(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3814459号
(P3814459)
(45)発行日　平成18年8月30日(2006.8.30)　(24)登録日　平成18年6月9日(2006.6.9)

(51)Int.Cl.　F I
G10L 15/14 (2006.01)　G10L 3/00 535B

請求項の数 7 (全 9 頁)

(21)出願番号　特願2000-99536(P2000-99536)
(22)出願日　平成12年3月31日(2000.3.31)
(65)公開番号　特開2001-282283(P2001-282283A)
(43)公開日　平成13年10月12日(2001.10.12)
審査請求日　平成16年12月10日(2004.12.10)

早期審査対象出願

前置審査

(73)特許権者　000001007
キヤノン株式会社
東京都大田区下丸子３丁目３０番２号
(74)代理人　100076428
弁理士　大塚　康徳
(74)代理人　100112508
弁理士　高柳　司郎
(74)代理人　100115071
弁理士　大塚　康弘
(74)代理人　100116894
弁理士　木村　秀二
(72)発明者　山本　寛樹
東京都大田区下丸子３丁目３０番２号　キヤノン株式会社内

最終頁に続く

(54)【発明の名称】音声認識方法及び装置と記憶媒体

(57)【特許請求の範囲】
【請求項１】
入力された音声信号から特徴ベクトル系列を抽出するステップと、
Ｎ個の特徴ベクトル系列ごとにＨＭＭ間の遷移を許可し、それ以外の特徴ベクトル系列ではＨＭＭ間の遷移を許可しない探索空間を用いて、前記音声信号を音声認識するステップとを有し、
前記Ｎの値は、２≦Ｎ≦４の範囲にあることを特徴とする音声認識方法。
【請求項２】
前記探索空間は、１つ以上の認識対象語に対応するＨＭＭの状態系列と前記特徴ベクトル系列との二軸で規定される空間であることを特徴とする請求項１に記載の音声認識方法。
【請求項３】
前記ＨＭＭは、音素、音節、単語、diphoneのいずれかに対応することを特徴とする請求項１に記載の音声認識方法。
【請求項４】
請求項１乃至３のいずれか１項に記載の音声認識方法をコンピュータに実行させるためのプログラムを記憶したことを特徴とする、コンピュータにより読み取り可能な記憶媒体。
【請求項５】
入力された音声信号から特徴ベクトル系列を抽出する抽出手段と、

Ｎ個の特徴ベクトル系列ごとにＨＭＭ間の遷移を許可し、それ以外の特徴ベクトル系列ではＨＭＭ間の遷移を許可しない探索空間を用いて、前記音声信号を音声認識する音声認識手段とを有し、

前記Ｎの値は、２≦Ｎ≦４の範囲にあることを特徴とする音声認識装置。

【請求項６】

前記探索空間は、１つ以上の認識対象語に対応するＨＭＭの状態系列と前記特徴ベクトル系列との二軸で規定される空間であることを特徴とする請求項５に記載の音声認識装置。

【請求項７】

前記ＨＭＭは、音素、音節、単語、diphoneのいずれかに対応することを特徴とする請求項５に記載の音声認識装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、隠れマルコフモデルを用いた音声認識方法及びその装置と、その音声認識を実行するプログラムを記憶したコンピュータにより読み取り可能な記憶媒体に関するものである。

【０００２】

【従来の技術】

近年、音声認識の有効な方法として、隠れマルコフモデル（Hidden Markov Model、以下、ＨＭＭと記す）を用いた方法の研究、応用が進み、多くの音声認識システムなどで用いられている。

【０００３】

【発明が解決しようとする課題】

図６は、ＨＭＭを用いた従来の音声認識方法の一例を示すフローチャートである。

【０００４】

まずステップＳ１の音声入力において、マイクロホンなどから入力された音声信号をＡ／Ｄ変換してデジタル信号に変換する。次にステップＳ２では、ステップＳ１で変換した音声信号を音響分析し、特徴ベクトルの時系列を抽出する。この音響分析では、時間的に変化する連続波形である音声信号に対して、３０ミリ秒程度の窓幅の分析窓を設け、この分析窓を窓幅の１／３～１／２程度（１０ミリ秒～１５ミリ秒）ずらしながら音響分析する。各窓内の分析結果は特徴ベクトルとして出力するため、ｔをフレーム番号とすると、音声信号は特徴ベクトル系列Ｏ(t)（１≦ｔ≦Ｔ）に変換される。

