

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2008-186171  
(P2008-186171A)

(43) 公開日 平成20年8月14日(2008.8.14)

(51) Int.Cl.	F I	テーマコード (参考)
<b>G06N 3/00 (2006.01)</b>	G06N 3/00 560G	5D015
<b>G10L 15/06 (2006.01)</b>	G10L 15/06 310Z	
<b>G10L 15/18 (2006.01)</b>	G10L 15/18 200E	
<b>G10L 15/14 (2006.01)</b>	G10L 15/14 200Z	

審査請求 未請求 請求項の数 4 O L 外国語出願 (全 25 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)

(71) 出願人 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  
Fターム(参考) 5D015 HH23 JJ07

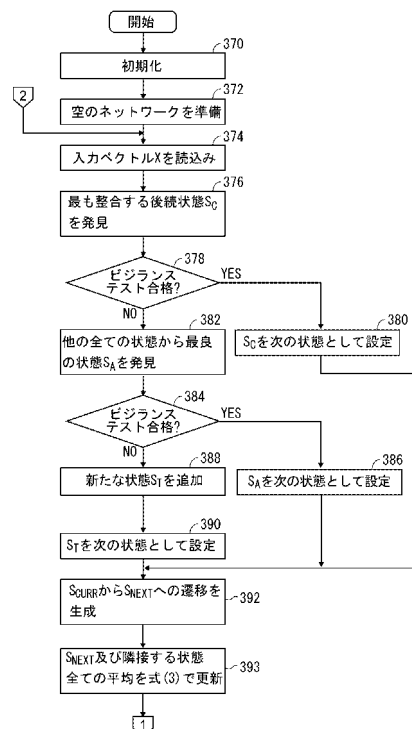
(54) 【発明の名称】 空間-時間パターンを同時に学習し認識するためのシステム

(57) 【要約】

【課題】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びし認識する能力を有するシステムを提供する。

【解決手段】システムは、ダイナミック隠れマルコフモデルネットワーク(DHMネット)モデルを表すデータの集合を記憶するモデル記憶部と、現在の状態の変数  $s_{CURR}$  と、入力特徴ベクトル  $X$  に最も良く整合し、特徴量空間において新たな特徴ベクトルからの距離がしきい値より小さい状態が存在すればこれを次の状態と判断し(378 - 386)、存在しない場合はネットワークに新たな状態を付加する(388 - 390)ユニットを含む。新たな状態は新たな特徴ベクトル  $X$  と、現在の状態  $s_{CURR}$  によって定義され、状態  $s_{CURR}$  と  $s_{NEXT}$  との間で次の遷移が定義される(392)。

【選択図】 図7



## 【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】

前記決定するための手段は、

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

10

20

30

40

50

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、統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

10

20

30

40

50

報のみであり、話者が誰であるか ( i d e n t i t y : I D )、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。

【 0 0 0 6 】

A S R 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。

【 0 0 0 7 】

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 ( I D、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 I D 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。

【 0 0 0 8 】

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。

【 0 0 0 9 】

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。

【 0 0 1 0 】

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、ヘップ則等のいくつかの基本的な原則が定式化されている。ヘップ則は、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。

【 0 0 1 1 】

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 ( A d a p t i v e R e s o n a n c e T h e o r y : A R T ) [ 非特許文献 1 ]、生涯学習セル構造 ( L i f e - l o n g L e a r n i n g C e l l S t r u c t u r e s ) [ 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 ( S e l f - O r g a n i z i n g I n c r e m e n t a l N e u r a l N e t w o r k ) [ 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。

【 0 0 1 2 】

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

10

20

30

40

50

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。

10

【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 - 壊滅的干渉無しに連続した学習」、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.)

20

【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-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.)

30

【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. Martinez and K. Schulten, "Topology representing networks," Neural Networks, vol. 7, no. 3, pp. 507-522, 1994.)

