

(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2004-310034  
(P2004-310034A)

(43) 公開日 平成16年11月4日(2004.11.4)

(51) Int. Cl. <sup>7</sup>	F I	テーマコード (参考)
G 1 0 L 15/22	G 1 0 L 3/00 5 7 1 U	4 C 0 3 8
G 1 0 L 13/00	G 1 0 L 3/00 5 3 1 N	5 D 0 1 5
G 1 0 L 13/08	G 1 0 L 3/00 5 6 1 C	5 D 0 4 5
G 1 0 L 15/10	G 1 0 L 3/00 5 7 1 G	
G 1 0 L 15/24	G 1 0 L 3/00 5 6 1 D	

審査請求 未請求 請求項の数 14 O L (全 25 頁) 最終頁に続く

(21) 出願番号 特願2003-394663 (P2003-394663)  
 (22) 出願日 平成15年11月25日 (2003.11.25)  
 (31) 優先権主張番号 特願2003-81463 (P2003-81463)  
 (32) 優先日 平成15年3月24日 (2003.3.24)  
 (33) 優先権主張国 日本国 (JP)

(71) 出願人 000005832  
 松下電工株式会社  
 大阪府門真市大字門真1048番地  
 (74) 代理人 100087767  
 弁理士 西川 恵清  
 (74) 代理人 100085604  
 弁理士 森 厚夫  
 (72) 発明者 大林 史明  
 大阪府門真市大字門真1048番地 松下  
 電工株式会社内  
 (72) 発明者 西山 高史  
 大阪府門真市大字門真1048番地 松下  
 電工株式会社内

最終頁に続く

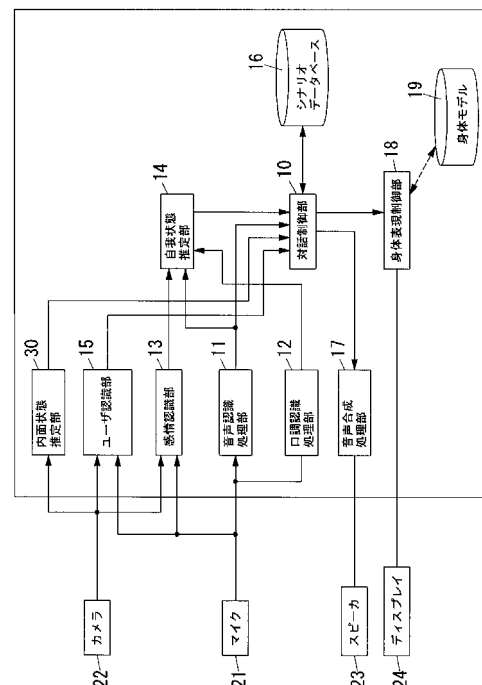
(54) 【発明の名称】 対話エージェントシステム

(57) 【要約】

【課題】 話者の自我状態とエージェントとの自我状態に着目しエージェントの応答を変化させ、話者に受け入れやすくかつ違和感のない自然な対話を行う。

【解決手段】 音声認識処理部11は話者の音声からテキストを抽出し、口調認識処理部12は話者の音声から韻律的特徴を抽出する。自我状態推定部14は、テキストと韻律的特徴とを用いて、話者の自我状態、刺激されたエージェントの自我状態を推定する。対話制御部10は、自我状態推定部14の出力と話者からのテキストとをシナリオデータベース16に照合して得られる対話シナリオに従ってエージェントの自我状態、刺激する話者の自我状態を自動的に決定し、エージェントの発話するテキストを自動的に決定する。音声合成処理部17は、対話制御部10で得られたテキストを用いて合成音声を生じスピーカ23から出力する。

【選択図】 図1



## 【特許請求の範囲】

## 【請求項 1】

話者の音声が入力される音声入力手段と、音声入力手段から入力された話者の音声からテキストを抽出する音声認識処理部と、音声入力手段から入力された話者の音声から韻律的特徴を抽出する口調認識処理部と、少なくとも音声認識処理部により抽出したテキストと口調認識処理部により抽出した韻律的特徴とを用いて話者の自我状態を推定するとともに話者に対応するエージェントにおいて刺激された自我状態を推定する自我状態推定部と、話者の自我状態と刺激された自我状態と話者からのテキストとをエージェントの対話戦略である対話シナリオに対応付けたシナリオデータベースと、自我状態推定部により推定した話者の自我状態と刺激された自我状態と音声認識処理部により抽出したテキストとをシナリオデータベースに照合して得られる対話シナリオに従ってエージェントの自我状態および刺激する話者の自我状態を決定するとともにエージェントの発話するテキストを自動的に決定する対話制御部とを備え、エージェントは、対話制御部において決定したテキストに対話制御部において決定した自我状態から得られる韻律制御用の韻律パラメータを適用した合成音声を作成する音声合成処理部と、音声合成処理部で生成した合成音声を入力する音声出力手段とを備えることを特徴とする対話エージェントシステム。

10

## 【請求項 2】

話者の顔を撮像する画像入力手段と、画像入力手段により撮像した話者の顔の各部位の時間変化と前記音声入力手段から入力された話者の音声における韻律的特徴とから話者の感情を分類する感情認識部とが付加され、前記自我状態推定部は、感情認識部で分類された話者の感情を、話者の発話から抽出したテキストおよび韻律的特徴とに加味して話者の自我状態および刺激された自我状態の推定に用いることを特徴とする請求項 1 記載の対話エージェントシステム。

20

## 【請求項 3】

前記感情認識部は、音声の韻律的特徴を複数種類の感情に対応付けた韻律データベースと、前記音声入力手段から入力される話者の音声の韻律的特徴を韻律データベースに照合して感情の候補を抽出する音声感情認識処理部と、顔の各部位の時間変化のパターンを複数種類の表情に対応付けた表情データベースと、前記画像入力手段により撮像される話者の顔の各部位の時間変化を表情データベースに照合して表情の候補を抽出する表情推定処理部と、音声感情認識処理部で抽出した感情の候補と表情推定処理部で抽出した表情の候補とを用いて話者の感情を決定する感情判断部とから構成されることを特徴とする請求項 2 記載の対話エージェントシステム。

30

## 【請求項 4】

前記表情データベースにおける表情は、「無表情」、「驚き」、「恐怖」、「嫌悪」、「怒り」、「幸福」、「悲しみ」の 7 種類であることを特徴とする請求項 3 記載の対話エージェントシステム。

## 【請求項 5】

話者の音声および話者の顔の特徴とを話者に対応付けて登録したユーザデータベースを含み、前記音声入力手段から入力される話者の音声の特徴量と前記画像入力手段により撮像される話者の顔の画像の特徴量とをユーザデータベースに照合して話者を特定するユーザ認識部を備え、前記シナリオデータベースは、話者の自我状態と刺激された自我状態と話者からのテキストとのほかに話者が併せて対話シナリオに対応付けられ、前記対話制御部は、自我状態推定部により推定した話者の自我状態と刺激された自我状態と音声認識処理部により抽出したテキストとのほかにユーザ認識部で特定された話者を併せてシナリオデータベースに照合して得られる対話シナリオに従ってエージェントの自我状態およびエージェントの発話するテキストを自動的に決定することを特徴とする請求項 2 ないし請求項 4 のいずれか 1 項に記載の対話エージェントシステム。

40

## 【請求項 6】

前記ユーザ認識部は、前記音声入力手段から入力される話者の音声の特徴量を前記ユーザデータベースに照合して話者の候補を抽出する話者認識処理部と、前記画像入力手段に

50

より撮像される話者の顔の画像の特徴量を前記ユーザデータベースに照合して話者の候補を抽出する顔画像認識処理部と、話者認識処理部で抽出した話者の候補と顔画像認識処理部で抽出した話者の候補とを用いて話者を決定する話者判断部とから構成されることを特徴とする請求項5記載の対話エージェントシステム。

【請求項7】

前記ユーザ認識部により特定された話者に対応付けて前記自我状態推定部が推定した話者の自我状態を蓄積して記憶する自我状態履歴記憶部と、自我状態履歴記憶部に蓄積された話者の自我状態の出現頻度の分布パターンにより話者の性格を推定する自我状態特徴抽出部と、自我状態特徴抽出部により推定された性格を話者に対応付けて記憶する自我状態特徴記憶部とが付加され、前記対話制御部は前記自我状態特徴記憶部に格納された話者の性格を用いて前記シナリオデータベースから対話シナリオを選択することを特徴とする請求項5または請求項6記載の対話エージェントシステム。

10

【請求項8】

前記ユーザ認識部には、前記画像入力手段により撮像された画像から前記顔画像認識処理部が複数人を認識したときに、前記話者認識処理部は前記音声入力手段から入力される話者の音声により発話した話者を特定する機能が付加され、さらに前記画像入力手段により撮像された画像内の人物の視線の向きを用いて対話相手を認識する対話相手認識部と、対話における発話者、発話時、対話相手、発話のテキスト、発話者と対話相手との自我状態からなる対話データを蓄積する対話記録部とが付加され、前記対話制御部は前記シナリオデータベースの対話シナリオと対話記録部の対話データとを参照することにより発話のタイミングおよびテキストを決定することを特徴とする請求項6記載の対話エージェントシステム。

20

【請求項9】

前記対話制御部は、話者の発話によって刺激されるエージェントの自我状態を応答時のエージェントの自我状態とし、エージェントの応答時に刺激する話者の自我状態を話者の発話時の自我状態とすることを特徴とする請求項8記載の対話エージェントシステム。

【請求項10】

前記自我状態推定部は、前記音声認識処理部により抽出したテキストに話者の自我状態に対応付けた特定の付帯語が含まれているときには、当該付帯語から推定される自我状態を話者の自我状態とすることを特徴とする請求項1ないし請求項9のいずれか1項に記載の対話エージェントシステム。

30

【請求項11】

話者の覚醒度を判断する覚醒度判断部が付加され、対話制御部は、覚醒度判断部において判断される話者の覚醒度が低いほど音声出力手段から出力する合成音声の速度を低下させることを特徴とする請求項1ないし請求項10のいずれか1項に記載の対話エージェントシステム。

【請求項12】

話者の感情により変化する生体情報を計測する生体情報計測手段と、前記自我状態推定部で推定した自我状態の適否を生体情報計測手段で計測された生体情報を用いて判断する適否判断部とが付加されたことを特徴とする請求項1ないし請求項11のいずれか1項に記載の対話エージェントシステム。

40

【請求項13】

前記エージェントは、身体動作を伴う表現を行う身体モデルと、前記対話制御部で決定されたテキストおよびエージェントの自我状態を身体モデルの身体動作に変換する身体表現制御部とを備えることを特徴とする請求項1ないし請求項12のいずれか1項に記載の対話エージェントシステム。

