

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4641389号
(P4641389)

(45) 発行日 平成23年3月2日(2011.3.2)

(24) 登録日 平成22年12月10日(2010.12.10)

(51) Int. Cl.	F I	
G 0 6 T 7/00 (2006.01)	G O 6 T 7/00	3 5 O C
G 1 0 L 15/16 (2006.01)	G 1 0 L 15/16	
G 1 0 L 15/10 (2006.01)	G 1 0 L 15/10	5 0 O N
G 1 0 L 11/00 (2006.01)	G 1 0 L 11/00	4 0 2 H
G O 6 N 3/00 (2006.01)	G O 6 N 3/00	5 6 O C
請求項の数 11 (全 18 頁) 最終頁に続く		

(21) 出願番号	特願2004-166135 (P2004-166135)	(73) 特許権者	000001007
(22) 出願日	平成16年6月3日(2004.6.3)		キヤノン株式会社
(65) 公開番号	特開2005-346471 (P2005-346471A)		東京都大田区下丸子3丁目30番2号
(43) 公開日	平成17年12月15日(2005.12.15)	(74) 代理人	100076428
審査請求日	平成19年6月4日(2007.6.4)		弁理士 大塚 康德
		(74) 代理人	100112508
			弁理士 高柳 司郎
		(74) 代理人	100115071
			弁理士 大塚 康弘
		(74) 代理人	100116894
			弁理士 木村 秀二
		(74) 代理人	100130409
			弁理士 下山 治
		(74) 代理人	100134175
			弁理士 永川 行光
最終頁に続く			

(54) 【発明の名称】 情報処理方法、情報処理装置

(57) 【特許請求の範囲】

【請求項1】

映画のデータを再生する情報処理装置であって、
 前記映画のシーン毎に、該シーンを見る人が持つと思われる感情を示す感情情報を予め
 関連付けて記憶保持している記憶保持手段と、
 前記映画のデータを再生する再生手段と、
 前記映画のデータを見ているときのユーザの画像を撮像する撮像手段と、
 前記画像から得られる前記ユーザの特徴量に基づいて、前記映画のシーンを見ていると
 きの前記ユーザの感情を、予め設定された認識モデルを用いて推定する推定手段と、
 前記ユーザが前記シーンを見ているときに前記推定手段により推定される感情が、該シ
 ーンに関連付けて前記記憶保持手段に記憶保持されている感情情報が示す感情として推定
 されるように、前記認識モデルのパラメータを修正する修正手段と
 を備えることを特徴とする情報処理装置。

【請求項2】

前記推定手段は、前記特徴量を多層型ニューラルネットワークにおける入力層のニュー
 ロン群に入力し、出力層のニューロン群からの出力を前記ユーザの感情を示す情報として
 得ることを特徴とする請求項1に記載の情報処理装置。

【請求項3】

前記修正手段は、前記シーンを見ているときの前記ユーザの画像から得られる前記特徴
 量を前記入力層のニューロン群に入力した場合に、該シーンに関連付けて前記記憶保持手

段に記憶保持されている感情情報が出力されるように、ニューロン間の重み係数を修正することを特徴とする請求項 2 に記載の情報処理装置。

【請求項 4】

前記修正手段は、前記重み係数の修正を、バックプロパゲーションによって修正することを特徴とする請求項 3 に記載の情報処理装置。

【請求項 5】

前記推定手段は、感情毎に予め求めた前記特徴量の許容範囲を参照し、前記ユーザの画像から得られる前記特徴量を許容範囲に含む感情を推定結果として出力することを特徴とする請求項 1 に記載の情報処理装置。

【請求項 6】

更に、前記ユーザの音声情報を収集する収集手段を備え、

前記推定手段は前記画像から得られる特徴量、及び前記音声情報から得られる特徴量に基づいて、前記シーンを見ているときの前記ユーザの感情を、予め設定された認識モデルを用いて推定することを特徴とする請求項 1 乃至 5 の何れか 1 項に記載の情報処理装置。

【請求項 7】

ゲームを再生する情報処理装置であって、

前記ゲームの各進展具合を示す情報毎に、該進展具合において前記ゲームをしている人が持つと思われる感情を示す感情情報を予め関連付けて記憶保持している記憶保持手段と

前記ゲームを行っているときのユーザの画像を撮像する撮像手段と、

前記ユーザが行っているゲームの進展具合を監視する監視手段と、

前記画像から得られる前記ユーザの特徴量に基づいて、前記ゲームの各進展具合における前記ユーザの感情を、予め設定された認識モデルを用いて推定する推定手段と、

前記ゲームの各進展具合について前記推定手段により推定される感情が、前記記憶保持手段に該進展具合を示す情報に関連付けられて保持された感情情報が示す感情として推定されるように、前記認識モデルのパラメータを修正する修正手段と

を備えることを特徴とする情報処理装置。

【請求項 8】

映画のデータを再生する情報処理装置が行う情報処理方法であって、

前記情報処理装置が有する記憶保持手段が、前記映画のシーン毎に、該シーンを見る人が持つと思われる感情を示す感情情報を予め関連付けて記憶保持部に記憶保持する記憶保持工程と、

前記情報処理装置が有する再生手段が、前記映画のデータを再生する再生工程と、

前記情報処理装置が有する撮像手段が、前記映画のデータを見ているときのユーザの画像を撮像する撮像工程と、

前記情報処理装置が有する推定手段が、前記画像から得られる前記ユーザの特徴量に基づいて、前記映画のシーンを見ているときの前記ユーザの感情を、予め設定された認識モデルを用いて推定する推定工程と、

前記情報処理装置が有する修正手段が、前記ユーザが前記シーンを見ているときに前記推定工程により推定される感情が、該シーンに関連付けて前記記憶保持部に記憶保持されている感情情報が示す感情として推定されるように、前記認識モデルのパラメータを修正する修正工程と

を備えることを特徴とする情報処理方法。

【請求項 9】

ゲームを再生する情報処理装置が行う情報処理方法であって、

前記情報処理装置が有する記憶保持手段が、前記ゲームの各進展具合を示す情報毎に、該進展具合において前記ゲームをしている人が持つと思われる感情を示す感情情報を予め関連付けて記憶保持部に記憶保持する記憶保持工程と、

