HIV Health Information Access using Spoken Dialogue Systems: Touchtone vs. Speech

Aditi Sharma Grover, Madelaine Plauché, Etienne Barnard, Christiaan Kuun

Abstract— **This paper presents our work in the design of a SDS for the provision of health information to caregivers of HIV positive children. We specifically address the frequently debated question of input modality in speech systems; touchtone versus speech input, in a new context of low literacy users and a health information service. We discuss our experiences and fieldwork which includes needs assessment interviews, focus group sessions, and user studies in Botswana with semi and low-literate users. Our results indicate user preference for touchtone over speech input although both systems were comparable in performance based on objective metrics.**

Index Terms— **Spoken dialogue systems, DTMF, Touchtone, Speech interfaces, Health information, HIV, Illiterate users, Semi-literate users, Low literate users, ICT, Developing regions, Information access, Africa.**

I. Introduction

There is a widespread belief that spoken dialogue systems (SDSs) will have a significant impact in the developing world [1]. This belief is based on a number of factors. Firstly, illiteracy is predominantly a problem of the developing world (according to a recent estimate, about 98% of the illiterate people on earth live in the developing world [2]), and speech-based access to information may enable illiterate or semi-literate people to participate in the information age. Also, the availability of traditional computer infrastructure is low in the developing world, but telephone networks (especially mobile/cellular networks) are spreading rapidly [3]. (For example, a recent community survey in South Africa [4] found that 73% of households owned at least one mobile phone, but only 7% of homes had internet access.) A further factor is the strong oral culture that exists in many traditional societies, which is likely to render such systems more acceptable than text-based or graphical information sources. Finally, the availability of relevant services and alternative information sources is often low in the developing world.

Manuscript received September 22 2008. This research was supported by OSI, OSISA and the NRF under the Key International Research Capacity (KISC) programme, UID no. 63676.
(Corresponding author) A. Sharma, E. Barnard and C. Kuun are with the Human Language Technologies (HLT) Group in the Meraka Institute, CSIR, Pretoria, South Africa. (phone: +27128413028, email: <asharma1, ebarnard, ckuun>@csir.co.za)
M Plauché was a visiting researcher at the Meraka Institute during the project. (email: mad@brainhotel.org)

Based on this perceived value of SDSs in the developing world, a number of exploratory studies have been performed in recent years. Barnard et al. [5] report on preliminary experiments performed to assess the usability of a telephone-based information service for access to government information in South Africa. A kiosk-based SDS for agricultural information was developed by Plauche et al. [6], and evaluated with semi-literate users in rural Tamil Nadu, India. Nasfors [7] also developed an agricultural information service, aimed at mobile telephone users and deployed in Kenya. The most sophisticated speech technology in this category was employed in the telephone-based information service for community health workers developed by Sherwani et al. [8]; this was piloted in Sindh, Pakistan. Agarwal et al. [9] have implemented a telephone-based kiosk system, which they call "VoiKiosk"; this is being trialled in rural villages in Andhra Pradesh, India.

Each of these pioneering studies was primarily aimed at assessing the feasibility of using speech technology in various settings in the developing world. However, in the process of determining feasibility, a number of practical lessons were also learnt. For example, it was found that user acceptance of such systems is proportional to the difficulty that users would have to access the same information through other mechanisms [6] (thereby confirming the concept of the "motivated user"), and two studies [5, 8] found that it may be preferable to use more verbose, less efficient user interfaces to guide inexperienced users for whom time pressure is not a primary concern.

The current contribution similarly has a twofold aim: both to explore the use of an SDS in a new environment (namely, by caregivers of HIV positive children in Botswana, Southern Africa), and to contrast different input modalities used in such a system. In particular, we compare systems using key-presses ("DTMF or Touchtone") with those that use automatic speech recognition (ASR) for user input in this application. To this end, we start with the most basic variant of speech systems (key-press replacement), and compare such systems with DTMF input. The motivation for this choice is twofold: on the one hand, key-press replacement is likely to be more acceptable in the developing world, where general numeracy is less common; on the other, such systems are much easier to develop than natural-language systems, and are therefore more attainable in the resource-constrained environments that typically characterize the developing world.

To our current knowledge DTMF and ASR input modalities have not been compared systematically in the developing world. However, in a review of developed-world applications

comparing DTMF to ASR input by Lee and Lai [10], both user preference and performance are found to depend on the nature of the task, the personality of the user, and the capabilities of the speech-recognition system. With the exception of only one call-routing system [11], all studies found that simply replacing key-presses with speech does not improve user performance or perception. Conversely, well-engineered speech recognition systems with natural-language input are often preferred to DTMF for tasks that are not easily accomplished with DTMF. In terms of user preference, studies in laboratory settings [10], [12] report that users find speech input more interesting and enjoyable to use, but contrastingly in a recent informal poll of over a thousand users of real-world information-access systems [13], almost half of the users responded that they would prefer to use a speech input modality "as little as possible" and only 8% would do so "most of the time". Despite these reservations, numerous applications that are completely reliant on speech input are currently in use in the developed world - examples are the health-management systems described by Migneault et al. [14] and commercial voice portal systems, such as the Tellme portal (which serves 40 million phone calls per month, according to its providers [15]).

Below, we first provide background on the health-care application selected for our study and describe the system that was developed as well as the experimental protocol employed (Section II). Section III contains our experimental results, including user profiles, usability measurements and task completion rates. Finally, we discuss the scope and generalizability of our results, and conclude with thoughts on next steps to be taken along this research trajectory.

II. Open Phone: HIV/Aids Health Information Line in Botswana

A. Background

HIV/Aids is perhaps the gravest health pandemic to face the world, and Southern Africa has been the worst hit region. Of the 33 million people infected with HIV worldwide, approximately two-thirds are inhabitants of sub-Saharan Africa [16]. Within sub-Saharan Africa, Botswana has one of highest HIV prevalence rates, with 1 in every 4 adults being HIV positive [16]. The hardest hit in Botswana are women; nearly 40% of pregnant women (ages 25-39) are living with HIV and infection levels are increasing amongst pregnant women aged 30-34 years, with nearly one in two living with HIV [17]. Aids deaths have orphaned approximately 120 000 children (ages 0-17) and another 14 000 children (aged 0-14) are living with HIV in Botswana.

During our study a partnership was established with The Botswana-Baylor Children's Clinical Centre of Excellence (hence forth referred to as Baylor). Baylor is a specialised paediatric institute serving the Botswana capital, Gaborone and its neighbouring areas since June 2003. The centre is staffed collaboratively by U.S. and Botswana health professionals. The services provided by Baylor range from primary and specialty medical care for HIV/Aids, to catering for the psychosocial needs of HIV patients and their families [18]. Baylor provides treatment to over 2100 children infected with HIV and 260 families across Botswana. The centre is involved in a number of support activities such as community outreach programmes for patients in rural areas, servicing 20 communities outside Gaborone and a "Teen Club" to provide moral support and counselling to teenagers living with HIV [18].

