A Voice User Interface for Low-literacy Users in a Rural Community

EDIM, A. EMMANUEL†
Department of Maths/Stats., & Computer Science, University of Calabar, Nigeria
MUYINGI, H. NSUNG-NSA,
Department of Software Engineering, The Polytechnic, Namibia
SIBANDA, K.
Department of Computer Science, University of Fort Hare, South Africa

ABSTRACT

The mobile phone offers a medium for harnessing the potentials of information and communication technology services to contribute to the social and economic development of low-literacy users in rural underserved communities. The design of user interfaces for information and communication services for low-literacy users remain a big challenge. User interfaces that enable the low-literacy users to interact with the services provided should be made simple and meet the experience and preferences of users. The aim of this study is to design a voice user interface that meets the experience and preferences of low-literacy users in a rural community. We applied a user centered approach and ethnographic research methods to collect quantitative and qualitative data for the design of two varieties of voice user interface. The voice user interface varieties were evaluated with users from the rural community and the results show that the interface varieties were simple and easy to use and they meet the natural experience of the users. A higher percentage of the users prefer to use the dual-tone multi-frequency input user interface than the voice input user interface for interacting with a mobile commerce application.

Categories and Subject Descriptors: H.5.2 [Information Interfaces and Presentation]: User Interfaces—*User-centered design, Voice I/O, Natural language, Interaction style;* **H.5.3 [Information Interface and Presentation]**: Group and Organization Interfaces—*Web-based interaction*

General Terms: Voice Interaction, User Interface

Keywords: ICT, Banking, Cashless Transaction, Nigeria, Economy

Additional Key Words and Phrases: Voice user interface, mobile commerce, dual-tone multi-frequency, low-literacy, rural community

IJCIR Reference Format:

Edim A. Emmanuel, Muyingi, H. Nsung-Nsa, Sibanda K. A Voice User Interface for Low-literacy Users in a Rural Community. International Journal of Computing and ICT Research, Vol. 7 Issue 1, pp 56-69. http://www.ijcir.org/volume7-issue1/article6.pdf.

1. INTRODUCTION

User interface and interactions design have received tremendous attention in the research community. These efforts have lead to different proposals on different types of user interfaces and interaction techniques. Very few of these user interface innovations have been tested in low literacy communities. Information accessibility remains difficult for

†Author's address: Edim A. Emmanuel, Department of Maths/Stats., & Computer Science, University of Calabar, Nigeria; Muyingi H. Nsung-Nsa, Department of Software Engineering, The Polytechnic, Namibia; Sibanda K. Department of Computer Science, University of Fort Hare, South Africa.

International Journal of Computing and ICT Research, ISSN 1818-1139 (Print), ISSN 1996-1065 (Online), Vol. 7, Issue 1, pp. 56-69, June 2013.

low-literacy people living in underserved communities [Sherwani et al. 2009]. Poverty and lack of infrastructures challenges also pose as obstacles to information access in rural communities in Africa. Desktop computers are not common in rural communities. User interfaces designed for the highly literates pose usability challenges to low-literacy users. In places where they have been introduced, the community find it difficult to use them due to literacy challenges and lack of motivation as a result of not being aware of any benefit this technology will add to their livelihood [Edim and Muyingi 2010a]. With the proliferation of mobile phones to rural communities, easy information access is possible to the rural users. Mobile phones are easy to use and the users in rural communities are already witnessing the added value these devices offer to the users. These ubiquitous devices provide platform for information access and delivery of services for development of large population of rural dwellers [Medhi et al. 2009]. Mobile phones offer different services to the users, such as multimedia messages, short message service, voice and data services.

Huang and Deng [2008] noted that mobile phones as an ICT tool have changed the way people interact with each other and between communities and cultures. The mobile phone has affected both social and cultural lives, the manner in which people interact with one another and their environment at their work places, at home and other places [Sato and Chen 2008]. These changes occur in people because of the deep attachment they have with these technologies both socially and economically. Also, users derive added value as they use the technology on daily basis. The low-literacy inhabitants in rural underserved communities can also derive these benefits and also have access to information for socio-economic development. In order to achieve this, the user interface which serves as the medium of interaction should be designed to meet the natural experience and preferences of low-literacy users in rural communities.

Low-literacy users in rural communities use ICTs in ways that are different from people who are functionally literate and living in urban centres [Lalji and Good 2008]. The expectations and preferences of low-literacy users are different. And so, user interfaces designed in a language (e.g. English) different from the natural/home language of the target users, usually have negative impact on the development of ICTs in rural communities [Kam et al. 2007]. Due to low literacy and technology inexperience in underserved communities text-based interfaces are often difficult to use among these groups of users [Medhi et al. 2006]. Among low literacy rural users, the use of the mobile phone is often restricted to voice calls only due to literacy challenges, cost and lack of adequate experience to use most of the technologies mobile communication provide. Voice interaction is the commonest medium of interaction that cuts across all users of the device irrespective of literacy level and cultural background. Speech interfaces on mobile phones for information access in rural communities may be able to provide intuitive interaction and ease of use to low-literacy users who are technologically inexperienced.

Several speech recognition and synthesis systems have been designed to aid speech-based interactions. Applications that use these technologies have demonstrated the potential of speech technology for communication between users and machines. Voice-based interfaces are accessible and provide easy interactions to every category [Sherwani et al. 2009]. This interaction technique conforms to oral form of communication that is a very common among rural dwellers [Pade et al. 2009]. Well adapted voice interface that meets the experience and preferences of the users will motivate the users and help them to have access to information.

