

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第3676969号  
(P3676969)

(45) 発行日 平成17年7月27日(2005.7.27)

(24) 登録日 平成17年5月13日(2005.5.13)

(51) Int. Cl.<sup>7</sup>

F I

G 1 0 L 15/10

G 1 0 L 3/00 5 3 1 N

G 1 0 L 15/00

G 1 0 L 3/00 5 5 1 L

請求項の数 14 (全 16 頁)

(21) 出願番号	特願2000-278397 (P2000-278397)	(73) 特許権者	500430877
(22) 出願日	平成12年9月13日(2000.9.13)		株式会社エイ・ジー・アイ
(65) 公開番号	特開2002-91482 (P2002-91482A)		東京都品川区北品川5丁目13-15-1
(43) 公開日	平成14年3月27日(2002.3.27)		01
審査請求日	平成13年12月12日(2001.12.12)	(74) 代理人	100072718
			弁理士 古谷 史旺
		(72) 発明者	光吉 俊二
			東京都品川区北品川5丁目17-24 秀
			和御殿山レジデンス104
		審査官	荏原 雄一

最終頁に続く

(54) 【発明の名称】感情検出方法及び感情検出装置ならびに記録媒体

(57) 【特許請求の範囲】

【請求項1】

被験者の感情を検出するための感情検出方法であって、

音声信号を入力し、

入力した音声信号から音声の強度及び音声の出現速度を表すテンポをそれぞれ検出するとともに、音声の各単語内の強度変化パターンに出現する同一周波数成分の領域の時間間隔を抑揚として検出し、

検出された音声の強度の時間軸方向の変化を表す第1の変化量、音声のテンポの時間軸方向の変化を表す第2の変化量及び音声の抑揚の時間軸方向の変化を表す第3の変化量をそれぞれ求め、

前記第1の変化量、第2の変化量及び第3の変化量のパターンと、感情状態とを関連付ける情報を予め保持し、

前記関連付ける情報を参照することで、前記第1の変化量、第2の変化量及び第3の変化量から、感情状態を表す信号を生成する

ことを特徴とする感情検出方法。

【請求項2】

被験者の感情を検出するための感情検出装置であって、

音声信号を入力する音声入力手段と、

前記音声入力手段が入力した音声信号から音声の強度を検出する強度検出手段と、

前記音声入力手段が入力した音声信号から音声の出現速度をテンポとして検出するテン

が検出手段と、

前記音声入力手段が入力した音声信号から、音声の単語内の強度変化パターンに出現する同一周波数成分の領域の時間間隔を抑揚として検出する抑揚検出手段と、

前記強度検出手段が検出した音声の強度の時間軸方向の変化を表す第1の変化量、前記テンポ検出手段が検出した音声のテンポの時間軸方向の変化を表す第2の変化量及び前記抑揚検出手段が検出した音声の抑揚の時間軸方向の変化を表す第3の変化量をそれぞれ求める変化量検出手段と、

前記第1の変化量、第2の変化量及び第3の変化量のパターンと、感情状態とを関連付ける情報を予め保持する感情パターンデータベースと、

前記感情パターンデータベースの前記関連付ける情報を参照することで、前記第1の変化量、第2の変化量及び第3の変化量から、感情状態を表す信号を生成する感情検出手段と

10

を設けたことを特徴とする感情検出装置。

#### 【請求項3】

請求項2の感情検出装置において、前記抑揚検出手段に、

単語毎に分離されて入力される音声信号から特定の周波数成分を抽出するバンドパスフィルタ手段と、

前記バンドパスフィルタ手段により抽出された信号のパワースペクトルをその強度に基づいて複数の領域に分離する領域分離手段と、

前記領域分離手段により分離された複数の領域の各々の中心位置の時間間隔に基づいて抑揚の値を算出する抑揚計算手段と

20

を設けたことを特徴とする感情検出装置。

#### 【請求項4】

請求項2の感情検出装置において、

被験者の少なくとも顔の画像情報を入力する撮像手段と、

前記撮像手段が入力した画像情報から顔面各部に関する位置情報を検出する画像認識手段と、

顔面各部の特徴量の基準情報を保持する画像基準情報保持手段と、

前記画像認識手段の検出した位置情報と前記画像基準情報保持手段の保持する基準情報とに基づいて画像特徴量を検出する画像特徴量検出手段と、

30

を更に設けるとともに、前記感情検出手段が、前記画像特徴量検出手段の検出した画像特徴量の変化を判断材料に加えて感情状態を総合的に判断することを特徴とする感情検出装置。

#### 【請求項5】

請求項2の感情検出装置において、

前記感情検出手段の検出した感情状態の情報を逐次入力して蓄積する感情情報蓄積手段と、

前記感情情報蓄積手段に蓄積された過去の感情状態の情報のうち、記憶時点から所定の時間が経過した情報を削除するとともに、削除対象の情報のうち、予め定めた変化パターンに適合する情報については削除対象から除外する忘却処理手段と

40

を更に設けたことを特徴とする感情検出装置。

#### 【請求項6】

請求項5の感情検出装置において、

被験者の発した音声もしくは被験者の入力した文字の情報を処理して文法解析を行い文章の意味を表す発言情報を生成する文章認識手段と、

前記文章認識手段の生成した発言情報を、前記感情状態の情報と同期した状態で感情情報蓄積手段に蓄積する蓄積制御手段とを更に設けて、

感情状態の情報だけでなく、そのときの状況を表す発言情報を取り出すことを可能にした

ことを特徴とする感情検出装置。

50

## 【請求項 7】

請求項 2 の感情検出装置において、  
 検出された感情状態に基づいて基準無音時間を決定する無音時間決定手段と、  
 前記無音時間決定手段の決定した基準無音時間を利用して、音声の文章の区切りを検出する文章区切り検出手段と  
 を更に設けたことを特徴とする感情検出装置。

## 【請求項 8】

被験者の感情を検出するための計算機で実行可能な感情検出プログラムを記録した記録媒体であって、

前記感情検出プログラムには、

音声信号を入力する手順と、

入力した音声信号から音声の強度及び音声の出現速度を表すテンポをそれぞれ検出するとともに、音声の各単語内の強度変化パターンに出現する同一周波数成分の領域の時間間隔を抑揚として検出する手順と、

検出された音声の強度の時間軸方向の変化を表す第 1 の変化量、音声のテンポの時間軸方向の変化を表す第 2 の変化量及び音声の抑揚の時間軸方向の変化を表す第 3 の変化量をそれぞれ求める手順と、

