

(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2017-111760

(P2017-111760A)

(43) 公開日 平成29年6月22日(2017.6.22)

(51) Int.Cl.	F I	テーマコード (参考)
G 0 6 F 17/30 (2006.01)	G O 6 F 17/30 2 1 O D	
G 1 0 L 15/10 (2006.01)	G O 6 F 17/30 1 7 O E	
G 1 0 L 25/63 (2013.01)	G O 6 F 17/30 2 1 O A	
	G 1 0 L 15/10 5 O O N	
	G 1 0 L 25/63	
審査請求 未請求 請求項の数 14 O L (全 22 頁)		

(21) 出願番号	特願2015-247885 (P2015-247885)	(71) 出願人	000001443
(22) 出願日	平成27年12月18日 (2015.12.18)		カシオ計算機株式会社
			東京都渋谷区本町1丁目6番2号
		(74) 代理人	100095407
			弁理士 木村 満
		(72) 発明者	中込 浩一
			東京都羽村市栄町3-2-1 カシオ計算機株式会社 羽村技術センター内
		(72) 発明者	佐藤 勝彦
			東京都羽村市栄町3-2-1 カシオ計算機株式会社 羽村技術センター内
		(72) 発明者	山谷 崇史
			東京都羽村市栄町3-2-1 カシオ計算機株式会社 羽村技術センター内

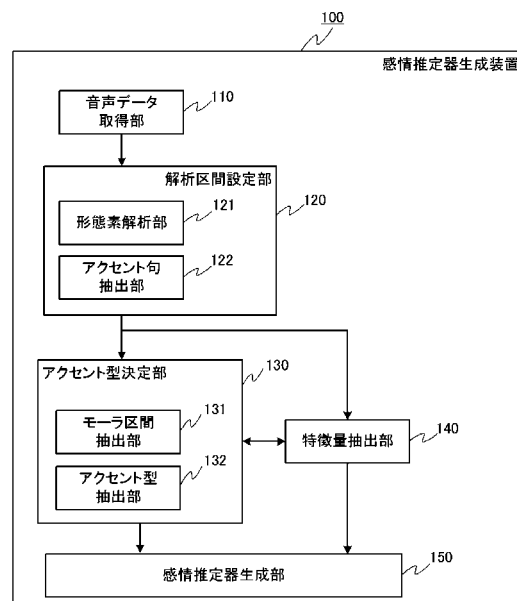
(54) 【発明の名称】 感情推定器生成方法、感情推定器生成装置、感情推定方法、感情推定装置及びプログラム

(57) 【要約】

【課題】 音声データから発話者の感情を推定する推定精度を向上する。

【解決手段】 感情推定器生成装置100は、教師データの元となる音声データの特徴量を解析する解析区間を設定する解析区間設定部120と、解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、解析区間に含まれる音声データの特徴量の変化パターンとして決定するアクセント型決定部130と、特徴量の変化パターンごとに分類された音声データを教師データとして、特徴量の変化パターンごとに、音声が発話したときの発話者の感情を推定する感情推定器を生成する感情推定器生成部150と、を備える。

【選択図】 図2



【特許請求の範囲】

【請求項 1】

教師データの元となる音声データの特徴量を解析する解析区間を設定する解析区間設定ステップと、

前記解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、前記解析区間に含まれる音声データの特徴量の変化パターンとして決定する変化パターン決定ステップと、

前記特徴量の変化パターンごとに分類された音声データを教師データとして、前記特徴量の変化パターンごとに、音声を発話したときの発話者の感情を推定する感情推定器を生成する感情推定器生成ステップと、

を含む感情推定器生成方法。

10

【請求項 2】

前記変化パターン決定ステップは、

前記解析区間に含まれる音声データを、音節の単位であるモーラ区間に分割するモーラ区間抽出ステップと、

前記解析区間における音声データの特徴量の平均値と、前記モーラ区間における音声データの特徴量の平均値と、をモーラ区間ごとに比較した比較結果に基づいて、前記解析区間の音声データを発話したときのモーラ区間ごとに变化する音声の特徴量の変化パターンを抽出する変化パターン抽出ステップと、

を含むことを特徴とする請求項 1 に記載の感情推定器生成方法。

20

【請求項 3】

前記変化パターン抽出ステップでは、音声の特徴量として音声データから抽出した音声の基本周波数を用い、前記解析区間における音声の平均基本周波数と、前記モーラ区間における音声の平均基本周波数と、をモーラ区間ごとに比較し、モーラ区間の音声の平均基本周波数が解析区間の音声の平均基本周波数よりも高い場合には High を、低い場合には Low を付与し、モーラ区間ごとに High と Low に変化する音声の特徴量の変化パターンを抽出する、

ことを特徴とする請求項 2 に記載の感情推定器生成方法。

【請求項 4】

前記変化パターン抽出ステップでは、音声の特徴量として音声データから抽出した音声の強度を用い、前記解析区間における音声の平均強度と、前記モーラ区間における音声の平均強度と、をモーラ区間ごとに比較し、モーラ区間の音声の平均強度が解析区間の音声の平均強度よりも高い場合には High を、低い場合には Low を付与し、モーラ区間ごとに High と Low に変化する音声の特徴量の変化パターンを抽出する、

ことを特徴とする請求項 2 に記載の感情推定器生成方法。

30

【請求項 5】

前記変化パターン抽出ステップでは、音声の特徴量として音声データから抽出した音素の継続時間長を用い、前記解析区間における音素の平均継続時間長と、前記モーラ区間における音素の平均継続時間長と、をモーラ区間ごとに比較し、モーラ区間の音素の平均継続時間長が解析区間の音素の平均継続時間長よりも長い場合には High を、短い場合には Low を付与し、モーラ区間ごとに High と Low に変化する音声の特徴量の変化パターンを抽出する、

ことを特徴とする請求項 2 に記載の感情推定器生成方法。

40

【請求項 6】

前記変化パターン抽出ステップでは、音声の特徴量として、音声の基本周波数、音声の強度、音素の継続時間長の少なくとも何れか 1 つを使用して音声の特徴量の変化パターンを抽出する、

ことを特徴とする請求項 2 から 5 の何れか一項に記載の感情推定器生成方法。

【請求項 7】

前記解析区間設定ステップでは、音声データを、言語の意味を持つ最小の単位である形

50

態素に分割し、当該形態素の後で発話された助詞又は助動詞と結合したアクセント句の区間を前記解析区間として設定する、

ことを特徴とする請求項 1 から 6 の何れか一項に記載の感情推定器生成方法。

【請求項 8】

前記モーラ区間抽出ステップでは、音声データをテキスト表示した場合に、仮名文字 1 文字を 1 モーラ区間とし、小書きの仮名文字はその前の仮名文字と一緒にして 1 モーラ区間とし、長音は独立して 1 モーラ区間とする、

ことを特徴とする請求項 2 に記載の感情推定器生成方法。

【請求項 9】

前記感情推定器は、発話者の発話時の感情を、悲しみ、退屈、怒り、驚き、落胆、嫌悪、喜び、の何れかの感情であると推定する、

ことを特徴とする請求項 1 から 8 の何れか一項に記載の感情推定器生成方法。

【請求項 10】

前記複数のクラスに分類された変化パターンを設定する変化パターン設定ステップを含む、

ことを特徴とする請求項 1 から 9 の何れか一項に記載の感情推定器生成方法。

【請求項 11】

教師データの元となる音声データの特徴量を解析する解析区間を設定する解析区間設定手段と、

前記解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、前記解析区間に含まれる音声データの特徴量の変化パターンとして決定する変化パターン決定手段と、

前記特徴量の変化パターンごとに分類された音声データを教師データとして、前記特徴量の変化パターンごとに、音声を発話したときの発話者の感情を推定する感情推定器を生成する感情推定器生成手段と、

を含む感情推定器生成装置。

【請求項 12】

コンピュータを

教師データの元となる音声データの特徴量を解析する解析区間を設定する解析区間設定手段、

前記解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、前記解析区間に含まれる音声データの特徴量の変化パターンとして決定する変化パターン決定手段、

前記特徴量の変化パターンごとに分類された音声データを教師データとして、前記特徴量の変化パターンごとに、音声を発話したときの発話者の感情を推定する感情推定器を生成する感情推定器生成手段、

として機能させるためのプログラム。

【請求項 13】

解析対象とする音声データの特徴量を解析する解析区間を設定する解析区間設定ステップと、

前記解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、前記解析区間に含まれる音声データの特徴量の変化パターンとして決定する変化パターン決定ステップと、

前記特徴量の変化パターンごとに、同じ特徴量の変化パターンを有する教師データに基づいて生成された感情推定器を用いて、前記解析区間の音声を発話した時の発話者の感情を推定する感情推定ステップと、

を含む感情推定方法。

【請求項 14】

解析対象とする音声データの特徴量を解析する解析区間を設定する解析区間設定手段と、

10

20

30

40

50

前記解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、前記解析区間に含まれる音声データの特徴量の変化パターンとして決定する変化パターン決定手段と、