This study was conducted to design a voice user interface that meets capability and preference of low-literacy users in Dwesa rural community in South Africa. The interface will enable the users to have access to their mobile online shops in a mobile commerce application. In this community, the inhabitants are subsistence farmers to a large extend and few of them also practice local crafting. Due to the absence of developmental infrastructures and poverty in this community, local crafters find it difficult to project their trade outside the community and to attract potential customers. This study considered the development of a mobile commerce application with a voice interface that will be easy to use by low-literacy users in the community and will enable the crafters and other community members to have easy access to the mobile commerce application that is hosted in the Siyakhula Living Lab (SLL) project within the community. The SLL provides Information and Communication (ICT) facility installed in the community to enhance ICT awareness and provide services that will improve rural livelihood in the community. We have designed and evaluated the voice user interface with participants from the community. The results show that the participants recorded high tasks completion rate with few errors when they used the two varieties (DTMF input and voice input) of the voice interface even though they were inexperience in terms of using voice applications.

2. RELATED WORK

Voice based interfaces have the advantage of being able to accommodate a wide variety of users irrespective of their physical disabilities and level of education and technology literacy. Different research efforts have been made in the design of voice-enabled interface for different groups of users in different regions. For instance, a voice-based interface and application that provides the users access to Web service is discussed in Tsai [2006]. The voice interface enables the users to interact with the web service using the Mandarin language and through a VoIP telephone service.

The application provides Automatic Speech Recognition (ASR) and Text-To-Speech (TTS) synthesis and telephone lines for voice inputs and feedbacks.

Several attempts have also been made to use voice interfaces to assist the low-literacy communities to have access to information. Healthline [Sherwani et al. 2007] is a speech-based user interface designed to enable low-literacy rural health workers in Pakistan perform health services to their community. The user study shows that the low literate community health workers were able to interact with the interface and had a high level of tasks success. The users found the system useful and less difficult to use. Patel et al. [2009] conducted a user study to determine users' interaction preference between two input modes among low-literacy users in rural India. The prototype system is a Gujarati language application that helps farmers in this region to have access to agricultural information. The outcome of the user study showed that the participants recorded a higher task completion rate with the DTMF input interface than the voice input interface. Patel et al. observed that different reasons were cited by the participants for poor performance with the voice input interface such as the slow and unnatural way of speaking input into the system, and difficulty in handling errors during interactions. The participants were however excited about the application and interface and feel that it will help them to improve their farming practices. In a similar study, Sherwani et al. [2009] designed a speech and touch tone telephone interfaces for low literate users in rural Pakistan to have access to health information. The comparative study carried out by Sherwani et al. [2009] to determine the user preference between the touch tone and speech input showed that the touch tone outperformed the speech interface. The user studies showed that the users found the interfaces less difficult and the recognition accuracy was very high. Our work is similar to the studies described above in terms of the literacy level of the users, technology experience and the type of user interface design for user interactions with the application. Where we differ is the application domain which also imposes a different design approach and the level of user interactions or utterances provided. In our previous study, we designed and tested a localized mobile graphic user interface [Edim and Muyingi, 2010a] on a mobile commerce application for low-literacy users in Dwesa. This study is connected to the previous study in the sense that the voice user interface will be used by the same user group to have access the mobile commerce application.

Plauche et al. [2006] discussed a voice user interface designed to enable low-literacy farmers in the Tamil region in India to have access to market information on agricultural produce. In this design, the prompts were restricted to simple yes or no answers so that the interactions were simple enough for higher accuracy. The application was a Tamil spoken language system designed by Berkeley's TIER group to test speech interface in low-literacy environment. The system gave low-literacy users access to crop information in order to help improve crop production in the community. Lo and Meng [2002] discussed a WAP-based mobile information system for accessing Web-based services. The users could gain access to information such as stock alert service, weather information, and news. Parikh et al. [2006] described an approach for capturing data from paper documents using mobile phones, while the users receive voice feedback from the application. The application was designed to assist low literate users in rural India to manage the finances of micro-finance groups. Voice interactions are capable of providing every category of user irrespective of educational background or technological experience with easy access to information. Simple and spoken commands enable the user to establish the needed communication with the application [Zhou 2007].

2.1. ICT Services and UIs Design for Low-literacy Users

ICTs are a driving force for sustainable economic development in the 21st century [id21 2007]. The rural communities are yet to experience these services to a level that they will begin to experience economic values in their livelihood. This is because they are still viewed as risky markets and are yet to be explored to a large extent [Heeks 2008]. ICTs can bring positive impact in the lives of people even among the rural underserved communities. This has been proven to a little extent in places where ICT services have been implemented. Few examples are worthy of note. Medhi et al. [2006] describes a text free user interface on an ICT service that was used to support semi-literate and illiterate

domestic labourers in rural India. The users use the user interface to search for domestic jobs within their neighborhood. Patel et al. [2008] also discussed a mobile voice user interface to help subsistence farmers in rural India to access and share farming information with each other.

Gosh et al. [2003] designed a user interface to enable community-based microfinance in rural India. The members of a self help group use the user interface and application for saving and lending money to the group members. They also used the application to keep track of their savings and to attract government attention. Patel et al. [2009] designed and tested a speech versus Dual-Tone Multi-frequency (DTMF) input interfaces. In the study, they found that low-literacy users performed better when the interacted with the application using DTMF input interface. The CAM user interface [Parikh et al. 2006], is a successful implementation of ICT project for development. The application helps semi-illiterate mobile phone users to use the phone camera to capture financial data from paper documents. The camera reads the visual codes on the paper to produce individual's financial statement. The low-literacy users in this study were able to understand the user interface and used it with ease.

Successes have been recorded where ICT services have been provided for low-literacy users in rural communities. A lot of effort is still desired in this direction due to the large population of people living in underserved communities in developing countries. These communities do not have the technology found in urban cities. Low-literacy users have limited experiences with ICTs. In order to increase their skills and experiences, the natural knowledge and preferences of the users should be elicited and applied in the user interface design. These groups of users perform better when the user interface is designed to conform to their experiences and capabilities [Edim and Muyingi 2010a].