前記第 1 の変化量、第 2 の変化量及び第 3 の変化量のパターンと、感情状態とを関連付ける情報を予め保持しておく手順と、

前記関連付ける情報を推定規則として参照することで、前記第 1 の変化量、第 2 の変化量及び第 3 の変化量から、感情状態を表す信号を生成する手順と

を設けたことを特徴とする記録媒体。

## 【請求項 9】

請求項 1 に記載の感情検出方法において、

怒り、悲しみ、および喜びのいずれか 1 つの感情状態を表す信号を生成することを特徴とする感情検出方法。

## 【請求項 10】

請求項 2 ないし請求項 7 のいずれか 1 項に記載の感情検出装置において、

前記感情検出手段は、

怒り、悲しみ、および喜びのいずれか 1 つの感情状態を表す信号を生成することを特徴とする感情検出装置。

## 【請求項 11】

請求項 8 に記載の記録媒体において、

怒り、悲しみ、および喜びのいずれか 1 つの感情状態を表す信号を生成することを特徴としたプログラムを記録した記録媒体。

## 【請求項 12】

被験者の感情の検出に使用する抑揚を、音声信号から検出する検出方法であって、

音声信号を入力し、

入力した前記音声信号の単語内の強度変化パターンから同一周波数成分の領域を検出し

、  
 前記同一周波数成分の領域が出現する時間間隔を検出して抑揚とする

ことを特徴とする抑揚の検出方法。

## 【請求項 13】

被験者の感情の検出に使用する抑揚を、音声信号から検出する検出装置であって、

音声信号を入力する音声入力手段と、

入力した前記音声信号の単語内の強度変化パターンから同一周波数成分の領域を検出する手段と、

前記同一周波数成分の領域が出現する時間間隔を検出して抑揚とする手段と、

を備えたことを特徴とする抑揚の検出装置。

## 【請求項 14】

10

20

30

40

50

被験者の感情の検出に使用する抑揚を、音声信号から検出するためのプログラムを記録した記録媒体であって、

コンピュータに、

音声信号を入力する手順と、

入力した前記音声信号の単語内の強度変化パターンから同一周波数成分の領域を検出する手順と、

前記同一周波数成分の領域が出現する時間間隔を検出して抑揚とする手順と

を実行させるためのプログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、人間の感情を検出するために用いる感情検出方法及び感情検出装置ならびに記録媒体に関する。本発明は、医療分野における感情検出にも利用できるし、人工知能や人工感性の一部として様々なシステムに利用することもできる。

【0002】

【従来の技術】

本発明に関連のある従来技術は、例えば特開平5-12023号公報、特開平9-22296号公報及び特開平11-119791号公報に開示されている。特開平5-12023号公報においては、音声の特徴量として、音声の継続時間、音声のフォルマント周波数及び音声の周波数毎の強度をそれぞれ検出している。また、各々の特徴量について基準信号とのずれを検出し、検出したずれ量からファジー推論により感情の検出を行うことを開示している。

【0003】

特開平9-22296号公報においては、音声の特徴量として、音声の発生速度（単位時間あたりのモーラ数）、音声ピッチ周波数、音量及び音声スペクトルを検出している。また、検出した音声の特徴量と、HMM（隠れマルコフモデル：Hidden Markov Model）の統計処理を行った結果とを用いて感情を検出することを開示している。

【0004】

特開平11-119791号公報においては、HMMを用いて音素スペクトルの遷移状態の確率に基づいて感情を検出することを開示している。

【0005】

【発明が解決しようとする課題】

しかしながら、従来の感情検出方法では感情の検出精度が低く、特定の限定された言葉について感情を検出できたとしても、実際の人間の感情を正確に検出できるものではない。従って、例えば比較的単純なゲーム装置の限定的な用途においてのみ感情検出方法が実用化されているのが実情である。

【0006】

本発明は、被験者である人間の感情をより正確に検出可能な感情検出方法及び感情検出装置ならびに記録媒体を提供することを目的とする。

【0007】

【課題を解決するための手段】

請求項1は、被験者の感情を検出するための感情検出方法であって、音声信号を入力し、入力した音声信号から音声の強度及び音声の出現速度を表すテンポをそれぞれ検出するとともに、音声の各単語内の強度変化パターンに出現する同一周波数成分の領域の時間間隔を抑揚として検出し、検出された音声の強度の時間軸方向の変化を表す第1の変化量、音声のテンポの時間軸方向の変化を表す第2の変化量及び音声の抑揚の時間軸方向の変化を表す第3の変化量をそれぞれ求め、第1の変化量、第2の変化量及び第3の変化量のパターンと、感情状態とを関連付ける情報を予め保持し、関連付ける情報を参照することで、第1の変化量、第2の変化量及び第3の変化量から、感情状態を表す信号を生成することを特徴とする。

10

20

30

40

50

## 【 0 0 0 8 】

請求項 1 においては、被験者から入力される音声の強度、テンポ及び抑揚の各々の変化量を、感情状態に対応付けて感情を検出している。このような方法を用いることにより、従来よりも正確に感情を検出することが可能である。

請求項 2 は、被験者の感情を検出するための感情検出装置であって、音声信号を入力する音声入力手段と、前記音声入力手段が入力した音声信号から音声の強度を検出する強度検出手段と、前記音声入力手段が入力した音声信号から音声の出現速度をテンポとして検出するテンポ検出手段と、前記音声入力手段が入力した音声信号から、音声の単語内の強度変化パターンに出現する同一周波数成分の領域の時間間隔を抑揚として検出する抑揚検出手段と、前記強度検出手段が検出した音声の強度の時間軸方向の変化を表す第 1 の変化量、前記テンポ検出手段が検出した音声のテンポの時間軸方向の変化を表す第 2 の変化量及び前記抑揚検出手段が検出した音声の抑揚の時間軸方向の変化を表す第 3 の変化量をそれぞれ求める変化量検出手段と、第 1 の変化量、第 2 の変化量及び第 3 の変化量のパターンと、感情状態とを関連付ける情報を予め保持する感情パターンデータベースと、感情パターンデータベースの前記関連付ける情報を参照することで、第 1 の変化量、第 2 の変化量及び第 3 の変化量から、感情状態を表す信号を生成する感情検出手段とを設けたことを特徴とする。

10

## 【 0 0 0 9 】

請求項 2 の感情検出装置においては、音声入力手段、強度検出手段、テンポ検出手段、抑揚検出手段、変化量検出手段及び感情検出手段を設けることにより、請求項 1 の感情検出方法を実施することができる。