【請求項14】

前記自我状態は、交流分析に基づく心のモデルである「批判的な親」、「保護的な親」、「大人」、「自由な子供」、「順応する子供」の5種類に分類されることを特徴とする請求項1ないし請求項13のいずれか1項に記載の対話エージェントシステム。

50

## 【発明の詳細な説明】

## 【技術分野】

## 【0001】

本発明は、音声によって話者と対話する対話エージェントシステムに関するものである。

## 【背景技術】

## 【0002】

従来から、コンピュータを用いて音声による話者との対話を可能とする対話システムが種々提案されており、この種の対話システムでは、人間同士の対話と同様な自然な対話を行うことが要望されている。たとえば、話者の音声による認識情報を静的な情報と動的な情報とに分けて保持し、動的な情報を認識対象の項目別に管理する構成によって、対話に用いる内容の絞り込みを迅速に行えるようにする技術が提案されている（たとえば、特許文献1参照）。

10

【特許文献1】特開平6-208389号公報（第0023-0046段落、図3）

## 【発明の開示】

## 【発明が解決しようとする課題】

## 【0003】

上述した特許文献1に記載された技術では、対話における応答時間を短縮することによって応答の遅れによる違和感を抑制する技術であって、話者の発話内容が同じであれば同じ応答になるから、画一的な応答しかできないものである。たとえば、話者が大人か子供かにかかわらず、いかにも機械が応答しているという対話しか行えないという問題がある。

20

## 【0004】

本発明は上記事由に鑑みて為されたものであり、その目的は、話者の自我状態とエージェントにおいて刺激されている自我状態を推定するとともに、エージェントの持つ自我状態とエージェントが刺激する話者の自我状態を設定して応答を変化させることにより、話者にとって受け入れやすかつ違和感のない自然な対話を行うことができる対話エージェントシステムを提供することにある。

## 【課題を解決するための手段】

## 【0005】

請求項1の発明は、話者の音声が入力される音声入力手段と、音声入力手段から入力された話者の音声からテキストを抽出する音声認識処理部と、音声入力手段から入力された話者の音声から韻律的特徴を抽出する口調認識処理部と、少なくとも音声認識処理部により抽出したテキストと口調認識処理部により抽出した韻律的特徴とを用いて話者の自我状態を推定するとともに話者に対応するエージェントにおいて刺激された自我状態を推定する自我状態推定部と、話者の自我状態と刺激された自我状態と話者からのテキストとをエージェントの対話戦略である対話シナリオに対応付けたシナリオデータベースと、自我状態推定部により推定した話者の自我状態と刺激された自我状態と音声認識処理部により抽出したテキストとをシナリオデータベースに照合して得られる対話シナリオに従ってエージェントの自我状態および刺激する話者の自我状態を決定するとともにエージェントの発話するテキストを自動的に決定する対話制御部とを備え、エージェントは、対話制御部において決定したテキストに対話制御部において決定した自我状態から得られる韻律制御用の韻律パラメータを適用した合成音声を作成する音声合成処理部と、音声合成処理部で生成した合成音声を出力する音声出力手段とを備えることを特徴とする。

30

40

## 【0006】

この構成によれば、話者の音声に含まれるテキストと韻律的特徴とを用いて話者の自我状態を推定するとともに、エージェントにおいて刺激された自我状態を推定し、これらの情報をシナリオデータベースに照合することによって抽出される対話シナリオに従ってエージェントの自我状態およびエージェントの発話するテキストを自動的に決定するのであって、さらにエージェントでは対話シナリオにより決定したエージェントの自我状態およ

50

び刺激する話者の自我状態から得られる韻律パラメータを発話用のテキストに適用した合成音声を生成するから、エージェントは、話者の自我状態と話者の音声に含まれるテキストとに応じて、話者に対応する際の内容が同じであってもテキストおよび韻律パラメータを変化させることができ、話者に応じて変化する自然な対話が可能になる。また、話者の自我状態とエージェントにおいて刺激される自我状態とを知ることによって、違和感のない対話が可能な対話シナリオを選択することが可能になる。

【0007】

請求項2の発明では、請求項1の発明において、話者の顔を撮像する画像入力手段と、画像入力手段により撮像した話者の顔の各部位の時間変化と前記音声入力手段から入力された話者の音声における韻律的特徴とから話者の感情を分類する感情認識部とが付加され、前記自我状態推定部は、感情認識部で分類された話者の感情を、話者の発話から抽出したテキストおよび韻律的特徴とに加味して話者の自我状態および刺激された自我状態の推定に用いることを特徴とする。

10

【0008】

この構成によれば、話者の顔の表情および話者の音声における韻律的特徴とから話者の感情を分類する感情認識部とが付加され、話者の自我状態およびエージェントが刺激された自我状態を推定するにあたって、感情認識部により得られる感情も利用するから、適正な対話シナリオを選択することができる。しかも、話者の自我状態を音声および表情によって決めるから、自我状態を決定する精度が高くなる。

【0009】

請求項3の発明では、請求項2の発明において、前記感情認識部は、音声の韻律的特徴を複数種類の感情に対応付けた韻律データベースと、前記音声入力手段から入力される話者の音声の韻律的特徴を韻律データベースに照合して感情の候補を抽出する音声感情認識処理部と、顔の各部位の時間変化のパターンを複数種類の表情に対応付けた表情データベースと、前記画像入力手段により撮像される話者の顔の各部位の時間変化を表情データベースに照合して表情の候補を抽出する表情推定処理部と、音声感情認識処理部で抽出した感情の候補と表情推定処理部で抽出した表情の候補とを用いて話者の感情を決定する感情判断部とから構成されることを特徴とする。

20

【0010】

この構成によれば、話者の音声の韻律的特徴を韻律データベースに照合して話者の感情の候補を抽出するとともに、話者の顔の各部位の時間変化のパターンを表情データベースに照合して話者の表情の候補を抽出し、感情の候補と表情の候補とを用いて話者の感情を決定するから、韻律と表情とに関してパターンマッチングを行って比較的簡単な方法で話者の感情を決定することができる。

30

【0011】

請求項4の発明では、請求項3の発明において、前記表情データベースにおける表情は、「無表情」、「驚き」、「恐怖」、「嫌悪」、「怒り」、「幸福」、「悲しみ」の7種類であることを特徴とする。

【0012】

この構成によれば、表情データベースにおける表情として、「無表情」、「驚き」、「恐怖」、「嫌悪」、「怒り」、「幸福」、「悲しみ」の7種類を用いるのであって、これらの7種類の表情を用いれば自我状態との対応付けは比較的容易である。

40

【0013】

請求項5の発明は、請求項2ないし請求項4の発明において、話者の音声および話者の顔の特徴とを話者に対応付けて登録したユーザデータベースを含み、前記音声入力手段から入力される話者の音声の特徴量と前記画像入力手段により撮像される話者の顔の画像の特徴量とをユーザデータベースに照合して話者を特定するユーザ認識部を備え、前記シナリオデータベースは、話者の自我状態と刺激された自我状態と話者からのテキストとのほかに話者が併せて対話シナリオに対応付けられ、前記対話制御部は、自我状態推定部により推定した話者の自我状態と刺激された自我状態と音声認識処理部により抽出したテキス

50

トとのほかにユーザ認識部で特定された話者を併せてシナリオデータベースに照合して得られる対話シナリオに従ってエージェントの自我状態およびエージェントの発話するテキストを自動的に決定することを特徴とする。

【0014】

この構成によれば、あらかじめユーザデータベースに登録されている特定多数の話者について自我状態を決定するから、対話シナリオを決定する際に話者について既知の情報を利用することが可能になり、不特定多数の話者について自我状態を決定する場合に比較すると、違和感を生じさせない対話シナリオを決定できる可能性が一層高くなる。また、エージェントが対話する話者を特定することによって許可されていない話者との対話を禁止することも可能である。

【0015】

請求項6の発明では、請求項5の発明において、前記ユーザ認識部は、前記音声入力手段から入力される話者の音声の特徴量を前記ユーザデータベースに照合して話者の候補を抽出する話者認識処理部と、前記画像入力手段により撮像される話者の顔の画像の特徴量を前記ユーザデータベースに照合して話者の候補を抽出する顔画像認識処理部と、話者認識処理部で抽出した話者の候補と顔画像認識処理部で抽出した話者の候補とを用いて話者を決定する話者判断部とから構成されることを特徴とする。

【0016】

この構成によれば、話者の音声の特徴量と話者の顔の画像の特徴量とをユーザデータベースに照合して話者の候補を抽出するとともに、抽出した話者の候補を用いて話者を決定するから、話者の認識率が高くなる。

【0017】

請求項7の発明では、請求項5または請求項6の発明において、前記ユーザ認識部により特定された話者に対応付けて前記自我状態推定部が推定した話者の自我状態を蓄積して記録する自我状態履歴記憶部と、自我状態履歴記憶部に蓄積された話者の自我状態の出現頻度の分布パターンにより話者の性格を推定する自我状態特徴抽出部と、自我状態特徴抽出部により推定された性格を話者に対応付けて記憶する自我状態特徴記憶部とが付加され、前記対話制御部は前記自我状態特徴記憶部に格納された話者の性格を用いて前記シナリオデータベースから対話シナリオを選択することを特徴とする。

【0018】

この構成によれば、話者の自我状態の出現頻度の履歴によって話者の性格を推定することができるから、カウンセラーの診断や自己診断テストなどを行うことなく、話者の性格推定が可能になる。性格の推定結果は自我状態特徴記憶部に格納され、話者の性格の推定結果を用いて対話シナリオを選択するから、話者に応じたスムーズな対話が可能になる。なお、話者の自我状態の履歴を記憶しまた性格の推定結果を記憶しているから、話者のカウンセリングのためにカウンセラが利用することも可能である。

【0019】

請求項8の発明では、請求項6の発明において、前記ユーザ認識部には、前記画像入力手段により撮像された画像から前記顔画像認識処理部が複数人を認識したときに、前記話者認識処理部は前記音声入力手段から入力される話者の音声により発話した話者を特定する機能が付加され、さらに前記画像入力手段により撮像された画像内の人物の視線の向きを用いて対話相手を認識する対話相手認識部と、対話における発話者、発話時、対話相手、発話のテキスト、発話者と対話相手との自我状態からなる対話データを蓄積する対話記録部とが付加され、前記対話制御部は前記シナリオデータベースの対話シナリオと対話記録部の対話データとを参照することにより発話のタイミングおよびテキストを決定することを特徴とする。

【0020】

この構成によれば、話者が複数人存在する場合でも対話が可能になり、たとえば看護ロボットに対話エージェントを設けている場合に、被看護人と来訪者（あるいは見舞客）と看護ロボットとの3者での対話が可能になる。ここに、対話制御部において対話記録部の

