(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5050175号
(P5050175)

(45)発行日　平成24年10月17日(2012.10.17)　(24)登録日　平成24年8月3日(2012.8.3)

(51)Int.Cl.		F I		
G10L 15/28	(2006.01)	G10L	15/28	210A
G10L 15/06	(2006.01)	G10L	15/28	370E
G10L 15/22	(2006.01)	G10L	15/06	400V
		G10L	15/22	470Z

請求項の数 8　(全 18 頁)

(21)出願番号　特願2008-173551 (P2008-173551)
(22)出願日　平成20年7月2日 (2008.7.2)
(65)公開番号　特開2010-14885 (P2010-14885A)
(43)公開日　平成22年1月21日 (2010.1.21)
審査請求日　平成23年5月26日 (2011.5.26)

(73)特許権者　393031586
株式会社国際電気通信基礎技術研究所
京都府相楽郡精華町光台二丁目２番地２
(74)代理人　100099933
弁理士　清水　敏
(72)発明者　松田　繁樹
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内
(72)発明者　中村　哲
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内
(72)発明者　葦苅　豊
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内

最終頁に続く

(54)【発明の名称】音声認識機能付情報処理端末

(57)【特許請求の範囲】
【請求項１】
音声信号から所定の音響特徴量を抽出して記憶するための特徴量記憶手段と、
前記所定の音響特徴量を予め定められた音声認識サーバに送信するための特徴量送信手段と、
前記サーバから前記所定の音響特徴量に対する音声認識の結果のテキストを受信するための受信手段と、
音声認識のための統計的音響モデルと、音声認識のための、カテゴリ別に編成された複数個のカテゴリ別言語モデルとを記憶するためのカテゴリ別モデル記憶手段と、
前記受信手段が受信した前記テキスト中の、未知語のタグ付けがされた区間に対応する音響特徴量を前記特徴量記憶手段から読出し、前記モデル記憶手段に記憶された前記統計的音響モデル、及び前記カテゴリ別言語モデルの内で前記未知語のカテゴリに対応する言語モデル、を使用して音声認識を行なうための音声認識手段と、
前記受信手段が受信した前記テキスト中の前記未知語のタグ付けがされた区間を、前記音声認識手段の出力で置換するための置換手段とを含む、音声認識機能付情報処理端末。
【請求項２】
前記受信手段が受信した前記テキスト中に未知語のタグ付けがされた区間があるか否かを判定するための判定手段と、
前記判定手段の判定結果に応答して、前記受信手段が受信した前記テキストと、前記置換手段によって前記未知語が置換された前記テキストとを選択的に出力するための選択手

段とをさらに含む、請求項１に記載の音声認識機能付情報処理端末。
【請求項３】
前記特徴量記憶手段は、
前記音声信号を所定時間ごとに所定長のフレームにフレーム化するためのフレーム化手段と、
前記フレーム化手段によりフレーム化されたフレームの各々の音声信号から、所定の複数個の音響特徴量を抽出するための特徴量抽出手段と、
前記フレーム化手段によりフレーム化されたフレームの各々に対して前記特徴量抽出手段により抽出された前記所定の複数個の音響特徴量を所定の圧縮アルゴリズムにより圧縮するための圧縮手段と、
前記フレーム化手段によりフレーム化されたフレームの各々に対して、前記圧縮手段により圧縮された音響特徴量を記憶するための記憶手段とを含み、
前記特徴量送信手段は、前記圧縮手段により圧縮された前記音響特徴量を送信するための手段を含む、請求項１又は請求項２に記載の音声認識機能付情報処理端末。
【請求項４】
前記音声認識手段は、
前記受信手段が受信した前記テキスト中の、未知語のタグ付けがされた区間に対応するフレーム列の各々の音響特徴量を前記特徴量記憶手段から読出し、前記所定の圧縮アルゴリズムに対応する伸長アルゴリズムを用いて伸長するための伸長手段と、
前記伸長手段により伸長されたフレーム列の前記複数個の音響特徴量を入力として、前記モデル記憶手段に記憶された前記統計的音響モデル、及び前記カテゴリ別言語モデルの内で前記未知語のカテゴリに対応する言語モデルを使用して音声認識を行なうための手段とを含む、請求項３に記載の音声認識機能付情報処理端末。
【請求項５】
前記圧縮手段は、
前記複数個の所定の音響特徴量から予め組合された２つずつの音響特徴量の組合せの各々に対して予め準備されたコードブックを記憶するためのコードブック記憶手段と、
前記フレーム化手段によりフレーム化されたフレームの各々について、前記複数個の所定の音響特徴量から予め組合された２つずつの音響特徴量の組合せの各々を、前記コードブックのうちで対応するものを用いて符号化するための符号化手段とを含み、
前記送信するための手段は、前記フレーム化手段によりフレーム化されたフレームの各々について、前記符号化手段により得られた符号からなる符号列を送信するための手段を含む、請求項３又は請求項４に記載の音声認識機能付情報処理端末。
【請求項６】
前記複数個の所定の音響特徴量は、各フレームの音声信号の第０次から第１２次のＭＦＣＣパラメータと、パワーとを含む、請求項１～請求項５のいずれかに記載の音声認識機能付情報処理端末。
【請求項７】
前記音声認識機能付情報処理端末で実行可能なアプリケーションプログラムにより、前記音声認識機能付情報処理端末の使用者に関連して集積された情報を記憶するための関連情報記憶手段と、
前記関連情報記憶手段に記憶された前記情報を、カテゴリに分類するための分類手段と、
前記分類手段により分類されたカテゴリごとに統計的言語モデルを作成することにより、前記複数個のカテゴリ別言語モデルを作成するための言語モデル作成手段とをさらに含む、請求項１～請求項６のいずれかに記載の音声認識機能付情報処理端末。
【請求項８】
前記未知語のタグ付けがされた区間は、未知語のタグ付けがされた音節列を含む、請求項１～請求項７のいずれかに記載の音声認識機能付情報処理端末。
【発明の詳細な説明】

【技術分野】
【０００１】
この発明は通信機能を有する情報処理端末に関し、特に、携帯に便利な小さな筐体で、しかも音声認識による入力が可能な情報処理端末に関する。
【背景技術】
【０００２】
計算機の性能が向上し小型化するに伴い、携帯型情報端末が普及している。現代の携帯型情報端末は、例えば携帯電話のように、主たる機能の他にインターネットに接続する機能も持っており、電子メールによるコミュニケーションの有力なツールとなっている。
【０００３】
携帯型情報端末を電子メールなどのテキストベースのコミュニケーションにおいて使用する場合の最大の問題は、入力インターフェイスである。大きな情報端末であればフルキーボードを装備することも可能であるが、携帯電話のような小型の装置ではそれは難しい。そのため、テンキーパッドを複数回押すことによって一文字を入力するようなインターフェイスが一般的である。その結果、通常の情報端末と比較して入力に時間がかかるという問題がある。
【０００４】
こうした問題に対処すべく、あえてフルキーボードを備えた小型の情報端末もあるが、キートップが非常に小さくなってしまうために、結局は入力がしづらいという欠点がある。
【０００５】
一方、携帯型情報端末の高能力化に伴い、いわゆる音声認識技術を携帯型情報端末における入力に使用することも考えられている。ＣＰＵ（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）の処理能力の向上に伴い、そのようなことも不可能ではないと考えられる。
【０００６】
しかし、現代の音声認識技術の場合、音響モデル、単語辞書、及び統計的言語モデルなどを装備する必要がある。音声認識の性能を高めるためには、これらモデルは大容量にせざるを得ない。その結果、現在のところは携帯型情報端末装置単体で十分な性能の音声認識を行なうことは難しいか、コストが非常に高くなってしまう。
【０００７】
そこで、携帯型情報端末では音声認識を行なわず、携帯型情報端末から送られた音声をサーバ側で音声認識する音声認識システムが特許文献１に記載されている。特許文献１に記載された音声認識システムでは、予め、サーバの検索データベースに、氏名等と、住所等と、関連情報等とを関連づけて蓄積しておく。携帯型情報端末から音声を受取ると、住所等の一部若しくは全部、又は関連情報等を取得し、この取得された住所等の一部若しくは全部、又は関連情報等に基づいて検索データベースを検索し、この検索結果を用いて、氏名等の音声データを認識する。
【特許文献１】特開2008-015439号公報
【特許文献２】特開2008-129318号公報
【非特許文献１】山本博史他４名、「複数のマルコフモデルを用いた階層化言語モデルによる未登録語認識」、電子情報通信学会論文誌、Ｄ－ＩＩ、Ｖｏｌ．Ｊ８７－Ｄ－ＩＩ，Ｎｏ．１２，ｐｐ．２１０４－２１１１、２００４年１２月
【発明の開示】
【発明が解決しようとする課題】
【０００８】
特許文献１に記載の技術によれば、音声認識は携帯型情報端末装置ではなくサーバ側で行なわれる。そのため、携帯型情報端末装置に音声認識のシステムを搭載する必要はない。音声認識に必要なリソースはサーバ側に十分確保できるため、音声認識の精度は確保できる。

