HealthLine: Speech-based Access to Health Information by Low-literate Users

J. Sherwani, N. Ali, S. Mirza, A. Fatma, Y. Memon, M. Karim, R. Tongia, R. Rosenfeld

Abstract—Health information access by low-literate community health workers is a pressing need of community health programs across the developing world. We present results from a needs assessment we conducted to understand the health information access practices and needs of various types of health workers in Pakistan. We also present a prototype for speech-based health information access, as well as discuss our experiences from a pilot study involving its use by community health workers in a rural health center.

Index Terms—speech recognition, dialog systems, developing regions, health information, community health, Pakistan

I. INTRODUCTION

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

J. Sherwani is a graduate student in the Computer Science Department at Carnegie Mellon University (email: jsherwan@cs.cmu.edu).

N. Ali is a graduate student in the Department of Development Sociology at Cornell University (e-mail: nosheen.ali@cornell.edu).

S. Mirza is the Deputy Executive Coordinator at the Health and Nutrition Development Society, Pakistan (e-mail: sarwat.mirza@hands.org.pk).

A. Fatma is the Senior Manager of Monitoring and Evaluation at the Health and Nutrition Development Society, Pakistan (e-mail: anjum.fatma@hands.org.pk).

Y. Memon is a Senior Instructor at the Community Health Sciences Department at Aga Khan University, Pakistan (e-mail: yousuf.memon@aku.edu)

M. Karim is a Professor at the Community Health Sciences Department at Aga Khan University, Pakistan (e-mail: mehtab.karim@aku.edu)

R. Tongia is a Senior Systems Scientist in Program in Computation, Organizations, and Society (COS), School of Computer Science, and the Dept. of Engineering and Public Policy at Carnegie Mellon University (e-mail: tongia@cmu.edu). He is also a Senior Fellow at the Center for Study of Science, Technology, and Policy (CSTEP) at Bangalore, India.

R. Rosenfeld is a Professor in the School of Computer Science at Carnegie Mellon University (e-mail: roni@cs.cmu.edu).

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. 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" [1].

Given the low-literate background of many community health workers (CHWs), as well as the difficulties of carrying around large health manuals by literate CHWs, speech-based information access becomes a potentially viable alternative. By making a telephone call to an automated system, and conversing with it in her local language, the CHW could gain access to the information she seeks. The system would need to be easy to learn, and easy to use, yet should enable access to a potentially vast amount of health information, as quickly and as effectively as possible. Designing, developing, and evaluating such a system, and generating general principles of dialog design for low-literate users are the core goals of the HealthLine project, which we present here.

Before working to create a solution, the first research task is to understand the nature of the health information needs of CHWs, and how well these needs are met by existing methods. Alternative mechanisms for information access would then be designed according to these needs.

Additionally, the deployment of spoken language interfaces in developing countries is known to be a daunting task, because of the scarcity of various types of language specific resources [2]. Our design would need to deal with these scarcities, and present some solutions for each of them.

This paper is organized as follows. We discuss related work in Section II. In Section III, we give an overview of community health in Pakistan. Section IV details our needs assessment study, involving field-based interviews of various CHWs. In Section V, we present the first HealthLine prototype, and in Section VI, we present results from a pilot study involving the use of this prototype by CHWs in a rural health center in Sindh, Pakistan. We discuss future directions in Section VII.

II. Related Work

There have been a number of approaches to GUI design for low-literate users. [3] 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. [4] focuses on extending access to digital libraries by non-literate users, and also gives a short list of recommendations for such interfaces. However, usability tests reveal that users were not able to navigate information effectively, and result in recommendation for keyword search, audio-based help, and limiting the information set to lessen the cognitive load on users during navigation. [5] describes interface design guidelines, and a text-free interface that performed well in a usability test.

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, Carnegie Mellon University's Communicator travel information system [6] and MIT's Jupiter weather information system [7] 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, acceptable user input is limited at each step. Most commercial systems tend to be more constrained, since these are cheaper to build, although exceptions do exist, such as Amtrak's "Julie" system 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. 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 (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.

The Berkeley's TIER group's Tamil Market project was the first to design, develop and test a spoken language system with low-literate users in a domain (crop information access) relevant to them [8]. Results from a usability study of the speech interface suggest a difference in task success rates as well as in task completion times between groups of literate and non-literate users. However, the sample size used in the study was too small for statistical significance. 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 derive principles of dialog design targeted towards such users.

[9] 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 is on information entry, while ours is on information access. Furthermore, their interface is entirely GUI-based – ours is entirely speech-based.

[10] 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 is GUI-based.

