

# Spoken Web: A Parallel WWW in Developing Regions

Sheetal K. Agarwal, Arun Kumar, Amit Anil Nanavati, Nitendra Rajput  
IBM India Research Lab,  
4, Block C, ISID Campus, Vasant Kunj, India

*Abstract* — The Spoken Web is a voice-based equivalent of the World Wide Web (WWW), developed by IBM Research Laboratory, India, primarily designed for rural and semi-urban people to provide information of value to them through their mobile or landline phones. It will also help the government / industry / micro-business to reach out to rural population with their offerings and help the people at the Bottom of the Pyramid. The vision is to create an information ecosystem that helps provide Internet-like information services through phones.

## MOTIVATION

The World Wide Web (WWW) has become a rich source of accessing information and services over the last decade. Hardly anyone reading this paper would have not accessed the WWW today, for accessing some kind of information. The Internet therefore is one of the most significant technologies that have changed our daily lives in the recent past. This has been made possible through the numerous information sources and applications available over the WWW. However the impact of the WWW is still not at the level of some basic facilities such as the railroad, electricity

Because of the low literacy rates, and the low Internet penetration in India, the PC based model to deliver information and services in the rural areas in India has not been as effective as in the western world.. There is a significant percentage of population that is still untouched by the WWW revolution and are either unaware of or are unable to catch the momentum. Even today, barely 17% of the world's population has access to the Internet. There are a variety of reasons that act as a hindrance for this technology to impact the remaining 83% section of the human population. Firstly, 53% of the world population lives below USD 2 per day – so they cannot afford a PC or high end phones and hence cannot access the Internet. Secondly, a significant portion of the remaining 30% is illiterate and semi-literate people who do not know how to operate a computer. Thirdly, most of the information and applications available on the Internet is hardly relevant to this section of the society. However, for this technology to become a commodity such as a road or

electricity, a significant shift in paradigm is needed. Incremental improvements in terms of the sophisticated services, advanced user interfaces, easier application authoring techniques for the WWW do not appear to be helping in providing a shift of such a large scale. Interestingly, the telecommunication network does not face some of the challenges of the Internet world – from an acceptance perspective. The cost of a phone is significantly lower than a PC and, the learning required to operate a phone is negligible as compared to a PC, especially when the phone is used as a device to communicate in free speech.

Thus telecommunications have become a commodity for the common man and are a step closer towards achieving that status for the underprivileged as well. The Spoken Web, in our vision, has the potential to deliver to underprivileged, what WWW delivers to IT literate users today. In this paper, we present the concept of Spoken Web and illustrate a specific usage scenario to explain the working model. We also present the details of a specific service called VoiceSite to create services for the individuals.

## INTRODUCTION

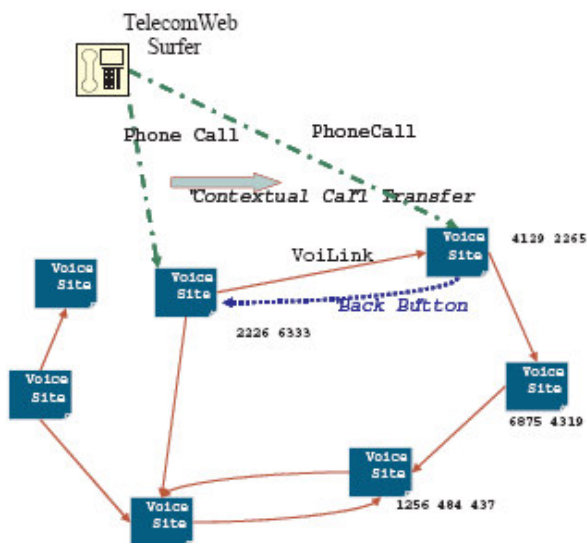
The Spoken Web [1] is a voice-based equivalent of the World Wide Web (WWW), developed by IBM Research Laboratory, India, primarily designed for rural and semi-urban people to provide information of value to them through their mobile or landline phones. It will also help the government / industry / micro-business to reach out to rural population with their offerings and help the people at the Bottom of the Pyramid. The vision is to create an information ecosystem that helps provide Internet-like information services through phones.

