# A Voice-enabled Interactive Services (VòIS) Architecture for e-Learning

Luvai F. Motiwalla
*Department of Operations & Information Systems*
*University of Massachusetts Lowell*
*1 University Ave,*
*Lowell, MA 01854*
*luvai_motiwalla@uml.edu*

***Abstract---Mobile devices have gained global popularity for business and government uses. Yet, these devices are not commonly used in e-learning environments. One reason is the difficulty of interacting with e-learning tools like forums or chats from mobile devices which usually have small screens and limited keyboards. The goal of this paper is to discuss a voice enablement approach that can provide a solution to this problem. This approach can enhance mobile learning with an architecture that can make communication and collaboration easy from mobile devices. The paper discusses a pilot study with voice-enabled mobile learning project which was evaluated with blind students from higher education institutions in USA. The lessons learned from this project enable us to propose a voice-enabled interactive services (VòIS) architecture which can increase use of voice recognition technology in mobile learning. Results from this exploratory study, though small, specific, and culturally biased, will hopefully lead to proper investments in mobile learning applications and more research on understanding the role of mobile devices in education.***

***Keywords--Distance education, e-learning architecture, speech recognition, interactive voice response, access for the disabled.***

## I. INTRODUCTION

This paper is an extended version of the presentation at the LMPCNA 2009 conference [1]. According to Gartner survey [2], mobile devices sales were 32.2 million units in second quarter of 2008 with North American market showing fast growth by 78% over the year. Today, smart phones sales have outstripped PC sales. Similarly, computing and mobile devices have become ubiquitous on today's college campuses. According to Harris Interactive College Explorer Study [3], today's college students are connected via digital communication devices throughout the day spending on average eleven hours with their devices and roughly 1.3 million students have smart phones with which they communicate daily. The massive infusion of computing devices and rapidly improving Internet capabilities are altering the learning and teaching pedagogies in higher education [4]. A 2000 Campus Computing Survey revealed that the majority of college professors use email to communicate with their students, and approximately one-third of college courses use Web resources or have a class Web page (National Survey of Information Technology in US Higher Education). Similarly, Jones [5] reports that a great majority of college students using digital media 80 percent believing that Internet use has enhanced their learning experience.

Despite the tremendous growth and potential of the mobile devices and networks, wireless e-learning or mobile learning (*m-learning*) is still in its infancy and in an embryonic stage. *M-Learning* intersects mobile computing and e-learning; it combines individualized (or personal) learning with anytime and anywhere learning [6]. The relationship between a person and their mobile device is usually one-to-one, always on, always there, location aware and personalized [7]. The location independence of mobile devices provides several benefits for learning environments like allowing learners and instructors to utilize their spare time while traveling in a bus or train to finish their homework or class preparations [9]. It also has the potential to change the way students behave and interact with each other because wireless access technology "takes e-learning to the field, where the best hands-on learning takes place." [13]. But, one key problem with these devices is the limited space for data/text display and telephone key-pads for text input. This makes them cumbersome for interactive pedagogies used in learning environments [14].

Online learning tools such as discussion forums, blogs, and wikis are often considered good surrogates for classroom interaction [12]. According to a National Center for Education Statistics (NCES) study [8], 90% of institutions offering e-learning courses used one or more communication technologies such as online discussion forums, blogs, wikis, and chat rooms to facilitate classroom interaction among students and instructors to discuss course materials. This means that interactive tools add value in online learning pedagogy. Research on the introduction of technology in education [9] has shown that it is effective only when developers understand the strengths and weaknesses of technology. One of the major concerns of implementing m-learning technology is the accessibility issue. Mobile devices usually have simplified input devices and smaller screens than normal desktop computers. Typing messages using a cell phone keypad or reading articles on small screen devices can be cumbersome for users. While these tools are relatively easy on computers, they are difficult for people without computers. Voice or speech enablement has the potential for overcoming the cumbersome user interface issues and unlocking the potential use of these devices for mobile learning and other application areas. This paper discusses outcomes from a pilot project on voice enablement of m-learning environment. The lessons learned from the pilot study enabled us to define the technical requirements of voice-enabled e-learning services platform from which m-learning applications, like forums and other Web 2.0 tools, could be launched to increase the usage of mobile devices for interactive learning. Before discussing our *VòIS* architecture for e-learning, a brief background on interactive voice recognition technology is provided below, followed by a discussion on our pilot study and proposed architecture design for voice-enablement of e-learning (interaction) tools. The paper concludes by articulating some key issues underlying the usage of voice recognition technology and its impact on m-learning environments.

## II. INTERACTIVE VOICE RECOGNITION TECHNOLOGY

From making airline reservations to confirming postage rates, consumers are increasing their acceptance of applications that utilize synthesized speech or voice recognition. While the public can be unforgiving when it comes to the naturalness of synthesized speech, demand for speech applications has been steadily increasing. According to Forrester Research [10], the speech recognition market has reached an inflection point. After many years of slow growth, the year 2006 showed increased spending on speech technology with a compound annual growth rate of 52% versus 11% for IVR spending over the last five years. The percentage of companies who have deployed speech grew four-fold – from 10% in 2000 to 40% in 2003 – with an additional 32% planning to deploy speech every year. This growth suggests that for most online learning programs, the question is no longer deciding whether to adopt speech, but, rather, determining which applications are most suitable for speech and developing a speech strategy that most effectively complements and integrates with an overall e-learning strategy.

M-learning has not yet fully embraced voice recognition technology to complement its existing learning technologies. Online interaction tools such as discussion forums, blogs, and wikis are often considered good surrogates for classroom interaction [11, 12]. According to a National Center for Education Statistics (NCES) study [8] 90% of institutions offering online Internet courses used one or more communication technologies such as online forums, blogs, wikis, and chat rooms to facilitate classroom interaction among students and instructors to discuss course materials. This demonstrates the value of interaction tools for improving the online pedagogy.

