

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5799733号  
(P5799733)

(45) 発行日 平成27年10月28日(2015.10.28)

(24) 登録日 平成27年9月4日(2015.9.4)

(51) Int.Cl. F I  
**G 1 0 L 15/18 (2013.01)** G 1 0 L 15/18 3 0 0 G  
**G 1 0 L 15/06 (2013.01)** G 1 0 L 15/06 3 0 0 C

請求項の数 7 (全 32 頁)

<p>(21) 出願番号 特願2011-225344 (P2011-225344)                  (22) 出願日 平成23年10月12日(2011.10.12)                  (65) 公開番号 特開2013-83897 (P2013-83897A)                  (43) 公開日 平成25年5月9日(2013.5.9)                  審査請求日 平成26年6月3日(2014.6.3)</p>	<p>(73) 特許権者 000005223                  富士通株式会社                  神奈川県川崎市中原区上小田中4丁目1番                  1号                  (74) 代理人 100089118                  弁理士 酒井 宏明                  (72) 発明者 原田 将治                  神奈川県川崎市中原区上小田中4丁目1番                  1号 富士通株式会社内                  審査官 千本 潤介</p>
---	--

最終頁に続く

(54) 【発明の名称】 認識装置、認識プログラムおよび認識方法

(57) 【特許請求の範囲】

【請求項1】

文章に含まれる単語と該単語の文章内の位置を示す位置情報とを記憶する記憶部と、  
 入力された音声信号と、前記記憶部に記憶された複数の単語を接続した文字列の読み情報とを比較して、類似度を算出する第一の算出部と、

前記記憶部に記憶された各単語の位置情報に基づいて、接続した複数の単語間の近さを示す接続スコアを算出する第二の算出部と、

前記類似度および前記接続スコアに基づいて、前記音声信号に対応する文字列を決定する決定部と

を有し、

前記第二の算出部は、2つの単語が接続された文字列の先頭側の第一の単語と末尾側の第二の単語において、

前記第一の単語が前記第二の単語よりも文章中で先頭側に位置し、かつ、前記第一の単語と前記第二の単語とが文章中で隣接している場合に、第一の値を前記接続スコアとして算出し、

前記第一の単語が前記第二の単語よりも文章中で末尾側に位置している場合に、前記第一の値よりも低い第二の値を前記接続スコアとして算出し、

前記第一の単語が前記第二の単語よりも文章中で先頭側に位置し、かつ、前記第一の単語と前記第二の単語とが文章中で隣接していない場合に、前記第一の値と前記第二の値の間の値であって、前記第一の単語と前記第二の単語とが文章中で近い位置にあるほど高い

値となる第三の値を接続スコアとして算出することを特徴とする認識装置。

【請求項 2】

前記記憶部は、さらに、複数の単語が結合された単語列と該単語列の文章内の位置を示す位置情報とを記憶し、

前記第一の算出部は、さらに、前記単語列を単語として、前記類似度を算出し、

前記第二の算出部は、さらに、前記単語列を単語として、前記接続スコアを算出することを特徴とする請求項 1 に記載の認識装置。

【請求項 3】

前記記憶部は、さらに、前記文章に含まれる文に対応させて該文を識別するための識別情報を記憶し、

前記第二の算出部は、前記記憶部に記憶された各単語の位置情報および識別情報に基づいて、文ごとに、接続した複数の単語間の近さを示す接続スコアを算出する

ことを特徴とする請求項 1 または 2 に記載の認識装置。

【請求項 4】

前記記憶部は、さらに、第一の単語に第二の単語が接続される可能性を示す情報を記憶し、

前記第二の算出部は、さらに、前記記憶部に記憶された第一の単語に第二の単語が接続される可能性を示す情報に基づいて、前記接続スコアを補正する

ことを特徴とする請求項 1、2、3 のいずれか一つに記載の認識装置。

【請求項 5】

前記記憶部は、さらに、接続される複数の単語のそれぞれの品詞の組み合わせの適否にかかる情報を記憶し、

前記第二の算出部は、さらに、前記記憶部に記憶された接続される複数の単語のそれぞれの品詞の組み合わせの適否にかかる情報に基づいて、前記接続スコアを補正する

ことを特徴とする請求項 1 ~ 4 のいずれか一つに記載の認識装置。

【請求項 6】

コンピュータに、

入力された音声信号と、文章に含まれる単語と該単語の文章内の位置を示す位置情報とを記憶する記憶部に記憶された複数の単語を接続した文字列の読み情報とを比較して、類似度を算出し、

前記記憶部に記憶された各単語の位置情報に基づいて、接続した複数の単語間の近さを示す接続スコアを算出し、

前記類似度および前記接続スコアに基づいて、前記音声信号に対応する文字列を決定する

各処理を実行させ、

前記接続スコアを算出する処理は、2つの単語が接続された文字列の先頭側の第一の単語と末尾側の第二の単語において、

前記第一の単語が前記第二の単語よりも文章中で先頭側に位置し、かつ、前記第一の単語と前記第二の単語とが文章中で隣接している場合に、第一の値を前記接続スコアとして算出し、

前記第一の単語が前記第二の単語よりも文章中で末尾側に位置している場合に、前記第一の値よりも低い第二の値を前記接続スコアとして算出し、

前記第一の単語が前記第二の単語よりも文章中で先頭側に位置し、かつ、前記第一の単語と前記第二の単語とが文章中で隣接していない場合に、前記第一の値と前記第二の値の間の値であって、前記第一の単語と前記第二の単語とが文章中で近い位置にあるほど高い値となる第三の値を接続スコアとして算出することを特徴とする認識プログラム。

【請求項 7】

コンピュータが、

入力された音声信号と、文章に含まれる単語と該単語の文章内の位置を示す位置情報とを記憶する記憶部に記憶された複数の単語を接続した文字列の読み情報とを比較して、類

10

20

30

40

50

似度を算出し、

前記記憶部に記憶された各単語の位置情報に基づいて、接続した複数の単語間の近さを示す接続スコアを算出し、

前記類似度および前記接続スコアに基づいて、前記音声信号に対応する文字列を決定する

処理を実行し、

前記接続スコアを算出する処理は、2つの単語が接続された文字列の先頭側の第一の単語と末尾側の第二の単語において、

前記第一の単語が前記第二の単語よりも文章中で先頭側に位置し、かつ、前記第一の単語と前記第二の単語とが文章中で隣接している場合に、第一の値を前記接続スコアとして算出し、

前記第一の単語が前記第二の単語よりも文章中で末尾側に位置している場合に、前記第一の値よりも低い第二の値を前記接続スコアとして算出し、

前記第一の単語が前記第二の単語よりも文章中で先頭側に位置し、かつ、前記第一の単語と前記第二の単語とが文章中で隣接していない場合に、前記第一の値と前記第二の値の間の値であって、前記第一の単語と前記第二の単語とが文章中で近い位置にあるほど高い値となる第三の値を接続スコアとして算出することを特徴とする認識方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、認識装置、認識プログラム、認識方法、生成装置、生成プログラムおよび生成方法に関する。

【背景技術】

【0002】

ファイルやウェブページに含まれるテキストを用いて、入力された音声認識する装置がある。かかる装置の一例としては、入力された音声を示す音声信号とテキストに含まれる単語や単語が接続された文字列との類似度を算出し、算出した類似度が閾値を超える場合の単語または文字列を音声信号に対応する単語または文字列として決定する装置が挙げられる。

【0003】

また、テキストを用いて、入力された音声認識する装置の他の例としては、テキストに含まれる単語の全ての接続パターンを生成し、生成した接続パターンを、音声を認識する際に用いられる辞書に登録することで、辞書を生成する装置が挙げられる。辞書を生成する装置は、辞書に登録された接続パターンと、入力された音声を示す音声信号とを比較して、音声認識を行う。なお、辞書を生成する装置は、テキストに含まれる単語の個数が  $n$  個である場合には、1 から  $n$  までの和の個数分の接続パターンを生成する。

【0004】

さらに、テキストを用いて、入力された音声認識する装置の他の例としては、 $N$  グラムにより言語モデルを再学習する装置が挙げられる。この言語モデルを再学習する装置は、コーパスから学習した言語モデルに対して、テキストに存在する、単語が接続された単語列について、接続された単語の確率を高くする。この際、言語モデルを再学習する装置は、テキストに存在する単語数の  $N$  乗の個数のパターンを生成し、生成したパターンを用いて、言語モデルに対して、接続された単語の確率を高くする。

【先行技術文献】

【特許文献】

【0005】

【特許文献1】特開2002-41081号公報

【特許文献2】特開2002-342323号公報

【発明の概要】

【発明が解決しようとする課題】

10

20

30

40

50

## 【 0 0 0 6 】

しかしながら、従来の装置では、精度良く音声の認識ができないという問題がある。具体例を挙げて説明すると、上述した辞書を生成する装置では、接続パターンとして辞書に登録されていない音声が入力された場合には、入力された音声の認識結果の精度が低くなる。これは、辞書に登録されている接続パターンには、テキストに含まれる隣接する単語の組み合わせのパターンが含まれるが、テキストに含まれる単語であっても、隣接しない単語の組み合わせのパターンが含まれないからである。

## 【 0 0 0 7 】

なお、上述した言語を再学習する装置では、テキストに存在する単語数のN乗の個数のパターンを生成するため、生成するパターンの情報量が多いという問題がある。

10

## 【 0 0 0 8 】

開示の技術は、上記に鑑みてなされたものであって、精度良く音声の認識を行うことができる認識装置、認識プログラム、認識方法、生成装置、生成プログラムおよび生成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 0 9 】

本願の開示する認識装置は、一つの態様において、記憶部と、第一の算出部と、第二の算出部と、決定部とを有する。記憶部は、文章に含まれる単語と単語の文章内の位置を示す位置情報とを記憶する。第一の算出部は、入力された音声信号と、記憶部に記憶された複数の単語を接続した文字列の読み情報とを比較して、類似度を算出する。第二の算出部は、記憶部に記憶された各単語の位置情報に基づいて、接続した複数の単語間の近さを示す接続スコアを算出する。決定部は、類似度および接続スコアに基づいて、音声信号に対応する文字列を決定する。

20

## 【発明の効果】

## 【 0 0 1 0 】

本願の開示する認識装置の一つの態様によれば、精度良く音声の認識を行うことができる。

## 【図面の簡単な説明】

## 【 0 0 1 1 】

【図1】図1は、実施例1に係る認識装置および生成装置を有するシステムの構成の一例を示す図である。

30

【図2】図2は、実施例1に係る生成装置の機能構成の一例を示す図である。

【図3】図3は、辞書の一例を示す図である。

【図4】図4は、実施例1に係る認識装置の機能構成の一例を示す図である。

【図5】図5は、実施例1に係る生成処理の手順を示すフローチャートである。

【図6】図6は、実施例1に係る認識処理の手順を示すフローチャートである。

【図7】図7は、実施例2に係る生成装置の機能構成の一例を示す図である。

【図8】図8は、辞書の一例を示す図である。

【図9】図9は、実施例2に係る認識装置の構成の一例を示す図である。

【図10】図10は、実施例2に係る生成処理の手順を示すフローチャートである。

40

【図11】図11は、実施例2に係る認識処理の手順を示すフローチャートである。

【図12】図12は、実施例3に係る認識装置の構成の一例を示す図である。

【図13】図13は、言語モデルの一例を示す図である。

【図14】図14は、単語辞書の一例を示す図である。

【図15】図15は、実施例3に係る認識処理の手順を示すフローチャートである。

【図16】図16は、実施例4に係る生成装置の構成の一例を示す図である。

【図17】図17は、品詞番号表の一例を示す図である。

【図18】図18は、辞書の一例を示す図である。

【図19】図19は、実施例4に係る認識装置の構成の一例を示す図である。

【図20】図20は、確率表の一例を示す図である。

50

【図 2 1】図 2 1 は、実施例 4 に係る生成処理の手順を示すフローチャートである。

【図 2 2】図 2 2 は、実施例 4 に係る認識処理の手順を示すフローチャートである。

【図 2 3】図 2 3 は、生成プログラムを実行するコンピュータを示す図である。

【図 2 4】図 2 4 は、認識プログラムを実行するコンピュータを示す図である。

【発明を実施するための形態】

【0012】

以下に、本願の開示する認識装置、認識プログラム、認識方法、生成装置、生成プログラムおよび生成方法の実施例を図面に基づいて詳細に説明する。なお、この実施例は開示の技術を限定するものではない。そして、各実施例は、処理内容を矛盾させない範囲で適宜組み合わせることが可能である。

10

【実施例 1】

【0013】

実施例 1 に係る認識装置および生成装置について説明する。図 1 は、実施例 1 に係る認識装置および生成装置を有するシステムの構成の一例を示す図である。本実施例に係るシステム 1 は、生成装置 10 と、認識装置 20 とを有する。生成装置 10 と、認識装置 20 とは、ネットワーク 30 を介して接続されている。生成装置 10 と、認識装置 20 とは、通信可能である。なお、図 1 の例では、システム 1 に、生成装置 10 および認識装置 20 が各々 1 台の場合が示されているが、生成装置 10 および認識装置 20 の台数はこれに限られない。例えば、生成装置 10 および認識装置 20 のそれぞれの台数は、複数台であってもよい。

20

【0014】

生成装置 10 は、文章に含まれる単語と、単語の文章内の位置を示す位置情報とが対応付けられて登録された辞書を生成する。例えば、生成装置 10 は、まず、入力されたテキストファイルが示す文章を単語ごとに分割する。かかる文章は、テキストとも称される。そして、生成装置 10 は、分割された単語の文章内の位置を示す位置情報を生成する。続いて、生成装置 10 は、単語と、単語の文章内の位置を示す位置情報とが対応付けられて登録された辞書を生成する。そして、生成装置 10 は、生成した辞書を認識装置 20 へ送信する。