[11] describes the iterative & collaborative design process for and evaluation of a GUI targeted to low-literate users for managing community-based financial institutions in rural India. While the principles of GUI design do not carry across well to speech interface design, the collaborative design process described has lessons highly relevant to all interface design in such contexts.

[12] describes VoicePedia, a purely telephone-based speech interface for searching, navigating and accessing the entire Wikipedia web-site. An evaluation comparing VoicePedia with a GUI-based smartphone equivalent shows comparable task success across interface conditions, although the (highly literate) users in the evaluation invariably preferred the GUI alternative.

[13] gives an excellent review of the potential contributions of CHWs in the developing world.

III. Community Health in Pakistan

In Pakistan, there are various community health programs managed by both governmental and non-governmental organizations, which vary in the number of CHWs employed, the population each CHW serves, the length of training and refresher courses given to the CHWs, the health domains covered, as well as the literacy requirements for inducting CHWs.

The government's Ministry of Health manages the largest community health program in the country – the Lady Health Worker Program – which has trained around 80,000 Lady Health Workers (LHWs), each serving a population of 1,000. These health workers form the backbone of community-level health services in the areas of maternal and child health in communities across Pakistan. However, there are many communities that do not have LHWs. A minimum of 8 years of schooling is a prerequisite to joining the LHW program, although in some cases this is overlooked. LHWs receive 3 months of basic training.

The knowledge level of LHWs has been evaluated as 69% of the total range of information that they are supposed to have command over, and hence, improvement in the LHW knowledge level is suggested as a foremost factor for increasing their service delivery [14].

In addition to the LHW program, various NGOs manage other community health programs. The Health and Nutrition

Development Society (HANDS) manages three such programs: Community Mid-Wives (CMWs), Community Health Attendants (CHAs) and Community Health Workers in the PAIMAN project (PAIMAN-CHWs). CMWs are female, while CHAs are male[1] – both have a requirement of 10 years of schooling, and receive 1 year of training, which is much more in-depth than the LHW training. HANDS has trained 140 CMWs since their program started in 2002. PAIMAN-CHWs on the other hand are often taken from communities without access to schooling, and hence there is no programmatic requirement of education for these health workers. They receive 6 days of training consisting of much simpler messages than any other category of health workers – although 171 such workers have been trained since the program started in 2006.

Access to health information for community health workers in Pakistan is primarily provided through trainings and workshops employing a variety of written material. However, these trainings are impeded by a number of barriers such as:

- Non-availability of trainers as they are frequently transferred or move to better opportunities
- Non-availability of training material for all participants
- Addition of new technical content does not easily filter down to trainings at the community-level
- Trainers themselves have inadequate knowledge and skills
- Low literacy skills

IV. HEALTH WORKER NEEDS ASSESSMENT

A. Overview

Our needs assessment began in February, 2007, and lasted 3 months. Its primary purpose was to help us understand CHWs' background, their health information practices and needs, their literacy and comprehension skills, and their perceptions of their informational needs.

The key goals of the study were:

- To understand the range of health workers operating at the community-level in Karachi, and the nature of the health services that they provide
- To grasp the perceptions of health workers regarding the prominent health problems facing their community
- To identify the specific areas of health in which information is most needed by community health workers
- To assess the Urdu reading and comprehension abilities of health workers, as a means to evaluate their literacy skills
- To explore the preferences of health workers regarding the design and use of HealthLine, and gauge its overall need and suitability in the local context

B. Methodology

A cross-sectional survey based on a sample of convenience was conducted, using two structured questionnaires with both close and open-ended questions – one was a detailed questionnaire used in the initial in-depth interviews with 12 health workers to get a richer sense of the lives and practice of health workers, while the other was a shorter questionnaire administered to 23 health workers which focused mainly on the key goals of the study. Hence, a total of 35 community health workers were interviewed (32 female, 3 male).

These 35 were drawn from the following groups:
- 22 government LHWs
- 7 UHP-CHWs: UHP[2] community health workers
- 3 HANDS CMWs
- 3 HANDS CHAs

These health workers belonged to the low-income settlements of Rehri Goth, Sultanabad, Hijrat Colony, Jamkanda, and Shedi Goth, all of which are located in or near Karachi. As the emphasis in the study was on LHWs, the questions were designed around the LHW curriculum.

C. Profile

The mean age of health workers was 30.3 years. Almost 70% had completed schooling till at least 10th grade, pointing to a potentially high level of literacy. Health workers had been serving in the area of community health for an average of 6.1 years. The mean number of people whom each health worker provided health services to was 1100, and the mean number of hours that they worked in a day was 4.2. According to health workers' perceptions, on average 10 people visited their house for advice or treatment in the previous week. 77.1% had access to some form of phone service, which was more likely to be a cell phone than a land-line.