前記情報処理装置が有する撮像手段が、前記ゲームを行っているときのユーザの画像を撮像する撮像工程と、

10

20

30

40

50

前記情報処理装置が有する監視手段が、前記ユーザが行っているゲームの進展具合を監視する監視工程と、

前記情報処理装置が有する推定手段が、前記画像から得られる前記ユーザの特徴量に基づいて、前記ゲームの各進展具合における前記ユーザの感情を、予め設定された認識モデルを用いて推定する推定工程と、

前記情報処理装置が有する修正手段が、前記ゲームの各進展具合について前記推定工程により推定される感情が、前記記憶保持部に該進展具合を示す情報に関連付けられて保持された感情情報が示す感情として推定されるように、前記認識モデルのパラメータを修正する修正工程と

を備えることを特徴とする情報処理方法。

10

【請求項 10】

コンピュータを請求項 1 乃至 7 の何れか 1 項に記載の情報処理装置が有する各手段として機能させる為のコンピュータプログラム。

【請求項 11】

請求項 10 に記載のコンピュータプログラムを格納した、コンピュータ読みとり可能な記憶媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ユーザの表情や動作等といった反応を検出し、その検出した反応からユーザの感情を認識する為の技術に関するものである。

20

【背景技術】

【0002】

近年、機械と人間とのインターフェースとして、人間の意志や感情を認識し、人間とのコミュニケーションを行う人工エージェントの研究が進んでいる。また、様々なペットロボットが登場し、使用者の行動や感情等を認識し、その認識結果によりペットロボットの行動を変化させるものもあり、このようなペットロボットは一種の人工エージェントと言える。このような人工エージェントにおいては、カメラやマイクを用いて、人物の状態を検出し、検出した状態に基づいて人物の喜怒哀楽等といった感情を認識することが必要とされている。

30

【0003】

例えば、CCDカメラ等で撮影された画像から顔の表情を検出する装置が従来から開示されている（例えば特許文献 1 を参照）。これは、入力画像に対してウェーブレット変換を行い、各周波数領域での平均電力と無表情の時の平均電力との差分から、表情を検出するものである。

【0004】

また、音声と画像の両方を用いて、被写体の顔の表情と感情を検出する技術が従来から開示されている（例えば非特許文献 1 を参照）。

【0005】

さらに、顔の変位に応じた特徴ベクトルと、表情の各カテゴリごとに用意したベクトル量子化を行うためのコードブックとを使用し、ベクトル量子化後のシンボル列に基づいてカテゴリを決定する技術が従来から開示されている（例えば特許文献 2 を参照）。

40

【0006】

さらにまた、笑顔画像に対して、眼や口や眉の位置やサイズ等の変動から、その笑顔を、快の笑顔・不快の笑顔・社交的な笑顔に分類する検討を行なう技術も開示されている（例えば非特許文献 2 を参照）。

【特許文献 1】特開平 8 249447 号公報

【特許文献 2】特許第 2839855 号

【非特許文献 1】「Lawrence S.Chen, Thomas S.Huang, Tsutomu Miyasato, Ryohei Nakatsu : "Multimodal Human Emotion/Expression Recognition", Proceedings of the Sec

50

ond International Conference on Automatic Face and Gesture Recognition pp366-371 (1998) 」

【非特許文献2】「片岡、岩口、佐治：“目と口の動きの追跡による笑顔の分類”、情報処理学会研究報告、2001-HI-95、pp109-116(2001) 」

【発明の開示】

【発明が解決しようとする課題】

【0007】

上記説明したように、カメラやマイクを用いて人物の状態（反応）を検出し、さらに検出した状態（反応）に基づいて人物の喜怒哀楽等といった感情を認識する研究が進められている。しかし、検出した状態（反応）と感情とを正確に対応させるのは容易ではない。

10

【0008】

例えば、反応の一つとして表情を考えれば、表情の豊かな人もいれば、そうでない人もいる。つまり、同じ感情を持った人が全員同じ表情をするわけではない。逆に言うと、同じ表情をした人たちが、同じ感情を持っているとは限らない。そのため、表情といった人物の状態（反応）と感情との対応は、個人毎に設定する必要がある。上記従来例においては、ユーザ個々人の個人差に関しては言及がない。

【0009】

本発明は以上の問題に鑑みてなされたものであり、表情に代表される人物の状態（反応）と感情との対応を、個人毎に簡便に設定する為の技術を提供することを目的とする。

【課題を解決するための手段】

20

【0010】

本発明の目的を達成するために、例えば本発明の情報処理装置は以下の構成を備える。

【0011】

すなわち、映画のデータを再生する情報処理装置であって、
前記映画のシーン毎に、該シーンを見る人が持つと思われる感情を示す感情情報を予め関連付けて記憶保持している記憶保持手段と、
前記映画のデータを再生する再生手段と、
前記映画のデータを見ているときのユーザの画像を撮像する撮像手段と、
前記画像から得られる前記ユーザの特徴量に基づいて、前記映画のシーンを見ているときの前記ユーザの感情を、予め設定された認識モデルを用いて推定する推定手段と、
前記ユーザが前記シーンを見ているときに前記推定手段により推定される感情が、該シーンに関連付けて前記記憶保持手段に記憶保持されている感情情報が示す感情として推定されるように、前記認識モデルのパラメータを修正する修正手段と
を備えることを特徴とする。

30

【0012】

本発明の目的を達成するために、例えば本発明の情報処理装置は以下の構成を備える。

【0013】

すなわち、ゲームを再生する情報処理装置であって、
前記ゲームの各進展具合を示す情報毎に、該進展具合において前記ゲームをしている人が持つと思われる感情を示す感情情報を予め関連付けて記憶保持している記憶保持手段と
、
前記ゲームを行っているときのユーザの画像を撮像する撮像手段と、
前記ユーザが行っているゲームの進展具合を監視する監視手段と、
前記画像から得られる前記ユーザの特徴量に基づいて、前記ゲームの各進展具合における前記ユーザの感情を、予め設定された認識モデルを用いて推定する推定手段と、
前記ゲームの各進展具合について前記推定手段により推定される感情が、前記記憶保持手段に該進展具合を示す情報に関連付けられて保持された感情情報が示す感情として推定されるように、前記認識モデルのパラメータを修正する修正手段と
を備えることを特徴とする。

40

【0014】

50

本発明の目的を達成するために、例えば本発明の情報処理方法は以下の構成を備える。

【0015】

すなわち、映画のデータを再生する情報処理装置が行う情報処理方法であって、

前記情報処理装置が有する記憶保持手段が、前記映画のシーン毎に、該シーンを見る人が持つと思われる感情を示す感情情報を予め関連付けて記憶保持部に記憶保持する記憶保持工程と、

