CollabGen®
Research · Analysis · Consulting

877.7Collab
877.726.5522
www.collabgen.com

Publication UTM-1005-3

Unified Telecommunication Model © White Paper

Third in a Series on
The Major Components of Voice and Data Convergence in the Enterprise

Self-service & Automated Response Systems – Legacy vs IP-based

Interactive Voice and Web Response (IVR / IWR) and Automated Speech Recognition (ASR)

Ed LaBanca, President & Principal Analyst
CollabGen Inc.
877.7Collab (877.726.5522)
elabanca@collabgen.com

Contributor:
Mark Levinson, President
VoxMedia Consulting

Abstract:

This generic paper is the third in a series on migrating toward a ***Unified Telecommunication Model (UTM©)*** in the enterprise. The first paper was on *UTM©* for Communication Servers & Contact Distribution Systems, and the second paper was on *UTM©* for VoIP Network Design, including WiFi & Mobility. Earlier papers serve as prerequisites to the topic at hand. Please contact CollabGen for copies of these or other white papers in the series.

Each part in the series will categorize a major component or set of components as an integral part of an overall technology architecture (infrastructure, systems and applications) and will cover the ***major trends in telecommunications including networks convergence, TDM / IP Telephony, portals, presence / proximity, multi-modal and collaborative communications.*** The central thesis is to look at traditional areas such as Telephony and Media Processing Systems (voice, fax, email, text, video etc.) and provide a holistic view into the current state-of-the-art for integrated / unified telecommunications. On the practical side, for most people it all comes down to the user interfaces and how well these systems work for them. For enterprise executives and staff, the goal of this series is to provide a better understanding of which components may need to be addressed and in what order.

This series will also cover issues associated with this new model including the efficacy of implementing various components in-house or via a service provider, and mitigating concerns including network and systems security, quality and reliability.

Each White Paper is generally structured as follows:

I. Executive Overview

- Scope
- Benefits
- Processes
- Technology Issues

II. Analysis

- Encapsulates the component's technological domain
    - Primary enabling technologies
    - Standards
    - Applications
    - Configurations
- Next Steps

III. Optional section

- Specific vendor(s) solution set

The above sections contain perspectives for **enterprise, contact centers, and remote / mobile applications.**

Note to Readers

- **This symbol** (the CollabGen logo) designates **analysis checkpoints**^SM that may be particularly significant to your organization's effectiveness and may warrant further professional services discussion.

Terms and acronyms identified in *italics* can be found in the attached Glossary.

Customized versions of this White Paper or Series are available upon request based on your specific environment and industry. Please contact CollabGen for other White Papers in this series.

Other reports include detailed side-by-side comparisons of telecom and contact center vendor solutions based on CollabGen's proprietary analysis criteria.

About CollabGen

CollabGen Inc. (www.collabgen.com) performs research and analysis of telecommunication technology, systems and applications. We provide management consulting services for clients to improve communications and customer service in contact centers and across the enterprise regardless of size or complexity. In addition to knowledge transfer and independent consulting, we are certified to provide Voice Response and Web Usability Audits and re-design services based on video-based testing to establish Best Practices. We are also certified in Benchmarking for Contact Centers to determine efficiency and effectiveness in comparison with your peer organizations and provide recommendations for improvement. CollabGen is an alliance-based firm comprised of independent analysts and consultants. Independent telecom planning, procurement assistance and project management is provided in conjunction with its subsidiary, eTelecom Consultants (www.4etel.com).

About the Authors

Ed LaBanca – CollabGen

Ed LaBanca is President and Principal Analyst / Consultant for CollabGen. He has over 30 years experience working for both large and small equipment manufacturers and software development organizations in the computing and telecommunications industries including 6 years as an independent analyst / consultant. Experience relevant to this paper includes positions as Product Manager of Voice Processing / IVR for Toshiba Telecommunication Systems Division and Senior Product Manager of Voice Processing / Messaging for Fujitsu Business Communication Systems. At Toshiba, his responsibilities included product, marketing and business planning for the company's internally developed Stratagy ™ Voice Processing System. Specific individual contributions included drafting the initial product specifications for the company's IVR and voice processing enterprise server

Mark Levinson – VoxMedia Consulting

Mark Levinson is President of VoxMedia Consulting. He has over fifteen years software and telecommunications industry experience, including stints at Bell Laboratories and GTE, where he did research, software development, system design, and project management. During five years as Director of Systems Development and Director of Partner Support at SpeechWorks International (since acquired by ScanSoft), he managed and contributed to many ground-breaking projects covering almost every aspect of voice technology. He holds S.B. and Sc.D. degrees from MIT, and an MBA from Boston University.

October 2005

I. Executive Overview

Scope

With the emergence of new media processing standards in conjunction with *Internet Protocol (IP)* for voice, data and video – applications can now be combined in new and more cost-effective ways in the context of a *Unified Telecommunication Model (UTM©)*. This paper addresses ***UTM© for Self-service & Automated Response Systems*** which until recently have been implemented extensively on first generation stand-alone systems in conjunction with *PBX's* within an enterprise.

This paper also provides an introduction to current *Interactive Web Response (IWR)* and speech-enabled *Interactive Voice Response (IVR)* self-service systems.

Interactive Web Response (IWR) Systems have been deployed using *Hyper-text Markup Language (HTML)* and *Extensible Markup Language (XML)* for customer specific data interaction via a web browser. Initially systems used the back-end host database integration of *IVR* systems to pass customer-specific data to the *IWR* system via a *common gateway interface (CGI)* script. This replicated the same functions as the *IVR* system. However, with the emergence of *Voice XML* this function reversed so the *IWR* system feeds back to the *IVR* system, since the *IWR* application can be more elaborate via the *graphical user interface (GUI)* and the *IVR* system functions as a subset.

Interactive Voice Response (IVR) Systems have been deployed for quite some time for customer specific data interaction via a phone with optional *Text-to-Speech (TTS)* and *Automated Speech Recognition (ASR)* more commonly referred to as speech recognition. An enormous number of different speech self-service applications have been successfully deployed. They include customer service applications in financial services, travel, catalog, and telecom. Many internal enterprise applications are also common, such as password reset and human resource benefits enrollment.

The concentration for speech recognition is on call / contact center applications, but the material applies equally well to any function that can be processed by telephone. Such applications are dependent on specific organizational requirements but warrant research relative to the capabilities that may already be in place within the contact center, and to extend these capabilities throughout the enterprise including mobile workers. New telecommunication standards such as *Session Initiated Protocol (SIP),* which incorporates new communication features such as *presence,* enables real-time multi-media communications over *IP*, including toll quality voice, in the same way that *Hyper-text Transfer Protocol (HTTP)* provides the exchange of files over *IP* either within the enterprise or via the Internet. The fact that these two capabilities can now be combined over the same network for voice and data provides opportunities for new self-service applications.

