(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5039024号
(P5039024)
(45)発行日　平成24年10月3日(2012.10.3)　(24)登録日　平成24年7月13日(2012.7.13)

(51)Int.Cl.　　F I
H04M 3/493 (2006.01)　　H04M 3/493
G06F 13/00 (2006.01)　　G06F 13/00　353C
H04M 11/06 (2006.01)　　H04M 11/06

請求項の数 10　(全 11 頁)

(21)出願番号　特願2008-504770 (P2008-504770)
(86)(22)出願日　平成18年4月6日 (2006.4.6)
(65)公表番号　特表2008-537669 (P2008-537669A)
(43)公表日　平成20年9月18日 (2008.9.18)
(86)国際出願番号　PCT/EP2006/061380
(87)国際公開番号　WO2006/108795
(87)国際公開日　平成18年10月19日 (2006.10.19)
審査請求日　平成21年1月23日 (2009.1.23)
(31)優先権主張番号　0507148.5
(32)優先日　平成17年4月8日 (2005.4.8)
(33)優先権主張国　英国 (GB)

(73)特許権者　390009531
インターナショナル・ビジネス・マシーンズ・コーポレーション
ＩＮＴＥＲＮＡＴＩＯＮＡＬ　ＢＵＳＩＮＥＳＳ　ＭＡＳＣＨＩＮＥＳ　ＣＯＲＰＯＲＡＴＩＯＮ
アメリカ合衆国１０５０４　ニューヨーク州　アーモンク　ニュー　オーチャード　ロード
(74)代理人　100108501
弁理士　上野　剛史
(74)代理人　100112690
弁理士　太佐　種一
(74)代理人　100091568
弁理士　市位　嘉宏

最終頁に続く

(54)【発明の名称】多モード音声及びウェブ・サービスのための方法及び装置

(57)【特許請求の範囲】
【請求項１】
オーディオ・インターフェースとオーディオ・プロセッサ備えるコンピュータ・システムに適用され、前記オーディオ・インターフェースから前記オーディオ・プロセッサへのオーディオ接続を制御する方法であって、この方法は、
前記オーディオ・プロセッサ及び前記オーディオ・プロセッサを含むサーバのスイッチとのオーディオ・データのためのプロセッサ・リンクをセットアップするステップと、
オーディオ・インターフェース・セッションのセットアップに応答してインターフェース・セッションにおいて前記オーディオ・インターフェース及び前記スイッチとのオーディオ・データのためのインターフェース・リンクをセットアップするステップと、
前記インターフェース・セッション内でのオーディオ・インタラクションの開始を示す信号に応答して前記プロセッサ・リンクと前記インターフェース・リンクとを前記スイッチにより結合させるステップであって、これによりオーディオ・データが前記オーディオ・インターフェース・セッションと前記オーディオ・プロセッサとの間を流れることができるステップと、
前記オーディオ・インタラクションの終了を示す信号に応答して前記プロセッサ・リンクと前記インターフェース・リンクとを前記スイッチにより切り離すステップと、
前記インターフェース・セッションの終了に応答して前記インターフェース・リンクを停止させるステップと、
を含むことを特徴とする方法。

【請求項２】
前記コンピュータ・システムは、音声認識又はテキストから音声への変換或いはその両方のためのオプションの音声サーバを有する対話型音声応答システムであることを特徴とする請求項１に記載の方法。
【請求項３】
前記オーディオ・インターフェースは、多モード動作のための音声機能を有するＸＭＬブラウザであることを特徴とする請求項１又は２に記載の方法。
【請求項４】
前記インターフェース・リンクを前記プロセッサ・リンクと接続する遅延があるならば前記オーディオ・データをバッファリングするステップを更に含むことを特徴とする請求項１，２又は３に記載の方法。
【請求項５】
前記オーディオ・プロセッサとのプロセッサ・リンクをセットアップする前記ステップは、ＳＩＰプロトコルを用いてＲＴＰ接続を折衝するステップを含むことを特徴とする請求項１ないし４のうちのいずれかに記載の方法。
【請求項６】
オーディオ・インタラクションの開始をトリガーする前記信号は、多モード対話の開始を示す既存の信号であることを特徴とする請求項１ないし５のうちのいずれか１つに記載の方法。
【請求項７】
オーディオ・インタラクションの終了をトリガーする前記信号は、多モード対話におけるフィールドの同期化を示す既存の信号であることを特徴とする請求項１ないし６のうちのいずれか１つに記載の方法。
【請求項８】
オーディオ・インターフェースからオーディオ・プロセッサへのオーディオ接続を制御するためのスイッチであって、このスイッチは、
前記オーディオ・プロセッサとのオーディオ・データのためのプロセッサ・リンクをセットアップするための手段と、
オーディオ・インターフェース・セッションのセットアップに応答してインターフェース・セッションにおいて前記オーディオ・インターフェースとのオーディオ・データのためのインターフェース・リンクをセットアップするための手段と、
前記インターフェース・セッション内でのオーディオ・インタラクションの開始を示す信号に応答して前記プロセッサ・リンクと前記インターフェース・リンクとを結合させるための手段であって、これによりオーディオ・データが前記オーディオ・インターフェース・セッションと前記オーディオ・プロセッサとの間を流れることができる手段と、
前記オーディオ・インタラクションの終了を示す信号に応答して前記プロセッサ・リンクと前記インターフェース・リンクとを切り離すための手段と、
前記インターフェース・セッションの終了に応答して前記インターフェース・リンクを停止させるための手段と、
を含むことを特徴とするスイッチ。
【請求項９】
多モード・ウェブ・サーバであって、この多モード・ウェブ・サーバは、
ＸＨＴＭＬ及び音声文書データベースと、
ＸＨＴＭＬ及び音声フィルタと、
音声ＸＭＬブラウザと、
対話型音声応答サーバと、
音声サーバと、
クライアント・オーディオ・インターフェースから前記対話型音声応答サーバ（ＩＶＲ）へのオーディオ接続を制御するためのスイッチと、を含んでおり、前記スイッチは、
前記ＩＶＲとのオーディオ・データのためのＩＶＲリンクをセットアップするための手

