A voice service for user feedback on school meals

Aditi Sharma Grover[1], Karen Calteaux[1], Etienne Barnard[2], Gerhard van Huyssteen[3]
[1]CSIR Meraka Institute, P.O. Box 395, Pretoria, 0001, SOUTH AFRICA
[2]Multilingual Speech Technologies Group, North-West University, Vanderbijlpark, 1900, SOUTH AFRICA
[3]Centre for Text Technology, North-West University, Potchefstroom, 2520, SOUTH AFRICA
Email: asharma1@csir.co.za, kcalteaux@csir.co.za, etienne.barnard@gmail.com, gerhard.vanhuyssteen@nwu.ac.za

ABSTRACT
Research using voice-based services as a technology platform for providing information access and services within developing world regions has shown much promise. The results for design and deployment of such voice-based services have varied depending on the application domain, user community and context. In this paper we describe our work on developing a voice-based service for obtaining feedback from school children, a previously unexplored user community. Through a user study, focus group discussions and observations of learners' interaction with multiple design prototype versions, we investigated several factors around input modality preference, language preference, performance and overall user experience. Whilst no significant differences were observed for performance across the prototypes, there were strong preferences for speech (input modality) and English (language). Focus group discussions revealed rich information on learner's perceptions around trust, confidentiality and general system usage. We highlight several design changes made and provide further recommendations on designing for this user community.

Categories and Subject Descriptors
H.5.2 User Interfaces: Voice I/O User Interfaces; H.5.2 User Interfaces: Evaluation; H.1.2 User/Machine Systems: Human Factors

General Terms
Design, Human Factors, Languages

Keywords
Voice user interface, DTMF, speech interface, children, ICTD, developing world.

1. INTRODUCTION
It is widely acknowledged that mobile voice-based services are effective channels for bridging communication gaps caused by challenges such as literacy, language, distance and infrastructure in developing countries [5, 30, 20], although their uptake is as yet limited in these contexts. While the design and development of voice-based services for developing world contexts is well-established in the literature, an understanding of how such technologies are used by children and the design changes required to adapt such technologies for use by children, is lacking.

Approximately 7 million learners in 20 000 schools in South Africa benefit daily from a school feeding scheme implemented by the country's Department of Basic Education (DBE). School feeding schemes in South Africa started in the early 1940s with the supply of free milk in certain schools. This was later broadened to include the provision of fortified biscuits, nutrient supplements or full meals [38]. The National School Nutrition Programme (NSNP) came into being in 1994 and at present, provides (at least) 1 warm (cooked) meal to learners in impoverished primary and secondary schools for a minimal amount (less than USD 0.50 per learner per meal per day) [34]. This programme is seen as a fundamental component in the drive towards improved education in South Africa, since research has shown that good health and nutrition are prerequisites for effective learning [6, 40]. However, a programme of this magnitude is inevitably vulnerable to abuse, and DBE has identified feedback from the intended beneficiaries as an important route towards improved service delivery in the NSNP.

Our study involved the development and deployment of a voice-based service aimed at obtaining feedback from child users on their experiences of the NSNP. This paper reports on the findings of a user study we conducted to evaluate our preliminary designs prior to deployment in the field, reports on changes we made to the preliminary designs and provides recommendations based on our experiences.

Section 2 summarises related work on designing voice-based services in developing world contexts, for child users in particular. We then describe the systems we developed with reference to their design and implementation. In section 4 we discuss the results of the user study conducted to evaluate our preliminary designs and indicate the changes we made to the design based on said study. This is followed by a discussion of the lessons learnt from our study, which may be useful for others involved in designing voice-based services for child users. We conclude with an indication of future work we plan to conduct in this field.

2. RELATED WORK
Voice-based services in domains such as education, health, agriculture, finance, etc. are discussed in studies by [1, 19, 26, 31, 33, 36, 39]. Voice-based services are often referred to as spoken dialog systems (SDS) or interactive voice response (IVR) (typically touchtone input and speech output) systems in literature. Such systems enable a user to access information or services using the voice channel on their telephones – which in the developing world are most frequently mobile phones. Users

navigate the systems through a set of menus, using their voices or dual tone multi frequency (DTMF) keypresses. Access to the system is through a telephone call to the system's number.

Research on designing voice-based services for developing world contexts, relevant to our study, has focused on topics such as *input modality* (speech vs. DTMF) [19, 27, 31, 36], *dialog design* [18], *mobile user interfaces* (text-based vs. text-free interfaces) [24, 30, 32], and *user-centered design* issues including designing for low literate/oral users [27, 28, 29, 35], designing for users without access to specialized software [27], designing for users without prior IVR experience [30], and designing within socio-cultural contexts unique to the developing world [1] or distant (geographically, culturally, cognitively or operationally – in terms of language, literacy, disability and others) [15]. Due to the differences between developing for literate and low literate users, Sherwani *et al.* [35] suggest testing deployed interfaces in the field instead of conducting user studies.

Results of research on the topic of *input modality* in voice-based services have varied in terms of user performance (task success) and user preference for any particular modality. Sherwani *et al.* [36] and Lerer *et al*. [27], found speech input provided a significantly higher task success rate than DTMF. Conversely Patel *et al.* [31] found that user performance was better with DTMF input, whilst Grover *et al.* [19], report no significant difference in user performance between speech and DTMF input. For user preference, both Grover *et al.* [19] and Patel *et al.* [31] report that users preferred DTMF over speech input, whilst Sherwani *et al.* [35] report no significant difference in user preference, and Lerer *et al.* [27] – although not explicitly reporting user preference – suggest that users did not like the DTMF aspects of the system. It is worth mentioning that while all these studies targeted developing world users, they were conducted in different domains and contexts with different types of users; HIV info (Grover *et al.*), agriculture info (Patel *et al.*), general health info (Sherwani *et al.*) and an audio survey (Lerer *et al.*). Hence, the apparently contradictory results may be indicative of domain and user community differences.

