

(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2010-96899

(P2010-96899A)

(43) 公開日 平成22年4月30日(2010.4.30)

(51) Int. Cl.
G 1 O L 15/06

F I

G I O L	15/06	3 0 0 Y
G I O L	15/06	3 0 0 D

テーマコード (参考)
5D015

審査請求 未請求 請求項の数 11 O L (全 20 頁)

(21) 出願番号 特願2008-266288 (P2008-266288)
(22) 出願日 平成20年10月15日 (2008.10.15)

(71) 出願人	000004226 日本電信電話株式会社 東京都千代田区大手町二丁目3番1号
(74) 代理人	100121706 弁理士 中尾 直樹
(74) 代理人	100066153 弁理士 草野 卓
(74) 代理人	100128705 弁理士 中村 幸雄
(72) 発明者	小橋川 哲 東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内
(72) 発明者	山口 義和 東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内

[最終頁に続く](#)

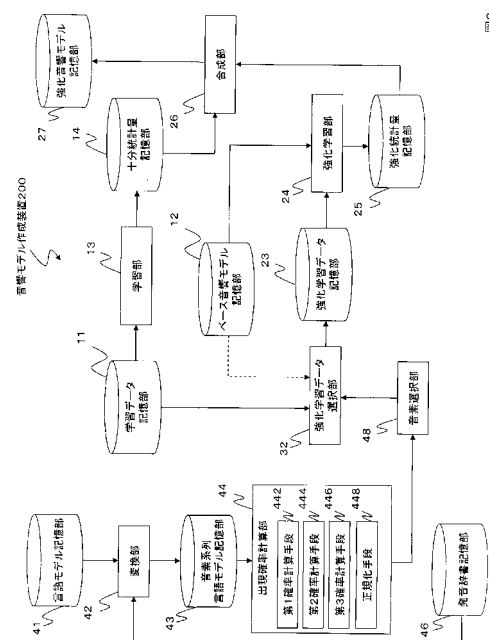
(54) 【発明の名称】 音響モデル作成方法、その装置、プログラム、その記録媒体

(57) 【要約】

【課題】 全体的な認識精度が向上された音響モデルを作成する。

【解決手段】 発音辞書記憶部に記憶されている発音辞書を用いて、言語モデル記憶部に記憶されている言語モデル中の単語を音素系列に変換し、全ての対象音素ごとに、当該対象音素を含む音素系列の出現確率の総和を当該対象音素の音素出現確率を求め、音素出現確率が高い高出現確率音素を選択し、学習データ記憶部から、高出現確率音素を含む音素系列についての学習データを抽出して、強化学習データとし、強化学習データとベース音響モデルから強化統計量を計算し、学習データ記憶部からの学習データについての十分統計量と強化統計量とから強化音響モデルを作成する。

【選択図】 図2



【特許請求の範囲】

【請求項 1】

発音辞書記憶部に記憶されている発音辞書を用いて、言語モデル記憶部に記憶されている言語モデル中の単語を音素系列に変換する変換過程と、

全ての対象音素ごとに、当該対象音素を含む音素系列の出現確率の総和を当該対象音素の出現確率（以下、「音素出現確率」という。）として求める出現確率計算過程と、

音素出現確率が高い対象音素（以下、「高出現確率音素」という。）を選択する音素選択過程と、

学習データ記憶部から、前記高出現確率音素を含む音素系列についての学習データを抽出して、強化学習データとする強化学習データ選択過程と、

前記強化学習データと前記ベース音響モデルから強化統計量を計算する強化学習過程と、

学習データ記憶部からの学習データについての十分統計量と前記強化統計量とから強化音響モデルを作成する合成過程と、

を備える音響モデル作成方法。

【請求項 2】

請求項 1 記載の音響モデル作成方法であって、

前記出現確率計算過程は、前記対象音素を環境独立音素として、前記音素系列中の全ての環境独立音素ごとに、当該環境独立音素を含む単語のユニグラム確率の総和、当該環境独立音素を含む単語のバイグラム確率の総和、当該環境独立音素を含む単語のトライグラム確率の総和、のうち少なくとも 1 つを用いて、音素出現確率を求める第 1 確率計算ステップを有するものであることを特徴とする音響モデル作成方法。

【請求項 3】

請求項 1 または 2 記載の音響モデル作成方法であって、

前記出現確率計算過程は、前記対象音素を片側環境依存音素として、前記音素系列中の全ての片側環境依存音素ごとに、当該片側環境依存音素を含む単語のユニグラム確率の総和、当該片側環境依存音素を含む単語のバイグラム確率の総和、当該片側環境依存音素を含む単語のトライグラム確率の総和、のうち少なくとも 1 つを用いて、音素出現確率を求める第 2 確率計算ステップを有するものであることを特徴とする音響モデル作成方法。

【請求項 4】

請求項 1 ～ 3 何れかに記載の音響モデル作成方法であって、

前記出現確率計算過程は、前記対象音素を両側環境依存音素として、前記音素系列中の全ての両側環境依存音素ごとに、当該両側環境依存音素を含む単語のユニグラム確率の総和、当該両側環境依存音素を含む単語のバイグラム確率の総和、当該両側環境依存音素を含む単語のトライグラム確率の総和、のうち少なくとも 1 つを用いて、音素出現確率を求める第 3 確率計算ステップを有するものであることを特徴とする音響モデル作成方法。

【請求項 5】

請求項 1 ～ 4 何れかに記載の音響モデル作成方法であって、

前記出現確率計算過程は、予め定められた閾値以上の音素数を持つ音素系列の出現確率の総和を演算することを特徴とする音響モデル作成方法。

【請求項 6】

請求項 1 ～ 5 何れかに記載の音響モデル作成方法であって、

前記確率計算過程は、計算された音素出現確率を、対象音素の全ての出現確率で除算することで、正規化音素出現確率を求める正規化ステップを有し、

前記音素選択過程は、正規化音素出現確率が大きい対象音素を選択することを特徴とする音響モデル作成方法。

【請求項 7】

請求項 1 ～ 6 何れかに記載の音響モデル作成方法であって、

更に、音素毎に、前記強化統計量に対して、当該音素の音素出現確率から求まる値を重み係数として、重み付けして、前記強化統計量として出力する統計量強化過程を有するこ

10

20

30

40

50

とを特徴とする音響モデル作成方法。

【請求項 8】

言語モデルを記憶する言語モデル記憶部と、

発音辞書記憶部に記憶されている発音辞書を用いて、前記言語モデル中の単語を音素系列に変換する変換部と、

全ての対象音素ごとに、当該対象音素を含む音素系列の出現確率の総和を当該対象音素の出現確率（以下、「音素出現確率」という。）として求める出現確率計算部と、

音素出現確率が高い対象音素（以下、「高出現確率音素」という。）を選択する音素選択部と、

学習データ記憶部から、前記高出現確率音素を含む音素系列についての学習データを抽出して、強化学習データとする強化学習データ選択部と、

前記強化学習データと前記ベース音響モデルから強化統計量を計算する強化学習部と、

学習データ記憶部からの学習データについての十分統計量と前記強化統計量とから強化音響モデルを作成する合成部と、

を備える音響モデル作成装置。

【請求項 9】

請求項 8 記載の音響モデル作成装置であって、

更に、音素毎に、前記強化統計量に対して、当該音素の音素出現確率から求まる値を重み係数として、重み付けして、前記強化統計量として出力する統計量強化部を有することを特徴とする音響モデル作成装置。

