Data Entry Errors in Rural Context: Evaluation and Design of Efficient Error Limiting Intelligent Interface for Rural and Semi-urban Indian Data Entry Operators

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

By

Shrikant Salve
(Reg. No. 11610506)

Under the Supervision of
Prof. Pradeep G. Yammiyavar FDRS

Department of Design
Indian Institute of Technology Guwahati
Guwahati - 781039, INDIA

March 2017

Department of Design
Indian Institute of Technology Guwahati
Guwahati Assam -781039
India

DECLARATION

I hereby declare that the work contained in this thesis entitled "Data entry errors in rural context: Evaluation and design of efficient error limiting intelligent interface for Rural and semi-urban Indian data entry operators" is my own work done under the supervision of Professor Pradeep G. Yammiyavar, at the Department of Design, Indian Institute of Technology Guwahati (IITG), Assam. I hereby declare that to the best of my knowledge, it contains no materials previously published or written by another person, or substantial proportion of material which have been accepted for the award of any other degree or diploma at IITG or any other educational institute, except where due acknowledgement is made in this thesis. Any contribution made to the research made by others, with whom I have worked at IITG or elsewhere, is explicitly acknowledged in the thesis. I also hereby declare that the intellectual content of this thesis is the product of my own work, and as per general norms of reporting research findings, due acknowledgements have been made wherever the research findings of other researchers have been cited in this thesis.

Place: Guwahati
Date:

Shrikant Salve
Department of Design
IIT Guwahati, Assam - 781039

Department of Design
Indian Institute of Technology Guwahati
Guwahati Assam -781039
India

CERTIFICATE

This is to certify that the work contained in this thesis entitled "Data entry errors in rural context: Evaluation and design of efficient error limiting intelligent interface for Rural and semi-urban Indian data entry operators" submitted by Mr. Shrikant Salve to the Indian institute of Technology Guwahati, Assam (India) for the award of the degree of Doctor of Philosophy has been carried out under my supervision. This work has not been submitted elsewhere for the award of any other degree or diploma.

Place: Guwahati
Date:

Prof. Pradeep G. Yammiyavar FDRS
Supervisor, Department of Design
IIT Guwahati, Assam - 781039

DEDICATION

I would like to dedicate my thesis to my beloved father 'Vitthalrao' and mother 'Sindhu'.

Acknowledgements

There are many people who have directly or indirectly contributed to the completion of this thesis. Before presenting this work in the conscious circle, I would like to take the opportunity to humbly and solemnly acknowledge all of them for their constant support and guidance.

First and foremost, I would like to thank my supervisor Professor Pradeep G. Yammiyavar, Department of Design, IIT Guwahati for his guidance and intellectual inputs throughout the completion of this thesis. He was the one who inspired me to pursue research and have nurtured me to whatever I am today as a researcher. He showed tremendous confidence in me and was very patient, for which I shall always be indebted to him. Apart from research, his solemn care and support have helped me in becoming a better person. In his active guidance I have learnt a lot and I express my deep gratitude and respect for the same.

I pay my heartiest and sincere thanks to the members of doctoral committee Prof. Debkumar Chakrabarti, Dr. Sougata Karmakar, Department of Design, IIT Guwahati and Dr. Vinayak Kulkarni, Department of Mechanical Engineering, IIT Guwahati for their valuable advice, guidance and suggestions which helped me to grow during the doctoral research period.

My sincere thanks and due reverence to three rural-BPOs- *Source2Rural*, *RuralShores* and *Maitreya*, who supported and allowed me to carry out experiment with their premises. My sincere thanks to Dr. Patil, Principal, Gautam Polytechnic Institute Kolpewadi and Dr. R. A. Kapgate, Principal, Sanjivani K.B.P. Polytechnic Kopargaon for permitting us to collect data in their institutes. I thank to Shanu Shukla, Department of Psychology, IIT Indore for helping in conduction of psychological experiment.

I would like to thank Rashmi Jain, Pankaj Deharia, Abhishek Vahadane for helping me out with the technicalities during the development of the user interface (experiment tool). I owe to Priya Salve, Abhijit Jadhav, Prashant Salve and Pravin Bagde for proving hole hearted help in data collection work.

I am sincerely thankful to all the faculty and staff members of Department of Design, IIT Guwahati who had helped me in whatever way they could.

My friends and ‘UE - HCI lab’-group at IIT Guwahati have contributed immensely in completion of this thesis. My special thanks go to Yogesh Deshpande, Debayan Dhar, Vikas Kumar, Satish Shivarudriah, Vikramjit Kakati, Sai Prasad Ojha, Venkateshwarlu Varala, Anmol Srivastava, Ravi Ligannavar, Deepshikha, S. Yadav, Gajanan Shelake and Nilesh Patil. Their advice, co-operation and support have made my stay at IITG a memorable experience.

Lastly, I would like to thank my parents, in-laws, sisters and brothers for their love and encouragement. Without their support this journey would not have been this easy. I pay my humblest sense of gratitude and respect to my wife Priya for her love, care, sacrifice and encouragement have made it possible for me to come so far. I appreciate the courage, understanding and dedicated support shown by all of them despite many testing times at their end. The timing of the Ph.D. tenure made me so happy that another family member joined us in life’s journey. He is none other than my son Shreyank.

Last but not least, I pay my sincere thanks, love and respect to all who have directly or indirectly helped during my doctoral thesis work.

Funding Acknowledgment

The undersigned hereby duly acknowledges the Ministry of Social Justice and Empowerment, Government of India for awarding ‘Rajiv Gandhi National Fellowship’ for this research study.

Place: Guwahati Shrikant Salve
Date:

Abstract

With the stupendous rise of Rural- Business Process Outsourcing (Rural-BPOs) in India, employment opportunities have increased greatly for the rural/ village youth as data entry operators – which is one of the essential source of earning for them. Typical services offered by rural-BPOs include data based services and voice based services to outsourcing agencies such as banks, insurance, telecom, micro finance and information technology enabled service companies. The data based services involves digitization, data entry, converting document to different format and many other similar ones. The main focus of this thesis is on data entry performed by operators in Rural based BPOs from India. The data entry work done by an operator (also called as 'data entry operator') at rural-BPO involve transcribing information from paper forms into computer databases. It is challenging task for many smaller rural-BPOs working in developing country like India to maintain high quality during data entry (also called as transcription). One of the reason of extra effort is lower usability factor of software employed for data entry. There is also lack of expertise in designing user interfaces for such data entry software, especially failing to address localised specific field constraints that can, if incorporated, ensure high quality of transcription with low rate of errors. There are other issues related to data entry work done at rural-BPOs, like transcription process (paper to digital) for double entry is expensive and time consuming, the poor quality of mobile data entry and failure to rectify specific field constraints. There may be cultural issues / challenges like differences between local spoken language and input language (English) by data entry operators working at rural-BPOs - all of which needs to be investigated. In this theses such factors are under investigation.

It is imperative to measure the usability of user interfaces used by rural-BPOs by factors such as: time to learn, speed of performance, rate of error by users, retention (memory) over time and subjective satisfaction. From these factors, human error (error made by 'user' – in this case rural data entry operator) is identified as one of important usability test factor. Therefore, modelling of these errors through experimentation has been of interest in Usability Engineering (UE) research as evidenced by number of papers published in this area. This thesis reports studies that have been done to answer question such as what is the effect of interface designed features on the efficiency (in terms of errors i.e. accuracy and time i.e. speed) of data entry operators? What is the effect of local language on data entry?

The research literature also highlights that sensitive variables like- perceived cognitive load, perceived system usability, user interface satisfaction, willingness to continue usage and relative advantage, have been ignored while designing the interfaces for data entry.

Therefore, to address the above challenges we have conceived, prototyped and developed *ELIIDE* - tool after studying existing literature, data collection and usability aspects of rural-BPOs. We named it as *ELIIDE*- tool, it is supported with intelligent features like- (i) display of autocomplete suggestion for text field by ranking strategy based on likelihood, (ii) predictive text entry widget, (iii) radio button pointed with most likely options and (iv) dynamic drop-down split-menu. The interface uses local Marathi language to communicate with user / operator. The communication happens in terms of error and feedback messages. This additional feature may support rural users to get emotionally attached to interface.

The experiments were conducted to compare two user interfaces, one is newly designed interface (*ELIIDE*- tool) and second is the existing user interface the operator currently uses for data entry. The participants including 224 professionals (rural-BPO operators and polytechnic students training to become BPO operators) were volunteered for the study. Prior to the actual experiment, the participants were explained about the design and purpose of user interface and also provided practice session on it. Before going for the actual experiment the participants were told to fill pre-test questionnaires which include demographic information. Each participant performed four tasks, two tasks were data entry on existing interface (having static widgets) and other two were on the intelligent interface (having dynamic widgets). The sequence of the task was random to avoid learning effect. The tasks consist of a transcription of given data entry form (refer Figure 4.2) (also called as paper form) into electronic form using both interfaces. Participants were instructed to perform the tasks as quickly and accurately as possible. The computer based background recording of each participant interaction with the designed user interface have taken for calculation of the accuracy and speed. After completion of the experiment the participants were instructed to fill the post-task questionnaires to express their opinion and experience about the user interface. The subjective experience was recorded in terms of cognitive load, perceived system usability, user interface satisfaction, willingness to continue usage and relative advantage.

The t-test and ANOVA analysis technique were adopted for the analysis of the data. Results highlight there is significant difference between intelligent user interface and existing user interface for errors and time. It has also been observed that *ELIIDE* -tool can affect operators subjective experience.

Therefore, we conclude that, intelligent user interface design features do affect the operator's performance in terms of accuracy and speed. Also, it decreases operators' cognitive load, increases system usability and user satisfaction. This is because the intelligent features of *ELIIDE* like dynamic, predictive, adaptive and probabilistic. It may be due to the incorporation of user specific features related to local language and error prompting specific to errors for the rural group of operators. Also it is inferred by feedback that users are willing to continue using this interface for data entry. The thesis argues for incorporating user specific prompting local features that intelligently cater to the group's error patterns over general prompts that user software normally provides. It posits that local language prompts with voice over are more acceptable to rural operators over screen based visual prompts alone.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction: Improving Work Efficiency of Rural-Business Process Outsourcings'

This Chapter introduces the research work reported in this thesis. It describes the background and motivation behind this research work. The Chapter starts with a brief introduction to the area of research and its background. The research issues were introduced and placed in the content of their multidisciplinary background of human-computer interaction, usability engineering and information technology. Research gaps of the thesis are highlighted which lead towards formulation of research questions. The boundaries and scope of the thesis is laid out along with definitions and taxonomy. The Chapter concludes with summaries of all the Chapters.

1.1.Introduction

TO ERR IS HUMAN. It is the nature of the human being to make errors such as mistakes, slips, lapses, miscalculations etc. while using everyday devices like turning on the heat under on an empty kettle or pressing buttons on remote controllers or computers. People have a tendency to blame themselves for human errors, but human error is often invoked in the absence of technological explanations (Sears & Jacko, 2009). Chapanis (1999) wrote back in the 1940s that 'pilot error' was real 'designer error'. This was a challenge to contemporary thinking and showed that design is all important in human error reduction. Half a century after Chapanis's original observations, the idea that one can design error-tolerant devices is beginning to gain acceptance (Baber & Stanton, 1994). One can argue that human error is not a simple matter of one individual making one mistake, so much as the product / system of a design which has permitted the existence and continuation of specific activities which could lead to errors (Reason, 1990). However, newer Graphical User Interfaces (GUIs) with newer modes of interaction such as gestures lead to new learning and adapting situation for the user or operator of complex interactive systems. Coping by the user has always been a matter of training. Therefore, costs of training and costs of unintended error causing grave situations are always involved whenever new software has been introduced. In life critical systems like hospitals, aviation, railway systems and nuclear power plant disastrous outcomes have been reported regularly in the daily press worldwide due to the consequence of human error or operator error. The analysis of these human errors is very vital for helping future error prevention and build error recovery as part of the system. Human-computer interaction research reports many investigations done in this area in the past decade. With the advent of software as an important component of work processes over the past two decades - Usability of software in terms of addressing 'error' is the area of interest in this thesis.

Human-Computer Interaction (HCI) is a discipline concerned with the design, implementation and evaluation of interactive computing systems for human use and with the study of major phenomenon surrounding them (Hewett, *et al*., 1992). HCI is an interdisciplinary field which involves various disciplines like computer science, psychology sociology and anthropology and industrial design etc. (Shneiderman & Plaisant, 1987). The interaction between user and computer occurs at the user interface, which includes both software and hardware. The software interfaces consist of application software, system software and development software. Therefore, every software engineer and designer wants to build high-quality interfaces that are admired by users and are easy to use. The graphical user interface (GUI) is not only to be appreciated for flamboyant aesthetics or stylish visuals, but rather for inherent quality features such as usability, functionality, universality and usefulness. It becomes imperative to measure the usability of GUI's by means of five factors such as time to learn, the speed of performance, a rate of error by users, retention (memory) over time and subjective satisfaction (Dix, Finlay, Abowd, & Beale, 2003). From these factors, human error (error by the user) is identified as one of important usability measurement test factor. Therefore, modelling of these errors through experimentation has been of interest in Usability Engineering (UE) research as evidenced by a number of papers published in this area. It is imperative to study- (i) what these errors are and how frequently they occur. (ii) How severe is their impact (iii) What contributes to forming these errors (iv) How can GUIs contribute in reducing errors.

Usability: Dix, Finlay, Abowd, & Beale (2003) have stated that there are three 'use' words that must all be true for a product (user interface) to be successful; it must be:
useful - accomplish what is required: play music, cook dinner, format a document;
usable - do it easily and naturally, without danger of error;
used - make people want to use it, be attractive, engaging, fun, etc. (Dix, Finlay, Abowd, & Beale, 2003)

According to Jakob Nielsen (Nielsen, 2012), usability is a quality attribute that assesses how easy user interfaces are to use. The word 'usability' also refers to methods for improving ease of use during the design process. Usability is defined by five quality components as given below.

1. *Learnability*: How easy is it for users to accomplish basic tasks the first time they encounter the design?
2. *Efficiency*: Once users have learned the design, how quickly can they perform tasks?
3. *Memorability*: When users return to the design after a period of not using it, how easily can they re-establish proficiency?
4. *Errors*: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
5. *Satisfaction*: How pleasant is it to use the design?

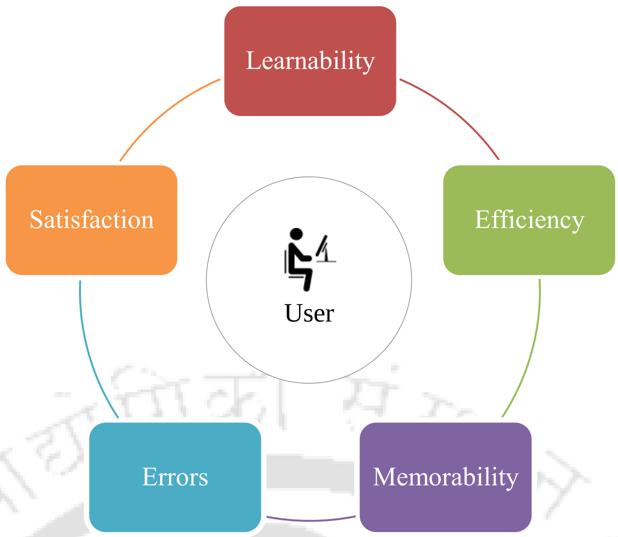


Figure 1-1: Figure depicting Nielsen's five component of usability

As shown in Figure 1-1 above, all usability attributes are interconnected with each other, but the errors can have a major influence on all other. So our usability research interest is in 'human errors' that are important in the context of rural computer users' environments.

1.2. Human Error

The human-being is the one whom computer systems are designed to assist and not the other way round. Requirements of the user should be given first priority while designing the user interface of computer systems. However, human performance can be affected by different factors like age, state of mind, physical health, attitude, emotions and tendency of making errors or mistakes. It is the nature of human being to make errors, and any system should be designed to reduce such errors and to minimize the consequences when errors happen. Traditional approaches have attributed errors to individuals. Over the last several decades' specialists like psychologists, system developers and much more, have been involved in discussions about the error for different perspectives. Therefore, human error is an emotional topic too and psychologists have been investigating its origins and causes since the beginning of the discipline (Reason, 1990).

Cognitive psychologists have considered the issues of error classification and explanation (Senders & Moray, 1991). The taxonomic approaches of (Norman, 1988) and Reason (1990) have fostered the development and formal definition of several categories of human error (e.g., capture errors, description errors, data-driven errors, associated activation errors, and loss of activation errors) while the work of Reason (1990) and

Wickens (1992) attempted to understand the psychological mechanisms which combine to cause errors (e.g., failure of memory, poor perception, errors of decision making, and problems of motor execution). Reason (1990) in particular has argued that we need to consider the activities of the individual if we are to be able to identify what may go wrong. Activity analysis could be a source of identifying and rating errors. Rather than viewing errors as unpredictable events, this approach of activity analysis regards them to be wholly predictable occurrences based on an analysis of an individual's actions.

1.2.1. Defining Human Error

While there are many working definitions of errors in literature a working definition of error representing the essential psychological characteristics and its principal types is relevant to this thesis. One of the most widely accepted definition of human error is, *Errors are all those occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome and when these failures cannot be attributed to the intervention of some change agency* (Reason, 1990). According to James Reason (Reason, 1990), human errors are divided into two major categories: (a) slips & lapses and (b) mistakes. Slips and lapses are errors which result from some failure in the execution and/or storage stage of an action sequence, regardless of whether or not the plan which guided them was adequate to achieve its objective. For example, typing an incorrect word or number, typing a number twice. Mistakes may be defined as deficiencies or failures in the judgemental and/or inferential processes involved in the selection of an objective or in the specification of the means to achieve it, irrespective of whether or not the action directed by this decision scheme run according to plan. For example, mistakenly typing both first and last name in the first name field.

Table 1-1: Definitions of error from literature

Definitions of error	Authors
The state of believing what is untrue. Something incorrectly done.	Webster's new world dictionary
A divergence between the action actually performed and the action that should have been performed An action or event whose effect is outside specific tolerances required by a particular system	Senders and Moray (1991)
Experiments in an unkind environment	Rasmussen (1982)
The debit side of what are useful and essential mental process	Reason (1990)

The Table 1-1, illustrates few selected definitions of error according to different aspects like cognitive, physical, consequence of actions and philosophical issues, as they appear in the literature. A study on errors conducted by (Sauro, 2012), has broadly classified errors as (a) slips and (b) mistakes. 'Slips' are the 'unintended action' a user makes while trying to do something on an interface even though the goal is correct (e.g., a typo) as types of errors. Following are some further examples of slips cited by Sauro. When the goal is wrong it is a mistake, even if that goal was accomplished. Following Table 1-2 shows examples of slips and mistakes.

