(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3936266号
(P3936266)
(45)発行日　平成19年6月27日(2007.6.27)　(24)登録日　平成19年3月30日(2007.3.30)

(51)Int.Cl.　　F I
G10L 15/10 (2006.01)　　G10L 15/10 3OOZ
G10L 15/14 (2006.01)　　G10L 15/14 2OOA

請求項の数 10　外国語出願　(全 13 頁)

(21)出願番号　特願2002-265510 (P2002-265510)
(22)出願日　平成14年9月11日 (2002.9.11)
(65)公開番号　特開2004-163445 (P2004-163445A)
(43)公開日　平成16年6月10日 (2004.6.10)
審査請求日　平成16年7月7日 (2004.7.7)

(73)特許権者　393031586
株式会社国際電気通信基礎技術研究所
京都府相楽郡精華町光台二丁目２番地２
(74)代理人　100099933
弁理士　清水　敏
(72)発明者　マルコフ　コンスタンチン
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内
(72)発明者　中村　哲
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内

審査官　樫本　剛

最終頁に続く

(54)【発明の名称】音声認識装置およびプログラム

(57)【特許請求の範囲】
【請求項１】
音声認識のためのハイブリッド隠れマルコフ／ベイジアンネットワーク（ＨＭＭ／ＢＮ）モデルを記憶する記憶装置を含み、隠れマルコフモデル（ＨＭＭ）は時間的な音声の特徴をモデリングするのに用いられ、ベイジアンネットワーク（ＢＮ）は状態確率モデルをあらわすのに用いられ、さらに
当該記憶装置に記憶されたＨＭＭ／ＢＮモデルを用いて入来する音声データをデコードする音声デコーダを含む、音声認識装置。
【請求項２】
ＢＮの状態確率モデルが、変数Ｘと、以下で計算される条件付確率Ｐ（Ｙ｜Ｑ）とをさらに含み、
【数１】

$$P(Y|Q) = \frac{1}{N(x)} \sum_{x} P(Y|X = x, Q)$$

ただし、Ｙは一連の観測パラメータを表わし、Ｑは状態変数を表わし、ＸはＱとは独立してＹの値に影響を与える所定の要素を反映する変数を表わし、ｘはＸがとり得る値の一つであり、Ｎ（ｘ）はＸがとり得る値の数である、請求項１に記載の音声認識装置。
【請求項３】
前記変数Ｘが環境ノイズを表わす、請求項２に記載の音声認識装置。

【請求項４】
前記変数Ｘが話者の識別情報を表わす、請求項２に記載の音声認識装置。
【請求項５】
前記変数Ｘが話者の母語を表わす、請求項２に記載の音声認識装置。
【請求項６】
状態確率モデルが変数ＮとＳとをさらに含み、条件付確率Ｐ（Ｙ｜Ｑ）は以下で計算され、
【数２】

$$P(Y|Q)=\frac{1}{N(n,s)}\sum_{n,s}P(Y|N=n,S=s,Q)$$

ただし、Ｙは一連の観測パラメータを表わし、Ｑは状態変数を表わし、ＮおよびＳは、Ｑとは独立してＹの値に影響を与える所定の要素を反映する変数を表わし、ｎおよびｓはＮおよびＳがそれぞれとり得る値の一つであり、Ｎ（ｎ，ｓ）はＮおよびＳがとり得る値の組合せの数である、請求項１に記載の音声認識装置。
【請求項７】
前記変数Ｎがノイズの種類を表わす、請求項６に記載の音声認識装置。
【請求項８】
前記変数Ｓが入来する音声データの信号対雑音比を表わす、請求項６または請求項７に記載の音声認識装置。
【請求項９】
請求項１から請求項８のいずれかに記載の音声認識装置としてコンピュータを動作させるための、コンピュータで実行可能な音声認識プログラム。
【請求項１０】
請求項１から請求項８のいずれかに記載の音声認識装置としてコンピュータを動作させるための、コンピュータで実行可能な音声認識プログラムを記憶した、コンピュータ読取可能な記憶媒体。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
この発明は音声認識システムに関し、特に、ベイジアンネットワーク（ＢＮ）を採用したＨＭＭ（隠れマルコフモデル）に基づく音声認識システムに関する。
【０００２】
【従来の技術】
多年にわたり、ＨＭＭの音声認識への導入以来、各状態Ｑに対する観測の、条件付き分布Ｐ（ｙ｜Ｑ）は確率密度関数の混合によりモデル化されてきた（離散ＨＭＭはここでは考慮しない）。ガウスｐｄｆ（確率密度関数）およびラプラシアンｐｄｆがこの目的でよく使用されている。後には、ハイブリッドＨＭＭ/ＮＮ（隠れマルコフモデル／ニューラルネットワーク）システムが提案されたが、ここでは所与の入力観測に対しＨＭＭの状態の尤度を推定するのにニューラルネットワークが用いられている。
【０００３】
多くの場合、音声スペクトルから抽出された特徴がこれらの観測値を形成する。しかしながら、音声認識に関する研究により、これらの特徴のみを用いるだけでは高いシステム性能を達成するには不十分である事が示された。このため、多くの研究者が、ＨＭＭシステムに何か他の知識を表わす付加的な特徴を含めようとしてきた。
【０００４】
トクダら（後掲の非特許文献１）は、付加的なピッチ情報をモデル化するため、多元空間での確率分布を提案している。しかしほとんどの場合、付加的特徴の特性により種々の方策がとられている。この問題に対処すべき共通の、十分に柔軟性の高いフレームワークはこれまでには存在しなかった。

