(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
2013 年 10 月 17 日(17.10.2013)

WIPO | PCT

(10) 国際公開番号
WO 2013/154010 A1

(51) 国際特許分類:
G10L 15/32 (2013.01) *G10L 15/22* (2006.01)
G10L 15/06 (2013.01) *G10L 15/30* (2013.01)

(21) 国際出願番号: PCT/JP2013/060238

(22) 国際出願日: 2013 年 4 月 3 日(03.04.2013)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:
特願 2012-088230 2012 年 4 月 9 日(09.04.2012) JP

(71) 出願人: クラリオン株式会社 (CLARION CO., LTD.) [JP/JP]; 〒3300081 埼玉県さいたま市中央区新都心７番地２ Saitama (JP).

(72) 発明者: 大淵 康成 (OBUCHI Yasunari); 〒1858601 東京都国分寺市東恋ヶ窪一丁目２８０番地 株式会社日立製作所 中央研究所内 Tokyo (JP). 本間 健 (HOMMA Takeshi); 〒3300081 埼玉県さいたま市中央区新都心７番地２ クラリオン株式会社内 Saitama (JP).

(74) 代理人: ポレール特許業務法人 (POLAIRE I.P.C.); 〒1040032 東京都中央区八丁堀二丁目７番１号 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, BN, 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, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

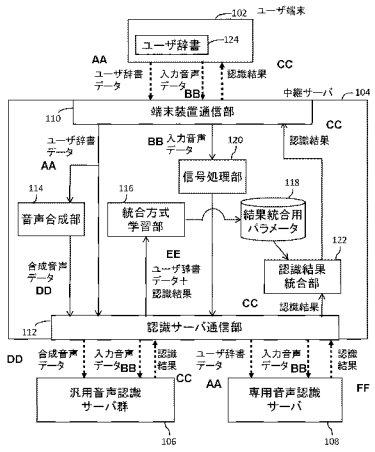
(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI

[続葉有]

(54) **Title**: VOICE RECOGNITION SERVER INTEGRATION DEVICE AND VOICE RECOGNITION SERVER INTEGRATION METHOD

(54) 発明の名称 : 音声認識サーバ統合装置および音声認識サーバ統合方法

図1



102 User terminal
104 Relay server
106 General purpose voice recognition server group
108 Dedicated voice recognition server
110 Terminal device communication unit
112 Recognition server communication unit
114 Voice synthesis unit
116 Integration method learning unit
118 Parameter for result integration
120 Signal processing unit
122 Recognition results integration unit
124 User dictionary
AA User dictionary data
BB Input voice data
CC Recognition results
DD Synthesized voice data
EE User dictionary data + recognition results
FF Recognition results

(57) **Abstract**: Provided is a voice recognition function for integrating in an optimal manner the voice recognition results of a general purpose voice recognition server and the voice recognition results of a dedicated voice recognition server, and which produces few errors in the end. A dedicated voice recognition server (108) is constructed using phrases contained in user dictionary data, and using such data, the functions of a general purpose voice recognition server (106) are evaluated in advance. On the basis of these evaluation results, for recognition results obtained from dedicated and general use voice recognition servers, information related to which results to utilize, and with respect to those results, what weighting to perform thereon to obtain optimal recognition results is preliminarily held as a parameter for result integration (118) in the form of a database. When conducting recognition, optimal recognition results are obtained by comparing recognition results according to dedicated and general purpose voice recognition servers with the parameter for result integration (118).

(57) 要約: 汎用音声認識サーバの音声認識結果と専用音声認識サーバの音声認識結果とを最適な形で統合し、最終的に間違いの少ない音声認識機能を提供する。 ユーザ辞書データに含まれる語句を使って専用の音声認識サーバ１０８を構築するとともに、それらのデータを用い、汎用の音声認識サーバ１０６の性能を事前評価する。この評価結果をもとに、専用および汎用の音声認識サーバから得られた認識結果に対し、どれを採用し、それらに対してどのような重み付けを行えば最適な認識結果が得られるかを、結果統合用パラメータ１１８としてデータベースの形で保持しておく。認識実行時には、専用および汎用の音声認識サーバによる認識結果を、結果統合用パラメータ１１８と比較することにより、最適な認識結果を得る。

WO 2013/154010 A1

(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)）

明　細　書

発明の名称 ：

音声認識サーバ統合装置および音声認識サーバ統合方法

技術分野

[0001] 本発明は、ユーザが音声を用いて操作を行うための端末装置と、音声データを認識してその結果を返す複数の音声認識サーバとを繋ぎ、複数の音声認識サーバにより得られる認識結果を統合して最適な音声認識結果をユーザに提供するための装置および方法に関する。

背景技術

[0002] 車載情報機器や携帯電話などの操作を、ユーザの声により行うための音声認識機能が、数多くの機器に搭載されている。さらに近年では、データ通信技術の発展により、音声データをサーバに送信し、サーバの潤沢な計算資源を用いてより高精度の音声認識を行う方式が普及しつつある。また、特許文献１に示されるように、こうした音声認識サーバをより有効に用いるため、個人用端末と音声認識サーバとの間に中継サーバを置き、その中で付加的な処理を行う方式も提案されている。

先行技術文献

特許文献

[0003] 特許文献1：特開２００８－２４２０６７号公報

特許文献2：特開２００２－１１６７９６号公報

特許文献3：特開２０１０－２２４３０１号公報

発明の概要

発明が解決しようとする課題

[0004] 音声認識サーバが汎用のサービスとして運営される例が増えており、ユーザが扱う端末向けのサービスを提供する主体と、音声認識サーバを運営する主体とが異なる場合が多くなっている。また、運営主体が同一である場合においても、音声認識サーバの開発と音声認識アプリケーションの開発を別個

に行い、それらが相互に最適化されていない場合もある。このような状況で、汎用の音声認識サーバが全般的には高い性能を示しながら、特定の語句に対しては必ずしも十分な性能を示さないということがある。

[0005] 一方、特定のアプリケーションを使用する特定のユーザに着目した場合、そのユーザの知人の名前や好きな音楽の楽曲名など、一般的ではないが重要度の高い語句が存在する。こうした語句を認識するためには、専用の音声認識サーバを設けることが望ましいが、専用音声認識サーバの開発には十分なコストをかけられないことが多く、一般的な語句に対する性能では、汎用音声認識サーバに劣ることになる。このように、汎用音声認識サーバと専用音声認識サーバとでは、語句によって認識の得手不得手があり、音声認識性能が異なる。したがって、ユーザの発した語句によってこれらを使い分けることが求められるが、音声認識というのが「内容のわからない発話に対して内容を推定する」タスクである以上、事前に発話内容を知ってサーバの使い分けを行うことは原理的に不可能である。

[0006] 本発明は、汎用音声認識サーバの音声認識結果と専用音声認識サーバの音声認識結果とを最適な形で統合し、最終的に間違いの少ない音声認識機能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

[0007] 本発明においては、ユーザが持つ端末装置に含まれる特定単語のリストをあらかじめ入手し、それらの単語のデータをもとに、専用の音声認識サーバを構築する。また、それらのデータを用い、汎用音声認識サーバの性能を事前評価する。その評価結果をもとに、専用および汎用の音声認識サーバから得られた認識結果の中で、どれを採用し、それらに対してどのような重み付けを行えば最適な認識結果が得られるかを、データベースの形で保持しておく。ユーザが実際に音声認識機能を用いる際には、専用および汎用の音声認識サーバにより入力音声を認識した後、そこで得られる結果を、先に述べたデータベースの内容と比較することにより、最適な音声認識結果を得る。また、事前評価の基準として、音声認識の正しさに加えて応答速度を用いるこ

とで、なるべく正しい音声認識結果を、なるべく早く得ることができるようにする。

[0008] 本発明の音声認識サーバ統合装置の一例を挙げるならば、ユーザが音声を用いて操作を行うための端末装置と、音声データを認識してその結果を返す音声認識サーバとの間を中継する装置であって、ユーザが登録した語句もしくはユーザがよく使う語句のリストに基づいて認識結果統合用のパラメータを学習し保存する統合方式学習部と、ユーザが音声認識を意図して発した音声のデータを前記端末装置から受信する手段と、前記受信した音声データを汎用音声認識サーバおよび専用音声認識サーバに送信する手段と、前記音声データの前記汎用音声認識サーバおよび専用音声認識サーバによる認識結果を受信する手段と、前記汎用音声認識サーバおよび専用音声認識サーバによる認識結果を、前記保存された認識結果統合用のパラメータと比較し、最適な認識結果を選択する認識結果統合部と、前記選択された認識結果を前記端末装置に送信する手段とを備えるものである。

[0009] 本発明の音声認識サーバ統合装置において、更に、前記端末装置からユーザが登録した語句もしくはユーザがよく使う語句のリストを受信する手段と、前記受信した語句をもとに合成音声を生成する音声合成部と、前記生成された合成音声を前記汎用音声認識サーバおよび専用音声認識サーバに送信する手段と、前記合成音声の前記汎用音声認識サーバおよび専用音声認識サーバによる認識結果を受信する手段を備え、前記統合方式学習部は、前記合成音声の基となった語句と前記認識結果とを合わせて解析し、認識結果統合用パラメータを学習し保存するものでよい。

[0010] また、本発明の音声認識サーバ統合装置において、更に、前記端末装置からユーザが登録した語句もしくはユーザがよく使う語句のリストを受信する手段と、前記汎用音声認識サーバから認識用の語句リストを受信する手段と、前記認識用の語句リストを前記端末装置から受信した語句リストと比較し、類似度を推定する語句比較・類似度推定部とを備え、前記統合方式学習部は、前記推定結果を認識結果統合用のパラメータとして保存するものでよい

。

[0011] 本発明の音声認識サーバ統合方法の一例を挙げるならば、ユーザが登録した語句もしくはユーザがよく使う語句のリストに基づいて認識結果統合用のパラメータを学習して保存するステップと、ユーザが音声認識を意図して発した音声のデータを汎用音声認識サーバおよび専用音声認識サーバに送信するステップと、前記音声データの前記汎用音声認識サーバおよび専用音声認識サーバによる認識結果を受信するステップと、汎用音声認識サーバの認識結果および専用音声認識サーバの認識結果と、前記認識結果統合用パラメータとを比較して、最適な音声認識結果を選択するステップと、から成るものである。

発明の効果

[0012] 本発明により、一般的な語句に関しては汎用音声認識サーバの認識結果を重要視し、ユーザ固有の語句に関しては専用音声認識サーバの結果を重要視するなど、個々の入力に対して最適な形で認識結果の統合が行われ、最終的に間違いの少ない音声認識機能をユーザに提供することが可能となる。また、間違いが少ないだけでなく、応答速度の点でも利便性の高いシステムを実現することができる。

図面の簡単な説明

[0013] [図1]本発明の実施例１の音声認識サーバ統合装置の構成図である。

[図2]本発明の実施例１の音声合成を用いた結果統合用パラメータ推定の処理を示す図である。

[図3]本発明の単一汎用音声認識サーバを用いた結果統合用パラメータの一例を示す図である。

[図4]本発明の複数汎用音声認識サーバを用いた結果統合用パラメータの一例を示す図である。

[図5]本発明の実施例１の複数サーバの認識結果の統合方法の一例を示す図である。

[図6]本発明の複数汎用音声認識サーバの認識結果信頼度を用いた結果統合用

パラメータの一例を示す図である。

[図7]本発明の複数汎用音声認識サーバの認識結果信頼度と誤認識結果を用いた結果統合用パラメータの一例を示す図である。

[図8]本発明の同音異表記の変換を利用した認識結果の統合方法の一例を示す図である。

[図9]本発明を実現するためのユーザ端末の構成例を示す図である。

[図10]本発明におけるユーザ辞書の作成方法の一例を示す図である。

[図11]本発明における音声合成部の構成の一例を示す図である。

[図12]本発明の応答時間を考慮した結果統合用パラメータの一例を示す図である。

[図13]本発明の実施例２の音声認識サーバ統合装置の構成図である。

[図14]本発明の実施例３の音声認識サーバ統合装置の構成図である。

[図15]本発明の実施例３の認識用語句リストを用いた結果統合用パラメータ推定の処理を示す図である。

[図16]本発明の実施例４の音声認識サーバ装置の構成図である。

[図17]本発明の実施例５の音声認識サーバ装置の構成図である。

発明を実施するための形態

[0014] 以下、図面を用いて本発明の実施例を説明する。なお、発明を実施するための形態を説明するための全図において、同一の機能を有する要素には同一の名称、符号を付して、その繰り返しの説明を省略する。