Patnaik *et al.* [32], Medhi *et al.* [28] and Kote *et al.* [24] compare a range of *mobile user interfaces* (UIs) which include voice-based services. In [24], it was found that error rates for data collection through a live operator (voice) were significantly lower as compared to SMS and electronic forms (via mobile phone). Medhi *et al.* [28], compare text-based interfaces such as electronic forms, SMS and USSD with text-free interfaces such as an SDS, graphical UI and a live operator. They report that "textual interfaces were unusable by low literacy users and difficult to use by novice users". In the case of text-free interfaces, the live operator (voice) was found to be the most effective, with varying results for voice and graphical UIs respectively. In particular for voice UIs, Medhi *et al.* [28] suggest that users, who are somewhat familiar with the concept and the general terminology (prompts) of the voice UI, were faster and more independent in their task execution. However, overall, graphical UIs had a higher task completion rate but users took significantly more time to complete the task and required more prompting and encouragement during the study. In a similar vein, Kote *et al.* [24] found that users preferred an IVR over SMS for a service that crowd sources water availability information in India.

User-centered design methodologies have contributed significantly to ensuring that technologies are designed to meet the needs of users in the context of their work or life [15, 35]. Gorman *et al.*'s study of usability testing for oral, rural users indicated that issues such as memory retention, training, testing in groups and localization to suit each unique context are important variables when measuring the success of systems designed for people living in oral cultures. Although not typically low literate in terms of reading and writing, the users in our study generally did not have previous exposure to an IVR, leading us to adopt a user-centered approach to designing our systems.

Research on the *use of technology by children* has mainly led to projects aimed at developing both hardware and software systems to enhance learning. The One Laptop per Child (OLPC) project [http://one.laptop.org/about/mission], a topic of much debate, aims to provide children in developing countries with rugged, low-cost, low-power, connected laptops, software, tools and content designed for collaborative, fun, self-empowered learning. Mobile- and web-based games for enhancing language learning [8, 11, 22, 23] and mathematics education [7], for example, have also received significant attention in the e-learning field. Negative experiences of e-learning and m-learning technologies, which have limited their use and benefits, have led to research aimed at understanding mobile user experiences in m-learning environments [4]. This research reiterated that designers cannot design an enhanced user experience but rather need to design *for* an enhanced user experience through extensive user involvement.

There is limited evidence of *VUI services* being developed *for use by children*. Stritzke *et al.* [36] investigated the use of IVRs as a data collection method for obtaining information on children's daily attitudes toward alcohol and tobacco use. However, they [36] indicate that a 2002 review by Corkrey and Parkinson [10] of such use of Interactive Voice Response (IVRs) systems in 54 studies published between 1989 and 2000 indicated no applications of IVRs involving children. A number of such voice-based services exist, however, to gather and share information relating to children [9, 19, 21]. In most instances, such services are used by adults, such as caregivers of HIV positive children [19]. Likewise, an IVR-based system deployed in India to gather data on the number of children fed on a particular day, made use of outbound calls to teachers [21]. A (mostly) operator-backed IVR service for use by children for a tele-counseling service in Delhi is described in [9].

From a comparison of two developmental applications of speech technologies, Grover *et al.* [16] distill a set of dimensions which we considered when conceptualizing the design, development and implementation of our system. These include the *nature of the user community* (with issues such as literacy, technology experience and openness of the user community being considered); *content source* (where relevant and timely content, local language, sensitivity of the content and trustworthiness was considered); *application complexity* (where we considered the degree of technical difficulty appropriate for the target population); and the *business model and deployment* (where aspects such as cost, stakeholder support, alignment with existing channels and sustainability were considered).

In contrast to previous research on designing, implementing and testing voice user interfaces (VUIs) for adults, this paper makes a unique contribution by focusing on VUI design for children who, in addition, have limited or no prior exposure to VUI services. We do this by describing our research process and discussing the findings with regard to language preference, input modality, socio-cultural issues, persona, prompt design and implementation, to be considered when designing for this target population. To our knowledge, the aspects affecting design decisions for systems aimed at this target population have not been considered before.

3. SCHOOL MEALS LINE

3.1 Background

The NSNP aims to contribute to improving learner capacity; to promote self-supporting food gardens and other production initiatives; and to promote healthy lifestyles among learners [12, 13]. Implementation guidelines for the NSNP include menus [14] designed to ensure that learners receive a balanced meal, which meets at least 30% of their daily nutritional requirements, and recipes [13] which also provide measurements to guide the preparation of the meals. The times at which learners should be fed, and the conditions under which they should be fed, are also prescribed.

Facilities for storing the foodstuff and preparing the meals differ from school to school, but are typically under-resourced. The meals are prepared by volunteer food handlers (mostly parents of the learners) who are paid a monthly stipend for their efforts. The implementation of the feeding scheme is monitored at school level by teachers designated as NSNP School Coordinators. They, in turn, are monitored by officials at the district, provincial and national tiers of government. Monitoring compliance with the prescriptions for implementation of the NSNP is a challenge, as is evaluating the quality and quantity of the meals being served. There is an acute need for evaluation by end users: to obtain feedback and provide more direct access to the NSNP's "clients" – the school learners.

An IVR, henceforth referred to as the School meals line was developed to address some of these challenges. In this application, learners and their caregivers are able to make a free call to give feedback on whether or not a meal has been received; what the meal entailed; whether the quality of the food was satisfactory; and whether the meal was to their liking.

In addition to the School meals line, the research team also developed a School Coordinator Reporting Service (an IVR-based daily report on the number of learners fed and what the meal consisted of) and a Provincial Coordinator Report Reminder Service (a web-based notification system using SMSs to remind officials to submit progress reports). All three applications are linked to a monitoring web interface through which the researchers and the NSNP can monitor the data emanating from their deployment.

3.2 Design & Implementation

The School meals line allows learners and their caregivers to make a free call to provide feedback on the quality of the last meal received at their school; whether they liked the meal, if it was enough, on time, and tasty. The learner simply has to give a "missed call" to the system and the system calls the learner back.

A sample interaction is presented below for one of the prototype versions (English speech) we used our user study.

Sys 0.1: Hola! Thanks for calling the school meals line.

Sys 0.2: Which school do you go to? Say the name after the tone. <Tone>

User: XYZ Secondary School

Sys 0.3: Alright, was the food good.., ok.., or bad?

User: Ok

Sys 0.4: Let's chat about that. Was the food tasty? You can say yes.. or..no.

User: Yes

Sys 0.5: Was the food enough? Say yes.. or.. no.

User: No

Sys 0.6: Did you eat on time? Say yes.. or.. no.