【０００５】

次にステップＳ３に進み、所定の構成単位からなるＨＭＭを保持するＨＭＭデータベース５と、認識対象語とＨＭＭの状態系列との対応関係を記述した辞書６を用いて、ＨＭＭの状態系列と入力音声の特徴ベクトル系列とを二軸とする探索空間を生成し、この探索空間上で音響尤度が最大となる最適パスをビタビ探索を用いて求める。

【０００６】

この最適パス探索の詳細な手順を図７を用いて以下で述べる。

【０００７】

図７は、音素を構成単位としたＨＭＭを用いて、２つの単語『あき』『あか』を連続音声認識する場合の探索空間および探索の様子を表している。この図７において、横軸は特徴ベクトル系列の一例で、縦軸はＨＭＭの状態系列の一例である。

【０００８】

まず、ＨＭＭデータベース５と、認識対象語とＨＭＭ５の状態系列との対応関係を記述した辞書６から、１つ以上の認識対象語に対応するＨＭＭの状態系列を生成する。こうして生成されたＨＭＭの状態系列は図７の縦軸のようになる。

【０００９】

こうして生成されたＨＭＭの状態系列と特徴ベクトル系列とから二次元の格子状の探索空

間が形成される。

【００１０】

次に、図７に示した探索空間上の「START」から出発し「END」に到達する全ての経路（パス）について、各格子点における状態出力確率と各格子点間の遷移に対応するＨＭＭの状態遷移確率とから累積音響尤度が最大となる最適パスを探索する。

【００１１】

まず、探索空間の各格子点（状態仮説）について、その格子点に到達するまでの累積音響尤度（状態仮説尤度）をｔ＝１からｔ＝Ｔまで順番に求める。第ｔフレームの状態ｓの状態仮説尤度Ｈs,tは以下の式で求める。

【００１２】

$$H(s,t)=\max_{s'\in S'(s)} H(s',t-1)\times a(s',s)\times b(s,O(t)) \quad \cdots\text{式（1）}$$

ここで、Ｓ'(s)は、状態ｓに接続する状態の集合、ａ(ｓ',ｓ）は、状態ｓ'から状態ｓへの遷移確率、ｂ(ｓ,(Ｏ(t))は、特徴ベクトルＯ(t)に対する状態ｓの状態出力確率である。

【００１３】

以上で求めた状態仮説尤度を用いて、「END」に到達する最適パスの音響尤度は以下の式で求める、

$$\max_{s\in Sf} H(s,T)\times a(s,s') \quad \cdots\text{式（2）}$$

ここで、Ｓfは「END」に到達可能な音素ＨＭＭの状態の集合、即ち、各認識対象語を表すＨＭＭの最終状態の集合を表す。またａ(ｓ,s')は、状態ｓから他の状態へ遷移する確率である。

【００１４】

以上の計算の過程で、各状態仮説の状態仮説尤度を求める際に、状態仮説尤度が最大となる遷移元の状態（式（１）におけるｓ'）を記憶しておき、これを辿ることで音響尤度が最大となる最適パスが求まる。

【００１５】

以上の手順で求めた最適パスに対応するＨＭＭの状態系列を求め、さらにその状態系列に対応する認識対象語を認識結果とする。図７で太線で示したパスが累積音響尤度を最大にする最適パスであった場合、このパスは音素ＨＭＭ /a//k//a/の状態を通るため、この場合の音声認識結果は「あか」となる。最後にステップＳ４に進み、その認識結果を表示装置等に表示したり、或いは他の処理に渡したりする。

【００１６】

しかしながら、図７に示した探索空間は、認識対象語数、発声時間に比例して大きくなり、このような探索空間の拡大に伴って最適パスの探索処理の処理量が飛躍的に増加する。これにより、大語彙の音声認識を実現する場合や、処理能力が劣る計算機で音声認識を実現する場合に音声認識の応答速度が遅くなるという問題があった。