A child is typically brought to Baylor to be tested for HIV/Aids by a caregiver. A caregiver is any individual who takes care of an HIV positive child; it may be the child's parents (who themselves might be HIV positive), other family members or an unrelated community member. Children who test as positive receive free treatment from Baylor for the remainder of their infancy and adolescence. Caregivers of such children are also counselled and trained. Baylor provides free lectures for caregivers three times a week, where many aspects of HIV/Aids, antiretroviral (ARV) medication are explained and advice is given on how to live with the condition. Each caregiver on average attends two lecture sessions. The primary focus of the lecture sessions in Baylor is on adherence to ARV medication, with topics such as the principles of HIV, universal precautions, basics of ARV therapy, medication dosage, side effects and storage, and importance and strategies for adherence, being covered.

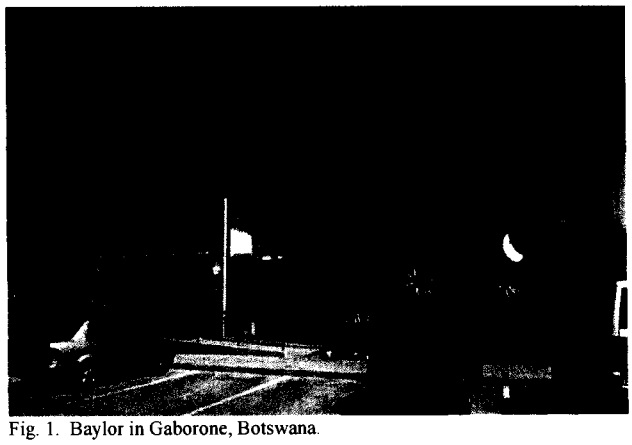

Fig. 1. Baylor in Gaborone, Botswana.

B. Open Phone Development

Open Phone is a pilot HIV/Aids community-oriented SDS service that makes use of language technologies to address acute informational needs of caregivers of children with HIV.

1) Preliminary Investigations

Our initial investigations started in April 2007, where we conducted interviews and discussions with 2 doctors, 4 nurses, 9 caregivers, a social worker and the technical manager. We also accompanied 3 community outreach workers on visits to 2 caregivers' homes. The intention of this investigation was to identify specific needs of the various user groups and their day-to-day tasks.

This trip also served to acquaint us with HIV domain-specific terminology used by the interviewees to describe their situations, tasks and other notable work processes. These 'hidden' pieces of information allowed us to build a profile of the individuals interviewed and make informed choices later in the process of application design.

This initial field trip highlighted a number of challenges and issues for consideration in providing health information to caregivers:

- Caregivers often struggle to recall material covered in the lecture sessions.
- Caregivers often have questions regarding general health information, for example, how to deal with infections, nutritional needs, hygiene requirements and commonly held misconceptions about HIV.
- Travelling to Baylor to address health information queries is not always possible for caregivers due to family/work responsibilities, transportation costs and time constraints.
- The majority of caregivers are semi and low literate populations. No written material is used and thus there is a lack of reinforcement and support for remembering material learnt in the Baylor lectures.
- Most caregivers are uncomfortable with English, thus Baylor lectures and all interactions with caregivers are in Setswana. Baylor staff explains complex health information in accessible terms in local language.
- Although caregivers are encouraged to call Baylor with any questions they may have, most are reluctant (and unable) due to the high costs of mobile phone calls.

These challenges and issues formed the basis of our motivation for the design of a health information SDS that provided not only adherence education from Baylor lectures but also general health information tailored to the needs of caring for HIV positive children. An SDS in the local language Setswana, toll-free and accessible at any time through a simple telephone call, could greatly support Baylor's services to caregivers of children with HIV.

2) Content Development

Our design process started with the identification of relevant content and development of a framework (Fig.2) which detailed the broad health topics that needed to be covered by the SDS. Interviews with Baylor staff (nurses, doctors, nutritionists) and the following printed sources were used to create locally relevant accessible HIV health related content:

- Baylor Adherence lecture materials
- HIV Aids Care & Counselling, A Multidisciplinary Approach [19]
- Where there is no Doctor: A Village Care Handbook [20]
- HIV, Health & your Community, A Guide for Action [21]

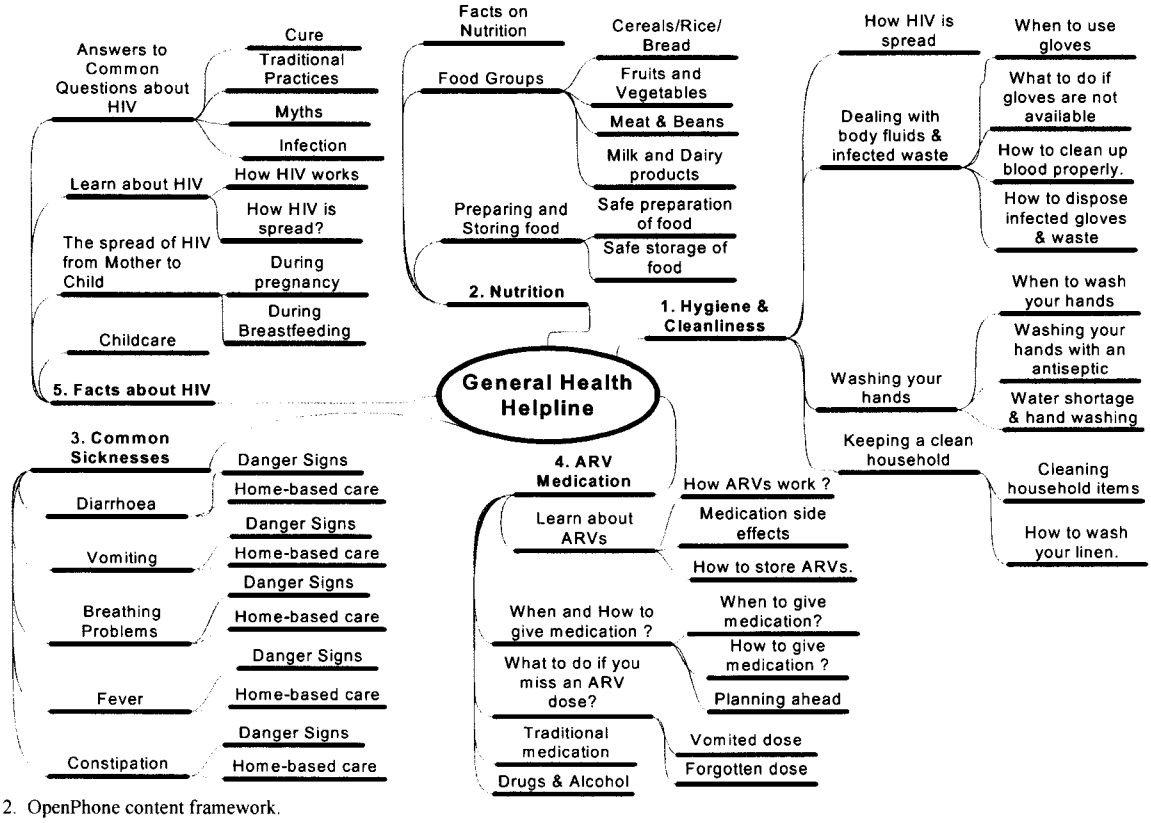

Fig. 2. OpenPhone content framework.