> "*Session Initiated Protocol (SIP),* ... enables communications over *IP, including toll quality voice,* in the same way that *Hyper-text Transfer Protocol (HTTP)* provides the exchange of files over *IP* either within the enterprise or via the Internet."

Benefits

In addition to the technology aspects which present significant cost-saving opportunities, we will also address some of the major principles and methods for achieving best practices for *IVR* and

IWR at the application level to satisfy and retain customers as well as improving intra / inter-enterprise communications and collaboration.

Current generation ***Media and Application Processors / Servers*** provide *IP* streaming, forwarding and integration of voice, data and video in various combinations in conjunction with *web services*. Therefore traditional voice processing applications such as Auto-attendant / Voice Mail; Fax; *IVR*; etc. can now be implemented and managed using browsers and thin-clients under a single (unified) environment for all media.

Note: Although this paper references *IP* Communication Platforms in conjunction with distinct Media and Application Servers, these logical functions can be physically combined in a single chassis or distributed in separate chassis depending on the specific design and business requirements.

Media Server

Media Servers provide *IP* real-time, streaming and store-and-forward capabilities for data, voice and video in applications such as those provided with *Communication Servers*, *IVR* and *Unified Messaging*. Current systems are employing *host media processing (HMP)* software in lieu of specialized cards using *digital signal processors (DSP's)*.

Call and Application Server

Application Servers provide control capabilities to answer and direct calls and handle signaling for *IP* media using industry standards including *SIP* and *H.323* (as discussed in the first and second papers in the series). They also enable access to *web services* such as *HTTP*, *XML*, *VoiceXML*, *WML, etc.,* in addition to the required resources of the *Media Server(s)*. A considerable new converged capability is the recent specification for *Computer Supported Telecommunication Applications (CSTA) – a standard which combines computing and switching applications* – in conjunction with *SIP*. Essentially, this enables built-in *Computer-telephony Integration (CTI)* via *SIP* peer-to-peer voice and multi-media communications with *web services* (to be addressed in a subsequent paper in this series for *CTI*).

> "A considerable new converged capability is the recent specification for *Computer Supported Telecommunication Applications (CSTA) – a standard which combines computing and switching applications* – in conjunction with *SIP."*

There are benefits for consolidating and combining Self-service and Automated Response Systems within *Communication Servers* and migrating from stand-alone application servers to *Media and Application Servers* which provides significant cost and overhead savings:

- Infrastructure is simplified and technology can be centralized which reduces or eliminates the need for maintaining disparate systems at different locations, but includes the option for mirrored hot-back-up distributed systems for multi-sites

- Specialized voice cards and dedicated circuits are no longer needed for voice mail, *unified messaging* and *IVR* systems

- *Automated Speech Recognition (ASR)* has matured, and includes *Natural Language Processing (NLP)*. When implemented using best methodologies *ASR* does much more than replace touch-tone entry, which will be addressed in more detail in subsequent sections

> "When implemented using best methodologies *ASR* does much more than replace touch-tone entry, which will be addressed in more detail in subsequent sections."

- The primary benefits of self-service and automated response are cost reduction through automation of routine telephone interactions, thereby saving labor costs and associated overhead

- Secondary benefits are 24x7 availability, and, in many cases, the opportunity to offer customers easy-to-use self-service options

- Further benefits may derive from specific applications, like call-direction. Call directors allow caller to access the desired party by speaking a name, rather than having to remember individual telephone numbers or extensions

- Applications are implemented or easily changed and extended using browser technology including drop-down menus, radio buttons, check boxes and wizards

- *Computer-telephony Integration (CTI)* projects and programming requirements are reduced and implemented more quickly so technology efforts can be directed to improve effectiveness, and customer retention / expansion

- Contact Centers are able to overcome obstacles to become Profit Centers

- Systems can be scaled quickly to meet the needs of ad hoc campaigns and surges in activity

- Increased built-in flexibility for continuity of business / disaster recovery, including outsourcing options

- Comprehensive end-to-end real-time and historical reporting and analytics

- Administrative and supervisory functions via web browsers

- Web multi-media services and standards, including those which enable toll-quality voice for the enterprise, provide richer and broader set of added capabilities for new and enhanced converged applications.

Processes

Although this paper focuses on Self-service and Automated Response Systems, including *Interactive Web Response (IWR), Interactive Voice Response (IVR)* and eMail, the following is an overview of interrelated systems and applications, including *Computer-telephony Integration (CTI), Enterprise Application Integration (EAI), Customer Relationship Management (CRM) and Unified Contact Distribution (UCD©) for* contact centers and across the enterprise, including mobile workers:

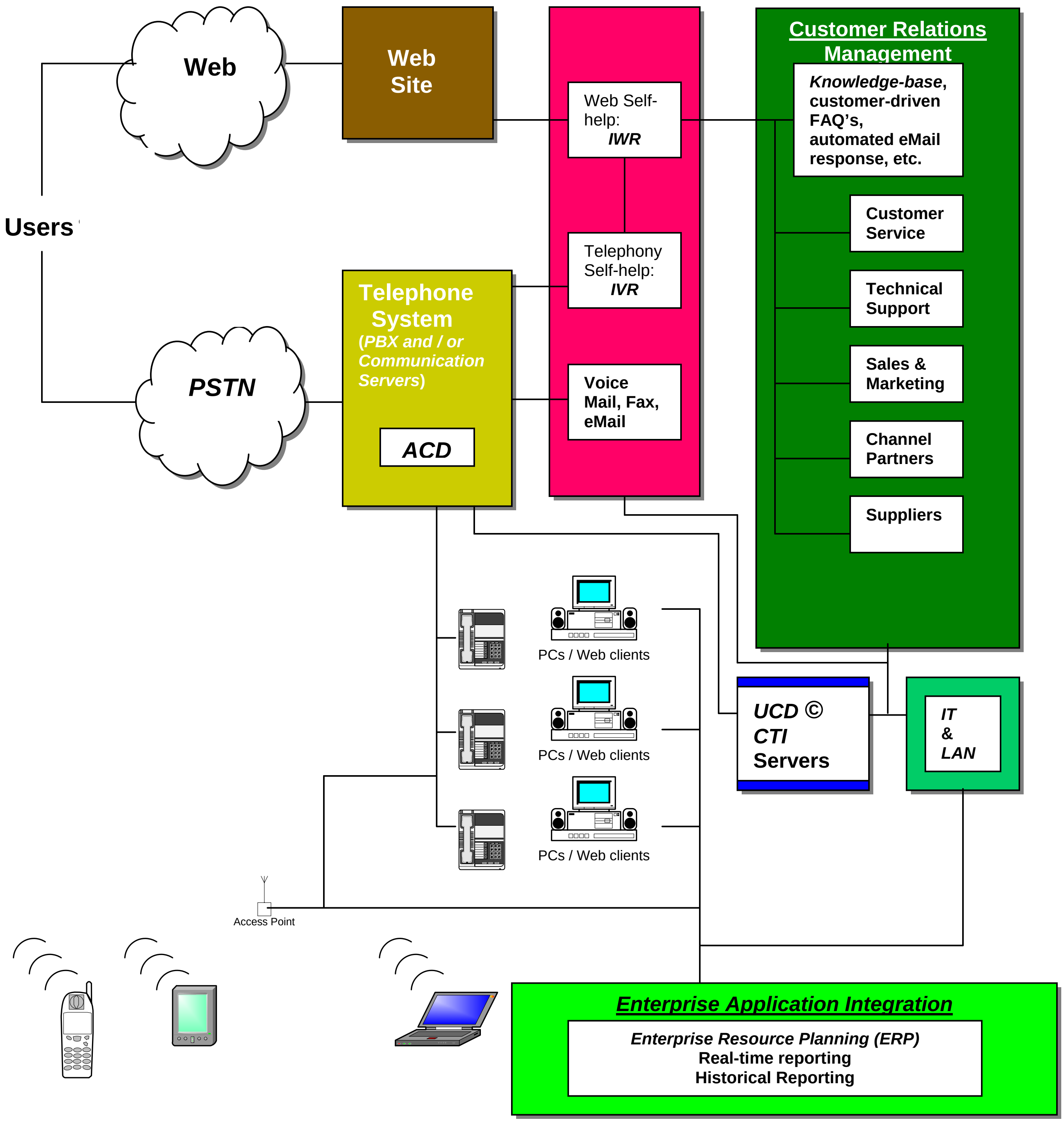


Figure 1

In the first paper in the series, we covered Communication Servers and Contact Distribution Systems as part of a *Unified Telecommunications Model (UTM©)*, in which a major benefit of converged networks is that the above systems and applications can be incorporated within *Communication Servers* and no longer need to reside at individual locations. Instead they can be centralized and distributed based on business requirements and the new opportunities being presented with networks convergence.

In the context of self-service and automated response systems, current generation systems consolidate individual servers that needed specialized voice cards into *Media and Application processors* (that can be part of *Communication Servers*). Together these servers can more efficiently accommodate all forms of multi-channel, multi media communications (refer to the following Analysis / Configuration section for more detail).

IVR or voice-response systems are normally used to process incoming telephone calls for customer service or internal enterprise applications. A voice-response application answers calls and attempts to process through completely automated self-service mechanisms if possible. In a typical call center environment, some calls are not automated (either because the application doesn't have the ability to handle the caller's request, or because callers opt-out of the automated system). Those calls are then routed to agents. In the ideal case, data collected by the automated system is simultaneously passed to the agents via *Computer-telephony Integration (CTI)*, so the agent has the caller's information on their screen before answering the call.

Another notable aspect of network convergence for *IVR* has been the rapid adoption in recent years of web-based standards, primarily *Voice XML*. Voice applications no longer require proprietary hardware and software. They now just reside on web servers like any other web application and can leverage the same business logic, back-end interfaces, and development tools.

In this standards-based environment, *CTI* is also becoming much simpler. Like *IVR*, it can just be implemented as a web application. The convergence process will be complete when *Voice-over IP (VoIP)* is deployed and there's no longer any need to interface with proprietary telephone switches and *PBX's*.

Customer Relationship Management

The diagram above (Figure 1) also shows the correlation of self-service and automated response to *IWR*, *IVR*, and *CTI* functions with *Customer Relationship Management (CRM)*. There are many aspects of *CRM* which will be addressed in future papers. The point here is that *knowledge-base (KB) systems* can provide users with additional options to further facilitate self-service containment. Modern *knowledge-base systems* are self-building by organizing inquiries or questions from customers and publishing company authorized answers in a *knowledge base*. These add to self-service containment by enabling users to search for answers via their browser or through *automated speech recognition (ASR)* and to automatically reply with possible answers via web / email or voice. The *knowledge base* can also be used by partners, suppliers, internal staff or agents.

Knowledge-base queries, which automate more complex technical service calls with speech, inherently involve more free-form language than more structured interactions like account status or travel reservations. For voice applications, these interactions generally require that the system recognize and extract meaning from natural language phrases

"*Knowledge-base* queries, which automate more complex technical service calls with speech, inherently involve more free-form language than more structured interactions like account status or travel reservations."

and sentences. This can be accomplished with *Statistical Language Models* and *Natural Language Processing*, as discussed below. Although these technologies are powerful, they are also subtle, and a good deal of skill and experience is necessary to deploy them successfully.

Technology and Application Issues

The principal technology used in current voice automation systems is *Automated Speech Recognition (ASR)*, or speech recognition.

Speech recognition technology has made enormous strides in the last ten years, and now offers robust capabilities. Highly accurate results can be achieved for large vocabularies (up to tens of thousands of words) and most speakers, despite person-to-person vocal variations and accents. However, to achieve these results, the recognition process must be constrained to limited domains at any one time.

For example, a travel application might first ask the caller for their destination and then the date of travel. In the first case, only city names would be recognized. They would be specified in a grammar built into the application. If the caller says something other than a city (or a city which isn't in the grammar) it won't work, and the application will re-prompt the caller to say an appropriate city. Similarly, for dates, the application will use a date grammar, and won't recognize phrases that are not dates within the expected range.

Therefore, speech recognition applications only work well when the dialog with the caller is limited to a sequence of specific prompts and responses where the speech recognition is tailored specifically to each step. Complex and useful applications can be constructed in this way, such as stock trading and catalog ordering.

The technology does have the ability to process more complex, free-form phrases through what is called *Statistical Language Models* and *Natural Language Processing (NLP)*. These systems are often used for call direction, and referred to as "How can I help you?" applications. They are used for categorizing incoming calls to determine which can be processed by the voice response system, and where to transfer the remainder. The grammars are developed by recording thousands of examples of real callers, and constructing grammars based on statistical predictions of what callers will say. Although the resulting recognition process won't work 100% of the time, it's not a problem for this type of application, since when callers are not understood, their calls are simply sent to the default agent group.

Problem: Widespread Deficiencies in Current Voice Response & Web Self-service Applications

Based on research of hundreds of live systems with access to a library of video-based testing to establish best practices, the main conclusion is that many leading organizations, including those with valuable brands were seen to have deficiencies in their self-service systems from the user or customer perspective.

> "Early deployment was technology driven…there was little or no science or methodology available for self-service best practices. Oftentimes, new systems would mimic the poor practices of other systems in operation."

