

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3549681号

(P3549681)

(45) 発行日 平成16年8月4日(2004.8.4)

(24) 登録日 平成16年4月30日(2004.4.30)

(51) Int. Cl.⁷

F I

G 1 0 L 15/28

G 1 0 L 3/00 5 6 1 B

G 1 0 L 15/06

G 1 0 L 3/00 5 2 1 F

G 1 0 L 15/14

G 1 0 L 3/00 5 3 5 Z

請求項の数 19 (全 24 頁)

(21) 出願番号	特願平8-242436	(73) 特許権者	390035493
(22) 出願日	平成8年9月13日(1996.9.13)		エイ・ティ・アンド・ティ・コーポレーション
(65) 公開番号	特開平9-127972		AT&T CORP.
(43) 公開日	平成9年5月16日(1997.5.16)		アメリカ合衆国 10013-2412
審査請求日	平成12年3月30日(2000.3.30)		ニューヨーク ニューヨーク アヴェニュー
(31) 優先権主張番号	08/528902		オブ ジ アメリカズ 32
(32) 優先日	平成7年9月15日(1995.9.15)	(74) 代理人	100064447
(33) 優先権主張国	米国 (US)		弁理士 岡部 正夫
		(74) 代理人	100085176
			弁理士 加藤 伸晃
		(72) 発明者	ウー チョー
			アメリカ合衆国 07922 ニュージャ
			ーシ、パークレイ ハイツ、グリーンブ
			ライア ドライヴ 22

最終頁に続く

(54) 【発明の名称】 連結数字の認識のための発声識別立証

(57) 【特許請求の範囲】

【請求項1】

既知の音声の発声を表わす1つ又は2つ以上の既知の仮定ストリング信号と現在の立証モデルのセットに基づいて音声認識装置立証モデルデータベースを作成する音声信号処理方法であって、

(A) 既知の音声の発声を表わす既知の仮定ストリング信号を受信するステップと、

(B) 既知の音声の発声を表わす既知の仮定ストリング信号と、その信号のための現在の立証モデルとに基づいて第1のストリング - ベース信用測定信号を生成するステップと、

(C) 各々が、既知の音声発話を表わす既知の仮定ストリング信号と、別の現在の立証モデルとに基づくものである1つ又は2つ以上の他のストリング - ベース信用測定信号を生成するステップと、

(D) 前記第1のストリング - ベース信用測定信号と前記他のストリング - ベース信用測定信号に基づいて誤立証信号を計算するステップと、

(E) 前記誤立証信号と既知の音声の発声を表わす前記既知の仮定ストリング信号に基づいて、前記現在の立証モデルのうちの1つ又は2つ以上を修正して未知の音声の発声を表わす未知の仮定ストリング信号を誤立証する尤度を減少させるステップと、及び

(F) 1つ又は2つ以上の修正された立証モデルをメモリーに格納するステップとを具備する方法。

【請求項2】

前記立証データベースの前記立証モデルを初期化するステップを更に具備する請求項1に

10

20

記載の方法。

【請求項 3】

前記誤立証信号に基づいて損失関数信号を計算するステップと、
前記損失関数信号の導関数を計算するステップを含めて、前記損失関数信号の勾配を決定するステップとを更に具備する請求項 1 に記載の方法。

【請求項 4】

前記ステップ (B) は、
前記既知の仮定ストリング信号を一連の単語信号に分割するステップと、
前記一連の単語信号の各々ごとに 1 つの単語ベース信用スコアを計算するステップと、
各単語信号に対して計算された単語ベース信用スコアを平均するステップとを具備する請求項 1 に記載の方法。 10

【請求項 5】

前記ステップ (C) は、
前記既知の仮定ストリング信号に基づいて混同しやすいストリングモデルのセットを生成するステップと、前記混同しやすいストリングモデルは、1 つ又は 2 つ以上の単語モデルを具備して、
前記セットの混同しやすいストリングモデルのうちのそれぞれに対してストリング - ベース信用測定信号を生成するステップと、及び
前記セットの混同しやすいストリングモデルのそれぞれに対するストリング - ベース信用測定信号の平均を計算するステップとを具備する請求項 1 に記載の方法。 20

【請求項 6】

前記ステップ (C) は、
前記セットの混同しやすいストリングモデルのそれぞれの単語モデルごとに単語ベース信用スコアを決定するステップと、
前記セットの混同しやすいストリングモデルのそれぞれの各単語モデルに対して決定された単語ベース信用スコアを平均化して前記セットの混同しやすいストリングモデルのそれぞれに対する前記ストリング - ベース信用測定信号を生成するステップとを具備する請求項 5 に記載の方法。

【請求項 7】

ステップ (D) は、 30
(a) その仮定ストリング信号に対する現在の立証モデルに基づく前記既知の仮定ストリング信号に対する信用測定信号の発生を、
(b) 1 つ又は 2 つ以上の他の現在の立証モデルに基づく前記既知の仮定ストリング信号に対する信用測定信号の発生に
関連づける関数の勾配を決定するステップを具備する請求項 1 に記載の方法。

【請求項 8】

ステップ (E) は、
前記勾配に基づいて前記現在の立証モデルの 1 つ又は 2 つ以上のパラメーターを調整するステップを具備する請求項 7 に記載の方法。

【請求項 9】

ステップ (D) は、 40
(a) 前記第 1 のストリング - ベース信用測定信号と
(b) 前記 1 つ又は 2 つ以上の他の信用測定信号の平均との差を形成するステップ
を具備する請求項 1 に記載の方法。

【請求項 10】

期待値をとった損失関数信号を評価するステップと、及び前記評価された損失関数信号に
応答して繰り返し前記立証モデルを調整して立証エラーの尤度を最小にするステップとを
具備する請求項 3 に記載の方法。

【請求項 11】

接続され発声された単語のストリングを認識するための音声認識システムであって、 50

接続され発声された単語の音響入力ストリングに応答して未知の音声の発声を表わす仮定ストリング信号を生成するための認識プロセッサと、
 複数の認識モデルを格納するための認識データベースと、
 未知の音声の発声を表わす前記仮定ストリング信号に
 応答してストリング - ベース信用測定信号を生成するための立証プロセッサと、
 複数の立証モデルを格納するための立証データベースとを具備する音声認識システム。

【請求項 1 2】

前記立証プロセッサに接続され、ストリング - ベース信用測定信号の立証しきい値信号値を計算するための手段と、
 前記ストリング - ベース信用測定信号を測定して測定された信号値を生成するための手段と、
 前記測定された信号値をしきい値と比較するための手段とを更に具備する請求項 1 1 に記載の音声認識システム。

10

【請求項 1 3】

前記立証モデルは、隠れマルコフモデルパラメータのセットを具備する請求項 1 1 に記載の音声認識システム。

【請求項 1 4】

前記立証モデルは、キーワードモデル、反キーワードモデル、音響のフィルタモデル、およびそれらの組合せから成るグループから選択される請求項 1 1 に記載の音声認識システム。

20

【請求項 1 5】

前記立証モデルは、識別訓練を反映する請求項 1 1 に記載の音声認識システム。

【請求項 1 6】

前記仮定ストリング信号は、分割された一連の単語信号からなり、
 前記立証プロセッサは、
 前記分割された一連の単語信号の各々に対する単語ベース信用スコアを計算するための手段と、
 前記分割された一連の単語信号の各々に対する単語ベース信用スコアの平均に
 応答して前記ストリング - ベース信用測定信号を生成するための手段とを具備する請求項 1 1 に記載の音声認識システム。

30

【請求項 1 7】

音声認識システムの立証プロセッサにより生成された信用測定信号の立証しきい値信号サンプルを発生する音声信号処理方法であって、

(A) (i) 既知の音声の発声を表わす既知の仮定ストリング信号と、既知の音声の発声を表わす該既知の仮定ストリング信号のための現在の立証モデルとに基づいて第 1 のストリング - ベース信用測定信号を生成し、

(i i) 各々が既知の音声の発声を表わす前記既知の仮定ストリング信号と、別の立証モデルに基づく 1 つ以上の他のストリング - ベース信用測定信号を生成するよう機能する立証プロセッサの一巡の動作を実行するステップと、

(B) 予め選択された回数についてステップ (A) を繰り返すステップと、

40

(C) 前記予め選択された回数について第 1 のストリング - ベース信用測定信号の値の第 1 の分布を記録するステップと、

(D) 前記予め選択された回数について、前記 1 つ又は 2 つ以上の他のストリング - ベース信用測定信号の値の第 2 の分布を記録するステップと、

(E) 前記第 1 の分布のための第 1 の平均を計算するステップと、

(F) 第 2 の分布のための第 2 の平均を計算するステップと、

(G) 前記第 1 の平均と、前記第 2 の平均に基づいて前記立証しきい値信号サンプルを生成するステップとを具備する方法。

【請求項 1 8】

誤り拒絶立証エラーレートを選択するステップと、

50

誤り受け入れ立証エラーレートを選択するステップと、前記選択された誤り拒絶立証エラーレートと前記選択された誤り受け入れ立証エラーレートに基づいて初期立証しきい値信号値を評価するステップとを更に具備する請求項 17 に記載の方法。

