Supporting Voice Content Sharing among Underprivileged People in Urban India

Christian Remy[1], Sheetal Agarwal[2], Arun Kumar[2], and Saurabh Srivastava[2]

[1] University of Zurich, Binzmühlestrasse 14, 8050 Zurich, Switzerland
remy@ifi.uzh.ch
[2] IBM Research-India, Plot 4, Block C, Vasant Kunj, Phase II, New Delhi, India
{sheetaga, kkarun, saurabhsrivastava}@in.ibm.com

Abstract. Recent advances in voice-based telecom information systems enable underprivileged and low-literacy users to access and offer online services without expensive devices or specialized technical knowledge. We propose SRLs (speech resource locators), a mechanism that facilitates the creation, access, and sharing of online voice content. To test the interaction with SRLs, we developed a proof-of-concept application that allows for simple sharing of voice content. We subsequently created a smartphone application for the same service that provided a graphical user interface to the online voice application. Our findings show that literate underprivileged people were able to share online voice content on feature phones and smart phones whereas in low-literacy people were unable to access shared content over feature phones but able to do so on smart phones. We conclude by highlighting opportunities and challenges for the design of voice-based applications that support information sharing.

Keywords: HCI4D, Information Sharing, ICTD, User-Centered Design, Interactive Voice Systems, Smartphones, India.

1 Introduction

In recent years, a variety of voice-based telecom information systems have been created, allowing people in developing regions to access information they would not be able to retrieve otherwise. Among the many barriers to access a few stand out: low computer ownership, low levels of literacy, and lack of Internet access [8]. Interactive voice systems are services that users can call using easily accessible and affordable commercial mobile phones and navigate by speech or dual-tone multi-frequency (DTMF) input in order to retrieve information. Voice-based systems have been proven to be useful and accessible, through research projects [1, 19, 23] as well as large-scale deployments for public use [4, 13, 26]. Examples where these systems have been put into practice include but are not limited to: information for farmers in rural areas [19], education for school children [17], and job opportunities in urban areas

[26]. One of the aspects that most of these applications have in common is the concept of knowledge distribution beyond face-to-face communication. An important issue when developing such applications is the sharing aspect, e.g., if a user wants to notify a friend about specific voice content she came across. We derived the following scenario that shows the need for sharing and the interaction breakdowns in current voice-based systems:

Raj is a migrant from a small village in Punjab who works part-time at a big market in Delhi. He comes across an interactive voice application about health information and posts a query on behalf of his cousin, who has recently been diagnosed with a new skin disease. A doctor records a response to the query, informing Raj of a new drug along with its usage instructions. Raj believes that this could help to improve his cousin's condition. Due to the complicated drug name and usage description, Raj is hesitant to call his cousin and paraphrase the information. Instead he would prefer to point him to the specific information.

There are several barriers illustrated in this example which are typical of voice-based telecom information systems. Currently, such applications do not offer easy ways to access or forward a specific piece of content; simply telling another person about the service results in a difficult navigation task to get to the information. Furthermore, everyday interactions with text content such as searching, browsing, or navigating voice content are still unsolved problems. In this particular example, direct access to a specific piece of voice content is the biggest challenge; this becomes even more complicated for information retrieval of dynamic content, such as user-generated content.

Further, given that pragmatic software for natural speech recognition is not yet available for most languages prevalent in developing countries, searching for audio content is in a nascent stage. Therefore, alternate mechanisms such as sharing content explicitly assume much more importance in the context of voice applications.

In analogy to URLs (uniform resource locators) for accessing Internet content, we therefore propose *SRLs* (speech resource locators) for voice-based systems. An SRL enables users to point to a specific item of voice content in an interactive voice application, thus enabling explicit sharing of online voice content. An SRL consists of a phone number (the voice-based application) and a *LinkCode* (a unique identifier for the online voice content). SRLs are novel as they offer users a way to create persistent links to content, dynamically while interacting with the site. The links are persistent as they can be shared with others even after the call has ended. This ability to link and provide direct access to content is so valued that startup companies[1,2] have created a business out of it. These companies essentially offer 'links' to a service (for instance, extension of a particular department or employee) desired by a caller. Instead of calling the enterprise Interactive Voice Response System (IVR), the caller calls one of these services which automatically traverses the touch tone hierarchy of the target IVR and connects the user directly to the desired option. SRLs are meant to fill the

[1] http://deepdial.com/
[2] http://gethuman.com/

gap in existing techniques by enabling a mechanism for sharing specific online voice content with a set of intended recipients. Furthermore, they enable it both for pre-created and user generated content. The links created by users themselves are usable across applications and users.

In order to test the accessibility, usability, and usefulness of SRLs, we developed a proof-of-concept application and conducted field studies in urban areas of two states in India. We highlight insights from the observations and the informal interviews during the user tests. Based on the findings from the first study, we conducted a second study, in which we evaluated the same interactive voice application with a graphical user interface augmenting the interaction with a smartphone application. Our findings suggest opportunities for the design of new interactive voice applications that allow content sharing, but also challenges that need to be addressed in the future.

We propose that SRLs can simplify the interaction required for knowledge distribution of voice content. Furthermore, we present a study that explores sharing of online voice content through low end mobiles and smartphones. The results of our studies provide empirical evidence that sharing of voice content can be supported by using SRLs. The findings highlight that smartphone applications can improve the user experience for sharing of voice content, in particular for low-literacy populations that were unable to receive SRLs on feature phones.