While interactive tools are relatively easy to use for people with full eye-sight, they are incompatible for people without eye-sight (or blind-disabled). Similarly, the American Foundation for Blind (http://www.afb.org) estimates roughly 10 million people in the United States with visually impairments, only 15% use computers and Internet. This leaves 8.5 million without access to the online interaction tools. Speech or voice enabling technologies, such as text-to-speech (TTS) technology and automatic speech recognition (ASR), has the potential of overcoming access and user interface deficiencies for the disabled. Currently, however, speech technology usage is limited to integrating Web browsers with screen readers, zoom text, and alternative media such as audiotapes or audio descriptions [12]. These applications require disabled to be sophisticated with computer usage and have the ability to afford computers and the assistive technology; thus, limiting their full integration into online learning and employment. Therefore, an added benefit of speech technology is that it expands mobile learning to the disabled. Another goal of this project is to develop a voice-enabled platform that works seamlessly with interactive learning tools such as discussion forums, wikis, blogs, instant messaging and others, to increase the access and participation for the disabled population in online and mobile learning environments.

Speech technology has traditionally been an esoteric and expensive technology available only to big businesses and research labs. Combining speech recognition with the simplicity of markup languages like VoiceXML makes it dramatically simpler to develop a Voice User Interface (VUI). Furthermore, the wide proliferation of speech applications, until recently, has been impeded by the fact that it is error-prone and still cannot handle natural language. The Web provides a relatively simpler framework for interaction with a computer or phone. Within this framework, several self-service applications have been developed in customer service and help desk environments accessible from computers or telephones.

Today, interactive voice response (IVR) applications can adequately handle the basic self-service requirements of navigating menus, traversing links, and entering data into forms. Combining recordings with speech synthesis is sufficient to develop intelligible voice interfaces that are hands-free and can be seamlessly integrated with other Web-based systems. The next section discusses our pilot study which was conducted to test the technical feasibility of voice-enable online learning tools like discussion forums and making them voice accessible from mobile phones.

## III. PILOT STUDY ON VOICE ENABLEMENT OF A DISCUSSION FORUM

A pilot study was conducted to determine the feasibility of voice-enabled architecture in higher-education learning environment. The goal was to enable us to learn the requirements for a broader voice enabled learning platform discussed later in this paper. The prototype, called Voice-enabled Discussion Forum (VeDF), was tested with 11 blind users to access online discussion forums for over 45 days after which they were surveyed providing both quantitative and qualitative support for our study. The reason for selecting only blind students was that these users would provide us with the best feedback on voice interaction as they have some experience of using voice enabled devices and the best motivation of using this technology as it enables them to participate in online learning. Our plan is to expand to both disabled and non-disabled students in the larger study with the full-scale applications later.

Specifically, the objectives of our study were to:

- setup an online discussion forum on our learning Architecture
- setup and design voice recognition system accessible from a telephone and link it to the online discussion forum
- provide seamless integration between voice-enabled and regular web discussion forums
- implement and test the prototype VeDF system
- create real-world forums for evaluating the prototype
- select a sample group of blind and/or visually impaired users for the evaluation study
- test the IVR system with the sample group
- collect preliminary feedback from sample users on the system and
- determine the potential of VòIS architecture.

In general, users found our application friendly and were able to accomplish their discussions in acceptable time frame. In this pilot study, students used an online discussion forum to exchange text messages from handheld devices [13]. The same forum was accessible from a traditional PC computer and a Smartphone. An earlier study [14] had evaluated the same online (text only) forum with students from several courses in both online and traditional classroom environments. This study found the anytime/anywhere convenience of accessing the forum very useful but the user interface very cumbersome because it required them to type text messages using the phone keypad. The feedback from the students in this study led to the development of a voice-enabled architecture for mobile-and other phone devices [15]. This application used a server-based screen reader which converts all text messages to voice on-demand and also facilitates the storage of voice recorded messages on the server. The project's focus is on using text-to-speech (TTS) technology to convert the text messages to voice format on the web server and make them available to the students through the discussion forum application. These projects shifted our focus on making online learning more accessible by using speech recognition technologies through a hybrid of computer and telephone devices.

A hosted telephone/speech gateway solution from a Voice Service Provider (VSP) company was used to help us quickly develop a prototype and test our architecture with students at our university. This VSP provided us with an environment equivalent to Internet Service Provider (ISP) for accessing the Internet. They provide voice gateway and speech Architecture software which was linked to our host server via a secure hyper-text transfer protocol (HTTPS) connection. We have developed a customized VeDF application using the VoiceXML standard and interfaced it with an existing online learning discussion forum on our server. More details on this are provided in the sections below.

### *A. Prototype Configuration*

Our voice-enabled discussion forum has several unique features that can appeal to the disabled and non-disabled users alike. The success of this technology has the potential to attract a wide variety of applications that can be voice-enabled in the future. Here are some of these features of our application:

- No training requirements – user needs to know how to use phone
- Synchronized phone and web browsing of message boards
- Voice and data integrated discussion board
- One voice command functionality
- Message thread recording and replay
- Cross Architecture (telephone and computer) access
- Caller-ID login: registered users can login only with password
- Customizable voice user interface: your welcome messages
- Scalable (1000's of participants in a single forum)
- Instant messaging with searchable message log
- VoIP or PSTN phone access
- Individual user accounts and passwords
- Superb audio conferencing
- Best in class use speech recognition and TTS technology
- On demand help availability
- Available 24x7 for unlimited use

