Phonologies Media Services Framework

Copyright © 2004 Phonologies (India) Private Limited

1. Overview

Phonologies Media Services Framework provides standards based interface for the development of media rich and feature rich, interactive applications and services. It is a collection of software components implementing the industry standards in the Voice and Telephony Space.

Important components of the framework:
1. Oktopous Media Server – CCXML 1.0 browser
2. InterpreXer – VoiceXML 2.0 browser
3. Speech Recognition and Text to Speech
4. Phonologies Conferencing Server
5. VoIP Gateways

1.1 Oktopous Media Server

Oktopous™ Media Server is a "multi-service" Voice Platform built around some of the existing and emerging standards and technologies. It offers a full range of VoIP carrier requirements while providing a unique, open architecture through which new services can be quickly created and easily offered by Call Centers, Information Centers, Telcos, ISPs and Enterprises.

Support for multiple standards (VoiceXML, CallControlXML, SIP) ensures that it can work in legacy as well as next generation networks. For instance, it uses Session Initiation Protocol (SIP) for communications and a VoiceXML browser for information retrieval to achieve a seamless convergence of communication and information. Single platform solution provisions for multitude of services like IP Telephony, In bound / Out bound calling, Call distribution, IVR, IP based Fax and Audio conferencing, Call recording and Web integration.

Oktopous Media Server consists of Phonologies Open-source CCXML 1.1 Interpreter running on top of the SIP based Telephony platform. The SIP platform can seamlessly communicate with Phonologies VoiceXML Interpreter, Conferencing server and third-party VoIP gateways to provide rich interactive experience for the caller.

1.2 InterpreXer™

InterpreXer is a VoiceXML 2.0 compliant browser. InterpreXer™ provides convenient and easy access to voice enabled web services. InterpreXer™ is an abstract implementation of the VoiceXML specification and is not tied to any particular platform. It can be integrated with any ASR / TTS / Telephony platform of choice. InterpreXer has been fine tuned to work well with the Oktopous Media Server to provide interactive media services. (To find out more information on Phonologies InterpreXer Architecture please visit http://www.phonologies.com/platform.html)

1.3 Speech Recognition and Text to Speech

Phonologies InterpreXer used Speech Recognition Engines for understanding user utterances and, Text-to-Speech to convert the text-data to Speech on the fly. InterpreXer is independent of ASR and TTS vendors, and provides a uniform interface to integrate any ASR and TTS of choice. Phonologies provides a proprietary API for integrating any ASR and TTS engines. Work is underway to provide MRCP based interface, which is a open standard for Media resource control. Phonologies recommends Loquendo ASR and TTS for use with InterpreXer and Oktopous Media Server.

1.4 Phonologies Conferencing Server

Phonologies Conferencing Server is a software based conferencing mixer that is capable of handling hundreds of full/half duplex audio streams. It offers considerable cost and scalability advantages over hardware based conferencing boards.

1.5 VoIP Gateways

Media Server is a software based solution based on the SIP protocol. You need VoIP Gateways to connect to the PSTN / GSM networks. This enables users to call from any device and interact with the media applications or participate in conference discussions.

2. Architecture

All the important components of the framework are designed as servers to provide reliability and redundancy. Oktopous Media Server is the main entry point into the framework. Media Server can receive calls from any SIP enabled device.

Media Server, after receiving the call, can transfer the call to various destinations depending on the business logic. You can connect the user to an interactive speech application, transfer him into the call center queue, add him to a conference etc.

Media Server should be configured to work with one or more InterpreXer servers and Conference servers. When a Speech Dialog is required, Media Server connects the user to one of the available InterpreXer channels. In the same way, Conference Server will be required to create and manage conferences.

InterpreXer communicates with ASR and TTS engines through an abstract API. The ASR interface is based on a client / server model. For each ASR engine supported, we need to make a server component (RecServer.nuance for example) using the API provided by Phonologies as well as the ASR vendor. This Server will act as a mediator between the InterpreXer and ASR engine.

TTS interface is implemented as a shared object (plug-in). For each TTS engine supported we need to build a plug-in using the Phonologies API.