20

請求項 3 は、請求項 2 の感情検出装置において、前記抑揚検出手段に、単語毎に分離されて入力される音声信号から特定の周波数成分を抽出するバンドパスフィルタ手段と、前記バンドパスフィルタ手段により抽出された信号のパワースペクトルをその強度に基づいて複数の領域に分離する領域分離手段と、前記領域分離手段により分離された複数の領域の各々の中心位置の時間間隔に基づいて抑揚の値を算出する抑揚計算手段とを設けたことを特徴とする。

## 【 0 0 1 0 】

バンドパスフィルタ手段は、単語毎に分離されて入力される音声信号から特定の周波数成分を抽出する。領域分離手段は、検出されたパワースペクトルをその強度に基づいて複数の領域に分離する。抑揚計算手段は、前記領域分離手段により分離された複数の領域の各々の中心位置の時間間隔に基づいて抑揚の値を算出する。

30

## 【 0 0 1 1 】

請求項 3 においては、音声の特定の周波数成分に関する単語内のエネルギー分布パターンを複数の領域の間隔を表す時間の値として検出し、その時間の長さを抑揚として利用している。

請求項 4 は、請求項 2 の感情検出装置において、被験者の少なくとも顔の画像情報を入力する撮像手段と、前記撮像手段が入力した画像情報から顔面各部に関する位置情報を検出する画像認識手段と、顔面各部の特徴量の基準情報を保持する画像基準情報保持手段と、前記画像認識手段の検出した位置情報と前記画像基準情報保持手段の保持する基準情報とに基づいて画像特徴量を検出する画像特徴量検出手段とを更に設けるとともに、前記感情検出手段が、前記画像特徴量検出手段の検出した画像特徴量の変化を判断材料に加えて感情状態を総合的に判断することを特徴とする。

40

## 【 0 0 1 2 】

請求項 4 においては、音声だけでなく、被験者の顔の表情に基づいて感情状態を推定している。一般に、人間の感情状態はその人の顔の表情に反映されるので、顔の表情を検出することにより感情状態を把握することができる。そこで、請求項 4 では、前記感情検出手段は画像特徴量検出手段の検出した画像特徴量の変化に基づいて感情状態を推定している。

## 【 0 0 1 3 】

50

請求項5は、請求項2の感情検出装置において、前記感情検出手段の検出した感情状態の情報を逐次入力して蓄積する感情情報蓄積手段と、前記感情情報蓄積手段に蓄積された過去の感情状態の情報のうち、記憶時点から所定の時間が経過した情報を削除するとともに、削除対象の情報のうち、予め定めた変化パターンに適合する情報については削除対象から除外する忘却処理手段とを更に設けたことを特徴とする。

【0014】

請求項5においては、検出された過去の感情状態の情報を感情情報蓄積手段に蓄積しておくことができる。また、検出してから長い時間の経過した古い情報については感情情報蓄積手段から自動的に削除されるので、感情情報蓄積手段に必要なとされる記憶容量を減らすことができる。

10

【0015】

但し、感情変化が所定以上に大きい情報や、予め定めた変化パターンに適合する情報のように特徴的な情報については削除対象から自動的に除外される。このため、特徴的な情報は古くなってもそのまま感情情報蓄積手段に保持される。従って、人間の記憶と同じように、後で役に立つ印象的な情報については古くなっても感情情報蓄積手段から読み出して再生することができる。

【0016】

請求項6は、請求項5の感情検出装置において、被験者の発した音声もしくは被験者の入力した文字の情報を処理して文法解析を行い文章の意味を表す発言情報を生成する文章認識手段と、前記文章認識手段の生成した発言情報を、前記感情状態の情報と同期した状態

20

【0017】

文章認識手段は、被験者の発した音声もしくは被験者がキーボードなどを用いて入力した文字の情報を処理して文法解析を行い文章の意味を表す発言情報を生成する。

文法解析により、例えば「5W3H」、すなわち「誰が」、「何を」、「いつ」、「どこで」、「なぜ」、「どうやって」、「どのくらい」、「いくら」を表す発言情報を得ることができる。

【0018】

蓄積制御手段は、前記文章認識手段の生成した発言情報を、前記感情状態の情報と同期した状態で感情情報蓄積手段に蓄積する。

30

請求項6においては、感情情報蓄積手段を参照することにより、過去の任意の時点における感情情報だけでなく、そのときの状況を表す発言情報を取り出すことができる。

【0019】

感情情報蓄積手段に保持された情報については、様々な用途で利用することができる。例えば、感情検出装置自体の感情推定機能が不正確であった場合には、感情情報蓄積手段に保持された過去の検出結果に基づいて感情推定に利用されるデータベースを修正することができる。

【0020】

請求項7は、請求項2の感情検出装置において、検出された感情状態に基づいて基準無音時間を決定する無音時間決定手段と、前記無音時間決定手段の決定した基準無音時間を利用して、音声の文章の区切りを検出する文章区切り検出手段とを更に設けたことを特徴とする。

40

音声の認識や感情の検出などを行う場合には、文章毎の区切りを検出してそれぞれの文章を抽出する必要がある。一般的には、文章と文章との区切りには無音区間が存在するので、無音区間が現れたタイミングで複数の文章を分離すればよい。

【0021】

しかしながら、無音区間の長さは一定ではない。特に、話者の感情の状態に対応して無音区間の長さは変化する。このため、無音区間の判定のために一定の閾値を割り当てた場合には、文章の区切りの検出に失敗する可能性が高くなる。

請求項7においては、例えば直前に検出された感情状態を利用して基準無音時間を決定し

50

、この基準無音時間を用いて音声の文章の区切りを検出するので、話者の感情が変化した場合であっても正しく文章の区切りを検出できる。

【 0 0 2 2 】

請求項 8 は、被験者の感情を検出するための計算機で実行可能な感情検出プログラムを記録した記録媒体であって、前記感情検出プログラムには、音声信号を入力する手順と、入力した音声信号から音声の強度及び音声の出現速度を表すテンポをそれぞれ検出するとともに、音声の各単語内の強度変化パターンに出現する同一周波数成分の領域の時間間隔を抑揚として検出する手順と、検出された音声の強度の時間軸方向の変化を表す第 1 の変化量、音声のテンポの時間軸方向の変化を表す第 2 の変化量及び音声の抑揚の時間軸方向の変化を表す第 3 の変化量をそれぞれ求める手順と、第 1 の変化量、第 2 の変化量及び第 3 の変化量のパターンと、感情状態とを関連付ける情報を予め保持しておく手順と、関連付ける情報を推定規則として参照することで、第 1 の変化量、第 2 の変化量及び第 3 の変化量から、感情状態を表す信号を生成する手順とを設けたことを特徴とする。

