

(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2014-194480

(P2014-194480A)

(43) 公開日 平成26年10月9日(2014.10.9)

(51) Int.Cl.
G10L 15/22 (2006.01)

F1
G10L 15/22 460D

テーマコード (参考)

審査請求 未請求 請求項の数 8 O L (全 20 頁)

(21) 出願番号 特願2013-70682(P2013-70682)
(22) 出願日 平成25年3月28日(2013.3.28)

(71) 出願人 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 外川 太郎
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
(72) 発明者 石川 千里
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
(72) 発明者 大谷 猛
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

(54) 【発明の名称】 音声処理装置、音声処理システムおよび音声処理方法

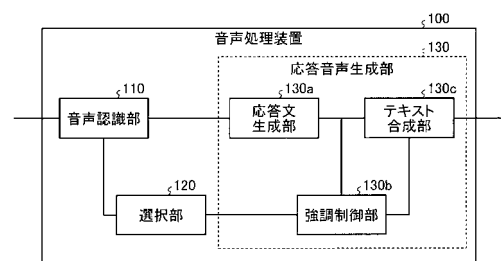
(57) 【要約】

【課題】 音声認識結果の間違いを判りやすくすること。

【解決手段】 音声処理装置100は、音声認識部110、選択部120、応答音声生成部130を有する。音声認識部110は、音声から複数の単語候補を抽出する。選択部120は、音声認識部110によって認識された複数の単語候補を基にして、各単語候補間で相違する相違箇所を選択する。応答音声生成部130は、単語候補のうち、相違箇所の音の大きさを増加させた応答音声出力する。

【選択図】 図1

本実施例1に係る音声認識装置の構成を示す図



【特許請求の範囲】**【請求項 1】**

入力された音声を単語に変換する音声認識を行い、前記音声に対応する複数の単語候補を抽出する音声認識部と、

前記音声認識部によって抽出された複数の単語候補を基にして、各単語候補間で相違する箇所を選択する選択部と、

前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、前記選択部によって選択された相違する箇所を強調して、前記単語候補の応答音声を出力する出力部と

を有することを特徴とする音声処理装置。

10

【請求項 2】

前記音声認識部は、前記音声に対応する複数の単語候補毎に信頼度を算出し、前記選択部は、信頼度が閾値以上となる複数の単語候補を基にして、各単語候補のうち相違する箇所を選択することを特徴とする請求項 1 に記載の音声処理装置。

【請求項 3】

前記音声認識部は、前記音声に対する複数の単語候補毎に信頼度を算出し、前記選択部は、最大の信頼度となる単語候補と、最大の信頼度との差が閾値未満となる単語候補とを基にして、各単語候補間で相違する箇所を選択することを特徴とする請求項 1 に記載の音声処理装置。

【請求項 4】

20

前記出力部は、前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、前記選択部によって選択された相違する箇所の強度を、選択されていない箇所の強度よりも大きくして、前記単語候補の応答音声を出力することを特徴とする請求項 1、2 または 3 に記載の音声処理装置。

【請求項 5】

前記出力部は、前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、前記選択部によって選択された相違する箇所の発話速度を、選択されていない箇所の発話速度よりも遅くして、前記単語候補の応答音声を出力することを特徴とする請求項 1、2 または 3 に記載の音声処理装置。

【請求項 6】

30

前記出力部は、前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、前記選択部によって選択された相違する箇所の音声の基本周期を長く、または短くして、前記単語候補の応答音声を出力することを特徴とする請求項 1、2 または 3 に記載の音声処理装置。

【請求項 7】

端末装置とサーバとを備えた音声処理システムであって、

前記端末装置は、入力された音声の情報を前記サーバに送信し、

前記サーバは、

前記端末装置から音声の情報を受信し、該音声を単語に変換する音声認識を行い、前記音声に対応する複数の単語候補を抽出する音声認識部と、

40

前記音声認識部によって抽出された複数の単語候補を基にして、各単語候補間で相違する箇所を選択する選択部と、

前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、前記選択部によって選択された相違する箇所を強調する前記単語候補の応答音声の情報を前記端末装置に送信する送信部と

を有することを特徴とする音声処理システム。

【請求項 8】

コンピュータが実行する音声処理方法であって、

入力された音声を単語に変換する音声認識を行い、前記音声に対応する複数の単語候補を抽出し、

50

抽出した複数の単語候補を基にして、各単語候補間で相違する箇所を選択し、
複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、
相違する箇所を強調して、前記単語候補の応答音声を出力する
各処理を実行することを特徴とする音声処理方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声処理装置等に関する。

【背景技術】

【0002】

ユーザとの間で音声による対話を繰り返して、情報検索をはじめとする各種タスクを実行する音声対話システムが存在する。この音声対話システムは、ユーザから入力される音声を単語に変換する音声認識技術を用いる。現状の音声対話システムでは、音声認識結果が正しいか否かを独自に判定できない。このため、音声対話システムは、音声認識結果をディスプレイ等に表示して音声認識結果が正しいか否かをユーザに確認する。

【0003】

ここで、音声対話システムが、音声認識結果の確認を頻繁にユーザに対して実行すると、ユーザに対する負担が増加するため、音声認識結果の確認を効率良く実行することが求められる。

【0004】

例えば、音声認識の信頼度が低い単語全体をゆっくり再生して、音声認識結果の確認をユーザに対して実行する従来技術がある。例えば、ユーザが「岡山県の天気は？」と発音し、音声対話システムが「和歌山県の天気は？」と音声認識し、「和歌山県」の信頼度が低くなった場合に、音声認識結果のうち「和歌山県」をゆっくり再生して、音声認識結果が正しいか否かをユーザに確認する。

【先行技術文献】

【特許文献】

【0005】

【特許文献1】特開2003-208196号公報

【特許文献2】特開2006-133478号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

しかしながら、上述した従来技術では、音声認識結果の間違いが判りづらいという問題がある。

【0007】

従来技術のように、音声認識の信頼度が低い単語全体をゆっくり再生すると、正しい認識結果との区別がつきにくく、ユーザが誤認識であるか否かを判断できない場合がある。例えば、上記の例に関して、音声対話システムが、信頼度の低い「和歌山県」をゆっくり再生したとしても、単語全体で聞いた場合には「和歌山県」と「岡山県」とが似ており、ユーザが違いを判別できない場合がある。

【0008】

一つの側面では、上記に鑑みてなされたものであって、音声認識結果の間違いを判りやすくする音声処理装置、音声処理システムおよび音声処理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】

【0009】

一つの案では、音声処理装置は、音声認識部と、選択部と、出力部とを有する。音声認識部は、入力された音声を単語に変換する音声認識を行い、音声に対応する複数の単語候補を抽出する。選択部は、音声認識部によって抽出された複数の単語候補を基にして、各

10

20

30

40

50

単語候補間で相違する箇所を選択する。出力部は、複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、選択部によって選択された相違する箇所を強調して、前記単語候補の応答音声を出力する。

