Comparing Two Developmental Applications of Speech Technology

Aditi Sharma Grover[1]
[1]CSIR Meraka Institute,
P.O. Box 395,
Pretoria, 0001,
South Africa
asharma1@csir.co.za

Etienne Barnard[1,2]
[2]Multilingual Speech Technologies
Group, North-West University,
Vanderbijlpark, 1900
South Africa
etienne.barnard@gmail.com

Abstract

Over the past decade applications of speech technologies for development (ST4D) have shown much potential for enabling information access and service delivery. In this paper we review two deployed ST4D services and posit a set of dimensions that pose a conceptual space for the design, development and implementation of ST4D applications. We also reflect on these dimensions based on our experiences with the above-mentioned services and some other well-known projects.

1 Introduction

The use of speech technology for developmental purposes (ST4D) has seen slow but steady growth during the past decade (for an overview, see Patel, 2010). However, much of that growth has been opportunistic in nature: researchers or development agencies determine an environment where speech technology (ST) may potentially be useful, and then deploy a variety of methods and tools in order to deliver useful services (and ask appropriate research questions) in that environment. To date, no overarching framework that could assist in this process of service identification – design – implementation – analysis has emerged.

Having been involved in a number of such efforts ourselves, we have experienced the lack of theoretical guidance on the application of ST4D as a significant challenge. Hence, we here attempt some initial steps in that direction. In particular, we review two services that we have deployed, sketch a number of important ways in which they differed from one another, and use that comparison to derive some abstract dimensions that partially span the space of ST4D applications. Of course, these are early days for ST4D, and much remains to be discovered. We nevertheless believe that the explication of this space will be useful guidance for further research in this field.

2 Two deployed ST4D services

We describe in this section two telephone-based services that were designed for developing world regions with varying contexts and goals, and use them in subsequent sections to illustrate the various dimensions related to design and deployment of ST4D.

2.1 Lwazi

The Lwazi Community Communication Service (LCCS) is a multilingual automated telephone-based information service. The service acts as a communication and dissemination tool that enables managers at local community centres known as Thusong community centres (TSCs) to broadcast information (e.g. health, employment, social grants) to community development workers (CDWs) and the communities they serve. The LCCS allows the recipients to obtain up-to-date, relevant information in a timely and efficient manner, overcoming the obstacles of transportation, time and costs incurred in trying to physically obtain information from the TSCs. At a high-level, the LCCS can be viewed as consisting of 3 distinct parts as illustrated in Figure 1.

During our investigations we found that TSCs often need to communicate announcements to the CDWs, who in turn disseminate the information to the relevant communities. The TSC managers and local municipal (government) communication officers most often have regular Internet connectivity and use emails in their daily routine to communicate with government departments and non-profit organisations (NPOs).

In the majority of the communities we investigated, communication with the CDWs is mostly through face-to-face meetings and the telephone. CDWs on average had grade 8-12 or higher level of education (vocational certificate courses),

Conference on Human Language Technology for Development, Alexandria, Egypt, 2-5 May 2011.

ranged between 20-45 years in age, with no significant differences in the gender proportions. They were familiar with mobile phone usage and some even used computers and the Internet. Almost none of the community members have access to the Internet, while most of the households have access to at least a mobile phone.

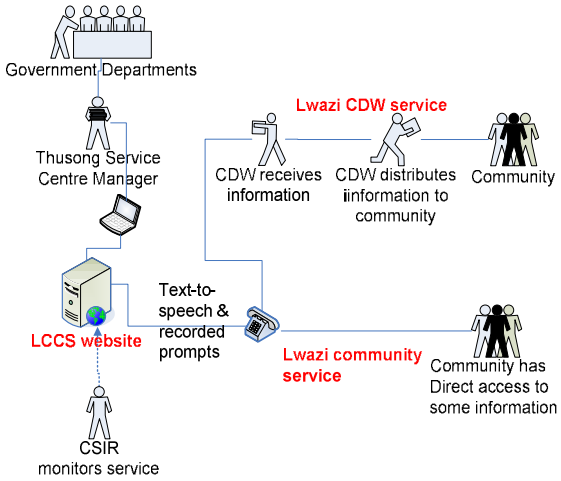

Figure 1. Overview of the LCCS.

