を共用する場面を例示したが、必ずしも話者と聴講者が 1 つの表示装置を共用せずともかまわず、複数の表示装置の間で同一の表示内容が共有される場面にも上記の画像表示サービスを適用できる。例えば、会議等のコミュニケーションにおいて各参加者が話者及び聴講者の少なくとも一方または両方の立場で参加する状況が挙げられる。この場合、互いの表示装置に接続されるコンピュータがネットワークを介して接続されていれば互いが遠隔地に存在してもかまわない。

【0140】

図 9 は、画像表示サービスの会議システムへの適用例を示す図である。例えば、図 9 に示すように、図 1 に示した画像表示装置 10 と同様の機能を有するクライアント端末 10 A 及び 10 B がネットワーク NW を介して接続されると共にクライアント端末 10 A 及び 10 B 上でコミュニケーションツール、例えば画面共有用のアプリケーションプログラムが実行される場面に適用できる。これによって、クライアント端末 10 A 及び 10 B が有する各表示装置の間で同一の表示内容、例えばプレゼンテーションソフト用の文書ファイルが共有される。このような状況の下、クライアント端末 10 A 及び 10 B のうち少なくとも一方の端末が図 3 ~ 図 6 に示した処理を実行することにより、クライアント端末 10 A または 10 B の利用者の発話および視線を利用して、文書ファイルに含まれるスライドのうち説明箇所に対応する領域をハイライト表示することができる。

【0141】

図 10 は、画像表示サービスの会議システムへの適用例を示す図である。例えば、図 10 に示すように、図 8 に示したクライアント端末 20 と同様の機能を有するクライアント

10

20

30

40

50

端末 20A 及び 20B と、図 8 に示したサーバ装置 200 とがネットワーク NW を介して接続されると共に、サーバ装置 200 上でコミュニケーションツール、例えば画面共有用のアプリケーションプログラムが実行される場面に適用できる。これによって、クライアント端末 20A 及び 20B が有する各表示装置の間で同一の表示内容、例えばプレゼンテーションソフト用の文書ファイルが共有される。このような状況の下、サーバ装置 200 が図 3 ~ 図 6 に示した処理を実行することにより、クライアント端末 20A または 20B の利用者の発話および視線を利用して、文書ファイルに含まれるスライドのうち説明箇所に対応する領域をハイライト表示することができる。

【0142】

[ 画像表示プログラム ]

また、上記の実施例で説明した各種の処理は、予め用意されたプログラムをパーソナルコンピュータやワークステーションなどのコンピュータで実行することによって実現することができる。そこで、以下では、図 11 を用いて、上記の実施例と同様の機能を有する画像表示プログラムを実行するコンピュータの一例について説明する。

【0143】

図 11 は、実施例 1 及び実施例 2 に係る画像表示プログラムを実行するコンピュータのハードウェア構成例を示す図である。図 11 に示すように、コンピュータ 100 は、操作部 110a と、スピーカ 110b と、カメラ 110c と、ディスプレイ 120 と、通信部 130 とを有する。さらに、このコンピュータ 100 は、CPU 150 と、ROM 160 と、HDD 170 と、RAM 180 とを有する。これら 110 ~ 180 の各部はバス 140 を介して接続される。

【0144】

HDD 170 には、図 11 に示すように、上記の実施例 1 で示した抽出部 15a、音声認識部 15b、視線検出部 15c、話速算出部 15d、対応付け部 15e、推定部 15f 及び表示制御部 15g と同様の機能を発揮する画像表示プログラム 170a が記憶される。また、HDD 170 には、上記の実施例 2 で示した音声認識部 241、話速算出部 242、対応付け部 243、推定部 244 及び表示制御部 245 と同様の機能を発揮する画像表示プログラム 170a が記憶されることとしてもよい。この画像表示プログラム 170a は、図 11 に示した抽出部 15a、音声認識部 15b、視線検出部 15c、話速算出部 15d、対応付け部 15e、推定部 15f 及び表示制御部 15g の各構成要素と同様、統合又は分離してもかまわない。また、画像表示プログラム 170a は、図 8 に示した音声認識部 241、話速算出部 242、対応付け部 243、推定部 244 及び表示制御部 245 の各構成要素と同様、統合又は分散してもかまわない。すなわち、HDD 170 には、必ずしも上記の実施例 1 または上記の実施例 2 で示した全てのデータが格納されずともよく、処理に用いるデータが HDD 170 に格納されればよい。