2.2. Mobile Services

Mobile ICT services like mobile commerce provide added value to the users. ICT services bring new markets and business opportunities to IT professionals, ICT vendors and service providers [Heeks 2008]. In rural underserved communities, it gives the mobile phone users added value and promote micro-enterprises. Low-literacy users will derive these benefits and many more if the user interfaces to ICT services are adequately designed and the usability problems are greatly reduced. Although the design of ICT services for rural underserved communities comes with different challenges that require different views and approaches. The challenges also require that the developers understand the new environment. This is because the socio-cultural environment is different from that of the urban cities and also has a great influence on the behavior and the experience of the people. As the people interact with their environment on daily basis, it affects their way of life and how they interact with new technology. When the users' cultural experience within their cultural environment is properly applied to user interface design, it will enhance user experience and user interface acceptance [Shen et al. 2006].

The penetration of mobile phones to rural communities had made it possible to make mobile ICT services accessible (e.g. mobile commerce) in these communities. This study looked at the design of an acceptable voice user interface that meets low-literacy user experience and preference. The voice user interface was designed to serve as an additional interaction technique to a localized graphical user interface designed to enable rural micro-entrepreneurs to have access to their online shops in a mobile commerce application [Edim and Muyingi, 2010b].

3. METHODOLOGY

This study was conducted in Dwesa rural community with the involvement of the community members. Ethnographic research methods were conducted to have contact with participants from the community. Interactions with different groups of participants took place in the community (e.g. school premises and art and craft centers in the community). Regular site visits to the community were conducted at least once or twice in a month during the course of the study. With these visits, we were able to build relationship with the community.

We started data collection for the user interface design during the period we were designing the graphic user interface for the mobile commerce application [Edim and Muyingi 2010a]. Data collected for the graphic user interface also served as input to the design of the voice user interface. Since we do not understand the local language (isiXhosa language), we wanted to have a tangible prototype to lead us into the elicitation of data and the design of the voice user interface.

A questionnaire was distributed to members of the community to find out the level of mobile phones services usage in the community. The data include the level of awareness and use of the internet, short message service, multimedia services and voice mails. These services if they are used by community members will influence their understanding of mobile commerce and user interfaces. 120 respondents returned completed questionnaires. We held the first focus groups meetings with crafters and educators at different times to elicit information about interaction elements on the interface. The prototype graphic user interface was used to facilitate the interactions with the groups. The crafters consist of 9 participants with an average of 49 years. Among the crafters were 6 illiterates and 3 semi-literates. A facilitator was used to serve as a connection between us and the participants. The interactions were held in Ngwane art and craft centre. The educators were the functional literate members of the community. These were teachers in the schools in the community and the interactions were conducted in the school's laboratory. The average age of the educator was 38 years. We were able to interact with this group fluently and useful data concerning interface voice metaphors, commands, and text in the local language were collected for verification. Interviews were also carried out with members of the community in their homes. The interview was conducted to get correct pronunciation of certain isiXhosa words according to the dialect of the region. The data was captured using an audio recorder for analysis.

We also conducted focus groups meetings with postgraduate and undergraduate students of computer science who were native isiXhosa speaking participants. These groups were considered computer experts and also knowledgeable in the local language of implementation [Yeo, 1996]. They consisted of 5 postgraduate students and 5 undergraduate students. The sessions were held in the master students' computer laboratory. These sessions were necessary to collect further data, verify the data collected from the community and make corrections where necessary.

We also conducted voice recording sessions with a professional female participant who speaks the isiXhosa language fluently. The participant was also a computer literate person. A voice recorder was used for this purpose. The voice prompts that were recorded were some of the language clicks that were used on the interface as metaphors and commands.

A culturally adapted prototype voice user interface was designed using the data collected from the community and the computer experts. The prototype voice user interface consists of two varieties. A voice input user interface and a DTMF input user interface. The voice user interface varieties are the shop-owner voice interfaces on the mobile commerce application. The Dwesa community shop-owners will interact with the mobile commerce application through any of the interface to perform tasks such as update of product in a shop, creating an online shop, adding product details, etc. After the design, the two voice interface flavours were presented to the community members for user evaluation.

3.1. Pilot Testing

After the design of the prototype voice user interface, a pilot test was conducted with members of the community. The pilot test was meant to introduce the users early to the application and also serve as a small training procedure for system evaluation to users who have no prior knowledge about the process of evaluation. The test was a formative evaluation process and also to assess users' capability in learning and understanding the voice user interface. Participants were asked to interact with menu items, navigations and check if the interface feedbacks were acceptable. Few participants at this first contact were able to understand the process. A short training was given to the participants. The few of them who understood the process interacted with the interfaces. We observed and noted where they had problems like the feed back from the interface not being the right word based on their language dialect. We used the information collected and made adjustment to the interfaces.

3.2. Design Evaluation

User evaluation of a software design is very important because it helps to ascertain if the design meets user requirements and the objective of the design. Evaluation with the actual users have always produced good outcome [Patel et al. 2009; Sherwani et al. 2007]. This kind of evaluation involving the intended users is common and very effective because it helps to uncover errors that the designer or expert may never notice or in most cases overlook such errors.

