

(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2012-244285

(P2012-244285A)

(43) 公開日 平成24年12月10日(2012.12.10)

(51) Int.Cl.

H04N 7/15 (2006.01)

F I

H04N 7/15 630Z

テーマコード (参考)

5C164

審査請求 未請求 請求項の数 7 O L (全 9 頁)

(21) 出願番号 特願2011-110379 (P2011-110379)
 (22) 出願日 平成23年5月17日 (2011.5.17)

(71) 出願人 000004226
 日本電信電話株式会社
 東京都千代田区大手町二丁目3番1号
 (74) 代理人 100083806
 弁理士 三好 秀和
 (74) 代理人 100129230
 弁理士 工藤 理恵
 (72) 発明者 玉木 秀和
 東京都千代田区大手町二丁目3番1号 日
 本電信電話株式会社内
 (72) 発明者 東野 豪
 東京都千代田区大手町二丁目3番1号 日
 本電信電話株式会社内
 Fターム(参考) 5C164 FA10 UB41S UB88S VA42P YA12

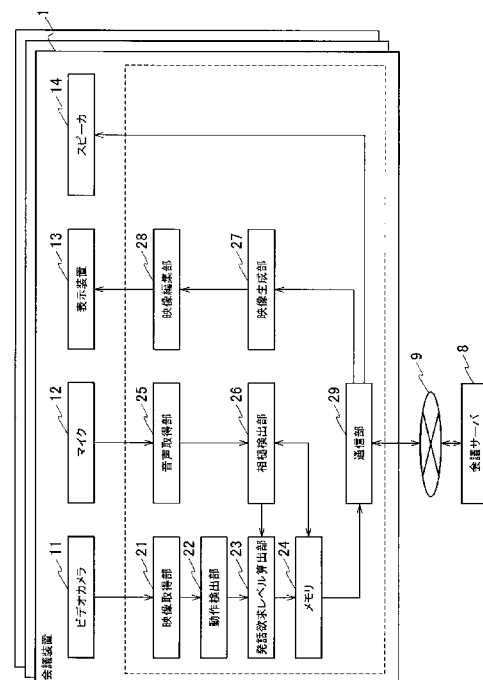
(54) 【発明の名称】 会議装置、会議方法および会議プログラム

(57) 【要約】

【課題】参加者の表情や仕草が読み取れない場合であっても、各参加者の発話欲求を容易に認識させる。

【解決手段】ネットワークを介して遠隔会議を行う会議装置1であって、自参加者の映像データを取得する映像取得手段21と、自参加者の映像データから発話欲求が想定される動作を検出する動作検出手段22と、検出した動作に基づいて発話欲求レベルを算出する発話欲求レベル算出手段23と、他の会議装置から他参加者の映像データおよび発話欲求レベルを受信する受信手段29と、各参加者の映像データを配置した会議映像を生成する映像生成手段27と、各参加者の発話欲求レベルを示す情報を会議映像に設定する映像編集手段28と、編集した会議映像を表示する表示手段13とを有する。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

ネットワークを介して遠隔会議を行う会議装置であって、
当該会議装置を使用する自参加者の映像データを取得する映像取得手段と、
前記自参加者の映像データから発話欲求があると想定される所定の動作を検出する動作
検出手段と、
前記動作検出手段が検出した動作に基づいて発話欲求レベルを算出する発話欲求レベル
算出手段と、
ネットワークを介して他の会議装置から他参加者の映像データおよび発話欲求レベルを
受信する受信手段と、
自参加者および他参加者の映像データを配置した会議映像を生成する映像生成手段と、
自参加者および他参加者の発話欲求レベルを示す情報を、前記会議映像に設定する映像
編集手段と、
前記映像編集手段が編集した会議映像を表示する表示手段と、を有すること
を特徴とする会議装置。

10

【請求項 2】

請求項 1 記載の会議装置であって、
前記映像編集手段は、発話欲求レベルが所定の値以上の参加者の映像データの近傍に、
所定のマークを設定すること
を特徴とする会議装置。

20

【請求項 3】

請求項 1 または請求項 2 記載の会議装置であって、
当該会議装置を使用する自参加者の音声データを取得する音声取得手段と、
前記自参加者の音声データから相槌を検出する相槌検出手段と、さらに有し、
前記発話欲求レベル算出手段は、前記動作検出手段が検出した動作および前記相槌検出
手段が検出した相槌に基づいて発話欲求レベルを算出すること
を特徴とする会議装置。