【0015】

認識装置 20 は、生成装置 10 から送信された辞書を受信する。そして、認識装置 20 は、入力された音声信号と、辞書に登録された複数の単語を接続した文字列の読み情報とを比較して、類似度を算出する。続いて、認識装置 20 は、辞書に登録された各単語の位置情報に基づいて、接続した複数の単語間の近さを示す接続スコアを算出する。そして、認識装置 20 は、類似度および接続スコアに基づいて、入力された音声信号に対応する文字列を決定する。続いて、認識装置 20 は、決定した文字列を出力する。

30

【0016】

このように、認識装置 20 は、類似度を算出する際に用いられる文字列の複数の単語間の近さを示す接続スコアを算出する。例えば、認識装置 20 は、類似度を算出する際に用いられる文字列の複数の単語間が近いほど、値が大きくなるような接続スコアを算出する。そして、認識装置 20 は、類似度だけでなく、接続スコアをも加味して、入力された音声信号に対応する文字列を決定する。例えば、認識装置 20 は、類似度と接続スコアとを加算した値のうち、閾値を超えた値に対応する文字列を、入力された音声信号に対応する文字列として決定する。そのため、認識装置 20 は、文章に含まれる複数の単語であって、文章内で隣接しない複数の単語を接続した文字列を発話者が発話して、音声信号として入力された場合であっても、発話者が発話した音声認識することができる。したがって、認識装置 20 によれば、精度良く音声の認識を行うことができる。

40

【0017】

また、生成装置 10 は、上述したように精度良く音声の認識を行うことが可能な辞書を生成することができる。

【0018】

50

[ 生成装置の機能構成の一例 ]

図 2 は、実施例 1 に係る生成装置の機能構成の一例を示す図である。図 2 の例では、生成装置 1 0 は、入力部 1 1 と、出力部 1 2 と、通信部 1 3 と、記憶部 1 4 と、制御部 1 5 とを有する。

【 0 0 1 9 】

入力部 1 1 は、各種情報を制御部 1 5 に入力する。例えば、入力部 1 1 は、ユーザから、他の装置から所定の話題に関する文章を含むテキストファイルを取得する指示を受け付け、受け付けた指示を制御部 1 5 に入力する。入力部 1 1 のデバイスの一例としては、マウスやキーボードなどが挙げられる。

【 0 0 2 0 】

出力部 1 2 は、各種情報を出力する。例えば、出力部 1 2 は、制御部 1 5 から画像が入力された場合には、画像を表示出力する。出力部 1 2 のデバイスの一例としては、C R T ( Cathode Ray Tube ) や液晶ディスプレイなどが挙げられる。

【 0 0 2 1 】

通信部 1 3 は、認識装置 2 0 および外部の装置と通信を行うためのネットワークインターフェースである。例えば、通信部 1 3 は、外部の装置へ所定の話題に関する文章を含むテキストファイルを生成装置 1 0 へ送信する旨の指示を制御部 1 5 から受信すると、受信した指示をネットワーク 3 0 を介して外部の装置へ送信する。また、通信部 1 3 は、外部の装置からテキストファイルを受信すると、受信したテキストファイルを、制御部 1 5 へ送信する。

【 0 0 2 2 】

記憶部 1 4 は、テキストファイル 1 4 a と、辞書 1 4 b とを記憶する。

【 0 0 2 3 】

テキストファイル 1 4 a は、後述する取得部 1 5 a により記憶部 1 4 に格納される。テキストファイルは、文章を示す。テキストファイル 1 4 a は、発話者が発話する音声を認識する際に用いられる後述の辞書 1 4 b を生成する際に用いられる。したがって、発話者が発話する内容に合ったテキストファイル 1 4 a が記憶部 1 4 に記憶されることが好ましい。例えば、発話者が天気に関する内容を発話する場合には、テキストファイルに含まれる文章の内容が天気に関する内容であることが好ましい。

【 0 0 2 4 】

辞書 1 4 b は、後述する生成部 1 5 c により生成され、後述する格納部 1 5 d により記憶部 1 4 に格納される。

【 0 0 2 5 】

図 3 は、辞書の一例を示す図である。図 3 の例では、辞書 1 4 b は、「表記」の項目、「読み情報」の項目、「位置情報」の項目を有する。「表記」の項目には、入力されたテキストファイルが示す文章に含まれる単語、および、複数の単語を結合した単語列が登録される。また、「読み情報」の項目には、「表記」の項目に登録された単語または単語列の読み方、すなわち、単語または単語列のひらがなの表記が登録される。「位置情報」の項目には、「表記」の項目に登録された単語または単語列の文章内での位置を示す位置情報が登録される。これらの各項目の内容は、生成部 1 5 c によって生成される。

【 0 0 2 6 】

図 3 の例は、一番目のレコードに、テキストファイルが示す文章「衆議院で改正法が審議入り。山川総理が法案を通したいと言及」に含まれる単語「衆議院」が「表記」の項目に登録された場合を示す。また、図 3 の例は、一番目のレコードに、衆議院の読み方「しゅうぎいん」が「読み情報」の項目に登録された場合を示す。また、図 3 の例は、一番目のレコードに、単語「衆議院」の文章内での位置を示す位置情報「0」が、「位置情報」の項目に登録された場合を示す。

【 0 0 2 7 】

また、図 3 の例は、二番目のレコードに、テキストファイルが示す文章「衆議院で改正法が審議入り。山川総理が法案を通したいと言及」に含まれる単語「衆議院」と単語「で

10

20

30

40

50

」とを結合した単語列「衆議院で」が「表記」の項目に登録された場合を示す。また、図3の例は、二番目のレコードに、「衆議院で」の読み方「しゅうぎいんで」が「読み情報」の項目に登録された場合を示す。また、図3の例は、二番目のレコードに、単語列「衆議院で」の文章内での位置を示す位置情報「1」が、「位置情報」の項目に登録された場合を示す。

【0028】

記憶部14は、例えば、フラッシュメモリなどの半導体メモリ素子、または、ハードディスク、光ディスクなどの記憶装置である。なお、記憶部14は、上記の種類の記憶装置に限定されるものではなく、RAM(Random Access Memory)、ROM(Read Only Memory)であってもよい。

10

【0029】

制御部15は、各種の処理手順を規定したプログラムや制御データを格納するための内部メモリを有し、これらによって種々の処理を実行する。制御部15は、取得部15aと、分割部15bと、生成部15cと、格納部15dとを有する。

【0030】

取得部15aは、他の装置から所定の話題に関する文章を含むテキストファイル14aを取得する指示を入力部11から受け付けると、図示しない外部の装置へかかるテキストファイル14aを生成装置10へ送信する旨の指示を通信部13へ送信する。これにより、外部の装置から、所定の話題に関する文章を含むテキストファイル14aが生成装置10へ送信される。そのため、取得部15aは、外部の装置から、所定の話題に関する文章を含むテキストファイル14aを取得することができる。そして、取得部15aは、取得したテキストファイル14aを記憶部14に格納する。

20

【0031】

分割部15bは、テキストファイル14aが示す文章を単語ごとに分割する。例えば、分割部15bは、形態素解析を行うためのソフトウェア「ChaSen(茶筌)」、「MeCab(和布蕪)」などを用いて、テキストファイル14aが示す文章に対して形態素解析を行い、形態素ごとに分割する。そして、分割部15bは、分割された各形態素のそれぞれを単語とみなすことで、文章を単語ごとに分割することができる。以下では、形態素を単語とみなす場合を例に挙げて説明するが、複数の形態素、例えば、語幹、活用語尾、助動詞などの3つの形態素を1つの単語とみなしてもよい。例えば、3つの形態素が、語幹「通」、活用語尾「し」、助動詞「たい」である場合には、「通したい」を1つの単語とみなすことができる。また、分割部15bは、上述したソフトウェア「ChaSen(茶筌)」、「MeCab(和布蕪)」などを用いることで、単語ごとの読み情報を得ることができる。すなわち、分割部15bは、文章を単語ごとに分割し、単語の読み情報を取得する。分割部15bにより得られた単語の総数がM個である場合には、文章の先頭の単語は、1番目の単語であり、文章の最後の単語は、M番目の単語となる。

30

【0032】

例えば、テキストファイル14aが示す文章が、「衆議院で改正法が審議入り。山川総理が法案を通したいと言及」である場合には、分割部15bは、「衆議院で改正法が審議入り。」の部分について、次のように、単語ごとに分割し、単語の読み情報を取得する。すなわち、分割部15bは、「衆議院+しゅうぎいん/で+で/改正法+かいせいほう/が+が/審議入り+しんぎいり。」というように、文章を単語ごとに分割し、単語の読み情報を取得する。ここで、「/」は、単語の区切り位置を示す記号であり、「+」は、単語と単語に対応する読み情報とを接続する記号である。また、分割部15bは、「山川総理が法案を通したいと言及」の部分について、次のように、単語ごとに分割し、単語の読み情報を取得する。すなわち、分割部15bは、「山川+やまかわ/総理+そうり/が+が/法案+法案/を+を/通+とお/し+し/たい+たい」というように、文章を単語ごとに分割し、単語の読み情報を取得する。

40

【0033】

生成部15cは、分割部15bにより分割された単語の文章内の位置を示す位置情報を

50

生成する。これに加えて、生成部 15c は、単語のそれぞれについて、その単語に続く 1 個から (N - 1) 個の単語までのそれぞれの単語列を生成する。そして、生成部 15c は、単語列の文章内の位置を示す位置情報を生成する。

【0034】

具体例を挙げて説明する。例えば、生成部 15c は、まず、位置情報 n、変数 i、変数 j の各値を初期化する。すなわち、生成部 15c は、位置情報 n の値を 0、変数 i の値を 1、変数 j の値を 0 に設定する。そして、生成部 15c は、i 番目の単語から (i + j) 番目の単語までの単語列を生成する。なお、j = 0 の場合は、i 番目の単語を単語列とみなして下記の処理を行う。また、(i + k ; k j) 番目の単語が存在しない場合には、生成部 15c は、(i + k) 番目の単語を空欄として扱って、i 番目の単語から (i + j) 番目の単語までの単語列を生成する。

10

【0035】

そして、生成部 15c は、格納部 15d により新たに生成された単語列と、読み情報と、位置情報 n の値とが辞書 14b に登録された場合に、変数 j の値を 1 つインクリメントし、変数 n の値を 1 つインクリメントする。そして、生成部 15c は、変数 j の値が、単語列に結合する単語の数の最大値である最大結合数 N よりも小さいか否かを判定する。小さい場合には、生成部 15c は、再び、i 番目の単語から (i + j) 番目の単語までの単語列を生成し、上記で説明した処理と同様の処理を行う。

【0036】

一方、変数 j の値が、最大結合数 N 以上である場合には、生成部 15c は、変数 i の値を 1 つインクリメントし、変数 j の値を 0 に設定する。続いて、生成部 15c は、変数 i の値が、分割部 15b により得られた単語の総数である総単語数 M 以下であるか否かを判定する。変数 i の値が総単語数 M 以下である場合には、生成部 15c は、再び、i 番目の単語から (i + j) 番目の単語までの単語列を生成し、上記で説明した処理と同様の処理を行う。

20

【0037】

格納部 15d は、単語および単語列と、対応する読み情報と、対応する位置情報 n とを対応付けて記憶部 14 に格納する。具体例を挙げて説明する。例えば、格納部 15d は、生成部 15c により、i 番目の単語から (i + j) 番目の単語までの単語列が生成された場合に、次のような処理を行う。すなわち、格納部 15d は、新たに生成された単語列と、新たに生成された単語列に対応する読み情報と、位置情報 n の値とを、それぞれ、辞書 14b の「表記」、「読み情報」、「位置情報」の各項目に登録する。なお、格納部 15d は、新たに生成された単語列中の単語に、句点「。」が含まれる場合や、空欄がある場合には、単語列と読み情報と位置情報 n の値とを辞書 14b へ登録する処理を省略できる。これは、単語列中の単語に、句点「。」が含まれる場合や、空欄がある場合には、その単語列は、音声を認識する際の音声との比較対象として不適切であるという理由による。したがって、格納部 15d によれば、辞書 14b を生成する速度が速くなる。また、格納部 15d によれば、より少ない情報量の辞書 14b を生成することができる。

30

【0038】

以上説明したように、生成部 15c および格納部 15d などによって、辞書 14b が生成される。例えば、テキストファイル 14a が示す文章が「衆議院で改正法が審議入り。山川総理が法案を通したいと言及」である場合には、図 3 の例に示すような辞書 14b が生成される。図 3 の例では、最大結合数 N の値は 3 である。また、図 3 の例では、句点や空欄を示す単語が含まれる単語列も辞書 14b に登録されているように示されている。しかしながら、これは説明の便宜上のためである。図 3 の例では、句点や空欄を示す単語が含まれる単語列にかかるレコードについては、「表記」の項目に登録された単語列の先頭部分に「(」の記号が付加されており、「(」の記号が付加された単語列にかかるレコードは、辞書 14b には登録されない。すなわち、格納部 15c は、総単語数 M × 最大結合数 N - n (n = 1, 2, …, N - 1) の個数程度のレコードを辞書 14b に登録するため、少ない情報量の辞書 14b を生成することができる。