前記情報処理装置が有する再生手段が、前記映画のデータを再生する再生工程と、

前記情報処理装置が有する撮像手段が、前記映画のデータを見ているときのユーザの画像を撮像する撮像工程と、

前記情報処理装置が有する推定手段が、前記画像から得られる前記ユーザの特徴量に基づいて、前記映画のシーンを見ているときの前記ユーザの感情を、予め設定された認識モデルを用いて推定する推定工程と、

前記情報処理装置が有する修正手段が、前記ユーザが前記シーンを見ているときに前記推定工程により推定される感情が、該シーンに関連付けて前記記憶保持部に記憶保持されている感情情報が示す感情として推定されるように、前記認識モデルのパラメータを修正する修正工程と

を備えることを特徴とする。

【0016】

本発明の目的を達成するために、例えば本発明の情報処理方法は以下の構成を備える。

【0017】

すなわち、ゲームを再生する情報処理装置が行う情報処理方法であって、

前記情報処理装置が有する記憶保持手段が、前記ゲームの各進展具合を示す情報毎に、該進展具合において前記ゲームをしている人が持つと思われる感情を示す感情情報を予め関連付けて記憶保持部に記憶保持する記憶保持工程と、

前記情報処理装置が有する撮像手段が、前記ゲームを行っているときのユーザの画像を撮像する撮像工程と、

前記情報処理装置が有する監視手段が、前記ユーザが行っているゲームの進展具合を監視する監視工程と、

前記情報処理装置が有する推定手段が、前記画像から得られる前記ユーザの特徴量に基づいて、前記ゲームの各進展具合における前記ユーザの感情を、予め設定された認識モデルを用いて推定する推定工程と、

前記情報処理装置が有する修正手段が、前記ゲームの各進展具合について前記推定工程により推定される感情が、前記記憶保持部に該進展具合を示す情報に関連付けられて保持された感情情報が示す感情として推定されるように、前記認識モデルのパラメータを修正する修正工程と

を備えることを特徴とする。

【発明の効果】

【0018】

本発明の構成により、表情に代表される人物の状態（反応）と感情との対応を、個人毎に簡便に設定することができる。

【発明を実施するための最良の形態】

【0019】

以下添付図面を参照して、本発明を好適な実施形態に従って詳細に説明する。

【0020】

[第1の実施形態]

図1は、本実施形態に係る情報処理システムの機能構成を示す図である。

【0021】

同図において100はユーザで、以下では、このユーザの表情と感情との対応関係を設定する処理について説明する。本実施形態に係るシステムは、カメラ、101、マイク102、制御部103、反応計測部104、感情認識部105、刺激データ提示部106、

10

20

30

40

50

感情データ保持部 107、感情モデル修正部 108、感情モデル 109 により構成されている。

【0022】

以下、システムを構成する各部について簡単に説明する。

【0023】

カメラ 101 はユーザの様子を動画像として撮像するものである。マイク 102 は、ユーザの音声を収集するものである。カメラ 101、マイク 102 により得られた画像情報、音声情報は、後段の反応計測部 104 に出力される。

【0024】

制御部 103 は、同図に示したシステムを構成する各部を制御するものである。反応計測部 104 は、カメラ 101 から得られた画像情報、もしくはマイク 102 から得られた音声情報を用いて、ユーザ 100 の反応を計測する。この反応とは例えば、ユーザ 100 の顔における眼や口等の位置や大きさ、また声の大きさ等といったものである。

【0025】

感情認識部 105 は、反応計測部 104 が計測したユーザ 100 の反応を示す情報、及び感情モデル 109 を用いて、ユーザ 100 の感情を認識する。この認識方法については後述する。

【0026】

刺激データ提示部 106 は、刺激データに従った画像や音声をユーザ 100 に提示（再生）するものである。刺激データには、この刺激データに従った画像や音声を一般のユーザに提示した場合に、一般のユーザが抱くとおぼしき感情を示すデータ（情報）が関連付けられている。この刺激データについては詳しくは後述する。感情データ保持部 107 は、上記刺激データ、及びこの刺激データに関連付けられている感情を示すデータを保持する。

【0027】

感情モデル修正部 108 は、刺激データ提示部 106 によってユーザ 100 に刺激データを提示した場合に、感情認識部 105 が認識した感情が、提示した刺激データに関連付けられた感情として認識されるように、感情モデル 109 のパラメータを修正するものである。感情モデル 109 は、反応計測部 104 からの情報を入力とし、この入力に基づいて、ユーザ 100 の感情を認識する為のモデルである。

【0028】

図 7 は、本実施形態に係るシステムのハードウェア構成を示す図である。本実施形態では、カメラ 101、マイク 102 を PC（パーソナルコンピュータ）や WS（ワークステーション）などのコンピュータに接続し、一連の上記認識処理、感情モデル 109 のパラメータ修正処理をこのコンピュータにおいて行うとする。

【0029】

701 は CPU で、RAM 702 や ROM 703 に格納されているプログラムやデータを用いてコンピュータ全体の制御を行うと共に、I/F 707 を介してカメラ 101 から入力される各フレームのデータ、マイク 102 から入力されるアナログ信号を受ける処理をも行う。また、後述する一連の認識処理、感情モデル 109 のパラメータ修正処理をも行う。CPU 701 は、図 1 では制御部 103 として機能するものである。

【0030】

702 は RAM で、外部記憶装置 706 に保存されており、CPU 701 の制御によりロードされたプログラムやデータを一時的に記憶するためのエリアを備えると共に、CPU 701 が各種の処理を行うために使用するワークエリアも備える。また、I/F 707 を介してカメラ 101、マイク 102 から入力される画像情報、音声情報を一時的に記憶するためのエリアを備える。

【0031】

703 は ROM で、コンピュータを起動するためのプログラムやデータなどを格納する。

10

20

30

40

50

【 0 0 3 2 】

7 0 4 は操作部で、キーボードやマウスなどにより構成されており、各種の指示を C P U 7 0 1 に入力することができる。

【 0 0 3 3 】

7 0 5 は表示部で、C R T や液晶画面などにより構成されており、画像や文字などを表示することができる。また、表示部 7 0 5 は音声出力も可能である。表示部 7 0 5 は、図 1 では刺激データ提示部 1 0 6 として機能するものである。