【００１７】

本発明は上記従来例に鑑みてなされたもので、音声認識のための探索処理に要する処理量を削減して高速な音声認識を可能にした音声認識方法及びその装置と記憶媒体を提供することを目的とする。

【００１８】

【課題を解決するための手段】

上記目的を達成するために本発明の音声認識方法は以下のような工程を備える。即ち、

入力された音声信号から特徴ベクトル系列を抽出するステップと、

Ｎ個の特徴ベクトル系列ごとにＨＭＭ間の遷移を許可し、それ以外の特徴ベクトル系列ではＨＭＭ間の遷移を許可しない探索空間を用いて、前記音声信号を音声認識するステップとを有し、前記Ｎの値は、２≦Ｎ≦４の範囲にあることを特徴とする。

【００１９】
上記目的を達成するために本発明の音声認識装置は以下のような構成を備える。即ち、
入力された音声信号から特徴ベクトル系列を抽出する抽出手段と、
Ｎ個の特徴ベクトル系列ごとにＨＭＭ間の遷移を許可し、それ以外の特徴ベクトル系列ではＨＭＭ間の遷移を許可しない探索空間を用いて、前記音声信号を音声認識する音声認識手段とを有し、前記Ｎの値は、２≦Ｎ≦４の範囲にあることを特徴とする。
【００２０】
【発明の実施の形態】
以下、添付図面を参照して本発明の好適な実施の形態を詳細に説明する。
【００２１】
図１は本発明の実施の形態に係る音声認識装置のハードウェア構成を示すブロック図である。
【００２２】
図１において、１０１は出力部で、例えば表示部や印刷部等を有し、音声認識の結果、或いはその音声認識の結果から得られた応答（文書データ）を出力する。１０２は入力部で、ここでは例えばマイクロフォンなどの音声を入力するための構成を備えている。またこの入力部１０２は、キーボードやマウス等のように、オペレータにより操作されて各種データを入力するための構成を備えている。１０３は中央処理部（ＣＰＵ）で、数値演算やこの音声認識装置全体の動作制御等を行なう。１０４は記憶部で、ディスク装置等の外部メモリ装置や、ＲＡＭ、ＲＯＭ等の内部メモリを含み、この記憶部には、本実施の形態の手順や処理を実行するための制御プログラム、更にはこの処理に必要な一時的データおよび認識対象語とＨＭＭの対応関係を示した辞書、ＨＭＭ等が格納されている。１０５は音声認識ユニットである。
【００２３】
以上の構成を備える音声認識ユニット１０５の動作を以下に詳しく説明する。
【００２４】
本実施の形態では、ＨＭＭの構成単位を音素とし、「あか」「あき」を認識対象語とする、ＨＭＭの状態系列を用いて、入力音声を連続音声認識を行う場合について説明する。
【００２５】
図５は、本実施の形態における認識対象語を示す図で、各認識対象語は音素ＨＭＭで構成されている。また、各音素ＨＭＭは、図４に示すように複数の状態の連結として構成されている。
【００２６】
図４では３つの状態（Ｓ1、Ｓ2、Ｓ3）が示されており、状態Ｓ1におけるループ確率はａ11、状態Ｓ1から状態Ｓ2への遷移確率がａ12で示されている。本実施の形態では、ビタビアルゴリズムを用いて、これら状態間での遷移確率等に基づいて、探索空間内の各経路スコアを求め、そのスコア値の累積値（尤度）が最も大きい経路を探索する。
【００２７】
図２は、本実施の形態に係る音声合成ユニット１０５における音声認識処理を示すフローチャートである。なお、図６のＨＭＭを用いた音声認識方法と同様の過程については、詳細な説明を省略し、ステップＳ１１の音声入力処理（ステップＳ１に対応する）、ステップＳ１２の音響分析処理（ステップＳ２に対応する）により、入力音声から特徴ベクトル系列Ｏ(t)（１≦ｔ≦Ｔ）を抽出した後の探索処理から説明を行う。
【００２８】
ステップＳ１３の探索空間生成処理において、音素を構成単位とするＨＭＭを保持するＨＭＭデータベース１６と、認識対象語とＨＭＭの状態系列との対応関係を記述した辞書１７とを用いて、１つ以上の認識対象語に対応するＨＭＭの状態系列を生成し、特徴ベクトル系列Ｏ(t)と、このＨＭＭの状態系列とからなる二軸の探索空間を生成する（図３及び図７参照）。
【００２９】