【請求項 19】

前記音声認識システムが動作している間に前記立証しきい値信号値を発生するステップと、前記認識システムが動作している間、前記誤り拒絶立証エラーレートと前記誤り受け入れ立証エラーレートを維持するステップとをさらに具備する請求項 18 に記載の方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、一般的に音声認識と立証の分野に関し、特に、音声立証訓練の分野に関する。

【0002】

【従来の技術】

音声認識は、未知の話された発声識別されるプロセスである。訓練として知られているプロセスを通して、知られている単語または単語ストリングがテストされ、単語または単語ストリングの特徴は音声認識装置メモリー内の認識モデルとしてサンプリングされて、記録される。認識モデルは、知られている発声の典型的な音響的解釈を表している。訓練プロセスにおいて、訓練アルゴリズムが認識モデルに適用され、将来の未知の単語および単語のストリングを認識するために利用されるこれらの格納された表現を形成する。

【0003】

動作において、音声認識装置は未知の発声を受信し、未知の発声を認識するために、未知の発声から特徴を抽出する。未知の発声の抽出された特徴は、テストパターンとして参照される。

【0004】

認識装置は、その後、未知の発声のためのテストパターンとメモリー内の 1 つ以上の認識モデルの組合せとを比較する。認識モデルの各組合せがテストパターンとどれほどよくマッチするかの相対的な測定を提供するために、スコアをカウントする技術が用いられる。未知の発声と最もはつきりとマッチする 1 つ以上の認識プロセッサモデルの組合せと関連する単語として未知の発声は認識される。

【0005】

従来の音声認識の実際には、音声認識モデル化とモデル訓練のために妥当な可能性またはクラスの数「N」が利用された。そのような N - クラスモデルにおいては、認識されるべきすべての発声入力は妥当であると仮定される。「N」個の存在可能なクラスの各々に対する認識モデルは、認識装置メモリーに格納されている。妥当であるかどうかにかかわらず、認識されるべき全ての音声入力は、Nクラスのうちの 1 つとして分類される。認識装置は、N個のモデルの各々に対する発声のスコアを計算し、発声を、最もよいスコアを持つ(N個のクラスの内の)あるクラスにマッチさせる。

【0006】

付加的な、または「N + 1」のクラス概念が、N個の妥当なクラスの認識に加えて、(「フィルター」モデルと指定される)関連した N + 1 モデルを提供するように開発されている。そのようなシステムにおいては、認識されるべき入力発声は N 個の妥当なクラスの入力に対する N 個のモデルの各々に対して、また付加的に、N + 1 フィラークラスモデルに対して、スコアが数えられる。N + 1 モデルは、すべての無効の入力を表すように設計されている。N + 1 フィラーモデルの使用により、認識プロセスはさらに洗練され、入力発声または入力発声のセグメントが妥当ではないかもしれないことを考慮するようになった。

【0007】

いったん未知の発声が音声認識プロセスによって分類されると、音声認識装置によりなさ

10

20

30

40

50

れる分類決定の確実性を評価することがしばしば、望ましい。これは、2パスプロセスにより達成され、未知の発声は最初にN個の可能な妥当なクラスを表すN個のモデルに対してそのスコアを数えることにより認識される(分類される)。最もよいスコアを提供する認識モデルは、その関連したクラスにマッチされ、それは、後で立証されるべき仮定発声として認識される。仮定認識発声は、その後、フィルア立証モデルに対して第2のパスにおいて立証され、そのモデルは、上述のN+1認識フィルアモデルと同様であるはずである。

【0008】

2パス音声認識プロセスは、未知の発声の認識を可能にし、続いて、システムが認識決定についてどれほど信用できるかの計算される判定が続く。例えば、認識装置は未知の発声を特定の単語として分類し、立証器は、フィルアモデルを使って分類が正しいことを示す。立証器は、認識装置が適切に分類したかどうかにかかわらず、立証器は、2進の決定、「はい」、または「いいえ」を作成するために使用される。

10

【0009】

発声立証は音声認識システム開発の重要な観点を表している。発声立証は、計算された信用スコアに基づいて全体の発声のうちの拒絶する部分またはすべてを含んでいる。信用スコアは、尤度比距離に基づくことができ、それは、単語が音声のあるセグメント中に存在するかどうかをテストする。尤度比距離は、発声が妥当な単語なしでなされ、認識される発声が乏しい状況で、あるいは重要な混乱が高いエラーレートを発生する単語間に存在するとき、特に有益である。

20

【0010】

音声認識システムの文脈での発声立証の成功した性能は、立証モデルを訓練するために用いられる技術の有効性と密接に関連し、そのモデルは立証プロセスにおいて使用される。従来の訓練方法には欠陥が存在し、それは、訓練と立証エラーレート間の直接的な関係の不足に起因している。立証エラーレートを直接最小化する訓練方法を見つけることが、目下の問題である。

【0011】

【発明の概要】

音声認識技術の進歩が、本発明により提供され、それでは、接続された単語のストリングを認識するために2つのパスの認識と立証戦略を採用する。本発明の図示される実施例によれば、音声認識システムは、認識プロセッサを具備するか、または仮定ストリングを生成し、未知の接続された単語の音響的入力ストリングに基づいて接続された単語の仮定ストリングを発生するための認識プロセッサと、接続された単語の仮定ストリングに回答するストリングベース信用測定信号を生成する立証プロセッサとを具備する。認識データベースは、認識プロセッサによって使用される認識モデルを格納する。立証データベースは、立証プロセッサによって使用される立証モデルを格納する。

30

【0012】

第1のパスでの認識は、隠れマルコフモデルのセットを使用して従来のビタービ(Viterbi)ビームサーチアルゴリズムで実行され、そのモデルは、ストリング分類エラーを最小化するように訓練されている。第2のパスでの発声立証は、本発明による最小ストリング立証エラー(MSVE)訓練プロセスで訓練された隠れマルコフモデルのセットを使用して、仮定認識ストリングを立証するために実行される。

40

【0013】

立証エラーを最小限にするために立証データベースに格納された立証モデルを訓練する方法が利用される。訓練方法は、予期されるストリング立証エラーを計算し、最小化することを含んでいる。

【0014】

訓練プロセスにおいて、最初に、立証データベースの立証モデルが初期設定される。既知の仮定ストリングが選ばれる。立証プロセッサは、既知の仮定ストリングのための現在の立証モデルに基づいて第1のストリングベース信用測定信号を生成する。

50

【0015】

既知の仮定ストリングのための現在の立証モデルと混同するほど似た1つ以上の競合ストリングモデルが生成される。1つ以上の競合ストリングモデルは立証プロセッサに提供され、それは、既知の仮定ストリングと1つ以上の競合ストリングモデルに基づいて1つ以上のストリングベース信用測定信号を生成する。

【0016】

誤立証信号は、既知の仮定ストリングのための第1のストリングベース信用測定信号に基づいて立証プロセッサにより生成され、1つ以上の競合ストリングモデルに基づいてストリングベース信用測定信号が生成される。誤立証信号に基づく損失関数信号は、その既知の仮定ストリングのための現在の立証モデルに基づく既知の仮定ストリングの立証を、1つ以上の他の競合立証モデルをもつ既知の仮定ストリングの立証に関係づける勾配に提供する。勾配決定は、損失関数信号の導関数を利用して達成される。

10

【0017】

立証モデルは、未知の仮定ストリングを誤立証する尤度を最小化するために更新される。更新プロセスは、予期される損失関数信号を評価し、立証エラーの尤度を最小化するために、評価される損失関数信号に回答する立証モデルを繰り返し調整する。

【0018】

本発明の音声認識システムは、立証プロセッサにより生成される信用測定信号を立証するために最適立証しきい値信号値を生成するように動作させられることができる。立証しきい値信号値は、仮定認識ストリングを拒絶する決定のための基礎を提供し、立証エラーレートを維持するように選ばれる。

20

【0019】

既知の仮定ストリングは、立証プロセッサに入力され、それは、立証データベースに格納された立証モデルをアクセスするように結合されている。1つ以上の競合ストリングモデルが既知の仮定ストリングに基づいて生成される。1つ以上の競合ストリングモデルは立証プロセッサに提供される。

【0020】

立証プロセッサは、予め選択された数の場に対して、既知の仮定ストリングのための現在の立証モデルに基づいて第1のストリングベース信用測定信号を生成し、既知の仮定ストリングに基づいて1つ以上の競合ストリングモデルのうちの各々に対するストリングベース信用測定信号を生成するように動作させられる。既知の仮定ストリングに対する現在の立証モデルに基づくストリングベース信用測定信号値の第1の分布は記録される。既知の仮定ストリングに対するストリングベース信用測定信号値の第2の分布と競合モデルは記録される。立証しきい値信号値は、第1の平均と第2の平均とに基づいて選択され、調整される。

