(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4199927号
(P4199927)
(45)発行日　平成20年12月24日(2008.12.24)　(24)登録日　平成20年10月10日(2008.10.10)

(51)Int.Cl.　　F I
G10L 15/10 (2006.01)　　G10L 15/10 2OOW
G10L 15/08 (2006.01)　　G10L 15/08 3OOZ

請求項の数 8　(全 9 頁)

(21)出願番号　特願2000-550095(P2000-550095)
(86)(22)出願日　平成11年5月3日(1999.5.3)
(65)公表番号　特表2002-516419(P2002-516419A)
(43)公表日　平成14年6月4日(2002.6.4)
(86)国際出願番号　PCT/DE1999/001322
(87)国際公開番号　WO1999/060560
(87)国際公開日　平成11年11月25日(1999.11.25)
審査請求日　平成12年11月14日(2000.11.14)
審判番号　不服2005-22480(P2005-22480/J1)
審判請求日　平成17年11月22日(2005.11.22)
(31)優先権主張番号　198 21 900.8
(32)優先日　平成10年5月15日(1998.5.15)
(33)優先権主張国　ドイツ(DE)

(73)特許権者　390039413
シーメンス　アクチエンゲゼルシヤフト
Ｓｉｅｍｅｎｓ　Ａｋｔｉｅｎｇｅｓｅｌｌｓｃｈａｆｔ
ドイツ連邦共和国　Ｄ－８０３３３　ミュンヘン　ヴィッテルスバッハープラッツ　２
Ｗｉｔｔｅｌｓｂａｃｈｅｒｐｌａｔｚ　２，　Ｄ－８０３３３　Ｍｕｅｎｃｈｅｎ，　Ｇｅｒｍａｎｙ
(74)代理人　100061815
弁理士　矢野　敏雄
(74)代理人　100094798
弁理士　山崎　利臣

最終頁に続く

(54)【発明の名称】発声言語における少なくとも１つのキーワードを計算器により認識する方法および認識装置

(57)【特許請求の範囲】
【請求項１】
発声言語における少なくとも１つのキーワードを計算器により識別するための方法において、
ａ）キーワードを複数の代表子の形態で記憶し、
ｂ）キーワードをセグメントに分割し、このとき各セグメントに種々異なる前記代表子の基準特徴の集合を配属し、
ｃ）発声言語中のテストパターンをセグメントに分割し、このときテストパターンの各セグメントに、当該セグメントに類似する基準特徴を前記基準特徴の集合から、キーワードの相応するセグメントに対して配属し、
ｄ）キーワードの基準特徴とテストパターンとのセグメントごとの対応関係を累積した類似性尺度が所定の閾値以下にあることが検出された場合、当該キーワードとしてテストパターンを識別する、ことを特徴とする方法。
【請求項２】
テストパターンは閉じた形態素、とりわけ単語である、請求項１記載の方法。
【請求項３】
キーワードに対するセグメントの数とテストパターンに対するセグメントの数はそれぞれ同じである、請求項１または２記載の方法。
【請求項４】
テストパターンを複数のキーワードと比較し、もっとも類似するキーワードを出力する

