

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2010年11月11日(11.11.2010)

PCT

(10) 国際公開番号
WO 2010/128560 A1

- (51) 国際特許分類:
G10L 15/28 (2006.01)
- (21) 国際出願番号: PCT/JP2009/058707
- (22) 国際出願日: 2009年5月8日(08.05.2009)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人(米国を除く全ての指定国について): パイオニア株式会社(PIONEER CORPORATION) [JP/JP]; 〒2120031 神奈川県川崎市幸区新小倉1番1号 Kanagawa (JP).
- (72) 発明者; および
- (75) 発明者/出願人(米国についてのみ): 川添 佳洋(KAWAZOE, Yoshihiro) [JP/JP]; 〒3502288 埼玉県鶴ヶ島市富士見6丁目1番2号 パイオニア株式会社 総合研究所内 Saitama (JP). 吉田 実(YOSHIDA, Minoru) [JP/JP]; 〒3502288 埼玉県鶴ヶ島市富士見6丁目1番2号 パイオニア株式会社 総合研究所内 Saitama (JP).
- (74) 代理人: 中村 聡延, 外(NAKAMURA, Toshinobu et al.); 〒1040031 東京都中央区京橋一丁目16番10号 オークビル京橋3階 東京セントラル特許事務所内 Tokyo (JP).

- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

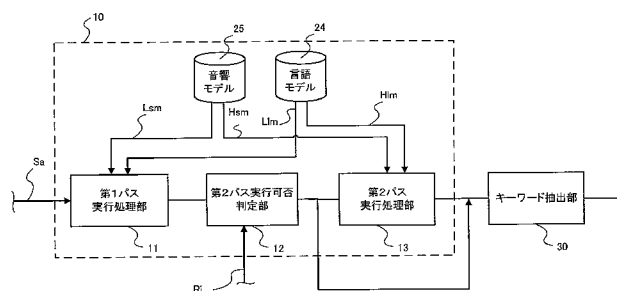
添付公開書類:

- 国際調査報告(条約第21条(3))

(54) Title: VOICE RECOGNITION DEVICE, VOICE RECOGNITION METHOD, AND VOICE RECOGNITION PROGRAM

(54) 発明の名称: 音声認識装置、音声認識方法、及び音声認識プログラム

[図1]



- 11 First pass execution unit
12 Second pass execution/non-execution determining unit
13 Second pass execution unit
24 Language model
25 Acoustic model
30 Keyword extractor

(57) Abstract: Provided is a voice recognition device that is equipped with: an acoustic model storage unit; a language model storage unit; a first pass processing means; a second pass execution/non-execution determining means; and a second pass processing means. The first pass processing means determines the candidates and scores for an array of words from the inputted speech signal, based on the acoustic model and the language model. The second pass execution/non-execution determining means determines whether the second pass processing should be executed, based on the recognition results of the first pass processing means and/or the recognition environment data. If the second pass execution/non-execution determining means determines that the second pass processing should be implemented, the second pass processing means re-determines the candidates and scores for an array of words based on the acoustic model and language model.

(57) 要約:

[続葉有]

WO 2010/128560 A1

音声認識装置は、音響モデル記憶部と、言語モデル記憶部と、第１パス処理手段と、第２パス実行可否判定手段と、第２パス処理手段と、を備える。第１パス処理手段は、音響モデルと言語モデルに基づき、入力された発話信号から単語列の候補及びスコアを決定する。第２パス実行可否判定手段は、第１パス処理手段の認識結果及び／又は認識環境の情報に基づき、第２パス処理を実行すべきか否かを判定する。そして、第２パス処理手段は、第２パス実行可否判定手段が第２パス処理を実行すべきと判定した場合、音響モデル及び前記言語モデルに基づき、単語列の候補及びスコアを再決定する。

明 細 書

発明の名称：

音声認識装置、音声認識方法、及び音声認識プログラム

技術分野

[0001] 本発明は、音声認識処理を複数回実行するマルチパス探索を用いた音声認識技術に関する。

背景技術

[0002] ディクテーション認識に代表される大語彙連続音声認識のアルゴリズムの１つとして、入力された発話に対して音声認識を複数回実行するマルチパス探索が既知である。例えば、特許文献１には、簡易な音響モデル及び簡易な言語モデルに基づき連続音声に対し認識処理を実行する第１パス処理手段と、第１パス処理の認識結果と詳細な音響モデル及び詳細な言語モデルとに基づき単語列を生成する第２パス処理手段と、を備えた音声認識装置が開示されている。

先行技術文献

特許文献

[0003] 特許文献１：特開２００３－１４０６８５号公報

発明の概要

発明が解決しようとする課題

[0004] 一般に、第１パス処理は、入力音声とほぼ並行してリアルタイムに処理をすることが可能である。しかし、第２パス処理は、その処理時間分だけ音声認識処理全体のレスポンスに影響を与える。即ち、第２パス処理の処理時間分だけ結果出力の遅延が発生する。一方、第１パス処理の認識結果が十分に信頼できるとみなせる場合には、音声認識装置は、第２パス処理以降の認識処理を必ずしも実行する必要がない。特許文献１には、上記の問題は、何ら検討されていない。

[0005] 本発明は、上記のような課題を解決するためになされたものであり、マル

チパス探索において処理量を削減し、結果出力までの処理速度を向上させることが可能な音声認識装置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

- [0006] 請求項 1 に記載の発明は、音声認識装置は、1 又は複数の音響モデルを記憶する音響モデル記憶部と、1 又は複数の言語モデルを記憶する言語モデル記憶部と、前記音響モデルと前記言語モデルに基づき、入力された発話信号から単語列の候補及びスコアを決定する第 1 パス処理手段と、前記第 1 パス処理手段の認識結果及び／又は認識環境の情報に基づき、第 2 パス処理を実行すべきか否か判定する第 2 パス実行可否判定手段と、前記第 2 パス実行可否判定手段が第 2 パス処理を実行すべきと判定した場合、前記音響モデル及び前記言語モデルに基づき、前記候補及び前記スコアを再決定する第 2 パス処理手段と、を備えることを特徴とする。
- [0007] 請求項 1 1 に記載の発明は、1 又は複数の音響モデルを記憶する音響モデル記憶部と、1 又は複数の言語モデルを記憶する言語モデル記憶部と、を利用する音声認識方法であって、前記音響モデルと前記言語モデルに基づき、入力された発話信号から単語列の候補及びスコアを決定する第 1 パス処理工程と、前記第 1 パス処理手段の認識結果及び／又は認識環境の情報に基づき、第 2 パス処理を実行すべきか否か判定する第 2 パス実行可否判定工程と、前記第 2 パス実行可否判定工程が第 2 パス処理を実行すべきと判定した場合、前記音響モデル及び前記言語モデルに基づき、前記候補及び前記スコアを再決定する第 2 パス処理工程と、を備えることを特徴とする。
- [0008] 請求項 1 2 に記載の発明は、1 又は複数の音響モデルを記憶する音響モデル記憶部と、1 又は複数の言語モデルを記憶する言語モデル記憶部と、を利用するコンピュータにより実行される音声認識プログラムであって、前記音響モデルと前記言語モデルに基づき、入力された発話信号から単語列の候補及びスコアを決定する第 1 パス処理手段と、前記第 1 パス処理手段の認識結果及び／又は認識環境の情報に基づき、第 2 パス処理を実行すべきか否か判定する第 2 パス実行可否判定手段と、前記第 2 パス実行可否判定手段が第 2

パス処理を実行すべきと判定した場合、前記音響モデル及び前記言語モデルに基づき、前記候補及び前記スコアを再決定する第2パス処理手段と、を備えることを特徴とする。

図面の簡単な説明

- [0009] [図1]言語モデルを用いた音声認識装置の概略構成図である。
- [図2]音声認識装置が実行する処理のブロック図を示す。
- [図3]第1パスマッチング処理部11cの認識結果を表す単語グラフの例を示す。
- [図4]信頼度Tに基づき音声認識装置が実行する処理を示した図である。
- [図5]本実施例の処理手順を示すフローチャートの一例である。
- [図6]変形例1に係る音声認識装置の概略構成図である。

発明を実施するための形態

- [0010] 本発明の1つの観点では、音声認識装置は、1又は複数の音響モデルを記憶する音響モデル記憶部と、1又は複数の言語モデルを記憶する言語モデル記憶部と、前記音響モデルと前記言語モデルに基づき、入力された発話信号から単語列の候補及びスコアを決定する第1パス処理手段と、前記第1パス処理手段の認識結果及び／又は認識環境の情報に基づき、第2パス処理を実行すべきか否か判定する第2パス実行可否判定手段と、前記第2パス実行可否判定手段が第2パス処理を実行すべきと判定した場合、前記音響モデル及び前記言語モデルに基づき、前記候補及び前記スコアを再決定する第2パス処理手段と、を備える。
- [0011] 上記の音声認識装置は、音響モデル記憶部と、言語モデル記憶部と、第1パス処理手段と、第2パス実行可否判定手段と、第2パス処理手段と、を備える。第1パス処理手段は、音響モデルと言語モデルに基づき、入力された発話信号から単語列の候補及びスコアを決定する。第2パス実行可否判定手段は、第1パス処理手段の認識結果及び／又は認識環境の情報に基づき、第2パス処理を実行すべきか否か判定する。「認識環境の情報」とは、音声認識装置が認識処理を実行する環境に関する情報を指し、例えば、S/N比、発

話スピード、入力音声の大きさ、車両の情報等が該当する。そして、第2パス処理手段は、第2パス実行可否判定手段が第2パス処理を実行すべきと判定した場合、音響モデル及び前記言語モデルに基づき、単語列の候補及びスコアを再決定する。このように、音声認識装置は、認識結果及び／又は認識環境の情報に基づき、適切に第2パス処理を実行すべきか否か判定することで、不要に第2パス処理以降の認識処理を実行するのを抑制することができる。従って、音声認識装置は、処理量を削減し、結果出力までの処理速度を向上させることができる。

