

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号  
特許第6276570号  
(P6276570)

(45) 発行日 平成30年2月7日 (2018.2.7)

(24) 登録日 平成30年1月19日 (2018.1.19)

(51) Int.Cl.	F I	
HO 4 N 21/472 (2011.01)	HO 4 N 21/472	
G 1 O L 19/00 (2013.01)	G 1 O L 19/00	3 1 2 F
HO 4 N 5/91 (2006.01)	HO 4 N 5/91	
HO 4 N 5/76 (2006.01)	HO 4 N 5/76	
HO 4 N 21/44 (2011.01)	HO 4 N 21/44	

請求項の数 13 (全 18 頁)

(21) 出願番号	特願2013-252416 (P2013-252416)	(73) 特許権者	000001007
(22) 出願日	平成25年12月5日 (2013.12.5)		キヤノン株式会社
(65) 公開番号	特開2015-109612 (P2015-109612A)		東京都大田区下丸子3丁目30番2号
(43) 公開日	平成27年6月11日 (2015.6.11)	(74) 代理人	100076428
審査請求日	平成28年12月5日 (2016.12.5)		弁理士 大塚 康德
		(74) 代理人	100112508
			弁理士 高柳 司郎
		(74) 代理人	100115071
			弁理士 大塚 康弘
		(74) 代理人	100116894
			弁理士 木村 秀二
		(74) 代理人	100130409
			弁理士 下山 治
		(74) 代理人	100134175
			弁理士 永川 行光

最終頁に続く

(54) 【発明の名称】 画像音声再生システム及び画像音声再生方法とプログラム

(57) 【特許請求の範囲】

【請求項 1】

画像オブジェクトが描画先に描画されていない状態から前記画像オブジェクトが前記描画先に描画された状態になるまでの前記描画先を撮像した時系列の静止画像データを取得する画像取得手段と、

ユーザーによる発話の音声データを含む、前記時系列の静止画像データに対応する時系列の音声データを取得する音声取得手段と、

前記画像オブジェクトを発話のタイミングと関連付ける関連付け手段と、

前記画像オブジェクトを含む静止画像データを表示する表示手段と、

前記表示された画像オブジェクトがユーザーの指示により選択された場合に、前記関連付けにより前記画像オブジェクトに関連付けられた前記発話のタイミングから、前記時系列の静止画像データおよび前記時系列の音声データを再生する再生手段と、  
を有することを特徴とする画像音声再生システム。

【請求項 2】

前記画像取得手段は、動画像から所定のタイミングで静止画像を抽出することで前記時系列の静止画像データを取得し、

前記音声取得手段は、前記動画像に同期して記録された音声から、無音区間で区切られた発話区間ごとに音声を抽出することで前記音声データを取得し、

前記関連付け手段は、先行する静止画像データとそれに続く静止画像データとの差分に含まれた画像オブジェクトを、前記先行する静止画像データが抽出されたタイミングから

前記それに基づく静止画像データが抽出されたタイミングまでの描画期間と発話区間が重複する前記音声データに関連付けることを特徴とする請求項 1 に記載の画像音声再生システム。

【請求項 3】

前記関連付け手段はさらに、前記描画期間を前後に拡張した拡張描画期間と発話区間が重複する前記音声データも前記画像オブジェクトに関連付けることを特徴とする請求項 2 に記載の画像音声再生システム。

【請求項 4】

前記関連付け手段は、前記差分に含まれた画像オブジェクトが誤記であるか判定し、誤記であった場合には、誤記であると判定された前記画像オブジェクトに関連付けられた音声データを、誤記であると判定された画像オブジェクトの消去の後に書き込まれた画像オブジェクトに関連付けることを特徴とする請求項 2 又は 3 に記載の画像音声再生システム。

10

【請求項 5】

前記関連付け手段は、前記差分に含まれた画像オブジェクトが下地色であり、前記先行する静止画像データの更に先行する静止画像データに、前記画像オブジェクトと重複する画像オブジェクトが含まれている場合に、当該画像オブジェクトを、誤記であると判定された前記画像オブジェクトとすることを特徴とする請求項 4 に記載の画像音声再生システム。

【請求項 6】

20

前記関連付け手段は、前記画像オブジェクトに関連付けた音声データが複数あった場合には、そのうちのひとつを最優先の音声データとして前記画像オブジェクトに関連付け、前記再生手段は、最優先の音声データを再生することを特徴とする請求項 1 乃至 5 のいずれか一項に記載の画像音声再生システム。

【請求項 7】

前記関連付け手段は、前記最優先の音声データとして、前記画像オブジェクトの描画期間と最も長く重複する発話区間の音声データを前記画像オブジェクトに関連付けることを特徴とする請求項 6 に記載の画像音声再生システム。

【請求項 8】

前記関連付け手段は、前記最優先の音声データとして、前記画像オブジェクトの描画期間に最も近い発話区間の音声データを前記画像オブジェクトに関連付けることを特徴とする請求項 6 に記載の画像音声再生システム。

30

【請求項 9】

前記画像取得手段は、動画像から所定のタイミングで静止画像を抽出することで前記時系列の静止画像データを取得し、

前記音声取得手段は、前記動画像に同期して記録された音声から、無音区間で区切られた発話区間ごとに音声を抽出することで前記音声データを取得し、

前記関連付け手段は、先行する静止画像データとそれに基づく静止画像データとの差分に含まれた画像オブジェクトを、前記先行する静止画像データが抽出されたタイミングから前記それに基づく静止画像データが抽出されたタイミングまでの描画期間の中の所定の特徴点時刻と発話区間の中の所定の特徴点時刻が最も近い前記音声データに関連付けることを特徴とする請求項 1 に記載の画像音声再生システム。

40

【請求項 10】

前記特徴点時刻は、前記描画期間及び前記発話区間の、開始時刻または終了時刻または中間時刻のいずれかであることを特徴とする請求項 9 に記載の画像音声再生システム。

【請求項 11】

動画像データに含まれたフレームの間の差分に基づいて、画像オブジェクトの現れるフレームと、該画像オブジェクトが現れる前のフレームとにより前記画像オブジェクトの書込み期間を特定し、前記動画像データに同期して記録された音声のうちから、発話区間の少なくとも一部が、前記書込み期間に対して一定時間内である音声を、前記画像オブジェ

50

クトに関連付ける関連付け手段と、

前記画像オブジェクトが選択された場合に、前記関連付けにより前記画像オブジェクトに関連付けられた前記音声の発話区間の開始時刻から、前記動画画像データを、同期して記録された音声とともに再生する再生手段と  
を有することを特徴とする画像音声再生システム。

【請求項 1 2】

画像オブジェクトが描画先に描画されていない状態から前記画像オブジェクトが前記描画先に描画された状態になるまでの前記描画先を撮像した時系列の静止画像データを取得する画像取得工程と、

ユーザーによる発話の音声データを含む、前記時系列の静止画像データに対応する時系列の音声データを取得する音声取得工程と、

前記画像オブジェクトを発話のタイミングと関連付ける関連付け工程と、

前記画像オブジェクトを含む静止画像データを表示する表示工程と、

前記表示された画像オブジェクトがユーザーの指示により選択された場合に、前記関連付けにより前記画像オブジェクトに関連付けられた前記発話のタイミングから、前記時系列の静止画像データおよび前記時系列の音声データを再生する再生工程と、  
を有する画像音声再生方法。

【請求項 1 3】

請求項 1 乃至 1 1 のいずれか一項に記載の画像音声再生システムをコンピュータにより実現するためのプログラム。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、画像音声再生システム及び画像音声再生方法とプログラムに関する。