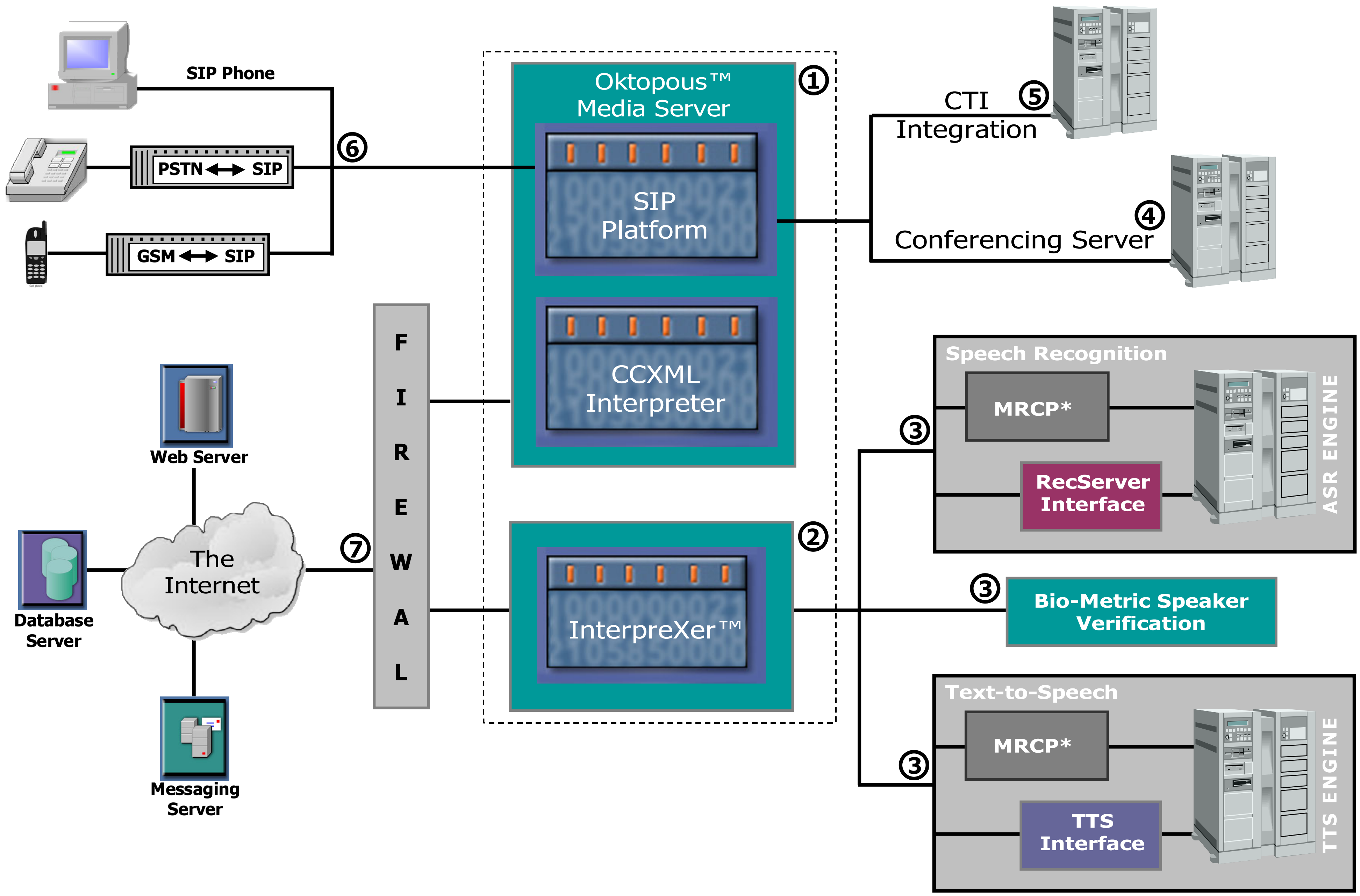
2.1 Setup

InterpreXer default configuration supports only DTMF based input and Pre-recorded audio playback. As explained above, Speech Recognition and Text-to-speech are implemented as optional components that can be plugged in. The Media Server base configuration does not include a Conference Server. Conference Server has to be added as a separate component.

