RI 07001, 01 January 2007 Computer Science

IBM Research Report

VOIGEN : A Technology for Enabling Data Services in Developing Regions

Arun Kumar, Nitendra Rajput, Dipanjan Chakraborty, Sandeep Jindal, Amit Anil Nanavati
IBM Research Division
IBM India Research Lab
Block I, I.I.T. Campus, Hauz Khas
New Delhi - 110016. India.

IBM Research Division
Almaden - Austin - Beijing - Delhi - Haifa - T.J. Watson - Tokyo - Zurich

VOIGEN - A Technology for Enabling Data Services in Developing Regions

Arun Kumar, Nitendra Rajput, Dipanjan Chakraborty,
Sandeep Jindal, Amit Anil Nanavati

IBM India Research Laboratory
Block 1, Indian Institute of Technology
Hauz Khas, New Delhi 110016
{kkarun, rnitendra, cdipanjan, sajindal, namit}@in.ibm.com

ABSTRACT

World Wide Web has made information accessible to computer users in various ways not imagined before. However, there is a huge pool of people, especially in developing countries, who are still untouched by this revolution and are either unaware of or are unable or to join this bandwagon. Strong penetration of mobile phones into these masses as a powerful yet cheap device comes as a breather. It is increasingly empowering the underprivileged to utilize data and services beyond the basic voice communication. However, various factors particular to these masses, present an impedance to enable such access. These include user interface issues, cultural issues, cost factors, illiteracy, infrastructural issues etc. To this effect, we have developed VoiGen – a voice-driven generator of voice-based applications that addresses some of these issues while enabling underprivileged users to derive benefits from data and services accessible to WWW users. VoiGen enables ordinary telephone subscribers to create, deploy and offer customized voice-driven applications through a simple voice-based interface accessible from telephony devices. Results from a user study of our prototype implementation, suggest a tremendous potential of this technology for IT access in developing regions.

1. INTRODUCTION

Information Delivery through WWW

Access to information and services today has been brought as close and made as simple as the click of a mouse or push of a few buttons. The success of World Wide Web in making this possible, however, has been limited to a section of the society. There is a huge pool of population that is still untouched by this revolution and are either unaware of or are unable or to join this bandwagon. These include various illiterate and semi-literate people owning small individual businesses or are workers, either in cities or rural areas, who are unable to afford computers or high end web-enabled handheld devices. As seen in Figure 1, the Internet penetration in developing countries is still below 10%. The low rate of internet penetration in these regions can be attributed to the fact that most common access

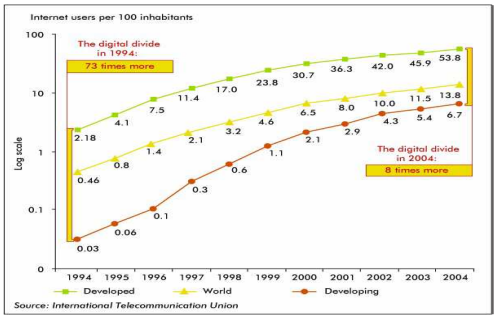

Figure 1: Growth of Internet users (1994–2004).

mechanism to the Internet has been through the PC. Even though the connectivity costs have reduced, a PC still costs about Rs 10000 (USD 220). This is very costly for people in developing regions (56% people in developing countries live below USD 700 per year [2]). Furthermore, using a PC requires IT skills beyond language reading and writing, leading to a low acceptance rate. Due to these limiting factors, the benefits of the WWW have not been able to impact a majority of population in developing regions. The use of WWW to such sections has been identified over the last decade and efforts have been made to reduce this digital divide. Projects such as eChoupal [13], eSAGU[1] are steps taken in this direction. However these solutions aim to increase web access through low cost shared personal computers and shared kiosks. These solutions involve intermediaries (such as human agents) that connect the end user to the web, thus reducing the skill requirement for using a PC. The use of intermediaries however, prevents these users from a *direct* access to data and services. This does not let these users explore the various services available to them and the benefits they can obtain from the IT infrastructure. In addition, there are infrastructural issues such as non-availability of electricity for long hours. It would be beneficial to bring access to information services as close to the user as possible and this needs a change in the current delivery model. Other researchers have also

[1]http://www.esagu.in

proposed delivery models [16] alternative to current WWW which might be more suited for developing regions.

Mobile Phones as Alternate Delivery Model

The telecom world is undergoing a paradigm shift with increased convergence of voice and data communications. It is slowly moving towards adoption of Internet Protocol based networks with more and more data services being deployed and planned. While the Internet users are seeing increased integration with the telecom world in the form of services such as Click to Dial, Video Conferencing, VoiceOverIP etc., the reverse integration is minimal. Only a few applications such as web access over phone [20], email over phone have allowed traditional telecom users to access information and services in the Internet world.

Mobile phones have provided a ray of hope as advancements in technology have transformed them into powerful yet cheap communication devices. This has lead to an explosive growth in their penetration into masses, primarily as a voice-based communication tool. The mobile penetration in developing countries is far more than the Internet penetration.

Mobile phones have the potential to empower the underprivileged to utilize services beyond the basic voice communication. However, various factors present an impedance to making that happen. There are several user-interface issues that need to be addressed while designing services that can be accessed through a mobile device. A survey done by the authors (across 6 cities in India – involving mobile phone users that earn below Rs. 40000, i.e. USD 1000, in a year) revealed that these users use the mobile phone only to make phone calls. They are not able to use any other feature on their mobile phone, such as address-book, reminders and short message service (SMS). We interpret this as a problem with the learning-curve that is required for such text-based menu-driven applications. Even the low cost of SMS is not attractive enough to ride this learning-curve. On the other hand it has been observed that voice-driven interfaces not only attract illiterate users but are able to build a rapport with semi-literate users as well [17].

Interactive Voice Response Systems

Interactive Voice Response (IVR) systems have played the role of an enabler in delivering services such as airline/rail reservation, tele-banking etc, through voice channel (phone). Currently, these systems are primarily used by IT-savvy users as the typical applications offered through the IVRs are not the ones that are needed by the underprivileged. Secondly, current deployment of IVRs is primarily enabled by governmental or non-governmental organizations with investments of infrastructure and programming effort. Some examples include systems for agricultural inquiry [19], electronic mail delivery, and even banking through voice driven ATMs [11].

