(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第4474013号
(P4474013)

(45)発行日　平成22年6月2日(2010.6.2)　(24)登録日　平成22年3月12日(2010.3.12)

(51)Int.Cl.		F I		
G10L 15/24	(2006.01)	G10L	15/24	Q
G10L 15/28	(2006.01)	G10L	15/28	400
G10L 15/00	(2006.01)	G10L	15/00	200U
G06F 3/16	(2006.01)	G06F	3/16	320H
G06F 17/30	(2006.01)	G06F	17/30	170B

請求項の数 78　外国語出願　(全 54 頁)

(21)出願番号　特願2000-86807(P2000-86807)
(22)出願日　平成12年3月27日(2000.3.27)
(65)公開番号　特開2000-352996(P2000-352996A)
(43)公開日　平成12年12月19日(2000.12.19)
審査請求日　平成19年3月27日(2007.3.27)
(31)優先権主張番号　9907103.7
(32)優先日　平成11年3月26日(1999.3.26)
(33)優先権主張国　英国(GB)
(31)優先権主張番号　9908546.6
(32)優先日　平成11年4月14日(1999.4.14)
(33)優先権主張国　英国(GB)

(73)特許権者　000001007
キヤノン株式会社
東京都大田区下丸子３丁目３０番２号
(74)代理人　100076428
弁理士　大塚　康徳
(74)代理人　100112508
弁理士　高柳　司郎
(74)代理人　100116894
弁理士　木村　秀二
(74)代理人　100115071
弁理士　大塚　康弘

最終頁に続く

(54)【発明の名称】情報処理装置

(57)【特許請求の範囲】

【請求項１】

画像データ及び音声データを処理する装置であって、

少なくとも１台のカメラにより記録され且つ複数の人物の動きを示す画像データを処理して、各人物を三次元で追跡する画像処理手段と、

音声データを処理して、音声の到達方向を判定する音声処理手段と、

画像処理手段により実行される処理の結果と、音声処理手段により実行される処理の結果とに基づいて、どの人物が話しているかを判定する話者識別手段と、

受信した音声データを処理して、話者識別手段により実行される処理の結果に従って音声データからテキストデータを生成する音声認識処理手段とを備えることを特徴とする装置。

【請求項２】

音声認識処理手段は、人物ごとの対応する音声認識パラメータを格納する記憶手段と、話者識別手段により話していると判定された人物に従って音声データを処理するために使用すべき音声認識パラメータを選択する手段とを含むことを特徴とする請求項１記載の装置。

【請求項３】

画像処理手段は、処理される画像データを提供する各カメラの位置と向きを定義するカメラ校正データを使用して画像データを処理することにより各人物を追跡するように構成されていることを特徴とする請求項１又は２記載の装置。

【請求項４】
画像処理手段は、各人物の頭部を追跡することにより各人物を追跡するように構成されていることを特徴とする請求項１乃至３のいずれか１項に記載の装置。
【請求項５】
画像処理手段は、少なくとも話をしている各人物がどこを見ているかを判定するために画像データを処理するように構成されていることを特徴とする請求項１乃至４のいずれか１項に記載の装置。
【請求項６】
話者識別手段は、受信した画像データの所定のフレームについて画像処理手段及び音声処理手段により実行される処理の結果を使用して話者を識別できない場合、少なくとも１つの他のフレームに対して画像処理手段及び音声処理手段により実行される処理の結果を使用して所定のフレームにおいて話をしている人物を識別するように構成されていることを特徴とする請求項１乃至５のいずれか１項に記載の装置。
【請求項７】
受信した画像データ、音声データ、音声認識処理手段により生成されるテキストデータ、及び少なくとも話をしている各人物がどこを見ているかを定義する視線データの少なくとも一部を格納するデータベースを更に具備し、前記データベースは、対応するテキストデータと視線データとが互いに関連すると共に、対応する画像データ及び音声データとも関連するようにデータを格納する構成であることを特徴とする請求項１乃至６のいずれか１項に記載の装置。
【請求項８】
データベースに格納するために、画像データ及び音声データを圧縮する手段を更に具備することを特徴とする請求項７記載の装置。
【請求項９】
画像データ及び音声データを圧縮する手段は、画像データ及び音声データをMPEGデータとして符号化する手段であることを特徴とする請求項８記載の装置。
【請求項１０】
所定の期間にわたり、その所定の期間中に所定の人物がその他の人物の各々を見るのに費やした時間の割合を定義するデータを生成する手段を更に具備し、データベースは、そのデータが対応する画像データ、音声データ、テキストデータ及び視線データと関連するようにデータを格納する構成であることを特徴とする請求項７乃至９のいずれか１項に記載の装置。
【請求項１１】
所定の期間は所定の人物が話をしていた期間である請求項１０記載の装置。
【請求項１２】
画像データ及び音声データを処理する装置において、
少なくとも１台のカメラにより記録され且つ複数の人物の動きを示す画像データを処理して、各人物を三次元で追跡する画像処理手段と、
音声データを処理して、音声の到達方向を判定する音声処理手段と、
画像処理手段により実行される処理の結果と、音声処理手段により実行される処理の結果とに基づいて、どの人物が話しているかを判定する話者識別手段とを具備することを特徴とする装置。
【請求項１３】
画像処理手段は、処理される画像データを提供する各カメラの位置と向きを定義するカメラ校正データを使用して画像データを処理することにより各人物を追跡するように構成されていることを特徴とする請求項１２記載の装置。
【請求項１４】
画像処理手段は、各人物の頭部を追跡することにより各人物を追跡するように構成されていることを特徴とする請求項１２又は１３に記載の装置。
【請求項１５】

画像処理手段は、少なくとも話をしている各人物がどこを見ているかを判定するために画像データを処理するように構成されていることを特徴とする請求項１２乃至１４のいずれか１項に記載の装置。

【請求項１６】

話者識別手段は、受信した画像データの所定のフレームについて画像処理手段及び音声処理手段により実行される処理の結果を使用して話者を識別できない場合、少なくとも１つの他のフレームに対して画像処理手段及び音声処理手段により実行される処理の結果を使用して所定のフレームにおいて話をしている人物を識別するように構成されていることを特徴とする請求項１２乃至１５のいずれか１項に記載の装置。

【請求項１７】

画像データ及び音声データを処理する方法において、

少なくとも１台のカメラにより記録され且つ複数の人物の動きを示す画像データを処理して、各人物を三次元で追跡する画像処理ステップと、

音声データを処理して、音声の到達方向を判定する音声処理ステップと、

画像処理ステップで実行される処理の結果と、音声処理ステップで実行される処理の結果とに基づいて、どの人物が話しているかを判定する話者識別ステップと、

受信した音声データを処理して、話者識別ステップで実行される処理の結果に従って音声データからテキストデータを生成する音声認識処理ステップとを備えることを特徴とする方法。

【請求項１８】

音声認識処理ステップは、人物ごとに格納されている音声認識パラメータから、話者識別ステップで話していると判定された人物に従って音声データを処理するために使用すべき音声認識パラメータを選択することを含むことを特徴とする請求項１７記載の方法。

【請求項１９】

画像処理ステップでは、処理される画像データを提供する各カメラの位置と向きを定義するカメラ校正データを使用して画像データを処理することにより各人物を追跡することを特徴とする請求項１７又は１８記載の方法。

【請求項２０】

画像処理ステップでは、各人物の頭部を追跡することにより各人物を追跡することを特徴とする請求項１７乃至１９のいずれか１項に記載の方法。

【請求項２１】

画像処理ステップでは、少なくとも話をしている各人物がどこを見ているかを判定するために画像データを処理することを特徴とする請求項１７乃至２０のいずれか１項に記載の方法。

【請求項２２】

話者識別ステップでは、受信した画像データの所定のフレームについて画像処理ステップ及び音声処理ステップにより実行される処理の結果を使用して話者を識別できない場合、少なくとも１つの他のフレームに対して画像処理ステップ及び音声処理ステップにより実行される処理の結果を使用して所定のフレームにおいて話をしている人物を識別することを特徴とする請求項１７乃至２１のいずれか１項に記載の方法。

【請求項２３】

音声認識処理ステップで生成されるデータを搬送する信号を発生するステップを更に含むことを特徴とする請求項１７乃至２２のいずれか１項に記載の方法。

【請求項２４】

受信した画像データ、音声データ、音声認識処理ステップにより生成されるテキストデータ、及び少なくとも話をしている各人物がどこを見ているかを定義する視線データの少なくとも一部をデータベースに格納するステップを更に含み、データは、対応するテキストデータと視線データとが互いに関連すると共に、対応する画像データ及び音声データとも関連するようにデータベースに格納されることを特徴とする請求項１７乃至２３のいずれか１項に記載の方法。

【請求項２５】
画像データ及び音声データは圧縮された形態でデータベースに格納されることを特徴とする請求項２４記載の方法。
【請求項２６】
画像データ及び音声データはMPEGデータとして格納されることを特徴とする請求項２５記載の方法。
【請求項２７】
所定の期間にわたり、その所定の期間中に所定の人物がその他の人物の各々を見るのに費やした時間の割合を定義するデータを生成するステップと、そのデータが対応する画像データ、音声データ、テキストデータ及び視線データと関連するようにデータをデータベースに格納するステップとを更に含むことを特徴とする請求項２４乃至２６のいずれか１項に記載の方法。
【請求項２８】
所定の期間は所定の人物が話をしていた期間であることを特徴とする請求項２７記載の方法。
【請求項２９】
データベースを格納されているデータと共に搬送する信号を発生するステップを更に含むことを特徴とする請求項２４乃至２８のいずれか１項に記載の方法。
【請求項３０】
信号の記録を生成するために信号を直接に又は間接的に記録するステップを更に含むことを特徴とする請求項２９記載の方法。
【請求項３１】
画像データ及び音声データを処理する方法において、
少なくとも１台のカメラにより記録され且つ複数の人物の動きを示す画像データを処理して、各人物を三次元で追跡する画像処理ステップと、
音声データを処理して、音声の到達方向を判定する音声処理ステップと、
画像処理ステップにより実行される処理の結果と、音声処理ステップにより実行される処理の結果とに基づいて、どの人物が話しているかを判定する話者識別ステップとを備えることを特徴とする方法。
【請求項３２】
画像処理ステップでは、処理される画像データを提供する各カメラの位置と向きを定義するカメラ校正データを使用して画像データを処理することにより各人物を追跡することを特徴とする請求項３１記載の方法。
【請求項３３】
画像処理ステップでは、各人物の頭部を追跡することにより各人物を追跡することを特徴とする請求項３１又は３２記載の方法。
【請求項３４】
画像処理ステップでは、少なくとも話をしている各人物がどこを見ているかを判定するために画像データを処理することを特徴とする請求項３１乃至３３のいずれか１項に記載の方法。
【請求項３５】
話者識別ステップでは、受信した画像データの所定のフレームについて画像処理ステップ及び音声処理ステップにより実行される処理の結果を使用して話者を識別できない場合、少なくとも１つの他のフレームに対して画像処理ステップ及び音声処理ステップにより実行される処理の結果を使用して所定のフレームにおいて話をしている人物を識別することを特徴とする請求項３１乃至３４のいずれか１項に記載の方法。
【請求項３６】
話者識別ステップで識別された話者のアイデンティティを搬送する信号を発生するステップを更に含むことを特徴とする請求項３１乃至３５のいずれか１項に記載の方法。
【請求項３７】

プログラム可能処理装置を請求項１乃至１６の少なくとも１項に記載の装置として構成させるための命令を格納することを特徴とする記憶装置。

【請求項３８】

プログラム可能処理装置を請求項１７乃至３６の少なくとも１項に記載の方法を実行するように動作可能にさせるための命令を格納することを特徴とする記憶装置。

【請求項３９】

画像データ及び音声データを処理する装置であって、

少なくとも１台のカメラにより記録され且つ複数の人物の動きを示す画像データを処理して、各人物がどこを見ているかを判定すると共に、複数の人物がどこを見ているかに基づいてどの人物が話しているかを判定する画像処理手段と、

人物が話した言葉を定義する音声データを処理して、画像処理手段により実行される処理の結果に従って音声データからテキストデータを生成する音声処理手段とを備えることを特徴とする装置。

【請求項４０】

音声処理手段は、人物ごとの対応する音声認識パラメータを格納する記憶手段と、画像処理手段により話をしていると判定された人物に従って音声データを処理するために使用すべき音声認識パラメータを選択する手段とを含むことを特徴とする請求項３９記載の装置。

【請求項４１】

画像処理手段は、処理される画像データを提供する各カメラの位置と向きを定義するカメラ校正データを使用して画像データを処理することにより、各人物がどこを見ているかを判定するように構成されていることを特徴とする請求項３９又は４０記載の装置。

【請求項４２】

画像処理手段は、各人物の頭部の位置と向きを三次元で追跡するために画像データを処理することにより、各人物がどこを見ているかを判定するように構成されていることを特徴とする請求項３９乃至４１のいずれか１項に記載の装置。

【請求項４３】

画像処理手段は、各々の人物を見ている人物の数に基づいてどの人物が話をしているかを判定するように構成されていることを特徴とする請求項３９乃至４２のいずれか１項に記載の装置。

【請求項４４】

画像処理手段は、各人物が誰を見ているかを定義する値を人物ごとに生成し且つそれらの値を処理して、話をしている人物を判定するように構成されていることを特徴とする請求項４３記載の装置。

【請求項４５】

画像処理手段は、話をしている人物が他の人物の大半が見ている人物であることを判定するように構成されていることを特徴とする請求項３９乃至４４のいずれか１項に記載の装置。

【請求項４６】

画像データ、音声データ、音声処理手段により生成されるテキストデータ、及び各人物がどこを見ているかを定義する視線データを格納するデータベースを更に備え、前記データベースは、対応するテキストデータと視線データとが互いに関連すると共に、対応する画像データ及び音声データとも関連するようにデータを格納する構成であることを特徴とする請求項３９乃至４５のいずれか１項に記載の装置。

【請求項４７】

データベースに格納するために、画像データ及び音声データを圧縮する手段を更に備えることを特徴とする請求項４６記載の装置。

【請求項４８】

画像データ及び音声データを圧縮する手段は、画像データ及び音声データをMPEGデータとして符号化する手段であることを特徴とする請求項４７記載の装置。

【請求項４９】
所定の期間にわたり、その所定の期間中に所定の人物がその他の人物の各々を見るのに費やした時間の割合を定義するデータを生成する手段を更に備え、データベースは、そのデータが対応する画像データ、音声データ、テキストデータ及び視線データと関連するようにデータを格納する構成であることを特徴とする請求項４６乃至４８のいずれか１項に記載の装置。
【請求項５０】
所定の期間は所定の人物が話をしていた期間であることを特徴とする請求項４９記載の装置。
【請求項５１】
画像データを処理する装置において、少なくとも１台のカメラにより記録され且つ複数の人物の動きを示す画像データを処理して、各人物がどこを見ているかを判定すると共に、複数の人物がどこを見ているかに基づいてどの人物が話しているかを判定する画像処理手段を備えることを特徴とする装置。
【請求項５２】
画像処理手段は、処理される画像データを提供する各カメラの位置と向きを定義するカメラ校正データを使用して画像データを処理することにより、各人物がどこを見ているかを判定するように構成されていることを特徴とする請求項５１記載の装置。
【請求項５３】
画像処理手段は、各人物の頭部の位置と向きを三次元で追跡するために画像データを処理することにより、各人物がどこを見ているかを判定するように構成されていることを特徴とする請求項５１又は５２記載の装置。
【請求項５４】
画像処理手段は、各々の人物を見ている人物の数に基づいてどの人物が話をしているかを判定するように構成されていることを特徴とする請求項５１乃至５３のいずれか１項に記載の装置。
【請求項５５】
画像処理手段は、各人物が誰を見ているかを定義する値を人物ごとに生成し且つそれらの値を処理して、話をしている人物を判定するように構成されていることを特徴とする請求項５４記載の装置。
【請求項５６】
画像処理手段は、話をしている人物が他の人物の大半が見ている人物であることを判定するように構成されていることを特徴とする請求項５１乃至５５のいずれか１項に記載の装置。
【請求項５７】
画像データ及び音声データを処理する方法において、
少なくとも１台のカメラにより記録され且つ複数の人物の動きを示す画像データを処理して、各人物がどこを見ているかを判定すると共に、複数の人物がどこを見ているかに基づいてどの人物が話しているかを判定する画像処理ステップと、
人物が話した言葉を定義する音声データを処理して、画像処理手段により実行される処理の結果に従って音声データからテキストデータを生成する音声処理ステップとを備えることを特徴とする方法。
【請求項５８】
音声処理ステップは、人物ごとに格納されているそれぞれの音声認識パラメータから、画像処理ステップで話をしていると判定された人物に従って音声データを処理するために使用すべき音声認識パラメータを選択するステップを含むことを特徴とする請求項５７記載の方法。
【請求項５９】
画像処理ステップでは、処理される画像データを提供する各カメラの位置と向きを定義するカメラ校正データを使用して画像データを処理することにより、各人物がどこを見て

いるかを判定することを特徴とする請求項５７又は５８記載の方法。
【請求項６０】
画像処理ステップでは、各人物の頭部の位置と向きを三次元で追跡するために画像データを処理することにより、各人物がどこを見ているかを判定することを特徴とする請求項５７乃至５９のいずれか１項に記載の方法。
【請求項６１】
画像処理ステップでは、各々の人物を見ている人物の数に基づいてどの人物が話をしているかを判定することを特徴とする請求項５７乃至６０のいずれか１項に記載の方法。
【請求項６２】
画像処理ステップでは、各人物が誰を見ているかを定義する値を人物ごとに生成し且つそれらの値を処理して、話をしている人物を判定することを特徴とする請求項６１記載の方法。
【請求項６３】
画像処理ステップでは、話をしている人物が他の人物の大半が見ている人物であることを判定することを特徴とする請求項５７乃至６２のいずれか１項に記載の方法。
【請求項６４】
画像データ、音声データ、音声処理手段により生成されるテキストデータ、及び各人物がどこを見ているかを定義する視線データをデータベースに格納するステップを更に含み、前記データベースは、対応するテキストデータと視線データとが互いに関連すると共に、対応する画像データ及び音声データとも関連するようにデータを格納することを特徴とする請求項５７乃至６３のいずれか１項に記載の方法。
【請求項６５】
画像データ及び音声データは圧縮された形態で格納されることを特徴とする請求項６４記載の方法。
【請求項６６】
画像データ及び音声データはMPEGデータとして格納されることを特徴とする請求項６５記載の方法。
【請求項６７】
所定の期間にわたり、その所定の期間中に所定の人物がその他の人物の各々を見るのに費やした時間の割合を定義するデータを生成するステップと、そのデータが対応する画像データ、音声データ、テキストデータ及び視線データと関連するようにデータをデータベースに格納するステップとを更に含むことを特徴とする請求項６４乃至６６のいずれか１項に記載の方法。
【請求項６８】
所定の期間は所定の人物が話をしていた期間であることを特徴とする請求項６７記載の方法。
【請求項６９】
データベースを格納されているデータと共に搬送する信号を発生するステップを更に含むことを特徴とする請求項６４乃至６８のいずれか１項に記載の方法。
【請求項７０】
信号の記録を生成するために信号を直接に又は間接的に記録するステップを更に含むことを特徴とする請求項６９記載の方法。
【請求項７１】
画像データを処理する方法において、少なくとも１台のカメラにより記録され且つ複数の人物の動きを示す画像データを処理して、各人物がどこを見ているかを判定すると共に、複数の人物がどこを見ているかに基づいてどの人物が話しているかを判定するステップを備えることを特徴とする方法。
【請求項７２】
処理される画像データを提供する各カメラの位置と向きを定義するカメラ校正データを使用して画像データを処理することにより、各人物がどこを見ているかを判定することを

特徴とする請求項７１記載の方法。
【請求項７３】
各人物の頭部の位置と向きを三次元で追跡するために画像データを処理することにより、各人物がどこを見ているかを判定することを特徴とする請求項７１又は７２記載の方法。
【請求項７４】
各々の人物を見ている人物の数に基づいてどの人物が話をしているかを判定することを特徴とする請求項７１乃至７３のいずれか１項に記載の方法。
【請求項７５】
各人物が誰を見ているかを定義する値を人物ごとに生成し且つそれらの値を処理して、話をしている人物を判定することを特徴とする請求項７４記載の方法。
【請求項７６】
話をしている人物が他の人物の大半が見ている人物であることを判定することを特徴とする請求項７１乃至７５のいずれか１項に記載の方法。
【請求項７７】
プログラム可能処理装置を請求項３９乃至５６の少なくとも１項に記載の装置として構成させるための命令を格納することを特徴とする記憶装置。
【請求項７８】
プログラム可能処理装置を請求項５７乃至７６の少なくとも１項に記載の方法を実行するように動作可能にさせるための命令を格納することを特徴とする記憶装置。
【発明の詳細な説明】
【０００１】
【発明が属する技術分野】
本発明は、画像データのアーカイビングを補助するためのデータを生成する画像データの処理に関する。
【０００２】
更に、本発明は、画像データ及び音声データのアーカイビングを補助するためのデータを生成する画像データ及び音声データの処理に関する。
【０００３】
【従来の技術】
データを格納するためのデータベースは数多く存在している。しかし、既存のデータベースには、データベースから情報を検索するためにデータベースを問い合わせる方法が限られているという問題がある。
【０００４】
【発明が解決しようとする課題】
本発明は、上記の問題に留意してなされた。
【０００５】
【課題を解決するための手段】
本発明によれば、複数の人物の動きと話し言葉を記録した画像データ及び音声データを、画像処理と音声処理の組み合わせを利用して処理し、画像データ中に示されているどの人物が話をしているかを識別すると共に、音声データを処理し、識別された話者である参加者に従って選択された処理パラメータを使用して、話された言葉に対応するテキストデータを生成する装置又は方法が提供される。
【０００６】
データベースからの情報の検索を容易にするために、この後、テキストデータを画像データ及び／又は音声データと共にデータベースに格納しても良い。
【０００７】
また、本発明は、画像データを処理することにより複数の人物の三次元位置を判定し、それらの人物が話した言葉を伝達する音声データを処理して音源の方向を三次元で判定し、生成された位置情報を使用して、それらの言葉の話者を識別し、且つ識別された話者に関