【 0 0 3 4 】

7 0 6 は外部記憶装置で、ハードディスク装置などの大容量情報記憶装置として機能するものであり、ここに O S (オペレーティングシステム) や C P U 7 0 1 に後述する一連の処理を実行させるためのプログラムやデータを保存しておくことができる。本実施形態では、反応計測部 1 0 4、感情認識部 1 0 5、感情モデル 1 0 9、感情モデル修正部 1 0 8 は全てプログラムやデータなどにより構成され、外部記憶装置 7 0 6 に保存されており、C P U 7 0 1 の制御に従って R A M 7 0 2 にロードされるものであるとする。また、外部記憶装置 7 0 6 は、図 1 では、感情データ保持部 1 0 7 として機能するものである。

【 0 0 3 5 】

7 0 7 は I / F で、カメラ 1 0 1 やマイク 1 0 2 を接続することができ、この I / F 7 0 7 を介してカメラ 1 0 1 から入力される各フレームのデータ、マイク 1 0 2 から入力されるアナログ信号を受けることができる。なお、I / F 7 0 7 には、カメラ 1 0 1 やマイク 1 0 2 からのアナログ信号をデジタルデータに変換するための A / D 変換器が内蔵されているものとする。

【 0 0 3 6 】

7 0 8 は上述の各部を繋ぐバスである。

【 0 0 3 7 】

以下では、このコンピュータが行う感情認識処理、及びユーザ 1 0 0 の感情を認識するための感情モデル 1 0 9 のパラメータの修正処理について、これらの処理のフローチャートを示す図 3 を参照して詳細に説明する。

【 0 0 3 8 】

ここで、本実施形態では、表示部 7 0 5 によってユーザ 1 0 0 に提示するものを映画とする。すなわち、上記刺激データは、映画のデータとする。従って刺激データは、映画の各フレームの画像データ、及び音声データにより構成される。また、刺激データには、映画の各シーン毎に、そのシーンを見ている人が(一般に)抱くとおぼしき感情を示す情報が関連付けられている。

【 0 0 3 9 】

このように刺激データとして映画を使用することに対してはいくつかの長所がある。一つは、多数の人が同じ体験をすることが出来るので、平均的な反応と個人の特異な反応との比較がしやすい。つまり、日常生活の中での経験に基づいて個人ごとの設定を行おうとすると、状況がまったく同じであることはまれであるので比較がしにくい。映画を使用することで、そのような問題を回避できる。

【 0 0 4 0 】

また、一般的に映画は多くのシーンがあり映画を一通り見ると、基本的な感情全てに対する反応が得られる。さらに、ユーザが同じ感情を持つ複数のシーンがあるので、ある特異なシーンのみで感情モデルが作成されるのではなく、ユーザが同じ感情をもつ複数のシーンに対する平均的な反応を使用して、感情モデルを作成するため、精度の高い感情モデルが作成できる。

【 0 0 4 1 】

図 2 は、各シーン毎に感情を示すデータを関連付けた感情データシートの構成例を示す図である。例えば映画開始から 1 0 分 2 1 秒目のフレームから 1 2 分 2 0 秒目のフレームまでのシーンを見た場合、多くの人には「怒り」の感情を抱くというように統計が取れているので、同図に示す如く、1 0 分 2 1 秒目のフレームから 1 2 分 2 0 秒目のフレームまで

10

20

30

40

50

のシーンには、「怒り」の感情を示すデータが関連付けられている。これは他のシーンについても同様である。

【 0 0 4 2 】

このような感情データシートのデータは予め作成され、外部記憶装置 7 0 6 に保存されており、必要に応じて R A M 7 0 2 にロードされる。

【 0 0 4 3 】

そして C P U 7 0 1 は外部記憶装置 7 0 6 に保存されている映画のデータ、各シーンに関連付けられている感情を示すデータ、上記感情データシートのデータを R A M 7 0 2 にロードし、映画のデータを再生して表示部 7 0 5 に映画の各フレームの画像情報、音声情報を出力（再生）する（ステップ S 3 0 1）。

10

【 0 0 4 4 】

ユーザ 1 0 0 はこの提示された映画の画像を見、提示された音声情報を聞くのであるが、このときのユーザの様子（反応）はカメラ 1 0 1、マイク 1 0 2 により収集される。カメラ 1 0 1 はユーザ 1 0 0 の動画像を撮像しており、撮像した各フレームの画像のデータは I / F 7 0 7 を介して順次 R A M 7 0 2 に出力される。同様にマイク 1 0 2 はユーザ 1 0 0 から発せられる音声を収集しており、収集した音声信号は I / F 7 0 7 を介して R A M 7 0 2 に出力される。

【 0 0 4 5 】

C P U 7 0 1 は、カメラ 1 0 1、マイク 1 0 2 から R A M 7 0 2 に出力された画像情報、音声情報を用いて、映画を見ているユーザ 1 0 0 の反応を計測する（ステップ S 3 0 2）。

反応の計測とは上述の通り、ユーザ 1 0 0 の顔の眼・口・鼻といった顔を構成する部品の画像上における位置や大きさ、またそれらの変化、その他にも手や足の画像上における位置やその位置の変化、またユーザ 1 0 0 の発する音生の大きさ等といった特徴量が、後述する感情モデル 1 0 9 への入力に使用されるパラメータ計測値として計測される。なお、このような特徴量の内容については特に限定するものではないし、またその計測、算出方法については周知のものであるので、これ以上の説明は省略する。

20

【 0 0 4 6 】

そして C P U 7 0 1 は、得られた計測値に基づいてユーザ 1 0 0 の感情を認識（推定）する（ステップ S 3 0 3）。すなわち、ステップ S 3 0 1 で提示したシーンを見、聞いているユーザ 1 0 0 の感情を認識する。この認識のために、感情モデル 1 0 9 を用いる。感情モデル 1 0 9 は、プログラムやデータの形態で外部記憶装置 7 0 6 から R A M 7 0 2 にロードされており、C P U 7 0 1 がこれを用いることで、以下説明する認識処理を行うことができる。

30

【 0 0 4 7 】

感情モデル 1 0 9 としては例えば周知のニューラルネットワークを用いることができる。図 4 は、周知の階層型ニューラルネットワークの構成例を示す図である。同図に示す如く、階層型ニューラルネットワークは、入力層、中間層、出力層の 3 層構造でもって構成されるもので、入力層に入力された情報は中間層でもって処理され、その処理結果は出力層でもって更に処理されて、出力層を構成する各ニューロンから出力される。