【０００５】
最近、ＨＭＭに対する別の選択肢として、ベイジアンネットワーク（ＢＮ）が研究者の関心を集めている。ＢＮは人工知能の研究分野では周知でありよく研究されている。しかし、音声認識においては、これらは比較的新しい研究課題である。
【０００６】
ベイジアンネットワークとは有向非巡回的グラフであり、そのノードは事象を表わす。ベイジアンネットワークでは、第一のノード（ノードＡ）から第二のノード（ノードＢ）への枝は、ＡがＢの原因である事を示す。各枝（Ａ　Ｂ）には、確率Ｐ（Ｂ｜Ａ）を示す「リンク」行列が割当てられ、この確率は、Ａの各値に対するＢの各値の確率を特定する。
【０００７】
ベイジアンネットワークは、多くの互いに異なる（離散的または連続的な）ランダムな変数の複合的な同時確率分布を、よく構造化され容易に表現可能な仕方でモデル化できる。特に、時間的音声特徴をモデル化するのに適しているのは、ディーンらの提案するダイナミックＢＮ（ＤＢＮ）である（後掲の非特許文献２）。
【０００８】
音声認識におけるＤＢＮの最初の報告のいくつかでは、ツバイクら、およびダウディらの報告でのように（後掲の非特許文献３、４）、これらは単独の語認識作業での語モデルとして用いられていた。これらの著作においては、ＤＢＮは、音声スペクトル情報に加えて、調音的特徴、サブ帯域相関、話し方のスタイル等の付加的な知識を容易に組み入れる事が可能な、ＨＭＭの一般化と考えられている。スティーブンソンらはＤＢＮのフレームワーク内で、音声学的な特徴をピッチ情報で容易に補う事ができる事を報告している。
【０００９】
ベイジアンネットワークの別の利点は、認識中に信頼性をもって推定するのが困難な付加的な特徴を、隠されたまま、すなわち観測不能な状態にしておける事である。
【００１０】
【非特許文献１】
Ｋ.トクダ、Ｔ.マスコ、Ｎ.ミヤザキ、Ｔ.コバヤシ、『ピッチパターンモデリングのための多元空間確率分布に基づく隠れマルコフモデル』、ＩＣＡＳＳＰ予稿集、ｐｐ．２２９－２３２，１９９９年。（K. Tokuda, T. Masuko, N. Miyazaki, and T. Kobayashi, “Hidden Markov Models based on multi-space probability distribution for pitch pattern modeling.” In Proc. ICASSP, pp. 229-232, 1999.）
【非特許文献２】
Ｔ.ディーン、Ｋ.カナザワ、「確率的な時間的推論」、ＡＡＡＩ，ｐｐ．５２４－５２８，１９８８年。（T. Dean and K. Kanazawa, “Probabilistic temporal reasoning,” in AAAI, pp. 524-528, 1988.）
【非特許文献３】
Ｇ. ツヴァイク、Ｓ.ラッセル、『自動音声認識のためのベイジアンネットワークによる確率モデリング』、ＩＣＳＬＰ予稿集、ｐｐ．３０１０－３０１３，１９８８年。（G. Zweig and S. Russell, “Probabilistic modeling with Bayesian Networks for automatic speech recognition,” in Proc. ICSLP, pp. 3010-3013, 1998.）
【非特許文献４】
Ｋ.ダウディ、Ｄ.フォア、Ｃ.アントワーヌ、『確率的グラフモデルに基づく多帯域音声認識の新たなアプローチ』、ＩＣＳＬＰ予稿集，ｖｏｌ．１，ｐｐ．３２９－３３２，２０００年。（K. Daoudi, D. Fohr, and C. Antoine, “A new approach for multi-band speech recognition based on probabilistic graphical models,” in Proc. ICSLP, vol. 1, pp. 329-332, 2000.）
【非特許文献５】
Ｔ.スティーブンソン、Ｍ.マシュウ、Ｈ. ボラード、『ベイジアンネットワークに基づ

くＡＳＲでの補助情報のモデル化』、ユーロスピーチ予稿集、ｐｐ．］２７６５－２７６８，２００１年。（T. Stephenson, M. Mathew, and H. Bourlard, “Modeling auxiliary information in Bayesian Network based ASR,” in Proc. Eurospeech, pp. 2765-2768, 2001.）