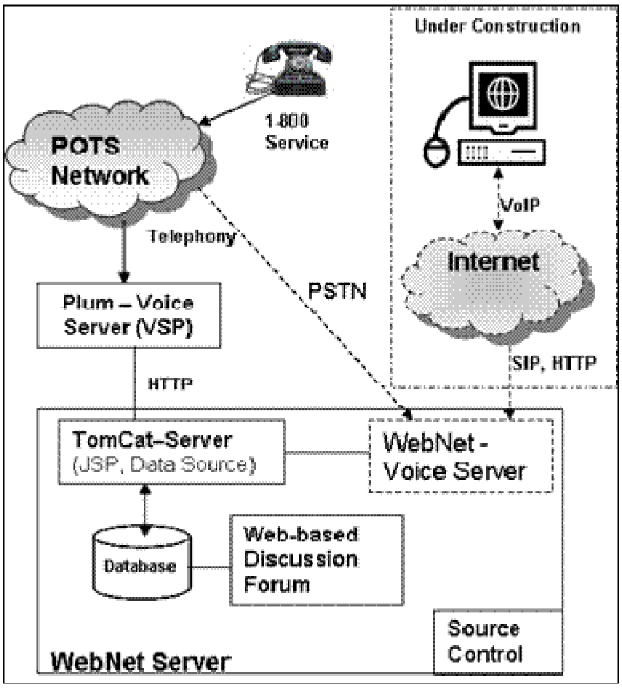


Figure 1: Pilot Prototype Configuration

As mentioned earlier, this project's accomplished its technical goal of developing and testing a voice-enabled discussion forum (VeDF) application by integrating a commercially available speech Architecture with the online discussion forum software. Our IVR application was developed with the *Open Source* software model. The use of Plum hosting services™ as the VSP is a temporary arrangement to let us quickly demonstrate the feasibility of the application. We plan to replace the Plum-VSP Architecture with our own voice server Architecture in the next phase. The WebNet voice server Architecture has been configured and internally tested with the Nuance Voice Architecture™ and NMS Communications™ CG series telephony card for developing IVR system with telephony access via VoIP/SIP and PSTN communications protocols. This Architecture will replace the Plum VSP services for full-scale development and testing environment in next phase. We will use Plum only for the production environment as it provides a better 24x7 support, maintenance and uptime. The overall configuration of our Pilot prototype application, with indication for future changes in dotted-line, is shown in Figure below.

In the above configuration, a user can call a toll-free number provided by the Plum hosting service from any telephone device. This call is processed by Plum and re-directed to our IVR system. Plum provides the voice gateway to the WebNet server, a Dell PowerEdge™ 3.0GH$_z$ RAID-1 server with dual-processor capability. This server hosts our IVR system designed with the VoiceXML standard. Plum provides a text-to-speech engine (TTS) from AT&T Natural Voice™ and automated speech recognition (ASR) software from SpeechWorks™ which are used by our IVR system. Our IVR system is developed on the Java™ Architecture with the Tomcat web server using J2EE software language, JSP scripting language, and JDBC database environment. The IVR system was connected to the web-based discussion forum through the database which provided the content for the IVR application. A source control environment, with Subversion™, has been installed on the WebNet server for multiple programmers to check-in their software and quality control. We plan to replace the Plum service with our own voice server Architecture for the development and staging environment in next phase. This voice server is developed with the Nuance Voice Architecture and will have both SIP (for VoIP phones) and PSTN (analog phones) connectivity with the NMS Communications CG series telephony card installed on our server, as mentioned earlier.

### *B. Voice Forum Application Architecture*

Our IVR system is developed using technology that supports the open public standards and in particular, uses the VoiceXML scripting language. VoiceXML was created to simplify the development of Interactive Voice Response (IVR) applications. Thanks to active support from America's leading technology companies, including AT&T, IBM, Lucent, Microsoft, Motorola, and Sun, the VoiceXML language standard entered the public domain in 2000 as the accepted lingua franca for voice and IVR applications. VoiceXML provides a mature feature set - a superset of traditional IVR features - for handling conventional telephony input, output and call control, including: touch-tone input, automatic speech recognition support, audio recording (e.g., for voice mail), the ability to play recordings (such as .wav files), speech synthesis from plain or annotated text, call transfer, and other advanced call management features.

General benefits of fully standards-compliant VoiceXML IVR system include:

- *Elimination of Vendor Risk* -- VoiceXML applications can be easily ported among VoiceXML IVR machines. This portability

™ Plum Voice Portals (http://www.plumvoiceportals.com/)
™ Nuance, Corp.

™ Dell Computers, Inc.
™ AT&T, Corp.
™ Nuance, Corp.
™ Sun Corp.
™ Subversion, Inc.

eliminates reliance on any one vendor, and Plum has to earn the right for continued business.

- *Standardization of Applications* -- VoiceXML is a well documented and very popular public standard similar to HTML -- avoids reliance on proprietary APIs or marginal solutions.
- *Flexible Deployment Options* -- VoiceXML apps can run within self-contained systems or via hosted gateways.

Our pilot VoiceXML IVR system architecture consists of the Plum *Telephony Gateway* controlled by the Plum *VoiceXML Interpreter;* these components together form the heart of a *VoiceXML IVR Gateway*. The gateway seeks VoiceXML instructions from any *Application Server*, which was hosted on the WebNet server. This application server provided a dynamic *VoiceXML Software Suite* consisting of a *Controller Servlet*, *Voice User Interface (VUI)* server pages and *Data Source* objects which integrated our IVR with the online forums *Database*. Plum's *Admin Tools* enable browser-based configuration of the gateway, application setup, modification, user management, reporting, etc. This modular implementation allows for *Future Plug-ins* of other *ASR* and *TTS* engines into the VoiceXML IVR Gateway and Admin Tools can be expanded to allow for enhanced application control.

For this pilot implementation, as mentioned before, we used Plum Hosting Services™ to provide a quick way to get started while eliminating large capital expenses. Plum provided a per-minute and per-port hosting environment to develop and test this early version of our prototype.