40

50



## 【 0 0 3 9 】

なお、最大結合数  $N$  の値は、音声を認識する認識装置 2 0 にとって、認識精度が良好となる読みの長さに合わせるのが好ましい。例えば、認識精度が良好となる読みの長さが 8 音節であり、単語の平均音節長が 3 である場合には、 $8 / 3 = 2.6 \dots$ となるので、最大結合数  $N$  の値として 2 または 3 を用いるのが好ましい。以下の説明では、最大結合数  $N$  の値が 3 である場合について説明するが、最大結合数  $N$  の値はこれに限られない。

## 【 0 0 4 0 】

また、格納部 1 5 d は、認識装置 2 0 から辞書 1 4 b の送信要求を受信すると、記憶部 1 4 から辞書 1 4 b を読み込み、辞書 1 4 b を通信部 1 3 に送信する。辞書 1 4 b を受信した通信部 1 3 は、辞書 1 4 b をインターネット 3 0 を介して、認識装置 2 0 へ送信する。

10

## 【 0 0 4 1 】

制御部 1 5 は、例えば、A S I C (Application Specific Integrated Circuit) や F P G A (Field Programmable Gate Array) などの集積回路、または、C P U (Central Processing Unit) や M P U (Micro Processing Unit) などの電子回路である。

## 【 0 0 4 2 】

## [ 認識装置の機能構成の一例 ]

図 4 は、実施例 1 に係る認識装置の機能構成の一例を示す図である。図 4 の例では、認識装置 2 0 は、音声入力部 2 1 と、出力部 2 2 と、通信部 2 3 と、記憶部 2 4 と、制御部 2 5 とを有する。

20

## 【 0 0 4 3 】

音声入力部 2 1 は、ユーザなどの発話者によって発話された音声を A / D 変換して、音声信号として制御部 2 5 に入力する。音声入力部 2 1 のデバイスの一例としては、マイクロフォンなどが挙げられる。

## 【 0 0 4 4 】

出力部 2 2 は、各種情報を出力する。例えば、出力部 2 2 は、制御部 2 5 から音声認識結果の文字列を示す画像が入力された場合には、かかる画像を表示する。出力部 2 2 のデバイスの一例としては、C R T (Cathode Ray Tube) や液晶ディスプレイなどが挙げられる。

30

## 【 0 0 4 5 】

通信部 2 3 は、生成装置 1 0 と通信を行うためのネットワークインターフェースである。例えば、通信部 2 3 は、制御部 2 5 から、生成装置 1 0 に対して認識装置 2 0 に辞書 1 4 b を送信する旨の送信要求を受信すると、かかる送信要求をネットワーク 3 0 を介して生成装置 1 0 へ送信する。また、通信部 2 3 は、生成装置 1 0 から辞書 1 4 b を受信すると、受信した辞書 1 4 b を制御部 2 5 へ送信する。

## 【 0 0 4 6 】

記憶部 2 4 は、辞書 1 4 b と、文字列 2 4 a を記憶する。

## 【 0 0 4 7 】

辞書 1 4 b は、生成装置 1 0 から取得され、後述する第一の算出部 2 6 a により記憶部 2 4 に格納される。また、文字列 2 4 a は、音声認識の結果、入力された音声信号に対応する文字列として後述する決定部 2 6 c により決定され、後述する出力部 2 7 により記憶部 2 4 に格納される。

40

## 【 0 0 4 8 】

また、記憶部 2 4 は、辞書 1 4 b の「読み情報」の項目に登録された読み情報に対応する、隠れマルコフモデルなどの音響モデルを記憶する。音声を認識して類似度 (確率値) などを算出する際には、この音響モデルと、音声信号との比較が行われる。

## 【 0 0 4 9 】

記憶部 2 4 は、例えば、フラッシュメモリなどの半導体メモリ素子、または、ハードディスク、光ディスクなどの記憶装置である。なお、記憶部 2 4 は、上記の種類の記憶装置に限定されるものではなく、R A M (Random Access Memory)、R O M (Read Only M

50

emory)であってもよい。

【 0 0 5 0 】

制御部 2 5 は、各種の処理手順を規定したプログラムや制御データを格納するための内部メモリを有し、これらによって種々の処理を実行する。制御部 2 5 は、照合部 2 6 と、出力部 2 7 とを有する。

【 0 0 5 1 】

照合部 2 6 は、入力された音声信号と、辞書 1 4 b に登録された読み情報の複数の組み合わせに対応する音響モデルの複数の組み合わせのそれぞれとを比較し、類似度を音響モデルの組み合わせごとに算出する。なお、「読み情報の組み合わせ」とは、読み情報が 1 つである場合には、その読み情報を指し、読み情報が複数である場合には、読み情報を順番に並べて、各読み情報を接続した場合の文字列を指す。並べる順番を入れ替えることで様々な読み情報の組み合わせが可能である。「音響モデルの組み合わせ」についても同様である。また、照合部 2 6 は、類似度を算出する際に組み合わせられた読み情報に対応する単語または文字列の位置情報に基づいて、次のような処理を行う。すなわち、照合部 2 6 は、組み合わせられた複数の読み情報のそれぞれに対応する単語または文字列について、単語間または文字列間もしくは単語と文字列間の文章中の位置の近さを示す接続スコアを読み情報の組み合わせごとに算出する。そして、照合部 2 6 は、複数の算出された和のうち、閾値を超え、かつ、最も値が大きい和に対応する文字列を、音声信号に対応する文字列として決定する。照合部 2 6 によれば、文章に含まれる複数の単語であって、文章内で隣接しない複数の単語を接続した文字列を発話者が発話して、音声信号として入力された場合であっても、発話者が発話した音声認識することができる。したがって、照合部 2 6 によれば、精度良く音声の認識を行うことができる。

10

20

【 0 0 5 2 】

照合部 2 6 は、第一の算出部 2 6 a と、第二の算出部 2 6 b と、決定部 2 6 c とを有する。

【 0 0 5 3 】

第一の算出部 2 6 a は、入力された音声信号と、辞書 1 4 b に登録された複数の単語および単語列を接続した文字列の読み情報に対応する音響モデルとを比較して、類似度を算出する。

【 0 0 5 4 】

具体例を挙げて説明する。例えば、第一の算出部 2 6 a は、認識装置 2 0 に備えられたキーボードやマウスなどの操作受付装置（図示せず）を介して、ユーザから認識処理を実行する旨の指示を受信した場合に、次のような処理を行う。すなわち、第一の算出部 2 6 a は、生成装置 1 0 に対して認識装置 2 0 に辞書 1 4 b を送信する旨の送信要求を通信部 2 3 へ送信する。続いて、第一の算出部 2 6 a は、生成装置 1 0 から辞書 1 4 b を受信したか否かを判定する。辞書 1 4 b を受信した場合には、第一の算出部 2 6 a は、受信した辞書 1 4 b を記憶部 2 4 へ格納する。

30

【 0 0 5 5 】

その後、第一の算出部 2 6 a は、音声入力部 2 1 を介して音声信号を受信したか否かを判定する。音声信号を受信した場合には、第一の算出部 2 6 a は、音声信号をフレームごとに分割する。そして、第一の算出部 2 6 a は、全てのフレームに対して、下記で説明する処理を行ったか否かを判定する。未処理のフレームがある場合には、第一の算出部 2 6 a は、未処理のフレームに対して、辞書 1 4 b に登録された読み情報の組み合わせを複数生成し、生成された複数の組み合わせのそれぞれに対応する音響モデルの組み合わせを生成する。

40

【 0 0 5 6 】

そして、第一の算出部 2 6 a は、フレームと、音響モデルの複数の組み合わせのそれぞれとを比較し、音響モデルの組み合わせごとに類似度を算出する。

【 0 0 5 7 】

第二の算出部 2 6 b は、辞書 1 4 b を参照し、類似度を算出する際に組み合わせられた読

50

み情報に対応する単語または文字列の位置情報に基づいて、次のような処理を行う。すなわち、第二の算出部 26 b は、組み合わされた複数の読み情報のそれぞれに対応する単語または文字列について、単語間または文字列間もしくは単語と文字列間の文章中の位置の近さを示す接続スコアを読み情報の組み合わせごとに算出する。

【 0 0 5 8 】

第二の算出部 26 b が行う接続スコアの算出方法の一例について説明する。ここでは、先の図 3 の例に示す辞書 14 b が記憶部 24 に記憶された場合を想定する。第二の算出部 26 b は、第一の算出部 26 a で類似度を算出する際に、音声信号との比較に用いられた音響モデルの組み合わせを全て特定する。そして、第二の算出部 26 b は、辞書 14 b を参照し、特定した音響モデルのそれぞれについて、対応する読み情報の組み合わせを特定する。続いて、第二の算出部 26 b は、辞書 14 b を参照し、組み合わされた各読み情報に対応する位置情報を、特定した全ての読み情報の組み合わせごとを取得する。

10

【 0 0 5 9 】

その後、第二の算出部 26 b は、組み合わされた各読み情報に対応する位置情報が示す値が、先頭から昇順であるか否かを判定する。例えば、組み合わされた各読み情報が、「しゅうぎいん」、「で」、「かいせいほう」である場合には、図 3 の例に示す辞書 14 b に登録された位置情報がそれぞれ「0」、「3」、「6」であるため、第二の算出部 26 b は、昇順であると判定する。また、組み合わされた各読み情報が、「しゅうぎいで」、「で」、「かいせいほう」である場合には、図 3 の例に示す辞書 14 b に登録された位置情報がそれぞれ「1」、「3」、「6」であるため、第二の算出部 26 b は、昇順であると判定する。また、組み合わされた各読み情報が、「しゅうぎいん」、「かいせいほう」、「しんぎいり」である場合には、図 3 の例に示す辞書 14 b に登録された位置情報がそれぞれ「0」、「6」、「13」であるため、第二の算出部 26 b は、昇順であると判定する。

20

【 0 0 6 0 】

続いて、第二の算出部 26 b は、位置情報が示す値が昇順であると判定された組み合わせについて、次のような処理を行う。すなわち、組み合わせされた読み情報のうち、結合された 2 つの読み情報、すなわち隣接する読み情報について、先頭に近いほうの読み情報に対応する位置情報が示す値を変数  $x$  に代入し、先頭から遠い方の読み情報に対応する位置情報が示す値を変数  $y$  に代入する。そして、第二の算出部 26 b は、下記の式 ( 1 )、式 ( 2 ) に基づいて、 $X$ 、 $Y$  を算出する。

30

$$X = (x / N) + (x \% N) + 1 \quad \text{式 ( 1 )}$$

$$Y = (y / N) \quad \text{式 ( 2 )}$$

【 0 0 6 1 】

ただし、 $N$  は、上述した最大結合数である。また、「/」は、小数点以下を切り捨てて商を整数とするための除算演算を意味する。また、「%」は、剰余演算を意味する。

【 0 0 6 2 】

例えば、組み合わされた各読み情報が、「しゅうぎいん」、「で」、「かいせいほう」である場合には、位置情報がそれぞれ「0」、「3」、「6」である。このため、「しゅうぎいん」および「で」の隣接する読み情報について、変数  $x$  に「0」、変数  $y$  に「3」が代入されて、 $X$ 、 $Y$  が算出される。さらに、「で」および「かいせいほう」の隣接する読み情報について、変数  $x$  に「3」、変数  $y$  に「6」が代入され、 $X$ 、 $Y$  が算出される。また、組み合わされた各読み情報が、「しゅうぎいで」、「で」、「かいせいほう」である場合には、位置情報がそれぞれ「1」、「3」、「6」である。このため、「しゅうぎいで」および「で」の隣接する読み情報について、変数  $x$  に「1」、変数  $y$  に「3」が代入されて、 $X$ 、 $Y$  が算出される。さらに、「で」および「かいせいほう」の隣接する読み情報について、変数  $x$  に「3」、変数  $y$  に「6」が代入され、 $X$ 、 $Y$  が算出される。

40

【 0 0 6 3 】

50

第二の算出部 26 b は、このようにして X、Y を算出する処理を、位置情報が示す値が昇順であると判定された組み合わせに含まれる、全ての隣接する読み情報について行う。

【0064】

続いて、第二の算出部 26 b は、 $X > Y$  の場合、隣接する読み情報同士で同じ読み情報（読み情報に対応する単語または文字列）が重複して用いられており、これらの隣接する読み情報の組み合わせは適切でないと考えられるため、接続スコアを最低の点数とする。例えば、接続スコアが 0 ~ 30 点である場合には、第二の算出部 26 b は、0 点とする。位置情報が「1」、「3」となる「しゅうぎいんで」および「で」が隣接する読み情報として組み合わせられた場合、 $X = 2$ 、 $Y = 1$  となり、「で」の読み情報に対応する単語「で」が重複しているため、第二の算出部 26 b は、接続スコアを 0 点とする。

10

【0065】

また、第二の算出部 26 b は、 $X = Y$  の場合、入力されたテキストファイル 14 a が示す文章内で隣接している単語または文字列であるため、これらの読み情報の組み合わせは適切であると考えられるため、接続スコアを最高の点数とする。例えば、接続スコアが 0 ~ 30 点である場合には、第二の算出部 26 b は、30 点とする。位置情報が「0」、「3」となる「しゅうぎいん」および「で」が隣接する読み情報として組み合わせられた場合、 $X = Y = 1$  となり、第二の算出部 26 b は、接続スコアを 30 点とする。また、位置情報が「3」、「6」となる「で」および「かいせいほう」が隣接する読み情報として組み合わせられた場合、 $X = Y = 2$  となり、第二の算出部 26 b は、接続スコアを 30 点とする。