D. Use of Training Manual

Health workers were asked what health topics they most recently consulted their manual for. The responses are shown in Table I.

TABLE I
HEALTH TOPICS CONSULTED IN THE MANUAL

Health Topic	Responses
Management of Diarrhea - e.g. dose of O.R.S.	5
Management of Pneumonia - e.g. dose of cotrimaxazole	3
Pregnancy care - e.g. what to do if a pregnant woman's BP goes high and she has asthma?	3
Family Planning - e.g. under what conditions should a woman abstain from taking injections for birth control	3
Hepatitis	1
Treatment of worms	1
Eclampsia	1
Delivery complications	1

[1] The underlying philosophy is to train one man and one woman from the same community on complementary health topics.

[2] Aga Khan University's Dept. of Community Health Sciences has an Urban Health Project (UHP) which has trained a number of health workers over its 10 year history. However, many of them are now involved in data gathering roles for research projects, and do not provide health services.

E. Preferred Use of HealthLine

In designing the system, we needed to know whether CHWs would prefer to use it for immediate help when a patient presents them with an information need, or for offline self-learning so that she may be better equipped for future situations. The results are shown in Fig 1.
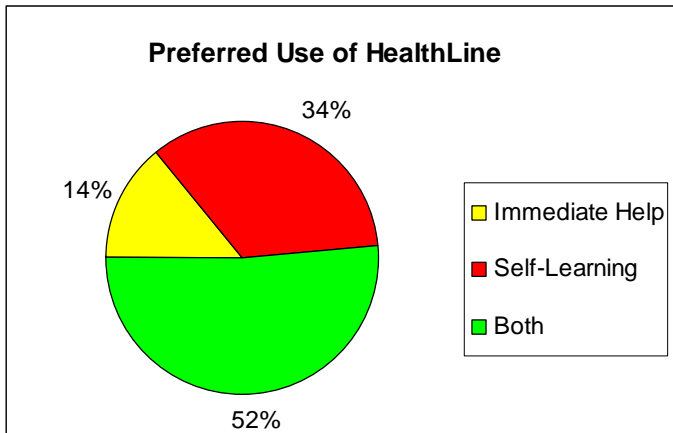
Preferred Use of HealthLine

Fig 1. Preferred use of HealthLine.

Health workers who preferred a system that would provide on-the-spot information for addressing patient queries cited reasons such as:

"In many circumstances, facing patients can be overwhelming. Sometimes, my mind goes blank and I can't think properly. I can't even remember the information that I already know." – LHW, Hijrat Colony

At the same time, 75% of these respondents said that they would prefer not to call HealthLine in front of the patient, reasoning that the patient would think them incompetent, and that the information might be troubling or not relevant, making the patient all the more anxious. The remaining 25% felt that it would be better in front of the patient as they might be needed to answer certain questions during the interaction, and also the patient would feel grateful that the health worker is double-checking her knowledge.

F. Urdu Reading and Comprehension Ability

Health workers were asked to read a paragraph in Urdu from the LHW manual during the interview as a literacy test. 25% had difficulty in reading the text, which meant that one or more of the following factors applied to them:

- Took a much longer time to read the text as well as made mistakes
- Took reasonable time to read the text but skipped 1-3 words at a time, while making mistakes.
- Inability to read slightly difficult Urdu words in the text (e.g. "منفی" which means negative, and "ردِ عمل" which means reaction).
- Replaced the end of sentences (where the verb occurs in Urdu) with their best guess, suggesting that these health workers try to get the gist of the sentence by concentrating on the middle and fill in the rest.

Health workers were then asked to answer two straight-forward questions based on the text that they read – 30% exhibited significant difficulty, suggesting weak comprehension ability. However, the inability to give correct answers can perhaps be partly attributed to a tendency to answer questions based on instinct or prior knowledge, rather than on the reference material. It should also be noted that the health workers who showed average-to-poor comprehension ability did not necessarily have lower levels of schooling: 21.7% of such health workers had completed school till 10th or 12th grade.

G. Self-perceived Information Needs of Health Workers

Health workers were asked what topics they would like more information on – regardless of whether it was in their existing curriculum or not. While the actual list had 24 unique topics, only 7 topics had five or more requests, as shown in Table 2.

TABLE II
PERCEIVED KNOWLEDGE GAPS

Topic	Requests
Hepatitis	18
Diabetes	12
Sexual diseases	10
HIV/AIDS	9
Deliveries	8
Antenatal Care	7
High blood pressure	5

We analyzed the data to check if a particular information need was limited to a specific place or type of health worker. Our conclusion was that the information on the above topics was required across the board.