VoiceSites are voice driven applications that are created by the subscribers and hosted in the telecom network. They are represented by a unique phone number and can be accessed from any phone instrument, mobile or landline through an ordinary phone call to that number. The phone does not require any extra features or software to be installed on the device. VoiceSites are therefore analogous to websites in the WWW but can be accessed by dialing a phone number

and information can be listened rather than being read or seen. Creation of a VoiceSite is made easy by the VoiGen system through which anyone can call up the VoiGen system and interact with it through voice. This can enable any illiterate person to create her VoiceSite. Such a system enables easy local-content creation. All information in the VoiceSite is stored as audio messages that are recorded by making a phone call to the system. A blind person can create his/her VoiceSite by a simple phone interaction with the VoiGen system.

By answering simple questions as the ones shown in this interaction, a visually challenged user can easily create her VoiceSite. A person who needs a classical vocal teacher can access this VoiceSite and get her information. Such a simple mechanism can improve the business of the VoiceSite creator. Users can dial into Spoken Web through mobile phones or landline phones, create their own Voicemails (analogues to websites) in local languages, and browse voicemails created by others. An example in the rural setting is of a Voicemails Owner who gathers local information that has relevance to a specific cluster of villages, and uploads it onto his Voicemails. The villagers dial the Voicemails number and receive the information that they want. All this is possible just by talking over the phone – in their own local language and dialect.

Spoken web has the ability to link multiple voicemails together [2]. User can call in to any of these voice sites and move from one to the other and the context get transferred seamlessly.



**Figure 1.** The Spoken Web

The Spoken Web has tremendous implications: For the visually impaired this means that they do not have to spend months learning how to use a computer and the various accessibility software such as screen readers. It also enables the non-PC literate people to access

information and services that were hitherto unavailable to them through IT systems [3].

### MOTIVATING SCENARIO

*Sanjay has just moved to a house in a new city. He needs to fix several taps and electricity fittings. However since he is new in the locality, he does not know where to find the plumbers and electricians. He does not want to rely on the yellow pages since they have mostly stale limited information.*

*At the other end, Tom is a skilled electrician who works in the area where Sanjay lives. Tom has a lot of free time but is not able to get business since not many people know about him. Tom is an individual businessman who is not IT-savvy, so he cannot afford to advertise his skills through the Internet. However, since Tom works in different houses when he gets work, he carries a mobile phone so that people who know him can reach him for business.*

*To increase his business reach, Tom creates his VoiceSite. This can be done by making a phone call to the system and then specifying the different features about his service. When the system asks, Tom replies by mentioning his work area, charges, working hours and a list of customer references who can vouch for his good work. This information is used to create his VoiceSite. Sanjay makes a phone call to Tom's VoiceSite and gets all the details about Tom's work. The service charges and work times suit him. He then calls a couple of Tom's references to verify if he has done good work earlier. Satisfied with the responses, Sanjay finally asks Tom to provide the service to him.*

*The VoiceSite helps Tom in getting the business and it helps Sanjay in getting a right person for the services that he required.*

We observe some salient capabilities of our system from the above scenario:

1. Using this system does not require to learn any new interaction modalities – just a voice conversation with the system is good enough. Thus semi-literate people create their VoiceSite through a voice-driven interface.
2. At no extra cost (other than the cost of making a phone call), the service providers and service requesters will be able to exchange the information through the VoiceSite.
3. It requires no extra device or hardware to create a voice avatar for a person. They can do this just by making calls through their mobile phones.
4. It makes the vendor accessible without having to establish a direct communication channel with the service providers.

A sample interaction of the service provider with the VoiceSite system is presented in Figure 2. The number of questions are made very few and the system asks the

user to record her details. The system is not heavily dependent

on the speech recognition errors since very few utterances of the end-user are converted to text – most is recorded. Moreover, since the generated VoiceSite of the service provider is in his/her own voice, callers can easily relate the VoiceSite to the actual person.

