

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2015-43507  
(P2015-43507A)

(43) 公開日 平成27年3月5日(2015.3.5)

(51) Int.Cl.	F I	テーマコード (参考)
HO4N 7/15 (2006.01)	HO4N 7/15 630Z	5C164
HO4M 3/56 (2006.01)	HO4M 3/56 C	5K201

審査請求 未請求 請求項の数 10 O L (全 22 頁)

(21) 出願番号 特願2013-174578 (P2013-174578)  
(22) 出願日 平成25年8月26日 (2013.8.26)

(71) 出願人 000006747  
株式会社リコー  
東京都大田区中馬込1丁目3番6号  
(74) 代理人 100110607  
弁理士 間山 進也  
(72) 発明者 島中 涉  
東京都大田区中馬込1丁目3番6号 株式会社リコー内  
Fターム(参考) 5C164 FA10 GA07 VA08P VA33S VA36P  
VA52P VA56S  
5K201 AA05 BA09 BB09 BC27 CA06  
CC06 CC10 DC05 EF04 EF10

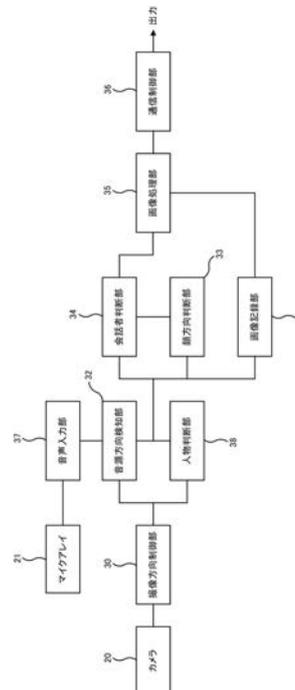
(54) 【発明の名称】 情報処理装置、通信システムおよびプログラム

(57) 【要約】 (修正有)

【課題】 会話の当事者に応じた画像を表示させ、会議の臨場感を高めることができる装置等を提供する。

【解決手段】 カメラ20の撮像方向を変更して、同じ拠点にいる1以上のユーザの画像を記録し、音源方向検知部32により、同じ拠点内のユーザの音源方向を検知し、顔方向判断部33は、その音源方向へ撮像方向を変更して撮像されたユーザの顔の向きを判断し、会話者判断部34はその顔の向きと同じ方向を、音源方向として検知したかどうかに応じて、同じ拠点内の会話か、他拠点との会話かを判断する。画像処理部35は、他拠点との会話と判断された場合、撮像された画像に対して、同じ拠点内の会話と判断された場合、撮像されたユーザの画像、または撮像されたユーザの画像もしくは記録されたユーザの画像と記録された他のユーザの画像とに対して画像処理を行い、他拠点へ送信し、ユーザの画像を、またはユーザの画像と他のユーザの画像とを並べて表示させる。

【選択図】 図3



## 【特許請求の範囲】

## 【請求項 1】

撮像手段と複数の音入力手段とを備える情報処理装置であって、  
前記撮像手段の撮像方向を制御する撮像方向制御手段と、  
前記撮像方向制御手段により撮像方向を変更し、前記撮像手段により撮像された同じ拠点にいる 1 以上のユーザの画像を記録する画像記録手段と、  
前記複数の音入力手段に入力された前記同じ拠点にいるユーザの音声に基づき、音源方向を検知する音源方向検知手段と、  
前記音源方向検知手段により検知された音源方向へ前記撮像方向制御手段によって前記撮像方向を変更し、前記撮像手段により撮像された前記ユーザの画像に基づき、当該ユーザの顔の向きを判断する顔方向判断手段と、  
前記顔方向判断手段により判断された顔の向きと同じ方向を、前記音源方向検知手段が音源方向として検知したかどうかに応じて、前記ユーザと同じ拠点にいる他のユーザとの会話か、他拠点との会話かを判断する会話者判断手段と、  
前記会話者判断手段により他拠点との会話と判断された場合、前記撮像手段により撮像された画像に対して画像処理を行い、前記同じ拠点にいる他のユーザとの会話と判断された場合、前記撮像手段により撮像された前記ユーザと前記他のユーザとを含む画像、または前記撮像手段により撮像された前記ユーザの画像もしくは前記画像記録手段に記録された前記ユーザの画像と前記画像記録手段に記録された前記他のユーザの画像とに対して画像処理を行う画像処理手段と、  
前記画像処理手段により画像処理して得られた画像を前記他拠点へ送信し、前記ユーザの画像を、または前記ユーザの画像と前記他のユーザの画像とを並べて表示させる通信制御手段とを含む、情報処理装置。

10

20

## 【請求項 2】

前記会話者判断手段は、前記同じ拠点にいる他のユーザとの会話と判断した場合、前記他のユーザが複数人であるか否かを、前記複数の音入力手段に入力された音声に基づき判断し、  
前記会話者判断手段により前記他のユーザが一人と判断された場合に、前記撮像手段により撮像された画像に前記ユーザと前記他のユーザとが含まれるとき、前記画像処理手段が前記ユーザの顔を含む所定領域と前記他のユーザの顔を含む所定領域とを切り抜き、切り抜いた 2 つの所定領域の画像を並べて表示するための画像を生成し、前記撮像手段により撮像された画像に前記ユーザのみが含まれるとき、前記撮像手段により撮像された前記ユーザの画像もしくは前記画像記録手段に記録された前記ユーザの画像と前記画像記録手段に記録された前記他のユーザの画像を並べて表示するための画像を生成する、請求項 1 に記載の情報処理装置。

30

## 【請求項 3】

前記会話者判断手段は、前記同じ拠点にいる他のユーザとの会話と判断した場合、前記他のユーザが複数人であるか否かを、前記複数の音入力手段に入力された音声に基づき判断し、  
前記会話者判断手段により前記他のユーザが複数人と判断された場合に、前記複数の音入力手段に対して一定時間内に音声の入力があり、前記撮像手段により撮像された画像に前記ユーザと他の複数のユーザ全員が含まれるとき、前記画像処理手段が前記ユーザの顔を含む所定領域と前記他の複数のユーザの各顔を含む各所定領域を切り抜き、切り抜いた全ての所定領域の画像を並べて表示するための画像を生成し、前記撮像手段により撮像された画像に前記他の複数のユーザの少なくとも一人が含まれないとき、前記撮像手段により撮像された前記ユーザの画像もしくは前記画像記録手段に記録された前記ユーザの画像と前記画像記録手段に記録された前記他の複数のユーザ全員の画像を並べて表示するための画像を生成し、前記一定時間内に音声の入力がないとき、前記撮像手段により撮像された前記ユーザの画像を拡大して表示するための画像を生成する、請求項 1 に記載の情報処理装置。

40

50

**【請求項 4】**

前記会話者判断手段は、前記同じ拠点にいる他のユーザとの会話と判断した場合、前記他のユーザが複数人であるか否かを、前記複数の音入力手段に入力された音声に基づき判断し、

前記会話者判断手段により前記他のユーザが複数人と判断された場合に、前記複数の音入力手段に対して一定時間内に音声の入力があり、前記撮像手段により撮像された画像に前記ユーザと他の複数のユーザ全員が含まれるとき、前記画像処理手段が前記ユーザの顔画像と前記他の複数のユーザの各顔画像を切り抜き、切り抜いた全ての顔画像を並べて表示した画像を生成し、前記撮像手段により撮像された画像に前記他の複数のユーザの少なくとも一人が含まれないとき、前記撮像手段により撮像された前記ユーザの画像もしくは前記画像記録手段に記録された前記ユーザの画像と前記画像記録手段に記録された前記他の複数のユーザ全員の画像とを並べて表示するための画像を生成し、前記一定時間内に音声の入力がないとき、前記撮像手段により撮像された前記ユーザの画像を拡大して表示するための画像を生成する、請求項 1 に記載の情報処理装置。

10

**【請求項 5】**

前記画像記録手段に記録された前記 1 以上のユーザの画像は、当該 1 以上のユーザの静止画であり、検知された音源方向へ前記撮像方向を変更し、前記撮像手段により撮像された画像は、動画である、請求項 1 ~ 4 のいずれか 1 項に記載の情報処理装置。

**【請求項 6】**

地理的に離れた 2 以上の拠点の各々に配置される、請求項 1 ~ 5 のいずれか 1 項に記載の 2 以上の情報処理装置がネットワークに接続された、通信システム。

20

**【請求項 7】**

撮像手段と複数の音入力手段とを備える情報処理装置に実行させるためのプログラムであって、

前記撮像手段の撮像方向を変更し、前記撮像手段により撮像された同じ拠点にいる 1 以上のユーザの画像を画像記録手段に記録するステップと、

前記複数の音入力手段に入力された前記同じ拠点にいるユーザの音声に基づき、音源方向を検知するステップと、

検知された音源方向へ前記撮像方向を変更し、前記撮像手段により撮像された前記ユーザの画像に基づき、当該ユーザの顔の向きを判断するステップと、

30

判断された顔の向きと同じ方向を、前記音源方向として検知したかどうかに応じて、前記ユーザと同じ拠点にいる他のユーザとの会話か、他拠点との会話かを判断するステップと、