前記特徴量の変化パターンごとに、同じ特徴量の変化パターンを有する教師データに基づいて生成された感情推定器を用いて、前記解析区間の音声を発話した時の発話者の感情を推定する感情推定手段と、

を備えた感情推定装置。

【発明の詳細な説明】

【技術分野】

【0001】

10

本発明は、感情推定器生成方法、感情推定器生成装置、感情推定方法、感情推定装置及びプログラムに関する。

【背景技術】

【0002】

感情をラベリングした音声データ群を教師データとして機械学習により生成された感情推定装置を用いて、発話者の感情を推定する技術の開発が進められている。例えば、特許文献1は、音声の強度、音声のテンポ、音声の抑揚のそれぞれの変化量を求め、求めた変化量に基づいて発話者の感情を推定する技術を開示している。

【先行技術文献】

【特許文献】

20

【0003】

【特許文献1】特開2002-91482号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

一般に、興奮した状態で発話すると、通常の前話時よりも話し方が早くなり、声が高くなる傾向がある。また、落胆した状態で発話すると、通常の前話時よりも話し方が遅くなり、声が低くなる傾向がある。このように、前話時の前話者の感情と音声の特徴量とは相関性がある。特許文献1は、このような音声データの特徴量の変化を解析することにより、前話者の感情を推定する技術を開示している。

30

【0005】

ところで、通常の前話状態で前話された音声の特徴量と怒った感情状態で前話された音声の特徴量とを比較した場合、短い言葉と長い言葉とでは特徴量の変化の傾向が異なる場合がある。例えば、前話しやすい短い言葉は、前話時の感情状態によって音声の特徴量の変化が大きい場合が多い。これに対して、早口言葉のように前話しにくい長い言葉は、前話時の感情状態によって音声の特徴量の変化が小さい場合がある。特許文献1が開示する技術は、このように前話時の感情状態によって音声の特徴量に変化が少ない言葉と変化が大きい言葉とを一律にして前話者の感情推定を行うので、推定精度が上がりにくいという問題があった。

【0006】

40

本発明は、このような状況を鑑みてなされたものであり、音声データから前話者の感情を推定する推定精度を向上することができる感情推定器生成方法、感情推定器生成装置、感情推定方法、感情推定装置及びプログラム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】

【0007】

上記目的を達成するため、本発明の第1の観点に係る感情推定器生成方法は、

教師データの元となる音声データの特徴量を解析する解析区間を設定する解析区間設定ステップと、

前記解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、前記解析区間に含まれる音声データの特徴量の変化パ

50

ターンとして決定する変化パターン決定ステップと、

前記特徴量の変化パターンごとに分類された音声データを教師データとして、前記特徴量の変化パターンごとに、音声を発話したときの発話者の感情を推定する感情推定器を生成する感情推定器生成ステップと、

を含むことを特徴とする。

【0008】

また、本発明の第2の観点に係る感情推定器生成装置は、

教師データの元となる音声データの特徴量を解析する解析区間を設定する解析区間設定手段と、

前記解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、前記解析区間に含まれる音声データの特徴量の変化パターンとして決定する変化パターン決定手段と、

前記特徴量の変化パターンごとに分類された音声データを教師データとして、前記特徴量の変化パターンごとに、音声を発話したときの発話者の感情を推定する感情推定器を生成する感情推定器生成手段と、

を含むことを特徴とする。

【0009】

また、本発明の第3の観点に係るプログラムは、

コンピュータを

教師データの元となる音声データの特徴量を解析する解析区間を設定する解析区間設定手段、

前記解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、前記解析区間に含まれる音声データの特徴量の変化パターンとして決定する変化パターン決定手段、

前記特徴量の変化パターンごとに分類された音声データを教師データとして、前記特徴量の変化パターンごとに、音声を発話したときの発話者の感情を推定する感情推定器を生成する感情推定器生成手段、

として機能させることを特徴とする。

【0010】

また、本発明の第4の観点に係る感情推定方法は、

解析対象とする音声データの特徴量を解析する解析区間を設定する解析区間設定ステップと、

前記解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、前記解析区間に含まれる音声データの特徴量の変化パターンとして決定する変化パターン決定ステップと、

前記特徴量の変化パターンごとに、同じ特徴量の変化パターンを有する教師データに基づいて生成された感情推定器を用いて、前記解析区間の音声を発話した時の発話者の感情を推定する感情推定ステップと、

を含むことを特徴とする。

【0011】

また、本発明の第5の観点に係る感情推定装置は、

解析対象とする音声データの特徴量を解析する解析区間を設定する解析区間設定手段と、

前記解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、前記解析区間に含まれる音声データの特徴量の変化パターンとして決定する変化パターン決定手段と、

前記特徴量の変化パターンごとに、同じ特徴量の変化パターンを有する教師データに基づいて生成された感情推定器を用いて、前記解析区間の音声を発話した時の発話者の感情を推定する感情推定手段と、

を備えることを特徴とする。

10

20

30

40

50

【発明の効果】

【0012】

本発明によれば、音声データから発話者の感情を推定する推定精度を向上することができる。

【図面の簡単な説明】

【0013】

【図1】本発明の実施形態1に係る感情推定器生成装置の物理構成を示すブロック図である。

【図2】本発明の実施形態1に係る感情推定器生成装置の機能構成を示すブロック図である。

10

【図3】形態素について説明するための図である。

【図4】モーラ区間について説明するための図である。

【図5】特徴量の解析方法について説明するための図である。

【図6】特徴量の解析方法について説明するための図である。

【図7】クラス分けについて説明するための図である。

【図8】生成された感情推定装置の識別閾値のイメージについて説明するための図である。

【図9】感情推定装置の機能構成を示すブロック図である。

【図10】感情推定器の生成処理について説明するためのフローチャートである。

【図11】感情推定処理について説明するためのフローチャートである。

20

【図12】変形例1に係る特徴量の解析区間について説明するための図である。

【図13】変形例1に係る特徴量の解析方法について説明するための図である。

【図14】変形例2に係る音声の強度による特徴量の解析について説明するための図である。

【図15】変形例5に係る複数の感情の度合いを推定する技術について説明するための図である。

【発明を実施するための形態】

【0014】

以下、本発明の実施形態に係る感情推定器生成方法、感情推定器生成装置、感情推定方法、感情推定装置及びプログラムについて、図面を参照しながら説明する。なお、図中同一又は相当する部分には同一符号を付す。

30

【0015】

(実施形態1)

本実施形態では、音声データから発話者の感情を推定する感情推定器を生成する感情推定器生成装置について説明した後、音声を発話した時の発話者の感情を推定する感情推定装置について説明する。本実施形態では、感情推定装置が、発話者の感情を悲しんでいる状態（悲しみ）、退屈している状態（退屈）、怒っている状態（怒り）、驚いている状態（驚き）、落胆している状態（落胆）、嫌悪感を抱いている状態（嫌悪）、喜んでいる状態（喜び）、の基本的な7種類の感情状態のいずれかであると推定する場合について説明する。

40

なお、以下の実施形態では、音声データの特徴量の変化パターンをアクセント型と称する。

【0016】

実施形態1に係る感情推定器生成装置100は、物理的には、図1に示すように、制御部1と、記憶部2と、入出力部3と、バス4と、を備える。

【0017】

制御部1は、ROM（Read Only Memory）と、RAM（Random Access Memory）と、CPU（Central Processing Unit）と、を備える。ROMは、本実施形態に係る感情推定器生成プログラム、及び、各種初期設定、ハードウェアの検査、プログラムのロード等を行うための初期プログラム等を記憶する。RAMは、CPUが実行する各種ソフトウェア

50

プログラム、これらのソフトウェアプログラムの実行に必要なデータ等を一時的に記憶するワークエリアとして機能する。CPUは、各種ソフトウェアプログラムを実行することにより、様々な処理及び演算を実行する中央演算処理部である。

【0018】

記憶部2は、ハードディスクドライブ、フラッシュメモリ等の不揮発性メモリを備える。記憶部2は、教師データとする音声データ等を記憶する。

【0019】

入出力部3は、教師データとする音声データを取得するための音声入力装置、CD（Compact Disc）ドライブ、USB（Universal Serial Bus）インタフェースを備える。入出力部3は、教師データとする音声データを取得する。また、入出力部3は、生成した感情推定器をプログラムもしくは感情推定器の特性を決めるパラメータを外部装置に出力する。

10

【0020】

バス4は、制御部1と、記憶部2と、入出力部3と、を接続する。

【0021】

感情推定器生成装置100は、機能的には、図2に示すように、音声データ取得部110と、解析区間設定部120と、アクセント型決定部130と、特徴量抽出部140と、感情推定器生成部150と、を含む。また、解析区間設定部120は、形態素解析部121とアクセント句抽出部122と、を含む。また、アクセント型決定部130は、モーラ区間抽出部131と、アクセント型抽出部132と、を含む。

