

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5408631号  
(P5408631)

(45) 発行日 平成26年2月5日(2014.2.5)

(24) 登録日 平成25年11月15日(2013.11.15)

(51) Int. Cl.	F 1	
<b>G 0 6 F 17/30</b>	<b>(2006.01)</b>	G O 6 F 17/30 3 1 O Z
<b>G 1 0 L 15/00</b>	<b>(2013.01)</b>	G O 6 F 17/30 1 7 O E
<b>G 1 0 L 15/10</b>	<b>(2006.01)</b>	G 1 0 L 15/00 2 0 O T
		G 1 0 L 15/10 2 0 O W

請求項の数 8 (全 16 頁)

(21) 出願番号	特願2011-501548 (P2011-501548)	(73) 特許権者	304027349
(86) (22) 出願日	平成22年2月10日 (2010.2.10)		国立大学法人豊橋技術科学大学
(86) 国際出願番号	PCT/JP2010/051937		愛知県豊橋市天伯町雲雀ヶ丘1-1
(87) 国際公開番号	W02010/098209	(74) 代理人	100082773
(87) 国際公開日	平成22年9月2日 (2010.9.2)		弁理士 柴田 肇
審査請求日	平成24年2月27日 (2012.2.27)	(74) 代理人	100149320
(31) 優先権主張番号	特願2009-44842 (P2009-44842)		弁理士 井川 浩文
(32) 優先日	平成21年2月26日 (2009.2.26)	(72) 発明者	桂田 浩一
(33) 優先権主張国	日本国 (JP)		愛知県豊橋市天伯町雲雀ヶ丘1-1 国立 大学法人豊橋技術科学大学内
		(72) 発明者	新田 恒雄
			愛知県豊橋市天伯町雲雀ヶ丘1-1 国立 大学法人豊橋技術科学大学内

最終頁に続く

(54) 【発明の名称】 音声検索装置および音声検索方法

(57) 【特許請求の範囲】

【請求項1】

音声を入力として、前記入力された音声をサンプリングして得られる音声データを検索する音声検索装置であって、

音声データベースに記録される音声を認識するデータベース用音声認識器と、

前記データベース用音声認識器によって認識された単語列から音素列を生成する音声用音素列生成部と、

前記音声用音素列生成部によって生成された音素列から接尾辞配列を生成する S u f f i x A r r a y 生成部と、

検索キーワードを入力する入力装置と、

前記入力装置により入力された検索キーワードから音素列を生成する入力音素生成部と、

前記接尾辞配列上で検索キーワードを動的計画法により検索する音声検索部と、

前記音声検索部により検索された結果を出力する出力装置とを備え、

前記音声検索部は、検索で用いる第1の閾値を設定する手段と、

前記第1の閾値を用いて動的計画法により検索対象を検索する手段と、

前記第1の閾値を逐次的に増加させながら繰り返し検索し、検索結果を逐次的に提示する閾値調整手段と

を備えた音声検索部であることを特徴とする音声検索装置。

【請求項2】

請求項1に記載の音声検索装置であって、

10

20

前記音声検索部は、さらに、  
 検索キーワードが所定長さ以上であるとき、該検索キーワードを音素により分割する手段と、  
 前記検索キーワードの分割手段によって分割されたキーワードに対する検索に用いる第2の閾値を前記第1の閾値から決定する手段とを備え、  
 前記検索対象を検索する手段は、前記第2の閾値を用いて動的計画法により検索対象を検索する手段である  
 ことを特徴とする音声検索装置。

【請求項4】

請求項1または2に記載の音声検索装置であって、  
 前記音声検索部は、さらに、  
 検索キーワードの長さによりキーワード分割の有無を判定する手段と、  
 前記キーワード分割後の音素数の決定を行うキーワード分割手段と  
 を備えた音声検索部であることを特徴とする音声検索装置。

【請求項5】

請求項1ないし4のいずれかに記載の音声検索装置であって、  
 前記動的計画法により検索対象を検索する手段は、さらに、  
 動的計画法において音素弁別特徴に基づく音素間距離を用いて音素間の類似性を算出する手段を備えた検索対象を検索する手段であることを特徴とする音声検索装置。

【請求項6】

音声を入力として、前記入力された音声をサンプリングして得られる音声データを検索する音声検索方法であって、  
 音声データを音素列に変換し、接尾辞配列を作成するステップと、  
 検索キーワードを受け付け、音素列に変換するステップと、  
 検索で用いる第1の閾値を設定するステップと、  
 前記第1の閾値を用いて動的計画法により検索対象を検索するステップと、  
 前記第1の閾値を逐次的に増加させながら繰り返し検索する閾値調整ステップと、  
 前記検索のステップにより検索された結果を出力するステップと  
 を含むことを特徴とする音声検索方法。

【請求項7】

請求項6に記載の音声検索方法であって、さらに、  
 検索キーワードが所定長さ以上であるとき、該検索キーワードを音素により分割するステップと、  
 前記検索キーワードの分割ステップによって分割されたキーワードに対する検索に用いる第2の閾値を前記第1の閾値から決定するステップとを備え、  
 前記検索対象を検索するステップは、前記第2の閾値を用いて動的計画法により検索対象を検索するステップである  
 ことを特徴とする音声検索方法。

【請求項9】

請求項6または7に記載の音声検索方法であって、さらに、  
 前記検索キーワードの長さによりキーワード分割の有無を判定するステップと、  
 前記キーワード分割後の音素数の決定を行うキーワード分割ステップと  
 を含むことを特徴とする音声検索方法。

【請求項10】

請求項6ないし9のいずれかに記載の音声検索方法であって、  
 前記検索対象を検索するステップは、前記動的計画法において音素弁別特徴に基づく音素間距離を用いて音素間の類似性を算出するステップを有することを特徴とする音声検索方法。

【発明の詳細な説明】

【技術分野】

10

20

30

40

50

## 【0001】

本発明は、音声検索装置および音声検索方法に関する。より詳細には、音声に対する検索を高速かつ効率的に行う装置ならびに方法に関するものである。

## 【背景技術】

## 【0002】

ブロードバンド回線の普及など情報通信技術の発展により、Web上で音声や動画のコンテンツを利用する機会が増え、コンテンツの数も急激に増加している。このため、Web上の音声データや動画データを効率的に探索し、利用するには、音声検索技術が必要不可欠である。特に高速に検索するシステムが求められている。

## 【0003】

このようなニーズを満たすために、非特許文献1および2に記載の先行技術では、音声データベースから索引データを作成し、これを用いて高速な音声検索を実現する手法が提案されている。

10

## 【0004】

また、特許文献1に記載の先行技術では、接尾辞配列と動的計画法を組み合わせ、文書検索において表記揺れの解消と高速化について記述がある。

## 【先行技術文献】

## 【特許文献】

## 【0005】

【特許文献1】国際公開2005/122002号

20

## 【非特許文献】

## 【0006】

【非特許文献1】N. Kanda, et al., "Open-Vocabulary Keyword Detection from Super-Large Scale Speech Database," IEEE MMSP 2008, pp. 939-944, 2008.