【０００９】
これと同様の技術として、音声認識の前段である特徴量の抽出までを携帯型情報端末で行ない、特徴量のみをサーバに送信するという、分散型音声認識という考え方もある。送信されるデータ量は音声そのものよりも特徴量の方が少なくなるため、通信容量が少なくなるという効果がある。また、サーバ側の負荷が軽くなるという効果もある。情報処理装置が高性能化していることに鑑みると、分散型音声認識システムがこれからの音声認識システムとしては有力である。
【００１０】
しかし、音声そのものを送信するにせよ、特徴量を送信するにせよ、音声認識の精度を十分に高めるためには、サーバ側の辞書に非常にたくさんの固有名詞を登録する必要がある。例えばある個人にとって、友人の名前又は愛称（ニックネーム）、行きつけのお店、通学する学校、よく利用する施設、駅名などの固有名詞はコミュニケーションを行なう上で非常に重要な意味を持つ。これらが十分な精度で認識されるためには、サーバ側にそれらの固有名詞を正しく登録しなければならない。万が一、ある固有名詞が登録されていない場合には、その固有名詞については正しい音声認識結果が得られない。固有名詞は無数に存在し、しかも新しく生ずるものがある。したがって、それら無数の固有名詞について最新の状態にサーバのリソースを維持する作業は大変なものになる。
【００１１】
しかも、登録される固有名詞の数が多ければ音声認識の精度が高まるかというと、必ずしもそうではない。例えば同じようで少し異なるような固有名詞が辞書又は言語モデルに複数個登録されている場合には、音声認識の精度が低くなる可能性がある。したがって仮に固有名詞を１００パーセント登録できたとしても、音声認識の精度が高まるとは限らないという問題がある。
【００１２】
このように、辞書に登録されていない単語をどのように扱うかは、未知語の問題として知られている。特許文献２には、未知語をカタカナ文字列として出力できるような言語モデルを作成するシステムが開示されている。
【００１３】
しかし、未知語がカタカナ文字列で出力されても、音声認識が正しく行なわれているわけではない。カタカナ文字列自体に誤りがあるかも知れず、仮にカタカナ文字列が正しくとも、日本語の場合には固有名詞としての文字列に変換されなければ正しい認識が行なわれたとはいえない。このように未知語を未知語として出力するだけでは、音声認識の精度を高めたことにならず、結局、サーバ側に十分なリソースを準備する必要があり、サーバ側のリソースの肥大化を招くことになる。
【００１４】
それゆえに本発明の目的は、分散型の音声認識を利用する情報処理端末であって、使用者にとって音声認識の精度が十分に高く、かつ音声認識を行なうサーバ側のリソースの極端な肥大化を防止できる音声認識機能付情報処理端末を提供することである。
【課題を解決するための手段】
【００１５】
本発明の第１の局面に係る音声認識機能付情報処理端末は、音声信号から所定の音響特徴量を抽出して記憶するための特徴量記憶手段と、所定の音響特徴量を予め定められた音声認識サーバに送信するための特徴量送信手段と、サーバから所定の音響特徴量に対する音声認識の結果のテキストを受信するための受信手段と、音声認識のための統計的音響モデルと、音声認識のための、カテゴリ別に編成された複数個のカテゴリ別言語モデルとを記憶するためのカテゴリ別モデル記憶手段と、受信手段が受信したテキスト中の、未知語のタグ付けがされた区間に対応する音響特徴量を特徴量記憶手段から読出し、モデル記憶手段に記憶された統計的音響モデル、及びカテゴリ別言語モデルの内で未知語のカテゴリに対応する言語モデル、を使用して音声認識を行なうための音声認識手段と、受信手段が受信したテキスト中の未知語のタグ付けがされた区間を、音声認識手段の出力で置換する

ための置換手段とを含む。

【００１６】

この情報処理端末では、特徴量記憶手段が、音声信号から所定の音響特徴量を抽出し、記憶する。この音響特徴量は、特徴量送信手段により音声認識サーバに送信される。音声認識サーバでの音声認識結果であるテキストは受信手段により受信される。このテキスト中の未知語部分には、未知語であることを示すタグと、その未知語が属するカテゴリを示すタグとが付されている。音声認識手段は、特徴量記憶手段に記憶されている音響特徴量のうち、未知語に対応する部分を読出し、未知語に付されていたタグのカテゴリに対応するカテゴリ別言語モデルを使用して音声認識を行なう。置換手段は、音声認識の結果で未知語部分を置換する。

【００１７】

カテゴリ別言語モデルはこの情報処理端末に固有のものである。したがってこれらカテゴリ別言語モデルは、利用者にとって特に関連ある固有名詞などから生成される。サーバで未知語として認識された音響特徴量の部分を、情報処理端末でこのカテゴリ別言語モデルを用いて音声認識し直すことにより、サーバでは未知語であった固有名詞が正しく認識される可能性が大きくなる。そのために情報処理端末に多くのリソースを準備する必要はない。また、サーバでも未知語の音声認識を行なうためにリソースを肥大化させる必要がない。その結果、分散型の音声認識を利用する情報処理端末であって、使用者にとって音声認識の精度が十分に高く、かつ音声認識を行なうサーバ側のリソースの極端な肥大化を防止できる音声認識機能付情報処理端末を提供できる。

【００１８】

好ましくは、音声認識機能付情報処理端末は、受信手段が受信したテキスト中に未知語のタグ付けがされた区間があるか否かを判定するための判定手段と、判定手段の判定結果に応答して、受信手段が受信したテキストと、置換手段によって未知語が置換されたテキストとを選択的に出力するための選択手段とをさらに含む。

【００１９】

サーバから受けた音声認識結果に未知語がなければそれを選択し、未知語がある場合だけ情報処理端末での音声認識を行なう。情報処理端末において余分な処理をする必要がなく、音声認識の結果をより早く提示することが可能になる。

