VOISERV: Creation and Delivery of Converged Services through Voice for Emerging Economies

Arun Kumar, Nitendra Rajput, Dipanjan Chakraborty, Sheetal K Agarwal, Amit Anil Nanavati
IBM India Research Laboratory
Plot No. 4, Phase-II, Institutional Area,
Vasant Kunj, New Delhi - 110 070, India,
{kkarun, rnitendra, cdipanjan, sheetaga, namit}@in.ibm.com

Abstract

WWW has made information accessible to computer users in various ways not imagined before. However, there is a huge pool of people, especially in emerging economies, still untouched by this revolution and are either unaware of, or are unable or to join this bandwagon. Mobile phones are increasingly empowering the under-privileged to utilize data and services beyond the basic voice communication. However, factors such as high illiteracy rate, cost sensitivity, and user interface issues prevent these users from deriving benefits of available infrastructure and services. We have developed a novel system – VOISERV *that enables ordinary telephone subscribers to create, deploy and offer their own customized voice-driven applications called VoiceSites. The generated VoiceSites get hosted in the network for low cost of ownership and maintenance, and are integrated with advanced services available in the converged networks of today.*

Keywords *Service Creation, IVRs, VoiceXML, developing regions, voice based applications, IMS*

1 Introduction

Access to information and services today in the IP world, has been 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 semi-literate and illiterate people in cities and rural areas of emerging economies. They are unable to afford computers or high end web-enabled handheld devices. The Internet penetration in developing countries is still below 10%. This is partially because, cheap PCs (about 220 USD) prove to be an expensive proposition for people in emerging economies. (56% people in developing countries live below USD 700 per year [12]). Furthermore, using a PC requires IT skills beyond language reading and writing, leading to a low acceptance rate.

Kiosk based solutions are being used to enable access to information and services on the Internet infrastructure for the non-IT savvy users. They 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 obtaining a *direct* access to the infrastructure. It does not let these users explore the various services available to them and the benefits they can obtain from them.

Mobile phones, on the other hand, being cheap and affordable communication devices, have had a far greater penetration among this population. Due to the convergence of IP world with the telephony world through intermediaries such as IP Multimedia Subsystem (IMS) [3], 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.

First, there are several user-interface issues that need to be addressed while designing services that can be accessed through a mobile phone. 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[1] is not attractive enough to ride this learning-curve.

Secondly, even though the telecom world is steadily

[1] A one minute call can cost twice as much as a domestic SMS

converging with the IP world, the benefits of converged services are not reaching traditional telecom users (PSTN and mobile subscribers with low end devices without Internet access). The benefits of converged services such as ClickToDial, Video Conferencing are reaching users accessing Internet through computers, while services such as VoiceOverIP, web access over phone [13], email over phone are being utilized by affluent and educated users who own high-end devices.

Third, is the lack of features, facilities and usable applications available to legacy telephone users. Users in IP world are allowed to host and offer their own services such as online stores, blogs, homepages etc., without actually being required to own the entire infrastructure. In contrast, users possessing a voice-based phone as the only medium to *access* technology, are deprived of the such benefits that they can derive from the available infrastructure. Voice-driven interfaces to applications have been found to have immense appeal for the illiterate and semi-literate users [10]. Yet, current voice-driven applications (such as airline/railway reservations, tele-banking solutions) are primarily meant to be used by the well-off IT savvy users. For instance, most of the underprivileged do not make reservations for domestic travel and do not have bank accounts either. Moreover, developing IVRs is a costly and time consuming proposition with investments in the infrastructure and programming effort. In other words, 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*.

