Investigating Design Strategies for Audio-Visual Interfaces for Emergent Users

Submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY
by

Abhishek Shrivastava
(Roll no. 10413001)

Supervisor
Professor Anirudha Joshi

Industrial Design Centre
INDIAN INSTITUTE OF TECHNOLOGY BOMBAY
(2018)

Approval Sheet

Thesis entitled "**Investigating Design Strategies for Audio-Visual Interfaces for Emergent Users**" by Mr. Abhishek Shrivastava (Roll. No. 10413001) is approved for the degree of Doctor of Philosophy.

Examiners

Supervisor

Chairman

Date: MARCH 09, 2018

Place: I.I.T. Bombay

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources, which have thus not been properly cited, or from whom proper permission has not been taken when needed.

Abhishek Shrivastava
(Name of the student)
10413001
(Roll No.)

Date: MAR 09, 2018

Dedicated to my grandparents and to my mother's lost village "Bijaygadh".

Abstract

Emergent users in developing countries (Devanuj & Joshi, 2013) are fast securing access to Information and Communication Technologies (ICTs) due to massive expansion of mobile phones in developing regions. However, they still need appropriate interfaces to perform better with interactive products. Many regard that audio interfaces like Interactive Voice Response systems (IVRs) can be the best-fit interfaces for emergent users on account of easier deployments, and a strong presence of vocal culture in developing regions (Barnard, E.; Plauche, M.; Davel, 2008; Plauché & Nallasamy, 2007). In addition, IVRs are known to upgrade system usability and task completion rates by preventing users from losing track of interface features, functions and limitations (Pieraccini & Lubensky, 2005). For the same reason, Balentine (Balentine, 2007, p. 102; Balentine & Morgan, 1999, p. 146; Marics & Engelbeck, 1997) and Suhm (Suhm, 2008) insist on using directed dialog IVRs for novice (first-time) users. It seems possible that by using IVRs based interfaces with directed dialog for emergent users; the advantages that exist for first-time users can be transferred to emergent users.

IVRs, however, posit serious usability challenges to their users because of inherent transience and temporality of audio. Tatchell (Tatchell, 1996) finds IVRs based services difficult to learn, easy to forget and confusing. Users must pay an attentive ear to the audio prompts presenting menu choices and system control features. This puts heavy demands on user's working memory. Consequently, user interactions with directed dialog IVRs suffer from 'poor referability' and 'absence of memory aid'. Recent studies with emergent users in focus (Grover, Stewart, & Lubensky, 2009; Joshi et al., 2012) have reconfirmed these usability difficulties. A lesser-explored approach aimed at addressing usability barriers with IVRs is the use of coordinated visuals along with audio prompts (Yin & Zhai, 2005, 2006). Our efforts in the current research work are primarily based on this approach.

An important facet of IVRs is the number of menu items and the depth of the menu hierarchy. Literature is often vocal about the one at the expance of other leading to a gap in the wholistic understanding of IVRs menu structure. In addition, we are not aware of any study with suggestions on menu depth and number of menu items in audio-visual interface design for emergent users. IVRs design guidelines propose a number close to 4 or fewer (Cohen, 2004; Miller, 1956; Suhm, 2008, p. 13); to as much as 9 menu items for interruptible menus (Balentine & Morgan, 1999, p. 159). In a subsequent follow-up by Suhm et al. (Suhm, Freeman, & Getty, 2001), users are seen routing themselves more efficiently using longer menus than shorter menus, provided items in the longer menu are specific and sufficiently detailed. In addition, Commarford (Commarford et al., 2008) demonstrates that broader menu depths are preferred over deeper menu depths for traditional users. In a different study with emergent users, Medhi et al. (I Medhi et al., 2013) favor using multi-page list of items over a four level deep menu hierarchy in graphical user interface. With this background, the current thesis comprises of two major studies aimed at refining our understanding of audio-visual interfaces for emergent users.

In our first study, we hypothesized that audio-visual interfaces would do better than audio-only interfaces in terms of enabling users to exhibit better task completion across different variations of menu depths and menu positions. We hoped that shallow menu depth would continue resulting in higher task completion than deep menu depth not only in case (IVRs based) audio-only interface (Commarford et al., 2008; Suhm et al., 2001) but also for audio- visual interface. Regarding menu position, we hoped that early menu position would result in higher task completion than late menu position across both audio-only and audio-visual interfaces. We organized test prototypes based on “Agricultural commodity market pricing service” with variations in the use of visuals (audio-visual vs. audio-only), menu depths (deep vs. late) and menu positions (early vs. late). With “use of visual” being the between group variable, we organized four different combinations of menu depths and menu positions as test tasks namely, shallow-early (SE), shallow-late (SL), deep-early (DE) and deep-late (DL). Our test tasks typically required emergent users to find information on prices of a specific agricultural commodity for a given market location. Our findings demonstrate that emergent users performed better with significant differences in task success for all task

types with audio-visual interfaces than with audio-only interfaces. Further, emergent users were significantly more successful with deep menus than with shallow menus in audio-visual interfaces. Even in case of audio-only interfaces, this trend continued although the difference was not statistically significant. This is contrary to prior studies which show that shallow menus do better than deeper menus in IVRs with traditional users (Cohen, 2004; Commarford et al., 2008; Suhm et al., 2001) and in graphical user interfaces for emergent users (Medhi, Toyama, Joshi, Athavankar, & Cutrell, 2013). The directedness of audio seems to be helping emergent users navigate hierarchies better than graphical user interfaces. Users were also seen making menu selection independent of the position of the menu items.

In the second study, we wanted to test the relevance of *directedness* of audio and *persistence* of visuals in audio-visual interfaces for emergent users. We are not aware of any substantial research evidence endorsing their combined use. In addition, two other factors motivated us further. *First*, our results from last study have favoured use of audio-visuals interfaces over audio-only interfaces but test tasks assigned to the users were only informational in nature - requiring them to find pricing information for agricultural commodities across different market locations. We questioned if our results were generalizable for transactional tasks, which expected users to make multiple data entries during indermediate tasks like confirming one's identity as in banking services. *Second*, we had only compared audio-visual interfaces with audio-only interfaces in the last study. And, our results were significant for audio-visual interfaces over audio-only interfaces across variations in menu depth and menu positions. We realized that our results would be more relevent if we could also include a graphical user interfaces in our comparisions involving audio-only and audio-visual interfaces. Subsequently we designed three different test prototypes of a banking application, namely audio-only interface (A), a graphical user interface (G) and an audio-visual interface (AV). These were balanced for both directedness of audio and persistence of visuals. Audio prompts were kept the same across A and AV, and hence both of A and AV had identical directedness. While graphical elements of the interface were kept the same across G and AV, and hence both of G and AV had identical persistence. We asked users to perform two different test tasks: informational and transactional. An informational task required them to find out the address of a nearest

ATM bank for a given location. A transactional task required them to transfer a specified amount of money to one of the pre-registered payees. These tests tasks were carried out across all the three test prototypes A, G and AV.

Our results demonstrate that emergent users exhibited significantly greater task success with audio-visual interfaces than with both graphical user interface and audio-only interface for both transactional and informational tasks. Audio-visual interfaces were also faster to use than audio-only interface, and while they were slower than graphical user interfaces, they were better appreciated than graphical user interfaces. We claim that for (first- time use by) emergent users, audio-visual interface offers a good balance between task completion and speed in comparison to audio-only interfaces and audio-visual interfaces. From a design perspective, it implies that if the task times with audio-visual interfaces need to be further reduced, then perhaps designers could provide functionality to switch the directed dialog "on/off", or could decide to provide audio prompts only if the user is unable to proceed. Subsequently with any such functionality, audio-visual interfaces may exhibit improved task time for frequent (returning) users, while staying valuable for first time emergent users.

Overall our research demonstrates that audio-visual interfaces have specific advantages for emergent users over both audio-only interfaces and graphical user interfaces. These utilize the best of both- *directedness* of audio and *persistence* of visuals to have more usable interfaces for emergent users.

Table of contents

List of figures

List of tables

Chapter 1 Introduction

1.1 Motivation

We are living in interesting times in the context of technology-mediated communications. By the moment we reach towards the end of this paragraph, the total number of wireless telephony subscribers in India would be close to 1200 millions (Figure 1-1). Within these the plot for urban subscribers may exhibit some temporary depression. But for rural subscribers, the curve would fairly be a steady positive gradient aiming to touch a mark close to 500 million. Overall across the entire India, there would be more than 90% of people with access to wireless telephony (Figure 1-2). We would have more than 240 million people in India with access to broadband Internet services (Figure 1-3). With these markers of the growth of wireless mobile telephony, India is expected to have more than 340 million smartphone users by the year 2018 (Figure 1-4). Information and Communications Technologies (or ICTs) are spreading in an unprecedented manner across all the quarters of our everyday life. This growth of ICTs is seemingly more and more independent of identities, locations, literacy or occupation and earnings of the subscribers of ICTs. We realize that it sounds immensely encouraging for proponents of using ICTs for the betterment of developing regions and their people, in particular the less materially advantaged members (Walsham, 2017). For people in the discipline, it has taken close to 3 decades for ICTs to come to their current stage of penetration in developing regions. Patra et al. (Patra, Pal, & Nedevschi, 2009) records first few instances of use of ICTs in Scandinavian villages in 1985. Starting with the village of Fjaltring in Denmark, the

ICTs deployments in the form of tele-cottages grew rapidly across Sweden, Norway and Denmark. It was not until a decade later, by 1990s, that the world paid significant emphasis on the role of ICTs in bringing changes in the developing world. International agencies including United Nations and World Bank made deliberate efforts by formulating policies and funding mechanisms aimed at deploying ICTs for development. These initiatives led to an increasing interest of the academic and industry partners alike to carry out sustained efforts in ICT research, development and deployment. Subsequently by mid-1990s we get to see an increased number of field deployment of ICTs in developing regions. International institutions, governments, and non-governmental organizations supported most of these deployments during this time. In parallel computing itself also transformed from command line interfaces to a family of personal computers with graphical user interfaces and from bigger bulkier screens to smaller handheld screens and form factors (Beaudouin-Lafon, 2004). A part of this shift in technology led to re-use of computer donated by corporations and even installation of newer computers in community information centres in developing world (James, 2001; Warschauer, 2004).

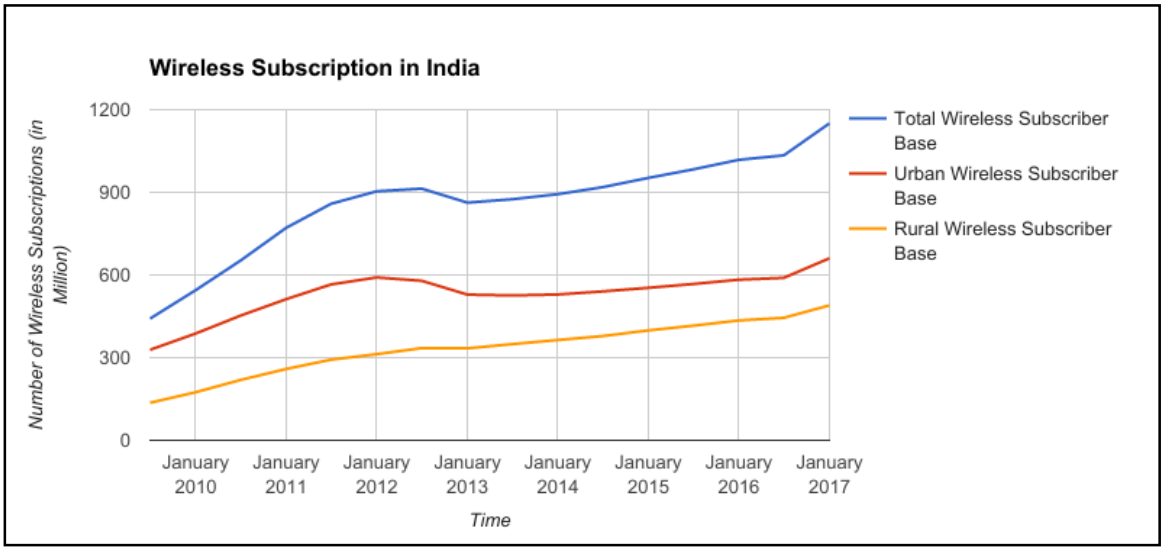


Figure 1-1. Wireless Subscription in India (TRAI, 2017).

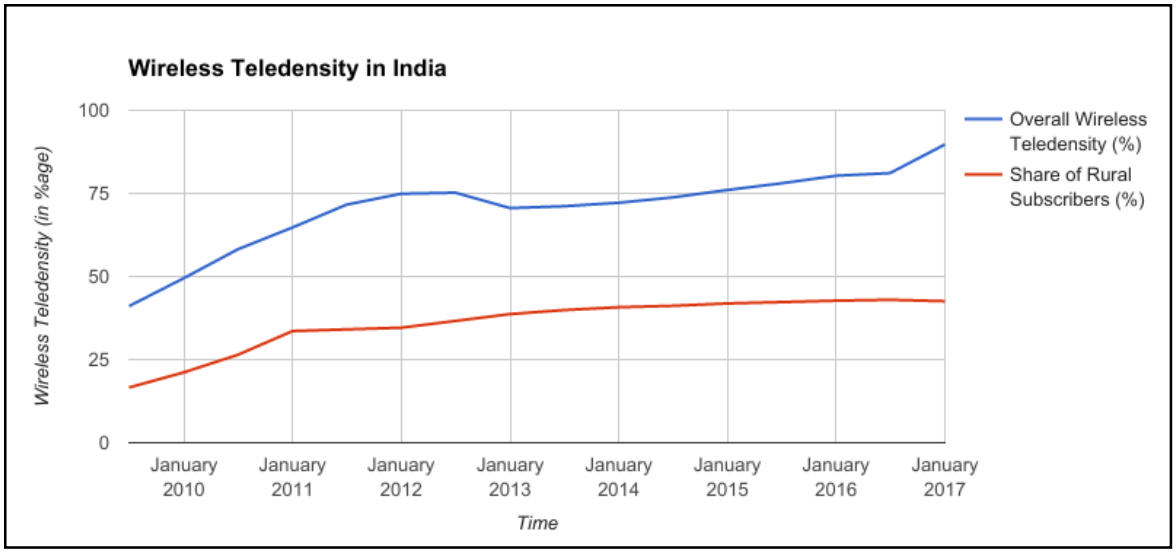


Figure 1-2. Share of rural subscribers (in %) in total wireless teledensity in India (TRAI, 2017).

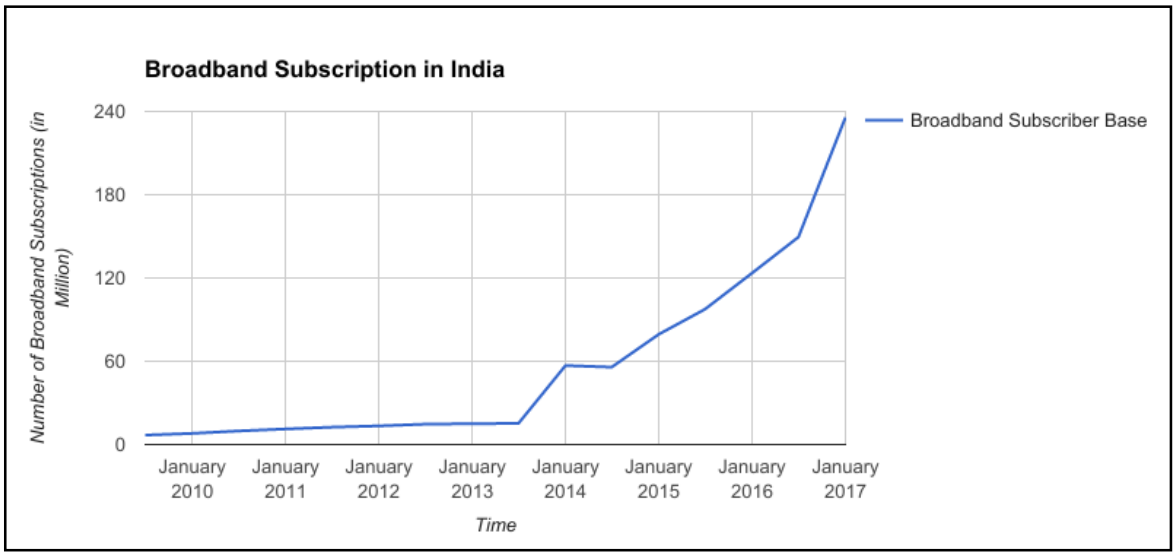


Figure 1-3. Broadband Subscription in India (TRAI, 2017).

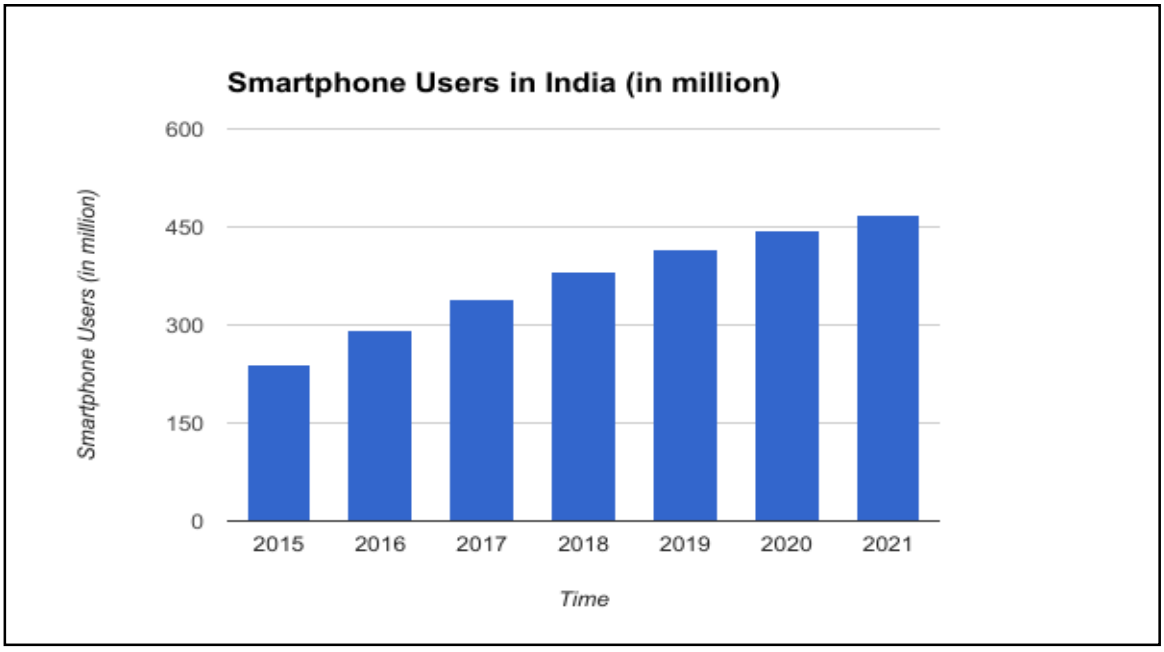


Figure 1-4. Smartphone users in India with predictive figures till 2021 (Statista, 2017).

However, within an interesting relationship between ICTs and their deployment in developing regions, certain specific and distinct field are growing steadily and concurrently. At one end, sociologists and anthropologists acknowledge presence of mobile telephony amongst masses in developing regions while shaping their own canonical perspectives. At another end, we get to hear economists interpreting economic impact of proliferation of ICTs. In between all of these, there is a growing body of work which raises questions relevant to the appropriate design of interactive products, applications and systems with respect to users' needs in the developing regions. The questions raised covers wide-range of issues from micro-level interface design issues to macro-level social interactions. Ho et al. (Ho, Smyth, Kam, & Dearden, 2009) would term this stream of work as "Human Computer Interaction for Development (HCI4D)". While stating the scope of HCI4D, Ho et al. (Ho et al., 2009) writes:

"Any HCI research that addresses the needs or aspirations of people in developing regions, or that addresses specific social, cultural, and/or infrastructural challenges of developing regions."

It appears to us that Ho et al.'s deliberation on defining HCI4D's scope is clearly suggestive of certain core intents and approaches that HCI4D promotes and

itself relies on. These are proposals to understand users' needs, their contexts and their engagements with technology artefacts (existing as well as recently introduced) prior to designing information technologies and artefacts in developing regions. While there are indications of approach being highly human-centered here, another perspective of HCI4D methodology is given by Anokwa et al. (Anokwa et al., 2009). Yaw Anokwa and eight other HCI4D researchers from different regions examine challenges in HCI4D due to differences in cultures, language, ethnicity and socioeconomic status. In their writing, they attempt to reflect on their own experiences in HCI4D while talking about research methods and practices. These researchers suggest prominence of no single method over distinct others. What is rather proposed is a form of mixed method research involving both qualitative and quantitative research methods. Another aspect, which is distinctly discussed, is the importance of studying existing technologies (or their traces) in communities where newer information and communication technologies (ICTs) needs to be deployed. For these HCI4D researchers, it is only after paying attention to studying users' interaction with existing technologies, can newer ICTs interventions be studied appropriately. They also talk about issues related with securing access to users in the developing regions. They acknowledge that communities in these regions may be relatively 'closed' to outsiders if not conservative. Under these situations it may sometimes be difficult to secure access to less privileged member of these communities viz. females and children. The role of facilitators is crucial while performing HCI4D research, and a HCI4D researcher may have to negotiate constantly while during field trials.

We identify the current thesis and the research enquiries it makes as part of HCI4D research work. We address specific issues born out of inconsistencies in the literature and practice with respect to the contemporary designs of the audio-only interfaces like Interactive Voice Response System (IVRs). We intervene through design artefacts, the visually augmented design variants of audio-only interfaces, and conduct systematic empirical experimentation to establish their relevance over audio-only interfaces and graphical user interfaces for users in the developing regions.

1.2 Audio-Visual Interface Design and Emergent Users

Unlike an educated, western or westernised, office going, urban and globalized users of technology, many users from developing countries are characterised by low levels of education, low literacy, low incomes, social inequality and less exposure to technology. Devanuj et al. call these users as "emergent users" (Devanuj & Joshi, 2013). Researchers have paid a continual attention to designing appropriate interfaces for emergent users (Grover, Stewart, & Lubensky, 2009; T. Parikh, Ghosh, & Chavan, 2003; N Patel, Chittamuru, Jain, Dave, & Parikh, 2010).

Since the mid 1980s, graphical user interfaces (GUIs) have been the predominant form of human-computer interaction (HCI). However, in the context of emergent users we could argue that audio-based interfaces such as interactive voice response systems (IVRs) have several advantages over GUIs. Firstly, IVRs do not require literacy. Further, for obvious reasons, IVRs are particularly suitable in oral cultures. From a technical and infrastructural point of view, IVRs work on all available phones, including smartphones, feature phones, basic mobile phones, and even landline phones. In contrast, GUIs are more dependent on the type of phone in use. GUIs almost always involve reading text, which requires literacy. While some work has been done for text-free GUIs for emergent users (Marsden, 2007; Indrani Medhi, Prasad, & Toyama, 2007; Neil Patel et al., 2009), as discussed in chapter 3, graphics and icons are dependent on culture and interpretation by the user. In contrast, audio, especially verbal audio "directs" the users to do tasks. Hence, researchers in HCI for development (HCI4D) have a growing interest in using IVRs in context of emergent users (Joshi, Emmadi, et al., 2012; N Patel, Agarwal, & Rajput, 2008; N Patel et al., 2010; Plauché & Nallasamy, 2007; Stritzke & Dandy, 2005). These approaches have been summarized in chapter 3.

However, at the same time IVRs interface designs are in particular problematic in nature. These are not known to have high levels of usability. In terms of the context of use, IVRs have been traditionally used for customer relationship management (CRM). They are typically used to "triage" calls to the call centre, to let customers avail services in a "self-service" mode, thereby reducing the costs of running call centres. However, customers often want someone to hear them out, and for them IVRs

just “get in the way”. In addition, the differences in terminologies used by the customers and the companies are often a reason of disconnect between IVRs and their users. Along with such a typical case of IVRs deployment and usage, the IVRs designs suffers from ephemeral and sequential nature of audio, and limitations of the user’s short term memory, IVR menu hierarchies are difficult to navigate. Users’ interactions with IVRs are easy to forget and hard to learn (Tatchell, 1996). IVRs based on automatic speech recognition (ASR) technologies and flexible dialog structures seem to improve the user experience somewhat, but even so, IVRs have been used sparingly. In any case, ASR technologies are still not sufficiently robust in context of languages used in developing countries. This is the problem space this thesis seeks to address.

One approach to improve the usability of IVRs is to support their interface designs with visuals. In our previous work done so far (Joshi, Emmadi, et al., 2012; Joshi, Chakravarty, & Shrivastava, 2012; Shrivastava & Joshi, 2014), we found that such audio-visual interfaces can overcome some usability issues in context of emergent users. Often use of visuals is favored because of literacy issues which may exist. Audio-visual interfaces are particularly promising in the context of increasing popularity of smartphones in developing countries. As smartphones are “phones”, audio-based interfaces are natural to them. As they have a large touchscreen, displaying visuals is easily possible. Further, the ecosystem of deploying applications on smartphones is becoming increasingly easy, and a wide variety of “apps” are becoming available. Although encouraging, but our efforts in the previous work have led us to realize certain specific aspects of audio-visual interface designs which need further research investigation. The next section lists details of these aspects considered in this thesis.

1.3 Research contributions

This thesis presents research that investigates both styles of user interfaces, audio-based and graphical user interfaces, while designing audio-visual interfaces for emergent users. This is done systematically by first acknowledging and identifying relative merits of both the modalities of interaction, audio and visuals, in sustaining task oriented interactions with emergent users. It is then followed by specific enquiry based studies with emphasis on menu hierarchies (see Chapter 4), and on the ability of

individual modalities for bringing *directedness* and *persistence* to the audio-visual interface design (Chapter 5).

Menu hierarchies in IVRs design is a much debated topic amongst HCI researchers and designers. Certain other issues in relation with the *optimal number to items present in an audio menu*, and *how users exercise menu selection while interacting with audio interfaces like IVRs* also keep surfacing in this discussion. Literature (Commarford, Lewis, Smither, & Gentzler, 2008; I Medhi, Toyama, Joshi, Athavankar, & Cutrell, 2013) generally recommends using broader menu-hierarchies over deeper menu-hierarchies in IVRs design (Chapter 3) while keeping 4 or fewer menu items at each audio menu level (Cohen, 2004; Marics & Engelbeck, 1997; Miller, 1956). In our opinion this is only a partial suggestion. It does not address the issue of appropriately designing menu-hierarchies in audio-based interfaces. We are not aware of any guidelines for audio-visual interfaces. In addition, it also does not make any explicit remark with reference to users making selection in an audio-menu where too the literature is inconsistent. Some believe (Marics & Engelbeck, 1997; Miller, 1956) that fewer items in an audio menu work better than larger number of items. More the number of items in a menu, more is the load on users' working memory which is known to have limited capacity. Contrary to this, few others (Balentine & Morgan, 1999; Commarford, 2006; Schnelle-walka, 2011) posit that menu item selection is independent of numbers of items in any given menu. Users do not keep items in their working memory rather they look for 'best fit' choices while making the selection.

In response to inconsistencies in the existing literature on menu hierarchies for IVRs design, our first major study (chapter 4) demonstrates that users exhibit better task completion behaviour with deeper menu-hierarchies over shallower menu-hierarchies in audio-visual interfaces. We also illustrated that any change in menu-depth and/or menu-item-position is less likely to bring significantly different success scores in case of audio-only IVRs. In other words, audio-only IVRs users are more likely to show consistently 'poor' success scores even if menu-depth and menu-item-position are changed. While on the other hand if the same changes occur in audio-

visual IVRs, emergent users are more likely to exhibit significantly different success scores.

We realize that an appropriate use of "audio" can bring *directedness* in the designed interface, be it audio-only IVRs or the proposed audio-visual interface. With the use of audio, the interface can help direct emergent users towards successful completion of the task. It can present users with explicit suggestions through audio prompts and can help users sustaining their interaction with the interface (Acomb et al., 2007). Emergent users can be presented with a series of audio prompts mentioning explicitly interface features and functionalities. "Visual", contrary to audio, can help bringing *persistence* to the interface. There is unique to visuals as these can help presenting information parallelly before the users. If needed these can stay for a desired time before the users and therefore, unlike audio, are not always temporal unless desired deliberately with some specific intent in designer's mind. With the use of visuals in interface design letting information to be present parallelly before the users, users' working memory does not get taxed while interacting with the interface. We record that such a knowledge regarding relative advantages of both the modalities seems to be present for long (Hartman, 1961; Nugent, 1982). However we would see (in chapter 3) that its' use in interface design for emergent users is fairly absent for reasons beyond the scope of this thesis.

In our second major study (chapter 5), we present empirical evidence in support of audio-visual interfaces on parameters of *directedness* and *persistence* by comparing one of its test variant against variants of graphical user interface and audio-only IVRs. This is done through rigorous testing of test prototypes with a sample of emergent users. We demonstrate that emergent users were significantly more successful in completing test tasks with audio-visual interfaces than with graphical user interface and audio-only IVRs. We also demonstrated that audio-visual interface enables emergent users to perform tasks significantly much faster with audio-only interface, and with tasks times close enough to graphical user interface. This indicates that use of "audio" and "visuals" in designing audio-visual interfaces for emergent users lead to substantially improved task performance and task time than when these modalities are used individually.

1.4 Organization of the thesis

This thesis reports research work carried out at the cross-section of three different entities: (1) Emergent users and the relevance of audio interfaces (like IVRs) in designing appropriate interfaces for them, (2) Issues inherent to audio-only interfaces, and (3) Visual augmentation of audio-only interfaces to result in audio-visual interfaces for emergent users. This thesis brings together relevant commentary and research perspectives on all of these three entities along with original research work undertaken by us. The eventual outcome of this thesis and the research enquiries it raises, is a set of prepositions supported through empirical evidences favouring audio-visual interfaces for emergent users. We now present the outlines of the upcoming chapters of this thesis.

Chapter 2

Chapter 2 presents four different studies as our early explorations with audio interfaces. In the first study, we gauge the limits of interactions possible over an audio channel in a situation when the users are devoid of contextual information in any possible visual form. We use think aloud protocol to record and analyze users' responses and saw our participants experiencing limits of such interactions. In the second study, we provided participants with two different audio scripts with identical content: continuous script and quiz script. In response, we measured their comprehension of the content through pre-test and post-test assessments. In third study, we experimented with the use of audio prompts along with number-pad based interface in a concept phone called *Shree Ganesha* designed specifically for emergent users. Towards the end in the fourth study, we mention details of usability evaluation of a limited scope audio-visual interface.

Chapter 3

This chapter presents background of this research project. This includes review of literary sources including researcher papers, patents, project reports and critical commentary available in the domain of HCI4D with respect to audio-visual interface design for emergent users. We begin with a discussion on understanding "emergent

users" and their context. It brings essential commentary pertaining to the relevance of directed dialog IVRs as interfaces for deploying services of interest for emergent users. Recent instances of case studies and field deployments with implications are brought to light. We also cover details of long standing usability issues prevalent with IVRs and audio interfaces. Subsequently, the chapter includes details on supporting IVRs with visuals as possible strategies to improve their usability. Towards the end, we bring relevant discussion on menu hierarchies, directedness of audio and persistence of visuals, which we address as part of the current research.

Chapter 4

In this chapter, we report our first major experiment where we used visuals along with audio prompts to design audio-visual interface. We varied use of modality (audio-visual and audio-only), depth of the menu hierarchy (shallow and deep), and position of the menu items (early and late) in our test prototypes. Effects of these variations were studied for task success, task time, choice errors and menu repetition as dependent variables with emergent users as participants.

Chapter 5

Chapter 5 reports our second major experiment. We bring attention on audio and visual modalities for their inherent abilities of "directedness" and "persistence" respectively. We hypothesize that using visuals along with audio prompts to design audio-visual interfaces would work with emergent users owing to directedness and persistence. In other words, we believed that both the modalities, audio and visuals, in audio-visual interfaces (AV) would work better than when they are used individually as interface modalities in audio-only interface (A) and graphical user interface (G) respectively for emergent user population.

Chapter 6

In this chapter, we reflect upon our research efforts towards designing audio-visual interfaces for emergent users. We discuss relevance of our results from our major experiments, reported in chapter 4 and chapter 5, with respect to the already known knowledge. We discuss the claims that the current research work makes. We

mention implications of our research findings for a community of designers, researchers and institutions working toward designing appropriate interfaces for emergent users. Towards the end, we acknowledge the limitations of the current research project. We also identify certain opportunities for conducting future research enquiries in the domain of audio-visual interface design for emergent users.

Chapter 2 Early exploration in IVRs research space

2.1 Introduction

In this chapter, we discuss four of ours research studies carried out in the Interaction Design for International Development (IDID) laboratory at Industrial Design Centre, IIT Bombay. It is in the same laboratory that this thesis work has originated and successively shaped up. These studies took place during the initial phase of our research journey and before our main experiments (see Chapter 4 and Chapter 5) began. We find it important to mention these because of novelty of explorations in these studies. We were motivated to touch these diverse themes to help us gain a first-hand experience of IVRs oriented research space for emergent users.

2.2 Study 1: Gauging limits of interaction over an audio channel

2.2.1 Motivation

Our motivation in this study was to observe limits of interaction solely supported over an audio-only channel. We wanted to record participant's reactions and the challenges that they face in a communication environment, supported only over an audio channel with no presence of context specific visual information. We hoped that such an experimentation would deliver interesting insights which will help us understand both audio and visuals as interaction modalities.

2.2.2 Research question

In this study we considered an activity which was usually a visually rich experience. We chose *a game of cards to be played amongst three players* as one such activity. As part of this experiment, we transformed this activity and made our participants play the game from a distant location over an audio channel. Under these conditions, we raised the following research question.

What elements of visual information be required by a player, deliberating over a distance through the use of audio channel only, to play a game of cards with competence and ease similar to the natural setting of the game?

Note that although we dictated usage of only audio channel for all possible interactions in this experiment, we encouraged our participants to verbalize their thoughts while they performed the given activity. They were also encouraged to make notes on a piece of paper during the experiment. We hoped that we could analysis participants' thoughts and their paper records to deduce knowledge of critical elements of visual information which may be required to complement audio based interactions.

2.2.3 Experiment setup and design

2.2.3.1 Activity design

A 3-2-5 game of cards was selected as an activity for this study. We had two reasons behind this choice: (a) popularity of the game, and (b) the number of players involved, which was three in this case. It helped us optimize the resources required for the study. Note that this activity, in its usual form, imparted a visually rich experience (Figure 2-1). However, *under experiment settings*, the participants sat at two different locations (Figure 2-1). Location A housed two players and one facilitator. The facilitator had to play on behalf of the participant. The other location B housed only the participant. The locations A and B were acoustically as well as visually isolated from each other. Moreover, participant in location B was connected to the facilitator in location A via an audio-channel over two identical feature phones. The participant, in location B, had to act through the facilitator to play the game with the other two players at location A.