【請求項 10】

請求項 1～7 何れかに記載の音響モデル作成方法の各過程をコンピュータに実行させるためのプログラム。

【請求項 11】

請求項 10 に記載のプログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、入力された音声信号から計算した音響特徴パラメータ系列に対して、各音素の音響的特徴を表現した確率モデルを作成する音響モデル作成方法、装置、プログラム及びその記録媒体に関する。

【背景技術】

【0002】

図 1 に従来 of 音響モデル作成装置 100 の機能構成例を示す。音響モデル作成装置 100 は、変換部 31、強化学習データ選択部 32、学習データ記憶部 11、学習部 13、十分統計量記憶部 14、強化学習データ記憶部 23、強化学習部 24、強化統計量記憶部 25、合成部 26、強化音響モデル記憶部 27 で構成されている。また、各構成部の処理内容、十分統計量、強化統計量については、この「背景技術」では簡単に説明し、以下の「発明を実施するための最良の形態」で詳細に説明する。

【0003】

学習データ記憶部 11 には、音声データとこの音声データに音素等を対応付けたラベルとから構成される学習データが複数予め格納されている。ベース音響モデル記憶部 12 には、学習前の音響モデルであるベース音響モデルが予め格納されている。学習部 13 は、学習データ記憶部 11 から読み出した学習データと、ベース音響モデル記憶部 12 から読み出したベース音響モデルとから、十分統計量を計算して十分統計量記憶部 14 に格納する。

【0004】

一方、強化キーワードリスト記憶部 21 には、所定のキーワードのリストが記憶されている。所定のキーワードとは、例えば、認識性能を改善したいキーワードや、誤認識となったキーワードである。

【0005】

強化学習データ選択部32は、強化キーワードリスト記憶部21から読み出したキーワードと同一及び／又は類似のキーワードを含む学習データを、学習データ記憶部11から抽出して、強化学習データを作成する。作成された強化学習データは、強化学習データ記憶部23に記憶される。

【0006】

強化学習部24は、強化学習データ記憶部23から読み出した強化学習データから、強化統計量を計算する。計算された強化統計量は、強化統計量記憶部25に格納される。強化統計量は、強化学習データ記憶部23から読み出した強化学習データに基づいて計算された十分統計量のことである。

10

【0007】

合成部26は、十分統計量記憶部14から読み出した十分統計量と、強化統計量記憶部25から読み出した強化統計量に重みをかけたものから強化音響モデルを作成する。作成された強化音響モデルは、強化音響モデル記憶部27に格納される。なお、音響モデル作成装置100の詳細については、特許文献1、非特許文献1に記載されている。

【特許文献1】特開2008-129527号公報

【非特許文献1】小橋川哲、小川厚徳、政瀧浩和、高橋敏、「キーワードに関する十分統計量増強による精度向上の検討」、日本音響学会講演論文集、2008年3月、pp213-214

【発明の開示】

20

【発明が解決しようとする課題】

【0008】

従来の音響モデル作成装置は、キーワードに含まれる音素のみの認識精度を向上させるものであった。従って、全体的な単語や音素の認識精度の向上を行うことができないという問題があった。また、強化するキーワードリストが必要であるという問題があった。

この発明は、キーワードリストを必要とすることなく、全体的な単語や音素の認識精度の向上を行う音響モデル作成方法、音響モデル作成装置、そのプログラム、記録媒体を提供するものである。

【課題を解決するための手段】

【0009】

30

この発明の音響モデル作成方法は、変換過程と、出現確率計算過程と、音素選択過程と、強化学習データ選択過程と、強化学習過程と、合成過程と、を有する。変換過程は、発音辞書記憶部に記憶されている発音辞書を用いて、言語モデルに記憶されている言語モデル中の単語を音素系列に変換する。出現確率計算過程は全ての対象音素ごとに、当該対象音素を含む音素系列の出現確率の総和を当該対象音素の音素出現確率として求める。音素選択過程は、音素出現確率が高い高出現確率音素を選択する。強化学習データ選択過程は、学習データ記憶部から、前記高出現確率音素を含む音素系列についての学習データを抽出して、強化学習データとする。強化学習過程は、前記強化学習データと前記ベース音響モデルから強化統計量を計算する。合成過程は、学習データ記憶部からの学習データについての十分統計量と前記強化統計量とから強化音響モデルを作成する。

40

【発明の効果】

【0010】

この発明によれば、キーワードリストではなく、言語モデルを用いるので、当該言語モデルに含まれている全体的な単語や音素の認識精度の向上を行う。また、キーワードリストを必要としないという効果も有する。

【発明を実施するための最良の形態】

【0011】

以下に、発明を実施するための最良の形態を示す。なお、同じ機能を持つ構成部や同じ処理を行う過程には同じ番号を付し、重複説明を省略する。

【実施例1】

50

【0012】

図2に実施例1の音響モデル作成装置200の機能構成例を示し、図3に処理フローを示す。この実施例の音響モデル作成装置200は、言語モデル記憶部41、変換部42、音素系列言語モデル記憶部43、出現確率計算部44、発音辞書記憶部46、音素選択部48、学習データ記憶部11、強化学習データ記憶部23、ベース音響モデル記憶部12、学習部13、強化学習部24、強化統計量記憶部25、合成部26、強化音響モデル記憶部27、とで構成されている。

【0013】

まず、言語モデル記憶部41には、予め言語モデルを記憶させる。ここで言語モデルについて説明する。この音響モデル作成装置200から出力される（後述する強化音響モデル記憶部27に記憶される）強化音響モデルを用いて、音声認識処理を行うのであるが、当該音声認識処理に係る分野（例えば、電話番号案内）の言語的な出現頻度を反映した言語モデルを用いる。

【0014】

図4に言語モデルの一例を示す。この例では、言語モデルは、Nグラムで表されとし、Nグラムの例として、ユニグラム（uni gramまたは1-gram）、バイグラム（bi gramまたは2-gram）、トライグラム（tri gramまたは3-gram）であるとし、つまり、単語の頻度（出現確率）が、ユニグラム確率、バイグラム確率、トライグラム確率により表されているとする。また、言語モデルには、複数の文章が格納されているが、当該複数の文章のうちの一文の例を「私は発明をします」とし、「私／は／発明／を／します」というように、単語で区切られているとする。以下では、単語とは、1つの単語（例えば、「私」）を示し、単語系列とは、1つの単語または1つ以上の単語の組（例えば、「私→は」や「私→は→発明」等）を示す。また、図4の例では、例えば「私」のユニグラム確率は0.0001とし、「私→は」のバイグラム確率は0.000008とし、「私→は→発明」のトライグラム確率は0.0000004とする。

