(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4849303号
(P4849303)
(45)発行日　平成24年1月11日(2012.1.11)　(24)登録日　平成23年10月28日(2011.10.28)

(51)Int.Cl.　F I
G06N 3/00 (2006.01)　G06N 3/00 550Z
G06N 5/04 (2006.01)　G06N 5/04 550M

請求項の数 5　(全 18 頁)

(21)出願番号　特願2005-243656(P2005-243656)
(22)出願日　平成17年8月25日(2005.8.25)
(65)公開番号　特開2007-58615(P2007-58615A)
(43)公開日　平成19年3月8日(2007.3.8)
審査請求日　平成20年3月27日(2008.3.27)

(73)特許権者　393031586
株式会社国際電気通信基礎技術研究所
京都府相楽郡精華町光台二丁目２番地２
(74)代理人　100099933
弁理士　清水　敏
(72)発明者　ニック・キャンベル
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内

審査官　長谷川　篤男

最終頁に続く

(54)【発明の名称】行動指針決定装置及びコンピュータプログラム

(57)【特許請求の範囲】
【請求項１】
人間が存在する可能性のある場において、周囲の状況から、とるべき行動の指針を決定するための行動指針決定装置であって、
周囲の画像から抽出される人間の動きに関する所定の特徴情報の時系列と、とるべき行動の指針との関係を予め学習した行動指針決定モデルを格納するためのモデル格納手段と、
周囲の動画像から前記所定の特徴情報の時系列を作成し、前記モデル格納手段に格納された行動指針決定モデルを参照する事により、とるべき行動の指針を決定するためのモデル参照手段とを含み、
前記モデル参照手段は、
前記動画像の各フレームから肌色の部分を検出するための肌色検出手段と、
前記動画像に対し、前記肌色検出手段によって検出された肌色の部分の中で、同一人物の顔と手との組合せを識別し、各組合せについて、顔と手とをトラッキングするためのトラッキング手段と、
前記トラッキング手段によりトラッキングされた各人物の顔及び手の動きから、前記所定の特徴情報をフレームごとに作成し、複数フレームに対する前記所定の特徴情報を用い、前記モデル格納手段に格納された前記行動指針決定モデルを参照する事により、行動の指針を決定するための手段とを含む、行動指針決定装置。
【請求項２】

前記行動指針決定装置は、人間の顔と手との動きに関する動き予測モデルを格納するための動き予測モデル格納手段をさらに含み、
前記トラッキング手段は、
前記動画像に対し、前記肌色検出手段によって検出された肌色の部分の中で、同一人物の顔と手との組合せを識別するための手段と、
前記識別するための手段により識別された各組合せについて、前記動き予測モデル格納手段に格納された動作モデルに基づいて予測される、当該組合せに含まれる顔と手との動きに基づいて顔と手とをトラッキングするための手段とを含む、請求項１に記載の行動指針決定装置。
【請求項３】
人間が存在する可能性のある場において、周囲の状況から、とるべき行動の指針を決定するための行動指針決定装置であって、
周囲の画像から抽出される人間の動きに関する所定の特徴情報の時系列と、とるべき行動の指針との関係を予め学習した行動指針決定モデルを格納するためのモデル格納手段と、
周囲の動画像から前記所定の特徴情報の時系列を作成し、前記モデル格納手段に格納された行動指針決定モデルを参照する事により、とるべき行動の指針を決定するためのモデル参照手段とを含み、
前記行動指針決定モデルは、周囲の画像から抽出される人間の動作に関する前記所定の特徴情報と、周囲の音声から得られる非言語的音声情報に基づいて作成される所定の音響特徴情報とを統合した画像・音声統合型の特徴情報の時系列と、とるべき行動の指針との関係を予め学習した音声統合型の行動指針決定モデルを含み、
前記モデル参照手段は、
前記動画像の各フレームから肌色の部分を検出するための肌色検出手段と、
前記動画像に対し、前記肌色検出手段によって検出された肌色の部分の中で、同一人物の顔と手との組合せを識別し、各組合せについて、顔と手とをトラッキングするためのトラッキング手段と、
前記トラッキング手段によりトラッキングされた各人物の顔及び手の動き、並びに、前記音声から、前記所定の特徴情報をフレームごとに作成し、複数フレームに対する前記所定の特徴情報を用い、前記モデル格納手段に格納された前記音声統合型の行動指針決定モデルを参照する事により、行動の指針を決定するための画像・音声統合型モデル参照手段を含む、行動指針決定装置。
【請求項４】
前記周囲の音声は、マイクロフォンで受音され、音声信号に変換され、
前記所定の音響特徴情報は、
前記音声信号に基づいて推定される、発話の有無を示す情報、及び
前記音声信号に基づいて推定される発話の持続時間、
の任意の組合せを含む、請求項３に記載の行動指針決定装置。
【請求項５】
コンピュータにより実行されると、当該コンピュータを請求項１～請求項４のいずれかに記載の行動指針決定装置として動作させる、コンピュータプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
この発明は、音声と画像とから周囲の状況を推定し、それに従って行動指針を決定する装置に関し、特に、人間の音声に関する非言語的情報と、画像から得られる人間の動きに関する情報とを有効に利用して周囲の状況に対する適切な行動を行なうための行動指針を決定するための行動指針決定装置及びコンピュータプログラムに関する。
【背景技術】
【０００２】

近年、生活の更なる簡便性のために家庭用電機製品等のオートメーション化が進んでいる。例えば、人が部屋に入った事に反応して点灯する照明、人が前に立った事に反応して蓋が開く電動式の便座等がある。こうした技術により、人々の生活はますます便利なものになっている。

【０００３】

これらの他に、ユーザの操作なしに所定の処理を行なうテレビジョン受像機システムも存在する。このようなテレビジョン受像機として、特許文献１に記載のものがある。特許文献１に記載のテレビジョン受像機は、ユーザの視聴履歴を蓄積しておき、ユーザがその番組の放送時間帯と同じ時間帯に別の番組を見ている事が判明すると、同じ時間帯に好みの番組が別チャンネルで放送されている事をユーザに通知する。この事により、ユーザは当該好みの番組を見逃す事なく確実に視聴できる。

【０００４】

この様な技術をさらに進歩させると、ユーザの視聴履歴に基づき、ユーザの視聴パターンに従って決定される時間に、ユーザの視聴パターンに適合したチャンネルに合わせて自動的に電源の投入又は切断をする自動制御テレビジョン受像機が考えられる。

【０００５】

例えば月曜日から金曜日まで毎日夜１０時から１１時まで決まったチャンネルの番組を見るユーザがいるものとする。通常のテレビジョン受像機であれば、ユーザは番組を見るために電源を入れて、チャンネルを合わせなければならない。しかし、自動制御テレビジョン受像機の場合には、ユーザの過去の番組視聴歴から「月曜日～金曜日の夜１０時～１１時には○チャンネルを見る」という情報が自動的に得られる。そして、その視聴履歴を元にしてテレビジョン受像機の電源を自動的に入れたり切ったりできる上に、チャンネル合わせも自動的に行なう事ができる。それゆえ、例えば火曜日の夜１０時にユーザがいつも見ている番組を見るために、自らテレビジョン受像機の電源を入れて該当チャンネルに合わせる必要はない。毎週火曜日の夜の１０時になれば、テレビジョン受像機の電源が自動的に入り、チャンネルが該当チャンネルに自動的に合わされるからである。

【特許文献１】特開２００５－０３９６８１

【発明の開示】

【発明が解決しようとする課題】

【０００６】

この様に、自動制御テレビジョン受像機においては、ユーザの視聴パターンに合わせて、自動的にテレビジョン受像機がついたり消えたりする。それゆえ、ユーザが習慣的に見る事にしている番組を見逃す可能性は格段に減ると思われる。しかし、視聴パターンのみでテレビジョン受像機の電源のオンオフを行なうと、場合によっては様々な問題が生じる。