When a service requester needs the service of a carpenter, it calls the VoiceSite of Ratan and interacts with the system as shown in Figure 4 to get information about the carpenter.

---

S: Welcome to the VOI Avatar system. Please speak your welcome prompt  
 C: My name is Ratan and I work as a carpenter in south extension, malviya nagar and nearby areas  
 S: The welcome prompt has been recorded.  
 S: Say 1 if you wish to specify your references, Say 2 to specify the charges or Say 3 to specify your working hours.  
 C: One  
 S: Say the people and their contacts who can vouch for your work.  
 C: You can ask about me from Mr. Sanjay, his number is 9484042433. Or you can ask from Mr. Gaurav - 9844740248.  
 S: References have been recorded. What next?  
 C: Two  
 S: What are the charges that you ask for your service.  
 C: I charge \$1 for an hour and \$10 for a day. However the rates really depend on the type of job.  
 S: Charges have been saved. What next?  
 C: Three  
 S: What are your working hours.  
 C: I work from 9 am to about 8 pm - all seven days a week.  
 S: You have specified all the preferences. Should I save them?  
 C: Yes.  
 S: Your VOI Avatar has been created. Thanks for using the system.

---

**Figure 2:** A sample interaction of service provider with the Spoken Web system.

Both, the service provider and the service requester are able to provide and access information through the normal telephony network. The VoiceSite is able to provide the information about the mobile worker Ratan and thus can enhance his business prospects.

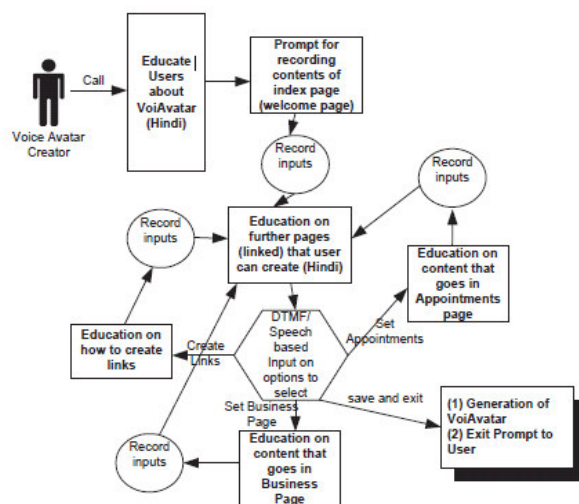
## SPOKEN WEB DEPLOYMENTS

In all, there have been many applications deployed on the Spoken Web platform. In all the settings, end-users call up the VoiceSites by dialing a specific phone number from a regular phone. The VoiceSites are hosted on a centralized server. The VoiceSites are authored in VoiceXML and use a VoiceXML browser (Genesys Voice Platform, GVP, in our case) and a speech recognition server (Websphere Voice Server, WVS, in our case). Built on standard languages and supporting standard protocols, the Spoken Web VoiceSites themselves do not depend on the specific servers mentioned above.

The Spoken Web applications have been used by more than 10,000 users across several states in India [4,5]. In this section, we will provide a specific test deployment that highlights the ease of VoiceSite creation in the Spoken Web.

The specific application presented in this section, on receiving a call, carries out the following activities:

1. It uses custom recorded prompts (spoken in the local language) to educate the caller about the content he/she can put in his/her VoiceSite.
2. It prompts the caller to specify his/her preferences and records them.
3. It browses through the template and guides the creator to provide information to customize the template. E.g. a template of a mobile micro-businessman such as a plumber, allows the caller to provide basic information about him, followed by the ability to create reference/links to other users as well as professional information about his/her business.
4. On receiving all the inputs, it parses through the data obtained, and automatically generates a VoiceSite for the caller using a generation engine [4].



