(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2015-11170
(P2015-11170A)
(43)公開日　平成27年1月19日(2015.1.19)

(51)Int.Cl.		F I		テーマコード（参考）
G10L 15/30	(2013.01)	G10L 15/28	210A	
G10L 15/00	(2013.01)	G10L 15/00	200A	
G10L 15/10	(2006.01)	G10L 15/10	200W	

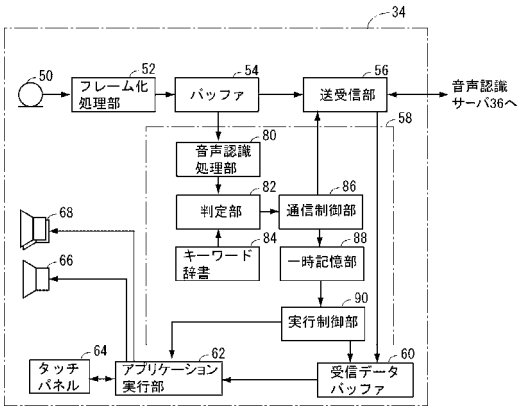
審査請求 未請求　請求項の数 6　OL　（全 16 頁）

(21)出願番号　特願2013-136306 (P2013-136306)
(22)出願日　平成25年6月28日 (2013.6.28)

(71)出願人　307041344
株式会社ＡＴＲ－Ｔｒｅｋ
神奈川県川崎市川崎区砂子２－４－１０
(74)代理人　100099933
弁理士　清水　敏
(72)発明者　古谷　利昭
大阪府大阪市淀川区西中島６丁目１番１号

(54)【発明の名称】ローカルな音声認識を行なう音声認識クライアント装置

(57)【要約】
【課題】ローカルにも音声認識機能を持ち、音声認識サーバの音声認識機能の起動を自然に行なえ、通信回線の負荷を抑えながら精度も高く維持できるクライアントを提供する。
【解決手段】音声認識クライアント装置３４は、音声認識サーバ３６との通信により、音声認識サーバ３６による音声認識結果を受信するクライアントであり、音声を音声データに変換するフレーム化処理部５２と、音声データに対する音声認識を行なうローカルな音声認識処理部８０と、音声データを音声認識サーバに送信し、当該音声認識サーバによる音声認識結果を受信する送受信部５６と、音声データに対する音声認識処理部８０の認識結果により、送受信部５６による音声データの送信を制御する判定部８２及び通信制御部８６とを含む。
【選択図】図２



【特許請求の範囲】
【請求項１】
音声認識サーバとの通信により、当該音声認識サーバによる音声認識結果を受信する音声認識クライアント装置であって、
音声を音声データに変換する音声変換手段と、
前記音声データに対する音声認識を行なう音声認識手段と、
前記音声データを前記音声認識サーバに送信し、当該音声認識サーバによる音声認識結果を受信する送受信手段と、
前記音声データに対する前記音声認識手段の認識結果により、前記送受信手段による音声データの送信を制御する送受信制御手段とを含む、音声認識クライアント装置。
【請求項２】
前記送受信制御手段は、
前記音声認識手段による音声認識結果中にキーワードが存在することを検出して、検出信号を出力するキーワード検出手段と、
前記検出信号に応答して、前記音声データのうち、前記キーワードの発話区間の先頭と所定の関係にある部分を前記音声認識サーバに送信するよう前記送受信手段を制御する送信開始制御手段とを含む、請求項１に記載の音声認識クライアント装置。
【請求項３】
前記送信開始制御手段は、前記検出信号に応答して、前記音声データのうち、前記キーワードの発話終了位置を先頭とする部分を前記音声認識サーバに送信するよう前記送受信手段を制御する手段を含む、請求項２に記載の音声認識クライアント装置。
【請求項４】
前記送信開始制御手段は、前記検出信号に応答して、前記音声データのうち、前記キーワードの発話開始位置を先頭とする部分を送信するよう前記送受信手段を制御する手段を含む、請求項２に記載の音声認識クライアント装置。
【請求項５】
前記送受信手段が受信した前記音声認識サーバによる音声認識結果の先頭部分が、前記キーワード検出手段が検出したキーワードと一致するか否かを判定する一致判定手段と、
前記一致判定手段による判定結果にしたがって、前記送受信手段が受信した前記音声認識サーバによる音声認識結果を利用する処理と、前記音声認識サーバによる音声認識結果を破棄する処理とを選択的に実行する手段とをさらに含む、請求項４に記載の音声認識クライアント装置。
【請求項６】
前記送受信制御手段は、
前記音声認識手段による音声認識結果中に第１のキーワードが存在することを検出して第１の検出信号を、何らかの処理を依頼することを表す第２のキーワードが存在することを検出して第２の検出信号を、それぞれ出力するキーワード検出手段と、
前記第１の検出信号に応答して、前記音声データのうち、前記第１のキーワードの発話区間の先頭と所定の関係にある部分を前記音声認識サーバに送信するよう前記送受信手段を制御する送信開始制御手段と、
前記送受信手段により前記音声データの送信が開始された後に前記第２の検出信号が発生されたことに応答して、前記音声データの前記第２のキーワードの発話の終了位置で前記送受信手段による音声データの送信を終了させる送信終了制御手段とを含む、請求項１に記載の音声認識クライアント装置。
【発明の詳細な説明】
【技術分野】
【０００１】
この発明は音声認識サーバと通信することにより音声を認識する機能を備えた音声認識クライアント装置に関し、特に、サーバとは別にローカルな音声認識機能を備えた音声認識クライアント装置に関する。

【背景技術】
【０００２】
ネットワークに接続される携帯電話等の携帯型端末装置の数が爆発的に増加している。携帯型端末装置は、事実上、小型のコンピュータということができる。特に、いわゆるスマートフォン等では、インターネット上のサイトの検索、音楽・ビデオの視聴、メールの交換、銀行取引、スケッチ、録音・録画等、デスクトップコンピュータと同等の充実した機能が利用できる。
【０００３】
しかしこのように充実した機能を利用するための１つのネックが、携帯型端末装置の筐体の小ささである。携帯型端末装置はその宿命として筐体が小さい。そのため、コンピュータのキーボードのように高速に入力をするためのデバイスを搭載することができない。タッチパネルを使用した様々な入力方式が考えられており、以前と比較して素早く入力できるようにはなっているが、依然として入力はそれほど容易でない。
【０００４】
こうした状況で入力のための手段として注目されているのが音声認識である。音声認識の現在の主流は、多数の音声データを統計的に処理して作成した音響モデルと、大量の文書から得た統計的言語モデルとを使用する統計的音声認識装置である。こうした音声認識装置は、非常に大きな計算パワーを必要とするため、大容量で計算能力が十分に高いコンピュータでのみ実現されていた。携帯型端末装置で音声認識機能を利用する場合には、音声認識サーバと呼ばれる、音声認識機能をオンラインで提供するサーバが利用され、携帯型端末装置はその結果を利用する音声認識クライアントとして動作する。音声認識クライアントが音声認識をする際には、音声をローカルに処理して得た音声データ、符号データ、又は音声の特徴量（素性）を音声認識サーバにオンラインで送信し、音声認識結果を受け取ってそれに基づいた処理を行なっている。これは、携帯型端末装置の計算能力が比較的低く、利用できる計算資源も限られていたためである。
【０００５】
しかし、半導体技術の進歩により、ＣＰＵ（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）の計算能力は非常に高くなり、また、メモリ容量も従来と比較して桁違いに大きくなってきた。しかも消費電力は少なくなっている。そのため、携帯型端末装置でも音声認識が十分に利用可能となっている。しかも、携帯型端末装置では使用するユーザが限定されるため、音声認識の話者を予め特定し、その話者に適合した音響モデルを準備したり、特定の語彙を辞書に登録したりすることで、音声認識の精度を高めることができる。
【０００６】
もっとも、利用できる計算資源の点では音声認識サーバの方が圧倒的に有利であるため、音声認識の精度の点では、携帯型端末装置よりも音声認識サーバで行なわれる音声認識の方が優れている点は間違いない。
【０００７】
このように、携帯型端末装置に搭載される音声認識の精度が比較的低い、という欠点を補うための提案が、後掲の特許文献１に開示されている。特許文献１は音声認識サーバと交信するクライアントに関する。このクライアントは、音声を処理して音声データに変換し、音声認識サーバに送信する。音声認識サーバからその音声認識結果を受信すると、その音声認識結果には、文節の区切り位置、文節の属性（文字種）、単語の品詞、文節の時間情報等が付されている。クライアントは、サーバからの音声認識結果に付されているこのような情報を利用して、ローカルに音声認識を行なう。この際、ローカルに登録されている語彙又は音響モデルを使用できるので、語彙によっては音声認識サーバで誤って認識された語を正しく認識できる可能性がある。
【０００８】
特許文献１に開示されたクライアントでは、音声認識サーバからの音声認識結果と、ローカルに行なった音声認識結果とを比較し、両者の認識結果が異なった箇所についてはユーザによりいずれかを選択させる。