【０００７】

例えば、部屋の中にいる人間が緊急かつ深刻な会話をしている最中に、テレビジョン受像機の電源が入ると、結局、すぐにテレビジョン受像機の電源が切られる事になる可能性が高い。このような場合には、再度テレビジョン受像機を切るという作業を無駄に行なわなければならない。さらに、会話に加わっていた人の気分を害するという可能性も考えられる。これは、テレビジョン受像機が、部屋の中の雰囲気を感じ取る事ができないために起こる問題である。

【０００８】

場の雰囲気と無関係に会話をさえぎる形で何らかの行動をするために、会話に加わっていた人間の気分を害するという状況が考えられるのは、自動制御テレビジョン受像機の使用時に限定されない。これは、例えば多くの人間のいる場で何らかのサービスをするロボットの場合にも起こり得る。

【０００９】

例えば、大勢の人間が集まっている場で、皆が楽しめる様に自律的に行動するロボットを考える。このロボットは例えば、誰とも会話せずに孤立している人間を見つけると、話

しかけるという機能を持っているとする。しかし、ある参加者が孤立しているか否かを判定する際に、単に会話をしているか否かのみを参考にすると問題が生じる。

【００１０】

具体的には、例として３人の人間が会話をしている場面を想定する。このうち２名が対話を行なっていて３人目の人間は言葉を発さずに、傍らに立っているだけであるとする。このとき、３人目の人間が会話に参加せずに孤立していると判定したロボットがその人間に話しかけると、不都合が生じる場合がある。なぜならば、その人間は積極的に対話の輪には入っていない様に見えたとしても、対話している２人の会話を聞いているかもしれないからである。

【００１１】

複数人が会話をしている際にはこのような状況は一般によく見られる。つまり、グループ中の一部の人間のみが対話を行ない、その他の人間は明確な言語を発せず、一見会話に加わっていない様に見えるが、そのような人間でも確かに話を聞いており、会話に参加しているという状況である。

【００１２】

このような状況は一般的であるにも関わらず、ロボットがこの状況に適切に対応できず、いわば無頓着にその場に介入してしまうとすれば、それは問題である。一方、人間の場合には、種々の情報から判断してそのような行動をとるべきかつつしむべきかについて、適切な判断を行なう事ができる。人間と係わり合いを持つロボットについても、そのような能力を備える事が望ましい。しかも、人間との係わり合いを持つという観点から、こうした能力は、リアルタイムの動作に適用される事が望ましい。

【００１３】

そこで、本発明は、その場の空気を読んで、どのような動作をすべきかを適切に判断して行動指針を決定する事ができる行動指針決定装置及びコンピュータプログラムを提供する事を目的とする。

【００１４】

この発明の他の目的は、その場の人間の間の空気を読んで、どのような動作をすべきかを適切に判断して行動指針を決定する事ができる行動指針決定装置及びコンピュータプログラムを提供する事である。

【００１５】

この発明の他の目的は、その場の人間の間の空気を読んで、どのような動作をすべきかを適切に、かつリアルタイムで判断して行動指針を決定する事ができる行動指針決定装置及びコンピュータプログラムを提供する事である。

【課題を解決するための手段】

【００１６】

本発明の第１の局面に係る行動指針決定装置は、人間が存在する可能性のある場において、周囲の状況から、とるべき行動の指針を決定するための行動指針決定装置であって、周囲の画像から抽出される人間の動きに関する所定の特徴情報の時系列と、とるべき行動の指針との関係を予め学習した行動指針決定モデルを格納するためのモデル格納手段と、周囲の動画像から所定の特徴情報の時系列を作成し、モデル格納手段に格納された行動指針決定モデルを参照する事により、とるべき行動の指針を決定するためのモデル参照手段とを含む。

【００１７】

この行動指針決定装置によると、人間の動きから抽出される所定の特徴量の時系列によってモデルを参照する事により、予め学習されたルールに従って、とるべき行動指針を決定する。それゆえ、自動制御により動作する装置が、人間の動きを参照して行動指針を決定し、それに従い動作を行なう事ができる。従って、その場の雰囲気に応じた適切な行動をとる事ができ、場に相応しくない行動をとったり、無駄な動作を行ったりする事が防止できる。その結果、その場の空気を読んで、どのような動作をすべきかを適切に判断して行動指針を決定する事ができる行動指針決定装置を提供できる。

【００１８】
好ましくは、モデル参照手段は、動画像の各フレームから肌色の部分を検出するための肌色検出手段と、動画像に対し、肌色検出手段によって検出された肌色の部分の中で、同一人物の顔と手との組合せを識別し、各組合せについて、顔と手とをトラッキングするためのトラッキング手段と、トラッキング手段によりトラッキングされた各人物の顔及び手の動きから、所定の特徴情報をフレームごとに作成し、複数フレームに対する所定の特徴情報を用い、モデル格納手段に格納された行動指針決定モデルを参照する事により、行動の指針を決定するための手段とを含む。
【００１９】
この行動指針決定装置によると、トラッキングされた各人の顔及び手の動きを参照し、とるべき行動指針を決定するための特徴情報を抽出する事ができる。ある人の感情、発話時の意図、会話への参加状況などは、顔及び手の位置を時系列に追う事で得られるそれらの動きから推測できる。ある人物のそのときの感情、発話意図、会話への参加状況などを考慮して行動指針を決定できる。その結果、その場の人間の間の空気を読んで、どのような動作をすべきかを適切に判断して行動指針を決定する事ができる行動指針決定装置を提供できる。
【００２０】
さらに好ましくは、行動指針決定装置は、人間の顔と手との動きに関する動き予測モデルを格納するための動き予測モデル格納手段をさらに含み、トラッキング手段は、動画像に対し、肌色検出手段によって検出された肌色の部分の中で、同一人物の顔と手との組合せを識別するための手段と、識別するための手段により識別された各組合せについて、動き予測モデル格納手段に格納された動き予測モデルに基づいて予測される、当該組合せに含まれる顔と手との動きに基づいて顔と手とをトラッキングするための手段とを含む。
【００２１】
この行動指針決定装置によると、動き予測モデルに基づいて予測される組合せに含まれる顔と手との動きを参照して顔と手とをトラッキングする。予測に従ってトラッキングを行なう事により、適切で迅速なトラッキング処理が行なわれる様になる。その結果、その場の人間の間の空気を読んで、どのような動作をすべきかを適切に、かつリアルタイムで判断して行動指針を決定する事ができる行動指針決定装置を提供できる。
【００２２】
さらに好ましくは、行動指針決定装置はさらに、肌色検出手段により検出されたトラッキングするための手段によりトラッキングされた顔の各々の向いている方向を検出するための顔方向検出手段を含み、所定の特徴情報は、各フレームにおける各組合せの顔の位置及び向きと、手の位置とを示す情報を含む。
【００２３】
この行動指針決定装置によると、顔の向いている方向を検出する事ができる。言葉を発していない人物がいても、顔の向きからも会話に加わっているか否かを判定できる。また、それらの時系列情報によって、顔の上下動（うなづき）、左右への往復運動（かぶりをふる事）など、各人の感情、判断などを表わす情報が得られる。従って、その場の人間の間の空気を読んで、どのような動作をすべきかを適切に、かつリアルタイムで判断して行動指針を決定する事ができる行動指針決定装置を提供できる。
【００２４】
さらに好ましくは、行動指針決定モデルは、周囲の画像から抽出される人間の動きに関する所定の特徴情報と、周囲の音声から得られる非言語的音声情報に基づいて作成される所定の音響特徴情報とを統合した画像・音声統合型の特徴情報の時系列と、とるべき行動の指針との関係を予め学習した音声統合型の行動指針決定モデルを含み、モデル参照手段は、周囲の動画像及び音声から画像・音声統合型の特徴情報を作成し、画像・音声統合型の特徴情報の時系列を用いてモデル格納手段に格納された音声統合型の行動指針決定モデルを参照する事により、とるべき行動の指針を決定するための画像・音声統合型モデル参照手段を含む。