Table 1-2: Broad categories of errors (Sauro, 2012)

Slips	**Mistakes**
Mistyping an email address	Clicking on a heading that isn't clickable
Mistyping a password	Intentionally double clicking a link or button
Picking the wrong month when making a reservation	Typing both first and last name in the first name field
Clicking Reset instead of Submit button	Entering today's date instead of the date of birth
Mistyping an email address in the re-enter email address field	Replying to all in an email instead of just one person
Accidentally clicking an adjacent link	Entering hyphens in your bank account number
Accidentally double clicking a button	

From Sauro's study, it is observed that even a 'taken for granted' task such as feeding information into a computer has many types of situations and errors. This has implications for training new computer users. Can the interface be designed to avoid such errors being made? Though (Sauro, 2012) has made an interesting study on human error he did not state elaborately on, how one can mitigate the situation (i.e. error situation)? In this thesis, Sauro's broader view of slips and mistakes has been adopted for the basis of discussion.

1.3.The Taxonomy of Errors

There are many schemes suggested by researchers (Rumelhart & Norman, 1982; Salthouse, 1986; Lang, Graesser, & Hemphill, 1991; MacKenzie & Tanaka-Ishii, 2007; Oladimeji et al., 2011) for classifying errors. Each one has constructed the 'taxonomy of error' for their specific purpose. Therefore, there is no single universally agreed classification of human error. Error can be classified according to whether they occur at a skill, rule or knowledge-based level (Rasmussen, 1983; 1986); whether they are slips or lapses (automaticity errors) or mistakes (conceptual errors) (Norman, 1983); and according to whether the error occurs at a task, semantic or interactional level (Maran, 1981; Devis, 1983). The text entry by physical keyboard typing has been studied by many researchers (Rumelhart & Norman, 1982; Grudin, 1984). Gentner *et al.*, 1984 have found that there is large percentage of typing errors such as substitutions, insertions and omissions. The other errors like transposition error, doubling error, alternation error, homologous error, capture error, phonetic swap; type of errors found in transcription typing.

Table 1-3: Classification of errors found in literature

Classification of Errors	**Authors**
Transposition error, Doubling error, Alternation reversal error, Homologous error, Capture error, Omission error, Misstroke error	Rumelhart & Norman, 1982
Substitutions, insertions, omissions, transposition error, doubling error, alternation error, homologous error, capture error, phonetic swap	Gentner *et al.*, 1984
Affordance Errors, Message misinterpretation errors, Goal induced errors Option identification errors (menu option), Status acquisition, Incomplete procedure, Pre-requisite action not performed, Generic command, Mode errors, Superstitious	Lang, Graesser, & Hemphill, 1991
Errors of omission, Errors of commission, Errors of Selection, Errors of Sequence, Errors of Timing, Errors of Quantity	Byren, 1997
Lexical error, Syntax error, Semantic error	Yeum *et al.*, 2005
Transposition Error / reverse digit error, Doubling error / double entry error, Alternation error, Homologous error Capture error, Phonetic swap	MacKenzie & Tanaka-Ishii, 2007
Missing decimal, Skipped, Transposition, Wrong digit, Missing digit	Oladimeji et al., 2011

The Table 1-3 depicts the classifications error in data entry which are found during literature study. A comprehensive list of relevant classification of data entry errors found in this study is prepared. The items in this list are sorted into two broad groups as- text entry errors and numerical entry errors. The text entry errors are classified into six types as- (i) Mistype/ Spelling/ Incorrect: substitutions and intrusions, (ii) transposition, (iii) doubling, (iv) case, (v) capture, phonetic, misinterpretation and (vi) omission/ wrong field.

The numerical entry errors are classified into four types as- wrong, reverse, double and missing.

J. Reason (1990), distinguishes three levels of classification as the (a) behavioural, (b) contextual and (c) conceptual levels,

1. The behavioural (what?): This level classifies errors according to some easily observable features which include formal characteristics of errors like- omission-commission, repetition, misordering or its consequences like nature and extent of damage, injury.

2. The contextual (where?): Acknowledges the critical relationship between error type and the character of the situation or task in which it appears.

3. The conceptual (how?): This level predicts on assumptions about the cognitive mechanism involved in error production. The conceptual level of classification is considered as the most fruitful because they seek to identify the underlying cause in an activity analysis.

Wickens (1992), considered the implications of psychological mechanisms at work in error formation. He discussed that with mistakes- the situation assessment and/or planning are poor while the retrieval action execution is good; with slips- the action execution is poor whereas the situation assessment and planning are good; and finally with lapses- the situation assessment and action execution are good but memory is poor. A summary of these distinctions is shown in Table 1-4.

Table 1-4: Error types and associated psychological mechanisms; Source: Wickens (1992)

Error Type	Associated Psychological Mechanism
Slip	Action execution
Lapse and mode errors	Memory
Mistake	Planning and intention of action
Mistake	Interpretation and situation assessment

Wickens (1992) was also concerned with mode errors, with particular reference to technological domains. He suggested that a pilot raising the landing gear while the aircraft is still on the runway is an example of a mode error. He proposed that mode errors are the result of poorly conceived system design that allows the mode confusion to occur and the operation in an inappropriate mode.

Taxonomies of errors can be used to anticipate what might go wrong in any task. Potentially, every task or activity could be subject to a slip, lapse, or mistake. We have classified 'human-computer interaction errors' as shown in Figure 1-2. Three types of errors reported in it, first computer errors- occurred due to user interface defects, second human error- occurred due to human and last interaction error- error occurred due to wrong interaction or hardware defect. The study reported in this thesis mainly concentrates on human error and what is the effect of user interface design on it?

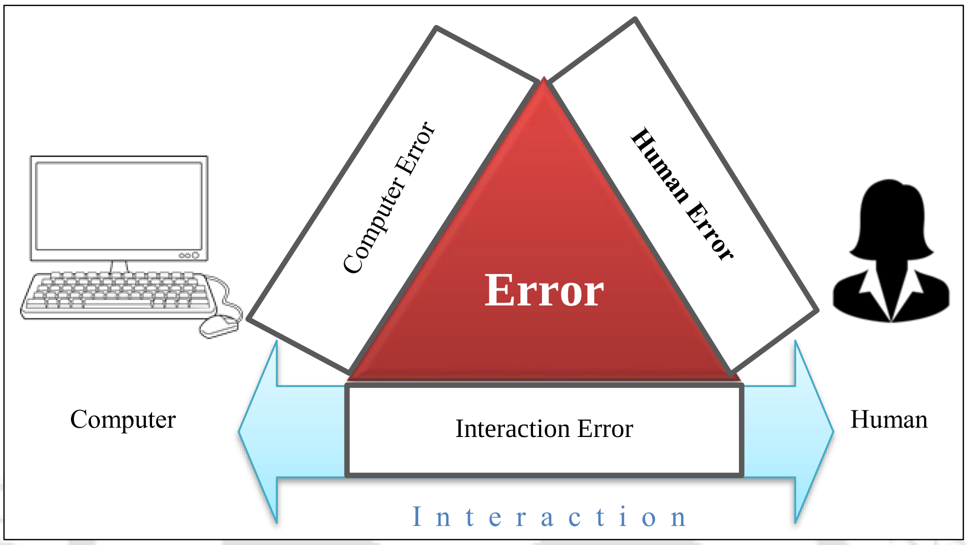


Figure 1-2: Errors in human-computer interaction; Source: Author-generated

1.3.1. Error Types

Simple slips or mistakes distinction is not sufficient to capture all of the basic error types (Reason, 1990). So, mistakes were divided into two kinds: rule-based mistakes and knowledge-based mistakes. Therefore, three error types are (a) slips or lapses, (b) rule-based mistakes and (c) knowledge-based mistakes. Reason and Rasmussen (1983) have stated different error types observed in carrying out an action sequence. As shown in Table 1-5, the three errors types are classified according to the cognitive stages at which they occur. First, planning refers to identifying the goal and deciding the means to achieve it. The second is storage stage of some variable duration required to formulate the intended actions and running them off. The last execution stage consists of the actual implementation of stored plan. The Table depicts the relationship between these three stages and the error types.

Table 1-5: Classifying the error types according to the cognitive stages

Cognitive Stage	Error Type
Planning	Mistakes (rule-based and knowledge-based)
Storage	Lapses
Execution	Slips

During the time of data entry, the cognitive stage is at an execution level hence slips occur- which are the common error types that were captured and reported in this study.

1.3.2. Performance Level and Error Type

Rasmussen (1983) distinguished error types according to human performance level, as shown in Table 1-6 below. The performance level is the ability of the individual to engage in problem-solving at the time an error occurred (Rasmussen, 1986). Data entry (or transcription typing) is highly skilled cognitive-motor (Rumelhart & Norman, 1982; Salthouse, 1986) activity, therefore, considerable training has been required to mitigate errors.

Table 1-6: Relating three basic error types to Rasmussen's three performance level

Performance level	Error type
Skill-based level	Slips and lapses
Rule-based level	Rule-based mistakes
Knowledge-based level	Knowledge-based mistakes

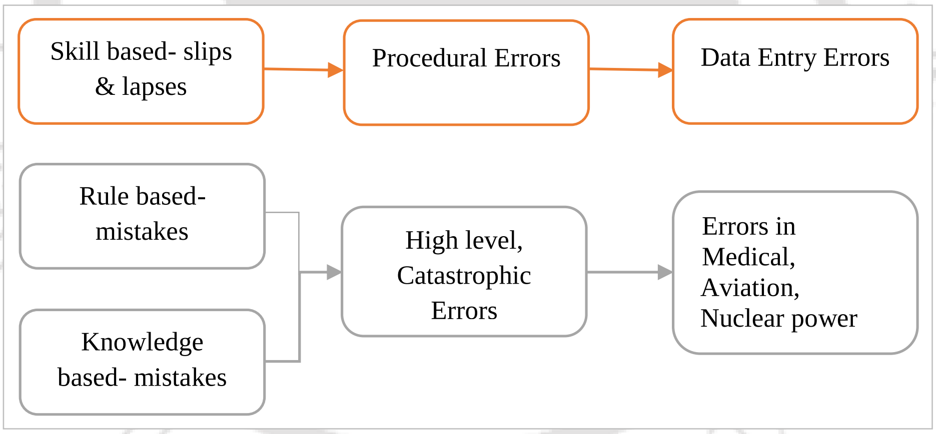


Figure 1-3. Classification of errors adopted from Byrne & Bovair (1997) and Rasmussen (1983)

The skill-based slips and lapses type of errors is procedural errors, for example, data entry errors as shown in Figure 1-3. Procedural errors have received relatively little attention from cognitive psychologists. One reason for this is that, error is frequently considered only as a result or measure of some other variable, and not as a phenomenon in its own right (Byrne & Bovair, 1997). Our aim is to study the operator's skilled motor performance errors that is slips and lapses which are procedural errors.

As depicted in Figure 1-3, mistakes can be high level or catastrophic errors; for example, errors in medical, nuclear power plant, aviation errors.

The generic error modelling system (GEMS) was developed based on Rasmussen's (1983) skill-rule-knowledge classification of human performance. This conceptual framework consists of three basic error types as, skill-based slips (and lapses), rule-based mistakes and knowledge-based mistakes.

Table 1-7: The distinction between skill-based, rule-based and knowledge-based errors

Dimension	Skill-based Errors	Rule-based Errors	Knowledge-based Errors
Types of activity	Routine actions	Problem-solving activities	
Focus of attention	On something other than the task in hand	Directed at problem-related issues	
Predictability of error type	Largely predictable 'strong-but-wrong' errors (actions)	(rules)	Variable
Ratio of error to opportunity of error	Though absolute numbers may be high, these constitute a small proportion of the total number of opportunities for error		Absolute numbers small, but opportunity ratio high
Influence of situational factors	Low to moderate; intrinsic factors (frequency of prior use) likely to expert the dominant influence		Extrinsic factors likely to dominate
Ease of detection	Detection usually fairly rapid and effective	Difficult and often only achieved through external intervention	

The GEMS presents an integrated picture of error mechanisms operating at these three level of performance i.e. skill-based, rule-based and knowledge-based. In the context of data entry at rural-BPOs, we are concentrating on skill-based errors. Therefore, GEMS gives us an understanding of skill-based errors according to different dimensions like types of activity, focus of attention, influence of situational factors and so on, as mentioned in Table 1-7. The 'focus of attention' during skill-based data entry error is something other than the task in hand. This generic error modelling scenario helps us to gain knowledge about influence of situational factors on making data entry slips and lapses. The situational factors like intrinsic factors of operator for example, frequency of prior use, knowledge, skills, can have influence on data entry errors (that is slips and lapses).

1.4.Factors Affecting Performance of Data Entry

Performance Shaping Factor (PSF) provides a measure to account the human performance (Boring, Griffith, & Joe, 2007). PSFs are categorised as internal or external, corresponding to the individual versus situational or environmental circumstances (Rooney, Heuvel, & Lorenzo, 2002).

Therefore, we have extracted few performance shaping factors suggested in literature (Boring, Griffith, & Joe; 2007 and Rooney, Heuvel, & Lorenzo; 2002), that might influence the performance operator during data entry.

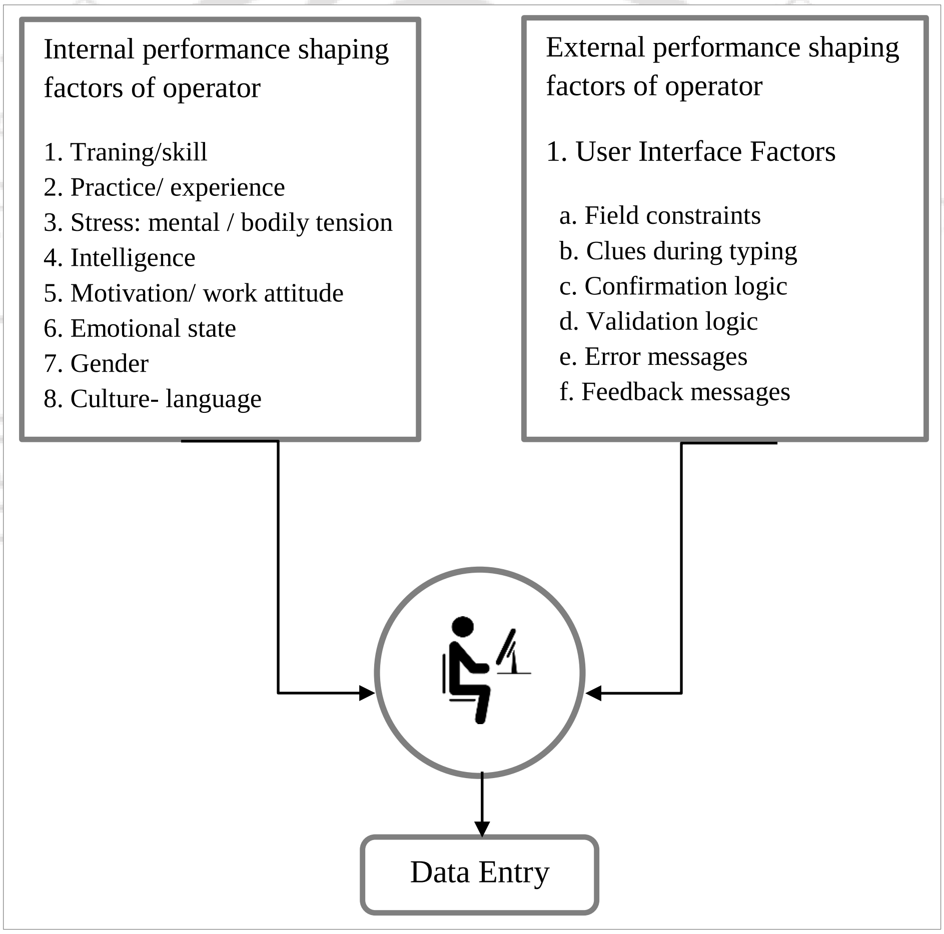


Figure 1-4: Schematic diagram of factors affecting performance of operator during data entry; Source: Author generated

Nielsen J. , 1993 defined the categories of user and individual user differences as depicted in Figure 1-5. He stated two most important issues for usability (a) users' task and (b) their individual characteristics and differences. In Figure 1-5, he called it as 'user cube' of three dimensions along which users experience differs- experience with the system, with the computer in general and with the task domain. We have categorised the operators based on these three dimentions of user experice, which are also related to direct performance shaping factors (Boring, Griffith, & Joe; 2007) shown in Figure 1-4.

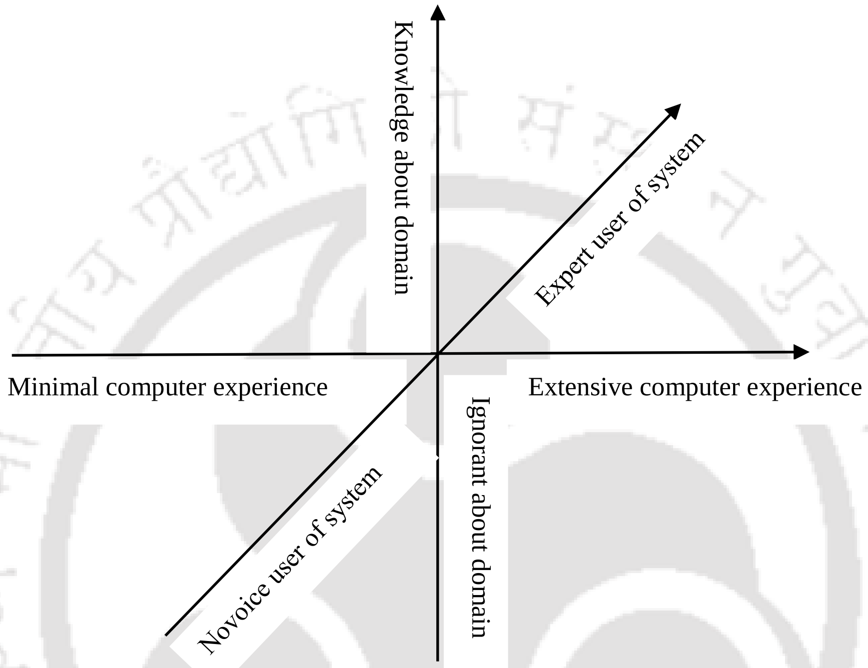


Figure 1-5: Three dimensions of user experience (Nielsen J. , 1993)

Figure 1-4 illustrates the direct (internal) and indirect (external) performance shaping factors those are investigated in this thesis work. Therefore, the internal performance factor like (a) Language- what is effect of local language on data entry? (b) Emotional state- Does emotional state affects the data entry work? (c) Gender- Does female makes less errors compared to male? (d) Skill- skilled operator can be called as expert user of the system (refer Figure 2-5)- Do the expert operator commit less errors compare to novice operator? (e) Experience (refer Figure 2-5)- Do experience operator commit less errors compare to less experienced operator. The external performance shaping factors are related to design of user interface (UI). The UI factor like field constraints, clues during typing, confirmation logic, validation logic, error messages, feedback messages are investigated in this thesis.

1.5.Effect of Language on Rural Computer Users

We can learn to make use of information technology which will empower people with little or no exposure to electronic media. The user is the one whom computer systems are designed to assist. The requirement of the user should be our first priority. In the Indian context, one-way computer users can be categorised as rural / semi-urban and urban users. Many user interfaces have developed and proposed by researchers (Grisedale, Graves, & Grünsteidl, 1997 and Parikh, Ghosh, & Chavan, 2002) for rural Indian users in their local language. The study in this thesis has been conducted on rural/ semi-urban users specifically working in rural areas because of several reasons stated below, which are a strong motivation behind this research work. Researchers (Kam, Kumar, Jain, Mathur, & Canny, 2009; World Bank, 2013), have reported that in India, almost 72% of the population stays in villages with twenty-two regional and two national languages that are Hindi and English being spoken. Although English is widely spoken, it is not comfortable official language (Smith, et al., 2007). About 92.39% schools in rural areas teach in the medium of a regional language (mother tongue) (Meganathan, 2009).