【請求項 4】

ネットワークを介して遠隔会議を行う会議方法であって、
会議装置は、
当該会議装置を使用する自参加者の映像データを取得する映像取得ステップと、
前記自参加者の映像データから発話欲求があると想定される所定の動作を検出する動作
検出ステップと、
前記動作検出ステップで検出した動作に基づいて発話欲求レベルを算出する発話欲求レ
ベル算出ステップと、
ネットワークを介して他の会議装置から他参加者の映像データおよび発話欲求レベルを
受信する受信ステップと、
自参加者および他参加者の映像データを配置した会議映像を生成する映像生成ステッ
プと、
自参加者および他参加者の発話欲求レベルを示す情報を、前記会議映像に設定する映像
編集ステップと、
前記映像編集ステップで編集した会議映像を表示する表示ステップと、を行うこと
を特徴とする会議方法。

30

40

【請求項 5】

請求項 4 記載の会議方法であって、
前記映像編集ステップは、発話欲求レベルが所定の値以上の参加者の映像データの近傍
に、所定のマークを設定すること
を特徴とする会議方法。

【請求項 6】

請求項 4 または請求項 5 記載の会議方法であって、

50

当該会議装置を使用する自参加者の音声データを取得する音声取得ステップと、前記自参加者の音声データから相槌を検出する相槌検出ステップと、さらに行い、前記発話欲求レベル算出ステップは、前記動作検出ステップで検出した動作および前記相槌検出ステップで検出した相槌に基づいて発話欲求レベルを算出すること
を特徴とする会議方法。

【請求項 7】

請求項 4 から請求項 6 のいずれか 1 項に記載の会議方法をコンピュータに実行させるための会議プログラム。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は、ネットワークを介した遠隔会議を行う会議装置、会議方法および会議プログラムに関する。

【背景技術】

【0002】

環境問題（エコロジー）や、企業活動における出張費を抑えるなどの目的から、遠隔会議の需要が高まっている。従来の会議システムには、例えば、電話（音声）会議システムや、ブラウザ上で動作する Web（映像）会議システムなどがある。Web 会議システムでは、ブラウザを起動し、Web カメラを設置し、ヘッドセットを装着すれば、互いの参加者の映像をみながら会議を行うことができる。例えば、特許文献 1 には、発言権の取得
や移譲が可能な多地点通信システムが記載されている。

20

【先行技術文献】

【特許文献】

【0003】

【特許文献 1】特開 2004-248145 号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

このような会議システムでは、映像や音声の質が低いという欠点がある。例えば、映像の解像度は低く、映像表示サイズは小型ディスプレイサイズにより制限され、映像と音声に伝送には遅延が生じる。このため、遠隔に存在する他の会議参加者の表情や仕草を読み取ることが困難になる。

30

【0005】

これに起因して発生する問題の 1 つとして、誰が次に発言しそうなかがわからず、同時に複数の会議参加者が発話を開始する発話の衝突が発生してしまう。発話の衝突が頻発する会議では、会議参加者に精神的なストレスが蓄積するとともに、会議の進行を停滞させてしまう。

【0006】

特許文献 1 の技術では、発言権を取得するための意識的な行動が必要であり、会議の円滑な進行が妨げられる可能性がある。対面コミュニケーションでは、発話したい旨を明確に意思表示して発言する場合と、自然な仕草（非言語情報のやりとり）から周囲に発話欲求を伝達して発言する場合とがあるが、特許文献 1 では後者の場合については考慮されていない。

40

【0007】

本発明は、上記事情に鑑みてなされたものであり、本発明の目的は、参加者の表情や仕草が読み取れない場合であっても、各参加者の発話欲求を容易に認識させる会議装置、会議方法および会議プログラムを提供することにある。

【課題を解決するための手段】

【0008】

上記目的を達成するため、本発明は、ネットワークを介して遠隔会議を行う会議装置で

50

あって、当該会議装置を使用する自参加者の映像データを取得する映像取得手段と、前記自参加者の映像データから発話欲求があると想定される所定の動作を検出する動作検出手段と、前記動作検出手段が検出した動作に基づいて発話欲求レベルを算出する発話欲求レベル算出手段と、ネットワークを介して他の会議装置から他参加者の映像データおよび発話欲求レベルを受信する受信手段と、自参加者および他参加者の映像データを配置した会議映像を生成する映像生成手段と、自参加者および他参加者の発話欲求レベルを示す情報を、前記会議映像に設定する映像編集手段と、前記映像編集手段が編集した会議映像を表示する表示手段と、を有する。