【非特許文献2】K. Thambiratnam & S. Sridharan, "Dynamic Match Phone-Lattice Searches For Very Fast And Accurate Unrestricted Vocabulary Keyword Spotting," ICASSP 2005, vol. 1, pp. 465-468, 2005.

30

## 【発明の概要】

## 【発明が解決しようとする課題】

## 【0007】

上記特許文献1に記載の先行技術では、文書検索を対象としており、本発明の係る音声認識による音素列に対するあいまい検索ではない。とくに単純な接尾辞配列と動的計画法を組み合わせるだけでは計算時間が著しく増大することは自明である。

## 【0008】

一方、上述した先行技術（非特許文献1ならびに2）では、音声データベースが大規模になると、従来の音声検索高速化手法ではデータベースに見合う規模の索引データを作成しなければならない。このため高速な二次記憶装置が必要となり、コスト面からは望ましくない。

40

## 【0009】

また、主記憶装置と比較して二次記憶装置はアクセスに時間がかかるため、検索速度も低下するデメリットも存在する。

## 【0010】

さらに、索引データを単語、あるいはサブワードから作成することが一般的であるが、その場合、検索キーワード（あるいはサブキーワード）と単語、あるいはサブワードとの完全一致が前提になるため、誤認識の多い現状の音声認識では十分な検索性能が得られない可能性がある。

50

## 【 0 0 1 1 】

本発明は、上記の課題に対し、二次記憶装置を必要とせず、検索速度が高速であり、かつ、検索に要するコストが低コストであり、さらに、良好な検索性能を発揮する、あいまい検索を行う音声検索装置および音声検索方法を提供することを目的とする。

## 【課題を解決するための手段】

## 【 0 0 1 2 】

本発明に係る音声検索において、あいまい検索を行う音声検索装置および音声検索方法は、具体的には以下のように構成される。

## 【 0 0 1 3 】

請求項 1 に係る発明は、音声を入力として、前記入力された音声をサンプリングして得られる音声データを検索する音声検索装置であって、音声データベースに記録される音声を認識するデータベース用音声認識器と、前記データベース用音声認識器によって認識された単語列から音素列を生成する音声用音素列生成部と、前記音声用音素列生成部によって生成された音素列から接尾辞配列を生成する *Suffix Array* 生成部と、検索キーワードを入力する入力装置と、前記入力装置により入力された検索キーワードから音素列を生成する入力音素生成部と、前記接尾辞配列上で検索キーワードを動的計画法により検索する音声検索部と、前記音声検索部により検索された結果を出力する出力装置とを備え、前記音声検索部は、検索で用いる第 1 の閾値を設定する手段と、前記第 1 の閾値を用いて動的計画法により検索対象を検索する手段と、前記第 1 の閾値を逐次的に増加させながら繰り返し検索し、検索結果を逐次的に提示する閾値調整手段とを備えた音声検索部であることを特徴とする音声検索装置を要旨としている。

10

20

## 【 0 0 1 4 】

この音声検索装置は、音声を入力として、前記入力された音声をサンプリングして得られる音声データを検索する音声検索装置であって、接尾辞配列と、動的計画法をとともに用いてあいまい検索を行うものである。音素単位で検索キーワードとのマッチングを行うため、索引に登録された単語、あるいはサブワードに完全一致しなくとも検索を可能としている。

また、上記構成は、前記検索の第 1 の閾値を *iterative lengthening search* (一種の反復深化探索) によって逐次的に増加させながら繰り返し検索し、検索結果を逐次的に提示する閾値調整機能を備えたものである。装置の利用者が初期に提示された検索結果を確認している間に、閾値を更新しつつ新たな検索結果を逐次的に提示することにより、体感的な検索速度の向上を実現している。

30

## 【 0 0 1 5 】

請求項 2 に係る発明は、請求項 1 に記載の音声検索装置であって、前記音声検索部は、さらに、検索キーワードが所定長さ以上であるとき、該検索キーワードを音素により分割する手段と、前記検索キーワードの分割手段によって分割されたキーワードに対する検索に用いる第 2 の閾値を前記第 1 の閾値から決定する手段とを備え、前記検索対象を検索する手段は、前記第 2 の閾値を用いて動的計画法により検索対象を検索する手段であることを特徴とする音声検索装置を要旨とする。

## 【 0 0 1 6 】

上記構成の音声検索装置は、請求項 1 に係る発明に加えて、処理時間の指数爆発を防ぐために、検索キーワードの分割、2 箇所以上マッチするための検索第 1 の閾値の変更、キーワードの長さに応じた分割音素数、分割 / 非分割の決定を行い、高速な検索を実現するものである。前記 2 箇所以上マッチするための検索第 1 の閾値の変更は、次の式 1 (すなわち数 1) に従って行う。

40

## 【 0 0 1 7 】

【数 1】

$$t' = \frac{p}{p-1} t$$

【0018】

ただし、 $p$ は分割数、 $t$ は上記複数の分割された検索キーワードに対する検索の閾値を決定する手段により求められる元の第1の閾値、 $t'$ は前記閾値変更手段による変更後の第2の閾値である。

10

【0021】

請求項4に係る発明は、請求項1または2に記載の音声検索装置であって、前記音声検索部は、さらに、検索キーワードの長さによりキーワード分割の有無を判定する手段と、前記キーワード分割後の音素数の決定を行うキーワード分割手段とを備えた音声検索部であることを特徴とする音声検索装置を要旨としている。

