

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3702978号
(P3702978)

(45) 発行日 平成17年10月5日(2005.10.5)

(24) 登録日 平成17年7月29日(2005.7.29)

(51) Int. Cl.⁷

F I

G 1 0 L 15/24
G 0 6 N 3/00
G 1 0 L 15/06

G 1 0 L 3/00 5 7 1 Q
G 0 6 N 3/00 5 6 0 G
G 1 0 L 3/00 5 2 1 C

請求項の数 7 (全 21 頁)

<p>(21) 出願番号 特願平8-347312 (22) 出願日 平成8年12月26日(1996.12.26) (65) 公開番号 特開平10-187186 (43) 公開日 平成10年7月14日(1998.7.14) 審査請求日 平成15年2月21日(2003.2.21)</p>	<p>(73) 特許権者 000002185 ソニー株式会社 東京都品川区北品川6丁目7番35号 (74) 代理人 100082131 弁理士 稲本 義雄 (72) 発明者 近藤 哲二郎 東京都品川区北品川6丁目7番35号 ソ ニー株式会社内 審査官 石丸 昌平</p>
--	--

最終頁に続く

(54) 【発明の名称】 認識装置および認識方法、並びに学習装置および学習方法

(57) 【特許請求の範囲】

【請求項1】

所定の認識対象を、複数の異なる種類の入力データから認識する認識装置であって、
前記複数の異なる種類の入力データを分析し、その分析結果の少なくとも一部に基づいて、
所定のクラスに分類するクラス分類を行うクラス分類手段と、

前記複数の異なる種類の入力データの分析結果の少なくとも一部を統合した統合パラメータを構成する統合パラメータ構成手段と、

前記クラス分類手段が出力するクラスごとに、所定の標準のパラメータが登録されているテーブルを記憶している標準パラメータ記憶手段と、

前記クラス分類手段が出力するクラスに対応するテーブルに登録された標準のパラメータと、前記統合パラメータとに基づいて、前記認識対象を認識する認識手段と
を備え、

前記標準パラメータ記憶手段に記憶されている前記所定の標準のパラメータは、前記認識対象に対応するパラメータであって、複数の異なる種類の学習用データの分析結果の少なくとも一部を統合した学習用統合パラメータを前記クラスごとに求める学習により得られたものである

ことを特徴とする認識装置。

【請求項2】

所定の認識対象を、複数の異なる種類の入力データから認識する認識装置であって、
前記複数の異なる種類の入力データからクラス分類用データを構成し、前記クラス分類

10

20

用データを所定のクラスに分類し、その結果得られる第1のクラス分類結果を出力する第1のクラス分類手段と、

前記入力データを、前記第1のクラス分類結果に応じて分析し、前記入力データの特徴パラメータを求める分析手段と、

前記分析手段からの特徴パラメータの少なくとも一部に基づいて、前記入力データを所定のクラスに分類し、その結果得られる第2のクラス分類結果を出力する第2のクラス分類手段と、

前記特徴パラメータの少なくとも一部を統合した統合パラメータを構成する統合パラメータ構成手段と、

前記第1のクラス分類手段により分類されたクラスごとであって、前記第2のクラス分類手段により分類されたクラスごとに、所定の標準のパラメータが登録されているテーブルを記憶する標準パラメータ記憶手段と、

前記第1および第2のクラス分類手段により分類された両方のクラスに対応するテーブルに登録された標準のパラメータと、前記統合パラメータとに基づいて、前記認識対象を認識する認識手段と

を備え、

前記標準パラメータ記憶手段に記憶されている前記所定の標準のパラメータは、前記認識対象に対応するパラメータであって、前記第1のクラス分類手段によるクラスごとであって前記第2のクラス分類手段によるクラスごとに、複数の異なる種類の学習用データの特徴パラメータの少なくとも一部を統合した学習用統合パラメータを求める学習により得られたものである

ことを特徴とする認識装置。

【請求項3】

前記認識対象は、所定の話者が発した音声であり、

前記複数の異なる種類の入力データは、少なくとも、その音声データと、話者の口を撮影して得られる画像データとを含む

ことを特徴とする請求項1に記載の認識装置。

【請求項4】

所定の認識対象を、複数の異なる種類の入力データから認識する認識方法であって、

前記複数の異なる種類の入力データを分析し、その分析結果の少なくとも一部に基づいて、所定のクラスに分類するクラス分類を行うとともに、前記複数の異なる種類の入力データの分析結果の少なくとも一部を統合した統合パラメータを構成し、

前記クラス分類の結果得られるクラスに対応し、前記認識対象に対応するパラメータであって、複数の異なる種類の学習用データの分析結果の少なくとも一部を統合した学習用統合パラメータを前記クラスごとに求める学習により得られた標準のパラメータが登録されたテーブルと、前記統合パラメータとに基づいて、前記認識対象を認識する

ことを特徴とする認識方法。

【請求項5】

所定の認識対象を、複数の異なる種類の入力データから認識するのに用いる標準のパラメータを求める学習装置であって、

学習用の複数の異なる種類の入力データを分析し、その特徴パラメータを求める分析手段と、

前記学習用の複数の異なる種類の入力データを、前記特徴パラメータの少なくとも一部に基づいて、所定のクラスに分類するクラス分類を行うクラス分類手段と、

前記特徴パラメータの少なくとも一部を統合した統合パラメータを構成する統合パラメータ構成手段と、

前記統合パラメータを、前記クラス分類手段が出力するクラスごとに分類して記憶する記憶手段と

を備えることを特徴とする学習装置。

【請求項6】

10

20

30

40

50

所定の認識対象を、複数の異なる種類の入力データから認識するのに用いる標準のパラメータを求める学習装置であって、

学習用の複数の異なる種類の入力データからクラス分類用データを構成し、前記クラス分類用データを所定のクラスに分類し、その結果得られる第1のクラス分類結果を出力する第1のクラス分類手段と、

前記学習用の複数の異なる種類の入力データを、前記第1のクラス分類結果に応じて分析し、特徴パラメータを求める分析手段と、

前記分析手段からの特徴パラメータの少なくとも一部に基づいて、前記入力データを所定のクラスに分類し、その結果得られる第2のクラス分類結果を出力する第2のクラス分類手段と、

前記特徴パラメータの少なくとも一部を統合した統合パラメータを構成する統合パラメータ構成手段と、

前記統合パラメータを、前記第1のクラス分類手段が出力するクラスごとであって、前記第2のクラス分類手段が出力するクラスごとに分類して記憶する記憶手段と

を備えることを特徴とする学習装置。

【請求項7】

所定の認識対象を、複数の異なる種類の入力データから認識するのに用いる標準のパラメータを求める学習方法であって、

学習用の複数の異なる種類の入力データを分析し、その特徴パラメータを求め、

前記学習用の複数の異なる種類の入力データを、前記特徴パラメータの少なくとも一部に基づいて、所定のクラスに分類するクラス分類を行うとともに、前記特徴パラメータの少なくとも一部を統合した統合パラメータを構成し、

前記統合パラメータを、前記クラス分類の結果得られるクラスごとに分類して記憶することを特徴とする学習方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、認識装置および認識方法、並びに学習装置および学習方法に関する。特に、例えば、音声や、物体などを、その音声データや画像データだけでなく、他のデータをも用いて認識することにより、その認識精度を向上させることができるようにする認識装置および認識方法、並びに学習装置および学習方法に関する。

【0002】

【従来の技術】

従来の、例えば、音声を認識する音声認識装置では、マイク（マイクロフォン）から出力される音声データを音響分析し、その分析結果を用いて、ユーザの発した音声を認識するようになされている。

【0003】

【発明が解決しようとする課題】

しかしながら、従来の音声認識装置では、マイクからの音声データの分析結果だけを用いて音声認識を行うようになされているため、その認識性能を向上させるのに、ある程度の限界があった。

【0004】

即ち、話者の発話の認識にあたっては、マイクからの音声データだけでなく、その相手の表情や口の動きなども重要な要素であると考えられ、従って、そのような種々の要素を考慮すべきである。

【0005】

また、音声認識装置は、防音室などのように、認識対象である音声だけがマイクに入力されるような環境で使用されることはほとんどなく、それ以外の音である様々な雑音のある環境下で使用されることが多い。即ち、最近では、音声認識装置を備えた、例えば、ナビゲーションシステムが実用化されつつあるが、この場合、音声認識装置には、認識すべき

10

20

30

40

50

話者の音声の他、自動車に装備されたCD (Compact Disc) プレーヤが出力する音や、エンジン音、エアコンディショナの音、その他の雑音が入力される。そして、このような雑音が混ざった音声データから、その雑音だけを取り除くのは困難であるため、認識性能を向上させるには、認識対象の音声以外の音である雑音も加味する必要がある。

