Evaluation of IVR data collection UIs for untrained rural users

Evaluation of IVR Data Collection UIs for Untrained Rural Users

Adam Lerer
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA, USA
alerer@mit.edu

Molly Ward
Project WET Foundation
Bozeman, MT, USA
molly.ward@projectwet.org

Saman Amarasinghe
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA, USA
saman@csail.mit.edu

ABSTRACT

Due to the rapid spread of mobile phones and coverage in the developing world, mobile phones are being increasingly used as a technology platform for developing-world applications including data collection. In order to reach the vast majority of mobile phone users without access to specialized software, applications must make use of interactive voice response (IVR) UIs. However, it is unclear whether rural users in the developing world can use such UIs without prior training or IVR experience; and if so, what UI design choices improve usability for these target populations.

This paper presents the results of a real-world deployment of an IVR application for collecting feedback from teachers in rural Uganda. Automated IVR data collection calls were delivered to over 150 teachers over a period of several months. Modifications were made to the IVR interface throughout the study period in response to user interviews and recorded transcripts of survey calls. Significant differences in task success rate were observed for different interface designs (from 0% to over 75% success). Notably, most participants were not able to use a touchtone or touchtone-voice hybrid interface without prior training. A set of design recommendations is proposed based on the performance of several tested interface designs.

1. INTRODUCTION

In the past several years, there has been a growing adoption of mobile phone technology as a tool for solving a variety of challenges in international development, including health delivery, disaster management, microbanking, sanitation, and education. This new focus on technology is a result of the explosive growth of mobile phone usage and coverage throughout the developing world. As of 2008, there were 4.1 billion worldwide mobile phone subscribers, compared to just 1.5 billion internet users [26]. 60% of these mobile phone users live in the developing world [25]. Millions of people, many of whom have never used a computer and earn only a couple dollars a day, now own their own mobile phone; this trend is enabling a wide range of potential technological solutions that were not possible a decade ago.

One important use of mobile technology in the developing world is data collection. Collecting data in the developing world presents a number of unique challenges: a diffuse rural population, low literacy and education, and a lack of financial resources. Recently, a number of organizations and projects have successfully used mobile phone and PDA software in place of paper-based methods for data collection [10, 6, 22, 1, 2]. Unfortunately, these existing solutions require access to particular mobile phones running particular software. This presents limitations in every application area: health reporting and advice, disaster reporting, microfinance, and project feedback must all be intermediated by specially trained and equipped 'surveyors', limiting the usefulness and scalability of these services.

Expanding the reach of mobile data collection to all mobile phone users requires the use of either voice or SMS modalities, since these are available on nearly all mobile phones. Of these, only voice is suitable for answering an extended series of questions (although SMS can be used for very simple data collection protocols). Therefore, an interactive voice response (IVR) platform for rendering data collection protocols is the natural choice for expanding the reach of data collection beyond customized smartphones and PDAs.

In addition to expanded reach, voice-based data collection has several additional advantages. First, using voice-based communication circumvents the serious incentive hurdles in more common, SMS-based ICTD programs (e.g. those using FrontlineSMS [3]), since a phone call initiated by the survey application does not incur a cost to the respondent. Second, there is preliminary evidence that data collected over voice in resource-poor areas may be more accurate than data collected by either SMS or custom mobile applications [17]. Finally, studies have shown that data collection through an automated voice system is significantly more effective at obtaining sensitive information than a live interviewer [14, 24].

However, even in the best of circumstances, voice interfaces present usability challenges such as the conventions of spoken language, limitations of speech recognition, limitations of human cognition and working memory, and differences between users [23]. These usability problems are

ACM DEV'10, December 17–18, 2010, London, United Kingdom.

exacerbated by a user population who lacks experience using voice interfaces or even other automated interfaces, and who often have a low level of education and literacy [13].

The investigation of these usability challenges and their solutions in a live IVR UI deployment without user training is the main contribution of this work. Evaluation of several IVR UIs was performed through interviews with volunteers and observation of recorded calls from over 150 survey participants, using a platform we developed for rapid development of IVR data collection applications in the developing world.

Section 2 summarizes related work on voice interfaces and evaluation of these interfaces in the developing world. Section 3 describes our study methodology, the study participants, and details of our IVR platform. Section 4 presents the results of our evaluation of several IVR interfaces. Section 5 discusses the outcomes of the study and proposes several general design principles for IVR interfaces targeted at these users. Section 6 provides concluding remarks and suggests areas of further research.