【背景技術】

【0 0 0 2】

一般に、ホワイトボードなどの盤面を用いた打ち合わせや会議などでは、盤面に書かれた内容と共に、会議参加者が発話した内容も重要である。会議の後にこれらの内容を振り返る際に、盤面の内容と共に発話の内容を同時に再生し、会議等の振り返りを容易にする手法はいくつか知られている（特許文献 1 参照）。例えば特許文献 1 の技術では、盤面上に書き込まれた文字や図等の画像と、それらが書き込まれた際に発話された音声とを再生することにより、会議の内容を図形画像と共に動画議事録のように参照する事ができる。

【先行技術文献】

【特許文献】

【0 0 0 3】

【特許文献 1】特開 2 0 0 3 - 2 6 0 8 9 6 号公報

【発明の概要】

【発明が解決しようとする課題】

【0 0 0 4】

特許文献 1 の手法では、盤面上に文字や図形等が書込まれた時刻の音声を時刻情報と共に書込まれた文字や図形と関連付けて保存しておく。後から音声再生をする際には、盤面画像の中から文字や図形をユーザーが選択すると、文字や図形が書き込まれた時刻の音声再生される。しかしながら、本手法では盤面への書込みと発話が同時に行われるような場合には問題はないが、盤面への書込みと発話が同時で無い場合は盤面画像の中から文字や図形を選択しても、音声は正しく再生されとは限らない。例えば会議の参加者が文字や図形を盤面に書込みを行った後に発話を行う際には、その文字や図形を選択しても音声再生される箇所に到達するまでに時間が掛かってしまう。また逆に会議の参加者が文字や図形を盤面に書込む前に発話を行った場合は、その文字や図形を選択しても音声は終了した後の部分が再生されてしまう。

【0 0 0 5】

本発明は上記従来例に鑑みて成されたもので、盤面上へ書き込まれた文字や図形と音声との、それぞれの内容に応じて同期させ、記録を検討する際のユーザーの生産性や利便性を向上させることを目的とする。

【課題を解決するための手段】

【0006】

上記目的を達成するために本発明は以下の構成を有する。

【0007】

すなわち、本発明は、画像オブジェクトが描画先に描画されていない状態から前記画像オブジェクトが前記描画先に描画された状態になるまでの前記描画先を撮像した時系列の静止画像データを取得する画像取得手段と、

ユーザーによる発話の音声データを含む、前記時系列の静止画像データに対応する時系列の音声データを取得する音声取得手段と、

前記画像オブジェクトを発話のタイミングと関連付ける関連付け手段と、

前記画像オブジェクトを含む静止画像データを表示する表示手段と、

前記表示された画像オブジェクトがユーザーの指示により選択された場合に、前記関連付けにより前記画像オブジェクトに関連付けられた前記発話のタイミングから、前記時系列の静止画像データおよび前記時系列の音声データを再生する再生手段と、

を有することを特徴とする画像音声再生システムにある。

あるいは他の側面によれば、本発明は、動画像データに含まれたフレームの間の差分に基づいて、画像オブジェクトの現れるフレームと、該画像オブジェクトが現れる前のフレームとにより前記画像オブジェクトの書込み期間を特定し、前記動画像データに同期して記録された音声のうちから、発話区間の少なくとも一部が、前記書込み期間に対して一定時間内である音声を、前記画像オブジェクトに関連付ける関連付け手段と、

前記画像オブジェクトが選択された場合に、前記関連付けにより前記画像オブジェクトに関連付けられた前記音声の発話区間の開始時刻から、前記動画像データを、同期して記録された音声とともに再生する再生手段と

を有することを特徴とする画像音声再生システムにある。

【発明の効果】

【0008】

本発明によれば、盤面上への文字や図形の書込み時刻と発話時刻を参考にしながら、盤面上への書込み内容に該当する内容の発話時刻を検出することで、盤面上へ書き込まれた文字や図形と映像や音声の同期を行う事が可能になる。盤面上へ書き込まれた文字や図形と映像や音声を同期させて表示する事により、ユーザーにとって知りたい内容がすぐに再生されるような映像再生システムを提供する事が可能となり、ユーザーの利便性が向上する。

【図面の簡単な説明】

【0009】

【図1】映像撮影再生システムのハードウェア構成例を示した図である。

【図2】情報処理装置10のハードウェア内部構成を示した図である。

【図3】映像撮影再生システムを機能構成で示した図である。

【図4】動画像から静止画像を抽出した例の概略である。

【図5】差分領域情報を生成する処理のフローチャートである。

【図6】動画像から抽出された静止画像と差分画像を一覧にした図である。

【図7】差分領域情報と発話情報の例を示した図である。

【図8】領域差分の例を示した図である。

【図9】動画像の再生を行う処理のフローチャートである。

【図10】再生処理の初期画面表示例を示した図である。

【図11】再生処理中の表示例を示した図である。

【図12】差分領域情報と発話情報を紐付ける処理のフローチャートである。

【図13】本実施形態における領域への書込みと発話の開始、終了時刻の一例を示す図である。

10

20

30

40

50

【図 1 4】図 1 3 の例における、書込みと発話の紐付け結果を示す図である。

【図 1 5】実施形態 2 における差分領域情報と発話情報を紐付ける処理のフローチャートである

【図 1 6】図 1 3 の例の前半部における書込みと発話の特徴点の算出結果を示す図である。

【発明を実施するための形態】

【 0 0 1 0 】

[ 実施形態 1 ]

始めに本発明の実施の形態の適用が想定される場面について説明する。例えば会議参加者がそれぞれホワイトボードなどの盤面（以下、ホワイトボードあるいは描画先）に文字や図形などを書き込む会議を行うことがある。このような会議において、会議の内容を記録する方法としては、ホワイトボードの様子を撮影（あるいは撮像）して静止画として残す方法がある。しかし会議中の細かな会話の内容や議論の経緯などは文字や図形で残りにくく、後から記録を参照しても思い出すことが困難であることが多い。そこで本実施形態では、会議の内容を音声付きの動画像で全て記録しておき、後から参照するシステムを提供する。これにはホワイトボードの静止画像から任意の知りたい書込み（すなわち書き込まれた画像オブジェクト）の領域を選択し、動画像の任意の再生開始時刻から再生されるようにすることで動画参照や音声視聴の補助を行う。

【 0 0 1 1 】

< 映像撮影再生システムの構成 >

図 1 は本実施形態における映像撮影再生システム（あるいは画像音声再生システム）のハードウェア構成例を示したものである。図示するようにこの映像撮影再生システムは、情報処理装置 1 0 と、ホワイトボード 1 1 と、ビデオカメラ 1 2 と、マイクロフォン 1 3 によって構成されている。情報処理装置 1 0 は、ビデオカメラ 1 2 がホワイトボード 1 1 を撮影した動画像、およびマイクロフォン 1 3 から送られてくる音声を保存し、後述する処理を行い内蔵する表示装置で静止画像や動画像の表示を行う。この情報処理装置 1 0 は、ビデオカメラとマイクロフォンが接続できるパーソナルコンピュータやワークステーションだけでなく、それらを内蔵するタブレットに代表される一体型コンピュータシステムを用いても良い。

【 0 0 1 2 】

ホワイトボード 1 1 は、専用のマーカーペンで文字や図形などを書き込むのに用いる板で、専用のイレーサーを使うことで書込みを何度でも消去できる。壁面に貼るタイプの巻取り式のシート状の物等もあり、それを用いても良い。ビデオカメラ 1 2 は、ホワイトボード 1 1 に書き込まれた文字や図形を電子情報に変換し情報処理装置 1 0 へ送信するもので、これ自身で動画像を 2 次記憶等へ保存する機能は備えていない。マイクロフォン 1 3 は、会議参加者の会話を電子情報に変換し、情報処理装置 1 0 へ送信する。ここではビデオカメラ 1 2 とマイクロフォン 1 3 は別体型を用いているが、マイクロフォンの機能を内蔵した一体型のビデオカメラを用いるようにしても良い。