30

【0021】

ここに、教示される技術は、ストリング立証エラーレートを最小化することと直接関連する。

本発明の他の目的、利点、および顕著な特徴は、本発明の好適実施例を開示し、添付図面と関連してなされる以下の詳細な説明から明らかとなるであろう

40

【0022】

【発明の実施の形態】

説明の明瞭化のために、本発明の図示される実施例は、(「プロセッサ」とラベルされた機能ブロックを含めて)個々の機能ブロックを具備するとして提供される。これらのブロックが表している機能は、それに制限されないが、ソフトウェアを実行可能なハードウェアを含めて、共有されたまたは専用とされたハードウェアのいずれかの使用によって提供される。例えば、図1から図4において提供されるプロセッサの機能は、単一の共有プロセッサによって提供されても良い。(ソフトウェアを実行可能なハードウェアを排他的に参照するように、用語「プロセッサ」の使用が解釈されるべきでない。)

【0023】

50

図示される実施例は、A T & T DSP 16またはDSP 32Cなどのデジタル・シグナル・プロセッサ(DSP)ハードウェア、下で議論される動作を実行するソフトウェアを格納するためのリード・オンリー・メモリ(ROM)、DSP結果を格納するためのランダム・アクセス・メモリ(RAM)を具備する。汎用DSP回路と組合せてのカスタムVLSI回路だけでなく、超大規模集積(VLSI)ハードウェア実施例が、また提供されてもよい。

【0024】

図1は、HMMベース音声認識システムの文脈の一般的説明の目的のために本発明の図示される実施例を提供する。音声認識システム8は、認識プロセッサ10、別個の隠れマルコフモデル("HMM")認識モデルのセットを含む認識データベース12、立証プロセッサ14および別個のHMM立証モデルのセットを含む立証データベース16を具備する。

10

【0025】

認識プロセッサは、単語の未知の音声ストリング18(発声)を入力として受信する。認識プロセッサ10は、未知の音声ストリング18入力に応答して認識データベース12をアクセスし、未知の単語ストリングを分類して、仮定認識ストリング信号を生成するために、認識データベース12の認識モデルに対する未知の単語音声ストリングのスコアを数える。立証プロセッサは、仮定ストリング信号20を、立証されるべき入力として受信する。立証プロセッサ14は、立証データベース16をアクセスして、立証データベースに格納されている立証モデルに対する仮定ストリング信号をテストする。立証テストに基づいて、立証プロセッサ14は、信用測定信号22を測定して、測定された信号値を生成する。測定された信号値は、立証しきい値信号値に対して比較されるべきしきい値コンパレータ24に通され、認識プロセッサ10によりなされた分類決定の精度を決定する。

20

【0026】

図2は、音声認識システム8の立証訓練およびテストプロセスを説明するためのより詳細な図を提供する。図2を参照して、音声認識システムは、訓練音声データベース26、特徴抽出器28、第1のモードスイッチ30、第2のモードスイッチ32、単語ベース信用スコア計算器34、平均単語ベース信用スコアプロセッサ36、ストリングベース信用測定信号発生器38、しきい値比較器40、従来のHMM訓練器42、HMMパラメータデータベース44、N最良ストリングモデル発生器46、および強化された訓練器48を

30

【0027】

訓練音声データベース26は、デジタル形式での既知の音声信号のサンプルを具備する。各音声信号は、1語以上の話された単語のストリングに対応する。これらのサンプルは従来の訓練器42と強化された訓練器48のためにそれぞれ使われる。訓練音声データベース26からの既知の音声ストリングのサンプルは、第1のモードスイッチ30を介してシステムの他の構成要素に提供される。

【0028】

第1のモードスイッチ30は、システムの2つの動作モードを反映する：訓練(T)モードとテストのための立証(V)モードである。第1のモードスイッチ30が(図示のように)T位置にあるとき、訓練データベース26からの訓練ストリング信号は、訓練がシステムの訓練部50によって実行できるように、システムのバランスに提供される。第1のモードスイッチがV位置にあるとき、デジタル形式の仮定認識音声ストリング信号が、システムの立証部52により立証のためにシステムのバランスに提供される。

40

【0029】

システムの特徴抽出器28は、第1のモードスイッチ30に結合されている。第1のモードスイッチの状態に依存して、特徴抽出器28は、訓練ストリング信号または仮定音声ストリング信号を受信する。これらの受信ストリング信号に基づいて、特徴抽出器は、立証または訓練の目的でストリング信号を特徴付ける1つ以上の特徴ベクトル0を作成する。

【0030】

50

特徴抽出器からの特徴ベクトル出力は、第2のモードスイッチ32に提供される。第2のモードスイッチは、第1のモードスイッチとタンデム形式で動作する。すなわち、両方のスイッチは、両方のスイッチに提供されるモード信号54に基づいて、同時に同じ状態（TまたはV）にセットされる。第2のモードスイッチ32は、特徴抽出器28の出力をシステムの立証部52または訓練部50に向ける。

【0031】

システムの立証部52は、認識プロセッサ10によりなされた認識決定をテストするために選択されることができる（図1）。図2に示される立証部52は、単語ベース信用スコア計算器34、平均単語ベース信用スコアプロセッサ36、およびストリングベース信用測定信号発生器38を具備する。

10

【0032】

特徴抽出器28は、一連のフレーム特徴ベクトルを出力する。特徴ベクトルのフレームに基づく単語信号は、単語ベース信用スコア計算器34に提供される。単語ベース信用スコア計算器は、離散的な単語信号について計算を実行する。単語ベース信用スコア計算器34は、1つ以上の立証モデル（この場合隠れマルコフモデル「HMM」）のストリングを、仮定認識音声ストリング信号の各単語信号にマッチさせる。信用スコア計算器34は、仮定ストリング信号の各単語信号に対して単語ベース信用スコアを生成する。ストリングの各単語セグメントに対して計算される単語ベース信用スコアは、信用のレベルを示し、それは、仮定ストリングの各特定の単語が正しく認識され、音声認識システムに提供された入力発声の単語に正確に対応しているレベルである。

20

【0033】

平均単語ベース信用スコアプロセッサ36は、仮定ストリング信号の各単語セグメント信号に対して生成される単語ベース信用スコアのセットについて算術平均演算を実行して、平均単語ベース信用スコア信号を生成するための手段を提供する。平均信用スコア信号は、ストリングベース信用測定信号発生器38に通される。

【0034】

ストリングベース信用測定信号発生器38は、平均信用スコア信号を受信するように結合されていて、ストリングベース信用測定信号を測定して、測定された信号値を生成するように応答する。ストリングベース信用測定信号の測定された信号値は、立証テスト結果を具体化し、立証プロセッサに提供される仮定ストリング信号が実際の話された単語ストリングを正確に表しているという信用度の測定を提供する。

30

【0035】

立証プロセッサに提供される仮定（の）認識（される）音声ストリング信号が音声認識システムに元々入力される未知の音声単語ストリングであることを立証するために、信用測定信号の測定された信号値は、立証しきい値信号値40と比較される。この測定された信号値がしきい値信号値以下ならば、仮定ストリング信号は、実際の話された入力発声を間違っていると拒絶される。すなわち、立証プロセッサに提供される仮定ストリング信号は立証されることができないであろう。

【0036】

システムの訓練部50は、従来のHMM訓練器42、HMMパラメータデータベース44、N最良ストリングモデル発生器46、および強化された訓練器48を具備する。従来のHMM訓練器42は、米国特許番号4,783,804により説明されるようであり、HMMの従来の訓練を提供する。従来の訓練器の出力は、HMMのセットに基づく訓練音声ストリングの特徴ベクトルだけでなく、1つ以上の従来のように訓練された複数のHMMの組 I を具備する。従来の訓練器42の出力は、初期化パラメータとしてN最良ストリング発生器46および強化された訓練器48による使用のためにHMMパラメータデータベース44に記録される。

40

【0037】

強化された訓練器48は、初期のパラメータ I を受信して、HMMパラメータデータベースに強化されたまたは更新されたHMMのセット U を戻す。強化されたモデ

50

ル U は、最終的に、立証モード処理の間に単語ベース信用スコア計算器 34 に提供される。従来のように訓練された HMM_I よりもむしろ強化された HMM_U の単語ベース信用スコア計算器の使用は、ストリング立証エラーレートの減少を達成するために提供される。

【0038】

N最良ストリングモデル発生器 46 は、 HMM パラメータデータベース 44 から、隠れマルコフモデル (HMM) パラメータ x と訓練音声ストリング O を受信するように結合されている。N最良ストリングモデル発生器 46 は、ストリングモデルのセットを生成するように動作し、そのモデルは、仮定認識ストリング信号のための現在の立証モデルと高度に競合する。これらの競合するストリングモデルは、強化された訓練器 48 の区別的な訓練のための基礎を提供する。即ち、立証プロセッサへの正しい仮定ストリング信号入力が立証プロセッサに極めて高い信用測定信号を生成させるように、また、間違っ

10

【0039】

N最良ストリング発生器 46 は、"N" 個の最良の (競合する) 単語ストリングと、各そのような単語ストリングのための最良の HMM 立証モデル列とを決定することにより、"N" 個の最良の (あるいは最も競合する) 単語ストリングを発生する。N最良単語ストリングの決定は、修正されたビタービデオコーダの DSP 実行の使用によってなされる。

