

(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
2011年2月10日(10.02.2011)

PCT

(10) 国際公開番号
WO 2011/016129 A1

- (51) 国際特許分類:
G10L 15/06 (2006.01) G10L 15/22 (2006.01)
G10L 15/18 (2006.01)
- (21) 国際出願番号: PCT/JP2009/063996
- (22) 国際出願日: 2009年8月7日(07.08.2009)
- (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): 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).

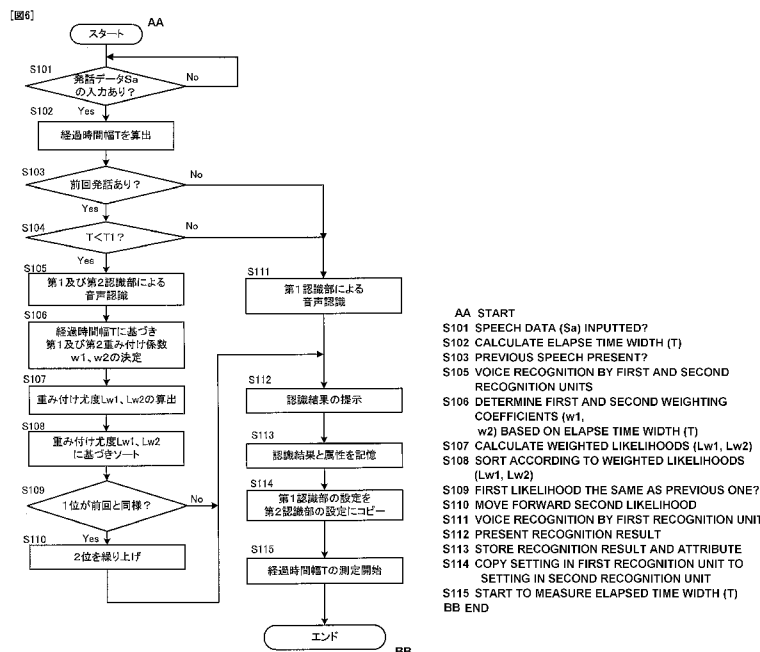
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, 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), ユーロシア (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, SM, 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



(57) Abstract: A voice recognition device voice-recognizes a hierarchical command and comprises a dictionary storage means, a first voice recognition means, a second voice recognition means, and a recognition result processing means. The dictionary storage means stores dictionaries used for voice recognition in each hierarchy for recognizing the hierarchical command. The first voice recognition means, when voice data is inputted, performs a voice recognition based on a dictionary to be used when it is assumed that previously inputted voice data has been correctly recognized. The second voice recognition means, when voice data is inputted, performs a voice recognition based on a dictionary used when the previously inputted voice data has been voice-recognized. The recognition result processing means weights the recognition result of the first voice recognition means and the recognition result of the second voice recognition means based on the elapse time width elapsed from the time when the previously inputted voice data has been processed and then determines the final recognition result.

(57) 要約:

[続葉有]

WO 2011/016129 A1



音声認識装置は、階層コマンドを音声認識し、辞書記憶手段と、第1音声認識手段と、第2音声認識手段と、認識結果処理手段と、を備える。辞書記憶手段は、階層コマンドを認識するための各階層の音声認識で使用する辞書を記憶する。第1音声認識手段は、音声データの入力があった場合、前回入力された音声データを正しく認識したと仮定した場合に使用すべき辞書に基づき音声認識を行う。第2音声認識手段は、音声データの入力があった場合、前回入力された音声データの音声認識を実行した際に使用した辞書に基づき音声認識を行う。認識結果処理手段は、第1音声認識手段の認識結果と、第2音声認識手段の認識結果と、を前回入力された音声データの処理時から起算した経過時間幅に基づき重み付けをして最終認識結果を決定する。

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、階層化された発話コマンドを認識する音声認識技術に関する。

背景技術

[0002] 従来から階層化された発話コマンドを認識する音声認識装置において、発話者の明示的な戻り操作又は初期化操作を受け付けることにより誤認識や誤操作に対処する技術が周知である。例えば、特許文献１には、音声入力を受け付けた後、所定の時間幅までにさらに音声入力があった場合には、前に入力された音声の言い直しがあったとして音声認識する技術が開示されている。その他、本発明に関連する技術が、特許文献２に開示されている。

先行技術文献

特許文献

[0003] 特許文献１：特開平８－１９０３９８号公報

特許文献２：特開２００１－０６３４８９号公報

発明の概要

発明が解決しようとする課題

[0004] 発話者の明示的な戻り操作又は初期化操作を受け付けることにより誤認識や誤操作に対処する場合、当該対処により浪費する時間が大きい。また、このような明示的な操作が要求される場合、発話者は煩わしさを感じるようになる。

[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、誤認識等があった場合でも、訂正動作に相当する発話や操作、又は最初からやり直しをすることなく、階層コマンドを認識す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、階層コマンドを音声認識する音声認識装置であって、前記階層コマンドを認識するための各階層の音声認識で使用する辞書を記憶する辞書記憶手段と、音声データの入力があった場合、前回入力された音声データを正しく認識したと仮定した場合に使用すべき辞書に基づき音声認識を行う第 1 音声認識手段と、前記音声データの入力があった場合、前記前回入力された音声データの音声認識を実行した際に使用した辞書に基づき音声認識を行う第 2 音声認識手段と、前記第 1 音声認識手段による認識結果と、前記第 2 音声認識手段による認識結果と、を前記前回入力された音声データの処理時から起算した経過時間幅に基づき重み付けをして最終認識結果を決定する認識結果処理手段と、を備えることを特徴とする。
- [0007] 請求項 8 に記載の発明は、階層コマンドを認識するための各階層の音声認識で使用する辞書を記憶する音声認識装置が使用する音声認識方法であって、音声データの入力があった場合、前回入力された音声データを正しく認識したと仮定した場合に使用すべき辞書に基づき音声認識を行う第 1 音声認識工程と、前記音声データの入力があった場合、前記前回入力された音声データの音声認識を実行した際に使用した辞書に基づき音声認識を行う第 2 音声認識工程と、前記第 1 音声認識手段による認識結果と、前記第 2 音声認識手段による認識結果と、を前記前回入力された音声データの処理時から起算した経過時間幅に基づき重み付けをして最終認識結果を決定する認識結果処理工程と、を備えることを特徴とする。
- [0008] 請求項 9 に記載の発明は、階層コマンドを認識するための各階層の音声認識で使用する辞書を記憶する音声認識装置により実行される音声認識プログラムであって、音声データの入力があった場合、前回入力された音声データを正しく認識したと仮定した場合に使用すべき辞書に基づき音声認識を行う第 1 音声認識手段と、前記音声データの入力があった場合、前記前回入力された音声データの音声認識を実行した際に使用した辞書に基づき音声認識を行う第 2 音声認識手段と、前記第 1 音声認識手段による認識結果と、前記第

2 音声認識手段による認識結果と、を前記前回入力された音声データの処理時から起算した経過時間幅に基づき重み付けをして最終認識結果を決定する認識結果処理手段、として前記音声認識装置を機能させることを特徴とする。

図面の簡単な説明

[0009] [図1]非階層コマンド及び階層コマンドを音声認識する際に使用する辞書構成の一例を示す図である。

[図2]第1実施例に係る音声認識装置の概略構成の一例を示す図である。

[図3]第1実施例に係る認識結果処理部の機能ブロックの一例を示す図である。

[図4]重み付け関数 F_w の例を示す図である。

[図5]第1認識部の認識結果、第2認識部の認識結果、及び最終的な認識結果の一例を示す図である。

[図6]第1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7]第1実施例の変形例1の概略構成の一例を示すフローチャートの一例である。

[図8]第2実施例に係る音声認識装置の概略構成の一例を示す図である。

発明を実施するための形態

[0010] 本発明の1つの観点では、階層コマンドを音声認識する音声認識装置であって、前記階層コマンドを認識するための各階層の音声認識で使用する辞書を記憶する辞書記憶手段と、音声データの入力があった場合、前回入力された音声データを正しく認識したと仮定した場合に使用すべき辞書に基づき音声認識を行う第1音声認識手段と、前記音声データの入力があった場合、前記前回入力された音声データの音声認識を実行した際に使用した辞書に基づき音声認識を行う第2音声認識手段と、前記第1音声認識手段による認識結果と、前記第2音声認識手段による認識結果と、を前記前回入力された音声データの処理時から起算した経過時間幅に基づき重み付けをして最終認識結果を決定する認識結果処理手段と、を備える。

[0011] 上記の音声認識装置は、階層コマンドを音声認識し、辞書記憶手段と、第1音声認識手段と、第2音声認識手段と、認識結果処理手段と、を備える。

ここで、「階層コマンド」とは、制御対象となる装置への1の動作命令又は出力命令を2以上の発話で実現するコマンドを指す。即ち、階層コマンドの場合、1の動作命令又は出力命令は、音声認識されたコマンドに基づき徐々に階層的に特定される。辞書記憶手段は、階層コマンドを認識するための各階層の音声認識で使用する辞書を記憶する。ここで、「各階層の音声認識で使用する辞書」とは、任意の1の動作命令又は出力命令を実現するのに必要な各音声認識処理で使用する辞書を指す。なお、ここでの「音声認識処理」とは、1の音声データを認識するのに必要な処理全体を指す。また、辞書には、入力された音声データとパターンマッチングを行う複数の認識語が格納されている。第1音声認識手段は、音声データの入力があった場合、前回入力された音声データを正しく認識したと仮定した場合に使用すべき辞書に基づき音声認識を行う。即ち、第1音声認識手段は、一つ前に入力された音声データに基づき使用すべき辞書を決定し、新たに入力された音声データの音声認識を行う。第2音声認識手段は、音声データの入力があった場合、前回入力された音声データの音声認識を実行した際に使用した辞書に基づき音声認識を行う。即ち、第2音声認識手段は、1つ前に入力された音声データの音声認識が誤認識であったことを前提に新たに入力された音声データの音声認識を行う。認識結果処理手段は、第1音声認識手段の認識結果と、第2音声認識手段の認識結果と、を前回入力された音声データの処理時から起算した経過時間幅に基づき重み付けをして最終認識結果を決定する。ここで、「認識結果」とは、使用した辞書に格納された各認識語と新たに入力された音声データとの類似度を指す。「経過時間幅」とは、1つ前に入力された音声データの処理開始時若しくは終了時から新たに入力された音声データの処理開始時までの時間幅、又はこれに相当する時間幅を指す。また、「最終認識結果」とは、音声認識装置が最終的に出力する認識結果を指す。例えば最新認識結果は、類似度が高い順に並べられた所定個数分の認識語及びその類似

度である。

[0012] 一般に、再発話までの経過時間幅が短い場合、前に入力された音声の言い直しである可能性が高い。従って、音声認識装置は、経過時間幅に基づき、2つの認識処理の認識結果の重み付けを行う。これにより、音声認識装置は、誤認識があった場合でも、訂正動作に相当する発話や操作、又は最初からやり直しをすることなく、最小限のステップで再発話を認識することができる。

[0013] 上記の音声認識装置の一態様では、前記認識結果処理手段は、前記経過時間幅が所定幅以上の場合、前記第1音声認識手段の認識結果のみに基づき前記最終認識結果を決定する。上述の所定幅は、実験等に基づき適切な値に設定される。具体的には、この所定幅は、例えば新たに入力された音声データが前回入力された音声データの言い直しである可能性がない又は極めて低いと判断される経過時間幅に設定される。このように、この態様では、音声認識装置は、再発話である可能性がない又は低い場合に、不要な処理を削減することができる。

