

(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2012-137777

(P2012-137777A)

(43) 公開日 平成24年7月19日(2012.7.19)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 0 L 15/04 (2006.01)	G 1 0 L 15/04 3 0 0 Z	5 D 0 1 5
G 1 0 L 15/28 (2006.01)	G 1 0 L 15/28 2 1 0 A	
G 1 0 L 15/22 (2006.01)	G 1 0 L 15/22 3 0 0 Z	

審査請求 有 請求項の数 19 O L 外国語出願 (全 36 頁)

(21) 出願番号 特願2012-60252 (P2012-60252) (22) 出願日 平成24年3月16日 (2012. 3. 16) (62) 分割の表示 特願2001-528975 (P2001-528975) の分割 原出願日 平成12年10月4日 (2000. 10. 4) (31) 優先権主張番号 09/412, 202 (32) 優先日 平成11年10月5日 (1999. 10. 5) (33) 優先権主張国 米国 (US) (特許庁注：以下のものは登録商標) 1. B L U E T O O T H	(71) 出願人 500043574 リサーチ イン モーション リミテッド Research In Motion Limited カナダ国 エヌ2エル 3ダブリュー8 オンタリオ, ウォータールー, フィリ ップ ストリート 295 295 Phillip Street, Waterloo, Ontario N2L 3W8 Canada (74) 代理人 100107489 弁理士 大塩 竹志 (72) 発明者 アイラ・エイ・ガーソン アメリカ合衆国、イリノイ60143、ア イタスカ、パーク・ブルーバード500 Fターム(参考) 5D015 DD00
---	--

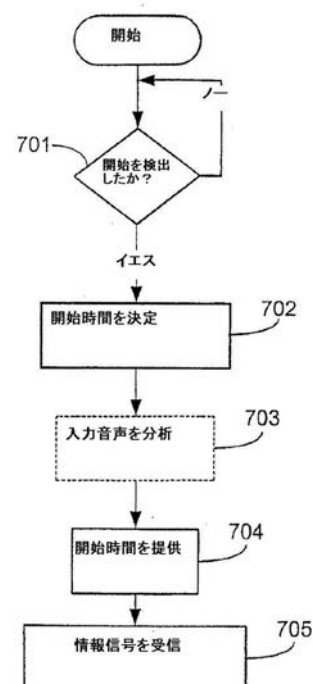
(54) 【発明の名称】 出力オーディオ信号が生ずる間に入力音声信号を処理する方法および装置

(57) 【要約】

【課題】入力音声信号を処理する方法および装置を提供すること。

入力音声信号の開始は、決定される(701)際の出力オーディオ信号に対する、出力オーディオ信号と入力開始時間との生成の間に検出される。入力開始時間は、次に、入力音声信号に応答するのに使用されるために提供される(704)。入力音声信号が、出力オーディオ信号が生ずる間に検出されるとき、出力オーディオ信号の識別は、入力音声信号に応答するのに使用されるために提供される。データおよび/または制御信号を備えている情報の信号(705)は、少なくとも提供されるコンテキスト上の情報、すなわち、入力開始時間および/または出力オーディオ信号の識別に応じて提供される。本発明は、基礎をなす通信システムの遅延特性にかかわらず、出力オーディオ信号に対する入力音声信号のコンテキストを精密に確立する。

【選択図】 図7



【特許請求の範囲】

【請求項 1】

無線通信加入者ユニットにおいて、入力音声信号を処理する方法であって、前記無線通信加入者ユニットは、スピーカとマイクロフォンとを含み、前記スピーカは、出力オーディオ信号を提供し、前記マイクロフォンは、入力音声信号を提供し、

前記方法は、

前記出力オーディオ信号を提示する間に前記入力音声信号の開始を検出することと、

前記出力オーディオ信号に対して、前記入力音声信号の開始の入力開始時間を決定することと、

前記入力音声信号を分析することにより、パラメータで表示された音声信号を提供することと、

前記入力開始時間を制御パラメータとして音声認識サーバに提供することと、

前記パラメータで表示された音声信号を前記音声認識サーバに提供することと、

前記入力開始時間および前記パラメータで表示された音声信号に少なくとも基づいて、前記音声認識サーバから少なくとも 1 つの情報信号を受信することと、

前記入力音声信号に応答するように、前記少なくとも 1 つの情報信号に基づいて動作することと

を含む、方法。

【請求項 2】

前記入力開始時間を決定することというステップは、

前記出力オーディオ信号の開始より遅く、かつ、その後の出力オーディオ信号の開始よりも早く、前記入力開始時間を決定することをさらに含む、請求項 1 に記載の方法。

【請求項 3】

前記入力開始時間は、前記出力オーディオ信号のリアルタイム参照に対するタイムスタンプと、前記出力オーディオ信号が複数のオーディオサンプルとして再構成および／または符号化された場合には、前記出力オーディオ信号の開始サンプルに対するサンプルインデックスと、前記出力オーディオ信号のフレームコンテキストに対するフレームインデックスとのうちのいずれか 1 つである、請求項 1 に記載の方法。

【請求項 4】

無線通信加入者ユニットにおいて、入力音声信号を処理する方法であって、前記無線通信加入者ユニットは、スピーカとマイクロフォンとを含み、前記スピーカは、出力オーディオ信号を提供し、前記マイクロフォンは、入力音声信号を提供し、

前記方法は、

前記出力オーディオ信号を提示する間に前記入力音声信号を検出することと、

前記出力オーディオ信号を識別する識別データを決定することと、

前記入力音声信号を分析することにより、パラメータで表示された音声信号を提供することと、

前記識別データを制御パラメータとして音声認識サーバに提供することと、

前記パラメータで表示された音声信号を前記音声認識サーバに提供することと、

前記識別データおよび前記パラメータで表示された音声信号に少なくとも基づいて、前記音声認識サーバから少なくとも 1 つの情報信号を受信することと、

前記入力音声信号に応答するように、前記少なくとも 1 つの情報信号に基づいて動作することと

を含む、方法。

【請求項 5】

前記出力オーディオ信号は、出力音声信号を含む、請求項 4 に記載の方法。

【請求項 6】

1 つ以上の加入者ユニットと無線通信する音声認識サーバから情報信号を提供する方法であって、

前記方法は、

10

20

30

40

50

前記加入者ユニットにおいて、出力オーディオ信号を提示させることと、
前記加入者ユニットにおける前記出力オーディオ信号に対する入力音声信号の開始に
応する入力開始時間を前記加入者ユニットから少なくとも受信することと、
前記入力音声信号に対応するパラメータで表示された音声信号を受信することと、
前記入力開始時間および前記パラメータで表示された音声信号に少なくとも基づいて、
前記情報信号を生成することと、
前記情報信号を前記加入者ユニットに提供することと
を含み、前記情報信号は、前記入力音声信号に応答するように、前記加入者ユニットの
動作を制御する、方法。

【請求項 7】

10

前記入力開始時間は、前記出力オーディオ信号のリアルタイム参照に対するタイムスタ
ンプと、前記出力オーディオ信号が複数のオーディオサンプルとして再構成および／また
は符号化された場合には、前記出力オーディオ信号の開始サンプルに対するサンプルイン
デックスと、前記出力オーディオ信号のフレームに対するフレームインデックスとのうち
のいずれか 1 つである、請求項 6 に記載の方法。

【請求項 8】

前記出力オーディオ信号を提示させることは、
前記加入者ユニットに音声信号を提供することをさらに含む、請求項 6 に記載の方法。

【請求項 9】

1 つ以上の加入者ユニットと無線通信する音声認識サーバから情報信号を提供する方法
であって、

20

前記方法は、

前記加入者ユニットにおいて出力オーディオ信号を提示させることであって、前記出力
オーディオ信号は、識別データによって識別される、ことと、

前記出力オーディオ信号を提示する間に入力音声信号が前記加入者ユニットにおいて検
出された場合には、前記加入者ユニットから少なくとも前記識別データを受信することと

、
前記入力音声信号に対応するパラメータで表示された音声信号を受信することと、
前記識別データおよび前記パラメータで表示された音声信号に少なくとも基づいて、前
記情報信号を生成することと、

30

前記情報信号を前記加入者ユニットに提供することと

を含み、前記情報信号は、前記入力音声信号に応答するように、前記加入者ユニットの
動作を制御する、方法。

【請求項 10】

前記出力オーディオ信号を提示させることは、

前記加入者ユニットに音声信号を提供することをさらに含む、請求項 9 に記載の方法。

【請求項 11】

音声認識サーバを含む無線通信システムと無線通信する加入者ユニットであって、前記
加入者ユニットは、

出力オーディオ信号を提供するように構成されたスピーカと、

40

入力音声信号を提供するように構成されたマイクロフォンと、

無線トランシーバと、

プロセッサと

を含み、

前記プロセッサは、前記無線トランシーバと協働することにより、

前記入力音声信号の開始を検出することと、

前記出力オーディオ信号に対して、前記入力音声信号の開始の入力開始時間を決定す
ることと、

前記入力音声信号を分析することにより、パラメータで表示された音声信号を提供す
ることと、

50

前記入力開始時間を前記音声認識サーバに提供することと、
前記パラメータで表示された音声信号を前記音声認識サーバに提供することと、
前記入力開始時間および前記パラメータで表示された音声信号に少なくとも基づいて
、前記音声認識サーバから、少なくとも1つの制御信号を受信することと、
前記入力音声信号に応答するように、前記少なくとも1つの制御信号に基づいて動作
することと
を行うように構成されている、加入者ユニット。

【請求項12】

前記プロセッサは、前記出力オーディオ信号の開始より遅く、かつ、その後の出力オーディオ信号の開始よりも早く、前記入力開始時間を決定するように構成されている、請求項11に記載の加入者ユニット。

【請求項13】

前記入力開始時間は、前記出力オーディオ信号のリアルタイム参照に対するタイムスタンプと、前記出力オーディオ信号が複数のオーディオサンプルとして再構成および／または符号化された場合には、前記出力オーディオ信号の開始サンプルに対するサンプルインデックスと、前記出力オーディオ信号のフレームコンテキストに対するフレームインデックスとのうちのいずれか1つである、請求項11に記載の加入者ユニット。

【請求項14】

音声認識サーバを含む無線通信システムと無線通信する加入者ユニットであって、前記加入者ユニットは、

出力オーディオ信号を提供するように構成されたスピーカと、
入力音声信号を提供するように構成されたマイクロフォンと、
無線トランシーバと、
プロセッサと
を含み、

前記プロセッサは、前記無線トランシーバと協働することにより、

前記入力音声信号の開始を検出することと、

前記出力オーディオ信号を識別する識別データを決定することと、

前記入力音声信号を分析することにより、パラメータで表示された音声信号を提供することと、

前記識別データを前記音声認識サーバに提供することと、

前記パラメータで表示された音声信号を前記音声認識サーバに提供することと、

前記識別データおよび前記パラメータで表示された音声信号に少なくとも基づいて、前記音声認識サーバから少なくとも1つの制御信号を受信することと、

前記入力音声信号に応答するように、前記少なくとも1つの制御信号に基づいて動作することと

を行うように構成されている、加入者ユニット。

【請求項15】

前記出力オーディオ信号は、出力音声信号を含む、請求項14に記載の加入者ユニット。

【請求項16】

1つ以上の加入者ユニットと無線通信する無線通信システムの音声認識サーバであって、

前記音声認識サーバは、プロセッサを含み、

前記プロセッサは、

入力音声信号の開始を検出することと、

出力オーディオ信号に対して、前記入力音声信号の開始の入力開始時間を決定することと、

前記入力音声信号を分析することにより、パラメータで表示された音声信号を提供することと、

10

20

30

40

50

前記入力開始時間を前記音声認識サーバに提供することと、
前記パラメータで表示された音声信号を前記音声認識サーバに提供することと、
前記入力開始時間および前記パラメータで表示された音声信号に少なくとも基づいて
、前記音声認識サーバから少なくとも1つの制御信号を受信することと、
前記入力音声信号に応答するように、前記少なくとも1つの制御信号に基づいて動作
することと
を行うように構成されている、音声認識サーバ。

【請求項17】

前記入力開始時間は、前記出力オーディオ信号のリアルタイム参照に対するタイムスタ
ンプと、前記出力オーディオ信号が複数のオーディオサンプルとして再構成および／また
は符号化された場合には、前記出力オーディオ信号の開始サンプルに対するサンプルイン
デックスと、前記出力オーディオ信号のフレームコンテキストに対するフレームインデッ
クスとのうちのいずれか1つである、請求項16に記載の音声認識サーバ。

【請求項18】

1つ以上の加入者ユニットと無線通信する無線通信システムの音声認識サーバであって
、
前記音声認識サーバは、プロセッサを含み、
前記プロセッサは、
入力音声信号の開始を検出することと、
出力音声信号を識別する識別データを決定することと、
前記入力音声信号を分析することにより、パラメータで表示された音声信号を提供す
ることと、
前記識別データを前記音声認識サーバに提供することと、
前記パラメータで表示された音声信号を前記音声認識サーバに提供することと、
前記識別データおよび前記パラメータで表示された音声信号に少なくとも基づいて、
前記音声認識サーバから少なくとも1つの制御信号を受信することと、
前記入力音声信号に応答するように、前記少なくとも1つの制御信号に基づいて動作
することと
を行うように構成されている、音声認識サーバ。

【請求項19】

前記出力オーディオ信号は、出力音声信号を含む、請求項18に記載の音声認識サーバ

【発明の詳細な説明】

【技術分野】

【0001】

(発明の属する技術分野)

本発明は、一般に、音声認識を組み込んでいる通信システムに関し、特に、出力オーディオ信号が生ずる間の入力音声信号を“バージイン”(barge-in)処理する方法および装置に関する。

【背景技術】

【0002】

(従来技術)

音声認識システムは、特に電話システムに関して一般的に公知の技術である。米国特許4,914,692、5,475,791、5,708,704と、また5,765,130は、音声認識システムを組み込んだ電話網を例示している。各システムの共通する特徴は、音声認識エレメント(即ち、音声認識を実施するデバイス)は、加入者の通信デバイス(即ち、ユーザーの電話)とは逆に、一般的に、電話網の構成の中心に位置されている。典型的には、音声の合成と音声認識エレメントとの組み合わせは、電話網あるいはインフラストラクチャ内で展開されている。呼者は、そのシステムにアクセスし、音声合成エレメントを介して、合成された音声の形での情報のプロンプトあるいはクエリーの提供を