[0012] 上記の音声認識装置の一態様では、前記第2パス処理手段は、前記第2パス実行可否判定手段が前記第2パス処理を実行すべきと判定した場合、前記第1パス処理手段で使用された音響モデル及び言語モデル以上の精度を具備する音響モデル及び言語モデルに基づき、前記候補及び前記スコアを再決定する。ここで、「第1パス処理手段で使用された音響モデル及び言語モデル以上の精度」とは、第1パス処理手段で使用された音響モデル及び言語モデルより精度の高い音響モデル及び言語モデルはもとより、第1パス処理手段で使用された音響モデル及び言語モデルと同じ音響モデル及び言語モデルをも含む意味である。このように、音声認識装置は、第1パス処理の処理量を低減し、かつ、必要なときのみ第2パス処理を実行することで、全体の処理量を削減し、結果出力までの処理速度を向上させることができる。

[0013] 上記の音声認識装置の他の一態様では、前記第2パス実行可否判定手段は、前記認識結果及び／又は前記情報に基づき前記認識結果の信頼度を算出し、当該信頼度が第1の閾値よりも高い場合又は第2の閾値よりも低い場合、前記第2パス処理を継続すべきではないと判定する。第1の閾値及び第2の閾値は、実験等に基づき適切な値に設定される。この態様では、音声認識装置は、第1パス処理の認識結果に対する信頼度を計算し、当該信頼度に基づき第2パス処理を実行すべきか否か判定する。そして、音声認識装置は、信頼度が第1の閾値より大きい場合、第1パス処理の認識結果が十分信頼でき、正解の可能性が高いと判定する。一方、音声認識装置は、信頼度が第2の

閾値より小さい場合、第１パス処理の認識結果の信頼性が低く、第２パス処理を実行しても正解の単語列を得られる可能性が極めて低いと判定する。そして、音声認識装置は、信頼度が第１の閾値よりも高い場合又は第２の閾値よりも低い場合、前記第２パス処理を継続すべきではないと判定し、不要に第２のパス処理以降の認識処理を実行するのを抑制する。このように、音声認識装置は、信頼度に基づき第２パス処理を実行すべきか否か適切に判定することができ、結果出力までの処理速度を向上させることができる。

[0014] 上記の音声認識装置の他の一態様では、前記第２パス実行可否判定手段は、前記第１パス処理により決定された前記単語列を構成する各単語の候補数に基づき前記信頼度を決定する。一般に、第１パス処理により得られた単語列を構成する各単語の候補数が少ない場合、第２パス処理を実行しても認識結果が全く変わらない可能性が高い。一方、認識できない未知語が入力された場合、上述の候補数が多くなる傾向がある。従って、この態様により、音声認識装置は、信頼度を適切に設定することができる。

[0015] 上記の音声認識装置の他の一態様では、前記第２パス実行可否判定手段は、前記第２パス実行可否判定手段は、前記第１パス処理により決定されたキーワードの候補数が第１所定個数以下の場合に前記信頼度を第１の閾値より高い値に設定し、当該候補数が第２所定個数以上の場合に前記信頼度を第２の閾値より低い値に設定する。ここで、キーワードとは、音声認識装置が単語列中で特に認識する必要がある単語を指す。第１所定個数は、第２所定個数未満の値に設定される。具体的には、第１所定個数及び第２所定個数は、実験等に基づき適切な値に設定される。この態様では、音声認識装置は、キーワードの候補が第１所定個数以下の場合、十分にキーワードが絞られており、第１パス処理手段の認識結果の信頼性が高いと判断する。一方、音声認識装置は、キーワードの候補が第２所定個数以上の場合、未知語の入力等に起因して第１パス処理手段の認識結果の信頼性が低いと判断する。このように、音声認識装置は、キーワードの候補数に基づき信頼度を決定することで、適切に第２パス処理を実行すべきか否かについて判定することができる。

[0016] 上記の音声認識装置の他の一態様では、前記第2パス実行可否判定手段は、S/N比が第1所定値より大きい場合に前記信頼度を第1の閾値より高く設定し、S/N比が第2所定値より小さい場合に前記信頼度を第2の閾値より低く設定する。第1所定値は、第2所定値以上の値に設定される。具体的には、第1所定値及び第2所定値は、実験等に基づき適切な値に設定される。一般に、S/N比と認識結果の正解率とは、相関関係を有する。従って、S/N比が高い場合、第1パス処理のみの認識結果でも十分である可能性が高い。一方、S/N比が低い場合、第2パス処理を実行しても正しく単語列を認識する可能性が低い。従って、音声認識装置は、S/N比に基づき信頼度を設定することで、適切に第2パス処理を実行すべきか否かについて判定することができる。

[0017] 上記の音声認識装置の他の一態様では、前記第2パス実行可否判定手段は、前記スコアのうち最良スコアと2番目に良いスコアとの差が所定値より大きい場合、前記信頼度を前記第1の閾値よりも高く設定する。上述の所定値は、実験等に基づき設定される。一般に、認識結果が正しい場合、上述のスコア差が大きくなる傾向がある。したがって、この態様では、音声認識装置は、信頼度を上述のスコア差に基づき設定することで、第2パス処理を実行すべきか否かについて適切に判定することができる。

[0018] 上記の音声認識装置の他の一態様では、前記第2パス実行可否判定手段は、発話スピード、音声の大きさ、及び突発性雑音の有無の少なくとも1つに基づき前記信頼度を決定する。一般に、発話スピード、音声の大きさ、及び突発性雑音の有無によって、認識結果の正解率は大きく左右される。従って、音声認識装置は、上述の各要素を考慮して信頼度を設定することで、第2パス処理を実行すべきか否かについて適切に判定することができる。

[0019] 上記の音声認識装置の他の一態様では、車両に搭載され、前記第2パス実行可否判定手段は、前記車両の状態を示す情報に基づき前記信頼度を決定する。車両の状態を示す情報とは、例えば、車速パルスに基づく走行速度、エアコンのオン又はオフの情報、窓が空いているか否かの情報等が挙げられる。

。このように、音声認識装置は、車両の状態を示す情報に基づき認識環境を適切に推定し、信頼度を適切に設定することができる。

[0020] 上記の音声認識装置の他の一態様では、前記第 1 パス処理手段と並行して実行され、前記発話信号に基づきサブワード単位の解析を行うことでスコアを算出するサブワード認識手段をさらに備え、前記第 2 パス実行可否判定手段は、サブワード認識手段により得られた最良スコアと第 1 パス処理手段により得られた最良スコアとのスコア差に基づき前記信頼度を決定する。この態様では、音声認識装置は、第 1 パス処理と並行してサブワード認識を行い、そのスコア差を監視することで、第 1 パス処理の認識結果が信頼できるか否か判断する。このようにすることで、音声認識装置は、適切に信頼度を設定し、第 2 パス処理を実行すべきか否かについて適切に判定することができる。

[0021] 上記の音声認識装置の他の観点では、1 又は複数の音響モデルを記憶する音響モデル記憶部と、1 又は複数の言語モデルを記憶する言語モデル記憶部と、を利用する音声認識方法であって、前記音響モデルと前記言語モデルに基づき、入力された発話信号から単語列の候補及びスコアを決定する第 1 パス処理工程と、前記第 1 パス処理手段の認識結果及び／又は認識環境の情報に基づき、第 2 パス処理を実行すべきか否か判定する第 2 パス実行可否判定工程と、前記第 2 パス実行可否判定工程が第 2 パス処理を実行すべきと判定した場合、前記音響モデル及び前記言語モデルに基づき、前記候補及び前記スコアを再決定する第 2 パス処理工程と、を備える。音声認識装置は、この方法を使用することで、適切に第 2 パス処理を実行すべきか否か判定し、不要に第 2 パス処理以降の認識処理を実行するのを抑制することができる。

[0022] 上記の音声認識装置の他の一態様では、1 又は複数の音響モデルを記憶する音響モデル記憶部と、1 又は複数の言語モデルを記憶する言語モデル記憶部と、を利用するコンピュータにより実行される音声認識プログラムであって、前記音響モデルと前記言語モデルに基づき、入力された発話信号から単語列の候補及びスコアを決定する第 1 パス処理手段と、前記第 1 パス処理手

段の認識結果及び／又は認識環境の情報に基づき、第２パス処理を実行すべきか否か判定する第２パス実行可否判定手段と、前記第２パス実行可否判定手段が第２パス処理を実行すべきと判定した場合、前記音響モデル及び前記言語モデルに基づき、前記候補及び前記スコアを再決定する第２パス処理手段と、を備える。音声認識装置は、このプログラムを搭載することで、適切に第２パス処理を実行すべきか否か判定し、不要に第２パス処理以降の認識処理を実行するのを抑制することができる。なお、好適な例では、上記プログラムは、記憶媒体に記録される。

実施例

[0023] 以下、図面を参照して本発明の好適な実施例について説明する。

[0024] [基本説明]

まず、本発明に係る音声認識装置の概要について図１及び図２を用いて説明する。

[0025] 図１は、言語モデルを用いた音声認識装置の概略構成図である。言語モデルを用いた音声認識装置は、ユーザによる発話を単語の組合せとして認識する。発話を単語の組合せとして認識し、テキスト化する処理を「ディクテーション」と呼ぶ。発話を単語の組合せとして認識することにより、予め用意された文章以外の文章、即ち、複数の単語を任意に組み合わせて作られた文章を認識することが可能となる。

[0026] 図１に示すように、音声認識装置は、ディクテーションを行うディクテーション部１０と、キーワード抽出部３０と、を備える。また、ディクテーション部１０は、第１パス実行処理部１１と、第２パス実行可否判定部１２と、第２パス実行処理部１３と、言語モデルを記憶する言語モデルデータベース２４（以下、「データベース」を「ＤＢ」と略記することがある。）と、音響モデルを記憶する音響モデルＤＢ２５と、を有する。図１に示すように、ディクテーション部１０は、マイク等を通じて入力された発話データ（以後、「発話データＳａ」と呼ぶ。）に対し音声認識処理を複数回行うマルチパス探索を行う。

[0027] ここで、「発話データ S a」とは、音声を含む入力信号を指す。たとえば、カーナビゲーション装置に実装されている音声認識装置の場合、発話データ S aはユーザが発話ボタンを押下してから一定時間の間にマイクから録音された入力信号を指す。