20

【0022】

音声データ取得部110は、入出力部3を介して感情推定器を生成するために教師データとして使用する音声データを取得する。教師データは、例えば、悲しみ、退屈、怒り、驚き、落胆、嫌悪、喜び、の7種類の感情状態で発話された音声から構成される。また、教師データは、十分に多くの種類の語句を含む音声データで構成される。教師データを発話する人数及び教師データに含まれるアクセント句の種類は多い方が好ましい。アクセント句とは、名詞と助詞、もしくは動詞と助動詞を結合した音声データを区分する単位である。例えば、教師データとして、500人程度の多人数が7種類の感情状態で発話した、1000種類以上のアクセント句を含む音声データを準備する。

【0023】

解析区間設定部120は、教師データとする音声データの特徴を解析する単位である解析区間を設定する。そのために、解析区間設定部120は、形態素解析部121とアクセント句抽出部122とを備える。

30

【0024】

形態素解析部121は、取得した音声データを形態素に分割する。形態素とは、言語としての意味を有する最小単位である。例えば、「坊主が屏風に上手に坊主の絵を描いた」という音声は、図3に示すように、「坊主」、「が」、「屏風」、「に」、「上手」、「に」、「坊主」、「の」、「絵」、「を」、「描い」、「た」の12個の形態素に分割される。

【0025】

アクセント句抽出部122は、取得した音声データからアクセント句を抽出する。アクセント句とは、形態素に分割した名詞又は動詞に、それに続く助詞又は助動詞を結合した区間である。上述の例では、アクセント句は、「坊主が」、「屏風に」、「上手に」、「坊主の」、「絵を」、「描いた」となる。本実施形態では、このアクセント句の単位で音声データの特徴を解析する場合について説明する。アクセント句の単位で音声データを解析する理由は、アクセント句の単位で発話者の感情状態が変化する場合が多いからである。

40

【0026】

アクセント型決定部130は、アクセント句のアクセント型を決定する。アクセント型とは、アクセント句を構成する音節が発話されている区間であるモーラ区間ごとに、音声

50

の特徴量が平均特徴量に対して大きい場合には「H」、小さい場合には「L」を付与して得られる「H」と「L」の組み合わせのパターンである。アクセント型を決定するために、アクセント型決定部 130 は、モーラ区間抽出部 131 とアクセント型抽出部 132 とを備える。

【0027】

モーラ区間抽出部 131 は、図 4 に示すように、解析対象のアクセント句区間の音声データから、モーラ区間を抽出する。モーラ区間は 1 つの音節が発話されている区間である。アクセント句「坊主が」の場合で説明すると、「ボ」、「ウ」、「ズ」、「ガ」のそれぞれの音節がモーラ区間である。

【0028】

アクセント型抽出部 132 は、モーラ区間のそれぞれが「H」もしくは「L」のいずれに該当するかを判別し、アクセント句のアクセント型を抽出する。アクセント型の抽出方法には、音声の強度、音声のピッチ、音素の発話時間長等の特徴量を使用する方法がある。ここでは、音声のピッチに着目した抽出方法について、図 5 と図 6 を参照しながら説明する。

【0029】

アクセント型抽出部 132 は、図 5 に示すように、モーラ区間をさらに細分する所定時間の時間窓を設定する。1 つのモーラ区間に対して、窓 1 を設定し、その窓内の音声データを FFT (Fast Fourier Transform) 変換する。次に、窓 1 を所定時間 dt ずらした窓 2 内の音声データを FFT 変換する。以下、同様に窓 n 内の音声データを FFT 変換する。時間窓の設定方法は、例えば、モーラ区間内に 10 以上の時間窓を構成するように時間窓とずらす時間幅 dt を設定する。時間窓の数が少なすぎると、計算精度が低下するからである。

【0030】

図 6 は、上記の FFT 変換により得られた各窓内の音声データのスペクトル分布を示した例である。横軸は周波数であり、縦軸はスペクトルの強度である。このスペクトルの中で最も低い周波数領域に存在するピーク周波数を f_0 とする。この f_0 は、その窓区間の音声データから得られた発話者固有の基本周波数を示す。窓 1 から得られた f_0 を f_{0_1} 、窓 2 から得られた f_0 を f_{0_2} 、とする。同様にして、窓 n から得られた f_0 を f_{0_n} とする。そして、アクセント型抽出部 132 は、 f_{0_1} から f_{0_n} までの平均値を計算し、第 1 モーラ区間の平均基本周波数 1_f_0 とする。

【0031】

アクセント型抽出部 132 は、アクセント句に含まれる全てのモーラ区間について同様の計算をする。第 m モーラ区間の平均基本周波数 m_f_0 は、式 1 を用いて算出することができる。

【0032】

$$m_f_0 = 1/n \cdot \sum f_{0_n} \quad (\text{式 1})$$

【0033】

次に、アクセント型抽出部 132 は、アクセント句区間における平均基本周波数 m_t_h を式 2 を用いて求める。

【0034】

$$m_t_h = \{\max(1_f_0, \dots, n_f_0) - \min(1_f_0, \dots, n_f_0)\} / 2 \quad (\text{式 2})$$

【0035】

次に、アクセント型抽出部 132 は、モーラ区間の平均基本周波数 m_f_0 とアクセント句区間の平均基本周波数 m_t_h とを比較し、 $m_f_0 \geq m_t_h$ であれば「H」、 $m_f_0 < m_t_h$ であれば「L」をそれぞれのモーラ区間に付与する。アクセント型抽出部 132 は、このようにアクセント句を構成するモーラ区間ごとに「H」と「L」を付与することにより、H と L の組み合わせで構成されるアクセント型を抽出する。

【0036】

アクセント型決定部 130 は、教師データから生成されたアクセント句の全てについて

、この処理を行う。図7に示すアクセント型の例は、大量の教師データから得られたアクセント型の中で発生頻度が高い順に20種類のアクセント型を選択した、モーラ区間数が6以下の例である。解析対象のアクセント句に含まれるモーラ区間の数が6以下である場合、この20種類のアクセント型に対して順にクラス1からクラス20までのクラス名を付与する。アクセント型とクラスとは1対1に対応している。アクセント型決定部130は、教師データとするアクセント句単位の音声データとアクセント型（クラス）とを対応付けて記憶部2に記憶する。なお、この20種類のアクセント型に該当しなかったアクセント句は、教師データから除外する。以後、本実施形態では、解析対象とするアクセント句に含まれるモーラ区間数が6以下である場合について説明する。

【0037】

ここで、アクセント句に含まれるモーラ区間数が6以下の場合における20種類のアクセント型の選択方法は、7種類の感情で発話された大量の日本語を統計処理した実験結果に基づいて、発生頻度が高い順に20種類のアクセント型を選択する。アクセント型のクラス数を減らすと、アクセント型に該当しない教師データの頻度が高くなり、生成された感情推定器を内蔵した感情推定装置の推定精度が低下することになる。一方、アクセント型のクラス数を増やすと、生成する感情推定器の種類が増えるので感情推定装置の製造コストが高くなることになる。したがって、この2つの兼ね合いでアクセント型のクラス数を決定する。なお、アクセント型を20種類としたのはアクセント句に含まれるモーラ区間数が6以下の場合である日本語の場合の例である。アクセント句に含まれるモーラ区間数が7以上である場合、もしくは他の言語の場合は、アクセント型の発生頻度についてさらに統計処理して決める必要がある。

【0038】

図2に戻って、特徴量抽出部140は、アクセント句ごとの音声の特徴量を抽出する。音声の特徴量とは、音声の大きさ、音声のピッチ、音素の発話時間長等である。そして、抽出した特徴量にアクセント型決定部130で決定したクラス1からクラス20のクラス名を付与し、教師データとして記憶部2に記憶する。

【0039】

感情推定器生成部150は、教師データをクラスごとに記憶部2から取得し、それぞれのクラスに適応した感情推定器を生成する。具体的には、感情推定器生成部150は、クラス1に分類された教師データを取得し、その教師データの発話時の感情状態を、7種類の感情状態である悲しんでいる状態（悲しみ）、退屈している状態（退屈）、怒っている状態（怒り）、驚いている状態（驚き）、落胆している状態（落胆）、嫌悪感を抱いている状態（嫌悪）、喜んでいる状態（喜び）、に分類するクラス1用の感情推定器を生成する。図8は、7種類の感情に識別する識別閾値を2次元で表現したイメージ図である。教師データに基づいて感情推定器を生成する方法には公知の技術を用いることができる。次に、感情推定器生成部150は、クラス2に分類された教師データを取得し、その教師データの発話時の感情状態を、7種類の感情状態である悲しみ、退屈、怒り、驚き、落胆、嫌悪、喜び、に分類するクラス2用の感情推定器を生成する。同様にして、感情推定器生成部150は、クラス20までの感情推定器を生成する。

【0040】

次に、上記の説明により生成した20種類の感情推定器を内蔵する感情推定装置200の構成について、図9を参照しながら説明する。感情推定装置200は、物理的には、図1に示す構成と同じである。