実施例 1

[0015] 図１は、本発明の実施例１に基づく音声認識サーバ統合装置の構成例を示す図である。音声認識機能は、ユーザ端末１０２、中継サーバ１０４、汎用音声認識サーバ群１０６、専用音声認識サーバ１０８を用いて提供される。なお、汎用音声認識サーバ群１０６は、単一の汎用音声認識サーバでも構わない。

[0016] ユーザ端末１０２は、ユーザ個人が持つ端末装置で、入力音声データの取得および音声認識結果に基づくサービスの提供を行う他に、アドレス帳や楽

曲名リストなどのユーザに固有の語句リストを保持している。以下では、これらのユーザに固有の語句リストのことを「ユーザ辞書」と呼ぶ。ユーザ辞書には、ユーザが登録した語句もしくはユーザがよく使う語句のリストが保持されている。

[0017] 汎用音声認識サーバ群１０６は、本発明により実現されるサービスのみにより使用されることを想定していない、１台以上の音声認識サーバである。一般に、大規模な語句リストを内蔵し、様々な言葉に対する認識性能が高い一方、ユーザ辞書に含まれる一部の語句については、正しく認識できない可能性がある。

[0018] 専用音声認識サーバ１０８は、本発明により実現されるサービスに特化した音声認識サーバであり、ユーザ辞書に含まれる語句のすべてもしくは大半を認識するように設計されている。専用音声認識サーバ１０８は、ユーザ辞書に含まれない語句が入力された場合には、「認識結果なし」という結果が出力されるよう設計されている。専用音声認識サーバは、サーバとして構成されるものにかぎらず、専用音声認識装置でも良いし、また、実施例２や実施例５のように、ユーザ端末や中継サーバに内蔵されるものでもよい。

[0019] 中継サーバ１０４は、本発明の「音声認識サーバ統合装置」に該当するもので、ユーザ端末１０２と音声認識サーバ１０６，１０８とを繋ぎ、音声認識結果の統合などを行う。ユーザ端末１０２とのデータのやりとりは、端末装置通信部１１０を介して行う。また、音声認識サーバ１０６，１０８とのデータのやりとりは、認識サーバ通信部１１２を介して行う。中継サーバ１０４は、端末装置通信部１１０、音声合成部１１４、統合方式学習部１１６、信号処理部１２０、認識結果統合部１２２、認識サーバ通信部１１２などから構成されている。

[0020] 中継サーバ１０４の動作を説明する。はじめに、ユーザがユーザ端末１０２を通信可能な状態にセットすると、ユーザ辞書１２４のデータが端末装置通信部１１０を経由して送信される。このデータは直接認識サーバ通信部１１２に送られ、さらに専用音声認識サーバ１０８に送られる。専用音声認識

サーバ１０８では、送られてきたユーザ辞書データに基づき、そこに含まれる語句を正しく認識できるようチューニングを行う。一方、端末装置通信部１１０で受信されたユーザ辞書データは、音声合成部１１４にも送られる。ここでは、文字列として送られてきたユーザ辞書データをもとに、合成音声データが作られる。一つの語句に対する合成音声データは、一つでも良いし、音質の違う複数のものであっても良い。作成された合成音声データは、認識サーバ通信部１１２を介して、汎用音声認識サーバ群１０６および専用音声認識サーバ１０８に送られる。これらに対する認識結果が各サーバから返されると、認識サーバ通信部１１２がそれを受信し、統合方式学習部１１６に送る。統合方式学習部１１６では、合成音声のもととなったユーザ辞書データと認識結果とを合わせて解析し、認識結果統合のためのパラメータを学習する。得られたパラメータは、結果統合用パラメータ１１８として保存される。この時点で、本発明を用いたシステムの事前学習処理が終了する。

[0021] ユーザが実際に音声インタフェースを使う際には、ユーザ端末１０２で取得した入力音声データが、端末装置通信部１１０により受信される。受信されたデータは、信号処理部１２０に送られ、必要な処理が施される。ここで、必要な処理とは、例えば雑音を含む入力音声から雑音を取り除くこと等を指すが、必ずしも必須ではなく、何も処理をしなくても良い。信号処理部１２０から出力されたデータは、認識サーバ通信部１１２を経て、汎用音声認識サーバ群１０６および専用音声認識サーバ１０８に送られる。これらのサーバから返された認識結果は、認識サーバ通信部１１２を経て、認識結果統合部１２２に送られる。認識結果統合部１２２では、複数の認識結果と、結果統合用パラメータ１１８に含まれるパラメータとを比較して、最適な認識結果を選択する。選択された認識結果は、端末装置通信部１１０を経て、ユーザ端末１０２に送られる。ユーザ端末１０２では、この結果をもとに、ナビゲーション機能の目的地を設定する、電話をかける、楽曲を再生するなどのサービスを提供する。

[0022] 図２は、図１に示した構成において、ユーザ辞書データを使って結果統合

用パラメータを作成するまでの処理の手順を示す図である。まず、ユーザ辞書データは、そのまま専用音声認識サーバに送られる。専用音声認識サーバでは、送られてきた語句を認識対象とするよう、音声認識エンジンをチューニングする。従って、ユーザ辞書に含まれない語句を発声したデータが送られてきた場合、専用音声認識サーバは、間違った結果を返すか、もしくは認識不能という結果を返すことになる。一方、ユーザ辞書データは、音声合成部にも送られ、そこで合成音声データが作成される。通常、一つの語句に対しては一つの合成音声が作られるが、音声合成部が、話者や話速、声の高さなどを選択できるような機能を持っている場合は、それらを変化させて、同じ語句に対して複数の合成音声データを作成すれば、後段で行う統合方式学習の性能をより高めることができる。

[0023] こうして得られた合成音声データは、各汎用音声認識サーバおよび専用音声認識サーバに送られる。それらのサーバからは、認識結果が返される。また、認識結果だけではなく、それに付随する信頼度スコアが一緒に帰される場合もある。これらを元に、統合方式学習部で統合方式を学習し、その結果を結果統合用パラメータに保存する。

[0024] 図３は、結果統合用パラメータの最も簡単な構成の例を示す図である。この例では、汎用音声認識サーバが1台だけ存在すると仮定し、そのサーバで、ユーザ辞書の各語句が正しく認識されたかどうかだけを○と×で保持している。即ち、「鈴木一郎」「山田二郎」という語句は汎用音声認識サーバで正しく認識されたが、それ以外は正しく認識されなかったということを、この図は表している。図４は、同様の学習を、３台の汎用音声認識サーバを用いて行った際の例である。

[0025] 図３、図４に示したような結果を使って実際に認識を行う際の処理の手順を、図５に示す。入力音声データは、はじめに信号処理部で事前処理される。信号処理部での処理の代表的なものとして、特許文献１に示されるような雑音抑圧処理が挙げられる。信号処理部での処理の結果、一つの入力音声データに対して一つの音声データが得られるのが普通であるが、設定を変えて

複数の音声データが得られる場合もある。そのような場合には、以下に述べる処理を、音声データの数だけ繰り返す。また、信号処理部での処理が不要と思われる場合には、入力音声データをそのまま信号処理部の出力データとする。

[0026] 信号処理部の出力データは、汎用音声認識サーバおよび専用音声認識サーバに送られる。これらの結果がすべて認識結果統合部に送られる。認識結果統合部では、まず専用音声認識サーバの認識結果をチェックする。専用認識サーバの認識結果が、「認識結果なし」であった場合、汎用音声認識サーバの認識結果のみから最終的な認識結果を決定する。すなわち、汎用音声認識サーバが1台しかない場合は、その結果をそのまま採用する。複数台ある場合には、それらの認識結果のあいだで多数決を取る。多数決を取る際、各認識サーバが信頼度スコアを付与する場合であれば、その値で重み付けをした多数決とすることもできる。また、事前に各認識サーバの性能を推定して、重み付けの係数とすることもできる。このような、一般的な語句に対する複数の音声認識サーバの認識結果の統合については、特許文献２に示されるような公知の技術を用いることが可能である。

[0027] 一方、専用音声認識サーバの認識結果として、ユーザ辞書データに含まれる語句が得られた場合、図３や図４に示したような結果統合用パラメータを参照する。例えば、図３の例で、専用音声認識サーバの認識結果が「日立太郎」であった場合、結果統合用パラメータの該当する行を見ると、この語句は汎用音声認識サーバでは認識できないはずだということがわかるので、専用音声認識サーバの結果をそのまま採用する。一方、専用音声認識サーバの認識結果が「鈴木一郎」であった場合、結果統合用パラメータの該当する行を見ると、この語句は専用音声認識サーバでも認識されうるということがわかる。そこで次に汎用音声認識サーバの認識結果をチェックする。汎用音声認識サーバの認識結果も「鈴木一郎」である場合にはそのまま「鈴木一郎」を最終的な認識結果とすれば良いが、そうでない場合には、一般的に性能が高いと思われる汎用音声認識サーバの結果を優先するか、もしくは汎用音声

認識サーバと専用音声認識サーバの認識結果のうち、信頼度スコアの高い方を最終認識結果として採用する。これにより、「鈴木一郎」と似た発音の言葉が、専用音声認識サーバにより誤認識されてしまった場合であっても、汎用音声認識サーバの認識結果に基づきこれを棄却することができるようになる。図４の例でも同様であり、「日立太郎」については無条件で専用音声認識サーバの結果を採用する。「鈴木一郎」については３台の汎用音声認識サーバすべてが認識可能な語句であるので、これらの認識結果での多数決、もしくはこれらすべてに専用音声認識サーバも加えての多数決により最終認識結果を決定する。また、専用音声認識サーバの認識結果が「山田二郎」であった場合には、これを正しく認識できる可能性のある汎用音声認識サーバは１番のみであることから、このサーバと専用音声認識サーバとの間で、図３の例と同じ処理を行うことにより最終認識結果を得る。

[0028] 図６は、図３や図４とは異なるもう一つの結果統合用パラメータの実現例である。ここでは、ある語句が各汎用音声認識サーバで認識可能な場合に、その語句が正しく認識される確率を重みの数値に置き換えて保持している。ここで、正しく認識される確率は、たとえば「鈴木一郎」という語句に対し、音声合成用パラメータを様々に変えて作った合成音声による認識を行い、それらに対する認識結果のうち何個が正しいものであったかにより推定することができる。また、汎用音声認識サーバが複数の認識結果候補を返す仕様になっている場合には、正解単語の平均順位や平均信頼度スコアなどを用いることもできる。これらの値を適当な非線形変換により重み値に変換したものを、結果統合用パラメータに保持する。この例では、専用音声認識サーバの認識結果が「鈴木一郎」、汎用サーバ１の結果が「佐々木一郎」、汎用サーバ２と３の結果が「鈴木一郎」だった場合、「佐々木一郎」の重みが３．０、「鈴木一郎」の重みが１．４と１．２の和で２．６となり、前者の方が大きいことから、「佐々木一郎」を最終認識結果とする。

[0029] 図７は、図３，４，６とは異なるもう一つの結果統合用パラメータの実現例である。ここでは、ユーザ辞書データに含まれる語句を汎用音声認識サー

バで認識して、正しく認識されなかった場合においても、そのときに得られた認識結果を結果統合用パラメータとして保存しておく。それぞれのサーバの重みを設定するのは図６の例と同様である。複数回の実験を行った際には、最も多かった結果のみか、もしくは複数の認識結果を保存しておいても良い。また、実験の回数にかかわらず、２位以下の認識結果も併せて保存しておいても良い。認識実行時には、これまでの例と同じように、専用音声認識サーバの認識結果に基づき結果統合用パラメータを参照する。その際、汎用音声認識サーバの認識結果が、結果統合用パラメータに保存されているものと一致するかどうかをチェックする。例えば、専用音声認識サーバの認識結果が「日立太郎」で、汎用サーバ１の結果が「日立市」、汎用サーバ２の結果が「二十歳」、汎用サーバ３の結果が「日立」だった場合、汎用サーバ１の結果は「日立太郎」に変換した上で、各認識結果での多数決を行い、最終的に「日立太郎」が選択される。