2. RELATED WORK

There is a large body of work on voice interfaces in the developed world. Commercial interfaces tend to focus on simple task completion, particularly for call center operation. Several authors have provided guidelines for creating usable voice interfaces (e.g. [7, 15, 23]), with many ideas drawn from the field of computer-human interaction, such as the iterative design process, rapid prototyping, and heuristic and user evaluation techniques. However, most existing IVR systems designed for the developed world target neither the needs nor the usability challenges of people in resource- and literacy-poor regions [21].

A number of previous studies have designed and evaluated voice interfaces in the developing world for applications such as health reference [19, 20, 8], microbanking [13], real-time information dissemination [9, 16, 18], citizen reporting [11], and data collection [17]. Berkeley's Tamil Market project [18] was the first speech interface that was rigorously evaluated with low-literacy users in the developing world. Developers performed user studies and interviews and recorded target users. The study suggests that there are differences in task success between literate and illiterate users, but the sample sizes were too small to be conclusive.

Subsequent studies have evaluated IVR UI designs for illiterate or semi-literate users in the developing world. In particular, several studies have compared voice and touchtone input, with mixed results. Patel et al. found that subjects in Gujarat with less than an eighth grade education performed significantly better using touchtone input than speech recognition [16], and the OpenPhone team also found that 'most of the low literacy caregivers...preferred the touchtone system' [12]. Sherwani et al., however, found that task success in a speech interface was significantly higher than a touchtone interface [20], and Grover et al. reported similar user performance for a touchtone and key-press replacement voice interface. These conflicting results show that even basic IVR UI choices are highly context-dependent and require careful consideration and study.

Recent studies have also compared different types of mobile UIs for developing world users. One study involving health workers in Gujarat compared data collection accuracy using a mobile phone electronic forms interface, an SMS data encoding scheme, and transcription via a live voice operator [17]. Live operator transciption was found to be an order of magnitude more accurate than electronic forms or SMS. In a similar comparison study, the ability of low-literacy users to complete banking transactions was evaluated on a text UI, a 'rich' audiovisual UI, and a voice UI [13]. Task completion rates were highest for the rich UI, but the voice UI was faster and more accurate. Users were hesitant to press buttons on the phone in the rich UI and preferred a voice interaction, but they were confused in the voice UI by the inflexible system responses. These studies suggest that voice-based interactions may be preferable for users in the developing world, if they can imitate human interactions sufficiently.

Almost all previous IVR evaluations have provided training to participants, and several have cited effective training as crucial for task success with speech interfaces. Sherwani et al. and Patnaik et al. had participants complete a series of instructor-guided practice tasks to learn the interface [20, 17]. Patel et al. and Medhi et al. provided participants with a verbal explanation of the system before evaluation [16, 13]. Grover et al. had each user 'watch a five minute video showing how a caregiver could use the system' [8]. Only the Tamil Market project reported user success without training on their Wizard-of-Oz IVR prototype [18]. They report that even inexperienced users were successful because they 'provided information even when no input [was] given'. This strategy is not suitable for data collection applications.

In contrast with previous work, our work examines what IVR interactions are possible *without* training, in a real-world environment. The most promising attributes of IVR applications in the developing world are their reach and scalability, which are hampered by a dependence on prior user training. Survey/census applications such as ours would be rendered pointless by a dependence on prior live training, and citizen applications such as health information, agricultural information, and citizen reporting would ideally be spread virally without a need for individual training of each user. Furthermore, we do not predict a rise in IVR-savvy in the developing world obviating the current need for IVR training; if anything, IVR-savvy will only improve *after* IVR interfaces are developed that can be used without training and thus reach untrained users. Therefore, we have attempted to elucidate if and how IVR interfaces can be used successfully without prior training.

3. METHODS

3.1 Deployment

The Project WET Foundation is a non-profit organization whose mission is to "reach children, parents, educators, and communities of the world with water education." [5]. Project WET conducted a teacher training program throughout rural Northern Uganda in July and August 2009. Teachers were trained and given materials to teach students proper sanitation and personal hygiene practices.

The Project WET organizers were interested in obtaining feedback from participating teachers about if and how they had used the Project WET materials in their schools and communities. The teachers were located throughout rural Uganda and were difficult to reach in person, but approximately 250 of the teachers provided mobile phone numbers at which they could be reached. Project WET originally planned to collect feedback with an SMS campaign, but did

not want teachers to have to pay for SMS usage to provide feedback. Calling teachers with an automated voice survey circumvented this problem because mobile phone users are not charged for received calls in Uganda (and most other countries). Furthermore, a voice survey could collect more detailed information than could be sent in a 160-character SMS message.

