Alternative information Web for visually impaired users in developing countries

S. K. Agarwal
A. Kumar
A. A. Nanavati
N. Rajput

Websites in the World Wide Web are primarily meant for visual consumption. Accessibility tools such as screen readers that render the visual content in audio format enable the visually impaired to access information on websites. Despite standards that are available to make websites more amenable for screen reading software, not many website authors embed the required metadata information that feeds into such tools. Moreover, the wide variety of visual controls available makes it more difficult to interpret websites with screen readers. This problem of accessing information and services on the Web escalates even further for visually impaired users in developing regions, since they are either semiliterate or illiterate or cannot afford computers and high-end phones with screen-reading capability. In this paper, we present an alternative platform: the World Wide Telecom Web (WWTW), which can be used for delivering information and services to the visually impaired. WWTW is a network of VoiceSites that can be created and accessed by voice interaction over an ordinary phone. We present user studies that demonstrate that using applications on the Telecom Web does not require extensive training. The study leads us to believe that the Telecom Web can be the mainstream Web for blind users, particularly in developing countries.

Introduction

For a typical person, access to information is a key requirement today. Over the last decade, the World Wide Web has grown tremendously to become the largest source of information. It is also being used by governments and enterprises to provide services to their citizens and customers. There are several existing efforts at making the content on the Web accessible to visually impaired users. These include software tools such as screen readers, Web accessibility standards [1], and government laws [2].

However, of the 37 million blind people worldwide, 90% live in developing regions [3], and more than half are in India (9 million), Africa (7 million), and China (6 million). Only 26.6% of the entire world population has access to the Internet, and for developing regions, the fraction is even smaller (8.7% in Africa and 20.1% in Asia) [4]. Fifty-three percent of the remaining world population has an income of less than $2 (U.S.) per day [5] and cannot afford a computer or high-end phones required to access the Internet. A significant portion of the remaining 30% are illiterate and semiliterate people [6] and are not information technology (IT) savvy. These statistics clearly suggest that the World Wide Web is not accessible to a very significant segment of visually impaired users.

Compared to Internet penetration, the last few years have seen a tremendous growth in mobile phone penetration in these developing regions. Phone penetration in developing countries had reached 32.4% in 2006 [7]. The actual penetration is even higher when we consider the fact that a phone is usually shared in a family (which has an average size of four people) in such countries [8]. In this paper, we present an alternative mechanism for delivering information services to the visually impaired that leverages the pervasive reach of the phone network. Since the majority of the visually impaired population is from developing countries, it is important that the solution be affordable and easy to learn and use. We use the concept of VoiceSites [9] that can be accessed through voice by a phone as a mechanism for

Digital Object Identifier: 10.1147/JRD.2010.2066350

0018-8646/10/$5.00 © 2010 IBM

information upload and access for visually impaired users. The contribution of this paper is threefold:

- We propose a VoiceSites-based World Wide Telecom Web as more suited to the visually impaired in developing regions than the existing World Wide Web.
- We perform user studies with 43 blind users to evaluate the following: 1) initial learning difficulties; 2) ease of use of VoiceSites; and 3) accessibility of VoiceSites for blind users.
- We summarize the user studies and provide insights gained from the interaction of blind users with the VoiceSites.

In this paper, we first describe the underlying concepts of VoiceSites and the Telecom Web and then describe the application scenario that was used to perform the user study. We present the profile of the subjects, the user study method, and the results of the user study. Finally, we discuss some potential applications that can be hosted on VoiceSites before concluding this paper.

Related work

The work presented in this paper falls in the unique intersection of existing work in 1) providing Web accessibility to blind users and 2) providing affordable access to the poor in developing countries. Although there is a huge population that is blind and poor, not much work has been done to address this space. In this section, we draw implications and position this paper with respect to the existing work in the two different research communities.

PC-based information access

Providing IT benefits to blind people has been a very well studied field of research. WebinSitu [10] provides an extensive study of the browsing behavior of blind users and compares it with that of sighted users. The IBM Accessibility Internet Browser for Multimedia (aiBrowser) [11] provides uninterrupted accessibility to multimedia content such as Flash on the websites. However, the benefits of these innovations are limited to blind computer users. Similarly, Home Page Reader [12] enables easy access and navigation of Web content to users who have access to computers.

Voice Diary [13] tries to combine various aids used by the visually impaired such as a Braille slate, a Braille watch, and a talking calculator into a single unit. Although these systems have been very successful in providing accessibility to the Web, they are not relevant to most of the world's blind population.

Mobile-based information access

With the increasing penetration of mobile phone devices, researchers have started to investigate the use of mobile phones to provide services to blind users. Pharos [14] and Drishti [15] are services that use Global Positioning System and a talking map software to provide location-based services to blind users. In FETCH (finding everything using technology convenient and handy) [16], the authors propose the use of cell phones for visually impaired people to find lost items. Handheld audio devices have been used to generate audio books that blind users can use to learn science [17]. The Digital Accessible Information System book format [18] is a standard format to enable navigation of audio books by features such as book skimming, providing jumps to pages, chapters, section headers, and note taking. Although several point applications have been built on the mobile phone, the Telecom Web is a good enabling platform where applications such as VoiceSites can be hosted and thus can have a far-reaching impact on blind users.

Systems for low-literacy users

Since affordability is a major issue in developing countries, particularly in rural areas, several interesting information systems for this community are phone based rather than PC based. Studies from a spoken dialog system for rural India [19] show that users are able to navigate the dialog system irrespective of their literacy levels. A speech-based system for providing health information to low-literate users has been presented in [20]. This paper reports that there is a need to educate the users by creating a video to explain and demonstrate the system to a new user.

Information systems based on low-end mobile devices provide an affordable alternative to PC-based systems. A system for providing information to farmers using mobile phones has been presented in [21]. Kolko et al. [22] conclude that mobile social software holds significant potential as an information tool in rural areas. The work presented in this section leads us to believe that creation and dissemination of locally relevant content using a speech interface through mobile devices can be used as an effective system to fulfill the information needs of the blind population in developing countries. System-level technology development work in this domain has been presented in [9] and [23], where the authors present technologies for content creation and linking. A solution based on these technologies [24] proposes to organize the unorganized urban poor businesses. In this paper, we extend the concepts of the Telecom Web [25] and apply them to provide information solutions to the blind population in developing countries.

Technology background

In this section, we briefly describe our vision of a Web of VoiceSites, which we propose can become the mainstream information Web for the visually impaired in developing regions.

VoiceSites are voice-driven applications that are created by the subscribers and hosted in the telecom network [9]. They are represented by a unique phone number and can be

S: Welcome to the VOIGEN system. Please speak your welcome prompt.
C: My name is Ratan and I work as a hundastani classical vocal teacher 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 my teaching abilities 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 fifty rupees for an hour and one thousand rupees for a month. 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 am available from 3 p.m. to about 6 p.m. all seven days a week.
S: You have specified all the preferences. Should I save them?
C: Yes.
S: Your VoiceSite has been created. Thanks for using the system.

(a)

S: My name is Ratan and I work as a hundastani classical vocal teacher in south extension, malviya nagar and nearby areas.
S: Say 1 if you wish to know my references, say 2 to know my charges, or say 3 to schedule an appointment.
C: One.
S: You can ask about my teaching abilities from Mr. Sanjay, his number is 9484042433. Or you can ask from Mr. Gaurav: 9844740248.
S: What next?
C: Two.
S: I charge fifty rupees for an hour and one thousand rupees for a month. However, the rates really depend on the type of job.
S: What next?
C: Three.
S: I am available from 3 p.m. to about 6 p.m. all seven days a week.
C: Schedule an appointment from 10:00 a.m. today.
S: I am not free that time. Please choose another time slot.
C: Schedule an appointment from 11:00 a.m. today.
S: Appointment is scheduled. Thanks for using my VoiceSite.

(b)

Figure 1

Sample interactions of (a) service provider with the VOIGEN system and (b) a customer with the VoiceSite.