[0030] 図８は、同音異表記の検出を利用した、音声認識結果統合方式の例を示す図である。図に示すように、専用音声認識サーバの認識結果が「左藤一郎」である場合、これを汎用音声認識サーバの各認識結果と比較し、同音異表記が含まれないかをチェックする。ここで、表記から発音を推定するには、日本語であれば、個々の漢字の読みをデータとして保持しておき、当該語句を構成する漢字の読みを連結することにより発音表記を得る。英語であれば、部分的な綴りに対する読み付与のルールを保持しておき、これらを順次適用することにより発音表記を得る。その他の言語の場合であっても、一般にＧｒａｐｈｅｍｅ　ｔｏ　Ｐｈｏｎｅｍｅと呼ばれる技術によって、発音表記を得ることができることは良く知られている。また、ユーザ辞書データの中に、漢字表記とカナ表記のように、発音情報が含まれている場合もあり、そのような場合にはそれを活用する。上述のチェックにより、同音異表記が含まれている場合には、当該認識結果の表記を専用音声認識サーバによる認識結果の表記に変換して用いる。図の例では、汎用音声認識サーバ１の認識結果「佐藤一郎」が、専用音声認識サーバの認識結果と同音であることから、

これを「左藤一郎」に変換する。その結果、３台の汎用音声認識サーバによる多数決の結果は「左藤一郎」となり、これが最終結果として採用される。

[0031] 図９は、自動車内でのナビゲーション機能やハンズフリー通話機能などを提供する場合を例に、ユーザ端末の具体的な実現形態の例を示した図である。（ａ）では、マイク装置９０４、アプリケーション９０６、通信モジュール９０８などのすべての機能をカーナビゲーション装置９０２内に実装している。（ｂ）では、カーナビゲーション装置９０２とスマートフォン９１０とを連結し、マイク装置９０４はカーナビゲーション装置９０２のものを、通信部９０８はスマートフォン９１０のものを用いている。アプリケーション９１２，９１４は、それぞれの機能に応じて、カーナビゲーション装置内とスマートフォン内に分散して配置するか、もしくはどちらか片方のみに配置する。（ｃ）では、スマートフォン９１０内に、すべての機能を実装する。

[0032] 図１０は、本発明を構成するユーザ辞書１２４の作成方法の例を示した図である。例えば、ユーザ端末１０２内にアドレス帳１００２が存在する場合には、そこに含まれる人名をユーザ辞書に登録する。同様に、音楽プレーヤーの楽曲リスト１００４が存在する場合には、そこに含まれる楽曲名やアーティスト名をユーザ辞書に登録する。また、ウェブブラウザのブックマーク１００６として登録されたページタイトルをユーザ辞書に登録することもできる。その他に、ユーザ端末内に蓄積されたメール１００８やショートメッセージなどのデータを解析し、そこに頻出する語句をユーザ辞書に登録するという方式も可能である。これらのデータに関しては、ユーザ端末がはじめて本発明によるシステムに接続された際には、ユーザ端末に含まれる全ユーザ辞書データをシステムに送信するのに加えて、アドレス帳や楽曲リストなどへの新規エントリの追加時には、新規追加データのみをシステムに追加送信し、結果統合用パラメータの更新を促すという方式を採ることもできる。このとき、結果統合用パラメータだけでなく、専用音声認識部の照合用辞書も同時に更新する必要がある。

[0033] 図１１は、一般的な音声合成部の構成を変更し、本発明に特化した特殊な構成の一例を示す図である。一般に音声合成部１１４は、合成音声作成部１１０２と、音声素片データ１１０６～１１１０とから成る。ここで、素片データとは、データを直接つなぎあわせて合成音声を作る方式で用いるためのデータの名称であるが、直接つなぎあわせる代わりに、統計処理と信号処理により波形を合成する方式を用いる場合でも、個々の音素や音節などの処理単位に対し、類似のデータ集合を用いるため、以下に述べる方式を適用することは可能である。合成音声作成部１１０２では、音声素片データを繋ぎ合わせ、必要であれば適切な信号処理を行い、標準合成音声を作成する。しかし、本発明においては、ユーザ端末の所有者である特定ユーザの声に対して各汎用音声認識サーバ群がどのように反応するかを知ることが重要であるので、音声合成部で作成される合成音声も、ユーザの声に似たものであることが望ましい。そこで、ユーザが音声認識機能を使用するたびに、あるいはそれ以外の音声機能や音声通話を使用するたびに、その声をユーザ音声データ１１１２として蓄積しておき、これを活用して音声変換部１１０４により標準合成音声からユーザ適応音声への変換を行う。こうして変換した音声を汎用音声認識サーバ群への入力とすることにより、より精度の高い性能予測を行うことが可能となり、結果統合用パラメータの値もより適切なものになることが期待できる。

[0034] 図１２は、音声認識の正しさに加えて、応答速度を評価基準とする場合の結果統合用パラメータの例を示す図である。この例では、ユーザ辞書データに含まれる各語句に対応する合成音声を用いた認識を実行し、その処理にかかった平均時間をパラメータとして保持しておく。この例でいうと、専用音声認識サーバの認識結果が「鈴木一郎」であった場合、汎用サーバ２の認識結果は０．５秒で得られると期待されるが、汎用サーバ１の認識結果を得るには１．５秒も待たなければならない。この応答時間がアプリケーションで想定される上限値を上回る場合、汎用サーバ２の結果が得られた時点で結果統合処理を行う。これにより、結果統合処理にほとんど時間がかからないと

仮定すると、約０．５秒の応答時間で最終認識結果を得ることができることになり、ユーザの利便性を向上させることができる。

実施例 2

[0035] 図１３は、ユーザ端末内に組み込まれた専用音声認識部１０８を用いて、図１に示した例と同等の機能を実現するような例の構成を示した図である。ここでは、ユーザ端末１０２が、中継サーバ１０４を介することなく、内部に存在する専用音声認識部１０８を用いてユーザ辞書１２４に含まれる語句の認識を行う。ユーザ辞書データを用いて、汎用音声認識サーバ群１０６の性能評価を事前に行う方法は、図１の場合に示したものと同様である。認識実行時には、中継サーバ１０４を介して汎用音声認識サーバ１０６による認識を実行すると同時に、ユーザ端末内の専用音声認識部１０８でも認識を実行する。このような、端末内の音声認識部と、通信装置を介して接続された音声認識部とを併用する方式は、特許文献３にも示されているが、特許文献３記載の発明が、通信経路が確立されているかどうかという点に着目して結果の取捨選択を行うのに対し、本発明では、事前に行った音声認識の結果に基づき求めた結果統合用パラメータを用いるという点が異なっている。

実施例 3

[0036] 図１４は、本発明に基づく音声認識サーバ統合装置のもう一つの構成例を示す図である。ここでは、汎用音声認識サーバ群１０６の機能として、そこで用いられている認識用語句リストが入手可能である場合を想定する。そのような条件のもとで、ユーザ端末１０２から中継サーバ１０４に送られたユーザ辞書データは、語句比較・類似度推定部１２６に送られる。当該部では、汎用音声認識サーバ群１０６から入手した認識用語句リストとユーザ辞書データとを比較し、ユーザ辞書１２４に含まれる各語句が、各々のサーバで正しく認識されうるかどうかを判定する。判定結果は統合方式学習部１１６に送られ、パラメータとして整理されたものが結果統合用パラメータ１１８に保持される。一方、ユーザ辞書データがそのまま専用音声認識サーバ１０８に送られ、専用音声認識サーバがチューニングされるのは、図１に示した

例と同じである。

[0037] このような準備が済んだ状態で、ユーザ端末１０２から入力音声データが送られてくると、図１に示した例と同様に、信号処理部１２０を経由して、汎用音声認識サーバ１０６および専用音声認識サーバ１０８に当該データが送られる。それらのサーバから返された認識結果は、認識結果統合部１２２に送られ、そこで、結果統合用パラメータ１１８との比較により、最適な認識結果が選択される。選択された認識結果がユーザ端末１０２に送信されて後は、図１に示した例と同様である。

[0038] 図１５は、図１４に示した構成において、ユーザ辞書データを使って結果統合用パラメータを作成するまでの処理の手順を示す図である。この例では、合成音声を作成することも、それを使って音声認識を実行してみることもなく、単に各汎用音声認識サーバから認識用語句リストを取得する。これらのリストと、ユーザ辞書データに含まれる語句とを比較し、ユーザ辞書データの各語句が、どの汎用音声認識サーバの語句リストに含まれているかをデータ化する。ここでは、認識用語句リストに含まれている（○）か、含まれていない（×）かのどちらかしか有り得ないことから、得られた結果をまとめた結果統合用パラメータは、図３ないし図４と同じものになる。従って、実際の認識を行う際の使い方も、前述した例と同じになる。また、各汎用音声認識サーバから、語句リストのみならず、それらの語句の認識されやすさを表す言語モデルを入手することが可能な際には、図６のような重み付きの結果統合用パラメータを作成することもできる。たとえば、代表的な言語モデルであるＮグラム言語モデルを用いる場合、ユニグラムの値をその単語の認識されやすさとする、もしくは倍グラムやトライグラムの最大値をその単語の認識されやすさとするなどの方式が考えられる。

実施例 4

[0039] 図１６は、ユーザとの間の入出力機能と音声認識サーバ統合機能とを単一の装置の中に組み込んだ装置により、図１に示した例と同等の機能を実現するような例の構成を示した図である。ここでは、音声認識サーバ統合装置１

０４の内部に蓄積されているユーザ辞書１２４に含まれるユーザ辞書データが、装置内の音声合成部１１４および認識サーバ通信部１１２に転送される。ユーザが話した声は、マイク装置１２８により取り込まれ、信号処理部１２０に転送される。これらを用いた処理の進め方は、図１の例において説明したものと同等であり、結果として認識結果統合部１２２にて認識結果が確定させられる。この認識結果は、装置内の表示部１３２に転送され、ユーザに提示される。

実施例 5

[0040] 図１７は、図１６に示した例をもとに、さらに専用音声認識サーバが担っている機能を音声認識サーバ統合装置に組み込んだ場合の構成を示した図である。音声認識サーバ統合装置１０４に含まれるマイク装置１２８から入力音声が取り込まれ、ユーザ辞書１２４からユーザ辞書データが転送される部分は図１６の例と同様であるが、それらに加えて、専用音声認識部１０８が装置内に組み込まれており、ユーザ辞書の内容を直接読み出した上で、マイク装置から送られてくる音声データを認識する。そこで得られた単体認識結果は、認識結果統合部１２２に送られ、汎用音声認識サーバ群１０６によって得られた認識結果と統合される。統合された認識結果は、装置内に存在するアプリケーション１３０に送られ、そこで各々のアプリケーションの目的に沿って活用される。