【先行技術文献】
【特許文献】
【０００９】
【特許文献１】特開２０１０－８５５３６号公報、特に段落００４５～００５０、図４
【発明の概要】
【発明が解決しようとする課題】
【００１０】
特許文献１に開示されたクライアントは、音声認識サーバによる認識結果をローカルな音声認識結果で補完できるという優れた効果を奏する。しかし、現在の携帯型端末装置における音声認識の利用方法を見ていると、こうした機能を持つ携帯型端末の操作に関しては、いまだ改善の余地があると思われる。１つの問題点は、音声認識処理をどのようにして携帯型端末装置に開始させるか、という点である。
【００１１】
特許文献１には、ローカルでどのようにして音声認識を開始するかについての開示はない。現在利用可能な携帯型端末装置では、音声認識を開始するためのボタンを画面に表示させ、このボタンがタッチされたら音声認識機能を起動するものが主流である。又は、音声認識を開始させるための専用のハードウェアボタンを設けたものもある。ローカルな音声認識機能を持たない携帯電話で動作するアプリケーションの中には、ユーザが発話姿勢をとったとき、すなわち携帯電話を耳にあてたときをセンサで感知し、音声入力とサーバへの音声データの送信とを開始するものもある。
【００１２】
しかし、これらはいずれも音声認識機能を起動するにあたって特定の動作をユーザに要求するものである。これからの携帯型端末装置では、多様な機能を利用するために、音声認識機能を従来以上に活用することが予測され、そのためには音声認識機能の起動をより自然なものにする必要がある。一方で、携帯型端末装置と音声認識サーバとの間の通信量はできるだけ抑える必要があるし、音声認識の精度は高く維持する必要もある。
【００１３】
それゆえにこの発明の目的は、音声認識サーバを利用するとともに、ローカルにも音声認識機能を持つ音声認識クライアント装置であって、音声認識機能の起動を自然に行なえ、通信回線の負荷を抑えながら音声認識の精度も高く維持できる音声認識クライアント装置を提供することである。
【課題を解決するための手段】
【００１４】
本発明の第１の局面に係る音声認識クライアント装置は、音声認識サーバとの通信により、当該音声認識サーバによる音声認識結果を受信する音声認識クライアント装置である。この音声認識クライアント装置は、音声を音声データに変換する音声変換手段と、音声データに対する音声認識を行なう音声認識手段と、音声データを音声認識サーバに送信し、当該音声認識サーバによる音声認識結果を受信する送受信手段と、音声データに対する音声認識手段の認識結果により、送受信手段による音声データの送信を制御する送受信制御手段とを含む。
【００１５】
ローカルな音声認識手段の出力に基づいて、音声データを音声認識サーバに送信するか否かが制御される。音声認識サーバを利用するためには、発話することを除き特別な操作は必要ない。音声認識手段の認識結果が特定のものでなければ音声認識サーバへの音声データの送信が行なわれない。
【００１６】
その結果、本発明によれば、音声認識機能の起動を自然に行なえ、通信回線の負荷を抑えながら音声認識の精度も高く維持できる音声認識クライアント装置を提供できる。
【００１７】
好ましくは、送受信制御手段は、音声認識手段による音声認識結果中にキーワードが存

在することを検出して、検出信号を出力するキーワード検出手段と、検出信号に応答して、音声データのうち、キーワードの発話区間の先頭と所定の関係にある部分を音声認識サーバに送信するよう送受信手段を制御する送信開始制御手段とを含む。

【００１８】

ローカルな音声認識手段の音声認識結果中にキーワードが検出されると、音声データの送信が開始される。音声認識サーバの音声認識を利用するために、特別なキーワードを発話するだけでよく、ボタンを押す等、音声認識を開始するための明示的な操作をする必要がない。

【００１９】

より好ましくは、送信開始制御手段は、検出信号に応答して、音声データのうち、キーワードの発話終了位置を先頭とする部分を音声認識サーバに送信するよう送受信手段を制御する手段を含む。

【００２０】

キーワードの次の部分から音声認識サーバに音声データを送信することにより、キーワード部分の音声認識を音声認識サーバでは行なわずに済む。音声認識結果にキーワードが含まれないため、キーワードに続けて発話した内容に関する音声認識結果をそのまま利用できる。

【００２１】

さらに好ましくは、送信開始制御手段は、検出信号に応答して、音声データのうち、キーワードの発話開始位置を先頭とする部分を送信するよう送受信手段を制御する手段を含む。

【００２２】

キーワードの発話開始位置を先頭として音声認識サーバに送ることにより、音声認識サーバで再びキーワード部分の確認を行なったり、音声認識サーバの音声認識結果を利用して携帯型端末でローカルな音声認識の結果の正確さを検証したりできる。

【００２３】

音声認識クライアント装置は、送受信手段が受信した音声認識サーバによる音声認識結果の先頭部分が、キーワード検出手段が検出したキーワードと一致するか否かを判定する一致判定手段と、一致判定手段による判定結果にしたがって、送受信手段が受信した音声認識サーバによる音声認識結果を利用する処理と、音声認識サーバによる音声認識結果を破棄する処理とを選択的に実行する手段とをさらに含む。

【００２４】

ローカルな音声認識結果と、音声認識サーバによる音声認識結果とが異なる場合、より精度が高いと思われる音声認識サーバの結果を用いて発話者の発話を処理するか否かを判定する。ローカルな音声認識結果が誤っている場合には、音声認識サーバの音声結果は何ら利用されず、携帯型端末は何事もなかったように動作する。したがって、ローカルな音声認識による音声認識結果の誤りにより、ユーザの意図しないような処理を音声認識クライアント装置が実行することが予防できる。