20

【0066】

また、 $X < Y$  の場合、 $Y - X$  の値が小さいほど、2つの隣接する読み情報に対応する単語または文字列が、テキストファイル 14 a が示す文章内で近い位置に存在すると考えられる。そこで、第二の算出部 26 b は、 $Y - X$  の値が小さいほど、接続スコアが大きくなるように、接続スコアの点数をつける。例えば、接続スコアが 0 ~ 30 点である場合には、第二の算出部 26 b は、 $MAX(0, 30 - (Y - X))$  とする。ただし、 $MAX(A, B)$  は、A と B のうち、大きい値を選択する関数である。

【0067】

第二の算出部 26 b は、X、Y が算出された隣接する読み情報の全てについて、同様に、接続スコアを算出する。

30

【0068】

なお、第二の算出部 26 b は、位置情報が昇順である場合についての組み合わせについてのみ接続スコアを算出する場合について例示したが、開示の装置はこれに限られない。例えば、第二の算出部 26 b は、位置情報が降順である場合には、次のような処理を行ってもよい。すなわち、組み合わせされた読み情報のうち、隣接する読み情報について、先頭に近いほうの読み情報に対応する位置情報が示す値を変数  $y$  に代入し、先頭から遠い方の読み情報に対応する位置情報が示す値を変数  $x$  に代入するようにしてもよい。この場合、算出された X、Y に対して所定の減点、例えば、3 点の減点をするようにして、昇順の場合との接続スコアの点数に差を付けてもよい。このように、昇順だけでなく降順の場合も接続スコアを算出することで、類似度を算出する際に組み合わせられた読み情報に対応する単語または文字列が、テキストファイル 14 a が示す文章内で近い場合には、高い接続スコアの点数が算出される。たとえば、テキストファイル 14 a が示す文章が、「衆議院で改正法が審議入り。山川総理が法案を通したいと言及」である場合に、「山川総理が通したい改正法」といったテキストファイル 14 a が示す文章とは単語の語順が異なるような単語または文字列の組み合わせについても、接続スコアが単語同士の近さに応じて算出される。そのため、発話者が、「山川総理が通したい改正法」と発話した場合であっても、音声認識をすることができる。

40

【0069】

決定部 26 c は、類似度および接続スコアに基づいて、入力された音声信号に対応する文字列を決定する。例えば、決定部 26 c は、類似度を算出する際に用いられた音響モデ

50

ルの組み合わせごとに、すなわち、音響モデルの組み合わせに対応する読み情報の組み合わせごとに、対応する類似度と、対応する接続スコアとの和を算出する。そして、決定部 26c は、複数の算出された和のうち、閾値を超えた和があるか否かを判定する。閾値を超えた和がある場合には、閾値を超えた和のうち、最も値が大きい和に対応する文字列を、音声信号に対応する文字列として決定する。このようにして、決定部 26c は、各フレームに対して、文字列を決定する。

【0070】

出力部 27 は、各フレームに対して決定された文字列を音声の認識結果として画面に表示するように出力部 22 に送信する。また、出力部 27 は、各フレームに対して決定された文字列を記憶部 24 に格納する。なお、記憶部 24 に格納された文字列は、上述した、  
10

【0071】

制御部 25 は、例えば、ASIC (Application Specific Integrated Circuit) や FPGA (Field Programmable Gate Array) などの集積回路、または、CPU (Central Processing Unit) や MPU (Micro Processing Unit) などの電子回路である。

【0072】

[処理の流れ]

次に、本実施例に係る生成装置 10 の処理の流れを説明する。図 5 は、実施例 1 に係る生成処理の手順を示すフローチャートである。この生成処理は、入力部 11 から制御部 15 に生成処理を実行する指示が入力された場合に実行される。  
20

【0073】

図 5 に示すように、分割部 15b は、テキストファイル 14a が示す文章を単語ごとに分割する (S101)。生成部 15c は、位置情報 n の値を 0、変数 i の値を 1、変数 j の値を 0 に設定する (S102)。生成部 15c は、i 番目の単語から (i + j) 番目の単語までの単語列を生成する (S103)。

【0074】

格納部 15d は、新たに生成された単語列と、新たに生成された単語列に対応する読み情報と、位置情報 n の値とを、それぞれ、辞書 14b の「表記」、「読み情報」、「位置情報」の各項目に登録する (S104)。生成部 15c は、変数 j の値を 1 つインクリメントし、変数 n の値を 1 つインクリメントする (S105)。生成部 15c は、変数 j の  
30

値が、単語列に結合する単語の数の最大値である最大結合数 N よりも小さいか否かを判定する (S106)。小さい場合 (S106 肯定) には、S103 へ戻る。

【0075】

一方、変数 j の値が、最大結合数 N 以上である場合 (S106 否定) には、生成部 15c は、変数 i の値を 1 つインクリメントし、変数 j の値を 0 に設定する (S107)。生成部 15c は、変数 i の値が、分割部 15b により得られた単語の総数である総単語数 M 以下であるか否かを判定する (S108)。変数 i の値が総単語数 M 以下である場合 (S108 肯定) には、S103 へ戻る。一方、変数 i の値が総単語数 M より大きい場合 (S108 否定) には、処理を終了する。

【0076】

次に、本実施例に係る認識装置 20 の処理の流れを説明する。図 6 は、実施例 1 に係る認識処理の手順を示すフローチャートである。この認識処理は、操作受付部 (図示せず) から制御部 25 に認識処理を実行する指示が入力された場合に実行される。  
40

【0077】

図 6 に示すように、第一の算出部 26a は、生成装置 10 に対して認識装置 20 に辞書 14b を送信する旨の送信要求を通信部 23 へ送信する (S201)。続いて、第一の算出部 26a は、生成装置 10 から辞書 14b を受信したか否かを判定する (S202)。辞書 14b を受信していない場合 (S202 否定) には、第一の算出部 26a は、再び同様の判定を行う。一方、辞書 14b を受信した場合 (S202 肯定) には、第一の算出部 26a は、受信した辞書 14b を記憶部 24 へ格納する (S203)。  
50

## 【 0 0 7 8 】

第一の算出部 2 6 a は、音声入力部 2 1 を介して音声信号を受信したか否かを判定する ( S 2 0 4 )。音声信号を受信していない場合 ( S 2 0 4 否定 ) には、第一の算出部 2 6 a は、再び同様の判定を行う。一方、音声信号を受信した場合 ( S 2 0 4 肯定 ) には、第一の算出部 2 6 a は、音声信号をフレームごとに分割する ( S 2 0 5 )。第一の算出部 2 6 a は、全てのフレームに対して、下記の S 2 0 7 以降の処理を行ったか否かを判定する ( S 2 0 6 )。未処理のフレームが無い場合 ( S 2 0 6 肯定 ) には、処理を終了する。一方、未処理のフレームがある場合 ( S 2 0 6 否定 ) には、第一の算出部 2 6 a は、次のような処理を行う。すなわち、第一の算出部 2 6 a は、未処理のフレームに対して、辞書 1 4 b に登録された読み情報の組み合わせを複数生成し、生成された複数の組み合わせのそれぞれに対応する音響モデルの組み合わせを生成する ( S 2 0 7 )。

10

## 【 0 0 7 9 】

第一の算出部 2 6 a は、フレームと、音響モデルの複数の組み合わせのそれぞれとを比較し、音響モデルの組み合わせごとに類似度を算出する ( S 2 0 8 )。

## 【 0 0 8 0 】

第二の算出部 2 6 b は、辞書 1 4 b を参照し、類似度を算出する際に組み合わせられた読み情報に対応する単語または文字列の位置情報に基づいて、次のような処理を行う。すなわち、第二の算出部 2 6 b は、組み合わせられた複数の読み情報のそれぞれに対応する単語または文字列について、単語間または文字列間もしくは単語と文字列間の文章中の位置の近さを示す接続スコアを読み情報の組み合わせごとに算出する ( S 2 0 9 )。

20

## 【 0 0 8 1 】

決定部 2 6 c は、類似度を算出する際に用いられた音響モデルの組み合わせごとに、すなわち、音響モデルの組み合わせに対応する読み情報の組み合わせごとに、対応する類似度と、対応する接続スコアとの和を算出する ( S 2 1 0 )。決定部 2 6 c は、複数の算出された和のうち、閾値を超えた和があるか否かを判定する ( S 2 1 1 )。ない場合 ( S 2 1 1 否定 ) には、S 2 0 6 へ戻る。一方、閾値を超えた和がある場合 ( S 2 1 1 肯定 ) には、決定部 2 6 c は、閾値を超えた和のうち、最も値が大きい和に対応する文字列を、音声信号に対応する文字列として決定する ( S 2 1 2 )。

## 【 0 0 8 2 】

出力部 2 7 は、各フレームに対して決定された文字列を音声の認識結果として画面に表示するように出力部 2 2 に送信し、各フレームに対して決定された文字列を記憶部 2 4 に格納し ( S 2 1 3 )、処理を終了する。

30

## 【 0 0 8 3 】

## [ 実施例 1 の効果 ]

上述してきたように、認識装置 2 0 は、入力された音声信号と、辞書 1 4 b に登録された読み情報の複数の組み合わせに対応する音響モデルの複数の組み合わせのそれぞれとを比較し、類似度を音響モデルの組み合わせごとに算出する。また、認識装置 2 0 は、類似度を算出する際に組み合わせられた読み情報に対応する単語または文字列の位置情報に基づいて、次のような処理を行う。すなわち、認識装置 2 0 は、組み合わせられた複数の読み情報のそれぞれに対応する単語または文字列について、単語間または文字列間もしくは単語と文字列間の文章中の位置の近さを示す接続スコアを読み情報の組み合わせごとに算出する。そして、認識装置 2 0 は、複数の算出された和のうち、閾値を超え、かつ、最も値が大きい和に対応する文字列を、音声信号に対応する文字列として決定する。認識装置 2 0 によれば、文章に含まれる複数の単語であって、文章内で隣接しない複数の単語を接続した文字列を発話者が発話して、音声信号として入力された場合であっても、発話者が発話した音声を認識することができる。したがって、認識装置 2 0 によれば、精度良く音声の認識を行うことができる。

40

## 【 0 0 8 4 】

また、生成装置 1 0 は、認識装置 2 0 で精度良く音声の認識を行うことができる辞書 1 4 b を生成することができる。

50

## 【 0 0 8 5 】

また、生成装置 1 0 は、総単語数  $M \times$  最大結合数  $N - n$  ( $n = 1, 2, \dots, N - 1$ ) の個数程度のレコードを辞書 1 4 b に登録するため、少ない情報量の辞書 1 4 b を生成することができる。

## 【 実施例 2 】

## 【 0 0 8 6 】

実施例 2 では、テキストファイル 1 4 a に含まれる文章の文ごとに識別子が登録された辞書 1 4 a を生成し、かかる辞書 1 4 a を用いて、文ごとに、接続スコアを算出する場合について説明する。

## 【 0 0 8 7 】

## [ 生成装置 4 0 の機能構成 ]

図 7 は、実施例 2 に係る生成装置の機能構成の一例を示す図である。図 7 に示すように、生成装置 4 0 は、記憶部 1 4 および制御部 4 5 を有する。かかる制御部 4 5 は、図 2 に示す実施例 1 に係る制御部 1 5 と比較して、生成部 4 5 c および格納部 4 5 d を有する点異なる。また、記憶部 1 4 は、実施例 1 に係る記憶部 1 4 と比較して、辞書 4 4 b を記憶する点異なる。なお、以下では、上記の実施例 1 と同様の機能を果たす各部については図 2 と同様の符号を付し、その説明は省略することとする。

## 【 0 0 8 8 】

辞書 4 4 b は、後述する生成部 4 5 c により生成され、後述する格納部 4 5 d により記憶部 1 4 に格納される。

## 【 0 0 8 9 】

図 8 は、辞書の一例を示す図である。図 8 の例では、辞書 4 4 b は、実施例 1 と同様の「表記」の項目、「読み情報」の項目、「位置情報」の項目に加え、「文識別子」の項目を有する。「文識別子」の項目には、文を識別するための識別子が登録される。これらの各項目の内容は、生成部 4 5 c によって生成される。

## 【 0 0 9 0 】

図 8 の例は、一番目のレコードに、テキストファイルが示す文章「決勝に海山学院が進出。山川総理が法案を通したいと言及」に含まれる単語「決勝」が「表記」の項目に登録された場合を示す。また、図 8 の例は、一番目のレコードに、決勝の読み方「けっしょう」が「読み情報」の項目に登録された場合を示す。また、図 8 の例は、一番目のレコードに、単語「決勝」の文章内での位置を示す位置情報「0」が、「位置情報」の項目に登録された場合を示す。また、図 8 の例は、一番目のレコードに、単語「決勝」を含む文の識別子「10」が、「文識別子」の項目に登録された場合を示す。

## 【 0 0 9 1 】

生成部 4 5 c は、分割部 1 5 b により分割された単語の文章内の位置を示す位置情報を生成する。これに加えて、生成部 4 5 c は、単語のそれぞれについて、その単語に続く 1 個から  $(N - 1)$  個の単語までのそれぞれの単語列を生成する。そして、生成部 4 5 c は、単語列の文章内の位置を示す位置情報を生成する。さらに、生成部 4 5 c は、分割された単語を含む文を識別するための識別情報を生成する。

## 【 0 0 9 2 】

具体例を挙げて説明する。例えば、生成部 4 5 c は、まず、位置情報  $n$ 、変数  $i$ 、変数  $j$ 、変数  $s$  の各値を初期化する。すなわち、生成部 4 5 c は、位置情報  $n$  の値を 0、変数  $i$  の値を 1、変数  $j$  の値を 0、変数  $s$  の値を 10 に設定する。そして、生成部 4 5 c は、 $(i - j)$  番目の単語から  $i$  番目の単語までの単語列を生成する。なお、 $j = 0$  の場合は、 $i$  番目の単語を単語列とみなして下記の処理を行う。また、 $(i - w ; w \ j)$  番目の単語が存在しない場合には、生成部 4 5 c は、 $(i - w)$  番目の単語を空欄として扱って、 $(i - j)$  番目の単語から  $i$  番目の単語までの単語列を生成する。