【0009】

本発明は、ネットワークを介して遠隔会議を行う会議方法であって、会議装置は、当該会議装置を使用する自参加者の映像データを取得する映像取得ステップと、前記自参加者の映像データから発話欲求があると想定される所定の動作を検出する動作検出ステップと、前記動作検出ステップで検出した動作に基づいて発話欲求レベルを算出する発話欲求レベル算出ステップと、ネットワークを介して他の会議装置から他参加者の映像データおよび発話欲求レベルを受信する受信ステップと、自参加者および他参加者の映像データを配置した会議映像を生成する映像生成ステップと、自参加者および他参加者の発話欲求レベルを示す情報を、前記会議映像に設定する映像編集ステップと、前記映像編集ステップで編集した会議映像を表示する表示ステップと、を行う。

【0010】

本発明は、前記会議方法をコンピュータに実行させるための会議プログラムである。

【発明の効果】

【0011】

本発明によれば、参加者の表情や仕草が読み取れない場合であっても、各参加者の発話欲求を容易に認識させる会議装置、会議方法および会議プログラムを提供することができる。

【図面の簡単な説明】

【0012】

【図1】本発明の実施形態に係る会議システムの全体構成図である。

【図2】本実施形態の動作を示すフローチャートである。

【図3】会議映像の一例を示す図である。

【発明を実施するための形態】

【0013】

以下、本発明の実施の形態について、図面を参照して説明する。

【0014】

図1は、本発明の実施形態に係る会議システムの全体構成図である。本実施形態の会議システムは、ネットワーク9に接続された複数の会議装置1を用いて、遠隔地にいる複数の会議参加者（以下、「参加者」という）が遠隔会議を行うためのシステムである。

【0015】

図示する会議システムは、複数の参加者の各々が使用する複数の会議装置1と、これらの会議装置1とネットワーク9を介して接続された会議サーバ8とを備える。

【0016】

会議装置1は、例えばPCなどであって、当該会議装置1を使用する自参加者の映像を撮影するビデオカメラ11と、自参加者が発言した音声を集音するマイク12と、会議映像を表示する表示装置（ディスプレイ）13と、会議サーバ8から送信された他の参加者の音声を出力するスピーカ14とを備える。

【0017】

また、会議装置1は、映像取得部21と、動作検出部22と、発話欲求レベル算出部23と、メモリ24と、音声取得部25と、相槌検出部26と、映像生成部27と、映像編集部28と、通信部29とを備える。

【0018】

10

20

30

40

50

映像取得部 2 1 は、ビデオカメラ 1 1 が撮像した映像データを取り込む。動作検出部 2 2 は、自参加者の映像データから発話欲求があると想定される所定の動作を検出する。

【0019】

音声取得部 2 5 は、マイク 1 2 から集音された音声データを取り込む。相槌検出部 2 6 は、自参加者の音声データから相槌を検出する。

【0020】

発話欲求レベル算出手段 2 3 は、動作検出部 2 2 が検出した動作および相槌検出部 2 6 が検出した相槌に基づいて発話欲求レベルを算出し、メモリ 2 4 に記憶する。

【0021】

通信部 2 9 は、映像取得部 2 1 により取り込まれた映像データを、ネットワーク 9 を介して会議サーバ 8 に送信するとともに、映像生成部 2 7 に送出する。また、通信部 2 9 は、音声取得部 2 5 により取り込まれた音声データをネットワーク 9 を介して会議サーバ 8 に送信する。また、通信部 2 9 は、他の参加者の音声データを、ネットワーク 9 を介して会議サーバ 8 から受信し、スピーカ 1 4 に出力・再生するとともに、他の参加者の映像データをネットワーク 9 を介して会議サーバ 8 から受信し、映像生成部 2 7 に送出する。また、通信部 2 9 は、メモリ 2 4 に記憶された情報を、ネットワーク 9 および会議サーバ 8 を介して、他の会議装置 1 と送受信する。

【0022】

映像生成部 2 7 は、映像取得部 2 1 が取り込んだ自参加者の映像データと、通信部 2 9 が受信した他参加者の映像データとを配置した会議映像を生成する。映像編集部 2 8 は、自参加者および他参加者の発話欲求レベルを示す情報を、映像生成部 2 7 が生成した会議映像に設定し、表示装置 1 3 に表示する。