【００２０】

より好ましくは、特徴量記憶手段は、音声信号を所定時間ごとに所定長のフレームにフレーム化するためのフレーム化手段と、フレーム化手段によりフレーム化されたフレームの各々の音声信号から、所定の複数個の音響特徴量を抽出するための特徴量抽出手段と、フレーム化手段によりフレーム化されたフレームの各々に対して特徴量抽出手段により抽出された所定の複数個の音響特徴量を所定の圧縮アルゴリズムにより圧縮するための圧縮手段と、フレーム化手段によりフレーム化されたフレームの各々に対して、圧縮手段により圧縮された音響特徴量を記憶するための記憶手段とを含み、特徴量送信手段は、圧縮手段により圧縮された音響特徴量を送信するための手段を含む。

【００２１】

サーバには、圧縮された音響特徴量が送信される。その結果、情報処理端末から音声認識のためのサーバへの送信データ量を少なく抑えることができる。

【００２２】

さらに好ましくは、音声認識手段は、受信手段が受信したテキスト中の、未知語のタグ付けがされた区間に対応するフレーム列の各々の音響特徴量を特徴量記憶手段から読出し、所定の圧縮アルゴリズムに対応する伸長アルゴリズムを用いて伸長するための伸長手段と、伸長手段により伸長されたフレーム列の複数個の音響特徴量を入力として、モデル記憶手段に記憶された統計的音響モデル、及びカテゴリ別言語モデルの内で未知語のカテゴリに対応する言語モデルを使用して音声認識を行なうための手段とを含む。

【００２３】

カテゴリ別言語モデルの内で、未知語に付されていた、カテゴリを表すタグに対応する

ものが選択され、それを使用して音声認識が行なわれる。情報処理端末の利用者に特に関連する情報であって、かつサーバであるカテゴリに属すると推定された単語を、そのカテゴリの単語の言語モデルを使用して音声認識するので、音声認識の結果の精度がより高くなる。

【００２４】

圧縮手段は、複数個の所定の音響特徴量から予め組合された２つずつの音響特徴量の組合せの各々に対して予め準備されたコードブックを記憶するためのコードブック記憶手段と、フレーム化手段によりフレーム化されたフレームの各々について、複数個の所定の音響特徴量から予め組合された２つずつの音響特徴量の組合せの各々を、コードブックのうちで対応するものを用いて符号化するための符号化手段とを含んでもよい。送信するための手段は、フレーム化手段によりフレーム化されたフレームの各々について、符号化手段により得られた符号からなる符号列を送信するための手段を含んでもよい。

【００２５】

一実施の形態では、複数個の所定の音響特徴量は、各フレームの音声信号の第０次から第１２次のＭＦＣＣパラメータと、パワーとを含む。

【００２６】

好ましくは、音声認識機能付情報処理端末は、音声認識機能付情報処理端末で実行可能なアプリケーションプログラムにより、音声認識機能付情報処理端末の使用者に関連して集積された情報を記憶するための関連情報記憶手段と、関連情報記憶手段に記憶された情報を、カテゴリに分類するための分類手段と、分類手段により分類されたカテゴリごとに統計的言語モデルを作成することにより、複数個のカテゴリ別言語モデルを作成するための言語モデル作成手段とをさらに含む。

【００２７】

未知語のタグ付けがされた区間は、未知語のタグ付けがされた音節列であってもよい。

【発明の効果】

【００２８】

以上のようにこの発明によれば、分散型の音声認識を利用するシステムにおいて、サーバでは未知語であった固有名詞を情報処理端末で正しく認識できる可能性が大きくなる。そのために情報処理端末に多くのリソースを準備する必要はない。また、サーバでも未知語の音声認識を行なうためにリソースを肥大化させる必要がない。さらに、サーバからの音声認識結果に、未知語のカテゴリを示すタグを挿入することで、そのタグに対応した言語モデルを用いて情報処理端末で未知語に対して音声認識をし直すことができる。その結果、使用者にとって音声認識の精度が十分に高く、かつ情報処理端末側でも、音声認識を行なうサーバ側でも、リソースの極端な肥大化を防止できる音声認識機能付情報処理端末を提供できる。

【発明を実施するための最良の形態】

【００２９】

以下の説明において、全図を通じ、同一の部品には同一の参照番号を付してある。それらの名称及び機能も同一である。したがってそれらについての詳細な説明は繰返さない。

【００３０】

＜構成＞

図１に、本発明の第１の実施の形態に係る音声認識システム１０の概略構成を示す。図１を参照して、音声認識システム１０は、携帯型情報処理装置の一例であり、利用者の音声３０から音響特徴量３２を抽出する機能を持つ携帯電話機２０と、携帯電話機２０が抽出した音響特徴量３２を受けると、この音響特徴量３２に対して音声認識を行ない、認識結果のテキスト３４を携帯電話機２０に返信する機能を持つ音声認識サーバ２２とを含む。

【００３１】

音声認識サーバ２２は、音声認識結果に未知語が存在する場合には、その未知語を認識結果のテキスト内に音節列として挿入し、かつその音節列が未知語であることを示すタグ

と、その未知語が、予め分類されたいくつかのカテゴリの中のどのカテゴリに属するか示すタグとをその音節列に付与する機能を持つ。音声認識サーバ２２は、例えば周知の音声認識技術と、特許文献２に記載されているような未知語の認識技術及び非特許文献１に記載されているような、階層化言語モデルによるクラス推定とを組合せることにより実現できる。

【００３２】

再び図１を参照して、携帯電話機２０は、音声認識サーバ２２から送信されてくる認識結果のテキスト３４を受けると、この中に未知語が含まれている場合には、元の音声信号から得た音響特徴量の、その未知語部分に対して音声認識を行なって、その結果で未知語を置換する処理をして最終結果のテキスト３６を出力する。携帯電話機２０で行なわれるこの未知語の音声認識には、この携帯電話機２０の使用者に関連して各種アプリケーションプログラムによって集積された情報から作成された、カテゴリ別言語モデルのうち、未知語に付されたカテゴリのタグに対応したものが使用される。このカテゴリ別言語モデルは、この携帯電話機２０の利用者に特に関連した情報から作成されたものである。音声認識の結果として得られる固有名詞としては、この携帯電話機２０の利用者の友人、知人、よく利用する施設、学校などに関するものが大部分であるから、携帯電話機２０におけるこの音声認識での認識精度は高くなる。音声認識サーバ２２のように多数の利用者による音声を処理する必要はないので、携帯電話機２０の言語モデルに登録すべき単語は少なくて済む。

【００３３】

図２に、携帯電話機２０のうち、本発明に関連する部分の機能的構成を示す。図２を参照して、携帯電話機２０は、マイクロフォン５０と、マイクロフォン５０からの音声信号に対して所定の音響処理を行なって音声信号の特徴量を抽出し、さらにコードブックを用いて符号化して符号列を時系列で出力する音響信号処理部５４と、音響信号処理部５４が符号化時に使用するコードブックを記憶したコードブックメモリ５２と、音響信号処理部５４が出力する符号列を一時記憶するための送信バッファ５６と、送信バッファ５６に記憶された符号列をパケット化して音声認識サーバ２２に送信するための送信処理部５８とを含む。

【００３４】

携帯電話機２０はさらに、音響信号処理部５４が出力する符号列をフレームごとに順次記憶するための符号記憶部６０と、音声認識サーバ２２から音声認識結果のテキスト３４のパケットを受信するための受信処理部６２と、受信処理部６２により受信された音声認識結果のテキスト３４を一時記憶するための受信バッファ６４と、受信バッファ６４に記憶された音声認識結果のテキストに未知語が含まれていれば、その部分をコードブックを用いて復号し、改めて音声認識を行なって、未知語をその音声認識結果の単語で置換する未知語処理部７０と、未知語処理部７０が音声認識の際に利用する音響モデルを記憶する音響モデル記憶部６８及び複数のカテゴリ別言語モデルを記憶する言語モデル記憶部６６と、未知語処理部７０が出力するテキストを携帯電話機２０上で稼動している他のアプリケーションに渡す処理を行なうための出力部７２とを含む。