【発明の効果】

【0010】

開示の態様では、音声認識結果の間違いを判りやすくすることができるという効果を奏する。

【図面の簡単な説明】

【0011】

【図1】図1は、本実施例1に係る音声認識装置の構成を示す図である。 10

【図2】図2は、本実施例1に係る選択部の構成を示す図である。

【図3】図3は、本実施例1に係る有力候補抽出部の処理を説明するための図である。

【図4】図4は、本実施例1に係る評価部の処理を説明するための図(1)である。

【図5】図5は、本実施例1に係る評価部の処理を説明するための図(2)である。

【図6】図6は、本実施例1に係る評価部の処理を説明するための図(3)である。

【図7】図7は、本実施例1に係る強調制御部の構成を示す図である。

【図8】図8は、本実施例1に係るモーラ位置照合部の処理を説明するための図である。

【図9】図9は、本実施例1に係る音声認識装置の処理手順を示すフローチャートである。

【図10】図10は、本実施例1に係る選択部の処理手順を示すフローチャートである。 20

【図11】図11は、本実施例2に係る音声認識装置の構成を示す図である。

【図12】図12は、本実施例2に係る選択部の構成を示す図である。

【図13】図13は、本実施例2に係る有力候補抽出部の処理を説明するための図である。

【図14】図14は、本実施例3に係る音声認識装置の構成を示す図である。

【図15】図15は、本実施例3に係る選択部の構成を示す図である。

【図16】図16は、本実施例3に係る有力候補抽出部が抽出した単語候補と信頼度との一例を示す図である。

【図17】図17は、本実施例3に係る評価部の処理を説明するための図(1)である。

【図18】図18は、本実施例3に係る評価部の処理を説明するための図(2)である。 30

【図19】図19は、本実施例3に係る評価部の処理を説明するための図(3)である。

【図20】図20は、本実施例3に係る強調制御部の構成を示す図である。

【図21】図21は、本実施例3に係るモーラ位置照合部の処理を説明するための図である。

【図22】図22は、本実施例4に係る音声処理システムの一部を示す図である。

【図23】図23は、本実施例4に係るサーバの構成を示す図である。

【図24】図24は、音声処理プログラムを実行するコンピュータの一部を示す図である。

【発明を実施するための形態】

【0012】 40

以下に、本願の開示する音声処理装置、音声処理システムおよび音声処理方法の実施例を図面に基づいて詳細に説明する。なお、この実施例によりこの発明が限定されるものではない。

【実施例1】

【0013】

本実施例1に係る音声処理装置について説明する。図1は、本実施例1に係る音声認識装置の構成を示す図である。図1に示すように、この音声処理装置100は、音声認識部110、選択部120、応答音声生成部130を有する。応答音声生成部130は、応答文生成部130a、強調制御部130b、テキスト合成部130cを有する。

【0014】 50

音声認識部 110 は、マイク等から入力される音声を単語に変換する音声認識を行い、音声に対応する複数の単語候補を抽出する処理部である。また、音声認識部 110 は、各単語候補の信頼度をそれぞれ求める。音声認識部 110 は、単語候補と信頼度とを対応付けた情報を、選択部 120 および応答文生成部 130 a に出力する。以下の説明において、マイク等から入力される音声を適宜、入力音声と表記する。

【0015】

音声認識部 110 の処理の一例について具体的に説明する。音声認識部 110 は、複数の単語と、該単語に対応する音声の標準パターンとを対応付けた標準テーブルを保持する。音声認識部 110 は、入力音声の周波数特性を基にして入力音声の特徴ベクトルを算出し、算出した特徴ベクトルと標準テーブルの各標準パターンとを比較して、特徴ベクトルと各標準パターンとの類似度をそれぞれ算出する。この特徴ベクトルと各標準パターンとの類似度を、上記信頼度とする。

10

【0016】

音声認識部 110 は、特徴ベクトルとの信頼度が限りなく 0 に近い標準パターン以外の標準パターンを、単語候補として抽出する。例えば、音声認識部 110 は、特徴ベクトルとの信頼度が 0.1 以上となる標準パターンを、単語候補として抽出する。音声認識部 110 は、抽出した単語候補と信頼度とを対応付けた情報を、選択部 120 および応答音声生成部 130 に出力する。

【0017】

音声認識部 110 が、信頼度を算出する処理は、上記のものに限られず、如何なる周知技術を用いても良い。例えば、音声認識部 110 は、特開平 4 - 255900 号公報の技術を用いて、単語候補の信頼度を算出しても良い。

20

【0018】

選択部 120 は、複数の単語候補間で相違する箇所を選択する処理部である。図 2 は、本実施例 1 に係る選択部の構成を示す図である。図 2 に示すように、選択部 120 は、有力候補抽出部 120 a と、評価部 120 b を有する。

【0019】

有力候補抽出部 120 a は、複数の単語候補の信頼度を基にして、信頼度が所定の閾値以上となる単語候補を抽出する。有力候補抽出部 120 a は、抽出した単語候補と該単語候補の信頼度との組みを、評価部 120 b に出力する。

30

【0020】

図 3 は、本実施例 1 に係る有力候補抽出部の処理を説明するための図である。例えば、音声認識部 110 から取得した単語候補と信頼度との関係が、図 3 に示す関係となっており、所定の閾値を「0.6」とする。この場合には、有力候補抽出部 120 a は、候補 1 ~ 3 の単語候補および信頼度の組みを抽出する。有力候補抽出部 120 a は、候補 1 ~ 3 の単語候補および信頼度の組みの情報を、評価部 120 b に出力する。

【0021】

評価部 120 b は、各単語候補をそれぞれ比較して、単語候補間で相違する箇所を選択する処理部である。以下の説明において、各単語候補のうち、信頼度が最大となるものを、第 1 単語候補と表記し、その他の単語候補を第 2 単語候補と表記する。例えば、図 3 に示す例では、信頼度「0.80」の単語候補「わかやま(和歌山)」が、第 1 単語候補となる。また、信頼度「0.75」の単語候補「おかやま(岡山)」および信頼度「0.65」の単語候補「とやま(富山)」が、それぞれ第 2 単語候補となる。

40

【0022】

評価部 120 b は、第 1 単語候補と各第 2 単語候補との一致スコアをそれぞれ算出し、算出した各一致スコアを合計することで、最終的な第 1 単語候補の一致スコアを算出する。例えば、評価部 120 b は、第 1 単語候補「わかやま(和歌山)」と、第 2 単語候補「おかやま(岡山)」とを比較して、一致スコアを算出する。また、評価部 120 b は、第 1 単語候補「わかやま(和歌山)」と、第 2 単語候補「とやま(富山)」とを比較して、一致スコアを算出する。そして、評価部 120 b は、各一致スコアを合計することで、第

50

1 単語候補の一致スコアを算出する。

【0023】

評価部120bは、例えば、DPマッチング法を用いて、一致スコアを算出する。図4、図5、図6は、本実施例1に係る評価部の処理を説明するための図である。図4について説明する。図4は、第1単語候補「わかやま(和歌山)」と、第2単語候補「おかやま(岡山)」とを比較する処理を示す。評価部120bは、第1単語候補および第2単語候補の文字をそれぞれ比較し、一致していれば、該当文字にスコア「0」を与え、一致しなければ、該当文字にスコア「-1」を与えるテーブル10aを生成する。

【0024】

評価部120bは、テーブル10aを基にして、第1単語候補の各文字に対応するスコアのうち、大きい方のスコアを通る経路を優先して選択することで、各文字のスコアを特定する。図4に示す例では、経路11aが選択され、第1単語候補の各文字のスコアは、スコアテーブル20aに示すものとなる。すなわち、「わ」に対するスコアが「-1」となる。「か」、「や」、「ま」に対するスコアがそれぞれ「0」となる。

【0025】

図5について説明する。図5は、第1単語候補「わかやま」と第2単語候補「とやま」とを比較する処理を示す。評価部120bは、第1単語候補および第2単語候補の文字をそれぞれ比較し、一致していれば、該当文字にスコア「0」を与え、一致しなければ、該当文字にスコア「-1」を与えるテーブル10bを生成する。

【0026】

評価部120bは、テーブル10bを基にして、第1単語候補の各文字に対応するスコアのうち、大きい方のスコアを通る経路を優先して選択することで、各文字のスコアを特定する。図5に示す例では、経路11bが選択され、第1単語候補の各文字のスコアは、スコアテーブル20bに示すものとなる。すなわち、「わ」および「か」に対するスコアがそれぞれ「-1」となる。また、「や」、「ま」に対するスコアがそれぞれ「0」となる。

【0027】

図6について説明する。評価部120bは、スコアテーブル20aとスコアテーブル20bとをそれぞれ第1単語候補の文字毎に合計することで、第1単語候補のスコアテーブル30を算出する。

【0028】

評価部120bは、スコアテーブル30を基にして、第1単語候補の相違箇所を選択する。例えば、評価部120bは、スコアテーブル30の各スコアのうち、スコアが「0」未満となるスコアを選択し、選択したスコアに対応する文字を、相違箇所として選択する。図6に示す例では、評価部120bは、第1単語候補「わかやま」のうち「わか」を相違箇所として選択する。選択部120bは、相違箇所の情報を、強調制御部130bに出力する。

