EXPERIENCES DESIGNING A VOICE INTERFACE FOR RURAL INDIA

Neil Patel [1], Sheetal Agarwal [2], Nitendra Rajput [2], Arun Kumar [2], Amit Nanavati [2], Paresh Dave [3], Tapan S. Parikh [4]

[1]Stanford University HCI Group
Computer Science Department, Stanford, CA 94305
neilp@cs.stanford.edu

[2]IBM India Research Laboratory
New Delhi, India, 110070
{sheetaga, rnitendra, kkarun, namit}@in.ibm.com

[3]Development Support Center
Ahmedabad, Gujarat, India
pdave68@gmail.com

[4]UC Berkeley
School of Information, Berkeley, CA
parikh@ischool.berkeley.edu

ABSTRACT

In this paper we describe our experiences designing a voice interface in rural India. We outline our design process from initial contextual inquiry to a formal user evaluation, and use this discussion to motivate research guidelines for others designing voice interfaces in developing regions. Our three guidelines are to *build around existing information systems*, to *iterate on the design through user testing*, and to *explore design alternatives through empirical analysis*. We also share some practical lessons learned in designing, implementing, and evaluating information systems for developing regions in general.

Index Terms — Voice Interfaces, India, Rural Development, Agriculture, Semi-literate, ICTD

1. INTRODUCTION

Voice-based interfaces have been identified for their potential to increase access to information services in developing countries like India, where 480 million illiterate people reside [13]. However, much of the prior work in voice user interface (UI) design has been with users in developed countries [2, 3, 9]. In addition, the last fifty years of research in speech technologies has focused on a small number of languages, and assumed the availability of costly speech resources such as pronunciation dictionaries and annotated speech corpora [11].

In rural areas, low literacy combined with intermittent power and connectivity makes the possibility of widespread usage of Internet-connected PCs still a distant vision. Yet at the same time, the mobile phone has rapidly penetrated even the most remote areas of the world. Currently, there are 3.5 billion active mobile phone subscriptions worldwide, or over half the world's population [6]. Mobile, voice-based interfaces have the potential to address both the literacy and connectivity constraints of rural populations simultaneously.

In this paper we describe our experiences designing, implementing, and evaluating a mobile voice interface for farmers in Gujarat, India to access and share agricultural information. We highlight three guidelines which we hope will inform the design process of researchers or practicitioners working with spoken language technologies for rural areas in the developing world.

2. AVAAJ OTALO

Working with Development Support Center (DSC) in Ahmedabad, Gujarat, we designed *Avaaj Otalo* (roughly translated as "a voice-based community forum"), a Gujarati-language application allowing farmers to receive timely and relevant agricultural information over the phone. A user interacts with the system by navigating a hierarchical menu using isolated-word spoken commands. Through a joint needs-finding process with DSC, we identified three features for the initial prototype system to support. After several rounds of design iteration including informal testing with farmers, we implemented a prototype version of Avaaj Otalo. The system was built using IBM Research's WWTW [8] platform. The Gujarati commands were converted to lexicons with phonemes from the US English set. Although there are some Gujarati phonemes that do not have exact English equivalents, this technique has been found acceptable for simple command vocabularies [12]. After implementing the system, we ran a controlled user study to evaluate user performance based on input modality (speech vs. dialed input) [10]. We found that dialed input performed significantly better than speech input in terms of task completion rate. In terms of the application's concept and functionality, the response from farmers was unanimously enthusiastic. Based on the positive response, we will be piloting Avaaj Otalo this winter in preparation for a full

launch to serve over 500,000 farmers across the state of Gujarat.

3. GUIDELINES

In this section, we offer an initial set of guidelines for researchers designing voice interfaces for developing regions. The main goal of these guidelines is to initiate a discussion about how to design for this domain; they are not an exhaustive set, and may not apply to all scenarios. However, we believe that successful implementation and evaluation of Avaaj Otalo would not have occurred without these three factors. The first two guidelines, to leverage existing information systems and to iterate with users, pertain to practical matters of identifying needs and creating a working, usable system. The final guideline, evaluation through empirical methods, addresses how we can demonstrate our results in a way that can inform future work.

