

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5318230号  
(P5318230)

(45) 発行日 平成25年10月16日(2013.10.16)

(24) 登録日 平成25年7月19日(2013.7.19)

(51) Int.Cl. F I  
**G 1 O L 15/06 (2013.01)** G 1 O L 15/06 3 O O E  
 G 1 O L 15/06 5 O O Z

請求項の数 11 (全 14 頁)

(21) 出願番号	特願2011-552580 (P2011-552580)	(73) 特許権者	000006013 三菱電機株式会社 東京都千代田区丸の内二丁目7番3号
(86) (22) 出願日	平成22年2月5日(2010.2.5)	(74) 代理人	100123434 弁理士 田澤 英昭
(86) 国際出願番号	PCT/JP2010/000709	(74) 代理人	100101133 弁理士 濱田 初音
(87) 国際公開番号	W02011/096015	(74) 代理人	100173934 弁理士 久米 輝代
(87) 国際公開日	平成23年8月11日(2011.8.11)	(74) 代理人	100156351 弁理士 河村 秀央
審査請求日	平成24年1月17日(2012.1.17)	(72) 発明者	山崎 道弘 東京都千代田区丸の内二丁目7番3号 三菱電機株式会社内

最終頁に続く

(54) 【発明の名称】 認識辞書作成装置及び音声認識装置

(57) 【特許請求の範囲】

【請求項1】

言語同定の対象となる複数の言語のうち、入力された登録対象のテキストの言語としての確からしさを示すスコアの上位から所定数の言語を同定結果として出力する言語同定部と、

前記言語同定部によって同定された前記所定数の各言語の音素で前記登録対象のテキストに読みをそれぞれ付与する読み付与部と、

前記登録対象のテキストの読みを、前記言語同定部によって同定された前記所定数の各言語の音素から音声認識で扱う認識対象言語の音素へそれぞれ変換する読み変換部と、

前記読み変換部によってそれぞれ変換された前記登録対象のテキストの読みを登録した認識辞書を生成する認識辞書生成部

とを備えた認識辞書作成装置。

【請求項2】

前記言語同定部は、前記スコアが所定の閾値未満である場合、前記認識対象言語を同定結果として出力することを特徴とする請求項1記載の認識辞書作成装置。

【請求項3】

言語同定の除外対象の文字又は文字列を記憶する除外対象記憶部と、

前記登録対象のテキストから、前記除外対象記憶部に記憶した前記除外対象の文字又は文字列に相当する部分を除外する言語同定前処理部と、

前記言語同定前処理部によって前記登録対象のテキストに施された前記除外対象の文字

10

20

又は文字列の除外処理の内容を記憶する処理内容記憶部とを備え、

前記言語同定部は、前記言語同定前処理部により前記除外対象の文字又は文字列が除外された前記登録対象のテキストの読みの言語を同定し、

前記認識辞書生成部は、前記処理内容記憶部に記憶された除外処理の内容を参照して、前記除外対象の文字又は文字列の読み及び前記除外対象の文字又は文字列を除外した前記登録対象のテキストの読みから、当該登録対象のテキストの読みを求め、これを登録した認識辞書を生成する

ことを特徴とする請求項 1 記載の認識辞書作成装置。

【請求項 4】

分割対象の文字又は文字列を記憶する分割対象記憶部を備え、

前記言語同定前処理部は、前記分割対象記憶部に記憶した前記分割対象の文字又は文字列で前記登録対象のテキストの文字列を分割し、

前記処理内容記憶部には、前記言語同定前処理部によって前記登録対象のテキストに施された分割処理の内容が記憶され、

前記言語同定部は、前記言語同定前処理部によって分割された前記登録対象のテキストの分割部分ごとに読みの言語を同定し、

前記認識辞書生成部は、前記処理内容記憶部に記憶された分割処理の内容を参照して、前記分割部分ごとの読みから前記登録対象のテキストの読みを求め、これを登録した認識辞書を生成する

ことを特徴とする請求項 3 記載の認識辞書作成装置。

【請求項 5】

前記分割対象記憶部は、数字又は大文字列を含む分割対象の文字又は文字列及びその認識対象言語の読みを記憶しており、

前記言語同定前処理部は、前記登録対象のテキストから、前記数字又は大文字列を含む分割対象の文字又は文字列を除外して、当該登録対象のテキストの文字列を分割し、

前記言語同定部は、前記言語同定前処理部によって分割された前記登録対象のテキストの分割部分ごとに読みの言語を同定し、

前記読み付与部は、前記数字又は大文字列を含む分割対象の文字又は文字列に対して、前記言語同定部によって同定された前記分割部分の言語の読みを付与し、

前記認識辞書生成部は、前記処理内容記憶部に記憶された除外処理の内容を参照して、前記分割部分ごとの読みと、前記分割対象の文字又は文字列の認識対象言語の読み及び前記分割部分の言語から前記認識対象言語へ変換された読みから、前記登録対象のテキストの読みを求め、これを登録した認識辞書を生成する

ことを特徴とする請求項 4 記載の認識辞書作成装置。

【請求項 6】

前記言語同定部は、言語同定の対象となる言語ごとの N - g r a m 及びその出現確率を含む言語同定モデルを用いて、前記登録対象のテキストの N - g r a m の連鎖確率を言語ごとに算出し、前記連鎖確率値に基づく尤度から読みの言語を同定することを特徴とする請求項 1 記載の認識辞書作成装置。

【請求項 7】

前記言語同定部は、前記登録対象のテキストのうち、言語同定の対象となる複数の言語で共通に記述されて言語同定に寄与しない文字又は文字列については特殊文字に置換して N - g r a m を生成することを特徴とする請求項 6 記載の認識辞書作成装置。

【請求項 8】

前記言語同定部は、前記登録対象のテキストに使用言語が限定される文字又は文字列が含まれる場合、前記使用言語のうち、尤度が最も高い言語を同定結果として出力することを特徴とする請求項 6 記載の認識辞書作成装置。

【請求項 9】

前記言語同定部は、認識対象語彙が使用される処理ごとに言語同定モデルを備え、前記登録対象のテキストが使用される処理に対応する言語同定モデルを言語同定に用いること

10

20

30

40

50

を特徴とする請求項 6 記載の認識辞書作成装置。

【請求項 10】

言語同定の対象となる複数の言語のうち、入力された登録対象のテキストの言語としての確からしさを示すスコアの上位から所定数の言語を同定結果として出力する言語同定部と、