10

20

30

40

50

対話データを参照しているのは、対話エージェントの応答が要求されているか否かを判断するためであって、対話相手に対話エージェントを含む場合にのみ対話エージェントによる発話を行うことになる。

【0021】

請求項9の発明では、請求項8の発明において、前記対話制御部は、話者の発話によって刺激されるエージェントの自我状態を応答時のエージェントの自我状態とし、エージェントの応答時に刺激する話者の自我状態を話者の発話時の自我状態とすることを特徴とする。

【0022】

この構成によれば、話者とエージェントとにおいて発話の際の自我状態と刺激される自我状態とが一致する結果、違和感のないスムーズな対話が可能になる。 10

【0023】

請求項10の発明では、請求項1ないし請求項9の発明において、前記自我状態推定部は、前記音声認識処理部により抽出したテキストに話者の自我状態に対応付けた特定の付帯語が含まれているときには、当該付帯語から推定される自我状態を話者の自我状態とすることを特徴とする。

【0024】

この構成によれば、話者の自我状態を音声認識処理部により抽出したテキストの中の特定の付帯語によって推定するのであって、自我状態を推定する条件としてこの種の付帯語を用いると信頼性の高い推定が可能であることが知られているから、この種の付帯語がテキストに含まれるときには、これを利用することによって話者の自我状態を高い精度で推定することが可能になる。しかも、話者の発話したテキスト中に特定の付帯語が含まれているか否かを判断するだけであるから、比較的簡単な処理ながら話者の自我状態を精度よく推定することが可能になる。 20

【0025】

請求項11の発明では、請求項1ないし請求項10の発明において、話者の覚醒度を判断する覚醒度判断部が付加され、対話制御部は、覚醒度判断部において判断される話者の覚醒度が低いほど音声出力手段から出力する合成音声の速度を低下させることを特徴とする。

【0026】

この構成によれば、話者の覚醒度に応じて合成音声の速度が調節され、話者の覚醒度が低ければ遅いテンポで応答することにより応答内容を話者に聞き取りやすくし、逆に話者の覚醒度が高ければ速いテンポで応答することにより話者を待たせることなく応答することが可能になる。 30

【0027】

請求項12の発明では、請求項1ないし請求項11の発明において、話者の感情により変化する生体情報を計測する生体情報計測手段と、前記自我状態推定部で推定した自我状態の適否を生体情報計測手段で計測された生体情報を用いて判断する適否判断部とが付加されたことを特徴とする。

【0028】

この構成によれば、自我状態推定部で推定した自我状態が話者の感情を害していないか否かを適否判断部において判断することができる。したがって、適否判断部の判断結果を自我状態の推定結果にフィードバックすれば、話者の感情を害することのない適正な自我状態の推定が可能になる。 40

【0029】

請求項13の発明では、請求項1ないし請求項12の発明において、前記エージェントは、身体動作を伴う表現を行う身体モデルと、前記対話制御部で決定されたテキストおよびエージェントの自我状態を身体モデルの身体動作に変換する身体表現制御部とを備えることを特徴とする。

【0030】

この構成によれば、身体モデルを備えるエージェントについて、対話制御部で決定されたテキストおよびエージェントの自我状態を身体モデルの身体動作に反映させるから、対話の際にエージェントの身振りや手まねを付与することができ、エージェントから話者へのメッセージの伝達がスムーズになる。

【0031】

請求項14の発明では、請求項1ないし請求項13の発明において、前記自我状態は、交流分析に基づく心のモデルである「批判的な親」、「保護的な親」、「大人」、「自由な子供」、「順応する子供」の5種類に分類されることを特徴とする。

【0032】

この構成によれば、自我状態として、交流分析に基づく心のモデルである「批判的な親」、「保護的な親」、「大人」、「自由な子供」、「順応する子供」を用いるから、交流分析に従って対話シナリオを比較的容易に設定することができる。

【発明の効果】

【0033】

本発明の構成によれば、話者の音声に含まれるテキストと韻律的特徴とを用いて話者の自我状態を推定するとともに、エージェントにおいて刺激された自我状態を推定し、これらの情報をシナリオデータベースに照合することによって抽出される対話シナリオに従ってエージェントの自我状態およびエージェントの発話するテキストを自動的に決定するのであって、さらにエージェントでは対話シナリオにより決定したエージェントの自我状態および刺激する話者の自我状態から得られる韻律パラメータを発話用のテキストに適用した合成音声を生成するから、エージェントは、話者の自我状態と話者の音声に含まれるテキストとに応じて、話者に応対する際の内容が同じであってもテキストおよび韻律パラメータを変化させることができ、話者に応じて変化する自然な対話が可能になるという利点がある。また、話者の自我状態とエージェントにおいて刺激される自我状態とを知ることによって、違和感のない対話が可能な対話シナリオを選択することが可能になるという利点がある。

【発明を実施するための最良の形態】

【0034】

(実施形態1)

本実施形態は、コンピュータ装置を用いて実現されるエージェントシステムが話者との間で自然な対話を行うために、交流分析(たとえば、杉田峰泰:「交流分析」,日本文化科学社,1985)に基づく心のモデルである自我状態に着目するものである。自我状態は、親(P)、大人(A)、子供(C)の3状態に大きく分類され、さらに親は批判的な親(CP)と保護的な親(NP)に分類され、子供は自由な子供(FC)と順応する子供(AC)とに分類されている。以下において自我状態を説明する際には、CP、NP、A、FC、ACの符号を用いる。

【0035】

上述のように、交流分析において自我状態は5種類に分類されており、対話する両者の自我状態が適正な関係であるとスムーズな対話が可能になることが知られている。すなわち、発話者と応答者との関係においては、発話者の自我状態と発話者の発話によって応答者の中で刺激される自我状態との関係が適正な関係であれば、応答者の応答は発話者にとって違和感を生じることがなく、応答内容が効率よく伝達され、あるいは発話者に暖かい印象を与える応答が可能になる。一方、発話者の自我状態と応答者の自我状態との関係が適正でないときには、発話者は応答者に対して、なれなれしさを感じて不快感を生じたり、あるいは発話者に冷たい印象を受ける応答になる。要するに、発話者と応答者との自我状態の関係が適正でなければ、スムーズな対話が行えないことが知られている。また、対話によって相手に要求するような場合には、対話戦略に従って自我状態を制御することにより、相手を得心させて要求を受け入れやすくすることも可能になる。このように対話をスムーズに行ったり要求を受け入れさせたりするための自他状態の制御手順を、以下では対話シナリオと呼ぶ。

10

20

30

40

50



## 【0036】

本実施形態においては、コンピュータ装置を用いて実現されるエージェントシステムと話者である人との間で対話を行うこととし、原則として話者が発話するものとする。なお、エージェントシステムが人に対して行動を促すような場合には、エージェントシステムから発話する場合もあり、この種の対話においても以下に説明する対話シナリオに従って自我状態を制御することが可能である。また、本実施形態では仮想的な身体を備えるエージェントを想定するが、実体的な身体を備えるロボットのようなエージェントに本発明の対話エージェントシステムを適用することも可能である。

## 【0037】

対話エージェントシステムの基本的な構成を図1に示す。図示する対話エージェントシステムは、話者の音声が入力される音声入力手段としてマイクロホン21を備えるとともに、話者の顔を撮像する画像入力手段としてTVカメラからなるカメラ22を備える。カメラ22は話者の顔付近を撮像するように視野が設定される。また、対話エージェントシステムは、後述する合成音声を出力する音声出力手段としてスピーカ23を備えるとともに、文字情報を出力したりエージェントの身体表現(身振り・手振りなど)を行うためのディスプレイ24とを備える。つまり、ディスプレイ24には、人や動物などにより表現された仮想的な身体を備えるエージェントが表示され、スピーカ23からの音声出力に合うようにエージェントが身体表現を行う構成を採用している。

## 【0038】

マイクロホン21から入力される音声は、音声認識処理部11と口調認識処理部12と感情認識部13とユーザ認識処理部15とに入力される。音声認識処理部11は、マイクロホン21から入力された話者の音声から話者が話したテキスト(音列)を抽出する。音声認識処理部11における音声の認識には隠れマルコフモデルを用いた周知の技術を採用することができる。口調認識処理部12は、マイクロホン21から入力された話者の音声から韻律的特徴を口調として抽出する。つまり、口調認識処理部12では、話者の音声のパワー、ピッチ、抑揚、速度などの韻律情報を抽出することによって、上述した自我状態の区分に用いる。ちなみに、自我状態と声の調子との関係は、CPでは「やさしい」「安心感を与える」、NPでは「断定的」「疑いがこもる」「非難めく」「押し付け調」、Aでは「落ち着いた低い声で」「単調」「一定の音調」「冷静」「相手に合わせる」、FCでは「開放的」「大声で」「興奮調」「明るい」「無邪気」「楽しそう」、ACでは「自身がない」「くどい」「遠慮がち」「かみつく」「恨みがましい」などの特徴を有することが知られており、この種の情報を韻律的特徴から抽出することによって、話者の自我状態の推定に用いることができる。

## 【0039】

ところで、感情認識部13にはマイクロホン21から入力された話者の音声だけではなく、カメラ22により撮像した話者の顔の画像も入力される。図2に示すように、感情認識部13には、話者の音声における韻律的特徴から話者の感情を推定する音声感情認識処理部13aと、話者の顔の各部位の時間変化から話者の感情を推定する表情推定処理部13bとが設けられる。

## 【0040】

音声感情認識処理部13aはマイクロホン21から入力された話者の音声の韻律的特徴を抽出し、抽出した韻律的特徴を韻律データベース13dと照合する機能を有する。韻律データベース13dには、音声の韻律的特徴が複数種類の感情に対応付けて格納されており、音声感情認識処理部13aは、韻律的特徴を韻律データベース13dと照合することによって、韻律データベース13dから感情を分類する。ここにおいて、音声感情認識処理部13aにおいて韻律的特徴を抽出する技術としては、隠れマルコフモデルを用いたり感情による基本周波数の変化を用いる。また、上述した音声認識処理部11とともに音声感情認識処理部13aにおいても認識精度を高めるために、種々の音声サンプルを収集するとともに各音声サンプルをテキストや感情に対応付けて格納した音声データベースを作成し、音声データベースに収集した音声サンプルを分析することによって得られる情報を

10

20

30

40

50

、音声感情認識処理部13aでのテキストの抽出および音声感情認識処理部13aでの感情の候補の抽出の際の学習情報として用いるようにすれば、テキストおよび感情の候補の分類精度を高めることができる。感情の分類にあたっては1種類の候補を抽出するだけでなく複数種類の候補を尤度に対応付けて抽出するのが望ましい。