【非特許文献６】

Ｋ.ダウディ、Ｄ.フォア、Ｃ.アントワーヌ、「ベイジアンネットワークを用いた連続多帯域音声認識」、Proc. ASRU, 2001年。（K. Daoudi, D. Fohr, and C. Antoine, “Continuous multi-band speech recognition using Bayesian Networks,” in Proc. ASRU, 2001.）

【発明が解決しようとする課題】

ＢＮのこれらの魅力的な特性にも関わらず、音声認識への応用は依然として、小さな、単独の語認識作業に対するものに限られている。その理由は、ＢＮパラメータの学習および推論のための既存のアルゴリズムが、連続した音声認識（ＣＳＲ）および、特に語彙の多いＣＳＲ作業にはそれほど適していないからである。連続して発話された数字の認識を可能にするＤＢＮ語モデルの拡張が、ダウディらの非特許文献６に報告されてはいるものの、作業対象となる語彙がわずか数百に増加するだけでも、あまりにも計算量が多くなり過ぎる。

【００１１】

従って、この発明の目的の一つは、ＢＮを採用したものであって、かつＣＳＲに適した装置を提供する事である。

【００１２】

この発明の別の目的は、ＢＮを採用したものであって、かつ語彙量の多いＣＳＲ作業に適した装置を提供する事である。

【００１３】

この発明のさらに別の目的は、ＢＮを採用したものであって、かつＢＮのための煩瑣なパラメータ学習を必要としない装置を提供する事である。

【課題を解決するための手段】

この発明のある局面に従った音声認識装置は、音声認識のためのハイブリッド隠れマルコフ／ベイジアンネットワーク（ＨＭＭ／ＢＮ）モデルを記憶するための記憶装置を含む。隠れマルコフモデル（ＨＭＭ）は時間的な音声の特徴をモデリングするのに用いられ、ベイジアンネットワーク（ＢＮ）は状態確率モデルをあらわすのに用いられる。この装置はさらに、記憶装置に記憶されたＨＭＭ／ＢＮモデルを用いて入来する音声データをデコードするための音声デコーダを含む。

【００１４】

ＢＮの状態確率モデルは、変数Ｘを含んでもよく、条件付確率Ｐ（Ｙ｜Ｑ）は以下の式によって計算され、

【００１５】

【数３】

$$P(Y|Q)=\frac{1}{N(x)}\sum_{x}P(Y|X=x,Q)$$

ただし、Ｙは一連の観測パラメータを表わし、Ｑは状態変数を表わし、ＸはＱとは独立してＹの値に影響を与える所定の要素を反映する変数を表わし、ｘはＸがとり得る値の一つであり、Ｎ（ｘ）はＸがとり得る値の数である。

【００１６】

好ましくは、変数Ｘは環境ノイズ、話者認識情報、または話者の母語を表わしてもよい。

【００１７】

状態確率モデルは変数ＮとＳとをさらに含んでもよく、条件付確率Ｐ（Ｙ｜Ｑ）は以下で計算され、

【００１８】
【数４】

$$P(Y|Q)=\frac{1}{N(n,s)}\sum_{n,s}P(Y|N=n,S=s,Q)$$

ただし、Ｙは一連の観測パラメータを表わし、Ｑは状態変数を表わし、ＮおよびＳはＱとは独立してＹの値に影響を与える所定の要素を反映する変数を表わし、ｎおよびｓはＮおよびＳがそれぞれとり得る値の一つであり、Ｎ（ｎ，ｓ）はＮおよびＳがとり得る値の組合せの数である。

【００１９】
変数Ｎはノイズの種類を表わし、変数Ｓは入来する音声データの信号対雑音比を表わしてもよい。

【００２０】
この発明の別の局面は、上述の音声認識装置としてコンピュータを動作させる、コンピュータで実行可能な音声認識プログラムに関する。

【００２１】
この発明のさらに別の局面は、上述の音声認識装置としてコンピュータを動作させる、コンピュータで実行可能な音声認識プログラムを記憶する、コンピュータ読取可能な記憶媒体に関する。

【００２６】
【発明の実施の形態】
－ハイブリッドＨＭＭ／ＢＮモデリング－
多くの場合、音声認識でＢＮを使用するのは、ＨＭＭをダイナミックなＢＮとして表わそうとする考えに基づいている。このような表現を図１に示す。ここで、Q_tは状態変数であり、Y_tは時刻ｔ＝1，2，3，4，…での連続な観測変数である。枝は変数間の確率的依存を表わす。状態インスタンス間の枝はＨＭＭ遷移確率を表わし、状態インスタンスと観測インスタンスとの間の枝はＨＭＭ状態の条件付分布を表わす。以下の図において、四角で囲った変数は離散的であり、丸で囲った変数は連続である。ハッチングした丸／四角は観測可能な変数を示す。

【００２７】
Ｙ＝y_1，…，y_T、Ｍは本発明のモデルとしたとき、入力される観測シーケンスＰ（Ｙ｜Ｍ）の尤度を得る必要がある場合、ＨＭＭおよびＢＮの表現にＢＮ推論アルゴリズムを用いる必要がある。この作業の間に、ネットワークのサイズが入力シーケンスＴのサイズに合わせて調整され、その後、ネットワーク全体からＰ（Ｙ｜Ｍ）が推論される。

