DESIGNING INTERACTIVE VOICE RESPONSE (IVR) INTERFACES: LOCALISATION FOR LOW LITERACY USERS

Aditi Sharma Grover[1], Osamuyimen Stewart[2], David Lubensky[2]
Meraka Institute, CSIR[1] and IBM Research[2]
Building 43B, Meiring Naude Road, Brummeria, Pretoria, 0184, South Africa[1]
IBM T.J.Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, NY 10598, USA[2]
asharma1@csir.co.za[1], ostewart@us.ibm.com[2], davidlu@us.ibm.com[2]

ABSTRACT
In this paper, we address the issue of how to design IVR interfaces for the developing world. Against the backdrop of the following idiosyncratic observations including, the majority of users being either semi-literate or non-literate, and the impact of a different set of social-cultural, linguistic, and domestic challenges, amongst others, we advocate the enculturation of IVR interfaces different from the developed world. This requires the tailoring of functionalities and interactive modalities to the cultural values and context of use. Thus, we propose a dialog (user interface) design model consisting of three components: Get input, Error-recovery, and Play results (output). These are shown to be critical for implementing usable and culturally-suitable IVR interfaces for low-literacy user populations.

KEY WORDS
Interactive Voice Response (IVR), speech and natural language interfaces, user interface development, low literacy, developing world, localisation.

1. Introduction

In general, Interactive Voice Response (IVR) interfaces, ipso facto, are fraught with typical design issues including (a) recognition errors, (b) auditory and cognitive challenges, (c) user control of the interaction, (d) ambiguities, etc., all of which present usability challenges that greatly impact voice automated services [1, 2]. Despite these challenges, examples abound of several successfully deployed IVR systems (speech and touch-tone input) in the developed world - health-management systems [3] and commercial voice portal systems, such as Google's Goog-411 service [4] and the TellMe service [5].

In the developing world (note that many regions e.g. India, South Africa, Brazil, may have hubs of technological advancement with small sub-sets of highly literate populations, however our reference to the 'developing world' in this paper refers to larger populations in these regions who are low literate and technologically in-experienced) IVR interfaces have also been argued to be the logical and natural means for implementing automated services or solutions [6, 7, 8, 9, 10, 11]. This proposal is supported by several reasons and two of which will suffice; the first is the pervasive use and penetration of mobile phones in the developing world [12], and the second is the observation that use of IVR-based interfaces requires minimal skills compared to the usability of other computer-based interfaces. These are important considerations since the majority (98%) of the world's non-literate population lives in the developing world [13]. As a result, telephone-based access to information can afford such populations access to information and the opportunity to participate in the digital age. In addition, telephone-based IVR services in contrast to PC-based solutions have lesser infrastructure costs and no maintenance requirements from the user since the system can easily be centrally maintained at more geographically convenient locations [8, 9, 14].

To date, there have been a few pockets of research activities to explore the feasibility of IVR interfaces for low literacy users [15, 16, 17, 7, 8, 18]. There have been a number of preliminary findings and practical lessons learnt but much still needs to be done in terms of fully exploiting the possibilities of IVR interfaces in a context very different from the developed world [19, 20]. In light of the foregoing, the question addressed in this paper is how we should design IVR interfaces for low literacy users in the developing world. We focus on the range of currently deployed systems (primarily in the healthcare and information retrieval domains) and propose a dialog design model shown in Fig. 1 that should be considered valuable for designing optimal IVR interfaces for low literacy users in the developing world.

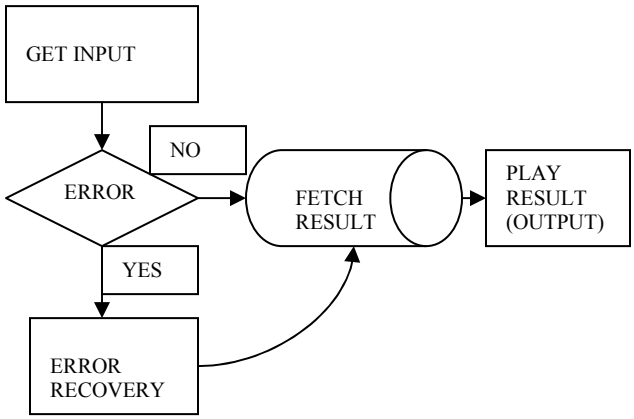

Fig. 1 Dialog design model for IVR interfaces

GET INPUT: this offers the principled basis for user-system interaction. We propose natural language interfaces should be avoided and that semantic options (what the user can do or request) should be presented to the user in the form of numerical choices that they can select(press) on their mobile phones, i.e., Dual Tone Multi Frequency (DTMF) or touchtone.