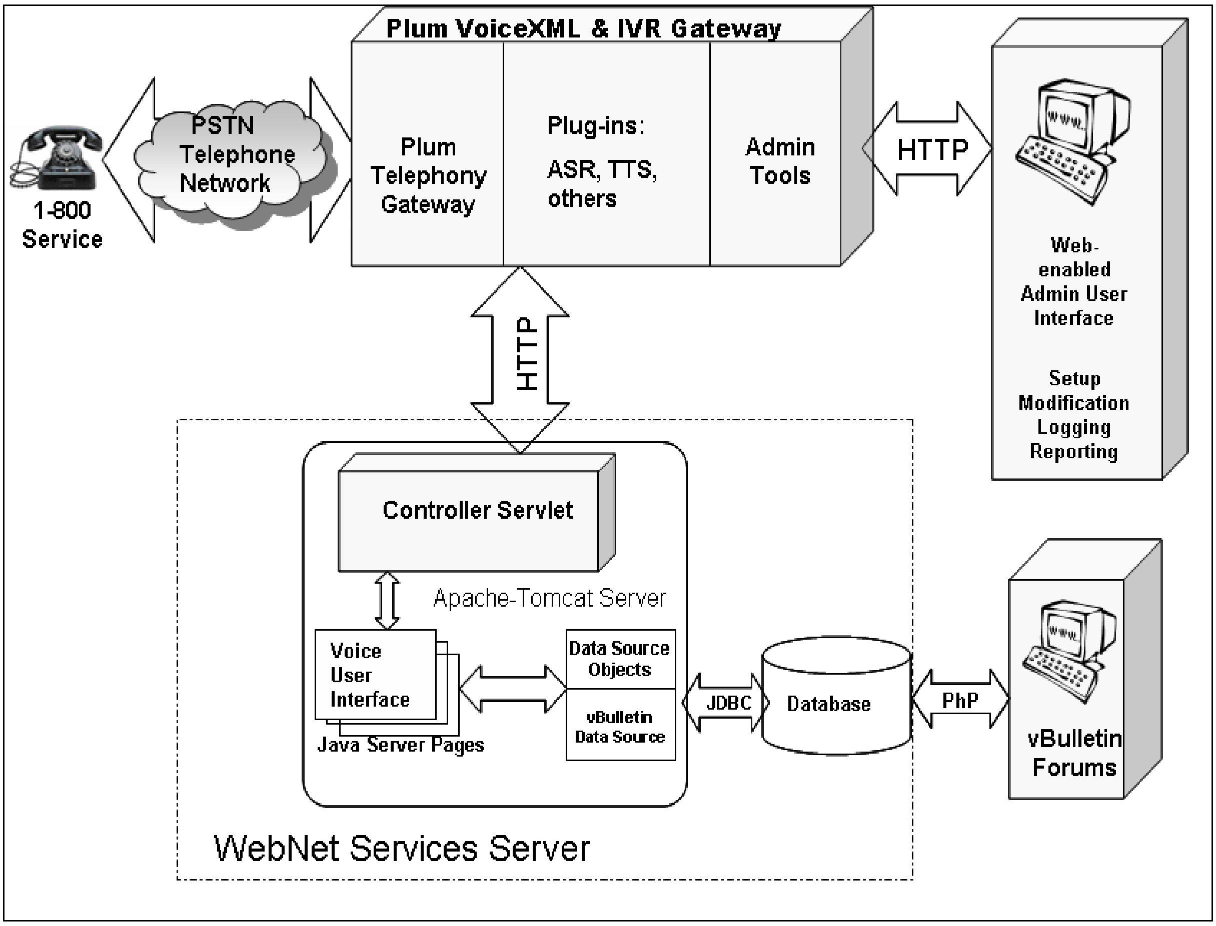

Figure 2: Voice Forum Application Architecture

The VeDF application architecture, shown in Figure 2, was designed using a layered and modular architectural design to support for future enhancements and integration of other application tools, while allowing for quick and dynamic code flow changes, as is required in a typical initial proof of concept application. This prototype application consisted of Java Server Pages (JSP), interacting with java class objects which implemented a simple interface to a given data sources. Interactions with the existing web based software where broken down into distinct steps, such as 'Login', 'Post Message', 'Get Forum List'. After an initial usable prototype, it was decided to also include the use of a controller servlet that dispatched the given state of the application to the associated JSP page. This allowed the controller servlet to dynamically change the applications path at run time, allowing for global user authentication, posting of new threads, and providing greater security to the end users, as security verification was able to take place with every interaction with the system.

Generally, this environment allowed us strict call flow control within the application, while still allowing dynamic code deployment and execution utilizing JSP pages. These JSP pages could change dynamically at run-time; give us much greater flexibility, while still maintaining security and control. The technical goal of this pilot study was to develop a prototype tool for students to access online discussion forums; this can potentially reduce access barriers that prevent them from fully integrating, participating and excelling in online education programs.

### C. *VeDF - Voice User Interface*

1) The user connects through a phone and the public switched telephony network (PSTN) to the VoiceXML Interpreter through the Plum telephony gateway.
2) The VoiceXML Interpreter conducts the call interaction with a caller based on the instructions of a VoiceXML script supplied by the application server. The Interpreter natively detects touchtone input and can manage pre-recorded audio prompts or files. The Interpreter also calls the TTS (text-to-speech) and ASR (automatic speech recognition) for enhanced voice functionality.
3) The VoiceXML Interpreter communicates via web protocols (HTTP) to our remote VoiceXML application server.
4) The VoiceXML application server delivers the application, including VoiceXML text pages and binary audio files. The application server receives spoken, touchtone, and recorded input from the VoiceXML Interpreter.
5) The VoiceXML application server queries the database via the data source objects to dynamically drive VoiceXML to the interpreter (hence "speak" with the caller).
6) Both Human and Machine Personas were used to provide the audio prompts to the end-users. The human persona was professionally developed through the recording studio using a local radio station.

In sum, the key variation from the originally proposed architecture was that the speech recognition functionality was accessed from the voice service provider (VSP) host, Plum Voice Portals, Inc. and integrated with our server which provided the IVR functionality through VoiceXML scripts. These scripts then used the Java™ Architecture to access the discussion forum messages and served them to the user with the VoiceXML gateway of Plum Voice Portal. This architecture is designed to be scalable for growth and flexible to be integrated with other forums at the back-end and voice Architectures at the front-end. The Voice User Interface (VUI) was developed with the VXML 2.1 standard from W3C organization. This standard permits the use of voice Architecture from any vendor that supports the VXML standard. The data from the discussion forum were extracted using the Java Data Base Coding (JDBC) without making any modifications to the vBulletin™ discussion forum application or the DF database. This permits quick flexibility to integrate our voice forum application with other discussion forums software vendors that support the JDBC standard.

### D. *Pilot Study Results*

The pilot study has allowed us to investigate the feasibility for a voice-enabled discussion forum application, as well as design, develop and evaluate a prototype application with blind and visually impaired learners from different organizations. The prototype provides the basis for the architecture for developing a commercial voice-enabling services Architecture that can be extended to other communication and interaction tools such as blogs, wikis, pod casting, and instant messaging services current available for use on over the Internet with minimal effort. End-users were an integral part of the design and prototype development effort; this user involvement allowed us to discover new requirements to change application functionality and system components. Changes to the application architecture where required, as the initial request for a feature was often found to be an indicator of an additional requirement. Because this prototype was primarily created for research and feasibility studies, these changes were made fairly quickly. Some of these ideas and methods will be helpful in developing the VòIS architecture.