【００３５】

音響信号処理部５４は、マイクロフォン５０からの音声信号を、所定時間おきに所定時間長でフレーム化するためのフレーム化モジュール８０と、フレーム化モジュール８０から出力されるフレーム列の各々のフレームに対し、雑音抑圧及び特徴量抽出処理を行なって特徴量ベクトルを出力するための雑音抑圧・特徴量抽出部８２と、雑音抑圧・特徴量抽出部８２から出力される特徴量ベクトル列の各ベクトルに対し、コードブックメモリ５２に記憶されたコードブックを用いた符号化を行ない、符号列を送信バッファ５６及び符号記憶部６０に格納するための符号化処理部８４とを含む。

【００３６】

本実施の形態では、雑音抑圧・特徴量抽出部８２が抽出する音響特徴量は、ＭＦＣＣ（Ｍｅｌ　Ｆｒｅｑｕｅｎｃｙ　Ｃｅｐｓｔｒｕｍ　Ｃｏｅｆｆｉｃｉｅｎｔ）の第１次～

第１２次の係数、Ｃ０（第０次のＭＦＣＣ係数）、及び音声信号のパワーを含む。すなわち、特徴量ベクトルは１４次元である。

【００３７】

未知語処理部７０は、受信バッファ６４に記憶された、音声認識結果のテキスト列の中で未知語のタグが付された音節列（カタカナ列）を抽出し、符号記憶部６０に記憶された符号列の中から、この未知語に対応する符号列部分を切出す処理を行なう未知語切出処理部９０と、未知語切出処理部９０によって切出された符号列をコードブックメモリ５２に記憶されたコードブックを用いて音響特徴量列に戻し、言語モデル記憶部６６に記憶された複数個の言語モデルの内で、未知語に付されたカテゴリタグに対応するものと、音響モデル記憶部６８に記憶された音響モデルとを用いて音声認識処理を行ない、音声認識結果の単語を出力する未知語認識処理部９２と、受信バッファ６４に記憶されたテキストを読込み、未知語のタグが付された音節列を、未知語認識処理部９２により出力される音声認識後の単語で置換したテキストを出力するための未知語入替処理部９４とを含む。

【００３８】

未知語処理部７０はさらに、受信バッファ６４に記憶された音声認識後のテキストに、未知語のタグが付された音節列があるか否かを判定し、ある場合にはＴＲＵＥを、ない場合にはＦＡＬＳＥをとる判定結果信号を出力するための判定部９６と、受信バッファ６４に記憶されたテキストを受ける第１の入力と、未知語入替処理部９４の出力するテキストを受ける第２の入力とを有し、判定部９６から出力される判定信号がＴＲＵＥのときには未知語入替処理部９４からのテキストを、ＦＡＬＳＥのときには受信バッファ６４に格納されたテキストを、それぞれ選択して出力部７２に与えるための選択部９８とを含む。なお、判定部９６からの判定結果信号は、未知語切出処理部９０、未知語認識処理部９２及び未知語入替処理部９４にも与えられており、これら回路は判定結果信号がＴＲＵＥのときには動作し、ＦＡＬＳＥであるときには停止する。

【００３９】

図３は、図２に示す言語モデル記憶部６６に記憶されたカテゴリ別言語モデルを作成するためのカテゴリ別言語モデル作成部１００のブロック図である。図３を参照して、図２に示す携帯電話機２０には、住所録プログラムにより集積された住所録１０２と、メールプログラムにより集積されたメールアドレスＤＢ１０４と、ＧＰＳ（Ｇｌｏｂａｌ Ｐｏｓｉｔｉｏｎｉｎｇ Ｓｙｓｔｅｍ）などの地図ソフトで使用される地図データ１０６とが含まれる（いずれも図２では図示していない。）。カテゴリ別言語モデル作成部１００は、これらからカテゴリ別言語モデルを作成する。図３に示すように、本実施の形態では、カテゴリ別言語モデルとしては、施設名言語モデル（ＬＭ）と、日本人の姓に関する姓言語モデルと、日本人の名前に関する名前言語モデルと、日本人のニックネームに関するニックネーム言語モデルと、場所名に関する場所言語モデルとを有する。

【００４０】

図３を参照して、カテゴリ別言語モデル作成部１００は、住所録１０２、メールアドレスＤＢ１０４、及び地図データ１０６から言語モデル作成のためのデータを抽出し分類して、施設名データファイル１１２、姓データファイル１１４、名データファイル１１６、ニックネームデータファイル１１８、及び場所データファイル１２０等、カテゴリ別のファイルに出力するための抽出部１１０と、抽出部１１０により作成されたデータファイル１１２～１２０をそれぞれ用いて、施設名言語モデル、姓言語モデル、名言語モデル、ニックネーム言語モデル、場所言語モデルなど、カテゴリ別言語モデルを言語モデル記憶部６６に作成するための言語モデル作成部１２２とを含む。

【００４１】

住所録１０２などでは、予め所定の見出しとそれに対するデータという形でデータが集積されている。内部的には、これらデータは例えばＸＭＬ（ｅＸｔｅｎｄｅｄ Ｍａｒｋｕｐ Ｌａｎｇｕａｇｅ）などで保持されていることが多く、各タグをキーワードにして対応するデータを集めることにより、カテゴリ別のデータファイル１１２～１２０を集めることができる。

【００４２】
本実施の形態では、抽出部１１０を１本のコンピュータプログラムで実現し、住所録１０２、メールアドレスＤＢ１０４及び地図データ１０６から一度に言語モデル作成用のデータファイルを作成するが、アプリケーション別に抽出用のコンピュータプログラムを作成するようにしてもよい。
【００４３】
言語モデル記憶部６６に記憶されるカテゴリ言語モデルはいずれも同一のフォーマットである。データファイル１１２～１２０も同一フォーマットである。したがってここでも言語モデル作成部１２２は１本のコンピュータプログラムで実現できる。言語モデルの作成時に、入力ファイル名及び言語モデル名を引数として与えれば、言語モデル作成部１２２はそれら引数にしたがって別々のデータファイルからデータを読み、指定された言語モデルを作成する。
【００４４】
図４は、音声認識サーバ２２の機能ブロック図である。音声認識サーバ２２のハードウェア構成は公知であるため、その詳細については述べない。音声認識サーバ２２は、概略的には、任意の情報処理端末から音声認識の要求とともに音声認識の対象データである符号列をパケット形式で受信するための受信処理部１３０と、受信処理部１３０により受信されたパケットを一時的に記憶するための受信バッファ１３２と、図２に示すコードブックメモリ５２に記憶されたコードブックと同一のコードブックを記憶したコードブックメモリ１３４と、受信バッファ１３２に記憶されたパケット列から、音声認識の対象となる符号列を抽出し、コードブックメモリ１３４に記憶されたコードブックを用いて音響特徴量に戻す処理を行なうためのデコーダ１３６とを含む。
【００４５】
音声認識サーバ２２はさらに、音声認識に使用される、隠れマルコフモデル（ＨＭＭ）からなる音響モデルを記憶した音響モデル記憶部１３８と、予め所定のコーパスから作成された、クラス（品詞）別のバイグラムの統計的言語モデルを記憶するためのクラス言語モデル記憶部１４０と、予め所定のコーパスから作成された、単語トライグラムからなる統計的言語モデルを記憶するための単語言語モデル記憶部１４４と、携帯電話機２０に記憶されているカテゴリ別の言語モデルと同様、カテゴリ別に予め作成された複数個のカテゴリ別音節モデルを記憶するためのカテゴリ別音節モデル記憶部１４６とを含む。音節モデルとは、音節単位で前後の音節との文脈を考慮して作成された言語モデルである。同一の言語では、姓、名、地名、施設名など、単語が属するカテゴリによって音韻列の生起確率は異なっている。したがって、音声認識の過程で未知語に遭遇した場合、これら音節モデルを参照してその未知語の音節列が生ずる尤度を各モデルを使用して算出し、最も高い尤度を示す音節モデルのカテゴリをその未知語のカテゴリとすることができる。
【００４６】
クラス言語モデル記憶部１４０に記憶されたクラス言語モデル（バイグラム）とは、二つの連続する単語の品詞について、どのような順序付組合せがどの程度の確率で生ずるかを表す言語モデルである。
【００４７】
音声認識サーバ２２はさらに、音響モデル記憶部１３８に記憶された音響モデル、クラス言語モデル記憶部１４０に記憶されたクラスバイグラム、単語言語モデル記憶部１４４に記憶された単語トライグラムを用いて音声認識を行なってテキストに変換し、未知語はカタカナ列で出力するための音声認識処理部１４２を含む。音声認識処理部１４２は、未知語部分については、クラスバイグラムから算出される尤度と、音節モデルから算出される音節列の尤度とを乗算することにより、各音節列の候補の尤度を算出し、最も尤度が高い音節列を、未知語のタグを付して出力するとともに、その音節列を与える音節モデルのカテゴリを示すタグをその音節列に付与する。
【００４８】
なお、通常の音声認識処理と同様、音声認識処理部１４２が出力するテキストの各単語