In Indian context, opportunities in rural-BPO sector has been grown up exponentially. The NASSCOM (National Association of Software and Services Companies) report says that in 2010, about 50 rural BPOs employ 5000 people. According to projection in 2015, the 11 rural are staggering 1000 centres and 150000 employees (Ravi & Venkatrama Raju, 2013).

Therefore, leading BPO companies in India such as Infosys, Wipro, TCS are searching their talents from small cities in India to achieve cost efficiency in performing transactional jobs like data entry and form filling. There are more than 50 successful rural centres in India providing BPO services to both domestic and global clients (Ernst & Young, 2011). Ravi & Venkatrama Raju (2013), report says that in Tier II and Tier III cities in India have cornered 38.8% (total 17 cities Tier II cities) and 23% (total 33 Tier III cities) share of the job space respectively in the financial year 2011-12.

There are multiple types of rural-BPOs in India such as BPOs in 'Tier II / Tier III' cities and BPOs that run in villages (village BPO). The Table 1-8 summerises the growth of rural-BPO in India reported by Ravi & Venkatrama Raju (2013).

Table 1-8: Rural BPOs Growth projection report (by 2013-2015) by Ravi & Venkatrama Raju (2013)

	Current			Projection	
Companies	**Customers**	**Centres**	**Employees**	**Centres**	**Employees**
ADF	1	2	550	NA	NA
B2R	3	2	100	100	6,000
DesiCrew	12	5	225	50	5,000
Drishtee	6-7	2	30	NA	NA
eGramIT	15	4	700	30	3,000
Harva	5	3	30	70-100	10,000
NextWealth	NA	2	200	40	1,000
RuralShores	12	6	500	500	100,000
Source For Change	4	1	70	200	10,000
SourcePilani	7	1	60	5	500
Tata Group	4	NA	2000	---	10,000
Total		26	4,465	925	145,500

In the rural and semi-urban area of India, computers are used by people in many places like banks, railways, bus stands, hospitals, factories, marketplaces / shops, government offices, Non-Government Organizations (NGOs) (provides data entry jobs) and Rural-Business Process Outsourcing (Rural-BPOs) for data entry. The task of feeding information into a computer (called as data entry task) has many types of errors called as data entry errors. Simple data entry errors such as typing incorrect number / text, typing a number / text twice or skipping a line can give wrong results.

As shown in Figure 1-6, rural-BPO is one of the few avenues of employment for rural India. Typical services offered by rural-BPOs include data based services and voice-based services to outsourcing agencies such as banks, insurance, telecom, microfinance and information technology enabled service companies (Ravi & Venkatrama Raju, 2013). The data based services involves digitization services, data entry, converting document to different format and much more. The main focus of this thesis is on 'data entry' because which is predominantly observed in rural-BPOs from India. The data entry work is done by the operator (also called as 'data entry operator') at rural-BPO, which involve transcribing information from paper forms into computer databases. It is a challenging task for many smaller rural-BPOs working in developing country like India to maintain high quality during transcription typing. One of the reason is because lack of expertise in designing user interfaces, especially failing to correct specific field constraint and other validation logic. Also the transcription process (paper to digital) for double entry is expensive and time consuming (Chen K., Chen, Conway, Hellerstein, & Parikh, 2011), is of poor quality of mobile data entry (Patnaik, Brunskill, & Thies, 2009) and fails to correct specific field constraints (Broeck, et al., 2007).

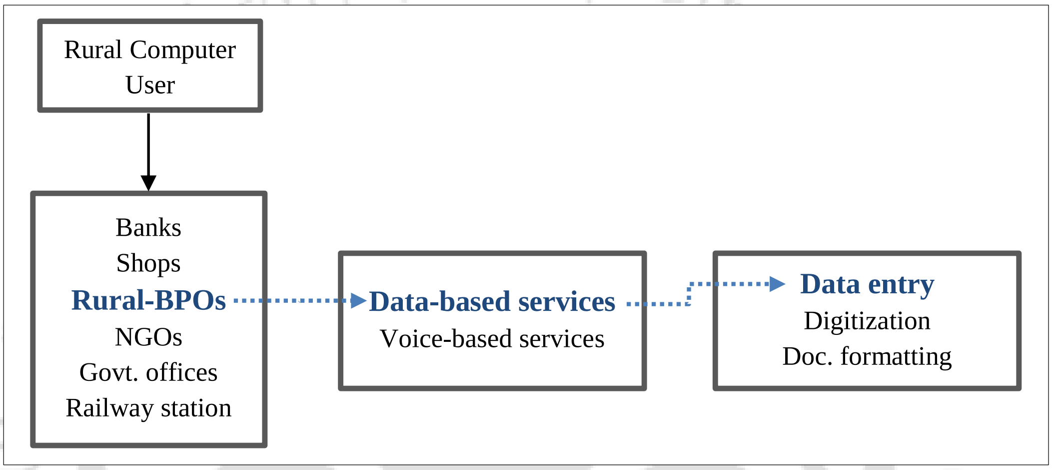


Figure 1.6: Block diagram showing research area as 'data entry at rural-BPOs'

1.5.1. Language versus Cognitive Thinking Strategy of Rural Users during Data Entry

Transcription typing (or data entry) involves complex interaction of perceptual, cognitive and motoric processes (Salthouse, 1986). It has been also observed that, there may be cultural issue/ challenge like local language being different from transcribing language. During data collection, we observed that majority of operators/ workers educated in their mother tongue (local language) and the graphical user interface (GUI) used for data entry is completely in the English language. Therefore, their cognitive thinking is in local language but work language is different. The operators speak in their local language when they are socialising and also at work. The language used during data entry task on computer is English. This means their cognitive thinking and talking language (usually local personal language like *Marathi*, *Assamese* etc.) is different from the transcribing content which is in English. Wanting to work seamlessly between two languages often cause them to make errors and also requires extra time during data entry. Sometimes, the operator gets confused when error or feedback messages appear in English which take time for them to read and understand it. Another issue is a poor design of the GUI which does not provide specific field constraint, clues and confirmation logic in consort with their flow of thoughts. Therefore, to overcome these limitations and overcome data quality challenges, this study introduces a concept of user interface involving intelligent data entry widgets based on quantitative probability. Other researcher like Schlimmer & Wells (1996) have advocated this approach. In the approach taken in this thesis the users are from rural and semi-urban background and work in outsourcing computer data entry enterprises.

1.6. Emotion and Design of User Interfaces

Our 'emotional response' to situations affects our performance level. For example, positive emotions enable us to think more creatively, to solve complex problems, whereas negative emotion pushes us into narrow, focused thinking. A problem that may be easy to solve when we are relaxed will become difficult if we are frustrated or anxious. Psychologists have studied emotional response for decades and there are many theories as to what is happening when we feel an emotion and why such a response occurs (Dix, Finlay, Abowd, & Beale, 2003). Emotion is a salient feature of a human being; and so, it is important to scientifically understand its influence on human behaviour. The effect of emotion on performance has been studied by several researchers (Gray, 2001; Chepenik, Comew & Farah, 2007; Zhu *et al*, 2013) who have concluded that emotion contributes significantly to the performance of different tasks (For example working memory task, decision making task, etc).

Emotion is a temporary fleeting state that emerges from the environment, situation or person himself. Different emotional states subconsciously/consciously exert different effects on the information processing style of the person (Forgas, 2013). For instance, positive emotion signals the familiarity in the environment and hence direct individual towards assimilative processing style. On the contrary, negative emotion identifies a challenging situation and hence calls for externally focused, bottom up and accommodative processing style (Forgas, 2013). In another study, researchers (Clore & Storbeck, 2006) demonstrate that positive affect encourages interpretive or relational processing style and negative affect leads to detailed, stimulus- bound, or referential processing strategy. In terms of visual processing style, researchers (Nath & Pradhan, 2012) found that the positive emotion leads person towards global processing where the focus is on whole stimulus. On the other hand, a person induced by negative emotion focuses on detailed or component part of the stimulus (called, local processing) (Nath & Pradhan, 2012).

To sum up this subsection, one can state that positive and negative emotions have a contrasted processing style - positive emotion activate individuals towards faster response in comparison with that rendered by negative emotion. This viewpoint, however, should be scrutinized under the lens of 'task contextualization'. In a study on motor and movement task (Coombes, et al., 2009), for instance, reaction time of participants' in a negative emotional state is seen to decrease as compared with those with positive and neutral state. Similar findings were highlighted in a study led by Coombes, Janelle and Duley (2005) in the performance of square tracing task where 'approach' and 'avoidance' behaviour style were employed.

Therefore, Youth in developing semi-urban (and rural) India, for whom data entry jobs at rural-BPO provide employment, get emotionally attached to their jobs and become either complacent or anxious due to their performance. This job is their only means of livelihood. In this thesis, we are also investigating the effect of 'emotional state' on the performance of operator in terms of the errors that may creep in. When emotionally vulnerable it is observed that human beings revert back to what is familiar and comfortable. In the BPO operator can they are often observed using / conversation in mother tongue / local language. Does emotional variation lead to language preference? and when switching does it result in errors? is the question.

1.7.Graphical User Interface (GUI) or Software for data entry

During our initial investigation at several rural-BPOs in India, we found that it is challenging task for many smaller rural-BPOs working in developing country like India to maintain high quality during data entry. The process of data entry is also called as transcription. For many reasons, one of the reason of extra effort is lower usability factor of software (or GUIs) employed for data entry. There is also lack of expertise in designing user interfaces for such data entry software, especially failing to address localised specific field constraints that can, if incorporated, ensure high quality of transcription with low rate of errors. There are many user interface factors and individual factors as shown in Figure 1-7 could be involved, which leads to errors. The factors are listed below the Figure 1-7.

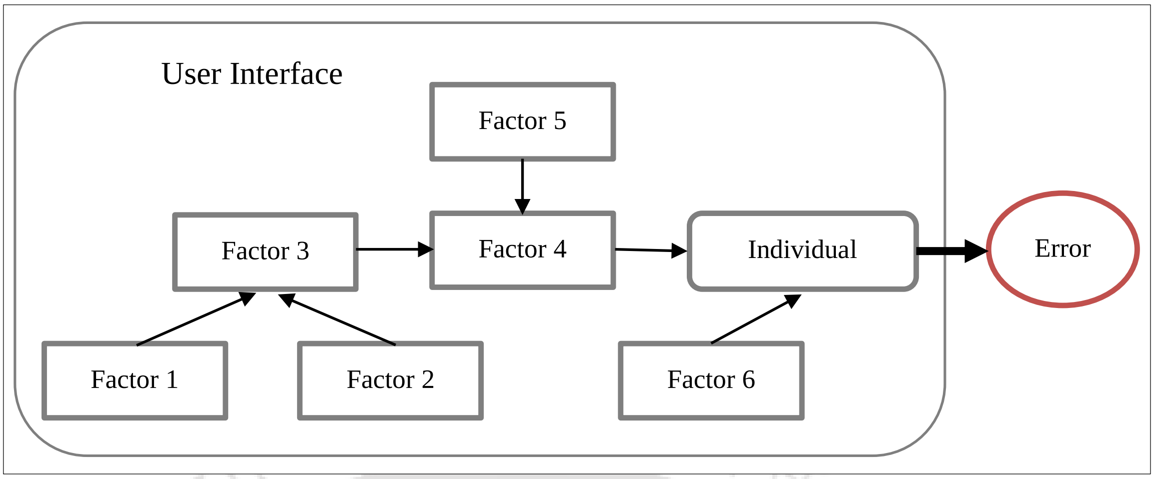


Figure 1-7: Block diagram showing chain of events (factors) leading to an error

Issues (or factors shown in Figure 1-7) with existing data entry user interfaces / software used at rural-BPOs are noted as follows:

A. Poor design of user interface: Lack of expertise in designing user interface especially for operators in rural-BPOs. The existing user interface has some drawbacks mentioned below:

a. Failed to correct specific field constraint.

b. Does not provide clues during typing

c. Fails to provide confirmation logic

d. Does not provide validation logic for fields.

B. Double entry is costly and time-consuming.

C. Data entry using a mobile phone were ever used is of poor quality. Viewing windows are small. Screens are not ergonomic compliant in terms of font size, view window, pixels, layout, colour etc.

D. User interface was in the English language:

a) The majority of operators were educated in their mother tongue (local language) and the user interface (UI) used for data entry is completely in the English language.

b) The operators speak in their local language when they are socialising at work. The language used during data entry is English, this means their thinking and conversing language is different therefore working seamlessly between two languages cause them to make errors and also takes extra time during data entry. This is a potential context for errors.

c) Error or feedback message: The operator gets confused when error or feedback message appear in English which take extra time for them to read and understand it and internalize it.

There are other issues related to data entry work done at rural-BPOs, like transcription process or data entry process (paper to digital) for double entry is expensive and time consuming (Chen K., Chen, Conway, Hellerstein, & Parikh, 2011), the poor quality of mobile data entry (Patnaik, Brunskill, & Thies, 2009) and failure to rectify specific field constraints (Broeck, *et al*., 2007). In this theses such factors are under investigation.

Therefore, to address the issues like local language, emotions, data entry errors and poor design of user interface for data entry, this research study proposed a new graphical user interface (GUI) that is designed with intelligent widgets. The embedded intelligent methods like- machine learning, probabilistic approach and artificial intelligence are proposed to be used in design and development of 'error limiting user interface' for data entry. Further, this new GUI with embedded intelligence to prompt the operator as well as train the operator so as to reduce the errors is intended to be validated by testing.

1.8.Broad Research Gap

The block diagram Figure 1-8, depicts the broad research gaps formulated after studying the background and context of this research work. It infers that a rural-BPO is one of the few avenues of employment for rural India and data entry is the typical service provided by these BPOs.

There are many factors that affect the data entry performed by operator. During our initial background investigated we find that factors like- local language, emotions and design of the graphical user interface, may have an influence on data entry as depicted in Figure 1-8. The number of errors one makes is one of the performance measuring criteria. Therefore, this study investigates the influence of local language factor and emotional state on rural data entry operators.

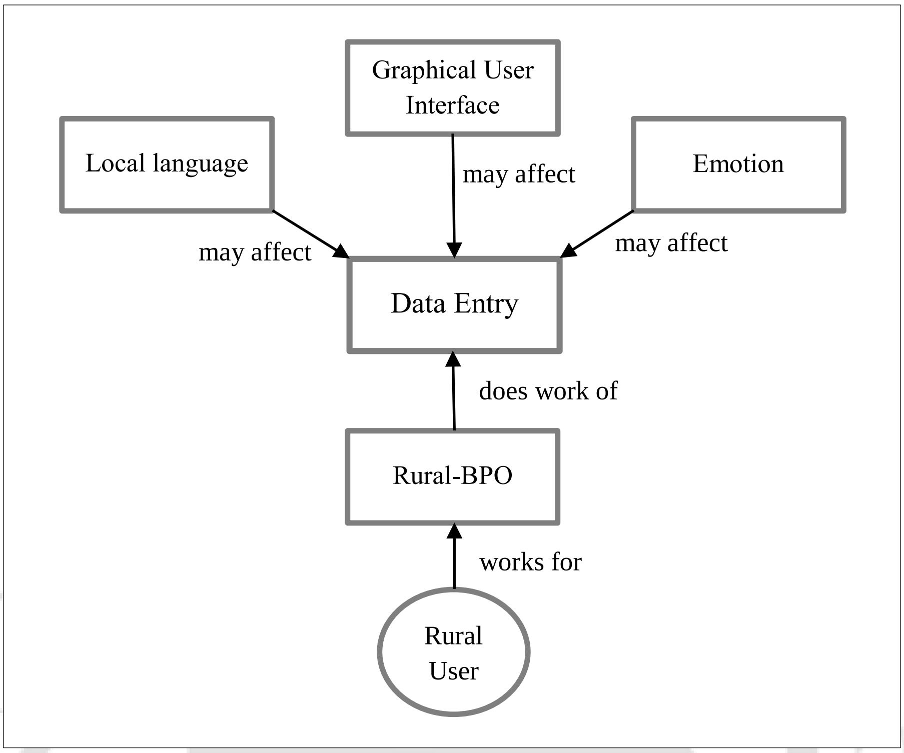


Figure 1-8: Block diagram represent the research gap; Source: Author-generated

1.9. Scope of the thesis

The work in this thesis is in the area of human-computer interaction. It spreads across disciplines such as of human-computer interaction, usability engineering and information technology. Though the current research work reviews literature in "human error" that is 'data entry errors' studies, but the research investigation argues that in order to increase the performance of data entry specifically for operators working in rural-BPO of India the other areas like GUI design, language, emotions should be investigated.

The **data entry error** is the focal point around which research questions have been formulated, hypotheses postulated, experiments conducted, results analysed and conclusions inferred. In this thesis it is argued that there are several factors listed below, which may affect the performance (error/accuracy and time/speed) of these operators.

1. Effect of lower usability factor of software employed for data entry: There is lack of expertise in designing user interfaces for such data entry software, especially failing to address localised specific field constraints that can, if incorporated, ensure high quality of transcription (data entry) with low rate of errors. what is the

effect of interface designed features on the efficiency (in terms of errors i.e. accuracy and time i.e. speed) of data entry operators?

2. There may be cultural issues / challenges like differences between local spoken language and input language (English) by data entry operators - all of which needs to be investigated. In this theses such factors are under investigation. What is the effect of local language on data entry?
3. We are also investigating the effect of ‘emotional state’ on the performance of operator in terms of the errors that may creep in. When emotionally vulnerable it is observed that human beings revert back to what is familiar and comfortable. In the BPO operator can they are often observed using / conversation in mother tongue / local language. Does emotional variation lead to language preference? and when switching does it result in errors? is the question.

During our initial observation we found that in rural-BPO’s the number of female operators are slightly more compared to male. Therefore, we are also investigating that, women are more accurate compared to men during data entry.

1.9.1. Research Questions

The research gaps highlighted above from the perspective of the data entry errors, graphical user interface and the factors identified to capture their effects have been modelled in terms of research questions that would guide this research investigation. The research questions are listed below:

RQ1: What is the effect of a newly configured user interface designed with intelligent features like- (i) display of autocomplete suggestion for text field by ranking strategy based on likelihood, (ii) predictive text entry widget, (iii) radio button pointed with most likely options and (iv) dynamic drop-down split-menu, on accuracy of data entry?

RQ2: What is the effect of user interface designed with intelligent features like- (i) display of autocomplete suggestion for text field by ranking strategy based on likelihood, (ii) predictive text entry widget, (iii) radio button pointed with most likely options and (iv) dynamic drop-down split-menu, on speed of data entry?

RQ3: What is the effect of user interface designed with intelligent features on the variables like- (i) perceived system usability, (ii) perceived cognitive load, (iii) user interface satisfaction, (iv) willingness to continue the usage and (v) relative advantage?

RQ4: Are female operator more accurate in data entry as compared to male?