The goal of our work is to enable the underprivileged population to exploit the capabilities and services offered through the convergence of IP and telephony world (referred to as converged infrastructure). To enable this, we propose a novel system - VOISERV, that enables individual phone subscribers to *create and offer* their own customized voice driven converged services[2]. VOISERV enables creation, deployment and hosting of *VoiceSites* that are developed by the users themselves. A *VoiceSite* is a set of interconnected VoicePages (i.e. Vxml files), hosted and co-located in the telecom infrastructure. One novelty in our system is that the users create these *VoiceSites* through a voice guided system called VOIGEN. Furthermore, VOISERV offers *VoiceSites* to be integrated with services offered by the converged infrastructure. Such integration and delivery is enabled through open IT standards such as Web Services [5] or through emerging services offered by 3G networks and IP Multimedia Subsystem (IMS) [3].

We envision that this approach would open up a plethora of opportunities for the underprivileged by enabling them to join the existing digital information highway. It would also prepare them to utilize the benefits of advanced services that would get enabled through Web services and IMS. Furthermore, till IMS gets widely deployed and the networks converge to all-IP, the telecom service providers can continue to leverage their existing GSM/CDMA/PSTN infrastructure and still be able to offer IMS based services to existing customers and increasing base of new ones in emerging economies.

The contributions of the paper are as follows:

- We present a technology and system for individual phone subscribers to create, host and deploy customized voice driven services.
- We enable access to data and services residing in the converged infrastructure to masses in emerging economies, through VOISERV.

The rest of the paper is organized as follows. Section 2 provides the motivation for VOISERV. Section 3 presents VOISERV system architecture and explains the hosting and the deployment model for *VoiceSites* along with the role played by network convergence. Section 4 presents our prototype implementation detailing the process of creation of *VoiceSites* and their integration with back-end services. Section 5 presents related work in the domain and we conclude in Section 6.

2 Motivating Scenario

Enabling the underprivileged population to create *VoiceSites* impacts various aspects of their life, specially in their business or daily activities. These people are primarily micro-businessmen having small jobs. Examples of such micro-businesses are plumbers, electricians, shopkeepers etc. To illustrate the benefits and features of our VOISERV, we present the following scenario of a plumber who uses VOISERV to create and maintain his *VoiceSite*.

A semi-literate plumber Sam, creates his VoiceSite by calling into a well-known phone number. VOIGEN *(the creation component of* VOISERV*) guides the plumber through a simple voice-driven interface, prompting for inputs whenever necessary. Sam provides basic information about himself, such as his service description, working hours etc.* VOIGEN *also offers him the ability to receive appointments for the next working day. Sam does not understand how the system enables this, but is very interested to use it and selects the option through a voice-prompted "yes".* VOIGEN *creates the VoiceSite and prompts Sam stating that his phone has been enabled with a VoiceSite.*

Next day, Dipti - a local citizen in the area encounters a plumbing problem in her house. She uses a telephone directory or an online yellow pages service to locate plumbers in

[2]The term *converged services* in our paper includes traditional information available through WWW, technologies such as databases and applications on top of it, IT services available through open standards such as web services, as well as next generation telecom services such as Presence based services available through the converged infrastructure.

the vicinity and then calls Sam for help. Sam is busy working on another appointment at that time and is unable to take calls. The call goes to the VoiceSite*. The VoiceSite determines Sams current availability status by using back-end services. It presents a voice prompt to Dipti stating that Sam is busy currently and provides her with the option of scheduling an appointment with Sam after 2 hours. Dipti happily schedules an appointment with Sam through voice driven prompts. Sam (after finishing his appointment) promptly calls his/her VoiceSite and receives Dipti's order.*

We observe some salient capabilities of our system from the above scenario: (1) Semi-literate people create their *VoiceSites* through a voice-driven interface (2) The generated *VoiceSite* integrates with a back-end calendaring application for maintaining working hours, schedules and appointments (3) The generated *VoiceSite* exploits location and presence components of IMS and applications running in the infrastructure to become aware of Sam's location and current status.