前記言語同定部によって同定された前記所定数の各言語の音素で前記登録対象のテキストに読みをそれぞれ付与する読み付与部と、

前記登録対象のテキストの読みを、前記言語同定部によって同定された前記所定数の各言語の音素から音声認識で扱う認識対象言語の音素へそれぞれ変換する読み変換部と、

前記読み変換部によってそれぞれ変換された前記登録対象のテキストの読みを登録した認識辞書を生成する認識辞書生成部と、

前記認識辞書生成部によって生成された前記認識辞書を参照して、入力音声の音声認識を行う音声認識部

とを備えた音声認識装置。

【請求項 11】

言語同定の除外対象の文字又は文字列を記憶する除外対象記憶部と、

分割対象の文字又は文字列を記憶する分割対象記憶部と、

前記除外対象記憶部及び前記分割対象記憶部に記憶された内容に基づいて、入力された登録対象のテキストから、前記除外対象の文字又は文字列を除外するとともに、前記分割対象の文字又は文字列で分割する言語同定前処理部と、

前記言語同定前処理部により前記登録対象のテキストに施された処理の内容を記憶する処理内容記憶部と、

言語同定の対象となる複数の言語のうち、前記言語同定前処理部により前記処理が施された前記登録対象のテキストの言語としての確からしさを示すスコアの上位から所定数の言語を同定結果として出力する言語同定部と、

前記言語同定部によって同定された前記所定数の各言語の音素で前記登録対象のテキストに読みをそれぞれ付与する読み付与部と、

前記登録対象のテキストの読みを、前記言語同定部によって同定された前記所定数の各言語の音素から音声認識で扱う認識対象言語の音素へそれぞれ変換する読み変換部と、

前記処理内容記憶部に記憶された前記処理の内容を参照して、前記読み変換部によってそれぞれ変換された前記登録対象のテキストの読みを求めて、これを登録した認識辞書を生成する認識辞書生成部と、

前記認識辞書生成部によって生成された前記認識辞書を参照して、入力音声の音声認識を行う音声認識部

とを備えた音声認識装置。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、音声認識の対象となる語彙の辞書を作成する認識辞書作成装置及びこれを用いた音声認識装置に関するものである。

【背景技術】

【0002】

特許文献 1 には、音声認識の対象となる複数の言語の音響モデルを同時に使用することにより、多言語に対応した音声認識を行うことができる音声認識装置が開示されている。

【0003】

しかしながら、この特許文献 1 の発明では、音声認識の対象となる複数の言語の全てに対応した多言語用の音響モデルを持つ必要があり、一般的な一言語のみに対応した音声認識には適用できないという問題点があった。

また、特許文献 1 の発明は、認識語彙がどの言語で記載されているのかを予め特定して読みを付与しておく必要がある。

10

20

30

40

50

一方、一言語のみに対応する音声認識では、読みが付与されていない登録対象テキストに対して、認識対象言語の読みを自動で生成して認識が行われる。このとき、認識対象言語とは別の言語で記載されたテキストについては読みを付与できない。

【0004】

この発明は、上記のような課題を解決するためになされたもので、認識辞書へ登録する対象となる語彙の言語が不明な場合であっても、当該語彙から音声認識の言語の音素体系に読みを変換した認識辞書を作成できる認識辞書作成装置及びこれを用いた音声認識装置を得ることを目的とする。

【先行技術文献】

【特許文献】

10

【0005】

【特許文献1】特許第3776391号公報

【発明の概要】

【0006】

この発明に係る認識辞書作成装置は、言語同定の対象となる複数の言語のうち、入力された登録対象のテキストの言語としての確からしさを示すスコアの上位から所定数の言語を同定結果として出力する言語同定部と、言語同定部によって同定された所定数の各言語の音素で登録対象のテキストに読みをそれぞれ付与する読み付与部と、登録対象のテキストの読みを、言語同定部によって同定された所定数の各言語の音素から音声認識で扱う認識対象言語の音素へそれぞれ変換する読み変換部と、読み変換部によってそれぞれ変換された登録対象のテキストの読みを登録した認識辞書を生成する認識辞書生成部とを備えるものである。

20

【0007】

この発明によれば、言語同定の対象となる複数の言語のうち、入力された登録対象のテキストの言語としての確からしさを示すスコアの上位から所定数の言語を同定結果として出力し、同定された所定数の各言語の音素で登録対象のテキストに読みをそれぞれ付与するとともに、登録対象のテキストの読みを、同定された所定数の各言語の読みから、音声認識で扱う認識対象言語へそれぞれ変換した読みを登録した認識辞書を生成する。このようにすることで、認識辞書への登録対象のテキスト（語彙）がどの言語で記述されているか不明である場合であっても、音声認識の言語の音素体系に即した認識辞書を得ることができるという効果がある。

30

【図面の簡単な説明】

【0008】

【図1】この発明の実施の形態1による認識辞書作成装置を用いた登録型の音声認識装置の構成を示すブロック図である。

【図2】実施の形態1の認識辞書作成装置による認識辞書の作成動作の流れを示すフローチャートである。

【図3】ドイツ語と英語で発音が類似する音素の対応表の一例を示す図である。

【図4】実施の形態1の認識辞書作成装置による認識辞書の作成動作の流れを示すフローチャートである。

40

【図5】この発明の実施の形態2による認識辞書作成装置を用いた登録型の音声認識装置の構成を示すブロック図である。

【図6】実施の形態2の認識辞書作成装置による認識辞書の作成動作の流れを示すフローチャートである。

【発明を実施するための形態】

【0009】

以下、この発明をより詳細に説明するために、この発明を実施するための形態について、添付の図面に従って説明する。

実施の形態1.