【0040】

修正されたビタービデオコーダは、部分的なパスマップ、即ち瞬間毎にいずれかの文法ノード (即ち単語の結合点) に通じるすべての部分的なパスのすべてのスコアのリスト (または、マトリクス) を準備する。その後、修正されたビタービ前検索の終了時に、デコーダは、当業界でよく知られた A^* トレースバックツリー検索を使用する。発生器は、「スタック」 (位付けられたデータリスト) の先頭にある先頭 (ベスト) 部分的パスを育てることによりツリー検索を実行する。スタック上のすべての部分的パスの位付けは、完了されるときに部分的パスが達成できる最良の可能なスコアにより決定される。ビタービ前検索において準備される部分的パスマップは、後方ツリー検索においていかなる不完全な部分的パスに対してもスコアを提供する。 A^* 手続の「最良が1番」という性質のため、先頭のNストリングは逐次生成される。

20

30

【0041】

スタック深さがいずれの検索エラーも回避するように十分に深いとすると、 A^* アルゴリズムの許容性、または最適なパスの確実な発見は、部分的なパスの不完全な部分の上限がスタックデコーダとして使用されるときに保証される。ツリートレリス手順において、同じ HMM が両方向に沿っての検索において使われれば、最もしっかりとした上限、あるいはそのパスの実際のスコアが得られる。

【0042】

発生器は、N最良ストリングリスト中の各単語ストリングに対するモデルのセットに対して入力音声ストリング O のビタービ整列によりN最良単語ストリングに基づいてN最良ストリングモデルを生成する。このステップは、同じ単語ストリングに対応する多くのスト

40

【0043】

発生器が、未知の競合ストリングをモデル化する能力を提供することに注意すべきである。これは、N最良デコーディングによって生成される競合する単語ストリングモデルが、訓練資料により制限されないという事実起因している。N最良ストリングリストは、動的で、音声立証モデルの組に基づいてストリングレベル音響分解能を反映する。

【0044】

従って、 HMM パラメータデータベース 44 の基礎的な立証モデルは、N最良ストリングモデル発生器 (一種の音声認識装置) により生成されるストリングモデルにより訓練さ

50

れる。生成されたストリングモデルは、元の訓練資料に無い単語ストリングに基づくことができるが、認識装置が、既知の仮定ストリング信号と混同してそれを見つける。認識装置が既知の仮定ストリング信号と混同して見つけるこれらの知られていない競合単語ストリングは、そのような知られていない混同しやすい同様なストリングに対する立証モデル化を改善するように識別的にモデル化される。(正しいストリングと混同されそうである)知られていないストリングを識別するための基礎を提供するように立証モデルを準備するモデル化は、訓練資料の貧しいストリング範囲によりそのようなモデル化が連続的な音声認識と立証において遭遇する困難さを減少させるので、有益である。

【0045】

N最良ストリングモデル発生器の動作は、同時継続の米国特許出願番号08/030,895に詳細に説明されており、それは本願の所有者に譲渡され、ここにそれに述べられているまま引用により組み込まれる。

【0046】

最小ストリング立証エラーレート訓練では、N最良ストリングモデルは、ストリングエラーを表し、減らすために特に設計されている識別関数のセットに組み込まれる。これは、以下に説明するように、強化された訓練器48の動作によりなされる。

【0047】

発声立証プロセッサ14(図1)は、仮定ストリング信号という形で提案された仮定認識発声を立証するために、統計的仮定テストを公式化し、適用するように動作する。立証プロセスのタスクは、仮定のキーワードまたはキーワードの組が、未知の発声に存在しない(すなわち、提案された認識された仮定発声20は不正確である)という他の仮定に対して、仮定のキーワードまたはキーワードのセットが、未知の発声内に存在するという無効仮定をテストすることである。

【0048】

立証プロセスは、提案された仮定発声の精度を立証するために、尤度比距離計算を用いるように設計されている。本発明によれば、発声立証モデルは、ストリングレベルについて作られ(訓練され)、立証テストは、ストリングレベルについて実施される。ここで教示される発声立証は、仮定ストリング信号に対するストリングベース信用測定信号を生成するために単語信号セグメントの単語ベース信用スコアの寄与を結合する幾何学的平均化の形式を使用する。ストリングベース信用測定信号は、仮定ストリング信号のすべてまたは一部分を受け入れ、あるいは拒絶するために、しきい値信号値40(図2)に対して立証される。

【0049】

図2を参照して、立証テストモードでは、仮定された単語ストリング55は、立証部52への提供のために第1のモードスイッチ30を介して特徴抽出器28に通される。単語iの仮定されたストリング(信号55)は、認識プロセスの結果としてN個の単語{O_q}に分割される。

【0050】

立証部52は、立証のために、与えられたモデル組 θ を利用する。立証部52は、HMMパラメータデータベース44の現在のモデルのセットに基づいて仮定単語ストリングに基づいて信用測定信号を生成する。仮定単語ストリングは、生成された測定信号

【数1】

$$s_i(O; \theta) = \log \left[(1/N) \sum_{q=1}^N \exp \{ k \cdot L(O_q; \theta, 1) \} \right]^{1/k} \quad (1)$$

が、予め定義された立証しきい値信号値以下であれば、拒絶される。ここで、kは、負の定数であり、 $L(O_q; \theta, 1)$ は、数字(フレーム)1と認識される音声セグメントqの信用スコアを示す。従って、信用測定信号は、ストリング信号の各音声セグメントq(単語フレーム)の信用スコアの平均と数学的に関連付けられる。

10

20

30

40

50

【0051】

HMMベース立証モデルは、立証部52によりアクセスされて、利用されるための立証データベース44に格納されている。立証データベース44内の立証モデルは、3つの異なる組からなる。即ち、キーワード間の混同可能性を扱うキーワード $\{ \theta_1^{(k)} \}$ 、と反キーワード $\{ \theta_1^{(a)} \}$ 、及び無キーワード（無効の入力）を識別するための一般的音響フィルタモデル $\theta_1^{(f)}$ である。

【0052】

音声セグメント q 、 $L(O_q; \theta, l)$ の単語ベース信用スコア34は、キーワード仮定とその競合する他の仮定の間で優度比のログ(log)を用いて構成されることができ

10

【数2】

$$L(O_q; \theta, l) = g_1(O_q) - G_1(O_q), \quad (2)$$

ここで、 $g_1(O_q) = \log p(O_q | \theta_1^{(k)})$ であり、

【数3】

$$G_1(O_q) = \log[\sum_p p(O_q; \theta_1^{(p)}) + \sum_f p(O_q; \theta_1^{(f)})]. \quad (3)$$

従って、音声セグメント q に対する信用スコア計算は、単語モデルスコアとフィルタモデルを用いて反単語モデルで計算されたスコアの間での比較を関連づける。

【0053】

式(1)で信用測定信号 $s_i(O; \theta)$ を最大化する方法の1つは、 θ のパラメーターが、データベース44に格納されるように訓練するために、最大尤度手続を適用することである。しかしながら、音声認識実験に基づいてこの種の訓練は、立証または認識エラーレートを最小化することについて最良の性能を与えないことが示されている。

20

【0054】

このため、最小ストリング分類エラー(MSCE)訓練プロセスが、認識データベース12(図1)に認識モデルを創造するために開発された。そのような音声認識訓練方法は、1つ以上の既知の音声信号と現在の認識モデルのセットとに基づいて音声認識データベースを提供するために用いられる。MSCE訓練プロセスにおいて、第1の認識装置スコア信号は、既知の音声信号と、その信号のための現在の認識モデルに基づいて生成される。混同しやすい仮定のセットは、各々既知の音声信号および別の現在の認識モデルに基づいて1つ以上の他の認識装置スコア信号を生成するために生成されて、認識装置に適用される。

30

【0055】

認識プロセッサは、第1の認識装置スコア信号および他の競合認識装置スコア信号に基づいて誤分類(誤認識)信号を生成する。認識モデルパラメーターは、誤認識信号に基づいて修正されて、訓練における既知の音声信号またはテスト動作における未知の音声信号を誤認識する優度を減少させる。実施例の説明と手続は、本願の所有者に譲渡された上記同時継続米国特許出願08/030,895に提供される。

【0056】

ストリングベース最小ストリング分類エラー(MSCE)訓練におけるゴールは、予期されるストリング分類エラーレートを最小化することである。対照的に、最小ストリング立証エラー(MSVE)訓練におけるゴールは、予期されるストリング立証エラーレートを最小化することである。従って、MSCEアプローチにおいて使われる数学的な戦略と採用される特定の目的関数は、最適な発声立証性能とは一致しない。

40

【0057】

本発明は、MSCEと同じ精神で最小ストリング立証エラー(MSVE)訓練プロセスを利用するが、HMMパラメーターデータベース44の立証モデルに対して設計されている。誤立証測定は計算されて、予期されるストリング立証エラーレートを最小化するために使用される。