[0014] 上記の音声認識装置の他の一態様では、前記認識結果処理手段は、前記経過時間幅が所定幅未満の場合、前記第2音声認識手段の認識結果のみに基づき前記最終認識結果を決定する。上述の所定幅は、実験等に基づき適切な値に設定される。具体的には、この所定幅は、例えば新たに入力された音声データが前回入力された音声データの言い直しである可能性が極めて高いと判断される経過時間幅に設定される。このように、この態様では、音声認識装置は、再発話である可能性が高い場合に不要な処理を削減することができる。

[0015] 上記の音声認識装置の他の一態様では、前記第1音声認識手段及び前記第2音声認識手段は、認識結果として尤度を算出し、前記認識結果処理手段は、前記第1音声認識手段が算出した尤度に前記経過時間幅に対して非増加関数となる第1パラメータを乗じると共に、前記第2音声認識手段が算出した尤度に前記経過時間幅に対して非減少関数となる第2パラメータを乗じるこ

とで前記重み付けを実行する。ここで、「尤度」とは、対数尤度も含む。この態様では、音声認識装置は、第1音声認識手段が算出した尤度に経過時間幅に対して非増加関数となる第1パラメータを乗じることで、徐々に第1音声認識手段の認識結果の重み、即ち最終認識結果に対する重要度を上げる。また、音声認識装置は、第2音声認識手段が算出した尤度に経過時間幅に対して非減少関数となる第2パラメータを乗じることで、徐々に第2音声認識手段の認識結果の重みを下げる。このようにすることで、音声認識装置は、誤認識があった場合でも、訂正動作に相当する発話や操作、又は最初からやり直しをすることなく、最小限のステップで再発話を認識することができる。

[0016] 上記の音声認識装置の他の一態様では、前記認識結果処理手段は、前記第1パラメータと前記第2パラメータとの和を一定値にする。上述の一定値は、例えば1に設定される。このようにすることで、音声認識装置は、第1パラメータ又は第2パラメータの一方に基づき他方のパラメータを容易に決定し、かつ、これらのパラメータを適切な値に設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、認識環境に関する情報を取得する環境情報取得手段をさらに備え、前記認識結果処理手段は、前記情報に基づき認識環境が劣悪であると判断した場合、前記認識環境が劣悪でない場合に比べ、前記第2音声認識手段の認識結果の重みを大きくする。「認識環境」とは、音声認識装置が音声認識を行う環境であって、当該音声認識の精度に影響を与えるものを指す。一般に、認識環境が劣悪の場合、誤認識が発生する可能性が高い。即ち、この場合、発話者は言い直しをする可能性が高い。従って、音声認識装置は、認識環境が劣悪であると判断した場合には、第2音声認識手段の認識結果の重みを通常より大きくすることで、より精度高く音声認識を実行することができる。

[0018] 上記の音声認識装置の他の一態様では、車両に搭載され、前記環境情報取得手段は、前記車両の車速又は／及び音声データに含まれる雑音の度合いを取得し、前記認識結果処理手段は、前記車速又は／及び前記雑音の度合いが

大きいほど、前記第2音声認識手段の認識結果の重み付けを大きくする。ここで、「雑音の度合い」とは、例えばS/N比が該当する。一般に、車速が大きい場合、運転者は運転に集中するため、音声認識装置に対する反応が遅くなり経過時間も大きくなる傾向がある。また、車速が大きい場合、走行に伴う雑音も大きくなることが予想される。従って、この態様では、音声認識装置は、車速又は／及び雑音の度合いが大きいほど、誤認識の発生確率が高く、再発話が行われる可能性が高いと判断する。従って、この態様では、音声認識装置は、より精度高く音声認識を実行することができる。

[0019] 本発明の他の観点では、階層コマンドを認識するための各階層の音声認識で使用する辞書を記憶する音声認識装置が使用する音声認識方法であって、音声データの入力があった場合、前回入力された音声データを正しく認識したと仮定した場合に使用すべき辞書に基づき音声認識を行う第1音声認識工程と、前記音声データの入力があった場合、前記前回入力された音声データの音声認識を実行した際に使用した辞書に基づき音声認識を行う第2音声認識工程と、前記第1音声認識手段による認識結果と、前記第2音声認識手段による認識結果と、を前記前回入力された音声データの処理時から起算した経過時間幅に基づき重み付けをして最終認識結果を決定する認識結果処理工程と、を備えることを特徴とする。音声認識装置は、この方法を使用することにより、誤認識があった場合でも、訂正動作に相当する発話や操作、又は最初からやり直しをすることなく、最小限のステップで再発話を認識することができる。

[0020] 本発明のさらに別の観点では、階層コマンドを認識するための各階層の音声認識で使用する辞書を記憶する音声認識装置により実行される音声認識プログラムであって、音声データの入力があった場合、前回入力された音声データを正しく認識したと仮定した場合に使用すべき辞書に基づき音声認識を行う第1音声認識手段と、前記音声データの入力があった場合、前記前回入力された音声データの音声認識を実行した際に使用した辞書に基づき音声認識を行う第2音声認識手段と、前記第1音声認識手段による認識結果と、前

記第2音声認識手段による認識結果と、を前記前回入力された音声データの処理時から起算した経過時間幅に基づき重み付けをして最終認識結果を決定する認識結果処理手段、として前記音声認識装置を機能させる。音声認識装置は、このプログラムを搭載することで、誤認識があった場合でも、訂正動作に相当する発話や操作、又は最初からやり直しをすることなく、最小限のステップで再発話を認識することができる。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

[0021] 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。以下では、まず〔基本説明〕のセクションで本発明の音声認識手法に関連する辞書の構成について基本説明を行った後、〔第1実施例〕及び〔第2実施例〕の各セクションで本発明を適用した実施例について説明する。

[0022] 〔基本説明〕

具体的な実施例の説明に先立って、本発明の音声認識手法に関連する辞書の構成について基本説明を行う。図1は、音声認識に用いられる2種類の辞書構成を示す概念図である。具体的には、図1(a)は、比較例となる非階層コマンドを認識する場合の辞書構成の一例を示し、図1(b)は、本発明の対象となる階層コマンドを認識する場合の辞書構成の一例を示す。本発明の音声認識装置は、後述するように、階層コマンドを認識する。なお、「非階層コマンド」とは、制御対象となる装置への1の動作命令又は出力命令を1の発話で実現する発話コマンドを指し、「階層コマンド」とは、発話コマンドのうち、制御対象となる装置への1の動作命令又は出力命令を2以上の発話で実現する発話コマンドを指す。また、「辞書」とは、認識対象の単語（以後、「認識語」と呼ぶ。）を格納したデータベースを指す。

[0023] まず、非階層コマンドを認識する場合について図1(a)を用いて説明する。音声認識装置は、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S_a」と呼ぶ。）に対し所定の解析を実行した後、コマンド用辞書11に基づき認識結果を出力する。ここで、「認識結果」とは、例えば、発

話データ S a に対し最も類似度が高い認識語、又は、発話データ S a に対し類似度が高い順に並べられた認識語及びその類似度のリスト等が該当する。また、「発話データ S a」とは、音声を含む入力信号を指す。例えば、カーナビゲーション装置に音声認識装置が実装された場合では、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。

[0024] 次に、階層コマンドを認識する場合について図 1 (b) を用いて説明する。図 1 (b) の例では、階層コマンドは、2つの階層から構成される。まず、音声認識装置は、一連の階層コマンドの認識処理のうち最初に入力される発話データ S a を、第 1 コマンド用辞書 12 を用いて音声認識する。ここでの音声認識を、以後「第 1 階層での音声認識」と呼ぶ。即ち、第 1 階層での音声認識とは、一連の階層コマンドを認識する処理のうち、初期状態、即ち発話コマンドを受理していない状態で用いる辞書により発話データ S a を音声認識する処理を指す。このように、階層コマンドを認識する場合、音声認識装置は、その処理状態（階層）ごとに使用する辞書（以後、「待ち受け辞書」とも呼ぶ。）を変える。そして、音声認識装置は、第 1 コマンド用辞書 12 に基づき、第 1 階層での音声認識で「名称検索」を認識したものとする。即ち、音声認識装置は、第 1 コマンド用辞書 12 に格納された認識語のうち、「名称検索」が最も類似度が高いと判断したものとする。

[0025] 次に、音声認識装置は、認識した「名称検索」のコマンドに基づき、第 2 コマンド用辞書 13 及び地名用辞書 14 を次の待ち受け辞書として音声認識を行う。ここで、第 2 コマンド用辞書 13 は、「戻る」、「中止」、「訂正」等の発話者が明示的に前の状態（即ち、第 1 階層での音声認識）に処理を戻したい場合又は階層コマンドの一連の認識処理を最初からやり直したい場合に入力すべき発話コマンドを含む。また、地名用辞書 14 は、第 1 階層での音声認識で認識された「名称探索」の対象となる地名を含む。そして、音声認識装置は、次に入力された発話データ S a（ここでは、「東京タワー」が発話されたとする。）に基づき、第 2 コマンド用辞書 13 及び地名用辞書

14に格納された認識語の類似度を計算する。以後、ここでの音声認識を、以後「第2階層での音声認識」と呼ぶ。即ち、第2階層での音声認識とは、第1階層での音声認識結果に基づき発話データS_aを認識する処理を指す。なお、第3階層での音声認識についても、同様に定義される。

[0026] そして、音声認識装置は、第2階層での音声認識の結果を最終的な認識結果として出力する。その後、音声認識装置は、新たな発話データS_aに対し再び第1コマンド用辞書12を用いて第1階層での音声認識を行う。

[0027] なお、上述の階層コマンドの例では、2階層の発話コマンドの場合について説明したが、3階層以上の階層コマンドであっても同様に、音声認識装置は、各階層での音声認識で、前の階層で音声認識された認識語に基づき待ち受け辞書を選択し音声認識を行う。

[0028] 本発明では、音声認識装置は、上述したように、階層コマンドの音声認識を実行する。さらに、本発明に係る音声認識装置は、以後説明するように、前の階層での音声認識からの経過時間に応じて認識結果の出力を変える。これにより、本発明に係る音声認識装置は、一連の階層コマンドの認識処理の途中で誤認識が生じた場合であっても、利用者が明示的にコマンド（上述の例では、第2コマンド用辞書13に格納されたコマンドに相当する。）を入力することなく、最小限のステップで再発話を音声認識することができる。

[0029] [第1実施例]

まず、本発明に係る第1実施例について説明する。以後では、音声認識装置の概略構成について説明した後、音声認識方法、及びその処理フローについて説明する。その後、第1実施例の各変形例について説明する。

[0030] (概略構成)

まず、第1実施例に係る音声認識装置の概略構成について図2を用いて説明する。

[0031] 図2は、本発明に係る音声認識装置100の概略構成図である。音声認識装置100は、音声分析部1と、音声区間検出部2と、第1認識部3と、第2認識部4と、認識結果処理部5と、提示部6と、主制御部7と、時間計測

部 8 と、記憶部 9 と、を備える。なお、破線は、制御信号の流れを示し、実線は、データの流れを示す。

[0032] 音声分析部 1 は、主制御部 7 の制御に基づき、発話データ S a の音響特徴量（即ち、音声の特徴パラメータ）を算出する。具体的には、音声分析部 1 は、図示しないマイク等から入力された発話データ S a を A/D 変換し、例えば周知の音響分析手法又はこの組み合わせを用いることで、音響特徴量を算出する。