図1は、この発明の実施の形態1による認識辞書作成装置を用いた登録型の音声認識装

50

置の構成を示すブロック図である。図1において、実施の形態1による音声認識装置100は、言語同定部101、読み付与部102、読み変換部103、認識辞書生成部104、認識辞書記憶部105及び音声認識部106を備える。これら構成部のうち、言語同定部101、読み付与部102、読み変換部103、認識辞書生成部104及び認識辞書記憶部105が、実施の形態1による認識辞書作成装置の構成である。

#### 【0010】

言語同定部101は、認識辞書への登録対象となる語彙のテキスト文字列（以下、登録対象テキストと呼ぶ）の言語を同定する構成部である。登録対象テキストには、言語を特定しにくい語彙のテキスト文字列も含まれ、例えば、携帯音楽プレーヤに登録されている楽曲のタイトルやアーティスト名等の書誌データ、携帯電話に登録されている地名や人名等が挙げられる。読み付与部102は、言語同定部101によって同定された言語で登録対象テキストに読みを音素で付与する構成部である。読み変換部103は、読み付与部102によって付与された読みを、音声認識部106が行う音声認識で使用する言語へ音素変換する構成部である。

10

#### 【0011】

認識辞書生成部104は、読み変換部103により変換された音素を、音声認識の対象となる語彙（以下、認識対象語彙と呼ぶ）として生成し、認識辞書記憶部105の認識辞書へ登録する構成部である。認識辞書記憶部105は、認識辞書生成部104及び音声認識部106によって読み書き可能な記憶部であって、認識辞書生成部104により生成された認識対象語彙を登録する認識辞書が記憶される。音声認識部106は、認識辞書記憶部105に記憶された認識辞書の認識対象語彙を用いて音声認識を実行し、認識結果を出力する構成部である。

20

#### 【0012】

また、言語同定部101、読み付与部102、読み変換部103、認識辞書生成部104、認識辞書記憶部105及び音声認識部106は、この発明の趣旨に従う認識辞書作成プログラムをコンピュータに記憶し、CPUに実行させることにより、ハードウェアとソフトウェアが協働した具体的な手段として当該コンピュータ上で実現することができる。さらに、認識辞書記憶部105及び音声認識部106で用いる記憶領域は、上記コンピュータに搭載された記憶装置、例えばハードディスク装置や外部記憶メディア等に構築される。

30

#### 【0013】

次に動作について説明する。

図2は、実施の形態1の認識辞書作成装置による認識辞書の作成動作の流れを示すフローチャートであり、1つの登録対象テキストに対する、言語同定部101、読み付与部102、読み変換部103及び認識辞書生成部104による動作の詳細を述べる。

まず、言語同定部101が、登録対象テキストの文字列に言語同定処理を開始し、当該文字列がどの言語で記載されたものであるかを判定する（ステップS T 2 0 1）。具体的には、言語同定部101に設定される複数の言語のうち、入力された登録対象テキストの文字列が、どの言語に該当するかが判定される。

例えば、言語同定部101において、英語、ドイツ語、フランス語、イタリア語、スペイン語、オランダ語の欧州6言語が、言語同定の対象として設定されている場合、入力された登録対象テキストの文字列が“Guten Morgen”であると、言語同定部101は、当該文字列の言語がドイツ語であるとの言語同定の結果を出力する。

40

言語の同定に失敗する等して言語が同定できなかった場合、言語同定部101は、音声認識部106が音声認識の対象として受理可能な言語を同定結果として出力する。

#### 【0014】

また、言語同定部101の言語同定には、例えば、文字のN-gramが使用される。N-gram自体は、Claude Elwood Shannonにより確立された言語情報に関する分析モデルであり、音声認識のための言語モデルや全文検索等に使用されている。これらの一般的なN-gramの利用方法については、下記の参考文献1に記載

50

されている。

(参考文献1) "A Mathematical Theory of Communication", C.E. SHANNON, The Bell System Technical Journal, Vol. 27, pp. 379-423, 623-656, July, October, 1948.

【0015】

このように、N-gramの概念をテキスト文字列の言語同定に取り入れることで、精度良く言語を同定することが可能である。例えば、Nを3として、文字のTrigramを用いた言語同定について説明する。

言語同定部101は、言語同定の対象となる各言語で記載された学習用テキスト文字列を有しており、学習用文字列中に現れる文字の3連鎖の出現確率を求める。例えば、文字列“MITSUBISHI”の場合、“\$\$M”、“\$MI”、“MIT”、“ITS”、  
10 “TSU”、“SUB”、“UBI”、“BIS”、“ISH”、“SHI”、“HI\$”、“I\$\$”の各連鎖に分解できる。なお、“\$”は、語頭、語尾を表す文字を意味する。

【0016】

言語 $i$  ( $i = 1, 2, \dots, N$ )の全学習文字列中に各連鎖が出現する頻度を求めた後、言語同定部101は、学習された言語ごとに連鎖の出現確率の上位 $n$ 個をモデルとして採用し、文字の連鎖とその出現頻度を各言語の言語同定モデルに記憶する。例えば、言語 $i$ の言語同定モデルには、文字連鎖(Trigram)“\$\$M”と、そのときの出現確率 $P_i(\$ , \$ , M)$ 等が記憶される。

【0017】

言語同定部101は、登録対象テキストの文字列に対して、各言語の言語同定モデルを用いて言語ごとに文字連鎖(Trigram)の連鎖確率を求めていき、連鎖確率値が最も大きくなる言語を言語の同定結果として採用する。例えば、“Best”に対する言語 $i$ の連鎖確率 $P_i$ は、 $P_i = P_i(\$ , \$ , B) \times P_i(\$ , B , E) \times P_i(B , E , S) \times P_i(E , S , T) \times P_i(S , T , \$) \times P_i(T , \$ , \$)$ で求められる。この連鎖確率 $P_i$ が最大となる言語 $i$ が、言語の同定結果となる。

【0018】

なお、言語同定部101は、N-gramが言語同定モデルに記憶されていない場合には、出現確率として予め定めた一定確率を与えて計算を行う。

N-gramを求めるにあたり、言語同定の対象となる複数の言語で共通に記述され、言語同定に寄与しない文字(例えば、数字、括弧やピリオド等の記号)については、これらの文字を特殊文字を表す文字に予め置き換えてN-gramを求めても構わない。例えば、#、@等の特殊文字を使用する。

【0019】

また、言語同定において、使用される言語が限定される文字が使用されていた場合は、当該文字が使用されている言語のうち、尤度が最も高い言語を同定結果として出力してもよい。使用される言語が限定される文字としては、例えば、ウムラウトを用いた文字等が挙げられる。

