(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5065693号
(P5065693)

(45)発行日　平成24年11月7日(2012.11.7)　(24)登録日　平成24年8月17日(2012.8.17)

(51)Int.Cl.		F I		
G06N 3/00	(2006.01)	G06N	3/00	560G
G10L 15/06	(2006.01)	G10L	15/06	310Z
G10L 15/18	(2006.01)	G10L	15/18	200E
G10L 15/14	(2006.01)	G10L	15/14	200Z

請求項の数 4　外国語出願　(全 24 頁)

(21)出願番号　特願2007-18135(P2007-18135)
(22)出願日　平成19年1月29日(2007.1.29)
(65)公開番号　特開2008-186171(P2008-186171A)
(43)公開日　平成20年8月14日(2008.8.14)
審査請求日　平成22年1月27日(2010.1.27)

(73)特許権者　301022471
独立行政法人情報通信研究機構
東京都小金井市貫井北町４－２－１
(74)代理人　100099933
弁理士　清水　敏
(72)発明者　コンスタンティン・マルコフ
東京都小金井市貫井北町４－２－１　独立行政法人情報通信研究機構内
(72)発明者　中村　哲
東京都小金井市貫井北町４－２－１　独立行政法人情報通信研究機構内

審査官　新井　寛

最終頁に続く

(54)【発明の名称】空間－時間パターンを同時に学習し認識するためのシステム

(57)【特許請求の範囲】
【請求項１】
ネットワークモデルを用いて物理的測定値から導出された特徴ベクトルのシーケンスの空間－時間パターンを同時に学習し認識するためのシステムであって、
前記特徴ベクトルは予め定められた特徴量空間内で定義されており、
前記ネットワークモデルは、前記特徴量空間に定義される一組の状態と、前記状態間の遷移と、前記状態間の横方向接続とを含み、
前記状態の各々は、出力値の確率密度関数を規定し、前記遷移の各々は、状態から状態への遷移を規定するとともに前記遷移の発生頻度と関連付けられており、前記横方向接続の各々は、隣接する状態の対を規定するとともに、前記横方向接続によって接続された状態のいずれかが前記システムによって最後に経由されてからの経過時間の測定値と関連付けられており、
前記システムは、
前記ネットワークモデルを表すデータセットを記憶するためのモデル記憶手段と、
現在の状態の識別子を記憶するための現在状態記憶手段と、
新たな特徴ベクトルに応答して、新たな特徴ベクトルに最も良く整合する状態であって、かつ前記特徴量空間において前記新たな特徴ベクトルからあるしきい値距離内にある状態が存在する場合は、それを次の状態と決定し、存在しない場合は前記ネットワークに新たな状態を追加するための決定手段とを含み、前記新たな状態は、前記新たな特徴ベクトルと現在の状態の識別子によって特定される現在の状態とによって規定され、かつ前記新

たな状態は、前記現在の状態からの次の遷移を規定し、
前記システムはさらに、
前記次の状態が決定されたことに応答して、前記モデル記憶手段に記憶された前記モデルにおける次の遷移の頻度を更新するための手段と、
前記次の状態が決定されたことに応答して、前記次の状態と、それに隣接する状態との前記確率密度関数の各々を、予め定められた更新関数によって更新するための手段と、
前記次の状態が決定されたことに応答して、前記次の状態と、その隣接する状態との接続に関連付けられた経過時間の測定値が予め定められた初期値にリフレッシュされ、かつ他の接続に関連付けられた経過時間の測定値が増分されるように、前記ネットワークモデル内の横方向接続を更新するための手段と、
前記横方向接続が更新されたことに応答して、予め定められたしきい値より大きい経過時間の測定値と関連付けられている接続を削除するための手段と、
前記接続のいずれかが削除されたことに応答して、何の接続も有していない状態を前記ネットワークモデルから除去するための手段と、
前記現在の状態の識別子を出力状態シーケンスの末尾に追加するための手段と、
前記現在状態記憶手段に記憶された前記現在の状態の識別子を、前記次の状態の識別子で置換するための手段とを含む、システム。
【請求項２】
前記決定手段は、
前記現在の状態からの遷移を有する状態の組にあって、かつ前記新たな特徴ベクトルから前記しきい値距離内にある、前記新たな特徴ベクトルに最も近い、次の状態の候補を発見するための第１の発見手段と、
前記第１の発見手段が次の状態の候補を発見できなかったことに応答して、前記現在の状態からの遷移を持たず、前記新たな特徴ベクトルから前記しきい値距離内にある、前記新たな特徴ベクトルに最も近い、次の状態の候補を発見するとともに、前記ネットワークモデルを、前記現在の状態から前記次の状態の候補への新たな遷移が生成されるように更新するための、第２の発見手段と、
前記第１又は第２の発見手段によって次の状態の候補が発見されたことに応答して、前記次の状態の識別子を前記次の状態の候補の識別子に設定するための手段と、
前記第１又は第２の発見手段によって次の状態の候補が発見されなかったことに応答して、前記ネットワークモデルに新たな状態を追加するための手段とを含み、前記新たな状態は前記新たな特徴ベクトルによって規定される確率密度関数と前記現在の状態からの遷移とを有し、前記新たな状態の前記遷移は初期頻度の値と関連付けられている、請求項１に記載のシステム。
【請求項３】
前記横方向接続を更新するための手段は
前記次の状態と、前記新たな特徴ベクトルに次に近い状態との間の接続が生成されるように前記ネットワークモデルを更新するための手段と、
前記次の状態とそれに隣接する状態との間の接続の経過時間の測定値をリフレッシュするための手段と、
前記ネットワークモデル内の接続の経過時間の測定値を増分するための手段とを含む、請求項１に記載のシステム。
【請求項４】
コンピュータ上で実行されると、コンピュータを請求項１～請求項３のいずれかに記載のシステムとして機能させる、コンピュータプログラム。

【発明の詳細な説明】
【技術分野】
【０００１】
この発明はパターン認識システムに関し、特に、測定値又は観察値の空間－時間パター

ンを同時に学習し認識するための、教師無しの適応学習能力を有するパターン認識装置に関する。

【背景技術】

【０００２】

＜はじめに＞

現在の自動音声認識システムは、２つの別々の動作モードを有する。トレーニングと、認識とである。トレーニングの後、システムのパラメータは固定され、トレーニング条件とテスト条件との間にミスマッチが生じると、通常は適応手順が行なわれる。

【０００３】

生物学的及び技術的観点から見れば、ライフスパンを学習段階と認識段階とに人為的に分離することは現在の自動音声認識（Ａｕｔｏｍａｔｉｃ　Ｓｐｅｅｃｈ　Ｒｅｃｏｇｎｉｔｉｏｎ：ＡＳＲ）システムの欠点である。こうした方法は、整合した環境で動作するシステムでは可能であるが、環境が変わればうまくいかない。費用のかかる再トレーニングを避けるため、最近の研究では高速適応化とオンライン適応学習とに焦点をあてている。

【０００４】

しかし、このような方法は必然的に、それまでよく学習してきたパターンを破壊してしまう。これは、認知科学において壊滅的忘却（ｃａｔａｓｔｒｏｐｈｉｃ　ｆｏｒｇｅｔｔｉｎｇ）として知られる現象である。インテリジェントなシステムであれば、変化する環境に適応するのみならず、その知識を保存することもできなければならない。これは、壊滅的忘却なしの、生涯にわたる、すなわち終わりのない学習能力を示唆する。もちろん、段階的な干渉（知識の消去）は不可避であり、望ましくさえある。これがなければ、このようなシステムは遅かれ早かれそのメモリ資源を使い果たしてしまうからである。現実の応用では、環境を制御できることは稀であり、またその特徴について前もって知識を得ることも稀である。このため、システムには新たな要求が生じることになる。すなわち、このシステムは教師無しの適応学習が可能でなければならない、ということであり、これはニューラルネットワークの文献では自己組織化と称されている。

【０００５】

現在のＡＳＲシステムの主たる目標は、所与の音声信号について最も確率の高い単語シーケンスを見出すことである。言い換えれば、興味があるのはその信号が持つ語彙的な情報のみであり、話者が誰であるか（ｉｄｅｎｔｉｔｙ：ＩＤ）、話し方のスタイル、感情的な状態等の存在する他の情報は、信号特性に望ましくない変動を生じる「ノイズ」であると考えられる。これは、このような変動に対し頑健なシステムを要求する。信号の可変性が非静止環境、通信チャンネル、付加的ノイズ等によってももたらされる場合、この課題は特に困難となる。

【０００６】

ＡＳＲシステムの頑健性を改善するために、多くの方法とアルゴリズムとが提案されてきた。しかしながら、依然として、可能な状況の全てにおいて一貫してうまく働くような、この問題に対する効果的な解決策はない。

【０００７】

人間との自然なコミュニケーションが可能な機械を構築するにあたっては、発話の語彙的内容のみでなく、話者（ＩＤ、アクセント、感情）及び環境（オフィス、街路等）の情報も重要となる。現在は、このような情報を得るために、通常は単一のファクタのみ、例えば話者のＩＤ又は発話された言語のみを認識又は特定可能な、別個のシステムが用いられる。この場合、言語学的内容から来る変動性は「不所望」であり、これに対処しなければならない。このような方法は、実務的な観点からはきわめて非効率である。

【０００８】

別の選択肢は、音声信号の変動性を正規化又は減少させる代わりに、これを学習して、語彙的情報だけでなく、興味のある他の何らかの情報とともに、同時に出力するようなシステムを設計することである。このようなシステムは、教師無しのやり方で連続して学習

を行なうことができなければならない。変動性の元となるものすべてについて、前以って知識を得ることは不可能だからである。このこともまた、自己組織化する終わりのない学習システムを持つ、という思想につながる。