【0145】

このような環境の下、CPU 150 は、HDD 170 から画像表示プログラム 170a を読み出した上で RAM 180 へ展開する。この結果、画像表示プログラム 170a は、図 11 に示すように、画像表示プロセス 180a として機能する。この画像表示プロセス 180a は、RAM 180 が有する記憶領域のうち画像表示プロセス 180a に割り当てられた領域に HDD 170 から読み出した各種データを展開し、この展開した各種データを用いて各種の処理を実行する。例えば、画像表示プロセス 180a が実行する処理の一例として、図 3 ~ 図 6 に示す処理などが含まれる。なお、CPU 150 では、必ずしも上記の実施例 1 で示した全ての処理部が動作せずともよく、実行対象とする処理に対応する処理部が仮想的に実現されればよい。

【0146】

なお、上記の画像表示プログラム 170a は、必ずしも最初から HDD 170 や ROM 160 に記憶されておらずともかまわない。例えば、コンピュータ 100 に挿入されるフレキシブルディスク、いわゆる FD、CD-ROM、DVD ディスク、光磁気ディスク、IC カードなどの「可搬用の物理媒体」に画像表示プログラム 170a を記憶させる。そ

10

20

30

40

50

して、コンピュータ 100 がこれらの可搬用の物理媒体から画像表示プログラム 170 a を取得して実行するようにしてもよい。また、公衆回線、インターネット、LAN、WANなどを介してコンピュータ 100 に接続される他のコンピュータまたはサーバ装置などに画像表示プログラム 170 a を記憶させておき、コンピュータ 100 がこれらから画像表示プログラム 170 a を取得して実行するようにしてもよい。

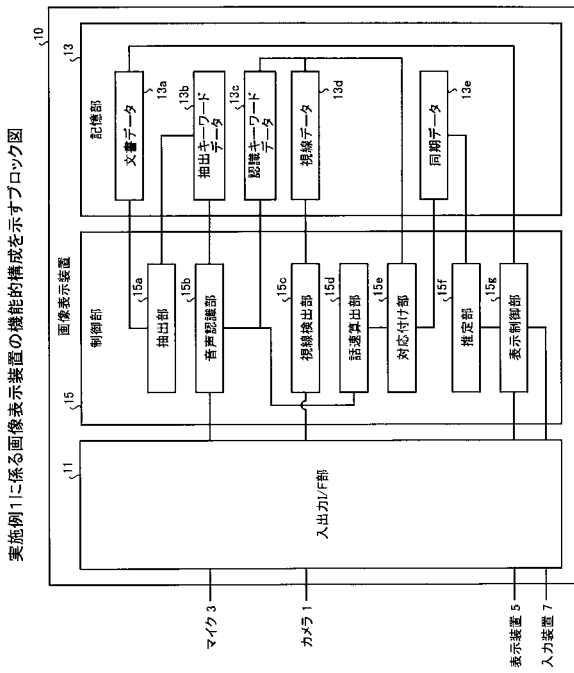
【符号の説明】

【0147】

1	カメラ	
3	マイク	
5	表示装置	10
7	入力装置	
10	画像表示装置	
11	入出力 I / F 部	
13	記憶部	
13 a	文書データ	
13 b	抽出キーワードデータ	
13 c	認識キーワードデータ	
13 d	視線データ	
13 e	同期データ	
15	制御部	20
15 a	抽出部	
15 b	音声認識部	
15 c	視線検出部	
15 d	話速算出部	
15 e	対応付け部	
15 f	推定部	
15 g	表示制御部	