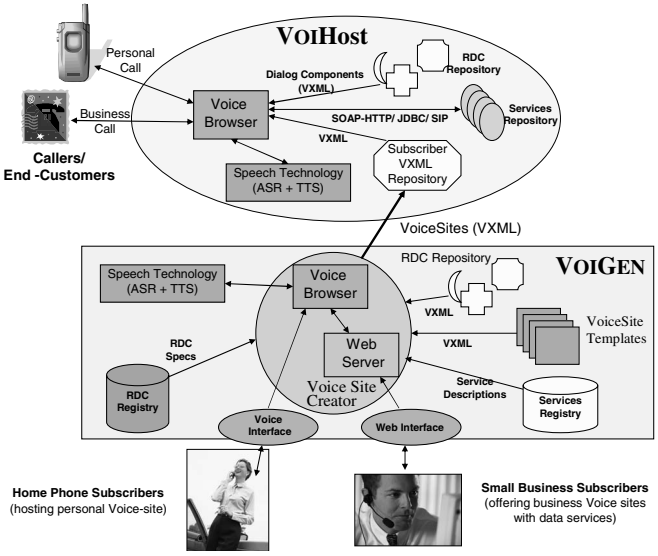

Figure 1. VOISERV **System Architecture**

3 VOISERV: Architecture and Components

Figure 1 depicts the architecture of VOISERV system consisting of VOIGEN – a voice driven generator of voice based applications, along with VOIHOST – a *VoiceSite* hosting engine.

3.1 VOIGEN: A Voice Driven Generator of Voice Based Applications

VOIGEN simplifies the process of creation of voice-based applications by enabling it through a voice-driven interaction. A phone subscriber could call in to VOIGEN and compose an application by navigating through the custom options offered to her. This application is then deployed in the form of a *VoiceSite*. VOIGEN makes use of existing components (reusable dialogs as well as IT components such as databases and web services) to compose custom applications. By virtue of having a voice-driven interface, the services get exposed to all telephony devices.

VOIGEN consists of the following components:

- **VoiceSite Creator:** It presents a voice-based (and an optional web-based) interface to the subscribers through which they can create new *VoiceSites* and edit previously created ones. It contains a Voice Browser that presents the voice interface to subscribers Voice Browser interacts with a Speech Technologies server that provides speech recognition and synthesis technologies required to capture and render voice input/output during the conversation with the subscriber.

  The Web Server presents a web based graphical user interface for creating *VoiceSites*. While micro-businessmen and individual users would typically use the voice interface for creating their *VoiceSites*, web based interface presents an alternative for situations where the *VoiceSite* becomes too complex to be edited easily through the voice interface.

- **VoiceSite Templates:** VOIGEN consists of predefined *VoiceSite* templates suited to different user profiles. A *VoiceSite* template is a VXML file that presents relevant menu options based on the type of *VoiceSite* being created. These templates are used by the *VoiceSite* Creator to help subscribers create their own *VoiceSite*. The actual content of the *VoiceSite* is governed by the subscriber. The *VoiceSite* Templates 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. *VoiceSite* Creator also supports the ability to create arbitrary *VoiceSites* not conforming to any template.

- **Reusable Dialog Component (RDC) Repository and Registry:** RDCs [1] are predefined Voice User Interface components 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. For example, a Date RDC could accept dates as spoken input in a particular format such as "Month day and year" and return a string in the MMDDYYYY format. The RDCs have configurable options and the prompts can be customizable based on the application. Depending on the options selected by the subscriber, appropri-

ate RDC references from RDC registry get included in the target *VoiceSite*.

- **Services Registry:** The *VoiceSite* being composed may also need to utilize other services such as some specific database, or a web service for calendar or appointments, or even a SIP-enabled [14] application. The Services Registry maintains a list of all services and their specifications. Based on the options selected by the subscriber, corresponding service stubs (i.e. VXML and java snippets) get included in the target *VoiceSite*.