【0015】

変換部42は、発音辞書記憶部46に記憶されている発音辞書を用いて、言語モデル中の単語を音素系列に変換する（ステップS2）。この例では、発音辞書は、「単語」と「その単語の読み」とが対応されて格納されているものである。図5に発音辞書の一例を示す。図5に示すように、例えば、単語「私」とその単語の読み「わたし」とが対応された発音辞書が、発音辞書記憶部46に格納されている。変換部42は言語モデル中の全ての単語について、読みに変換される。そして変換部42は、変換部42自身が保持している文節表を用いて、この読みを音素系列に変換する。文節表については、例えば「鹿野清宏 他著「IT TEXT 音声認識システム」、2001年 p45の表3.1」に記載されている。このようにして言語モデル中の全ての単語が、読み→音素系列、に変換され、音素系列言語モデルとして音素系列言語モデル記憶部43に記憶される。図6に音素系列言語モデルの一例を示す。図6に示すように、例えば、音素系列「watashi」と出現確率「0.00010」とが対応して記憶されている。また、図6の記載「*」は、単語と他の単語の区切りを示す。

【0016】

次に、出現確率計算部44は、全ての対象音素ごとに、当該対象音素を含む音素系列の出現確率の総和を当該対象音素の出現確率（以下、「音素出現確率」という。）として求める（ステップS4）。ここで、音素系列とは、1つの音素、または、2つ以上の音素の系列を意味する。また、対象音素とは、音素出現確率を求める対象となる音素であり、存在する全ての音素をいう。そして、対象音素は、環境独立音素（monophone）、片側環境依存音素（biphone）、両側環境依存音素（triphone）などがある。例えば、対象音素を環境独立音素とすると、対象音素はa、i、u・・・であり、音素体系にも依るが、約30～50個程度存在する。対象音素の数を Ω （例えば、30～50）個とし、対象音素を片側環境依存音素とすると、対象音素は $a-a+*$ 、 $a-i+*$

10

20

30

40

50

、．．．や $*-a+a$ 、 $*-a+i$ 、などであり、約 $\Omega \times \Omega \times 2$ 個存在する。また、対象音素が、両側環境依存音素 (triphone) とすると、対象音素は $a-a+a$ 、 $a-a+i$ 、．．．や $n-a+a$ 、 $n-a+i$ 、．．．等があり、約 $\Omega \times \Omega \times \Omega$ 個存在する。この発明では、言語モデルをNモデル、対象音素をNphone (monophoneやbiphone) として、出現確率計算部44は演算可能であるが、この実施例1では対象音素を環境独立音素として、説明をする。

【0017】

対象音素を環境独立音素とした場合には、出現確率計算部44は第1確率計算手段442を有する。第1確率計算手段442は、音素系列中の全ての環境独立音素ごとに、当該環境独立音素を含む単語のユニグラム確率の総和、当該環境独立音素を含む単語のバイグラム確率の総和、当該環境独立音素を含む単語のトライグラム確率の総和、のうち少なくとも1つを用いて、音素出現確率を求める。また、以下の説明では、対象音素の一般的な表記を「 a 」とし、音素 a の出現確率を $P(a)$ とし、音素 a を含む単語のユニグラム確率の総和を $\Sigma uni(a)$ とし、音素 a を含む単語のバイグラム確率の総和を $\Sigma bi(a)$ とし、音素 a を含む単語のトライグラム確率の総和を $\Sigma tri(a)$ とする。そうすると、音素 a の出現確率は以下の式で表すことができる。

$$P(a) = w_1 \Sigma uni(a) + w_2 \Sigma bi(a) + w_3 \Sigma tri(a) \quad (1)$$

ここで、 w_1 、 w_2 、 w_3 はそれぞれ、重み係数を示す。状況に応じて、重み係数 w_1 、 w_2 、 w_3 を調整する。例えば、 $w_1 = 1$ 、 $w_2 = w_3 = 0$ とすると、ユニグラム確率のみを使用して求めることができる。対象音素 a を例えば音素「 a 」とし、重み係数を $w_1 = 1$ 、 $w_2 = w_3 = 0$ （つまり、ユニグラム確率のみ用いる）とし、図5記載の音素系列言語モデルを用いる場合を考える。そうすると、ユニグラム確率に記載されている「 a 」を含む単語は、図5記載の音素系列の単語では、

「 $watashi$ (出現確率は0.0001)」

「 wa (出現確率は0.00025)」

「 $hatsumei$ (出現確率は0.00002)」

「 $shimasu$ (出現確率は0.00005)」

である。そして、これらの音素系列の出現確率の合計（つまり、 $0.0001 + 0.00025 + 0.00002 + 0.00005$ ）が対象音素「 a 」の音素出現確率として、出現確率計算部44から出力される。このように、出現確率計算部44は、その他の全ての環境独立音素を対象音素として「 i 」「 u 」・・・の音素出現確率を求める。図7に出現確率計算部44から出力される音素出現確率の例を示す。図7に記載のように、全ての対象音素と、当該対象音素の音素出現確率と、が対応されたものが、音素選択部48に入力される。図7の記載では、「 $w-a*$ 」「 $h-a+t$ 」の音素出現確率が記載されているが、これらは、それぞれ対象音素を両側環境依存音素、片側環境依存音素にした場合の音素出現確率である（それぞれ実施例3、実施例2で説明）。

【0018】

音素選択部48は、音素出現確率が大きい対象音素（以下、「高出現確率音素」という。）を選択する（ステップS6）。音素出現確率の大小の判定手法例として、予め音素出現確率についての閾値 Th_1 を定め、音素出現確率が閾値 Th_1 より大きい対象音素のみを選択する。選択された高出現確率音素は強化学習データ選択部32に入力される。

【0019】

一方、学習データ記憶部11には、音声データとこの音声データに音素等を対応付けたラベルとから構成される学習データが複数予め格納されている。ベース音響モデル記憶部12には、学習前の音響モデルであるベース音響モデルが予め格納されている。学習部13は、学習データ記憶部11から読み出した学習データと、ベース音響モデルから読み出したベース音響モデルとから、十分統計量を計算して十分統計量記憶部14に記憶させる。

【0020】

以下、十分統計量について説明する。十分統計量とは、HMMを特徴付けるパラメータ

であり、遷移確率に関わる統計量、混合正規分布の混合重み、平均、分散に関わる統計量等である。具体的には、音響モデル λ と、観測系列 $\mathbf{O}$ が与えられたとき、時刻 t に状態 i に存在し、時刻 $t+1$ に状態 j に存在する確率 $\xi_t(i, j)$ 、また音響モデル λ と観測系列 $\mathbf{O}$ が与えられ、状態は N 個の状態から構成されるとき、時刻 t に状態 i に存在する確率 $\gamma_t(i)$ をそれぞれ、