We also held focus group sessions with 27 caregivers at Baylor on a second visit. The discussions were led by a Setswana facilitator and were aimed at soliciting input from caregivers on what kind of health information they typically require under the broad topics of the framework. The caregivers were first given a demonstration of a health SDS prototype and then asked what kind of information they would like the system to provide to them. For instance, the facilitator would introduce the topic of "Nutrition" and then ask e.g. "what kind of food does a child with HIV require? Caregivers' responses were used to fuel discussion and direct further questions on sub-topics of the framework. Thereafter caregivers were asked to prioritize (rank) topics they had listed in order of need in a group consensus setting. In many cases caregivers debated and gave reasons for their choices.

We found that the topics in our content were mostly inline with what caregivers proposed in the focus groups. However, the topics' ranking done by the caregivers varied from ours, for example, they ranked Hygiene & Cleanliness as most important which we had thought was secondary to ARV medication. The rankings allowed us to determine the topics the SDS should focus on providing first and foremost and were later also used to determine to order of main menu options in the SDS. The focus group sessions allowed us to revise the content and framework to build a SDS that caters for the health information needs expressed by the caregivers themselves.

3) System Design

Once the initial content had been created and approved, the dialogue flow and audio output of the SDS had to be designed in a user-friendly manner. We carefully crafted the primarily text-based content for spoken speech and simple language. For instance, words like "lower cranium" will have little meaning for semi & low literate users; such references were translated into more accessible language, using terminology and metaphors used by Baylor staff in their lecture sessions. For example, the white blood cells are the 'soldiers' of the body, and ARV medication is the 'ammunition' for these soldiers.

Since the SDS was aimed at low literacy users and users who may never have used a SDS before, the design goal was to create an easy-to-use interface that placed a low cognitive load on the user. We grouped content under logical menus and sub-menus, so that the amount of audio output under each node was balanced with the number of menu options at any given point in time.

The design team set the following constraints based on guidelines for speech user interface design [22, 23]:

- The maximum depth of menus in the SDS should not be more than 3 levels.
- The maximum breath or number of options at any menu should be no more than 5.

A sample system-user interaction follows.

User: [Dials number...]
System (Introduction): *Hello and Welcome to the Health Helpline, I am Nurse Lerato and I know you probably have many questions about caring for someone with HIV.*
System (Overview): *I can tell you about Hygiene & Cleanliness, Nutrition, Common Sicknesses, ARV Medication, and Facts about HIV. If at any time during your call you want to start our conversation again, you can press 0.*
System (Main Menu): *For Hygiene & Cleanliness, please press 1, for Nutrition, press 2, for Common Sicknesses, press 3, for ARV medication, press 4 or for Facts about HIV, please press 5.*
User: [Presses 2.]
System: *Eating a balanced diet of different foods helps people with HIV stay healthy and strong. A healthy diet does not have to be costly and contains food from all the different food groups. Healthy food is always prepared and stored in a clean environment...*

As a final step, the design team decided on the specific wording of all audio prompts in Setswana. This required careful consideration for the dialect, the register (informal or formal), the cognitive load (short audio prompts, but long enough to provide contextual "anchoring"), and use of appropriate user interface metaphors. For instance, a system that suggests to the user "say Main Menu" is unlikely to make sense to users who have never interacted with visual or audio interfaces before. Instead, we chose the metaphor of a "conversation" and asked the user to "Start our conversation again". The Setswana-speaking linguist on the design team played an essential role in the selection of ASR keywords that were locally relevant and logical to users, yet acoustically dissimilar.

4) Final SDS Development

All the prompts and the health content were translated using a registered translation service into the dialect of Setswana spoken in Botswana. We specified that the content would be spoken aloud and the intended audience would likely be low-literate. Since user interaction with an SDS is based on audio modality, the voice of the system plays a crucial role. With our target audience we felt that this was an essential element in creating a persona that would not only make users comfortable in their interactions with the system but also make the user experience enjoyable. Thus, our ideal voice talent would:

- Sound like a caring nurse willing to answer questions.
- Be a mother-tongue Setswana speaker.
- Have a full, mature female, well-articulated voice.
- Instil a sense of confidence and trust.

Our recruited voice talent was a well-regarded local soap celebrity, which meant her voice would be familiar to many of the target users. All the system prompts and content recordings were done in a professional recording studio.

The SDS was built using the Asterisk telephony platform, the current set up in Botswana is a free standing PC, connected through an Asterisk card to an ISDN line. The ISDN line allows up to two calls simultaneously and users dial a local telephone number to access the system. While the system is operational, all aspects of the calls are logged. From start to finish all key-presses are monitored and all audio is recorded.

C. User Study

The goal of the study was to compare the most basic variant of speech systems (key-press replacement) with DTMF input. Thus, we built two identical systems that differed only at the menu prompts in choice of input modality i.e. in one system the user would press a key to chose a menu option and in the other they would say a keyword or phrase. For example, a DTMF menu option would be; "to hear about Nutrition, press 1," whereas the ASR menu option would say, "to hear about Nutrition, say Nutrition."

The ASR was simulated using the Wizard-of-Oz (WOZ) methodology [24] where a researcher played the role of the speech recogniser. The 'wizard' listened to the speech input of the user and chose the next state of the system on this basis. The 'wizard' only accepted the exact keyword or phrase that the user was allowed to say at a particular menu option, any other input was directed to a No-Match state. In the case of the DTMF system, key presses were handled by the back-end telephony platform. Both systems ran from a PC laptop which was connected to standard telephone through a voice over internet protocol (VOIP) gateway.

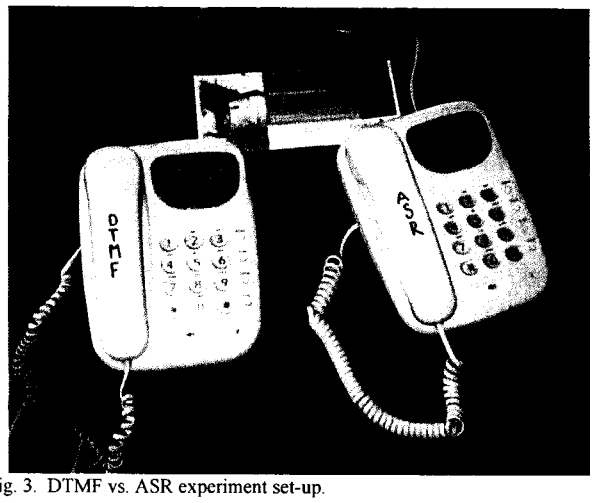

Fig. 3. DTMF vs. ASR experiment set-up.

The user study was held over a period of five days in April 2008 at the Baylor premises. A total of 33 caregivers were part of the study, of which 27 tried both the DTMF and ASR systems. The remaining 6 caregivers did not try both systems due to either user's time constraints or a technology failure experienced in system set-up. The experimental set-up included a facilitator, an observer and a WOZ operator. The facilitators were local graduate students who were trained by the authors to facilitate in the local language, Setswana. The observers took notes on user behaviour, number of verbal prompts needed by a user, any user comments, and general body language of the user.

Each user was introduced to the system and asked to sign a consent form. Emphasis was placed on communicating that users were not being tested but rather the system for purposes of improvement. Thereafter, each user watched a five minute video showing how a caregiver could use the system to find typical health information that they might need (Fig. 4).