Taking a look at the IT users in the Internet world, they currently host and offer their own services such as online stores, blogs, homepages etc. without actually having to own the entire infrastructure. In contrast, users having voice-based phone as the only medium to *access* technology, are deprived of the such benefits that they can derive from the available infrastructure. Even though voice driven interfaces can currently be utilized as access mechanisms, they do not empower the end user to *create and offer her own voice based applications*. Moreover, existing efforts for users in developing regions are aimed at utilizing voice interfaces for delivery of applications and information access. This forces the user to remain a passive consumer rather than becoming an active participant in development and deployment of voice based services.

In this paper, we approach the problem from a different perspective and enable individual phone subscribers to create and offer their own customized voice driven data services. At the core of the proposed system is a voice driven user interface generator coupled with a voice driven service composition module. These along with a supporting infrastructure allow voice-driven applications to be composed from existing components and deployed on a hosting platform so that they become available to other subscribers. By virtue of these users having a voice-driven interface, they can expose their services through all telephony devices. Further, hosting in the network brings two-fold benefits. First, it enables the subscribers to pay per use rather than investing in a huge upfront cost. Second, it helps them connect to other applications thus enabling creation of a world wide ecosystem of telecom web [8].

The contributions of the paper are summarized as follows:

- Propose VoiGen - a Voice-driven generator of voice-based applications.
- Enable a mechanism to enable individual phone subscribers to create, host and deploy customized voice driven services.
- Enable access to data and services residing in the IT infrastructure to masses in developing regions, through VoiGen.
- Present user study of a sample population on the acceptability and use of this technology.

The rest of the paper is organized as follows. Section 2 provides the motivation for VoiGen. Section 3 presents VoiGen system architecture and explains the hosting and the deployment model of these voice-sites, with a motivation through several potential applications. The implementation of VoiGen is explained in Section 4 and the system evaluation through user-study is presented in Section 5.

2. MOTIVATION

Employing Voice to penetrate into developing regions

Voice-driven telephony applications have been increasingly used across the world to provide services to the end-users. With the developments in the underlying speech recognition [12] and speech synthesis [18] technologies over the past decade, use of voice has emerged as a much pervasive and easier alternative to accessing services. In developing countries, voice-driven applications are even more attractive means of accessing services and information. This attraction is partially due to the easier and more human-like interface that voice applications can provide. Further, the lower cost of accessing such services in a telephony infrastructure is another attraction. High volume and friendly government policies have enabled a low call rate in developing regions. India enjoys the lowest call rates in the world at an average price of 2 cents per minute, compared to average prices of 33 cents in Japan, 11 cents in Brazil, and 24 cents in Australia [4]. China has comparable call rates of 2 cents per minute though incoming calls are charged too.To justify the ease-of-use of voice-driven applications, we performed a preliminary survey. Our

surveys of low-end mobile phone users in tier-1 and second tier-2 cities of India revealed that 42 of the 50 participants surveyed had not used any feature on their mobile handsets other than making voice calls. This includes *zero-cost* handset features such as address-book, reminders and also the *low-cost* paid services such as SMS. This reflects on their low levels of comfort with GUI/text-driven interfaces that are available on current low-cost handsets.

Literature also supports that voice-driven interfaces are effective in providing information to illiterate and semi-literate population [5, 19]. Authors in [5] have observed that variation in dialects of spoken languages, presence of multiple languages and lack of linguistic resources, among other factors lead to difficulties in employing speech recognition for illiterate people. However, they also report that despite speech technology problems, users with little education could navigate through a dialog system with very little training.

It has also been observed that spoken language based interaction not only enables illiterate users but even literate users feel more comfortable with audio feedback in local language [17].

VoiGen *for Enabling Tele-Online Personification*

As we envision it, VoiGen can open up a plethora of opportunities for users in developing regions. We start with the notion of a tele-online identity. Through our surveys, we realized that small businessmen such as street vendors are willing to be more reachable but do not wish to actively reach out to their customers in order to keep their operating costs low. For instance, a fruit-seller in New Delhi whom we interviewed sells fruits and vegetables through his makeshift setup located at the junction of two streets in a residential colony. Residents in nearby multi-storey apartments and row houses prefer to buy from him as compared to a nearby grocery store simply due to convenience of doorstep availability. As some of the items on sale begin to rot, he drops his prices and would prefer to inform of this drop to his regular customers which is not feasible currently unless he chooses to call or text message his customers. Having the ability to upload this information which could then be notified to his customers appealed very much to him.

Similar situation occurs in scenarios where some service providers such as plumbers, electricians, carpenters, etc. are highly mobile. Most of them carry a mobile phone, since this is the only way to reach them. Most often, they end up spending time in attending clients resulting in loss of valuable work time. During festive seasons and other peak times due to increased query and follow-up calls these people become virtually unreachable since their phones are either perpetually busy or switched off to perform their work. This results into loss of business.

This leads us to believe that a major step needed for users to be initiated into realizing the benefits of IT is to enable their tele-online personification. As an analogy in the World Wide Web, users either have an id (such as an email-id) through which they are identified or a homepage where they can be reached. Similarly, a custom *VoiceSite* created through VoiGen enables individual phone subscribers to have a tele-online persona. As mentioned earlier, a *VoiceSite* is equivalent to a WWW homepage that can be browsed through a voice based interaction via a Voice Browser. In simple form *VoiceSites* could be used to advertise one's business by creating menu options that describe what is on offer such as "*Welcome! My name is Shyam and I am a plumber. Press 1 to know about my services, 2 for charges, 3 for making an appointment.*" Calls made to Shyam are diverted to his *VoiceSite* and is connected to the actual instrument of the subscriber only if needed. In more advanced scenarios, these *VoiceSites* could take the shape of tele-online stores, tele-blogs, information delivery systems, tele-online catalogs etc.

3. VOIGEN: A VOICE DRIVEN GENERATOR OF VOICE BASED APPLICATIONS

VoiGen simplifies the process of creation of voice-based applications. It enables creation of voice-based applications through a voice-driven interaction. It has two intertwined components – a user interface generator and an application composition system. A phone subscriber could call in to VoiGen and can compose an application by navigating through the custom options offered to her. This application is then deployed in the form of a *VoiceSite*, which is a VoiceXML representation of the created application. VoiGen makes use of existing components (reusable dialogs as well as IT components such as databases, web services etc.) to compose custom applications. A key aspect is that the generated application can be hosted in the network and for the subscriber it virtually resides on the phone.