前記他拠点との会話と判断された場合、前記撮像手段により撮像された画像に対して画像処理を行い、前記同じ拠点にいる他のユーザとの会話と判断された場合、前記撮像手段により撮像された前記ユーザと前記他のユーザとを含む画像、または前記撮像手段により撮像された前記ユーザの画像もしくは前記画像記録手段に記録された前記ユーザの画像と前記画像記録手段に記録された前記他のユーザの画像とに対して画像処理を行うステップと、

画像処理して得られた画像を前記他拠点へ送信し、前記ユーザの画像を、または前記ユーザの画像と前記他のユーザの画像とを並べて表示させるステップとを前記情報処理装置に実行させる、プログラム。

40

**【請求項 8】**

前記同じ拠点にいる他のユーザとの会話と判断された場合、前記他のユーザが複数人であるか否かを、前記複数の音入力手段に入力された音声に基づき判断するステップをさらに含み、

前記他のユーザが一人と判断された場合、前記画像処理を行うステップでは、前記撮像手段により撮像された画像に前記ユーザと前記他のユーザとが含まれるとき、前記ユーザの顔を含む所定領域と前記他のユーザの顔を含む所定領域とを切り抜き、切り抜いた 2 つの所定領域の画像を並べて表示するための画像を生成し、前記撮像手段により撮像された

50

画像に前記ユーザのみが含まれるとき、前記撮像手段により撮像された前記ユーザの画像もしくは前記画像記録手段に記録された前記ユーザの画像と前記画像記録手段に記録された前記他のユーザの画像とを並べて表示するための画像を生成する、請求項7に記載のプログラム。

【請求項9】

前記同じ拠点にいる他のユーザとの会話と判断された場合、前記他のユーザが複数人であるか否かを、前記複数の音入力手段に入力された音声に基づき判断するステップをさらに含み、

前記他のユーザが複数人と判断された場合、前記画像処理を行うステップでは、前記複数の音入力手段に対して一定時間内に音声情報の入力があり、前記撮像手段により撮像された画像に前記ユーザと他の複数のユーザ全員が含まれるとき、前記ユーザの顔を含む所定領域と前記他の複数のユーザの各顔を含む各所定領域を切り抜き、切り抜いた全ての所定領域の画像を並べて表示するための画像を生成し、前記撮像手段により撮像された画像に前記他の複数のユーザの少なくとも一人が含まれないとき、前記撮像手段により撮像された前記ユーザの画像もしくは前記画像記録手段に記録された前記ユーザの画像と前記画像記録手段に記録された前記他の複数のユーザ全員の画像を並べて表示するための画像を生成し、前記一定時間内に音声の入力がないとき、前記撮像手段により撮像された前記ユーザの画像を拡大して表示するための画像を生成する、請求項7に記載のプログラム。

10

【請求項10】

前記同じ拠点にいる他のユーザとの会話と判断された場合、前記他のユーザが複数人であるか否かを、前記複数の音入力手段に入力された音声に基づき判断するステップをさらに含み、

前記他のユーザが複数人と判断された場合、前記画像処理を行うステップでは、前記複数の音入力手段に対して一定時間内に音声情報の入力があり、前記撮像手段により撮像された画像に前記ユーザと他の複数のユーザ全員が含まれるとき、前記ユーザの顔画像と前記他の複数のユーザの各顔画像を切り抜き、切り抜いた全ての顔画像を並べて表示するための画像を生成し、前記撮像手段により撮像された画像に前記他の複数のユーザの少なくとも一人が含まれないとき、前記撮像手段により撮像された前記ユーザの画像もしくは前記画像記録手段に記録された前記ユーザの画像と前記画像記録手段に記録された前記他の複数のユーザ全員の画像を並べて表示するための画像を生成し、前記一定時間内に音声の入力がないとき、前記撮像手段により撮像された前記ユーザの画像を拡大して表示するための画像を生成する、請求項7に記載のプログラム。

20

30

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、撮像手段と複数の音入力手段とを備え、複数の音入力手段に入力された音声情報に基づいて音源方向を検知し、その音源方向へ撮像方向を変更して撮像を行い、得られた画像情報をその音声情報とともに出力する情報処理装置、2以上の情報処理装置を含む通信システム、およびその処理を情報処理装置に実行させるためのプログラムに関する。

40

【背景技術】

【0002】

地理的に離れた拠点間で、画像および音声によってコミュニケーションを取るための通信システムの1つとして、テレビ会議システムが利用されている。このテレビ会議システムには、一台のカメラで複数人の会議参加者を撮影するようなシステムが存在している。このようなシステムでは、カメラを左右に回転させることにより、会話している参加者にカメラが向くように制御している。

【0003】

このようなシステムの1つとして、会話している参加者をその音声方向から検知し、その方向にカメラを向け、その参加者の顔を検知して表示させるシステムが知られている（

50

例えば、特許文献 1 参照)。このシステムでは、会話している参加者の顔画像を、その音声に合わせて確実に伝達するべく、カメラが回転している間、音声を検出した参加者の顔画像を表示させ、口元の形を変えた画像を合成することにより音声に合わせて口を動かしている。

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、このような処理は、他拠点との間の自然なコミュニケーションを取るための画像処理であり、同じ拠点内で行われている会話については考慮されていない。すなわち、同じ拠点内で複数の参加者が会話していても、そのとき会話している一人の参加者の顔画像を表示させ、口を音声に合わせて動かすのみである。これでは、他拠点の参加者は、その顔画像が表示されている参加者が誰と会話しているのか分からないという問題があった。

10

【0005】

また、同じ拠点にいる参加者同士が交互に会話をしていると、カメラが左右に振れ続けるため、振れるたびに顔画像が切り替わり、それを見ている他拠点の参加者は、会議に集中することができないという問題もあった。

【0006】

そこで、同じ拠点内で誰と会話しているかが分かり、また、会議に集中することができるように、会話の当事者に応じた画像を表示させ、会議の臨場感を高めることができる装置等の提供が望まれている。

20

【課題を解決するための手段】

【0007】

本発明は、上記課題に鑑み、撮像手段と複数の音入力手段とを備える情報処理装置であって、撮像手段の撮像方向を制御する撮像方向制御手段と、撮像方向制御手段により撮像方向を変更し、撮像手段により撮像された同じ拠点にいる 1 以上のユーザの画像を記録する画像記録手段と、複数の音入力手段に入力された同じ拠点にいるユーザの音声に基づき、音源方向を検知する音源方向検知手段と、音源方向検知手段により検知された音源方向へ撮像方向制御手段によって撮像方向を変更し、撮像手段により撮像されたユーザの画像に基づき、当該ユーザの顔の向きを判断する顔方向判断手段と、顔方向判断手段により判断された顔の向きと同じ方向を、音源方向検知手段が音源方向として検知したかどうかに応じて、ユーザと同じ拠点にいる他のユーザとの会話か、他拠点との会話かを判断する会話者判断手段と、会話者判断手段により他拠点との会話と判断された場合、撮像手段により撮像された画像に対して画像処理を行い、同じ拠点にいる他のユーザとの会話と判断された場合、撮像手段により撮像されたユーザと他のユーザとを含む画像、または撮像手段により撮像されたユーザの画像もしくは画像記録手段に記録されたユーザの画像と画像記録手段に記録された他のユーザの画像とに対して画像処理を行う画像処理手段と、画像処理手段により画像処理して得られた画像を他拠点へ送信し、ユーザの画像を、またはユーザの画像と他のユーザの画像とを並べて表示させる通信制御手段とを含む、情報処理装置が提供される。

30

40

【発明の効果】

【0008】

本発明によれば、会話の当事者に応じた画像を表示させ、会議の臨場感を高めることが可能となる。

【図面の簡単な説明】

【0009】

【図 1】本実施形態の通信システムの構成例を示した図。

【図 2】図 1 に示す通信システムに用いられる情報処理装置のハードウェア構成図。

【図 3】図 1 に示す通信システムに用いられる情報処理装置の機能ブロック図。

【図 4】情報処理装置により実行される処理の流れを示したフローチャート。

50

【図5】情報処理装置が備える顔方向判断部および画像処理部により実行される処理の詳細な流れを示したフローチャート。

【図6】任意の拠点に配置された情報処理装置が備えるカメラの向きと、そのカメラにより撮像されたユーザの画像および他拠点にて表示されるユーザの画像の一例を示した図。

【図7】任意の拠点に配置された情報処理装置が備えるカメラの向きと、そのカメラにより撮像されたユーザの画像および他拠点にて表示されるユーザの画像の別の例を示した図。

【図8】任意の拠点に配置された情報処理装置が備えるカメラの向きと、そのカメラにより撮像されたユーザの画像および他拠点にて表示されるユーザの画像のさらに別の例を示した図。