ERROR RECOVERY: this component validates the accuracy of what the system has understood as the user's request. When there is an error, it guides users to correctly state their request. To recover from errors, we propose a multimodal process combining audio with images or icons with appropriate use of explicit vs. implicit confirmation prompts.

OUTPUT: this offers the principled basis for presenting the requested information back to the user. We propose that this interaction should be done in the local language with attention paid to user interface metaphors, persona and prompt translation.

This proposed dialog design model is predicated on a crucial underlying assumption that centers on the need for enculturation of IVR interfaces in the developing world—due to the vastly different social-cultural and domestic environment. The enculturation of IVR interfaces suggests that such interfaces should be balanced against the existing social-cultural and domestic environment of the users, which may influence the usability of the applications in terms of its domain and nature of task. This may significantly affect the usage and trustworthiness of these systems amongst low literacy users.

For example, in a recent study done on kiosk usage amongst farmers in India [21], it was found that the provision of information was itself not enough to promote usage of an information source (kiosk), but rather the perceived quality of the information determined its usage. Although the farmers had some use for the service, they did not visit the kiosk because they did not consider certain kinds of information at the kiosk 'trustworthy' enough. The study found that users did not regard trustworthiness as only the "correctness" of the information being presented but also the "information source": if the information required by the farmers involved high stakes (urgent, confidential or crucial) then trustworthiness was a crucial factor in determining service usage and in the absence of regular interactions with kiosks the farmers reverted to more traditional channels even if they were more expensive (time and money). Similarly, it was observed during the heuristic evaluation of a demo telephony application for tracking Unemployment Insurance Fund (UIF) payments in South Africa, that some evaluators felt "even though the system cannot be used to obtain money directly, that it was not sufficiently 'secure' or 'private'"[22, 23]. These two examples underscore the fact that the perception of risk with technology adoption in the developing world may be tied to the culture surrounding the nature of request and not so much the perceived value of the interface.

From a design perspective, enculturation of IVR interfaces implies tailoring the functionalities and interactive modalities to the cultural values of the user population. For example, the fact that IVR interfaces with speech are popular in the developed world does not guarantee its adoption and usability in the developing world. Given the typical design problems with standard IVR interfaces, and in addition to the peculiarities of developing regions such as (a) literacy (text/writing versus speech/orality in local languages), (b) phatic communion (domestic and social-cultural aspects of rural life), (c) usability (user interface suitability and familiarity with technology,) etc., one can not mindlessly impose IVR design practices on the developing world without significant customization to suit the needs and behavior of users.

In the sections that follow, we will describe each element of the dialog design model mentioned above with motivation and various examples from our own experiences and previous work.

2. Get Input

In this section we will address the question of what could be the preferred input mode in IVR interfaces used by semi and non-literate users in the developing world.

There has been much debate around which modality is preferred by users for interaction with speech interfaces: DTMF or speech input? [2, 24, 25, 26]. The general conclusion from previous work comparing speech to DTMF in developed world contexts favor the use of natural language (NL), which uses open-ended or free-form dialog, or at least some form of speech recognition, over DTMF. We note that all of these comparison studies [2, 24, 25, 26] targeted business-typical applications, like voicemail, call routing and banking, which are more applicable to users on the higher end of the literacy spectrum in the developed world.

Another noteworthy issue from a technical feasibility perspective is that the above-mentioned studies are based on IVR systems in the common world languages (English and European languages) and they assume that NL technologies are easily available in all languages. This is, however, not the case for under-resourced languages in developing countries where linguistic knowledge and resources, technical expertise, computing infrastructure and funding to build NL systems are limited [27]. Thus, the conclusion on the use of NL in IVR systems is questionable in the context of developing regions of the world where the users are naturally unfamiliar with such systems and the technology. With an unconstrained interface like NL (e.g. "How may I help you today?" type

of interfaces), which may not have optimal recognition/understanding rates for new languages, the likelihood is that users may get confused and this may result in user frustration, leading to call abandonment [1, 28]. We conjecture that such initial bad experiences may result in the development of a fear/aversion towards the use of automated systems with this user population.