【0006】

さらに、従来の音声認識装置では、マイクから出力される音声データに対して、一定の音響分析処理を施して、その特徴パラメータを求め、その特徴パラメータのパラメータ空間における距離などに基づいて、音声認識されるようになされている。しかしながら、認識にあたって重要な特徴パラメータは、音声認識装置が使用される環境によって変化すると考えられる。

10

【0007】

本発明は、このような状況に鑑みてなされたものであり、音声その他の認識を行う認識装置の認識性能を向上させることができるようにするものである。

【0008】

【課題を解決するための手段】

請求項1に記載の認識装置は、複数の異なる種類の入力データを分析し、その分析結果の少なくとも一部に基づいて、所定のクラスに分類するクラス分類を行うクラス分類手段と、複数の異なる種類の入力データの分析結果の少なくとも一部を統合した統合パラメータを構成する統合パラメータ構成手段と、クラス分類手段が出力するクラスごとに、所定の標準のパラメータが登録されているテーブルを記憶している標準パラメータ記憶手段と、クラス分類手段が出力するクラスに対応するテーブルに登録された標準のパラメータと、統合パラメータとに基づいて、認識対象を認識する認識手段とを備え、標準パラメータ記憶手段に記憶されている所定の標準のパラメータは、認識対象に対応するパラメータであって、複数の異なる種類の学習用データの分析結果の少なくとも一部を統合した学習用統合パラメータをクラスごとに求める学習により得られたものであることを特徴とする。

20

請求項2に記載の認識装置は、複数の異なる種類の入力データからクラス分類用データを構成し、クラス分類用データを所定のクラスに分類し、その結果得られる第1のクラス分類結果を出力する第1のクラス分類手段と、入力データを、第1のクラス分類結果に応じて分析し、入力データの特徴パラメータを求める分析手段と、分析手段からの特徴パラメータの少なくとも一部に基づいて、入力データを所定のクラスに分類し、その結果得られる第2のクラス分類結果を出力する第2のクラス分類手段と、特徴パラメータの少なくとも一部を統合した統合パラメータを構成する統合パラメータ構成手段と、第1のクラス分類手段により分類されたクラスごとであって、第2のクラス分類手段により分類されたクラスごとに、所定の標準のパラメータが登録されているテーブルを記憶する標準パラメータ記憶手段と、第1および第2のクラス分類手段により分類された両方のクラスに対応するテーブルに登録された標準のパラメータと、統合パラメータとに基づいて、認識対象を認識する認識手段とを備え、標準パラメータ記憶手段に記憶されている所定の標準のパラメータは、認識対象に対応するパラメータであって、第1のクラス分類手段によるクラスごとであって第2のクラス分類手段によるクラスごとに、複数の異なる種類の学習用データの特徴パラメータの少なくとも一部を統合した学習用統合パラメータを求め

30

40

【0009】

請求項4に記載の認識方法は、複数の異なる種類の入力データを分析し、その分析結果の少なくとも一部に基づいて、所定のクラスに分類するクラス分類を行うとともに、複数の異なる種類の入力データの分析結果の少なくとも一部を統合した統合パラメータを構成し、クラス分類の結果得られるクラスに対応し、認識対象に対応するパラメータであって、複数の異なる種類の学習用データの分析結果の少なくとも一部を統合した学習用統合パラメータをクラスごとに求める学習により得られた標準のパラメータが登録されたテーブルと、統合パラメータとに基づいて、認識対象を認識することを特徴とする。

【0010】

50

請求項5に記載の学習装置は、学習用の複数の異なる種類の入力データを分析し、その特徴パラメータを求める分析手段と、学習用の複数の異なる種類の入力データを、特徴パラメータの少なくとも一部に基づいて、所定のクラスに分類するクラス分類を行うクラス分類手段と、特徴パラメータの少なくとも一部を統合した統合パラメータを構成する統合パラメータ構成手段と、統合パラメータを、クラス分類手段が出力するクラスごとに分類して記憶する記憶手段とを備えることを特徴とする。

請求項6に記載の学習装置は、学習用の複数の異なる種類の入力データからクラス分類用データを構成し、クラス分類用データを所定のクラスに分類し、その結果得られる第1のクラス分類結果を出力する第1のクラス分類手段と、学習用の複数の異なる種類の入力データを、第1のクラス分類結果に応じて分析し、特徴パラメータを求める分析手段と、分析手段からの特徴パラメータの少なくとも一部に基づいて、入力データを所定のクラスに分類し、その結果得られる第2のクラス分類結果を出力する第2のクラス分類手段と、特徴パラメータの少なくとも一部を統合した統合パラメータを構成する統合パラメータ構成手段と、統合パラメータを、第1のクラス分類手段が出力するクラスごとであって、第2のクラス分類手段が出力するクラスごとに分類して記憶する記憶手段とを備えることを特徴とする。

10

【0011】

請求項7に記載の学習方法は、学習用の複数の異なる種類の入力データを分析し、その特徴パラメータを求め、学習用の複数の異なる種類の入力データを、特徴パラメータの少なくとも一部に基づいて、所定のクラスに分類するクラス分類を行うとともに、特徴パラメータの少なくとも一部を統合した統合パラメータを構成し、統合パラメータを、クラス分類の結果得られるクラスごとに分類して記憶することを特徴とする。

20

【0012】

請求項1に記載の認識装置においては、クラス分類手段は、複数の異なる種類の入力データを分析し、その分析結果の少なくとも一部に基づいて、所定のクラスに分類するクラス分類を行い、統合パラメータ構成手段は、複数の異なる種類の入力データの分析結果の少なくとも一部を統合した統合パラメータを構成するようになされている。標準パラメータ記憶手段は、クラス分類手段が出力するクラスごとに、所定の標準のパラメータが登録されているテーブルを記憶しており、認識手段は、クラス分類手段が出力するクラスに対応するテーブルに登録された標準のパラメータと、統合パラメータとに基づいて、認識対象を認識するようになされている。ここで、標準パラメータ記憶手段に記憶されている所定の標準のパラメータは、認識対象に対応するパラメータであって、複数の異なる種類の学習用データの分析結果の少なくとも一部を統合した学習用統合パラメータをクラスごとに求める学習により得られたものであるようになされている。

30

請求項2に記載の認識装置においては、第1のクラス分類手段は、複数の異なる種類の入力データからクラス分類用データを構成し、クラス分類用データを所定のクラスに分類し、その結果得られる第1のクラス分類結果を出力し、分析手段は、入力データを、第1のクラス分類結果に応じて分析し、入力データの特徴パラメータを求め、第2のクラス分類手段は、分析手段からの特徴パラメータの少なくとも一部に基づいて、入力データを所定のクラスに分類し、その結果得られる第2のクラス分類結果を出力するようになされている。統合パラメータ構成手段は、特徴パラメータの少なくとも一部を統合した統合パラメータを構成し、標準パラメータ記憶手段は、第1のクラス分類手段により分類されたクラスごとであって、第2のクラス分類手段により分類されたクラスごとに、所定の標準のパラメータが登録されているテーブルを記憶し、認識手段は、第1および第2のクラス分類手段により分類された両方のクラスに対応するテーブルに登録された標準のパラメータと、統合パラメータとに基づいて、認識対象を認識するようになされている。ここで、標準パラメータ記憶手段に記憶されている所定の標準のパラメータは、認識対象に対応するパラメータであって、第1のクラス分類手段によるクラスごとであって第2のクラス分類手段によるクラスごとに、複数の異なる種類の学習用データの特徴パラメータの少なくとも一部を統合した学習用統合パラメータを求める学習により得られたものであるようにな

40

50

されている。

【0013】

請求項4に記載の認識方法においては、複数の異なる種類の入力データを分析し、その分析結果の少なくとも一部に基づいて、所定のクラスに分類するクラス分類を行うとともに、複数の異なる種類の入力データの分析結果の少なくとも一部を統合した統合パラメータを構成し、クラス分類の結果得られるクラスに対応し、認識対象に対応するパラメータであって、複数の異なる種類の学習用データの分析結果の少なくとも一部を統合した学習用統合パラメータをクラスごとに求める学習により得られた標準のパラメータが登録されたテーブルと、統合パラメータとに基づいて、認識対象を認識するようになされている。

【0014】

請求項5に記載の学習装置においては、分析手段は、学習用の複数の異なる種類の入力データを分析し、その特徴パラメータを求め、クラス分類手段は、学習用の複数の異なる種類の入力データを、特徴パラメータの少なくとも一部に基づいて、所定のクラスに分類するクラス分類を行い、統合パラメータ構成手段は、特徴パラメータの少なくとも一部を統合した統合パラメータを構成するようになされている。記憶手段は、統合パラメータを、クラス分類手段が出力するクラスごとに分類するようになされている。