なお、N-gramの学習に用いるテキスト文字列(言語同定モデルに用いるテキスト文字列)を、認識対象語彙と同じタスクの語彙を用いて学習を行うことにより、言語の同定精度を向上させることが可能である。タスクとは、例えば音楽検索や住所認識等の認識対象語彙が使用される処理を意味する。言語同定部101は、タスクごとの学習用文字列を備え、登録対象テキストが使用されるタスクに対応する学習用文字列を言語同定に用いる。

【0020】

次に、読み付与部102は、言語同定部101から同定結果を入力すると、音声認識装置100に設定されている複数の言語のうち、当該同定結果がいずれの言語であるかを判定し(ステップST202)、入力された登録対象テキストの文字列に対して、判定結果の言語による音素で読みを付与する(ステップST203)。音声認識部106にて現在音声認識の対象とされている認識対象言語であると、この認識対象言語による音素で読み  
50

が付与される。同様に、判定結果の言語が、認識対象言語以外の言語 1, 2, . . . , N のいずれかであっても、図 2 に示すように、当該言語の音素で読みが付与される。

なお、音素の付与には、例えば G 2 P (Grapheme to Phoneme) を用いる。また、この読み付与処理において、省略語の判定や記号等の処理のように言語に依存した処理も併せて行う。

#### 【 0 0 2 1 】

読み変換部 1 0 3 は、ステップ S T 2 0 3 において、認識対象言語以外の言語の音素で読みが付与された登録対象テキストに対して、各言語による音素の読みを、認識対象言語による音素の読みに変換する (ステップ S T 2 0 4)。

このように音素体系を変換する理由は、音声認識部 1 0 6 にて受理可能な音素体系が、音声認識の対象としている認識対象言語のみであり、言語が異なる音素体系は受理されない音素が存在するためである。

このような音素 (読み) の変換方法としては、例えば、読み変換部 1 0 3 が、音声認識部 1 0 6 で受理できない言語の音素又は音素の系列に対して、認識対象言語のうち、最も近い音素又は音素の系列を、対応表として予め用意しておき、ステップ S T 2 0 3 で得られたテキスト文字列の読みに対して、上記対応表に応じて読みの変換 (音素マッピング) を行う。

#### 【 0 0 2 2 】

図 3 は、上述のような対応表の一例を示す図であり、ドイツ語と英語の間での対応関係を示している。ドイツ語における発音 / a / (非円唇前舌広母音) や / Y / (円唇前舌め広めの狭母音) 等は、イギリス英語の発音体系にない。このため、音声認識部 1 0 6 が、イギリス英語を受理する場合、当該読みに対応していない。

そこで、ドイツ語の発音 / a / や / Y / については、図 3 に示す対応表のように、音声認識部 1 0 6 にて受理可能なイギリス英語に存在する音素のうち、発音が最も近い音素、例えば、/ { / (非円唇前舌狭めの広母音) や / } / (円唇前舌半狭母音) と対応付けておく。なお、ここでの発音表記は、X - S A M P A 表記を用いている。

この対応表は、言語的に近いもの同士を対応付けてもよいが、例えば、各言語の発音がどの音素表記で認識しやすいか等に基づいて対応関係を決定してもよい。

#### 【 0 0 2 3 】

認識辞書生成部 1 0 4 は、ステップ S T 2 0 3 で読み付与部 1 0 2 によって登録対象テキストの文字列に付与された音素、又はステップ S T 2 0 4 で読み変換部 1 0 3 によって変換された音素を入力し、音声認識部 1 0 6 が参照可能な形式の認識辞書を生成する (ステップ S T 2 0 5)。例えば、認識語彙をバイナリデータに変換する他、必要に応じて形態素解析、単語分割を行って言語制約の作成等を行って認識辞書を得る。

登録対象テキストである語彙が複数ある場合、各登録対象テキストに対して、上述までの処理を繰り返す。なお、認識辞書の生成は、一語彙ずつの追加登録ではなく、全ての登録対象テキストの語彙に読みを付与した後にまとめて行うようにしても構わない。

認識辞書生成部 1 0 4 によって生成された認識辞書は、認識辞書記憶部 1 0 5 に記憶される。

#### 【 0 0 2 4 】

音声認識部 1 0 6 は、認識辞書記憶部 1 0 5 に記憶された認識辞書に記載の認識語彙や文法を参照して入力音声の音声認識を行い、認識結果を出力する。例えば、音声認識部 1 0 6 が、特定の言語の音素体系の音素で記載された認識辞書を読み込み、特定の言語の入力音声を音声認識する。音声認識のアルゴリズムとしては、例えば H M M (隠れマルコフモデル) や D P (Dynamic Programming) マッチング等を用いる。

#### 【 0 0 2 5 】

上述まででは、言語同定部 1 0 1 で同定された言語が 1 言語である場合の動作を説明したが、続いて、入力された登録対象テキストの文字列から、言語同定部 1 0 1 により複数の言語が同定された場合の動作について説明する。

図 4 は、実施の形態 1 の認識辞書作成装置による認識辞書の作成動作の流れを示すフロ

10

20

30

40

50

ーチャートであり、言語同定部 101 に N 個の言語が同定された場合を示している。

言語同定部 101 は、登録対象テキストの文字列に対して言語同定処理を開始し、当該文字列がどの言語で記載されたものであるかを判定して、当該文字列の言語として確からしい上位 N 個の言語を言語同定結果とする（ステップ ST301）。

なお、ここで、N 個は固定値としてもよいが、言語同定部 101 で信頼度を表すスコアを出力し、このスコアが予め定めた閾値以上の信頼度である候補の数や、1 位の信頼度に対する一定の信頼度差以内の候補の数としても構わない。

例えば、登録対象テキストとして“Hamburg”が入力された場合、言語同定部 101 は、このテキストから同定される言語がドイツ語と英語である（ $N = 2$ ）との結果を出力する。また、言語の同定に失敗する等して言語が同定できなかった場合や、信頼度のスコアが閾値未満である場合、言語同定部 101 は、音声認識部 106 で受理可能な言語（認識対象言語）を、言語の同定結果として出力する。