10

【 0 0 2 3 】

請求項 8 の記録媒体に記録された感情検出プログラムを計算機を用いて実行することにより、請求項 1 の感情検出方法を実施することができる。

請求項 9 の感情検出方法は、請求項 1 に記載の感情検出方法において、怒り、悲しみ、および喜びのいずれか 1 つの感情状態を表す信号を生成することを特徴とする。

請求項 10 の感情検出装置は、請求項 2 ないし請求項 7 のいずれか 1 項に記載の感情検出装置において、感情検出手段は、怒り、悲しみ、および喜びのいずれか 1 つの感情状態を表す信号を生成することを特徴とする。

20

請求項 11 の記録媒体は、請求項 8 に記載の記録媒体において、怒り、悲しみ、および喜びのいずれか 1 つの感情状態を表す信号を生成することを特徴としたプログラムを記録する。

請求項 12 の抑揚の検出方法は、被験者の感情の検出に使用する抑揚を、音声信号から検出する検出方法であって、音声信号を入力し、入力した前記音声信号の単語内の強度変化パターンから同一周波数成分の領域を検出し、前記同一周波数成分の領域が出現する時間間隔を検出して抑揚とすることを特徴とする。

請求項 13 の抑揚の検出装置は、被験者の感情の検出に使用する抑揚を、音声信号から検出する検出装置であって、音声信号を入力する音声入力手段と、前記音声信号の単語内の強度変化パターンから同一周波数成分の領域を検出する手段と、前記同一周波数成分の領域が出現する時間間隔を検出して抑揚とする手段とを備えたことを特徴とする。

30

請求項 14 の記録媒体は、被験者の感情の検出に使用する抑揚を、音声信号から検出するためのプログラムを記録した記録媒体であって、コンピュータに、音声信号を入力する手順と、前記音声信号の単語内の強度変化パターンから同一周波数成分の領域を検出する手順と、前記同一周波数成分の領域が出現する時間間隔を検出して抑揚とする手順とをコンピュータに実行させるためのプログラムが記録されていることを特徴とする。

【 0 0 2 4 】

【 発明の実施の形態 】

本発明の感情検出方法及び感情検出装置の 1 つの実施の形態について、図 1 ~ 図 6 を参照して説明する。この形態は全ての請求項に対応する。

40

【 0 0 2 5 】

図 1 は、この形態の感情検出装置の構成を示すブロック図である。図 2 は抑揚検出部の構成を示すブロック図である。図 3 は感情の状態の変化と音声の強度、テンポ及び抑揚との関係を示すグラフである。図 4 は抑揚検出部における音声信号処理の過程を示すタイムチャートである。図 5 は忘却処理部の動作を示すフローチャートである。図 6 は感情記憶 DB に記憶された情報の構成例を示す模式図である。

【 0 0 2 6 】

この形態では、請求項 2 の音声入力手段、強度検出手段、テンポ検出手段、抑揚検出手段、変化量検出手段及び感情検出手段は、それぞれマイク 11、強度検出部 17、テンポ検

50

出部 18, 抑揚検出部 19, 感情変化検出部 22 及び音声感情検出部 23 に対応する。また、請求項 3 のバンドパスフィルタ手段, 領域分離手段及び抑揚計算手段は、それぞれバンドパスフィルタ 51, 比較部 53 及び領域間隔検出部 55 に対応する。請求項 4 の撮像手段, 画像認識手段, 画像基準情報保持手段, 画像特徴量検出手段及び感情検出手段は、それぞれテレビカメラ 31, 画像認識部 32, 顔パターン DB 33, 顔感情検出部 34 及び顔感情検出部 34 に対応する。

【0027】

更に、請求項 5 の感情情報蓄積手段及び忘却処理手段はそれぞれ感情感性記憶 DB 41 及び忘却処理部 42 に対応する。請求項 6 の文章認識手段及び蓄積制御手段は、それぞれ文章認識部 26 及び同期処理部 43 に対応する。請求項 7 の無音時間決定手段及び文章区切り検出手段は文章検出部 16 に対応する。

10

図 1 を参照すると、この感情検出装置にはマイク 11, A/D 変換器 12, 信号処理部 13, 音声認識部 20, 強度検出部 17, テンポ検出部 18, 抑揚検出部 19, 一時記憶部 21, 感情変化検出部 22, 音声感情検出部 23, 感情パターン DB (データベースの略: 以下同様) 24, キーボード 25, 文章認識部 26, テレビカメラ 31, 画像認識部 32, 顔パターン DB 33, 顔感情検出部 34, 文字認識部 39, 感情感性記憶 DB 41, 忘却処理部 42, 同期処理部 43, 人間性情報 DB 44, 個人情報 DB 45, 専門情報 DB 46 及び感情認識部 60 が備わっている。

【0028】

また、音声認識部 20 には信号処理部 13, 音素検出部 14, 単語検出部 15 及び文章検出部 16 が設けてある。音声認識部 20 には、市販の音声認識 (事前言語) デバイスの機能も含まれている。

20

図 1 において、音声認識部 20, 強度検出部 17, テンポ検出部 18, 抑揚検出部 19, 一時記憶部 21, 感情変化検出部 22 及び音声感情検出部 23 は、音声から感情を検出するための回路である。

【0029】

この感情検出装置は、感情の検出対象となる相手の人間の情報を読み取るための入力手段として、マイク 11, キーボード 25 及びテレビカメラ 31 を備えている。すなわち、マイク 11 から入力される音声, キーボード 25 から入力される文字情報及びテレビカメラ 31 から入力される顔の表情などの情報を利用して相手の人間の感情を検出する。

30

【0030】

なお、実際にはマイク 11 から入力される音声だけに基づいて感情を検出することも可能であり、キーボード 25 から入力される文字情報だけに基づいて感情を検出することも可能であり、テレビカメラ 31 から入力される顔の表情だけに基づいて相手の人間の感情を検出することも可能である。しかし、複数の情報源から得られる情報を総合的に判断した方が感情の検出精度を高めるうえで効果的である。

【0031】

まず、音声に関する処理について説明する。マイク 11 から入力された音声信号は、A/D 変換器 12 でサンプリングされ、デジタル信号に変換される。A/D 変換器 12 の出力に得られる音声のデジタル信号は、音声認識部 20 に入力される。

40

【0032】

信号処理部 13 は、音声の強度検出に必要な周波数成分を抽出する。強度検出部 17 は、信号処理部 13 の抽出した信号からその強度を検出する。例えば、音声信号の振幅の大きさを平均化した結果を強度として利用することができる。

