Customized IVR Implementation Using Voicexml on SIP (Voip) Communication Platform

Kranti Kumar Appari[1]
(ECE Department, JNT University, India)

ABSTRACT : *An innovative application for communication platform based on SIP (VoIP) protocol is presented in this paper.*

The Voice Server is an Open-standards-based voice services framework that interprets the VoiceXML dialog markup language. It is designed to serve as a VoiceXML interpreter implementation for VoIP platform. Although it is perfectly suitable for PC desktop applications, it can be integrated with any telephony platform, messaging suite or communications solution intended to implement the VoiceXML functionality to execute feature rich voice enabled applications like: auto-attendant, email-by-phone, voice dialing, message notification and reminders, contact book look-up, business transaction enablement, customer relationship management and utility applications (driving directions, flight tracking, audio newsmagazines, prescription refilling).

It also allows input via speech recognition (SR) or "touch tone" DTMF and dialog prompting via synthesized speech (TTS) or recorded audio playback. And the experience with the platform shows that, it could be widely utilized in enterprises, groups and organizations with low-cost because of those improvements.

Keywords*: Voice Over Internet Protocol (VoIP), Session Initiation Protocol (SIP), VoiceXML, Text-To-Speech (TTS), Dual-tone multi-frequency (DTMF), IPPBX (Internet Protocol Private Branch Exchange)*

I. INTRODUCTION

A browser is a client application program that takes one or more input streams on a platform and executes an application that lives on one or more document servers by interpreting markup. In the case of VoiceXML, the application consists of the call flow logic, the prompts for the application, and any associated grammars, the document server executes portions of the application dialog by delivering VoiceXML markup to the browser in response to a document request. The markup interpreter renders the VoiceXML markup within an interpreter context, perhaps changing the context, and then makes calls into the implementation platform. The implementation platform contains all of the resources needed by the markup interpreter to render the dialog. This application deployed in IPPBX.

A private branch exchange (PBX) is a telephone routing system that directs all calls from outside lines and routes them to the appropriate phone. This type of system is most commonly used in an office space. PBXs make connections among the internal telephones of a private organization usually a business and also connect them to the public switched telephone network (PSTN) via trunk lines.

IP Communication solutions offer migration at an organization's preferred pace. By integrating with most of the major legacy PBXs and voicemail systems, as well as other business applications, most leading IP players empower customers to migrate to full IP based on their business needs, instead of being forced to adopt technologies due to limitations like interoperability of the various business applications. Successful customer migration to IPPBX communications is as much about processes as it is about technologies. Understanding this, leading industry players have developed detailed plans and processes that make migration smoother, faster, and easier for companies of all sizes.

In the IP communications world, telephony is just one of the services in the network. And, this service is available from anywhere in the network, independent of location. For example, a multisite business may deploy the call control (IP PBX) software only at the central site, then enable the remote sites to access the service remotely over the network.

II. SYSTEM ANALYSIS

2.1 Existing Systems

Till date different kinds of browsers are being used to browse the web content like Internet explorer, Mozilla, Netscape etc. which needs a typical computer and network connectivity to the web content. A web browser is a Software Application which enables a user to display and interact with text, images, videos, music and other information typically located on a Web Page at a website on the World Wide Web or a Local Area Network. Text and images on a Web page can contain hyperlinks to other Web pages at the same or different website.

2.2 Proposed System

The existing internet protocol network is connecting VOIP (Voice over Internet Protocol) together enabling users to make a call in a hassle free environment. Just a stable internet connection is required and a user can make a call from anywhere in the world independent of the location. Proprietary systems are easy to outgrow: Adding more phone lines or extensions often requires expensive hardware modules. In some cases an entirely new phone system is required. Not so with an IP PBX: a standard computer can easily handle a large number of phone lines and extensions just add more phones to your network to expand.

2.3 Feasibility Study

a) Operation Feasibility:

It is much easier to operate unlike the traditional web browsers. This project has been proposed in a user-friendly environment where the hardware requirement is very less. No proprietary software is required.

b) Technical Feasibility:

VoiceXML scripting is much easier to design. Any layman can understand VXML code. The hardware complexity is very less. All this project requires is a stable Internet connection, a processor with 1GB of RAM.

c) Cost Feasibility:

This project has one of the major benefits i.e. that most of the software that are being used here all available for Open Sources. The set up cost is very less. Just a stable Internet connection is required and a user can access it from anywhere in the world independent of the location.