User: No

Sys 0.7: Thanks! If you want to tell us more about the meal, leave your message after the tone and press hash when you're done. Or you can now hang up. <Tone>

User: The food was not well-cooked..

Sys 0.8: Ok. If you want us to call you back, leave your name and number after the tone. Or you can now hang up.<Tone>

User: ABCD, 0123456789

Sys 0.9: Please repeat your number after the tone. <Tone>

User: 0123456789

Sys 0.10: Thanks for calling! Bye!

From the stakeholder's perspective (DBE) the feedback provided by the children through the School meals line is captured on a monitoring web interface which can be viewed by any DBE official with an authorised login. The web interface allows DBE to obtain daily information on the details of how many children received breakfast/lunch on per province basis (figure 1).

Figure 1. School meals line's monitoring web interface

The web interface also allows the DBE to get a breakdown of the learners' responses to the questions regarding meal satisfaction (on time, enough, tasty, well-cooked?) and to listen to any audio messages left by the learners (figure 2).

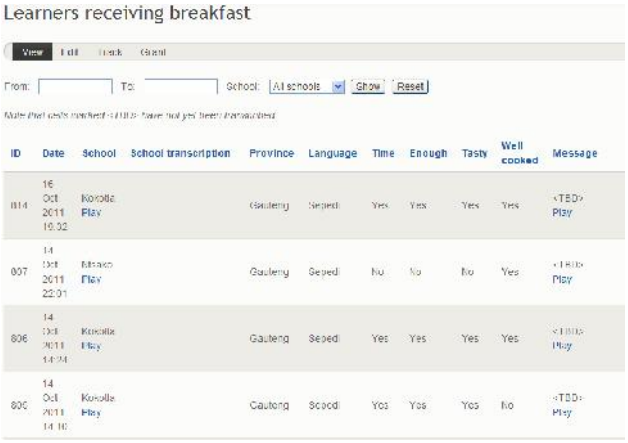

Figure 2. Monitoring web interface view: learners who reported receiving breakfast.

Currently the audio messages are transcribed manually by an in-house system administrator. This provides the DBE with the additional facility to not only listen to messages but also, at a glance, be able to read them.

In designing the questions to ask the children and subsequent system prompts, we worked with a qualified child therapist to ensure that the prompts were appropriate for interaction with learners. The School meals line's system architecture is illustrated in figure 3. The School meals website was built using the Drupal web content management system, which enabled the re-use of existing open-source components. The School meals line's IVR was built using the open-source Lwazi telephony platform (http://sourceforge.net/projects/lwazi), which facilitates speedy development and provision of multilingual IVR applications.

The telephony platform builds upon the well-established Asterisk software private branch exchange (PBX) by providing an IVR application programming interface (API) and runtime engine in the Python programming language, MobilIVR. This enables application developers to easily create telephony-based information services. The IVR is provided over a standard ISDN line, which in turn interfaces with the Asterisk software PBX via an ISDN-SIP gateway with the SIP protocol. All incoming calls are serviced by the Lwazi telephony platform's call-back mechanism, which interfaces directly with Asterisk. The call-back mechanism queues all missed calls and services them sequentially, one at a time. When the service calls the user back, it hands the call over to be handled by the School meals IVR dialog application, which also interfaces directly with Asterisk. All voice messages left by learners using the IVR, are stored in the School meals database which was implemented using the open-source MySQL relational database management system.

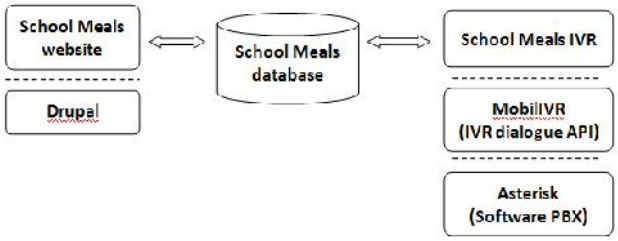

Figure 3. System architecture.

4. RESULTS

In this section we describe the user study we conducted with learners to investigate their interaction with the service. Amongst other factors we investigated 1) performance, 2) input modality preference (DTMF vs. speech), 3) language preference (local language vs. English), and 4) user experiences.

4.1 User study

4.1.1 Participants

We tested the School meals line with 35 learners (15 male, 20 female) recruited from two schools where meals are served, from an urban township region (usually located on the peripheries of towns and cities where housing varies from formal to informal) in Gauteng. Participation was voluntary, and both the parents' and children's consent was obtained. The learners were of middle and secondary school age ranging from 12-18 years old with the majority being 13-15 years old. Figure 4 illustrates the age distribution of the children.

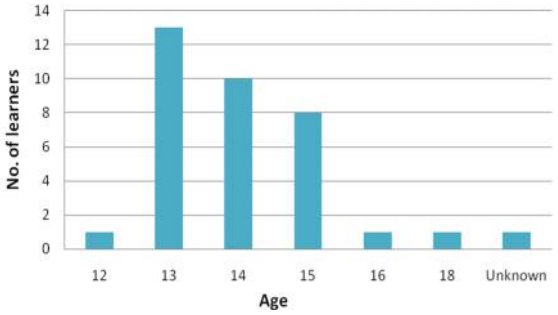

Figure 4. Learner's age distribution.

Prior to beginning the study, the children were asked to indicate the local languages they spoke, read and wrote (up to 3 languages), as well English. Figure 5 below shows the spoken language profile of the participants. Some children indicated English as part of their 3 languages (shown below) but for the specific follow-up question regarding English, 27 children explicitly indicated that they could speak it.

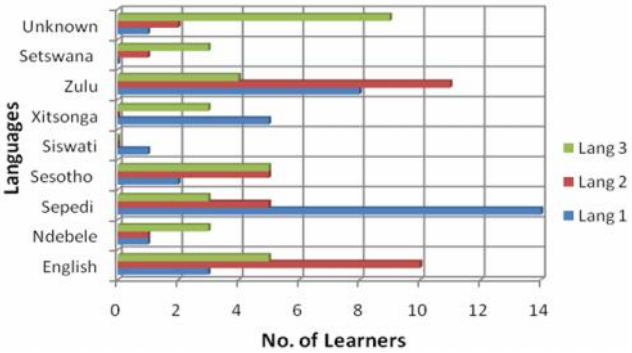

Figure 5. Language (spoken) distribution for learners.