音声の強度を検出するための平均化の周期については、例えば 10 秒程度に定める。但し、10 秒以内であっても文章毎の区切りを検出した場合には、文章の最初から区切りを検出した時点までの平均化を行う。すなわち、音声の文章毎にそれぞれの強度を検出する。

【0033】

音声認識部 20 に備わった音素検出部 14 は、入力される音声の音素毎の区切りを検出する。例えば、「今日はいい天気ですね」の文章が音声で入力された場合には、「きょ/う

50

／は／い／い／て／ん／き／で／す／ね」のように音素毎の区切りを検出する。

また、音声認識部 20 に備わった単語検出部 15 は、入力される音声の単語毎の区切りを検出する。例えば、「今日はいい天気ですね」の文章が音声で入力された場合には、「きょう／は／いい／てんき／ですね」のように単語毎の区切りを検出する。

#### 【0034】

また、音声認識部 20 に備わった文章検出部 16 は、入力される音声の文章毎の区切りを検出する。特定の長さ以上の無音状態を検出した場合に、文章毎の区切りが現れたものとみなす。無音状態の長さの閾値には、(0.1~2)秒程度の値が割り当てられる。また、この閾値は一定ではなく、直前に検出された感情の状態を反映するように自動的に変更される。

10

#### 【0035】

テンポ検出部 18 は、音素検出部 14 から出力される音素毎の区切りの信号を入力して、単位時間に現れた音素の数をテンポとして検出する。テンポの検出周期については、例えば 10 秒程度の時間が割り当てられる。しかし、文章の区切りを検出した場合には、10 秒以内であってもその時点まで音素数のカウントを中止してテンポの値を計算する。つまり、文章毎にテンポが検出される。

#### 【0036】

抑揚検出部 19 には、単語検出部 15 が区切りを検出した単語毎に区分されて、音声信号が入力される。抑揚検出部 19 は、入力される音声信号から各単語内及び文章検出部 16 における文章毎の区切り内の音声の強度変化パターンを表す抑揚を検出する。これにより、抑揚検出部 19 は区切りの中での特徴的な強度パターンを検出する。

20

#### 【0037】

抑揚検出部 19 の内部には、図 2 に示すように、バンドパスフィルタ 51, 絶対値変換部 52, 比較部 53, 領域中心検出部 54 及び領域間隔検出部 55 が備わっている。また、抑揚検出部 19 における各部の信号 SG1, SG2, SG3, SG4 の波形の例が図 4 に示されている。なお、図 4 における各信号の縦軸は振幅又は強度を表している。また、図 4 の例では音声から取り出された 1 つの単語の長さが約 1.2 秒になっている。

#### 【0038】

バンドパスフィルタ 51 は、入力された信号 SG1 の中から抑揚の検出に必要な周波数成分だけを抽出する。この例では、800 Hz ~ 1200 Hz の範囲内の周波数成分だけがバンドパスフィルタ 51 の出力に信号 SG2 として現れる。図 4 を参照すると、単語内の抑揚による強度変化のパターンが信号 SG2 に現れていることが分かる。

30

#### 【0039】

信号の計算処理を容易にするために、抑揚検出部 19 には絶対値変換部 52 を設けてある。絶対値変換部 52 は、入力される信号の振幅をその絶対値に変換する。従って、絶対値変換部 52 の出力には図 4 に示す信号 SG3 が現れる。

比較部 53 は、信号 SG3 の大きさを閾値と比較して閾値よりも大きい成分だけを信号 SG4 として出力する。すなわち、比較部 53 は信号 SG3 のパワースペクトルの中で値の大きな成分だけを出力する。なお、比較部 53 に印加する閾値については、判別分析法と呼ばれる方法を用いて適応的に決定している。

40

#### 【0040】

図 4 を参照すると、信号 SG4 には音声の単語における抑揚パターンに相当する 2 つの領域 A1, A2 が明確に現れている。領域中心検出部 54 は、2 つの領域 A1, A2 のそれぞれの中心に相当する位置が現れた時間 t1, t2 を検出する。

領域間隔検出部 55 は、領域中心検出部 54 の検出した 2 つの時間 t1, t2 に関する時間差を領域間隔 Ty として検出する。この領域間隔 Ty の値は、音声の単語における抑揚パターンに相当する。実際には、領域間隔 Ty の値を平均化した結果を抑揚の値として利用している。

#### 【0041】

なお、1 つの単語の中で信号 SG4 に 3 つ以上の領域が現れる場合もある。3 つ以上の領

50

域が現れた場合には、互いに隣接する2つの領域について領域間隔 $T_y$ をそれぞれ計算し、求められた複数の領域間隔 $T_y$ を平均化した結果を抑揚の値として利用する。

人間の感情の状態は、例えば図3に示すように変化する。また、怒り、悲しみ、喜びなどの感情を正しく把握するためには、強度、テンポ、抑揚のような特徴量の変化を検出することが重要である。

#### 【0042】

図1に示す感情検出装置においては、過去の特徴量の参照を可能にするため、強度検出部17が出力する強度、テンポ検出部18が出力するテンポ及び抑揚検出部19が出力する抑揚の値を一時的に一時記憶部21に記憶しておく。

また、感情変化検出部22は、強度検出部17が出力する現在の強度、テンポ検出部18が出力する現在のテンポ及び抑揚検出部19が出力する現在の抑揚の値と、一時記憶部21に保持された過去の（現在よりも少し前の時刻の）強度、テンポ及び抑揚の値とを入力して、感情状態の変化を検出する。つまり、音声の強度の変化、テンポの変化及び抑揚の変化をそれぞれ検出する。

10

#### 【0043】

音声感情検出部23は、感情変化検出部22が出力する音声の強度の変化、テンポの変化及び抑揚の変化を入力し、現在の感情の状態を推定する。感情の状態として、この例では怒り、悲しみ及び喜びの3種類の状態をそれぞれ推定している。

#### 【0044】

感情パターンDB24には、音声の強度の変化、テンポの変化及び抑揚の変化のパターンと怒りの状態とを関連付ける情報と、音声の強度の変化、テンポの変化及び抑揚の変化のパターンと悲しみの状態とを関連付ける情報と、音声の強度の変化、テンポの変化及び抑揚の変化のパターンと喜びの状態とを関連付ける情報とが予め保持されている。

20

#### 【0045】

音声感情検出部23は、感情パターンDB24に保持された情報を推定規則として参照しながら、感情変化検出部22が出力する強度の変化、テンポの変化及び抑揚の変化のパターンに基づいて現在の感情の状態を推定する。

音声感情検出部23によって推定された怒り、悲しみ及び喜びの3種類の各々の状態を表す情報は、感情認識部60及び感情感性記憶DB41に入力される。感情感性記憶DB41は、音声感情検出部23から入力される現在の感情の状態を逐次記憶され、蓄積される。