10

【発明を実施するための形態】

【0010】

図1は、本実施形態の通信システムの構成例を示した図である。通信システムは、この構成に限定されるものではないが、地理的に離れた各拠点に配置される2以上の情報処理装置10、11がネットワーク12に接続された構成とされる。地理的に離れた拠点としては、東京本社と北海道事業所、東京本社とニューヨーク支店等を例示することができる。

【0011】

情報処理装置10、11は、それぞれ有線または無線によりネットワーク12に接続することができる。無線により接続する場合、アクセスポイントと呼ばれる基地局を介してネットワーク12に接続することができる。ここでは、2つの情報処理装置10、11のみが示されているが、3以上の情報処理装置がネットワーク12に接続されたものであってもよい。また、ネットワーク12は、有線ネットワーク、無線ネットワークのいずれであってもよく、WAN(Wide Area Network)やインターネット等を利用することができる。

20

【0012】

情報処理装置10、11は、同じハードウェア構成とされ、同じ機能を有するものとされる。ハードウェア構成および機能の詳細については後述する。任意の拠点に配置される情報処理装置10は、撮像機能を有し、その拠点にいる会議の参加者である1以上のユーザを撮像して得られた画像を、他拠点に配置される情報処理装置11へ送信し、その画像を表示させる。実際には、撮像して得られる画像データまたは画像情報を送信し、画像データまたは画像情報に基づき画像を表示させるが、説明を容易にするために、本願では「画像」という用語を使用する。

30

【0013】

また、情報処理装置10は、音入出力機能を有し、そのユーザが発した声等の入力を受け付け、それを音声等として上記の画像とともに情報処理装置11へ送信する。この音声も、実際には音声データまたは音声情報であるが、説明を容易にするために、本願では「音声」という用語を使用する。図1では、情報処理装置10が備えるカメラにより撮像を行い、複数のマイクにより音の入力を受け付ける。

【0014】

情報処理装置10は、情報処理装置11が撮像して得られた画像および入力された音声を、当該情報処理装置11から受信し、画像を表示し、音声を出力する。図1では、情報処理装置10の表示装置が備えるスピーカから音声を出力し、表示装置のディスプレイ上に画像を表示する。

40

【0015】

情報処理装置10は、音を発生させている物体、すなわち音源がある方向(音源方向)を検知し、その音源方向へ撮像方向を変更し、撮像を行うことができるように構成されている。また、情報処理装置10は、撮像方向を変更して撮像された1以上のユーザの画像を記録し、任意のユーザの画像を送信してそのユーザの画像を表示させることができるように構成されている。このため、情報処理装置10は、会話しているユーザを撮像するために撮像方向を変更している間、記憶されているそのユーザの画像情報を読み出し、送信

50

して、情報処理装置 11 にそのユーザの画像を表示させることができる。

【0016】

情報処理装置 10 は、そのユーザの画像内の顔画像と口元の形を変えた画像とを合成した画像を生成し、送信することができるように構成されている。このため、情報処理装置 10 は、その画像を送信し、情報処理装置 11 に音声に合わせて口が動く擬似画像を表示させることができる。この擬似画像を表示させることにより、実際に喋っているように見せることができる。

【0017】

また、情報処理装置 10 は、撮像されたユーザの顔の向きを判断し、その顔の向きと同じ方向を、音源方向として検知したかどうかに応じて、同じ拠点にいる他のユーザと会話しているか、他拠点にいるユーザと会話しているかを判断するように構成されている。また、情報処理装置 10、11 は、同じ拠点にいる他のユーザと会話している場合、その二人のユーザの画像を並べて表示させるための画像を生成し、それを送信するように構成されている。このため、他拠点にいるユーザは、その二人のユーザが会話していることを知ることができる。

10

【0018】

同じ拠点にいる二人のユーザが会話を行う場合、撮像方向がそのつど変更になるが、撮像方向を変更している間、上記の二人のユーザの画像を並べて表示させるための画像を生成し、送信している。このため、画像に振れが生じることはなく、それを見ている他拠点にいるユーザは、会議に集中することが可能となる。

20

【0019】

上記のような処理や機能を実現するためのハードウェア構成を、図 2 に例示する。情報処理装置 10、11 は同じ構成であるため、情報処理装置 10 についてのみ説明する。情報処理装置 10 は、撮像機能および音入出力機能を実現するために、カメラ 20 といった撮像手段と、マイクアレイ 21 といった複数の音入力手段と、スピーカ 22 といった音出力手段とを備える。

【0020】

カメラ 20 は、入力する光を集束させるレンズと、レンズにより集束された光を電気信号に変換する撮像素子とを含み、静止画、または連続して撮像を行い、得られた静止画を時系列に並べて動画とし、その動画を出力する。カメラ 20 は、静止画または動画を設定するためのモード設定ボタン等を備え、その設定に応じていずれかの画像を出力する。カメラ 20 としては、例えば、デジタルカメラやビデオカメラを挙げることができる。

30

【0021】

このカメラ 20 は、例えば、このカメラ 20 を支持する支持部を備え、支持部を中心として左右の方向に回転可能とされる。回転する角度は、会議に参加している複数のユーザの各々に向けて撮像可能な角度であれば、120度、180度や360度等、いかなる角度であってもよい。この構成に限られるものではなく、カメラ 20 は、ターンテーブルに載置され、ターンテーブルを回転させることにより左右の方向に回転させることも可能である。

【0022】

マイクアレイ 21 は、複数のマイクロフォンから構成され、複数のマイクロフォンが筐体内に水平方向に一列に配列したものを使用してもよいし、各マイクロフォンが自在に配置できるようになっていて、各マイクロフォンを水平方向に一列に並べて使用してもよい。また、一列ではなく、各ユーザの前に各マイクロフォンを1つずつ配置して使用することもできる。会話をしているユーザをより正確に検知するためには、各ユーザの前に各マイクロフォンを配置して使用することが望ましい。

40

【0023】

マイクアレイ 21 に使用されるマイクロフォンは、磁石、コイル、振動板を含み、振動板が音波を受けて振動し、振動板の振動を、磁石に近隣して配置されるコイルが受け、コイル内の磁束を変化させることにより起電力が発生し、それを音信号として出力する。

50

## 【 0 0 2 4 】

スピーカ 2 2 は、他拠点から送信されてきた音情報を再生して出力する。スピーカ 2 2 も、マイクロフォンと同様、磁石、コイル、振動板を含むものとしてすることができる。この場合、スピーカ 2 2 は、入力された音声信号により、磁石に近隣して配置されるコイル内の磁束が変化し、それによってコイルが振動し、振動板が振動して、その振動板に接する空気が振動して音波を発生させることにより音を出力する。

## 【 0 0 2 5 】

情報処理装置 1 0 は、そのほか、表示装置 2 3 およびコントローラ 2 4 を含んで構成される。表示装置 2 3 は、情報処理装置 1 0 へ送られてきた画像を表示する。通信システムにおいては、他拠点において会話しているユーザの画像が送られてくるので、そのユーザの画像として、そのユーザの顔の静止画または動画を表示する。表示装置 2 3 としては、ディスプレイを用いることができ、スクリーンおよびプロジェクタを用いることもできる。

10

## 【 0 0 2 6 】

コントローラ 2 4 は、CPU 2 5 と、ROM 2 6 と、RAM 2 7 と、HDD 2 8 と、ネットワーク 1 2 に接続するためのネットワーク I / F 2 9 とを含んで構成される。

## 【 0 0 2 7 】

ROM 2 6 は、情報処理装置 1 0 の起動時に実行される BIOS (Basic Input / Output System) 等のプログラムを記憶する。RAM 2 7 は、CPU 2 5 が作業を行うために必要とされる記憶領域を提供する。HDD 2 8 は、アプリケーションや OS 等のプログラム、それらに関連するデータ等を記憶する。ここでは、HDD 2 8 を使用しているが、SSD (Solid State Drive) 等のその他の記憶装置を用いてもよい。

20

## 【 0 0 2 8 】

このプログラムは、ネットワーク 1 2 を介して、または図示しない記録媒体を介して提供され、HDD 2 8 に格納される。情報処理装置 1 0 は、この記録媒体を接続可能にするために外部記憶装置 I / F を備えることができる。記録媒体としては、CD - ROM、DVD、SD カード等を挙げることができ、外部記憶装置 I / F としては、これらを読み書き可能にする CD ドライブ、DVD ドライブ、SD カードスロット等を挙げることができる。

## 【 0 0 2 9 】

CPU 2 5 は、情報処理装置 1 0 内の各手段を制御し、データの演算や加工を行う。各手段としては、上記の撮像手段、音入力手段、音出力手段、HDD 2 8 等の記憶手段等である。CPU 2 5 は、カメラ 2 0 やマイクアレイ 2 1 等からデータを受け取り、また、HDD 2 8 等からデータを読み出し、演算や加工を行い、それをネットワーク 1 2 上や表示装置 2 3 へ出力し、また、HDD 2 8 等に記憶させる処理を実行する。