3.2 VOIHOST: VoiceSite Hosting Engine

VoiceSites created by VoiceSite Creater component of VOIGEN get deployed in VOIHOST – a *VoiceSite* hosting engine. VOIHOST can reside in telecom provider's infrastructure and can be shared among its many subscribers. Hosting *VoiceSites* in the network brings several benefits. First and foremost, it enables the subscribers to pay per use rather than investing in a huge cost of owning and maintaining the *VoiceSite* infrastructure. Second, it provides connections to other applications available in the network thus enabling a channel for delivering existing services to traditional telephone subscribers. For the subscriber, the *VoiceSite* virtually resides on the phone.

The deployment process consists of assigning a telephone number to the *VoiceSite*. This is similar to a URL for a webpage. A VOIHOST deployment consists of a Voice Browser, Speech Technologies server and existing dialog components repository (such as a RDC repository) that have same function as in VOIGEN. In addition it contains the following components:

- **Subscriber VXML Repository:** This repository consists of all *VoiceSites* generated and deployed by various subscribers through VOIGEN.

- **Services Repository:** This repository contains various services on which the subscriber *VoiceSite* relies for its functionality. Examples include databases such as customer directory, appointment diary, and other advanced services such as presence information services and online banking services.

The calls arriving at the VOIHOST are from end-callers trying to reach the subscribers. Based on the called subscriber's preference, the incoming calls could always be transferred to her *VoiceSite*. Alternatively, the calls may be diverted to her *VoiceSite* only when the subscriber is unavailable or decides not to pick the call. VOIHOST runs on traditional IP networks.

3.3 Delivering Converged Services

As shown in Figure 2, *VoiceSites* provide a window to several categories of services in the converged networked world. Services and applications accessible to *VoiceSites* include (1) local applications in VOIHOST, (2) Remote applications exposed through Web Services [5], (3) Converged Services [14] and applications accessible through IMS.

Local applications are accessed by conventional protocols over TCP/IP. For example, a database application storing details of customers of the *VoiceSite* owner, can be connected to it using traditional client protocols such as Java Database Connectivity (JDBC) etc.

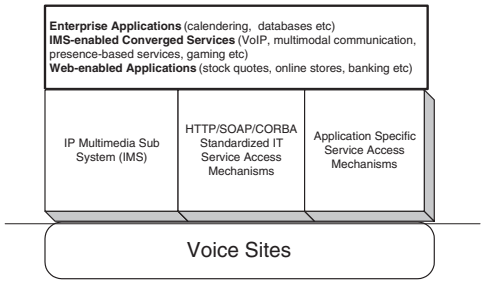

Figure 2. Architectural Integration of VOISERV with WWW through IMS

Web services enable easy integration of remote services with *VoiceSites*. They have become popular as the dominant technology for distributed systems integration. More and more enterprises are moving towards web services based software infrastructures as they enable easy interoperability. Functionality provided by web service enabled applications can easily be offered through *VoiceSites* by including appropriate web service client stubs in the target *VoiceSites*. This integration would enable subscribers to utilize services ranging from simple stock tickers to advanced banking through their *VoiceSites*.

VoiceSites also allow traditional telephony users to exploit the services offered by converged infrastructure. Users typically need 2.5G/3G enabled phones with IP support, to be able to effectively leverage services offered by converged networks. VOICESERV enables even non-IP based telephony users to access data and services in the 3G IMS infrastructure. For example, *VoiceSites* and services available through them can be configured to adapt based upon presence and other information available from appropriate IMS components. *VoiceSites* can exploit advanced call-control and multi-media control components of IMS and expose voice-driven services that might need them. IMS also increases the access modalities through which the *VoiceSites* can be accessed. Connecting to a *VoiceSite* could be as simple as clicking on a *VoiceSites* link, and a soft

phone application plug-in being triggered to browse through it. This is possible since addressing in converged networks moves away from traditional telephonic addresses to Session Initiation Protocol (SIP) [14] addresses (e.g. sip:manoj@in.ibm.com). Online representation of the huge population in developing regions through effective creation of *VoiceSites*, enables IP-based users to discover and avail the information and services offered by these users.