Therefore, we suggest that IVR interfaces for low literacy users in the developing world should exclude the use of NL. This leaves us with only one other intuitive design option with speech, i.e., directed dialog (where a speaker says a keyword or command). In contrast to DTMF, it has been suggested that speech recognition input may be more suited to low literacy user groups [16, 22]. Some reasons offered include the need for numerical literacy to operate a DTMF interface, and the strong oral cultures that exist in these groups. Here is an example of IVR interface with speech:

System: Welcome to the Nation's Rail System. For Fares, say "fares" or for the list of trains, say "schedule".
User: "Fares"

In contrast, IVR interfaces using DTMF to elicit user input would have the following structure:

System: Welcome to the Nation's Rail System. For Fares, press 1 or for the list of trains, press 2.
User: [user presses 1 on mobile phone keypad]

To date there have only been a few user studies conducted to experimentally evaluate the modality preference of low literacy users [15, 16, 17] and we will describe these below. The study by Sherwani et al [16] compared DTMF with speech input (directed dialog design) in the context of an information-line for community health workers in Pakistan (HealthLine). They report higher task success rates with speech input, although, subjectively, there was no significant difference in user preference for DTMF or speech input.

In the Avaaj Otalo study by Patel et al [17] in India, an IVR-based community forum which allows farmers to obtain agricultural information over the phone was tested for DTMF and speech input. Their results indicate that task success was significantly higher for DTMF compared to speech input (directed dialog). The subjective 'user perception of difficulty in task' was also measured: DTMF users found it much easier to provide input to the system as opposed to speech-input users.

Lastly, a study conducted by Sharma et al [15] compared two identical systems (DTMF and speech) for user preference and task completion, in the context of OpenPhone – a health information IVR service for caregivers of HIV-positive children in Botswana. The results indicate user preference for DTMF (59%) over speech input (19%) (22% indicated no preference), although both systems were comparable in performance. This was based on objective metrics such as task completion rates, average response time and the number of user turns taken per call. Users in the OpenPhone study cited reasons such as DTMF being faster, quicker to use and more confidential. The study did not investigate which of these (speed or confidentiality) carries more weight.

On closer examination of these three studies, we observe that the results of HealthLine seemingly disagree with those of Avaaj Otalo and OpenPhone. However, we note a significant consideration regarding user training or familiarity with the system. In the case of HealthLine, Sherwani et al [16] report that their users had been exposed to the system earlier and there was significant time spent training the users before the study, but in OpenPhone and Avaaj Otalo, this was not the case. From a practical perspective, it would appear that the latter studies might be more representative since, in many instances, user training on a large scale may not be a feasible option in developing world regions due to geographical and resource constraints. In addition, another important observation is also underscored in the results of [15, 17] which is the observation that numerical literacy is not an overwhelming argument against using DTMF in the developing world.

Furthermore, in comparing these studies, one must acknowledge that the specifics of the interface (e.g. type of prompting), the cultural context, and the domain of use (health, agricultural, finance) need to be taken into account when choosing the modality type. A similar observation has been made by Barnard et al in [19] where the interaction between 'user ability' and 'application complexity' are used to describe the 'utility of speech' in the developing world. They also conjecture (based on developed world studies) that simple linear tasks (which can serve as a good starting point for introducing IVR services in the developing world) may be easily performed using non-speech modalities [19].

Below we describe two additional factors which support using DTMF for IVR interaction, especially in the healthcare domain.

2.1. Usage Culture

Mobile phone usage and ownership models in the developing world tend to be different from those of the developed world due to cultural and socio-economic disparities. The latter tends to have a single ownership model, but in the former we observe multiple users and a shared ownership model [29]. For example, in many parts of Africa, it is very common to see kiosks where people resell mobile minutes by allowing the population to come in and pay to use the phone service (charged per minute). This shared use has implications for the design of IVR

interfaces. For example, the systems with speech commands will have to deal with possible speech recognition issues from multiple users, noisy environments, and interruptions from the context of use. Quite often, on-lookers are keen to participate as experienced in the Tamil market user study in India [7, 8], and noise from the surroundings increase the chances for greater recognition errors with speech. In contrast, culturally-based usage favors DTMF since its recognition is not typically affected by a noisy background.