10

20

30

40

50

受けることができる。呼者は、一般的に、合成された音声に対して口頭による応答を行い、音声認識エレメントは、呼者にさらなるサービスを提供するために、呼者の口頭による応答を処理する。

【発明の概要】

【発明が解決しようとする課題】

【0003】

人間性およびいくつかの音声合成／認識システムの設計を所与のものとするれば、発呼者によって提供される口頭の応答は、出力オーディオ信号、たとえば、合成された音声プロンプトが生ずる間に生ずることが多い。このような発生セグメントの処理は、“バージイン”処理と呼ぶことが多い。米国特許第4,914,692号、第5,155,760号、第5,475,791号；第5,708,704号および第5,765,130号すべては、バージイン処理のための技術について記述している。一般に、これらの特許のそれぞれに記述されている技術は、バージイン処理の間のエコーキャンセルの必要性と取り組んでいる。すなわち、合成された音声プロンプト（すなわち、出力オーディオ信号）が生ずる間、音声認識システムは、音声認識分析を効果的に行うために、ユーザによって提供されるあらゆる口頭の応答（すなわち、入力音声信号）に存在するプロンプトからの未処理の作為を説明する必要がある。したがって、これらの従来技術技術は、一般に、バージイン処理の間の入力音声信号の質に関している。音声電話技術システムに見出される比較的小さな待ち時間、あるいは、遅延により、これらの従来技術の技術は、一般に、入力音声信号を特定の出力オーディオ信号や出力オーディオ信号内の特定のモーメントに関連づけるといったバージイン処理のコンテキスト決定について考慮されてない。

【0004】

従来技術のこの欠点は、無線システムに関していっそう著しい。従来技術の主要部が、前記で説明されたシステムに関連して存在してはいるが、無線通信システムへの音声認識システムの組み込みは、新しく開発されたばかりである。無線通信環境の中での音声認識の応用の標準化に努めるために、オーロラ・プロジェクトと呼ばれる欧州通信規格協会（ETSI）による取り組みが最近開始された。オーロラ・プロジェクトの目標は、音声認識システムの普及のためのグローバルな規格である。このプロジェクトは、クライアント／サーバ構成を確立することを提案している。この構成では、特徴抽出あるいはパラメータ化などのフロントエンドの音声認識処理が、加入者ユニット（例えば、セルラ電話のような携帯無線デバイス）で実施される。その後、フロントエンドにより提供されるデータは、サーバに転送され、バックエンドの音声認識処理が行われる。

【0005】

オーロラ・プロジェクトによって提案されるクライアントーサーバ構成が、普及される音声認識システムのための必要性和適切に取り組んでいることは予測される。とはいえ、バージイン処理が、オーロラ・プロジェクトによって、なされたとしても、取り組まれる方法は、この時には不明確である。これは、無線システムにおいて一般に直面する欠点と、このような欠点が、バージイン処理に関してあるという効果とにおけるより広い変化を所与のものとするれば特に懸念される。たとえば、ユーザの音声に基づく応答の処理が、音声認識プロセッサによって受けられた特定の時点に一部基づいて行われることは、まれではない。すなわち、一連の複数の連続しないプロンプトが提供される場合に、ユーザの応答が、所定の合成プロンプトの特定の部分の間に受信されたのか、あるいは、どのプロンプトの間に応答が受信されたかどうかによって、相違を生じうる。要するに、ユーザの応答のコンテキストは、ユーザの応答の情報の内容の認識と同じほど重要である。とはいえ、いくつかの無線システムの不特定の遅延特性は、このようなコンテキストを適切に決定するのに障害となる。したがって、特に、パケットデータ通信を利用するなどの不確実な、および／または広範に変化する遅延特性を有するシステムにおいて、出力オーディオ信号が生ずる間に入力音声信号のコンテキストを決定するための技術を提供することは有利なことである。

【課題を解決するための手段】

【0006】

本発明は、出力オーディオ信号が生ずる間に入力音声信号を処理するための技術を提供する。主として、無線通信システムに適用可能であるが、本発明の技術は、有益なことには、不確実な、および／または広範囲に変化する遅延特性を有するあらゆる通信システム、たとえば、Internet（インターネット）などのパケットデータシステムに適用されることができる。本発明の1つの実施の形態によれば、入力音声信号の開始は、出力オーディオ信号が生ずる間に検出され、そして、出力オーディオ信号に対して入力開始時間が、決定される。入力開始時間は、次に、入力音声信号に応答するのに使用されるために提供される。別の実施の形態において、出力オーディオ信号は、対応する識別を有している。入力音声信号が、出力オーディオ信号が生ずる間に検出されるとき、出力オーディオ信号の識別は、入力音声信号に応答するのに使用されるために提供される。データおよび／または制御信号を備えている情報の信号は、少なくとも提供されるコンテキスト情報、すなわち、入力開始時間および／または出力オーディオ信号の識別に応じて提供される。このように、本発明は、基礎をなす通信システムの遅延特性にかかわらず、出力オーディオ信号に対して入力音声信号のコンテキストを精密に確立するための技術を提供する。本発明は、例えば、以下を提供する。

10

（項目1） 入力音声信号の開始を検出するステップと、

出力オーディオ信号に対して、入力音声信号の開始の入力開始時間を決定するステップと、

入力音声信号に応答するのに使用される入力開始時間を提供するステップとを備える、出力オーディオ信号が生ずる間に入力音声信号を処理する方法。

20

（項目2） 入力開始時間が、出力オーディオ信号の一時的なコンテキストに対するタイムスタンプ、出力オーディオ信号のサンプルコンテキストに対するサンプルインデックス、および出力オーディオ信号のフレームコンテキストに対するフレームインデックスのいずれか1つを備える、項目1に記載の方法。

（項目3） 項目1に記載のステップを行うコンピュータ実行可能命令を有するコンピュータ読出し可能媒体。

（項目4） 入力音声信号を検出するステップと、

出力オーディオ信号に対応する識別を決定するステップと、

入力音声信号に応答するのに使用される識別を提供するステップとを備える、出力オーディオ信号が生ずる間に入力音声信号を処理する方法。

30

（項目5） 項目4に記載のステップを行うコンピュータ実行可能命令を有するコンピュータ読出し可能媒体。

（項目6） 音声認識サーバを備えるインフラストラクチャとの無線通信における加入者ユニットにおいて、加入者ユニットがスピーカおよびマイクロフォンを備え、スピーカが出力オーディオ信号を提供し、マイクロフォンが入力音声信号を提供し、入力音声信号を処理する方法であって、

出力音声信号が生ずる間に入力音声信号の開始を検出するステップと、

出力オーディオ信号に対して、入力音声信号の開始の入力開始時間を決定するステップと、

40

制御パラメータとして音声認識サーバに上記入力開始時間を提供するステップとを備える、入力音声信号を処理する方法。

（項目7） 少なくとも一部の入力開始時間に基づき、音声認識サーバから少なくとも1つの情報信号を受信するステップをさらに備える、項目6に記載の方法。

（項目8） オンセットマーカを決定するステップが、

出力オーディオ信号の開始より遅く、かつその後の出力オーディオ信号の開始よりも早い入力開始時間を決定するステップをさらに備える、項目6に記載の方法。

（項目9） 入力開始時間が、出力オーディオ信号の一時的なコンテキストに対するタイムスタンプ、出力オーディオ信号のサンプルコンテキストに対するサンプルインデックス、および出力オーディオ信号のフレームコンテキストに対するフレームインデックス

50

のいずれか 1 つである、項目 6 に記載の方法。

(項目 10) 出力オーディオ信号が、インフラストラクチャにより提供された音声信号を備える、項目 6 に記載の方法。

(項目 11) 出力オーディオ信号が、インフラストラクチャにより提供された制御シグナリングに応答して加入者ユニットにより合成された音声信号を備える、項目 6 に記載の方法。

(項目 12) パラメータで表示された音声信号を提供するために入力音声信号を分析するステップと、

パラメータで表示された音声信号を音声認識サーバに提供するステップと、

少なくとも一部の入力開始時間およびパラメータで表示された音声信号に基づき音声認識サーバから少なくとも 1 つの情報信号を受信するステップとをさらに備える、項目 6 に記載の方法。

(項目 13) 音声認識サーバを備えるインフラストラクチャとの無線通信における加入者ユニットにおいて、加入者ユニットがスピーカおよびマイクロフォンを備え、スピーカが出力オーディオ信号を提供し、マイクロフォンが入力音声信号を提供し、入力音声信号を処理する方法であって、

出力オーディオ信号が生ずる間に入力音声信号を検出するステップと、

出力オーディオ信号に対応する識別を決定するステップと、

制御パラメータとして音声認識サーバに識別を提供するステップとを備える、入力音声信号を処理する方法。

(項目 14) 少なくとも一部の識別に基づき、音声認識サーバから少なくとも 1 つの情報信号を受信するステップをさらに備える、項目 13 に記載の方法。

(項目 15) 出力オーディオ信号が、インフラストラクチャにより提供された音声信号を備える、項目 13 に記載の方法。

(項目 16) 出力オーディオ信号が、インフラストラクチャにより提供された制御シグナリングに応答して加入者ユニットにより合成された音声信号を備える、項目 13 に記載の方法。

(項目 17) パラメータで表示された音声信号を提供するために入力音声信号を分析するステップと、

パラメータで表示された音声信号を音声認識サーバに提供するステップと、

少なくとも一部の識別およびパラメータで表示された音声信号に基づき、音声認識サーバから少なくとも 1 つの情報信号を受信するステップとをさらに備える、項目 13 に記載の方法。

(項目 18) 1 つ以上の加入者ユニットと無線通信するインフラストラクチャの一部を形成する音声認識サーバにおいて、1 つ以上の加入者ユニットの加入者ユニットに情報信号を提供する方法であって、

加入者ユニットで出力オーディオ信号を生じさせるステップと、

加入者ユニットにおける出力オーディオ信号に対する入力音声信号の開始に対応する少なくとも入力開始時間を加入者ユニットから受信するステップと、

少なくとも一部の開始時間に応答して、加入者ユニットに情報信号を提供するステップとを備える、1 つ以上の加入者ユニットの加入者ユニットに情報信号を提供する方法。

(項目 19) 入力開始時間が、出力オーディオ信号の一時的なコンテキストに対するタイムスタンプ、出力オーディオ信号のサンプルコンテキストに対するサンプルインデックス、および出力オーディオ信号のフレームコンテキストに対するフレームインデックスのいずれか 1 つである、項目 18 に記載の方法。

(項目 20) 出力オーディオ信号を生じさせるステップが、

加入者ユニットに音声信号を提供するステップをさらに備える、項目 18 に記載の方法

。

(項目 21) 情報信号を提供するステップが、

加入者ユニットに情報信号を向けるステップをさらに備え、情報信号が、上記加入者ユ

10

20

30

40

50

ニットの動作を制御する、項目 18 に記載の方法。

(項目 22) 上記加入者ユニットが、少なくとも 1 つの装置に結合され、情報信号を提供するステップが、

少なくとも 1 つの装置に情報を向けるステップをさらに備え、情報信号が、少なくとも 1 つの装置の動作を制御する、項目 18 に記載の方法。

(項目 23) 出力オーディオ信号を生じさせるステップが、

加入者ユニットに制御シグナリングを提供するステップをさらに備え、制御シグナリングが、加入者ユニットに出力オーディオ信号として音声信号を合成させる、項目 18 に記載の方法。

(項目 24) 入力音声信号に対応するパラメータで表示された音声信号を受信するステップと、

少なくとも一部の開始時間およびパラメータで表示された音声信号に応答して、加入者ユニットに情報信号を提供するステップとをさらに備える、項目 18 に記載の方法。

(項目 25) 1 つ以上の加入者ユニットと無線通信するインフラストラクチャの一部を形成する音声認識サーバにおいて、1 つ以上の加入者ユニットの加入者ユニットに情報信号を提供する方法であって、

加入者ユニットで出力オーディオ信号を生じさせるステップであって、出力オーディオ信号が対応する識別を有する、加入者ユニットで出力オーディオ信号を生じさせるステップと、

出力オーディオ信号が生ずる間に入力音声信号が加入者ユニットで検出されると、加入者ユニットから少なくとも識別を受信するステップと、

少なくとも一部の識別に応答して、加入者ユニットに情報信号を提供するステップとを備える、1 つ以上の加入者ユニットの加入者ユニットに情報信号を提供する方法。

(項目 26) 出力オーディオ信号を生じさせるステップが、

加入者ユニットに音声信号を提供するステップを備える、項目 25 に記載の方法。

(項目 27) 情報信号を提供するステップが、

加入者ユニットに情報信号を向けるステップをさらに備え、情報信号が、加入者ユニットの動作を制御する、項目 25 に記載の方法。

(項目 28) 加入者ユニットが、少なくとも 1 つの装置に結合され、情報信号を提供するステップが、

少なくとも 1 つの装置に情報を向けるステップをさらに備え、情報信号が、少なくとも 1 つの装置の動作を制御する、項目 25 に記載の方法。

(項目 29) 出力オーディオ信号を生じさせるステップが、

加入者ユニットに制御シグナリングを提供するステップをさらに備え、制御シグナリングが、加入者ユニットに出力オーディオ信号として音声信号を合成させる、項目 25 に記載の方法。

(項目 30) 入力音声信号に対応するパラメータで表示された音声信号を受信するステップと、

少なくとも一部の識別およびパラメータで表示された音声信号に応答して、加入者ユニットに情報信号を提供するステップとをさらに備える、項目 25 に記載の方法。

(項目 31) 音声認識サーバを備えるインフラストラクチャと無線通信する加入者ユニットであって、加入者ユニットがスピーカおよびマイクロフォンを備え、スピーカが出力オーディオ信号を提供し、マイクロフォンが入力音声信号を提供し、加入者ユニットが、

入力音声信号の開始を検出する手段と、

出力オーディオ信号に対して、入力音声信号の開始の入力開始時間を決定する手段と、

制御パラメータとして音声認識サーバに入力開始時間を提供する手段とを備える、加入者ユニット。

(項目 32) 少なくとも一部の入力開始時間に基づき、音声認識サーバから少なくとも 1 つの情報信号を受信する手段をさらに備える、項目 31 に記載の加入者ユニット。