It is interesting to see how the health information needs revealed by this study are covered in the LHW manual. Diabetes is completely absent from it, as it is not part of the LHW curriculum. Hepatitis – which constitutes the single largest area in which health workers say they need information – is devoted less than half a page in the 361-page manual. A wide variety of sexual diseases are covered in two pages, with the exception of AIDS, which gets 5 pages by itself. This lack of written material and coverage might be one reason why health workers most feel the need for information on these topics. An alternative interpretation is that the LHW program has effectively trained its workers on the core components of the curriculum, and hence their information needs are primarily in topics that they haven't been trained on.

H. Self-Perceived Difficulty Level of Key Health Topics

Health workers were asked to rate the difficulty level of key health topics: maternal health, nutrition, family planning, immunization, diarrhea, pneumonia, and sexual diseases, on a 5-point Likert scale ranging from "Very Easy" through "Very Difficult". These topics were chosen because of their importance to LHW practice, and hence, this was a close-ended question. "Difficulty level" was defined in terms of how tough and challenging different topics are for health workers to understand, remember, and practice. The results are summarized in Fig. 2.
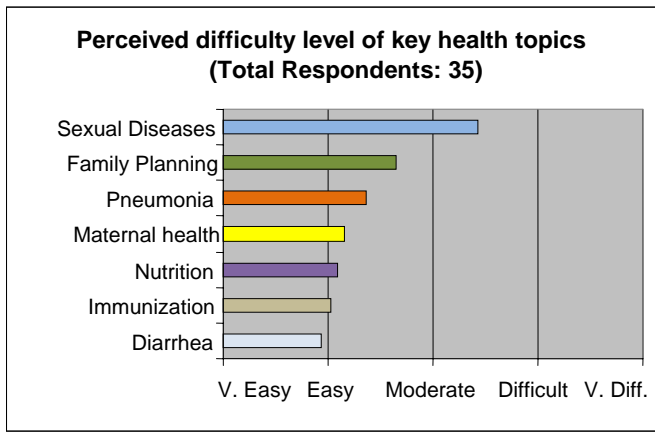
Fig. 2. Perceived difficulty level of key health topics.

Sexual diseases emerged as the most difficult topic with a mean value of 3.4. Put differently, 83% of health workers said that the difficulty level of sexual diseases was average, difficult, or very difficult. This corroborates the results of the previous section, in which sexual diseases emerged as a key area in which health workers perceive a need for more information.

I. Performance-based Difficulties in Addressing Health Issues

In addition to asking health workers what they found difficult, we looked at performance-based metrics of where they fail in their knowledge level. An evaluation of the LHW program administered a knowledge test to 500 LHWs in 2002 [14]. Through this test, specific areas of weakness in the knowledge base of LHWs were identified, mainly:

- Inability to remember the immunization schedule – less than half the LHWs interviewed could name the four EPI vaccines (BCG, DPT, polio, measles), give the correct number of doses, and identify the correct age at which the doses are given.
- Lack of knowledge about the correct doses of medicines for common illnesses. For example, less than one quarter of LHWs interviewed could state the correct dose and duration for a course of Cotrimoxazole for a child with pneumonia, even though they were encouraged to use their training manual and medicine box to answer the question.
- Inability to correctly conduct case-based analysis of diarrhea and pneumonia using the classification and management protocols that LHWs have been trained in. For example, 79% could not identify simple pneumonia. 35% could not identify a case of severe dehydration, while 29% could not identify severe pneumonia.

We also included knowledge-based questions during in-depth interviews of 12 health workers. They were asked about danger signs during pregnancy, danger signs of newborns, vaccination schedules, and case-based management of diarrhea and pneumonia. Case-based analysis was most difficult for health workers, with 66% of health workers failing to diagnose a case of severe pneumonia and a different 66% failing to diagnose a case of simple pneumonia. Health

workers tended to classify severe pneumonia as simple, and simple pneumonia as a case of cough/cold. Health workers fared better in diarrhea management, as only a small percentage (16.7%) failed to classify a case of mild to moderate dehydration, and the same percentage failed to identify a case of severe dehydration. Even though the numbers are the same, it was not the same health workers that had difficulty in classifying simple vs. severe cases of these diseases.

J. Summary

While community health workers in Pakistan are required to be literate, their reading and comprehension skills in Urdu show a variation that suggests that book-based material may not be as easily understood and learnt as normally considered. Additionally, during the reading test it was discovered that a number of workers were unable to understand some Urdu words used in their health manual. Hence, it is imperative that alternative information content should use simple Urdu. Finally, there are significant gaps in their health knowledge. The needs assessment highlighted a number of topics within their existing curriculum that workers would like to have reinforced, as well as other topics that they are keen to learn.