【０００９】

人間と機械との学習能力のギャップを埋めるために、多くの研究者が、このようなシステムを設計するための思想の源として人間の能力の研究に目を向けてきた。日常の経験から、人間は生涯を通じて学習が可能であり、新たな知識を獲得しても、先に学習したことの記憶が流し去られるわけではない、ということができる。

【００１０】

人間の脳がどのように働いているかについては、多くが依然としてよく分かっていないが、ニューロンレベルの学習には、ヘッブ則等のいくつかの基本的な原則が定式化されている。ヘッブ則は、シナプス前後のニューロンにおいて同時に発生する活動が、これらのニューロン間の接続を強化するにあたって決定的に重要である、という仮説である。脳の研究によって、神経系はトポロジー的な構造を有することが示された。類似の刺激は脳の中でトポロジー的に近い区域を活性化させるのである。この観察が、いくつかのニューラルネットワークアーキテクチャの開発につながった。

【００１１】

終わりのない、又は生涯にわたる学習の原理は、いわゆる安定性－柔軟性のジレンマを生み出す。システムは如何にして、それまでに学習した知識を保存しながら、新たな事物の学習を続けることができるか、という問題である。この問題については、ニューラルネットワークの研究分野において、以下を含むいくつかの解決策が提案されている。すなわち、適応共振理論（Ａｄａｐｔｉｖｅ　Ｒｅｓｏｎａｎｃｅ　Ｔｈｅｏｒｙ：ＡＲＴ）[非特許文献１]、生涯学習セル構造（Ｌｉｆｅ－ｌｏｎｇ　Ｌｅａｒｎｉｎｇ　Ｃｅｌｌ　Ｓｔｒｕｃｔｕｒｅｓ）[非特許文献２]、及び自己組織化漸次的ニューラルネットワーク（Ｓｅｌｆ－Ｏｒｇａｎｉｚｉｎｇ　Ｉｎｃｒｅｍｅｎｔａｌ　Ｎｅｕｒａｌ　Ｎｅｔｗｏｒｋ）［非特許文献３］である。通常は、新たな知識を受容するために新たなノードを追加することによって保証され、一方で接続の重みによって学習率を減じることによって、必要とされるネットワークの安定性を提供する。

【００１２】

空間－時間パターンの学習と認識とを同時に行ない、これらを思い出すことのできるシステムが、非特許文献４で提案されている。このシステムは自己組織化マップ（Ｓｅｌｆ－Ｏｒｇａｎｉｚｉｎｇ　ｍａｐ：ＳＯＭ）と、同様の有限長さの入力パターンのみをとるＡＲＴネットワークとの組合せである。加えて、入力空間におけるシステムの動作範囲を決定する最初のＳＯＭ層を学習するために、オフラインの前処理ステップが必要とされる。

【００１３】

いわゆるガイド付伝播ネットワーク（Ｇｕｉｄｅｄ　Ｐｒｏｐａｇａｔｉｏｎ　Ｎｅｔｗｏｒｋｓ：ＧＰＮｓ）に基づく、終わりのない学習システムが、非特許文献５に示されている。音声及び自然言語処理を含む、このシステムの様々な考え得る応用が提示されている。

【非特許文献１】Ｇ．カーペンター及びＳ．グロスバーグ、「自己組織化ニューラルネットワークによる適応パターン認識のＡＲＴ」、コンピュータ、７７－８８ページ、１９８８年３月。（G. Carpenter and S. Grossberg, "The ART of adaptive pattern recognition by a self-organizing neural network," Computer, pp. 77-88, Mar. 1988.）

【非特許文献２】Ｆ．ハムカー、「生涯学習セル構造－壊滅的干渉無しの連続した学習」、ニューラルネットワークス、第１４巻、５５１－５７３ページ、２００１年。（F. Hamker, "Life-long learning Cell Structures -continuously learning without catastrophic interference," Neural Networks, vol. 14, pp. 551-573, 2001.）

【非特許文献３】Ｓ．フラオ及びＯ．ハセガワ、「オンラインの教師無し分類及びトポロジー学習のための漸次的ネットワーク、ニューラルネットワークス、第１９巻、９０－１０６ページ、２００６年。（S. Furao and O. Hasegawa, "An incremental network for on-line unsupervised classification and topology learning," Neural Networks, vol. 19, pp. 90-106, 2006.）
【非特許文献４】Ｎ．スリニバサ及びＮ．アージャ、「空間時間パターン学習、認識及び想起のためのトポロジー的時間的相関器ネットワーク、ＩＥＥＥトランザクション、ニューラルネットワーク、第１０巻、第２号、３５６－３７１ページ、１９９９年３月。（N. Srinivasa and N. Ahuja, "A topological and temporal correlator network for spatiotemporal pattern learning, recognition and recall," IEEE Trans. Neural Networks, vol. 10, no. 2, pp. 356-371, Mar. 1999.）
【非特許文献５】Ｄ．ベロー、「時間的符号化に依拠した一致検出アーキテクチャの例」、ＩＥＥＥトランザクション、ニューラルネットワークス、第１５巻、第５号、９６３－９７９ページ、２００４年９月。（D. Beroule, "An instance of coincidence detection architecture relying on temporal coding," IEEE Trans. Neural Networks, vol. 15, no. 5, pp. 963-979, Sept. 2004.）
【非特許文献６】Ｔ．マルチネス及びＫ．シュルテン、「トポロジー表現ネットワーク」ニューラルネットワークス、第７巻、第３号、５０７－５２２ページ、１９９４年。（T. Martinetz and K. Schulten, "Topology representing networks," Neural Networks, vol. 7, no. 3, pp. 507-522, 1994.）

【発明の開示】
【発明が解決しようとする課題】
【００１４】
真正で効果的な終わりのない学習システムは、実時間の適応学習が望ましい広範な分野で用いることができる。このようなシステムが利用可能となれば、人と機械との対話は全く違ったものとなるであろう。残念ながら、先行技術のニューラルネットワークは、音声パターンのような空間－時間データでは動かない。非特許文献４で提案されたシステムは、入力空間におけるシステムの動作範囲を決定する最初のＳＯＭ層を学習するために、オフラインの前処理ステップを必要とする。従って、これは真正の終わりのない学習システムではない。非特許文献５で提案されたＧＰＮシステムは、確証となる実験結果を欠いている。さらに、ＧＰＮの実際的な欠点は、空間－時間的入力データを、２進パターンに変換する必要があるということである。
【００１５】
従って、この発明の目的の一つは、測定値又は観測値の所与の空間的－時間的パターンを実時間で、かつ教師無しで適応学習及し認識する能力を有するシステムを提供することである。
【００１６】
この発明の別の目的は、オフラインのトレーニング無しで、測定値又は観測値の所与の空間的－時間的パターンを同時に学習し認識する能力を有するシステムを提供することである。
【課題を解決するための手段】
【００１７】
この発明の第１の局面は、ネットワークモデルを用いて物理的測定値から導出された特徴ベクトルのシーケンスの空間－時間パターンを同時に学習し認識するためのシステムに関する。特徴ベクトルの各々は予め定められた特徴量空間内に定義されている。ネットワークモデルは、特徴量空間に定義される一組の状態と、状態間の遷移と、状態間の横方向接続とを含む。状態の各々は、出力値の確率密度関数を規定する。遷移の各々は、状態から状態への遷移を規定するとともに遷移の発生頻度と関連付けられている。横方向接続の各々は、隣接する状態の対を規定するとともに、横方向接続によって接続された状態のいずれかがシステムによって最後に経由されてからの経過時間の測定値と関連付けられてい

る。
【００１８】
このシステムは、ネットワークモデルを表すデータセットを記憶するためのモデル記憶手段と、現在の状態の識別子を記憶するための現在状態記憶手段と、新たな特徴ベクトルに応答して、新たな特徴ベクトルに最も良く整合する状態であって、かつ特徴量空間において新たな特徴ベクトルからしきい値距離内にある状態が存在する場合は、それを次の状態と決定し、存在しない場合はネットワークに新たな状態を追加するための手段とを含む。新たな状態は、新たな特徴ベクトルと現在の状態の識別子によって特定される現在の状態とによって規定される。新たな状態は、現在の状態からの次の遷移を規定する。
【００１９】
このシステムはさらに、次の状態が決定されたことに応答して、モデル記憶手段に記憶されたモデルにおける次の遷移の頻度を更新するための手段と、次の状態が決定されたことに応答して、次の状態と、それに隣接する状態との確率密度関数の各々を、予め定められた更新関数によって更新するための手段と、次の状態が決定されたことに応答して、次の状態と、その隣接する状態との接続に関連付けられた経過時間の測定値が予め定められた初期値にリフレッシュされ、かつ他の接続に関連付けられた経過時間の測定値が増分されるように、ネットワークモデル内の横方向接続を更新するための手段と、横方向接続が更新されたことに応答して、予め定められたしきい値より大きい経過時間の測定値と関連付けられている接続を削除するための手段と、接続のいずれかが削除されたことに応答して、何の接続も有していない状態をネットワークモデルから除去するための手段と、現在の状態の識別子を出力状態シーケンスの末尾に追加するための手段と、現在状態記憶手段に記憶された現在の状態の識別子を、次の状態の識別子で置換するための手段とを含む。
【００２０】
決定するための手段は、現在の状態からの遷移を有する状態の組にあって、かつ新たな特徴ベクトルからしきい値距離内にある、新たな特徴ベクトルに最も近い、次の状態の候補を発見するための第１の発見手段と、第１の発見手段が次の状態の候補を発見できなかったことに応答して、現在の状態からの遷移を持たず、新たな特徴ベクトルからしきい値距離内にある、新たな特徴ベクトルに最も近い、次の状態の候補を発見するとともに、ネットワークモデルを、現在の状態から次の状態の候補への新たな遷移が生成されるように更新するための、第２の発見手段と、第１又は第２の発見手段によって次の状態の候補が発見されたことに応答して、次の状態の識別子を次の状態の候補の識別子に設定するための手段と、第１又は第２の発見手段によって次の状態の候補が発見されなかったことに応答して、ネットワークモデルに新たな状態を追加するための手段とを含んでもよく、新たな状態は新たな特徴ベクトルによって規定される確率密度関数と現在の状態からの遷移とを有し、新たな状態の遷移は初期頻度の値と関連付けられている。
【００２１】
横方向接続を更新するための手段は、次の状態と、新たな特徴ベクトルに次に近い状態との間の接続が生成されるようにネットワークモデルを更新するための手段と、次の状態とそれに隣接する状態との間の接続の経過時間の測定値をリフレッシュするための手段と、ネットワークモデル内の接続の経過時間の測定値を増分するための手段とを含んでもよい。
【００２２】
この発明の第２の局面に従ったコンピュータプログラムは、コンピュータ上で実行されると、コンピュータを上述のシステムのいずれかとして機能させる。
【発明を実施するための最良の形態】
【００２３】
［第１の実施の形態］
我々は、終わりのない学習原理を実現し、既存の生涯学習構造の限界を避けようと試みた。そうするにあたって、目標としたのは、自己組織化する、かつトポロジーを表す、終わりのない学習システムであって、発話パターンの持続時間、ダイナミックレンジ又はパ