【 0 0 1 3 】

図 2 は、情報処理装置 1 0 のハードウェア内部構成を示したものである。情報処理装置 1 0 は、CPU 2 0 0、ROM 2 0 1、RAM 2 0 2、HDD 2 0 3、ディスプレイ 2 0 4、カメラ I / F 2 0 5、マイク I / F 2 0 6、キーボード 2 0 7、ポインティングデバイス 2 0 8 からなり、それらはシステムバス 2 0 9 によって接続されている。CPU（プロセッサ）2 0 0 は、情報処理装置 1 0 全体の制御を行うものであり、ROM 2 0 1 や HDD 2 0 3 などに記憶されたプログラムを RAM 2 0 2 に読み込み、各種処理を実行する。すなわち、情報処理装置 1 0 の CPU 2 0 0 は、コンピュータ読み取り可能な記憶媒体に記憶されたコンピュータプログラムを実行することにより、後述するフローチャートの各ステップを実行する処理部（処理手段）として機能する。ROM 2 0 1 は、CPU 2 0 0 を起動するためのプログラムを含む各種データを記憶している。RAM 2 0 2 は、ワークメモリとして用いられる。HDD 2 0 3 はカメラ I / F 2 0 5 やマイク I / F 2 0 6 か

10

20

30

40

50

ら入力されたデータやCPU200が処理したデータを保持している。ディスプレイ204はHDD203やRAM202に格納されているコンピュータプログラムやデータをCPU200の指示に従って処理された結果を表示するものである。カメラI/F205はビデオカメラ12から受信したデータをHDD203やRAM202等に格納する際のデータ変換を行う。マイクI/F206はマイクロフォン13から受信したデータをHDD203やRAM202等に格納する際のデータ変換を行う。キーボード207、ポインティングデバイス208はユーザーの指示動作をCPU200に伝達する。

#### 【0014】

##### <映像撮影再生システムの機能構成>

図3は、映像撮影再生システムを機能構成で示したものである。制御部30は、後述する関連情報生成処理300と映像再生処理301で構成されている。関連情報生成処理300は、以下の処理ユニットによって構成されている。動画受信部3000は、時系列で画像取得する動画撮影部32から動画像を受信して2次記憶部31へ動画像データ310として保存する。静止画抽出部3001は動画像データ310から所定のタイミングで静止画像を取り出し、静止画像データ311として図6のようにIDを1から順に付与して保存する。静止画像を取り出すタイミングについては、例えばユーザーが操作部34を用いて手動で指示したタイミングを用いて取り出したり、図示しない時刻計数部を用いて一定時刻間隔で取り出したりすることが考えられる。本実施形態では、ユーザーが操作部34を用いて手動で指示したタイミングを用いて動画像データ310から静止画像311を抽出すると仮定する。静止画抽出部3001は、撮影対象である描画先を撮像した時系列の画像データを取得する画像取得手段として機能する。なお、動画撮影部32は、動画像と同期した音声データを記録する。

#### 【0015】

静止画時刻抽出部3002は、前述の静止画像抽出手段3001が静止画像データを取り出した、動画像中の静止画時刻を抽出する。発話時刻抽出部3005は動画受信部3000で受信した動画の音声から発話が行われている時刻を抽出する。発話区間の抽出は、所定時間以上連続する無音区間の開始時間と終了時間とを抽出することにより行うことができる。即ち、無音区間の終了時間が発話区間の終了時間であり、発話区間は無音区間で区切られている。発話時刻抽出部3005は検出した発話区間の開始及び終了時刻を発話時刻情報313として2次記憶部31に保存する。図7(b)がこの発話時刻情報を具体的に示したものである。発話時刻抽出部3005は、時系列の画像データである抽出された静止画像に対応する時系列の音声データを、動画像とともに記録された音声から取得する音声取得手段として機能する。

#### 【0016】

差分抽出部3003は、前述の静止画像抽出部3001が取り出した静止画像を時系列順に比較し、差分を抽出する。この差分が、静止画像すなわちフレーム間に書き込まれた画像オブジェクトを含む。すなわち、前のフレームはいまだ書き込まれていない状態を、後のフレームは書きこまれた状態を示す。関連情報生成部3004は、前述の静止画時刻抽出部3002が抽出した静止画時刻情報と、差分抽出部3003が抽出した差分情報に関連付け、後述する、差分領域と時刻とを関連付けた差分領域情報312を生成する。この生成方法は後述する。図7(a)がこの差分領域情報312の具体例を示したものである。また関連情報生成部3004はこの差分領域情報312と発話時刻抽出部3005が抽出した発話時刻情報313とを紐付ける、すなわち関連付ける。この紐付け方法については後述する。

#### 【0017】

映像再生処理301は、以下の処理ユニットによって構成されている。動画再生部3010は差分領域情報312の情報を元に、動画像データ310を指定された再生開始時刻から読み出し、表示部33へ表示を行う。静止画再生部3011は静止画像データ311を読み出し、表示部33へ表示を行う。画像合成部3012は前述の選択情報生成部が生成した表示画面など複数の静止画像を同時に表示するために合成処理を行い、表示部33

10

20

30

40

50

へ表示を行う。選択情報生成部 3013 は前述の差分領域情報 312 の走査の結果、再生開始時刻が複数存在した場合に、該当する再生開始時刻の静止画像を複数読みだしてユーザーに選択を促すための表示画面を生成する。領域選択部 3014 は、ユーザーが操作部 34 を操作して得られた領域選択情報を読み取り、差分領域情報 312 を走査して動画の再生開始時刻を決定し、動画再生部 3010 へ動画再生の指示を行う。ユーザーが行う操作部 34 の操作は、座標情報の入力可能なポインティングデバイス 208 のクリック操作を用いる。また、表示部 33 に図示しない座標情報が得られるタッチ式の入力装置を付加し、画像を見ながらのタッチ入力によって情報を得るようにしても良い。

#### 【0018】

図 4 は動画像から静止画像を抽出した例の概略である。この例では、動画像データ 310 の中から動画の最初のフレームの静止画像 500、動画の最後のフレームの静止画像 504 の抽出を指示し、さらにその間に、ユーザーが 3:32、7:09、12:49 の時刻で静止画像の抽出指示を行った。その結果、最初のフレーム 500 と最後のフレーム 504 とに加えて静止画像 501、502、503 が抽出され、合計 5 つの静止画像が得られたことを示している。これら静止画像には ID として、例えば時系列にシリアル番号が付与される。ID は固有であればシリアル番号でなくともよい。なお以下の説明では、ユーザーにより動画から静止画像としてフレームが抽出されており、そのフレームが記録された時刻と関連付けて HDD 203 に保存されていることを前提とする。

#### 【0019】

##### < 差分領域情報の生成 >

図 5 は、前述の静止画抽出処理によって得られた静止画像を時系列順に比較し、差分領域情報を生成する処理をフローチャートにしたものである。S601 から S605 までの処理を静止画像の ID の値を 1 で初期化し、繰り返しごとに ID の値を加算しながら、比較する静止画像が存在する範囲で繰り返し実行する。S602 では、先行する ID 番目の静止画像と、それに続く (ID + 1) 番目の静止画像とを比較することで静止画像の差分抽出を行う。S603 では抽出結果から差分の有無を判断する。なお静止画像の ID がシリアル番号でない場合には、ID として静止画像の時系列の順序を示す番号を利用すればよい。