The purpose of the survey was to collect information from teachers about their use of the Project WET materials and training. The survey asked whether and how the Project WET materials had been used, and what results had been observed from using the materials. Teacher names were also recorded to verify that the correct user had been contacted. The survey was delivered in English, which was spoken by all of the participating teachers.

Figure 1: A Project WET teacher training in Uganda. Photo courtesy of Project WET Foundation.

Calls were scheduled between 10AM and 7PM local time, with unanswered calls being retried up to 4 times in intervals of 2 hours. Each survey call was preceded by the following text message sent 24 hours in advance of the call.

> Hello! This is Project WET in the USA. Please expect a recorded survey call on [Thursday]. Help us by answering the questions. Thank you!

The first version of the survey was designed and tested by us and members of the Project WET teams. Feedback was then solicited from volunteer testers in Uganda - the supervisors of the teachers being surveyed - and used to improve the UI. Finally, calls were delivered to teachers, and several additional UI iterations were performed based on listening to recordings of those calls. The users for each UI iteration were completely non-overlapping.

Call recordings were all listened to in their entirety, and call outcomes were classified into one of the following categories: success, partial success[1], user failure (i.e. interface failure), early hangup[2], environmental factors and call quality[3], and wrong person. Calls that were never answered were excluded from the count.

[1]Calls in which some of the questions were answered correctly.

[2]Calls in which the user hung up near the beginning of the survey instructions. It is not clear why the users hung up in these cases, but many were likely unavailable to take the call, since calls were delivered during working hours.

3.2 Ethnography

Since this work is based on a real-world deployment necessitated by the difficulty of collecting data, we cannot have precise ethnographic data for our participants. The following information was provided by a Project WET Coordinator in Uganda who works regularly with the Ugandan teachers.

The teachers are from mostly rural schools in central and northern Uganda. The majority of teachers have English as a second language. "Some teachers speak good English but may not write good English"; likewise, "they may understand...English spoken by the people from their locality but find it difficult to get the accent from other parts of the world." On the whole, primary teachers have a moderate understanding of English.

Teachers undergo eleven years of education; some may have gone to college or technical training. Most teachers have used computers during their schooling, but most do not have computers of their own. Almost all teachers, however, own a mobile phone. The majority of teachers have never used an IVR interface before; even voicemail is not common among these users.

Approximately 85% of surveyed teachers were male.

3.3 IVR Technology

The IVR survey application was built using an IVR data collection platform called ODK Voice that we have developed for use by organizations in the developing world. ODK Voice was designed as both a platform for creating IVR data collection interfaces suited to the needs of developing-world organizations, and as a prototyping tool for IVR interfaces.

ODK Voice was developed as part of the open source Open Data Kit (ODK) mobile data collection project [4]. ODK is a set of open-standards-based mobile data collection applications for the developing world, centred around the OpenROSA/XForms form specification language. ODK Voice is therefore able to operate with the same forms used by other OpenROSA mobile data collection applications, and integrates with existing XForms data aggregation and analysis tools such as ODK Aggregate.

ODK Voice allowed us to iterate and evaluate fully-functional IVR data collection applications very rapdily, because it achieved a separation of concerns between protocol specification and rendering. Data collection protocols were specified generically with the XForms specification language, and rendering of each question was handled automatically by ODK Voice based on question type. A number of complex protocol features could be encoded as part of the XForm with no changes to the IVR software, including multi-lingual support, a variety of touchtone input types and audio recording, branching, constraints, and other form logic. Customizations specific to the IVR rendering of a form could be achieved without software modifications or changes to the underlying XForm behavior using an extensible set of rendering attributes specific to ODK Voice.

As a result of this design, we were able to accomplish most of our UI modifications by simply modifying the pro-

[3]Calls that failed because the connection was extremely poor or intermittent, or in which the user said 'I'm busy, call me back later' before hanging up.

tocol XForm. In cases where a new IVR rendering feature did have to be added, this rendering feature was then externalized as an XForm attribute so that it could be 'tweaked' from within the XForm. ODK Voice also automatically determined the set of recorded prompts necessary to render a new or modified form and guided recording of these prompts over the phone.

To support our deployment, an outbound call scheduling system was incorporated into ODK Voice. This system automated scheduling of outbound calls in certain time windows and automatic retry for unanswered calls.

ODK Voice can be hosted on any server and can connect to regional cell networks through either a voice-over-IP (VoIP) provider or a hardware gateway. ODK Voice specifies voice dialogues using VoiceXML, a W3C standard with many client implementations. The only operating expenses for an ODK Voice instance are for the server and cellular network usage charges. Figure 2 illustrates the choices of voice and data infrastructure that can be used with ODK Voice.

