Assessing Designs of Interactive Voice Response Systems for Better Usability

K. Gokul[1], P. Ajitha[2], S. Mohanraj[3], Dineshpaul[4], K. Dinesh Kumar[5]

[1,2,3,4] Student, Department of Computer Science and Engineering, K.S.R. College of Engineering, Tiruchengode-637215, India
Email:gokulkrishkbk@gmail.com[1],ajithakrsna16@gmail.com[2],mohanraj13saravanan@gmail.com[3], gdineshpaul@gmail.com[4]

[5]Assistant Professor, Department of Computer Science and Engineering, K.S.R. College of Engineering, Anna University,Tiruchengode-637215, Tamilnadu, India. Email:gokulkbkkrish@gmail.com[5]

Abstract - Interactive Voice Response Systems (IVR) have arose as a current medium to access data over phones. Despite the low usability of IVR systems, they are widely used by profitable organizations due to high reach of phones. Several studies have absorbed on improving the usability and design of IVR systems. An IVR can be designed in several ways which container have one or more features like touch-tone, speech recognition, content penetrating etc. However, selecting an appropriate design needs comparison of dissimilar designs. In this paper, we propose an data space with three dimensions to study the usability of IVR design as an Data System. We study two dissimilar IVR plans real world deployment and controlled experiment. We additional link these with the traditional IVR design over the future dimensions of Data space.

Keywords: IVR, Dataspace, usability, communication, cisco;

I. INTRODUCTION

Interactive Voice Response (IVR) technology is used for retrieving info via phones. IVR systems automate the call handling and are used in customer care, call routing etc. and as a medium for data dissemination [10]. High reach of telecommunication services and the need to industrialize data exchange and communication has made it imperious for the commercial organizations to use IVR systems. Although IVR systems have been in use for some period, they are frequently well-thought-out as exasperating and time consuming [1].

The usability issues in IVR have drawn attention of the research community. Several improvements in design have been future. In particular, interface design, especially navigation [7], [4], has received a lot of attention from the research community. Several system designs have been future which are exact either to an ethnic group like illiterate [13], rural [9] etc. Orto an application scenario like browsing web pages on voice interface, retrieving health data 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 suitable system design is highly contextual which involves data of target user base and application scenario. Thus, for a known scenario and a user base, selecting an suitable design requires comparing several designs on dissimilar usability aspects and understanding trade-offs among the usability aspects of each design.

Though now computerized tools [6] exist that allow testing of an IVR system before actual positioning, It is important to have design parameters which help in analysing the usability of IVR systems as Data Systems. To define plan parameters that quantity usability of an info system requires focus on data delivery mechanisms. The data delivery mechanism has access to a repository with data content that the system is expected to deliver. This amount of data in the repository can be termed as the data delivery capacity of the system. The next step requires gathering data need of the user which may be either selecting appropriate menu or sub-menu options or query formulation through speech utterances. After gathering the requirement of user's data needs, the system must respond to the user with appropriate data content from the repository.

Hence, the usability of data system depends upon the three factors:
– Data capacity of the system
– Period taken in expressing the data need of the user to the system
– The quality of data delivered to the user

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

II. DATA SPACE

Perugino et al. have studied dissimilar 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 are absorbed on exploring the alternatives for dissimilar design aspects of IVR rather than analysing its usability. In our work, we propose

the idea of an info space to study usability aspects of dissimilar IVR designs. Data space is a 3-Dimensional space with each dimension measuring usability aspect of a data system as shown in Figure 1.

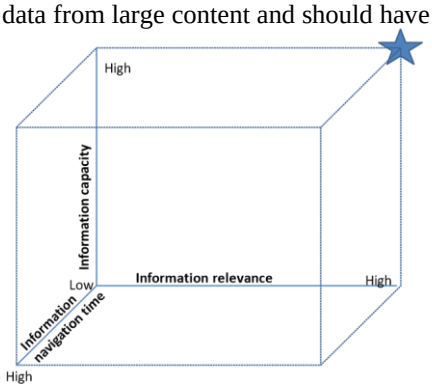
The three dimensions of data space are as follows:

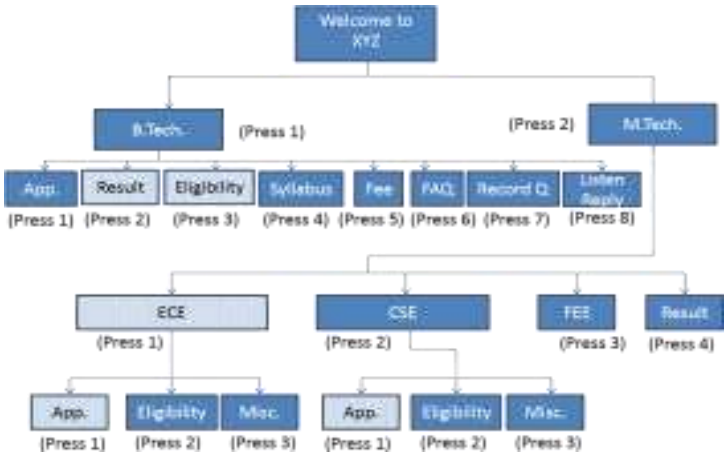
– **Data navigation time:** It refers to time spent on steering to access data in IVR systems. Lesser the navigation time the quicker a user can access data.

– **Data relevance:** It refers to the relevance of data provided by the system as measured through standard metrics like precision and recall. Traditional IVR systems, provide related data as data content is prepared manually. However, computerized technique as practiced by the upcoming system may result in providing irrelevant data.

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

Fig.1. Data space to measure the usability of data dissemination system Ideally, a usable system should provide relevant data 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 data space. Skip and 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 data retrieval (IR) technique have been proposed to reduce navigation time and to increase the data capacity in voice based system [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.

III. EXPERIMENTAL DESIGN

We did our experiment in two phases. In the first phase we organized two IVR systems in the real world and in a second phase, we conducted a measured experiment. IVR systems in the first phase were designed, developed and deployed to serve data 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 intended a traditional static IVR system with static menu options. With the intent to decrease 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 data to applicants for admission to undergraduate and postgraduate 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 reducing the waiting time for the desired option to appear in menu sequence. In this IVR, nodes (menu options) at the same level are reorganized in descendant 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 data content. The menu structure of IVR grows with the increase in data content. Thus, accessing data through

a huge number of options in IVR becomes difficult. This is the primary reason for IVR being kept for delivering small amount of data. Thus, to support data access in IVR from a bigger set of data is a challenging task.

To overcome this challenge, we have used Data retrieval (IR) techniques. We built a data corpus from the text sources which were used in audio recording of conventional IVR system. Results retrieved by IR component of IR-IVR were 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 data corpus was the official website of IIIT-Delhi.

Data relevance

- Low
- High
- High
- High

– IR-IVR: An IVR system using IR technique to inevitably create retorts to the user query. The system was designed to accomplish high data capacity in IVR system. This IVR was tested under controlled experiment as impact of using IR technique on data relevance was not known and mistaken or unrelated data 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 data 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.

IV. SYSTEM USAGE

Traditional and Dynamic IVR were up from 28th May to 7thJuly, 2012. In this period we received 188 calls on both the systems. Calls were audio recorded and users were knowledgeable about this at the beginning of each call. We also logged the phone number, received as caller-ID data, with each call. Users of the system accessed various content on the IVR. Number of applicants for undergraduate course was 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 favourite 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 data search task. We had 16 participants in our study. Twelve out of 16participants were from IIIT-Delhi. These 12 participants had good data about the IIIT-Delhi. The rest 4participant 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 Web Score of our system came out to be 59.8%.Web Score 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 data 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 show the system was able to handle complex queries both in terms of recognition and retrieval. Table 1 represents system usage 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

We define user Table 1. Data collection of each IVR system query as a expression of data need by the user response to which system gives specific data 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 data is treated as1 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, 16participants are treated as 16 dissimilar caller to the system.

V. RESULTS AND ANALYSIS

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

A. Navigation

As defined earlier navigation time is the time spent on navigation to access data 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 IRIVR, 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.

Traditional Dynamic IR-IVR
User Queries
174 202 64
Unique callers
127 132 16

All age group, All age group, Students in age **User type** across the India across the
India group of 18-24 with dissimilar background1-month deployment1-month deployment Conducted a control **Remarks** for admission for admission experiment with process 16 users.

B. Relevance

Relevance of data delivered to the user is a subjective decision and requires human intelligence to assess it correctly. Traditional IVR and its contents System Navigation Time Traditional 51.59
Dynamic 44.41IR 5.2

System	Navigation Time
Traditional	51.59
Dynamic	44.41
IR	5.2

Table 2. Average Navigation time (in seconds) of each system were prepared manually. Hence they are assumed to be highly relevant. We are giving traditional IVR a normalized score of 1 (or 100% relevant data) and other system will be evaluated in comparison to this core for calculating their data relevance. Similarly, the content of Dynamically rearranging IVR was also prepared manually but they try to assess the data 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 data relevance as the data needed by user is dissimilar from data delivered to the user. Hence a repeated user who has tendency to select menu ahead of its announcement is likely to get wrong data 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 data to new callers. We evaluate data relevance by evaluating the following expression:

New Caller/Total Call= 132/202 = 0.65 (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 scale of 1 to 5. A score of 1signifies the responses 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 data retrieval based on user relevance feedback as average score given by user. In our experiment we found that, the value for data retrieval accuracy as assessed by users came out to be 3.46 (out of 5) and a normalized score of 0.69.