2.2. Privacy

The use of speech feedback or input in kiosks as well as the use of personal mobile phones in public places compromises user privacy since anybody within hearing distance can hear the user's input. This is particularly important because, for example, many of the current applications or services available via mobile phones in Africa focus on the health domain, e.g. AIDS/HIV outpatient care. As was noted in the study of OpenPhone (a HIV health information IVR service) in Botswana [15], a few participants reported privacy as a reason for their preference of DTMF over speech input – "*nobody can hear what I say* [health-related keywords] *and what I'm looking for* [the kind of health information]".

We observe that although the HealthLine service [16] provided health information, privacy was not mentioned as a primary concern by its users unlike OpenPhone. We note and reflect from [16] that the community health workers receive training in health topics and use health-related terms (potentially private) in their day-to-day work of helping community members. The care-givers in Botswana, on the other hand, do not receive much health-related training (something in the order of 1 hour to 1.5 hours of a basic lecture) and only deal with health-related topics in the context of caring for their own child, thus making them possibly more sensitive to privacy issues.

One may also argue that this is an extreme case since HealthLine and OpenPhone were health-related applications. However, privacy as a design consideration may not be a concept that is important to just the developed world, especially if/when financial services become available on mobile phones in the developing world. For example, in an assessment for kiosk usage by low literacy users in India [21], it was found that some farmers did not want to use an account balance checking service at the kiosk because they did not want the kiosk operator to know their account information (the operator typically had to help them since they didn't know how to use a computer). In another example, a mobile phone usage study was conducted in Uzbekistan by Wei et al [30]. They report that mobile phone use in public spaces was 'conservative', where people generally tend to minimise public use of their phones.

In conclusion, when designing IVR interfaces for low literacy users, designers should take into account the environment of use, the privacy implications for the particular task domains and the technical feasibility of the speech processing tasks required. In fact, from a technical feasibility and usability perspective a sub-optimal speech recognition system in an IVR service can lead to user frustration and call abandonment [28]. In a recent industry survey [31], it was found that 45% of consumers would like to use speech "as little as possible" in contrast to 9% of vendors; this highlights the fact that while researchers and the speech industry may be more focused on spoken interaction, consumers are *"more concerned with limitations that cause [system] fragility and unexpected behaviour"* [19]. In this regard, on the issue of whether to use DTMF or speech, we have observed that natural language understanding (NLU) is not suitable, and that DTMF may be better suited for healthcare and possibly for financial services.

3. Error Recovery

Given the challenges of temporality and ambiguity of speech (audio prompting) [1, 28], whenever there is an error in the use of the IVR, listening to the error re-prompt may exasperate the low-literate user. Also, the unfamiliarity of low literacy users with IVR systems in general may lead to more errors (at least in the initial stages). Plauche et al [7] found in their field experiments with a speech interface prototype that users often did not limit their speech to the allowed keywords leading to errors.

Thus, an important element of audio prompt design for low literacy users is providing robust error management. Without clear error recovery feedback for low literacy users, the learning curve for the user interface may become much harder and perhaps a somewhat frustrating experience. This may discourage low-literacy users in making use of the IVR service or at the very least, it could cause unnecessary apprehension towards the technology (or any ICT intervention). Thus, amongst the elements of error recovery, an important aspect is to provide targeted help at each point in the dialogue. All error feedback messages should be clear and context-dependent so that users have unambiguous instructions on exactly what they can say/press at any point.

In this regard for IVR services, implicit and explicit confirmation prompts can play an important role in error prevention and recovery. Implicit confirmations are more open-ended, where the system does not explicitly tell the user what to say. For example:
System: Which city are you travelling to?
User: Pretoria
System: What day of the week would you like to leave Victoria ?
User: No, Pretoria !

The use of implicit confirmations reduces the number of turns taken by the user, which makes the interaction less cumbersome but it places the onus on the user to respond to an error in an expected manner [32]. In the case of low literacy users (or any user unfamiliar with such interfaces), they may not be fully aware that they can say something or, in fact, what to say to correct the error, resulting in a high likelihood of call abandonment.

Explicit confirmations, on the other hand, compel the user to respond through direct instructions, but of course these result in a longer interaction and are less conversationally natural. For example:
System: Which city are you travelling to?
User: Pretoria
System: Did you say Victoria ? Please say 'Yes' or 'No'.
User: No !

In general, some useful guidelines for determining the use of implicit or explicit confirmations have been proposed [33]: (a) for actions that can be easily undone and are non-critical, use implicit confirmations; (b) for critical actions (e.g. a bank transaction) use explicit prompts. However, in the context of the developing world, the IVR designer will have to maintain a careful balance between the use of implicit and explicit confirmations. We conjecture that explicit confirmations may hold a higher weight in this compromise; however this research question remains yet to be empirically answered for low literacy user groups.