30

## 【 0 0 3 0 】

情報処理装置 1 0 は、電源が投入されると、ROM 2 6 から BIOS を取り出して実行し、カメラ 2 0、マイクアレイ 2 1、スピーカ 2 2、表示装置 2 3、HDD 2 8 等が使用できることをチェックする。そして、情報処理装置 1 0 は、HDD 2 8 から OS を RAM 2 7 に読み出し、実行することにより起動する。その後、情報処理装置 1 0 は、OS による制御の下、アプリケーション等のプログラムを実行し、所望の処理を実現する。

40

## 【 0 0 3 1 】

図 3 を参照して、情報処理装置 1 0 が備える機能について詳細に説明する。情報処理装置 1 1 は、情報処理装置 1 0 と同様の機能を備えるため、ここでは説明を省略する。なお、図 3 には、説明を分かりやすくするため、撮像手段としてのカメラ 2 0、音入力手段としてのマイクアレイ 2 1 も図示されている。

## 【 0 0 3 2 】

情報処理装置 1 0 は、機能部として、撮像方向制御部 3 0、画像記録部 3 1、音源方向検知部 3 2、顔方向判断部 3 3、会話者判断部 3 4、画像処理部 3 5、通信制御部 3 6 を含んで構成される。情報処理装置 1 0 は、上記機能部のみであってもよいが、そのほか、

50

音声入力部 37、人物判断部 38 を含むことができる。これらの機能部は、HDD 28 やネットワーク I/F 29 のほか、CPU 25 が HDD 28 からプログラムを読み出し実行することにより実現される。

【0033】

撮像方向制御部 30 は、カメラ 20 のレンズの向きを変えることにより撮像方向を制御する。カメラ 20 のレンズの向きは、カメラ 20 を支持する支持部を中心として左右の方向に回転させることにより変えることができる。また、ターンテーブル上にカメラ 20 を載置した構成の場合、撮像方向制御部 30 は、ターンテーブルをいずれかの方向に回転させることにより撮像方向を変えることができる。また、撮像方向制御部 30 は、音源方向検知部 32 により検知された音源方向の情報や、人物判断部 38 の判断結果を受けて、い

10

【0034】

人物判断部 38 は、情報処理装置 10 が動作を開始したタイミングや、予め設定されたタイミングにおいて、カメラ 20 により撮像された画像に人物が含まれるか否かを判断する。人物が含まれるか否かは、公知の顔認識アルゴリズムを使用して判断することができる。顔認識アルゴリズムでは、目、鼻、口、あご等の顔のパーツの相対位置、大きさ、形等の特徴を抽出し、また、肌色を検出することにより、顔認識を行う。ここでは、顔認識アルゴリズムを使用して人物の有無を判断しているが、人物の有無を判断することができれば、その他の方法を採用することもできる。

20

【0035】

画像記録部 31 は、HDD 28 等により実現され、人物判断部 38 により人物が含まれることを判断し、撮像方向制御部 30 がその判断結果から決定した撮像方向にカメラ 20 の向きを変更し、カメラ 20 により撮像して得られた人物であるユーザの画像を記録する。画像記録部 31 は、このようにして得られたその拠点にいるユーザ全員の画像を記録する。記録する画像は、静止画の画像である。画像記録部 31 は、各ユーザの画像を記録する際、各ユーザの位置情報と関連付けて記録される。

【0036】

ユーザの位置情報は、例えば、カメラ 20 が正面を向いたときの角度を、基準の 0 度とし、左右の方向へ回転させたときの角度の情報とすることができる。なお、角度には誤差が生じるので、左右に 5° 程度の誤差範囲を設けることが望ましい。また、ユーザの位置情報は、角度に限らず、東西南北のような方位を用いてもよい。このため、情報処理装置 10 は、角度を測定するためのセンサ、ロータリエンコーダ、方位磁針等を備えることができる。

30

【0037】

音声入力部 37 は、マイクアレイ 21 から入力された音が、情報処理装置 10 が配置されている拠点にいるユーザからの音声入力であるか否かを検知し、音声入力を検知した場合、その音声を音源方向検知部 32 へ送る。入力される音には、雑音も含まれるが、一般に音声の方が、音量が大きいことから音量により音声入力かどうかを検知することができる。これは一例であるので、音声を検知する方法としては、これまでに知られたいかなる方法でも使用することができる。

40

【0038】

音源方向検知部 32 は、音声入力部 37 からの音声情報から音源方向を検知する。マイクアレイ 21 は、複数のマイクロフォンから構成され、各マイクロフォンは異なる位置に配置されるため、入力される音声は、各マイクロフォンによって時間差が生じる。最先に音声が入力されたマイクロフォンは、音源に最も近いことを示すため、どのマイクロフォンに最先に音声が入力されたかを判断することにより、その音源方向を検知することができる。音源方向検知部 32 は、検知した音源方向を、音源方向の情報として撮像方向制御部 30 へ入力する。なお、音源方向は、上記ユーザの位置情報と同様の情報とすることができる。

【0039】

50

顔方向判断部 33 は、音源方向検知部 32 により検知された音源方向に向けて撮像方向制御部 30 が撮像方向を変更し、その方向に向けられたカメラ 20 により撮像されたユーザの画像に基づき、そのユーザの顔の向きを判断する。顔方向判断部 33 は、例えば、画像記録部 31 に記録されたそのユーザの画像が正面を向いた顔であるので、その顔の向きを基準とし、撮像された画像の顔の角度を推定することにより顔の向きを判断することができる。

【0040】

顔方向判断部 33 は、この角度も誤差が生じるので、例えば左右に 5° 程度の誤差範囲を設けることが望ましい。また、顔方向判断部 33 は、角度に限らず、東西南北のような方位により顔の向きを判断してもよい。

10

【0041】

会話者判断部 34 は、顔方向判断部 33 により判断された顔の向きと同じ方向を、音源方向検知部 32 が音源方向として検知したかどうかに応じて、いずれの会話であるかを判断する。会話者判断部 34 は、顔の向きと同じ方向を音源方向として検知した場合、そのユーザと同じ拠点にいる他のユーザとの会話と判断する。また、会話者判断部 34 は、顔の向きと同じ方向を音源方向として検知しない場合、他拠点との会話と判断する。

【0042】

会話者判断部 34 は、同じ拠点にいる他のユーザとの会話と判断した場合、当該他のユーザが複数人であるか否かを、マイクアレイ 21 に入力された音声に基づき判断する。すなわち、複数人の音声が入力されていれば、複数人と判断し、特定のユーザの音声のみが入力されていれば、そのユーザー一人と判断することができる。これらの情報は、画像処理部 35 が画像処理を行う際に利用される。

20

【0043】

このとき会話を行っている他のユーザは、音源方向検知部 32 により音源方向として検知され、得られたその音源方向の情報からその位置が特定される。

【0044】

画像処理部 35 は、カメラ 20 により撮像された画像あるいは画像記録部 31 に記録された画像に対して画像処理を行い、他拠点にて表示させるための画像を生成する。例えば、カメラ 20 を回転させている間に他拠点で表示させる画像を、画像記録部 31 に記録されたそのユーザの画像に口元の形を変えた画像を合成することにより生成する。

30

【0045】

また、カメラ 20 により撮像された画像から 2 以上のユーザの顔画像等を切り抜き、並べる処理や、画像記録部 31 に記録された 2 以上のユーザの画像の、表示する位置や大きさを変更する処理等を行うことにより生成する。このような合成する処理、切り抜き、並べる処理、表示する位置や大きさを変更する処理等については、従来から知られているいかなる方法を採用して実施することができ、ここではその詳細な方法について省略する。

【0046】

会話者判断部 34 により他のユーザが一人と判断された場合、カメラ 20 により撮像された画像にユーザと当該他のユーザの二人が含まれているとき、画像処理部 35 は、その画像からその二人のユーザの顔画像あるいは顔を含む所定領域を切り出す処理を行う。カメラ 20 が回転し、振れたとしても、小さな振れであり、他拠点にいるユーザが会議に集中できなくなるような振れではないためである。

40

【0047】

画像処理部 35 は、切り抜いた顔画像あるいは所定領域の画像を並べて表示するための画像を生成する。このときの画像は、動画であるため、動画における顔や所定領域を並べて表示するための画像が生成される。所定領域としては、例えば、顔を含む最小の矩形の領域とすることができる。また、並べ方としては、上下あるいは左右に並べることができる。一般に、表示画面は、横長であるため、左右に並べることが望ましい。

【0048】

一方、他のユーザが含まれていない場合は、カメラ 20 が大きく振れることになるため

50

、画像記録部 3 1 に記録されたユーザの画像と他のユーザの画像とを用い、それらを並べて表示するための画像を生成する。画像記録部 3 1 に記録される画像は、静止画であるため、その静止画を並べた画像が生成される。このとき、単に並べただけではなく、口元の形を変えた画像を合成した画像が生成される。

【 0 0 4 9 】

他のユーザが複数人と判断された場合に、一定時間内に音声の入力があり、カメラ 2 0 により撮像された画像にユーザと他の複数のユーザ全員が含まれているとき、画像処理部 3 5 は、その画像からユーザ全員の顔画像あるいは顔を含む所定領域を切り出す。画像処理部 3 5 は、各ユーザにつき切り出す処理を行う。そして、画像処理部 3 5 は、切り抜いた顔画像あるいは所定領域の画像を並べて表示するための画像を生成する。この場合、複数の画像を並べて表示することになるため、表示画面上の上下左右に並ぶように縮小（ズームアウト）して配置した画像を生成することができる。