【0023】

会議サーバ 8 は、各参加者が使用する各会議装置 1 から入力される、当該参加者の映像データ、音声データおよびメモリ 2 4 の情報をそれぞれ受信し、他の会議装置 1 に送信する。

【0024】

会議装置 1 および会議サーバ 8 は、例えば、CPU と、メモリ と、HDD 等の外部記憶装置と、入力装置と、出力装置とを備えた汎用的なコンピュータシステムを用いることができる。このコンピュータシステムにおいて、CPU がメモリ上にロードされた所定のプログラムを実行することにより、各装置の各機能が実現される。例えば、会議装置 1 および会議サーバ 8 およびの各機能は、会議装置 1 用のプログラムの場合には会議装置 1 の CPU が、そして、会議サーバ 8 用のプログラムの場合には会議サーバ 8 の CPU が、それぞれ実行することにより実現される。

【0025】

また、会議装置 1 用のプログラムおよび会議サーバ 8 用のプログラムは、ハードディスク、フレキシブルディスク、CD-ROM、MO、DVD-ROM などのコンピュータ読取り可能な記録媒体に記憶することも、ネットワークを介して配信することもできる。

【0026】

次に、本実施形態の動作について説明する。

【0027】

各参加者は、それぞれ、会議装置 1 に前に座り、ネットワークを介した遠隔会議を行う。

【0028】

映像取得部 2 1 は、ビデオカメラ 1 1 で撮像された自参加者の映像データを取り込み、動作検出部 2 2 に送出する。

【0029】

動作検出部 2 2 は、映像データから自参加者の所定の動作を検出する。ここで検出される所定の動作は、例えば、「手を口元へ動かす」、「挙手」、「頷き」、「体を横に動かす」など、発話欲求があると想定される動作である。

10

20

30

40

50

【0030】

動作検出部22は、映像データを画像処理することで、このような動作を検出する。例えば、「頷き」動作は、顔向き検出を行い、顔の向きが上下に動いた場合に、「頷き」と判別する。「手を口元へ動かす」動作は、例えば、顔領域と隣接していない肌色の領域を手の領域であると識別し、この手の領域が顔領域と重なった場合、「手を口元へ動かす」と判別する。

【0031】

「挙手」動作は、例えば、顔領域と隣接していない肌色の領域を手の領域であると識別し、この手の領域の重心が画像（映像データ）のy軸方向に、所定の高さ（閾値）を超えた場合、「挙手」と判別する。「体を横に動かす」動作は、例えば顔検出を行い、検出した顔領域の中心点が、単位時間当たりの所定の閾値を越えて、画像のx軸方向へ移動した場合、「体を横に動かす」と判別する。

10

【0032】

音声取得部25は、マイク12から入力された自参加者の音声を集音し、相槌検出部26に送出する。相槌検出部26は、音声データが入力され、かつ、自参加者が話者でない場合は、入力された音声データは全て「相槌」とであると判別する。例えば、各会議装置は、自参加者が話者であるかどうかを判別するための話者フラグがメモリ24に記憶される。話者フラグの初期値は「0」（非話者）であり、これが「1」（話者）である場合、その自参加者は話者であると判別される。

【0033】

20

会議サーバ8では、全会議装置（全参加者）の話者フラグを管理するとともに、誰が現在の話者であるかを決定する。例えば、全ての参加者の話者フラグが「0」であるときに、ある会議装置から音声データが入力された場合、この入力を行った会議装置の参加者を話者に指定し、当該参加者の話者フラグを「1」にする。すなわち、会議サーバ8は、当該会議サーバ8内の話者フラグを更新するとともに、話者に指定した会議装置に話者指定通知を送信し、当該会議装置のメモリ24の話者フラグを「0」から「1」に更新させる。なお、当該参加者の音声データ入力がなくなった時点で、話者の指定を解除し、話者フラグを「1」から「0」に更新する。

【0034】

また、相槌検出部26は、メモリ24の自参加者の話者フラグが「0」（非話者）であって、自参加者以外の参加者の話者フラグが「1」であるときに、音声データが入力された場合、これを「相槌」とであると判別する。このとき、この参加者の話者フラグは「0」のままである。相槌検出部26は、音声データが入力された場合、会議サーバ8に話者フラグが「1」の他の参加者が存在するか否かを問い合わせ、話者フラグが「1」の他の参加者が存在する場合、「相槌」とであると判別する。