【0022】

上記構成の装置は、検索キーワードの長さによりキーワード分割の有無を判定し、前記キーワード分割後の音素数の決定を行うことができる。

【0023】

請求項5に係る発明は、請求項1ないし4のいずれかに記載の音声検索装置であって、前記動的計画法により検索対象を検索する手段は、さらに、動的計画法において音素弁別特徴に基づく音素間距離を用いて音素間の類似性を算出する手段を備えた検索対象を検索する手段であることを特徴とする音声検索装置を要旨としている。

20

【0024】

上記の音素間距離には、例えば、音素弁別特徴の差異のハミング距離がある。そこで、上記構成では、このハミング距離を用いることによって音素間の類似性を算出するものである。

【0025】

請求項6に係る発明は、音声を入力として、前記入力された音声をサンプリングして得られる音声データを検索する音声検索方法であって、音声データを音素列に変換し、接尾辞配列を作成するステップと、検索キーワードを受け付け、音素列に変換するステップと、検索で用いる第1の閾値を設定するステップと、前記第1の閾値を用いて動的計画法により検索対象を検索するステップと、前記第1の閾値を逐次的に増加させながら繰り返し検索する閾値調整ステップと、前記検索のステップにより検索された結果を出力するステップとを含むことを特徴とする音声検索方法を要旨としている。

30

【0026】

上記構成の音声検索方法は、接尾辞配列（以下、`Suffix Array`と記述する場合がある）と動的計画法（以下、`DP (Dynamic Programming)`マッチングと記述する場合がある）をともに用いてあいまい検索を行うものである。音素単位で検索キーワードとのマッチングを行うため、索引に登録された単語、あるいはサブワードに完全一致しなくとも検索を可能としている。

40

また、上記構成の音声検索方法は、前記検索の第1の閾値を `iterative lengthening search`（一種の反復深化探索）によって逐次的に増加させながら繰り返し検索し、検索結果を逐次的に提示する閾値調整機能を備えたものである。第1の閾値が小さい初期の探索では `Suffix Array` の特徴により二分探索に近い検索になることから、非常に高速な検索が可能である。

【0027】

請求項7に係る発明は、請求項6に記載の音声検索方法であって、さらに、検索キーワードが所定長さ以上であるとき、該検索キーワードを音素により分割するステップと、前記検索キーワードの分割ステップによって分割されたキーワードに対する検索に用いる第

50

2の閾値を前記第1の閾値から決定するステップとを備え、前記検索対象を検索するステップは、前記第2の閾値を用いて動的計画法により検索対象を検索するステップであることを特徴とする音声検索方法を要旨とする。

【0028】

上記構成の音声検索方法は、処理時間の指数爆発を防ぐために、検索キーワードの分割、2箇所以上マッチする手法、検索第1の閾値の変更、キーワードの長さに応じた分割音素数、分割/非分割の決定を行い、高速な検索を実現するものである。なお、分割されたキーワードに対する検索の第2の閾値は数1に示す式1に基づいて決定することができる。

【0031】

請求項9に係る発明は、請求項6または7に記載の音声検索方法であって、さらに、前記検索キーワードの長さによりキーワード分割の有無を判定するステップと、前記キーワード分割後の音素数の決定を行うキーワード分割ステップとを含むことを特徴とする音声検索方法を要旨としている。

【0032】

上記構成の音声検索方法は、検索キーワードの長さによりキーワード分割の有無を判定し、前記キーワード分割後の音素数の決定を行うことができるように処理される。

【0033】

請求項10に係る発明は、請求項6ないし9のいずれかに記載の音声検索方法であって、前記検索対象を検索するステップは、前記動的計画法において音素弁別特徴に基づく音素間距離を用いて音素間の類似性を算出するステップを有することを特徴とする音声検索方法を要旨としている。

【0034】

上記構成の音声検索方法は、前記動的計画法において音素弁別特徴に基づく音素間距離（例えば、音素弁別特徴の差異のハミング距離）を用いて音素間の類似性を算出するような処理が可能となっている。

【発明の効果】

【0035】

本発明によれば、データ領域を多く必要としないため高速な二次記憶装置を必要とせず、二次記憶の準備に必要なコストを低減できる。つまり、検索速度が高速かつ低コストであり、かつ、良好な検索性能も備えた音声検索装置ならびに音声検索方法を提供することが可能となる。

【図面の簡単な説明】

【0036】

【図1】本発明に係る音声検索のフローチャートである。

【図2】本発明に係る音声検索の内部構成ブロック図である。

【図3】本発明に係る音声データベースからの接尾辞配列 ( S u f f i x A r r a y ) の作成の説明図である。

【図4】本発明に係る接尾辞配列 ( S u f f i x A r r a y ) 上での動的計画法 ( D P マッチング ) によるあいまい検索の説明図である。

【図5】本発明に係るキーワードの分割と音声検索の説明図である。

【図6】本発明の実施例に係る6音素の検索キーワードを対象に第1の閾値を最も低くした場合に最初の検索結果が出力されるまでの時間の図である。グラフの横軸は第1の閾値、縦軸は検索の再現率、適合率および処理時間を表す。

【図7】本発明の実施例に係る12音素の検索キーワードを対象に第1の閾値を最も低くした場合に最初の検索結果が出力されるまでの時間の図である。グラフの横軸は第1の閾値、縦軸は検索の再現率、適合率および処理時間を表す。

【図8】本発明の実施例に係る18音素の検索キーワードを対象に第1の閾値を最も低くした場合に最初の検索結果が出力されるまでの時間の図である。グラフの横軸は第1の閾値、縦軸は検索の再現率、適合率および処理時間を表す。

10

20

30

40

50

【図 9】本発明の実施例に係る 24 音素の検索キーワードを対象に第 1 の閾値を最も低くした場合に最初の検索結果が出力されるまでの時間の図である。グラフの横軸は第 1 の閾値、縦軸は検索の再現率、適合率および処理時間を表す。

【図 10】本発明の実施例に係る 6 音素の検索キーワードを対象に半分の正解キーワードを検出するまでの時間の図である。グラフの横軸は第 1 の閾値、縦軸は検索の再現率、適合率および処理時間を表す。