3.1. Guideline #1: Leverage Existing Systems

There is a long list of past computing projects for the developing world that have not sustained beyond pilot deployments [5]. An oft-mentioned explanation is insufficient incorporation of the externally developed technology artifact with existing practices and processes. Information in the developed world is currently accessed and shared through "low tech" media: human interaction, the phone, radio, television. While designing Avaaj Otalo, we learned that community-based radio is an effective means of providing localized information about topics such as agriculture. While the radio broadcasts provide a highly effective mechanism for delivering information, they do not easily support ongoing, multi-way dialogue. This inspired us to introduce Avaaj Otalo as a feedback mechanism for radio listeners. The system enables a new level of interactivity for the radio program; conversations can develop around topics discussed on the air, and can also be highlighted on a later radio broadcast. The feedback loop enabled by combining a mobile feedback channel with the radio broadcast is illustrated in Figure 1.

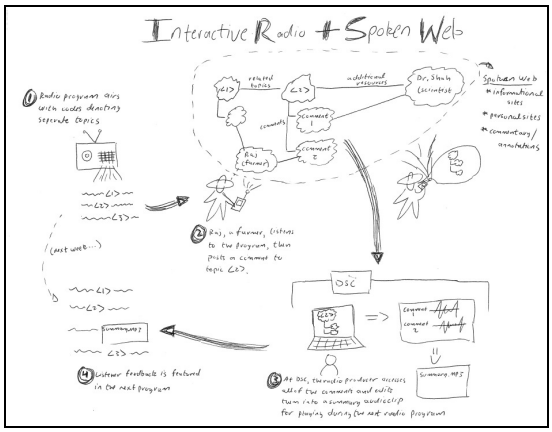

Figure 1: A diagram of Avaaj Otalo sketched during initial brainstorming

Through our design process, we found that integrating our system with the existing radio program provided us with an advantage over a standalone application. From a content perspective, the radio programs themselves serve as a basis for conversations through Avaaj Otalo. Farmers would be able to contribute questions and comments about issues and recommendations given on air. Secondly, the credibility of DSC and the popularity of the radio program would have a positive effect on uptake of the new system. Finally, the existing audience of the radio program immediately becomes potential users of the new system.

3.2. Guideline #2: Iterate with Users

We attempted to include farmers in the design process of Avaaj Otalo from an early stage, following the participatory design imperative [1, 4]. This was especially important because interactive voice interfaces are unfamiliar to rural populations in Gujarat. Through study in the field, we were able to gain insights that wouldn't have otherwise been possible. For example, in early design of the interface's prompts, we took the approach recommended by existing literature. This was to make interaction with the system as conversational as possible [1]. However, we found that a prompt such as "Welcome to Avaaj Otalo! What would you like to do?" did not work as well as "Welcome to Avaaj Otalo! To record a question, say 'question'; to listen to announcements, say 'announcements'…". Prompts that require users to infer something about the system's properties were confusing. Subsequently, we changed to explicit, directive-style prompting.

This learning would have had a disastrous consequence if we discovered it after implementing a working prototype. However, we were able to get such feedback through early, low-fidelity prototype scenarios. During a field visit, we had potential users "act out" an interaction with the system in Wizard-of-Oz [7] style using humans and flashcards (see Figure 2). Later in the implementation process, we also benefitted from developing simple features early and testing

them with famers that already felt comfortable with the technology. This served as a sanity check that the application could be used under the best possible task/user scenarios.

Figure 2: Testing prompts in the field using paper flashcards read to the user from behind.

Finally, we caution that this guideline should not discount the value of indirect means of learning about the needs of the target population. We found that observation and interaction with DSC field staff was invaluable in identifying the conceptual definition and features of Avaaj Otalo. Often, gathering ideas for features from farmers was constrained by their inexperience with computing systems; they didn't know what is possible. NGO staff and field agents have the unique combination of understanding the realities on the ground as well as significant exposure to computers. We found our partner's field agents were helpful in validating our design ideas and suggesting alternatives.

3.3. Guideline #3: Evaluate Design Choices Empirically

In order to advance our understanding of how to design voice interfaces for developing regions, we must also develop objective knowledge about voice UIs' usability and effectiveness. The design space for voice interfaces includes dimensions such as input modality, prompt style, call flow structure, and use of sound effects. There is a need to quantitatively and qualitatively compare alternatives within each of these dimensions. Aggregating knowledge through experiments that are replicated in a variety of rural settings will enable the community as a whole to begin developing theories about the "right" design for voice UIs for this domain.

As one data point, we ran a comparative study of voice and DTMF (dialed) input for Avaaj Otalo. We conducted a controlled, between-subjects experiment with 45 participants, most of whom had less than an eighth grade education (see Figure 3). The goal of our study was to compare performance and user preference between the two modalities. We found that DTMF dominates in terms of performance and learnability, and that users found they had much less difficulty providing input using DTMF. This result is useful for practitioners who are choosing between input alternatives, where high-capability ASR technology may be costly in terms of time, effort, and resources.