段と、
クライアント・オーディオ・インターフェース・セッションのセットアップに応答してインターフェース・セッションにおいて前記クライアント・オーディオ・インターフェースとのオーディオ・データのためのクライアント・インターフェース・リンクをセットアップするための手段と、
前記クライアント・インターフェース・セッション内でのオーディオ・インタラクションの開始を示す信号に応答して前記ＩＶＲリンクと前記クライアント・インターフェース・リンクとを結合させるための手段であって、これによりオーディオ・データが前記クライアント・オーディオ・インターフェース・セッションと前記ＩＶＲとの間を流れることができる手段と、
前記クライアント・オーディオ・インタラクションの終了を示す信号に応答して前記ＩＶＲリンクと前記クライアント・インターフェース・リンクとを切り離すための手段と、
前記クライアント・インターフェース・セッションの終了に応答して前記クライアント・インターフェース・リンクを停止させるための手段と、
を含むことを特徴とする多モード・ウェブ・サーバ。
【請求項１０】
オーディオ・インターフェースからオーディオ・プロセッサへのオーディオ接続を制御するためのコンピュータ・プログラムであって、前記コンピュータ・プログラムは、コンピュータにおいてロードされて実行されたときに前記コンピュータに
前記オーディオ・プロセッサ及び前記オーディオ・プロセッサを含むサーバのスイッチとのオーディオ・データのためのプロセッサ・リンクをセットアップするステップと、
オーディオ・インターフェース・セッションのセットアップに応答してインターフェース・セッションにおいて前記オーディオ・インターフェース及び前記スイッチとのオーディオ・データのためのインターフェース・リンクをセットアップするステップと、
前記インターフェース・セッション内でのオーディオ・インタラクションの開始を示す信号に応答して前記プロセッサ・リンクと前記インターフェース・リンクとを前記スイッチにより結合させるステップであって、これによりオーディオ・データが前記オーディオ・インターフェース・セッションと前記オーディオ・プロセッサとの間を流れることができるステップと、
前記オーディオ・インタラクションの終了を示す信号に応答して前記プロセッサ・リンクと前記インターフェース・リンクとを前記スイッチにより切り離すステップと、
前記インターフェース・セッションの終了に応答して前記インターフェース・リンクを停止させるステップと、
を実行させるコンピュータ・プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、多モード音声及びウェブ・サービスのための方法及び装置に関する。
【背景技術】
【０００２】
装置が小型化すると、キーボード及びスタイラス以外のインタラクションのモードが必要になる。特に、携帯電話機及びＰＤＡのような小型ハンドヘルド装置は、多くの機能の役に立ち、多様なタスクを処理する充分な処理能力を包含する。現在及び将来の装置は多モードのアクセス方法の使用から大いに利益をこうむるであろう。
【０００３】
多チャネル・アクセスは、電話機、ラップトップ又はＰＤＡのような多様な方法又はチャネルからエンタープライズ・データ及びアプリケーションにアクセスする能力である。例えば、ユーザは、オフィス又は家庭にいるときにインターネット・ブラウザを用いてウェブ上の自分の銀行口座残高にアクセスすることができ、また路上にいるときに音声認識及びテキストから音声への変換を用いてダム・フォーン（dump phone）を介して同じ情報