[0028] 音響モデル D B 2 5 は、音節や音素の単位の音の特徴を格納したデータベースである。発話に含まれる各単語の音の特徴は、音響モデルに記録されている音の特徴との比較により決定され、音響スコアとして算出される。また、音響モデル D B 2 5 は、精度が高い、即ちモデルの規模が大きい音響モデル（以後、「高精度音響モデル H s m」と呼ぶ。）と、高精度音響モデル H s mより精度が低い、即ちモデルの規模が小さい音響モデル（以後、「低精度音響モデル L s m」と呼ぶ。）と、を備える。音響モデル D B 2 5 は、本発明の音響モデル記憶部に相当する。

[0029] 言語モデル D B 2 4 は、隣り合う単語の組合せの出現確率を記憶したデータベースである。本発明では、例えば統計的言語モデルの一つである単語 N グラムモデルが使用される。言語モデル D B 2 4 は、精度が高い言語モデル（以後、「高精度言語モデル H l m」と呼ぶ。）と、高精度言語モデル H l mより精度が低い言語モデル（以後、「低精度言語モデル L l m」と呼ぶ。）を備える。言語モデルを用いて言語スコアが計算される。「言語スコア」とは、隣り合う単語の組合せの出現確率（出現頻度）を示す値である。言語モデル D B 2 4 は、本発明の言語モデル記憶部に相当する。

[0030] 第 1 パス実行処理部 1 1 は、低精度音響モデル L s mと低精度言語モデル L l mとに基づき、認識結果として単語列の候補及びこれに対応するスコア（トータルスコア）を出力する。トータルスコアについては後述する。第 1 パス実行処理部 1 1 は、発話データ S aの入力と並行して処理を行い、発話データ S aの入力の終了と同時に認識結果を出力する。第 1 パス実行処理部 1 1 の詳細な説明は、後述する図 2 の説明でさらに述べる。

[0031] 第 2 パス実行可否判定部 1 2 は、第 1 パス実行処理部 1 1 の認識結果と、認識を実行する環境に関する情報（以後、「認識環境情報 R i」と呼ぶ。）

と、に基づき、第２パス実行処理部１３により引き続き認識処理を行うか否か判定する。この処理については、後述する「実行可否判定処理」にて詳細を説明する。そして、第２パス実行可否判定部１２は、第２パス実行処理部１３で継続して認識処理を実行すべきと判断した場合、第１パス実行処理部１１の認識結果を第２パス実行処理部１３へ供給する。一方、第２パス実行可否判定部１２は、第２パス実行処理部１３で継続して認識処理を実行すべきでないとして判断した場合、第１パス実行処理部１１の認識結果をキーワード抽出部３０へ供給する。

[0032] 第２パス実行処理部１３は、高精度音響モデル Hsm と高精度言語モデル Hlm とに基づき、第１パス実行処理部１１で得られた単語列の候補のトータルスコアを再計算する。第２パス実行処理部１３は、上述したように、第２パス実行可否判定部１２が引き続き認識処理を実行すべきと判断した場合のみ認識処理を行う。なお、第２パス実行処理部１３は、発話データ Sa の入力終了後に実行される。

[0033] キーワード抽出部３０は、認識結果として得られたトータルスコアが最大となる単語列から所定のキーワードを抽出する。キーワードは予め決められており、キーワードと非キーワードとを識別する情報が辞書に記憶されている。例えば、辞書に記憶された多数の単語のうち、キーワードにはそれを示すキーワードフラグが付加されている。なお、好適には、本発明の音声認識手法が適用される機器の操作コマンドなどがキーワードとして設定される。

[0034] 次に、図２を参照して音声認識装置が実行する処理について引き続き説明する。

[0035] 図２は、音声認識装置が実行する処理のブロック図を示す。図２では、第１パス実行処理部１１は、音声区間切り出し部１１ａと、特徴パラメータ計算部１１ｂと、第１パスマッチング処理部１１ｃとを備える。第２パス実行処理部１３は、第２パスマッチング処理部１３ａを備える。また、認識結果出力処理部３１は、図１のキーワード抽出部３０に相当する。

[0036] 音声区間切り出し部１１ａは、発話データ Sa の中から音声区間を検出し

、音声区間内の音声データを出力する。つまり、「音声データ」とは、発話データ S a の中から音声に該当する区間だけが切り出されたものを指す。

[0037] 特徴パラメータ計算部 11 b は、音声区間切り出し部 11 a で切り出された音声データを単位時間毎に分割し、それぞれにおいて特徴パラメータを計算し、第 1 パスマッチング処理部 11 c へ供給する。

[0038] 次に、第 1 パスマッチング処理部 11 c は、単位時間毎に得られた特徴パラメータを低精度言語モデル L l m、低精度音響モデル L s m にあてはめることによって認識結果を出力するマッチング処理を行う。

[0039] 具体的には、第 1 パスマッチング処理部 11 c は、図示しない辞書 DB などに登録されている単語の組み合わせのうち、音声データに最も適合するものを、音声データの始端から時系列順に探索する。この探索により複数の単語列の候補（以下、「候補パターン」とも呼ぶ。）が作成される。ここで、音声データの始端から途中までをマッチング処理した結果、スコアの低かった組合せについては、以降マッチング処理しないようにする枝刈り処理も行う。第 1 パスマッチング処理部 11 c は、複数の候補パターンについて、音響スコア及び言語スコアを計算してトータルスコアを求める。

[0040] 第 1 パスマッチング処理部 11 c が出力する候補パターンの例について図 3 を用いて例示する。図 3 は、第 1 パスマッチング処理部 11 c の認識結果を表す単語グラフの例を示す。具体的には、図 3 は、ナビゲーション装置を操作する際に入力された発話データ S a の認識結果を単語グラフで表示したものである。図 3 では、横軸は時間軸を表し、白丸はノードを表す。また、図 3 (a) は、候補パターンが少ない場合、即ち、各ノードに終端する矢印の数が少ない場合を示し、図 3 (b) は、候補パターンが多い場合、即ち、各ノードに終端する矢印の数が多い場合を示す。第 1 パスマッチング処理部 11 c は、単位時間幅ごと（フレームごと）に得られた特徴パラメータを低精度言語モデル L l m、低精度音響モデル L s m にあてはめることによって、図 3 (a) または図 3 (b) に示すような候補パターンを生成する。そして、第 1 パスマッチング処理部 11 c が生成した候補パターンは、第 2 パス

実行可否判定部 12 に供給される。

[0041] 第 2 パス実行可否判定部 12 は、上述したように、第 1 パス実行処理部 11 の認識結果と、認識環境情報 R_i とに基づき、第 2 パス実行処理部 13 により認識処理を行うか否か判定する。そして、第 2 パス実行可否判定部 12 は、第 2 パス実行処理部 13 による処理が不要と判断した場合、第 1 パス実行処理部 11 の認識結果を認識結果出力処理部 31 へ供給する。一方、第 2 パス実行可否判定部 12 は、第 2 パス実行処理部 13 による処理が必要と判断した場合、第 1 パス実行処理部 11 の認識結果を第 2 パスマッチング処理部 13a に供給する。

[0042] 第 2 パスマッチング処理部 13a は、第 2 パス実行可否判定部 12 が認識処理を継続すべきと判断した場合に、第 1 パス実行処理部 11 で得られた候補パターンまたはトータルスコアが上位の候補パターンについて、高精度音響モデル H_{sm} と高精度言語モデル H_{lm} とを用いて、トータルスコアを再計算する。そして、第 2 パスマッチング処理部 13a は、再計算したトータルスコアに基づき最終結果として出力すべき単語列を確定し、その認識結果を認識結果出力処理部 31 に供給する。

[0043] 認識結果出力処理部 31 は、供給された認識結果に基づき、ディスプレイやスピーカなどの出力装置により、所定の画像または音声を出力する。

[0044] [実行可否判定処理]

次に、第 2 パス実行可否判定部 12 が実行する処理について詳しく説明する。第 2 パス実行可否判定部 12 は、第 1 パス実行処理部 11 の認識結果と認識環境情報 R_i とに基づき、第 1 パス実行処理部 11 の認識結果が信頼できるか否か判断するための信頼度（以後、「信頼度 T 」と呼ぶ。）を算出する。そして、第 2 パス実行可否判定部 12 は、信頼度 T が所定の閾値（以後、「第 1 の閾値 T_{th1} 」と呼ぶ。）より大きい場合、または、信頼度 T が所定の閾値（以後、「第 2 の閾値 T_{th2} 」と呼ぶ。）より小さい場合、第 2 パス実行処理部 13 の処理は不要であると判断する。このようにすることで、音声認識装置は、処理量を削減し、レスポンスを向上させる。

[0045] これについて図4を用いてさらに説明する。図4は、信頼度 T に基づき音声認識装置が実行する処理を示した図である。図4において、「第1パス処理」とは、第1パス実行処理部11が実行する処理を示し、「第2パス処理」とは、第2パス実行処理部13が実行する処理を示す。図4に示す第1の閾値 T_{th1} 及び第2の閾値 T_{th2} は、実験等により適切な値に設定される。なお、信頼度 T の算出方法については、別途詳しく説明する。

[0046] 図4に示すように、音声認識装置は、信頼度 T が第2の閾値 T_{th2} より小さい場合、第1パス処理のみを実行する。即ち、この場合、第2パス実行可否判定部12は、信頼度 T が第2の閾値 T_{th2} より小さいことから、第2パス処理を実行しても正しい認識結果を得ることができないと判断する。従って、この場合、第2パス実行可否判定部12は、第2パス処理を実行すべきではないと判断する。このように、第2の閾値 T_{th2} は、第2パス処理を実行することにより正しい認識結果を得る可能性がある信頼度 T の下限値に設定される。以上のように、音声認識装置は、信頼度 T が第2の閾値 T_{th2} より小さい場合には第1パス処理のみを実行することで、無駄な処理を削減し、レスポンスを向上させることができる。

