Visual Conversational Interfaces to Empower Low-Literacy Users

Sheetal Agarwal, Jyoti Grover, Arun Kumar, Monia Puri, Meghna Singh, Christian Remy

Visual Conversational Interfaces to Empower Low-literacy Users

Sheetal Agarwal[1], Jyoti Grover[1], Arun Kumar[1], Monia Puri[1], Meghna Singh[1], and Christian Remy[2]

[1] IBM India Research Lab, ISID Campus, Block C, Vasant Kunj, New Delhi, India-110070

[2] University of Zurich, Binzmühlestrasse 14, 8050 Zurich, Switzerland

{sheetaga,jyogrove,kkarun,monipur,meghna.singh}@in.ibm.com, remy@ifi.uzh.ch

Abstract. Mobile phones have come a long way from being plain voice calling devices to becoming multipurpose handy tools powered by ever increasing new applications available on-the-go. For many, the mobile phone of today has become the essential device one does not leave home without. However, for a large percentage of human population mobile phone apps are not of much use as they are not literate or IT savvy enough to be able to benefit from them. Recent advances in voice-based telecom information systems enable underprivileged and low-literacy users to access and offer online services without requiring expensive devices or specialized technical knowledge. However, voice applications are limited in their capability due to their time consuming nature. In this paper, we demonstrate an interaction modality that combines the power of voice communication with graphical interfaces in smartphones to break the barrier of illiteracy.

Keywords: Diversity**,** HCI4D, Information Sharing, ICTD, User-Centered Design, Interactive Voice Systems, Smartphones, India.

1 Introduction

Over the last couple of years, smartphones have proliferated at a large scale even in developing countries. Emergence of new local manufacturers has led to easy availability of smartphones in the market costing less than USD 100 [2]. The plethora of mobile apps available today have brought applications in domains ranging from personal management to healthcare, collaboration to education, at users' fingertips. However, even though a lower upfront cost has helped extend the reach of smartphones, their utility still remains limited to voice calls for a large section of their

users in developing countries. The primary reason for this is that the running cost of downloading and using several apps is high if an active internet connection is required. Furthermore, another barrier is the low level of literacy or English literacy that prevents many needy users from making use of their powerful handsets effectively.

On the other hand, advances in interaction experience of telephony voice user interfaces have seen good uptake by this underserved population in developing countries [1, 5, 10]. Traditionally, they have not had any access to online information systems due to reasons of affordability, local relevance, and illiteracy.

In this paper, we propose to marry voice user interfaces available on ordinary phones with the power of graphical user interfaces for Visual Conversational interfaces that achieve a two-pronged effect. First one is to enable textually illiterate people to harness the power of mobile apps available on their smartphone devices. Second is to exploit the rich user interface available through mobile apps to make existing applications usable by such users. Examples include even basic device based services such as SMS and address book among others that are currently not usable by low-literate. In this paper, we demonstrate the use of smartphones to enable a voice content sharing telephony application for textually illiterate users without the need for an Internet connection.

2 Voice Interface intermingled with visual interface

To demonstrate the synergy between voice user interfaces and smartphone based graphical user interfaces, we took the scenario of voice content sharing over telephony voice applications. The voice interface part of the application allows callers to identify and generate a link to online voice content of interest. This link is received by the caller as SMS which can then be shared with others. A corresponding mobile application makes use of intuitive icons that let the user call online voice application for accessing voice content. The application also allows navigation of local repository containing links received from others and enables commands for accessing those links without having to open the address book or SMS to punch in the code for those links.

Figure 1 shows screenshots of the application that enables users to share funny messages in a modulated voice, with their friends. On calling the application, the caller is asked to record any content. This content is then voice modulated to resemble a cat voice and is played back to the user. At this point the user can press the *7 key combination to save this recording and generate a link to this recording. If not, the user can simply continue using the application by recording more content. If the *7 key combination is pressed while the modulated content is being played back or a few seconds after it, the application logs the request to save it, generates a link for it and informs the user that he will get the link via an SMS shortly. The received link allows direct access to the stored voice content bypassing any voice menu navigation.