【0029】

図1の説明に戻る。応答文生成部130aは、音声認識結果が正しいか否かをユーザに確認するための応答文を生成する処理部である。例えば、応答文生成部130aは、複数種類の文字列のテンプレートを保持しており、音声認識部110から得られる単語候補と、テンプレートとを組み合わせることで、応答文を生成する。応答文生成部130aは、生成した応答文の情報を、強調制御部130b、テキスト合成部130cに出力する。

【0030】

例えば、応答文生成部130aは、複数の単語候補を取得した場合には、信頼度が最大となる単語候補を選択して、応答文を生成する。例えば、信頼度が最大となる単語候補が「わかやま」の場合には、応答文生成部130aは、テンプレート「ですね」と組み合わせ、応答文「わかやまですね」を生成する。

【0031】

強調制御部130bは、応答文中の強調箇所を選択し、選択した強調箇所と、該強調箇

10

20

30

40

50

所を強調するためのパラメータを、テキスト合成部 130c に通知する処理部である。図 7 は、本実施例 1 に係る強調制御部の構成を示す図である。図 7 に示すように、この強調制御部 130b は、モーラ位置照合部 131 および強調パラメータ設定部 132 を有する。

【0032】

モーラ位置照合部 131 は、評価部 120b から取得した相違箇所の情報を基にして、応答文中の強調箇所を選択する処理部である。図 8 は、本実施例 1 に係るモーラ位置照合部の処理を説明するための図である。図 8 に示すように、モーラ位置照合部 131 は、応答文 40 における開始モーラ位置 40a と、単語候補 50 の相違箇所 50a とを照合することで、応答文 40 中の強調箇所を算出する。図 8 に示す例では、相違箇所 50a に対応する応答文 40 の文字は、1 文字目の「わ」と 2 文字目の「か」となる。このため、強調箇所は、1 ~ 2 モーラとなる。

10

【0033】

強調パラメータ設定部 132 は、予め設定された増幅量のパラメータを、テキスト合成部 130c に出力する。例えば、強調パラメータ設定部 132 は、「強調箇所を 10dB 増幅する」旨の情報を、テキスト合成部 130c に出力する。

【0034】

テキスト合成部 130c は、応答文の情報と、強調箇所の情報と、強調するためのパラメータとを基にして、応答文のうち強調箇所を強調した応答音声を生じ、生成した応答音声を出力する処理部である。例えば、テキスト合成部 130c は、応答文を言語解析して、各単語に対応した韻律を特定し、特定した各韻律を組み合わせることで、応答音声を生成する。テキスト合成部 130c は、応答音声のうち、強調箇所の文字に対応する音声の韻律を強調することで、強調箇所を強調した応答音声を生成する。

20

【0035】

例えば、強調箇所が「1 ~ 2 モーラ」で、パラメータが「強調箇所を 10dB 増幅する」であれば、テキスト合成部 130c は、応答文「わかやまです」の「わか」の音声のパワーを「10dB」増幅することで、応答音声を生成する。テキスト合成部 130c が生成した応答音声は、スピーカなどから出力される。例えば、応答文「わかやまです」のうち、「わか」の箇所が、他の文言よりも強調されて出力される。

【0036】

ところで、選択部 120 において、複数の単語候補が抽出されなかった場合には、応答音声生成部 130 は、応答文の情報をそのまま応答音声に変換し、出力する。

30

【0037】

次に、本実施例 1 に係る音声処理装置 100 の処理手順について説明する。図 9 は、本実施例 1 に係る音声認識装置の処理手順を示すフローチャートである。図 9 に示す処理は、例えば、音声入力を受け付けたことを契機として実行される。図 9 に示すように、音声処理装置 100 は、音声入力を受け付け（ステップ S101）、音声認識を実行して単語候補を抽出する（ステップ S102）。

【0038】

音声処理装置 100 は、単語候補の信頼度を算出し（ステップ S103）、信頼度が所定値以上の単語候補を選択する（ステップ S104）。音声処理装置 100 は、応答文を生成し（ステップ S105）、相違箇所を選択する（ステップ S106）。

40

【0039】

音声処理装置 100 は、パラメータの設定を行い（ステップ S107）、言語解析を実行する（ステップ S108）。音声処理装置 100 は、韻律生成を行い（ステップ S109）、強調箇所の韻律を変更する（ステップ S110）。音声処理装置 100 は、波形処理を行い（ステップ S111）、応答音声を出力する（ステップ S112）。

【0040】

次に、図 1 に示した選択部 120 の処理手順の一例について説明する。図 10 は、本実施例 1 に係る選択部の処理手順を示すフローチャートである。選択部 120 は、複数の単

50

語候補から、所定値以上の信頼度の単語候補を抽出する（ステップS201）。

【0041】

選択部120は、単語候補が2つ以上であるか否かを判定する（ステップS202）。選択部120は、単語候補が2つ以上でない場合には（ステップS202, No）、相違箇所無しと判定する（ステップS203）。

【0042】

一方、選択部120は、単語候補が2つ以上である場合には（ステップS202, Yes）、第1単語候補に対する第2単語候補の一致度スコアを算出する（ステップS204）。選択部120は、各単語候補の合計のスコアを算出する（ステップS205）。選択部120は、合計のスコアが低い箇所を相違箇所として選択する（ステップS206）。 10

【0043】

次に、本実施例1に係る音声処理装置100の効果について説明する。音声処理装置100は、音声認識部110によって認識された複数の単語候補を基にして、各単語候補間で相違する相違箇所を選択する。音声処理装置100は、単語候補のうち、相違箇所の音の大きさを増加させた応答音声を入力する。このように、本実施例1に係る音声処理装置100によれば、単語全体ではなく、相違箇所のみを強調して出力するので、音声認識結果の間違いが判りやすくすることができる。また、この技術を音声対話システムに適用することで、誤認識した箇所にユーザが気づいて訂正の発話を行いやすくなり、誤認識時の対話を効率化することもできる。 20

【実施例2】

【0044】

本実施例2に係る音声処理装置について説明する。図11は、本実施例2に係る音声認識装置の構成を示す図である。図11に示すように、この音声処理装置200は、音声認識部210、選択部220、応答音声生成部230を有する。応答音声生成部230は、応答文生成部230a、強調制御部230b、テキスト合成部230cを有する。

【0045】

音声認識部210は、マイク等から入力される音声を単語に変換する音声認識を行い、音声に対応する複数の単語候補を抽出する処理部である。また、音声認識部210は、各単語候補の信頼度をそれぞれ求める。音声認識部210は、単語候補と信頼度とを対応付けた情報を、選択部220および応答音声生成部230に出力する。音声認識部210に関する具体的な説明は、実施例1で説明した音声認識部110と同様である。 30

【0046】

選択部220は、複数の単語候補間で相違する箇所を選択する処理部である。図12は、本実施例2に係る選択部の構成を示す図である。図12に示すように、選択部220は、有力候補抽出部220aと、評価部220bとを有する。

【0047】

有力候補抽出部220aは、複数の単語候補の信頼度を基にして、最も高い信頼度との差が所定の閾値以下となる単語候補を抽出する。有力候補抽出部220aは、抽出した単語候補と該単語候補の信頼度との組みを、評価部220bに出力する。

【0048】

図13は、本実施例2に係る有力候補抽出部の処理を説明するための図である。図13に示す例では、候補と、単語候補と、信頼度と、信頼度の差とを対応付けて示している。信頼度の差は、最も高い信頼度と該当する信頼度との差を示す。所定の閾値を「0.2」とすると、信頼度の差が閾値以下となる単語候補は、候補1~3となる。このため、有力候補抽出部220aは、候補1~3の単語候補および信頼度の組みの情報を、評価部220bに出力する。 40

