Assessing Designs of Interactive Voice Response Systems for Better Usability

Siddhartha Asthana, Pushpendra Singh, and Amarjeet Singh

Indraprastha Institute of Information Technology
{siddharthaa,psingh,amarjeet}@iiitd.ac.in

Abstract. Interactive Voice Response Systems (IVR) have emerged as a popular medium to access information over phones. Despite the low usability of IVR systems, they are widely used by commercial organizations due to high reach of phones. Several studies have focused on improving the usability and design of IVR systems. An IVR can be designed in several ways which can have one or more features like touch-tone, speech recognition, content searching etc. However, selecting an appropriate design requires comparison of different designs. In this paper, we propose an information space with three dimensions to study the usability of IVR design as an Information System. We study two different IVR designs - real world deployment and controlled experiment. We further compare these with the traditional IVR design over the proposed dimensions of Information space.

Keywords: IVR, Information space, usability.

1 Introduction

Interactive Voice Response (IVR) technology is used for accessing information via phones. IVR systems automate the call handling and are used in customer care, call routing etc. and also as a medium for information dissemination [10]. High reach of telecommunication services and the need to automate information exchange and communication has made it imperative for the commercial organizations to use IVR systems.

Although IVR systems have been in use for some time, they are often considered as frustrating and time consuming [1]. The usability issues in IVR have drawn attention of the research community. Various improvements in design have been proposed. In particular, interface design, especially navigation [7,4], has received a lot of attention from the research community. Several system designs have been proposed which are specific either to an ethnic groups like illiterate [13], rural [9] etc. or to an application scenario like browsing web pages on voice interface, accessing health information on phones [14], etc.

A design which is focused on improving one aspect of usability may adversely affect the other. Prior studies suggest that an appropriate system design is highly contextual which involves knowledge of target user base and application scenario.

A. Marcus (Ed.): DUXU/HCII 2013, Part I, LNCS 8012, pp. 183–192, 2013.

Thus, for a known scenario and a user base, selecting an appropriate design requires comparing several designs on different usability aspects and understanding trade-offs among the usability aspects of each design. Though now automated tools [6] exist that allow testing of an IVR system before actual deployment, it is important to have design parameters which help in analyzing the usability of IVR systems as Information Systems.

To define design parameters that measure usability of an information system requires focus on information delivery mechanisms. The information delivery mechanism has access to a repository with information content that the system is expected to deliver. This amount of information in the repository can be termed as the information delivery capacity of the system. The next step requires gathering information need of the user which may be either selecting appropriate menu and sub-menu options or query formulation through speech utterances. After gathering the requirement of users information needs, the system must respond to the user with appropriate information content from the repository. Hence, the usability of information system depends upon the three factors:

- Information capacity of the system
- Time taken in expressing the information need of the user to the system
- The quality of information delivered to the user

Based on this, we propose an information space to measure the usability of an IVR system as an information system. The next section describes the information space in detail and explores the existing literature across the three dimensions of it.

2 Information Space

Perugini et al. [11] have studied different aspects (e.g. Interaction style, input modality) of IVR system's design through a 3-Dimensional conceptual design space. The dimensions of the design space in their work is focused on exploring the alternatives for different design aspects of IVR rather than analyzing its usability. In our work, we propose the concept of an information space to study usability aspects of different IVR designs. Information space is a 3-Dimensional space with each dimension measuring usability aspect of an information system as shown in Figure 1. The three dimensions of information space are as follows:

- **Information navigation time:** It refers to time spent on navigation to access information in IVR systems. Lesser the navigation time the quicker a user can access information.
- **Information relevance:** It refers to the relevance of information provided by the system as measured through standard metrics like precision and recall. Traditional IVR systems, provide relevant information as information content is prepared manually. However, automated technique as practiced by the upcoming system may result in providing irrelevant information.

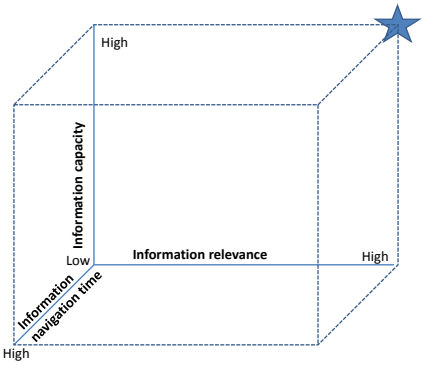

Fig. 1. Information space to measure the usability of information dissemination system

– **Information capacity:** It refers to the breadth of information content provided by the system for a variety of user queries. Traditional IVR systems provide limited information as browsing and navigating through large content is difficult on audio interfaces.