It is to be noted that integration with IMS components, even though not absolutely necessary, enables *VoiceSites* to be readily integrable with capabilities of next generation networks.

3.4 Applications and Implications

VOISERV enables several interesting applications and has business implications for telecom service providers.

The VOIGEN component empowers ordinary telephone subscribers with the ability to create several voice based applications such as Tele-Homepages (personal VoiceSite against their phone number), Tele-Blogging (voice based blogging accessible through phone) and personalized call handling (call routing logic can be embedded in *VoiceSite* options).

As described above, *VoiceSites* can be augmented to connect to web services available in the IT infrastructure and services in the IMS infrastructure. This facilitates creation of applications requiring access to other services. Examples include tele-online stores – shops selling merchandise over phone through a voice based interface. Such applications exist today, however, they are typically created and maintained by large enterprises and have a complex creation process. VOISERV enables the same functionality for individual phone subscribers and small/micro businesses allowing them to offer services such as appointment and scheduling as well as setting up businesses over phone such as tele-online shops mentioned above.

With acceptance over time, VOISERV can be expected to enable a web of interconnected *VoiceSites* that provide a multitude of services to the population that does not have access to the Internet. The resulting increase in penetration of IT into masses and the benefits of convergence outlined above, should enable an ecosystem [7] that allows both non-IT savvy as well as IT users to access information and services exposed by the underprivileged users.

VOISERV presents new business opportunities for the telecom operators. They could charge a subscription fee for hosting the basic *VoiceSite* applications whereas for back-end services and applications they could follow a usage based model. The telecom operator could charge for supporting multiple incoming calls against a single phone number since all of them could terminate in the subscriber's *VoiceSite* and can remain active simultaneously. *VoiceSites* acting as a gateway to various converged services, would also lead to a further increase in traffic and increased usage of services offered by the operator and others.

4 Implementation

We have designed, implemented and deployed a prototype of VOISERV at IBM India Research Lab. In this section, we provide detailed description of the prototype implementation of VOIGEN system, followed by the process of generating *VoiceSites*, deployment of the generated *VoiceSites* through VOIHOST and finally details of integration of *VoiceSites* with IT services and IMS infrastructure. Figure 3 shows the various implementation components of VOISERV.

4.1 VOIGEN

On receiving a call from the subscriber, VOIGEN carries out the following activities:

- It uses custom recorded prompts (spoken in a local language – Hindi) to describe to the *VoiceSite* Creators (mostly semi-literate or illiterate people who know only the local language) about the content they can put in their *VoiceSite*.
- VOIGEN prompts the caller to specify her preferences and loads the appropriate *VoiceSite* template (e.g. a specific template is used for a plumber who wants to create a site to advertise her services and schedule appointments with clients).
- VOIGEN stores user preferences and service options selected in the Subscriber Repository. It stores recorded prompts in the Voice Prompts Repository, as shown in Figure 3. We employ a combination of VXML, java scripts (with JDBC connectors) to capture and store the preferences in the respective repositories.
- VOIGEN, on receiving all the input, parses through the data obtained, and generates the *VoiceSite* for the caller. Depending on the caller's preferences, the generation either employs pure VXML or a combination of VXML, JDBC, TCP/IP constructs and JSP (e.g. JDBC/JSPs are employed if the VXML needs to integrate with a calendar service). The created *VoiceSites* are stored in the VXML Repository (indexed by caller's phone number).