[0047] また、音声認識装置は、信頼度 T が第1の閾値 T_{th1} より大きい場合、第1パス処理のみを実行する。即ち、この場合、第2パス実行可否判定部12は、信頼度 T が第1の閾値 T_{th1} より大きいことから、第1パス処理の認識結果は正しい可能性が高いと判断する。従って、この場合、第2パス実行可否判定部12は、第2パス処理を実行すべきではないと判断する。このように、第1の閾値 T_{th1} は、第2パス処理を実行することで、第1パス処理よりも精度の高い認識結果を得る可能性がある信頼度 T の上限値に設定される。以上のように、音声認識装置は、信頼度 T が第1の閾値 T_{th1} より大きい場合には第1パス処理のみを実行することで、不要な処理を削減し、レスポンスを向上させることができる。

[0048] 一方、音声認識装置は、信頼度 T が第2の閾値以上であり、かつ、第1の閾値以下の場合、第1パス処理に加えて、第2パス処理を実行する。即ち、

第２パス実行可否判定部１２は、第１パス処理の認識結果に基づき第２パス処理を実行することで、より精度の高い認識結果を得ることができると判断する。従って、第２パス実行可否判定部１２は、第２パス処理を実行すべきと判断する。このように、音声認識装置は、信頼度Ｔが第２の閾値以上であり、かつ、第１の閾値以下の場合に第２パス処理を実行することで、より精度の高い認識結果を得ることができる。

[0049] （信頼度の算出方法）

次に、信頼度Ｔの算出方法について具体例を用いて説明する。上述したように、信頼度Ｔは、第１パス実行処理部１１の認識結果と認識環境情報Ｒ_iとに基づき決定される。ここで、第２パス実行可否判定部１２は、後述する具体例でも示すように、第１パス実行処理部１１の認識結果として各候補パターンのトータルスコアの差、又は／及びキーワードの候補数などを使用する。また、第２パス実行可否判定部１２は、認識環境情報Ｒ_iとして、ＳＮ比などの音響情報や、その他発話データＳ_aの取得環境を推定可能な外部情報（以後、単に「外部情報」と呼ぶ。）を使用する。外部情報は、例えば音声認識装置が車両に搭載されていた場合には、車両から送信されるエアコンのオンまたはオフに関する情報及び走行速度に関する情報が該当する。

[0050] このように、第２パス実行可否判定部１２は、マイクなどの音声入力装置から音響情報を取得すると共に、音声認識装置が搭載されている機器又はその機器と電氣的に接続している機器等から外部情報を取得する。そして、第２パス実行可否判定部１２は、所定の式又はマップに基づき、これらの情報から信頼度Ｔを算出する。上述の式又はマップは、実験等により適切に作成され、音声認識装置のメモリ内に保持される。このようにすることで、音声認識装置は、適切に信頼度Ｔを設定し、これに基づき第２パス処理を実行すべきか否かについて判定することができる。

[0051] 以下、信頼度Ｔを高く設定する場合及び信頼度Ｔを低く設定する場合について具体例を示す。なお、以下に示す項目に複数該当する場合、第２パス実行可否判定部１２は、例えば、各項目間に予め定められた優先順位に従い最

も優先順位が高い項目に基づき信頼度 T を算出してもよく、または、所定の重み付け等を行うことにより各項目を加味して信頼度 T を算出してもよい。その他、第 2 パス実行可否判定部 12 は、上述した所定のマップ又は式に基づき信頼度 T を算出する。

[0052] 1. 信頼度 T を高く設定する例

以下、第 2 パス実行可否判定部 12 が信頼度 T を高く設定する例を以下の 1-1 乃至 1-3 に挙げる。

[0053] 1-1. SN 比

第 2 パス実行可否判定部 12 は、 SN 比が大きい場合、信頼度 T を高く設定する。例えば、第 2 パス実行可否判定部 12 は、 SN 比が所定値以上である場合には、第 1 パス処理のみを実行し、第 2 パス処理は実行する必要がないと判定する。上述の所定値は、実験等に基づき適切な値に設定される。

[0054] これについて補足説明する。一般に、音声認識率は SN 比によって変動することが知られている。従って、 SN 比が所定値より大きい場合、第 1 パス処理による認識率（正解率）は高いことが推定される。以上を考慮し、第 2 パス実行可否判定部 12 は、 SN 比が所定値以上である場合には、第 2 パス処理を実行する必要がないと判定する。これにより、第 2 パス実行可否判定部 12 は、不要な処理量を削減することができる。この場合、音声認識装置は、第 1 パス処理の認識結果のみに基づきトータルスコアが最大である候補パターンを出力すべき単語列として特定する。

[0055] 1-2. トータルスコア

第 2 パス実行可否判定部 12 は、第 1 パス処理により得られた候補パターンのうち、トータルスコアが最大となる候補パターンと、2 番目に大きくなる候補パターンとのトータルスコアの差分値が大きい程、信頼度 T を高く設定する。例えば、第 2 パス実行可否判定部 12 は、上述の差分値が所定値以上である場合、信頼度 T を第 1 の閾値 T_{th1} より高く設定する。上述の所定値は、実験等に基づき適切な値に設定される。

[0056] これについて補足説明する。一般に、トータルスコアが最大となる候補パ

ターンが正しい場合、当該候補パターンのトータルスコアと他の候補パターンのトータルスコアとの差が大きいという傾向がある。以上を考慮し、第2パス実行可否判定部12は、トータルスコアが最大となる候補パターンと2番目に大きくなる候補パターンとのトータルスコアの差分値が所定値以上の場合には、信頼度 T を第1の閾値 T_{th1} より高く設定する。これにより、第2パス実行可否判定部12は、第2パス処理を実行する必要があると判定し、不要な処理量を削減することができる。この場合、音声認識装置は、第1パス処理の認識結果のみに基づき、トータルスコアが最大である候補パターンを出力すべき単語列として特定する。

[0057] 1-3. キーワードの候補数

第2パス実行可否判定部12は、第1パス処理により得られたキーワードに相当する単語の候補が少ない程、信頼度 T を高く設定する。例えば、第2パス実行可否判定部12は、上述の候補が所定値（例えば1）以下の場合、信頼度 T を第1の閾値 T_{th1} より高く設定する。上述の所定値は、実験等に基づき適切な値に設定される。

[0058] これについて、図3の例を用いて説明する。図3（a）の例では、キーワードに候補として「200メータスケール」が一意に認識されている。即ち、第1パス処理のみによって目的となるキーワードが一意に認識されている。従って、この場合、第2パス実行可否判定部12は、信頼度 T を第1の閾値 T_{th1} より高く設定する。一方、図3（b）の例では、キーワードに相当する候補として「100メータスケール」、「200メータスケール」、「500メータスケール」というように複数の候補が認識されている。即ち、第1パス処理によって多数の候補が認識されている。この場合、第2パス実行可否判定部12は、信頼度 T を第1の閾値 T_{th1} より低く設定する。これにより、音声認識装置は、不要な処理量を削減し、結果出力までの処理速度を向上させることができる。

[0059] 2. 信頼度 T を低く設定する例

以下、第2パス実行可否判定部12が信頼度 T を低い値に設定する例を2

－ 1 乃至 2 － 3 に示す。

[0060] 2 － 1. 音響上の要因

第 2 パス実行可否判定部 1 2 は、上述したように、音響情報に基づき信頼度 T を設定する。例えば、第 2 パス実行可否判定部 1 2 は、音響上の要因に起因して認識率が極めて低くなる可能性が高いと判断した場合、信頼度 T を第 2 の閾値 T_{th2} より低い値に設定する。以下、音響上の要因として主要な例を挙げる。

[0061] 2 － 1 － 1. SN 比

第 2 パス実行可否判定部 1 2 は、SN 比が低い程、信頼度 T を低い値に設定する。例えば、第 2 パス実行可否判定部 1 2 は、SN 比が所定値よりも低い場合、第 1 パス処理の認識結果は信頼性が極めて低いと判断し、信頼度 T を第 2 の閾値 T_{th2} より低い値に設定する。所定値は、実験等に基づき適切な値に設定される。即ち、この場合、第 2 パス実行可否判定部 1 2 は、第 2 パス処理を実行しても認識結果を向上することはできないと判断し、第 2 パス処理を実行しないと判定する。これにより、音声認識装置は、無駄な処理量を削減することができる。

[0062] 2 － 1 － 2. 発話スピード

第 2 パス実行可否判定部 1 2 は、検出された発話スピードと想定している発話スピードとの差が大きい程信頼度 T を低い値に設定する。例えば、第 2 パス実行可否判定部 1 2 は、発話のスピードが所定速度よりも速い場合（「第 1 所定速度」と呼ぶ。）、又は発話のスピードが所定速度（「第 2 所定速度」と呼ぶ。）よりも遅い場合、信頼度 T を第 2 の閾値 T_{th2} より低い値に設定する。第 1 及び第 2 所定速度は、実験等に基づき適切な値に設定される。即ち、この場合、第 2 パス実行可否判定部 1 2 は、音響モデル L_{sm} 、 H_{sm} で想定されている発話スピードと差があり、正しい認識結果が得られる可能性が極めて低いと判断する。従って、この場合、音声認識装置は、信頼度 T を第 2 の閾値 T_{th2} より低い値に設定することで、無駄な処理量を削減することができる。

[0063] なお、第2パス実行可否判定部12は、例えば発話データS_aの入力時間幅を認識した文字数で除することで、発話スピードを算出する。

[0064] 2-1-3. 音声の大きさ

第2パス実行可否判定部12は、入力された音声の大きさ（即ち入力された音声データの信号レベル）と想定している音声の大きさととの差が大きい程、信頼度Tを低い値に設定する。例えば、第2パス実行可否判定部12は、音声の大きさが所定値（「第1所定値」と呼ぶ。）より大きい場合、または、所定値（「第2所定値」と呼ぶ。）より小さい場合、信頼度Tを第2の閾値T_{th2}より低い値に設定する。第1及び第2所定値は、実験等に基づき適切な値に設定される。即ち、この場合、第2パス実行可否判定部12は、各モデルで想定されている音声の大きさと差があり、第2パス処理を実行しても正しい認識結果が得られる可能性が極めて低いと判断する。これにより、音声認識装置は、信頼度Tを適切に設定し、無駄な処理量を削減することができる。