【００２５】

好ましくは、送受信制御手段は、音声認識手段による音声認識結果中に第１のキーワードが存在することを検出して第１の検出信号を、何らかの処理を依頼することを表す第２のキーワードが存在することを検出して第２の検出信号を、それぞれ出力するキーワード検出手段と、第１の検出信号に応答して、音声データのうち、第１のキーワードの発話区間の先頭と所定の関係にある部分を音声認識サーバに送信するよう送受信手段を制御する送信開始制御手段と、送受信手段により音声データの送信が開始された後に第２の検出信号が発生されたことに応答して、音声データの第２のキーワードの発話の終了位置で送受信手段による音声データの送信を終了させる送信終了制御手段とを含む。

【００２６】

音声データを音声認識サーバに送信するにあたり、ローカルな音声認識手段による音声認識結果に第１のキーワードが検出されたときには、その第１のキーワードの発話開始位

置と所定の関係にある部分の音声データが音声認識サーバに送信される。その後、ローカルな音声認識手段による音声認識結果に、何らかの処理を依頼することを表す第２のキーワードが検出されたときには、それ以後の音声データの送信は行なわれない。音声認識サーバを利用するにあたり、第１のキーワードを発話するのみでよいだけでなく、第２のキーワードを発話することにより音声データの送信をその時点で終了できる。発話の終了を検知するために所定の無音区間を検出したりする必要はなく、音声認識のレスポンスを向上させることができる。

【図面の簡単な説明】

【００２７】

【図１】本発明の第１の実施の形態に係る音声認識システムの概略構成を示すブロック図である。

【図２】第１の実施の形態に係る携帯端末装置である携帯電話の機能的ブロック図である。

【図３】逐次方式の音声認識の出力の仕方の概略を説明する模式図である。

【図４】第１の実施の形態において、音声認識サーバへの音声データの送信開始及び送信終了タイミングと送信内容とを説明するための模式図である。

【図５】第１の実施の形態において、音声認識サーバへの音声データの送信開始及び終了を制御するプログラムの制御構造を示すフローチャートである。

【図６】第１の実施の形態において、音声認識サーバの結果とローカルな音声認識結果とを利用して携帯型端末装置を制御するプログラムの制御構造を示すフローチャートである。

【図７】本発明の第２の実施の形態に係る携帯型端末装置である携帯電話の機能的ブロック図である。

【図８】第２の実施の形態において、音声認識サーバへの音声データの送信開始及び送信終了タイミングと送信内容とを説明するための模式図である。

【図９】第２の実施の形態において、音声認識サーバへの音声データの送信開始及び終了を制御するプログラムの制御構造を示すフローチャートである。

【図１０】第１及び第２の実施の形態に係る装置の構成を示すハードウェアブロック図である。

【発明を実施するための形態】

【００２８】

以下の説明及び図面では、同一の部品には同一の参照番号を付してある。したがって、それらについての詳細な説明は繰返さない。

【００２９】

＜第１の実施の形態＞

［概略］

図１を参照して、第１の実施の形態に係る音声認識システム３０は、ローカルな音声認識機能を持つ音声認識クライアント装置である携帯電話３４と、音声認識サーバ３６とを含む。両者はインターネット３２を介して相互に通信可能である。この実施の形態では、携帯電話３４はローカルな音声認識の機能を持ち、音声認識サーバ３６との間の通信量を抑えながら、自然な形でユーザによる操作に対する応答を実現する。なお、以下の実施の形態では、携帯電話３４から音声認識サーバ３６に送信される音声データは音声信号をフレーム化したデータであるが、例えば音声信号を符号化した符号化データでもよいし、音声認識サーバ３６で行なわれる音声認識処理で使用される特徴量でもよい。

【００３０】

［構成］

図２を参照して、携帯電話３４は、マイクロフォン５０と、マイクロフォン５０から出力される音声信号をデジタル化し、所定フレーム長及び所定シフト長でフレーム化するフレーム化処理部５２と、フレーム化処理部５２の出力である音声データを一時的に蓄積するバッファ５４と、バッファ５４に蓄積された音声データを音声認識サーバ３６に送信す

る処理と、音声認識サーバ３６からの音声認識結果等を含むネットワークからのデータを無線により受信する送受信部５６とを含む。フレーム化処理部５２の出力する各フレームには、各フレームの時間情報が付されている。

【００３１】

携帯電話３４はさらに、バッファ５４に蓄積された音声データによるローカルな音声認識をバックグラウンドで行ない、音声認識結果の中に所定のキーワードが検出されたことに応答して、送受信部５６による音声認識サーバ３６への音声信号の送信開始及び送信終了を制御する処理と、音声認識サーバからの受信結果とローカルな音声認識の結果とを照合し、その結果にしたがって携帯電話３４の動作を制御するための制御部５８と、送受信部５６が音声認識サーバ３６から受信した音声認識結果を一時的に蓄積する受信データバッファ６０と、ローカルな音声認識結果と音声認識サーバ３６からの音声認識結果との照合に基づいて制御部５８が実行指示信号を発生したことに応答して、受信データバッファ６０の内容を用いたアプリケーションを実行するアプリケーション実行部６２と、アプリケーション実行部６２に接続されたタッチパネル６４と、アプリケーション実行部６２に接続された受話用のスピーカ６６と、同じくアプリケーション実行部６２に接続されたステレオスピーカ６８とを含む。

【００３２】

制御部５８は、バッファ５４に蓄積された音声データに対してローカルな音声認識処理を実行する音声認識処理部８０と、音声認識処理部８０の出力する音声認識結果に、音声認識サーバ３６への音声データの送受信を制御するための所定のキーワード（開始キーワード及び終了キーワード）が含まれているか否かを判定し、含まれている場合には検出信号をそのキーワードとともに出力する判定部８２と、判定部８２が判定の対象とする開始キーワードを１又は複数個記憶するキーワード辞書８４とを含む。なお、音声認識処理部８０は、無音区間が所定のしきい値時間以上続くと発話が終了したとみなし、発話終了検出信号を出力する。判定部８２は、発話終了検出信号を受信すると、通信制御部８６に対して音声認識サーバ３６へのデータの送信を終了する指示を出すものとする。

キーワード辞書８４に記憶される開始キーワードは、通常の発話とできるだけ区別するために、名詞を用いるものとする。携帯電話３４に何らかの処理を依頼することを考えると、この名詞としては特に固有名詞を使用することが自然であり好ましい。固有名詞でなく、特定のコマンド用語を用いるようにしてもよい。

終了キーワードとしては、日本語の場合には、開始キーワードとは異なり、より一般的に動詞の命令形、動詞の基本形＋終止形、依頼表現、又は疑問表現等、通常の日本語で他人に何かを依頼する表現を採用する。すなわち、これらのいずれかを検出したときに、終了キーワードを検出したものと判定する。こうすることにより、ユーザが自然な話し方で携帯電話に処理を依頼することが可能になる。こうした処理を可能とするためには、音声認識処理部８０が、認識結果の各単語にその単語の品詞、動詞の活用形、助詞の種類等を示す情報を付すようなものであればよい。