【図 11】本発明の実施例に係る 12 音素の検索キーワードを対象に半分の正解キーワードを検出するまでの時間の図である。グラフの横軸は第 1 の閾値、縦軸は検索の再現率、適合率および処理時間を表す。

【図 12】本発明の実施例に係る第 1 の閾値を初期値 0.0 に設定して 6 音素から 24 音素の検索キーワードを検索し、検索結果群がユーザに提示されるまでの時間の図である。グラフの横軸は擬似音声データベース（毎日新聞コーパス）の音声換算の時間（単位：時間）、縦軸は検索の処理時間（単位：ミリ秒）である。

10

【図 13】本発明の実施例に係る第 1 の閾値を図 12 の状態から 0.2 に更新し、改めて 6 音素から 24 音素の検索キーワードを検索し、検索結果群がユーザに提示されるまでの時間の図である。グラフの横軸は擬似音声データベース（毎日新聞コーパス）の音声換算の時間（単位：時間）、縦軸は検索の処理時間（単位：ミリ秒）である。

【図 14】本発明の実施例に係る第 1 の閾値を図 13 の状態からさらに 0.4 に更新し、改めて 6 音素から 24 音素の検索キーワードを検索し、検索結果群がユーザに提示されるまでの時間の図である。グラフの横軸は擬似音声データベース（毎日新聞コーパス）の音声換算の時間（単位：時間）、縦軸は検索の処理時間（単位：ミリ秒）である。

20

【図 15】本発明の実施例に係る図 12 とは異なる実施環境において、第 1 の閾値を初期値 0.0 に設定して 6 音素から 24 音素の検索キーワードを検索し、検索結果群がユーザに提示されるまでの時間の図である。グラフの横軸は擬似音声データベース（毎日新聞コーパス）の音声換算の時間（単位：時間）、縦軸は検索の処理時間（単位：ミリ秒）である。

【図 16】本発明の実施例に係る第 1 の閾値を図 15 の状態から 0.2 に更新し、改めて 6 音素から 24 音素の検索キーワードを検索し、検索結果群がユーザに提示されるまでの時間の図である。グラフの横軸は擬似音声データベース（毎日新聞コーパス）の音声換算の時間（単位：時間）、縦軸は検索の処理時間（単位：ミリ秒）である。

30

【図 17】本発明の実施例に係る第 1 の閾値を図 16 の状態からさらに 0.4 に更新し、改めて 6 音素から 24 音素の検索キーワードを検索し、検索結果群がユーザに提示されるまでの時間の図である。グラフの横軸は擬似音声データベース（毎日新聞コーパス）の音声換算の時間（単位：時間）、縦軸は検索の処理時間（単位：ミリ秒）である。

【図 18】本発明の実施例に係る第 1 の閾値を図 17 の状態からさらに 1.0 に更新し、改めて 6 音素から 24 音素の検索キーワードを検索し、検索結果群がユーザに提示されるまでの時間の図である。グラフの横軸は擬似音声データベース（毎日新聞コーパス）の音声換算の時間（単位：時間）、縦軸は検索の処理時間（単位：ミリ秒）である。

【発明を実施するための形態】

【0037】

40

以下、本発明の実施の形態における音声検索装置および音声検索方法について、図面を参照して説明する。なおこれらの図面は、本発明が採用し得る技術的特徴を説明するために用いられるものであり、記載されている装置の構成、各種処理のフローチャートなどは、特に特定の記載がない限り、そのみに限定する趣旨ではなく、単なる説明例である。

【0038】

本発明に係る音声検索方法の実施形態について、図 1 のフローチャートに従って説明する。本実施形態は、スタート時に音声を入力として、前記入力された音声をサンプリング（例えば、サンプリングビット数 16 bit、サンプリング周波数 44.1 kHz）して得られる音声データを S u f f i x A r r a y と D P マッチングをともに用いてあいま

50

い検索を行うものである。

【0039】

まず、音声データベースに記録される音声データを音素列(a)に変換し、この音素列(a)からSuffix Arrayを作成する(S11)。次に、検索キーワードを受け付け、その検索キーワードを音素列に変換する(S12)。この処理と同時に、検索で用いる第1の閾値(図1において閾値1と表記)を設定する(S12)。上記処理において、音素列に変換した検索キーワードの長さが所定以上(例えば、音素の数で9以上)である場合には、これを分割するが、それ以下である場合には分割しないものとし、当該分割/非分割を判断する(S13)。分割すべきと判断された場合は、当該検索キーワードを所定の音素数に分割する(S14)。このとき、分割後の音素の数を予め決定しておくことができる。例えば、分割後の音素の数を3とすることにより、検索キーワードの音素数が9である場合に、3個の音素ごとに3分割することができる。そして、検索キーワードの音素数が10~12では、4分割することができる。

10

【0040】

分割されたキーワードは、音声データに含まれる音素弁別特徴間の距離を算出して類似性が判断される。すなわち、数1に示した数式によって第1の閾値から第2の閾値(図1において閾値2と表記)を決定するとともに、この第2の閾値を用いて分割されたキーワードをSuffix Array上でDPマッチングするのである(S15)。この結果は第1段階候補(b)として一時保存し(S15)、この第1段階候補(b)の位置関係から最終候補(c)を決定するのである(S16)。また、この最終候補(c)の結果

20

【0041】

本実施形態では、第1次的な検索の終了後に、第1の閾値を少し高い値に更新(例えば、0.2を加算)して、再度検索ステップを繰り返すように処理される(S17, S18)。第1の閾値が少し高い値に変更されていることから、当該第1の閾値に基づいて算出される第2の閾値も少し高い値となる。これにより、音素弁別特徴間の距離がやや離れた類似語(類似する音素列の語)の検索が実施されることとなる。なお、上記検索ステップの繰り返しは、第1の閾値が所定の値に到達する場合、または、検索結果の総数が所定の

30

【0042】

次に、検索キーワードが短い(例えば、音素の数が8以下となる)場合には、検索キーワードを分割せず、Suffix Array(a)上において、第1の閾値を用いてDPマッチングする(S19)。このとき、使用される閾値は第1の閾値のみである。そして、ここで得られた結果はそのまま出力(ユーザに提示)されるのである(S19)。検索キーワードが分割されていないことから、マッチングされた結果の位置関係を参照する必要はないのである。