C. Capacity

Increasing content on menu based IVR increases its menu size. As menu cannot grow infinite so it imposes an upper limit on the size content of menu. In our experiment, Traditional IVR and Dynamic rearrangement IVR had 16dissimilardatacontents 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 dissimilar queries Data content of IR-IVR system were prepared from FAQ available on IIIT-Delhi website. Data corpus had answers of70 FAQ available on website. Similar to menu based IVR, IRI VR cannot have infinite data content in it. On increasing the data content on IR-IVR may result in low relevance of data as it increases 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 data corpus increase the chance of selecting irrelevant content for a given user query. Thus, for comparing IR-IVR with other systems for data capacity we gave it a normalized score of 1. Based on this relative capacity of menu based IVR which had capacity to answer 12 dissimilar user query in comparison to IR-IVR which can answer 70dissimilarqueries had a data capacity score of *Query* (*Traditional or Dynamic*)*/Query* (*IR − IV R*) = 12/70= 0.171 (2)

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

Table 3. Usability score of IVR system on each dimension Space. This suggests that the techniques used for reducing navigation time or increasing the data capacity may result in low data relevance.

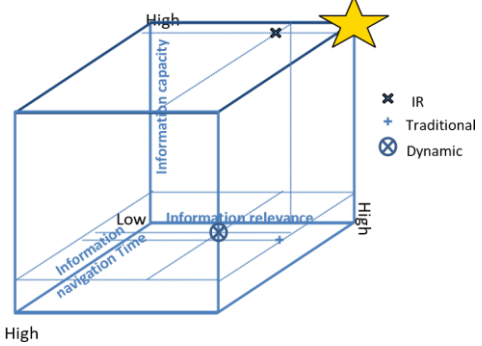


Fig.3. Comparison of IVR design in Data Space

VI. CONCLUSION

In this work, we have extended our initial work [5] and proposed and studied data 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 dissimilar dimension of data space. We showed that improving a design for one usability aspect may affect other usability aspects of data system. Our proposed data space may help n designing an IVR system based on desired usability on dissimilar dimensions.

ACKNOWLEDGMENT

We would like to thank Servion Global Solution for supporting this research.

REFERENCES

[1] Consumer Study Finds Overwhelming Dissatisfaction with IVR(June2011),http://ibmnews.tmcnet.com/news/2011/06/23/5593905.htm

[2] Asthana, S., Singh, P., Kumaraguru, P., Singh, A., Vinayak, N.: Tring!tring! - anexploration and analysis of interactive voice responsesystems. In: 4th International Conference on Human Computer Interaction. IndiaHCI '12 (2012)

[3] Asthana, S., Singh, P., Singh, A.: Design and evaluation of adaptiveinterfaces forever systems. In: CHI '13 Extended Abstracts onHuman Factors in Computing Systems. CHI EA '13, ACM, New York,NY, USA (2013)

[4] Asthana, S., Singh, P., Singh, A.: Exploring adverse effects of adaptive voice menu. In: CHI '13 Extended Abstracts on Human Factors in Computing Systems. CHI EA '13, ACM, New York, NY,USA (2013)

[5] Asthana, S., Singh, P., Singh, A.: Exploring the usability of interactive voice response system's design. In: Proceedings of the3rd ACM Symposium on Computing for Development. pp. 36:1–36:2. ACM DEV '13, ACM, New York, NY, USA (2013), http://doi.acm.org/10.1145/2442882.2442924

[6] Asthana, S., Singh, P., Singh, A.: Mock tell: Exploring challenges of user emulation in interactive voice response testing. In: 4thACM/SPEC International Conference on Performance Engineering .ICPE '13, ACM, New York, NY, USA (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. pp.34:1–34:2. ACM DEV '13, ACM, New York, NY, USA (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. pp. 134–146. Zurich (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 inrural India. In: Proceedings of the 27th international conference on Human factors in computing systems. pp. 51–54. CHI '09, ACM(2009)

[10] Patel, N., Chittamuru, D., Jain, A., Dave, P., Parikh, T.S.: Avaajotalo: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. pp. 733–742. CHI '10,