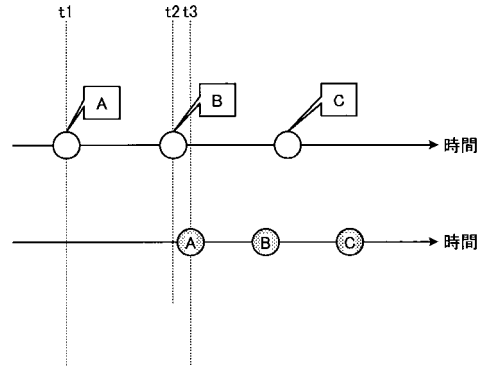


【 図 1 】



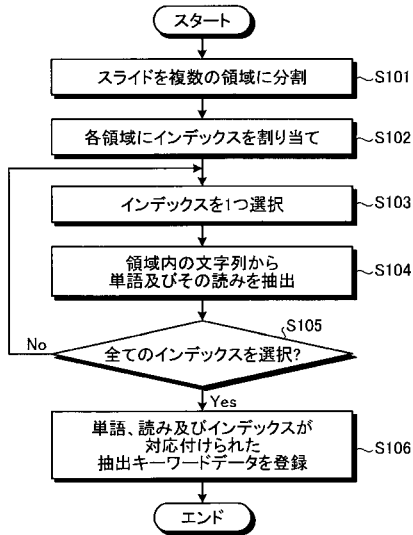
【 図 2 】

視線の検出時刻とキーワードの認識時刻の関係の一例を示す図



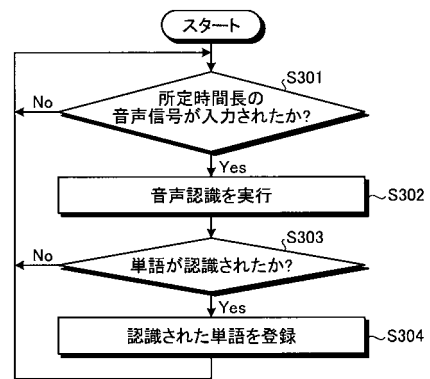
【 図 3 】

実施例1に係る抽出キーワードデータの生成処理の手順を示すフローチャート



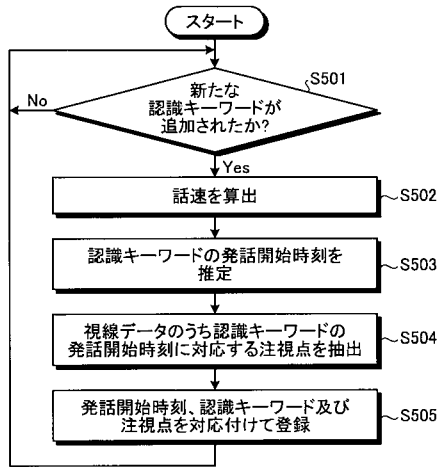
【 図 4 】

実施例1に係る音声認識処理の手順を示すフローチャート



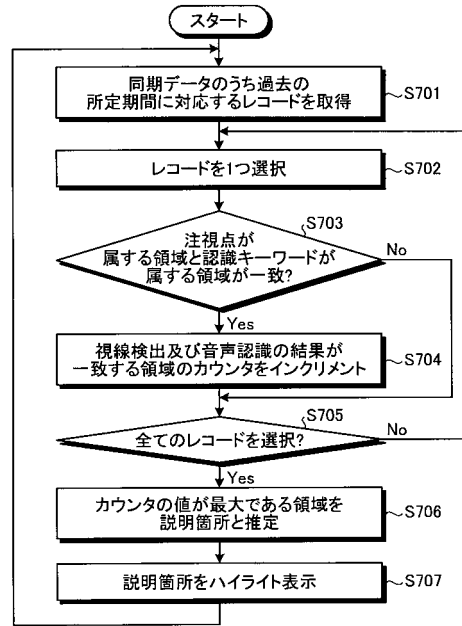
【 図 5 】

実施例1に係る同期データの生成処理の手順を示すフローチャート



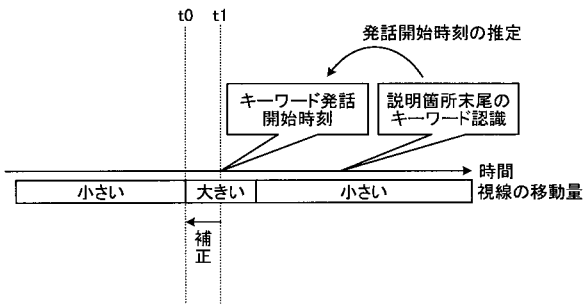
【 図 6 】

実施例1に係る表示制御処理の手順を示すフローチャート



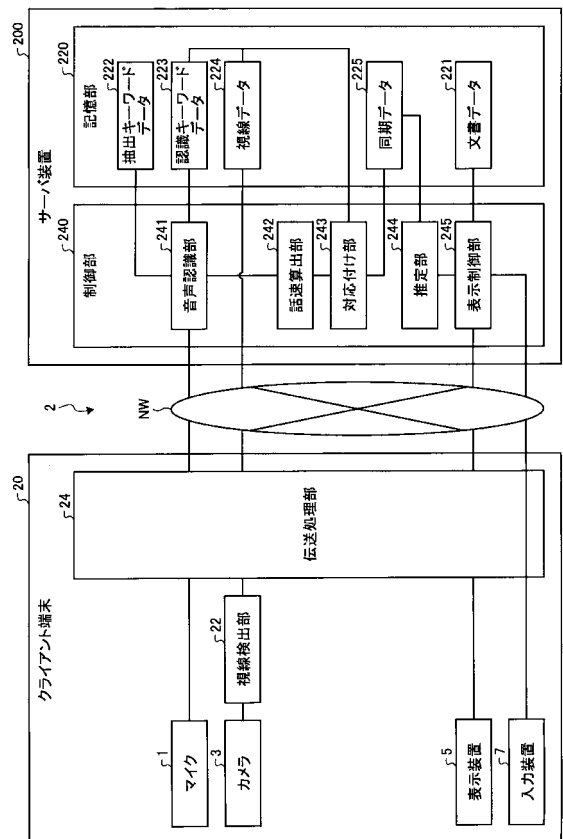