10

【 0 0 5 0 】

他の複数のユーザが含まれていない場合は、画像記録部 3 1 に記録されたユーザの画像と他の複数のユーザ全員の画像とを用い、それらユーザの画像を並べて表示するための画像を生成する。画像記録部 3 1 に記録される画像は、静止画であるため、その静止画を並べて表示するための画像情報が生成される。この場合も、生成される画像は、口元の形を変えた画像を合成した画像とされる。

【 0 0 5 1 】

一定時間内に音声の入力がない場合は、他のユーザと会話していないものとみなし、カメラ 2 0 により撮像されたユーザの画像を拡大（ズームアップ）して表示するための画像を生成する。ここではズームアップして表示するための画像を生成しているが、ズームアップしなくてもよいし、顔画像のみを切り抜いて表示してもよい。また、顔を含む所定領域を切り抜いて表示してもよい。

20

【 0 0 5 2 】

通信制御部 3 6 は、画像処理部 3 5 により画像処理して得られた画像を、マイクアレイ 2 1 に入力された音声とともに他拠点へ送信し、また、他拠点から画像および音声を受信するための通信制御を行う。このように会話の当事者に応じた画像を表示させることができるため、他拠点にいるユーザが、誰に向かって会話をしているかが分かり、また、他拠点にいるユーザも、会議に集中することができ、さらには、会議の臨場感を高めることができる。

30

【 0 0 5 3 】

図 4 に示すフローチャートを参照して、情報処理装置 1 0 が行う処理について詳細に説明する。情報処理装置 1 0 は、情報処理装置 1 0 が配置される拠点において、電源が投入され、プログラムが読み出されて実行されることにより、ステップ 4 0 0 から処理を開始する。ここでは自動的にプログラムが読み出されて実行されるようになっているが、自動的に実行されない場合、そのプログラムが起動された段階でこの処理を開始することができる。

【 0 0 5 4 】

ステップ 4 0 5 では、その拠点にいる会議の参加者であるユーザ全員をカメラ 2 0 により撮像する。このとき、撮像方向制御部 3 0 によりカメラ 2 0 の向きを変え、ユーザ全員の顔を撮像する。そして、撮像して得られた顔画像において口元を検出する。後の処理において、口元の形を変えた画像と合成し、実際に話しているように見せる処理を行うためである。

40

【 0 0 5 5 】

ステップ 4 1 0 では、全員の顔が撮像された後、実際に会議が始まり、マイクアレイ 2 1 が音の入力を受け付ける。音には、音声のほか、エアコンや外を走る車等の雑音がある。ステップ 4 1 5 では、実際に会話が始まり、音声入力部 3 7 が、ユーザからの音声入力であるか否かを検知する。そして、音声入力を検知した場合、その音声を音源方向検知部 3 2 へ送る。

50

## 【0056】

ステップ420では、音源方向検知部32が、音声入力部37からの音声に基づき音源方向を検知する。音源方向検知部32は、検知した音源方向の情報を、撮像方向制御部30へ入力する。ステップ425では、撮像方向制御部30は、入力された音源方向の情報から、撮像方向の変更が必要かどうかを判断する。必要と判断した場合、ステップ430へ進み、必要でないと判断した場合、ステップ455へ進む。

## 【0057】

ステップ430では、撮像方向制御部30が、音源方向の情報に基づき撮像方向を変更する。そして、画像記録部31から音源方向のユーザの画像を読み出す。画像処理部35は、読み出した画像と口元の形を変えた画像とを合成し、その合成した画像を生成する。その合成した画像を送信して表示させる。

10

## 【0058】

ステップ435では、カメラ20の向きが、音源方向になったかどうかを判断する。この判断は、カメラ20の向きが、音源方向になるまで繰り返される。音源方向になった場合、ステップ440へ進み、顔方向判断部33が、会話をしているユーザの顔の向きを判断する。すなわち、カメラ20の方向に向いているか、それ以外の方向に向いているかを判断し、それ以外の方向である場合、どのユーザの方向に向いているかを判断する。

## 【0059】

ステップ445では、ステップ440での判断結果から、同じ拠点での会話であるかどうかを判断する。上記のそれ以外の方向に向いているという判断結果である場合、同じ拠点での会話と判断する。カメラ20の方向に向いているという判断結果である場合、他拠点との会話と判断する。同じ拠点での会話と判断した場合、ステップ450へ進み、他拠点との会話と判断した場合、ステップ455へ進む。

20

## 【0060】

ステップ450では、カメラ20により撮像された画像内のユーザと会話している他のユーザの顔画像を切り抜き、切り抜いた2つの顔画像を並べて表示させるための画像を生成し、その画像を送信する。一方、ステップ455では、カメラ20に向いているユーザをズームアップした画像を生成し、その画像を送信する。これらの送信が終了したところで、ステップ460へ進み、この処理を終了する。

## 【0061】

次に、ステップ440～ステップ455にて顔方向判断部33および画像処理部35が行う処理について、図5に示すフローチャートを参照して詳細に説明する。顔方向判断部33は、上記のステップ435でカメラ20の向きが音源方向になったときに、ステップ500からこの処理を開始する。ステップ505では、会話しているユーザの顔の向きを判断する。顔の向きは、上記に例示した方法により判断することができる。

30

## 【0062】

ステップ510では、ステップ505にて判断した顔の向きに基づき、その方向に他のユーザがいるかどうかを判断する。例えば、画像記録部31に記録された画像に関連付けられた位置情報を用いて、その方向に他のユーザがいるかどうかを判断することができる。他のユーザがいると判断した場合、ステップ515にて同じ拠点で会話していると判断し、ステップ520にてその方向にいる他のユーザが複数人かどうかを判断する。複数人であると判断した場合、ステップ525へ進み、一人であると判断した場合、直接ステップ530へ進む。

40

## 【0063】

ステップ525では、一定期間内、すなわち予め設定された期間内に他のユーザの音声を検知したかどうかを判断する。検知した場合、ステップ530へ進み、検知しなかった場合、ステップ560へ進む。ステップ530では、カメラ20で撮像する撮像範囲内に他のユーザが存在するかどうかを判断する。すなわち、会話しているユーザと他のユーザが存在するかどうかを判断する。

## 【0064】

50

他のユーザが存在すると判断した場合、ステップ535へ進み、画像処理部35が、カメラ20により撮像された画像内からユーザおよび他のユーザの顔画像を切り抜き、それらを並べて表示するための画像を生成する。そして、ステップ565へ進み、この処理を終了する。

【0065】

ステップ530において他のユーザが存在しないと判断した場合、ステップ540へ進み、画像処理部35は、画像記録部31から会話をしているユーザおよび他のユーザの画像を取得する。ステップ545では、取得したユーザの顔画像と口元の形の画像とを合成し、合成したユーザの顔画像を並べて表示するための画像を生成する。そして、ステップ560へ進み、この処理を終了する。

10

【0066】

ステップ510でその方向に他のユーザがいないと判断した場合、ステップ550へ進み、他拠点にいるユーザと会話していると判断する。そして、ステップ555で、表示すべきユーザは、他のユーザがいないので、当該ユーザのみであり、カメラ20により撮像された当該ユーザの画像を処理し、ズームアップした画像を生成する。ステップ560で、この処理を終了する。

【0067】

実際に情報処理装置10は、任意の拠点である本社あるいは事業所内のある会議室の所定位置に配置される。図6(a)では、会議室に置かれたテーブル40上の縁部にカメラ20を中央とし、その左右にマイクアレイ21を配置した構成とされている。テーブル40を挟んで左右およびカメラ20の正面に3人のユーザA、B、Cが会議に参加している。

20

【0068】

カメラ20のレンズは、正面にいるユーザBの方を向いている。このため、カメラ20により撮像して得られた画像は、図6(b)のような画像となる。すなわち、正面にユーザB、その左右にユーザA、Cがテーブル40を挟んで座っている画像となる。

【0069】

この画像を他拠点に配置された情報処理装置11へ送信し、表示させた場合の画像は、図6(c)のような画像となる。この画像は、図6(b)と同様、正面にユーザB、その左右にユーザA、Cがテーブル40を挟んで座っている画像である。

30

【0070】

ユーザCが会話を開始すると、音源方向検知部32によりそのユーザCの方向を音源方向として検知するため、カメラ20の向きがその方向に変更される。このときの様子を、図7に例示する。図7(a)に示すように、カメラ20の向きが、ユーザCがいる方向に変更される。

【0071】

カメラ20により撮像して得られた画像は、ユーザCを中心とした画像となるため、図7(b)のような画像となる。この画像は、ユーザB、Cのみを含み、ユーザAは含まれない。また、図7に示す実施形態では、ユーザCが正面を向き、他拠点にいるユーザと会話しているため、他拠点に配置された情報処理装置11へ送信し、表示させた場合、そのユーザCをズームアップした画像が送信され、そのズームアップした画像が表示される。このため、その画像は、図7(c)に示すようなユーザCをズームアップした画像となる。このように、会話の当事者を特定し、それを画像として表示することで、臨場感を高めることができる。