30

#### 【0046】

従って、感情感性記憶DB41に記憶された情報を読み出すことにより、過去の感情の状態を再生することができる。

一方、音声としてマイク11から入力された文章の内容（相手の発言内容）は、文章認識部26で認識される。文章認識部26の入力には、音声認識部20で認識された各音素に対応する文字情報や、単語の区切り及び文章の区切りを表す情報が入力される。また、キーボード25から入力された文字情報も文章認識部26に入力される。

#### 【0047】

文章認識部26は、入力される文字列の単語毎の認識及び構文解析を行い、文章の内容を自然言語として把握する。実際には、「5W3H」、すなわち「誰が」、「何を」、「いつ」、「どこで」、「なぜ」、「どうやって」、「どのくらい」、「いくら」を表す発言情報を認識する。文章認識部26が認識した発言情報は感情認識部60に入力される。

40

#### 【0048】

次に、相手の顔の表情から感情を検出するための処理について説明する。テレビカメラ31は、図1の感情検出装置の被験者となる人間の少なくとも顔の部分撮影する。テレビカメラ31の撮影した画像、すなわち人間の顔の表情が含まれる画像が画像認識部32に入力される。

なお、テレビカメラ31の撮影した画像の情報は文字認識部39に入力される。すなわち、文章の映像をテレビカメラ31で撮影した場合には、文字認識部39は撮影された映像

50

から文章の各文字を認識する。文字認識部 3 9 の認識した文字情報は文章認識部 2 6 に入力される。

【 0 0 4 9 】

画像認識部 3 2 は、入力される画像の中から特徴的な要素を認識する。具体的には、被験者の顔における目、口、眉毛、頬骨の部分をそれぞれ認識し、顔の中における目、口、眉毛、頬骨のそれぞれの相対的な位置を検出する。また、画像認識部 3 2 は顔の表情の変化に伴う目、口、眉毛、頬骨のそれぞれの位置の変化及び首を振るなどの表現を検出するために位置の追跡を常に行う。

【 0 0 5 0 】

顔パターン D B 3 3 には、顔の中における目、口、眉毛、頬骨のそれぞれの位置に関する基準位置の情報（被験者の平常時の顔の表情に相当する情報）が予め保持されている。なお、顔パターン D B 3 3 の内容を任意に変更することも可能である。また、顔パターン D B 3 3 には顔の表情の変化と 6 種類の感情（喜び、怒り、悲しみ、恐れ、楽しみ、驚き）のそれぞれとの対応関係を表す規則情報が予め保持されている。

10

【 0 0 5 1 】

顔感情検出部 3 4 は、画像認識部 3 2 が認識した目、口、眉毛、頬骨のそれぞれの位置と顔パターン D B 3 3 に保持された基準位置の情報とを用いて特徴量、すなわち平常時の位置に対する表情の違いを検出する。

また、顔感情検出部 3 4 は検出した特徴量の変化量及び変化の速さと、顔パターン D B 3 3 に保持された規則情報とに基づいて、6 種類の感情（喜び、怒り、悲しみ、恐れ、楽しみ、驚き）のそれぞれの状態を推定する。推定された 6 種類の感情の状態を表す情報は、顔感情検出部 3 4 から出力されて感情認識部 6 0 及び感情感性記憶 D B 4 1 に入力される。

20

【 0 0 5 2 】

感情認識部 6 0 は、音声感情検出部 2 3 から入力される感情（怒り、悲しみ、喜び）の状態を表す情報と、文章認識部 2 6 から入力される発言情報と、顔感情検出部 3 4 から入力される感情（喜び、怒り、悲しみ、恐れ、楽しみ、驚き）の状態を表す情報とを総合的に判断して最終的な感情の状態を推定する。発言情報については、その文章の内容（5 W 3 H）を予め定めた規則に従って判断することにより、発言情報に含まれている感情（喜び、怒り、悲しみ、恐れ、楽しみ、驚き）の状態を推定することができる。

30

【 0 0 5 3 】

音声感情検出部 2 3 が音声から推定した感情の状態を表す情報と、文章認識部 2 6 が音声又はキーボード 2 5 から入力された文字から認識した発言内容の情報と、顔感情検出部 3 4 が顔の表情から推定した感情の状態を表す情報とが、それぞれ感情感性記憶 D B 4 1 に入力されて逐次記憶される。感情感性記憶 D B 4 1 に記憶されたそれぞれの情報には、それが検出された時刻あるいは時間ならびに年月日が付加される。

【 0 0 5 4 】

感情感性記憶 D B 4 1 に入力される情報のうち、音声感情検出部 2 3 から入力される感情の情報と、文章認識部 2 6 から入力される発言内容の情報と、顔感情検出部 3 4 から入力される感情の情報とは互いに関連付けて把握しなければならない。

40

そこで、同期処理部 4 3 は感情感性記憶 D B 4 1 に蓄積された複数種類の情報を、それらの検出された時間（入力された時間）及び年月日によって互いに関連付ける。例えば、図 6 に示されるように、音声感情検出部 2 3 の推定した怒り、悲しみ及び喜びの感情の状態を表す情報と発言の内容（5 W 3 H）の情報とを、それらの時間によって互いに関連付ける。

【 0 0 5 5 】

ところで、感情感性記憶 D B 4 1 には比較的大量の情報を蓄積できる十分な記憶容量が備わっている。しかしながら、記憶容量には限りがあるのでこの装置を長期間に渡って使い続けるためには蓄積する情報の量を抑制する必要がある。

【 0 0 5 6 】

50

そこで、忘却処理部 4 2 が設けてある。忘却処理部 4 2 は、古くなった情報を感情感性記憶 DB 4 1 上から自動的に削除する。但し、特定の条件に適合する情報については古くなった場合でも削除せずに保存される。

忘却処理部 4 2 の動作について、図 5 を参照しながら説明する。

図 5 のステップ S 1 1 においては、感情感性記憶 DB 4 1 に蓄積されている多数のデータのそれぞれについて、記憶された時刻（あるいは検出された時刻）及び年月日の情報を参照する。

【 0 0 5 7 】

ステップ S 1 2 では、現在の時刻とステップ S 1 1 で参照したデータの時刻とに基づいて、該当するデータが記憶されてから予め定めた一定の期間が経過したか否かを識別する。記憶してから一定の期間が経過した古いデータを処理する場合には、ステップ S 1 3 以降の処理に進む。一定の期間が経過していない比較的新しいデータについては、そのまま保存される。

