Voice Interfaces for Visually Impaired and Low-Literate Communities in the Developing World

Aditya Vashistha
adityav@cs.washington.edu

Sam Sudar
sudars@cs.washington.edu

INTRODUCTION

According to the World Health Organization, 285 Million people in the world are visually impaired (VI) and 90% of them live in developing countries [26]. VI people in developing countries have limited access to assistive technologies because of various socio-economic, device-related and network related constraints. In India, only 5% of all mobile phone subscribers own a smartphone [18]. The remaining 95% use a basic or feature phone;these typically have only one in-built assistive tool for VI people, a speaking clock. Lack of screen readers on these devices and tactile interfaces force VI users to access information and provide data using voice based communication on their phones.

The literacy rate in most developing countries is quite low. For example, 26% people in India are illiterate and the rate of functional illiteracy is surprisingly high. Even in a developed country like US where the claimed literacy rate is 99%, around 20% of adults are functionally illiterate. Thus, text based interfaces are out of reach of sighted low-literate communities as well. In India, there are millions of speakers of tribal languages like Kui, Kurukh, Gondi. However, the font support for many tribal languages is still non-existent. Thus many literate tribal people also use voice medium for producing and accessing information. The limited usage of computing devices like smartphone, tablet and desktop limits adoption of graphical user interfaces designed for low-literate low-income communities. As a result, these communities also use their basic or feature phones to access informational services on the voice channel.

In recent years, various researchers and practitioners have used Interactive Voice Response (IVR) systems to provide health information, citizen journalism, educational services, feedback channels, data collection systems, social media platform etc. for low-literate marginalized communities. IVR systems have also been used as a means to provide information to and collect data from VI people with low socio-economic status. Though there are a lot of success stories of organizations appropriating IVR systems for societal development, there are a few challenges that impede the sustainable usage of such systems. The first challenge is the high cost of calls for accessing information. The second challenge is inefficient audio content management. Hierarchical call tress are used in IVR system to structure the information much like a menu in graphical user interfaces. However, auditory interfaces are quite different from textual and graphical interfaces. Voice interfaces force users to access information sequentially, which results in frustration, high cost, and low-usability. Moreover, searching and indexing audio recordings on an IVR system are non-trivial operations because of un-availability of high efficacy speech recognition engines for resource constrained languages spoken in developing countries. Because of the above reasons, concentrated efforts are needed for accurate and swift retrieval of information to decrease the time required by callers and to increase usability.

Our aim is to understand ways in which voice interfaces on an IVR platform can be organized so as to optimize information seeking time, information absorption and call duration. Through this project, we are evaluating expressiveness and effectiveness of various voice interface designs for providing information to low-literate and VI communities. Our research will answer the following research questions:

- Is hierarchical voice menu better than a linear list?

- How many number of items in a linear list are acceptable?

- Is shallow hierarchy better than deep hierarchy?

- How deep or shallow the hierarchy should be?

- How abstraction capabilities, literacy level and prior exposure to IVR impacts different designs of voice user interfaces?

RELATED WORK

Interfaces for Low-Literate People
A lot of work has been done in designing user interfaces for low-literate low-income communities in the developing world. Most of the work explore usage of graphical interfaces [8, 15, 16, 7] and auditory interfaces [6, 1, 20] for collecting data and providing information to marginalized low-literate communities. Various researchers have explored challenges faced by literate users while accessing hierarchical menus [2, 4]. However, the most relevant work to our project is done by Medhi et. al. where they presented how limited education affect hierarchical user interface navigation on desktops and mobile phones [14, 17]. Medhi measured the textual literacy level, and abstract reasoning level for low-income low-literate people in India. The participants were then asked to

find items from graphical user interfaces consisting of a linear list, shallow, and deep hierarchy. Medhi found that low-income low-literate communities performed best when navigating a linear structure on both desktops as well as mobile phones.

Interfaces for VI People

Although VI people can use screen readers to access information on the Internet, this requires a computer/smartphone, expensive screen reader software, and an Internet connection which is beyond the reach of the majority of the low-income VI people in the developing world [13]. Our parallel research (under review in MobileHCI 2014) has demonstrated that many low-income VI people do not use screen reader software but use alternate information and communication tools for accessing information. Many participants of the study indicated using IVR systems for accessing educational and entertainment content. Some work has also been done to augment or replace visual menus on desktop, mobile phone and other electronic gadgets like music player for improved usability and accessibility [27, 28, 11]. However, no work has been done yet to understand the trade-offs between using list and different kind of hierarchical interfaces to organize information in IVR systems.