#### 【0041】

表情推定処理部13bは、カメラ22で撮像した話者の顔の画像（たとえば、30フレーム/秒で画像を取り込む）のうち、眉、目、口などの各部位について時系列的な変化から表情を分類する。表情の種類としては、エクマン：「表情分析入門」などに記述されているように、「無表情」、「驚き」、「恐怖」、「嫌悪」、「怒り」、「幸福」、「悲しみ」の7種類を用いる。すなわち、表情推定処理部13bでは顔の各部位の時間変化のパターンを抽出するとともに、抽出した時間変化のパターンを表情データベース13eと照合する機能を有する。表情データベース13eには、顔の各部位の時間変化のパターンが複数種類の表情に対応付けて格納されており、表情推定処理部13bは、顔の各部位の時間変化のパターンを表情データベース13eと照合することによって、表情データベース13eからパターンマッチングの技術によって表情を分類する。表情データベース13eの作成にあたっては、種々の表情の画像サンプルを収集するとともに各画像サンプルを感情に対応付けて格納した画像データベースを作成し、画像データベースに収集した画像サンプルを分析することによって得られる情報を、表情推定処理部13bでの感情の候補の抽出の際の学習情報として用いるようにすれば、感情の候補の抽出精度を高めることができる。表情推定処理部13bにおいても音声感情認識処理部13aと同様に、感情の分類にあたっては1種類の候補を抽出するだけでなく複数種類の候補を尤度に対応付けて抽出するのが望ましい。

10

20

#### 【0042】

音声感情認識処理部13aで抽出した感情の候補と表情推定処理部13bにおいて抽出した表情の候補とは感情判断部13cに入力される。感情判断部13cでは、音声感情認識処理部13aで抽出した感情の候補と表情推定処理部13bから入力された表情の候補とをファジー論理などによって組み合わせることにより話者の感情を決定する。

#### 【0043】

上述のようにして、音声認識処理部11で得られたテキストD1と、口調認識処理部12で得られた韻律的特徴D2と、感情認識部13で得られた感情D3とは自我状態推定部14に入力される。すなわち、自我状態推定部14は、話者の発話から抽出したテキストD1および韻律的特徴D2とに、感情認識部13で分類された話者の感情を加味して、話者の自我状態を推定するとともに、話者がエージェントのどの自我状態を刺激しているかを推定する。テキストD1に含まれるキーワード以外の付帯語は、話者の自我状態と刺激を与える相手の自我状態とに深く関係するから、自我状態推定部14での自我状態の推定には、テキストD1の内容（とくに、キーワード以外の付帯語）を主として採用する。また、上述したように韻律的特徴D2も自我状態と関係し、感情D3も自我状態に関連するから、韻律的特徴D2や感情D3をテキストD1に組み合わせることにより自我状態の推定を行う。組合せ方については後述する。

30

#### 【0044】

自我状態推定部14では、話者の自我状態と、刺激を受けたエージェントの自我状態と、話者の感情D3とを対話制御部10に与える。また、対話制御部10には音声認識処理部11で得られたテキストD1も入力される。対話制御部10では、話者の自我状態と刺激されたエージェントの自我状態とテキストD1とをシナリオデータベース16に照合することによって、シナリオデータベース16から対話戦略としての対話シナリオを抽出する。つまり、シナリオデータベース16には、話者の自我状態と刺激されたエージェントの自我状態とテキストD1との組が、対話シナリオに対応付けて格納してある。対話制御部10ではシナリオデータベース16から対話シナリオを抽出すると、テキストD1に対して適切な応答が可能となるように、抽出した対話シナリオに従ってエージェントの自我状態と刺激する話者の自我状態を自動的に決定し、さらに話者からのテキストD1と対話

40

50

制御部 10 において決定した自我状態とに基づいて、エージェントの発話するテキストを自動的に決定する。エージェントの発話するテキストは、話者の発話によるテキスト D 1 中のキーワードに対応して応答用のキーワードが決められており（シナリオデータベース 16 に対話シナリオの一部として格納されている）、対話正よ部 10 において、自我状態に応じて決まる付帯語を応答用のキーワードに付加してテキストを組み立てることにより、エージェントが応答するテキストを生成することができる。

【 0 0 4 5 】

また、対話シナリオによってエージェントの自我状態と刺激する話者の自我状態と応答用のテキストとが決めれば応答用の韻律が決まるから、対話制御部 10 において生成した韻律制御用の韻律パラメータを応答用のテキストに適用するように、音声合成処理部 17 を制御することによって、音声合成処理部 17 において応答用の合成音声を生成することができる。音声合成処理部 17 で生成した合成音声はスピーカ 23 を通して話者に対する応答音声として出力される。

10

【 0 0 4 6 】

ところで、自我状態推定部 14 では、上述のように基本的にはテキスト D 1 と韻律的特徴 D 2 と感情 D 3 との 3 種類の情報を用いて話者の自我状態と刺激されたエージェントの自我状態との推定を行っているが、必ずしも 3 種類の情報のすべてが揃わない場合もある。そこで、本実施形態では、図 3 に示すように、自我状態推定部 14 に入力される情報に優先順位を設定し、少なくとも音声認識処理部 11 からテキスト D 1 が入力されれば、自我状態が推定可能になるようにしてある。図 3 は上段が感情認識部 13 での感情 D 3 の取得の有無、中段が口調認識処理部 12 での韻律的特徴 D 2 の取得の有無、下段が音声認識処理部 11 でのテキスト D 1 の取得の有無を示しており、テキスト D 1 が取得できれば他の情報がなくとも自我状態を推定することができるように処理手順が設定されている。また、テキスト D 1 の次に必要な情報は韻律的特徴 D 2 であって、テキスト D 1 と感情 D 3 との組合せでは、テキスト D 1 のみが用いられることになる。

20

【 0 0 4 7 】

図 3 に示している「自我状態推定表」とは図 4 の表のことであって、自我状態推定表では、感情 D 3 を「喜」「怒」「哀」「恐」「無」の 5 種類の分類してあり、図 4 に示す例では、感情 D 3 が「怒」であるときに韻律的特徴 D 2（図 4 では示していない）を加味し（韻律的特徴 D 2 は重み付けなどによって組み合わせる）、さらにテキスト D 1 にキーワードとは別に含まれる付帯語が「～すべき」であるときに、話者の自我状態が CP であり、刺激されたエージェントの自我状態が AC であることを示している。自我状態推定部 14 ではこの種の対応付けによって話者の自我状態と刺激されたエージェントの自我状態とを推定する。

30

【 0 0 4 8 】

上述のように、自我状態推定部 14 ではテキスト D 1 があれば自我状態を推定することができるが、とくに音声認識処理部 11 により抽出したテキスト D 1 の中に、表 1 に示す特定の付帯語が含まれているときには、この付帯語から推定される自我状態を話者の自我状態に用いる。つまり、自我状態推定部 14 には付帯語を自我状態に対応付けた付帯語データベースが設けられ、テキスト D 1 の中に付帯語が含まれているか否かを付帯語データベースと照合し、付帯語が含まれていれば対応する自我状態を抽出する。また、テキスト D 1 には必ずしも付帯語が含まれているわけではないから、付帯語が含まれていなければ他の情報を用いて話者の自我状態を推定する。

40

【 0 0 4 9 】

【表 1】

自我状態	付帯語
CP	～しなくてはいけない；～しなさい；だめねえ
NP	～してあげよう；よくできたよ；よかったね
A	～と思う；具体的に言うと
FC	ほしい；やって；助けて；～がしたい；好きだ；嫌いだ；できない
AC	困るんです；よくわかりません；もういいです

10

## 【0050】

対話制御部 10 では、自我状態推定部 14 から話者の感情 D3 と話者の自我状態および刺激されたエージェントの自我状態とが入力され、音声認識処理部 11 からはテキスト D1 が入力される。対話制御部 10 ではこれらの 4 種類の情報を用いて対話シナリオを決定しているが（本実施形態では、ユーザ認識部 15 から話者の識別情報 D4 も対話制御部 10 に入力されるが、識別情報 D4 については後述する）、必ずしも 4 種類の情報のすべてが揃わない場合もある。そこで、自我状態推定部 14 と同様に、対話制御部 10 においても入力される情報に図 5 に示すような優先順位を設定し、少なくとも音声認識処理部 11 からテキスト D1 が入力されれば、対話シナリオを決定できるようにしてある。図 5 は最上段がユーザ認識部 15 から入力される識別情報 D4 の取得の有無、2 段目が自我状態推定部 14 からの感情 D3 の取得の有無、3 段目が自我状態推定部 14 からの話者の自我状態および刺激されたエージェントの自我状態の取得の有無、最下段がテキスト D1 の取得の有無を示しており、テキスト D1 が取得できれば他の情報がなくとも対話シナリオを決定することができるように処理手順が設定されている。テキスト D1 の次に重要な情報は、話者の自我状態および刺激されたエージェントの自我状態であり、感情 D3 と話者の識別情報 D4 とは付加的な情報になっている。

20

30

## 【0051】

図 5 に示している「エージェント応答パターン表」とは図 6 の表のことであって、図 6 では話者の自我状態および刺激されたエージェントの自我状態の組合せを、応答におけるエージェントの自我状態と刺激する話者の自我状態との組合せに対応付けている（図示する対応関係は自我状態の基本的な組合せを示しているが、これ以外の組合せも可能である）。図 6 に示すエージェント応答パターン表はシナリオデータベース 16 に格納されたデータの一部であって、実際には感情 D3 や話者の識別情報 D4 も加味される。また、シナリオデータベース 16 では話者から得たテキスト D1 に応答するテキストの対応付けも行われる。

## 【0052】

本実施形態が目指している対話シナリオの一例を図 7 ないし図 9 に示す。図 7 ないし図 9 に示す例は、対話エージェントシステムが駅の案内を行う場合を想定したものである。ここでは、対話がスムーズに進行するように、対話シナリオとして交流分析において知られている相補的交流を採用している。また、説明を簡単にするために、話者の発話によって話者の自我状態とエージェントの自我状態とが決まるものとする。つまり、話者の発話によって刺激されるエージェントの自我状態が、応答時のエージェントの自我状態になり、また話者の発話時の自我状態がエージェントからの応答時に刺激する話者の自我状態になるものとする。

40

## 【0053】

図 7 は話者とエージェントとの自我状態がともに A の場合であって、自我状態を A とす

50

る話者が「駅はどちらですか」という発話を行うことによって、エージェントはAの自我状態が刺激される。ここで、エージェントは対話シナリオによって、エージェントの自我状態をAとし、話者のAの自我状態を刺激するような応答を行うために、「次の信号を右に曲がってまっすぐです」という応答を行うことになる。