【００２５】
この行動指針決定装置によると、モデルを参照し、人間の動きと周囲の音声とを参照してとるべき行動指針を決定できる。動きだけではなく、音声も統合する事により、周囲の状況を判断するための情報がより多くなる。また、音声と動きとを統合する事により、動きのみから、又は音声のみからは得られなかった情報も得られる。従って、周囲の状況に関し、より適切な判断を行なう事ができる。その結果、その場の人間の間の空気を読んで、どのような動作をすべきかを適切に、かつリアルタイムで判断して行動指針を決定する事ができる行動指針決定装置を提供できる。
【００２６】
さらに好ましくは、周囲の音声は、マイクロフォンで受音され、音声信号に変換され、所定の音響特徴情報は、音声信号に基づいて推定される、発話の有無を示す情報、及び音声信号に基づいて推定される発話の持続時間、の任意の組合せを含む。
【００２７】
この行動指針決定装置によると、音響特徴情報は、上記した任意の情報の組合せからなっている。発話の有無、発話の持続時間などは、その場で会話が行われているか否か、行なわれているとしてそれが活発か否か、等を表わす。それゆえ、それらの情報を適切に組合せる事により、その場の雰囲気を的確にとらえる事ができる。その結果、その場の人間の間の空気を読んで、どのような動作をすべきかを適切に、かつリアルタイムで判断して行動指針を決定する事ができる行動指針決定装置を提供できる。
【００２８】
さらに好ましくは、周囲の音声は、指向性を有する複数のマイクロフォンで別々に受音されて複数の音声信号に変換され、所定の音響特徴情報は、複数の音声信号に基づいて推定される発話方位を示す情報、複数の音声信号に基づいて推定される、各方位における発話の有無を示す情報、及び複数の音声信号に基づいて推定される、各方位における発話の持続時間、の任意の組合せを含む。
【００２９】
この行動指針決定装置によると、音声の生じた方向である発話方位、各発話方位で発話があったか否か、又は各発話方位での発話時間を音響特徴量の要素の一つとして得る事ができる。これらと画像情報とを総合する事で、その場にいる複数人の発話状況を区別する事ができる。その結果、複数人が存在する場の空気を読んで、どのような動作をすべきかを適切に、かつリアルタイムで判断して行動指針を決定する事ができる行動指針決定装置を提供できる。
【００３０】
さらに好ましくは、画像・音声統合型の特徴情報は、発話者別に推定される発話の有無を示す情報、発話者別に推定される発話の持続時間、発話者別に推定される発話音声の大きさ、及び周囲の人間の間に、所定の関係があると推定されるか否かに関する情報、の任意の組合せを含む。
【００３１】
この行動指針決定装置によると、その場にいる人間に関し、各発話者の発話に関する情報と、発話者間の関係に関する情報とを用いて行動指針を決定できる。それゆえ、音声のみもしくは動きのみから特徴情報を抽出するよりも広い範囲の特徴情報を得る事ができる。その結果、行動指針決定をさらに適切に行なう事ができる。
【００３２】
本発明の第２の局面に係る行動指針決定装置は、人間が存在する可能性のある場において、周囲の状況から、とるべき行動の指針を決定するための行動指針決定装置であって、周囲の音声から抽出される人間の発する音声に関する所定の音響特徴情報の時系列と、とるべき行動の指針との関係を予め学習した行動指針決定モデルを格納するためのモデル格納手段と、周囲の音声から所定の音響特徴情報の時系列を作成し、モデル格納手段に格納された行動指針決定モデルを参照する事により、とるべき行動の指針を決定するためのモデル参照手段とを含む。

【００３３】
この行動指針決定装置によると、モデルを参照し、周囲の音声を参照してとるべき行動指針を決定する情報を抽出する事ができる。それゆえ、自動制御により動作する装置が、音を参照して行動指針を決定し、それに従い動作を行なう事ができる。その結果、その場にいる人間の気分を害する事がなくなったり、自動制御にあたって無駄な動作を省く事ができる様になったりする。
【００３４】
好ましくは、周囲の音声は、マイクロフォンで受音して音声信号に変換され、所定の音響特徴情報は、単数の音声信号に基づいて推定される、発話の有無を示す情報、及び単数の音声信号に基づいて推定される発話の持続時間、の任意の組合せを含む。
【００３５】
この行動指針決定装置によると、音響特徴情報は、任意の情報の組合せからなっている。それゆえ、それらの情報を適切に組合せる事により、必要な行動指針決定情報を得る事ができる。その結果、適切な行動指針に基づいて、装置の自動制御がされる。
【００３６】
さらに好ましくは、周囲の音声は、指向性を有する複数のマイクロフォンで別々に受音されて複数の音声信号に変換され、所定の音響特徴情報は、複数の音声信号に基づいて推定される発話方位を示す情報、複数の音声信号に基づいて推定される、発話の有無を示す情報、及び複数の音声信号に基づいて推定される発話の持続時間、の任意の組合せを含む。
【００３７】
この行動指針決定装置によると、音声の生じた方向である発話方位を音響特徴量の要素の一つとして得る事ができる。それゆえ、発話方位を参照して複数人の音響情報を区別する事ができる。その結果、複数人の反応を適切に予測して行動指針を決定する事ができる。
【００３８】
本発明の第３の局面に係るコンピュータプログラムは、コンピュータにより実行されると、当該コンピュータを上記したいずれかの行動指針決定装置として動作させるものである。
【発明を実施するための最良の形態】
【００３９】
以下、図面を参照し、本発明の実施の形態を説明する。本実施の形態は、視聴履歴に基づいて電源のオンオフやチャンネル合わせを制御するテレビジョン受像機において、周囲の状況から得られる非言語音声情報及び動き情報に基づいて、ユーザの視聴パターンにあった番組の放送時に、電源を自動的にオンするか否かに関する行動指針を決定する装置に関するものである。
【００４０】
＜構成＞
図１に、本実施の形態に係る自動制御テレビジョン受像機に搭載された行動指針決定装置４２と自動制御テレビジョン受像機システムとについての機能ブロック図を示す。
【００４１】
図１を参照して、このテレビジョン受像機システムは、図には示さないテレビジョン受像機本体に取付けられたスイッチ又はリモートコントロールパネルを用いたユーザの操作に応じて、ユーザの視聴履歴を取得する視聴履歴取得部３０と、視聴履歴の取得に使用するための時間情報をカウントするタイマ３２と、視聴履歴取得部３０によって取得された視聴履歴を記憶するための視聴履歴記憶部３４と、視聴履歴記憶部３４に記憶された視聴履歴からユーザの視聴パターンを学習するための視聴パターン学習部３６と、視聴パターン学習部３６で学習された視聴パターンを記憶するための視聴パターン記憶部３８と、視聴パターン記憶部３８に記憶された視聴パターンを参照してテレビジョン受像機のチャンネル合わせをするための信号及び電源のオン又はオフを制御する信号を出力するためのテ

レビ制御信号出力部４０とを含む。

【００４２】

このテレビジョン受像機システムはさらに、テレビジョン受像機のある部屋の全景を撮影できるよう設置されたカメラ４４と、テレビジョン受像機のある部屋の音声がどの方向から生じたのかが明確になるような方法で収録できる指向性のある複数のマイクロフォン４６と、カメラ４４で録画された画像とマイクロフォン４６で収録された音とを使用して、当該テレビジョン受像機の設置してある部屋の中にいる人が発する音声に付随する非言語音声情報と、人物の頭部及び両手の動きに関する情報とを同期させ両者の関係を抽出して、人間の反応を予測して、テレビジョン受像機の電源をオンすべきか否かに関する行動指針を示す情報を抽出する行動指針決定装置４２とを含む。

【００４３】

ここで、非言語音声情報とは、人間の発する音声のうちから話の内容に関する情報を除いたものである。つまり、音声の生じている方位である発話方位、音声の有無、音声の継続している時間の長短といったものである。