【００３３】

制御部５８はさらに、判定部８２から検出信号と検出されたキーワードとを受信したことに応答し、検出されたキーワードが開始キーワードか終了キーワードかにしたがって、バッファ５４に蓄積された音声データを音声認識サーバ３６に送信する処理を開始又は終了するための通信制御部８６と、判定部８２が音声認識処理部８０による音声認識結果内に検出したキーワードのうち、開始キーワードを記憶する一時記憶部８８と、受信データバッファ６０が受信した音声認識サーバ３６の音声認識結果のテキストの先頭部分と、一時記憶部８８に記憶された、ローカル音声認識結果の開始キーワードとを比較し、両者が一致したときには受信データバッファ６０に記憶されたデータの内、開始キーワードの後に続く部分を使用して所定のアプリケーションを実行するようアプリケーション実行部６２を制御するための実行制御部９０とを含む。本実施の形態では、どのようなアプリケーションを実行するかはアプリケーション実行部６２が受信データバッファ６０に記憶された内容によって判定する。

【００３４】
音声認識処理部８０が、バッファ５４に蓄積された音声データに対する音声認識をするにあたり、音声認識結果を出力する仕方には２通りある。発話ごと方式と逐次方式とである。発話ごと方式は、音声データ内に所定時間を超える無音区間があったときに、それまでの音声の音声認識結果を出力し、次の発話区間から新たに音声認識を開始する。逐次方式は、随時バッファ５４に蓄積されている音声データ全体に対する音声認識結果を所定時間間隔（たとえば１００ミリ秒ごと）で出力する。したがって、発話区間が長くなると音声認識音結果のテキストもそれにつれて長くなる。本実施の形態では、音声認識処理部８０は逐次方式を採用している。なお、発話区間が非常に長くなると、音声認識処理部８０による音声認識が困難になる。したがって音声認識処理部８０は、発話区間が所定時間長以上になると、強制的に発話が終了したものとしてそれまでの音声認識を終了し、新たな音声認識を開始するものとする。なお、音声認識処理部８０による音声認識の出力が発話ごとの方式である場合でも、以下の機能は本実施の形態のものと同様に実現できる。

【００３５】
図３を参照して、音声認識処理部８０の出力タイミングについて説明する。発話１００が、第１の発話１１０と第２の発話１１２とを含み、両者の間に無音区間１１４があるものとする。音声認識処理部８０は、バッファ５４に音声データが蓄積されていくと、音声認識結果１２０で示されるように、１００ミリ秒ごとに、バッファ５４に蓄積された音声全体に対する音声認識結果を出力する。この方式では、音声認識結果の一部が途中で修正される場合もある。例えば、図３に示す音声認識結果１２０の場合、２００ミリ秒時点で出力された「熱い」という単語が３００ミリ秒時点では「暑い」に修正されている。この方式では、無音区間１１４の時間長が所定のしきい値より大きい場合には、発話が終了したものとみなされる。その結果、バッファ５４に蓄積されていた音声データはクリアされ（読捨てられ）、次の発話に対する音声認識処理が開始される。図３の場合には、次の音声認識結果１２２が新たな時間情報とともに音声認識処理部８０から出力される。判定部８２は、音声認識結果１２０又は音声認識結果１２２等の各々について、音声認識結果が出力されるごとに、キーワード辞書８４に記憶された開始キーワードのいずれかと一致しているか、又は終了キーワードの条件を充足しているか否かを判定し、開始キーワード検出信号又は終了キーワード検出信号を出力する。ただし、本実施の形態では、開始キーワードは音声認識サーバ３６への音声データの送信が行なわれていないときにしか検出されず、終了キーワードは開始キーワードが検出された後でなければ検出されない。

【００３６】
［動作］
携帯電話３４は以下のように動作する。マイクロフォン５０は常に周囲の音声を検知して音声信号をフレーム化処理部５２に与える。フレーム化処理部５２は、音声信号をデジタル化及びフレーム化し、バッファ５４に順次入力する。音声認識処理部８０は、バッファ５４に蓄積されていく音声データの全体について、１００ミリ秒ごとに音声認識を行ない、その結果を判定部８２に出力する。ローカルな音声認識処理部８０は、しきい値時間以上の無音区間を検知するとバッファ５４をクリアし、発話の終了を検出したことを示す信号（発話終了検出信号）を判定部８２に出力する。

【００３７】
判定部８２は、音声認識処理部８０からローカルな音声認識結果を受信すると、その中にキーワード辞書８４に記憶された開始キーワードがあるか、又は終了キーワードとしての条件を充足する表現があるかを判定する。判定部８２は、音声認識サーバ３６に音声データを送信していない期間にローカルな音声認識結果内に開始キーワードを検出した場合、開始キーワード検出信号を通信制御部８６に与える。一方、判定部８２は、音声認識サーバ３６に音声データを送信している間にローカルな音声認識結果内に終了キーワードを検出すると、終了キーワード検出信号を通信制御部８６に与える。判定部８２はまた、音声認識処理部８０から発話終了検出信号を受信したときには、音声認識サーバ３６への音声データの送信を終了するよう通信制御部８６に対して指示を与える。

【００３８】
通信制御部８６は、判定部８２から開始キーワード検出信号が与えられると、送受信部５６を制御してバッファ５４に蓄積されているデータのうち、検出された開始キーワードの先頭位置からデータを読出して、音声認識サーバ３６に送信する処理を開始させる。このとき、通信制御部８６は、判定部８２から与えられた開始キーワードを一時記憶部８８に保存する。通信制御部８６は、判定部８２から終了キーワード検出信号が与えられると、送受信部５６を制御して、バッファ５４に蓄積されているデータのうち、検出された終了キーワードまでの音声データを音声認識サーバ３６に送信させた後に送信を終了させる。判定部８２から発話終了検出信号による送信終了の指示が与えられると、通信制御部８６は、送受信部５６を制御して、バッファ５４に記憶されている音声データのうち、発話の終了が検出された時間までの音声データを全て音声認識サーバ３６に送信させた後に送信を終了させる。

【００３９】
受信データバッファ６０は、通信制御部８６によって音声認識サーバ３６への音声データの送信が開始された後、音声認識サーバ３６から送信されてくる音声認識結果のデータを蓄積する。実行制御部９０は、受信データバッファ６０の先頭部分が、一時記憶部８８に保存されている開始キーワードと一致するか否かを判定する。両者が一致していると、実行制御部９０は、アプリケーション実行部６２を制御し、受信データバッファ６０のうちで、開始キーワードと一致した部分の次からのデータを読出すようにさせる。アプリケーション実行部６２は、受信データバッファ６０から読出したデータに基づいてどのようなアプリケーションを実行するかを判定し、そのアプリケーションに音声認識結果を渡して処理させる。処理の結果は、例えばタッチパネル６４に表示されたり、スピーカ６６又はステレオスピーカ６８から音声の形で出力されたりする。

【００４０】
例えば図４を参照して、具体的な例を説明する。ユーザが発話１４０を行なったものとする。発話１４０は、「ｖＧａｔｅ君」という発話部分１５０と、「このあたりのラーメン屋さん調べて」という発話部分１５２とを含む。発話部分１５２は、「このあたりのラーメン屋さん」という発話部分１６０と、「調べて」という発話部分１６２とを含む。

【００４１】
ここでは、開始キーワードして例えば「ｖＧａｔｅ君」、「羊君」等が登録されているものとする。すると、発話部分１５０が開始キーワードと一致しているため、発話部分１５０が音声認識された時点で音声データ１７０を音声認識サーバ３６に送信する処理が開始される。音声データ１７０は、図４に示すように発話１４０の音声データの全体を含み、その先頭は開始キーワードに対応する音声データ１７２である。