産業上の利用可能性

[0041] 本発明は、車載端末と音声認識サーバとの間に介在して、高精度の音声認識機能を提供するための音声データ中継装置として利用可能である。

符号の説明

[0042] １０２ ユーザ端末

１０４ 中継サーバ

１０６ 汎用音声認識サーバ群

１０８ 専用音声認識サーバ

１１０ 端末装置通信部

１１２　認識サーバ通信部

１１４　音声合成部

１１６　統合方式学習部

１１８　結果統合用パラメータ

１２０　信号処理部

１２２　認識結果統合部

１２４　ユーザ辞書

１２６　語句比較・類似度推定部

１２８　マイク装置

１３０　アプリケーション

１３２　表示部

請求の範囲

［請求項1］ ユーザが音声を用いて操作を行うための端末装置と、音声データを認識してその結果を返す音声認識サーバとの間を中継する装置であって、

ユーザが登録した語句もしくはユーザがよく使う語句のリストに基づいて認識結果統合用のパラメータを学習し保存する統合方式学習部と、

ユーザが音声認識を意図して発した音声のデータを前記端末装置から受信する手段と、

前記受信した音声データを汎用音声認識サーバおよび専用音声認識サーバに送信する手段と、

前記音声データの前記汎用音声認識サーバおよび専用音声認識サーバによる認識結果を受信する手段と、

前記汎用音声認識サーバおよび専用音声認識サーバによる認識結果を、前記保存された認識結果統合用のパラメータと比較し、最適な認識結果を選択する認識結果統合部と、

前記選択された認識結果を前記端末装置に送信する手段と

を備える音声認識サーバ統合装置。

［請求項2］ 請求項１記載の音声認識サーバ統合装置において、更に、

前記端末装置からユーザが登録した語句もしくはユーザがよく使う語句のリストを受信する手段と、

前記受信した語句をもとに合成音声を生成する音声合成部と、

前記生成された合成音声を前記汎用音声認識サーバおよび専用音声認識サーバに送信する手段と、

前記合成音声の前記汎用音声認識サーバおよび専用音声認識サーバによる認識結果を受信する手段を備え、

前記統合方式学習部は、前記合成音声の基となった語句と前記認識結果とを合わせて解析し、認識結果統合用パラメータを学習し保存す

ることを特徴とする音声認識サーバ統合装置。

[請求項3] 請求項１記載の音声認識サーバ統合装置において、更に、

前記端末装置からユーザが登録した語句もしくはユーザがよく使う語句のリストを受信する手段と、

前記汎用音声認識サーバから認識用の語句リストを受信する手段と、

前記認識用の語句リストを前記端末装置から受信した語句リストと比較し、類似度を推定する語句比較・類似度推定部とを備え、

前記統合方式学習部は、前記推定結果を認識結果統合用のパラメータとして保存することを特徴とする音声認識サーバ統合装置。

[請求項4] ユーザが音声を用いて操作を行うための装置であって、

ユーザが登録した語句もしくはユーザがよく使う語句のリストに基づいて認識結果統合用のパラメータを学習し保存する統合方式学習部と、

ユーザが音声認識を意図して発した音声のデータを汎用音声認識サーバおよび専用音声認識サーバに送信する手段と、

前記音声データの前記汎用音声認識サーバおよび専用音声認識サーバによる認識結果を受信する手段と、

前記汎用音声認識サーバおよび専用音声認識サーバの認識結果を、前記保存された認識結果統合用パラメータと比較し、最適な認識結果を選択する認識結果統合部と、

前記選択された認識結果を表示する表示部を備える音声認識サーバ統合装置。

[請求項5] 請求項４記載の音声認識サーバ統合装置において、更に、

ユーザが登録した語句もしくはユーザがよく使う語句を記憶するユーザ辞書と、

前記ユーザ辞書に記憶した語句をもとに合成音声を生成する音声合成部と、

前記生成された合成音声を前記汎用音声認識サーバおよび専用音声認

識サーバに送信する手段と、

前記合成音声の前記汎用音声認識サーバおよび専用音声認識サーバによる認識結果を受信する手段とを備え、

前記統合方式学習部は、前記合成音声の基となった語句と前記認識結果とを合わせて解析し、認識結果統合用パラメータを学習し保存することを特徴とする音声認識サーバ統合装置。

[請求項6] 請求項４記載の音声認識サーバ統合装置において、更に、

ユーザが登録した語句もしくはユーザがよく使う語句のリストを記憶するユーザ辞書と、

前記汎用音声認識サーバから認識用の語句リストを受信する手段と、

前記認識用の語句リストを前記ユーザ辞書の語句リストと比較し、類似度を推定する語句比較・類似推定部とを備え、

前記統合方式学習部は、前記推定結果を認識結果統合用のパラメータとして保存することを特徴とする音声認識サーバ統合装置。

[請求項7] 請求項１～６の何れか１つに記載の音声認識サーバ統合装置において、

前記専用音声認識サーバは、ユーザが登録した語句もしくはユーザがよく使う語句のリストをもとに認識対象語句リストを作成し、このリストに含まれる語句を高い精度で認識することが可能なことを特徴とする音声認識サーバ統合装置。

[請求項8] 請求項１～６の何れか１つに記載の音声認識サーバ統合装置において、

前記専用音声認識サーバは、前記音声認識サーバ統合装置或いは端末装置内に音声専用認識部として組み込まれていることを特徴とする音声認識サーバ統合装置。

[請求項9] 請求項２または請求項５に記載の音声認識サーバ統合装置において、

前記認識結果統合用パラメータは、ユーザが登録した語句もしくは

ユーザがよく使う語句に対する音声認識サーバの認識結果の正誤を蓄積するものであり、

前記認識結果統合部は、前記専用音声認識サーバによる認識結果をもとに、その単語に対する音声認識サーバの認識結果を前記認識結果統合用パラメータから取り出し、かつ、

前記取り出した結果が正であるような音声認識サーバによる音声認識結果のみを取り出し、前記取り出した認識結果に基づき最適な認識結果を選択することを特徴とする音声認識サーバ統合装置。

[請求項10] 請求項２または請求項５に記載の音声認識サーバ統合装置において、

前記認識結果統合用パラメータは、ユーザが登録した語句もしくはユーザがよく使う語句に対する音声認識サーバの認識結果の正誤、および、個々の語句に対する前記音声認識サーバの認識結果の信頼度を表す値を蓄積するものであり、

前記認識結果統合部は、前記専用音声認識サーバによる認識結果をもとに、前記認識結果統合用パラメータからその単語に対する前記音声認識サーバの認識結果およびその信頼度を取り出し、かつ、前記取り出した認識結果が正であるような音声認識サーバによる音声認識結果および信頼度のみを取り出し、前記取り出した音声認識結果を前記信頼度の重み付けをして統合することを特徴とする音声認識サーバ統合装置。

[請求項11] 請求項２または請求項５に記載の音声認識サーバ統合装置において、

前記認識結果統合用パラメータは、ユーザが登録した語句もしくはユーザがよく使う語句に対する音声認識サーバの認識にかかる時間を測定し測定値を蓄積するものであり、

前記認識結果統合部は、前記専用音声認識サーバによる認識結果をもとに、前記認識結果統合用パラメータからその単語に対する前記音

声認識サーバの認識所要時間を取り出し、アプリケーションに依存して決まる認識所要時間の許容上限値を取得し、音声認識サーバのうち認識所要時間が前記許容上限値を下回るもののみの認識結果を取り出すことにより、前記取り出した認識結果をもとに最適な認識結果を選択することを特徴とする音声認識サーバ統合装置。

[請求項12] 請求項２または請求項５に記載の音声認識サーバ統合装置において、

前記認識結果統合用パラメータは、ユーザが登録した語句もしくはユーザがよく使う語句に対する音声認識サーバの認識結果の正誤および１つもしくは複数の誤認識結果を蓄積するものであり、

前記認識結果統合部は、前記専用音声認識サーバによる認識結果をもとに、前記認識結果統合用パラメータからその単語に対する前記音声認識サーバの認識結果の正誤および誤認識結果を取り出し、前記取り出した認識結果が誤である場合には前記取り出した誤認識結果を実行時の認識結果と比較し、前記比較の結果が同一であると判定された場合のみ当該認識結果を有効とすることにより、前記有効とされた認識結果をもとに最適な認識結果を選択することを特徴とする音声認識サーバ統合装置。

[請求項13] ユーザが登録した語句もしくはユーザがよく使う語句のリストに基づいて認識結果統合用のパラメータを学習して保存するステップと、

ユーザが音声認識を意図して発した音声のデータを汎用音声認識サーバおよび専用音声認識サーバに送信するステップと、

前記音声データの前記汎用音声認識サーバおよび専用音声認識サーバによる認識結果を受信するステップと、

汎用音声認識サーバの認識結果および専用音声認識サーバの認識結果と、前記認識結果統合用パラメータとを比較して、最適な音声認識結果を選択するステップと、

から成る音声認識サーバ統合方法。

［請求項14］ 請求項１３記載の音声認識サーバ統合方法において、更に、

ユーザが登録した語句もしくはユーザがよく使う語句をもとに合成音声を生成するステップと、

前記生成された合成音声を前記汎用音声認識サーバおよび専用音声認識サーバに送信するステップと、

前記合成音声の前記汎用音声認識サーバおよび専用音声認識サーバによる認識結果を受信するステップとを備え、

認識結果統合用のパラメータを学習して保存するステップは、前記合成音声の基となった語句と前記認識結果とを合わせて解析し、認識結果統合用パラメータを学習し保存することを特徴とする音声認識サーバ統合方法。

