Are Spoken Dialog Systems Viable for Under-served Semi-literate Populations?

PhD Thesis Proposal

Jahanzeb Sherwani

Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract. Spoken dialog systems have been widely researched, and are in everyday use (e.g., Amtrak's automated reservation system at 1-800-USA-RAIL). This research, however, has been targeted at Western users – comparatively affluent, literate, with many different options for information access, and with very different realities than those of the typical user in the developing world. Further research is needed to understand the users that comprise the other half of the world – the 2.8 billion for whom such speech interfaces may be the *only* realistically viable option. It is the use of dialog systems by such users that I propose to study. Specifically, the research question I wish to address is: **are spoken dialog systems an effective interface choice for information access by semi-literate users in the developing world?** My proposed research plan is to design, develop and test a spoken dialog system for information access in the domain of community health, tailored to the skills and needs of community health workers in the developing world, who are often semi-literate and have a dire need for access to information. The system will be evaluated relative to existing mechanisms for information access (such as handbooks), on standard HCI metrics such as task completion rate, task completion time, and user satisfaction.

Table of Contents

1. Motivation

1.1 Information & Communication Technology for Development (ICT4D)

Just under half the world's population, or around 2.8 billion people, currently live on less than 2 dollars a day. There exist a number of domains where Information and Communication Technologies (ICTs) can provide real value to such populations, in a way that is both sustainable and appropriate [14, 36]. There are hundreds of such projects (as cataloged by the World Bank [28]), most of which use existing, off-the-shelf technology [1]. However, the direct transfer of "First World" technology has not been successful in most cases, primarily because of the mismatch between the intended environment the technology was designed for, and the ground realities of the environments in which they are deployed [29, 30, 31, 32]. Brewer et al. describe this situation concisely: "Although it is clear that there are large differences in assumptions related to cost, power, and usage, there has been little work on how technology needs in developing regions differ from those of industrialized nations. We argue that Western market forces will continue to meet the needs of developing regions accidentally at best" [1]. Research on technology design tailored to the specific needs of emerging regions is needed to address this issue [1, 15, 32, 36].

Domains such as health, education, agriculture, e-governance, and commerce have great potential for the application of appropriate technology [14, 36]. Within these domains, there are multiple avenues where technology can play a role, such as information access [19], information entry [24, 25], (decentralized) information sharing, access to services, and access to automated decision-making systems. These applications in the aforementioned domains can lead to higher efficiency, lower costs, greater reliability, more accessibility, greater localization and better quality of information & content, decentralized communication, which in turn directly lead to better income, health, education, and ultimately, in the quality of life [16].

Most ICT-for-development (ICT4D) initiatives involve the use of standard PCs as the form factor, and standard web-based forms or Windows-based GUIs as the primary interface, and the Internet for connectivity. However, PCs and current GUIs were designed with a specific ("First

World") user in mind: a user who can afford a roughly $500 machine as well as Internet connectivity, has access to a stable electricity supply, is literate, uses a language that has a written form, and finally can access and afford technical support when something fails or needs to be upgraded. These requirements are unrealistic for major parts of the developing world, where users in many cases cannot afford such costly technology, do not have access to continuous electricity, are not literate, may be fluent only in a language without a written form, and do not have access to any ongoing support for using unintuitive technology. While there are successful ICT projects using PCs (for example, the widely publicized e-Choupal initiative in India [33]), there is a large part of the developing world for which such design is not viable. It is no surprise that rates of PC use in the developing world are dwarfed by those in the developed world.

Cell phones, on the other hand, are a huge ICT success at the time of this writing [15]. Cell phone use across the developing world is increasing remarkably – according to a report by the International Telecommunications Union titled "Mobile overtakes Fixed":

> "The greatest impact of mobile communications on access to communication services – in other words, increasing the number of people who are in reach of a telephone connection of any kind – can be seen in developing countries." [20]

Across Sub-Saharan Africa, access to cell phones surpassed access to fixed-line phones in 2000 – the same occurred in South Asia in 2002 [18]. Given that cell phones in the developing world often follows a shared model of usage, the actual number of people with access to telephony is much higher than the number of subscriptions suggest [3]. The extensive use of cell phones suggests that this is a fundamental mechanism through which underdeveloped regions are benefiting from ICTs, because they are easy to use, affordable, and suitable for non-literate populations. Furthermore, the sustained use of cell phones in these regions also implies the existence of widespread ecosystems of supply, maintenance and technical support, which do not exist for other types of devices. For these reasons, it is believed that cell phones have great potential for facilitating ICT projects of a wide variety [36].

Most usage of cell phones is for human-human communication – however, telephony also enables the possibilities of human-computer interaction. The question then is: what kind of cell phone-based applications and interfaces are most appropriate? There is promising work in the field of mobile GUIs targeted towards developing regions for rural self-help groups involved in micro-credit [4]. However, GUIs largely depend on literacy, and with literacy rates of less than 50% in developing regions, this is not a mechanism that can work for all. Furthermore, while literacy statistics seem to suggest high overall levels (76% in developing countries) [26], the methodology of these statistics reveals that the data is derived from individual or household declaration (and not through any standardized testing), and that the definition of literacy is stretched in some cases to "the ability to read easily or with difficulty a letter or a newspaper" [27]. Unfortunately, when it comes to the use of interfaces, "difficulties" with the interface often spell the end of any use of that interface.

The core technologies of speech recognition and speech synthesis, on the other hand, do not require literacy and in fact even work for languages that have no written form. Thus, interfaces that use speech as the underlying modality – otherwise known as spoken dialog systems – hold great promise as an interface choice for such users.

These users are not one homogenous group. Income, literacy, and other factors vary widely within regions, although in general, it is the case that the affluent and literate are the minority, while the poor and semi-literate are the majority. SLTs may not be the answer for those at either extreme of the income & literacy spectra. For the resource-rich, the realities are similar to those in the West, for whom speech has not been appealing, and for whom other technologies such as Internet-through-the-PC may be more affordable and accessible, and so are less motivated to use SLTs. For the extremely resource-starved the situation is completely the opposite: they may not be able to easily learn to use SLTs, and might have more pressing needs, such as food and water, instead of information access (although it has been argued that information can reduce the price paid for such commodities greatly [2, 19]). *It is my hypothesis that in between these two extremes there is a middle ground, where users have the motivation and the skills to be able to master the use of SLTs, yet for whom accessing "richer" interfaces to information is not an option.* I aim to

investigate this hypothesis in the context of community health, which I will now give a brief overview of.