、請求項１から３までのいずれか１項記載の方法。
【請求項５】
音声をキーワード並びにテストパターンに対して記憶するために特徴ベクトルを使用し、
所定のサンプリング時点で音声をデジタル化し、各特徴ベクトルを当該音声を表すデータと共に記憶する、請求項１から４までのいずれか１項記載の方法。
【請求項６】
各セグメントに対して、当該セグメントの全ての特徴ベクトルにわたって平均した特徴ベクトルを記憶し、当該セグメントを表すものとしてさらに使用する、請求項５記載の方法。
【請求項７】
所定の閾値は、常にキーワードが識別されるように決められている、請求項１から６までのいずれか１項記載の方法。
【請求項８】
発声言語中の少なくとも１つのキーワードを識別する装置であって、プロセッサユニットが設けられており、該プロセッサユニットは次のように構成されている、すなわち、
ａ）キーワードが複数の代表子の形態で記憶可能であり、
ｂ）キーワードがセグメントに分割可能であり、ここで各セグメントには種々異なる前記代表子の基準特徴の集合が配属可能であり、
ｃ）発声言語中のテストパターンがセグメントに分割可能であり、ここでテストパターンの各セグメントには該セグメントに類似する、前記基準特徴の集合からの基準特徴が、当該キーワードの相応のセグメントに対して配属可能であり、
ｃ）キーワードの基準特徴とテストパターンとのセグメントごとの対応関係を累積した類似性尺度が所定の閾値以下にある場合、テストパターンがキーワードとして識別可能である、ように構成されていることを特徴とする識別装置。
【発明の詳細な説明】
【０００１】
本発明は、発声言語における少なくとも１つのキーワードの計算器による認識方法および認識装置に関する。
【０００２】
発声言語の認識方法および認識装置は、［１］から公知である。そこには音声認識装置ないし音声認識方法の構成要素への基礎的導入部、並びに音声認識で重要な通常技術が記載されている。
【０００３】
キーワードとは、発声言語において音声認識装置により識別すべき所定の単語である。このようなキーワードとは少なくとも所定のアクションが結合しており、キーワードの識別後にこのアクションが実行される。
【０００４】
［２］にも、発声言語の認識装置が記載されている。ここでは隠れマルコフモデルを用いたモデル化により話者速度の変動への適合が可能となり、認識の際に所定の言語構成要素を、発声された言語にダイナミックに適合することができる。この適合はとりわけ、時間軸の圧縮ないし伸張を実行することにより行われる。これは例えばビタビアルゴリズムにより保証されたダイナミック適合（ダイナミックプログラミング：とも）に相当する。
【０００５】
音韻ないし音列間の間隔（距離）は例えば、言語の音韻をデジタル形態で表す特徴ベクトル間の（多次元）間隔の検出により求められる。この間隔は、音韻または音列間の類似性尺度に対する例である。
【０００６】
本発明の課題は、キーワードを識別するための方法および装置を、とりわけ認識がノイズに対して非常に頑強であり、不感であるように構成することである。

【０００７】
この課題は、独立請求項の構成によって解決される。
【０００８】
発声言語における少なくとも１つのキーワードを計算器により識別するための方法が提供される。ここではキーワードがセグメントに分割され、各セグメントに基準特徴の集合が配属される。発声言語に含まれるテストパターンがセグメントに分割され、ここでテストパターンの各セグメントには、このセグメントに類似する基準特徴が、基準特徴の集合からキーワードの相応のセグメントに対して配属される。テストパターンは、キーワードの基準特徴とテストパターンとの累積されたセグメントごとの対応関係に対する類似性尺度が所定の閾値より下にある場合、キーワードとして識別される。類似性尺度が所定の閾値より下になければ、テストパターンはキーワードとして識別されない。ここでは類似性尺度の低いことが、キーワードの基準特徴とテストパターンとの良好な一致を示す。
【０００９】
以下簡単に種々の概念およびその意味について述べる：
テストパターンは、発声言語に含まれるパターンであり、キーワードと比較され、場合によりキーワードとして識別される。類似性尺度は、テストパターンとキーワードとの一致程度、ないしはテストパターンの一部とキーワードの一部との一致程度を表す。セグメントは、テストパターンないしキーワードの区間であり、所定の持続時間を有する。基準特徴は、セグメントに関連するキーワードの部分特徴である。基準パターンは、キーワードの表現形態を表す基準特徴を含む。ワードクラスは、基準特徴の種々の合成により発生し得る全ての基準パターンを含み、ここではとりわけ複数の基準特徴がセグメントごとにキーワードに対して記憶される。トレーニングフェーズでは、それぞれのキーワードの基準特徴に対する代表子が検出され、記憶される。一方、認識フェーズでは、テストパターンとキーワードの可能な基準パターンとの比較が実行される。
【００１０】
トレーニングフェーズでは、有利には所定量Ｍの代表子が基準特徴に対して記憶される。基準特徴以上の空きスペースＭがあれば、例えば滑らかな平均値の形態で基準特徴の平均を取ることができ、これにより付加的基準特徴の情報を代表子で考慮することができる。
【００１１】
本発明の改善形態では、テストパターン（および／またはキーワード）は閉じた形態素、とりわけ単語である。テストパターンおよび／またはキーワードはそれぞれ１つの音素、ディフォン、その他の複数の音素から合成された音韻または単語の集合とすることができる。
【００１２】
別の改善形態では、キーワードに対するセグメントの数とテストパターンに対するセグメントの数は同じである。
【００１３】
付加的改善形態の枠内では、テストパターンが複数のキーワードと比較され、テストパターンに類似のキーワードが出力される。これは個別語の識別システムに相当する。ここでは複数のキーワードが発声言語中の識別すべき個別語である。発声言語に含まれるテストパターンにそれぞれもっとも適合するキーワードが出力される。
【００１４】
改善形態では、キーワード並びにテストパターンを記憶するのに特徴ベクトルが使用され、ここでは所定のサンプリング時点で音声がデジタル化され、各特徴ベクトルが音声を表すデータと共に記憶される。音声信号のこのデジタル化は、前処理の枠内で行われる。有利には１０ｍｓ毎に特徴ベクトルが音声信号から検出される。
【００１５】
別の改善形態では、各セグメントに対し、このセグメントの全ての特徴ベクトルについて平均された特徴ベクトルが記憶され、このセグメントを表すものとしてさらに使用される。例えば各１０ｍｓ毎に発生するデジタル化された音声データは、有利にはオーバラップ