Ideally, a usable system should provide relevant information from large content and should have low navigation-time as shown by "⋆" in Figure 1. Prior work has shown to improve the usability of IVR in one or more dimensions of information space. Skip and Scan [12], ZAP and ZOOM [8] have tried to improve on navigation time. These techniques allow caller easily navigate back and forth through menus or jump directly to another location using shortcuts without first listening to all of the prompts for a particular menu. IVR systems with dynamic rearrangement of menu and use of information retrieval (IR) technique have been proposed to reduce navigation time and to increase the information capacity in voice based system [15,2,4,3]. In the next section we will study two system designs for IVR system that are based on dynamic rearrangement and automatic IR technique in detail through real world deployment and control experiment. Further, we will compare the system usage of IVR based on dynamic rearrangement and automatic IR technique with a traditional IVR system.

3 Experiment Design

We did our experiment in two phases. In the first phase we deployed two IVR systems in the real world and in a second phase, we conducted a controlled experiment.

IVR systems in the first phase were designed, developed and deployed to serve information to the applicants for admission in undergraduate and post-graduate courses at IIIT-Delhi (a state university in India). This deployment helped us to create real world usage of our system [4]. We designed a traditional static IVR system with static menu options. With the intent to reduce navigation time in the

IVR system, we also designed an adaptive system that dynamically rearranges the menu options based on their relative popularity. Details of Traditional and Dynamic IVR is given below:

– Traditional IVR: An IVR system with a static menu deployed in practice to provide information to applicants for admission to undergraduate and post-graduate courses in the institute. Figure 2 shows the various menu options available in IVR system.

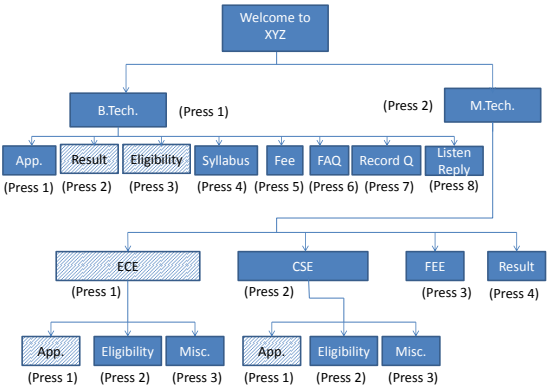

Fig. 2. Menu options in Traditional and Dynamic Rearrangement IVR

– Dynamic IVR: An IVR system deployed in parallel with traditional IVR but the menu sequence gets rearranged based on the relative popularity of menu options among the callers. The goal of this system design is to reduce the navigation time by decreasing the waiting time for the desired option to appear in menu sequence. In this IVR, nodes (menu options) at the same level are rearranged in descending order of the number of times a node has been accessed in the past by the users. The menu items placed at the same level were rearranged automatically without any manual intervention. Hierarchical scheme decides the new order of options based on the historical data of calls made to system.

In the second phase, we deployed an IVR (named IR-IVR) which was aimed to address the inability of IVR system to serve large information content. The menu structure of IVR grows with the increase in information content. Thus, accessing information through a huge number of options in IVR becomes difficult. This is the primary reason for IVR being kept for delivering small amount of information. Thus, to support information access in IVR from a bigger set of information is a challenging task. To overcome this challenge we have used Information retrieval (IR) techniques.

We built an information corpus from the text sources which were used in audio recording of conventional IVR system. Results retrieved by IR component

of IR-IVR was compared against the text played in human recorded voice by the traditional IVR system. To assess the usability of the IR-IVR system, we conducted a controlled user experiment with 16 participants. The information corpus was the official website of IIIT-Delhi.

– IR-IVR: An IVR system using IR technique to automatically generate responses to the user query. The system was designed to achieve high information capacity in IVR system. This IVR was tested under controlled experiment as impact of using IR technique on information relevance was not known and inaccurate or irrelevant information may have had adverse effects in real world deployments. IR-IVR takes the speech input as user query and process this voice data to satisfy the information need of the caller. This allows user to make a free form query. It helped us to do away with menu-based navigation of traditional IVR system.