[0033] 音声区間検出部 2 は、主制御部 7 の制御に基づき、音声分析部 1 から供給される音声特徴量を用いて、発話データ S a から発話区間（以後、「音声データ」と呼ぶ。）のみを切り出す。そして、音声区間検出部 2 は、第 1 認識部 3 及び第 2 認識部 4 に音声データに該当する音声特徴量を供給する。

[0034] 第 1 認識部 3 は、音声特徴量に基づき、HMM（Hidden Markov Model）などの識別器を用いて所定の待ち受け辞書に格納される認識語の尤度を算出する。そして、第 1 認識部 3 は、類似度が高い順に並べられた認識語及びその尤度のリストを出力する。

[0035] 具体的には、第 1 認識部 3 は、前回入力された音声データを正しく認識したと仮定した場合に使用すべき辞書に基づき音声認識を行う。即ち、第 1 認識部 3 は、前回入力された音声データを処理した階層の次の階層（当該次の階層がない場合には、第 1 階層）で音声認識を行う。具体的には、図 1（b）の例では、第 1 認識部 3 は、前に入力された音声データを第 1 階層で音声認識した場合には、第 2 コマンド用辞書 13 及び地名用辞書 14 を用いて認識結果を算出し、前に入力された音声データを第 2 階層で音声認識した場合には、第 1 コマンド用辞書 12 を用いて認識結果を算出する。そして、第 1 認識部 3 は、認識結果である認識語（以後、「認識語 WR 1」と呼ぶ。）及びこれに対応する対数化された尤度（以後、「尤度 L 1」と呼ぶ。）を認識結果処理部 5 に供給する。ここでは、第 1 認識部 3 は、上位となった所定個数（以後、「所定個数 N」と呼ぶ。）の認識語 WR 1 及びその尤度 L 1 を供給するものとする。上述の所定個数 N は、実験等により適切な値に設定され

る。

[0036] 第2認識部4は、音声特徴量に基づき、HMMなどの識別器を用いて、第1認識部3とは異なる所定の待ち受け辞書に格納される認識語の尤度を算出する。そして、第2認識部4は、類似度が高い順に並べられた認識語及びその尤度のリストを出力する。

[0037] 具体的には、第2認識部4は、前回入力された音声データの音声認識を実行した際に使用した待ち受け辞書に基づき音声認識を行う。即ち、第2認識部4は、前に入力された発話データS_aを処理した階層と同一階層で音声認識を行う。言い換えると、第2認識部4は、第1認識部3の前の階層の音声認識を行う。より具体的には、図1(b)の例では、第2認識部4は、前に入力された発話データS_aを第1階層で音声認識した場合には、再び第1コマンド用辞書12を用いて第1階層で音声認識を行い、前に入力された発話データS_aを第2階層で音声認識した場合には、再び第2コマンド用辞書13及び地名用辞書14を用いて第2階層で音声認識を行う。そして、第2認識部4は、認識結果である認識語（以後、「認識語WR2」と呼ぶ。）及びこれに対応する対数尤度（以後、「尤度L2」と呼ぶ）を認識結果処理部5に供給する。ここでは、第2認識部4は、上位となった所定個数Nの認識語WR2及びその尤度L2を供給するものとする。

[0038] 認識結果処理部5は、第1認識部3及び第2認識部4で出力された尤度L1、L2を、前回の認識結果の出力時から起算した経過時間幅（以後、「経過時間幅T」と呼ぶ。）に基づき所定の重み付けを行う。これにより、認識結果処理部5は、重み付けされた尤度L1（以後、「重み付け尤度L_w1」と呼ぶ。）及び重み付けされた尤度L2（以後、「重み付け尤度L_w2」と呼ぶ。）を算出する。そして、認識結果処理部5は、重み付け尤度L_w1、L_w2を一体にしてソートすることで、類似度が高い順に認識語WR1、WR2（以後、「認識語WR」と総称する。）及びこれに対応する重み付け尤度L_w1、L_w2（以後、「重み付け尤度L_w」と総称する。）を提示部6へ供給する。なお、認識結果処理部5での具体的な処理については、次の（

認識処理方法)のセクションで詳しく説明する。

[0039] 提示部6は、認識結果処理部5が出力した認識結果を最終的な認識結果として提示する。具体的には、提示部6は、ディスプレイ又はスピーカ等であり、認識結果処理部5が出力した認識結果を画像又は音声により出力する。

[0040] 主制御部7は、それぞれ図示しないCPU (Central Processing Unit)、ROM (Read Only Memory) 及びRAM (Random Access Memory)などを備え、音声認識装置100の各要素に対して種々の制御を行う。

[0041] 時間計測部8は、経過時間幅Tを計測するためのタイマである。具体的には、時間計測部8は、1つ前の発話データS_aの認識結果を提示部6が出力した時刻から起算した経過時間幅Tを計測する。

[0042] 記憶部9は、第1認識部3及び第2認識部4等が使用する音響モデル、及び各階層での音声認識ごとに使用される待ち受け辞書、その他認識結果処理部5が出力した認識結果等を記憶するメモリである。記憶部9に記憶されたデータは、主制御部7の制御に基づき、音声認識装置100の各要素へ必要に応じて供給される。

[0043] (認識処理方法)

次に、図3乃至図5を参照して、認識結果処理部5が実行する処理について具体的に説明する。図3は、認識結果処理部5の処理内容を具体的に示した機能ブロック図である。図3に示すように、認識結果処理部5は、乗算部5_x、5_yと、ソート部5_zと、スイッチ部5_vと、を備える。

[0044] 乗算部5_xは、第1認識部3から供給される尤度L₁に所定の重み付け係数(以後、「第1重み付け係数w₁」と呼ぶ。)を乗じて、その乗算結果である重み付け尤度L_{w1}をソート部5_zに供給する。即ち、乗算部5_xは、重み付け尤度L_{w1}を以下の式(1)により求める。

$$L_{w1} = L_1 \times w_1 \quad \text{式(1)}$$

ここで、第1重み付け係数w₁は、経過時間幅Tに応じて変動する係数である。具体的には、第1重み付け係数w₁は、経過時間幅Tが大きくなるほ

ど、第1認識部3の認識結果の重要度（重み）を上げるように設定される。即ち、第1重み付け係数 w_1 は、経過時間幅 T を変数かつ尤度 L_1 を定数とした場合に、経過時間幅 T の増加と共に重み付け尤度 L_{w_1} の絶対値を小さくするように減少（即ち、非増加）する係数となる。なお、音声認識装置100は、尤度 L_1 、 L_2 及び重み付け尤度 L_{w_1} 、 L_{w_2} を、それぞれ負値が小さいほど、即ち絶対値が小さいほど、類似度が高いとみなす。従って、乗算部5xは、第1重み付け係数 w_1 が小さいほど、第1認識部3の認識結果の重みを上げるように重み付け尤度 L_{w_1} を設定することになる。第1重み付け係数 w_1 の設定方法については、後述する。

[0045] 同様に、乗算部5yは、第2認識部4から供給される尤度 L_2 に所定の重み付け係数（以後、「第2重み付け係数 w_2 」と呼ぶ。）を乗じて、その乗算結果である重み付け尤度 L_{w_2} をソート部5zに供給する。即ち、乗算部5yは、第2重み付け尤度 L_{w_2} を以下の式（2）により求める。

$$L_{w_2} = L_2 \times w_2 \quad \text{式（2）}$$

ここで、第2重み付け係数 w_2 は、経過時間幅 T に応じて変動する係数である。具体的には、第2重み付け係数 w_2 は、経過時間幅 T が大きくなるほど、第2認識部4の認識結果の重要度（重み）を下げるように設定される。即ち、第2重み付け係数 w_2 は、経過時間幅 T を変数かつ尤度 L_2 を定数とした場合に、経過時間幅 T の増加と共に重み付け尤度 L_{w_2} の負値を大きくするように増加（即ち、非減少）する係数となる。従って、乗算部5yは、第2重み付け係数 w_2 が小さいほど、第2認識部4の認識結果の重みを上げるように重み付け尤度 L_{w_2} を設定することになる。第2重み付け係数 w_2 の算出方法については、後述する。

[0046] また、本実施例では、第1重み付け係数 w_1 と第2重み付け係数 w_2 とは、それぞれ0より大きく1より小さい値に設定され、かつ和が1になるように設定される。即ち、第1重み付け係数 w_1 と第2重み付け係数 w_2 とは、以下の式（3）乃至（5）を満たす。

$$w_1 + w_2 = 1 \quad \text{式（3）}$$

$$0 < w_1 < 1 \quad \text{式 (4)}$$

$$0 < w_2 < 1 \quad \text{式 (5)}$$

これにより、認識結果処理部 5 は、第 1 重み付け係数 w_1 と第 2 重み付け係数 w_2 とのいずれか一方を求めることで、他方の係数を適切に設定することができる。即ち、認識結果処理部 5 は、経過時間幅 T が大きくなるほど、第 2 認識部 4 の認識結果を第 1 認識部 3 の認識結果に比べ、相対的にその重要度を下げるように設定することができる。

[0047] なお、以後では、便宜的に、第 1 及び第 2 重み付け係数 w_1 、 w_2 を所定の重み付け係数「 w 」を用いて以下の式 (6)、式 (7) により表す。

$$w_1 = w \quad \text{式 (6)}$$

$$w_2 = 1 - w \quad \text{式 (7)}$$

ここで、重み付け係数 w は、式 (4) 及び式 (6) より、0 より大きく 1 より小さい値に設定される。即ち、以下の式 (8) を満たす。

$$0 < w < 1 \quad \text{式 (8)}$$

次に、第 1 重み付け係数 w_1 及び第 2 重み付け係数 w_2 を決定する重み付け係数 w の設定方法について図 4 を用いて詳しく説明する。図 4 は、重み付け係数 w と経過時間幅 T との関係を示すグラフの一例である。グラフ「 G_1 」及びグラフ「 G_2 」は、経過時間幅 T から重み付け係数 w を決定する関数（以後、「重み付け関数 F_w 」と呼ぶ。）の一例である。また、図 4 中の「 T_1 」は、第 1 認識部 3 の認識結果を最終的な認識結果とするか、又は第 1 認識部 3 の認識結果に加え第 2 認識部 4 の認識結果を用いて最終的な認識結果を導出するかをスイッチ部 5 v が判断するための経過時間幅 T に対する閾値である。これについては、スイッチ部 5 v の説明で詳しく説明する。

[0048] 図 4 に示すように、重み付け関数 F_w は、グラフ G_1 、 G_2 のいずれで表される場合も、経過時間幅 T を変数とした減少関数となる。具体的には、グラフ G_1 は、経過時間幅 T が時間幅「 $T_1 \alpha$ 」までは同一の傾きを保って減少し、その後、下に凸となる減少関数である。一方、グラフ G_2 は、閾値 T_1 まで下に凸の減少関数となる。他の例では、重み付け係数 w を決定する重

み付け関数 F_w は、経過時間幅 T によらず一定値でもよい。具体的には、重み付け関数 F_w に相当するマップ又は式は、ユーザの使用環境や認識語数、当該音声認識装置 100 の性能等を考慮し、実験等に基づき予め生成され、記憶部 9 に記憶される。