3.1 System Architecture

Figure 2 depicts the overall architecture of VoiGen system consisting of a core VoiGen component along with a *VoiceSite* hosting engine. The VoiGen core component consists of the following:

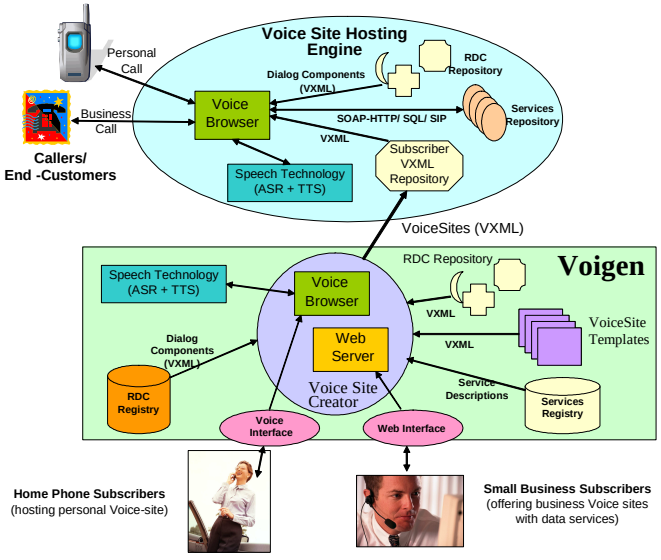

Figure 2: VoiGen **System Architecture**

- **VoiceSite Creator:** This is the central component of VoiGen as it serves the requests from subscribers. It presents a voice-based (and a web based) interface to the subscribers through which they can create new *VoiceSites* and edit previously created ones. The Voice Browser presents the voice interface to subscribers

consisting of a sequence of voice prompts based upon the profile for the site specified by the subscriber. It interacts with a Speech Technologies server that provides speech recognition and synthesis technologies required to capture and render voice output/input during the conversation with the subscriber. The Web server on the other hand presents a web based graphical user interface for creating the *VoiceSite*.

Once the *VoiceSite* has been created it gets deployed onto a *VoiceSite* hosting engine to make it accessible to end-callers. Small businessmen and individual users would typically use the voice interface for creating their *VoiceSites*. Web based interface presents an alternative by enabling creation of *VoiceSite* through a GUI for situations where the *VoiceSite* becomes too complex to be edited through the voice interface.

- **Voice Templates:** VoiGen consists of predefined Voice Application templates suited to different user profiles such as plumbers, electricians, home users, shop owners etc., which can be used by the *VoiceSite* Creator to help subscribers create their own *VoiceSite*. The actual content of the *VoiceSite* is governed by the subscriber. The Voice Templates may also retrieve relevant information from appropriate databases to populate their menu options. For instance, a plumber *VoiceSite* template would retrieve the list of plumbing services (from a preconfigured database) so that the subscriber could select the services specific to her.

- **Reusable Dialog Components (RDCs):** These [3] are predefined Voice User Interface modules that can be used as ready-made dialog components to develop voice applications. RDCs encapsulate prompts, grammars and dialog strategies for a lot of common UI components. Thus generation of *VoiceSites* will not require creation of new grammars and authoring of new dialog strategies. Using RDCs help in building complex voice-based applications without much skill and effort. The Voice Templates may make use of these RDCs and when needed the desired RDC component can be pulled out from the RDC repository. Appropriate RDC references might also need to be inserted in the newly composed *VoiceSites*. For that purpose a registry maintaining a list of all such RDC's is also available.

- **Services Registry:** The *VoiceSite* being composed may also need to utilize other services such as some specific database, or a web service such as for calendar or appointments, or even a SIP-enabled [10] application. The Services Registry maintains a list of all services and their specification (signatures with protocols used to communicate with them) so that the *VoiceSite* Creator could include appropriate references into the new *VoiceSite* description.

VoiceSite Hosting Engine

VoiceSites created get deployed in a hosting engine. The deployment process essentially consists of assigning a telephone number to the *VoiceSite*. This is similar to a URL for a webpage and providing the server that hosts these pages and interacts with the Voice Browser. The design of the *VoiceSite* Hosting engine is similar to that of the VoiGen core component itself since both are essentially Interactive Voice Response systems. The *VoiceSite* hosting engine contains a subscriber *VoiceSite* repository (as VXML files) instead of *VoiceSite* templates and it does not have an RDC or a service registry. However, it has an RDC repository and a service repository which contain RDC and services respectively that can be used while executing a *VoiceSite*. The calls arriving at the *VoiceSite* hosting engine are typically from end-callers trying to reach the subscribers as against subscriber calls to VoiGen core component. The voice call terminates at the hosting engine rather than the actual dialed phone. The call can be forwarded to the subscribers actual phone based upon the interaction of the caller with the subscribers *VoiceSite*. The engine runs on traditional IP networks.

3.2 Delivery of IT Services through VOIGEN

IT services and applications accessible to *VoiceSites* are (1) local applications hosted in the *VoiceSite* hosting engine (e.g. local database application), (2) Remote (web-enabled) applications exposed through Web Service [9] based interfaces, (3) SIP-enabled [10] applications accessible through converged networks.

Local applications are accessed by conventional protocols over TCP/IP. For example, a database application storing personal requests received by the *VoiceSite* can be connected to it using traditional programming constructs available (e.g. java database connectivity). The integration of *VoiceSites* with backend web service based applications is achieved by integrating web service clients with VoiGen (and the deployed *VoiceSites*). Accessing web services is driven by standards like WSDL [9] and SOAP [1]. Implementations of both these standards are available for IP-based applications. A huge number of web-based applications are exposing web service interfaces for them to be invoked by other applications programmatically, following the SOAP/WSDL standard.

Enabling *VoiceSites* to access SIP-enabled applications happen through the newly emerging IP Multimedia Subsystem [6], that enables the convergence between services and applications available on IP networks with traditional telephony networks. Session Initiation Protocol (SIP) [10] is a critical protocol that allows the convergence of IT applications with telephony services. Figure 3 describes the integration of VoiGen with the SIP-based IMS [6] converged network architecture. This enables these users to access backend data and services existing on the IP network. Please note that the figure does not depict a network stack, rather shows a logical distinction between the various components of IMS, Web(IP) applications and logical positioning of VoiGen in the infrastructure.