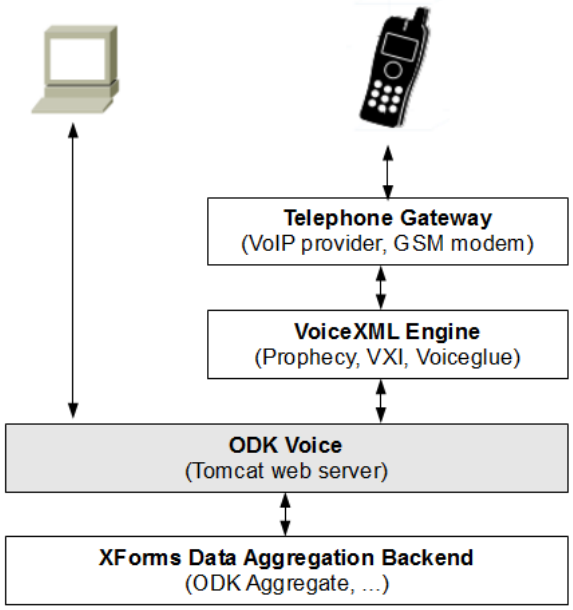

Figure 2: A diagram of the hardware/software infrastructure for ODK Voice. ODK Voice uses VoiceXML to specify the audio dialogues that are played by a VoiceXML engine and transmitted via a telephone gateway to the cellular network. Collected data is sent to an XForms backend for viewing.

4. RESULTS

The first version ***V0*** of the Project WET survey UI contained 4 questions including touchtone-based multiple-choice, numeric, and recorded audio answer formats. The survey was tested with a small group of volunteers in Uganda immediately followed by interviews. These calls and interviews led to a number of qualitative conclusions. First, users who didn't receive a text message 'warning' in advance were completely unable to make sense of the survey call when it arrived. Furthermore, even with a text message, users tried to 'get the speakers attention' when a call was received by saying things like 'Hello? Hello?', missing the initial instructions. A 2-second 'chime' sound effect was added at the beginning of calls to encourage users to listen, followed immediately by 'This is a recorded call from Project WET. You are not talking to a real person.' Based on our initial tests, we estimate the success rate without text message or chime to be close to 0%.

Users reported that the most difficult part of the interface was using the keypad during a phone call; they said it would be much better if they didn't have to use the keys. Users also had difficulty understanding when to speak and when to use the keypad. We modified the instructions to clearly tell users to either 'Press the 1 button on your phone to ...' or 'Please *say* your answer after the beep'.

This improved survey ***V1***, which used both touchtone and voice input, was delivered to 20 participants. The success rate was 8%, with an additional 23% answering some questions correctly. 55% of participants failed to succeed at even the first input task - pressing 1 to begin the survey - even when the instructions explicitly said to 'press the 1 button to begin the survey'[4]. Based on these results, we chose to switch to an entirely voice-based UI.

The voice-based survey ***V2*** contained 3 recorded audio questions that attempted to capture information similar to the previous version. Of 70 participants who received this version of the survey, the overall complete and partial success rates are 19% and 11% (30% total). If we exclude the calls that failed due to factors external to the interface (hang-ups, environmental factors, wrong person), the complete and partial success rates are 33% and 20% (52% total). This success rate is several times higher than that of ***V1***, and we saw a dramatic qualitative improvement in user performance with this interface.

The voice-only UI was then redesigned based on observations of recorded calls from ***V2*** to produce ***V3***. The initial instructions and the question prompts were reduced in length, and the question prompts were rewritten to be conversational rather than instructional, with a focus on turn-taking conventions. For example, ***V2*** contained the following prompt: 'After you hear the beep, please say your name and the name of the school where you work. When you stop talking, the survey will continue. [beep]' which was replaced in ***V3*** by 'What is your name? [beep]'. ***V3*** was tested with only 9 users with a 37% success and 50% partial success rate (87% total) excluding external factors.

Finally, survey ***V4*** was identical to ***V3*** except that the prompts were recorded by a native Ugandan who spoke with a Ugandan accent and dialect. Of 49 participants who received this version of the survey, the overall complete and partial success rates were 63% and 18% (82% total). Excluding external factors, the complete and partial success rates were 78% and 23% (100% total). No participants fell into the 'user failure' category; every user answered at least 2 of 3 questions satisfactorily.