Following are the steps requited to setup the system:

1. Start the Phonologies RecServer designed for your ASR vendor (RecServer.nuance for example) (optional)
2. Configure the InterpreXer to use this RecServer for recognition. You can use multiple RecServers in the configuration for fail-over recovery etc. (optional)
3. Configure the InterpreXer TTS plug-in to use the TTS of your choice. (optional)
4. Start the Licensing Server.
5. Start the InterpreXer Server with the desired number of channels
6. Start the Conference Server with the desired number of channels (optional)
7. Configure the Media Server to use the InterpreXer and Conference Server configured above. You may specify multiple options for redundancy.
8. Specify the default CCXML application. Also configure the DNIS based routing table in order to execute different applications based on the DNIS of the incoming call.
9. Now, Start the Oktopous Media Server
10. Media Server is ready to accept calls and serve your callers.

Phonologies Media Services Architecture



3. Usage scenarios

Phonologies Media Services Framework can be used by Call Centers, Unified Messaging Service Providers, Conferencing Service Providers, Telecom and Internet Service Providers. Build and deploy innovative services and solutions to the end user, for example, conferencing, unified messaging, unified communication, announcement servers, virtual PBX or virtual assistant, that can be hosted or sold bundled or unbundled to a customer or as their business requirements scale.

3.1 Outbound Call Campaign

Outbound calling is an important feature for service providers who provide reminder services, tele-marketing services, conditional alerts & notification, etc. Phonologies CCXML Interpreter gives total control over the call from the beginning to the end to the service provider, when making an outbound call.

Phonologies provides a simple API to the user to communicate with Media Server and initiate the outbound call. Setup for outbound dialing:

1. User program implementing the outbound calling API.
2. Media Server
3. SIP-PSTN gateway

Generally, the Media Server executes a CCXML application after receiving a call, but in case of outbound dialing, the Media Server starts the outbound dialer CCXML application without any incoming call. In this case there has to be some method of notification to the Media Server to execute the application. The notification is done by sending a simple text message to the Media Server with the details of the outbound call.

This message is of the form

`Createcall|`*`ccxmlurl`*

The first part is the command "`Createcall`", instructs the Media Server that this is an outbound call request. The second part, "*`ccxmlurl`*", is the URL of the Outbound dialer CCXML application, that contains the code to create and manage the outbound call.

How to make an outbound call:
Assume that you have a database of user's phone numbers and the time at which you have to call them. Also assume that you have a program that constantly monitors this database and triggers an event when a reminder is to be sent. When this event is triggered, your program should create a CCXML application to make the outbound call to that user. This CCXML application should have the logic to connect to a VoiceXML Interpreter after the call is answered, and also implement properties like "`connecttimeout`", "`maxtime`", "`retries`", "`maxretries`" etc. Once the CCXML application is ready, you can send a notification to the Media Server, in the above mentioned format.

We have provided a sample application with the Media Server distribution. This application uses a template CCXML document to create an outbound CCXML application for each user and then send an outbound request to the Media Server.

Note: Please have a look at section "3.7-Interactive Voice Response Systems" to see some more examples and usage scenarios of "Outbound Call Campaigns"

3.2 Advanced Conferencing

Media Server provides advanced conferencing facilities like, fully automated conference setup, floor control, conference management etc. Users can setup various conference types like – reserved, ad-hoc, system initiated setup, meet-me etc. Administrator of the conference can control the conference simply by issuing spoken commands. Every user of the conference can be attached to a dialog session, to provide speech access to various commands.

Different conference types:

1. Reserved

   In a reserved conferencing system, users go to a subscriber portal hosted by the service provider, and create a conference session. They will specify the date and time at which the conference should be created and names / phone numbers of the participants. Service provider will have a program which will be constantly monitoring the conference database, and trigger an event when a conference has to be created. After receiving the event, this program will send a notification to the Media Server to create the conference, similar to the way outbound call is handled.

   a. System Initiated: In a reserved conference, the system can call out all the participants and join them to the conference after proper authentication.
   b. Meet Me: In a "meet-me" conference, all the participants call the system at a specified time and get into the conference after proper authentication.