10

20

30

40

50

(項目 3 3) パラメータで表示された音声信号を提供するために入力音声信号を分析する手段をさらに備え、

提供する手段が、さらに、パラメータで表示された音声信号を音声認識サーバに提供するように機能し、受信する手段が、さらに、少なくとも一部の入力開始時間およびパラメータで表示された音声信号に基づき音声認識サーバから少なくとも 1 つの制御信号を受信するように機能する、項目 3 2 に記載の加入者ユニット。

(項目 3 4) 入力開始時間を決定する手段が、出力オーディオ信号の開始より遅く、かつその後の出力オーディオ信号の開始よりも早い入力開始時間を決定するよう機能する、項目 3 1 に記載の加入者ユニット。

(項目 3 5) 入力開始時間が、出力オーディオ信号の一時的なコンテキストに対するタイムスタンプ、出力オーディオ信号のサンプルコンテキストに対するサンプルインデックス、および出力オーディオ信号のフレームコンテキストに対するフレームインデックスのいずれか 1 つである、項目 3 1 に記載の加入者ユニット。

(項目 3 6) インフラストラクチャから出力オーディオ信号として提供される音声信号を受信する手段をさらに備える、項目 3 1 に記載の加入者ユニット。

(項目 3 7) インフラストラクチャから出力オーディオ信号に関する制御シグナリングを受信する手段と、

制御シグナリングに応答して出力オーディオ信号として音声信号を合成する手段とをさらに備える、項目 3 1 に記載の加入者ユニット。

(項目 3 8) 音声認識サーバを備えるインフラストラクチャと無線通信する加入者ユニットであって、加入者ユニットがスピーカおよびマイクロフォンとを備え、スピーカが出力オーディオ信号を提供し、マイクロフォンが入力音声信号を提供し、

出力オーディオ信号が生ずる間に入力音声信号を検出する手段と、

出力オーディオ信号に対応する識別を決定する手段と、

制御パラメータとして音声認識サーバに識別を提供する手段とをさらに備える、加入者ユニット。

(項目 3 9) 少なくとも一部の識別に基づき、音声認識サーバから少なくとも 1 つの制御信号を受信する手段をさらに備える、項目 3 8 に記載の加入者ユニット。

(項目 4 0) パラメータで表示された音声信号を提供するために入力音声信号を分析する手段をさらに備え、

提供する手段が、さらに、パラメータで表示された音声信号を音声認識サーバに提供するように機能し、受信する手段が、さらに、少なくとも一部の識別およびパラメータで表示された音声信号とに基づき音声認識サーバから少なくとも 1 つの制御信号を受信するように機能する、項目 3 9 に記載の加入者ユニット。

(項目 4 1) インフラストラクチャから出力オーディオ信号として提供される音声信号を受信する手段をさらに備える、項目 3 8 に記載の加入者ユニット。

(項目 4 2) インフラストラクチャから出力オーディオ信号に関する制御シグナリングを受信する手段と、

制御シグナリングに応答して出力オーディオ信号として音声信号を合成する手段とをさらに備える、項目 3 8 に記載の加入者ユニット。

(項目 4 3) 1 つ以上の加入者ユニットと無線通信するインフラストラクチャの一部を形成する音声認識サーバであって、

1 つ以上の加入者ユニットの加入者ユニットで出力オーディオ信号を生じさせる手段と、

加入者ユニットにおける出力オーディオ信号に対する入力音声信号の開始に対応する少なくとも入力開始時間を加入者ユニットから受信する手段と、

少なくとも一部の入力開始時間に応答して加入者ユニットに情報信号を提供する手段とを備える、音声認識サーバ。

(項目 4 4) 入力開始時間が、出力オーディオ信号の一時的なコンテキストに対するタイムスタンプ、出力オーディオ信号のサンプルコンテキストに対するサンプルインデ

10

20

30

40

50

ックス、および出力オーディオ信号のフレームコンテキストに対するフレームインデックスのいずれか1つである、項目43に記載の音声認識サーバ。

(項目45) 情報信号を提供する手段が、

加入者ユニットに情報信号を向けるよう機能し、情報信号が、加入者ユニットの動作を制御する、項目43に記載の音声認識サーバ。

(項目46) 加入者ユニットが、少なくとも1つの装置に結合され、情報信号を提供する手段が、さらに、少なくとも1つの装置に情報を向けるよう機能し、情報信号が、少なくとも1つの装置の動作を制御する、項目43に記載の方法。

(項目47) 出力オーディオ信号を生じさせる手段が、さらに、出力オーディオ信号として提供される音声信号を提供するよう機能する、項目43に記載の音声認識サーバ

。

(項目48) 出力オーディオ信号を生じさせる手段が、さらに、加入者ユニットに制御シグナリングを提供するよう機能し、制御シグナリングが、加入者ユニットに出力オーディオ信号として音声信号を合成させる、項目43に記載の音声認識サーバ。

(項目49) 受信する手段が、さらに、入力音声信号に対応するパラメータで表示された音声信号を受信するよう機能し、提供する手段が、さらに、少なくとも一部の入力開始時間およびパラメータで表示された音声信号に応答して加入者ユニットに情報信号を提供するよう機能する、項目43に記載の音声認識サーバ。

(項目50) 1つ以上の加入者ユニットと無線通信するインフラストラクチャの一部を形成する音声認識サーバであって、

1つ以上の加入者ユニットの加入者ユニットで出力オーディオ信号を生じさせる手段であって、出力オーディオ信号が対応する識別を有する、1つ以上の加入者ユニットの加入者ユニットで出力オーディオ信号を生じさせる手段と、

入力音声信号が、出力オーディオ信号が生ずる間に加入者ユニットにおいて検出されると、加入者ユニットから少なくとも識別を受信する手段と、

少なくとも一部の識別に応答して、加入者ユニットに情報信号を提供する手段とをさらに備える、音声認識サーバ。

(項目51) 出力オーディオ信号を生じさせる手段が、さらに、出力オーディオ信号として提供される音声信号を提供するよう機能する、項目50に記載の音声認識サーバ

。

(項目52) 出力オーディオ信号を生じさせる手段が、さらに、加入者ユニットに制御シグナリングを提供するよう機能し、制御シグナリングが、加入者ユニットに出力オーディオ信号として音声信号を合成させる、項目50に記載の音声認識サーバ。

(項目53) 受信する手段が、さらに、入力音声信号に対応するパラメータで表示された音声信号を受信するよう機能し、提供する手段が、さらに、少なくとも一部の入力開始時間およびパラメータで表示された音声信号に応答して加入者ユニットに情報信号を提供するよう機能する、項目50に記載の音声認識サーバ。

(項目54) 情報信号を提供する手段が、さらに、加入者ユニットに情報信号を向けるよう機能し、情報信号が、加入者ユニットの動作を制御する、項目50に記載の音声認識サーバ。

(項目55) 加入者ユニットが、少なくとも1つの装置に結合され、情報信号を提供する手段が、さらに、少なくとも1つの装置に情報を向けるよう機能し、情報信号が、少なくとも1つの装置の動作を制御する、項目50に記載の方法。

【図面の簡単な説明】

【0007】

【図1】本発明による無線通信システムのブロック図である。

【図2】本発明による加入者ユニットのブロック図である。

【図3】本発明による加入者ユニット内の音声およびデータ処理機能の概略図である。

【図4】本発明による音声認識サーバのブロック図である。

【図5】本発明による音声認識サーバ内の音声およびデータ処理機能の概略図である。

10

20

30

40

50

【図 6】本発明によるコンテキスト決定を図示している。

【図 7】本発明による出力オーディオ信号が生ずる間に入力音声信号を処理する方法を図示しているフローチャートである。

【図 8】本発明による出力オーディオ信号が生ずる間に入力音声信号を処理する別の方法を図示しているフローチャートである。

【図 9】本発明による音声認識サーバ内で実施されることができする方法を図示しているフローチャートである。

【発明を実施するための形態】

【0008】

(発明の実施の形態)

本発明は、図 1 乃至図 9 を参照すると、いっそう完全に記述されることができる。図 1 は、加入者ユニット 102-103 を備える無線通信システム 100 の全体のシステムアーキテクチャを示している。その加入者ユニット 102-103 は、無線システム 110 によってサポートされる無線チャンネルを媒介としてインフラストラクチャと通信する。本発明のインフラストラクチャは、無線システム 110 に加えて、小さなエンティティシステム 120、コンテンツプロバイダシステム 130 およびデータ網 150 いずれかを媒介として互いに結合される企業システム 140 を備えている。

【0009】

加入者ユニットは、通信インフラストラクチャと通信することが可能な、ハンドヘルドセル電話 103、自動車 102 内にある無線通信装置などのあらゆる無線通信装置を備えることができる。図 1 に示されるもの以外の様々な加入者ユニットが使用されることが可能であることは分かる；本発明は、この点については限定されるものではない。加入者ユニット 102-103 は、ハンドフリー音声通信と、ローカル音声認識および合成システムと、クライアント-サーバ音声認識および合成システムのクライアント部分とのためのハンドフリーセルラー電話のコンポーネントとを含むことが好ましい。これらのコンポーネントは、図 2 および図 3 を参照として以下にいっそう詳細に記述されている。

【0010】

加入者ユニット 102-103 は、無線チャンネル 105 を媒介として無線システム 110 と無線で通信する。本発明が音声通信をサポートする無線システムの他のタイプに有益に適応されることを通常の技量の者が認めているが、無線システム 110 は、セルラーシステムを備えることが好ましい。無線チャンネル 105 は、一般に、デジタル送信技術を実施し、かつ、加入者ユニット 102-103 への、また、加入者ユニット 102-103 からの音声および／またはデータを伝達することが可能な無線周波数 (RF) キャリアである。アナログ技術などの別の送信技術も使用されることが分かる。好ましい実施の形態において、無線チャンネル 105 は、European Telecommunications Standards Institute (欧州電気通信標準化協会) (PTSI) によって定義される General Packet Data Radio Service (ジェネラルパケットデータ無線サービス) (GPRS) などの無線パケットデータチャネルである。無線チャンネル 105 は、クライアント-サーバ音声認識および合成システムのクライアント部分と、クライアント-サーバ音声認識および合成システムのサーバ部分との間の通信を促進するために、データを搬送する。ディスプレイ、制御、ロケーション、ステータス情報などの他の情報は、さらに、無線チャンネル 105 全域で搬送されることが可能である。

【0011】

無線システム 110 は、加入者ユニット 102-103 から無線チャンネル 105 によって伝達される送信を受信するアンテナ 112 を備えている。アンテナ 112 は、さらに、無線チャンネル 105 を媒介として加入者ユニット 102-103 に伝送する。アンテナ 112 を媒介として受信されるデータは、データ信号に変換され、そして、無線網 113 に搬送される。逆に、無線網 113 からのデータは、送信用のアンテナ 112 に送信される。本発明のコンテキストにおいて、無線網 113 は、一般に技術上周知のように、ベ

10

20

30

40

50

ースステーション、コントローラ、リソース割当装置、インターフェイス、データベースなどの無線システムを実施するのに必要なこれらの装置を備えている。通常の当業者に理解できるように、無線網 113 に組み込まれる特定の要素は、使用される無線システム 110 の特定のタイプ、たとえば、セルラーシステム、中継陸上移動システムなどによって決まる。

【0012】

クライアントサーバ音声認識および合成システムのサーバ部分を提供する音声認識サーバ 115 は、無線網 113 に結合されることができ、それによって、無線システム 110 のオペレータが加入者ユニット 102-103 のユーザに音声利用サービスを提供することを可能とする。制御エンティティ 116 は、さらに、無線網 113 に結合されることができ、制御エンティティ 116 は、加入者ユニット、あるいは、加入者ユニットに相互に接続される装置を制御するために、音声認識サーバ 115 によって提供される入力に応答して、制御信号を加入者ユニット 102-103 に送信するのに使用されることが可能である。示されるように、あらゆる適切にプログラムされた汎用コンピュータを備えることができる制御エンティティ 116 は、無線網 113 を介してか、あるいは、ダッシュ記号の相互接続によって示されるように、直接的にかのいずれかで、音声認識サーバ 115 に結合されることができ、

10

【0013】

上記に述べられるように、本発明のインフラストラクチャは、データ網 150 を媒介として互いに結合される様々なシステム 110, 120, 130, 140 を備えることが可能である。適切なデータ網 150 は、周知の網テクノロジーを使用するプライベートデータ網、Internet（インターネット）などの公衆網、あるいは、それらの組み合わせを備えることができる。別の方法として、あるいは、追加として、無線システム 110 内の音声認識サーバ 115、リモート音声認識サーバ 123, 132, 143, 145 は、音声利用サービスを加入者ユニット 102-103 に提供するために、様々な方法で、データ網 150 に接続されることができ、リモート音声認識サーバは、提供されるとき、同様に、データ網 150 およびあらゆる介在通信パスを介して制御エンティティ 116 と通信することが可能である。

20

【0014】

デスクトップパーソナルコンピュータ、小さなエンティティシステム 120（小規模ビジネス、あるいは、ホームなどの）内の他の汎用処理装置などのコンピュータ 122 は、音声認識サーバ 123 を実施するのに使用されることが可能である。加入者ユニット 102-103 へのデータおよび加入者ユニット 102-103 からのデータは、無線システム 110 とデータ網 150 とを介してコンピュータ 122 にルートされる。保存されたソフトウェアアルゴリズムおよびプロセスを実行して、コンピュータ 122 は、好ましい実施の形態において、音声認識システムと音声合成システムとの両方のサーバ部分を含む音声認識サーバ 123 の機能を提供する。たとえば、コンピュータ 122 が、ユーザのパーソナルコンピュータである所では、そのコンピュータの音声認識サーバソフトウェアは、ユーザの電子メール、電話帳、カレンダー、他の情報などのコンピュータにあるユーザの個人情報に結合されることが可能である。この構成により、加入者ユニットのユーザが、音声利用インターフェイスを利用する加入者のパーソナルコンピュータの個人情報にアクセスすることを可能とする。本発明によるクライアントサーバ音声認識および音声合成システムのクライアント部分は、図 2 および図 3 を参照として以下に記述される。本発明によるクライアントサーバ音声認識および音声合成システムのサーバ部分は、図 4 および図 5 を参照として以下に記述される。