、及び未知語を構成するカタカナ列を構成するカタカナ（音節）の各々には、元の音声信号における開始時間と終了時間とを示す情報が付加されている。

【００４９】

音声認識サーバ２２はさらに、音声認識処理部１４２の出力する時間情報付のテキストを一時記憶するための出力バッファ１４８と、出力バッファ１４８に記憶されたテキスト列を、音声認識要求を送信してきた情報処理端末に送信するための送信処理部１５０とを含む。図２に示す受信処理部６２が受信するのは、この送信処理部１５０により送信された、時間情報付のテキストである。

【００５０】

次に、図２に示す携帯電話機２０の音声認識機能のうち、未知語処理部７０の機能を実現するためのコンピュータプログラムのフローチャートを図５に示す。携帯電話機２０の音声認識機能のうち、音響信号処理部５４の部分については公知で、通常の分散処理型音声認識システムで採用されているものであるため、ここではその詳細については述べない。

【００５１】

図５を参照して、このプログラムは、音声認識結果の時間情報付のテキストを音声認識サーバ２２から受信するステップ１６０と、受信した時間情報付のテキストを受信バッファ６４に一時保存するステップ１６２と、受信したテキスト内に未知語のタグが付された部分があるか否かを判定し、判定結果に応じて制御の流れを分岐させるステップ１６４と、ステップ１６４において未知語タグが付された部分がないと判定されたことに応答して、音声認識サーバ２２から受信したテキストをそのままアプリケーションに渡して処理を終了するステップ１８０とを含む。

【００５２】

このプログラムはさらに、ステップ１６４において、テキスト内に未知語のタグが付された部分があると判定されたときに実行され、その未知語のタグが付された部分の時間情報に基づいて、符号記憶部６０に記憶された符号列の中で、その時間に対応する部分を切出す、すなわち読出す処理を実行するステップ１６６と、ステップ１６６に続き、その符号列をコードブックを用いて音響特徴量に伸長する処理を行なうステップ１６８と、ステップ１６８に続き、未知語部分に付されている、その未知語が属するカテゴリを示すタグに対応した言語モデルを言語モデル記憶部６６（図２参照）から選択するステップ１７０と、ステップ１７０で選択された言語モデルと、音響モデル記憶部６８（図２参照）に記憶された音響モデルとを使用して音声認識し、最尤の単語を出力するステップ１７２と、ステップ１７２で音声認識により得られた単語で、音声認識サーバ２２から受信したテキスト列の内の未知語タグが付された部分を置換するステップ１７４と、ステップ１７４で未知語部分が音声認識の結果で置換されたテキストをアプリケーションに渡して処理を終了するステップ１７６とを含む。

【００５３】

＜動作＞

以上、図１～図５に示した構成を有する音声認識システム１０は以下のように動作する。最初に、利用者が例えばメールプログラムを起動し、メールテキストを音声で入力する場合を想定する。利用者の音声はマイクロフォン５０により音声信号に変換され、フレーム化モジュール８０によって所定時間おきに所定長でフレーム化される。フレーム化モジュール８０が出力するフレーム列は雑音抑圧・特徴量抽出部８２に与えられる。

【００５４】

雑音抑圧・特徴量抽出部８２は、入力されるフレーム列の各々に対し、雑音抑圧処理を行なった後、先に述べたとおり、第１～第１２次のＭＦＣＣ係数、Ｃ０（第０次のＭＦＣＣ係数）、及びエネルギーを算出して１４次の音響特徴量ベクトルを生成し、符号化処理部８４に与える。

【００５５】

符号化処理部８４は、雑音抑圧・特徴量抽出部８２から与えられる音響特徴量ベクトル

の各々に対し、特徴量を示す要素を２つずつ組合せてコードブックメモリ５２に記憶されたコードブックのうちでその組合せに対応するものを用いて符号化し出力する。一つの音響特徴量ベクトルの要素は１４個であり、２つずつの組合せで符号化が行なわれるので、１４個の音響特徴量が全部で７個の符号からなる符号列に変換される。例えば１特徴量について８ビットが使用され、コードブックにより既定される符号が１６個であれば、全部で１６ビットの情報が４ビットに圧縮されることになる。これが７組あるので、全体では７×１６ビット＝１１２ビットの情報が４×７＝２８ビットに削減されることになる。

【００５６】

符号化処理部８４は、このように圧縮された符号列を送信バッファ５６及び符号記憶部６０に格納する。

【００５７】

送信処理部５８は、送信バッファ５６に２０フレーム分の符号列が格納されると、それらから１つのパケットを組立てて音声認識サーバ２２に送信する。

【００５８】

音声認識サーバ２２の受信処理部１３０は、受信したパケットを受信バッファ１３２に格納する。デコーダ１３６は、受信バッファ１３２に格納されたパケットから各フレーム毎の符号列を順次読出して、コードブックメモリ１３４に記憶されたコードブックを用いて音響特徴量に戻す。この場合、元の音響特徴量を完全に復元することはできないが、符号列をある程度の長さにしておけば、十分な精度で音声認識を行なうことができる。

【００５９】

音声認識処理部１４２は、デコーダ１３６が出力する各フレームの音響特徴量に基づいて、さらにＭＦＣＣ係数の差分（「Δ」と呼ぶ。）を算出して、１２次のＭＦＣＣとそれらの差分、Ｃ０，及びパワーからなる２６次元の音響特徴量ベクトルを生成する。音声認識処理部１４２は、このようにして生成された音響特徴量ベクトルの列に対し、音響モデル記憶部１３８に記憶された音響モデル、クラス言語モデル記憶部１４０に記憶されたクラス言語モデル、及び単語言語モデル記憶部１４４に記憶された単語言語モデルを用いて音声認識処理を実行する。音声認識処理部１４２はこの際、未知語部分については、クラス言語モデル記憶部１４０によって算出された尤度と、候補の音節列についてカテゴリ別音節モデル記憶部１４６によって算出された尤度とを乗算することによって候補の音節列の尤度を算出し、最尤の音節列を表すカタカナ列を未知語に対応する音声認識結果として出力する。音声認識処理部１４２は、この未知語部分には、未知語を示すタグと、さらに、最大尤度を与えた音節モデルのカテゴリを示すタグとを付して出力する。なおこのとき、音声認識処理部１４２は、各単語及び未知語部分の各音節について、その開始時間と終了時間とからなる時間情報を付す。