図３は、本実施の形態に係る音声認識装置における１方向の探索処理経路を説明する図である。

【００３０】

図３に示す本実施の形態と、図７の構成との相違点は、図３では、探索空間を生成する際に、特定のフレームのみにＨＭＭ間の遷移を許可するパスを用意し、それ以外のフレームではＨＭＭ間の遷移を許可しない点にある。これにより、探索空間における最適パスを探索する際の探索すべきパスの数を減少させることができ、これにより処理速度を高めることができる。ここでは、例えば、ＨＭＭ間の遷移を許すフレームを、例えばＮ（２≦Ｎ≦４）フレーム間隔というように設定する。

【００３１】

図３の例では、Ｎ＝３として、３フレームごとにＨＭＭ間での遷移を許すように探索空間が設定されている。この図３と図７とを比較すると、図７の音声認識方法で生成される探索空間に比べ，ＨＭＭ間を遷移するパスの数が大幅に削減されているのがわかる。

【００３２】

図３の例では、各音素（/a/，/k/，/i/）のＨＭＭは３つの状態を有し、それぞれ所定の遷移規則に従って他のＨＭＭに遷移する。本実施の形態では、これらＨＭＭ間での遷移をＮ（＝３）フレーム毎に許可している。即ち、特徴ベクトル系列Ｏ(2)，Ｏ(5)，Ｏ(8)，…，Ｏ(T-1)でのみＨＭＭ間での遷移が許可されている。

【００３３】

次にステップＳ１４に進み、探索処理において、図３の「START」から出発して「END」に到達する全ての経路（パス）について、各状態仮説の累積尤度における状態出力確率と各格子点間の遷移に対応するＨＭＭの状態遷移確率とから累積音響尤度を計算し、その計算した累積音響尤度が最も大きくなるような最適パスを探索する。なお、この最適パスの探索方法は、図６のステップＳ３と同様の処理で求まるので、その説明を省略する。

【００３４】

こうして求めた最適パス上のＨＭＭ系列の認識対象語を認識結果とし、ステップＳ１５の認識結果出力処理で、その認識結果を出力部１０１の表示装置に表示したり、他の処理に渡したりする。

【００３５】

以上説明したように本実施の形態によれば、特徴ベクトル系列とＨＭＭの状態系列とを用いて探索空間を生成する際に、ＨＭＭ間での遷移をＮフレーム毎にのみ許可することにより、探索するパスの数を減らして、認識処理速度をより高めることができる。

【００３６】

本実施の形態によれば、図２のステップＳ１３において、ＨＭＭ間の遷移を許可するフレームをＮ（Ｎ＝３）フレーム単位とする場合について説明したが、これに限るものではない。例えば、上述の探索空間は、認識対象語の増加や発声時間の増加に伴って拡大することを考慮し、認識対象語の増加または発声時間の増加に応じて、ＨＭＭ間の遷移を許可するフレームの間隔を２≦Ｎ≦４の範囲において段階的に広げることも可能である。また、認識対象語の増加と発声時間の増加の双方を考慮して、フレームの間隔を２≦Ｎ≦４の範囲において段階的に変更することも可能である。このように構成することにより、探索空間の規模に応じて適応的に探索パスの削減を行え、認識処理の速度を高めることができる。

【００３７】

また、本実施の形態のステップＳ１３では、ＨＭＭの状態系列内に存在する全てのＨＭＭに対して、ＨＭＭ間の遷移を許可するフレームをＮ（Ｎ＝３）フレーム単位とする場合について説明したが、これに限るものではない。例えば、ＨＭＭ間の遷移を許可するフレームの間隔を、所定のＨＭＭ間において２≦Ｎ≦４の範囲で変更することも可能である。また、所定数の特徴ベクトル系列ごとに、フレーム間隔を可変とすることも可能である。これにより、他のＨＭＭへ遷移する頻度が高いＨＭＭと、他のＨＭＭへ遷移する頻度が低いＨＭＭとでフレーム間隔を変更することができる。このように構成することにより、認識