ラメータ化に何ら制限を課さないシステムを生成することである。

【００２４】

＜ダイナミック隠れマルコフネットワーク＞

１．一般的構造

上述の問題への解決策を求め、さらに最近の神経学的－生物学的研究結果から刺激を受けて、教師無しでオンラインの適応学習が可能であり、一方で、以前に獲得した知識を保存できる、隠れマルコフ状態のネットワークを開発した。発話パターンは、ネットワークを通る状態のシーケンス、すなわち経路として表される。ネットワークは以前に見たことのないパターンを検出することができ、もしこのような新たなパターンに遭遇すると、これは新たな状態と遷移とをネットワークに追加することで学習される。不要なイベント又は「ノイズ」に対応する経路及び状態を経由することは、従って、稀にしかないので、これらは段階的に除去される。従って、ネットワークは必要に応じて成長したり収縮したりする。すなわち、ダイナミック隠れマルコフネットワークはその構造をダイナミックに変化させる。

【００２５】

学習プロセスは、ネットワークが存続する限り、すなわち理論的には永久に続くので、これは終わりのない学習と呼ばれる。発話パターンの認識は、学習と同時に行なわれ、従ってネットワークは常に、単一の学習／認識モードで動作する。

【００２６】

先に説明したとおり、この学習及び認識の新たな枠組に従ったネットワークは隠れマルコフモデル（Ｈｉｄｄｅｎ　Ｍａｒｋｏｖ　Ｍｏｄｅｌ：HMM）を基本とする。これは、測定値又は観測値の入力シーケンスに応じて、その構造をダイナミックに変化させる。従って、これを、ダイナミック隠れマルコフネットワーク（Ｄｙｎａｍｉｃ　Ｈｉｄｄｅｎ　Ｍａｒｋｏｖ　ｎｅｔｗｏｒｋ、略して「ＤＨＭネット」）と呼ぶことにする。

【００２７】

分離して綴った文字からなる小規模データベースでの初期の実験では、ＤＨＭネットは終わりのない学習が可能であることを示し、以前に学習した発話パターンを完璧に認識した。

【００２８】

ＤＨＭネットは自己ループと、それらの間の遷移とを備えた、隠れマルコフ状態を含む。

【００２９】

図１は、簡単な左から右へのHMM構造を概略的に示す。なお、これはＤＨＭネットではない。図１を参照して、このHMMは３個のHMM状態８０、８２、８４を含む。HMM状態８０、８２、８４の各々は他の状態への１又は複数の遷移エッジ９２、９６及び１００と、自己ループ９０、９４及び９８とを有する。各HMM状態の遷移の各々について、遷移確率が割当てられる。同様に、HMM状態８０、８２、８４の各々は、可能な出力値に関する確率分布を有する。

【００３０】

HMMにおいては、モデルの挙動を規定するパラメータ（確率）は不可視であり、不明である。これらのパラメータは統計学的に学習される。

【００３１】

同様に、ＤＨＭネットにおける可能な出力に関する状態遷移の確率と確率分布も、統計学的に学習される。一例を図３に示す。

【００３２】

図３を参照して、ＤＨＭネット１４０はHMM状態１５０、１５２、１５４、１５６、１５８、１６０、１６２及び１６４と、実線の矢印で示す状態間の学習済み経路（状態遷移）とを含む。図３において、HMM状態１６０は削除された状態である。従って、状態１６０と、状態１６０へ／からの遷移２００及び２０２（長い破線矢印で示す。）とは、削除されている。これに対して、HMM状態１６２及び１６４は新たにＤＨＭネット１４

０に追加されたものであり、これらの状態へ／からの遷移２１０、２１２及び２１４（短い破線矢印で示す。）もまた、新たに追加されたものである。

【００３３】

さらに、ＤＨＭネット１４０において、隣接する状態は横方向接続で接続されている。図３において、横方向接続は、矢印でない破線１８０、１８２、１８４、１８６、１８８、１９０、１９２及び１９４で示される。

【００３４】

各状態は多変量ガウス関数によってモデル化された入力特徴量空間の一部を表す。従って、これらの状態はそれぞれ平均ガウスベクトルを有する。ネットワークを通る状態シーケンスすなわち経路は、学習された発話パターン又はパターンのクラスに対応する。これを図２に示す。

【００３５】

図２を参照して、特徴量空間が座標のＸ、Ｙ及びＺ軸で規定されると仮定する。観察された状態は超空間１２０上にある。状態の各々は入力特徴ベクトルによって特定される。例えば、状態１２２は入力ベクトル１２４に対応し、物理的測定値の所与の観察パターンにおいて状態１２２に隣り合う状態１２６は入力ベクトル１２８によって規定される。状態１２２から状態１２６への遷移１３０は入力パターンの経路の一部となる。状態間の遷移を接続することにより、入力パターンに対応する経路が特定される。

【００３６】

他の方法と同様、ＤＨＭネットのネットワークの柔軟性は、新たなパターンに遭遇するたびに新たな状態及び遷移を付加していくことで保証される。

【００３７】

ＤＨＭネットにおける実際的な問題は、何をもって「新たな」パターンと定義し、それをいかにして検出するか、ということである。偽イベントやノイズは、必然的に状態を割当てるが、その経路が再び経由されることはないであろう。このような状態（及び経路）は「死んだ」と考えられ、ネットワークから段階的に除去されるべきものである。

【００３８】

２．「新しさ」の検出

一般に、すでに学習済みのものから十分に異なるパターンはいずれも、新たなパターンと考えることができる。何をもって十分に異なると判断するかに関して、再び、人間の聴覚系の研究に目を向ける。

【００３９】

音圧レベルの変化に対する人間の感受性には限界があることが知られている。多くの心理学的－生物学的研究がこの調査を行なっているが、広帯域のノイズについては、強度の検出可能な最小の変化ΔＩは刺激の強度Ｉにほぼ比例することが分かっている。すなわち、ΔＩ／Ｉは一定である（ウェーバーの法則）。対数の領域では、検出可能な最小変化はΔＬ＝ｌｏｇ（１＋ΔＩ／Ｉ）であり、これは全ての強度値について一定で、約０．２３であると推定される。

【００４０】

発話音声に対してもウェーバーの法則がほぼ当てはまると仮定し、かつＡＳＲシステムフロントエンドで推定される発話スペクトルパワーが発話強度に比例すると仮定すれば、概念的には、同じように「聞こえる」全ての発話パターンはΔL^2に等しい固定された分散を持つガウス関数でモデル化できることになる。従って、対数パワースペクトルが（それまでに学習された全てのパターンを表す）ガウス平均のいずれからもΔＬより遠くにあるパターンはいずれも、新たな、すなわち異なる、発話パターンであると考えられる。このため、ΔＬは新しさを検出する基準として好適である。

【００４１】

しかし、全帯域のパワースペクトルで作業するのは好ましくない。なぜなら、実際のところ、パワースペクトルは、通常であればフィルターバンク（ＦＢ）で推定されるからである。この場合、ΔＬは平均ＦＢパワー差に適用されることになり、これは単一のフィル

タ出力より大きくなる可能性がある。

【００４２】

知覚的な差を生じさせないようなＦＢエネルギ変動の上方の境界を推定するために、以下の実験を行なった。５秒の音声発話を、標準的な前処理手順に従って４８チャンネルのＦＢ対数エネルギベクトルのシーケンスに変換した。その後、平均が０．２３、分散が０．２から３．０の範囲のガウスノイズが特徴ベクトルに付加された。修正されたＦＢエネルギから音声波形を再構築し、これを何人かの被験者に提示して、知覚的評価を行なった。変化に気づいたのは、ガウスノイズの分散が２．０より大きい場合のみであった。

【００４３】

上述の考察に従い、ＤＨＭネット状態確率密度モデルに、固定対角共分散行列を伴う、単一の多変量ガウス関数を選択した。ＤＨＭネットは入力ベクトルが条件付きで独立であると仮定される一次のマルコフ鎖であるので、パターンレベルの新しさの検出は、複数のフレームレベルでの新しさの検出と置換えることができる。従って、所与の入力ベクトルｘはいずれも、もし（ｘ－μ_b）^{2}＞θであれば、「新しい」と考えることができる。ただし、μ_bは最も良く整合する状態の平均であり、θはいわゆるビジランスしきい値である。ここで、「最も良く整合する」状態とは、入力ベクトルに最も近い状態を意味する。