RQ5: What is the effect of language used on error rate in the case of, (i) English language in forms used for data entry, (ii) Indian (*Marathi*) language and (iii) Mixed language (i.e. both English and Indian (*Marathi*) language combined)?

RQ6: What is the effect of emotions on the data entry error rate?

RQ7: In the case of data entry operators what is the influence of level of knowledge about computer in general on the speed and accuracy of data entry?

RQ8: What is the effect of learned expertise in using a particular system- on the speed and accuracy of data entry?

RQ9: What is the influence of level of understanding of the task domain on the speed and accuracy of data entry?

RQ10: Do experienced operator commit more errors compared to less experienced data entry operator in case of being under pressure such as 'limited time' data entry?

1.10. Overview of the Thesis

The thesis is structured into the following seven Chapters.

Chapter 1: Introduction: Work Efficiency of Rural- Business Process Outsourcings'- describes the background and motivation behind this research work. The research issues were introduced and placed in the content of their multidisciplinary background of human-computer interaction, usability engineering and information technology. Aims and objectives of the thesis are highlighted. The boundaries and scope of the thesis are laid out along with definitions and taxonomy. Summaries of all the Chapters are outlined.

Chapter 2: State of the Art Literature Survey: Understanding Nature of the Problem- A state of the art review of the literature and related work on human errors, data entry. Specifically, previous studied on user interface used for data entry and human errors in data entry. Use of different types of intelligent widgets in designing user interface, to mitigate human errors are identified from past studies. The issues in existing interfaces used at rural-BPOs have been reported. The research gaps and research questions have been highlighted.

Chapter 3: Exploring the Potential of Influence of Errors during Data Entry - This Chapter includes the three pilot studies conducted for exploring the potential of data entry task and influence of errors on data entry in the context of rural Indians. It gives details of a first pilot study conducted on numerical data entry. The second pilot study is on text data entry and the third and last pilot study conducted on the influence of emotional factor on numerical data entry by rural people. Finally, the Chapter mentioned about inferences drawn from the pilot studies.

Chapter 4: Error Limiting Intelligent Interface for Date Entry (ELIIDE)- a Tool- includes the design and implementation of the newly built intelligent error limiting user interface tool. The design of this interface and the comparisons with the existing UIs, is carried out to investigate how much better the new approach is, in terms of usability. The design of the intelligent tool as a metric and as a validating medium for its suitability to rural-BPOs is an important contribution of this thesis. The development of this tool and it's use as an instrument to collect experimental performance data specifically for the rural Indian context is attempted in this thesis as a novel approach to reduce rate of error and thereby contribute to the efficiency of work done by rural-BPOs.

Chapter 5: Experimental Methodology: Research Methods, User Testing, Experiment Design- This Chapter presents the overall experimental methodology used to address the research questions. Initially, the working hypotheses, independent and dependent measures have been listed. The methodology involving- an instrument used for data collection, sampling framework and the procedure adopted for data collection is reported. The experiments and the methodology for them are fully described. All details of how the empirical side of the research has been conducted.

Chapter 6: User Testing / Validation of Designed User Interface- ELIIDE tool: Results and Analysis- involves the evaluation of the newly built user interface that has error prompting in local language capability. This Chapter reports the statistical analysis of the data collected in the experiments. The effect of intelligent features in the user interface on data entry operators have been reported.

Chapter 7: Discussion- discusses the result and analysis done on previous Chapter 6. It also summarises the major findings of this research work and discusses its relation with the theory presented in Chapter 2.

Chapter 8: Conclusion, Contribution and Future Work- Consolidated findings of the experiment and investigation have been reported and the implications of the finding for designers of GUIs have been highlighted. Limitations of the current study have been presented. It also enumerates avenues of future work for further development of the concept and it's more applications.

Chapter 2

State of the Art Literature Survey: Understanding Nature of the Problem

This Chapter reviews the available published literature concerning the broad research gaps illustrated in the introduction Chapter. The Chapter starts with the literature on 'data entry errors' involving numerical and text entry errors and influence of errors on the performance of data entry. The next part Section 2.3 discusses literature on use of interactive devices in the context of rural India. Then Section 2.4 extends the literature study in intelligent features used in user interface. Section 2.4 demonstrates literature on how emotional factor plays role in making data entry. The last Section 2.7 provides the consolidation of theory/ concepts reported from literature.

2.1. Introduction

This Chapter reviews available published literature and identifies and discusses the research gaps. The broad areas reviewed are from human-computer interaction, usability engineering, human errors, cognitive psychology, machine learning and artificial intelligence domains.

The Chapter starts with literature on 'data entry errors' involving numerical and text entry errors and influence of errors on the performance of data entry. The next part Section 2.3 discusses literature on use of interactive devices in the context of rural India. Then Section 2.4 extends the literature study in intelligent features used in user interface. Section 2.4 demonstrates literature on how emotional factor plays role in making data entry. The last Section 2.7 provides the consolidation of theory/ concepts reported from literature. Toward the end, the Chapter revisits the research gaps in the light of literature reviewed and concludes with listing research question and objectives of the study.

2.2. Data Entry Error

Errors inevitably occur every day at work and incur economic costs. Therefore, it is imperative to observe errors during data entry. A detailed study on analysis of single error was done by Smelcer (1989). He estimated of $58 million lost in the US per year due to this single error situation. Card, Newell, & Moran (1983) found in one of their experiment that 26 percent of the total time for text editing was spend dealing with the rectification of errors.

There are two types of data entered- numerical and text. Therefore, this Section is divided into two parts, first involves the literature on numerical data entry and second on text entry. The data entry work has been carried out at different work contexts like banks, medical, aviation, petrochemical plant, nuclear power plant, rural-BPOs etc.

2.2.1. Numerical Data Entry and Errors

Owing to its ubiquitousness, researchers have address several issues related to numerical data entry performance. For instance, Oladimeji *et. al*. (2011) have proposed the study of number entry interface found on medical devices. They reported an experiment that investigates the effect of interface design on error detection in number entry tasks using two number entry interfaces. One serial interface with 12 key numeric keypad and another incremental interface that use a knob or a pair of keys to increase or decrease numbers. 22

participants aged 18-55 years took part in the experiment. A computer was used with an integrated eye-tracker to present the instructions and number entry interfaces. Each participant used both number entry interfaces (independent variable). The dependent variables were the number of undetected errors, number of corrected errors, total eye fixation time and task completion time. The participant was required to enter 100 numbers using both interfaces according to the instruction shown on the right half of the screen. They identified six categories of number entry errors (skipped, transposition, wrong digit, missing decimal, missing digit and other). Their study suggests giving priority to research number entry styles and their relation to error rate, behavior and performance in the context of safety critical number entry systems. However, their study was restricted to medical number entry systems.

Thimbleby et. al. (2010) have proposed the user interface to prevent number entry errors in medical devices. They used three error analysis methods, first a Monte Carlo simulation of number entry with varying error rates, second an exhaustive method where each target number in the range is considered in turn and third is symbolic analysis, where the proportion of blocked out by '*r*' errors is calculated as a function of the underlying keystroke error rates. They proposed new interface which blocks the entry of numbers that do not conform to the specific guidance (ISMP guidance) for the reliable formatting of the number.

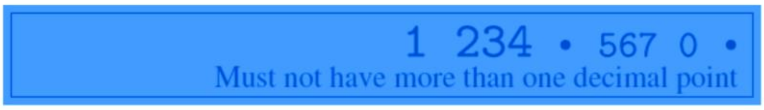


Figure 2-1: Snapshot of error-blocking user interface after an error has occurred
Source: Adopted from Thimbleby et. al. (2010)

The Figure 2-1 depicts, how error blocking mechanism can be implemented on numerical entry interfaces. Their work has shown that user errors are ignored by many number entry systems in user interfaces from interactive devices to desktop applications in all domains which cause confusion, problems and possibly leads to damage. For an operator the twice entered (.) could mean a full stop after a sentence as in language. In maths, it has a meaning of multiplying. A rural operator in a non-English speaking country like India may overlook such difference while being trigger (keyboard) happy by pressing full stop key repeatedly by sheer force of habit. Indian languages use a vertical line to indicate a full stop – not a dot as in English.

Of the various data entry methods found in literature, a research was conducted on the efficiency of three different data entry methods. Barchard et. al. (2011) have projected the study on the impact of human data entry errors on statistical results and calculations. 195 undergraduate students were assigned to three data entry methods: double entry, visual checking and single entry. Participants entered 30 data sheets each containing six types of data. Their results found that visual checking resulted in 2958% more errors than double entry. Also in the double entry, there are significantly fewer errors than single data entry task. This study shows that double entry method is more accurate as compared to visual checking and single entry.

2.2.2. Text Data Entry and Errors

Data entry mechanism becomes pervasive in the area of HCI as it is the first interface for human interaction. Data entry technologies are designed and evaluated through an empirical evaluation of intervention of the user. The literature published in Indian context (i.e on Indian language) is very few in number. One of them is by Ghosh et. al. (2012) on the cost of error correction during *Bengali* text entry. They have proposed algorithms on the error correction quantification problem for Indian language (*Bengali*) transcribed text typed by any single stroke or tap text entry tool. They identify the unequal character (error) positions in both transcribed and presented text both by using longest common subsequence algorithm. They proposed two algorithms first to identify whether the error in simple character or in a complex character. Second to calculate the minimum number of operations to renew transcribed to presented text depending upon error positions and type (error in simple or complex character).

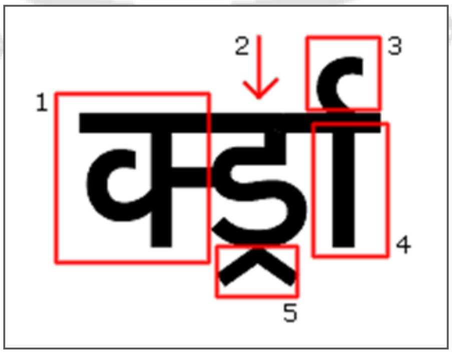


Figure 2-2: Example in Devanagari script
Source: Adopted from (Ghosh, Samanta, & Sarma, 2012)

They also define correction cost per error metric to calculate average correction cost for an erroneously transcribed text. This is the only study based on text entry errors for Indian language. This paper also highlighted the complexity of Indian languages because of complex font typography styles. Figure:2-2 is an example of the complex typography of Indian language *Hindi* in *Devanagari* script. Combining multiple characters like *vattu* (5) a below-base form of a consonant as in Figure:2-2 and *matra* (4) makes it difficult to detect and measure the errors in transcribed texts in terms of the number of characters and similarity of characters. Indian *Devanagari* script is phonetic- each character contributing to a different sound.

MacKenzie & Tanaka-Ishii (2007) have provided the excellent consolidation of text entry systems explaining its mobility, accessibility and universality in research conducted by them. This study helps us to know about different measures of text entry performance like (a) entry rates- words per minute, keystrokes per second; (b) error rates- keystrokes per character performance measure, minimum string distance.

Soukoreff & MacKenzie (2001) have proposed a technique to measure errors in text entry based on the 'Levenshtein minimum string distance statistic'. The proposed algorithm calculates the minimum distance between two strings (i.e. presented string and transcribed string) defined in terms of three edit primitives- insertion, deletion and substitution. The minimum string distance denoted *MSD(A,B)*, where *A* and *B* are character strings.

- Well defined zero: $MSD(A,B)=0$, if and only if $A=B$
- It is bounded: $0 \leq MSD(A,B) \leq max(|A|,|B|)$, where $|A|$ denotes the length of A.
- It is commutative: $MSD(A,B)= MSD(B,A)$

Algorithm for calculating Minimum String Distance (Soukoreff & MacKenzie, 2001):

```
function r(x,y)
  if x=y return 0
  otherwise return 1

function MSD(A,B)
  for i = 0 to |A|
    D[i, 0] = i
  for j = 0 to |B|
    D[0, j] = j
  for i = 1 to |A|
    for j = 1 to |B|
```

D

B \ A=		a	b	c	d
	0	1	2	3	4
a	1	0	1	2	3
c	2	1	1	1	2
b	3	2	1	2	2
d	4	3	2	2	**2**

Lev(A,B)=2

$$\mathbf{D}[\mathbf{i},\mathbf{j}] = \mathbf{min}\begin{bmatrix} \mathbf{D}[\mathbf{i}-\mathbf{1},\mathbf{j}]+\mathbf{1} \\ \mathbf{D}[\mathbf{i},\mathbf{j}-\mathbf{1}]+\mathbf{1} \\ \mathbf{D}[\mathbf{i}-\mathbf{1},\mathbf{j}-\mathbf{1}]+\mathbf{r}(\boldsymbol{A}[\boldsymbol{i}],\boldsymbol{B}[\boldsymbol{j}]) \end{bmatrix}$$

return D[|A|, |B|]

Computing the entries in the matrix 'D' starts in the top-left cell and proceeds to the bottom-right. The value in the bottom-right cell is the minimum string distance. We have adopted this algorithm to measure text entry errors for our newly designed interface meaning tool *ELIIDE* reported in detail in Chapter 5.

Salthouse (1986), reviewed the transcription typing work by using four component heuristic model. The four components (shown in Figure 2-3) consist of an input phase in which to-be-typed text is grouped into familiar chunks, a parsing phase in which the chunks are decomposed into discrete characters, a translation phase in which characters are converted into movement specifications and finally as execution phase in which the actual movements are produced.

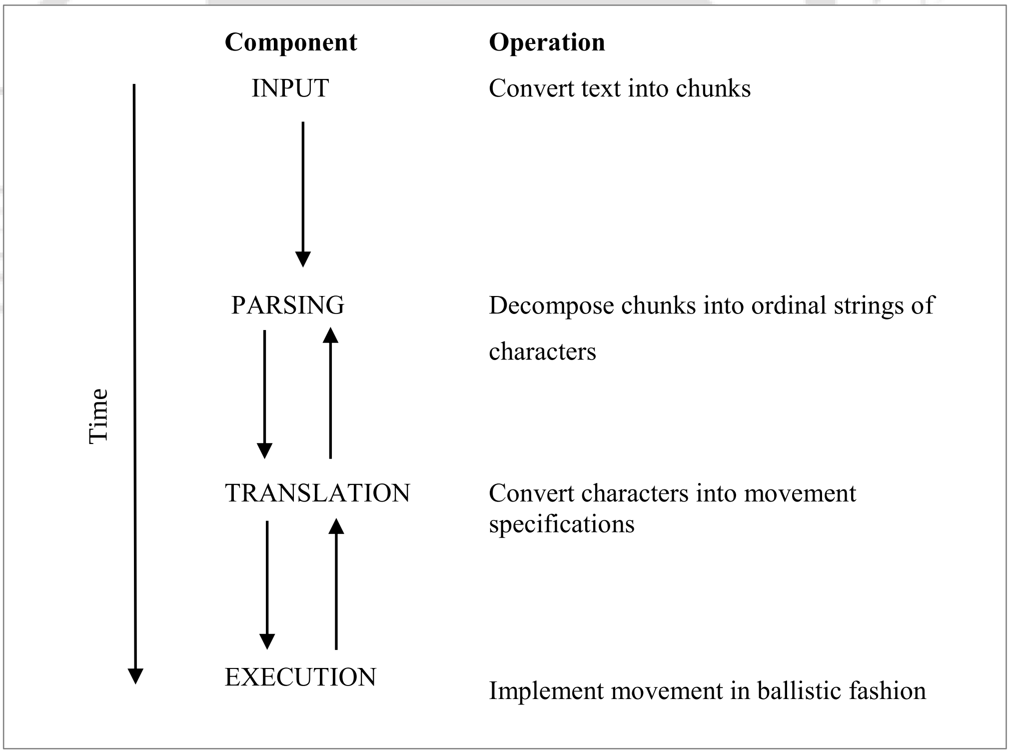


Figure 2-3: Four typing components proposed by (Salthouse, 1986)

Salthouse, (1986) classified text entry errors into four categories: substitutions, intrusions, omissions and transpositions. He illustrated possible determinants for each type of errors according to four processing components. This model helped us to distinguish the possible cause of error for each cognitive stage like input, parsing, translation and execution (shown in Figure 2-3). As we have discussed on page number 8 (Table 1-5), during data entry or text entry the cognitive stage is at an execution level hence slips occur.

Data entry also called as 'text input' research has been carried out to create and evaluate novel text input technologies. Empirical evaluations are conducted to measure speed and accuracy under the controlled environment. Therefore, repeated trials are necessary to generate the volume of paired data consisting of presented text (what subject were asked to enter) and transcribed text (what they actually entered). Transcription typing involves an intricate and complex interaction of perceptual, cognitive and motoric processes.

It is difficult to measure the errors in Indian languages because combination of multiple characters like *vattu* and *matra.* The correction cost per error metric proposed by (Ghosh, Samanta, & Sarma, 2012), can help to calculate average correction cost for erroneous transcribed text in Indian language. For measuring the text entry errors in English language, we have adopted the algorithm proposed by (Soukoreff & MacKenzie, 2001). The study proposed by (Salthouse, 1986), helps us to distinguish errors in transcription typing work done by (data entry) operators into four components (or cognitive stages) like input, parsing, translation and execution. During data entry or text entry the cognitive stage is at an execution level hence slips occur- which are the common error types that were captured and reported in this study.

2.3. Use of Interactive Devices in Rural Indian Context

This section of literature study reports research done on the use of interactive devices like- Kiosk, Mobile phone, ATM machine and Computer and their issues in the context of rural and semi-urban India. Chand (2002) has discerned issues pertaining to the designing of the interface for computer driven kiosks used in rural areas of India. He analyses the kiosk interface based on factors like- motivation, visual interface, mouse-based interaction, navigation, multimedia (e.g. video, animation, text and images). His study raises the importance of use of multilingual text and video contents while developing an interface for illiterate and multilingual rural people. In relation to rural users as a sample population, there is another study by Patel *et al*. (2008) and Patel *et al*. (2009). They have stressed upon the importance of audio and spoken based modality in designing computer-based interface. They designed and developed a voice-based community forum (named *Avaaj Otalo*) interface for Indian rural *Gujarati* users. This application was developed in the *Gujarati* language which allows farmers to receive timely and relevant agriculture information over the mobile phone. They have given importance to the research on the use of spoken language in interactive devices for rural areas of the developing world. The study reported by Gore *et. al*. (2012) on mobile- based application for rural Indian users. They developed a mobile-based collaborative system (social networking) to exchange information amongst rural mobile users in their local language. This application supports voice, video clip and image-based information dissemination. This study showed that voice-based or text-free interaction is appreciated by rural users. Voice as a mean of feedback may have more relevance in a multi-language user scenario such as rural and semi-urban India.