#### 【0026】

次に、言語同定部 101 は、同定結果として得られた言語の数、すなわち言語ごとに行われる後述の処理の回数を示すカウンタ  $i$  を  $i = 0$  に初期化（ステップ ST301 - 1）し、 $i$  番目の同定結果に相当する言語を読み付与部 102 に設定する。これにより、読み付与部 102、読み変換部 103 及び認識辞書生成部 104 が、 $i$  ( $i = 0 \sim N$ ) 番目の同定結果に相当する言語で、ステップ ST302 からステップ ST306 までの処理を実行する。なお、ステップ ST302 は、図 2 で示したステップ ST202 と同様の処理であり、ステップ ST303 は、図 2 で示したステップ ST203 と同様の処理、ステップ ST304 は、図 2 で示したステップ ST204 と同様の処理、ステップ ST305 は、図 2 で示したステップ ST205 と同様の処理である。

#### 【0027】

ステップ ST306 において、言語同定部 101 は、カウンタ  $i$  を +1 インクリメントして次の同定結果の言語で上記一連の処理を繰り返す。ステップ ST307 において、言語同定部 101 が、カウンタ  $i$  のカウント値に基づいて、全ての同定結果の言語に対して上記一連の処理が完了（ $i = N + 1$ ）したと判定すると、入力された登録対象テキストに対する登録処理を終了する。これにより、1 つの登録対象テキストが、複数の言語で記述される場合であっても、これらの言語を同定しその音素で読みをそれぞれ付与してから、認識対象言語の音素による読みをそれぞれ変換することで、認識語彙として認識辞書に登録できる。従って、ユーザが、言語同定部 101 が同定したいいずれの言語で当該テキスト文字列の発話を行っても、認識辞書に登録された対応する認識語彙で音声認識することが可能である。

#### 【0028】

なお、登録対象テキストが複数ある場合は、同定結果として 1 つの言語が得られた場合と同様に、各登録対象テキストに対して、上述の処理を繰り返す。

認識辞書の生成は、1 つの登録対象テキストに対する言語同定結果として得られた言語ごとに追加登録するのではなく、1 つの登録対象テキストの語彙で同定された全ての言語をまとめて追加登録してもよい。若しくは、全ての登録対象テキストの語彙に対して読みを付与した後にまとめて行っても構わない。

#### 【0029】

以上のように、この実施の形態 1 によれば、入力された登録対象テキストの読みの言語を同定し、同定された言語の音素で登録対象テキストに読みを付与するとともに、登録対象テキストの読みを、同定された言語の読みから、音声認識で扱う認識対象言語へ変換した読みを登録した認識辞書を生成する。

このようにすることで、登録対象テキストがどの言語が不明である場合（例えば、携帯音楽プレーヤに登録されている楽曲に対するタイトルやアーティスト名などの書誌データや、携帯電話に登録されている地名や人名等）であっても、N-gram を用いた言語同定モデルで登録対象テキストの言語を同定し、同定された言語による音素の付与を行い、音声認識で受理可能な言語の音素に変換することにより、音声認識で参照される認識語彙

10

20

30

40

50



として登録することができる。

【0030】

また、この実施の形態1によれば、言語同定部101が、登録対象テキストから複数の言語を同定することで、登録対象テキストが複数の言語に該当する可能性がある場合に、各言語での読みを付与して認識語彙として登録する。このようにすることで、ユーザが、どちらの言語での発音をしても音声認識が可能である。

【0031】

さらに、この実施の形態1によれば、言語同定部101が、言語の同定結果に対して、各言語に対する信頼度を表すスコアを設定し、このスコアに関する所定の閾値と比較して信頼度が高い言語を最終的な同定結果として出力する。このようにすることで、ユーザに発声される可能性が高い言語の発音のみを採用でき、音声認識性能を向上させることができる。

10

【0032】

さらに、この実施の形態1によれば、N-gramを用いて言語同定を行うので、言語ごとに単語辞書等を用意して言語判定を行う場合と比較して、安定した言語同定性能を得ることができる。また、同時に、辞書サイズを小さくすることができ、演算量やメモリ消費量を削減することが可能となる。

【0033】

さらに、この実施の形態1によれば、N-gramを用いた言語同定において、言語の同定に寄与しない文字（例えば、数字や、括弧、ピリオド等の記号）を、特殊文字を表す1文字に置き換えたN-gramを用いる。これにより、言語同定モデルの記憶領域のサイズを低減でき、言語同定モデルの検索時間やメモリ消費量を削減することができるので、組み込み型の装置等の演算速度やメモリサイズに制限がある装置への本発明の適用が容易である。

20

【0034】

さらに、この実施の形態1によれば、ウムラウト等のように、使用される言語が限定される文字が登録対象テキストに存在する場合には、当該文字が使用される言語のうちから言語を同定することにより、言語同定の精度を向上させることが可能となる。

【0035】

さらに、この実施の形態1によれば、N-gramを用いた言語同定において、言語同定モデルを、認識対象語彙と同じタスク（認識対象語彙が使用された処理）の語彙を用いて作成することにより、言語同定の精度を向上させることが可能である。

30

【0036】

実施の形態2.

図5は、この発明の実施の形態2による認識辞書作成装置を用いた登録型の音声認識装置の構成を示すブロック図である。図5において、実施の形態2による音声認識装置100Aは、上記実施の形態1による音声認識装置の構成に加え、言語同定前処理部107、定型文字列記憶部（除外対象記憶部）108、分割文字列記憶部（分割対象記憶部）109及び変換処理記憶部（処理内容記憶部）110を備える。

言語同定前処理部107は、言語同定部101の前段に配置され、登録対象テキストが入力される構成部であり、言語同定処理を行う前処理として、入力された登録対象テキストの文字列における、特定の文字又は文字列（以下、定型文字又は定型文字列と呼ぶ）を言語同定の対象から除外したり、所定の文字又は文字列（以下、分割文字又は分割文字列と呼ぶ）を基準として、当該登録対象テキストを分割する。

