

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5243325号
(P5243325)

(45) 発行日 平成25年7月24日(2013.7.24)

(24) 登録日 平成25年4月12日(2013.4.12)

(51) Int.Cl.	F I	
G 1 O L 15/00 (2013.01)	G 1 O L 15/00	2 O O B
G 1 O L 15/187 (2013.01)	G 1 O L 15/18	2 O O D
G 1 O L 15/197 (2013.01)	G 1 O L 15/18	2 O O Z
G 1 O L 15/183 (2013.01)	G 1 O L 15/18	3 O O G
G 1 O L 15/18 (2013.01)		

請求項の数 9 (全 14 頁)

(21) 出願番号	特願2009-81489 (P2009-81489)	(73) 特許権者	000208891 K D D I 株式会社 東京都新宿区西新宿二丁目3番2号
(22) 出願日	平成21年3月30日(2009.3.30)	(74) 代理人	100135068 弁理士 早原 茂樹
(65) 公開番号	特開2010-231149 (P2010-231149A)	(72) 発明者	内部 利明 埼玉県ふじみ野市大原二丁目1番15号 株式会社K D D I 研究所内
(43) 公開日	平成22年10月14日(2010.10.14)	審査官	山下 剛史
審査請求日	平成23年8月16日(2011.8.16)		

最終頁に続く

(54) 【発明の名称】 音声認識に仮名漢字変換システムを用いた端末、方法及びプログラム

(57) 【特許請求の範囲】

【請求項1】

音声認識で用いられる音響モデルを蓄積する音響モデル蓄積手段と、
単語辞書と、N-gram及び格文法に対応した言語モデルとを蓄積する言語モデル蓄積手段と、

利用者によって発声された音声信号を取得するマイク手段と、
前記音声信号から音素フレームを抽出する音響分析手段と、
前記音素フレームに対する複数候補の読み仮名毎に、前記音響モデルが与える音響スコアを算出する音響スコア算出手段と、

候補となる読み仮名系列に対して、前記言語モデルが与える単語間の接続強度に応じて N-gramスコアを算出するN-gramスコア算出手段と、

候補となる読み仮名系列に対して、文節 - 文節間の接続強度に応じて格文法スコアを算出する格文法スコア算出手段と、

候補となる読み仮名系列に対して、前記音響スコア、前記N-gramスコア及び前記格文法スコアに基づく統合スコアを算出する統合スコア算出手段と、

形態素解析に基づいて複数候補の仮名漢字を選別すると共に、前記統合スコアが最も高い仮名漢字を探索するために、前記音響スコア算出手段、前記N-gramスコア算出手段、前記格文法スコア算出手段及び前記統合スコア算出手段を繰り返す探索制御手段と、

探索された仮名漢字を含む文章の格構造によって変換候補を選択する格フレーム処理手段と

10

20

を有し、音声認識機能を実現することを特徴とする端末。

【請求項 2】

過去に変換された仮名漢字を記憶する履歴蓄積手段と、
前記履歴蓄積手段を用いて、候補となる読み仮名系列に対して、過去の学習履歴に応じて履歴スコアを算出する履歴スコア算出手段と
を有し、

前記統合スコア算出手段は、更に、前記履歴スコアを前記統合スコアに算入することを特徴とする請求項 1 に記載の端末。

【請求項 3】

キーを用いて前記利用者によって入力された平仮名系列を取得するキー操作手段を更に有し、

前記N-gramスコア算出手段、前記格文法スコア算出手段、前記統合スコア算出手段、前記探索制御手段及び前記格フレーム処理手段によって、仮名漢字変換機能を実現することを特徴とする請求項 1 又は 2 に記載の端末。

【請求項 4】

前記利用者に対して複数の変換候補の中から、いずれか 1 つの仮名漢字を選択させる候補選択画面制御手段を更に有し、

候補選択画面制御手段は、前記音声認識機能及び前記仮名漢字変換機能を、利用者に対して同一の候補選択画面を視認させることを特徴とする請求項 3 に記載の端末。

【請求項 5】

端末における音声認識方法であって、
前記端末は、
音声認識で用いられる音響モデルを蓄積する音響モデル蓄積部と、
単語辞書と、N-gram及び格文法に対応した言語モデルとを蓄積する言語モデル蓄積部と
を有し、

利用者によって発声された音声信号を、マイク部によって取得する第 1 のステップと、
前記音声信号から音素フレームを抽出する第 2 のステップと、

前記音素フレームに対する複数候補の読み仮名毎に、前記音響モデルが与える音響スコアを算出する第 3 のステップと、

候補となる読み仮名系列に対して、前記言語モデルが与える単語間の接続強度に応じて N-gramスコアを算出する第 4 のステップと、

候補となる読み仮名系列に対して、文節 - 文節間の接続強度に応じて格文法スコアを算出する第 5 のステップと、

候補となる読み仮名系列に対して、前記音響スコア、前記N-gramスコア及び前記格文法スコアに基づく統合スコアを算出する第 6 のステップと、

形態素解析に基づいて複数候補の仮名漢字を選別すると共に、前記統合スコアが最も高い仮名漢字を探索するために、第 3 から第 6 のステップを繰り返す第 7 のステップと、