【００４４】

これを図４に概略的に示す。図４を参照して、ＤＨＭネット内に５個のＨＭＭ状態２３２、２３４、２３６、２３８及び２４０があり、新たな特徴ベクトルが与えられたと仮定する。この新たなベクトルは特徴量空間内で新たなデータ点２３０を規定する。もしＨＭＭ状態２３２、２３４、２３６、２３８及び２４０のうちいずれかがこの新たなデータ点２３０からある距離θの範囲内（円２５０で示す）にある場合、この入力データは新しいとは考えられない。図４において、状態点２３４が新たなデータ点と最もよく整合し、かつこれが円２５０内にあるため、この入力パターンは新しいものではないと判断される。

【００４５】

３．安定な学習

「はじめに」の部分で検討した型のニューラルネットワークでは、各学習の繰返しにおいて、重みの更新ΔＷｎは一般に次のように設定される。

【００４６】

【数１】

$$\Delta W_n = \alpha_n \left(X_n - W_{n-1} \right) \qquad (1)$$

ここでX_nは入力ベクトルであり、α_nはｎ回目の繰返しにおける学習率である。安定な学習は、α_nが以下の制約（非特許文献３）に従った場合に保証される。

【００４７】

【数２】

$$\sum_{n=1}^{\infty} \alpha_n = \infty, \qquad \sum_{n=1}^{\infty} \alpha_n^2 < \infty. \qquad (2)$$

ＤＨＭネットの状態確率密度関数（Ｐｒｏｂａｂｉｌｉｔｙ Ｄｅｎｓｉｔｙ Ｆｕｎｃｔｉｏｎ：ＰＤＦ）学習としては、最大尤度推定アルゴリズムをシーケンシャルにしたものを用いる。この場合、入力ベクトルX_nの後のガウス平均更新Δμ_nは以下のようになる。

【００４８】

【数３】

$$\Delta\mu_n = \mu_n - \mu_{n-1}$$
$$= \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{n}{n}\mu_{n-1}$$
$$= \frac{(n-1)\mu_{n-1} + x_n - n\mu_{n-1}}{n}$$
$$= \frac{1}{n}(x_n - \mu_{n-1}), \qquad (3)$$

これは式（１）と全く同じである。学習率はα$_n$＝１／$_n$であり、これは明らかに式（２）の制約を満足している。

【００４９】

４．トポロジーの表現

ＤＨＭネットの状態は、入力特徴量空間の異なる領域を表すため、図２に示すように、隣接する状態が隣接する領域に対応することが重要である。すなわち、状態ネットワークはトポロジーを表すネットワークでなければならない。ニューラルネットワークのノード（ＤＨＭネットの場合は状態）間の横方向接続が、競合ヘッブ則（非特許文献６）を用いて構築される場合、結果として得られるネットワークは完全にトポロジーを表すネットワークである。横方向接続の各々が、特徴量空間におけるトポロジー的に隣接した状態の対を規定している。

【００５０】

競合ヘッブ則は、以下のように説明できる。すなわち、入力ベクトルの各々について、最も近い２個のノードをエッジによって互いに接続する。このようなネットワークは、２つの非常に有用な特性を有する。すなわち、１）入力空間において互いに隣接するベクトルは、互いに隣接するノードによって表される。２）入力空間において２つのベクトル間に経路がある場合、これらのベクトルを表す２個のノードを接続する経路がある筈である。これらの特性はしばしば、隣接性及び経路保存特性と称される。

【００５１】

５．「死んだ」状態の除去

ネットワークがダイナミックにその構造を変化させるとき、状態の隣接性関係もまた変わる。これらの変化に対処するため、横方向接続の各々には年齢が与えられる。これは接続が生成されたか、リフレッシュされた場合にゼロとなる。その他の場合、接続年齢は、接続の状態の一つが経由されるたびに増加する。従って、年齢は、その接続のいずれかの状態をシステムが最後に経由してからの経過時間の測定値として機能する。このようにして、ある年齢に達した接続、すなわち、ある程度の期間にわたってリフレッシュされていないものは、除去される。

【００５２】

ＤＨＭネットは多くの横方向接続を持つことができ、ある状態について、その全ての接続が除去された場合、この状態は「死んだ」と宣言され、その状態に入る遷移、及びその状態から遷移の全てとともに、除去される。

【００５３】

６．復号

特徴ベクトルのシーケンスによって表されるいずれかの入力発話パターンに関して、ネットワークを通る最良の状態シーケンスすなわち経路を発見することが目標である。これは以下のように定式化できる。

【００５４】

【数４】

$$\overline{S} = \max_{S} P\left(S\middle|X\right), \quad X = \left\{x_i\right\}_1^T, \quad S = \left\{s_i\right\}_1^T. \qquad (4)$$

ネットワークの隣接性及び経路保存特性は、所与の現在のベクトルｘ$_t$に対し、現在の状態ｓ$_t$の各々が最良の状態であることを保証する。最良の状態シーケンスは、再帰的な手順を用いて見出すことができる。Ｓ$_t$は時間ｔまでの最良の経路であると仮定する。すると、以下が成り立つ。

【００５５】

【数５】

$$P\left(S_1^{t+1}\middle|X_1^{t+1}\right) = \max_{s_j \subset Succ(s_t)} P\left(s_j S_1^t \middle| x_{t+1} X_1^t\right)$$

$$= \max_{s_j \subset Succ(s_t)} P\left(s_j \middle| S_1^t x_{t+1} X_1^t\right) P\left(S_1^t \middle| x_{t+1} X_1^T\right)$$

$$= \max_{s_j \subset Succ(s_t)} P\left(s_j \middle| s_t x_{t+1}\right) P\left(S_1^t \middle| X_1^t\right)$$

$$= \left[\max_{s_j \subset Succ(s_t)} P\left(s_j \middle| s_t\right) P\left(x_{t+1} \middle| s_j\right)\right] P\left(S_1^t \middle| X_1^t\right) \qquad (5)$$

ここで、Ｓｕｃｃ（ｓ$_t$）は状態ｓ$_t$に後続する状態の集合、すなわち、状態ｓ$_t$から入来する遷移を有する状態の集合である。この集合は（自己ループがあるため）ｓ$_t$自身を含み、さらに、おそらくは新たに追加された状態を含む。上の再帰は、最良の状態シーケンスは、次の入力ベクトルの各々について最良となる次の状態を発見することによって、シーケンシャルなフレーム同期の方法で得られることを示している。

【００５６】

７．認識

ＤＨＭネットでの認識は、復号された最良の状態シーケンスを適切に解釈することによって行なわれる。人間がこの課題を遂行するのと同じやり方で、ネットワーク中の経路が、それらが表すパターンの特性と関連付けられる。最初の近似では、各経路と、それに対応する状態とが、この経路が生成されたか又は再び経由されたときの情報の全てでラベル付けされることを意味する。これは、語彙的内容、話者の情報、環境情報等を含み得る。

【００５７】

音声発話がネットワークに提示されるとき、一般には次の２つの事例が生じうる。１）復号された状態シーケンスが「古い」状態のみからなる場合。これは、全ての発話パターン又はその全てのセグメントがすでに見たことのあるものであって学習済みであることを意味する。この場合、経路と状態のラベルとから、入力発話を認識することができる。２）復号された状態シーケンスが、完全に、又は部分的に、新たに追加された状態からなる場合。この場合、新たな状態の各々について、それに最も近接する状態からラベルを得て、新たな状態をその隣接するものと「同じように聞こえる」と解釈する。

【００５８】

この認識原理は極めて一般的なものであって、大規模な音声認識を可能にするためには、明らかに、別のインテリジェントなシステム、例えばより高度なＤＨＭネット層であって最良の状態シーケンスについて最良の解釈を自動的に発見できるようなものが必要となるであろう。

【００５９】

８．ＤＨＭネットアルゴリズム

完全なＤＨＭネットのアルゴリズムを以下に要約して述べる。

（１）空のネットワークから開始する。

（２）現在の状態をｓ$_{CURR}$として与えられているとき、次の入力ベクトルｘ$_T$について、最も良く整合する後続の状態ｓ$_C$を見つける。もしこれがビジランス試験に合格すれ

ば、これを次の状態として設定して、すなわち$s_{NEXT}=s_C$として、（５）に進む。
（３）他の全ての状態から、最良の状態s_Aを見出す。もしこれがビジランス試験に合格すれば、$s_{NEXT}=s_A$として、（５）に進む。
（４）新たな状態s_Tを末尾に付加する、すなわち$s_{NEXT}=s_T$とし、その平均をx_Tに設定する。
（５）遷移を現在の状態s_{CURR}からs_{NEXT}にする（更新する）。
（６）s_{NEXT}とそれに隣接するもの全ての平均を、式（３）に従って更新する。
（７）s_{NEXT}と次に最良の状態との接続を生成（又はリフレッシュ）する。全てのs_{NEXT}の接続の年齢を増加させる。
（８）いずれかの接続の年齢が年齢しきい値TH_{AGE}に達したら、その接続を除去する。接続のない状態を除去する。
（９）最良の状態シーケンスの末尾にs_{NEXT}を付加する。現在の状態$s_{CURR}=s_{NEXT}$に設定し、（２）に進む。

【００６０】

このアルゴリズムを実現するコンピュータプログラムの制御フローは、図７に関連して後で説明する。

【００６１】

＜音声認識フロントエンドユニットの構造＞

図５は、上述の復号アルゴリズムを組入れた音声認識フロントエンドユニット２６０の機能を示すブロック図である。音声認識フロントエンドユニット２６０はマイクロフォン２６２からのオーディオ信号を受け、ＤＨＭネット音響モデルを構築してこれをトレーニングし、音響モデルを利用して音声信号を復号し、復号された（推定された）状態シーケンスを出力する。音声認識フロントエンドユニット２６０は例えば、より高度な音声認識システムのフロントエンドとして用いることもできる。