にアクセスすることができる。
【０００４】
対照的に、多モード・アクセスは同じインタラクション又はセッションにおいて多様なモード又はチャネルを組み合わせる能力である。入力の方法は、音声認識、キーボード、タッチ・スクリーン、及びスタイラスを含む。状況及び装置により、入力モードの組み合わせは小型装置の使用を容易にするであろう。例えば、ＰＤＡ上のウェブ・ブラウザでは、軽く叩くことにより又は音声入力を与えることにより、アイテムを選択することができる。同様に、音声又はスタイラスを用いて情報をフィールドに入力することができる。多モード技術では、装置上の情報は表示され且つ発音され得る。
【０００５】
ＸＨＴＭＬ＋音声を用いる多モード・アプリケーションは、今日の音声ＸＭＬベースの音声アプリケーション及びＸＨＴＭＬベースのビジュアル・アプリケーションから、これらの環境と多モードのものとの両方に役立つ単一のアプリケーションへの自然な移行経路を提供する。多モード・アプリケーションは、２つのチャネル（１つはグラフィカル・ユーザ・インターフェースのためのもので、他の一つは音声のためのもの）をセットアップすることによって音声インターフェース及びグラフィカル・ユーザ・インターフェース・インタラクションを統合する。執筆時に、ＸＨＴＭＬ＋音声（Ｘ＋Ｖ）　Ｐｒｏｆｉｌｅ　１．２が２００４年3月１６日にｗｗｗ．ｖｏｉｃｅｘｍｌ．ｏｒｇで公開された。
【０００６】
リモート音声処理を伴う多モード・ブラウザの１つの公知インプリメンテーションでは、クライアントと音声サーバとの間に音声チャネルがセットアップされて、Ｘ＋Ｖセッション内での音声インタラクションの持続時間にわたって音声データを運ぶために割り当てられる。音声チャネルは音声インタラクション後に切断され、Ｘ＋Ｖセッションは継続する。Ｘ＋Ｖセッション内の各々のインタラクションのために別々の音声チャネルがセットアップされなければならない。なぜならば、そのようにすればＸ＋Ｖセッションがアイドルであるときにサーバ上の高価な音声リソースを消費せずにすむからである。
【発明の開示】
【発明が解決しようとする課題】
【０００７】
各々の音声インタラクションのために音声チャネルをセットアップし閉じることには、現在のプロトコル（ＳＩＰ及びＲＴＰ）を用いて音声チャネルを開き閉じるためにかかる時間に起因して各音声インタラクションの応答時間を増大させるという不利益がある。付加される待ち時間は、装置とサーバとの間で利用可能なネットワーク帯域幅の直接の関数である。これは、低速インターネット接続のような狭帯域幅ネットワーク及び低速無線ネットワークにおいて問題を生じさせる。例えば、３Ｇより前の無線ネットワークにおけるネットワーク帯域幅は限られている。
【課題を解決するための手段】
【０００８】
本発明の第１の側面に従って、オーディオ・インターフェースからオーディオ・プロセッサへのオーディオ接続を制御する方法が提供され、この方法は、オーディオ・プロセッサとのオーディオ・データのためのプロセッサ・リンクをセットアップするステップと、オーディオ・インターフェース・セッションのセットアップに応答してインターフェース・セッションにおいてオーディオ・インターフェースとのオーディオ・データのためのインターフェース・リンクをセットアップするステップと、インターフェース・セッション内でのオーディオ・インタラクションの開始を示す信号に応答してプロセッサ・リンクとインターフェース・リンクとを結合させるステップであって、これによりオーディオ・データがオーディオ・インターフェース・セッションとオーディオ・プロセッサとの間を流れることができるステップと、オーディオ・インタラクションの終了を示す信号に応答してプロセッサ・リンクとインターフェース・リンクとを切り離すステップと、インターフェース・セッションの終了に応答してインターフェース・リンクを取り壊すステップと、

を含む。

【０００９】

仮想化ユニットと装置との一定の接続を維持しながらパー・ターン・ベースで（ｏｎａ　ｐｅｒ　ｔｕｒｎ　ｂａｓｉｓ）使用されるようにＩＶＲポートを仮想化することは確かに新規である。狭帯域幅ネットワーク上でＳＩＰプロトコルを用いてＲＴＰチャネルをセットアップすることは、ハンドシェークにおいてセッション番号とＩＶＲポート番号とを識別する必要があるので、過度の容量を使用する。更に、新しいコールのためにＳＩＰコールが使用され、ＲＴＰチャネルがセットアップされる前にリソース（音声ＸＭＬブラウザ、スピーチ・エンジンなど）が割り当てられなければならない。新しい接続プロトコルは同じ接続を達成するために単一の信号トリガーを使用し、その信号は現存するセッション・チャネルで送られるのでセッション番号は不要である。ポートは既に割り当てられているので、ポート番号、セッション、或いはボイス・エンジン・リソース割り当ては不要である。必要な唯一の割り当ては、利用可能な設定済みＲＴＰチャネルである。更に、新しい接続プロトコルは仮想回路をトリガーするために現存する信号を使用するので、セットアップ・オーバーヘッドは全く存在しない。