探索された仮名漢字を含む文章の格構造によって変換候補を選択する第 8 のステップと
を有し、音声認識機能を実現することを特徴とする音声認識方法。

【請求項 6】

前記端末は、過去に変換された仮名漢字を記憶する履歴蓄積部を更に有し、
前記履歴蓄積部を用いて、候補となる読み仮名系列に対して、過去の学習履歴に応じて履歴スコアを算出するステップを更に有し、

第 6 のステップは、更に、前記履歴スコアを前記統合スコアに算入することを特徴とする請求項 5 に記載の音声認識方法。

【請求項 7】

前記端末は、キーを用いて前記利用者によって入力された平仮名系列を取得するキー操作部を更に有し、

第 4 のステップから第 8 のステップまでによって、仮名漢字変換機能を実現することを特徴とする請求項 5 又は 6 に記載の音声認識方法。

10

20

30

40

50

【請求項 8】

前記利用者に対して複数の変換候補の中から、いずれか 1 つの仮名漢字を選択させるステップを更に有することによって、前記音声認識機能及び前記仮名漢字変換機能を、利用者に対して同一の候補選択画面を視認させることを特徴とする請求項 7 に記載の音声認識方法。

【請求項 9】

端末に搭載されたコンピュータを機能させる音声認識プログラムであって、音声認識で用いられる音響モデルを蓄積する音響モデル蓄積手段と、単語辞書と、N-gram及び格文法に対応した言語モデルとを蓄積する言語モデル蓄積手段と、

利用者によって発声された音声信号を取得するマイク手段と、前記音声信号から音素フレームを抽出する音響分析手段と、前記音素フレームに対する複数候補の読み仮名毎に、前記音響モデルが与える音響スコアを算出する音響スコア算出手段と、

候補となる読み仮名系列に対して、前記言語モデルが与える単語間の接続強度に応じてN-gramスコアを算出するN-gramスコア算出手段と、

候補となる読み仮名系列に対して、文節 - 文節間の接続強度に応じて格文法スコアを算出する格文法スコア算出手段と、

候補となる読み仮名系列に対して、前記音響スコア、前記N-gramスコア及び前記格文法スコアに基づく統合スコアを算出する統合スコア算出手段と、

形態素解析に基づいて複数候補の仮名漢字を選別すると共に、前記統合スコアが最も高い仮名漢字を探索するために、前記音響スコア算出手段、前記N-gramスコア算出手段、前記格文法スコア算出手段及び前記統合スコア算出手段を繰り返す探索制御手段と、

探索された仮名漢字を含む文章の格構造によって変換候補を選択する格フレーム処理手段と

してコンピュータを機能させることを特徴とする音声認識プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声認識に仮名漢字変換システムを用いた端末、方法及びプログラムに関する。

【背景技術】

【0002】

従来、アプリケーションに文字を入力するために、キー操作によって利用者に入力された平仮名を、仮名漢字の文字列に変換する仮名漢字変換システムがある。通常、ワードプロセッサソフトウェアやテキストエディタソフトウェアに組み込まれた日本語入力機能として動作する。

【0003】

これに対し、利用者の発声した音声信号をマイクで取得し、その音声信号を仮名漢字の文字列に変換する音声認識システムがある。利用者は、キーを操作することなく、文字列を入力することができる。しかしながら、人が発声する単語の語彙数が膨大である。そのために、音声認識の精度を高めることが難しい。

【0004】

音声認識によって予測される文字列の精度を向上させるために、文章から話題予測データを作成する技術がある（例えば特許文献 1 参照）。話題予測データと平仮名とに基づいて、仮名漢字の文字列が予測される。

【0005】

また、予測される複数候補の文字列を、特定の認識対象語に絞り込むことによって、認識精度を高める技術もある（例えば特許文献 2 参照）。しかしながら、認識対象語が絞り込まれているために、音声認識システムの用途も限定される。

10

20

30

40

50

【0006】

図1は、従来技術における仮名漢字変換機能及び音声認識機能を有するシステムである。

【0007】

図1によれば、端末1は、ハードウェアとして、ディスプレイ100と、キー操作部111と、マイク121とを有する。ディスプレイ100は、テキストエディタのようなアプリケーションの入力インタフェースを表示すると共に、利用者によって入力された平仮名及び予測候補の仮名漢字を表示する。キー操作部111は、利用者に対して、キーを用いて平仮名を入力させる。マイク121は、利用者によって発声された音声信号を取得する。