#### 【0020】

差分があった場合には、さらに S604 において、S602 で抽出した差分が下地色であるかの判断を行う。すなわち差分のあった領域内について、たとえば ID 番目の静止画像と、ID + 1 番目の静止画像との比較を画素ごとに行い、画素値の差分が所定の値以下の画素が、比較対象となった総画素数の所定のパーセント以上であるかの判定を行う。所定のパーセント以上であれば、差分のあった領域は下地色であると判断し、そうでない場合は下地色でないとする。差分が下地色でない場合は S605 において、領域情報として、差分のあった領域を特定する座標、例えば図 8 の (X1, Y1) 及び (X2, Y2) のように、矩形領域の対角点の座標の生成を行う。差分領域情報は、差分の領域情報と、静止画像の ID とを抽出した時刻とを関連付けたものである。一方、S604 で下地と判断された場合には、S607 において、差分領域は、ユーザーがホワイトボード上に誤って記入した誤記を消去したあとかどうかの判断を行う。この下地領域が誤記の消去かどうかの判断の詳細は後述する。この判断結果を S608、S609 で誤記フラグの生成を行い、S605 の差分領域情報生成に移行する。差分領域情報は画像の差分情報、誤記フラグと、差分があった時の静止画像の ID 番目を抽出した時刻とを関連付けたものである。その例を図 7 (a) に示す。一方、S603 で差分が無いと判断された場合には、差分領域情報を生成せずに S606 へ処理を移す。

#### 【0021】

S606 では静止画像の ID の値を 1 加算して、その値を抽出された静止画像の数と比較し、同じ値になったら繰り返し処理を終了する。ID の値が静止画像の数に達していなかった場合には S601 から再度処理を行う。なお、図 5 では 1 つの静止画像から 1 つの差分領域を抽出するかのよう記載したが、互いに連結していない複数の差分領域が 2 つの隣接する静止画像 (すなわち ID 番目と ID + 1 番目の静止画像) の間にある場合には、そ

10

20

30

40

50

れらすべての差分領域についてS 6 0 4 ~ S 6 0 5の工程が実行される。

【 0 0 2 2 】

図6は動画像データ310から抽出された静止画像71を入力した場合に領域差分抽出S 6 0 2によって得られる静止画像の差分72を一覧にしたものの一例である。図6におけるIDを引数として、静止画像をP ( ID ) と、差分領域をD ( ID ) と表す。S 6 0 2を実行する時にIDの値が1だった場合は、IDが1である静止画像P ( 1 ) と、IDが2の静止画像P ( 2 ) とを比較し、その結果として差分領域D ( 2 ) が得られ、差分領域情報が生成される。またIDの値が3だった場合は、静止画像P ( 3 ) とP ( 4 ) とを比較し、その結果として差分領域D ( 4 ) が得られる。このときの差分領域はP ( 3 ) に在った書込みが消された下地色であるため、その情報を差分領域情報に追加するとともに、図5のS 6 0 7の下地領域が消去された誤記かどうかの判断を行う。たとえば、下地色が発生する1つ前の差分領域情報を参考に、今回の差分領域(着目差分領域)の座標と1つ前の差分領域の座標とが所定の割合以上オーバーラップする場合には、着目差分領域は消去された誤記であると判断する。例えば着目差分領域と、下地色が発生する1つ前の差分領域とが50%以上オーバーラップするのであれば、1つ前の差分領域に描かれた誤記の消去であるとの判断が行われる。例えば、差分領域D ( 4 ) では、領域B-3に下地色が発生した。ここで、そのひとつ前の差分領域D ( 3 ) でどの座標に書込みが行われたかを参照し、差分領域D ( 4 ) とのオーバーラップの割合を判断する。この場合は差分領域D ( 3 ) の差分領域は、差分領域D ( 4 ) と同じ範囲を占める領域B-2であるため、オーバーラップの割合は100%であり、この下地色が誤記の消去を示すと判断し、差分領域情報の誤記フラグを'1'に設定する。領域Aと領域Bは位置的に異なる箇所であるが、差分領域D ( 3 ) の領域B-2と差分領域D ( 4 ) の領域B-3は、位置的に同一箇所である判断したものとする。

【 0 0 2 3 】

一方、差分領域D ( 7 ) において差分領域A, B-4はいずれも下地色と判断される。この場合においても下地色が発生する前の書込み領域を参考にする。下地色が発生する前の差分領域D ( 6 ) では領域Cに書込みが行われており、領域A、領域B-4と異なる領域のため、この下地色は誤記の消去でないと判断し、差分領域情報の誤記フラグに'0'を設定する。領域Aと領域B-1とは位置的に異なる箇所であるが、領域B-2と領域B-3、領域B-4は、位置的に同一箇所である判断したものとする。

【 0 0 2 4 】

図7(a)は、図5の処理に対して図6の静止画像71を入力した場合に得られる、差分領域情報を示したものの一例である。ID 8 0 は、静止画像の抽出時に当該静止画像に付されたIDである。領域情報81は画像差分の領域を示す情報であり、領域を一意に特定できればどのようなデータ構造でも良い。例えば図8にある、領域90のように画像差分に外接する矩形領域であれば、左上の座標であるX1,Y1、右下であるX2,Y2の情報を格納すれば良いし、領域91のように画像差分の輪郭領域であれば輪郭を示す座標の集合の情報を格納すれば良い。再生開始時刻情報82には、該当する領域に対して書込みが始まる時刻情報、すなわち図5の手順の処理中に着目しているIDを持つ静止画像の、元の動画像における時刻が再生開始時刻情報82として格納される。また静止画像時刻情報83には、該当する領域に対して書込みが終了した時刻情報、すなわち図5の手順の処理中に着目しているIDの次のID(すなわちID+1)を持つ静止画像の、元の動画像における時刻が格納される。非下地領域情報84は書込みが消されたかどうかを示し、消された領域の場合はNoが、書き込みがされた場合にはYesが記録される。もちろん表示は逆であってもよい。誤記フラグ情報85は前述したフローに従って下地領域情報が誤記であると判断された場合に'1'が記録される。

【 0 0 2 5 】

これら情報について図6を用いて具体的に説明する。IDの値が1であるときには、静止画像P ( 1 ) と静止画像P ( 2 ) の比較を行い、差分として領域721が検出される。この領域721の情報を領域Aとして、書込みが始まる時刻情報はIDが1であるときの撮影時刻であるため0:00が再生開始時刻として記録される。書込みが終了した時刻情

10

20

30

40

50



報はIDが2であるときの撮影時刻であるため3:22が静止画像時刻情報としてそれぞれ記録される。また同様にIDが2~6の領域情報81、再生開始時刻情報82、静止画像時刻情報83の値が該当する静止画像から生成される。IDが3、6の場合はこれらに加えて、下地領域情報と誤記フラグ情報が付与される。

#### 【0026】

<差分領域情報と発話時刻情報との関連付け>

図12は前述した差分領域情報312と発話期間情報313を関連付けるフローである。図12においては、ID=1の差分領域情報(図7(a)参照)から順次着目し、S1701~S1714までの処理を行う。なお発話情報が、処理対象の動画情報から別途抽出されているものとする。発話情報は、一連の音声が始まる開始時刻と当該音声が終わる終了時刻とから構成される発話時刻と、IDを含む。さらに音声情報そのものを含んでもよい。まずS1701において、着目した差分領域情報を取り出す(すなわち参照する)。S1702において着目差分領域のひとつ前のID(ID=1の場合は存在しない)の着目差分領域の誤記フラグ85を参照し、誤記フラグ85が無いもしくは'0'の場合はS1704に進み、着目差分領域が非下地領域かの判断を行う。非下地領域であればS1705に進み、書込みの開始時刻から終了時刻まで(以下、書込み期間あるいは描画期間と呼ぶ)と、開始時刻から終了時刻まで(以下、発話期間と呼ぶ)の重なる発話があるかを判断する(S1705)。なお書込みの開始時刻とは、着目差分領域情報の再生開始時刻82に相当し、書込み終了時刻とは着目差分領域情報の静止画像時刻情報83に相当する。発話期間が書込み時間に重なるような発話がある場合は、該当する発話すべてを着目差分領域に紐付ける(S1706)。この関連付けは図14に示すような書込み-音声表に記録される。書込み-音声表では、関連付けの対象となる差分領域のID80とその領域情報81、関連付けられる発話を特定する情報(例えばID)とその最優先の発話識別情報とが1レコードを構成する。発話情報のIDは、図7(b)の発話情報のID160に相当する。最優先の発話情報とは、一つの差分領域に複数所発話が関連付けられる場合に最優先される優先順位であり、書込み期間と重複している期間が最も長い発話が最優先の発話に指定される。そのためにひとつのIDに紐付いている発話がひとつかどうかの判断を行う(S1707)。複数ある場合には、紐付けられる複数の発話の書き込み期間との重複の長さを比較する。最も書込み期間との重複期間が長い発話のIDが最優先の発話欄に記録される(S1708)。この紐づけは、差分領域に表れた画像オブジェクトを、その書込み期間に発話された音声に関連付けることとすることができる。