【００６２】

図５を参照して、音声認識フロントエンドユニット２６０は、マイクロフォン２６２からのオーディオ信号を採取し、オーディオ信号を、１０ミリ秒のレート、２０ミリ秒のスライド量で移動するウィンドウで、入力オーディオ信号のディジタル形式の音声フレームのストリームに変換する音声キャプチャブロック２８０を含む。

【００６３】

音声認識フロントエンドユニット２６０はさらに、入来する音フレームをウィンドウ処理し、ウィンドウ処理されたフレームにＦＦＴ（Ｆａｓｔ　Ｆｏｕｒｉｅｒ　Ｔｒａｎｓｆｏｒｍ：高速フーリエ変換）を施すＦＦＴブロック２８２と、ＦＦＴブロック２８２の出力を受けるように接続されたＦＢ２８４と、ＦＢ２８４のエネルギビンの各々の対数をとり、特徴ベクトルのシーケンスを出力するための対数関数ブロック２８６とを含む。

【００６４】

音声認識フロントエンドユニット２６０はさらに、特徴ベクトルのシーケンスを受け、ＤＨＭモデルを生成してトレーニングし、ＤＨＭネットモデルを利用して、特徴ベクトルのシーケンスを同時に復号するためのデコーダ２８８と、デコーダ２８８によって生成されトレーニングされたＤＨＭネットモデルを記憶するための記憶部２９０と、ＤＨＭネットのトレーニングに用いる定数TH_{AGE}、TH_{VIGI}及び他の変数を記憶するための記憶部２９２とを含む。TH_{AGE}は横方向接続を削除すべきか否かを判断するために用いられるしきい値であり、TH_{VIGI}は入力ベクトルが特徴量空間において新たな状態を規定するか否かを判断するのに用いられる、図４に示されるビジランスしきい値θである。

【００６５】

デコーダ２８８の出力は、ＤＨＭネットにおけるＨＭＭ状態のシーケンスであり、そのパターンが音声認識に用いられる。

【００６６】

図６はＤＨＭネットで生成される状態の各々に関する状態レコード３００の構造を示す

。状態レコードのデータセットは全体としてＤＨＭネットを定義し、これを表している。
【００６７】
図６を参照して、状態レコード３００は、状態レコード３００を特定する２進値を記憶するための識別子（ＩＤ）フィールド３０２と、この状態からの出力トークンのＰＤＦの平均ベクトルを記憶するための平均ベクトルフィールド３０４と、ＤＨＭネットにおいてこの状態に後続する１又は複数の状態のリンクトリストである、後続状態リスト３０６と、この状態との間で横方向接続を有する１又は複数の状態の、これもまたリンクトリストである隣接状態リスト３０８とを含む。
【００６８】
後続状態リスト３０６は状態識別子項目３２０のリストを含む。状態識別子項目３２０の各々は後続状態のうち１つを特定する後続状態ＩＤフィールド３３０と、状態レコード３００によって規定された状態から後続状態ＩＤフィールド３３０によって特定された状態への遷移の発生頻度を記憶する遷移頻度フィールド３３２とを含む。この頻度は、後続状態への遷移の確率を計算するのに用いることができる。
【００６９】
状態が自己ループを有する場合、状態レコード３００の状態のＩＤ、すなわちＩＤフィールド３０２の値もまた、状態識別項目３２０のうち１つに記憶される。
【００７０】
隣接状態リスト３０８は、横方向接続項目３４０のリストを含む。項目３４０の各々は、この状態と横方向接続を有する状態を特定する隣接状態ＩＤフィールド３５０と、この接続の年齢を記憶するための接続年齢フィールド３５２とを含む。
【００７１】
図７はデコーダ２８８を実現するコンピュータプログラムの制御構造を示す。図７を参照して、このプログラムは、このプログラムで用いられる変数、インデックス、及びデータベース接続を初期化する初期化ステップ３７０と、図５に示される記憶部２９０に空のＤＨＭネットワークを準備するステップとを含む。このプログラムで用いられる変数は、s_{CURR}，s_T，s_A，s_{NEXT}及びs_Cを含み、これらについては全て後述する。変数s_{CURR}は現在の状態のＩＤを示し、ステップ３７２で初期化される。最良の状態シーケンス、すなわち最も確からしい状態のシーケンスもまた、空のリストとしてステップ３７２で準備される。
【００７２】
プログラムはさらに、対数関数ブロック２８６から供給される入力ベクトルＸを読むステップ３７４と、s_{CURR}の状態レコード３００の後続状態リスト３０６に列挙された状態の中から、入力特徴ベクトルＸに最も良く整合する後続状態s_Cを見出すステップ３７６と、ステップ３７６で見出された最も良く整合する後続状態がビジランス試験に合格するか否かを判定し、テストの結果に従って命令実行シーケンスのフローを制御するステップ３７８とを含む。
【００７３】
この実施の形態では、「ビジランス試験に合格する」とは、当該状態と入力ベクトルとの特徴量空間内の距離が、ビジランスしきい値θ（又は「TH_{VIGI}」）に等しいかそれより小さいことを意味する。
【００７４】
プログラムはさらに、ステップ３７８での判断が「ＮＯ」であった場合に実行され、ＤＨＭネットの他の全ての状態から最良の状態s_Aを見出すステップ３８２と、ステップ３８２の後、状態s_Aがビジランス試験に合格するか否かを判断し、試験結果に従って命令実行シーケンスのフローを制御するステップ３８４と、ステップ３８４の結果が「ＮＯ」であった場合に実行され、ＤＨＭネットに新たな状態s_Tを付加する、すなわち状態s_Tの新たなレコード３００を生成するステップ３８８とを含む。
【００７５】
プログラムはさらに、ステップ３８８の後、新たな状態s_Tを次の状態s_{NEXT}とし

て設定するステップ３９０を含む。ここで、新たな状態識別子項目３２０がs_{CURR}の状態レコード３００の後続状態リスト３０６に追加される。後続状態ＩＤフィールド３３０には$s_T = s_{NEXT}$のＩＤが書込まれ、遷移頻度フィールド３３２はゼロに設定される。

【００７６】

プログラムはさらに、ステップ３８０、３８６及び３９０の後に、状態s_{CURR}からs_{NEXT}への遷移を行なうステップ３９２を含み、ここでは状態s_{CURR}のレコード３００の後続状態ＩＤフィールド３３０でs_{NEXT}のＩＤを有する状態識別子項目３２０の遷移頻度フィールド３３２に１が加算される。プログラムはさらに、上述の式（３）により、s_{NEXT}とその全ての隣接する状態との平均を更新するステップ３９３を含む。ステップ３９２において、s_{CURR}からs_{NEXT}への遷移がない場合、状態s_{CURR}の状態レコード３００の後続状態リスト３０６に新たな状態識別子項目３２０が追加される。ここでs_{NEXT}の値（次の状態のＩＤ）が、後続状態ＩＤフィールド３３０内に書込まれ、遷移頻度フィールド３３２はゼロに設定される。

【００７７】

プログラムはさらに、ステップ３７８の判断が「ＹＥＳ」であった場合に実行され、状態s_Cを次の状態s_{NEXT}として設定し、制御をステップ３９２に移すステップ３８０と、ステップ３８４の判断が「ＹＥＳ」であることに応答して、状態s_Aを次の状態s_{NEXT}として設定し、制御をステップ３９２に移すステップ３８６とを含む。

【００７８】

図８を参照して、プログラムはさらに、ステップ３９３に続いて、s_{NEXT}と次に最良の状態との接続をリフレッシュするステップ３９４を含む。すなわち、次に最良の状態と同じＩＤを有するs_{NEXT}の状態レコードの横方向接続項目３４０（図６を参照）の隣接状態リスト３０８において、接続年齢フィールド３５２が「０」にリフレッシュされ、同様に、s_{NEXT}と同じＩＤを有する次に最良の状態の横方向接続項目３４０の隣接状態リスト３０８において、接続年齢フィールド３５２が「０」にリフレッシュされる。もし次に最良の状態とs_{NEXT}との間に接続がない場合には、次に最良の状態とs_{NEXT}との状態レコード３００の各々に新たな横方向接続項目３４０が生成される。ここで、次に最良の状態とs_{NEXT}とのＩＤが、次に最良の状態とs_{NEXT}との状態レコード３００のそれぞれの隣接状態ＩＤフィールド３５０に書込まれる。

【００７９】

プログラムはさらに、ステップ３９４に続いて、s_{NEXT}の全ての横方向接続の年齢を増加させるステップ３９６と、ステップ３９６に続いて、接続年齢のうちしきい値TH_{AGE}と等しいものがある状態レコードが存在するか否かによって条件付きで分岐するステップ３９８と、いずれかの接続年齢＝TH_{AGE}である状態レコードが存在する場合に実行され、その状態レコードの接続を除去するステップ４００と、ステップ４００に続いて、接続無しの状態レコードが存在するか否かによって条件付きで分岐するステップ４０２と、接続無しの状態レコード３００が存在する場合に実行され、その状態レコード３００を、記憶部２９０に記憶されたＤＨＭネットから除去するステップ４０４とを含む。

【００８０】

プログラムはさらに、ステップ４０４に続いて、最良の状態シーケンスの末尾にs_{NEXT}を付加するステップ４０６と、ステップ４０６に続いて、s_{CURR}にs_{NEXT}を設定し、その後図７に示すステップ３７４に進むステップ４０８とを含む。ステップ３９８で、接続年齢がTH_{AGE}と等しい状態レコードがないと判断された場合、又はステップ４０２で接続無しの状態がないと判断された場合には、制御はステップ４０６に進む。