Figure 1 : Screenshots of the Android application: a) home screen, b) list of recordings, c) single recording selected, d) screen during the phone call.

To access a link, a feature phone user needs to dial the phone number embedded in the link and enter a numeric code (also embedded in the link) when prompted. When using the smartphone application the process is seamless and it requires only a tap from the user to access the link. On link access, the voice application fetches and plays the corresponding content and continues with the regular voice application interaction (i.e., the caller can record and share more such recordings).

This application enables textually illiterate users to make use of device features such as SMS and address book through the use of an intuitively designed mobile app. Also, it enables mobile apps to take the aid of voice interfaces in local language that these users are comfortable with.

A second example scenario where this concept can be applied is in the context of navigating Interactive Voice Response (IVR) menus of various organizations. Services such as Gethuman[1] and Deepdial[2] provide a mechanism to callers to navigate directly to a particular portion of an IVR's deep navigation menu without having to go through the entire process manually. A mobile application on the user's device could either prefetch or obtain via Short Message Service (SMS) on demand, an outline of the target IVR menu. On connecting a call, it could present a visual interface constructed from that IVR meta data. This would enable the caller to navigate the IVR menu through the visual interface rather than having to speak or punch in a digit everytime. Further, with the concept of voice links, the individual menu options of the IVR could be hyperlinked allowing the user to jump directly to a portion rather than wait for the entire navigation to take place sequentially. Augmenting IVR menus with corresponding visual interfaces was also explored by FonePal system [14] where an Instant Messaging (IM) client was used to present the visual interface.

[1] http://gethuman.com
[2] http://deepdial.com

Figure 2 : Researchers demonstrating the visual conversational interface to subjects from target population**.**

3 System Design and Implementation

Even though the processing power and memory available on mobile phones has increased drastically in recent years, yet current speech recognition systems available on mobile devices are not powerful enough to perform sophisticated recognition tasks. Applications such as Siri[3] virtual assistant from Apple and Nina[4] from Nuance make use of cloud based speech recognition software to deliver their service. Since our target users may not have Internet subscription required to utilize such services, our system makes use of ordinary voice channel based interaction with server side voice application deployment and hosting.

To offer these telephony voice services coupled with visual mobile app interface, we make use of applications built over Spoken Web Application Framework [4]. It is a platform that enables easy creation of new voice applications even without the need for any programming skills or even textual literacy. It has been deployed in several applications meant for serving the underprivileged that have language or affordability as a barrier to access or offer information services.

Figure 3 shows a typical deployment infrastructure where a smartphone user without a data connection invokes a mobile application. This mobile application provides its functionality locally and makes a call to a server based voice application deployment. The user could interact with the voice call even as supportive content or icons are displayed on the screen. The server side telephony infrastructure makes use of a Voice Gateway to convert Public Switched Telephone Network (PSTN) calls into voice-over-IP (VoIP). This is then delivered to the application platform which renders Voice XML (VXML) pages to interact with the user.