【００２８】
状態ノード間の枝を切断した場合を想定する。こうする事により、各時刻ｔに対応する、図２に示すように複数の独立したＢＮが得られる。時間遷移（切断された枝）が従来のＨＭＭにより支配されるとすれば、これらのＢＮを適切なＨＭＭ状態に割当てる事で時間指標をなくす事ができる。ＢＮは全て同じ共通の構造を有するので、それらを図３のように単一のＢＮとして表わす事ができる。図３では、変数Ｑは音声学的モデル内の全てのＨＭＭの状態指標（S_{ij}）の値をとり、状態確率分布Ｐ（Ｙ｜Ｑ＝S_{ij}）は枝によって表わされる。

【００２９】
こうして、ガウスの混合ではなく、状態分布モデルとしてＢＮを有するように従来のＨＭＭを修正した。ＨＭＭとＢＮとをこのように組合せる事で、ＨＭＭ／ＢＮモデルが階層的となる。ＢＮは最下層にあり、ＨＭＭは最上層にある。なお、状態変数Ｑ（図３）はＢＮに対しては観測可能となるが、上のＨＭＭレベルでは、依然として隠されたままである。

【００３０】
状態ＢＮは付加的な知識を表わす他のランダム変数に容易に拡張可能である。簡単な作業とは言いがたいデータからの学習によってではなく、変数間の関係に関する発明者らの知

識に従って、拡張ＢＮのグラフィック構造を課する事ができる。この実施例は煩瑣なパラメータ学習とは無縁である。

【００３１】

拡張状態ＢＮの可能な構造のいくつかを図４に示す。たとえば、変数Ｘはこの図では環境ノイズの種類を表わす事ができ、他のＷおよびＺ変数は話者のＩＤ（識別情報）および話者の母語を表わす事ができる。

【００３２】

このＨＭＭ／ＢＮモデルで認識を行なう場合には、従来のＨＭＭと同様に、各状態$Q=q_{ij}$についてＰ（ｙ｜Ｑ）を計算する必要がある。ただし、「ｉ」はＨＭＭ指標であり、「ｊ」はｉ番目のＨＭＭの状態指標である。この値をＢＮ確率モデルから推論する事ができ、これを行うアルゴリズムとして、正確なものと近似的なものとを含め多数の推論アルゴリズムがある。単純なＢＮでは、図４（ａ）に示すように、「力任せ」法さえも適用可能である。このＢＮに対する同時確率モデルは、以下のようにチェーンルールで表わす事ができる。

【００３３】

【数５】

$$P(X,Y,Q)=P(Y|X,Q)*P(X|Q)*P(Q) \quad (1)$$

ＸおよびＱは独立した変数なので、「Ｐ（Ｘ｜Ｑ）＝Ｐ（Ｘ）」が成立し、従って上の式は以下のように書換える事ができる。

【００３４】

【数６】

$$P(X,Y,Q)=(P|X,Q)*P(X)*P(Q) \qquad (2)$$

従って、求める確率Ｐ（Ｙ｜Ｑ）はＸに対するマージナライゼーションにより、以下のようになる。

【００３５】

【数７】

$$P(Y|Q)=\frac{P(Y,Q)}{P(Q)}=\frac{\sum_x P(Y,X=x,Q)}{P(Q)}$$

$$=\frac{\sum_x P(Y|X=x,Q)*P(X=x)*P(Q)}{P(Q)}$$

$$=\sum P(Y|X=x,Q)*P(X=x) \qquad (3)$$

実用上、多くの場合には全てのＸ＝ｘについてＰ（Ｘ）は同じであると仮定できるので、式（３）を次のように変形できる。

【００３６】

【数８】

$$P(Y|Q)=\frac{1}{N(x)}\sum_x P(Y|X=x,\ Q) \qquad (4)$$

ここで、Ｎ（ｘ）は変数Ｘがとり得る値の数である。

【００３７】

ＢＮパラメータのトレーニングは、従来のＨＭＭ状態パラメータのトレーニングとほぼ同様に、各状態について独立して行なう事ができる。トレーニングの詳細は図１０を参照して後述する。

【００３８】

－ノイズの多い音声認識システムにおけるＨＭＭ／ＢＮモデル－

音声がノイズを含む場合、音声の特徴ベクトルはその分布を変え、その変化はノイズの種類とともにＳＮＲ（信号対雑音比）の値にも依存する。このため、この依存性を図５に示すような種類の状態ＢＮで表わす事ができる。ここで、ＮとＳとはそれぞれ、ノイズの種類とＳＮＲ値とを示す、隠れた離散的変数である。この場合、状態の尤度は式（３）を導出したのと同じ方法で解析的に表現できる。