して、音声／テキスト変換処理を実行するための音声認識パラメータを選択する装置又は方法を提供する。

【０００８】

このようにして、話者である参加者を容易に識別し、音声データを処理することができる。

【０００９】

各人物の位置は、各人物の少なくとも頭部を追跡するために画像データを処理することにより判定されるのが好ましい。

【００１０】

更に、本発明は、そのようなシステムにおいて画像データ及び音声データを処理し、話者である参加者を識別する装置又は方法を提供する。

【００１１】

更に、本発明は、信号の形態をとる命令及び記録された形態をとる命令を含めて、プログラム可能処理装置をそのようなシステムにおいて装置として構成させるため又は方法を実行可能にさせるように構成するための命令を提供する。

【００１２】

本発明によれば、画像データを処理し、どの人物が画像中でその他の人物の注目を集めているかを判定することにより、画像中のどの人物が話しているかを判定し、且つ音声データを処理し、画像データを処理することにより識別された話者である参加者に従って選択される処理パラメータを使用して、その人物により話された言葉に対応するテキストデータを生成する装置又は方法も提供される。

【００１３】

また、本発明は、画像データを処理して、画像中の人物が誰を見ているかを判定すると共に、それに基づいてどの人物が話をしているかを判定し、且つ音声データを処理して、話者である参加者について音声認識を実行する装置又は方法を提供する。

【００１４】

このようにして、話者である参加者を容易に識別し、音声データを処理することができる。

【００１５】

更に、本発明は、そのようなシステムにおいて画像データを処理する装置又は方法を提供する。

【００１６】

更に、本発明は、信号の形態をとる命令及び記録された形態をとる命令を含めて、プログラム可能処理装置をそのようなシステムにおいて装置として構成させるため又は方法を実行可能にさせるように構成するための命令を提供する。

【００１７】

【発明の実施の形態】

以下、添付の図面を参照して、単なる例示として、本発明の実施形態を説明する。

【００１８】

＜第１の実施形態＞

図１を参照して説明すると、複数のビデオカメラ（図１に示す例では３台であるが、これとは異なる台数であっても良い）２－１、２－２、２－３と、マイクロホンアレイ４とを使用して、複数の人物６、８、１０、１２の間で行われる会議から画像データと音声データをそれぞれ記録する。

【００１９】

マイクロホンアレイ４は、例えば、英国特許第２１４０５５８号、米国特許第４３３３１７０号及び米国特許第３３９２３９２号に記載されているような、入って来る音の方向を判定できるように配列されたマイクロホンのアレイから構成されている。

【００２０】

ビデオカメラ２－１、２－２、２－３からの画像データと、マイクロホンアレイ４からの

音声データは、ケーブル（図示せず）を介してコンピュータ２０に入力され、コンピュータ２０は受信したデータを処理し、データベースにデータを格納して、会議のアーカイブ記録を作成する。後に、このデータベースから情報を検索することができる。

【００２１】

コンピュータ２０は、従来のように、表示装置２６や、この実施形態ではキーボード２８及びマウス３０であるユーザ入力装置と共に、１つ又は複数のプロセッサ、メモリ、サウンドカードなどを含む処理装置２４を有する従来通りのパーソナルコンピュータである。

【００２２】

コンピュータ２０の構成要素と、それらの構成要素に対し入出力されるデータの流れを図２に概略的に示す。

【００２３】

図２を参照すると、処理装置２４は、例えば、ディスク３２などのデータ記憶媒体に格納されたデータとして及び／又は例えば、インターネットなどの通信ネットワーク（図示せず）を介する送信又は無線送信により遠隔データベースから処理装置２４に入力され且つ／又はキーボード２８などのユーザ入力装置又は他の入力装置を介してユーザにより処理装置２４に入力される信号３４として入力されるプログラミング命令に従って動作するようにプログラムされている。

【００２４】

プログラミング命令によりプログラムされると、処理装置２４は、処理動作を実行するための複数の機能ユニットに有効に構成される。そのような機能ユニットの例とそれらの配線を図２に示す。しかし、図２に示すユニットと配線は概念的なもので、単に理解を助けるために例示を目的として示されているにすぎない。従って、図２のユニットと配線とは、処理装置２４のプロセッサ、メモリなどが構成される実際のユニットと接続とを必ずしも表してはいない。

【００２５】

図２に示す機能ユニットについて説明すると、中央制御装置３６はユーザ入力装置２８、３０からの入力を処理すると共に、ユーザによりディスク３８などの記憶装置に格納されたデータとして、又は処理装置２４へ送信される信号４０として処理装置２４に入力されるデータを受信する。また、中央制御装置３６はその他の機能ユニットに対して制御と処理を実行する。メモリ４２は、中央制御装置３６及びその他の機能ユニットにより使用されるメモリである。

【００２６】

頭部追跡装置５０はビデオカメラ２－１、２－２、２－３から受信した画像データを処理して、会議のそれぞれの参加者６、８、１０、１２の頭部の位置と向きを三次元で追跡する。この実施形態では、この追跡を実行するために、頭部追跡装置５０は、後述するように、各々の参加者の頭部の三次元コンピュータモデルを定義するデータと、その特徴を定義するデータとを使用する。これらのデータは頭部モデル記憶装置５２に格納されている。

【００２７】

方向プロセッサ５３はマイクロホンアレイ４からの音声データを処理して、マイクロホンにより記録された音が来た方向を判定する。そのような処理は、例えば、英国特許第２１４０５５８号、米国特許第４３３３１７０号及び米国特許第３３９２３９２号に記載されているような従来の方式で実行される。

【００２８】

音声認識プロセッサ５４はマイクロホンアレイ４から受信された音声データを処理して、そこからテキストデータを生成する。すなわち、音声認識プロセッサ５４は、「Dragon Dictate」又はIBMの「ViaVoice」などの従来の音声認識プログラムに従って動作し、参加者６、８、１０、１２により話された言葉に対応するテキストデータを生成する。音声認識処理を実行するために、音声認識プロセッサ５４は、音声認識パラメータ記憶装置５６に格納されている、参加者６、８、１０、１２ごとの音声認識パラメータを定義するデー

タを使用する。すなわち、音声認識パラメータ記憶装置５６に格納されるデータは、音声認識プロセッサを従来の方式で訓練することにより生成される各参加者の音声プロファイルを定義するデータである。例えば、このデータは、訓練後にDragon Dictateの「ユーザファイル」に格納されるデータである。

【００２９】

アーカイブプロセッサ５８は、頭部追跡装置５０、方向プロセッサ５３及び音声認識プロセッサ５４から受信したデータを使用して、会議アーカイブデータベース６０に格納すべきデータを生成する。すなわち、後述するように、カメラ２－１、２－２及び２－３からの映像データと、マイクロホンアレイ４からの音声データとを、音声認識プロセッサ５４からのテキストデータ及び所定の時点で会議の各参加者が誰を見ていたかを定義するデータと共に会議アーカイブデータベース６０に格納するのである。

【００３０】

テキストサーチャ６２は、中央制御装置３６と関連して、会議アーカイブデータベース６０を探索し、後に更に詳細に説明するように、ユーザにより指定される探索基準に適合する会議の１つ又は複数の部分を見出し、その部分の音声データ及び映像データを再生するために使用される。

【００３１】

表示プロセッサ６４は、中央制御装置３６の制御の下に、ユーザに対し表示装置２６を介して情報を表示すると共に、会議アーカイブデータベース６０に格納された音声データと映像データを再生する。

【００３２】

出力プロセッサ６６はアーカイブデータベース６０のデータの一部又は全てを、例えば、ディスク６８などの記憶装置に又は信号７０として出力する。

【００３３】

会議を始める前に、処理装置２４が必要な処理動作を実行できるようにするために必要なデータを入力することにより、コンピュータ２０を初期設定する必要がある。

【００３４】

図３は、この初期設定中に処理装置２４により実行される処理動作を示す。

【００３５】

図３を参照して説明すると、ステップＳ１では、中央制御装置３６は表示プロセッサ６４に、ユーザが会議に参加するであろう各人物の名前を入力することを要求するメッセージを表示装置２６に表示させる。

【００３６】

ステップＳ２では、例えば、ユーザがキーボード２８を使用して入力した、名前を定義するデータを受信して、中央制御装置３６は各参加者に独自の識別番号を割り当て、識別番号と参加者の名前との関係を定義するデータ、例えば、図４に示すテーブル８０を会議アーカイブデータベース６０に格納する。

【００３７】

ステップＳ３では、中央制御装置３６は表示プロセッサ６４に、会議中のかなり長い時間にわたり人物が見ると考えられ、会議アーカイブデータベース６０にアーカイブデータを格納することが望まれる物体それぞれの名前をユーザが入力することを要求するメッセージを表示装置２６に表示させる。そのような物体としては、例えば、図１に示すフリップチャート１４などのフリップチャート、ホワイトボード又は黒板、又はテレビなどが挙げられる。

【００３８】

ステップＳ４では、例えば、ユーザがキーボード２８を使用して入力した、物体の名前を定義するデータを受信して、中央制御装置３６は各物体に独自の識別番号を割り当て、識別番号と物体の名前との関係を定義するデータ、例えば、図４に示すテーブル８０を会議アーカイブデータベース６０に格納する。

【００３９】

ステップＳ６では、中央制御装置３６は頭部モデル記憶装置５２を探索して、会議の参加者ごとに頭部モデルを定義するデータが既に格納されているか否かを判定する。

【００４０】

ステップＳ６で、１人または複数の参加者について頭部モデルがまだ格納されていないと判定されたならば、ステップＳ８で、中央制御装置３６は表示プロセッサ６４に、頭部モデルがまだ格納されていない各参加者の頭部モデルを定義するデータをユーザが入力することを要求するメッセージを表示装置２６に表示させる。

【００４１】

これに応答して、ユーザは、例えば、ディスク３８などの記憶媒体にあるデータを入力するか、又は接続している処理装置から信号４０としてデータをダウンロードすることにより、必要な頭部モデルを定義するデータを入力する。このような頭部モデルは、従来の方式により、例えば、Valente他の「An Analysis／Synthesis Cooperation for Head Tracking andVideo Face Cloning」（ProceedingsECCV’９８Workshop on Perceptionof Human Actionに掲載、ドイツ、フライブルク大学、１９９８年6月6日）に記載されている方法で生成されれば良い。

【００４２】

ステップＳ１０では、中央制御装置３６は、ユーザにより入力されたデータを頭部モデル記憶装置５２に格納する。

【００４３】

ステップＳ１２では、中央制御装置３６及び表示プロセッサ６４はユーザにより入力された各三次元コンピュータ頭部モデルをレンダリングして、ユーザが各モデルにおいて少なくとも７つの特徴を識別することを要求するメッセージと共に、ユーザに対し表示装置２６を介してモデルを表示する。

【００４４】

これに応答して、ユーザは、各々のモデルの中で、参加者の頭部の正面、側面及び（可能であれば）背面にある顕著な特徴、例えば、目尻、鼻孔、口、耳又は参加者が掛けている眼鏡の特徴などに対応する３０個の点をマウスを使用して指定する。

【００４５】

ステップＳ１４では、中央制御装置３６は、ユーザにより識別された特徴を定義するデータを頭部モデル記憶装置５２に格納する。

【００４６】

これに対し、ステップＳ６で、参加者ごとに頭部モデルが既に頭部モデル記憶装置５２に格納されていると判定された場合には、ステップＳ８からＳ１４を省略する。

【００４７】

ステップＳ１６では、中央制御装置３６は音声認識パラメータ記憶装置を探索して、参加者ごとに音声認識パラメータが既に格納されているか否かを判定する。

【００４８】

ステップＳ１６で、全ての参加者については音声認識パラメータを利用できないと判定されたならば、ステップＳ１８で、中央制御装置３６は表示プロセッサ６４に、パラメータがまだ格納されていない各参加者について音声認識パラメータを入力することをユーザに要求するメッセージを表示装置２６に表示させる。

【００４９】

これに応答して、ユーザは、例えば、ディスク３８などの記憶媒体のデータを入力するか、又は遠隔処理装置からの信号４０として入力することにより、必要な音声認識パラメータを定義するデータを入力する。先に述べた通り、これらのパラメータはユーザの話す音声のプロファイルを定義し、従来の方式で音声認識プロセッサを訓練することにより生成される。従って、例えば、Dragon Dictateを組み込んだ音声認識プロセッサの場合、ユーザにより入力される音声認識パラメータは、Dragon Dictateの「ユーザファイル」に格納されるパラメータに相当する。

【００５０】

ステップＳ２０では、中央制御装置３６は、ユーザにより入力された音声認識パラメータを定義するデータを音声認識パラメータ記憶装置５６に格納する。

【００５１】

これに対し、ステップＳ１６で、参加者ごとに音声認識パラメータを既に利用できる状態にあると判定された場合には、ステップＳ１８からS２０を省略する。

【００５２】

ステップＳ２２では、中央制御装置３６は表示プロセッサ６４に、カメラ２－１、２－２及び２－３の校正(キャリブレーション）を可能にするためのステップをユーザが実行することを要求するメッセージを表示装置２６に表示させる。

【００５３】

これに応答して、ユーザは必要なステップを実行し、ステップＳ２４では、中央制御装置３６はカメラ２－１、２－２及び２－３を校正するための処理を実行する。すなわち、この実施形態においては、ユーザにより実行されるステップ及び中央制御装置３６により実行される処理は、Wiles及びDavisonの「Calibrating and 3D Modelling with a Multi－Camera System」（１９９９IEEE Workshop on Multi－View Modelling and Analysis of Visual Scenes, ISBN ０７６９５０１１０９）に記載されているような方式で実行される。これは、会議室に対する各カメラ２－１，２－２及び２－３の位置及び向きを定義する校正データ（キャリブレーションデータ）と、各カメラ固有のパラメータ（横縦比、焦点距離、主点及び一次半径方向ひずみ係数）とを生成する。カメラ校正データ（カメラキャリブレーションデータ）は、例えば、メモリ４２に格納される。

【００５４】

ステップＳ２５では、中央制御装置３６は表示プロセッサ６４に、ステップＳ４で識別データが格納された物体それぞれの位置と向きを判定できるようにするためのステップをユーザが実行することを要求するメッセージを表示装置２６に表示させる。

【００５５】

これに応答して、ユーザは必要なステップを実行し、ステップＳ２６では、中央制御装置３６は、各物体の位置と向きを判定するための処理を実行する。すなわち、この実施形態においては、ユーザは、会議の参加者が見ると思われる物体の面の周囲、例えば、フリップチャート１４の紙の平面にカラーマーカーを置く。次に、中央制御装置３６は、カメラ２－１、２－２及び２－３の各々により記録された画像データをステップＳ２４で格納されたカメラ校正データを使用して処理し、従来の方式で、各々のカラーマーカーの三次元位置を判定する。この処理はカメラ２－１、２－２及び２－３ごとに実行されるので、各カラーマーカーの位置は別個に推定され、各カメラ２－１、２－２及び２－３からのデータを使用して計算された位置から、各マーカーの位置について平均位置が判定される。各マーカーの平均位置を使用して、中央制御装置３６は、従来の方式により、物体面の中心と、物体面の向きを定義するための面垂線とを計算する。物体ごとに判定された位置と向きは、例えば、メモリ４２に物体校正データとして格納される。

【００５６】

ステップＳ２７では、中央制御装置３６は表示プロセッサ６４に、会議の次の参加者（初めてステップＳ２７を実行する場合には、これは最初の参加者である）が着席することを要求するメッセージを表示装置２６に表示させる。

【００５７】

ステップＳ２８では、要求された参加者に着席する時間を与えるために、処理装置２４は所定の期間待機し、ステップＳ３０では、中央制御装置３６は各カメラ２－１、２－２及び２－３からのそれぞれの画像データを処理して、カメラごとに、着席した参加者の頭部の推定位置を判定する。すなわち、この実施形態においては、中央制御装置３６は従来の方式でカメラごとに別個に処理を実行し、参加者の肌の色に対応する色（この色は、頭部モデル記憶装置５２に格納されている参加者の頭部モデルを定義するデータから判定される）を有する、カメラからの画像データの１つのフレームにおける位置をそれぞれ識別し、次に、（頭部は人体の中で最も高い位置にある肌色の部分であると想定されるので）会

議室内の最も高い位置に相当する部分を選択する。画像中の識別された部分の位置と、ステップＳ２４で判定されたカメラ校正パラメータとを使用して、中央制御装置３６は従来の方式により頭部の三次元推定位置を判定する。この処理はカメラ２－１、２－２及び２－３ごとに実行され、カメラごとに別個の推定頭部位置が得られる。

【００５８】

ステップＳ３２では、中央制御装置３６は、カメラ２－１、２－２及び２－３ごとに、参加者の頭部の三次元推定向きを判定する。すなわち、この実施形態においては、中央制御装置３６は、頭部モデル記憶装置５２に格納されている参加者の頭部の三次元コンピュータモデルをそのモデルの複数の異なる向きについてレンダリングして、向きごとに対応するモデルの二次元画像を作成する。この実施形態では、参加者の頭部のコンピュータモデルを１０８の異なる向きでレンダリングするので、１０８の対応する二次元画像が得られる。これらの向きは、頭部モデルを０°（正面を向いている場合）、＋４５°（上を向いている場合）及び－４５°（下を向いている場合）のそれぞれについて１０°ずつ３６回回転させた向きに相当する。次に、中央制御装置３６は、モデルの各々の二次元画像を参加者の頭部を示す、カメラ２－１、２－２、２－３からの映像フレームの部分と比較し、モデルの画像が映像データと最も良く整合する向きを選択する。この比較と選択はカメラごとに実行されるので、カメラごとに推定頭部向きが得られる。頭部モデルをレンダリングすることにより生成される画像データをカメラからの映像データと比較するときには、例えば、Schodl、Haro及びEssaの「Head Tracking Using a Textured Polygonal Model」（Proceedings１９９８Workshop on Perceptual User Interfacesに掲載）に記載されているような従来の技法を使用する。

【００５９】

ステップＳ３４では、ステップＳ３０で生成された参加者の頭部のそれぞれの推定位置と、ステップＳ５２で生成された参加者の頭部のそれぞれの推定向きとを頭部追跡装置５０に入力し、各々のカメラ２－１、２－２及び２－３から受信した画像データのフレームを処理して、参加者の頭部を追跡する。すなわち、この実施形態においては、頭部追跡装置５０は、例えば、Valente他の「An Analysis／Synthesis Cooperation for Head Tracking andVideo Face Cloning」（ProceedingsEECV’９８Workshop on Perceptionof Human Action、ドイツ、フライブルク大学、１９９８年６月６日）に記載されているような従来の方式で頭部を追跡するための処理を実行する。

【００６０】

図５は、ステップＳ３４で頭部追跡装置５０により実行される処理動作の概要を示す。

【００６１】

図５を参照すると、ステップＳ４２－１からS４２－n（この実施形態では、カメラは３台であるので、「n」は３である）の各々においては、頭部追跡装置５０は会議を記録しているカメラのうち対応する１台からの画像データを処理して、そのカメラからの画像データにおける参加者の頭部の特徴（ステップＳ１４で格納された）の位置を判定すると共に、それに基づき、そのカメラからの画像データの現在フレームについて参加者の頭部の三次元位置と向きを判定する。

【００６２】

図６は、ステップＳ４２－１からS４２－nの所定の１つで実行される処理動作を示す。この処理動作は各ステップで同一であるが、異なるカメラからの画像データに対して実行されることになる。

【００６３】

図６を参照すると、ステップＳ５０では、頭部追跡装置５０は参加者の頭部の現在推定３D位置及び現在推定3D向きを読み取る。初めてステップＳ５０を実行する場合、これらは図３のステップＳ３０及びS３２で生成される推定位置と推定向きである。

【００６４】

ステップＳ５２では、頭部追跡装置５０はステップＳ２４で生成されたカメラ校正データを使用して、ステップＳ５０で読み取られた推定位置と推定向きに従って、頭部モデル記

憶装置５２に格納されている参加者の頭部の三次元コンピュータモデルをレンダリングする。

【００６５】

ステップＳ５４では、頭部追跡装置５０は、カメラから受信された映像データの現在フレームについて画像データを処理して、ユーザにより識別され、ステップＳ１４で格納された頭部の特徴の中の１つの特徴の期待位置を取り囲む各領域からの画像データを取り出す。この期待位置はステップＳ５０で読み取られた推定位置及び推定向きと、ステップＳ２４で生成されたカメラ校正データとから判定される。

【００６６】

ステップＳ５６では、頭部追跡装置５０はステップＳ５２で生成された、レンダリングされた画像データと、ステップＳ５４で取り出されたカメラ画像データとを整合し、レンダリングされた頭部モデルに最も良く整合するカメラ画像データを見出す。

【００６７】

ステップＳ５８では、頭部追跡装置５０は、ステップＳ５６で識別された、レンダリングされた頭部モデルに最も良く整合するカメラ画像データを、ステップＳ２４（図３）で格納されたカメラ校正データと共に使用して、映像データの現在フレームについて参加者の頭部の３D位置と３D向きを判定する。