2.1.1 LCCS website

This site provides a TSC manager with the ability to upload new announcements to the service for CDWs and/or community members. Managers may choose to send an announcement to all the CDWs registered in their area, or select group, and/or the rest of the community. The manager provides information on relevant fields (date, time, etc.) and selects the languages he/she would like the message to appear in. For each additional language selected, additional textboxes are provided to type the message text for those languages. The audio structure of the announcements as heard by the community workers is shown in figure 2. A 'View Messages' page also provides managers with a voicemail-like facility to listen to messages (new and old) on the LCCS website left by CDWs through the Lwazi CDW service.

2.1.2 LCCS telephone services

Two IVR services, namely the Lwazi CDW service and the Lwazi community service were developed. The former allows CDWs to phone in and get access to the announcements uploaded by the TSC managers and leave voice messages for the TSC managers. Each CDW in a targeted area is registered on the CDW service and obtains a PIN code that allows them to access their announcements.

Within each targeted area the service was provided in the most commonly spoken languages of that area. Daily at 5 pm, all the recipients receive an SMS notification if they have new message(s) on the LCCS. The service is free in the sense that a CDW only gives a missed call to the service (calls the service number and hangs up after a ring), and the service then calls them back. The CDW also gets the option in the main menu to record a voice message for the TSC manager. This feature was envisaged to save the CDW the cost of a phone call to leave a voice message for the TSC manager using their mobile phone (i.e. to directly dial the TSC manager's number vs. using the LCCS). The TSC manager, in turn, receives all her/his community workers' related messages through the LCCS website where he/she can easily store and retrieve them as required.

The announcement is played back using a combination of pre-recorded voice prompts for the dates and time fields and TTS (in blue) for the less predictable fields like the message text.

Lwazi (user is called back on registered number): *Welcome to the Lwazi Service. You have 2 new messages. To hear your New Message, press 1. To hear your Old Messages, press 2. To Leave a Message for the Thusong centre manager, press 3. Note that you can return to the 'Start' of the service at any time, by pressing 0.*
User: *(presses 1)*
Lwazi: *New Messages. Note that you can skip to the next message, by pressing 1. First message:*
Lwazi: *Description of Meeting: <(TTS:) Weekly progress report meeting>*
Date: <29th> of <March>
Venue: <(TTS:) Thusong Service centre>
Starting Time: <3> <30> <PM>
End time: <5> <PM>
Message: <(TTS:) All CDWs are requested to attend this weekly meeting where community participation strategies will be discussed>.

Figure 2. Lwazi CDW service- sample script.

The Lwazi community service (second IVR) extends from the Lwazi CDW service; if a TSC manager wishes to make announcements to the community in general, community members are marked as recipients on the LCCS website. Therefore, the same announcement which goes out to the CDWs is then accessible to the community members as well. Community members similarly give the service a missed call, which calls them back making the service free for their usage. However, the community line differs from the CDW line as no registration is required by the communities (i.e. no PIN codes thus any user

can call in or SMS notifications of a new message on the service).

The LCCS was piloted at 6 sites in South Africa across the 11 official South African languages. The pilot areas included rural (3), semi-rural (2), and urban areas (1). Note, these pilots were intended as short-term deployments running on average 4-12 weeks to determine the uptake and usage trends in each area.

2.2 OpenPhone

The OpenPhone service is a general health information line designed for caregivers of HIV positive children in Botswana. A caregiver is any individual who takes care of an HIV positive child, e.g. parents, family or community member. The vast majority of caregivers are females and range between the ages of 18 and 65 with most being semi and low literate. Most caregivers tend to have low-income jobs and many are often unemployed. Baylor, a HIV paediatrics hospital where children receive free treatment provides the caregivers with free lectures on various aspects of living with HIV and caregiving. Each caregiver on average attends two lecture sessions. It was observed that caregivers often forget the material covered in the lectures. Reinforcement through written material is not viable as many caregivers are semi-to-low literate. The Baylor lectures and all interactions with caregivers are in Setswana (local language) since most caregivers are uncomfortable with English. Baylor staff explains complex health information in accessible terms in the local language. Mobile phone usage and ownership was widely prevalent (up to 90%).