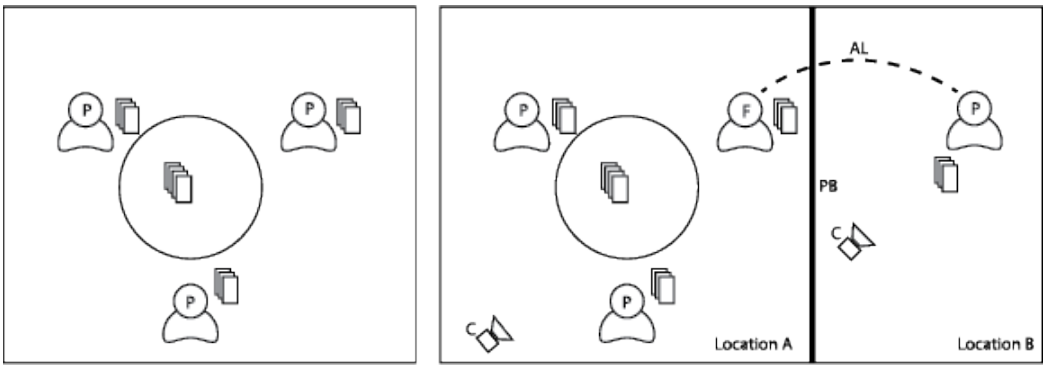


Figure 2-1. (LEFT) Usual game settings with three players (P). (RIGHT) Experiment game settings with two players (P) and a facilitator (F) at location A. The third player (P), at a distant location B, is connected with facilitator (F) over an audio channel (AL).

2.2.3.2 Participants

A total of 9 participants, 4 males and 5 females (M = 27.7 years, S.D. = 2.43), participated in this study. All the participants were volunteers and were drawn from a population of graduate and doctoral students aged between 25-32 years through convenient sampling.

2.2.3.3 Protocol

The *participants* were briefed to ask the facilitator for any context specific information but not suggestions besides throwing cards through him. Participants were detailed to not to team up with the facilitator. Instead he was briefed of the role of facilitator as mere imitation of a machine which followed his will. The *facilitator* was briefed to act strictly on behalf of the participants. He was told not to push any explicit or implicit suggestion or render help.

2.2.3.4 Data collection and analysis

The experiment setup mimicked interactions supported by an audio-based interface where users had to deliberate explicitly to navigate and complete their tasks. We asked our participants to verbalize their thoughts during the task performance, leading to 'concurrent verbalization'. We videotaped these verbal reports and made other observations. If required, we probed participants in retrospection for showing a distinctly specific behavior. We analyzed these records using protocol analysis, where researchers collect and analyze verbal reports contributed by the subjects (Schmidt,

1995, p. 188). The method has been in vogue in a diversity of domains including cognitive science, psychology, translation studies, language learning and education (Ericsson & Simon, 1993). We come across several studies where researchers have tried gaining insights into the 'mind's eye' during creative processes including roles of gestures (U. Athavankar, 1999; U. A. Athavankar, 1997; U. Athavankar, Bokil, Guruprasad, & Patsute, 2008; Bilda & Gero, 2005; Pillai, Athavankar, & Schmidt, 2013). Ericsson et al. suggests that if administered carefully, protocol analysis maintains the course and structure of the subject's thought process without any alteration (Ericsson & Simon, 1993). Another apparent limitation of this method is that parts of the thought process during a concurrent task performance are often the only verbalizable (Halliday, Gibbons, & Nicholas, 2012, p. 381) elements. Accordingly we allowed our participants to take pauses in their verbalizations. They could also reflect back and add more details (in retrospection) to complement this limitation.

2.2.4 Findings and their implications

We saw several instances of regression suggesting that participants faced problems memorizing relevant factual information during the game like sequence of cards and of participants' turns. Another instance when we observed regression in the verbal reports was connected with the weakening of communication between participant players and the facilitators as the game progressed. Participants, challenged by their capacity to memorize, entered into an endless loop of requesting the facilitator for relevant information. Initially facilitators received requests from participant players limited only to queries regarding cards played during the deals. However, we later saw these queries turning into requests for verifying the beliefs of the participant players, which they developed during the game.

Our findings suggest and at times, reiterate certain implications. *First*, audio interface designers must pay attention to the limits of human working memory. Suhm (Suhm, 2008) distinctly proposes using 7+-2 chunks of information in an audio prompt. In his argument, he cautions designers to weigh the amount of information which a speech interface can present to the users. On a similar note, Balentine (Balentine, 2007, p. 117) finds memory challenges to be critical to tasks where knowledge of information at intermediate steps is crucial to complete the task at hand.

Second, we observed nearly all of the participants making visual records of relevant information during the game (Figure 2-2). Verbal report analysis constantly suggested that participants tried devising methods to overcome the limitations posed by the experiment. However, their notes on sheets of paper suggested a clear preference for drawing "pictures" with very little text. We, therefore, bend toward suggesting preference for using visuals to complement audio prompts in order to make interactions less taxing for users' working memory.

Figure 2-2. Instances of different players recording game related information on the paper. Distant player keeping cards at different locations so as to be able to distinguish between the cards where deals are made and lost, and the card which she is currently pursuing in the deal.

2.3 Study 2: Audio Information Delivery - Continuous vs. Quiz script

2.3.1 Background

In this study we wanted to weigh the effects of changing audio prompt's script on users' comprehension of the content. As we discuss in chapter 3, audio prompts are transient and non-persistent in nature. They exist in time and appear sequentially before the users. In IVRs, these prompts not only present users with system features and functionalities but with relevant information content too. Common sense suggests that an audio prompt with passively played information content, let us call it *continuous* script, may instill lesser content comprehension in the users. It may prove taxing for users' working memory, and may bring aural fatigue. We wonder if incorporating some kind of interactively in the script itself may improve this picture for IVRs users. We took inspiration from "situational judgment test" (Whetzel &

Wheaton, 2016) to conceptualize such an audio script. In this technique, participants are presented with scenarios and are subsequently asked to mark their response by making a choice from an available set of possible responses. The scenarios depict the situations that the participants might encounter in a particular experience. The participants, in turn, mark their response by evaluating probable measures as response to the situation mentioned in the scenario. This technique of testing participants therefore simulates *situations* to measure *knowledge* of a particular kind. In terms of interaction, this is akin to quizzing the participant. We called the audio script based on situational judgement test as *quiz* script in this study.

2.3.2. Hypothesis

We used two different script types with identical content. The first one, continuous script (Appendix 7), was more conventional. Whereas the second one, quiz script, was made interactive by bringing in aspects of quizzing (Appendix 8). We hypothesized that quiz script would yield users' comprehension of the content better than continuous script.

2.3.3. Content

We chose Human Immunodeficiency Virus (HIV) and Acquired Immune Deficiency Syndrome (AIDS) as topics of the audio scripts. We looked into accredited resources of both government and non-governmental organizations ("HIV Aids Topical Information," 2011, "NACP," 2011); expert interviews; and awareness campaigns - both online and offline. Besides we considered relevant documents prepared by a panel of competent doctors consulting us in an another live project in IDID lab. The information gathered on HIV-AIDS included the following topics: (a) definition of HIV-AIDS, (b) cause of spread of HIV-AIDS amongst human beings, (c) ways in which HIV-AIDS attack a human body, (d) HIV-AIDS treatment regimen, and (e) importance of adherence to the treatment. Note that the language chosen for audio scripts was Hindi.

2.3.4 Method

3.3.4.1 Experiment design

This study employed a between-group design. All the 30 participants were divided into two independent groups of 15 participants each. They were randomized to be assigned only one of quiz script and continuous script. The between group variable was the script design with two factors-continuous and quiz scripts. The dependent variable was the gain, calculated as a difference between the post-test and pre-test.

2.3.4.2 Participants and protocol

A total of 35 participants, both 28 males and 7 females between 25-50 years of age, appeared for this study. Out of these, first 5 participants were considered for pilot trials.

We *first* briefed participants about the objective of the study, its stages (Figure 2-3) and their roles. The first stage of the test was to ask the participants to respond to a set of multiple choice questions (Appendix 5). Each question was printed on cards to help us randomize their assignment. During the *second* stage, we asked participants to listen to audio on a mobile phone. We intend to bring all the participants at the same level of understanding of HIV- AIDS (Appendix 6). During the *third* stage, we asked participant to randomly listen to only one of continuous script (Appendix 7) and quiz script (Appendix 8). During the *fourth* stage, we asked participants to respond to the same set of questions as in stage 1. We changed the order of these questions to eliminate easy recognition of the questions.

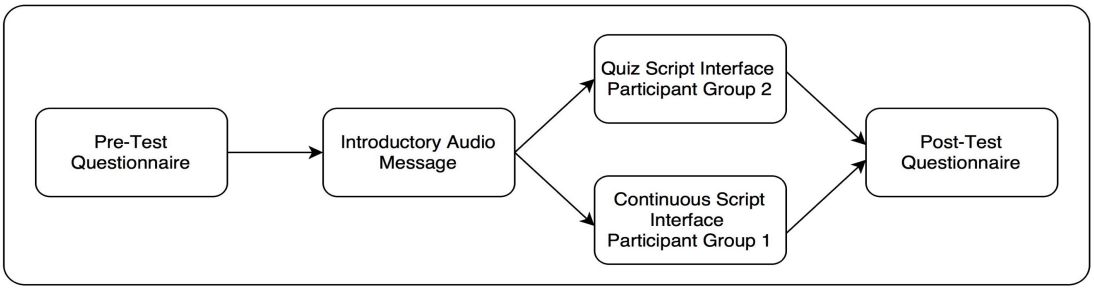


Figure 2-3. Study design.

2.3.4.3 Measuring instrument and data collection

We used a questionnaire of 9 questions on topics related with HIV-AIDS as measuring instrument. This questionnaire was prepared in consultation with an onboard team of medical professionals. We administered it twice, in stage 1 and 2, to constitute pre-test and post-test scores respectively.

2.3.5 Result and its implication

The mean of the difference between post-test and pre-test for continuous and quiz script participants are 2.53 and 3.73 respectively. The gain was therefore better for continuous script (Table 2-1). In order to establish the significance of this observation, a 2-tailed student *t-test* was conducted to compare scores corresponding to quiz script and continuous script. Test statistics suggest no significant difference in the scores for quiz script (M=2.53, SD=1.55) and continuous script (M=3.73, SD=1.75) conditions; t (28)=0.057, p = 0.05.

Table 2-1. Descriptive statistics of gain (difference between post-test and pre-test).

Method	Number of participant (N)	Mean	Standard Deviation	Standard error
Quiz script	15	2.53	1.55	0.4
Continuous script	15	3.73	1.75	0.45

Our results imply that engaging users over an audio prompt design based on an interaction like quizzing might not help users in developing a better understanding of the underlying information content over a conventional continuous audio prompt. This is counter intuitive. We speculate that perhaps there were other unknown variables which led to this outcome. May be it is not just about the design of the audio script but also the nature of the content as well that decides its comprehension. May be we could have seen a different result in the favor of quiz script had the topic of the content was something other than HIV-AIDS. Our participants might have chosen to attend the content sequentially (in continuous script) than as episodes and scenarios (in quiz script) considering its value to their lives. We therefore caution IVRs system designers and others to pay attention to both the nature of the content as well as its manner of scripting in audio prompts. No one solution seems to work when it comes to audio prompt script design.

2.4 Study 3: 'Shree Ganesha', a phone for illiterate users

2.4.1. Motivation

This study was based on the observation that members of the emergent user population were numeral literate more often than textual literate. They dealt with numbers as part of their daily experiences. While they might not have time, opportunity and resources to attain formal education, but they had experience managing money, time, their schedules and dates. They had interest in knowing cricket scores, market rates of commodities etc. Hence, numeral literacy seemed to have come naturally to emergent users. This motivated us to design a number-pad based interface with audio instructions as an illustrative mobile phone design for illiterate users. We hoped to draw useful insights in order to inform our understanding of emergent users and subsequently to suggest improvements in the design of mobile phones to suit their requirements.

2.4.2 Salient details of the test prototype

We used *Ganesha* as a metaphor for the landing page or homepage of the phone. It was found appropriate for couple of reasons. *Ganesha*, the elephant-headed god of knowledge, is one of the best-known and most-worshipped Hindu deities. He is the first to be worshipped, and is the remover of obstacles and creator of happiness. In addition, *Shree Ganesha* is written at the top of an important document. We hoped that use of a well-received metaphor like *Shree Ganesha* would sound exciting to our users.

We designed the phone with at least three levels of audio prompts corresponding to each functionality viz. dial a number, see received calls etc. The default mode had the most detailed instructions, while the rest two modes had lesser information prompts. For dialling a specified number, the typical user journey began with encountering a locked screen (Figure 2-4). The user had to unlock the phone by pressing three number keys 1,3, and 7, at three corners of the screen. On pressing a wrong key, the expected key blinked to generate visual cues corresponding to unlock code. If the user still pressed wrong keys, an audio prompted to suggest (in Marathi)

the right combination of keys. It said, “The phone is locked. To unlock the phone, press 137”.

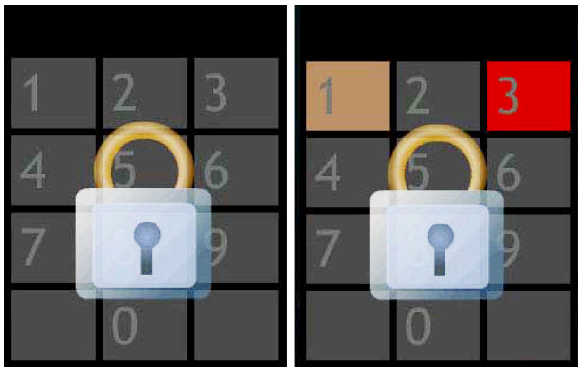


Figure 2-4. Locked phone screen.

Once the user unlocked the phone, he saw the landing page of the phone with menu suggesting phone features and functionalities (Figure 2-5). In addition, an audio prompted user of probable actions he could take. It said, “Welcome. To dial a number, press 1. To see previously dialled numbers, press 2. To see received calls, press 3. To open the phonebook, press 4… To lock the phone press 0”. On pressing 1, the user navigated to “Dial a Number” screen. He heard the corresponding audio prompt, “Please punch the numbers you wish to dial”. The user punched the keys, and with each input, an audio prompt read the number to him. On punching the keys associated with the specified number, the phone dialled it automatically. In case the user punched wrong keys, and didn’t realize it in another 10 seconds, another audio prompted the user to press the green button to initiate the call. On initiating the call, he was shown visual feedback corresponding to the status of the call.

Figure 2-5. (Left) Landing page, (Middle) Dial a number screen, and (Right) Confirm dialing screen.

2.4.3 Usability Evaluation

We were interested in finding out if first-time users, illiterate or lowly literate, could carry out the two basic tasks of unlocking the phone and making a call without help. In addition, we compared the test prototype with a contemporary mobile phone (in Marathi). We recruited 12 users, 6 men and 6 women - all aged between 30-50 years, from *Kashele,* a village 130km away from Mumbai. Users had low literacy of up to 4^{th} standard and with 3 of them being complete illiterate. 4 out of 9 users with low literacy could read Marathi, and not English. Rest 5 could read English numbers, apart from Marathi.

In the beginning of evaluation session, we briefed every user about the study objectives and his contribution. We then gave him the test prototype, and asked him to dial a specified number. The test prototype, at the time of assignment, was in locked screen mode. Therefore, the user could dial the number only after unlocking the phone. We counted a "successful attempt" when the following conditions were met: User could realize that the phone was locked, he could unlock it further and he could dial the specified number. If these conditions were not met, he was offered manual help in two different format. *First,* he could be helped in suggesting what should be done to perform the task. *Second,* he could he helped in reading out the numbers or the text displayed on the screen. We counted an "unsuccessful" attempt against the first format and a "successful with help" attempt against the second format. Once user finished performing task with test prototype, he was asked to repeat the same with a contemporary mobile phone. Assignment of the contemporary mobile phone and of the test prototype was randomized across the users.

2.4.4 Results and their implications

A total of 75% users unlocked the phone successfully without help in their first attempt with *Shree Ganesha* test prototype (Table 2-2). This existed in contrast with only 16.7% users doing the same with contemporary mobile phone (Table 2-3). The percentage of users completing the task successfully with help in *Shree Ganesha* and contemporary mobile phone were 25% and 16.7% respectively. While we had no

unsuccessful user in case of *Shree Ganesha* phone, we had 66.7% users failing in contemporary mobile phone.

Table 2-2. Task "Unlock the phone", *Shree Ganesha* phone.

Evaluation outcome	User profile			Total	
	Illiterate	Cannot read English numbers	Can read English numbers	Count	%age
Successful	1	3	5	9	75.0
Successful with help	2	1	0	3	25.0
Unsuccessful	0	0	0	0	0.0

Table 2-3. Task "Unlock the phone", contemporary mobile phone.

Evaluation outcome	User profile			Total	
	Illiterate	Cannot read English numbers	Can read English numbers	Count	%age
Successful	0	0	2	2	16.7
Successful with help	0	1	1	2	16.7
Unsuccessful	3	3	2	8	66.7

Given the task of dialling a specific number, we recorded that only 33.3% with either of the mobile phone could do it successfully without any help (Table 2-4, and Table 2-5). While we had more unsuccessful users in contemporary mobile phone, we had 66.7% users doing the task successfully with help in *Shree Ganesha* phone than 41.7% in contemporary mobile phone.

Table 2-4. Task "Dial the given number", Shree Ganesha phone.

Evaluation outcome	User profile			Total	
	Illiterate	Cannot read English numbers	Can read English numbers	Count	%age
Successful	0	0	4	4	33.3
Successful with help	3	4	1	8	66.7
Unsuccessful	0	0	0	0	0.0

Table 2-5. Task "Dial the given number", contenporary mobile phone.

Evaluation outcome	User profile			Total	
	Illiterate	Cannot read English numbers	Can read English numbers	Count	%age
Successful	0	0	4	4	33.3
Successful with help	2	2	1	5	41.7
Unsuccessful	1	2	0	3	25.0

Users were evidently completing tasks both “successfully” and “successfully with help” when they tried *Shree Ganesha* phone. Incorporation of audio prompts along with a simplified interface with visual cues, did show promise in terms of helping the illiterate or low-literate members of emergent user population. We, therefore, recommend investing an increased design and research efforts in the combined usage of audio and visuals for emergent users. In addition, there were certain interesting observation corresponding to the use of audio prompts. *Shree Ganesha* users didn’t interrupt the audio. They could barge-in but they refrained deliberately even when they knew their way through the interface. In their first attempts when audio prompts were played, some of our users brought the phone closer to their ears. During this time, they missed seeing visual interface and its highlights. Hence while designing audio prompts, perhaps it was important to time it with respect to the timing of the elements of visual interface. An audio-visual interface designer needed to account for this eventual disparity in attending audio and visual signals. Perhaps content redundancy could be brought in during the audio-visual interface design. An identical content on both the audio channel and visual channel might help avoiding user experiences where they tend to miss content on one of the channels.

2.5 Study 4: Usability evaluation of audio-visual interface

2.5.1 Motivation

This study was one of our first attempts to augment audio prompts with visuals. We made deliberate efforts towards designing audio-visual interfaces by creating test

prototypes in two different variations, audio-only interface and audio-visual interface for two different product types, banking and railway ticket enquiry. We conducted usability evaluation of these test prototypes for test tasks which imitated real world users' requirements, like gathering transaction history for banking accounts and looking for railway ticket enquiry between two locations. Presented further are the details of different segments of this study.

2.5.2 Design of the prototypes

We designed two different flash applications for two different exiting IVRs – banking and railway ticket enquiry. These flash applications worked to provide visual skinning to the existing IVRs (Figure 2-6). That is to say, they were designed to run in real time with the existing IVRs and provided visual aids to the users in one to one correspondence with the spoken audio prompts. While running in real time, these applications generate dual tome multi-frequency (DTMF) signals against the user inputs on audio-visual interface. The users pressed soft keys on the screen in order to make choices. But these keypresses converted into DTMF inputs for existing IVRs running in the background (Figure 2-7). The IVRs then detected these DTMF signals generated by the flash applications and accepted these are user inputs. One could appreciate that flash applications were designed to run locally on the mobile phone used during the evaluation. We did not need to download it, and hence we did not require connectivity through internet. We believe that this could be one of the optimum ways of generating audio-visual interfaces with existing IVRs without disturbing or bring any change in the later.

Figure 2-6. Audio-visual interfaces for (Left) Banking and (Right) Railway ticket enquiry.

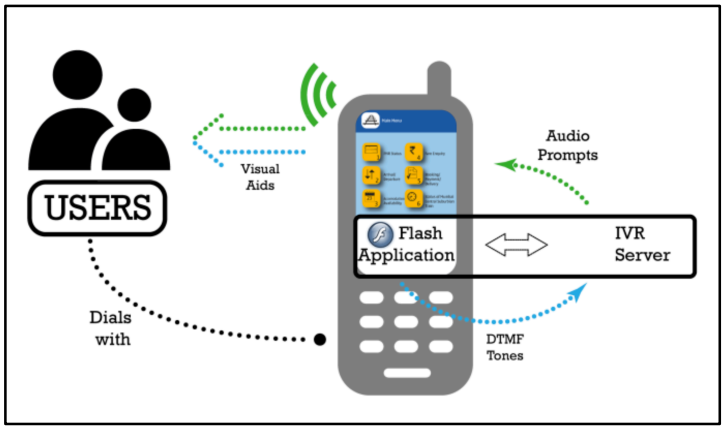


Figure 2-7. Concept model of the audio-visual interface.

2.5.3 Participants

A population of 40 users (19 males, 21 females, mean age 39.1 years) participated in this study. Amongst these, 7 had studied up to 12th standard, 26 had bachelor degrees, and 7 had attained a higher education at master's level. They were non-native speakers of English but could follow the language reasonably well. They had exposure of using IVRs at least once, and many of them had checked their banking transactions or had enquired about railway reservation through IVRs based services. However, as we mention in chapter 3, they were usually secondary users (T. S. Parikh & Ghosh, 2006) of ICTs. That is to say, they had experienced interacting with technology through human proxies, or someone with skills to operate technology

systems. Despite this, these users including women had experience carrying out tasks like checking talk time balance of their prepaid mobile connections.

2.5.4 Method

The study employed a within subject design. Each user performed two different test tasks, one each on banking (task A) and railway ticket enquiry (task B). By choosing within subject design, we tried avoiding errors induced due to individual differences. However, we still had to deal with carry-over effects resulting out of fatigue and learning. Assignment of test tasks A and B was, therefore, not only randomised but also counter- balanced within two products (banking and railway ticket enquiry) and interface styles (audio-visual and audio-only).

The protocol of the experiment was as follows. We briefed the participants first about the evaluation study and its broader objectives. The participants were then allotted warm up tasks to help them familiarize with the mobile phone usage, and with IVRs. The warm up task included calling a toll free IVRs service hosted by a mobile telephone operator to gather information on talk time balance. The facilitators provided help in the warm up task if required by the participants. This was followed by the assignment of test tasks A and B. As task A, we asked our participants to gather information on last 5 banking transaction against a given savings account. We provided participants with the banking IVRs toll free number, along with ID and access details. As task B, we asked our participants to enquire about the possibility of a reserved ticket for a journey with predetermined sources and destination, and for a given date and train. Here too, the participants were provided with account access details, the train number and station codes. We chose these tasks A and B, and warm up task to suit the set of activities which our participants might be performing on a regular basis.

The date recorded included task success score and time taken. For task success, we assigned 3 when users successfully completed the task in their first attempt without any help. We assigned 2, if users successfully completed the task in their second attempt. We assigned 1, if users completed the task successfully in their second attempt but with help. And we assigned 0, if users failed in their second attempt with

or without help. In addition to these performance markers, users also marked their response in a post-test evaluation form provided towards the end of the testing session. This questions helped us collecting users' subjective evaluation of *ease of use, confidence, time* consuming and *system complexity* corresponding to the test prototypes.

2.5.5 Results and their implications

Users with test tasks assigned on audio-visual interface exhibited better task completion rates over those with audio-only interface (Table 2-6). The mean success score for audio-visual interface (2.52, n = 40) was significantly higher than success score for audio-only (1.82, n = 40, $p < 0.05$). However, the difference between the mean time taken for successful task completion in audio-only (190 s, no. of users who finished the task: n = 34) and in audio-visual interfaces (177 s, no. of users who finished the task: n = 40) was only marginal. Note that 3 users declined finishing the test tasks on audio-only interface. Another 8 users exhibited poor performance when they were assigned first the audio-only interface and then the audio-visual interface. This observation came in contrast with those who were assigned the interfaces in a reverse sequence. Further we conducted a paired sample t-test to establish the significance for difference in mean time taken for users who did both the test tasks successfully. For these users, we found the difference in time taken for audio-only and audio-visual interfaces to be statistically significant.

Table 2-6. Average performance scores and time taken. A (✓) denotes significant difference for $p<0.05$.

Dependent variables	**Audio-only**	**Audio-visual**
Success score (out of 3) ✓	1.82	2.52
Time for successful task (seconds)	190	177
Time for users successful in both tasks (seconds) ✓	190	169

We now report analysis of *ease of use, confidence, time* consuming and *system complexity*, as markers of users' perception of the audio-visual and audio-only interfaces. We record that users found the task performance on audio-visual interface (6.8/10, n = 40) significantly easier than on audio-only interface (5.7/10, n = 40, p <

0.05). Their perception of the system complexity was also significantly less in case of audio-visual interface (6.4/10, n = 40) over audio-only interface (5.0/10, n = 40, $p < 0.05$). While we did not locate any significant difference between the confidence and the time consuming scores for audio-visual and audio-only interfaces, we still found direction of these results in favour of audio-visual interface (Table 2-7).

Table 2-7. Visual analog scale mean scores about user perception (0=worst, 10=best). A (✓) denotes significant difference for p<0.05.

User perception scores	Audio-only	Audio-visual
Ease of use ✓	5.7	6.8
Confidence	6.3	6.9
Time consuming	5.2	6.2
System complexity ✓	5	6.4

Now we detail the results of ANOVA analysis. It helped us access effects of individual factors (style of interface- audio-visual vs. audio-only; sequence of assignment-audio-visual first vs. audio-only first; and product- banking vs. railway ticket enquiry) and their interactions on the results of the study. The effect of the style of interface, audio-visual vs. audio-only, was found to be statistically significant for success score ($p < 0.0005$, $F = 25.333$), perception of ease of use ($p = 0.032$, $F = 4.703$), perception of time consumed ($p = 0.035$, $F = 4.505$) and perception of system complexity ($p = 0.047$, $F = 4.560$). In addition, there was a significant effect of the interaction between the style of interface and the product on time taken ($p = 0.004$, $F = 8.373$). The effect of the using audio-visual first was found to be statistically significant for time taken ($p < 0.0005$, $F = 17.770$). The interaction between style of interface and sequence on the perception of "how time consuming a task is" was also significant ($p = 0.021$, $F = 5.439$). No statistical significance was found for any other variables including its factors or their interactions (Table 2-8).

Table 2-8. Results of ANOVA. A: Style of interface (audio-visual, audio-only). B: Sequence (audio-visual first, audio-only first). C: Product (bank, railway). A (✓) denotes significant difference for p<0.05.

	A	B	C	A*B	B*C	A*C	A*B*C
Performance scores							
Success score	✓	-	-	-	-	-	-
Time taken	-	✓	-	-	-	✓	-
Perception scores							
Ease of use	✓	-	-	-	-	-	-
Confidence	-	-	-	-	-	-	-
Time consuming	✓	-	-	✓	-	-	-
System complexity	✓	-	-	-	-	-	-

Apart from quantitative analysis, we made some **qualitative observations** too. We often probed users, and encouraged them to give us more details of any specific behaviour shown by them during the evaluation. We saw that users, whom we first assigned test tasks on audio-visual interface, were frustrated when they were later assigned test tasks on audio-only interface. We had 3 users unwilling to finish the tasks in such scenarios. Their perception dictated that audio-visual interface made the entire system sound easy to understand. This was in stark contrast to the perception delivered by the audio-only interface which sounded limiting, and devoid of visuals denoting interface features and functionalities. Another interesting point, indicated by the users, connects use of audio-visual interfaces with real world contexts. Our users believed that they might have to access IVRs services even while commuting or in noisy environments. In such conditions, if they had audio-visual interface instead of audio-only interface, they would be able to make immediate references with the visuals even if they missed audio prompt. This alleviated their perception of *ease-of-use* and lesser *system complexity* for audio-visual interface.

2.6 Conclusion

Studies mentioned in this chapter touch diverse themes in IVRs oriented research space. In study 1, we setup a novel experiment to gauge limits of audio-only channel interaction with the users. In study 2, we experimented with the design of the IVRs audio scripts itself. We didn't alter the interface rather our approach was more content centric in this study. In study 3, we presented a keypad based interface with

audio prompts for illiterate users. And in study 4, we presented a short study evaluating audio-visual interfaces for performance and subjective satisfaction measures. Users distinctly mentioned possibilities of scoping up the audio-visual interface. They hoped that we could have used more items in an audio-visual interface, and could have imagined it with longer menu lists. In addition, using more number of items in an audio-visual interface would have helped achieving shallower menu depths. We consider this topic in chapter 4 of this thesis. We conduct a related study aimed at studying use of visuals with audio prompts for varying menu depths and menu positions with emergent users.

Over all, our experience across these studies got us motivated about studying use of visuals with audio prompts to design audio-visual interfaces for emergent users, as the research direction to pursue.

Chapter 3 Background

3.1 Emergent Users

The term "emergent users" was coined by Devanuj et al. (Devanuj & Joshi, 2013). They position modelling emergent users as their motivation on identifying an absence of adequate focus on modelling emergent users by members of the research community. Their understanding of emergent users is invariably linked with the trajectory of growth of Information and Communication technologies (ICTs) in developing regions of the world. The initial use of information and communication technologies (ICTs) shaped as tools for office and factory automation in predominantly urban and developed contexts. This involved use of computers (or, computing machines in general) for carrying out daily activities in offices, factories, hospitals, control rooms, banks and in locations with similar work requirements. Devanuj et al. call this *the traditional use of ICTs*. They contrast this *traditional use* with *non-traditional use of ICTs* by bringing use of ICTs in non-work environment and with intents other than automation viz. socialization, leisure and education. They mention that even if ICTs were gradually making their way to include non-traditional domains, their deployments stayed limited only to developed regions of the world for a long time. In line with the observations made by Ho et al. (Ho et al., 2009) and Anokwa et al. (Anokwa et al., 2009) which we discuss in chapter 1, they contend that adoption of ICTs in contexts of developing countries is a recent phenomenon.

Speaking of emergent users in this context, Devanuj et al. write;

> *The prime factors contributing to the human disadvantage are poverty, illiteracy, social inequity, poor healthcare and environmental degradation - issues that ICTs have not dealt with traditionally. Only within the last decade have ICTs reached beyond the traditional users and have shown a promise to enable human development at large by reaching new users, who may have less education (not reached college), who may be poor (for example, marginal farmers, very small business owners, village artisans catering to local markets), who are often located away from commercial and political centres, and are culturally different not only from the traditional (that is, urban and educated) users but from each other as well. In this paper, we refer to these as the emergent users of ICTs.*

This elaboration consists of several relevant markers to identify emergent users. It encompasses people with low incomes as well as with lesser education but with functional or numerical literacy. It speaks of extents to which people are exposed to ICTs, and related technologies. It doesn't include only the rural areas, which has been a case in most of HCI4D literature, but also considers urban locations. However, in either case emergent users may have limited access to infrastructure such as roads, drinking water, sanitation, and to services such as banking and healthcare. Devanuj et al. mention specific (but not limited to) personas of members of emergent users like marginal farmers, field labourers, small store owners, drivers, flour millers, tailors and carpenters, artisans and similar others.

While the term "emergent users" was coined fairly recent, research on critically understanding the users with stakes in ICTs led interventions, has been going on in an earnest over several years. Brewer et al. (Brewer et al., 2005) would consider emergent users as users of the "shared technologies". As part of their projects carried out in developing regions, including India, Banglades and Brazil, they had come across shared technology deployments viz. kiosks, telecentres doing better than personally owned technology like desktop computers. On similar lines, Gamage and Halpin (Gamage & Halpin, 2007) in an evaluation of impact of telecentres in North East Sri Lanka, identifed a differential in ICTs access amongst their users. In an another

writing, Parikh et al. (T. S. Parikh & Ghosh, 2006) call emergent users as secondary users. They suggest that emergent users experience interacting with technology through human proxies, or someone with skills to operate technology systems.

Interestingly apart from concerns of technology access amongst emergent users, we also come across relevant discussions on how to design interfaces for them. This discussion, as we would see in the following section, attempts understanding emergent users through the lenses of education, their capabilities and needs.

3.2 GUIs for Emergent Users

Emergent users may pose unique challenges to designers of interactive artefacts because of their contexts. There are many constraints which affect design of an appropriate interface for emergent users. Majority of studies count illiteracy as a limiting factor and propose designing text-free interfaces. Parikh et al. (T. Parikh et al., 2003) studied a Self Employed Women Association (SEWA) bank where most of the women had very little or no formal education at all. Out of a group of 25 women whom they interviewed, they discovered that only two or three of them were literate to any level. During interviews, they also discovered that they could understand numbers upto the extents of being able to perform simple calculations. It was with words and text in general that they had problem. Born out of this observation, Parikh et al. subsequently designed a numeric interface for these women. The conducted formative evaluation using paper prototypes (Figure 3-1) and found that numeric data including dates, percents and interest rates, etc. were successful as interface cues with their users. Medhi et al. (I Medhi, Sagar, & Toyama, 2007) considers this suggestion of avoiding text but also made use of semi-abstracted graphic in their interface designs for illetrates and semi-literate users. They designed an employment search portal for women from marginalized locations in Bangalore. In their evaluation, they demonstrated the relevance of text-free interface with voice feeback, where voice was readily helping users in comprehending the content. Further ahead, Joshi et al. (Joshi, Emmadi, et al., 2012) also observed that emergent users are quite likely to possess numeral literacy, if not formal education. All these attempts indicate an important concern. The text-centric interfaces pose difficulty for the illiterate or semi-literate

users. However, if the interfaces can be freed completely of text, they may offer easier understanding of interface features and functionalities to emergent users.