【００６８】

再び図５に戻ると、ステップＳ４４では、頭部追跡装置５０は、ステップＳ４２－１からＳ４２－nの各々で識別された、レンダリングされた頭部モデルに最も良く整合するカメラ画像データ（図６のステップＳ５８で識別される）を使用して、映像データの現在フレームについて参加者の頭部の平均３D位置と平均３D向きを判定する。

【００６９】

ステップＳ４４を実行するのと同時に、ステップＳ４６では、ステップＳ４２－１からＳ４２－nの各々で判定されたカメラ画像データにおける頭部の特徴の位置（図６のステップＳ５８で識別される）を従来のカルマンフィルタに入力して、映像データの次のフレームについて参加者の頭部の推定３D位置及び推定３D向きを生成する。ビデオカメラ２－１、２－２及び２－３から映像データのフレームが受信されている間、その参加者についてステップＳ４２からＳ４６を繰り返し実行する。

【００７０】

再び図３に戻ると、ステップＳ３６では、中央制御装置３６は、会議に他の参加者がいるか否かを判定し、参加者ごとに先に説明したように処理が実行され終わるまでステップＳ２７からＳ３６を繰り返す。しかし、参加者ごとにこれらのステップが実行されている間、ステップＳ３４では、頭部追跡装置５０は既に着席した各参加者の頭部を追跡し続けている。

【００７１】

ステップＳ３６で、会議に他の参加者はなく、従って、各参加者の頭部は頭部追跡装置５０により追跡されていると判定されたならば、ステップＳ３８で、中央制御装置３６は、参加者の間で会議を始めて良いことを指示するために、処理装置２４から可聴信号を出力させる。

【００７２】

図７は、参加者間で会議が行われている間に処理装置２４により実行される処理動作を示す。

【００７３】

図７を参照すると、ステップＳ７０では、頭部追跡装置５０は会議中の各参加者の頭部を追跡し続ける。ステップＳ７０で頭部追跡装置５０により実行される処理は先にステップＳ３４に関して説明した処理と同じであるので、ここでは繰り返し説明しない。

【００７４】

頭部追跡装置５０がステップＳ７０で各参加者の頭部を追跡しているのと同時に、ステップＳ７２では、データを生成し、会議アーカイブデータベース６０に格納するための処理

を実行する。
【００７５】
図８は、ステップＳ７２で実行される処理動作を示す。
【００７６】
図８を参照すると、ステップＳ８０では、アーカイブプロセッサ５８は、参加者ごとに、その参加者がどの人物又はどの物体を見ているかを定義するいわゆる「視線パラメータ」を生成する。
【００７７】
図９は、ステップＳ８０で実行される処理動作を示す。
【００７８】
図９を参照すると、ステップＳ１１０では、アーカイブプロセッサ５８は頭部追跡装置５０から各参加者の頭部の現在三次元位置を読み取る。これは、ステップＳ４４（図５）で頭部追跡装置５０により実行される処理で生成された平均位置である。
【００７９】
ステップＳ１１２では、アーカイブプロセッサ５８は頭部追跡装置５０から次の参加者（初めてステップＳ１１２を実行する場合、これは最初の参加者である）の頭部の現在向きを読み取る。ステップＳ１１２で読み取られる向きは、ステップＳ４４（図５）で頭部追跡装置５０により実行される処理で生成された平均向きである。
【００８０】
ステップＳ１１４では、アーカイブプロセッサ５８は、参加者がどこを見ているかを定義する線（いわゆる「視線」）と、参加者の頭部を他の参加者の頭部の中心と結ぶ概念上の各々の線とが成す角度を判定する。
【００８１】
更に詳細に説明するため、図１０及び図１１を参照すると、１人の参加者、すなわち、図１の参加者６についてステップＳ１１４で実行される処理の一例が示されている。図１０を参照すると、ステップＳ１１２で読み取られる参加者の頭部の向きは、参加者の両目の中央の一点から、参加者の頭部に対し垂直に延びる視線９０を定義する。同様に、図１１を参照すると、ステップＳ１１０で読み取られる全ての参加者の頭部の位置は、参加者６の両目の中央の点からその他のそれぞれの参加者８、１０、１２の頭部の中心に至る概念上の線９２、９４、９６を定義する。ステップＳ１１４で実行される処理においては、アーカイブプロセッサ５８は視線９０と、それぞれの概念上の線９２、９４、９６とが成す角度９８、１００、１０２を判定する。
【００８２】
再び図９に戻ると、ステップＳ１１６では、アーカイブプロセッサ５８は、最小値を有する角度９８、１００又は１０２を選択する。すなわち、図１１に示す例でいえば、角度１００が選択されることになるであろう。
【００８３】
ステップＳ１１８では、アーカイブプロセッサ５８は、ステップＳ１１６で選択した角度が１０°より小さい値を有するか否かを判定する。
【００８４】
ステップＳ１１８で、角度が１０°より小さいと判定されれば、ステップＳ１２０で、アーカイブプロセッサ５８は参加者の視線パラメータを、視線と最小の角度を成す概念上の線により結ばれている参加者の識別番号（図３のステップＳ２で割り当てられている）に設定する。すなわち、図１１に示す例でいえば、角度１００が１０°より小さければ、角度１００は視線９０と、参加者６を参加者１０と結ぶ概念上の線９４とが成す角度であるので、視線パラメータは参加者１０の識別番号に設定されるであろう。
【００８５】
これに対し、ステップＳ１１８で、最小の角度が１０°以上であると判定された場合には、ステップＳ１２２で、アーカイブプロセッサ５８はステップＳ２６（図３）で先に格納された各物体の位置を読み取る。

【００８６】
ステップＳ１２４では、アーカイブプロセッサ５８は、参加者の視線９０がいずれかの物体の平面と交わるか否かを判定する。
【００８７】
ステップＳ１２４で、視線９０が１つの物体の平面と交わると判定されたならば、ステップＳ１２６で、アーカイブプロセッサ５８は参加者の視線パラメータを視線と交わる物体の識別番号（図３のステップＳ４で割り当てられている）に設定する。視線９０と交わる物体が２つ以上ある場合には、これは、視線と交わる物体のうち、参加者に最も近い物体ということになる。
【００８８】
これに対し、ステップＳ１２４で、視線９０が物体の平面と交わらないと判定されたならば、ステップＳ１２８で、アーカイブプロセッサ５８は参加者の視線パラメータを「０」に設定する。これは、（視線９０が概念上の線９２、９４、９６のいずれにも十分近接していないために）参加者はその他の参加者の誰をも見ておらず、また、（視線９０が物体と交わらないために）どの物体をも見ていないと判定されたことを示している。このような状況は、例えば、参加者が会議室内の、ステップＳ４でデータが格納されず且つステップＳ２６で校正が実行されなかった何らかの物体（例えば、図１に示す例において参加者１２が手に持っているメモ）を見ている場合などに起こりうるであろう。
【００８９】
ステップＳ１３０では、アーカイブプロセッサ５８は会議に他の参加者がいるか否かを判定し、参加者ごとに先に説明した処理が実行され終わるまでステップＳ１１２からＳ１３０を繰り返す。
【００９０】
再び図８に戻ると、ステップＳ８２では、中央制御装置３６及び音声認識プロセッサ５４は、映像データの現在フレームに対応する音声データがマイクロホンアレイ４から受信されたか否かを判定する。
【００９１】
ステップＳ８２で、音声データが受信されたと判定されれば、ステップＳ８４で、会議中の参加者のうち誰が話をしているかを判定するための処理を実行する。
【００９２】
図１２は、ステップＳ８４で実行される処理動作を示す。
【００９３】
図１２を参照すると、ステップＳ１４０では、方向プロセッサ５３はマイクロホンアレイ４からの音声データを処理して、音声が来ている方向を判定する。この処理は、例えば、英国特許第２１４０５５８号、米国特許第４３３３１７０号及び米国特許第３３９２３９２号に記載されているような従来の方式で実行される。
【００９４】
ステップＳ１４２では、アーカイブプロセッサ５８は、画像データの現在フレームについてステップＳ４４（図５）で頭部追跡装置５０により判定された各参加者の頭部の位置を読み取り、それに基づいて、どの参加者の頭部がステップＳ１４０で判定された方向、すなわち、音声が来ている方向に対応する位置にあるかを判定する。
【００９５】
ステップＳ１４４では、アーカイブプロセッサ５８は、音声が来ている方向に２人以上の参加者がいるか否かを判定する。
【００９６】
ステップＳ１４４で、音声が来ている方向には１人しか参加者がいないと判定されれば、ステップＳ１４６で、アーカイブプロセッサ５８は、音声が来ている方向にいる参加者を画像データの現在フレームの話者として選択する。
【００９７】
これに対し、ステップＳ１４４で、音声が来ている方向に対応する位置に２人以上の参加

者の頭部があると判定された場合には、ステップＳ１４８で、アーカイブプロセッサ５８は、画像データの直前フレームでそれらの参加者のうち１人が話者として識別されていたか否かを判定する。

【００９８】

ステップＳ１４８で、音声が来ている方向にいる参加者の１人が画像データの直前フレームで話者として選択されていたと判定されれば、ステップＳ１５０で、アーカイブプロセッサ５８は画像データの直前フレームで識別されていた話者を画像データの現在フレームについても話者として選択する。これは、画像データの直前フレームの話者が現在フレームの話者と同１人物である確率が高いからである。

【００９９】

これに対し、ステップＳ１４８で、音声が来ている方向にいる参加者がいずれも直前フレームで話者として識別された参加者ではないと判定された場合、又は直前フレームでは話者が識別されなかった場合には、ステップＳ１５２で、アーカイブプロセッサ５８は、音声が来ている方向にいるそれぞれの参加者を話者に「なりうる」参加者として選択する。

【０１００】

再び図８に戻ると、ステップＳ８６では、アーカイブプロセッサ５８は話者である参加者ごとの視線パラメータ値、すなわち、ステップＳ８０で生成された、話者である各参加者が誰を又は何を見ているかを定義する視線パラメータ値を後の解析に備えて、例えば、メモリ４２に格納する。

【０１０１】

ステップＳ８８では、アーカイブプロセッサ５８は、ステップＳ８４で判定された話者である各参加者のアイデンティティを音声認識プロセッサ５４に報知する。これに応答して、音声認識プロセッサ５４は話者である参加者の音声認識パラメータを音声認識パラメータ記憶装置５６から選択し、選択されたパラメータを使用して、受信した音声データに対して音声認識処理を実行し、話者である参加者が話した言葉に対応するテキストデータを生成する。

【０１０２】

他方、ステップＳ８２で、受信した音声データが話し言葉を含まないと判定されたならば、ステップＳ８４からステップＳ８８を省略する。

【０１０３】

ステップＳ８９では、アーカイブプロセッサ５８は、会議アーカイブデータベース６０にどの画像データを格納すべきか、すなわち、どのカメラ２－１、２－２及び２－３からの画像データを格納すべきかを判定する。

【０１０４】

図１３は、ステップＳ８９でアーカイブプロセッサ５８により実行される処理動作を示す。

【０１０５】

図１３を参照すると、ステップＳ１６０では、アーカイブプロセッサ５８は、画像データの現在フレームについてステップＳ８２（図８）で何らかの話し言葉が検出されたか否かを判定する。

【０１０６】

ステップＳ１６０で現在フレームについては話し言葉が存在しないと判定されれば、ステップＳ１６２で、アーカイブプロセッサ５８は、画像データを格納すべきカメラとしてデフォルトカメラを選択する。すなわち、この実施形態においては、アーカイブプロセッサ５８は直前フレームで画像データが記録されたカメラを選択する。処理中の現在フレームが全く初めてのフレームである場合には、アーカイブプロセッサ５８はカメラ２－１、２－２、２－３のうち１台を無作為に選択する。

【０１０７】

他方、ステップＳ１６０で、処理中の現在フレームに話し言葉があると判定された場合には、ステップＳ１６４で、アーカイブプロセッサ５８は、次の話者である参加者（初めて

ステップＳ１６４を実行するときには、これは最初の話者である参加者である）についてステップＳ８６で先に格納された視線パラメータを読み取り、その話者である参加者が見ている人物又は物体を判定する。

【０１０８】

ステップＳ１６６では、アーカイブプロセッサ５８は、現在考慮されている話者である参加者の頭部の位置と向き（図５のステップＳ４４で判定された）を、話者である参加者の視線の先にいる参加者の頭部の位置と向き（図５のステップＳ４４で判定された）又は話者である参加者の視線の先にある物体の位置と向き（図３のステップＳ２６で格納された）と共に読み取る。

【０１０９】

ステップＳ１６８では、アーカイブプロセッサ５８はステップＳ１６６で読み取られた位置と向きを処理して、カメラ２－１、２－２、２－３のうちどのカメラが話者である参加者と、話者である参加者が見ている参加者又は物体の双方を最も良く示しているかを判定し、且つこのカメラを現在フレームの画像データを会議アーカイブデータベース６０に格納すべきカメラとして選択する。

【０１１０】

図１４は、ステップＳ１６８でアーカイブプロセッサ５８により実行される処理動作を示す。

【０１１１】

図１４を参照すると、ステップＳ１７６では、アーカイブプロセッサ５８は次のカメラ（初めてステップＳ１７６を実行するときには、これは第１のカメラである）の三次元位置と視野方向を読み取る。この情報は先に図３のステップＳ２４で生成、格納されている。

【０１１２】

ステップＳ１７８では、アーカイブプロセッサ５８は、ステップＳ１７６で読み取られた情報を、話者である参加者の頭部の三次元位置と向き（図５のステップＳ４４で判定された）を定義する情報及び話者である参加者が見ている参加者の頭部の三次元位置と向き（図５のステップＳ４４で判定された）又は話者である参加者が見ている物体の三次元位置と向き（図３のステップＳ２６で格納された）を定義する情報と共に使用して、話者である参加者と、話者である参加者が見ている参加者又は物体の双方が現在考慮されているカメラの視野の中に入るか否か（すなわち、現在考慮されているカメラが話者である参加者と、話者である参加者が見ている参加者又は物体の双方を捉えることができるか否か）を判定する。すなわち、この実施形態においては、アーカイブプロセッサ５８は下記の式を評価し、全ての不等式が成立すれば、カメラは話者である参加者と、話者である参加者が見ている参加者又は物体の双方を捉えることができると判定する。

【０１１３】

【数１】

$$\left|\arccos\left[\frac{1}{\sqrt{(X_{p1}-X_c)^2+(Y_{p1}-Y_c)^2}}\begin{pmatrix}X_{p1}-X_c\\Y_{p1}-Y_c\end{pmatrix}\bullet\begin{pmatrix}dX_c\\dY_c\end{pmatrix}\right]\right|<\theta_h \qquad \cdots(1)$$

【０１１４】

【数２】

$$\left|\arccos\left[\frac{1}{\sqrt{(X_{p1}-X_c)^2+(Y_{p1}-Y_c)^2+(Z_{p1}-Z_c)^2}}\begin{pmatrix}X_{p1}-X_c\\Y_{p1}-Y_c\\Z_{p1}-Z_c\end{pmatrix}\bullet\begin{pmatrix}dX_c\\dY_c\\dZ_c\end{pmatrix}\right]\right|<\theta_v \qquad \cdots(2)$$

【０１１５】
【数３】

$$\left|\arccos\left[\frac{1}{\sqrt{(X_{p2}-X_c)^2+(Y_{p2}-Y_c)^2}}\begin{pmatrix}X_{p2}-X_c\\Y_{p2}-Y_c\end{pmatrix}\bullet\begin{pmatrix}dX_c\\dY_c\end{pmatrix}\right]\right|<\theta_h \qquad \cdots(3)$$

【０１１６】
【数４】

$$\left|\arccos\left[\frac{1}{\sqrt{(X_{p2}-X_c)^2+(Y_{p2}-Y_c)^2+(Z_{p2}-Z_c)^2}}\begin{pmatrix}X_{p2}-X_c\\Y_{p2}-Y_c\\Z_{p2}-Z_c\end{pmatrix}\bullet\begin{pmatrix}dX_c\\dY_c\\dZ_c\end{pmatrix}\right]\right|<\theta_v \qquad \cdots(4)$$

【０１１７】
式中、（X_c，Y_c，Z_c）は、それぞれ、カメラの主点のx座標、y座標及びz座標（図３のステップＳ２４で先に判定、格納されている）であり、
（dX_c、dY_c，dZ_c）は、それぞれ、x方向、y方向及びz方向のカメラの視野方向（同様に、図３のステップＳ２４で先に判定、格納されている）を表し、
θ_h及びθ_vは、それぞれ、水平方向と垂直方向のカメラの角視野（同様に、図３のステップＳ２４で判定、格納されている）であり、
（X_{p1}、Y_{p1}，Z_{p1}）は、それぞれ、話者である参加者の頭部の中心のx座標、y座標及びz座標（図５のステップＳ４４で判定されている）であり、
（dX_{p1}，dY_{p1}，dZ_{p1}）は、それぞれ、話者である参加者の視線９０の向き（同様に、図５のステップＳ４４で判定されている）を表し、
（X_{p2}，Y_{p2}，Z_{p2}）は、それぞれ、話者である参加者が見ている参加者の頭部の中心のx座標、y座標及びz座標（図５のステップＳ４４で判定されている）又は話者である参加者が見ている物体の面の中心のx座標、y座標及びz座標（図３のステップＳ２６で判定されている）であり、
（dX_{p2}，dY_{p2}，dZ_{p2}）は、それぞれ、話者である参加者が見ている参加者の視線９０のx方向、y方向及びz方向の方向（同様に、図５のステップＳ４４で判定されている）又は話者である参加者が見ている物体面に対する垂線のx方向、y方向及びz方向の方向（図３のステップＳ２６で判定されている）を表す。
【０１１８】
ステップＳ１７８で、カメラが話者である参加者と、話者である参加者が見ている参加者又は物体の双方を捉えることができる（すなわち、上記の式（１）、（２）、（３）及び（４）の不等式が成立する）と判定されれば、ステップＳ１８０で、アーカイブプロセッサ５８は、現在考慮されているカメラが話者である参加者を捉えている視野の画質を表す値を計算し、格納する。すなわち、この実施形態においては、アーカイブプロセッサ５８は、下記の式を使用して画質値Ｑ１を計算する。
【０１１９】
【数５】

$$Q1=\frac{1}{\sqrt{(X_c-X_{p1})^2+(Y_c-Y_{p1})^2+(Z_c-Z_{p1})^2}}\begin{pmatrix}X_c-X_{p1}\\Y_c-Y_{p1}\\Z_c-Z_{p1}\end{pmatrix}\bullet\begin{pmatrix}dX_{p1}\\dY_{p1}\\dZ_{p1}\end{pmatrix} \qquad \cdots(5)$$

【０１２０】

式中、各項の定義は先の式（１）及び（２）に関して挙げた定義と同じである。
【０１２１】
ステップＳ１８０で計算される画質値Q１は、－１から＋１の値をとるスカラであり、話者である参加者の頭部の背面がカメラに直接向いている場合、その値は－１であり、話者である参加者の顔面が直接カメラに向いている場合には＋１である。話者である参加者の頭部がその他の向きである場合には、－１と＋１の間の値をとる。
【０１２２】
ステップＳ１８２では、アーカイブプロセッサ５８は、現在考慮されているカメラが話者である参加者が見ている参加者又は物体を捉えている視野の画質を表す値を計算し、格納する。すなわち、この実施形態においては、アーカイブプロセッサ５８は下記の式を使用して、画質値Q２を計算する。
【０１２３】
【数６】

$$Q2 = \frac{1}{\sqrt{(X_c - X_{p2})^2 + (Y_c - Y_{p2})^2 + (Z_c - Z_{p2})^2}} \begin{pmatrix} X_c - X_{p2} \\ Y_c - Y_{p2} \\ Z_c - Z_{p2} \end{pmatrix} \bullet \begin{pmatrix} dX_{p2} \\ dY_{p2} \\ dZ_{p2} \end{pmatrix} \qquad \cdots (6)$$

【０１２４】
式中、パラメータの定義は先の式（３）及び（４）に関して挙げた定義と同じである。
【０１２５】
Q２も、同様に、参加者の頭部の背面又は物体の面の背面が直接カメラに向いている場合に－１、参加者の顔面又は物体の正面が直接カメラに向いている場合には＋１の値をとるスカラである。参加者の頭部又は物体の面がその他の向きである場合には、－１と＋１の間の値をとる。
【０１２６】
ステップＳ１８４では、アーカイブプロセッサ５８はステップＳ１８０で計算した画質値Q１と、ステップＳ１８２で計算した画質値Q２とを比較し、最小値を選択する。この最小値は、カメラが話者である参加者、あるいは話者である参加者が見ている参加者又は物体を捉えている「最悪の視野」を示す（Q１がQ２より小さい場合、最悪の視野は話者である参加者の視野であり、Q２がQ１より小さい場合は、最悪の視野は話者である参加者が見ている参加者又は物体の視野である）。
【０１２７】
他方、ステップＳ１７８で、式（１）、（２）、（３）及び（４）の不等式の１つ又は２つ以上が成立しない（すなわち、カメラが話者である参加者と、話者である参加者が見ている参加者又は物体の双方を捉えることができない）と判定された場合には、ステップＳ１８０からS１８４を省略する。
【０１２８】
ステップＳ１８６では、アーカイブプロセッサ５８は、画像データを提供していたカメラが他にあるか否かを判定する。カメラごとに上記の処理が実行され終わるまで、ステップＳ１７６からS１８６を繰り返す。
【０１２９】
ステップＳ１８８では、アーカイブプロセッサ５８は、ステップＳ１８４で処理を実行したときにカメラごとに格納された「最悪の視野」の値（すなわち、ステップＳ１８４でカメラごとに格納されたQ１又はQ２の値）を比較し、格納されているそれらの値の中で最大の値を選択する。この最大値は「最良の最悪の視野」を表し、そこで、ステップＳ１８８で、アーカイブプロセッサ５８は、ステップＳ１８４でこの「最良の最悪の視野」値が格納されていたカメラを会議アーカイブデータベースに画像データを格納すべきカメラとして選択する。これは、このカメラが話者である参加者と、話者である参加者が見ている参加者又は物体の双方を最良の視野で捉えているからである。