【００４４】

図２に、行動指針決定装置４２の詳細を示す。図２を参照して、この行動指針決定装置４２は、カメラ４４によって録画された部屋の全景の画像をフレーム単位で格納する画像格納部６０と、画像格納部６０に格納された画像の各フレームから肌色の部分をかたまりでかつリアルタイムで検出する肌色検出部６２と、肌色検出部６２によって検出された肌色のかたまりの中で、一般的に上位にあり大きいものを顔とし下位にあり小さいものを手として区別し、それぞれに予め決められたルールに則って識別番号を付与する肌色部位区別部６４と、人間の顔と手との動きに関する動き予測モデルを格納するための動き予測モデル格納部６８と、動き予測モデル格納部６８に格納された動き予測モデルを用い、肌色部位区別部６４によって区別され識別番号を付与された顔及び手の動きを１フレームずつトラッキングする肌色部位トラッキング部６６とを含む。

【００４５】

ここで、肌色を検出するとは、人間の皮膚の色であると考えられる肌色を、画像の他の部分から分離する処理を行なう事である。この処理を行なう事によって、肌色の部分がいくつかのかたまりとして画像の他の部分から区別される。一般的に、１人の人間の画像に対してこの処理を行なうと、顔の部分が一つの肌色のかたまり、両手の部分が二つの肌色のかたまりとして検出される。

【００４６】

この行動指針決定装置４２はさらに、マイクロフォン４６によって収録された、部屋で生じる様々な音声をフレーム化して格納するための音声格納部７０と、音声格納部７０に格納された音声から、人間の発する音の有無、音声の生じた方位、及び、音声が継続する時間の長短に関する音響特徴情報をフレームごとに抽出処理するための音声関連処理部７２と、音声関連処理部７２で処理された音声関連の情報と肌色部位トラッキング部６６でトラッキング処理された動作情報とを同期させ、所定時間ごとに一組のデータとしてフレーム化して画像・音声の特徴情報を統合するための音声－動作統合部７４とを含む。

【００４７】

行動指針決定装置４２はさらに、画像・音声統合型の特徴情報ととるべき行動の指針との関係を予め学習したマッチング用モデルを格納するためのマッチング用モデル格納部７８と、音声－動作統合部７４によりフレーム化された画像・音声統合型の特徴情報を、所定フレームだけ先入れ先出し方式で一時的に蓄積するためのバッファ８０と、バッファ８０によって蓄積された動作・音声情報のうち、最新の複数フレーム分を使用して、マッチング用モデル格納部７８に格納されたモデルを参照する事により、周囲の人間により発生する音声及び画像から得られた顔及び手の動きに関する情報を用い、発話内容に依存せずに場の状況について判断し、それに従ってテレビジョン受像機の電源をオンすべきかすべきでないかという行動に関する行動指針をＹＥＳ／ＮＯの形で決定し、テレビジョン受像機に与えるためのモデル参照部７６とを含む。

【００４８】
図３に、カメラ４４で撮影された画像の一例を示す。図３を参照して、この例は、部屋の中にいる人が会話をするためにテーブルについている場合を撮影した例である。この撮影にあたっては、テーブルの中央に全方位カメラを設置している。このカメラの解像度はそれほど高くないので、個々の人物の視線の動きを明らかにする事はできない。しかし、上記図２の説明で触れた様に、後の肌色部位トラッキング部６６での処理では顔及び手の動きを明らかにする必要があるのみで、視線の動きを明らかにする必要はない。それゆえ、この程度の解像度を持つカメラで撮影された映像で十分である。
【００４９】
図４に、肌色検出部６２で肌色検出処理をされた画像の一例を示す。図４を参照して、この例では一人の人物を例にとって説明する。ここで、肌色の領域９０は顔を示し、肌色の領域９２は左手、肌色の領域９４は右手をそれぞれ示す。この例からわかる様に、一般的に顔は手よりも上位にあり、かつ、大きい事が多い。また、顔は一つであるのに対し、一般的に手は二つある。そこでこの上下関係と大小関係と数的関係とを用いて肌色部位区別部６４での部位区別処理が行なわれる。
【００５０】
図５に、マッチング用モデル格納部７８に格納されるマッチング用モデルを作成する際に使用される手動ラベリングの方法の一例を示す。図５を参照して、まず、セクション１００は音声関連情報への手動ラベリング結果を示すものである。枠の種類が６種類あるのは、会話への参加者が６人いるからである。ａは参加者Ａ、ｂは参加者Ｂ、ｃは参加者Ｃ、ｄは参加者Ｄ、ｅは参加者Ｅ、及び、ｆは参加者Ｆをそれぞれ表わす。Ａ～Ｆの詳細については後述する。図中に示された他のアルファベットは、音声関連情報の種類を示す。ここで使用されているアルファベットが具体的に何を示しているのかについては後述する。
【００５１】
次に、セクション１０２は動作情報を得るために参照される会話の様子を録画した画像である。この画像を参照しながら次に述べる動作の手動でのトラッキングを行なう。
【００５２】
セクション１０４はこの手動でのトラッキング結果を示すものである。枠が６種類あるのは、セクション１００同様、会話への参加者がＡ～Ｆの６人いるためである。図中に示された他のアルファベットは、音声関連情報の種類を示す。ここで使用されているアルファベットが具体的に何を示しているのかについては後述する。
【００５３】
図６に、セクション１００（図５参照）で示された音声関連情報への手動ラベリング結果を参加者ごとに経時的に配列したものを示す。図６を参照して、ここでの会話への参加者１１０は、Ａ、Ｂ、Ｃ、Ｄ、Ｅ、Ｆの計６名である。
【００５４】
セクション１１６では参加者の性別を示す。ここで、アルファベットｍは男性、ｆは女性をそれぞれ表わす。これによると、Ａは男性、Ｂは女性、Ｃは男性、Ｄは女性、Ｅは女性、Ｆは男性である事がわかる。
【００５５】
また、セクション１１８は参加者の年代を示す。ここで、ｓは年長者、ｊは年少者、ｍはその中間の年代をそれぞれ表わす。これによると、Ａは年長者、Ｂは年少者、Ｃは年少者、Ｄは中間の年代、Ｅは年長者、Ｆは中間の年代である事がわかる。
【００５６】
さらに、セクション１２０は参加者の使用する言語を示す。ここで、ｊは日本語、ｅは英語をそれぞれ表わす。これによると、Ａ～Ｅは日本語を使用して、Ｆは英語を使用する事がわかる。
【００５７】
さらに、列１１２は、発話開始時間からの経過時間を秒で示したものである。

【００５８】
セクション１１４は、音声関連情報を手動でラベリングした結果を示すものである。ここで、ｙは肯定、ｐは会話に加わっている人の一部による局所的な会話、ｔは通訳もしくは説明、ｗは笑い声をそれぞれ示す。また、図中の縦の棒線は発話中、横の棒線は沈黙を、それぞれ示す。つまり、縦の棒線が続いているとその人物が話し続けている事がわかる。また、横の棒線が続いていると、沈黙し続けている事がわかる。セクション１１４中の６列の記号の並びは、参加者６名の発話情報をそれぞれ示している。例えば、参加者Ｂの発話時間３３１５秒から３３４１秒までを見ると、３３１５秒で肯定する事を表わす音声を発した後、３３１６秒で一旦黙り、その後３３１７秒から３３４１秒まで話し続け、最後の３３４１秒では笑い声をあげているという事がわかる。