40

【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

50

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び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。

10

【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。

20

【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回答して、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回答して、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回答して、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回答して、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回答して、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回答して、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。

30

40

【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回答して、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

50

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。

10

【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。

20

【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。

30

【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、学習と同時に進められ、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。

40

【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

50

。

## 【 0 0 2 9 】

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 8 0、8 2、8 4 を含む。HMM 状態 8 0、8 2、8 4 の各々は他の状態への 1 又は複数の遷移エッジ 9 2、9 6 及び 1 0 0 と、自己ループ 9 0、9 4 及び 9 8 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 8 0、8 2、8 4 の各々は、可能な出力値に関する確率分布を有する。

## 【 0 0 3 0 】

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。

10

## 【 0 0 3 1 】

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。

## 【 0 0 3 2 】

図 3 を参照して、DHM ネット 1 4 0 は HMM 状態 1 5 0、1 5 2、1 5 4、1 5 6、1 5 8、1 6 0、1 6 2 及び 1 6 4 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 1 6 0 は削除された状態である。従って、状態 1 6 0 と、状態 1 6 0 へ / からの遷移 2 0 0 及び 2 0 2（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 1 6 2 及び 1 6 4 は新たに DHM ネット 1 4 0 に追加されたものであり、これらの状態へ / からの遷移 2 1 0、2 1 2 及び 2 1 4（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。

20

## 【 0 0 3 3 】

さらに、DHM ネット 1 4 0 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 1 8 0、1 8 2、1 8 4、1 8 6、1 8 8、1 9 0、1 9 2 及び 1 9 4 で示される。

## 【 0 0 3 4 】

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。

30

## 【 0 0 3 5 】

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 1 2 0 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 1 2 2 は入力ベクトル 1 2 4 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 1 2 2 に隣り合う状態 1 2 6 は入力ベクトル 1 2 8 によって規定される。状態 1 2 2 から状態 1 2 6 への遷移 1 3 0 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。

## 【 0 0 3 6 】

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。

40

## 【 0 0 3 7 】

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。

## 【 0 0 3 8 】

## 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かと判断するかに関して、再び、人間の聴

50



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  $L$  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、 $L/I$  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  $L = \log(1 + I/I)$  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  $L^2$  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  $L$  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、 $L$  は新しさを検出する基準として好適である。

10

【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、 $L$  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。

20

【0042】

知覚的な差を生じさせないような FB エネルギー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 対数エネルギー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 エネルギー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  $(x - \mu_b)^2 >$  であれば、「新しい」と考えることができる。ただし、 $\mu_b$  は最も良く整合する状態の平均であり、 $L$  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。

30

【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  $L$  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。

40

【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  $W_n$  は一般に次のように設定される。

【0046】

【数 1】

$$\Delta W_n = \alpha_n (X_n - W_{n-1}) \quad (1)$$

ここで  $X_n$  は入力ベクトルであり、 $\alpha_n$  は  $n$  回目の繰返しにおける学習率である。安定な学習は、 $\alpha_n$  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】

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

10

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  $\mu_n$  は以下のようになる。

【0048】

【数 3】

$$\begin{aligned} \Delta \mu_n &= \mu_n - \mu_{n-1} \\ &= \frac{1}{n} \sum_{i=1}^n x_i - \frac{n}{n} \mu_{n-1} \\ &= \frac{(n-1)\mu_{n-1} + x_n - n\mu_{n-1}}{n} \\ &= \frac{1}{n} (x_n - \mu_{n-1}), \end{aligned} \quad (3)$$

20

これは式（1）と全く同じである。学習率は  $\alpha_n = 1/n$  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

30

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、競合ヘップ則（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】

競合ヘップ則は、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。

40

【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

50

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。

【 0 0 5 2 】

D H M 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。

【 0 0 5 3 】

#### 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。

【 0 0 5 4 】

【 数 4 】

$$\bar{S} = \max_S P(S|X), \quad X = \{x_i\}_1^T, \quad S = \{s_i\}_1^T. \quad (4)$$