10

【 0 0 5 8 】

ステップ S 1 3 では、データが感情の状態を表す情報である場合に、その感情の変化量を感情状態を表す情報の前後の違いから検出する。感情の変化量が予め定めた閾値を超える場合にはステップ S 1 3 から S 1 7 に進むので、そのデータが古い場合であってもそのままデータは保存される。感情の変化量が閾値以下の場合には、ステップ S 1 3 から S 1 4 に進む。

【 0 0 5 9 】

20

ステップ S 1 4 では、そのデータに関する感情のパターンを検出し、そのパターンが予め定めた特定のパターンと一致するか否かを識別する。すなわち、複数の感情の状態及び発言内容の組み合わせが、「印象が強い」状態を表す特定のパターンと一致するか否かを調べる。検出したパターンが特定のパターンと一致した場合には、ステップ S 1 4 から S 1 7 に進むので、そのデータが古い場合であってもそのままデータは保存される。パターンが一致しない場合にはステップ S 1 4 から S 1 5 に進む。

【 0 0 6 0 】

ステップ S 1 5 では、データが発言内容である場合に、その内容と予め定めた発言内容（印象に残りやすい発言）とが一致するか否かを識別する。なお、完全に一致しなくても、類似性が高い場合には「一致」とみなすこともできる。データの発言内容が予め定めた発言内容と一致した場合には、ステップ S 1 5 から S 1 7 に進むので、そのデータが古い場合であっても、そのままデータは保存される。

30

【 0 0 6 1 】

ステップ S 1 5 で一致しない場合には、ステップ S S 1 6 において当該データは削除される。

上記の処理は感情感性記憶 DB 4 1 上の全てのデータについて実行される。また、図 5 に示す忘却処理は定期的に繰り返し実行される。この忘却処理を実行留周期については、個人の個性として任意に変更することができる。なお、ステップ S 1 4 , S 1 5 では予め容易されたパターン DB（図示せず）を参照して処理を行う。このパターン DB については、入力情報を学習することにより自動的に内容が更新される。

40

【 0 0 6 2 】

なお、図 5 では処理を簡略化して表してある。実際には、感情の変化量、感情のパターン及び発言の内容の全てを総合的に判断する。すなわち、感情の変化量が大きい情報と、感情のパターンが一致した情報と、発言内容が同一もしくは近似する情報とが存在する場合には、総合的に優先順位を判断する。具体的には、発言内容が同一もしくは近似する情報の優先順位が最も大きく、感情のパターンが一致した情報の優先順位が 2 番目に高く、感情の変化量が大きい情報の優先順位は低い。従って、発言内容が同一もしくは近似する情報は忘却処理で削除されにくく、古くなっても記憶として残る。

【 0 0 6 3 】

上記のような忘却処理部 4 2 の処理によって、感情感性記憶 DB 4 1 上の古くなったデー

50

タについては、感情の変化が大きいもの、「印象が強い」とみなされるパターンであるもの、幾度も入力を繰り返されたもの、及び発言の内容が印象に残りやすいもののみがその強度と内容に合わせて順位をつけてそのまま保存される。その結果、感情感性記憶DB41上の古いデータについては、一部分のみが残った不完全なデータとなる。このようなデータは、人間の記憶における過去の曖昧な記憶と同じような内容になる。

#### 【0064】

感情感性記憶DB41に蓄積された過去の感情の状態及び発言内容を読み出してデータを分析することにより、例えばこの感情検出装置が正しく動作しているか否かを判断したり、感情の推定に利用される各部のデータベースの内容を改良するように更新することも可能になる。

10

感情感性記憶DB41に蓄積されたデータについては、その内容に応じて更に振り分けられ、人間性情報DB44、個人情報DB45又は専門情報DB46に記憶される。

#### 【0065】

人間性情報DB44には、性別、年齢、攻撃性、協調性、現在の感情などのように被験者の性格を決定付ける情報や行動の決定パターンの情報が保持される。また、個人情報DB45には、個人の住所、現在の状況、環境、発言内容(5W3H)などの情報が保持される。専門情報DB46には、職業、経歴、職業適性格、職業的行動決定パターンなどの情報が保持される。

#### 【0066】

人間性情報DB44、個人情報DB45及び専門情報DB46から出力されるのは、個人のモラルパターン情報である。このモラルパターン情報と過去の相手の感情とに基づいて相手の感性を察知することができる。

20

なお、図1に示す感情検出装置の機能をコンピュータのソフトウェアにより実現する場合には、コンピュータが実行するプログラム及び必要なデータを、例えばCD-ROMなどの記録媒体に記録しておけばよい。

#### 【0067】

なお、図1に示すマイク11を電話機の受話器に置き換えてもよいし、文字などの情報を入力する手段としてマウスを設けてもよい。

また、図1に示すテレビカメラ31については、光学式カメラ、デジタルカメラ、CCDカメラのような様々な撮像手段のいずれでも置き換えることができる。

30

#### 【0068】

##### 【発明の効果】

本発明の感情検出方法及び感情検出装置ならびに記録媒体によれば、より正確に被験者の感情を検出することができる。

##### 【図面の簡単な説明】

【図1】実施の形態の感情検出装置の構成を示すブロック図である。

【図2】抑揚検出部の構成を示すブロック図である。

【図3】感情の状態の変化と音声の強度、テンポ及び抑揚との関係を示すグラフである。

【図4】抑揚検出部における音声信号処理の過程を示すタイムチャートである。

【図5】忘却処理部の動作を示すフローチャートである。

40

【図6】感情感性記憶DBに記憶された情報の構成例を示す模式図である。

##### 【符号の説明】

- 11 マイク
- 12 A/D変換器
- 13 信号処理部
- 14 音素検出部
- 15 単語検出部
- 16 文章検出部
- 17 強度検出部
- 18 テンポ検出部