した時間窓で２５ｍｓの時間的伸張によりさらに処理される。このためにＬＰＣ分析、スペクトル分析、またはCepstral分析を使用することができる。それぞれの分析結果として、各１０ｍｓ区間毎に、ｎ個の係数を備えた特徴ベクトルが得られる。セグメントの特徴ベクトルは有利には平均化され、これによりセグメントごとに１つの特徴ベクトルが得られる。キーワード識別のためのトレーニングの枠内で、セグメントごとに種々異なるソースから、発生された言語に対して複数の種々異なる基準特徴が記憶される。これにより複数の平均化された基準特徴（キーワードに対する特徴ベクトル）が得られる。

【００１６】

さらに、発生された言語中の少なくとも１つのキーワードを識別するための装置が記載されており、この装置はプロセッサユニットを有し、プロセッサユニットは次のステップを実行するように構成されている：

－キーワードをセグメントに分割するステップ；ここで各セグメントには基準特徴の集合を割り当てることができる。

【００１７】

－発声言語中のテストパターンをセグメントに分割するステップ；ここでテストパターンの各セグメントに、このテストパターンに類似の基準特徴をキーワードの相応するセグメントに対する基準特徴の集合から割り当てることができる。

【００１８】

－キーワードの基準特徴とテストパターンとの累積されたセグメントごとの対応関係に対する類似性尺度が所定の閾値より下にある場合、テストパターンをキーワードとして識別するステップ；

－類似性尺度が所定の閾値より下になければ、キーワードは識別されないステップ。

【００１９】

本発明のさらなる改善形態は従属請求項から明らかである。

【００２０】

本発明の装置はとりわけ、本発明の方法または前記に説明した改善形態を実施するのに適する。

【００２１】

以下図面に基づき、本発明の実施例を詳細に説明する。

【００２２】

図１は、発生言語中の少なくとも１つのキーワードを識別するための方法ステップを概略的に示す図である。

【００２３】

図２は、キーワードを識別するための２つの可能な実施例の概略図である。

【００２４】

図３は、テストパターンのキーワードへの複写、および類似性尺度の検出を説明するための図である。

【００２５】

図４は、発声言語中のキーワードを識別するための装置の概略図である。

【００２６】

図１は、発声言語中の少なくとも１つのキーワードを識別するための方法を概略的に示す。

【００２７】

発生された言語は有利には１０ｍｓ毎に、２５ｍｓの長さのオーバラップする時間窓でデジタル化され、場合により前処理される（フィルタリング）。このためにＬＰＣ分析、スペクトル分析、またはCepstral分析が適用される。前処理の結果として各１０ｍｓ区間毎にｎ個の係数を備えた特徴ベクトルが得られる。