請求項6に記載の学習装置においては、第1のクラス分類手段は、学習用の複数の異なる種類の入力データからクラス分類用データを構成し、クラス分類用データを所定のクラスに分類し、その結果得られる第1のクラス分類結果を出力し、分析手段は、学習用の複数の異なる種類の入力データを、第1のクラス分類結果に応じて分析し、特徴パラメータを求め、第2のクラス分類手段は、分析手段からの特徴パラメータの少なくとも一部に基づいて、入力データを所定のクラスに分類し、その結果得られる第2のクラス分類結果を出力し、統合パラメータ構成手段は、特徴パラメータの少なくとも一部を統合した統合パラメータを構成し、記憶手段は、統合パラメータを、第1のクラス分類手段が出力するクラスごとであって、第2のクラス分類手段が出力するクラスごとに分類して記憶するようになされている。

【0015】

請求項7に記載の学習方法においては、学習用の複数の異なる種類の入力データを分析し、その特徴パラメータを求め、学習用の複数の異なる種類の入力データを、特徴パラメータの少なくとも一部に基づいて、所定のクラスに分類するクラス分類を行うとともに、特徴パラメータの少なくとも一部を統合した統合パラメータを構成し、統合パラメータを、クラス分類の結果得られるクラスごとに分類して記憶するようになされている。

【0016】

【発明の実施の形態】

以下に、本発明の実施の形態を説明するが、その前に、特許請求の範囲に記載の発明の各手段と以下の実施の形態との対応関係を明らかにするために、各手段の後の括弧内に、対応する実施の形態（但し、一例）を付加して、本発明の特徴を記述すると、次のようになる。

【0017】

即ち、請求項1に記載の認識装置は、所定の認識対象を、複数の異なる種類の入力データから認識する認識装置であって、複数の異なる種類の入力データを分析し、その分析結果の少なくとも一部に基づいて、所定のクラスに分類するクラス分類を行うクラス分類手段（例えば、図1に示すクラス分類部25や、図6、図8に示すクラス分類部52など）と、複数の異なる種類の入力データの分析結果の少なくとも一部を統合した統合パラメータを構成する統合パラメータ構成手段（例えば、図2や図6、図8に示す統合パラメータ構成部26など）と、クラス分類手段が出力するクラスごとに、所定の標準のパラメータが登録されているテーブルを記憶している標準パラメータ記憶手段（例えば、図2に示す標準パラメータ記憶部28や、図6、図8に示す標準パラメータ記憶部28₁乃至28_nなど）と、クラス分類手段が出力するクラスに対応するテーブルに登録された標準のパラメータと、統合パラメータとに基づいて、認識対象を認識する認識手段（例えば、図2や図

10

20

30

40

50

6、図8に示すマッチング部29など）とを備え、標準パラメータ記憶手段に記憶されている所定の標準のパラメータは、認識対象に対応するパラメータであって、複数の異なる種類の学習用データの分析結果の少なくとも一部を統合した学習用統合パラメータをクラスごとに求める学習により得られたものであることを特徴とする。

【0019】

請求項2に記載の認識装置は、所定の認識対象を、複数の異なる種類の入力データから認識する認識装置であって、複数の異なる種類の入力データからクラス分類用データを構成し、クラス分類用データを所定のクラスに分類し、その結果得られる第1のクラス分類結果を出力する第1のクラス分類手段（例えば、図6や図8に示すクラス分類用データ構成部51およびクラス分類部52など）と、入力データを、第1のクラス結果に応じて分析し、入力データの特徴パラメータを求める分析手段（例えば、図6や図8に示す前処理部20乃至23など）と、分析手段からの特徴パラメータの少なくとも一部に基づいて、入力データを所定のクラスに分類し、その結果得られる第2のクラス分類結果を出力する第2のクラス分類手段（例えば、図6や図8に示すクラス分類部25など）と、特徴パラメータの少なくとも一部を統合した統合パラメータを構成する統合パラメータ構成手段（例えば、図6や図8に示す統合パラメータ構成部26など）と、第1のクラス分類手段により分類されたクラスごとであって、第2のクラス分類手段により分類されたクラスごとに、所定の標準のパラメータが登録されているテーブルを記憶する標準パラメータ記憶手段（例えば、図6や図8に示す標準パラメータ記憶部28₁乃至28_nなど）と、第1および第2のクラス分類手段により分類された両方のクラスに対応するテーブルに登録された標準のパラメータと、統合パラメータとに基づいて、認識対象を認識する認識手段（例えば、図6や図8に示すマッチング部29など）とを備え、標準パラメータ記憶手段に記憶されている所定の標準のパラメータは、認識対象に対応するパラメータであって、第1のクラス分類手段によるクラスごとであって第2のクラス分類手段によるクラスごとに、複数の異なる種類の学習用データの特徴パラメータの少なくとも一部を統合した学習用統合パラメータを求める学習により得られたものであることを特徴とする。

【0020】

請求項5に記載の学習装置は、所定の認識対象を、複数種類の入力データから認識するのに用いる標準のパラメータを求める学習装置であって、学習用の複数の異なる種類の入力データを分析し、その特徴パラメータを求める分析手段（例えば、図4に示す前処理部40乃至43など）と、学習用の複数の異なる種類の入力データを、特徴パラメータの少なくとも一部に基づいて、所定のクラスに分類するクラス分類を行うクラス分類手段（例えば、図4に示すクラス分類部45や、図7、図9に示すクラス分類部62など）と、特徴パラメータの少なくとも一部を統合した統合パラメータを構成する統合パラメータ構成手段（例えば、図4や図7、図9に示す統合パラメータ構成部46など）と、統合パラメータを、クラス分類手段が出力するクラスごとに分類して記憶する記憶手段（例えば、図4に示すメモリ47₁や、図7、図9に示すメモリ47₁乃至47_nおよびセクタ63など）とを備えることを特徴とする。

【0022】

請求項6に記載の額種装置は、所定の認識対象を、複数の異なる種類の入力データから認識するのに用いる標準のパラメータを求める学習装置であって、学習用の複数の異なる種類の入力データからクラス分類用データを構成し、クラス分類用データを所定のクラスに分類し、その結果得られる第1のクラス分類結果を出力する第1のクラス分類手段（例えば、図7や図9に示すクラス分類用データ構成部61およびクラス分類部62など）と、学習用の複数の異なる種類の入力データを、第1のクラス分類結果に応じて分析し、特徴パラメータを求める分析手段（例えば、図7や図9に示す前処理部40乃至43など）と、分析手段からの特徴パラメータの少なくとも一部に基づいて、入力データを所定のクラスに分類し、その結果得られる第2のクラス分類結果を出力する第2のクラス分類手段（例えば、図7や図9に示すクラス分類部45など）と、特徴パラメータの少なくとも一部を統合した統合パラメータを構成する統合パラメータ構成手段（例えば、図7や図9に

10

20

30

40

50

示す統合パラメータ構成部46など)と、統合パラメータを、第1のクラス分類手段が出力するクラスごとであって、第2のクラス分類手段が出力するクラスごとに分類して記憶する記憶手段(例えば、図7や図9に示すメモリ47₁乃至47_nおよびセレクトラ63など)とを備えることを特徴とする。

【0023】

なお、勿論この記載は、各手段を上記したものに限定することを意味するものではない。

【0024】

図1は、本発明を適用したナビゲーションシステムの一実施の形態の構成例を示している。

【0025】

このナビゲーションシステムは、システム制御部1、位置測定装置2、データベース装置3、入力装置4、および出力装置5によって構成されており、自動車などに装備され、例えばボタンなどの操作による他、対話的に発せられる入力音声に対応して、装置の制御が行われるようになされている。なお、このナビゲーションシステムは、例えば携帯して用いることもできるようになされている。

10

【0026】

システム制御部1は、システムを構成する各ブロックとの間で情報の送受を行い、システム全体の動作を制御するようになされている。位置測定装置2は、GPS(Global Positioning System)衛星からの電波を受信したり、例えばジャイロスコープや車速センサなどの測定装置を用いて現在地を測定するようになされている。データベース装置3には、例えば複数のスケールの、電子化された地図情報、その他のナビゲーションに必要な情報が記録(記憶)されており、そこからは、システム制御部1からの指令に基づいて必要な情報が検索され、システム制御部1に送信されるようになされている。

20

【0027】

入力装置4には、ナビゲーションシステムに所定の処理を行わせるためのボタンやジョイスティック、音声を入力するためのマイク、ユーザを撮像するためのCCD(Charge Coupled Device)カメラ、さらには、自動車の振動を測定するための加速度センサや、湿度や温度を測定するためのセンサ、その他の各種のセンサなどが装備されている。入力装置4からは、ボタンやジョイスティックなどの操作に対応した信号がシステム制御部1に送信されるようになされている。さらに、入力装置4は、音声認識装置を有しており、入力された音声を音声認識し、その音声認識結果を、システム制御部1に送信するようにもなされている。