2. Ad-hoc

   Using Ad-hoc conferencing feature, users can create and manage conferences without the need to reserve in advance. The administrator / creator of the conference is already accessing some service of the media server (email-by-phone, voicemail etc) and decides to have a conference with a few people. Then the user will press a designated key (DTMF) to bring-in the conferencing application. He can speak names or phone numbers of the persons he want to conference with, and the system calls those people and join them to the conference automatically.

   You can also have a mix of reserved / ad-hoc conferences. Once a reserved conference is setup, the administrator can call out extra participants and join them to the conference.

   You can have many more conference types like, Coach-Pupil, where Administrator can communicate in full-duplex and all the participants will be only able to listen to him (half-duplex). It is also possible to join two existing/running conferences.

<u>Management and Control:</u>
The Administrator of the conference (optionally, other participants too) will have the ability to control and manage the conference. The Administrator can press a pre-configured DTMF key to "bring-up" the control-panel dialog. Using the control-panel, the Administrator can add / remove / mute / un-mute any participant using speech commands. The entire conference conversation can be recorded optionally and may be played back at the end of the conference for review.

<u>Participants Calling from geographically separated places</u>:
In a conference, assume that we have 4 participants calling from city A, 3 participants from city B and 5 participants from city C. In an ordinary conference, all the individual participants will independently call the conference server hosted in another city, example city A. In this case, all the participants from city B and city C will have to pay long distance call charges individually for the duration of the call.

If we setup independent Media Servers in different cities, one each in city A, B and C. All the participants will call the media server located in their respective cities, participants from city B call Media Server B, and participants from city C call Media Server C and so on. In turn the Media Servers located in geographically different cities exchange / mix media streams using only one channel, leading to only one long-distance call irrespective of number of participants in the call. This saves a lot of long-distance minutes for the participants.

3.3 Automated Call Distribution (ACD)

Automated Call Distribution is a common, but important feature in Call Center and Enterprise Telephony systems. ACDs route incoming calls to the final destination depending on various parameters and business rules. CCXML offers great flexibility to fully control and manage both incoming and outbound calls effortlessly.

Few examples of ACD feature:

1. Customer calls a company with a sales query.
2. The ACD (Media Server) answers the call and starts an interactive dialog with the caller to find out which department the caller would like to speak with.
3. The caller selects "Sales" (either says "sales" or uses the "touch tone" option).
4. The Media Server sends a request to the business intelligence module, to determine the appropriate destination (sales agent or automated dialog) for this customer. The appropriate destination is selected based on ANI, geography, time of the day etc.
5. Media Server transfers the call to the appropriate destination (sales agent or automated dialog), or to a voicemail if none available.

1. Customer calls a Call center.
2. The Media Server answers the call and puts the customer "on-hold" in a call queue. (Optional "music-on-hold")
3. While the call is in the queue, the Media Server can connect the customer to an "interactive dialog", in an attempt to address the customer's problem, or, to collect information related to the customer's profile, nature of query etc.
4. If the customer is satisfied with the response from the "interactive dialog", he / she can simply hang-up. Upon hanging-up, the customer's call will be removed from the call queue.
5. Alternatively, the Media Server receives external events notifying availability of an agent. It then takes the first call from the queue the matches the capabilities of the available agent, and connects the call to the agent. The Media Server then sends a message to the CTI server to show a pop-up menu on the agent's computer screen based on the information captured while in the call was in queue.
6. Media Server can also bring-in the supervisor into a conference call to assist novice agents.

Note: Please have a look at section "3.7-Interactive Voice Response Systems" to see some more examples and usage scenarios of "Automated Call Distribution"

3.4 Calling Card Application

Calling Card services are widely deployed by Telecom and Internet Service Providers. Of the Calling Card services, the most frequently accessed services are those for Long Distance calling. CCXML offers great flexibility to set-up, manage, route calls and retrieve information effortlessly. VoiceXML Dialog Server eliminates the need for specialized equipment and additional telephony ports on the VoIP Bridge to support IVR capabilities for retrieving calling card user account information, PIN authentication, and other information.