Survey ***V4*** had significantly higher success and success / partial-success rates than both ***V1*** and ***V2*** ($p < 10^{-4}$). The improvement from ***V1*** to ***V2*** was not statistically significant - ***V1*** was abandoned after a small sample size -

[4]These results may be overly pessimistic since the prompts were not recorded by a native speaker. Nonetheless, these results measure poorly even against the first voice-only UI, which was not recorded by a native speaker.

OV: [Intro Music] This is a recorded call from Project WET. You are not talking to a real person. This call will record your answers to three questions about your Project WET training. After each question, you will hear this sound: [beep]. After this sound, say your answer. When you are finished, stop talking and wait for the next question.
OV: Question 1: What is your name?
User: My name is Emuru Richard.
OV: Thank you. Question 2: How have you used the Project WET materials since the training?
User: Since the training, we divided the Project WET materials to all schools. They are displayed in the schools and they are used for reading and for practicing in the schools.
OV: Thank you. Question 3: What results or changes in student behavior have you noticed after using the Project WET materials?
User: More students are now cleaning their hands before eating and after eating. And they now know ...

Call 1: A sample call from ***V4***.

but we observed a dramatic qualitative improvement in user performance between these two versions.

Table 1 provides a complete breakdown of call outcomes. Figure 3 compares task success rates for the three UIs quantitatively evaluated.

	V0	V1	V2	V3	V4
Success	0	1	13	3	31
Partial Success	1	3	8	4	9
User Failure	5-10	9	19	1	0
Early Hangup		4	22	1	4
Call Quality		1	6	0	3
Wrong Person		0	2	0	2
Total	**5-10**	**20**	**70**	**9**	**49**
Not Answered		1	17	5	23

Table 1: Call outcomes for Project WET survey by interface.

Women had greater task success than men. In the first voice interface, men were at least partially successful 45% of the time; women were at least partially successful 85% of the time[5]. The gender discrepancy in success rate in a Fisher's exact test ($p = 0.06$ one-tailed; $p = 0.09$ two-tailed) did not meet the standard significance criterion ($p < 0.05$). Results are shown in Table 2.

	Success / Partial Success	User Failure
Male	14	18
Female	5	1

Table 2: Call success rate for Project WET survey by gender.

We make no claims on the external validity of the survey methodology. Reporting bias could have been introduced if the participating teachers felt that it was in their interest to report positive results, particularly since the survey was not

[5]We are only considering V2, because the sample size is too small in V0, V1 and V3, and there are no user failures in V4.

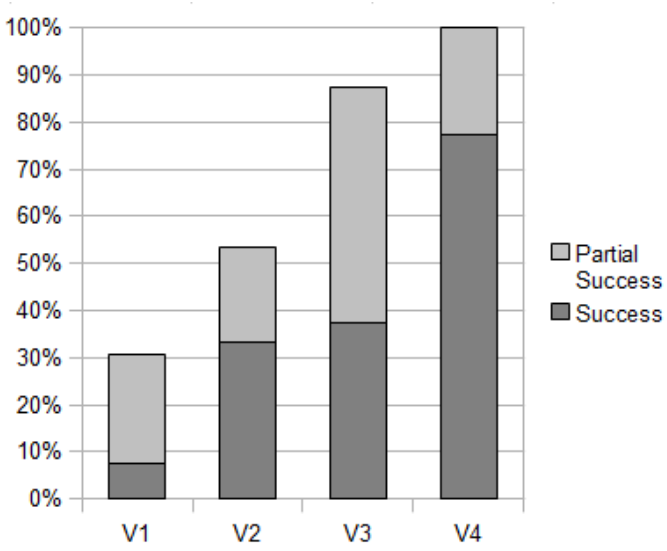

Figure 3: Call success rate for Project WET survey by interface version.

anonymous, since Project WET provided training, materials and funding. Positive results were in fact reported from nearly all of the respondents. Most participants reported that the materials had been 'rolled out' to students and other teachers, and that students had begun to wash their hands properly, clean water containers, etc.

The aggregate conclusions of the phone survey *were* at least in accordance with those observed directly. In approximately 40 of the surveyed schools, written feedback from teachers was elicited and direct observations of Project WET implementation and results was performed. All but one of the schools visited had been using the Project WET materials, and major changes such as handwashing were observed.

5. DISCUSSION

This work demonstrates a variation in task success rate - from practically 0% initially to 75-100% in the final version - as a result of UI design choices. Qualitative and quantitative analysis provides several insights and suggests a number of UI design principles for IVR applications targeted at a particular class of users; namely, users in the developing world without prior IVR experience or training, with real-world connectivity problems and distractions, and for whom English is not a first language.