1.2 Community Health Programs in Developing Countries

Healthcare is a fundamental, under-serviced need of citizens in developing countries. In addition to having the highest maternal mortality and neonatal mortality ratios, these regions have the largest unmet need for health service providers in the world. Given the high cost of training doctors and nurses, and the low number of medical schools in these parts of the world, many governments have begun community health worker (CHW) programs, where people (usually women) are chosen from their own communities, trained in basic health service provision for a few months, and sent back to provide health services in their communities. These CHWs vary greatly in literacy levels and receive little refresher training, if any. It is not surprising that the need for better information access by CHWs is widely agreed upon: “Providing access to reliable health information for health workers in developing countries is potentially the single most cost effective and achievable strategy for sustainable improvement in health care” [6].

The Pakistani government, for example, has initiated a community health worker program with the same logic – called the “Lady Health Worker Programme” (LHWP). This program employs 100,000 LHWs across Pakistan (a country with a population of around 160 million). These LHWs receive 3 months’ training, with no refresher courses in most cases. A recent evaluation of the LHWP gave a strong recommendation for the improvement in the quality of knowledge of the LHWs [5, 7]. Many other countries have similar programs [37, 12].

Traditional mechanisms for health information access by LHWs have not been adequate. The easiest such mechanism for health workers is to ask someone who is better-informed: a doctor, a nurse, or even the health worker’s supervisor. Unfortunately, there are not enough doctors and nurses to satisfy the information demands of the health workers. Furthermore, there are interpersonal dynamics that limit the effectiveness of supervisor-worker training: some supervisors have the same training as the health workers they supervise, and are afraid of losing their job to a well-performing health worker [7, 10].

Another mechanism used for information access is that of written text: handbooks and training manuals are difficult to distribute, costly to update, and are seldom referred to by health workers (see Section 3 for details). Solutions involving information access through PCs (using the Internet or even CD-ROMs) are not viable for reasons of cost, literacy, and access. Thus, there is a strong need for information that is not being addressed through existing mechanisms.

1.3 Types of Health Information

The book "Where There Is No Doctor" [38] is a classic example of the community handbooks used by CHWs across the world. It contains the basic health information that would be useful in communities which do not have access to a doctor, and has been translated and adapted into more than 100 languages.

The information presented in the book ranges in terms of its level of structure. For topics such as family planning, the information is largely unstructured text. For all specific diseases such as diarrhea, the information is very structured, including consistent subtopics such as mechanisms for prevention, signs & symptoms, mechanisms for diagnosis and classification, and treatment. In some cases, complex flow charts are presented for the classification of some diseases.

1.4 Range of Health Worker types

There are many different levels of health workers, varying from the two extremes of the highly educated MDs to the completely untrained Traditional Birth Assistants (TBAs). TBAs are women whose family's women have traditionally been the midwives of their community, and their only source of knowledge is what has been passed down through the oral tradition. While some of their practices are safe and effective, much of their knowledge is actually harmful to their clients [10, 11].

For an MD, accessing a dialog system for health information may be needless and cumbersome – MDs are very well educated, can easily refer to other doctors and professional journals, and in

many cases have access to the Internet as well. For a completely non-literate TBA, it may be prohibitively difficult to learn the use of the dialog system, even though it may provide real value. My hypothesis is that there exists a group of users in between both these extremes in education and income levels, who are semi-literate and are able to learn the use of a dialog system without much difficulty, but at the same time cannot access, afford or use other information mechanisms, and hence have an optimal mix of ability and motivation to use such a system.

Trained community health workers (such as LHWs) described previously fit this description quite well. They have some limited education, and have limited health training. They clearly have the motivation to access health information (and have expressed this desire unequivocally). Yet, they are unable to use existing mechanisms well (e.g., books) and are not provided the support they can use most easily (e.g., doctors). A well-designed dialog system, tailored to their needs, has the potential to give them the information they want, and their motivation levels and lack of reasonable alternatives should enable them to be more forgiving of the system's flaws.

Furthermore, books are costly to print and distribute, difficult to carry around, and impossible to implement usage monitoring with. Updates to such books are also very expensive. Speech systems, on the other hand, have a much lower cost of scale-up and maintenance. Only a cell phone needs to be carried around, and an update to the system is reflected in all interactions thereafter by all health workers. Also, it is easy to monitor usage of the system, and to use this information to improve and/or localize specific content that is most used. Thus, if speech systems are even comparable to handbooks (and not significantly better) in usability, there are large external advantages they bring as well.

In the next section, I will review related work and present the current state of the art. In Section 3, I will present findings from the preliminary phase of the needs assessment process I have initiated to understand the needs and abilities of this target group. In Section 4, I present the proposed research program. Section 5 presents the timeline for the project. References are listed in Section 6.

2. Related Work and Current State of the Art

I will now review work from various related fields: design for semi-literate users and/or users in cross-cultural contexts, research on user responses and preferences on dialog systems, research on the cognitive differences brought about by literacy, and speech interfaces for semi-literate users.

2.1 Design for Semi-literate & Different Cultural Contexts

There has been a modest amount of research on principles of design for semi-literate users in the developing world. Hofstede [43, 44] developed a framework of five fundamental cultural dimensions based on anthropological research on IBM employees from 53 countries:

- Power distance: the extent to which people accept power inequality and social hierarchies
- Individualism vs. collectivism: the orientation towards individual or group achievements
- Masculinity vs. femininity: the degree to which a culture separates gender roles
- Uncertainty avoidance: the degree to which uncertainty is considered uncomfortable
- Long-term time orientation: the degree of devotion to long term traditions

There have been suggested mechanisms for concretizing such higher-level theories by mapping each dimension to specific user interface components [45, 46, 49]. For instance, [49] suggests that web interface design for cultures with high power distance (i.e., where power inequalities are more accepted) should use more formal text. These recommendations are based on the authors' experience in designing web-sites in different countries, using Hofstede's underlying theoretical framework.