Before development of prototype, two blind users were interviewed and observed as they used commercial IVR/Call center applications. As the prototype was being developed it was tested with the core group of ten users from higher education programs. Users were shown a demo and provided an instruction sheet on how to use the application and were given tasks like software training, joining a social club and course material discussion. Case scenarios were developed for each organization to provide a realistic testing environment. Problems

™ vBulletin, Inc (www.vbulletin.com)

encountered were recorded and fixed within 48-72 hours time period. This process continued for over 45 days, after which users were surveyed on various aspects, including system performance and user satisfaction, with a standardized questionnaire. There were some technical problems initially with the system and breakdowns. But after the initial learning curve, users liked accessing information from the phone, and were generally satisfied with the systems performance and interface.

An empirically validated survey instrument [16] for measuring learner satisfaction for e-learning systems was customized for our pilot study. The students were given written instructions on how to access the voice forum through a handout which listed the steps on how to use the phone to login, access the website, and participate in the discussion board and chat. A fixed chat time and day to participate in the chat and discussion questions were posted on the bulletin board. Students were given the incentive to participate in form of class participation grade for graduate students and extra-credit for undergraduate students.

Of the ten users who participated in our prototype evaluation study, two users did not respond to the survey and therefore, the results are based on sample of eight users. The survey was divided into two parts: 1) 12 survey questions on the Voice-enabled discussion forum (VeDF) system which required users to circle their answer on a five-point Likert scale with 5=Strongly Agree, 3=Neutral, 1=Strongly Disagree). 2) 12 questions on users' background. The users were provided with a help to fill-out the survey either by the researcher or through the Blind center where they visited. Surveys were e-mailed or faxed along with an Informed Consent Form. **Error! Reference source not found.Error! Reference source not found.** below presents an analysis of the feedback received from the users on the survey conducted at the end of the evaluation period.

TABLE 1: RESULTS FROM ALL END USER SURVEY

| Survey Questions | *AVG* | *STD DEV* | VAR |
|---|---|---|---|
| VeDF did a good job in providing access | 4.00 | 1.10 | 1.20 |
| VeDF met all my expectations | 3.50 | 1.38 | 1.90 |
| VeDF user interface was easy | 3.33 | 1.21 | 1.47 |
| VeDF prompts were easy to understand | 4.33 | 0.52 | 0.27 |
| Average time it took to accomplish a task was acceptable | 3.67 | 1.03 | 1.07 |
| VeDF is useful for disabled students to access online learning | 4.00 | 1.26 | 1.60 |
| VeDF can be useful tool to access the other online materials | 4.33 | 0.82 | 0.67 |
| VeDF makes the process of discussing with other student easier | 4.00 | 0.00 | 0.00 |
| I prefer using VeDF instead of screen readers like JAWs | 3.00 | 1.10 | 1.20 |
| I prefer using VeDF instead of BRAILLE to access online forums | 4.00 | 1.22 | 1.50 |
| I foresee a good demand for this technology from the blind | 4.33 | 0.82 | 0.67 |
| I'm willing to use phones with IVR technology instead of computers | 4.17 | 0.41 | 0.17 |

The survey on VeDF prototype indicates a general satisfaction on the use of this system for accessing online materials from the users. Users were neutral on some survey items like substituting VeDF for JAWs. On questioning the users later we found that they misunderstood the question. Our survey did not include the "to access online forums" in this question which resulted in general comparison of the two systems. They were also more familiar with JAWs having used it for a longer time and it can be used with many other programs. Secondly, we found that they were unhappy with the system audio prompts because we used the machine personas available from AT&T Natural Voice system. To fix this problem, we contacted a local radio station and asked them to record the system's audio prompts. These human prompts were much clearer and as you can see from the results, the VeDF prompts received a score of 4.13 out of 5.

TABLE 2: RESULTS FROM SURVEY WITHOUT THE TWO USERS

| Survey Questions | *AVG* | *STD DEV* | VAR |
|---|---|---|---|
| VeDF did a good job in providing access | 3.75 | 1.49 | 2.21 |
| VeDF met all my expectations | 3.50 | 1.41 | 2.00 |
| VeDF user interface was easy | 3.25 | 1.49 | 2.21 |
| VeDF prompts were easy to understand | 4.13 | 0.99 | 0.98 |
| Average time it took to accomplish a task was acceptable | 3.63 | 1.19 | 1.41 |
| VeDF is useful for disabled students to access online learning | 3.88 | 1.36 | 1.84 |
| VeDF can be useful tool to access the other online materials | 4.25 | 0.89 | 0.79 |
| VeDF makes the process of discussing with other student easier | 3.75 | 1.16 | 1.36 |
| I prefer using VeDF instead of screen readers like JAWs | 3.13 | 1.25 | 1.55 |
| I prefer using VeDF instead of BRAILLE to access online forums | 4.14 | 1.07 | 1.14 |
| I foresee a good demand for this technology from the blind | 4.13 | 1.13 | 1.27 |
| I'm willing to use phones with IVR technology instead of computers | 4.13 | 0.64 | 0.41 |

The reason for lower averages on other items was because of some of the technical issues with the Plum voice portal gateway which we did not have any control. Plum's ASR and TTS engines were inconsistent in their performance. For example, their ASR was very sensitive to speech barge-ins and when we tried to adjust the sensitivity of this barge-in the system did not respond to these functions. Another problem was clarity with their TTS personas often produced *slurry* speech and some of the VoiceXML parameters used to speed-up or slow-down the TTS conversation speed the system did not respond to these functions.

Nonetheless, the survey results were very encouraging; especially, if you consider the last question which asked the user whether they were willing to use phones instead of computers to access online information. This question got a score of 4.13. The standard deviation and variance were normal, except we found two users had answered all the questions with *one* value (5 for one user and 1 or 2 for the other). This looked suspicious to us. So, when we removed their responses from the survey analysis, the results looked as follows:

Comparing the average scores with Table 1, you will notice average support for the system was even better. In addition to the questions on the VeDF application, users were asked a few background questions. Table below provides a glimpse on the users who participated in this study. Please note all the scores are not an average; some are just simple counts of user responses.