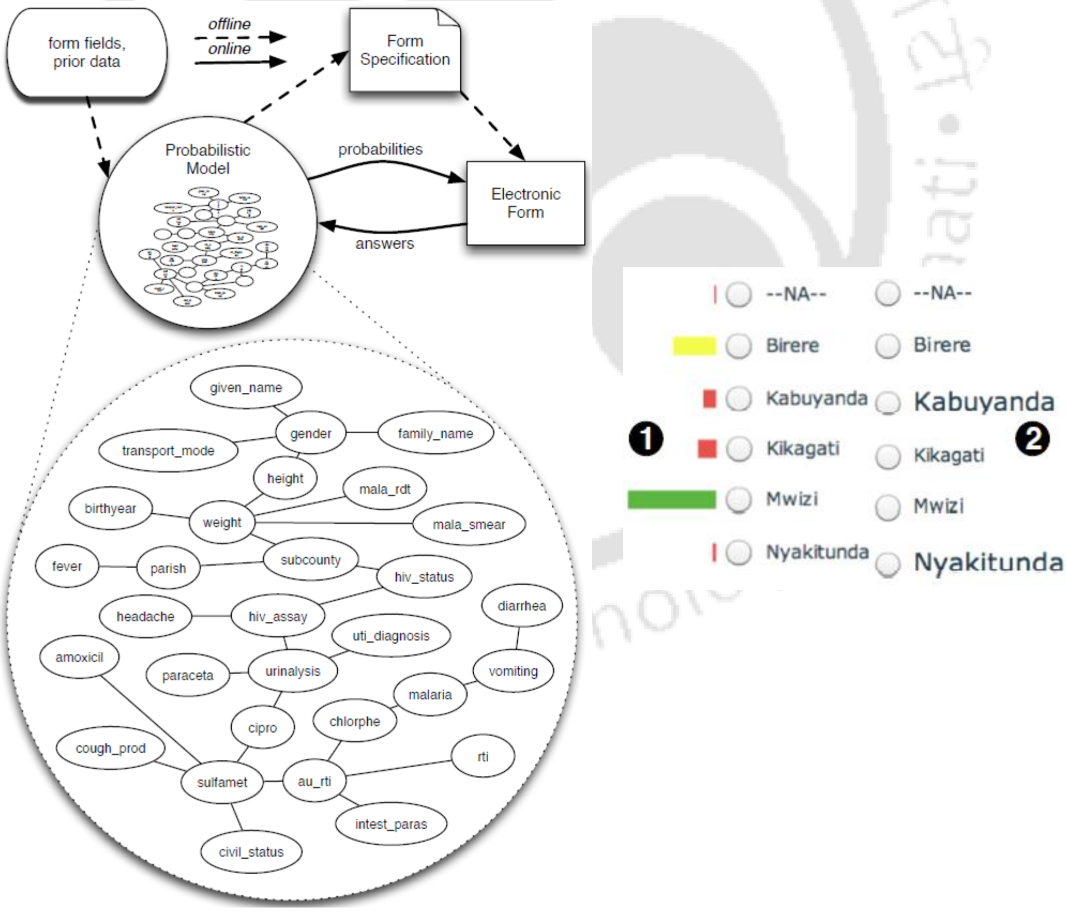
Chand & Day (2006) have proposed *Jadoo*- A paper user interface for people living in rural India. It is a prototype system used by computer literate to create and distribute paper user interface which can be used by computer illiterate to access online information. He stated that illiteracy, the user of a non-native language and fear of technology are big hurdles for rural users in India. Singh *et. al*. (2009) have proposed study on the numeric paper forms used by the NGOs (Non-Government Organizations) for data collection in rural India. They have investigated NGO's form filling requirements which were used to interact with rural people. They proposed the numeric input method for different NGO's form filling requirements which is easy to use for rural people and also machine-readable. The context of this study is 'data entry jobs in both local and English languages provided by NGOs for rural India people', which has provided motivation for working towards the context of data entry job by rural users. But this paper did not reported information about, what kind of data entry job rural people does? what type of user interfaces used for data entry? which need to be explored. This is one of the few literatures we found on data entry in rural India.

2.4.Extended Literature Study on Intelligent features in User Interface

The observational field study and above literature suggest that there is need to redesign the interface used by rural users for data entry. Due to the poor design of the user interface (UI) of the systems used by them which does not provide specific field constraint, clues and confirmation logic. Therefore, we found and reported several areas of related work like managing, improving data entry efficiency and designing of adaptive, dynamic widgets for user interfaces.

Kleinman (2001) has stated the adaptive method for double data entry based on probabilistic approach. In this technique, a probability-based algorithm is proposed that will select the form for double entry based on most likely to contain errors. The algorithm uses ADDER (Adaptive Double Data Entry) to estimate the probability that a particular data enterer has made an error on a given form, then to use that probability as a basis for deciding whether to have that form re-entered. The probability is updated after each re-entered form. The simulation shows that much of the re-entry can be avoided by detecting many errors. This study gives rise to the development of the probabilistic approach for the data entry.

Chen, Hellerstein, & Parikh (2010) and Chen, Chen, Conway, Hellerstein, & Parikh (2011) have proposed a system (named as *USHER*) for data entry form design, entry and data quality assurance.



(a) (b)

Figure 2-4: Components of *USHER* system (a) the arrows showing data flow and zoom circle shows the Bayesian network. (b) different design of radio buttons (1) radio buttons with bar-chart overlay (2) radio buttons with scaled labels, Chen *et. at*. (2010).

USHER learns a probabilistic model over the questions of the form using previous form submissions. Then it applies this model at every step of the data entry process so as to improve the data entry quality. Before the entry *USHER* induces a form layout that captures the most important data values of a form instance. Once *USHER* has been learned, it dynamically adapts the form to the values being entered and enables the real-time feedback to guide the data entry operators toward their intended values. Their results demonstrate considerable improvement in data quality for each component / widget compares to existing practice. We have adapted the technique of probabilistic model based on Baysian Network to develop the relationship between form field.

Lee & Tsatsoulis (2005) have implemented intelligent data entry assistant (called as *SmartXAutofill*) for predicting and automating inputs during entry for XML document. *SmartXAutofill* consists of multiple internal classification algorithms integrated into an ensemble classifier to form single architecture. Each internal classifier uses approximate techniques to propose a value for an empty XML field and through voting the ensemble classifier determines which value to accept. The *SmartXAutofill* system was evaluated using data from eleven different XML domains. This study is limited to XML document domains only.

The machine learning tools can be used to assist repetitive form filling tasks by providing default values for a particular section of the form, which thereby reduces the number of keystrokes necessary to complete a form and also reduces the risk of errors. Hermens & Schlimmer (1994) have developed the user interface (learning apprentice) for repetitive form filling task of 'leave report form'. The authors evaluated the efficiency of this system by measuring keystroke error and prediction errors observed during typing. The results indicate that their method (ID4) reduces number of keystrokes required by 87% compare to non-learning methods. Another empirical study was conducted by Warren (1996) to show the development of an adaptive interface for physician's data entry of electronic medical record. In this interface, he developed short menus that provide a likely selection to user using machine learning technique. The results of this paper indicate the use of machine learning for the development of data entry applications. The usability study conducted by Sears & Shneiderman (1994) indicates that split menus reduce the performance time by 17 to 58%. Thirteen participants were involved in this study. Two different menus design, traditional menu and later slit menus were used by participants for four weeks each. The program created split menus for the font menus (containing 28 items) in MacWrite and Microsoft Word were installed on Macintosh computers at two sites. The statistical t-test shows that split menus resulted in faster mean selection time for each menu and faster selection time for several individual fonts. Also during a usability test, out of 13 participants, nine preferred the split menus. This literature suggests that intelligent methods (like machine learning, probabilistic approach and artificial intelligence) can be used in design and development of user interfaces for data entry.

2.5.Literature on Influence of Emotion on Data Entry

Several studies have proved the influence of emotions on the complex as well as simple tasks in giving rise to human errors. For example, Jeon, Yim, & Walker, (2011) proved that there are behaviour changes when a driver is in different emotional states such as anger and fear, even when those states share the same emotional valence. Similarly, Causse, Dehais, Péran, Sabatini, & Pastor, (2013) confirmed in their study that negative emotional states can provoke plan continuation errors in pilots. This means that they are more likely to continue acting even when available evidence indicates them to stop. Another study, Cairns, Pandab, & Power, (2014) have reported the influence of emotions on number data entry on devices like infusion pumps in hospitals. Their hypothesis was that people who are in the negative affective state will make more errors than those in a positive affective state. The sample size of the experiment consists of 28 participants. The first part involves emotion inducement procedure where participants were shown 24 images for their particular experimental condition and asked to rate each one as they went along. The Microsoft PowerPoint presentation was set up to display 24 images of either positive or negative valence depending on the experimental condition. A standard International Affective Picture System (IAPS) – a database of images was used for inducing a different level of affect in both the valence and arousal dimensions. In the second step, the participants moved on to the number entry task. For this, the Microsoft Excel was used to randomly generate numbers and display them. Participant had to enter the displayed number into a Google Nexus tablet using number pad touch interface. Finally, they concluded that the users in negative affective state are more likely to make number entry errors. This study focuses only on safety-critical environment of number entry in the healthcare domain which involve devices like infusion pump, ventilators having touch screen user interface. Besides, the study did not take into account the arousal dimension of emotions. This study can be applied for rural-BPO context to find out what is the influence of emotion on data entry operators? Because there may be factors like language, job anxiety related to operator which may affect their emotional state which intern affect their performance during data entry work.

2.6. Study of Sensitive Variables

The research literature also highlights that sensitive variables like- perceived cognitive load, perceived system usability, user interface satisfaction, willingness to continue usage and relative advantage, have been ignored while designing the interfaces for data entry.

Stanton, Salmon, Walker, Baber, & Jenkins, (2006), have indicated that mental workload assessment techniques are used to assess the level of demand imposed on an operator by a task or scenario. The NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988) is a multi-dimensional subjective rating tool that is used to derive a mental workload rating based upon a weighted average of six workload subscale ratings. The six sub-scales are mental demand, physical demand, temporal demand, effort, performance and frustration level. The NASA-TLX is the most commonly used subjective mental workload assessment technique. The system usability scale (SUS) offers a very quick and simple to use a questionnaire designed to assess the usability of a particular interface, device or product. The SUS consists of ten usability statements that are rated on a Likert scale of 1 (strongly agree with the statement) to 5 (strongly disagree with the statement). Answers are coded and a total usability score is derived for the interface under analysis.

The interface analysis techniques are used to assess a particular interface in terms of usability, user satisfaction (Stanton, Salmon, Walker, Baber, & Jenkins, 2006). The questionnaire for user interface satisfaction (QUIS) (Chin, Diehl, & Norman, 1988) is a questionnaire method that is used to assess user acceptance and opinions of human-computer interfaces. The QUIS method is used to elicit subjective user opinions on all usability-related aspects of an interface, including ease of use, system capability, consistency and learning. There are a number of different versions of the QUIS method available. QUIS Version 5.5 was selected in this thesis for subjective measurement of interface satisfaction. Each question has an associated rating scale, typically ascending from 1 to 10. Another two scales used in this thesis were defined here. The relative advantage defined as, the degree to which a new system is perceived as being better than old one (Moore & Benbasat, 1991). The extent to which the operator/ user intends to continue to use as data entry interface.

Therefore, the NASA- TLX scale was used to measure the perceived cognitive load. The SUS scale was adopted to find out perceived system usability. The user interface satisfaction was measured using QUIS scale. The measurement of willingness to continue usage and relative advantage were taken by respective scales.

2.7.Consolidated theory/ concept from Literature

The broad literature based review in this study was from usability engineering, human errors in human-computer interaction, artificial intelligence and human psychology domains. A consolidation is presented below.

1. Several researchers (Chand, 2002; Chand & Dey, 2006; Gore, Lobo, & Doke, 2012; Kam, Kumar, Jain, Mathur, & Canny, 2009; Patel, 2008) have reported that in rural parts of India people have minimum access and familiarity with computers because of illiteracy and spoken language problems, most information systems being in English. The development cost of applications with community partners that meet their local language learning needs, is beyond the budgets of community development projects. In such a scenario the reliability and quality of rural based data entry services may also become questionable in terms of output quality.

2. Rural- Business Process Outsourcing (Rural-BPO) is one of the few avenues of employment for rural India. The data based services involve digitization, data entry, converting document to different format, transcribing, cross checking, collating, amalgamating, merging and many more.

3. Issues with existing data entry user interfaces used at rural-BPOs are noted as follows:

A. Poor design of user interface: Lack of expertise in designing user interface especially for operators in rural-BPOs. The existing user interface has some drawbacks mentioned below:

   a. Failed to correct specific field constraint.

   b. Does not provide clues during typing

   c. Fails to provide confirmation logic

   d. Does not provide validation logic for fields.

B. Double entry is costly and time-consuming.

C. Data entry using a mobile phone were ever used is of poor quality. Viewing windows are small. Screens are not ergonomic compliant in terms of font size, view window, pixels, layout, colour etc.

D. User interface was in the English language:

a) The majority of operators were educated in their mother tongue (local language) and the user interface (UI) used for data entry is completely in the English language.

b) The operators speak in their local language when they are socialising at work. The language used during data entry is English, this means their thinking and conversing language is different therefore working seamlessly between two languages cause them to make errors and also takes extra time during data entry. This is a potential context for errors.

c) Error or feedback message: The operator gets confused when error or feedback message appear in English which take extra time for them to read and understand it and internalize it.

E. Female operators are steadily increasing in semi-urban pockets. Issues regarding their performance and pay do exist.

4. A new user can spend up to 30-50 % of activity time on dealing with errors while interacting with the computer (Lazonder & Meij, 1994). The error occurs during interaction due to some computer software protocols or due to human actions. Researchers (Busse, 1999; Walia & Carver, 2009; Shwartz, et al., 2010; Weyers, Burkolter, Kluge, & Luther, 2010; Madduri, Gupta, De, & Anand, 2010) argued that studies on human errors should place greater emphasis rather than computer / system errors, because human errors are inevitable, even if we can design perfect systems.

5. The literature studies also focused on highlighting the influence of different emotional states on the performance giving rise to human errors. Language too plays a role.

6. Apart from the gaps highlighted above, the research literature also notices the fact that sensitive variables like- perceived system usability, perceived cognitive load, user interface satisfaction, willingness to continue usage and relative advantage, have been ignored while designing the interfaces for data entry, on systems being used by BPO organizations.

7. Therefore, to address above issues like local language, emotions, data entry errors and poor design of user interface for data entry, this research study proposed a new user

interface that is designed with intelligent widgets. The embedded intelligent methods like-machine learning, probabilistic approach and artificial intelligence are proposed to be used in design and development of 'error limiting user interface' for data entry.

8. Further, this new GUI with embedded intelligence to prompt the operator as well as train the operator so as to reduce the errors is intended to be validated by testing.

2.8.Research Questions and Objectives

2.8.1. Research Questions

Following were the research questions addressed in this thesis based on the research gaps identified and discussed above.

RQ1: What is the effect of a newly configured user interface designed with intelligent features like- (i) display of autocomplete suggestion for text field by ranking strategy based on likelihood, (ii) predictive text entry widget, (iii) radio button pointed with most likely options and (iv) dynamic drop-down split-menu, on accuracy of data entry?

RQ2: What is the effect of user interface designed with intelligent features like- (i) display of autocomplete suggestion for text field by ranking strategy based on likelihood, (ii) predictive text entry widget, (iii) radio button pointed with most likely options and (iv) dynamic drop-down split-menu, on speed of data entry?

RQ3: What is the effect of user interface designed with intelligent features on the variables like- (i) perceived system usability, (ii) perceived cognitive load, (iii) user interface satisfaction, (iv) willingness to continue the usage and (v) relative advantage?

RQ4: Are female operator more accurate in data entry as compared to male?

RQ5: What is the effect of language used on error rate in the case of, (i) English language in forms used for data entry, (ii) Indian (*Marathi*) language and (iii) Mixed language (i.e. both English and Indian (*Marathi*) language combined)?

RQ6: What is the effect of emotions on the data entry error rate?

RQ7: In the case of data entry operators what is the influence of level of knowledge about computer in general on the speed and accuracy of data entry?

RQ8: What is the effect of learned expertise in using a particular system- on the speed and accuracy of data entry?

RQ9: What is the influence of level of understanding of the task domain on the speed and accuracy of data entry?

RQ10: Do experienced operator commit more errors compared to less experienced data entry operator in case of being under pressure such as ‘limited time’ data entry?

In this thesis RQ1 to RQ5 have taken up for research.

2.8.2. Objective of the Study

Aim: To improve the usability in terms of work efficiency of rural and semi-urban data entry operators working in rural-BPOs of India.

Objectives:

OB1: To collect computer knowledge and usage patterns of rural and semi- urban users of age 18 to 30.

OB2: To understand their attitudes, difficulties, errors and usability issues towards computers.

OB3: To find out user interface problems, errors while using different data entry software / tools.

OB4: To conceive, propose, model, simulate and test an “Intelligent error limiting user interface” so as to increase the usability and user experience in terms of work efficiency- of data entry operators.

2.9. Conclusion

This Chapter reviewed the literature about data entry errors i.e. numerical entry errors and text entry errors. It also gave us an understanding of types of errors performed by operators in a different context like medical, aviation, rural-BPOs etc. The next part dealt with literature on use of interactive devices in the context of rural Indian user. This section provides difficulties, problem and issues rural user (computer operator in BPOs) have while interacting with computers. Later sections described related work on the role of emotions during data entry.

Based on this literature we have formulated research gaps and accordingly extended our literature survey on proposed intelligent techniques used to improve data entry. The extended literature survey is on machine learning techniques, Bayesian network and a probabilistic model. The identified research gaps lead to a set of ten research questions.

The following Chapter presents methodology adopted for the research and how the research questions were planned to be investigated.

Chapter 3

Exploring the Potential and Influence of Errors during Data Entry

This Chapter includes the three pilot studies conducted for exploring the potential of data entry process and influence of errors on data entry in the context of rural Indians as not many earlier reports exist. Section 3.1 gives details of first pilot study conducted on numerical data entry. The second pilot study is on text data entry reported in Section 3.2. Section 3.3 demonstrates the third and last pilot study conducted on influence of emotional factor on numerical data entry by rural people. The last part of the Chapter mentioned about inferences drawn from the pilot studies.

3.1.Introduction

Several researchers (Chand, 2002; Chand & Dey, 2006; Patel, 2008; Kam, *et al.*, 2009; Gore, Lobo, & Doke, 2012) have investigated the spoken language challenges while designing the user interface for interactive device for rural Indian context. Therefore, our concentration is on consideration of language on usage. So we take general rural population in order to understand in what form language takes to affect the human cognitive scheme i.e. study of language, behaviour and habits. The literature study (Singh, et al., 2009) also indicates that consideration of the data entry errors are crucial for evaluating efficiency of rural users while interacting with computers.

Therefore, to investigate the effect of language on data entry errors and influence of emotion on data entry, the following three pilot studies have been conducted. First pilot study conducted on numerical data entry, second pilot study is on text data entry and third is on influence of emotional factor on numerical data entry by rural people.

Table 3-1: Details of pilot studies conducted

Pilot Study	Hypothesis	Page no.
Pilot Study 1 Numerical Data Entry	**H_1**- The rural users make more errors in English numerical data entry compared to local language (*Marathi* and Assamese) numerical data entry. **H_2**- The rural users require more time in typing English numerical during data entry then if they do it using their local language (*Marathi* and *Assamese*).	43
Pilot Study 2 Text Data Entry	**H_1-** Rural users make more errors in text entry using the English language as compared to local language (*Assamese*). **H_2**- Rural users require more time in text entry in the English language as compared to local language (*Assamese*).	49
Pilot Study 3 Effect of Emotion on Data Entry	**H_1**- Rural users make more errors in local language numerical data entry without time limit during negative state of emotions rather than positive state of emotions. **H_2**- Rural users make more errors in English language numerical data entry without time limit during negative state of emotions as compared to being in positive state of emotions. **H_3**- Rural users require more time in local language numerical data entry during negative emotions as compared to positive emotions. **H_4**- Rural users require more time in English language numerical data entry during negative state of emotions when compared to being in positive state of emotions.	55

3.2.Pilot Study 1: Numerical Data Entry

3.2.1. Research Hypotheses

The hypotheses which were tested in this pilot study are given below:

H_1- The rural users make more errors in English numerical data entry compared to local language (*Marathi* and *Assamese*) numerical data entry.