Example of a typical call:

1. The Customer (pre-paid or post-paid) calls the Long Distance Service Number.
2. The Calling Card Application Server (Media Server) answers the call and starts an interactive dialog (VoiceXML session is invoked) with the caller to capture Account Number and PIN. The customer maybe interested in account balance, account refill, make a long distance call, etc.
3. The customer selects "Long Distance Call" (either says "call John Smith in the UK" or uses the "touch tone" option).
4. Calling Card Application Server (Media Server) determines the appropriate destination (using customers address book) for the call and establishes a VoIP call across the IP backbone directly from the customer to the called party.
5. The customer completes the long distance call and is interested in details of the call, account balance, etc. It is important for the caller not to be disconnected from the Media Server.
6. Once the customer obtains the information, he / she may hang-up or place another call.

Media server eliminates the need to transfer to an external VoIP Bridge for calling and redirection back to the IVR from the VoIP Server. This results in a single voice circuit (for the VoIP call and IVR session) being utilized, eliminating "call tromboning", which decreases transport costs while reducing the number of telephony processing ports.

Optional features of Calling Card Application Servers include reseller management, real-time call billing, online credit card recharge and customizable user interfaces.

3.5. Customer Call Back Application

Customer Call Back services are provided by a specialized group of service providers. Through these services, sellers, marketeers and e-sellers are connected (speak) with their online customers with just one click at the point of sale, i.e. their web-site, banner advertisement or email link. CCXML offers great flexibility to set-up, bridge and route calls, and, VoiceXML provides an easy IVR interface to announce customer information, authentication, and other information. The Customer Call Back application (The Media Server) offers full control to manage both completed and missed call information effortlessly.

Example of a typical call:

1. A visitor to your customers web-site wants to ask a question related to your customers products & services. The visitor clicks the "Call Back" link / button on the web-site, after which a customized form of your customers services appears on the screen. The visitor enter his / her phone number and other optional data and clicks on the "Initiate Call Back".
2. The Media Server initiates a call with your customer at the number they have registered with. If your customer answers the call, an interactive dialog (VoiceXML) is invoked announcing details of the visitor requesting information on the web-site (information entered in online form).
3. Your customer requests to be connected with the visitor. Media Server initiates a call to the visitor (appropriate destination) and bridges the call (your customer + visitor). Your customers number is not displayed to the visitor. Your customer now has a live voice connection at no cost to their customers (visitor).

Optional Features:

1. Incase your customer is not available, the Customer Call Back application (Media Server) initiates a call to the visitor and announces the "unavailable" status of your customer. The Customer Call Back application can additionally carry out an interactive dialog with the visitor and gather more information related to the query.

2. If your customer is not available and misses a call, the application can optionally notify via sms or email. The message, email and sms will contain the call data entered in the form on the website. Additionally, incase your customer is not available, or the visitor is not available the Customer Call Back application (Media Server) can leave a message in voicemail box.

3. The number at which your customer wishes to be reached at can be modified at any time (edit / modify can be done through the phone or web interface) or optionally the application can "hunt" for your customer (findme-followme on office, home, mobile phone).

Management

1. Completed and missed call data are logged and stored on a webserver. These records can be accessible through a web-interface control panel for reference.

2. "Call Back Link" can be inserted in email messages, websites, banner advertisements, search engine listings, search engine advertisements, auction site listings, newsletters or any other areas where HTML links can be inserted.

3. Optional features of Customer Call Back application include reseller management, real-time call billing, online credit card recharge and customizable user interfaces.

Note: Please have a look at section “3.7-Interactive Voice Response Systems” to see some more examples and usage scenarios of “Customer Services and Customer Notification Services”

3.6 IP-PBX, IP Centrex & Virtual Assistant

Media Server enables carriers, system integrators and enterprises to build and deploy their custom Unified Communications Systems. These include the ability to provide:

1. IP Centrex Services (PBX like functions to a group of users, IP or PSTN-based, without the need for a CPE PBX)
2. IP-PBX (CPE PBX functions to a group of users, IP or PSTN-based)
3. Virtual Assistant: (Offer sophisticated call and voice control of incoming and outgoing calls and messages, unified communications by using voice commands)

IP based Unified Communications offer complete interoperability between users with TDM access or IP access, allowing a seamless migration of existing TDM-based users & customers, irrespective of their location to feature rich VoIP solutions.