4 System Usage

Traditional and Dynamic IVR were up from 28^{th} May to 7^{th} July, 2012. In this duration we received 188 calls on both the systems. Calls were audio recorded and users were informed about this at the beginning of each call. We also logged the phone number, received as caller-ID information, with each call. Users of the system accessed various content on the IVR. Number of applicants for undergraduate course were higher than the number of applicants for postgraduate course and it was clearly reflected in system usage as well. Contents of B.Tech (Undergraduate equivalent to BS program) were accessed far more than the M.tech (Postgraduate equivalent to MS program). Among the B.Tech submenu as shown in Figure 2 Application process was all time favorite among the callers. B.Tech result saw a steep rise in number of times it is accessed after the announcement for the results of selected candidate as was expected. It is this dynamic nature of menu sequence that we wanted to capture through our menu rearrangement [4].

In the second phase, we conducted a controlled study in which participants engaged in an informational search task. We had 16 participants in our study. Twelve out of 16 participants were from IIIT-Delhi. These 12 participants had good information about the IIIT-Delhi. The rest 4 participant were from outside the institute and had no or little knowledge about the IIIT-Delhi. All the 16 participants were told to ask 4 queries on IR-IVR about IIIT-Delhi.Accuracy of speech recognizer used in IR-IVR in terms of Word-Error-Rate (WER) came out to be 0.267.We also found that WebScore of our system came out to be 59.8%. WebScore measures semantic quality of the recognizer. A higher value indicates better performance. It is calculated by measuring the number of times the search result as queried by the recognition hypothesis varies from the search result as queried by a human transcription. The average response time of IR-IVR for the queries came out to be 1.257 seconds. We measure the response time of IR-IVR as the time taken by it in fetching the results from information corpus. We also found that the average response length for words in the query response.

The average response length for our experiment came out to be 98.46 words. IR-IVR allowed user to make free form of query e.g. "What is the last date of application form submission". The query length is the length of query text in words. The average query length for our experiment came out to be 7.96 words. The longest query had 13 words in it. Higher values of query length shows the system was able to handle complex queries both in terms of recognition and retrieval. Table 1 represents system usage of each IVR system. We define user query as a expression of information need by the user response to which system gives specific information related to IIIT-Delhi admission. In menu based IVR (i.e. Traditional and dynamic) selection of menu and sub-menu options for which IVR system gives a piece of information is treated as 1 user query. On the other hand, in IR-IVR a user input given by speaking into the system is treated as 1 user query. Further, number of unique callers for menu based IVR is assessed based on number of unique caller-ID (i.e. received as phone number) received by us. In the case of IR-IVR, 16 participants are treated as 16 different caller to the system.

Table 1. Data collection of each IVR system

	Traditional	**Dynamic**	**IR-IVR**
User Queries	174	202	64
Unique callers	127	132	16
User type	All age group, across the India	All age group, across the India	Students in age group of 18-24 with different background
Remarks	1 month deployment for admission process	1 month deployment for admission process	Conducted a control experiment with 16 users

5 Results and Analysis

In this section, we will show that how different designs have performed on each dimension of information space. We will compare each system usability and will try to place them at correct position in information space.

5.1 Navigation

As defined earlier navigation time is the time spent on navigation to access information in IVR systems. In the menu based IVR, it is the time spent on the announcement and selection of menu and submenu options whereas in IR-IVR, it is the time taken by user to speak into the system followed by the time taken by system to recognize this speech. Based on this we calculated the average navigation time of each system as shown in Table 2.

Table 2. Average Navigation time (in seconds) of each system

System	Navigation Time
Traditional	51.59
Dynamic	44.41
IR	5.2

5.2 Relevance

Relevance of information delivered to the user is a subjective decision and requires human intelligence to assess it correctly. Traditional IVR and its contents were prepared manually. Hence they are assumed to be highly relevant. We are giving traditional IVR a normalized score of 1 (or 100% relevant information) and other system will be evaluated in comparison to this score for calculating their information relevance.