In terms of mobile phone usage, 26 children said they used a mobile phone (74%) and of these, 19 owned a mobile phone. Typical activities cited for mobile phone usage included making calls, and playing music and games. Four children indicated that they used the phone to browse the Internet and 3 mentioned they use MXit (a free online mobile instant messenger and social

network). For computer usage, 11 children had used a computer before, with 5 saying that they owned it. Most children mentioned that computer access was not frequent since it was mostly at school, through a family member or at an Internet café and was typically for finding information for homework and playing games.

4.1.2 Methodology

We conducted a 2x2 within-subjects experiment. Four similar-in-logic prototype versions of the system were designed viz. English DTMF, English speech, isiZulu DTMF and isiZulu speech. The School meals line prototype asked the learners a number of questions on whether a meal was received and then some questions on the various aspects of the meal to indicate their satisfaction. This was followed by some open-ended questions to get more general feedback.

We simplified the logic of our design by ensuring that there were no dependencies in the design so the next state was independent of the user input, thus Wizard of Oz (WOZ) speech input was easily simulated. Barge-in on input was allowed for all 4 versions of the design. Four different stations (stations A-D) were set-up to host the 4 versions of the system. Each participant tried all 4 versions consecutively, with randomization across all learners to vary the sequence of stations per learner (4! = 24 distinct sequences). Prior to the trials, the children were given an interactive demonstration of another IVR around learning new languages (unrelated to school feeding).

At each station the child was asked to think about the last meal they had at school and then use the service to answer some questions about it. Similar mobile phone handsets were used across all 4 stations, with the children being asked to dial the number themselves. At the end of the task at each station, the child was given a post-questionnaire to answer 6 questions ranging on content, task ease, interaction pace, user expertise, length and overall experience. It was emphasized to the children that they were evaluating the system they had just tried and not those at previous stations. The questions were in a multiple choice format that the children are familiar with in school exams and were designed in conjunction with a qualified child psychologist. The experiment was conducted on two consecutive Saturdays in semi-formal lab setting (with dedicated children's activity area). The 35 learners were spread across the two days: 18 and 17 learners respectively in each group.

At the end of the day, both groups of learners participated in a focus group conducted by a qualified child psychologist and a social worker. The environment was kept informal and conversational. The children were asked questions on various aspects of the systems such as: preferences on using DTMF vs. speech input; language, viz. English vs. isiZulu; system design features; future use; and motivators.

4.2 Performance

The four systems were compared by measuring a number of objective variables, including task completion (state 0.8 onwards, where the user has answered all the food related questions and has a choice to hang up), call duration and the number of error states (timeouts and no match inputs) entered (table 1). By these objective measures, there were no significant differences between the different systems (single factor ANOVA tests with 3 degrees of freedom showed no significant differences at p=0.01 significance level).

Table 1. Average summary statistics per system.

System	Task Completion	Call duration	Total no of call states	No. of barge-ins	No. of error states
Eng DTMF	85.7%	00:01:33	9.514	0.629	0.314
Eng speech	91.4%	00:01:26	9.400	0.200	0.229
Zul DTMF	91.2%	00:01:35	9.206	1.735	0.118
Zul speech	88.2%	00:01:31	9.324	0.294	0.265

We found for the post-questionnaire results for all aspects (content, task ease, interaction pace, user expertise, length and overall experience) there were positive results reported by a large majority of the users with no significant statistical differences across the 4 systems.

To observe how children interact with speech input, we analyzed the utterances of the children for the four dialog states where there was a defined vocabulary (yes/no, or good/ok/bad). As illustrated in figure 6, we observed that the majority of the utterances were in-vocabulary i.e. the children used the correct keywords for speech input as per the dialog state. About 13% and 14% utterances were on-task (but out-of–vocabulary, e.g. any logical answer that could complete the task such as a "yeah") for English speech and isiZulu speech respectively. For English speech we found there was 6% of silence across the utterances (user did not speak at all) whilst for isiZulu speech it was 9%. "Other" constituted any other input that was out of context for that dialog state.

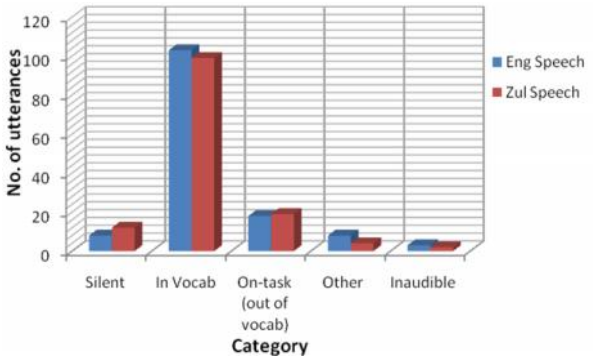

Figure 6. Utterance categorization: Defined vocabulary.

The remaining dialog states allowed more free-form input with no defined vocabulary; thus, here we analyzed the utterances in terms of the content. The majority of the utterances across all 4 systems contained "silences" (figure 7), followed by more or less an equal spread (except for isiZulu speech) across valid in-context messages (e.g. learner left their name & number when asked to or left a general message about the meals at their school) and out-of-context messages.

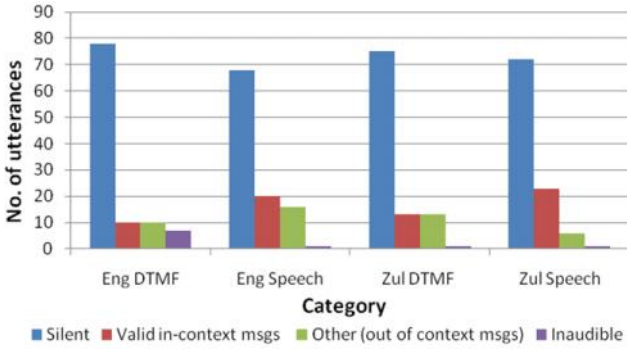

Figure 7. Utterance categorization: Free-form input.

When we examined the out-of-context utterances we found in some cases the user had left a message about a previous question in a succeeding state, e.g. the system first asks the user to leave a general message about their school meals, followed by giving them a choice to leave their name and number. Here we noticed a few participants were silent at the general message but then left a message about the school meals when they got to the next state (providing name & number).