40

【0037】

定型文字列記憶部108は、言語同定の対象から除外する定型文字又は定型文字列及びその記述言語と読みが記憶された記憶部である。分割文字列記憶部109は、登録対象テキストを分割する際の分割位置となる分割文字又は分割文字列及びその記述言語と読みが記憶された記憶部である。変換処理記憶部110は、言語同定前処理部107によって登録対象テキストに施された前処理の内容（文字列除外や分割）を記憶する記憶部である。

50

## 【 0 0 3 8 】

なお、図5において、図1と同一又は同様に動作する構成部には、同一符号を付して説明を省略する。ただし、言語同定部101は、言語同定前処理部107によって前処理が施された登録対象テキストの文字列に対して言語の同定を行う。また、認識辞書生成部104は、認識辞書を生成するにあたり、変換処理記憶部110に記憶されている前処理内容、つまり登録対象テキストの分割された各分割部分間の接続関係や、登録対象テキストから除外された文字列及びその読みを用いて、当該登録対象テキストの文字列にあった認識辞書を生成する。

## 【 0 0 3 9 】

また、言語同定前処理部107、定型文字列記憶部108、分割文字列記憶部109及び変換処理記憶部110は、この発明の趣旨に従う認識辞書作成プログラムをコンピュータに記憶し、CPUに実行させることにより、ハードウェアとソフトウェアが協働した具体的な手段として当該コンピュータ上で実現することができる。さらに、定型文字列記憶部108、分割文字列記憶部109及び変換処理記憶部110で用いる記憶領域は、上記コンピュータに搭載された記憶装置、例えばハードディスク装置や外部記憶メディア等に構築される。

## 【 0 0 4 0 】

次に動作について説明する。

図6は、実施の形態2の認識辞書作成装置による認識辞書の作成動作の流れを示すフローチャートであり、言語同定前処理部107、言語同定部101、読み付与部102、読み変換部103及び認識辞書生成部104による動作の詳細を述べる。

先ず、言語同定前処理部107は、登録対象テキストを入力すると、定型文字列記憶部108の記憶内容を参照して、当該登録対象テキストの文字列に含まれる定型文字又は定型文字列を検出し、言語同定の対象から除外する(ステップS T 5 0 1)。定型文字又は定型文字列としては、複数の言語に共通して現れる特定の言語による記述が挙げられる。

## 【 0 0 4 1 】

楽曲の書誌データを対象とするシステムにおいて、“Disc”や“Best of”等は、言語に依らずに、複数の言語で共通に記載(英語表記)される。そこで、このような文字又は文字列及びその記述言語(上記例では英語)を定型文字列記憶部108に記憶しておく。例えば、“Best Of C e n ' e s t q u e m o i ”という登録対象テキストが入力された場合、“Best Of”は、定型文字列記憶部108の記憶内容から英語とその読みが得られるため、言語同定前処理部107により言語同定の対象から削除され、言語同定対象の文字列として、“C e n ' e s t q u e m o i ”が残る。

## 【 0 0 4 2 】

続いて、言語同定前処理部107は、分割文字列記憶部109の記憶内容を参照して、登録対象テキストの文字列に含まれる分割文字又は分割文字列を検出し、分割文字又は分割文字列を基準として、登録対象テキストの文字列を分割する(ステップS T 5 0 2)。分割文字又は分割文字列としては、登録対象テキストの記述を区切る文字である“( ”や“ ) ”、“ - ”等が挙げられる。

例えば、“J e v i v r a i s a n s t o i ( I W i l l S a y G o o d b y e ) ”という登録対象テキストが入力された場合、分割文字列記憶部109の記憶内容を参照して、言語同定前処理部107は、当該登録対象テキストの文字列に含まれる、括弧“( ”や“ ) ”を検出し、これらの文字を基準として当該登録対象テキストの文字列を分割する。これにより、“J e v i v r a i s a n s t o i ”と“ I W i l l S a y G o o d b y e ”の2つの文字列に分割される。

## 【 0 0 4 3 】

次に、言語同定前処理部107は、分割文字列記憶部109の記憶内容を参照して、登録対象テキストの文字列に含まれる、例えば、大文字列(スペル)や数字等のように表記言語によらず、認識対象言語で読み上げられる文字列を特定し、この文字列を登録対象テ

10

20

30

40

50

キストから除外して分割する（ステップ S T 5 0 3）。この文字列部分では、言語同定の結果として言語が特定されず、認識対象言語となるため、認識対象言語による読みが付与される。

【 0 0 4 4 】

なお、この文字列部分（大文字列（スペル）や数字等の文字列）に対して、分割文字列記憶部 1 0 9 に予め認識対象言語での読みを付与して記憶しておくとともに、登録対象テキストにおける、この文字列部分を境とした前後の文字列に対して言語同定部 1 0 1 が同定した言語で読み付与部 1 0 2 が当該文字列部分に読みを付与するようにしてもよい。

このように、大文字列（スペル）や数字等の文字列部分に対して、認識対象言語と同定結果の言語との 2 種類の読みを付与することで、登録対象テキストにおける、この文字列部分以外の部分を記述する言語（同定結果の言語）に依らず、認識対象言語で発話される文字列部分に正しく読みを付与することが可能となる。

10

【 0 0 4 5 】

言語同定前処理部 1 0 7 は、ステップ S T 5 0 1 からステップ S T 5 0 3 までの前処理の内容を変換処理記憶部 1 1 0 に記憶する。前処理の内容として、登録対象テキストから除外された定型文字又は定型文字列、その記述言語及びこの言語の音素による読み、登録対象テキストの分割位置となる分割文字又は分割文字列、各分割部分間の接続関係等が、変換処理記憶部 1 1 0 に記憶される。

【 0 0 4 6 】

この後、言語同定前処理部 1 0 7 は、登録対象テキストを分割した分割数  $K$  をカウントするカウンタ  $j$  を初期化（ $j = 0$ ）する（ステップ S T 5 0 4）。続いて、言語同定前処理部 1 0 7 は、 $j$  番目の分割部分を言語同定の対象となる文字列に設定して言語同定部 1 0 1 へ出力する（ステップ S T 5 0 5）。

20

【 0 0 4 7 】

言語同定部 1 0 1 は、言語同定前処理部 1 0 7 から入力した  $j$  番目の分割部分の文字列に対して、上記実施の形態 1 と同様な手順で言語同定処理を開始し、当該文字列がどの言語で記載されたものであるかを判定して、当該文字列の言語として確からしい上位  $N$  個の言語（尤度の高い上位  $N$  個の言語）を言語同定結果とする（ステップ S T 5 0 6）。