It is important to note that *IVR* application development has gone through various stages since its general availability in the early 1980's. Early deployment was technology driven and relied primarily on the end-user organization to draft the requisite script and associated flow chart, and there was little or no science or methodology available for self-service best practices. Oftentimes, new systems would mimic the poor practices of other systems in operation. Unfortunately, all this lead to much frustration by users in adapting to these systems. Later, vendors understood the need to provide professional services support, especially with the convergence of web-based applications such as eMail. Applications improved, but still did not always conform to best practices. So now enterprises are accepting the need for professional third-party assistance, especially when deploying speech recognition.

Similar problems exist with web self-service applications, since many applications were developed without an understanding of human factors and associated methodologies.

Solution: Benchmarking Web and Voice User-Interface Design

The authors of this paper have worked with a leading research organization for performing auditing and benchmarking voice and web self-service applications in critical areas including:

- **Navigation**
- **Content**
- **Usability**
- **Interactivity**
- **Credibility**

Key benefits of professional auditing and analysis include:

- Establish standard metrics for usability and customer satisfaction
- Establish realistic goals by benchmarking performance against a nationwide database
- Identify specific areas for improvement based on observation of peer group performance
- Leverage best practices for short term improvements without requiring a capital investment in replacing existing platforms

Enterprises are continually balancing appropriate self-service and automated response systems with efficient and effective operations to achieve customer satisfaction, retention, and expansion. Benchmarking is also available including people and processes to compare your operations to other contact centers in your industry and identify quantifiable areas for improvement.

Voice User-Interface Design and Tuning

For voice self-service applications, the benchmarking process feeds into the design of the *Voice User-Interface (VUI)*, which is defined as the logical flow of the dialogs, together with the wording, tone, and “sound and feel” of the audio prompts played to callers, and the wording of expected responses. The understanding gained from benchmarking is combined with business requirements and caller demographics to create *VUI’s* that are easy for callers to understand and use, and give an experience that’s consistent with the organizations branding and marketing goals.

The *VUI* design process doesn’t end when applications are completed. Because the final application engages in dialogs with large numbers of callers, there will inevitably be areas where at least some people have difficulty. So a deliberate “tuning” process of monitoring, analysis, and improvement should be conducted over the first few months of production deployment to catch and resolve these issues.

II. Analysis

Technological Domain

Self-service & Automated Response Systems are categories within *media processing* systems and applications. These systems enable the use of ubiquitous tools including various forms of telephones, notification devices and web clients for real-time, streaming and store-and-forward communications and collaboration.

Primary Enabling Technologies

- Media (voice / sound, image / fax, and video) *digitization* and *compression*
- Touch-tone recognition and generation
- Voice and video streaming; and store-and-forward / retrieval
- *Automated Speech Recognition (ASR)* – Speech Recognition interprets callers' spoken phrases and matches them to sequences of words allowed by a grammar.
- *Speech Synthesis / Text-to-Speech (TTS)* – Speech synthesis (the reverse of speech recognition), is where computer-generated audio is created from text. The best current products sound fairly natural (compared with earlier, robotic-sounding voices), but still lack the intonation and subtlety of human speech. Text-to-speech is used primarily where it is impractical to pre-record the information, either because the number of possible variations is too large (like names and addresses) or changes too frequently (like weather reports). However, it is widely used in *Unified Messaging* for reading eMail via the phone.

Standards

There are various standards and *protocols* used for the *media* and *call processing* systems and technologies itemized above which need to be considered in detail when specifying associated systems and applications to ensure adaptability and inter-operability (reference the first paper on *UTM© for Communication Servers & Contact Distribution Systems*).

For *Automated Response* and *media processing* / *call processing servers* the following are the most relevant protocols and standards:

- ***Voice XML (VXML) –*** a revolution in voice response application development has occurred with the widespread adoption of *VXML*. It leverages web-derived standards, including *XML* and the three-tier web architecture of back-end, web / application server, and browser. It brings welcome standardization to voice response hardware and software, and enables application portability between different vendor's platforms. It also enables the use of widely available web development tools and methodologies. Microsoft has been promoting an alternate standard, Speech Application Language Tags (SALT), but it has not been widely adopted.

- ***Session Initiated Protocol (SIP)*** – the initial proposal developed for *VoIP* and multi-media communications to define the technical parameters and details required for peer-to-peer session management (more detail is provided in the previous white papers in the series).

- ***H.323*** – the precursor, and relatively mature multi-media *protocol* standard, to *SIP*. There are many functions not yet defined in *SIP*, such as negotiating capabilities and communications between devices based on compression, etc. The vast majority of *VoIP* networks interoperating with the *PSTN* today are using *H.323*.

- ***Host Media Processing (HMP)*** – performs media processing functions via the CPU without any special-purpose telephony hardware.

Applications

Automated Response and *Media Processing Systems* can be combined to include the following point solutions under a single server (including mirrored and *carrier-grade servers*) as opposed to having separate systems and applications in multiple locations:

- **Traditional Telephony and Web Communications Applications being combined:**

    - Telephone self-service / *Interactive Voice Response (IVR) and Interactive Fax Response (IFR)*
    - Web self-service / *Interactive Web Response (IWR)*
    - Auto-attendant (call director) / Voice Mail
    - *Audiotex*
    - Fax
    - E-mail
    - *Unified Messaging*
    - Web call-back, etc.

The following table provides a cross section of these applications relative to degrees of automation and timing for communication to take place between or among the various parties.

Voice Processing & Computer - Telephony / Web Integration		
Application	**Real-time communications**	**Store-and-forward / retrieval**
Fully Automated	Automated-attendant, *Interactive Voice Response (IVR) / Interactive Fax Response (IFR), Interactive Web Response (IWR), Knowledge Base* / key word search / *natural language processing*, text / email / fax / multi-media response	Voice / video mail, *Audiotex*, fax back, fax-on-demand, fax distribution (standard documents & forms)
Semi-automated	Screen / data pops to agent desktop (*CTI*), Web text chat, Web call-back, Web co-browsing, Web page push to browser	Voice mail reply, email reply