To demonstrate the delivery of IT services through VoiGen, consider a semi-literate plumber Sam, creating his *VoiceSite* through VoiGen. VoiGen offers him the ability to store his availability hours and also receive appointments for the next working day. VoiGen integrates the *VoiceSite* with a *web-based* application capable of receiving orders for small businesses in a locality. Through *SIP-enablement*, the web-based application becomes aware of the plumber's location and availability hours. SIP also allows the *VoiceSite* to receive notifications when new appointments come in (both from IT users accessing the web portal as well as through voice calls to the *VoiceSite*) and stores the orders in a

local database application. Sam periodically calls his/her *VoiceSite* to check for outstanding orders.

We briefly attempt to explain how *VoiceSites* are architecturally integrated with the SIP infrastructure. Extensions to the SIP [10] protocol offers a uniform signaling substrate in the IMS plane to deliver rich dynamic information of end users to backend applications. The integration happens in two stages.

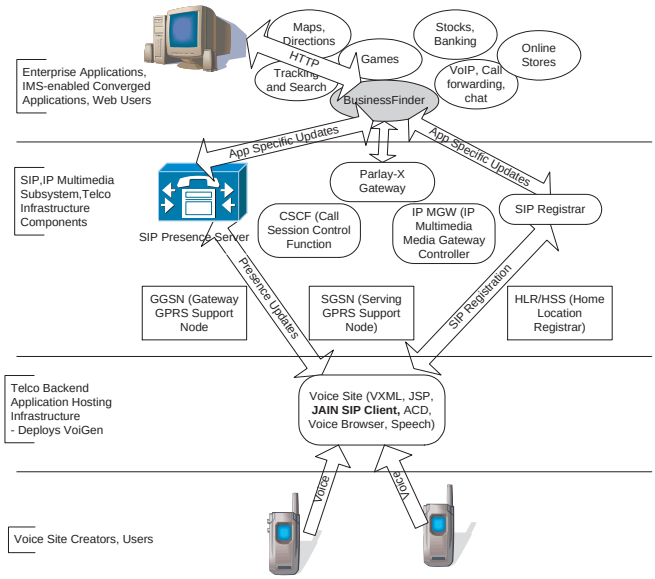

Figure 3: Architectural Integration of VoiGen with WWW through IMS

1. When VoiGen creates a new *VoiceSite*, it uses speech-to-text conversion to extract out relevant information of the user (e.g. phone number, location, availability) and *publishes* [14] it to the SIP Presence Server. Thereafter, enterprise and web applications enabled with support to converse with the SIP infrastructure, access information about these users from the Presence Server (by *subscribing* to the information).

2. Users accessing the web-based service place appointments using traditional web-based portals. These appointments come back through the SIP Presence Server as *notifications* to the local database application integrated with the *VoiceSite*. Traditional telephony users can also call the *VoiceSite* and place appointment requests. VoiGen enables this by allowing the callers to specify their appointment hours and storing them in the backend database application after checking for conflicts.

3.3 Discussion

The strength of VoiGen comes from the fact that a simple voice interface can be used to generate an entire voice-based application that consists of several interactions with the caller. The key characteristics include:

- **Usability:** It is a basic requirement considering that the target population consists of illiterate, semi-literate users. Initially, we implemented a generic voice application generator with the aim of catering to a wide variety of users while giving them maximum flexibility. That implementation allowed any arbitrary voice interface to be created. We explored a depth first approach as well as the breadth first approach for creating the voice interaction tree. However, it soon became clear that without having a guided interaction, the user is left clueless with respect to the purpose, capability and assumptions of the system. Even if an educated user tried to create an application, it became difficult to retain the context of different options specified at different levels since there is no visual interface to view the entire application. This is similar to remembering all the streets and connections in a neighborhood as compared to looking at a map. Other researchers have made similar observations in the context of building a speech based personal information management system [22].

  It has been suggested [16] that a temporally laid out interface consisting of a sequential presentation of tasks works well with our target population. In addition, a spoken dialog based input and output and/or numeric keypad based input increases accessibility and satisfaction of these people. VoiGen's interface for creating voice-based applications is a voice-based application itself. The lessons learnt from the generic system led to the use of predefined templates for different user profiles. These templates are populated with subscriber specific content obtained during *VoiceSite* creation session.

- **Customizability:** Providing the ability to customize user creations results in successful applications since it supports creativity and personalization. The *VoiceSite* creation process of VoiGen neither forces the user to specify all the options of the template nor does it restrict the users to preconfigured menu options alone. The generic *VoiceSite* creator comes in handy and allows addition of custom menu options to existing templates. Customizations are however governed by how generically we can specify access to databases and services. In the simplest case these could simple prompts.

- **Economic Viability:** From the point of view of end-users, VoiGen comes with the ability of getting hosted by a service provider. Ideally, this could be the local telecom operator since all the calls come to its infrastructure anyway and it would have other services (such as databases, Web services, SIP infrastructure etc.) that a VoiGen deployment could make use of. The telco could charge a subscription fee for hosting the basic Voice applications and for heavy traffic applications it could follow a usage based model.

  From the point of view of Telecom operator, apart from the hosting fee there are other channels that can add to its revenue growth. It could introduce a service for supporting multiple incoming calls against a single number since all of them terminate in the subscriber's *VoiceSite* and can remain active simultaneously. These *VoiceSites* can act as a gateway to many data services as described in later sections, leading to a further increase in traffic and increased usage of data services offered by the operator and others.

3.4 Applications of VoiGen

Since VoiGen allows creation of *VoiceSites* by individual phone users, it can be expected that a large number of *VoiceSites* will be created over time. Interconnections within these *VoiceSites* will lead to a further advantage in connecting small businesses. VoiGen thus enables a vision of a World Wide Telecom Web (WWTW) which consists of interconnected *VoiceSites* that provide a multitude of services to the population that does not have access to the Internet. With increasing growth in mobile penetration in developing regions, small businesses and their customers are likely to benefit tremendously through the enablement of this WWTW. WWTW can be promising to people in the developing regions to (1) Expose themselves to the IT world (2) Get access to several IT services and applications (that are available through WWW) (3) Enable WWW to leverage WWTW to deliver interesting applications (4) Enable an ecosystem that allows both non-IT savvy as well as IT users to access information and services exposed by the masses. Though additional technologies (such as search, links) will have to be defined for WWTW, we envision this vision in a separate paper [8]. We validate the potential of WWTW through user surveys later in this paper. In this section, we explain our vision of the World Wide Telecom Web (WWTW) that is realized through VoiGen, describing some key properties, inter-connections with the World Wide Web, and applications that can be hosted over WWTW. WWTW, exploits the concept of online presence and enables a huge mass of people in the developing regions to (1) Expose themselves to the IT world (2) Get access to several IT services and applications (that are available through WWW) (3) Enable WWW to leverage WWTW to deliver interesting applications (4) Enable an ecosystem that allows both non-IT savvy as well as IT users to access information and services exposed by the masses.