[0049] 次に、再び図 3 に戻り、ソート部 5 z が実行する処理について説明する。ソート部 5 z は、乗算部 5 x から供給された重み付け尤度 L_{w1} 及び乗算部 5 y から供給された重み付け尤度 L_{w2} に基づき、類似度が高い順に、重み付け尤度 L_{w1} 、 L_{w2} に対応する認識語をソートする。そして、ソート部 5 z は、上位の所定個数 N の認識語 WR 及びその重み付け尤度 L_w を提示部 6 へ供給する。

[0050] これについて図 5 の具体例を用いて説明する。図 5 は、図 1 (b) の例の第 2 階層の音声認識で誤認識が発生し、発話者が誤認識された単語の言い直し（以後、単に「再発話」と呼ぶ。）を行った場合の第 1 認識部 3 の認識結果、第 2 認識部 4 の認識結果、及びソート部 5 z の出力結果のリストの具体例を示す。前提として、図 5 では、発話者が「名称検索」を発話した後「東京タワー」と発話したのに対し、音声認識装置 100 は、第 1 階層の音声認識で「名称検索」と正しく認識した後、第 2 階層の音声認識で「東京港」と誤認識したとする。また、発話者は、その後、「東京タワー」と再発話を行ったものとする。

[0051] この場合、第 1 認識部 3 は、再発話の音声データに対し、第 1 コマンド用辞書 12 を用いて第 1 階層での音声認識を行うとともに、第 2 認識部 4 は、再発話の音声データに対し、第 2 コマンド用辞書 13 及び地名用辞書 14 を用いて第 2 階層での音声認識を行う。また、ここでは、重み付け係数 w は、経過時間幅 T に基づき「0.7」、即ち第 1 認識部 3 の認識結果より第 2 認識部 4 の認識結果に重みを置く値に設定されたものとする。

[0052] 図 5 (a) は、再発話に対する第 1 認識部 3 の認識結果を示すリストの一例である。具体的には、第 1 認識部 3 は、再発話の音声データの音響特徴量に基づき、第 1 コマンド用辞書 12 に格納された単語の尤度 L_1 を算出する

。そして、第1認識部3は、当該尤度 L_1 に基づき類似度が高い順に認識語 WR_1 及びその尤度 L_1 を所定個数 N だけ出力する。また、図5(a)に示すように、乗算部5xは、第1重み付け係数 w_1 ($=0.7$)に基づき、各尤度 L_1 に対する重み付け尤度 L_{w_1} を算出する。

[0053] 図5(b)は、再発話に対する第2認識部4の認識結果を示すリストの一例である。具体的には、第2認識部4は、再発話の音声データの音響特徴量に基づき、第2コマンド用辞書13及び地名用辞書14に格納された認識語の尤度 L_2 を算出する。そして、第2認識部4は、当該尤度 L_2 に基づき図5(b)のリストを類似度が高い順に認識語 WR_2 及びその尤度 L_2 を所定個数 N だけ出力する。また、図5(b)に示すように、乗算部5yは、第2重み付け係数 w_2 ($=0.3$)に基づき、各尤度 L_2 に対する重み付け尤度 L_{w_2} を算出する。

[0054] 図5(c)は、再発話に対するソート部5zの出力を示すリストの一例である。図5(c)に示すように、ソート部5zは、図5(a)、(b)に示す重み付け尤度 L_{w_1} 、 L_{w_2} に基づき、認識語 WR_1 、 WR_2 を並び替え、所定個数 N ($=4$)だけ出力している。このとき、第2重み付け係数 w_2 が第1重み付け係数 w_1 より小さいこと等に起因して、第2認識部4で認識された「東京タワー」が1位にランクし、同様に第2認識部4で認識された「東京港」が2位にランクしている。一方、第1認識部3で1位及び2位として認識された「ルート探索」及び「ルート消去」は、第2認識部4で認識された単語よりも下位にランクしている。

[0055] このように、ソート部5zは、経過時間幅 T に基づき適切に設定した重み付け係数 w を用いて重み付け尤度 L_{w_1} 、 L_{w_2} を算出し、これらを一体にソートする。これにより、音声認識装置100は、発話者が明示的に再発話を行う旨の操作をしたり、発話者が再び第1階層から音声入力をやり直すことなく、最小限の手続きで、誤認識が生じた場合の再発話を正しく認識することができる。

[0056] また、ソート部5zは、上述の処理に加え、1位の認識語 WR が前回入力

された発話データ S_a に対する認識結果が 1 位の認識語と一致する場合、2 位にランクされた認識語 WR を 1 位に繰り上げる。このようにすることで、ソート部 5 z は、再発話に対し再び同一の認識語であると誤認識して前回と同じ認識語を 1 位として出力するのを防ぐことができる。

- [0057] 次に、スイッチ部 5 v が実行する処理について再び図 3 に戻り説明する。スイッチ部 5 v は、経過時間幅 T に基づき、第 1 認識部 3 の出力を最終的な認識結果として提示部 6 に供給するか、または、ソート部 5 z の出力を最終的な認識結果として提示部 6 に供給するか切り替える。具体的には、スイッチ部 5 v は、経過時間幅 T が閾値 T_1 未満の場合、再発話の可能性があると判断し、第 2 認識部 4 の認識結果が考慮されたソート部 5 z の出力を提示部 6 へ供給する。一方、スイッチ部 5 v は、経過時間幅 T が閾値 T_1 以上の場合、前回の発話から相当の時間が経過していることより再発話の可能性はない又は低いと判断し、第 1 認識部 3 の認識結果を提示部 6 に供給する。このように、スイッチ部 5 v は、経過時間幅 T が閾値 T_1 未満の場合、第 1 認識部 3 の認識結果のみに基づき最終的な認識結果を出力することで、不要な処理を削減することができる。

- [0058] (処理フロー)

次に、第 1 実施例における処理の手順について説明する。図 6 は、第 1 実施例において音声認識装置 100 が実行する処理手順を表すフローチャートの一例である。音声認識装置 100 は、図 6 に示すフローチャートの処理を繰り返し実行する。

- [0059] まず、音声認識装置 100 は、発話データ S_a の入力があるか否か判定する（ステップ $S101$ ）。そして、発話データ S_a の入力がある場合（ステップ $S101$; Yes）、音声認識装置 100 は、ステップ $S102$ へ処理を進める。一方、発話データ S_a の入力がない場合（ステップ $S101$; No）、音声認識装置 100 は、引き続き発話データ S_a の入力があるか否か監視する。

- [0060] 次に、音声認識装置 100 は、経過時間幅 T を算出する（ステップ $S10$

2)。即ち、音声認識装置100は、例えば提示部6により認識結果が提示された後次の発話データS_aの入力があるまでの時間幅を経過時間幅Tに定める。

[0061] そして、音声認識装置100は、前回発話があるか否か判定する（ステップS103）。即ち、音声認識装置100は、階層コマンドの音声入力を初めて受け付けたか否か判定する。そして、前回発話がある場合（ステップS103；Yes）、音声認識装置100はステップS104の処理を行う。一方、前回発話がない場合（ステップS103；No）、音声認識装置100は、第1認識部3により音声認識を行う（ステップS111）。

[0062] 次に、音声認識装置100は、経過時間幅Tが閾値T₁より小さいか否か判定する（ステップS104）。そして、経過時間幅Tが閾値T₁より小さい場合（ステップS104；Yes）、音声認識装置100は、ステップS105乃至ステップS110の処理を行う。即ち、この場合、音声認識装置100は、誤認識に起因した再発話の可能性があると判断し、第2認識部4の認識結果も考慮して最終的な認識結果を出力する。

[0063] 一方、音声認識装置100は、経過時間幅Tが閾値T₁以上の場合（ステップS104；No）、第1認識部3により音声認識を行う（ステップS111）。即ち、この場合、音声認識装置100は、誤認識に起因した再発話の可能性がない又は極めて低いと判断する。そして、第1認識部3は、尤度L₁を算出し、類似度が高い順に所定個数Nの認識語WR₁及びその尤度L₁のリストを出力する。

[0064] 次に、ステップS105では、音声認識装置100は、第1認識部3及び第2認識部4により音声認識を行う（ステップS105）。具体的には、第1認識部3は、尤度L₁を算出し、類似度が高い順に所定個数Nの認識語WR₁及びその尤度L₁のリストを出力する。また、第2認識部4は、尤度L₂を算出し、類似度が高い順に所定個数Nの認識語WR₂及びその尤度L₂のリストを出力する。このとき、第2認識部4は、第1認識部3の前の階層の音声認識を行う。

- [0065] 次に、音声認識装置 100 は、経過時間幅 T に基づき第 1 及び第 2 重み付け係数 w_1 、 w_2 を決定する（ステップ S 106）。例えば、音声認識装置 100 は、記憶部 9 に記憶したグラフ G_1 又はグラフ G_2 等を参照し、経過時間幅 T に基づき第 1 及び第 2 重み付け係数 w_1 、 w_2 を決定する。
- [0066] そして、音声認識装置 100 は、重み付け尤度 Lw_1 、 Lw_2 を算出する（ステップ S 107）。即ち、音声認識装置 100 は、尤度 L_1 に第 1 重み付け係数 w_1 を乗じることで重み付け尤度 Lw_1 を算出するとともに、尤度 L_2 に第 2 重み付け係数 w_2 を乗じることで重み付け尤度 Lw_2 を算出する。
- [0067] 次に、音声認識装置 100 は、重み付け尤度 Lw_1 、 Lw_2 に基づきソートを行う（ステップ S 108）。ここで、重み付け尤度 Lw_1 、 Lw_2 は第 1 又は第 2 重み付け係数 w_1 、 w_2 により経過時間幅 T に応じて重み付けられている。これにより、音声認識装置 100 は、誤認識に起因した再発話の場合であっても、経過時間幅 T に基づき適切に音声認識することができる。
- [0068] 次に、音声認識装置 100 は、ソート後の 1 位となる認識語が前回と同様か否か判定する（ステップ S 109）。ここで、音声認識装置 100 は、前回の 1 位となった認識語等の情報を、記憶部 9 に保持しているものとする。そして、ソート後の 1 位が前回と同様の場合（ステップ S 109；Yes）、音声認識装置 100 は 2 位を 1 位に繰り上げる（ステップ 110）。これにより、音声認識装置 100 は、前回と同様の誤認識を行うのを防ぐことができる。一方、ソート後の 1 位が前回と異なる場合（ステップ S 109；No）、音声認識装置 100 は、ステップ S 112 へ処理を進める。
- [0069] 次に、音声認識装置 100 は、認識結果を提示する（ステップ S 112）。例えば、音声認識装置 100 は、認識結果を図 5（c）に示すようにリストにして表示し、または、最も類似度が高いと判断した認識語 WR を音声出力する。
- [0070] そして、音声認識装置 100 は、認識結果及び属性を記憶する（ステップ S 113）。ここで、「属性」とは、例えば重み付け尤度 Lw などの算出値

又は／及び第 1 認識部 3 と第 2 認識部 4 とのいずれの認識結果を 1 位として出力したか等の認識処理に付随した情報を指す。

[0071] 次に、音声認識装置 100 は、第 1 認識部 3 の設定を第 2 認識部 4 の設定にコピーする（ステップ S 114）。ここで、設定とは、例えば、いずれの辞書を待ち受け辞書として使用したか等の情報が該当する。そして、音声認識装置 100 は、経過時間幅 T の計測を開始する（ステップ S 115）。