【0054】

一方、図8は話者の自我状態がFCであり、エージェントの自我状態がNPになる場合の例であって、自我状態をFCとする話者が「駅に行きたいの」という発話を行うことによって、エージェントはNPの自我状態が刺激される。エージェントは対話シナリオによって、エージェントの自我状態をNPとし、話者のFCの自我状態を刺激するような応答を行うために、「次の信号を右に曲がってごらん」という応答を行う。

10

【0055】

話者の自我状態がFCであっても発話の仕方によっては、刺激されるエージェントの自我状態がNPではなく他の自我状態になる場合もある。図9はこの種の例であって、話者の自我状態がFCであり、エージェントの自我状態もFCになっている。つまり、自我状態をFCとする話者が「駅に行きたーいなー」という発話を行うことによって、エージェントではFCの自我状態が刺激される。対話シナリオによって、エージェントは自我状態をFCとし、話者のFCの自我状態を刺激するような応答を行うために、「次の信号を右に曲がってごらんよ」という応答を行う。

【0056】

図7ないし図9は対話シナリオを例示したものであるが、対話エージェントシステムの使用目的によって対話シナリオは変化し、また応答内容も変化するのは言うまでもない。いずれにせよ、話者の自我状態と話者が刺激するエージェントの自我状態との組合せによって、エージェントが応答に用いるテキストや韻律情報に変化が生じるから、話者にとって受け入れやすく、かつ違和感のない自然な対話が可能になるのである。ここに、話者の発話により刺激されるエージェントの自我状態とは異なる自我状態でエージェントが応答したり、エージェントの応答時に刺激する話者の自我状態を話者の発話時の自我状態とは異ならせたりすると、話者にとっては違和感を生じることが多く、話者が感情を害する場合もあることが知られている。これに対して、本実施形態では、上述のような相補的交流を行うからスムーズな対話が可能になる。

20

【0057】

ところで、上述したユーザ認識部15は、話者があらかじめ登録されているか否かを認識することによって、特定の話者の特徴を利用して認識率を高めたり、話者の認証を行ったりするために設けられている。ユーザ認識部15は、図2に示すように、マイクロホン21から入力される話者の音声の特徴量を話者データベース15dに照合して話者の候補を抽出する話者認識処理部15aと、カメラ22により撮像される話者の顔の画像の特徴量を顔画像データベース15eに照合して話者の候補を抽出する顔画像認識処理部15bとを備える。ここに、図では便宜上、話者データベース15dと顔画像データベース15eとを別に分けて記述しているが、両者を一括してユーザデータベースとするのが望ましい。話者認識処理部15aで抽出した話者の候補と顔画像認識処理部15bで抽出した話者の候補とは話者判断部15cに入力され、話者判断部15cではファジー論理などを用いて話者の候補を組み合わせることにより、話者を決定するとともに話者ごとに付与した識別情報D4を出力する。

30

40

【0058】

ユーザ認識部15から出力される識別情報D4を対話制御部10で採用するために、シナリオデータベース16には、対話における話者の好みや話者の性格などの情報が識別情報D4に対応付けて格納されており、対話制御部10では、ユーザ認識部15から識別情報D4が得られるときには、話者の好みや性格などを加味して対話シナリオを決定する。また、話者の識別情報D4が得られることによって、エージェントが話者の名前を呼びかけに用いることも可能になる。

【0059】

50

上述のようにして対話制御部 10 で対話シナリオが決定されることによって、エージェントの自我状態が決定されるとともに、刺激する話者の自我状態が決定され、さらにテキスト D 1 に対応して応答するテキストの内容が決定されると、上述したように音声合成処理部 17 により合成音声を生成し、スピーカ 23 を通して応答音声を出力する。ここで、エージェントはディスプレイ 24 に表示された仮想的な身体を用いて身体動作を伴う表現が可能になっており、身体表現を行うことによって親近感のある応答が可能になる。

【0060】

このような仮想的な身体による身体表現を行うために、本実施形態の対話エージェントシステムには、仮想的な身体を表現するためのデータ群を記憶装置に格納した身体モデル 19 と、対話制御部 10 で決定した応答内容（テキスト、エージェントの自我状態、刺激する話者の自我状態）を身体モデル 19 の身体動作に変換する身体表現制御部 18 とを設けてある。ここに、身体モデル 19 の身体動作に変換するとは、身体モデル 19 に格納されている身体表現のデータ群から、表現に必要なデータ群を抽出するとともに、抽出したデータ群をディスプレイ 24 に表示された仮想的な身体に適用することを意味する。この処理によって、ディスプレイ 24 に表示されたエージェントの身振り・手振り（手を挙げる、手を伸ばす、首を振る、頷くなど）を対話エージェントシステムの応答内容に応じて制御することが可能になる。

10

【0061】

なお、上述した実施形態では、話者の音声と顔の画像とを入力とし、エージェントは音声と身体表現との出力を行うように構成しているが、話者とエージェントとは少なくとも音声による対話が可能であれば、話者の画像やエージェントの身体表現は必須ではない。

20

【0062】

ところで、上述した例では自我状態の推定に際して話者の感情も用いており、感情の推定にはカメラ 22 で撮像した話者の顔の画像に関する情報を話者の音声から得られる情報と併せて用いている。つまり、カメラ 22 から得られる非言語的な情報は、自我状態の推定に用いることを目的として 5 種類に分類した感情を推定するために用いている。一方、カメラ 22 から得られる画像を用いて目の動きなどを監視すれば、上述のような 5 種類の感情だけではなく、話者の内面状態を推定することが可能である。そこで、感情認識部 13 とは別にカメラ 22 により撮像される顔画像から話者の内面状態を推定する内面状態推定部 30 を設け、話者の特定の内面状態を検出してもよい。

30

【0063】

たとえば、内面状態として「焦り」の状態では視線の移動が激しくなるから、カメラ 22 で撮像した話者の顔の画像における瞳の位置変化（移動量および移動頻度）を監視することにより、話者に「焦り」があるか否かを判断することが可能になる。つまり、内面状態推定部 30 ではカメラ 22 により撮像した顔の画像から画像処理の技術によって両目と瞳との位置を検出し、目の位置に対する瞳の位置の相対的な変化を監視して、移動量が規定値以上である状態が単位時間内に規定頻度以上生じるときには、内面状態が「焦り」であると判断する。

【0064】

内面状態推定部 30 において特定の内面状態（「焦り」のような内面状態）が検出されたときには、自我状態に優先して内面状態を用いるように対話制御部 10 に指示する。対話制御部 10 では内面状態を自我状態に優先させる指示があると、シナリオデータベース 16 から特定の内面状態に対応した対話シナリオを抽出する。内面状態に対応する対話シナリオとしては一般に A A あるいは A C C P を選択し、自我状態によって F C F C という応答が望ましいと判断されても、内面状態による判断を優先して用いる。

40

【0065】

いま、話者に対して料理の献立や調理方法のアドバイスを行う目的で対話エージェントを用いる場合を想定する。ここで、「焦り」がない場合には〔1〕のように対話を行い、内面状態として「焦り」が検出されたときには〔2〕のように対話を行う。

〔1〕

50

話者：簡単にお客様に出せる料理を教えてくださいの、助けて！

対話エージェント：はい、わかりました。まずどんな材料を使いますか？

〔 2 〕

話者：簡単にお客様に出せる料理を教えてくださいの、助けて！

対話エージェント：はい、とりあえず次のようなメニューはいかがですか？

また、上述した構成において、マイクロホン 2 1 により得られる音声およびカメラ 2 2 により撮像される画像に基づいて年齢層や性別を判断する機能をユーザ認識部 1 5 に持たせてもよい。話者の年齢層や性別を推定すれば、対話制御部 1 0 において年齢層や性別に合わせた対話シナリオを選択することが可能になる。たとえば、低年齢層の子供の場合に、対話エージェントから話者に対して C P A C という威圧的な応答を行ったとしても、話者から対話エージェントに対して A C C P という服従的な応答を得ることは難しい。この場合、対話エージェントから話者に対して A A という平静、あるいは N P F C という教育的な応答が望ましい。あるいはまた、子供向けに語尾を変えることも考えられる。一方、話者が高齢者であれば、対話エージェントが応答するテキストをかしこまった文言とするのが望ましい。このように年齢層や性別の情報を取得することによって、より適切な対話シナリオを選択することが可能になる。

10

【 0 0 6 6 〕

さらにまた、話者の性格的傾向は大人と子供とでは異なる場合があるから（大人は自己否定的、子供は自己肯定的な傾向がある）、大人と子供とで応答の内容を変化させるようにしてもよい。たとえば、看護ロボットに対話エージェントを用いるような場合には、大人の話者に対しては励ますように応答することが望ましい。

20

【 0 0 6 7 〕

本実施形態では、対話エージェントの出力手段としてスピーカ 2 3 とディスプレイ 2 4 とを用いているが、香料噴出装置（図示せず）を付加してもよい。一般に香りによる各種の心理的効果が知られており、たとえばジャスミンの香りは、リラックス効果、情緒安定効果などがあり、また抑鬱状態にもよいと言われている。そこで、話者から対話エージェントに対して C P A C であって話者が怒気を含んだ心理状態であるときや、話者が抑鬱的な状態であるときには、香料噴出装置からジャスミンの香りを噴出することにより、対話エージェントの応答だけではなく、香りの効果によっても話者の心理状態を改善することが可能になる。この機能は看護ロボットや介護ロボットに組み込むのに適している。

30

【 0 0 6 8 〕

上述した説明では、話者の発話に対して対話エージェントが応答する例を示したが、話者の発話がない場合でも対話エージェントから発話する機能を設けることによって、対話エージェントの擬人性を高めることができるから、対話エージェントに話者の存在を検出する機能を付加するのが望ましい。この機能を実現するには、カメラ 2 2 の視野内の人の存否を検知することができる人感センサ（図示せず）を付加する。人感センサには、人体から放射される熱線を検知する焦電型赤外線センサなどを用いた周知のものを用いる。対話エージェントでは、人感センサが人を検知すると、カメラ 2 2 で撮像されている画像内の人を顔画像データベース 1 5 e に照合することによって登録されている人物か否かを判断する。ここで、ユーザ認識部 1 5 において登録されている人物か否かを判断する際に、音声を利用することができないから、顔画像データベース 1 5 e のみを用いて人物を判断することになる。カメラ 2 2 で撮像された画像からユーザ認識部 1 5 に登録された人物と判断されると、対話制御部 1 0 では当該人物の名前を呼びかけて挨拶を行い、人物が非登録であると挨拶のみを行う。また、呼びかけに回答して人物が発した音声から人物の誤認に気付いたときには、名前を訂正する回答を行えばよい。なお、カメラ 2 2 の画像で人物を認識したときに対話エージェントが挨拶の発話を行うのは一例であって、たとえば、許可者以外は侵入禁止であるような場所であれば、登録された人物には侵入許可を通知し、登録されていない人物には侵入禁止を通知するように発話するなど、発話内容は目的に応じて適宜に設定される。