Comparison of Touchtone and Recorded Voice Interactions

The most serious usability problems with the initial Project WET survey involved understanding how and when to use the touchtone keys. In our initial interviews with participants in Uganda, we received feedback such as "It was very good, but the buttons were very hard. It would be better if you could get rid of the buttons", and "Pressing the buttons did not work for me." Many participants did not press any keys or did so only with significant prompting, and most participants who did press keys made a number of mistakes throughout the interaction. This observation is in line with Medhi et al., who found that subjects responded well to

an ASR-based voice UI prototype but "required significant prompting and encouragement [from the experimenter] to press any key" in a graphical mobile interface [13].

Combining touchtone and audio input made matters even worse: once participants learned that they were supposed to enter information on the keypad, they often did not say anything during audio input questions.

Based on our observations, we speculate that difficulties with hearing and/or comprehending the touchtone instructions, the added difficulty of moving the phone from one's ear and finding a button to press (possibly missing further instructions), unfamiliarity with using the keypad during a phone call, and failing to recognize the automated nature of the UI, all contributed to the failure of the touchtone interface.

Despite a number of attempts at improving the touchtone interface, 55% of the participants receiving a touchtone survey did not even succeed in pressing the 1 button to begin the survey, even when they were told to "Please press the 1 button on your phone to begin the survey." Instead, they said "Yes" or "1" or "Yes, I am ready" or simply hung up after hearing the instructions. In the cases where calls were at least carried out to completion (successfully or unsuccessfully), they typically took 6-8 minutes for 4 questions (versus about 3 minutes for the voice versions) because participants had to hear each question multiple times before they would press a button. This may have been a useful learning experience for participants, but was almost certainly also a frustrating one. Finally, considering the low success rate, it is likely that even the successfully completed surveys had a low degree of accuracy.

These results suggest that without at least some initial training, a touchtone interface is infeasible for this target population. We should emphasize that this work makes no claims about the usability of ASR-based UIs, which present a host of challenges themselves such as recognition accuracy and limited vocabulary (see e.g. [20], [16]). Clearly, *recorded audio* UIs are feasible, but how they can be further automated and scaled is not addressed here.

Outbound vs. Inbound Calling

Although having an IVR system call participants - rather than having participants initiate the call - was financially advantageous, we found that it introduced additional usability problems, which were only partially offset by the use of text message warnings. First, participants were often in an environment not conducive to a successful IVR interaction. These environmental factors included loud background noise, external distractions such as conversations with third parties, and intermittent call quality. Second, participants generally did not understand that they were interacting with an automated system, and tried to initiate conversation with the system. These problems were partially offset by the strategies described below.

Automated Call Preparation and Initiation

One thing that became clear from the initial testing was the importance of the text message 'warning'. Each of the Ugandans interviewed cited the importance of the text message to 'prepare' them for the call. Participants who were sent a call without receiving a text message warning were confused and would hang up after a few seconds of failed attempts to start a conversation with the recorded voice.

Despite the text message warning, participants generally did not immediately realize the nature of the IVR calls; we found that no matter how we began the survey dialogue, participants repeatedly said "Hello, hello? Who is this?", trying to establish a conversation, and thus missed the instructions. We found that beginning the call with a sound effect such as a chime, followed by 'This is a recorded call. You are not talking to a real person.' effectively captured the attention of users and compelled them to listen to the instructions.

Leveraging Implicit Dialogue and Turn-Taking Conventions

The success of the second voice interface (***V4***) suggests that leveraging conversational and turn-taking conventions of normal conversation are much more successful than detailed instructions in eliciting desired user behavior. In the first version of the voice survey, detailed instructions were provided at the beginning of the survey and questions were asked as statements (e.g. 'After the beep, please say your name and the name of the school where you work'). In the final version, we asked questions as questions (e.g. 'What is your name?') and relied on turn-taking to signal when the user was supposed to speak. This turned out to be much more successful. Users with limited understanding of English have a hard time understanding complex instructions, and the 'talk after the beep' convention is not understood in Uganda, where voicemail is rarely used; conversely, all users know to speak when they are asked a question.

Interestingly, in contrast to previous versions, participants were usually able to answer the questions even if they did not hear or understand the instructions due to call quality or background noise, because the expected response was implicit in the conversational nature of the survey.

The responses to ***V4*** were spoken more slowly and clearly enunciated than in previous versions. The literature reports that people tend to emulate the speaking style of their conversational partner in a voice dialogue[15]. Therefore, since the prompts were recorded more slowly and in a more understandable accent, the responses were also spoken more slowly and clearly[6].