[0072] 以上説明したように、本実施例による音声認識装置は、階層コマンドを音声認識し、辞書記憶手段と、第 1 音声認識手段と、第 2 音声認識手段と、認識結果処理手段と、を備える。辞書記憶手段は、階層コマンドを認識するための各階層の音声認識で使用する辞書を記憶する。第 1 音声認識手段は、音声データの入力があった場合、前回入力された音声データを正しく認識したと仮定した場合に使用すべき辞書に基づき音声認識を行う。即ち、第 1 音声認識手段は、前回入力された音声データに基づき使用すべき辞書を決定し、次の音声データの入力の音声認識を行う。第 2 音声認識手段は、音声データの入力があった場合、前回入力された音声データの音声認識を実行した際に使用した辞書に基づき音声認識を行う。即ち、第 2 音声認識手段は、前回入力された音声データの音声認識が誤認識であったことを前提に次に入力される音声データの音声認識を行う。認識結果処理手段は、第 1 音声認識手段の認識結果と、第 2 音声認識手段の認識結果と、を前回入力された音声データの処理時から起算した経過時間幅に基づき重み付けをして最終認識結果を決定する。一般に、再発話までの経過時間幅が短い場合、前に入力された音声の言い直しである可能性が高い。従って、音声認識装置は、経過時間幅に基づき、2 つの認識処理の認識結果の重み付けを行う。これにより、音声認識装置は、誤認識があった場合でも、訂正動作に相当する発話や操作、又は最初からやり直しをすることなく、階層コマンドを認識することができる。

[0073] （変形例 1）

図 2 の音声認識装置 100 の構成では、音声認識装置 100 は、第 1 認識部 3 と第 2 認識部 4 との 2 つの音声認識部を備え、第 2 認識部 4 が第 1 認識

部 3 の前の階層の音声認識を実行していた。しかし、本発明が適用可能な音声認識装置 100 の構成は、これに限定されない。これに代えて、音声認識装置 100 は、1 体の音声認識部のみで上述の処理を実行してもよい。

[0074] この具体例について、図 7 を用いて説明する。図 7 は、変形例 1 に係る音声認識装置 100 a の概略構成を示す図の一例である。音声認識装置 100 a は、図 2 の第 1 認識部 3 及び第 2 認識部 4 に代えて認識部 20 を備える点、及び、認識結果保持部 21 を備える点で図 2 に示す音声認識装置 100 の構成と異なる。

[0075] 認識部 20 は、第 1 認識部 3 及び第 2 認識部 4 の処理を逐次的に両方実行する。具体的には、認識部 20 は、まず、第 1 認識部 3 が行う処理を実行することで、尤度 L_1 を算出する。そして、認識部 20 は、尤度 L_1 及びこれに対応する認識語 WR_1 を認識結果保持部 21 へ供給する。次に、認識部 20 は、第 2 認識部 4 が行う処理を実行することで、尤度 L_2 を算出する。そして、認識部 20 は、尤度 L_2 及びこれに対応する認識語 WR_1 を認識結果処理部 5 へ供給する。

[0076] 認識結果保持部 21 は、認識部 20 から尤度 L_1 及びこれに対応する認識語 WR を受信した後、認識部 20 が尤度 L_2 を算出するまでこれらの認識結果を保持する。そして、認識結果保持部 21 は、認識部 20 が尤度 L_2 及びこれに対応する認識語 WR_2 を認識結果処理部 5 へ供給すると同時又はその前後に、保持した尤度 L_1 及びこれに対応する認識語 WR_1 を認識結果処理部 5 へ供給する。その後、認識結果処理部 5 は、図 3 で説明した処理を実行する。

[0077] 以上のように、音声認識装置 100 a は、音声認識部である認識部 20 を 1 つのみ備える場合であっても、第 1 実施例と同様の処理を実行することができる。また、これにより、音声認識装置 100 a の小型化を実現することができる。

[0078] (変形例 2)

図 2 の説明では、時間計測部 8 は、提示部 6 により認識結果が出力された

時から次の発話データ S_a の入力がある時までの時間幅を経過時間幅 T と設定していた。しかし、本発明が適用可能な設定は、これに限定されない。例えば、これに代えて、時間計測部 8 は、前回の発話データ S_a が入力された時から次の発話データ S_a が入力された時までの時間幅、又は、前回の発話データ S_a を入力するための発話ボタンが押下された時から次の発話ボタンが押下される時までの時間幅又はこれに相当する時間幅等を経過時間幅 T に設定してもよい。

[0079] なお、時間計測部 8 は、音声認識装置 100 が車両に搭載されているときに当該車両のエンジンがオフになった場合等、経過時間幅 T を適宜リセットしてもよい。

[0080] （変形例 3）

図 4 のグラフ G_1 及びグラフ G_2 では、音声認識装置 100 は、重み付け係数 w が「0」付近になる時間幅に閾値 T_1 を設定していた。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 100 は、例えば、重み付け係数 w が「0、5」付近になる時間幅に閾値 T_1 を設定してもよい。その他、閾値 T_1 は、用いられる重み付け関数 F_w ごとに、又は、ユーザからの入力等に基づき設定されてもよい。これによっても、音声認識装置 100 は、不要な処理を削減しつつ、誤認識に起因した再発話を最小限のステップで認識することができる。

[0081] （変形例 4）

図 2 等の説明では、音声認識装置 100 は、第 1 認識部 3 及び第 2 認識部 4 を備え、第 2 認識部 4 は第 1 認識部 3 の前の階層の音声認識を実行した。これに加え、音声認識装置 100 は、第 2 認識部 4 のさらに前の階層での音声認識を実行する第 3 の認識部をさらに備えてもよい。この場合、認識結果処理部 5 は、所定の式又はマップを参照し、経過時間幅 T から各認識部が算出した尤度の重み付けを決定し、当該重み付けされた尤度に基づきソートを行う。上述の式又はマップは、実験等に基づき予め作成され、記憶部 9 に記憶される。4 以上の認識部を用いる場合についても、同様に拡張される。

[0082] 以上のように、音声認識装置 100 は、3 以上の階層にわたる認識処理についても、実行することができる。

[0083] (変形例 5)

第 1 実施例の説明では、音声認識装置 100 は、類似度として尤度を用いていた。しかし、本発明が適用可能な類似度の指標は、これに限定されない。これに代えて、音声認識装置 100 は、他の類似度を示す指標を用いてもよい。この場合であっても、音声認識装置 100 は、経過時間幅 T が短いほど誤認識があった可能性が高いと判断し、第 2 認識部 4 の認識結果の重みを大きくする。これによっても、音声認識装置 100 は、誤認識に起因した再発話を最小限のステップで認識することができる。

[0084] (変形例 6)

図 3 の説明では、認識結果処理部 5 は、経過時間幅 T が閾値 T_1 未満の場合、第 1 認識部 3 の認識結果と第 2 認識部 4 の認識結果とに基づき最終的な認識結果を決定していた。これに代えて、認識結果処理部 5 は、経過時間幅 T が閾値 T_1 未満の場合、第 2 認識部 4 の認識結果のみに基づき最終的な認識結果を決定してもよい。即ち、この場合、音声認識装置 100 は、再発話の可能性が高いと判断し、前の階層での音声認識のみを行い、尤度 L_2 のみに基づき最終的な認識語 WR の順位を決定する。これによっても、音声認識装置 100 は、誤認識に起因した再発話をより確実に最小限のステップで認識することができる。

[0085] [第 2 実施例]

次に、第 2 実施例で音声認識装置が実行する処理について説明する。第 2 実施例では、音声認識装置は、経過時間幅 T に加え、音声認識の環境を示す情報（以後、「環境情報 I_c 」と呼ぶ。）に基づき重み付け係数 w を決定する。これにより、音声認識装置は、音声認識を実行する環境に応じてよりの確に再発話を認識することができる。

[0086] 以後では、まず、第 2 実施例の音声認識装置の概略構成について説明した後、音声認識方法及びその処理フローについて順に説明する。

[0087] (概略構成)

図8は、第2実施例に係る音声認識装置100bの概略構成を示す図の一例である。音声認識装置100bは、車両に搭載される。音声認識装置100bは、ECU23、GPS受信機24、及びルート情報取得部25を備える点で第1実施例の音声認識装置100と異なる。なお、以後では、音声認識装置100bが搭載される車両を、「搭載車両」と呼ぶ。

[0088] ECU23は、図示しないCPU、ROM及びRAMなどを備え、搭載車両内の各構成要素に対して種々の制御を行う。また、ECU23は、主制御部7と電氣的に接続し、搭載車両の状態を示す情報（以後、「車両情報」と呼ぶ。）を主制御部7へ送信する。車両情報としては、例えば、ACC（アクセサリ電源：エンジンキー）の状態、車速パルスの状態、搭載車両が備える窓の開閉状態、エンジンの状態、トランスミッションの状態などの情報が該当する。

[0089] GPS受信機24は、複数のGPS衛星から、測位用データを含む下り回線データを搬送する電波を受信する。測位用データは、緯度及び経度情報等から搭載車両の絶対的な位置を検出するために用いられる。GPS受信機24は、主制御部7と電氣的に接続し、主制御部7に搭載車両の緯度経度情報を送信する。

[0090] ルート情報部25は、VICS（Vehicle Information Communication System）センタなどから配信される情報（以下、「VICS情報」と呼ぶ。）を電波より取得する。また、ルート情報部25は、地図情報を保持すると共に、搭載車両の運転者が目的地設定をしていた場合にはそのルート案内に関する情報（ルート案内情報）を保持する。ルート情報部25は、主制御部7と電氣的に接続し、主制御部7にこれらの情報の授受を行う。

[0091] 主制御部7は、図示しないマイク等から入力された発話データSaのS/N比などの音響情報を図示しないセンサ等を介して取得し、環境情報Icとして認識結果処理部5に供給する。また、主制御部7は、ECU23、GP

S 受信機 24、及びルート情報部 25 から供給される車両情報、緯度経度情報、ルート案内情報を取得し、環境情報 I c として認識結果処理部 5 に供給する。これについては、（音声認識方法）のセクションにて詳しく説明する。

[0092] （音声認識方法）

次に、第 2 実施例で音声認識装置 100b が実行する処理について説明する。認識結果処理部 5 は、主制御部 7 の制御に基づき、供給された環境情報 I c から重み付け関数 F_w の傾きを変更する。具体的には、認識結果処理部 5 は、音声認識の環境が劣悪であるほど、誤認識の可能性が高くなり再発話となされる可能性が高いと判断する。従って、この場合、認識結果処理部 5 は、当該傾きを緩めることで、第 2 認識部 4 の認識結果の重要度を大きくする。一方、認識結果処理部 5 は、音声認識の環境が良好であるほど、誤認識の可能性は低くなると判断する。従って、この場合、認識結果処理部 5 は、重み付け関数 F_w の傾きを急にすることで、第 1 認識部 3 の認識結果の重要度を大きくする。これにより、認識結果処理部 5 は、認識環境に応じて、適切に重み付け係数 w を設定することができる。

[0093] 以下では、各種の環境情報 I c から重み付け関数 F_w の傾きを変更する具体例について説明する。これらの具体例は、組み合わせて適用してもよい。

[0094] 1. 車両情報を用いた場合

認識結果処理部 5 は、車両情報から搭載車両の走行速度が速い場合、低速時よりも重み付け関数 F_w の傾きを緩くする。具体的には、認識結果処理部 5 は、例えば、各走行速度の値ごとに連続して傾きを変化させる。他の例では、認識結果処理部 5 は、走行速度を予めいくつかの範囲に区分けしておき、当該範囲ごとに傾きを定めてもよい。