【００１０】

接続可能で切断可能なオーディオ・チャネル（仮想回路）を作ることの利点は、１）同時仮想回路の数が利用可能な物理的ＩＶＲポート以下であるときにシステムがインタラクションの応答時間に実際上待ち時間を付け加えないこと、２）クライアントとＩＶＲとにおいてコード変更が不要であること、３）仮想回路の数が利用可能な物理的ＩＶＲポートの数を上回るときに、システム性能が潔く低下すること、４）装置とスイッチとの間に永久的音声回路があるので、どんなユーザ・オーディオ入力でもスイッチにおいて捕獲され、後に、仮想回路が利用可能になるとＩＶＲに再生され得ること、である。

【００１１】

仮想音声チャネル接続及び切断のコントロールを中間コントローラに与えることは、クライアント及びサーバの両方が接続及び切断をトリガーすることを可能にする。

【００１２】

オーディオ・プロセッサは、音声認識、又はテキストから音声への変換、或いはその両方のためのオプションの音声サーバを有する対話型音声応答システムであり得る。

【００１３】

オーディオ・インターフェースは、音声ＸＭＬブラウザ又は多モード動作のための音声機能を有するＸＭＬブラウザであり得る。

【００１４】

この方法は、もしインターフェース・リンクをプロセッサ・リンクと接続する遅延があればオーディオ・データをバッファリングするステップを更に含む。

【００１５】

オーディオ・プロセッサとのプロセッサ・リンクをセットアップするステップは、ＳＩＰプロトコルを用いてＲＴＰ接続を折衝するステップを含む。

【００１６】

オーディオ・インタラクションの開始をトリガーする信号は、多モード対話の開始を示す既存の信号である。

【００１７】

オーディオ・インタラクションの終了をトリガーする信号は、多モード対話におけるフィールドの同期化を示す既存の信号である。

【００１８】

次に、添付図面を参照して、単に例を用いて、本発明の実施態様を説明する。

【発明を実施するための最良の形態】

【００１９】

好ましい実施態様のシステムは、ネットワーク１１を通して多モード・サーバ１２に接続された複数のクライアント１０．１～１０．ｎを含む。各クライアント１０は、ＸＨＴ

ＭＬブラウザ１４と、インタラクション・マネージャ１６と、オーディオ・ハードウェア１８とを含む。サーバ１２は、Ｘ＋Ｖ文書データベース２０と、ＸＨＴＭＬ＋音声（Ｘ＋Ｖ）フィルタ２２と、音声ＸＭＬ（ＶＸＭＬ）ブラウザ２４と、対話型音声応答システム（ＩＶＲ）２６と、音声エンジン２８と、スイッチ３０とを含む。

【００２０】

ネットワーク１１は、ＸＨＴＭＬ（Ｘ）チャネル３２と、ＣＴＲＬ（制御）チャネル３４と、ＳＩＰ（セッション初期化プロトコル）チャネル３６と、ＲＴＰ（リアルタイム・プロトコル）チャネル３８とを伝える。Ｘチャネル３２は、Ｘ＋ＶインタラクションのためのＸＨＴＭＬアプリケーションを伝える。好ましい実施態様では、ＣＴＲＬチャネル３４は、対応するＸＨＴＭＬフィールドとＶＸＭＬフィールドとを同期化するための同期信号を伝える。Ｘ＋Ｖは、ＸＨＴＭＬ形式とＶＸＭＬ形式とにおける対応するフィールドの概念を定義する。多モード・インタラクションが行われ得るときに同期データ・イベント信号が送られ、そのインタラクションの後に１つのＸＭＬフィールドを対応するＸＭＬフィールドと同期化するために同期データ信号が送られる。ＲＴＰチャネル３８は、オーディオのためのデータを伝える。ＳＩＰチャネル３６は、ＲＴＰチャネルをセットアップするために使われる。ＣＴＲＬ、ＲＴＰ及びＳＩＰチャネルは、クライアント１０をＶＸＭＬブラウザ２４／ＩＶＲ２６に直接にではなくてスイッチ３０を介して接続する。スイッチ３０は、リソースを殆ど消費しないので、利用可能なＩＶＲ容量より大きな、多数の同時並行のクライアント１０をサポートすることができる。

【００２１】

好ましい実施態様は１つのクライアント１０と１つのＲＴＰチャネル３８とに関して記述されるけれども、本発明の利点は音声チャネルより多数のクライアントがあるときに明白になる。

【００２２】

クライアントＸＨＴＭＬブラウザ１４は、Ｘチャネル３２を介して受信されたＸＨＴＭＬ文書を解釈する。ＸＨＴＭＬブラウザ１４は、ＶＸＭＬブラウザとインタラクションし、オーディオ・ハードウェアとインタラクションする追加された機能を有する既知のＸＨＴＭＬブラウザ１４である。