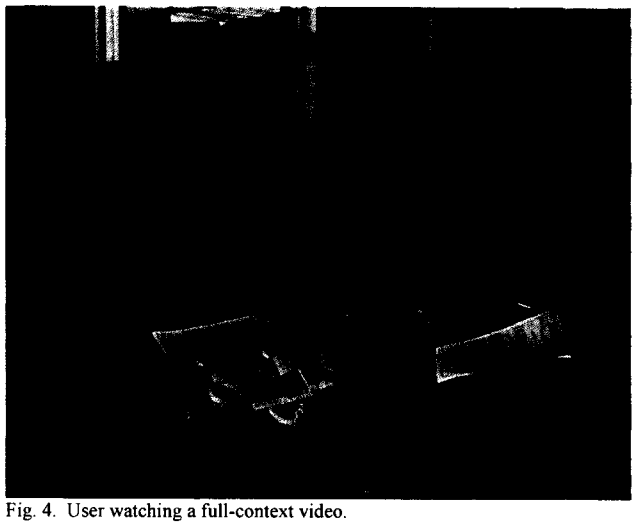

Fig. 4. User watching a full-context video.

The context of the video was carefully matched to depict typical scenarios where a caregiver could use the SDS to obtain health information. For example, in the first part of the video, a caregiver with a sick child who has just thrown up his ARV medication is unsure of whether to give the medication again. A friend of the caregiver arrives and tells her about a telephone information system that she can call and learn when and how the next dose of ARV medication can be given if a child throws up his medication. Research [25, 8] on designing interfaces for low-literacy users has shown that such a full-context video greatly improves user studies by offering not only a demonstration of how to use the interface, but also the source (and therefore trustworthiness) of the content, the context in which you might use such a service, and the potential impact in your day-to-day life.

After this video, the facilitator gave a short demonstration to the caregiver on how to retrieve "Nutrition" information as shown in the second part of the video. For each task, the caregiver was asked to dial a number themselves to access the SDS. This simple action served as a quick check to verify the caregiver's ability to recognise numbers and use a phone. In order to show the difference between the DTMF and ASR systems we used separate telephone sets for each system (Fig.

3). After the caregiver finished the tasks using the DTMF system, for example, the facilitator explained that they would now try a similar system (on the other handset) but one where they would now say a keyword to obtain information. An example of a task explanation is shown on the following page.

Facilitator: *What's the name of any good friend in your neighbourhood?*
Caregiver: *Kabelo*
Facilitator: *Your friend Kabelo says she must wash her hands frequently to keep her family safe from disease but she has very little water at home so she doesn't know what to do. You've heard that Baylor has a phone number (help line) that can answer many health questions. So you decide to call the phone number, can you help your friend?*

Since the study was a within-subject comparison, we refrained from using the same tasks in both DTMF and ASR to prevent bias of previous knowledge. Thus, we created two sets of tasks; Set A and Set B with two tasks each (an easy and a difficult task; Task 1 and 2). The tasks were designed to be similar across Sets and to require a user to make three correct menu selections (Menu levels 1, 2, 3) to reach the specified information. Recall that our HIV health helpline is only three menu levels deep. The correct paths for each task are shown in Table I.

TABLE I
TASK DESCRIPTIONS WITH CORRECT SDS MENU OPTIONS.

Task Description	Menu Level 1	Menu Level 2	Menu Level 3
Task 1, Set A *Find out how to wash hands during water shortage.*	Hygiene and cleanliness	Washing your hands	Water shortage and hand washing
Task 1, Set B *Find out how to protect hands when gloves are not available.*	Hygiene and cleanliness	Body fluids & infected waste	What to do if gloves are not available
Task 2, Set A *Find out how to care for a child with fever.*	Common Sicknesses	Fever	Home Care *
Task 2, Set B *Find out how to care for a child with diarrhoea.*	Common Sicknesses	Diarrhoea	Home Care *

* Home Care can only be reached after the user has heard the Danger Signs of the selected illness. For this reason, Task 2 is slightly more difficult than Task 1.

Additionally, to minimize any possible bias due to the order of trial of the DTMF and ASR systems or the use of a Task Set (A or B) with a particular modality, we systematically ensured that our data covered all possible combinations of Order of Modality and Task Set (illustrated as Quadrants 1-4 in Fig. 5). For example, if a caregiver started with DTMF using Task Set A he/she would then proceed to do ASR with Task Set B (Quadrant 1). Each caregiver thus did a total of 4 tasks and users were approximately evenly assigned between Set A (15 caregivers) or Set B (18 caregivers). The permutations in the order of trials also help to counter the impact of subject fatigue and learning effects within the small sample size.

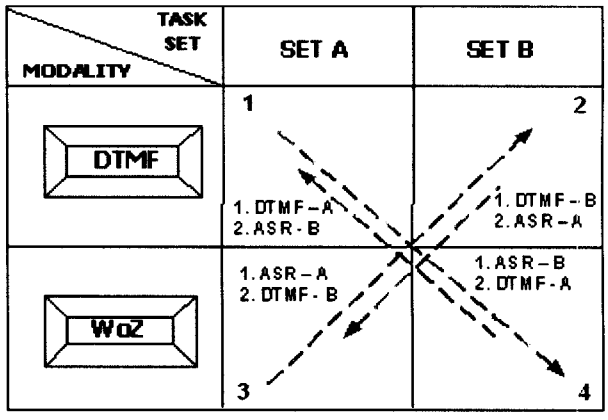

Fig. 5. Order of Trials and Task Sets in experiment.

After completion of each modality's trial, a post questionnaire was administered verbally to the caregiver. It consisted of ten questions adapted from the PARADISE evaluation framework [26]. The facilitator recorded the response in a 5-point Likert scale format based on the strength of the user response.

We then verbally interviewed the caregivers to gather demographic data on education levels, language, and occupation, telephone usage (mobile and landline). Caregivers were also interviewed on their familiarity with technology (computer, mobile phone, TV, radio, video/DVD machine) based on factors such as use, ownership, frequency of use, place of use, and reason for use. Data was also gathered on the number of children they take care of, how often they visit Baylor and how they usually resolve their information queries. At the end of each session caregivers were provided with small non-monetary incentives (juice, potato chips and fruit for the child and gloves & household disinfectant for the caregiver) to thank them for their participation.

Fig. 6. A user taking part in the study.

III. Results

In this section we present results from our user study, including a description of our users (Section A), their task completion scores, other usability metrics from the study and a comparison of their performances on the DTMF and ASR parallel systems (Section B).

A. User Profile

During our user study, we had 33 caregivers who participated, of which 27 tried both DTMF and ASR systems. The caregivers were all female with the exception of one male caregiver. The age of our users ranged between 22 and 61 years old with the average age of 34 years. The average number of years of schooling amongst our users was 9 years but 2 users had 0 years of schooling. All of the users could read and write the local language Setswana and approximately 79% of them knew some English. In terms of occupation, 47% were unemployed and of the 53% who were employed, the majority were in low-income occupations such as cook, cleaner, house maid, hair dresser, or security guard. Caregivers reported that they visit Baylor between 1-3 times a month, with average travelling distance and time at 28 km and 1 hour respectively, some travelling from as far as 130 km, with average cost of travel at 18 Pulas (approx. $3 USD). The average waiting time at Baylor reported by the caregivers was 2.5 hrs, also travel time to Baylor ranges from 30 minutes to 3 hours; together, these represent a significant portion of a working day and are a substantial burden to the caregivers.