#### 【0027】

一方、S1705において、書込み時刻と重なる発話がないと判定された場合、S1709に進み、書込み期間の前後の一定時間内に発話があるかどうかの判断を行う。このためには、例えば着目差分領域の再生開始時刻をその一定時間だけ前にずらし、静止画像時刻をその一定時間だけ後ろにずらして時間を前後に拡張した拡張書込み期間を一時的に求め、その拡張書込み期間と発話期間が重複する発話を探す。該当する発話があればS1710に進み、該当する会話を着目差分領域と紐づけて書込み-音声表に登録する。この紐づけは、差分領域に表れた画像オブジェクトを、その書込み期間に近い時刻に発話された音声に関連付けることとすることができる。次にS1711において、紐付けられる発話がひとつかの判断を行い、複数の場合はその中で時刻情報が近いものを第一優先とする。この場合の時刻情報の近いものの判断は、たとえば拡張書込み期間(あるいは拡張描画期間ともいう)と重複する期間が最も長い発話を第一優先とすることで実現できる。なお重複もまた一定時間内であることから、S1708とS1710との関連付けは、書込み期間から一定時間内の拡張書込み期間に少なくとも発話期間の一部が重複する音声を、当該書込み(すなわち画像オブジェクト)に関連付けることとすることもできる。

#### 【0028】

また、S1709において該当する発話がないと判定された場合には、S1714に進む。ここで紐付く発話がないと判断し、その旨を再生処理の初期画面の表示部110の該当領域中に表示させる。図12のS1712、S1713は単に説明のための工程であり

10

20

30

40

50

、何も行う必要はない。会議録の再生時に、会話が紐づけられていない書込み（すなわち差分領域）については、関連付けられる音声がない旨のメッセージが重複して表示される。図10（b）がその例であり図10（a）の領域1100に紐付けられた発話が無かった場合に、ユーザーにその旨を例えば1102のように表示する。

【0029】

さてS1702においてひとつ前の差分領域の誤記フラグが1である場合、さらにその前の差分領域に既に紐づけられている発話情報を紐付ける。なお誤記が連続する場合には、連続する誤記の更に前の差分領域に紐づけられている発話情報を紐付ける。その後S1705に分岐する。

【0030】

次に図12のフローを、図13に示すように書込みと発話が行われた場合を例として、各書込み領域に対して、どの発話が紐付けられるかを説明する。

【0031】

図13は時系列的にIDが1～12までの12個の差分領域A～G-3が図5の手順で抽出され、発話として発話1～12が存在しているものとする（以下書込みをIDで、発話をその番号で表現する）。なお「発話」とは記録された音声のことであり、たとえば一方的な説明や会話などの音声を含む。CPU200は2次記憶部31に保持されている差分領域情報312及び発話時刻情報313を取り出し、差分領域情報のID1から順に以下の処理を差分領域情報の最終IDまで繰り返す。

【0032】

図12のS1701において、まず、取り出した領域情報が非下地情報かどうかの判断を行う。S1702においてひとつ前のID（ID=1の場合は存在しない）の誤記フラグを確認し、誤記フラグが無いもしくは'0'の場合はS1704に進み、ID1の領域（領域A）が非下地領域かどうかの判断を行う。非下地領域であればS1705に進み、ID1の書込みの書込み期間と各発話の発話期間の重なる部分があるかを判断する（S1705）。重なる発話がある場合は、その発話を紐付ける（S1706）。この場合はID1の領域情報に発話1が紐付けられる。次にひとつのIDに紐付いている会話がひとつかどうかの判断を行う（S1707）。図13の場合はID1には会話1のみが紐付けられているので、次のIDの処理に向かう。ID2（領域B）の処理では、ID1の同様のフローでS1707まで処理が進み、ID2の場合は会話2と会話3が紐付いているのでS1708に進み、書込み時刻と紐付く発話時刻の重な長さを比較する。この場合、会話2の方が重なる時間が長いので、会話2を第一優先とする。

【0033】

次にID3（領域C）においては、ID1と同様のフローで処理が進み、S1705において、書込み期間と重なる発話がないため、S1709に進み、書込み期間の一定時間内に発話があるかどうかの判断を行う。ここでの一定時間内は予め定められた可変の時間であり、ここでは1分（1:00）とする。ID3の場合は会話4と会話5がその区間内で発話されている。従って、S1710に進みこの会話4と会話5をID3に紐付ける。次にS1711において、紐付く発話がひとつかの判断を行い、紐付く発話が複数の場合はその中で時刻情報が近いものを第一優先とする。この場合の時刻情報の近いものの判断は、それぞれの開始時刻と終了時刻の差のいずれをとっても良いものとする。

【0034】

ID4やID5においてはID3と同様のフローでS1709まで処理が進み、一定区間内で発話がないのでS1714に進む。ここで紐付く発話がないと判断し、その旨を再生処理の初期画面の表示部110の該当領域中に表示させる。図10（b）がその例であり図10（a）の領域1100に紐付けられた発話が無かった場合に、ユーザーにその旨を例えば1102のように表示する。

【0035】

ID8（領域F-3）の処理においては、S1702において、ひとつ前のID7の領域F-2の誤記フラグが'1'であるためS1703に進み、2つ前のID6の領域F-1

10

20

30

40

50

に紐付けられている会話 8 を紐付ける。また、ID 1 2 (領域 G - 3) においては、S 1 7 0 2 の判断でひとつ前の ID の誤記フラグが '0' のため S 1 7 0 4 に進み、非下地情報なので S 1 7 0 5 からの処理へと進む。一方、ID 7 や ID 1 1 などの下地情報自体にはどの会話とも紐付けを行わない。

#### 【0036】

以上の手順により、撮影したホワイトボードへの書き込みと音声とを関連付けた書込み - 音声表が作成される。

#### 【0037】

##### < 動画像再生処理 >

図 9 は、前述の差分領域情報と動画像とを利用して任意の再生開始時刻から動画像の再生を行う処理をフローチャートにしたものである。まず S 1 0 0 1 において、最終静止画像の表示を行う。この最終静止画像とは、保存した動画像の最終静止画像のことであり、静止画像例の図 6 では P ( 8 ) であるため、最終的な表示例は図 1 0 のようになる。次に S 1 0 0 2 で、ユーザーの操作部 3 4 に対する操作による座標情報の入力を待つ。入力される座標情報は、表示されている静止画像上の座標である。入力があった場合には S 1 0 0 3 で、入力された座標情報を基に差分領域情報 ( 図 7 参照 ) を走査する。より具体的には、入力された座標情報が差分領域情報内の領域情報 8 1 に保存された各領域内に含まれているか否かを走査する。すなわち、差分領域情報に登録された静止画像すべてについて、入力された座標を含む領域情報を検索する。次に S 1 0 0 4 で、その走査の結果該当する領域情報がないと判断された場合には、S 1 0 0 2 で再びユーザーの操作入力を待つ。