【００２３】

インタラクション・マネージャ１６は、ＣＴＲＬ（制御）チャネル３４及びＳＩＰ（セッション初期化プロトコル）チャネル３６で制御情報を送受信することによってＸＨＴＭＬブラウザ１４とＶＸＭＬブラウザ２４との間のインタラクションを制御する。ＸＨＴＭＬブラウザ１４とＶＸＭＬブラウザ２４との間のインタラクションの重要な側面は、音声インタラクションの直前にインタラクション・マネージャ１６から送られる同期イベント信号と音声インタラクション後に送られる同期データ信号とである。同期イベント信号はＶＸＭＬブラウザ２４における音声対話をトリガーする。同期データ信号は、音声インタラクション後に対応するフィールド・データを同期化する。

【００２４】

オーディオ・コンテンツは、インタラクション・マネージャ１６の制御下でオーディオ・ハードウェアによりＲＴＰ（リアルタイム・プロトコル）チャネル３８で送受信される。

【００２５】

Ｘ＋Ｖ文書データベース２０は、Ｘ＋Ｖ文書を記憶し、それらをリクエストに応じてＸ＋Ｖフィルタ２２に送る。

【００２６】

Ｘ＋Ｖフィルタ２２は、ＸＨＴＭＬブラウザ１４からのリクエストに応じてＸ＋Ｖ文書データベース２０からＸ＋Ｖ文書を取得する。Ｘ＋Ｖ文書は、ＸＨＴＭＬコンポーネント部分とＶＸＭＬコンポーネント部分とにフィルタリングされる。ＸＨＴＭＬコンポーネント部分はＸＨＴＭＬブラウザ１４に送られ、ＶＸＭＬコンポーネント部分はＶＸＭＬブラウザ２４に送られる。ＸＨＴＭＬコンポーネント部分は、対応するＸフィールド及びＶフ

ィールドを示し、ＸＨＴＭＬの、ＶＸＭＬブラウザとのインタラクションが必要とされる部分を標示する音声ハンドラを含む。

【００２７】

ＶＸＭＬブラウザ２４は在来のＶＸＭＬブラウザである。ＶＸＭＬブラウザは、Ｘ＋Ｖセッションの間にＸ＋Ｖ文書のＶＸＭＬコンポーネント部分を用いて音声インタラクションを行うリクエストを受信する。ＶＸＭＬブラウザは、そのＸ＋Ｖセッション内の音声インタラクションを管理する。個々の音声機能の処理はＩＶＲに渡される。

【００２８】

好ましい実施態様ではＶＸＭＬブラウザ２４とＸ＋Ｖフィルタとはサーバ１２内に示されているが、これらはクライアント１０において実現されてもよい。

【００２９】

ＩＶＲ２６は音声インタラクションを処理する。ブラウズされたＶＸＭＬ文書部分からのリクエストと、ブラウズされたＶＸＭＬ文書部分への入力として受信されたデュアル・トーン・マルチフリーケンシー信号（ＤＴＭＦ）とに応じて、予め記録されたプロンプトが再生され得る。ＩＶＲ２６は、音声エンジン２８ともインターフェースする。ＩＶＲ２６は、スイッチ３０を通してＳＩＰチャネル３６及びＲＴＰチャネル３８に接続する。

【００３０】

音声エンジン２８は、ＩＶＲ２６のために音声認識入力及びテキストから音声への変換出力を実行する。

【００３１】

スイッチ３０は、マルチプレクサ４０とバッファ４２とを含む。マルチプレクサ４０は、（自分自身と、数千個に及び得るモバイル装置クライアントとの間の）多数のクライアント音声リンクのうちの１つを（自分自身とＩＶＲとの間の）より少数のＩＶＲ音声リンクのうちの１つに接続する。スイッチ３０は、制御チャネル（ＣＴＲＬ）上の同期イベント信号を傍受すると、クライアントとＩＶＲポートとの間に仮想音声回路を生じさせるために、対応するクライアント音声リンクをＩＶＲ音声リンクに接続する。

【００３２】

生成されると、同期イベント信号は、音声インタラクションの処理のためにＶＸＭＬブラウザ２４に渡される。ＶＸＭＬブラウザ２４は、ＶＸＭＬフィールドを更新することができ、その後に、プロンプトを再生し且つ仮想回路を介して音声入力を受け取るようにＩＶＲ２６に命じることができる。仮想回路は、単一の音声インタラクション（’会話ターン（ｃｏｎｖｅｒｓａｔｉｏｎａｌ　ｔｕｒｎ）’）の持続時間の間だけ存続し、そのインタラクションの終了は同期データ信号によって信号される。同期データ信号の開始時に、スイッチ３０は仮想回路を切断する。仮想回路のための音声リソースは、その後、切断直後に他のクライアント装置により再使用され得る。スケーラビリティを得るために、スイッチは高速ネットワークによりＩＶＲに接続され得る。