2 Related Work

By enabling the underprivileged to share voice content, we are providing alternative mechanisms for information retrieval and knowledge distribution among this population. Our background research focuses on two main aspects in previous works: firstly, to what extent sharing of information has been investigated in emerging and developing countries; secondly, what other approaches exist that could lend themselves helpful to achieve sharing and better information distribution for the underprivileged.

2.1 Content Sharing for the Underprivileged

Smyth et al. [25] investigated media sharing on mobile phones and discovered that people in urban India adapt quickly to new interaction techniques if the application addresses their needs while saving them money at the same time. Particularly, if the motivation is high enough, users will overcome smaller interaction issues and find a way to use it even in their everyday practice. However, Densmore [11] pointed out that overly high barriers for sharing result in failure, as in their case relying on Internet connectivity for smartphone applications in a project in Uganda.

Dhir et al. [12] conducted ethnographic studies about information sharing in South Africa, observing that face-to-face communication was by far the most commonly used communication channel. However, especially young people or those involved in businesses express a strong desire for new technology.

Polly [21] is a service designed to facilitate communication among underprivileged users by implementing a service that calls users back, reducing telephony costs signif-

icantly. Another system that has been implemented and evaluated in depth is Message Phone [15], which focuses on exploring the usefulness of voice messages for underprivileged people in Uganda by comparing it to SMS messages. Having similar functionality, Bubbly[3] and Kirusa[4] are two systems that deploy fully functional solutions for voice messaging in developing countries. The proliferation of such voice-based services yields strong evidence for the need of alternative means of sharing, as also pointed out in the introduction.

2.2 Improving Voice-based Information Accessibility

Dhanesha et al. [10] demonstrated bookmarking content in voice-based applications for users and how to re-visit that content. However, sharing of bookmarks is not available. Bookmarking entails assigning an identifier, which must be unique for a caller. The application session information required to reach that point in the application is saved in order to revisit the bookmark. Bookmarking audio content enables the caller to retrieve content of interest to him quickly when he calls the voice application. However bookmarks are typically not shareable and are valid only for the user who created them.

A different concept is to augment existing websites with voice navigation, such as TeleWeb [7], Hearsay [20], or the introduction of voice anchors [22]. This voice anchors allow users to navigate HTML websites with a feature phone by converting information automatically into VoiceXML. By placing a voice anchor via voice input, the user can then go back to the auditory representation of the website later on. Furthermore, the text-to-speech output does not offer the same user experience as voice content recorded by a real person.

To provide search within voice-based systems, Ajmera et al. [2] introduced algorithms that allow for automatic tagging of audio documents in voice-based information systems. Similarly, Srivastava et al. [24] proposed SWAicons that assign auditory cues to improve navigation in voice-based information systems.

These approaches facilitate navigation in voice applications, allow for easier accessibility, and address problems of information retrieval. We present a different concept that enables sharing of any online voice content. While it addresses similar issues as previous projects, it does so by enabling not only direct access for one user to static content, but also allows users to point others to particular items of online voice content, including static and user-generated content.

3 The Billi Mausi Voice Application

We developed a voice-based telephony system that allowed for simple tasks focusing on the sharing mechanics. Our main goal was to implement a system that served as a vehicle to see whether SRLs are useful in lowering barriers to interacting with and

[3] http://bubblemotion.com
[4] http://www.kirusa.com

sharing voice content. We therefore opted for a single-service application that was easy to understand and not tied to any specific use case. In particular, the application let us focus on the three important parts of the sharing aspect: sending, receiving, and accessing SRLs through SMS, without heavy navigation or content-wise explanation in the voice application.

Furthermore, we wanted to attract a wide range of possible users, and therefore chose to use an application that would appeal to many people and was not constrained to a particular domain such as healthcare or education. To serve this purpose, we developed a voice application called *Billi Mausi*[5]. It is a playful service that makes use of the character of a talking cat as the anchor for the voice application[6]. The Billi Mausi character prompts the caller to record her voice, which it then plays back with a modulated pitch (meant to represent a cat's voice). Recent studies conducted by HCI researchers also have shown that fun applications lend themselves well to designing for underprivileged users and encouraging them to participate in on-the-street user tests [17, 25].

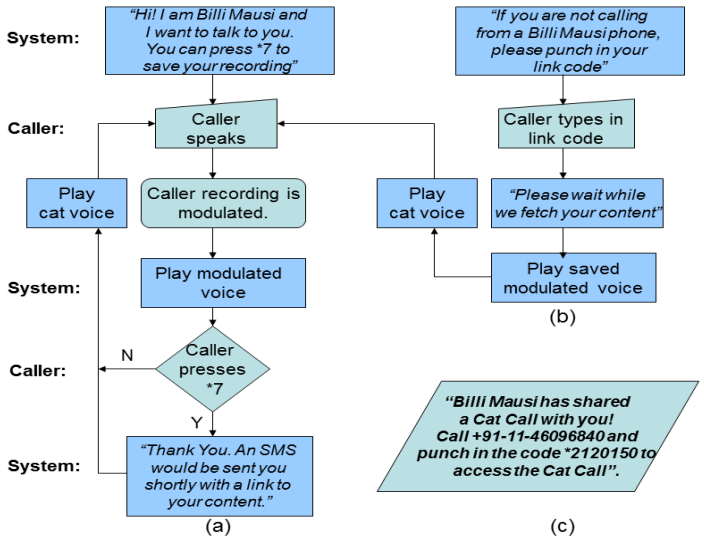

Fig. 1. Interaction tree of sharing voice site content, a) calling the general application and creating a recording b) calling the application to retrieve stored content, c) example SMS as received by our users.