【0049】

評価部220bは、各単語候補をそれぞれ比較して、単語候補間で相違する箇所を選択する処理部である。実施例1と同様にして、各単語候補のうち、信頼度が最大となるものを、第1単語候補と表記し、その他の単語候補を第2単語候補と表記する。評価部220 50

b は、実施例 1 の評価部 1 2 0 b と同様の処理を実行して、相違箇所を選択し、選択した相違箇所の情報を、強調制御部 2 3 0 b に出力する。

【 0 0 5 0 】

応答文生成部 2 3 0 a は、音声認識結果が正しいか否かをユーザに確認するための応答文を生成する処理部である。応答文生成部 2 3 0 a が応答文を生成する処理は、実施例 1 の応答文生成部 1 3 0 a と同様である。応答文生成部 2 3 0 a は、生成した応答文の情報を、強調制御部 2 3 0 b、テキスト合成部 2 3 0 c に出力する。

【 0 0 5 1 】

強調制御部 2 3 0 b は、応答文中の強調箇所を選択し、選択した強調箇所と、該強調箇所を強調するためのパラメータを、テキスト合成部 2 3 0 c に通知する処理部である。強調制御部 2 3 0 b は、実施例 1 の強調制御部 1 3 0 b と同様にして、強調箇所を特定する。なお、強調制御部 2 3 0 b は、パラメータとして「強調箇所の持続長を 2 倍にする」旨の情報を、テキスト合成部 2 3 0 c に出力する。

10

【 0 0 5 2 】

テキスト合成部 2 3 0 c は、応答文の情報と、強調箇所の情報と、強調するためのパラメータとを基にして、応答文のうち強調箇所を強調した応答音声を生じ、生成した応答音声を出力する処理部である。例えば、テキスト合成部 2 3 0 c は、応答文を言語解析して、各単語に対応した韻律を特定し、特定した各韻律を組み合わせることで、応答音声を生成する。テキスト合成部 2 3 0 c は、応答音声のうち、強調箇所の文字に対応する音声の韻律を強調することで、強調箇所を強調した応答音声を生成する。

20

【 0 0 5 3 】

例えば、強調箇所が「1～2 モーラ」で、パラメータが「強調箇所の持続長を 2 倍にする」であれば、テキスト合成部 2 3 0 c は、応答文「わかやまです」の「わか」の韻律部分の持続長を 2 倍にすることで、応答音声を生成する。テキスト合成部 2 3 0 c が生成した応答音声は、スピーカなどから出力される。例えば、応答文「わかやまです」のうち、「わか」の箇所が、他の文言よりも長めに強調されて出力される。

【 0 0 5 4 】

次に、本実施例 2 に係る音声処理装置 2 0 0 の効果について説明する。音声処理装置 2 0 0 は、音声認識部 2 1 0 によって認識された複数の単語候補を基にして、各単語候補間で相違する相違箇所を選択する。音声処理装置 2 0 0 は、単語候補のうち、相違箇所の持続長を長めにした応答音声を出力する。このように、本実施例 2 に係る音声処理装置 2 0 0 によれば、単語全体ではなく、相違箇所のみを長めにして出力するので、音声認識結果の間違いが判りやすくすることができる。また、この技術を音声対話システムに適用することで、誤認識した箇所にユーザが気づいて訂正の発話を行いやすくなり、誤認識時の対話を効率化することもできる。

30

【 0 0 5 5 】

ところで、本実施例 2 に示した音声処理装置 2 0 0 は、パラメータとして「強調箇所のピッチを 2 倍にする」旨の情報をを用いて、相違箇所を強調しても良い。ピッチは例えば、基本周期に対応する。例えば、強調箇所が「1～2 モーラ」で、パラメータが「強調箇所のピッチを 2 倍にする」であれば、テキスト合成部 2 3 0 c は、応答文「わかやまです」の「わか」の韻律部分のピッチ 2 倍にすることで、強調箇所の声を通常時よりも低くした応答音声を生成する。このように、本実施例 2 に係る音声処理装置 2 0 0 によれば、相違箇所のみを声の高さを低くして出力するので、音声認識結果の間違いが判りやすくすることができる。なお、音声処理装置 2 0 0 は、ピッチを 1 / 2 にして強調しても良い。

40

【 実施例 3 】

【 0 0 5 6 】

本実施例 3 に係る音声処理装置について説明する。図 1 4 は、本実施例 3 に係る音声認識装置の構成を示す図である。図 1 4 に示すように、この音声処理装置 3 0 0 は、音声認識部 3 1 0、選択部 3 2 0、応答音声生成部 3 3 0 を有する。応答音声生成部 3 3 0 は、応答文生成部 3 3 0 a、強調制御部 3 3 0 b、テキスト合成部 3 3 0 c を有する。

50

【 0 0 5 7 】

音声認識部 3 1 0 は、マイク等から入力される音声を単語に変換する音声認識を行い、音声に対応する複数の単語候補を抽出する処理部である。また、音声認識部 3 1 0 は、各単語候補の信頼度をそれぞれ求める。音声認識部 3 1 0 は、単語候補と信頼度とを対応付けた情報を、選択部 3 2 0 および応答文生成部 3 3 0 a に出力する。以下の説明において、マイク等から入力される音声を適宜、入力音声と表記する。

【 0 0 5 8 】

音声認識部 3 1 0 の処理の一例について具体的に説明する。音声認識部 3 1 0 は、複数の単語と、該単語に対応する音声の標準パターンとを対応付けた標準テーブルを保持する。音声認識部 3 1 0 は、入力音声の周波数特性を基にして入力音声の特徴ベクトルを算出し、算出した特徴ベクトルと標準テーブルの各標準パターンとを比較して、特徴ベクトルと各標準パターンとの類似度をそれぞれ算出する。この特徴ベクトルと各標準パターンとの類似度を、上記信頼度とする。

10

【 0 0 5 9 】

音声認識部 3 1 0 は、特徴ベクトルとの信頼度が限りなく 0 に近い標準パターン以外の標準パターンを、単語候補として抽出する。例えば、音声認識部 3 1 0 は、特徴ベクトルとの信頼度が 0.1 以上となる標準パターンを、単語候補として抽出する。音声認識部 3 1 0 は、抽出した単語候補と信頼度とを対応付けた情報を、選択部 3 2 0 および応答音声生成部 3 3 0 に出力する。

【 0 0 6 0 】

選択部 3 2 0 は、複数の単語候補間で相違する箇所を選択する処理部である。図 1 5 は、本実施例 3 に係る選択部の構成を示す図である。図 1 5 に示すように、選択部 3 2 0 は、有力候補抽出部 3 2 0 a と、評価部 3 2 0 b を有する。

20

【 0 0 6 1 】

有力候補抽出部 3 2 0 a は、複数の単語候補の信頼度を基にして、信頼度が所定の閾値以上となる単語候補を抽出する。有力候補抽出部 3 2 0 a は、抽出した単語候補と該単語候補の信頼度との組みを、評価部 3 2 0 b に出力する。各単語候補のうち、信頼度が最大となるものを、第 1 単語候補と表記し、その他の単語候補を第 2 単語候補と表記する。

【 0 0 6 2 】

図 1 6 は、本実施例 3 に係る有力候補抽出部が抽出した単語候補と信頼度との一例を示す図である。なお、図 1 6 に示すように、第 1 単語候補の「seven」の音節は「sev・en」となる。第 2 単語候補の「eleven」の音節は「e・lev・en」となる。第 2 単語候補の「seventeen」の音節は「sev・en・teen」となる。

30

【 0 0 6 3 】

評価部 3 2 0 b は、第 1 単語候補と各第 2 単語候補との一致スコアをそれぞれ算出し、算出した各一致スコアを合計することで、最終的な第 1 単語候補の一致スコアを算出する。例えば、評価部 3 2 0 b は、第 1 単語候補「seven」と、第 2 単語候補「eleven」とを比較して、一致スコアを算出する。また、評価部 3 2 0 b は、第 1 単語候補「seven」と、第 2 単語候補「seventeen」とを比較して、一致スコアを算出する。評価部 3 2 0 b は、各一致スコアを合計することで、第 1 単語候補の一致スコアを算出する。