2.4 Model Used

When the user places a call to the designated VoiceXML extension the call gets handed over to the pre-recorded hunt group. The particular code for a call flow gets activated when a call is placed. After the session is established successfully Asterisk gateway interface invokes the VoiceXML browser and calls the initial VoiceXML script hosted on the same server. Voice Glue starts interpreting the VoiceXML code for audio output and the user input by way of DTMF signals or voice commands.

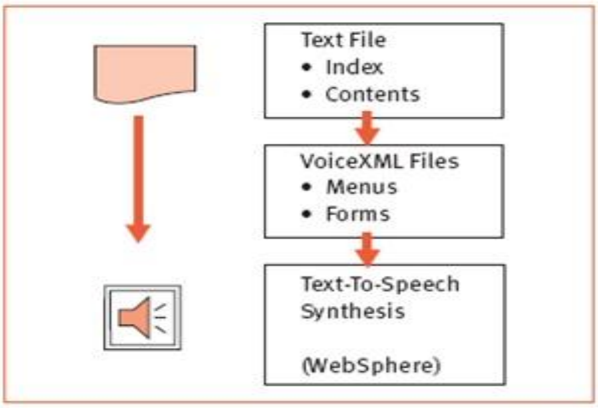

Fig 1: Text-To-Speech (TTS)

User can control the navigation of VoiceXML code with user's commands. Once the user is finished with his operation on VoiceXML the control is handed over to Asterisk (IPPBX). Asterisk comes as a part of preconfigured code terminates the call after the control is handed over from Voice Glue.

VoiceXML is a key used to transfer text to speech or database entries to speech in a very flexible way. Any text with its index of contents can be transferred from the document file into forms and menus of VoiceXML files that can be read out by text-to-speech synthesis tools like Web Sphere.

III. TECHNOLOGY OVERVIEW

3.1 Voice Over Internet Protocol (VoIP)

In just a few years, the old circuit-switched voice-centric communications network will give way to a data-centric, packet-oriented network that seamlessly supports data, voice, and video with a high quality of service. The switching equipment, protocols, and links are already being put into place. A transition network is currently in place that joins the packet data world with the circuit-switched world. Integrated access solutions are being installed that support integrated data, voice, and other media into the Internet or the PSTN.

Voice over Internet Protocol (VoIP) is a protocol optimized for the transmission of voice through the Internet or other packet switched networks. VoIP is often used abstractly to refer to the actual transmission of voice (rather than the protocol implementing it). VoIP is also known as IP Telephony, Internet telephony, Broadband telephony, Broadband Phone and Voice over Broadband. "VoIP" is pronounced voyp.

Despite a number of technological issues, real-time multimedia transmission (voice and video) over IP networks and the Internet has largely been worked out. Advanced compression techniques have reduced voice data transfer rates from 64 Kbits/sec to as little as 6 Kbits/sec. Voice over IP or VoIP can potentially allow users to call worldwide at no charge (except for the fee paid to service providers for Internet access). A user's IP address basically becomes a phone number. Additionally, computer-based phone systems can be linked to servers that run a variety of interesting telephony applications, including PBX services and voice messaging.

One of the best reasons to support packet telephony can be seen in the service limitations of the traditional telephone system. The switches are mostly proprietary with embedded call control functions and service logic. That makes it difficult to add new services. In addition, the end devices-telephones-are limited in functionality to a 12-key pad! In contrast, new services are easy to add in the IP telephony world because users simply add new telephony applications on their computers and communicate with other users who are running the same telephony applications.

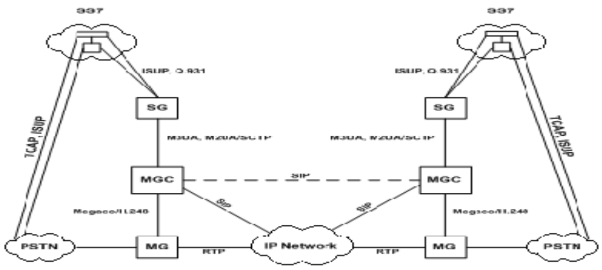

Fig 2: VoIP Architecture

3.2 Session Initiation Protocol (SIP)

There are many applications of the Internet that require the creation and management of a session, where a session is considered an exchange of data between an association of participants. The implementation of these applications is complicated by the practices of participants: users may move between endpoints, they may be addressable by multiple names, and they may communicate in several different media sometimes simultaneously. Numerous protocols have been authored that carry various forms of real-time multimedia session data such as voice, video, or text messages.