【００２８】

発生された言語の個々の要素、有利には単語は、単語の間の休止エネルギーまたは休止スペクトルに基づいて求められた休止によって検出される。このようにして発生された言語

内の個々の単語が識別される。

【００２９】

図１では、２つの要素、すなわちトレーニングフェース１０１と識別フェーズ１０２とが大きく区別されている。トレーニングフェーズ１０１でも識別フェーズ１０２でも、検出された単語、すなわちキーワードまたはテストパターンが所定数のセグメントに分割される。有利にはセグメントの特徴ベクトルは平均化される。平均化された特徴ベクトルを備えるセグメントの連続はワードパターンを提供する。

【００３０】

発生された言語で識別された単語（キーワードまたはテストパターン）はそれぞれ所定数のセグメントに分割される。セグメント内の複数の特徴ベクトルは平均化される。このときこの平均化された特徴ベクトルは全体で単語（ワードパターン）を表す。キーワードは後での識別のために記憶される。このときとりわけこのキーワードの複数の代表子が記憶される。複数の発声のキーワードを複数回記録し、キーワードを表すのに最適の記録をそれぞれ記憶するのが有利である。ここでこの最適の記録は、セグメントごとにそれぞれ平均化された特徴ベクトルの形態で記憶される。これにより、キーワードのそれぞれのセグメントに関連する所定数の基準特徴が得られる。このようにして記憶された基準特徴に基づいて、単語をセグメントの連続により検出されたシーケンスから種々異なる基準パターンに統合することができる。ここでは、キーワードの種々の代表子の基準特徴が１つの基準パターンに統合される。このようにして、複数の可能性が基準パターンに対してキーワードに対する元の代表子として記憶される。トレーニングフェーズ１０１に続く識別フェーズ１０２では、それぞれ次にある基準特徴（セグメントを基準にして）がテストパターンの相応のセグメントに配属される。

【００３１】

トレーニングフェーズ１０１は、キーワードの所定数のセグメントへの分割を含む（ブロック１０３参照）。ステップ１０４では、k_i基準特徴がセグメントｉ毎に記憶される。ここでｋはキーワードに対して検出された代表子の数を表す。ワードクラスはステップ１０５でセグメントの連続により規定された基準特徴の連続によって表される。記憶された全ての変形を備えたキーワードを記述するワードクラスは、基準特徴と基準パターンとの種々の結合によって表される。ここで基準パターンはワードクラスの基準パターン事例である。

【００３２】

識別フェーズ１０２では、ここではテストパターンと称される単語がキーワードに配属することができるか否かが検出される。このためにテストパターンは上の実施例に相応してステップ１０６でセグメント化される。ステップ１０７では、テストパターンのセグメントがキーワードのセグメントに複写され、ここでキーワードの各セグメントにキーワードの類似のセグメントがそれぞれ配属される。この配属は全てのセグメントに対して実行され、各セグメントごとに算出された類似性尺度が全体類似性尺度に累積される（ステップ１０８参照）。累積された類似性尺度の値が所定の閾値以下であれば、このことはテストパターンとキーワードとの類似性が高いことに相当し、テストパターンはキーワードとして識別される（ステップ１０９参照）。

【００３３】

類似性はとりわけ間隔に基づいて検出される。２つのパターンが類似であれば、これらは相互に小さい間隔を有し、特徴ベクトルの差も相応に小さい。したがってそれぞれのセグメントに対して検出された類似性尺度は特徴ベクトルの間隔に相応し、この間隔はまたセグメントごとに実行された複写の際の複写エラーに相応する。類似性尺度の累積は、セグメントごとに生じる複写エラーの加算に相応し、全体類似性尺度はしたがってテストパターンをキーワードに配属する際に生じる全体エラーに対する値である。とりわけ１つ以上のキーワードを識別すべきであるから、テストパターンは複数のキーワードに複写され、それぞれセグメントごとの類似性尺度が検出され、各キーワードへの各配属ごとに累積された類似性尺度が算出される。そして、累積された類似性尺度が複数のキーワードへの全