40

【 0 0 6 4 】

評価部 3 2 0 b は、例えば、DP マッチング法を用いて、一致スコアを算出する。図 1 7、図 1 8、図 1 9 は、本実施例 3 に係る評価部の処理を説明するための図である。図 1 7 について説明する。図 1 7 は、第 1 単語候補「seven」と、第 2 単語候補「eleven」とを比較する処理を示す。評価部 3 2 0 b は、第 1 単語候補および第 2 単語候補の文字をそれぞれ比較し、一致していれば、該当文字にスコア「0」を与え、一致しなければ、該当文字にスコア「-1」を与えるテーブル 1 0 c を生成する。

【 0 0 6 5 】

評価部 3 2 0 b は、テーブル 1 0 c を基にして、第 1 単語候補の各文字に対応するスコアのうち、大きい方のスコアを通る経路を優先して選択することで、各文字のスコアを特

50

定する。図 17 に示す例では、経路 11c が選択され、第 1 単語候補の各文字のスコアは、スコアテーブル 20c に示すものとなる。すなわち、「s」に対するスコアが「-1」となる。「e」、「v」、「e」、「n」に対するスコアがそれぞれ「0」となる。

【0066】

図 18 について説明する。図 18 は、第 1 単語候補「seven」と、第 2 単語候補「seventeen」とを比較する処理を示す。評価部 320b は、第 1 単語候補および第 2 単語候補の文字をそれぞれ比較し、一致していれば、該当文字にスコア「0」を与え、一致しなければ、該当文字にスコア「-1」を与えるテーブル 10d を生成する。なお、第 1 単語候補の文字数が、第 2 単語候補の文字数未満の場合には、評価部 320b は、第 2 単語候補のうち第 1 単語候補の文字数分だけ、比較対象とする。例えば、第 1 単語候補「seven」と、第 2 単語候補「seventeen」とを比較する場合には、第 2 単語候補の「seventeen」のうち、「seven」を比較対象とする。

10

【0067】

評価部 320b は、テーブル 10d を基にして、第 1 単語候補の各文字に対応するスコアのうち、大きい方のスコアを通る経路を優先して選択することで、各文字のスコアを特定する。図 18 に示す例では、経路 11d が選択され、第 1 単語候補の各文字のスコアは、スコアテーブル 20d に示すものとなる。「s」、「e」、「v」、「e」、「n」に対するスコアがそれぞれ「0」となる。

【0068】

図 19 について説明する。評価部 320b は、スコアテーブル 20c とスコアテーブル 20d とをそれぞれ第 1 単語候補の文字毎に合計することで、第 1 単語候補のスコアテーブル 35 を算出する。

20

【0069】

評価部 320b は、スコアテーブル 35 を基にして、第 1 単語候補の相違箇所を選択する。例えば、評価部 320b は、スコアテーブル 35 の各スコアのうち、スコアが「0」未満となるスコアを選択し、選択したスコアに対応する文字を、相違箇所として選択する。図 19 に示す例では、評価部 320b は、第 1 単語候補「seven」のうち「s」を相違箇所として選択する。選択部 320b は、相違箇所の情報を、強調制御部 330b に出力する。

【0070】

図 14 の説明に戻る。応答文生成部 330a は、音声認識結果が正しいか否かをユーザに確認するための応答文を生成する処理部である。例えば、応答文生成部 330a は、複数種類の文字列のテンプレートを保持しており、音声認識部 310 から得られる単語候補と、テンプレートとを組み合わせることで、応答文を生成する。応答文生成部 330a は、生成した応答文の情報を、強調制御部 330b、テキスト合成部 330c に出力する。

30

【0071】

例えば、応答文生成部 330a は、複数の単語候補を取得した場合には、信頼度が最大となる単語候補を選択して、応答文を生成する。例えば、信頼度が最大となる単語候補が「seven」の場合には、応答文生成部 330a は、テンプレート「o'clock?」と組み合わせ、応答文「Seven o'clock?」を生成する。

40

【0072】

強調制御部 330b は、応答文中の強調箇所を選択し、選択した強調箇所と、該強調箇所を強調するためのパラメータを、テキスト合成部 330c に通知する処理部である。図 20 は、本実施例 3 に係る強調制御部の構成を示す図である。図 20 に示すように、この強調制御部 330b は、モーラ位置照合部 331 および強調パラメータ設定部 332 を有する。

【0073】

モーラ位置照合部 331 は、評価部 320b から取得した相違箇所の情報を基にして、応答文中の強調箇所を選択する処理部である。図 21 は、本実施例 3 に係るモーラ位置照合部の処理を説明するための図である。図 21 に示すように、モーラ位置照合部 331 は

50

、応答文 4 5 における開始モーラ位置 4 5 a と、単語候補 5 5 の相違箇所 5 5 a とを照合することで、応答文 4 5 中の強調箇所を算出する。図 2 1 に示す例では、相違箇所 5 0 a に対応する応答文 4 5 の文字は、1 文字目の「S」となる。このため、強調箇所は、1 モーラとなる。なお、モーラ位置照合部 3 3 1 は、音節単位で、強調箇所を特定しても良い。例えば、1 文字目の「s」は、音節「Sev」に含まれるので、モーラ位置照合部 3 3 1 は、「Sev」を強調箇所としても良い。この場合には、強調箇所は 1 ~ 3 モーラとなる。

【0074】

強調パラメータ設定部 3 3 2 は、予め設定された増幅量のパラメータを、テキスト合成部 3 3 0 c に出力する。例えば、強調パラメータ設定部 3 3 2 は、「強調箇所を 1 0 d B 増幅する」旨の情報を、テキスト合成部 3 3 0 c に出力する。

10

【0075】

テキスト合成部 3 3 0 c は、応答文の情報と、強調箇所の情報と、強調するためのパラメータとを基にして、応答文のうち強調箇所を強調した応答音声を生成し、生成した応答音声を出力する処理部である。例えば、テキスト合成部 3 3 0 c は、応答文を言語解析して、各単語に対応した韻律を特定し、特定した各韻律を組み合わせることで、応答音声を生成する。テキスト合成部 3 3 0 c は、応答音声のうち、強調箇所の文字に対応する音声の韻律を強調することで、強調箇所を強調した応答音声を生成する。

【0076】

例えば、強調箇所が「1 ~ 3 モーラ」で、パラメータが「強調箇所を 1 0 d B 増幅する」であれば、テキスト合成部 3 3 0 c は、応答文「Seven o'clock?」の「Sev」の音声のパワーを「1 0 d B」増幅することで、応答音声を生成する。テキスト合成部 3 3 0 c が生成した応答音声は、スピーカなどから出力される。例えば、応答文「Seven o'clock?」のうち、「Sev」の箇所が、他の文言よりも強調されて出力される。

20

【0077】

強調するパラメータは上記のものに限られない。例えば、パラメータが「強調箇所の持続長を 2 倍にする」であれば、テキスト合成部 3 3 0 は、応答文「Seven o'clock?」の「Sev」の韻律部分の持続長を 2 倍にすることで、応答音声を生成する。例えば、パラメータが「強調箇所のピッチを 2 倍にする」であれば、テキスト合成部 3 3 0 は、応答文「Seven o'clock?」の「Sev」の韻律部分のピッチ 2 倍にすることで、強調箇所の声を通常時よりも低くした応答音声を生成する。

30

【0078】

次に、本実施例 3 に係る音声処理装置 3 0 0 の効果について説明する。音声処理装置 3 0 0 は、音声認識部 3 1 0 によって認識された複数の単語候補を基にして、各単語候補間で相違する相違箇所を選択する。音声処理装置 3 0 0 は、単語候補のうち、相違箇所の音の大きさを増加させた応答音声を出力する。このように、本実施例 3 に係る音声処理装置 3 0 0 によれば、単語全体ではなく、相違箇所のみを強調して出力するので、音声認識結果の間違いが判りやすくすることができる。また、この技術を音声対話システムに適用することで、誤認識した箇所にユーザが気づいて訂正の発話を行いやすくなり、誤認識時の対話を効率化することもできる。

【実施例 4】

40

【0079】

本実施例 4 に係る音声処理システムについて説明する。図 2 2 は、本実施例 4 に係る音声処理システムの一例を示す図である。図 2 2 に示すように、この音声処理システムは、端末装置 4 0 0 とサーバ 5 0 0 とを有する。端末装置 4 0 0 およびサーバ 5 0 0 はネットワーク 8 0 を介して相互に接続される。