The localized voice user interface varieties (DTMF input and voice input) were evaluated by participants drawn from the user groups identified in the community. The experiment was conducted to examine user understanding and how much the users are able to learn the interfaces and improve the interactions. We also try to find out tasks completion rate among the different user groups that participated. The focus groups include 12 school learners, 8 crafters, and 6 clerical staff. Selected tasks were performed by each participant in each group. The first task asks the users to update any chosen attribute of a given product (e.g. the price or weight of a product). The second task asks the users to create an online shop using the two varieties of the voice user interface. Tasks completion time, tasks successes and failures as well as difficulties encountered were recorded. A short interview was also conducted at the end of each group session. Participants were also given a post test Likert-scale questionnaire to fill. The Likert-scale questionnaire was positively rated with a 1-5 point scale.

Other participants who were functionally literate members of the community also took part in the evaluation process. These participants are educators (n = 12) in the community schools and they evaluated the interfaces through single user evaluation process. Interaction with these members of the community was very easy because of their level of education. Similar tasks used during focus groups were also performed. During each evaluation session, we also observed and recorded each user performance. Tasks completion time, failures and successes, problems encountered by the participants were recorded on field note for analysis. Post test interviews were also conducted with each participant as well as the participants completing a post evaluation questionnaire.

4. RESULTS AND DATA ANALYSIS

During the study, quantitative and qualitative data was collected and analyzed. In some instances, more than one method of data collection was applied. Data was collected from different user groups in the community in order to determine the similarities and differences in the outcome across the groups. Data collected for voice user interface adaptation passed through a validation process performed by computer experts who share the same culture with the Dwesa people.

The demographic data of the respondents in Dwesa community is presented in figure 1. The result shows the different occupation of the people in the community. The result identified occupations such as farming, crafting, teaching (educators), school learners, and others (e.g. clerical staff, traders, clerks, unemployed, etc.). 19.7% of the respondents were farmers, who are mostly semi or illiterate population. 15.83% were crafters, consisting of respondents who did not attend school more than the grade 12 level of education. The school learners (26.67%) are young people who are in school either in the grade 11 or 12. The educators are the functional literate members of the community. This group made up 21.67% of the respondents. The educators have among them those who are computer literate and others who are not. In the data collection, this group made significant contribution on providing information on user interface elements translation into the local language (e.g. text, and metaphor meaning).

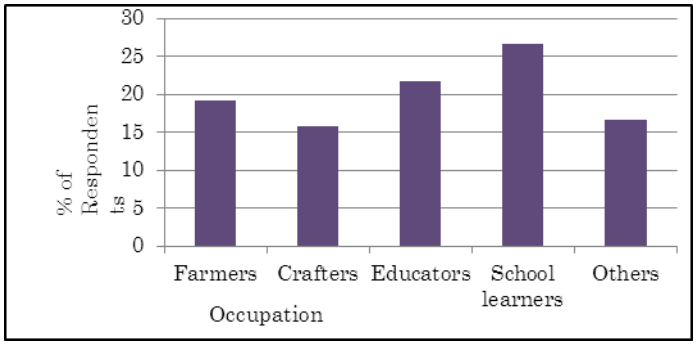

Fig. 1. Occupation of respondents

The survey conducted also provided data on the use of mobile services in the community. The services include short message service (SMS), multimedia service (MMS), voice mail and mobile internet access. Due to the low level of economic activities in the community, low brand mobile phones are common. Also, the cost of some mobile services prevents the people from using such services. Figure 2 shows the result on the level of usage of the common mobile services apart from voice calls. 17.01% of the respondents do use the Internet on their mobile phones. The functional literate members of the community may account for this percentage. 40.02% of the respondentds do send and receive

SMS, while 8.05% only receive SMS. 24.04% also use the MMS service, while 8.05% can only receive MMS but they do not send MMS to other people. 2.65% say the use the voice mail service provided by the mobile operators in their phones. The results indicated a low level of usage of these services. Listening to voice mail is a form of voice interface that can have a positive impact on users performance on the voice user interface. The results shows that these groups of users have little experience in terms of using voice mail service or listening to audio feedback from the mobile phone service. The results indicate that these user groups have little or no experience with voice user interface and may not be familiar with it.

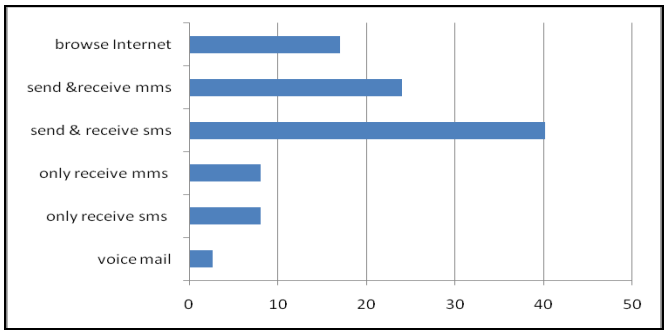

Fig. 2. Mobile phone technologies use in Dwesa community

The design of user interfaces for low-literacy users in rural communities is a big challenge. This study outcome illustrates an attempt to design usable interface by adapting the user interface to the socio-cultural experience of the target users. During the user evaluations, the participants performed two separate tasks using each of the two interface varieties. The results are presented next.

4.1. Tasks Completion Rate

The evaluation of the voice user interface varieties (DTMF input and voice input) was conducted to determine among other things the tasks completion rate among the various user groups in Dwesa community. The first selected task asked the participants to update any chosen attribute of a given product (e.g. the price or weight). The subtasks for this task include navigating to the shop-owner/admin menu, select the update menu item, indicated the online shop to enter, identify the product to update, select the field to update and enter the new data, and then submit the update. The second task for the evaluation required the users to create a new online shop. It includes the following subtasks: navigate to the admin menu, select create new online shop menu item, input data into each of the fields for the new shop one after the other, and then submit the information.