[0095] これについて、補足説明する。搭載車両の走行速度が速い場合、運転者の運転への集中力が高まることに起因して音声認識装置 100b に対する反応が遅くなることが予想される。また、この場合、走行に起因した騒音レベルも上昇することが予想される。以上を考慮し、認識結果処理部 5 は、車両情

報から搭載車両の走行速度が速い場合、低速時よりも重み付け関数 F_w の傾きを緩くする。これにより、認識結果処理部 5 は、適切に重み付け係数 w を設定することができる。

[0096] 2. 緯度経度情報及びルート案内情報を用いた場合

認識結果処理部 5 は、緯度経度情報及びルート案内情報に基づき、走行中の経路が運転上注意を要する地点の場合には、重み付け関数 F_w の傾きを緩くする。具体的には、認識結果処理部 5 は、走行中の経路が都市部か否か、事故多発地点か否か、交差点であって搭載車両が左折又は右折する地点であるか否か、等を判断する。そして、認識結果処理部 5 は、走行経路が都市部、事故多発地点、又は／及び右左折地点等に該当する場合、誤認識する可能性が高いと判断し、重み付け関数 F_w の傾きを緩くする。

[0097] 以上のように、認識結果処理部 5 は、緯度経度情報及びルート案内情報を用いることで、適切に重み付け係数 w を設定することができる。

[0098] (処理フロー)

次に、第 2 実施例で音声認識装置 100b が実行する処理手順について説明する。ここでは、第 1 実施例の処理フローの説明で用いた図 6 を再び参照して説明する。以後では、第 1 実施例と処理が異なる部分について説明する。

[0099] 第 2 実施例では、音声認識装置 100b は、ステップ S105 の実行後、環境情報 I_c を ECU23 等から取得する。そして、音声認識装置 100b は、環境情報 I_c に基づき重み付け関数 F_w を決定する。例えば、音声認識装置 100b は、環境情報 I_c に応じた複数の重み付け関数 F_w を予め複数用意し、所定のマップ等を参照することで、環境情報 I_c から使用する重み付け関数 F_w を決定する。上述のマップ等は、実験等により予め作成される。他の例では、音声認識装置 100b は、重み付け関数 F_w の傾きを決定するためのパラメータを環境情報 I_c に応じてマップ等を参照して変更する。上述のマップ等は、実験等により予め作成される。そして、音声認識装置 100b は、ステップ S106 で、経過時間幅 T に基づき重み付け関数 F_w か

ら第1及び第2重み付け係数 w_1 、 w_2 を決定する。これにより、音声認識装置100bは、音声認識の環境を考慮して、よりの確に再発話を認識することができる。

[0100] (変形例1)

上述の第2実施例の説明では、認識結果処理部5は、認識環境が劣悪であると判断した場合、重み付け関数 F_w の傾きを緩やかに設定した。これに代えて、又は、これに加えて、認識結果処理部5は、認識環境が劣悪であると判断した場合、例えば経過時間幅 T の計測を一時的に停止させてもよい。他の例では、認識結果処理部5は、認識環境が劣悪であると判断した場合、経過時間幅 T を所定の値で割る又は減算することで小さくしてもよい。これらによっても、音声認識装置100bは、認識環境に応じて適切に音声認識処理を実行することができる。

[0101] (変形例2)

第1実施例の(変形例1)乃至(変形例6)は、第2実施例にも適用することができる。即ち、この場合、音声認識装置100bは、上述した第2実施例の処理に加え、第1実施例の(変形例1)乃至(変形例6)から選択した1又は複数の処理を行う。

産業上の利用可能性

[0102] 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

- [0103]
- 1 音声分析部
 - 2 音声区間検出部
 - 3 第1認識部
 - 4 第2認識部
 - 5 認識結果処理部

- 6 提示部
- 7 主制御部
- 8 時間計測部
- 9 記憶部
- 20 認識部
- 21 認識結果保持部
- 23 ECU
- 24 GPS
- 25 ルート情報取得部

請求の範囲

[請求項1]

階層コマンドを音声認識する音声認識装置であって、
前記階層コマンドを認識するための各階層の音声認識で使用する辞書を記憶する辞書記憶手段と、
音声データの入力があった場合、前回入力された音声データを正しく認識したと仮定した場合に使用すべき辞書に基づき音声認識を行う第1音声認識手段と、
前記音声データの入力があった場合、前記前回入力された音声データの音声認識を実行した際に使用した辞書に基づき音声認識を行う第2音声認識手段と、
前記第1音声認識手段による認識結果と、前記第2音声認識手段による認識結果と、を前記前回入力された音声データの処理時から起算した経過時間幅に基づき重み付けをして最終認識結果を決定する認識結果処理手段と、
を備えることを特徴とする音声認識装置。

[請求項2]

前記認識結果処理手段は、前記経過時間幅が所定幅以上の場合、前記第1音声認識手段の認識結果のみに基づき前記最終認識結果を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]

前記認識結果処理手段は、前記経過時間幅が所定幅未満の場合、前記第2音声認識手段の認識結果のみに基づき前記最終認識結果を決定することを特徴とする請求項1又は2に記載の音声認識装置。

[請求項4]

前記第1音声認識手段及び前記第2音声認識手段は、認識結果として尤度を算出し、

前記認識結果処理手段は、前記第1音声認識手段が算出した尤度に前記経過時間幅に対して非増加関数となる第1パラメータを乗じると共に、前記第2音声認識手段が算出した尤度に前記経過時間幅に対して非減少関数となる第2パラメータを乗じることによって前記重み付けを実行することを特徴とする請求項1乃至3のいずれか一項に記載の音声

認識装置。

[請求項5] 前記認識結果処理手段は、前記第1パラメータと前記第2パラメータとの和を一定値にすることを特徴とする請求項4に記載の音声認識装置。

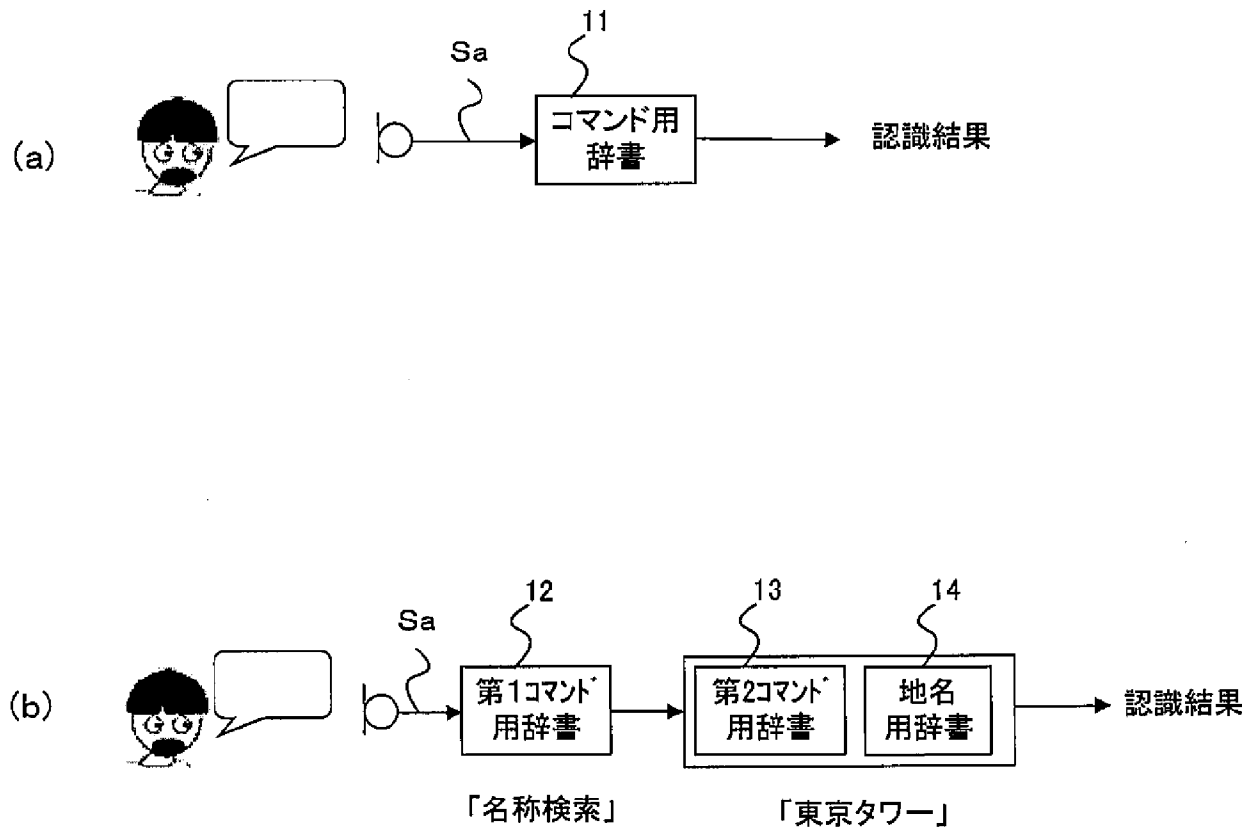
[請求項6] 認識環境に関する情報を取得する環境情報取得手段をさらに備え、
前記認識結果処理手段は、前記情報に基づき認識環境が劣悪であると判断した場合、前記認識環境が劣悪でない場合に比べ、前記第2音声認識手段の認識結果の重みを大きくすることを特徴とする請求項1乃至5のいずれか一項に記載の音声認識装置。

[請求項7] 車両に搭載され、
前記環境情報取得手段は、前記車両の車速又は／及び音声データに含まれる雑音の度合いを取得し、
前記認識結果処理手段は、前記車速又は／及び前記雑音の度合いが大きいほど、前記第2音声認識手段の認識結果の重み付けを大きくすることを特徴とする請求項6に記載の音声認識装置。