【0080】

端末装置 4 0 0 は、マイク等を用いてユーザから音声の入力を受け付け、受け付けた音声の情報をサーバ 5 0 0 に送信する。また、端末装置 4 0 0 は、サーバ 5 0 0 から応答音声の情報を受信し、受信した応答音声をスピーカ等から出力する。

【0081】

50

サーバ500は、実施例1～3に示した音声処理装置と同様の機能を有する。図23は、本実施例4に係るサーバの構成を示す図である。図23に示すように、このサーバ500は、通信制御部500aおよび音声処理部500bを有する。音声処理部500bは、音声認識部510、選択部520、応答音声生成部530を有する。応答音声生成部530は、応答文生成部530a、強調制御部530b、テキスト合成部530cを有する。

【0082】

通信制御部500aは、端末装置400とデータ通信を実行する処理部である。通信制御部500aは、端末装置から受信した音声の情報を、音声認識部510に出力する。また、通信制御部500aは、テキスト合成部530cから出力される応答音声の情報を、端末装置400に送信する。

10

【0083】

音声認識部510は、通信制御部500aから音声の情報を取得し、音声を単語に変換する音声認識を行い、音声に対応する複数の単語候補を抽出する処理部である。また、音声認識部510は、各単語候補の信頼度をそれぞれ求める。音声認識部510は、単語候補と信頼度とを対応付けた情報を、選択部520および応答文生成部530aに出力する。

【0084】

選択部520は、複数の単語候補間で相違する箇所を選択する処理部である。選択部520に関する具体的な説明は、実施例1～4に示した選択部120, 220, 230と同様である。

20

【0085】

応答文生成部530aは、音声認識結果が正しいか否かをユーザに確認するための応答文を生成する処理部である。応答文生成部530aが応答文を生成する処理は、実施例1の応答文生成部130aと同様である。応答文生成部530aは、生成した応答文の情報を、強調制御部530b、テキスト合成部530cに出力する。

【0086】

強調制御部530bは、応答文中の強調箇所を選択し、選択した強調箇所と、該強調箇所を強調するためのパラメータを、テキスト合成部530cに通知する処理部である。強調制御部530bは、実施例1の強調制御部130bと同様にして、強調箇所を特定する。なお、強調制御部530bは、パラメータとして「強調箇所の持続長を2倍にする」旨の情報を、テキスト合成部530cに出力する。なお、強調制御部530bは、パラメータとして「強調箇所を10dB増幅する」旨の情報を、テキスト合成部530cに出力する。なお、パラメータは、実施例2と同様に、「強調箇所の持続長を2倍にする」または「強調箇所のピッチを2倍にする」旨の情報であってもよい。

30

【0087】

テキスト合成部530cは、応答文の情報と、強調箇所の情報と、強調するためのパラメータとを基にして、応答文のうち強調箇所を強調した応答音声を生成し、生成した応答音声を出力する処理部である。例えば、テキスト合成部530cは、応答文を言語解析して、各単語に対応した韻律を特定し、特定した各韻律を組み合わせることで、応答音声を生成する。テキスト合成部530cは、応答音声のうち、強調箇所の文字に対応する音声の韻律を強調することで、強調箇所を強調した応答音声を生成する。テキスト合成部530cは、生成した応答音声の情報を、通信制御部500aに出力する。

40

【0088】

次に、本実施例4に係るサーバ500の効果について説明する。サーバ500は、音声認識部510によって認識された複数の単語候補を基にして、各単語候補間で相違する相違箇所を選択する。サーバ500は、単語候補のうち、相違箇所の音の大きさを増加させた応答音声を出力する。このように、本実施例4に係るサーバ500によれば、単語全体ではなく、相違箇所のみを強調して出力するので、音声認識結果の間違いが判りやすくすることができる。また、この技術を音声対話システムに適用することで、誤認識した箇所にユーザが気づいて訂正の発話を行いやすくなり、誤認識時の対話を効率化することでも

50

きる。

【0089】

次に、上記の実施例に示した音声処理装置と同様の機能を実現する音声処理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、音声処理プログラムを実行するコンピュータの一例を示す図である。

【0090】

図24に示すように、コンピュータ600は、各種演算処理を実行するCPU601と、ユーザからのデータの入力を受け付ける入力装置602と、ディスプレイ603を有する。また、コンピュータ600は、記憶媒体からプログラム等を読み取る読み取り装置604と、ネットワークを介して他のコンピュータとの間でデータの授受を行うインターフェース装置605とを有する。また、コンピュータ600は、各種情報を一時記憶するRAM606と、ハードディスク装置607を有する。そして、各装置601～607は、バス608に接続される。

10

【0091】

ハードディスク装置607は、例えば、音声認識プログラム607a、選択プログラム607b、出力プログラム607cを有する。CPU601は、各プログラム607a～607cを読み出してRAM606に展開する。

【0092】

音声認識プログラム607aは、音声認識プロセス606aとして機能する。選択プログラム607bは、選択プロセス606bとして機能する。出力プログラム607cは、出力プロセス606cとして機能する。

20

【0093】

例えば、音声認識プロセス606aは、音声認識部110, 210, 310, 510に対応する。選択プロセス606bは、選択部120, 220, 320, 520に対応する。出力プロセス606cは、応答音声生成部130, 230, 330, 530に対応する。

【0094】

なお、各プログラム607a～607cについては、必ずしも最初からハードディスク装置607に記憶させておかなくてもよい。例えば、コンピュータ600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」に各プログラムを記憶させておく。そして、コンピュータ600がこれらから各プログラム607a～607cを読み出して実行するようにしてもよい。

30

【0095】

以上の各実施例を含む実施形態に関し、さらに以下の付記を開示する。

【0096】

(付記1) 入力された音声を変換する音声認識を行い、前記音声に対応する複数の単語候補を抽出する音声認識部と、

前記音声認識部によって抽出された複数の単語候補を基にして、各単語候補間で相違する箇所を選択する選択部と、

40

前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、前記選択部によって選択された相違する箇所を強調して、前記単語候補の応答音声を出力する出力部と

を有することを特徴とする音声処理装置。

【0097】

(付記2) 前記音声認識部は、前記音声に対応する複数の単語候補毎に信頼度を算出し、前記選択部は、信頼度が閾値以上となる複数の単語候補を基にして、各単語候補のうち相違する箇所を選択することを特徴とする付記1に記載の音声処理装置。

【0098】

(付記3) 前記音声認識部は、前記音声に対する複数の単語候補毎に信頼度を算出し、前

50

記選択部は、最大の信頼度となる単語候補と、最大の信頼度との差が閾値未満となる単語候補とを基にして、各単語候補間で相違する箇所を選択することを特徴とする付記 1 に記載の音声処理装置。

【0099】

(付記 4) 前記出力部は、前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、前記選択部によって選択された相違する箇所の強度を、選択されていない箇所の強度よりも大きくして、前記単語候補の応答音声を出力することを特徴とする付記 1、2 または 3 に記載の音声処理装置。

【0100】

(付記 5) 前記出力部は、前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、前記選択部によって選択された相違する箇所の発話速度を、選択されていない箇所の発話速度よりも遅くして、前記単語候補の応答音声を出力することを特徴とする付記 1、2 または 3 に記載の音声処理装置。

10

【0101】

(付記 6) 前記出力部は、前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、前記選択部によって選択された相違する箇所の音声の基本周期を長く、または短くして前記単語候補の応答音声を出力することを特徴とする付記 1、2 または 3 に記載の音声処理装置。

【0102】

(付記 7) 端末装置とサーバとを備えた音声処理システムであって、
前記端末装置は、入力された音声の情報を前記サーバに送信し、
前記サーバは、

20

前記端末装置から音声の情報を受信し、該音声を変換する音声認識を行い、前記音声に対応する複数の単語候補を抽出する音声認識部と、

前記音声認識部によって抽出された複数の単語候補を基にして、各単語候補間で相違する箇所を選択する選択部と、

前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、前記選択部によって選択された相違する箇所を強調する前記単語候補の応答音声の情報を前記端末装置に送信する送信部と