V. THE HEALTHLINE PROTOTYPE

A. Health material

The needs assessment presented a clear list of topics on which to focus the development of the system. The primary sources for health material were two Karachi-based public health organizations, the Health and Nutrition Development Society (HANDS) and Aga Khan Health Services Pakistan (AKHSP). The list of material accumulated is as follows:

1. HANDS' CHW manual: "Messages for CHWs" (3 pages)
2. AKHSP pamphlets on diarrhea and pneumonia (4 + 3 pages)
3. The LHW manual (350 pages)
4. A summary of main messages for LHWs (20 pages)
5. AKHSP pamphlets on hypertension, diabetes, leucorrhea (8 + 9 + 3 pages)
6. HANDS' CMW manual (526 pages)

AKHSP's pamphlets had been designed for literate, urban readers, and had been tested extensively on them. AKHSP staff cautioned against the use of the materials on audiences that the pamphlets had not been tested on, such as low-literate CHWs. As a precaution, prior to putting the text in the system, minor changes were made to the text to simplify meaning and sentence construction. For the prototype, only the first two items in the above list were put into the system.

B. Prototype System Architecture

As mentioned earlier, the creation of a speech interface in a new language requires the existence of a number of resources [2], specifically:

1) Basic linguistic and cultural knowledge
2) Linguistic resources

3) Software tools
4) Guidelines for the design of user interfaces

We noted that the Tamil Market project [8] spent considerable effort in creating acoustic models for Tamil, involving recording Tamil speakers speaking out certain words, which proved challenging with low-literate speakers. We did not wish to have to go through such a process unless absolutely necessary. Additionally, their approach led to a small vocabulary, single-word recognizer – which would limit dialog design choices considerably. One significant benefit of their approach was low word error rate (as low as 2%), which would be difficult to improve on.

Our solution was to use a commercial package, Microsoft Speech Server 2007 Beta (MSS), which includes a speech recognition engine, a speech synthesis engine, and a dialog management architecture, along with dialog authoring tools. However, MSS supports only a few languages, not including Urdu. To solve this problem for speech synthesis, all audio was recorded as individual prompts from an Urdu-speaking voice talent. It was then possible to encode prompt text in Urdu in the code (through Unicode), so that there was no extra layer of mapping required – running an MSS function such as `Prompt.AppendText("سلام")` would work correctly.

For speech recognition, we tried a "poor man's speech recognizer" approach. Most (if not all) speech recognition systems allow the definition of new words in a lexicon file – this file contains each word's textual representation, and its phonetic representation(s). Our approach was to define each of the Urdu words that needed to be recognized by the system, and hand-code pronunciation was defined using US English phonemes. For instance, the word "سلام" (salaam) would have been described in the lexicon as سلام→S AH L AA M. Once all the words were defined this way, context-free GRXML grammars could be created to recognize single word utterances, more complicated utterances, or even conversational recognition using the MSS Conversational Grammar Builder, depending on the dialog need. Again, since the words had been defined in Urdu Unicode, there was no need for an arbitrary mapping – which simplified coding and debugging.

One major limitation of this solution is that it cannot recognize phonemes that do not exist in the language whose phoneme set is used. For example, the غ sound (a guttural "g" fricative phoneme, called GH) has no equivalent in the US English phoneme set, and so a fallback such as AH might be used. Preliminary results indicate that using such fallbacks still leads to correct recognitions, although the recognizer confidence is lower on these words. However, this limitation should primarily prove a problem in situations where the missing phoneme has to be differentiated from the fallback equivalent – for example, if trying to recognize the difference between words such as "GH IH Z AA" and "AH IH Z AA", as these would be represented identically using the US English phoneme set.

C. SIP server

To run a dialog system over the phone network, MSS requires a SIP-over-TCP VOIP end-point. All of the MSS-recommended hardware telephony products were digital line-based (e.g., T1/E1), and comparatively expensive ($3000+). Additionally, the costs of setting up an E1 line in Pakistan are significantly higher than setting up an analog line. Asterisk, on the other hand, is a widely used, open-source VOIP platform, which significantly lower hardware costs (<$1000), and also supports analog phone lines. However, Asterisk implements only SIP-over-UDP, and none of the available SIP TCP/UDP proxies we tried were able to connect Asterisk and MSS. Our solution was to write a proxy for this purpose, sending SIP messages over UDP to Asterisk, and over TCP to MSS. Due to complexities in the SIP protocol, it was not adequate to simply relay a message from one end-point to the other – this is why available proxies did not work. Instead, the entire SIP conversation had to be orchestrated as two separate conversations: one between Asterisk and the proxy, and the other between the proxy and MSS. All audio is passed directly between Asterisk and MSS, however. The system architecture is shown in Fig. 3.
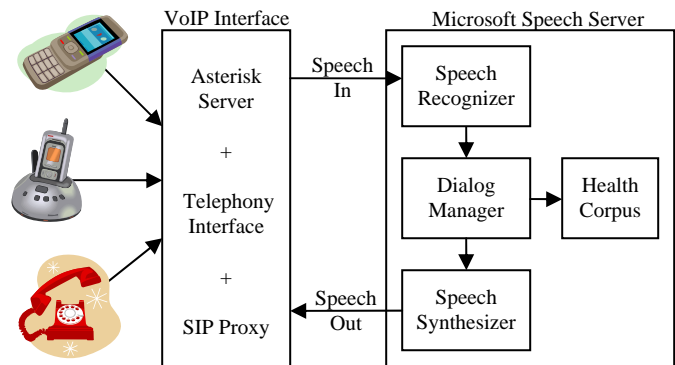