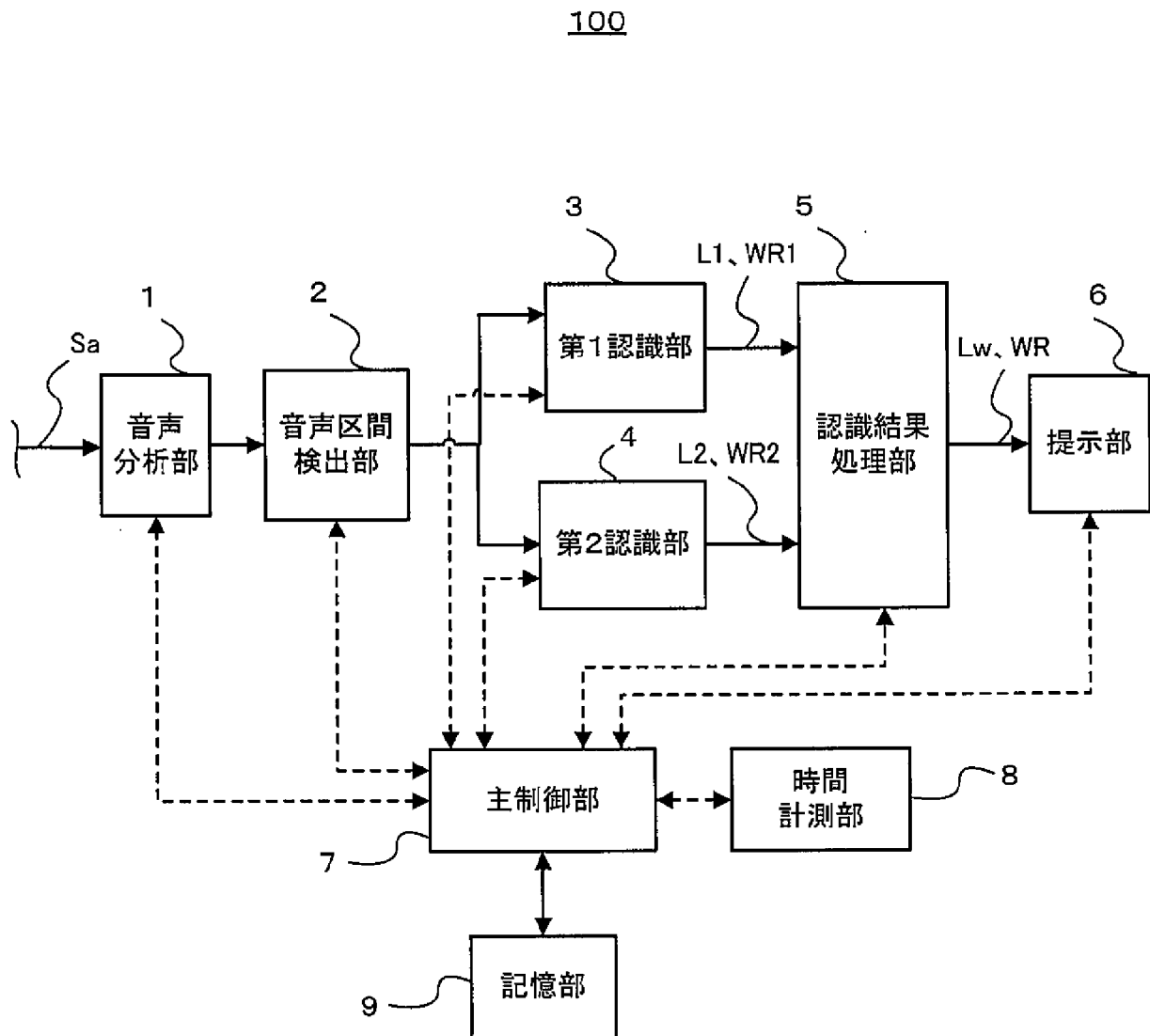
[請求項8] 階層コマンドを認識するための各階層の音声認識で使用する辞書を記憶する音声認識装置が使用する音声認識方法であって、
音声データの入力があった場合、前回入力された音声データを正しく認識したと仮定した場合に使用すべき辞書に基づき音声認識を行う第1音声認識工程と、
前記音声データの入力があった場合、前記前回入力された音声データの音声認識を実行した際に使用した辞書に基づき音声認識を行う第2音声認識工程と、
前記第1音声認識手段による認識結果と、前記第2音声認識手段による認識結果と、を前記前回入力された音声データの処理時から起算した経過時間幅に基づき重み付けをして最終認識結果を決定する認識結果処理工程と、
を備えることを特徴とする音声認識方法。

- [請求項9] 階層コマンドを認識するための各階層の音声認識で使用する辞書を記憶する音声認識装置により実行される音声認識プログラムであって、
- 、
- 音声データの入力があった場合、前回入力された音声データを正しく認識したと仮定した場合に使用すべき辞書に基づき音声認識を行う第1音声認識手段と、
- 前記音声データの入力があった場合、前記前回入力された音声データの音声認識を実行した際に使用した辞書に基づき音声認識を行う第2音声認識手段と、
- 前記第1音声認識手段による認識結果と、前記第2音声認識手段による認識結果と、を前記前回入力された音声データの処理時から起算した経過時間幅に基づき重み付けをして最終認識結果を決定する認識結果処理手段、
- として前記音声認識装置を機能させることを特徴とする音声認識プログラム。
- [請求項10] 請求項9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。

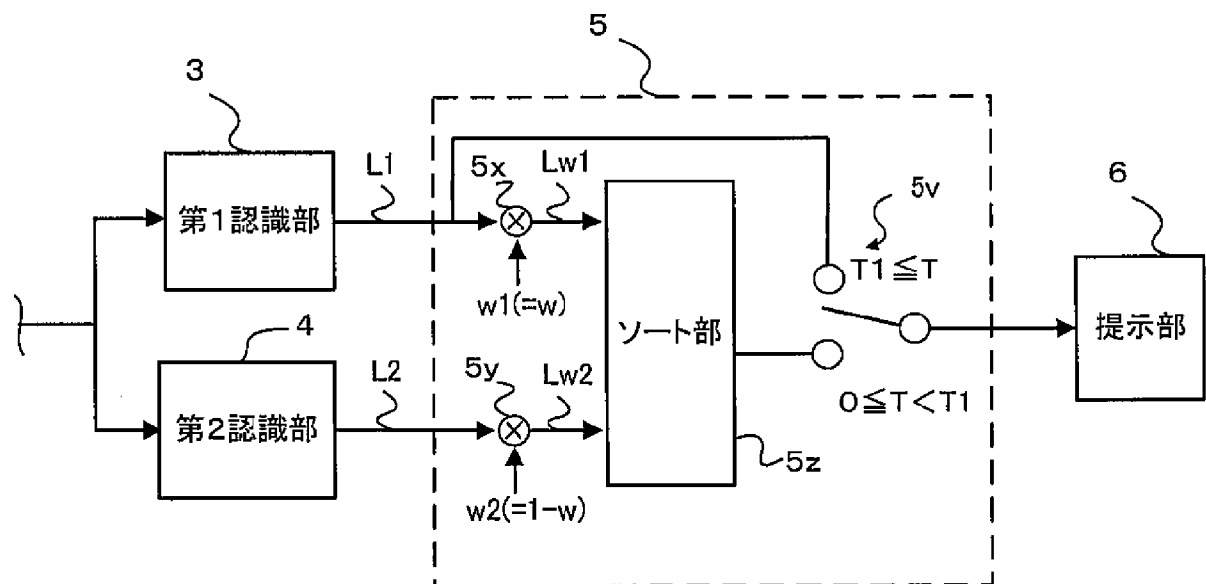
[図1]



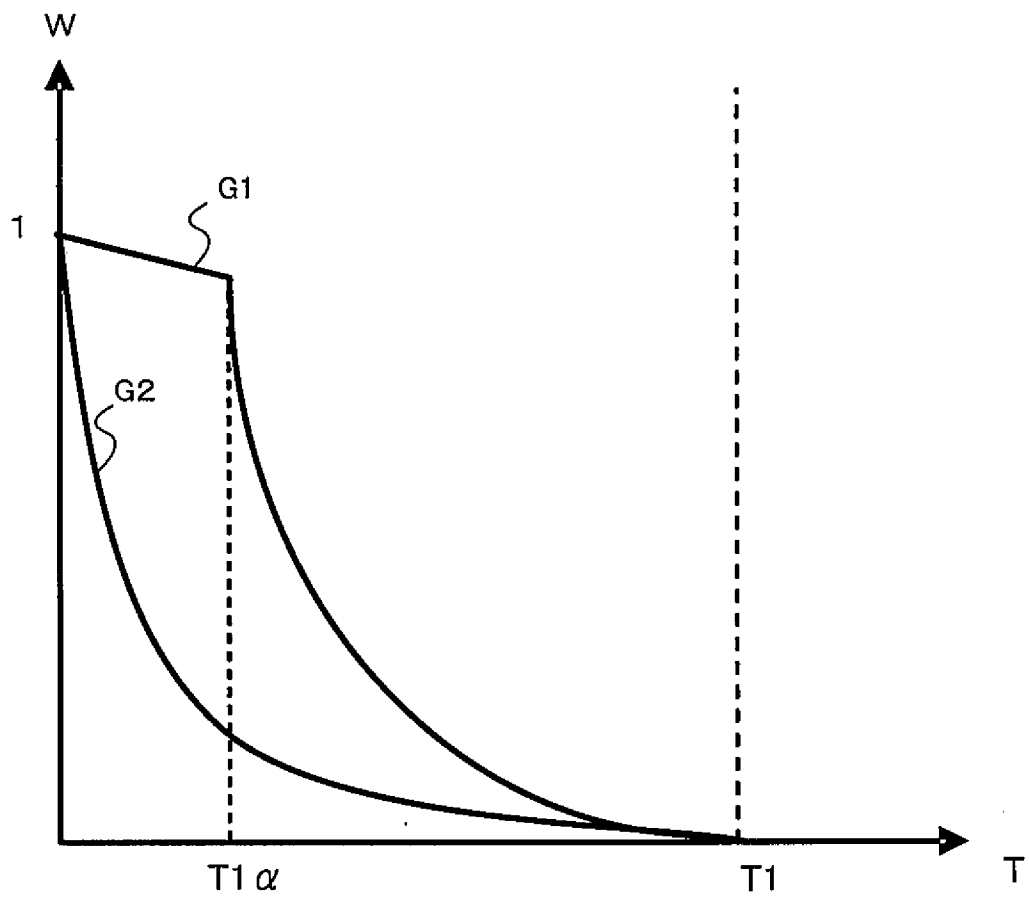
[図2]



[図3]



[図4]



[図5]

(a)

順位	認識語WR1	尤度L1	尤度Lw1
1	ルート探索	-4	-2.8
2	ルート消去	-6	-4.2
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮

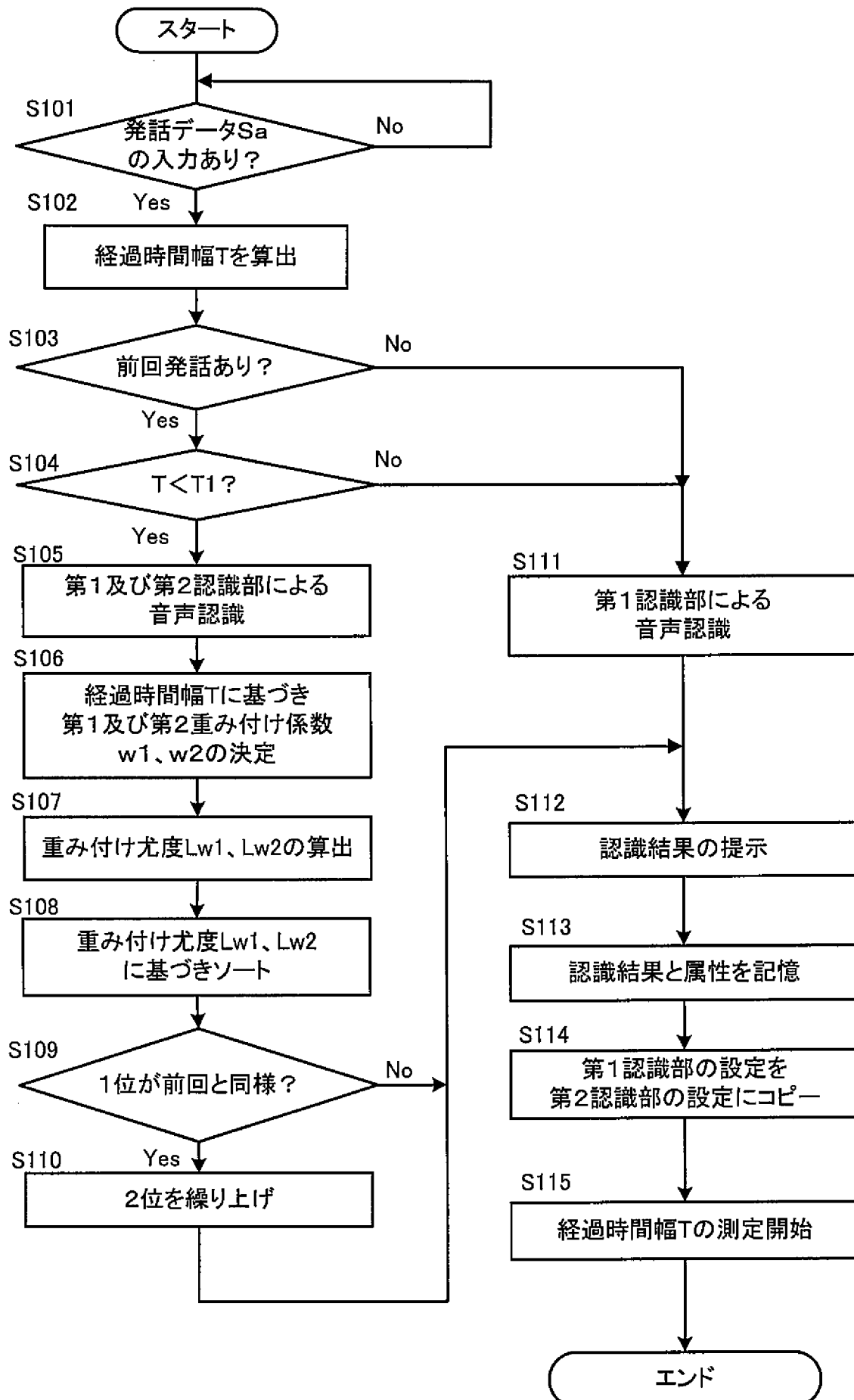
(b)