率の向上と探索空間の縮小とを同時に実現することができる。

【００３８】

また、本実施の形態では、ＨＭＭの構成単位を音素として説明したが、これに限るものではない。音節、単語、diphone等の音韻を構成単位としてもよい。

【００３９】

また本実施の形態では、日本語の単語を認識する例について説明したが、これに限るものではない。日本語以外の言語にも適用できる。

【００４０】

なお本発明は、複数の機器（例えばホストコンピュータ、インターフェース機器、リーダ、プリンタなど）から構成されるシステムに適用しても、一つの機器からなる装置（例えば、複写機、ファクシミリ装置など）に適用してもよい。

【００４１】

また、本発明の目的は、前述した実施の形態の機能を実現するソフトウェアのプログラムコードを記録した記憶媒体（または記録媒体）を、システム或いは装置に供給し、そのシステム或いは装置のコンピュータ（またはＣＰＵやＭＰＵ）が記憶媒体に格納されたプログラムコードを読み出し実行することによっても達成される。この場合、記憶媒体から読み出されたプログラムコード自体が前述した実施の形態の機能を実現することになり、そのプログラムコードを記憶した記憶媒体は本発明を構成することになる。また、コンピュータが読み出したプログラムコードを実行することにより、前述した実施の形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているオペレーティングシステム（ＯＳ）などが実際の処理の一部または全部を行い、その処理によって前述した実施の形態の機能が実現される場合も含まれる。

【００４２】

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

【００４３】

以上説明したように本実施の形態によれば、最尤状態系列を探索する探索空間を形成する際に、特定のフレームのみＨＭＭ間の遷移を許すことによって、探索すべきパスを削減し、最適パスの探索に要する処理量を削減できる。この結果、従来の方法よりも高速な音声認識を実現することが可能となる。

【００４４】

【発明の効果】

以上説明したように本発明によれば、最適パスを探索する際の探索すべきパスを減らすことができ、音声認識のための探索処理に要する処理量を削減して高速な音声認識を可能にできるという効果がある。