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. Therefore, VoiceSites are analogous to websites in the World Wide Web but can be accessed by dialing a phone number, and information can be listened to rather than being read or seen. The creation of a VoiceSite is made easy by the VoiGen system [9], through which anyone can call the VoiGen system and interact with it through voice. This can enable any illiterate person to create a 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 a VoiceSite by a simple phone interaction with the VoiGen system. A sample interaction is shown in **Figure 1(a)**. By answering simple questions such as the ones shown in this interaction, a blind user can easily create a VoiceSite. A person who needs a classic vocal teacher can access this VoiceSite and get information. Such a simple mechanism can improve the business of the VoiceSite creator. A sample interaction of a user with the created VoiceSite is shown in **Figure 1(b)**. A VoiceSite can link to other VoiceSites through Hyperspeech Transfer Protocol [23]. A network of such interconnected VoiceSites results in a Telecom Web [25], as shown in **Figure 2**.

This WWTW has tremendous implications. The visually impaired do not have to spend months learning how to use a computer and the various accessibility software such as screen readers. It also enables non-PC-literate people to access information and services that were unavailable to them through IT systems.

VoiceSites can be created through a simple voice-driven interface, i.e., VoiGen [9], over a phone call. The ease of creation of VoiceSites enables the subscribers to become information providers, as opposed to being simply information consumers. Several applications of Telecom Web and similar voice-based systems are emerging [25, 26]. Next, we describe a sample VoiceSite that we created for the purpose of studies done in this paper.

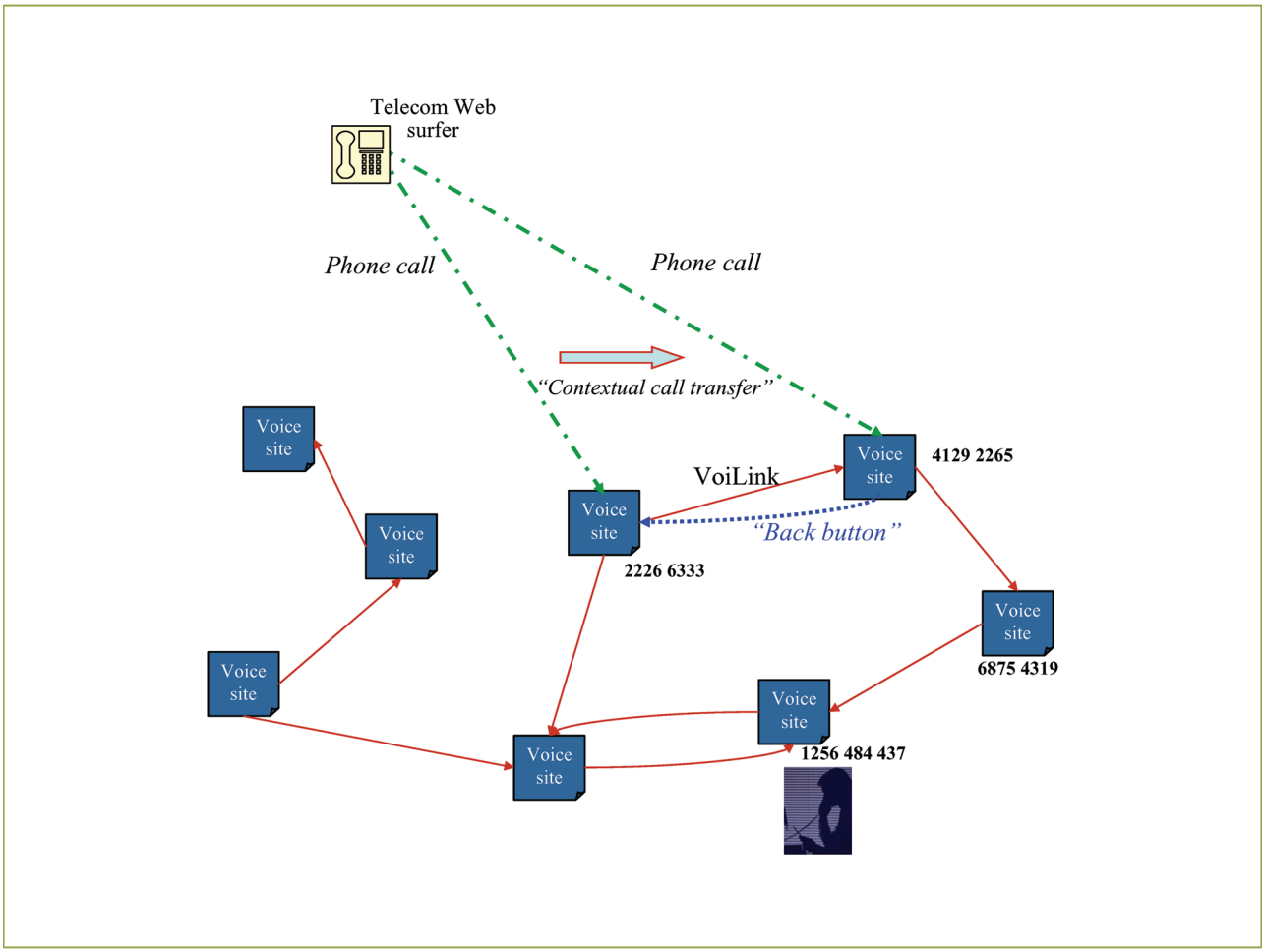

Figure 2

World Wide Telecom Web.

Application scenario

To demonstrate the usability of the concept of VoiceSites, we developed a sample VoiceSite based on the website of the Ministry of Social Justice, Government of India, which enables people to lodge their complaints at the ministry. In current practice, complaints are lodged either through written applications or through the ministry website.

The VoiceSite emulates the Web-based form that is used to lodge complaints. It takes the caller's personal details before asking the caller to record his complaint. After the complaint is recorded, a complaint reference number is given to the caller, which can be used to call the site later and check on the status of the complaint. **Figure 3** shows the structure of the VoiceSite. The VoiceSite does not take all of the data that a Web-based form asks the user for but only takes in the mandatory information. **Figure 4** shows a snapshot of the Web-based grievance registration form.

The VoiceSite records the audio of some inputs such as the caller name and address, as well as the actual grievance. It takes number inputs such as pincode (zip code) and password through the phone keypad. It also performs speech recognition for simple inputs such as the gender and the name of the government department.

Study methodology

For our study, vision impairment was the only criterion used to select the subjects. The study was conducted at two institutes: the National Association for the Blind (NAB), New Delhi, and the Blind Relief Association (BRA), New Delhi. The staff at both institutes quickly grasped the concept of VoiceSites and readily agreed to help us with the user study of our prototype. The subjects were interviewed at the institute locations in the computer laboratories on campus.

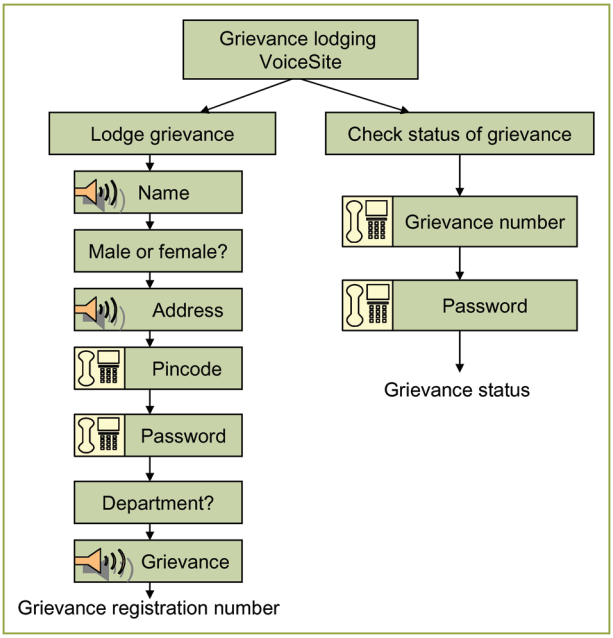

Figure 3

Call flow for the Grievance VoiceSite.