【 図 7 】

視線の移動量の一例を示す図



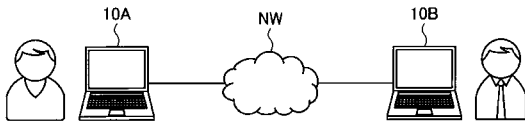
【 図 8 】

実施例2に係る画像表示システムの構成例を示す図



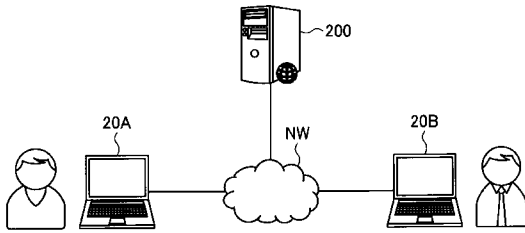
【 図 9 】

画像表示サービスの会議システムへの適用例を示す図



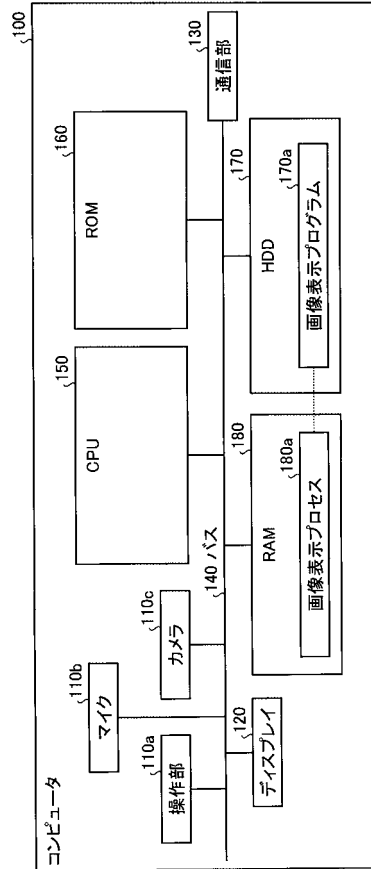
【 図 10 】

画像表示サービスの会議システムへの適用例を示す図



【 図 11 】

実施例1及び実施例2に係る画像表示プログラムを実行するコンピュータのハードウェア構成例を示す図



フロントページの続き

(51)Int.Cl.

F I

テーマコード(参考)

G 0 6 F 3/0481

Fターム(参考) 5C058 BA23 BA35 EA02

5E555 AA22 BA02 BA13 BB02 BC04 CA42 CA47 CB64 CB65 DB41

DC19 DC31 DC75 DD01 EA23 FA00