As shown in Figure 1 (a), on receiving a call the Billi Mausi application plays a short introduction (*"Hi, my name is Billi Mausi and I want to talk with you!"* in Hindi). It also informs the caller that by pressing a key combination (*7) on the phone's dial pad, a recording can be stored, and plays a "meow" cat sound as soon as it is ready to

[5] Billi Mausi (Hindi for *cat aunt*) is a well-known fictional representation of cat as a character in many children's stories and poems in India.

[6] A popular service with a similar functionality is Talking Tom, an application available for iPhone and Android.

record user input. Once the input is completed, a different cat sound is played while the server processes the input (which usually takes about a second) and then plays back the modulated voice to the user.

If the caller presses *7 while the modulated voice is played back, or within a few seconds after the playback (we had it configured to three seconds), the voice application (an Apache Tomcat server) registers that as a request to save the recording. At this point, the voice application logic stores the recording, generates an SRL for it, and announces the scheduled delivery of an SMS to the caller. These SRLs allow users to bypass cumbersome and time-consuming voice menu navigation and connect directly to the desired content in an online voice application. An SRL is a pair consisting of a phone number followed by a numeric code (called *LinkCode*) preceded by an asterisk ('*') that can be entered on the voice application as a dual-tone multi-frequency (DTMF) signal. An example SRL would be <+911146096840, *21250> as shown in the sample SMS in Figure 1 (c), that is sent to the caller on successful saving of recording and generation of SRL.

As shown in Figure 1 (b), an SRL can be accessed by simply calling the specified phone number and entering the LinkCode on the phone's keypad, once connected. If the caller is calling from a 'Billi Mausi phone', i.e., an Android smartphone with the Billi Mausi Android application installed, the manual step of punching the LinkCode is not required. Once the LinkCode has been received by the voice application, it fetches and plays the corresponding content and continues with the regular Billi Mausi voice application interaction (i.e., the caller can record and share more such recordings). Thus SRLs not only point to online voice content, they enable callers to continue interacting with the application from that point.

4 Study #1: Sharing with Feature Phones

The user studies were conducted in urban areas of two states of Northern India, namely Uttar Pradesh and New Delhi. At the onset user tests were conducted in the field as well as in the laboratory. Preliminary observations showed that feedback in the lab setting was not as helpful or as successful in eliciting honest participant feedback as field studies. Similar issues have been observed previously in research projects in India [5, 3]. We therefore decided to conduct the remaining user tests entirely in the field to get more realistic feedback and let participants use the system in their everyday environment.

4.1 User Test Protocol

The user test started with an introduction into what our service does, with an emphasis on the sharing part, but building on the funny interaction with Billi Mausi's cat voice to keep participants engaged. We provided them with a short demo of the system and asked them to use it themselves, with a task consisting of four different steps: 1) call the Billi Mausi application, 2) record a sentence and store it by pressing *7, 3) read the incoming SMS, and 4) follow the instructions in the SMS to access the recording.

The tasks were intentionally kept simple to focus on the sharing mechanics only and not be overshadowed by other usability issues or content-related questions.

We closely observed the participants while they used the system and took notes of all interaction breakdowns as well as any comments they made. Once the participant completed the task and listened to the recording, we conducted a short interview, asking for demographic background questions (age, occupation, technical background, as in phone model and Internet experience, if any) as well as feedback on the system. The debriefing interview had two separate goals: firstly, we wanted to ensure that the participants understood the sharing concept by asking clarifying questions and/or letting them explain in their own words what the system does. Secondly, we asked questions designed to elicit: 1) whether they liked the system, 2) if they would use the system and if so what for, 3) what ideas they might have for potential uses, and 4) what overall comments they had.

All user tests were conducted in the participants' native language (Hindi) by researchers who were fully proficient in the area's specific local dialect. There were at least two researchers for each user test: one to conduct the interview and one to observe the interaction and take notes. An average user test took about 20-30 minutes, in three parts of roughly equal length: the first part was about sparking participants' interest, explaining the system, and showing the demo; the second part involved participants completing the tasks; the third part concluded the user test with the debriefing interview.

4.2 User Population

We conducted the studies in urban areas, where we recruited participants from malls, street markets, residential area and their vicinities, and taxi stands. The user set primarily consisted of migrant workers from the unorganized sector, which accounts for 90% of India's workforce [18].

Our study comprised 38 users, of which 33 were male and five female. Participants' ages ranged from 19 to 55 (average 31.2). The five female participants were housemaids in the communities we conducted our studies in, but the atmosphere of these user tests was more comparable to a lab study than to a field study. A group of men approaching a female stranger is not considered culturally appropriate in India and the housemaids acted almost intimidated during the user tests. Therefore, we decided against recruiting further female participants. Our observations in this regard parallel previous experiences with the difficulty with interviewing female users in underprivileged communities (e.g., [3]). Thus, the user population in our studies comprised occupations such as several kinds of drivers (auto rickshaw, cab, and bus), plumbers, carpenters, shop keepers, cleaning staff, and security guards in addition to the aforementioned housemaids. For most of these occupations, the gender distribution is close to 100% male in Delhi.