【0041】

制御部1が備えるROMは、本実施形態に係る感情推定器生成装置100で生成された感情推定プログラムを記憶する。記憶部2は、解析対象とする音声データ等を記憶する。入出力部3は、解析対象とする音声データを取得するための音声入力装置、CDドライブ、USBインタフェースを備える。また、入出力部3は、感情推定器生成装置100で生成された感情推定器の特性を決定するパラメータを取得するようにしてもよい。また、入出力部3は、感情を推定した結果を出力するための表示装置もしくは音声出力装置を備え

10

20

30

40

50

る。

【0042】

感情推定装置200は、図9に示すように、音声データ取得部210と、話者分割部220と、解析区間設定部230と、アクセント型決定部240と、選択部250と、特徴量抽出部260と、感情推定部270と、統合部280と、の機能を含む。また、解析区間設定部230は、形態素解析部231とアクセント句抽出部232との機能を含む。また、アクセント型決定部240は、モーラ区間抽出部241とアクセント型抽出部242との機能を含む。

【0043】

音声データ取得部210は、ユーザが発話した解析対象とする音声を取得する。音声データ取得部210は、マイク等の音声取得装置から構成される。また、音声データ取得部210は、CDドライブ、USBインタフェースを備え、音声データとしてユーザの音声を取得することもできる。

10

【0044】

話者分割部220は、取得した解析対象の音声データを話者ごとに分割する。音声データの中に複数人の音声データが存在する場合、1人の話者が発話した文ごとに発話者の感情を推定するためである。音声データを話者ごとに分割する方法は、公知の技術を用いて行う。例えば、音声の強度、音声のピッチ、音素の発話時間長等の相関性に基づいて分割することができる。

20

【0045】

解析区間設定部230、アクセント型決定部240は、感情推定器の生成時と同じ条件下で解析対象の音声データを解析するために、感情推定器生成装置100と同じ構成を有している。つまり、解析区間設定部230は、音声データからアクセント句を抽出し、アクセント型決定部240は、アクセント句ごとにアクセント型（クラス）を決定する。

【0046】

選択部250は、クラス分けされたアクセント句ごとに、該当するクラスに対応する感情推定器を選択する。具体的には、感情推定装置200に内蔵している20種類の感情推定器の中から、解析対象のアクセント句のクラスに対応する感情推定器を選択する。

【0047】

特徴量抽出部260は、感情推定器生成装置100と同じ構成を有しており、同じ条件下で音声データから特徴量を抽出する。そして、特徴量抽出部260は、抽出した特徴量とアクセント型を示すクラス名とを対応付けて記憶部2に記憶する。

30

【0048】

感情推定部270は、選択部250が選択した感情推定器を用いて、アクセント句ごとに発話者の感情を推定する。具体的には、感情推定部270は、クラス1に分類されたアクセント句の感情を推定する場合には、クラス1用の感情推定器を選択して発話者の感情を推定する。感情推定部270は、クラスnに分類されたアクセント句の感情を推定する場合には、クラスn用の感情推定器を選択して発話者の感情を推定する。そして、感情推定部270は、解析対象のアクセント句を発話したときの発話者の感情状態が、悲しみ、退屈、怒り、驚き、落胆、嫌悪、喜び、のいずれの感情状態に該当するかを推定する。

40

【0049】

統合部280は、発話者の感情を音声データの文単位で推定する。具体的には、統合部280は、1文の中で最も多かった感情をその文を発話した発話者の感情として推定する。例えば、「坊主が」、「屏風に」、「上手に」、「坊主の」、「絵を」、「描い」、「た」の7つのアクセント句から構成される「坊主が屏風に上手に坊主の絵を描いた」という文において、「喜び」と判別されたアクセント句の数が4であり、「怒り」と判別されたアクセント句の数が2であり、「驚き」と判別されたアクセント句の数が1であった場合、一番多い「喜び」をこの「坊主が屏風に上手に坊主の絵を描いた」を発話したときの発話者の感情として推定する。

【0050】

50

次に、以上の構成を有する感情推定器生成装置 100 が感情推定器を生成する処理について、図 10 を参照しながら説明する。教師データとして使用する悲しみ、退屈、怒り、驚き、落胆、嫌悪、喜び、の 7 種類の感情状態で発話された音声データは、予め記憶部 2 に記憶されているものとする。解析対象の音声データに含まれるアクセント句のモーラ区間数は 6 以下であると仮定する。感情推定器を生成する担当者が、感情推定器生成装置 100 に予めインストールされている感情推定器生成プログラムを起動することにより、図 10 に示すフローチャートは開始される。

【0051】

制御部 1 は、感情推定器生成プログラムが起動されると、記憶部 2 に記憶されている教師データを音声データ取得部 110 に取得する（ステップ S11）。そして、形態素解析部 121 は、取得した音声データを形態素の単位で分割する（ステップ S12）。次に、アクセント句抽出部 122 は、音声データの特徴を解析する単位であるアクセント句を抽出し、音声データをアクセント句に分割する（ステップ S13）。

【0052】

次に、モーラ区間抽出部 131 は、アクセント句に含まれるモーラ区間を抽出する（ステップ S14）。そして、アクセント型抽出部 132 は、アクセント句のアクセント型を抽出する（ステップ S15）。具体的には、アクセント型抽出部 132 は、図 7 を用いて説明したように、教師データとして使用するアクセント句を 20 のクラスに分類する（ステップ S16）。アクセント型抽出部 132 は、その分類をするために、図 5 と図 6 を用いて説明したように、モーラ区間ごとの平均基本周波数 m_f0 とアクセント句区間の平均基本周波数 m_th とを比較し、 $m_f0 \geq m_th$ であれば「H」、 $m_f0 < m_th$ であれば「L」をそれぞれのモーラ区間に付与する。アクセント型抽出部 132 は、このようにして教師データとして使用するアクセント句に対して、アクセント句を構成するモーラ区間ごとに H と L を付与し、H と L のパターンによりアクセント型を抽出する。そして、アクセント型決定部 130 は、教師データのアクセント型を図 7 に示す 20 のアクセント型（クラス）の何れかに決定する。

【0053】

次に、特徴量抽出部 140 は、教師データとする音声データの特徴量をアクセント句ごとに抽出し、抽出した特徴量のデータと分類されたクラスとを対応付けて教師データとして記憶部 2 に記憶する（ステップ S17）。

【0054】

感情推定器生成部 150 は、アクセント型（クラス）ごとに分類された教師データに基づいて、それぞれのクラスごとに感情推定器を生成する（ステップ S18）。具体的には、感情推定器生成部 150 は、クラス 1 に分類された教師データ（アクセント句）を取得して、その教師データの発話時の感情状態を、7 種類の感情状態である悲しみ、退屈、怒り、驚き、落胆、嫌悪、喜び、に分類することが可能な感情推定器を生成する。より具体的には、図 8 に示すような 7 種類の感情に分類するための識別閾値（分類器を構成する数式のパラメータ）を生成する。次に、感情推定器生成部 150 は、クラス 2 に分類された教師データ（アクセント句）を取得して、その教師データの発話時の感情状態を、7 種類の感情状態である悲しみ、退屈、怒り、驚き、落胆、嫌悪、喜び、に分類することが可能な 2 つめの感情推定器を生成する。感情推定器生成部 150 は、このように 20 種類の感情推定器を生成する。以上で、感情推定器生成装置 100 の感情推定器生成処理の説明を終了する。

【0055】

次に、感情推定器生成装置 100 が生成した 20 種類の感情推定器を内蔵する感情推定装置 200 が発話者の感情を推定する感情推定処理について、図 11 を参照しながら説明する。ユーザが、感情推定装置 200 に予めインストールされている感情推定プログラムを起動し、解析対象とする音声データを感情推定装置 200 に入力することにより、図 11 に示すフローチャートは開始される。

【0056】

制御部 1 は、感情推定プログラムが起動され、ユーザが解析対象の音声データを感情推定装置 200 に入力すると、話者分割部 220 は、取得した音声データを話者ごとに分割して記憶部 2 に記憶する。次に、解析区間設定部 230 は、話者ごとに音声データを記憶部 2 から取得する（ステップ S 31）。次に、形態素解析部 231 は、任意の話者の音声データを形態素に分解し（ステップ S 32）、アクセント句抽出部 232 は、解析単位であるアクセント句を決定する（ステップ S 33）。

【0057】

次に、アクセント型決定部 240 は、感情推定器生成装置 100 の動作説明と同様に、アクセント句ごとにアクセント型（クラス）を決定する。具体的には、モーラ区間抽出部 241 が、該当するアクセント句に含まれるモーラ区間を抽出し（ステップ S 34）、アクセント型抽出部 242 が、そのアクセント句のアクセント型を抽出する（ステップ S 35）。そして、アクセント型決定部 240 は、抽出したアクセント型からそのアクセント句が属するクラスを決定する。そして、選択部 250 は、該当するアクセント句を発話したときの発話者の感情を推定するために使用する感情推定器として、同じアクセント型（クラス）の教師データに基づいて感情推定器生成装置 100 が生成した感情推定器を選択する（ステップ S 36）。