Similarly the content of Dynamically rearranging IVR was also prepared manually but they try to assess the information need of the user based on past system usage. This creates a chance for system to make error with some caller. We have found that repeated caller (i.e., the callers who had called to the system before) has tendency to select menu options based on their past interaction with the system. They often select a menu option before it is announced by IVR. But because menu changed from their last interaction to current interaction they end up in selecting wrong menu options. This effects the information relevance as the information needed by user is different from information delivered to the user. Hence a repeated user who has tendency to select menu ahead of its announcement is likely to get wrong information in dynamically rearranging IVR. In our experiment, Dynamic rearrangement IVR received 70 repeated calls. However, some of the repeated user who behaved like first time caller did not faced the problem of selecting wrong menu options but as soon as they get acquainted with the system they are likely to be more prone to such error. Hence, assuming that system will only provide relevant information to new callers. We evaluate information relevance by evaluating the following expression:

$$NewCaller/TotalCall = 132/202 = 0.65 \quad (1)$$

This expression will tend to normalized score of 1 which shows dynamically rear-ranging IVR will perfect in the situation where system does not have any repeated caller.

In IR-IVR system we asked user to rate the retrieved response from IR-IVR on a of scale of 1 to 5. A score of 1 signifies the response were extremely poor and has no relevance to the query where as a score of 5 signifies the response generated by IR-IVR were extremely good and relevant to the user query. Here user were told to rate the responses they listen on Mega-IVR for its relevance to the query asked. We then computed the accuracy of information retrieval based on user relevance feedback as average score given by user. In our experiment we found that, the value for information retrieval accuracy as assessed by users came out to be 3.46 (out of 5) and a normalized score of 0.69.

5.3 Capacity

Increasing content on menu based IVR increases its menu size. As menu can not grow infinite so it imposes an upper limit on the size content of menu. In our experiment, Traditional IVR and Dynamic rearrangement IVR had 16 different information content out of which 12 are general content and other 4 were specific to the caller. The option specific to the callers are the option provided by system where human intervention is needed to answer the query e.g. user can record their voice message for admission authorities at IIIT-Delhi. Thus our menu based IVR is capable of answering 10 different queries Information content of IR-IVR system were prepared from FAQ available on IIIT-Delhi website. Information corpus had answers of 70 FAQ available on website. Similar to menu based IVR, IR-IVR can not have infinite information content in it. On increasing the information content on IR-IVR may result in low relevance of information as it increase the chance of error. On increasing the size of content a system need to increase the size of vocabulary used for speech recognition. Bigger the vocabulary larger are the chance of error in speech recognition. Similarly IR component of such system is also affected because increasing the size of information corpus increase the chance of selecting irrelevant content for a given user query. Thus for comparing IR-IVR with other systems for information capacity we gave it a normalized score of 1. Based on this relative capacity of menu based IVR which had capacity to answer 12 different user query in comparison to IR-IVR which can answer 70 different query had a information capacity score of

$$Query\ (Traditional\ or\ Dynamic)/Query\ (IR-IVR) = 12/70 = 0.171 \quad (2)$$

5.4 Analysis

Earlier, we have computed the values of each IVR design on different dimensions of Information space. We have measured navigation time in seconds. Information relevance is assumed to be highest when prepared information content is manually prepared. We gave the Traditional IVR maximum score on a scale of 0 to 1 as its content were manually prepared and placed in the IVR. We calculated information relevance of dynamic rearrangement IVR based on the number of time it is expected to deliver relevant information. We found that the dynamic rearrangement IVR may give wrong or irrelevant information to repeated callers. Hence relevant information is guaranteed only for new callers. Based on this we calculate the information relevance of dynamic rearrangement IVR. Information relevance of IR-IVR is directly assessed by the participants. The third dimension of Information space i.e. information capacity is measured in terms of ability to handle different user queries. A menu based IVR which gave different information about the admission process is assume to answer exactly one user query for each information. Similarly, IR-IVR was prepared from 70 FAQ so it is assumed to answer 70 different user queries. In Table 3 we show the score of each IVR system on different dimensions of information space. Improving a usability aspects affects other usability aspects. Figure 3 shows three system in Information

Table 3. Usability score of IVR system on each dimension

	Traditional	Dynamic	IR
Navigation	51.59	44.41	5.2
Relevance	1	0.65	0.69
Capacity	0.17	0.17	1

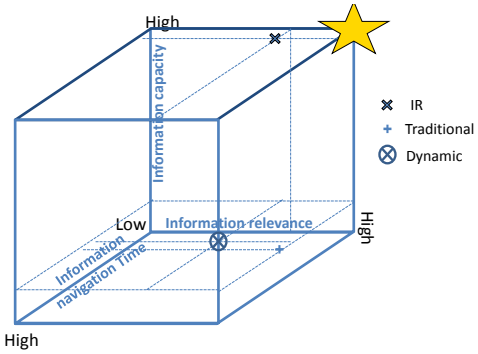

Fig. 3. Comparison of IVR design in Information Space

Space. This suggests that the techniques used for reducing navigation time or increasing the information capacity may result in low information relevance.