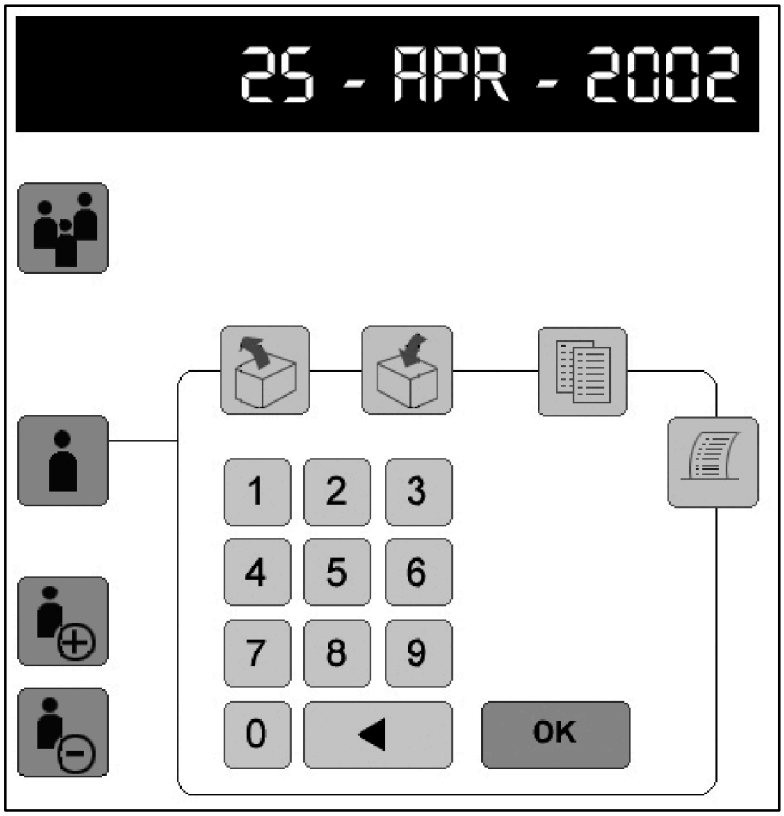


Figure 3-1. An early keypad based prototype by Parikh et al. Adapted from Parikh el al (T. Parikh et al., 2003).

In addition to relevant work done in designing numeric text-free interfaces, there were parallel efforts in the direction of designing interfaces with graphical aids for emergent users. Götze et al. (Götze & Thomas, 2001) call elements of a text-free interface as graphical aids. They illustrated use of graphical aids through replacing individual words in a web browser by pictures representing their meaning. Parikh (T. Parikh et al., 2003) and Ávila et al. (Ávila & Gudwin, 2009), in India and in Brazil respectively, further explored the inclusion of graphical aids. They used icons along with numerals for semi-literate women of microcredit groups. They found that a combined use of icons and numerals contributes towards an overall interface comprehension.

Literature, apparently, is also suggestive of limitations of text-free user interfaces, in particular purely graphical interfaces and numeric-only interfaces.

Marsden (Marsden, 2007) has illustrated how buttons and icons may be interpreted differently by the users in South Africa. While presenting a case of designing an educational interface for teachers and students in South Africa in his formative years, he comes across variations in the cultural meanings and prior knowledge of the users and the designers. What designers may find working as an 'interface button' could simply be reduced to a clueless or a misinterpreted graphic for users with no prior knowledge (Figure 3-2). In a different writing, Grover at al. (Grover et al., 2009) express similar caution while using graphics. They suggest that users may even find it difficult to make any definite sense of interface graphics which are too abstract. They also suggest using a relevant interface metaphor in order to invoke prior knowledge while designing interfaces for emergent users (Grover et al., 2009). Looking at both of Marsden's and Grover et al.'s work, it appears to us that the argument is essentially taking about the Semiotics of Interface design. We discuss it further along with other theories in section 2.7.

Figure 3-2. This is not a button, adapted from Marsden (Marsden, 2007).

3.3 Directed dialog IVRs for Emergent users

3.3.1 Directed dialog IVRs

A directed dialog interactive voice response system (IVR) is a system where each dialog consists of a menu of finite number of choices at each level. In the most common interface, the user receives these choices in an audio prompt over a phone, and is promoted to indicate their choice through numerical keypress or touchtone. These user inputs are interpreted by the system to enable navigation through the interface hierarchy. A more deliberate definition of directed dialog IVRs is given by Acomb et al. (Acomb et al., 2007) as following:

A specific prompt enumerates all the possible reasons for a call, and the user would choose one of them.

Over the years, some of the touch-tone interfaces have been replaced by automatic speech recognition (ASR) based interfaces (Balentine & Morgan, 1999; Suhm, 2008). ASR based interfaces allow users to "speak out" their choices instead of choosing them from a menu, and are therefore considered more "natural" (Boyce, 2000). However, primarily for the reason that these technologies are still not available in languages spoken in developing countries, in this thesis we restrict our attention to touch-tone based IVRs.

Traditionally, IVRs have been in use for a variety of self-service customer care applications, typically as a triage before call centre operator provides manual service. As they present users with finite choices, IVRs may sound restrictive. But they are known to upgrade system usability and task completion rates by preventing users from losing track of interface features, functions and limitations (Roberto Pieraccini & Lubensky, 2005). For the same reason, Balentine (Balentine, 2007, p. 102; Balentine & Morgan, 1999, p. 146; Marics & Engelbeck, 1997) and Suhm (Suhm, 2008) insist on using directed dialog IVRs as relevant interfaces for novice (first-time) users.

3.3.1 IVRs for emergent users

Directed dialog IVRs are known to have particular advantages for emergent users. Directed dialog IVRs have ability to direct the users through a series of potential actions, and can lead to fewer misinterpretations of interface elements. IVRs do not require the user to read any text and are thus literacy-free, another advantage in case of emergent users. Further, as IVRs are known to benefit novice users, the same benefit to be transferred to emergent users. There have been several examples where usability issues of touch-tone IVR interfaces have been overcome in the context of emergent users (Joshi et al., 2014; Marathe, O'Neill, Pain, & Thies, 2015; N Patel et al., 2010; Rashinkar et al., 2011). Yet, most of these applications were of limited complexity and limited deployment in a research context. Widespread use of IVR interfaces for emergent users is yet to happen.

While IVRs are relevant interfaces for emergent users, their reputation is severely affected by poor usability (Kamm & Helander, 1997; Marics & Engelbeck, 1997; Resnick & Virzi, 1992; Tatchell, 1996; Yin & Zhai, 2005). Tatchell (Tatchell, 1996) states that IVRs based services are difficult to learn, easy to forget and confusing. Audio prompts, though may be explicit in directing users, are ephemeral and transient. Users must pay an attentive ear to the audio prompts presenting menu choices and system control features. This puts heavy demands on user's working memory while navigating through a sequential and hierarchical menu structure. Consequently, user interactions with directed dialog IVRs are temporal with 'poor referability' and 'absence of memory aid'. Recent studies with emergent users in focus (Grover et al., 2009; Joshi, Emmadi, et al., 2012; Joshi, Chakravarty, et al., 2012; N Patel et al., 2008; N Patel, Agarwal, Rajput, & Nanavati, 2009) have reconfirmed these usability difficulties.

In order to overcome these difficulties, researchers have suggested few approaches: (a) providing user control over the playback rate of the audio prompts while listening (Gardner-Bonneau, 1999, p. 207; Schnelle-walka, 2011), (b) use speech recognition to incorporate an open-ended dialog style (Plauche & Prabaker, 2006), (c) usage of effective metaphors in auditory interface design (Brandt, 2008; Dutton, Foster, & Jack, 1999), and (d) complementing audio prompts with coordinated visuals

(Yin & Zhai, 2005, 2006). This thesis focuses on 'supporting IVRs with visuals' to have more usable IVRs based audio interfaces.

3.4 Supporting IVRs with visuals

In this thesis, we are concerned with audio-visual interfaces. Audio-visual interfaces refer to a traditional IVRs with audio prompts based interactions and supported by the presence of visuals. Here, audio prompts carry out a directed dialog with the users by (sequentially) presenting the menu choices along with synchronized visuals. These visuals not only depict the same menu choices on the screen but highlight them as well. Additionally, the visual menu choices are already visible before they are called in an audio prompt, and stay visible even after they have been called out, till the user makes a choice. This makes audio-visual interfaces interactions "persistent" and free from temporality.

The idea of using visuals to support IVRs is not new. Fawcett and Brown (Fawcett, Blomfield-Brown, & Storm, 1998), in a 1998 patent illustrate the possibility of displaying menu information on the user's visual display unit or computer monitor as he interacted with a IVRs on a phone line. In their proposal, the user's phone and computer were two distinct devices. The user could not only receive menu information on both the channels but could choose either of the two to provide input. He could navigate "Up" and "Down" in the IVRs cascaded menu tree, and could skip intermediate steps if required. Interestingly the inventors proposed displaying two different boxes on the user's computer terminal. The left box was called a menu box while the right box was called a choice box (Figure 3-3). The menu choices appeared in the menu box and user's selection of menu items appeared in the choice box. Any step which marked the user going down in the menu added an entry to the choice box while a step going up deleted one. In case of the user wanting to revert to an earlier choice he could do so by selecting the associated entry from the choice box. This way he could exercise a 'menu choice' but only in one direction-from lower hierarchy menu to higher hierarchy menu. There was no extensive use of icons. Instead text was used to complement messages, information and calls to actions provided through audio prompts.

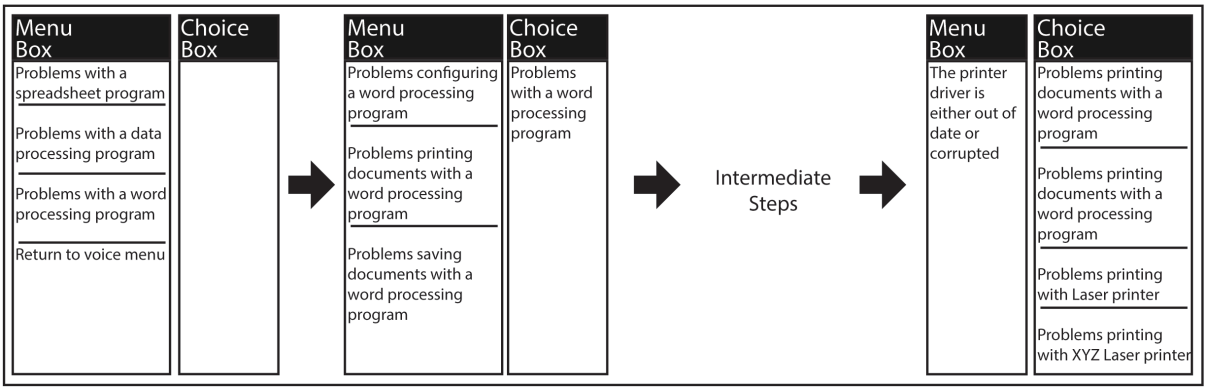


Figure 3-3. Menu box and Choice box, adapted from Fawcett and Brown (Fawcett et al., 1998).

A similar visual augmentation of audio prompts was detailed by Chandrasekhar (Narayanaswami, 2000) and Hiller (Rock & Hiller, 2002). Apart from certain upgrades in technology, their proposed interaction with the IVRs was similar to the one suggested by Fawcett and Brown (Fawcett et al., 1998). In their design, the users could move up and down across the entire width of the cascaded menus and could exercise skipping the menu in varying degrees.

Later Blankenhorn (Blankenhorn, 2008) and Haifeng (Haifeng Bi, 2013) proposed integration of both the audio and visual channel in a single device. Yin and Zhai (Yin & Zhai, 2005, 2006) provided evidence establishing merits of visually augmented audio IVRs over audio-only IVRs. They designed and tested FonePal - visually augmented audio IVRs for two different application types - Medical centre and IBM HR services. From a hardware standpoint, the FonePal still had used telephone and a computer screen separately for carrying audio prompts and visual menu respectively. However, what it did in particular was to provide means to visually browse and search the IVRs menu. The users had the ability to control the amount of menu they would like to see. They could choose to see the entire menu hierarchy, a menu and its submenu, or just a single menu. They could also provide a keyword to search through the menu. Yin and Zhai found that users had better task completion times and lesser error rate; and that they felt more satisfied after their interaction with the visually augmented IVRs. In spite of this variety of work, GUIs have continued to remain the predominant mode in HCI in the context of traditional users.

3.5 Menu hierarchies in IVRs based interfaces

Audio interfaces designers and researchers have been deeply concerned with menu hierarchies in IVRs design. It appears that IVRs menu design is initially perceived as relying on number of menu items in any given menu. Terms like 'deep', 'shallow' or 'broader' menu depth do not appear much in the guidelines. However lately the focus seems to shift towards levels of menu hierarchies present in the design. IVRs design guidelines for menu items propose a number close to 4 or fewer (Cohen, 2004; Miller, 1956; Suhm, 2008, p. 13); to as much as 9 menu items for interruptible menus (Balentine & Morgan, 1999, p. 159). An interruptible menu is the one where a user can make a selection while the audio prompt continues playing. To quote Marics and Engelbeck;

"*The majority opinion is that four items per menu is about right. If more than four menu items are offered, users can forget which items were offered, and which one they wanted. Although menus should be limited to four items, global commands such as Help and Exit need not be included in this count. Also, menus may have additional items that are not stated in the menu for expert users." (Marics & Engelbeck, 1997, p. 1087).*

This range of 4-9 is further analysed by Suhm et al.(Suhm, Freeman, & Getty, 2001). He studied the routing behaviour of the users as they interacted with two different version of a commercial automated customer support IVRs; one with a longer menu with seven specific and highly detailed choices and the other with a shorter menu with four broad choices. Their results indicate that users route themselves more efficiently using longer menus than shorter menus, provided items in the longer menu are specific and sufficiently detailed.

A similar scepticism on the number of menu items is expressed by Commarford (Commarford et al., 2008). In a comprehensive study, he argues that the guideline for keeping only four or fewer menu items in a single menu is often a misinterpretation of the working memory capacity (Miller, 1956). Rather a relatively recent version of working memory (Baddeley, 1981) suggests that it is also involed in processing and retrieval, along with storage of the information. Therefore, an IVRs user listens to an

audio menu with intent or a goal. He searches and opts for the choice that 'best' matches his goal. Such a selection can happen at any time during menu browsing; beginning, middle or towards the end of an interruptible menu. In his experiment, Commarford carefully crafted participant tasks involving both the storage and the processing of information and asked his participants to accomplish these tasks on two different versions of the IVRs enabled interfaces; one with a deep menu structure and the other with a broad menu structure. The results of the experiment indicated higher performance and satisfaction scores for the broad menu structure in comparison to the deep menu structure. His study seems to shift the emphasis of the discussion from number to menu items to number of menu levels. It illustrates that broader menu depths are preferable over deeper menu depths.

In a different context of designing GUI for mobile phone, Medhi et al. (I Medhi et al., 2013) favor using multi-page list of items organized in a four level deep menu hierarchy. Their results indicated that emergent users using multi-page list (40 items spread over 7 pages) performed better both in terms of time taken and percent correct than a 4-level deep menu with abstract categories.

We note an interesting point here. A deep menu IVRs is more likely to have content organized as categories or information chunks. Organizing information in chunks is most often a crucial part of designing applications. In fact, it can be said that with applications getting more complex, information chunking becomes not only useful but almost essential. Therefore, if IVRs menu design refrains from using deep menus in favor of shallow menus; creating services for different application domains in developing regions would seem less plausible. In this condition, we wonder if visual augmentation of IVRs audio prompts can help integrating deeper menu depth as well in IVRs designs, along with shallower menu depth. We conducted a systematic enquiry to weigh the effects of visuals, menu depths, and menu positions on IVR and audio-visual usage by emergent users. Chapter 4 reports thorough details of this study. In this study, we hypothesized that audio-visual IVRs would do better than audio-only IVRs in terms of enabling users to exhibit better task completion across different variations of menu depths and menu positions. We hoped that shallow menu depth would continue resulting in higher task completion than deep menu depth not only in case of

audio-only IVRs (Commarford et al., 2008; Suhm et al., 2001) but also for audio-visual IVRs. Talking of menu position, we believed that early menu position would result in higher task completion than late menu position.

3.6 Directedness and Persistence

In addition to menu hierarchies of IVRs and audio visual interfaces, this thesis also tests some fundamental theoretical perspectives with respect to audio-visual interface design for emergent users. This set of hypotheses are based on inherent and intrinsic merit of "audio" and "visual" as interface modalities which qualify the two for being used as complementary modalities in audio-visual interface design. Although the literature rarely speaks on "directedness" and "persistence", but we find few recurring voices (Balentine & Morgan, 1999; Shneiderman & Plaisant, 2005; Suhm, 2008, p. 4). It appears to us that any mention of persistence is closely linked with human cognition, in particular with the limits of human working memory.

3.6.1 "Directedness" of Audio

As discussed in section 2.3.1, a directed dialog is a system driven audio dialog where a user is presented with finite options (as menu) and is prompted to make a selection. It is seen that use of directed dialog would increase accuracy in the interaction (Acomb et al., 2007). Since the users are provided with explicit choices, there are lesser chances of committing errors during menu selection. One more preposition suggests that directed dialog is a preferred interface for first time or novice users (Balentine & Morgan, 1999, p. 218). Users with no knowledge of interface or task, may use a directed dialog with good enough chances of success. They get routed step-by-step through the entire interface across different menu hierarchies to the information terminal of interest. In this process, interface features and functionalities are slowly disclosed to them. User actions, even with a repeated use, are more selection oriented. We call this "directedness" of audio and deem it of specific advantage in desining audio-visual interfaces for emergent users.

While a directed dialog audio may well direct users towards task completion without being an expert in handling the system itself, it is known to have a major

problem. Literature identifies it as *problem of persistence* (Balentine & Morgan, 1999, p. 11). Balentine writes (Balentine & Morgan, 1999, p. 2),

With telephony interfaces, there is no margin of error. Unlike the computer screen that we can revisit when we're lost; unlike the bookmark that returns us to our previ- ous position; unlike the simple restaurant menu that confines our choices, the words transmitted across the telephony interface have no persistence.

Audio is inherently transient and temporal. It exists in time. What gets said at a given moment of time, can not be scanned again or listened parallely with others. Hence, listening to an audio menu requires attention and time. It may put load on short term memory (Suhm, 2008, p. 4) and may eventually lead to a break in user's interactions with the system.

3.6.2 "Persistence" of Visual

Visual, unlike audio, can stay in front of the users until he dismisses it, or a new data replaces it. Two or more visuals can be presented parallely. It expands in space. Unlike audio, users need not to remember visuals but can scan it at different times. It doesn't put any load on user's short term memory as it does not require users to memorize any data. Rather it works as memory aids when used in the interface. Visuals, when used in interface design, bring persistence in the interactions. To quote Balentine (Balentine & Morgan, 1999, p. 11), visuals when used in interface design brings the following:

Return to a task after interruption, Review—by scanning back and forth among several possible menu choices, Eliminate or minimize the effects of time by scrolling freely between the past and the present, and Maintain context—even when confronted with multiple tasks.

We therefore consider "persistence" of visuals (Shneiderman & Plaisant, 2005, p. 304) as an ability to present information parallely, non-temporally and decision making without taxing user's working memory. At the same time, visuals, typically used in graphical user interfaces, do not inherently direct users towards task completion. Audio, as we see, has complementary qualities. It can direct users step-by-

step through the interface but it lacks persistence. When used together in audio-visual interfaces for emergent users, we believe that these will complement each other as interface modalities.

3.7 Relevant theories for designing audio-visual interfaces

In this section, we identify certain basic theories which provide ground to our understanding of audio and visuals, and of their combined use in audio-visual interafces for emergent users. These are the information theory, the general model of human information processing, semiotics, dual code theory, and gestalt theory.

The information theory contends that the process of a user interacting with an audio-visual interface is a communication process (Gallager, 1968, p. 2). It starts with either the interface or the user initiating the dialog by transmitting a message to the other. However once initiated, both the interface and the user reciprocate to each other's call by transmitting further messages. This message transmission occurs through a channel – the interface modality. This setting can further be closely viewed as consisting of three processes: stimulus identification, response selection and response programming as per the general model of human information processing (Sears & Jacko, 2008, p. 29). In this framework, stimulus identification suggests processes aimed at the perception of information. Response selection suggests translation between stimuli and responses. And, response programming corresponds to the organization of the final output.

The response selection or translation phase, in particular, affects human performance (Hommel & Prinz, 1997; Wallace, 1971) and is known as stimulus response (S-R) compatibility. The errors or choice reaction times against an incoming stimuli is known to be dependent on (a) whether or not a set of stimuli shares one or more features with the response sets; and (b) the manner in which stimuli and responses are mapped onto each other. Better the mapping and the manner of coding, faster is the S-R translation. Or in other words, users would exhibit optimal performance in calibrating their responses against a set of stimuli if the stimuli and response sets are appropriately mapped and coded. We spot several mentions (Bayerl,

Millen, & Lewis, 1988; Chapanis & Lindenbaum, 1959; Proctor & Van Zandt, 1994) of S-R compatibility in context of GUIs but almost none in context of speech interfaces or IVRs.

Consider an usual IVRs prompt of the kind "*For banking related services, press 1*". One may argue that a mention of menu items works as stimuli against corresponding key presses as (expected) responses. In addition, one would note that their association is fairly arbitrary. Because it is spoken as an audio prompt, an explicit mapping or suggestion is introduced between the two. We call this the *directedness* of audio. In addition to establishing explicit mapping, *directedness* of audio attempt influencing users towards completing the task by prompting them to take actions in response. This is akin to persuading users to change their behavior or attitude or both without coercion or deception (Fogg, 2003).

Similar to Information theory, Semiotics in Human Computer Interaction interprets an interactive system as a communication artifact (De Souza, 2005). Both the designers and the users are seen as interlocutors. It is imperative for the designers to appropriatly convey messages about the artifact to the users to help them understand the same. The content of these messages is the intended use of the artifact by the users- what purpose does the artifact address, and how users can interact with it. Words, graphics, behavior, online help and assistive information are all different ways of encoding such messages. The underlying concepts here is signs. To quote Peirce, sign is *something which stands to somebody for something in some respect or capacity* (Peirce, 2011, p. 99). Note that Peirce is clearly indicative of a triadic model of sign (Figure 3-4) consisting of the representation (the representamen), its referent (the object) and its meaning (the interpretant). There is no necessary connection between the referent and the representamen. Critical is the role of interpretant who accepts a representation to mean what referant is trying to establish. The 'meaning' is therefore negotiated by the interpretant.

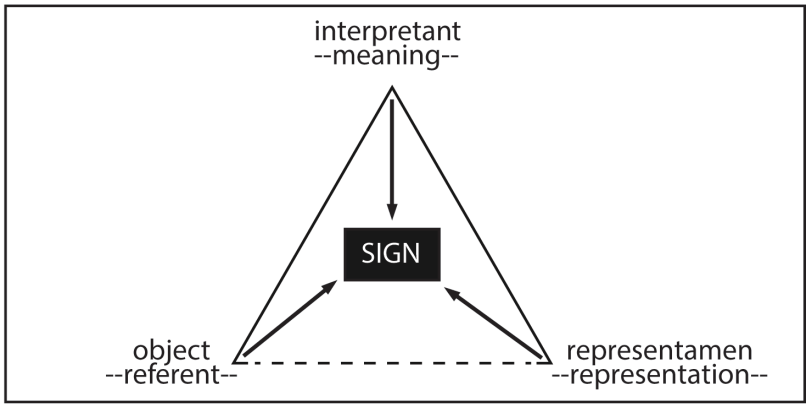


Figure 3-4. Peirce's Triadic model of Sign.

An important theory with respect to a combined use of audio and visual is the Dual Code theory (Paivio, 1990). It speaks specifically of imagery and verbal processes. It suggests that humans can simultaneously accommodate both visual and verbal thinking. And, imagery and verbal processes are handled cognitively by separate subsystems. Paivio further identifies that imagery is specialized in dealing with information about non-verbal objects and events; and that the verbal system is specialized in handling linguistic information. Additionally, the imagery and the verbal processes are independent but interconnected i.e. activities corresponding to both the systems can happen independently. The symbolic transfer of information from one system to another is possible. Apparently one of two systems can initiate or enhance an activity in the other system. This is of particular advantage in audio based interactions where working memory is a primary limitation in user interactions with the system. Audio being temporal is ephemeral, while visuals exhibit persistence by expanding spatially. A combined use of audio and visual would help increasing effective working memory and subsequently the user performance.

On similar lines, Gestalt theory would embrace a combined use of audio and visual in designing audio-visual interfaces. It would see individual modalities forming a unique and qualitatively different whole when combined (Oviatt et al., 2003). In particular, Gestalt principle of proximity (Koffka, 1999) states that an existence of temporal or spatial proximity would cause elements to be perceived as related. We know that audio, as a verbal symbol system, has temporal dimension. While the visual, as non-verbal symbol system, has spatial dimension but can also be presented temporally. Gestalt would reason that the common temporal dimension between these

two symbol system would cause these systems to be perceived as related. With a structurally identical content, a common temporal dimension between the two systems would enable proximity. And, one can vary proximity by inducing changes to temporality i.e. by varying timing between the two symbol systems. At it would be possible at some critical timing that both the audio and visual (as two different symbol system) would work as one unitary whole. This unitary whole, audio-visual, would have properties different from both audio and visual alone.

3.8 Conclusion

- There is a substantial growth in the understanding of emergent user population. From earlier known literature where a major parameter was the level of literacy, we now have an elaborate understanding of their identities in terms of other socio-economic markers as well. Along with formal education, lately we have noticed a growing emphasis on technology exposure in using ICTs (Devanuj & Joshi, 2013).

- We learn from available literature including field deployments that efforts in the direction of designing appropriate interfaces for emergent users are contantly ongoing. Since mid 1980s, these efforts have taken various shapes. Lately imagining IVRs as best-fit interfaces for emergent users is gaining widespread attention.

- Even when IVRs or similar audio based interfaces are attended for their perceived usefulness for emergent users, a list of long standing usability issues inherent to these interfaces also surface up. Although audio may direct users towards task completion, but simultaneously it poses unique challenges as interface modality on account of inherent temporality and a transient nature. While several strategies have been in vogue, visual augmentation of audio-based interfaces like IVRs has received relatively lesser research attention.

- Literature is inconsistent with respect to knowledge on menu hierarchies in audio interfaces like IVRs, and on relevance of existing guidelines in designing audio-visual interfaces for emergent users. We, therefore, require research efforts aimed at studying effects of changing menu depths, and with possibilities of realizing deeper menu depth along with shallower hierarchies in audio-visual interfaces for emergent users. We discuss our work relevant to this topic in chapter 4.

- While directedness and persistence can be theoretically highlighted as inherent qualities of audio and visual modalities when use in interfaces for emergent users, we have no substantial research evidence endorsing their combined use. We require research efforts aimed at comparing audio-visual interfaces with traditional audio-only IVRs and graphical user interfaces on grounds of directedness and persistence amongst emergent users. We discuss our work relevant to this topic in chapter 5.

Chapter 4 Effects of visuals, menu depths, and menu positions on IVR usage by emergent users

4.1 Introduction

In this chapter, we talk specifically about menu depths, use of visuals and variations in the positions of the menu-items with respect to audio-only and audio-visual interface design. We realize that there is an absence of any systematic study done earlier with emergent users which speaks of above mentioned variables and their effects on users' performance. Besides, as we see in chapter 3, there lies inconsistency in the literature on the number of items in a given IVRs audio menu.

"*Four or fewer is better* (Cohen, 2004; Marics & Engelbeck, 1997, p. 1087; Suhm, 2008, p. 13)*, or upto 9 in an interruptible menu (Balentine & Morgan, 1999, p. 159)"* says most of literature on IVRs design. The reason put forth is the limitated capacity of human working memory (Miller, 1956). Commarford (Commarford et al., 2008), dismisses this as a misinterpretation of human working memory. In an updated understanding (Baddeley, 2003), working memory is a far more active system responsible for storage, retrieval, rehearsal and processing of verbal and visuospatial information. Based on this approach, Commarford has demonstrated higher performance and satisfaction scores for the broad menu structure over deep menu structures.

In our own understanding, these developments, with respect to number of items in an audio menu to a preference for using shallow menu depths, are still limiting in their scope. These yield no explicit suggestion when we attempt imagining menu depths for audio-visual interfaces for emergent users. From a designer's perspective, we hope that the choice of using either deep or shallow menu depths in audio-visual interfaces for emergent may well be function of domain and user requirements. Hence, our main motivation is to see if addition of visuals to audio-only interfaces can help realizing both deep and shallow menu-depths.

The rest of the chapter is organized as follows: We follow with a mention of hypothesis with details of design and development of the test prototypes based on agriculture market service. This explains our systematic approach for creating content and the interface components of the IVRs prototypes. Next we write a comprehensive account of the method used in the experiment. It covers details related with participants' recruitment, the experimental protocol and the data collected. Finally, we list results of statistical analysis and end with a summary of our work.

4.2 Hypothesis

The pivot of this study rests on the visual augmentation of audio-only IVRs to result in a more desirable audio-visual IVRs with respect to menu-depths. We, therefore, hypothesized that audio-visual IVRs would do better than audio-only IVRs in terms of enabling users to exhibit better task completion across different variations of menu depths and menu positions. For us, this had been a central hypothesis which had driven this study. That is to say,

Hypothesis 1 (H1)*: The audio-visual IVRs performs better than audio-only IVRs.*

Now we speak in terms of variation in menu-hierarchies as exercised in the current study. We had test prototypes which not only varied on account of use of visuals, but also in terms of underlying menu-hierarchies. To this effect, we believed that prototypes with shallow menu-depth would continue yielding better results than prototypes with deep menu-depth. This belief is deep rooted in the IVRs literature

which endorses use of shallow menu-depth over deep menu-depth in audio based interfaces (Commarford et al., 2008; Suhm et al., 2001). We hypothesized that our shallow menu-depth variants of test prototypes, both audio-only and audio-visual, would continue following directives as mentioned in the existing literature. That is to say;

Hypothesis 2.1 (H2.1)*: The audio-only-shallow IVRs performs better than audio-only-deep IVRs.*

Hypothesis 2.2 (H2.2)*: The audio-visual-shallow IVRs performs better than audio-visual-deep IVRs.*

At last we include a mention of the third variable, the position of the menu-item in a given menu, in order to detail the third pair of hypotheses. Common sense dictates that while listening to a series of menu items in a transient audio menu, participants would perhaps choose the (desired) one if appearing early on over the scenario when it appears later in the menu. This thought appealed to us while looking at how users make menu selection in audio-based interfaces. As discussed in chapter 3, users listen to an audio menu with a specific intent or goal (Baddeley, 1981). They would opt for the menu item that best fits their goal. Hence if this were happening, early positions of desired menu items in our prototype variants, both audio-only and audio-visual, would result in better performance of the users than late positions of menu items. That is to say;

Hypothesis 3.1 (H3.1)*: The audio-only IVRs exhibits better user performance for early position of the menu items than for the late position.*

Hypothesis 3.2 (H3.2)*: The audio-visual IVRs exhibits better user performance for early position of the menu items than for the late position.*

4.3 Design of test prototypes

4.3.1 Content

We chose information content on agricultural commodity market pricing for this study in order to suit the interest for our participants. The content was collated from online repositories maintained by state led institutions and other accredited organizations ("Agricultural Marketing Information Network," 2014, "Agriwatch - Commodity Prices India,Commodity Markets," 2014, "State Agricultural Produce Marketing Board, Uttar Pradesh," 2014). We collected information on daily pricing reports including market trends against five or more varieties of grains, vegetables, and fruits. We also included miscellaneous other produce including seasonal crops, cash crops and animal fodder. In order to contextualize the content and to make it more relevant for our participants, we consulted a group of farmers with the same demographic details as of the participants of the study. They gave invaluable insights into their selection of crops which helped us scale down the content to a manageable volume. They mentioned their preferred market locations and their reasons for choosing one location over the other. They also suggested colloquial terms of several commodities, which we used later in our audio scripts.

4.3.2 Organization

We designed a total of four different test prototypes for this study (Table 4-1). This included two audio-visual prototypes and two audio-only prototypes with menu-depths of 3 and 5 respectively. In case of both *audio-visual-shallow* (Figure 4-1) and *audio-visual-deep* (Figure 4-2) prototypes, visuals were used along with audio prompts. Users consistently saw either a picture or text corresponding to different menu items spoken in audio prompts. The other two, *audio- only shallow* and *audio-only deep* prototypes mimicked conventional *audio- only* IVRs (Figure 4-3).

Table 4-1. Test prototype design with variation in use of visuals and menu depths.

Test Prototypes	**Audio-Visual**	**Audio-only**
Shallow (3-level menu depth)	Audio-visual shallow	Audio-only shallow
Deep (5-level menu depth)	Audio-visual deep	Audio-only deep

4.3.2.1 Audio-visual shallow

An audio-visual shallow prototype had only 3 levels of menu depths (Figure 4-1). We, therefore, had to organize the entire content of the agricultural commodity market pricing service to fit within only 3 levels of menu depth. We presented menu items as discrete large sets of entities spread across the entire menu. The 1st level menu consisted of a long list of 9 agricultural commodities: peas, potato, wheat, jwaar, rice, cashewnuts, papaya, fodder and chickpeas. These menu items were randomly placed next to each other to eliminate possibilities of category formation. The 2nd level menu consisted of sub-varieties of commodities presented at level 1. At the last and 3rd level of menu-depth, the prototype mentioned selected commodity sub-variety's pricing and its trend across three different time-stamps for three different market locations. Figure 4-1 shows this workflow for a sub-variety of peas. At level 1, user selects peas. This selection is followed by a mention of sub-varieties of peas: *Rachna*, *Muchchad*, *Pusa 10*, *PK3* and *Azaad C1* at level 2. He chooses *Muchchad* peas at level 2. Subsequently at level 3 of menu depth, he finds *Muchchad* peas' pricing details for three different markets (Delhi, Kanpur and Jhansi) across three different time stamps (last week, yesterday and today). For a detailed account of audio prompts used in this prototype, and the associated call flow, see appendix 1 and 2.