50

【 0 0 5 8 】

MSVE 訓練は、ストリングレベルにおいて適用されて、式 (1) において使われる目的関数と一致している。MSVE 訓練は、予期される関数を最小化するように実行される。

【 数 4 】

$$E[S\{-s_i(O;\theta) + S_i(O;\theta)\}], \quad (4)$$

ここで、 $S(\)$ は、なめらかな 0 - 1 シグモイド非線形関数である。

10

【 0 0 5 9 】

図 3 を参照して、強化された訓練器 4 8 は、ストリング誤立証測定プロセッサ 5 6、ストリングモデル損失関数プロセッサ 5 8、及び HMM 更新プロセッサ 6 0 を具備する。強化された訓練プロセスは、目的として、式 (4) について説明された予期される損失関数を最小化して、ストリング立証エラーを実質的に減少させる。好適実施例において、強化された訓練プロセスは、音声認識システムにおいて上記プロセッサにより実行される 2 つのプロセスステップとして実現される。

【 0 0 6 0 】

A . ストリング誤立証測定プロセッサ。

ストリング誤立証測定プロセッサ 5 6 は、N 最良ストリングモデル発生器 4 6 から HMM 20
パラメータ θ_i 、N ストリングモデル S_j 、および訓練音声ストリングサンプル O を受信する。ストリングサンプル O 、モデル S_j 、及び θ_i に基づいて、誤立証プロセッサ 5 6 は、第 1 のプロセスステップとして、 i 番目の仮定ストリング $d_i(O; \theta_i)$ ためのストリング誤立証測定値を決定する

【 数 5 】

$$d_i(O;\theta) = -s_i(O;\theta) + S_i(O;\theta), \quad (5)$$

ここで、 $s_i(O; \theta_i)$ は、式 (1) により定義される発声ベース識別関数であり、 $S_i(O; \theta_i)$ は、以下のように定義される反識別関数である。

30

【 数 6 】

$$S_i(O;\theta) = \log [(1/N-1) \sum_{j,j \neq i} \exp\{\eta \cdot s_j(O;\theta)\}]^{1/\eta}, \quad \eta > 0, \quad (6)$$

ここで、 N は、競合ストリング仮定の全体数であり、 η は、図では 2 に設定される正の数である。

【 0 0 6 1 】

ストリング誤立証プロセッサ 5 6 は、図 3 に示される後続のプロセッサによって使用されるためのスカラー d の値を決定し、以下の間の差を説明する誤立証信号を生成する：(a) 既知の仮定ストリング信号に対する現在の立証モデルと既知の仮定ストリング信号に基づく信用測定信号、および (b) その既知の仮定ストリング信号に対する N 最良競合立証モデルに基づく 1 つ以上の他の信用測定信号の平均。

40

【 0 0 6 2 】

B . ストリングモデル損失関数プロセッサ

第 2 のプロセスステップは、立証エラーカウントを近似する。これは、シグモイド非線形関数を用いてなされる。ストリングモデル損失関数プロセッサ 5 8 は、誤立証測定プロセッサ 5 6 から受信されるスカラー値 $d_i(O; \theta_i)$ に基づいてなめらかなストリング損失関数 $l_i(O; \theta_i)$ を評価する。この損失関数は以下のように定義される。

50

【数7】

$$l_i(O; \theta) = \frac{1}{1 + \exp\{-\alpha d_i(O; \theta) + \beta\}} \quad (7)$$

ここで、 α と β は、それぞれ、シグモイド非線形平滑化関数の傾きとシフトを制御する定数である。

【0063】

ストリングモデル損失関数プロセッサ58は、出力として、スカラー1を損失関数信号の形でHMM更新プロセッサ60に提供する。 10

【0064】

上記のパラメーター α と β は、ストリング損失関数 $l_i(O; \theta)$ に平滑化された近似を提供するのを助ける。ストリング損失関数の傾きは、HMM更新プロセッサ60により使用され、現在の立証モデルHMMパラメーター θ を更新する。その傾きは、その仮定ストリング信号に対する現在の立証モードモデルに基づく既知の仮定ストリング信号の信用測定を、1つ以上の他の現在の立証モデルに基づく既知の仮定ストリング信号の信用測定に関連づける。従って、この損失関数の予期されるストリング損失の最小化は、ストリング立証エラー確率の最小化と直接結び付けられる。 20

【0065】

C. HMM更新プロセッサ

立証モデルパラメーター更新の説明に先立ち、HMM更新プロセッサ60の動作の背景が提供される。最小ストリング立証エラー(MSVE)訓練は、式(7)の予期される損失関数を最小化する1組のパラメーターを見つけることに導かれ、それは $E[l_i(O; \theta)]$ として指定される。 20

【0066】

最小化パラメーターの組を見つけるために、パラメーターセット θ は、以下の式に従って、各繰り返し n 毎に更新される。

【数8】

$$\theta_{n+1} = \theta_n - \epsilon_n \nabla E[l_i(O; \theta)], \quad \epsilon_n > 0. \quad (8)$$

式(8)では、 θ_n は初期HMMパラメーターセット評価である。 $E[l_i(O; \theta)]$ は、予期される損失関数である。 ∇ は、数学的導関数演算を表している。 θ_{n+1} は、予期される損失関数 $E[l_i(O; \theta)]$ を最小にするHMMパラメーターセットに対する最良評価である。 30

【0067】

この式は、立証モデルパラメーターを調整するための勾配下降繰り返しを遂行して、予期される損失関数の最小化を達成する。ここで、 ϵ_n は、繰り返しの際に使用される一連のステップサイズパラメーターあるいは学習レートであり、 V_n は、以下に定義される正の有限の学習マトリクスである。 40

【数9】

$$\sum_{n=1}^{\infty} \epsilon_n = \infty \text{ and } \sum_{n=1}^{\infty} \epsilon_n^2 < \infty.$$

ならば、予期される損失の静止点に収束する。

【0068】

図3の実施例では、HMMパラメータは、HMM更新プロセッサ60により(8)に従って適応的に調整される。HMMの構造により、それらのパラメータは、ある制約条件を満たさなければならない。HMM更新プロセッサは、すべてのそのような制約条件を満たすように、パラメータ更新プロセスの一部として、変換されたHMMパラメータを使用する。以下の変換は、HMM更新プロセッサにより使われる。

【0069】

(1) 分散の対数

【数10】

$$\sigma_{i,j,k,d}^2 = \log \sigma_{i,j,k,d}^2 \quad (9)$$

10

ここで、各立証モデルが、例えば単語またはサブ単語を反映する場合、 $\sigma_{i,j,k,d}^2$ は i 番目の立証モデル、 j 番目の状態、 k 番目の混合要素、及び d 特徴ベクトルの d 次元の分散である。

【0070】

(2) 混合重みの変換された対数

元の混合重み $C_{i,j,k}$ は、次の通りに、変換された混合重み $C_{i,j,k}$ と関連づけられる。

【数11】

$$C_{i,j,k} = \frac{e^{C_{i,j,k}}}{\sum_{l=1}^L e^{C_{i,j,l}}} \quad (10)$$

20

ここで、 L は、 i 番目の立証モデル中の j 番目の状態にある混合重みの全体数である。

【0071】

(3) 遷移確率の変換された対数

元の遷移確率 $a_{i,j}$ は、以下のようにして、変換された遷移確率に関連づけられる。

【数12】

$$a_{i,j} = \frac{e^{a_{i,j}}}{\sum_{k=1}^M e^{a_{i,k}}} \quad (11)$$

30

ここで、 M は、 i 番目の立証モデルの全体の状態数である。

【0072】

図示される実施例の重要な観点は、小さい分散を扱うことに関する。正しくなく扱われると、逆効果になるので、過去に提案されたいくつかの補正訓練アルゴリズムにおいては、分散調整は避けられていた。

40

【0073】

HMMの分散は、 10^4 から 10^6 程異なる。これらの分散は、観察確率密度関数 $b_{i_1}(x)$ の指数項に生じ、HMMの尤度スコアの支配的な効果を持つ。図示される実施例において、これは、HMMパラメータ調整特に、HMMの観察確率密度関数の平均パラメータの調整への異なる感度に導く。

【0074】

この巨大な感度の差を補償するために、実施例は、正の有限のマトリクス V_n を使用する。正の有限のマトリクス V_n は、各状態に対して、

【数13】

$$\text{diag}(\sigma_1^2(n), \dots, \sigma_D^2(n)), \tag{12}$$

である対角マトリクスである。ここで、 $\sigma_i^2(n)$ は、時間 n における HMM I の分散である。

【0075】

図4は、HMM更新プロセッサ60のブロック図を示す。図に示すように、HMM更新プロセッサ60は、 I 、 O 、 l に基づいて、パラメーター μ 、 σ^2 、 c 、 a を更新し、HMMパラメーターデータベース44に更新された U を戻す。 I と U は、パラメーター μ 、 σ^2 、 c 、 a からなり、 I と U は、それぞれ更新されない、及び更新された量を表している。 10