Figure 2

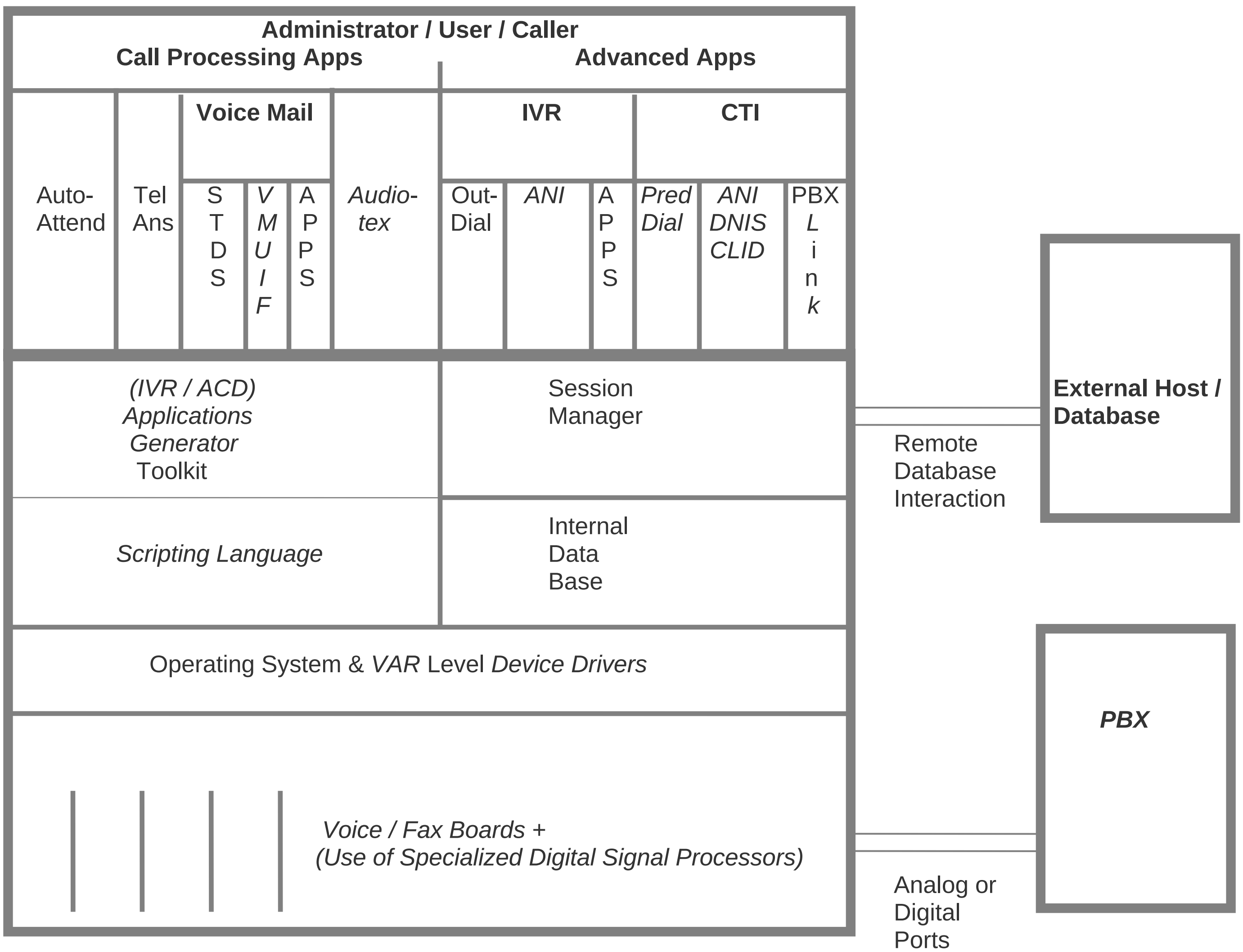
Store-and-forward applications are being mediated through the use of ***presence*** (a feature of *SIP*) and *multi-modal* communications.

It is important to re-iterate that many ‘front-door’ applications do not adhere to best practices. As mentioned previously in the Executive Overview, studies including the use of video clips of users indicate a clear and compelling need for improvement in most enterprise self-service applications. The following section addresses Self-service and Automated Response Systems, specifically *IVR* for telephone users and *IWR* for Internet users.

Configurations

Most organizations are using first generation systems with stand-alone servers connected to a *PBX* or *Communication Server* on one end, and an external host or database on the other, as illustrated in Figure 3 below:

<u>First Generation *Voice Processing System*</u>
(Includes legacy / non-IP design elements)



<u>Figure 3</u>

Today's systems utilize open standards providing *Modular Communications Platform (MCP)* architecture. *IVR* can now be included within current generation *IP Communication / Media and Application Server* configurations as shown below in Figure 4.

Next-Generation Call and Multi-media Processing System

Call Processing **and** ***IP Multi-media Services and Applications***
Standard ***API's and Web Services***
Application Generator **Toolkit and Database Interaction via IP** (including the new *CSTA* specification for *CTI* with *SIP*)
Host Media Processing (HMP)
Operating System *(including hardened or Carrier-Grade)*
PCI, *Compact PCI*, or *Advanced Telecom Computing Architecture (Advanced TCA)* ***Media and Application Processor Servers and Blades*** [No specialized *Digital Signal Processor (DSP)*-based cards required for *HMP*]

Figure 4

Working from the lower layers of the above configuration, the following provides further description:

PCI, Compact PCI, and Advanced Telecom Computing Architecture (Advanced TCA) – initiatives of PICMG (PCI Industrial Computer Manufacturers Group) that is a consortium of over 450 companies who collaboratively develop open specifications for high performance telecommunications and industrial computing applications and provide high-availability. *Advanced TCA* is an open industry specification designed the requirements for current and next-generation carrier-grade communications equipment primarily based on *IP*.

Operating Systems – include Linux and Windows. Carrier-grade versions are hardened to ensure against security attacks.

Host Media Processing (HMP) – With the increased performance of general-purpose processors, voice processing that once required dedicated digital signal processors (DSP's) can now be performed with software-only solutions on standard servers running Windows or Linux. *HMP* performs media processing functions without any special-purpose telephony hardware.

***Application Generator* Toolkit and Database Interaction via *IP* –** First generation systems relied on cryptic scripting languages. Some best-of-breed systems incorporated drag-and-drop application generators based on object-oriented programming. Few vendors with legacy systems were able to offer a common end-to-end development following the entire call process. With *IP*, all of that is changing since telecom vendors are now shifting to modular platforms using *IP*, and to embrace *web services* and common development environments to stay competitive. Moreover, with major standards converging around *IP* communications including *CSTA* and *SIP*, end-to-end application development has taken on a whole new dimension.

> "Few vendors with legacy systems were able to offer a common end-to-end development following the entire call process. With *IP* all of that is changing, since telecom vendors are now shifting to modular platforms using *IP* and embracing *web services* and common development environments to stay competitive."

Standard *Application Program Interfaces (API's) and Web Services* – *API's* have been used for quite some time, including with first generation client-server architectures to enable a programmer to make requests from the operating system or another program. *Web Services* are a third-generation services-oriented, component-based application architecture using standard Internet protocols (*XML*, *Voice XML*, etc.) to make requests of other web applications. The client does not need to be a browser, but can be any type of Internet device, including a PDA or cell phone using *Wireless Markup Language (WML)*. Machine-to-machine transactions are done using *Simple Object Access Protocol (SOAP)*.