【0058】

一方、特徴量抽出部 260 は、解析対象のアクセント句の音声の強度、音声のピッチ、音素の継続時間長といった音声の特徴量を抽出し、抽出した特徴量と判別したクラスとを対応付けて記憶部 2 に記憶する（ステップ S 37）。

【0059】

次に、感情推定部 270 は、選択部 250 が選択した感情推定器を用いて、該当するアクセント句を発話したときの発話者の感情を推定する（ステップ S 38）。

【0060】

アクセント句の 1 つについて感情推定が完了すると、感情推定装置 200 は、まだ解析が完了していないアクセント句が存在するか否かを判別する（ステップ S 39）。解析が完了していないアクセント句が存在する場合（ステップ S 39：No）、解析が完了していない他のアクセント句を抽出し（ステップ S 40）、そのアクセント句に該当する感情を推定する。

【0061】

すべてのアクセント句の解析が完了している場合（ステップ S 39：Yes）、感情推定装置 200 は、解析した文単位で統合処理を行う（ステップ S 41）。具体的には、統合部 280 は、解析対象の文に含まれるアクセント句ごとの感情推定結果に基づいて、最も多かった感情をその文を発話したときの発話者の感情として推定する。

【0062】

次に、感情推定装置 200 は、最初に取得した任意の人が発話したすべての文について解析が完了したか否かを判別する（ステップ S 42）。すべての文について解析が完了していない場合は（ステップ S 42：No）、他の文を抽出し（ステップ S 43）、他の文について感情推定処理を継続する。

【0063】

一方、感情推定装置 200 は、すべての文について解析が完了している場合は（ステップ S 42：Yes）、音声データに含まれているすべての人について感情推定が完了しているか否かを判別する（ステップ S 44）。すべての人について解析が完了していない場合は（ステップ S 44：No）、他の人の音声データを抽出して感情推定処理を継続する（ステップ S 45）。すべての人について解析処理が完了している場合は（ステップ S 44：Yes）、感情推定処理を終了する。

【0064】

以上に説明したように感情推定器生成装置 100 は、アクセント型ごとに分類した教師データに基づいて、アクセント型ごとに感情推定器を生成する。そして、感情推定装置 200 は、アクセント型ごとに生成された感情推定器を使用して、発話者の感情を推定する

10

20

30

40

50

。具体的には、感情推定装置 200 は、解析対象の音声データをアクセント型ごとに分類し、同じアクセント型を有する教師データに基づいて生成された感情推定器を用いて発話者の感情を推定する。これにより、音声データから発話者の感情を推定する推定精度を向上することができる。

【0065】

また、アクセント型抽出部 132 は、モーラ区間の単位で音声の特微量の変化を抽出するので、感情推定器生成装置 100 は、発話者の感情をより細かく解析することが可能な感情推定器を生成することができる。

【0066】

また、アクセント型抽出部 132 は、音声の基本周波数の変化に基づいてアクセント型を抽出する。発話時の感情状態により音声の基本周波数は変化する傾向がある。したがって、感情推定器生成装置 100 は、発話者の感情をより正確に推定することが可能な感情推定器を生成することができる。また、同じ理由により、感情推定装置 200 は、発話者の感情をより正確に推定することができる。

【0067】

解析区間設定手段 120 は、形態素の単位で音声を解析するので、感情推定器生成装置 100 は、発話者の感情をより正確に解析することが可能な感情推定器を生成することができる。

【0068】

感情推定器生成装置 100 は、発話者の発話時の感情を、悲しみ、退屈、怒り、驚き、落胆、嫌悪、喜び、に分類する感情推定器を生成する。この推定器を内蔵する感情推定装置 200 は、発話者の発話時の感情状態を、悲しみ、退屈、怒り、驚き、落胆、嫌悪、喜び、に分類することができる。

【0069】

（変形例 1）

実施形態 1 では、アクセント型を判別するために特微量をモーラ区間の単位で解析する説明をした。変形例 1 では、モーラ区間の中の母音区間に限定して特微量を解析する説明を行う。具体的には、図 12 に示すように、母音区間のみの音声データを取り出して、図 13 に示すように特微量の解析を行う。基本周波数の解析方法は実施形態 1 の説明と同じである。

【0070】

母音区間にのみ着目する理由は、子音区間よりも母音区間の方が音素の継続時間長が長く、含まれる音声のエネルギーも大きいので、感情の変化による特微量の変化は、子音区間よりも母音区間の方に顕著に現れるからである。

【0071】

このように、変形例 1 に係る感情推定装置 200 は、母音区間に限定して特微量の解析を行うことにより、感情推定の推定精度を向上することができる。

【0072】

（変形例 2）

実施形態 1 の説明では、アクセント型抽出部 132 が、音声の特微量として音声のピッチ情報（音声の基本周波数）を利用する場合について説明した。変形例 2 では、音声の特微量として音声の強度情報を利用する場合について説明する。ここでは、発話時の感情状態によって母音の発話区間における音声のエネルギー分布が変化することに着目した技術について説明する。

【0073】

具体的には、アクセント型抽出部 132 は、図 14 に点線の丸印で示した音声のエネルギーのピークが、母音区間の前半に存在するか後半に存在するかを判別する。例えば、前半にピークが存在した場合には「H」を付与し、後半にピークが存在した場合には「L」を付与する。これにより、アクセント型決定部 130 は、アクセント型を決定する。音声の強度によりアクセント型进行分类する場合は、実験データに基づいてクラス分けの仕方を

10

20

30

40

50

検討する必要がある。その他の説明は実施形態 1 の説明と同じである。

【0074】

なお、変形例 2 の説明では、音声エネルギーのピーク点の時間位置の変化に着目する解析方法を説明したが、音声の強度の変化を用いてアクセント型を抽出することもできる。怒った状態で発話すると音声の強度は高くなる傾向があり、悲しい状態で発話すると音声の強度は低くなる傾向があるので、この傾向を利用するものである。この場合、例えば、アクセント句区間に含まれるモーラ区間ごとの音声のピーク強度を計測して、アクセント句区間の平均ピーク強度を求める。そして、モーラ区間の音声のピーク強度と平均ピーク強度とを比較して、モーラ区間ごとに「H」又は「L」を付与することにより、アクセント型を抽出することもできる。

10

【0075】

このように、変形例 2 に係る感情推定装置 200 は、音声の発話時の感情状態を音声の強度の変化情報を利用して解析するので、感情推定の推定精度を向上することができる。

【0076】

(変形例 3)

変形例 3 では、音声の特徴量として音素の継続時間長を利用する場合について説明する。怒ったり喜んだりした状態で発話すると音素の継続時間長は短くなる傾向があり、退屈な状態や悲しい状態で発話すると音素の継続時間長が長くなる傾向があるので、この傾向を利用するものである。

【0077】

具体的には、アクセント型抽出部 132 は、モーラ区間に含まれる母音の継続時間長と、教師データに含まれる同じ母音の平均継続時間長とを比較し、モーラ区間に含まれる母音の継続時間長が平均継続時間長よりも長い場合は「H」を、短い場合は「L」を付与する。これにより、アクセント型決定部 130 は、アクセント型を決定する。

20

【0078】

実施形態 1、変形例 1、変形例 2 の説明では、解析区間であるアクセント句の区間における音声の特徴量の平均値とモーラ区間の平均値とを比較した。しかし、音素の継続時間長で比較する場合、感情推定器生成装置 100 のアクセント型抽出部 132 は、平均継続時間長を解析区間内の音声データの平均ではなく、教師データ全体の平均継続時間長と比較する。母音によって継続時間長は異なるので、異なる母音の継続時間長と比較することはできない。アクセント句に含まれる同じ母音の数が少ないため、平均継続時間長のバラツキが大きくなり、誤判定の要因となるので、教師データ全体の平均をとることが好ましい。

30

【0079】

一方、感情推定装置 200 のアクセント型抽出部 242 は、話者分類部 220 が分類した話者ごとの音声データについて、母音ごとに平均継続時間長を計算することが好ましい。

【0080】

音素の継続時間長によりアクセント型を分類する場合は、実験データに基づいてクラス分けの仕方を検討する必要がある。その他の説明は実施形態 1 の説明と同じである。

40

【0081】

このように、変形例 3 に係る感情推定装置 200 は、音声の発話時の感情状態を音素の発話時間長の変化情報を利用して解析するので、感情推定の推定精度を向上することができる。

【0082】

(変形例 4)

実施形態 1 と変形例 1 では、音声の特徴量として音声のピッチ情報を利用してアクセント型を抽出する技術の説明をした。また、変形例 2 では、音声の強度情報を利用してアクセント型を抽出する技術を紹介し、変形例 3 では、音素の継続時間長を利用してアクセント型を抽出する技術を紹介した。アクセント型を抽出する場合、これらの技術を単独で使