[47] presents a large list of design recommendations for non-literate users of a proposed PDA-like device, with many recommendations for speech interfaces. However, these recommendations are not derived from empirical evidence from evaluations with actual semi- or non-literate users – they are more the result of a literature review of interface issues with Western users. Most of the given design recommendations assume that such users would respond

very similarly to speech interfaces as Western users do (e.g., “[users] of speech-only interfaces like to feel that they are in control of the interface, and allowing users to interrupt the system's speech output is an important way to convey this sense of control”). However, according to [49], users with high power distance *prefer* interfaces where the system maintains control.

A project by the creators of the Greenstone Digital Library platform focused on extending access to digital libraries by non-literate users [51]. The authors enumerate a list of user requirements for the interface, including high learnability, habitability, usability, robustness to errors, and non-textual as far as possible. Testing paper prototypes provided the further guidelines of the need for an easily accessible main menu, as well as audio-based help. Usability studies of the actual system showed that users (non-literate entrants to an adult school in New Zealand) were not able to adequately navigate the information, mostly because of the high memory demands of the text-free browsing interface. The authors suggested that keyword search and/or browsing of a limited information set could be a potentially more feasible approach.

A similar attempt at a text-free interface is described in [50]. A number of GUI-specific design principles are given, including the need for audio feedback on all functional units on the display, and the importance of providing help at any time during the interaction. Usability tests on the actual applications showed that users (non-literate sweepers in an Indian urban slum) were generally successful (task success between 50-100%) in completing the given tasks using the text-free interfaces, and were completely unable to do so using the text-based one. However, the sample size was small, and these results are not statistically significant.

There are a number of other projects investigating interfaces for non-literate users, such as Livestock Guru [48], which provides livestock information to rural users in India and Bolivia, and PCtvt, which seeks to create a compelling interface that enables access to television, video, telephony and regular PC use, for semi-literate and non-literate users [61].

2.2 Speech Interfaces

There are many models of user behavior that speech interface research has created. For instance, it is widely reported that users speak differently with systems than with humans, in various domains. [47] suggests that users prefer being able to barge in, because it gives them a feeling of control. Somewhat contradicting this finding, [13] shows strong results indicating users' preference for directed dialog-based speech interfaces over user initiative-based interfaces. However, none of this has been evaluated in the context of non- and semi-literate users in the developing world, where variance along economic, sociocultural, and cognitive lines may mean completely different user behavior compared with Western users, who have been the focus of most contemporary speech interface research.

Cross-cultural interface studies show that computer interfaces are often anthropomorphized by users [53, 54]. Also, there is promising work on earcons that have cross-cultural relevance: for instance, the sound of a breaking glass was perceived universally as an event that had to be stopped instinctively, much more so than a ringing alarm sound, which worked only for users from a specific cultural context [56].

2.3 Cognitive Differences due to Literacy

Literate and non-literate or semi-literate users are similar in most ways, although there are some fundamental differences. [57] presents results from a study comparing task success by literate and non-literate participants in 7 visuospatial tasks and 13 memory tasks. In all but one task (immediate memory of a sentence), non-literate participants were significantly worse than literate participants. [58] presents evidence of differences in brain activation during a word repetition task in literate and non-literate participants, while [59] shows that non-literates perform comparatively poorly in repeating pseudo-words. Thus, there is evidence in the literature that there are fundamental differences in literate and non-literate populations. How this impacts speech interface usage remains to be investigated.

2.4 Speech Interfaces for Semi-Literate Users

The deployment of spoken language interfaces in developing countries is a daunting task, specifically because of the scarcity of: 1) basic linguistic and cultural knowledge, 2) linguistic resources, 3) software tools, and 4) guidelines for the design of user interfaces [21]. While there is currently no easy way to deal with linguistic gaps, there has been considerable work on efficient bootstrapping mechanisms for linguistic resources (e.g., corpora, dictionaries, grammars, parsers) for resource-scarce languages [22, 23]. Also, while there are some freely available software tools, such as Festival, Sphinx, and HTK, the scarcity of readily available run-time environments and authoring tools makes the rapid development of speech interfaces difficult. Current work on the creation of toolkits targeted towards content-providers in the developing world [8] is promising – although to be effective, such initiatives require a clearer understanding of user interface design guidelines, so that best practices can be built into the toolkits themselves.

After presenting the above four problems, the authors in [21] go on to surmise: “taken together, the obstacles listed above may seem insurmountable”. However, they point to a number of examples where similar problems were solved through a combination of technological and sociological means. They conclude with the following: “If spoken interfaces can be shown to have [high] value in the developing world, one can be confident that the hurdles to their development will be overcome”.

There has been only one study evaluating speech systems for non- and semi-literate users. [8, 9] describe a speech interface called Tamil Market, that was evaluated on 13 men and women, 3 of whom were non-literate. There was a difference in task completion rates and task completion time. However, these differences were not significant since the sample size was too small.

The main anecdotal lesson from this study was that users were not interested in using an interface that did not directly address a need of theirs. For example, the system could answer questions on banana crop prices, and users enthusiastically used the system in villages where banana crops were grown. However, in other nearby villages, users were unwilling to test the

interface since they grew other crops, and did not perceive the interface as giving any value, even when the researchers said that their feedback would help them customize it to the crops they were interested in.

To summarize, there have not yet been any studies testing speech interfaces with significant sample sizes of non-literate users. There are many indications that such users can use such interfaces successfully. However, there is also evidence that there are fundamental differences between literate and non-literate users, and that design principles that work for the former may be successful for the latter. Research targeted towards understanding these differences is required, and is one of the contributions of this thesis.

3. Initial Findings

During the summer of 2006, I conducted preliminary discussions with community health workers, as well as their supervisors and with researchers in public health. The goals of this endeavor were to get a basic understanding of the health information needs of community health workers, as well as the views of their supervisors regarding these needs.