4. IMPLEMENTATION

In this section, we provide detailed description of implementation of VoiGen system, followed by how VoiGen is integrated with IT services through the IMS telecom infrastructure.

We have designed, implemented and deployed a prototype of VoiGen at IBM India Research Lab. The system accepts voice calls through regular PSTN connections and guides the caller in creating his/her own voicesite. VoiGen on receiving a call, carries out the following activities:

- It educates the caller with the functionality offered by VoiGen. It uses custom recorded prompts (spoken in the local language here - Hindi) to describe the caller (callers are mostly semi-literate or illiterate people knowing the local language here) about the content s/he can put in his/her voicesite.
- VoiGen prompts the caller to specify his/her preferences and records them. We employ a combination of VXML, java and java scripts to receive the preferences and determine the control flow of VoiGen.
- We first allow the caller to provide basic information about him (similar to index pages on the web). Subsequently, VoiGen provides the caller with the ability to create links to other voicesites that the user might want to connect himself to. The destination voicesite can represent several entities like individual people, micro businessmen, local shops to large enterprise driven IVR systems. This is an important for enabling World Web Telecom Web [8]. VoiGen also allows the caller to specify detailed information about his/her business (business location, service charges, appointment schedule, business hours etc). This is similar to constructing a detailed "Profession" page that is connected to the index page.
- VoiGen, on receiving all the inputs, parses through the data obtained, and automatically generates a voice page (primarily VXML) for the caller. Depending on the caller's preferences, the automatic generation either employs pure VXML or a combination of VXML and JSP (e.g. JSPs are employed if the VXML needs to integrate with a calendar service).

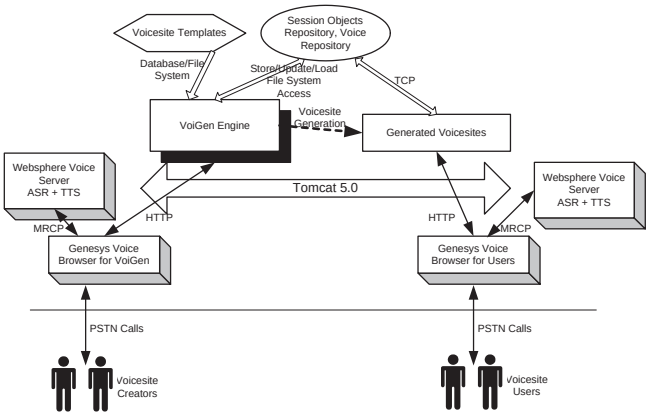

Figure 4: VoiGen Prototype Components

Figure 4 shows the various components involved in the creation of VoiGen. VoiGen is deployed as an application on Tomcat 5.0 and connected to a PSTN phone through the Genesys voice browser. Genesys utilizes Websphere Voice Server to enable speech recognition and text-to-speech conversions whenever necessary.

4.1 VoiceSite Generation

The VoiGen Engine loads customized voicesite templates depending on user preferences (e.g. a businessman wants to generate a voicesite for his shop) and uses Genesys to talk to the caller. VoiGen engine keeps track of entries being made (preferences as well as recorded messages) through java objects in a backend repository. When the caller has finished customizing the template (filling out all the inputs required), VoiGen engine employs a template-based parser to parse through the user inputs and automatically generate a set of VXML and JSP files. These files are then deployed at the Genesys browser that is connected to the caller's personal phone.

VoiGen keeps track of ongoing voicesite creation by keeping a persistent session of the ongoing call. The session variables (java object) also keeps track of preferences being saved, while the audio messages being spoken by the caller are saved in the backend repository.

We have also implemented a generic voicesite generation engine, based on a generic tree-based template. This engine

essentially allows a user to create arbitrary number of voice pages, with arbitrary nesting. The principle employed to generate voicesites are the same as that used by VoiGen (VXML and JSPs for gathering user requirements, internal data structure and repository for storing information, the nesting structure, automatic generation of VXML/JSPs using the information stored). Each node in the generalized tree has an audio prompt recorded and the set of choices and options entered by the user. The options enable links to other nodes in the tree. We implemented a depth-first-algorithm to enable users store their voice pages and define links in each node and associating enumerated variables with each node to distinguish between nodes that are empty, leaf and non-leaf nodes. Apart from generating the VXML/JSP files, we also generate grammar files that are used by the Voice Server to parse the inputs when users browse through the voicesite. When a user browses the voicesite generated by the generic *VoiceSite* generation engine, each node plays out the audio prompt (content of that page), lists the available options to the user (available links) and moves to the next document (or previous document of exits) depending on the options given by the user.

As pointed out in section 3, we realized that human capacity of keeping track of all branches is very difficult, especially when the browsing/creation mechanism is voice. Hence, for VoiGen, we resorted to template-based creation of voicesites. Here, the user is provided with the template that essentially guides him through the options (links) available for him to fill up once s/he finishes creating a particular page. We are in the process of integrating the generic voicesite generation engine with VoiGen, so as to allow advanced users with the ability to provide customized browsing options (instead of template driven options).