4.3 Preferences

To investigate the preference for DTMF vs. speech input we started with a more qualitative approach during week 1. In the focus group with the participants, we again played each of the 4 systems to the children, with an emphasis on how they were different. This was followed by asking them to choose which they preferred – DTMF or speech input. Here 6 participants preferred speech and 7 preferred DTMF (five subjects were unable to participate in the focus groups). On the language preference, 9 learners indicated preference for English and 4 for isiZulu. However, observers of the group session noted significant interaction between participants in this qualitative evaluation, thus suggesting that the results may not be fully valid.

During week 2 we therefore revised our strategy and created a simple ranking scheme where each participant individually ranked all 4 systems from 1-4 (1 being 1st preference). Table 2 below shows the results of the second group's rankings. Here we see that 15 (of the 17) participants gave English speech the 1st rank and of these 2 participants (in brackets) had indicated that isiZulu was Language 1 in their demographic questionnaire.

Table 2. Comparative preferences through ranking.

Rank	Eng DTMF	Eng speech	Zul DTMF	Zul speech
1	0 (0)	15 (2)	0 (0)	2 (1)
2	10 (1)	2 (1)	1 (0)	4 (1)
3	5 (1)	0 (0)	3 (1)	9 (1)
4	2 (1)	0 (0)	13 (2)	2 (0)

Figure 8 illustrates the above data, from which it can be seen that there was strong preference for the English speech system (1st rank), followed by English DTMF, isiZulu speech and finally isiZulu DTMF. Both binary distinctions were statistically significant for DTMF vs. speech input at the $p<0.0001$ level (using a binomial distribution) for English vs. isiZulu at the $p<0.01$ level.

In the focus group some of children indicated that "it was easier to speak into the phone" referring to the speech system. With regards to the language preference, an interesting observation was also made during the focus group, where some learners mentioned that they should be given an option to choose their language before they start (in principle – a language menu upfront).

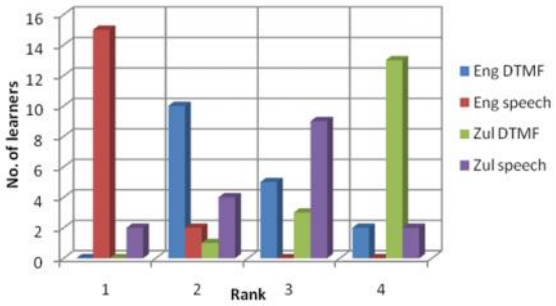

Figure 8. Comparative ranking of the various systems.

4.4 Usage Observations

In this section we highlight a number of issues around usage which were observed during the focus group discussions.

System design features

We found through the focus groups that the learners were quite astute in providing practical design recommendations. We prompted them through basic seed concepts such as "which language do you prefer?" and they easily came up with suggestions. For example, as indicated earlier, learners specifically mentioned "*there should be choice between the different languages at the beginning of the conversation* (with the system)". They also indicated that the system should be able to "*repeat the question*" in case they get confused.

Both groups wanted a *longer timeslot to leave messages* (state 0.7) regarding their experiences and needs. They also indicated that *more questions* on the type and quality of the food, the ability of the kitchen staff to prepare hygienic, healthy and well-cooked food should be added to the system. In order to keep the system interaction short and not cumbersome, we included the most important questions about the school meals that affect learners, and kept questions that could be addressed by school meals coordinators out of the learner meals line. The "leave a message" state (0.7) was intended to provide the facility for additional feedback related to the learners' meals in general. However, it seems that the *learners did not explicitly make the connection that through leaving a message they could tell us more about the other issues that they had wanted addressed* through additional questions.

Through this interaction with the learners on design aspects we find that *learners can indeed make perceptive and practical design choices*, perhaps more so than adults, based on our previous experiences [17].

Motivators

The learners from both groups indicated that the system was "fun" to use and although the learners from the second group indicated that games could be added to make it "more fun", and they

brainstormed on it; no final suggestions were made. Learners in the second group mentioned that they want the system to be presented at school so that the other children can also give ideas on how to make it more fun.

Here the challenge lies in *ensuring the system was engaging enough for the children to capture their attention whilst conveying the gravity of providing legitimate feedback* (e.g. no pranks). A potential option would be to provide some form of incentive for the children but this may not necessarily set the right motivation for children to provide feedback about their meals. Both groups also indicated that the fact that this is a free call service is a motivator to make use of it.

It was noted that some children in the first group indicated in terms of the engageability of the system that "*when they* [system] *talk to you, you just feel interested to them"* trying to convey that they felt that the 'service cared'. During the system trial sessions as well, a number of learners mentioned that the lady (system voice) was talking to them and asking them questions "nicely". Given the learners' feelings that their concerns are not always attended to, we feel the persona of the system (caring and wanting to listen to them) and the child-centric style of the interaction will be crucial in making them feel at ease and gain their trust in the efficacy of the service.

Future use – Trust and Confidentiality

In both focus groups we tried to investigate the factors that would affect the learners' decision to use the system to voice their feedback to the DBE. Initially both groups indicated that they would use the system again and that they thought that DBE would use the system to improve the meals served at their school. Later in both group discussions, the learners admitted that there was a lack of trust towards the educational system as the learners' concerns were not always addressed, and indicated that they were worried about possible victimisation if they reported negative aspects to DBE. They further indicated that there might be other learners who would prefer not to use the system as they "…won't feel ok to talk to a person they never met on a phone line".

The learners from both groups indicated that they would leave an anonymous message on the system as it might lead to improvement of the NSNP. They were ambivalent about leaving their names and numbers. Some indicated that they would like DBE to contact them, whilst others were worried about possible victimisation. The discussions finally revealed that learners were not convinced that using the system to report on the meals they received at school, would lead to improvement of the NSNP. However, they expressed the hope that it would make a positive difference. It was also clear from the discussion that *training provided for the use of the system* and *reassurance about the confidential use of the information provided* by learners *would motivate them to make use* of it. The learners of the second group also indicated in the focus group that they wanted the system to be available every day and wanted it to be introduced as soon as possible.

In terms of practical usage, learners from both groups indicated that they would call the system after school (afternoon). The learners from the first group indicated that they would use mostly public telephones or their parents' cell phones (although cell phone usage and ownership (section 4.1) was spread equally across the two groups), while the learners from the second group indicated that they own cell phones, which they would use.