My discussions were primarily with people associated with the following institutions:

1) Lady Health Workers (LHW) from the Government of Pakistan's Ministry of Health

   I spoke with 20 LHWs from Sultanabad (an urban slum in Karachi) in a group discussion. I also spoke with the supervisor of LHWs in that area, as well as the manager of supervisors in the district.

2) Male Community Health Attendants (CHAs) and Female Community Midwives (CMWs) from the Health & Nutrition Development Society (HANDS), a Karachi-based NGO

   I held individual interviews with 3 CHAs and 3 CMWs in Jamkhanda, a rural area two hours by road from Karachi. I also spoke with the executive director of HANDS, and the doctor involved with training health workers.

3) Lady Health Visitors (LHVs) from the Family Planning Association of Pakistan (FPAP), a national NGO

   I spoke with 8 LHVs in a group at the FPAP regional office in Karachi. I also spoke with the FPAP regional director, as well as the doctor involved in quality assurance and training of the LHVs.

4) Researchers from Aga Khan University's Community Health Sciences Department's (CHS) Urban Health Program

I spoke with various researchers with vast experience in community health interventions and research.

Early interactions showed that since most people in Pakistan have never used a dialog system, it was nearly impossible to discuss the possibility of using such systems for information access by health workers. I realized the need for clearly presenting what was meant by "spoken dialog system" to the various people I spoke with, and created video prototypes for this purpose. Specifically, I developed designs illustrating three different interaction styles, wrote scripts for each style, translated them into Urdu, and produced videos in which a "health worker" interacts with each system. I showed these videos [42] to every person I spoke with, which greatly helped ground the concept of spoken dialog systems for health information access, including their strengths and weaknesses. The three interaction styles consisted of:

1. System-initiative dialog with an expert system to help with the classification and treatment of a disease
2. Natural language question-answer system where user asks *any* question to the system
3. User-initiative dialog with keyword search-based information retrieval system

One misrecognized utterance was present in each of the three systems. Furthermore, the system's response to the natural language question in the second video prototype was purposefully made slightly inaccurate, to reflect the limitations of question-answer systems' actual accuracy.

Informal interviews and group discussions with health workers revealed the following:

a) Every single health worker I spoke with expressed a desire for more refresher courses, and more mechanisms for accessing information. One group of health workers even said this in front of their supervisor, who was technically in charge of giving them the refresher training as and when needed -- a very significant indication in this deference-based culture.

b) Every health worker asserted that the systems we described would be useful and that they would use them – although they varied in the extent and enthusiasm with which they said so.

c) Most health workers were worried that the system would only give information on diarrhea (since all of our videos only dealt with diarrhea), and gave us a long list of areas that they would want information on.

d) Of the 34 health workers interviewed, fewer than 5 reported ever referring to a book. Only 1 said that she read regularly (one chapter of the reference handbook every night). Most of the reasons the others gave for not reading books was that they weren't used to it, that it's difficult to read and to find what one is looking for, and that it's hard to carry around a heavy book.

e) Female health workers are not comfortable discussing female reproductive health issues in the company of males. In one focus group with female CHWs that I co-ran with a female colleague, participants engaged deeply but only after I, a male, left the room – even though the my colleague typed notes into the same laptop I was using, and even though it was evident that the two of us would confer on the discussion later on. This incident underscored the need for a female-mediated needs assessment strategy.

Discussions with public health experts revealed the following:

- The underlying principle of health worker programs is to provide a basic level of service that does not require highly trained and skilled personnel such as doctors and nurses.

- Health workers are given training in neither biochemistry nor pharmacology. Instead, they are trained on simple protocols that have been established by organizations such as the World Health Organization. The most commonly practiced set of protocols is the WHO's Integrated Management of Childhood Illnesses (IMCI) guidelines, which specify a set of algorithms that health workers need to follow in order to correctly assess and treat

children. This information is usually expressed as flowcharts to convey the big picture, with details given in longer pieces of text (paragraph, bulleted, or even tables).

4. Proposed Research

4.1 Thesis Statement

Spoken dialog systems are a viable interface choice for information access by semi-literate users in the developing world.

4.2 General Approach to Testing the Thesis

My proposed approach is to design, develop and evaluate a spoken dialog system for health information access by community health workers to understand whether such interfaces are viable for such users.

Viability can be defined in absolute or relative terms. I define viability in terms of *relative efficiency* – namely, with respect to comparable alternatives. Thus, in the context of health information access by community health workers, I will be looking at the relative efficiency of a speech interface compared to existing mechanisms for information access – namely, health handbooks.

The information that will be accessible through the interface will be the same as the information found in the text alternatives I will be comparing against. The purpose of this research is to understand the use of the *interface* by which information is accessed, and this requires that both interfaces (speech and text handbook) are balanced with respect to the information they contain. However, since speech interfaces allow much more flexibility in terms of the manner in which the information is presented, there will be a potentially large difference in the style of presentation between both alternatives – in fact, it is specifically this difference that I believe will give speech interfaces a comparative advantage. For instance, instead of displaying a complete flowchart showing the steps needed to classify a disease, which may be daunting to many semi-literate users, a speech system can ask the user one question after the other, and lead them through the same flowchart in a conversational manner.

As outlined in Section 2, speech systems bring many external advantages beyond usability issues when compared with books, in terms of time and cost for dissemination, as well as accessibility. Thus, my aim is to show that speech systems are at least comparable to books, and I consider this a proof of viability. However, if the evaluation shows that speech interfaces do not perform even comparably with books, my research will highlight the underlying causes for this, and present future lines of inquiry in this realm.

I now present the proposed sub-goals / milestones for the research. Secion 4.3 presents the preparatory work both in terms of field partnerships as well as on system development. Section 4.4 details the evaluation goals and methodology for this research proposal. Section 4.5 outlines some clarifications on the research goals, and Section 4.6 lists thesis contributions.