Fig. 3. Block diagram of HealthLine system components.

D. Urdu language text tools

While Unicode-based Urdu language support has recently been standardized and made available, a proprietary application called InPage has been the *de facto* standard for Urdu word processing for the past decade, with its own encoding format. These needed to be converted to Unicode, although none of the available online tools for this purpose were completely accurate. Some of the discrepancies occurred as a result of fundamentally different methods of representing the same letter with its diacritics between InPage and Unicode. Again our solution was to write such a converter ourselves.

E. Large corpora Urdu language audio recording

Recording the reading out of multiple pages of Urdu text required another component to present each sentence to the voice talent and record them speaking it aloud. Such software is commonly found in speech synthesis packages (as well as in MSS) – however, none of the available options were ideal for our use. First, these applications displayed each sentence in

isolation – however, we wanted the voice talent to be able to see the rest of the text, so that she could have an idea of how to place emphasis based on context (paragraph end, bulleted list, etc.), even though the audio for each sentence was recorded separately. Next, the text had to be in a visually pleasing Unicode Urdu font – and many applications either didn't support Unicode, or if they did, did not allow for changing the font. Our solution again was to write our own component. All audio was recorded in 44.1kHz, 16-bit, and downsampled to 8kHz for use in a telephony environment. Recordings were done using an M-Audio Fast Track Pro USB audio device, with a Shure SM-58 microphone, and a pop filter.

F. Dialog design

While our primary research interest is in investigating various alternative dialog strategies, we decided to create a very simple prototype for the purpose of pilot testing. Thus, the aim was to introduce CHWs to a simple system, and gradually grow it based on their feedback – rather than starting off with a complex system that is unable to meet their needs. We now present the dialog design used in the prototype. All prompts and valid recognition options have been translated from Urdu to English below.

The interaction begins with the system saying: *Hello, I'm Dr Jameela, and I'm here to give you whatever health information you need. What topic do you want information on: Diarrhea, Pneumonia, or Important Messages for Health Workers?* The system will then accept any of the above three topics as a valid response.

Once the topic is chosen, the system says (for example): *Diarrhea: This topic has 7 sections, when you hear the one you want, just say it: 1 What is Diarrhea [pause], 2 Causes of Diarrhea [pause], …, 7 The Third Principle of Treatment.* At this point, users can either repeat the number, or the section name.

Once the section has also been chosen, the system begins reading out the text, and gives the user options on how to continue after all the text has been read out. By design, the sections were kept purposefully small in the prototype, since there were no options for navigating within the section. Specifically, the system would say: *What is Diarrhea? Diarrhea is a …[continues to the end of the section]. To hear this again, say "repeat"; to choose a different section, say "different section", or for a different topic, say "different topic". If you're done, say "goodbye".* Thus, at the final step, the only valid options are "repeat", "different section", "different topic" and "goodbye".

In addition to the above dialog options, there are a few globally active commands and behaviors. First, the user can say "what can I say?" at any time, which the system responds to by giving a list of currently valid options, as well as tersely explains the current status of the dialog.

At any step in the dialog, if the recognizer has low confidence on what was spoken, the system asks for a confirmation, such as: *I think you said Pneumonia, am I correct?* – to which a valid response would be "yes" or "no"

including a few variations, equivalent to "yeah", "that's correct", etc.

Finally, if sound input level is too low, the system replies with *I'm sorry, I didn't hear you*, as well as an escalating silence prompt (i.e., on two successive silences) of *I'm sorry, I still didn't hear you, please speak louder*. Similarly, on instances where the recognizer generates no valid hypothesis, the system says *I'm sorry, I didn't understand you, please say that again*, and on an escalating non-recognition, *I'm sorry, I still didn't understand you, please say that again clearly*.