30

【0028】

出力装置5は、画像その他を表示するための、例えば液晶モニタやCRT(Cathod Ray Tube)、音声その他を出力するためのスピーカ、テキスト情報から合成音を生成する音声合成装置などを有しており、システム制御部1から送られてくる情報のうちの、例えば地図情報や現在位置情報などを表示したり、また音声情報を出力したりするようになされている。さらに、出力装置5は、システム制御部1からテキスト情報が送信されてきた場合には、そのテキスト情報に対応する音声を音声合成装置に音声合成させて出力するようにもなされている。

40

【0029】

以上のように構成されるナビゲーションシステムには、例えば、ユーザが目的地としての地名を発話すると、その音声は、入力装置4が内蔵する音声認識装置で音声認識され、その音声認識結果が、システム制御部1に供給される。システム制御部1は、目的地としての地名の音声認識結果を受信すると、位置測定装置2の出力から現在地を認識し、さらに、データベース装置3に記憶された地図情報を参照することで、現在地から目的地までのルートを求める。そして、システム制御部1は、そのルートを表した地図情報を、出力装置5に供給して表示させるとともに、そのルートを案内する合成音を出力装置5が内蔵する音声合成装置に生成させて出力させる。

【0030】

50

従って、この場合、ユーザは、容易に、目的地にたどりつくことができる。

【0031】

次に、図2は、図1の入力装置4が内蔵する音声認識装置の第1の実施の形態の構成例を示している。

【0032】

入力部10は、マイク11、CCDカメラ12、マイク13、センサ14、アンプ15、A/D変換器16、アンプ17、A/D変換器18、および19から構成され、そこからは、運転者であるユーザの音声を認識するために用いられる各種の入力データが出力される。

【0033】

即ち、マイク11は、例えば、指向性を有するもので、運転者であるユーザの方向に向けて設置されている。従って、マイク11には、主としてユーザの音声が入力される。マイク11に入力された音声は、そこで電気信号としての音声信号に変換され、アンプ15を介することにより増幅されて、A/D変換器18に供給される。A/D変換器18では、アンプ15を介して供給されるアナログ信号である音声信号が、所定のサンプリングクロックにしたがってサンプリングされ、さらに、所定の量子化ステップで量子化されることにより、デジタル信号である音声データとされる。A/D変換器18が出力する音声データは、前処理部20に供給される。

【0034】

CCDカメラ12は、ユーザの口の部分を撮影するように設置されており、このCCDカメラ12で撮像されたユーザの口の部分は、電気信号としての画像信号とされ、A/D変換器16に供給される。A/D変換器16では、A/D変換器18における場合と同様に、アナログ信号である画像信号が、デジタル信号である画像データとされ、前処理部21に供給される。

【0035】

マイク13は、例えば、無指向性のもので、主としてユーザの音声以外の音が入力される。即ち、例えば、エンジン音や、自動車に装備されたラジオ受信機やCDプレーヤなどから出力される音、エアコンディショナの音、さらに、窓が開いている場合には、外部の雑音などが入力される。マイク13に入力された音は、アンプ17またはA/D変換器19において、上述のアンプ15またはA/D変換器18における場合とそれぞれ同様に処理され、これにより音声データとされて、前処理部22に供給される。

【0036】

センサ14は、自動車の振動を測定するための加速度センサや、湿度や温度を測定するためのセンサなどで、その出力は、前処理部23に供給される。なお、加速度センサの出力からは、例えば、自動車が振動することにより生じる騒音の程度(大きさ)がわかる。また、湿度や温度を測定するためのセンサの出力からは、例えば、雨が降っているかどうか分かり、さらに、雨が降っていれば、その雨の音の程度がわかる。

【0037】

前処理部20乃至23では、それぞれに供給されるデータが分析され、その特徴を表す特徴パラメータが求められる。

【0038】

即ち、前処理部20および22では、音声データから、例えば、そのゼロクロス数や、所定の周波数帯域ごとのパワー、線形予測係数、ケプストラム係数などが、所定の時間としての音声フレーム単位で求められ、特徴パラメータとして、クラス分類用データ構成部24および統合パラメータ構成部26に出力される。

【0039】

前処理部21では、ユーザの口の部分の画像データから、例えば、図3に示すように、ユーザの口の横の長さ L_1 および縦の長さ L_2 が求められ、その比 L_1/L_2 が、特徴パラメータとして、クラス分類用データ構成部24および統合パラメータ構成部26に出力される。あるいは、また、前処理部21では、ユーザの口の部分の画像データから動きベクトル

10

20

30

40

50

や、エッジ、DCT（離散コサイン変換）係数などが求められ、特徴パラメータとして、クラス分類用データ構成部24および統合パラメータ構成部26に出力される。

【0040】

前処理部23では、センサ14の出力が分析されることにより、上述したように、例えば、自動車が振動することにより生じる騒音の程度（大きさ）を反映する特徴パラメータや、雨の音の程度を反映する特徴パラメータなどが求められ、やはり、クラス分類用データ構成部24および統合パラメータ構成部26に出力される。

【0041】

クラス分類用データ構成部24は、前処理部20乃至24が出力する特徴パラメータの少なくとも一部を、クラス分類に用いるクラス分類用データとして、クラス分類部25に出力する。クラス分類部25では、クラス分類用データ構成部24からのクラス分類用データを、その性質に応じて、所定のクラスに分類するクラス分類を行う。

10

【0042】

即ち、クラス分類部25は、クラス分類用データを構成する特徴パラメータの値のパターンにあらかじめ割り当てられた値を、そのクラス分類用データのクラスとして、適応判定部27に供給する。

【0043】

ここで、特徴パラメータを表すのに、例えば、Aビットが割り当てられている場合に、そのような特徴パラメータをB個集めてクラス分類用データを構成すると、そのようなクラス分類用データを構成する特徴パラメータの値のパターン数は、 $(2^A)^B$ 通りとなる。従って、AまたはBが大きいと、クラス数が莫大な数となり、その後の処理の迅速化が困難となる。

20

【0044】

そこで、クラス分類を行う前の前処理として、クラス分類用データには、それを構成する特徴パラメータのビット数を低減するための処理である、例えばADRC（Adaptive Dynamic Range Coding）処理などが施される。

【0045】

即ち、ADRC処理では、まず、クラス分類用データを構成するB個の特徴パラメータから、その値の最大のもの（以下、適宜、最大特徴パラメータという）と最小のもの（以下、適宜、最小特徴パラメータという）とが検出される。そして、最大特徴パラメータの値MAXと最小特徴パラメータの値MINとの差分DR（ $=MAX - MIN$ ）が演算され、このDRをクラス分類用データの局所的なダイナミックレンジとし、このダイナミックレンジDRに基づいて、クラス分類用データを構成する各特徴パラメータが、元の割り当てビット数Aより少ないCビットに再量子化される。つまり、クラス分類用データを構成する各特徴パラメータから最小特徴パラメータの値MINが減算され、各減算値が、 $DR / 2^C$ で除算される。

30

【0046】

その結果、クラス分類用データを構成する各特徴パラメータはCビットで表現されるようになる。従って、例えばC=1とした場合、B個の特徴パラメータののパターン数は、 $(2^1)^B$ 通りになり、ADRC処理を行わない場合に比較して、パターン数を非常に少ないものとすることができる。

40

【0047】

また、クラス分類用データを構成する特徴パラメータのパターン数を少なくする観点からは、クラス分類用データを構成する特徴パラメータの数Bは、あまり大きな値にしない方が望ましい。但し、特徴パラメータの数Bは、少なすぎても、適切なクラス分類が行われないおそれがある。従って、特徴パラメータの数Bは、これらをバランスさせて決めるのが望ましい。

【0048】

一方、統合パラメータ構成部26は、前処理部20乃至24が出力する特徴パラメータの少なくとも一部を統合して（集めて）統合パラメータとして、適応判定部27に出力する

50

。

【0049】

ここで、統合パラメータは、クラス分類用データを構成する特徴パラメータと同一のものを統合して構成しても良いし、また、クラス分類用データを構成する特徴パラメータとは異なるものを統合して構成しても良い。

【0050】

適応判定部27は、標準パラメータ記憶部28およびマッチング部29から構成されており、クラス分類部25または統合パラメータ構成部26からクラスまたは統合パラメータをそれぞれ受信すると、まずクラス分類部25から受信したクラスに対応する標準パラメータが登録された標準パラメータテーブルを、標準パラメータ記憶部28から選択する。

10

【0051】

即ち、標準パラメータ記憶部28は、後述する学習装置(図4)による学習処理によって得られる、例えば音素に対応する標準パラメータが登録されている標準パラメータテーブルを、クラスごとに分けて記憶しており、マッチング部29では、そのクラスごとの標準パラメータテーブルの中から、クラス分類部25が出力したクラスに対応するものが選択される。