4.3 Preparatory Work & System Development

Preparatory work and system development consist of the following:

1. Field preparation
2. Health information digitization
3. Urdu speech component creation
4. Prototype creation

4.3.1 Field Preparation

Goal: Build partnerships with various organizations involved in community health

Proposed Approach: Form a partnership with a research university working in the field of public health, and in collaboration with them, form a partnership with community health organizations. Discuss speech-based health information access with administrators and health workers and get initial feedback.

Action Items:

- Form partnership with local university in public health – *completed*
- Form partnership with community health organizations – *completed*
- Conduct initial discussions with community health workers – *completed*

4.3.2 Health Information Digitization

Goal: Compile health information resources and digitize health content in Urdu

Proposed Approach: I will request access to the training manuals and handbooks used by the aforementioned health organizations' health workers. Where soft copies are available, I will convert the format they are given in – usually in the InPage word processor's proprietary format – to Unicode. Where only hard copies are available, I will oversee the process of transcription by professional transcribers I will hire.

Action Items: (see Section 3 for explanation of acronym names)

- HANDS – access and convert soft copies of manuals for health workers – c*ompleted*
- FPAP – access hard copies and oversee transcription
- LHW – access hard copies and oversee transcription

4.3.3 Urdu Speech Component Creation

Goal: Build speech recognition and speech synthesis components for Urdu

Proposed Approach: For Urdu speech recognition, I will attempt to use an existing US English speech recognizer and using phoneme mapping between US English to Pakistani Urdu, will test its word error rate in a short study. If such an approach is not effective in the intended domain, I will initiate a modest Urdu speech data collection effort and train a speech recognizer on this data. However, if this approach does not lead to acceptable accuracy, I will use a Wizard-of-Oz set up for my experiments instead of an actual speech recognizer, with artificially injected error

rates comparable to existing state-of-the-art error rates in commercial speech recognizers. Additionally, I will hand-create a phonetic dictionary for the intended domain.

For Urdu speech synthesis, I will use mostly pre-recorded prompts and audio, and will use concatenative synthesis sparingly, to minimize intelligibility issues with speech output. Since the domain consists mostly of static information, such an approach is feasible. I will recruit a voice talent to read out the handbooks used in the system in their entirety. In the parts of the interaction where dynamic information is presented (e.g., the number of search results found), concatenative synthesis will be used.

Action Items:

- Build an English-to-Urdu phoneme mapping component
- Build a phonetic dictionary for the health domain in Urdu
- Test the accuracy of the recognizer with phoneme mapping
- Collect Urdu acoustic and linguistic data
- Train speech recognition engine for Urdu
- Record Urdu audio for entire handbooks & concatenative prompts
- Create speech synthesis voice for Urdu

4.3.4 Prototype Creation

Goal: Design prototypes of speech interfaces and get initial feedback from potential users

Proposed approach: Design prototypes of potential systems, and create video prototypes to show health workers to elicit initial feedback. Then design and build actual prototypes, test on users, and re-iterate design and tests until convergence.

Action Items:

- Design and create video prototypes – *completed*
- Design, develop and test first prototype
- Reiterate design, develop and test cycle successively

4.4 Research Strategy

I now present three main themes of my research. The first involves a study comparing the usability of a speech interface with a textbook, with the task of directed information access. The second involves a longitudinal study. The third covers a number of dimensions I will explore in terms of similarities and differences in the behavior of semi-literate users in the developing world compared to users described in traditional speech interface research literature.

4.4.1 Evaluation of speech system using textbook control

Goal: Evaluate comparative efficiency of speech interface with respect to comparable handbooks

Proposed Approach:

Participant profile: I will recruit 50 community health workers from the various organizations outlined in Section 3. I will select users who are trained on similar aspects of health service provision (e.g., maternal health), with large variance in factors such as:

- Literacy
- Years of schooling
- Years of health training
- Years of health service provision
- Complexity of job role / training

Methodology: I will record all relevant participant bio-data, including the factors listed above. I will first conduct a written pre-test to evaluate their individual existing knowledge levels, using test instruments from public health evaluation resources. I will also test factors such as comfort with reading text, as well as, comfort with technology. A randomly selected half the participants will first be given training on the use of the speech system for searching for and accessing health information, while the other half will be trained on the use of handbooks and indexes for the

same task. They will then be presented with another knowledge test, although this time they will be asked to use the mechanism they were trained on (speech interface / handbook) to answer these questions. After this, they will be given training on the other system, and given another test to be answered with this other system. Users will be incentivized for the quality of their answers.

Evaluation Metrics: I will evaluate the system on ISO guidelines of usability measurement [40], namely effectiveness, efficiency and satisfaction.

- Effectiveness is the accuracy and completeness of the task performed by the user. This is often measured by task completion rates and task error rates. In this study, I will focus on the correctness of users' answers to the post-test as the primary measure of effectiveness. I will also look at the difference in the pre- and post tests in both conditions.

- Efficiency is the measure between user effectiveness and the cost incurred to achieve that effectiveness. In this study, I will be measuring efficiency in terms of task completion time.

- Satisfaction is the users' subjective rating of the system. In this study, I will design a questionnaire based on the Subjective Assessment of Speech System Interfaces (SASSI) project [39]. This will contain 34 subjective response items on a 7-point Likert scale, with

System Response Accuracy
- The system is accurate
- The system is dependable
- The system makes few errors
- The interaction with the system is consistent
- The interaction with the system is efficient
- *The system is unreliable
- *The interaction with the system is unpredictable
- *The system didn't always do what I wanted
- *The system didn't always do what I expected

Likeability
- The system is useful
- The system is pleasant
- The system is friendly
- I was able to recover easily from errors
- I enjoyed using the system
- It is clear how to speak to the system
- It is easy to learn how to use the system
- I would use this system
- I felt in control of the interaction with the system

Cognitive Demand
- I felt confident using the system
- I felt calm using the system
- The system is easy to use
- *I felt tense using the system
- *A high level of concentration is required when using the system

Annoyance
- *The interaction with the system is repetitive
- *The interaction with the system is boring
- *The interaction with the system is irritating
- *The interaction with the system is frustrating
- *The system is too inflexible