30

40

【0015】

別の方法として、加入者ユニットのユーザに使用可能とさせる情報を有するコンテンツプロバイダ 130 は、音声認識サーバ 132 をデータ網に接続することが可能である。特徴、あるいは、特別のサービスとして提案されると、音声認識サーバ 132 は、コンテンツプロバイダの情報（図示せず）へのアクセスを望む加入者ユニットのユーザに音声利用

50

インターフェイスを提供する。

【0016】

音声認識サーバのための別の可能なロケーションは、大企業、あるいは、同様のエンティティなどの企業140内にある。Intranet（イントラネット）などの企業の内部網146は、セキュリティゲートウェイ142を媒介としてデータ網150に接続される。セキュリティゲートウェイ142は、加入者ユニットと関連して、企業の内部網146への安全なアクセスを提供する。技術上周知のように、このように提供される安全なアクセスは、一般に、一部分、認証および暗号化テクノロジーに依存にしている。このように、安全にされていないデータ網150を媒介とする加入者ユニットと内部網146との間の安全な通信が行われる。企業140内で、音声認識サーバ145を実施するサーバソフトウェアは、所定の従業員のワークステーションなどのパーソナルコンピュータに提供されることが可能である。小さなエンティティシステムにおける使用について上記に記述される構成と同様に、ワークステーションアプローチにより、従業員が、音声利用インターフェイスを介して作業関連情報、あるいは、他の情報にアクセスすることを可能とする。さらに、コンテンツプロバイダ130モデルと同様に、企業140は、企業データベースへのアクセスを行うために、内部で使用可能な音声認識サーバ143を提供することが可能である。

10

【0017】

本発明の音声認識サーバが設置される所にかかわらず、それらは、様々な音声利用サービスを実施するのに使用されることが可能である。たとえば、制御エンティティ116と関連して操作すると、提供されるとき、音声認識サーバは、加入者ユニットに結合される加入者ユニット、あるいは、装置の動作制御を可能とする。本明細書を通じて使用されるように、音声認識サーバという用語は、音声合成機能をも含むことを意図されるということは留意されるべきである。

20

【0018】

本発明のインフラストラクチャは、さらに、加入者ユニット102-103と通常の電話技術システムとの間の相互接続を提供する。これは、POTS（プレーンな旧来の電話システム）への無線網113の結合により、図1において示されている。技術上周知のように、POTS網118、あるいは、同様な電話網は、陸上通信線電話受話器、あるいは、他の無線装置などの複数の呼び局119への通信アクセスを提供する。このように、加入者ユニット102-103のユーザは、呼び局119の別のユーザと音声通信を行うことが可能である。

30

【0019】

図2は、本発明による加入者ユニットを実施するのに使用されることができハードウェアアーキテクチャを示している。示されるように、2つの無線トランシーバ、無線データトランシーバ203と無線音声トランシーバ204とが、使用されることができ。技術上周知のように、これらのトランシーバは、データおよび音声の機能の両方を行うことが可能である単一のトランシーバに組み合わせられることができる。無線データトランシーバ203および無線音声トランシーバ204は、両方ともアンテナ205に接続される。別の方法として、各トランシーバのための独立したアンテナが、さらに、使用されることができ。無線音声トランシーバ204は、無線音声通信を行うために、すべての必要な信号処理、プロトコル終了、変復調などを行い、そして、好ましい実施の形態において、セルラートランシーバを備えている。同様な方法で、無線データトランシーバ203は、インフラストラクチャとのデータ連結性を提供する。好ましい実施の形態において、無線データトランシーバ203は、European Telecommunication Standards Institute（欧州電気通信標準化協会）（ETSI）によって定義されるGeneral Packet Data Radio Service（ジェネラルパケットデータ無線サービス）（GPRS）などの無線パケットデータをサポートする。

40

【0020】

50

本発明が、以下に論じられるように、自動車システム内に特定の利点を有して適用されることが可能であることは予測される。自動車内で使用されるとき、本発明による加入者ユニットは、さらに、一般に、自動車の一部であるか、加入者ユニットの一部でないとして理解される処理コンポーネントを含む。本発明を説明するために、このような処理コンポーネントは、加入者ユニットの一部であると想定する。加入者ユニットの実際の実施は、設計の考慮すべき問題によって指図されるように、このような処理コンポーネントを含んでも、あるいは、含まなくてもよい。好ましい実施の形態において、処理コンポーネントは、IBM Corp. , による“POWER PCなどの”汎用プロセッサ(CPU) 201と、Motorola Inc. によるDSP56300 シリーズプロセッサなどのデジタル信号プロセッサ(DSP) 202とを備えている。CPU 201およびDSP 202は、それらが技術上周知のように、データおよびアドレスバスと、他の制御接続とを媒介として互いに結合されることを示すために、図2に隣接して示されている。別の実施の形態は、CPU 201とDSP 202との両方の機能を単一のプロセッサに組み合わせるか、あるいは、それらをいくつかのプロセッサに分割することが可能である。CPU 201およびDSP 202は、その関連したプロセッサにプログラムおよびデータストレージを提供するそれぞれのメモリー240, 241に結合される。格納されたソフトウェアルーチンを使用して、CPU 201および/またはDSP 202は、本発明の機能の少なくとも一部を実施するためにプログラムされることが可能である。CPU 201およびDSP 202のソフトウェア機能は、以下に図3および図7を参照として少なくとも一部記述されている。

【0021】

好ましい実施の形態において、加入者ユニットは、さらに、アンテナ207に結合される全地球測位システム(GPS)受信機206を含む。GPS受信機206は、受信したGPS情報を提供するためにDSP 202に結合される。DSP 202は、GPS受信機206から情報を受け、そして、無線通信装置のロケーション座標を計算する。別の方法として、GPS受信機206は、直接CPU 201にロケーション情報を提供することができる。

【0022】

CPU 201およびDSP 202の様々な入力および出力は、図2に示されている。図2に示されるように、太い実線は、音声関連情報に相当し、そして、太い鎖線は、制御/データ関連情報に相当する。オプション要素および信号バスは、点線を使用して図示されている。DSP 202は、以下にさらに詳細に記述されるように、電話(セル電話)会話および音声入力の両方のための音声入力を、ローカル音声認識装置およびクライアントサーバ音声認識装置のクライアントサイド部分との両方に提供するマイクロフォン270からマイクロフォンオーディオ220を受信する。DSP 202は、さらに、電話(セル電話)会話および音声入力の音声入力を、ローカル音声シンセサイザおよびクライアントサーバ音声シンセサイザのクライアントサイド部分との両方から提供する少なくとも1つのスピーカ271に向けられる出力オーディオ211に結合される。マイクロフォン270およびスピーカ271は、ハンドヘルド装置におけるように、隣接して互いに位置されることができると、あるいは、遮光板取り付けマイクロフォンおよびダッシュ、または、ドア取り付けスピーカを有する自動車の適用におけるように、互いに最も近くに位置されることができるとに留意してください。

【0023】

本発明の1つの実施の形態において、CPU 201は、双方向性インターフェイス230を介して自動車内のデータバス208に結合されている。このデータバス208により、制御およびステータス情報が、セル電話、エンターテインメントシステム、環境制御システムなどの自動車内の様々な装置209a-nとCPU 201との間に通信されることを可能とする。適切なデータバス208は、一般に、Society of Automotive Engineers (自動車エンジニア協会)によって標準化されるプロセスにおけるITS Data Bus (ITSデータバス)であることが予期される。

Bluetooth Special Interest Group (ブルトウス 特別利害グループ) (SIG) によって定義される短距離の無線データ通信システムなどの様々な装置間の制御およびステータス情報を伝える別の手段が使用されることができる。データバス 208 により、CPU 201 が、ローカル音声認識装置によって、あるいは、クライアントサーバ音声認識装置によってかのいずれかによって認識される音声コマンドに応答して自動車データバスの装置 209 を制御することを可能とする。

【0024】

CPU 201 は、受信データ接続 231 および送信データ接続 232 を媒介として無線データトランシーバ 203 に結合される。これらの接続 231-232 により、CPU 201 が、無線システム 110 から送信される制御情報および音声合成情報を受信すること
10
を可能とする。音声合成情報は、無線データチャネル 105 を媒介としてクライアントサーバ音声合成システムのサーバ部分から受信される。CPU 210 は、次に、DSP 202 に引き渡される音声合成情報を復号化する。DSP 202 は、次に、出力音声を合成し、そして、それをオーディオ出力 211 に引き渡す。受信データ接続 231 を媒介として受信されるあらゆる制御情報は、加入者ユニット自体の動作を制御するのに使用されるか、あるいは、それらの動作を制御するために、1 つ以上の装置に送信するの
20
に使用されることができる。そのうえ、CPU 210 は、ステータス情報と、クライアントサーバ音声認識システムのクライアント部分からの出力データとを無線システム 110 に送信することが可能である。クライアントサーバ音声認識システムのクライアント部分は、以下にいっそう詳細に記述されるように、DSP 202 および CPU 201 のソフトウェアにおいて実施されることが好ましい。音声認識をサポートするとき、DSP 202 は、マイクロフォン入力 220 から音声を受信し、そして、パラメータで表
示された音声信号を CPU 201 に提供するために、このオーディオを処理する。CPU 201 は、パラメータで表示された音声信号を符号化し、そして、無線データチャネル 105 全体にインフラストラクチャの音声認識サーバに送信されるべく、送信データ接
続 232 を媒介として、この情報を無線データトランシーバ 203 に送信する。

【0025】

無線音声トランシーバ 204 は、双方向性データバス 233 を媒介として CPU 201 に結合される。このデータバスにより、CPU 201 が、無線音声トランシーバ 204 の動作を制御することを可能とし、そして、無線音声トランシーバ 204 からステータ
30
ス情報を受信する。無線音声トランシーバ 204 は、さらに、送信オーディオ接続 221 と受信オーディオ接続 210 とを媒介として DSP 202 に結合される。無線音声トランシーバ 204 が、電話 (セルラー) コールを促進するのに使用されるとき、オーディオは、DSP 202 によってマイクロフォン入力 220 から受信される。マイクロフォンオーディオは、処理され (たとえば、フィルター、圧縮されるなど)、そして、セルラーインフラストラクチャに伝送されるべく、無線音声トランシーバ 204 に提供される。逆に、無線音声トランシーバ 204 によって受信されるオーディオは、受信オーディオ接続 210 を媒介として、オーディオが処理され (たとえば、解凍、フィルターされるなど)、そして、スピーカ出力 211 に提供される DSP 202 に送信される。DSP 202 によ
40
って行われる処理は、図 3 を参照としていっそう詳細に記述される。

【0026】

図 2 に示される加入者ユニットは、任意に、音声通信の間に割り込みインジケータ 251 を手動で提供するのに使用される入力装置 250 を備えることができる。すなわち、音声会話の間、加入者ユニットのユーザは、割り込みインジケータを提供するために、手動で入力装置を始動することが可能であり、それによって、音声認識機能を引き起こすためにユーザの所望を信号する。たとえば、音声通信の間、加入者ユニットのユーザは、音声利用コマンドを電子付随部に提供するために、たとえば、電話をかけて、そして、第三者をコールに加えるために、会話に割り込むことを希望することができる。入力装置 250 は、特に、たとえば、単一の、あるいは多目的ボタン、マルチ位置決めセレクト、あるいは入力
50
ケイパビリティを有するメニュー駆動ディスプレイなどを含む、あらゆるタイプの

ユーザ始動入力メカニズムを実質的に備えることができる。別の方法として、入力装置 250 は、双方向性インターフェイス 230 と自動車内データバス 208 とを媒介として CPU 201 に接続されることができる。とにかく、このような入力装置 250 が提供されるとき、CPU 201 は、割り込みインジケータの発生セグメントを識別するために、検出器として作動する。CPU 201 が、入力装置 250 のための検出器として作動するとき、CPU 201 は、参照符号 260 で識別される信号パスによって図示されるように、DSP 202 への割り込みインジケータの存在を表示する。逆に、別の実施が、割り込みインジケータを提供するために、検出器アプリケーションに結合されるローカル音声認識装置 (DSP 202 および／または CPU 201 内で実施されることが好ましい) を使用する。その場合、CPU 201 か、あるいは、DSP 202 かのいずれかが、参照符号 260 a によって識別される信号パスによって表わされるように、割り込みインジケータの存在を信号で知らせる。とにかく、割り込みインジケータの存在が検出されると、音声認識要素の部分 (このましくは、加入者ユニットと関連して、あるいは加入者ユニットの一部として実施されるクライアント部分) は、音声利用コマンドを処理し始めるために始動される。そのうえ、音声認識要素の部分が始動されたという表示は、さらに、ユーザと音声認識サーバとに提供されることができる。好ましい実施の形態において、このような表示は、音声認識要素を提供するために、送信データ接続 232 を媒介として、音声認識クライアントと協働する音声認識サーバへの送信のための無線データトランシーバ 203 に搬送される。

10

【0027】

20

最後に、加入者ユニットは、音声認識機能が割り込みインジケータに応答して始動されたというアナンスエータ制御 256 に応答して加入者ユニットのユーザに表示を提供するアナンスエータ 255 を装備することが好ましい。アナンスエータ 255 は、割り込みインジケータの検出に応答して始動され、そして、制限した持続期間のトーン、あるいは、ビーッという音などの音響表示を提供するのに使用されるスピーカを備えることができる。(もう一度、割り込みインジケータの存在は、入力装置利用信号 260 か、あるいは、音声利用信号 26 a のいずれかを使用して、信号で知らされることが可能である。) 別の実施において、アナンスエータの機能は、オーディオをスピーカ出力 211 に向ける DSP 202 によって実行されるソフトウェアプログラムを媒介として提供される。スピーカは、オーディオ出力 211 を可聴させるのに使用されるスピーカ 271 から独立しているか、あるいは、同一のものでよい。別の方法として、アナンスエータ 255 は、可視インジケータを提供する、LED、あるいは、LCD ディスプレイなどのディスプレイ装置を備えることができる。アナンスエータ 255 の特定の形状は、設計選択の問題であり、そして、本発明は、この点では限定される必要はない。さらに、アナンスエータ 255 は、双方向性インターフェイス 230 と自動車内データバス 208 とを媒介として、CPU 201 に接続されることができる。