IVR Design Guidelines

Various researchers have provided design guidelines to create effective IVR systems and decrease cognitive load on the callers [21, 23, 19, 9]. However, a few guidelines focus on the breadth and depth of navigation IVR trees. Various researchers have investigated deep and shallow hierarchies in IVR systems for high-literate users in the developed world [10, 25, 5]. For instance, Commarford et.al. compared deep and shallow IVR hierarchies and concluded that deep hierarchies take longer to navigate and have high cognitive load [5]. However, the participants of the study were undergraduate students with at least 3-months of email experience and were based out in United States.

Researchers have also studies various facets of IVR systems to improve information consumption and production by low-income low-literate population in the developing world. Chakraborty et.al. has provided design guidelines for creating IVR surveys for low-literate low-income marginalized communities in India [3]. Kummamuru et.al. has analyzed the relationship between cognitive load and various navigational factors like breadth, depth, type of information, and numerical order [12]. They asked semi-literate participants in India to find job information from a synthetic IVR job repository. Each participant had to complete eight tasks. The tasks were designed such that the impact of depth, breadth and other parameters can be measured on the cognitive load. They found that the time to choose right option decreases as the breadth and depth increases. This is because practice and learning effects impact the time taken to navigate the depth. Moreover, cognitive load is high while listening to the first few options in the breadth and thus, the navigation time is lesser for later options in the breadth. The biggest weakness of the paper is the absence of direct comparisons between linear lists, shallow hierarchies and deep hierarchies.

To the best of our knowledge, our work will be the first to directly compare linear list, shallow hierarchy and deep hierarchy for maximizing information absorption by low-literate, low-income communities and visually impaired people in India.

DESIGN

We have conducted 3 between-subjects experiment design to compare three designs of IVR interfaces: linear list, shallow hierarchical interface, and deep hierarchical interface. We have used *Household items* dataset used by researchers for comparing performance of low-literate communities on list, shallow and deep hierarchy graphical user interfaces [14, 17]. We chose the same dataset in order to directly compare our finding with the findings for graphical visualization. The list voice user interface is a sequential list of the household items. Shallow hierarchical voice user interface has a branching factor of five and the depth of the tree is two (see figure 1). The branching factor for deep hierarchical interface is up to four, and the depth of the tree is four (see figure 2). We have implemented the design structures in VoiceXML using Voxeo Prophecy and IVR Junction [24]. We have also logged user's interaction with the IVR system while performing tasks to review the actions user took while performing the tasks.
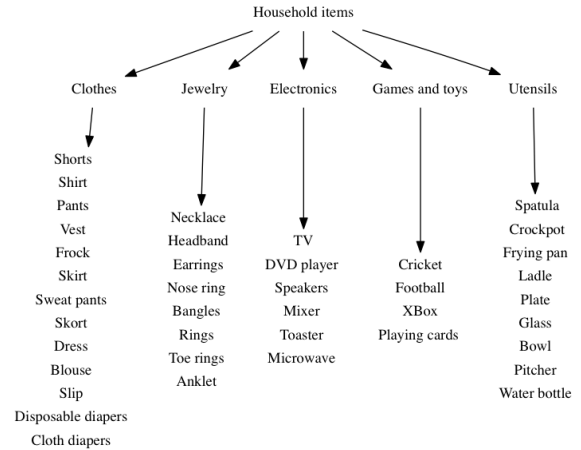

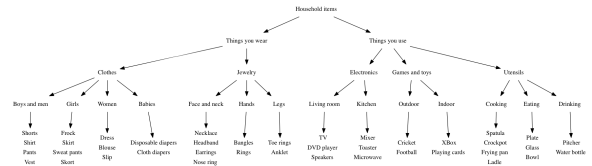**Figure 1. Shallow hierarchy arrangement of household items**

Figure 2. Deep hierarchy arrangement of household items

Participants were randomly divided into 3 groups where each group contained equal number of participants. Each participant performed five tasks, one after another. Each task consisted of selecting a target item from the assigned design structure. All participants searched for five target items (Shirt, Blouse, Rings, Football, and Plate) in the same order. We did not randomize the tasks for participants to ensure that the learning effects for each participant is same. We selected

two easy tasks (*find Shirt, Blouse*), one task of intermediate difficulty (*find Rings*), and two complex tasks (*find Football, Plate*). The comparative complexity of performing a particular task (say *find Rings*) across different hierarchies is same as the number of hops required to reach the target item (excluding navigational prompts) in a brute force manner is same for each design structure.

EVALUATION

We recruited 18 participants (12 Male, 6 Female) to evaluate three design structure. All the participants were either students or employee of the Computer Science department of the University of Washington. The average age of the participants was 26.7 years (max=52, min=19) and the average education level was masters. Sixteen participants were heavy users and two were moderate users of IVR systems. None of the participant was visually impaired.