【0076】

最小ストリング立証エラー訓練では、式(8)に示されるモデルパラメーターを更新することは、損失関数の導関数を見つけることを含んでいる。項 l / L は、各モデルパラメーターに対して特に計算され、目下のHMMがキーワードモデルか、反キーワードモデル、またはフィルモデルであるかに依存して異なっている。このステップは、導関数 l / L が全てのモデルに共通であれば、最小ストリング分類エラー訓練において使われるものとは異なっている。 20

【0077】

MSVEでは、 l / L は、項 l / L が全ての立証モデルに共通であれば、 $l / L = l / L \cdot L / L$ のようなチェーン規則として書かれることができる。 l / L は、以下の部分導関数からなる。

【数14】

$$\frac{\partial l}{\partial d} \cdot [\frac{\partial d}{\partial s_i} + \frac{\partial d}{\partial s_i} \cdot \frac{\partial s_i}{\partial s_j}] \cdot \frac{\partial s}{\partial L}$$

$$\frac{\partial l}{\partial d} = \alpha l_i (1 - l_i) \quad ;$$

$$\frac{\partial d}{\partial s_i} = -1 \quad ;$$

$$\frac{\partial d}{\partial s_i} = 1 \quad ;$$

$$\frac{\partial s_i}{\partial s_j} = \frac{\exp[\eta s_j(O; \theta)]}{\sum^N \exp[\eta s_j(O; \theta)]}$$

$$\frac{\partial s}{\partial L} = \frac{\exp[\kappa L(O_q; \theta; l)]}{\sum^N \exp[\kappa L(O_q; \theta; l)]}$$

【0078】

導関数 l / L は、更新されたモデルがキーワード (k) であるか、反キーワード (a)、またはフィルモデル (f) であるかに依存して異なる。

(k) では、

【数15】

10

20

30

40

$$\frac{\partial L}{\partial \theta^{(k)}} = \frac{1}{p(O_q \theta^{(k)})} \frac{\partial p(O_q \theta^{(k)})}{\partial \theta^{(k)}}$$

(a) では、
【数 1 6】

$$\frac{\partial L}{\partial \theta^{(a)}} = \frac{-1}{p(O_q \theta^{(a)}) + p(O_q \theta^{(f)})} \frac{\partial p(O_q \theta^{(a)})}{\partial \theta^{(a)}}$$

10

(f) では、
【数 1 7】

$$\frac{\partial L}{\partial \theta^{(f)}} = \frac{-1}{p(O_q \theta^{(a)}) + p(O_q \theta^{(f)})} \frac{\partial p(O_q \theta^{(f)})}{\partial \theta^{(f)}}$$

20

【 0 0 7 9 】

項 $p(\cdot)$ は、全ての立証モデルに共通であり、更新されたモデルの特定のパラメータだけが異なる。この確率 $p(\cdot)$ は、以下のように定義される。

【数 1 8】

$$p(O_q \theta) = \sum_{i_1 \dots i_q} b_{i_1}(O_1) \cdot a_{i_1 i_2} \cdot b_{i_2}(O_2) \cdot \dots \cdot b_{i_q}(O_q)$$

30

【 0 0 8 0 】

HMM更新プロセッサ60により提供されるパラメータ更新は以下の通りである。

(1) 平均更新

【数 1 9】

$$\mu_{i,j,k,d}(n+1) = \dots \quad (13)$$

40

$$\mu_{i,j,k,d}(n) - \epsilon_n \sum_{n \rightarrow j}^{j+1} [(\partial l(O, \theta) / \partial \log b_{i,j}(o_{nn})) Y_{i,j,k}(o_{nn}) (o_{nn}(d) - \mu_{i,j,k,d}(n))],$$

ここで、 $O_{nn}(d)$ は、特徴ベクトル O_{nn} の d 次元要素であり、 j は、最適パスが最初に状態 j に入るタイムフレームに対応し、 $j+1$ は、最適パスが状態 $j+1$ に入るタイムフレームに対応し、 $l(O, \theta)$ は (7) に従って構成されている損失関数であり、 n はステップサイズで有り、

【数 2 0】

50

$$Y_{i,j,k}(O_{nn}) = (c_{i,j,k}(2\pi)^{-D/2} (\prod_{d=1}^D \sigma_{i,j,k,d})^{-1} \prod_{d=1}^D e^{-(O_{nn}(d) - \mu_{i,j,k,d}) / 2\sigma_{i,j,k,d}}) / b_{i,j}(O_{nn}), \quad (14)$$

ここで、Dは特徴ベクトルの次元である。平均更新は、図4のブロック62により提供される。

【0081】

(2) 分散更新

$\sigma^2 = \log^2$ の更新は、以下の公式に従って行われる。

【数21】

$$\sigma_{i,j,k,d}^2(n+1) = \sigma_{i,j,k,d}^2(n) - \epsilon_n \sum_{m=0}^{j+1-1} (\partial l(O, \theta) / \partial \log b_{i,j}(O_{nn})) Y_{i,j,k}(O_{nn}) \quad (15)$$

$$\cdot ((O_{nn}(d) - \mu_{i,j,k,d})^2 / 2\sigma_{i,j,k,d}^2(n) - 0.5),$$

ここで、 $Y_{i,j,k}(O_{nn})$ は、(14)に示されるように定義される。従って、時間N+1における分散は以下の通りである。

【数22】

$$\sigma_{i,j,k,d}^2(n+1) = e^{\sigma_{i,j,k,d}^2(n+1)} \quad (16)$$

分散は、 10^{-6} で下にクリップされ、制約条件 $\sigma_{i,j,k,d}^2(n) > 10^{-6}$ を満足する。分散更新は、図4のブロック64により提供される。

【0082】

(3) 混合重み更新

パラメータ化された混合重みは、以下の式(17)に従って調整される。

【数23】

$$c_{i,j,k}(n+1) = c_{i,j,k}(n) - \epsilon_n \sum_{m=0}^{j+1-1} (\partial l(O, \theta) / \partial \log b_{i,j}(O_{nn})) \quad (17)$$

$$\cdot [c_{i,j,k}(n) / b_{i,j}(O_{nn})] [N(O_{nn}, \mu_{i,j,k,d}, V_{i,j,k,d}) - b_{i,j}(O_{nn})].$$

従って、時間n+1における混合重みは、式(18)により与えられる。

【数24】

$$c_{i,j,k}(n+1) = \frac{e^{c_{i,j,k}(n+1)}}{\sum_{l=1}^L e^{(c_{i,j,t}(n+1))}} \quad (18)$$

従って、調整された混合重みは、本発明の訓練プロセスの間に以下の制約条件に合致する。

$\sum_k C_{i,j,k}(n) = 1$ と $C_{i,j,k}(n) > 0$
 混合重み更新は、図4のブロック66により提供される。

【0083】

(4) 状態遷移確率更新

左から右へのHMMにおいて、1番目の単語モデルのパラメータ化された遷移確率は以下のように調整される。

【数25】

$$\bar{a}_{i,i}^1(n+1) = \bar{a}_{i,i}^1(n) - e_n [\partial l(O, \theta) / \partial g_1] \quad (19)$$

$$\left(N_{i,i}^1 - (N_{i,i}^1 + 1) \left[\frac{e^{a_{i,i}^1(n)}}{e^{a_{i,i}^1(n)} + e^{a_{i,i+1}^1(n)}} \right] \right)$$

10

$$\bar{a}_{i,i+1}^1(n+1) = \bar{a}_{i,i+1}^1(n) - e_n [\partial l(O, \theta) / \partial g_1]$$

$$\left(1 - (1 + N_{i,i}^1) \left[\frac{e^{a_{i,i+1}^1(n)}}{e^{a_{i,i}^1(n)} + e^{a_{i,i+1}^1(n)}} \right] \right)$$

ここで、 g_1 は1番目の立証モデルのOの尤度スコアであり、 i は終了状態であり、状態 i 内の自己遷移の全体数は以下の式により示される。

【数26】

$$N_{i,i}^1 = \{ \# \text{ of } t | (\omega_{t-1}, \omega_t) = (i, i) \text{ in } l\text{-th verification model } \lambda_l \}$$

(20)

20

結論として、時間 $(n+1)$ における遷移確率は、以下のように与えられる。

【数27】

$$a_{i,i}^1(n+1) = \left[\frac{e^{a_{i,i}^1(n+1)}}{e^{a_{i,i}^1(n+1)} + e^{a_{i,i+1}^1(n+1)}} \right] \quad (21)$$

30

$$a_{i,i+1}^1(n+1) = \left[\frac{e^{a_{i,i+1}^1(n+1)}}{e^{a_{i,i+1}^1(n+1)} + e^{a_{i,i+2}^1(n+1)}} \right],$$

これはまた、 $\sum_j a_{i,j}^1(n) = 1$ 、 $a_{i,i}^1 > 0$ と、 $a_{i,i+1}^1 > 0$ の制約条件を満足する。状態遷移確率更新は図4のブロック68により提供される。