【００５９】
図７に、セクション１０４（図５参照）で示された動作情報への手動ラベリング結果を参加者ごとに経時的に配列したものを示す。

【００６０】
図７を参照して、ここでの会話への参加者は図６と同一人物であるＡ～Ｆの６名である。

【００６１】
列１３０は発話時間を秒で示したものである。

【００６２】
セクション１３２は参加者の顔の動きを示すものである。アルファベットｕは上、ｄは下、ｒは右、ｌは左に顔を動かした事をそれぞれ示す。

【００６３】
セクション１３４は参加者の手の動きを示すものである。アルファベットｒは右手、ｌは左手、ｂは両手を動かす事をそれぞれ表わす。

【００６４】
図７に示されたセクション１３２の６列もセクション１３４の６列も、図６のセクション１１４同様、参加者６名の動作情報をそれぞれ示す。例えば、セクション１３２と１３４とを参照すると、参加者Ｂは発話時間３３１４～３３１７秒では顔を右に向けながら両手を動かしているが、３３１８～３３１９秒では顔を左に向けながら左手を動かしているという事がわかる。

【００６５】
これら図７で示される顔の動き、手の動き、及び、図６で示される非言語音声情報とを同期させて、発話内容に依存しない音声－動作情報を得る事ができる。そして、この音声－動作情報から本発明の実施の一形態である自動制御テレビジョン受像機システムを制御するためのモデルを構築する事ができる。例えば、ある非言語音声情報及び動作がある場合に、部外者が会話に割込むのが適切か否かを判定するモデルを考える。このモデル作成にあたっては、会話中に、適宜部外者を会話に割込ませてそのときに会話に加わっている人の各々がどう感じたか、つまり、会話に割込まれて不快だったかという事をリサーチする必要がある。そして、このリサーチ結果を、既に得られている音声－動作情報に付与し、そうした情報を集積し、集積された情報によって、さらに何らかの機械学習を行なう事により、会話がなされている場の雰囲気がいかなるものであれば部外者が口を挟んでもよいかを判定するためのモデルを作成する事ができる。このモデルが、実際の会話の場で、自動制御テレビジョン受像機が会話に口を挟む、つまり会話中に電源を入れても良いか否かを判定する際に参照される。

【００６６】
このモデルに使用されるものとしては、ニューラルネットワーク、ＨＭＭ（Hidden Markov Model）、ＳＶＭ（Support Vector Machine）及び、ＭＬＰ（Multi Layered Perceptron）等が考えられる。

【００６７】
＜動作＞

図１を参照して、まずユーザが、図には示さないテレビジョン受像機本体に取付けられたスイッチ又はリモートコントロールパネルを用いて、ある番組を見るためにテレビジョン受像機の電源を入れたりチャンネルを変えたりする。

【００６８】

次に、電源が入れられ、チャンネルが決定された事に応じて、視聴履歴取得部３０が、そのチャンネル番号とそのチャンネルに決定されたときの日時を含む視聴履歴を取得する。この視聴履歴の取得には、チャンネルが決定されたときの日時が、何月何日の何時何分であるかを計測するためのタイマ３２での計測結果を使用する。この視聴履歴取得処理は、その後、ユーザが別のチャンネルの番組を見るためにチャンネルを変えたときにも同様に行なわれる。そして、ユーザがテレビジョン受像機を見終わって、テレビジョン受像機の電源を切ると、その電源が切られた時点での日時も視聴履歴に関する情報として同様に取得される。この視聴履歴の取得により、何月何日何時何分からいつまで何チャンネルの番組が選択されていたかという事が、明らかになる。

【００６９】

視聴履歴取得部３０によって取得された、視聴履歴は視聴履歴記憶部３４に記憶され蓄積される。この視聴履歴記憶部３４に記憶された視聴履歴から、視聴パターン学習部３６が、ユーザの視聴パターンを学習する。ここで視聴パターンとは、具体的には、ユーザが週に５回月曜日～金曜日の決まった時間、又は、週に１回日曜日に決まったチャンネルで放映される番組を視聴するという様なパターンの事である。この視聴パターン学習部３６によって学習された視聴パターンを、視聴パターン記憶部３８が記憶する。

【００７０】

視聴パターン記憶部３８に記憶された視聴パターンとタイマ３２でのカウント結果とを参照して、テレビ制御信号出力部４０がテレビジョン受像機の電源を入れてチャンネルを合わせる。つまり、ユーザの視聴パターンに一致する日時に視聴パターンに一致するチャンネルに合わせて、自動的にテレビジョン受像機の電源が入れられる。

【００７１】

テレビジョン受像機の電源が入れられる際に、カメラ４４によって撮影された映像とマイクロフォン４６によって収録された音声とを使用して行動指針決定装置４２で決定された行動指針決定情報が参照される。

【００７２】

カメラ４４は、本発明の一実施の形態であるテレビジョン受像機が置いてある部屋の全景が撮影できる様な位置に設置されている。一例として、天井に全方位カメラを設置するという様な方法が考えられる。

【００７３】

マイクロフォン４６で録音された音はカメラ４４で撮影された画像と一致する様なものである必要がある。つまり、同じ場の音と画像とをそれぞれカメラとマイクロフォンで収録する必要がある。それゆえ、マイクロフォン４６はカメラ４４に近い場所に設置されるのが望ましい。一例として、上述した様に天井に取付けられた全方位カメラの周囲を囲む様にマイクロフォンを配置する事が考えられる。

【００７４】

図２を参照して、まず、カメラ４４によって撮影された画像が、毎秒１５フレームでフレーム化され、画像格納部６０に格納される。各フレームに対し、肌色検出部６２で、格納された画像から肌色領域をリアルタイムで検出する。

【００７５】

次に、肌色部位区別部６４で、肌色検出部６２で画像の他の部分から分離された肌色領域の種類を区別する。具体的には、肌色領域の中から、どの領域が顔でどの領域が手であるかを区別する。一般的に、肌色のかたまりのうちで、上位にあり比較的大きい部分が顔、下位にあり比較的小さい部分が手であると考えられる。それゆえ、肌色のかたまりの上下関係、大小関係を比較する事によりこの部位区別処理を行なう。

【００７６】

次に、肌色部位トラッキング部６６で、肌色部位区別部６４により顔と手に区別された肌色領域を領域ごとにトラッキングする。すなわち、あるフレームで検出された肌色領域と、次のフレームで検出された肌色領域との間に対応関係を付ける処理を、各フレームの肌色領域に関する情報が与えられるたびに繰返す。肌色部位トラッキング部６６での処理にあたっては、予め人間の様々な動作を集積したデータから作成された動き予測モデルを参照する。このデータモデルを格納するのが、動き予測モデル格納部６８である。肌色部位トラッキング部６６でトラッキングした結果、次のフレームでの顔及び手が存在する可能性の高い位置が、動き予測モデルにより予測される。

【００７７】

このような動き予測モデルを利用する事によって、ある程度、次に続く動作の予測がつく様になるので、肌色部位トラッキング部６６での処理を速く行なう事ができる。例えば、ある人物の動作をトラッキングしている最中に、手が家具の陰に隠れてカメラ４４の視野から外れてしまった場合を考える。動き予測モデルを参照する事ができれば、この様にトラッキングが不完全にしか行なわれない場合にも、トラッキング対象の顔及び手の位置がある程度予測できる。また、トラッキング結果と動き予測モデルとの比較を行なう事により、トラッキング結果の明らかな誤りを検出し、トラッキングの精度を上げる事もできる。

【００７８】

この様に肌色部位トラッキング部６６でのトラッキング処理によって得られた動作を使用して音声－動作統合部７４でのマッチング処理を行なう。その処理の詳細については後述する。

【００７９】

一方、マイクロフォン４６によって収録された音声は、フレーム化されて音声格納部７０に格納される。音声関連処理部７２は、音声格納部７０に格納された音声から、所定の音響特徴情報を抽出する。

【００８０】

具体的には、音声関連処理部７２は、その音声が部屋のどの方角から聞こえてきたか、つまり発話方位を明らかにする。マイクロフォン４６は指向性のあるマイクロフォンなので、このマイクロフォンから得られる複数の音声信号のレベルを比較する事により、収録された音声の発話方位を明らかにする事ができる。その方位を、予め決められた規則に従ってラベリングする。例えば、ある方位を基点として、そこから時計回りに何度ずれた場所で音声が生じたかを明らかにし、その角度を各音声に対しラベリングする。