Before beginning the task, participants were shown a demo task where they had to find an item from the randomly assigned design structure. We used a different dataset (animal and birds) to illustrate the task and design structure to participants. Participants were allowed to perform demo task at most two times before starting the actual experimental task. We also observed participants while they were performing the task for understanding the challenges they faced while navigating design structures. After completing the task, we interviewed participants to collect demographic information. Participants also completed a short survey where they rated their performance, frustration, and physical, temporal and mental efforts exuded while performing the tasks.

The following criteria were used for evaluating information absorption and swiftness in navigating the design structures:

- Task completion rate

- Task correctness

- Average task completion time

TO DO

- Tell the overall correctness and time taken for List, Deep and Shallow.

- Graphs: Box plots of time taken for all tasks in List, Shallow and Deep. Graphs created by Aditya in poster with error bars. Graphs plotted by Sam depicting results of NASA study.

- explain the graphs

ANALYSIS AND RECOMMENDATION
TO DO

- What are the salient observations?

- What did you find out - Try to rationally explain the findings

- how is it different from the hierarchical comparisons done for graphical interfaces for US population? Why? (Read literature survey, some people have done the same for graphical interfaces where test subjects were university students)

- tell problems in our hierarchy - ambiguity in Rings, Shirt, Pants

- tell participants frustration, confusion while navigating hierarchies

- tell how two participants while navigating deep hierarchy chose a wrong path but then figured it out and selected the right option eventually.

Sam's recommendations

- Lists have a wide distribution of completion times and are frustrating to users

- Hierarchies provide the opportunity for performance feedback while completing the task

- Real world hierarchies will have classification ambiguity. Shallow hierarchies limit the impact of hierarchy decisions and keep the choices intuitive for a broader group of users

- Shallow hierarchies outperform the other two categories in frustration and performance

- Use a shallow hierarchy

Aditya's recommendations

- List of size 9 is better than deep or shallow of 9 items because time taken to traverse a list of this size is less than navigating hierarchical structure

- Within deep and shallow hierarchies, recommend using the list of items less than 9 items at the last level

FUTURE WORK

We will recruit visually impaired people, and low-literate people from India to participate in our study. We will work with a local organization in Madhya Pradesh to get access to the marginalized communities. We will measure the recruited participants on three variables:

- Literacy level (low, medium/high)

- Cognitive skills (low, medium/high)

- Familiarity to IVR usage (limited, medium/high)

Previous research has indicated strong correlation between literacy level and cognitive skills [22, 14]. Thus, we will group participants in 4 groups on the basis of the level of cognitive skills and familiarity with IVR:

- People with low cognitive skills and limited IVR experience

- People with low cognitive skills and medium/high IVR experience

- People with medium/high cognitive skills and limited IVR experience

- People with medium/high cognitive skills and medium/high IVR experience

We will also measure **(TO DO)**: **Add Sam's point here**

- Impact of hierarchy appropriateness on usability

- Items that are more easily categorizable (e.g. animal species broken by taxa)

- Greater than 40 items–when does a shallow hierarchy become insufficient?

- Is shallow preferable to deep simply because there were fewer questions asked? What if shallow was a single hop but had 15 intermediate questions? Would it still be found preferable? Perhaps users don't like having to stay alert and make decision in the hierarchy

- What is the ideal number of items in a list at any given point? E.g. some users seemed confused by having to press two buttons on the keypad

- What is the ideal number of questions in intermediary levels? E.g. perhaps a deep binary tree would out perform our deep tree with a variable branching factor

- Compare different shallow and deep hierarchies with proportional branching factor and depth of the tree. In our experiment, the ratio of depth was 2: 4 for shallow and deep. It would be great proportions like 4: 8 and 3:6 to understand which is better among deep and shallow.

CONCLUSION
TO DO

WRITE THE CONCLUSION

MISCELLANEOUS TO-DOS
- GitHub page for the project

- Upload paper and other artifacts in the repository.

REFERENCES
1. Agarwal, S. K., Kumar, A., Nanavati, A. A., and Rajput, N. User-Generated Content Creation and Dissemination in Rural Areas. *Information Technologies and International Development 6*, 2 (2010), 21–37.

2. Allen, R. Cognitive Factors in the Use of Menus and Trees: An Experiment. *Journal on Selected Areas in Communications SAC 1*, 2 (1983), 333–336.

3. Chakraborty, D., Medhi, I., Cutrell, E., and Thies, W. Man versus machine: evaluating IVR versus a live operator for phone surveys in India. In *DEV* (New York, New York, USA, Jan. 2013).