H_2- The rural users require more time in typing English numerical during data entry then if they do it using their local language (*Marathi* and *Assamese*).

The experiment designed to test the above hypotheses is as below,

3.2.2. Methodology

3.2.2.1. Participants

Forty-eight (48) participants (male / female) belonging to age group 18 to 30 years working in the Indian Institute of Technology Guwahati (IITG) campus that is – people working at shopping complex, vegetable market and security guards as well as and twenty-four (24) rural village based users from the state of Maharashtra India were selected.

Figure 3-1: User's participation in experiment

All participants had a minimum of six months of experience of using computer and had primary education (up to 10th standard) in local (i.e. *Marathi* or *Assamese*) language. Also they use computers or laptops at least one hour in a week. The following Figure 3-1 shows the participants performing given experiment.

3.2.2.2. Instruments

A software (named as *CALCI*) interface was designed specifically for this experiment. It was designed to take number entry input in English and two local languages (*Marathi* and *Assamese*) using an input device such as mouse and keyboard.

Figure 3-2: Screenshot of software interface designed, which does calculations in three languages- English, *Marathi* and *Assamese*

This *CALCI*- interface does arithmetic operations such as addition, subtraction, multiplication and division on numbers with and without decimal point. Figure 3-2 depicts the three screen shots of English, *Marathi* and *Assamese* language calculator respectively.

3.2.2.3. Experiment Design and Variables

The experiment was a 'within subject repeated measures' design. The participant used *CALCI* interface six times to perform six tasks for calculation in English and local languages. The participants perform same calculation in English and local languages using two input devices mouse and keyboard. The number entry language was the independent variable and it had three types: English language interface and two local languages interface. They are *Marathi*- a language spoken in the state of Maharashtra (capital Mumbai) and *Assamese*- language spoken in the state of Assam in the North-East part of India. The dependent variables were the task completion time and errors made in number data entry.

3.2.2.4. Procedure in Detail

The *Marathi* language keyboard was provided to *Marathi* language speaker natives of Maharashtra and *Assamese* to natives of Assam. All participants were tested individually. The *CALCI* interface was used for each participant and they were briefed about the stages and purpose of the experiment before starting. The experiment was divided into six parts (tasks) as given in Table 3.2 below.

Table 3-2: Task Design

Task No.	Language	Input Device	Time Allotted
Task 1(T1)	Local	Mouse	Without time limit
Task 2(T2)	Local	Mouse	Within one minute
Task 3(T3)	Local	Keyboard	Within one minute
Task 4(T4)	English	Mouse	Without time limit
Task 5(T5)	English	Mouse	Within one minute
Task 6(T6)	English	Keyboard	Within one minute

In Table 3-2: Task 1 consists of local language number data entry by using input device mouse without time limit, Task 2 consists of local language number data entry by using mouse within specified limited time (i.e. one minute) and so on. Each participant had to perform all tasks, but the sequence / order of the tasks was different. Table 3-3 shows how sample distribution was done among each task and sequence of tasks to perform.

Table 3-3: Equal distribution of samples among each task

Sample Distribution	Local Language			English Language		
	T1	T2	T3	T4	T5	T6
1-8	1	3	5	2	4	6
9-16	5	1	3	6	2	4
17-24	3	5	1	4	6	2
25-32	2	4	6	1	3	5
33-40	6	2	4	5	1	3
41-48	4	6	2	3	5	1

The first row of Table 3-3 consist of 1-8 samples were performed task1 to task 6 in sequence/order first task1, second task4, third task2, fourth task5, fifth task3 and sixth task6 (i.e. 135246, see Table 3-3's first row) and likewise.

Prior to each stage of the experiment, the participants were given orientation session where they could enter two or three simple calculations and get familiar with operating the interface as it was new to them. When the participants were comfortable with how the interface worked, they were allowed to proceed to the experiment. The participants were required to enter given mathematical calculations having three different difficulty levels (like very easy, easy and hard) using two interfaces (English and *Marathi* or

Assamese) in the defined order in Table 3-3. The participants were provided with the experiment sheet including mathematical calculations in English and the local language they speak. The participants were instructed to perform the mathematical calculation as quickly and as accurately as possible. The computer based background recording of each participant interaction with designed interface (*CALCI*) was enabled to collect data of speed of entry and errors of each participant. This was logged into the data of *CALCI* and retrieved for analysis.

3.2.3. Result and Discussion

3.2.3.1. Types of Errors

Analysis of the total number data entry errors for both interfaces (English and local language) using paired t-test indicates that the mean errors for the local languages number data entry (mean=0.35, sd=0.49) was significantly lower than that of the English language number data entry (mean=0.52, sd=0.69), $t(144)=-3.45$, $p=0.001$. Below, we report the different types of error that occurred in our experiment.

Wiseman, Cairns, & Cox (2011) and Oladimeji, Thimbleby, & Cox (2011) have proposed a classification of number entry errors. They reported the occurrence of certain error types between the two number entry interface styles. Therefore, we have done the classification and showing frequency of each group of number entry errors as depicted in Figure 3-3.

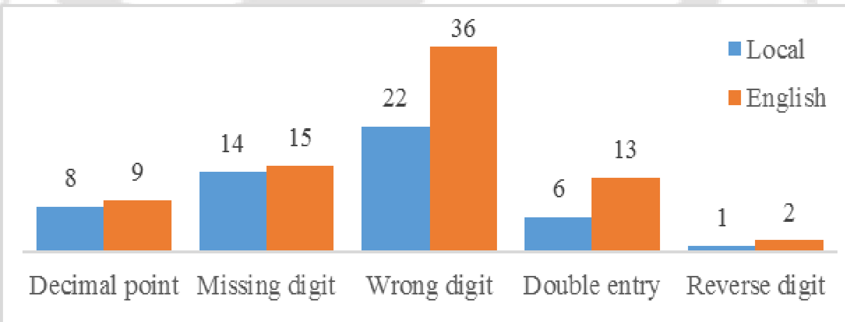


Figure 3-3: The classification of number entry errors in different tasks

Error is unintended action (slip), mistake or omission a user makes while attempting a task. There are common user errors observed in number entry interfaces experiment.

Decimal point: This error occurs when a decimal point is absent or misplaced from the transcribed number but is present or appropriate in the instruction. There are 8 instances of errors in local language and 9 instances of English language.

Missing digit: This refers to occurrences of errors where one digit, one digit before and after decimal from the intended value missing from the transcribed value. For example, a participant entered 6.25 instead of 68.25.

Wrong digit: Wrong digit errors occur when one of the digits in the written value is incorrect. The most cases of the wrong digit error happened whenever the whole calculation wrong part of the operand number is wrong. For example, a participant entered 30.75 instead of 30.25. This type of errors was more frequent on English language number entry.

Double entry: This type of errors occurs when double or repeated entry of numbers is found. The most cases when the participants were given limited time to perform given calculation. For instance, a participant entered 4996.75 instead of 496.75.

Reverse digit: Reverse digit or transposition happen when the user switches two adjacent digits in a number. For instance, instead of entering 1586.50 (in English language 1586.50), a user might enter 1568.50 (in English language 1568.50).

3.2.3.2. Task completion time

It is the time taken by participants to complete a task. Task completion time of Task 1 for English and Task 4 for local language was measured. The task completion time for English language number entry (mean=159.9, sd=80.67) was significantly slower than the task completion time for local language (*Marathi* and *Assamese*) number data entry (mean=152.5, sd=80.13), $t(48)=2.69$, $p=0.01$.

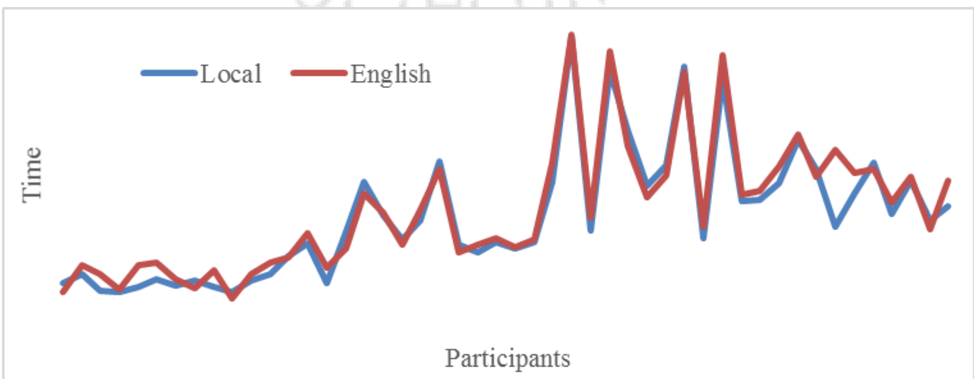


Figure 3-4: Time required for data entry

The graph shown in Figure 3-4 represents the time taken by each participant to complete the given task in both local and English languages. By looking at the graph, it has been observed that, the red lines are dominating blue one. This means that the time taken (represented using red lines in Figure 3-4) for data entry in English language is more compared to local language (represented using blue lines).

3.2.3.3. Other observations and findings

The number of errors per participants within the limited time by input device mouse (mean=0.53, sd=0.65) was significantly more than without time limitation (mean=0.28, sd=0.54), $t(96)=3.14$, $p=0.002$.

Another analysis of the total number data entry errors for both local languages number data entry (*Marathi* and *Assamese*) using independent t-test shows that total number of errors in *Assamese* language number entry (mean=0.46, sd=0.56) were more in comparison with *Marathi* language number entry (mean=0.25, sd=0.44), $t_{(72)}=2.50$, $p=0.01$.

3.2.3.4. Discussion

The results show a considerably higher number of errors on the English number data entry task in comparison to the local language number data entry task. This is because as we have stated above in Section 1.5, about 92.39% schools in rural area teach in the medium of a regional language. It was also observed that rural users were slightly slower (37.5 seconds) during data entry using English. Therefore, during English number data entry their thinking (local language) and typing language (English language) is different. So switching between these two cause them to make more errors and take more time compare to local language where their thinking and typing language is same i.e. local language. Also in rural area of India people generally use their local language numerical for daily routine math/ calculations e.g. for buy and sale. So rural users are more habituated to local language numerical which make then less errors and take less time during local language number data entry. This study shows that there is influence of local language on number data entry.

There is large difference in ‘Wrong digit’ errors group which may due to higher familiarity and understanding of local language numerical by rural users.

For both languages number entry task (English and local), number of errors within limited time entry was significantly more than without the constraint of time limit. This was probably because the participants try to key in the entries faster within the given limited time. This results in more errors in comparison to when there is no time limit prescribed. The results also show that users of *Assamese* language make more errors compared to users in *Marathi*. The possible explanation to this is that *Marathi* language users are more familiar with device operation than *Assamese* subjects.

3.2.4. Conclusion from Pilot Study 1

There are significant differences in the error rates and slight different in speed of entry for the two experimental conditions of number entry. This upholds both hypothesis (H_1 and H_2). The rural users made more errors and required more time in English numbers data entry then data entry using local language (*Marathi* and *Assamese*). The result suggests that for a designer involved in designing interfaces or navigation for predominantly rural users who are more comfortable with the local language, influence of local language needs to be taken into account while determining the information architecture in an application.

3.3. Pilot Study 2: Text Data Entry

Text data entry errors are vital for evaluating the efficiency of rural users while interacting with computers. We are investigating the influence and extent of contribution of local language in triggering these errors.

3.3.1. Research Hypotheses

The hypotheses which were tested in this pilot study are given below:

H_1- Rural users make more errors in text entry using the English language as compared to local language (*Assamese*).

H_2- Rural users require more time in text entry in the English language as compared to local language (*Assamese*).

The usability experiment research design set up to test the above hypotheses is given below.

3.3.2. Research Design

This section gives details about participants, instrument used and detail procedure followed to test the above hypotheses.

3.3.2.1. Participants

Total sample size of forty-four (44) participants (male and female) were approached randomly with a request to participate. The subjects belonged to age group 18 to 30 years. They worked in shops, vegetable market and as security guards in the campuses of Indian Institute of Technology Guwahati (IITG). Figure 3-5 depicts pictures of the participants performing the experiment. Pictures of participants are used with their consent.

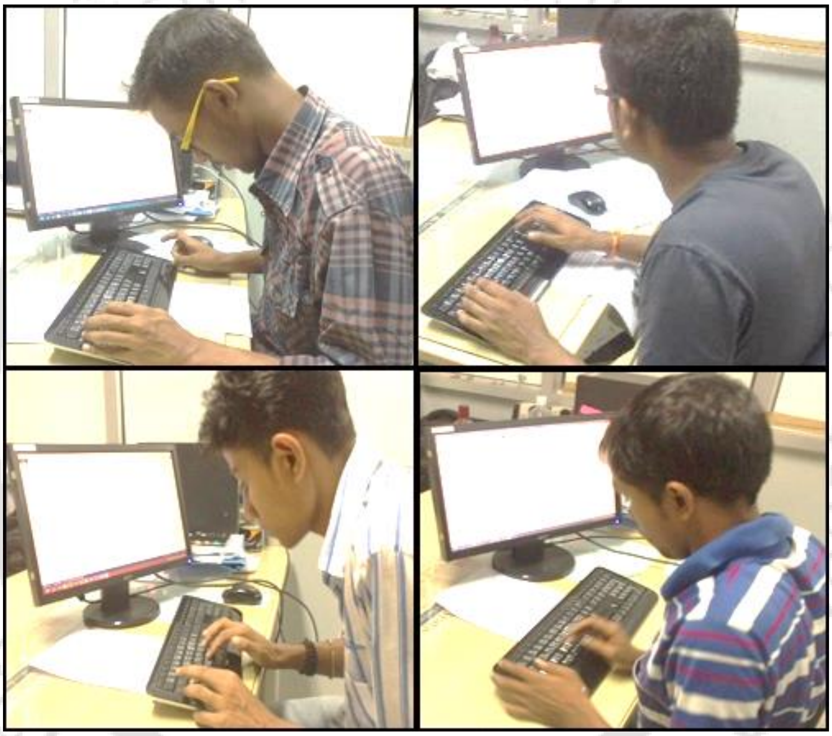


Figure 3-5: Users performing the data entry operation

Demographics

Participants (39 Males + 5 Females) of age group 18 to 30 years had a primary education (up to 10th standard) in local (i.e. *Assamese*) language and used computers or laptops at least one hour in a week as part of their jobs or for personal communication use.

Local Language and English Baseline

All the participants had completed schooling in their mother tongue language (that is *Assamese*). Their proficiency in English was ascertained before the test. This was done by giving a sentence in English and asking subjects to translate it into their local language. Following are the errors and problems observed in the translation of given English sentence into the *Assamese* language by rural users with their percentages expressing their proficiency in English. This proficiency will be used later on to compare with the test results.

- Spelling errors (35%)
- Grammatical errors (55%)
- Difficulty in phrasing complete sentences (20%)
- Difficulty in locating the keys for typing complex words (combining multiple characters together) in the case of *Assamese* language (75%).

Technology Baseline

Among the 44 participants, only 10 of them have access to computer / laptop at their homes on a regular basis. All the participants use computer / laptop at least one hour in a week.

3.3.2.2. Instruments Used

Microsoft Word for English language text entry was adopted and for *Assamese* language text entry- a locally developed Notepad equivalent software named as *BarahaPad* (http://www.baraha.com) was utilized. Figure 3-6 shows the keyboard design which can be used to type in both English as well as the *Assamese* language. The keyboard of *Baraha* system (a local adaptation of keyboard for Indian language) used is depicted in Figure 3-6.

Figure 3-6: Keyboard for *Assamese* and English language text entry

3.3.2.3. Experiment Design and Variables

The text entry language was the independent variable and it had two types: English language interface (that is Microsoft Word) and *Assamese* languages interface (*BarahaPad*). The dependent variables were the task completion time and errors made in text entry.

3.3.2.4. Procedure

All participants were tested individually. Each participant was used *BarahaPad* and Microsoft Word for *Assamese* and English language text entry respectively. The sequence of the tasks was different. The twenty-two participants were performed task1 first and then task2, were as remaining performed task2 first and then task1. The samples distribution was done equally among each sequence of tasks. Table 3-4 depicts the task design for text entry experiment.

Table 3-4: Task Design

Task No.	Task name	Experiment tool
Task 1	Type given *Assamese* language sentence	*BarahaPad*
Task 2	Type given English language sentence	Microsoft Word

To get familiar with the interfaces the participants were given an orientation session where they could enter given sentences in both languages (English & their local language) so that the participants were comfortable with the experimental instrument interface.

The *Assamese* language contains thirty-one phonemes, eight vowel and other twenty-three consonant phonemes (Sarma & Sarma, 2012). The phonetic keyboard was designed by sticking the *Assamese* language phonemes on keys of regular English language keyboard (see Figure 3-5). Writing *Assamese* words using phonetic keyboard is as easy as writing in English. The participants were instructed to perform the tasks (data entry) as quickly and as accurately as possible. The computer based background recording of each participant interaction with interfaces have taken to collection of the speed of entry and errors.

3.3.3. Results and Discussion

The Table 3-5 illustrates the statistical analysis of results obtained from experiments.

Table 3-5: Statistical analysis of results

Hypotheses	*t*- value (paired *t*-test)	mean		sd	
		English	Local	English	Local
H_1	3.56	5.75	4.95	2.57	2.46
H_2	1.90	32.20	32.85	4.79	4.88

3.3.3.1. Types of Errors and Error Rate

Levenshtein minimum string distance statistic (Soukoreff & MacKenzie, 2001; Calculate Levenshtein Distance, 2013) was used for measuring error rates in text entry. Statistical analysis of the total text entry errors for both language entry (English and *Assamese*) using paired t-test indicates that the mean errors for the English language text entry (mean=5.75, sd=2.57) was significantly higher than that of the *Assamese* language text entry (mean=4.95, sd=2.46), $t(88) = 3.56$, $p < 0.002$.

Types of errors observe in text entry by rural users.

1. *Spelling, Incorrect/ Missing Word*: This error occurs when the user forgets characters within words or whole word in transcribed text.

2. *Double character*: In this type user tries to create an unwarranted duplicate character after a target character.

3. *Unrelated*: This error means creating unrelated characters in relation to the presented text.

4. *Related*: This error means deleting characters that are related to the presented text.

5. *Case*: This refers to entering a target character in the wrong case (that is creating a uppercase letter when it is supposed to be in uppercase, or vice versa).

6. *Layer switching*: User needlessly switching between the upper / lower case layer of the keyboard (e.g. by pressing SHIFT / CAP lock).