**Figure 3:** Control flow diagram of a voice template used to create the VoiceSite mentioned in this section.

Figure 3 shows the control flow diagram of a concrete template that we used in VoiceSite to enable small and micro businesses (plumbers, electricians, servants, home delivery services) in metropolitan cities to create their avatars. This voice template allows these classes of 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.

The name of the author and affiliation should follow on separate lines in upper and lower case letters. Use the Times or Times Roman 10 point typeface.

## SYSTEM EVALUATION

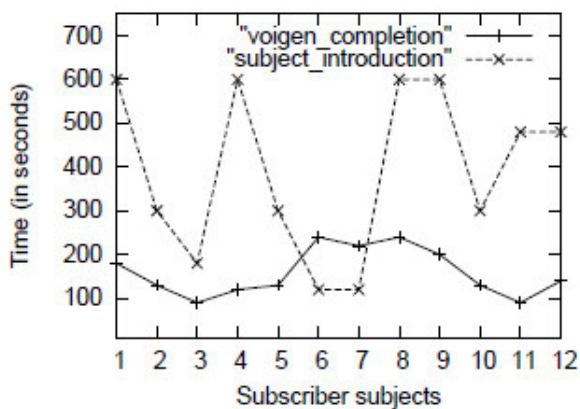
While the different Spoken Web applications have been vastly used by end-users, we present results from a

specific study that indicates the ease of creation of VoiceSites for the abovecategory of users.

The subjects of this study were typically skilled labourers (such as electricians, plumbers, carpenters) who 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 labourers work independently. Most of them have had only about five to ten years of formal school education. Their yearly earnings are below Rs. 75000. 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 reachable (available on call as far as possible). 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.



**Figure 4:** Time taken by each subject to understand the concept of Spoken Web and to build their VoiceSites.

Figure 4 shows the time that was spent to build the voice avatars for the 12 subjects. As seen, most subjects were able to generate their own voice avatars within 4 minutes. A 4 minute phone call in India costs less than 4 rupees. 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 voice avatar 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 4. All showed tremendous interest in the concept of voice avatars and the fact that their work can be advertised without them actually requiring to purchase any additional equipment.

## CONCLUSION

Spoken Web is an attempt to envision a service for the underprivileged communities, similar in theme to

what WWW is to the IT literate users today. It enables masses to access information and services through voice driven channels. Information and services could be community created as well as leveraged from existing Internet infrastructure. We have summarized key technology enablers for this vision to be successful as well as attempted to articulate interesting research problems in this vision. Enabling voice-driven front-ends to websites and WWW services would only enable the underprivileged to access global information.

Spoken Web builds a vision of a service for users in developing regions that harnesses WWW services as well as the ones in the converged networks — under one umbrella. Further, it provides the means to create and sustain an ecosystem of local (and global) services, information and communities relevant to these underprivileged users.

## REFERENCES

- [1] A. Kumar, N. Rajput, D. Chakraborty, S. K. Agarwal, and A. A. Nanavati, "VOISERV: Creation and Delivery of Converged Services through Voice for Emerging Economies," *IEEE WoWMoM*, Helsinki, Finland, June 2007.
- [2] S. K. Agarwal, D. Chakraborty, A. Kumar, A. A. Nanavati, N. Rajput, "HSTP: Hyperspeech Transfer Protocol," *ACM Hypertext*, Manchester, UK, 10-12 September 2007.
- [3] A. Kumar, N. Rajput, S. K. Agarwal, D. Chakraborty, A. A. Nanavati, "Organizing the Unorganized - Employing IT to Empower the Under-privileged," *WWW*, April 2008.
- [4] S. Agarwal, A. Kumar, A. Nanavati, N. Rajput, "Content Creation and Dissemination by-and-for Users in Rural Areas," *ICTD 2009*.
- [5] N. Patel, D. Chittamuru, A. Jain, P. Dave, and T. S. Parikh, "Avaaj Otalo ? A Field Study of an Interactive Voice Forum for Small Farmers in Rural India," *CHI 2010*.