[3] http://*www.apple.com/ios/**siri/***

[4] http://*www.**nuance**.com/meet-**nina**/index.htm*

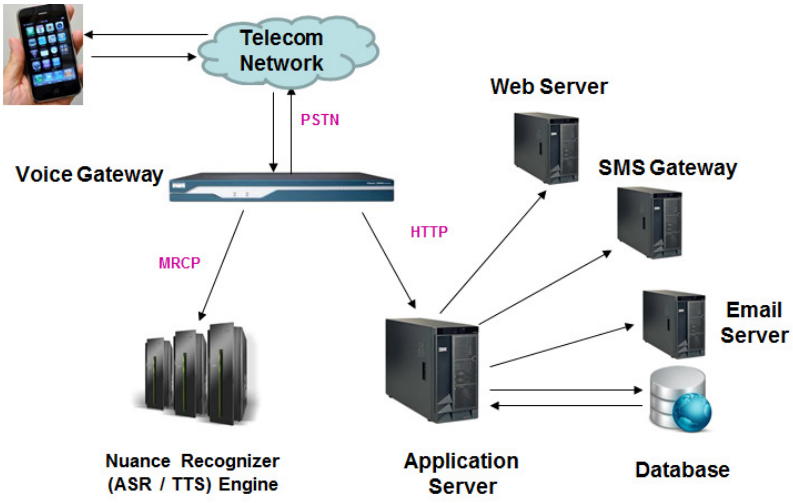

Figure 3 : Infrastructure setup for providing conversational interface to mobile applications.

Nuance Recognizer (or similar system) is used for Speech Recognition through its Automatic Speech Recognition (ASR) function. The database server manages all recordings, content and configuration, while SMS gateway and Email server provide a mechanism to send and/or receives SMS and email respectively. The mobile application can interact with server on any of these channels – voice, SMS, email. A web server is shown since in some scenarios, the stakeholders may like to see the reports and other function through a web based interface.

Since the voice interface is available over ordinary telephone call, it does not requiring any sophisticated device capabilities at the client's end.

4 Related Work

The concept of combining a conversational interface with a visual one has been applied in the context of visualization tools [11]. To create effective visualizations the Articulate system provides a semi-automated visual analytic model coupled with a conversational user interface. Using natural language processing and some heuristics it tries to create a suitable visualization as desired by the user.

Employing conversational interfaces for software agents was employed as early as in the 90s [12] and some of those concepts are visible in intelligent automated assistant systems of today such as CoCo system [6] that uses conversational interface to automate tasks on the web as well as Siri and Nina. They are also being applied to other tasks such as information retrieval [5].

Interplay of a conversational interface with animated personas has been studied by Oviatt et al. [7]. In their study with children, they learnt that children's speech con-

verged with the text-to-speech (TTS) heard from the animated persona. The participating children adapted several acoustic-prosodic features of their speech based upon what they heard from animated personas. Further, children readapted when exposed to new personas.

However, the focus for such conversational systems primarily has been to enable multiple channels of communication with a software system or to study the influence of multiple channels of communications. Recent popular tools such as Siri or Nina, focus primarily on the conversational aspect of speech based interface and do not attempt to leverage the conversation interface with a corresponding visual interface.

This paper, on the other hand, proposed an interaction modality in which applications leverage conversational interfaces side by side and intermixed with visual interfaces to help several users overcome their accessibility problem. These could be textually-illiterate people or older adults unable to read or type properly especially on mobile phones. An entertainment application with such a visual conversation interface has been presented in [9].

5 Discussion

Due to the increased proliferation of smartphones, it is only a logical step to leverage the power of visual interfaces and provide illiterates with additional means of communication. Friscira et al. [3] used a smartphone application to augment SMS messages with icons that let illiterate users make use of basic SMS messaging functionality. This marks just the beginning of a range of new possibilities that open up for researchers and practitioners. As we found out in a study focusing on the aspect of sharing information [9], smartphone applications can contribute to the understanding of interactive voice application services and facilitate the interaction with such voice applications.

The example content sharing implementation we presented in this paper could be extended to allow for sharing to multiple contacts simultaneously by simply clicking on pictures of contacts stored in the address book. This would not only simplify sharing, it would also save time and therefore money for the caller of such a service, and make the process less error-prone by removing the need to enter all recipients' numbers via phone keyboard or voice input.

An area of growing interest in research in developing countries is that of job opportunities for the underprivileged [8, 13]. Browsing such jobs by voice navigation only is cumbersome and time-consuming, eventually reducing effectiveness and success of such applications. Not only does this make such services more useful, the visualization might also contribute to the understanding of the hierarchical structure of voice systems.