Habitability
- *I sometimes wondered if I was using the right word
- *I was not always sure what the system was doing
- *It is easy to lose track of where you are in an interaction with the system
- I always knew what to say to the system

Speed
- The interaction with the system is fast
- *The system responds too slowly

Figure 1: Subjective response items and their corresponding user satisfaction factors [41]. Items with a star (*) imply a negative relationship of the response with the factor.

each item related to a specific factor. These items and their corresponding factors are shown in Figure 1. [41]

Analysis: I will analyze the above three usability metrics and their relation to various factors, such as the demographics of the users (literacy level, years of training, etc.) and the conditions of specific users in the experimental setup (order of presentation, information access mechanism used), similar to the strategy used in [60]. I compare the differences in pre- and post test usability metrics between both conditions, as well as just the post-test metrics between both conditions. By looking at the difference between the post-test metrics for speech interfaces and books, I will be able to factor out issues such as the ability to take tests and the ability to look for information in general – and this will give a quantifiable measure of the direct benefit of speech interfaces (if any).

Action items:

- Compile test instruments for pre- and post-tests of user knowledge and information access
- Design training procedure for both conditions (speech and book)
- Recruit 50 health workers with varying levels of literacy and training
- Conduct user studies with the 50 participants, including training sessions

4.4.2 Longitudinal Evaluation

Goal: Evaluate the effects of continued use versus sporadic use of the speech interface

Proposed Approach:

I plan to take a smaller subset of around 20 users from the previous study, and give a randomly selected half the opportunity to use and learn the speech and handbook mechanisms continually. The other half will not be given any opportunity to learn the speech system or the handbook. I will then carry out a study similar to the previous study, using the same evaluation metrics as

before, where participants use either the speech system or the handbook to answer questions in a knowledge test.

In this study, I am interested in longer term issues, to simulate conditions outside the lab. Due to the uncertainty of the longer-term availability of participants, this study is not yet concretized. It may last up to 8 weeks, where participants are given a few hours of exposure and practice with the system and handbook every week, or I may give participants concentrated exposure to the system and handbook every day for a week.

The main question I wish to answer with this study is: does the viability of the interface increase with usage, and decrease with lack of exposure and supervision?

4.4.3 Investigating Speech Interfaces Issues with Semi-literate Users

Goal: Test findings of speech interface user research in the West with semi-literate users in developing countries to understand how universal such findings are

Proposed Approach: As detailed in Section 2, there is a large body of work on the response of users in the developed world to speech systems. However, there have been no studies done to date comparing these results in the developing world – specifically with semi- and non-literate users. I will investigate as many as possible of the following aspects of speech systems by users in the developing world, subject to time limitations:

- Anthropomorphization of speech systems [53, 54]
- Tendency to speak tersely with speech systems
- Preference for system-driven dialog if new users [13]
- Preference for speech-systems over alternatives when given a choice
- Desire to barge-in and have control [47]
- Perception of “intuitive” ear-cons [56]

4.5 Clarifications Regarding Scope of Proposed Work

As outlined earlier, I hope to eventually contribute towards *improved health outcomes* in developing countries through the use of speech interfaces. However, the specific goal of my research is much more modest. I will be focusing entirely on the user – i.e., the health worker – and will be looking exclusively at user-centric measures of usability.

Furthermore, while the measure of *value* that speech-based information access systems give to health workers is important, "value" depends on many external factors, including user motivation, the perceived importance of the specific information being accessed, and the perceived importance of the role of health workers in the community. I consider these issues important, yet outside the scope of my thesis, and I will instead be focusing exclusively on standard usability metrics listed above.

Also, I do not intend to investigate the proper niche for speech systems in public health ecosystems – this is a question for public health policy research. I hope to show, however, that such interfaces are *useable* by community health workers, and my research will hopefully be useful in guiding public health policy decisions in the near future.

Finally, it is generally accepted in the speech technology community that given enough time, money and effort, speech recognition and speech synthesis capabilities for any language can be built, with error rates comparable to systems for English. Given the kind of evidence my research hopes to discover, international institutions and governments could be persuaded to invest in the creation of language resources for such technology. Thus, my user studies will be surmised on achievable state-of-the-art language technology capabilities. I see little value in investigating usability issues with dialog systems based on highly error-prone speech recognition and synthesis components, which is why I will be using higher accuracy substitutes where applicable.

4.6 Thesis contributions

This research will contribute the following to the scientific community:

- An understanding of the viability of speech interfaces for semi-literate users in the developing world
- A more thorough investigation of speech interface issues for such users
- Design recommendations for speech interfaces for such users
- A research platform for speech-based information access of unstructured information

5. Timeline

The plan I developed for this research includes the following (items with a checkmark indicate they have been completed):

January – August 2006

- ✓ Literature review of speech interfaces for semi-literate users, speech interfaces for information access, ICT-for-development initiatives, and health information access interfaces for community health workers

- ✓ Setting up a research collaboration with a health sciences university in Pakistan

- ✓ Building a partnership with health-worker-based NGOs in Pakistan, whose health workers would be potential users of the research system

- ✓ Identifying health information resources that are in current use by these health workers

- ➢ Acquiring and digitizing text resources where necessary into a machine-readable text format (behind schedule)

September – December 2006

[Internship at Microsoft Research, Redmond]

- ➢ Familiarize myself with core technology that may be useful in the field, subject to approval by Microsoft

January – June 2007

- ➢ Developing and implementing all the necessary components of the system, such as telephony interface, Urdu speech recognition (including development of Urdu acoustic

models), appropriate lexicon, grammar, and language model, Urdu language generation and Urdu speech output.

- Building an end-to-end system that is capable of speech-based health information access in Urdu, and which allows a "Wizard-of-Oz" mode, where an operator is able to choose dialog actions on behalf of the dialog manager

- Testing the system with health workers in a lab environment in Pakistan

July – December 2007

- Iterating the design and development of the system, guided by frequent tests of the system with health workers

- Evaluating the usability of the system compared to the usability of text resources such as handbooks, in a lab-based, between-subjects study measuring task completion rate/quality, task completion time, and user satisfaction.