We implemented a sample template in VOIGEN to enable small and micro businesses (plumbers, electricians, servants, home delivery services) in Indian metropolitan cities to create their *VoiceSites*. This VoiceSite template allows these classes of users to create their welcome page and

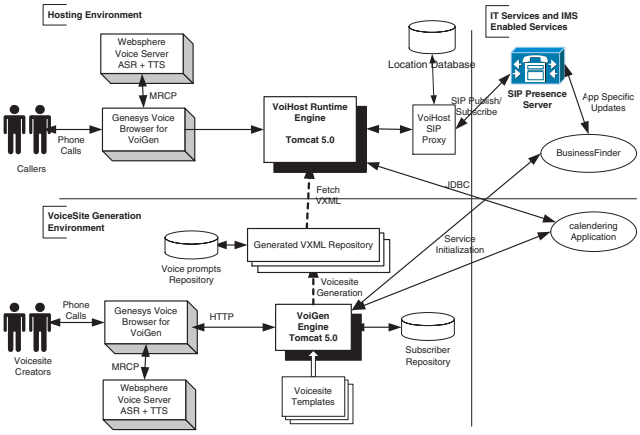

Figure 3. VOISERV **Prototype Components**

a business page. The business page provides information about their business and enable the users to interface with a calendaring service to provide scheduling features to their callers. The template also allows these users to create links to other *VoiceSites*. The destination *VoiceSite* can represent different entities like individual people, micro businessmen, local shops to large enterprise driven IVR systems. In our sample template the destination VoiceSites represented individuals that the subscriber might want to connect to or provide as references for their business.

We have also implemented a generic *VoiceSite* generation engine. Each node in a generic VoiceSite has an associated audio prompt and the set of options entered by the user. The options form the edges to other nodes in the tree. We implemented a depth-first-algorithm to enable users to create the nodes and edges of the tree. Thus the user has the freedom to define the structure of the tree. 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*. As pointed out in Section 3, we realized that human capacity of keeping track of all branches is very limited, especially when the creation mechanism is voice. Hence, for VOIGEN, we resorted to template-based creation of *VoiceSites*.

VOIGEN is deployed as an application on Tomcat 5.0 and connected to the PSTN network through the Genesys Voice Browser. Genesys utilizes Websphere Voice Server to enable speech recognition and text-to-speech conversions. The speech recognition errors have been made minimal by using input through telephone keypad (DTMF) for complex user input such as a *date*.

4.2 VOIHOST

VoiceSites created through VOIGEN are stored in a VXML repository. The VOIHOST runtime engine is connected to the PSTN network (through the Genesys Voice Browser) and is capable of receiving phone calls. Each *VoiceSite* has a phone number associated with it. On receiving a call, the appropriate VXML representation of the *VoiceSite* is loaded (from the VXML Repository) and rendered by the Genesys Voice Browser as seen in Figure 3.

The VOIHOST runtime engine also executes the code snippets to interface with back-end services such as the calendaring service or services available over the converged infrastructure. In the next subsection, we explain the details on how *VoiceSites* are integrated with converged services.

4.3 Integration with Converged Services

One goal of VOISERV is to allow users to create and offer their own services through their *VoiceSites*, by exploiting services available in the converged infrastructure. To demonstrate the integration of *VoiceSites* with local applications and back-end converged services, *VoiceSites* in our implementation interface with (a) local Calendaring Service and, (b) an advanced web-based presence-information driven matchmaking service called BusinessFinder [4] developed at IBM India Research Lab.

The Calendaring Service provides appointment scheduling features to the users using a simple database implementation for storing appointments. VOIHOST, while accepting appointments from callers (for a certain *VoiceSite*) uses JDBC connectivity code snippets to store appointments and also check for busy hours of the creator. The VoiceSite creator can herself modify, delete appointments or add busy hours. VoiceSite creators are free to choose which optional features of the template they wish to include in their VoiceSite. For instance, the calendaring service was implemented as an optional feature in our prototype.

BusinessFinder [4] is an IMS-enabled service that utilizes Location and Presence capabilities of IMS infrastructure to provide efficient, on-demand, context-aware matching of customer requests to nomadic vendors. Web-enabled users can use a web portal to find out *nearest and available* vendors through this service.

In our current prototype, we have enabled *VoiceSites* of plumbers (who are essentially nomadic vendors) to be subscribed to BusinessFinder - hence making web users discover *VoiceSites* through a web-based portal, after they are created. To achieve this, VOIGEN sends a subscription request (using a simple stub code) to BusinessFinder through direct subscription channels exposed by BusinessFinder.