［請求項15］ 請求項１３記載の音声認識サーバ統合方法において、更に、

ユーザが登録した語句もしくはユーザがよく使う語句のリストを得るステップと、

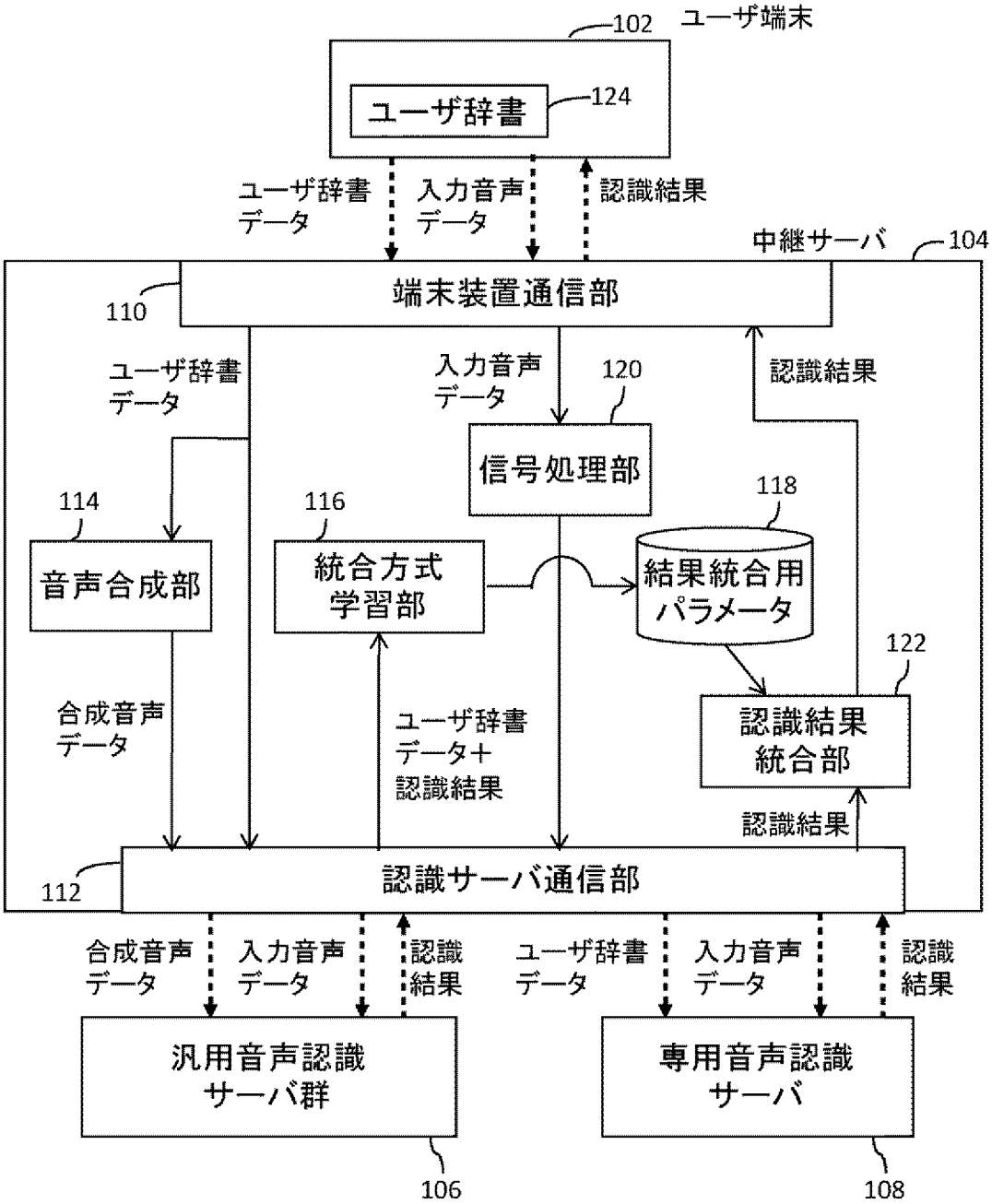
前記汎用音声認識サーバから認識用の語句リストを受信するステップと、

前記認識用の語句リストを、前記ユーザが登録した語句もしくはユーザがよく使う語句のリストと比較し、類似度を推定するステップとを備え、

前記認識結果統合用のパラメータを学習して保存するステップは、前記推定結果を認識結果統合用のパラメータとして保存することを特徴とする音声認識サーバ統合方法。

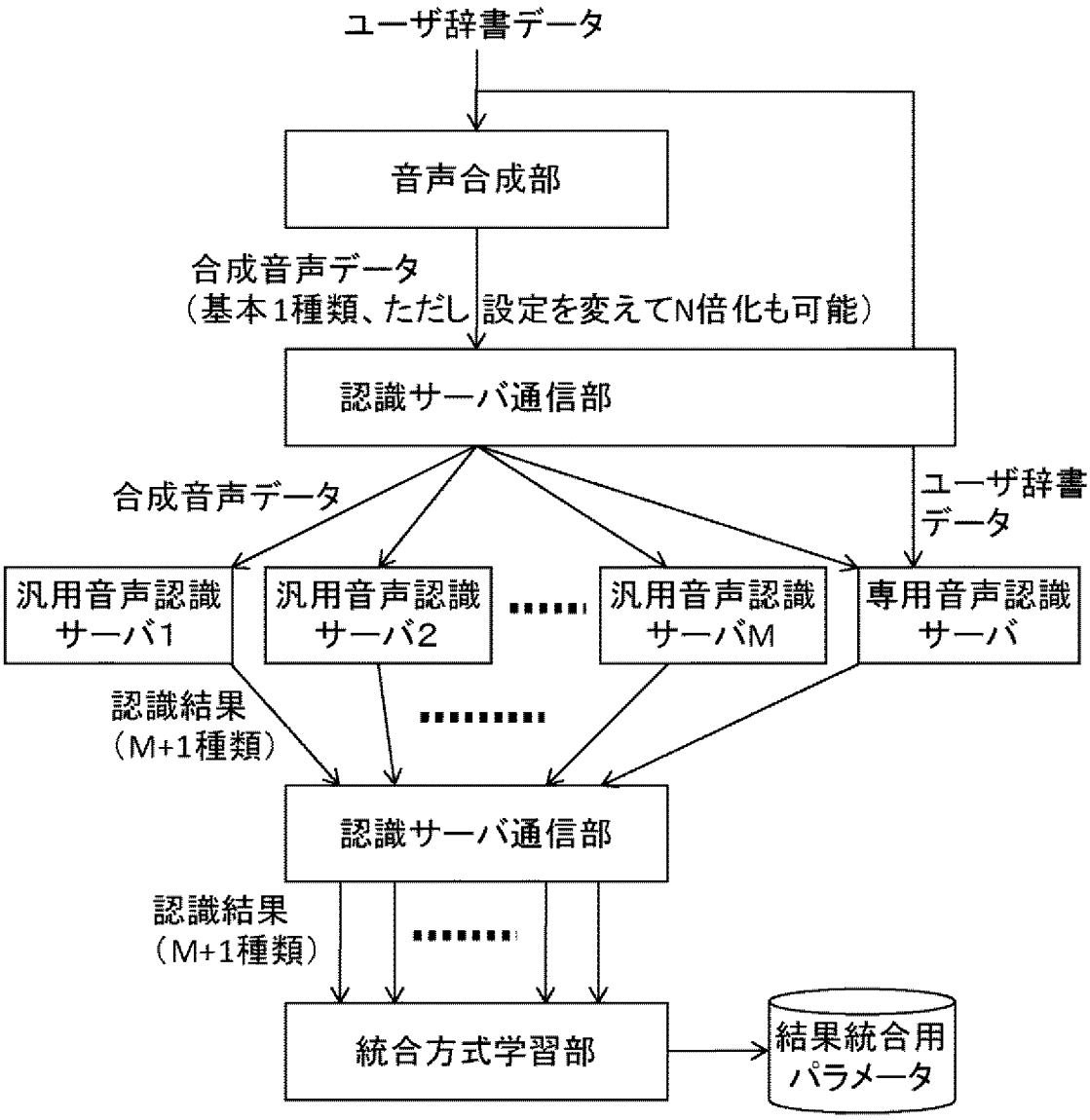
[図1]

図1


102
ユーザ端末
ユーザ辞書
124
ユーザ辞書データ
入力音声データ
認識結果
中継サーバ
104
端末装置通信部
110
ユーザ辞書データ
入力音声データ
120
認識結果
信号処理部
114
116
118
音声合成部
統合方式学習部
結果統合用パラメータ
122
合成音声データ
ユーザ辞書データ＋認識結果
認識結果統合部
認識結果
112
認識サーバ通信部
合成音声データ
入力音声データ
認識結果
ユーザ辞書データ
入力音声データ
認識結果
汎用音声認識サーバ群
専用音声認識サーバ
106
108

[図2]

図2


ユーザ辞書データ
音声合成部
合成音声データ
（基本1種類、ただし設定を変えてN倍化も可能）
認識サーバ通信部
合成音声データ
ユーザ辞書データ
汎用音声認識サーバ1
汎用音声認識サーバ2
汎用音声認識サーバM
専用音声認識サーバ
認識結果
（M+1種類）
認識サーバ通信部
認識結果
（M+1種類）
統合方式学習部
結果統合用パラメータ

[図3]

図3

語句	汎用サーバ
日立太郎	×
日立花子	×
鈴木一郎	○
山田二郎	○
佐藤三郎	×

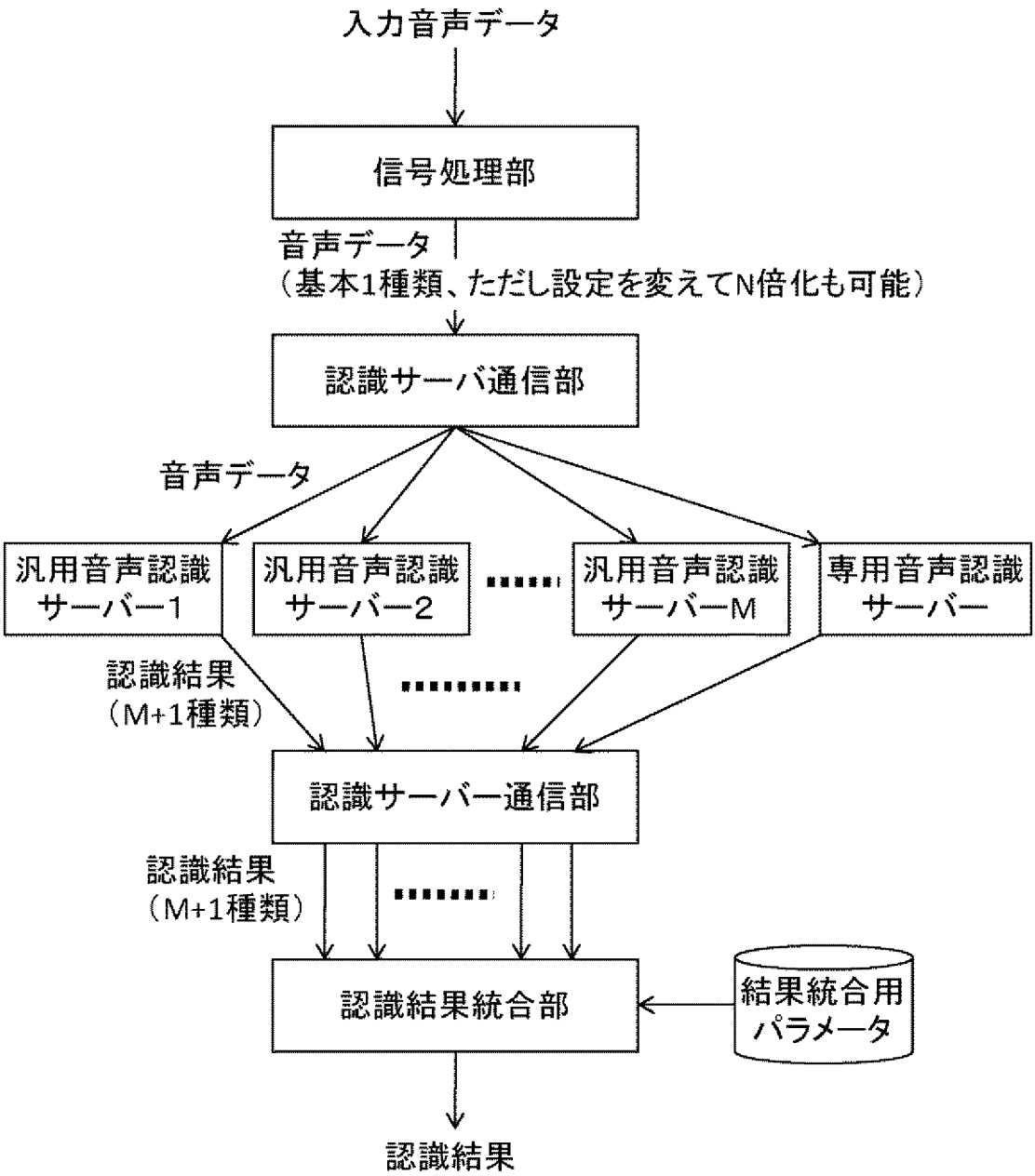
[図4]

図4

語句	汎用サーバ1	汎用サーバ2	汎用サーバ3
日立太郎	x	x	x
日立花子	x	x	x
鈴木一郎	○	○	○
山田二郎	○	x	x
佐藤三郎	x	x	○

[図5]

図5


入力音声データ
信号処理部
音声データ
（基本1種類、ただし設定を変えてN倍化も可能）
認識サーバ通信部
音声データ
汎用音声認識サーバー1
汎用音声認識サーバー2
汎用音声認識サーバーM
専用音声認識サーバー
認識結果
（M+1種類）
認識サーバー通信部
認識結果
（M+1種類）
認識結果統合部
結果統合用パラメータ
認識結果

[図6]

図6

語句	汎用サーバ1	汎用サーバ2	汎用サーバ3
日立太郎	×	×	×
日立花子	×	×	×
鈴木一郎	3.0	1.4	1.2
山田二郎	2.5	×	×
佐藤三郎	×	×	1.3

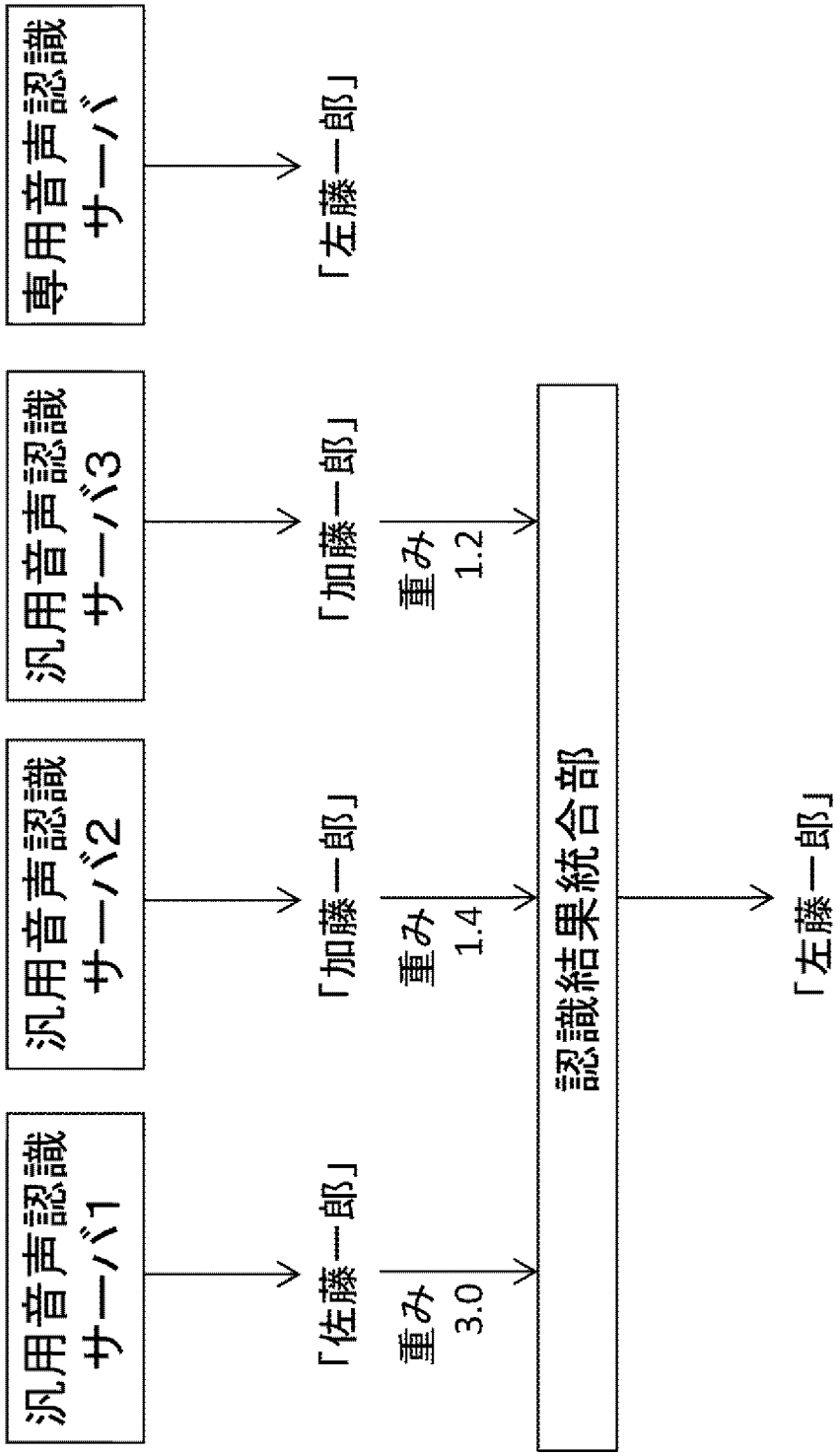
[図7]