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。

【 0 0 5 5 】

【 数 5 】

$$\begin{aligned} P(S_1^{t+1}|X_1^{t+1}) &= \max_{s_j \in Succ(s_t)} P(s_j S_1^t | x_{t+1} X_1^t) \\ &= \max_{s_j \in Succ(s_t)} P(s_j | S_1^t x_{t+1} X_1^t) P(S_1^t | x_{t+1} X_1^t) \\ &= \max_{s_j \in Succ(s_t)} P(s_j | s_t x_{t+1}) P(S_1^t | X_1^t) \\ &= \left[ \max_{s_j \in Succ(s_t)} P(s_j | s_t) P(x_{t+1} | s_j) \right] P(S_1^t | X_1^t) \quad (5) \end{aligned}$$

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。

【 0 0 5 6 】

#### 7 . 認識

D H M 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。

【 0 0 5 7 】

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

10

20

30

40

50

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み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マイクロ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のエネルギーピンの各々の対数をとる、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、

D H Mモデルを生成してトレーニングし、D H M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によって生成されたD H Mネットモデルを記憶するための記憶部290と、D H Mネットのトレーニングに用いる定数 $T H_{A G E}$ 、 $T H_{V I G I}$ 及び他の変数を記憶するための記憶部292とを含む。 $T H_{A G E}$ 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、 $T H_{V I G I}$ 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である。

【0065】

デコーダ288の出力は、D H MネットにおけるH M M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。

10

【0066】

図6はD H M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D H M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(I D)フィールド302と、この状態からの出力トークンのP D F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、D H M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、これもまたリンクリストである隣接状態リスト308とを含む。

20

【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I D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I D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I D、すなわちI D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。

30

【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I D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D H M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、 $S_{C U R R}$ 、 $S_T$ 、 $S_A$ 、 $S_{N E X T}$ 及び $S_C$ を含み、これらについては全て後述する。変数 $S_{C U R R}$ は現在の状態のI D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。

40

【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、 $S_{C U R R}$ 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。

50

## 【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（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。

10

## 【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、388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。

20

## 【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。

30

## 【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。

40

## 【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40

50

2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\_N\_E\_X\_Tを付加するステップ406と、ステップ406に続いて、S\_C\_U\_R\_RにS\_N\_E\_X\_T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T\_H\_A\_G\_E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。

10

【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、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。

20

【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、及びハードディスク474とを含む。

30

【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。

40

【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。

50

## 【 0 0 8 8 】

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状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。

## 【 0 0 8 9 】

< 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ジタル形式に変換し、10 ミリ秒のレートで、20 マイクロ秒のスライド幅で移動するウィンドウのデジタル音声信号フレームのシーケンスを出力する。

10

## 【 0 0 9 0 】

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 FB 284 に供給される。各音声信号フレームについて、FB 284 は 24 ピ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。

20

## 【 0 0 9 1 】

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 RAM 480）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。

## 【 0 0 9 2 】

- 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み出す。すなわち、デコーダ 288 は、対数関数ブロック 286 から特徴ベクトルを受け、このベクトルを読み込む。

30

## 【 0 0 9 3 】

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。

40

## 【 0 0 9 4 】

- 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み込む。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。

## 【 0 0 9 5 】

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。

50



## 【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<sub>0</sub>」 (= 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 10  
では、後続状態 id は  $id = \text{「ID}_0\text{」}$  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 20  
表す  $ID = \text{「ID}_0\text{」}$  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 30  
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 40  
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 40  
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 50  
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<sub>0</sub>」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<sub>1</sub>」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。THAG<sub>E</sub>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<sub>1</sub>)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<sub>0</sub> ID<sub>1</sub>}となる。ステップ408で変数SCURRの値は $s_1$ に設定され、制御はステップ374(図7)に戻る。

10

#### 【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SCURRに後続する状態の中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。

20

#### 【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THAG<sub>E</sub>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。

30

#### 【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を読み出し、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。

40

#### 【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

50

、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ー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 = 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。

10

【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が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。