There was significant tasks completion successes recorded by the different user groups as they interacted with the two voice interface varieties. Figure 3 presents the task (task 1) completion levels for each of the user groups for the DTMF input and voice input interfaces. During the performance of the first task, school learners and clerical staff demonstrated higher performances when they used the DTMF input interface. A total of 75% of the school learners were able to perform and complete task 1, while 58.33% were able to complete the same task when they interacted with the voice input user interface. Also, 46.78% of the crafters were able to carry out the task successfully when they used the DTMF input interface, and 48.23% of the crafters performed and completed the task with the voice input interface. We observed that the short training impacted positively on the participants. The training contributed to the level of tasks completion successes the users attained. Participants who received prior training performed better than those who had no prior training. The level of task completion successes recorded by the low-literacy participants (e.g crafters, school learners) was significant. The educators and clerical staff also performed very well with the two interface varieties. The DTMF input recorded a higher task success rate. We observed that the participants were more familiar with DTMF input. This is an indication that the interface meets user experience and preference. Other factors also contributed to the level of performance with the voice input interface. These are noise interference and the nature of entering voice input into the system.

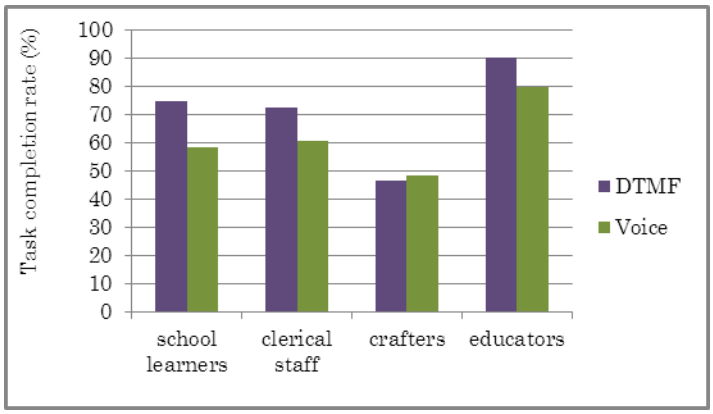

Fig.3. Task completion rate among different user groups (task1)

There was also significant task performance success among the different user groups with the second task. The results (see Figure 4) indicate a higher task completion with the DTMF input than the voice input interface for the different user groups. The second task performed by the participants was more tedious in terms of the number of subtasks than the first task. It was necessary to begin the evaluation with an easier task in order to allow the participants get familiar with the interfaces and then gradually raise their confidence and performance. It was observed that during the second tasks, the participants had become familiar with the operations of the interfaces and their responses were improved. Among the school learners, 66.6% task completion rate was achieved while using the DTMF input interface as against 53% success recorded with the voice input interface. The crafters also had 58.33% success with the DTMF input and 41.67% task completion success with the voice input interface. During the study, it was observed that problems associated with poor or inaudible inputs affected the users' performance with the voice input interface. The participants were excited interacting with the system using their local language. Getting system voice responses in their language made it look natural to them. This form of interaction with the user interface conforms to their social form of interactions.

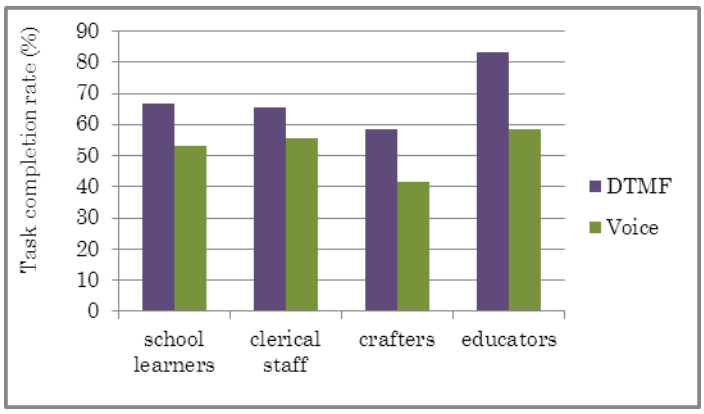

Fig.4. Task Completion Rate among the Different User Groups (task2)

We found that the natural accent of the people lead to poor pronunciation of words and delayed task completion especially among the low-literacy groups. For instance, while some users pronounced "wani" for "one", others used "uwani" or "u one". This will always cause recognition errors and also lead to frustration. The right pronunciation of such a word will depend on which region or dialect of isiXhosa is being considered and then which dialect or region the user comes from. Training of the users before they interact with the system is also very important. The training of low-literacy users helps them to overcome some of these limitations and challenges [Sherwani et al. 2009]. The preference for any interaction medium for any voice user interface depends on the user groups, their capabilities and cultural background, literacy level and the type of application. The study found that low-literacy users in Dwesa performed better with higher tasks completion successes when they used DTMF input than with voice input. Although speech input have the capability of recording high successes, even more than DTMF input. This depends

largely on the type of application area and the users involved. Similarly, it also depends on how fluent the users are when pronouncing words as inputs into the system for effective and fast recognition.

The localized voice user interfaces (DTMF input and voice input) design for low-literacy users in Dwesa community meets their capabilities and interaction experiences. Cultural adaptation of the interfaces enabled the interfaces to conform to the cultural expectations of the users. Voice interface for low-literacy rural users presents a lot of challenges during design. Finding suitable metaphors and being able to translate voice prompts into the natural language of the users are some of the challenges. When a language is spoken with clicks instead of plain words, it makes it more difficult. We suggest a lot of effort should be put in understanding the metaphors within the language of implementation and how to apply the metaphors in order to overcome some of the challenges the users may face during interactions.

4.2. Tasks Performance Time and Errors

Table I shows the average time to complete the first task (update an attribute of a product). Participants using DTMF input interface recorded an average time of 38.07 seconds and a standard deviation of 7.38. With the voice input interface, participants recorded an average time of 49.75 seconds and a standard deviation of 11.44. The results show a faster completion time and smaller variations with DTMF vs. voice input among all the participants. It also indicates a better precision in task performance output for DTMF. The results also indicate that the DTMF input will provide the users faster access to information, better performance and it is easier for the participants to use. Voice inputs also proved to be successfully used from the results. It will be able to provide an effective means for accessing information by the community.