【0052】

さらに、マッチング部29は、選択された標準パラメータテーブルに登録されている標準パラメータそれぞれと、統合パラメータ構成部26からの統合パラメータとの、例えば、ユークリッド距離を算出し、そのユークリッド距離を最も小さくする標準パラメータに対応する音素を、音声認識結果として出力する。

20

【0053】

従って、この音声認識装置によれば、主としてユーザの音声が入力されるマイク11から出力される音声データだけでなく、CCDカメラ12が出力する画像データから得られる、例えばユーザの口の状態や動き、マイク13が出力する音声データやセンサ14の出力から得られる、例えば雑音の種類や周波数帯域成分などをも用いて、ユーザの音声認識されるので、その認識率を向上させることができる。

【0054】

また、クラスごとに、標準パラメータテーブルを用意し、入力部10が出力する複数種類のデータのクラスに対応する標準パターンテーブルを選択するようにしたので、入力部10が出力する複数種類のデータに基づいて、ユーザの音声を認識するのに最適な標準パターンテーブルを用いることができ、その結果、認識率をさらに向上させることが可能となる。

30

【0055】

次に、図4は、図2の標準パラメータ記憶部28に記憶させる、クラスごとの標準パラメータテーブルに登録する標準パラメータを求める学習処理を行う学習装置の構成例を示している。

【0056】

入力部30(マイク31、CCDカメラ32、マイク33、センサ34、アンプ35、A/D変換器36、アンプ37、A/D変換器38、および39)、前処理部40乃至43、クラス分類用データ構成部44、クラス分類部45、または統合パラメータ構成部46は、図2の音声認識装置における入力部10(マイク11、CCDカメラ12、マイク13、センサ14、アンプ15、A/D変換器16、アンプ17、A/D変換器18、および19)、前処理部20乃至23、クラス分類用データ構成部24、クラス分類部25、または統合パラメータ構成部26とそれぞれ同様に構成されている。メモリ47は、そのアドレス端子(AD)に、クラス分類部45が出力するクラスがアドレスとして与えられるようになされており、統合パラメータ構成部46が出力する統合パラメータを記憶するようになされている。

40

【0057】

以上のように構成される学習装置では、入力部30に学習を行うための学習データが入力

50

される。即ち、マイク 3 1 には、例えば、ある話者が発話した音声が入力される。また、その時の話者の口の部分が、CCDカメラ 3 2 で撮像される。さらに、マイク 3 3 には、例えば、各種の自動車のエンジン音や、CDプレーヤーで再生された曲、雨の音、エアコンディショナの音、その他の雑音が入力される。センサ 3 4 に対しては、各種の振動や、また、マイク 3 3 に雨の音が入力されるときには、その雨が降っている状態における温度、湿度などが与えられる。

【0058】

入力部 3 0 に入力された学習データとしての各種のデータは、前処理部 4 0 乃至 4 3、クラス分類用データ構成部 4 4、クラス分類部 4 5、または統合パラメータ構成部 4 6 において、図 2 の前処理部 2 0 乃至 2 3、クラス分類用データ構成部 2 4、クラス分類部 2 5、または統合パラメータ構成部 2 6 における場合とそれぞれ同様に処理され、これにより、メモリ 4 7 には、クラス分類部 4 5 からクラスが供給されるとともに、統合パラメータ構成部 4 6 から統合パラメータが供給される。

10

【0059】

メモリ 4 7 は、クラス分類部 4 5 からのクラスに対応するアドレスに、統合パラメータ構成部 4 6 からの統合パラメータを、標準パラメータとして記憶する。

【0060】

以上の処理が、多くの話者が発話した各音素について、マイク 3 3 に入力する雑音、およびセンサ 3 4 への入力を様々なものに変えて行われる。

【0061】

その結果、メモリ 4 7 には、同一のクラスに属する統合パラメータが、各アドレスに分類されて記憶される。

20

【0062】

図 2 の標準パラメータ記憶部 2 8 には、メモリ 4 7 の各アドレスに記憶された統合パラメータ(の集合)が、標準パラメータとして、各クラスに対応する標準パラメータテーブルに登録されたものが記憶されている。

【0063】

この学習装置によれば、例えば、マイク 3 3 に雑音を入力しなかった場合と、入力した場合とでは、クラス分類部 4 5 において異なるクラスに分類され、その結果、雑音がある場合に最適な標準パラメータテーブルと、雑音がない場合に最適な標準パラメータテーブルとが作成されることになる。そして、図 2 の音声認識装置では、クラス分類部 4 5 の出力と同一のクラスがクラス分類部 2 5 から出力されるから、雑音がある場合には、その場合に最適な標準パラメータテーブルが、また、雑音がない場合にも、その場合に最適な標準パラメータテーブルが選択されることになる。

30

【0064】

なお、学習装置においては、標準パラメータは、上述のように、雑音がある場合とない場合とで異なるクラスに分類され、また、雑音の種類やレベル、さらには、話者が発話した音素の種類や、話者が男性であるか、あるいは女性であるかなどによっても異なるクラスに分類されることが予想されるが、標準パラメータが、どのように分類されるかは、特に問題とならない。これは、例えば、音声認識装置の入力部 1 0 への入力が、学習装置の入力部 3 0 と同一であれば、音声認識装置のクラス分類部 2 5 におけるクラス分類結果は、学習装置のクラス分類部 4 5 におけるクラス分類結果と同一になり、従って、音声認識装置では、入力部 1 0 への入力と同一の入力から得られた標準パラメータ、即ち、その入力に対して最適な標準パラメータを用いて音声認識が行われるからである。

40

【0065】

また、図 4 の学習装置において、メモリ 4 7 には、各クラスごとに、ある音素についての統合パラメータが複数記憶される場合がある。即ち、学習においては、1 の話者に、様々な雑音環境下で、同一の音素を発話してもらい、さらに、このような発話を、多数の話者に行ってもらうため、あるクラスのある音素について、パラメータ空間のある程度の範囲に散らばる複数の統合パラメータが得られる場合がある。

50

【0066】

具体的には、例えば、図5(左側)は、簡単のために、統合パラメータが3つの成分 P_1 、 P_2 、 P_3 を有するものとして、3次元のパラメータ空間を表しているが、このようなパラメータ空間に、あるクラスに対して得られた音素「ア」や「イ」についての統合パラメータをプロットすると、その点は、ある程度の範囲に散らばる。

【0067】

そこで、このような、ある程度の範囲に散らばるすべての点を、その音素についての標準パラメータとしても良いが、その他、例えば、図5(右側)に示すように、その範囲の重心などを求め、これを、その音素の標準パラメータとすることなども可能である。

【0068】

次に、図6は、図1の入力装置4が内蔵する音声認識装置の第2の実施の形態の構成例を示している。なお、図中、図2における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。即ち、この音声認識装置は、標準パラメータ記憶部28に代えて、標準パラメータ記憶部28₁乃至28_Mが設けられ、さらに、クラス分類用データ構成部51およびクラス分類部52が新たに設けられている他は、図2の音声認識装置と基本的に同様に構成されている。

【0069】

クラス分類用データ構成部51は、入力部10が出力する複数種類のデータから、クラス分類用データを構成し、クラス分類部52に出力する。クラス分類部52は、クラス分類用データ構成部51が出力するクラス分類用データをクラス分類し、そのクラス分類結果としてのクラスを、前処理部20乃至23に供給する。

【0070】

この場合、前処理部20乃至23では、クラス分類部52が出力するクラスに対応した前処理が行われる。即ち、例えば、マイク11に入力された音声は母音などの有声音である場合は、ゼロクロス数などよりも、線形予測係数やケプストラム係数などの方が、また、マイク11に入力された音声が無声音である場合は、線形予測係数やケプストラム係数などよりも、ゼロクロス数や周波数帯域ごとのパワー、無音の期間などの方が、音声の特徴をとらえ易い。さらに、マイク13に入力される雑音のレベルが小さい場合は、その雑音による影響は考慮する必要がないが、そのレベルが大きい場合には、雑音による影響を考慮して音声認識を行うのが望ましい。また、ユーザの口の動きがほとんどない場合は、その動きベクトルは考慮する必要がなく、逆に、動きがある場合には、その動きベクトルを考慮して音声認識を行うのが望ましい。さらに、自動車の振動がほとんどない場合や、雨が降っていない場合などには、センサ14の出力を考慮する必要はないが、その逆の場合は、センサ14の出力を考慮して音声認識を行うのが望ましい。

【0071】

つまり、認識の対象である音声は勿論、その他の要因に関しても、音声を認識するにあたって最適な特徴パラメータ(より正答率の高い認識結果を得るための特徴パラメータ)は、場合によって変化すると考えられる。

【0072】

そこで、図6の音声認識装置では、入力部10が出力するデータからクラス分類用データを構成し、それをクラス分類することで、幾つかの場合に分け、各場合について最適な特徴パラメータが、前処理部20乃至23で求められる。