Figure 5 shows the control flow diagram of a concrete template that we used in VoiGen to enable small and micro businesses (plumbers, electricians, servants, home delivery services) in Indian metropolitan cities to create their *VoiceSites*. This voice template allows these classes of

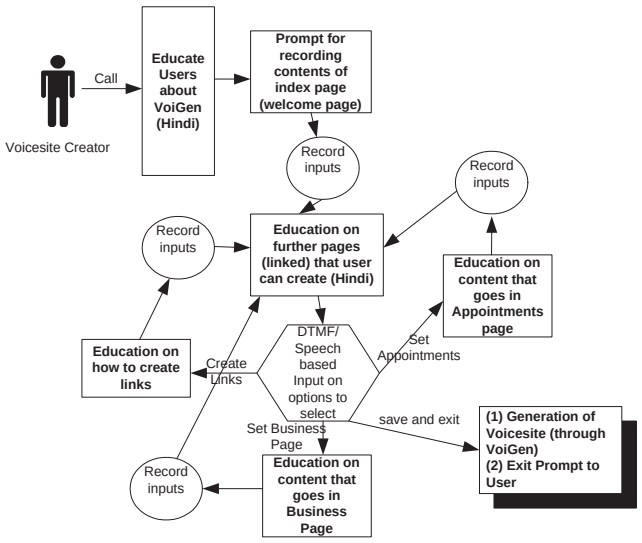

Figure 5: Control Flow diagram of a voice template used in VoiGen for survey

users to create their welcome page, create their business page where they provide information about their business, and also offers the ability to create a page to specify appointment hours. Apart from these, it allows these users to create links to other users (1) they might want to connect to; (2)they use as references for their business. Figure 6 shows a snippet of a sample VXML/JSP voicesite that was generated by VoiGen during our survey.

```
<!DOCTYPE vxml PUBLIC "-//W3C//DTD VOICEXML 2.0//EN" "vxml20-1115.dtd">
....
<vxml version="2.0" xml:lang="en-US" xmlns="http://www.w3.org/2001/vxml" application="rootApp.vxml">
<!-- Initial Welcome prompt. This would be generated by a generic statementif the plumber doesn't specify any
welcome prompt -->
<form id="main">
  <block>
  <prompt bargein="true">
  <audio src="<%=request.getContextPath()%>/1123456789/WelcomeMessage.vox" /></prompt>
  <goto next="<%=request.getContextPath()%>/1123456789.jsp#serviceOptions"/></block>
</form>
<form id="serviceOptions">
<field name="choiceOption">
  <!-- Enter customized prompts by considering preferences selected by user
during voice page generation -->
  <prompt>
    <audio src="audio/final_ServicePrompt.vox"/>
  </prompt>
  <help>
    <audio src="audio/final_ServicePrompt.vox"/></help>
  <grammar type="application/srgs+xml" src="grammar/servicePromptVoice1.grxml"/>
  <filled>
    <if cond="choiceOption &lt; 11">
      <prompt>
        <audio src="<%=request.getContextPath()%>/1123456789/ReferenceListMessage.vox" />
      </prompt>
            <goto next="<%=request.getContextPath()%>/1123456789.jsp#serviceOptions"/>
    </if>
    <if cond="choiceOption &lt; 21">
    ........
    </if>
    <if cond="choiceOption &lt; 41"><prompt>Thank you for accessing my I V R. </prompt>
      <exit/>
    </if>
.....
</vxml>
```

Figure 6: Sample VXML/JSP snippet generated by a survey participant through VoiGen

To demonstrate the integration of voicesites with backend IT services, we have augmented an advanced web-based presence-information driven matchmaking service called BusinessFinder [7] developed at IBM India Research Lab, and integrated it with the SIP Infrastructure. BusinessFinder essentially enables a web-based search (and matchmaking) solution to discover both static (shops, restaurants) and mobile businesses (plumbers, electricians etc) in a person's vicinity. It collects dynamic location and availability information available in the various IMS components. The use case is very pertinent to the decentralized marketplace (markets with several shops owned by individual businesses) observed in developing regions of several countries (e.g. India). In our current prototype, plumbers, electricians are capable of providing SMS-based updates of their work schedules (which is used to determine availability). Through VoiGen, these businesses simply need to speak out their preferences (while creating their voicesites) and update them as and when necessary. These updates are transmitted to the Presence Server, which come as *notifications* to BusinessFinder [2]. A web-based user simply needs to open the BusinessFinder portal to check for available businesses nearby. We are in the process of composing BusinessFinder with VoiGen to offer it as a service that can be connected to voicesites of interested voicesite creators (typically plumbers, electricians and other mobile micro businesses). We are also carrying out surveys to determine the best set of

[2]Information sharing through *VoiceSites* poses similar privacy threats that are there in the WWW. While we are aware of this, our attempt in this paper is to show the utility of VoiGen, and we plan to handle privacy issues as a part of our future work.

candidate web-based services that can be offered to WWTW users.

5. SYSTEM EVALUATION

Since the VoiGen system is intended to be used by the masses, we invited people such as carpenters, electricians, plumbers (*subscribers*) to use VoiGen. They created their own *VoiceSites*. We did a user study to find out the comfort level of them using the VoiGen system. Secondly, we invited people who are typical *users* of such services and asked them to call the *VoiceSites* that were generated by the *subscribers*. We further did a user study to evaluate the usability of the *VoiceSites* that are generated by VoiGen. Thus we evaluated the entire process of generation of *VoiceSites* (by *subscribers*) and then the use of these *VoiceSites* (by *users*). In this section, we elaborate on the profile of the subjects who took part in the evaluations. Later we describe the evaluation process and conclude by providing the evaluation results and insights gained from user study.

5.1 Profile of survey subjects

The system was evaluated on two processes. The first process focused on the usability aspects of VoiGen, which is used to create *VoiceSites*. The target population chosen for this task are the people who work as freelancers and have a specific region of operation (a few kilometers). They are typically skilled laborers (such as electricians, plumbers, carpenters) and charge on the basis of the amount of work that is required of them. This profile is typically observed in developing countries where a decentralized and a disconnected set of labors work at an individual level. Most of them have had only about five to ten years of formal school education. Their yearly earnings are below Rs. 75000 (USD 1600). They get business when households call them for work to fix things in their home/office. Their advertisement largely depends on the social network and is based on the word-of-mouth. Despite their low income, most of them carry a cell-phone since it helps in their business to be contactable. They do not have a phone at their residence. We surveyed 12 subjects of which 3 were carpenters, 5 were plumbers, 3 were electricians and one was a drilling person.

The second process involves use of the *VoiceSites* that are generated by VoiGen in the first process. These set of people, who we foresee would use these *VoiceSites*, are from a well-to-do middle-class family who need services to fix water taps, electricity problems, among other work items. This section of the population is relatively more exposed to IT services. Some of them have used IVRs and know about the Internet.