VI. PILOT STUDY

A. Study Design

The goal of the pilot study was to test the system with a handful of low-literate health workers to understand the issues that arise in the use of a relatively simple speech interface. CHWs that were part of the Pakistan Initiative for Mothers and Newborns (PAIMAN)[3] were chosen for this study. These workers are located in rural areas, and have a much lower literacy level than the workers we interviewed in and near Karachi. They are trained on 7 themes, each of which has a number of messages they need to deliver to their communities. These were the lowest-literate of the health workers, and hence were ideal for an initial pilot test.

The study design was as follows: each participant was given an introduction to the study, with an emphasis on the fact that it was the *system* being tested, not the CHWs themselves. Demographic information was taken, and then the participant was asked to read one line from their training manual, as a quick literacy evaluation. CHWs that were able to read were then asked to answer a simple health question using the text pamphlets. Participants were then given a brief verbal explanation of the working of the system, and then asked to find the answer to a health question using the system. Finally, they were asked to rate the system along 6 dimensions on a 5-point Likert scale, based on the SASSI set [15].

B. Results

The actual study took place at the Basic Health Unit in Wahi Pandi, in Dadu District, Sindh. Due to a miscommunication with local coordinators, there wasn't an available telephone to dial the system, and so we used a USB audio device that closely resembles a phone plugged into a laptop running MSS as the interface mechanism.

9 CHWs participated in the pilot. 3 of them had not received any formal schooling, and hence had minimal fluency in Urdu since their native language was Sindhi. These 3 were not able to continue with the study.

Of the remaining 6, 1 had minimal reading ability, while 5 were able to read reasonably well or effectively. When given a question to answer from the pamphlet, these literate 5 were successful in finding the answer within 60 seconds, as only one was able to skim the text and jump to section of interest –

---

the other 4 used their finger to point to the first word on the first page, and read linearly from that point on until they found the answer in the relevant section (which was the second section out of seven sections). Thus, in a task with the section placed at the end, these 4 would have taken significantly longer.

When these 6 participants used the speech interface, 5 of them were able to successfully hear and report the information they heard. The 1 participant who did not succeed failed mostly due to an echo cancellation problem, which meant that the informative prompts that would have explained the dialog to her were cut short time and again because of the system hearing itself. Most significantly, the low-literate participant who was not able to read pamphlets at all *was* able to use the speech interface successfully.

Finally, the questions asked after the interaction were as follows: How useful did you find the system? How difficult was it to use? How annoyed were you while using it? How many errors did you feel it made? How well did you know what to say to the system? Was the interaction too long? The mean responses are shown in Fig. 4.
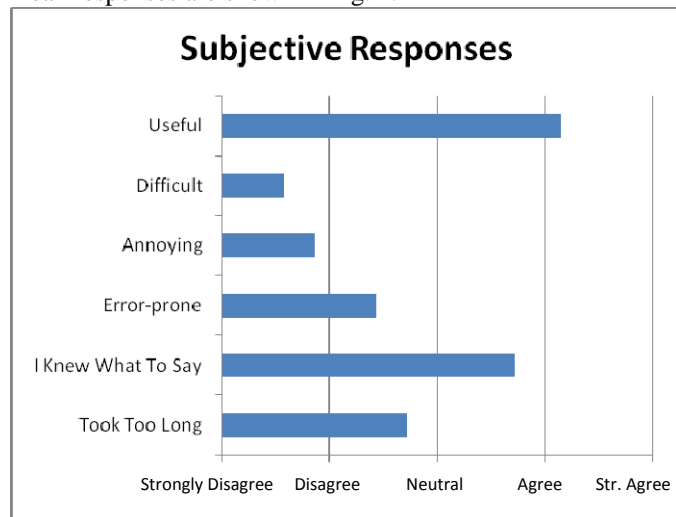

Fig. 4. Subjective responses to Likert-scale questions.

C. Discussion

In a pilot setting, our inability to hear the conversation was very detrimental, as we couldn't tell until post-hoc analysis whether a conversation was failing because the user was confused, or because the system was hearing extraneous noise (including its own voice). In a more rigorous experiment, this would not be a concern, but for a pilot, suspending an interaction, explaining the issue, and then restarting the dialog could be useful.

Environmental noise was a frequently-occurring problem, and one method we tried showed promise: by lowering the input volume on the microphone, the system would only hear the loudest of sounds. Many times, this would mean that it would not even hear the user speak, but after triggering the silence prompt even once, CHWs were quickly entrained to speaking louder so as to avoid hearing the prompt again.