6 Conclusions

In this work, we have extended our initial work [5] and proposed and studied information space for identifying the usability of IVR systems. We also evaluated three system designs through real world experiment and controlled lab studies. The system designs were analyzed on different dimension of information space. We showed that improving a design for one usability aspect may affect other usability aspects of information system.

Our proposed information space may help in designing an IVR system based on desired usability on different dimensions.

Acknowledgment. We would like to thank Tata Consultancy Services for supporting this research.

References

1. Consumer Study Finds Overwhelming Dissatisfaction with IVR (June 2011), `http://ibm-news.tmcnet.com/news/2011/06/23/5593905.htm`
2. Asthana, S., Singh, P., Kumaraguru, P., Singh, A., Vinayak, N.: Tring! tring! - an exploration and analysis of interactive voice response systems. In: 4th International Conference on Human Computer Interaction, IndiaHCI 2012 (2012)

3. Asthana, S., Singh, P., Singh, A.: Design and evaluation of adaptive interfaces for ivr systems. In: CHI 2013 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2013. ACM, New York (2013)
4. Asthana, S., Singh, P., Singh, A.: Exploring adverse effects of adaptive voice menu. In: CHI 2013 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2013. ACM, New York (2013)
5. Asthana, S., Singh, P., Singh, A.: Exploring the usability of interactive voice response system's design. In: Proceedings of the 3rd ACM Symposium on Computing for Development, ACM DEV 2013, pp. 36:1–36:2. ACM, New York (2013), http://doi.acm.org/10.1145/2442882.2442924
6. Asthana, S., Singh, P., Singh, A.: Mocktell: Exploring challenges of user emulation in interactive voice response testing. In: 4th ACM/SPEC International Conference on Performance Engineering, ICPE 2013. ACM, New York (2013)
7. Asthana, S., Singh, P., Singh, A.: A usability study of adaptive interfaces for interactive voice response system. In: Proceedings of the 3rd ACM Symposium on Computing for Development, ACM DEV 2013, pp. 34:1–34:2. ACM, New York (2013), http://doi.acm.org/10.1145/2442882.2442922
8. Hornstein, T.: Telephone Voice Interfaces on the Cheap. In: Union Bank of Switzerland, UVK Informatik, Zurich, pp. 134–146 (1994)
9. Patel, N., Agarwal, S., Rajput, N., Nanavati, A., Dave, P., Parikh, T.S.: A comparative study of speech and dialed input voice interfaces in rural India. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, pp. 51–54. ACM (2009)
10. Patel, N., Chittamuru, D., Jain, A., Dave, P., Parikh, T.S.: Avaaj otalo: A field study of an interactive voice forum for small farmers in rural india. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, pp. 733–742. ACM, New York (2010)
11. Perugini, S., Anderson, T.J., Moroney, W.F.: A study of out-of-turn interaction in menu-based, IVR, voicemail systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2007, pp. 961–970. ACM, New York (2007)
12. Resnick, P., Virzi, R.A.: Skip and scan: Cleaning up telephone interface. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1992, pp. 419–426. ACM, New York (1992)
13. Sharma Grover, A., Stewart, O., Lubensky, D.: Designing interactive voice response (IVR) interfaces: localisation for low literacy users. In: Proceedings of Computers and Advanced Technology in Education, CATE 2009, St Thomas, US Virgin Islands (2009)
14. Sherwani, J., Ali, N., Mirza, S., Fatma, A., Memon, Y., Karim, M., Tongia, R., Rosenfeld, R.: Healthline: Speech-based access to health information by low-literate users. In: International Conference on Information and Communication Technologies and Development, ICTD 2007., pp. 1–9. IEEE (December 2007)
15. Sherwani, J., Yu, D., Paek, T., Czerwinski, M., Ju, Y.C., Acero, A.: Voicepedia: Towards speech-based access to unstructured information. In: INTERSPEECH 2007, pp. 146–149 (2007)