30

【0028】

ここでは図 3 を参照すると、加入者ユニット内で行われる処理の部分 (本発明による動作) が、概略的に図示されている。図 3 に示される処理は、CPU 201 および／または DSP 202 によって実行される格納された機械可読インストラクションを使用して実施される。以下に呈される論議は、自動車の車両内に採用される加入者ユニットの動作について記述している。とはいえ、図 3 に全体として示され、そして、ここに記述される機能は、同様に、音声認識を使用する、あるいは、音声認識の使用で利益を得る非自動車利用アプリケーションに適用可能である。

40

【0029】

マイクロフォンオーディオ 220 は、加入者ユニットに入力として提供される。自動車環境において、マイクロフォンは、自動車の遮光板、あるいは、ステアリングコラムに、または、近くに、一般に取り付けられるハンドフリーマイクロフォンである。マイクロフォンオーディオ 220 は、デジタル形状でエコーキャンセルおよび環境処理 (ECEP) ブロック 301 に到達することが好ましい。スピーカオーディオ 211 は、あらゆる必要

50

な処理を行った後、E C E Pブロック301によって、1つまたは複数のスピーカに引き渡される。自動車内において、このようなスピーカは、ダッシュボードの下に取り付けられることが可能である。別の方法として、スピーカオーディオ211は、エンターテインメントシステムのスピーカシステムを介してプレイされるべく、自動車内のエンターテインメントシステムを介してルートされることが可能である。スピーカオーディオ211は、デジタルフォーマットであることが好ましい。セルラー電話コールが、たとえば、処理中であるとき、セルラー電話からの受信オーディオは、受信オーディオ接続210を媒介としてE C E Pブロック301に到達する。同様に、送信オーディオは、送信オーディオ接続221全体にわたりセル電話に引き渡される。

【0030】

E C E Pブロック301は、送信オーディオ接続221を媒介として、無線音声トランシーバ204に引き渡す前に、マイクロフォンオーディオ220からスピーカオーディオ211のエコーキャンセルを行う。エコーキャンセルのこの形状は、音響エコーキャンセルとして周知であり、そして、技術上周知である。たとえば、Amanoらに発行され、そして、“Sub-band Acoustic Echo Canceller (サブバンド音響エコーキャンセラー)”と題する米国特許第5,136,599号およびGenterに発行され、そして、“Echo Canceller with Subband Attenuation and Noise Injection Control (サブバンド減衰およびノイズインジェクション制御を有するエコーキャンセル)”と題する米国特許第5,561,668号は、音響エコーキャンセルを行うための適切な技術を教示し、その特許の教示は、この結果、ここに参照として含まれている。

【0031】

E C E Pブロック301が、さらに、エコーキャンセルに加えて、いっそう快い音声信号を加入者ユニットによって伝送されるオーディオを受信する関係者に提供するために、環境処理をマイクロフォンオーディオ220に提供する。一般に使用されている1つの技術は、ノイズ抑制と呼ばれる。自動車内のハンドフリーマイクロフォンは、一般に、別の関係者によって聞かされる多くのタイプの音響ノイズをピックアップする。この技術は、別の関係者が聞く知覚バックグラウンドノイズを減少し、そして、たとえば、Vilмуらに発行された米国特許第4,811,404号に記述されており、その特許の教示は、この結果、参照としてここに含まれている。

【0032】

E C E Pブロック301は、さらに、第1のオーディオパス316を媒介として音声合成バックエンド304によって提供される合成された音声のエコーキャンセル処理を行い、その合成された音声は、オーディオ出力211を媒介として1つまたは複数のスピーカに引き渡される。1つまたは複数のスピーカにルートされる受信された音声を有する場合におけるように、マイクロフォンオーディオパス220に到達するスピーカオーディオ“エコー”は、キャンセルされる。これにより、アコースティックにマイクロフォンに結合されるスピーカオーディオが、音声認識フロントエンド302に引き渡される前に、マイクロフォンオーディオから削除されることを可能とする。このタイプの処理は、“バージイン”として技術上周知であることを可能にする。バージインにより、音声認識システムが、出力音声と同時にシステムによって生成される間に、入力音声に応答することを可能とする。“バージイン”実施の実施例は、たとえば、米国特許第4,914,692号、第5,475,791号、第5,708,704号および第5,765,130号に見られる。バージイン処理についての本発明の適用は、以下にいっそう詳細に記述されている。

【0033】

エコーがキャンセルされたマイクロフォンオーディオは、音声認識処理が行われるたびに、第2のオーディオパス326を媒介として音声認識フロントエンド302に提供される。任意に、E C E Pブロック301は、第1のデータパス327を媒介として音声認識フロントエンド302にバックグラウンドノイズ情報を提供する。このバックグラウンド

ノイズ情報は、ノイズのある環境における音声認識システムオペレーティングのための認識性能を改善するのに使用されることが可能である。このような処理を行うための適切な技術は、Gersonらに発行された米国特許第4,918,732号に記述されており、その特許の教示は、この結果、参照としてここに含まれている。

【0034】

エコーがキャンセルされたマイクロフォンオーディオと、任意に、ECEPブロック301から受信されるバックグラウンドノイズ情報とに基づき、音声認識フロントエンド302は、パラメータで表示された音声情報を生成する。全体として、音声認識フロントエンド302および音声合成バックエンド304は、クライアントサーバ利用音声認識および合成システムのクライアントサイド部分のコア機能を提供する。パラメータで表示された音声情報は、一般に、特徴ベクトルの形をとり、新しいベクトルは、10msecから20msec毎に計算される。音声信号のパラメータ表示化のための1つの一般に使用される技術は、“Comparison Of Parametric Representations For Monosyllabic Word Recognition In Continuously Spoken Sentences,” IEEETransactions on acoustics Speech and Signal Processing, ASSP-28(4), pp. 357-366, 1980年8月に、Davisらによって記述されるようにmel cepstraであり、その公報の教示は、この結果、参照としてここに含まれている。

【0035】

音声認識フロントエンド302によって計算されるパラメータベクトルは、ローカル音声認識処理のための第2のデータパス325を媒介としてローカル音声認識ブロック303に受け渡される。パラメータベクトルは、さらに、任意に、第3のデータパス323を媒介として、音声適用プロトコルインターフェイス(API's)およびデータプロトコルとを備えるプロトコル処理ブロック306に受け渡される。周知の技術によれば、処理ブロック306は、送信データ接続232を媒介として、パラメータベクトルを無線データトランシーバ203に送信する。順番に、無線データトランシーバ203は、クライアントサーバ利用音声認識の一部として機能するサーバにパラメータベクトルを搬送する。(加入者ユニットは、パラメータベクトルを送信するのではなく、その代わりに、無線データトランシーバ203か、あるいは、無線音声トランシーバ204のいずれかを使用して、サーバに音声情報を送信することは明らかである。これは、加入者ユニットから電話網への音声の伝送をサポートするのに使用されるのと類似の方法で、あるいは、音声信号の他の適切な表現を使用して、行われることができる。すなわち、音声情報は、様々なパラメータで表示されない表現：未使用のデジタル化されたオーディオ、セルラー音声コーデックによって処理されたオーディオ、IP(インターネットプロトコル)などの特定のプロトコルによる伝送に適切なオーディオデータ等の様々なパラメータで表示されない表現のいずれかを備えることができる。順番に、サーバは、パラメータで表示されない音声情報を受信すると、必要なパラメータ表示化を行うことが可能である。)単一の音声認識フロントエンド302が示されるのに対して、ローカル音声認識装置303およびクライアントサーバ利用音声認識装置は、実際には、異なる音声認識フロントエンドを利用することができる。

【0036】

ローカル音声認識装置303は、音声認識フロントエンド302からパラメータベクトル325を受信し、そして、たとえば、パラメータで表示された音声内に何か認識可能な発話があるかどうかを決定するために、その結果として、音声認識分析を行う。1つの実施の形態において、認識された発話(一般に、ワード・言葉)は、第4のデータパス324を媒介として、ローカル音声認識装置303からプロトコル処理ブロック306に送信され、それは、順番に、さらなる処理のため様々なアプリケーション307に認識された発話を受け渡す。CPU201およびDSP202のいずれか、あるいは、両方を使用して実施されることができアプリケーション307は、認識された発話に基づき、音声利

用割り込みインジケータが受信されたことを確かめる検出器アプリケーションを含むことが可能である。たとえば、検出器は、同一性を検索する予め定められた発話（たとえば、“ウェークアップ”）のリストに対して、認識された発話を比較する。同一性が検出されるとき、検出器アプリケーションは、割り込みインジケータの存在を表わす信号260aを発する。割り込みインジケータの存在は、順番に、音声利用コマンドを処理し始めるために、音声認識要素の部分を始動するのに使用される。これは、音声認識フロントエンドに送られる信号260aによって、図3に概略的に示されている。応答中、音声認識フロントエンド302は、パラメータで表示されたオーディオを、ローカル音声認識装置にか、あるいは、好ましくは、追加の処理のため音声認識サーバへの伝送のためのプロトコル処理ブロック306にかのいずれかに、ルーティングし続ける。（さらに、入力装置250によって任意に提供される入力装置利用信号260は、さらに、同一の機能を作用することに留意してください。）そのうえ、割り込みインジケータの存在は、音声認識装置のインフラストラクチャ利用要素に警告するために、送信データ接続232に送信されることができる。

【0037】

音声合成バックエンド304は、音声のパラメトリック表現を入力とみなし、そして、パラメトリック表現を、次に、第1のオーディオパス316を媒介としてECPブロック301に引き渡される音声信号に変換する。使用される特定のパラメトリック表現は、設計選択の考慮すべき問題である。1つの一般に使用されているパラメトリック表現は、Klattの“Software For A Cascade/Parallel Formant Synthesizer”, Journal of the Acoustical society of America, Vol. 67, 1980, pp. 971-995に記述されるように、フォルマントパラメータである。線形予測パラメータは、MarkelらのLinear Prediction of Speech, Springer Verlag, New York, 1976に記述されるように、別の一般に使用されるパラメトリック表現である。KlattおよびMarkelらの公報のそれぞれの教示は、参照としてここに含まれている。

【0038】

クライアントーサーバ利用音声合成の場合、音声のパラメトリック表現は、無線チャンネル105、無線データトランシーバ203およびプロトコル処理ブロック306を媒介として、回線網から受信され、それは、第5のデータバス313を媒介として音声合成バックエンドに転送される。ローカル音声合成の場合、アプリケーション307は、話されるテキストストリングを生成する。このテキストストリングは、第6のデータバス314を媒介としてプロトコル処理ブロック306からローカル音声シンセサイザ305に受け渡される。ローカル音声シンセサイザ305は、テキストストリングを音声信号のパラメトリック表現に変換し、そして、第7のデータバスを媒介としてこのパラメトリック表現を音声信号への変換のため音声合成バックエンド304に受け渡す。

【0039】

受信データデータ接続231が、音声合成情報に加えて、他の受信された情報を搬送するのに使用されることが可能であることは留意されるべきである。たとえば、他の受信された情報は、インフラストラクチャから受信されるデータ（ディスプレイ情報などの）および／または制御情報、およびシステムにダウンロードされるコードを含むことができる。同様に、送信データ接続232は、音声認識フロントエンド302によって計算されるパラメータベクトルに加えて、他の伝送情報を搬送するのに使用されることが可能である。たとえば、他の伝送情報は、装置ステータス情報、装置可能出力およびバージンタイミングに関連する情報を含むことができる。

【0040】

ここでは、図4を参照すると、本発明によるクライアントーサーバ音声認識および合成システムのサーバ部分を提供する音声認識サーバのハードウェアの実施の形態が示されている。このサーバは、図1を参照として上記に記述されるように、いくつかの環境にいる

ことが可能である。加入者ユニット、あるいは、制御エンティティとのデータ通信は、インフラストラクチャ、あるいは、網接続 411 を介して可能にされる。この接続 411 は、たとえば、図 1 に示されるように、無線システムに局部的であり、および直接無線網に接続されることができる。別の方法として、接続 411 は、パブリックデータ網にか、あるいは、プライベートデータ網にか、または、なにか他のデータ通信リンクにかである；本発明は、この点では限定されるものではない。

【0041】

網インターフェイス 405 は、CPU 401 と網接続 411 との間の接続を行う。網インターフェイス 405 は、受信バス 408 を媒介として、網 411 から CPU 401 に、そして、送信バス 410 を媒介として、CPU 401 から網接続 411 にデータをルートする。クライアントーサーバ構成の一部として、CPU 401 は、網インターフェイス 405 と網接続 411 とを媒介として、1 つ以上のクライアント（加入者ユニットにおいて実施されることが好ましい）と通信する。好ましい実施の形態において、CPU 401 は、クライアントーサーバ音声認識および合成システムのサーバ部分を実施する。図示されていないが、図 4 に示されるサーバは、さらに、サーバへのローカルアクセスを可能とするローカルインターフェイスを備えることができ、それによって、たとえば、サーバメンテナンス、ステータスチェックングおよび他の同様の機能を促進する。

【0042】

メモリー 403 は、機械可読インストラクション（ソフトウェア）と、クライアントーサーバ構成のサーバ部分を実施する際に、CPU 401 による実行および使用のためのプログラムデータとを格納する。このソフトウェアの動作および構造は、さらに、図 5 を参照として記述される。

【0043】

図 5 は、音声認識および合成サーバ機能の実施を示している。少なくとも 1 つの音声認識クライアントと協働して、図 5 に示される音声認識サーバ機能は、音声認識要素を提供する。加入者ユニットからのデータは、受信バス 408 を媒介として、受信機（RX）502 に到達する。受信機は、データを復号化し、そして、音声認識クライアントから音声認識分析器 504 に音声認識データ 503 をルートする。装置ステータス情報、装置可能出力、バージョンコンテキストに関連する情報などの加入者ユニットからの他の情報 506 は、受信機 502 によってローカル制御プロセッサ 508 にルートされる。1 つの実施の形態において、他の情報 506 は、音声認識要素（たとえば、音声認識クライアント）の部分が、始動されたという加入者ユニットからの表示を含む。このような表示は、音声認識サーバにおける音声認識処理を始動するのに使用されることが可能である。