## 【 0 0 9 3 】

そして、生成部 4 5 c は、格納部 4 5 d により新たに生成された単語列と、読み情報と、位置情報  $n$  の値と、文識別子  $s$  の値とが辞書 4 4 b に登録された場合に、変数  $j$  の値を

10

20

30

40

50

1つインクリメントし、変数  $n$  の値を1つインクリメントする。そして、生成部 45c は、変数  $j$  の値が、単語列に結合する単語の数の最大値である最大結合数  $N$  よりも小さいか否かを判定する。小さい場合には、生成部 45c は、再び、 $i$  番目の単語から  $(i + j)$  番目の単語までの単語列を生成し、上記で説明した処理と同様の処理を行う。

【0094】

一方、変数  $j$  の値が、最大結合数  $N$  以上である場合には、生成部 45c は、 $i$  番目の単語が文の区切りを示す記号、例えば、句点「。」であるか否かを判定する。 $i$  番目の単語が文の区切りを示す記号である場合には、生成部 45c は、変数  $s$  の値を1つインクリメントする。そして、生成部 45c は、変数  $i$  の値を1つインクリメントし、変数  $j$  の値を0に設定する。続いて、生成部 45c は、変数  $i$  の値が、分割部 15b により得られた単語の総数である総単語数  $M$  以下であるか否かを判定する。変数  $i$  の値が総単語数  $M$  以下である場合には、生成部 45c は、再び、 $(i - j)$  番目の単語から  $i$  番目の単語までの単語列を生成し、上記で説明した処理と同様の処理を行う。

【0095】

格納部 45d は、単語および単語列と、対応する読み情報と、対応する位置情報  $n$  と、文識別子  $s$  とを対応付けて記憶部 14 に格納する。具体例を挙げて説明する。例えば、格納部 45d は、生成部 45c により、 $(i - j)$  番目の単語から  $i$  番目の単語までの単語列が生成された場合に、次のような処理を行う。すなわち、格納部 45d は、新たに生成された単語列と、新たに生成された単語列に対応する読み情報と、位置情報  $n$  の値と、文識別子  $s$  の値とを、それぞれ、辞書 44b の「表記」、「読み情報」、「位置情報」、「文識別子」の各項目に登録する。なお、格納部 45d は、新たに生成された単語列中の単語に、句点「。」が含まれる場合や、空欄がある場合には、単語列と読み情報と位置情報  $n$  の値とを辞書 14b へ登録する処理を省略できる。これは、単語列中の単語に、句点「。」が含まれる場合や、空欄がある場合には、その単語列は、音声認識する際の音声との比較対象として不適切であるという理由による。したがって、格納部 45d によれば、辞書 14b を生成する速度が速くなる。また、格納部 45d によれば、より少ない情報量の辞書 14b を生成することができる。

【0096】

以上説明したように、生成部 45c および格納部 45d などによって、辞書 44b が生成される。例えば、テキストファイル 14a が示す文章が「決勝に海山学院が進出。山川総理が法案を通したいと言及」である場合には、図8の例に示すような辞書 14b が生成される。図8の例では、最大結合数  $N$  の値は3である。格納部 15c は、総単語数  $M \times$  最大結合数  $N - n$  ( $n = 1, 2, \dots, N - 1$ ) の個数程度のレコードを辞書 44b に登録するため、少ない情報量の辞書 44b を生成することができる。

【0097】

また、格納部 45d は、後述の認識装置 50 から辞書 14b の送信要求を受信すると、記憶部 14 から辞書 44b を読み込み、辞書 44b を通信部 13 に送信する。辞書 44b を受信した通信部 13 は、辞書 44b をインターネット 30 を介して、後述の認識装置 50 へ送信する。

【0098】

制御部 45 は、例えば、ASIC (Application Specific Integrated Circuit) や FPGA (Field Programmable Gate Array) などの集積回路、または、CPU (Central Processing Unit) や MPU (Micro Processing Unit) などの電子回路である。

【0099】

[ 認識装置 50 の機能構成 ]

図9は、実施例2に係る認識装置の構成の一例を示す図である。図9に示すように、認識装置 50 は、記憶部 24 および制御部 55 を有する。かかる制御部 55 は、図4に示す実施例1に係る制御部 25 と比較して、照合部 56 の第二の算出部 56b を有する点異なる。また、記憶部 24 は、実施例1に係る記憶部 24 と比較して、辞書 44b を記憶する点異なる。なお、以下では、上記の実施例1と同様の機能を果たす各部については図

10

20

30

40

50



4と同様の符号を付し、その説明は省略することとする。

【0100】

辞書44bは、第一の算出部26aにより記憶部24に格納される。

【0101】

第二の算出部56bは、実施例1の第二の算出部26bとほぼ同様の機能を有するが、下記の点で、第二の算出部26bと異なる。第二の算出部56bは、辞書44bを参照し、文識別子が同一の読み情報間でのみ、実施例1と同様に、昇順であるか否かの判定を行う。また、第二の算出部56bは、位置情報が示す値が昇順であると判定された組み合わせについて、次のような処理を行う。すなわち、組み合わせされた読み情報のうち、結合された2つの読み情報、すなわち隣接する読み情報について、先頭に近いほうの読み情報  
10  
に対応する位置情報が示す値を変数xに代入し、先頭から遠い方の読み情報に対応する位置情報が示す値を変数yに代入する。そして、第二の算出部56bは、下記の式(3)、式(4)に基づいて、X、Yを算出する。

$$X = (x / N) + 1 \quad \text{式(3)}$$

$$Y = (y / N) - (y \% N) \quad \text{式(4)}$$

【0102】

ただし、Nは、上述した最大結合数である。また、「/」は、小数点以下を切り捨てて商を整数とするための除算演算を意味する。また、「%」は、剰余演算を意味する。

【0103】

なお、XとYとを比較し、接続スコアを算出する方法は、実施例1と同様である。  
20

【0104】

このように、第二の算出部56bによれば、同一の文内でのみ、接続スコアを算出する。したがって、内容が異なる文をまたがった場合などに接続スコアが算出されないの  
より精度良く音声を認識することができる。

【0105】

制御部55は、例えば、ASIC (Application Specific Integrated Circuit) や FPGA (Field Programmable Gate Array) などの集積回路、または、CPU (Central Processing Unit) や MPU (Micro Processing Unit) などの電子回路である。

【0106】

[処理の流れ]

次に、本実施例に係る生成装置40の処理の流れを説明する。図10は、実施例2に係る生成処理の手順を示すフローチャートである。この生成処理は、入力部11から制御部45に生成処理を実行する指示が入力された場合に実行される。  
30

【0107】

図10に示すように、分割部15bは、テキストファイル14aが示す文章を単語ごとに分割する(S301)。生成部45cは、位置情報nの値を0、変数iの値を1、変数jの値を0、変数sの値を10に設定する(S302)。生成部45cは、(i-j)番目の単語からi番目の単語までの単語列を生成する(S303)。

【0108】

格納部45dは、新たに生成された単語列と、新たに生成された単語列に対応する読み情報と、位置情報nの値と、文識別子sとを、それぞれ、辞書44bの「表記」、「読み情報」、「位置情報」、「文識別子」の各項目に登録する(S304)。生成部45cは、変数jの値を1つインクリメントし、変数nの値を1つインクリメントする(S305)。生成部45cは、変数jの値が、単語列に結合する単語の数の最大値である最大結合数Nよりも小さいか否かを判定する(S306)。小さい場合(S306肯定)には、S303へ戻る。  
40

【0109】

一方、変数jの値が、最大結合数N以上である場合(S306否定)には、生成部45cは、i番目の単語が文の区切りを示す記号、例えば、句点「。」であるか否かを判定する(S307)。i番目の単語が文の区切りを示す記号である場合(S307肯定)には  
50

、生成部 45c は、変数  $s$  の値を 1 つインクリメントする (S308)。生成部 45c は、変数  $i$  の値を 1 つインクリメントし、変数  $j$  の値を 0 に設定する (S309)。生成部 45c は、変数  $i$  の値が、分割部 15b により得られた単語の総数である総単語数  $M$  以下であるか否かを判定する (S310)。変数  $i$  の値が総単語数  $M$  以下である場合 (S310 肯定) には、S303 へ戻る。一方、変数  $i$  の値が総単語数  $M$  より大きい場合 (S310 否定) には、処理を終了する。また、 $i$  番目の単語が文の区切りを示す記号でない場合 (S307 否定) には、S309 へ進む。

#### 【0110】

次に、本実施例に係る認識装置 50 の処理の流れを説明する。図 11 は、実施例 2 に係る認識処理の手順を示すフローチャートである。この認識処理は、操作受付部 (図示せず) から制御部 55 に認識処理を実行する指示が入力された場合に実行される。

10

#### 【0111】

図 11 に示す S401 ~ S408、S410 ~ S413 は、上述した S201 ~ S208、S210 ~ S213 と同様の処理であるため説明を省略する。図 11 に示すように、第二の算出部 56b は、辞書 44b を参照し、文識別子が同一の読み情報間でのみ、接続スコアを算出する (S409)。

#### 【0112】

##### [実施例 2 の効果]

上述してきたように、認識装置 50 は、入力された音声信号と、辞書 44b に登録された読み情報の複数の組み合わせに対応する音響モデルの複数の組み合わせのそれぞれとを比較し、類似度を音響モデルの組み合わせごとに算出する。また、認識装置 50 は、類似度を算出する際に組み合わせられた読み情報に対応する単語または文字列の位置情報に基づいて、次のような処理を行う。すなわち、認識装置 50 は、組み合わせられた複数の読み情報のそれぞれに対応する単語または文字列について、単語間または文字列間もしくは単語と文字列間の文章中の位置の近さを示す接続スコアを読み情報の組み合わせごとに算出する。そして、認識装置 50 は、複数の算出された和のうち、閾値を超え、かつ、最も値が大きい和に対応する文字列を、音声信号に対応する文字列として決定する。認識装置 50 によれば、文章に含まれる複数の単語であって、文章内で隣接しない複数の単語を接続した文字列を発話者が発話して、音声信号として入力された場合であっても、発話者が発話した音声を認識することができる。したがって、認識装置 50 によれば、精度良く音声の認識を行うことができる。

20

30

#### 【0113】

また、生成装置 40 は、認識装置 50 で精度良く音声の認識を行うことができる辞書 44b を生成することができる。

#### 【0114】

また、生成装置 40 は、総単語数  $M \times$  最大結合数  $N - n$  ( $n = 1, 2, \dots, N - 1$ ) の個数程度のレコードを辞書 44b に登録するため、少ない情報量の辞書 44b を生成することができる。

#### 【0115】

また、認識装置 50 によれば、同一の文内でのみ、接続スコアを算出する。したがって、内容が異なる文をまたがった場合などに接続スコアが算出されないので、より精度良く音声を認識することができる。

40

#### 【実施例 3】

#### 【0116】

実施例 3 では、さらに、予め定められたオートマトンなどを用いて、接続スコアを補正する場合について説明する。

#### 【0117】

##### [認識装置 60 の機能構成]

図 12 は、実施例 3 に係る認識装置の構成の一例を示す図である。図 12 に示すように、認識装置 60 は、記憶部 24 および制御部 65 を有する。かかる制御部 65 は、図 4 に

50

示す実施例 1 に係る制御部 2 5 と比較して、照合部 6 6 の第二の算出部 6 6 b を有する点が異なる。また、記憶部 2 4 は、辞書 6 4 b を記憶するが、かかる辞書 6 4 b は、実施例 1 に係る辞書 1 4 b または実施例 2 に係る辞書 4 4 b と同一である。また、記憶部 2 4 は、言語モデル 6 4 c、単語辞書 6 4 d を記憶する。なお、以下では、上記の実施例 1、2 と同様の機能を果たす各部については図 4、図 9 と同様の符号を付し、その説明は省略することとする。

【 0 1 1 8 】

辞書 6 4 b は、第一の算出部 2 6 a により記憶部 2 4 に格納される。

【 0 1 1 9 】

図 1 3 は、言語モデルの一例を示す図である。図 1 3 の例では、言語モデル 6 4 c は、例えば、オートマトンである。図 1 3 の例は、ある単語または文字列に、「に関する」、「に関連する」、「についての」の文字列が付加される場合に、状態 1 から状態 2 へと遷移することを示す。また、図 1 3 の例は、状態 2 の状態において、「に関する」、「に関連する」、「についての」の文字列に、「ニュース」、「解説」、「ブログ」、「話題」の単語が付加される場合に、状態 2 から最終状態へと遷移することを示す。また、図 1 3 の例は、ある単語に「を検索」、「のニュース」の文字列が付加される場合に、状態 1 から最終状態へと遷移することを示す。

10

【 0 1 2 0 】

図 1 4 は、単語辞書の一例を示す図である。図 1 4 の例では、単語辞書 6 4 d は、「表記」の項目と、「読み情報」の項目とを有する。図 1 4 の例の単語辞書 6 4 d には、単語モデル 6 4 c が示す各状態を遷移する場合の単語または文字列の表記、および、読み情報が対応付けて登録されている。

20

【 0 1 2 1 】

第二の算出部 6 6 b は、実施例 1 の第二の算出部 2 6 b または実施例 2 の第二の算出部 5 6 b の機能を有するが、下記の点で、第二の算出部 2 6 b、5 6 b と異なる。すなわち、第二の算出部 6 6 b は、実施例 1 または実施例 2 と同様に、接続スコアを算出するが、かかる接続スコアを、言語モデル 6 4 c および単語辞書 6 4 d に基づいて補正する点が、実施例 1 および実施例 2 と異なる。