【 0 0 4 8 】

入力層を構成するニューロンの数は、ステップ S 3 0 2 で求めた特徴量の数だけ用意される。従って入力層を構成する各ニューロンには、ステップ S 3 0 2 で求めたそれぞれの特徴量（ユーザ 1 0 0 の顔の眼・口・鼻といった顔を構成する部品の位置や大きさ、またそれらの変化、その他にも手や足の位置やその位置の変化、またユーザ 1 0 0 の発する音生の大きさ等）が入力される。

40

【 0 0 4 9 】

一方、出力層を構成する各ニューロンは、それぞれが異なる感情を示す為のもので、入力層を構成する各ニューロンに、ステップ S 3 0 2 で求めたそれぞれの特徴量を入力すると、出力層を構成するニューロンの何れか 1 つが発火する。従って、発火したニューロンに割り当てられた感情が、認識結果とする。例えば出力層を構成するニューロンの数を 7

50

つとすると、各ニューロンに“幸福”“怒り”“悲しみ”“嫌悪”“驚き”“恐れ”“感情なし”の何れか1つを重複なしに割り当てる。そして例えば“驚き”を割り当てられたニューロンのみが発火した場合には、認識結果は“驚き”となる。なお、感情の種類はこれに限定するものではないし、これに伴って出力層を構成するニューロンの数もこれに限定するものではない。

【0050】

なお、周知の通り、各層間のニューロン間の重み係数はすでに設定されているのであるが、この重み係数は、平均的な人の反応データのパラメータ計測値が入力されたときに、平均的な人の感情を出力するように初期設定されている。この初期設定は、多数の人に同じ刺激データを提示し、その時の反応データのパラメータ計測値を入力層への入力信号とし、そしてそのときの平均的な感情を教師信号として学習することにより、行われる。

10

【0051】

これにより、様々な人の反応データをニューラルネットワークに入力しても、出力層からは、おおよその人が抱くであろう感情を示すニューロンのみが発火するようになる。

【0052】

しかし、人によっては、例えば怒っている場合の顔の特徴量をニューラルネットワークに入力しても、悲しみの感情を割り当てられたニューロンが発火してしまうことがある。これは、この人の表情が平均的な表情ではない場合等に生ずる。

【0053】

従って以下で詳しく説明するが、感情モデル109としてニューラルネットワークを用いる場合には、あるシーンを見たユーザ100の反応データをニューラルネットワークに入力した場合の出力層におけるニューロンの発火パターンが、このシーンに関連付けられている感情を示す出力層におけるニューロンの発火パターンに一致するように、ニューロン間の重み係数を修正する必要がある。

20

【0054】

一方、感情モデル109としてニューラルネットワーク以外のモデルを用いた場合について説明する。ニューラルネットワーク以外のモデルとして、ステップS302で計測されるユーザ100の反応データのパラメータ計測値の存在範囲を、各感情毎に規定している感情モデルを使用する。

【0055】

図5は、このようなモデルが規定する各反応パラメータの範囲の例を示す図である。同図に示した感情モデルシート501には、前述した“幸福”・・・といった各感情ごとに、眼や口等、及びそれらを構成するさらに小さな部位の位置やその変化といった反応データのパラメータ計測値の存在範囲が記述されている。例えば、“幸福”という感情は、眼に関しては、目の中心位置と目尻の位置関係が規定されている。原点を画像の左上にとり、眼の垂直方向の長さを E_v 、眼の中心位置を E_x, E_y 、とすると、目尻の垂直方向の位置 E_{ey} は、

$$E_{ey} = E_y + a \times E_v \quad \text{且つ} \quad E_{ey} < E_y + b \times E_v \quad (a = 0.1, b = 1.0)$$

という条件式に従っている。つまり、この式から、“幸福”という感情をしめす時には目尻が下がっているということがわかる。

40

【0056】

逆にいうと、ユーザ100の反応データにおいて、眼の中心位置及び目尻の位置というパラメータ計測値がこの式を満足したときに、ユーザ100が“幸福”という状態である可能性があるということである。そして、この感情モデル109を用いて、この他にも、眼・口等の構成部品それぞれに対して求められたパラメータ計測値が、それぞれの位置関係を示す式を満足しているかどうかを調べ、満足した式が最も多い感情をユーザ100の感情として認識する。

【0057】

このように各感情時の顔の構成部品の位置や変化、また手や足の動作や声の大きさ等の存在範囲を規定することで、モデルシート上で各感情を表現することができる。なお、上

50

記式におけるパラメータ（係数）（ a ， b ）の値は、平均的な人の反応データのパラメータ計測値が入力されたときに、平均的な人の感情を出力するように初期設定されている。当然、他の関係式におけるパラメータについても同様に初期設定されている。

【0058】

これにより、様々な人の反応データを図5に示すようなモデルシートに記述された各条件式に当てはめてみても、およそその人が抱くであろう感情に該当する各条件式を満たすようになる。

【0059】

しかし、人によっては、例えば怒っている場合の顔の特徴量を各条件式に代入してみても、悲しみの感情に該当する条件式を満たすようになることがある。これは、この人の表情が平均的な表情ではない場合等に生ずる。

10

【0060】

従って以下で詳しく説明するが、感情モデル109として感情毎に設けた条件式群を用いる場合には、あるシーンを見たユーザ100の反応データを感情毎に設けた条件式群に代入した場合に、このシーンに関連付けられている感情に対して設けた条件式群を最も満たすように、各条件式群におけるパラメータを修正する必要がある。

【0061】

以上のようにして、感情モデル109に何れのモデルを用いたとしても、本実施形態では、あるシーンを見ているユーザ100の反応データに基づいて、ユーザ100の感情がこのシーンに関連付けられている感情として認識されるように、感情モデル109のパラメータを修正する必要がある。

20

【0062】

従ってステップS304では、感情モデル109のパラメータの修正処理を行う。

【0063】

先ず、感情モデル109としてニューラルネットワークを用いた場合のステップS304における処理について説明する。

【0064】

図8は、感情モデル109としてニューラルネットワークを用いた場合のステップS304における処理を説明する図である。ここでは出力層を構成するニューロンの数を3とするが、これに限定するものではない。

30

【0065】

同図において出力層を構成するニューロン801，802，803にはそれぞれ「怒り」、「悲しみ」、「幸福」の感情を割り当てているものとする。従ってニューロン801のみが発火する場合（「1 0 0」の出力パターンが出力層から出力された場合）、ニューラルネットワークは認識結果として「怒り」を出力したことになる。一方、ニューロン802のみが発火する場合（「0 1 0」の出力パターンが出力層から出力された場合）、ニューラルネットワークは認識結果として「悲しみ」を出力したことになる。一方、ニューロン803のみが発火する場合（「0 0 1」の出力パターンが出力層から出力された場合）、ニューラルネットワークは認識結果として「幸福」を出力したことになる。