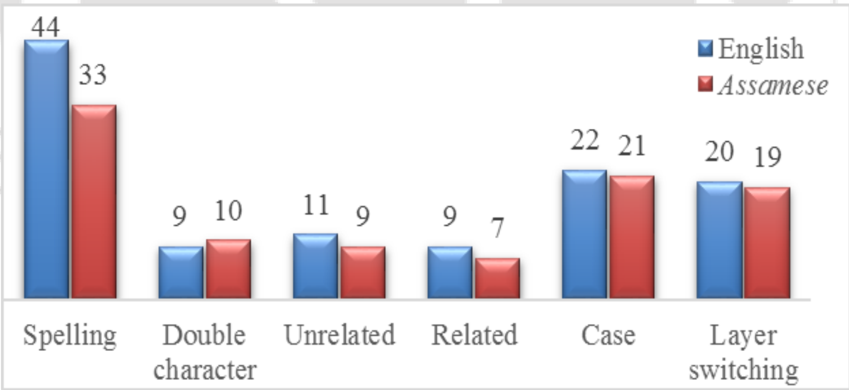


Figure 3-7: Taxonomy of text entry errors

Figure 3-7 depicts the taxonomy of text entry errors with their frequencies observed in English and *Assamese* language by rural users.

3.3.3.2. Task completion time

It is the time taken by participants to complete a task. Task completion time of Task 1 for *Assamese* and Task 2 for English language was measured. The task completion time for *Assamese* language text entry (mean=32.85, sd=4.88) was not significant compared to the task completion time for English local language text entry (mean=32.20, sd=4.79), $t(88) = 1.90$, $p > 0.073$.

3.3.3.3. Discussion

The results confirm a noticeably higher number of errors on the English language text entry task in comparison to the local language (*Assamese*) text entry task. It was also observed that rural users were slower during text entry using *Assamese* language as compared to while using English language. It was noticed that during typing rural user reads the presented text and then types it using the given interface. While reading the presented *Assamese* word from sentence, he attempts to memories 2-3 words and then type it on the screen. But in case of English language words (especially long and difficult words) the rural user is unable to remember the full spelling while transcribing and makes more errors as more entry strokes are involved accompanied by eye movement.

The structure of Indian languages is different from English containing simple, complex and *matra* characters. As explained on page number 19, the complex character is made by combining multiple characters together and error in one single character may be required multiple edit operation to fix it. In such case fixing of a single error requires additional edit operation to non-erroneous characters. So for Indian language (*Assamese*) the time required for number of edit primitives to transform transcribed text from presented text is more as compared to English.

The results of statistical analysis indicate that significant differences in the error rates exists which upholds hypotheses (H_1) (Table 3-5). However, we found no significant difference in speed of entry for the two experimental conditions of text entry (hypotheses (H_2)) thereby resulting in failure to reject the null hypotheses.

3.3.4. Conclusion from Pilot Study 2

There is significant difference in the error rates for the two experimental conditions of text entry. The rural users made more errors in English text entry then local language (*Assamese*). The result suggests that for a designer involved in designing interfaces or navigation for predominantly rural users more comfortable with local language, influence of local language needs to be taken into account while determining the information architecture in an application.

3.4. Pilot Study 3: Effect of Emotion on Data Entry

This section focuses on the role of emotions and their influence on errors in computer data entry work. Emotions are important and most pervasive aspect of human behavior including during work. This pilot study explored the role of 'internal performance shaping factor' like 'emotions' which may affect the work performance during numerical data entry work.

3.4.1. Hypotheses

In this section we are attempting to find the extent of influence of emotions on making these errors in local and English language. The hypotheses are stated as below:

H_1- Rural users make more errors in local language numerical data entry without time limit during negative state of emotions rather than positive state of emotions.

H_2- Rural users make more errors in English language numerical data entry without time limit during negative state of emotions as compared to being in positive state of emotions.

H_3- Rural users require more time in local language numerical data entry during negative emotions as compared to positive emotions.

H_4- Rural users require more time in English language numerical data entry during negative state of emotions when compared to being in positive state of emotions.

The experiment designed to test the above hypotheses is given below.

3.4.2. Methods

3.4.2.1. Participants

Forty-Eight (48) participants (male / female) belonging to the age group of 16 to 30 years were selected for the experiment. They were students of 11th and 12th standard, people working in the coffee shop, stationary shop, grocery shop, vegetable market and security guards in the campuses of Indian Institute of Technology Guwahati (IITG). All participants from rural background had educational qualification of 10th to 12th standard (that is non-graduate) and used computers or laptops at least one hour in a week. The following Figure 3.8 shows the participants performing given experiment.

Figure 3-8: (a) picture depicts, the process of experiment being explained to the participant and (b) and (c) pictures showing participants performing the numerical entry operation task assigned to them.

3.4.2.2. Instrument and Materials

A software interface involving a calculator was designed specifically for this experiment. It was designed to input numerical data in *Assamese* and English language using keyboard and mouse. This interface can perform arithmetic operations like addition, subtraction, multiplication and division - with and without a decimal point. Figure 3-9 depicts the screenshot of the software interface in *Assamese* and English language.

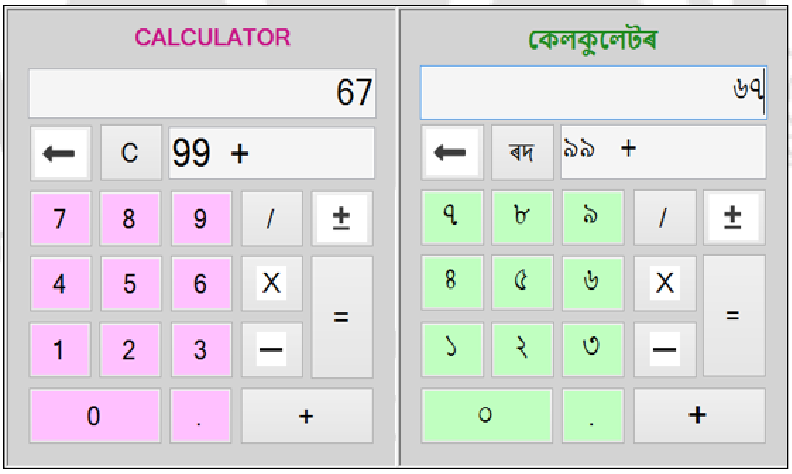


Figure 3-9: Screenshot of software interface, which does calculations in *Assamese* and English language

3.4.2.3. Stimuli

Videos can be used to induce emotions artificially in the participants (Neerincx & Streefkerk, 2003; Spering, Wagener, & Funke, 2005). We used, three video clips to influence affective states and for inducing positive and negative types of emotions in both the valence and arousal dimensions. We used three types of video clips like violence video for negative emotion and comedy video for positive emotions each of five minutes duration.

The video for control group was an abstract images video for duration of two minutes. To assess whether right emotions have been induced, an emotion measuring Self-Assessment Manikin (SAM) scale, proposed by Bradley & Lang, 1994; was used. Figure 3-10 depicts the SAM scale having a valence (top) and arousal (bottom). The valence was used as a typical dimension for checking whether emotional state is positive or negative. At one extreme of the scale one can felt happy, pleased, satisfied, contented and hopeful. The other end it reflects unhappy, annoyed, unsatisfied, melancholic, depressed and/or bored. Similarly, the degree of arousal reflects from an excited wide-eyed figure to a relaxed sleepy figure. At one end of this scale one physically felt relaxed, calm, sluggish, dull, sleepy and unaroused. The other end it reflects stimulated, excited, frenzied, Jittery, wide awake and aroused. So a highly aroused negative state corresponds to anger whereas a low aroused positive state would be contentment (Bradley & Lang, 1994).

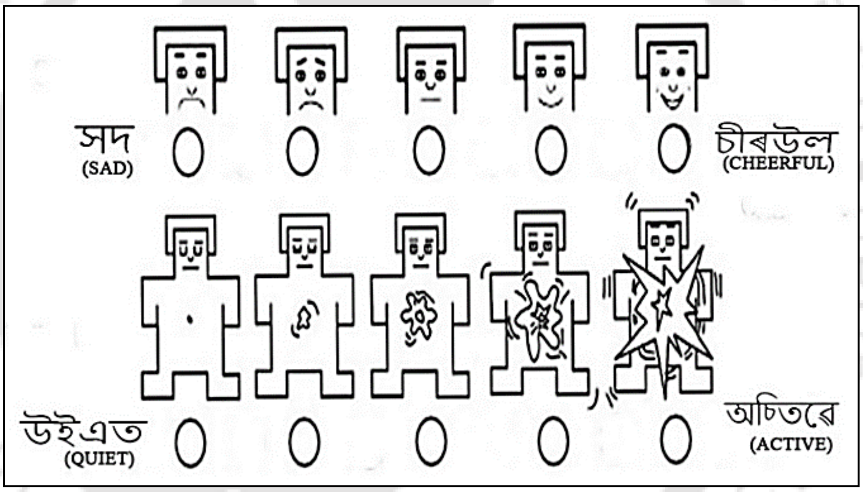


Figure 3-10: The valence (top) and arousal (bottom) scales of Self-Assessment Manikin (SAM) (Bradley & Lang, 1994)

3.4.3. Research Design

3.4.3.1. Experimental Variables

The experiment was a between subject design. The participants used software interface four times to perform four tasks for calculation in *Assamese* and English languages. The participants perform calculation in *Assamese* and English languages using input device-keyboard only. The emotion (positive, negative and neutral/control group), number entry language (*Assamese* and English) were the independent variables. The dependent variables were the task completion time and errors made in numerical data entry.

3.4.3.2. Experimental Design

The experiment was divided into four tasks as given in Table 3-6 below, Task 1 consists of local language numerical data entry without time limit, Task 2 consists of English language numerical data entry without time limit and so on. Each participant has to perform all tasks, but the sequence/order of the tasks may be different. Table 3-7 shows how samples distribution was done among each task and sequence of tasks to perform. As shown in table 2, the samples were equally distributed among three emotional states like positive, negative and control group. The first row of Table 3-7- positive emotion induced for sample 1,2, control/neutral emotion induced for sample 3,4, negative emotion induced for sample 5,6 and all samples 1 to 6 perform task 1 (T1) to task 4 (T4) in sequence T1, T2, T3 and T4.

Table 3-6: Task Design

Task No.	Time Allotted	Language
Task 1 (T1)	Without time limit	Local
Task 2 (T2)	Without time limit	English
Task 3 (T3)	Within one minute	Local
Task 4 (T4)	Within one minute	English

Table 3-7: Distribution of Samples among Emotion Affective State and Task

Sample Distribution			Tasks			
Positive	**Neutral/control**	**Negative**				
1,2	3,4	5,6	T1	T2	T3	T4
7,8	9,10	11,12	T1	T2	T4	T3
13,14	15,16	17,18	T2	T1	T3	T4
19,20	21,22	23,24	T2	T1	T4	T3
25,26	27,28	29,30	T3	T4	T1	T2
31,32	33,34	35,36	T3	T4	T2	T1
37,38	39,40	41,42	T4	T3	T1	T2
43,44	45,46	47,48	T4	T3	T2	T1

3.4.4. Procedure

All participants were tested individually. They were briefed about the stages and purpose of the experiment before starting. They were also instructed on the use of SAM scale and the software interface for numerical data entry.

To study empirically, the experiment was divided into two parts. Part one consists of emotional inducement process. When the participants were ready, their emotion was checked by SAM scale before starting the actual experiment. Then the participants were shown the five minutes video clip for their particular experimental condition. To assess whether right emotions have been induced, an emotional measured through SAM scale of participants were taken. In second part, participants moved on to the numerical entry task. Prior participants were given orientation session where they could enter two or three simple calculations and get familiar with the interface. When the participants were comfortable with how the interface worked, they were allowed to proceed to the experiment. The participants were required to enter given mathematical calculations having three different difficulty levels (like very easy, easy and hard) using two interfaces (*Assamese* and English) in the defined order in Table 3-6. The participants were provided the experimental sheet including mathematical calculations in English and local language they speak (i.e. *Assamese*). The participants were instructed to perform the mathematical calculation as quickly and as accurately as possible. The computer based background recording of each participant interaction with designed software interface have taken as a data collection for entry typing speed and errors.

3.4.5. Results and Discussion

In this experiment we have observed three types of errors like interface error, interaction error and data entry error. Only the data entry errors (that is numerical data entry) are considered for this study. The forty-eight participants entered around 6528 numbers in total and made 877 errors, or approximately 13.43% (overall error rate). Thus, participants made a mean number of 4.66 errors and SD= 5.49.

3.4.5.1. Emotion Manipulation

As expected with the experiment, a paired t-test showed significant difference in SAM scale ratings for before (pre) and after (post) in case of positive (t (32) = 7.97, p = 0.001) and negative (t (32) = 5.59, p = 0.001). Also we found the significant difference in between positive and negative emotions for valence (t (16) = 18.75, p = 0.001) and arousal (t (16) = 3.46, p = 0.002).

Table 3-8: Results manipulation by statistical analysis

Hypothesis	t- value (independent t-test)	Mean		SD	
		Positive	Negative	Positive	Negative
H_1	5.90	0.47	8.20	0.74	4.75
H_2	7.76	2.27	15	3.37	4.74
H_3	6.12	118.7	176.33	32.01	35.81
H_4	6.30	124.80	191.80	33.26	33.28

Table 3-8 illustrates the statistical analysis done by independent t-test to find the significance. As shown in Table 3, all hypothesis from H_1 to H_4 with their mean, standard deviation (SD) values for both positive and negative affective states and t-values. According to results from Table 3-8, the hypotheses H_1 and H_2 are significant, that is rural users make more errors in numerical data entry by both *Assamese* and English language without time limit during negative emotion than in positive emotion. Also, rural users require more time in numeric data entry by both *Assamese* and English language without time limit during negative emotion than positive emotion which proves hypotheses H_3 and H_4.

3.4.5.2. Discussion

The participants were able to appropriately attribute the expected valence by watching video clips measured by SAM scale. This suggests that the video clips were influencing the affective state of the participant and the experimental manipulation had worked. As a consequence, there was a significant effect of the affective state of the participants on the number of errors made by them. The number of errors made by participants including within and without time limit including two different languages (*Assamese* and English) are quite high. The use of multiple languages and time pressure component made this is somewhat challenging task to the participants.

The experiment had limitations. There are several issues of environmental validity which were compromised. Participants were required to enter several calculations including addition, subtraction, multiplication and division. This is not the normal job of number entry usually performed in the rural-BPOs and NGOs in India. The calculator interface is also not the only style of visual interface seen in rural Indian workplace data entry screens. This experiment was conducted as a pilot study level overlooking the limitations.

3.4.6. Conclusion from Pilot Study 3

Several studies (Jeon, Yim, & Walker, 2011; Causse *et al*., 2013; Cairns, Pandab, & Power, 2014) from the literature have proved the influence of emotion on the tasks which cause them to make errors. The results also show a significant difference in error rate and speed of entry for two emotional states - both in the *Assamese* language as well as the English language. The rural (middle and lower-middle-income category) users in negative state of emotion made more errors and required more time as compared to positive emotional state in both *Assamese* and English language numerical data entry. The study is helpful from the designer's aspects where it will guide them to consider emotional aspects in designing user interfaces to mitigate number data entry error. Could this mean that language alone may not be a factor? This needs another set of experiments to be planned in the future. Some language had more complex set of variables, further work in this thesis from language specific influence was curtailed.

3.5.Conclusion

Many user interfaces have developed and proposed by researchers (Grisedale, Graves, & Grünsteidl, 1997; Parikh, Ghosh, & Chavan, 2002; Chand & Day 2006; Gore *et al*. 2012) for rural Indian users in their local language. Therefore, the literature suggests that there is influence of local language on rural Indian users. The pilot studies reported in this thesis also concludes the influence of local language on data entry.

From the pilot studies one concludes that, first, one need to provide alternative for local language numerical audio support along with English designing numerical entry fields; second, design feedback and error messages for use of local language during text entry must be considered so as to understand and learn from mistakes / slips; and third, incorporate emotional aspects by using audio feedback in local language and using visual metaphors involving culture and habits. These factors may be considered in designing user interfaces for rural Indian especially doing data entry.

The pilot studies were conducted to study if there is effect of language, behaviour (emotion) and habits on data entry in the context of rural Indian. We noticed that most of the data entry on computers happens in the English language. Therefore, we did not continue with the local language influence line of investigation which also had logistical limitations. For this research, as India is a widely distributed geographical entity with each language separated by up to 3000 km. So we ended our investigations on emotional factors due to language variable.

3.6.Consolidation of all Pilot Studies

	Pilot Study - 1	**Pilot Study - 2**	**Pilot Study - 3**
Hypothesis	**H_1**- The rural users make more errors in English numerical data entry compared to local language (*Marathi* and *Assamese*) numerical data entry. **H_2**- The rural users require more time in typing English numerical during data entry then if they do it using their local language (*Marathi* and *Assamese*).	**H_1-** Rural users make more errors in text entry using the English language as compared to local language (*Assamese*). **H_2**- Rural users require more time in text entry in the English language as compared to local language (*Assamese*).	**H_1**- Rural users make more errors in local language numerical data entry without time limit during negative state of emotions rather than positive state of emotions. **H_2**- Rural users make more errors in English language numerical data entry without time limit during negative state of emotions as compared to being in positive state of emotions. **H_3**- Rural users require more time in local language numerical data entry during negative emotions as compared to positive emotions. **H_4**- Rural users require more time in English language numerical data entry during negative state of emotions when compared to being in positive state of emotions.
Subjects	48 Participants (40M+8F)	44 Participants (39M + 5F)	48 Participants (43M+5F)
	• age group 18 to 30 years • participant had a primary education (up to 10th standard) in local (i.e. *Assamese* or *Marathi*) language • participants from rural and semi-urban Indian computer users • uses computers or laptops at least one hour in a week		
Instrument	Designed Software interface- named as *CALCI*	Microsoft word- English language text entry *BarahaPad* – *Assamese* language text entry	• Calculator User Interface • Violence video for negative emotions • Comedy video for positive emotions • Self-Assessment Manikin

Experimental variables

Independent Variable
- English language interface
- *Marathi* language interface
- *Assamese* languages interface

Dependent Variables
- task completion time
- errors

Independent Variable
- English language interface (Microsoft Word)
- *Assamese* languages interface (*BarahaPad*)

Dependent Variables
- task completion time
- errors

Independent Variable
- positive emotion
- negative emotion
- Neutral/ control emotion
- English language numerical entry
- *Assamese* lang. numerical entry

Dependent Variables
-task completion time and errors

Tasks

Task No.	Language	Input Device	Time Allotted
Task 1(T1)	Local	Mouse	Without time limit
Task 2(T2)	Local	Mouse	Within one minute
Task 3(T3)	Local	Keyboard	Within one minute
Task 4(T4)	English	Mouse	Without time limit
Task 5(T5)	English	Mouse	Within one minute
Task 6(T6)	English	Keyboard	Within one minute

Task No.	Task name	Experiment tool
Task 1	Type given Assamese language sentence	BarahaPad
Task 2	Type given English language sentence	Microsoft Word

Task No.	Time Allotted	Language
Task 1 (T1)	Without time limit	Local
Task 2 (T2)	Without time limit	English
Task 3 (T3)	Within one minute	Local
Task 4 (T4)	Within one minute	English

Results

Hypothesis	t- value (paired t-test)	mean		sd	
		English	Local	English	Local
Error (H_1)	-3.45	0.52	0.35	0.69	0.49
Time(H_2)	-2.69	159.9	152.5	80.67	80.13

Hypothesis	t- value (paired t-test)	mean		sd	
		English	Local	English	Local
Error (H_3)	3.56	5.75	4.95	2.57	2.46
Time(H_4)	-1.90	32.20	32.85	4.79	4.88

Hypo.	t- value (independent t-test)	Mean		SD	
		+ve	-ve	+ve	-ve
H_1	-5.90	0.47	8.20	0.74	4.75
H_2	-7.76	2.27	15	3.37	4.74
H_3	-6.12	118.7	176.33	32.01	35.81
H_4	-6.30	124.80	191.80	33.26	33.28

Conclusion

Provide alternative of local language numerical audio support along with English designing numerical entry fields

Design feedback & error messages by use of local language during text entry so as to understand & learn from mistakes/ slips

Incorporate emotional aspects by using audio feedback in local language and using visual metaphors involving culture & habits

Chapter 4

Error Limiting Intelligent Interface for Date Entry (*ELIIDE*)- a Tool

This Chapter reports the different phases of development of a tool- named as 'error limiting intelligent interface for data entry' that was used in the previous (pronounced as *e-lide*). Section 4.2 reports the block diagram of *ELIIDE*. The details of development phases include formation of Bayesian network, formation of probabilistic dependencies, calculation of Conditional Probability Table (CPT), applying chain rule- formation of production rule based on CPT and marginalization. The later part illustrates the screen shots of *ELIIDE* - tool.