【００３９】

多くの場合、事前確率Ｐ（Ｎ）およびＰ（Ｓ）は、ノイズの各種類および各ＳＮＲ値について等しいと合理的に仮定できるので、以下のとおりとなる。

【００４０】

【数９】

$$P(Y|Q) = \frac{1}{N(n,s)} \sum_{n,s} P(Y|N=n, S=s, Q) \qquad (5)$$

語モデルおよびサブワードモデルも、従来のＨＭＭの場合と同様に作成される。デコードもまた、デコーダを変更する事なく、標準的なＨＭＭベースのシステムと同様に行なう事ができる。Ｎ（ｎ，ｓ）はＮとＳのとり得る組合せの数を示す。

【００４１】

この実施例のＨＭＭ／ＢＮモデルの構造を図６に要約して示す。

【００４２】

－コンピュータでの実現例－

図７はこの実施例の全体図である。図７を参照して、このシステムは、トレーニングデータ１１０とＨＭＭ／ＢＮモデル１１２とに基づいてＨＭＭ／ＢＮハイブリッドモデル１１４をトレーニングするためのトレーニングシステム１００と、ＨＭＭ／ＢＮハイブリッドモデル１１４を記憶するための媒体１０４と、コンピュータシステムで実現され、音声データ１２０を媒体１０４に記憶されたＨＭＭ／ＢＮハイブリッドモデル１１４でデコード（音声認識１２２）し、認識された音声１２４を出力するための音声認識システム（音声デコーダ）１０２とを含む。

【００４３】

図８はこのコンピュータシステム３０を概略的に示し、図９はシステム３０をブロック図形式で示す。図８を参照して、このコンピュータシステム３０は、ＦＤ（フレキシブルディスク）ドライブ５２およびＣＤ－ＲＯＭ（コンパクトディスク読出専用メモリ）ドライブ５０を有するコンピュータ４０と、キーボード４６と、マウス４８と、モニタ４２とを含む。

【００４４】

図９を参照して、コンピュータ４０は、ＦＤドライブ５２およびＣＤ－ＲＯＭドライブ５０に加えて、ＣＰＵ（中央処理装置）５６と、ＣＰＵ５６、ＦＤドライブ５２およびＣＤ－ＲＯＭドライブ５０に接続されたバス６６と、ブートアッププログラムおよびＨＭＭデコードプログラム等を記憶する読出専用メモリ（ＲＯＭ）５８と、バス６６に接続され、プログラム命令、システムプログラム、およびＨＭＭ／ＢＮハイブリッドモデルデータを記憶するランダムアクセスメモリ（ＲＡＭ）６０とを含む。

【００４５】

ここでは示さないが、コンピュータ４０はさらにローカルエリアネットワーク（ＬＡＮ）への接続を提供するネットワークアダプタボードを含んでもよい。音声認識をリアルタイムで行なう場合には、コンピュータシステム３０はさらにマイクロフォンとオーディオキャプチャボードとを含むことになる。

【００４６】

コンピュータシステム３０に音声認識を行なわせるプログラムは、ＣＤ－ＲＯＭドライブ５０またはＦＤドライブ５２に挿入されるＣＤ－ＲＯＭ６２または図示しないＦＤに記憶され、さらにハードディスク５４に転送される。またはこれに代えて、プログラムは図示しないネットワークを通じてコンピュータ４０に送信されハードディスク５４に記憶さ

れてもよい。プログラムは実行の際にＲＡＭ６０にロードされる。ＣＤ－ＲＯＭ６２、ＦＤ、またはネットワークを介してＲＡＭ６０にプログラムを直接ロードしてもよい。

【００４７】

プログラムは、コンピュータ４０にこの実施例の音声認識を行なわせるいくつかの命令を含む。この方法を行なわせるのに必要な基本的機能のいくつかはコンピュータ４０のオペレーティングシステム（ＯＳ）またはサードパーティのプログラム、もしくはコンピュータ４０にインストールされるＨＭＭツールキット等のモジュールにより提供されるので、このプログラムはこの実施例の方法を実現するのに必要な機能全てを必ずしも含まなくてよい。このプログラムは、命令のうち、所望の結果が得られるように制御されたやり方で適切な機能または「ツール」を呼出す事により音声認識プロセスを実行する命令のみを含んでいるだけでよい。コンピュータシステム３０の動作は周知であるので、ここでは繰り返さない。

【００４８】

音声認識システム１０２は従来のＨＭＭデコードプログラムで実現される。この実施例で新規な点は、ＨＭＭ／ＢＮハイブリッドモデル１１４が媒体１０４上に記憶されている事である。

【００４９】

－ＨＭＭ／ＢＮモデルでのトレーニングと認識－

ＨＭＭ／ＢＮモデルのトレーニングには、ＨＭＭ／ＮＮトレーニングと同じアプローチを採用する事ができる。これはビタビトレーニングアルゴリズムに基づくものである。全体の制御フローを図１０に示す。まず始めに、ＨＭＭと状態ベイジアンネットワークとのトポロジーを選択する。モデルを初期化し、特徴抽出ステップ１５０で入力音声１４０から抽出した特徴１４４に対し、状態分離ステップ１５２で、ビタビアライメントが基本ＨＭＭモデル１４２を用いた基本認識部を使用して行なわれる。