50

用することもできるが、音声のピッチ情報と音声の強度情報のように2つ以上の技術を組み合わせてアクセント型を抽出することもできる。2つ以上の情報を組み合わせるとアクセント型の種類が増えることになるが、感情推定の精度を向上させることができる。

【0083】

なお、上記の説明では、音声の特徴量として、音声の強度、音声のピッチ、音素の継続時間長を例にして説明したが、これに限定する必要はない。例えば、音声の強度の変化量、音声のピッチの変化量、音素の継続時間長の変化量等を抽出してアクセント型を決定することもできる。

【0084】

（変形例5）

実施形態1の説明では、解析対象の文に含まれるアクセント句ごとの感情推定結果に基づいて、最も多かった感情をその文を発話したときの発話者の感情として推定する技術について説明を行った。しかし、統合処理の仕方はこれに限定する必要はない。例えば、「少し驚きを伴った喜び」のように、複数の感情を含む推定を行うこともできる。感情推定器を構成する分類器では、特徴量をベクトルとして取得し、そのベクトルと識別閾値との距離に基づいて、いずれの感情に分類するかを決める場合が多い。例えば、「坊主が」、「屏風に」、「上手に」、「坊主の」、「絵を」、「描いた」の7つのアクセント句に対応する特徴量を、図15に示す1から7に示す位置ベクトルで表し、7つの位置ベクトルを合成した平均ベクトルが、図15に「平均」で示した位置ベクトルであったとする。この場合、位置ベクトル「平均」は、喜びの領域に属しているが、喜びと驚きの境界に近い位置に存在する。このような場合には、「少し驚きの感情が混在している可能性がある」というニュアンスを含めた感情推定結果を出力するようにしてもよい。

【0085】

図8と図15とは、7次元の識別空間を2次元でイメージ表現した図であるので、複雑な例を表現することは困難である。しかし、感情推定器を構成する分類器の中では、それぞれの識別境界との距離を数値で計算することが可能である。したがって、「怒りと悲しみ」、「怒りと落胆」のように、複数の感情の組み合わせと、その感情の度合い（識別境界との距離）を数値計算することが可能である。さらに、複数の閾値を設定することにより、「怒り、悲しみ、落胆」のように2つ以上の感情を含めた感情推定も可能である。また、複数の感情の複合度合いも推定することができる。

【0086】

変形例5で説明した構成および処理を設けることにより、感情推定器生成装置100は、複数の感情の度合いを推定可能な感情推定器を生成することが可能となる。また、複数の感情の度合いを推定可能な感情推定器を内蔵する感情推定装置200は、発話者の複数の感情度合いを推定することができる。

【0087】

なお、実施形態1の説明では、発話者の感情状態を悲しみ、退屈、怒り、驚き、落胆、嫌悪、喜び、の7種類に分類する説明をしたが、感情の分類方法はこれに限定する必要はない。例えば、喜、怒、哀、楽の4種類に分類してもよい。

【0088】

また、実施形態1の説明では、発話者の発話時の感情状態を、悲しみ、退屈、怒り、驚き、落胆、嫌悪、喜び、の7つの感情状態の何れかに分類する場合について説明し、いずれにも該当しない教師データは除外する説明をした。しかし、発話者の感情として、「普通」という感情状態を設け、7つの感情に分類できなかった教師データを感情「普通」に分類するようにしてもよい。これにより、感情推定器生成装置100は、発話者の感情を7つの感情に「普通」を加えた8つの感情に推定可能な感情推定器を生成することができる。また、感情推定装置200は、発話者の感情を7つの感情に「普通」を加えた8つの感情に推定することができる。

【0089】

また、実施形態1の説明では、解析区間をアクセント句の区間単位とする説明をしたが

、解析区間はこれに限定する必要はない。例えば、解析区間を単語の発話区間としてもよいし、息継ぎ区間である呼気段落区間としてもよいし、文の発話区間としてもよい。解析区間を文の発話区間とした場合には、統合部280は、文単位で発話者の感情を推定してもよいし、さらに複数の文をまとめた単位で発話者の感情を推定するようにしてもよい。

【0090】

また、式1の説明では、平均値を用いてその区間の特徴量を代表する処理について説明したが、平均値の代わりに中央値を用いて処理を行ってもよい。また、最も低い周波数を代表値として処理を行うようにしてもよい。

【0091】

また、式2の説明では、中央値を用いてその区間の特徴量を代表する処理について説明したが、中央値の代わりに平均値を用いて処理を行ってもよい。

【0092】

また、本発明に係る機能を実現するための構成を予め備えた感情推定器生成装置100、感情推定装置200として提供できることはもとより、プログラムの適用により、既存のパーソナルコンピュータや情報端末機器等を、本発明に係る感情推定器生成装置100、感情推定装置200として機能させることもできる。すなわち、上記実施形態で例示した感情推定器生成装置100、感情推定装置200による各機能構成を実現させるためのプログラムを、既存のパーソナルコンピュータや情報端末機器等を制御するCPU等が実行できるように適用することで、本発明に係る感情推定器生成装置100、感情推定装置200として機能させることができる。また、本発明に係る感情推定器生成方法及び感情推定方法は、感情推定器生成装置100、感情推定装置200を用いて実施できる。

【0093】

また、このようなプログラムの適用方法は任意である。プログラムを、例えば、コンピュータが読取可能な記録媒体（CD-ROM（Compact Disc Read-Only Memory）、DVD（Digital Versatile Disc）、MO（Magneto Optical disc）等）に格納して適用できる他、インターネット等のネットワーク上のストレージにプログラムを格納しておき、これをダウンロードさせることにより適用することもできる。

【0094】

以上、本発明の好ましい実施形態について説明したが、本発明は係る特定の実施形態に限定されるものではなく、本発明には、特許請求の範囲に記載された発明とその均等の範囲が含まれる。以下に、本願出願の当初の特許請求の範囲に記載された発明を付記する。

【0095】

（付記1）

教師データの元となる音声データの特徴量を解析する解析区間を設定する解析区間設定ステップと、

前記解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、前記解析区間に含まれる音声データの特徴量の変化パターンとして決定する変化パターン決定ステップと、

前記特徴量の変化パターンごとに分類された音声データを教師データとして、前記特徴量の変化パターンごとに、音声が発話したときの発話者の感情を推定する感情推定器を生成する感情推定器生成ステップと、

を含む感情推定器生成方法。

【0096】

（付記2）

前記変化パターン決定ステップは、

前記解析区間に含まれる音声データを、音節の単位であるモーラ区間に分割するモーラ区間抽出ステップと、

前記解析区間における音声データの特徴量の平均値と、前記モーラ区間における音声データの特徴量の平均値と、をモーラ区間ごとに比較した比較結果に基づいて、前記解析区間の音声データを発話したときのモーラ区間ごとに変化する音声の特徴量の変化パターン

を抽出する変化パターン抽出ステップと、

を含むことを特徴とする付記 1 に記載の感情推定器生成方法。

【0097】

(付記 3)

前記変化パターン抽出ステップでは、音声の特徴量として音声データから抽出した音声の基本周波数を用い、前記解析区間における音声の平均基本周波数と、前記モーラ区間における音声の平均基本周波数と、をモーラ区間ごとに比較し、モーラ区間の音声の平均基本周波数が解析区間の音声の平均基本周波数よりも高い場合には H i g h を、低い場合には L o w を付与し、モーラ区間ごとに H i g h と L o w に変化する音声の特徴量の変化パターンを抽出する、

10

ことを特徴とする付記 2 に記載の感情推定器生成方法。

【0098】

(付記 4)

前記変化パターン抽出ステップでは、音声の特徴量として音声データから抽出した音声の強度を用い、前記解析区間における音声の平均強度と、前記モーラ区間における音声の平均強度と、をモーラ区間ごとに比較し、モーラ区間の音声の平均強度が解析区間の音声の平均強度よりも高い場合には H i g h を、低い場合には L o w を付与し、モーラ区間ごとに H i g h と L o w に変化する音声の特徴量の変化パターンを抽出する、

ことを特徴とする付記 2 に記載の感情推定器生成方法。

【0099】

20

(付記 5)

前記変化パターン抽出ステップでは、音声の特徴量として音声データから抽出した音素の継続時間長を用い、前記解析区間における音素の平均継続時間長と、前記モーラ区間における音素の平均継続時間長と、をモーラ区間ごとに比較し、モーラ区間の音素の平均継続時間長が解析区間の音素の平均継続時間長よりも長い場合には H i g h を、短い場合には L o w を付与し、モーラ区間ごとに H i g h と L o w に変化する音声の特徴量の変化パターンを抽出する、

ことを特徴とする付記 2 に記載の感情推定器生成方法。

【0100】

(付記 6)

30

前記変化パターン抽出ステップでは、音声の特徴量として、音声の基本周波数、音声の強度、音素の継続時間長の少なくとも何れか 1 つを使用して音声の特徴量の変化パターンを抽出する、