【０１３０】
ステップＳ１７０では、アーカイブプロセッサ５８は、話者に「なりうる」参加者を含めて、他に話者である参加者がいるか否かを判定する。話者である参加者ごとに、また、話者に「なりうる」参加者ごとに上記の処理が実行され終わるまで、ステップＳ１６４からＳ１７０を繰り返す。
【０１３１】
再び図８に戻ると、ステップＳ９０では、アーカイブプロセッサ５８はステップＳ８９で選択したカメラから受信された映像データの現在フレームと、マイクロホンアレイ４から受信された音声データとを従来の方式によりMPEG２データとして符号化し、符号化されたデータを会議アーカイブデータベース６０に格納する。
【０１３２】
図１５は、会議アーカイブデータベース６０へのデータの格納を概略的に示す。図１５に示す格納構造は概念上のものであり、格納される情報間のリンクを示すことにより理解を助けることを目的としている。従って、これは、会議アーカイブデータベース６０を構成するメモリにデータが厳密にどのように格納されるかを必ずしも表してはいない。
【０１３３】
図１５を参照すると、会議アーカイブデータベース６０は水平軸２００により表される時間情報を格納している。水平軸２００上の各単位は所定の量の時間、例えば、カメラから受信される映像データの１フレーム分の周期を表す。（会議アーカイブデータベース６０が一般には図１５に示す数より多くの数の時間単位を含むことは言うまでもなく了解されるであろう。）ステップＳ９０で生成されたMPEG２データは、タイミング情報（このタイミング情報は図１５では水平軸２００に沿ったMPEG２データ２０２の位置により概略的に表されている）と共に、データ２０２として会議アーカイブデータベース６０に格納されている。
【０１３４】
再び図８に戻ると、ステップＳ９２では、アーカイブプロセッサ５８は、現在フレームについてステップＳ８８で音声認識プロセッサ５４により生成されたテキストデータを会議アーカイブデータベース６０に格納する（図１５には２０４で示す）。すなわち、テキストデータは対応するMPEG２データへのリンクを伴って格納される。図１５においては、このリンクは、テキストデータがMPEG２データと同じ縦列に格納されることによって表されている。話をしていない参加者からは格納すべきテキストデータが得られないことは理解されるであろう。図１５に示す例では、参加者１については初めの１０個のタイムスロットにテキストが格納され（２０６で示す）、参加者３については１２番目から２０番目のタイムスロットに格納され（２０８で示す）、参加者４については２１番目のタイムスロットに格納されている（２１０で示す）。この例では、参加者２は図１５に示すタイムスロットの間は話をしていないので、参加者２のテキストは格納されていない。
【０１３５】
ステップＳ９４では、アーカイブプロセッサ５８は、ステップＳ８０で現在フレームについて参加者ごとに生成された視線パラメータ値を会議アーカイブデータベース６０に格納する（図１５には２１２で示す）。図１５を参照すると、視線パラメータ値は、参加者ごとに、関連するMPEG２データ２０２及び関連するテキストデータ２０４へのリンクと共に格納されている（このリンクは、図１５では、視線パラメータ値が関連するMPEG２データ２０２及び関連するテキストデータ２０４と同じ縦列にあることによって表されている）。従って、一例として、図１５の第１のタイムスロットに関していえば、参加者１の視線パラメータ値は、参加者１が参加者３を見ていることを指示する３であり、参加者２の視線パラメータ値は、参加者２がフリップチャート１４を見ていることを指示する５であり、参加者３の視線パラメータ値は、参加者３が参加者１を見ていることを指示する１であり、参加者４の視線パラメータ値は、参加者４が他の参加者の誰も見ていないことを指示する「０」である（図１に示す例では、１２で示される参加者は他の参加者ではなく、自分のメモを見ている）。

【０１３６】
ステップＳ９６では、中央制御装置３６及びアーカイブプロセッサ５８は、会議の参加者の１人が話を止めたか否かを判定する。この実施形態においては、この検査は、所定の参加者のテキストデータが直前のタイムスロットには存在したが、現在タイムスロットには存在しないことを判定するためにテキストデータ２０４を検査することにより実行される。いずれかの参加者についてこの条件が満たされれば（すなわち、参加者が話を止めていれば）、ステップＳ９８で、アーカイブプロセッサ５８は、話を止めた参加者ごとに、先にステップＳ８６を実行したときに格納されていた視線パラメータ値を処理し（それらの視線パラメータ値は、この時点で止まった話をしていた期間中にその参加者が誰を又は何を見ていたかを定義する）、視線ヒストグラムを定義するデータを生成する。すなわち、参加者が話をしていた期間の視線パラメータ値を処理して、その期間中に話者である参加者がその他の参加者及び物体のそれぞれを見ていた時間の割合（％）を定義するデータを生成するのである。
【０１３７】
図１６A及び図１６Bは、図１５のテキスと２０６及び２０８の期間にそれぞれ相当する視線ヒストグラムを示す。
【０１３８】
図１５及び図１６Aを参照すると、参加者１が話していた期間２０６の間、参加者１は、図１６に３００で示すように、１０個のタイムスロットのうち６個のタイムスロット（すなわち、参加者が話をしていた期間全体の長さの６０％）にわたり参加者３を見ており、図１６Aに３１０で示すように、１０個のタイムスロットのうち４個のタイムスロット（すなわち、時間の４０％）にわたり参加者４を見ていた。
【０１３９】
同様に、図１５及び図１６Bを参照すると、期間２０８の間、参加者３は、図１６Bに３２０で示すように、時間の約４５％にわたり参加者１を見ており、図１６Bに３３０で示すように、時間の約３３％にわたり物体５（すなわち、フリップチャート１４）を見ており、図１６Bに３４０で示すように、時間の約２２％にわたり参加者２を見ていた。
【０１４０】
再び図８に戻ると、ステップＳ１００では、ステップＳ９８で生成した各視線ヒストグラムを、それを生成する元になったテキストの関連する期間とリンクさせて、会議アーカイブデータベース６０に格納する。図１５を参照すると、格納される視線ヒストグラムは２１４で示され、テキスト期間２０６のヒストグラムを定義するデータは２１６で示され、テキスト期間２０８のヒストグラムを定義するデータは２１８で示されている。図１５においては、視線ヒストグラムと関連するテキストとのリンクは、視線ヒストグラムがテキストデータと同じ縦列に格納されることにより表されている。
【０１４１】
他方、ステップＳ９６で、現在時限について参加者の１人が話を止めていないと判定された場合には、ステップＳ９８及びS１００を省略する。
【０１４２】
ステップＳ１０２では、アーカイブプロセッサ５８は、映像フレームの直前フレーム（すなわち、ステップＳ８０からS１００でデータが生成、格納されたばかりのフレームの直前のフレーム）及び他の先行フレームについて、会議アーカイブデータベース６０に格納されているデータを必要に応じて修正する。
【０１４３】
図１７は、ステップＳ１０２でアーカイブプロセッサ５８により実行される処理動作を示す。
【０１４４】
図１７を参照すると、ステップＳ１９０では、アーカイブプロセッサ５８は、次の先行フレーム（初めてステップＳ１９０を実行する場合には、これは現在フレームの直前のフレームであり、すなわち、現在フレームが「ｉ」番目のフレームであれば、「ｉ－１」番目

のフレームである）について、話者に「なりうる」参加者のデータを会議アーカイブデータベース６０に格納するかどうかを判定する。

【０１４５】

ステップＳ１９０で、考慮されている先行フレームについて話者に「なりうる」参加者のデータが格納されていないと判定されれば、会議アーカイブデータベース６０のデータを修正する必要はない。

【０１４６】

他方、ステップＳ１９０で、考慮されている先行フレームについて話者に「なりうる」参加者のデータが格納されていると判定された場合には、ステップＳ１９２で、アーカイブプロセッサ５８は、先行フレームについてデータが格納された話者に「なりうる」参加者の１人が現在フレームについて識別された話者である参加者（話者に「なりうる」参加者ではない）、すなわち、図１２のステップＳ１４６で識別された話者である参加者と同１人物であるか否かを判定する。

【０１４７】

ステップＳ１９２で、先行フレームの話者に「なりうる」参加者がいずれも現在フレームについてステップＳ１４６で識別された話者である参加者と同じではないと判定されれば、考慮されている先行フレームについて会議アーカイブデータベース６０に格納されているデータの修正を実行しない。

【０１４８】

他方、ステップＳ１９２で、先行フレームの話者に「なりうる」参加者が現在フレームについてステップＳ１４６で識別された話者である参加者と同１人物であると判定された場合には、ステップＳ１９４で、アーカイブプロセッサ５８は、現在フレームの話者である参加者と同じではない話者に「なりうる」参加者のそれぞれについて、考慮されている先行フレームのテキストデータ２０４を会議アーカイブデータベース６０から削除する。

【０１４９】

以上説明したようにステップＳ１９０、S１９２及びS１９４の処理を実行することにより、現在フレームについて画像データ及び音声データを処理することによって話者が明確に識別された場合、現在フレームの話者は先行フレームの話者と同１人物であるという仮定を利用して、話者に「なりうる」参加者（すなわち、曖昧さなく話者を識別することが不可能であったため）について格納された直前フレームのデータを更新するのである。

【０１５０】

ステップＳ１９４を実行した後、次の先行フレームについてステップＳ１９０からS１９４を繰り返す。すなわち、現在フレームが「ｉ」番目のフレームであれば、初めてステップＳ１９０からS１９４を実行するときに「ｉ－１」番目のフレームを考慮し、２度目にステップＳ１９０からS１９４を実行するときには「ｉ－２」番目のフレームを考慮する。これ以降も同様である。ステップＳ１９０で、考慮されている先行フレームについて話者に「なりうる」参加者のデータが格納されていないと判定されるか、またはステップＳ１９２で、考慮されている先行フレームにおける話者に「なりうる」参加者がいずれも現在フレームについて曖昧さなく識別された話者である参加者と同じではないと判定されるまで、ステップＳ１９０からS１９４を繰り返し実行し続ける。このようにして、いくつかの連続するフレームにわたり話者に「なりうる」参加者が識別された場合には、次のフレームで話者に「なりうる」参加者の中から実際の話者である参加者が識別されれば、会議アーカイブデータベースに格納されているデータを修正する。

【０１５１】

再び図８に戻ると、ステップＳ１０４では、中央制御装置３６は、カメラ２－１、２－２、２－３から映像データの別のフレームが受信されたか否かを判定する。カメラ２－１、２－２、２－３から画像データが受信されている間は、ステップＳ８０からS１０４を繰り返し実行する。

【０１５２】

会議アーカイブデータベース６０にデータが格納されている場合、会議アーカイブデータ

ベース６０を問い合わせて、会議に関連するデータを検索しても良い。

【０１５３】

図１８は、ユーザにより指定される探索基準を満たす会議の各部分に関連するデータを検索する目的で会議アーカイブデータベース６０を探索するために実行される処理動作を示す。

【０１５４】

図１８を参照すると、ステップＳ２００では、中央制御装置３６は表示プロセッサ６４に、要求する会議アーカイブデータベース６０の探索を定義する情報をユーザが入力することを求めるメッセージを表示装置２６に表示させる。すなわち、この実施形態においては、中央制御装置３６は図１９Aに示す表示を表示装置２６に表示させる。

【０１５５】

図１９Aを参照すると、ユーザは、会議アーカイブデータベース６０の中で見出すことを臨む会議の部分を定義する情報を入力することを求められる。すなわち、この実施形態においては、ユーザは、話をしていた参加者を定義する情報４００、情報４００の中で識別される参加者が口に出した１つ又は複数のキーワードから成る情報４１０、及び情報４００の中で識別される参加者が話している間に見ていた参加者又は物体を定義する情報４２０を入力することを求められる。更に、ユーザは、探索を実行すべき会議の部分を定義する時間情報を入力することができる。すなわち、ユーザは、その時間を越えたら探索を中止すべきである会議中の時間（すなわち、指定される時間に至るまでの会議の期間を探索すべきである）を定義する情報４３０と、その時間から探索を実行すべきである会議中の時間を定義する情報４４０と、探索を実行すべき期間の開始時間と終了時間をそれぞれ定義する情報４５０及び４６０とを入力できる。この実施形態においては、情報４３０、４４０、４５０及び４６０は、例えば、分単位で絶対期限として時間を指定するか、又は会議時間全体に占める割合を指示する小数値を入力することにより相対期限で時間を指定することにより入力されれば良い。例えば、情報４３０として値０．２５を入力した場合、探索は会議の初めの四分の一に限定されるであろう。

【０１５６】

この実施形態では、ユーザは１回の探索で情報４００、４１０及び４２０の全てを入力する必要はなく、この情報のうち１つ又は２つを省いても良い。ユーザが情報４００、４１０及び４２０の全てを入力すれば、会議の中で、情報４００の中で識別される参加者が情報４２０の中で識別される参加者又は物体に向かって話していた部分及び情報４００の中で識別される参加者が情報４１０の中で定義されるキーワードを話した部分をそれぞれ識別するための探索が実行されることになる。これに対し、情報４１０を省いた場合には、会議の中で、情報４００の中で識別される参加者が何を言ったかに関わらず、情報４２０の中で定義される参加者又は物体に向かって話していた部分をそれぞれ識別するための探索が実行される。情報４１０及び４２０を省いた場合には、会議の中で、情報４００の中で識別される参加者が何を誰に向かって話したかに関わらず、話していた部分をそれぞれ識別するための探索が実行される。情報４００を省いた場合には、会議の中で、いずれかの参加者が情報４２０の中で定義される参加者又は物体を見ながら情報４１０の中で定義されるキーワードを話した部分をそれぞれ識別するための探索が実行される。情報４００及び４１０を省いた場合には、会議の中で、いずれかの参加者が情報４２０の中で定義される参加者又は物体に向かって話した部分をそれぞれ識別するための探索が実行される。情報４２０を省いた場合には、会議の中で、情報４００の中で定義される参加者が情報４１０の中で定義されるキーワードを誰に向かって話したかに関わらず、キーワードを話した部分をそれぞれ識別するための探索が実行される。同様に、情報４００及び４２０を省いた場合には、会議の中で、誰が誰に向かって言ったかに関わらず、情報４１０の中で識別されるキーワードが話された部分をそれぞれ識別するための探索が実行される。

【０１５７】

更に、ユーザは時間情報４３０、４４０、４５０及び４６０の全てを入力しても良いし、あるいはそのうちいくつかを省いても良い。

【０１５８】
また、探索者が言葉の組み合わせ又はある言葉に代わる言葉を探索できるようにするために、情報４１０の中で入力されるキーワードと組み合わせて周知のブール演算子及び探索アルゴリズムを使用しても良い。
【０１５９】
探索を定義するためにユーザが必要な全ての情報を入力したならば、マウス３０などのユーザ入力装置を使用して領域４７０をクリックすることにより探索を開始する。
【０１６０】
再び図１８に戻ると、ステップＳ２０２では、ユーザが入力した探索情報を中央制御装置３６により読み取り、命令された探索を実行する。すなわち、この実施形態においては、中央制御装置３６は情報４００又は４２０の中で入力された参加者又は物体の名前をテーブル８０（図４）を使用して識別番号に変換し、情報４００で定義される参加者（情報４００が入力されなかった場合は全ての参加者）についてテキスト情報２０４を考慮する。ユーザにより情報４２０が入力されていれば、テキストの期間ごとに、中央制御装置３６は対応する視線ヒストグラムを定義するデータを検査して、情報４２０の中で定義される参加者又は物体のヒストグラムにおける注目時間の割合がこの実施形態では２５％である閾値以上であるか否かを判定する。このように、話者である参加者が話をしている時間の少なくとも２５％にわたって情報４２０の中で定義される参加者又は物体を見ていれば、話者である参加者が話しの間に他の参加者又は物体を見たとしても、話し言葉（テキスト）の各期間を考慮して、情報４００の中で定義される参加者は情報４２０の中で定義される参加者又は物体に話しかけていたという基準を満たす。従って、例えば、情報４２０の中で２人以上の参加者が識別されていれば、視線ヒストグラムの値が２人以上の参加者について２５％以上であるような話の期間が識別されるであろう。ユーザが情報４１０を入力した場合、中央制御装置３６及びテキストサーチャ６２は、先に情報４００及び４２０に基づいて識別されたテキストの各部分（情報４００及び４２０が入力されていなければ、テキストの全ての部分）を探索して、情報４１０の中で定義されるキーワードを含む各部文を識別する。ユーザが時間情報を入力していれば、上記の探索はそれらの期限により定義される会議時間に限られる。
【０１６１】
ステップＳ２０４では、中央制御装置３６は表示プロセッサ６４に、探索中に識別された関連話題のリストを表示装置２６を介してユーザに対し表示させる。すなわち、中央制御装置３６は、図１９Bに示すような情報をユーザに対し表示させる。図１９Bを参照すると、探索パラメータを満足させる各々の話題のリストが作成され、その話の開始時間を定義する情報が絶対期限と、会議時間全体に占める割合の双方で表示されている。そこで、ユーザは、例えば、リスト中の必要な話題をマウス３０を使用してクリックすることにより、話題の１つを選択して、再生させることができる。
【０１６２】
ステップＳ２０６では、中央制御装置３６はステップＳ２０４でユーザが行った選択を読み取り、格納されている会議の関連部分のMPEG２データ２０２を会議アーカイブデータベース６０から再生させる。すなわち、中央制御装置３６及び表示プロセッサ６４はMPEG２データ２０２を復号し、画像データと音声を表示装置２６を介して出力するのである。再生すべき話の一部又は全てについて２台以上のカメラからの画像データが格納されている場合には、そのことを表示装置２６によりユーザに指示し、ユーザは、例えば、キーボード２８を使用して中央制御装置３６に命令を入力することにより、再生すべき画像データを選択することができる。
【０１６３】
ステップＳ２０８では、中央制御装置３６は、ユーザが会議アーカイブデータベース６０の問い合わせを中止することを望むか否かを判定し、望まないのであれば、ステップＳ２００からS２０８を繰り返す。
【０１６４】

以上説明した本発明の実施形態に対しては、様々な変形や変更を実施することができる。

【０１６５】

上記の実施形態では、ステップＳ３４（図３）及びステップＳ７０（図７）においては、会議中の各参加者の頭部を追跡していた。しかし、これに加えて、ステップＳ４及びＳ２６でデータを格納した物体が移動する場合（そのような物体としては、例えば、参加者により回覧されるようなメモ又は参加者間で手渡されるべき物体などが考えられる）それらの物体を追跡することも可能であろう。

【０１６６】

上記の実施形態では、複数台のビデオカメラ２－１、２－２、２－３からの画像データを処理していた。しかし、その代わりに、１台のビデオカメラからの画像データを処理しても良い。この場合、例えば、ステップＳ４２－１（図５）のみを実行し、ステップＳ４２－２からＳ４２－nを省略する。同様に、ステップＳ４４を省略し、ステップＳ４２－１で実行される処理の間、画像データの現在フレームに関わる参加者の頭部の３D位置及び向きをステップＳ５８（図６）で判定される３D位置及び向きであるとみなす。ステップＳ４６では、カルマンフィルタに入力される頭部の特徴の位置はその１台のカメラからの画像データにおける位置になるであろう。更に、会議アーカイブデータベース６０に画像データを記録すべきカメラを選択するためのステップＳ８９（図８）も省略されるであろう。

【０１６７】

上記の実施形態では、ステップＳ１６８（図１３）において、話者である参加者と、話者である参加者が見ている参加者又は物体とを最も良く捉えるカメラを識別するための処理を実行していた、しかし、上記の実施形態において説明したようにカメラを識別する代わりに、処理装置２４の初期設定中に、会議テーブルを囲む２つずつの着席位置を最も良く捉え且つ／又は各々の着席位置と所定の物体（フリップチャート１４など）を最も良く捉えるのはカメラ２－１、２－２、２－３のうちどれであるかをユーザが定義することも可能である。このようにして、話者である参加者と、話者である参加者が見ている参加者があらかじめ定義された着席位置にいると判定されれば、それらのあらかじめ定義された着席位置を最も良く捉えるとユーザにより定義されたカメラを画像データを格納すべきカメラとして選択することができる。同様に、話者である参加者があらかじめ定義された位置にあり且つある物体を見ている場合、そのあらかじめ定義された着席位置と物体を最も良く捉えるとユーザにより定義されたカメラを画像データを格納すべきカメラとして選択することができる。

【０１６８】

上記の実施形態では、ステップＳ１６２（図１３）において、直前フレームで画像データが格納されたカメラとしてデフォルトカメラを選択していた。しかし、その代わりに、例えば、処理装置２４の初期設定中に、ユーザがデフォルトカメラを選択しても良い。

【０１６９】

上記の実施形態では、ステップＳ１９４（図１７）において、その時点で実際には話者である参加者として識別されなかった話者に「なりうる」参加者について、テキストデータ２０４を会議アーカイブデータベース６０から削除していた。しかし、これに加えて、関連する視線ヒストグラムデータ２１４も共に削除して良い。更に、カメラ２－１、２－２、２－３のうち２台以上からのMPEG２データ２０２が格納されていた場合、話者に「なりうる」参加者に関連するMPEG２データも削除して良い。

【０１７０】

上記の実施形態では、話者である参加者を一意性をもって識別することが不可能である場合、話者に「なりうる」参加者を定義し、話者になりうる参加者についてデータを処理して会議アーカイブデータベース６０に格納し、その後、会議アーカイブデータベース６０に格納されたデータを修正していた（図８のステップＳ１０２）。しかし、話者に「なりうる」参加者についてデータを処理し、格納するのではなく、カメラ２－１、２－２及び２－３から受信した映像データと、マイクロホンアレイ４から受信した音声データとを、