図7

語句	汎用サーバ1	汎用サーバ2	汎用サーバ3
日立太郎	日立市3.0	日立1.2	(結果なし)1.1
日立花子	日立二校2.8	日立0.9	(結果なし)1.0
鈴木一郎	3.0	1.4	1.2
山田二郎	2.5	山田0.8	(結果なし)1.0
佐藤三郎	佐藤二郎2.2	佐藤0.9	1.3

[図8]

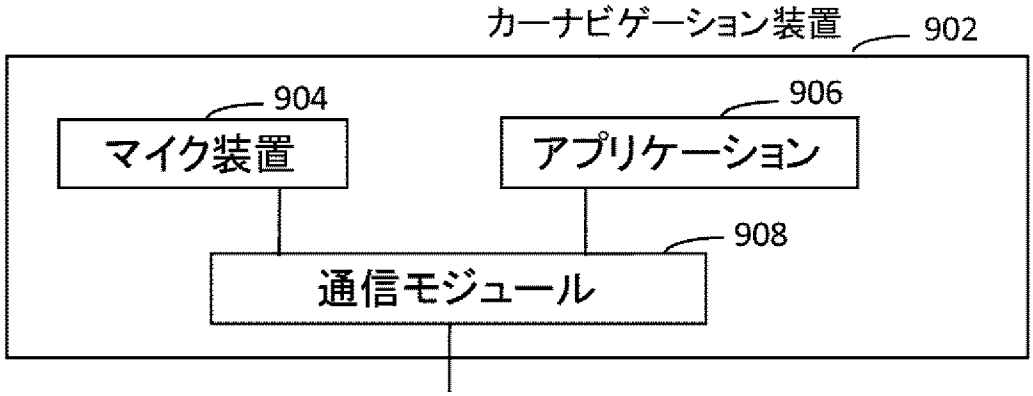
図8


専用音声認識サーバ
「左藤一郎」
汎用音声認識サーバ3
「加藤一郎」
重み
1.2
汎用音声認識サーバ2
「加藤一郎」
重み
1.4
汎用音声認識サーバ1
「佐藤一郎」
重み
3.0
認識結果統合部
「左藤一郎」

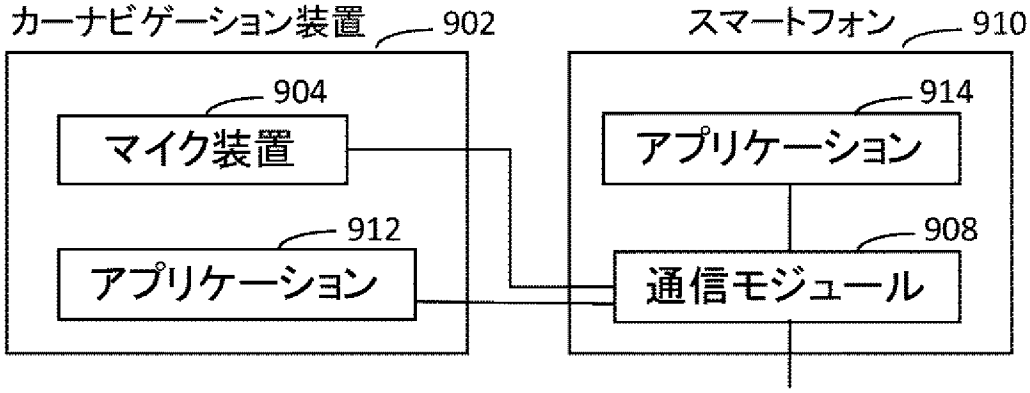
[図9]

図9

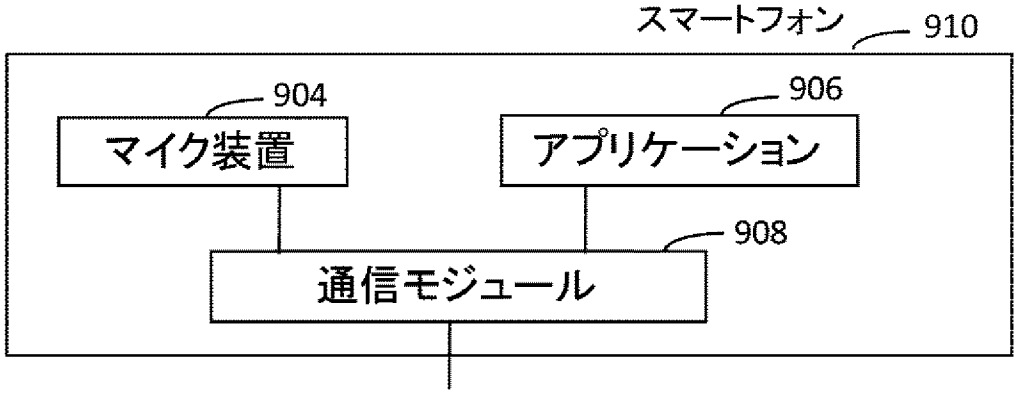
(a)


カーナビゲーション装置
902
904
マイク装置
906
アプリケーション
908
通信モジュール

(b)

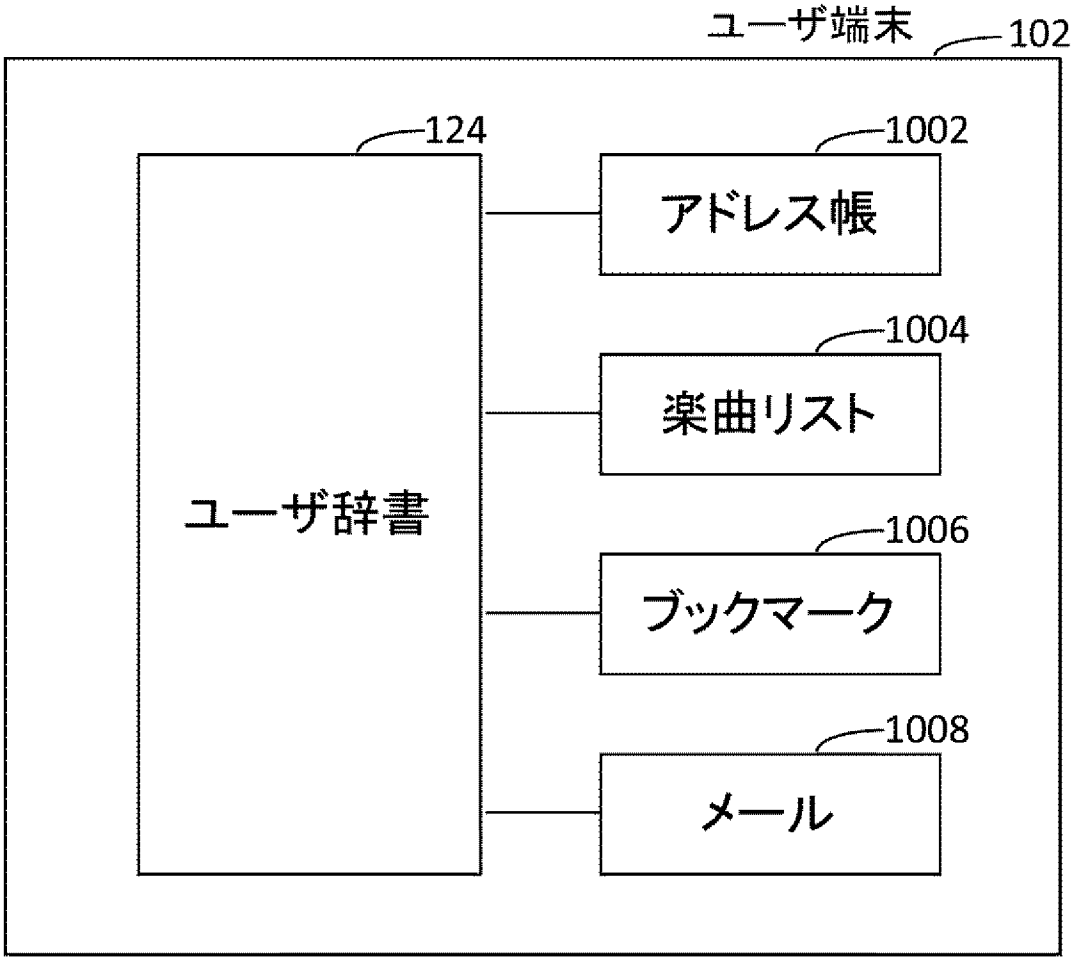

カーナビゲーション装置
902
スマートフォン
910
904
マイク装置
914
アプリケーション
912
アプリケーション
908
通信モジュール

(c)


スマートフォン
910
904
マイク装置
906
アプリケーション
908
通信モジュール

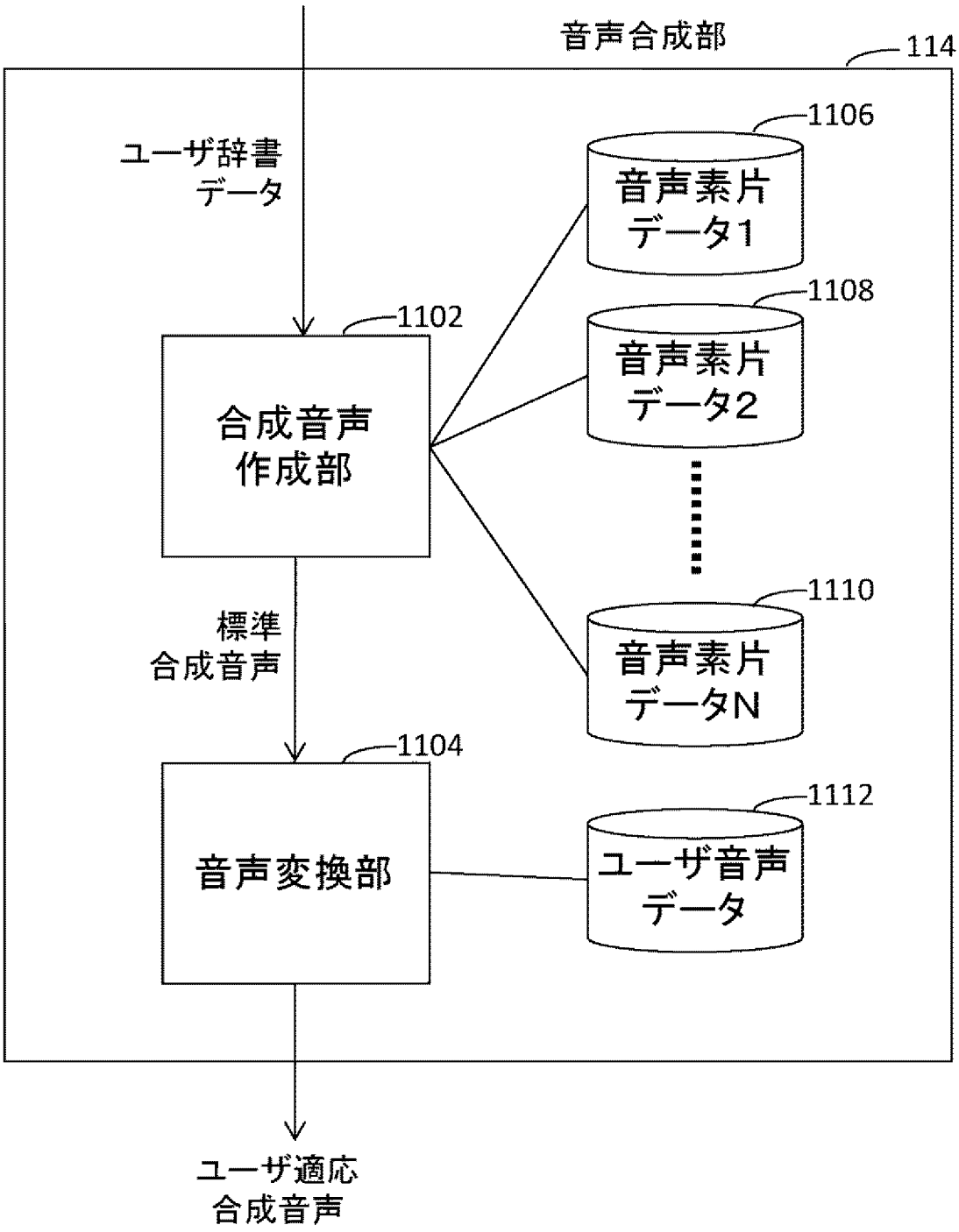
[図10]