Caregivers were also asked what they usually do if they have questions regarding the child's health and when the last time was that they had such a question. One third of the caregivers usually go the local clinic to resolve their health queries and another quarter go to Baylor for this purpose (Fig. 7). Forty percent of the caregivers had a query regarding the child's health within the last 6 months, and another 21% had more recent queries; in contrast, only 12 % could not remember specific queries (Fig. 8).

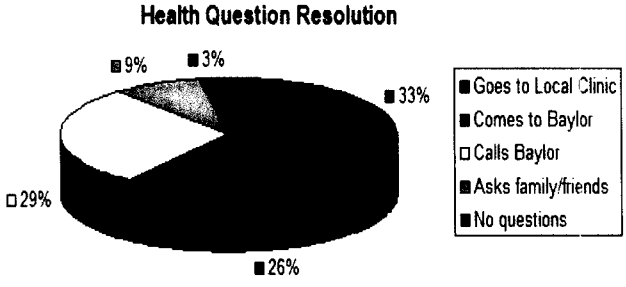

Fig. 7. Methods of Health Resolution by caregivers.

In terms of technology familiarity, 30 out of 33 (91%) caregivers owned mobile phones and 85% of these knew how to load their mobile phones with 'airtime' (pre-paid phones which require users to load money by calling the network provider's service number and entering a sequence of digits from the pre-paid calling card). Average mobile phone costs per month were 68 Pulas ($10.5 USD) with an average cost per call being reported as 4.5 Pulas ($0.75 USD). Only 30% of caregivers reported having access to a landline telephone and of these only 9% had the landline at home (Table II).

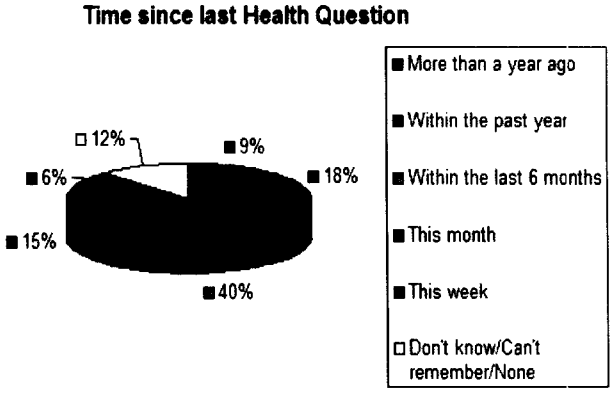

Fig. 8. Time since last Health Question by caregivers.

TABLE II
SUMMARY OF TECHNOLOGY OWNERSHIP AND USAGE BY CAREGIVERS.

Technology	Use	Ownership
Mobile Phone	91%	91%
Landline	30%	9%
Computer	15%	3%
TV	76%	71%
Radio	91%	91%
DVD/Video Machine	41%	38%
ATM	35%	N/A

B. Broad Usability Metrics and Observations

System usability was determined using objective (Task completion rate, response time, routing time) and subjective (user preference, Likert ratings) metrics [27].

1) Response Time

Response Time measures the time it takes someone to respond to the system for the first time. It is usually measured not from the system start, but from the end of the Main Menu prompt, which in our case was 29 seconds long. This results in three basic categories of responses: (1) people who barge-in will have a negative response time, (2) people who respond in the 4 seconds of silence after the end of the Main Menu prompt will have a response time between 0 and 4 seconds, and (3) people who listen to one or several timeouts before responding will have a response time greater than 4 seconds. The Mean Response Times of our users (Fig. 9) indicates many barge-ins, roughly corresponding to the time the correct option at the first menu level was played: *Hygiene and Cleanliness* for Task 1, *Common Sicknesses* for Task 2 (Table I). Although we expected that our users would exhibit some short term learning effects, we found that response time did not decrease from the first call to subsequent calls.

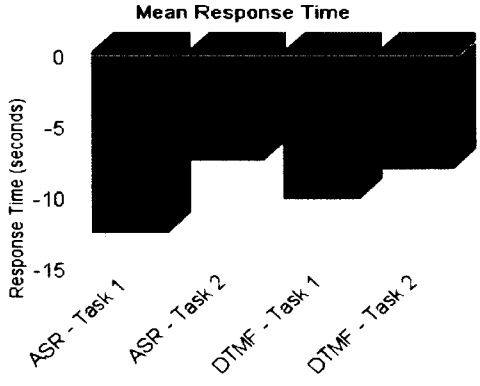

Fig. 9. Average Response Time (time from end of the Main Menu prompt until the user's first input).

2) Task Completion Rate

Task completion rate measures how frequently users were able to reach the node that provided the correct information for their assigned task.

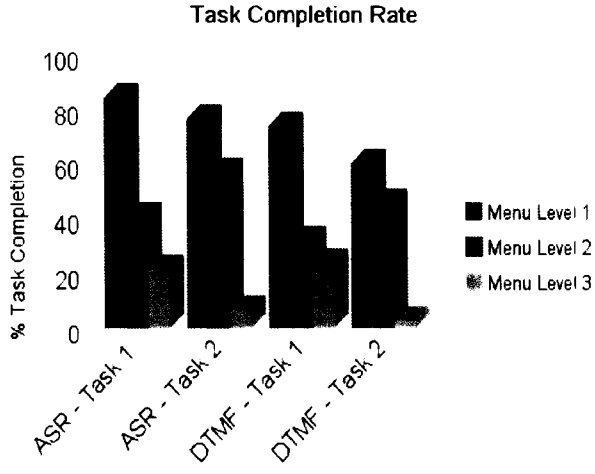

Fig. 10. Task Completion Rates for DTMF and ASR.

For this study, we provide the task completion rate at each menu level of the SDS (Fig. 10). Across both tasks and both input systems, 60% to 84.38% of caregivers selected the correct first menu option (Menu Level 1). Approximately half of those people were able to select the appropriate sub-menu option (Menu level 2) when the task was to find out about "Common Illnesses" (Task 2), with the ASR system yielding the highest task completion rate for this menu level (56.67%). Caregivers seemed to have had more trouble correctly selecting Menu level 2 options for Task 1, with ASR again yielding the highest task completion rate of the two systems (40.63%). Task completion rate for level 3 ranged from 2.85 % to 23.68%, with Task 2 causing the most difficulty. Rates did not vary significantly by input mode (DTMF vs. ASR) or by task set (A vs. B).

3) Routing Time

Routing Time measures the time it takes a user to reach the beginning of the node which contains the correct information for the assigned task. Fig. 11 shows the mean routing time from our user study. Users' routing time was similar across both tasks and both systems for both Level 1 and Level 2. As expected, due to the forced loop through *Danger Signs* for *Common Sickness*, caregivers took longer at Level 3 for Task 2.