【 0 1 2 2 】

例えば、組み合わせられた読み情報「けっしょうにうみやまがくいんがしんしゅつ」（表記：決勝に海山学院が進出）について、接続スコア S が算出された場合を想定する。この場合、第二の算出部 6 6 b は、言語モデル 6 4 c および単語辞書 6 4 d を参照し、組み合わせられた読み情報に対応する状態が、「状態 1」であるため、接続スコア S については補正を行わない。

30

【 0 1 2 3 】

また、例えば、組み合わせられた読み情報「けっしょうにうみやまがくいんがしんしゅつのニュース」（表記：決勝に海山学院が進出のニュース）について、接続スコア Q が算出された場合を想定する。この場合、第二の算出部 6 6 b は、接続スコア Q でなく、最終状態に遷移する前の状態 1 に対応する読み情報「けっしょうにうみやまがくいんがしんしゅつ」について算出された接続スコア S を補正する。例えば、第二の算出部 6 6 b は、言語モデル 6 4 c および単語辞書 6 4 d を参照し、組み合わせられた読み情報に対応する状態が、「状態 1」から「最終状態」へと遷移したため、接続スコア S に所定値を加算して接続スコアを補正する。なお、接続スコア S に、所定値、例えば、100%を示す値を乗算して、接続スコアを補正することもできる。

40

【 0 1 2 4 】

また、例えば、組み合わせられた読み情報「やまかわそうりがとおしたいほうあん」（表記：山川総理が通したい法案）について、接続スコア S' が算出された場合を想定する。この場合、第二の算出部 6 6 b は、言語モデル 6 4 c および単語辞書 6 4 d を参照し、組み合わせられた読み情報に対応する状態が、「状態 1」であるため、接続スコア S' については補正を行わない。

50

## 【 0 1 2 5 】

また、例えば、組み合わされた読み情報「やまかわそうりがとおしたいほうあんについてのブログ」（表記：山川総理が通したい法案についてのブログ）について、接続スコア  $Q'$  が算出された場合を想定する。この場合、第二の算出部 66b は、接続スコア  $Q'$  でなく、状態 1 から状態 2、状態 2 から最終状態に遷移する前の状態 1 に対応する読み情報「やまかわそうりがとおしたいほうあん」について算出された接続スコア  $S'$  を補正する。例えば、第二の算出部 66b は、言語モデル 64c および単語辞書 64d を参照し、組み合わされた読み情報に対応する状態が、「状態 1」から「状態 2」、「状態 2」から「最終状態」へと遷移したため、接続スコア  $S'$  に所定値を 2 回加算して接続スコアを補正する。なお、接続スコア  $S$  に、所定値、例えば、100% を示す値を 2 回乗算して、接続スコアを補正することもできる。

10

## 【 0 1 2 6 】

制御部 65 は、例えば、ASIC (Application Specific Integrated Circuit) や FPGA (Field Programmable Gate Array) などの集積回路、または、CPU (Central Processing Unit) や MPU (Micro Processing Unit) などの電子回路である。

## 【 0 1 2 7 】

## [ 処理の流れ ]

次に、本実施例に係る認識装置 60 の処理の流れを説明する。図 15 は、実施例 3 に係る認識処理の手順を示すフローチャートである。この認識処理は、操作受付部（図示せず）から制御部 65 に認識処理を実行する指示が入力された場合に実行される。

20

## 【 0 1 2 8 】

図 15 に示す S501 ~ S509、S511 ~ S514 は、上述した S401 ~ S413、または、S201 ~ S213 と同様の処理であるため説明を省略する。図 15 に示すように、第二の算出部 66b は、言語モデル 64c および単語辞書 64d を参照し、状態が遷移した場合には、状態 1 の場合に算出された接続スコアを補正する (S510)。

## 【 0 1 2 9 】

## [ 実施例 3 の効果 ]

上述してきたように、認識装置 60 は、入力された音声信号と、辞書 64b に登録された読み情報の複数の組み合わせに対応する音響モデルの複数の組み合わせのそれぞれとを比較し、類似度を音響モデルの組み合わせごとに算出する。また、認識装置 60 は、類似度を算出する際に組み合わされた読み情報に対応する単語または文字列の位置情報に基づいて、次のような処理を行う。すなわち、認識装置 60 は、組み合わされた複数の読み情報のそれぞれに対応する単語または文字列について、単語間または文字列間もしくは単語と文字列間の文章中の位置の近さを示す接続スコアを読み情報の組み合わせごとに算出する。そして、認識装置 60 は、複数の算出された和のうち、閾値を超え、かつ、最も値が大きい和に対応する文字列を、音声信号に対応する文字列として決定する。認識装置 60 によれば、文章に含まれる複数の単語であって、文章内で隣接しない複数の単語を接続した文字列を発話者が発話して、音声信号として入力された場合であっても、発話者が発話した音声認識することができる。したがって、認識装置 60 によれば、精度良く音声の認識を行うことができる。

30

40

## 【 0 1 3 0 】

また、本実施例にかかる生成装置は、認識装置 60 で精度良く音声の認識を行うことができる辞書 64b を生成することができる。

## 【 0 1 3 1 】

また、本実施例にかかる生成装置は、総単語数  $M \times$  最大結合数  $N - n$  ( $n = 1, 2, \dots, N - 1$ ) の個数程度のレコードを辞書 64b に登録するため、少ない情報量の辞書 64b を生成することができる。

## 【 0 1 3 2 】

また、認識装置 60 によれば、同一の文内でのみ、接続スコアを算出する。したがって、内容が異なる文をまたがった場合などに接続スコアが算出されないため、より精度良く

50

音声を認識することができる。

【 0 1 3 3 】

また、認識装置 6 0 によれば、オートマトンなどの言語モデルを用いて接続スコアを補正するため、特定の発話パターンを言語モデルに登録することで、特定の発話パターンの認識率を向上させることができる。

【 実施例 4 】

【 0 1 3 4 】

実施例 4 では、さらに、読み情報（読み情報に対応する単語または文字列）の組み合わせにおいて、隣接する 2 つの読み情報の接続部分の 2 つの品詞の接続の適否を示す情報を用いて、接続スコアを補正する場合について説明する。

10

【 0 1 3 5 】

[ 生成装置 7 0 の機能構成 ]

図 1 6 は、実施例 4 に係る生成装置の機能構成の一例を示す図である。図 1 6 に示すように、生成装置 7 0 は、記憶部 1 4 および制御部 7 5 を有する。かかる制御部 7 5 は、図 2 に示す実施例 1 に係る制御部 1 5 と比較して、生成部 7 5 c および格納部 7 5 d を有する点が異なる。また、記憶部 1 4 は、実施例 1 に係る記憶部 1 4 と比較して、辞書 7 4 b および品詞番号表 7 4 c を記憶する点が異なる。なお、以下では、上記の実施例 1、2、3 と同様の機能を果たす各部については図 2、図 7 と同様の符号を付し、その説明は省略することとする。

【 0 1 3 6 】

辞書 7 4 b は、後述する生成部 7 5 c により生成され、後述する格納部 7 5 d により記憶部 1 4 に格納される。

20

【 0 1 3 7 】

図 1 7 は、品詞番号表の一例を示す図である。図 1 7 の例では、品詞番号表 7 4 c は、先頭の単語の品詞が名詞類で、末尾の単語の品詞が名詞類である場合には、品詞番号が「1」である場合が示されている。

【 0 1 3 8 】

図 1 8 は、辞書の一例を示す図である。図 1 8 の例では、辞書 7 4 b は、実施例 1 と同様の「表記」の項目、「読み情報」の項目、「位置情報」の項目に加え、「品詞番号」の項目を有する。「品詞番号」の項目には、「表記」の項目に登録された単語または単語列の先頭の単語の品詞と、末尾の単語の品詞との組み合わせを示す番号である品詞番号が登録される。かかる品詞番号は、後述の生成部 7 5 c によって品詞番号表 7 4 c から取得され、後述の格納部 7 5 d によって、辞書 7 4 b に登録される。

30

【 0 1 3 9 】

生成部 7 5 c は、分割された単語または生成した単語列について、先頭の単語の品詞と、末尾の単語の品詞との組み合わせを示す品詞番号を、品詞番号表 7 4 c を参照して取得する。

【 0 1 4 0 】

格納部 7 5 d は、実施例 1 ~ 3 のいずれかの格納部の機能に加え、下記の機能を有する。すなわち、格納部 7 5 d は、生成部 7 5 c によって取得された品詞番号を辞書 7 4 b の「品詞番号」の項目に登録する。

40

【 0 1 4 1 】

制御部 7 5 は、例えば、A S I C (Application Specific Integrated Circuit) や F P G A (Field Programmable Gate Array) などの集積回路、または、C P U (Central Processing Unit) や M P U (Micro Processing Unit) などの電子回路である。

【 0 1 4 2 】

[ 認識装置 8 0 の機能構成 ]

図 1 9 は、実施例 4 に係る認識装置の構成の一例を示す図である。図 1 9 に示すように、認識装置 8 0 は、記憶部 2 4 および制御部 8 5 を有する。かかる制御部 8 5 は、図 4 に示す実施例 1 に係る制御部 2 5 と比較して、照合部 8 6 の第二の算出部 8 6 b を有する点

50

が異なる。また、記憶部 2 4 は、実施例 1 に係る記憶部 2 4 と比較して、辞書 7 4 b と、確率表 8 4 a を記憶する点が異なる。なお、以下では、上記の実施例 1 ~ 3 と同様の機能を果たす各部については図 4、図 9、図 1 2 と同様の符号を付し、その説明は省略することとする。

【 0 1 4 3 】

辞書 7 4 b は、第一の算出部 2 6 a により記憶部 2 4 に格納される。

【 0 1 4 4 】

図 2 0 は、確率表の一例を示す図である。図 2 0 の例の確率表 8 4 a では、縦軸が、2 つの単語または文字列もしくは単語と文字列との組み合わせのうち、前の単語または文字列の品詞番号を示す。また、図 2 0 の例の確率表 8 4 a では、横軸が、2 つの単語または文字列もしくは単語と文字列との組み合わせのうち、後ろの単語または文字列の品詞番号を示す。図 2 0 の例の確率表 8 4 a には、前の単語の品詞番号が 1 であり、後ろの単語の品詞番号が 1 である場合には、1 0 0 % の確率が登録されている。

10

【 0 1 4 5 】

第二の算出部 8 6 b は、実施例 1 ~ 3 のいずれかの第二の算出部とほぼ同様の機能を有するが、下記の点で、実施例 1 ~ 3 のいずれの第二の算出部とも異なる。すなわち、第二の算出部 8 6 b は、まず、組み合わせされた隣接する単語または文字列を特定する。そして、第二の算出部 8 6 b は、確率表 8 4 a を参照し、特定された隣接する単語または文字列もしくは単語と文字列の 2 つの品詞番号から、品詞の組み合わせの適否を示す値、例えば、2 つの単語または文字列もしくは単語と文字列が接続される確率を取得する。そして、第二の算出部 8 6 b は、取得した確率と、対応する組み合わせの接続スコアとを乗算して、接続スコアを補正する。

20

【 0 1 4 6 】

制御部 8 5 は、例えば、A S I C (Application Specific Integrated Circuit) や F P G A (Field Programmable Gate Array) などの集積回路、または、C P U (Central Processing Unit) や M P U (Micro Processing Unit) などの電子回路である。

【 0 1 4 7 】

[ 処理の流れ ]

次に、本実施例に係る生成装置 7 0 の処理の流れを説明する。図 2 1 は、実施例 4 に係る生成処理の手順を示すフローチャートである。この生成処理は、入力部 1 1 から制御部 7 5 に生成処理を実行する指示が入力された場合に実行される。

30

【 0 1 4 8 】

図 2 1 に示す S 6 0 1 ~ S 6 0 3、S 6 0 5 ~ S 6 0 8 の処理は、上述した S 1 0 1 ~ S 1 0 3、S 1 0 5 ~ S 1 0 8 の処理と同様であるので、説明を省略する。図 2 1 に示すように、格納部 7 5 d は、実施例 1 ~ 3 の格納部の機能に加えて、品詞番号を辞書 7 4 b に登録する ( S 6 0 4 ) 。

【 0 1 4 9 】

次に、本実施例に係る認識装置 8 0 の処理の流れを説明する。図 2 2 は、実施例 4 に係る認識処理の手順を示すフローチャートである。この認識処理は、操作受付部 ( 図示せず ) から制御部 8 5 に認識処理を実行する指示が入力された場合に実行される。

40

【 0 1 5 0 】

図 2 2 に示す S 7 0 1 ~ S 7 0 9、S 7 1 1 ~ S 7 1 4 は、上述した S 5 0 1 ~ S 5 0 9、S 5 1 1 ~ S 5 1 4 と同様の処理であるため説明を省略する。図 2 2 に示すように、第二の算出部 8 6 b は、確率表 8 4 a を参照し、品詞番号を用いて、接続スコアを補正する ( S 7 1 0 ) 。

【 0 1 5 1 】

[ 実施例 4 の効果 ]

上述してきたように、認識装置 8 0 は、入力された音声信号と、辞書 7 4 b に登録された読み情報の複数の組み合わせに対応する音響モデルの複数の組み合わせのそれぞれとを比較し、類似度を音響モデルの組み合わせごとに算出する。また、認識装置 8 0 は、類似