It was also observed that caregivers often travel large distances (average 28 km and as far as 500 km) with high costs and time spent during a working day. Also they often have general health information queries (e.g. nutritional needs, hygiene, etc.) beyond the material covered in the lectures and although caregivers are encouraged to call Baylor with any questions they may have, most are reluctant (and unable) due to the high costs of mobile phone calls. These challenges and issues formed the basis of the design for OpenPhone, an IVR health information service in Setswana and accessible at any time through a simple telephone call. A sample system-user interaction is shown in figure 3.

Openphone was piloted for a week at Baylor where the service was tested with 33 caregivers with one of the major questions being the preference of input modality between automatic speech recognition (ASR) and touchtone (DTMF), where ASR was simulated by using the wizard of Oz (WoZ) methodology (Fraser & Gilbert 1991). Two identical systems were built that differed only at the menu prompts in choice of input modality, e.g. a DTMF menu option would be; "to hear about Nutrition, press 1," whereas the ASR menu option would say, "to hear about Nutrition, say Nutrition."

On average caregivers reported that mobile phone costs per month were 68 Pulas ($10.5 USD) with an average cost per call being reported as 4.5 Pulas ($0.75 USD). Only 30% of caregivers reported having access to a landline telephone and of these only 9% had the landline at home.

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

Figure 3. OpenPhone – sample script.

3 Charting the ST4D space

In this section we posit a set of dimensions that pose a conceptual space for the design, development and implementation of ST4D. We reflect on these dimensions based on our experiences with the above-mentioned projects and some other well-known projects.

3.1 Nature of the user community

Various user-related factors are vital in ST4D. These include literacy, the technology experience of the user and the 'openness' of the user community.

Undoubtedly *literacy* of the target users is one of the major considerations in any information and communications technology for development (ICTD) project, and speech applications have often been proposed to address user interface problems with other modalities experienced by low and semi-literate users (Plauche et al, 2007; Medhi et al, 2007). However, Barnard et al (2008) highlighted that, without training or prior exposure, a large class of speech applications

may not be usable for the developing world user. In OpenPhone for example, at the end of one call, a user proceeded to ask the 'nurse' (system persona) a question when prompted by the system to leave a comment (and waited for the answer), highlighting that some users within low and semi-literate populations' may not fully realise that a service is automated.

Technology experience coupled with literacy may also affect an application's uptake. The uptake of mobile phones in the developing world is widespread (Heeks, 2009); users who are comfortable with related features of mobile technologies (e.g. SMS, USSD, a service provider's airtime loading system) may perform better on speech applications. In OpenPhone we found that experience 'loading airtime' [1] correlated positively with higher task completion. We also found that loading airtime was the sole significant factor in user preference of DTMF over the ASR system i.e. those people who loaded airtime regularly preferred DTMF over ASR. These experiences perhaps indicate that prior technological experience may be just as important a factor in adopting new technology as literacy.

Another factor with great variegation is the *'openness'* of the user community. Here we refer to how membership to the user community is determined; e.g. in Lwazi the user community was essentially closed: a set of CDWs who were paid government workers. This leads to a set of repeat users, who are also trainable if required. In contrast, in OpenPhone the user community was more open in that any caregiver (or anyone with the application's phone number) could call the application. However, it would be hard to identify repeat users (except through caller ID) and even harder to provide such a user community with training.

This factor also affects design choices; in catering for a closed community, known commonalities between users allow the designer to fine-tune the application towards their specific needs. In contrast, a more open community would have various types of users (e.g. cultural, language, age, literacy differences) where adapting the application becomes much more difficult due to the user diversity. For example, in Lwazi our pilot sites varied from rural to urban as well as across 11 languages, which made the system persona design quite a challenge.

3.2 Content source

Generation of content is a critical issue for not only ST4D, but any ICTD application. Here it is of essence to provide the users with *relevant and timely content*. The *content source* may be locally (user/community) generated or externally generated by the system designers themselves. This of course, has implications on *content updates and its timeliness*. For example, in Lwazi the content was generated within the community by the TSC manager (and sometimes CDWs) – thus, it was very locally relevant, could be easily updated and was timely in nature but this led to a large reliance on the role of the TSC managers to provide an announcements 'feed' to the application. In contrast, in OpenPhone the content was purely externally generated removing the reliance factor but placing a large burden in the design process to investigate and ensure that content provided was relevant, useful and timely.