【0044】

クライアントーサーバ音声認識構成の一部として、音声認識分析器 504 は、加入者ユニットからの音声認識パラメータベクトルを受け、そして、認識処理を完了する。認識されたワード、あるいは、発話 507 は、次に、ローカル制御プロセッサ 508 に受け渡される。パラメータベクトルを認識された発話に変換するのに必要な処理の記述は、Lee らの“自動音声認識：The Development of the Sphinx System”，1988 に見られることが可能であり、その公報の教示は、参照としてここに含まれている。上記に述べられるように、加入者ユニットからパラメータベクトルを受信するのではなく、サーバ（すなわち、音声認識分析器 504）は、パラメータで表示されない音声情報を受信することができることも分かる。もう一度、音声情報は、上記に記述されるようにいくつかの形状のいずれかをとり、この場合、音声認識分析器 504 は、第一に、たとえば、mel cepstra 技術を使用して、音声情報をパラメータで表示する。結果として生ずるパラメータベクトルは、次に、上記に記述されるように、認識された発話に変換されることができる。

【0045】

ローカル制御プロセッサ 508 は、音声認識分析器 504 と他の情報 508 とから認識された発話 507 を受信する。一般に、本発明は、認識された発話に作用し、そして、認

識された発話に基づき、制御信号を提供するために、制御プロセッサを必要とする。好ましい実施の形態において、これらの制御信号は、加入者ユニット、あるいは、加入者ユニットに結合される少なくとも1つの装置の動作を実質的に制御するのに使用される。このために、ローカルコントロールプロセッサは、2つの方法の1つで作動されることが好ましい。第一に、ローカル制御プロセッサ508は、アプリケーションプログラムを実施することが可能である。典型的なアプリケーションの1つの実施例は、米国特許第5,652,789号に記述されるように電子アシスタントである。別の方法として、このようなアプリケーションは、リモート制御プロセッサ516でリモート制御によりランすることが可能である。たとえば、図1のシステムにおいて、リモート制御プロセッサは、制御エンティティ116を備えている。この場合、ローカル制御プロセッサ508は、データ網接続515を媒介として、リモート制御プロセッサ516と通信することにより、データを受け渡して、受信することによってゲートウェイのように作動する。データ網接続515は、公衆（たとえば、インターネット）、プライベート（たとえば、イントラネット）、他のデータ通信リンクなどでよい。実際に、ローカル制御プロセッサ508は、ユーザによって利用されるアプリケーション／サービスによって異なるが、データ網にある様々なリモート制御プロセッサと通信することができる。

【0046】

リモート制御プロセッサ516か、あるいは、ローカル制御プロセッサ508かのいずれかでランするアプリケーションプログラムは、認識された発話507および／または他の情報506への応答を決定する。その応答は、合成されたメッセージおよび／または制御信号を備えることが好ましい。制御信号513は、ローカル制御プロセッサ508からトランシーバ(TX)510に中継される。合成される情報514、一般に、テキスト情報は、ローカル制御プロセッサ508から、テキストから音声への分析器512に送信される。テキストから音声への分析器512は、入力テキストストリングをパラメトリック音声表現に変換する。このような変換を行う適切な技術は、Sproat（編集者）の“Multilingual Text-to-Speech Synthesis: The Bell Labs Approach”，1997に記述され、その公報の教示は、参照としてここに含まれている。テキストから音声への分析器512からのパラメトリック音声表現は、必要に応じて、加入者ユニットへの伝送のための送信パス410全体にわたりパラメトリック音声表現511および制御情報513を多重送信するトランスミッタ510に提供される。ほんの今記述されたのと同じ方法のオペレーティングで、テキストから音声への分析器512は、さらに、加入者ユニットで出力オーディオ信号としてプレイされる合成されたプロンプトなどを提供するのに使用されることができる。

【0047】

本発明によるコンテキスト決定は、図6に示されている。図6に示されるアクティビティのための基準のポイントは、加入者ユニットのものであることは留意されるべきである。すなわち、図6は、加入者への、また、加入者からの可聴信号のタイム進行を示している。特に、出力オーディオ信号601のタイムを通じての進行が図示されている。出力オーディオ信号601は、出力サイレンス・消音604aの第1の周期によって分離される先行の出力オーディオ信号602によって処置されることができ、そして、出力サイレンス604bの第2の周期によって、結果として生ずる出力オーディオ信号が続く。出力オーディオ信号601は、音声信号などのあらゆるオーディオ信号、合成された音声信号、あるいは、プロンプト、可聴トーン、あるいは、ビーツという音等を備えることができる。本発明の1つの実施の形態において、各出力オーディオ信号601-603は、ちょうどよいあらゆる所定のモーメントで出力される信号を識別するのを促進するために、それに割り当てられる関連したユニークな識別子を有している。このような識別子は、非リアルタイムで、様々な出力オーディオ信号（たとえば、合成されたプロンプト、トーンなど）予め割り当てられることができ、あるいは、リアルタイムで作成され、そして、割り当てられることができる。さらに、識別子自体は、たとえば、帯域内信号方式、あるいは、帯域外周波信号方式を使用して、出力オーディオ信号を提供するのに使用される情報とと

もに伝送されることができる。別の方法として、予め割り当てられた識別子の場合、識別子自体は、加入者ユニットに提供されることが可能であり、そして、その識別に基づいて、加入者ユニットは、出力オーディオ信号を合成することが可能である。通常の技量の者には、出力オーディオ信号のための識別子を提供し、そして、使用する様々な技術が、難なく考案され、そして、本発明に適用されることができるといことは認められるであろう。

【0048】

示されるように、入力音声信号605は、出力オーディオ信号601のプレゼンテーションに対してちょうどよいあるポイントで発生する。これは、たとえば、出力オーディオ信号601-603が、一連の合成された音声プロンプトであり、そして、入力音声信号605が、音声プロンプトのいずれか1つへのユーザの応答である場合である。同様に、出力オーディオ信号は、さらに、加入者ユニットに通信される合成されない音声信号でもありうる。とにかく、入力音声信号が検出され、そして、入力開始時間608が、入力音声信号605の開始をメモリアリズするために確立される。入力音声信号の開始を決定するための様々な技術がある。1つのこのような方法は、米国特許第4,821,325号に記述されている。入力音声信号の開始を決定するのに使用されるあらゆる方法は、1/20秒より多い分解で開始を決定できることが好ましい。

【0049】

入力音声信号の開始は、入力音声信号が出力オーディオ信号に対して検出された精密なポイントを表わすインターバル609を生じさせると、2つの連続する出力開始時間607, 610間のあらゆるタイムで検出されることが可能である。したがって、入力音声信号の開始は、任意に、出力オーディオ信号に続くサイレンス・消音の周期（すなわち、出力オーディオ信号が提供されないとき）を含むことができる出力オーディオ信号が生ずる間のあらゆるポイントで効果的に検出されることが可能である。別の方法として、出力オーディオ信号の終了に続く任意の長さのタイムアウト周期611は、出力オーディオ信号が生ずる終りを画定するのに使用されることができる。このように、入力音声信号の開始は、個々の出力オーディオ信号と関連することが可能である。有効な検出周期を確立する他のプロトコルが確立されることが可能であることがわかる。たとえば、一連の出力プロンプトが、互いにすべて関連される所では、有効な検出周期は、一連のプロンプトのための第1の出力開始時間で開始し、そして、連続した最後のプロンプトの後のタイムアウト周期でか、あるいは、連続に直ぐ続く出力オーディオ信号のための第1の出力開始時間で終る。

【0050】

入力開始時間を検出するのに使用されるのと同じ方法は、出力開始時間607, 610を確立するのに使用されることができる。これは、特に、出力オーディオ信号がインフラストラクチャから直接提供される音声信号であるこれらの例には真実である。出力オーディオ信号が、たとえば、合成されたプロンプト、あるいは、他の合成された出力である所では、出力開始時間は、以下にいっそう詳細に記述されるように、クロックサイクル、サンプル、境界、フレーム境界などの使用によりいっそう直接に確実にされることができる。とにかく、出力オーディオ信号は、入力音声信号が処理されることが可能なコンテキストを確立する。

【0051】

上記に述べられるように、各出力オーディオ信号は、識別と関連することができ、それによって、出力オーディオ信号間の微分を提供する。したがって、入力音声信号が、出力オーディオ信号のコンテキストに対して開始したときを決定する別の方法として、さらに、入力音声信号のコンテキストを記述する方法としてだけ、出力オーディオ信号の識別を使用することが可能である。これは、たとえば、入力音声信号が、出力オーディオ信号に対して開始する精密な時間を知るのに重要でない所では、単に、入力音声信号が、実際に、出力オーディオ信号が生ずる間のある時期に開始した場合である。このような出力オーディオ信号識別は、除外とは対照的に、入力オーディオ開始時間の決定に関連して使用さ

ることができることも分かる。

【0052】

入力開始時間および／または出力オーディオ信号識別が使用されるかどうかにかかわらず、本発明は、不確実な遅延特性を有するこれらのシステムにおける精密なコンテキスト決定を可能にする。上記に記述されるコンテキスト決定技術を実施し、そして、使用方法は、図7および図8を参照としてさらに示されている。

【0053】

図7は、出力オーディオ信号が生ずる間に入力音声信号を処理するため、加入者ユニット内で実施されることが好ましい方法を示している。たとえば、図7に示される方法は、格納されたソフトウェアルーチンおよび図2に示されるCPU201および／またはDSP202などの適切なプラットフォームによって実行されるアルゴリズムを使用して実施されることが好ましい。網でカバーされたコンピュータなどの他の装置が、図7に示されるステップを実施するのに使用されることが可能であること、そして、図7に示されるステップのいくつか、あるいは、すべてが、ゲートアレイ、カスタマイズされた集積回路などの特殊化ハードウェア装置を使用して実施されることが可能であることが分かる。

【0054】

出力オーディオ信号が生ずる間に、入力音声信号の開始が検出されたかどうか、ステップ701で、連続して決定される。もう一度、音声信号の開始を決定するための様々な技術は、技術上周知であり、そして、設計の選択の考慮すべき問題として本発明によって同時に採用されることができる。好ましい実施の形態において、入力音声信号の開始を検出する効果的な周期は、出力オーディオ信号の開始が、結果として生ずる出力オーディオ信号の開始でか、あるいは、カレント出力オーディオ信号の終結で開始されるタイムアウトタイマの終結でかのいずれかで終了するや否や開始する。入力音声信号の開始が検出されるとき、出力オーディオ信号によって確立されるコンテキストに対する入力開始時間が決定される。入力開始時間を決定するための様々な技術のいずれかが採用されることができる。1つの実施の形態において、リアルタイム参照は、CPU201（秒などのあらゆる便宜なタイムベース、あるいは、クロックサイクルを使用して）によって維持されることができ、それによって、特定の時間のコンテキストを確立する。この場合、入力開始時間は、出力オーディオ信号のコンテキストに対するタイム特質として表わされる。別の実施の形態において、可聴信号は、サンプル1つずつのベースで、再構成され、および／または、符号化される。たとえば、8kHzオーディオサンプリングレートを使用するシステムにおいて、各オーディオサンプルは、オーディオ入力、あるいは、出力の125マイクロセカンドに相当する。したがって、ちょうどよいあらゆるポイント（たとえば、入力開始時間）は、出力オーディオ信号（サンプルコンテキスト）の開始サンプルに対するオーディオサンプルのインデックスによって表わされることができる。この場合、入力開始時間は、出力オーディオ信号の第1のサンプルに対するサンプルインデックスとして表わされる。さらに別の実施の形態において、可聴信号は、フレーム1つずつのベースで、再構成され、各フレームは、マルチプルサンプル周期を備えている。この方法において、出力オーディオ信号は、フレームコンテキストを確立し、そして、入力開始時間は、フレームコンテキスト内のフレームインデックスとして表わされている。入力開始時間が表わされる方法にかかわらず、入力開始時間は、入力音声信号が出力オーディオ信号に関して開始したとき、精確に、分解の変化する程度で、メモライズする。

【0055】

いずれにせよ、入力音声信号の開始の検出から、入力音声信号は、任意に、ステップ703によって表わされるように、パラメータで表示された音声信号を提供するために分析されることが可能である。音声信号のパラメータによる表示化のための特定の技術は、図3に対して上記に論じられた。ステップ704で、いずれにせよ、入力開始時間は、入力音声信号に対応するために提供される。図7の方法が、無線電話加入者ユニット内で実施されるとき、このステップは、音声認識／合成サーバへの入力開始時間の無線伝送を含む。

【0056】

最後に、ステップ705で、情報信号は、任意に、少なくとも入力開始時間に応じて、そして、提供されるとき、パラメータで表示された音声信号に応じて、受信される。本発明のコンテキストにおいて、このような“情報信号”は、加入者ユニットが操作するデータ信号を含んでいる。たとえば、このようなデータ信号は、ユーザディスプレイを生成するディスプレイデータ、あるいは、加入者ユニットが自動的にダイヤルすることが可能な電話番号を備えている。他の実施例は、通常の技量の者によって難なく同一とみなしうる。本発明の“情報信号”は、さらに、加入者ユニット、あるいは、加入者ユニットに結合されるあらゆる装置の動作を制御するのに使用される制御信号を備えている。たとえば、制御信号は、ロケーションデータ、あるいは、ステータスアップデートを提供するために加入者ユニットに指示することが可能である。もう一度、通常の技量の者は、多数のタイプの制御信号を考案することができる。音声認識サーバによってこのような情報信号を提供する方法は、図9を参照としてさらに記述されている。とはいえ、入力音声信号を処理する別の実施の形態は、図8を参照としてさらに示されている。

10

【0057】

図8の方法は、図2に示されるCPU201および／またはDSP202などの適切なプラットフォームによって実行される格納されたソフトウェアルーチンおよびアルゴリズムを使用して、加入者ユニット内で実施されることが好ましい。網でカバーされたコンピュータなどの他の装置は、図8に示されるステップを実施するのに使用されることが可能であり、そして、図8に示されるステップのいくつか、あるいは、すべては、ゲートアレイ、あるいは、カスタマイズされた集積回路などの特殊化ハードウェア装置を使用して実施されることが可能である。

20

【0058】

出力オーディオ信号が生ずる間に、入力音声信号が検出されたかどうか、ステップ801で、連続して決定される。音声信号の存在を決定する様々な技術は、技術上周知であり、そして、設計の選択の考慮すべき問題として本発明によって同時に採用されることができる。図8に示される技術は、このような決定が、入力音声信号の存在を検出するステップに含まれることができるが、入力音声信号の開始を検出することに特に関係がないことに留意してください。