ことを特徴とする付記 2 から 5 の何れか一つに記載の感情推定器生成方法。

【0101】

(付記 7)

前記解析区間設定ステップでは、音声データを、言語の意味を持つ最小の単位である形態素に分割し、当該形態素の後で発話された助詞又は助動詞と結合したアクセント句の区間を前記解析区間として設定する、

ことを特徴とする付記 1 から 6 の何れか一つに記載の感情推定器生成方法。

40

【0102】

(付記 8)

前記モーラ区間抽出ステップでは、音声データをテキスト表示した場合に、仮名文字 1 文字を 1 モーラ区間とし、小書きの仮名文字はその前の仮名文字と一緒にして 1 モーラ区間とし、長音は独立して 1 モーラ区間とする、

ことを特徴とする付記 2 に記載の感情推定器生成方法。

【0103】

(付記 9)

前記感情推定器は、発話者の発話時の感情を、悲しみ、退屈、怒り、驚き、落胆、嫌悪、喜び、の何れかの感情であると推定する、

50

ことを特徴とする付記 1 から 8 の何れか一つに記載の感情推定器生成方法。

【0104】

(付記 10)

前記複数のクラスに分類された変化パターンを設定する変化パターン設定ステップを含む、

ことを特徴とする付記 1 から 9 の何れか一つに記載の感情推定器生成方法。

【0105】

(付記 11)

教師データの元となる音声データの特徴量を解析する解析区間を設定する解析区間設定手段と、

前記解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、前記解析区間に含まれる音声データの特徴量の変化パターンとして決定する変化パターン決定手段と、

前記特徴量の変化パターンごとに分類された音声データを教師データとして、前記特徴量の変化パターンごとに、音声を発話したときの発話者の感情を推定する感情推定器を生成する感情推定器生成手段と、

を含む感情推定器生成装置。

【0106】

(付記 12)

コンピュータを

教師データの元となる音声データの特徴量を解析する解析区間を設定する解析区間設定手段、

前記解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、前記解析区間に含まれる音声データの特徴量の変化パターンとして決定する変化パターン決定手段、

前記特徴量の変化パターンごとに分類された音声データを教師データとして、前記特徴量の変化パターンごとに、音声を発話したときの発話者の感情を推定する感情推定器を生成する感情推定器生成手段、

として機能させるためのプログラム。

【0107】

(付記 13)

解析対象とする音声データの特徴量を解析する解析区間を設定する解析区間設定ステップと、

前記解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、前記解析区間に含まれる音声データの特徴量の変化パターンとして決定する変化パターン決定ステップと、

前記特徴量の変化パターンごとに、同じ特徴量の変化パターンを有する教師データに基づいて生成された感情推定器を用いて、前記解析区間の音声を発話した時の発話者の感情を推定する感情推定ステップと、

を含む感情推定方法。

【0108】

(付記 14)

解析対象とする音声データの特徴量を解析する解析区間を設定する解析区間設定手段と、

前記解析区間に含まれる音声データの特徴量の変化するパターンを、複数のクラスに分類された変化パターンに基づいて、前記解析区間に含まれる音声データの特徴量の変化パターンとして決定する変化パターン決定手段と、

前記特徴量の変化パターンごとに、同じ特徴量の変化パターンを有する教師データに基づいて生成された感情推定器を用いて、前記解析区間の音声を発話した時の発話者の感情を推定する感情推定手段と、

10

20

30

40

50

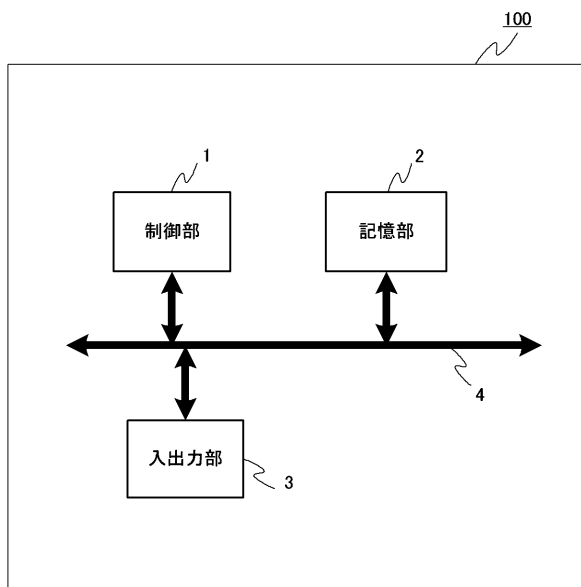
を備えた感情推定装置。

【符号の説明】

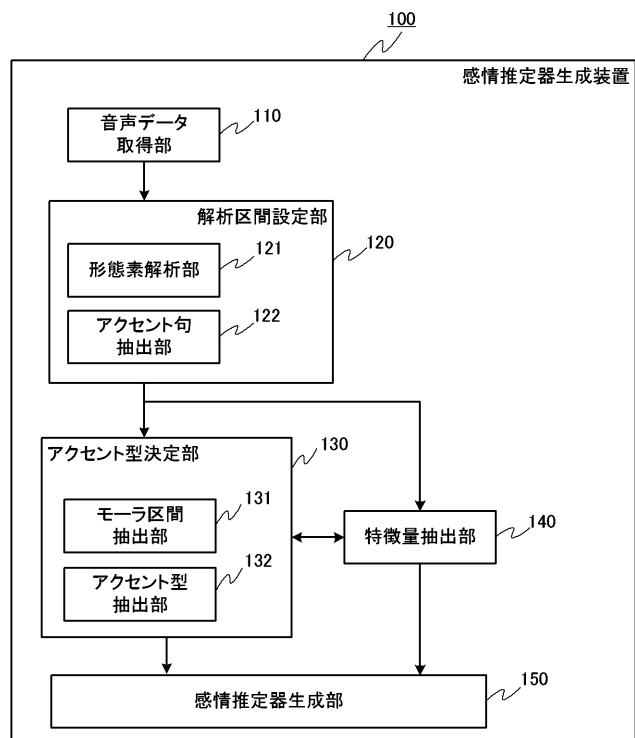
【0109】

1…制御部、2…記憶部、3…入出力部、4…バス、100…感情推定器生成装置、110、210…音声データ取得部、120、230…解析区間設定部、121、231…形態素解析部、122、232…アクセント句抽出部、130、240…アクセント型決定部、131、241…モーラ区間抽出部、132、242…アクセント型抽出部、140、260…特徴量抽出部、150…感情推定器生成部、200…感情推定装置、220…話者分割部、250…選択部、270…感情推定部、280…統合部