10

【0008】

端末1は、仮名漢字変換機能として、仮名漢字変換エンジン112と、言語モデル蓄積部113と、履歴蓄積部114と、仮名漢字選択画面制御部115とを有する。仮名漢字変換エンジン112は、言語モデル蓄積部113及び履歴蓄積部114を用いて、キー操作部111によって入力された平仮名を、複数候補の仮名漢字に変換する。仮名漢字選択画面制御部115は、利用者に対して、複数候補の仮名漢字から1つの仮名漢字を選択させる。

【0009】

言語モデル蓄積部113は、単語辞書と言語モデルとを蓄積する。言語モデルは、単語間の接続関係の確率を規定する。単語辞書に未登録の単語には、変換することができない。そのために、言語モデル蓄積部113は、キー操作によって入力される全ての単語を含む単語辞書と、それらの言語モデルとを蓄積しておく必要がある。仮名漢字変換システムに対応する言語モデル蓄積部113は、係り受け文法の確率と、格文法の確率とを蓄積する。

20

【0010】

履歴蓄積部114は、過去に利用者によって選択された仮名漢字(単語、文節等)の学習履歴を蓄積する。

【0011】

また、端末1は、通信インタフェースを介して、音声認識サーバ2へ接続する。そして、端末1は、マイク121によって取得された音声信号を、ネットワークを介して、音声認識サーバ2へ送信する。音声認識では、人が発声する単語の語彙数が膨大であって、大容量の言語モデルと高い処理負荷とを要する。そのために、音声認識エンジンを、音声認識サーバとして別途備えることは好ましい。

30

【0012】

音声認識サーバは、音声認識エンジン122と、音響モデル蓄積部123と、言語モデル蓄積部124とを有する。音声認識エンジン122は、音響モデル蓄積部123及び言語モデル蓄積部124を用いて、受信された音声信号から複数候補の仮名漢字に変換する。複数候補の仮名漢字は、ネットワークを介して、端末1へ送信される。端末1の音声認識選択画面制御部125は、複数候補の仮名漢字から、利用者によって1つの仮名漢字を選択させる。

40

【0013】

音響モデル蓄積部123は、音響モデルを蓄積する。音響モデルとは、音響特徴パラメータと音素との対応関係を蓄積する。音素とは、単語をローマ字で書いたときのアルファベットの単位にほぼ相当する。音響モデル蓄積部123は、音響特徴パラメータに対するその音素の確率を蓄積する。

【0014】

言語モデル蓄積部124は、単語間の接続関係の確率を規定する。音声認識システムの言語モデル蓄積部124は、隣接するN個の単語間の接続関係の確率を利用するN-gramを採用する。これに対し、仮名漢字変換システムの言語モデル蓄積部113は、隣接する2個の単語間の接続関係の確率のみを利用する係り受け文法を採用する。そのために、音声

50

認識システムの言語モデル蓄積部 1 2 4 は、仮名漢字システムの言語モデル蓄積部 1 1 3 よりも、リソースが膨大となる。勿論、端末 1 が、音声認識エンジン 1 2 2 と、音響モデル蓄積部 1 2 3 と、言語モデル蓄積部 1 2 4 とを更に含むものであってもよい。

【 0 0 1 5 】

尚、音声認識システムによって人の発声した音声信号を平仮名に変換し、次に、仮名漢字変換システムによってその平仮名を仮名漢字に変換する技術もある（例えば非特許文献 1 及び 2 参照）。

【先行技術文献】

【特許文献】

【 0 0 1 6 】

【特許文献 1】特開 2 0 0 0 - 2 8 5 1 1 2 号公報

【特許文献 2】特許第 4 0 1 2 1 4 3 号

【非特許文献】

【 0 0 1 7 】

【非特許文献 1】音声認識エンジン「VORERO (VOice REcognition RObustR)」、旭化成、[online]、[平成 2 1 年 3 月 1 9 日検索]、インターネット<URL:http://www.asahi-kasei.co.jp/vorero/jp/vorero/index.html>

【非特許文献 2】日本語かな漢字変換システム「モバイルWnn」、オムロン、[online]、[平成 2 1 年 3 月 1 9 日検索]、インターネット<URL:http://www.omronsoft.co.jp/SP/mobile/>

【発明の概要】

【発明が解決しようとする課題】

【 0 0 1 8 】

前述した従来技術によれば、音声認識システムと仮名漢字変換システムとは別々に備えられ、互いに独立のリソースを用いている。音声認識システムにおける音響モデル及び言語モデルは、流行語のような全ての口語文字に対応する必要があり、そのリソースも膨大である。一方で、音声認識システムでは、仮名漢字変換システムに組み込まれた格文法、形態素解析及び格フレーム処理のような処理を実行しないため、仮名漢字変換における誤変換も生じやすい。

【 0 0 1 9 】

そこで、本発明は、音声認識に仮名漢字変換システムを用いることによって、システム全体の言語モデルのリソースを統合すると共に、音声認識の誤変換を削減することができる端末、方法及びプログラム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 2 0 】