【0059】

ステップ802において、出力オーディオ信号に対応する識別が決定される。図6に関して上記に述べられるように、その識別は、出力オーディオ信号から独立するか、あるいは、出力オーディオ信号に組み込まれることができるかである。最も重要なことは、出力オーディオ信号識別は、出力オーディオ信号とすべての他の出力オーディオ信号とを独特に区別する必要がある。合成されたプロンプトなどの場合、これは、各そのような合成されたプロンプトに独特のコードを割り当てることによって、達成されることが可能である。リアルタイム音声の場合、インフラストラクチャ利用タイム特質などの反復性のないコードが使用されることができる。識別が表わされる方法にかかわらず、それは、加入者ユニットによって確かめ得る必要がある。

30

【0060】

ステップ803は、ステップ703に相当し、そして、さらに詳細に論じられる必要はない。ステップ804において、入力音声信号に対応する識別が提供される。図8の方法が、無線電話加入者ユニット内で実施されるとき、このステップは、音声認識／合成サーバへの識別の無線伝送を含んでいる。ステップ705と本質的に同一である方法で、加入者ユニットは、いずれにせよ、識別に基づき、ステップ805においてインフラストラクチャから情報信号を受信することが可能である。

40

【0061】

図9は、音声認識サーバによって情報信号を提供する方法を示している。述べられている点を除いては、図9に示される方法は、図4および図5に示されるCPU401および／またはリモート制御プロセッサ516などの適切な1つまたは複数のプラットフォームによって実行される格納されたソフトウェアルーチンおよびアルゴリズムを使用して実施

50

されることが好ましい。もう一度、他のソフトウェアおよび／またはハードウェア利用の実施は、設計の選択の考慮すべき問題として可能である。

【0062】

ステップ901において、音声認識サーバは、加入者ユニットで提供される出力オーディオ信号を生じさせる。これは、たとえば、独特の識別された音声プロンプトあるいは連続のプロンプトを合成することを加入者ユニットに指示する加入者ユニットに制御信号を提供することによって達成されることが可能である。別の方法として、たとえば、テキストから音声への分析器512によって提供されるパラメトリック音声表現は、音声信号の結果として生ずる再構成のために加入者ユニットに送信されることが可能である。本発明の1つの実施の形態において、リアルタイム音声信号は、音声認識サーバが存在する（音声認識サーバの介入を有して、あるいは、持たないで）インフラストラクチャによって提供される。これは、たとえば、加入者ユニットがインフラストラクチャを媒介として別の関係者との音声通信にかかわる場合である。

10

【0063】

加入者ユニットで出力オーディオ信号を生じさせるのに使用される技術にかかわらず、上記に記述されるタイプのコンテキスト情報（入力開始時間および／または出力オーディオ信号識別子）が、ステップ902で受信される。好ましい技術において、入力開始時間および出力オーディオ信号識別子の両方は、入力音声信号に対応するパラメータで表示された音声信号と共に提供される。

【0064】

20

ステップ903において、少なくともコンテキスト上の情報に基づいて、加入者ユニットに搬送される制御信号および／またはデータ信号を備える情報信号が決定される。もう一度、図5を参照すると、これは、ローカル制御プロセッサ508および／またはリモート制御プロセッサ516によって遂行されることが好ましい。最低限度で、コンテキスト上の情報は、出力オーディオ信号に対する入力音声信号のコンテキストを確立するのに使用される。コンテキストは、入力音声信号が、間隔を決定するのに使用される出力オーディオ信号に応答したかどうかを決定するのに使用されることが可能である。特定の出力オーディオ信号に対応する独特の識別子は、どの特定の出力オーディオ信号が入力音声信号のためのコンテキストを確立したかについて、あいまいさが可能であるコンテキストを確立するのに使用されることが好ましい。これは、たとえば、ユーザが電話帳のだれかに電話をかけようとする場合である。システムは、オーディオ出力を媒介として電話をかけるために、いくつか可能な人の名前提供することが可能である。ユーザは、“呼び出し”などのコマンドで出力オーディオに割り込むことが可能である。システムは、次に、独特の識別子および／または入力開始時間に基づいて、ユーザが割り込んだときどの名前が出力されたかを決定し、そして、その名前に関連する電話番号に電話をかけることが可能である。さらに、コンテキストを確立すると、パラメータで表示された音声信号は、提供される場合、認識された発話を提供するために分析されることが可能である。認識された発話は、何か入力音声信号に応答するのに必要とされる場合、順番に、制御信号、あるいは、データ信号を確かめるのに使用される。あらゆる制御、あるいは、データ信号がステップ903で決定される場合、それらは、ステップ904で、コンテキスト上の情報のソースに提供される。

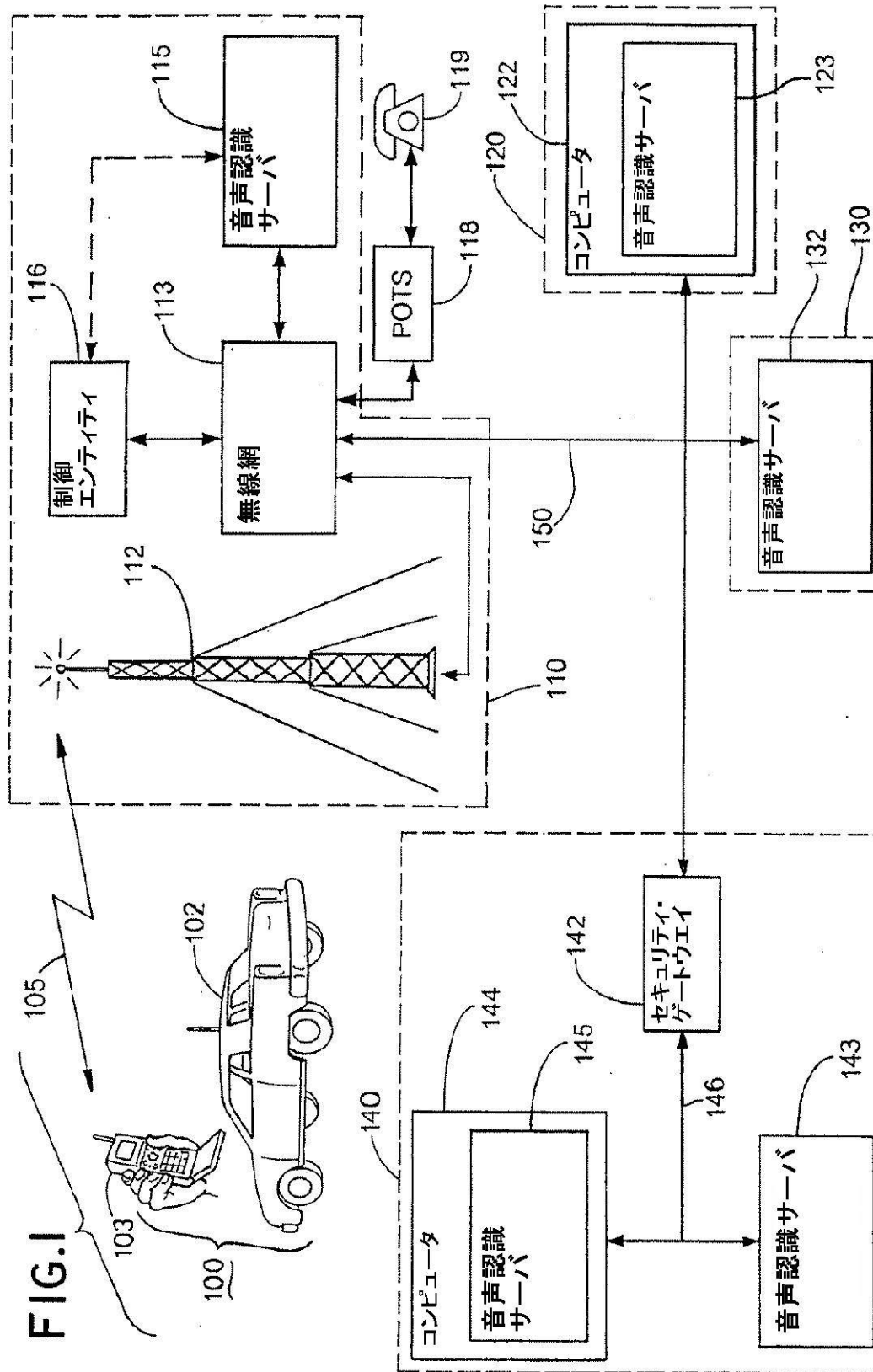
30

40

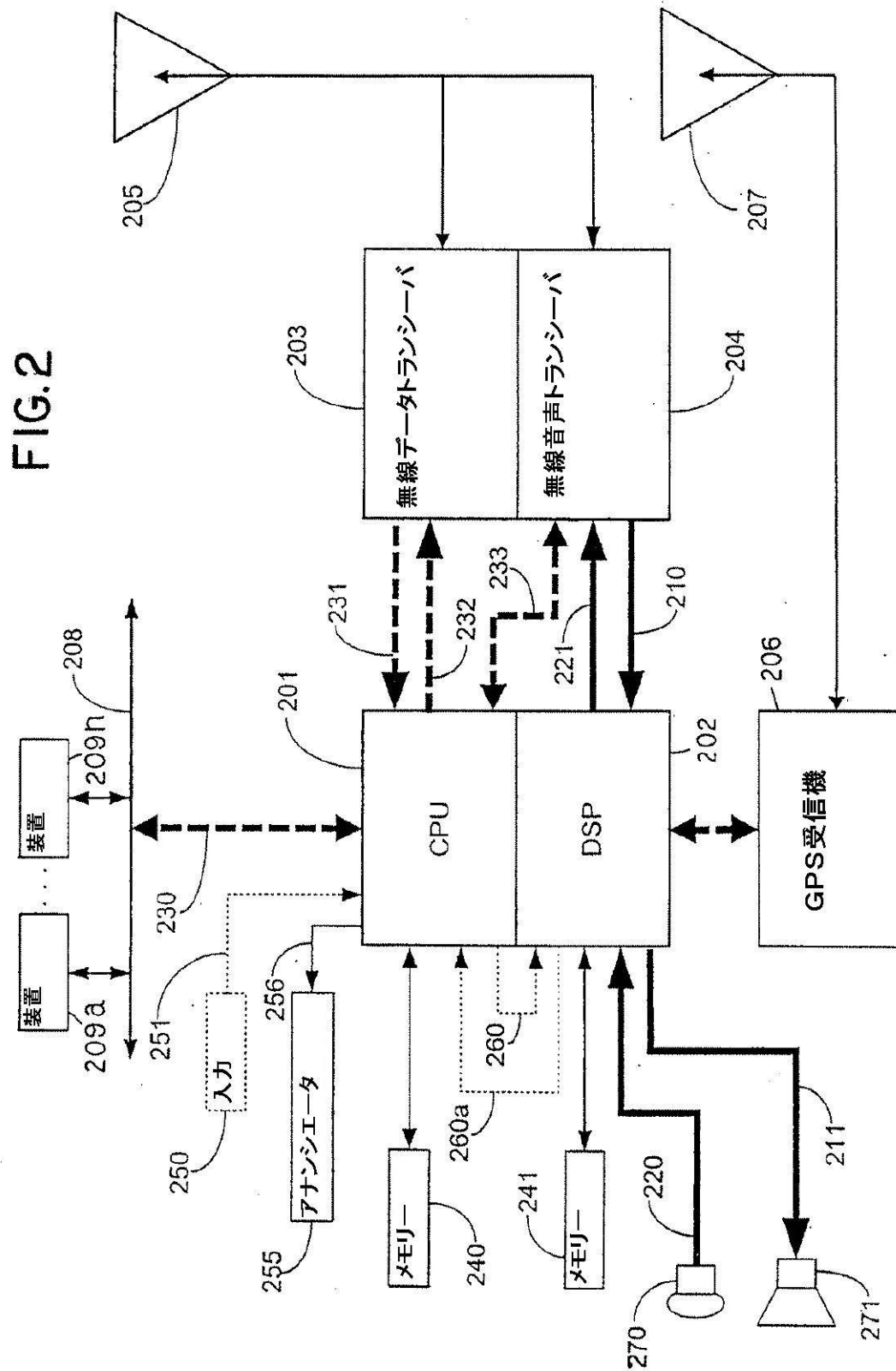
【0065】

上記に記述される本発明は、出力オーディオ信号が生ずる間に入力音声信号を処理する独特の技術を提供する。入力音声信号のための適切なコンテキストは、入力開始時間および／または出力オーディオ信号識別子の使用により確立される。このように、加入者ユニットに送信される情報信号は、入力音声信号に適切に応答するというより大きな確実性が、提供される。上記に記述されたことは、本発明の原理の応用について単に示しているものである。他の構成および方法は、本発明の精神と範囲とから逸脱することなく、当業者によって実施されることが可能である。