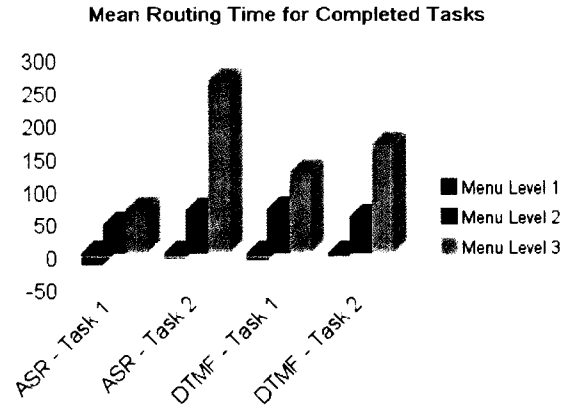

Fig. 11. Routing Time for completed tasks. Recall from Task Completion Rates (above) that the Level 1 mean response is averaged over approximately 75% of our users, whereas Level 2 and Level 3 times are averaged over fewer users, in some cases only one or two.

Only two users correctly completed Task 2 with the ASR system; both users took over 4 minutes to do so. With the DTMF system only one person completed Task 2, doing so in less than 3 minutes. All users that correctly completed Task 1 using the ASR system did so in about 60 seconds. For Task 1 on the DTMF system, most users completed the task in just over 60 seconds, but one user took over 6 minutes.

4) *Other Usability Metrics*

Results from our user study show that user interaction with the ASR and DTMF system are very similar across a wide range of additional usability metrics, illustrated in Fig. 12. Use of the Main Menu global (press '0' or say 'Simolola') and the Exit global (press '9' or say "Fetsa"), were similar for both systems. There were almost the same average number of timeouts (when 4 seconds elapsed with no user response) and repeats (when user chooses to repeat an information node) for the ASR and the DTMF system. On average, caregivers used the barge-in function one more time when using the ASR system but the total number of turns taken by users was similar for both systems. Note, all the usability metrics are means for each call, whether the task was completed or not.

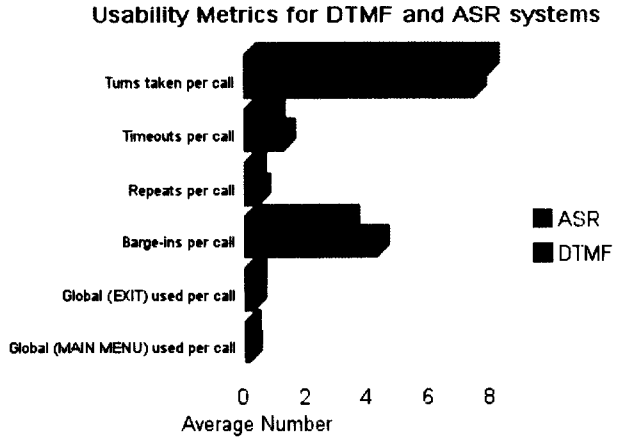

Fig. 12. Usability Metrics for DTMF and ASR systems.

5) *User Preference*

Systems were rated separately using the PARADISE [26] framework after use. The Likert scores were found to be unreliable, however. Despite efforts to elicit honest, critical feedback to the system (for example, we had a different person in a separate room conduct the post-study evaluations), all caregivers gave the system the highest marks possible across all categories and were hesitant to provide any criticism. For those 27 (out of 33 total) caregivers who tried both the DTMF and the ASR system, we were able to elicit feelings of preference for one system over the other (Table III). Most caregivers (59%) preferred the DTMF system over the ASR one (19%) and 22% indicated no preference. Both the DTMF and the ASR systems were judged to be the faster system by those who preferred it. Our measurements of routing time and task completion, however, show the systems are comparable.

TABLE III
USER PREFERENCE FOR DTMF OR SPEECH INPUT

Preferred System	**Number**	**Reasons Given**	**Example Remark**
DTMF	16	Clearer instructions (7) Faster to use (4) More private (2)	*"Its quick and the doctor gives you the instructions, you just have to follow them."*
ASR	5	More accessible (2) Faster (1) Clearer instructions (1) Hands free (1)	*"Its faster and old people can also use it. The button system takes too long. "*
Both or None	6	Similar (2)	*"They are similar. With one you press buttons, with the other, you say*

			things."

6) Social Factor Correlations

We examined several social factors based on our users' responses to questionnaires to see if they correlated with either their interaction with the system, based on objective metrics, or their reported system preference (DTMF vs ASR).

Employment and experience loading airtime on a mobile phone were significant factors in the overall task completion rate (p=0.09 and 0.02, respectively). Those users who were employed or who had experience loading airtime completed more tasks during the study. Age, education, use of landlines, mobile phones, or ATM machines were not significant factors. Only previous use of a landline was a significant factor in use of system globals (p=0.01). Use of landlines was also a factor, along with loading airtime, in whether a user barged-in during system use (p=0.1 and 0.1 respectively). Caregivers who use a landline and load airtime were more likely to barge-in during system use. No questionnaire responses were found to be significant factors for system repeats or overall response time.

Employment, previous use of an ATM, and experience loading airtime on a mobile phone were significant factors in the overall correct response (p= 0.07, 0.09, and 0.03)[1]. Caregivers who were employed, used ATM machines and loaded airtime had more correct responses to tasks during the user study. Age, education, and amount of mobile phone use were not factors in correct response.

Loading airtime was the sole significant factor in user preference of DTMF over the ASR system (p=0.1). Those people who load airtime regularly preferred DTMF over ASR. Employment significantly correlated (p<0.1) with Overall Task Completion and Overall Correct Response. Age and monthly mobile phone costs were not found to correlate significantly with user performance or user preference for either the DTMF or ASR system.

7) Other Observations

We observed based on body language and explicit remarks that several caregivers were nervous at first, and then became more relaxed during the first few minutes of the study. Many caregivers (and their children) showed signs of fatigue while trying the second system (sometimes an hour later). Caregivers sometimes had trouble understanding the task. They would often try to find information about 'Nutrition', for example, which was the topic of the demonstration, instead of searching for information on their assigned task. Caregivers also often clarified with the interpreter what the task was, what the keywords were, and what they should press or say during their trial. The interpreter would nod or say 'yes' but would not help the caregiver any further. In some cases, they would ask to try again, which we allowed if time permitted. Most were very interested in the content and many referred to the voice they heard as 'the doctor'. Only one caregiver recognized the voice of the SDS as the celebrity soap star although most commented that "the 'doctor' explained very nicely".

During the interviews, all of our users enthusiastically indicated that they would like to use the service again; many said that it would be very valuable for educating themselves and their family/friends on caregiving aspects for children with HIV. A SDS such as OpenPhone could also serve as persuasion tool for caregivers trying to educate others, as explicitly reaffirmed by a caregiver "*now I can tell them at home that the doctor* (SDS voice) *says the same thing* (referring to a HIV related topic) *that I'm telling them*".

IV. DISCUSSION

From our pilot study, we found that there were no significant differences between task completion rates (ASR only performed slightly better) or other usability metrics for both systems. This agrees with a number of previous studies in the developed world [12, 28, 29] where no major differences were found in terms of performance. However, subjectively the majority of our users preferred DTMF (59%) over ASR (19%), which is in contrast to formal studies in the developed world [9]-[11] (where user preferences largely favour speech), but correlates with the observation that simple key-press replacement with keywords is generally not viewed favourably. The users who did prefer ASR did not as in developed world studies comment on the aesthetics of speech input, that "speech is more entertaining or enjoyable" but rather on the utility of speech "more accessible for older people or faster".

Our finding that users' employment and experience loading airtime correlated with higher task completion but that education level does not, indicates that technological literacy is a more important factor in adopting new technology than literacy itself. This may also contribute to the finding that 'loading of airtime' was the sole significant factor found in DTMF preference over ASR.