An interesting example of the use of implicit confirmations with low literacy users was reported in [9], where implicit confirmation of users' input resulted in a "Yes" (from the users) causing the system to misunderstand this unexpected utterance. Examples like these further motivate the need for focused investigations to determine the application (and effectiveness) of implicit and explicit confirmation prompts.

We further propose that the process of recovery from errors should go beyond mere audio prompting and use multimodality, i.e., combine audio with icons or images, to offer contextually-relevant help. This proposal is supported by previous studies that have argued that semi-literate users needed to be ***shown*** the proposed computer intervention in the context of how it could be useful to help them solve the problem, rather than simply ***telling*** them [9, 10, 11, 16, 34].

This general strategy of providing context to assist low literacy users can be applied to IVR interfaces such that graphical content (icon or image) is combined with audio. As a result, instead of just providing speech instructions to the user, they would see a set of icons or images and clicking or selecting a particular one triggers speech to guide them to fix the error in the relevant context. Moreover, some studies have shown that the inclusion of speech with graphical content (or vice versa) increases the usability of interfaces designed for low literacy users.

In particular, it has been observed that providing voice feedback (audio message playback when a user hovers over a graphical icon) for a text-free graphical user interface was crucial to the success of the user interface (UI) [34]. Without voice feedback, users struggled despite significant prompting. The importance of voice feedback has also been noted by Parikh et al. [34] who found that providing audio feedback enforces a structured 'dialogue' like style, making the interaction proceed like a conversation between the user and the device. This 'dialogue' like conversation could also help in shaping the speed of the user interaction in that the user listens to the audio and then responds accordingly.

In conclusion, we postulate that error recovery in IVR interfaces for the developing world should be based on clear instructions such as explicit confirmations where needed. In cases where it is relevant, the application should use a multimodal strategy which combines graphical content on the user's mobile phone with audio to offer context-sensitive information. This will help or guide the user to easily and gracefully recover from an error.

4. Play Results (Output)

The final component of our proposed dialog design model deals with the way that results or information are presented to low literacy users. In our discussion of error recovery and multimodality, we adopted a strategy which combines graphical content (icons or images) with speech or voice feedback. We would like to point out that the crucial success element of voice feedback in the examples from previous literature [8, 34, 35] and our experiences in designing OpenPhone in Botswana [15], is the fact that it was done in the local vernacular. This did not only make the interface on the mobile phone easier to use but also generated a lot of interest and excitement amongst participants in the user study [7, 8].

Thus, it is important that the voice output in IVR interfaces in the developing world be in the local language. Having voice feedback in the local language could also play an important role in improving usability of the interface by making it faster and easier for users to understand the system. For instance, it has been observed based on use of a PDA-based health intervention in Tanzania that reading English, even for indigenous literate users (who were supposed to be fluent in English), was a rather slow process [36]. They suggest that text should rather be translated into the local language Swahili. We extrapolate that the ability to interact with the IVR interface in the local language may mitigate the literacy effects in these populations, while ensuring that mobile devices offering automated services are consistent with the domestic and social-cultural life of these groups [34, 36, 37]. A further benefit of using local language is that users will be more likely to trust such systems; a vital

element in the adoption curve for new technologies such as IVR-based interfaces.

Since IVR services rely primarily on the audio medium to provide results to the users, prompt design also plays a crucial role in ensuring successful user interaction. Well-designed prompts should take into account not only linguistic concepts (e.g. discourse structure) but also aspects such as user interface metaphors, persona and cultural dimensions. These aspects are an essential part of the prompt design process for developing regions.

For instance, user interface metaphors help the user in getting oriented with and learning a UI, and they facilitate the formation of the user's mental model of the UI [38] (e.g. a computer "desktop" is similar to a real desktop with 'files' and 'folders' or the use of "CC" in emails to indicate 'carbon copy'). User interface metaphors allow the user to draw on their existing knowledge to create a mental model of the unknown/new concept or object (interface) that the metaphor illustrates.