5 Results

We categorized participants into three rough levels of literacy: illiterate, low-literacy, and literate. In order to estimate their literacy, we asked them to read and paraphrase the SMS that they received during the interview. We also asked for their level of literacy, but often discovered mismatches in which people underestimated or overestimated their ability to read and understand the SMS. In our study, we refer to participants who could read some words and had a basic understanding of written language, but were unable to comprehend the SMS as a whole as "low-literacy." Out of 38 participants, four were illiterate and seven low-literacy (29% of all participants). We acknowledge that this is not a scientific approach to measuring literacy. However, it provided us with useful, albeit approximate, context about our participants without requiring more rigorous and extensive literacy testing that would have made their study participation inconvenient and burdensome.

5.1 Performance Results

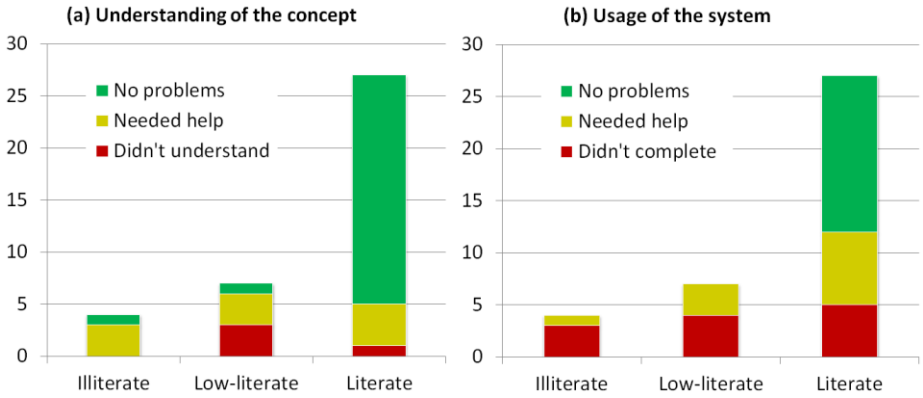

Fig. 2. Participants' understanding of the concept and ability to use the system.

Figure 2b shows the participants' ability to use the system including the SMS reception part, Figure 2a the level of understanding that participants had about the system and its sharing functionality, by level of literacy. The category "didn't complete" encompasses two different reasons: firstly, some participants were not able to use the system – even with our help. Secondly, some participants declined to use the system (e.g., because they felt it was beyond their abilities) after they had patiently listened to the introduction and looked at the demo. They read and correctly paraphrased the SMS and are therefore regarded as literate in our study, but did not complete the user test; e.g., one such participant said: *"I don't want to use it, but you can ask me questions."* Since these users answered all our questions, expressed interest, and provided insightful comments, we decided to continue the interview with them and did not exclude them from the study.

Low-literacy and illiterate participants had significant problems in using the system and understanding its use. Seven of the low-literacy or illiterate participants could not complete the last task (accessing the recording by following the instructions in the SMS) or did not even try since they could not read the SMS. This difficulty was also reflected in the interviews when participants described their issues with the system, which we elaborate on in the next section. Altogether, 26 out of 38 users (68%) completed all tasks, from creating a recording through following the instructions in the SMS to accessing the online voice content; 11 of them with help of the researchers.

In the interviews, participants revealed a much better understanding of the system than their use would suggest. Overall, 24 out of 38 participants were able to correctly paraphrase what the system did and proposed a simple use case that clearly indicated that they understood the sharing concept. Even the majority of low-literacy and illiterate participants were able to abstract the proposed idea, which was probably due to the more elaborate explanation and heavier assistance in using the application, giving them additional hints and examples. This is reflected by the category labeled as "needed help" to understand the system in Figure 2a.

5.2 Usability Issues

We received mixed feedback in our interviews. Although participants liked the system in general and only a few saw no use in it, the interaction was questioned by many. The following comment expresses a frequent response: *"This is a nice idea, but it's so difficult to use… I need a piece of paper to write down the LinkCode, call the number and then enter the LinkCode number… this is too complicated." (Shopkeeper, 26)*

These concerns, however, addressed accessing the voice content and not the sharing concept itself. While the idea of recording voice content was well received, the need to enter the LinkCode was criticized. Our participants had several ideas on how to overcome this issue, e.g.: *"Just let me enter the phone number of a person I want to share this with within the voice application directly, without this SMS." (Auto rickshaw driver, 38)*

Even though this solution is feasible, it breaks down when a user wants to share the content with multiple people. In this case the user must enter the number of each recipient, which not only results in an increased memory load on the user, but it might also be a more expensive solution. Moreover, the probability of entering the numbers incorrectly remains high and there is no way for the user to know if the content was indeed shared with all recipients. Our system allows storing the link to the content and sharing it at a later point in time; this would not be possible with an immediate sharing solution.