【0084】

HMM更新プロセッサ60により計算された更新表現(15-18)は、式(22)に従って、 j 番目の状態の i 番目の立証モデルの更新された観察確率密度関数と関係付けられる。

【数28】

$$b_{i,j}(o) = \sum_{k=1}^M c_{i,j,k} N(O, \mu_{i,j,k}, V_{i,j,k}) \quad (22)$$

40

ここで、 $c_{i,j,k}$ は混合重みであり、 $N(O, \mu_{i,j,k}, V_{i,j,k})$ は、 i 番目の認識ユニットモデル、 j 番目の状態、対角共分散マトリクス $V_{i,j,k}$ との k 番目の混合の D 次元の通常のガウス分布である。ブロック62、64、66から出力される $\mu(n+1)$ 、 $V(n+1)$ 及び $c(n+1)$ の値は、ブロック68によ

50

り提供される状態遷移確率 $a(n+1)$ に対する値と共に、HMMパラメータデータベース44に戻される。図4に例示するように、 μ は、 μ 、 σ^2 、 c 及び a のための更新された値を具備する。

【0085】

上に提供されるHMMパラメータ更新のための表現は、HMMパラメータデータベース44からの単一音声訓練ストリング信号の使用を関連づけHMMパラメータを強化する。しかしながら、HMMパラメータを改善するために、強化された訓練器48が複数回繰り返される。例えば、強化された訓練器48は、完全なHMMパラメータデータベースを通るいくつかのパスまたは繰返し動作してもよい。強化された訓練器は、いくつかの方法のうちの処理サンプルを止めるようにされてもよい。例えば、強化された訓練器48は、サンプルのHMMパラメータデータベースを通して固定数の繰返しが行われるときには、停止してもよい。強化された訓練器は、現在の繰返しの全ての音声訓練ストリング信号を処理するため、ストリング立証エラーレートの増加的改良の記録も維持し得る。増分的改良がしきい値以下に落ちるとき、処理は停止してもよい(図4の70、及びSTOP信号を参照)。

10

【0086】

MSVE訓練は、実行され、好適音声認識システム実施例に適用されることができ、サブ単語特定キーワードおよび反キーワードのモデルパラメータを設計する。各訓練では、式(1)により定義された識別関数 $S_i(O; \theta)$ が、現在の立証モデルに基づいて既知の仮定ストリング信号のための信用測定信号値に設定され、式(6)により定義される反識別関数 $S_i(O; \theta)$ は、 N 最良ストリング仮定を用いて1つ以上の競合ストリングモデルに基づいて既知の仮定ストリング信号に対する信用測定信号値により近似される。誤立証測定信号は、式(5)におけるように決定される。訓練場に対する誤立証測定と信用測定信号値の分布は、メモリーに記録されて、臨界しきい値を決定するために利用される。

20

【0087】

発声立証システムの性能は、適切な臨界しきい値の選択により強化できる。しきい値は、最小合計エラーレートを得るか、または誤り拒絶エラー(タイプI)が誤り受け入れエラー(タイプII)に等しくなる等しいエラーレートを達成するために、あらかじめ決められた基準に従ってセットできる。立証しきい値は、誤り拒絶エラー(タイプI)と誤り受け入れエラー(タイプII)の間に望ましいトレードオフを得るために初めに設定されてもよい。各エラーレートは、訓練データとテストデータの間の環境の不均衡によりもたらされるかもしれない。環境の不均衡は、異なるアクセントをもつ異なる話者、または異なる周辺ノイズも含んでいる。

30

【0088】

タイプIとタイプIIのエラーの両方のための等しい立証エラーレートは、本発明の音声認識システムにより達成できる。しかしながら、等しいエラーレートは、異なる選択されたしきい値信号値に対して変わることができる。異なる選択されたしきい値において、タイプIエラーレートは、タイプIIエラーレートより高いか、またはより低いに違いない。従って、特定の環境条件の下での仮定単語ストリング信号の立証を実行するための演算しきい値信号値は、異なる環境条件に対して適切ではないかもしれない。

40

【0089】

タイプIとタイプIIに対する異なるエラーレートは、商業的アプリケーションにおいて有利であるかもしれず、それでは誤り拒絶エラーが誤り受け入れエラーより有害であるかもしれず、あるいは逆もまた同様かもしれない。本発明は、タイプIとタイプII間の望ましいトレードオフを維持するために、立証プロセッサにより生成される信用測定信号の臨界しきい値を調整するためのメカニズムを提供する。

【0090】

本発明によるしきい値適応化技術は、既知の仮定ストリング信号に回答して生成される信用測定信号値を利用して、立証しきい値信号値を適応化する。しきい値動作点は、仮定ス

50

トリング信号のグループが利用可能になるにつれて、調整される。(式(1)からの)信用測定信号値の分布は、図5に示されるように、正しい仮定ストリング信号72のクラスと正しくない仮定ストリング信号74のために記録される。右の分布72は、単語ストリングが妥当で、正しく(認識されている)分類されているとき発生される信用測定信号を提供する。左の分布74は、無効の単語ストリング、および間違っ認識されていた他のストリングに対して生成された信用測定信号を提供する。

【0091】

2つの分布72と74は、同様な形を持ち、2つの分布の分散はほとんど等しい。この特徴は、しきい値適応化を実行するために、ここで教示されるように、利用される。

【0092】

0 Cであるように X_{H0} を $S(0;)$ の平均信用測定信号値に対応させ、 X_{H1} を要素0が集合Cに属さないように平均信用測定信号値 $S(0;)$ に対応させる。ここで、Cは正しく認識された仮定ストリング信号である。等しいエラーレートあるいは最小全体エラーレートを達成するために、両方の分布に対して等しい分散を仮定して、臨界しきい値 τ_s は、

【数29】

$$\tau_s = [X_{H0} + X_{H1}] / 2 \quad (23)$$

におかれる。

【0093】

しきい値適応化は、演算立証テストの間に2つの分布の間に τ_s を位置決めし、 $X_{H0} - \tau_s$ と $\tau_s - X_{H1}$ のとき τ_s の値を調整するように導かれる。実際、システムが走っている間に2つの分布の実際の瞬間的平均を決定することは難しい。結果として、立証処理は、 X_{H0} と X_{H1} に対するある初期評価値で開始される。これらの値は、十分な数のストリング("R")が受信され、 τ_s が適応化されることが度毎に、調整される。このプロセスは次の通りに実行される。

【0094】

ストリング信用測定信号値の数Rを用いて、 X_{H0} と X_{H1} の値は、評価され、以下のエラー測定を計算するために使用される。

【数30】

$$E_1 = (X_{H0} - \tau_s) - (\tau_s - X_{H1}). \quad (24)$$

図5に示される分布72と74の長い末部のために平均をバイアスすることを回避するために、あらかじめ定義された最大と最小のしきい値を越える信用測定は、クリップされる。さらに、エラー関数は、以下の形のシフトされたシグモイドを用いて、平滑化される。

【数31】

$$E_1 = \frac{1}{1 + \exp\{-\alpha_1 E_1\}} - 0.5, \quad (25)$$

ここで、 α_1 は、平滑化の程度を決定し、0.1に設定された。nストリングを処理した後の臨界しきい値 $\tau_s(n)$ は、以下のように更新される。

【数32】

10

20

30

40

$$\Delta\tau_s(n) = B_1 E_1, \quad (26)$$

ここで、 B_1 はステップサイズであり、 $s_s(n) = s_s(n) - s_s(n-1)$ 。適応の間に、 s_s の値は $s_s(0) \pm 2$ を越えることは許されない。

【0095】

本発明による立証しきい値の適応化の間に、ストリングは、平均を評価するために等しく、十分な情報を確保するために、ランダムな認識と立証命令のために提供される。臨界しきい値は、最初、最小のエラーレート点に設定され、20ストリング毎（即ち、 $R = 20$ ）に調整され、各分布の平均を計算するために十分な、しかし最小の数のストリングの利用可能性を確保する。立証しきい値信号値を調整する方法は、適応化の前後でタイプIとタイプIIのエラーレートの前選択と維持のために提供する。

10

【0096】

話者独立電話ベースの接続される数字データベースは、本発明による実験的な訓練と動作上のテストの際に使用された。1から16桁の長さの範囲の数字ストリングは、異なるフィールドトライアルでのコレクションから抽出された。フィールドトライアルでのコレクションは、異なる環境条件と音響トランスデューサー機器を表した。

【0097】

訓練セットは、16089数字ストリングから構成された。テストセットは、27702数字ストリングから成っていた。発生の約22%は、語彙外の単語、誤り開始、および重要な背景ノイズを含んでいた。3000の音声学上の豊かな文からなる別のデータベースが、フィルモデルを訓練するために提供された。立証結果は、99%を越える正確な拒絶であった。

20

【0098】

上記から、本発明は、発生立証のためのストリングベース最小立証エラー訓練プロセスを提供することは明らかである。立証プロセッサモデルを訓練することへの識別的なアプローチは、発声立証のために使われた目的関数と一致し、予期されるストリング立証エラーレートを最小化することに直接関連する。