40

【0072】

図7では他拠点との間で会話を行っている場合の様子を例示したが、同じ拠点で会話を行っている場合の様子を、図8を参照して説明する。ユーザAがユーザCと会話しているものとする。

【0073】

カメラ20の向きは、図8(a)に示すように、現在会話をしているユーザAに向けら

50

れる。このときに撮像して得られた画像が、図8(b)に示すような画像である。図8に示す実施形態では、ユーザAが正面を向いておらず、同じ拠点にいるユーザCがいる方向を向き、ユーザCと会話を行っている。このため、他拠点に配置された情報処理装置11では、図8(c)に示すようなユーザA、Cの画像を並べた画像が表示される。これらの画像は、カメラ20により撮像して得られたユーザAの動画と、画像記録部31に記録されたユーザCの画像を並べて表示した画像で、ユーザCの画像は口元の形の画像が合成されるので、口元が音声に合わせて動く。この場合も、会話の当事者を特定し、それを画像として表示することで、臨場感を高めることができる。

【0074】

ユーザA、Cが一定時間内の短い会話をやりとりしている場合、ユーザAについては動画を、ユーザCについては、静止画に口元の形の画像を合成したものを表示し続けることができる。ユーザCが会話を開始し、その会話が上記一定時間より長い場合、カメラ20がユーザCの方を向き、撮像を開始する。カメラ20が移動している間は、ユーザAについても、画像記録部31に記録された静止画の画像に口元の形の画像が合成されたものを表示させることができる。そして、カメラ20がユーザCの方を向き、撮像を開始すると、ユーザCの画像を静止画から動画へ切り替え、ユーザAについては画像記録部31に記録された静止画の画像に口元の形の画像が合成されたものを表示させる。再びユーザAが会話を開始し、その会話が長い場合は、再びカメラ20の向きが変えられるので、ユーザCの画像を静止画に、ユーザAの画像を動画に戻すことができる。

10

【0075】

これまで本発明の情報処理装置および通信システム、プログラムにより実行される処理について図面に示した実施形態を参照しながら詳細に説明してきたが、本発明は、上述した実施形態に限定されるものではない。したがって、他の実施形態や、追加、変更、削除など、当業者が想到することができる範囲内で変更することができ、いずれの態様においても本発明の作用・効果を奏する限り、本発明の範囲に含まれるものである。よって、本発明では、情報処理装置や通信システムにより実行される方法や、プログラムが記録された記録媒体も提供することができるものである。

20

【符号の説明】

【0076】

10、11...情報処理装置、12...ネットワーク、20...カメラ、21...マイクアレイ、22...スピーカ、23...表示装置、24...コントローラ、25...CPU、26...ROM、27...RAM、28...HDD、29...ネットワークI/F、30...撮像方向制御部、31...画像記録部、32...音源方向検知部、33...顔方向判断部、34...会話者判断部、35...画像処理部、36...通信制御部、37...音声入力部、38...人物判断部、40...テーブル