[0065] 2-1-4. 突発性雑音の有無

第2パス実行可否判定部12は、発話データS_aに非定常性の雑音、即ち、突発的な雑音が含まれる場合、信頼度Tを低い値に設定する。例えば、第2パス実行可否判定部12は、発話データS_aに非定常性の雑音が含まれる場合、第2の閾値T_{th2}より低い値に設定する。他の例として、第2パス実行可否判定部12は、発話データS_a中に所定個数以上の非定常性雑音が含まれていた場合、信頼度Tを第2の閾値T_{th2}より低い値に設定する。上述の所定個数は、実験等に基づき適切な値に設定される。これによっても、音声認識装置は、信頼度Tを適切に設定し、無駄な処理量を削減することができる。

[0066] 2-2. キーワードの候補数

第2パス実行可否判定部12は、第1パス処理により得られたキーワードに相当する単語の候補が多い程、信頼度Tを低い値に設定する。例えば、第2パス実行可否判定部12は、上述の候補が所定値以上の場合、信頼度Tを

第2の閾値 T_{th2} より低く設定する。上述の所定値は、実験等に基づき適切な値に設定される。一般に、辞書DB等に登録されていない未知語が入力された場合、単語の候補数が多くなる傾向がある。従って、第2パス実行可否判定部12は、キーワードに相当する単語の候補が所定値以上の場合、信頼度 T を第2の閾値 T_{th2} より低く設定することで、無駄な処理量を削減することができる。

[0067] これについて、図3(b)の例を用いて具体的に説明する。図3(b)の例では、キーワードの候補は、「500メータスケール」、「100メータスケール」、「200メータスケール」の3つが存在している。従って、この場合、第2パス実行可否判定部12は、上述の所定値を3と設定していた場合には、信頼度 T を第1の閾値 T_{th1} より低い値に設定する。

[0068] 2-3. 外部要因

第2パス実行可否判定部12は、取得した外部情報に基づき正しい認識結果が得られる可能性が低いことが推定される場合、信頼度 T を低い値に設定する。これによっても、第2パス実行可否判定部12は、適切に信頼度 T を設定することができる。

[0069] 以下、音声認識装置が車両に搭載されている場合を例にして説明する。第2パス実行可否判定部12は、車両からエアコンの作動の有無、車両の走行速度、車両に備わる窓の開閉に関する情報を取得する。そして、第2パス実行処理部13は、これらの情報に基づき信頼度 T を決定する。例えば、第2パス実行処理部13は、エアコンが作動している場合、又は／及び、走行速度が大きく窓が開いている場合、信頼度 T を第2の閾値 T_{th2} より低い値に設定する。

[0070] このように、第2パス実行処理部13は、外部情報に基づき適切に信頼度 T を設定することができる。

[0071] 以上説明したように、本実施例による音声認識装置は、1又は複数の音響モデルを記憶する音響モデル記憶部と、1又は複数の言語モデルを記憶する言語モデル記憶部と、低精度音響モデルと低精度言語モデルに基づき、入力

された発話信号から単語列の候補及びトータルスコアを決定する第１パス実行処理部と、第１パス実行処理部の認識結果及び／又は認識環境の情報に基づき、第２パス処理を実行すべきか否か判定する第２パス実行可否判定部と、第２パス実行可否判定部が第２パス処理を実行すべきと判定した場合、高精度音響モデル及び高精度言語モデルに基づき、単語列の候補及びスコアを再決定する第２パス実行処理部と、を備える。このように、音声認識装置は、認識結果及び／又は認識環境の情報に基づき、適切に第２パス処理を実行すべきか否か判定することで、不要に第２パス処理を実行するのを抑制することができる。従って、音声認識装置は、処理量を削減し、結果出力までの処理速度を向上させることができる。

[0072] [処理フロー]

次に、実施例における処理の手順について説明する。図５は、本実施例において音声認識装置が実行する処理の手順を表すフローチャートの一例である。音声認識装置は、図５に示すフローチャートの処理を発話データＳ_aが入力された際に繰り返し実行する。

[0073] まず、音声認識装置は、第１パス処理を実行する（ステップＳ１０１）。具体的には、音声区間切り出し部１１ａは、発話データＳ_aから音声データを切り出す。そして、特徴パラメータ計算部１１ｂは、音声区間切り出し部１１ａで切り出された音声データを単位時間毎に分割し、それぞれにおいて特徴パラメータを計算する。そして、第１パスマッチング処理部１１ｃは、単位時間毎に得られた特徴パラメータを低精度言語モデルＬ_l_m、低精度音響モデルＬ_s_mにあてはめることによって候補パターン及びトータルスコアを出力する。

[0074] 次に、音声認識装置は、第２パス処理を実行すべきか否か判定する（ステップＳ１０２）。具体的には、第２パス実行可否判定部１２は、第１パス処理による認識結果及び認識環境情報Ｒ_iに基づき、信頼度Ｔを算出する。そして、第２パス実行可否判定部１２は、信頼度Ｔに基づき、第２パス処理を実行すべきか否か判定する。

[0075] そして、第2パス処理を実行すべきと判定した場合（ステップS102；Yes）、即ち、信頼度Tが第1の閾値T_{th1}以下かつ第2の閾値T_{th2}以上の場合、音声認識装置は、第2パス処理を実行する（ステップS103）。具体的には、第2パスマッチング処理部13aは、第1パス実行処理部11で得られた候補パターンまたはトータルスコアが上位の候補パターンについて、高精度音響モデルH_{sm}と高精度言語モデルH_{lm}とを用いて、トータルスコアを再計算する。

[0076] 一方、第2パス処理を実行すべきでないと判定した場合（ステップS102；No）、即ち、信頼度Tが第2の閾値T_{th2}より小さい場合または信頼度Tが第1の閾値T_{th1}より大きい場合、音声認識装置は、ステップS104へ処理を進める。これにより、音声認識装置は、不要な処理を削減し、レスポンスを向上させることができる。

[0077] そして、音声認識装置は、認識結果を出力する（ステップS104）。即ち、音声認識装置は、認識結果として得られた単語列を合成音声により出力又はディスプレイ上に出力する。また、音声認識装置は、必要に応じて認識結果として得られた単語列からキーワードを抽出する。

[0078] [変形例1]

上述の実施例の説明では、第2パス実行可否判定部12は、第1パス処理の認識結果及び認識環境情報R_iに基づき、信頼度Tを設定した。しかし、本発明が適用可能な方法はこれに限定されない。これに代えて、またはこれに加えて、第2パス実行可否判定部12は、第1パス処理と並行して音節、音素単位のサブワード認識を行い、サブワード認識により得られたスコアと第1パス処理により得られたスコアとのスコア差（以後、単に「スコア差」と呼ぶ。）に基づき信頼度Tを設定してもよい。これにより、第2パス実行可否判定部12は、第2パス処理を実行すべきか否かについてより適切に判断をすることができる。

[0079] これについて、図6を用いて具体的に説明する。図6は、変形例1に係る音声認識装置のブロック図の一例である。図6に示すように、音声認識装置

は、サブワード認識処理部 4 1 と、音響モデル DB 4 2 と、を備える。

[0080] サブワード認識処理部 4 1 は、発話データ S_a に含まれる音声データを、音響モデル DB 4 2 に格納される音響モデルに基づきサブワード単位で解析を行い、音声データ全体を評価する。そして、サブワード認識処理部 4 1 は、所定のスコアを算出する。サブワード認識処理部 4 1 は、認識結果を第 2 パス実行可否判定部 1 2 へ供給する。

[0081] 音響モデル DB 4 2 は、サブワード認識を実行するための音響モデルを格納する。この種の音響モデルとして、例えばフィラーモデルなどの音響モデルが挙げられる。

[0082] 第 2 パス実行可否判定部 1 2 は、サブワード認識処理部 4 1 により得られた最良のスコアと第 1 パス処理により得られた最良のトータルスコアとのスコア差を算出する。そして、第 2 パス実行可否判定部 1 2 は、スコア差に基づき信頼度 T を設定する。

[0083] 例えば、第 2 パス実行可否判定部 1 2 は、スコア差が所定の閾値よりも大きい場合、信頼度 T を第 2 の閾値 T_{th2} より低い値に設定する。上述の閾値は、実験等に基づき適切な値に設定される。即ち、この場合、第 2 パス実行可否判定部 1 2 は、第 1 パス処理により得られた認識結果は信頼性が低く、第 2 パス処理を実行しても正しい結果を得られる可能性は低いと判断し、第 2 パス処理は実行しない。一方、第 2 パス実行可否判定部 1 2 は、スコア差が所定の閾値より小さい場合、信頼度 T を第 1 の閾値 T_{th1} より高い値に設定する。即ち、この場合、第 2 パス実行可否判定部 1 2 は、第 1 パス処理により得られた認識結果は信頼性が高いと判断し、第 2 パス処理は実行しない。

[0084] 以上のように、第 2 パス実行可否判定部 1 2 は、第 1 パス処理と並行してサブワード認識を行い、サブワード認識のスコアと第 1 パス処理のトータルスコアとのスコア差とに基づき信頼度 T を設定することで、不要な処理を削減することができる。

[0085] なお、図 6 の説明では、サブワード認識処理部 4 1 は、ディクテーション

部 10 が使用する音響モデルとは別の音響モデルを使用していたが、これに代えて、ディクテーション部 10 が使用する音響モデルと同一の音響モデルを使用してもよい。これにより、音声認識装置は、使用するメモリ量を削減することができる。

[0086] [変形例 2]

上述の実施例の説明では、音声認識装置は、第 1 パス実行処理部 11 と第 2 パス実行処理部 13 とによる 2 パス探索方式を実行していた。即ち、音声認識装置は、認識処理を 2 回実行していた。しかし、本発明が適用可能な方法はこれに限定されない。これに代えて、音声認識装置は、認識処理を 3 回以上実行してもよい。

[0087] この場合であっても、音声認識装置は、上述の実施例の構成に加えて、第 2 パス実行処理部 13 以降の各認識処理部間に、次の認識処理を実行すべきか否かを入力された認識結果及び認識環境情報 R_i に基づき判定する判定部を設ける。このとき、各認識処理部は、例えば後続の認識処理部ほど精度が高い言語モデル及び音響モデルを用いる。また、言語モデル DB 24 と音響モデル DB 25 は、例えば認識処理の個数に応じて精度の異なる言語モデル又は音響モデルを備える。そして、音声認識装置は、当該判定部が次の認識処理を実行すべきと判定した場合に限り後続の認識処理を実行する。このように、音声認識装置は、認識処理を 3 回以上実行する方式であっても、本発明を適用することで、不要な処理を削減し、レスポンスを向上させることができる。