January – April 2008

- Writing up the PhD dissertation

Given the variability of factors in the field out of my control, I estimate that there may be up to 6 months worth of cumulative delay in the above plan – and hence, I believe I may finish as late as October 2008.

6. References

[1] Eric Brewer, Michael Demmer, Bowei Du, Kevin Fall, Melissa Ho, Matthew Kam, Sergiu Nedevschi, Joyojeet Pal, Rabin Patra, and Sonesh Surana. ***The Case for Technology for Developing Regions***. IEEE Computer. Volume 38, Number 6, pp. 25-38, June 2005.

[2] Geoffrey Kirkman. ***Out of the Labs and into the Developing World: Using Appropriate Technologies to Promote Truly Global Internet Diffusion***, Journal of Human Development, Volume 2, Number 2, July 2001.

[3] Vodafone Policy Paper: **Africa: The Impact of Mobile Phones**. Vodafone Policy Paper Series Number 2 2005 [http://www.vodafone.com/assets/files/en/AIMP_09032005.pdf]. Accessed 30 August 2006.

[4] T. S. Parikh. ***Using Mobile Phones for Secure, Distributed Document Processing in the Developing World***. IEEE Pervasive Computing Magazine, 4(2):74–81, April 2005.

[5] ***Evaluation of the Prime Minister's Lady Health Worker Program***. Oxford Policy Management Institute. http://www.opml.co.uk/social_policy/health/cn1064_lhw.html Accessed 20th August 2006.

[6] Pakenham-Walsh N, Priestly C, Smith R. 1997. ***Meeting the information needs of health workers in developing countries***. British Medical Journal; 314:90.

[7] Afsar, H. and Younus, M. ***Recommendations to Strengthen the role of Lady Health Workers in the National Program for Family Planning and Primary Health Care in Pakistan: The Health Workers Perspective***. Journal of Ayub Medical College, Jan-Mar 2005. http://www.ayubmed.edu.pk/JAMC/PAST/17-1/HabibYounus.htm Accessed 30th Aug 2006.

[8] Plauche M, Nallasamy U, Pal U, Wooters C, and Ramachandran D. ***Speech Recognition for Illiterate Access to Information and Technology***. Proc. International Conference on Information and Communications Technologies and Development, 2006.

[9] Patwari AK, Raina N. *Integrated Management of Childhood Illness (IMCI): A Robust Strategy*. Indian J Pediatr. 2002 Jan; 69(1): 41-8.

[10] Personal interview with Dr Gul Rashida, Researcher at Population Council, Pakistan

[11] Personal interview with Dr Yousuf Memon, Researcher at Aga Khan University, Pakistan

[12] *The World Health Report 2006 – Working Together for Health.* World Health Organization. http://www.who.int/entity/whr/2006/whr06_en.pdf

[13] Williams J and Witt S. 2004. ***A Comparison of Dialog Strategies for Call Routing***. International Journal of Speech Technology, 7(1): 9-24.

[14] Mansell, R & and Wehn, U. 1998. ***Knowledge Societies: Information Technology for Sustainable Development.*** New York: Oxford University Press

[15] Tongia, R & Subrahmanian, E. 2006. ***Information and Communications Technology for Development (ICT4D) – A Design Challenge?***. Proc. International Conference on Information and Communications Technologies and Development, 2006

[16] Abraham, R. 2006. ***Mobile Phones and Economic Development: Evidence from the Fishing Industry in India***. Proc. International Conference on Information and Communications Technologies and Development, 2006

[17] Shneiderman, B. 2000. ***Universal Usability: Pushing Human-Computer Interaction Research to Empower Every Citizen***. Communications of the ACM. Vol 43, No. 5, Pp 85 – 91.

[18] Chaitali S. ***Effect of Mobile Telephony on Empowering Rural Communities in Developing Countries International Research Foundation for Development (IRFD)***. Conference on Digital Divide, Global Development and the Information Society. http://www.irfd.org/events/wf2005/papers/sinha_chaitali.pdf Accessed on 08/30/06

[19] Geertz C. ***The Bazaar Economy: Information and Search in Peasant Marketing***. American Economic Review, 1978, 68(2), pp.28-32.

[20] ITU. (2003). ***Mobile overtakes fixed: Implications for policy and regulation***. http://www.itu.int/osg/spu/ni/mobileovertakes/Resources/Mobileovertakes_Paper.pdf Accessed 30 October, 2005.

[21] Schultz, T, and Kirchhoff, K. (Ed.). 2006. ***Multilingual Speech Processing***. Elsevier, Academic Press

[22] Schultz, T, Waibel, A. ***The Globalphone project: Multilingual LVCSR with Janus-3***. Multilingual Information Retrieval Dialogs: 2nd SQEL Workshop. Plzen, Czech Republic, 1997, pp. 20-27

[23] Davel, M, Barnard, E. ***Bootstrapping for language resource generation***. Proc. Symposium of the Pattern Recognition Association of South Africa. South Africa, 2003, pp. 97-100

[24] Donner, J. ***Innovative approaches to public health information systems in developing countries: An example from Rwanda***. Presented at "Mobile Technology and Health: Benefits and Risks". Hosted by the Department of Economics, Society, and Geography at the University of Udine, Udine, Italy. (2004)

[25] Prahalad C. ***The Fortune at the Bottom of the Pyramid: Eradicating Poverty Through Profits.*** First printing ed. New Jersey: Wharton School Publishing; 2005:361-379.

[26] UN Educational, Scientific and Cultural Organization's Institute for Statistics http://www.uis.unesco.org Accessed 31 Aug 2006.