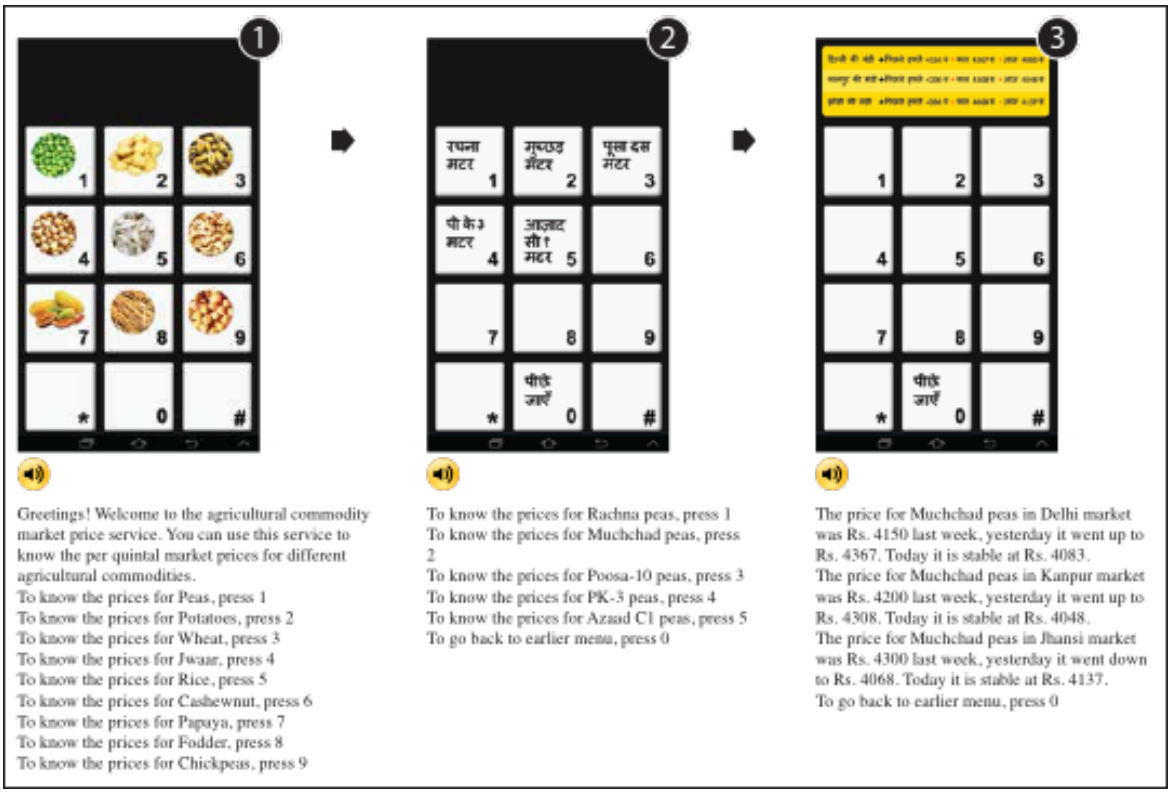


Figure 4-1. Audio-visual shallow prototype.

4.3.2.2 Audio-visual deep

An audio-visual deep prototype had 5 levels of menu depths (Figure 4-2). We organized the entire content of the agricultural commodity market pricing service to spread across 5 levels of menu depth. The 1st level menu consisted of a short list of 3 content categories of "grains", "fruits" and "vegetables". The 2nd level menu consisted of contents of these categories. The 3rd level menu further presented sub-varieties of menu items presented earlier at level 2. The 4th level menu mentioned the market locations. Towards the end at the 5th level, the prototype mentioned selected commodity sub-variety's pricing and its trend across three different time-stamps for a selected market location. Figure 4-2 illustrates this workflow for *Basmati* rice. At level 1, user selects "grains". He is then shown contents of the category "grain" at level 2. These are rice, wheat, chickpeas and peas. He selects "rice" at this level to move to see its sub-varieties at level 3. Out of the given sub-varieties, he selects *Basmati* rice to navigate to level 4 where he is shown three different markets locations. He opts for *Kanpur* market at this level. On this selection, he navigates to level 5 to find *Basmati* rice's pricing details for *Kanpur* market across three different time stamps (last week,

yesterday and today). For a detailed account of audio prompts used in this prototype, and the associated call flow, see appendix 3 and 4.

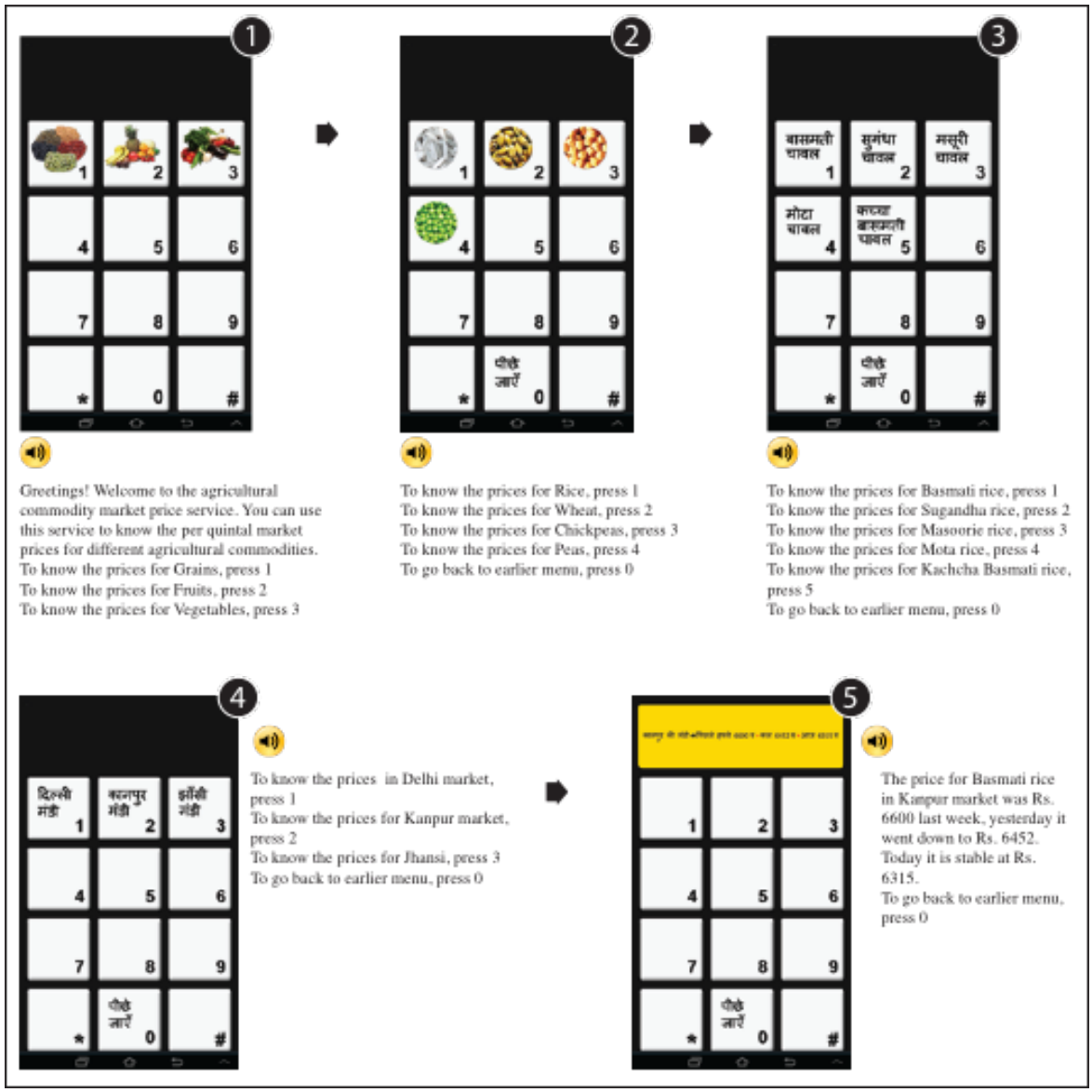


Figure 4-2. Audio-visual deep prototype.

4.3.2.3 Audio-only shallow and audio-only deep

Both the audio-only prototypes, *audio-only shallow* and *audio-only deep*, were designed with audio prompts identical to *audio-visual shallow* and *audio-visual deep* prototypes respectively. However, they were designed to imitate conventional IVRs with no visuals at all. The only visual users' encountered during their interactions with audio-only prototypes was of a touch sensitive number pad (Figure 4-3). They used it provide their inputs to the system through *keypresses*.

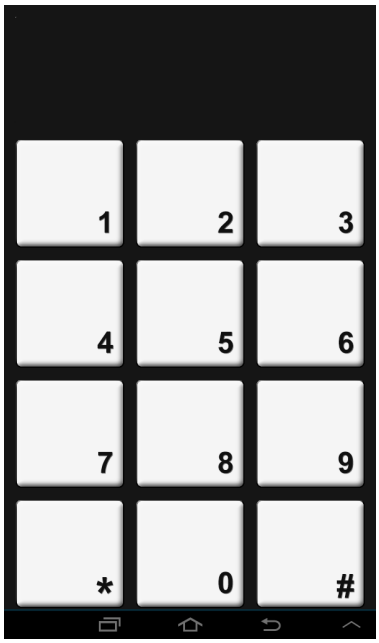


Figure 4-3. Audio-only shallow and audio-only deep prototypes.

4.3.3 Prototyping Environment and hardware used

The agriculture market service was prototyped using Eclipse ("Eclipse," 2012), an open source Integrated Development Environment (IDE) for Android applications. The code was written in Java. It supported flexible alterations and arrangement of components to form different interface elements. The prototypes could function on any tablet or an android based touch screen mobile phone. However, the preferred usage suggested using a screen resolution of 1024 x 600 pixels with android 4.1 (or later) operating system. The prototypes had ability to record data as text files.

4.4 Method

4.4.1 Experiment design

The experiment followed a mixed method design with three different independent variables: visual augmentation (*audio-visual* vs. *audio-only*), menu depth (*shallow* vs. *deep*), and menu position (*early* vs. *late*). The visual augmentation was a between-group factor while the rest were within-group factors. This resulted in two different groups of participants: Group **AV** with *audio-visual* IVRs prototypes and Group **A** with *audio-only* IVRs prototypes. Within each of groups **AV** and **A**, we realized four different variations of menu depth and menu position by means of

creating four different tasks. These were shallow-early (SE), shallow-late (SL), deep-early (DE) and deep-late (DL).

4.4.2 Pilot studies

As we did not have any bench data about such studies, we conducted three different pilot studies prior to the final study. These pilot studies had different scopes and were done sequentially based on need accessed at various stages of conceptualization, prototyping and testing. In this way every subsequent pilot study in this series helped experimenters moved incrementally towards the final study by improvising (and verifying) on their objectives, test prototype designs, experimental protocol and analysis of outcomes.

The first pilot was conducted in May 2013. We were mainly interested in evaluating feasibility for one such IVRs amongst the intended user population. The other motivations were to get a first-hand experience of the testing environments, strategies to sample users, refinement of the test tasks and their content, test prototypes and of the experiment protocol. It included a total of 20 farmers who volunteered as participants. They belonged to 5 different villages located in the district Gwalior of the state of Madhya Pradesh, India and were selected through convenient sampling. The test prototype was an agriculture market IVRs with ability to inform farmers of the price of a limited number of agricultural commodities. The prototyped IVRs although had some limitations. It could only suggest prices for the current day and for only one market location. Note that this was an audio-only IVRs prototype with no visual augmentation. It imitated contemporary IVRs design features with barge-in facility and hierarchical navigation. It was supported on a cloud telephony system supported by an enterprise called Exotel ("Exotel - Exotel, A Business Phone System from the Cloud," 2013).

The protocol followed during the first pilot study consisted of four different stages: (a) collection of profile details of the participant users, (b) briefing participants about the objectives of the study and the training task, (c) participants attempting test tasks on test prototype, and (d) a post-test open ended interview of the participants by experimenters to receive their feedback and to probe them for a particular response

behavior during the testing sessions. A stopwatch was used to record the *task completion time* by the experimenters. Along with this the experimenters kept making notes of any specific behavior shown by the participant during the prototype testing sessions. We found that the study yielded sufficient confidence in the usefulness of systems akin to the one realized in the test prototype for the farming community. We saw our participants being quite curious about trying the test prototype and about the agricultural commodity pricing system overall. We also realized that we needed to improve on our brief to the participants about the test tasks and the test prototypes. Most often even after careful briefing, they mistook our test prototype for a real system. At multiples instances the participants went around asking for the price of the commodity, which they had been growing in their fields. Consequently, they ended up in disappointment when the system proved incompetent against meeting their needs. We believe that this study enabled us to identify possible failures in our testing protocol, as well as in our method of recruitment of the participants. We learnt the importance of including commodities in the test prototypes which farmers would usually growing in their fields when we approach them for testing sessions. In addition, we realized that we should approach the participants possibly through a facilitator to gain trust of the user community. We made use of this knowledge gained from first pilot to design and conduct another pilot study.

The second pilot study was done in August 2013. We included a more relevant set of agricultural commodities and market locations in the test prototypes. As a result, the test prototypes presented opportunities for participant farmers to find and explore not simply one specific commodity (as mentioned in the test task) but other commodities as well. We designed two different versions of the test prototypes – one with a menu depth of 3 levels and the other with a menu depth of 5 levels. We approached farmers from three different villages in the district Ujjain of the state of Madhya Pradesh, India. A total of 35 participants were recruited. Noise-free environments whenever possible were preferred over conducting testing sessions in the field settings. All the trials were conducted either in a warehouse room, in a classroom or in a bank's reception room. We sought help from individuals with constructive influence in these villages – a cement dealer, a staff of district education committee and a banking officer - as session facilitators. They played a vital role by introducing

the experimenters with the community and in establishing the necessary rapport and trust. The experimenter would then interact with the participants individually. At several times the facilitator helped the experimenters in briefing the participants and suggesting them of different test tasks that they had to perform. Cautious of the observation from earlier pilot study that the participants had a tendency to look for commodity currently growing in their field and not the one mentioned in the task; the experimenter laid strong emphasis on explaining the participants the nature of the test prototypes, and how their contribution would help designing a better system in the future. This helped aligning participants' focus in the test tasks.

The test prototypes were again developed on a cloud telephony support provided by Exotel ("Exotel - Exotel, A Business Phone System from the Cloud," 2013). The test prototypes were still in the form of audio-only IVRs and essentially differed in terms of the menu-depth. We recorded task completion time taken by the participants in testing sessions. Our finding suggested that users exhibited better task completion rates while using IVRs with shallower (3 levels) menu depth than a deeper (5 levels) menu depth. This was consistent with the existing guidelines (Balentine & Morgan, 1999; Marics & Engelbeck, 1997) that a shallower menu-depths be realized for better user performance over deeper menu-depths.

We presented these findings to an expert peer review committee. The recommendation of the committee favored using an additional independent variable – the variation in the position of the menu item along with the menu-depth. This was suggested because it appeared to us collectively that the menu item position might play a critical role at the time of users selecting a specific menu item from a sequential menu. Following this, we also realized in retrospection that we could start analyzing the use of visuals along with variations in menu-depth and menu-item position. Corresponding the number of dependent variables was increased to include task success score, choice error count and menu-repetition count along with task completion time in the study. In this way the second pilot paved way for the current study.

The third pilot included hereby helped us fine-tune the current study as a mock trial of our final test prototypes. A total of 20 users were recruited and included people

working in the university campus. Although none of the users from the field were recruited here but those selected had similar profiles in terms of age, education and technology experience as the participants in the current study. Our main aim was to trouble shoot minor issues occurring with the design of the test prototypes and with the experiment method. This pilot did help us revealing certain essential flaws in the design of the test prototypes and with the protocol, which might have caused bigger problems if ignored. For example, prior to the pilot we did not hard coded our prototypes to record users' journey and time stamps. But we realized that we were making errors in recording data. We therefore card coded data recording capabilities in our test prototypes following this pilot.

4.4.3 Protocol

For the final study, each testing session began with the experimenter informing the participants about the objectives and goals of the study. They received information on their role during different stages of the session.

4.4.3.1 Stage I: Training (15-20 min.)

We used a touch-based tablet with test prototypes because of a larger real estate to incorporate visuals (Figure 4-3). As participants were new to tablets, we provided them with a standardized training. We designed two distinct training tasks to neutralize individual differences in tablet use. Sufficient care was taken for these training tasks to remain distinctly different from the actual tasks of the study. In the first task, the training prototype prompted participant to press a certain key. For example, if the audio prompt suggested "Press Seven", the participant was expected to touch '7' on the tablet's soft keypad. We designed the training prototype to commend the participants in case he pressed a correct key e.g. "Well done, you have pressed the right key". Otherwise he received feedback suggesting that he had pressed a wrong key, and was asked to try again. This feedback suggesting a right or a wrong key-touch was both played as an audio file as well as shown on tablet's to the participant. The training prototype also changed the color of the key-touch to green or red suggesting either a 'successful touch' or a 'failed touch' respectively (Figure 4-4). Typically, we asked participants to repeat this interaction couple of times with audio prompts

mentioning numbers in a random order. In the second training task, we required participant to identify categories of objects and their contents. We displayed an image consisting of more than one object of the same kind for the entire duration of the training session. Against this image, the training prototype asked three different questions from the participant. First he was asked to identify the object category e.g. utensils, electrical or agricultural equipment etc. On identifying the correct object category, he was asked to name the object. Finally, the IVRs asked him to indicate the number of objects in the image e.g. 1,2 or 3. Every question played to him as an audio prompt consisted of possible choices as answers. The participant pressed number keys associated with different choices to register his response. He received commendation for every correct choice and additional trial suggestions for wrong choices. Once fluent with the training task the participant was asked to take a short break of 2-5 minutes. The training helped familiarizing participants with the experimenter's device input methods and with the hierarchical nature of the test prototypes.

Figure 4-3. Participant during testing session.

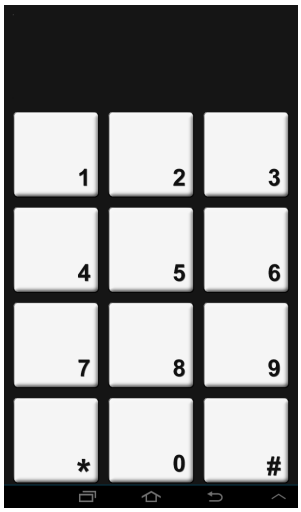
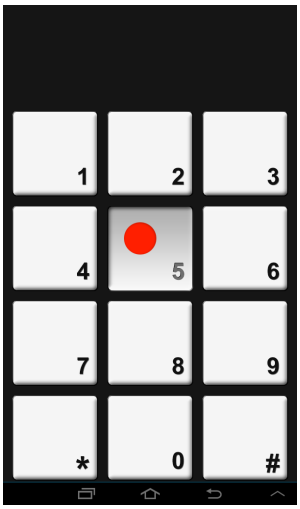
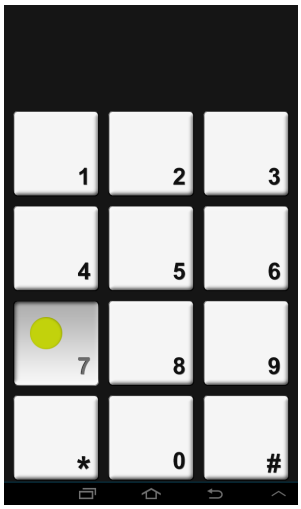
Training task	Wrong key touch	Right key touch
		
Audio prompt; *Please touch the number key 7 on your tablet.*	Audio prompt; *You have touched the wrong key, please try again.*	Audio prompt; *Well done, you have touched the right key.*

Figure 4-4. Training task 1 given to the participants.

4.4.3.2 Stage II: Test tasks trial

Each participant performed four test tasks. The tasks required participants to find rates for a specified agricultural produce amongst at a specified market location. Each participant was randomly assigned either to group **AV** (*audio-visual* IVRs) or to group **A** (*audio-only* IVRs). In each group, two tasks involved the use of the shallow menu depth prototypes, and two others involved the use of deep menu depth prototypes. Every participant in each group, therefore, performed tasks indicating shallow-early (SE), shallow-late (SL), deep-early (DE) and deep-late (DL). The task sequence was randomized to counter any learning effect. To eliminate the possibility of a participant forgetting the task, each task was printed on a sheet and was kept in front of the participant for the entire duration of the testing session.

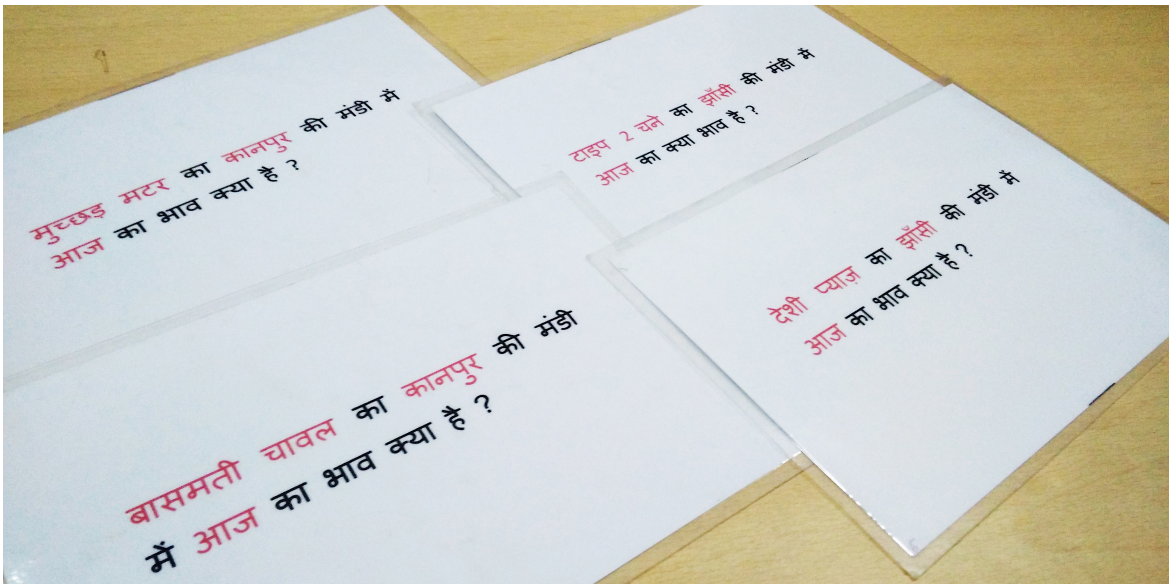


Figure 4-5. Printed cards corresponding to four test tasks.

Following completion of the test tasks, the participants were individually debriefed. They were often probed for their specific responses and were encouraged to contribute their suggestions, if any.

4.4.3 Participants and test environment

A total of 88 users (M=35.25 years; SD=3.8; Range=30 – 40 years) voluntarily participated in the study. The individuals belonged to a cluster of villages namely *Saalon, Bharroli, Khiriya-Riyasat, Tatarpur* and *Tatarpur- Patharya,* in the sub-district area *Bhander*, Madhya Pradesh, India. There were 8 individuals with 4th grade literacy, 17 with 5th grade literacy, 10 with 6th grade literacy and 53 with 7th grade literacy. In terms of occupation, 54 of these individuals owned agricultural land while the remaining 28 indirectly practiced agriculture as seasonal laborers. Only 6 users were female, and rest 82 were male participants.

The participant pool was created through a voluntary gathering around the enclosed premises provided by different village heads either in their own houses or in places like primary schools or community hospitals. To achieve this, the investigator contacted village heads, teachers posted in village schools, and members of group of farmers called ‘friends of farmers’- appointed by District Agriculture Officer, in advance to spread words about the study.

Each participant was individually screened and tested. We dropped 28 participants due to one of the following reasons: a literacy level which was either less than 4th grade or more than 7th grade, willing withdrawal from the study in between the testing sessions, partial completion of the training sessions, or interruptions by others (specifically in case of the 6 female participants, testing sessions were constantly interrupted by family members). Finally, 60 users (30 in group **AV** and 30 in group **A**) participated in the study.

4.4.4 Data collection

The IVRs prototypes were programmed to record and save quantitative data as text files for each of the dependent variable (Table 4-2) on exit. The data recorded revealed 'journey' details along with 'start', 'end' and 'date' time stamps individually for all the participants. In addition to this the experimenter had opportunity to record profile details, observation notes and retrospective accounts of participant's interaction during the testing sessions. Given the nature of the test tasks and the tech savviness of users, it is more critical to complete a task successfully. Hence in the analysis to follow, task success score is taken as the variable driving the analysis. In addition, we also analyzed task completion times, errors and menu repetition.

Table 4-2. Study variables.

Variable	Description
Task success	Determines the successful completion of the task by the participant. Task success is a dichotomous variable. A score of "1" is assigned for successful task completion otherwise "0".
Task time	The task completion time (in seconds) was counted from the moment of attempting the task trial to the moment when the trial completes.
Choice error	A choice error was said to have occurred at the moment when participant makes a wrong choice during menu selection.
Menu Repetition	Number of times a particular audio menu is played again.

4.5 Results

We begin with both between-group (**AV** vs. **A**) and within- group (individually for **AV** and **A**) analysis for task success scores, and then follow with a within group analysis for the other three dependent variables to gather supportive evidence.

4.5.1 Task success

4.5.1.1 Between group analysis

We plotted a frequency bar graph (Figure 4-6) between type of tasks (SE, SL, DE and DL) on X- axis and the task success on Y-axis. Note that vertical bars indicated task success count for a total of 30 participants in each group individually. The graph suggested presence of a larger number of successful participants in **AV** than in **A** across all the four different tasks SE, SL, DE and DL.

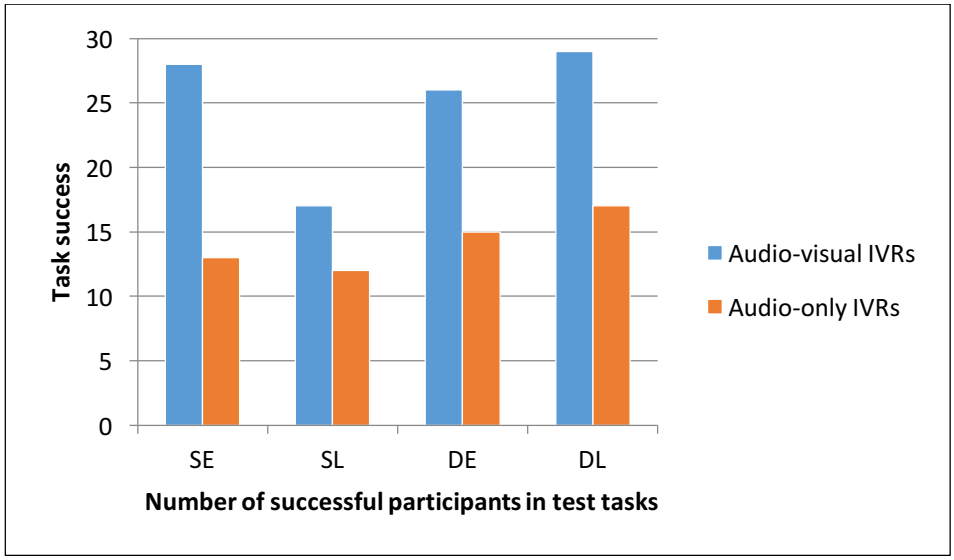


Figure 4-6. Task success between the two groups across 4 test tasks.

We plot another frequency bar graph (Figure 4-7) between variations in menu depth and menu item positions (shallow, deep, early and late) on X-axis, and task success on Y-axis. We obtained task success counts for both group **AV** and **A** by combining task success for tasks SE, SL, DE and DL collectively as shallow (SE and SL), deep (DE and DL), early (SE and DE) and late (SL and DL). We found that participants in group **AV** still scored better than group **A** even when task success

scores were collective grouped and compared for variations in menu depth (*shallow, deep)* and in menu-item positions (*early* and *late*).

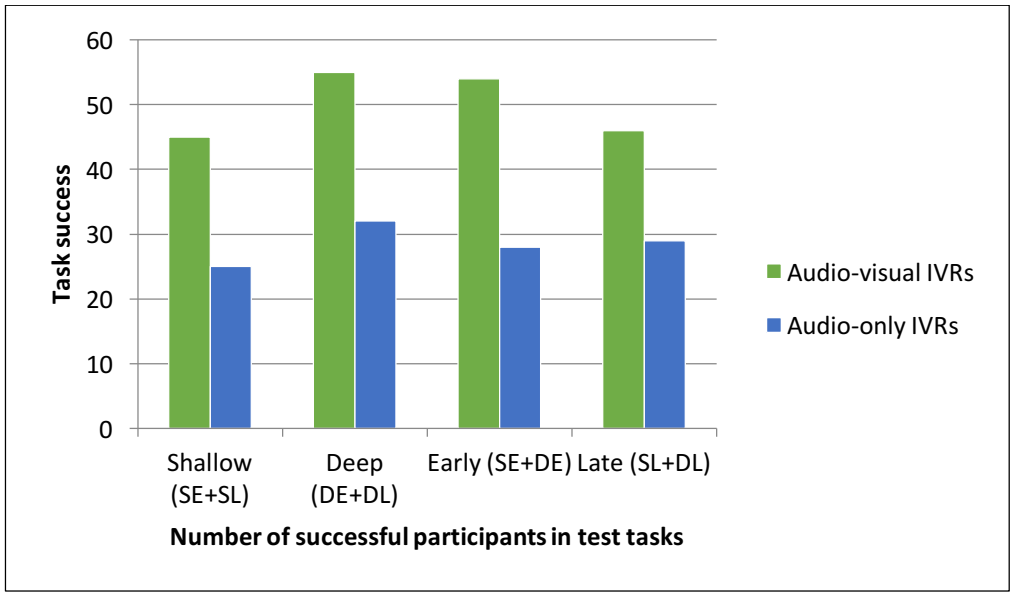


Figure 4-7. Task success between the two groups across 4 test tasks w.r.t menu depth and menu item position.

Task success, as mentioned earlier, was a dichotomous variable. We therefore used *two-independent proportions test* to establish the significance of our observations corresponding to task success in groups **AV** and **A** (Kurtz, 1983, p. 186). For each of task SE, SL, DE and DL, a *two-independent proportions* z-statistic was calculated (Table 4-3). The results were suggestive of significant difference between task success across groups **AV** and **A** for tasks SE, DE and DL. However, in case of SL the difference between task success across groups **AV** and **A** was only marginal. Even then the direction of results, in task SL, was in favor of group **AV** participants.

Table 4-3. 2- Independent proportions test between AV and A. n1, n2=30.

Task (N=30)	Task Success AV		Task Success A		Z	p-value
	Count	%age	Count	%age		
SE	28	93.3	13	43.3	4.16	0
SL	17	56.7	12	40.0	1.29	0.1
DE	26	86.7	15	50.0	3.05	0
DL	29	96.7	17	56.7	3.66	0

We further conducted another series of *two-independent proportions test* to establish significance of difference between groups AV and A across variations in menu depth (shallow vs. deep) and menu position (early vs. late) taken together across all the four tasks SE, SL, DE and DL (Table 4-4). We calculated task success counts corresponding to shallow and deep menu depths, and early and late menu positions by adding up the task success scores of corresponding pair of tasks. For example, in group **AV**, a task success counts of 45 for shallow menu depth was calculated after adding task success counts for tasks SE (28) and SL (17). We found that the participants of group **AV** scored significantly better task success than participants of group **A** across variations of menu depths (*shallow* vs. *deep*) and of menu positions (*early* vs. *late*). We therefore have enough evidence to support our hypothesis (H1) that *audio-visual* IVRs performs better than *audio- only* IVRs in terms of yielding higher task success scores.

Table 4-4. 2- Independent proportions test between AV and A. n1, n2=60.

Task (N=60)	**Task Success AV**		**Task Success A**		**Z**	**p-value**
	Count	**%age**	**Count**	**%age**		
Shallow (SE+SL)	45	75.0	25	41.7	3.7	0
Deep (DE+DL)	55	91.7	32	53.3	4.7	0
Early (SE+DE)	54	90.0	28	46.7	5.1	0
Late (SL+DL)	46	76.7	29	48.3	3.21	0

4.5.1.2 Within group analysis

We used a *two dependent proportions* test for within group analysis across three different groups: group **AV** with N=30, group **A** with N=30, and group **AV** and **A** taken together with N=60. Note that this test (Kurtz, 1983, p. 190) applies to dichotomous variables only, and is based on 2X2 contingency tables with change cells. The *two dependent proportions test* decides the significance of the change (across cells). For example, in the context of current within group analysis where task success was a known dichotomous variable, a *two dependent proportions test* carried out across menu depths “shallow” and “deep” would evaluate the significance of within group “change” occurred when participants tried shallow and deep menu depths prototypes sequentially in a certain order. In addition, it would be pertinent to make a mention of how we were calculating task success counts in this ongoing analysis. As

an illustration, consider the task success count of 16 for shallow menu depth in group **AV** (Table 4-5). The tasks with shallow menu depths were SE and SL. Their individual task success counts were 28 and 17 respectively (Table 4-3). Now, while calculating task success count for shallow menu depth, the *two dependent proportions test* dictated us consider not the sum of SE (28) and SL (17) which was 45. Rather, we considered only the task success count given by the participants who were successful in both the tasks SE and SL. This gave us a number of 16 task successes in shallow menu depth for group **AV**. Similar comments can be made regarding task success counts in *within group analysis* across all different forthcoming tables here.

We found that within group **AV** (Table 4-5), the *deep* menu-depth had resulted in significantly higher task success than the *shallow* menu-depth. However within group **A** (Table 4-6), the *deep* menu depth had resulted in marginally higher task success than the *shallow* menu depth.

Table 4-5. 2- Dependent proportions test within AV for Shallow and Deep menu depths.

Task pair (N=30)	Task Success Shallow		Task Success Deep		Z	p-value
	Count	%age	Count	%age		
Shallow vs. Deep	16	53.3	25	83.3	1.94	0.03

Table 4-6. 2- Dependent proportions test within A for Shallaw and Deep menu depths.

Task pair (N=30)	Task Success Shallow		Task Success Deep		Z	p-value
	Count	%age	Count	%age		
Shallow vs. Deep	6	20.0	12	40.0	1.58	0.06

Within-group comparison between *shallow* menu depth and *deep* menu depth for all the users taken together (Table 4-7) suggests that *deep* menu depth had resulted in significantly higher task success scores than *shallow* menu depth. Hence for *audio-only* IVRs we do not have enough evidence to reject hypothesis (H2.1) that *shallow* menu depth performs better than *deep* menu depth, but note that the direction of the scores was in favor of *deep* menu depth with calculated p-value (0.06) only slightly greater than critical value of 0.05. However, for *audio-visual* IVRs, results indicated that we cannot but reject the hypothesis (H2.2). In this case, *deep* menu depth resulted in significantly better task success scores than *shallow* menu depth.

Table 4-7. 2- Dependent proportions test within (AV+A) for shallow vs. deep menu depth.

Task pair (N=60)	Task Success Shallow		Task Success Deep		Z	p-value
	Count	%age	Count	%age		
Shallow vs. Deep	22	36.7	48	80.0	4.06	0

Within group **AV** (Table 4-8), participants trying *early* (SE and DE) menu position scored marginally higher task success than those trying *late* (SL and DL) menu positions. In addition, participants trying SE task scored significantly better than those trying SL task (Table 4-9). However, the trend reversed both in significance of result and its direction when we compare participants trying DE and DL tasks (Table 4-10).

Table 4-8. 2- Dependent proportions test within AV for Early and Late menu positions.

Task pair (N=30)	Task Success Early		Task Success Late		Z	p-value
	Count	%age	Count	%age		
Early vs. Late	24	80.0	17	56.7	1.46	0.07

Table 4-9. 2- Dependent proportions test within AV for SE and SL test tasks.

Task pair (N=30)	Task Success SE		Task Success SL		Z	p-value
	Count	%age	Count	%age		
SE vs. SL	28	93.3	17	56.7	2.77	0

Table 4-10. 2- Dependent proportions test within AV for DE and DL test tasks.