It is also interesting to note that our users who had minimal exposure to SDSs (except loading airtime), were relatively comfortable using our system for the first time, as indicated by their frequent, timely barge-ins. Also, our users noticed the value of speech (allowing hands-free operation, innumerate/older people being able to use it) and DTMF as well (provides privacy). This highlights that even if users may be technically inexperienced and unfamiliar with an ICT application, they have valuable and sound judgment on the utility of the interface. Also, both DTMF and the ASR systems were judged (subjectively) to be the faster system by those who preferred it, which indicates that time constraints, may also be of importance for low literacy users, in contrast to

[1] The following categorical groups (yes and no) were divided into two groups and the resulting dependent values analysed: Employment, previous use of an ATM, experience loading airtime, use of landlines, use of ATM machines. A p-value of 0.1 and less indicates a significant difference between the two groups, whereas a p-value higher than 0.1 indicate no significant difference.

earlier observations. The DTMF vs. speech input comparison could be improved in terms determining task performance by using a between-subjects experiment design; however this approach would not reveal user preference.

Our preliminary investigations indicated that a good fraction of caregivers were in the low and non–literate range. In our user study though, we encountered that many of our recruits were semi- to low literate. Also, whilst our user numbers represent a significant sample size for a developing world user study, it may in comparison to studies in the developed world be on the 'small' end of the spectrum. The above-mentioned issues emphasize some of the challenges faced in research for the developing world; that the very users for whom an SDS could be most useful may be the hardest for us to reach and also that user recruitment in developing regions can present significant obstacles [30].

Whilst this pilot study illustrated that telephony services could in fact be easily used by semi and low literacy users, and that an SDS in local language can be a powerful health education tool, the decisive factor in widespread uptake is likely to be the cost incurred by the caller for the service. The majority of caregivers said that even though the service would be useful to them they would only be able to make use of it, if the service is toll-free. An average phone call in Botswana of 5-10 minutes to the SDS would cost a mobile phone user $1-2 USD. From our questionnaires, the average cost per month for mobile phone usage was $10.5 USD. Thus, a single phone call to the SDS would consume, 10-20% of a caregiver's monthly mobile phone budget, making the case for a toll-free number all the more imperative.

Notwithstanding our findings that show promise for SDSs for low literacy users, we did encounter challenges in introducing the concept of a SDS to our users, for instance a few users did not fully realise that the system was automated. For example, at the end of one call, a user proceeded to ask the 'nurse' (system persona) a question when prompted by the system to leave a comment (and waited for the answer). Another user repeatedly acknowledged what the 'nurse' was saying by responding with "Yes, yes" or "I agree with you". An obvious solution here might be to use a text-to-speech (TTS) voice for the prompts. However, we run the risk of a mismatch between the target audience and the context (e.g. emotionally sensitive in terms of health care) or culturally-based communication norms of the community. This in turn may affect the willingness of users to interact with the system.

A general design challenge for SDSs is to ensure a minimal cognitive load on the user. This magnifies more so in the case of information access applications where lengthy pieces of information need to be provided (e.g. in a health care context). In our case, all the users indicated in the subjective questionnaire that the length of the content was not too long; however, during the experiment some users did mention the need to concentrate in order "not to miss all the things being said by the 'nurse' ". There is a need to address the above-mentioned issues by developing the dialog to be more interactive (perhaps with audio cues, getting intermittent user feedback) and conversational.

In addition, the nature of educational information services tends to be rather exploratory; where a user may peruse various topics related to his/her general query (e.g. a user may want to, in general know about "Dealing with body fluids and infected waste" which has 4 topics all related to that option). This in our case translated to some users struggling to find the exact menu option related to the task and exploring related sub-topics. One user even singled out that she would like to know the mapping of the menu options beforehand to help her locate the information she is looking for. This experience highlighted challenges of using hierarchical menus [8] and the importance of paying attention to the taxonomy and vocabulary of the system to enable easy navigation for the user.

SDS design for smaller languages also introduces challenges on other dimensions including prompting and persona. The prompts and content of a SDS application will typically be translated from a language such as English to the local language. Thus, great care has to taken in the prompt writing phase to ensure that intended meaning of the original prompt (English) is still preserved in the translated prompt (local language) and conveyed in the simplest and shortest way possible. Often, a concept described by a single word in English has no direct translation in another language. For instance, whereas a keyword in the English version of our application was "Safe food", it became the phrase, "Dijo tse di siameng" after translation, in order to adequately describe the concept.

Moreover, not only should the translated SDS prompts convey the intended meaning but the designer should ensure that the persona of the local language system is in line with cultural and contextual expectations of the intended audience. For instance we ensured our prompts not only had the right balance of formality and gravity appropriate for the message (HIV info) but were also understandable and conversational.

Our experience in employing multiple data-gathering techniques in the needs investigation phase (interviews and discussions, observations, field visits, and focus groups) better equipped us in trying to comprehend the needs of our users. The interviews and discussions helped us establish a rapport with our users and stakeholders (Baylor staff), whilst also providing the flexibility of follow-up questioning and recalibration of interviewer terminology when needed. Observations on the other hand helped to further reveal the issues that users are unable or unwilling to articulate or express. Field work allowed us to pick up on the cultural nuances and social-economic context of the users and enabled us to gain a deeper understanding of the sensitive environment (HIV/health) we were working in. Finally, through focus groups we were able to observe the interaction amongst caregivers and most importantly it enabled us to obtain specific targeted design information in a group setting and correct our earlier assumptions on the ranking of SDS health topics.

V. CONCLUSION

This paper has addressed the frequently debated question of input modalities of touchtone vs. speech in a completely new context; low literacy users and a domain (health) different from the usual business centric applications (call routing, voice mail systems or banking). Our pilot study also served to confirm the feasibility of SDS applications for semi and low literate populations.

It is interesting to speculate on the applicability of our findings to other resource-poor countries. Our intuitive judgment is that the only characteristic of our user population that strongly influenced our results was their relative familiarity with mobile telephones; however, the matter requires detailed investigation. In future work we would therefore like to further explore the space of non-literate users and the suitability of SDS applications for them, as well as investigate the interaction of task types (linear vs. non-linear) and application domain – informational (Openphone) vs. transactional (tracking a social services payment) with the input modality. Once the system is free of charge and has been in use for several months, we would like to study whether OpenPhone leads to a change in health habits and improves the ability of a caregiver in providing care to children, like change of hygiene habits, better nutrition, and fewer misconceptions about HIV.

Throughout the design and development process, we experienced that beyond usability and creating simple, accessible user interfaces for low literacy users, factors such as cultural and social context, establishing relationships with stakeholders and user communities, and localizing content, play a vital role in the success of an ICT intervention in the developing world. We intend to draw on these valuable experiences and carry our work forward in serving the information needs of citizens of developing regions with accessible and usable telephony based services.