本発明によれば、

音声認識で用いられる音響モデルを蓄積する音響モデル蓄積手段と、

単語辞書と、N-gram及び格文法に対応した言語モデルとを蓄積する言語モデル蓄積手段と、

利用者によって発声された音声信号を取得するマイク手段と、

音声信号から音素フレームを抽出する音響分析手段と、

音素フレームに対する複数候補の読み仮名毎に、音響モデルが与える音響スコアを算出する音響スコア算出手段と、

候補となる読み仮名系列に対して、言語モデルが与える単語間の接続強度に応じてN-gramスコアを算出するN-gramスコア算出手段と、

候補となる読み仮名系列に対して、文節 - 文節間の接続強度に応じて格文法スコアを算出する格文法スコア算出手段と、

候補となる読み仮名系列に対して、音響スコア、N-gramスコア及び格文法スコアに基づく統合スコアを算出する統合スコア算出手段と、

形態素解析に基づいて複数候補の仮名漢字を選別すると共に、統合スコアが最も高い仮

10

20

30

40

50

名漢字を探索するために、音響スコア算出手段、N-gramスコア算出手段、格文法スコア算出手段及び統合スコア算出手段を繰り返す探索制御手段と、

探索された仮名漢字を含む文章の格構造によって変換候補を選択する格フレーム処理手段と

を有し、音声認識機能を実現することを特徴とする。

【0021】

本発明の端末における他の実施形態によれば、

過去に変換された仮名漢字を記憶する履歴蓄積手段と、

履歴蓄積手段を用いて、候補となる読み仮名系列に対して、過去の学習履歴に応じて履歴スコアを算出する履歴スコア算出手段と

10

を有し、

統合スコア算出手段は、更に、履歴スコアを統合スコアに算入することも好ましい。

【0022】

本発明の端末における他の実施形態によれば、

キーを用いて利用者によって入力された平仮名系列を取得するキー操作手段を更に有し、

N-gramスコア算出手段、格文法スコア算出手段、統合スコア算出手段、探索制御手段及び格フレーム処理手段によって、仮名漢字変換機能を実現することも好ましい。

【0023】

本発明の端末における他の実施形態によれば、

利用者に対して複数の変換候補の中から、いずれか1つの仮名漢字を選択させる候補選択画面制御手段を更に有し、

20

候補選択画面制御手段は、音声認識機能及び仮名漢字変換機能を、利用者に対して同一の候補選択画面を視認させることも好ましい。

【0024】

本発明によれば、端末における音声認識方法であって、

端末は、

音声認識で用いられる音響モデルを蓄積する音響モデル蓄積部と、

単語辞書と、N-gram及び格文法に対応した言語モデルとを蓄積する言語モデル蓄積部と

30

を有し、利用者によって発声された音声信号を、マイク部によって取得する第1のステップと、音声信号から音素フレームを抽出する第2のステップと、

音素フレームに対する複数候補の読み仮名毎に、音響モデルが与える音響スコアを算出する第3のステップと、

候補となる読み仮名系列に対して、言語モデルが与える単語間の接続強度に応じてN-gramスコアを算出する第4のステップと、

候補となる読み仮名系列に対して、文節 - 文節間の接続強度に応じて格文法スコアを算出する第5のステップと、

候補となる読み仮名系列に対して、音響スコア、N-gramスコア及び格文法スコアに基づく統合スコアを算出する第6のステップと、

40

形態素解析に基づいて複数候補の仮名漢字を選別すると共に、統合スコアが最も高い仮名漢字を探索するために、第3から第6のステップを繰り返す第7のステップと、

探索された仮名漢字を含む文章の格構造によって変換候補を選択する第8のステップとを有し、音声認識機能を実現することを特徴とする。

【0025】

本発明の音声認識方法における他の実施形態によれば、

端末は、過去に変換された仮名漢字を記憶する履歴蓄積部を更に有し、

履歴蓄積部を用いて、候補となる読み仮名系列に対して、過去の学習履歴に応じて履歴スコアを算出するステップを更に有し、

第6のステップは、更に、履歴スコアを統合スコアに算入することも好ましい。

50

【0026】

本発明の音声認識方法における他の実施形態によれば、
 端末は、キーを用いて利用者によって入力された平仮名系列を取得するキー操作部を更に有し、
 第4のステップから第8のステップまでによって、仮名漢字変換機能を実現することも好ましい。

【0027】

本発明の音声認識方法における他の実施形態によれば、
 利用者に対して複数の変換候補の中から、いずれか1つの仮名漢字を選択させるステップを更に有することによって、音声認識機能及び仮名漢字変換機能を、利用者に対して同一の候補選択画面を視認させることも好ましい。