40

【 0 0 6 9 〕

50

## (実施形態2)

本実施形態は、ユーザ認識部15により話者を特定できることを利用して話者ごとに自我状態の履歴を記録し、話者の自我状態の履歴を用いて話者の性格を推定することにより話者の性格に合わせた応答を可能とするものである。

## 【0070】

本実施形態では、図10に示すように、自我状態推定部14が推定した話者の自我状態をユーザ認識部15において特定した話者に対応付けて蓄積して記憶する自我状態履歴記憶部31と、自我状態履歴記憶部31に記憶した話者の自我状態の出現頻度の分布パターンから話者の性格を推定する自我状態特徴抽出部32と、自我状態特徴抽出部32により推定した性格を話者に対応付けて記憶する自我状態特徴記憶部33とを付加している。自我状態履歴記憶部31では話者の自我状態を時系列で履歴として記憶するとともに各自我状態の出現頻度を記憶している。自我状態特徴抽出部32では、自我状態の出現頻度の分布パターンと性格とを対応付けて登録してある自我状態特徴データベース34とを照合し、一致度の高い性格を話者の性格として推定する。すなわち、自我状態特徴データベース34には、交流分析における5種類の自我状態の出現頻度の分布パターン(出現頻度を正規化した分布パターン)が性格に対応付けて登録してあり、自我状態履歴記憶部31に記憶した話者ごとの自我状態の出現頻度の分布パターンが、自我状態特徴データベース34に登録されている分布パターンと照合される。この照合はパターンマッチングであって類似度の高いものが選択され、選択された性格が話者の性格と推定される。自我状態特徴データベース34では、たとえば、自我状態としてNP, FCが低く、CP, ACが高い場合には、「自分を表現することができにくく、鬱になりやすい性格」「不登校タイプの性格」「責任感、現実検討能力、協調性は十分持っているが、思いやりに欠ける性格」などの性格が対応付けられる。自我状態特徴抽出部32での性格の推定結果は話者と対応付けて自我状態特徴記憶部33に記憶される。

10

20

## 【0071】

ところで、自我状態特徴記憶部33に話者の性格が記憶されているときには、当該話者との対話を行う際に、対話制御部10では自我状態特徴記憶部33から話者の性格を取得する。シナリオデータベース16には話者の性格に対応付けて対話シナリオが登録されているから、対話制御部10においてシナリオデータベース16から話者の性格に応じた対話シナリオを選択することができ、結果的に話者の性格に応じたスムーズな対話が可能になる。

30

## 【0072】

また、本実施形態の対話エージェントを看護ロボットなどに用いる場合に、話者の性格を推定することによって、対話エージェントをセラピストのように機能させることが可能になる。たとえば、話者の性格が悲観的あるいは自虐的な傾向であるときに、話者の心理状態を向上させるような応答を行うことが可能になる。いま、自我状態の出現頻度のうちNP, FCが低く、CP, ACが高い場合には上述したように、鬱傾向があることが知られている。そこで、一般の話者では対話エージェントと話者との間で以下に〔1〕で示す対話を行うとすれば、自我状態の出現頻度のうちNP, FCが低く、CP, ACが高い話者では以下に〔2〕で示すように肯定的な表現を用いて話者を励まし、話者の心理状態を向上させる対話が可能になる。

40

## 〔1〕

対話エージェント：お薬の時間ですよ。

話者：もうわかったよ。

対話エージェント：では、よろしくお願ひしますね。

## 〔2〕

対話エージェント：お薬の時間ですよ。

話者：もうわかったよ。

対話エージェント：これで良くなりますから、元気を出してください。もう少しですから。

50



## 【0073】

なお、上述のように、自我状態履歴記憶部31には話者ごとの自我状態の履歴（自我状態の出現頻度を含む）が記憶されており、また自我状態特徴記憶部33には話者ごとの性格の推定結果が記憶されているから、話者がカウンセリングを受ける際にはカウンセラにこれらのデータを提供することによって、カウンセリングの参考に用いることが可能になる。他の構成および機能は実施形態1と同様である。

## 【0074】

## （実施形態3）

上述した各実施形態では話者が1人である場合を例示したが、本実施形態は2人の話者が存在し対話エージェントを含めて3者での対話を行うことを可能とする構成について説明する。本実施形態では、図11に示すように、カメラ22により撮像された画像に含まれる人物の視線の向きを監視し、視線の向きによって対話相手を認識する対話相手認識部35と、対話における発話者、発話時、対話相手、発話のテキスト、発話者と対話相手との自我状態からなる対話データを蓄積する対話記録部36とを付加している。発話者および発話時はユーザ認識部35により取得でき、対話相手は対話相手認識部35から取得でき、発話のテキストおよび発話者と対話相手との自我状態は対話制御部10から取得することができる。なお、本実施形態では、カメラ22として話者の目の位置および瞳の位置を監視できる程度の解像度のものを用いることが必要である。

## 【0075】

また、ユーザ認識部15では、カメラ22で撮像された画像から顔画像認識処理部15bが2人の話者を認識したときに、対話制御部10、感情認識部13、自我状態推定部14、対話相手認識部35に通知することによって、2人の話者が存在する対話を行う動作に切り換える。その後、ユーザ認識部15では、マイクロホン21から入力される話者の音声を用いて話者認識処理部15aにおいて発話した話者を特定し、対話制御部10、感情認識部13、自我状態推定部14、対話相手認識部35に対して認識した話者を通知する。要するに、本実施形態ではカメラ22で撮像された画像を、話者の人数と話者の対話相手との特定に用い、マイクロホン21から入力される音声により発話した話者の感情を推定し、発話した話者と対話相手との自我状態を推定する。推定された自我状態と、音声認識処理部11で得られたテキストとは、対話制御部10に与えられ、対話制御部10では上述した対話データを対話記録部36に記録する。

## 【0076】

本実施形態におけるシナリオデータベース16には3者間の対話シナリオが上述した各実施形態のような2者間での対話シナリオとは別に格納されており、対話相手認識部35から2人の話者が存在することが対話制御部10に通知されると、対話制御部10ではシナリオデータベース16から3者間の対話シナリオを選択する。3者間の対話シナリオでは、対話相手が対話エージェントを含むときにのみ対話エージェントが発話するように設定され、話者同士の対話では対話エージェントから発話しないように設定される。つまり、対話制御部10は、音声認識処理部11が抽出した発話のテキストと、対話相手認識部35が認識した対話相手と、自我状態推定部14が推定した自我状態とを用い（対話記録部36に対話データとして記録されている）、対話エージェントが発話すべきか否かを判断する。対話エージェントが発話しないときにはカメラ22で撮像した画像から話者の人数を確認する処理に戻り、対話エージェントが発話するときには上述した各実施形態と同様に音声合成処理部17および身体表現制御部18を通して身体表現を伴う発話を行う。

## 【0077】

本実施形態を用いた対話シナリオの一例を以下に示す。以下に示す例は、対話エージェントが看護ロボットに設けられており、被看護人と来訪者との2人の話者が存在する場合を例示している。また、〔1〕は対話エージェントが発話しない場合であり、〔2〕は対話エージェントが発話する場合である。

## 〔1〕

被看護人 来訪者（FC FC）：学校では変わったことはない？

10

20

30

40

50

来訪者 被看護人 ( F C F C ) : いつも通りだよ。

〔 2 〕

被看護人 来訪者 ( F C N P ) : しんどいよ。

来訪者 被看護人 ( N P F C ) : もうすぐ良くなるよ。

看護ロボット 被看護人 ( N P F C ) : そうですよ、がんばってください。

【 0 0 7 8 】

なお、本実施形態では 2 人の話者と対話エージェントとの 3 者による対話を例示したが、本実施形態の技術を 3 人以上の話者に拡張することによって、さらに多人数での対話も可能である。他の構成および動作は実施形態 1 と同様である。

【 0 0 7 9 】

( 実施形態 4 )

本実施形態は話者の覚醒度を判断することによって、対話エージェントが応答する際の合成音声の速度を調節するものである。つまり、話者の覚醒度が低いときには、速いテンポで応答すると応答内容を話者が聞き取れない場合があり、逆に話者の覚醒度が高いときには、遅いテンポで応答すると話者がいらだつ場合があるから、話者の覚醒度を判断するとともに、話者の覚醒度に応じてスピーカ 2 3 から出力する合成音声の速度を調節している。

【 0 0 8 0 】

本実施形態では、話者の覚醒度の判断に皮膚電位水準 ( S P L ) を用いる。皮膚電位水準を求めるために、図 1 2 に示すように、話者の手のひらや足の裏のように精神状態による発汗量の変化が生じやすい部位に接触可能な電極 2 5 ( 電極 2 5 は一対ある ) を設けてあり、電極 2 5 は電位計測部 3 7 に接続され電位計測部 3 7 では一対の電極 2 5 の電位差を計測する。電極 2 5 が話者に装着されていると話者は煩わしく感じる可能性があるから、たとえば対話エージェントを看護ロボットに組み込むような場合には、話者との握手などの行為によって話者が電極 2 5 に触れるようにするのが望ましい。電位計測部 3 7 により計測された電位差は覚醒度判断部 3 8 に入力され、覚醒度判断部 3 8 では電位差を覚醒度に変換する。一般に、電位差 ( S P L ) が大きいほど覚醒度が高いから、適宜に設定した閾値に対して電位差が大きければ覚醒度判断部 3 8 において話者の覚醒度が高いと判断する。なお、本実施形態では話者の覚醒度を高低 2 段階で判断しているが、覚醒度を多段階で判断してもよい。ところで、電位計測部 3 7 で計測される電位差は話者によって個人差があるから、ユーザ認識部 1 5 により認識した話者に対応付けて電位計測部 3 7 での検出結果を記憶して蓄積しておき、話者別の蓄積結果の平均値を上述の閾値に用いるようにすればよい。

【 0 0 8 1 】

上述のようにして覚醒度判断部 3 8 において求めた話者の覚醒度は対話制御部 1 0 に与えられ、対話制御部 1 0 では話者の覚醒度が低い ( S P L が閾値より低い ) ときには、合成音声による応答の速度が比較的遅いテンポになるように、音声合成処理部 1 7 で生成する合成音声の出力速度を調節する。このように覚醒度の低い話者に対して遅いテンポで応答することにより、テキストの内容を聞き取りやすくすることができる。一方、話者の覚醒度が高い ( S P L が閾値より高い ) ときには、やや速いテンポで応答することで話者の覚醒度を保ち、話者が応答の遅さにいらだつことがないようにする。他の構成および動作は実施形態 1 と同様である。