40

【0043】

なお、検索キーワードを分割しない場合であっても、第1の閾値を少し高い値に更新(例えば、0.2を加算)して(S20)、再度検索ステップを繰り返すものである(S21)。音素弁別特徴間の距離が離れた類似語(類似する音素列の語)を検索するためである。そして、上記検索の繰り返しは、更新した閾値が所定の値に達する場合、または、検索結果数が所定数に達した場合に処理を終了させるように構築することができる。

【0044】

上記のような検索方法によれば、第1の閾値を小さくした初期の検索では、二分探索に近い検索条件となるため、検索キーワードに非常に近い音素列を高速に検索できることとなる。そして、この第1の閾値を徐々に大きくすることにより、一種の反復深化探索を可

50



能にする。さらに、第1の閾値を更新する前に逐次出力（ユーザに提示）することにより、検索キーワードに近似する音素列から順次出力することができる。キーワードの分割について、上記実施形態では、検索キーワードの長短を判断する際の音素数について9以上を例示したが、分割された後の音素数を6とする場合には、検索キーワードの長短の境界を18とすることができる。分割後の音素数が少ない場合には、第1段階候補（b）の数が膨大となり、処理速度が遅くなる可能性もあり得るからである。そこで、分割後の音素数を調整することによって、検索時間のさらに高速化することも可能である。

【0045】

なお、上記に示した検索方法において、第1の閾値を更新せずに処理を終了するような構成とすることもできる。この場合、検索により得られる音素列は検索キーワードに近似するものに限定されるが、予め第1の閾値を少し大きめに設定することにより、一度の検索ステップによって多くの音素列を検索することができる。

10

【0046】

また、キーワードの分割について、当該分割の有無を判定するステップ（S13）が設けられているが、このようなステップを割愛し、予め定めた音素数に分割するように処理するか、または、分割しないものとして処理することも可能である。所定の音素数に分割するような処理を選択する場合には、第1の閾値から第2の閾値を数1に示す数式に従って算出するには、3以上に分割すべきであるから、所定の音素数に分割した際の分割数が3未満となるか、3以上となるかを判断するステップが必要となる。

【0047】

20

一方、本発明に係る音声検索装置の実施形態は、図2の内部構成ブロック図のように構成される。本実施形態は、予めサンプリング（例えば、サンプリングビット数16bit、サンプリング周波数44.1kHz）された大規模音声データを音声データベース25に格納して、Suffix Array作成部28とDPマッチングを行う音声検索部29ともを用いてあいまい検索を行う手段を実現するものである。

【0048】

本実施形態の音声検索装置31は、音声データからSuffix Arrayを作成するために、音声データベース25、データベース用音声認識器26、音声用音素列生成部27およびSuffix Array作成部28が設けられている。他方、入力される検索キーワードの音素列を作成するために、入力装置21、24および音素列生成部23が設けられている。入力装置21、24は、一方が音声入力装置（例えば、マイク）21であり、他方が文字入力装置（例えば、キーボード）24である。両者は異なる入力手段であるが、両者をともに備える構成でもよいが、いずれか一方のみを備える構成でもよい。ただし、音声入力装置（例えば、マイク）21を備える場合には、音声認識器22を備える必要がある。そして、単語列として入力または音声を単語列に変換されたキーワードは、音素列生成部23によって音素列に変換されるのである。図2の「音声/文字用音素列生成部」との記載は、音声入力の場合または文字入力の場合のいずれについても対応することを意味する。

30

【0049】

図2に示しているように、音声データから作成されたSuffix Arrayの情報と、入力された検索キーワードの音素列の情報は、音声検索部29において検索処理される構成となっている。この音声検索部29には、検索で用いる第1の閾値を設定する手段と、検索キーワードが所定長さ以上であるとき、該検索キーワードを音素により分割する手段と、検索キーワードの分割手段によって分割されたキーワードに対する検索に用いる第2の閾値を第1の閾値から決定する手段と、第1および第2の閾値の少なくともいずれか一方を用いて動的計画法により検索対象を検索する手段とが備えられている。

40

【0050】

従って、音声データに含まれる音素弁別特徴間の距離を算出して類似性を判定する手段は、音声検索部29で実現される。一方、入力された検索キーワードを音素により分割する手段と、分割した検索キーワードを必ず2箇所以上マッチするために、上記式1（数1

50

)に従って第1の閾値を変更し、第2の閾値を求める手段と、前記第1の閾値および前記第2の閾値により決定される検索対象を検索する手段は、いずれも図2における音声検索部29で実現される。

【0051】

また、検索の第1の閾値を逐次的に増加させながら繰り返し検索する手段は、音声検索部29によって実現されることとなり、検索結果を逐次的に閾値調整手段についても音声検索部29において実現される。これと同時に、検索結果を逐次的に出力(ユーザに提示)する手段は、表示装置(例えば、ディスプレイ)30または音声出力装置(例えば、スピーカ)31によって実現される。

【0052】

さらに、検索キーワードの長さによりキーワード分割の有無を判定する手段は、音声検索部29において実現され、また、キーワード分割後の音素数の決定を行うキーワード分割手段は、音声/文字用音素列生成部23および音声検索部29より実現される。

【0053】

本実施形態の音声検索装置は、図2に示しているように、音声検索結果は、表示装置30(例えば、ディスプレイ)により検索に関する文字や画像等の情報が表示され、かつ、音声情報は、音声出力装置31(例えば、スピーカ)から音として再生される。これらは、いずれか一方のみ備えた構成としてもよい。

【0054】

図2に示された音声処理装置32は、ROM、RAM(以下、メモリ)、CPU、HDD、音声入出力インターフェイス(例えば、サンプリングビット数16bit、サンプリング周波数44.1kHzが処理可能なインターフェイス)がシステムバスにより電氣的に有意に接続されているパーソナルコンピュータを用いて実現してもよい。音声データベースをHDDに格納し、音声入力装置21と音声出力装置31を音声入出力インターフェイスに接続し、前記以外の手段をC#言語もしくはC++言語によりソフトウェアとして構成しHDDに格納して、前記ソフトウェアは起動時にメモリへ読み込まれ、システムバスを介して、おもにメモリとCPUを連動させ、必要な手段を実現することができる。