5.2 Survey Process

For the first process, we briefed the *subscribers* for about 10 minutes to motivate them of the use and advantages of VoiGen in their daily lives. Then we briefed them with the usage of VoiGen for about 5 minutes. Then we asked subjects to make a telephone call to the VoiGen. Subjects were asked to interact with VoiGen and respond to the commands and provide the relevant answers. Finally, we asked the subjects a set of questions to get an understanding of the usability of VoiGen and of the potential of VoiGen for the masses. The questions asked were the following:

- Have you ever uses an IVR?
- Was this method (VoiGen) of generating the *VoiceSites* easy to use?
- Are you interested in having your own *VoiceSites*?
- Do you think your business will improve with the use of your *VoiceSites*?

In addition, the following observations were made during their use of VoiGen:

- Time required to describe the concept of VoiGen.
- Time spend by the *subscriber* in creating the *VoiceSites* through VoiGen.

For the second process, we explained the concept of providing services through telephony infrastructure to the prospective *users* of *VoiceSites*. Then we asked them to make a phone call to the *VoiceSites*[3] that were generated as part of the first process. We asked the following questions to them after the phone call:

- Was this IVR informative?
- Do you normally have problems in getting touch with the plumber, electricians or such service providers?

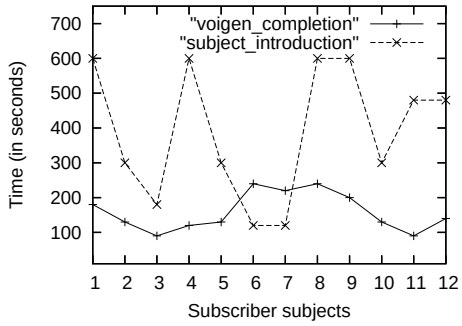

Figure 7: Time taken by each subscribers to understand VoiGen concept and to build their *VoiceSites*.

5.3 Survey results

Out of the 12 subjects that were used to generate their *VoiceSites*, 10 were able to successfully create the *VoiceSites*. Figure 7 shows the time that was spent to build the *VoiceSites* for the 12 subjects. As seen, most subjects were able to generate their own *VoiceSites* within 4 minutes. A 4 minute phone call in India costs less than 5 rupees (10 cents). This is despite the fact that none of these subjects knew or had used an IVR before. Surprisingly most of them were comfortable in using the IVR. More importantly, all of them were able to identify the potential that a *VoiceSite* can have in increasing their business. However each such phone call had to be preceded by an introduction to the concept which took about 5 minutes for each subject, as is seen in Figure 7. All showed tremendous interest in the concept of *VoiceSites* and the fact that their work can be advertised without them actually requiring to purchase any additional equipment.

About the usability of VoiGen, 2 subjects felt that the IVR menu was not designed properly. They highlighted that

[3]A sample *VoiceSite* can be accessed by making a phone call to 91-11-41654558

Figure 8: Evaluation in market - in high ambient noise

Figure 9: Evaluation shifted to inside a car

it asks difficult questions rather than just asking for simple things such as name,occupation or area-of-work. Initially we did these surveys in a open environment in the market as seen in Figure 8. However the speech recognizer had problems with the ambient noise and so we had to shift the survey to inside the car to avoid speech recognition errors (Figure 9).

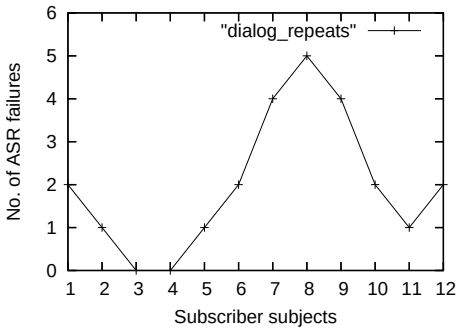

Figure 10: Number of dialog-repeats due to speech recognition errors.

The speech recognition accuracy was about 70%. Figure 10 shows that most speakers had to repeat the commands atleast once since the speech recogniser was not able to understand the user voice.

Overall, the concept of providing services over telephony infrastructure was found to be very important to the subscribers and they realized its potential. The concept of VoiGen was also found to be very easy to learn and use. Based on this survey study, we realize that since the access to *VoiceSites* is through a telephone channel, its reachability gains tremendously. This is especially important for developing regions where there is a huge mass of population that is neither computer savvy and nor can they afford such expensive machines. Secondly, since voice interface is used to access the services, the learning required to use such services deployed on the telephony infrastructure is minimal. This is critical since our target population is not expected to be educated enough to learn the complex and non-natural interfaces such as keyboards. Finally, the model of *VoiceSites* is such that it has minimal cost implications to the *subscribers* and *users*, this is extremely attractive option to use the benefits that have been so far limited to the WWW. The system evaluation through these user studies act as a proof point to the potential of VoiGen and *VoiceSites* in developing countries[4].

6. RELATED WORK

Automatic generation of user interfaces from a given specification has also been addressed in [15]. Their goal is to produce a universal controller that can act as a remote control to any appliance. The approach followed is to separate the interface of an appliance from the appliance itself. They focus on automatic generation of the user interface and have outlined various requirements that are needed for such systems. Our approach instead is to enable a voice-driven process for generation of voice-based interfaces with the end goal of enabling voice applications that deliver data services to underprivileged users.

The CAM system [17] also takes a non-PC centric approach and suggests a change in the current service delivery model [16]. It does so by leveraging the capabilities and reach of mobile phones in the rural regions of developing world. The focus is to enable custom development and deployment of mobile applications. CAM applications are accessed through barcodes captured from the phone camera or through numeric values from the keypad. It supports linkage to paper-based processes and enables offline usage by downloading the entire application on to the device. The authors note that CAM applications are similar to IVR systems as both consist of sequences of actions. IVRs require an online connection whereas CAM applications can be performed offline.

The Hearsay system [20, 23] enables voice-driven browsing of web pages. The approach followed is to partition existing web pages and generate voice dialogs from those partitions automatically. However, the input to our system are not webpages but voice based input from the users and the output is a voice-driven application that is composed from existing components, as configured by the subscriber.