【００４２】
一方、発話部分１６２のうち、「調べて」という表現は依頼表現であり終了キーワードとしての条件を充足する。したがって、この表現がローカル音声認識結果中に検出された時点で、音声データ１７０を音声認識サーバ３６に送信する処理は終了する。

【００４３】
音声データ１７０の送信が終了すると、音声データ１７０に対する音声認識結果１８０が音声認識サーバ３６から携帯電話３４に送信され、受信データバッファ６０に蓄積される。音声認識結果１８０の先頭部分１８２は、開始キーワードに対応する音声データ１７２の音声認識結果である。この先頭部分１８２が、発話部分１５０（開始キーワード）に対するクライアント音声認識結果と一致すると、音声認識結果１８０の内、先頭部分１８２の次の部分からの音声認識結果１８４がアプリケーション実行部６２（図１参照）に送信され、適切なアプリケーションにより処理される。先頭部分１８２が発話部分１５０（開始キーワード）に対するクライアント音声認識結果と一致していないと、受信データバッファ６０はクリアされ、アプリケーション実行部６２は何ら動作しない。

【００４４】
以上のようにこの実施の形態によれば、ローカル音声認識により発話中に開始キーワー

ドが検出されると音声データを音声認識サーバ３６に送信する処理が開始される。ローカル音声認識により発話中に終了キーワードが検出されると、音声認識サーバ３６への音声データの送信が終了される。音声認識サーバ３６から送信されてくる音声認識結果の先頭部分と、ローカル音声認識により検出された開始キーワードとが比較され、両者が一致していれば、音声認識サーバ３６の音声認識結果を用いて何らかの処理が実行される。したがって、この実施の形態では、携帯電話３４に何らかの処理を実行させようとする場合、ユーザは他に何もせず、単に開始キーワードと実行内容とを発話するだけでよい。開始キーワードがローカル音声認識で正しく認識されれば、携帯電話３４による音声認識の結果を用いた所望の処理が実行され、結果が携帯電話３４により出力される。音声入力の開始のためのボタンを押したりする必要はなく、携帯電話３４をより簡単に使用できる。

【００４５】

こうした処理で問題になるのは、開始キーワードが誤って検出された場合である。前述したように、一般的に、携帯型端末でローカルに実行される音声認識の精度は、音声認識サーバで実行される音声認識の精度よりも低い。したがってローカル音声認識で誤って開始キーワードが検出される可能性がある。そうした場合、誤って検出された開始キーワードに基づいて何らかの処理を実行し、その結果を携帯電話３４が出力すると、それはユーザが意図しない動作となってしまう。そのような動作は好ましくない。

【００４６】

本実施の形態では、仮にローカル音声認識で開始キーワードが誤検出されたとしても、音声認識サーバ３６からの音声認識結果の先頭部分が開始キーワードと一致していなければ携帯電話３４はその結果による処理は何も実行しない。携帯電話３４の状態は何も変化せず、見かけ上全く何もしていないように見える。したがって、ユーザは、上に記載したような処理が実行されたことには全く気付かない。

【００４７】

さらに、上記実施の形態では、開始キーワードがローカル音声認識で検出された場合に音声データを音声認識サーバ３６に送信する処理を開始し、終了キーワードがローカル音声認識で検出された場合に送信処理を終了する。音声の送信を終了するためにユーザが特別な操作をする必要がない。所定時間以上の空白を検出したときに送信を終了する場合と比較して、終了キーワードを検出すると直ちに音声認識サーバ３６への音声データの送信を終了できる。その結果、携帯電話３４から音声認識サーバ３６への無駄なデータ送信を防止できるし、音声認識のレスポンスも向上する。

【００４８】

［プログラムによる実現］

上記第１の実施の形態に係る携帯電話３４は、後述するような、コンピュータと同様の携帯電話ハードウェアと、その上のプロセッサにより実行されるプログラムとにより実現できる。図５に、図１の判定部８２及び通信制御部８６の機能を実現するプログラムの制御構造をフローチャート形式で示し、図６に、実行制御部９０の機能を実現するプログラムの制御構造をフローチャート形式で示す。ここでは両者を別プログラムとして記載しているが、両者をまとめることもできるし、それぞれさらに細かい単位のプログラムに分割することもできる。

【００４９】

図５を参照して、判定部８２及び通信制御部８６の機能を実現するプログラムは、携帯電話３４の電源投入時に起動されると、使用するメモリエリアの初期化等を実行するステップ２００と、システムからプログラムの実行を終了することを指示する終了信号を受信したか否かを判定し、終了信号を受信したときには必要な終了処理を実行してこのプログラムの実行を終わるステップ２０２と、終了信号が受信されていないときに、音声認識処理部８０からローカル音声認識結果を受信したか否かを判定し、受信していなければ制御をステップ２０２に戻すステップ２０４とを含む。前述したとおり、音声認識処理部８０は所定時間ごとに音声認識結果を逐次的に出力する。したがってステップ２０４の判定は、所定時間ごとにＹＥＳとなる。

【００５０】
このプログラムはさらに、ステップ２０４でローカル音声認識の結果を受信したと判定されたことに応答して、キーワード辞書８４に記憶された開始キーワードのいずれかがローカル音声認識結果に含まれるか判定し、含まれていない場合には制御をステップ２０２に戻すステップ２０６と、開始キーワードのいずれかがローカル音声認識結果にあったときに、その開始キーワードを一時記憶部８８に保存するステップ２０８と、バッファ５４（図２）に記憶されている音声データのうち、開始キーワードの先頭部分から音声認識サーバ３６への音声データの送信を開始させるよう送受信部５６に指示するステップ２１０とを含む。以後、処理は携帯電話３４への音声データ送信中の処理に移る。

【００５１】
音声データ送信中の処理は、システムの終了信号を受信したか否かを判定し、受信したときには必要な処理を実行してこのプログラムの実行を終了するステップ２１２と、終了信号が受信されていないときに、音声認識処理部８０からローカル音声認識結果を受信したか否かを判定するステップ２１４と、ローカル音声認識結果を受信したときに、その中に終了キーワードの条件を充足する表現があるか否かを判定し、なければ制御をステップ２１２に戻すステップ２１６と、ローカル音声認識結果中に終了キーワードの条件を充足する表現があったときに、バッファ５４に記憶されている音声データのうち、終了キーワードが検出された部分の末尾までを音声認識サーバ３６に送信して送信を終了し、制御をステップ２０２に戻すステップ２１８とを含む。

【００５２】
このプログラムはまた、ステップ２１４でローカル音声認識結果を音声認識処理部８０から受信していないと判定されたときに、発話なしで所定時間が経過したか否かを判定し、所定時間が経過していなければ制御をステップ２１２に戻すステップ２２０と、発話なしで所定時間が経過したときに、バッファ５４に記憶されている音声データの音声認識サーバ３６への送信を終了し、制御をステップ２０２に戻すステップ２２２とを含む。

【００５３】
図６を参照して、図２の実行制御部９０を実現するプログラムは、携帯電話３４の電源投入時に起動され、必要な初期化処理を実行するステップ２４０と、終了信号を受信したか否かを判定し受信したときにはこのプログラムの実行を終了するステップ２４２と、終了信号を受信していないときに、音声認識サーバ３６から音声認識結果のデータを受信したか否かを判定し、受信していなければ制御をステップ２４２に戻すステップ２４４とを含む。