【数 1】

$$\xi_t(i, j) = P(q_t = i, q_{t+1} = j | \mathbf{O}, \lambda) \quad (3)$$

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (4)$$

10

とする。また、式中のベクトル、行列を太字で記載するものとし、文中のベクトル、行列を標準書体で記載するものとする。このとき、音響モデルパラメータである、時刻 $t=1$ に状態 i に存在すると期待される頻度（回数） π_i 、状態遷移確率 a_{ij} 、状態 j で特徴ベクトル $\mathbf{v}_k$ が観測される確率 $b_j(k)$ 、状態 i の m 番目の混合分布に対する混合重み係数 c_{im} 、状態 i の m 番目の混合分布の平均ベクトル $\boldsymbol{\mu}_{im}$ 、共分散行列 $\mathbf{U}_{im}$ は、例えば、それぞれ次のように定義される。

【数 2】

$$\pi_i = \gamma_1(i)$$

20

$$a_{ij} = \frac{\sum_{t=1}^T \xi_t(i, j)}{\sum_{t=1}^T \gamma_t(i)}$$

$$b_j(k) = \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j) \text{ s.t. } o_t = v_k}$$

30

$$c_{im} = \frac{\sum_{t=1}^T \gamma_t(i, m)}{\sum_{t=1}^T \sum_{m=1}^M \gamma_t(i, m)}$$

$$\boldsymbol{\mu}_{im} = \frac{\sum_{t=1}^T \gamma_t(i, m) \cdot \mathbf{o}_t}{\sum_{t=1}^T \gamma_t(i, m)}$$

40

$$\mathbf{U}_{im} = \frac{\sum_{t=1}^T \gamma_t(i, m) \cdot (\mathbf{o}_t - \boldsymbol{\mu}_{im})(\mathbf{o}_t - \boldsymbol{\mu}_{im})'}{\sum_{t=1}^T \gamma_t(i, m)}$$

50

ここで、プライムはベクトルの転置を表し、 $\gamma_t(i, m)$ は、観測された観測信号の特徴ベクトル $\mathbf{o}_t$ を考慮したときに、時刻 t に状態 i の m 番目の混合要素に存在する確率である。

【0021】

このとき、前記式の分子と分母に記載された各数式で定義される値がそれぞれ十分統計量となる。すなわち、 a_{ij} については、 $\sum_{t=1}^T \xi_t(i, j)$ と、 $\sum_{t=1}^T \gamma_t(i)$ とがそれぞれ十分統計量となる。また、 $b_j(k)$ については、 $\sum_{t=1}^T (\mathbf{s}_t \cdot \mathbf{o}_t = v_k) \gamma_t(j)$ と、 $\sum_{t=1}^T \gamma_t(j)$ とがそれぞれ十分統計量となる。 c_{im} 、 μ_{im} 、 U_{jk} についても同様である。なお、 π_i については、 $\gamma_1(i)$ が十分統計量となる。これらの十分統計量は、学習データ記憶部 11 から読み出した学習データを用いて、例えば、Baum-Weilch アルゴリズム等の既存のアルゴリズムを用いて推定することができる。Baum-Weilch アルゴリズムは、文献「Lawrence Rabiner, Biing-Hwang Juang 共著、古井貞熙監訳、「音声認識の基礎（下）」、NTTアドバンステクノロジー、1995年」に詳しい。なお、十分統計量は、学習データ記憶部 11 に記憶された学習データ毎に計算される。

10

【数 3】

なお、状態毎に

$$\sum_{t=1}^T \xi_t(i, j) \quad (\text{上記 } a_{ij} \text{ の右辺の分子の式})$$

20

$$\sum_{t=1}^T \gamma_t(i, j) \quad (\text{上記 } a_{ij} \text{ の右辺の分母の式}) \text{ を保存し、}$$

分布 i 、 m 毎に、

$$\sum_{t=1}^T \gamma_t(i, m) \quad (\text{上記 } c_{im} \text{ の右辺の分子の式})$$

$$\sum_{t=1}^T \gamma_t(i, j) \cdot \mathbf{o}_t \quad (\text{上記 } \mu_{im} \text{ の右辺の分子の式})$$

$$\sum_{t=1}^T \gamma_t(i, j) \cdot \mathbf{o}_t \cdot \mathbf{o}_t' \quad (\text{上記 } U_{im} \text{ の右辺の分子の式の一部の項})$$

30

を保存していれば、各種統計量は計算により求めることができる。

【0022】

一方、強化学習データ選択部 32 は、学習データ記憶部 11 から高出現確率音素と同一の音素を含む音素系列についての学習データを抽出して、強化学習データとして強化学習データ記憶部 23 に記憶させる（ステップ S8）。ここで音素系列とは、音素系列（例えば、 $a-i+u$ ）の他に、単一の音素（例えば、 a ）も含むものとする。そして、強化学習部 24 は、強化学習データ記憶部 23 から読み出した強化学習データから、強化統計量を計算する（ステップ S10）。計算された強化統計量は、強化統計量記憶部 25 に格納される。強化統計量は、強化学習データ記憶部 23 から読み出した強化学習データに基づいて計算された十分統計量のことである。すなわち、十分統計量と強化統計量は、数式で表現した定義は同じであるが、十分統計量は学習データ記憶部 11 から読み出した学習データに基づいて計算されたものであるのに対し、強化統計量は強化学習データ記憶部 23 から読み出した強化学習データに基づいて計算されたものである点において異なる。強化統計量の計算方法は、学習部 13 によるものと同一なので省略する。

40

【0023】

また、強化学習データ選択部 32 は、ベース音響モデル記憶部 12 から読み出したベース音響モデルにおいて、高出現確率音素とベース音響モデル内の音素との分布間距離を求め、当該分布間距離の総和が閾値 Th_2 以上であるかどうかで判断してもよい。すなわち

50

、分布間距離の和を求めることにより状態間距離を求め、その状態間距離の和を求めることにより音素間の距離を求め、その音素間の距離から類似の判断を行う。

【 0 0 2 4 】

合成部 2 6 は、十分統計量記憶部 1 4 から読み出した十分統計量と、強化統計量記憶部 2 5 から読み出した強化統計量に重みをかけたものから強化音響モデルを作成する。作成された強化音響モデルは、強化音響モデル記憶部 2 7 に格納される。

【 0 0 2 5 】

学習部 1 3 により複数の学習データから G 個の十分統計量が得られたとする。このとき、十分統計量は、 $g = 1, \dots, G$ として、

【数 4】

10

$$\gamma_1^g(i)$$

$$\sum_{t=1}^{T^g} \xi_t^g(i, j) a_{ij}$$

$$\sum_{t=1}^{T^g} \gamma_t^g(i)$$

$$\sum_{t=1}^{T^g} \gamma_t^g(j)$$

$$s.t. \ o_t = v_k$$

$$\sum_{t=1}^{T^g} \gamma_t^g(j)$$

$$\sum_{t=1}^{T^g} \gamma_t^g(i, m)$$

$$\sum_{t=1}^{T^g} \sum_{m=1}^M \gamma_t^g(i, m)$$

$$\sum_{t=1}^{T^g} \gamma_t^g(i, m) \cdot \mathbf{o}_t$$

$$\sum_{t=1}^{T^g} \gamma_t^g(i, m) \cdot (\mathbf{o}_t - \boldsymbol{\mu}_{im})(\mathbf{o}_t - \boldsymbol{\mu}_{im})'$$