Task pair (N=30)	Task Success DE		Task Success DL		Z	p-value
	Count	%age	Count	%age		
DE vs. DL	26	86.7	29	96.7	0.89	0.19

Within group **A** (Table 4-11) participants trying *early* (SE and DE) menu position scored marginally lesser task success than those trying *late* (SL and DL) menu positions. Also, participants trying SE task scored marginally better than those trying SL task (Table 4-12). However, yet again, the trend reversed its direction when we talk of participants trying DE and DL tasks (Table 4-13).

Table 4-11. 2- Dependent proportions test within A for Early and Late menu positions.

Task pair (N=30)	Task Success Early		Task Success Late		Z	p-value
	Count	%age	Count	%age		
Early vs. Late	9	30.0	10	33.3	0	0.5

Table 4-12. 2- Dependent proportions test within A for SE and SL test tasks.

Task pair (N=30)	Task Success SE		Task Success SL		Z	p-value
	Count	%age	Count	%age		
SE vs. SL	13	43.3	12	40.0	0	0.5

Table 4-13. 2- Dependent proportions test within A for DE and DL test tasks.

Task pair (N=30)	Task Success DE		Task Success DL		Z	p-value
	Count	%age	Count	%age		
DE vs. DL	15	50.0	17	56.7	0.35	0.36

For a combined population of participants (**AV+A**), participants trying *early* (SE and DE) menu position scored marginally higher task success scores than those trying *late* (SL and DL) menu positions (Table 4-14). We, therefore, do not have sufficient evidence to support hypothesis H3.1 and H3.2. Users trying e*arly* menu positions were significantly not exhibiting better task success than those trying *late* menu positions.

Table 4-14. Dependent proportions test within (AV+A) for Early and Late menu positions.

Task pair (N=60)	Task Success Early		Task Success Late		Z	p-value
	Count	%age	Count	%age		
Early vs. Late	33	55.0	27	45.0	0.94	0.17

In addition to this, in order to locate supportive evidence, we wondered if the task success for tasks SE, SL, DE and DL in group **AV** were occurring independently of each other, or not. To this effect, we used the non-parametric *Chi-square test with Yates' continuity correction* (Kurtz, 1983, p. 216). We used Yates' continuity correction as the Chi-square test would risk overestimating associations if the number of observations is less than 75. We had observations less than 75 in this study. Hence, using Yates's continuity correction delivered a more conservative estimate of the

associations. The test indicated that for different within group permutation of task pairs (SE-SL, DE-DL, SE-DE, SL-DL, SE-DL and SL-DE), the calculated values of $\chi 2$ corrected (Table 4-15) were always less than the table values of $\chi 2$ (3.841 for d.f. = 1 at $\alpha = 0.05$). The task success counts in each pair were therefore occurring independent of each other. Additionally, we infer that the assignment of tasks to different participants had proven to be random without any learning effect.

Table 4-15. Between task analysis in group AV using Chi-square ($\chi 2$) test with Yates' continuity correction for $\chi 2$ table value of 3.841.

Task pair (X vs. Y)		d.f.	$\chi 2$ corrected
X (N= 30)	Y (N = 30)		
SE	SL	1	0.298
DE	DL	1	1.251
SE	DE	1	0.243
SL	DL	1	0.021
SE	DL	1	2.943
SL	DE	1	1.778

4.5.2 Within-group analysis for Task time, Choice error and menu repetitions in group AV

We now present results of the analysis for task time, choice errors and menu repetition. This was a within group analysis which considered only the group **AV**. We compared measures across different tasks when taken as pairs viz. SE-SL, DE-DL, SE-DE, SL-DL, SE-DL and SL-DE. Measures corresponding to task time, choice errors and menu repetition were first tested for their normality. It was seen that choice error and menu repetition measures were not normally distributed while task time measures were. Subsequently for task time, inferential statistics was computed using *Paired T-test.* And, the same was calculated using *Wilcoxon signed rank test* for choice error and menu repetition measures. Note that similar to within group analysis for task success for group **AV** in the previous section, we mentioned count of the participants successful in both the tasks in a given task pair wherever we used *"n"* in the forthcoming tables.

Paired T- test statistics for task time indicated that participants took significantly different time across all the task pairs except the pair SE-DE (Table

4-16). By looking at the descriptive statistics, we also inferred that task pairs which involved SL task i.e. SE-SL, SL-DL and SL-DE, participants significantly took more time to successfully attempt SL task in comparison to the rest. Another observation suggested that in the groups which included tasks with deep menu depth i.e. task pairs DE-DL, SE-DE and SE-DL, participants were taking significantly more time to successfully attempt tasks with *deep* menu depth.

Table 4-16. Between task analysis for group AV. Paired T- test statistics for task time.

Task pair	Descriptive statistics	Task time (in seconds)		
SE-SL (nSE = nSL = 16)		*SE*	*SL*	*Sig.*
	Sum	1635	2666	✓
	Mean	102.19	166.63	
	St. dev.	29.47	54.1	
DE-DL (nDE = nDL = 25)		*DE*	*DL*	*Sig.*
	Sum	2519	3017	✓
	Mean	100.76	120.68	
	St. dev.	29.12	30.93	
SE-DE (nSE = nDE = 24)		*SE*	*DE*	*Sig.*
	Sum	2229	2302	×
	Mean	92.88	95.92	
	St. dev.	31.21	24.52	
SL-DL (nSL = nDL = 17)		*SL*	*DL*	*Sig.*
	Sum	2799	2292	✓
	Mean	164.65	134.82	
	St. dev.	53.02	36.7	
SE-DL (nSE = nDL = 27)		*SE*	*DL*	*Sig.*
	Sum	2526	3311	✓
	Mean	93.55	122.62	
	St. dev.	29.05	32.55	
SL-DE (nSL = nDE = 13)		*SL*	*DE*	*Sig.*
	Sum	1969	1366	✓
	Mean	151.46	105.08	
	St. dev.	48.88	35.67	

Wilcoxon signed rank test statistics for menu repetition (Table 4-17) indicated that participants performed with significant differences in the menu repetitions for task pairs SE-DE, SL-DL and SE-DL. Looking at the individual means within these pairs,

we also inferred that *deep* menu depth caused more repetitions than *shallow* menu depth.

Table 4-17. Between task analysis for group AV. Wilcoxon signed rank test statistics for menu repetition.

Task pair	Descriptive statistics	Menu repetition		
SE-SL (nSE = nSL = 16)		*SE*	*SL*	*Sig.*
	Sum	9	21	×
	Mean	0.56	1.31	
	St. dev.	0.63	1.62	
DE-DL (nDE = nDL = 25)		*DE*	*DL*	*Sig.*
	Sum	48	34	×
	Mean	1.92	1.36	
	St. dev.	1.55	1.08	
SE-DE (nSE = nDE = 24)		*SE*	*DE*	*Sig.*
	Sum	11	40	✓
	Mean	0.46	1.66	
	St. dev.	0.66	1.15	
SL-DL (nSL = nDL = 17)		*SL*	*DL*	*Sig.*
	Sum	21	37	✓
	Mean	1.24	2.18	
	St. dev.	1.6	1.38	
SE-DL (nSE = nDL = 27)		*SE*	*DL*	*Sig.*
	Sum	12	43	✓
	Mean	0.44	1.59	
	St. dev.	0.58	1.42	
SL-DE (nSL = nDE = 13)		*SL*	*DE*	*Sig.*
	Sum	11	29	×
	Mean	0.85	2.23	
	St. dev.	1.41	1.92	

We did not find participants commit significantly different choice errors while attempting a task successfully across different task pairs (Table 4-18). Wilcoxon signed rank test statistics suggested only marginal differences across various task pairs (Table 4-18). However, we could observe indications of deep menu depth causing marginally more errors than the shallow menu depth. Similar marginal increase in error could also be seen for tasks where target menu item occurred late in the menu.

Table 4-18. Between task analysis for group AV. Wilcoxon signed rank test statistics for choice error.

Task pair	Descriptive statistics	Choice error		
		SE	*SL*	*Sig.*
SE-SL *(nSE = nSL = 16)*	**Sum**	3	4	×
	Mean	0.19	0.25	
	St. dev.	0.4	0.45	
		DE	*DL*	*Sig.*
DE-DL *(nDE = nDL = 25)*	**Sum**	7	10	×
	Mean	0.28	0.4	
	St. dev.	0.54	0.65	
		SE	*DE*	*Sig.*
SE-DE *(nSE = nDE = 24)*	**Sum**	2	7	×
	Mean	0.08	0.29	
	St. dev.	0.28	0.55	
		SL	*DL*	*Sig.*
SL-DL *(nSL = nDL = 17)*	**Sum**	4	8	×
	Mean	0.24	0.47	
	St. dev.	0.44	0.72	
		SE	*DL*	*Sig.*
SE-DL *(nSE = nDL = 27)*	**Sum**	3	11	×
	Mean	0.11	0.41	
	St. dev.	0.32	0.69	
		SL	*DE*	*Sig.*
SL-DE *(nSL = nDE = 13)*	**Sum**	3	3	×
	Mean	0.23	0.23	
	St. dev.	0.44	0.6	

4.6 Conclusion

We evaluated and analyzed four different designs of IVRs agricultural market service, namely *audio-visual-shallow*, *audio-visual-deep*, *audio-only-shallow* and *audio-only- deep*, with non-tech savvy users. We hypothesized that *audio-visual interfaces* would help emergent users to be significantly more successful than audio-only interface. Our hypothesis holds true for all the four combination of menu-depth and menu-item-position, namely shallow-early (SE), shallow-late (SL), deep-early (DE) and deep-late (DL). In case of shallow-early (SL), the task success is in the

direction of favor of audio-visual interface over audio-only IVRs although the difference is marginally significant.

IVRs design literature (Cohen, 2004; Commarford et al., 2008; Suhm et al., 2001) favors shallow menu depth over deep menu depth for better user performance. Note that this preference is mentioned for IVRs menu design, and we had hoped that the same will get confirmed in the case of current experiment. Shallow menu depth would be able to attract better task success scores that deep menu depth not only in case of audio-only interface but for audio-visual interface as well. However surprisingly, we discovered that our results countered the theoretical stance in this case. Talking of audio-only interface, deep menu depth had resulted in higher task success scores than shallow menu depth. Although this difference is not significant, but the calculated p-value is only slightly greater than the alpha value ($p=0.06$). Further, the task success score trend in favor of deep menu depth over shallow menu depth becomes rather statistically significant when we consider participants in audio-visual interface group and in the overall participant pool. In quest to locate probable reasons for this finding, we speculate that users were perhaps finding it more difficult to 'browse and select' a particular menu item amongst several others while using shallow menu depth IVRs in comparison to deep menu depth IVRs. While this seems a plausible argument explaining only the case for audio-visual interface. On an extended note, we find it more compelling to believe that users' ability to understand abstract categories, as present in deep menu depth IVRs of both audio-visual and audio-only groups, had improved due to training administered prior to the testing sessions. In an earlier work (Rashinkar et al., 2011) on training non-tech savvy users for category abstraction, we had demonstrated that a long-term training did result in better success scores, although not statistically significant, over a short-term training. The current study seems to reinstate the relevance of training. At the same time, it suggests that inclusion of visuals (along with training) helps achieving significantly better results with emergent users.

Any change in menu-depth and/or menu-item-position is less likely to bring significantly different success scores in case of audio-only interface. In other words, audio-only interface users are more likely to show consistently 'poor' success scores

even if menu-depth and menu-item-position are changed. While on the other hand if the same changes occur in audio-visual interface, participants are more likely to exhibit significantly different success scores. This is interesting when we start imagining ways to improve the usability of the existing IVRs. The current evidence supports using strategies involving visual augmentation of existing IVRs to improve usability. Retrospective accounts of users' experience of using the test prototypes also support this belief. Users testing audio-visual interface reported an 'in-control and confident' behavior while the users who tried audio-only interface kept feeling stressed and lost.

In case of participants using audio-visual IVRs, success score corresponding to deep-late (D) task is significantly better than the success score corresponding to deep-early (C) task. The menu-repetition score corresponding to D is also marginally less than C. This sounds counter intuitive; a participant can be assumed to perform better when the target item lies earlier in the menu. Under this assumption, participants should have shown better success scores with deep-early (C) instead of deep-late (D). But as we see this is not the case. We suspect participant's understanding of the appropriate content category as a possible reason here. This suspicion has emerged out of the observation data collected during testing sessions. Similar studies (Indrani Medhi, Menon, Cutrell, & Toyama, 2010; Rashinkar et al., 2011) have also suspected that emergent population of users might have specific usability challenges like ability to abstract concepts. In our case too, participant might have found one category more abstract than the other. Note that deep-late (D) and deep-early (C) tasks involved varieties of onion and rice respectively as the target menu-items. It seems that participants were able to better identify vegetables as category over grains, and subsequently onion over rice as the category contents.

One may argue that if users were learning to take advantage of the visual augmentation so much so that they could take actions even before the completion of the audio prompts, why not imagine a complete graphical user interface (GUI) to design such a system. We share a similar concern and probed our participants at specific moments and on their entire experience in general to get insights. Some of them reported that audio helped them out making sense of what they were seeing,

particularly when they had to interpret category titles. Additionally, at times when they would miss some visual information or a cue, audio helped aligning them to the relevant visual cue. For us, these are the instances when audio brings "directedness" and helps emergent users in their interaction with the interface. We present relevant work on directedness in chapter 5.

This study only compared audio-visual interface with audio-only interface. The spectrum of the results with respect to designing audio-visual interfaces for emergent users would only be richer if we can include an equally competent graphical user interface. In addition, all the tasks assigned to the participants were about seeking 'information'. We call these the "informational tasks". In these tasks, participants navigated across the interface through keypresses enabled selection of menu items to find prices for commodities in a market location of interest. Task performance with intents of seeking information did not expect users to provide any data inputs viz. profile access details or numbers. The only inputs provided to the system were numeral keypress from the user's side to navigation. We wonder if our results were generalizable to other tasks e.g transactional tasks which may involve relatively a complex interaction with the system. We discuss our relevant work on this topic in chapter 5.

Chapter 5 Directedness and Persistence in Audio-visual Interface for Emergent users

5.1 Introduction

As dicussed in chapter 3, Audio and visual as interaction modalities, have inherent advantages when used together in designing audio-visual interafces for emergent users. Use of audio prompts in an interface has the ability to direct users towards task completion. It can explicitly prompt users to take actions when presented with a menu. It can lead the users through system navigation and ease his journey in an interface (Acomb et al., 2007). We call this *directedness* of audio. Audio is also known to be transient, temporal and time-dependent. Visuals, on the other hand, have *persistence*. These can be made to stay in front of the users and can be made time – independent. Persistence is specifically known to provide graphical aids and memory assistance during users' interaction with the system.

In the current study, we pay emphasis on *directedness* and *persistence*, and design an experiment to demonstrate merits of their combined use in audio-visual interface for emergent users. We designed an experiment where participants had to use three different variants of a banking application, namely audio-only IVRs (A), a graphical user interface (G) and an audio-visual interface (AV). They used these test prototypes to perform two different test tasks- an informational task (Ti) and a transaction task (Tx). Note that audio prompts were kept same across A and AV, and hence both of these had identical *directedness*. Graphic elements of the interface, on

the other hand, were same across G and AV, and hence both of these had identical *persistence*.

The rest of the chapter is organized as follows. We follow with a mention of hypothesis with details of design and development of the banking applications in three different test prototypes: audio-only IVRs (A), a graphical user interface (G) and an audio-visual interface (AV). We list details of our approach in designing these test prototypes including their content, application specific features and user interactions. This is followed by a detailed mention of method used in the experiment including protocol and modes of data collection. Finally, we list results of statistical analysis and end with a summary of our work.

5.2 Hypotheses

Interface G, which is largely a mute interface, requires the user to interpret the visuals and take action. In contrast interface A explicitly directs the user to perform a task through audio prompts. Based on this, our hypothesis is that emergent users will successfully complete more tasks in case of A than in case of G.

Hypothesis 1 (H1) *Emergent users are able to complete tasks more successfully with A than G for both tasks Tx and Ti.*

Interface G is persistent because of the use of visuals. It presents options to users in parallel, without the limitations of sequentially. In contrast, the audio in interface A is ephemeral and temporal. Users need to pay continuous attention and wait for the appropriate option to be spoken. If the user misses the desired option in A, or if he forgets the number associated with the option, he needs to wait for the menu to repeat all over again before making a choice. Based on this, we hypothesize that users will need less time to complete the tasks with G than with A.

Hypothesis 2(H2)*: Emergent users will be able to complete the tasks faster with G than A for both tasks Tx and Ti.*

We expect that interfaces AV represent the best of both A and G. In AV interface, audio contributes directedness while visual contributes persistence. Also,

individually interface G is persistent but lack directedness and interface A is directed but lack persistence. At the least, we expect that emergent users will be more successful with tasks in the AV interface than they are with G because of directedness of audio; and that they will be faster with AV, than they are with A because of persistence of visuals.

***Hypothesis 3 (H3)**: Emergent users are able to complete tasks more successfully with AV than G for both tasks Tx and Ti.*

***Hypothesis 4 (H4)**: Emergent users will be able to complete the tasks faster with AV than A for both tasks Tx and Ti.*

In the best case, we hope that emergent users will be more successful with tasks in the AV interface as they are with A; and that they will be faster with AV than they are with G.

***Hypothesis 5 (H5)**: Emergent users are not significantly less successful in completing tasks with AV as with A for both tasks Tx and Ti.*

***Hypothesis 6 (H6)**: Emergent users are not significantly slower in completing tasks with AV as with G for both tasks Tx and Ti.*

5.3 Design of test prototypes

5.3.1 Content

We designed three different prototypes (A, G and AV) of a banking application for a fictitious bank. Designs of these interfaces were created after reviewing some of the widely used banking interfaces. This included existing IVRs of 3 banks, ATMs of 4 banks and mobile banking GUIs of 3 banks. The audio script (in IVRs) and screen content (in ATMs and mobile banking GUIs) were analyzed and a common set of features and functions applicable to banking were selected. These include account balance inquiry, transaction inquiry, and fund transfer to a registered beneficiary.

Choices related with content creation and its validation along with consistency checks were carried out after consulting stakeholders - banking officials from two banks and a group of representative users. This helped us in (a) considering banking priorities of the sample population and designing test tasks accordingly, (b) avoiding information gaps in task flows, and (c) ensuring the test prototypes' navigation structure to resemble their real world realizations.

5.3.2 Test tasks and prototypes

The study involved two test tasks. One task was "find the nearest ATM". This was an informational task (Ti). The second task was "transfer Rs. 5,000 to a beneficiary x", and it was called transactional task (Tx). Figure 5-1 and Figure 5-2 illustrate the abstracted logical flows of Ti and Tx. While these flows were identical for the three interfaces A, G and AV, their implementation details obviously varied. For each case, the design team tried to optimize on the type of modality in use. An example of implementation of Tx for A, G and AV is covered in figures 5-3, 5-4 and 5-5 respectively. Note that the test prototypes were designed in 'Hindi'. Figures 5-3, 5-4 and 5-5 provide merely the translations into English.

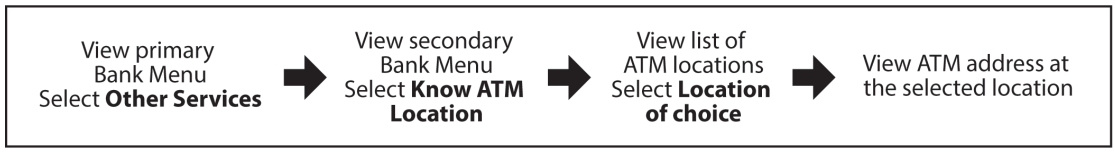


Figure 5-1. Informational Task (Ti).

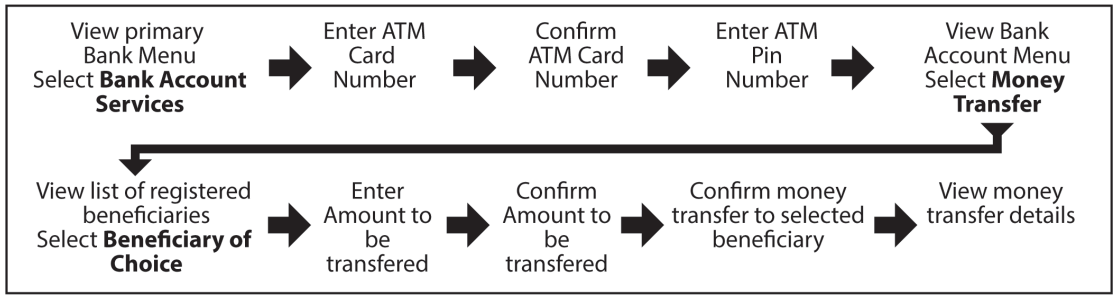


Figure 5-2. Transaction Task (Tx).

Menu level	Audio prompt	User Input
1	Greetings! Welcome to the XYZ Bank Mobile Banking. You can now avail different services offered by the bank. To access services related with bank savings account like inquiring about the last 10 transactions, or to pay bills etc.,	Key press 1
	For farmers' credit card, press 2. For information on other services like receiving a cheque book, or finding the address of the nearest ATM etc., press 3. For financial assistance to agricultural practices, press 4. For demat	
2	Enter your 16 digit ATM card number.	Enters 16 digit ATM card
	For assistance or help, press #. To go back, press *.	
3	The ATM number provided by you is ABCD VXYZ ABCD	Key press 1
	If this ATM number is right, press 1. If you like to change ATM number, press 2.	
	For assistance or help, press #. To go back, press *	
4	Enter your 4 digit ATM pin.	Enters 4 digit ATM pin number.
	For assistance or help, press #. To go back, press *.	
5	Your account balance is Rupees …….. Please hear again, your account balance is Rupees ……..	Key press 2
	To know last 10 transactions, press 1. To transfer money to a registered payee, press 2.	
	To pay electricity, water bills etc., press 3. For any other information, press 4.	
	For assistance or help, press #. To go back, press *.	
6	Choose the payee whom you would like to send money. To send money to registered payee (1,2,3,4,5,6,7), press	Key press 3
	To send money to an unregistered payee, press 8. For assistance or help, press #. To go back, press *	
7	You told us that you would like to send money to	Key press 1
	If you would like to do so, press 1. If you would like to change the beneficiary, press 2.	
	For assistance or help, press #. To go back, press *.	
8	Your account balance is Rupees (…). Press in numbers the amount which you like to send to registered payee 3. For example, to send Rupees 5000, you will have to once press the number 5 followed by pressing 0 thrice. Press in numbers the amount which you like to send to registered	Enters amount to be send to registered payee 3.
	For assistance or help, press #. To go back, press *.	
9	We are now ready to send Rupees (…) to registered payee 3. If you would like to do so, press 1. If you choose not to do so, press 2. For assistance or help, press #. To	Key press 1
10	We have sent Rupees (…) to registered payee 3. Your updated account balance is Rupees (…) .	Key press 2
	To continue mobile banking press 1. To exit mobile banking press 2. For assistance or help, press #.	

Figure 5-3. Transactional task (Tx) in prototype A. Audio prompts other than those meant for Transaction task are truncated for brevity.

Figure 5-4. Transaction task in prototype G.

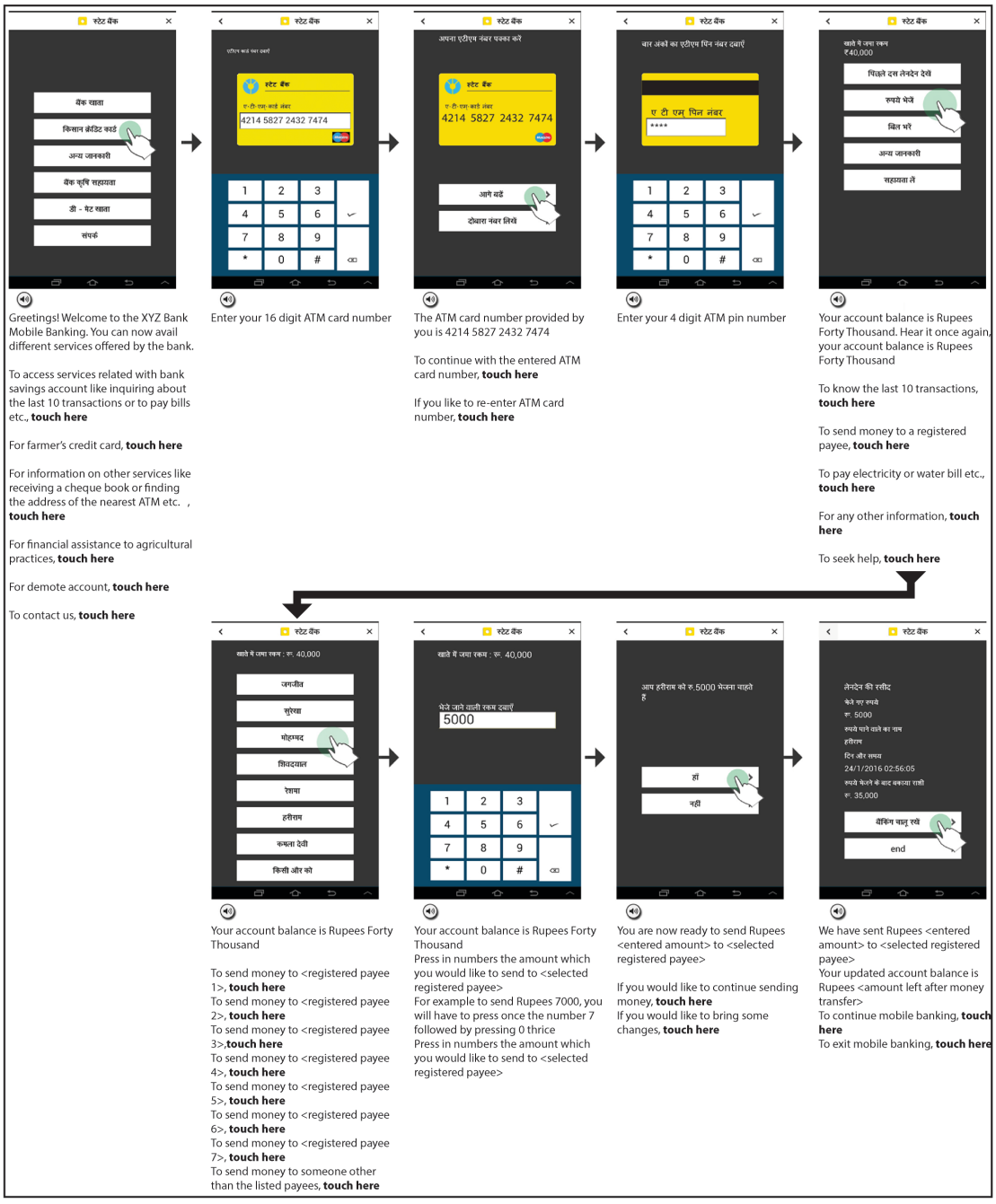


Figure 5-5. Transaction task (Tx) in prototype AV. Audio icon suggests presence of audio prompts.

We designed prototype A to use explicit audio prompts with adequate directedness. The navigation was akin to present day IVRs where users made choices through a sequential menu. As an interface, A was temporal and could only receive user inputs as number key presses. Whereas prototype G used graphical representations in the form of menu buttons, form fields and textual information wherever needed. Menu buttons suggested their actions as labels, form fields contained user entries; and textual information suggested appropriate system requirement and

feedback. G presents information with persistent visuals much akin to a Graphical user Interface. It was a 'prompt free and mute' interface where menu selection did not depend on time, rather menu could stay indefinitely until a user intervened.

In AV (see Figure 5-5) visuals on the screen complemented audio prompts played to the users. It was identical to G in use of visuals and A in use of audio prompts. Audio prompts described services offered through AV and its user control features like application exit. These prompted users to make menu choices. Audio prompts brought directedness to the interface. Visuals brought persistence and non-temporality to the interface. Visuals appeared before the beginning of the audio prompts and stayed till user inputs. These depicted menu choices and user control features like application exit. A 'hand cursor' with an animating circle underneath moved in sync with audio to indicate menu items. Such an use of visuals reinforced directedness offered by the audio prompts.

Note that the test prototypes A, G and AV were put together with the help of two teams: a design team, and a review team. The design team comprised of authors of this thesis, and 2 user interface designers. The review team comprised of 5 other user interface designers. The experience range for user interface designers varied from 2 to 6 years. The design team wrote audio prompt scripts and visualized menu flows for A and AV. For G and AV, they detailed paper prototypes. Working individually was encouraged in the beginning to generate sufficient design options. Later they circulated their designs amongst themselves for review and feedback.

During review process, A, G and AV were randomly assigned to the members of the review team. They accessed and inspected given designs against heuristics (Gómez, Caballero, & Sevillano, 2014; Neilsen, 1995). Their responses were collated and suggestions were distilled to evolve the final designs of the test prototypes.

Some of the crucial design modifications on recommendation by the review team included the following. *First*, we ensured that the audio prompts were sufficiently "directed". For example, the audio prompt "to access bank savings account, press 1" was replaced by "to access services related with the bank savings account like inquiring about the last ten transactions, or paying bills etc., press 1". The

later prompt sounded elaborated or expanded but emergent users find it easy to understand sufficiently 'directed' prompts (Rashinkar et al., 2011).

Second, for all the three prototypes we included an "ATM card" as a prop in the experiment. This was a printed graphical card (see Figure 5-6). In A and AV, at the time when audio prompt 'Enter your 16-digit ATM card number' was played, users had ATM card prop in their hands. They carried the same while interacting with G although with no audio prompts. This made the banking activity more tangible for the participants. It also reflected the reality a bit more closely. Third, we provided confirmatory feedback for critical inputs such as the 16-digit ATM card number, 4-digit PIN, amount to be transferred, and the account balance left after the transaction. And fourth, we eliminated ambiguity in audio prompts and button labels by improving the script.

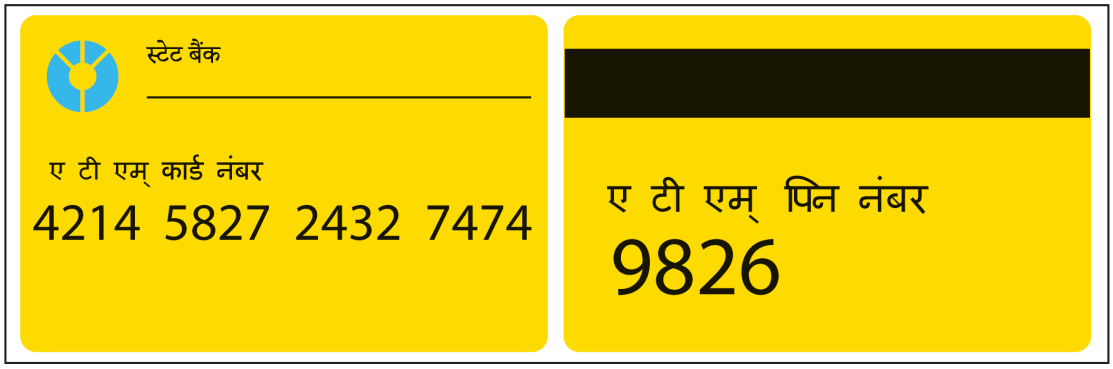


Figure 5-6. "ATM Card" as a prop in the experiment.

The review team also checked for "equivalence" in the three prototypes. Specifically, the same audio prompts were used across A and AV, while same visual elements were used across G and AV.

5.3.3 Prototyping environment and hardware used

The prototypes were developed using Eclipse ("Eclipse," 2012) and Java for Android tablets. The prototypes were designed to function on any android based (version 4.1 or later) tablet or mobile phone but could be best viewed with a screen resolution of 1024x600 pixels. Each prototype was designed to log usage including timestamps of interactions along with tags that depicted individual stages in the participant's journey through the tasks.

5.4 Method

5.4.1 Experiment design

This experiment was based on within-subject design. Individual participants were assigned three different test prototypes - A, G and AV over a period of three days. Each day, they performed two different test tasks - Ti and Tx. The sequence of assignment of the test prototypes and test tasks was randomized to minimize the learning effects.

5.4.2 Protocol

The testing session began with the experimenter briefing the participant about the goals of the study. They were informed of their roles during different stages of the experimentation. They had the choice to opt out of the study at any time. They provided a written consent to participate in the study.

5.4.2.1 Stage 1: Training (10-15 minutes)

The first training task asked the participant to input numbers (0-9) in a random sequence through an audio prompt. When the user touched the correct key, it got highlighted in green colour and the participant was commended for his effort. Else, the key that he touched was highlighted in red and he was asked to try again. The participant repeated this task till he was successful on two consecutive numbers.

As discussed above, emergent users are known to face difficulties while navigating hierarchies (I Medhi et al., 2013). The objective of the second training task was to help users learn to navigate hierarchies upto a depth of 3 menu levels. The experimenter showed the participant a printed photo consisting three identical images of either an agricultural implement (hand shovel) or a household utensil (deep cooking pot). For menu level 1, audio prompts asked the participant to identify the category of the objects, and then touch a corresponding key to mark their response. The audio prompt read as following, "*If you see images of an agricultural equipment, press 1. If you see images of an household item, press 2.*". Once the participant performed this task successfully, the audio prompts asked him to identify the agricultural implement

shown in the printed photo at menu level 2. "*You answered that you were seeing an agricultural implement. Very well! Identify the agricultural implement shown. If you see a sickle, press 1. If you see a hand shovel, press 2. If you see a rake, press 3.*" After identifying the agricultural equipment as hand shovel, for menu level 3, the audio prompt asked the participant about the number of hand shovels shown in the image. "*You answered that you were seeing a hand shovel. Very well! How many of hand shovels do you see? If you see one hand shovel, press 1. If you see two hand shovels, press 2. If you see three hand shovels, press 3.*" If the participant faced any difficulty in making menu-choices, help was provided until he could do the task successfully. Note that the two training tasks were sufficiently different from the actual test tasks. The training was intended to familiarise the user with a touchscreen and navigation of A, G and AV interfaces.

5.4.2.2 Stage 2: Prototype testing (10-15 minutes)

Each participant (see Figure 5-7) performed two different tasks Ti and Tx in a randomized order. In Ti, he had to find the address of the nearest ATM branch. For Tx, he was asked to transfer a specific amount of money to one of the registered payees in his account. The participant had to use one of the three test prototypes on any given day: A, G and AV in order to perform the test tasks.

