VoiKiosk – a Spoken Web based solution for Rural Communities in Developing Regions

Supplimentary Material for presentation made on 17th November 2008

Contact: Arun Kumar (kkarun@in.ibm.com)

Technology Background

Internet 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 World Wide Web (WWW). However, there is a significant percentage of population that is still untouched by this revolution and are either unaware of or are unable to catch the momentum.

Even today, barely 21% 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 79% section of the human population. Firstly, a large percentage 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 population consists of ***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. Essentially, people need *Information Literacy* rather than *Computer Literacy* to derive benefits out of the information based economy.

Interestingly, the telecommunication network does not face some of the challenges of the Internet world. The cost of obtaining and operating a phone is significantly lower than a PC. It does not require a consistent power source and the learning curve 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.

With the goal of enabling *information literacy* for the under-privileged, we have created World Wide Telecom Web (WWTW) (aka Spoken Web) -- our vision of a voice-driven ecosystem parallel to that of the WWW. WWTW is a network of VoiceSites that are voice driven applications *created* by the subscribers themselves and *hosted* in the telecom infrastructure. VoiceSites are accessed by calling up the associated phone number and interacting with its underlying application flow through an ordinary telephone. VoiceSites are analogous to websites in the World Wide Web and can be linked to other VoiceSites through hyperlinks called *VoiLinks* supported by Hyperspeech Transfer Protocol (HSTP).

We believe that Spoken Web has the potential to deliver to underprivileged, what WWW delivers to IT literate users today. Specifically, it

- enables the underprivileged to *create*, and *offer* information and services produced by themselves, *on the infrastructure already available today*.
- provides *simple and affordable access mechanisms* to let the masses exploit IT services and applications that are currently available to WWW users, and,
- provides *a cost effective ecosystem* that enables users to create and sustain a community parallel to the WWW.

Spoken Web acts as a platform that enables several applications for the population currently untouched by the information systems of today. VoiKiosk is one such solution that is meant primarily for rural population living in remote areas. For some of the other solutions based upon Spoken Web please refer Organizing the Unorganized and Visually Impaired.

VoiKiosk Solution

Existing efforts are aimed at setting up information kiosks that consist of a PC, internet connection, printer and related accessories. They are run by a kiosk operator and the goal is to facilitate access to internet based information and services at low cost. However, we see several problems with this approach, some of which are also identified in a Microsoft report.

First, end users do not have direct access to the kiosk. The kiosk operator typically acts as an intermediary as a majority of the end users are not computer literate. Second, a lot of the information that is required on a daily basis, especially locally relevant information, is simply not available on the Web. Information such as the timings for scheduled electricity blackouts, the local bus schedule, visiting hours of the doctor from the near-by town etc. is not available on the Web.

Third, people may have to travel a few kilometers or to the neighboring village to access the kiosk facilities. Kiosks are also susceptible to hardware failures, and the more than 8-9 hour power cuts make their use very difficult even with some power backup. Further, the problem of viruses multiplies the maintenance effort required. Lastly, the current kiosk models enable a one way interaction where the end users are mere consumers of information and services. Leveraging the increased mobile penetration and comfort of semi-literate and illiterate people with speech based interfaces, we present VoiKiosk as an alternate model to create and host village portal on phone.

A VoiKiosk acts as an information and service portal for a village. It can be a central point of access for a community where information relevant to the community can be posted/accessed directly by the users themselves. This solution doesn't rely on internet connectivity which is most often not available in the rural areas and most importantly it allows end users to directly interact with the services removing the dependence on the kiosk operator.

There are two types of users for this system. First is the kiosk operator who is responsible for supplying local content on the VoiKiosk. She configures the VoiKiosk for the village through a voice interface. Second are the end users, who can either access or post information on the VoiKiosk depending on the services that it supports.

The information on the kiosk is uploaded, maintained and consumed by the local community. Illiterate and semi-literate users find it very convenient to talk to the system in their own local language. The highly interactive nature of the system empowers the local community to become information providers instead of just consumers. The system makes use of already deployed telecom infrastructure and does not place any new infrastructural requirements. This enables NGOs, government and other agencies to reach out to the rural areas in a way that has not been possible before.

Frequently Asked Questions

Q: What level of local tech is needed to support a VoiceSite?

A: VoiKiosk (and other solutions based upon Spoken Web, for that matter) can be deployed in two primary modes. First, is the hosting mode where a service provider deploys and manages the necessary h/w and s/w resources and in turn offers VoiceSite hosting as a service to others. A subscriber could set up a VoiKiosk (or her personal VoiceSite) based upon the VoiceSite template(s) included in the Spoken Web deployment of the service provider. The phone number of this VoiKiosk could be advertised for others to call. This mode allows even non-IT savvy people to create and manage a VoiKiosk portal with minimal training.