20

30

40

と表現される。

【 0 0 2 6 】

また、強化学習部 2 4 は複数の強化学習データから H 個の強化統計量が得られたとする。このとき、強化統計量は、 $h = 1, \dots, H$ として、

【数 5】

$$\begin{aligned}
& \gamma'_1{}^h(i) \\
& \sum_{t=1}^{T^h} \xi'_t{}^h(i, j) a_{ij} \\
& \sum_{t=1}^{T^h} \gamma'_t{}^h(i) \\
& \sum_{t=1}^{T^h} \gamma'_t{}^h(j) \\
& s.t., o_t = v_k \\
& \sum_{t=1}^{T^h} \gamma'_t{}^h(j) \\
& \sum_{t=1}^{T^h} \gamma'_t{}^h(i, m) \\
& \sum_{t=1}^{T^h} \sum_{m=1}^M \gamma'_t{}^h(i, m) \\
& \sum_{t=1}^{T^h} \gamma'_t{}^h(i, m) \cdot \mathbf{o}_t \\
& \sum_{t=1}^{T^h} \gamma'_t{}^h(i, m) \cdot (\mathbf{o}_t - \boldsymbol{\mu}_{im})(\mathbf{o}_t - \boldsymbol{\mu}_{im})'
\end{aligned}$$

10

20

30

と表現される。

【0027】

このとき、合成部26は、下式で各音響パラメータを計算することにより、強化音響モデルを作成する（ステップS12）。

【数 6】

$$\begin{aligned}
\pi_i &= \sum_{g=1}^G \gamma_i^g(i) + \sum_{h=1}^H w^h \gamma_i^h(i) \\
a_{ij} &= \frac{\sum_{g=1}^G \sum_{t=1}^{T^g} \xi_t^g(i, j) + \sum_{h=1}^H w^h \sum_{t=1}^{T^h} \xi_t^h(i, j)}{\sum_{g=1}^G \sum_{t=1}^{T^g} \gamma_t^g(i) + \sum_{h=1}^H w^h \sum_{t=1}^{T^h} \gamma_t^h(i)} \\
b_i(k) &= \frac{\sum_{g=1}^G \sum_{t=1}^{T^g} \gamma_t^g(j) + \sum_{h=1}^H w^h \sum_{t=1}^{T^h} \gamma_t^h(j)}{\sum_{g=1}^G \sum_{t=1}^{T^g} \gamma_t^g(j) + \sum_{h=1}^H w^h \sum_{t=1}^{T^h} \gamma_t^h(j)} \\
c_{im} &= \frac{\sum_{g=1}^G \sum_{t=1}^{T^g} \gamma_t^g(i, m) + \sum_{h=1}^H w^h \sum_{t=1}^{T^h} \gamma_t^h(i, m)}{\sum_{g=1}^G \sum_{m=1}^M \gamma_t^g(i, m) + \sum_{h=1}^H w^h \sum_{t=1}^{T^h} \sum_{m=1}^M \gamma_t^h(i, m)} \\
\mu_{im} &= \frac{\sum_{g=1}^G \sum_{t=1}^{T^g} \gamma_t^g(i, m) \cdot \mathbf{o}_t + \sum_{h=1}^H w^h \sum_{t=1}^{T^h} \gamma_t^h(i, m) \cdot \mathbf{o}_t}{\sum_{g=1}^G \sum_{t=1}^{T^g} \gamma_t^g(i, m) + \sum_{h=1}^H w^h \sum_{t=1}^{T^h} \gamma_t^h(i, m)} \\
\mathbf{U}_{im} &= \frac{\sum_{g=1}^G \sum_{t=1}^{T^g} \gamma_t^g(i, m) \cdot (\mathbf{o}_t - \mu_{im})(\mathbf{o}_t - \mu_{im})' + \sum_{h=1}^H w^h \sum_{t=1}^{T^h} \gamma_t^h(i, m) \cdot (\mathbf{o}_t - \mu_{im})(\mathbf{o}_t - \mu_{im})'}{\sum_{g=1}^G \sum_{t=1}^{T^g} \gamma_t^g(i, m) + \sum_{h=1}^H w^h \sum_{t=1}^{T^h} \gamma_t^h(i, m)}
\end{aligned}$$

10

20

30

40

50

ここで、 w^h ($h = 1, \dots, H$) は、強化統計量にかけられる重みであり 1 以上の数である。重み w^h ($h = 1, \dots, H$) は、同一の値であってもよいし、互いに異なる値であってもよい。求められた強化音響モデルは強化音響モデル記憶部 27 に記憶される。

【0028】

また、学習部 13 を具備しない構成とすることが出来る。当該構成とするには、学習データ記憶部 11 に格納されている学習データについての十分統計量を別途の学習により求めておき、予め格納させておけばよい。

【0029】

このように、音声認識処理に係る分野の言語的な出現頻度を反映した言語モデルを用いて、言語モデルに含まれる単語の出現確率（頻度）を利用して擬似的な音素の出現確率を求めることで、ベース音響モデルの全体的な単語や音素の認識精度の向上を行うことが出来る。また、従来のようにキーワードリストを用いる必要もない。

【実施例 2】

【0030】

実施例 1 の出現確率計算部 44 は第 1 確率計算手段を備え、対象音素を環境独立音素として、対象音素の音素出現確率を求めた。この実施例 2 の出現確率計算部 44 は、対象音

素が両側環境依存音素 (t r i p h o n e 例えば a - i + i) として、音素出現確率を求める。この実施例 2 の出現確率計算部 4 4 は、第 3 確率計算手段 4 4 6 を有する。第 3 確率計算手段 4 4 6 は、対象音素を両側環境依存音素として、音素系列中の全ての両側環境依存音素ごとに、当該両側環境依存音素を含む単語のユニグラム確率の総和、当該両側環境依存音素を含む単語のバイグラム確率の総和、当該両側環境依存音素を含む単語のトライグラム確率の総和、のうち少なくとも 1 つを用いて、音素出現確率を求める。