[0088] [変形例 3]

上述の実施例の 1-3 及び 2-2 での説明では、第 2 パス実行可否判定部 12 は、信頼度 T を決定する方法の 1 つとして、キーワードの候補の数に基づき信頼度 T を決定していた。これに代えて、第 2 パス実行可否判定部 12 は、単語グラフの各ノードに終端する単語数に基づき信頼度 T を決定してもよい。

[0089] これについて、具体的に説明する。第 2 パス実行可否判定部 12 は、図 3

に示すような単語グラフから各ノードに終端する単語数、即ち、各ノードへ入力される矢印の数を算出する。そして、第2パス実行可否判定部12は、各ノードに終端する単語数の平均値又はこれに相当する計算値が所定の閾値より小さい場合、信頼度Tを第1の閾値 T_{th1} より高い値に設定する。一方、第2パス実行可否判定部12は、各ノードに終端する単語数の平均値等が、上述の閾値より大きい値に設定される所定の閾値より大きい場合、信頼度Tを第2の閾値 T_{th2} より低い値に設定する。これらの閾値は、実験等に基づき適切な値に設定される。これによっても、第2パス実行可否判定部12は、信頼度Tを適切に設定することができ、不要な処理を削減することができる。

[0090] [変形例4]

上述の実施例の説明では、第2パス実行可否判定部12は、第1パス処理の認識結果及び認識環境情報 R_i に基づき、信頼度Tを設定した。これに代えて、第2パス実行可否判定部12は、第1パス処理の認識結果または認識環境情報 R_i のいずれか一方に基づき、信頼度Tを設定してもよい。また、上述したように、第2パス実行可否判定部12は、認識環境情報 R_i を用いて信頼度Tを設定する場合、先に例示した複数の音響情報又は外部情報のうちいずれか1つ又は複数の情報に基づき信頼度Tを設定してもよい。

[0091] [変形例5]

上述の実施例の説明では、第2パス実行処理部13は、第1パス実行処理部11が使用する低精度音響モデル L_{sm} 及び低精度言語モデル L_{lm} より精度が高い高精度音響モデル H_{sm} 及び高精度言語モデル H_{lm} を使用した。しかし、本発明が適用可能な方法はこれに限定されない。これに代えて、第2パス実行処理部13は、第1パス実行処理部11が使用する音響モデル及び言語モデルと同一の音響モデル及び言語モデルを使用してもよい。この場合、言語モデルDB24と音響モデルDB25は、少なくとも1つの言語モデル又は音響モデルを備える。

[0092] [適用分野]

本発明は、音声認識処理を行う各種の機器に適用することができる。例えば、カーナビゲーション装置、携帯電話、パーソナルコンピュータ、ＡＶ機器、家電製品など、音声入力機能を備える各種の機器に適用することができる。

符号の説明

- [0093]
- １０ ディクテーション部
 - １１ 第１パス実行処理部
 - １２ 第２パス実行可否判定部
 - １３ 第２パス実行処理部
 - ２４ 言語モデルＤＢ
 - ２５、４２ 音響モデルＤＢ
 - ３０ キーワード抽出部
 - ３１ 認識結果出力処理部
 - ４１ サブワード認識処理部

請求の範囲

- [請求項1] 1又は複数の音響モデルを記憶する音響モデル記憶部と、
 1又は複数の言語モデルを記憶する言語モデル記憶部と、
 前記音響モデルと前記言語モデルに基づき、入力された発話信号から単語列の候補及びスコアを決定する第1パス処理手段と、
 前記第1パス処理手段の認識結果及び／又は認識環境の情報に基づき、第2パス処理を実行すべきか否か判定する第2パス実行可否判定手段と、
 前記第2パス実行可否判定手段が第2パス処理を実行すべきと判定した場合、前記音響モデル及び前記言語モデルに基づき、前記候補及び前記スコアを再決定する第2パス処理手段と、
 を備えることを特徴とする音声認識装置。
- [請求項2] 前記第2パス処理手段は、前記第2パス実行可否判定手段が前記第2パス処理を実行すべきと判定した場合、前記第1パス処理手段で使
 用された音響モデル及び言語モデル以上の精度を具備する音響モデル
 及び言語モデルに基づき、前記候補及び前記スコアを再決定すること
 を特徴とする請求項1に記載の音声認識装置。
- [請求項3] 前記第2パス実行可否判定手段は、前記認識結果及び／又は前記情
 報に基づき前記認識結果の信頼度を算出し、当該信頼度が第1の閾値
 よりも高い場合又は第2の閾値よりも低い場合、前記第2パス処理を
 継続すべきではないと判定する請求項1又は2に記載の音声認識装置
 。
- [請求項4] 前記第2パス実行可否判定手段は、前記第1パス処理により決定さ
 れた前記単語列を構成する各単語の候補数に基づき前記信頼度を決定
 することを特徴とする請求項3に記載の音声認識装置。
- [請求項5] 前記第2パス実行可否判定手段は、前記第1パス処理により決定さ
 れたキーワードの候補数が第1所定個数以下の場合に前記信頼度を第
 1の閾値より高い値に設定し、当該候補数が第2所定個数以上の場合

に前記信頼度を第2の閾値より低い値に設定することを特徴とする請求項4に記載の音声認識装置。

[請求項6] 前記第2パス実行可否判定手段は、S/N比が第1所定値より大きい場合に前記信頼度を第1の閾値より高く設定し、S/N比が第2所定値より小さい場合に前記信頼度を第2の閾値より低く設定する請求項3乃至5のいずれか一項に記載の音声認識装置。

[請求項7] 前記第2パス実行可否判定手段は、前記スコアのうち最良スコアと2番目に良いスコアとの差が所定値より大きい場合、前記信頼度を前記第1の閾値よりも高く設定することを特徴とする請求項3乃至6のいずれか一項に記載の音声認識装置。

[請求項8] 前記第2パス実行可否判定手段は、発話スピード、音声の大きさ、及び突発性雑音の有無の少なくとも1つに基づき前記信頼度を決定することを特徴とする請求項3乃至7のいずれか一項に記載の音声認識装置。

[請求項9] 車両に搭載され、
前記第2パス実行可否判定手段は、前記車両の状態を示す情報に基づき前記信頼度を決定する請求項3乃至8のいずれか一項に記載の音声認識装置。

[請求項10] 前記第1パス処理手段と並行して実行され、前記発話信号に基づきサブワード単位の解析を行うことでスコアを算出するサブワード認識手段をさらに備え、

前記第2パス実行可否判定手段は、サブワード認識手段により得られた最良スコアと第1パス処理手段により得られた最良スコアとのスコア差に基づき前記信頼度を決定する請求項3乃至9のいずれか一項に記載の音声認識装置。

[請求項11] 1又は複数の音響モデルを記憶する音響モデル記憶部と、1又は複数の言語モデルを記憶する言語モデル記憶部と、を利用する音声認識方法であって、

前記音響モデルと前記言語モデルに基づき、入力された発話信号から単語列の候補及びスコアを決定する第1パス処理工程と、

前記第1パス処理手段の認識結果及び／又は認識環境の情報に基づき、第2パス処理を実行すべきか否か判定する第2パス実行可否判定工程と、

前記第2パス実行可否判定工程が第2パス処理を実行すべきと判定した場合、前記音響モデル及び前記言語モデルに基づき、前記候補及び前記スコアを再決定する第2パス処理工程と、
を備えることを特徴とする音声認識方法。

[請求項12]

1又は複数の音響モデルを記憶する音響モデル記憶部と、1又は複数の言語モデルを記憶する言語モデル記憶部と、を利用するコンピュータにより実行される音声認識プログラムであって、

前記音響モデルと前記言語モデルに基づき、入力された発話信号から単語列の候補及びスコアを決定する第1パス処理手段と、

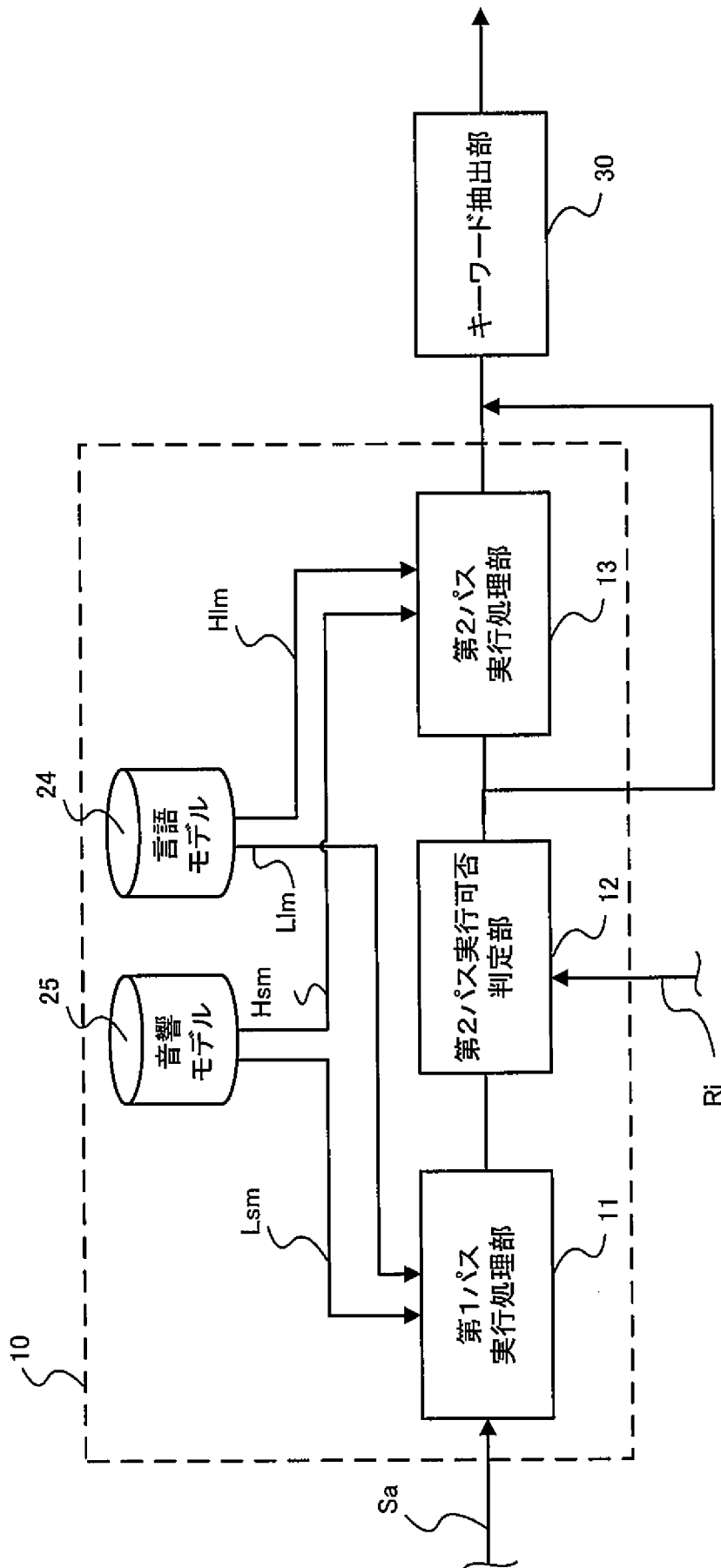
前記第1パス処理手段の認識結果及び／又は認識環境の情報に基づき、第2パス処理を実行すべきか否か判定する第2パス実行可否判定手段と、