4. Chaudry, B. M., Connelly, K. H., Siek, K. a., and Welch, J. L. Mobile interface design for low-literacy populations. *Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI '12* (2012), 91.

5. Commarford, P. M., Lewis, J. R., Smither, J. A.-A., and Gentzler, M. D. A Comparison of Broad Versus Deep Auditory Menu Structures. *Human Factors: The Journal of the Human Factors and Ergonomics Society 50*, 1 (Feb. 2008), 77–89.

6. Cuendet, S., Medhi, I., Bali, K., and Cutrell, E. VideoKheti: Making video content accessible to low-literate and novice users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13* (2013), 2833–2842.

7. Ghosh, K., Parikh, T. S., and Chavan, A. L. Design considerations for a financial management system for rural, semi-literate users. In *CHI '03 extended abstracts on Human factors in computer systems - CHI '03*, ACM Press (New York, New York, USA, 2003), 824.

8. Grisedale, S., Graves, M., and Grunsteidl, A. Designing a Graphical User Interface for Healthcare Workers in Rural India. In *Proceedings of the 1997 ACM SIGCHI Conference on Human Factors in Computing Systems* (1997), 471–478.

9. Halstead-Nussloch, R. The design of phone-based interfaces for consumers. *CHI '89 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (1989), 347–352.

10. Huguenard, B., Lerch, F., and Junker, B. Working-memory failure in phone-based interaction. *ACM Transactions on Computer-Human Interaction 4*, 2 (June 1997), 67–102.

11. Jeon, M., Walker, B. N., and Srivastava, A. Spindex (Speech Index) Enhances Menus on Touch Screen Devices with Tapping, Wheeling, and Flicking. *ACM Transactions on Computer-Human Interaction 19*, 2 (July 2012), 1–27.

12. Kummamuru, K., Jujjuru, A., and Duggirala, M. Generating facets for phone-based navigation of structured data. *Proceedings of the 21st ACM international conference on Information and knowledge management* (2012), 1283.

13. McCarthy, T., Pal, J., Marballi, T., and Cutrell, E. An analysis of screen reader use in India. In *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development - ICTD '12*, ACM Press (New York, New York, USA, 2012), 149.

14. Medhi, I., Lakshmanan, M., Toyama, K., and Cutrell, E. Some evidence for the impact of limited education on hierarchical user interface navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, ACM Press (New York, New York, USA, 2013), 2813.

15. Medhi, I., Patnaik, S., Brunskill, E., Gautama, S. N., Thies, W., and Toyama, K. Designing mobile interfaces for novice and low-literacy users. *ACM Transactions on Computer-Human Interaction 18*, 1 (Apr. 2011), 1–28.

16. Medhi, I., Sagar, A., and Toyama, K. Text-Free User Interfaces for Illiterate and Semiliterate Users. *Information Technologies and International Development 4*, 1 (2008), 37–50.

17. Medhi, I., Toyama, K., Joshi, A., Athavankar, U., and Cutrell, E. A comparison of list vs. hierarchical UIs on mobile phones for non-literate users. In *Human-Computer Interaction INTERACT 2013 in Lecture Notes in Computer Science* (2013).

18. Meeker, M., and Wu, L. 2013 Internet Trends.

19. MSDN. Best Practices in Designing Speech User Interfaces.

20. Mudliar, P., Donner, J., and Thies, W. Swara : A Voice Forum for Citizen Journalism in Rural India. *Information Technologies and International Development 9*, 2 (2013), 65–79.

21. Ndwe, T., Dlodlo, M., and Mashao, D. Usability engineering of an interactive voice response system in a diverse-cultured and multilingual setting. *Innovative Techniques in Instruction Technology, E-learning, E-assessment, and Education* (2008), 554–559.

22. Reis, a., Petersson, K. M., Castro-Caldas, a., and Ingvar, M. Formal schooling influences two- but not three-dimensional naming skills. *Brain and cognition 47*, 3 (Dec. 2001), 397–411.

23. Suhm, B. Ivr Usability Engineering Using Guidelines And Analyses Of End-to-End Calls. In *Signals and Communication Technology*. 2008, ch. Human Fact.

24. Vashistha, A., and Thies, W. IVR Junction.

25. Virzi, R. A., and Huitema, J. S. Telephone-Based Menus: Evidence that Broader is Better than Deeper. In *Proceedings of the Human Factors and Ergonomics Society* (1997), 315–319.

26. WHO. Factsheet on Visual impairment and blindness, 2013.

27. Yalla, P., and Walker, B. Advanced auditory menus: design and evaluation of auditory scroll bars. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility* (2008), 105–112.

28. Zhao, S., and Dragicevic, P. Earpod: eyes-free menu selection using touch input and reactive audio feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2007).