【００５０】

状態の分離は状態ベイジアンネットワークのためのトレーニングデータを生成するのに用いられ、つぎにこの状態ベイジアンネットワークはステップ１５４、１５６および１６０でトレーニングされる。このビタビトレーニング手法では、トレーニングの時間的な部分と静的な部分とが分離される。この処理では、ステップ１５８で終了条件が満たされるまで、ステップ１５４および１５６のＢＮトレーニングと、埋込まれたトレーニング処理であるステップ１６０の遷移確率の再推定とを交互に繰返す。

【００５１】

このＨＭＭ／ＢＮモデルで認識を行なう場合、従来のＨＭＭと同様に、通常のビタビデコードアルゴリズムが用いられる。ここで、各状態$Q=q_{ij}$（図６を参照）についてＰ（ｙ｜Ｑ）を計算する必要がある。ここでｉはＨＭＭ指標であり、ｊはｉ番目のＨＭＭの状態指標である。この値を、標準的推論アルゴリズムを用いてＢＮ確率モデルから推論する事ができる。

【００５２】

－Ａｕｒｏｒａ２タスクでの評価－

この実施例の音声認識システムをＡｕｒｏｒａ２タスクで評価する実験を行なった。これらの実験では、公式のＡｕｒｏｒａ２タスクで示唆されている評価用シナリオに忠実に従った。最も関心があったのは、ＨＭＭ／ＢＮシステムを複数条件でトレーニングされたＨＭＭシステムと比較する事であった。ＨＭＭ／ＢＮ状態の条件付き分布のトレーニングにあたっては、トレーニングデータをノイズの種類とＳＮＲ値とで分け、ＨＴＫ（ＨＭＭツールキット、ＨＭＭ処理のためのソフトウェアツールキット）を用いて各条件についてのパラメータを別個にトレーニングした。特徴ベクトル、語モデル、状態数、実験条件等の全ての他のシステムパラメータは同一にした。なお、ＨＭＭ／ＢＮシステムでは、何らかの適合やノイズに強い方法を用いたわけではない。二つのシステムでの主な機能的相違点は、ＨＭＭ／ＢＮシステムが音声の特徴およびノイズの隠れた依存性を探求する事である。

【００５３】
テストセットＡ（トレーニングデータと同じノイズ種類）とテストセットＢ（異なるノイズ）の認識結果を表１にまとめて示す。理解できるとおり、ＨＭＭ／ＢＮシステムの性能は閉じたノイズ条件（Ａセット）ではかなり高く、ずっと複雑なシステムでこの作業について得られる最新の結果に迫っている。
【００５４】
【表１】

表 1.　HMMおよびHMM／ＢＮシステムの性能（%）

SNR	テストセット A		テストセット B	
	HMM	HMM／ＢＮ	HMM	HMM／ＢＮ
Clean	98.54	98.83	98.54	98.83
20 dB	97.52	98.12	96.96	97.26
15 dB	96.94	97.65	95.38	95.05
10 dB	94.59	96.04	92.58	90.27
5 dB	87.51	91.7	83.5	78
0 dB	59.84	76.11	58.91	48.7
-5 dB	23.46	35.79	23.86	3.18
平均*	87.29	91.92	85.46	81.85
誤差	12.71	8.08	14.54	18.15
*20dBから 0dBの値で計算				

この評価は、ノイズの種類とＳＮＲ値とを付加的なパラメータとして加え、これらの依存性を検討する事で、結果として得られるスペクトル特徴パラメータの誤差が３６．４％（＝（１２．７１－８．０８）＊１００／１２．７１）減少する事を示している。
【００５５】
Ｂセットの条件についていえば、性能の劣化が見られる。これは、音声スペクトル特徴の分布のミスマッチに加えて、このＨＭＭ／ＢＮシステムには新たなノイズの依存性に関する知識が得られないという事実で説明がつく。他方で、複数条件ＨＭＭシステムでは、状態ガウス混合は多ノイズおよびＳＮＲ条件から複雑な分布をあまりうまくモデル化できていない事が明らかである。しかしながら、このデータとモデル分布のミスマッチにはある種の平滑化の効果があり、これによって、見られないデータから一般化するというこのモデルの能力が高まる。
【００５６】
[ハイブリッドＨＭＭ／ＢＮモデルの応用]
明らかに、提案されたハイブリッドＨＭＭ／ＢＮモデルはノイズの多い音声認識システムのみでなく、観測や隠れた特徴が新たに得られる事で性能上の利益が得られる他の多くの場合に適用可能である。このアプローチは、種々の空間からの特徴を組合せ、その間の依存性を検討する事により、システムのモデル化能力を高めるという、より一般的なフレームワークという方がより正確である。とくに興味深いのは、ＢＮの隠れた変数の確率が推論できる可能性である。このように、ＨＭＭ／ＢＮシステムはこれらの付加的パラメータの認識に用いる事ができる。
【００５７】
例えば、ある付加的な隠れ変数Ｘが多言語システムでの言語を表わす場合、各フレームについてＰ（Ｘ｜Ｑ）を計算し、これらの確率を、入力された発話全体にわたり累積する事ができる。その場合、$x = \mathrm{argmax}_x P(x \mid Q_s)$、Ｑｓは最高の仮定状態シーケンス、となるｘは、その発話がなされた言語として最も確率の高い言語を示す。従って、多言語の音声認識に加えて、このようなシステムは言語認識をも行なう事ができる。な