後続フレームに関連するデータから話者である参加者が識別されたときの後の処理及びアーカイビングに備えて格納しておいても良い。あるいは、ステップＳ１４４（図１２）で実行された処理の結果、音声が来ている方向に２人以上の参加者がいることが指示された場合には、カメラ２－１、２－２及び２－３からの画像データを処理して、参加者の唇の動きを検出すると共に、音声が来ている方向にいて、唇が動いている参加者を話者である参加者として選択しても良い。

【０１７１】

上記の実施形態では、各人物の頭部の位置と、各人物の頭部の向きと、各人物が誰を又は何を見ているかを定義する人物ごとの視線パラメータとを判定するための処理を実行していた。その後、画像データのフレームごとに、各人物の視線パラメータ値を会議アーカイブデータベース６０に格納する。しかし、全ての人物について視線パラメータを判定する必要はない。例えば、話者である参加者のみの視線パラメータを判定し、画像データのフレームごとにこの視線パラメータ値のみを会議アーカイブデータベース６０に格納することは可能である。従って、この場合、話者である参加者の頭部の位置を判定するだけで良いであろう。このようにすれば、処理及び格納に課される負担を軽減することができる。

【０１７２】

上記の実施形態では、ステップＳ２０２（図１８）において、テキストの特定の部分の視線ヒストグラムを考慮し、その視線ヒストグラムにおいて別の参加者又は物体に注目している時間の割合が所定の閾値以上である場合に、参加者は別の参加者と話していた又は別の物体を見ていたと判定していた。しかし、閾値を使用する代わりに、テキスト（話）の期間中、話者である参加者が見ていた参加者又は物体は、視線ヒストグラムの中で最も大きな割合の注目時間を有する参加者又は物体（例えば、図１６Aの参加者３及び図１６Bの参加者１）であると定義しても良い。

【０１７３】

上記の実施形態では、カメラ２－１、２－２及び２－３と、マイクロホンアレイ４とからデータが受信されている間、MPEG２データ２０２、テキストデータ２０４、視線パラメータ２１２及び視線ヒストグラム２１４をリアルタイムで会議アーカイブデータベース６０に格納していた。しかし、その代わりに、映像データと音声データを格納し、リアルタイムではなくデータ２０２、２０４、２１２０及び２１４を生成して、会議アーカイブデータベース６０に格納しても良い。

【０１７４】

上記の実施形態では、会議の定義された部分についてデータを検索するために会議アーカイブデータベース６０を問い合わせる前に、MPEG２データ２０２、テキストデータ２０４、視線パラメータ２１２及び視線ヒストグラム２１４を生成し、データベースに格納していた。しかし、探索の要求に先立ってデータを生成、格納するのではなく、会議アーカイブデータベース６０の探索がユーザにより要求されるのに応答して、既に会議アーカイブデータベース６０に格納されているデータを処理することにより、視線ヒストグラムデータ２１４の一部又は全てを生成しても良い。例えば、上記の実施形態では、視線ヒストグラム２１４はステップＳ９８及びS１００（図８）でリアルタイムで計算、格納されていたが、ユーザにより入力される探索要求に応答してそれらのヒストグラムを計算しても良い。

【０１７５】

上記の実施形態では、テキストデータ２０４は会議アーカイブデータベース６０に格納されていた。テキストデータ２０４の代わりに、音声データを会議アーカイブデータベース６０に格納しても良い。その場合、格納された音声データ自体を音声認識処理を利用してキーワードを求めて探索しても良いし、あるいは音声認識処理を利用して音声データをテキストに変換し、従来のテキストサーチャを使用してそのテキストを探索しても良い。

【０１７６】

上記の実施形態では、処理装置２４は、アーカイブすべきデータを受信し、生成するための機能構成要素（例えば、中央制御装置３６、頭部追跡装置５０、頭部モデル記憶装置５

２、方向プロセッサ５３、音声認識プロセッサ５４、音声認識パラメータ記憶装置５６及びアーカイブプロセッサ５８）と、アーカイブデータを格納するための機能構成要素（例えば、会議アーカイブデータベース６０）と、データベースを探索し、そこから情報を検索するための機能構成要素（例えば、中央制御装置３６及びテキストサーチャ６２）とを含む。しかし、これらの機能構成要素を別個の装置に設けても良い。例えば、アーカイブすべきデータを生成する１つ又は複数の装置と、データベースを探索する１つ又は複数の装置とをインターネットなどのネットワークを介して１つ又は複数のデータベースに接続しても良い。

【０１７７】

また、図２０を参照して説明すると、１つ又は複数の場所での会議５００、５１０、５２０から得られた映像データと音声データをデータ処理・データベース記憶装置５３０（アーカイブデータを生成し、格納するための機能構成要素を具備する）に入力し、データベースを問い合わせて、そこから情報を検索するために、１つ又は複数のデータベース問い合わせ装置５４０、５５０をデータ処理・データベース記憶装置５３０に接続しても良い。

【０１７８】

上記の実施形態では、プログラミング命令により定義される処理ルーチンを使用して、コンピュータにより処理を実行していた。しかし、処理の一部又は全てをハードウェアを使用して実行することも可能であろう。

【０１７９】

複数の参加者の間で行われる会議に関して上記の実施形態を説明したが、本発明はこの用途には限定されず、フィルムセットについて画像データ及び音声データを処理するなどの他の用途にも適用することができる。

【０１８０】

上記の変形の異なる組み合わせも言うまでもなく可能であり、本発明の趣旨から逸脱せずにその他の変更や変形を実施することができる。

【０１８１】

＜第２の実施形態＞

図２１を参照すると、この実施形態では、１台のビデオカメラ６０２と、１つ又は複数のマイクロホン６０４とを使用して、複数の人物６０６、６０８、６１０、６１２の間で行われている会議から画像データと音声データをそれぞれ記録している。

【０１８２】

ビデオカメラ６０２からの画像データと、マイクロホン６０４からの音声データはケーブル（図示せず）を介してコンピュータ６２０に入力され、コンピュータ６２０は受信したデータを処理し、データをデータベースに格納して、会議のアーカイブ記録を作成する。後に、このデータベースから情報を検索することができる。

【０１８３】

コンピュータ６２０は、従来のように、表示装置６２６や、この実施形態においてはキーボード６２８及びマウス６３０であるユーザ入力装置と共に、１つ又は複数のプロセッサ、メモリ、サウンドカードなどを含む処理装置６２４を有する従来通りのパーソナルコンピュータである。

【０１８４】

コンピュータ６２０の構成要素と、それらの構成要素に対し入出力されるデータの流れを図２２に概略的に示す。

【０１８５】

図２２を参照すると、処理装置６２４は、例えば、ディスク６３２などのデータ記憶媒体に格納されたデータとして及び／又は例えば、インターネットなどの通信ネットワーク（図示せず）を介する送信又は大気中を通る送信により遠隔データベースから処理装置６２４に入力され且つ／又はキーボード６２８などのユーザ入力装置又は他の入力装置を介してユーザにより処理装置６２４に入力される信号６３４として入力されるプログラミング

命令に従って動作するようにプログラムされている。

【０１８６】

プログラミング命令によりプログラムされると、処理装置６２４は処理動作を実行するための複数の機能ユニットとして有効に構成される。そのような機能ユニットの例と、それらの配線を図２２に示す。しかし、図２２に示すユニットと配線は概念上のものであり、単に理解を助けるために例示を目的として示されているにすぎない。従って、図２２の機能ユニット及び配線は、処理装置６２４のプロセッサ、メモリなどが実際に構成される厳密なユニットや接続関係を必ずしも表してはいない。

【０１８７】

図２２に示す機能ユニットに関して説明すると、中央制御装置６３６はユーザ入力装置６２８，６３０からの入力を処理し、且つユーザによりディスク６３８などの記憶装置に格納されたデータとして又は処理装置６２４へ送信される信号６４０として処理装置６２４に入力されるデータを受信する。また、中央処理装置６３６はその他の複数の機能ユニットに対して制御及び処理を実行する。メモリ６４２は、中央制御装置６３６及びその他の機能ユニットにより使用されるべきメモリである。

【０１８８】

頭部追跡装置６５０はビデオカメラ６０２から受信した画像データを処理して、会議中の各々の参加者６０６、６０８、６１０，６１２の頭部の位置と向きを三次元で追跡する。この実施形態では、この追跡を実行するために、頭部追跡装置６５０は各々の参加者の頭部の三次元コンピュータモデルを定義するデータと、頭部の特徴を定義するデータとを使用する。それらのデータは、後述するように、頭部モデル記憶装置６５２に格納される。

【０１８９】

音声認識プロセッサ６５４はマイクロホン６０４から受信される音声データを処理する。音声認識プロセッサ６５４は、「Dragon Dictate」又はIBMの「ViaVoice」などの従来の音声認識プログラムに従って動作し、参加者６０６、６０８、６１０、６１２により話された言葉に対応するテキストデータを生成する。音声認識処理を実行するために、音声認識プロセッサ６５４は、参加者６０６、６０８、６１０、６１２ごとの音声認識パラメータを定義するデータを使用する。このデータは音声認識パラメータ記憶装置６５６に格納される。すなわち、音声認識パラメータ記憶装置６５６に格納されるデータは、音声認識プロセッサを従来の方式で訓練することにより生成される各参加者の音声プロファイルを定義するデータである。例えば、このデータは、訓練後にDragon Dictateの「ユーザファイル」に格納されるデータである。

【０１９０】

アーカイブプロセッサ６５８は、頭部追跡装置６５０及び音声認識プロセッサ６５４から受信したデータを使用して、会議アーカイブデータベース６６０に格納すべきデータを生成する。すなわち、後述するように、カメラ６０２からの映像データとマイクロホン６０４からの音声データを、音声認識プロセッサ６５４からのテキストデータ及び会議中の各参加者が所定の時点で誰を見ていたかを定義するデータと共に会議アーカイブデータベース６６０に格納するのである。

【０１９１】

テキストサーチャ６６２は、中央制御装置６３６と関連して、会議アーカイブデータベース６６０を探索して、後に更に詳細に説明するように、ユーザにより指定される探索基準に適合する会議の１つ又は複数の部分に対応する音声データと映像データを見出し、再生するために使用される。

【０１９２】

表示プロセッサ６６４は、中央制御装置６３６の制御の下に、表示装置６２６を介してユーザに情報を表示すると共に、会議アーカイブデータベース６６０に格納されている音声データと映像データを再生する。

【０１９３】

出力プロセッサ６６６はアーカイブデータベース６６０からのデータの一部又は全てを、

例えば、ディスク６６８などの記憶装置へ出力するか、又は信号６７０として出力する。
【０１９４】
会議を始める前に、処理装置６２４が要求される処理動作を実行できるようにするために必要なデータを入力することによりコンピュータ６２０を初期設定しなければならない。
【０１９５】
図２３は、この初期設定中に処理装置６２４により実行される処理動作を示す。
【０１９６】
図２３を参照すると、ステップＳ３０２では、中央制御装置６３６は表示プロセッサ６６４に、ユーザが会議に参加する各人物の名前を入力することを要求するメッセージを表示装置６２６に表示させる。
【０１９７】
ステップＳ３０４では、中央制御装置６３６は、例えば、ユーザがキーボード６２８を使用して入力した名前を定義するデータを受信すると、各参加者に独自の参加者番号を割り当て、参加者番号と参加者の名前との関係を定義するデータ、例えば、図２４に示すテーブル６８０を会議アーカイブデータベース６６０に格納する。
【０１９８】
ステップＳ３０６では、中央制御装置６３６は頭部モデル記憶装置６５２を探索して、会議の参加者ごとに頭部モデルを定義するデータが既に格納されているか否かを判定する。
【０１９９】
ステップＳ３０６で、１人又は２人以上の参加者について頭部モデルがまだ格納されていないと判定されれば、ステップＳ３０８で、中央制御装置６３６は表示プロセッサ６６４に、モデルがまだ格納されていない各参加者の頭部モデルを定義するデータをユーザが入力することを要求するメッセージを表示装置６２６に表示させる。
【０２００】
これに応答して、ユーザは、例えば、ディスク６３８などの記憶媒体に格納されたデータとして要求された頭部モデルを定義するデータを入力するか、又は接続している処理装置から信号６４０としてこのデータをダウンロードすることによりデータを入力する。そのような頭部モデルは、例えば、Valente他の「An Analysis／Synthesis Cooperation for Head Tracking andVideo Face Cloning」（ProceedingsECCV’９８Workshop on Perceptionof Human Action、ドイツ、フライブルク大学、１９９８年６月６日に掲載）に記載されているような従来の方式で生成されれば良い。
【０２０１】
ステップＳ３１０では、中央制御装置６３６は、ユーザにより入力されたデータを頭部モデル記憶装置６５２に格納する。
【０２０２】
ステップＳ３１２では、中央制御装置６３６及び表示プロセッサ６６４は、ユーザにより入力された各々の三次元コンピュータ頭部モデルをレンダリングして、ユーザが各モデルにおいて少なくとも７つの特徴を識別することを要求するメッセージと共に、モデルをユーザに対し表示装置６２６を介して表示する。
【０２０３】
これに応答して、ユーザは、マウス６３０を使用して、参加者の頭部の正面、側面及び（可能であれば）背面にある顕著な特徴、例えば、目尻、鼻孔、口、耳又は参加者がかけている眼鏡の特徴などに対応する点を各モデルで指定する。
【０２０４】
ステップＳ３１４では、中央制御装置６３６はユーザにより定義された特徴を頭部モデル記憶装置６５２に格納する。
【０２０５】
他方、ステップＳ３０６で、参加者ごとに頭部モデルが頭部モデル記憶装置６５２に既に記憶されていると判定された場合には、ステップＳ３０８からS３１４を省略する。
【０２０６】

ステップＳ３１６では、中央制御装置６３６は音声認識パラメータ記憶装置６５６を探索して、参加者ごとに音声認識パラメータが既に格納されているか否かを判定する。

【０２０７】

ステップＳ３１６で、一部の参加者について音声認識パラメータを利用できないと判定されれば、ステップＳ３１８で、中央制御装置６３６は表示プロセッサ６６４に、パラメータがまだ格納されていない各参加者についてユーザが音声認識パラメータを入力することを要求するメッセージを表示装置６２６に表示させる。

【０２０８】

これに応答して、ユーザは、例えば、ディスク６３８などの記憶媒体に格納されたデータとして又は遠隔処理装置からの信号６４０として、必要な音声認識パラメータを定義するデータを入力する。先に述べた通り、これらのパラメータはユーザの音声のプロファイルを定義し、音声認識プロセッサを従来の方式で訓練することにより生成される。従って、例えば、Dragon Dictateを組み込んだ音声認識プロセッサの場合、ユーザにより入力される音声認識パラメータはDragon Dictateの「ユーザファイル」に格納されたパラメータに相当する。

【０２０９】

ステップＳ３２０では、中央制御装置６３６は、ユーザにより入力されたデータを音声認識パラメータ記憶装置６５６に格納する。

【０２１０】

他方、ステップＳ３１６で、参加者ごとに音声認識パラメータが既に利用可能な状態になっていると判定された場合には、ステップＳ３１８及びS３２０を省略する。

【０２１１】

ステップＳ３２２では、中央制御装置６３６は表示プロセッサ６６４に、ユーザがカメラ６０２の校正を可能にするためのステップを実行することを要求するメッセージを表示装置６２６に表示させる。

【０２１２】

これに応答して、ユーザは必要なステップを実行し、ステップＳ３２４では、中央制御装置６３６はカメラ６０２を校正するための処理を実行する。すなわち、この実施形態においては、ユーザにより実行されるステップ及び中央制御装置６３６により実行される処理は、Wiles及びDavisonの「Calibrating and３D Modelling with a Multi－Camera System」（１９９９IEEE Workshop on Multi－View Modelling and Analysis of Visual Scenes、ISBN ０７６９５０１１０９に掲載）に記載されているような方式で実行される。これにより、会議室に対するカメラ６０２の位置と向きを定義する校正データと、カメラの固有パラメータ（横縦比、焦点距離、主点及び一次半径方向ひずみ係数）とが生成される。校正データはメモリ６４２に格納される。

【０２１３】

ステップＳ３２６では、中央制御装置６３６は表示プロセッサ６６４に、会議の次の参加者（初めてステップＳ３２６を実行する場合には、これは最初の参加者である）が着席することを要求するメッセージを表示装置６２６に表示させる。

【０２１４】

ステップＳ３２８では、要求された参加者に着席する時間を与えるために、処理装置６２４は所定の時間待機し、その後、ステップＳ３３０で、中央制御装置６３６はカメラ６０２からの画像データを処理して、着席した参加者の頭部の推定位置を判定する。すなわち、この実施形態においては、中央制御装置６３６は、参加者の肌の色に対応する色（この色は、頭部モデル記憶装置６５２に格納されている参加者の頭部モデルを定義するデータから判定される）を有する、カメラ６０２からの画像データの１フレーム中の各部分を識別するために従来通りの処理を実行し、次に、会議室内の最も高い位置に相当する部分を選択する（頭部は人体の中で最も高い位置にある肌色の部分であると想定されるため）。画像中の識別された部分の位置と、ステップＳ３２４で判定されたカメラ校正パラメータとを使用して、中央制御装置６３６は従来の方式により頭部の三次元推定位置を判定する

。
【０２１５】
ステップＳ３３２では、中央制御装置６３６は参加者の頭部の三次元推定向きを判定する。すなわち、この実施形態においては、中央制御装置６３６は頭部モデル記憶装置６５２に格納されている参加者の頭部の三次元コンピュータモデルを複数の異なる向きについてレンダリングして、向きごとにそれぞれ対応するモデルの二次元画像を生成し、モデルの各二次元画像を参加者の頭部を示す、カメラ６０２からの映像フレームの部分と比較し、モデルの画像が映像データと最も良く整合する向きを選択する。この実施形態では、参加者の頭部のコンピュータモデルを１０８の異なる向きでレンダリングして、カメラ６０２からの映像データと比較すべき画像データを生成する。これらの向きは頭部モデルを０°（正面を向いている）、＋４５°（上を向いている）及び－４５°（下を向いている）に相当する３つの頭部の傾きのそれぞれについて１０°ずつのステップで３６回回転させた向きに相当する。頭部モデルをレンダリングすることにより生成された画像データをカメラ６０２からの映像データと比較するときには、例えば、Schodl、Haro及びEssaの「Head Tracking Using a Textured Polygonal Model」（Proceedings１９９８Workshop on Perceptual User Interfacesに掲載）に記載されているような従来の技法を使用する。
【０２１６】
ステップＳ３３４では、ステップＳ３３０で生成した参加者の頭部の推定位置と、ステップＳ３３２で生成した参加者の頭部の推定向きとを頭部追跡装置６５０に入力し、カメラ６０２から受信した画像データのフレームを処理して、参加者の頭部を追跡する。すなわち、この実施形態においては、頭部追跡装置６５０は、例えば、Valente他の「An Analysis／Synthesis Cooperation for Head Tracking andVideo Face Cloning」（Proceedings EECV’９８Workshop on Perceptionof Human Action、ドイツ、フライブルク大学、１９９８年６月６日）に記載されているような従来の方式で頭部を追跡するための処理を実行する。
【０２１７】
図２５は、ステップＳ３３４で頭部追跡装置６５０により実行される処理動作の概要を示す。
【０２１８】
図２５を参照すると、ステップＳ３５０では、頭部追跡装置６５０は参加者の頭部の現在推定3D位置及び現在推定3D向きを読み取る。ステップＳ３５０を初めて実行する場合には、これらは図２３のステップＳ３３０及びS３３２で生成される推定位置及び推定向きである。
【０２１９】
ステップＳ３５２では、頭部追跡装置６５０はステップＳ３２４で生成されたカメラ校正データを使用して、頭部モデル記憶装置６５２に格納されている参加者の頭部の三次元コンピュータモデルをステップＳ３５０で読み取った推定位置及び推定向きに従ってレンダリングする。
【０２２０】
ステップＳ３５４では、頭部追跡装置６５０はカメラ６０２から受信された映像データの現在フレームについて画像データを処理し、ユーザにより識別され、ステップＳ３１４で格納された頭部の特徴のうち１つの特徴の期待位置を取り囲む各々の領域から画像データを取り出す。それらの期待位置は、ステップＳ３５０で読み取った推定位置及び推定向きと、ステップＳ３２４で生成されたカメラ校正データとから判定される。
【０２２１】
ステップＳ３５６では、頭部追跡装置６５０はステップＳ３５２でレンダリングし、生成した画像データと、ステップＳ３５４で取り出したカメラ画像データとを整合し、レンダリングされた頭部モデルに最も良く整合するカメラ画像データを見出す。
【０２２２】
ステップＳ３５８では、頭部追跡装置６５０は、ステップＳ３５６でレンダリングされた

頭部モデルに最も良く整合すると識別されたカメラ画像データを使用して、映像データの現在フレームについて参加者の頭部の３D位置及び３D向きを判定する。

【０２２３】

ステップＳ３５８を実行するのと同時に、ステップＳ３６０では、ステップＳ３５６で判定されたカメラ画像データにおける頭部の特徴の位置を従来のカルマンフィルタに入力して、映像データの次のフレームについて参加者の頭部の推定３D位置及び推定３D向きを生成する。ビデオカメラ６０２から映像データのフレームが受信されている間、その参加者についてステップＳ３５０からS３６０を繰り加し実行する。