【００８１】

このプログラムでプログラムされたコンピュータにより、図５に示された音声認識フロントエンドユニット２６０のデコーダ２８８が実現される。

【００８２】

＜コンピュータによる実現＞

上述の実施の形態は、コンピュータシステムと、コンピュータシステム上で実行される上記コンピュータプログラムとによって実現できる。図１１はこの実施の形態で用いられるコンピュータシステム４５０の外観を示し、図１２はコンピュータシステム４５０のブロック図である。ここで示されるコンピュータシステム４５０は単なる例示であって、他の構成でも利用可能である。

【００８３】

図１１を参照して、コンピュータシステム４５０は、コンピュータ４６０と、全てコンピュータ４６０に接続された、モニタ４６２、キーボード４６６、スピーカ４５８、マイクロフォン４９０、及びマウス４６８とを含む。コンピュータ４６０はさらに、ＤＶＤ（Ｄｉｇｉｔａｌ　Ｖｅｒｓａｔｉｌｅ　Ｄｉｓｃ：デジタル多用途ディスク）ドライブ４７０とメモリポート４７２とを含む。

【００８４】

図１２を参照して、コンピュータ４６０はさらに、ＤＶＤドライブ４７０とメモリポート４７２とに接続されたバス４８６と、全てバス４８６に接続された、ＣＰＵ（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ：中央処理装置）４７６、コンピュータ４６０のブートアッププログラム等を記憶するＲＯＭ（Ｒｅａｄ　Ｏｎｌｙ　Ｍｅｍｏｒｙ：読出専用メモリ）４７８、ＣＰＵ４７６によって使用される作業領域を提供するとともにＣＰＵ４７６によって実行されるプログラムの記憶領域を提供するＲＡＭ（Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ：ランダムアクセスメモリ）４８０、スピーカ４５８及びマイクロフォン４９０が接続されるサウンドボード４８８、及びハードディスク４７４とを含む。

【００８５】

上述の実施の形態のシステムを実現するソフトウェアは、ＤＶＤ４８２又は着脱可能メモリ４８４等の記憶媒体上に記録されて配布され、ＤＶＤドライブ４７０又はメモリポート４７２等の読出装置を介してコンピュータ４６０に提供され、ハードディスク４７４に記憶される。ＣＰＵ４７６がプログラムの実行を開始すると、プログラムはハードディスク４７４から読出され、ＲＡＭ４８０に記憶される。ＣＰＵ４７６内の図示しないプログラムカウンタによって指定されたアドレスから命令がフェッチされ、命令が実行される。ＣＰＵ４７６は処理対象のデータをハードディスク４７４から読出し、処理の結果をこれもまたハードディスク４７４に記憶する。

【００８６】

コンピュータシステム４５０の一般的動作は周知であるので、ここではその詳細は説明しない。

【００８７】

ソフトウェアの配布の仕方については、これは必ずしもＤＶＤ４８２等の記録媒体上に固定されていなくてもよい。例えば、ソフトウェアはネットワークを介して接続された別のコンピュータから分配されてもよい。ソフトウェアの一部はハードディスク４７４に記憶されてもよく、残りの部分がネットワークを介してハードディスク４７４に入れられ実行の際に統合されてもよい。

【００８８】

典型的には、現代のコンピュータはコンピュータのオペレーティングシステム（ＯＳ）によって提供される一般的な機能を利用し、所望の目的に応じて制御された様態で機能を実行する。従って、ＯＳによって又はサードパーティによって提供されうる一般的な機能を含まないプログラムであって単に一般的機能を実行する命令の組合せのみを指定するプログラムもまた、そのプログラムが全体として所望の目的を達成する制御構造を有する限り、この発明の範囲に含まれることは明らかである。

【００８９】

＜音声認識フロントエンドユニット２６０の動作＞

音声認識フロントエンドユニット２６０は以下のように動作する。話者が１つ又は複数の文章を発話する。音声はマイクロフォン２６２によってアナログ音声信号に変換され、

音声キャプチャブロック２８０に供給される。音声キャプチャブロック２８０は入力音声信号をディジタル形式に変換し、１０ミリ秒のレートで、２０ミリ秒のスライド幅で移動するウィンドウのディジタル音声信号フレームのシーケンスを出力する。

【００９０】

ＦＦＴブロック２８２は供給された音声信号フレームの各々を周波数の領域に変換する。ＦＦＴブロック２８２の出力はＦＢ２８４に供給される。各音声信号フレームについて、ＦＢ２８４は２４ビンの出力スペクトルを出力し、これらは次に対数関数ブロック２８６に与えられて、これらのスペクトルの対数がとられ、それによって特徴ベクトルのシーケンスが出力される。

【００９１】

音声認識フロントエンドユニット２６０の開始時に、デコーダ２８８は記憶部２９２（すなわち図１２のＲＡＭ４８０）を初期化し、初期の空のＤＨＭネットを生成する（図７のステップ３７０及び３７２）。デコーダ２８８はさらに、変数s_{CURR}をヌルに設定し、これは、ＤＨＭネットがこれから構築されるべきことを示す。

【００９２】

－１回目の繰返し－

図７に示されるように、デコーダ２８８はステップ３７４で入力特徴ベクトルを読出す。すなわち、デコーダ２８２は、対数関数ブロック２８６から特徴ベクトルを受け、このベクトルを読込む。

【００９３】

ステップ３７６で、デコーダは最も良く整合する後続の状態s_Cを発見しようとする。開始時にはＤＨＭネットは空なので、最も良く整合する後続状態s_Cは存在しない。この場合、図示しないが、デコーダ２８８は第１のレコードに対し、新たなレコード３００を生成する。すなわち、デコーダ２８８は記憶部２９０に新たな状態レコード３００を生成する。この状態レコード３００のＩＤフィールド３０２には、新たに生成されたＩＤ＝ID_0を入れる。平均ベクトルフィールド３０４には入力された特徴ベクトルが入る。後続状態リスト３０６と隣接状態リスト３０８とは、この新たな状態（この状態s_Cを「s_0」と称する）が遷移を有していないことを意味する値であるヌルに設定される。横方向接続も存在しない。変数s_{CURR}は「s_0」に設定される。制御はステップ３７４に戻る。

【００９４】

－２回目の繰返し－

ステップ３７４で、デコーダ２８８は次の入力特徴ベクトルを読込む。デコーダ２８８は、最も良く整合する後続状態s_Cを発見しようとする。この段階で、ＤＨＭネットには状態が一つ、すなわちs_0しかない。従って、この例では状態s_0がここで発見される。

【００９５】

次に、ステップ３７８で、s_0が新たに入力されたベクトルＸに関しビジランス試験に合格するか否かが判断される。すなわち、状態s_0と入力ベクトルＸとの特徴量空間における距離がビジランスしきい値TH_{VIGI}以下であるか否かが判断される。

【００９６】

－ビジランス試験合格の場合－

状態s_0がビジランス試験に合格した場合、デコーダ２８８はステップ３８０を実行し、ここでs_0が次の状態として設定される。つまり、s_{NEXT}の値にs_0が代入される。これは、遷移が自己ループであることを意味する。

【００９７】

ステップ３９２で、s_0からs_0への遷移がなされる。すなわち、後続状態リスト３０６がヌルであるので、デコーダ２８８はs_0の状態レコード３００に新たな状態識別子項目３２０を生成し、ここで後続状態ＩＤフィールド３３０には「ID_0」（＝状態s_0のＩＤ）が入り、遷移頻度フィールド３３２は０に設定される。s_0の状態レコード３００では、後続状態ｉｄはｉｄ＝「ID_0」である状態識別子項目３２０の遷移頻度フィール

ド３３２に１が加算される。ステップ３９３で、s_0の状態レコード３００の平均が式（３）を用いて更新される。ステップ３９４で、デコーダ２８８はＤＨＭネット内の接続をリフレッシュしようとする。横方向接続がないので、ステップ３９４では何も行なわれない。

【００９８】

ステップ３９６で、デコーダ２８８は全てのs_0の接続の年齢を増加させようとする。s_0には接続がないので、ここでは何も行なわれない。

【００９９】

同様に、ステップ３９８から４０４までも行なわれず、ステップ４０６で、状態s_0を表すＩＤ＝「ID_0」が最良の状態シーケンスの末尾に添付される。こうして、最良の状態シーケンスは、｛ID_0　ID_0｝となる。

【０１００】

ステップ４０８で、s_{CURR}に再びs_0が設定され、制御はステップ３７４（図７）に戻る。

【０１０１】

－ビジランス試験に不合格の場合－

状態s_0がステップ３７８のビジランス試験に合格しない場合、入力ベクトルは状態s_0から十分異なるので、「新しい」と考えられる。ステップ３８２で、デコーダ２８８はＤＨＭネット内の他の全ての状態から、最良の状態s_Aを発見しようとする。動作のこの段階では、s_0以外の状態はないので、ステップ３８４での判断は「ＮＯ」となり、ステップ３８８で、デコーダ２８８はＤＨＭネットに新たな状態s_1を追加する。

【０１０２】

すなわち、状態s_1について新たな状態レコード３００が生成され、ここでＩＤフィールド３０２にはこの状態のための新たに生成されたＩＤであるID_1が入り、平均ベクトルフィールド３０４には第２の繰返しのステップ３７４で得られたベクトルが入り、後続状態リスト３０６及び隣接状態リスト３０８はヌルに設定される。ステップ３９０で、デコーダ２８８はs_1を次の状態として設定する。すなわち、s_{NEXT}の値はs_1に設定される。

【０１０３】

ステップ３９２で、s_0からs_1への遷移がなされる。s_0の状態レコード３００にはs_0からs_1への遷移がないので、s_0の状態レコード３００内の後続状態リスト３０６に新たな状態識別子項目３２０が追加され、後続状態ＩＤフィールド３３０にはID_1（＝状態s_1のＩＤ）が入り、遷移頻度フィールド３３２は０に設定される。その後、遷移頻度フィールド３３２に１が加算される。