【0055】

次に、音声検索の具体的手法について、図3から図5を参照して説明する。

【0056】

図3は、音声データベースからの接尾辞配列(Suffix Array)の作成の説明図である。データベース用音声認識器26を用いて音声データベース25に格納されている音声データを単語列に変換し、音声用音素列生成部27により更に単語列を音素列(a)に変換する。次に、前記音素列からSuffix Array生成部28によりSuffix Arrayを作成し、メモリもしくはHDD上に保存する。

【0057】

検索キーワードを音声で受け付けた場合(音声入力装置21により入力される)には、音声認識器22を用いて単語列に変換し、音声/文字用音素列生成部23により音素列に変換する。テキスト(文字列)で受け付けた場合(文字入力装置24により入力される)にも音声/文字用音素列生成部23により音素列に変換する。音声検索部29により検索で用いる音素あたりの平均第1の閾値を低い値(例えば0.0)に設定する。

【0058】

前記Suffix Array上でのDPマッチングによるあいまい検索の説明を図4に示す。最適の音素数に分割、あるいは非分割した後に、Suffix Array上でDPマッチングにより検索を行う。DPマッチングの閾値として、キーワードが非分割の場合には第1の閾値を用い、分割した場合には第1の閾値を上記式1(数1)で変更した値(第2の閾値)を用いる。これにより検索結果の第1段階候補(b)が得られる。非分割の場合には、(b)を結果としてユーザに表示装置30および音声出力装置31により提示する。

【0059】

10

20

30

40

50

キーワードの分割と音声検索の説明を図5に示す。音素を分割した場合には、分割されたキーワードのうち2つ以上が検索されるため、検索結果の位置関係から検索結果の最終候補(c)を求める。

【0060】

最終候補(c)を音素列(a)と第1の閾値を用いてDPマッチングし、検索結果をユーザに表示装置30および音声出力装置31により提示する。検索結果の提示後に、第1の閾値を少し高い値に更新(例えば0.2を加算する)した後、第1の閾値を用いたDPマッチングへ戻る。

【実施例1】

【0061】

CSJ(Corpus of Spontaneous Japanese)コーパス(男性話者、390時間)の音声データを対象に、パーソナルコンピュータ(Intel(登録商標) Pentium(登録商標) D 2.8GHz、メモリ4GB)上で図2の音声検索装置32を、C#言語で構築し、検索実験を行った結果、第1の閾値を最も低くした場合(0.0)に最初の検索結果が出力されるまでの時間を図6から図9に示す。図6から図9では、グラフの横軸は第1の閾値、縦軸は検索の再現率、適合率および処理時間を表している。検索キーワード6音素(図6を参照)、12音素(図7を参照)、18音素(図8を参照)、24音素(図9を参照)で、前記第1の閾値を最も低くした場合に最初の検索結果が出力されるまでの時間は、それぞれ数ミリ秒であった。

【実施例2】

【0062】

上記コーパス中に含まれる半分の正解キーワードを検出するまでの時間を図10と図11に示す。図10と図11では、グラフの横軸は第1の閾値、縦軸は検索の再現率、適合率および処理時間を表している。検索キーワード6音素(図10を参照)、12音素(図11を参照)で、前記半分の正解キーワードを検出するまでの時間は、それぞれ1秒以下であった。

【実施例3】

【0063】

音声に換算して10000時間相当の時間数の新聞記事データを対象に、パーソナルコンピュータ(Intel(登録商標) Pentium(登録商標) D 2.8GHz、メモリ4GB)上で図2の音声検索装置32をC#言語で構築し、検索実験を行った結果を図12から図14に示す。非特許文献1には2031時間の音声データベースから5.2モーラ(5から11音素の範囲内)の検索キーワードを検索するのに2.17秒を要すると記載されている。一方、本発明では図12に示すように、第1の閾値を0.0にして6音素から24音素の検索キーワードを検索し、最初の検索結果群をユーザに提示するまでの時間は数ミリ秒から120ミリ秒である。また、第1の閾値を0.2に更新して改めて6音素から24音素の検索キーワードを検索し、新たに得られた検索結果群をユーザに提示するまでの時間は図13に示すように、数ミリ秒から130ミリ秒である。さらに、第1の閾値を0.4に更新して改めて6音素から24音素の検索キーワードを検索し、新たに得られた検索結果群をユーザに提示するまでの時間は図14に示すように、数十ミリ秒から600ミリ秒である。以上から、高速に音声検索を行うことができていることが分かる。

【実施例4】

【0064】

音声に換算して10000時間相当の時間数の新聞記事データを対象に、パーソナルコンピュータ(Intel(登録商標) Core2 Duo E8600 3.3GHz、メモリ8GB)上で図2の音声検索装置32をC++言語で構築し、検索実験を行った結果を図15から図18に示す。図15に示すように、第1の閾値を0.0にして6音素から24音素の検索キーワードを検索し、最初の検索結果群をユーザに提示するまでの時間は数ミリ秒である。また、図16に示すように、第1の閾値を0.2に更新して改めて6

10

20

30

40

50

音素から24音素の検索キーワードを検索し、新たに得られた検索結果群をユーザに提示するまでの時間は数ミリ秒である。さらに、図17に示すように、第1の閾値を0.4に更新して改めて6音素から24音素の検索キーワードを検索し、新たに得られた検索結果群をユーザに提示するまでの時間は数ミリ秒から27ミリ秒である。さらに、図18に示すように、第1の閾値を1.0に更新して改めて6音素から24音素の検索キーワードを検索し、新たに得られた検索結果群をユーザに提示するまでの時間は十数ミリ秒から約1秒である。以上から、高速に音声検索を行うことができていることが分かる。