VOIGEN uses the SIP Presence Server (an IMS component) to receive real-time updates about the creator's location status. This information is cached locally and used by the VOIHOST engine. The VOIHOST SIP Proxy carries out this operation.

By integrating with SIP infrastructure, *VoiceSites* now

have access to dynamic location and availability information about a user available in the various IMS services. This rich presence information can be exploited in several ways to adapt the services offered by the creator. In the current implementation, we customize the scheduling service offered by the user based on his current location. For example, if the creator (plumber in this case) is in her home location, she is inferred to be available to take new appointments in that area. Otherwise a busy status is indicated to the caller.

5 Related Work

Automatic generation of user interfaces from a given specification has also been addressed in [8]. 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. Our end goal is that of enabling simple and easy creation of voice applications that deliver information and services to underprivileged users.

The CAM system [10] also takes a non-PC centric approach and suggests a change in the current service delivery model [9]. It proposes to leverage 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 bar-codes 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 authors also observed that spoken language based interaction not only enables illiterate users but even literate users feel more comfortable with audio feedback in local language [10]. However, the basic goal of their work is integration of mobile applications with paper based processes whereas we have proposed to enable integration with high end backend services.

The Hearsay system [13, 17] enables voice-driven browsing of web pages. They partition existing web pages and generate voice dialogs from those partitions automatically. However, the input to our system are not web pages 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 [16] 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 distinguished interpretable speech from the actual lists stored as non-interpretable speech. In the multi-modal 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 [15]. and VOISERV strives to enable integration of data services with voice interfaces in order to increase the benefits of voice-driven systems to the under-served.

Authors in [2] 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.

Authors in [11] 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.

6 Conclusion

In this paper, we presented VOISERV – an enabler for the underprivileged population in emerging economies to exploit the benefits of converged services available in the IP and telephony world. VOISERV enables underprivileged users to create and offer their own voice-driven applications and also utilize services available through the converged infrastructure, with little or no technical know-how.

The novelty of our approach comes from the fact that we use voice as the user-interface channel and provide a low cost mechanism for creating voice-driven applications and using converged services. This is important considering that our target population consists of non-IT savvy users. A preliminary survey [6] done with our prototype showed that most users in our target group did not have much problem in creating their own *VoiceSites* and they appreciated the benefits it can bring.

VOISERV presents new avenues for the Telecom operators to generate revenue by hosting these voice-based applications in the telco infrastructure. Hosting these applications result in increased incoming call traffic. Also this model relieves the users of incurring the high cost of owning and maintaining the infrastruture to support these applications. Thus users can create and offer services at a very low cost.

In future, we intend to explore ways to utilize VOISERV to enable rich applications toward realizing the vision

of a World Wide Telecom Web (WWTW) [7] 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.

Acknowledgments

The authors thank Sandeep Jindal, IBM, for building the initial prototype of the VOIGEN system. The authors also thank Udit Pareek, IIT Guwahati, for helping in implementation of the generic *VoiceSite* generator.

References

[1] 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.

[2] 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.

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

[4] 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.

[5] E. Christensen, F. Curbera, G. Meredith, and S. Weerawarana. Web services description language (WSDL) 1.1. W3C. http://www.w3.org/TR/wsdl, March 2001.

[6] A. Kumar, N. Rajput, D. Chakraborty, S. Jindal, and A. A. Nanavati. VOIGEN : A Technology for Enabling Data Services in Developing Regions. *IBM Research Report No. RI07001*, Jan 2007.

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

[8] 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.

[9] 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.

[10] 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.

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

[12] Population Reference Bureau. World Population Data Sheet, Aug 2006.

[13] 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.

[14] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. SIP: Session initiation protocol. RFC 3261. http://www.ietf.org/rfc/rfc3261.txt, June 2000.

[15] 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.

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

[17] 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.