【図面の簡単な説明】

【図１】本発明の実施の形態に係る音声認識装置のハードウェア構成を示すブロック図である。

【図２】本発明の実施の形態に係る音声認識装置における音声認識処理手順を示すフローチャートである。

【図３】本発明の実施の形態に係る探索処理を行う経路を説明する図である。

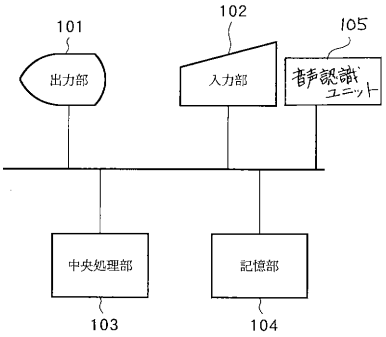
【図４】隠れマルコフモデルを説明する図である。

【図５】本発明の実施の形態における、認識対象語が複数の音素モデルで構成されている様子を示した図である。

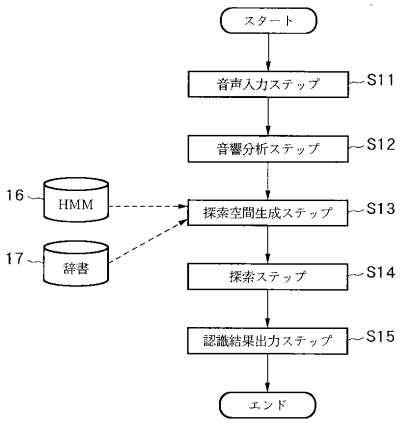
【図６】従来の音声認識処理の処理手順を示したフローチャートである。

【図７】従来の音声認識方法における探索処理経路を説明する図である。

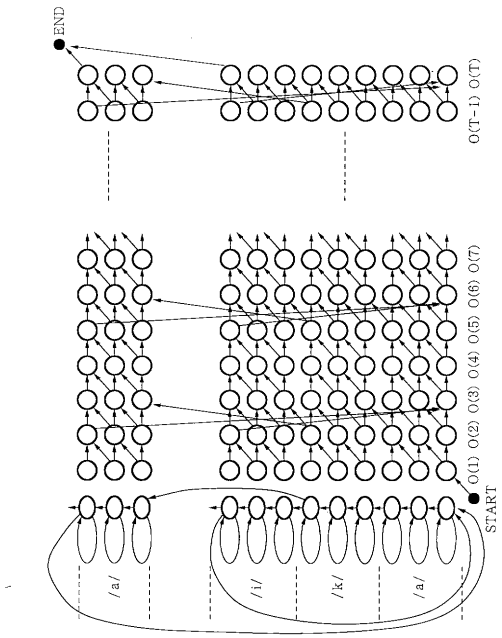
【図１】


101
出力部
102
入力部
105
音認識ユニット
中央処理部
103
記憶部
104

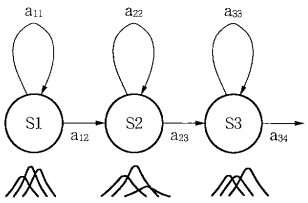
【図２】


スタート
音声入力ステップ
S11
音響分析ステップ
S12
16
HMM
探索空間生成ステップ
S13
17
辞書
探索ステップ
S14
認識結果出力ステップ
S15
エンド

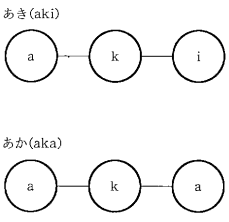
【図３】


END
O(T-1) O(T)
O(1) O(2) O(3) O(4) O(5) O(6) O(7)
START
/a/
/i/
/k/
/a/

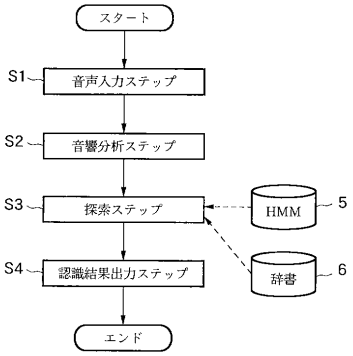
【図４】


a11
a22
a33
S1
S2
S3
a12
a23
a34

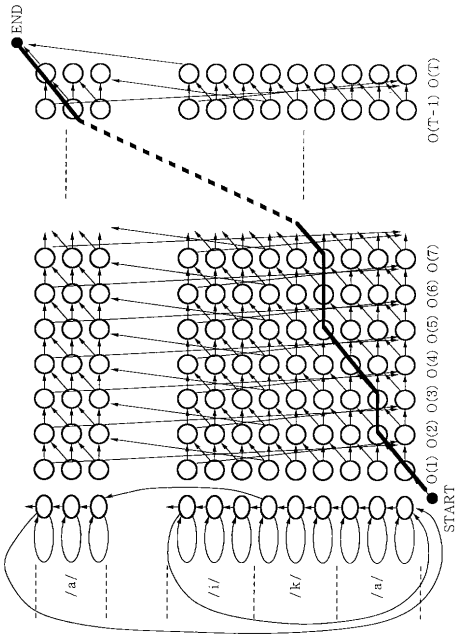
【図５】


あき(aki)
a
k
i
あか(aka)
a
k
a

【図６】


スタート
S1
音声入力ステップ
S2
音響分析ステップ
S3
探索ステップ
HMM
5
S4
認識結果出力ステップ
辞書
6
エンド

【図７】


END
O(T-1) O(T)
O(1) O(2) O(3) O(4) O(5) O(6) O(7)
START
/a/
/i/
/k/
/a/

フロントページの続き

審査官　江嶋　清仁

(56)参考文献　特開平０９－１２７９７８（ＪＰ，Ａ）
特開平０７－２６１７８６（ＪＰ，Ａ）
特開平０７－０８４５９３（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
G10L 15/14