10

【0028】

本発明によれば、端末に搭載されたコンピュータを機能させる音声認識プログラムであって、
 音声認識で用いられる音響モデルを蓄積する音響モデル蓄積手段と、
 単語辞書と、N-gram及び格文法に対応した言語モデルとを蓄積する言語モデル蓄積手段と、
 利用者によって発声された音声信号を取得するマイク手段と、
 音声信号から音素フレームを抽出する音響分析手段と、
 音素フレームに対する複数候補の読み仮名毎に、音響モデルが与える音響スコアを算出する音響スコア算出手段と、
 候補となる読み仮名系列に対して、言語モデルが与える単語間の接続強度に応じてN-gramスコアを算出するN-gramスコア算出手段と、
 候補となる読み仮名系列に対して、文節 - 文節間の接続強度に応じて格文法スコアを算出する格文法スコア算出手段と、
 候補となる読み仮名系列に対して、音響スコア、N-gramスコア及び格文法スコアに基づく統合スコアを算出する統合スコア算出手段と、
 形態素解析に基づいて複数候補の仮名漢字を選別すると共に、統合スコアが最も高い仮名漢字を探索するために、音響スコア算出手段、N-gramスコア算出手段、格文法スコア算出手段及び統合スコア算出手段を繰り返す探索制御手段と、
 探索された仮名漢字を含む文章の格構造によって変換候補を選択する格フレーム処理手段としてコンピュータを機能させることを特徴とする。

20

30

【発明の効果】

【0029】

本発明の端末、方法及びプログラムによれば、音声認識に仮名漢字変換システムを用いることによって、システム全体の言語モデルのリソースを統合すると共に、音声認識の誤変換を削減することができる。

【図面の簡単な説明】

【0030】

40

【図1】従来技術における仮名漢字変換機能及び音声認識機能を有するシステムである。

【図2】仮名漢字変換システム及び音声認識システムの機能構成図である。

【図3】本発明における端末の機能構成図である。

【図4】本発明におけるフローチャートである。

【図5】従来技術と本発明とを比較した、仮名漢字の予測変換シーケンスである。

【発明を実施するための形態】

【0031】

以下、本発明の実施の形態について、図面を用いて詳細に説明する。

【0032】

図2は、仮名漢字変換システム及び音声認識システムの機能構成図である。

50

文節「ぜいきんが」 文節「おさめられている」の接続強度は高く、且つ、文章構成のける重要度は低い。「ぜいきんがおさめられているので」を修飾語と判断し、「わたしは」「くにに」「おさめる」が、文章構成の重要度が高いと判断する。このように、従来の音声認識システムによれば言語モデルとしてN-gramしか考慮していないのに対し、自由度が高い文法の発声にも対応することができる。

【0040】

言語モデル蓄積部133は、単語辞書及び言語モデルを蓄積する。特に、単語辞書における単語間の接続関係を、点数化（スコア化）して蓄積する。ここで、本発明における言語モデル蓄積部133は、N-gramの確率及び格文法の確率を有し、N-gramスコア算出部1223及び格文法スコア算出部1122に対応したものである。その点で、仮名漢字変換システム及び音声認識システムにおける既存の言語モデル蓄積部とは異なる。

10

【0041】

履歴スコア算出部1123は、候補となる読み仮名系列に対して、履歴蓄積部114を用いて、過去に利用者によって選択された仮名漢字（単語、文節等）を、高いスコアに設定する。即ち、過去の学習結果を、言語スコアに反映させる。基本的に、その読み仮名で、直前に選択された仮名漢字を予測候補にする。また、その読み仮名に対して、過去に用いられた文章の内容に応じて仮名漢字を予測候補にする。

【0042】

従来技術における音声認識システムによれば、利用者に応じて提示する候補を変更することはできない。これに対し、本発明によれば、利用者に応じて使用頻度の高い単語を特定することによって、予測変換精度を向上させることができる。また、仮名漢字変換機能と音声認識機能との間で、履歴情報が相互に共有される。

20

【0043】

音響分析部1221は、マイク121から入力された音声信号を、音響特徴パラメータに変換する。その音響特徴パラメータは、音響スコア算出部1222へ出力される。音響特徴パラメータは、入力された音声信号を数十msecの音素フレーム単位（音素フレーム）で分析したLPCケプストラム又はMFCCのようなパラメータ系列である。

【0044】

音響スコア算出部1222は、音響モデル蓄積部123を用いて、音素フレームに対する複数候補の読み仮名毎に、音響モデルが与える音響スコアを算出する。

30

【0045】

統合スコア算出部130は、候補となる読み仮名系列に対して、言語スコア（N-gramスコア、格文法スコア及び履歴スコア）と音響スコアとを統合して、最も高いスコアの単語列を、認識結果として出力する。入力された音声信号に対して、音響的な類似度と言語的な妥当性とをスコア化することができる。

【0046】

探索制御部131は、形態素解析に基づいて探索候補を選別すると共に、統合スコアが最も高い仮名漢字列を探索するために、音響スコア算出部1222、N-gramスコア算出部1223、格文法スコア算出部1122、履歴スコア算出部1123及び統合スコア算出部130を繰り返す。