【0031】

両側環境依存音素の一般的な表記を「 $\alpha - \beta + \gamma$ 」とし、対象音素「 $\alpha - \beta + \gamma$ 」の出現確率を $P(\alpha - \beta + \gamma)$ とし、音素「 $\alpha - \beta + \gamma$ 」を含む単語のユニグラム確率の総和を $\Sigma uni(\alpha - \beta + \gamma)$ とし、音素「 $\alpha - \beta + \gamma$ 」を含む単語のバイグラム確率の総和を $\Sigma bi(\alpha - \beta + \gamma)$ とし、音素 α を含む単語のトライグラム確率の総和を $\Sigma tri(\alpha - \beta + \gamma)$ とする。そうすると、音素「 $\alpha - \beta + \gamma$ 」の出現確率は以下の式 (4) で表すことができる。

$$P(\alpha - \beta + \gamma) = w_1 \Sigma uni(\alpha - \beta + \gamma) + w_2 \Sigma bi(\alpha - \beta + \gamma) + w_3 \Sigma tri(\alpha - \beta + \gamma) \quad (4)$$

【0032】

上述のように、 w_1 、 w_2 、 w_3 は重み係数である。例えば、言語モデルが図 6 の例において、対象音素を「h - a + t」とした場合、対象音素「h - a + t」を含む単語は「h a t s u m e i」である。従って、図 6 に記載の単語のみで考えると、対象音素「h - a + t」の音素出現確率 $P(h - a + t)$ は前記式 (4) を用いて以下になる。

$$P(h - a + t) = w_1 \cdot 0.00002 + w_2 (0.000001 + 0.000005) + w_3 (0.0000004 + 0.0000002 + 0.0000005) \text{ になる。}$$

そして、出現確率計算部 4 4 からは、図 7 記載の両側環境依存音素 (図 7 の例では、「h - a + t」) と当該音素出現確率 (図 7 の例では、0.00008) とが対応付けられ、出力される。

【0033】

この実施例 2 のように、対象音素を両側環境依存音素とすることで、ベース音響モデルの両側環境依存音素についての全体的な認識精度の向上を図ることが出来る。

【実施例 3】

【0034】

この実施例 3 の出現確率計算部 4 4 は、対象音素が片側環境依存音素 (b i p h o n e 例えば a - i + *) として、音素出現確率を求める。この実施例 3 の出現確率計算部 4 4 は、第 2 確率計算手段 4 4 4 を有する。第 2 確率計算手段 4 4 4 は、対象音素を片側環境依存音素として、音素系列中の全ての片側環境依存音素ごとに、当該片側環境依存音素を含む単語のユニグラム確率の総和、当該片側環境依存音素を含む単語のバイグラム確率の総和、当該片側環境依存音素を含む単語のトライグラム確率の総和、のうち少なくとも 1 つを用いて、前記音素出現確率を求める。片側環境依存音素の一般的な表記を「 $\alpha - \beta + *$ 」とし、対象音素「 $\alpha - \beta + *$ 」の出現確率を $P(\alpha - \beta + *)$ とし、音素「 $\alpha - \beta + *$ 」を含む単語のユニグラム確率の総和を $\Sigma uni(\alpha - \beta + *)$ とし、音素「 $\alpha - \beta + *$ 」を含む単語のバイグラム確率の総和を $\Sigma bi(\alpha - \beta + *)$ とし、音素 α を含む単語のトライグラム確率の総和を $\Sigma tri(\alpha - \beta + *)$ とする。そうすると、音素「 $\alpha - \beta + *$ 」の出現確率は以下の式 (5) で表すことができる。

$$P(\alpha - \beta + *) = w_1 \Sigma uni(\alpha - \beta + *) + w_2 \Sigma bi(\alpha - \beta + *) + w_3 \Sigma tri(\alpha - \beta + *) \quad (5)$$

【0035】

上述のように、 w_1 、 w_2 、 w_3 は重み係数である。そして、言語モデルが図 6 の例において、対象音素を「w - a + *」とした場合、対象音素「w - a + *」

を含む単語は「w a t a s h i」と「w a」である。従って、図 6 に記載の単語のみで考えると、対象音素「w - a + *」の音素出現確率 $P(w - a + *)$ は上記式 (5) を用いて以下のようになる。

$$\begin{aligned} P(w - a + *) &= w_1 (0.00002 + 0.00025) \\ &\quad + w_2 (0.000008 + 0.000001) \\ &\quad + w_3 (0.0000004 + 0.0000002) \text{ になる。} \end{aligned}$$

この実施例 3 のように、対象音素を片側環境依存音素とすることで、ベース音響モデルの片側環境依存音素についての全体的な認識精度の向上を図ることが出来る。

【0036】

また、上述の、第 1 確率計算手段 442、第 2 確率計算手段 444、第 3 確率計算手段 446 のうちの 2 つの組み合わせを用いてもよく、3 つ用いてもよい。

【実施例 4】

【0037】

一般的に、音素数が少ない単語（例えば、助詞「は」や「を」等）は頻度（出現確率）が大きい。従って、音素系列言語モデルにおいて、音素数が少ない単語（「w a」や「w o」）は出現確率が高い傾向にある。図 6 のユニグラム確率の記載についても、単語「w a」や「w o」のユニグラム確率（出現確率）は他の単語のユニグラム確率（出現確率）よりも高い。従って、音素数が少ない単語の出現確率を対象にして、対象音素の音素出現確率を計算すると、正確でない音素出現確率が求められる可能性がある。従って、この実施例 4 の出現確率計算部 44 は、予め定められた閾値以上の音素数を持つ音素系列の出現確率の総和を演算することの特徴とすることで、より正確な音素出現確率を求めることが出来る。

【0038】

例えば、予め閾値 Th_3 を定めておき、出現確率計算部 44 は、閾値 Th_3 個以上の音素を含む音素系列のみを音素出現確率の計算の対象にする。図 6 の例でのユニグラム確率の場合において、閾値 Th_3 を「3」とすると、出現確率計算部 44 は、音素数が 3 以上である単語の出現確率のみを演算の対象とする。逆に言えば出現確率計算部 44 が音素閾値 Th_3 ($= 3$) 未満の単語（つまり、w a と w o）の出現確率を演算の対象としないことになる。

【0039】

このように、出現確率計算部 44 は、音素数が多い音素系列の出現確率の総和を演算するようにすると、より正確な音素出現確率を求めることが出来る。

【実施例 5】

【0040】

実施例 1～4 で説明した出現確率計算部 44 から出力される音素出現確率は、値の幅が不明である。従って、音素選択部 48 の音素選択処理に用いる閾値 Th_1 を定めがたいという問題が生じる場合がある。この実施例 5 の出現確率計算部 44 は、正規化手段 448 を有する。当該正規化手段 448 は求められた音素出現確率を正規化することにより、各対象音素について、全体に対する音素出現確率の比率を求めることが出来る。正規化手段 448 は、音素出現確率 $P(\alpha)$ 、 $P(\alpha - \beta + \gamma)$ 、 $P(\alpha - \beta + *)$ について、それぞれ正規化して、正規化音素出現確率 $P'(\alpha)$ 、 $P'(\alpha - \beta + \gamma)$ 、 $P'(\alpha - \beta + *)$ を求める。正規化の具体的手法は、例えば、計算された音素出現確率を、対象音素の全ての出現確率で除算する。つまり以下の式により正規化音素出現確率 $P'(\alpha)$ 、 $P'(\alpha - \beta + \gamma)$ 、 $P'(\alpha - \beta + *)$ は求められる。ただし x 、 y 、 z はそれぞれ全ての音素を示し、 $\Sigma(x)$ 、 $\Sigma(x - y + z)$ 、 $\Sigma(x - y + *)$ はそれぞれ音素 x 、音素 $x - y + z$ 、音素 $x - y + *$ の出現確率の和を示す。