Capturing high-quality information in a developing-world situation is a significant challenge in itself. The existence of electronic information sources that provide content feeds for the ST4D application (as are typically employed in the developed world) would be ideal when building such applications. However the nature of the information space in the developing world is such that these sources are generally unavailable. Thus, building them from scratch often takes bulk of the effort required in designing such speech applications. Also, the fact that the content needs to be available (or translated) in the *local language* further magnifies the issue.

It is also worthwhile to note that the *sensitivity of the content* may also affect the choice of content source. User generated content may not be feasible (and legally possible) in domains with highly sensitive information such as health as in the case of OpenPhone. Interestingly, even in other domains such as agriculture, Patel et al (2010) found that farmers preferred to obtain information from known and trusted experts rather than other framers. This goes to illustrate that content *trustworthiness* is also a significant factor.

3.3 Application complexity

In comparison to the HLT investment in the developed world, much advancement still needs to be made for the developing world resource-scarce languages. "How may I help you?" type

[1] Loading airtime refers pre-paid phones which require users to load money by calling the network provider's service number and entering a sequence of digits from the pre-paid calling card which can be done through IVR or USSD.

of applications are certainly a long way off, rather we propose that a *'human in the loop'* approach may be required in the interim such that technology constraints and inefficiencies do not prohibit the uptake of speech-based interfaces. In our initial pilots of Lwazi the CDWs said that they sometimes struggled to hear their messages (TTS parts) and felt that voice sounded somewhat 'funny' and robotic and as if a non-mother tongue speaker was trying to speak a language.

Thus, to ensure that a TTS audio announcement was sufficiently intelligible, we introduced a 'human-in-the-loop', where every time an announcement is posted, a notification was sent to our TTS support team which checked the quality of the rendered audio file for the announcement and, if required, normalised the input text (e.g. special characters and acronyms) to render better quality TTS audio that would be posted to the telephone services. Similarly an IVR-based service could be used in the front-end of question-answering service with humans in the back-end answering questions and channelling them back to the user via TTS or SMS.

3.4 Business model and deployment

The success of ST4D applications is subject to the same provisions as many other ICTD applications. One of the major factors is *cost*; applications that require even the cost of a local phone call may be prohibitively expensive for many users in the developing world. We found in OpenPhone that the majority of caregivers said that even though the service would be useful to them they would only be able to make use of it if the service is toll-free. An average phone call in Botswana of 5-10 minutes to the service would cost a mobile phone user $1-2 USD.The average cost per month for mobile phone usage was $10.5 USD. Thus, a single phone call to the service would consume, 10-20% of a caregiver's monthly mobile phone budget, making the case for a toll-free number all the more imperative.

In Lwazi, a notable observation was that community members sometimes struggled to understand how the service could be free if they needed airtime (prepaid balance) in order to give it a missed call. Community members would often be wary of using their own phones to test the service, afraid that their airtime would be used. We tried to address this by showing them (with our phones) that no charge was incurred upon calling the service. In some cases, we also noted that people did not have enough airtime to even give a missed call to the service.

Based on these experiences and those highlighted by Patel *et. al* (2010) we stress that users are extremely sensitive to call costs and the need to pay for such a service, which makes cost a decisive factor in the widespread uptake of a speech application. To our knowledge all ST4D efforts reported on to date have been 'free' initiatives (i.e. the user does not pay) with the exception of the BBC Janala project (Heatwole, 2010) where users pay subsidized call rates to access English lessons over an IVR.

Stakeholder support is a related aspect of great importance; one needs to ensure that the relevant stakeholders (users, community organizations, researchers, donors, etc.) involved in the project support the application and understand the practical roles they play in the ability of the technological intervention to solve a problem. In Lwazi we found that, of the six pilot sites, the one with the highest usage was where the keenness of the TSC manager and the government communication officers was the strongest. The latter, in particular, were young, open to trying new technologies and quite familiar with the Internet. They were enthusiastic to try out the service if it would assist in making their jobs easier and also contacted us to provide valuable feedback and suggestions for improvement. Thus the role of such intermediaries cannot be over-emphasized. Clear planning must be executed around the financial, organisational, operational support and human capacity required from these stakeholders.

Related closely to the above stakeholder support factor is the need to *align new ST4D applications with existing channels*. Rather than trying to replace an existing system (resistance from the status quo) it may be better to introduce a complementary service that rather leverages on existing channels and provides an added advantage. In OpenPhone the health info line was meant as a means to augment the Baylor lectures and not replace them.