図10


ユーザ端末
102
124
ユーザ辞書
1002
アドレス帳
1004
楽曲リスト
1006
ブックマーク
1008
メール

[図11]

図11


音声合成部
114
ユーザ辞書
データ
1106
音声素片
データ1
1108
音声素片
データ2
1102
合成音声
作成部
1110
音声素片
データN
標準
合成音声
1104
音声変換部
1112
ユーザ音声
データ
ユーザ適応
合成音声

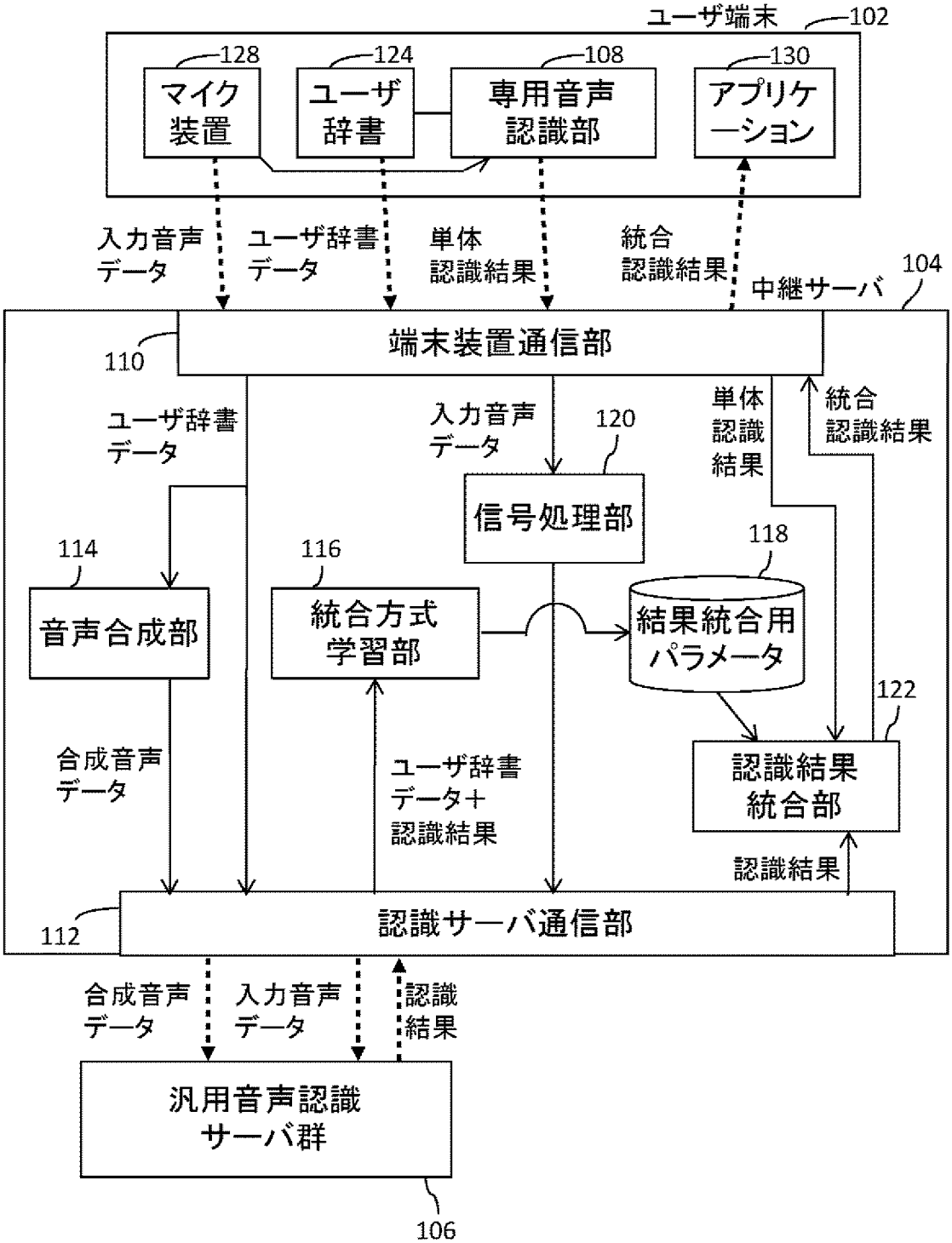
[図12]

図12

語句	汎用サーバ1	汎用サーバ2	専用サーバ
日立太郎	x	x	0.2秒
日立花子	x	x	0.1秒
鈴木一郎	1.5秒	0.5秒	0.1秒
山田二郎	1.8秒	x	0.2秒
佐藤三郎	x	x	0.2秒

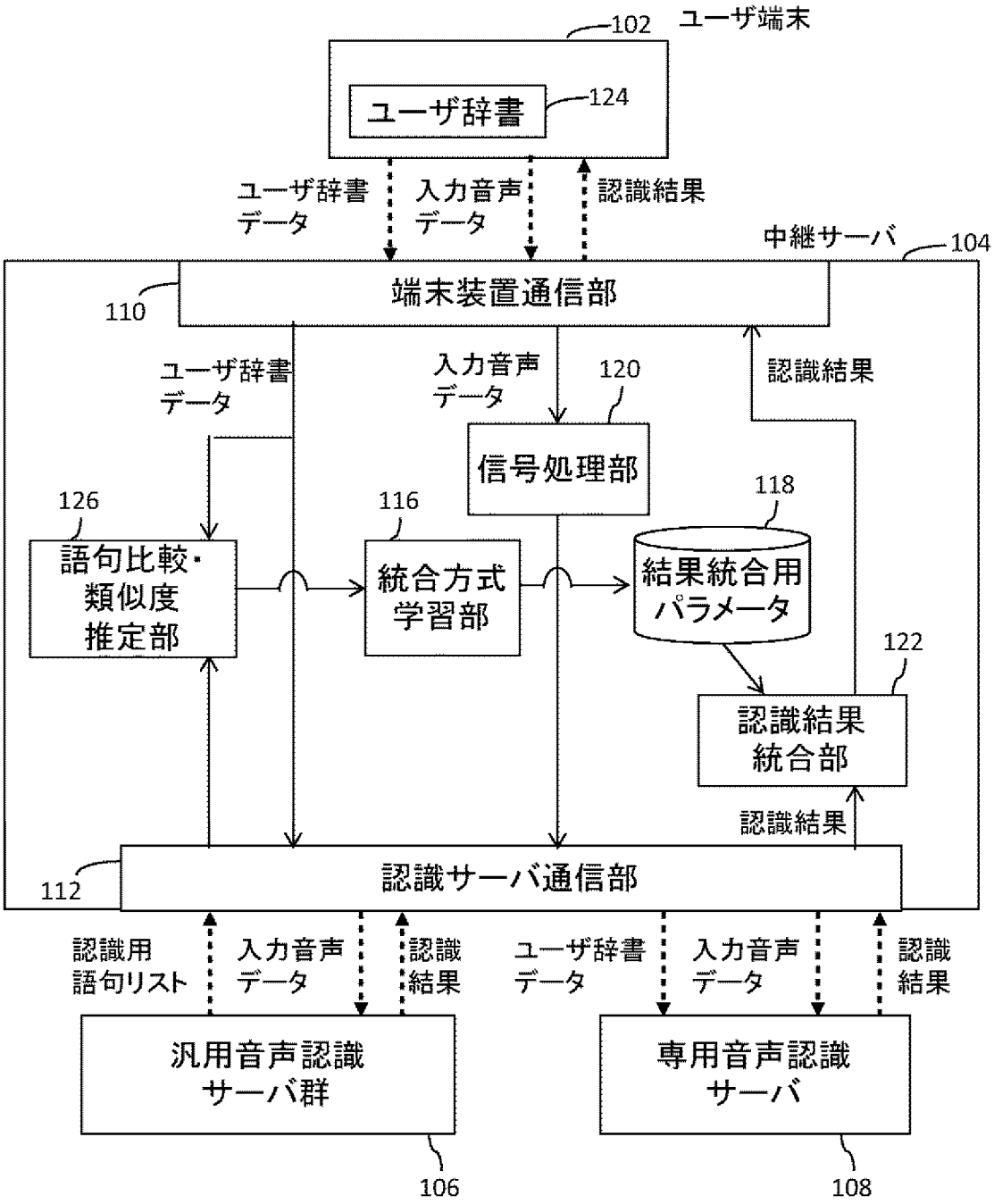
[図13]

図13


ユーザ端末
102
128
マイク装置
124
ユーザ辞書
108
専用音声認識部
130
アプリケーション
入力音声データ
ユーザ辞書データ
単体認識結果
統合認識結果
中継サーバ
104
端末装置通信部
110
ユーザ辞書データ
入力音声データ
120
単体認識結果
統合認識結果
信号処理部
114
音声合成部
116
統合方式学習部
118
結果統合用パラメータ
122
合成音声データ
ユーザ辞書データ＋認識結果
認識結果統合部
認識結果
112
認識サーバ通信部
合成音声データ
入力音声データ
認識結果
汎用音声認識サーバ群
106

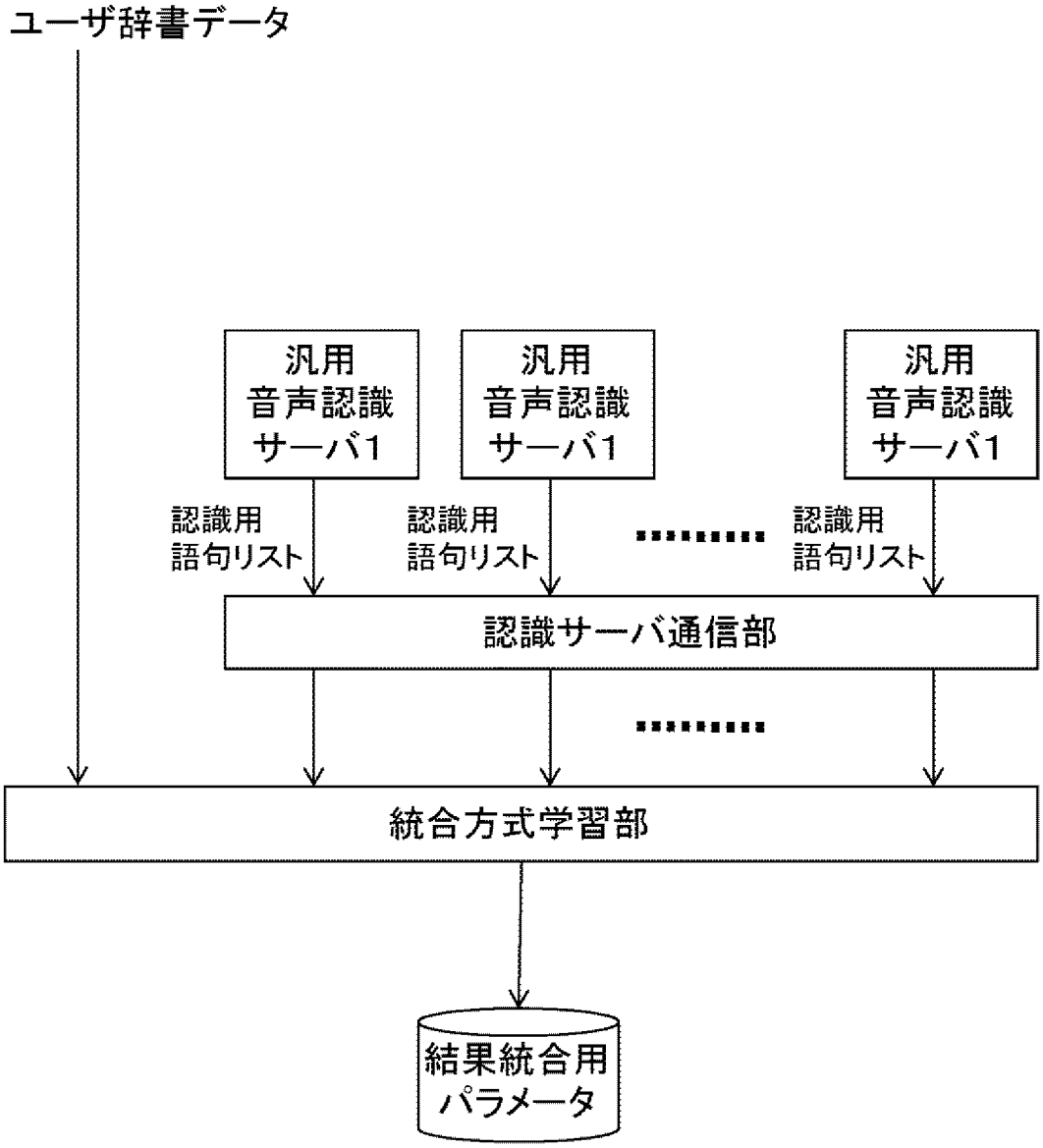
[図14]