【００３３】

好ましい実施態様では、ＲＴＰチャネル３８はＸ＋Ｖセッションの間に第１音声インタラクションのときに開かれる。或いは、スイッチは、Ｘ＋Ｖセッションの開始時に第１音声インタラクションの前にＲＴＰセッションを開くように構成され得る。予め開かれたＲＴＰチャネルの間、スイッチ３０とＩＶＲ２６との間の音声チャネルは接続されたままである。このことは、新しい仮想回路がセットアップされなければならないときに、それが単にスイッチ内でのルーティングをセットアップする問題となり、ＩＶＲとの付加的な物理的コールのセットアップが行われなくても良いということを意味する。実際には、これは、システムの応答性に無視できる待ち時間を付け加えることを意味する。

【００３４】

全ての物理的ＩＶＲポートが使用されている場合（すなわち、余りにも多数の装置が同時に音声インタラクションを実行しようと試みている場合）、スイッチ３０はオーディオをバッファ４２に格納することができる。後にＩＶＲポートが利用可能になったとき、仮想回路がセットアップされ、バッファリングされていたオーディオがＩＶＲに対して再生

されてインタラクションを完了させる（例えば、音声メッセージを残す）。もちろん、もしこれが余りに頻繁に起こるならば、それはシステムが１００％稼動していて追加のＩＶＲポートが必要とされることを意味する。

【００３５】

本実施態様の方法を具体的に説明するために、次に図２のイベント図と例とを参照して２つの代表的な音声インタラクションのイベントが記述される。

【００３６】

ハンドヘルドＰＤＡを用いてユーザが飛行便の見積もり到着時刻を見出すためにフライト情報ウェブサイトにネットサーフィンする。ＧＵＩは、２つのフィールド、すなわちフライト番号及びフライトの日付、を有する照会書式を表示する。ユーザは第１フィールドに焦点を合わせ、オーディオ・プロンプト“フライト番号を入力してください”が再生される。ユーザは、自分のＰＤＡのキーボードを用いてフライト番号を入力する（このインタラクションは、以下の’ウェブ同期’に関連して記載される）。その後、ユーザは次のフィールドに焦点を合わせ、オーディオ・プロンプト’日付を入力してください’が聞こえる。このときユーザはオーディオ・ハードウェアを用いてマイクロホンに’今日’と言う。その音声データが処理されて、その日の日付が自動的に日付フィールドに挿入される。このインタラクションは以下の’音声同期’イベント・シーケンスにおいて記述される。この例では、その２つのフィールドが満たされ、フライト情報サイトは見積もり到着時刻をテキストの形で又は音声で或いは両方で答える

【００３７】

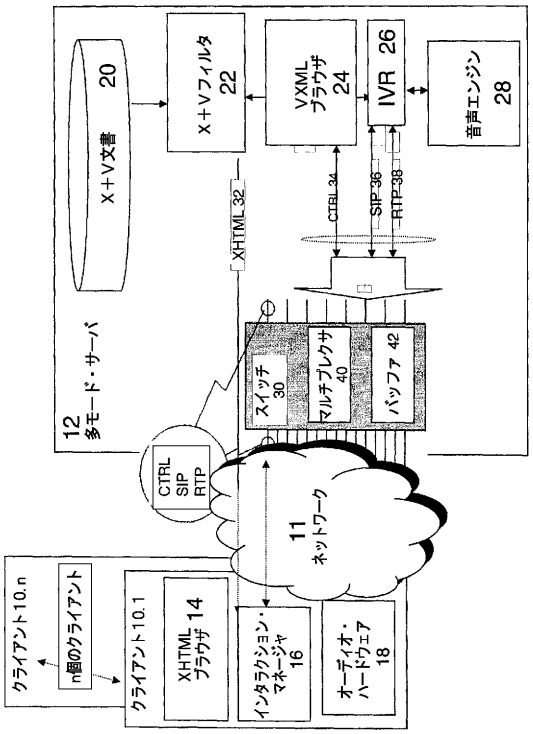
初めにクライアント１０は音声チャネル３８がセットアップされるように要求５０し、音声チャネル３８をセットアップするためにワン・タイムＳＩＰリクエストがＳＩＰチャネル３６でＩＶＲ２６に送られる。しかし、クライアント２０からサーバへの音声チャネル３８の代わりに、スイッチ３０とクライアントとの間の１つのクライアント・リンクがセットアップ５２され、スイッチとサーバとの間の少なくとも１つのサーバ・リンクがセットアップ５４される。このクライアント・リンクとサーバ・リンクとは、スイッチにより接続された後に切断され得る接続可能な音声回路を形成する。サーバ・リンクの数はＩＶＲ上のポートの数により制限され、クライアント・リンクの数はスイッチ３０のインプリメンテーションにより制限される。