前記第2パス実行可否判定手段が第2パス処理を実行すべきと判定した場合、前記音響モデル及び前記言語モデルに基づき、前記候補及び前記スコアを再決定する第2パス処理手段と、
を備えることを特徴とする音声認識プログラム。

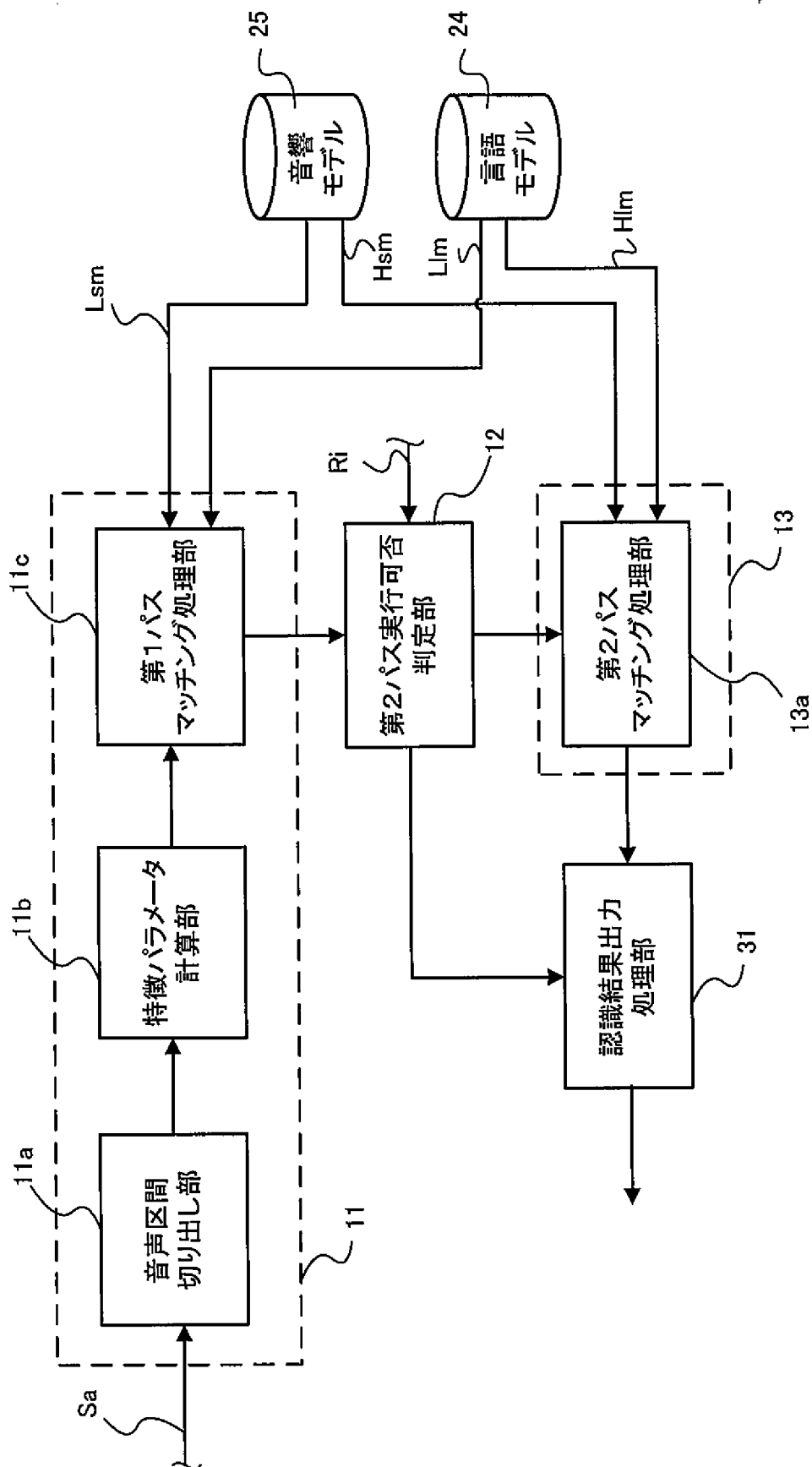
[請求項13]

請求項12に記載のプログラムを記憶したことを特徴とする記憶媒体。

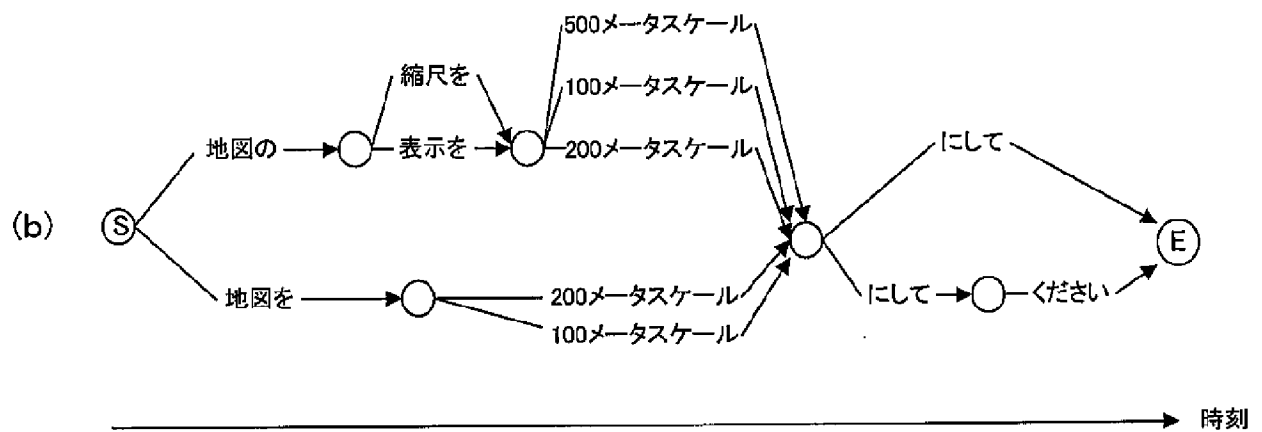
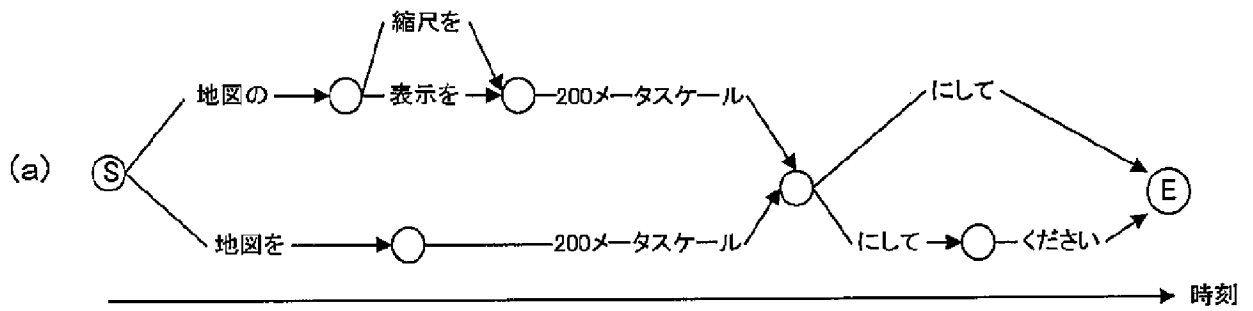
[図1]



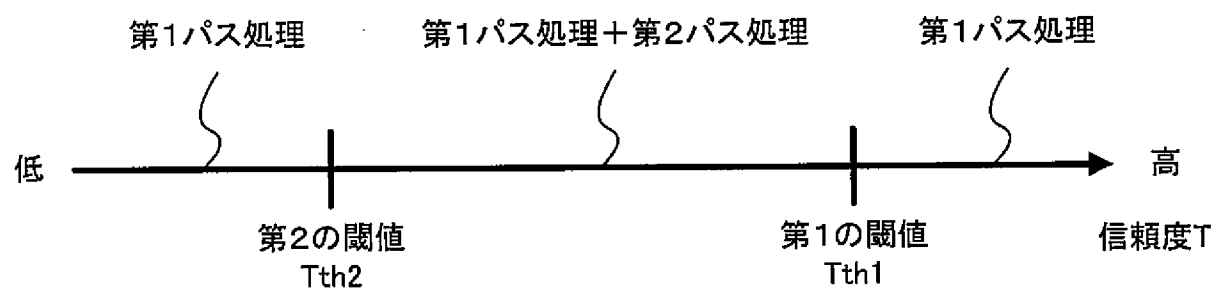
[図2]