40

【0066】

ここで、あるシーンを見ているユーザ100の反応データが入力層に入力され、出力層から「0.1 0.2 0.9」の出力パターンが出力され、更にこのシーンには「幸福」を示す情報が関連付けられているとする。

【0067】

この場合、このシーンを見て一般的にはユーザ100は「幸福」な感情を抱くと予想されるので、ユーザ100の感情の認識結果として「幸福」を示すパターン「0 0 1」が出力層から出力されることが好ましい。ところが、これとは異なる「0.1 0.2 0.9」の出力パターンが出力されたということは、これはすなわち、重み係数を修正し、ニューラルネットワークが上記反応データを入力した場合には所望の出力パターン（「

50

0 0 1」)を出力するようにする必要があるということである。

【0068】

従ってこのような場合には、上記反応データを入力層に入力することで出力層から出力される出力パターンと、教師信号としての上記所望の出力パターンとの差(誤差)を求め、この差に基づいて周知のバックプロパゲーション処理を行うことで、各層間のニューロン間の重み係数を修正する(学習処理)。このような処理を複数回繰り返して重み係数を修正することで、幸福なシーンを見ているユーザ100の反応データをニューラルネットワークに入力すると、「0 0 1」の出力パターンが出力層から出力される、すなわち、「幸福」の感情として認識されることになる。

【0069】

また、これは他のシーン(他の感情を抱かせるシーン)をユーザ100に見せ、そのときのユーザ100の感情を認識するために用いる重み係数の修正処理を行う場合についても同じである。なお、本実施形態では、1つのシーン(図2に示した各シーン)について上記誤差が所定値となるまで行ってから次にシーンに処理を移行しても良いし、各シーンについて1回ずつ順次行うようにしても良い。また、上記学習処理のアルゴリズムについては様々なものが適用可能であり、特に限定するものではない。また、学習処理の終了基準についても特に限定するものではない。

【0070】

次に、感情モデル109がパラメータ計測値の存在範囲を各感情毎に規定している感情モデルの場合のステップS304における処理について説明する。

【0071】

例えばユーザ100が「幸福」を示す情報が関連付けられたシーンを見ている場合、このときのユーザ100の反応データの存在範囲が、「幸福」の感情に該当する条件式を満たすように、条件式に含まれているパラメータを修正する。例えば幸福という感情として認識するための目に関する上記条件式

$$Eey = Ey + a \times Ev \quad \text{且つ} \quad Eey < Ey + b \times Ev$$

の場合、「幸福」を示す情報が関連付けられたシーンを見ているユーザ100の反応データに含まれる眼の垂直方向の長さ E_v 、眼の中心位置 E_x, E_y をこの式に代入し、この式を満たすように係数 a 、 b を修正する。

【0072】

なお、本実施形態では、1つのシーン(図2に示した各シーン)について条件式のパラメータ修正処理を行ってから次にシーンに処理を移行しても良いし、各シーンについて1回ずつ順次行うようにしても良い。

【0073】

以上の処理を映画の終了時まで行う(ステップS305)。

【0074】

以上説明した処理を行うことにより、各シーンに関連付けられた感情をユーザ100が抱いた場合に、どのような表情、手足の動き、声の大きさであるのかを感情モデル109に含まれるパラメータを修正することにより、学習することができ、その結果、学習後の感情モデル109は、ユーザ100の感情を、ユーザ100固有の視覚的な情報、聴覚的な情報をもって認識することができるモデルとなる。

【0075】

このようにユーザ個人毎の感情モデルを作成し、その個人ごとの感情モデルに基づいて、ユーザの反応に対して感情を認識することにより、正確にユーザの感情を認識することが出来る。つまり、一般的に同じ感情を抱いていても個人の反応(表情等)は異なるが、同じ刺激に対する各個人の反応を学習したため、個人の反応の差異に影響を受けない感情認識が可能になる。

【0076】

また、刺激データに映画を使用することで、前述のように、精度の高い感情モデルが作成できるため、精度の高い感情認識が可能になる。

10

20

30

40

50

【 0 0 7 7 】

また本実施形態では、感情モデル 1 0 9 のパラメータの初期値は、平均的な人の反応データのパラメータ計測値が入力されたときに、平均的な人の感情を出力するように初期設定されているので、初期値を全て 0 と設定するよりも、容易に個人ごとの感情モデルが作成できる。

【 0 0 7 8 】

また、各シーンごとに感情を認識するのではなく、ユーザ 1 0 0 の表情が変化したり、ユーザ 1 0 0 の動作が変化したとき等、ユーザ 1 0 0 の反応データのパラメータ計測値に変化が起こった時のみに、感情を認識するようにしてもよい。このようにすることで、反応の乏しい人に対して感情モデルを作成する時に、無反応の時の誤差を減少させることが出来る。

10

【 0 0 7 9 】

なお、本実施形態では、ユーザ 1 0 0 の画像情報と音声情報の両方を収集して感情を認識するようにしていたが、ユーザによっては如何なる感情を抱いても声を出さない場合もある。その場合には、画像情報のみを用いて感情を認識するようにしても良い。その場合には、マイク 1 0 2 は必要はないし、感情モデル 1 0 9 には、上記画像情報のみを入力することになる。

【 0 0 8 0 】

[第 2 の実施形態]

第 1 の実施形態では刺激データとして映画のデータとしたが、本実施形態ではテレビゲームのプログラム、及びデータとする。

20

【 0 0 8 1 】

つまり、ゲームを再生し、ゲームの進展具合と、その進展具合に対して人が抱く感情とに関して関連付けられた感情データを用意する。そして、ゲームの進展具合に対する感情データと、その進展具合に対するユーザの反応とから、そのユーザの感情モデルを第 1 の実施形態と同様にして作成する。

【 0 0 8 2 】

ここでいうゲームの進展具合とは、ある場面をクリアするのにどれくらい時間がかかっているか、またはある時間内にどれだけ敵を倒したか、またはある時間内に自分のキャラクタが敵に連続して負けたか等である。なお、本実施形態では、感情データ保持部 1 0 7 には、ゲームの進展具合とその時の感情とを関連付けた感情データを保持している。