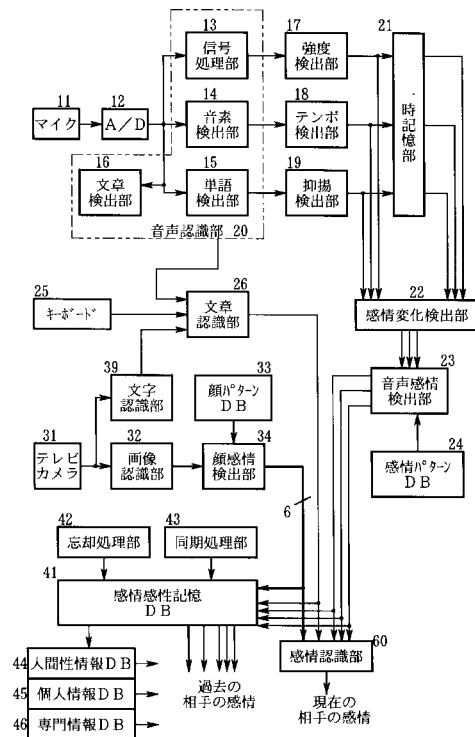
50

- 19 抑揚検出部
- 20 音声認識部
- 21 一時記憶部
- 22 感情変化検出部
- 23 音声感情検出部
- 24 感情パターンDB
- 25 キーボード
- 26 文章認識部
- 31 テレビカメラ
- 32 画像認識部
- 33 顔パターンDB
- 34 顔感情検出部
- 39 文字認識部
- 41 感情感性記憶DB
- 42 忘却処理部
- 43 同期処理部
- 44 人間性情報DB
- 45 個人情報DB
- 46 専門情報DB
- 51 バンドパスフィルタ
- 52 絶対値変換部
- 53 比較部
- 54 領域中心検出部
- 55 領域間隔検出部
- 60 感情認識部

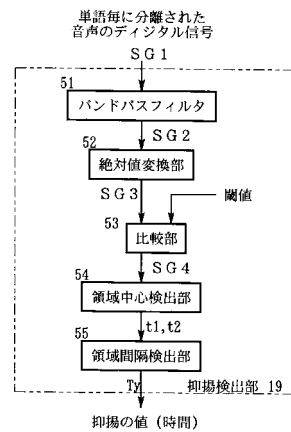
10

20

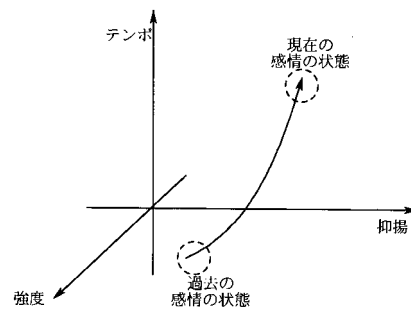
【図1】



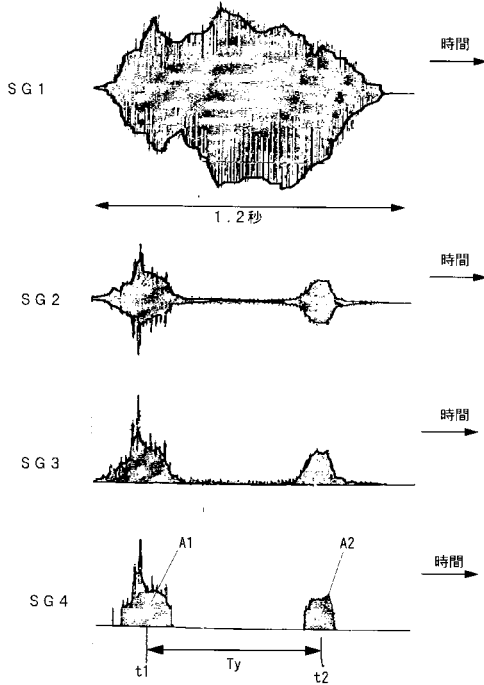
【図2】



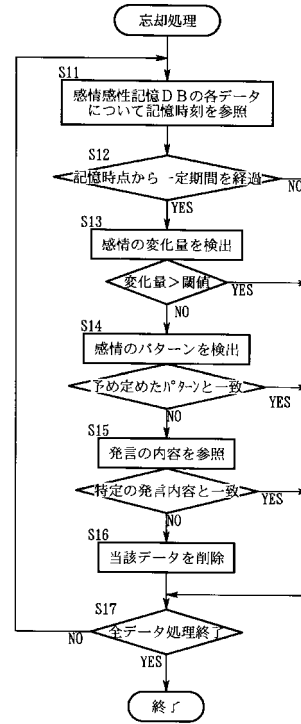
【図3】



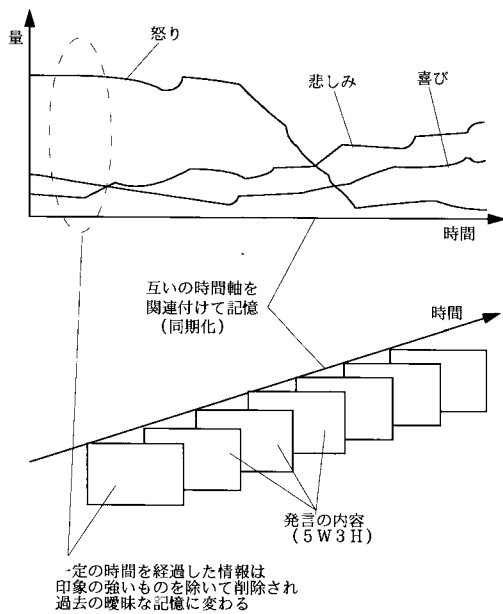
【 図 4 】



【 図 5 】



【 図 6 】



## フロントページの続き

- (56)参考文献 特開平05-012023(JP,A)  
特開2000-222378(JP,A)  
特開平08-339446(JP,A)  
特開2000-187435(JP,A)  
特開平10-143351(JP,A)  
特開平06-139044(JP,A)  
藤崎, 大野, 富田, 丸山, 話者の感情が音声の韻律的特徴に及ぼす影響, 日本音響学会平成7年度春季研究発表会講演論文集 I, 日本, 1995年 3月14日, 2-4-3, Pages 269-270  
小川, 中尾, 重永, 単語音声に含まれる感情情報について, 日本音響学会平成7年度春季研究発表会講演論文集 I, 日本, 1995年 3月14日, 2-4-2, Pages 267-268  
北原, 東倉, 音声の韻律情報と感情表現, 電子情報通信学会技術研究報告[音声], 日本, 1989年 3月24日, Vol.88, No.482, SP88-158, Pages 27-32  
羽尻, 山下, 小川, 天白, 日本語母音単音節における感情表現と韻律情報の関係, 日本音響学会平成7年度春季研究発表会講演論文集 I, 日本, 1995年 3月14日, 2-4-1, Pages 265-266  
森山, 斎藤, 小沢, 音声における感情表現語と感情表現パラメータの対応付け, 電子情報通信学会論文誌 D-II, 日本, 1999年 4月25日, Vol.J82-D-II, No.4, Pages 703-711  
北原義典, 感性情報とメディア処理技術, 電子情報通信学会誌, 日本, 1998年12月 5日, Vol.81, No.1, Pages 60-67

(58)調査した分野(Int.Cl.<sup>7</sup>, DB名)

G10L 15/00

G10L 15/10

JICSTファイル(JOIS)