30

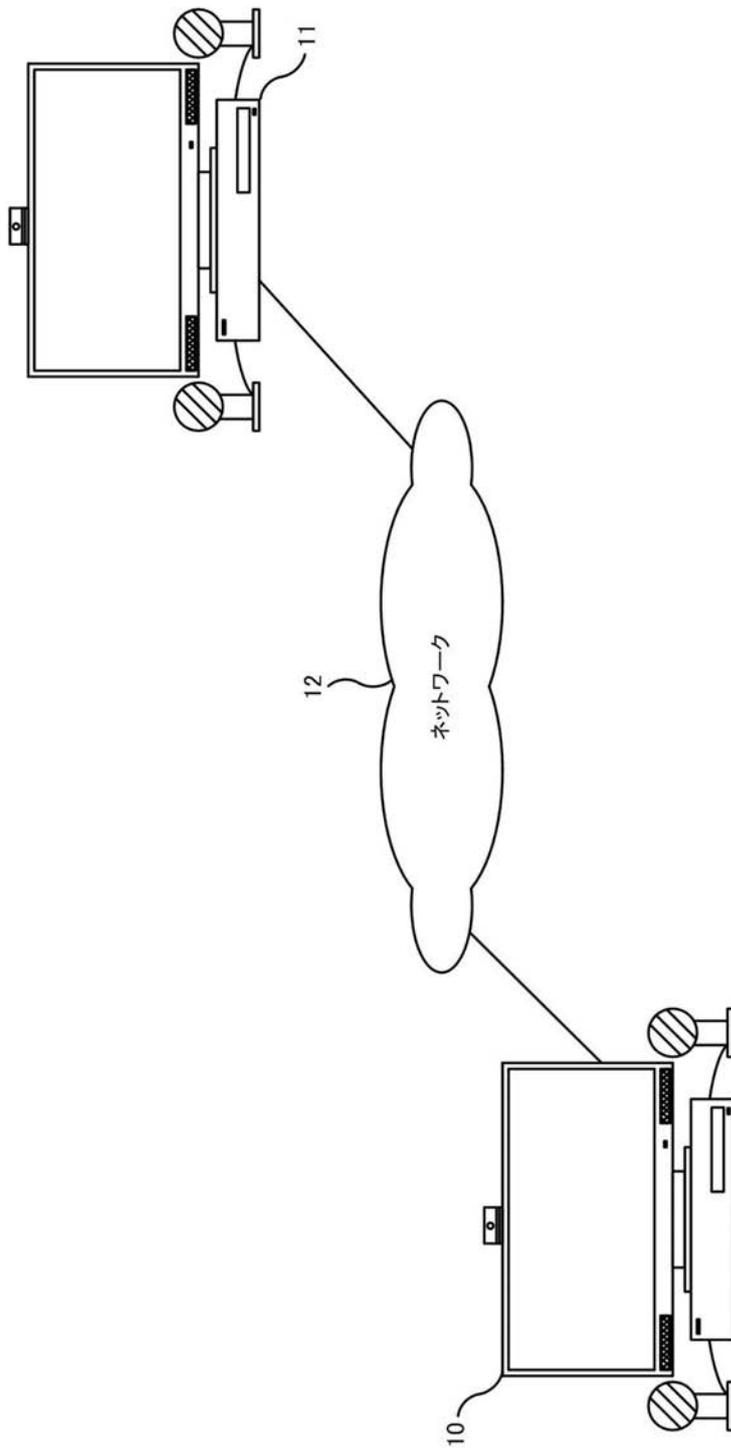
【先行技術文献】

【特許文献】

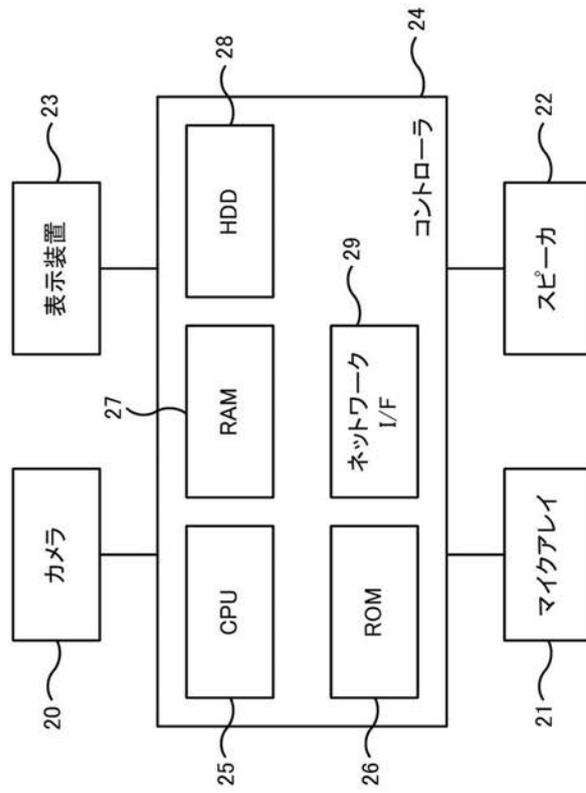
【0077】

【特許文献1】特開平5-122689号公報

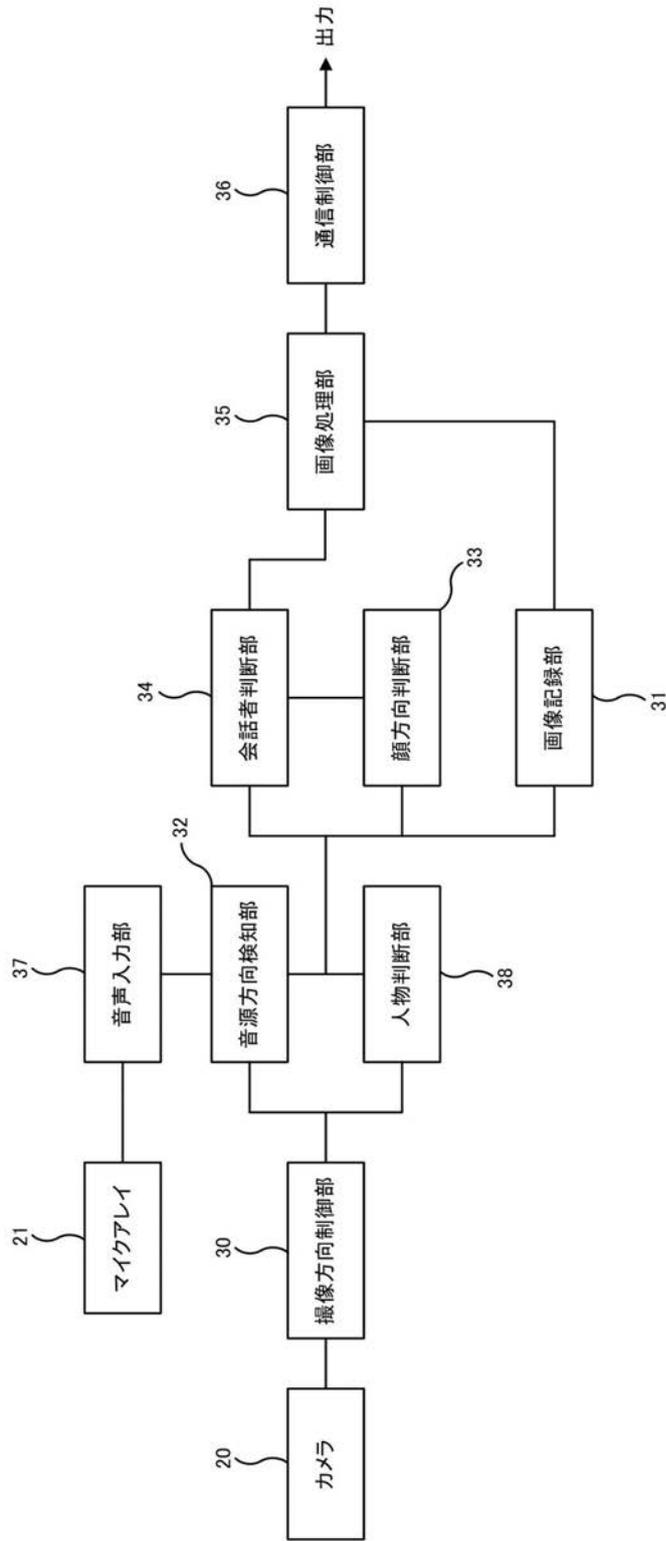
【 図 1 】



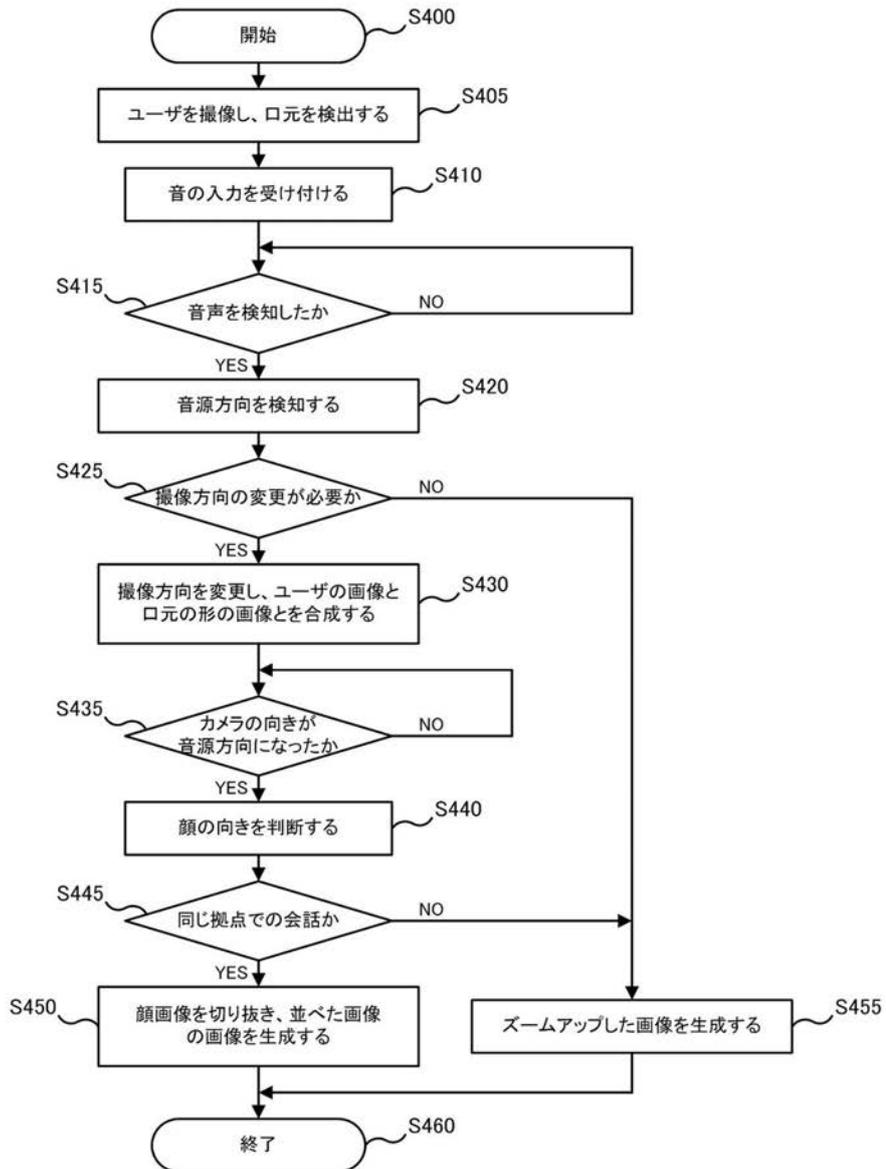
【 図 2 】



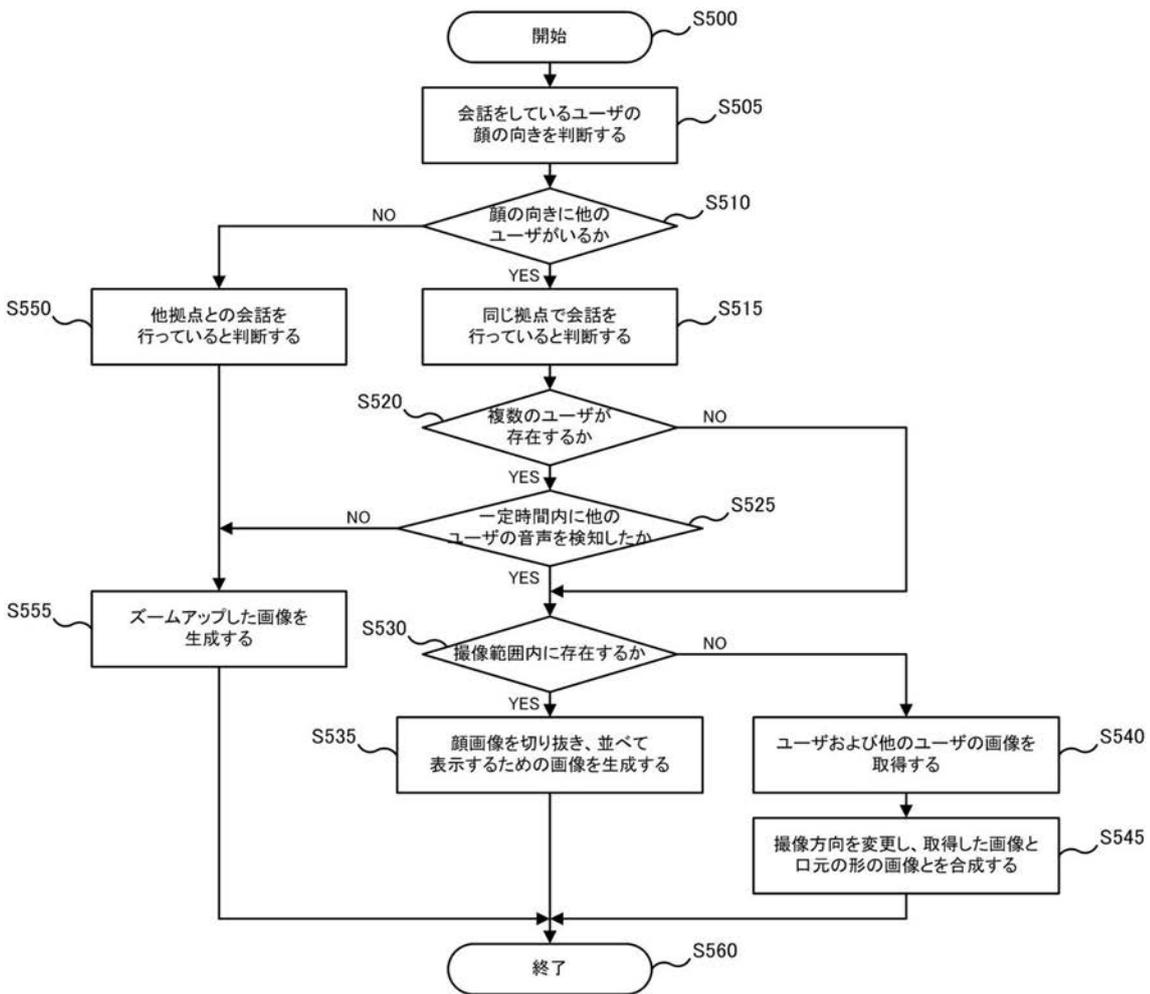
【図 3】



【 図 4 】

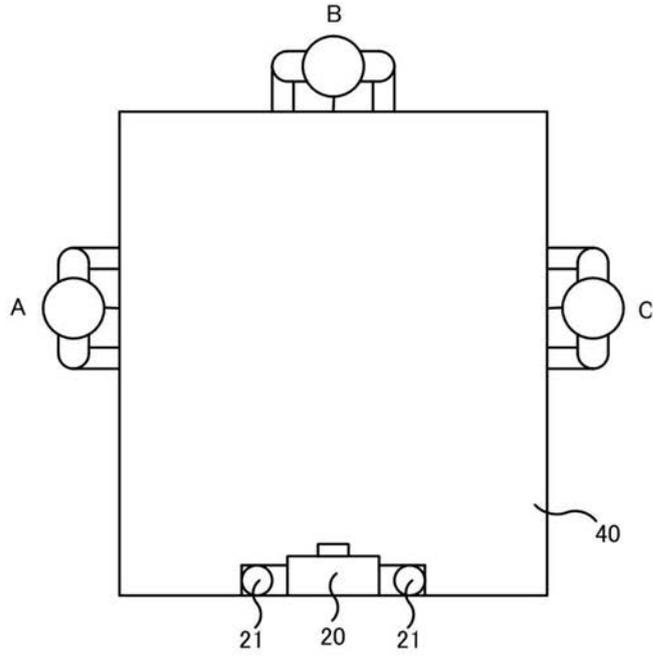


【 図 5 】