50

度を算出する際に組み合わせられた読み情報に対応する単語または文字列の位置情報に基づいて、次のような処理を行う。すなわち、認識装置 80 は、組み合わせられた複数の読み情報のそれぞれに対応する単語または文字列について、単語間または文字列間もしくは単語と文字列間の文章中の位置の近さを示す接続スコアを読み情報の組み合わせごとに算出する。そして、認識装置 80 は、複数の算出された和のうち、閾値を超え、かつ、最も値が大きい和に対応する文字列を、音声信号に対応する文字列として決定する。認識装置 80 によれば、文章に含まれる複数の単語であって、文章内で隣接しない複数の単語を接続した文字列を発話者が発話して、音声信号として入力された場合であっても、発話者が発話した音声を認識することができる。したがって、認識装置 80 によれば、精度良く音声の認識を行うことができる。

10

【0152】

また、生成装置 70 は、認識装置 80 で精度良く音声の認識を行うことができる辞書 74b を生成することができる。

【0153】

また、生成装置 70 は、総単語数  $M \times$  最大結合数  $N - n$  ( $n = 1, 2, \dots, N - 1$ ) の個数程度のレコードを辞書 74b に登録するため、少ない情報量の辞書 74b を生成することができる。

【0154】

また、認識装置 80 によれば、同一の文内でのみ、接続スコアを算出する。したがって、内容が異なる文をまたがった場合などに接続スコアが算出されないため、より精度良く音声の認識を行うことができる。

20

【0155】

また、認識装置 80 によれば、品詞の適否を加味された接続スコアに基づいて、音声信号に対応する文字列が決定されるので、より精度の高い音声の認識を行うことができる。

【0156】

さて、これまで開示の装置に関する実施例について説明したが、本発明は上述した実施例以外にも、種々の異なる形態にて実施されてよいものである。そこで、以下では、本発明に含まれる他の実施例を説明する。

【0157】

また、各実施例において説明した処理のうち、自動的に行われるものとして説明した処理の全部または一部を手動的に行うこともできる。また、各実施例において説明した処理のうち、手動的に行われるものとして説明した処理の全部または一部を公知の方法で自動的行うこともできる。

30

【0158】

また、各種の負荷や使用状況などに応じて、各実施例において説明した各処理の各ステップでの処理を任意に細かく分けたり、あるいはまとめたりすることができる。また、ステップを省略することもできる。

【0159】

また、各種の負荷や使用状況などに応じて、各実施例において説明した各処理の各ステップでの処理の順番を変更できる。

40

【0160】

また、図示した各装置の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、各装置の分散・統合の具体的状態は図示のものに限られず、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。

【0161】

[生成プログラム]

また、上記の各実施例で説明した生成装置の各種の処理は、あらかじめ用意されたプログラムをパーソナルコンピュータやワークステーションなどのコンピュータシステムで実行することによって実現することもできる。そこで、以下では、図 23 を用いて、上記の

50

実施例で説明した生成装置と同様の機能を有する生成プログラムを実行するコンピュータの一例を説明する。図23は、生成プログラムを実行するコンピュータを示す図である。

【0162】

図23に示すように、コンピュータ300は、CPU(Central Processing Unit)310、ROM(Read Only Memory)320、HDD(Hard Disk Drive)330、RAM(Random Access Memory)340を有する。これら310~340の各部は、バス350を介して接続される。

【0163】

ROM320には、OS(Operating System)などのプログラムが記憶される。

【0164】

HDD330には、上記のいずれかの実施例で示す生成装置の分割部、生成部、格納部と同様の機能を発揮する生成プログラム330aが予め記憶される。なお、生成プログラム330aについては、適宜分離しても良い。

【0165】

そして、CPU310が、生成プログラム330aをHDD330から読み出して実行する。

【0166】

また、HDD330には、上記のいずれかの実施例の記憶部に記憶された情報、例えば、テキストファイル、辞書、品詞番号表などが設けられる。

【0167】

そして、CPU310は、テキストファイル、辞書、品詞番号表などを読み出してRAM340に格納する。さらに、CPU310は、RAM340に格納された各種情報を用いて、生成プログラムを実行する。なお、RAM340に格納される各データは、常に全てのデータがRAM340に格納される必要はなく、処理に必要なデータのみがRAM340に格納されれば良い。

【0168】

なお、上記した生成プログラムについては、必ずしも最初からHDD330に記憶させておく必要はない。

【0169】

例えば、コンピュータ300に挿入されるフレキシブルディスク(FD)、CD-ROM、DVDディスク、光磁気ディスク、ICカードなどの「可搬用の物理媒体」にプログラムを記憶させておく。そして、コンピュータ300がこれらからプログラムを読み出して実行するようにしてもよい。

【0170】

さらには、公衆回線、インターネット、LAN、WANなどを介してコンピュータ300に接続される「他のコンピュータ(またはサーバ)」などにプログラムを記憶させておく。そして、コンピュータ300がこれらからプログラムを読み出して実行するようにしてもよい。

【0171】

[認識プログラム]

また、上記の各実施例で説明した認識装置の各種の処理は、あらかじめ用意されたプログラムをパーソナルコンピュータやワークステーションなどのコンピュータシステムで実行することによって実現することもできる。そこで、以下では、図24を用いて、上記の実施例で説明した認識装置と同様の機能を有する認識プログラムを実行するコンピュータの一例を説明する。図24は、認識プログラムを実行するコンピュータを示す図である。

【0172】

図24に示すように、コンピュータ400は、CPU410、ROM420、HDD430、RAM440を有する。これら410~440の各部は、バス450を介して接続される。

【0173】

10

20

30

40

50



ROM 420には、OSなどのプログラムが記憶される。

【0174】

HDD 430には、上記のいずれかの実施例で示す認識装置の第一の算出部、第二の算出部、決定部、出力部と同様の機能を発揮する認識プログラム330aが予め記憶される。なお、認識プログラム430aについては、適宜分離しても良い。

【0175】

そして、CPU 410が、認識プログラム430aをHDD 430から読み出して実行する。

【0176】

また、HDD 430には、上記のいずれかの実施例の記憶部に記憶された情報、例えば、辞書、言語モデル、単語辞書、確率表などが設けられる。

10

【0177】

そして、CPU 410は、辞書、言語モデル、単語辞書、確率表などを読み出してRAM 440に格納する。さらに、CPU 410は、RAM 440に格納された各種情報を用いて、認識プログラムを実行する。なお、RAM 440に格納される各データは、常に全てのデータがRAM 440に格納される必要はなく、処理に必要なデータのみがRAM 440に格納されれば良い。

【0178】

なお、上記した認識プログラムについては、必ずしも最初からHDD 430に記憶させておく必要はない。

20

【0179】

例えば、コンピュータ400に挿入されるフレキシブルディスク(FD)、CD-ROM、DVDディスク、光磁気ディスク、ICカードなどの「可搬用の物理媒体」にプログラムを記憶させておく。そして、コンピュータ400がこれらからプログラムを読み出して実行するようにしてもよい。

【0180】

さらには、公衆回線、インターネット、LAN、WANなどを介してコンピュータ400に接続される「他のコンピュータ(またはサーバ)」などにプログラムを記憶させておく。そして、コンピュータ300がこれらからプログラムを読み出して実行するようにしてもよい。

30

【符号の説明】

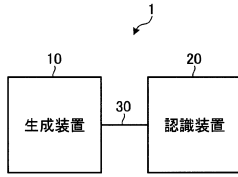
【0181】

- 14b 辞書
- 20 認識装置
- 24 記憶部
- 25 制御部
- 26 照合部
- 26a 第一の算出部
- 26b 第二の算出部
- 26c 決定部

40

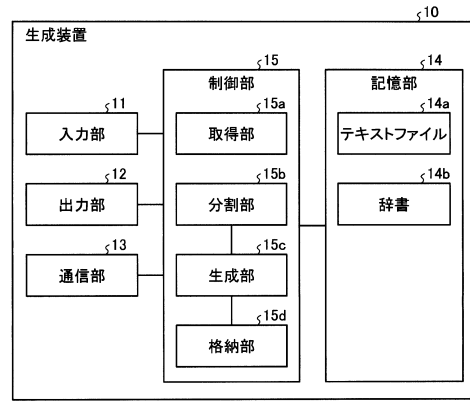
【図1】

実施例1に係る認識装置および生成装置を有するシステムの構成の一例を示す図



【図2】

実施例1に係る生成装置の機能構成の一例を示す図



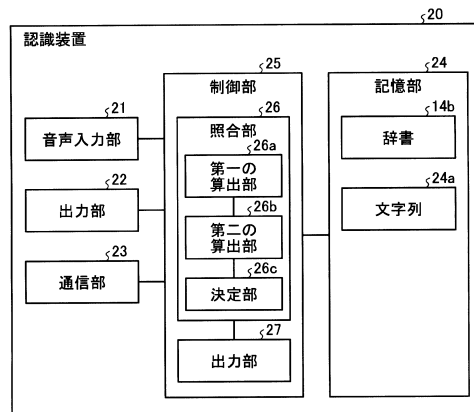
【図3】

辞書の一例を示す図

表記	読み情報	位置情報
衆議院	しゅうぎいん	0
衆議院で	しゅうぎいんで	1
衆議院で改正法	しゅうぎいんでかいせいほう	2
で	で	3
で改正法が	でかいせいほうが	4
改正法が	かいせいほうが	5
改正法が審議	かいせいほうがしんぎ	6
が	が	7
が審議	がしんぎ	8
が審議入り	がしんぎいり	9
審議	しんぎ	10
審議入り	しんぎいり	11
(審議入り)	(しんぎいり)	12
入り	いり	13
(入り、山川	(いり、やまかわ	14)
(	(	15)
、山川総理	、やまかわ	16)
山川	やまかわ	17)
山川総理	やまかわ	18)
山川総理が	やまかわ	19)
総理	そうり	20)
総理が法案	そうりがほうあん	21)
が	が	22)
が法案を	がほうあんとおし	23)
法案	ほうあん	24)
法案を	ほうあんとおし	25)
法案を通	ほうあんとおし	26)
を	を	27)
を通し	をとおし	28)
通し	とおし	29)
通したい	とおしたい	30)
と	と	31)
とを	とを	32)
とを	とを	33)
とを	とを	34)
とを	とを	35)
とを	とを	36)
とを	とを	37)
とを	とを	38)
とを	とを	39)
とを	とを	40)
とを	とを	41)
とを	とを	42)
とを	とを	43)
とを	とを	44)
とを	とを	45)
とを	とを	46)
とを	とを	47)
とを	とを	48)
とを	とを	49)
とを	とを	50)

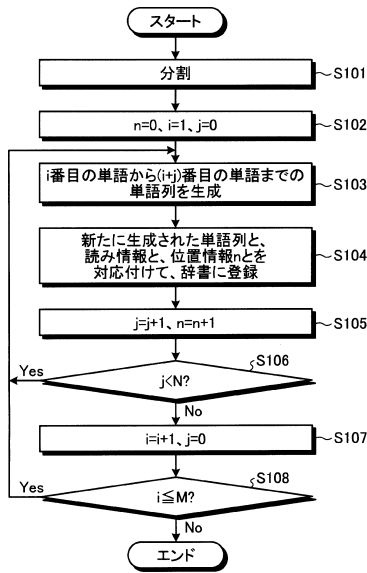
【図4】

実施例1に係る認識装置の機能構成の一例を示す図



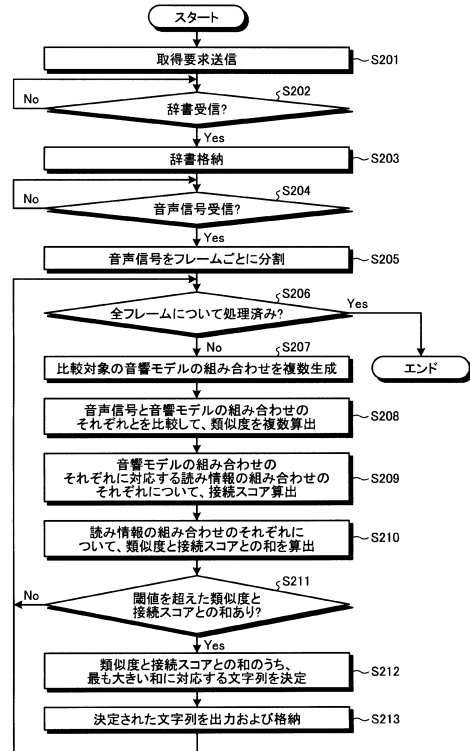
【図5】

実施例1に係る生成処理の手順を示すフローチャート



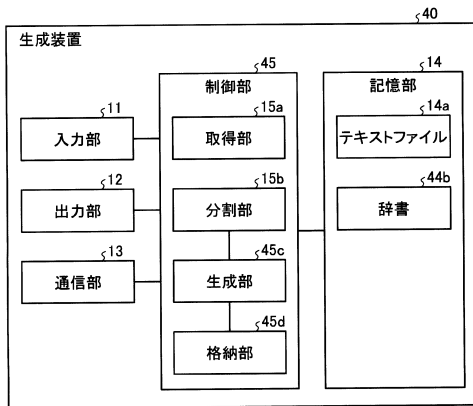
【図6】

実施例1に係る認識処理の手順を示すフローチャート



【図7】

実施例2に係る生成装置の機能構成の一例を示す図



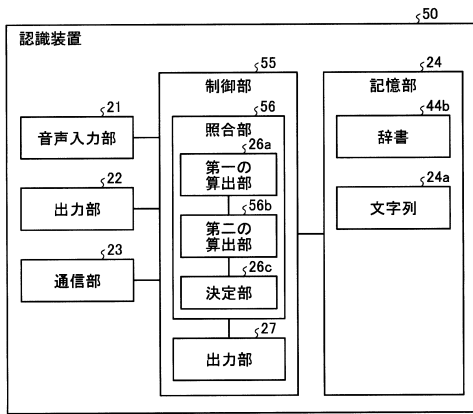
【図8】

辞書の一例を示す図

表記	読み情報	位置情報	文議別子
決勝	けっしょう	0	10
に	に	3	10
決勝に	けっしょうに	4	10
海山学院	うみやまがくいん	6	10
に海山学院	にうみやまがくいん	7	10
決勝に海山学院	けっしょうにうみやまがくいん	8	10
が	が	9	10
海山学院が	うみやまがくいんが	10	10
に海山学院が	にうみやまがくいんが	11	10
進出	しんしゅつ	12	10
が進出	がしんしゅつ	13	10
海山学院が進出	うみやまがくいんがしんしゅつ	14	10
山川	やまかわ	18	11
総理	そうり	21	11
山川総理	やまかわそうり	22	11
が	が	24	11
総理が	そうりが	25	11
山川総理が	やまかわそうりが	26	11
法案	ほうあん	27	11
が法案	がほうあん	28	11
総理が法案	そうりがほうあん	29	11
を	を	30	11
法案を	ほうあんを	31	11
が法案を	がほうあんを	32	11
通	とお	33	11
を通	とおし	34	11
法案を通	ほうあんをと	35	11
し	し	36	11
通し	とおし	37	11
を通し	をと	38	11
たい	たい	39	11
したい	したい	40	11
通したい	とおしたい	41	11
と	と	42	11
たいと	たいと	43	11
したいと	したいと	44	11
言及	げんきゆう	45	11
と言及	とげんきゆう	46	11
たいと言及	たいとげんきゆう	47	11