In terms of low literacy users, UI metaphors could play a crucial role in acquainting users with the concept of IVR systems, and they should thus be carefully considered at the onset of prompt design. For example, Sherwani et al [9] also suggest use of familiar metaphors such as "Ok, I've opened up the Pneumonia book, and I see 7 sections" for giving information to community health workers who refer to training manuals. In this scenario, the UI metaphor of a 'book and its sections' is the navigational metaphor rather than the use of 'menu's, sub-menus, and options' as the metaphor. A similar experience was reported in the design of a text-free graphical UI [34], where low-literacy users found it quicker to understand hypertext navigation when they were told to think of the application's pages as pages in a book.

For the developing world UI designers also need to take into account that a UI metaphor's interpretation is highly dependent on the user's culture or background. Barr et al [39] suggest designers should "*seek to understand the users' metaphoric world as deeply as possible to counter the problem of the deeply subjective nature of metaphor*". As an example, the "main menu" metaphor which signifies the starting point in audio interfaces would most likely be a foreign concept to a low literacy user who may have no familiarity with the concept of menus, sub-menus and options. In our experimental work on an Afrikaans IVR prototype, we found that "main menu" when translated refers to a "menu of food". Thus we used the metaphor of "home" derived from the popular "home page" metaphor which is a more easily understood concept in Afrikaans.

IVR design for under-resourced languages also introduces challenges on the dimensions of creating the system persona and translation of prompts into the local language. Persona can be defined as the perceived personality of the application and is collective of factors such as voice quality, gender, tone of voice and audio effects [1, 33]. Since persona reflects the perceived personality of an application, we believe that if designed carefully, it could be used to effectively take into account cultural dimensions that are important in communication. Thus in IVR design for the developing world, the designer needs to ensure that the persona of the local language IVR system is in line with cultural and contextual expectations of the intended audience.

In our experiences with OpenPhone, where health information was provided through an IVR system to caregivers of HIV positive children, the persona modeled was that of 'a caring nurse who was willing to patiently educate the caller on health issues related to HIV'. The aim of the persona was to represent the characteristics of the traditional channel through which caregivers obtain health information i.e. a nurse at a pediatric HIV institute.

Feedback from the pilot user study indicated that users felt that the 'nurse' spoke very clearly and explained the health information in a patient manner. Some participants even went on to say that they would use the system to educate their communities, and one in particular excitedly mentioned that "*now I can tell them at home that the nurse* (IVR voice) *says the same thing* (referring to a HIV related topic) *that I'm telling them*". This illustrates that persona in IVR systems may affect the usage and trustworthiness of an IVR system amongst low literacy users as pointed out earlier in section 1. Thus the use of an appropriate persona could play a significant role in a user's perceptions about (and subsequent usage of) an IVR service (or ICT application).

The prompts and content of an IVR system will typically be translated from a language such as English to the local language. Thus, great care has to be taken in the translation phase to ensure that the intended meaning of the original prompt (English) is still preserved in the translated prompt (local language) and conveyed in the simplest and shortest way possible. Often, a concept described by a single word in English has no direct translation in another language. For instance, whereas a keyword in the English version of our OpenPhone system was "safe food", it became the phrase, "dijo tse di siameng" after translation, in order to adequately describe the concept.

At other times, certain words in a target language may need to be left in English, since users are more comfortable with the English version, e.g. in the design of Openphone we chose to use English numbers i.e. "..press one" was "..tobetsa 1" in Setswana (rather than "..tobetsa ntlha"). This choice was based on our observations in Botswana with a prototype where we used Setswana numbers and through subsequent consultations with Setswana speakers. They suggested that using English numbers would be more appropriate since most people

already use English numbers in their daily lives (e.g. paying transport fares, buying groceries, etc.). Whereas the use of Setswana numbers may have sounded out of place since people do not use them as frequently in their day-to-day activities.

Taking into cognisance the above mentioned issues, we realise that designing multilingual IVR services from an English perspective or focus is not a simple one-to-one translation of the entire design into the target language. Rather the designer needs to revisit their translated prompts and apply a second iteration of the guidelines on prompt design, user interface metaphors and persona mentioned above. This will help to ensure that the local language prompts convey the message and intention of the original prompts and match the socio-cultural context of the target users.

5. Conclusion

We have proposed a framework for designing IVR interfaces for low literacy users in the developing regions around the world. This model consists of three components: Get Input, which is suited to DTMF in the healthcare domain; Error-recovery, which should be multimodal combining graphical content (icons or images) with speech and appropriate use of explicit vs. implicit confirmation prompts.; Playing Results (output), which should be in the local language with attention paid to translation, persona and user interface metaphors in prompt design. These have been shown to be critical aspects that must be considered in implementing usable and culturally-suitable IVR interfaces for the developing world.