40

【0047】

形態素解析とは、文法及び単語辞書を情報源として用いて、自然言語で書かれた文を形態素（Morpheme、言語で意味を持つ最小単位）の列に分割する。例えば、文「きのうわたしはしっていた。」に対して、多数の単語に区切ることができる。区切り方によっては、意味が通じない単語となる。

- ・きのう/わ/たし/は/しっ/て/いた
- ・きのう/わたし/は/しっ/て/いた
- ・きのう/わたし/はしっ/て/いた

【0048】

次に、探索制御部131は、最適な文節に区切るべく、スコアが高いN個の候補を選別

50

する。スコアを用いることによって、仮名漢字の出現頻度だけでなく、単語又は文節の接続強度によって、N個の候補が選別される。

$$\prod_{m=1}^M \prod_{n=1}^{N_m} P_m(x_{mn})$$

M：単語数

w_m ：m番目の単語

s_m ：単語 w_m の言語スコア

N_m ：単語 w_m に含まれる音素フレーム数

x_{mn} ：単語 w_m に含まれる n 番目の音素フレーム

$P_m(x_{mn})$ ：音素フレームの単語 w_m の音響モデルの音響スコア

【0049】

従来技術における音声認識システムでは、候補となる読み仮名系列の中から、音響スコア及びN-gramスコアが上位となる読み仮名に絞り込んでいる。これに対し、本発明によれば、形態素解析によって文構造の正しい候補を選別した上で、スコアが上位となる読み仮名に絞り込んでいる。これによって、予測精度も向上する。

【0050】

格フレーム処理部1126は、文の格構造によって変換候補を選択する。格フレームとは、動詞に、「どんな助詞(で、に、を、から、より、・・・)と一緒に使われるのか」という情報を持たせることを意味する。入力された読みと一致する格フレームを持つ単語を、変換候補の第1候補にする。例えば、「私はドレスを着る。」「私はナイフでドレスを切る。」のように、「着る」と「切る」の同音異義語を区別することができる。

【0051】

従来技術における音声認識システムでは、N-gramの性質上、隣接しない単語同士の間音異義語を正しい候補に変換することが難しい。これに対し、本発明によれば、格フレーム処理によって、文の格構造を概念的に認識し、名詞の後の助詞に基づいて「どの単語に係るか」を解析する。これによって、文節(格構造)が隣接しない(特に離れている)場合であっても、同音異義語の中から適切な候補を選択することができる。

【0052】

候補選択画面制御部132は、利用者に対して複数の変換候補の中から、いずれか1つの仮名漢字を選択させる。これにより、利用者は、音声認識機能及び仮名漢字変換機能を、同一の候補選択画面として視認することができる。

【0053】

図4は、本発明におけるフローチャートである。

【0054】

- (S401) マイクを用いて、利用者によって発声された音声信号を取得する。
- (S402) 音響分析によって、音声信号から音素フレームを抽出する。
- (S403) 以下、S408までを繰り返す。
- (S404) 音素フレームに対する複数候補の読み仮名毎に、音響モデルが与える音響スコアを算出する。
- (S405) 候補となる読み仮名系列に対して、言語モデルを用いて、単語間の接続強度に応じてN-gramスコアを算出する。
- (S406) 候補となる読み仮名系列に対して、言語モデルを用いて、文節 - 文節間の接続強度に応じて格文法スコアを算出する。
- (S407) 候補となる読み仮名系列に対して、履歴情報を用いて、過去の学習履歴に応じて履歴スコアを算出する。
- (S408) 候補となる読み仮名系列に対して、音響スコア及び言語スコア(N-gramスコア、格文法スコア及び履歴スコア)に基づく統合スコアを算出する。
- (S409) 形態素解析に基づいて複数候補の仮名漢字を選別すると共に、統合スコアが最も高い仮名漢字を探索するために、S403からS408までを繰り返す。
- (S410) 探索された仮名漢字を含む文章の格構造によって変換候補を選択する。
- (S411) 利用者に対して複数の変換候補の中から、いずれか1つの仮名漢字を選択さ

10

20

30

40

50

せる。利用者から見て、音声認識機能及び仮名漢字変換機能は、同一の候補選択画面として視認される。

【0055】

図5は、従来技術と本発明とを比較した、仮名漢字の予測変換シーケンスである。

【0056】

図5によれば、従来技術における音声認識の場合、音響スコア及びN-gramスコアによって探索を繰り返すために、隣接する単語に影響されやすく、誤変換されやすい。

【0057】

これに対し、本発明における音声認識の場合、格文法スコア及び履歴スコアを統合スコアに更に含めて探索を繰り返すことによって、最適な形態素に区分される。更に、格フレーム処理によって文章の格構造が判断され、最終的に、最適な仮名漢字に変換される。