【図1】



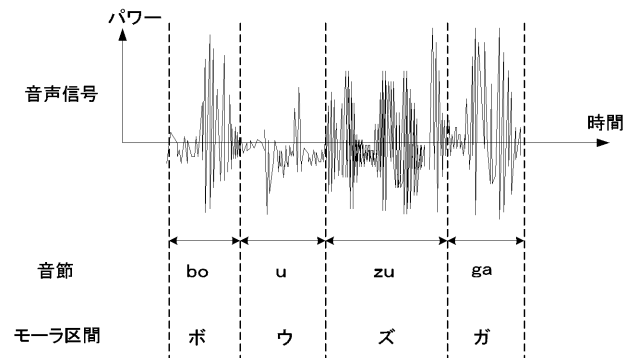
【図2】



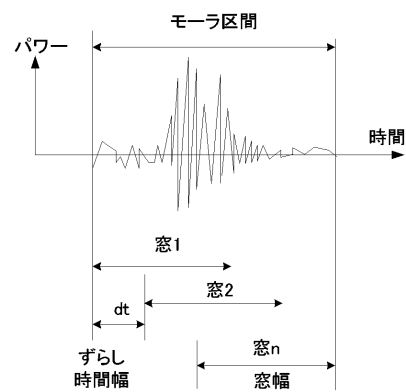
【図 3】

形態素	備考
坊主	名詞
が	助詞
屏風	名詞
に	助詞
上手	名詞
に	助詞
坊主	名詞
の	助詞
絵	名詞
を	助詞
書い	動詞
た	助動詞

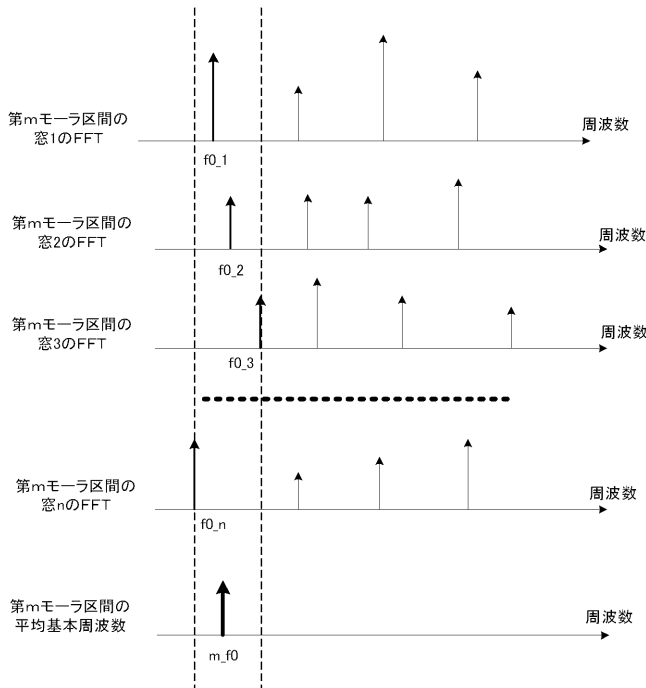
【図 4】



【図 5】



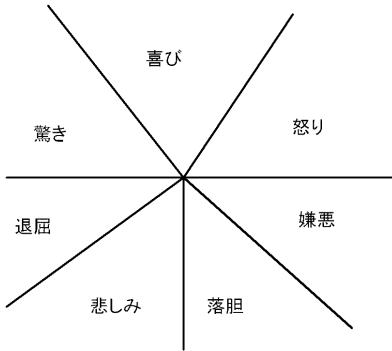
【図 6】



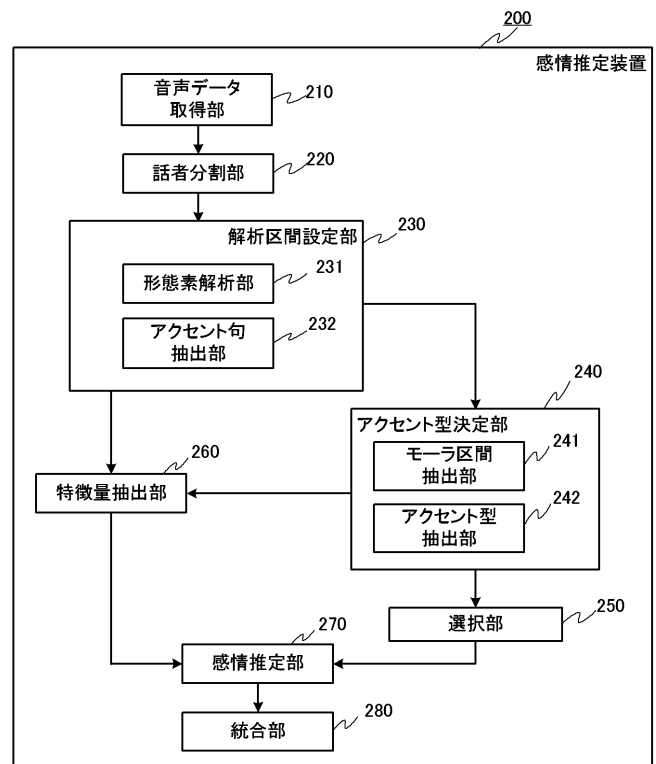
【図 7】

モーラ数	アクセント型	クラス
2	HL	1
	LH	2
3	HLL	3
	LHL	4
	LLH	5
4	HLLL	6
	LHLL	7
	LHHL	8
	LHHH	9
5	HLLLL	10
	LHLLL	11
	LHHLL	12
	LHHHL	13
	LHHHH	14
6	HLLLLL	15
	LHLLLL	16
	LHHLLL	17
	LHHHLL	18
	LHHHHL	19
	LHHHHH	20

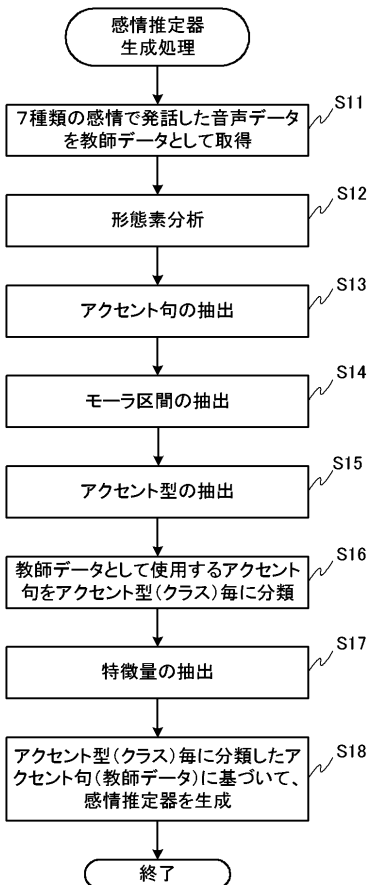
【図 8】



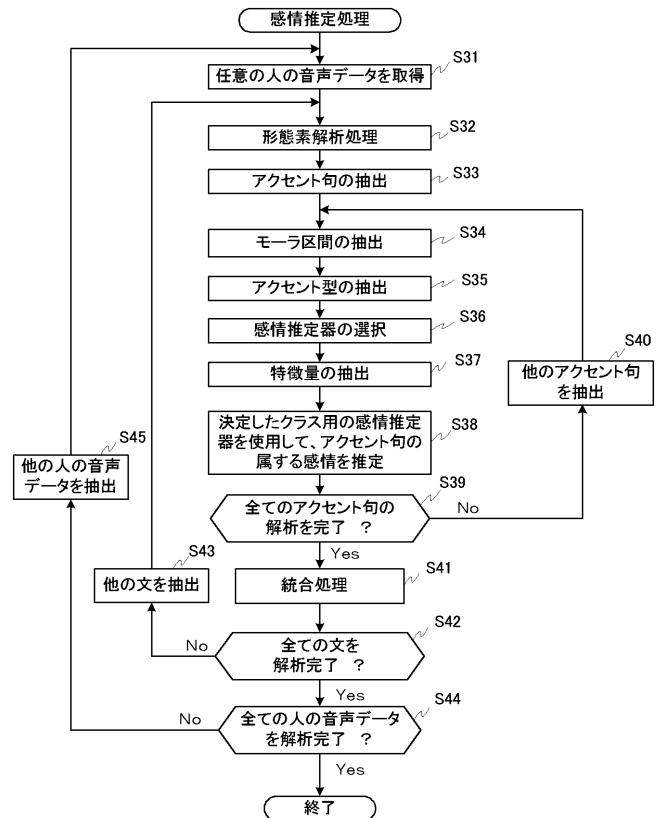
【図 9】



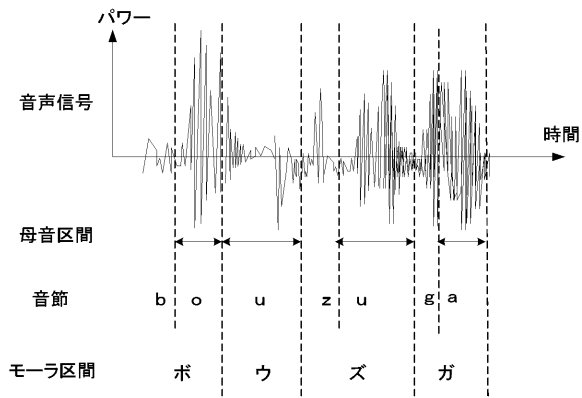
【図 10】



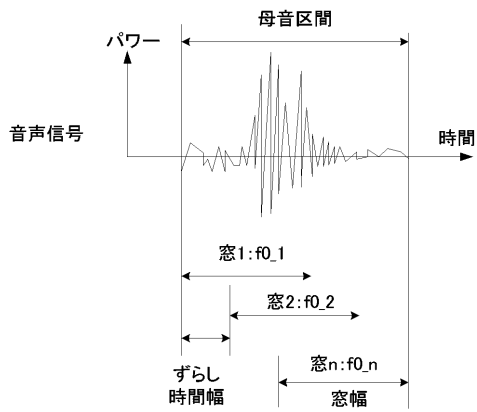
【図 11】



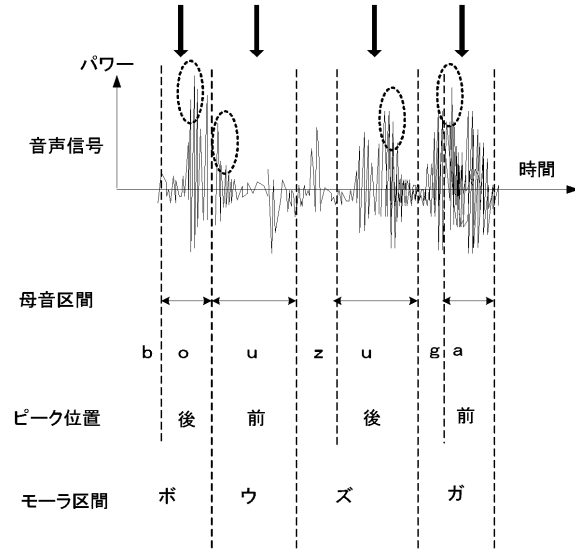
【図 1 2】



【図 1 3】



【図 1 4】



【図 1 5】