6 Conclusion

We presented a new approach of creating visual conversational interfaces that utilize the voice interface of telephony voice applications along with rich graphical interfaces of smartphones to help low literate users overcome their handicap. While providing this, the application does not assume the presence of a data connection on the phone. In the presence of a data connection, these interfaces can be made a lot richer and capable than what is possible with just voice connectivity and local processing power. This mode of interaction (with or without data connection) is also applicable to other users with accessibility challenges such as older adults. As mobile phone platforms become more powerful to host voice applications locally or Internet usage charges become affordable for this population, voice applications could be much more seamlessly integrated with mobile applications similar to emerging voice-on-cloud based mobile assistant applications of today.

References

1. Agarwal, S., Kumar, A., Nanavati, A.A., and Rajput, N. Content creation and dissemination by-and-for users in rural areas. In *Proceedings of International Conference on Information and Communication Technologies and Development (ICTD), Doha, Qatar, 2009*.
2. Datta, A.. "New low-cost smartphone from Micromax". *The Hindu Business Line*. *http://www.thehindubusinessline.com/industry-and-economy/info-tech/new-lowcost-smartphone-from-micromax/article3357066.ece. Last Accessed March 2013*.
3. Friscira, E., Knoche, H., and Huang, J. Getting in touch with text: designing a mobile phone application for illiterate users to harness SMS. In *Proceedings of ACM Second Annual Symposium on Computing for Development (DEV), Atlanta, Georgia, 2012*.
4. Kumar, A. Agarwal, S., Manwani, P., "The spoken web application framework: user generated content and service creation through low-end mobiles", In *Proceedings of the International Cross Disciplinary Conference on Web Accessibility (W4A), Raleigh, USA, 2010.*
5. Lau, T., Cerruti, J., Dixon, M., Nichols, J., "Towards conversational interfaces to web applications", In *Proceedings of Fifth Workshop on Human-Computer Interaction and Information Retrieval (HCIR), Mountain View, CA, October 2011.*
6. Lau, T., Cerruti, J., Manzato, G., Bengualid, M., Bigham, J. P., Nichols, J., "A Conversational Interface to Web Automation", In *Proceedings of 23rd ACM Symposium on User Interface Software and Technology (UIST), NY, USA, 2010.*
7. Oviatt, S., Darves, C. and Coulston, R., "Toward adaptive conversational interfaces: Modeling speech convergence with animated personas", *In ACM Transac-*

tion on Computer-Human Interaction, September 2004, Volume 11, Number 3, pp 300—328.

8. Raza, A.A., Haw, F., Tariq, Z., Pervaiz, M., Razaq, S., Saif, U., and Rosenfeld, R. "Spread and sustainability: the geography and economics of speech-based services." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 2013.*
9. Remy, C., Agarwal, S., Kumar, A., Srivastava S., "Supporting Voice Content Sharing among Underprivileged People in Urban India", To appear in *Proceedings of 14th IFIP TC13 Conference on Human-Computer Interaction, Cape Town, South Africa, September 2013.*
10. Sambasivan, N., Weber, J.S., and Cutrell, E. Designing a Phone Broadcasting System for Urban Sex Workers in India. In *Proceedings of ACM CHI Conference on Human Factors in Computing Systems (CHI), Vancouver, British Columbia, 2011*.
11. Sun, Y., Leigh, J., Johnson, A. E., Chau, D., "Articulate: a Conversational Interface for Visual Analytics", *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, *Atlantic City, New Jersey, 2009*.
12. Trower, T., "Creating Conversational Interfaces for Interactive Software Agents", In *Tutorial Proceedings of ACM CHI Conference on Human Factors in Computing Systems (CHI), 1997.*
13. White, J., Duggirala, M., Srivastava, S., and Kummamuru, K. Designing a Voice-based Employment Exchange for Rural India. In *Proceedings of International Conference on Information and Communication Technologies and Development (ICTD), Atlanta, Georgia, 2012.*
14. Yin, M., Zhai, S., "The benefits of augmenting telephone voice menu navigation with visual browsing and search" In *Proceedings of ACM conference on human factors in computing systems (CHI) 2006.*