ての配属においてもっとも小さい値を有するキーワードが識別される。

【００３４】

図２は、キーワードの識別に対する２つの実施例を概略的に示す。図１で説明した、各キーワード毎の類似性尺度の検出（ステップ２０１参照）は、もっとも類似するキーワードの識別ないし出力につながる（ステップ２０２ないし図１の説明参照）。最良の累積類似性尺度、すなわちテストパターンのそれぞれのキーワードへの複写の際のエラーが最小であっても、これが所定の閾値より上であれば（ステップ２０３参照）、第２の実施例はワードクラスを識別せず、またワードクラスを出力しない。このような場合、配属、すなわちテストパターンのキーワードへの複写は失敗であり、テストパターンはキーワードではないことを前提とする。もっとも適合するキーワードへの強制的配属は行われない。このキーワードはそのようにしてもほとんど適切でないからである。

【００３５】

図３には、テストパターンのキーワードへの複写、および類似性尺度の検出が示されている。

【００３６】

例として、テストパターンＴＥＭの５つのセグメントＳＧｉＴ（ｉ＝１，２，...、５）が基準パターンＲＭＵの５つのセグメントＳＧｉＳに複写されるとする。識別すべきキーワード（ワードクラス）は、基準特徴ＲＭｉの種々の組合せによって、セグメントの順序に従い示されている。基準特徴は、上に説明したように、キーワードのセグメントに対する特に良好な代表子として求められる（トレーニングフェーズ）。

【００３７】

始めに、テストパターンの第１のセグメントＳＧ１Ｔがキーワードの第１のセグメントＳＧ１Ｓに配属される。ここではテストパターンＴＥＭの第１のセグメントＳＧ１Ｔを表す平均特徴ベクトルが基準特徴ＲＭ１とＲＭ２のうちの良い方に複写される。これに続き、テストパターンＴＥＭの第２セグメントＳＧ２Ｔがキーワードの次の基準特徴に複写される。ここでは３つの異なる経路Ｗ１，Ｗ２，Ｗ３が可能である。経路Ｗ１は１セグメントだけ右へ勾配０で進む。経路Ｗ２は１セグメントだけ右へ勾配１で進み、経路Ｗ３は１セグメントだけ右へ勾配２で進む。このようにして最良の類似性尺度が検出される。ここではテストパターンＴＥＭの第２セグメントＳＧ２Ｔが基準特徴ＲＭ１，ＲＭ２と（経路Ｗ１に対して）、ＲＭ３，Ｒ４，ＲＭ５と（経路Ｗ２に対して）、そしてＲＭ１、ＲＭ２と（経路Ｗ３に対して）比較され、もっとも類似するものが検出される。相応にして、第１のセグメントから第２のセグメントへの移行の際に進んだ経路に依存して、第２のセグメントから第３のセグメントへの移行の際も同じように行われる。

【００３８】

キーワードを表す基準特徴は、例えば第１セグメントＳＧ１Ｓと第３セグメントＳＧ３Ｓに対して等しい。なぜなら、それぞれのセグメントが同じ音韻を表すからである。不要なメモリスペースを浪費せず、基準特徴をそれぞれに多重に、また各キーワードで個別に記憶するために、基準特徴を含むテーブルがファイルされる（テーブル参照）。キーワードのセグメントと共に、指示子だけがテーブルのメモリ領域にファイルされる。ここで指示子は基準特徴のそれぞれのデータを参照させる。指示子に対する必要メモリスペースは（テーブル内にオフセットがあっても）、それぞれの基準特徴に所属するデータよりは格段に小さい。

【００３９】

図４は、少なくとも１つのキーワードの音声識別装置を示す。発声された言語４０１から前処理４０２に基づいて特徴ベクトル（［２］参照）が検出される。これに基づき、単語開始部／単語終了部検出４０３が実行され、識別された単語がＮ個のセグメントに分割される（ブロック４０４参照）。トレーニングフェーズ（接続路４０５によって決められる）の間、セグメントはステップ４０６でデータバンク４０７に記憶され、１セグメントに対してＭ以上の代表子が、とりわけスムーズな平均値に従い、データバンク４０７での代表子の平均値となる。このためにセグメントの代表子は接続路４０８を介して平均過程（

クラスタリングにも）に供給される。識別フェーズ（接続路４０９により示されている）では、非線形複写が、テストパターンに最適に適合するセグメントの選択によって行われる（ブロック４１０参照）。このとき各セグメントごとに代表子がデータバンク４０７から求められる。これに続いて、ブロック４１１で分類および識別された単語が出力される（ブロック４１１参照）。

【００４０】

本発明の枠内で以下の刊行物が引用される。

【００４１】

【表１】

[1] A. Hauenstein: "Optimierung von Algorithmen und Entwurf eines Prozessors für die automatische Spracherkennung", Lehrstuhl für Integrierte Schaltungen, Technische Universität München, Dissertation, 19.07.1993, Kapitel 2, Seiten 13 bis 26.