【００８１】

音声関連処理部７２は、音声の有無に関する情報も処理する。例えば、いずれかの方向で何らかの音声が生じているときは音声有無情報として「１」をラベリングし、どの方位でも何の音声も生じていないときは「０」とラベリングする。

【００８２】

さらに、音声関連処理部７２は、音声の長短に関する情報も処理する。例えば、何らかの音声が継続している間、上記の音声の有無に関する処理で当該フレームに音声有無情報として「１」というラベルが付される。音声情報の各フレームには時間情報も含まれている。そこで、音声有無情報としてラベル「１」が付されているフレームの継続時間を計測する事により、音声の長短についての情報も得られる。

【００８３】

これら発話方位、音声の有無、及び音声の長短についての情報である音響特徴情報を統合する事によって、どの方位から音声が生じているかが明らかになる。さらに、その方位で生じた音声の継続時間も明らかになる。

【００８４】

この音響特徴情報と、肌色部位トラッキング部６６から得られた画像に関する情報とを音声－動作統合部７４で同期させるとともに統合する。この統合処理の際には、それぞれの情報に含まれている時間情報を参照し、音響と動作情報との間の関係を調べる。例えば

、まず、動作情報から、画面に写っている発話者の数と、その位置とが判る。この情報と、音声の方位とを比較する事により、どの発話者が発話しているかを推定できる。また、誰かが発話を開始した後、その発話者とは異なる互いに隣接する別の二人の一方が、その発話に少し遅れる形で小さな声で発話する場合、その二人の内の一方が他方に発話者の話の内容を説明しているという状況が推測できる。

【００８５】

また、ある人が発話しているときに、他の者の顔の動きから、発話者の話に同意しているか否かがわかる。発話者の発話のリズムに合わせて他の人が発する短い音が規則的に入っていたとすれば、誰かが相槌を打っているという事が推定できる。また、同様に発話のリズムに合わせて他の人の頭部が規則的に上下に動いていれば、頷いているという事が推定できる。さらに、ある人の顔が話者の方を向いていれば、上記のような音声から推測される相槌及び頭部の上下動から推測される頷きがなくとも、その人は話を聞いているという事が推測できる。顔の向きから他に推測する事のできる事の例としては、ある人が何の音声も発せずに、他の複数のある人に交互に顔を向けているとすれば、顔を向けられている複数の人が会話の主導権を握っているという事がわかる。以上の例の様に、発話者に対して他の人が明確な返事を返していなかったとしても、このような非言語的な情報から、誰が会話に参加していて誰が参加していないかという事が比較的容易に推測できる。

【００８６】

音声－動作統合部７４は、この様に、音響特徴情報と動作情報とを統合して処理する事により得られる情報を生成し、音響特徴情報及び動作情報に付加して、一定時間間隔ごとにフレーム化し、バッファ８０に出力する。以下、音響特徴情報、動作情報、それらを統合して得られる情報を音声・動作統合情報と呼ぶ。

【００８７】

バッファ８０は、音声－動作統合部７４から出力された音声・動作統合情報を先入れ先出し方式で所定フレーム分だけ記憶する。モデル参照部７６は、バッファ８０に記憶された音声・動作統合情報のフレームの時系列のうち、最新の所定フレーム数のフレームを読出し、これらフレームのデータを入力としてマッチング用モデル格納部７８に格納されたマッチング用モデルを参照し、マッチング用モデルの出力を得る。この出力が、現時点でテレビジョン受像機の電源をオンしてよいか否かに関する行動指針となる。すなわち、マッチング用モデルの出力は、電源をオンしてよい事を示す値と、そうでない事を示す値とのいずれか一方をとる。

【００８８】

マッチング用モデルは、予め複数人の会話の様子を録画及び録音したものを用いて作成したモデルである。ここでは、マッチング用モデルとして、複数人からなるグループ内で何らかの会話がなされているときに、話の途中で部外者が口を挟める状況であるか否かを判定するというモデルを想定する。このモデルを、テレビジョン受像機のオンしてよいかどうかの行動指針の決定にも使用できる。

【００８９】

まず、会話の最中には、誰かが何かの話をし、それについてグループの他の人が相槌を打ったり、頷いたり、笑い声を立てたりといった反応をするのが一般的である。このような、会話の内容には踏込まない音声情報である非言語音声情報及び動作情報は、それぞれマイクロフォンとカメラとによって収録された音声と画像とから得る事ができる。

【００９０】

ここで必要とされる情報は、具体的には、音の有無、音の方位、音の長短、顔の動き、手の動き、及びそれらの大きさ等である。これらの情報は場を読むための情報を作成するための要素となる。この様な要素の具体例を実際に会話の場を録音及び録画して、学習のためのデータを作成し集積する事によりモデル学習用の集積データを得る。そして、この具体例の各々について部外者に会話に割込まれてもよいか否かに関する実際の答えを参加者から得る。そして、その答えを各学習用データに付する。具体的には例えば、部外者が会話に割込んで良い場合には「１」及び、悪い場合には「０」という様に各学習用データ

に正解情報を付する。そして、この学習用データを用いた機械学習によりモデルの学習を行なう。

【００９１】

この様な処理により、いかなる会話の流れであれば部外者が口を挟んでもよいか否かを判定する、つまり、人間の反応を予測して行動指針を決定するためのモデルを作成する事ができる。

【００９２】

マッチング用モデル格納部７８に格納されたモデルを使用して、モデル参照部７６で、テレビ制御信号出力部４０へ出力するための行動指針決定情報をマッチング処理により作成する。この処理は、音声－動作統合部７４で同期させた非言語音声情報及び動作とマッチング用モデルとを使用して行なわれる。具体的には、音声－動作統合部７４で同期させた非言語音声情報及び動作情報からなる所定フレーム数のデータを入力としてマッチング用モデル格納部７８により格納されたモデルを参照し、その出力を得る。この出力がテレビの電源をオンしてよいか否かに関する行動指針となる。この行動指針を示す信号はテレビ制御信号出力部４０に与えられる。

【００９３】

会話がさえぎられてもよい状況であれば、テレビ制御信号出力部４０が、視聴パターン通りのチャンネルに合わせテレビジョン受像機の電源を入れるための信号を出力する。一方、会話がさえぎられると不都合な状況であれば、テレビ制御信号出力部４０は何の信号も出力しない。

【００９４】

この様に、会話がさえぎられると不都合な状況であればその部屋にあるテレビジョン受像機の電源が入らないので、部屋にいる人が会話をさえぎられて不愉快な思いをする事はない。また、会話を続けるために自動的に入った電源を再び手動で切る手間が省けるという利点もある。また会話をさえぎってもよいような場合には、テレビジョン受像機の電源が自動的に入り、視聴履歴に従って選ばれた番組のチャンネルが選択される。従って、好みの番組を見逃すというおそれが少なくなる。

【００９５】

なお、図１における自動制御テレビジョン受像機システムは、本発明の一実施の形態に係る行動指針決定装置４２を搭載したシステムの一例である。行動指針決定装置４２は、本実施の形態に係る自動制御テレビジョン受像機システムのみならず、議事の進行に直接関係のない会話を記録しない様にする事が可能な自動議事録システム、討論会の様子を撮影するために、次に発話したいという態度を示す発話者に向けてレンズを自動的に移動させるカメラシステム、及び人間に話しかける必要のある機能を持つロボット等、音情報と動作とに反応して自動で動く様々な装置の制御のために搭載する事ができる。

【００９６】

また、本発明に係る行動指針決定装置の構成は、上記実施の形態で示したものには限定されない。例えば、マッチング用モデル格納部７８に格納されるモデルへの入力として使用されるデータ形式に応じ、図２に示したもの以外の構成を採用する事ができる。