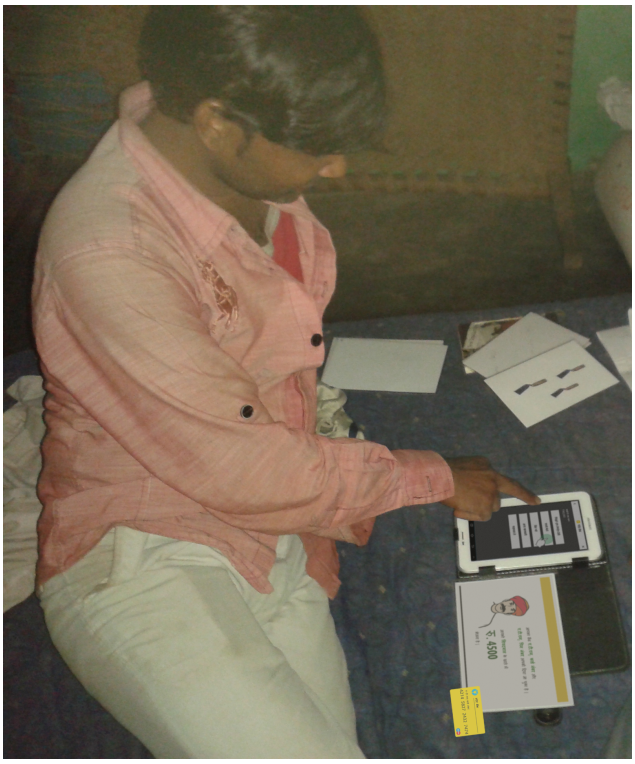


Figure 5-7. Participant during testing session.

5.4.2.3 Stage 3: Rating test prototypes on System Usability Scale

After performing test tasks, each participant was asked to rate his interaction with the day's prototype using a translated version of the System Usability Scale (Brooke, 1996). Note that sufficient care was taken by the experimenters while translating standard SUS questionnaire into *Hindi*, the participants' language of communication. In addition, participants were probed for specific responses if deemed necessary by the experimenters.

On completion of the three stages of the task, the experimenters debriefed the participants and answered any of their questions. They were encouraged to provide suggestions or feedback about the design of the prototype used on that day. Each participant used all the three interfaces in a random order. A gap of one day was provided between the use of two interfaces.

5.4.3 Participants and test environment

Our recruitment criteria were these: Users should be predominantly rural and they should have between 4 to 7 years of school education. They should have

sufficient exposure to physical banking, though should not have any exposure to virtual banking (such as IVRs, mobile banking, online banking or ATMs).

A total of 48 individuals volunteered for participating in the study. They were screened and tested to match certain other profile requirements. Consequently, 12 of these volunteers were dropped out of study for one of the following reasons: a literacy level of either below 4th grade or above 7th grade, willing withdrawal from the study in between the testing sessions, partial completion of training sessions focused at familiarizing users with touch based inputs, interruptions by others or presence of too many onlookers. In the end, 36 people (Females= 11; Males= 25; M= 36.2 years; SD=3.3; Range= 30-40 years) participated in the study.

This sample population belonged to 5 different villages, namely Saalon, Bharroli, Khiriya-Riyasat, Tatarpur and Tatarpur-Patharya, in the sub-district area of Bhander in the state of Madhya Pradesh, India. 9 participants were educated till 4th grade, 15 till 5th grade, 7 till 6th grade and 5 till 7th grade. 24 participants had a savings bank account for more than a decade; another 8 had a savings bank account for more than 5 years; and the remaining 4 had a savings bank account for more than two years. Everyone had the first-hand experience of visiting bank branches for tasks like money withdrawal and deposit, fund transfers and investing in bank savings schemes. 27 out of 36, had also taken agricultural loans against Farmer Credit Card scheme[1]. In this way, each of 36 individuals had fulfilled essential criteria of experiencing offline banking procedures before participating in the study.

Although the study was carried out in the villages, the experimenter could manage to seek a closed room premises for the sessions in each village. That allowed the users to perform tasks without distraction. The field facilitators who included village heads and school teachers, helped experimenter in securing access to these premises. The testing room typically included experimenter and the facilitator along with the user. In case of female participants, one of the family members was also allowed to accompany the user, as it would be considered culturally inappropriate for

[1] A part of Government of India's schemes.

women to be alone with strange men in a closed room. The family member was however requested to stay quiet during the sessions and to not prompt or influence the participant. For the same reason, some of the facilitators were deliberately chosen to be women.

5.4.4 Data collection

Similar to the study in chapter 4, our test prototypes could record and save timestamps and details of user journeys as text files on exit. The dependent variables in this study are Task success, Task time and SUS score (see Table 5-1). We also recorded profile details of the participants, and made observational notes during and after the testing sessions. Task success, was critical to us. We regarded that it is more important for users to complete a task successfully as they often fail.

Table 5-1. List of dependent variables.

Dependent variables	Description
Task success	Determines the successful completion of the task by the participant. Task success is a dichotomous variable. A score of "1" is assigned for successful task completion otherwise "0".
Task time	The task completion time (in seconds) was counted from the moment of attempting the task trial to the moment when the trial completes.
SUS score	SUS score assigned by the users to each interface.

5.5 Results

5.5.1 Task success

The frequency bar graph compares task success scores within a group of 36 participants; and each participant had performed test tasks Tx (Figure 5-8) and Ti (Figure 5-9) on all the three test prototypes A, G and AV.

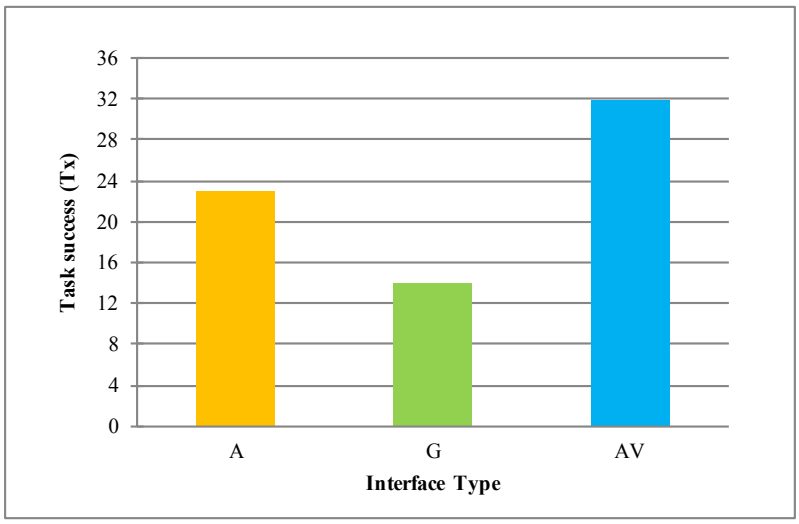


Figure 5-8. Task success in task Tx.

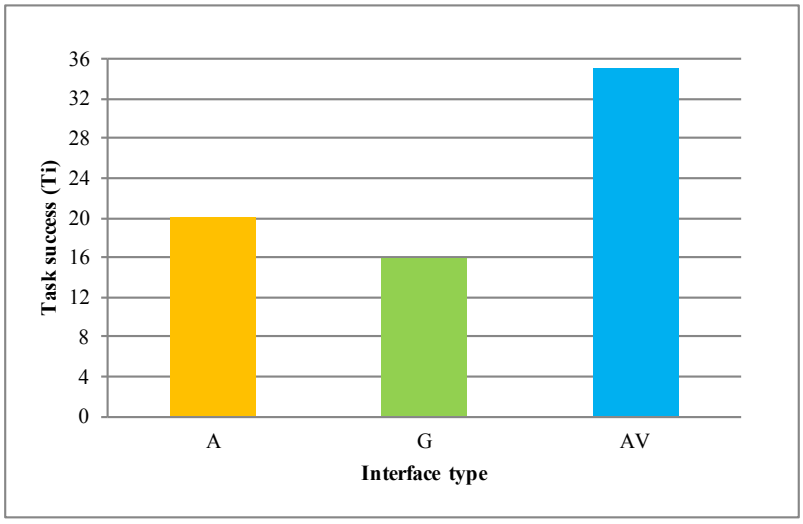


Figure 5-9. Task success in task Ti.

For Tx (n = 36), we observe task success scores of 23, 14 and 32 against A, G and AV respectively. For Ti (n = 36), the task success scores are 20, 16, and 35 corresponding to A, G and AV respectively (Table 5-2).

Table 5-2. Task Success Tx and Ti for interfaces A, G and AV.

Task type (N=36)	**Interface type**					
	A		*G*		*AV*	
	Count	*%age*	*Count*	*%age*	*Count*	*%age*
Tx	23	63.9	14	38.9	32	88.9
Ti	20	55.6	16	44.4	35	97.2

To establish significance of these observations a *two-dependent proportions test* was conducted. Task success scores were compared across all the test prototypes A, G and AV taken as pairs i.e. A vs. G, G vs. AV, and A vs. AV. Following are the results of this test:

(1) Emergent users had exhibited better task success with A over G for both Tx and Ti. In case of Tx, this difference was significant while in Ti, it was only marginal (Table 5-3). Therefore, we can not completely accept **hypothesis 1**, which says that *emergent users are able to complete tasks more successfully with A than G for both Tx and Ti*.

Table 5-3. Two Dependent Proportions test for task success in Tx and Ti across interfaces A and G.

Task type (N=36)	**Task success A**		**Task success G**		**Z**	**p-value**
	Count	*%age*	*Count*	*%age*		
Tx	23	63.9	14	38.9	2.22	0.03
Ti	20	55.6	16	44.4	0.87	0.38

(2) Emergent users had exhibited almost twice the task success with AV over G for both Tx and Ti. This difference is also significant for $p<0.05$ (Table 5-4). We can therefore accept **hypothesis 3**, which says that *emergent users are able to complete tasks more successfully with AV than G for both Tx and Ti.*

Table 5-4. Two Dependent Proportions test for task success in Tx and Ti across interfaces G and AV.

Task type (N=36)	**Task success G**		**Task success AV**		**Z**	**p-value**
	Count	*%age*	*Count*	*%age*		
Tx	14	38.9	32	88.9	3.8	$p<0.001$
Ti	16	44.4	35	97.2	3.93	$p<0.001$

(3) Emergent users had exhibited better task success with A over AV for both Tx and Ti. This difference is also significant for $p<0.05$ (Table 5-4). We can therefore accept **hypothesis 5**, which says that *emergent users are not significantly less successful in completing tasks with AV as with A for both Tx and Ti.* Note that although we can accept **hypothesis 5**, but we did not anticipate significance of differences here. This corresponds to our "best-case" situation for the AV interface, as discussed in section 5.2 (Hypotheses).

Table 5-5. Two Dependent Proportions test for task success in Tx and Ti across interfaces A and AV.

Task type (N=36)	**Task success A**		**Task success AV**		**Z**	**p-value**
	Count	*%age*	*Count*	*%age*		
Tx	23	63.9	32	88.9	2.07	0.04
Ti	20	55.6	35	97.2	3.4	p<0.001

In addition to this, in order to locate supportive evidence, we wanted to check if the task success for tasks Tx and Ti with interfaces A, G and AV were occurring independently of each other, or not. We used the non-parametric *Chi-square test with Yates' continuity correction* (Kurtz, 1983, p. 216). We used Yates' continuity correction as the Chi-square test would risk overestimating associations if the number of observations is less than 75. We had observations less than 75 in this study. Hence, using Yates's continuity correction delivered a more conservative estimate of the associations. The test indicated that for different pair of interfaces (A vs. G, G vs. AV, A vs. AV) individually in tasks Tx and Ti, the calculated values of $\chi 2$ corrected (Table 5-6) were always less than the table values of $\chi 2$ (3.841 for d.f. = 1 at $\alpha = 0.05$). The task success with interfaces A, G and AV for tasks Tx and Ti were therefore occurring independently of each other (Kurtz, 1983, p. 216). We infer that the assignment of test tasks and of the test prototypes to the participants has proved to be random with no significant learning effect.

Table 5-6. Between interface analysis for Tx and Ti using Chi-square ($\chi 2$) test with Yates' continuity correction for $\chi 2$ table value of 3.841.

Task type (N=36)	**Interface pair**		**d.f.**	**$\chi 2$ corrected**
Tx	A	G	1	3.309
	G	AV	1	0.004
	A	AV	1	0.806
Ti	A	G	1	3.106
	G	AV	1	0.013
	A	AV	1	0.013

5.5.2 Task time

5.5.2.1 For Transaction task (Tx)

The grand mean for time taken for successful tasks was 232.87 seconds. By Interface types, the means were 317.39 seconds for A, 165.86 seconds for G and 201.44 seconds for AV (Table 5-7, Figure 5-10). As evident in the means, the participants performing Tx on interface AV took 36.53 % less time than interface A, but 21.45 % more time than interface G. There was a statistically significant difference between groups ($F_{2, 32.289} = 92.92$, $p < .001$) as determined by Welch statistic (Table 5-8).

A Games-Howell post-hoc test (Table 5-9) showed that the task time Tx corresponding to A vs. AV and A vs. G differed significantly at $p < .05$; though the task time difference for AV vs. G was marginally significant (p=0.073).

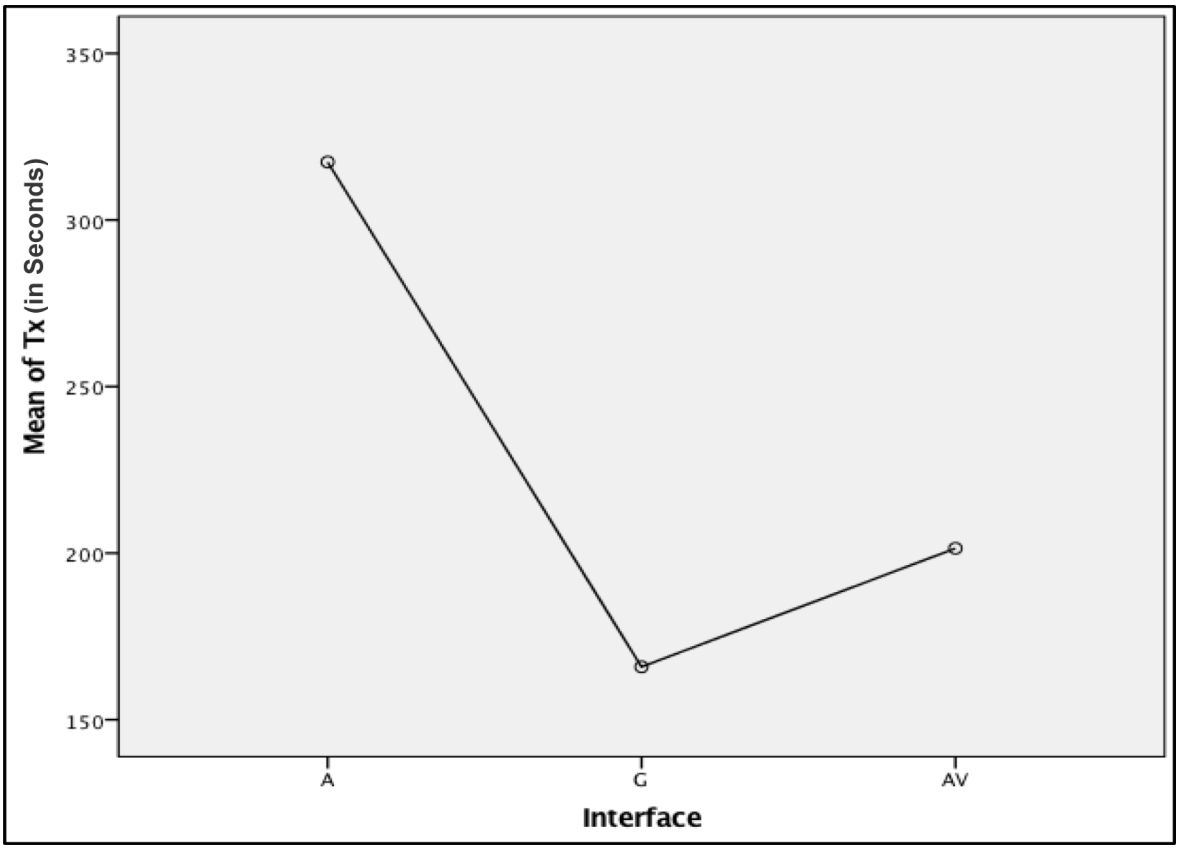


Figure 5-10. Mean plot: Time taken in Test Task Tx for Interfaces A, G and AV.

Table 5-7. Descriptive statistics for task time in task Tx across interfaces A, G and AV.

Task time for Tx (in seconds)								
	N	*Mean*	*Std. Deviation*	*Std. Error*	*95% Confidence Interval for Mean*		*Min*	*Max*
					Lower Bound	*Upper Bound*		
A	23	317.4	28.978	6.042	304.86	329.92	254	388
G	14	165.9	47.993	12.827	138.15	193.57	74	241
AV	32	201.4	48.586	8.589	183.92	218.95	111	285
Total	69	232.9	74.793	9.004	214.9	250.84	74	388

Table 5-8. Robust Tests of Equality of Means: Task time in task Tx for A, G and AV.

Tx				
	Statistic[a]	*df1*	*df2*	*Sig.*
Welch	92.92	2	32.29	0
Brown-Forsythe	70.677	2	41.22	0
a. Asymptotically F distributed.				

Table 5-9. Games-Howell Test: Task time in task Tx for A, G and AV.

Dependent Variable: Tx *Games-Howell*		*Mean Difference (I-J)*	*Std. Error*	*Sig.*	*95% Confidence Interval*	
(I) Interface					*Lower Bound*	*Upper Bound*
A	*G*	151.534*	14.179	0	115.49	187.58
	AV	115.954*	10.501	0	90.61	141.3
G	*A*	-151.534*	14.179	0	-187.58	-115.49
	AV	-35.58	15.437	0.073	-74.02	2.86
AV	*A*	-115.954*	10.501	0	-141.3	-90.61
	G	35.58	15.437	0.073	-2.86	74.02
** The mean difference is significant at the 0.05 level.*						

5.5.2.2 For Informational task (Ti)

The grand mean for time taken was 127.9 seconds. By Interface types, the means were 168.75 seconds for A, 70.75 seconds for G and 130.77 seconds for AV

(see Table 5-10, Figure 5-11). As evident in the means, the participants performing Ti on AV took 22.5 % less time than A, but 84.83 % more time than G. There was a statistically significant difference between groups ($F_{2,\ 38.286} = 50.245$, $p < 0.001$) as determined by Welch statistic (see Table 5-11) . A Games-Howell post-hoc test (see Table 5-12) showed that the task time Ti corresponding to A, G and AV for all combinations differed significantly at $p < 0.05$.

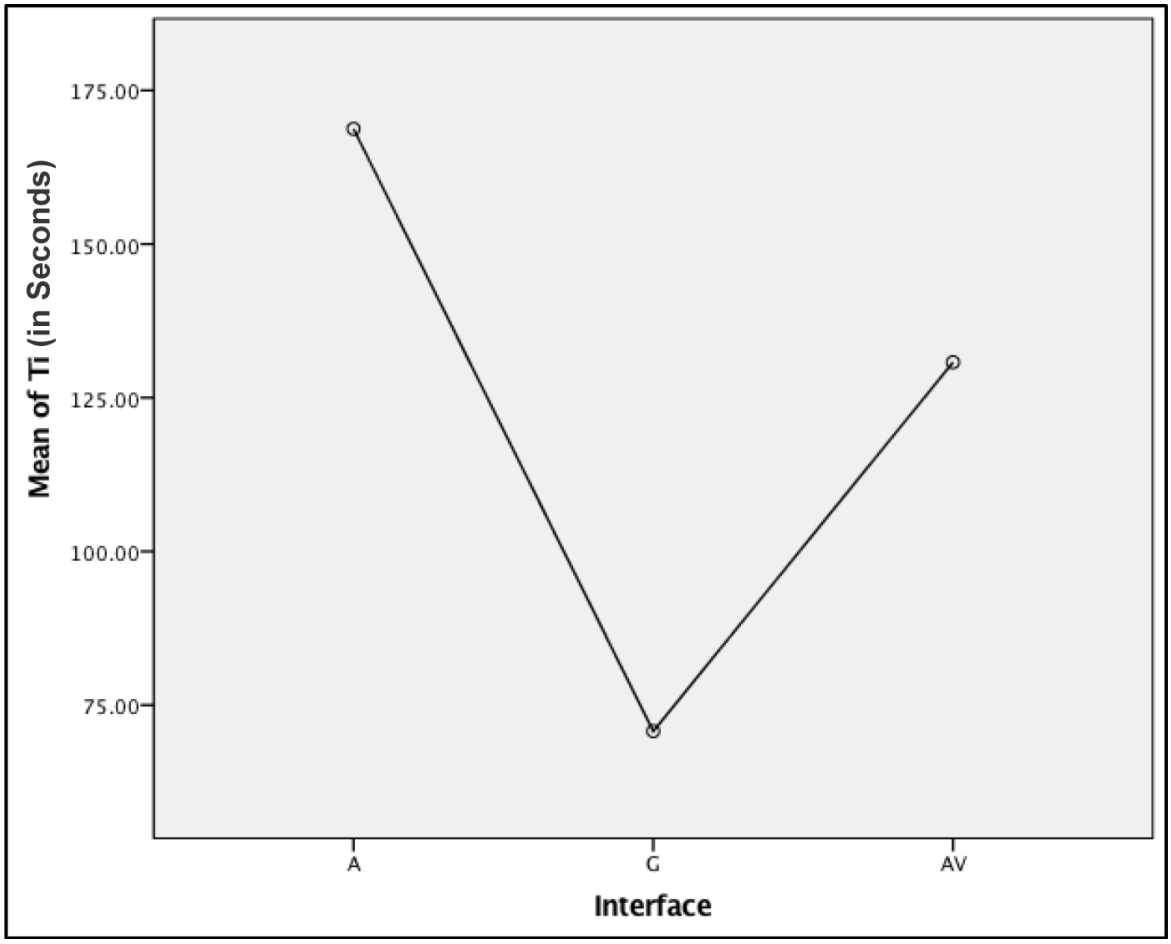


Figure 5-11. Mean plot: Time taken in Test Task Ti for Interfaces A, G and AV.

Table 5-10. Descriptive statistics for task time in task Ti across interfaces A, G and AV.

Task time for Ti (in seconds)								
	N	*Mean*	*Std. Deviation*	*Std. Error*	*95% Confidence Interval for Mean*		*Min*	*Max*
					Lower Bound	*Upper Bound*		
A	20	168.8	39.29762	8.78721	150.358	187.142	103	227
G	16	70.75	23.41652	5.85413	58.2722	83.2278	33	116
AV	35	130.8	33.36173	5.63916	119.311	142.232	71	189
Total	71	127.9	48.00741	5.69743	116.581	139.307	33	227

Table 5-11. Robust Tests of Equality of Means: Task time in task Ti for A, G and AV.

Ti				
	Statistic[a]	*df1*	*df2*	*Sig.*
Welch	50.245	2	38.29	0
Brown-Forsythe	40.946	2	51.11	0
a. Asymptotically F distributed.				

Table 5-12. Games-Howell Test: Task time in task Ti for A, G and AV.

Dependent Variable: Ti *Games-Howell*		*Mean Difference (I-J)*	*Std. Error*	*Sig.*	*95% Confidence Interval*	
(I) Interface					*Lower Bound*	*Upper Bound*
A	*G*	98.00000*	10.55869	0	72.0415	123.959
	AV	37.97857*	10.44104	0.002	12.4133	63.5438
G	*A*	-98.00000*	10.55869	0	-123.96	-72.042
	AV	-60.02143*	8.12841	0	-79.798	-40.245
AV	*A*	-37.97857*	10.44104	0.002	-63.544	-12.413
	G	60.02143*	8.12841	0	40.2451	79.7977
** The mean difference is significant at the 0.05 level.*						

We can see that users who have performed tasks successfully have done so faster with interface G than interface A for both Tx and Ti, thus confirming **hypothesis 2**. Likewise, users who have performed tasks successfully have done so faster with interface AV than interface A in both tasks Tx and Ti, as we had expected in **hypothesis 4**. Beyond that, users who have performed tasks successfully also

performed faster with interface G than with interface AV in both Tx and Ti. This difference is significant for the Ti, but only marginally significant for Tx. Therefore, we do not have complete evidence to say that G is faster than AV for both Tx and Ti. We can accept **hypothesis 6**, which says that *emergent users are not significantly slower in completing tasks with AV as with G.*

5.5.3 Subjective satisfaction using SUS

Unfortunately, the experimenters could gather a response of only 22 out of 36 participants against the standard SUS questionnaire (Brooke, 1996) for each of A, G and AV. Participants who didn't provide SUS skipped responding to the questionnaire for one of the following reasons: feeling tired and/ or uninterested by the end of test tasks, had some other work, perceiving a longer time involvement with the questionnaire, and additionally for reason unknown to the experimenters. For robust analysis we consider the 'true' case first where a total of 22 respondents are considered. Second to follow are the cases of considering a total of 36 participants with missing values replaced by 50 respectively.

5.5.3.1 SUS Analysis for N=22

The mean SUS score (Table 5-13) across participants using A is 64.36 (n = 22, SD 24.97, 95% CI 53.28 to 75.43). For participants using G, the mean SUS score is 71.25 (n = 22, SD 20.57, 95% CI 62.12 to 80.37). And, for those using AV, the mean SUS score is 84.97 (n = 22, SD 10.69, 95% CI 80.23 to 89.71). SUS scores of 81 converts to a percentile rank of above 90 % of reported studies and literature suggests that users are most likely to recommend the product to a friend (Tullis & Albert, 2008).

Table 5-13. Descriptive Statistics for SUS Scores against Interfaces A, G and AV (N=22).

Interface	**N**	**Mean**	**St.Dev.**	**95% confidence interval for mean**	
				Lower Bound	*Upper Bound*
A	22	64.36	24.97	53.28	75.43
G	22	71.25	20.57	62.12	80.37
AV	22	84.97	10.69	80.23	89.71

A Kruskal-Wallis test[2] (Table 5-14 and Table 5-15) showed that there was a statistically significant difference in SUS scores between the different interfaces, $\chi^2(2) = 14.117$, $p = 0.001$, with a mean rank of 25.25 for A, 29.43 for G and 45.82 for AV. Additionally calculation of effect size estimate suggests that 35.48% of the variability in SUS rank scores is accounted for by interface type. For post-hoc analysis, we conducted another set of Kruskal-Wallis to check the significance in difference in SUS scores between pairs A-G (Table 5-16 and Table 5-17), A-AV (Table 5-18 and Table 5-19), and G-AV (Table 5-20 and Table 5-21). We infer[3] that the interface AV scored significantly better than both the interfaces A and G. Where as within interfaces G and A, interface G scored marginally better than interface A.

Table 5-14. Kruskal-Wallis Test ranks for SUS Scores against Interfaces A, G and AV (N=22).

Kruskal-Wallis Test			
Ranks			
	Interface	*N*	*Mean Rank*
SUS	*A*	22	25.25
	G	22	29.43
	AV	22	45.82
	Total	66	

Table 5-15. Kruskal-Wallis Test statictics for SUS Scores against Interfaces A, G and AV (N=22).

Test Staistics[a,b]	
	SUS
Chi-Square	14.117
df	2
Asymp. Sig.	0.001
a. Kruskal Wallis Test	
b. Grouping Variable: Interface	

[2] SUS scores corresponding to A, G and AV are non-normal. Homogeneity of variance is confirmed using Levene statistics. Under these conditions, the Kruskal-Wallis test is a good non-parametric equivalent of one-way ANOVA (Kurtz, 1983, p. 243).

[3] The results are inferred for an adjusted value of $\alpha = 0.02$. This value is calculated using Bonferroni correction applied to reduce type I error in pairwise comparions amongst A, G and AV.

Table 5-16. Kruskal-Wallis Test ranks for SUS Scores against Interfaces A and G (N=22).

Kruskal-Wallis Test			
Ranks			
	Interface	*N*	*Mean Rank*
SUS	*A*	22	20.84
	G	22	24.16
	Total	44	

Table 5-17. Kruskal-Wallis Test statictics for SUS Scores against Interfaces A and G (N=22).

Test Staistics[a,b]	
	SUS
Chi-Square	0.734
df	1
Asymp. Sig.	0.391
a. Kruskal Wallis Test	
b. Grouping Variable: Interface	

Table 5-18. Kruskal-Wallis Test ranks for SUS Scores against Interfaces G and AV (N=22).

Kruskal-Wallis Test			
Ranks			
	Interface	*N*	*Mean Rank*
SUS	*G*	22	16.77
	AV	22	28.23
	Total	44	

Table 5-19. Kruskal-Wallis Test statictics for SUS Scores against Interfaces G and AV (N=22).

Test Staistics[a,b]	
	SUS
Chi-Square	8.752
df	1
Asymp. Sig.	0.003
a. Kruskal Wallis Test	
b. Grouping Variable: Interface	

Table 5-20. Kruskal-Wallis Test ranks for SUS Scores against Interfaces A and AV (N=22).

Kruskal-Wallis Test			
Ranks			
	Interface	*N*	*Mean Rank*
SUS	*A*	22	15.91
	AV	22	29.09
	Total	44	

Table 5-21. Kruskal-Wallis Test statictics for SUS Scores against Interfaces A and AV (N=22).

Test Staistics[a,b]	
	SUS
Chi-Square	11.595
df	1
Asymp. Sig.	0.001
a. Kruskal Wallis Test	
b. Grouping Variable: Interface	

5.5.3.2 SUS analysis for N=36, 14 missing values are replaced by 50.

The mean SUS score (Table 5-22) across participants using A is 58.77 (n = 36, SD 20.61, 95% CI 51.8 to 65.75). For participants using G, the mean SUS score is 62.98 (n = 36, SD 19.08, 95% CI 56.52 to 69.44). And, for those using AV, the mean SUS score is 71.37 (n = 36, SD 19.17, 95% CI 71.37 to 64.88).

Table 5-22. Descriptive Statistics for SUS Scores against Interfaces A, G and AV (N=36).

Interface	**N**	**Mean**	**St.Dev.**	**95% confidence interval for mean**	
				Lower Bound	*Upper Bound*
A	36	58.77	20.61	51.8	65.75
G	36	62.98	19.08	56.52	69.44
AV	36	71.37	19.17	71.37	64.88

A Kruskal-Wallis test[4] (Table 5-23 and Table 5-24) showed that there was statistically significant difference in SUS scores between the different interfaces, $\chi^2(2) = 6.154$, $p = 0.046$, with a mean rank of 47.13 for A, 52.01 for G and 64.36 for AV.

[4] SUS scores corresponding to A, G and AV are non-normal. Homogeneity of variance is confirmed.

For post-hoc analysis, we conducted another set of Kruskal-Wallis to check the significance in difference in SUS scores between pairs A-G (Table 5-25 and Table 5-26), A-AV (Table 5-27 and Table 5-28), and G-AV (Table 5-29 and Table 5-30). We infer[5] that the interface AV scored marginally better than both the interfaces A and G. Whereas within interfaces G and A, G scored marginally better than interface A.

Table 5-23. Kruskal-Wallis Test ranks for SUS Scores against Interfaces A, G and AV (N=36).

Kruskal-Wallis Test			
Ranks			
	Interface	*N*	*Mean Rank*
SUS	*A*	36	47.13
	G	36	52.01
	AV	36	64.36
	Total	108	

Table 5-24. Kruskal-Wallis Test statictics for SUS Scores against Interfaces A, G and AV (N=36).

Test Staistics[a,b]	
	SUS
Chi-Square	6.154
df	2
Asymp. Sig.	0.046
a. Kruskal Wallis Test	
b. Grouping Variable: Interface	

Table 5-25. Kruskal-Wallis Test ranks for SUS Scores against Interfaces A and G (N=36).

Kruskal-Wallis Test			
Ranks			
	Interface	*N*	*Mean Rank*
SUS	*A*	36	35.49
	G	36	37.51
	Total	72	

[5] The results are inferred for an adjusted value of $\alpha = 0.02$. This value is calculated using Bonferroni correction applied to reduce type I error in pairwise comparions amongst A, G and AV.

Table 5-26. Kruskal-Wallis Test statictics for SUS Scores against Interfaces A and G (N=36).

Test Staistics[a,b]	
	SUS
Chi-Square	0.18
df	1
Asymp. Sig.	0.672
a. Kruskal Wallis Test	
b. Grouping Variable: Interface	

Table 5-27. Kruskal-Wallis Test ranks for SUS Scores against Interfaces G and AV (N=36).

Kruskal-Wallis Test			
Ranks			
	Interface	*N*	*Mean Rank*
SUS	*G*	36	33
	AV	36	40
	Total	72	

Table 5-28. Kruskal-Wallis Test statictics for SUS Scores against Interfaces G and AV (N=36).

Test Staistics[a,b]	
	SUS
Chi-Square	2.14
df	1
Asymp. Sig.	0.144
a. Kruskal Wallis Test	
b. Grouping Variable: Interface	

Table 5-29. Kruskal-Wallis Test ranks for SUS Scores against Interfaces A and AV (N=36).

Kruskal-Wallis Test			
Ranks			
	Interface	*N*	*Mean Rank*
SUS	*A*	36	32.47
	AV	36	40.53
	Total	72	

Table 5-30. Kruskal-Wallis Test statictics for SUS Scores against Interfaces A and AV (N=36).

Test Staistics[a,b]	
	SUS
Chi-Square	2.834
df	1
Asymp. Sig.	0.092
a. Kruskal Wallis Test	
b. Grouping Variable: Interface	

5.6 Conclusion

We found that emergent users were significantly more successful in Tx with interface A, than they were with interface G. While they were also more successful in Ti with interface A than G, the difference was not significant. These results are quite contrary to the popular notion that IVRs provide a poor user experience, especially in comparison with GUIs. At least as far as emergent users are concerned, IVRs give better task completion than GUIs - an important user experience parameter. Our results confirm Pieraccinni and Huerta (R Pieraccini & Huerta, 2005) who had seen relative merit of using directed dialog IVRs over open-ended dialog IVRs. This opinion is also in line with Balentine (Balentine, 2007; Balentine & Morgan, 1999, p. 146) and Suhm (Suhm, 2008) who would express it when talking of novice users.

In both tasks, emergent users performed tasks significantly more successfully with interface AV than with both interface G and interface A. As far as task success is concerned, audio-visual interfaces seem to be not only "getting the best of the two media", they seem to be doing better than both. The use of audio and visuals does improve the task success substantially for emergent users. We attribute this gain to the simultaneous effect of the "directedness" of audio and the "persistence" of the visuals. These findings suggest that if audio interfaces of IVRs targeted at emergent users were to be enhanced with visuals, (as can be easily done on smartphones, which are becoming very popular amongst emergent users) this will lead to substantially improved task completion.

While many emergent users could not complete tasks with the interface G, those who could, took significantly less time with interface G than with interface A.

This is true for both Tx and Ti. This finding is on expected lines. An audio is temporal while a visual affords parallel perception. The labels on buttons in a GUI are much shorter than the spoken dialogues of the IVR, which are complete sentences. Further, a user must suffer through all the options of the IVR before the desired option appears, while in case of visuals, the user must just skim over the interface to identify the desired option. Moreover, in an IVR if the user misses the desired menu option or forgets the number associated with it, he must wait for the whole menu to repeat. Hence so long as the user can read and interpret the text and the visual, the GUI is faster than IVR.