The Session Initiation Protocol (SIP) works in concert with these protocols by enabling Internet endpoints (called user agents) to discover one another and to agree on a characterization of a session they would like to share. For locating prospective session participants, and for other functions, SIP enables the creation of an infrastructure of network hosts (called proxy servers) to which user agents can send registrations, invitations to sessions, and other requests. SIP is an agile, general-purpose tool for creating, modifying, and terminating sessions that works independently of underlying transport protocols and without dependency on the type of session that is being established. SIP is generic protocol for every IP capable access networks.

The Session Initiation Protocol (SIP) is an application-layer control (signaling) protocol for creating, modifying, and terminating sessions with one or more participants. It can be used to create two-party, multiparty, or multicast sessions that include Internet telephone calls, multimedia distribution, and multimedia conferences. SIP is designed to be independent of the underlying transport layer; it can run on TCP, UDP. It was originally designed by Henning Schulzrinne (Columbia University) and Mark Handley (UCL) starting in 1996. It is a 3GPP (Third Generation Partnership Project) signaling protocol. It is one of the major signaling protocols used in Voice over IP (VoIP).
SIP handles the signaling part of a communication session.

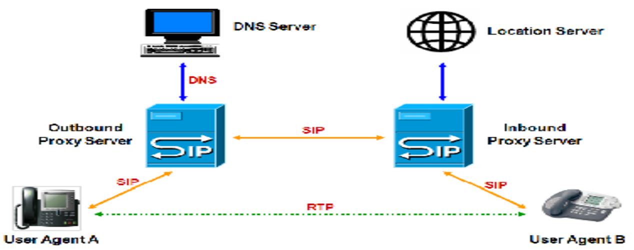

Fig 3: SIP Trapezoid Architecture

SIP handles the signaling part of a communication session. It serves as a carrier for the Session Description Protocol (SDP). SDP handles the media portion of the session. The transmission of voice and video content are done by the Real-time Transport Protocol (RTP). A SIP session thus involves packet streams of RTP. SIP is a part of the protocols involved in a multimedia session. The latest version of the specification is RFC 3261 from the IETF SIP Working Group.

3.3 Internet Protocol Private Branch Exchange (IPPBX)

An IP PBX or VOIP phone system replaces a traditional PBX or phone system and gives employees an extension number, the ability to conference, transfer and dial other colleagues. All calls are sent via data packets over a data network instead of the traditional phone network.

An IP PBX is a complete telephony system that provides telephone calls over IP data networks. Typically an IP PBX system is a piece of software running on a server. Depending on the workload, that server can also be performing other tasks, but usually it is dedicated and also acts as the VoIP system's connection to the internet.

An IP PBX is a telephone switching system inside an enterprise that switches calls between Voice over IP (VoIP) users on local lines and lets all users share a certain number of external telephone lines. The typical IP PBX can also switch calls between a VoIP user and a traditional telephone user, or between two traditional telephone users much like a conventional PBX does. The IP PBX is also able to connect to traditional PSTN lines via an optional gateway so upgrading day to day business communication to this most advanced voice and data network

Internet protocol private branch exchange (IP PBX) market offers a ray of hope in the otherwise depressed European telecommunications industry. Encouraging developments in this market have seen enterprises beginning to replace their time division multiplexing (TDM) voice networks with IP enabled/converged voice data networks.

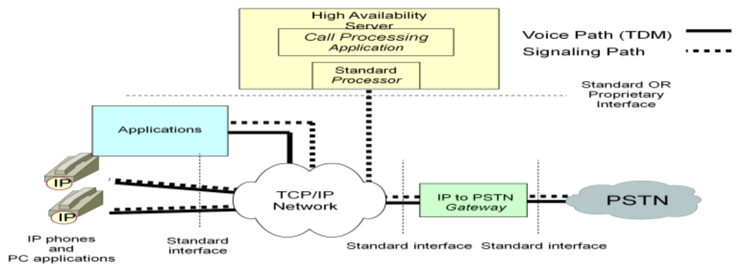

Fig 4: IPPBX Architecture

An IP PBX or IP Telephone System consists of one or more SIP phones, an IP PBX server and optionally a VOIP Gateway to connect to existing PSTN lines. The IP PBX server functions in a similar manner to a proxy server: SIP clients, being either soft phones or hardware-based phones, register with the IP PBX server, and when they wish to make a call they ask the IP PBX to establish the connection. The IP PBX has a directory of all phones/users and their corresponding SIP address and thus is able to connect an internal call or route an external call via either a VOIP gateway or a VOIP service provider.

3.4 VoiceXML

The Extensible Markup Language (XML) is a general-purpose markup language. It is classified as an extensible language because it allows its users to define their own elements. Its primary purpose is to facilitate the sharing of structured data across different information systems, particularly via the Internet, and it is used both to encode documents and to serialize data.