【符号の説明】

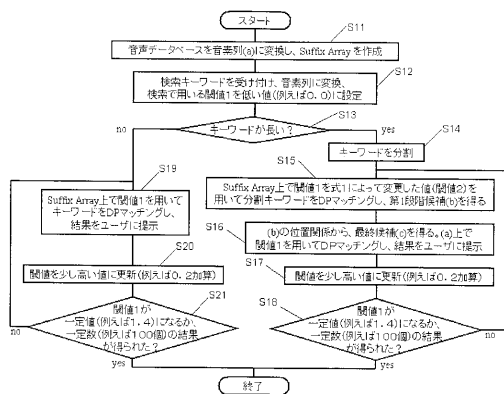
【0065】

- 21 音声入力装置
- 22 音声認識器
- 23 音声/文字用音素列生成部
- 24 文字入力装置
- 25 音声データベース
- 26 データベース用音声認識器
- 27 音声用音素列生成部
- 28 Suffix Array生成部
- 29 音声検索部
- 30 表示装置
- 31 音声出力装置
- 32 音声検索装置

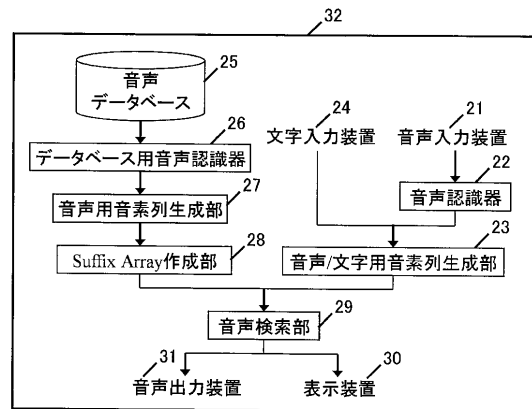
10

20

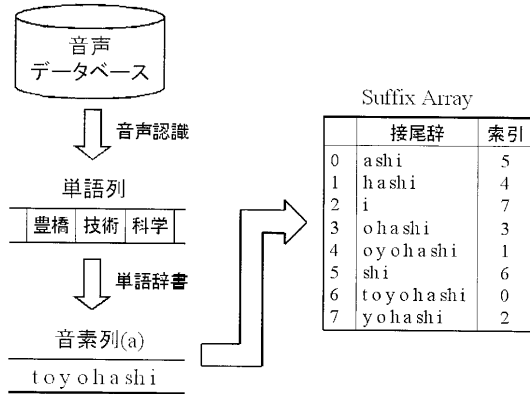
【図1】



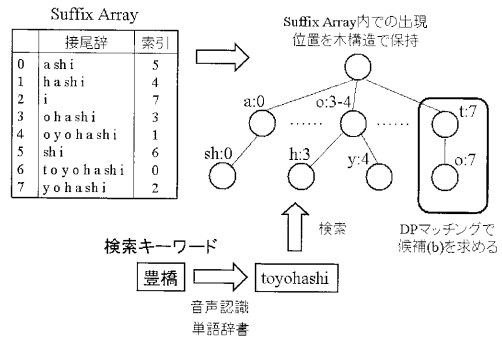
【図2】



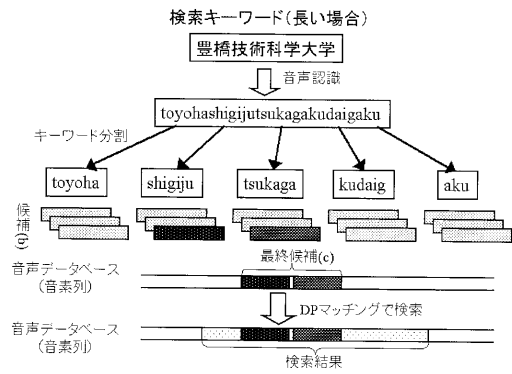
【図3】



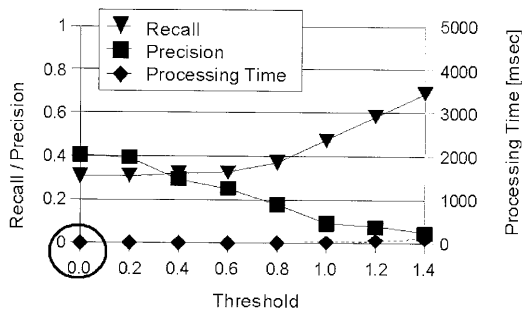
【図4】



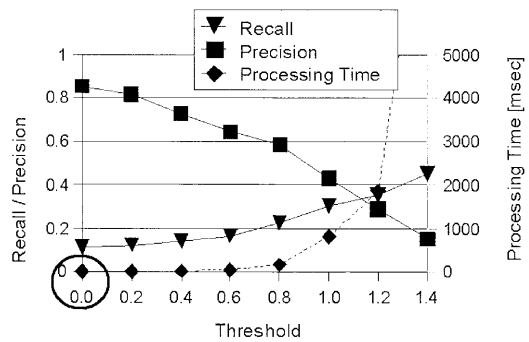
【図5】



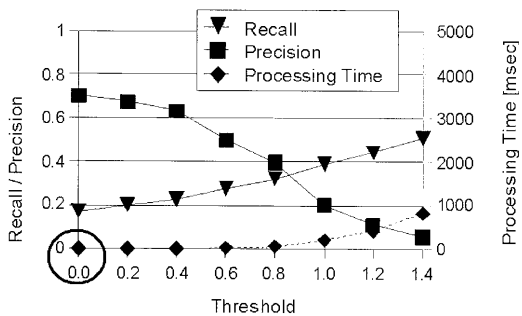
【図6】



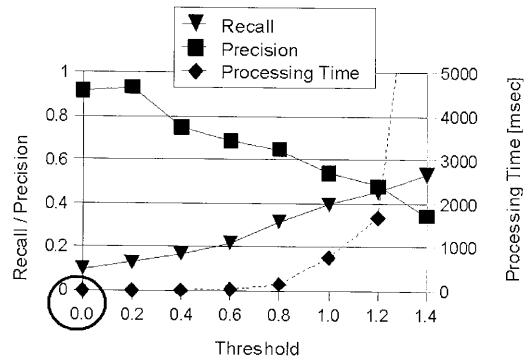
【図8】



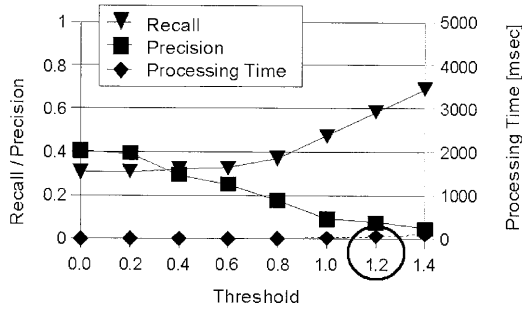
【図7】



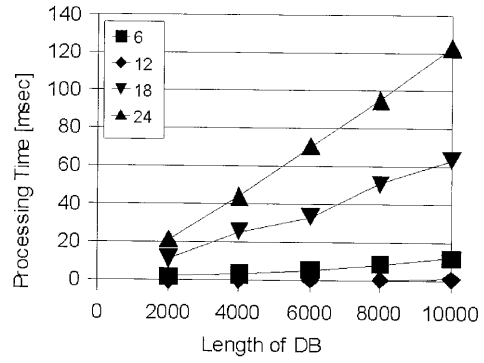
【図9】



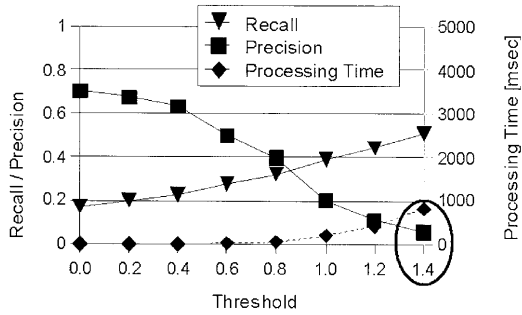