【００６０】

音声認識処理部１４２の音声認識結果は、未知語部分を含む場合も未知語部分を含まない場合も出力バッファ１４８（図４）に一旦格納される。

【００６１】

送信処理部１５０は、出力バッファ１４８に格納されたテキストを携帯電話機２０に送信する。

【００６２】

再び図２を参照して、受信処理部６２は、音声認識サーバ２２から音声認識結果のテキストを受信すると、時間情報とともに受信バッファ６４に格納する。未知語処理部７０は、受信バッファ６４に格納されたテキスト中に未知語を示すタグが存在するか否かを判定する。判定部９６は、もしも未知語を示すタグがなければ、ＦＡＬＳＥの判定結果信号を出力する。その結果、未知語切出処理部９０、未知語認識処理部９２、及び未知語入替処理部９４は動作せず、選択部９８は判定結果信号がＦＡＬＳＥであるため、第１の入力に与えられているテキスト、すなわち受信バッファ６４に記憶されている音声認識結果のテキストを選択して出力部７２に与える。出力部７２は、携帯電話機２０上で動作しているアプリケーション（現在の説明ではメールアプリケーション）にこのテキストを渡す。ア

プリケーションはこのテキストを、キーパッドから入力されたものと同様の入力として取り扱う。

【００６３】

もしも受信バッファ６４に記憶された音声認識結果のテキスト中に、未知語を示すタグが付された部分があれば、判定部９６は判定結果信号をＴＲＵＥとする。未知語切出処理部９０はこの判定結果信号に応答して、受信バッファ６４に記憶されたテキストの中の、未知語部分の開始時間及び終了時間を参照して、対応する符号列を符号記憶部６０から読出し、未知語認識処理部９２に与える。

【００６４】

未知語認識処理部９２は、未知語切出処理部９０から与えられた符号列の各々の符号を、コードブックメモリ５２に記憶されたコードブックを使用して音響特徴量ベクトルに伸長し、符号列に戻す。すなわち、未知語認識処理部９２は、圧縮時（符号化時）に対応する伸長アルゴリズムを用いて音響特徴量ベクトルを復元する。また、未知語認識処理部９２もサーバと同様にＭＦＣＣのデルタを算出する。ただし、コードブックを用いているため、ここでの復元は完全な復元ではない。

【００６５】

未知語認識処理部９２はさらに、受信バッファ６４中の未知語部分に付されている、カテゴリを現すタグを読出し、言語モデル記憶部６６に記憶されているカテゴリ別言語モデルのうちで、タグに対応するもの選択する。未知語認識処理部９２は、このようにして選択されたカテゴリ別言語モデルと、音響モデル記憶部６８に記憶された音響モデルとを使用して未知語の音声認識を行ない、認識結果の単語を未知語入替処理部９４に与える。未知語認識処理部９２での音声認識では、この携帯電話機２０の利用者に特に関連して、各種アプリケーションから抽出された固有名詞が音声認識結果の単語の候補となる。その結果、利用者が発話した確率の高い固有名詞が未知語の音声認識結果として得られる可能性が大きくなる。

【００６６】

未知語入替処理部９４は、受信バッファ６４に記憶されたテキストのうち、未知語のタグが付された音節列を、未知語認識処理部９２による音声認識の結果得られた単語で置換し、選択部９８の第２の入力に与える。選択部９８は、判定部９６からの判定結果信号がＴＲＵＥであるため、未知語入替処理部９４から与えられたテキストを選択し、出力部７２に与える。出力部７２にテキストが与えられた後の携帯電話機２０の動作は、音声認識サーバ２２からの音声認識結果のテキストに未知語が含まれていない場合と同様である。

【００６７】

＜例＞

図６に、この実施の形態に係る音声認識システム１０による音声認識の例を模式的に示す。図６を参照して、「私の名前は松田です」という音声に対する音声認識処理が携帯電話機２０で実行されるものとする。この携帯電話機２０がこの音声の符号列を音声認識サーバ２２に送信した後、音声認識サーバ２２から受信したテキスト２００が「私の名前はマツウダです」であったものとする。このテキストでは、本来は「松田」であった部分が、サーバでの音声認識では未知語として認識されている。すると、音声認識サーバ２２から送信されてきたテキスト２００のうち、「マツウダ」という音節列２０４の部分には、未知語を示すタグ２０６と、そのカテゴリとして日本人の「姓」を示すタグ２０８とが付されている。

【００６８】

携帯電話機２０では、符号記憶部６０に記憶されている符号列２０２のうち、未知語を示すタグ２０６が付されている音節列「マツウダ」に対応する部分符号列２１０を切出し、部分符号列２１０をコードブックを参照して伸長することで音響特徴量に戻し、未知語認識処理部９２で行なわれる音声認識の入力とする。

【００６９】

一方、「姓」を示すタグ２０８に対応するカテゴリ言語モデル、具体的には姓言語モデ

ル２１４が音声認識における言語モデルとして選択される。この姓言語モデル２１４には、「マツウダ」という姓はなく、例えば「松井」、「松田」、「松山」等という姓が存在しているものとすると、音声認識の結果、「マツウダ」ではなく正しい「松田」という単語２１２が選択される可能性が高い。

【００７０】

このように携帯電話機２０での音声認識処理で正しい固有名詞が選択される可能性が高いのは、この携帯電話機２０の使用者に特に関連した固有名詞のみを主に集め、それらをさらにカテゴリに分類してカテゴリ別言語モデルを作成しているためである。すなわち、使用者に関連のない固有名詞などが言語モデル中に含まれないため、使用者の発話に含まれる固有名詞に関する音声認識率が高くなる。また、音声認識を行なうために必要なリソースの量も少なくて済むという効果がある。

【００７１】

図７は、上記実施の形態に係る携帯電話機２０のハードウェア構成をブロック図形式で示す。図７を参照して、携帯電話機２０は、スピーカ２３６と、図２にも示したマイクロフォン５０と、液晶表示装置（ＬＣＤ）２３８と、テンキー及び特殊キーなどを含むキーパッド２４０と、アンテナ２３２と、着信及びアラームなどを振動により利用者に報知するための振動部２４２と、着信及びアラームなどを音声により利用者に報知するためのリンガ２４６と、携帯電話機２０の初期設定値、カテゴリ別言語モデル、音響モデル、及び種々のアプリケーションプログラム等を記憶するための不揮発性で書換可能なメモリ２４４と、スピーカ２３６、マイクロフォン５０、ＬＣＤ２３８、アンテナ２３２、振動部２４２、リンガ２４６及びメモリ２４４を用い、携帯電話機としての機能と、複数のアプリケーションを起動し、それらの出力をＬＣＤ２３８の表示面上に表示したり、キーパッド２４０からのユーザ入力を受けたりする機能とを実現するための制御回路２３０とを含む。