を有することを特徴とする音声処理システム。

30

【0103】

(付記 8) 前記音声認識部は、前記音声に対応する複数の単語候補毎に信頼度を算出し、前記選択部は、信頼度が閾値以上となる複数の単語候補を基にして、各単語候補のうち相違する箇所を選択することを特徴とする付記 7 に記載の音声処理システム。

【0104】

(付記 9) 前記音声認識部は、前記音声に対する複数の単語候補毎に信頼度を算出し、前記選択部は、最大の信頼度となる単語候補と、最大の信頼度との差が閾値未満となる単語候補とを基にして、各単語候補間で相違する箇所を選択することを特徴とする付記 7 に記載の音声処理システム。

【0105】

(付記 10) 前記送信部は、前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、前記選択部によって選択された相違する箇所の強度を、選択されていない箇所の強度よりも大きくした前記単語候補の応答音声の情報を前記端末装置に送信することを特徴とする付記 7、8 または 9 に記載の音声処理システム。

40

【0106】

(付記 11) 前記送信部は、前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、前記選択部によって選択された相違する箇所の発話速度を、選択されていない箇所の発話速度よりも遅くした前記単語候補の応答音声の情報を前記端末装置に送信することを特徴とする付記 7、8 または 9 に記載の音声処理システム。

【0107】

50

(付記 1 2) 前記送信部は、前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、前記選択部によって選択された相違する箇所音声の基本周期を長く、または短くした前記単語候補の応答音声の情報を前記端末装置に送信することを特徴とする付記 1、2 または 3 に記載の音声処理装置。

【0108】

(付記 1 3) コンピュータが実行する音声処理方法であって、
 入力された音声を単語に変換する音声認識を行い、前記音声に対応する複数の単語候補を抽出し、
 抽出した複数の単語候補を基にして、各単語候補間で相違する箇所を選択し、
 複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、相違する箇所を強調して、前記単語候補の応答音声を出力する
 各処理を実行することを特徴とする音声処理方法。

10

【0109】

(付記 1 4) 前記複数の単語候補を抽出する処理は、前記音声に対応する複数の単語候補毎に信頼度を算出し、前記相違する箇所を選択する処理は、信頼度が閾値以上となる複数の単語候補を基にして、各単語候補のうち相違する箇所を選択することを特徴とする付記 1 3 に記載の音声処理方法。

【0110】

(付記 1 5) 前記複数の単語候補を抽出する処理は、前記音声に対する複数の単語候補毎に信頼度を算出し、前記相違する箇所を選択する処理は、最大の信頼度となる単語候補と、最大の信頼度との差が閾値未満となる単語候補とを基にして、各単語候補間で相違する箇所を選択することを特徴とする付記 1 3 に記載の音声処理方法。

20

【0111】

(付記 1 6) 前記応答音声を出力する処理は、前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、相違する箇所の強度を、選択されていない箇所の強度よりも大きくして、前記単語候補の応答音声を出力することを特徴とする付記 1 3、1 4 または 1 5 に記載の音声処理方法。

【0112】

(付記 1 7) 前記応答音声を出力する処理は、前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、相違する箇所の発話速度を、選択されていない箇所の発話速度よりも遅くして、前記単語候補の応答音声を出力することを特徴とする付記 1 3、1 4 または 1 5 に記載の音声処理方法。

30

【0113】

(付記 1 8) 前記応答音声を出力する処理は、前記複数の単語候補から前記音声に対応する単語候補を選択し、選択した単語候補のうち、相違する箇所の音声の基本周期を長く、または短くして、前記単語候補の応答音声を出力することを特徴とする付記 1 3、1 4 または 1 5 に記載の音声処理方法。

【符号の説明】

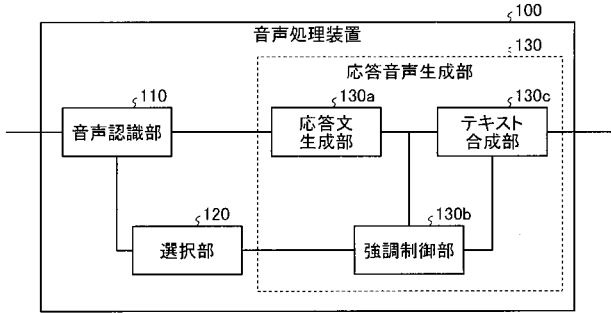
【0114】

100, 200, 300 音声処理装置
 110, 210, 310 音声認識部
 120, 220, 320 選択部
 130, 230, 330 応答音声生成部

40

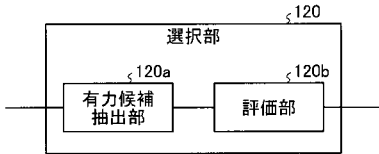
【 図 1 】

本実施例1に係る音声認識装置の構成を示す図



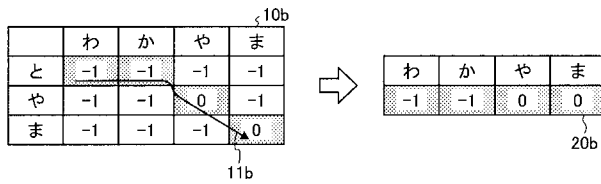
【 図 2 】

本実施例1に係る選択部の構成を示す図



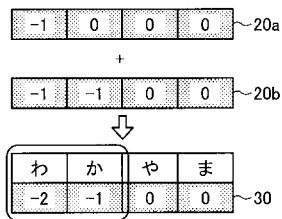
【 図 5 】

本実施例1に係る評価部の処理を説明するための図(2)



【 図 6 】

本実施例1に係る評価部の処理を説明するための図(3)



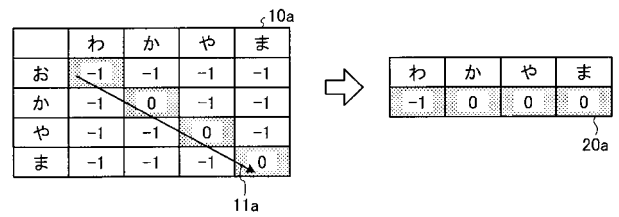
【 図 3 】

本実施例1に係る有力候補抽出部の処理を説明するための図

| 候補 | 単語候補 | 信頼度 |
|----|-----------|------|
| 1 | わかやま(和歌山) | 0.80 |
| 2 | おかやま(岡山) | 0.75 |
| 3 | とやま(富山) | 0.65 |
| 4 | とくしま(徳島) | 0.30 |
| 5 | かごしま(鹿児島) | 0.25 |

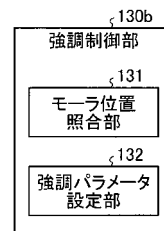
【 図 4 】

本実施例1に係る評価部の処理を説明するための図(1)