The following diagram (Figure 5) depicts an *IWR* and an *IVR* system [based on *Host Media Processing (HMP)]* using *Voice XML* in a *TDM* (legacy systems below the dashed line) and *IP* environment (above the dashed line):

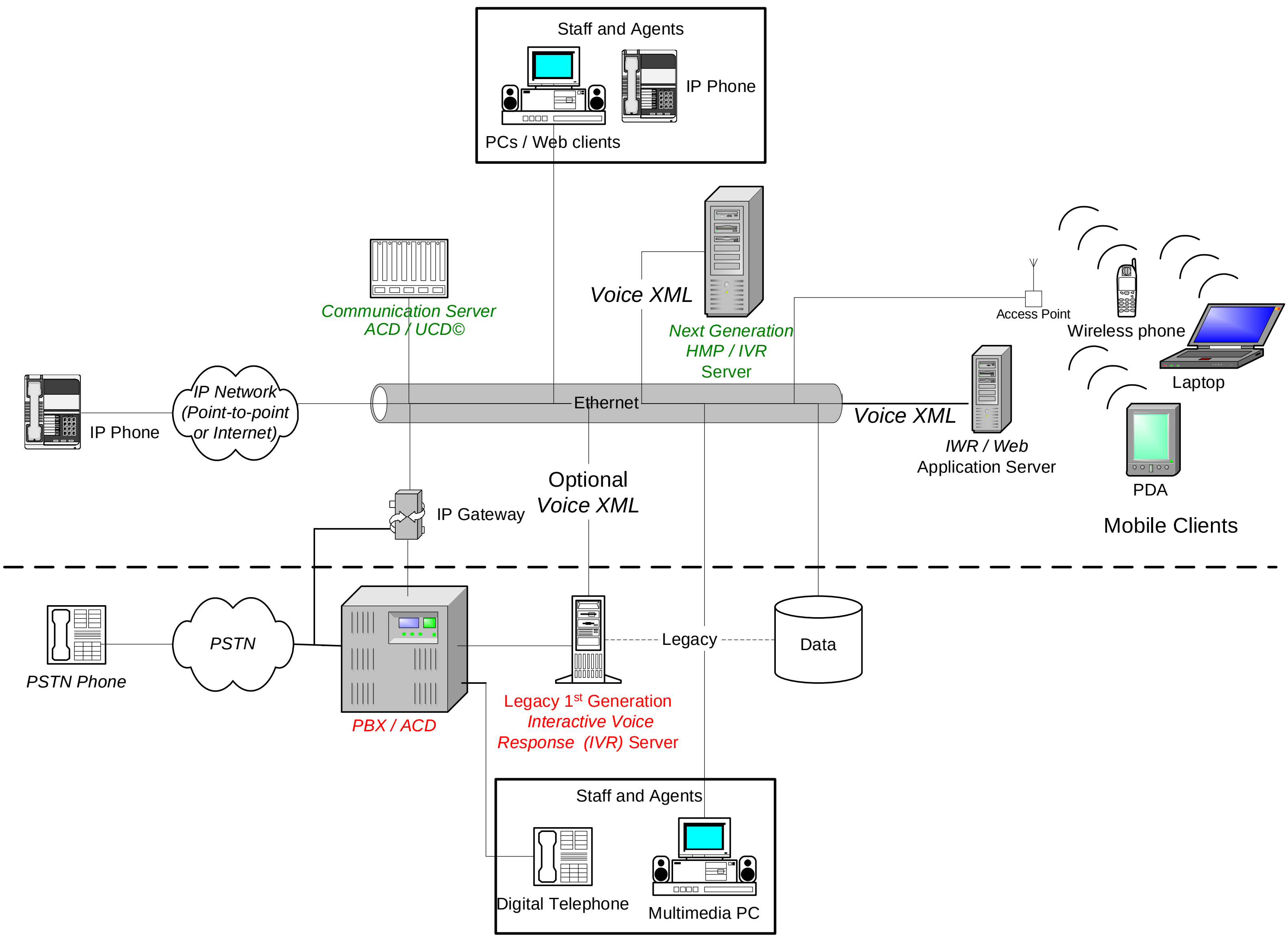


Figure 5

Normal configurations for *Voice XML*-based automated response systems conform to the web-based three-tier architecture; the application and associated business logic reside on a web server that serves *VXML* pages to “voice browsers”, also known as “voice gateways”. Unlike web-browsers, which are located with the user, voice browsers are normally housed in data centers, and users access them via telephone. The web servers connect to back-end data systems in exactly the same way as for web sites. In most cases, the same back-end data interfaces can be used for both web and voice response applications.

Next Steps

- Develop long term and short term strategy for going beyond traditional telephony and self-service and point solutions
- Re-evaluate your current telecom model vs. a unified model for your particular enterprise and environment
- Identify opportunities, impediments and solutions

Glossary

Advanced Telecommunication Architecture (ATCA) – is an industry initiative to provide standardized platform architecture for carrier-grade IP telecommunication applications, with support for carrier-grade features such as NEBS (Network Equipment-Building Systems) and 99.999% availability.

Access Point (AP) – a station that transmits and receives data and connects users to other users within the network and also can serve as the point of interconnection between the Wireless Local-Area Network (WLAN) and a fixed wire network.

Audiotex – *a* voice information bulletin board application that enables telephone users to hear pre-recorded information from a *voice processing system*.

Application Generator – a *graphical user interface (GUI)* using drag-and-drop objects to build an application and generate the required program code to run the application.

Application Processing System – provides operation and management / administrative features for various multi-media applications, which in the context of telecommunication architecture can reside on the same or a separate server from the media processing system that provides access to the required media resources.

Application Program Interface (API) – is a specific method prescribed by a computer operating system or by an application program by which a programmer writing an application program can make requests of the operating system or another application.

Automatic Call Distribution (ACD) – is an option that can integrate to most telephone systems which processes incoming calls to groups of agents. The ACD determines which group to transfer to, and if all agents are busy, plays a message and places the caller on hold and queues the call in sequence for the next available agent in that group. The system also provides management reports and information such as number of callers on hold, average hold time, etc.

Automatic Number Identification (ANI) – the telephone number passed thru the *PSTN* indicating the number of the phone calling into a *voice processing* application or directly to another phone.

Automated Speech Recognition (ASR) – is an option for a *voice processing* or *Interactive Voice Response (IVR)* system to enable input using voice. Types of ASR include speaker-dependent and independent; continuous [*Natural Language Processing (NLP) or Statistical Language Model (SLM)*] or discreet (limited to individual words or numbers).

Blade – A blade is a thin circuit board, containing one or more CPU's (making it a server) typically for a specific application within a modular rack which is energy and cost efficient.

Calling Line Identification (CLID) – similar to *ANI*, but typically is used in reference to the information provided by the Local Exchange Carrier (LEC) or *PBX*.