#### 【0038】

該当する領域情報が見いだされた場合には、次に S 1 0 0 5 において、領域情報に対する発話情報が複数存在するかを判定する。この判定は、該当する領域情報が含まれた静止画像の ID を索引として図 1 4 に示した書込み - 音声表のレコードを参照し、該当する領域に紐づけられた発話情報がいくつあるか判断することで行う。複数あった場合には S 1 0 1 0 に、なかった場合には S 1 0 0 6 に分岐する。ここで複数あった場合の処理については後述し、先に 1 つだけだった場合の処理の説明を行う。これは図 1 0 の表示例において領域 1 1 0 0 内の座標が指定された場合である。S 1 0 0 6 では、確認された領域情報に紐付けられた発話情報の発話開始時刻 1 6 2 から、動画の再生を開始する。次に S 1 0 0 7 でユーザーの操作部 3 4 に対する操作による入力を待つ。入力があった場合には S 1 0 0 8 で動画再生を中断した後に S 1 0 0 1 に処理を戻し、再び静止画像の表示を行う。入力がなかった場合には S 1 0 0 9 で動画の再生が最後まで到達したかの確認を行う。動画の再生が最後まで到達していなかった場合には S 1 0 0 7 で再びユーザーの操作入力を待ち、最後まで到達していた場合には S 1 0 0 1 に処理を戻し、再び静止画像の表示を行う。

#### 【0039】

次に S 1 0 0 5 でユーザー入力された座標情報に対する差分領域に対して複数の発話情報が紐付けされていた場合の処理について説明する。S 1 0 1 0 では、ユーザーが入力した座標情報に対応する差分領域に関連付けられた複数の発話情報の開始時刻と、第一優先の発話時刻を取得し、第一優先の発話時刻から再生を開始する。なお複数の静止画像に、入力座標を含む差分領域が含まれている場合には、たとえば時系列順に静止画像を選択する、そして選択した静止画像中の差分領域に紐づけられた発話情報のうち、最優先の発話情報の開始時刻から動画を再生する。更にその他の紐付けられた開始時刻を候補として再生画面に重畳して表示部 3 3 で表示する。この際の具体例が図 1 1 となる。動画の再生画面 1 1 0 1 の下に原座表示しているフレームの時刻を示すタイムバー 1 1 0 2 が表示される。表示部 3 3 では第一優先の発話時刻 1 2 0 0 から動画を再生すると共に、その他の紐付けられた発話の開始時刻も候補として 1 2 0 1 のように、発話された時刻を示すシンボルにより表示する。

#### 【0040】

また、図 1 4 のような領域情報が存在するような場合においては、領域 G - 1 と領域 G

- 3 は同一領域であるため、ユーザーがこの領域を指定した場合は領域 G - 1 もしくは領域 G - 3 のどちらかを再生するか選択できるようにユーザーに促してもよい。差分領域の同一性は、その範囲のみならず、そこに含まれた画像の画素単位での比較により判定される。その判定の結果同一であれば選択肢を表示する。領域 G - 1 が選択された場合には領域 G - 1 に紐付けられている会話 10 から再生が開始され、領域 G - 3 が選択された場合には領域 G - 3 に紐付けられている会話 12 が再生される。

#### 【 0 0 4 1 】

以上の処理により、表示された静止画像から注目したい箇所を指定することで、その領域に関連付けられた再生開始時刻から動画の再生が開始される。またその箇所に複数の再生開始時刻が関連付けられていた場合でも、再生開始候補の画像を表示し選択を促すことで複数の中から 1 つを選択可能になる。このことからホワイトボード各書込みに対して、書込みが開始されるところからの動画が参照でき、書き込みに対する動画中の音声を聞くことで会議中の会話や様子を容易に把握することが可能になる。

#### 【 0 0 4 2 】

##### [ 実施形態 2 ]

実施形態 2 では領域情報 8 1 と発話情報 1 6 1 の紐付けを行う際に、更に容易な方法で紐付けを行う方法について説明する。図 1 5 は本実施形態における領域情報と発話情報とを紐付ける処理をフローチャートにしたものであり、実施形態 1 の図 1 2 に代えて本実施形態で遂行される手順である。そのほかの構成は実施形態 1 に準ずる。

#### 【 0 0 4 3 】

図 1 5 において、実施形態 1 と同様の手順で CPU 2 0 0 は差分領域情報の取り出しを行う ( S 1 8 0 1 ) 。次に書込み時刻情報 ( すなわち再生開始時刻情報 8 2 ) から特徴点時刻の算出を行う ( S 1 8 0 2 ) 。この特徴点時刻とは開始時刻と終了時刻及び、それらから算出されるその中間点の中間時刻である。次に発話時刻情報からも同様に特徴点時刻の算出を行う ( S 1 8 0 3 ) 。これらの算出結果は図 1 3 の前半部を例にとると、図 1 6 のようになる。

#### 【 0 0 4 4 】

次に着目書込み領域 ( 着目差分領域であり、ID の値で特定される領域 ) の特徴点時刻と各発話の特徴点時刻の差分の算出を行う ( S 1 8 0 4 ) 。ここで、特徴点は開始時刻、終了時刻、中間時刻とあるが、どの特徴点間で差分の算出を行うかは予め定めておく。ここでは、それぞれの開始時刻を差分算出の特徴点とした場合を例にして説明する。領域 A の書込み開始時刻との会話 1 の発話開始時刻の時間差分は 0 : 1 0 、会話 2 との時間差分は 1 : 0 0 、会話 3 との時間差分は 2 : 4 0 となる。

#### 【 0 0 4 5 】

次に S 1 8 0 5 では、その算出した時間差分の中から最も小さいものの会話を選択する。すなわち着目書込み領域の特徴点時刻と最も近い特徴点時刻を持つ発話情報が選択される。領域 A が指定された場合には会話 1 が選択されその開始時刻から再生が開始される。次に S 1 8 0 6 では選択された発話時刻がひとつかの判断を行う。選択される発話時刻は通常ひとつであるが、前述した時間差分が同じ値であった場合は複数の発話を選択される。この場合は最も時刻情報が早い発話を選択することで、多くの発話をユーザーが再生できるようにする。同様に領域 B には会話 3 が、領域 C にも会話 3 が紐付けられる。

#### 【 0 0 4 6 】

このように本実施形態においては、領域情報と発話情報を紐付ける際に、少ない計算量で紐付けを行う事が可能となり、ユーザーが領域を指定した際に発話の開始時刻から再生を行え、ユーザーの利便性を向上させる事ができる。

#### 【 0 0 4 7 】

##### [ その他の実施形態 ]

なお本発明は、上記実施の形態に限定されず、本発明の要旨を逸脱しない範囲で種々の変形が可能である。例えば図 1 のハードウェア構成例に、ホワイトボードを撮影するビデオカメラだけでなく会議参加者の様子を撮影するビデオカメラを追加し動画像を 2 種類保

10

20

30

40

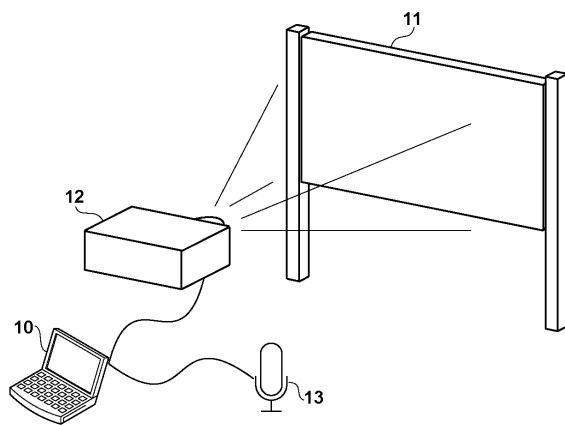
50

存し、動画像の再生時にピクチャーインピクチャーで両方の動画像を同時に再生してもよい。これにより会議の会話だけでなく参加者の様子も同時に確認できるため、より詳細な事後確認が可能になる。