Subject profile

Of the 43 people that were interviewed, five were partially blind, whereas the remaining were completely blind. The mean age of subjects was 25 years, with a standard deviation of eight years. The youngest subject was 12 years of age, and the eldest was 50 years, both male. There were six female subjects, all between 17 and 35 years of age. **Figure 5** shows the age histogram of the subject profile.

A total of 44% of the subjects had completed their college degree. **Figure 6** shows the histogram of the number of years of education of the subjects; 50% of the subjects were students, 22% were unemployed, and the remaining subjects were working in NAB, banks, or other government offices. About 60% of the subjects were familiar with the English language but preferred to interact in Hindi.

The average computer experience of this set was about 2.8 years, with a high standard deviation of 3.15 years. Therefore, the set had people who had no knowledge of computers as well as people who used computers for the last ten years. All computer-literate subjects use the job access with speech (JAWS) [27] screen reading software. Fifty-five percent of the subjects had used the Internet before. However, there were people who had used the computer (with software such as Word** and Excel**) but not the Internet.

Survey process

The usability study procedure was divided into three parts. All of the subjects were completely unknown to us, and similarly, the subjects did not know us. However, during the first part of the usability study, we did introduce ourselves briefly at the start of the process to ease the subjects through the course of the process. To gather the profile of the subjects, we asked demographic questions such as name, education, professional status, age, and their comfort levels in terms of using the PC. Subsequently, we introduced the concept of VoiceSites, and specifically the Grievance VoiceSite, and explained the need for conducting such a user study. Some subjects had several questions about the concept of the VoiceSites. Once we satisfied the subject queries, we proceeded to the second phase of the study.

In the second part of the usability study, we were silent observers, and the subjects were asked to interact with the VoiceSite. The grievances lodged by the subjects were open-ended comments. The application was configured to accept an input of up to 60 seconds. In a few cases, we had to provide some help to subjects when they had problems navigating the VoiceSite. We observed the number of times a user had to repeat a particular step and the total time of interaction with the VoiceSite. Once a subject was able to lodge a grievance, we concluded the study by gathering feedback from the subject through an interview session, which is the third part of the study. We specifically asked the following questions about the usability of the Grievance VoiceSite:

- Was the call flow simple to follow?
- Are you comfortable using the phone keypad?
- Was the audio clear and understandable?

We asked some open-ended questions about the usefulness of the Telecom Web for blind users. We also asked them to think about some of the applications they would be interested in browsing on the Telecom Web.

Although we had a dedicated computer laboratory and a private classroom at the NAB (**Figure 7**) and BRA, respectively, there was a lot of background noise at the NAB location. This noise was due to screen readers on a couple of machines in the room that were being used by students while we were conducting the user study. The subjects used a low-end Nokia 1600 mobile phone, which costs approximately $45 (U.S. dollars), in speaker mode to interact with the VoiceSite. Although the speaker mode was more sensitive to noise, it allowed us to listen to the entire conversation, and we could make detailed observations.

The entire interview process took about 20 minutes for each subject. Of this, we spent approximately 5–10 minutes on the initial introductory session, 4–6 minutes on the application, and 4–6 minutes on seeking the feedback. These interviews were conducted in four days with ten subjects interviewed per day.

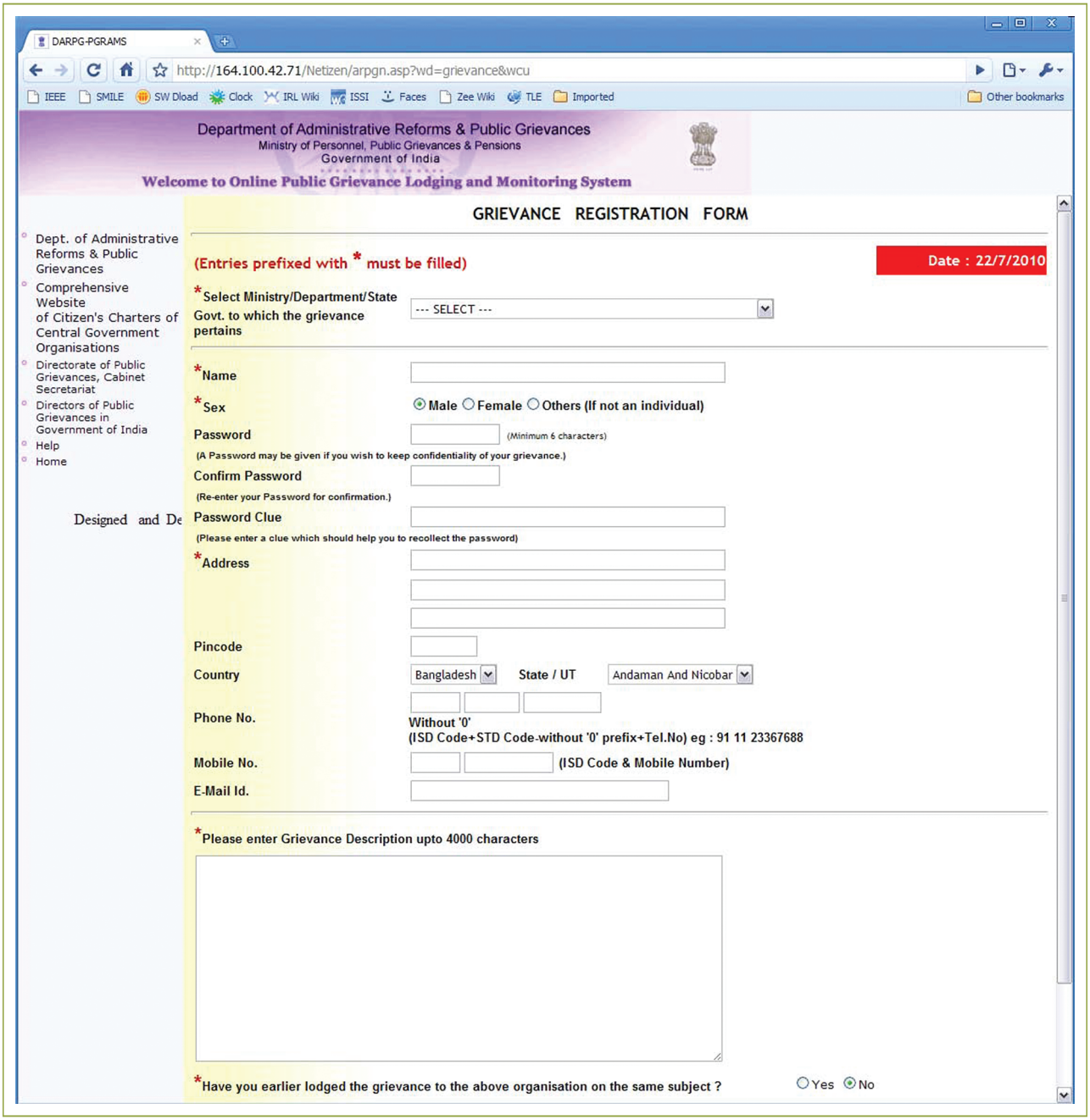

Figure 4

Web-based form for grievance registration.