Carrier-grade Operating System – an operating system that contains specific attributes to maximize security and uptime.

Common Gateway Interface (CGI) – a standard method for passing specific user information between a web server application and a user.

Communication Servers – also known as soft-switches are platforms that use software on industry standard servers to perform voice and multi-media communications over *IP* and via the *PSTN* through *gateways or SIP* trunks*.*

Compact PCI – a standard to define CPU form factor and backplane connectors

Computer Supported Telecommunication Applications (CSTA) – a standard which combines computing and switching *(CTI)* applications which includes a set of profiles that are especially tailored for CSTA applications that control and monitor an endpoint device (a *SIP* device, for example).

Computer Computer-Telephony (CTI) – Controlling the telephone system either from the client desktop (first party call control) or via a server (third party call control). CTI enables *voice processing* applications to monitor the status of telephone station sets and their availability within the switch. Call control functions include: answer call, make call / dial a number, release, hold, park, pick-up, call transfer (blind or supervised), conference, etc. Additionally, caller information such as *calling line identification (CLID)* can be passed through the telephone switch to a specified *voice processing* application. Commonly, it involves coordinating the transfer of a call in conjunction with sending screen pops to a desktop PC or thin client.

Customer Relationship Management (CRM) – a set of off-the-shelf or custom applications used to maintain customer information across the enterprise instead of disparate silos of information within departments. It also relates to best practices for improving customer satisfaction, retention and expansion.

Device Driver – is a program that controls a particular type of device that is attached to a computer. There are device drivers for voice cards, printers, displays, disk drives, voice cards and so on.

Dialed Number Identification Service – is the use of a specific incoming number to initiate a voice or multi-media processing application.

Digital Signal Processor – a CPU that performs specialized tasks, including voice and multi-media processing at a very fast rate compared to general-purpose CPU's. Essentially these chips process analog signals and convert them to digital form to facilitate analysis and manipulation.

Enterprise Application Integration (EAI) – is the integration of disparate or "best-of-breed" systems within the enterprise. For example, combining data or transactions that occur within a *Customer Relationship Management* application such as order entry with back-end applications such as inventory management.

Enterprise Resource Planning (ERP) – is a system that provides back-office applications including finance, human resources, inventory management, etc.

Extensible Markup Language (XML) – specifies content variables in terms of what data is being described. For example, the word "accountbal" placed within markup tags is a way to indicate that the data that follows is an account balance.

Gateway – a network point that acts as an entrance to another network and provides *protocol* conversion.

Graphical User Interface (GUI) – navigation of a computer screen using graphics (as opposed to text) and a pointing / selection device such as a mouse or touch-screen.

H.323 – describes how multimedia communications occur between devices, network equipment and services.

Host Media Processing (HMP) –performs media processing functions of a communications application without any special-purpose telephony hardware. In the HMP model, toll-grade audio is transported as *VoIP* through the Ethernet network interface and all media processing occurs in the host CPU.

Hyper-text Markup Language (HTML) – is the standard language for publishing hypertext on the World Wide Web. It is a non-proprietary format that can be created and processed by a wide range of tools, from simple plain text editors - you type it in from scratch- to sophisticated authoring tools. HTML uses tags such as <h1> and </h1> to structure text into headings, paragraphs, lists, hypertext links etc. Its successor for current and future document types and modules is XHTML, which is a reformulation of HTML using *XML*.

Hyper-text Transfer Protocol (HTTP) – is the set of rules for transferring files (text, graphic images, sound, video, and other multimedia files) on the World Wide Web.

Information Technology (IT) –is a term for encompassing both telephony and computer technology.

Interactive Fax Response (IFR) – similar to *IVR* below, but incorporates user-specific data in fax form (as opposed to voice) as a response to an automated request.

Interactive Voice Response (IVR) – enables the use of telephones or fax machines to effectively be used as data terminals for access to user-specific database information via a tone dialing keypad for navigation through menu prompts and entering data. *Automated Speech Recognition* can optionally be used for input. Output is via voice response using concatenated pre-recorded voice files, computer-generated text-to-speech, or fax [*Interactive Fax Response (IFR)*].

Interactive Web Response (IWR) – extending the functionality of *IVR* to the Web by leveraging the database access communications that is already in place with *IVR*. Additional functionality can be added using multi-media tools and content. Conversely, *IWR* can provide database information to the *IVR* system via *Voice XML.*

Internet Protocol (IP) – is the method or by which data is sent from one computer to another on the Internet. Each computer has at least one unique IP address. IP is a connectionless *protocol*, which means that there is no continuing connection between the end points that are communicating. Packets arrive in the correct sequence because of *Transport Control Protocol (TCP)*, the connection-oriented *protocol* that keeps track of the packet sequence in a message. IP Version 4 is prevalent with IP Version 6 now being deployed.

Knowledge-base (KB) System – is an application for the dissemination of information, generally online, but can be extended to a *voice user interface (VUI)* using *Automated Speech Recognition (ASR)*. Information access can be automated using search engines, parsing emails or by systems using artificial intelligent processes and a variety of media by users and / or agents in a contact center.

Media Processing Systems – interactive information systems comprised of computer-based hardware and software that integrate to *communication* servers / business telephone systems, directly to the Public Switched Telephone Network (*PSTN*), and / or data networks including the Internet for multi-media applications.

Modular Communication Platform (MCP) – are industry standards-based communications infrastructure platforms and building blocks.

Multi-media Services – various media and combinations of media including voice [*Voice-over IP (VoIP)]*, audio, video, text, fax, eMail, etc. used within telecommunication systems and applications.

Multi-modal communications – combines speech with other channels of access such as voice, keyboards and touch-screens to address business needs including mobility and extending customer service. Multi-modal communication applications include real-time and streaming voice or video, store-and-forward applications such as eMail, and collaborative applications such as conferencing. In the context of computer-generated speech, multi-modal systems incorporate a means of adding non-verbal cues to speech (for example, animated gestures such as nodding or winking by an avatar) to make the communication more clearly understood.

Natural Language Processing (NLP) – the recognition and generation of natural human language by a computer.

Operating System (OS) – the fundamental program that enables devices to be connected and applications to run on a computer.

PCI – a standard to define CPU form factor and backplane connectors

Point-to-point IP Network – allows corporations to extend their own corporate network through private and secure tunnels over the public Internet vs. leased lines for wide-area communication. Also known as a *Virtual Private Network (VPN)*.

Presence / proximity awareness – is a technique within a type of application that makes it possible to locate and identify a computing or communications device wherever it might be located when it is connected to the network. The device's location could be determined via a Global Positioning System (GPS). A user's proximity can be automatically determined, such as through *Radio Frequency Identification (RFID)*, indicating a readiness or ability to communicate in real-time via the connected device.