[2] N. Haberland, et. al.: "Sprachunterricht - Wie funktioniert die computerbasierte Spracherkennung?", c't 5/98, Heinz Heise Verlag, Hannover 1998, Seiten 120 bis 125.

【図面の簡単な説明】

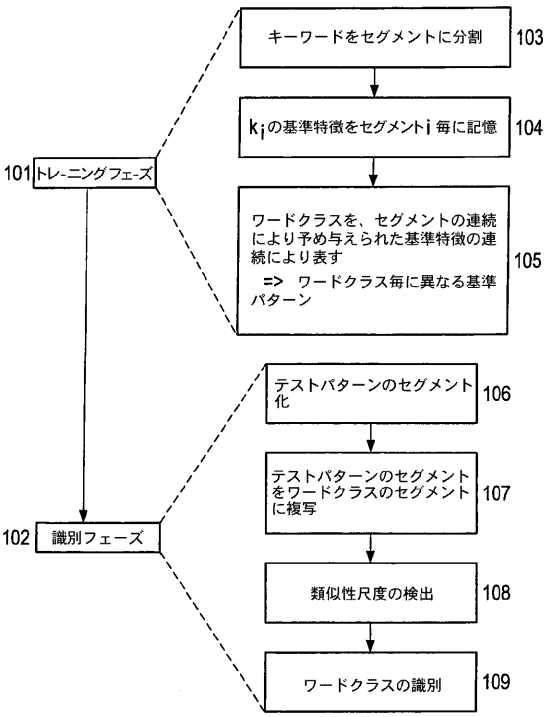
【図１】　図１は、発生言語中の少なくとも１つのキーワードを識別するための方法ステップを概略的に示す図である。

【図２】　図２は、キーワードを識別するための２つの可能な実施例の概略図である。

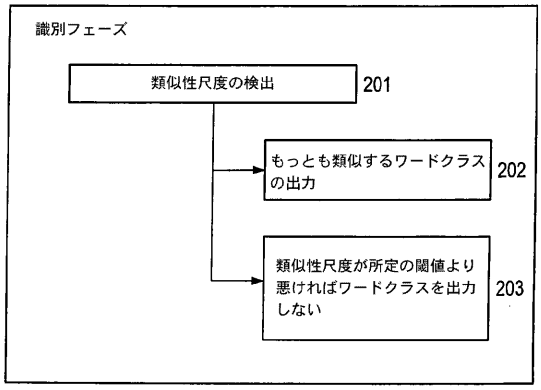
【図３】　図３は、テストパターンのキーワードへの複写、および類似性尺度の検出を説明するための図である。