In Lwazi we found that in the pilot sites where there was a fairly workable system in place for communicating with the CDWs, the service did not fare well. For example, in one site the TSC manager just preferred to communicate with the CDWs via telephone and did not mind the cost factor as the calls were government-sponsored. In another site, CDWs had government sponsored Internet access and laptops (a very rare occurrence). Their primary means of communication with the TSC manager was through face-to-face meetings or email. The TSC seemed to be

operating quite well, and the existing channels set up for communication between CDWs and the TSC manager were actively used and worked well for their needs.

Two of the CDWs mentioned that their offices were located (unusual for CDWs to have fully-fledged offices) close to that of the TSC manager and they may therefore not be using the service so much. Another suggested that the service be linked to their emails. From this, we surmised that the service only supplemented these existing channels rather than leverage on them and thus was not so widely used.

The above discussion illustrates that a *sustainability model* is essential for a speech applications' deployment. Here, the obvious potential options to consider include government subsidies, toll-free lines sponsored by telecommunications providers (this has been surprisingly rare in developing regions) but also consider models around marketing companies that target emerging markets or user-paid services (finding a strong value proposition for the user which makes them willing to pay for the service as in the case of BBC Janala). The sustainability challenge will prove to be a significant one in the long-term deployment of such applications.

Also relating to sustainability, it is important to take cognisance of the fact that *pilot deployments may be viewed exactly as that by the community* and therefore may not obtain their full buy-in, i.e. if the community knows that a service will only be available for a short period of time, they may be less inclined to use it than if it were introduced as a permanent solution. In general, ICTD researchers need to carefully balance this aspect with that of not creating false expectations when introducing ICT interventions into communities.

4 Conclusion

All of the dimensions discussed in Section 3, create a rather complex space which leads to very different classes of applications due to choices made on the various dimensions involved. These choices affect the design of a ST4D application in several different ways, including the *input modality* (touchtone vs. speech recognition), *menu design* (hierarchical vs. non-linear), *prompts* (TTS vs. pre-recorded) and system *persona* design.

As the application of ST4D becomes more widespread, this conceptual space will undoubtedly be understood in much more detail, and the characteristics of the most useable regions within the space will be defined with rigor and precision. We believe that such systematic knowledge will greatly enhance our ability to deliver impactful services.

Acknowledgments

This work was funded and supported by South African Department of Arts and Culture (Lwazi) and OSI/OSISA (OpenPhone). The authors wish to thank the numerous HLT research group members but especially Madelaine Plauche, Tebogo Gumede, Christaiaan Kuun, Olwethu Qwabe, Bryan McAlister and Richard Carlson who played various roles in the Lwazi and OpenPhone projects.

References

Agarwal, S., Kumar, A., Nanavati, A., and Rajput, N. 2009. Content creation and dissemination by-and-for users in rural areas. *In Proc. IEEE/ACM Int. Con. On ICTD (ICTD 09)*. April 2009, 56-65.

Barnard, E., Plauche, M.P., Davel, M. 2008. The Utility of Spoken Dialog Systems. In *Proc. IEEE SLT 2008*, 13-16.

Fraser, N.M., and Gilbert, G.N. 1991. Simulating speech systems, *Computer Speech and Language*, 5(2), 81-99.

Heeks, R., 2009. IT and the World's 'Bottom Billion'. Communications of the ACM, vol. 52(4), 22-24.

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

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

Plauche M.P., Nallasamy U., Pal J., Wooters C., and Ramachandran D. 2006. Speech Recognition for Illiterate Access to Information and Technology. In *Proc. IEEE Int Conf. on ICTD06.*

Plauche M.P. and Nallasamy U. Speech Interfaces for Equitable Access to Information Technology, Information Technologies and International Development (MIT Press), vol. 4, no. 1, pp. 69-86, 2007.

Sharma Grover, A., Plauche, M., Kuun, C., Barnard, E., 2009. HIV health information access using spoken dialogue systems: Touchtone vs. speech. *In Proc. IEEE Int. Conf. on ICTD* 2009 (ICTD 09), 95–107.

Sharma Grover, A., and Barnard, E., 2011, The Lwazi Community Communication Service: Design and Piloting of a Voice-based Information Service. *In Proc. WWW 2011 (to appear).*

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