Table I. Performance Result for First Task (time)

Performance result for first task (time) (n = 50)		
Method of interaction	Average time to complete task (seconds)	Standard deviation
DTMF	38.07	7.38
Voice	49.75	11.44

A t-T statistical analysis was performed of the data to compare the two voice interface varieties. the t-Test statistical analysis (paired t-Test) was peroformed on the completion times recorded by all the participants during tasks performance (first task). The t-Test results (T-Value = -3.83, P-Value = 0.001, two-tail) show that the difference between the DTMF interactions and the voice interaction was found to be statistically significant ($p <= 0.001$). This implies that the participants showed a better task completion time on the average with the DTMF input interface.

Table II presents the average error rate results for the first task. During the DTMF interactions the participants recorded an average error rate of 1.25 (standard deviation of 0.43), and with voice interactions, the participants' error rate was 2.44 (standard deviation of 0.87). Less errors were committed with the DTMF input. This shows that a higher precision was recorded for DTMF interactions among the participants.

Table II. Performance Error Rate for First Task

Performance error rate for first task (n = 50)		
Method of interaction	Average error rate	Standard deviation
DTMF	1.25	0.43
Voice	2.44	0.87

Participants encountered more errors when they used voice input when compared with the DTMF input. The overall number of errors was low. The t-Test statistical analyses performed on the errors encountered by each participant for the two input interfaces (T-Value = -4.84, P-Value = 0.000) shows that there is a significant statistical difference between the errors encountered using DTMF and voice interactions. The errors encountered were small for both varieties of interfaces. The interfaces meet their capabilities and they will be able to use any of the variety successfully. There was a higher level of accuracy with the DTMF input than the voice input.

Table III presents the average time to complete the second task (create an online shop). Participants recorded an average completion time of 26.25 seconds (standard deviation of 7.41) for the DTMF interactions. For the voice interactions, participants recorded an average time of 32.93 seconds (standard deviation of 7.56). The results show a close outcome in the performance time for both interface varieties. The variation also indicates that the participants completed the task with only a small time difference between them. There was significant improvement in task performance in the second task. The participants became more familiar with the voice input interface prompts and were able to respond faster. They were able to perform the given tasks with ease and a good level of success with the two interface varieties. It also shows that it was easier for the users to respond to the interface using DTMF input than the voice input. The DTMF matches the users' capabilities and preferences more than the voice input.

Table III. Performance Result for Second Task (time)

Performance result for second task (time) (n = 50)		
Method of interaction	Average time to comple the task (seconds)	Standard deviation
DTMF	26.25	7.41
Voice	32.93	7.56

T-Test statistical analyses (paired t-Test) were performed on the task performance time outcome recorded by all the participants for the interfaces. The results of the analyses (T-Value = -3.40, P-Value = 0.002) show that there is a significant statistical difference between the two interaction methods (DTMF vs. voice). Again the participants found DTMF input easier out of the two input interfaces.

Table IV displays the average error rate for the second task. Participants recorded an average error rate of 0.41 (standard deviation of 0.60) when they used DTMF interactions. For the voice interactions, participants recorded an average of 1.82 errors (standard deviation of 1.10). The results also show a general understanding of the interfaces and better performances among the participants.

Table IV. Performance Error Rate for Second Task

Performance error rate for second task (n = 50)		
Method of interaction	Average error rate	Standard deviation
DTMF	0.41	0.60
Voice	1.82	1.10

A t-Test statistical analysis was carried out on the number of errors encountered by all the participants as they performed the second tasks with both interfaces. The results of the analyses (T-Value = -6.20, P-Value = 0.000) show that there is a significant statistical difference between the error encountered while using DTMF and the error encountered using voice interactions. The level of accuracy was higher when the participants performed the task using the DTMF input interface. This could be as a result of familiarity or previous experience with key tone inputs or poor

recognition problems encountered. On the average, the localized interfaces meet the experiences and capabilities of the target users. This is because interface elements (menu, metaphors, feedbacks, etc.) reflect the culture and what is familiar to the target users.

Post Evaluation Interviews - We conducted interviews with the participants at the end of each session to find out their opinion about what they feel about the voice interfaces. Despite the difference in task completion time between the two interface flavours, the participants said they were satisfied with the interfaces. They were excited about the use of their local language on the interfaces. We asked the participants to identify with which of the interfaces they encountered problems when inputting data into the system - 58.28% of the respondents said voice, while 32.18% said DTMF. A total 70.23% feel that they were more comfortable using the DTMF interaction than the voice input. Those who preferred the DTMF input said they were conversant with keypad input based on their previous experience with phone keypad inputs.

The respondents said it was faster and easier for them to simply press the keys than saying the numbers one by one. Although, from our observation, among the school learners, when they encountered problems (e.g. if the system said "I did not hear you" and they needed the required word loudly) they felt they were being laughed at by their colleagues. A total of 29.77% felt that they were comfortable and had no problem using the voice input interface. This group consisted mostly of the educators who were functionally literate. Among those who prefer the use of voice, they said the only difficulty they encounter was that the system was repeating the input request even when they have uttered the word requested ("I said one and the system said I did not hear you so I shouted").

In the two voice interface varieties, we enabled barge-in during inputs so that the participants can reduce the time spent listening to the entire menu list before picking an option. That is, if they are already familiar with the menu or can pre-empt what is required. This was done to also reduce frustration and time wasting. We also observed that due to pronunciation problems, a few participants felt frustrated when they spoke into the system and it repeatedly responded "I did not hear you" and as such they could not complete their task or spent a longer time repeating what they had earlier said.