お、関数ｘ＝ａｒｇｍａｘ$_{x}$　Ｐ（ｘ｜Ｑ$_{s}$）は、Ｐ（ｘ｜Ｑ$_{s}$）を最大にするｘを示す。

【００５８】

[この実施の形態の効果]

この実施の形態では、ＨＭＭとＢＮとを単一のモデル内で組合せ、ＨＭＭとＢＮとの両者の長所を活かしている。ハイブリッドＨＭＭ／ＢＮモデルにより、音声認識システムに他の情報を容易に加える事が可能になり、最小限の費用でその性能を高める事ができる。さらに、ＨＭＭ／ＢＮモデルは従来のＨＭＭと同様に、サブワードの音声ユニットを表わす事ができる。こうして、ＢＮフレームワークを語彙数の多い連続した音声認識に用いる事が可能になる。

【図面の簡単な説明】

【図１】　ＨＭＭをＤＢＮとして表わす模式図である。

【図２】　各時刻ｔでの複数ＢＮの模式図である。

【図３】　通常のＢＮ構造を示す、状態ＢＮの模式図である。

【図４】　（ａ）は離散的変数を一つ付加した状態ＢＮの模式図であり、（ｂ）はより複雑な構造の状態ＢＮの模式図である。

【図５】　ノイズおよびＳＮＲ変数を有する状態ＢＮの模式図である。

【図６】　この発明の一実施の形態に係るＨＭＭ／ＢＮハイブリッドモデル構造の模式図である。

【図７】　この発明の一実施の形態の音声認識システムのブロック図である。

【図８】　この実施の形態の音声認識システムのプログラムを実行するコンピュータシステムの外観図である。

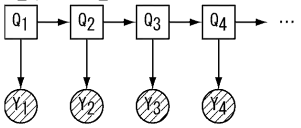
【図９】　図８のコンピュータシステムのブロック図である。

【図１０】　この実施の形態のＨＭＭ／ＢＮモデルのトレーニングプログラムの制御の流れを示すフローチャートである。

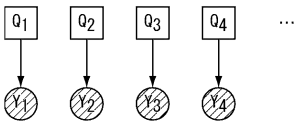
【符号の説明】

１００　トレーニングシステム、１０２　音声認識システム、１０４　記憶媒体、　１１０トレーニングデータ、１１２　ＨＭＭ／ＢＮモデル、１１４　ＨＭＭ／ＢＮハイブリッドモデル、１２０　音声データ、１２２　音声認識、１２４　出力