User study results

In this section, we present the results of the user study, which is based on the Grievance VoiceSite. This section covers results from parts 2 and 3 of the study process, where we observed the subjects interacting with the VoiceSite and interviewed them, respectively. Since the observations and insights from various users were significantly different, we present the results in three separate categories. In the first category, we present the statistical results that were derived from common observations and answers across all subjects. The second category of results has some specific observations that we noted when subjects were interacting with the VoiceSite. The third category of results is interesting and has important anecdotes that some subjects provided during the

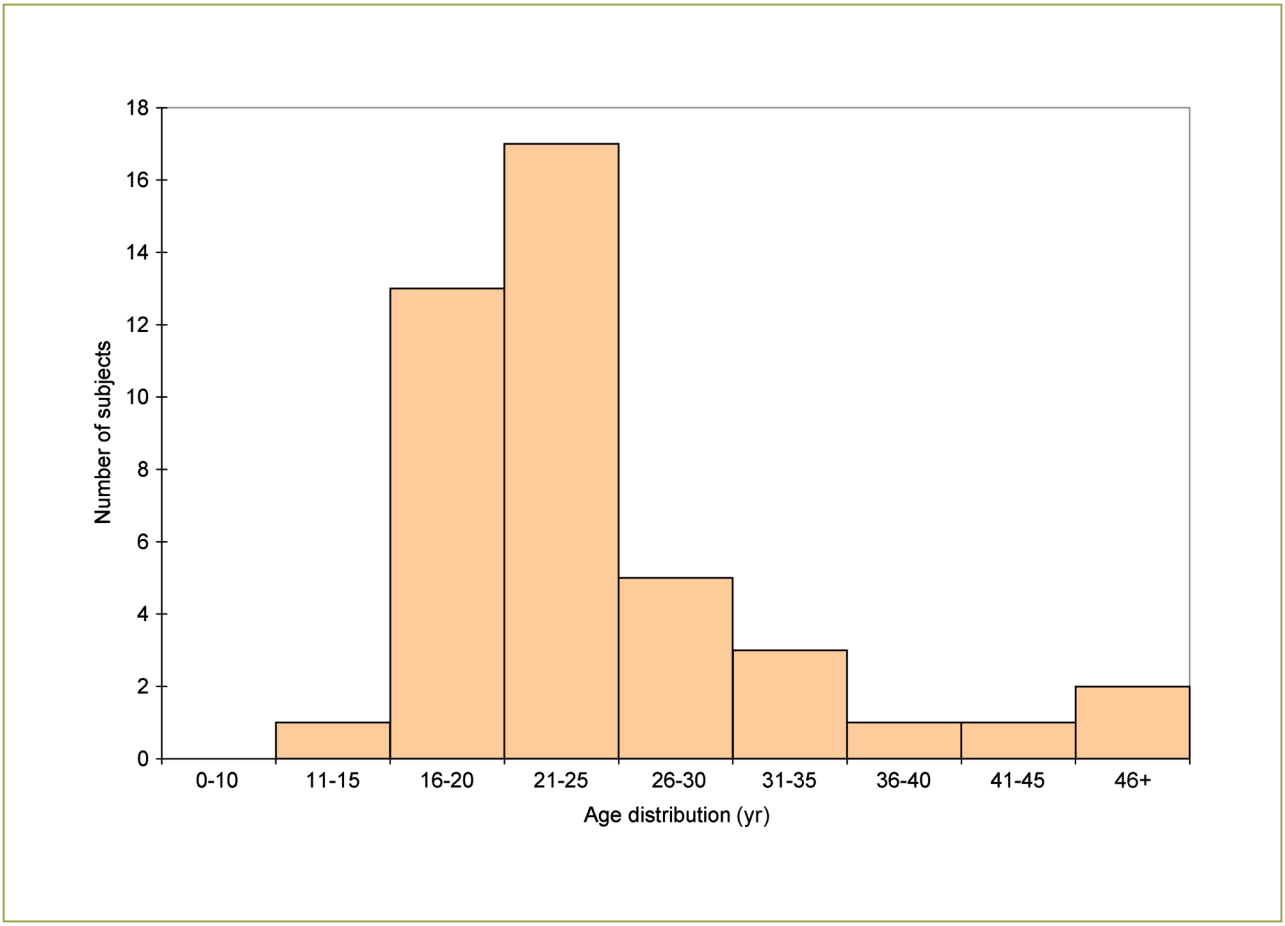

Figure 5

Age histogram of subjects.

interview process. We would like to clarify that the last two categories of results are not statistically driven, since they are influenced by just one subject in some cases. However, we believe that these observations provide valuable insights into the usability and applicability of VoiceSites, and hence, we have included them.

Statistical analysis

The average time for completing the grievance registration task for the subjects was 218 seconds, with a standard deviation of 60 seconds. This time ranged from a minimum of 120 seconds to a maximum of 380 seconds. The graph in **Figure 8** shows the histogram of the time taken by subjects.

An interesting observation is that the correlation between the number of years of computer experience of a subject and his task completion time is very low (0.28). This is reflected in **Figure 9**, which shows that if we sort the subjects in increasing order of the number of years of computer usage, then there is no corresponding decrease in the time that it took to complete the grievance task. Moreover, there were many subjects whose task completion time was high despite that they were experienced in using the computer (e.g., subjects numbered 40 and 41). We consider this lack of correlation as a proof point that using VoiceSites on the Telecom Web is not dependent on the computer skills of people.

This insight is particularly important considering that the ability of a novice user to browse a website requires a course for five months of computer training. Thus, if this computer training is not required for browsing a VoiceSite, this implies that the five months of training are also not required. It is arguable that the five months of computer training enables a user to not only browse but also use other computer applications. However, when we asked the staff at NAB, they mentioned that the five-month training period involves teaching of the interface such as the screen, the mouse, and the keyboard. A significant portion of the time is also involved in explaining the concepts of a screen reader

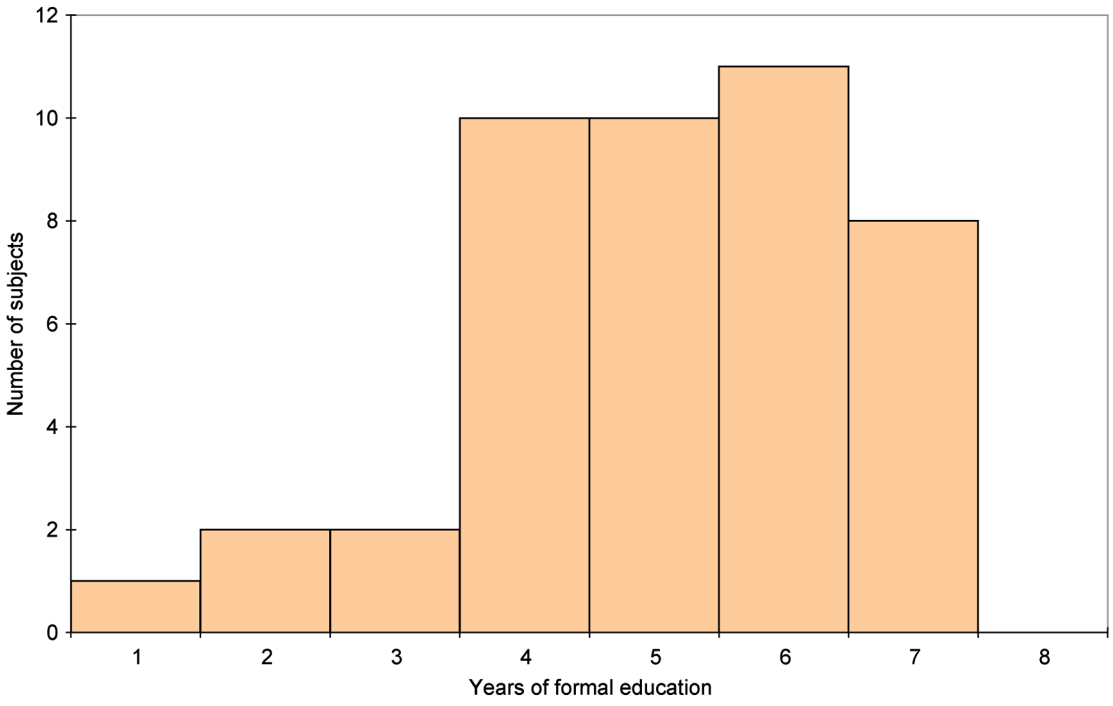

Figure 6

Education histogram of subjects.

and familiarizing the user with the accent of the screen reader. Only then did they teach browsing. Thus, a majority of the five months will anyway be required to teach browsing.