All users were blind-disabled and one also had mobility impairment. The average duration of their disability was around 25 years with most users being blind for over 12 years. In term of their technological sophistication, users in our sample have used computers for almost 14 years and Internet for 9 years and four of the users have taken online course or course with hybrid (in-class and online) components. In terms of their experience with IVR systems, seven out of eight have used an IVR system before with 3 users having used too many times, 3 users greater than 25 times and 2 users have used IVR less than 25 times. The average number of times the users used the VeDF application before the survey was 12.25. This is satisfactory in terms of exposure of the application to the users to make a critical analysis of our application.

TABLE 3: USER BACKGROUND INFORMATION

| **Background Questions** | **AVG** |
|---|---|
| ***User Disability*** | |
| Disabled (1=Yes, 2=No) | 8/0 |
| Disability (1=Blind, 2=Deaf, 3=Mobility Impaired, 4=other) | 8/0/0/0 |
| Disability Duration (Years) | ***24.38*** |
| Level of Disability (1=Total, 2=Partial) | 8/0 |
| ***User Experience with Computer Technology*** | |
| Computer Usage (Years) | ***13.75*** |
| Internet Usage (Years) | ***8.88*** |
| Online Education Participation (1=Yes, 2=No) | 4/4 |
| ***User Experience with IVR Technology*** | |
| Used IVR Before (1=Yes, 2=No) | 7/1 |
| IVR Usage (1= <+ 25 times; 2= >25 times; 3=too many) | 3/3/2 |
| | |
| Use of VeDF for study | ***12.25*** |

A deeper analysis of the results from our study reveals that even though the users were quite familiar with computer and Internet usage; yet, they were still willing to switch to a telephone-based application. The users preferred using our application instead of other assistive technologies like JAWs and BRAILLE which are also very expensive. Another look at the user background also reveals that although the users have considerable experience with computers and Internet they are also regular users of commercial IVR applications (7 out of 8) – this suggests that they may prefer to use the online forums from both computers and telephone depending on their convenience. Finally, most of users in our sample indicated very little exposure to online education participation; four of the users have never participated in online education course and of the remaining two users had participated only in one hybrid online course; only two users indicated some experience with online education courses. These results support the results of earlier research studies which indicated that online education is, generally speaking, not very user friendly to blind and disabled populations. Although the survey only measures learner satisfaction, we plan to develop other measures to determine the answers for our research questions in the next phase of our project.

The results from this study cannot be generalized to the larger disabled population at this time. The sample size was small and population was self-selected. However, in next phase, we plan to develop and test our application in a diverse setting with a much larger population. This will provide a better external validity for the study and allow us to generalize our results to the larger population. Nonetheless, this study makes a useful contribution to the knowledge of how voice-enabled applications would be received by the blind user population. The informal user feedback was crucial for next phase. We found the application was just as easily usable by technically sophisticated and novice computer users. Also, users with partial and total disability were able to use the system comfortably with minimal training. The results from this study, however, cannot be generalized to the large disabled population with multiple types of disabilities. Our sample size was small and type of disability was limited to make a full-scale commercialization decision yet. More research and development is necessary -- both on the technical side and testing on the human side with a bigger and more diversified sample population before making a final decision. However, results from this phase, users liked being part of an online system where they could share their discussions with other students (disabled and non-disabled). This was a promising development and gave us confidence in proposing Voice-enabled Interactive Service (VòIS) architecture for e-learning environment.

## IV. PROPOSING A VOICE-ENABLED INTERACTIVE SERVICE (VÒIS) FOR E-LEARNING

The goal of a VòIS platform for an e-learning system is to better meet the needs of students and mobile learning communities via speech recognition technology.

Our pilot study explored this goal by demonstrating how an interaction tools like discussion forums can be voice-enabled. However, enabling voice interaction in one online tool is not enough to increase usage of mobile devices. After voice enabling all interaction tools, our project long-term goal is to explore the use of these tools from mobile devices and measure its' impact on learning pedagogies.

The online education programs and e-learning software markets are experimenting with, or currently utilizing, a wide variety of communication and interaction tools ranging from live chats, blogs, wikis, instant messaging, pod-casting and others. One suggestion from our pilot study user group was to have the voice/audio messages searchable from the mobile devices by user or topic, have immediate access to the new messages, and others. These user requirements are embedded in the VòIS architecture to extend this functionality to all communication and interaction learning tools such as blogs, wikis, e-mail, and instant messaging services.

### *A. VòIS Design Specification*

Our design will include a Voice Services Core (VSC) component which will support:

- *A configurable and dynamic Call Flow*. The call flow in our pilot study prototype is currently fixed following standardized login path and menu navigation. However, many commercial applications may have different protocols for login, navigation, user input, etc. This service will allow easy integration with the existing login, navigation and other flows of the application.
- *A logging environment* that will capture all the call flows, configuration services and message logging. Provide audit log capabilities.
- *A data environment* which will provide connectivity via JDBC data sources to a wide variety of databases supporting the communication protocols.
- *An audio searching capability* that allows voice message searching through meta-tags assigned to messages by users. This service will utilize technologies such as Active Speech Recognition (ASR) and pre-parsed extracted meta-data.
- *A voice API* environment that can be easily deployed as a web package by IT staff in organizations to voice-enable their communication and interaction tools on their existing Intranet/Extranet Architectures. This allows our service to be integrations with existing applications with minimal cost and tight integration with any existing in-house applications an end customer may currently utilize.

This VSC can be utilized to develop the following VòIS functionality:

- *Integrated message searching capability*; namely, allow telephone users to search for textual and audio information by a variety of criteria, including user, thread topic, date, content, or other meta data.
- *Branding or content syndication capability* within our service to allow for easy to implement customization of the Voice User Interface, providing end users a transparent method of customizing all visible interfaces to the application.
- *Integration* with specific applications like vBulletin forums, Wikipedia, blogs, and other interaction tools, using add on 'modules' to extend the VSC capabilities.

To accomplish these goals, we will use the Java Enterprise Edition application environment. An inferred requirement of the objectives requires an application environment that can provide a phased and modular environment, while allowing dynamic changes to the run-time environment of the services provided. This modular design also allows for easy plug-in front ends (Voice Architectures, VoIP, TTY, etc) and easy component plug-in at the back ends with practically any database driven web applications without requiring any major coding effort.