【０２２４】

再び図２３に戻ると、ステップＳ３３６では、中央制御装置６３６は会議に他の参加者がいるか否かを判定し、参加者ごとに先に説明したように処理が実行され終わるまでステップＳ３２６からS３３６を繰り返す。しかし、参加者ごとにこれらのステップを実行している間、ステップＳ３３４では、頭部追跡装置６５０は既に着席した各参加者の頭部を追跡し続けている。

【０２２５】

ステップＳ３３６で、会議にそれ以上の参加者はなく、従って、各参加者の頭部が頭部追跡装置６５０により追跡されていることが判定されると、ステップＳ３３８で、中央制御装置６３６は、参加者間で会議を始めて良いことを指示するために、処理装置６２４から可聴信号を出力させる。

【０２２６】

図２６は、参加者間で会議が行われている間に処理装置６２４により実行される処理動作を示す。

【０２２７】

図２６を参照すると、ステップＳ３７０では、頭部追跡装置６５０は会議中の各参加者の頭部を追跡し続ける。ステップＳ３７０で頭部追跡装置６５０により実行される処理は、先にステップＳ３３４に関して説明した処理と同じであるので、ここでは説明を省略する。

【０２２８】

頭部追跡装置６５０がステップＳ３７０で各参加者の頭部を追跡しているのと同時に、ステップＳ３７２では、データを生成し、会議アーカイブデータベース６６０にデータを格納するための処理を実行する。

【０２２９】

図２７は、ステップＳ３７２で実行される処理動作を示す。

【０２３０】

図２７を参照すると、ステップＳ３８０では、アーカイブプロセッサ６５８は、参加者が誰を見ているかを定義するいわゆる「視線パラメータ」を参加者ごとに生成する。

【０２３１】

図２８は、ステップＳ３８０で実行される処理動作を示す。

【０２３２】

図２８を参照すると、ステップＳ４１０では、アーカイブプロセッサ６５８は各参加者の頭部の現在三次元位置を頭部追跡装置６５０から読み取る。これは、ステップＳ３５８（図２５）で頭部追跡装置６５０により実行される処理において生成された位置である。

【０２３３】

ステップＳ４１２では、アーカイブプロセッサ６５８は次の参加者（初めてステップＳ４１２を実行する場合には、これは最初の参加者である）の頭部の現在向きを頭部追跡装置６５０から読み取る。ステップＳ４１２で読み取られる向きは、ステップＳ３５８（図２５）で頭部追跡装置６５０により実行される処理において生成された向きである。

【０２３４】

ステップＳ４１４では、アーカイブプロセッサ６５８は、参加者がどこを見ているかを定義する線（いわゆる「視線」）と、参加者の頭部を別の参加者の頭部の中心と結ぶ概念上

の各々の線とが成す角度を判定する。

【０２３５】

図２９及び図３０を参照して更に詳細に説明する。図２９及び図３０には、１人の参加者、すなわち、図２１の参加者６１０についてステップＳ４１４で実行される処理の一例が示されている。図２９を参照すると、ステップＳ４１２で読み取られる参加者の頭部の向きは、その参加者の両目の中心の間の一点から出る、参加者の頭部に対し垂直な視線６９０を定義する。同様に、図３０を参照すると、ステップＳ４１０で読み取られる全ての参加者の頭部の位置は、参加者６１０の両目の中心の間の一点から他の各々の参加者６０６、６０８、６１２の頭部の中心に至る概念上の線６９２、６９４、６９６を定義する。ステップＳ４１４では、アーカイブプロセッサ６５８は視線６９０と、概念上の線６９２、６９４、６９６とがそれぞれ成す角度６９８、７００、７０２を判定する。

【０２３６】

再び図２８に戻ると、ステップＳ４１６では、アーカイブプロセッサ６５８は最小値を有する角度６９８、７００又は７０２を選択する。すなわち、図３０に示す例で言えば、角度７００が選択されることになるであろう。

【０２３７】

ステップＳ４１８では、アーカイブプロセッサ６５８は選択した角度が１０°より小さいか否かを判定する。

【０２３８】

ステップＳ４１８で、角度が１０°より小さいと判定されれば、アーカイブプロセッサ６５８は参加者の視線パラメータを、視線と最小の角度を成す概念上の線により結ばれている参加者の番号（図２３のステップＳ３０４で割り当てられている）に設定する。すなわち、図３０に示す例で言えば、角度７００が１０°より小さい場合には、角度７００は視線６９０と、参加者６１０を参加者６０６と結ぶ概念上の線６９４とが成す角度であるので、視線パラメータは参加者６０６の参加者番号に設定されることになるであろう。

【０２３９】

他方、ステップＳ４１８で、最小角度が１０°以上であると判定された場合には、ステップＳ４２２で、アーカイブプロセッサ６５８は参加者の視線パラメータを「０」に設定する。これは、視線６９０が概念上の線６９２、６９４、６９６のいずれにも十分に近接していないために、参加者はその他の参加者の誰も見ていないと判定されたことを示す。そのような状況は、例えば、参加者がメモ又は会議室内の他の何らかの物体を見ているときに起こりうるであろう。

【０２４０】

ステップＳ４２４では、アーカイブプロセッサ６５８は会議に他の参加者がいるか否かを判定し、参加者ごとに上記の処理がそれぞれ実行され終わるまでステップＳ４１２からＳ４２４を繰り返す。

【０２４１】

再び図２７に戻ると、ステップＳ３８２では、中央制御装置６３６及び音声認識プロセッサ６５４は、映像データの現在フレームについてマイクロホン６０４から音声データが受信されたか否かを判定する。

【０２４２】

ステップＳ３８２で、音声データが受信されていると判定されれば、ステップＳ３８４で、アーカイブプロセッサ６５８はステップＳ３８０で生成された視線パラメータを処理して、会議中のどの参加者が話をしているかを判定する。

【０２４３】

図３１は、ステップＳ３８４でアーカイブプロセッサ６５８により実行される処理動作を示す。

【０２４４】

図３１を参照すると、ステップＳ４４０では、ステップＳ３８０で生成された各視線パラメータ値の出現回数を判定し、ステップＳ４４２では、出現回数が最も多い視線パラメー

タ値を選択する。すなわち、図２７のステップＳ３８０で実行される処理は、会議中の参加者ごとに、映像データの現在フレームについて１つの視線パラメータ値を生成するのである（従って、図２１に示す例では、４つの値が生成されることになるであろう）。各視線パラメータは、その他の参加者のうち１人の参加者番号に相当する値又は「０」を有する。従って、ステップＳ４４０及びS４４２では、アーカイブプロセッサ６５８は、ステップＳ３８０で生成された視線パラメータ値の中で、映像データの現在フレームについて最も多くの回数で出現する値はどれであるかを判定する。

【０２４５】

ステップＳ４４４では、最も出現回数の多い視線パラメータが「０」の値を有するか否かを判定し、「０」の値であれば、ステップＳ４４６で、次に出現回数の多い視線パラメータ値を選択する。これに対し、ステップＳ４４４で、選択された値が「０」ではないと判定された場合には、ステップＳ４４６を省略する。

【０２４６】

ステップＳ４４８では、選択された視線パラメータ値（すなわち、ステップＳ４４２で選択された値、又はその値が「０」であれば、ステップＳ４４６で選択された値）を話をしている参加者として識別する。これは、会議中の参加者の大半は話者である参加者を見ているからである。

【０２４７】

再び図２７に戻ると、ステップＳ３８６では、アーカイブプロセッサ６５８は話者である参加者の視線パラメータ値、すなわち、ステップＳ３８０で生成された、話者である参加者が誰を見ているかを定義する視線パラメータ値を後の解析に備えて、例えば、メモリ６４２に格納する。

【０２４８】

ステップＳ３８８では、アーカイブプロセッサ６５８はステップＳ３８４で判定された話者である参加者のアイデンティティを音声認識プロセッサ６５４に報知する。これに応答して、音声認識プロセッサ６５４は話者である参加者の音声認識パラメータを音声認識パラメータ記憶装置６５６から選択し、選択したパラメータを使用して、受信された音声データに対して音声認識処理を実行し、話者である参加者が話した言葉に対応するテキストデータを生成する。

【０２４９】

他方、ステップＳ３８２で、受信された音声データが話し言葉を含まないと判定された場合には、ステップＳ３８４からS３８８を省略する。

【０２５０】

ステップＳ３９０では、アーカイブプロセッサ６５８はカメラ６０２から受信された映像データの現在フレームと、マイクロホン６０４から受信された音声データとを従来の方式でMPEG２データを符号化し、符号化されたデータを会議アーカイブデータベース６６０に格納する。

【０２５１】

図３２は、会議アーカイブデータベース６６０へのデータの格納状態を概略的に示す。図３２に示す格納構造は概念的なものであり、単に理解を助けるために例示を目的として提示されているにすぎない。従って、図３２に示す構造は、データが実際に会議アーカイブデータベース６６０に格納される厳密な状態を必ずしも表してはいない。

【０２５２】

図３２を参照すると、会議アーカイブデータベース６６０は水平軸８００により表される時間情報を格納している。水平軸８００に沿った各単位は所定の量の時間、例えば、カメラ６０２から受信される映像データの１つのフレームを表している。ステップＳ３９０で生成されるMPEG２データは、時間情報と共に、データ８０２として会議アーカイブデータベース６６０に格納されている（この時間情報は、図３２には、水平軸８００に沿ったMPEG２データ８０２の位置により概略的に表されている）。

【０２５３】

再び図２７に戻ると、ステップＳ３９２では、アーカイブプロセッサ６５８は、現在フレームについてステップＳ３８８で音声認識プロセッサ６５４により生成されたテキストデータを会議アーカイブデータベース６６０に格納する（図３２には８０４で示されている）。すなわち、テキストデータは対応するMPEG２データへのリンクを伴って格納される。このリンクは、図３２には、テキストデータがMPEG２データと同じ縦列に格納されることによって表されている。話をしていない参加者からは格納すべきテキストデータが得られないことは理解されるであろう。図３２に示す例では、初めの１０個のタイムスロットにわたり参加者１についてテキストが格納され（８０６で示す）、１２番目から２０番目のタイムスロットには参加者３のテキストが格納され（８０８で示す）、２１番目のタイムスロットには参加者４のテキストが格納されている（８１０で示す）。この例では、図３２に示すタイムスロットの間、参加者２は話をしなかったので、参加者２のテキストは格納されていない。

【０２５４】

ステップＳ３９４では、アーカイブプロセッサ６５８は、ステップＳ３８０で生成された参加者ごとの視線パラメータ値を会議アーカイブデータベース６６０に格納する（図３２には８１２で示す）。図３２を参照すると、視線パラメータ値は、参加者ごとに、関連するMPEG２データ８０２及び関連するテキストデータ８０４へのリンクと共に格納されている（このリンクは、図３２では、視線パラメータ値が関連するMPEG２データ８０２及び関連するテキストデータ８０４と同じ縦列に格納されることにより示されている）。従って、一例として第１のタイムスロットに関して言えば、参加者１の視線パラメータ値は、参加者１が参加者３を見ていることを指示する「３」であり、参加者２の視線パラメータ値は、参加者２が参加者１を見ていることを指示する「１」であり、参加者３の視線パラメータ値も、参加者３が同様に参加者１を見ていることを指示する「１」であり、参加者４の視線パラメータ値は、参加者４が他のどの参加者も見ていない（図２１に示す例では、６１２で示される参加者は他の参加者ではなく、自分のメモを見ている）ことを指示する「０」である。

【０２５５】

ステップＳ３９６では、中央制御装置６３６及びアーカイブプロセッサ６５８は、会議中の参加者の１人が話を止めたか否かを判定する。この実施形態においては、この検査は、所定の参加者のテキストデータが直前のタイムスロットには存在したが、現在タイムスロットには存在しないことを判定するためにテキストデータ８０４を検査することにより実行される。ある参加者についてこの条件が満たされれば（すなわち、参加者が話を止めたならば）、ステップＳ３９８で、アーカイブプロセッサ６５８は、話を止めた参加者について、先にステップＳ３８６を実行したときに格納されていた視線パラメータ値を処理して（それらの視線パラメータ値は、その時点で止まった話をしていた期間中にその参加者が誰を見ていたかを定義する）、視線ヒストグラムを定義するデータを生成する。すなわち、参加者が話をしていた期間の視線パラメータ値を処理して、その期間中に話者である参加者がその他の参加者の各々を見ていた時間の割合（％）を定義するデータを生成するのである。

【０２５６】

図３３A及び図３３Bは図３２のテキスと８０６及び８０８の期間にそれぞれ対応する視線ヒストグラムを示す。

【０２５７】

図３２及び図３３Aを参照して説明すると、参加者１が話していた期間８０６の間、図３３Aに９００で示すように、参加者１は１０個のタイムスロットのうち６個（すなわち、参加者１が話していた期間全体の長さの６０％）にわたり参加者３を見ており、また、図３３Aに９１０で示すように、４個のタイムスロット（すなわち、時間の４０％）にわたり参加者４を見ていた。

【０２５８】

同様に、図３２及び図３３Bを参照すると、期間８０８の間、図３３Bに９２０で示すよう

に、参加者３は時間の約４５％にわたり参加者１を見ており、図３３Bに９３０で示すように、時間の約３３％にわたり参加者４を見ており、図３３Bに９４０で示すように、時間の約２２％にわたり参加者２を見ていた。

【０２５９】

再び図２７に戻ると、ステップＳ４００では、ステップＳ３９８で生成された視線ヒストグラムをそれが生成された関連するテキストの期間にリンクさせて、会議アーカイブデータベース６６０に格納する。図３２を参照すると、格納された視線ヒストグラムは８１４で示されており、８１６で示されるテキスト期間８０６に対応するヒストグラムを定義するデータと、８１８で示されるテキスト期間８０８に対応するヒストグラムを定義するデータとを伴う。図３２において、視線ヒストグラムと関連するテキストとの間のリンクは、視線ヒストグラムがテキストデータと同じ縦列に格納されることにより示されている。

【０２６０】

他方、ステップＳ３９６で、現在時限について、参加者の１人が話を止めていないと判定された場合には、ステップＳ３９８及びS４００を省略する。

【０２６１】

ステップＳ４０２では、中央制御装置６３６は、カメラ６０２から映像データの別のフレームが受信されたか否かを判定する。カメラ６０２から画像データが受信されている間は、ステップＳ３８０からS４０２を繰り返し実行する。

【０２６２】

会議アーカイブデータベース６６０にデータが格納されている場合、会議に関連するデータを検索するために会議アーカイブデータベース６６０を問い合わせても良い。

【０２６３】

図３４は、ユーザにより指定される探索基準を満たす会議の各部分に関連するデータを検索する目的で会議アーカイブデータベース６６０を探索するために実行される処理動作を示す。

【０２６４】

図３４を参照すると、ステップＳ５００では、中央制御装置６３６は表示プロセッサ６６４に、要求される会議アーカイブデータベース６６０の探索を定義する情報をユーザが入力することを要求するメッセージを表示装置６２６に表示させる。すなわち、この実施形態においては、中央制御装置６３６は図３５Aに示す表示を表示装置６２６に出現させる。

【０２６５】

図３５Aを参照すると、ユーザは、会議アーカイブデータベース６６０の中で見出したい会議の１つ又は複数の部分を定義する情報を入力することを求められる。すなわち、この実施形態においては、ユーザは話をしていた参加者を定義する情報１０００と、情報１０００の中で識別される参加者が話した１つ又は複数のキーワードから成る情報１０１０と、情報１０００の中で識別される参加者が話しかけていた参加者を定義する情報１０２０とを入力することを求められる。更に、ユーザは、探索を実行すべき会議の１つ又は複数の部分を定義する時間情報を入力することができる。すなわち、ユーザは、その時間を越えたときに探索を中断すべき会議中の時間（すなわち、指定される時間以前の会議の期間を探索すべきである）を定義する情報１０３０と、その時間の後に探索を実行すべきである会議中の時間を定義する情報１０４０と、探索を実行すべき期間の開始時間と終了時間をそれぞれ定義する情報１０５０及び１０６０とを入力することができる。この実施形態では、情報１０３０、１０４０、１０５０及び１０６０は、例えば、分単位などの絶対期限で時間を指定するか、又は会議時間全体に占める割合を指示する小数値を入力することにより相対期限で時間を指定するかのいずれかにより入力されれば良い。例えば、情報１０３０として０．２５の値を入力すると、探索は会議の初めの四分の一に限られるであろう。

【０２６６】

この実施形態では、ユーザは１回の探索に際して全ての情報１０００、１０１０及び１０

２０を入力することを求められるわけではなく、そのうち１つ又は２つの情報を省いても良い。ユーザが情報１０００、１０１０及び１０２０の全てを入力すれば、会議中に情報１０００の中で識別される参加者が情報１０２０の中で識別される参加者に話しかけていた各部分及び情報１０１０の中で定義されるキーワードを話していた各部分を識別するための探索が実行される。これに対し、情報１０１０を省いた場合には、会議中に情報１０００の中で識別される参加者が、何を言ったかに関わらず、情報１０２０の中で定義される参加者に話しかけていた各部分を識別するための探索が実行されることになる。情報１０１０及び１０２０を省いた場合には、会議中に情報１０００の中で定義される参加者が何を誰に向かって話したかに関わらず、話をしていた各部分を識別するための探索が実行される。情報１０００を省いた場合には、会議中にいずれかの参加者が情報１０１０の中で定義されるキーワードを情報１０２０の中で定義される参加者に向かって話した各部分を識別するための探索が実行される。情報１０００及び１０１０を省いた場合には、会議中にいずれかの参加者が情報１０２０の中で定義される参加者に話しかけた各部分を識別するための探索が実行される。情報１０２０を省いた場合には、会議中に情報１０００の中で定義される参加者が、誰に向かって話したかに関わらず、情報１０１０の中で定義されるキーワードを話した各部分を識別するための探索が実行される。同様に、情報１０００及び１０２０を省いた場合には、会議中に、誰が誰に向かって話したかに関わらず、情報１０１０の中で識別されるキーワードが話された各部分を識別するための探索が実行される。

【０２６７】

更に、ユーザは時間情報１０３０、１０４０、１０５０及び１０６０の全てを入力しても良いし、そのうち１つ又は複数の情報を省いても良い。

【０２６８】

探索を定義するために必要な情報を全て入力したならば、ユーザは、マウス６３０などのユーザ入力装置を使用して領域１０７０をクリックすることにより探索を開始する。

【０２６９】

再び図３４に戻ると、ステップＳ８０２では、中央制御装置６３６はユーザにより入力された探索情報を読み取り、命令された探索を実行する。すなわち、この実施形態においては、中央制御装置６３６は情報１０００又は１０２０の中で入力された参加者の名前をテーブル６８０（図２４）を使用して参加者番号に変換し、情報１０００の中で定義される参加者（情報１０００が入力されていない場合には全ての参加者）についてテキスト情報８０４を考慮する。ユーザにより情報１０２０が入力されていれば、テキストの期間ごとに、中央制御装置６３６は対応する視線ヒストグラムを定義するデータを検査して、情報１０２０の中で定義される参加者のヒストグラムにおける注目時間の割合がこの実施形態では２５％である閾値以上であるか否かを判定する。このようにして、話し言葉（テキスト）の期間を考慮して、話者である参加者が話している時間の少なくとも２５％にわたって情報１０２０の中で定義される参加者を見ていたならば、情報１０００の中で定義される参加者が話をしている間に他の参加者を見たとしても、情報１０２０の中で定義される参加者に話しかけていたという基準を満たす。従って、情報１０２０の中で２人以上の参加者が指定されていれば、視線ヒストグラムの値が２人以上の参加者について２５％以上であるような話の期間が識別されるであろう。ユーザにより情報１０１０が入力されていた場合、中央制御装置６３６及びテキストサーチャ６６２は、先に情報１０００及び１０２０に基づいて識別されたテキストの各部分（情報１０００及び１０２０が入力されていなければテキストの全ての部分）を探索して、情報１０１０の中で識別されるキーワードを含む各部分を識別する。ユーザにより時間情報が入力されていた場合、上記の探索はそれらの期限により定義される会議の時間に限られる。

【０２７０】

ステップＳ５０４では、中央制御装置６３６は表示プロセッサ６６４に、探索中に識別された関連話題のリストを表示装置６２６を介してユーザに対し表示させる。すなわち、中央制御装置６３６は図３５Bに示すような情報をユーザに対し表示させる。図３５Bを参照

すると、探索パラメータを満足させるそれぞれの話題のリストが作成されており、その話題について開始時間を絶対期限と、会議時間全体に占める割合の双方で定義する情報が表示される。そこで、ユーザは、マウス６３０を使用してリスト中の必要な話題をクリックすることにより、話題の１つを選択し、再生することができる。

【０２７１】

ステップＳ５０６では、中央制御装置６３６はステップＳ５０４でユーザにより実行された選択を読み取り、格納されている会議の関連部分のMPEG２データ８０２を会議アーカイブデータベース６６０から再生させる。すなわち、中央制御装置６３６及び表示プロセッサ６６４はMPEG２データ８０２を復号し、画像データと音声を表示装置６２６を介して出力する。

【０２７２】

ステップＳ５０８では、中央制御装置６３６は、ユーザが会議アーカイブデータベース６６０の問い合わせを中止することを望むか否かを判定し、望まないのであれば、ステップＳ５００からS５０８を繰り返す。

【０２７３】

以上説明した本発明の実施形態に対し、様々な変形や変更を実施することができる。

【０２７４】

例えば、上記の実施形態では、マイクロホン６０４は会議室のテーブルの上に設けられていた。しかし、その代わりに、ビデオカメラ６０２のマイクロホンを使用して音声データを記録しても良い。