The second model is one where a small organization (such as an NGO) may wish to host its own set of VoiceSites. Here, the responsibility of obtaining, setting up and maintaining the setup lies with the organization. This mode requires significant technical resources since in addition to a web server, a couple of different kinds of servers (such as a speech server, voice browser) would need to be installed, deployed and maintained. Work is under progress to make this kind of deployment a bit simpler to manage and maintain.

Q: Would regional accents be a problem for voice recognition systems?

A: Dealing with different languages and several regional accents in those are definitely a challenge for voice recognition systems in general, especially if the aim is to be closer to natural language speech recognition. However, in the context of VoiKiosk (and Spoken Web, in general) local language specific speech recognition is not crucial for provide multilingual VoiceSites. The user interface for templates is designed such that minimal speech recognition is required which can be served with any good recognition engine in a

language independent fashion. Having said that, this approach does make the job of the VoiceSite template designer a bit harder for designing the user interface. But on the positive side the target population is one that is faced with the choice of few hours of travel with a whole day and money spent to get some piece of information/service as an alternative to listening to a few extra prompts. In practice, we have found out that these people are much more patient with the voice interface offered by VoiceSites since the benefits received outweigh the inconvenience faced. Also, Dual Tone Multi-Frequency (DTMF) input is allowed as a fallback option to speech recognition.

Q: What is required to run the service, is the software free/open source?

A: Details of a deployment are given here. The components such as a database, web server are available in open source. But other components such as a speech server that supports Text-to-speech and Speech-to-text are typically not available in open source. In addition, we also need an interface where a telephone line can terminate along with a Voice Browser that can render Vxml content. Asterisk is an open source software in this space but it has some limitations.

Q: What is accuracy/error rate for this in field tests?

A: As mentioned on the call, we have not done studies to determine this. However, the ever increasing usage of the deployed portal in the absence of any training or even advertisement gives us confidence that error rate/accuracy is not a major concern – at least for information, non-critical (e.g. financial) services.

Q: Is there a translation facility to access the same site as a normal web page?

A: Not yet, work is under progress on that front.

Q: From first pilots, what types of services are the most promising? Who is the most likely to use this system?

A: The pilots are limited by the goals of the partners we work with. So, it is hard to give an exhaustive, definitive list of types of services and people to be benefited. We have seen at least non-tech savvy/ illiterate people in rural areas, NGOs, visually impaired people, and micro-businesses in urban areas (such as plumbers, carpenters, electricians) as immediate beneficiaries. While there are other applications and users emerging, it would take some time to establish the benefits as observed in the field.

Q: The system would be useful for the http://www.questionbox.org/
A: Yes, definitely. It would help QuestionBox on at least three counts

1. to scale up by automating the information gathering and delivery component
2. enable it to provide locally relevant content rather than what is available on the Web
3. let villagers contribute to the information and content in addition to simply accessing what is available.

Q: Does it provide support to people wanting to create their own applications?

A: Not VoiKiosk which is a specific solution, but one of the key novelties of the Spoken Web platform (on which VoiKiosk is based) is that it enables individuals to create and offer their own applications without any technical know how. This is achieved through the use of a library of pre-built VoiceSite templates. VoiKiosk is one such template which different villages can use to create a kiosk for themselves.

Q: What about semantic/classification issues upon creation... controlled vocabulary used?

A: Not sure, if I understood the question correctly but yes, the template used to create the VoiceSite defines the vocabulary of the VoiceSite. It is therefore controlled in that fashion.

Q: Does a Voice Browser include Speech Recognition engine?

A: No, A Voice Browser is typically available separate from a Speech recognition engine.

Q: Are there simpler voice browsers without voice recognition?

A: There is a VoiceXML browser plugin available for Asterisk and Text to Speech facility. We have not tried it ourselves yet.

Q: Speech processing may not be mature enough to handle all kinds of nuances

A: Very True. And it is because of this reason that our design tries to work around the speech recognition problem by relying on it as little as possible. We use a single spoken word input recognition rather than lengthy phrases or natural language input. As and when speech recognition improves, the VoiceSites' UI capability can only benefit from it and get better.

Q: Is the system completely integrated? Is it modular?
A: The system is modular and needs to be integrated at deployment time. In fact, it basically follows the 3-tier web application architecture with the addition of a telephony

interface and Vxml (instead of Html) in the front end along with speech recognition capability.

Q: You said it is preferable to SMS. Why?

A: SMS is not favored by the illiterate and semi-literate population which constitutes a major portion in countries like India and, I believe, also in Africa. In addition, SMS is not available in local languages and given the diversity of the country, number of languages and dialects run into several hundreds. So, even literate, non-English speaking people find it difficult to use SMS. Another reason for popularity of voice compared to SMS is presence of very low call rates in India. These are barriers for SMS as an interface and even before it can be considered as a potential candidate. Whether SMS can serve the needs of the services in demand is yet another question. There are no SMS only accounts on offer in India, to the best of my knowledge.