【0073】

従って、図6の実施の形態では、適応判定部27(マッチング部29)において距離計算する際のパラメータ空間が、クラス分類部52のクラス分類結果によって変化することになる。即ち、適応判定部27では、クラス分類部52が出力するクラスに対応するパラメータ空間における距離が計算され、その距離に基づいて、音声認識結果が出力される。

【0074】

なお、ここでは、クラス分類部52において、クラス分類用データ構成部51が出力するクラス分類用データが、M個のクラスのうちのいずれかにクラス分類されるものとする

10

20

30

40

50

。

【 0 0 7 5 】

また、前処理部 2 0 乃至 2 3 には、クラス分類部 5 2 が出力するクラスに対応した種類の特徴パラメータを出力させる他、同一の特徴パラメータであっても、その次数を変化させたり（例えば、8 次や 1 2 次その他の線形予測係数を出力させたり）、また、特徴パラメータを出力させないようにする（例えば、自動車が静かな場所で停止している場合には、マイク 1 3 およびセンサ 1 4 の出力は考慮する必要がないと考えられるので、前処理部 2 2 および 2 3 には特に処理を行わせないようにする）ことなども可能である。

【 0 0 7 6 】

クラス分類部 5 2 が出力するクラスは、前処理部 2 0 乃至 2 3 の他、適応判定部 2 7 にも 10
供給される。適応判定部 2 7 は、上述したように、M 個の標準パラメータ記憶部 2 8₁ 乃至 2 8_M を有しており、この標準パラメータ記憶部 2 8₁ 乃至 2 8_M それぞれには、クラス分類部 5 2 が出力する M 個のクラスそれぞれに対応するパラメータ空間における標準パラメータが記憶されている。

【 0 0 7 7 】

さらに、標準パラメータ記憶部 2 8_m（但し、 $m = 1, 2, \dots, M$ ）には、図 2 の標準パラメータ記憶部 2 8 における場合と同様に、クラス分類部 2 5 が出力するクラスごとの標準パラメータテーブルが記憶されている。

【 0 0 7 8 】

なお、標準パラメータ記憶部 2 8₁ 乃至 2 8_M が記憶している標準パラメータテーブルは、 20
後述する学習装置（図 7）において学習により求められるようになされている。

【 0 0 7 9 】

適応判定部 2 7 では、クラス分類部 5 2 からクラスを受信すると、標準パラメータ記憶部 2 8₁ 乃至 2 8_M の中から、そのクラスに対応するものが選択される（標準パラメータ記憶部 2 8₁ 乃至 2 8_M の中から選択されたものを、以下、適宜、標準パラメータ記憶部 2 8 m_s と記載する）。

【 0 0 8 0 】

一方、前処理部 2 0 乃至 2 3 が出力する特徴パラメータは、クラス分類用データ構成部 2 4 を介してクラス分類部 2 5 に供給されることでクラス分類され、そのクラス分類結果としてのクラスが、適応判定部 2 7 に供給される。また、前処理部 2 0 乃至 2 3 が出力する 30
特徴パラメータは、統合パラメータ構成部 2 6 にも供給され、そこで、統合パラメータとされる。なお、統合パラメータ構成部 2 6 では、適応判定部 2 7 で選択された標準パラメータ記憶部 2 8 m_s が記憶している標準パラメータテーブルに登録されている標準パラメータと同一のパラメータ空間におけるパラメータが、統合パラメータとして構成される。

【 0 0 8 1 】

統合パラメータ構成部 2 6 で得られた統合パラメータは、適応判定部 2 7 に供給される。適応判定部 2 7 では、標準パラメータ記憶部 2 8 m_s が記憶している標準パラメータテーブルから、クラス分類部 2 5 が出力したクラスに対応するものが選択され、マッチング部 2 9 において、その標準パラメータテーブルに登録されている標準パラメータそれぞれと、統合パラメータ構成部 2 6 からの統合パラメータとの距離が算出される。そして、その 40
距離を最も小さくする標準パラメータに対応する音素が、音声認識結果として出力される。

。

【 0 0 8 2 】

以上のように、クラス分類部 5 2 が出力するクラスに対応した前処理を行い、各場合において最適な特徴パラメータを得るようにしたので、各場合に最適なパラメータ空間における距離に基づいて、より精度の高い音声認識結果を得ることが可能となる。

【 0 0 8 3 】

次に、図 7 は、図 6 の標準パラメータ記憶部 2 8₁ 乃至 2 8_M に記憶させる、クラスごとの標準パラメータテーブルに登録する標準パラメータを求める学習処理を行う学習装置の構成例を示している。

【 0 0 8 4 】

なお、図中、図 4 における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。即ち、この学習装置は、メモリ 4 7 に代えて、メモリ 4 7₁乃至 4 7_Mが設けられ、さらに、クラス分類用データ構成部 6 1、クラス分類部 6 2、およびセクタ 6 3 が新たに設けられている他は、図 4 の学習装置と基本的に同様に構成されている。

【 0 0 8 5 】

クラス分類用データ構成部 6 1 またはクラス分類部 6 2 では、図 6 のクラス分類用データ構成部 5 1 またはクラス分類部 5 2 における場合とそれぞれ同様の処理が行われ、その結果得られるクラスが、前処理部 4 0 乃至 4 3 およびセクタ 6 3 に供給される。前処理部 4 0 乃至 4 3 では、図 6 の前処理部 2 0 乃至 2 3 における場合と同様に、クラス分類部 6 2 が出力するクラスに対応した前処理が行われ、これにより、いまの場合に最適な特徴パラメータが出力される。

10

【 0 0 8 6 】

以下、クラス分類用データ構成部 4 4、クラス分類部 4 5、または統合パラメータ構成部 4 6 において、図 6 のクラス分類用データ構成部 2 4、クラス分類部 2 5、または統合パラメータ構成部 2 6 における場合とそれぞれ同様の処理が行われ、その結果、クラス分類部 4 5 からはクラスが、統合パラメータ構成部 4 6 からは統合パラメータが、それぞれ出力される。

【 0 0 8 7 】

一方、セクタ 6 3 では、クラス分類部 6 2 が出力するクラスにしたがって、メモリ 4 7₁乃至 4 7_Mのうちのいずれかのチップセレクト (CS) 端子に選択信号が出力される。即ち、これにより、メモリ 4 7₁乃至 4 7_Mのうち、クラス分類部 6 2 が出力するクラスに対応するものが選択される。

20

【 0 0 8 8 】

また、クラス分類部 4 5 が出力するクラスは、メモリ 4 7₁乃至 4 7_Mのアドレス (AD) 端子に供給されており、これにより、統合パラメータ 4 6 が出力する統合パラメータは、メモリ 4 7₁乃至 4 7_Mのうちの、クラス分類部 6 2 が出力するクラスに対応するものの、クラス分類部 4 5 が出力するクラスに対応するアドレスに記憶される。

【 0 0 8 9 】

図 6 の標準パラメータ記憶部 2 8₁乃至 2 8_Mそれぞれには、以上のようにしてメモリ 4 7₁乃至 4 7_Mそれぞれの各アドレスに記憶された統合パラメータ (の集合) が、標準パラメータとして、各クラス (クラス分類部 2 5 が出力するクラス) に対応する標準パラメータテーブルに登録されたものが記憶されている。

30

【 0 0 9 0 】

なお、この場合においても、図 6 の標準パラメータ記憶部 2 8₁乃至 2 8_Mそれぞれには、図 5 で説明したように、ある範囲にちらばる統合パラメータの重心などを標準パラメータとして記憶させるようにすることが可能である。

【 0 0 9 1 】

次に、図 8 は、図 1 の入力装置 4 が内蔵する音声認識装置の第 3 の実施の形態の構成例を示している。なお、図中、図 6 における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。即ち、この音声認識装置は、前処理部 7 1 乃至 7 4 が新たに設けられ、前処理部 2 0 乃至 2 3 に代えて、前処理部 7 1 乃至 7 4 の出力が統合パラメータ構成部 2 6 に供給される他は、図 6 の音声認識装置と同様に構成されている。

40

【 0 0 9 2 】

前処理部 7 1 乃至 7 4 には、前処理部 2 0 乃至 2 3 に対する入力と同一の入力が与えられるようになされている。さらに、前処理部 7 1 乃至 7 4 には、クラス分類部 5 2 が出力するクラスも供給されるようになされている。

【 0 0 9 3 】

50

前処理部 7 1 乃至 7 4 では、クラス分類部 5 2 が出力するクラスに対応した前処理が行われ、これにより、いまの場合に最適な特徴パラメータが、統合パラメータ構成部 2 6 に出力される。但し、ここでは、前処理部 7 1 乃至 7 4 それぞれと、前処理部 2 0 乃至 2 3 それぞれとは、基本的に異なる前処理が行われるようになされている。即ち、ここでは、前処理部 2 0 乃至 2 3 の出力からは、クラス分類部 2 5 が出力するクラスが、最終的に得られるのに対し、前処理部 7 1 の出力からは、統合パラメータ構成部 2 6 が出力する統合パラメータが得られる。そこで、前処理部 2 0 乃至 2 3 では、クラス分類部 2 5 においてクラス分類を行うのに最適な特徴パラメータが、クラス分類部 5 2 が出力するクラスに対応して求められ、他方、前処理部 7 1 乃至 7 4 では、音声認識に用いるのに最適な統合パラメータを構成するのに最適な特徴パラメータが、クラス分類部 5 2 が出力するクラスに対応して求められる。