【０２７５】

上記の実施形態では、１台のビデオカメラ６０２からの画像データを処理していた。しかし、各参加者の頭部の追跡精度を向上させるために、複数台のビデオカメラからの映像データを処理しても良い。例えば、複数台のカメラからの画像データを図２５のステップＳ３５０からS３５６におけるように処理し、全てのカメラから得られたデータをステップＳ３６０で従来のようにカルマンフィルタに入力して、カメラごとに映像データの次のフレームにおける各参加者の頭部の位置と向きを更に正確に推定しても良い。複数台のカメラを使用する場合、会議アーカイブデータベース６６０に格納されるMPEG２データ８０２は全てのカメラからの映像データということになり、図３４のステップＳ５０４及びS５０６では、ユーザが選択した１台のカメラからの画像データが再生される。

【０２７６】

上記の実施形態では、所定の参加者の視線パラメータは、その参加者が他のどの参加者を見ているかを定義していた。しかし、参加者が表示板、映写機のスクリーンなどのどの物体を見ているかを定義するために視線パラメータを使用しても良い。この場合、会議アーカイブデータベース６６０を問い合わせるときに、図３５Aの情報１０２０を利用して、参加者が話している間に誰を又は何を見ていたかを指定することが可能になるであろう。

【０２７７】

上記の実施形態では、ステップＳ５０２（図３４）で、テキストの特定の部分の視線ヒストグラムを考慮し、視線ヒストグラムにおける別の参加者への注目時間の割合が所定の閾値以上である場合に、参加者はその別の参加者に話しかけていたと判定していた。しかし、閾値を使用せずに、テキストの期間中に話者である参加者が見ていた参加者を視線ヒストグラムにおいて最も大きな割合の注目値を有する参加者（例えば、図３３Aの参加者３及び図３３Bの参加者１）であると定義しても良い。

【０２７８】

上記の実施形態では、カメラ６０２及びマイクロホン６０４からデータが受信されている間、MPEG２データ８０２、テキストデータ８０４、視線パラメータ８１２及び視線ヒストグラム８１４をリアルタイムで会議アーカイブデータベース６６０に格納していた。しかし、映像データと音声データを格納しておき、データ８０２、８０４、８１２及び８１４をリアルタイムではなく生成し、会議アーカイブデータベース６６０に格納しても良い。

【０２７９】

上記の実施形態では、会議の定義された部分のデータを検索するために会議アーカイブデータベース６６０を問い合わせる前に、MPEG２データ８０２、テキストデータ８０４、視線パラメータ８１２及び視線ヒストグラム８１４を生成し、会議アーカイブデータベース６６０に格納していた。しかし、探索の要求に先立ってデータを生成、格納するのではなく、会議アーカイブデータベース６６０の探索がユーザにより要求されるのに応答して、格納されているMPEG２データ８０２を処理することによりデータ８０４、８１２及び８１４の一部又は全てを生成しても良い。例えば、上記の実施形態では、ステップＳ３９８及びS４００（図２７）で視線ヒストグラム８１４をリアルタイムで計算、格納していたが、ユーザにより探索要求が入力されるのに応答してそれらのヒストグラムを計算することもできるであろう。

【０２８０】

上記の実施形態では、テキストデータ８０４を会議アーカイブデータベース６６０に格納していた。テキストデータ８０４の代わりに、音声データを会議アーカイブデータベース６６０に格納しても良い。その場合、格納されている音声データ自体を音声認識処理を使用してキーワードを求めて探索するか、又は音声認識処理を使用して音声データをテキストに変換し、従来のテキストサーチャを使用してテキストを探索すれば良い。

【０２８１】

上記の実施形態では、処理装置６２４はアーカイブすべきデータを受信し、生成するための機能構成要素（例えば、中央制御装置６３６、頭部追跡装置６５０、頭部モデル記憶装置６５２、音声認識プロセッサ６５４、音声認識パラメータ記憶装置６５６及びアーカイブプロセッサ６５８）と、アーカイブデータを格納するための機能構成要素（例えば、会議アーカイブデータベース６６０）と、データベースを探索し、そこから情報を検索するための機能構成要素（例えば、中央制御装置６３６及びテキストサーチャ６６２）とを含んでいた。しかし、これらの機能構成要素を別個の装置に設けても良い。例えば、アーカイブすべきデータを生成する１つ又は複数の装置と、データベース探索のための１つ又は複数の装置とを、インターネットなどのネットワークを介して１つ又は複数のデータベースに接続しても良い。

【０２８２】

また、図３６を参照して説明すると、一箇所又は複数箇所での会議１１００、１１１０、１１２０からの映像データと音声データをデータ処理・データベース記憶装置１１３０（アーカイブデータを生成し且つ格納するための機能構成要素を具備する）に入力し、データベースを問い合わせ、そこから情報を検索するために、１つ又は複数のデータベース問い合わせ装置１１４０、１１５０をデータ処理・データベース記憶装置１１３０に接続しても良い。

【０２８３】

上記の実施形態では、プログラミング命令により定義される処理ルーチンを使用して、コンピュータにより処理を実行していた。しかし、処理の一部又は全てをハードウェアを使用して実行しても良い。

【０２８４】

以上、複数の参加者の間で行われる会議に関して実施形態を説明したが、本発明はこの用途には限定されず、フィルムセットなどについて画像データ及び音声データを処理するなど、他の用途にも適用することができる。

【０２８５】

上記の変形例の異なる組み合わせも言うまでもなく可能であり、本発明の趣旨から逸脱せずにその他の変更や変形を実施することができる。

【図面の簡単な説明】

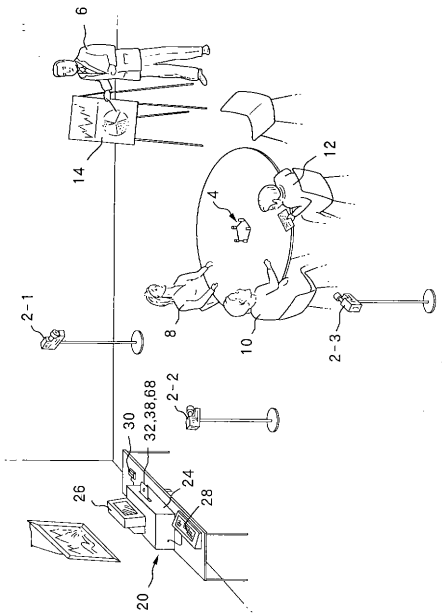
【図１】第１の実施形態における複数の参加者の間の会議からの音声データ及び映像データの記録を示す図である。

【図２】第１の実施形態の処理装置内部の概念上の機能構成要素の一例を示すブロック線図である。

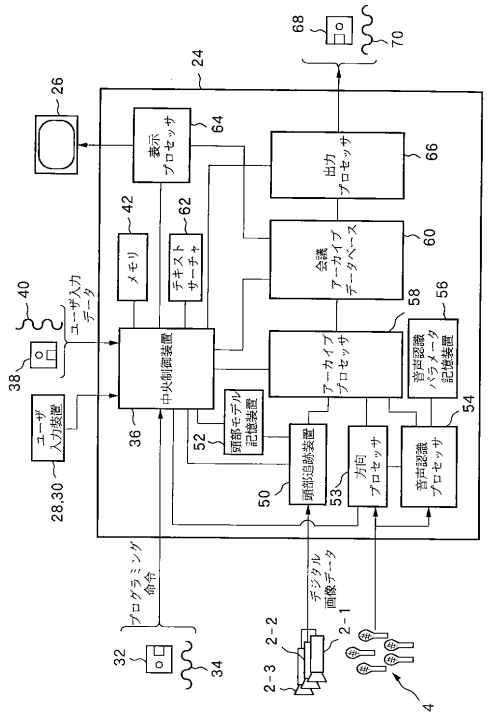
【図３Ａ】図１に示す参加者間の会議が始まる以前に図２の処理装置２４により実行される処理動作を示す図である。
【図３Ｂ】図１に示す参加者間の会議が始まる以前に図２の処理装置２４により実行される処理動作を示す図である。
【図３Ｃ】図１に示す参加者間の会議が始まる以前に図２の処理装置２４により実行される処理動作を示す図である。
【図４】　図３のステップＳ２及びステップＳ４で会議アーカイブデータベース６０に格納されるデータを概略的に示す図である。
【図５】図３のステップＳ３４及び図７のステップＳ７０で実行される処理動作を示す図である。
【図６】図５のステップＳ４２－１、S４２－２及びS４２－nのそれぞれで実行される処理動作を示す図である。
【図７】参加者間で会議が行われている間に図２の処理装置２４により実行される処理動作を示す図である。
【図８Ａ】図７のステップＳ７２で実行される処理動作を示す図である。
【図８Ｂ】図７のステップＳ７２で実行される処理動作を示す図である。
【図９Ａ】図８のステップＳ８０で実行される処理動作を示す図である。
【図９Ｂ】図８のステップＳ８０で実行される処理動作を示す図である。
【図１０】図９のステップＳ１１４及びステップＳ１２４で実行される処理で使用する参加者の視線を示す図である。
【図１１】図９のステップＳ１１４で実行される処理において計算される角度を示す図である。
【図１２】図８のステップＳ８４で実行される処理動作を示す図である。
【図１３】図８のステップＳ８９で実行される処理動作を示す図である。
【図１４】図１３のステップＳ１６８で実行される処理動作を示す図である。
【図１５】会議アーカイブデータベース６０への情報の格納を概略的に示す図である。
【図１６Ａ】会議アーカイブデータベース６０に格納されたデータにより定義される視線ヒストグラムの例を示す図である。
【図１６Ｂ】会議アーカイブデータベース６０に格納されたデータにより定義される視線ヒストグラムの例を示す図である。
【図１７】図８のステップＳ１０２で実行される処理動作を示す図である。
【図１８】会議アーカイブデータベース６０から情報を検索するために処理装置２４により実行される処理動作を示す図である。
【図１９Ａ】図１８のステップＳ２００でユーザに対し表示される情報を示す図である。
【図１９Ｂ】図１８のステップＳ２０４でユーザに対し表示される情報の一例を示す図である。
【図２０】１つのデータベースが複数の会議からの情報を格納し、１つ又は複数の遠隔装置からこのデータベースを問い合わせる第１の実施形態の変形例を概略的に示す図である。
【図２１】第２の実施形態における複数の参加者の間の会議からの音声データ及び映像データの記録を示す図である。
【図２２】第２の実施形態の処理装置内部の概念上の機能構成要素の一例を示すブロック線図である。
【図２３Ａ】図２１に示す参加者間の会議が始まる以前に図２２の処理装置６２４により実行される処理動作を示す図である。
【図２３Ｂ】図２１に示す参加者間の会議が始まる以前に図２２の処理装置６２４により実行される処理動作を示す図である。
【図２３Ｃ】図２１に示す参加者間の会議が始まる以前に図２２の処理装置６２４により実行される処理動作を示す図である。
【図２４】図２３のステップＳ３０４で会議アーカイブデータベース６６０に格納される

データを概略的に示す図である。
【図２５】図２３のステップＳ３３４で実行される処理動作を示す図である。
【図２６】参加者間で会議が行われている間に図２２の処理装置６２４により実行される処理動作を示す図である。
【図２７Ａ】図２６のステップＳ３７２で実行される処理動作を示す図である。
【図２７Ｂ】図２６のステップＳ３７２で実行される処理動作を示す図である。
【図２８】図２７のステップＳ３８０で実行される処理動作を示す図である。
【図２９】図２８のステップＳ４１４で実行される処理で使用する参加者の視線を示す図である。
【図３０】図２８のステップＳ４１４で実行される処理において計算される角度を示す図である。
【図３１】図２７のステップＳ３８４で実行される処理動作を示す図である。
【図３２】会議アーカイブデータベース６６０への情報の格納を概略的に示す図である。
【図３３Ａ】会議アーカイブデータベース６６０に格納されたデータにより定義される視線ヒストグラムの例を示す図である。
【図３３Ｂ】会議アーカイブデータベース６６０に格納されたデータにより定義される視線ヒストグラムの例を示す図である。
【図３４】会議アーカイブデータベース６６０から情報を検索するために処理装置６２４により実行される処理動作を示す図である。
【図３５Ａ】図３４のステップＳ５００でユーザに対し表示される情報を示す図である。
【図３５Ｂ】図３４のステップＳ５０４でユーザに対し表示される情報の一例を示す図である。
【図３６】１つのデータベースが複数の会議からの情報を格納し、１つ又は複数の遠隔装置からこのデータベースを問い合わせる第２の実施形態の変形例を概略的に示す図である。