Example of a typical call:

1. A call arrives and a speech-enabled "personal assistant" answers. The caller is presented with the option of the assistant either finding the called party or taking a message for them.
2. The caller chooses to find and speak with the called party.
3. The "Virtual Assistant" screens the calls through caller ID, or by asking the caller to speak their name.
4. The "virtual assistant" attempts to locate called party. The called party can be anywhere and can be contacted depending upon their preferences or rules denoted by the called party.

   For example:
   (a) During business hours, all calls diverted to office, or,
   (b) After business hours all calls diverted to voicemail, or,
   (c) After business hours calls from a particular list of people diverted to cell phone.

5. If the "Called Party" is available, the caller-id information or recorded name is presented to the "Called Party". The "Called Party" may accept or reject the call.

The synergy of speech and communications technologies, call control and messaging, in these applications provide the ability to create some powerful features including:

- Find Me - Follow Me (Single Number Access): "Called Party" gives out one number for all communications needs. The caller needs to dial one number and not multiple numbers. No separate numbers for business, personal cell-phones, home, etc. The system keeps track of the "Called Party" from one destination to the next based on flexible rules setup by user.

- Call Screening: "Virtual Assistant" screening calls through caller ID, or by asking the caller to speak their name. Further the information or recorded name is presented or played back to the "Called Party". The "Called Party" may accept or reject the call.

- Voice Activated Personal Information and Contact Management: Voice Activated Outbound dialing & Personal Information Management, Address Book lookup from personal or corporate directories or contact databases, and, calendaring capabilities.

- Call Back: "Virtual Assistant" places return calls to the callers for the "Called Party".

- Conferencing: Setup conferences based on conferencing party dialing-in or dialing-out. Initiate "on-the-fly" conferences with people listed in Address Books, corporate directories or contact databases through voice commands.

- Voice Messaging / Unified Inbox – Single "virtual box" for of all types messages (including email with an embedded voicemail, or simple voicemail, for example) to be stored and accessed over the telephone.

Management & Control
Media Server enables the ability to add, move, change and administer service features "on-the-fly" through the phone or through an interactive web-based interface. The application is device & location independent. Common IP infrastructure is shared by groups of users to improve network efficiencies.

Note: Please have a look at section "3.7-Interactive Voice Response Systems" to see some more examples and usage scenarios of "Auto Attendant, Call Distribution and Notification Services"

3.7 Interactive Voice Response Systems

Auto-attendant: Voice enabled "auto-attendant" to answer main, branch or remote office numbers. Callers simply speak the name of the person or department they want to reach. The auto-attendant then connects the caller with the appropriate person, desk, branch, group, department or location.

Bill Reminders & Collection: Billing call centers place reminder calls to customers of post-paid products and services before and after their bills are due.

Banking & Financial Application: Notification solutions inform customers of insufficient funds, potential fraud, and completed wire transfers, stock sales, and stock purchases in the Banking and Financial Sectors.

Customer Notification & Information: Application for delivery of time-critical information to customers. Inform flight delays and schedule changes, sale announcements. Retail vendors who ship products by air, ground, or mail can interact with customers as orders arrive or customers can check on the status of their orders as they desire.

Customer Surveys: Banks & Financial Institutions, Call Centers, Enterprises, Retail Vendors, Utilities, automatically survey customers by phone from time-to-time, regarding purchases, deliveries, after orders, services offered, installations and on-site repair services.

Employee Notification: Instant delivery of information to employees, divisions, or teams via phone calls to any desk or wireless phone. Application during emergencies, service outages, last-minute product updates, shift schedule updates. Remote employees, service technicians receive automatic dispatch calls. Remote employees call into an automated system to report inventory and sales information or report service status and completion updates.

Risk Management and Reduction: Call logging and Call recording, log whether calls were answered or not, record call audio for dispute resolution and minimize risk or for transcription or Call center agent training.

Store Locator: Provide prospective customer precise store location and directions, allowing retail vendors to let employees focus on in-store customers.