【 1 0 】



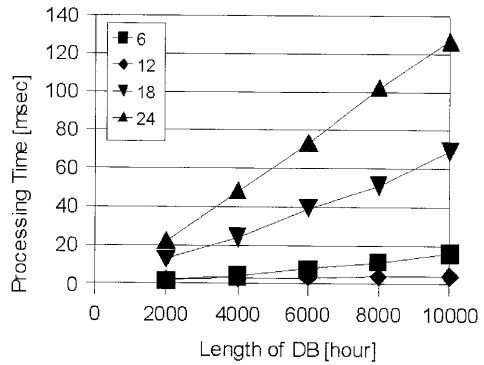
【 1 2 】



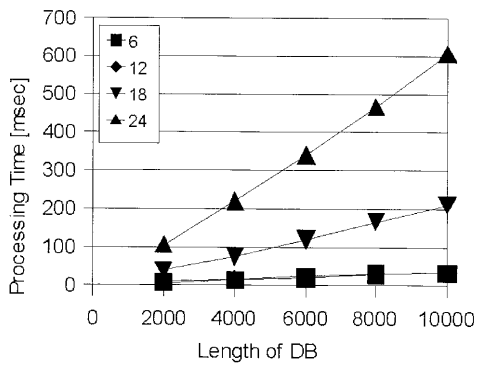
【 1 1 】



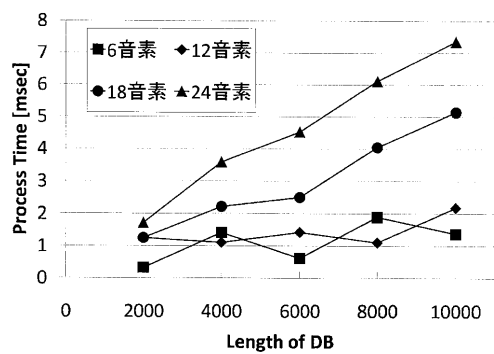
【 1 3 】



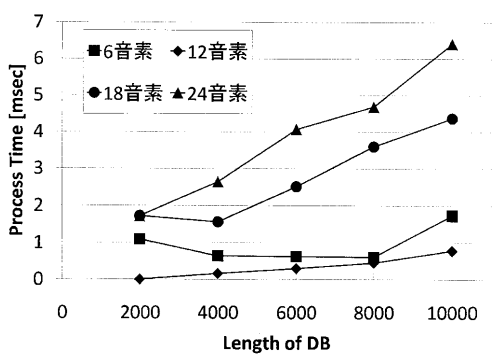
【 1 4 】



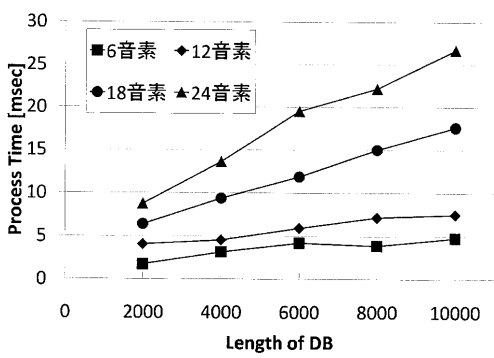
【 1 6 】



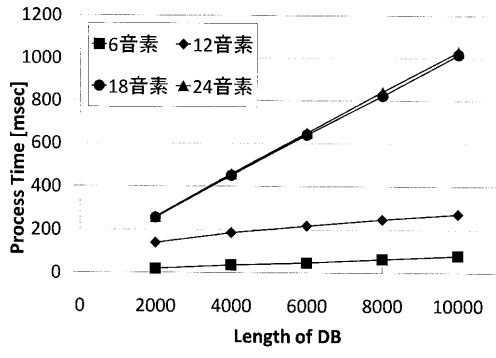
【 1 5 】



【 1 7 】



【 図 1 8 】



---

フロントページの続き

(72)発明者 手島 茂樹

愛知県豊橋市天伯町雲雀ヶ丘1-1 国立大学法人豊橋技術科学大学内

審査官 長 由紀子

(56)参考文献 特開平5-35292(JP,A)

特開2005-257954(JP,A)

山下 達雄 外1名, Suffix Arrayを用いたフルテキスト類似用例検索, 情報処理学会研究報告, 日本, 社団法人情報処理学会, 1997年 9月12日, 第97巻第86号, p.23-30

栗城 吾央 外5名, 連続単語音声認識結果の読み系列を利用した辞書未登録語の音声文書検索, 電子情報通信学会技術研究報告, 日本, 社団法人電子情報通信学会, 2008年 7月10日, 第108巻第142号, p.61-66

(58)調査した分野(Int.Cl., DB名)

G06F 17/30

G10L 15/00

G10L 15/10