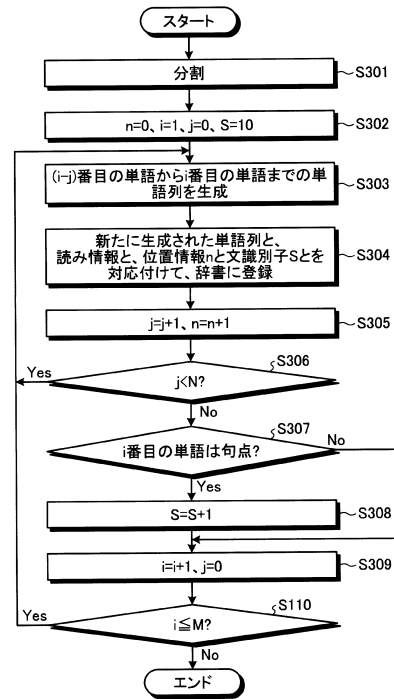
【図9】

実施例2に係る認識装置の構成の一例を示す図



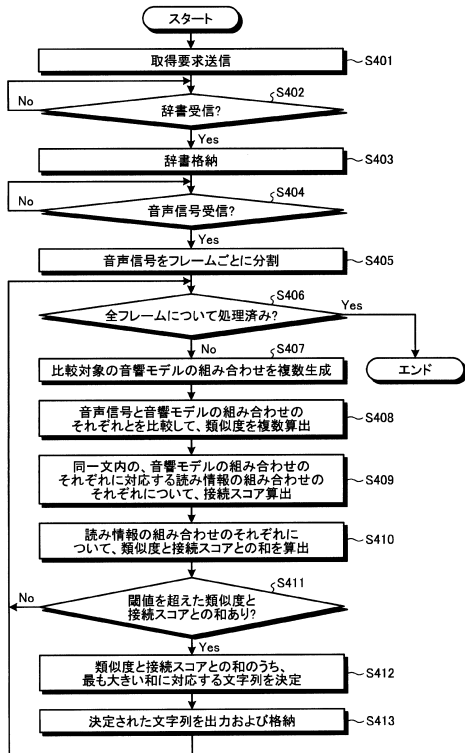
【図10】

実施例2に係る生成処理の手順を示すフローチャート



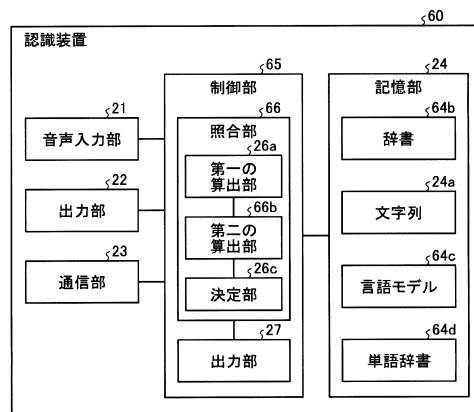
【図11】

実施例2に係る認識処理の手順を示すフローチャート



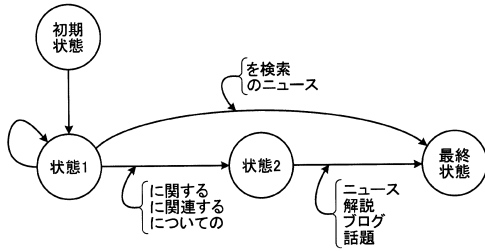
【図12】

実施例3に係る認識装置の構成の一例を示す図



【図13】

言語モデルの一例を示す図



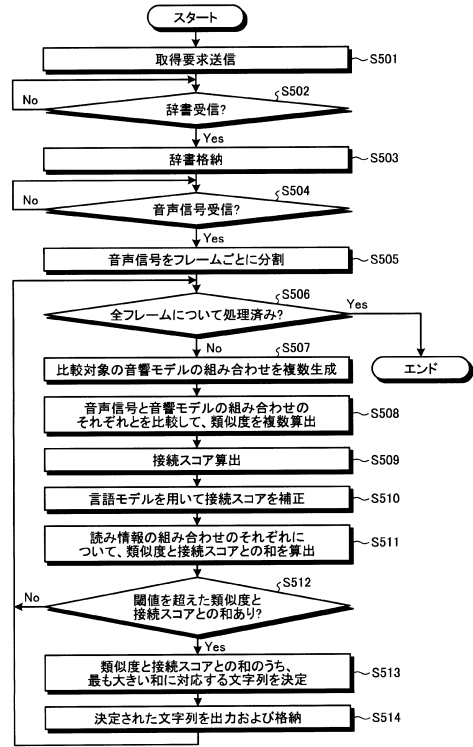
【図14】

単語辞書の一例を示す図

表記	読み情報
に関する	にかんする
に連するにの	にかんれんするにの
のニュース	にゆーす
解説	かいせつ
ブログ	ぶろぐ
話題	わだい
を探索	をけんさく
のニュース	のにゆーす

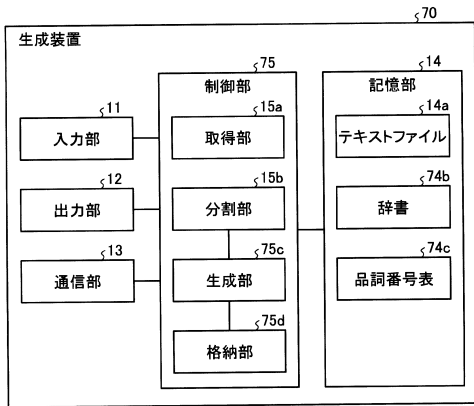
【図15】

実施例3に係る認識処理の手順を示すフローチャート



【図16】

実施例4に係る生成装置の構成の一例を示す図



【図17】

品詞番号表の一例を示す図

先頭の単語の品詞 \ 末尾の単語の品詞	品詞		
	名詞類	動詞類	格助詞
名詞類	1	2	3
動詞類	4	5	6
格助詞	7	8	9

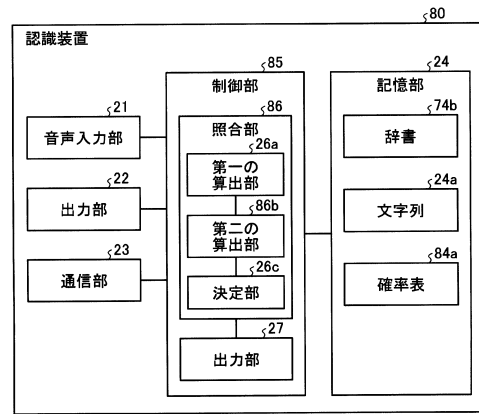
【図18】

辞書の一例を示す図

表記	読み情報	位置情報	品詞番号
決勝	けっしょう	0	1
に	に	3	9
決勝に	けっしょうに	4	3
海山学院	うみやまがくいん	6	1
に海山学院	にうみやまがくいん	7	7
決勝に海山学院	けっしょうにうみやまがくいん	8	1
が	が	9	9
海山学院が	うみやまがくいんが	10	3
に海山学院が	にうみやまがくいんが	11	9
進出	しんしゅつ	12	1
が進出	がしんしゅつ	13	7
海山学院が進出	うみやまがくいんがしんしゅつ	14	1
山川	やまかすわ	18	1
総理	そうり	21	1
山川総理	やまかわそうり	22	1
が	が	24	9
総理が	そうりが	25	3
山川総理が	やまかわそうりが	26	3
法案	ほうあん	27	1
が法案	がほうあん	28	7
総理が法案	そうりがほうあん	29	1
を	を	30	9
法案を	ほうあんを	31	3
が法案を	がほうあんを	32	9
通	とおし	33	5
を通	とおしを	34	8
法案を通	ほうあんをとおし	35	1
し	し	36	5
通し	とおし	37	5
を通し	をとおし	38	8
たい	たい	39	5
したい	したい	40	5
通したい	とおしたい	41	5
と	と	42	9
たいと	たいと	43	5
したいと	したいと	44	5
言及	げんきぎゅう	45	1
と言及	とげんきぎゅう	46	4
たいと言及	たいとげんきぎゅう	47	4

【図19】

実施例4に係る認識装置の構成の一例を示す図



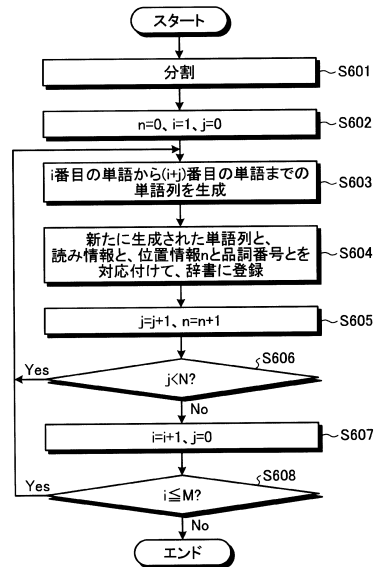
【図20】

確率表の一例を示す図

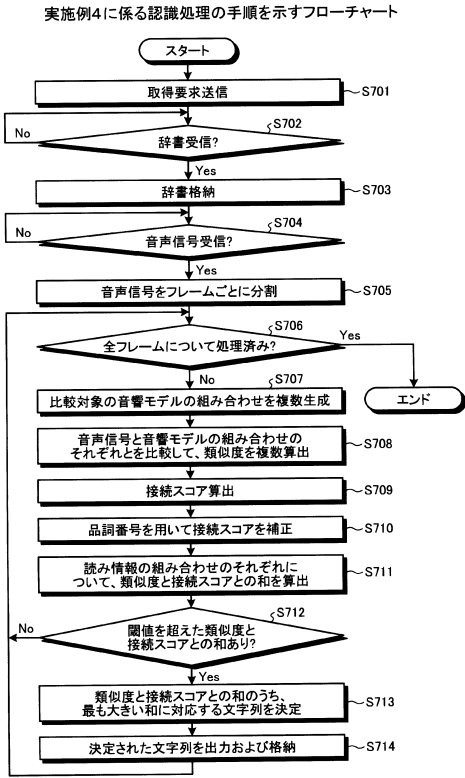
	1	2	3	4	5	6	7	8	9
1	100%	100%	100%	50%	50%	50%	100%	100%	100%
2	100%	100%	100%	100%	100%	100%	0%	0%	0%
3	100%	100%	100%	100%	100%	100%	0%	0%	0%
4	100%	100%	100%	50%	50%	50%	100%	100%	100%
5	100%	100%	100%	100%	100%	100%	0%	0%	0%
6	100%	100%	100%	100%	100%	100%	0%	0%	0%
7	100%	100%	100%	50%	50%	100%	100%	100%	100%
8	100%	100%	100%	100%	100%	100%	0%	0%	0%
9	100%	100%	100%	100%	100%	100%	0%	0%	0%
先頭	100%	100%	100%	0%	0%	0%	0%	0%	0%
末尾	100%	100%	0%	100%	100%	0%	100%	100%	0%

【図21】

実施例4に係る生成処理の手順を示すフローチャート

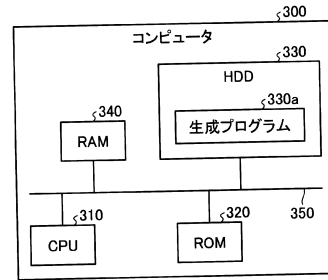


【図22】



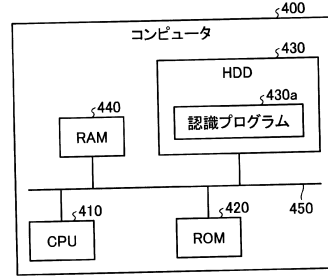
【図23】

生成プログラムを実行するコンピュータを示す図



【図24】

認識プログラムを実行するコンピュータを示す図



---

フロントページの続き

- (56)参考文献 特開2001-343994(JP,A)  
特開2005-227686(JP,A)  
特開2008-181537(JP,A)  
特開平10-097286(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 15/18  
G10L 15/06