【００５４】
このプログラムはさらに、音声認識サーバ３６から音声認識結果のデータを受信したときに、一時記憶部８８に保存されていた開始キーワードを読出すステップ２４６と、ステップ２４６で読出された開始キーワードが音声認識サーバ３６からの音声認識結果のデータの先頭部分と一致するか否かを判定するステップ２４８と、両者が一致したときに、音声認識サーバ３６による音声認識結果のうち、開始キーワードの終端部の次の位置から終了までのデータを受信データバッファ６０から読出すようアプリケーション実行部６２を制御するステップ２５０と、ステップ２４８で開始キーワードが一致しないと判定されたときに、受信データバッファ６０に記憶された音声認識サーバ３６による音声認識結果をクリアする（又は読捨てる）ステップ２５４と、ステップ２５０又はステップ２５４の後に、一時記憶部８８をクリアして制御をステップ２４２に戻すステップ２５２とを含む。

【００５５】
図５に示すプログラムによれば、ローカルな音声認識結果が開始キーワードとマッチしているとステップ２０６で判定されると、ステップ２０８でその開始キーワードが一時記憶部８８に保存され、ステップ２１０以後で、バッファ５４に記憶された音声データのうち、開始キーワードと一致した先頭部分からの音声データが音声認識サーバ３６に送信される。音声データの送信中にローカルな音声認識結果中に終了キーワードとしての条件を充足する表現が検出されると（図５のステップ２１６でＹＥＳ）、バッファ５４に記憶さ

れた音声データのうち、終了キーワードの部分の終端まで音声認識サーバ３６に送信された後、送信が終了する。

【００５６】

一方、音声認識サーバ３６から音声認識結果を受信したときに、図６のステップ２４８の判定が肯定なら、音声認識結果のうち、開始キーワードと一致した部分の末尾以後が受信データバッファ６０からアプリケーション実行部６２に読出され、アプリケーション実行部６２が音声認識結果の内容に応じた適切な処理を実行する。

【００５７】

したがって、図５及び図６に制御構造を示すプログラムを携帯電話３４で実行することにより、上記した実施の形態の機能を実現できる。

【００５８】

＜第２の実施の形態＞

上記実施の形態では、ローカル音声認識で開始キーワードを検出すると、その開始キーワードを一時的に一時記憶部８８に保存している。そして、音声認識サーバ３６から音声認識結果が返ってきたときに、音声認識結果の先頭部分と一時的に保存された開始キーワードとが一致するか否かにより、音声認識サーバ３６の音声認識結果を使用した処理を実行するか否かを判定している。しかし本発明はそのような実施の形態には限定されない。そのような判定を行なわず、音声認識サーバ３６の音声認識結果をそのまま利用する実施の形態も考えられる。これは、特にローカル音声認識でのキーワード検出の精度が十分に高いときに有効である。

【００５９】

図７を参照して、この第２の実施の形態に係る携帯電話２６０は、第１の実施の形態の携帯電話３４とほぼ同様な構成である。しかし、音声認識サーバ３６による音声認識結果と開始キーワードとの照合に必要な機能ブロックを含まず、より簡略となっている点で携帯電話３４と異なっている。

【００６０】

具体的には、携帯電話２６０は、図１に示す制御部５８を簡略化し、音声認識サーバ３６からの音声認識結果と開始キーワードとの照合を行なわないようにした制御部２７０を制御部５８に代えて持つ点と、制御部５８の制御によらず、音声認識サーバ３６からの音声認識結果を一時的に保持し、全て出力する受信データバッファ２７２を図１の受信データバッファ６０に代えて持つ点と、制御部２７０の制御を受けず、音声認識サーバ３６からの音声認識結果を全て処理するアプリケーション実行部２７４を図１のアプリケーション実行部６２に代えて持つ点で第１の実施の形態の携帯電話３４と異なっている。

【００６１】

制御部２７０は、図１に示す一時記憶部８８及び実行制御部９０を持たない点、及び、図１の通信制御部８６に代えて、ローカルな音声認識結果内に開始キーワードが検出されたときに、バッファ５４に記憶されている音声データの内で、開始キーワードに対応する位置の直後からのデータを音声認識サーバ３６に送信する処理を開始するよう送受信部５６を制御する機能を持つ通信制御部２８０を持つ点で図１の制御部５８と異なっている。なお、通信制御部２８０もまた、制御部５８と同様、ローカルな音声認識結果の中に終了キーワードが検出されたときには、音声認識サーバ３６への音声データの送信を終了するよう送受信部５６を制御する。

【００６２】

図８を参照して、この実施の形態に係る携帯電話２６０の動作の概略について説明する。発話１４０の構成は図４に示すものと同様であるものとする。本実施の形態に係る制御部２７０は、発話１４０中の発話部分１５０に開始キーワードが検出されたときに、音声データのうち、開始キーワードが検出された部分の次から終了キーワードが検出された直後（図８に示す発話部分１５２に相当）までの音声データ２９０を音声認識サーバ３６に送信する。すなわち、音声データ２９０には開始キーワード部分の音声データは含まれない。その結果、音声認識サーバ３６から返信される音声認識結果２９２にも開始キーワー

ドは含まれない。したがって、発話部分１５０の部分のローカル音声認識の結果が正しければ、サーバからの音声にも開始キーワードは含まれず、音声認識結果２９２の全体をアプリケーション実行部２７４が処理しても特に不都合は生じない。

【００６３】

図９に、この実施の形態に係る携帯電話２６０の判定部８２及び通信制御部２８０の機能を実現するためのプログラムの制御構造をフローチャート形式で示す。この図は、第１の実施の形態の図５に示すものに相当する。なおこの実施の形態では、第１の実施の形態の図６に制御構造を示すようなプログラムは必要ない。

【００６４】

図９を参照して、このプログラムは、図５に制御構造を示すものからステップ２０８を削除し、ステップ２１０に代えて、バッファ５４に記憶された音声データのうち、開始キーワードの終端の次の位置から音声認識サーバ３６に音声データを送信するように送受信部５６を制御するステップ３００を含む。その他の点では、このプログラムは図５に示すものと同じ制御構造を示す。このプログラムの実行時の制御部２７０の動作も、既に説明したものから十分に明らかである。

【００６５】

この第２の実施の形態では、音声データの送信を開始するためにユーザが何らかの操作を特に行なう必要がないという点と、音声データを音声認識サーバ３６に送信するにあたり、データ量を少なく抑えることができるという点で第１の実施の形態と同じ効果を得ることができる。またこの第２の実施の形態では、ローカル音声認識のキーワード検出の精度が高ければ、簡単な制御でサーバを用いた音声認識結果を利用した様々な処理を利用できるという効果を奏する。

【００６６】

［携帯電話のハードウェアブロック図］

図１０に、第１の実施の形態に係る携帯電話３４及び第２の実施の形態に係る携帯電話２６０を実現する携帯電話のハードウェアブロック図を示す。以下の説明では、携帯電話３４及び２６０を代表して携帯電話３４について説明する。