5.3 Use Cases and Opportunities for Sharing Content

Despite these issues, participants still saw a number of potential opportunities for them to use the application. Due to the playful context of our user test and the humor-

ous voice in which our application stored and played back the caller's voice, some suggested sharing fun content: *"I'd use it to send jokes to my friends." (Driver, 31)*

Many participants said they mainly use their phone for business purposes, to talk to their employers or to their family. However, the sharing functionality combined with the presented application made them aware of new ways and uses of communication. One participant, a property dealer with a stand in the middle of the street on the outskirts of Delhi, suggested using the application for his business: *"Many clients receive SMSs with information about new available offerings. Using this application, I would send them a recording of my voice, which is much better than SMS, as voice is more personal; but I would only have to record the information once and I could share it with multiple people." (Property dealer, 27)*

This use case would require recording the message in a non-modulated fashion, as it is about serious content rather than entertainment. In fact, many suggested uses were about storing non-humorous content and dealing with real-world problems. Participants not only mentioned that they would use this to store recordings themselves, but also that they would like to receive recordings by others, e.g.: *"If I'm not available, my friends can record something and send me the link to this recording, and I can listen to it later." (Auto rickshaw driver, 30)*

Of course, a service like this suggestion already exists: voicemail. However, it is important to note that many of the participants who described these use cases to us had never heard of voicemail and did not use it. Furthermore, while voicemail is a one-to-one service, online voice content sharing is a one-to-many phenomenon; it enables more than just saving one message for one specific person but aims to provide means of sharing different voice content (user-generated and pre-existing information) with multiple people. Therefore, it goes far beyond what traditional voicemail services can offer and requires a different user interaction.

6 Conclusions of the First Study

The basic idea of SRLs was perceived very well; many participants expressed enthusiasm while using our system, which is in line with the positive feedback. However, low-literacy and illiterate users had difficulties or were unable to access the recording as they could not read the SMS. When forwarding the SMS, users can make use of the contact list stored in their cell phone. However, especially the illiterate participants do not use the contact list on their phones; we oftentimes learned that they had their phone only to make calls to a few numbers they had memorized and mostly received calls from their employer, such as one participant who mentioned: *"I only use my phone so that my boss can call me whenever he needs me. I have no idea how to use any of the functions, and the only number that I can type in and call is the number of my family." (Construction worker, 25)*

This shows that our solution is of limited use for our intended user segment, as there are a large number of illiterate and low-literacy users who would be unable to access voice content via SRLs if they have to comprehend the information from an SMS. Our results showed that accessing the SRLs in this fashion is a solution only for

literate users. Unfortunately, there is no way of creating an SRL that uniformly combines a phone number and a LinkCode into one number. Most feature phones offer a way to make a phone call and enter a DTMF number with a certain delay; however, this is not standardized among all phone models. On some phones the letter "p" is used to add a pause and signal the beginning of a DTMF command, while on other phones the "#" sign is used for the same purpose.

To address this issue, we looked beyond feature phones and extended our application to another platform. During our user tests we noticed that even in our user target segment of low-income people, there were a significant number of smartphone owners (four out of 38). We asked these participants informal questions about their usage behavior and their experience using smartphones. While we did not collect enough data to draw quantitative conclusions or report detailed results, there was enough anecdotal evidence for smartphone usage in our user population to make us look more closely at the smartphone situation in India. Therefore, in the next section we present background information about smartphone usage in India and show why it warrants further attention.

7 Background: Smartphones in India

As recent studies of the mobile phone market highlight[7] [8, 9], underprivileged users are slowly switching to smartphones that are becoming affordable for low-income users in developing areas. Contrary to what one might assume, this does not imply that all these people will have Internet access. Despite strong efforts to increase Internet connectivity and campaigns by mobile network providers[8], many Indians are still hesitant to get mobile data contracts. Similar findings have been reported by other researchers looking at developing or emerging countries of the world, such as in Kenya [27] or South Africa [12].

A recent report states that Internet penetration in rural areas went from 2.6% in 2010 to 4.6% in 2012[9], and our encounters with smartphone owners during our study supported these numbers for urban areas as well. Many providers selling smartphones in India offered one month of complimentary Internet access. One participant said that he did not continue using Internet after this month as he saw no benefit: *"I used Internet for a month when I got my smartphone, but there was nothing interesting for me, so I'm not using it anymore." (Security guard, 24)*

The reasons for the lack of interest are many, with one major issue being the language barrier, as most Indians from low-income backgrounds are not literate in English. One of the participants offered additional support for an observation made earlier by Kam et al. [16], saying that the Internet did not offer him enough content in Hindi: *"Most stuff on the Internet that I came across was in English. There is not much you*

[7] http://www.ingentaconnect.com/content/routledg/ccon/2012/00000026/00000005/art00003

[8] http://sharesmartphone.com/2012/03/

[9] http://www.dnaindia.com/money/report_rural-internet-usage-grows-faster-than-urban_1734825

can do on the Internet on your smartphone if you can't read English, which is why I stopped using [the Internet]." (Cab driver, 25)

Another participant mentioned that he only used the Internet functionality on his phone to download songs, and that this was too expensive for him since he can get songs from his friends by sharing via Bluetooth for free. This practice is particularly common in communities of underprivileged users in urban areas in India: Bluetooth is used frequently to share content among people, as previous studies show [25]. Similarly, Bluetooth or local Wi-Fi connections are used to share or download applications or other data, accounting for the majority of all application downloads to smartphones in India according to one study [8]. The same study mentions other reasons why users do not purchase data contracts, such as fear of hidden data transfers incurring additional costs, incomplete Internet coverage across different areas especially when leaving popular regions, and unreliable and oftentimes slow connections.

We took the increased proliferation of smartphones into account, developing an Android application that augments the interactive voice application. It relies only on calls and SMS to communicate with the voice-based application and does not make use of data connectivity. Previous research projects, such as Claim Mobile [11], have shown that relying on GPRS can lead to failure for HCI4D applications.