【 0 0 8 2 】

( 実施形態 5 )

本実施形態は、話者の生体情報を検出することによって対話エージェントが推定した自我状態が対話に適切であったか否かを判断するものである。すなわち、話者と対話エージェントとが対話する際には自我状態推定部 1 4 において話者の発話に基づいて話者の自我状態と対話エージェントにおいて刺激される自我状態とを推定している。一般的には、推定した自我状態を用いて相補的交流を行えばスムーズな対話が可能になる。ここでスムーズな対話とは、話者が感情を害することなく対話することを意味する。一方、相補的交流

10

20

30

40

50

ではなく自我状態の交差が生じるような交流では、話者は感情を害して対話がスムーズに進行しなくなる可能性がある。

【0083】

そこで、本実施形態では、話者の感情を判断するために、話者の瞬時心拍率（1分間の心拍数）を用いている。一般に瞬時心拍率は、怒り、ストレス、恐怖などの防衛的あるいは攻撃的な感情が生じると上昇し、落ち着いているときには低下することが知られている。本実施形態では、このような生理現象を利用して対話中の話者の瞬時心拍率を求め、瞬時心拍率の変動を監視することによって、話者に防衛的ないし攻撃的な感情が生じていないか否かを判断している。瞬時心拍率の検出にはECGを用いており、実施形態4と同様に電極26（電極26は一对ある）を設けている。電極26は話者の胸部あるいは四肢に装着される。

10

【0084】

図13に示すように、電極26はECG計測部41に接続され、一对の電極26により検出される電位差がECG計測部41において検出される。ECG計測部41では、たとえば図14に示すような電位変化が検出される。この電位変化は、心臓における心房の興奮を示すP波と、心室の脱分極を示すQRSと、心室の細分極を示すT波とを含む。ここで、R-Rの間隔は心臓交換神経と心臓副交感神経との拮抗支配を受けており、R-R間隔を1分当たりの心拍数に換算することで瞬時心拍率を求めることができる。そこで、図14に示すようなECG計測部41の出力を瞬時心拍率検出部42に入力し、R-R間隔を求めることによって瞬時心拍率を求める。すなわち、電極26とECG計測部41と瞬時心拍率検出部42とにより生体情報計測手段が構成される。話者が防衛的ないし攻撃的な感情を持つと、瞬時心拍率が常時よりも上昇するから、適否判断部43では瞬時心拍率検出部42で求めた瞬時心拍率を適宜の閾値と比較することにより、話者の感情が防衛的ないし攻撃的であるか否かを判断する。つまり、適否判断部43は自我状態推定部14が推定した自我状態が適切であったか否かを判断することになる。適否判断部43では、瞬時心拍率が閾値よりも低いときには自我状態推定部14で推定した話者の自我状態および対話エージェントの刺激された自我状態が適切であったと判断し、瞬時心拍率が閾値以上であるときには自我状態推定部14の推定結果が適切でなかったと判断する。

20

【0085】

なお、図13に破線で示すように、適否判断部43による判断結果を自我状態推定部14にフィードバックして自我状態推定部14の学習に用いれば、推定した自我状態の信頼度を高めることができ、結果的に対話エージェントの応答によって話者の感情を害する可能性が低減されスムーズな対話が可能になる。また、瞬時心拍率の検出には電極26に変えて血流を監視する光学式のセンサを用いてもよい（この種のセンサは種々運動機械における心拍計に用いられている）。他の構成および動作は実施形態1と同様である。

30

【図面の簡単な説明】

【0086】

【図1】実施形態1を示すブロック図である。

【図2】同上の要部のブロック図である。

【図3】同上に用いる自我状態推定部の動作説明図である。

40

【図4】同上に用いる自我状態推定表の一例を示す図である。

【図5】同上に用いる対話制御部の動作説明図である。

【図6】同上に用いるエージェント対応パターン表を示す図である。

【図7】同上を用いた対話例を示す図である。

【図8】同上を用いた対話例を示す図である。

【図9】同上を用いた対話例を示す図である。

【図10】実施形態2を示すブロック図である。

【図11】実施形態3を示すブロック図である。

【図12】実施形態4を示すブロック図である。

【図13】実施形態5を示すブロック図である。

50

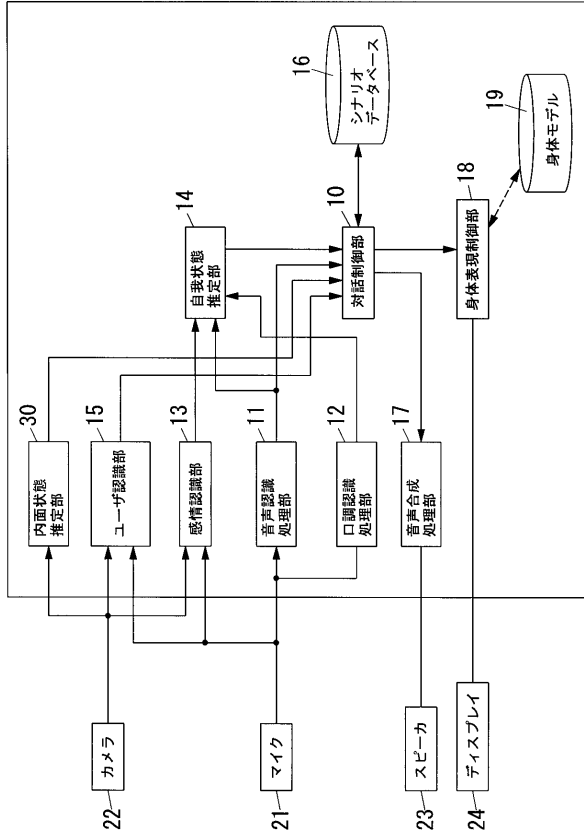
【図14】 同上に用いるECGの一例を示す図である。

【符号の説明】

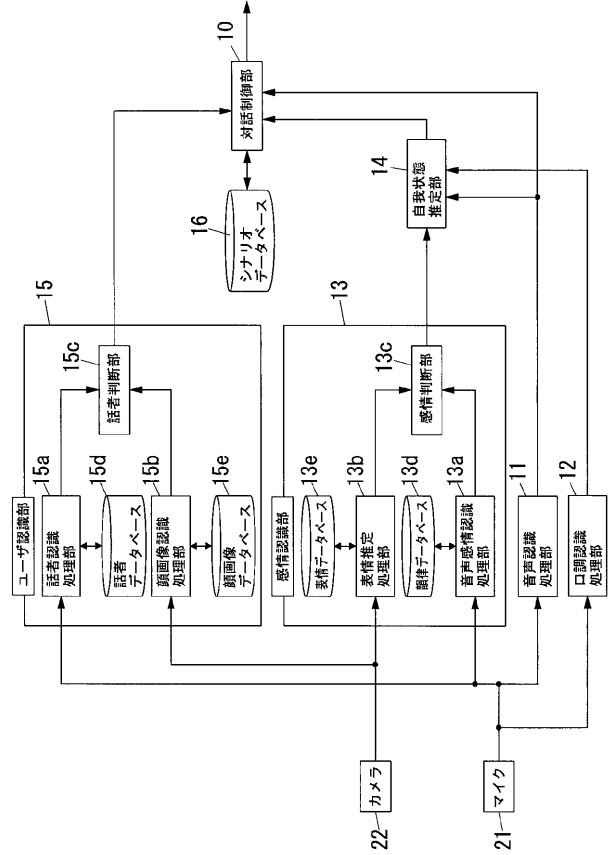
【0087】

- |     |            |    |
|-----|------------|----|
| 10  | 対話制御部      |    |
| 11  | 音声認識処理部    |    |
| 12  | 口調認識処理部    |    |
| 13  | 感情認識部      |    |
| 13a | 音声感情認識処理部  |    |
| 13b | 表情推定処理部    |    |
| 13c | 感情判断部      | 10 |
| 13d | 韻律データベース   |    |
| 13e | 表情データベース   |    |
| 14  | 自我状態推定部    |    |
| 15  | ユーザ認識部     |    |
| 15a | 話者認識処理部    |    |
| 15b | 顔画像認識処理部   |    |
| 15c | 話者判断部      |    |
| 15d | 話者データベース   |    |
| 15e | 顔画像データベース  |    |
| 16  | シナリオデータベース | 20 |
| 17  | 音声合成処理部    |    |
| 18  | 身体表現制御部    |    |
| 19  | 身体モデル      |    |
| 21  | マイクロホン     |    |
| 22  | カメラ        |    |
| 23  | スピーカ       |    |
| 24  | ディスプレイ     |    |
| 25  | 電極         |    |
| 26  | 電極         |    |
| 31  | 自我状態履歴記憶部  | 30 |
| 32  | 自我状態特徴抽出部  |    |
| 33  | 自我状態特徴記憶部  |    |
| 35  | 対話相手認識部    |    |
| 36  | 対話記録部      |    |
| 37  | 電位計測手段     |    |
| 38  | 覚醒度判断部     |    |
| 41  | ECG計測部     |    |
| 42  | 瞬時心拍率検出部   |    |
| 43  | 適否判断部      |    |

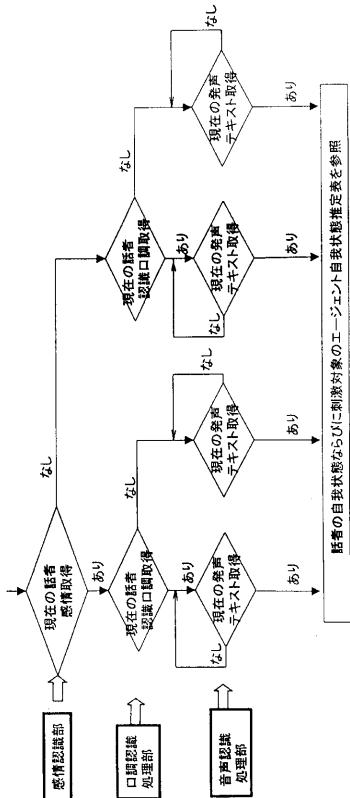
【 図 1 】



【 図 2 】



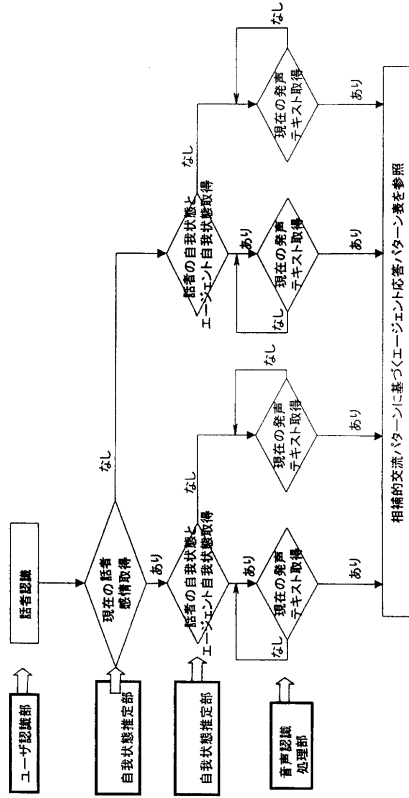
【 図 3 】



【 図 4 】

喜	～すばき ～りましたまえ	～してらよ ～してみたら?	～しましうよ	～ではありません ～して頂けません?
怒	～すばき ～りましたまえ	～してらよ ～してみたら?	～しないで! ～しないわよ!	～しませんか? ～致します
哀	～すばき ～りましたまえ	～してらよ ～してみたら?	～なのよねえ ～なんだよなあ	～致しますか? ～致します
恐	～すばき ～りましたまえ	～してらよ ～してみたら?	～してらよ ～してみたら?	～致しますか? ～致します
無	～すばき ～りましたまえ	～してらよ ～してみたら?	～してらよ ～してみたら?	～致しますか? ～致します
話者の自身体態 エージェントの自身体態 (刺激対象)	CP	NP AC or FC	A FC or A	AC CP