【００６７】

図１０を参照して、携帯電話３４は、マイクロフォン５０及びスピーカ６６と、マイクロフォン５０及びスピーカ６６が接続されたオーディオ回路３３０と、オーディオ回路３３０が接続されたデータ転送用及び制御信号転送用のバス３２０と、ＧＰＳ用、携帯電話回線用、及びその他規格にしたがった無線通信用のアンテナを備え、様々な通信を無線により実現する無線回路３３２と、無線回路３３２と携帯電話３４の他のモジュールとの間を仲介する処理を行なう、バス３２０に接続された通信制御回路３３６と、通信制御回路３３６に接続され、携帯電話３４に対する利用者の指示入力を受けて入力信号を通信制御回路３３６に与える操作ボタン３３４と、バス３２０に接続され、様々なアプリケーションを実行するためのＣＰＵ（図示せず）、ＲＯＭ（読出専用メモリ：図示せず）及びＲＡＭ（Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ：図示せず）を備えたアプリケーション実行用ＩＣ（集積回路）３２２と、アプリケーション実行用ＩＣ３２２に接続されたカメラ３２６、メモリカード入出力部３２８、タッチパネル６４及びＤＲＡＭ（Ｄｙｎａｍｉｃ　ＲＡＭ）３３８と、アプリケーション実行用ＩＣ３２２に接続され、アプリケーション実行用ＩＣ３２２により実行される様々なアプリケーションを記憶した不揮発性メモリ３２４とを含む。

【００６８】

不揮発性メモリ３２４には、図１に示す音声認識処理部８０を実現するローカル音声認識処理プログラム３５０と、判定部８２、通信制御部８６及び実行制御部９０を実現する発話送受信制御プログラム３５２と、キーワード辞書８４と、キーワード辞書８４に記憶されるキーワードを保守するための辞書保守プログラム３５６とが記憶されている。これらプログラムは、いずれもアプリケーション実行用ＩＣ３２２による実行時にはアプリケーション実行用ＩＣ３２２内の図示しないメモリにロードされ、アプリケーション実行用

ＩＣ３２２内のＣＰＵが持つプログラムカウンタと呼ばれるレジスタにより指定されるアドレスから読出され、ＣＰＵにより実行される。実行結果は、ＤＲＡＭ３３８、メモリカード入出力部３２８に装着されたメモリカード、アプリケーション実行用ＩＣ３２２内のメモリ、通信制御回路３３６内のメモリ、オーディオ回路３３０内のメモリのうち、プログラムにより指定されるアドレスに格納される。

【００６９】

図２及び図７に示すフレーム化処理部５２はオーディオ回路３３０により実現される。バッファ５４及び受信データバッファ２７２は、ＤＲＡＭ３３８若しくは通信制御回路３３６又はアプリケーション実行用ＩＣ３２２内のメモリにより実現される。送受信部５６は無線回路３３２及び通信制御回路３３６により実現される。図１の制御部５８及びアプリケーション実行部６２に、並びに図７の制御部２７０及びアプリケーション実行部２７４は、本実施の形態ではいずれもアプリケーション実行用ＩＣ３２２により実現される。

【００７０】

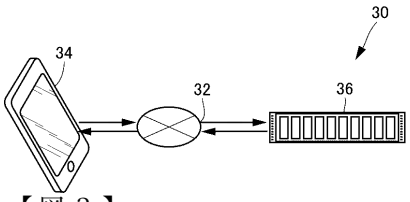
今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内での全ての変更を含む。

【符号の説明】

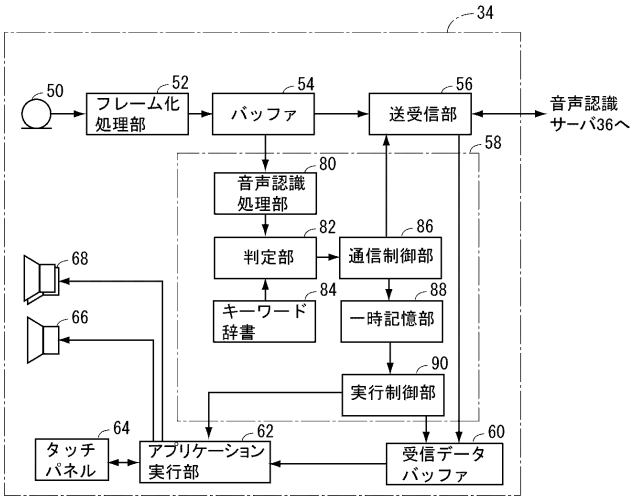
【００７１】

３０　音声認識システム
３４　携帯電話
３６　音声認識サーバ
５０　マイクロフォン
５４　バッファ
５６　送受信部
５８　制御部
６０　受信データバッファ
６２　アプリケーション実行部
８０　音声認識処理部
８２　判定部
８４　キーワード辞書
８６　通信制御部
８８　一時記憶部
９０　実行制御部