【 0 0 4 8 】

次に、言語同定部 1 0 1 は、同定結果として得られた言語の数、すなわち言語ごとに行われる後述の処理の回数を示すカウンタ  $i$  を  $i = 0$  に初期化（ステップ S T 5 0 7）し、 $i$  番目の同定結果に相当する言語を読み付与部 1 0 2 に設定する。

30

このようにして、読み付与部 1 0 2 及び読み変換部 1 0 3 が、 $i$ （ $i = 0 \sim N$ ）番目の同定結果に相当する言語でステップ S T 5 0 8 からステップ S T 5 1 0 までの処理を実行する。なお、ステップ S T 5 0 8 は、図 2 で示したステップ S T 2 0 2 と同様の処理であり、ステップ S T 5 0 9 は、図 2 で示したステップ S T 2 0 3 と同様の処理であり、ステップ S T 5 1 0 は、図 2 で示したステップ S T 2 0 4 と同様の処理である。

【 0 0 4 9 】

ステップ S T 5 1 1 において、言語同定部 1 0 1 は、カウンタ  $i$  を + 1 インクリメントして次の同定結果の言語で上記一連の処理を繰り返し、ステップ S T 5 1 2 において、全ての同定結果の言語に対して上記一連の処理が完了（ $i = N + 1$ ）したと判定されると、カウンタ  $j$  を + 1 インクリメントする（ステップ S T 5 1 3）。ステップ S T 5 1 4 で、言語同定部 1 0 1 が、カウンタ  $j$  のカウント値に基づいて、全ての分割部分の文字列に対して処理が完了（ $j = K$ ）したと判定するまで、分割部分の文字列のそれぞれに対して、ステップ S T 5 0 5 からステップ S T 5 1 4 までの一連の処理が繰り返される。

40

【 0 0 5 0 】

認識辞書生成部 1 0 4 は、変換処理記憶部 1 1 0 に記憶された前処理内容を参照して、登録対象テキストの文字列から除外された文字列に対する読みを特定し、当該言語同定の対象から除外された文字列の読みと、読み変換部 1 0 3 から入力した各分割部分の文字列に付与された読みとを合成し、音声認識部 1 0 6 が参照可能な形式の認識辞書を生成する

50

(ステップST515)。例えば、認識語彙をバイナリデータに変換する他、必要に応じて形態素解析、単語分割を行って言語制約の作成等を行って認識辞書を得る。

【0051】

以上のように、この実施の形態2によれば、定型文字列記憶部108、分割文字列記憶部109、これら記憶部108, 109に記憶された内容に基づいて、登録対象テキストに前処理を行う言語同定前処理部107と、言語同定前処理部107による前処理の内容を記憶する変換処理記憶部110とを備え、認識辞書生成部104が、変換処理記憶部110の記憶内容に基づいて、言語同定前処理部107によって前処理が施された登録対象テキストの読みを表す音素を求め、これを登録した認識辞書を生成する。

このように、特定文字列により入力テキストを分割し、分割部分に対してそれぞれ言語同定処理、読み付与処理を行うことで、登録対象テキストに複数の言語が含まれる場合であっても、登録対象テキストの各分割部分を正しく言語同定することが可能となる。

また、複数の言語に含んで使用される特定言語の単語/フレーズを、特定文字列として予め記憶し、言語同定の対象から除外することにより、言語に関係なく使用される言語のフレーズ(例えば、音楽のアルバムタイトルで“Disc 1”、“Best of”等)を含む場合においても、当該フレーズ以外の文字列部分で言語同定を行うことで各文字列部分の読みの言語を正しく言語同定することが可能となる。

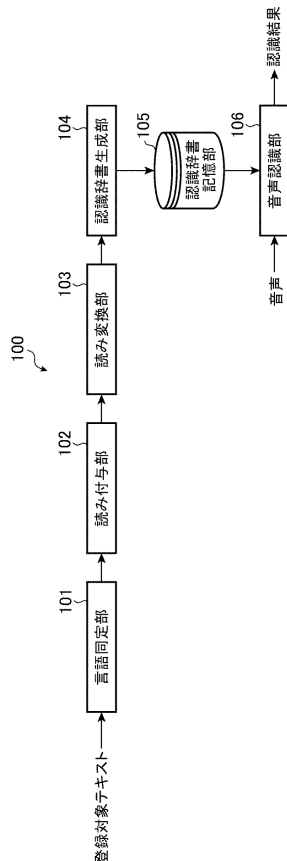
また、認識対象言語に依存した読みをする箇所を分割することにより、当該箇所についても、正しく読みを付与することが可能となる。

【産業上の利用可能性】

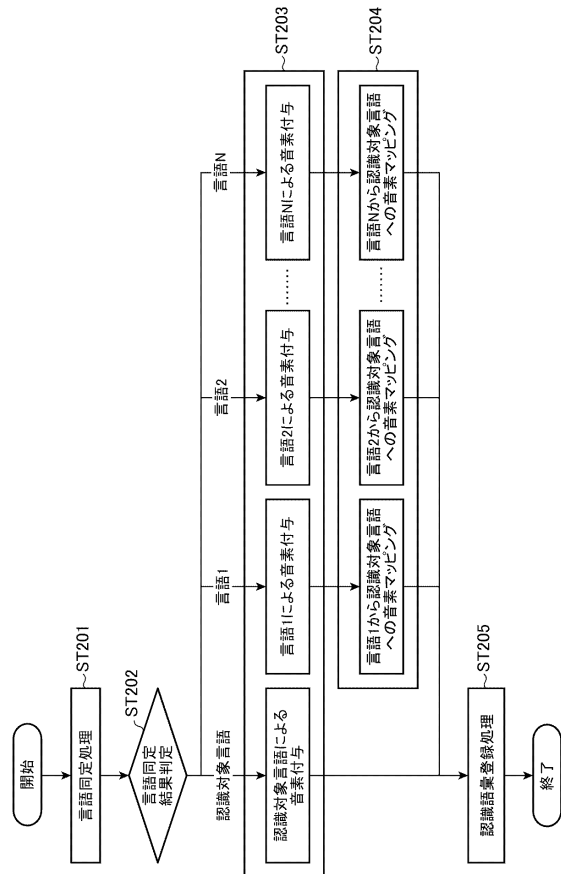
【0052】

この発明に係る認識辞書作成装置によれば、記述言語が不明な語彙から音声認識の音素体系における認識辞書を作成することができるので、複数の言語の語彙が混在するデータが扱われる、携帯音楽プレーヤ、携帯電話、車載ナビゲーションシステムなどの音声認識装置に適している。