As shown in **Table 1**, all 43 subjects were satisfied with the structure of the VoiceSite, which is a reflection of the sequence of questions that were asked by the VoiceSite. When asked whether they were comfortable in pressing digits on the mobile keypad (to enter zip code and password), 72% of the subjects answered affirmatively. Of the remaining 28% subjects who were not comfortable using the keypad, some mentioned that they found it difficult when using a mobile device other than their own. Nine percent of the subjects were using the mobile phone for the first time, so they did not know how to use the keypad; we had to input the keys for these subjects. Although 72% of subjects said that they were comfortable using the keypad to enter digits, only 39% were able to complete the task without having to repeat the keypad-input step. The Grievance VoiceSite was designed such that it waited 5 seconds to receive six digits from the subject. We identified this as a poor VoiceSite design, since the timeout could have been increased without affecting any other usability aspect of the VoiceSite. If a subject was unable to enter the digits within this time, then we counted this as a repeat attempt. Some subjects were generally comfortable using the keypad but still could not enter the digits in time. They took time to think about the answer, and that reaction time was enough to trigger the timeout.

We also observed the number of interactions each subject had to repeat, either because the system was unable to understand the user voice or because the user was unable to provide the input in time. **Figure 10** shows the histogram of the repeat attempts ("repeats") done by the subjects. A curve (rather than a bar) has been used for visualizing the histogram to highlight the (relatively) long tail of total repeats as compared with speech repeats.

The average number of repeats per subject was 2.7, with a standard deviation of 2. However, all subjects (except one)

Figure 7

Subjects interacting with the VoiceSite at the laboratory at the NAB.

were able to complete the grievance registration task in a single phone call. A total of 11.6% subjects were able to complete the task without having to repeat even a single utterance. The maximum repeats were eight, which was observed in one subject.

Not surprisingly, there was a high correlation[1] between the number of repeats and the task completion time (0.74). However, the correlation of the number of years of computer experience with the task completion time was very low (0.28). **Figure 11** shows that if we order the subjects in increasing order of the number of repeats that were observed during their conversation with the Grievance VoiceSite, then the correlation with task completion time is visible. It is also clear that computer experience is not related to the number of repeats.

[1]All correlation numbers reported in this paper are Pearson product-moment correlation.

Although the average number of repeats is high (2.7), most repeats were due to the fact that the timeout for digit input through the keypad was very low. This can be modified easily in the implemented VoiceSite. If we discard the repeats due to keypad input, then the average number of repeats was only 1.3 per subject, with a standard deviation of 1.2. This number is even more encouraging considering that most of these were conducted in a laboratory where people were working in normal conditions and so there was ample background noise and people movement. In addition, the recognition errors are expected to decrease when the users move from a speaker phone to regular voice mode.

Once the subjects were able to lodge the complaint on the Grievance VoiceSite, we had asked them whether they thought that the concept of VoiceSites and Telecom Web is useful. An overwhelming 90% of the subjects believed this concept to be very useful. The remaining 10% of the

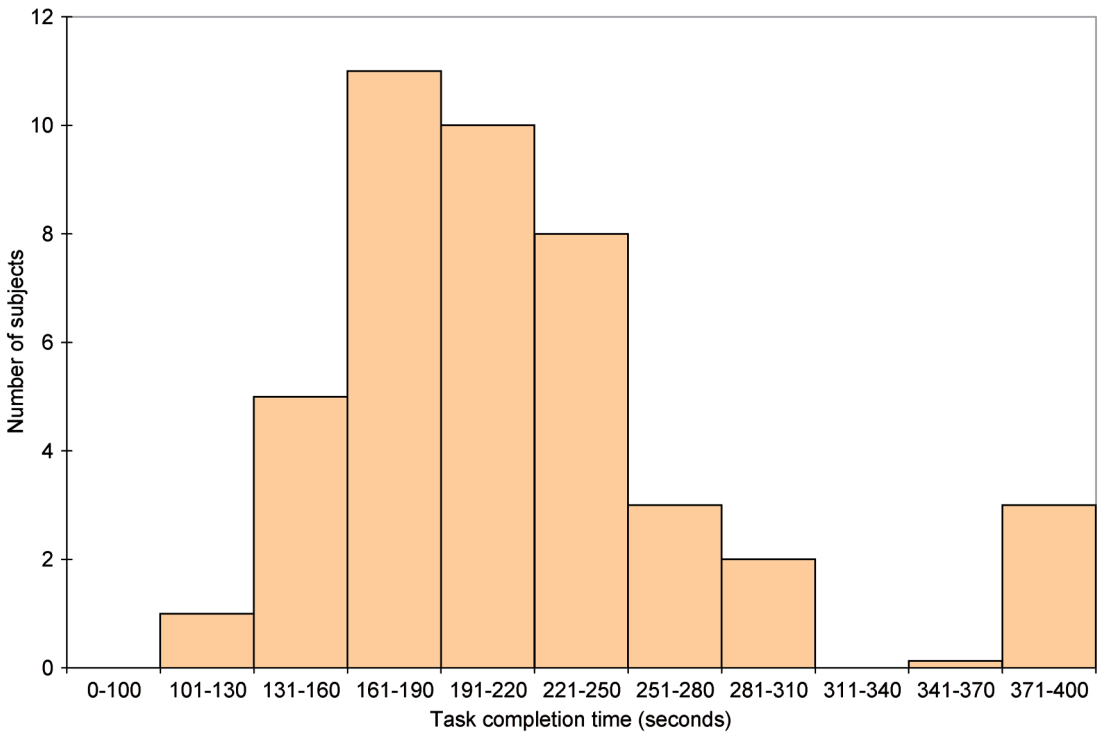

Figure 8

Task completion time histogram.

subjects mentioned that this may be useful for visually impaired people.

We derive the following conclusions from the statistical analysis of the usability study:

1. It was fairly easy to use the VoiceSite since everyone (except one subject) was able to register a complaint.
2. The ease of use of the VoiceSite is not related to the computer experience of the subject.
3. The most difficult task for subjects was to input numbers using the phone keypad, which was due to a bad design (short timeout) in the VoiceSite.

Observations

The statistical details explained above treat each study with a subject as an independent event. However, when we were conducting studies, some subjects were observing while others were interacting with the VoiceSite. There were eight such subjects. Interestingly, these subjects were able to complete the grievance registration task in a shorter time (174 seconds, 13% less than the general average). They made fewer repeats (average of 1.5) than the general subject average of 2.7 (44% less). Thus, the improvement in usability of the VoiceSite by subjects who were one-time observers leads us to claim that learning how to interact with the VoiceSite does not require special training. Users can have improved VoiceSite interaction by observing one user's interaction with it.

The above insight is supported by the fact that computer literacy usually involves training for five months before blind users could start browsing a website, whereas using VoiceSites does not need such training. Hence, Telecom Web is easier to learn than the PC-based Internet for a novice user.