We also realized that having a short, implicit confirmation after each step (the repetition of the chosen item, to inform the

user that it was correctly recognized, as in the bold text in the following dialog: *What topic do you want? Pneuomonia, Diarrhea, or—* [user: Pneumonia!] ***Pneumonia:*** *this topic has 7 sections*) was counter-productive, as users almost always said an enthusiastic "yes!" when they heard this, which the system did not expect, which led to a series of misunderstandings. In fact, this "feature" was put into the dialog with the assumption that our target audience was somewhat docile and not outspoken, and so might not respond to a prompt unless explicitly asked to do so. Since then we have realized the need to reverse our assumptions in this regard.

Instead, it appears that changing the interface metaphor used and making the response more concrete, if slightly lengthier, might be beneficial. Thus instead of saying "Pneumonia, this topic has 7 sections", the system would say "Ok, I've opened up Pneumonia book, and I see 7 sections". This might make it easier for users to understand the prompt, as well as the metaphor of opening a book on a certain topic.

Also, by re-using the same audio recordings for topic names (e.g. "Pneumonia") both in questions and in statements, most users believed that the question was being repeated, since the prosody was identical. This suggests re-recording prompts for different contexts, with appropriate coaching for the voice talent to record the prompts effectively.

We also realized the need for creating a video explaining and demonstrating the system for a new user. While a verbal explanation of the system's usage was adequate, having new users hear and see an actual conversation would make it much clearer, and might also make it easier to remember global commands such as "what can I say?".

While the initial steps in the dialog are conversational, the final one consists of a one-way lecture, with the system's voice reading the text out for the entire section, which seemed less than optimal. Three users employed the repeat command to hear the text again. One user tried to echo the system's reading out of a bulleted list, as a mechanism to help her remember it more effectively (presumably as compensation for inability to write), although this resulted in her interrupting the system's speech output twice. There is a need to develop this part of the dialog further.

VII. FUTURE DIRECTIONS

The prototype was the first step towards creating a larger system containing more information, and more functionality, targeted towards more categories of CHWs. Similarly, the pilot was the first step towards testing the system with more categories of health workers, and with different kinds of health information. We intend to test more complicated information with better-trained CHWs, and to understand the difference in use and preference between groups of CHWs and between literacy levels. We are interested in seeing whether the use of HealthLine by a CHW can result in an increase in her knowledge of health information that she has already been trained on. We also aim to study the possibility of HealthLine improving the potential of a given category of CHWs to perform health services beyond their training – as it may enable low-literate health workers to access more

information than was possible for them with traditional technology (e.g., whether or not a low-literate PAIMAN-CHW can correctly classify the severity of a given case of pneumonia using HealthLine, which is normally a task entrusted only to LHWs).

Within the dialog system, we will be looking at alternative mechanisms to present the information to the user, rather than reading it out sequentially and giving an option to repeat at the end. Similarly, instead of explicitly asking the user to choose the topic of interest (which would not scale well to a large number of topics), we aim to incorporate keyword-based search capabilities, and will test the use of voice-based search (as in [12]) as the main mechanism through which to locate the relevant information.

The relationship between schooling and linguistic competence in Urdu was unfortunate, as we had no Urdu-speaking, yet non-literate health workers in our study. It appears that this tight correlation might be the case across rural Sindh, and while this is not a significant issue in the long-term (creating a similar system for Sindhi is not difficult), for the purpose of our research, we intend to continue with Urdu, and so will be seeking out non-literate Urdu-speaking community health workers to explore the link between literacy and the ability of such users to use the system effectively.

REFERENCES

[1] N. Pakenham-Walsh, C. Priestley, and R. Smith. *Meeting the Information Needs of Health Workers in Developing Countries*. British Medical Journal, 314:90, January 1997.

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

[3] M. Huenerfauth. 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.

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

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

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

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

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

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

[10] V. Anantaraman, et al. *Handheld computers for rural healthcare, experiences in a large scale implementation*. In Proceedings of Development By Design, 2002.

[11] T. Parikh, G. Kaushik, and A. Chavan, *Design studies for a financial management system for micro-credit groups in rural India*. Proc. of the ACM Conference on Universal Usability, ACM Press (2003).

[12] J Sherwani, Dong Yu, Tim, Paek, Mary Czerwinski, Yun-Cheng Ju, Alex Acero, *VoicePedia: Towards Speech-based Access to Unstructured Information*, Interspeech 2007, Antwerp, Belgium.

[13] A. Haines, D. Sanders, U. Lehmann, AK Rowe, JE Lawn, S. Jan, DG Walker and Z Bhutta. *Achieving child survival goals: potential contribution of community health workers*. The Lancet 369(9579): 2121-2131. 2007

[14] S. Hunt. *Evaluation of the Prime Minister's Lady Health Worker Programme*. Oxford Policy Management Institute, http://www.opml.co.uk/social_policy/health/cn1064_lhw.html. March 2002.

[15] K. Hone, and R. Graham. *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