10

【 0 0 9 4 】

次に、図 9 は、図 8 の標準パラメータ記憶部 2 8₁ 乃至 2 8_M に記憶させる、クラスごとの標準パラメータテーブルに登録する標準パラメータを求める学習処理を行う学習装置の構成例を示している。

【 0 0 9 5 】

なお、図中、図 7 における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。即ち、この学習装置は、前処理部 8 1 乃至 8 4 が新たに設けられ、前処理部 4 0 乃至 4 3 に代えて、前処理部 8 1 乃至 8 4 の出力が統合パラメータ構成部 4 6 に供給される他は、図 7 の学習装置と基本的に同様に構成されている。

20

【 0 0 9 6 】

前処理部 4 0 乃至 4 3 では、図 8 の前処理部 2 0 乃至 2 3 における場合と同様に、クラス分類部 4 5 においてクラス分類を行うのに最適な特徴パラメータが、クラス分類部 6 2 が出力するクラスに対応して求められ、また、前処理部 8 1 乃至 8 4 では、図 8 の前処理部 7 1 乃至 7 4 における場合と同様に、音声認識に用いるのに最適な統合パラメータを構成するのに最適な特徴パラメータが、クラス分類部 6 2 が出力するクラスに対応して求められる。

【 0 0 9 7 】

なお、図 9 の学習装置による学習により得られる統合パラメータが、図 8 の標準パラメータ記憶部 2 8₁ 乃至 2 8_M に記憶されるが、この場合も、学習により得られた統合パラメータをすべて記憶させるのではなく、図 5 で説明したように、ある範囲にちらばる統合パラメータの重心などを標準パラメータとして記憶させるようにすることが可能である。

30

【 0 0 9 8 】

以上、本発明を適用した音声認識装置について説明したが、本発明は、音声以外の、例えば、画像や文字、人物その他を対象とした認識装置にも適用可能である。

【 0 0 9 9 】

なお、本実施の形態では、認識対象であるユーザの音声の他に、CCDカメラ 3 2、マイク 3 3、およびセンサ 3 4 の出力を用いて認識を行うようにしたが、この認識に用いるデータは、これらに限定されるものではない。

【 0 1 0 0 】

また、図 2 の実施の形態では、前処理部 2 0 乃至 2 3 の出力から、クラス分類用データを構成し、クラス分類部 2 5 に、そのクラス分類用データをクラス分類させるようにしたが、クラス分類部 2 5 には、その他、例えば、入力部 1 0 の出力から直接クラス分類用データを構成し、これをクラス分類させるようにすることなども可能である。

40

【 0 1 0 1 】

さらに、図 6 乃至図 9 の実施の形態においては、前処理部 2 0 乃至 2 3、4 0 乃至 4 3、7 1 乃至 7 3、8 1 乃至 8 3 に、クラスを供給し、そのクラスに対応した前処理を行わせるようにしたが、前処理部 2 0 乃至 2 3、4 0 乃至 4 3、7 1 乃至 7 3、8 1 乃至 8 3 には、クラスに対応した関数を供給するようにし、この関数にしたがった演算を行わせることで、そのクラスに対応した前処理を行わせるようにすることも可能である。

50

【 0 1 0 2 】

また、本実施の形態では、説明を簡単にするため、マッチング部 29 において、あるパラメータ空間における、統合パラメータと標準パラメータとの距離に基づいて、音声認識結果を得るようにしたが、マッチング部 29 には、その他、時系列に得られる統合パラメータと標準パラメータとの系列の間の距離や、そのような系列が観測される確率を算出させ、これに基づいて音声認識結果を求めさせるようにすることも可能である。さらに、マッチング部 29 には、クラス分類部 25 や 52 が出力するクラスごとに、異なる音声認識アルゴリズムを用いて、音声認識を行わせるようにすることも可能である。

【 0 1 0 3 】

さらに、図 2、図 6、および図 8 で説明した音声認識装置、並びに図 4、図 7、および図 9 で説明した学習装置は、ハードウェアで実現することは勿論、CPU やメモリなどを有するマイクロプロセッサのアプリケーションとしてソフトウェアで実現することも可能である。

10

【 0 1 0 4 】

【発明の効果】

本発明の認識装置および認識方法によれば、各場合ごとに適切なテーブルが用いられるので、認識性能を向上させることが可能となる。

【 0 1 0 5 】

本発明の学習装置および学習方法によれば、各場合ごとに、認識に用いるのに最適なパラメータを得ることが可能となる。

20

【図面の簡単な説明】

【図 1】本発明を適用したナビゲーションシステムの構成例を示すブロック図である。

【図 2】本発明を適用した音声認識装置の第 1 の実施の形態の構成例を示すブロック図である。

【図 3】前処理部 21 の処理を説明するための図である。

【図 4】本発明を適用した学習装置の第 1 の実施の形態の構成例を示すブロック図である。

【図 5】パラメータ空間を示す図である。

【図 6】本発明を適用した音声認識装置の第 2 の実施の形態の構成例を示すブロック図である。

30

【図 7】本発明を適用した学習装置の第 2 の実施の形態の構成例を示すブロック図である。

【図 8】本発明を適用した音声認識装置の第 3 の実施の形態の構成例を示すブロック図である。

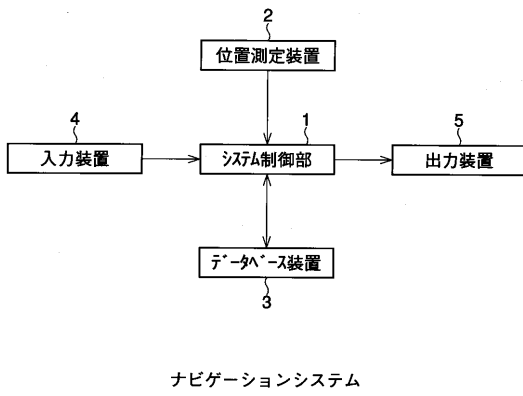
【図 9】本発明を適用した学習装置の第 3 の実施の形態の構成例を示すブロック図である。

【符号の説明】

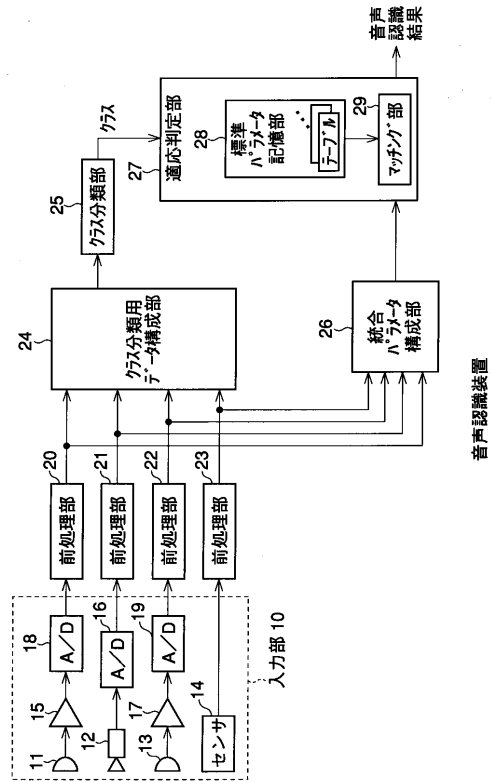
1 システム制御部, 2 位置測定装置, 3 データベース装置, 4 入力装置, 5 出力装置, 10 入力部, 11 マイク, 12 CCDカメラ, 13 マイク, 14 センサ, 15 アンプ, 16 A/D変換器, 17 アンプ, 18, 19 A/D変換器, 20乃至23 前処理部, 24 クラス分類用データ構成部, 25 クラス分類部, 26 統合パラメータ構成部, 27 適応判定部, 28, 28₁乃至28_n 標準パラメータ記憶部, 29 マッチング部, 30 入力部, 31 マイク, 32 CCDカメラ, 33 マイク, 34 センサ, 35 アンプ, 36 A/D変換器, 37 アンプ, 38, 39 A/D変換器, 40乃至43 前処理部, 44 クラス分類用データ構成部, 45 クラス分類部, 46 統合パラメータ構成部, 47, 47₁乃至47_n メモリ, 51 クラス分類用データ構成部, 52 クラス分類部, 61 クラス分類用データ構成部, 62 クラス分類部, 63 セレクタ, 71乃至74, 81乃至84 前処理部