8 Design of Smartphone Application

We chose Android as the platform on which to design the smartphone application. The application uses the same voice-based service that was also used in the feature phone study; it calls the same phone number and receives the same SMS. However, the visual interface augments the phone call and intercepts any incoming SMS based on a regular expression. The final interface with which the user tests were carried out can be seen in Figure 3.

Figure 3: Screenshots of the Billi Mausi Android application: a) home screen, b) list of recordings, c) single recording selected, d) screen during the phone call.

On the home screen (Figure 3 a), the application offers buttons to call the Billi Mausi service or access previously stored recordings. During the phone call, participants see

a hint on how to store a recording (Figure 3 b). The application reads the phone's SMS database, populating the list of recordings (Figure 3 c). Instead of the SMS, users see a screen that provides them with a variety of options: listen to the recording, share the recording by entering a phone number, or see whom they shared this recording with already (Figure 3 d). Pressing the "Play" button will initiate a phone call to the voice-based service, but it does not require the user to enter the LinkCode; after a short pause (to make sure that the telephony link has been established), the LinkCode is entered automatically.

Our Android application went through multiple iterations, and preliminary user tests, pilots, and informal focus groups highlighted several issues. In particular, we conducted several pilots with illiterate participants to ensure that the Android application was usable for them even if they had never used a touchscreen phone before. We chose to display as little text as possible but labeled buttons with icons, English text, and Hindi text at the same time. As Hindi symbols are still not officially supported by the Android platform, we added the buttons as images rather than Hindi text in the Android XML files.

For security reasons, telephony access is restricted in the Android SDK: once a call is established, it is decoupled from the application that initiated this call. It is also not generally possible across all devices to access the voice stream, detect DTMF signals, or send DTMF signals during the call. Therefore, we faced many limitations in what the application could do and how to synchronize the call interaction with our Android application. For example, it is currently not possible to implement a "save" button within the application that sends a command similar to *7 to the voice-based application – participants had to manually enter it via the smartphone dial pad. We therefore added a non-modal dialog (an Android *toast*) to display a help message that reminds the user of the *7 command to save a recording (Figure 3 b).

9 Study #2: Sharing with Smartphone

The target audience we aimed for in the second study was the same as in the first study. However, we tried to reach out more for smartphone users from low-income communities to reduce the novelty effect and usability issues due to inexperience. Several issues we observed stemmed from the fact that participants had never used a touchscreen and in particular had no experience with Android interaction – while potential users in the foreseeable future would be smartphone owners and have at least a minimum amount of experience in using it.

Our study comprised 30 participants (different from the first study), who were all male and aged between 19 and 65 (average 32), seven of whom had a smartphone and four of whom used mobile Internet. Typical occupations were similar to the first study, e.g., drivers, shopkeepers, cleaning staff, and security guards. The study protocol was also similar to the one in the first study: we first provided participants with a brief introduction of the study and the system, explained the general purpose, and gave a short demo of the application. We then handed the smartphone over to them and asked them to make a call, record a sentence, and store this recording by pressing

*7. After the recording was done and the call ended, participants would be taken to the screen (Figure 3 d). In contrast to the SMS reception in the feature phone study, participants did not need to input anything to access their recording; simply clicking on the "Play" button would call the voice-based application and enter the LinkCode automatically, playing back the recording without any user input required.

At the end of the user test, we asked participants to enter their own cell phone number and forward the recording to their own phone. Similar to the feature phone study, participants read the SMS and were supposed to access their recording. This task was introduced for three reasons: firstly, it enabled us to compare the user set with the one of the first study, making sure that participants had a similar technical background and level of literacy. Secondly, it gave participants the feeling of sharing in a more realistic fashion. Thirdly, it exposed a larger part of the Android application to the participants, ensuring that the whole interaction was possible for our entire target audience. We asked the same questions as for the feature phone study, but concluded the user test by showing them the one-click play button option on the smartphone – if they had not discovered this by themselves in the user test already.

9.1 Results

After the introduction and demo, all participants were able to use the system successfully with a little help depending on their technical background. No participant was completely unable to use the system and no one refused to use it – to some extent, this was due to the interest that the smartphone sparked among participants. Many users were exposed to smartphones for the first time and therefore we did not encounter as much criticism as in the first study.

The SMS task at the end allowed us to roughly categorize their literacy level in the same way as in the first study. The numbers were similar to those in the first study: three participants were illiterate and five low-literate, a combined 27% of all participants in our smartphone study. However, all participants understood the concept and were able to provide us with examples of use cases or ideas for which they would use the sharing functionality.

9.2 Possible Use Cases

Some of the feedback from our participants was similar to that from the feature phone study. In particular, when asked what they could use the service for, many participants came up with ideas similar to the voicemail-like use that was already mentioned in the first study. However, due to the Android interface even illiterate participants were able to use it and had a different experience in this trial, e.g.: *"I'm completely illiterate and can't write an SMS. But with this service I could send my supervisor a message when I'm on sick leave. This service is a very good service for people like us who don't know how to read or write SMS!" (Office support staff, 40)*

While comments like this suggest a one-to-one communication where content is only shared with one other user, some participants envisioned more complex use. One participant said that sharing something with a group of friends could start a conversa-

tion, and went further to describe how he would use it for threaded discussion: *"I could start a discussion and my friends could add comments to that and point others to their comments and so on… it could help organize this and provide an overview of the discussion, who replied to whom and so on." (Bus driver, 30)*

The idea expressed in the quote above could be further enhanced by taking advantage of the visual properties of smartphones. For example, by using a smartphone in conjunction with voice-based services, the structure and participants in a discussion could be represented visually, thus further enabling use by and providing value to illiterate users.