When we compare this learning curve with the time that it takes an average blind person to learn how to use a computer and the Internet, Telecom Web provides a significantly

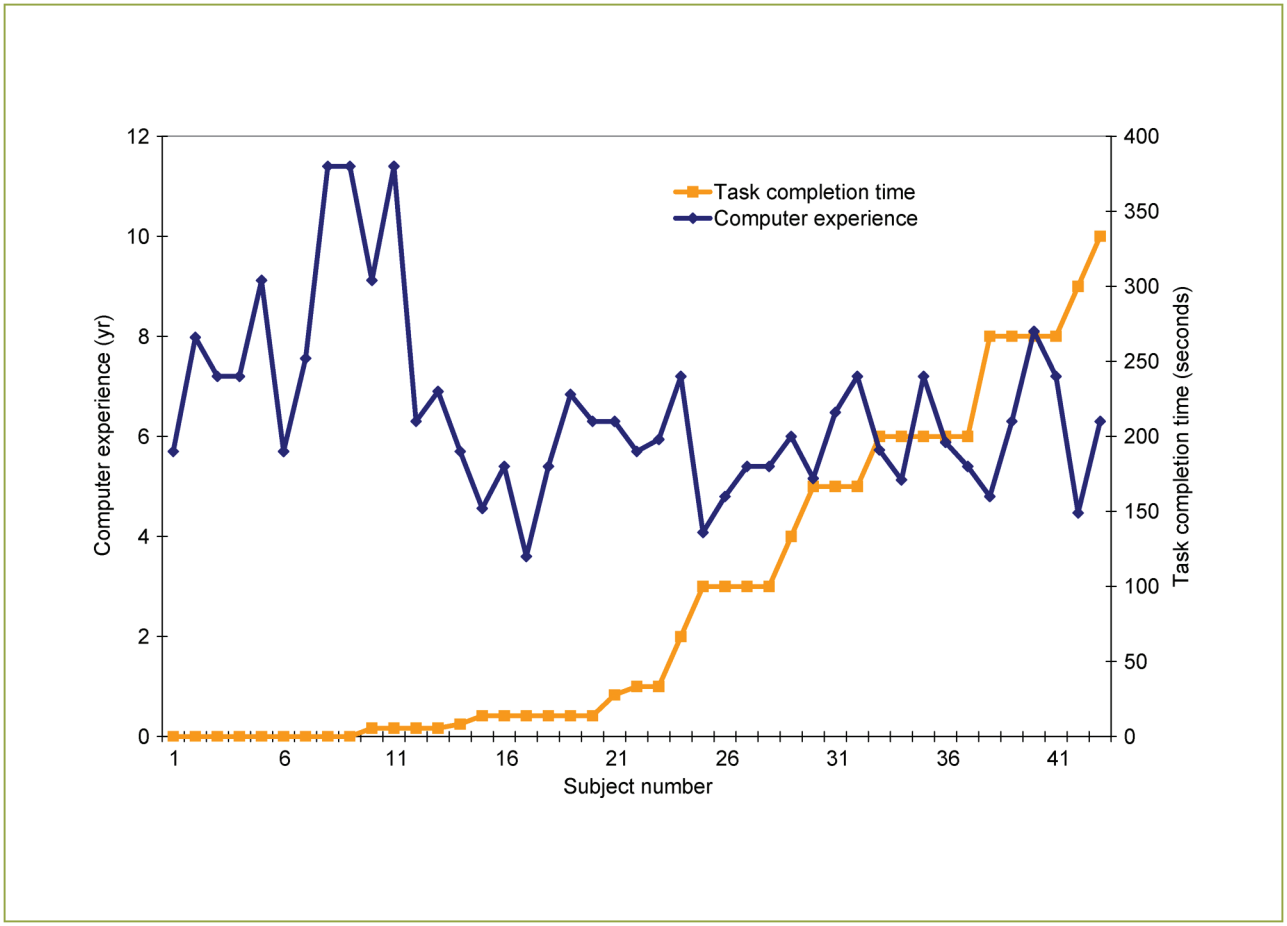

Figure 9

Correlation of computer experience with task completion time.

lower technology barrier. Most computer-literate subjects had taken a five-month course (offered at NAB) before they could operate the computer. Most of them mentioned that they had problems in understanding the synthetic voice that is used in the screen-reader software. This is particularly true for people in developing regions such as India, China, and Africa, where English is not the first language of most speakers, and where the accent of English is significantly different from the U.K. or U.S. accent that is available on screen readers.

Notably, Indian-accented English screen readers are not available in the market. In the initial part of the study, we had designed the VoiceSite such that it used a synthesized English voice in a U.S. accent to interact with the subjects. Of the nine subjects that interacted with the synthesized English voice, three did not understand English, and so we had to prompt them with the translations in Hindi. Five subjects were able to understand the voice but mentioned that they would prefer an Indian-accented English voice.

Based on these comments, we modified the VoiceSite for Hindi language, and this improved the usability significantly. Since creating VoiceSite in a local language is extremely easy, we believe that the literacy requirements will be far less in the Telecom Web world.

We also observed that unlike the case of a Web page where users respond through their keyboard, the VoiceSite has to be responded to through voice or through the phone keypad. To keep the interaction alive, the VoiceSite times out and repeats the question if there is no user input within the timeout limit. Therefore, a VoiceSite expects user input as soon as it provides some information or a question to the user. However, users take some time to understand the question and think of the correct answer. For some users, this time is greater than the VoiceSite timeout for that interaction. In our usability testing, some users had to repeat the interaction because of such timeouts. We observed that due to the different interaction styles, a VoiceSite is more spontaneous than a website. Perhaps this is why many

Table 1 Summary of responses from subjects.

Questions	*Yes*	*No*
Is VoiceSite structure okay?	100%	0%
Comfortable in punching digits?	72%	28%
Had used a mobile earlier?	91%	9%

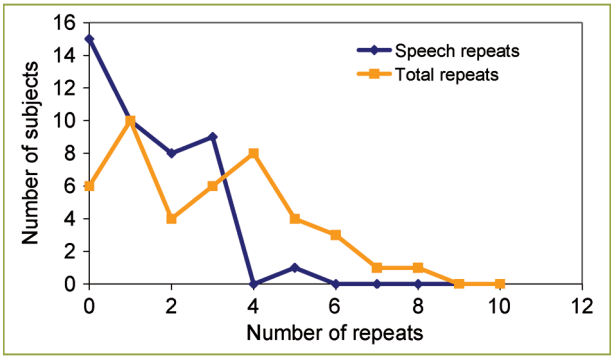

Figure 10

Repeats histogram.

subjects asked for a user-initiated repeat feature so that if they are unable to understand a VoiceSite utterance the first time, they can listen to it again before attempting to answer. Thus, it is important that VoiceSites give a user the option to pause and repeat the interaction at any time.

Finally, we identified the usability of the VoiceSite by subjects with varying literacy in terms of education, mobile usage, and computer usage. For this, we grouped the subjects into four categories: 1) subjects who are completely illiterate, that is, had no experience using mobile phones or computers; 2) subjects who are mobile and computer illiterate; 3) subjects who are mobile literate but computer illiterate; and 4) subjects who are computer and mobile literate. The task completion time for the four categories is shown in **Figure 12**. The height of different bars indicates the average task completion time for each category. The blue line shows the individual task completion time for subjects in that category. As the graph shows, there was significant variation within each category. We had only one subject who was completely illiterate. Since we helped him in punching digits, his task completion time was comparatively lower than that of other categories. However, for the other three categories, it is clear that as the literacy with different devices increases, the task completion time improves.

Anecdotal inputs

The accessibility efforts for the visually impaired mainly focus on how to deliver existing content. Although standards exist for making websites more amenable for screen-reading software, the standards are not necessarily followed by all websites. Moreover, websites are not really designed from an accessibility point of view. Although most software is designed for sighted users, the blind users become fairly adept at using them, barring some difficulties. During our interactions with the subjects, we received many useful and interesting comments about their existing "pain points." Some of the pain points that users mentioned regarding the use of websites and accessibility software are as follows:

"I had studied in a government school, so understanding the foreign-accent English in JAWS was a problem initially. It took me about two months to get used to a computer."
"The GUI should be more accessible. All menus should have alternate keys."
"We cannot work on all computers. We need computers that have JAWS installed on them."
"The structure of each website is different. This is a problem. There are too many different options."
"Screen readers convert everything in accessible format. Pictures are the only problem, not all have alternate text."

Some of the difficulties were related to the process of learning how to use a PC:

"I could never figure out how to type on the computer. I still do not know."
"Windows 98 and 2000 were not as comfortable. Now it is much better."

The subjects gave some specific comments when asked what they thought about the Telecom Web, in addition to unanimously agreeing that the concept is particularly useful for the visually impaired:

"We cannot see everything on screen, but with voice, we know that a VoiceSite is linear so we know that we are not missing any content."
"One or two attempts and you are all set. This is too easy to use."
"There should be a key for repeating whatever was asked."
"This is very useful. Women can lodge a grievance for police complaints."
"This could be a window to the world for us."
"I would prefer to answer through the keypad. Then, we will not have speech recognition errors."
"Speech recognition in crowded places will be an issue."
"A help button should be provided at every interaction. For example, people may not know what is a password, so that help key can tell you more details about that question."