$$P'(\alpha) = P(\alpha) / \Sigma(x)$$

$$P'(\alpha - \beta + \gamma) = P(\alpha - \beta + \gamma) / \Sigma(x - y + z)$$

$$P'(\alpha - \beta + *) = P(\alpha - \beta + *) / \Sigma(x - y + *)$$

である。

【0041】

求められた正規化音素出現確率は、音素選択部48に入力される。そして、音素選択部48は、正規化音素出現確率が高い対象音素を選択し、高出現確率音素を出力する。

このように、音素出現確率を正規化することで、音素出現確率の幅を0～1にすることができ、結果として閾値 Th_1 を定めやすくなるという効果を得ることができる。

【実施例6】

【0042】

図8に音響モデル作成装置300の機能構成例を示す。実施例6の音響モデル作成装置300は、音響モデル作成装置200と比較して、統計量強化部51、重み増強強化統計量記憶部52を具備する点で異なる。統計量強化部51は、音素毎に、強化統計量に対して、当該音素の音素出現確率から求まる値を重み係数として、重み付けして、前記強化統計量として出力する。このようにすることで、音素出現確率が高い対象音素の強化統計量を重みを付けることができ、実施例1～5で説明した音響モデル作成装置より、音素出現確率を反映した強化音響モデルを求めることができる。音素出現確率 $P(\alpha)$ から求まる値とは、音素出現確率 $P(\alpha)$ そのものでもよく、音素出現確率 $P(\alpha)$ と比例関係にある関数 $W(P(\alpha))$ でもよい。関数 $W(P(\alpha))$ とは例えば、以下の式で表される。

$$W(P(\alpha)) = P(\alpha) / \sum P(x)$$

ただし、 $\sum P(x)$ は、全ての対象音素 x の音素出現確率の総和を示す。

【0043】

上記式(2)の事後確率 r_t の記載について以下のように変更する。音素 α について、元の状態 j 、現在の状態 i 、混合分布 m の事後確率を $r(\alpha, i, j, m)$ とする。そうすると、統計量強化部51は、事後確率 $r(\alpha, i, j, m)$ に対して重み係数 $P(\alpha)$ または $W(P(\alpha))$ を乗算する(重み付けする)ことで重み増強事後確率を求め、当該重み増強事後確率からそのほかの統計量(例えば、分散や平均ベクトル)を求め、強化統計量として出力する。なお、以下の式は重み係数として関数 $W(P(\alpha))$ を用いた場合を示す。

$$r'(\alpha, i, j, m) = W(P(\alpha)) \cdot r(\alpha, i, j, m)$$

そして、重み増強強化統計量は重み増強強化統計量記憶部52に記憶される。合成部26は、十分統計量と重み増強強化統計量とから強化音響モデルを作成する。

【0044】

図9に、音素 α_1 と他の音素 α_2 の音素モデルの概念図を示す。図9に示すように音素 α_1 と音素 α_2 を音素モデルとして表した際に、音素 α_1 と音素 α_2 がそれぞれ、3つの状態を有している場合を考える。音素 α_1 が有する3つの状態を第1状態 S_{11} 、第2状態 S_{12} 、第3状態 S_{13} とし、音素 α_2 が有する3つの状態を第1状態 S_{21} 、第2状態 S_{22} 、第3状態 S_{23} とする。状態共有により、異なる音素 α_1 、 α_2 であっても、同じ状態として扱われる場合がある。また、上述のように、強化統計量に関わる重みは音素で計算される。また、強化統計量は状態毎に存在する。

【0045】

図9に示すように例えば、音素 α_1 についての第2状態 S_{12} と音素 α_2 についての第2状態 S_{22} が、状態共有により、同じ状態として扱われるとする。音素 α_1 の第2状態 S_{12} の重み係数を $W(S_{12})$ とし、音素 α_2 の第2状態 S_{22} の重み係数を $W(S_{22})$ とする。そうすると、音素 α_2 の第2状態 S_{22} の重み係数は以下のように新しく求められる。ただし、新しく求められた重み係数を $W'(S_{12} + S_{22})$ とする。

$$W'(S_{12} + S_{22}) = W(S_{12}) + W(S_{22})$$

【0046】

具体的な手法として、音素 α_1 と第 p 状態 S_{1p} の組から、状態ID: q へマッピングするテーブルを出現確率計算部44中の記憶部(図示せず)に記憶しておき、状態共有により同一の状態ID: q となる音素 α_2 と状態 S_{1q} があれば、それぞれの重み係数の和を計算すればよい。この概念は音素が2つ(α_1 と α_2)のみではなく、音素が3つ以上ある場合にも適用できる。また、環境独立音素についてのみ説明したが、片側依存音素、

10

20

30

40

50

両側依存音素についても適用できる。

【0047】

このように、音素出現確率から求まる値（音素出現確率 $P(\alpha)$ ）または音素出現確率に比例する関数 $W(P(\alpha))$ を重み係数として、強化統計量に乗算することで、実施例1～6で説明した音響モデル作成装置より、音素出現確率を反映した強化音響モデルを求めることができる。

【0048】

＜ハードウェア構成＞

本発明は上述の実施の形態に限定されるものではない。また、上述の各種の処理は、記載に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的にあるいは個別に実行されてもよい。その他、本発明の趣旨を逸脱しない範囲で適宜変更が可能であることはいうまでもない。

また、上述の構成をコンピュータによって実現する場合、音響モデル作成装置200が有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、処理機能がコンピュータ上で実現される。

【0049】

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記憶しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記憶装置、光ディスク、光磁気記録媒体、半導体メモリ等のようなものでもよいが、具体的には、例えば、磁気記憶装置として、ハードディスク装置、フレキシブルディスク、磁気テープ等を、光ディスクとして、DVD (Digital Versatile Disc)、DVD-RAM (Random Access Memory)、CD-ROM (Compact Disc Read Only Memory)、CD-R (Recordable) / RW (ReWritable) 等を、光磁気記録媒体として、MO (Magnetooptical disc) 等を、半導体メモリとしてEEPROM (Electrically Erasable and Programmable-Read Only Memory) 等を用いることができる。

【0050】

また、このプログラムの流通は、例えば、そのプログラムを記憶したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させる構成としてもよい。

【0051】

このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記憶されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶装置に格納する。そして、処理の実行時、このコンピュータは、自己の記録媒体に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。また、このプログラムの別の実行形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよく、さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆるASP (Application Service Provider) 型のサービスによって、上述の処理を実行する構成としてもよい。なお、本形態におけるプログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの（コンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等）を含むものとする。