Keeping in mind the vast variety of services possible with IVRs and the high mobile phone penetration in developing world regions, IVR services present many opportunities for further work in this area. Nevertheless, care must be taken to avoid a myopic focus on the technology, but rather holistically considering the usability, and cultural and social implications of such IVR services. Research efforts in this domain [7, 8, 11, 12 15, 16, 17] show great promise and our work is an attempt to tap into the lessons learnt based on practical and theoretical perspectives (keeping in mind the differences with the developed world), and present a framework for IVR design in the developing world regions. We believe that IVR-type of services can play a revolutionary role in the context of providing developing world citizens with access to essential information and services and bridging the digital divide.

References

[1] D. J. Gardner-Bonneau, and H. Blanchard (ed), *Human factors and voice interactive systems* (Boston, USA: Kluwer, 2007)

[2] K. M. Lee & J. Lai, Speech Versus Touch: A Comparative Study of the Use of Speech and DTMF Keypad for Navigation, *International Journal of Human–Computer Interaction*, *19* (3), 2005, 343–360.

[3] J.P. Migneault, R. Farzanfar, J.A. Wright & R.H. Friedman, "How to write Health Dialog for a Talking Computer", *Journal of Biomedical Informatics, 39*(5), 2006, 468 – 481.

[4] Google Inc. Goog-411 [Online] Available at: http://www.google.com/goog411/ (Last accessed 3 September 2009)

[5] Tellme Networks Inc. Who uses TellMe? [Online]. Available at: http://www.tellme.com/about (Last accessed 3 September 2009).

[6] R. Tucker & K. Shalonova, "The Local Language Speech Technology Initiative", *Proc. of SCALLA Conference*, Nepal, 2004.

[7] M.P. Plauche, U. Nallasamy, J. Pal, C. Wooters & D. Ramachandran, Speech Recognition for Illiterate Access to Information and Technology, *Proc. IEEE/ACM International Conference on Information and Communication Technologies and Development (ICTD2006),* USA, 2006.

[8] M.P. Plauche & U Nallasamy, Speech Interfaces for Equitable Access to Information Technology, *Information Technologies and International Development*, *4*(1),2007, 69-86.

[9] J. Sherwani, N. Ali, S. Mirza, A. Fatma, Y. Memon, M. Karim, R. Tongia & R. Rosenfeld, HealthLine: Speech-based Access to Health Information by Low-literate Users, *Proc. IEEE/ACM International Conference on Information and Communication Technologies and Development (ICTD2007)*, India, 2007.

[10] A. Kumar, N. Rajput, S. Agarwal, D. Chakraborty & A.A. Nanavati, Organizing the Unorganized: Employing IT to Empower the Underprivileged, *Proc. 17th International World Wide Web Conference (WWW 2008)* China, 2008.

[11] A. Kumar, N. Rajput, D. Chakraborty, S. Agarwal & A.A. Nanavati, WWTW: The World Wide Telecom Web, *Proc. 2007 Workshop on Networked Systems for Developing Regions (NSDR 07),* Japan, 2007.

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

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

[14] E. Brewer, M. Demmer, M. Ho, R.J. Honicky, J. Pal, M. Plauche, & S. Surana. The Challenges of Technology Research for Developing Regions, *IEEE Pervasive Computing*, *5*(2), 2006, 15-23.

[15] A. Sharma Grover, M. P. Plauche, C. Kuun, & E. Barnard, HIV health information access using spoken dialogue systems: Touchtone vs. Speech, *Proc. IEEE/ACM 3rd Int. Conf. on ICTD*, Doha, Qatar, Apr. 2009, 95–107.

[16] J. Sherwani, S. Palijo, S. Mirza, T. Ahmed, N. Ali, & R. Rosenfeld, Speech vs. touch-tone: Telephony interfaces for information access by low literate users, *Proc. IEEE/ACM 3rd Int. Conf. on ICTD*, Doha, Qatar, Apr. 2009, 447–457.

[17] N. Patel, S. Agarwal, N. Rajput, A. Nanavati, P. Dave & T. Parikh, A Comparative Study of Speech and Dialed Input Voice Interfaces in Rural India. *ACM CHI 2009*. April 4 - 9, 2009, Boston, Massachusetts, USA, 51-54.