Survey Design and Recording by Native Speakers

Even though our survey was delivered in English, we found the use of native speakers for designing and recording prompts to be extremely important. First of all, native speakers understand the vocabulary and mental model of target users. For example, we found that the phrases 'Press 1 to continue' or 'Press 1 on the keypad to continue' were much more difficult to understand than 'Press the 1 button to continue', because users did not know 'Press' referred to their phone, and 'keypad' was not a common word.

Perhaps even more importantly, native speakers are able to record prompts in an accent and speaking style that

[6]The responses also tended to be somewhat more concise in response to the shorter prompts, but we found that response length to open-ended questions was closely tied to the recording timeout (i.e. the length of silence before the recording ended), which we tuned to 3 seconds. Essentially, users continued talking with longer and longer pauses between sentences, and expected that the speaker would interrupt them when they had said enough. When the speaker finally interrupted with 'Thank you.', most participants appeared pleased.

is more understandable to users. Several Ugandans commented that our accent was hard for them to understand (just as their accent was hard for us to understand). Furthermore, we instructed native prompt recorders to record the prompts as if they were speaking to someone in Uganda with a poor cell connection. This resulted in a particular enunciation, intonation, and speaking rhythm that we could not have replicated, but which seemed to make the survey easier for users to follow.

Gender Differences in IVR Task Success

The gender discrepencies observed, although not conclusive, support the findings of Medhi et al. that women have a higher IVR UI task success rate than men [13]. Qualitatively, we found that (in line with Medhi) women generally listened more quietly to the instructions and answered more slowly and clearly, whereas men tended to talk during instructions (e.g. 'Hello? Hello?') and more often spoke at the wrong time, did not know what question they were supposed to answer, or were difficult to understand.

Remote IVR Application Hosting

This work demonstrates the feasibility of a 'cloud' approach to IVR application hosting. In this aproach, applications are hosted in a reliable 'cloud' or remote location, are administered over the internet, and connect to the country's phone network through VoIP. By hosting the application remotely, we eliminated the need for local hardware and technical expertise, and were not affected by power and network outages. The main disadvantage of this approach is cost: VoIP rates to Uganda can be over 10¢ per minute; however, for the scale of our deployment, we found the extra usage costs (<$100) were much less than procuring servers and technical support in-country.

We found no degradation in call quality when hosting our application in the United States and connecting to Ugandan mobile phones over VoIP. Therefore, remote hosting may be a good alternative to local hosting, especially for small-scale or prototype applications.

6. CONCLUSIONS AND FUTURE WORK

IVR applications in the developing world have the potential to extend ICT to the billions of developing-world users who own a mobile phone. The most serious challenge for IVR application development in this context is usability.

In this paper, we describe a study of IVR data collection UIs by untrained users in rural Uganda. Over 150 IVR survey calls were delivered to Ugandan teachers to collect feedback on a water education training. These calls were analyzed both qualitatively (listening to recorded call transcipts) and quantitatively (measuring task success rates), and informed UI changes in an iterative design process. Changes to the survey UI improved task success from nearly 0% to 75-100%. Our analysis of several UI designs suggests a number of more general design principles for IVR interfaces designed for similar populations.

We see several opportunities for further study of IVR data collection interfaces with untrained users.

First, further work is required to determine if and how conversational voice input can be used by an automated IVR interface. We found that UIs based on *recorded* voice input (rather than DTMF) were successful for untrained users, but it is unclear if and how this input could be interpreted using ASR.

Second, the accuracy of IVR-based data collection in the developing world has not yet been characterized. Patnaik et al. found that live operator data collection over voice outperformed graphical and SMS interfaces by an order of magnitude [17], but it remains unclear whether the improvements in data quality result from the voice modality or from the presence of a live operator. In order to answer this question, the accuracy of IVR interfaces in these environments must be determined experimentally.

There has also not been sufficient characterization of the effect of training on mobile data collection task success and accuracy. For example, Patnaik et al. observed over 95% accuracy on several UIs after hours of training, while we found that touchtone entry failed with no training. The tradeoff between training time and task success or accuracy on a particular interface has not been examined.

IVR applications in the developing world have the potential to connect billions of users to previously inaccessible automated services. For this potential to be realized, there remains much work to be done to develop the technology and the design principles necessary for these applications to be usable by these unreached populations.

7. ACKNOWLEDGMENTS

This work would not have been possible without the collaboration of the Project WET Foundation, particularly John Etgen and Teddy Tindamanyire. We would also like to thank Bill Thies and Neal Lesh; Gaetano Borriello, Yaw Anokwa, Carl Hartung and Waylon Brunette of Open Data Kit at University of Washington; and the JavaRosa team.