【0052】

また、この形態では、コンピュータ上で所定のプログラムを実行させることにより、本装置を構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

また、本実施例で説明した音響モデル作成装置 200 は、CPU (Central Processing Unit)、入力部、出力部、補助記憶装置、RAM (Random Access Memory)、ROM (Read Only Memory) 及びバスを有している (何れも図示せず)。

【0053】

CPU は、読み込まれた各種プログラムに従って様々な演算処理を実行する。補助記憶装置は、例えば、ハードディスク、MO (Magneto-Optical disc)、半導体メモリ等であり、RAM は、SRAM (Static Random Access Memory)、DRAM (Dynamic Random Access Memory) 等である。また、バスは、CPU、入力部、出力部、補助記憶装置、RAM 及び ROM を通信可能に接続している。

【0054】

10

<ハードウェアとソフトウェアとの協働>

本実施例の単語追加装置は、上述のようなハードウェアに所定のプログラムが読み込まれ、CPU がそれを実行することによって構築される。以下、このように構築される各装置の機能構成を説明する。

音響モデル作成装置 200 の図示しない入力部、出力部は、所定のプログラムが読み込まれた CPU の制御のもと駆動する LAN カード、モデム等の通信装置である。その他の出現確率計算部 44 などは、所定のプログラムが CPU に読み込まれ、実行されることによって構築される演算部である。記憶部は前記補助記憶装置として機能する。

【図面の簡単な説明】

【0055】

20

【図 1】従来の音響モデル作成装置の機能構成例を示したブロック図。

【図 2】実施例 1 の音響モデル作成装置の機能構成例を示したブロック図。

【図 3】本実施例の処理フローを示した図。

【図 4】発音辞書の一例を示した図。

【図 5】言語モデルの一例を示した図。

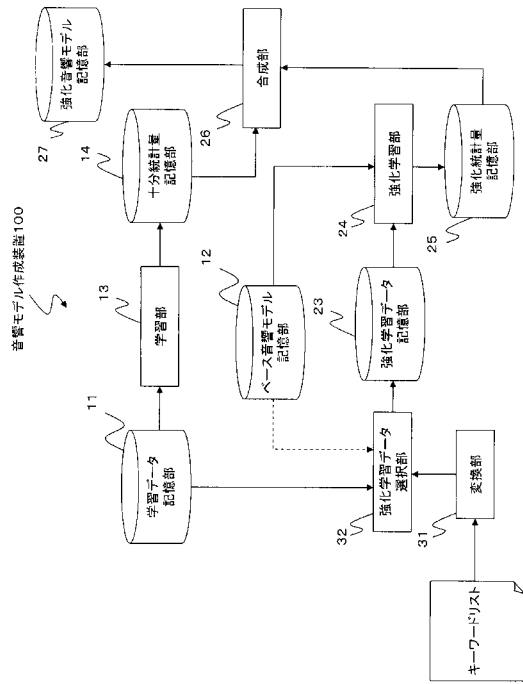
【図 6】音素系列言語モデルの一例を示した図。

【図 7】音素出現確率の一例を示した図。

【図 8】実施例 6 の音響モデル作成装置の機能構成例を示したブロック図。

【図 9】状態共有を示した概念図。

【図 1】



【図 2】

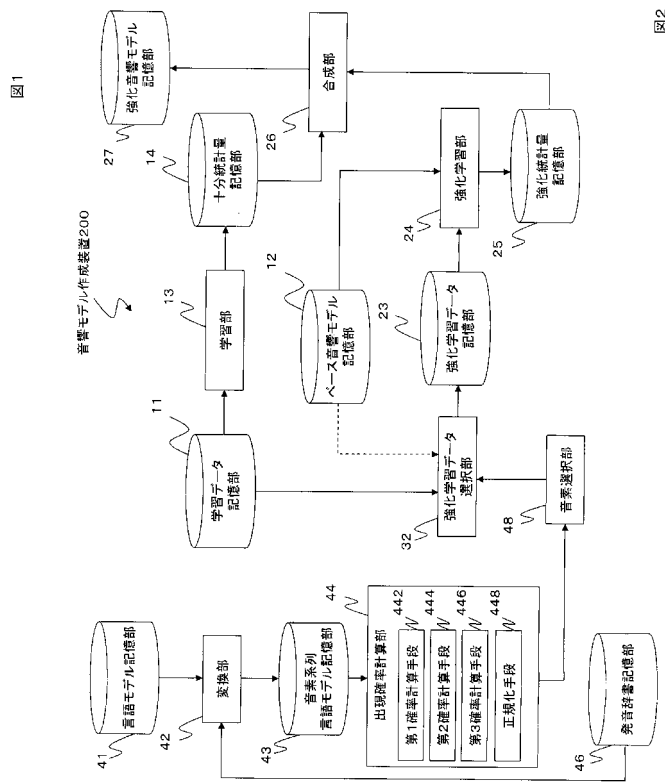


図2

【図 3】

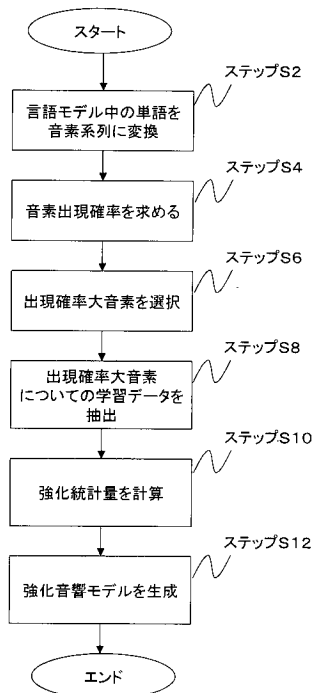


図3

【図 4】

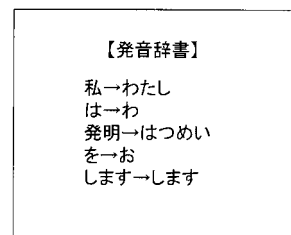


図4

【図 9】

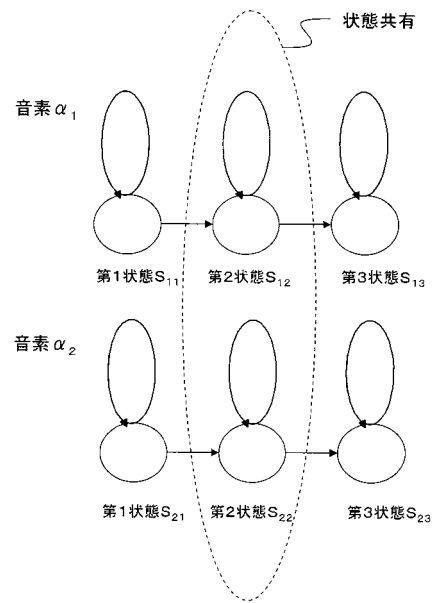


図9

フロントページの続き

(72)発明者 政瀧 浩和

東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内

Fターム(参考) 5D015 GG00