10 Discussion

Voice-based telephony applications have always been a promising alternative in technology design for low-income, low-literacy populations that do not have access to the Internet. Over time, these applications have become more powerful and versatile, and designers look into new ways to improve the user experience. We believe that the SRLs on smartphones can provide a new, viable means of interaction for underprivileged people. Based on the insights gained from our interviews and observations during our studies, we highlight opportunities for sharing content and how these can be applied to voice-based systems. As we discuss next, new challenges emerge that need to be addressed; in particular when implementing mechanisms for sharing, but also more general considerations as voice-based services become more popular.

10.1 New Opportunities through Sharing Voice Content

Smartphones for the Illiterate. After we limited the use of displayed text in our application to the bare minimum, participants were able to access SRLs due to its augmentation by the smartphone application. This concept allows for a variety of new interaction mechanics for voice-based systems, as it enables illiterate users to share voice content via SMS. Friscira et al. [14] previously highlighted the value of smartphones for low-literacy users as they enabled them to interact with basic amounts of text. Our study extends this work and enables illiterate, low-income users in developing countries to share any kind of voice content by using smartphones as a vehicle to share SRLs.

"Is that really you?" – Trust in Shared Content. Multiple participants in both studies mentioned that they preferred voice calls over texts as voice cannot be faked easily. Participants felt that voice messages are more authentic. One participant said, looking at the SMS after he listened to the recording: *"It tells me that the sender really is the sender – the combination of voice and [the sender's phone number] is really great!"* By sending an SMS that contains an SRL, the recipient can not only listen to the sender's voice, but also verify his or her number. Smartphones can enhance this effect by giving other visual cues, e.g., personalized icons [14] or symbols representing words [15].

Accessibility of Voice Content. One of the big challenges for voice-based telephony applications is the retrieval of desired content from the huge amount of voice content. Techniques such as searching are not very mature for voice content due to limitation of speech recognition engines for local languages and dialects. Alternate mechanisms of information retrieval such as browsing, different ways of navigation, or directly accessing a specific item in such an application assume more importance [2, 7, 10, 20, 22, 24]. As our second study showed, SRLs offered a quick and simple interaction for accessing voice content, regardless of how deep it is located within the navigation tree.

Multimodal Interfaces. The smartphone version of the Billi Mausi application essentially provides a multimodal interface to the users (visual and voice). None of the participants complained about having to deal with multiple modes; rather the multimodality was appreciated. This acceptance opens up opportunities for designing smartphone applications for this user segment that can rely on multiple modes of interaction simultaneously.

10.2 Challenges when Designing Applications for Sharing

Financial Concerns. Among underprivileged people, fear of hidden costs is a major concern when using telephony services [8]. In our application, a short popup message informed the user about SMS charges after they hit the "Share" button, keeping the costs transparent. If an application not only establishes trust in its affordability, but actually *saves* the user money, it may evolve into a thriving service, as participants pointed out, e.g.: "*If this service was free, I would contact a lot of people I usually don't have contact with."* A service does not need to be free in order to be used – if SRLs would be used to facilitate quick access to valuable content such as a health advisory or farming advice, offering a way to save money, this challenge can turn into an advantage. Especially among low-income users in developing countries, financial feasibility can be the difference between success and failure of an application.

Interface Customization for Voice-based Services. Participants sometimes mentioned a desire for customization in several stages of the interaction. Even though these comments were a response to SRLs, they apply to voice-based services in general. One participant mentioned that he would like to change the store command *7 to his *"lucky number"*. This brings up the question how to support customization and maintaining necessary consistency. The platform that our application was built upon used the * key to expose universal commands and 7 was allocated for saving a link to a content resource while other keys had different assignments. This follows established design guidelines for IVRs [6].

Recipient Unknown – Design for Diverse Technology. As smartphones become increasingly popular among underprivileged users, designers of voice-based systems have many options for interaction design, such as using visually augmented SMS or automatically dialed SRLs. However, these opportunities come with a challenge: the recipient's phone is an unknown device. It may be a feature phone, it may be a smartphone, or it may be a smartphone that is not running the same application as the sender. When designing voice applications, different client usage scenarios need to be

accounted for. For instance, in our study, the same voice application worked for users with a feature phone as well as through a multimodal interface via a smartphone.

11 Conclusion

We extend research on interactive voice-based telephony systems by proposing SRLs, a new means of sharing voice content. Our studies show that underprivileged users are able to use the service and understand the sharing functionality. At the same time, they exposed new issues that need to be taken into consideration when developing for sharing of voice content. Furthermore, we only investigated the sharing of user-generated content and looked at first-time users – but our findings suggest opportunities for a variety of different services and further investigations as we venture into a previously unexplored domain. Literacy issues pose one of the most difficult challenges for designing for underprivileged populations, and we provide evidence that smartphones can help to address these issues and even enable illiterate users to take advantage of the benefits SRLs offer, increasing the possibilities of sharing among underprivileged users. Looking back at our scenario in the introduction of this paper, Raj could use the functionality implemented in the Billi Mausi application to request information from a physician, and once he receives the answer, he could forward the information to his cousin by simply forwarding the SMS containing the SRL.