30

【 0 0 8 3 】

図 6 は、本実施形態における感情データシートの構成例を示している。

【 0 0 8 4 】

感情データシート 6 0 1 では、前述のようなゲームの進展具合と感情とを関連付けている。この関連付けは、多数の人にゲームをやってもらい、そのときのゲームの進展具合に応じた感情をまとめた平均的な感情データから作成される。

【 0 0 8 5 】

よって CPU 7 0 1 は、同図に示した各進展具合についてユーザ 1 0 0 のプレイを監視（クリア時間は？ある時間内にどの程度の数の敵を倒したか？等を監視）し、同図の感情データシートにおいて該当する進展具合を特定する。特定した各進展具合には感情が関連付けられているので、それぞれの進展具合に関連付けられた感情のうち、最も多いものを認識結果として出力する。すなわち、ここで認識したのは、「この進展具合なら、ユーザ 1 0 0 はこのような感情を抱くであろう」感情（推定感情）である。

40

【 0 0 8 6 】

一方で、第 1 の実施形態と同様にして、カメラ 1 0 1、マイク 1 0 2 からの画像情報、音声情報に基づいて感情モデル 1 0 9 によりユーザ 1 0 0 の感情を認識する。そして、認識した感情と上記推定感情（第 1 の実施形態では教師信号に該当）とを参照し、第 1 の実施形態と同様にして感情モデル 1 0 9 のパラメータを修正する。

【 0 0 8 7 】

50

以上の説明からもわかるとおり、第1の実施形態では、各シーン毎に感情を示す情報が予め関連付けられて感情データシートの形態で記憶保持されていたので、各シーン毎に認識されるべき感情はこの感情データシートを参照することにより得られたのであるが、本実施形態では、刺激データとしてゲームのプログラム、データを用いたことにより、感情データシートの構成、及びこの感情データシートを用いたユーザ100の感情を認識する方法は異なる。

【0088】

すなわち、CPU701は常にユーザ100のゲームの進展具合を監視し、所定の時間毎に感情データシートのどの進展具合に該当するのかを判断する。そして該当する進展具合に関連付けられた感情のうち、最も多いものを認識結果とする。これにより、CPU701は、「この進展具合なら、ユーザ100はこのような感情を抱くであろう」と判断して、感情を認識することができる。

10

【0089】

このように、現在のユーザ100の感情がどのようなものであるかを得るために映画の代わりにテレビゲームを用いることで、感情データシートの構成、及びこの感情データシートを用いたユーザ100の感情を認識する方法が異なる以外は、本実施形態は第1の実施形態と同じである。

【0090】

このように刺激データとしてテレビゲームを使用すると以下のような長所がある。すなわち、多数の人がゲームの進展に沿って、ほぼ同じ体験をすることが出来るので、平均的な反応と個人の特異な反応との比較がしやすい。さらに、同じ感情を持つ状態がいくつかあるので、ある特異な状態のみで感情モデルが作成されるのではなく、同じ感情をもつ複数の状態に対する平均的な反応を使用して感情モデルを作成するため、精度の高い感情モデルが作成できる。

20

【0091】

また、テレビゲームはインタラクティブ性があり、ユーザが映画よりも更に主体的に取り組むので、声を出したりして、ユーザの反応も強くなり、喜怒哀楽の感情が出やすい。そのため、ユーザの反応データのパラメータ計測が行いやすい。しかし、感情モデルの作成及び修正時には、入力されるユーザの反応データのパラメータ計測値に対して出力される感情の強さは最大のレベルとする必要がある。つまり、ゲームを実行中の反応データの

30

【0092】

なお、本実施形態では、ユーザ100の画像情報と音声情報の両方を収集して感情を認識するようにしていたが、ユーザによっては如何なる感情を抱いても声を出さない場合もある。その場合には、画像情報のみを用いて感情を認識するようにしても良い。その場合には、マイク102は必要はないし、感情モデル109には、上記画像情報のみを入力することになる。

【0093】

なお、ユーザ100の反応データを得る形態は上記実施形態以外にも考えられる。例えば、第1の実施形態ではシーンを見ているユーザ100の画像情報、音声情報を反応データとして収集したが、シーンを見ているユーザの脈拍数を計測し、計測した脈拍数が単位時間あたりどの程度の脈拍数であるかによって、感情を推定するようにしても良い。例えば興奮時には単位時間あたりの脈拍数は多くなる等、脈拍数は感情の部分情報になりうるので、この脈拍数を反応データとして用いることも可能である。また、反応データとして以上説明したものを適宜組み合わせる用いるようにしても良い。

40

【0094】

[その他の実施形態]

本発明の目的は、前述した実施形態の機能を実現するソフトウェアのプログラムコードを記録した記録媒体(または記憶媒体)を、システムあるいは装置に供給し、そのシステ

50

ムあるいは装置のコンピュータ（またはCPUやMPU）が記録媒体に格納されたプログラムコードを読み出し実行することによっても、達成されることは言うまでもない。この場合、記録媒体から読み出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコードを記録した記録媒体は本発明を構成することになる。

【0095】

また、コンピュータが読み出したプログラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているオペレーティングシステム（OS）などが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

10

【0096】

さらに、記録媒体から読み出されたプログラムコードが、コンピュータに挿入された機能拡張カードやコンピュータに接続された機能拡張ユニットに備わるメモリに書込まれた後、そのプログラムコードの指示に基づき、その機能拡張カードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0097】

本発明を上記記録媒体に適用する場合、その記録媒体には、先に説明したフローチャート（機能構成）に対応するプログラムコードが格納されることになる。

20

【図面の簡単な説明】

【0098】

【図1】本発明の第1の実施形態に係る情報処理システムの機能構成を示す図である。

【図2】各シーン毎に感情を示すデータを関連付けた感情データシートの構成例を示す図である。

【図3】本発明の第1の実施形態に係るコンピュータが行う感情認識処理、及びユーザ100の感情を認識するための感情モデル109のパラメータの修正処理のフローチャートである。

【図4】周知の階層型ニューラルネットワークの構成例を示す図である。

【図5】ステップS302で計測されるユーザ100の反応データのパラメータ計測値の存在範囲を、各感情毎に規定している感情モデルが規定する各反応パラメータの範囲の例を示す図である。