We observed that poor pronunciation was also responsible for the difficulty encountered (e.g. a participant saying "uwani" meaning one, "utuu" meaning two, or noise from the background). This situation made participants spend a longer time with voice input and was less successful compared to DTMF input. An instance where the user has to speak words one after the other also slowed the speed of interactions. As humans, we pronounce numbers continuously with very little gaps between pronunciations which is very different when the users of the voice input function speak numbers into the interface. It is very unnatural way of communication. Participants had to learn to say numbers one by one which we observed was not comfortable for everyone. Privacy was another factor we observed that made participants not to alter words audible enough into the application. A greater percentage of the participants prefer the DTMF dialed input/interactions and the results showed that DTMF dialed input recorded better and faster task completion time than voice inputs. The performance of any of the interaction modes for a voice user interface depends on the design, user study and the tasks performed during the study (Sherwani et al., 2009). The study also discovered that, despite the difficulty encountered during interactions, the participants were excited in their responses concerning the voice application.

Table V presents the results of the post test questionnaire. The purpose of the questionnaire was to assess the attitudes of the participants with regard to the ease of use and understanding of the voice user interfaces, and participants' perceptions. The results show that participants were satisfied with the interfaces (question 1, mean score = 4.10, standard deviation = 0.84). The low value in the different scores (variation) shows that the participants were consistent in their feelings about the interfaces. They also indicated that it was not difficult to use the interfaces (positive scores for questions 4 and 5).

The participants were also strongly positive that the isiXhosa language used in the system was easy to understand and provided adequate information to aid navigation on the interfaces (question 11 mean = 4.38, SD = 0.76). The voice prompts produced by the voice interfaces were not difficult for the participants to understand (question 7). Participants who prefer the voice input interface (question 8) scored a mean = 2.97, SD = 1.27, while participants who preferred the DTMF input interface scored a mean of 3.38 and SD = 1.32 (question 9). The results indicate that participants positively perceived both the DTMF input interface and the voice only input interface as easy to use and

simple for effective interactions. This shows that localized voice interfaces met local users' capabilities and preferences and made community participants to understand and use the interfaces with ease. Their satisfaction with the interfaces as indicated in the rating also shows that the interfaces will be useful within this community.

Table V. Shop-Owner VUI Evaluation Data and Analysis

Shop-owner VUI evaluation data and analysis (Likert-scale questions: 5 strongly agree – 1 strongly disagree) (n = 42)		
Question	Mean score	Standard deviation
1. I am satisfied with the interfaces	4.10	0.84
2. The interface with voice input is more easy to use	3.31	1.12
3. The interface with key input (DTMF) is more easy to use	3.69	1.18
4. The interface using voice input is difficult	2.03	1.00
5. The interface using key input (DTMF) is difficult to use	1.97	0.93
6. I did hear and understand whenever the system speaks	3.93	0.87
7. It was difficult to understand when the system speak	1.86	0.78
8. I prefer to used voice input when using the system	2.87	1.27
9. I prefer to use keypad (DTMF) input when using the system	3.38	1.32
10. The interfaces do not require special skills to use	4.21	0.80
11. The language (Isixhosa) spoken by the system is good and adequate	4.38	0.76

5. CONCLUSION

During user interface and interaction design and testing, culture and literacy plays a very vital role on the level of participation of the target users. In the study, we observed that the participants were more active when they are in groups. They also communicated fluently when they were expressing their opinion in their local language within the group. This was not the case when we tried to meet them on individual basis during the study. User interface design and testing methods that are not familiar to the socio-cultural background of the users affects the users' level of involvement [Smith et al. 2004]. Focus group method was mostly used due to the fact that the method is not different from the way the people interact within their socio-cultural environment. In a group interaction participants were able to motivate and help others in the group to become active.

Speech communication is an easy, natural and intuitive form of communication. Speech interaction is a common form of interaction among humans. The application of speech technology for human-to-machine interaction provides a natural sense of interaction to the users if properly adapted to suit the user environment. One prominent advantage of speech interactions is its ability to accommodate a variety of users irrespective of educational level and technological literacy. Speech user interfaces provide a medium for users to interact with the application using natural language.

Speech-based interactions reduce screen presentation, and overcome the constraints imposed by a small screen and cumbersome keypad on mobile devices. Speech interactions are simple to use and conform to users' common form of communication. The localized voice user interface varieties developed during this study has demonstrated that low-literacy users could interact with the interfaces. The users did experience difficulties when interacting with the interfaces, most of the problems were due to pronunciation and recognition problems. These problems need to be further looked at with the hope of resolving them. The study findings suggested that low-literacy users could use the voice interfaces with ease when performing tasks. The community users found the DTMF-input interface easier to

interact with. A higher percentage of low-literacy users in Dwesa prefer to use the DTMF-input instead of the voice-input interface. This may be due to pronunciation and recognition problems mentioned earlier. The findings from the study suggest that low-literacy users have a positive perception about the voice user interface and interactions. This indicates that the interface may have been within the limit of their capabilities experience and natural preference.

This study provides evidence that voice user interface and interactions can support effective use of ICT services in rural communities. The study found that, despite the recognition problems encountered, community users still feel that the voice user interface is better due to the ease with which they interacted with it. This is evidenced by the high task completion rate when they used the voice interface compared to when the used the graphic interface.

Voice user interfaces have the advantage that, if implemented in the local language of the target users, both the semi-literates and illiterates will make use of the service with minimal training.