【図１】



【図２】

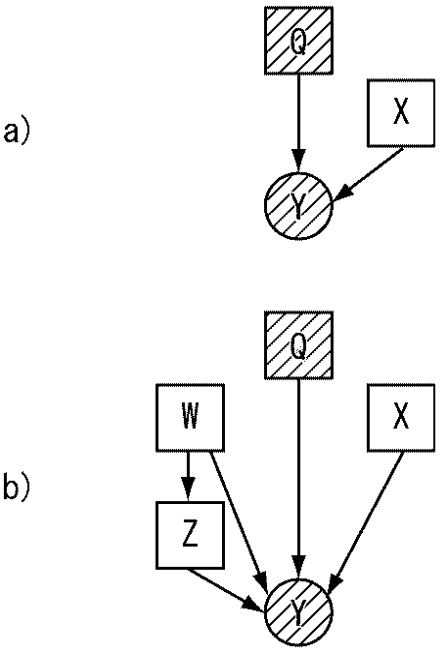


各時刻ｔにおける複数ＢＮ

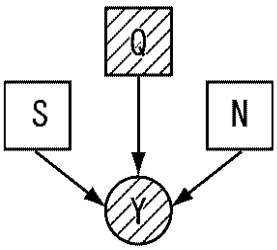
【図３】

状態ＢＮ

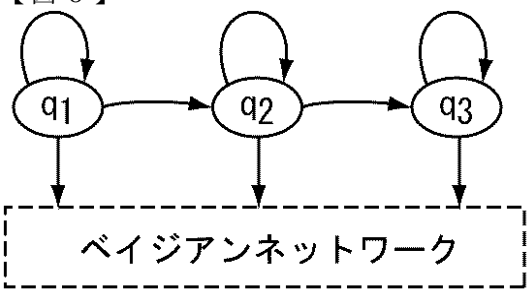
【図４】



【図５】

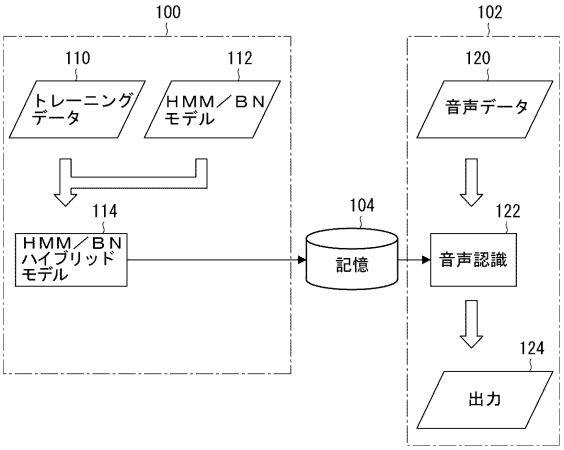


【図６】

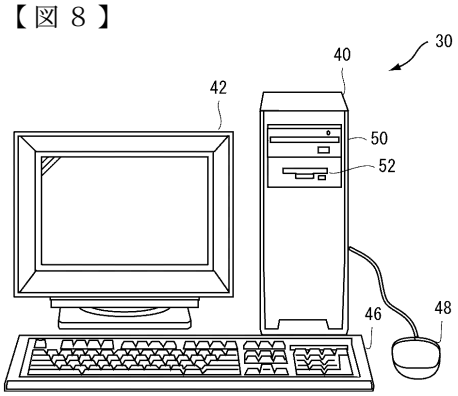


ＨＭＭ／ＢＮモデル構造

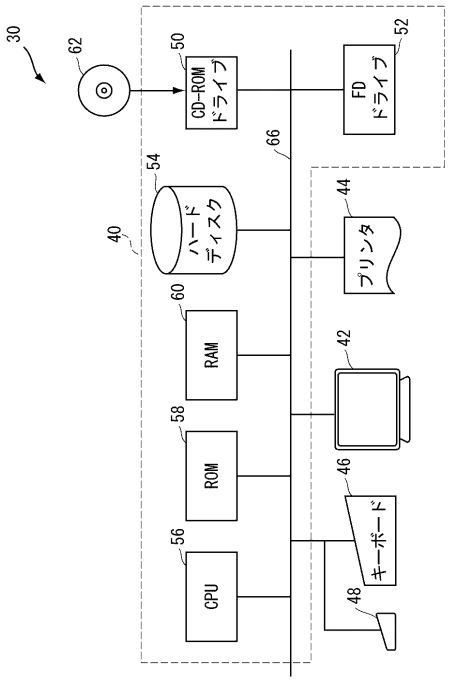
【図７】



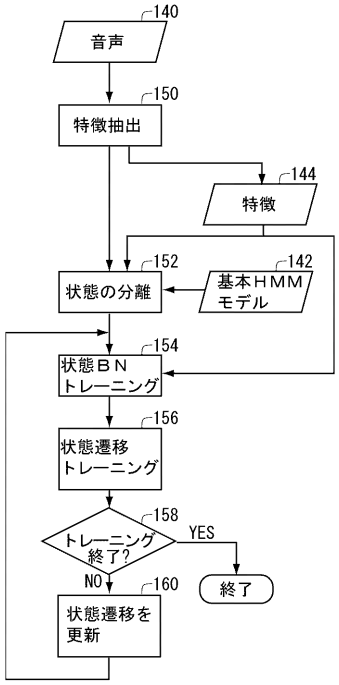
【図８】



【図９】


30
62
50
CD-ROM ドライブ
52
FD ドライブ
66
54
ハードディスク
40
44
プリンタ
60
RAM
42
58
ROM
46
キーボード
56
CPU
48

【図１０】


140
音声
150
特徴抽出
144
特徴
152
状態の分離
142
基本ＨＭＭ モデル
154
状態ＢＮ トレーニング
156
状態遷移 トレーニング
158
トレーニング 終了?
YES
終了
NO
160
状態遷移を 更新

フロントページの続き

(56)参考文献 特許第３５０８９７８（ＪＰ，Ｂ２）
中川，音声認識においてHMMとトライグラムを超えるもの，人工知能学会誌，日本，２００２年１月 １日，Vol.17、No.12，p.35-40
本村他，ベイジアンネットワーク －不確定性のモデリング技術－，人工知能学会誌，日本，２０００年 7月，Vol.15、No.4，p.575-582
中川，音声認識研究の動向，電子情報通信学会論文誌 D-II，日本，２０００年 2月２５日，Vol.J83-D-II、No.2，p.433-457
S. K. Riis et al., Hidden neural networks: a framework for HMM/NN hybrids, Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97), 米国，１９９７年 4月２１日，Vol.4，p.3233-3236

(58)調査した分野(Int.Cl.，ＤＢ名)
G10L 15/10
G10L 15/14
JSTPlus(JDream2)
IEEE