Although it is technically feasible to apply voice-enabling technology to web-based interaction tools, this project's broader goal is to determine whether this voice-enabling technology:

1. Will increase the access and usage rate of the learners in e-learning programs?
2. Can be seamlessly integrated with existing software applications on enterprise Architectures with minimal modifications?

### *B. VòIS Usability*

The pilot study provided us with the knowledge and data to extend the application in short and long-term future directions. However, the prototype is tightly integrated to the functionality of a discussion forum. While this allowed us to determine the feasibility of voice-enabling technology and test its usage with the disabled, its impact was limited to users and organizations using these architectures. Therefore, the challenge of VòIS architecture is to create application services and middleware tools that will make our voice Architecture modular and *plug-n-play* software, independent of any interaction tools and voice software. Our goal is to enable voice access via telephones to interactive learning tools, as shown in Figure 3 below.

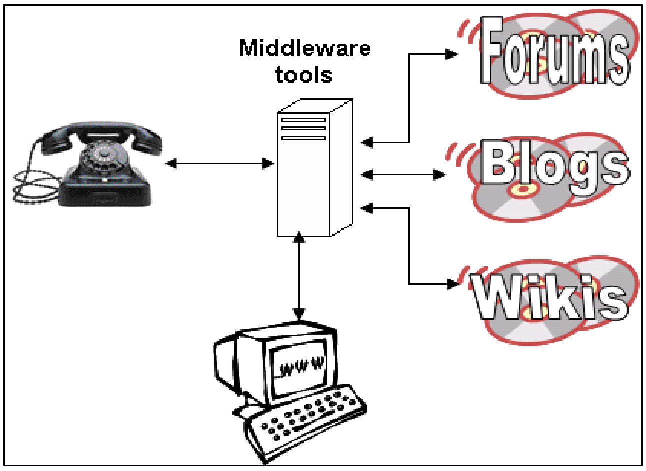


Figure 3: VòIS Middleware Tools

## V. VòIS Architecture

This architecture's goal is to support a *plug-n-play* environment with interactive learning tools and other commercial software. The software selected for integration will be leading software in their areas and work on a variety of operating systems such as Microsoft Windows™, Sun Solaris and Linux™. Universal programs design principles will be followed to integrate these applications. The Nuance Voice Architecture™ supports VoiceXML 2.1 the open, industry standard for speech applications and services developed through broad industry participation in an open forum managed by the W3 Consortium. Because VoiceXML leverages the Web infrastructure, such as application servers and databases, our application will be able to leverage the existing investment in online educational infrastructure such as e-learning servers, pedagogical tools, and discussion forum databases. Most notably, this allows our VòIS Architecture to interface with a wide range of existing systems.

This architecture (shown in *Figure 4*), referred to as 'VòIS' (Voice-enabled Interaction Service), will be separated into three primary components, all of which together, provide the voice interaction services. For the purposes of this document, '*Core*' refers to the basic services and components which provide the base operating capabilities of the system. This includes services such as logging, system management, application flow management, etc. '*Components*' refers to system resources that are used, primarily by services, to provide access to resources that the system requires. Examples of core components include database access, file access, telephony card access, VoIP access, etc.

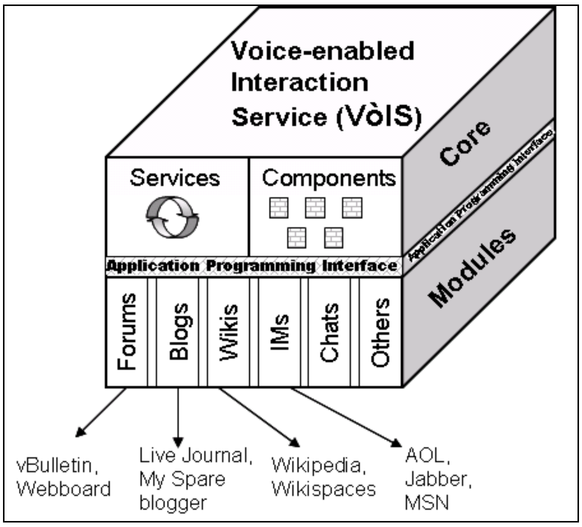


Figure 4: VòIS Architecture

### *A.* Core

1. *Services:* provide the basic functionality of the system. Examples of these core services include
   - Logging
   - Call Flow
   - Configuration
   - Deployment

The services are implemented as tasks that are constantly running within an application server, performing the basic work that is required by the VòIS. These services may operate based on configured parameters that allow them to interoperate with whatever given environment they are current deployed in. Examples of the above services configuration options may include:

- Logging to Database vs. Files.
- Call flow of Voice application vs. TTY application vs. Web Application. Different server may define different call flows based on the way they are configured.
- Services that provide special interfaces to a given J2EE run-time environment, such as connectors for Apache Tomcat, WebSphere, Nuance Voice Architecture, etc.

2. *Components: allow the application environment to interact with services that the server will utilize. These components can provide all access to resources that the system requires, such as:*
   - File Access
   - Database Access

- VoIP (SIP)Access
- PSTN Telephony access
- Web Access
- Messaging Access

Many of these components would simply be capabilities that are provided by the J2EE environment that the voice core services (VSC) will execute. These servers often provide custom capabilities depending on the provider of the environment. Components would allow the server access to these specialized capabilities in an abstract manner, without requiring modification of the VSC. As a simple example, WebSphere Application Server, a J2EE compliant environment, may provide its own file storage methods. A File Access module could provide access to this capability to any of the Services running, without the services being aware that the file is being treated different then if it was simply a file on the local hard drive.

### *B. Application Interface*

While the VSC itself provides everything required functionally, an application programming interface (API) is useful to allow modules to interact with the system as a whole. An API is a set of shared libraries, similar to DLL's in a Windows environment, or shared libraries under a Unix-based system. There are several methods and technologies for this, and we plan to provide to applications several different techniques to provide access to the VòIS. A local library that can be directly loaded by applications running on the same server will allow applications running on the same physical machine to interact with the server. Additionally, a SOAP or Web Services API will allow for applications to run remotely, and access the server by utilizing such web technologies as SOAP and Microsoft.NET.