ACKNOWLEDGEMENTS

The authors wish to thank the staff of Botswana Baylor Children's Clinical Centre of Excellence for the generous manner in which we were received, in particular Dr Paul Mullan, for the amount of help and time given to us. We would also like to thank members of the HLT group at Meraka Institute who provided valuable contributions throughout the project; Victor Zimu, Jama Ndwe, Mpho Kgampe, Louis Joubert, Richard Carlson, Bryan Mcalister, Mark Zsilavecz, Marelie Davel and Alta de Waal. We also greatly appreciate the time given to us by Connie Ferguson, our celebrity voice talent. This research would not have been possible without the joint support of OSI and OSISA. The project was also in part supported by the NRF under the Key International Research Capacity (KISC) programme, UID no. 63676.

REFERENCES

[1] R. Tucker and K. Shalonova. "The Local Language Speech Technology Initiative", *SCALLA Conference*, Nepal, 2004.

[2] SIL International. (2001). Facts about illiteracy. [Online] Available at:: http://www.sil.org/literacy/litfacts.htm (last accessed 15 Sept 2008).

[3] ITU: International Telecommunications Union. (2003). Mobile overtakes fixed: Implications for policy and regulation. [Online]. Available: http://www.itu.int/osg/spu/ni/mobileovertakes/Resources/Mobileovertakes_Paper.pdf, (last accessed 5 September 2008).

[4] Statistics South Africa. (2007). Community Survey 2007. [Online] Available at: http://www.statssa.gov.za/publications/P0301/P03011.pdf (last accessed 15 Sept 2008).

[5] E. Barnard, L. Cloete, and H. Patel. "Language and Technology Literacy Barriers to Accessing Government Services," *Lecture Notes in Computer Science*, vol. 2739, pp. 37-42, 2003.

[6] M. Plauche, U. Nallasamy, J. Pal, C. Wooters, and D. Ramachandran. "Speech Recognition for Illiterate Access to Information and Technology," in *Proc. IEEE International Conference on Information and Communications Technologies and Development '06*, pp. 83-92, May 2006.

[7] P. Nasfors. "Efficient Voice Information Services for Developing Countries", Master Thesis, Department of Information technology, Uppsala University, Sweden, 2007.

[8] J. Sherwani, N Ali, S. Mirza, A. Fatma, Y. Memon, M. Karim, R. TRongia and R. Rosenfeld, "Healthline: Speech-based Access to Health Information by low-literate users", in *Proc. IEEE International Conference on Information and Communications Technologies and Development '07*, Bangalore, India, Dec. 2007.

[9] S. Agarwal, A Kumar, AA Nanavati and N. Rajput, "VoiKiosk: Increasing Reachability of Kiosks in Developing Regions", in *Proc. of the 17th international conference on World Wide Web*, pp. 1123-1124, 2008.

[10] K. M. Lee and J. Lai "Speech Versus Touch: A Comparative Study of the Use of Speech and DTMF Keypad for Navigation" *International Journal of Human–Computer Interaction*, vol. 19, no. 3, pp.343–360, 2005.

[11] B. Suhm, J. Bers, D. McCarthy, B. Freeman, D Getty, K Godfrey and P Peterson, "A comparative study of speech in the call center: natural language call routing vs. touch-tone menus". In: Terveen, Loren (ed.) *Proc. of the ACM CHI 2002 Conference on Human Factors in Computing Systems Conference*, Minneapolis, Minnesota. pp. 283-290, April 2002.

[12] C. Delogu, A. Di Carlo, P Rotundi and D. Sartori, "Usability evaluation of IVR systems with DTMF and ASR", *in Proc. of the 5th International Conference on Spoken Language Processing (ICSLP 98)*, paper 0320, Australia,1998.

[13] T. Pearce and M Bergelson. (2008). Alignment index for speech self-service. Dimension Data Technical Report [online]. Available at http://www.dimensiondata.com/NR/rdonlyres/9191A848-5F35-459F-8239-8D9D2248414E/8791/mainstreamspeechalignmentindexreport2.pdf (Last accessed 15 Sept 2008).

[14] J.P. Migneault, R. Farzanfar, J.A. Wright, and R.H. Friedman. "How to write Health Dialog for a Talking Computer," *Journal of Biomedical Informatics*, vol. 39, no. 5, pp. 468 – 481, Oct. 2006.

[15] Tellme Networks Inc. Who uses TellMe? [Online]. Available at: http://www.tellme.com/about (Last accessed 15 Sept 2008).

[16] UNAIDS (August, 2008). Report on the Global AIDS Epidemic. Joint United Nations Programme on HIV/AIDS. [Online] Available at: http://www.unaids.org/en/KnowledgeCentre/HIVData/GlobalReport/2008/2008_Global_report.asp (last accessed 7 Sep 2008).

[17] K. Seipone, W. Jimbo, F.d.l.H. Gomez, K. Ampomah, J. Othwolo, O. Kaluwa, M. Busisiwe. "Trends in HIV Prevalence Among Pregnant Women in Botswana 2001-2005" 16th *Intl. Conf Aids*. Toronto, Canada, Aug. 2006.

[18] Botswana-Baylor Children's Clinical Centre of Excellence, Annual Report 2006, Gaborone, Botswana, 2006.

[19] A. Van Dyk, *HIV Aids Care & Counselling, a Multidisciplinary Approach*. 2005, Third Edition: Pearson Education, South Africa.

[20] D. Werner, C. Thuman and J. Maxwell. *Where there is no Doctor, a Village Care Handbook.*. 2007, New revised edition: Hesperian, CA, USA.

[21] R. Granich and J. Mermin. *HIV, Health & your Community, A Guide for Action*. 2001, Hesperian, CA, USA.

[22] G.A. Miller. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *The Psychological Review*. vol. 63, no. 2 pp. 81-97, 1956.

[23] Suhm B. "IVR Usability Engineering using Guidelines and Analyses of end-to-end calls" in D. Gardener-Bonneau and H.E. Blanchard (Eds). *Human Factors and Voice Interactive Systems*. 2008, pp. 1-41, Second Edition, Springer Science: NY, USA.

[24] N.M. Fraser and G.N. Gilbert, "Simulating speech systems", *Computer Speech and Language*, vol. **5, no.** 2, pp. 81-99, 1991.

[25] I. Medhi and K. Toyama, "Full-Context Videos for First-Time, Non-Literate PC Users" ", in *Proc. IEEE International Conference on Information and Communications Technologies and Development '07*, Bangalore, India, Dec. 2007.

[26] M.A. Walker, D. Litman, C.A. Kamm and A. Abella. "PARADISE: A general framework for evaluating spoken dialogue agents" *in Proc. of the 35th Annual Meeting of the Association of Computational Linguistics*. ACL/EACL 97, 1997.

[27] Nielsen, J., (1993). Usability Engineering. AP Professional, Boston, MA, USA.

[28] Foster, J. C., McInnes, F. R., Jack, M. A., Love, S., Dutton, R. T., Nairn, I. A., et al. (1998). An experimental evaluation of preference for data entry method in automated telephone services. Behaviour & Information Technology, 17, 82–92.

[29] Goldstein, M., Bretan, I., Sallnas, E.-L.,&Bjork, H. (1999). Navigational abilities in voice-controlled dialogue structures. *Behaviour & Information Technology, 18,* 83–95.

[30] E. Brewer, M. Demmer, M. Ho, R.J. Honicky, J. Pal, M. Plauche, and S. Surana. "The Challenges of Technology Research for Developing Regions," *IEEE Pervasive Computing*, vol. 5, no. 2, pp. 15-23, April-June 2006.