10

【0058】

以上、詳細に説明したように、本発明の端末、方法及びプログラムによれば、音声認識に仮名漢字変換システムを用いることによって、システム全体の言語モデルのリソースを統合すると共に、音声認識の誤変換を削減することができる。

【0059】

本発明によれば、音声認識における誤認識を、仮名漢字変換システムにおける言語スコア計算部によって補うことによって、音声認識における認識精度を向上させる。また、言語モデルを統合することによって、当該端末が備えるべきリソースを削減させ、携帯端末への組込みを可能とする。音声認識システムと仮名漢字変換システムとにおける予測候補の選択画面が共通化されるので、利用者から見て理解を容易にする。更に、仮名漢字変換システムで利用者によって選択された仮名漢字は、音声認識システムでも同様に、高いスコアで算出される。逆に、音声認識システムで利用者によって選択された仮名漢字は、仮名漢字変換システムでも同様に、高いスコアで算出される。即ち、利用者から見て、両システムともに、同様の仮名漢字変換が実行される。

20

【0060】

前述した本発明の種々の実施形態について、本発明の技術思想及び見地の範囲の種々の変更、修正及び省略は、当業者によれば容易に行うことができる。前述の説明はあくまで例であって、何ら制約しようとするものではない。本発明は、特許請求の範囲及びその均等物として限定するものにのみ制約される。

30

【符号の説明】

【0061】

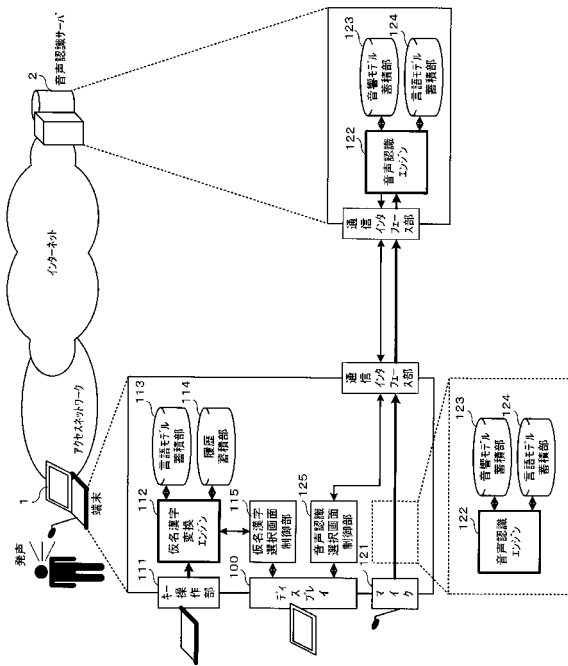
- 1 端末
- 100 ディスプレイ
- 111 キー操作部
- 112 仮名漢字変換エンジン
- 1121 係り受け文法スコア算出部
- 1122 格文法スコア算出部
- 1123 履歴スコア算出部
- 1124 統合スコア算出部
- 1125 形態素解析部
- 1126 格フレーム処理部
- 113 言語モデル蓄積部
- 114 履歴蓄積部
- 115 仮名漢字選択画面制御部
- 121 マイク
- 122 音声認識エンジン
- 1221 音響分析部
- 1222 音響スコア算出部
- 1223 N-gramスコア算出部

40

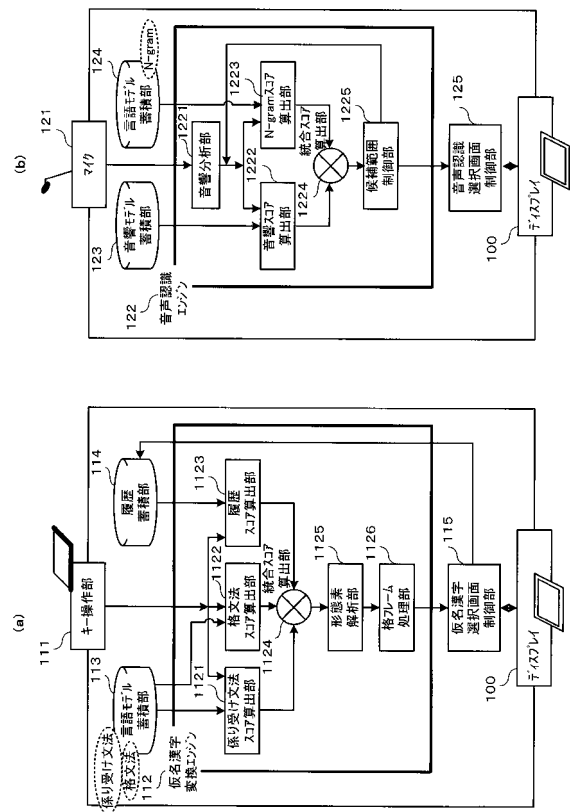
50

- 1 2 2 4 統合スコア算出部
- 1 2 2 5 候補範囲制御部
- 1 2 3 音響モデル蓄積部
- 1 2 4 言語モデル蓄積部
- 1 2 5 音声認識選択画面制御部
- 1 3 0 統合スコア算出部
- 1 3 1 探索制御部
- 1 3 2 候補選択画面制御部
- 1 3 3 言語モデル蓄積部