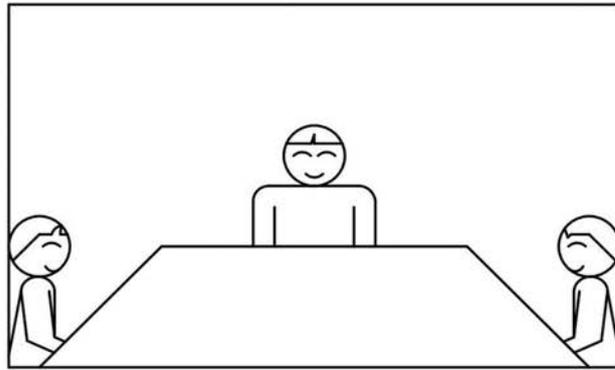


【 図 6 】

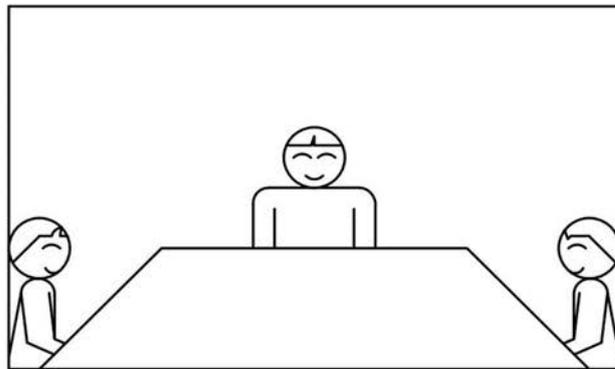
(a)



(b)

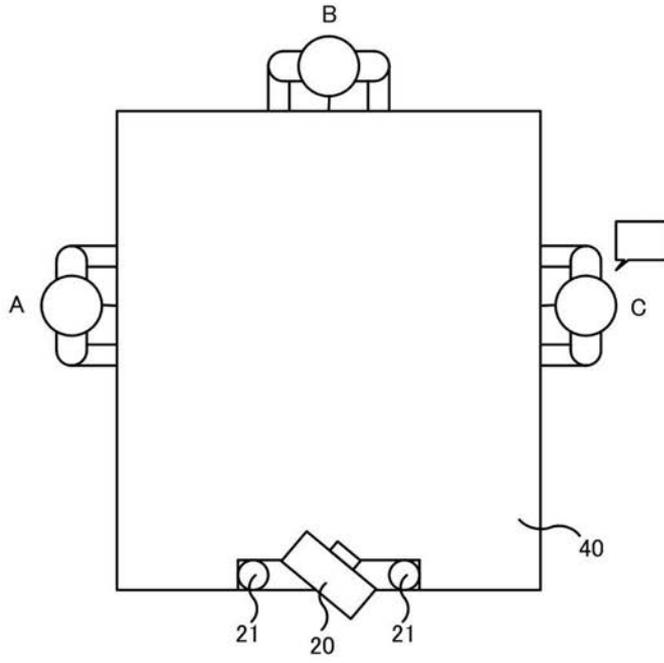


(c)

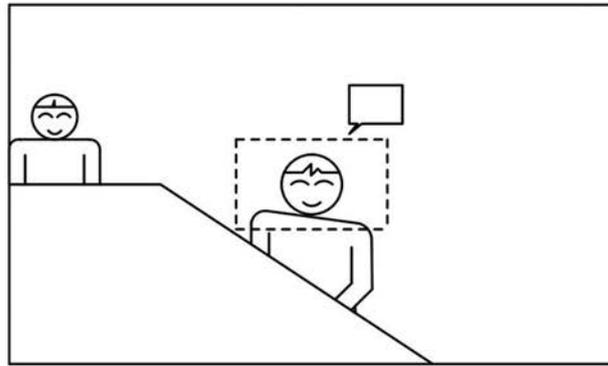


【 図 7 】

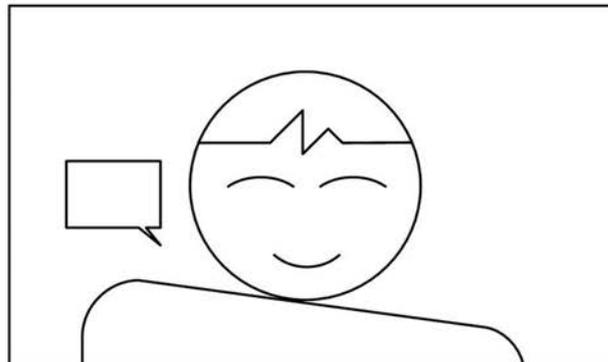
(a)



(b)

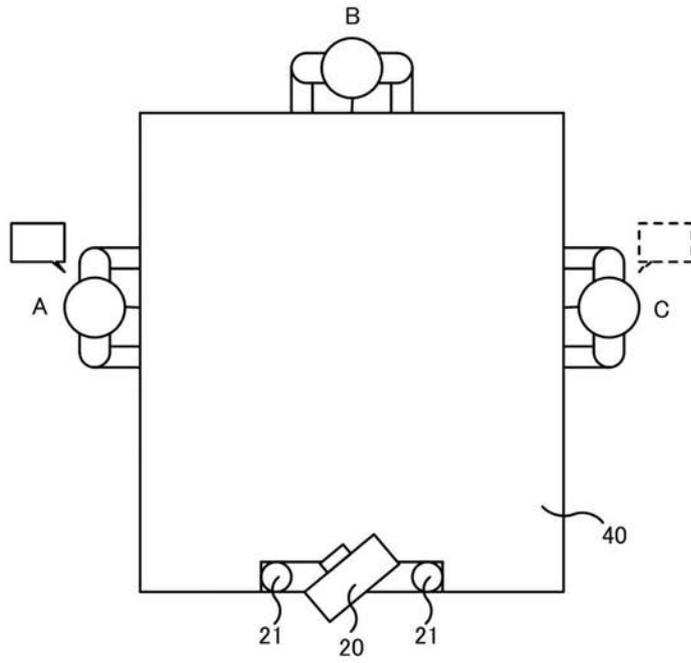


(c)

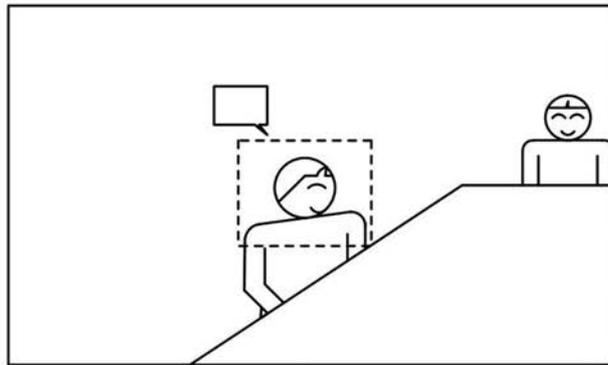


【 図 8 】

(a)



(b)



(c)