【０１０４】

ステップ３９４で、デコーダ２８８はＤＨＭネット内の接続をリフレッシュする。すなわち、s_0とs_1の状態レコード３００にはs_0とs_1との間の接続項目がないので、s_0とs_1の状態レコード３００の各々の隣接状態リスト３０８において新たな横方向接続項目３４０が生成され、それぞれの隣接状態ＩＤフィールド３５０にはID_1及びID_0がそれぞれ入る。その後、s_1の状態レコード３００の隣接状態ＩＤフィールド３５０において「ID_0」を有する横方向接続項目３４０（すなわち、s_0とs_1との接続）の接続年齢フィールドがゼロにリフレッシュされる。同様に、s_0の状態レコード３００の隣接状態ＩＤフィールド３５０において「ID_1」を有する横方向接続項目３４０の接続年齢フィールドがゼロにリフレッシュされる。こうして、s_0とs_1との接続の接続年齢がゼロにリフレッシュされる。

【０１０５】

ステップ３９６で、デコーダ２８８がs_0とs_1との接続の年齢を１だけ増加させる。TH_{AGE}が１より大きいと仮定して、ステップ３９８、４００、４０２又は４０４では何も行なわれない。新たに生成された状態s_1がステップ４０６で最良の状態シーケンスに付加される。すなわち、s_1のＩＤ（＝ID_1）が最良の状態シーケンスの末尾に付加

される。従って、最良の状態シーケンスは{ID_0 ID_1}となる。ステップ４０８で変数s_{CURR}の値はs_1に設定され、制御はステップ３７４（図７）に戻る。

【０１０６】

－第３及びそれ以降の繰返し－

第２の繰返しの後、ステップ３７４でデコーダ２８８によって読込まれたベクトルの各々について、デコーダ２８８はステップ３７６のs_{CURR}に後続する状態の内で最も良く整合する状態を発見しようとする。このような状態があり、かつその状態がビジランス試験に合格すれば、この状態が次の状態に設定される。このような状態がなければ、ステップ３８２で、他の状態の中から最も良く整合する状態を発見する。このような状態が存在し、その状態がビジランス試験に合格すれば、その状態が次の状態に設定される。そのような状態がなければ、新たな状態と、現在の状態からその新たな状態への遷移とが、ステップ３８８と３９２とでそれぞれ生成される。

【０１０７】

ある状態を経由するごとに、対応する遷移の頻度がステップ３９２で１づつ増分される。ある状態から出る遷移全ての頻度を用いれば、その状態の各々の遷移の遷移確率を計算することができる。

【０１０８】

ある状態を再経由するか、新たな状態が生成されるたびに、その状態と隣接する状態との接続がゼロにリフレッシュされ、他の接続の年齢は１だけ増分される。年齢がTH_{AGE}と等しい接続があれば、その接続はステップ４００で削除される。従って、稀にしか再経由されない状態の接続は、時間がたてば削除される。ある状態の接続全てが削除されると、その状態はそれに関連する遷移とともにＤＨＭネットから除去される。従って、偽イベント又は「ノイズ」に相当する経路や状態は段階的に除去される。

【０１０９】

この結果、ネットワークは必要に応じて成長したり収縮したりする。言換えれば、ネットワークはダイナミックにその構造を変える。

【０１１０】

典型的には、ＤＨＭネットは高速動作のため、状態レコード３００の集合の形でＲＡＭ４８０に記憶されることになる。しかし、音声認識フロントエンドユニット２６０がシャットダウンされる前に、ＤＨＭネットをハードディスク等の不揮発性記憶装置に保存してもよい。音声認識フロントエンドユニット２６０がその動作を再開する場合、ハードディスクから状態レコード３００を読出し、ＲＡＭにロードしても良い。この場合、音声認識フロントエンドユニット２６０はＤＨＭネットを何もないところから作成する必要がない。当業者には容易に理解されるように、このようにしてトレーニングされたＤＨＭネットを他のシステムに移植することもできる。

【０１１１】

＜実験＞

ＤＨＭネット等の終わりのない学習システムにとって、入手可能なデータをトレーニング、開発及びテスト、モデルトレーニング、チューニング及びテスト、に分割するという、伝統的な評価手法はあまり意味を成さない。

【０１１２】

実験のために、日本人の話者２０名（男性１０名、女性１０名）が発話した、２２の英語の文字の単一のサンプルからなる、スペルされた文字の発話の小規模データベースを選択した。合計発話数は４４０であった。発話の各々は、１０－ｍｓのレートで２０－ｍｓのスライドウィンドウで計算した２４個の対数フィルタバンクエネルギからなる特徴ベクトルのシーケンスに変換された。全てのＤＨＭネットの状態の共分散が単位行列に設定された。すなわちビジランスしきい値θ＝１．０に設定された。

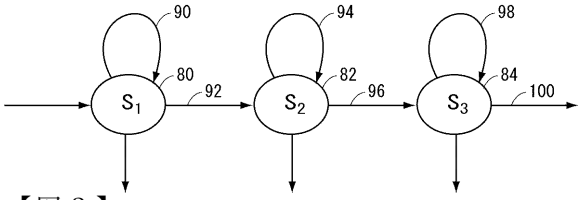
【０１１３】

第１の実験では、ネットワークの学習能力をテストした。全てのデータを用いた学習が２０回繰返された。図９は観察されたデータ尤度の変化を示す。図９を参照して、増加し

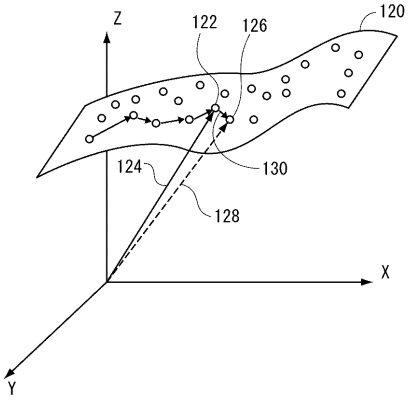
ている飽和曲線が、ＤＨＭネットは安定した学習が可能であることを明確に示している。
【０１１４】
次に、ネットワークが以前に学習した知識を忘れることなく新たな事柄を学習できるかを確認するために、以下の実験を行なった。始めに、「ＭＡＵ」という文字列で識別されるある話者のみによる学習の繰返しを１０回行なった。その後、次の１０回の繰返しに、別の話者によるデータ（「ＭＭＳ」という文字列で識別される。）を用いた。その後、ＭＡＵからのデータをさらに１０回繰返してネットワークに与えた。最後に、同じ手順をＭＭＳのデータでも繰返した。
【０１１５】
図１０は、このような学習の間の、データ尤度を示す。図１０を参照して、データがそれまでに見たのことのあるパターンに変わる２０回目と３０回目の繰返しで、尤度はそれらを最後に見たときの点からの上昇を続けた。これは、異なる話者のデータによる学習も、以前に記憶した知識を破壊しないこと、すなわち、ネットワークが終わりのない学習を可能とするものであることを意味する。
【０１１６】
最後の実験は、学習の繰返しごとに、ネットワークの認識能力を確認するために設計された。発話の各々について、デコードされた状態シーケンスが記憶され、話者と文字ＩＤでラベル付けされた。各学習の繰返しごとに、得られた状態シーケンスを先行する繰返しからのものと比較して、最も良く整合するシーケンスを発見した。ラベルが一致すれば、ヒットであると考えられた。
【０１１７】
わずか２回の繰返しで、認識率は９７．４４％となり、３回目以降の繰返しでは、１００％となった。これは、全く誤りなしに同時に音声及び話者の認識がされたことを意味する。
【０１１８】
上述の説明から理解されるように、ＤＨＭネットを利用したシステムは、現在の音声モデルとは対照的に、壊滅的忘却なしで、終わりのない、教師無しの適応学習が可能である。このネットワークを、同じ学習原理に従って構築されたフルスケールの音声認識用の階層的システムの最初の前処理層として利用することができる。上記したＤＨＭネットは単一の学習／認識モードで動作するが、これは、所与の経路に沿った状態のＰＤＦからのサンプリングにより、対応する音声パターンを再構築するような、パターンを再現（ｒｅｃａｌｌ）するモードに容易に拡張可能である。このような２つのモードを有するＤＨＭネットは、音声認識のみならず、音声合成、音声変換、音声強調等に用いることができる。
【０１１９】
今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味および範囲内でのすべての変更を含む。
【図面の簡単な説明】
【０１２０】
【図１】ＨＭＭの構造を概略的に示す図である。
【図２】特徴ベクトルによって規定される特徴量空間を概略的に示す図である。
【図３】ダイナミック隠れマルコフネットワークの概略構造を示す図である。
【図４】入力特徴ベクトルがどのようにして「新しい」と判断されるかを概略的に示す図である。
【図５】この発明の一実施の形態に従った音声認識フロントエンドユニット２６０の機能的ブロック図である。
【図６】状態レコード３００の構造を示す図である。
【図７】図５に示したデコーダ２８８を実現するプログラムのフローチャートの前半を示す図である。

【図８】デコーダ２８８を実現するプログラムのフローチャートの後半を示す図である。
【図９】２０回の繰返し学習の間の尤度の変化を示すグラフである。
【図１０】交互の話者によるデータ学習の間の尤度の変化を示すグラフである。
【図１１】コンピュータシステム４５０の外観を示す図である。
【図１２】コンピュータシステム４５０の構造を示すブロック図である。
【符号の説明】
【０１２１】
８０、８２、８４、１５０、１５２、１５４、１５６、１５８、１６０、１６２、１６４　ＨＭＭ状態
１４０　ＤＨＭネット
１８０、１８２、１８４、１８６、１８８、１９０、１９２、１９４　横方向接続
２６０　音声認識フロントエンドユニット
２６２　マイクロフォン
２８０　音声キャプチャブロック
２８２　ＦＦＴブロック
２８４　フィルタバンク（ＦＢ）
２８６　対数関数ブロック
２８８　デコーダ
２９０及び２９２　記憶部
３００　状態レコード
３０４　平均ベクトルフィールド
３０６　後続状態リスト
３３０　後続状態ＩＤフィールド
３０８　隣接状態リスト
３３２　遷移頻度フィールド
３５０　隣接状態ＩＤフィールド
３５２　接続年齢フィールド

