Towards Speech Interfaces for Health Information Access by Semi-literate Users

Jahanzeb Sherwani
Carnegie Mellon University

Rahul Tongia
Carnegie Mellon University

Roni Rosenfeld
Carnegie Mellon University

Yousuf Memon
Aga Khan University

Mehtab Karim
Aga Khan University

See next page for additional authors

Follow this and additional works at: http://repository.cmu.edu/compsci
Authors
Jahanzeb Sherwani, Rahul Tongia, Roni Rosenfeld, Yousuf Memon, Mehtab Karim, and Gregory Pappas
Towards Speech Interfaces for Health Information Access by Semi-literate Users

Jahanzeb Sherwani, Rahul Tongia, Roni Rosenfeld
Carnegie Mellon University
School of Computer Science
5000 Forbes Avenue, Pittsburgh, PA 15213
{jscherwan,rtongia,roni}+ijcai@cs.cmu.edu

Yousuf Memon, Mehtab Karim, Gregory Pappas
Aga Khan University
Community Health Sciences Dept.
Stadium Road, Karachi, Pakistan
{yousuf.memon,mehtab.karim,gregory.pappas}@aku.edu

Abstract

Community health programs in many developing countries are very similar: most involve semi-trained, semi-educated health workers (often female), who provide health services in their own communities in a large hierarchically managed system across the country. In Pakistan, recent evaluations of the government's flagship community health program have revealed the need for improvement in health workers' knowledge, and recommended frequent refresher training courses. However, these have not been possible to implement primarily for reasons of cost. It is our conjecture that telephone-based adaptive speech interfaces, tailored to the specific information needs of health workers, are a viable alternative to current information access mechanisms such as handbooks and manuals. We will be designing, developing and testing such an interface (“HealthLine”) in the coming year in Pakistan. Initial results from discussions with health workers (using video prototypes) indicate that speech interfaces may potentially be much more useful for health workers than traditional media.

1 Introduction

Just under half the world’s population, or around 2.8 billion people, currently live on less than 2 dollars a day.1 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 [Mansell, 1998 & Tongia, 2005]. There are hundreds of such projects (as cataloged by the World Bank at http://www.infodev.org), most of which use existing, off-the-shelf technology. 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 [Brand & Schwittay, 2006]. This is eloquently de-

1 This is in purchasing power parity (PPP) terms.
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 [Tongia, 2005].

Most usage of cell phones is for human-human communication, even when interacting via text (SMS). 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 [Parikh, 2005]. 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), 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” [UNESCO, 2006]. 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 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.

At a design level, 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. Spoken language technologies (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. It is our 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. We aim to investigate this hypothesis in the context of community health, of which we will now give a brief overview, after reviewing related research.

1.1 Related Work

There have been a number of approaches to GUI design for semi-literate users. [Huenerfauth, 2002] presents design recommendations for non-literate users of a proposed PDA-like device, with many recommendations involving speech. However, these recommendations are not derived from empirical evidence from evaluations with actual semi- or non-literate users – they are derived from a literature review of research on Western users. [Deo et al. 2004] focused on extending access to digital libraries by non-literate users, and also gave a short list of recommendations for such interfaces. However, usability tests revealed that users were not able to navigate information effectively, and recommended keyword search, audio-based help, and limited the information set to lessen the cognitive load on users during navigation. [Medhi et al., 2006] describes interface design guidelines, and a text-free interface that was performed well on a usability test, but with non-significant results due to the small sample size.

Speech interface research has resulted in a number of systems in various domains. While the most well known speech application is probably desktop dictation, this is just one point on a large multi-dimensional space of potential applications that can be made using speech. These dimensions include: choice of device (e.g., desktop, telephony, smartphone), task (e.g., information access, information entry), length of user training (often zero for commercial applications), vertical domain (e.g., stock prices, news, weather), acceptable user input (constrained, open-ended), interaction style (system initiative, user initiative, mixed initiative) and many others. For instance, CMU’s Communicator travel information system [Rudnicky et al. 1999] and MIT’s Jupiter weather information system [Zue et al. 2000] are two often-cited examples of speech-based information access systems usable over the telephone – these are mixed initiative systems that require zero user training, and accept a large range of user inputs, although as in all speech interfaces, the user is limited at each step in what they can say. Most commercial systems tend to be more constrained, since these are cheaper to build, although exceptions do exist, such as Amtrak’s “Julie” system (1-800-USA-RAIL) which is much more flexible. Contrasted to the above are call routing applications, which are used to direct a caller to a specific operator, given a few utterances (e.g., “I was overcharged last month and I don’t know why” might successfully be routed to the billing department). The major push for speech interfaces in the developed world has come from the call center market, and that is what most research has focused on. However, since the needs of the populations that such systems serve are very different, there are entire domains that are unexplored for which existing research on speech interfaces is woefully inadequate (e.g., access to books through speech). Thus, there is a need for research in domains relevant to emerging regions, targeted towards the specific needs and abilities of users in these regions.