30

【0035】

なお、相槌検出部26は、メモリ24の自参加者の話者フラグが「1」（話者）のときに音声データが入力された場合は、会議サーバ8への問い合わせを行わず、「相槌」とでないと判別する。このとき、メモリ24の話者フラグは「1」のままである。

【0036】

40

そして、動作検出部22で検出された動作の情報、および相槌検出部26が検出した相槌の情報は、発話欲求レベル算出部23に送出される。

【0037】

図2は、発話欲求レベル算出部23の処理を示すフローチャートである。発話欲求レベル算出部23は、所定のタイミングで繰り返し図2に示す処理を行う。

【0038】

発話欲求レベル算出部23は、動作検出部22が、「手を口元へ動かす」、「頷き」、「体を横に動かす」のいずれかの動作を検出した場合（S11：YES）、図示しない記憶部のカウンタの値に「1」を加算する（S12）。なお、カウンタの初期値は「0」である。

50

【0039】

また、発話欲求レベル算出部23は、動作検出部22が「挙手」の動作を検出した場合（S13：YES）、カウンタの値に「5」を加算する（S14）。

【0040】

また、発話欲求レベル算出部23は、相槌検出部26が「相槌」を検出した場合（S15：YES）、カウンタの値に「3」を加算する（S16）。

【0041】

そして、発話欲求レベル算出部23は、S11からS16で加算されたカウンタの値にもとづいて、発話欲求レベルを算出する。例えば、発話欲求レベルが、レベル0～レベル5までの6つのレベルに分かれている場合、カウンタの値をいずれかのレベルに変換し、メモリ24に記憶する（S17）。 10

【0042】

そして、発話欲求レベル算出部23は、変換したレベル値が所定の値（例えば「5」）以上の場合（S18：YES）、所定のマークの表示を決定し、マーク表示情報をメモリ24に記憶する（S19）。なお、所定のマークには、図形、記号、文字などを用いることができる。例えば、挙手を連想させる挙手マークを、所定のマークとして用いることとしてもよい。

図2で説明したように、メモリ24には、自参加者の発話欲求レベル、および、発話欲求レベルが「5」以上の場合にはマーク表示情報が記憶される。この自参加者の発話欲求レベルおよびマーク表示情報は、通信部29によりネットワーク9を介して他の全ての会議装置1に送信される。また、他の全ての会議装置1からネットワーク9を介して送信された他参加者の発話欲求レベルおよびマーク表示情報が通信部29により受信され、メモリ24に記憶される。このように、各参加者の発話欲求レベルおよびマーク表示情報は、全ての会議装置1で共有される。 20

【0043】

映像生成部27および映像編集部28は、ビデオカメラ11が撮像した自参加者の映像データと、ネットワーク9を介して会議サーバ8から受信した他参加者の映像データと、メモリ24に記憶された各参加者の発話欲求レベルおよびマーク表示情報とを用いて会議映像を生成・編集する。

【0044】

図3は、会議映像の一例を示す図である。映像生成部27は、図示するように、各参加者の映像データをタイル状に所定の位置に配置する。図示する例では、4人の参加者が会議に参加している。 30

【0045】

そして、映像編集部28は、各参加者の発話欲求レベルおよびマーク表示情報が記憶されたメモリ24を参照し、映像生成部27が生成した会議映像を編集する。具体的には、図示するように、発話欲求レベルを示す情報（図示する例では、インジケータ）を、各参加者の発話欲求レベルに応じて、対応する参加者の映像データの近傍に設定する。例えば、発話欲求レベルが「2」の参加者の場合に、下から2つインジケータを点灯するなど、発話欲求レベルに応じてインジケータを点灯する。 40

【0046】

また、マーク表示情報が記憶されている参加者については、当該参加者の映像データの近傍に所定のマーク（図示する例では、挙手マーク）を設定する。表示装置13は、映像編集部28が編集した会議映像を表示する。

【0047】

また、話者が存在する場合（話者フラグが「1」の参加者が存在する場合）、映像編集部28は、例えば話者の映像データを所定の色の枠で囲むなど、当該話者を目立たせるように表示させることとしてもよい。