【００９７】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内でのすべての変更を含む。

【図面の簡単な説明】

【００９８】

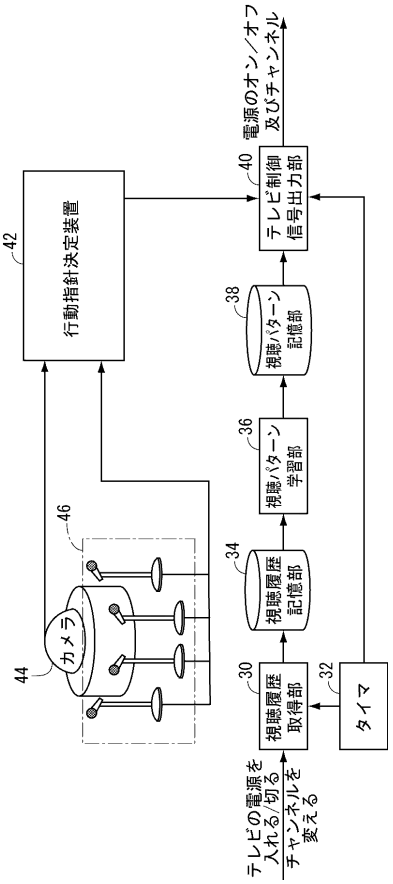
【図１】本発明の一実施の形態に係る行動指針決定装置と自動制御テレビジョン受像機システムの機能をブロック図形式で示した図である。

【図２】行動指針決定装置の詳細をブロック図形式で示した図である。

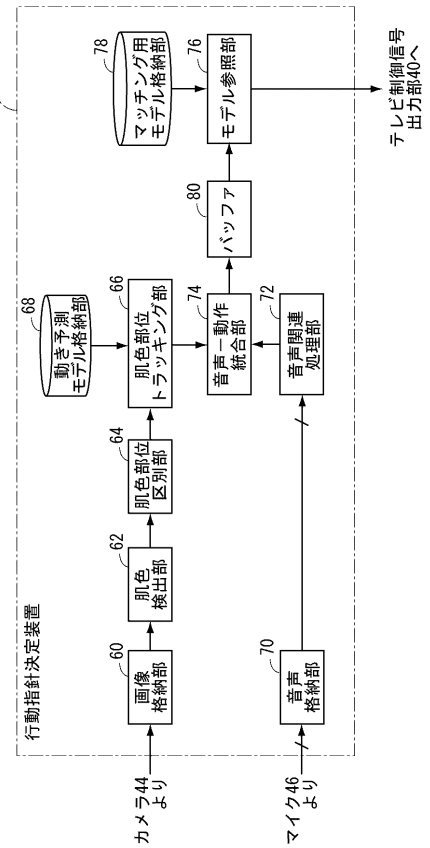
【図３】カメラで撮影された画像の一例を示す図である。

【図４】肌色検出部で肌色検出処理された画像の一例を示す図である。
【図５】マッチング用モデル格納部に格納されるマッチング用モデル作成の際に使用される手動ラベリングの方法の一例を示す図である。
【図６】音声関連情報への手動ラベリング結果を参加者ごとに経時的に配列したものを示す図である。
【図７】動作情報への手動ラベリング結果を参加者ごとに経時的に配列したものを示す図である。
【符号の説明】
【００９９】
４２　行動指針決定装置
６２　肌色検出部
６４　肌色部位区別部
６６　肌色部位トラッキング部
６８　動き予測モデル格納部
７４　音声－動作統合部
７６　モデル参照部
７８　マッチング用モデル格納部

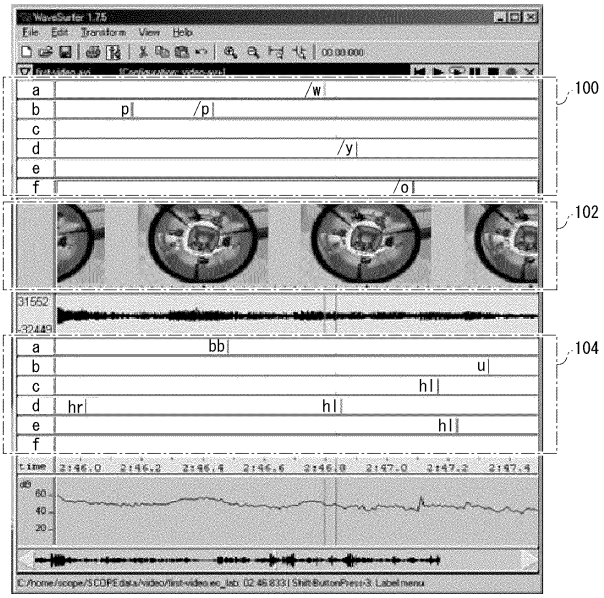
【図１】



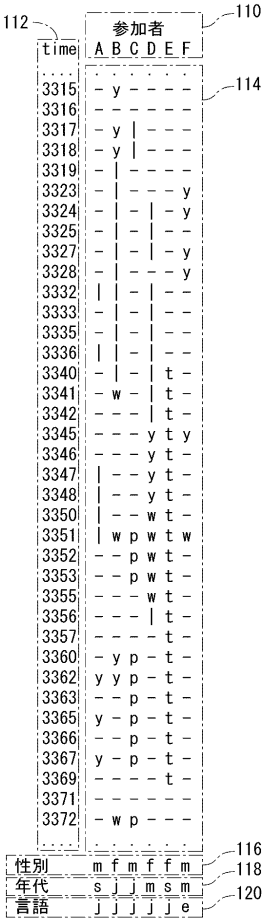
【図２】



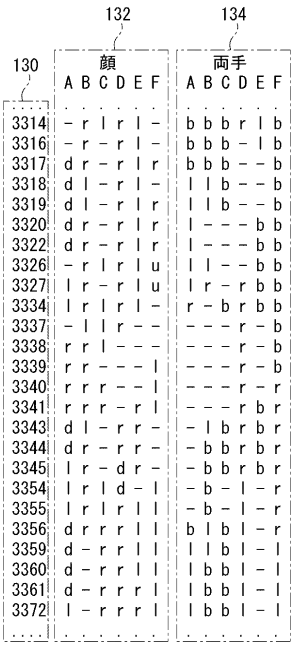
【図５】


WaveSurfer 1.75
File Edit Transform View Help
a
b
c
d
e
f
/w/
p/ /p/
/y/
/o/
100
102
31552
32449
a
b
c
d
e
f
bb/
u/
hl/
hr/ hl/
hl/
104
time
2:46.0 2:46.2 2:46.4 2:46.6 2:46.8 2:47.0 2:47.2 2:47.4
dB
60
40
20
C:/home/scope/SCOPE data/video/first-video.ec_lab: 02:46.833 | Shift-ButtonPress-3: Label menu