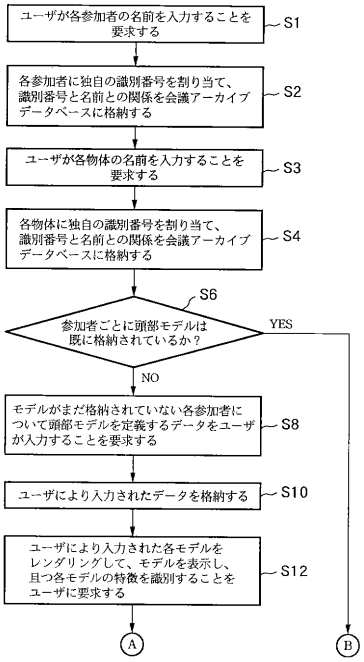
【図１】



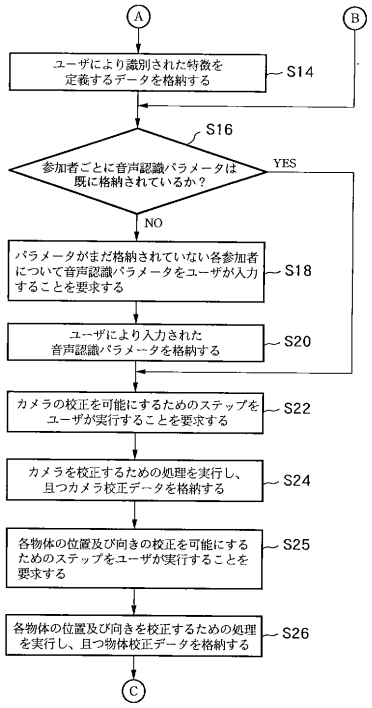
【図２】



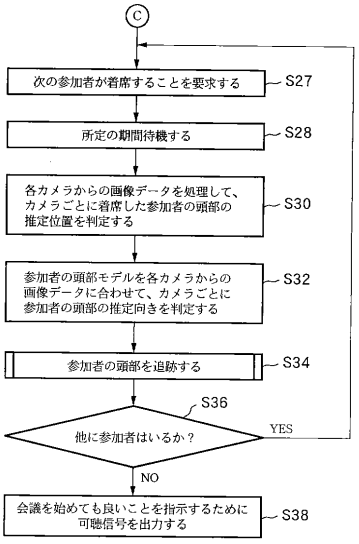
【図３Ａ】



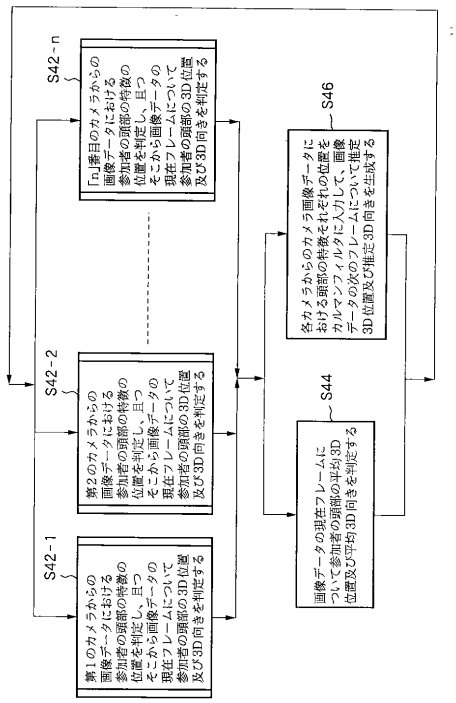
【図３Ｂ】



【図３Ｃ】



【図５】

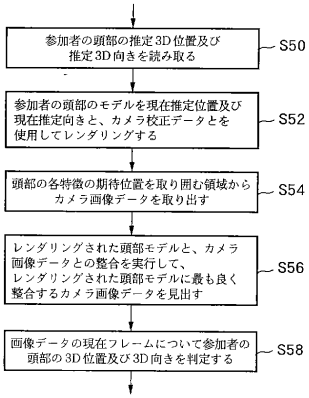


【図４】

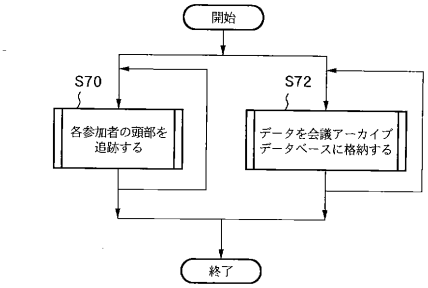
番号	名前
1	A氏
2	Bさん
3	C氏
4	Dさん
5	フリップチャート

80

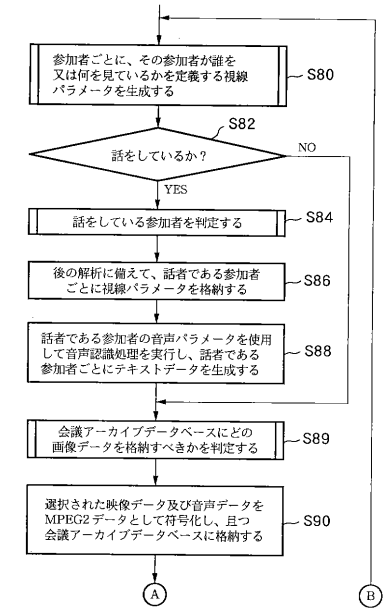
【図６】



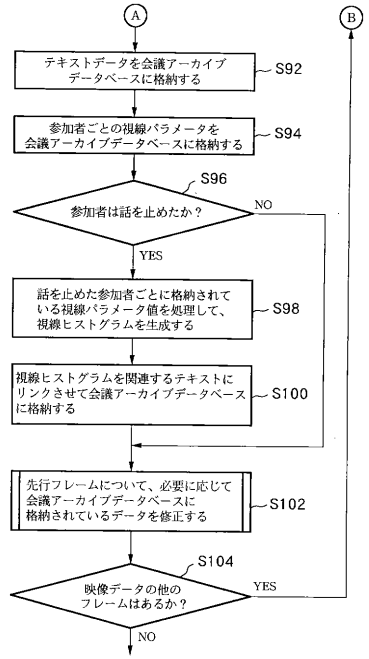
【図７】



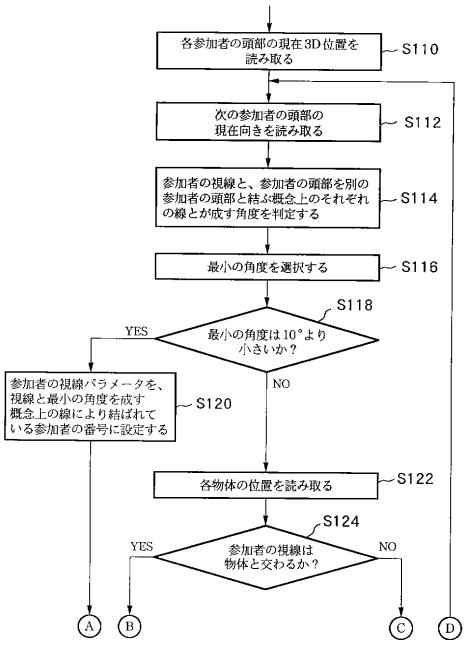
【図８Ａ】



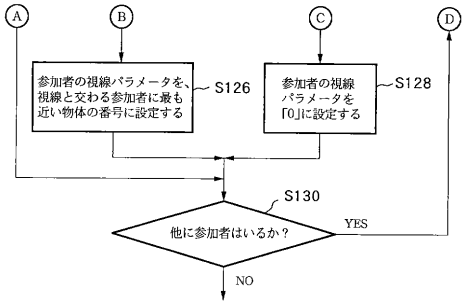
【図８Ｂ】



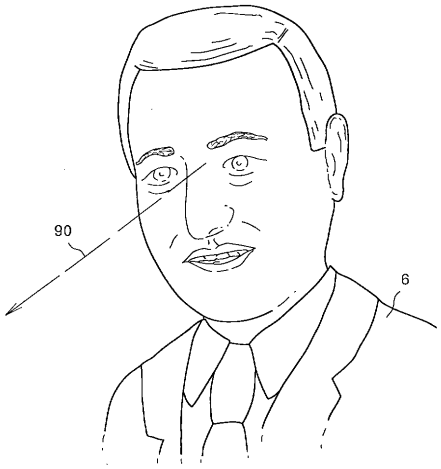
【図９Ａ】



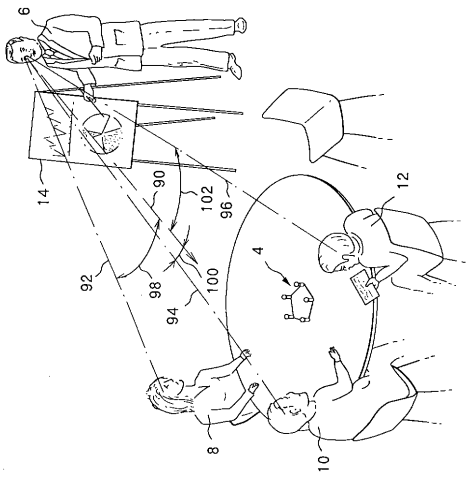
【図９Ｂ】



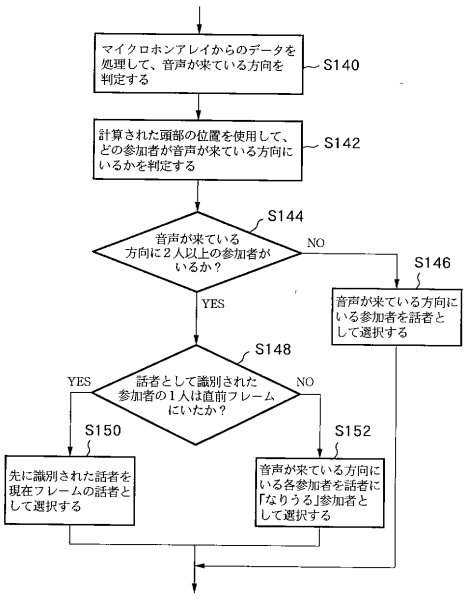
【図１０】



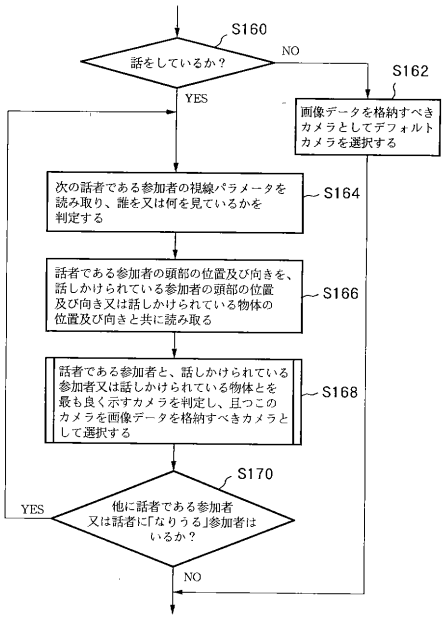
【図１１】



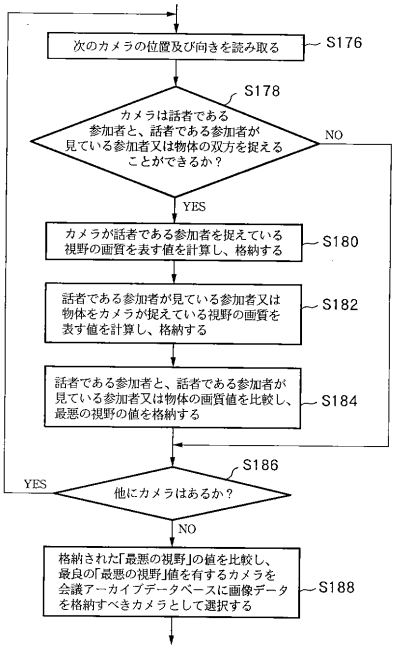
【図１２】



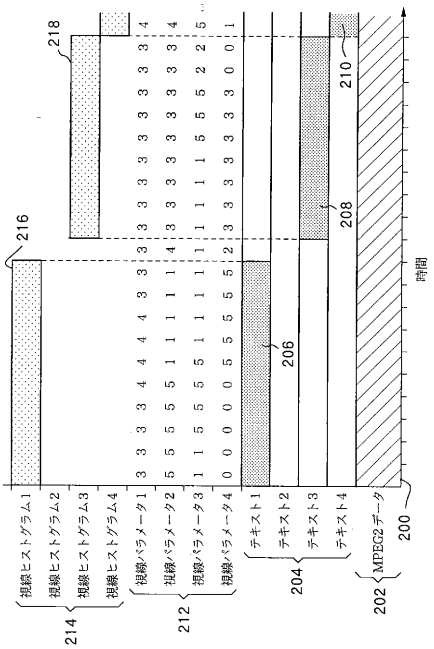
【図１３】



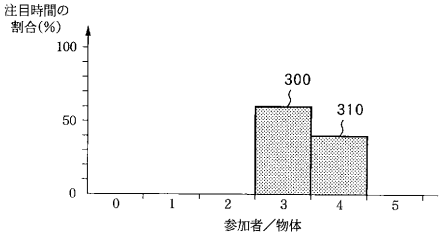
【図１４】



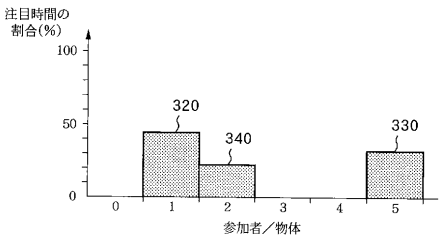
【図１５】



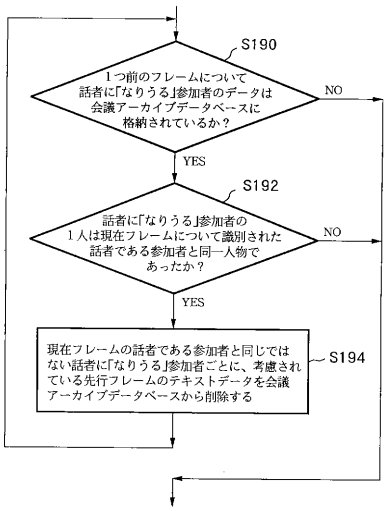
【図１６Ａ】



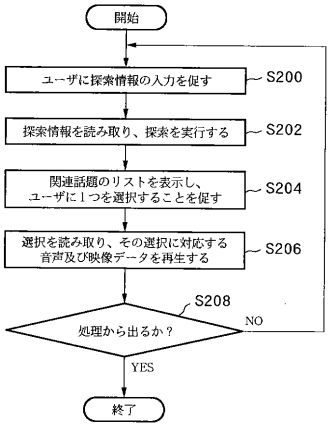
【図１６Ｂ】



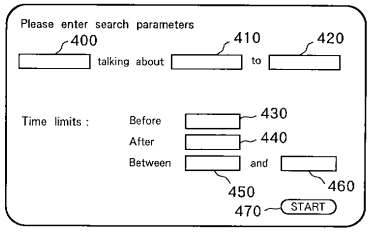
【図１７】



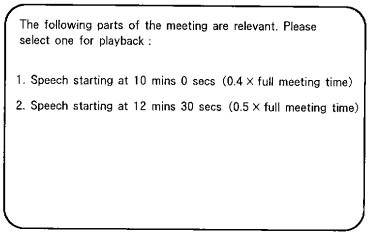
【図１８】



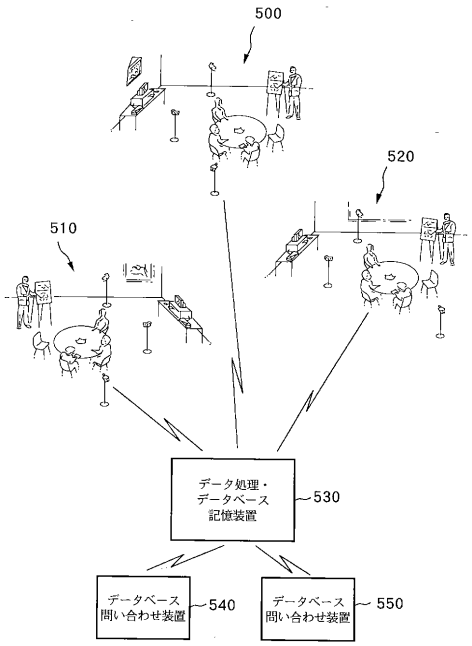
【図１９Ａ】



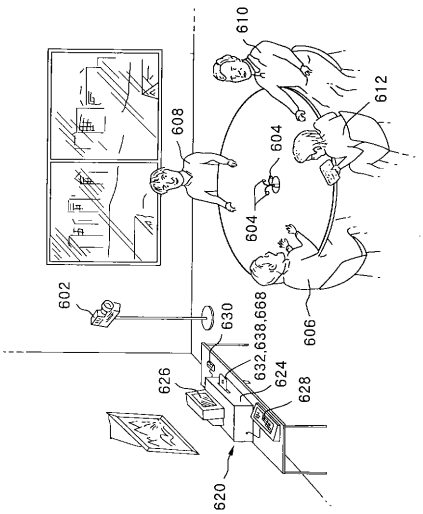
【図１９Ｂ】



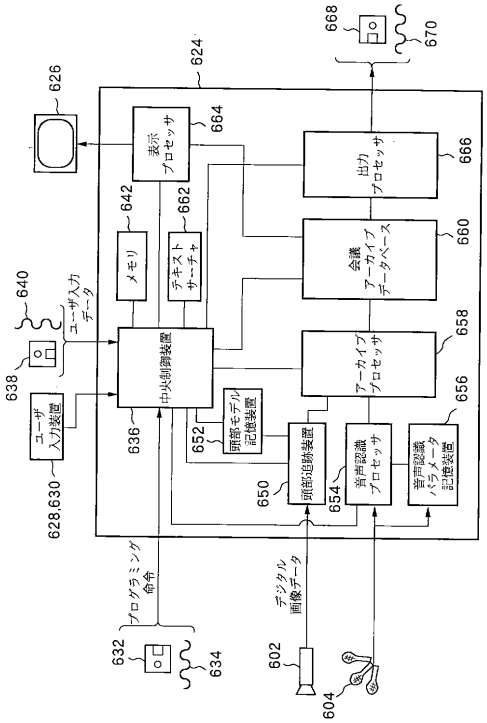
【図２０】



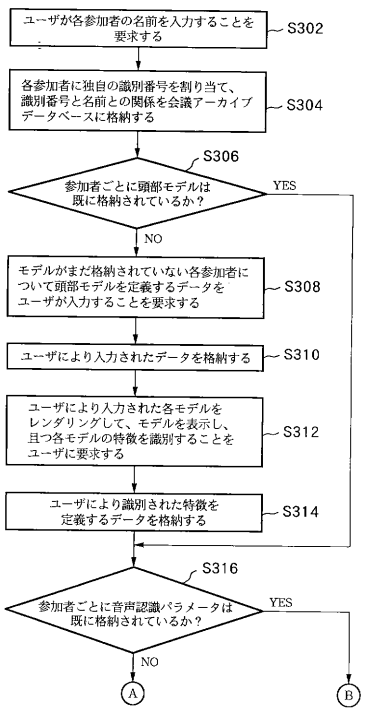
【図２１】



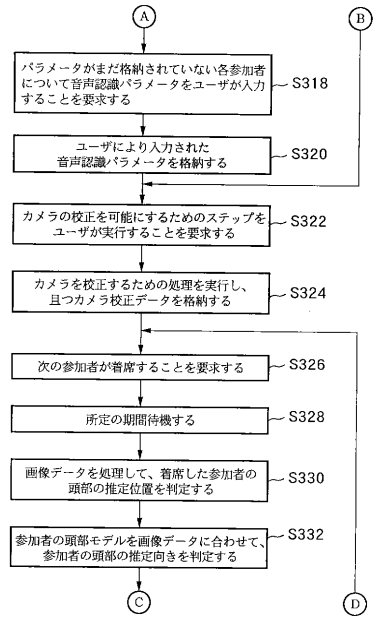
【図２２】



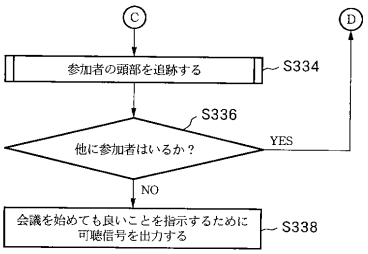
【図２３Ａ】



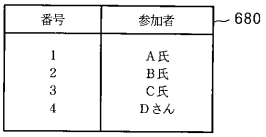
【図２３Ｂ】



【図２３Ｃ】



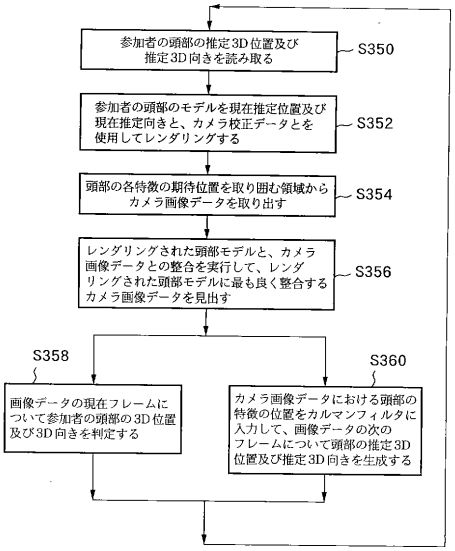
【図２４】



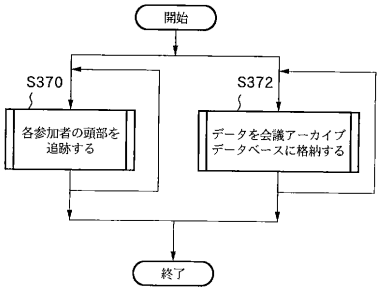
番号	参加者
1	A氏
2	B氏
3	C氏
4	Dさん

680

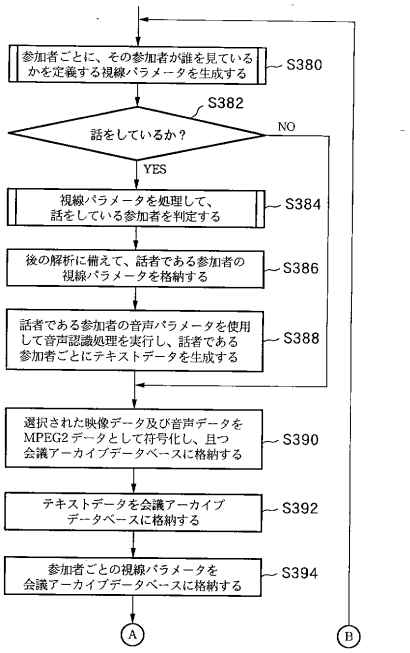
【図２５】



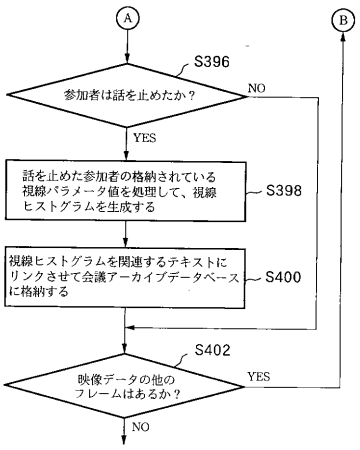
【図２６】



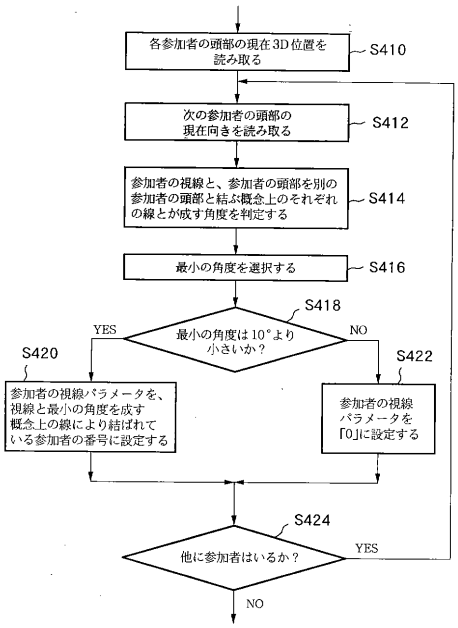
【図２７Ａ】



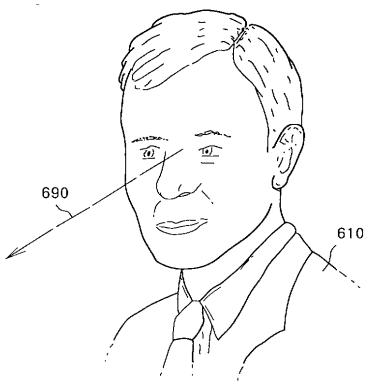
【図２７Ｂ】



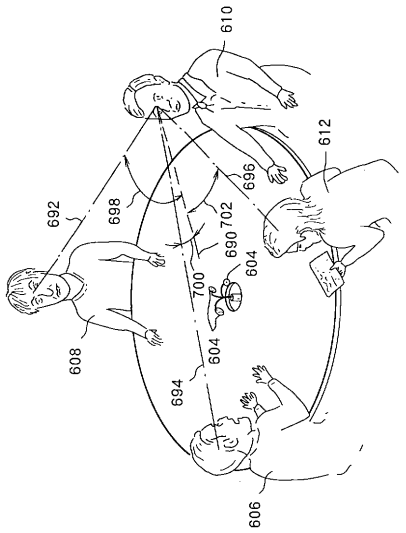
【図２８】



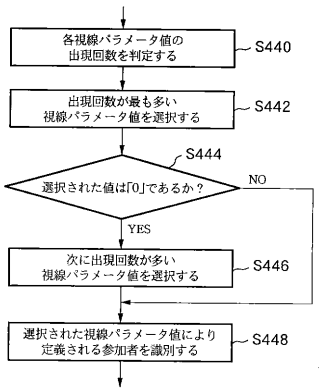
【図２９】



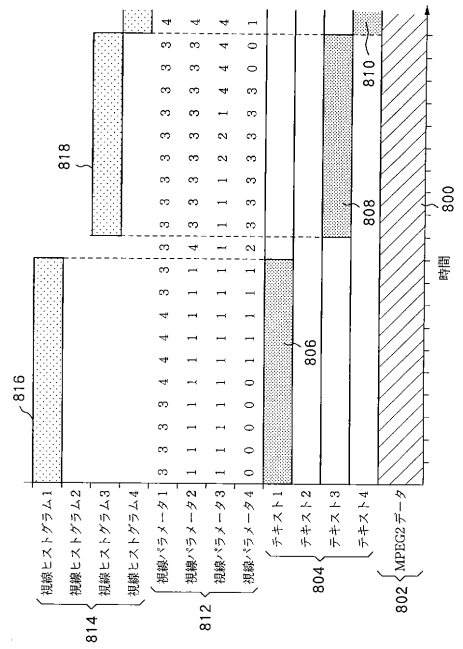
【図３０】



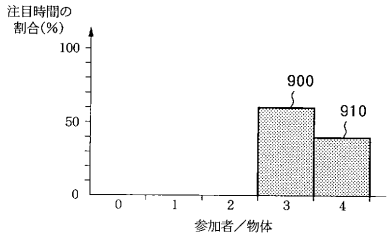
【図３１】



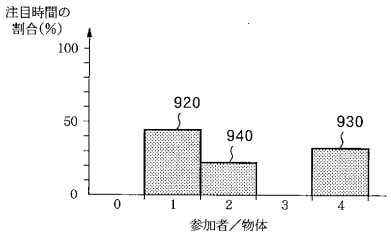
【図３２】



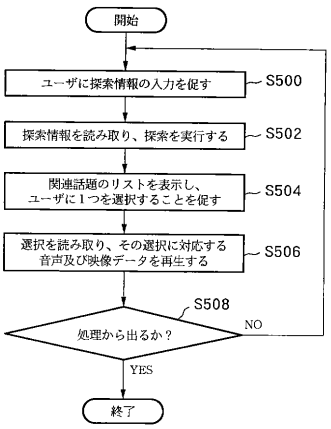
【図３３Ａ】



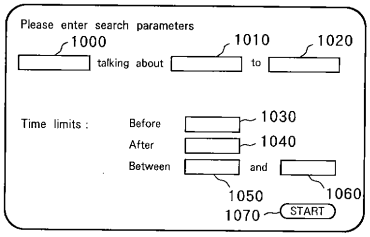
【図３３Ｂ】



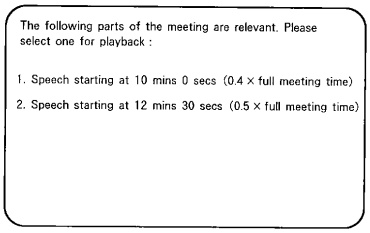
【図３４】



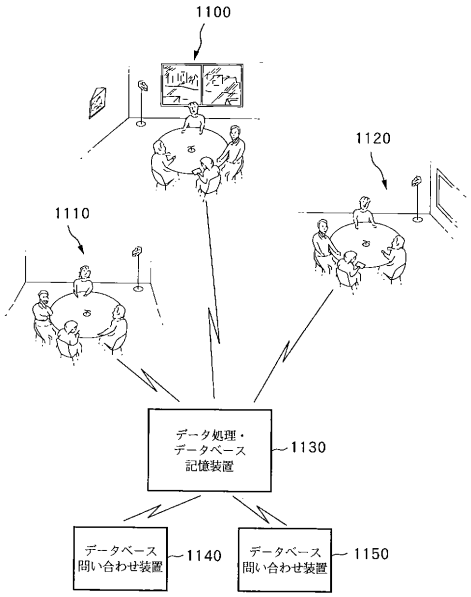
【図３５Ａ】



【図３５Ｂ】



【図３６】



フロントページの続き

(72)発明者　サイモン　マイケル　ロウ
イギリス国　ジーユー２　５ワイジェイ　サリー，　ギルドフォード，　サリー　リサーチ　パーク，　オッカム　ロード，　オッカム　コート　１　キヤノン　リサーチ　センター　ヨーロッパ　リミテッド　内
(72)発明者　マイケル　ジェームス　テイラー
イギリス国　ジーユー２　５ワイジェイ　サリー，　ギルドフォード，　サリー　リサーチ　パーク，　オッカム　ロード，　オッカム　コート　１　キヤノン　リサーチ　センター　ヨーロッパ　リミテッド　内
(72)発明者　ジェブ　ジェイコブ　ラジャン
イギリス国　ジーユー２　５ワイジェイ　サリー，　ギルドフォード，　サリー　リサーチ　パーク，　オッカム　ロード，　オッカム　コート　１　キヤノン　リサーチ　センター　ヨーロッパ　リミテッド　内

審査官　井上　健一

(56)参考文献　特開平１０－１４５７６３（ＪＰ，Ａ）
特開平０２－２０６８２５（ＪＰ，Ａ）
特開平０５－０３５４４１（ＪＰ，Ａ）
特開平０４－０８２３５７（ＪＰ，Ａ）
特開平０３－０２９５５５（ＪＰ，Ａ）
特開平０４－１８１３００（ＪＰ，Ａ）
特開平１１－２５９５０１（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
G10L　15/00-15/28