8. REFERENCES

[1] CommCare. `http://dimagi.com/commcare`.
[2] EpiSurveyor. `http://datadyne.org`.
[3] FrontlineSMS. `http://www.frontlinesms.com/`, Apr. 2010.
[4] Open Data Kit. `http://code.google.com/p/opendatakit/`, Apr. 2010.
[5] Project WET foundation. `http://www.projectwet.org/`, Apr. 2010.
[6] Y. Anokwa, C. Hartung, W. Brunette, A. Lerer, and G. Borriello. Open source data collection in the developing world. *IEEE Computer*, pages 97–99, 10 2009.
[7] M. H. Cohen, J. P. Giangola, and J. Balogh. *Voice User Interface Design*. Addison-Wesley, Boston, Massachusetts, first edition, 2004.
[8] A. S. Grover, M. Plauche, E. Barnard, and C. Kuun. Hiv health information access using spoken dialogue systems: Touchtone vs. speech. In *Proc. International Conference on Information and Communications Technologies and Development*, pages 95–107, 2009.
[9] Kabutana Trust of Zimbabwe. Freedomfone. `http://www.freedomfone.org/`, Apr. 2010.
[10] J. Klungsoyr, P. Wakholi, B. MacLeod, A. Escudero-Pascual, and N. Lesh. OpenROSA, JavaROSA, GloballyMobile - collaborations around open standards for mobile applications. In *Proceedings*

of The 1st International Conference on M4D Mobile Communication Technology for Development, pages 45–48, 2008.

[11] P. Kotkar, W. Thies, and S. Amarasinghe. An audio wiki for publishing user-generated content in the developing world. In *HCI for Community and International Development (CHI Workshop)*, 2008.

[12] C. Kuun. OpenPhone project piloted in Botswana. `http://www.csir.co.za/enews/2008_july/ic_05.html`, Apr. 2010.

[13] I. Medhi, S. N. N. Gautama, and K. Toyama. A comparison of mobile money-transfer UIs for non-literate and semi-literate users. In *CHI*, 2009.

[14] R. Millard and J. Carver. Cross-sectional comparison of live and interactive voice recognition administration of the SF-12 health status survey. *IEEE Computer*, 2(5):153–159, 1999.

[15] F. Oberle. *Who, Why and How Often? Key Elements for the Design of a Successful Speech Application Taking Account of the Target Groups*. Springer, Berlin Heidelberg, 2008.

[16] N. Patel, S. Agarwal, N. Rajput, A. Nanavati, P. Dave, and T. S. Parikh. A comparative study of speech and dialed input voice interfaces in rural india. In *Proc. ACM Conference on Computer Human Interaction*, 2009.

[17] S. Patnaik, E. Brunskil, and W. Thies. Evaluating the accuracy of data collection on mobile phones: A study of forms, SMS, and voice. In *Proc. International Conference on Information and Communications Technologies and Development*, pages 74–84, 2009.

[18] M. Plauche, U. Nallasamy, J. Pal, C. Wooters, and D. Ramachandran. Speech recognition for illiterate access to information and technology. In *Proc. International Conference on Information and Communications Technologies and Development*, 2006.

[19] 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. International Conference on Information and Communications Technologies and Development*, 2007.

[20] J. Sherwani, S. Palijo, S. Mirza, T. Ahmed, N. Ali, and R. Rosenfeld. Speech vs. touch-tone: Telephony interfaces for information access by low literate users. In *Proc. International Conference on Information and Communications Technologies and Development*, pages 447–457, 2009.

[21] J. Sherwani and R. Rosenfeld. The case for speech technology for developing regions. In *HCI*, 2008.

[22] K. Shirima, O. Mukasa, J. A. Schellenberg, F. Manzi, D. John, A. Mushi, M. Mrisho, M. Tanner, H. Mshinda, and D. Schellenberg. The use of personal digital assistants for data entry at the point of collection in a large household survey in southern tanzania. *Emerging Themes in Epidemiology*, 4:5+, June 2007.

[23] B. Suhm. *IVR Usability Engineering Using Guidelines And Analyses Of End-to-End Calls*. Springer, US, 2008.

[24] R. Tourangeau, M. P. Couper, and D. M. Steiger. Humanizing self-administered surveys: experiments on social presence in web and IVR surveys. *Computers in Human Behavior*, 19(1):1 – 24, 2003.

[25] UNCTAD. Information economy report 2007-2008: Science and technology for development - the new paradigm of ICT. In *United Nations Conference on Trade and Development*, 2008.

[26] I. T. Union. ICT statistics. `http://itu.int/ITU-D/ict/statistics`, Apr. 2010.