It is interesting to note that emergent users took significantly less time for both type of tasks with interface AV than with interface A. Adding visuals to an IVR-style interface does lead to substantial time saving on tasks. We could attribute this partly to the "persistence" of visuals and partly to fewer menu repetitions owing to better retention.

However, users consistently needed more time with the AV interface than with interface G. In case of the informational task Ti, the difference was significant. We can speculate two reasons why this might have happened. Firstly, if an audio is playing in the AV interface, emergent users might feel obliged to wait for it to get over before they provide an input (it might seem rude to interrupt a human voice). This is not the case in a GUI. We recall that during testing sessions users exercised deliberate care before interrupting an ongoing audio instruction. Whereas with GUI, they often acted instantly when presented with menu choices.

An analysis of post-task SUS scores shows that the AV interface was much better appreciated by the users followed by interfaces G and A. It is surprising to see that interface G did better on SUS than interface A despite lower task success rate. Clearly, the speed of use played a bigger role in the users' satisfaction than task success. Joshi et al. (Joshi, Chakravarty, et al., 2012) spotted similar findings in favor of audio-visual IVRs over audio-only IVRs. They reported an improved task completion rate, an enhanced perception of ease of use and of the system being less complex The current study extends its scope by including both a different set of user

group and application domain; non-tech savvy population of farmers and the agriculture market service.

In summary (Table 5-31), audio-visual interfaces have a significantly higher success rate than both IVRs and GUIs. They are also faster to use than A interfaces, and while they are slower than GUIs, they are better appreciated than GUIs. We claim that for (first-time use by) emergent users, audio-visual interface offers a good balance between task completion and speed in comparison to audio-only IVRs and GUIs.

Table 5-31. Summary of results for dependent variables across interfaces A, G and AV.

Dependent variable	**Summary of results**
Task Success	*Emergent users had exhibited better task success with A over G for both Tx and Ti. In case of Tx, this difference was significant while in Ti, it was only marginal (Table 5-3).*
	Emergent users had exhibited almost twice the task success with AV over G for both Tx and Ti. This difference is also significant for $p<0.05$ (Table 5-4).
	Emergent users had exhibited better task success with A over AV for both Tx and Ti. This difference is also significant for $p<0.05$ (Table 5-4).
Task Time Tx	*The participants performing Tx on interface AV took 36.53 % less time than interface A, but 21.45 % more time than interface G.There was a statistically significant difference between groups ($F2, 32.289 = 92.92$, $p < .001$) as determined by Welch statistic (see Table 5-8). A Games-Howell post-hoc test (Table 5-9) showed that the task time Tx corresponding to A vs. AV and A vs. G differed significantly at $p < .05$; though the task time difference for AV vs. G was marginally significant ($p=0.073$).*
Task Time Ti	*The participants performing Ti on AV took 22.5 % less time than A, but 84.83 % more time than G. There was a statistically significant difference between groups ($F2, 38.286 = 50.245$, $p < 0.001$) as determined by Welch statistic (see Table 5 11) . A Games-Howell post-hoc test (see Table 5-12) showed that the task time Ti corresponding to A, G and AV for all combinations differed significantly at $p < 0.05$.*
SUS Score	*The mean SUS score (Table 5-13) across participants using A is 64.36 ($n = 22$, SD 24.97, 95% CI 53.28 to 75.43). For participants using G, the mean SUS score is 71.25 ($n = 22$, SD 20.57, 95% CI 62.12 to 80.37). And, for those using AV, the mean SUS score is 84.97 ($n = 22$, SD 10.69, 95% CI 80.23 to 89.71). SUS scores of 81 converts to a percentile rank of above 90 % of reported studies and literature suggests that users are most likely to recommend the product to a friend*

Chapter 6 Summary and conclusion

6.1 Introduction

In our research we investigated the effects of using visuals to support audio in audio-only interfaces like IVRs. We call these audio-visual interfaces (AVIs). We also demonstrated that by building on the best of both the worlds- persistence of visuals and directedness of audio, AVIs can help emergent users in using interactive products and services.

We report work comprising two main studies, in Chapter 4 and Chapter 5, along with earlier explorations. In chapter 4, we study effects of visuals, menu depths, and menu positions on IVR usage by emergent users. Where as in chapter 5, we compare an audio-visual interface with a conventional audio-only interface and a graphical user interface for two different task types – transactional and informational.

6.2 Research claims

1. Emergent users exhibit better task success with audio-visual interfaces across variations in menu depths, and menu positions (see section 4.6). Further it is shown that this result is generalizable for both transactional as well as informational tasks (see section 5.6).

2. Emergent users exhibit better task success with deep menus over shallow menus in both audio-visual and audio-ony interfaces. However, we acknowledge that the result was only marginally significant across groups in audio-only interface ($p = 0.06$). This is contrary to prior studies which show that shallow menus do better than deeper menus in IVRs with traditional users (Cohen, 2004; Commarford et al., 2008; Suhm et al., 2001) and in graphical user interfaces for emergent users (I Medhi et al., 2013). The directedness of audio seems to be helping emergent users navigate hierarchies better than graphical user interfaces (see section 4.6).
3. Task success of emergent users does not degrade significantly with the late position and an early position of a menu item both in an audio-visual interface, and in an audio-only interface (see section 4.6). This is because users select a menu item which best-fits their intent, and not by memorizing all menus as was shown by (Baddeley, 2003).
4. Audio-visual interfaces are a good balance between an audio-only interface and a graphical user interface for emergent users. As discussed in chapter 3, task success comes across as a result of directedness of the audio (Balentine & Morgan, 1999). Visuals, though persistent, usually require interpretation which depends on prior knowledge and cultural background (Marsden, 2007). However, when used in coordination with audio, audio-visual interfaces get the best of both – *directedness* of audio and *persistence* of visuals. In our experiments, emergent users have exbihited significantly greater task success with audio-visual interfaces than with both graphical user interface and audio-only interface for both transactional as well as informational tasks (see section 5.6).
5. Audio-only interfaces result in longer task times due to the inherent nature of audio. Due to lack of persistence, audio-only interfaces require menu repetition which would result in longer task times too (Balentine & Morgan, 1999, p. 12). Further for emergent users, we had to use expansive menus (Rashinkar et al., 2011, p. 280) which will also result in longer task times. As expected, we found that audio-visual interfaces need longer task times than graphical user interfaces, though the difference was significant for informational task and not significant for transactional task. However, audio-visual interfaces are faster

than audio-only interfaces with significant differences because the persistent visuals reduce the need of menu repetition (see section 5.6).

6. Emergent users who were successful on both audio-visual interfaces and graphical user interfaces, were faster with graphical user interfaces (see section 5.6). As discussed below, this suggests useful design implication.
7. Emergent users prefer using audio-visual interfaces over audio-only and graphical user interface as indicated by significantly higher SUS scores (see section 5.6).

Note that our claims are limited to the use of 3 levels of menu depths with maximum 9 items, and of 5 levels of menu depths with maximum 5 items in any given menu.

6.3 Design implications

Our research demonstrates that by supporting audio-only interfaces with visuals, we can enable emergent users to use deeper menus. Our research also demonstrates that the task completion behavior is not degraded by the early and late positions of menu items in audio-visual interfaces. Interface designers can, therefore, utilize good number of menu items in a single menu of audio-visual interfaces. Deeper menu hierarchies with longer menus could enable emergent users to use more complex interactive products than what was previously possible.

Audio-visual interfaces direct emergent towards successful task completion at the expense of task time. Although task times recorded with audio-visual interfaces are better than audio-only interfaces, but they may still be slower than graphical user interfaces. Interface designers for emergent users may, at times, want to further reduce the task times for emergent users. This could be done by switching the directed dialog "on/off", or by providing audio prompts only if the user is unable to proceed. Subsequently with any such functionality, audio-visual interfaces may exhibit improved task time for frequent (returning) users, while staying valuable for first time users.

This thesis is focused on emergent users. However in our understanding, use of visuals should help in all kinds of audio interfaces to all kinds of users. We emphasize

on combined use of audio and visual modalities and don't really favor the use of one modality over the other. It is through the use of visuals and audio both, do we imagine audio-visual interfaces with enough *persistence* and *directedness* (as shown in our test prototypes) to help users sustain their interactions. The test prototypes designed and developed also illustrate non-traditional use of IVRs based audio interface. We have realized agricultural commodity market pricing system and banking system in our audio-visual test prototypes. In addition, early work in our lab with a focus on combined use of audio and visual modality (Joshi, Welankar, Naveen, Kanitkar, & Sheikh, 2008) realize an audio-visual based operating system for mobile phones. We would imagine much more diverse set of applications be designed and developed as audio-visual interfaces.

6.4 Future directions

In this section we acknowledge few interesting markers which inform us of the limitations of the thesis work and of the future research directions which can be considered. During different phases of prototype testing with emergent users, we kept observing participants and quizzing them for different moves (in retrospection) whenever essential. Subsequently we collated a series of incidents where we could derive certain other inferences. For example, some of the participants would take actions even before they hear audio prompts. At different times during the interaction, particularly at menu levels where there were no explicit or higher level categories, they would move ahead of voice menu. These interactions were more common for participants who had already tried completing a task or two with the audio-visual IVRs and were on the later series of tasks. This indicates that during initial trials some of them were concerned with the functioning of the system in relation to their inputs. They suspected that they could only provide their inputs once the voice menu stopped playing. Some of them also reported their concerns regarding information presented by the voice menu. They preferred listening to the voice menu to get a sense of what all was in store for them. We speculate if this is akin to users transforming from first time users to expert voice users (Resnick & Virzi, 1995). If at all this was happening, we believe that this transformation would be faster in case of audio-visual IVRs than

audio-only IVRs. A future study may be considered to generate more evidence to support or negate this speculation.

Talking of task time, common sense suggests that it will usually be more important if an application is to be used repeatedly. With a few exceptions (Joshi et al., 2014), little work is reported in literature where IVRs (or other types of interfaces) are used for frequent, repetitive, or time-critical tasks for emergent users. Hence speed of use of tasks has thus far not been a very important parameter of interest. However, if a wider set of applications are developed for emergent users, speed of use could become important. Supporting IVRs with visuals may improve task time in such scenarios. Since our studies were limited to one-time interaction with each interface, we wonder how repeated use affects the task time of A, AV and G interfaces. Further research work may be required in this direction.

This thesis provides evidence supporting use of visuals along with audio prompts for emergent users. We realize that perhaps more work is needed to understand the "kind" of visuals to be augmented with audio prompts. We may have to conduct future studies to understand these visuals in greater detail. In the current thesis, we use visuals only to complement audio prompts. We used visuals which signify directions to assist in navigation, and those which can suggest interface elements like actionable buttons to be pressed. We also used visuals which denoted the active interface elements versus the passive interface elements with respect to the audio prompts used. May be it is required to study these varied kinds of visuals in greater details including their scope of use and limitations in designing audio-visual interfaces for emergent users. Our results are indeed encouraging but are confined with a limited research scope much like other research projects. We are explicit in suggesting that a combined use of visual and audio prompts does lead to appropriate audio-visual interfaces for emergent users with both directedness and persistence. However, we need to generate more knowledge with respect to the kind of visuals used to expand the discipline.

An interesting point emerging out of the subsequent reviews of this document is the possibility of considering both the emergent and non-emergent groups of users in a single study with audio-visual interfaces as test prototypes. This would generate

opportunities for understanding contrasts or similarities across the two user groups. A future researcher with an inclination towards studying audio-visual interfaces from a universal design perspective is most likely to benefit from such a study.

On a generic note the field trials conducted during the course of the current research work had shown invaluable insights into the needs of the farming community. We realize that farmers could make use of information such as harvest activities, or seed and fertilizer procurement. They were interested in knowing about the agricultural benefits given by different state bodies, and in ways to resolve legal matters related with land. We observed long queues of people in front of bank kiosks and health centers. We hope that similar research projects with some of these application areas will contribute more examples of audio-visual interfaces.

Lastly we acknowledge an apparent limitation of this thesis. One may point it out that our studies are biased towards an all-male subject study, as there was not a substantial number of female subjects appearing in the study. Though the field experiment team consisted of female facilitators to help inducting an equal number of women participants, but we could never get an equal number of female participants in our studies. It seems to us that in the communities where a major part of studies was carried out, access to female members is rather difficult. In the future, we look forward to devise better strategies to engage them.

References

Acomb, K., Bloom, J., Dayanidhi, K., Krogh, P., Levin, E., & Pieraccini, R. (2007). Technical Support Dialog Systems : Issues , Problems , and Solutions. In *Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies* (pp. 25–31). ACM.

Agricultural Marketing Information Network. (2014). Retrieved January 2, 2013, from http://agmarknet.nic.in/

Agriwatch - Commodity Prices India,Commodity Markets. (2014). Retrieved January 2, 2013, from http://www.agriwatch.com

Anokwa, Y., Thomas, S. N., Ramachandran, D., Sherwani, J., Schwartzman, Y., Luk, R., … DeRenzi, B. (2009). Stories from the Field : Reflections on HCI4D Experiences. *Information Technologies & International Development*, *5*(4).

Athavankar, U. (1999). Gestures, mental imagery and spatial reasoning. In *International conference on Visual and spatial reasoning, MIT, Cambridge* (pp. 54–82).

Athavankar, U. A. (1997). Mental Imagery as a design tool. *Cybernetics and Systems*, *28*(1), 25–42. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/019697297126236

Athavankar, U., Bokil, P., Guruprasad, K., & Patsute, R. (2008). Reaching Out in the Mind's Space. In J. S. Gero & A. K. Goel (Eds.), *Design Computing and Cognition*. Springer Dordrecht Heidelberg.

Ávila, I. M. A., & Gudwin, R. R. (2009). Icons and Helpers in the Interaction of Illiterate Users With Computers. In K. Blashki (Ed.), *Proceedings of the IADIS International Conference on Interfaces and Human-Computer Interaction*.

Baddeley, A. (1981). The concept of working memory : A view of its current state and probable future development. *Cognition*, *10*, 17–23.

Baddeley, A. (2003). Working memory and language: an overview. *Journal of Communication*

Disorders, *36*(3), 189–208. http://doi.org/10.1016/S0021-9924(03)00019-4

Balentine, B. (2007). *It's better to be a good machine than a bad person*. Annapolis, Maryland: ICMI Press.

Balentine, B., & Morgan, D. (1999). *How to build a speech recognition application: a style guide for telephony dialogues* (1st ed.). San Ramon, California: Enterprise Integration Group, Inc.

Bayerl, J. P., Millen, D. R., & Lewis, S. H. (1988). Consistent Layout of Function Keys and Screen Labels Speeds User Responses. In *Proceedings of the Human Factors Society* (pp. 344–346). SAGE. http://doi.org/10.1177/154193128803200524

Beaudouin-Lafon, M. (2004). Designing interaction, not interfaces. In *Proceedings of the working conference on Advanced visual interfaces - AVI '04* (p. 15). New York, New York, USA: ACM Press. http://doi.org/10.1145/989863.989865

Bilda, Z., & Gero, J. (2005). Do We Need CAD during Conceptual Design? In B. Martens & A. Brown (Eds.), *Computer Aided Architectural Design Futures 2005* (pp. 155–164).

Blankenhorn, T. T. (2008). Visual Display of Automated Telephone System Menus. United States.

Boyce, S. (2000). Natural spoken dialogue systems for telephony applications. *Communications of the ACM*. Retrieved from http://dl.acm.org/citation.cfm?id=348974

Brandt, J. (2008). Interactive Voice Response Interfaces. In P. Kortum (Ed.), *HCI beyond the GUI: Design for haptic, speech, olfactory and other nontraditional interfaces* (pp. 229–266). Denise E. M. Penrose, Morgan Kaufmann.

Brewer, E., Demmer, M., Du, B., Ho, M., Kam, M., Nedevschi, S., … Fall, K. (2005). The case for technology in developing regions. *Computer*, *38*(6), 25–38. http://doi.org/10.1109/MC.2005.204

Brooke, J. (1996). SUS- A quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194), 4–7. Retrieved from https://scholar.google.com/citations?view_op=view_citation&hl=en&user=qjAGPUcAAAAJ&citation_for_view=qjAGPUcAAAAJ:u5HHmVD_uO8C

Chapanis, A., & Lindenbaum, L. E. (1959). A Reaction Time Study of Four Control-Display Linkages. *Human Factors*, *1*(4), 1–7.

Cohen, M. H. (2004). *Voice User Interface Design*. Addison-Wesley Professional.

Commarford, P. M. (2006). *Working memory, search and signal detection: Implications for IVRS menu design*. College of Arts and Sciences, Univeristy of Central Florida, Orlando, florida.

Commarford, P. M., Lewis, J. R., Smither, J. A.-A., & Gentzler, M. D. (2008). A Comparison of Broad Versus Deep Auditory Menu Structures. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(1), 77–89. Retrieved from http://hfs.sagepub.com/cgi/doi/10.1518/001872008X250665

De Souza, C. S. (2005). *The Semiotic Engineering of Human Computer Interaction*. The MIT Press.

Devanuj, & Joshi, A. (2013). Technology adoption by "emergent" users. In *Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction - APCHI '13* (pp. 28–38). Retrieved from http://dl.acm.org/citation.cfm?id=2525194.2525209

Dutton, R. T., Foster, J. C., & Jack, M. A. (1999). Please mind the doors — do interface metaphors improve the usability of voice response services? *BT Technology Journal*, *17*(1), 172–177.

Eclipse. (2012). Retrieved December 2, 2014, from https://www.eclipse.org/

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis : Verbal Reports as Data*. MIT Press.

Exotel - Exotel, A Business Phone System from the Cloud. (2013). Retrieved August 19, 2014, from http://exotel.in/

Fawcett, P. E., Blomfield-Brown, C., & Storm, C. P. S. (1998). System and method for graphically displaying and navigating through an interactive voice response menu. USA: United States Patent.

Fogg, B. J. (2003). *Persuasive Technology: Using Computers to Change what We Think and Do*. Retrieved from https://books.google.co.in/books?id=9nZHbxULMwgC&printsec=frontcover&dq=persuasive+technology+using+com&hl=en&sa=X&redir_esc=y#v=onepage&q=persuasive technology using com&f=false

Gallager, R. G. (1968). *Information theory and reliable communication*. John Wiley & Sons.

Gamage, P., & Halpin, E. (2007). E-Sri Lanka: bridging the digital divide. *The Electronic Library*. Retrieved from http://www.emeraldinsight.com/doi/abs/10.1108/02640470710837128

Gardner-Bonneau, D. (1999). Guidelines for speech-enabled IVR application design. *Human Factors and Voice Interactive Systems*. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4757-2980-1_7

Gómez, R. Y., Caballero, D. C., & Sevillano, J.-L. (2014). Heuristic Evaluation on Mobile Interfaces: A New Checklist. *The Scientific World Journal*, *2014*.

Götze, M., & Thomas, S. (2001). An Approach to Help Functionally Illiterate People with Graphical Reading Aids. In *Smart Graphics Symposium UK*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.14.9227

Grover, A. S., Stewart, O., & Lubensky, D. (2009). Designing interactive voice response (IVR) interfaces: localisation for low literacy users. In *Computers and Advanced Technology in Education*.

Haifeng Bi, X. B. (2013). Graphical interactive visual response system and method. United States.

Halliday, M. A. K., Gibbons, J., & Nicholas, H. (Eds.). (2012). *Learning, keeping, and using language*

(Vol 2): Selected papers from the 8th World Congress of Applied Linguistics, Sydney, 16-21 August 1987. John Benjamins Pub. Co. Retrieved from https://books.google.co.in/books?hl=en&lr=&id=cw4rWqyuPpsC&oi=fnd&pg=PA381&dq=think+aloud+protocol&ots=4NXVDV4BfV&sig=wRCa3LPb7Yk_Qop7RnvopKd4mPk#v=onepage&q=think aloud protocol&f=false

Hartman, F. R. (1961). Single and multiple channel communication: A review of research and a proposed model. *Audiovisual Communication Review*, *9*(6), 235–262. http://doi.org/10.1007/BF02769048

HIV Aids Topical Information. (2011). Retrieved from http://www.who.int/hiv/topics/en/

Ho, M. R., Smyth, T. N., Kam, M., & Dearden, A. (2009). Human-Computer Interaction for Development : The Past , Present , and Future, *5*(4), 1–18.

Hommel, B., & Prinz, W. (Eds.). (1997). *Theoretical Issues in Stimulus-Response Compatibility*. Elsevier Science B.V.

James, J. (2001). Bridging the digital divide with low-cost information technologies. *Journal of Information Science*, *27*(4), 211–217. http://doi.org/10.1177/016555150102700403

Joshi, A., Chakravarty, A., & Shrivastava, A. (2012). Usability Evaluation of Visual IVR Systems. Retrieved April 28, 2014, from http://www.amruthakrishnan.com/img/projects/visual-ivrs/apchi2012-visual-ivrs.pdf

Joshi, A., Emmadi, N., Rashinkar, P., Malandkar, P., Shrivastav, A., Srivastava, S., & Rajput, N. (2012). Visual IVRs for Low-literate , non-literate and Low-tech-savvy Users. In *CHI 2012*.

Joshi, A., Saple, D. G., Sen, K., Veldeman, E., Rutten, R., Rane, M., … Rodrigues, R. (2014). Supporting treatment of people living with HIV / AIDS in resource limited settings with IVRs. In *Proc. CHI '14* (pp. 1595–1604). New York, New York, USA: ACM Press. http://doi.org/10.1145/2556288.2557236

Joshi, A., Welankar, N., Naveen, B., Kanitkar, K., & Sheikh, R. (2008). Rangoli: a visual phonebook for low-literate users. *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '08*, 217. http://doi.org/10.1145/1409240.1409264

Kamm, C. A., & Helander, M. (1997). Design Issues for Interfaces using Voice Input. In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of Human Computer Interaction* (2nd ed., pp. 1043–1060). Elsevier.

Koffka, K. (1999). *Principles of Gestalt psychology*. Routledge.

Kurtz, N. R. (1983). *Introduction to Social Statistics*. Singapore: McGraw-Hill, Inc.

Marathe, M., O'Neill, J., Pain, P., & Thies, W. (2015). Revisiting CGNet Swara and its impact in rural India. In *Proceedings of the ICTD 2015*. Retrieved from

http://dl.acm.org/citation.cfm?id=2738026

Marics, A. M., & Engelbeck, G. (1997). Designing voice menu applications for telephones. In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of Human Computer Interaction* (2nd ed., pp. 1085–1102). Elsevier.

Marsden, G. (2007). Doing HCI Differently - Stories from the Developing World. In *CHI '07 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. http://doi.org/10.1145/1240624.2180966

Medhi, I., Menon, S. R., Cutrell, E., & Toyama, K. (2010). Beyond strict illiteracy: abstracted learning among low-literate users. In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development - ICTD '10* (pp. 1–9). New York, New York, USA: ACM Press. http://doi.org/10.1145/2369220.2369241

Medhi, I., Prasad, A., & Toyama, K. (2007). Optimal audio-visual representations for illiterate users of computers. In *Proc. WWW '07* (pp. 873–882). New York, New York, USA: ACM Press. http://doi.org/10.1145/1242572.1242690

Medhi, I., Sagar, A., & Toyama, K. (2007). Text-free user interfaces for illiterate and semi-literate users. *Information Technologies and International Development*, *4*(1), 37–50. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4085517

Medhi, I., Toyama, K., Joshi, A., Athavankar, U., & Cutrell, E. (2013). A comparison of list vs. hierarchical UIs on mobile phones for non-literate users. In *Proc. Interact 2013* (pp. 497–504).

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *The Psychological Review*, *63*, 81–97.

NACP. (2011). Retrieved from http://india.gov.in/sectors/health_family/national_aids.php

Narayanaswami, C. (2000). Graphical Voice Response System and Method therefor.

Neilsen, J. (1995). 10 Heuristics for User Interface Design. Retrieved September 30, 2014, from http://www.nngroup.com/articles/ten-usability-heuristics/

Nugent, G. C. (1982). Pictures, audio, and print: symbolic representation and effect on learning. *ECTJ*, *30*(3), 163–174. http://doi.org/10.1007/BF02766597

Oviatt, S., Coulston, R., Tomko, S., Xiao, B., Lunsford, R., Wesson, M., & Carmichael, L. (2003). Toward a Theory of Organized Multimodal Integration Patterns during Human-Computer Interaction. In *ICMI' 03* (pp. 44–51). Vancouver.

Paivio, A. (1990). *Mental Representations: A Dual Coding Approach (Google eBook)*. Oxford University Press.

Parikh, T., Ghosh, K., & Chavan, A. (2003). Design studies for a financial management system for micro-credit groups in rural india. In *CUU' 03*. http://doi.org/10.1145/960201.957209

Parikh, T. S., & Ghosh, K. (2006). Understanding and Designing for Intermediated Information Tasks in India. *Pervasive Computing, IEEE*, *5*(2), 32–39.

Patel, N., Agarwal, S., & Rajput, N. (2008). Experiences designing a voice interface for rural India. *SLT 2008*. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4777830

Patel, N., Agarwal, S., Rajput, N., & Nanavati, A. (2009). A comparative study of speech and dialed input voice interfaces in rural India. *Proceedings of the*. Retrieved from http://dl.acm.org/citation.cfm?id=1518709

Patel, N., Agarwal, S., Rajput, N., Nanavati, A., Dave, P., & Parikh, T. S. (2009). A comparative study of speech and dialed input voice interfaces in rural India. *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09*, 51. http://doi.org/10.1145/1518701.1518709

Patel, N., Chittamuru, D., Jain, A., Dave, P., & Parikh, T. (2010). Avaaj Otalo: a field study of an interactive voice forum for small farmers in rural India. In *CHI 2010*. Retrieved from http://dl.acm.org/citation.cfm?id=1753434

Patra, R., Pal, J., & Nedevschi, S. (2009). ICTD state of the union: where have we reached and where are we headed, 357–366. Retrieved from http://dl.acm.org/citation.cfm?id=1812530.1812567

Peirce, C. S. (2011). *Philosophical Writings Of Peirce*. (J. Buchler, Ed.). New York: Dover Publications, Inc.

Pieraccini, R., & Huerta, J. (2005). Where do we go from here? Research and commercial spoken dialog systems. In *6th SIGdial Workshop on Discourse and Dialog*. Lisbon, Portugal. Retrieved from http://www.isca-speech.org/archive_open/sigdial6/sgd6_001.html

Pieraccini, R., & Lubensky, D. (2005). Spoken Language Communication with Machines : The Long and Winding Road from Research to Business. In A. Moonis & E. Floriana (Eds.), *Innovations in Applied Artificial Intelligence* (pp. 6–15). Springer-Verlag Berlin Heidelberg.

Pillai, J., Athavankar, U., & Schmidt, C. (2013). Presence in Visual Mental Imagery. *International Visual*.

Plauché, M., & Nallasamy, U. (2007). Speech Interfaces for Equitable Access to Information Technology. *Information Technologies & International Development*, *4*(1), 69–86.

Plauche, M., & Prabaker, M. (2006). Tamil Market : A Spoken Dialog System for Rural India. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (pp. 1619–1624). New York, USA: ACM.

Proctor, R. W., & Van Zandt, T. (1994). *Human Factors in Simple and Complex Systems*. Prentice-Hall.

Rashinkar, P. G., Joshi, A., Rane, M., Sali, S., Badodekar, S., Emmadi, N., … Shrivastava, A. (2011). Healthcare IVRS for Non-Tech-Savvy Users. In A. Holzinger & K.-M. Simonic (Eds.),

Information Quality in e-Health, Lecture Notes in Computer Science (Vol. 7058, pp. 263–282). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-25364-5

Resnick, P., & Virzi, R. A. (1992). Skip and Scan: Cleaning up Telephone Interfaces. In *CHI 92* (pp. 419–426).

Resnick, P., & Virzi, R. A. (1995). Relief from the audio interface blues: expanding the spectrum of menu, list, and form styles. *ACM Transactions on Computer-Human Interaction*, *2*(2), 145–176. http://doi.org/10.1145/210181.210183

Rock, R., & Hiller, C. (2002). Text-Enhanced Voice Menu System.

Schmidt, R. W. (Ed.). (1995). *Attention and awareness in foreign language learning*. Second Language Teaching & Curriculum Center, University of Hawaii at Manoa. Retrieved from https://books.google.co.in/books?hl=en&lr=&id=P2gGD0HnjcYC&oi=fnd&pg=PA183&dq=think+aloud+protocol&ots=C27ywcZnAO&sig=MX88qipL1nQsYaBKsyf0afSFDdc#v=onepage&q=think aloud protocol&f=false

Schnelle-walka, D. (2011). I Tell You Something. In *16th European Conference on Pattern Languages of Programs* (p. 10). ACM.

Sears, A., & Jacko, J. A. (2008). *The Human – Computer Interaction Handbook* (Vol. 29). Lawrence Erlbaum Associates. http://doi.org/10.1201/9781410615862

Shneiderman, B., & Plaisant, C. (2005). *Designing the user interface: strategies for effective human-computer interaction*. *British Dental Journal* (Vol. 215). http://doi.org/10.1038/sj.bdj.2013.932

Shrivastava, A., & Joshi, A. (2014). Effects of visuals, menu depths, and menu positions on IVR usage by non-tech savvy users. In *Proceedings of the India HCI 2014 Conference on Human Computer Interaction - IHCI '14* (pp. 35–44). New York, New York, USA: ACM Press. http://doi.org/10.1145/2676702.2676707

State Agricultural Produce Marketing Board, Uttar Pradesh. (2014).

Statista. (2017). Smartphone users in India 2015-2021. Retrieved June 1, 2017, from https://www.statista.com/statistics/467163/forecast-of-smartphone-users-in-india/

Stritzke, W., & Dandy, J. (2005). Use of interactive voice response (IVR) technology in health research with children. *Behavior Research Methods*. Retrieved from http://link.springer.com/article/10.3758/BF03206405

Suhm, B. (2008). IVR Usability engineering using guidelines and anaylsis of end-to-end calls. In H. Gardner-Bonneau, Daryle; E. Blanchard (Ed.), *Human Factors and Voice Interactive Systems* (2nd ed., pp. 1–41). Springer US.

Suhm, B., Freeman, B., & Getty, D. (2001). Curing the menu blues in touch-tone voice interfaces. In *Ext. Abstracts CHI 2001* (pp. 131–132).

Tatchell, G. R. (1996). Problems with the Existing Telephony Customer Interface: The Pending Eclipse

of Touch-Tone and Dial-Tone. In *CHI 96* (pp. 242–243). ACM.

TRAI. (2017). Annual Reports | Telecom Regulatory Authority of India. Retrieved June 1, 2017, from http://www.trai.gov.in/about-us/annual-reports

Tullis, T., & Albert, B. (2008). *Measuring the User Experience*. Morgan Kaufmann Publishers Inc.

Wallace, R. J. (1971). S-R Compatability and the idea of a response code. *Journal of Experimental Psychology*, *88(3)*, 354–360. Retrieved from http://psycnet.apa.org/index.cfm?fa=buy.optionToBuy&id=1971-26092-001

Walsham, G. (2017). ICT4D research: reflections on history and future agenda. *Information Technology for Development*, *23*(1), 18–41. http://doi.org/10.1080/02681102.2016.1246406

Warschauer, M. (2004). *Technology and Social Inclusion*.

Whetzel, D. L., & Wheaton, G. R. (2016). *Applied Measurement: Industrial Psychology in Human Resource Management*. Taylor and Francis. Retrieved from https://books.google.co.in/books?id=B2eVCwAAQBAJ&dq=situational+judgement+test&lr=

Yin, M., & Zhai, S. (2005). Dial and see: tackling the voice menu navigation problem with cross-device user experience integration. In *18th annual ACM symposium on User interface software and technology* (pp. 187–190). ACM.

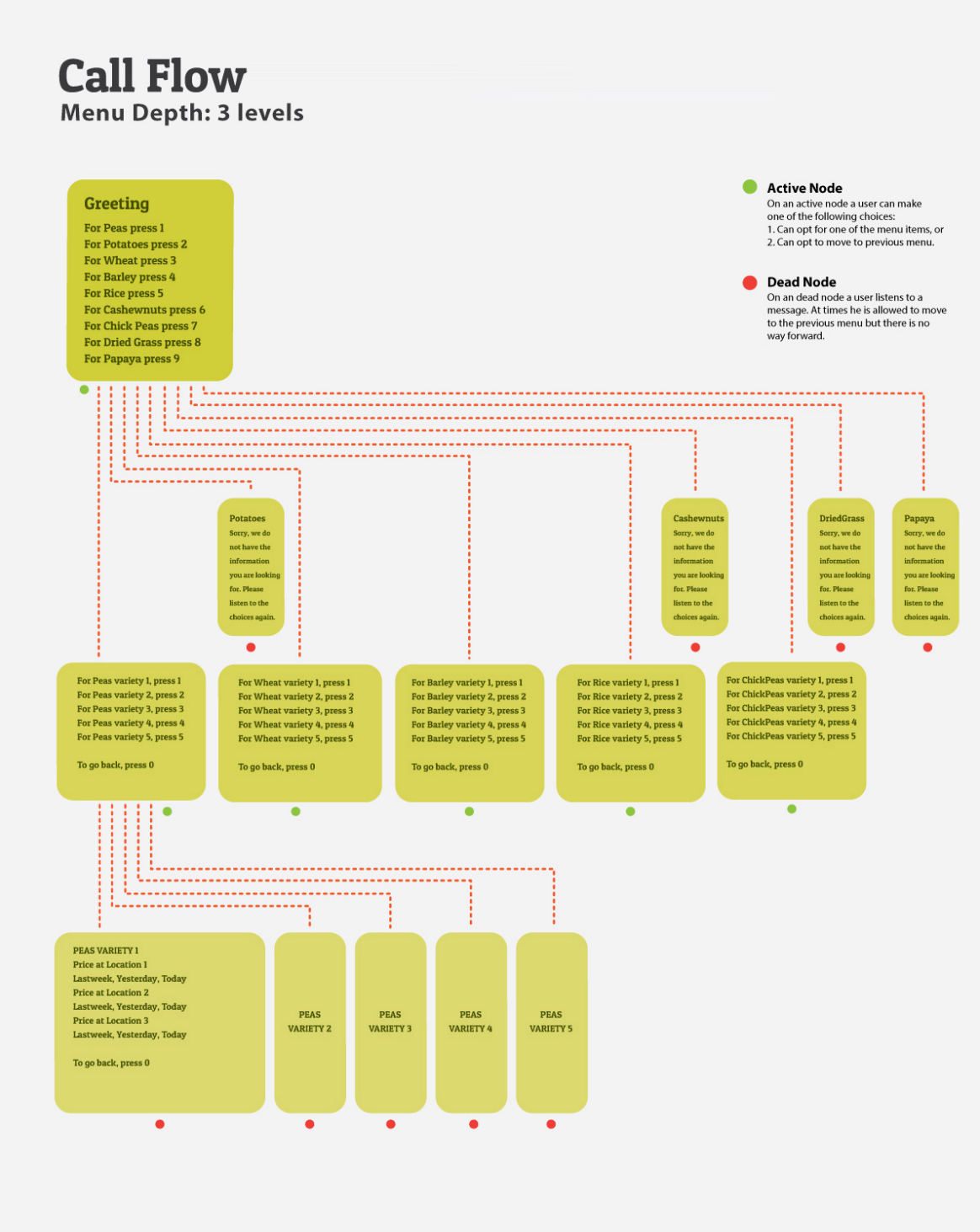
Yin, M., & Zhai, S. (2006). The benefits of augmenting telephone voice menu navigation with visual browsing and search. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '06*, 319. http://doi.org/10.1145/1124772.1124821

Appendices

Appendix 1: Example service prompts with shallow menu depth.