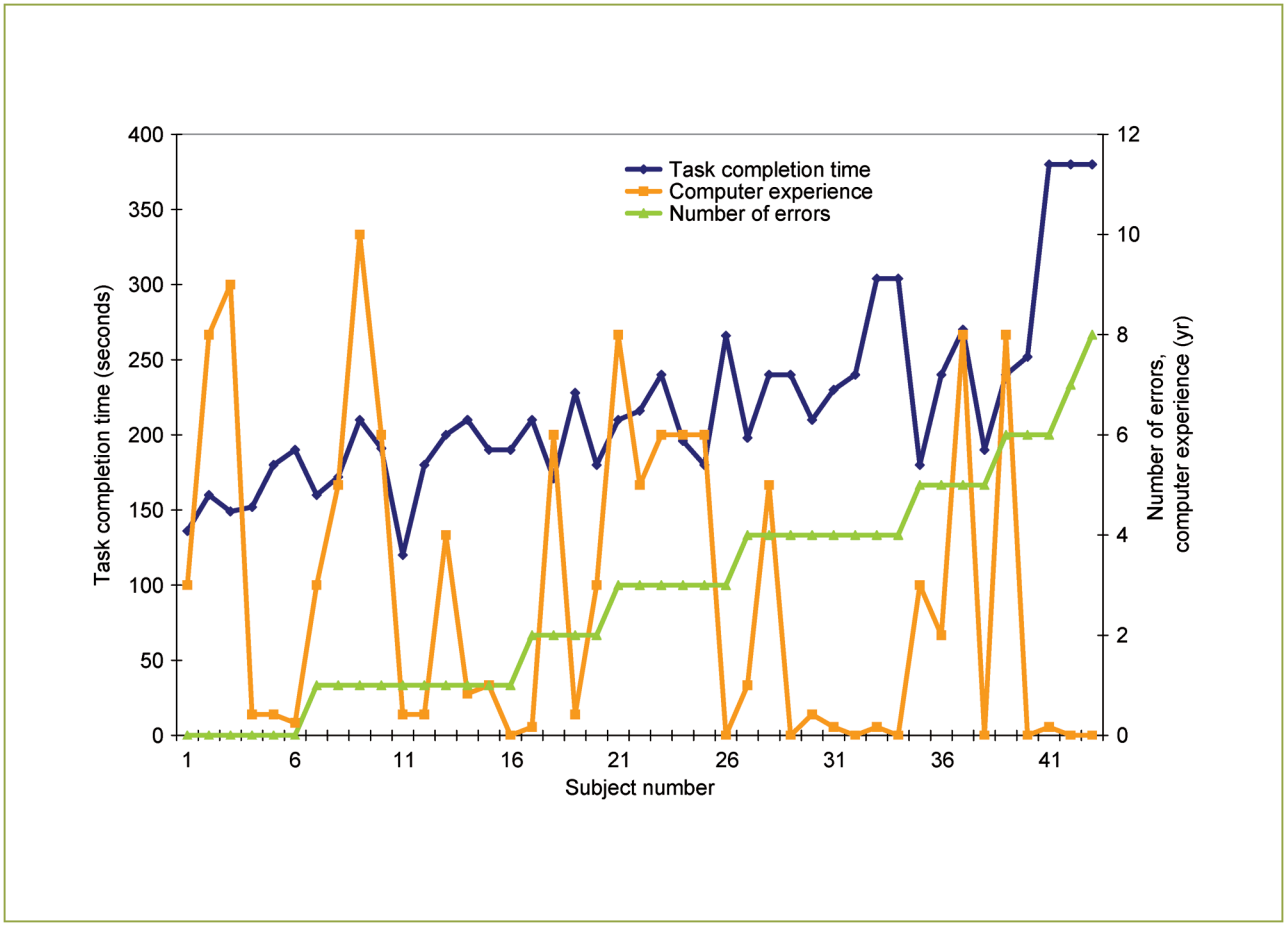

Figure 11

Correlation of repeats, task completion time, and computer experience.

"First time, it was a problem for me; else, I will be able to use it."
"Talking software on mobiles is very expensive; moreover, it is available only for high-end mobiles. Therefore, it is not good."
"I would prefer a language option to choose between Hindi and English."
"When I get to create a VoiceSite, I hope I will be able to edit to ensure that I can overwrite bad prompts."

In our discussions, they also raised other interesting points. For example, most of the subjects use their memory to make note of things. After the study, one of the subjects asked, "How can I remember my password? If I forget it, how can I recall my password." Another subject was an English major and very fond of writing, but he did not have a channel to publish his work or indulge in writing because he does not own a PC. Simple services that are available for sighted users such as reminders, scratchpads, and calendaring are not widely available for the blind. VoiceSites can become that channel to access such services.

Conclusion

In this paper, we have presented the Telecom Web as an alternative to the World Wide Web for delivering information services to visually impaired people. Telecom Web provides a low-cost completely accessible platform, particularly for people in developing countries. We performed a usability study with a sample VoiceSite and derived interesting insights. Based on the user feedback, we conclude that using VoiceSites requires no special training as compared with learning the PC and Internet for interacting with websites. We presented several potential applications that can be delivered to the blind population through the Telecom Web. In the future, we plan to develop some of these applications and study their benefits to the visually impaired. We also plan to investigate user-driven, rather than system-driven, input mechanisms for this population segment.

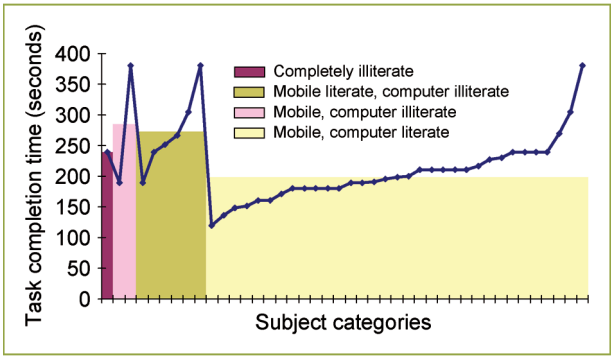

Figure 12

Correlation of repeats, task completion time, and computer experience.

Acknowledgments

The authors would like to thank the staff and students of the National Association for the Blind (NAB), New Delhi (http://www.nabdelhi.org), the Blind Relief Association (BRA), New Delhi (http://www.blindreliefdelhi.org/), who provided relentless support during the user study, and A. U. Joshi for setting up the user study with the NAB and BRA and for providing key statistical details on the visually impaired population.