The voice user interface designed during this study presented interaction elements (e.g. short menu hierarchy, cultural metaphors, feedbacks, etc.) that meet the cultural preferences, experiences and interaction expectation of the community users. We were deeply concerned with using metaphors and language translations within the cultural context of the Dwesa language during the voice interface design. The culture of the users influence the meaning the users attach to a metaphor. A metaphor helps the user to use familiar experience to create a mental model the information or object for which the metaphor is trying to illustrate. We believe that the cultural metaphors (e.g. welcome to Ngwane art and craft shop, Ok in the local language) provided on the user interface have played a significant role in making the participants acquaint themselves with the voice interface. User interface adaptation to users' culture and preferences will improve information accessibility among different user groups in different regions. The voice user interface varieties presented in this study can be adapted to different cultures to provide similar services for marginalized communities in developing countries. We intend to improve the voice user interface to make it self-learning and adaptive to user interactions.

REFERENCES

EDIM, A. E. AND MUYINGI, H. N. 2010a. A Cultural User Interface for a Rural Community in South Africa. In *Proceedings of SATNAC-2010*, Stellenbosch, South Africa, 231-236.

EDIM, A.E. AND MUYINGI, H. N. 2010b. A Mobile Commerce Application for Rural Economy Development: A Case Study for Dwesa, *Proceedings Of SAICSIT-2010*, ACM press, Bela Bela, South Africa, 58-66.

GHOSH, K., PARIKH, T. AND CHAVAN, A. 2003. Design Considerations for a Financial Management System for Rural, Semi-literate Users. In *ACM Conference on Computer-Human Interaction-2003.*

HEEKS, R. 2008. ICT4D 2.0: The Next Phase of Applying ICT for International Development. *Computer*, 41, 6, 26-33.

HUANG, K. H. AND DENG, Y. S. 2008. Social Interaction Design in Cultural Context: A Case Study of a Traditional Social Activity, *International Journal of Design*, 2, 2, 81-96.

ID21 INSIGHTS. 2007. Mobile phones and development, retrieved: 12 March, 2009 from www.id21.org/insights/insights69/ insights69.pdf. 2007.

KAM, M., RAMACHANDRAN, D., DEVANATHAN, V., TEWARI, A., AND CANNY, J. 2007. Localized iterative design for language learning in underdeveloped regions: The PACE framework. In *Proc. CHI 2007*, ACM Press, NY, USA, 31-40.

LALJI, Z. AND GOOD, J. 2008. Designing new technologies for illiterate populations: A study in mobile phone interface design, *Interacting with Computers*, 20, 574 – 586.

LO, T. AND MENG, H. M. 2002. WAP-Speech: Deriving Synergy between WAP and the Spoken Dialog Interface. In *Proc. of the Eleventh International World Wide Web Conference*. Honolulu.

MEDHI, I., SAGAR, A. AND TOYAMA, K. 2006. Text-Free User Interfaces for Illiterate and Semi-Literate Users. In Proc. International Conference on Information and Communications Technologies and Development, 72-82.

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 *Proceedings of CHI 2009*, ACM Press, Massachusetts, 1741-1750.

PADE, C., PALMER, R., KAVHAI, M. AND GUMBO, S. 2009. Siyakhula Living Lab Baseline study Report. (April, 2009).

PATEL, N., AGARWAL, S., RAJPUT, N., NANAVATI, A., DAVE, P. AND PARIKH, T. 2009. A Comparative Study of Speech vs. Dialed Input Voice Interfaces in Rural India. In *ACM CHI-2009*, Massachusetts, USA, 51-54.

PATEL, N., AGARWAL, S., RAJPUT, N., KUMAR, A., NANAVATI, A., DAVE, P. AND PARIKH, T. S. 2008. Experiences Designing a Voice Interface for Rural India. In *proceedings of the IEEE Workshop on Spoken Language Technology for Development (SLT)*, 21-25.

PLAUCHE, M., NALLASAMY, U., PAL, J., WOOTERS, C. AND RAMACHANDRAN, D. 2006. Speech Recognition for illiterate Access to Information and Technology. In *Proc. International Conference on Information and Communications Technologies and Development*.

PARIKH, T. JAVID, P. SASIKUMAR, K., GHOSH, K., AND TOYAMA, K. 2006. Mobile Phones and Paper Documents: Evaluating a New Approach for Capturing Microfinance Data in Rural India, *ACM Conference on Computer-Human Interaction (CHI)*, Montreal, Canada.

SATO, K. AND CHEN, K. 2008. Cultural Aspects of Interaction Design, *International Journal of Design*, 2, 2.

SHEN, S. WOOLLEY, M. AND PRIOR, S. 2006. Towards Culture-centered Design, *Interacting with Computers*, 18, 820-852.

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, *Proceedings IEEE/ACM 3rd Int. Conf. on ICTD*, Doha, Qatar, 447-457.

SHERWANI, J., ALI, N. MIRZA, S., FATMA, A., MEMON, Y., KARIM, M., TONGIA, R. AND ROSENFELD, R., 2007. HealthLine: Speech -based Access to Health Information by Low-literate Users. In *Proceedings International conference of Information & Communication Technologies and Development*, Bangalore, India.

SMITH, A., DUNCKLEY, L., FRENCH, T., MINOCHA, S., AND CHANG, Y. 2004. A Process Model for Developing Usable Cross-cultural Web Sites. *Interacting with Computers*, 16, 63-91.

TSAI, M. 2006. VoiceXML dialog system of the multimodal IP-telephony – The application for voice ordering service, *Journal of Expert systems with application*, 31, 684-696.

YEO, A. 1996. World-Wide CHI: Cultural user interface, a Silver lining in Cultural Diversity, *SIGCHI*, 28, 4-7.

ZHOU, L. 2007. Natural language interface for information management on mobile devices. *Behaviour and Information Technology*, 26, 3, 197–207.