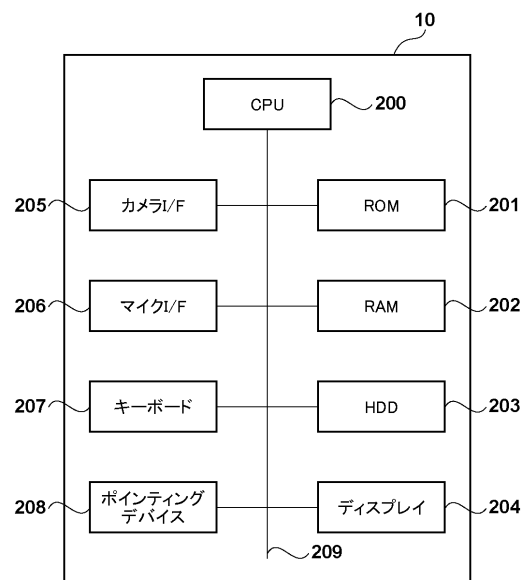
【 0 0 4 8 】

また、本発明は、以下の処理を実行することによっても実現される。即ち、上述した実施形態の機能を実現するソフトウェア（プログラム）を、ネットワーク又は各種記憶媒体を介してシステム或いは装置に供給し、そのシステム或いは装置のコンピュータ（またはCPUやMPU等）がプログラムを読み出して実行する処理である。

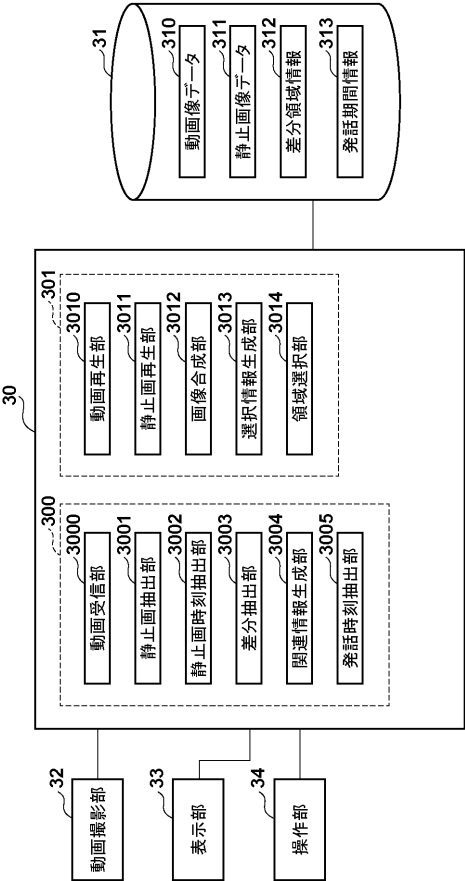
【 図 1 】



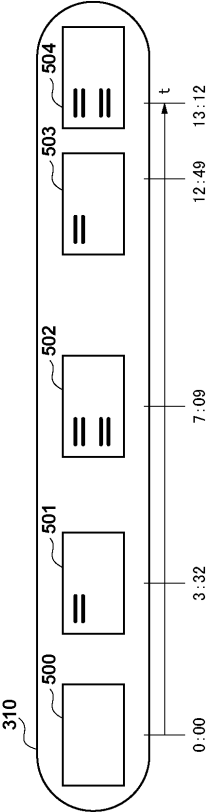
【 図 2 】



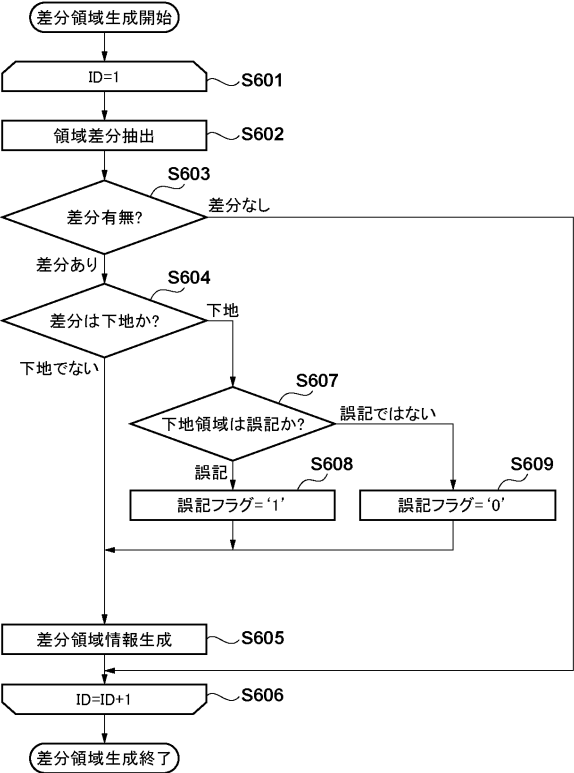
【図 3】



【図 4】



【図 5】



【図 6】

70		71	72
ID	撮影時刻(t)	静止画像(P)	画像差分領域(D)
1	0:00		
2	3:32		
3	7:09		
4	12:49		
5	13:12		
6	13:59		
7	14:49		
8	15:20		

【図 7】

80	81	82	83	84	85
ID	領域情報	再生開始時刻情報	静止画像時刻情報	非下地領域	誤記フラグ
1	領域A	0:00	3:32	Yes	
2	領域B-1	3:32	7:09	Yes	
3	領域B-2	12:49	13:12	No	1
4	領域B-3	13:12	13:59	Yes	
5	領域C	13:59	14:49	Yes	
6	領域A 領域B-4	14:49	15:20	No	0
7	領域D	15:20			

(a)

160	161	162	163
ID	発話情報	発話開始時刻情報	発話終了時刻情報
1	発話A	0:45	1:20
2	発話B	1:35	3:35
3	発話C	12:53	13:18

(b)

【図 8】

(X1, Y1)

90

1. ABCDEFG

2. abcdefg

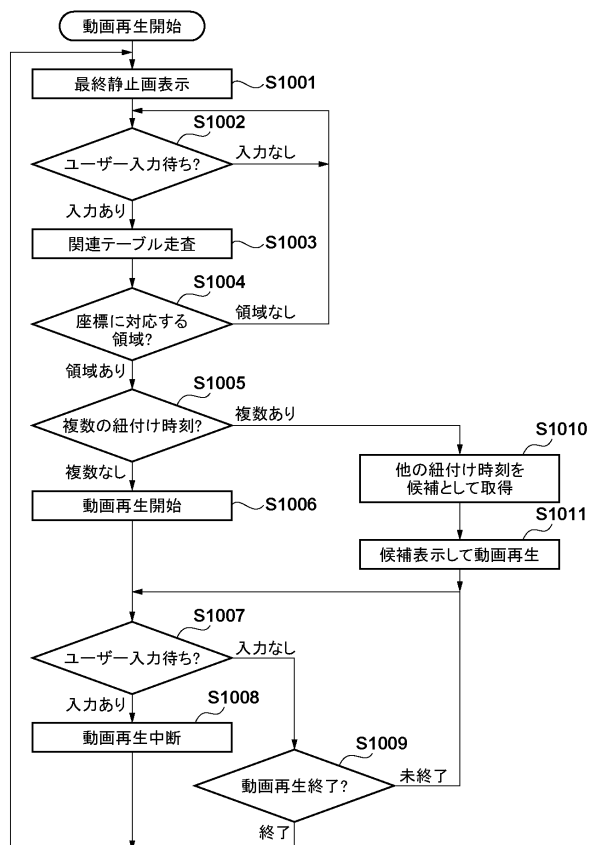
(X2, Y2)