---

2 Dictation is more of a speech recognition problem than a speech interface task. Speech interface research focuses mainly on issues such as task completion rate & time, prompt intelligibility, users’ cognitive load, and other user-level metrics, as opposed to speech recognition metrics such as word error rate.
<table>
<thead>
<tr>
<th>Task Summary</th>
<th>Benin</th>
<th>Botswana</th>
<th>Colombia</th>
<th>India</th>
<th>Jamaica</th>
<th>Liberia</th>
<th>Papua New Guinea</th>
<th>Phillipines</th>
<th>Sudan</th>
<th>Thailand</th>
<th>Yemen</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 First aid, treat accident and simple illness</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2 Dispense drugs</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3 Pre- and post-natal advice, motivation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4 Deliver babies</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Child-care advice, motivation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>6 Nutrition motivation, demonstration</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>7a Nutrition action, weigh children</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>7b Nutrition action, distribute supplements</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Immunization motivation, clinic assistance</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>9 Immunization-give injections</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Family planning motivation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>11 Family planning-distribute supplies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>12 Environmental sanitation, personal hygiene,</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>13 Communicable disease screening, referral</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>14 Communicable disease follow-up, motivation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>15 Communicable disease action</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>16 Assist health centre clinic activities</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 Refer difficult cases to health centre</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>18 Perform school health activities regularly</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 Collect vital statistics</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 Maintain records, reports</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>21 Visit homes on a regular basis</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>22 Perform tasks outside health sector (e.g., agriculture)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 Participate in community meetings</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 1. Roles of Community Health Workers in Various Countries.  
Source: [WHO, 1987]
The TIER group’s Tamil Market project is the first to design, develop and test a spoken language system with semi-literate users in a domain (crop information access) relevant to them [Plauche et al., 2005]. Results from a usability study of the speech interface indicated a difference in task success rates as well as task completion times between groups of literate and semi-literate users. However, the sample size used in the study was too small for significant results. Nonetheless, Tamil Market gives a strong indication that there are differences in skills and abilities between these two user groups, and further research is required to understand the nature of this difference, and to design principles of dialog design targeted towards such users.

[Grisedale et al., 1997] describes a PDA-based interface designed for rural community health workers in India. While this may appear to have similarities to our work, their focus was on information entry, while ours is on information access. Furthermore, their interface was entirely GUI-based – ours is entirely speech-based.

[Anantaraman, 2002] describes a system for data entry as well as access to decision support by community health workers in India. This is in the same domain as our project, and has many similarities to our work. However, our focus is on speech interfaces in this domain, while their approach was GUI-based.

1.2 Health Services in the Developing World

Healthcare is a fundamental, yet often under-serviced need of citizens in developing countries. These regions have the highest maternal mortality and neonatal mortality ratios in the world, and, not surprisingly, also 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. In some countries, especially in Latin America, their effectiveness is quite high, reducing infant mortality to below that of the US. 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” [Pakenham-Walsh, 1997].

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 [Hunt, 2002, & Afsar, 2005]. Many other countries have similar programs with similar issues [WHO, 2006, & Kahssay et al., 1998]. A slightly outdated summary is given in Table 1.

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 their health worker subordinates, and are afraid of losing their job to well-performing health workers [Afsar, 2005].

1.3 Types of Health Information

The book “Where There Is No Doctor” [Werner, 1992] 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 that do not have access to doctors, and has been translated and adapted into more than 100 languages.

The information presented in the book ranges widely 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 diagnosis and treatment of some diseases.