【００３８】

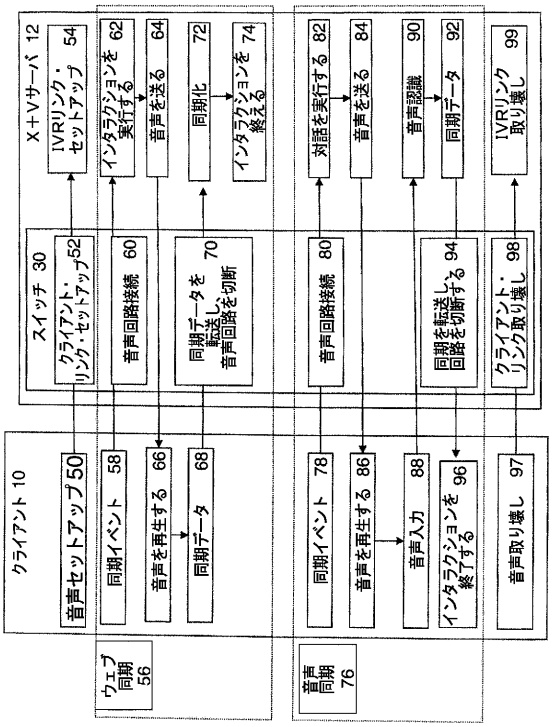
図２において、ウェブ同期インタラクション５６は、ＸＨＴＭＬフィールドがクライアントＸＨＴＭＬブラウザにおいて焦点に置かれて、対応するＶＸＭＬフィールドと同期化されるインタラクションにおけるイベントを示す。ＸＨＴＭＬフィールドは対応するＶＸＭＬフィールドを有するので、同期イベント信号がクライアントからサーバに送られる５８。同期イベントはスイッチにより傍受され、スイッチは、初めに、同期イベントをＸ＋Ｖサーバ上のＶＸＭＬブラウザに転送する前に音声回路を接続６０する。ＶＸＭＬブラウザは、同期イベントを受信したとき、対応するＶＸＭＬコンポーネントを見出して、それらのコンポーネントを音声インタラクションとして実行６２する。この例では、音声インタラクションは、新たに接続された音声チャネルを介して音声プロンプトを送る６４。クライアントは、その音声プロンプトを受信し、それをオーディオ・ハードウェアを用いて再生６６する。しかし、この例では、ユーザは応答を入力するためにＸＨＴＭＬインターフェースを用いてその音声プロンプトに応答することを選択するので、即時のＶＸＭＬ応答は無い。ユーザは、グラフィカル・インターフェースを用いてデータをＸＨＴＭＬフィールドに入力する。ＸＨＴＭＬフィールドは対応するＶＸＭＬフィールドを有し、同期化が続く。インタラクション・マネージャは同期データ信号をＣＴＲＬチャネルでＸ＋Ｖサーバに送り６８、これは、入力されたＸＨＴＭＬデータを含む同期データ信号である。この同期データ信号は、ＶＸＭＬフィールドへの入力のためにＶＸＭＬブラウザへ転送される前にスイッチにより傍受される。ＣＴＲＬ信号は同期データ信号であるので、スイッチは仮想音声回路を切断７０するように承知する。ＶＸＭＬブラウザは、ＶＸＭＬデータを同期化７２し、音声インタラクションを終了７４させる。