In 1998, W3C hosted a conference on voice browsers. By this time, AT&T and Lucent had different variants of their original PML, while Motorola had developed VoxML, and IBM was developing its own SpeechML. Many other attendees at the conference were also developing similar languages for dialog design; for example, such as HP's TalkML and Pipe Beach's VoiceHTML.

The Voice XML Forum was then formed by AT&T, IBM, Lucent, and Motorola to pool their efforts. The goal of the Voice XML Forum was to define a standard dialog design language that developers could use to build conversational applications.

In 2000, the Voice XML Forum released Voice XML 1.0 to the public. Shortly thereafter, Voice XML 1.0 was submitted to the W3C as the basis for the creation of a new international standard. Voice XML 2.0 is the result of this work based on input from W3C Member companies, other W3C Working Groups, and the public.

Voice XML is designed for creating audio dialogs that feature synthesized speech, digitized audio, recognition of spoken and DTMF key input, recording of spoken input, telephony, and mixed initiative conversations. Its major goal is to bring the advantages of Web-based development and content delivery to interactive voice response applications. A common architecture is to deploy banks of voice browsers attached to the Public Switched Telephone Network (PSTN) so that users can use a telephone to interact with voice application.

Here is a short example of Voice XML. This is a Hello World example:

```
<?xml version="1.0"?>
<vxml
version="2.0"xmlns="http://www.w3.org/2001/vxml">
<form>
<block>Hello World!</block>
</form>
</vxml>
```

The top-level element is <vxml>, which is mainly a container for dialogs. There are two types of dialogs: forms and menus. Forms present information and gather input; menus offer choices of what to do next. This example has a single form, which contains a block that synthesizes and presents "Hello World!" to the user. Since the form does not specify a successor dialog, the conversation ends.

The architectural model assumed by this document has the following components. A document server (e.g. a Web server) processes requests from a client application, the VoiceXML Interpreter, through the VoiceXML interpreter context. The server produces VoiceXML documents in reply, which are processed by the VoiceXML interpreter. The VoiceXML interpreter context may monitor user inputs in parallel with the VoiceXML interpreter.

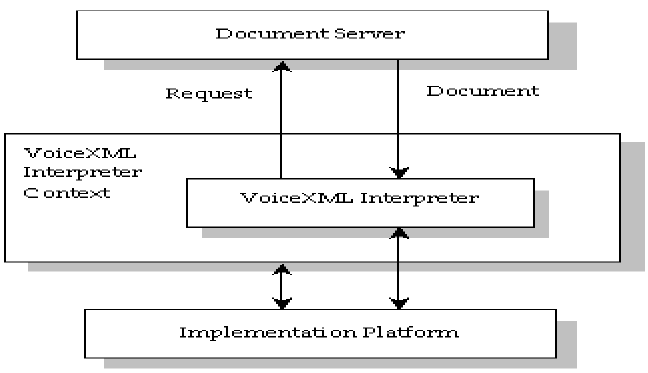

Fig 5: VoiceXML Architecture

For example, one VoiceXML interpreter context may always listen for a special escape phrase that takes the user to a high-level personal assistant, and another may listen for escape phrases that alter user preferences like volume or text-to-speech characteristics.

The implementation platform is controlled by the VoiceXML interpreter context and by the VoiceXML interpreter. For instance, in an interactive voice response application, the VoiceXML interpreter context may be responsible for detecting an incoming call, acquiring the initial VoiceXML document, and answering the call, while the VoiceXML interpreter conducts the dialog after answer. The implementation platform generates events in response to user actions (e.g. spoken or character input received, disconnect) and system events (e.g. timer expiration).

Some of these events are acted upon by the VoiceXML interpreter itself, as specified by the VoiceXML document, while others are acted upon by the VoiceXML interpreter context.

The language describes the human-machine interaction provided by voice response systems, which includes:

1) Output of synthesized speech (text-to-speech)
2) Recognition of spoken input
3) Recognition of DTMF input
4) Recording of spoken input
5) Control of dialog flow
6) Telephony features such as call transfer and disconnect

The language provides means for collecting character and/or spoken input, assigning the input results to document-defined request variables, and making decisions that affect the interpretation of documents written in the language. A document may be linked to other documents through Universal Resource Identifiers (URIs).