We believe that sharing of voice content can greatly enhance information dissemination among the underprivileged and that SRLs enable new ways of interacting with voice-based services. Therefore, our next steps are to apply SRLs to a variety of services that are more content-driven and reflective of current voice-based information systems in use today than Billi Mausi, which was built for the purpose of studying the basic sharing mechanics. We intend to design applications that lend themselves to long-term deployment studies in order to investigate the full potential of SRLs.

12 Acknowledgements

We would like to thank Elaine M. Huang and Gunnar Harboe for their invaluable comments and input on early drafts of this work. We also thank Meghna Singh, Jyoti Grover, and Monia Puri for helping implement the Billi Mausi application.

References

1. Agarwal, S., Kumar, A., Nanavati, A.A., and Rajput, N. Content creation and dissemination by-and-for users in rural areas. In Proc. ICTD 2009.
2. Ajmera, J., Joshi, A., Mukherjee, S., Rajput, N., Sahay, S., Shrivastava, M., and Shrivastava, K. Two Stream Indexing for Spoken Web Search. In Proc. WWW 2011.
3. Anokwa, Y., Smyth, T., Ramachandran, D., Sherwani, J., Schwartzman, Y., Luk, R., Ho, M., Moraveji, N., and DeRenzi, B. Stories from the Field: Reflections on HCI4D Experiences. Information Technologies & International Development, 5(4), 2012
4. Awaaz De Infosystems Pvt. Ltd. Awaaz.De, http://www.awaaz.de/, 2012.

5. Beaton, J. and Kumar, R. Indian Cultural Effects on User Research Methodologies. In Ext. Abstr. CHI 2010.
6. Blanchard, H.E. and Lewis, S.H. Voice messaging user interface. In D. Gardner-Bonneau (Ed.), Human Factors and Voice Interactive Systems, Kluwer 1999.
7. Borodin, Y., Dausch, G., and Ramakrishnan, I.V. TeleWeb: Accessible service for Web Browsing via Phone. In Proc. W4A 2009.
8. ConsumerLab, Ericsson. Emerging App Culture. An Ericsson Consumer Insight Summary Report, 2012.
9. Convergence Catalyst. India Smartphone Outlook 2012.
10. Danesha, K.A., Rajput, N., and Srivastava, K. User Driven Audio Content Navigation for Spoken Web. In Ext. Abstr. MM 2010.
11. Densmore, M. Claim Mobile: When to Fail a Technology. In Proc. CHI 2012.
12. Dhir, A., Moukadem, I., Jere, N., Kaur, P., Kujala, S., and Ylä-Jääski A. Ethnographic Examination for Studying Information Sharing Practices in Rural South Africa. In Proc. IARIA 2012.
13. Esoko, Esoko – For farmer, http://www.esoko.com/about/index.php, 2012.
14. Friscira, E., Knoche, H., and Huang, J. Getting in touch with text: designing a mobile phone application for illiterate users to harness SMS. In Proc. DEV 2012.
15. Heimerl, L., Honicky, R.J., Brewer, E., and Tapan, P. Message phone: A user study and analysis of asynchronous messaging in rural uganda. In Proc. NSDR 2009.
16. Kam, M., Ramachandran, D., Devanathan, V., Tewari, A., and Canny, J. Localized Iterative Design for Language Learning in Underdeveloped Regions: The PACE Framework. In Proc. CHI 2007.
17. Kumar, A., Tewari, A., Shroff, G., Chittamuru., D., Kam., M., and Canny, J. An Exploratory Study of Unsupervised Mobile Learning in Rural India. In Proc. CHI 2010.
18. National Statistical Commission, Government of India. Report of the Committee on Unorganised Sector Statistics, 2012.
19. Patel, N., Chittamuru, D., Jain, A., Dave, P., and Parikh, T.S. Avaaj Otalo - A Field Study of an Interactive Voice Forum for Small Farmers in Rural India. In Proc. CHI 2010.
20. Ramakrishnan,I.V., Stent, A., and Yang, G. Hearsay: enabling audio browsing in hypertext content. In Proc. WWW 2004.
21. Raza, A.A., Haq, F., Tariq, Z., Pervaiz, M., Razaq, S., Saif, U., and Rosenfeld, R. Job Opportunities through Entertainment: Virally Spread Speech-Based Services for Low-Literate Users. In Proc. CHI 2013.
22. Reddy, H., Annamalai, N., and Gupta, G. Listener-controlled dynamic navigation of voicexml documents. Computers Helping People with Special Needs, vol. 3118 of Lecture Notes in CS, 2004.
23. Sambasivan, N., Weber, J.S., and Cutrell, E. Designing a Phone Broadcasting System for Urban Sex Workers in India. In Proc. CHI 2011.
24. Srivastava, S., Rajput, N., and Mahajan, G. SWAicons: Spoken Web Audio icons: Design, Implications and Evaluation. In Proc. CSCW 2012.
25. Smyth, T.N., Kumar, S., Medhi, I., and Toyama, K. Where There's a Will There's a Way: Mobile Media Sharing in Urban India. In Proc. CHI 2010.
26. White, J., Duggirala, M., Srivastava, S., and Kummamuru, K. Designing a Voice-based Employment Exchange for Rural India. In Proc. ICTD 2012.
27. Wyche, S.P. and Murphy, L.L. "Dead China-Make'' Phones Off the Grid: Investigating and Designing for Mobile Phone Use in Rural Africa. In Proc. DIS 2012.