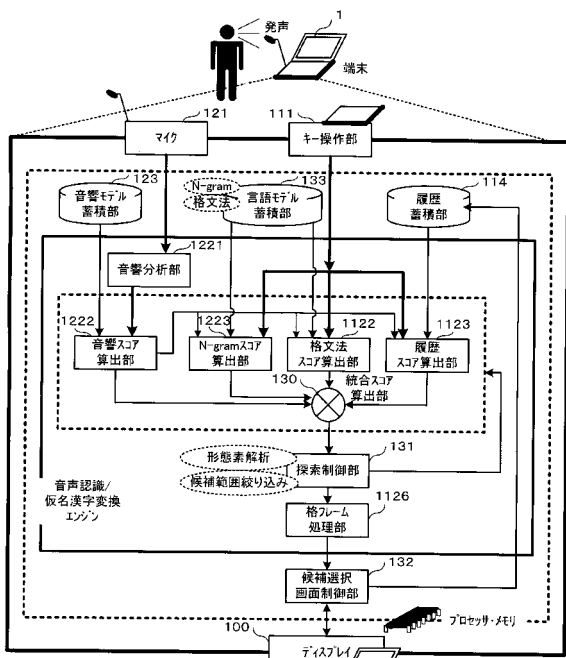
【図1】



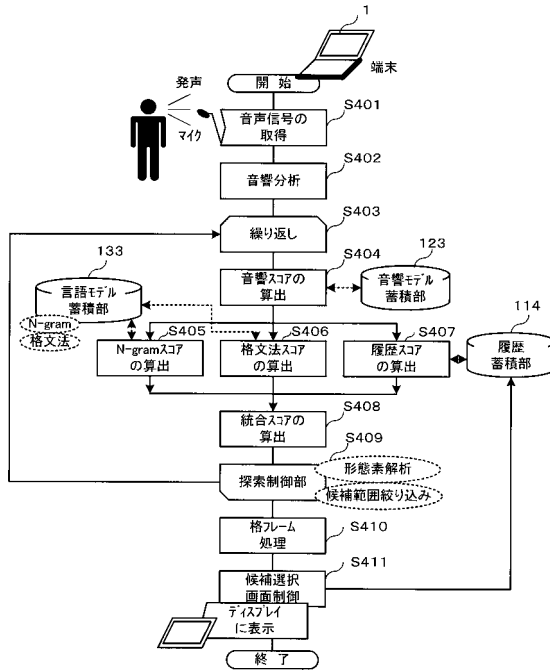
【図2】



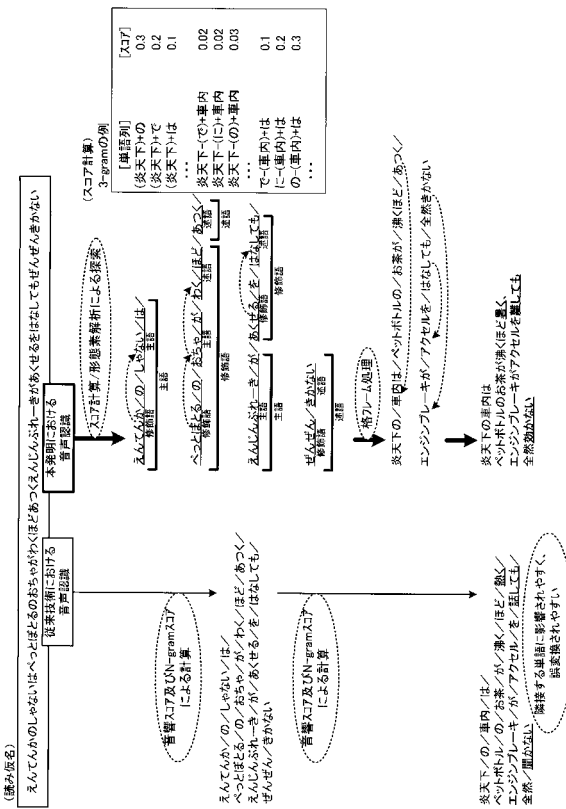
【図3】



【図4】



【図5】



フロントページの続き

- (56)参考文献 特開平2 - 22730 (JP, A)
特開2001 - 242885 (JP, A)
特開2005 - 275348 (JP, A)
特開2002 - 117025 (JP, A)
特開平6 - 102897 (JP, A)
特開平10 - 97285 (JP, A)
廣瀬良文他, "「読み」を考慮した言語モデルによる大語彙連続音声認識", 日本音響学会1999
年春季研究発表会講演論文集-I-, 1999年 3月, pp.55-56

(58)調査した分野(Int.Cl., DB名)

G10L 15/00 - 15/34
G06F 17/22