20

【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。

30

【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。

40

【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のすべての変更を含む。

【図面の簡単な説明】

50

## 【 0 1 2 0 】

【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 2 6 0 の機能的ブロック図である。

【図 6】状態レコード 3 0 0 の構造を示す図である。

【図 7】図 5 に示したデコーダ 2 8 8 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。 10

【図 8】デコーダ 2 8 8 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】2 0 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 1 0】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 1 1】コンピュータシステム 4 5 0 の外観を示す図である。

【図 1 2】コンピュータシステム 4 5 0 の構造を示すブロック図である。

## 【符号の説明】

## 【 0 1 2 1 】

8 0、8 2、8 4、1 5 0、1 5 2、1 5 4、1 5 6、1 5 8、1 6 0、1 6 2、1 6 4  
HMM 状態 20

1 4 0 DHM ネット

1 8 0、1 8 2、1 8 4、1 8 6、1 8 8、1 9 0、1 9 2、1 9 4 横方向接続

2 6 0 音声認識フロントエンドユニット

2 6 2 マイクロフォン

2 8 0 音声キャプチャブロック

2 8 2 FFT ブロック

2 8 4 フィルタバンク (FB)

2 8 6 対数関数ブロック

2 8 8 デコーダ

2 9 0 及び 2 9 2 記憶部 30

3 0 0 状態レコード

3 0 4 平均ベクトルフィールド

3 0 6 後続状態リスト

3 3 0 後続状態 ID フィールド

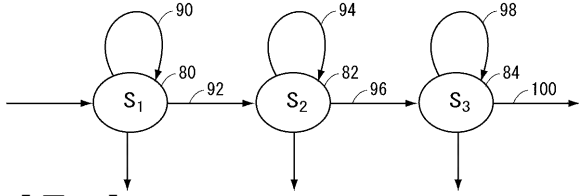
3 0 8 隣接状態リスト

3 3 2 遷移頻度フィールド

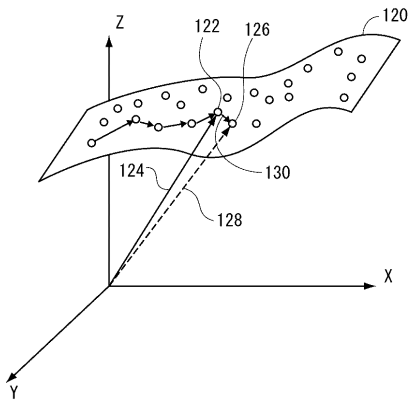
3 5 0 隣接状態 ID フィールド

3 5 2 接続年齢フィールド

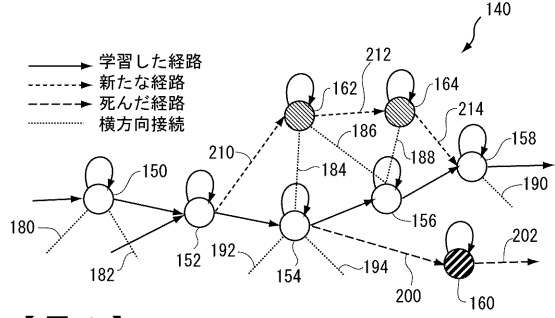
【図1】



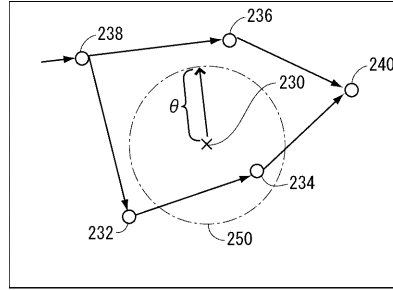
【図2】



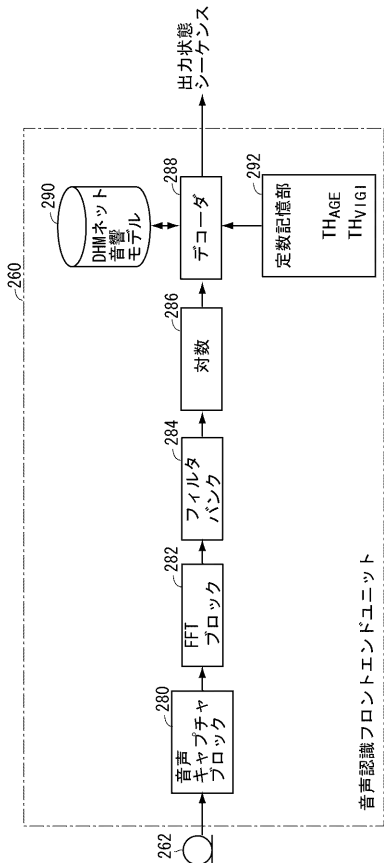
【図3】



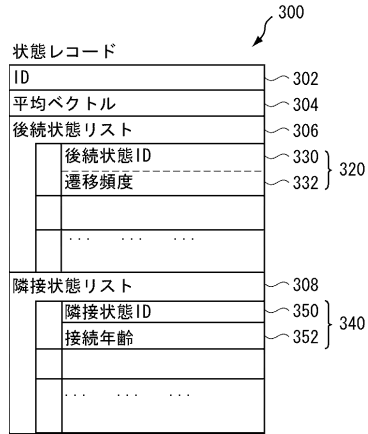
【図4】



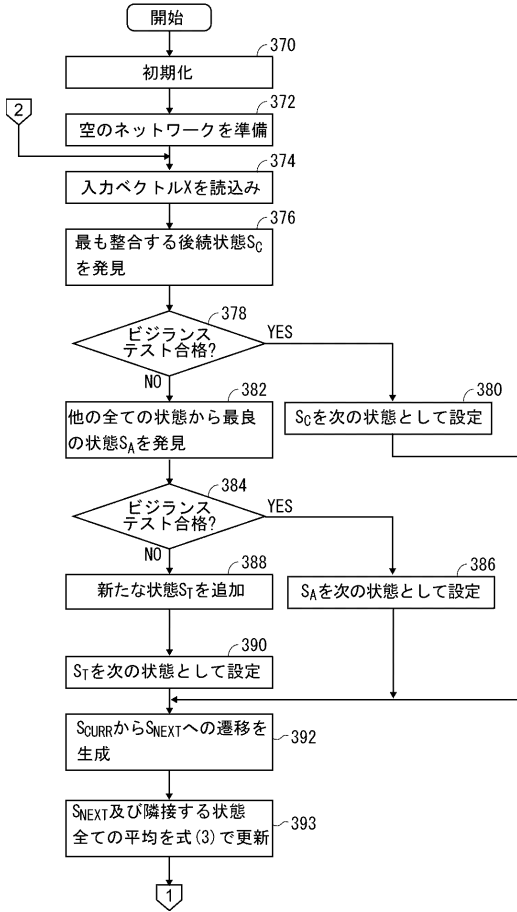
【図5】



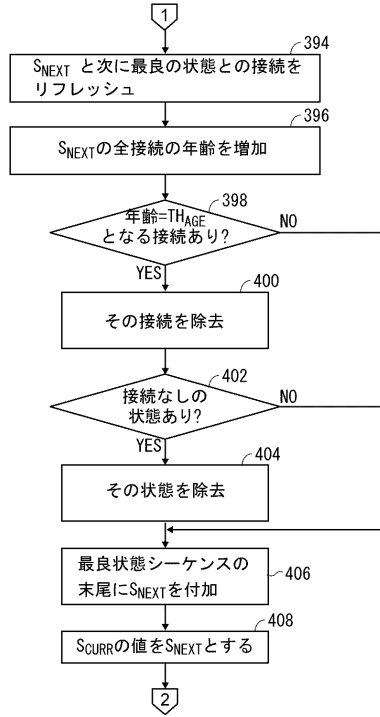
【図6】



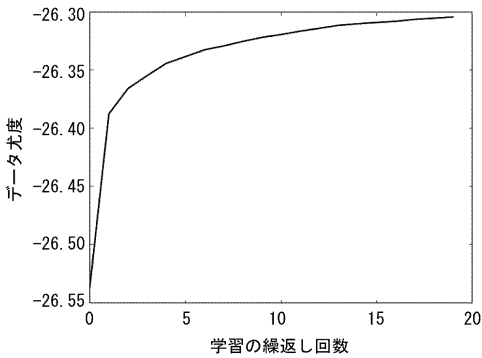
【 図 7 】



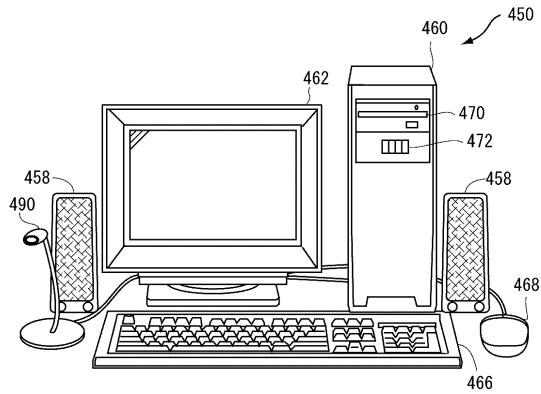
【 図 8 】



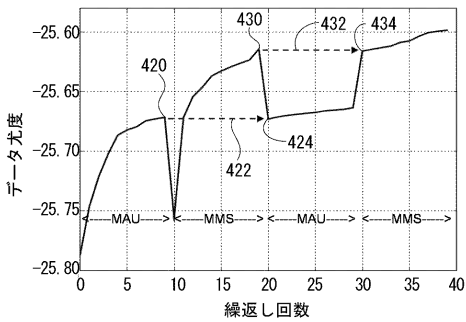
【 図 9 】



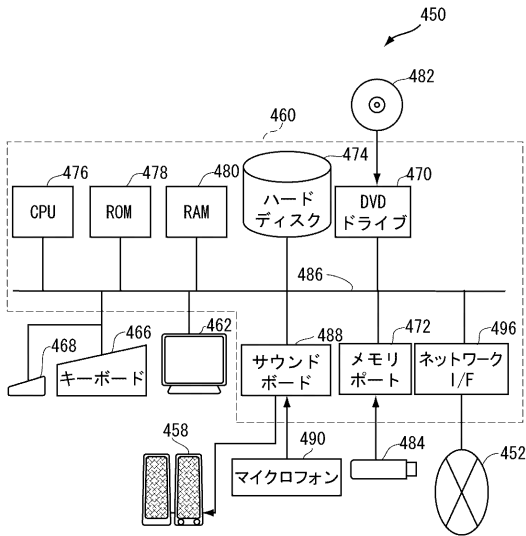
【 図 1 1 】



【 図 1 0 】



【図 1 2】



## 【手続補正書】

【提出日】平成19年3月16日(2007.3.16)

## 【手続補正 1】

【補正対象書類名】明細書

【補正対象項目名】0062

【補正方法】変更

【補正の内容】

【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。

【手続補正 2】

【補正対象書類名】明細書

【補正対象項目名】0076

【補正方法】変更

【補正の内容】

【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_N_E_X_T` 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。

【手続補正 3】

【補正対象書類名】明細書

【補正対象項目名】0083

【補正方法】変更

【補正の内容】

【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。

【手続補正 4】

【補正対象書類名】明細書

【補正対象項目名】0084

【補正方法】変更

【補正の内容】

【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、CPU 476 によって使用される作業領域を提供するとともに CPU 476 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。

【手続補正 5】

【補正対象書類名】明細書

【補正対象項目名】0089

【補正方法】変更

【補正の内容】

【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ジタル音声信号フレームのシーケンスを出力する。



【外国語明細書】

2008186171000001.pdf

2008186171000002.pdf

2008186171000003.pdf

2008186171000004.pdf