【図1】



【図2】



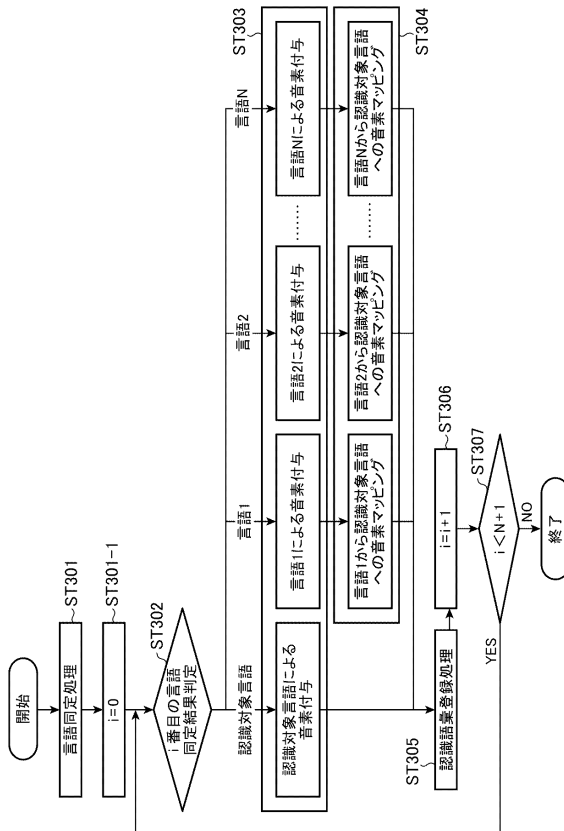
10

20

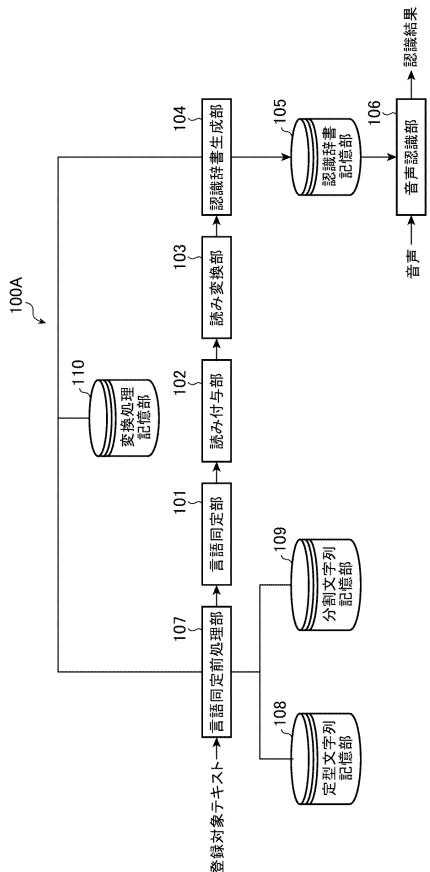
【図3】

German	English
/a/	/ʌ/
/E/	/E/
/I/	/I/
/O/	/O/
/Y/	/Y/
/E:/	/E@/
.....	.....
/Z/	/Z/
/x/	/h/
/h/	/h/
/m/	/m/
.....	.....

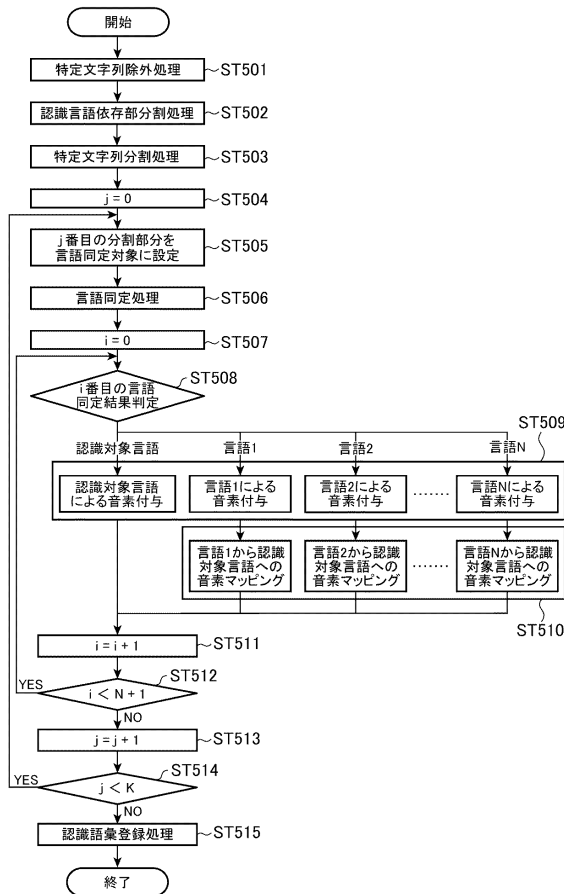
【図4】



【図5】



【図6】



---

フロントページの続き

- (72)発明者 石井 純  
東京都千代田区丸の内二丁目7番3号 三菱電機株式会社内
- (72)発明者 石川 泰  
東京都千代田区丸の内二丁目7番3号 三菱電機株式会社内

審査官 田川 泰宏

- (56)参考文献 特開2009-037633(JP,A)  
特開2004-271895(JP,A)  
特開2006-059105(JP,A)  
住友 徹,森田 和宏,泓田 正雄,青江 順一, "トライ構造を用いた多言語対訳辞書の効率的圧縮手法", 第61回(平成12年度後期)全国大会講演論文集(2), 日本, 社団法人情報処理学会, 2000年10月 3日, p.2-133 - 2-134

- (58)調査した分野(Int.Cl., DB名)  
G10L 15/06