Figure 3: Evaluating Avaaj Otalo with a farmer in her village home. The user is testing the DTMF version of the interface.

4. PRACTICAL LESSONS

Our experience designing Avaaj Otalo revealed some practical lessons that may extend beyond voice interfaces to the design of other information systems for the developing world. There were several factors that were critical for us to execute our design-implement-evaluate process. First, we began with a clear plan in mind. While user-centered design literature is justified in delaying prototype development until after user needs and capabilities have been assessed, we believe our design process benefitted from establishing a rough blueprint of our system as a starting point. At the beginning of our partnership with DSC, we presented the diagram in Figure 1 to DSC to communicate not only a possible mobile application, but also our underlying design philosophies, including the desire to enhance rather than replace existing systems. This early work served as useful scaffolding for subsequent brainstorming discussions.

Second, we depended heavily on the capabilities of our NGO staff partners. Coordination of field visits and the user study required relationships with farmers that would not have been possible to establish ourselves, especially in a short time period. In this sense, the NGO staff served as a proxy for directly building trust with farmers.

Finally, we believe that in order to avoid the fate of many research prototypes that flounder due to lack of resources, outlining a sustainability plan is critical. Such a plan should not only address financial sustainability, but clearly define

the roles of the various stakeholders (the NGO, funders, researchers, and end-users) in maintaining the system over the long term.

5. CONCLUSION

In this paper, we described our experiences designing Avaaj Otalo, a voice interface for farmers in Gujarat, India to access and share information over a phone. We offered three guidelines for further research in voice UI design, as a starting point for spoken language technology researchers interested in working in rural areas of the developing world. Two points are of particular interest to research in this domain. First, we found that the participatory design process is aided by "proxy" users that have both knowledge of the practice and the capabilities of new tools. While the Scandanavian PD tradition may prescribe a more direct approach, we believe that in this domain, other stakeholders such as grassroots organizations may be beneficially leveraged. Second, we emphasized the importance of rigorous quantitative evaluation in terms of usability and performance. Results in this vein, especially if validated through repeated experimentation, may allow useful heuristics for practitioners.

6. REFERENCES

[1] Beyer, Hugh, K. Holtzblatt. "Contextual Design – Defining Customer-centered Systems", Morgan Kaufmann, 1998

[2] Cohen, Michael, J.P. Giangola, J Balogh, "Voice User Interface Design," Addison-Wesley, 2004

[3] Delogu, Cristina, A.D. Carlo, P. Rotundi, D. Sartori, "A Comparison Between DTMF and ASR IVR Services through Objective and Sujective Evaluation," *IVTTA*, 1998.

[4] Ehn, Pelle. Work-Oriented Design of Computer Artifacts. Lawrence Erlbaum Assoc., January 1, 1988

[5] Fonseca, Rodrigo, J. Pal, "Computing Devices for All: Creating and Selling the Low-Cost Computer," *International Conference on Information and Communication Technologies and Development (ICTD2006)*, Berkeley, 2006.

[6] International Telecommunications Union. Press release, 2008. http://www.itu.int/newsroom/press_releases/2008/29.html

[7] Kelley, Jeff. An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Office Information Systems* 2(1) (1984), 26–41.

[8] Kumar, Arun, N. Rajput, S.K. Agarwal, D. Chakraborty, A. Nanavati, "WWTW: The World Wide Telecom Web," *SIGCOMM Workshop on Networked Systems for Developing Regions*, Japan Nov. 2007.

[9] Lee, Kwan Min, J. Lai, "Speech Versus Touch: A Comparitive Study of the Use of Speech and DTMF Keypad for Navigation," *International Journal of Human-Computer Interaction*, 2006.

[10] Patel, Neil, Sheetal Agarwal, Nitendra Rajput, Amit Nanavati, Paresh Dave, and Tapan S. Parikh, "A Comparative Study of Speech and Dialed Input Voice Interfaces in Rural India," http://www.stanford.edu/~neilp/pubs/otalo_submitted.pdf. In submission to CHI 2009.

[11] Plauche, Madelaine, Udhayakumar Nallasamy, Joyojeet Pal, Chuck Wooters, and Divya Ramachandran, "Speech Recognition for Illiterate Access to Information and Technology," *International Conference on Information and Communication Technologies and Development (ICTD2006)*, Berkeley, 2006.

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

[13] United Nations Development Program. Human Development Report, 2004. http://hdr.undp.org