Audio prompts for shallow menu depth prototypes.		
Menu depth	**In Hindi, as used in test prototype**	**Translated text**
1	नमस्कार | बाज़ार भाव सेवा में आपका स्वागत है | यहाँ आप विभिन्न कृषि उत्पादों का प्रति क्विंटल भाव जान सकते हैं | मटर का भाव जानने के लिए 1 दबाएँ | आलू का भाव जानने के लिए 2 दबाएँ | गेंहूँ का भाव जानने के लिए 3 दबाएँ | ज्वार का भाव जानने के लिए 4 दबाएँ | चावल का भाव जानने के लिए 5 दबाएँ | काजू का भाव जानने के लिए 6 दबाएँ | चने का भाव जानने के लिए 7 दबाएँ | सूखी घास का भाव जानने के लिए 8 दबाएँ | पपीते का भाव जानने के लिए 9 दबाएँ |	Greetings! Welcome to the agricultural commodity market price service. You can use this service to know the per quintal market prices for different agricultural commodities. To know the prices for Peas, press 1 To know the prices for Potatoes, press 2 To know the prices for Wheat, press 3 To know the prices for Jwaar, press 4 To know the prices for Rice, press 5 To know the prices for Cashewnuts, press 6 To know the prices for Chickpeas, press 7 To know the prices for Fodder, press 8 To know the prices for Papaya, press 9
2	देशी गेहूं का भाव जानने के लिए 1 दबाएँ | मध्यम गेहूं चावल का भाव जानने के लिए 2 दबाएँ | लाल गेहूं का भाव जानने के लिए 3 दबाएँ | शरबती गेहूं का भाव जानने के लिए 4 दबाएँ | सफ़ेद गेहूं का भाव जानने के लिए 5 दबाएँ | पीछे जाने के लिए 0 दबाएँ |	To know the prices for Deshi wheat press 1 To know the prices for Madhyam wheat, press 2 To know the prices for Red wheat, press 3 To know the prices for Sharbati wheat, press 4 To know the prices for White wheat, press 5 To go back to earlier menu, press 0
3	देशी गेंहू का दिल्ली की मंडी में पिछले हफ्ते भाव था 1750 रुपये | कल इसका भाव बढकर 1829 रुपये हो गया | आज इसका भाव घटकर 1597 रुपये हो गया है | देशी गेंहू का कानपुर की मंडी में पिछले हफ्ते भाव था 1800 रुपये | कल इसका भाव घटकर 1537 रुपये हो गया | आज इसका भाव बढ़कर 1719 रुपये हो गया है | देशी गेंहू का झाँसी की मंडी में पिछले हफ्ते भाव था 1825 रुपये | कल इसका भाव बढकर 2004 रुपये हो गया | आज इसका भाव और बढ़कर 2167 रुपये हो गया है | पीछे जाने के लिए 0 दबाएँ |	The price for Deshi wheat in Delhi market was Rs. 1750 last week, yesterday it went up to Rs. 1829. Today it came down to Rs. 1597. The price for Deshi wheat in Kanpur market was Rs. 1800 last week, yesterday it went up to Rs. 1537. Today it has gone upto to Rs. 1719. The price for Deshi wheat in Jhansi market was Rs. 1825 last week, yesterday it went up to Rs. 2004. Today it came down to Rs. 2167. To go back to earlier menu, press 0

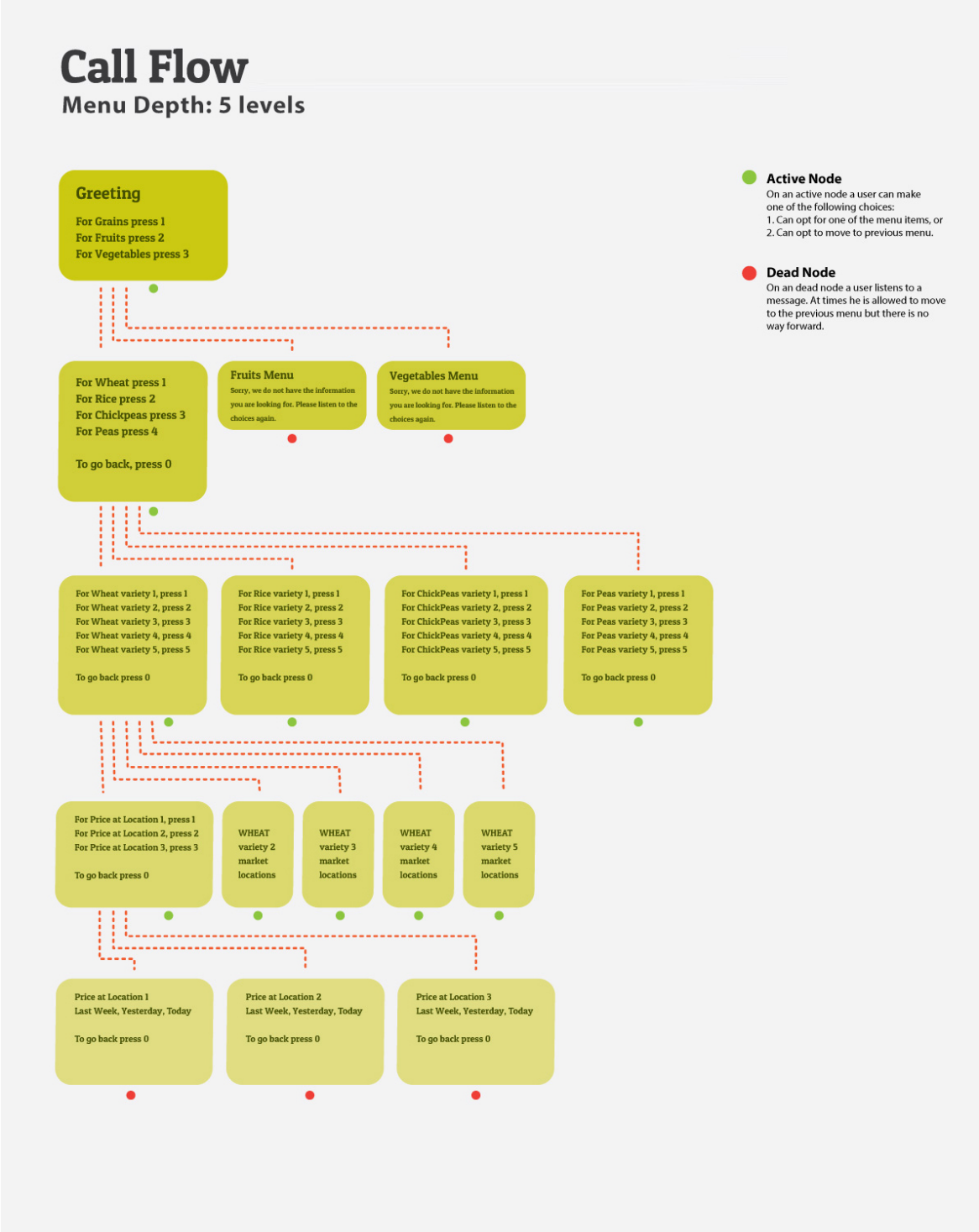
Appendix 2: Call flow for shallow menu depth prototypes.



Appendix 3: Example service prompts with deep menu depth.

Audio prompts for deep menu depth prototypes.		
Menu depth	**In Hindi, as used in test prototype**	**Translated text**
1	नमस्कार। बाज़ार भाव सेवा में आपका स्वागत है। यहाँ आप विभिन्न कृषि उत्पादों का प्रति क्विंटल भाव जान सकते हैं। अन्न-धान्य का भाव जानने के लिए 1 दबाएँ। फलों का भाव जानने के लिए 2 दबाएँ। सब्जियों का भाव जानने के लिए 3 दबाएँ।	Greetings! Welcome to the agricultural commodity market price service. You can use this service to know the per quintal market prices for different agricultural commodities. To know the prices for Grains, press 1 To know the prices for Fruits, press 2 To know the prices for Vegetables, press 3
2	चावल का भाव जानने के लिए 1 दबाएँ। गेहूं का भाव जानने के लिए 2 दबाएँ। चने का भाव जानने के लिए 3 दबाएँ। मटर का भाव जानने के लिए 4 दबाएँ। पीछे जाने के लिए 0 दबाएँ।	To know the prices for Rice, press 1 To know the prices for Wheat, press 2 To know the prices for Chickpeas, press 3 To know the prices for Peas, press 4 To go back to earlier menu, press 0
3	देशी गेहूं का भाव जानने के लिए 1 दबाएँ। मध्यम गेहूं का भाव जानने के लिए 2 दबाएँ। लाल गेहूं का भाव जानने के लिए 3 दबाएँ। शरबती गेहूं का भाव जानने के लिए 4 दबाएँ। सफ़ेद गेहूं का भाव जानने के लिए 5 दबाएँ। पीछे जाने के लिए 0 दबाएँ	To know the prices for Deshi wheat, press 1 To know the prices for Madhyam wheat, press 2 To know the prices for Lal wheat, press 3 To know the prices for Sharbati wheat, press 4 To know the prices for White wheat, press 5 To go back to earlier menu, press 0
4	दिल्ली की मंडी के भाव के लिए 1 दबाएँ। कानपुर की मंडी के भाव के लिए 2 दबाएँ। झाँसी की मंडी के भाव के लिए 3 दबाएँ। पीछे जाने के लिए 0 दबाएँ।	To know the prices in Delhi market, press 1 To know the prices for Kanpur market, press 2 To know the prices for Jhansi, press 3 To go back to earlier menu, press 0
5	देशी गेंहू का दिल्ली की मंडी में पिछले हफ्ते भाव था 1750 रुपये। कल इसका भाव बढकर 1829 रुपये हो गया। आज इसका भाव घटकर 1597 रुपये हो गया है। पीछे जाने के लिए 9 दबाएँ।	The price for Deshi wheat in Delhi market was Rs. 1750 last week, yesterday it went up to Rs. 1829. Today it came down to Rs. 1597. To go back to earlier menu, press 0

Appendix 4: Call flow for deep menu depth prototypes.



Appendix 5: Post-test questions used in exploratory study 2.

क्या एक एच आई वी पीड़ित माता पिता का बच्चा दुसरे बच्चों के साथ खेल सकता है ?

1. हाँ २. नहीं ३. पता नहीं

सुनने में आया है की एच आई वी होने पर पीड़ित व्यक्ति ज्यादा दिन जीवित नहीं रह सकता |

1. सही है २. गलत है ३. पता नहीं

एच आई वी पीड़ित को पालतू जानवर नहीं पालने चाहिए?

1. सही है २. गलत है ३. पता नहीं

ऐसी किसी स्थिति में जबकि हमें परिवार के किसी सदस्य को तुरंत ही डाक्टर के पास ले जाना पड़े तो आप कहाँ जायेंगे ?

1. घर के पास वाला अस्पताल जहाँ एच आई वी रोगी भी आते हैं,
2. घर से बहुत दूर एक सामान्य अस्पताल जहाँ एच आई वी का इलाज़ नहीं होता|

सावधानियां लेने के अलावा एक एच आई वी पीड़ित के परिवार के सदस्यों को प्रोफिलेक्तिक (prophylactic) दवाइयों की खुराक भी ले लेनी चाहिए ?

1. हाँ २. नहीं ३. पता नहीं

दुर्घटना कि स्थिति में एच आई वी पीड़ित के खून के संपर्क में आ जाने पर भी क्या एच आई वी से बचना संभव है ?

१. हाँ २. नहीं ३. पता नहीं

अगर आपके पिछले प्रश्न का उत्तर हाँ है तो बताइए कि यह कैसे संभव है ?

1. मुहं पर रुमाल और हांथों में दस्ताने पहन कर
2. समय पर सही दवाइयां लेने से
3. यौन सम्बन्ध न रखने से

क्या एक एच आई वी पीड़ित किसान अपने खेतों में काम कर सकता है ?

1. हाँ २. नहीं ३. पता नहीं

एच आय वी एड्स के फेलने की बात करते हुए आप नीचे दिए गई लाइनों में से किस लाइन के शब्दों का उपयोग करेंगे ?

1. एच आई वी, माँ, बच्चा, खाना
2. पिता, एच आई वी, बच्चा, खेलना
3. बच्चा, माँ, एच आई वी, माँ का दूध

एच आई वी शरीर की रोग प्रतिरोधक क्षमता को कम करता है | ऐसा होने पर भी क्या यह संभव है की हम रोज़ मर्रा में होने वाले सामान्य संक्रमण (इन्फेक्शनों) से बचे रहें ?

1. हाँ २. नहीं ३. पता नहीं

एक एच आई वी पीड़ित जो कि दवाइयां ले रहा है, उसके लिए निम्न में किसकी सम्भावना ज्यादा है ?

1. कि उसकी म्रत्यु निकट है
2. कि वह एक लंबा लेकिन रोगग्रस्त जीवन जीयेगा
3. कि वह एक सामान्य जीवन जीयेगा

क्या एक एच आई वी पीड़ित माँ अपने एच आई वी पीड़ित बच्चे को अपना दूध पीला सकती है ?

1. हाँ २. नहीं ३. पता नहीं

एक एच आई वी पीड़ित बच्चे को ऐ आर टी दवाइयां नहीं लेनी चाहिए क्यूंकि ये दवाइयां सिर्फ वयस्कों के लिए होती हैं ?

१. हाँ २. नहीं ३. पता नहीं

Appendix 6: Common audio script used in exploratory study 2.

नमस्कार! आज हम एच आई वी, एड्स और इससे जुड़े हुए पहलुओं पर बात करेंगे |

एच आई वी एड्स क्या है?

हम सभी के शरीर में बाहरी बीमारियों से लड़ने की शक्ति होती है | यही कारण है कि हम खांसी- ज़ुकाम या हलके बुखार जैसी छोटी मोटी बिमारियों होने पर भी कई बार खुद ही ठीक हो जाते हैं |

एच आई वी एक विषाणु या वायरस है जो शरीर की इसी शक्ति पर सीधा-सीधा हमला करता है | जिसके कारण कई सामान्य सी बीमारियाँ मौका पाते ही शरीर को अपना घर बना लेती हैं | शरीर की इस कमजोर दशा को हम एड्स कहते हैं |

एच आई वी विषाणु एक शरीर से दुसरे शरीर में कैसे पहुँच सकता है?

एच आई वी विषाणु सिर्फ मनुष्यों में फैलता है | यह एक शरीर से दुसरे शरीर में चार तरह से पहुँच सकता है –

पहला एच आई वी संक्रमित व्यक्ति के साथ बिना किसी सुरक्षा के यौन सम्बन्ध बनाने पर,

दूसरा संक्रमित व्यक्ति का खून दुसरे शरीर मैं पहुचने पर,

तीसरा संक्रमित माँ के द्वारा शिशु को जन्म देने, और

चौथा संक्रमित माँ के द्वारा शिशु को अपना दूध पिलाने पर |

एच आई वी विषाणु शरीर पर कैसे हमला करता है?

हमने पहले जाना की हम सभी के शरीर में बाहरी बीमारियों से लड़ने की शक्ति होती है | यह शक्ति उन अनगिनत कोशिकाओं या सेल्स से आती है जो खून में रहकर अलग- अलग काम करती हैं | इनमें से कुछ कोशिकाओं को हम CD4 कहते हैं | ये खून में पहुचने वाले रोगों के कीटाणुओं को मारने का काम करती हैं | एच आई वी विषाणु के खून में दिखते ही ये कोशिकाएं इससे लड़ने की भी कोशिश करती हैं | लेकिन एच आई वी विषाणु इन कोशिकायों से ज्यादा ताकतवर होता है | इसलिए बजाए इसके की ये कोशिकाएं एच आई वी को कमज़ोर कर पाएं, एच आई वी इनको नष्ट करना शुरू कर देता है | इस तरह इन कोशिकायों के कम होने पर शरीर की बाहरी बिमारियों से लड़ने की शक्ति कम होने लगती है और शरीर रोगग्रस्त हो जाता है |

एच आई वी का क्या इलाज़ है ?

दुर्भाग्यवश एक बार खून में पहुचने पर अभी तक एच आई वी विषाणु को पूरी तरह से मारना संभव नहीं है, लेकिन हम उसके खून में फेलने की गति को धीमा जरूर कर सकते हैं | इलाज़ के इस तरीके को ART कहा जाता है | इस इलाज़ के असर से शरीर की बाहरी रोगों से लड़ने की शक्ति फिर से बढ़ने लगती है | इस प्रकार एच आई वी पीड़ित भी दवाइयां लेते हुए एक स्वस्थ और स्थिर जीवन जी सकता है |

ऐ आर टी दवाइयों को नियम से लेना क्यों ज़रूरी है ?

एच आई वी विषाणु कुछ समय में ऐ आर टी दवाइयों से लड़ना सीख लेता है | यह समय हर एच आई वी पीड़ित के लिए अलग- अलग होता है | लेकिन हाँ, अगर एक एच आई वी पीड़ित दवाइयां ठीक से न ले या लेना बंद कर दे तब यह समय जल्दी भी आ सकता है | और ऐसा समय आने पर एच आई वी को खून में धीमा करना काफी महंगा और कठिन है | इस तरह पहले ली गई ऐ आर टी दवाइयाँ का नियमित पालन करना ही सबसे ज्यादा लाभदायक है |

Appendix 7: Continuous audio script used in exploratory study 2.

आईये अब एच आई वी से जुड़े हुए कुछ अनबुझे पहलुओं पर भी बात करें |

हम सबकी जिंदगी में ऐसा मौका भी आ सकता है की हम जाने-अनजाने ही किसी ऐसे व्यक्ति के संपर्क में आयें जिसके खून में एच आई वी हो | उदाहारण के तौर पर हम ऐसे व्यक्ति के साथ ट्रेन में ताश या मैदान में क्रिकेट भी खेल सकते हैं | ऐसी स्थिति में सही जानकारी ना होने पर पता चलते ही बहुत सारे लोग घबरा जाते हैं और पीड़ित व्यक्ति से दूर भागने की कोशिश करते हैं | लेकिन सच्चाई यह है की एच आई वी पीड़ित व्यक्ति के खांसने से, छींकने से, उसके साथ में खाना खाने से या ताश- क्रिकेट जैसा कोई खेल खेलने से एच आई वी नहीं फैलता |

लेकिन यह तब तक ही ठीक है जब तक कि एच आई वी पीड़ित के शरीर से खून ना बह रहा हो | खून बहने की स्थिति में आपको बहुत सावधान रहना होगा | ध्यान रहे कि आपके शरीर के उस भाग पर, जो कि एच आई वी पीड़ित के खून के संपर्क में आएगा, जैसे कि हथेली, पैर, चेहरा या गर्दन पर, कटने या छिलने का एक भी निशान ना हो | हो सके तो दस्ताने (gloves) पहेन लीजिए | आपको यह भी ध्यान देना होगा कि एच आई वी पीड़ित का खून आपकी आँखों में या आपके मुंह में नहीं जाए | क्यूंकि किसी भी खुले हुए निशान से और आँखों में या मुहं में जाने पर भी एच आई वी विषाणु एक स्वस्थ व्यक्ति के खून में पहुँच सकता है | अगर आप एच आई वी पीड़ित के खून के संपर्क में आ ही जाते हैं तो तुरंत एच आई वी डाक्टर से मिलिए | ऐसे मामलों में डाक्टर एक विशेष तरह की बचाव की दवाइयां देते हैं जो एच आई वी को आपके खून में जड पकडनेसे रोक सकती हैं | इन दवाइयों को Prophylactic कहा जाता है | याद रखिये कि ये दवाइयां एच आई वी खून के संपर्क में आने के कुछ ही घंटों में लेनी होती है |

बहुत सारे लोगों का यह भी मानना है कि एक एच आई वी पीड़ित जोड़े को स्वस्थ और एच आई वी रहित बच्चा नहीं हो सकता | सच्चाई यह है कि एक स्वस्थ बच्चा पाने के लिए उन्हें डाक्टर की सलाह मानते हुए कुछ सावधानीयां बरतनी होगी | बच्चे के माँ के पेट में आने तक माता-पिता दोनों को और फिर बच्चे के जन्म लेने तक माँ को, अपने खून में एच आई वी की संख्या पर दवाइयों की मदद से काबू रखना होगा | बच्चे का जन्म भी डाक्टरी निगरानी में होना बहुत ज़रूरी है | जन्म के बाद माँ को ध्यान रखना होगा की वह बच्चे को अपना दूध न पिलाये | इतनी सावधानी बरतने पर भी रुपये में एक पैसे के बराबर संभावना बनी रहती है की बच्चा एच आई वी पीड़ित पैदा होगा पर यह मान लेना की एच आई वी पीड़ित जोड़े के स्वस्थ और एच आई वी रहित बच्चा नहीं पैदा हो सकता, अपने आप में गलत ही है | एच आई वी के खून में होते हुए भी माता- पिता अपने बच्चों को सभी कि तरह स्वाभाविक प्यार दुलार दे सकते हैं |

और अंत में यह एक बार और दोहरा लें कि एच आई वी विषाणु का खून में पाए जाना कोई म्रत्युदंड नहीं है | यह सही है कि एच आई वी विषाणु शरीर की बाहरी बिमारियों से लड़ने की ताकत को कम करता है | जिसके कारण कई सामान्य सी बीमारियाँ मौका पाते ही शरीर पर हमला बोल देती हैं | ठीक से इलाज न होने पर यही छोटी मोटी बीमारियाँ गंभीर और जानलेवा रोगों का रूप ले लेती हैं | लेकिन इन रोगों का इलाज़ होने पर और व्यक्ति के ऐ आर टी दवाइयां नियमित रूप से लेने पर एच आई वी पीड़ित भी एक स्वस्थ और स्थिर जीवन जी सकता है |

आशा है यह जानकारी आपके लिए उपयोगी रहेगी | धन्यवाद |

Appendix 8: Script for Quiz audio script used in exploratory study 2.

अब हम आपसे एच आई वी एड्स से जुड़े हुए पांच सवाल पूछेंगे | हर सवाल के आगे कुछ उत्तर दिए जायेंगे| आप सही उत्तर का चुनाव करने की कोशिश कीजीए |

सोचिये की आप ट्रेन में सफर करते हुए एक यात्री के साथ ताश खेल रहे हैं | बातों ही बातों में पता चलता है की वह एच आई वी पीड़ित है | ऐसा पता चलने आप क्या करेंगे?

अगर आप सोचते हैं कि "आप ताश खेलते रहेंगे" तो एक दबाइए |

अगर आप सोचते हैं कि "आप ताश खेलना बंद कर देंगे"तो दो दबाइए |

अगर आप सोचते हैं की "आप अपनी सीट बदल लेंगे" तो तीन दबाइए |

आपका जवाब बिलकुल सही है / आपका जवाब गलत है | एच आई वी पीड़ित व्यक्ति के पास बैठने से, हाथ मिलाने से, उसके छिंकने से, या उसके साथ ताश या क्रिकेट जैसे खेल खेलने से एच आई वी नहीं फैलता |

एच आई वी क्लीनिक से बाहर आते ही एक आदमी या औरत की रास्ते पर दुर्घटना हो जाती है | वह घायल है और उसके शरीर से खून बह रहा है | ऐसी स्थिति में आप क्या करेंगे ?

अगर आप सोचते हैं कि "एम्बुलेंस आने तक आप उस व्यक्ति का खून रोकने की कोशिश करेंगे" तो एक दबाइए |

या अगर आप सोचते हैं कि "एच आई वी के डर से आप उस व्यक्ति की मदद नहीं करेंगे" तो दो दबाइए |

अ.. म... आपका जवाब सही है कि एम्बुलेंस आने तक आप स्वयं ही उसकी देखभाल कर सकते हैं पर ऐसा करते हुए आपको बहुत सावधान रहना होगा / वैसे तो किसी भी घायल व्यक्ति की मदद करना हर किसी का कर्तव्य है पर इस मामले में आपका डर जायज है | लेकिनआप अगर सावधानी बरतें, तो आप इन हालातों में भी मदद कर सकते हैं |

आपको ध्यान देना होगा कि आपके शरीर के उस भाग पर, जो कि घायल व्यक्ति के खून के संपर्क में आएगा, जैसे कि हथेली, पैर, चेहरा या गर्दन पर, कटने या छिलने का एक भी निशान ना हो | हो सके तो दस्ताने (gloves) पहेन लीजिए | आपको यह भी ध्यान देना होगा कि घायल व्यक्ति का खून आपकी आँखों में या आपके मुंह में नहीं जाए | क्यूंकि किसी भी खुले हुए निशान से और आँखों में या मुहं में जाने पर भी एच आई वी विषाणु एक स्वस्थ व्यक्ति के खून में पहुच सकता है |

अगर आप ऐसे खून के संपर्क में आ हि जाते हैं तो तुरंत एच आई वी डाक्टर से मिलिए | ऐसे मामलों में डाक्टर एक विशेष तरह की बचाव की दवाइयां देते हैं जो एच आई वी को आपके खून में जड पकडने से रोक सकती हैं | इन दवाइयों को Prophylactic कहा जाता है | याद रखिये कि ये दवाइयां एच आई वी पीड़ित के खून के संपर्क में आने के कुछ ही घंटों में लेनी होती हैं |

क्या एच आई वी अपने आप में एक जानलेवा विषाणु है?

अगर आप सोचते हैं कि “एच आई वी अपने आप में जानलेवा विषाणु है” तो एक दबाइए |

अगर आप सोचते हैं कि “एच आई वी अपने आप में जानलेवा विषाणु नहीं है” तो दो दबाइए |

आपका जवाब बिलकुल गलत तो नहीं है / अरे वाह! आपका जवाब सही है, लेकिन ध्यान रहे कि एच आई वी बहुत ही खतरनाक है | एच आई वी होने पर भी मरीज़ की मौत एच आई वी से नहीं होती | असल में एच आई वी विषाणु शरीर की बाहरी बिमारियों से लड़ने की ताकत को कम करता है | जिसके कारण कई सामान्य सी बीमारियाँ मौका पाते ही शरीर पर हमला बोल देती हैं | ठीक से इलाज न होने पर यही छोटी मोटी बीमारियाँ गंभीर और जानलेवा रोगों का रूप ले लेती हैं | लेकिन इन रोगों का इलाज़ होने पर और व्यक्ति के ऐ आर टी दवाइयां नियमित रूप से लेने पर एक एच आई वी पीड़ित भी एक स्वस्थ और स्थिर जीवन जी सकता है |

क्या एच आई वी पीड़ित जोड़े को एक स्वस्थ और एच आई वी रहित बच्चा हो सकता है?

अगर आप सोचते हैं कि “एच आई वी पीड़ित जोड़े को एक स्वस्थ और एच आई वी रहित बच्चा हो सकता है” तो एक दबाइए|

अगर आप सोचते हैं कि “एच आई वी पीड़ित जोड़े को एक स्वस्थ और एच आई वी रहित बच्चा नहीं हो सकता है” तो दो दबाइए|

आपका जवाब सही है/ आपका जवाब सही नहीं है | एच आई वी पीड़ित जोड़े को एक स्वस्थ और एच आई वी रहित बच्चा बिलकुल हो सकता है | लेकिन इसके लिए उन्हें डाक्टर की सलाह मानते हुए कुछ सावधानीयां बरतनी होगी | बच्चे के माँ के पेट में आने तक माता-पिता दोनों को और फिर बच्चे के जन्म लेने तक माँ को, अपने खून में एच

आई वी की संख्या पर दवाइयों की मदद से काबू रखना होगा | बच्चे का जन्म भी डाक्टरी निगरानी में होना बहुत ज़रूरी है | जन्म के बाद माँ को ध्यान रखना होगा की वह बच्चे को अपना दूध न पिलाये | इतनी सावधानी बरतने पर भी रुपये में एक पैसे के बराबर संभावना बनी रहती है की बच्चा एच आई वी पीड़ित पैदा होगा पर यह मान लेना की एच आई वी पीड़ित जोड़े के स्वस्थ और एच आई वी रहित बच्चा नहीं पैदा हो सकता, अपने आप में गलत ही है |

एक एच आई वी पीड़ित माँ को क्या नहीं करना चाहिए?

अगर आप सोचते हैं कि “एच आई वी पीड़ित माँ को अपने बच्चे को अपने हांथों से खाना नहीं खिलाना चाहिए” तो एक दबाइए|

अगर आप सोचते हैं कि “एच आई वी पीड़ित माँ को अपने बच्चे को अपनी गोद में नहीं सुलाना चाहिए” तो दो दबाइए|

अगर आप सोचते हैं कि “एच आई वी पीड़ित माँ को अपने बच्चे को अपना दूध नहीं पिलाना चाहिए” तो तीन दबाइए|

आपका जवाब गलत है / आपका जवाब बिलकुल सही है | एच आई वी वायरस बच्चे को अपने हाथ से खाना खिलाने से या अपनी गोद में सुलाने से नहीं फैलता | लेकिन अगर एच आई वी पीड़ित माँ बच्चे को अपना दूध पिलाये तो यह विषाणु बच्चे के खून में फैल सकता है |

इसी के साथ ही सवाल जवाब का यह क्रम यहीं समाप्त होता है| धन्यवाद |

Publications arising from this research

[1] Shrivastava, A. (2011). Spoken Dialog System: Gaining insights into developing a model from a game of cards. In IndiaHCI (pp. 111–114). New York, USA: ACM.

[2] Shrivastava, A., & Joshi, A. (2014). Effects of visuals, menu depths, and menu positions on IVR usage by non-tech savvy users. In Proceedings of the India HCI 2014 Conference on Human Computer Interaction - IHCI ’14 (pp. 35–44). New York, New York, USA: ACM Press.

[3] Rashinkar, P. G., Joshi, A., Rane, M., Sali, S., Badodekar, S., Emmadi, N., Shrivastava, A. (2011). Healthcare IVRS for Non-Tech-Savvy Users. In A. Holzinger & K.-M. Simonic (Eds.), Information Quality in e-Health, Lecture Notes in Computer Science (Vol. 7058, pp. 263–282). Berlin, Heidelberg: Springer Berlin Heidelberg.

[4] Joshi, A., Emmadi, N., Rashinkar, P., Malandkar, P., Shrivastav, A., Srivastava, S., & Rajput, N. (2012). Visual IVRs for Low-literate, non-literate and Low-tech-savvy Users. In CHI 2012.

Acknowledgements

I take the opportunity to reflect back on my research journey here, and to thank everyone who have made this journey not only possible but also memorable.

I would like to begin by thanking my supervisor Professor Dr. Anirudha Joshi for his continual support and encouragement throughout this research project. I remember the time when I appeared for the interview and suggested a preference to pursue research work in HCI4D domain. From that day till the time I pen down this section of the thesis, I cannot thank him enough for placing his trust in me. I thank him for giving me this opportunity and for his guidance and motivation at times when I needed it the most. I also thank him for making me design and code all the test prototypes used in this research. I learnt some really good lessons there. He has not just been my mentor and guide but also a harsh critique, a humble co-researcher and a keen facilitator. Thanks you for wearing these many hats, sir!

I wish to now thank Professor Ravi Poovaiah, Professor Gaur G. Ray and Professor Ramesh Bairy T.S. for their sustained interest in this research work. They have been a great panel with constructive comments and feedback all throughout this time. Their comments have made us consider interesting and important perspectives which we may have otherwise missed. They have not only questioned me to ignite my imagination, but also showed possible directions to achieve appropriate answers. I

thank them for all of their interest and guidance, and for all the time which they had given me individually as well as collectively.

I would like to make a special mention of Bruce Balentine, Senior Scientist, Enterprise Integration Group (Zurich, Switzerland) and Dr. Bernhard Suhm, IVRs design expert and researcher along with my friend Dr. Dirk Schnelle Walka. In particular, I came across Bruce first as a reader of his books on IVRs design. Later he obliged to participate in numerous discussions on matters of my research topic. I thank him sincerely for showing interest in my research work and for engaging with me so selflessly over distances and across time-zones at all different times. He has been an inspiration, and a true personification of someone with deep passion and commitment for his research.

I further express my deep gratitude to a group of humble men and women, whom I may not know personally, but who have played a critical and irreplaceable role in my research work. These are our participants, field facilitators, government and bank officials, teachers in rural schools including village committee heads, volunteers and others. They helped us recruit participants, secure locations to conduct assessments, moving across locations, and spreading the word amongst prospective participants. Within them, I call the names of Mr. Madan M. Shrivastava, Mr. Banke B. Shrivastava and Satendra Shrivastava, my cousins from field locations for giving me not just shelter and food but also love, support and company during field visits. Thank you all!

On the personal side, I cannot thank enough but can only express my deep gratitude and love for my siblings Nandini didi, Archana didi, Arvind dada, Ranjana didi, my bhabhi Chetna, my father Shri Brijesh C. Shrivastava, my mother Smt. Basanti D. Shrivastava, and my wife Rashmi. Thank you all for believing in me. You supported me in all possible ways throughout these years. You all loved me unconditionally and trusted me whole heartedly. You always suggested me what can be done, and how can I pull myself in difficult times. Thank you, thank you for encouraging me to chase my dreams and making me feel secure about things back home. Arvind dada, you always encouraged me throughout this journey and made me look at the brighter side of the picture. Rashmi, you were not there when I started this

endeavour. But since the time you were around, your constant push and motivation had helped me sum it up into a logical ending. I also owe thanks to my niblings Vini, Divyansh, Aagrah, Anugrah, Aryan, Sanika, Peehu and Ameya for their innocent questions and speculations about the research topic, and about my life at IIT Bombay in general. I also cannot thank enough Dr. Nina Sabnani, my mentor and friend, for her support and encouragement; and for being on my side all the times whenever needed. And to my friends Dr. Rajendra Patsute, Dr. Sachin Datt, Dr. Prasad Bokil, Dr. Sanjram P. Khanbanga, Nalin, Abhishek Pachori, Anuj and others, both far and near, those with whom I connected frequently as well as not so-frequently, a big thank you to all! Having you all nearby had certainly helped!

Towards the end, I realize that there might still be few who may have left in this acknowledgement. I thank them with all my sincerity. And, at last I express my gratitude to almighty for giving me the necessary strength and courage to pursue this journey.