2 Proposed Solution

Of all the forms of Information & Communications Technology (ICT), cell phones are by far the most common in the developing world. While most cell phones have some level of graphical displays, standardized mechanisms for programming these displays do not exist across different cell phone models of the same manufacturer, let alone across cell phones from different manufacturers using different platforms. However, the one standard that all phones (cell phones and landlines) do support is that of placing telephone calls. Based on this observation, our approach is to design, develop and test a telephone-based spoken language interface (HealthLine) for health information access by health workers. While limited, these interfaces may provide the most viable mechanism for health workers (or for that matter, for a large part of the population of developing regions) to access information-based services effectively and efficiently.
3 Initial Fieldwork

Through fieldwork in Pakistan during the summer of 2006, we 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.

3.1 Video Prototypes

Early interactions showed that since most people in Pakistan have never used a spoken dialog system, it was nearly impossible to discuss the possibility of using such systems for information access by health workers. We realized the need for clearly presenting what was meant by a “spoken dialog system” to the various people involved (ranging from researchers to health workers), and created video prototypes for this purpose. Specifically, we developed video prototypes illustrating three different interaction styles (below), wrote scripts for each style, translated them into Urdu, and produced videos in which a “health worker” interacts with each system. These videos 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 (Expert)

2. Natural language question-answer system where user asks any question to the system (QA)

3. User-initiative dialog with keyword search-based information retrieval system (IR)

One misrecognized utterance was present in each of the three systems to demonstrate its detection and rectification. 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 unconstrained natural language input in contemporary dialog systems.

3.2 Health Organizations

Our discussions were primarily with people associated with the institutions listed below. These discussions were preliminary, and we are in the process of designing and implementing a much larger and more quantitative study:

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

A focus group discussion with 20 LHWs from Sultanabad (an urban slum in Karachi) was conducted in the presence of their supervisor. Discussions were also held with various low-level program managers and supervisors in the Sultanabad district.

2) Male & female health workers from the Health & Nutrition Development Society (HANDS), a Karachi-based NGO

Individual interviews were conducted with 3 male and 3 female health workers in Jamkhanda, a rural area two hours by road from Karachi. Discussions were also held 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

A focus group discussion was conducted with 8 LHVs at the FPAP regional office in Karachi. Discussions were also held with the FPAP regional director, as well as the doctor involved in quality assurance and training of the LHVs.

3.3 Findings

The discussions with health workers and other public health personnel revealed the following:

a) Every single health worker interviewed expressed a desire for more refresher courses, and more mechanisms for accessing information. The LHWs 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) Everyone interviewed (health worker or otherwise) asserted that the systems we described would be useful and that health workers would use them – although they varied in the extent and enthusiasm with which they said so. More investigation is required to ascertain whether this is really the case, or whether this response is the result of the aforementioned deference.

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 a male researcher co-conducted with a female assistant, participants engaged deeply but only after the male left the room – even though the female assistant typed notes into the same laptop the male researcher was using, and that it was clear that the two would confer later. This incident underscored the need
for a female-mediated strategy for all interactions (interviews, contextual inquiry, prototype evaluation & feedback, usability testing, etc.). In the absence of appropriate female researchers, one potential methodology that may be employed is a setup where a female mediator is wired up and receives instructions from other researchers during these interactions as necessary; this needs to take into account issues of privacy, transparency, and data retention.

When asked to rank preference for the three video prototypes, QA was consistently ranked the lowest. Participants differed in their preferred rankings for Expert and IR – in general, the more trained a health worker was, the more likely she was to prefer IR over Expert. As most of these interactions occurred in one-time, group settings with a show of hands indicating preference, this data is not only preliminary. Further inquiry using a more rigorous testing methodology will be utilized for a more thorough needs assessment.

4 Next Steps

Based on the above experience, we are now designing a larger, more rigorous study to understand health workers’ contexts, backgrounds, information needs, comfort with technology, and literacy skills. However, initial feedback indicates that a hybrid solution incorporating both the Expert and IR models may be the optimal strategy. Through a process of design, development, testing and feedback, we will iterate the system to a point where field usability testing is feasible. We will then evaluate the system against existing mechanisms of health information access in use by these health workers (such as textbooks) and on both quantitative metrics (task completion rate, time to completion) and qualitative metrics (user satisfaction & preference). By analyzing these metrics with respect to bio-factors such as literacy and years of experience, we hope to show exactly under what circumstances a speech interface is most effective. Additionally, we hope to contribute to a better understanding of dialog design principles for semi-literate users, which would be useful for speech interface designers in different domains.

Acknowledgments

This work has been funded by the External Research & Programs group at Microsoft Research, through the Digital Inclusion RFP.

References