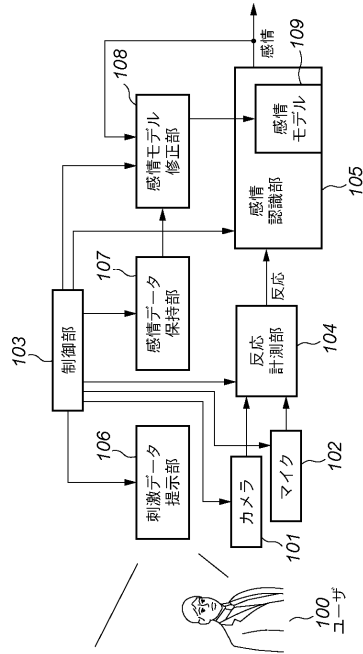
30

【図6】本発明の第2の実施形態における感情データシートの構成例を示している。

【図7】本発明の第1の実施形態に係るシステムのハードウェア構成を示す図である。

【図8】感情モデル109としてニューラルネットワークを用いた場合のステップS304における処理を説明する図である。

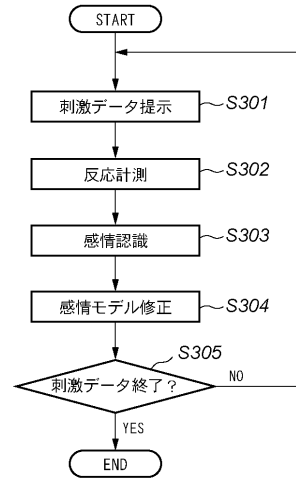
【図1】



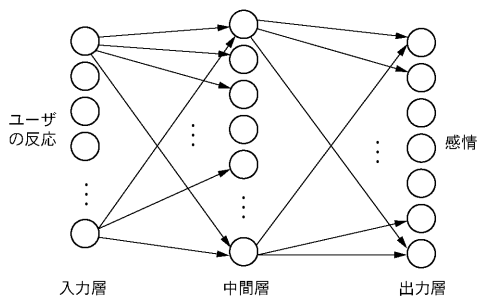
【図2】

シーン	感情
0:10:21~0:12:20	怒り
0:11:00~0:13:10	悲しみ
⋮	⋮
1:50:44~1:58:30	幸福感

【図3】



【図4】



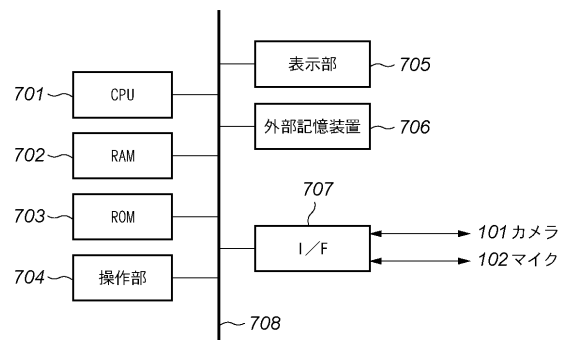
【図6】

進展具合	感情	
クリアする時間Tc	Tc < 0.3	幸福
	0.5 < Tc < 0.9	嫌悪
	Tc > 0.9	怒り
⋮	⋮	⋮
負けつづけているか	Yes	悲しみ 嫌悪 怒り

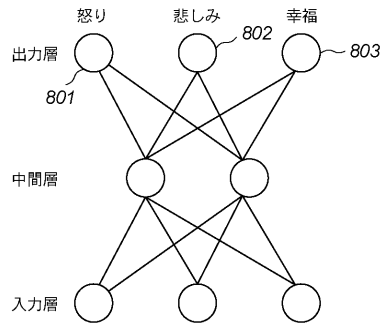
【図5】

感情	構成部品	位置関係
幸福	眼	・ 目尻の位置と眼の中心位置及び大きさとの関係 $Eey \geq Ey + a * Ev$ && $Eey < Ey + b * Ev$ (a=0.1, b=1.0)
	⋮	⋮
⋮	口	
⋮	⋮	⋮
悲しみ		

【図7】



【図 8】



フロントページの続き

(51)Int.Cl. F I
G 0 6 N 3/00 5 6 0 G

- (72)発明者 森 克彦
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
- (72)発明者 真継 優和
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
- (72)発明者 石井 美絵
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
- (72)発明者 御手洗 裕輔
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

審査官 松尾 淳一

- (56)参考文献 特開平02-165285(JP,A)
特開平06-231257(JP,A)
特開平07-019834(JP,A)
特開平08-249447(JP,A)
特開平11-219195(JP,A)
特開2001-051967(JP,A)
特開2001-209779(JP,A)
特開2003-099084(JP,A)
特開2003-162294(JP,A)
特開2003-178078(JP,A)
特開2003-330485(JP,A)
特開2004-064433(JP,A)
特開2004-144996(JP,A)
特開2004-153770(JP,A)
特開2005-063163(JP,A)
特開2005-128884(JP,A)
特開2004-178593(JP,A)
特許第2593625(JP,B2)
特許第2874858(JP,B2)
特許第3002136(JP,B2)
特許第3348956(JP,B2)
佐藤 貴志, 遠藤 祐樹, 金子 正人, 武内 惇, 藤本 洋, 音声に含まれる感情情報の認識に関する研究, 電子情報通信学会技術研究報告[ヒューマン情報処理], 日本, 電子情報通信学会, 2001年11月10日, 第101巻第512号, p.25~29, HIP2001-60
八尾 智幸, 小原 ゆう, 柴田 史久, 馬場口 登, 八木 康史, テレビ視聴行動からの個人的選好獲得のための行動認識, 電子情報通信学会技術研究報告[パターン認識・メディア理解], 日本, 電子情報通信学会, 2004年 1月16日, 第103巻第585号, p.65~70
佐藤 秀明, 赤松 則男, ニューラルネットワークによる感情音声の分類, 電子情報通信学会技術研究報告[ニューロコンピューティング], 日本, 電子情報通信学会, 2001年 6月22日, 第101巻第154号, p.85~90, NC2001-34

(58)調査した分野(Int.Cl., DB名)
G 0 6 F 3 / 0 1
G 0 6 F 3 / 0 4 8

G06F 15/18
G06F 19/00
G06N 1/00
G06N 3/00 - 3/12
G06N 7/08
G06Q 10/00
G06Q 30/00
G06Q 50/00
G06Q 90/00
G06T 7/00 - 7/60
G10L 11/00 - 11/06
G10L 13/00 - 13/08
G10L 15/00 - 15/28
G10L 17/00
G10L 19/00 - 19/14
G10L 21/00 - 21/06
H04N 7/10
H04N 7/14 - 7/173
H04N 7/20 - 7/22