【0099】

上記の観点で、本発明は話者依存及び話者独立の音声認識の両方に適用可能であることは当業者には明らかであろう。音声認識の文脈の中で説明されたが、モデルを訓練するために使用される技術は、一般にパターン認識にも適用可能であることも更に明らかであろう。

30

【0100】

上記実施例は、単に図示されたに過ぎず、特許請求の範囲の記載に基づいて定義される本発明の範囲から離れることなく、当業者により種々の変形がなされることが理解されよう。

【図面の簡単な説明】

【図1】本発明によるHMMベース音声認識システムを示す図である。

40

【図2】図1のHMMベース音声認識システムの立証部の詳細な図である。

【図3】図2の強化された訓練器を示す図である。

【図4】図3の強化された訓練器のHMM更新プロセッサを示す図である。

【図5】立証しきい値信号値決定のための信用測定信号値分布を示す図である。

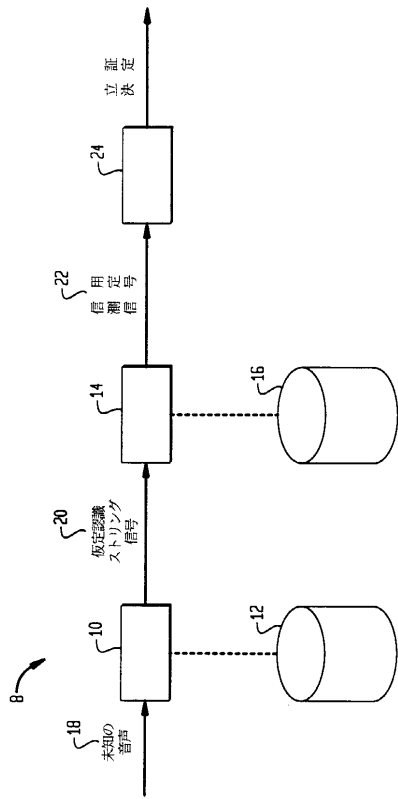
【符号の説明】

- 26 訓練音声データベース
- 28 特徴抽出器
- 30 第1のモードスイッチ
- 32 第2のモードスイッチ
- 34 スコア計算器

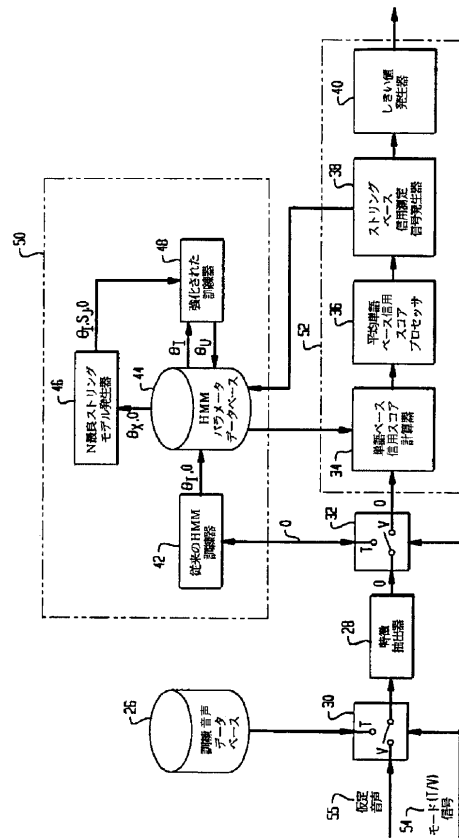
50

- 36 平均単語ベース信用スコアプロセッサ
- 38 ストリングベース信用測定信号発生器
- 40 しきい値比較器
- 42 従来のHMM訓練器
- 44 立証データベース
- 46 N最良ストリングモデル発生器
- 48 強化された訓練器
- 50 システムの訓練部
- 52 システムの立証部

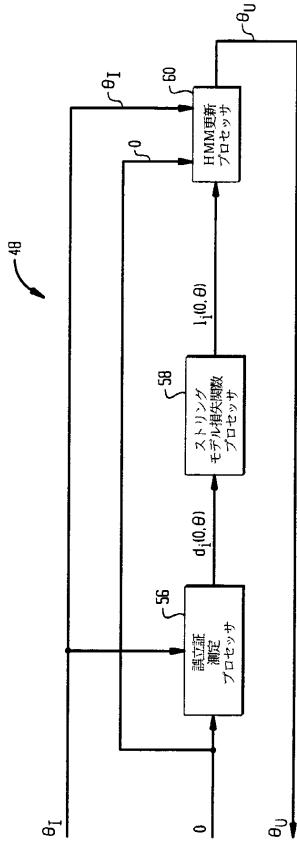
【図1】



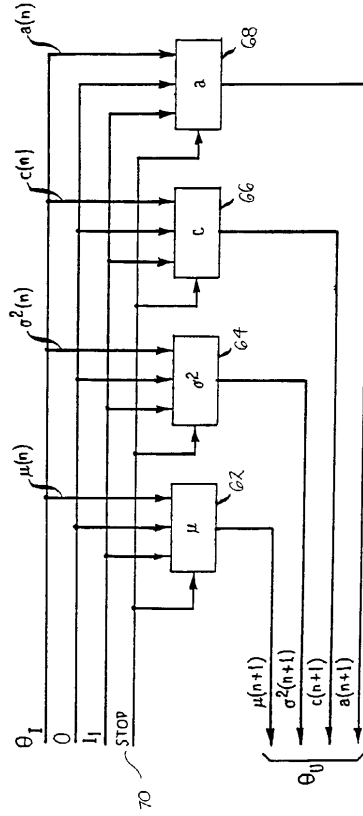
【図2】



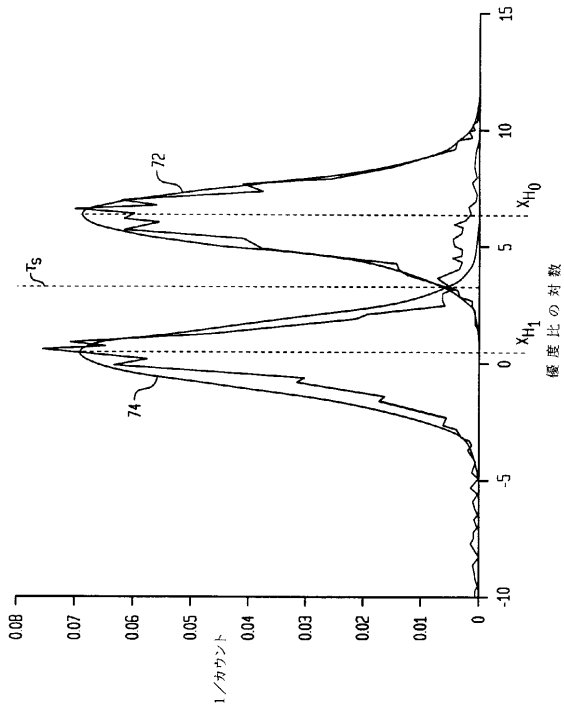
【 図 3 】



【 図 4 】



【 図 5 】



フロントページの続き

- (72)発明者 ビーイング - ホワン ジョアン
アメリカ合衆国 07059 ニュージャージー, ウォーレン, サウス レーン 8
- (72)発明者 チン - フィ リー
アメリカ合衆国 07974 ニュージャージー, ニュープロヴィデンス, ラニーメド パークウ
エイ 118
- (72)発明者 マズイン ジー . レイム
アメリカ合衆国 07726 ニュージャージー, マナラバン, キンパリー コート 31

審査官 榎本 剛

- (56)参考文献 特開平07 - 152394 (JP, A)
特開平08 - 263660 (JP, A)
特開平01 - 302297 (JP, A)
特開平02 - 310663 (JP, A)
Jung-Kuei Chen, Soong, F.K., Discriminative training of high performance speech recogn
izer using N best candidates, IEEE 1994 International Conference on Acoustics, Speech,
and Signal Processing (ICASSP-94.), 米国, 1994年 4月19日, Vol. 1, p.1/625 - 1
/628
Rahim, M.G., Chin-Hui Lee, Biing-Hwang Juang, Robust utterance verification for c
onconnected digits recognition, IEEE 1995 International Conference on Acoustics, Speech,
and Signal Processing (ICASSP-95.), 米国, 1995年 5月 9日, Vol. 1, p.285 - 288
Rahim, M.G., Chin-Hui Lee, Biing-Hwang Juang, Wu Chou, Discriminative utterance verifi
cation using minimum string verification error (MSVE) training, IEEE 1996 Internationa
l Conference on Acoustics, Speech, and Signal Processing (ICASSP-96.), 米国, 1996
年 5月 7日, Vol. 6, p.3585 - 3588
Chou, W., Lee, C.H., Juang, B.H., Minimum error rate training based on N-best string m
odels, IEEE 1993 International Conference on Acoustics, Speech, and Signal Processing
(ICASSP-93.), 米国, 1993年 4月27日, Vol.2, p.652 - 655

- (58)調査した分野(Int.Cl.⁷, DB名)
G10L 15/00-17/00
JICSTファイル(JOIS)
IEEE Xplore