References

1. *W3C: Web Accessibility Initiative*. [Online]. Available: http://www.w3.org/WAI/
2. *Section 508: The Road to Accessibility*. [Online]. Available: http://www.section508.gov/
3. *Global Initiative for the Prevention of Avoidable Blindness*, 1997, World Health Organization, WHO/PBL/97.61.
4. *Internet Usage Statistics for Africa*. [Online]. Available: http://www.Internetworldstats.com/stats1.htm
5. 2006 World Population Data Sheet, ''Population Reference Bureau,'' Aug. 2006, Technical Report ISSN 0085-8315.
6. R. Wedgeworth, "State of adult literacy," in *Proliteracy Resources*, Sep. 2004. [Online]. Available: http://www.proliteracy.org
7. *Mobile Telephone Subscribers Per 100 Inhabitants*, 1997–2007. [Online]. Available: http://www.itu.int/ITU-D/ict/statistics/ict/graphs/mobile.jpg
8. W. A. Kaplan. (2006, May). Can the ubiquitous power of mobile phones be used to improve health outcomes in developing countries? *Global. Health* [Online]. *2(9)*. Available: http://www.globalizationandhealth.com/content/2/1/9/abstract/
9. A. Kumar, N. Rajput, D. Chakraborty, S. Agarwal, and A. A. Nanavati, "Voiserv: Creation and delivery of converged services through voice for emerging economies," in *Proc. Int. Symp. WoWMoM*, Espoo, Finland, Jun. 2007, pp. 1–8.
10. J. P. Bigham, A. C. Cavender, J. T. Brudvik, J. O. Wobbrock, and R. E. Lander, "Webinsitu: A comparative analysis of blind and sighted browsing behavior," in *Proc. 9th Int. ACM SIGACCESS Conf. Comput. Accessibility ASSETS*, Tempe, AZ, Oct. 2007, pp. 51–58.
11. H. Miyashita, D. Sato, H. Takagi, and C. Asakawa, "Aibrowser for multimedia: introducing multimedia content accessibility for visually impaired users," in *Proc. 9th Int. ACM SIGACCESS Conf. Comput. Accessibility ASSETS*, Tempe, AZ, Oct. 2007, pp. 91–98.
12. C. Asakawa and T. Itoh, "User interface for home page reader," in *Proc. 3rd Int. ACM Conf. Assistive Technol.*, 1998, pp. 149–156.
13. P. V. Assistants, "Voice diary—Description of usability," in *Proc. California State Univ. Northridge Annu. Int. Technol. Persons Disabilities Conf.*, 1999, pp. 149–156. [Online]. Available: http://www.dinf.ne.jp/doc/english/Us_Eu/conf/csun_99/session0003.html
14. A. Marsha, M. May, and M. Saarelainen, "Pharos: Coupling GSM and GPS-talk technologies to provide orientation, navigation and location-based services for the blind," in *Proc. IEEE EMBS Int. Conf. Inf. Technol. Appl. Biomed.*, Sep. 2000, pp. 38–43.
15. A. Helal, S. E. Moore, and B. Ramachandran, "Drishti: An integrated navigation system for visually impaired and disabled," in *Proc. Int. Symp. Wearable Comput.*, Oct. 2001, p. 149.
16. J. A. Kientz, S. N. Patel, A. Z. Tyebkhan, B. Gane, J. Wiley, and G. D. Abowd, "Where's my stuff?: Design and evaluation of a mobile system for locating lost items for the visually impaired," in *Proc. 8th Int. ACM SIGACCESS Conf. Comput. Accessibility ASSETS*, Portland, OR, Oct. 2006, pp. 103–110.
17. J. Sanchez, H. Flores, and M. Senz, "Mobile science learning for the blind," in *Proc. Conf. Human Factors Comput. Syst. CHI*, Apr. 2008, pp. 3201–3206.
18. G. Kerscher and J. Sutton, "Daisy for all: Publishers' collaboration enabling print access," in *Information Technology and Disabilities E-Journal*. [Online]. Available: http://data.daisy.org/publications/docs/pub_print_access/ITD-2003.htm?q=publications/docs/pub_print_access/ITD-2003.htm
19. M. Plauche and M. Prabaker, "Tamil market: A spoken dialog system for rural India," in *Proc. Conf. Human Factors Comput. Syst. CHI*, Montreal, QC, Canada, Apr. 2006, pp. 1619–1624.
20. J. Sherwani, N. Ali, S. Mirza, A. Fatma, Y. Memon, M. Karim, R. Tongia, and R. Rosenfeld, "Healthline: Speech-based access to health information by low-literate users," in *Proc. Int. Conf. Inf. Commun. Technol. Develop.*, Bangalore, India, Dec. 2007. [Online]. Available: http://www.cs.cmu.edu/~jsherwan/pubs/ictd07.pdf
21. R. Veeraraghavan, N. Yasodhar, and K. Toyama, "Warana unwired: Replacing PCs with mobile phones in a rural sugarcane cooperative," in *Proc. Int. Conf. Inf. Commun. Technol. Develop. WWW*, Bangalore, India, Dec. 2007. [Online]. Available: http://itidjournal.org/itid/article/view/327/150
22. B. Kolko, E. Rose, and E. Johnson, "Communication as information-seeking: The case for mobile social software for developing regions," in *Proc. Int. Conf. WWW*, Banff, AB, Canada, May 2007, pp. 863–872.
23. S. Agarwal, D. Chakraborty, A. Kumar, A. A. Nanavati, and N. Rajput, "HSTP: Hyperspeech transfer protocol," in *Proc. ACM Hypertext*, Manchester, U.K., Sep. 2007, pp. 67–76.
24. A. Kumar, N. Rajput, S. Agarwal, D. Chakraborty, and A. A. Nanavati, "Organizing the unorganized—Employing it to empower the under-privileged," in *Proc. World Wide Web*, Beijing, China, Apr. 2008, pp. 935–944.
25. A. Kumar, N. Rajput, D. Chakraborty, S. Agarwal, and A. Nanavati, "WWTW: The world wide telecom web," in *Proc. SIGCOMM Workshop Netw. Syst. Develop. Regions*, Kyoto, Japan, Nov. 2007, p. 7.
26. P. Kotkar, W. Thies, and S. Amarasinghe, "An audio wiki for publishing user-generated content in the developing world," in *Proc. HCI Commun. Int. Develop. Workshop, CHI*, Florence, Italy, Apr. 2008. [Online]. Available: http://groups.csail.mit.edu/commit/papers/08/kotkar-hci08.pdf
27. Freedom Scientific, *JAWS for Windows Screen Reading Software*. [Online]. Available: http://www.freedomscientific.com/products/fs/jaws-product-page.asp

Received November 23, 2009; accepted for publication June 28, 2010

Sheetal K. Agarwal *IBM Research Division, Embassy Golf Links, Bangalore 560071, India (sheetaga@in.ibm.com)*.
Ms. Agarwal is an Advisory Software Engineer at IBM India Research Lab. She received her master's degree in computer science at University of Maryland, Baltimore County. Her research interests include pervasive and ubiquitous computing, mobile applications, and more recently information and communication technologies for emerging economies.

Arun Kumar *IBM Research Division, ISID Campus, Vasant Kunj 110070, India (kkarun@in.ibm.com)*. Mr. Kumar obtained a master's degree in CSE in 1999 and is currently pursuing a Ph.D. degree from IIT Madras, India. He is a Research Staff Member at IBM Research, India. His research interests include ICT for development, service-oriented computing, object-oriented programming, semantic web, and distributed systems. He represents IBM in W3C's Mobile Web for Social Development interest group and is a professional member of the ACM. He is originator and a key contributor to the IBM Spoken Web project. He is currently co-chair of WWW 2011 Warm-up Workshops, serves on various program committees, and has published papers in several reputed international conferences and journals.

Amit Anil Nanavati *IBM Research Division, ISID Campus, Vasant Kunj 110070, India (namit@in.ibm.com)*. Dr. Nanavati received the B.E. degree in computer science from the Maharaja Sayajirao University of Baroda and the M.S. and Ph.D. degrees in computer science (distributed computing) from Louisiana State University, Baton Rouge. Before completing his Ph.D. degree, he spent a summer in the Jet Propulsion Laboratory, California Institute of Technology, NASA. Prior to joining IBM, he was with Netscape Communications. He is currently a Research Staff Member at the IBM India Research Laboratory, New Delhi, India. His research interests include mobile computing and information mining, as applied to telecom service provider settings. He is currently working on solutions for emerging economies. He is particularly interested in applications of graph theory in various domains and also dabbles with speech in mobile and pervasive environments.

Nitendra Rajput *IBM Research Division, ISID Campus, Vasant Kunj 110070, India (rnitendra@in.ibm.com)*. Mr. Rajput has a master's degree in communications from Indian Institute of Technology, Bombay. He is a Research Staff Member at the IBM India Research Laboratory (IRL) in New Delhi since 1998. At IRL, he has been working in the field of conversational systems for the last ten years. He has worked on audiovisual speech-recognition systems for Indian languages and dialog systems. His interests are in statistical signal processing, dialog management, and speech and image processing.