【図４】　図４は、発声言語中のキーワードを識別するための装置の概略図である。

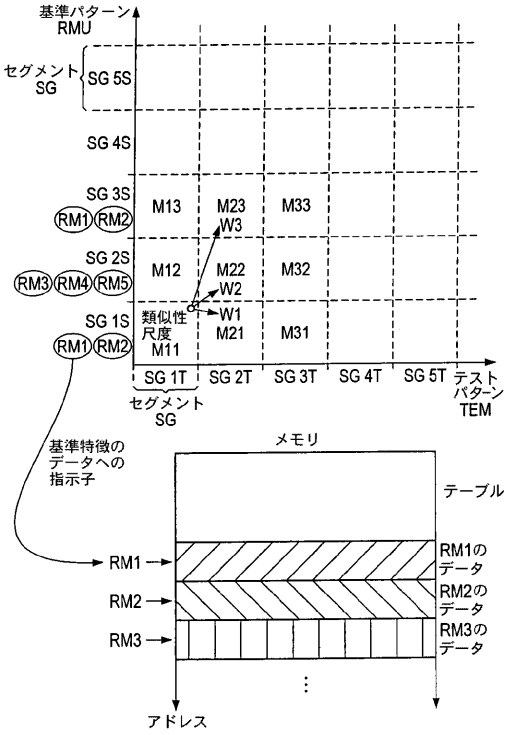
【図１】


キーワードをセグメントに分割
103
k_iの基準特徴をセグメントi毎に記憶
104
101 トレーニングフェーズ
ワードクラスを、セグメントの連続により予め与えられた基準特徴の連続により表す
=> ワードクラス毎に異なる基準パターン
105
テストパターンのセグメント化
106
テストパターンのセグメントをワードクラスのセグメントに複写
107
102 識別フェーズ
類似性尺度の検出
108
ワードクラスの識別
109

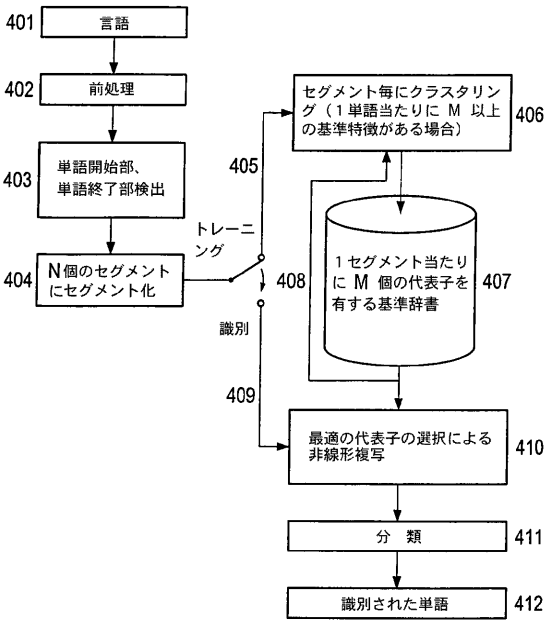
【図２】


識別フェーズ
類似性尺度の検出
201
もっとも類似するワードクラスの出力
202
類似性尺度が所定の閾値より悪ければワードクラスを出力しない
203

【図３】


基準パターン RMU
セグメント SG
SG 5S
SG 4S
SG 3S
SG 2S
SG 1S
RM1 RM2
RM3 RM4 RM5
M13 M23 M33
M12 M22 M32
M11 M21 M31
W3 W2 W1
類似性尺度
SG 1T SG 2T SG 3T SG 4T SG 5T
テストパターン TEM
セグメント SG
基準特徴のデータへの指示子
メモリ
テーブル
RM1 RM2 RM3
RM1のデータ
RM2のデータ
RM3のデータ
アドレス

【図４】


401 言語
402 前処理
403 単語開始部、単語終了部検出
404 N個のセグメントにセグメント化
トレーニング
識別
405
406 セグメント毎にクラスタリング（1単語当たりにM以上の基準特徴がある場合）
407 1セグメント当たりにM個の代表子を有する基準辞書
408
409
410 最適の代表子の選択による非線形複写
411 分類
412 識別された単語

フロントページの続き

(74)代理人 100099483
弁理士 久野 琢也
(74)代理人 100114890
弁理士 アインゼル・フェリックス＝ラインハルト
(74)代理人 230100044
弁護士 ラインハルト・アインゼル
(72)発明者 ベルンハルト ケンメラー
ドイツ連邦共和国 オットーブルン アム ビルケンガルテン ２４

合議体
審判長 原 光明
審判官 加藤 恵一
審判官 廣川 浩

(56)参考文献 特開昭６２－２７７９８（ＪＰ，Ａ）
特開昭６１－２４５１９８（ＪＰ，Ａ）
特開平５－５３５９７（ＪＰ，Ａ）
特開平７－２０８８９（ＪＰ，Ａ）
特公平７－６６２７３（ＪＰ，Ｂ２）
児島宏明，他，“区分線形セグメントラティスを用いた単語モデルの自動生成”，電子情報通信学会技術研究報告［音声］，ＳＰ９５－１０６，ｐ．９３－９８，平成7年１２月１５日発行