The above findings indicate two major underlying aspects that needed to be addressed: 1) *trust* – it was evident that the learners require a complete understanding of the context of the service that not only addresses how it could help them but also indicates the "where" and the "who", i.e. *addresses the concern regarding where the information is going and who is listening to it?*, and 2) *confidentiality* – this aspect relates to ensuring that the learners feel that the information they provide will lead to them being identified and possibly victimized, i.e. *"how" the information the learners have provided will be dealt with.*

Preference scales vs. forced-choice rankings

In general, we found that the richness of the inputs received through the focus groups as described above, did not correlate fully with the post-questionnaire results (which reported mostly positive experiences without providing any suggestions for improvement). We asked both groups about the use of the post-questionnaire, and learners from both groups indicated that it had been easy to use. Upon further discussion, some learners in the second group were able to explain that their answers differed somewhat because their understanding of the system's context and use had improved and had changed during the focus group discussion. (It is also possible that the children's levels of comfort with the facilitators increased over the span of the focus group discussion and that this contributed to them expanding on their initial impressions.) This highlights the challenge of using typical usability-type post-questionnaires (despite thorough simplification) when working with users unfamiliar with ICT-based solutions. Based on our experiences we found that the *value of using a triangulated approach for analysis* through using call logs, focus groups, questionnaires, and user observations cannot be overstated.

Another major methodological finding from the study was the decision to move from *preference scales* (similar to the post-questionnaire) *to a forced-choice ranking style to evaluate user preferences*. In contrast to the previous observation, the questionnaire was found to be much more informative in this case. This exemplifies a general principle of user testing in the developing world: since users often tend to be extremely polite, it may be necessary to design specific mechanisms that elicit useful responses without transgressing this cultural norm.

4.5 Design Recommendations and Revisions

We learnt some valuable lessons from the user study – especially from the focus group sessions and interactions with the participants. This enabled us to make a number of design changes to the system, and these were implemented for roll-out during the piloting phase.

As is clear from the previous section, the need to create a more trusting environment was paramount: the users needed to be ensured that they would not be victimized or exploited if they were to provide information – especially negative messages – to the system. In our re-design, we had to keep in mind that (a) the users are young, and therefore require not only direct, verbal assurance that they could trust the system, but also more indirect, tacit assertion; and (b) since the users typically do not have wide exposure to this kind of technology (i.e. telephone-based voice services), they should also be made comfortable with the technology, in order to enhance trust.

To this bring about, our main strategy was to make the persona more human-like. Initially, the driving design principle behind our initial design was one of de-anthropomorphism, or as per Balentine [2], we tried to develop a good machine rather than a bad person. Therefore, in our first design we tried to be as direct and to-the-point as possible, enabling the user to get on with his/her business as quickly as possible. However, keeping in mind that our users are young and have very limited exposure to similar machine-like applications, we learnt during the user study that we needed to take a less machine-like approach, and that we should try to humanize the system a bit more (without trying to pretend that it is actually a human). Some of the changes we implemented are summarized below:

We created a *fictitious persona*, called "Mama Nandi" (translated directly as "Mother Nice", and where Nandi is a common South African name). This persona is not used as the persona for the School meals line, but rather as a persona in the project, who also appears on marketing material, etc. In the School meals line, the system voice refers to Mama Nandi by stating that "*this is Mama Nandi's answering service. She is trying to improve the school nutrition programme, and this is a safe place where you can tell her about the meals at your school*".

Note that it is not Mama Nandi that speaks (which would have been a choice for an anthropomorphized system), but rather another person who represents her answering service, referring to her in the third person. The establishment of this persona aims to contribute towards making the experience more concrete for the learners. For example, in the initial application the voice said: "You can tell **us** about…", but learners in the user study said that they were not sure to whom they are providing this information – it could have been their teachers, or the government, or the police. In order to make this more concrete, we therefore changed the prompt into "You can tell **Mama Nandi** about…".

For the sake of making the service more human-like (and therefore trusting) we also made the dialog flow and prompts more natural.

- For example, *discourse markers* like "Let's start" and "OK, we're almost done" and "OK, finally…" were introduced to create a softer, more human-like application.
- We also introduced *explicit statements* to enhance trust; for example, the initial greeting, "Hola! Thanks for calling the school meals line", was rephrased as "*Hola! Welcome to the anonymous school meals line*" in the redesign.
- Also, in the initial design learners were requested (but not required) to leave their name and telephone number at the end of the call; during the user study it was pointed out by learners that this might be problematic, since learners would be afraid that the information might find its way back to one of their teachers. Therefore, we opted to have less rich information (i.e. names and telephones numbers of callers), but rather a more trusted service and removed the prompt to leave a name and number.
- Lastly, in the original design it was required of the user to state the name of their school immediately after the greeting prompt. However, during the user study we learnt that our typical users were a bit overwhelmed by this immediate action that was required of them. Since they were not that familiar with the technology, a number of them were silent at this question. In order to ease them into the service, and to first create trust (i.e. by not asking them immediately for personal information), in the re-design we postponed this question until the end of the dialog, again softening it with explicit verbal queues like "anonymously" and "you can help Mama Nandi" before asking for the school name e.g. "*You can help Mama Nandi by telling her anonymously about the meals at your school To start, say the name of your school after the tone and then press hash. <Tone.*". The idea that the user is helping to improve the NSNP is echoed in the marketing material through the slogan of the project: "Use your voice!" (Figure 9).

Figure 9. Marketing of the "School meals line".

In addition to these changes aimed at creating a more trusting environment, we introduced some tweaks and changes to *increase the user-friendliness and usability* of the application. These included the following:

- Since this application is intended to be implemented in multilingual environments, and since we cannot assume high levels of multilingualism in our target population, we introduced a *simple language menu* upfront (i.e. immediately after the greeting).
- *Changes to the call-flow sequence*: Such as, for questions to follow more logically and naturally on each other, or to replace some questions with others that would provide better information to the user (e.g. we removed the question around "How was the food?" to "Was the food cooked well?")
- We attempted to make *navigation easier* by introducing dialog markers such as "OK" and "Let's start", as well as making instructions easier to follow. For example, instead of "press hash when you're done" in the earlier design where some learners interpreted it as, hash was to be pressed when they were done (completed) with the call (as opposed to recording), we decided to use "do XYZ and then press hash".
- We introduced an error management strategy to assist users on time-outs (i.e. silence or no key pressed), erroneous input (e.g. when "4" is pressed if there are only three options), and on repeated errors (i.e. an exit strategy).