【００７２】

制御回路２３０は、アンテナ２３２を介して基地局から受信した信号に基づき、他の携帯通信端末からの着信を検出して着信検出信号を出力するための着信信号検出部２７０と、回線制御信号に応答して、アンテナ２３２を介した通信回線のオン／オフを制御するための回線閉結部２６８と、回線閉結部２６８及びアンテナ２３２を介して基地局との間で授受する信号の強度を制御するためのＲＦ（Ｒａｄｉｏ　Ｆｒｅｑｕｅｎｃｙ）処理部２６４と、基地局との信号の授受を安全に行なうために、ＲＦ処理部２６４に与える信号及びＲＦ処理部２６４を介して受ける信号に所定の信号処理を施すためのベースバンド処理部２６２と、ＤＡコンバータ及びＡＤコンバータを有し、マイクロフォン５０及びスピーカ２３６を介した音声の入出力を行なうためのオーディオインタフェース（オーディオＩ／Ｆ）２６０と、オーディオＩ／Ｆ２６０、ベースバンド処理部２６２、ＲＦ処理部２６４、回線閉結部２６８、ＬＣＤ２３８、振動部２４２、及びリンガ２４６を制御することにより、ユーザからの要求に応じて発呼したり、着呼を処理したりして、ユーザと他の携帯通信端末との間の音声通信を行なったり、文字通信を行なったり、ユーザの入力する文字列に対する処理を行なったりするための通信制御部２７２とを含む。

【００７３】

通信制御部２７２の機能は、実質的にはプロセッサとソフトウェアとにより実現される。ソフトウェアは本実施の形態ではメモリ２４４に記憶されていて、適宜通信制御部２７２内の図示しないメモリに読出され、実行される。本実施の形態では、詳細は説明しないが、メモリ２４４の内容を書き換えることが可能であり、それによって携帯電話機２０による種々の機能のアップグレード及び追加を行なうことができる。通信制御部２７２はまた、本実施の形態に係る携帯電話機２０の音声認識のためのプログラムを実行する。

【００７４】

以上のように本実施の形態に係る音声認識システム１０によれば、携帯電話機２０では音声認識の前処理に相当する特徴量の抽出が行なわれる。得られた特徴量ベクトルはコードブックを用いて符号化されて記憶されるとともに、サーバ２２に送信される。音声認識

サーバ２２は、この符号列を同じコードブックを用いて特徴量に戻した上で、音声認識サーバ２２に準備された豊富なリソースを使用して音声認識を行なう。音声認識の処理中に未知語に遭遇すると、音声認識サーバ２２は、その未知語を構成する音節列中の音節の遷移と予め準備されたカテゴリ別の音節モデルとに基づき、その未知語がどのカテゴリに属するかを推定し、未知語部分に未知語を示すタグとカテゴリを示すタグとを付して音声認識結果のテキスト中に挿入する。音声認識サーバ２２は、音声認識結果のテキストを携帯電話機２０に送信する。

【００７５】

携帯電話機２０では、このテキスト中に未知語があった場合、記憶されていた符号列の内、対応する部分を読出して特徴量に戻して音声認識を行なう。この音声認識では、言語モデルとして未知語に付されていたカテゴリに対応するカテゴリ別言語モデルが使用される。

【００７６】

携帯電話機２０において作成されたカテゴリ別言語モデルは、特にこの携帯電話機２０の使用者に関連する固有名詞から作成されている。その結果、音声認識サーバ２２では未知語として認識された単語でも、携帯電話機２０では利用者に特に関連する固有名詞として正しく認識される可能性が大きい。また、こうして言語モデルは、利用者に関連してアプリケーションによって集積された情報から作成されるものであり、その量が際限なく大きくなる可能性は極めて低い。そのため、携帯電話機２０に準備すべきリソースの量が際限なく大きくなるという心配もない。

【００７７】

その結果、分散型の音声認識を利用する情報処理端末であって、使用者にとって音声認識の精度が十分に高く、かつ音声認識を行なうサーバ側のリソースの極端な肥大化を防止できる音声認識機能付情報処理端末を提供できる。

【００７８】

なお、図示していないが音声認識サーバ２２側では、音声認識サービスを携帯電話機２０に対して提供するにあたって、利用者ごと（または携帯電話機２０ごと）に課金処理を行なうことが可能であることはいうまでもない。

【００７９】

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および範囲内でのすべての変更を含む。

【図面の簡単な説明】

【００８０】

【図１】本発明の一実施の形態に係る音声認識システム１０におけるデータの流れの概略を示す図である。

【図２】音声認識システム１０で使用される携帯電話機２０の機能ブロック図である。

【図３】カテゴリ別言語モデル作成部１００の機能ブロック図である。

【図４】音声認識サーバ２２の機能ブロック図である。

【図５】携帯電話機２０において、音声認識サーバ２２から音声認識結果を受けた後の未知語の音声認識及び入替処理を実現するプログラムのフローチャートである。

【図６】音声認識システム１０による音声認識の過程の一例を示す図である。

【図７】携帯電話機２０のハードウェアブロック図である。

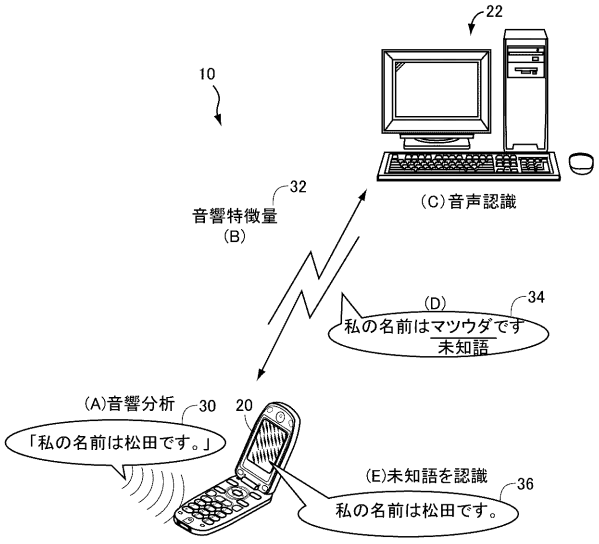
【符号の説明】

【００８１】

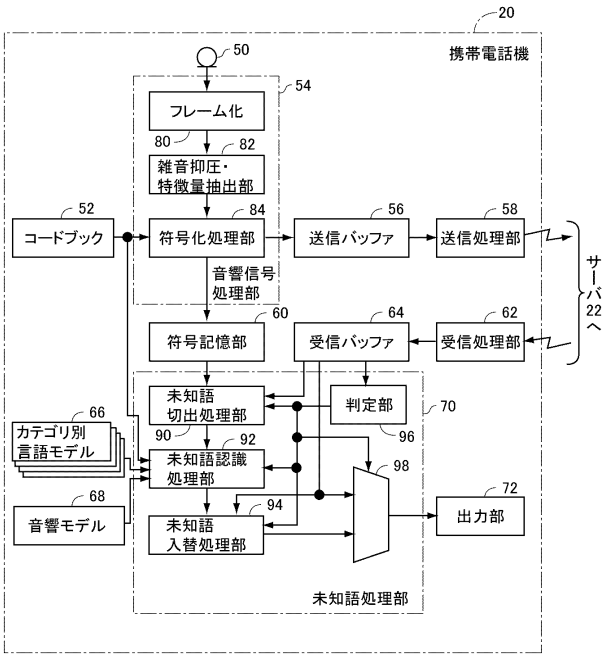
１０　音声認識システム
２０　携帯電話機
２２　音声認識サーバ
５２　コードブックメモリ

５４　音響信号処理部
６０　符号記憶部
６６　言語モデル記憶部
６８　音響モデル記憶部
７０　未知語処理部
９０　未知語切出処理部
９２　未知語認識処理部
９４　未知語入替処理部
９６　判定部