図14


102
ユーザ端末
ユーザ辞書
124
ユーザ辞書データ
入力音声データ
認識結果
中継サーバ
104
端末装置通信部
110
ユーザ辞書データ
入力音声データ
120
認識結果
信号処理部
126
116
118
語句比較・類似度推定部
統合方式学習部
結果統合用パラメータ
122
認識結果統合部
認識結果
認識サーバ通信部
112
認識用語句リスト
入力音声データ
認識結果
ユーザ辞書データ
入力音声データ
認識結果
汎用音声認識サーバ群
専用音声認識サーバ
106
108

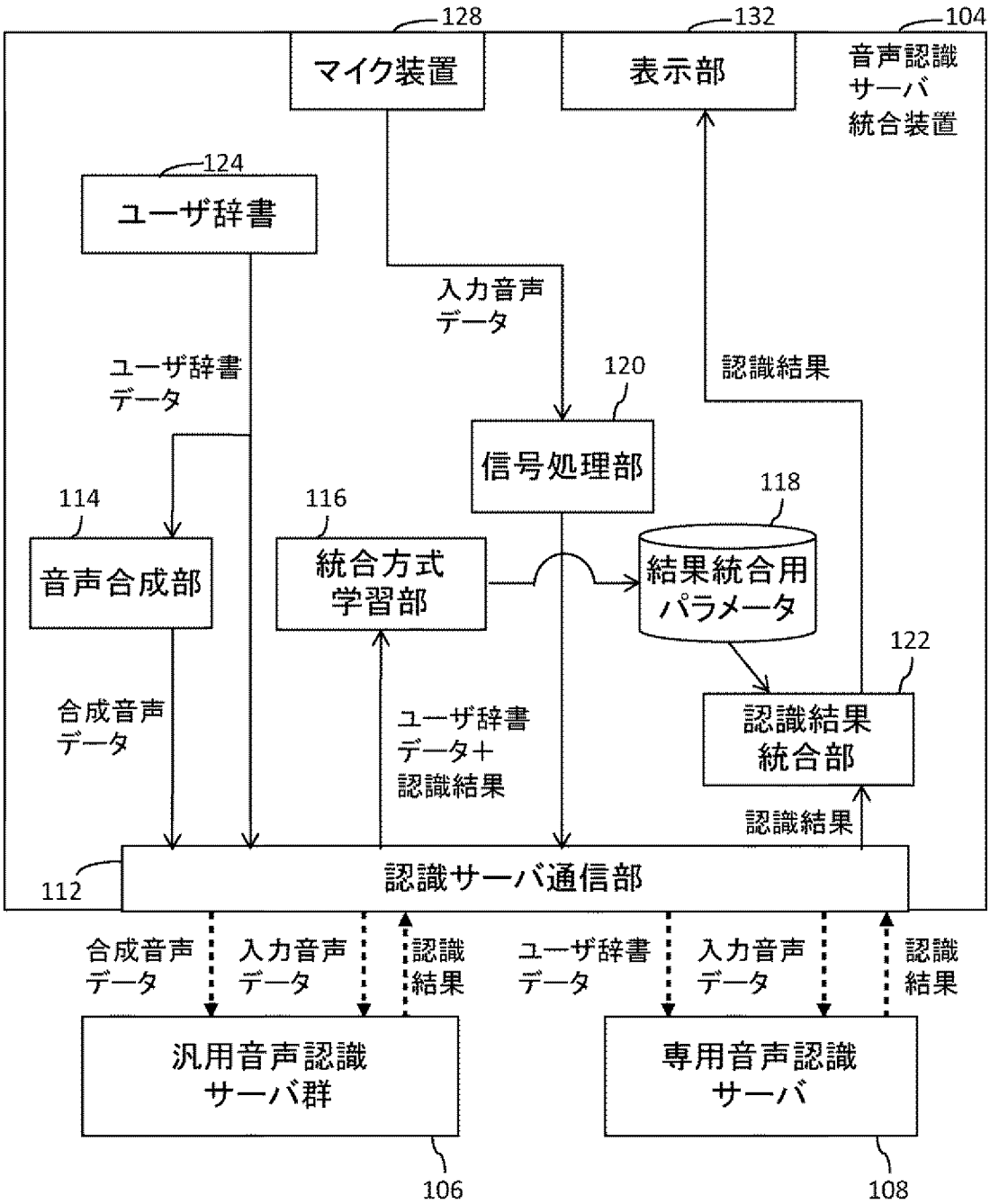
[図15]

図15


ユーザ辞書データ
汎用
音声認識
サーバ1
汎用
音声認識
サーバ1
汎用
音声認識
サーバ1
認識用
語句リスト
認識用
語句リスト
認識用
語句リスト
認識サーバ通信部
統合方式学習部
結果統合用
パラメータ

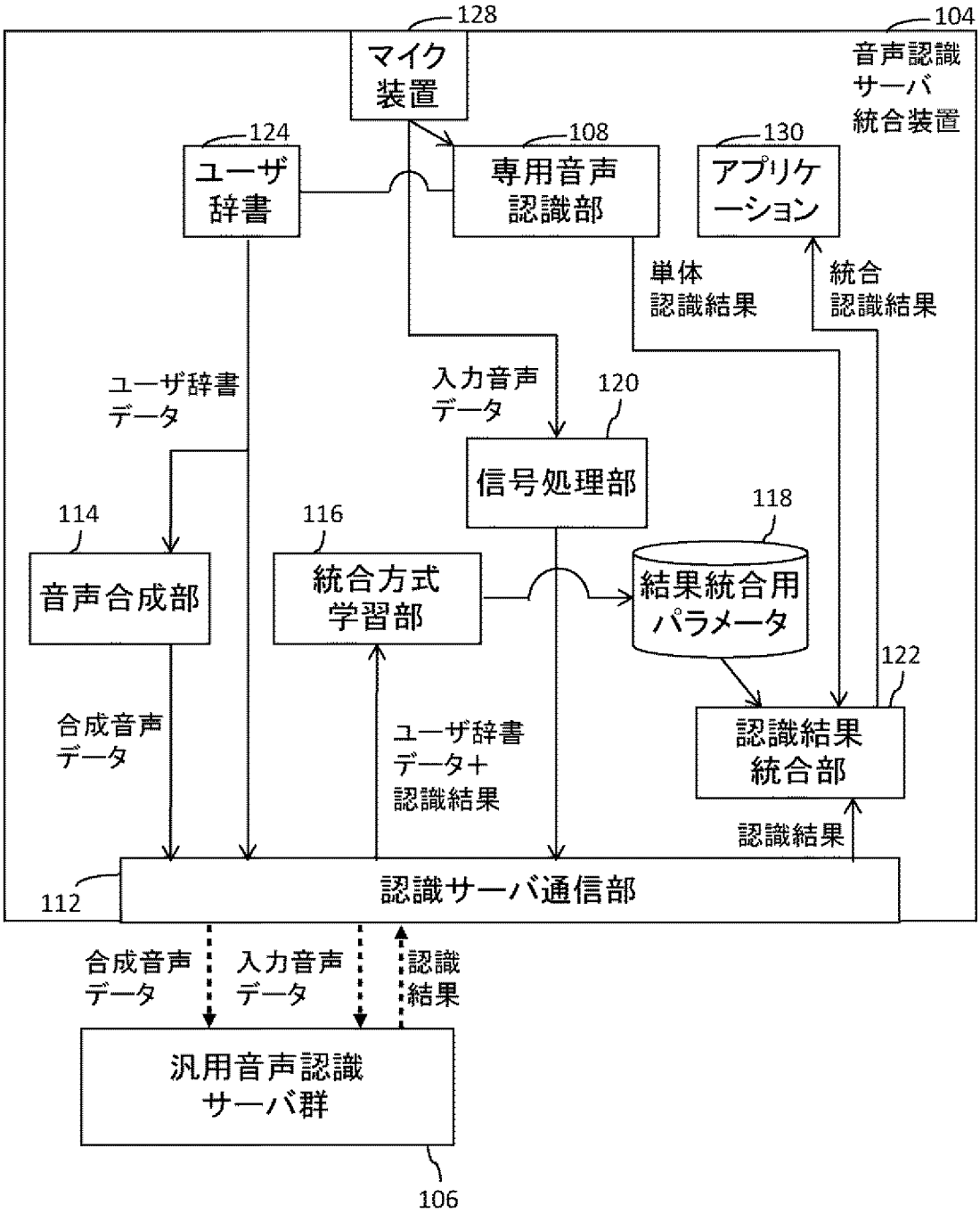
[図16]

図16


128
マイク装置
132
表示部
104
音声認識
サーバ
統合装置
124
ユーザ辞書
入力音声
データ
ユーザ辞書
データ
120
認識結果
信号処理部
118
114
116
音声合成部
統合方式
学習部
結果統合用
パラメータ
122
合成音声
データ
ユーザ辞書
データ＋
認識結果
認識結果
統合部
認識結果
112
認識サーバ通信部
合成音声
データ
入力音声
データ
認識
結果
ユーザ辞書
データ
入力音声
データ
認識
結果
汎用音声認識
サーバ群
専用音声認識
サーバ
106
108

[図17]

図17


128
マイク装置
104
音声認識サーバ統合装置
124
ユーザ辞書
108
専用音声認識部
130
アプリケーション
単体認識結果
統合認識結果
ユーザ辞書データ
入力音声データ
120
信号処理部
118
114
音声合成部
116
統合方式学習部
結果統合用パラメータ
122
合成音声データ
ユーザ辞書データ＋認識結果
認識結果統合部
認識結果
112
認識サーバ通信部
合成音声データ
入力音声データ
認識結果
汎用音声認識サーバ群
106

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2013/060238

A. CLASSIFICATION OF SUBJECT MATTER
G10L15/32(2013.01)i, *G10L15/06*(2013.01)i, *G10L15/22*(2006.01)i, *G10L15/30*(2013.01)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)
G10L15/32, G10L15/06, G10L15/22, G10L15/30

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 1996-2013
Kokai Jitsuyo Shinan Koho 1971-2013 Toroku Jitsuyo Shinan Koho 1994-2013

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.
A	WO 2011/121978 A1 (NEC Corp.), 06 October 2011 (06.10.2011), paragraph [0047] (Family: none)	1-15
A	JP 2010-85536 A (Fue Trek Co., Ltd. et al.), 15 April 2010 (15.04.2010), entire text; all drawings (Family: none)	1-15
A	JP 2006-243673 A (Canon Inc.), 14 September 2006 (14.09.2006), entire text; all drawings (Family: none)	1-15

☒ 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	Date of mailing of the international search report
04 July, 2013 (04.07.13)	16 July, 2013 (16.07.13)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2013/060238

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,A	JP 2013-7764 A (Clarion Co., Ltd.), 10 January 2013 (10.01.2013), entire text; all drawings & US 2012/0330651 A1 & EP 2538404 A1 & CN 102842304 A	1-15

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２０１３／０６０２３８

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

Int.Cl.　G10L15/32(2013.01)i, G10L15/06(2013.01)i, G10L15/22(2006.01)i, G10L15/30(2013.01)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（ＩＰＣ））

Int.Cl.　G10L15/32, G10L15/06, G10L15/22, G10L15/30

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	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２０１３年
日本国実用新案登録公報	１９９６－２０１３年
日本国登録実用新案公報	１９９４－２０１３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
A	WO 2011/121978 A1（日本電気株式会社）2011.10.06, 第0047段落（ファミリーなし）	1-15
A	JP 2010-85536 A（株式会社フュートレック他）2010.04.15, 全文、全図（ファミリーなし）	1-15
A	JP 2006-243673 A（キヤノン株式会社）2006.09.14, 全文、全図（ファミリーなし）	1-15

☑　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	国際調査報告の発送日
０４．０７．２０１３	１６．０７．２０１３

国際調査機関の名称及びあて先
日本国特許庁（ＩＳＡ／ＪＰ）
郵便番号１００－８９１５
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）　５Ｚ　３９９０
毛利　太郎

電話番号　０３－３５８１－１１０１　内線　３５９１

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
Ｐ，Ａ	JP 2013-7764 A (クラリオン株式会社) 2013.01.10, 全文、全図 & US 2012/0330651 A1 & EP 2538404 A1 & CN 102842304 A	1-15