順位	認識語WR2	尤度L2	尤度Lw2
1	東京タワー	-5	-1.5
2	東京港	-7	-2.1
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮

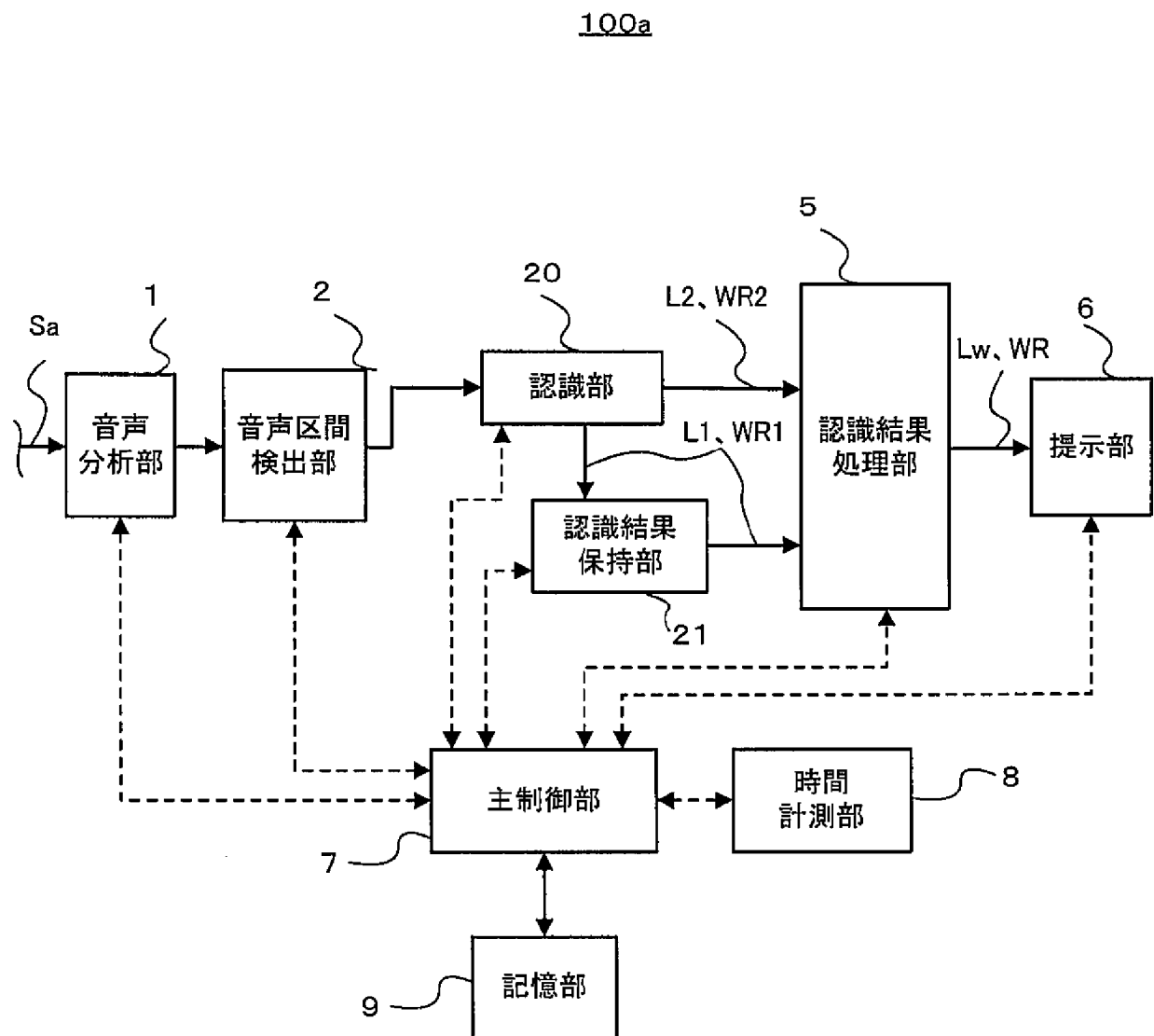
(c)

順位	認識語WR	尤度Lw
1	東京タワー	-1.5
2	東京港	-2.1
3	ルート探索	-2.8
4	ルート消去	-4.2

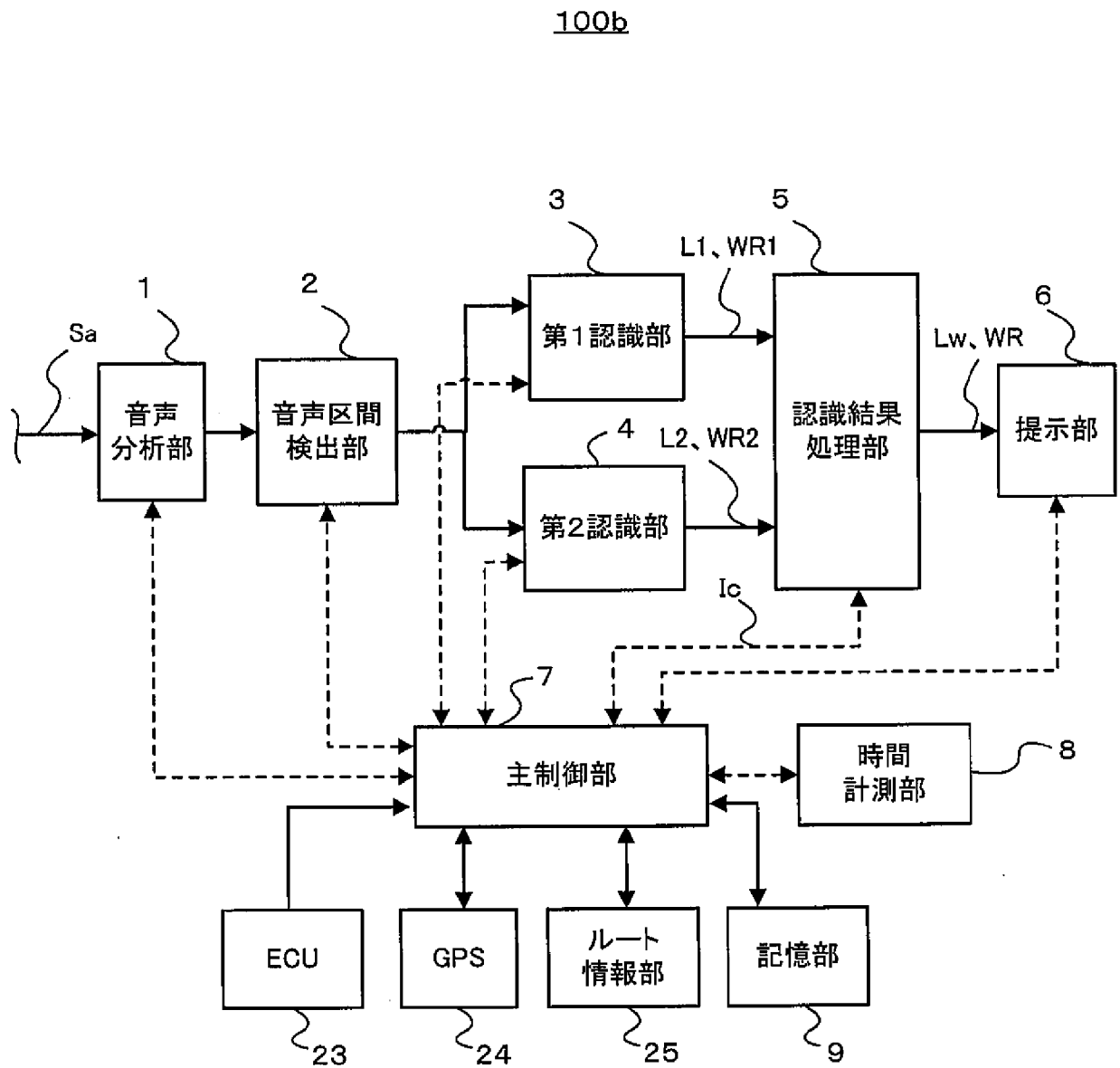
[図6]



[図7]



[図8]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/063996

A. CLASSIFICATION OF SUBJECT MATTER

G10L15/06(2006.01)i, G10L15/18(2006.01)i, G10L15/22(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)

G10L15/00-15/28

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)

JSTPlus(JDreamII)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2008-89625 A (Honda Motor Co., Ltd.), 17 April, 2008 (17.04.08), Full text; all drawings (Family: none)	1-10
A	JP 2002-41078 A (Sharp Corp.), 08 February, 2002 (08.02.02), Full text; all drawings (Family: none)	1-10
A	JP 2003-177788 A (Fujitsu Ltd.), 27 June, 2003 (27.06.03), Full text; all drawings (Family: none)	1-10

☒ 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
31 August, 2009 (31.08.09)

Date of mailing of the international search report
08 September, 2009 (08.09.09)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/063996

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 4-199198 A (Matsushita Electric Industrial Co., Ltd.), 20 July, 1992 (20.07.92), Full text; all drawings (Family: none)	1-10

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G10L15/06(2006.01)i, G10L15/18(2006.01)i, G10L15/22(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. G10L15/00-15/28

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）

JSTPlus(JDreamII)

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
A	JP 2008-89625 A（本田技研工業株式会社）2008.04.17, 全文, 全図（ファミリーなし）	1-10
A	JP 2002-41078 A（シャープ株式会社）2002.02.08, 全文, 全図（ファミリーなし）	1-10
A	JP 2003-177788 A（富士通株式会社）2003.06.27, 全文, 全図（ファミリーなし）	1-10

☒ 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

3 1 . 0 8 . 2 0 0 9

国際調査報告の発送日

0 8 . 0 9 . 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）

菊池 智紀

5 Z

3 3 5 2

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
A	JP 4-199198 A (松下電器産業株式会社) 1992. 07. 20, 全文, 全図 (ファミリーなし)	1-10