40

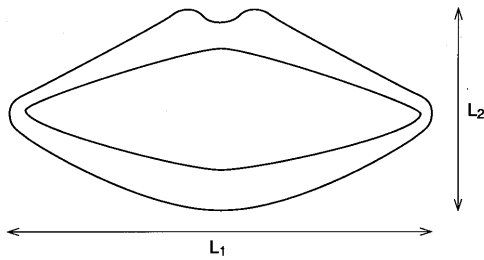
【 図 1 】



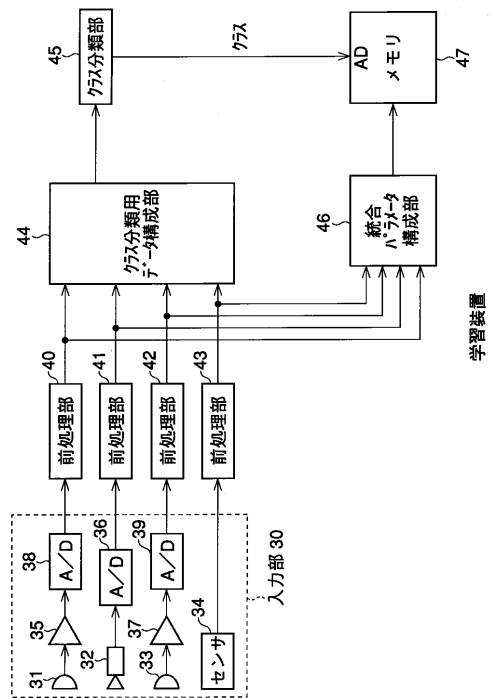
【 図 2 】



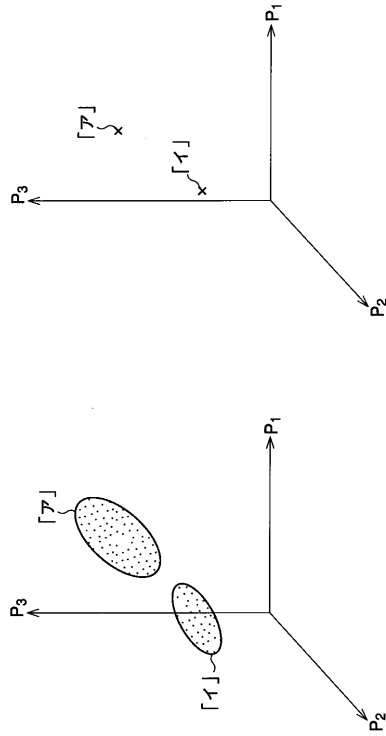
【 図 3 】



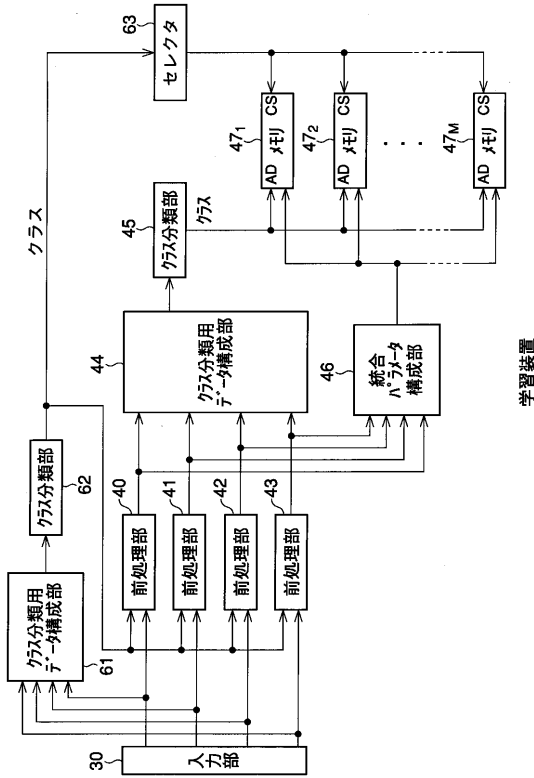
【 図 4 】



【図5】

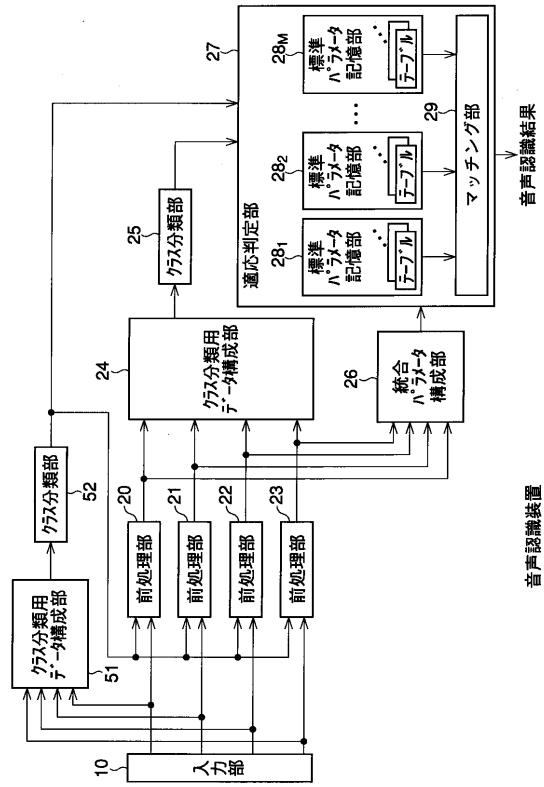


【図7】



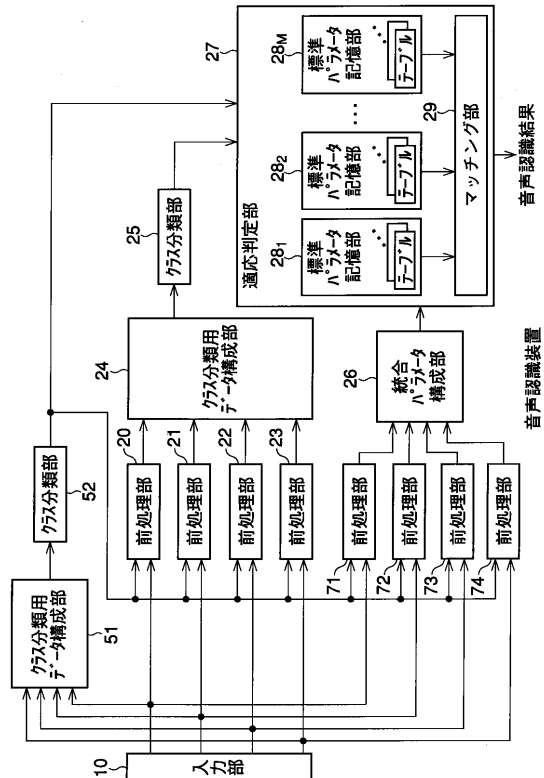
学習装置

【図6】



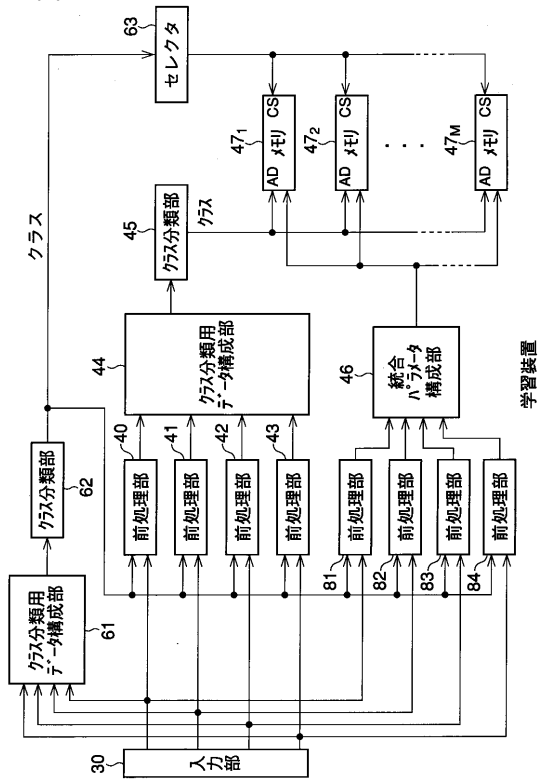
音声認識装置

【図8】



音声認識装置

【図 9】



フロントページの続き

- (56)参考文献 特開平09 - 134193 (JP, A)
特開平02 - 262199 (JP, A)
特開平03 - 129400 (JP, A)
特開平07 - 306692 (JP, A)
特開平08 - 009254 (JP, A)
特開昭62 - 239231 (JP, A)
特開昭53 - 142807 (JP, A)

- (58)調査した分野(Int.Cl.⁷, DB名)

G10L 15/00-15/28