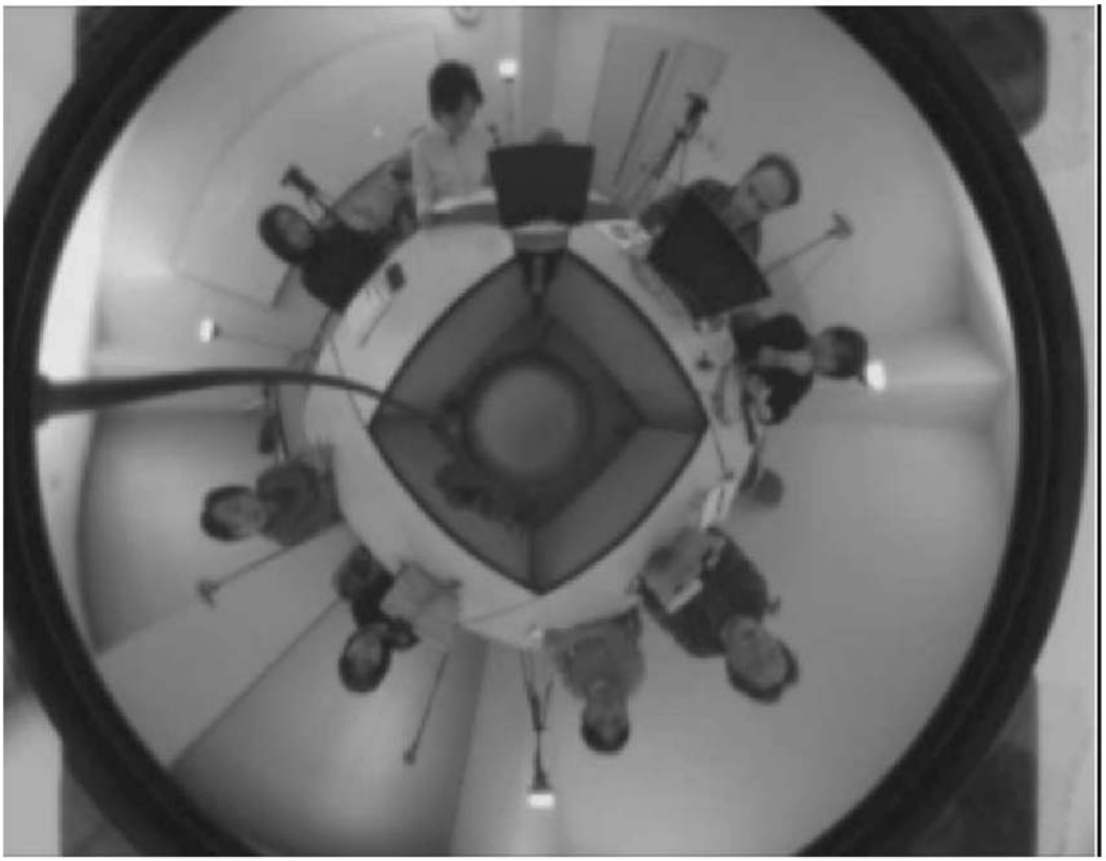
【図６】


112
参加者
110
time
A B C D E F
114
3315 - y - - - -
3316 - - - - - -
3317 - y | - - -
3318 - y | - - -
3319 - | - - - -
3323 - | - - - y
3324 - | - | - y
3325 - | - | - -
3327 - | - | - y
3328 - | - - - y
3332 | | - | - -
3333 - | - | - -
3335 - | - | - -
3336 | | - | - -
3340 - | - | t -
3341 - w - | t -
3342 - - - | t -
3345 - - - y t y
3346 - - - y t -
3347 | - - y t -
3348 | - - y t -
3350 | - - w t -
3351 | w p w t w
3352 - - p w t -
3353 - - p w t -
3355 - - - w t -
3356 - - - | t -
3357 - - - - t -
3360 - y p - t -
3362 y y p - t -
3363 - - p - t -
3365 y - p - t -
3366 - - p - t -
3367 y - p - t -
3369 - - - - t -
3371 - - - - - -
3372 - w p - - -
116
性別 m f m f f m
118
年代 s j j m s m
120
言語 j j j j e

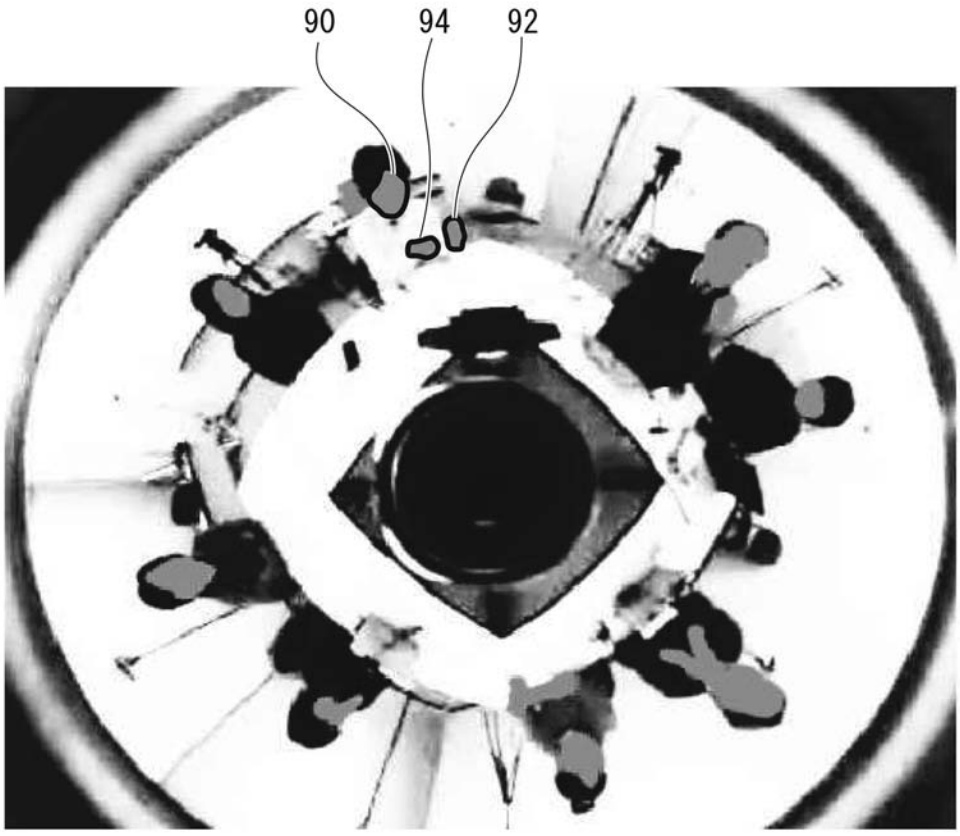
【図７】


132
134
130
顔
両手
A B C D E F
A B C D E F
3314 - r l r l - b b b r l b
3316 - r - r l - b b b - l b
3317 d r - r l r b b b - - b
3318 d l - r l - l l b - - b
3319 d l - r l r l l b - - b
3320 d r - r l r l - - - b b
3322 d r - r l r l - - - b b
3326 - r l r l u l l - - b b
3327 l r - r l u l r - r b b
3334 l r l r l - r - b r b b
3337 - l l r - - - - - r - b
3338 r r l - - - - - - r - b
3339 r r - - - l - - - r - b
3340 r r r - - l - - - r - r
3341 r r r - r l - - - r b r
3343 d l - r r - - l b r b r
3344 d r - r r - - b b r b r
3345 l r - d r - - b b r b r
3354 l r l d - l - b - l - r
3355 l r l r l l - b - l - r
3356 d r r r l l b l b l - r
3359 d - r r l l l l b l - l
3360 d - r r l l l b b l - l
3361 d - r r r l l b b l - l
3372 l - r r r l l b b l - l

【図３】



【図４】



フロントページの続き

(56)参考文献 国際公開第０３／０１５０２８（WO，Ａ１）
特開２００５－２１５９０１（ＪＰ，Ａ）
特開２００１－１５４６８１（ＪＰ，Ａ）
特開２００５－０７８２５７（ＪＰ，Ａ）
Towards reliable multimodal sensing in aware environments, Proceeding PUI '01 Proceedings of the 2001 workshop on Perceptive user interfaces, ２００１年, 第1-6頁
佐藤　洋平、杉原　厚吉，ＰＣへのマルチモーダルな入力手段としてのジェスチャ認識，情報処理学会研究報告，２００４年　9月１１日，Ｖｏｌ．２００４　Ｎｏ．９１，第171-178頁
藤江　真也、江尻　康、菊池　英明、小林　哲則，肯定的／否定的発話態度の認識とその音声対話システムへの応用，電子情報通信学会論文誌，２００５年　3月　1日，Ｊ８８－Ｄ－ＩＩ　第３号，第489-498頁

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｎ　　３／００
Ｇ０６Ｎ　　５／０４
ＩＥＥＥ Ｘｐｌｏｒｅ
ＪＳＴＰｌｕｓ（ＪＤｒｅａｍＩＩ）