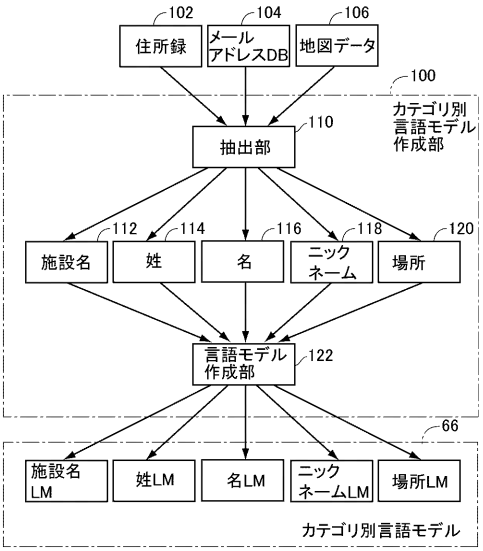
【図１】



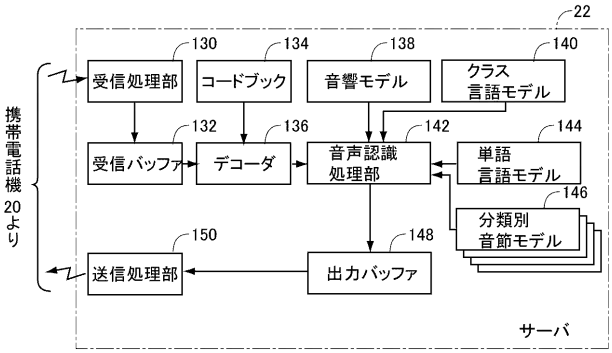
【図２】



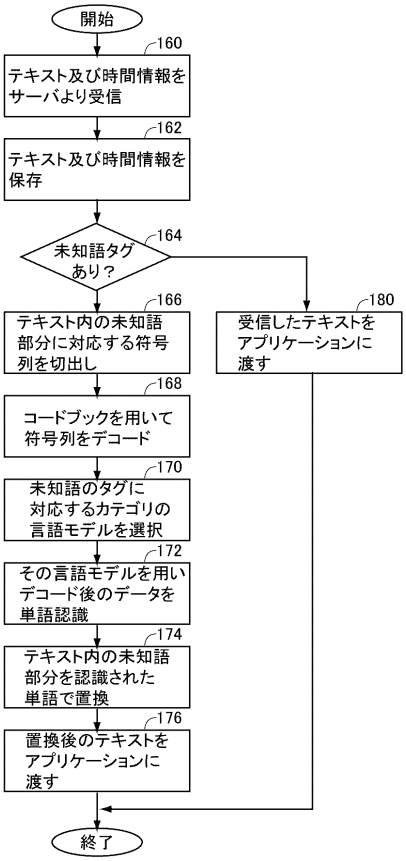
【図３】


102
104
106
住所録
メール
アドレスDB
地図データ
100
カテゴリ別
言語モデル
作成部
110
抽出部
112
114
116
118
120
施設名
姓
名
ニック
ネーム
場所
言語モデル
作成部
122
66
施設名
LM
姓LM
名LM
ニック
ネームLM
場所LM
カテゴリ別言語モデル

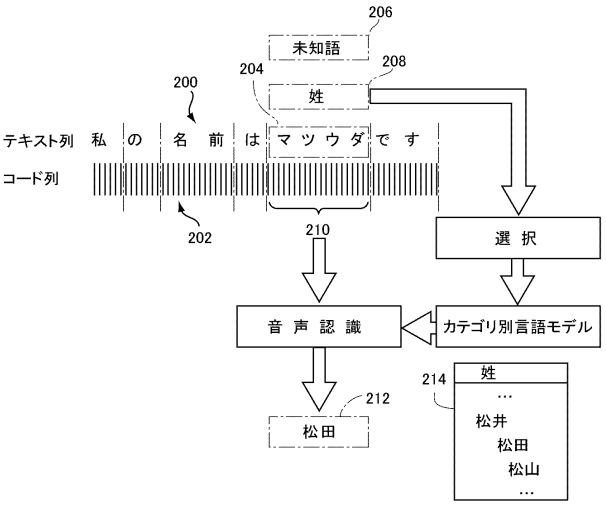
【図４】


22
携帯電話機20より
130
受信処理部
134
コードブック
138
音響モデル
140
クラス
言語モデル
132
受信バッファ
136
デコーダ
142
音声認識
処理部
144
単語
言語モデル
146
分類別
音節モデル
150
送信処理部
148
出力バッファ
サーバ

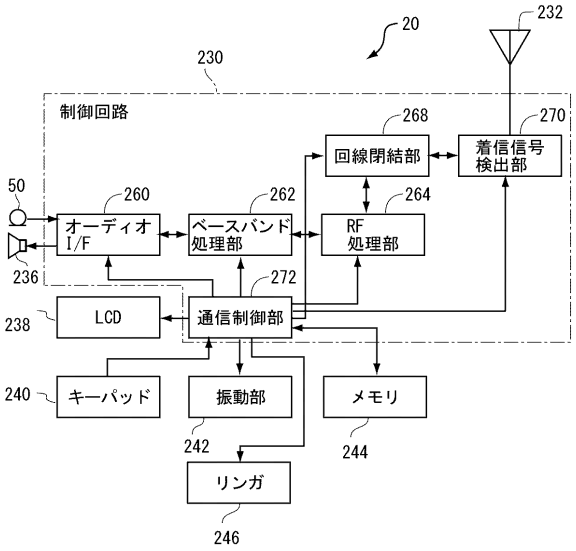
【図５】


開始
160
テキスト及び時間情報を
サーバより受信
162
テキスト及び時間情報を
保存
164
未知語タグ
あり？
166
テキスト内の未知語
部分に対応する符号
列を切出し
180
受信したテキストを
アプリケーションに
渡す
168
コードブックを用いて
符号列をデコード
170
未知語のタグに
対応するカテゴリの
言語モデルを選択
172
その言語モデルを用い
デコード後のデータを
単語認識
174
テキスト内の未知語
部分を認識された
単語で置換
176
置換後のテキストを
アプリケーションに
渡す
終了

【図６】


206
未知語
204
208
姓
200
テキスト列
私
の
名
前
は
マ
ツ
ウ
ダ
で
す
コード列
202
210
選 択
音 声 認 識
カテゴリ別言語モデル
214
姓
…
松井
松田
松山
…
212
松田

【図７】


20
232
230
制御回路
268
270
回線閉結部
着信信号
検出部
50
260
262
264
オーディオ
I/F
ベースバンド
処理部
RF
処理部
236
272
238
LCD
通信制御部
240
キーパッド
振動部
メモリ
242
244
リンガ
246

フロントページの続き

(72)発明者　山本　博史
京都府相楽郡精華町光台二丁目２番地２　株式会社国際電気通信基礎技術研究所内
(72)発明者　林　輝昭
京都府相楽郡精華町光台二丁目２番地２　株式会社国際電気通信基礎技術研究所内

審査官　間宮　嘉誉

(56)参考文献　特開２００８－９１５３（ＪＰ，Ａ）
特開２００３－１８６４９４（ＪＰ，Ａ）
特開平４－１８８２００（ＪＰ，Ａ）
特開２００４－３０９５２３（ＪＰ，Ａ）
特開２００７－２１３１０９（ＪＰ，Ａ）
国際公開第２００５／１２２１４４（WO，Ａ１）
特開平４－４９７１９（ＪＰ，Ａ）
特開２００１－１７５２８６（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１５／００－１５／２８