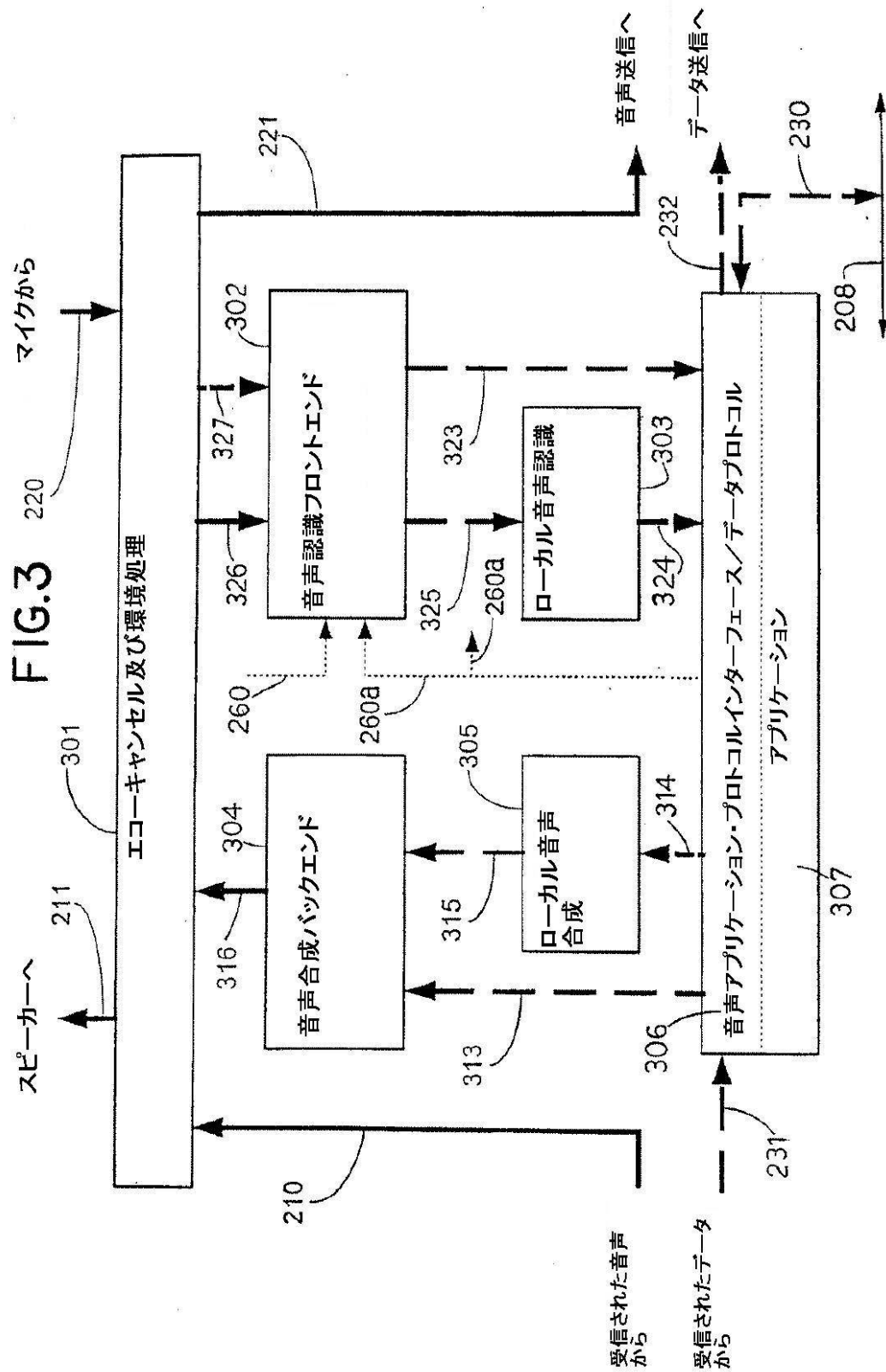
【図 1】



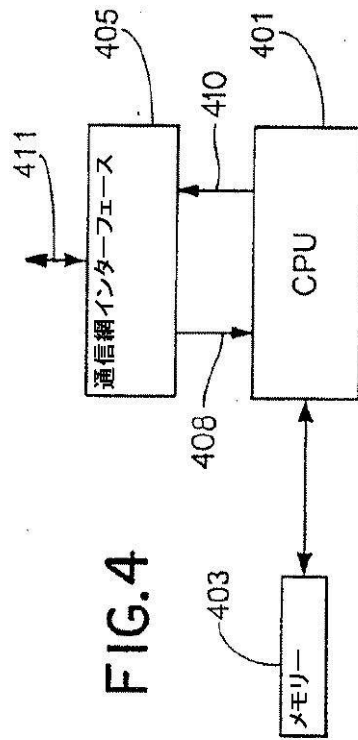
【図 2】



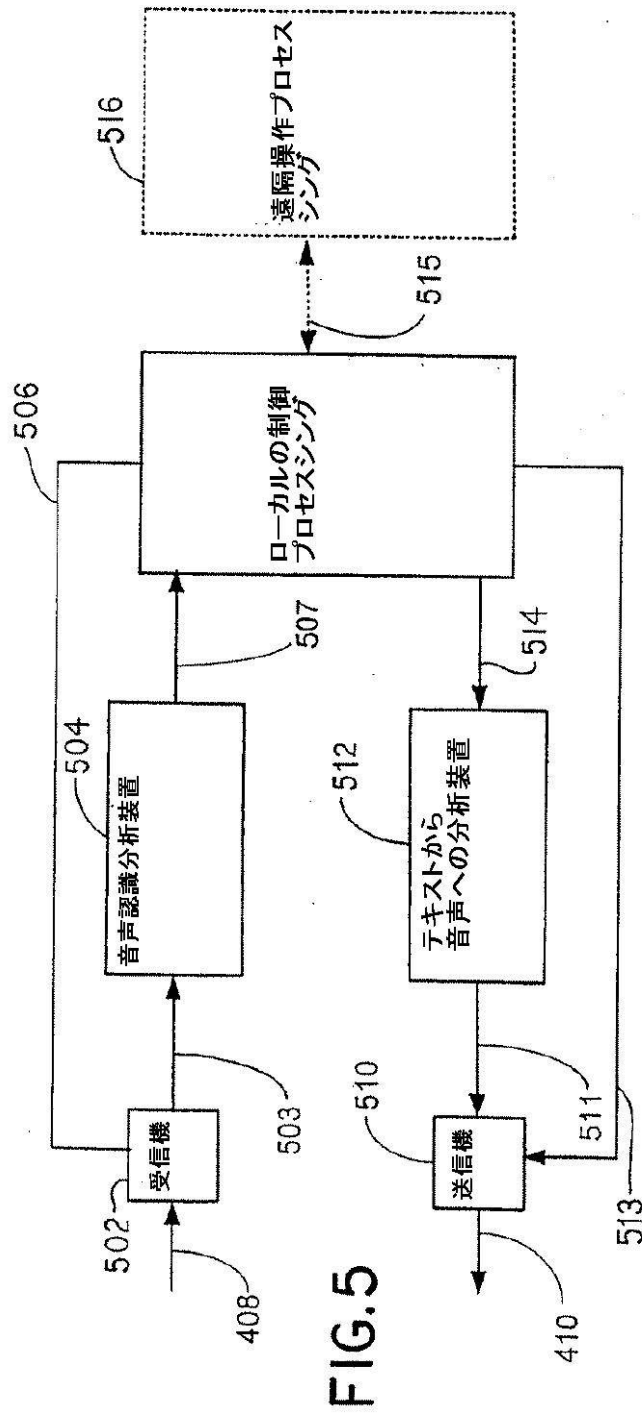
【図 3】



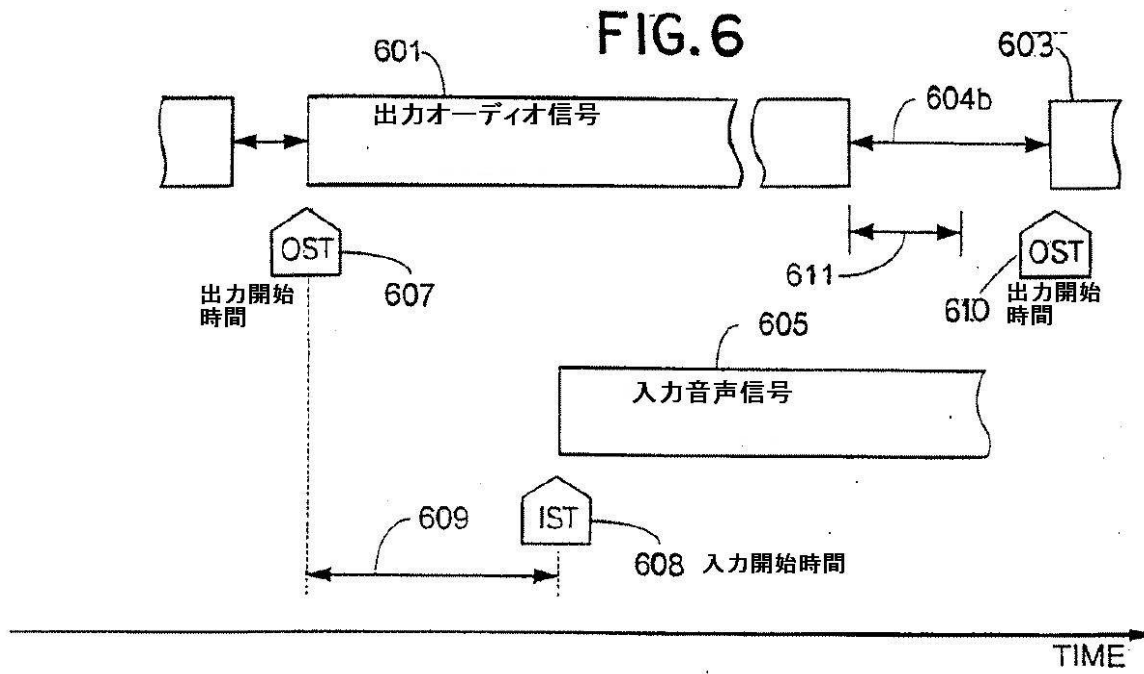
【図 4】



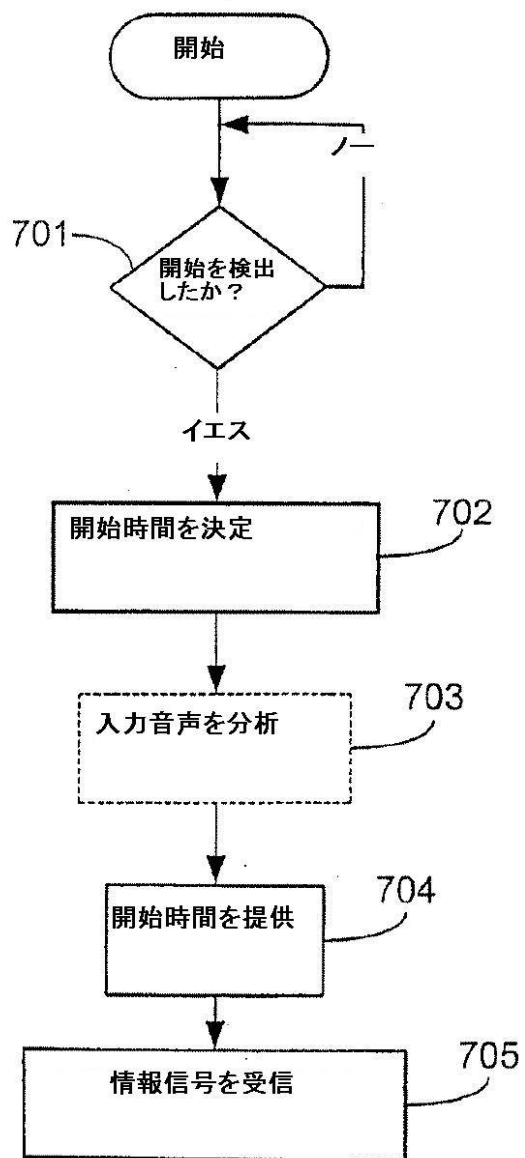
【図 5】



【図 6】

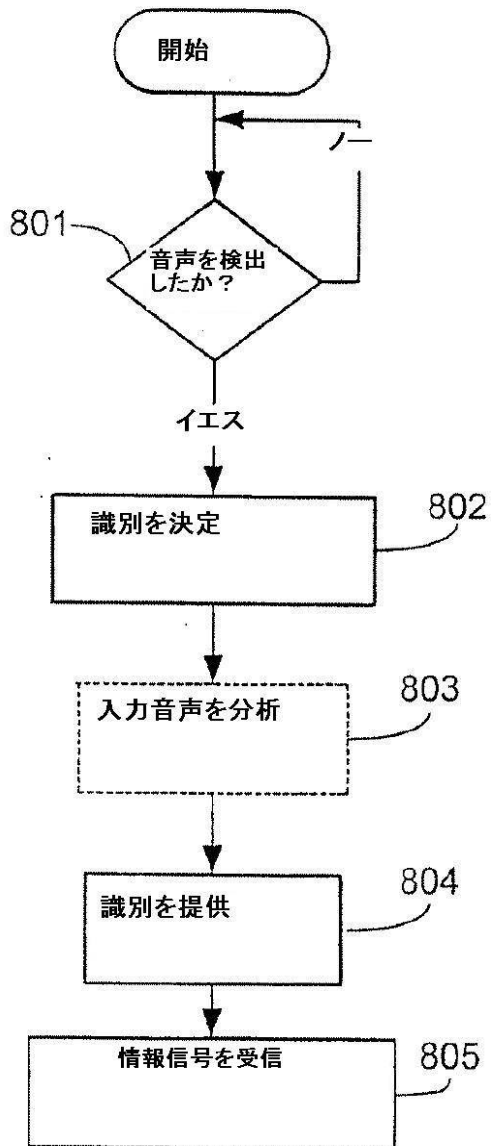


【図 7】

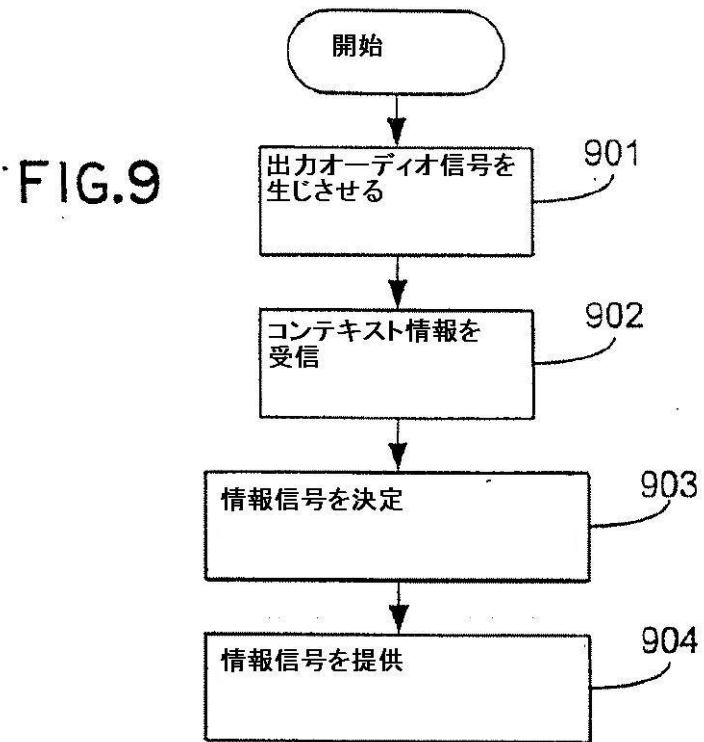


【図 8】

FIG.8



【図 9】



【外国語明細書】
2012137777000001.pdf