Private Branch Exchange (PBX) – also Private Automatic Branch Exchange. A customer premise switch that provides control and connection between station (telephone) sets within an organization based on *TDM and circuit-switching,* and provides numerous features and functions such as hold, transfer, conference, etc.

Protocol – is a specific set of rules that end points in a telecommunication connection used to communicate.

Public Switched Telephone Network (PSTN) – connection-oriented world-wide telephone network for voice and data.

Radio-frequency ID (RFID) – is a technology that incorporates the radio frequency (RF) portion of the electromagnetic spectrum to uniquely identify an object or individual. The advantage of RFID is that it does not require direct contact or line-of-sight scanning. An RFID system consists of an antenna, transceiver, and a transponder.

Scripting Language – uses text with specialized grammar and structure for developing an application.

Session Initiated Protocol (SIP) – is an Internet Engineering Task Force (IETF) standard *protocol* for initiating an interactive user session that involves multimedia elements such as video, voice, text chat, etc.

Simple Object Access Protocol (SOAP) – is a way for a program running in one kind of operating system to communicate with another using *web services*.

Statistical Language Model (SLM) – is a type of speech recognition grammar (a set of rules specifying which words and phrases will be recognized) constructed using statistical analyses of thousands of examples of real user spoken input. Distinct from conventional prescriptive grammars created by application developers based on a priority of what their expectations are of what users will say.

Text-to-speech – converting text to computer-generated voice.

Telecommunication Device for the Deaf (TDD) – text based communications over the telephone network or other transmission facility using special modems based on the Baudot signaling method. They are usually acoustically coupled and typically operate at a relatively slow 300 baud or bps.

Time-division multiplex (TDM) – is a method of putting multiple data or voice streams in a single signal (circuit) based on timing.

Unified Contact Distributor (UCD©) – is an evolution from the traditional *Automatic Call Distributor (ACD)*, which distributes real-time voice calls to agent queues. In addition to distributing voice calls. UCD's also control and distribute multi-channel, multi-media telecommunications in the form of real-time and non real-time communications. These include applications such as Voice Messaging, Fax, *Telecommunication Device for the Deaf (TDD)*, eMail, Text Chat, Web call-back, *Voice-over-IP (VoIP)*, Conferencing / Collaboration, and elements of *Computer-telephony Integration (CTI)* and Customer Relationship Management (CRM), etc., which are distributed to users and agent groups defined by skill sets and other business rules. Users and agents also have control over various functions within these applications via a *Graphical User Interface (GUI)* on thick or thin (browser-based) workstations and mobile devices. It follows the Integrated vs. *Unified Messaging* model of integrating stand-alone servers or having a single server that unifies / synchronizes the communications in real-time. Registered © CollabGen Inc. / eTelecom Consultants 2004 TX 6-066-696

Unified Messaging – combining voice / fax and email under a single server control. Access to all message types is done via the phone or PC / workstation. eMail headers and text content can be relayed over the phone using *text-to-speech* technology.

Unified Telecommunications Model © – planned convergence of telephony and web communication that migrates from *circuit-switched* to modern *packet-switched Internet Protocol (IP)* telecommunications systems and applications in contact centers and across the enterprise. The objective is to provide end-to-end monitoring, management, and reporting regardless of the telecommunications channel, media, application or location. Registered © CollabGen Inc. / eTelecom Consultants 2004 TX 6-066-696

Value-added Reseller (VAR) – is a general term for any organization that builds on an enabling technology to provide products and services thru various marketing channels to end-users.

Virtual Private Network (VPN) – is a way for an enterprise to use a public telecommunication infrastructure, including leased lines and the Internet, to provide remote offices or individual users with secure access to its network.

Voice-over-IP (VoIP) – Voice over a *packet-switched* network using *Internet Protocol (IP).*

Voice Mail User Interface Forum (VMUIF) – Bellcore and the local service telephone companies began a Voice Messaging Group which produced a user interface specification in 1989. Members of this group joined with others in this industry to form larger, independent group (with wider representation) called the Voice Messaging User Interface Forum (VMUIF) which produced another voice mail standard in 1990. VMUIF brought this activity to ANSI X3V1 in 1990, which began work on a voice messaging standards in its Text and Office Systems group (TG9). (Thus, even wider representation in the voice mail and information technology industries was brought to bear on this standards effort.) The voice mail standard gives specific recommendations on aspects of the telephone user interface, such as minimum response time, figures for inter-key time-outs, error handling and response to repeated time-outs. It specifies dial-ahead and dial-through behavior. Also specified is the details of the behavior of the # and * keys of the touch-tone keypad. Standardized assignments of touch-tone keys to functions are specified for (a) the control menu, a series of functions which should be available from any system state (e.g. *0 for help), (b) call answering before and after the record tone, (c) the main menu of voice mail applications, (d) the menu for listening to and administering messages, and (e) sending a message, before and after the record tone. A series of key assignments and requirements are also specified for bulletin board systems.

Voice Portal – provides entry to a voice application processor using *Automated Speech Recognition (ASR)* or touch-tone, which can be used for *IVR* and / or messaging and extended to web pages, general computing and *media processing.*

Voice Processing System – interactive telephone information systems, comprised of computer-based hardware and software, that integrate to business telephone systems or directly to the *Public Switched Telephone Network (PSTN)* or *VoIP* Network. Callers (or call recipients) can automatically be connected to digitally encoded voice recordings (sound), other telephone devices, and computer databases - all of which can be combined to provide an unlimited variety of fully and semi-automated applications.

Voice Processing systems allow the use of touch-tone phones and fax machines to enable the following categories of applications:

1) **Call Processing**
    a. Automated Attendant [call routing and automatic call distribution (ACD)]
    b. Voice Mail
    c. Voice (Audiotex) / Fax Information Bulletin Boards
2) ***Interactive Voice Response (IVR)***
    a. Enter and retrieve any information stored in a computer database
    b. Consultation with an internal or external *Knowledge-base (KB)* System.

Voice XML (VXML) – enables interactive access to Web applications and data through the telephone using touch-tone or *Automated Speech Recognition (ASR)*.

Web Portal – provides entry to a web site typically through a web browser [but can also include a *Voice User Interface (VUI])* which can be used for general computing and *media processing.*

Web Services – third-generation services-oriented, component-based application architecture using standard Internet protocols (*XML*, *Voice XML*, etc.) used to make requests of other web applications. The client does not need to be a browser, but can be any type of Internet device, including a PDA or cell phone using *Wireless Markup Language (WML)*. Machine-to-machine transactions are done using *Simple Object Access Protocol (SOAP).*

Wireless Markup Language (WML) – is part of the Wireless Application Protocol (WAP) standard which is a language that allows the text portions of Web pages to be presented on cellular phones and personal digital assistants (PDAs) via wireless connection.