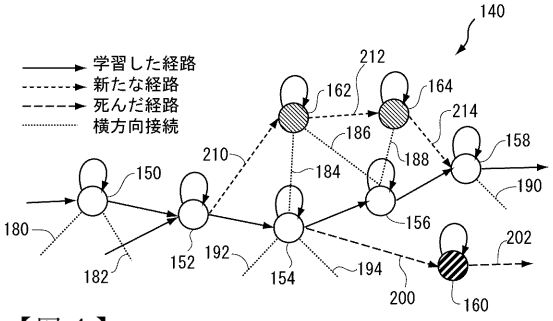
【図１】



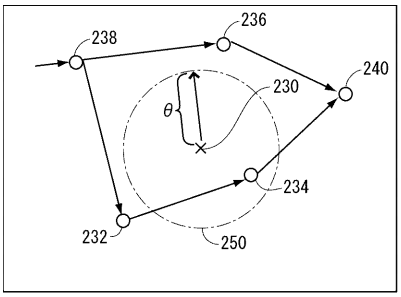
【図２】



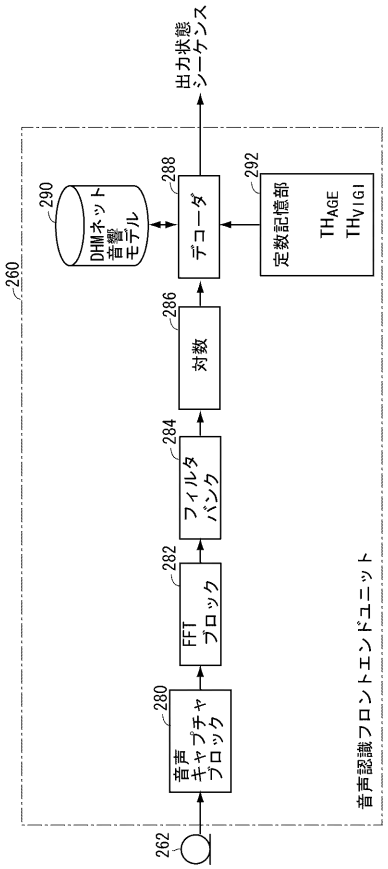
【図３】



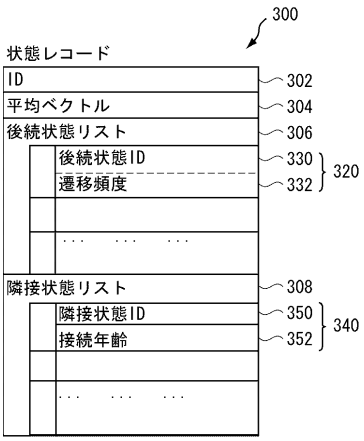
【図４】



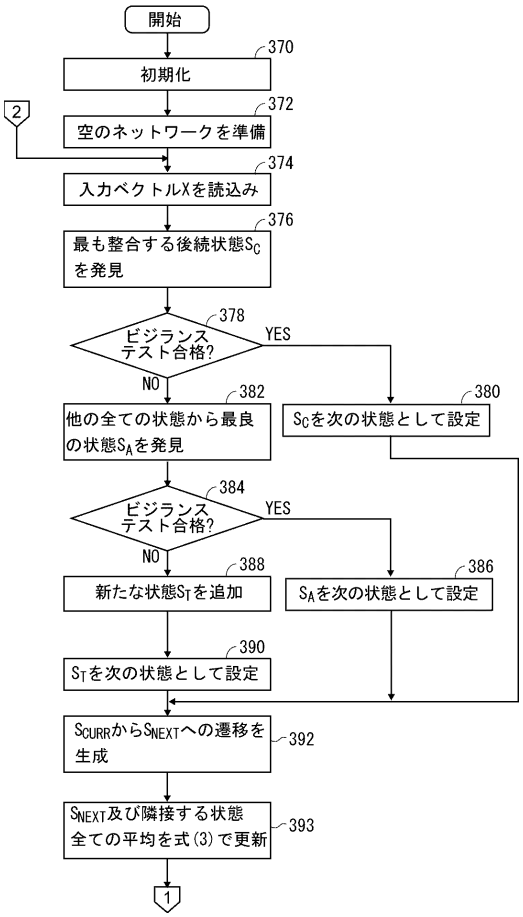
【図５】



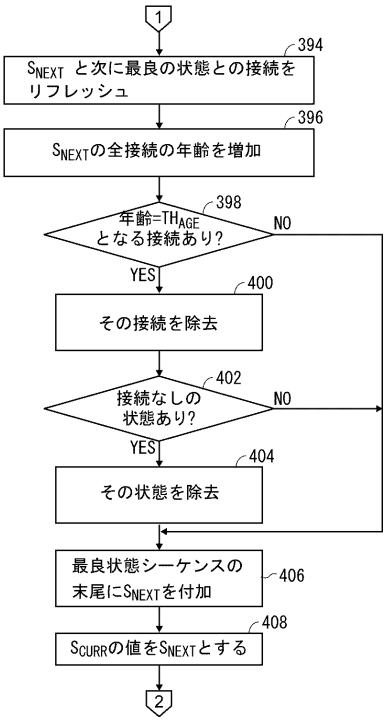
【図６】



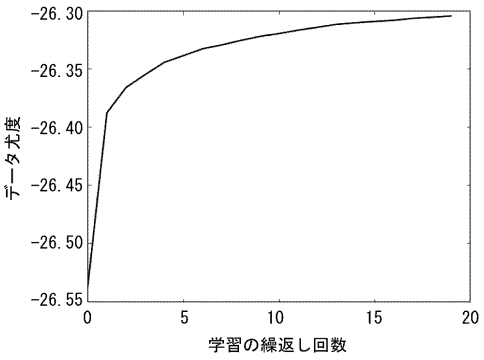
【図７】



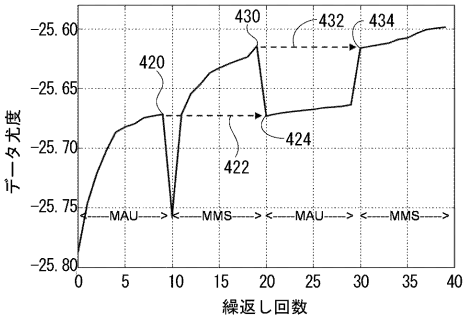
【図８】



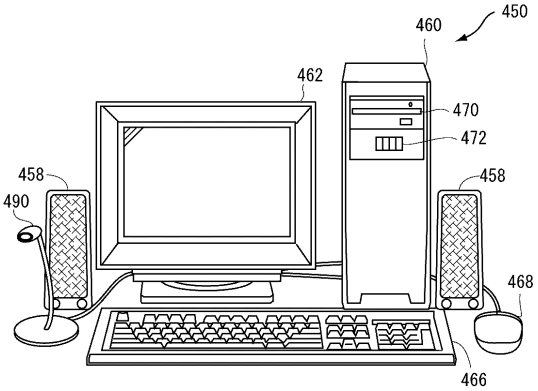
【図９】



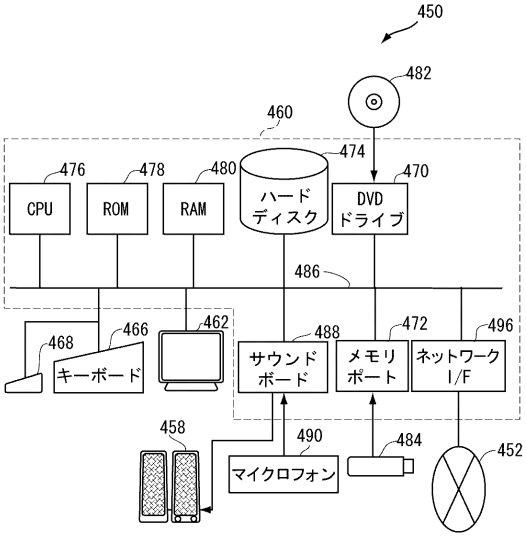
【図１０】



【図１１】



【図１２】


450
482
460
474
476
478
480
470
CPU
ROM
RAM
ハード
ディスク
DVD
ドライブ
486
462
466
468
488
472
496
キーボード
サウンド
ボード
メモリ
ポート
ネットワーク
I/F
458
490
484
452
マイクロフォン

フロントページの続き

(56)参考文献 特開２００６－２９３４８９（ＪＰ，Ａ）
冨田 仁志，他１名，隠れマルコフモデルの最良パラメータ推定，電子情報通信学会技術研究報告，社団法人電子情報通信学会，２０００年 3月１４日，第９９巻，第６８５号，ｐ．１０５－１１２
末永 寛，他１名，大きさ可変の競合層を用いた自己組織化マップ，電子情報通信学会技術研究報告，社団法人電子情報通信学会，２０００年 3月１４日，第９９巻，第６８５号，ｐ．１２９－１３６
Da Deng, et al., ESOM: An Algorithm to Evolve Self-Organizing Maps from On-line Data Streams, Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on, ２０００年, vol.6, p.3-8

(58)調査した分野(Int.Cl., ＤＢ名)
Ｇ０６Ｎ 3／００ － 3／１２
Ｇ１０Ｌ １５／００ － １５／２８
ＩＥＥＥ Ｘｐｌｏｒｅ