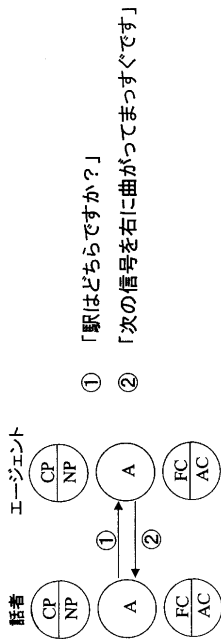
【 図 5 】



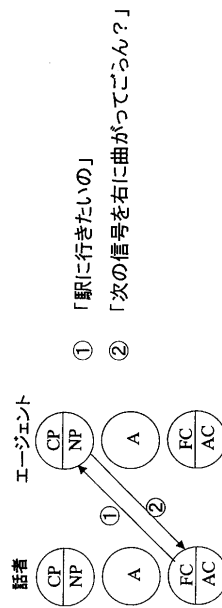
【 図 6 】

話者からの入力		エージェントの出力(応答)	
話者の自己状態	刺激対象のエージェント自己状態	エージェントが保持する自己状態	刺激対象の話者自己状態
CP	AC	AC	CP
NP	AC	AC	NP
NP	FC	FC	NP
A	A	A	A
FC	FC	FC	FC
FC	NP	NP	FC
FC	A	A	FC
AC	CP	CP	AC

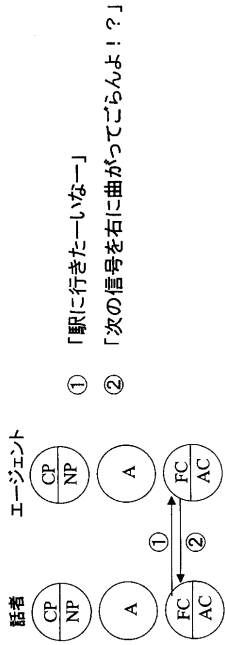
【 図 7 】



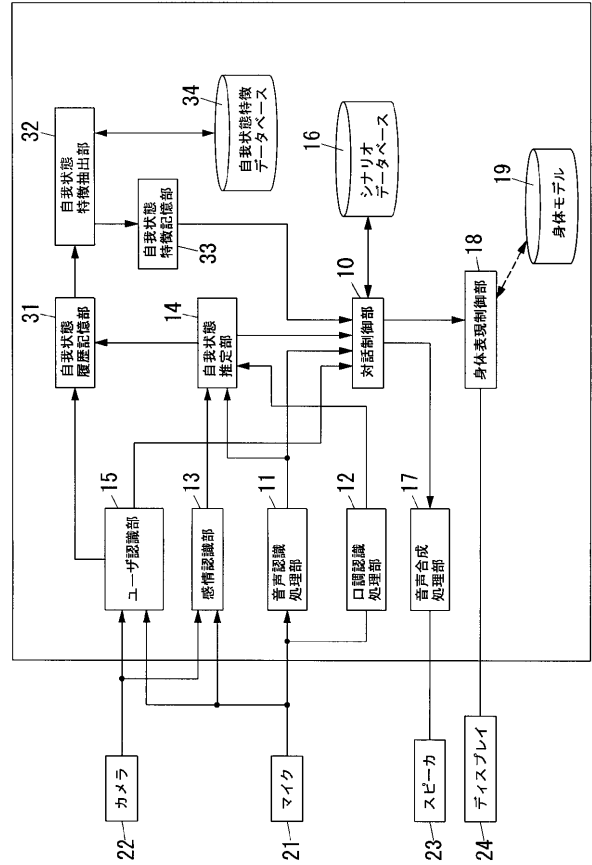
【 図 8 】



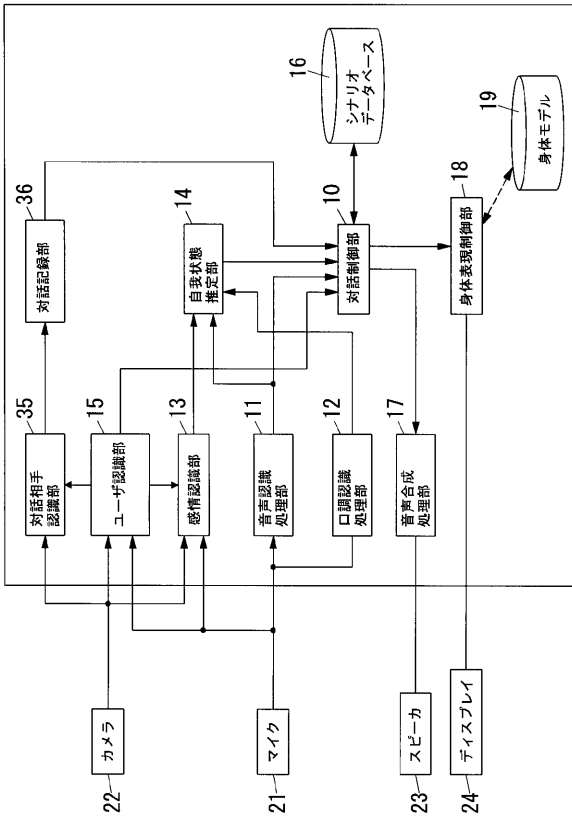
【 図 9 】



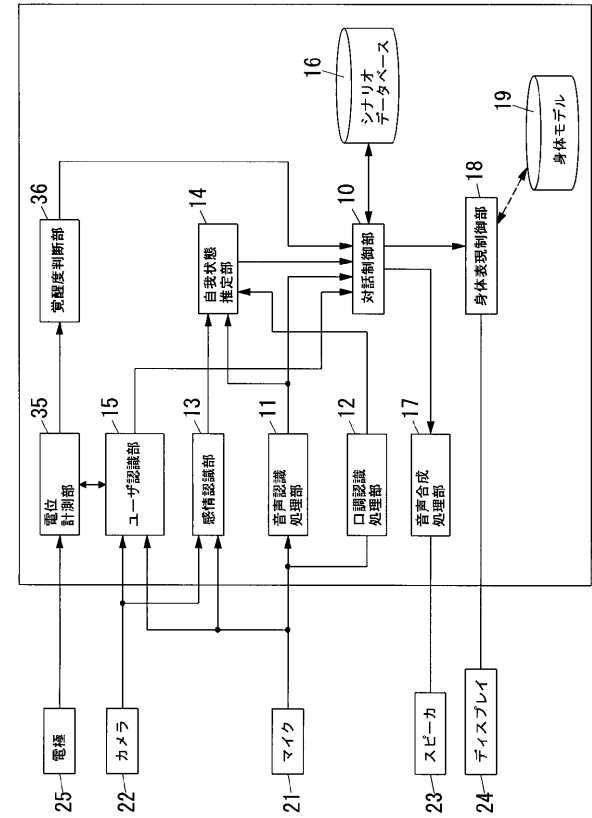
【 図 10 】



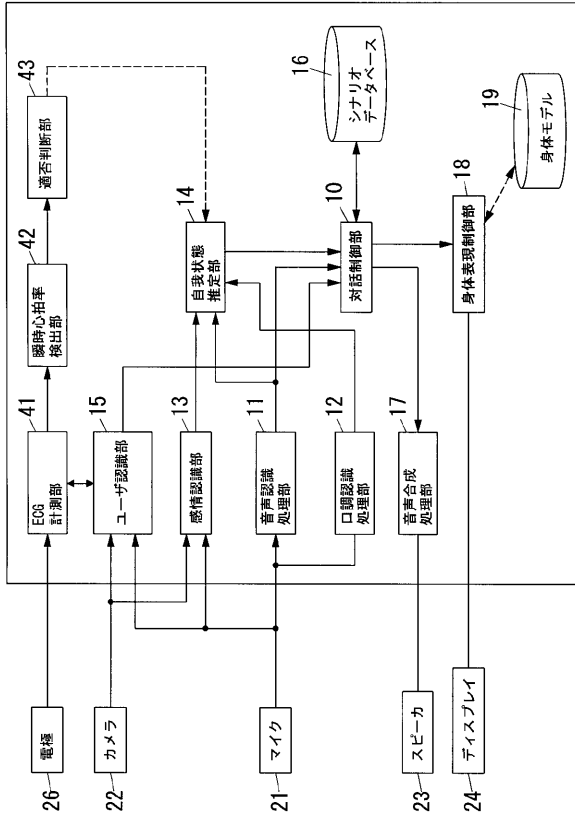
【 図 11 】



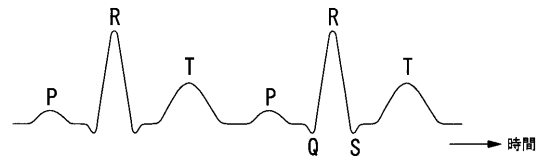
【 図 12 】



【 図 1 3 】



【 図 1 4 】





## フロントページの続き

(51)Int.Cl. <sup>7</sup>	F I		テーマコード(参考)
// A 6 1 B 5/16	G 1 0 L 3/00	H	
	G 1 0 L 9/00	L	
	G 1 0 L 3/00	R	
	A 6 1 B 5/16		

(72)発明者 馬場 朗

大阪府門真市大字門真 1 0 4 8 番地 松下電工株式会社内

Fターム(参考) 4C038 PP03 PS05 VC20

5D015 AA05 AA06 KK04 LL02 LL06 LL07

5D045 AA07 AB01