【００３９】
図２の音声同期インタラクション７６は、データがサーバＶＸＭＬブラウザでＶＸＭＬフィールドに入力されてクライアント・ブラウザで対応するＸＨＴＭＬフィールドと同期化されるインタラクションにおけるイベントを示す。グラフィカル・インターフェースにおいて対応するＶＸＭＬフィールドを有するＸＨＴＭＬフィールドに焦点が合わされると、同期イベント信号がスイッチを介してインタラクション・マネージャからサーバに送られる７８。同期イベントはスイッチにより傍受され、スイッチはクライアント・リンクをＩＶＲリンクに接続８０して音声回路を形成する。同期イベントはＸ＋Ｖサーバ上のＶＸＭＬブラウザに転送される。ＶＸＭＬブラウザは、同期イベントを受信すると、対応するＶＸＭＬインタラクションを実行８２する。この例では、音声インタラクションは、新たに接続された音声チャネルを介して音声プロンプトを送る８４。クライアントは、音声プロンプトを受信して、それをオーディオ・ハードウェアを用いて再生８６する。この例では、ユーザは、音声プロンプトに応答してオーディオ・ハードウェアに例えば’今日’などの応答を話す。音声データは、スイッチにより生成された仮想音声回路を通してＩＶＲに送られる８８。ＩＶＲは、音声データを認識９０のために音声エンジンに送り、その結果としてのテキスト・データをＶＸＭＬフィールドに入力する。ＶＸＭＬフィールドは対応するＸＨＴＭＬフィールドを有し、同期データ信号がＣＴＲＬチャネルに沿ってインタラクション・マネージャに送られる９２。同期データ信号をインタラクション・マネージャに転送した後、スイッチは音声回路を切断９４する。クライアントはインタラクションを終了９６する。Ｘ＋Ｖサーバはフィールド内のデータを処理し、結果を出力する。
【００４０】
クライアント装置がオフに切り替えられる寸前になって初めて音声取り壊し信号が送られる９７。この信号に応じて、スイッチはクライアント・リンクを取り壊し９８、サーバはＩＶＲリンクを取り壊す９９。
【００４１】
要約すると、本実施態様は、音声サーバの位置を突き止め、それを一時的に割り当て、それに“今日の６８３便は何時に到着する予定ですか？”と話す貴方のオーディオを送り、貴方が話したことの結果をブラウザで取り戻し、自分のブラウザに話す次の人に使用されるように音声サーバを割り当て解除することができることに基づいている。音声チャネルとＩＶＲポートとは初めに在来のオーディオ・プロトコルを用いることによりスイッチ及びＩＶＲによってセットアップされる。音声チャネルは初めはクライアントに接続されない。スイッチは、ＩＶＲとの連続的通信無しでＩＶＲ音声チャネルの割り当て及び割り当て解除を処理する。ユーザは、自分がＸ＋Ｖセッションの間に音声インタラクションを開始することを望んでいることを（普通はＰＴＴボタンを押すことによって）クライアント装置に示す。これはＸＨＴＭＬフォームとＶＸＭＬフォームとを同期化するＣＴＲＬチャネル上のリクエストと解釈され、これを本実施態様はＶＸＭＬブラウザが会話ターンを実行するためのトリガーとして使用する。マルチプレクサは、この制御コマンドを傍受し、装置と、開いているけれども接続されていない既存の音声ポートとの間に仮想音声回路を接続する。仮想回路は、ＲＴＰチャネルをセットアップすることを必要とすること無く接続される。ＣＴＲＬ信号は、その後、会話が行われ得るようにインタラクション・マネージャに転送される。会話の終わりに仮想回路は切断される。
【図面の簡単な説明】
【００４２】
【図１】ｎ個のクライアントと本発明の多モード・サーバとの図である。
【図２】１つのクライアントと本発明の多モード・システムの多モード・サーバとの間のシーケンス図である。

【図１】


X+V文書 20
X+Vフィルタ 22
VXML ブラウザ 24
IVR 26
音声エンジン 28
12 多モード・サーバ
XHTML 32
CTRL 34
SIP 36
RTP 38
スイッチ 30
マルチプレクサ 40
バッファ 42
CTRL SIP RTP
11 ネットワーク
クライアント10.n
n個のクライアント
クライアント10.1
XHTML ブラウザ 14
インタラクション・マネージャ 16
オーディオ・ハードウェア 18

【図２】


X+Vサーバ 12
IVRリンク・セットアップ 54
インタラクションを実行する 62
音声を送る 64
同期化 72
インタラクションを終える 74
対話を実行する 82
音声を送る 84
音声認識 90
同期データ 92
IVRリンク取り壊し 99
スイッチ 30
クライアント・リンク・セットアップ 52
音声回路接続 60
同期データを転送し、音声回路を切断 70
音声回路接続 80
同期を転送し回路を切断する 94
クライアント・リンク取り壊し 98
クライアント 10
音声セットアップ 50
同期イベント 58
音声を再生する 66
同期データ 68
同期イベント 78
音声を再生する 86
音声入力 88
インタラクションを終了する 96
音声取り壊し 97
ウェブ同期 56
音声同期 76

フロントページの続き

(74)代理人 100086243
弁理士 坂口 博
(72)発明者 ボウォーター、ロナルド、ジョン
イギリス国 エスオー５１ ８エスティ ラムジー・ハンプシャー ホワイトナップ アペルドーン
(72)発明者 クロス ジュニア、チャールズ、ウィラード
アメリカ合衆国３３４６７ フロリダ州 ウェリントン メインセイル・コート １１３００
(72)発明者 ウィルソン、レスリー、ロバート
アメリカ合衆国３３４８６ フロリダ州 ボカ・レイトン エス・ダブリュー １４番ストリート ６４９
(72)発明者 ウォン、ユックルン
イギリス国 エスオー５１ ７エイチユー ラムジー・ハンプシャー ステイプルフォード・クローズ ５７

審査官 仲間 晃

(56)参考文献 米国特許出願公開第２００３／００８８４２１（ＵＳ，Ａ１）
特開平１１－１０３３４８（ＪＰ，Ａ）
特開２００３－２８８３５６（ＪＰ，Ａ）

(58)調査した分野(Int.Cl., ＤＢ名)
H04M 3/493
G06F 13/00
H04M 11/06