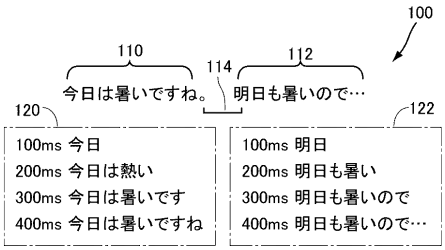
【図１】



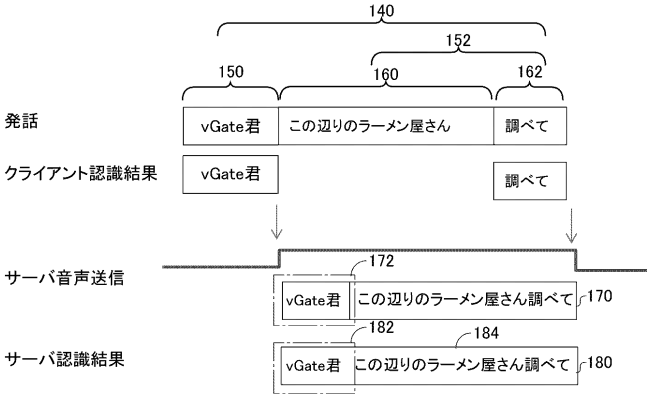
【図２】



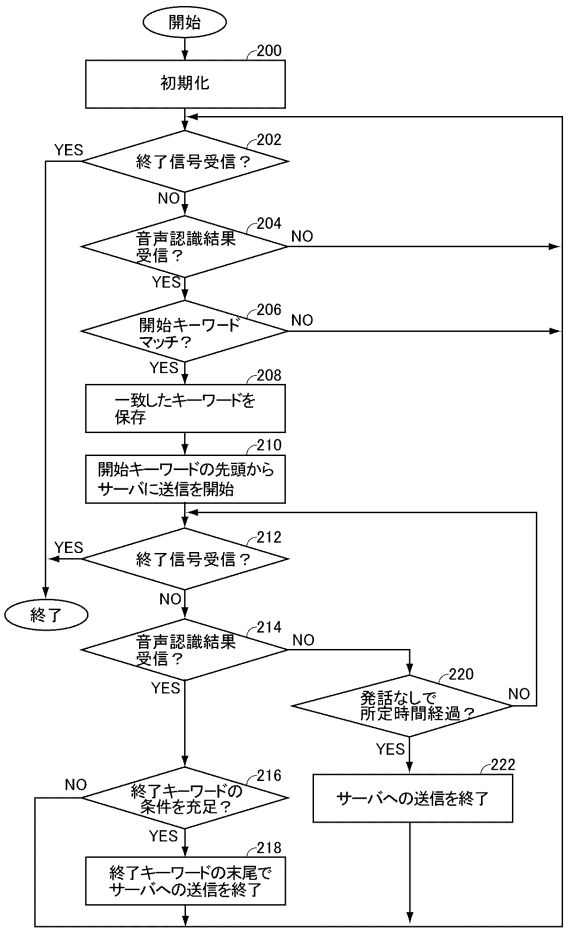
【図３】



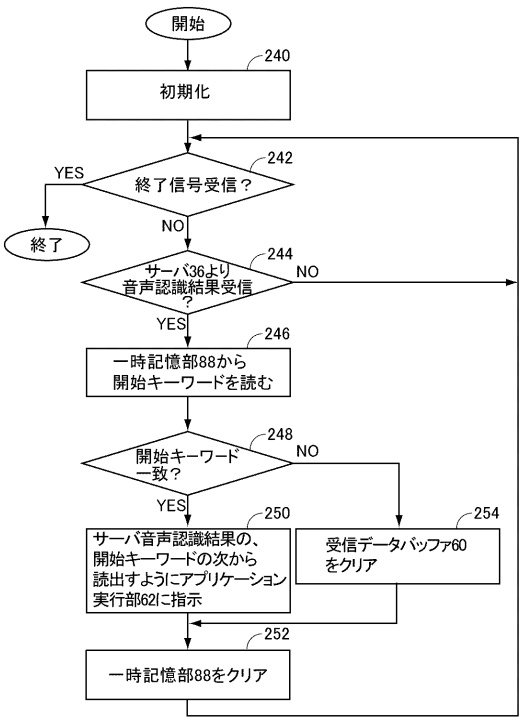
【図４】



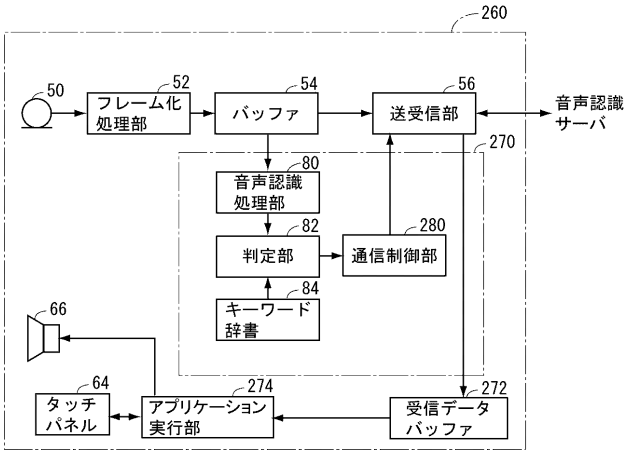
【図５】



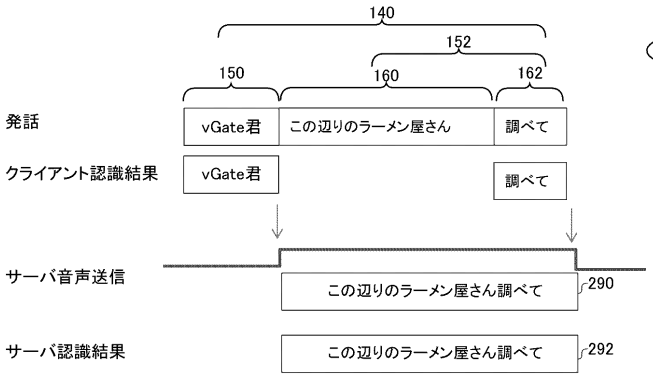
【図６】



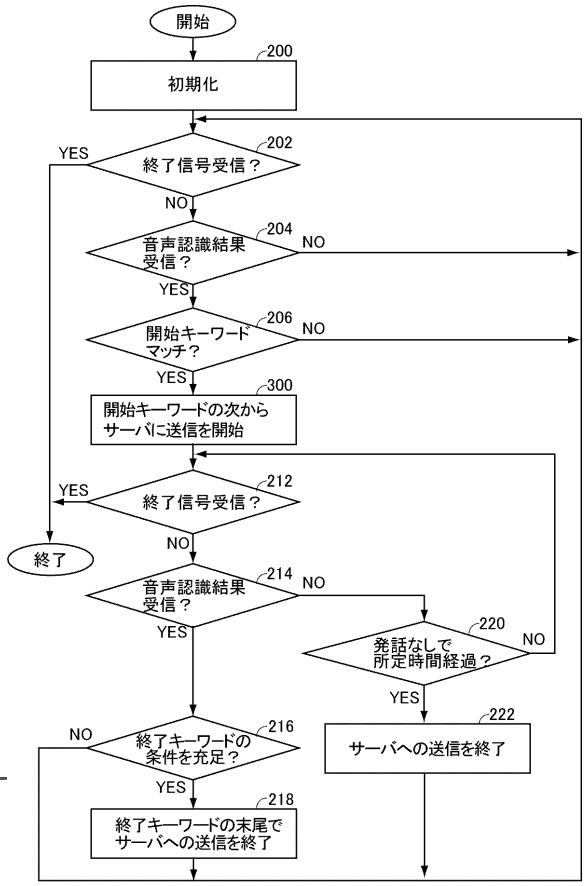
【図７】


260
50
フレーム化処理部
52
バッファ
54
送受信部
56
音声認識サーバ
270
音声認識処理部
80
判定部
82
通信制御部
280
キーワード辞書
84
66
64
タッチパネル
アプリケーション実行部
274
受信データバッファ
272

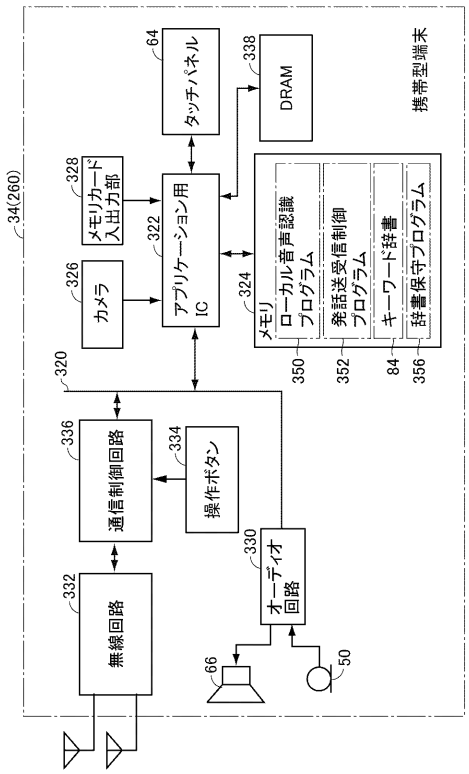
【図８】


140
152
150
160
162
発話
vGate君
この辺りのラーメン屋さん
調べて
クライアント認識結果
vGate君
調べて
サーバ音声送信
この辺りのラーメン屋さん調べて
290
サーバ認識結果
この辺りのラーメン屋さん調べて
292

【図９】


開始
200
初期化
202
終了信号受信？
YES
NO
204
音声認識結果受信？
NO
YES
206
開始キーワードマッチ？
NO
YES
300
開始キーワードの次からサーバに送信を開始
212
終了信号受信？
YES
NO
終了
214
音声認識結果受信？
NO
YES
220
発話なしで所定時間経過？
NO
YES
222
サーバへの送信を終了
216
終了キーワードの条件を充足？
NO
YES
218
終了キーワードの末尾でサーバへの送信を終了

【図１０】


携帯型端末
64
タッチパネル
338
DRAM
328
メモリカード入出力部
322
アプリケーション用IC
326
カメラ
324
メモリ
ローカル音声認識プログラム
発話送受信制御プログラム
キーワード辞書
辞書保守プログラム
350
352
84
356
34(260)
320
336
通信制御回路
334
操作ボタン
330
オーディオ回路
332
無線回路
66
50