Lastly, the design team also formulated some explicit directions for the voice artist with regard to where word stress should be placed (for example on "safe" in "this is a safe place"), tone and tempo. These directions were thought to contribute to developing a more trusting – and hence more efficient – service.

5. DISCUSSION

As highlighted in sections 4.4 and 4.5, various insights were obtained around the overall user experience of the learners where trust and confidentiality will play a crucial role in uptake. Based on these insights we suggested several design recommendations for designing voice-based services for this target user community.

In terms of modality and language preferences we found large differences between our first and second subject groups; however, as mentioned earlier, we have reason to doubt the outcomes from the initial group. We therefore assume that the results of the second group are more reliable for these preferences, but return to this matter below.

Based on the findings summarized in the Results section, we deduce that children aged 12-18 using a VUI-based service significantly prefer speech input over DTMF. Some reasons cited by a few participants were that it was easier to use and understand. This finding is somewhat consistent with [27] where in the case of an audio survey VUI service, adults' preference and task performance was positive for speech input. However, in our study we found that there were no significant task performance differences in the case of children, probably due to the simplicity of the task.

A major question in determining the feasibility of using speech input with children was around the use of limited keywords versus children being verbose with the system. As we see from the utterances for states with defined vocabulary input, 74-75% were in-vocabulary and another 13-14% were on-task. This shows that our target population, namely children of 12-18 years, was comfortable with speech input and was not verbose at all. This notwithstanding, we did face some of the same challenges noted by [27, 31] where, in the first few states of interaction some children were not sure whether they could speak to the system and when exactly to speak.

The high number of "silences" at the free-form input dialog states may be due to various reasons, such as the fact that the environment and task were staged, the user was being observed thus he/she may not have had a "real" message to leave or may have felt shy. We plan to observe this phenomenon further in a pilot deployment as the intention of allowing the free-form input states was to give the children an opportunity to voice their concerns.

In terms of language preference, it is striking to note the overwhelming preference for English over isiZulu (even though the majority speaks local languages that are more closely related to isiZulu than to English). We think that this may be due to various reasons. Firstly, in trying to improve the English language skills of learners, teachers often encourage them to choose English over any other local language for communicating with the external world. As one child put it succinctly, "my teacher says that we must speak English, otherwise we will not get jobs". This shows that for learners, English language skills are perceived as a prerequisite for social progression. Secondly, some children may have found it easier to use the system in English since English is employed in most of their interactions with technology. Finally, some of the children mentioned that they would like to have the system in another language (Sepedi or Sesotho) and indicated that they would have preferred the option to choose in which language to use the system.

6. CONCLUSION & FUTURE WORK

The redesigned system is currently being piloted in 2 schools in urban South African townships. These pilots will allow us to assess how successful we were in achieving trust and acceptance in the target population – issues which were seen to be important unknowns during our focus group studies. The other unknown that we have found to be crucial for services of this nature relates to the marketing campaign: user uptake of such a novel (for them) technology depends strongly on exterior factors such as community involvement, implicit and explicit rewards, etc.

We therefore intend to monitor both uptake and usage of the School meals line closely, and to assess its impact through interaction with several stakeholders – including learners, parents, teachers and personnel of the DBE. We hope to provide a useful efficiency gain for this vital service, and in the process to learn about the design of useable voice services aimed at children in the developing world.

7. ACKNOWLEDGMENTS

This work was funded by the South African Department of Arts and Culture under the Lwazi II project. We also thank the National School Nutrition Programme of the Department of Basic Education for the opportunity to work in this domain. We thank the various team members from CSIR Meraka Institute who provided valuable contributions throughout the project; Olwethu Qwabe, Richard Carlson, Bryan McAlister, and Tshepo Moganedi. And most importantly, we are very grateful to the school learners who participated in this user study for the school meals service.

8. REFERENCES

[1] Agarwal, S., Dhanesha, K., Jain, A., Kumar, A., Menon, S., Nanavati, A., Rajput, N., Srivastava,K., Srivastava, S. 2010. Organizational, Social and Operational Implications in Delivering ICT Solutions: A Telecom Web Case-study. In *Proc. ICTD 2010*, London, UK.

[2] Balentine , B. 2007. It's Better to Be a Good Machine Than a Bad Person: Speech Recognition and Other Exotic User Interfaces at the Twilight of the Jetsonian Age, ICMI Press.

[3] Barnard, E., Plauché, M. and Davel, M. 2008. The Utility of Spoken Dialog Systems. In *Proc. IEEE SLT 2008*, 13-16.

[4] Botha, A., Herselman, M. and Van Greunen, D. 2010. Mobile User Experience in an Mlearning Environment. In *Proc. SAICSIT 2010*, 29-38.

[5] Boyera, S. 2009. *Mobile Web for Social Development Roadmap*, in *W3C Interest Group Note 16 November 2009*. W3C11.

[6] Bundy, D.A.P., Shaeffer, S., Jukes, M. et al. 2006. School-Based Health and Nutrition Programs. In *Disease Control Priorities in Developing Countries*, Ch. 58, 1091-1108.

[7] Butgereit, L. 2009. How Dr Math reaches pupils with competitions and computer games by using MXit. In *Proc. IST-Africa 2009*, 7.

[8] Butgereit, L., Botha, A. and Van Niekerk, D. 2010. Using Cell Phones to Improve Language Skills: The Hadeda Project. In *Proc. AFRICOMM 2009*, LNICST 38 (2010), 11-19.

[9] Central Board of Secondary Education – post result CBSE annual counselling IInd phase. http://cbse.nic.in/Press%20Note%20Second%20Phase_2011.pdf.

[10] Corkrey, R., & Parkinson, L. 2002. Interactive voice response: Review of studies 1989–2000. *Behavior Research Methods, Instruments,& Computers*, 34, 342-353.

[11] CSIR, Meraka Institute. OpenSpell Educational Game. http://www.meraka.org.za/openspell.htm

[12] Department of Basic Education. Call for comments on Action Plan to 2014: Towards the Realisation of Schooling 2025. Government Gazette, Notice 752 of 2010.