【0048】

以上説明した本実施形態では、ネットワークを介した遠隔会議において、各参加者の発 50

話欲求レベルを示す情報を表示することで、各参加者の発話欲求を容易に（自然に）認識することができる。そのため、参加者は、発話意思を他の参加者に示すためにマウスやキーボードなどの入力デバイス进行操作することなく、各参加者の発話欲求を汲み取りながら、円滑な会議を進行することができる。例えば、発話の衝突を防止し、スムーズな話者交代を実現することができる。

【0049】

また、本実施形態では、動作検出部22が検出する動作の1つに「挙手」があるため、参加者が発話意思を明示的に示したい場合は、対面での会議と同様に挙手（手を挙げる）動作を行うことで、発話欲求レベルを上げることができる。

【0050】

また、本実施形態では、発話欲求レベルが所定の値以上の参加者の映像データの近傍に、所定のマーク（例えば、挙手マーク）を表示すること、発話欲求の高い参加者を一目で容易に認識させることができる。

【0051】

なお、本発明は上記実施形態に限定されるものではなく、その要旨の範囲内で数々の変形が可能である。例えば、本実施形態では、動作検出部22が検出した動作および相槌検出部26が検出した相槌に基づいて発話欲求レベルを算出することとした。しかしながら、動作検出部22が検出した動作のみに基づいて発話欲求レベルを算出することとしてもよい。また、相槌検出部26が検出した相槌のみに基づいて発話欲求レベルを算出することとしてもよい。

【符号の説明】

【0052】

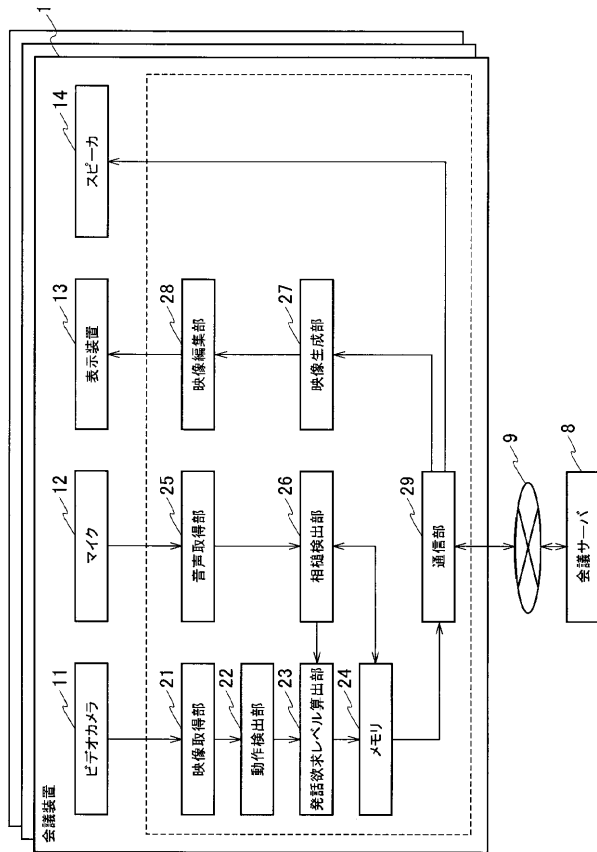
- 1 : 会議装置
- 11 : ビデオカメラ
- 12 : マイク
- 13 : 表示装置
- 14 : スピーカ
- 21 : 映像取得部
- 22 : 動作検出部
- 23 : 発話要求レベル算出部
- 24 : メモリ
- 25 : 音声取得部
- 26 : 相槌検出部
- 27 : 映像生成部
- 28 : 映像編集部
- 29 : 通信部
- 8 : 会議サーバ
- 9 : ネットワーク

10

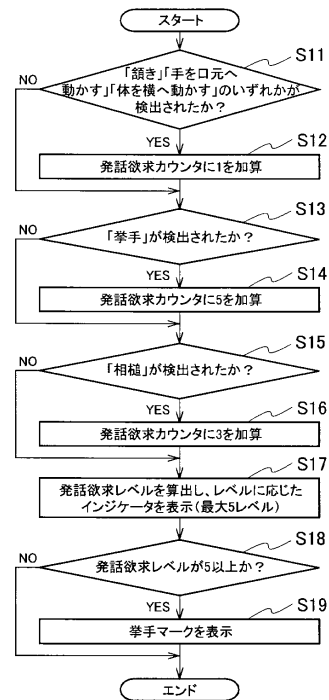
20

30

【図 1】



【図 2】



【図 3】