91

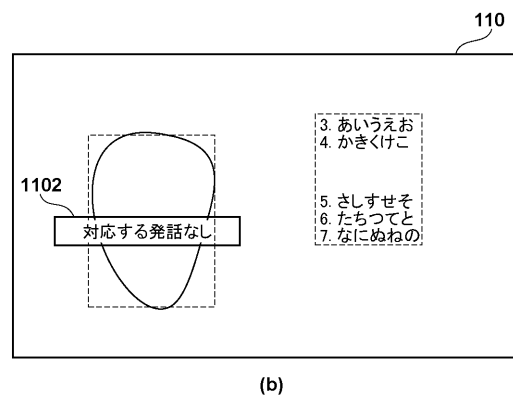
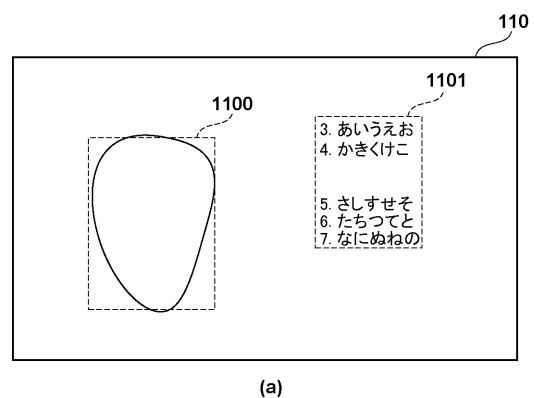
1. ABCDEFG

2. abcdefg

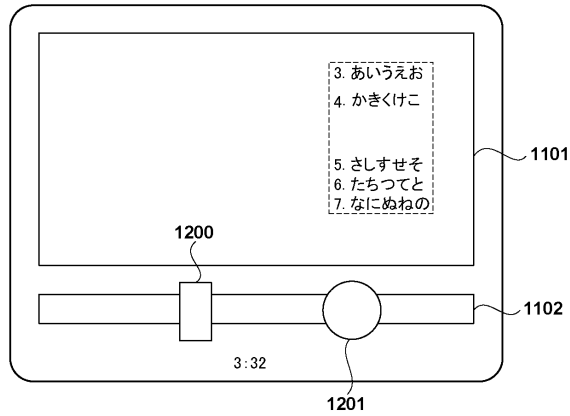
【図 9】



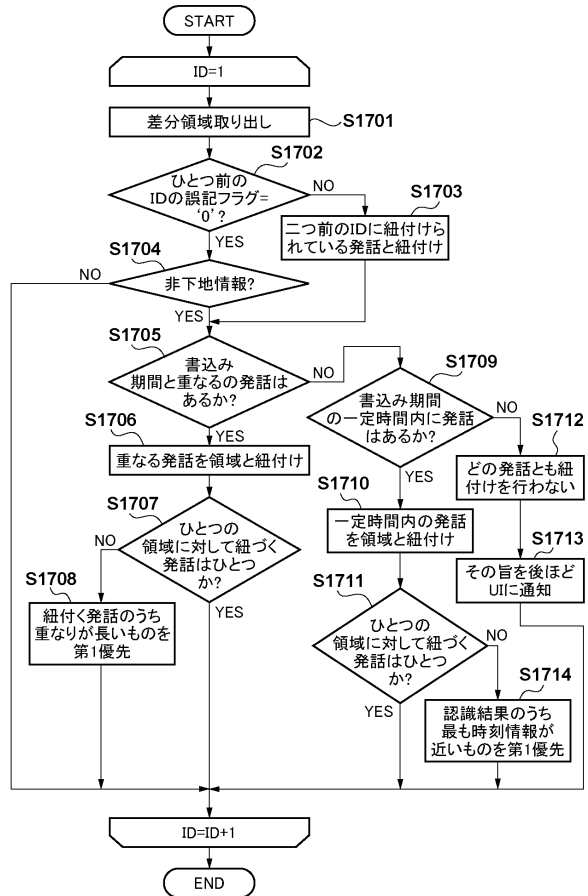
【図 10】



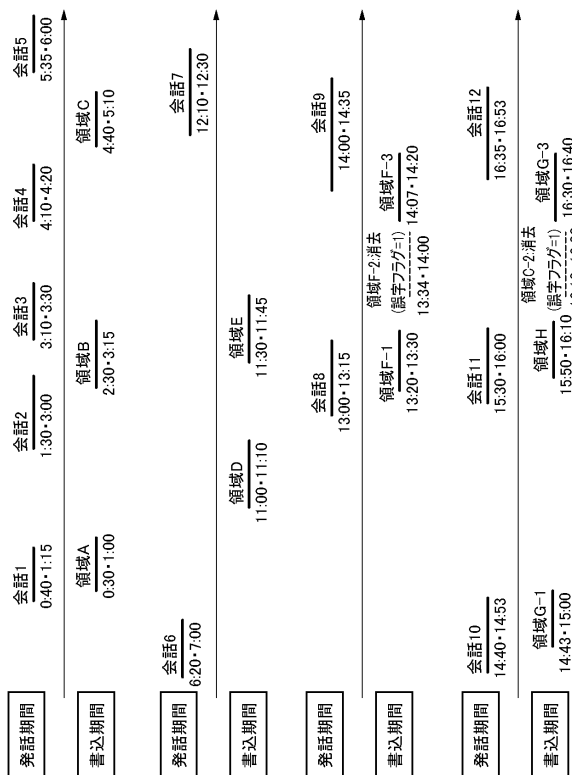
【図 1 1】



【図 1 2】



【図 1 3】

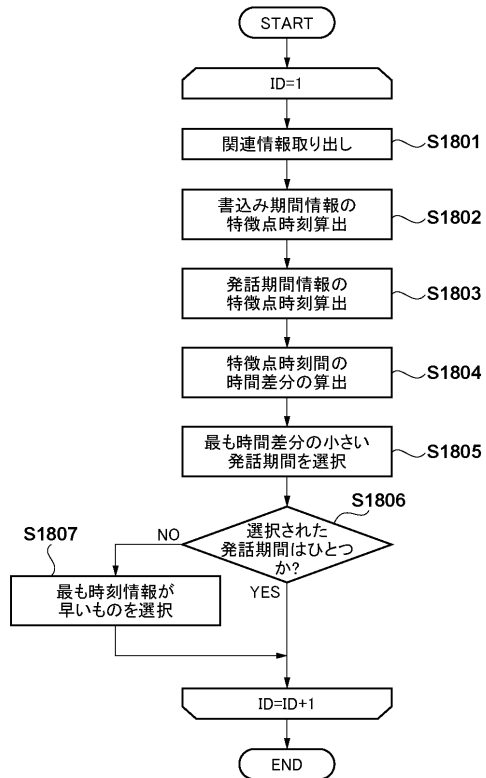


【図 1 4】

ID	領域情報	紐付けられる発話	最優先の発話
ID1	領域A	会話1	会話1
ID2	領域B	会話2、会話3	会話2
ID3	領域C	会話4、会話5	会話4
ID4	領域D	なし	なし
ID5	領域E	なし	なし
ID6	領域F-1	会話8	会話8
ID7	領域F-2	なし	なし
ID8	領域F-3	会話8、会話9	会話8
ID9	領域G-1	会話10	
ID10	領域H	会話11	
ID11	領域G-2	なし	
ID12	領域G-3	会話12	会話12



【図 15】



【図 16】

		書込みと発話の特徴点		
		開始時刻	終了時刻	中間時刻
書込み	領域A	0:30	1:00	0:45
	領域B	2:30	3:15	2:53
	領域C	4:40	5:10	4:55
発話	会話1	0:40	1:16	0:58
	会話2	1:30	3:00	2:15
	会話3	3:10	3:30	3:20

---

フロントページの続き

(72)発明者 松村 武士  
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

審査官 川中 龍太

(56)参考文献 国際公開第2010/025359(WO, A1)  
特開2011-135390(JP, A)  
特開2005-284490(JP, A)  
特開2005-167822(JP, A)  
特開2006-186786(JP, A)  
特開2003-260896(JP, A)  
特開2004-080750(JP, A)

(58)調査した分野(Int.Cl., DB名)  
H04N 21/00 - 21/858  
H04N 5/76 - 5/775  
H04N 5/91 - 5/956  
H04N 5/222 - 5/257  
G10L 19/00 - 99/00  
G06T 1/00  
G06F 17/30