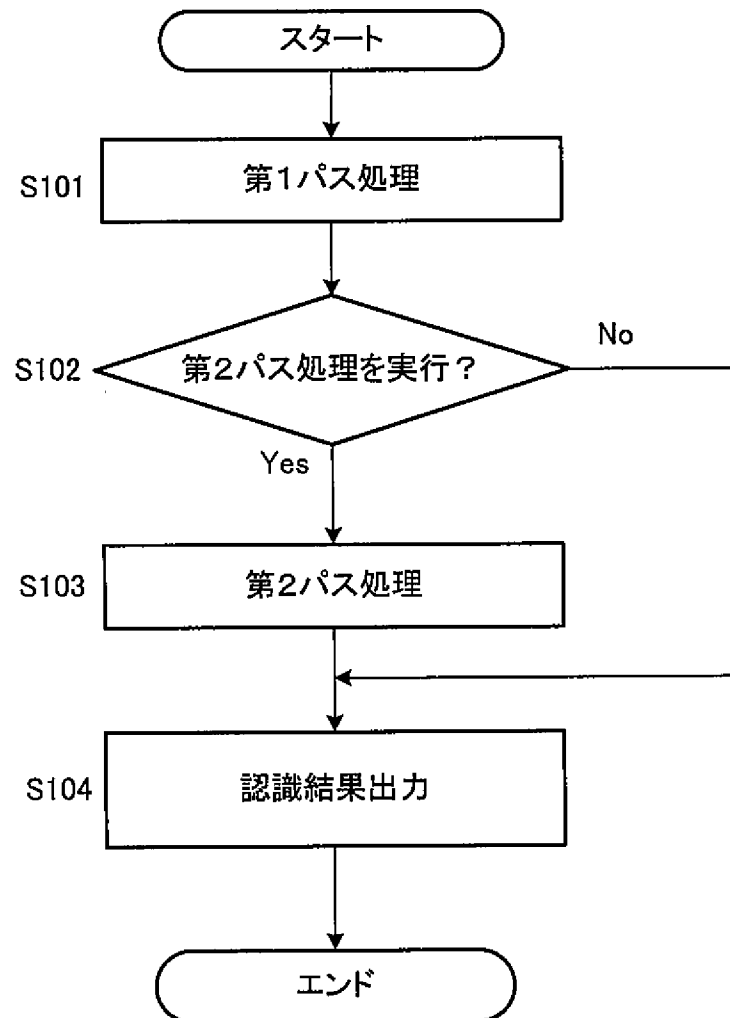
[図3]



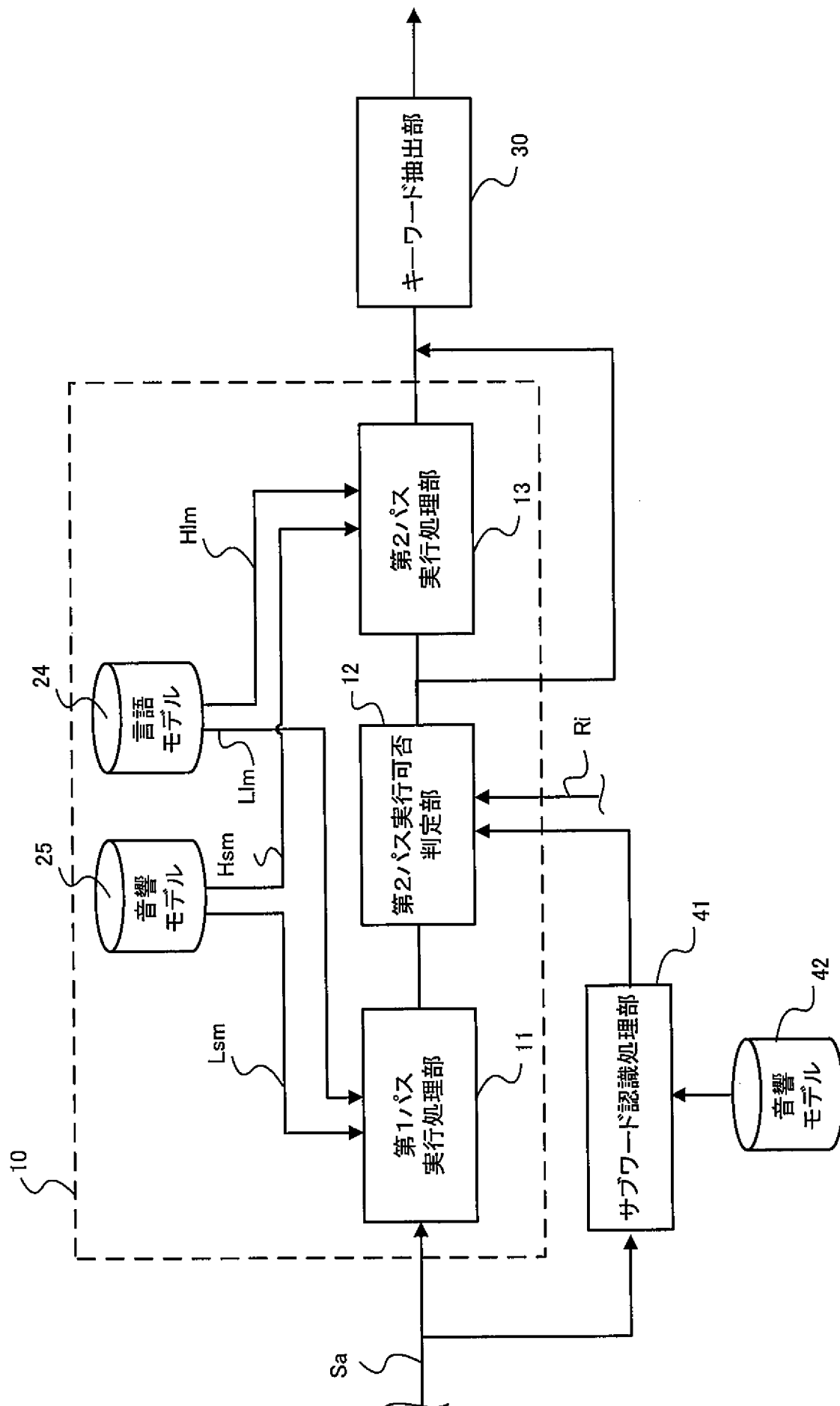
[図4]



[図5]



[図6]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/058707

A. CLASSIFICATION OF SUBJECT MATTER

G10L15/28 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L15/28

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2009
Kokai Jitsuyo Shinan Koho	1971-2009	Toroku Jitsuyo Shinan Koho	1994-2009

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	JP 2008-9153 A (Xanavi Informatics Corp., Nissan Motor Co., Ltd.), 17 January, 2008 (17.01.08), Full text; all drawings	1-3, 8, 11-13 4, 7, 9
X	JP 2002-6878 A (Sony International (Europe) GmbH), 11 January, 2002 (11.01.02), Full text; all drawings	1, 2
Y	JP 2007-108407 A (NEC Corp.), 26 April, 2007 (26.04.07), Par. No. [0081]	4, 7
Y	JP 2006-30908 A (Honda Motor Co., Ltd.), 02 February, 2006 (02.02.06), Par. Nos. [0046] to [0070]	9



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
11 June, 2009 (11.06.09)

Date of mailing of the international search report
23 June, 2009 (23.06.09)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/058707

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2001-92496 A (Nippon Hoso Kyokai), 06 April, 2001 (06.04.01), Full text; all drawings	1-13
A	JP 7-160822 A (Ricoh Co., Ltd.), 23 June, 1995 (23.06.95), Full text; all drawings	1-13
A	JP 3-266898 A (Fujitsu Ltd.), 27 November, 1991 (27.11.91), Full text; all drawings	1-13

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2009/058707

JP 2008-9153 A	2008.01.17	(Family: none)
JP 2002-6878 A	2002.01.11	EP 1162602 A1 US 2002/013706 A1 DE 60016722 T2
JP 2007-108407 A	2007.04.26	WO 2007/043566 A1
JP 2006-30908 A	2006.02.02	(Family: none)
JP 2001-92496 A	2001.04.06	US 6393398 B1
JP 7-160822 A	1995.06.23	(Family: none)
JP 3-266898 A	1991.11.27	(Family: none)

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G10L15/28(2006.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（I P C））

Int.Cl. G10L15/28

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 0 9 年
日本国実用新案登録公報	1 9 9 6 - 2 0 0 9 年
日本国登録実用新案公報	1 9 9 4 - 2 0 0 9 年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
X Y	JP 2008-9153 A(株式会社ザナヴィ・インフォマティクス、日産自動車株式会社)2008.01.17, 全文, 全図	1-3, 8, 11-13 4, 7, 9
X	JP 2002-6878 A(ソニー インターナショナル (ヨーロッパ) ゲゼル シャフト ミット ベシュレンクテル ハフツング) 2002.01.11, 全文, 全図	1, 2
Y	JP 2007-108407 A(日本電気株式会社) 2007.04.26, 段落【0081】	4, 7

☒ C欄の続きにも文献が列挙されている。☒ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日

1 1 . 0 6 . 2 0 0 9

国際調査報告の発送日

2 3 . 0 6 . 2 0 0 9

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）

郵便番号 1 0 0 - 8 9 1 5

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

5 Z

9 5 6 8

涌井 智則

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 4 1

C (続き) . 関連すると認められる文献		
引用文献の カテゴリ*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y	JP 2006-30908 A(本田技研工業株式会社) 2006.02.02, 段落【0046】－【0070】	9
A	JP 2001-92496 A(日本放送協会) 2001.04.06, 全文, 全図	1-13
A	JP 7-160822 A(株式会社リコー) 1995.06.23, 全文, 全図	1-13
A	JP 3-266898 A(富士通株式会社) 1991.11.27, 全文, 全図	1-13

JP 2008-9153 A	2008.01.17	ファミリーなし
JP 2002-6878 A	2002.01.11	EP 1162602 A1 US 2002/013706 A1 DE 60016722 T2
JP 2007-108407 A	2007.04.26	WO 2007/043566 A1
JP 2006-30908 A	2006.02.02	ファミリーなし
JP 2001-92496 A	2001.04.06	US 6393398 B1
JP 7-160822 A	1995.06.23	ファミリーなし
JP 3-266898 A	1991.11.27	ファミリーなし