[27] UN Educational, Scientific and Cultural Organization's Institute for Statistics. ***Literacy Statistics Metadata Table*** http://www.uis.unesco.org/TEMPLATE/html/Exceltables/education/MetadataLiteracy_April2006.xls

[28] The Information For Development Program. http://www.infodev.org

[29] WRI (2004) ***Lessons from the Field: Overview of ICT Use in Development***. http://www.digitaldividend.org/pubs/pubs_02_tele.htm

[30] ***Community Telecenters: Assuring Impact and Sustainability,*** Development Gateway, 2003 http://topics.developmentgateway.org/ict/sdm/previewDocument.do~activeDocumentId=441647

[31] Mardle E, Telecentres: ***How did we lose the Plot?*** 2003. http://topics.developmentgateway.org/ict/sdm/previewDocument.do~activeDocumentId=440944

[32] Brand P, Schwittay A. ***The Missing Piece: Human-Driven Design and Research in ICT and Development.*** Proc. International Conference on Information and Communications Technologies and Development, 2006

[33] Kumar R, ***E-Choupals: A Study on the Financial Sustainability of Village Internet Centers in Rural Madhya Pradesh***, Information Technologies and International Development, Vol. 1 issue 1, pp.45-73, Spring. 2004

[34] Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Hazen, T.J., Hetherington, L., 2000 – ***JUPITER: A Telephone-Based Conversational Interface for Weather Information***, in IEEE Transactions on Speech and Audio Processing, vol. 8, no. 1, January 2000.

[35] Rudnicky, A., Thayer, E., Constantinides, P., Tchou, C., Stern, R., Lenzo, K., Xu, W., Oh, A., 1999 – ***Creating natural dialogs in the Carnegie Mellon Communicator System***, in Proceedings of Eurospeech, 1999, pp 1531-1534

[36] Tongia R, Subrahmanian E, and Arunachalam V. S., ***Information and Communications Technology for Sustainable Development***, A Report based on two workshops organized by: Carnegie Mellon University, Pittsburgh and Indian Institute of Science, Bangalore; Washington, D.C., 2003, and Bangalore, 2004, Allied Publishers, Pvt. Ltd. Gandhi Nagar, Bangalore, India, c. 2005

[37] Kahssay, H.M., Taylor, M.E., Berman, P.A. *Community Health Workers: The Way Forward*. World Health Organization. http://www.who.int/bookorders/anglais/detart1.jsp

[38] Werner D. 1992. ***Where There Is No Doctor: A Village Health Care Handbook***. Hesperian Foundation.

[39] Hone, K. and Graham, R. 2001. ***Subjective Assessment of Speech-System Interface Usability***. Proceedings of Eurospeech.

[40] Bevan, N. ***Ergonomic Requirements for Office Work With VDT's***, Technical Report 9241-11, ISO. 1994.

[41] Hone, K, and Graham, R. ***Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI)***, Natural Language Engineering, Volume 6 , Issue 3-4, Pp 287 – 303, 2000, Cambridge University Press, New York, USA

[42] HealthLine project website: http://www.cs.cmu.edu/~healthline/

[43] Hofstede G, ***Cultures and Organizations: Software of the Mind***, McGraw-Hill, New York, 1997.

[44] Geert Hofstede Cultural Dimensions.
http://www.geert-hofstede.com/geert_hofstede_resources.shtml

[45] Marcus A. 2002. ***User-interface design, culture and the future***. In: De Marsico, M., Levialdi, S., Panizzi E., Proceedings of the Working Conference on Advanced Visual Interfaces AVI 2002, ACM Press, New York, pp. 15-27.

[46] Marcus A. 2003. ***User-interface design and China: A great leap forward***. Interactions 10(1), 21-25.

[47] Huenerfauth M. 2002. ***Developing Design Recommendations for Computer Interfaces Accessible to Illiterate Users***. Thesis. Master of Science (MSc). Department of Computer Science. National University of Ireland: University College Dublin.

[48] Heffernan C, ***The Livestock Guru: Demand-led knowledge transfer for poverty alleviation.*** Proc. International Conference on Information and Communications Technologies and Development, 2006.

[49] Ackerman S., 2002. ***Mapping User Interface Design to Culture Dimensions***. Proc. International Workshop on Internationalisation of Products and Systems

[50] Medhi I, Sagar A, Toyama K. ***Text-Free User Interfaces for Illiterate and Semi-Literate Users***. Proc. International Conference on Information and Communications Technologies and Development, 2006.

[51] Deo S, Nichols D, Cunningham S, Witten I, 2004. ***Digital Library Access For Illiterate Users***. Proc. 2004 International Research Conference on Innovations in Information Technology

[52] Grisedale S, Graves M, Grunsteidl A, 1997. ***Designing a Graphical User Interface for Healthcare Workers in Rural India***, ACM CHI 1997

[53] Jones D, Hapeshi K, Frankish C, 1989. ***Design Guidelines for Speech Recognition Interfaces***. Applied Ergonomics. 20(1):47-52.

[54] Nass C, Brave Scott. ***Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship***. MIT Press, 2006.

[55] Bickmore T, Giorgino T, ***Health Dialog Systems for Patients and Consumers***. Journal of Biomedical Informatics, 2006 Oct;39(5):556-71.

[56] Tannen R. 1998. ***Breaking the Sound Barrier: Cross Cultural Auditory Display of Browser Processes***. Proc. Fourth Conference on Human Factors and the Web.

[57] Ardila A, Rosselli M, Rosas P. ***Neuropsychological Assessment in Illiterates: Visuospatial and Memory Abilities***. Brain Cognition. 1989 Nov; 11(2):147-66

[58] Castro-Caldas A, Petersson K, Reis A, Stone-Elander S, Ingvar M. ***The Illiterate Brain: Learning to Read and Write During Childhood Influences the Functional Organization of the Adult Brain.*** Brain: A Journal of Neurology (1998), 121, 1053–1063.

[59] Petersson K, Reis A, Askelöf S, Castro-Caldas A, and Ingvar M. ***Language Processing Modulated by Literacy: A Network Analysis of Verbal Repetition in Literate and Illiterate subjects***. Journal of Cognitive Neuroscience 12:3, (2000), 364-382.

[60] Tomko S. 2004. ***Speech Graffiti: Assessing the User Experience***. Carnegie Mellon University, Language Technologies Institute Tech Report CMU-LTI-04-185.

[61] Reddy R. 2004. ***PCtvt: a Multifunction Information Appliance for Illiterate People.*** ICT4B retreat. http://www.rr.cs.cmu.edu/pctvt.ppt 26 August 2004.