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

[19] E. Barnard, M. Plauche & M. Davel, The Utility of Spoken Dialog Systems, *Proc. Spoken Language Technology for Development workshop*, *2nd IEEE SLT workshop,* Goa, India, 2008, 13-16.

[20] F. Weber, K. Bali, R. Rosenfeld & K. Toyama, Unexplored Directions in Spoken Language Technology for Development, *Proc. Spoken Language Technology for Development workshop, 2nd IEEE SLT workshop,*Goa, India, 2008, 1-4.

[21] Srinivasan J, The Role of Trustworthiness in Information Service Usage: The Case of Parry Information Kiosks, Tamil Nadu, India, *Proc. IEEE/ACM International Conference on Information and Communication Technologies and Development (ICTD2007)*, India, 2007.

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

[23] L. Cloete, E. Barnard & H. Patel, Initial Experiments on the Effectiveness of Telephone Access to Government Services, *Lecture Notes in Computer Science*, *3183*, 2004, 564–565.

[24] C. Delogu, A.D. Carlo, P. Rotundi & D. Sartori, Usability evaluation of IVR systems with DTMF and ASR, *Proc. International Conference on Spoken Language Processing (ICSLP 98),* Australia, 1998.

[25] M. Franzke, A.N. Marx., T.L. Roberts & G.E. Engelbeck. Is Speech Recognition Usable? An Exploration of the Usability of a Speech-Based Voice Mail Interface. *ACM SIGCHI Bulletin*, *25*(3). 1993, 49 – 51.

[26] B. Suhm, J. Bers, D. McCarthy, B. Freeman, D. Getty, K. Godfrey & P. Peterson, A Comparative Study of Speech in the Call Center: Natural Language Call Routing vs. Touch-Tone Menus, *Proc. 2002 Conference on Human Factors in Computing Systems,* (CHI-02), 2002, 283 – 290.

[27] E. Barnard, M. Davel & C. van Heerden, ASR Corpus Design for Resource-Scarce Languages, *10th Annual Conference of the International Speech Communication Association,* Interspeech 2009, Brighton, U.K., September 2009.

[28] O.T. Stewart & H. Banchard, Linguistics and Psycholinguistics in IVR Design. In D.J. Gardner-Bonneau, & H. Blanchard (ed), *Human factors and voice interactive systems*, 81-115 (Boston, USA: Kluwer, 2007)

[29] R. Patra, J. Pal, S. Nedevschi, M. Plauche & U. Pawar, Usage Models of Classroom Computing in Developing Regions, *Proc. IEEE/ACM International Conference on Information and Communication Technologies and Development (ICTD2007*), India, 2007.

[30] C. Wei & B. Kolko, Studying mobile phone use in context: Cultural, political, and economic dimensions of mobile phone use, *International Professional Communication Conference*, Limerick, Ireland, July 10-13, 2005.

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

[32] N. Yankelovich, How do users know what to say?, *Interactions (ACM), 3*(6), 1996, 32 – 43.

[33] M. H. Cohen, J.P. Giangola & J. Balogh, *Voice user interface design* (Boston: Addison- Wesley, 2004).

[34] I. Medhi, A. Sagar & K. Toyama, Text-Free User Interfaces for Illiterate and Semiliterate Users, *Information Technologies and International Development, 4*(1), 2007, 37–50.

[35] T.S. Parikh & E.D. Lazowska, Designing an Architecture for Delivering Mobile Information Services to the Rural Developing World. *Proc. 2006 International World Wide Web Conference (WWW 2006),* 791 – 800, Scotland, 2006.

[36] B. DeRenzi, N. Lesh, T. Parikh, C. Sims, M. Mitchell, W. Maokola, M. Chemba, Y. Hamisi, D. Schellenberg & G. Borriello, e-IMCI: Improving Pediatric Health Care in Low-Income Countries, *Proc. 2008 SIGCHI conference on Human factors in computing systems (CHI-2008),* 753-762, 2008.

[37] M. Huenerfauth, Developing design recommendations for computer interfaces accessible to illiterate users. Masters thesis, University College Dublin. 2002.

[38] A. Marcus, Metaphor Design for User Interfaces, *Proc. 1998 SIGCHI conference on Human factors in computing systems (CHI-1998),* 129-130, 1998.

[39] P. Barr, R. Biddle & J. Noble, A Taxonomy of User-Interface Metaphors, *Proc. of SIGCHI-NZ Symposium on Computer-Human Interaction,* New Zealand, 2002.