VoiceXML has become a standard and has the following advantages:

1) Reduces development costs
2) Separation between dialogue system components
3) Portability of application
4) Re-use of Internet infrastructure
5) VoiceXML is becoming a standard
6) Reduces dialogue system development time
7) Additional functions can be implemented
8) Can develop own dialogue system with free VoiceXML browsers

IV. SYSTEM DESIGN

4.1 UML Diagrams:

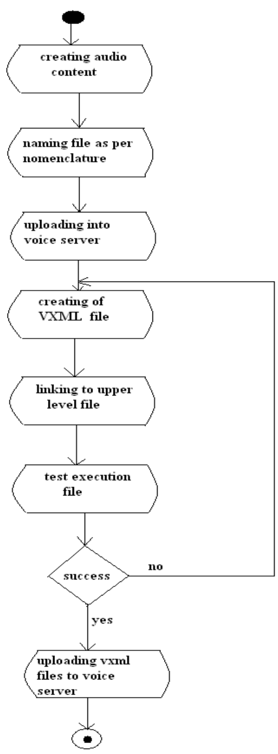

Fig 6: Creating Audio Database

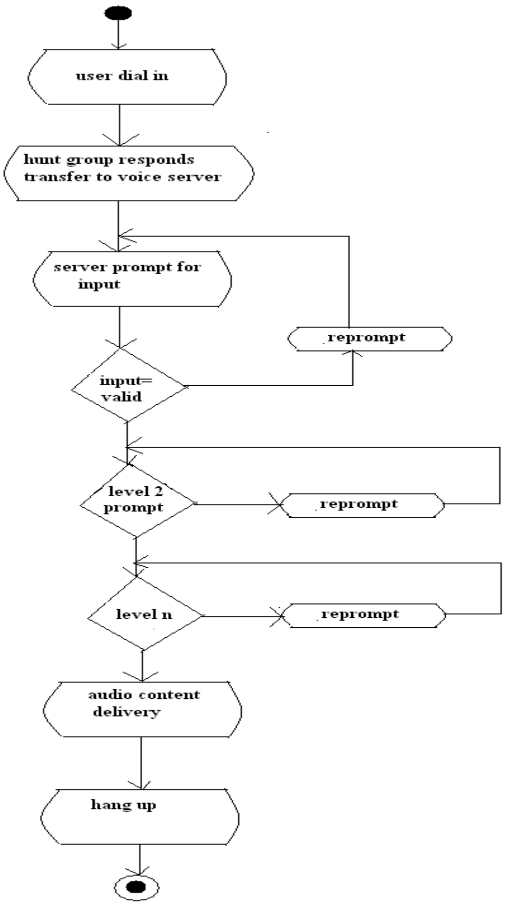

Fig 7: Flow of Execution

V. TESTING AND IMPLEMENTATION

Test 1:
Placing two concurrent call on the vxml system

Test 2:
Checking the code how when wrong option selected

5.1 Test Execution:
Case 1:
Call from two extensions 101,102 simultaneously

Case 2.1:
After the welcome message select ECE by pressing choice 3

CASE 2.2
After selecting ECE we choose first year by pressing 1.

5.2 Result
Case 1:
Two calls successfully able to hear the welcome message on both the calls

Case 2.1:
Able to hear the welcome message of ECE department and prompts to select year

Case 2.2:
The calls suppose to go to the year level menu but call get routed Thank you and it got hang up.

VI. CONCLUSION

In this paper, we proposed the Customized IVR implementation using VoiceXML on VoIP platform. VoiceXML provides flexibility at robust technology which delivers its output in form of audio which will be very much useful to all kinds of people starting from the very busy business man to layman who doesn't even know how to read.

VoiceXML enables all its robust features in lines with existing coding methodology used in today's very popular HTML and XML. It makes the development of VoiceXML very easier and adoptable to presently existing system.

References

[1] Benbin Chen, Wenchao Zhou, Yiyang Li ,Donghui Guo, Innovative application of SIP protocol for communication platform , Anti-Counterfeiting Security and Identification in Communication (ASID), 2010 International Conference, IEEE,2010
[2] Larson, J.A, W3C Speech Interface Languages: VoiceXML, Vol 24, IEEE, 2007
[3] Jianfeng Zhu, Zhuang Li, Yuchun Ma, Yulin Huang, Realization of Extended Functions of SIP-Based IP-PBX, Vol 3, IEEE, 2010
[4] Prasad, J.K., Kumar, B.A, Analysis of SIP and realization of advanced IP-PBX features , Vol 6, IEEE, 2011
[5] Gokhale, S.S , Jijun Lu, Signaling performance of SIP based VoIP: a measurement-based approach , Vol 2, IEEE, 2005