[13] Department of Basic Education. National School Nutrition Programme – A Guide for Secondary Schools, 2009. http://www.education.gov.za/LinkClick.aspx?fileticket=aUXWnnj3f5A%3d&tabid=93&mid=1130

[14] Department of Basic Education. NSNP: Approved 2010/2011 Provincial Menus. 2010. http://www.education.gov.za/LinkClick.aspx?fileticket=ZAhbLpd2HiY%3d&tabid=93&mid=1131

[15] Gorman, T., Rose, E., Yaaquobi, J., Bayor, A. and Kolko, B. 2011. Adapting Usability Testing for Oral, Rural Users. In *Proc. CHI 2011*, 1437-1440.

[16] Grover, A.S. and Barnard, E. 2011. Comparing Two Developmental Applications of Speech Technology. In *Proc. HLTD 2011*, 81-86.

[17] Grover, A.S. and Barnard, E. 2011. The Lwazi Community Communication Service: Design and Piloting of a Voice-based Information Service. In *Proc. WWW 2011*, IW3C2 (2011), 433-442.

[18] Grover, A.S., Stewart, O. and Lubensky, D. 2009. *Designing interactive voice response (IVR) interfaces: localisation for low literacy users*, in *Conference on Computers and Advanced Technology in Education (CATE 2009)*. 2009: St Thomas, US Virgin Islands. 8-15.

[19] Grover, A.S., Plauché, M., Barnard, E. and Kuun, C. 2009. HIV health information access using spoken dialog systems: Touchtone vs. Speech. In *Proc. IEEE Int. Conf. on ICTD 2009*, 95-107.

[20] Heeks, R. 2009. *The ICT4D Manifesto: Where next for ICTs and international development?* In *Development Informatics working paper no. 42*. 2009, Centre for Development Informatics: University of Manchester: Manchester, United Kingdom.

[21] IVRs based daily monitoring of the MDMS in UP. http://mbillionth.in/2011/08/05/ivrs-based-daily-monitoring-system-of-the-mdms-in-up/.

[22] Kam, M., Kumar, A., Jain, S., Mathus, A. and Canny, J. 2009. Improving Literacy in Rural India: Cellphone Games in and After-School Program. In *Proc.Int. Conf on ICTD 2009*, 139-149.

[23] Kam, M., Agarwal, A., Kumar, A., Lal, S., Mathur, A., Tewari, A. and Canny, J. 2008. Designing E-Learning Games for Rural Children in India: A Format for Balancing Learning with Fun. In *Proc. DIS 2008*, 58-67.

[24] Kote, T. and Barman, S. 2011. A Comparative Study of SMS and IVR Interfaces for Crowdsourcing Water Availability Information. In *Proc. CHI 2011*. ACM: Vancouver, BC, Canada.

[25] Kuts, E., Islas Sedano, C., Botha, A. and Sutinen, E. 2007. Communication and Collaboration in Educational Multiplayer Mobile Games. In *Proc. CELDA 2007, IADIS*, 295-298.

[26] Lall, A. and Sahi, S. 2009. Taking ICTs to the Grassroots: A Case Study of the LifeLines India Initiative. *Information Technology in Developing Countries*, 19 (2009).

[27] Lerer, A., Ward, M. and Amarasinghe, S. 2010. Evaluation of IVR data collection UIs for untrained rural users. In *Proc. ACM DEV 2010*, 1-8.

[28] Medhi, I., Patnaik, S., Brunskill, E., Gautama, S.N.N, Thies, W., Toyama, K. 2011. Designing mobile interfaces for novice and low-literacy users. In *ACM Trans. Comput.-Hum. Interact*. 18, 1, 1-28.

[29] Medhi, I., Menon, S.R., Cutrell, E. and Toyama, K. 2010. Beyond Strict Illiteracy: Abstracted Learning Among Low-Literate Users. In *Proc. IEEE Int. Conf. on ICTD 2010*, London UK.

[30] Medhi, I., Gautama, S.N.N. and Toyama, K. 2009. A Comparison of Mobile Money-Transfer UIs for Non-Literate and Semi-Literate Users. In *Proc. CHI 2009*, 1741-1750.

[31] Patel, N., Chittamuru, D., Jain, A., Dave, P. and Parikh, T.P. 2010. Avaaj Otalo – A Field Study of an Interactive Voice Forum for Small Farmers in Rural India. In *Proc. CHI 2010*, 733-742.

[32] Patnaik, S., E. Brunskill, and W. Thies. 2009. Evaluating the accuracy of data collection on mobile phones: a study of forms, sms, and voice. In *Proc. IEEE Int. Conf. on ICTD 2009*, 74-84.

[33] Plauché, M.P. and Nallasamy, U. 2007. Speech Interfaces for Equitable Access to Information Technology. *ITID Journal* 4, 1, 69-86.

[34] Public Service Commission. Report on the Evaluation of the National School Nutrition Programme (NSNP). 2008. Pretoria. http://www.info.gov.za/view/DownloadFileAction?id=88572

[35] Sherwani, J., Ali, N., Rosé, C.P. and Rosenfeld, R. 2009. Orality-Grounded HCID: Understanding the Oral User. *ITID Journal* 5, 4, 37-49.

[36] Sherwani, J., Palijo, S., Mirza, S., Ahmed, T., Ali, N. and Rosenfeld, R. 2009. Speech vs. Touch-tone: Telephony interfaces for information access by low-literate users. In *Proc. IEEE Int. Conf. on ICTD 2009*, 447-457.

[37] Stritzke, W., Dandy, J., Durkin, K. and Houghton, S. 2005. Use of interactive voice response (IVR) technology in health research with children. *Behaviour Research Methods* 37, 1, 119-126.

[38] Tomlinson, M. School feeding in east and southern Africa: Improving food sovereignty or photo opportunity? *Equinet Discussion Paper Number 46*. http://www.equinetafrica.org/bibl/docs/DIS46nutTOMLINSON.pdf

[39] Walsh, C.S., Shrestha, P. and Hedges, C. 2011. Leveraging Low-Cost Mobile Technologies in Bangladesh: A Case Study of Innovative Practices for Teacher Professional Development and Communicative English Language Teaching. In *Enhancing Learning Through Technology. Education Unplugged: Mobile Technologies and Web 2.0*, Kwan, R. et al., Editors. 2011, 152-166.

[40] World Food Programme – School Meals. http://www.wfp.org/school-meals.