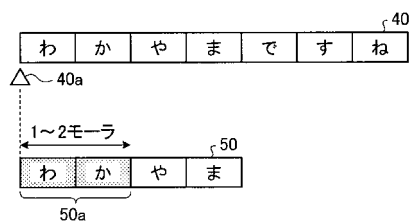
【 図 7 】

本実施例1に係る強調制御部の構成を示す図



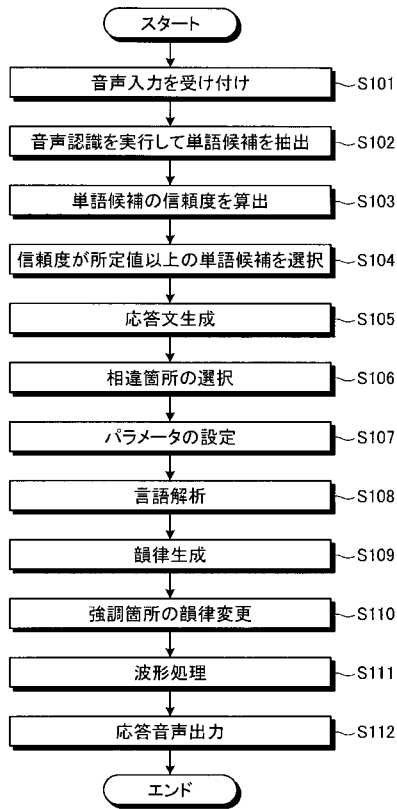
【 図 8 】

本実施例1に係るモーラ位置照合部の処理を説明するための図



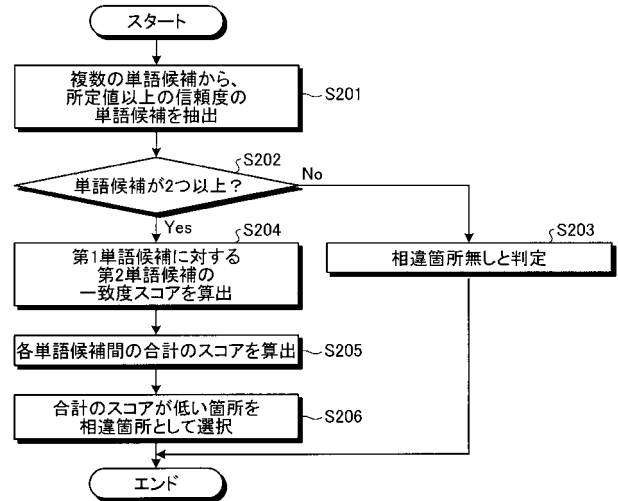
【 図 9 】

本実施例1に係る音声認識装置の処理手順を示すフローチャート



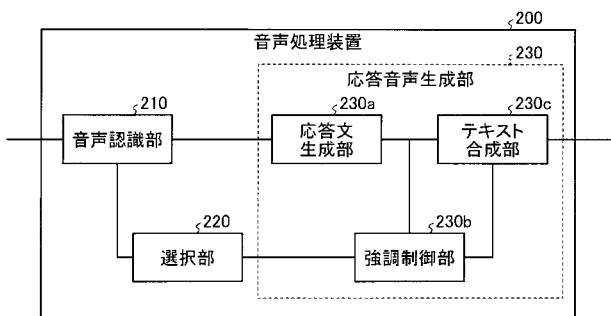
【 図 1 0 】

本実施例1に係る選択部の処理手順を示すフローチャート



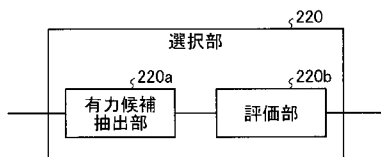
【 図 1 1 】

本実施例2に係る音声認識装置の構成を示す図



【 図 1 2 】

本実施例2に係る選択部の構成を示す図



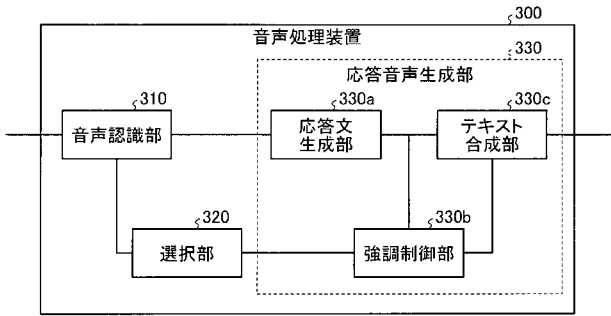
【 図 1 3 】

本実施例2に係る有力候補抽出部の処理を説明するための図

| 候補 | 単語候補 | 信頼度 | 信頼度との差 |
|----|-----------|------|--------|
| 1 | わかやま(和歌山) | 0.80 | 0.00 |
| 2 | おかやま(岡山) | 0.75 | 0.05 |
| 3 | とやま(富山) | 0.65 | 0.15 |
| 4 | とくしま(徳島) | 0.30 | 0.50 |
| 5 | かごしま(鹿児島) | 0.25 | 0.55 |

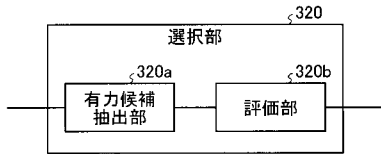
【図14】

本実施例3に係る音声認識装置の構成を示す図



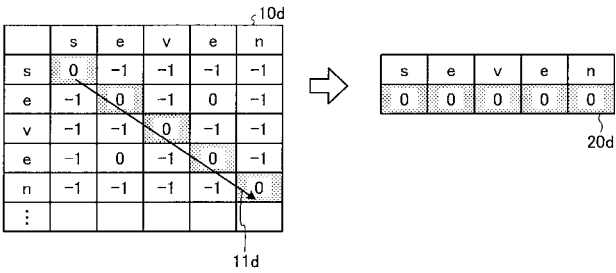
【図15】

本実施例3に係る選択部の構成を示す図



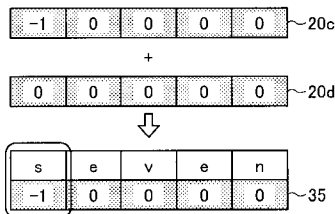
【図18】

本実施例3に係る評価部の処理を説明するための図(2)



【図19】

本実施例3に係る評価部の処理を説明するための図(3)



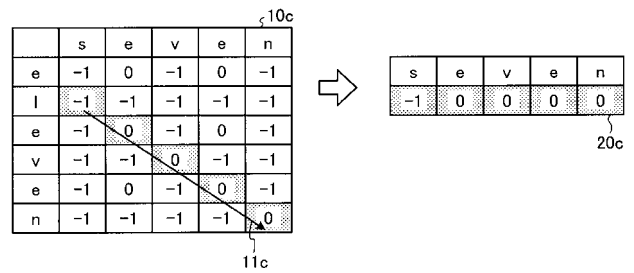
【図16】

本実施例3に係る有力候補抽出部が抽出した単語候補と信頼度との一例を示す図

| 候補 | 単語候補 | 音節 | 信頼度 |
|----|-----------|-------------|------|
| 1 | seven | sev·en | 0.80 |
| 2 | eleven | e·lev·en | 0.75 |
| 3 | seventeen | sev·en·teen | 0.65 |

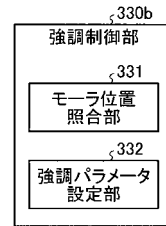
【図17】

本実施例3に係る評価部の処理を説明するための図(1)



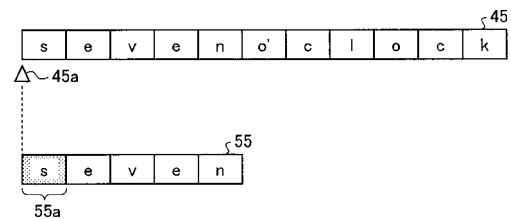
【図20】

本実施例3に係る強調制御部の構成を示す図



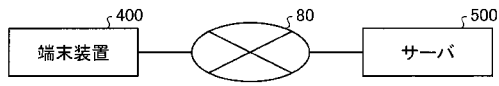
【図21】

本実施例3に係るモーラ位置照合部の処理を説明するための図



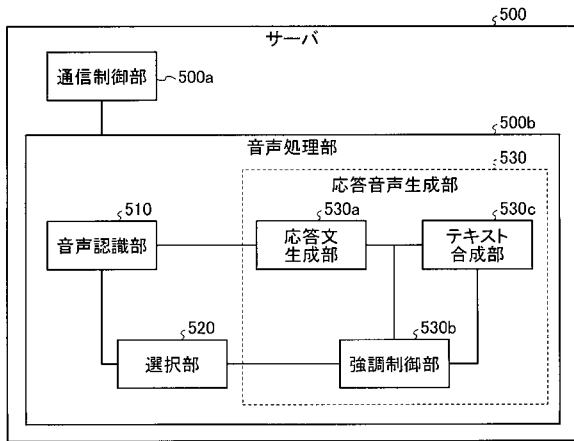
【 図 2 2 】

本実施例4に係る音声処理システムの一例を示す図



【 図 2 3 】

本実施例4に係るサーバの構成を示す図



【 図 2 4 】

音声処理プログラムを実行するコンピュータの一例を示す図