This API will expose many of the capabilities of the VòIS Components to external applications. For example, a module (documented below), may require database access to a MySQL database. By utilizing the data access portions of the VCS API, these modules could have managed access to a given database, without the need for the module to manage this particular data connection.

### *C. Modules*

Modules are the logical organization of any '*external*' interaction tools software. Modules will translate other applications concepts into those understood by the Voice Interaction Server. These modules may utilize any of the available APIs proposed earlier, as well as potentially utilize the same or similar components as the Core Services.

All of these interactions provide an abstraction of the concept of messaging. During initial research, it was found that nearly all sources of communications that we are targeting use the same or similar concepts when dealing with interactions and messaging. Forum users post messages, and response to messages, and these strings of communications are known as threads. An instant messaging application has a single thread between two individuals. The assumed reasoning behind these similarities is that they are basic concepts that humans have developed for dealing with any electronic medium. Any sort of message is generally a single piece of information. This is sent as one piece to an intended target. Sometimes, these interactions are shared with multiple individuals. Other times, they may be private and personal, from one person to another.

## VI. CONCLUSION

This paper has proposed voice-enabled interactive services architecture for e-learning systems that can be used from mobile phone devices. In general, this project has been a very useful exercise in understanding the role of speech recognition in mobile and e-learning. While the current device limitations reduce the usefulness of these devices, they are useful tools to complement existing computing and Internet environments. Speech-enabled interface can certainly enhance the usage of these devices as they minimize the negative impact of interacting with these small devices.

Voice enabled technology for user interactions is steadily growing. More than two-thirds of phone callers commonly use automated speech in their business interactions. Although early deployments for speech were relatively basic and resided on proprietary interactive voice response (IVR) Architectures, the introduction of new speech software standards has spawned new applications in improving the user interface. The relatively straightforward ROI for speech applications makes them an attractive investment opportunity. The inherent limitations of using the telephone keypad restrict the type of information that is collected because it only supports basic numeric commands. Speech applications provide a much greater range of options, eliminate menu trees, and allow callers to quickly go to their destination or information source. Many early adopters attest to the positive payback from speech applications by progressively adding more speech applications to support multiple aspects of their business operations. According to Harris Interactive report [10], sixty-one percent of customers were satisfied with their most recent speech encounter. This project has provided an application design to take advantage of this consumer sentiment and leverage the speech recognition and mobile networking technologies for electronic learning.

One problem with mobile computing is that most users in the U.S. have access to the Internet via PCs but a much lower percent of the learners in Asia, Africa and Europe have on-demand access to Internet from everywhere and anytime. On the other hand, mobile device penetration with access to data services and Internet is much higher globally than in the US. Therefore a mobile learning has much higher scope in Asian, African, and European countries than in US. This does not mean that US education institutions can ignore this form of learning. Learning, like other industries, is highly global and therefore, many of the e-learning programs in the US have students from countries around the globe. Voice-enablement can increase the access of higher education institutions to students who do not have unlimited access to Internet or PCs and disabled students who are disadvantaged by e-learning pedagogy. The success of VòIS architecture has the potential of increasing the usage of mobile devices in e-learning.

## REFERENCES


[1]. Motiwalla, L. “A Voice-Enabled E-Learning Service VòIS Platform,” The Fifth International Conference on Networking and Services (ICNS), April 2009, Valencia, Spain.

[2]. Pettey Christy, “Worldwide Smartphones Sales Grew 16% per cent in Second Quarter of 2008”, December, 2008. (http://www.gartner.com/it/page.jsp?id=754112 Accessed: Jan., 2008)

[3]. Smith Jodi, “College Students Wave of Digital Connections”, Retrieved via Web, December, 2008 (http://www.harrisinteractive.com/news/newsletters/clientnews/2006_alloy2.pdf Accessed: Nov., 2008)

[4]. Green, K. C., “Technology and instruction: Compelling, competing, and complementary visions for the instructional role of technology in higher education,” November, 2008. (http://www.campuscomputing.net Accessed: Oct., 2008)

[5]. Jones, S. & Madden, M., “The Internet goes to college: How students are living in the future with today’s technology,” Pew Internet & American Life Project. Sept. 2002, (http://www.pewinternet.org/~/media//Files/Reports/2002/PIP_College_Report.pdf.pdf, Accessed: Dec., 2008)

[6]. Quinn, C. Get Ready for M-Learning. Training & Development, Feb. 2001, pp 20-21.

[7]. Wood K. & Homan S, “Taming the Mega-Lecture: Wireless Quizzing,” Campus Technology, Sep. 2003.

[8]. Waits, T. & Laurie Lewis, L., U.S. Department of Education, National Center for Education Statistics. Distance Education at Degree-Granting Postsecondary Institutions: 2000–2001, NCES 2003-017, by Project Officer: Bernard Greene. Washington, DC: 2003.

[9]. Motiwalla, L., “Mobile Learning: A Framework and Evaluation,” Journal of Computers & Education, 2007, v49, pp 581-596.

[10]. Herrell, E., “Evaluating Speech Self-Service Architectures”, *Tech Choices*, August, 2006.

[11]. Mason, R. Models of Online Courses. ALN Magazine, 1998. V-2, I-2, (http://www.aln.org/alnweb/magazine/ Accessed: Jan., 2005)

[12]. Pearson, E & Koppi, T., “Essential Elements in Design and Development of Inclusive Online Courses,” Proceedings of ED-MEDIA Conference, June 2002, p.7.

[13]. Motiwalla, L. & Malik, R., “Exploration of Handheld/Mobile Devices for Distance Learning,” Austin Mobility Roundtable (AMR), an International Conference on Mobile Computing, Austin, Texas, March, 2004.

[14]. Motiwalla, L., “Speech-Enabled Mobile Learning Application,” Proceedings of the IEEE Wireless Technology Symposium, April 2005.

[15]. Abott, K., Voice enabling Web Applications: VoiceXML and Beyond, New York: a!press Publishers, 2002.

[16]. Wang, Y. Assessment of learner satisfaction with asynchronous electronic learning systems. Information & Management, 2001, v41 (1), 75-86.