The Sublime system [22] presents distributed, multimodal and mobile environment for voice-driven personal information management. It is aimed at improving self organization through voice-driven lists of tasks-to-do. The initial system created simple lists that distinguishes interpretable speech from the actual lists stored as non-interpretable speech. In the multimodal version, they added support for GUI interface to improve overall usability with lists stored as interpretable speech. For our target population, however, voice-driven interfaces hold the most promise as has also been suggested by [21] and VoiGen strives to enable integration of data services with voice interfaces in order

[4]We paid a token amount of Rs 100 (USD 2) to the subscribers who took part in the evaluation process. This payment was made as a consideration to their time that they spent during the evaluation. However we believe that the token amount did not bias their response to the system.

to increase the benefits of voice-driven systems to the underserved.

Authors in [19] performed a user study to evaluate the acceptance of voice-driven application in rural India based on a sample speech-driven agriculture query system. The study reveals that even illiterate users were able to navigate through the dialog system, though the number of errors for such users was higher than compared with literate people.

7. CONCLUSION

In this paper, we presented VoiGen – an enabler for delivering data services to under-privileged in developing countries. The novelty of our approach comes from the fact that we use voice as the channel for delivery of those services which is important considering that our target population is from developing countries. Further, we approached the problem from a direction different from the existing efforts and proposed a system for enabling underprivileged to *offer* their own data services rather simply easing their *access* to such services. Our user study shows that the approach has the potential to enable IT to make a difference to daily lives of millions of new users.

Keeping in mind that network connections in developing countries can be quite unreliable at times [17], we will build further on VoiGen to add support for state based interaction that can continue across disconnections. We will also explore ways in which we can complement VoiGen to enable richer applications and build upon the vision of a World Wide Telecom Web for developing regions that would be parallel and complementary to the WWW. We envision that a WWTW integrated with the WWW can help bridge the widening gap between the IT-savvy and non-IT-savvy population of today.

8. ACKNOWLEDGMENTS

The authors thank Udit Pareek, IIT Guwahati for helping in implementing the generic *VoiceSite* generator.

9. REFERENCES

[1] Simple object access protocol (SOAP) version 1.2. Standard. http://www.w3.org/TR/soap12-part1/.

[2] World Population DATA SHEET, Aug 2006.

[3] R. Akolkar, T. Faruquie, J. Huerta, P. Kankar, N. Rajput, T. V. Raman, R. Udupa, and A. Verma. Reusable Dialog Component Framework for Rapid Voice Application Development. In *SIGSOFT CBSE*, May 2005.

[4] B. Bremner. India's Great Leap Forward. *Business Week, http://www.businessweek.com*, Aug 2006.

[5] E. Brewer, M. Demmer, M. Ho, R. Honicky, J. Pal, M. Plauch, and S. Surana. The Challenges of Technology Research for Developing Regions. *IEEE Pervasive Computing*, 5(2):15–23, 2006.

[6] G. Camarillo, Miguel-Angel, and Garcia-Martin. The 3G IP multimedia subsystem (IMS): Merging the internet and the cellular worlds. John Wiley and Sons.

[7] D. Chakraborty, K. Dasgupta, S. Mittal, A. Misra, C. Oberle, A. Gupta, and E. Newmark. Businessfinder: Harnessing presence to enable live yellow pages for small, medium and micro mobile businesses. In *IEEE Communications, Issue on Networking Technologies in Emerging Economies. To appear.* January 2007.

[8] D. Chakraborty, A. Kumar, A. Nanavati, and N. Rajput. WWTW: A World Wide Telecom Web for Developing Regions. *IBM Research Report No. RI06010*, Nov 2006.

[9] E. C. et al. Web services description language (WSDL) 1.1. W3C. http://www.w3.org/TR/wsdl, March 2001.

[10] J. R. et al. SIP: Session initiation protocol. RFC 3261. http://www.ietf.org/rfc/rfc3261.txt, June 2000.

[11] R. Hernandez and Y. Mugica. What Works: PRODEM FFP's Multilingual Smart ATMs for Microfinance. *http://www.digitaldividend.org/ pdf/prodem.pdf*, August 2003.

[12] T. Kristjansson, J. Hershey, P. Olsen, S. Rennie, and R. Gopinath. Super-human multi-talker speech recognition: The ibm 2006 speech separation challenge system. In *International Conference on Spoken Language Processing ICSLP*, September 2006.

[13] R. Kumar. E-Choupals: A Study on the Financial Sustainability of Village Internet Centers in Rural Madhya Pradesh. *Information Technologies and International Development*, 1:45–73, 2004.

[14] A. Neimi. Sip extension for event state publication. http://www.ietf.org/rfc/rfc3903.txt, October 2004.

[15] J. Nichols, B. Myers, T. Harris, R. Rosenfeld, S. Shriver, M. Higgins, and J. Hughes. Requirements for automatically generating multi-modal interfaces for complex appliances. In *Proc. of IEEE International Conference on Multimodal Interfaces*, 2002.

[16] T. S. Parikh. Position Paper: Mobile Phones may be the Right Devices for Supporting Developing World Accessibility, but is the WWW the Right Service Delivery Model? In *International Cross-Disciplinary Workshop on Web Accessibility (W4A)*, May 2006.

[17] T. S. Parikh and E. D. Lazowska. Designing an Architecture for Delivering Mobile Information Services to the Rural Developing World. In *Proc. Intl. Conf. on World Wide Web (WWW)*, May 2006.

[18] J. F. Pitrelli, R. Bakis, E. M. Eide, R. Fernandez, W. Hamza, and M. A. Picheny. The IBM expressive Text-to-Speech synthesis system for American English. *IEEE Transaction on Audio, Speech and Language Processing*, July 2006.

[19] M. Plauch and M. Prabaker. Tamil Market: A Spoken Dialog System for Rural India. In *Working Papers in Computer-Human Interfaces (CHI)*, 2006.

[20] I. V. Ramakrishnan, A. Stent, and G. Yang. Hearsay: enabling audio browsing on hypertext content. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, USA, 2004.

[21] J. Sherwani. Are Spoken Dialog Systems Viable for Under-served Semi-literate Populations? *PhD Thesis Proposal, Carnegie Mellon University, http:// www.cs.cmu.edu/ jsherwan/JS-proposal.pdf*, 2005.

[22] J. Sherwani, S. Tomko, and R. Rosenfeld. Sublime: A Speech- and Language-based Information Management Environment. In *In Proc. ICASSP*, May 2006.

[23] Z. Sun, A. Stent, and I. Ramakrishnan. Dialog generation for voice browsing. In *Proc. of Intl. Cross-Disciplinary Workshop on Web Accessibility at WWW, Scotland*, May 2006.