This Chapter presented the development phases and screenshots of the newly developed interface for data entry, named as *ELIIDE* - tool. The *ELIIDE* has different features like dynamic, predictive, adaptive and probabilistic. The interface uses local *Marathi* language to communicate with user / operator. The communication happens in terms of error and feedback messages. This additional feature may support rural users to get emotionally attached to interface. The generation of different reports like user log record and user performance report may also help the manager to track the performance of particular user.

4.1.Introduction

After conclusion drawn from pilot studies, experimental based data collection, ethnographic study and detailed state of the art literature survey, we have proposed a tool named as 'Error Limiting Intelligent Interface for Data Entry (*ELIIDE*)'. The designed and developed procedure of the tool is mentioned in the Chapter. Figure 4-1 depicts the block diagram showing working of the tool named *ELIIDE* (pronounced as- 'e-*lide*')- tool.

As we have reported earlier in Chapter 1, Section 1.7, there are several issues with existing user interface used for data entry like (a) fails to correct specific field constraint (b) does not provide clues during typing (c) fails to provide confirmation logic and (d) does not provide validation logic for fields. The *ELIIDE* has provided solutions to these issues and gives additional features through intelligent mechanisms reported below.

Existing adaptive interfaces (Chen, Hellerstein, & Parikh., 2010; Kleinman, 2001; Mitchell & Shneiderman, 1989) have been in existence in literature since a decade. However, not many were found in literature that could address specific aspects of the rural-BPO scene in India which is to be experimented with. Hence it was decided to design the tool itself both, as a metric of errors for rural Indian BPO scenario and also act as validator during testing.

Following are features of the Tool:

Validation logic for fields: The *ELIIDE* does not allow user/ operator to enter special characters & numerical in text fields and special characters & texts in numerical fields. If user does so, *ELIIDE* consider them as errors and add it into errors list. The *ELIIDE* gives error message only when user left particular field with wrong / incorrect / blank entry (blank entry is not allowed for compulsory fields which are marked with '*' sign). For example, (1) special character and text entry is not allowed in 'mobile number field' and (2) if user mistakenly enters 'nine' digits which is not allowed, the *ELIIDE* gives error messages for both wrong entries.

Predictive mechanism: The *ELIIDE* provides predictive mechanism to text entries to predict future entries based on entries available in database.

Quantitative probabilistic approach*:* The *ELIIDE* is supported with the specially designed widgets having quantitative probability. For example, in Figure 4-12, "Date of Birth" field was provided with the 'bar chart' indicating quantitative probability of existing entries for different age groups like 'below 18 years', '18 to 60 years' and 'above 60 years' and another data entry widget or radio button named as "Gender / Sex" was supported with the numeric probability for particular gender in percentage. This type of practice can help operator to cross validate and hence speed up their performance while data entry.

Dynamic widgets: The *ELIIDE* implements the design of dynamic drop-down menu for data entry. This type of menu design is also known as elective split-menu design (Chen K., Chen, Conway, Hellerstein, & Parikh, 2011 and Warren & Bolton, 1999). In the first part of split-menu five most frequently used items (Miller, 1956) were displayed with quantitative probabilities and in second part the remaining items were shown in alphabetical order to the operator.

***Adaptive**:* The *ELIIDE* provides an adaptive audio support feature, which is optional for expert users. This feature was adopted by *ELIIDE* according to user's expertise which is calculated using users' performance index. The user/operator having performance index (PI) less than 5%, we called them as expert users. The performance index is calculated as,

$$\textbf{Performance Index (PI)} = \frac{\textbf{Error Made}}{\textbf{Error Opportunities}} * \textbf{100}$$

Here,

Error Made = number of errors made by the operator during data entry,

Error Opportunities= total number of possible errors an operator can make in one data entry form (Figure 5-2), here it is 40.

For example:

If operator made 5 errors in one data entry form, then PI will be

PI= 5 / 40 = 0.12 i.e. 12 % which is greater than 5%.

But if user made 2 errors in one data entry form, then

PI=2 / 40 = 0.048 i.e. 4.87 % which is less than 5%, we called him as expert user.

4.2. Block Diagram of *ELIIDE* - tool

The literature (Chen, Hellerstein, & Parikh (2010) and Hermens & Schlimmer (1994)), helped to derived key insights for the development of *ELIIDE* –tool. The Figure 4-1 depicts the block diagram showing working of *ELIIDE* tool. The transcription block (shown in Figure 4-1 as a thick-lined box) processes the input provided by operator on an instantiation of as electronic form. When user clicks the save button, each field value on the entire form is updated and forwarded to learning module and also saved in the database. These field values are processed by the learning mechanism and together with learned functions. Later these values are used by predictor module to suggest values for other form fields.

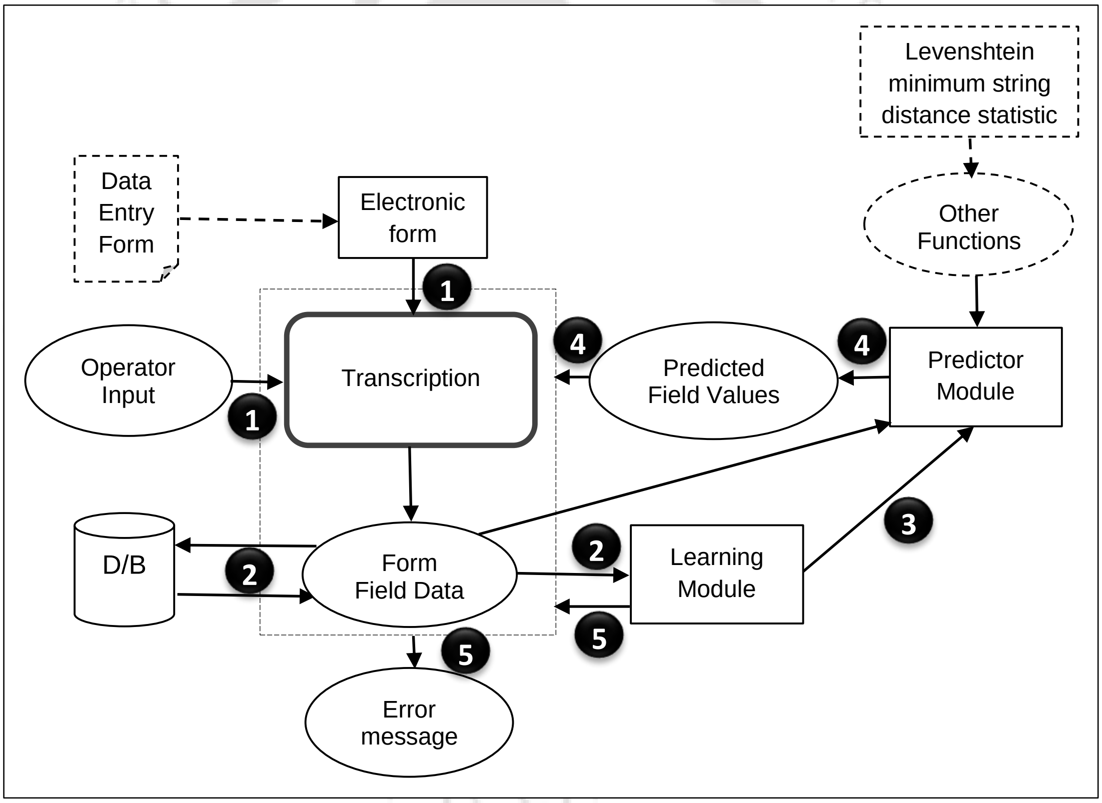


Figure 4-1: Block diagram showing components and working of *ELIIDE* - tool 1. the input provided by operator on an instantiation of as electronic form 2. When user clicks the save button, each field value on the entire form is updated, forwarded to learning module and also saved in the database 3. these values are used by predictor module 4. suggest values for other form fields 5. Feedback / error message

Suppose, a user begins working on a form and types values into a field. The system may use values from fields on the previous form or structure learned from those examples to find a relationship between fields. If the related field values do not match, it gives error message stored in the database. The error messages are in *Marathi* languages with audio support, so as to better understood without ambiguity and error.

4.3. Development Process of *ELIIDE* Tool

The development process of *ELIIDE* tool involves various stages given below (Black & Ertel, 2011; Russell & Norvig, 2003).

1. Formation of Bayesian network: Finding the probabilistic relation between form fields.

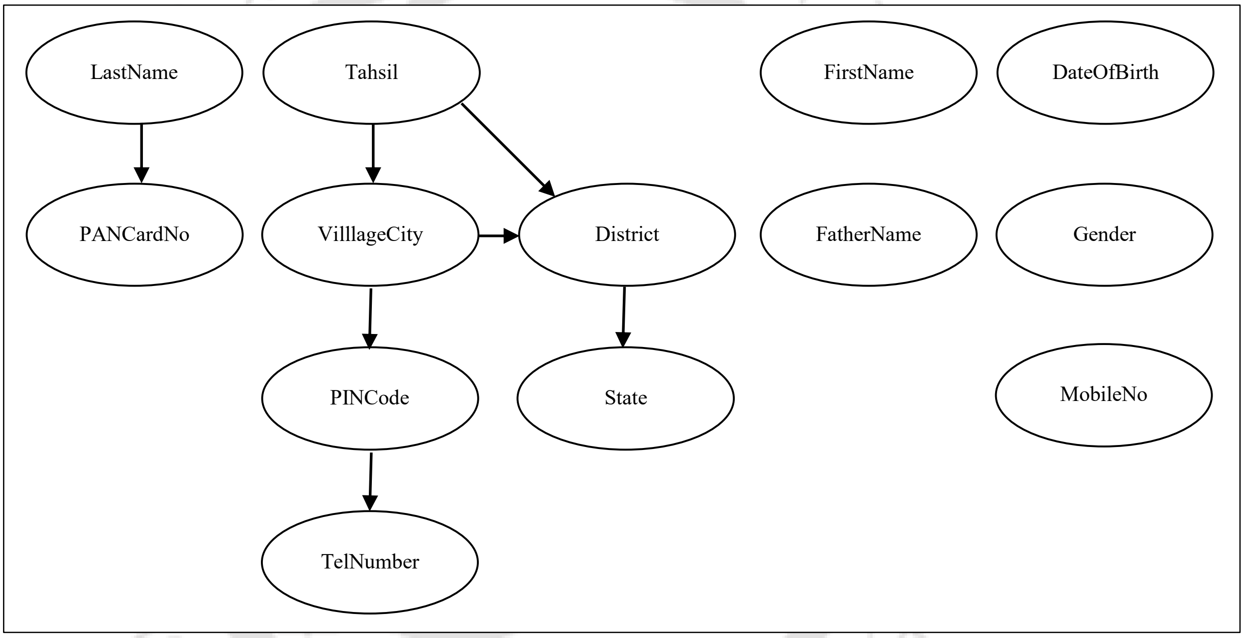


Figure 4-2: Bayesian network showing probabilistic relationship between form fields
Source: Author generated

The Figure 4-2 represents the Bayesian network drown from Finding the probabilistic relation between form fields. The variables shown by oval shapes represent the thirteen fields from the *ELIIDE's* data entry form. For example: 'Last Name' is represented using variable named as 'LastName', same for 'PAN card number' is represented using variable 'PANCardNo', and so on. Here, 'Tahsil' means the area of each sub-division / sub-district in a State. (http://www.censusindia.gov.in, 2016).

While forming the Bayesian network, we have identified the probabilistic relationship between form fields, for example, the fields 'LastName' and 'PANCardNo' are related with each other. The PAN means personal account number in India. The PAN structure is as follows: AAAPL1234C: First five characters are letters, next four numerals, last character letter (https://www.tin-nsdl.com/, 2016). The fifth character in PAN is the first letter of individual's last name. Therefore, if these two fields does not match with each other then it is identified as error. The same way, the Bayesian network of six fields like Tehsil, Village/ city, District, PIN code, State and Telephone number is formed.

2. Formation of probabilistic dependencies
3. Calculation of Conditional Probability Table (CPT)

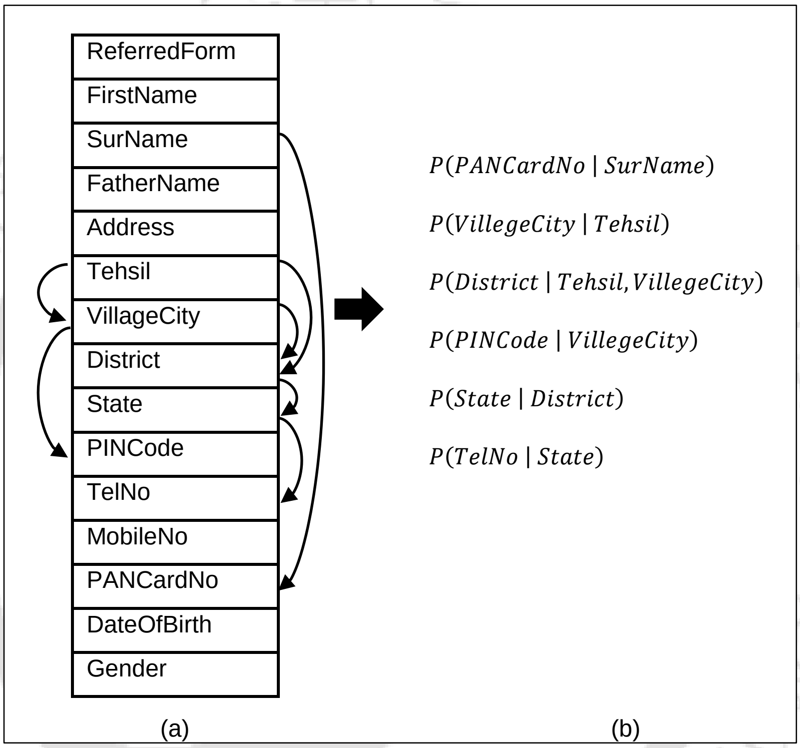


Figure 4-3: (a) Form field ordering layout, arrow showing probabilistic dependencies (b) Conditional Probability Table (CPT)

4. Applying chain rule- formation of production rule based on CPT
5. Marginalization/ Marginal distribution

The *ELIIDE* - tool was implemented using C#.NET, a Microsoft .NET framework toolkit. The SQL (Structured Query Language) language was used to manipulate database activities in the backend.

4.4.Screenshots of *ELIIDE* - Tool

4.4.1. Login Screen

Figure 4-4 shows screenshot of login screen for *ELIIDE* - tool. As *ELIIDE* has been used as a tool for conducting this experiment. So it is provided with two different form designs, one is according to existing UI and another is intelligent UI designed with intelligent widgets. Therefore, the screen-shot below showing one option button named as- 'Existing User Interface'. If we select the option button, then it will enter into the form designed according to existing UI and by default it will go to intelligent UI design.

The another option button named as- 'Time Bound' with 180 seconds. This option if selected can be used for experiment of limited time activity.

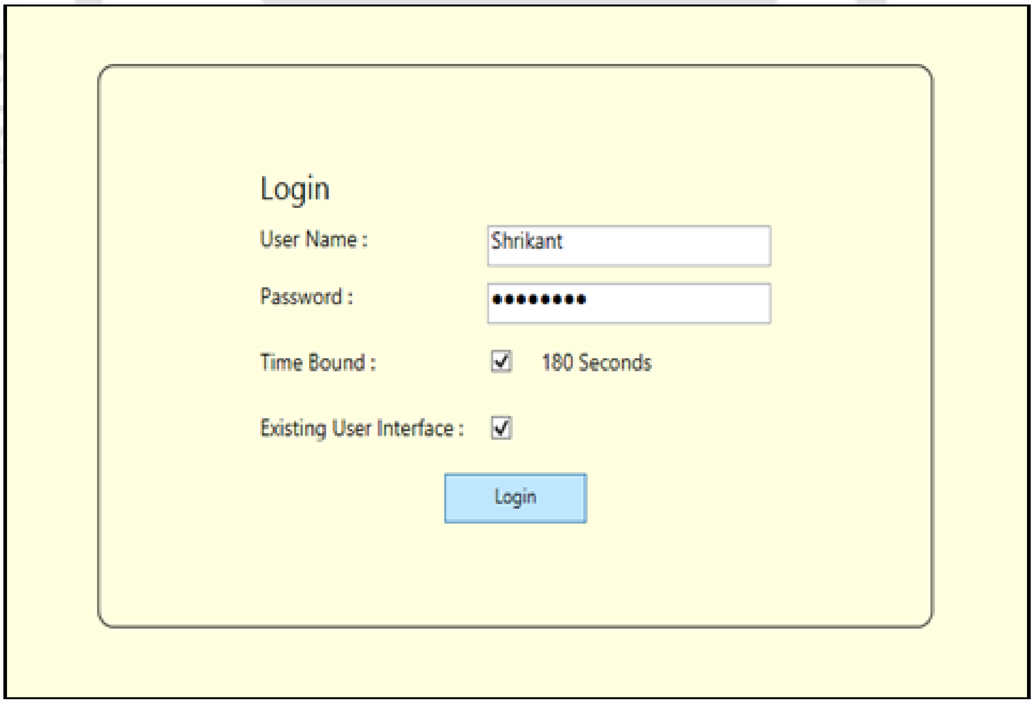


Figure 4-4: Screenshot of login screen of *ELIIDE* - tool

4.4.2. Data Entry Form

Figure 4-5 depicts the screenshot of *ELIIDE's* 'data entry form' that appears after user logged-in in intelligent UI. The form shows the dynamic widgets for particular fields like- State, DOB and Gender. The 'State'-field is implemented with dynamic drop-down menu design (Section 4.4.6), the 'Date of Birth'- field is supported with quantitative probability and bar graph (Section 4.4.8) and the 'Gender'-field is provided with numeric probability using percentage and bar graph (Section 4.4.8).

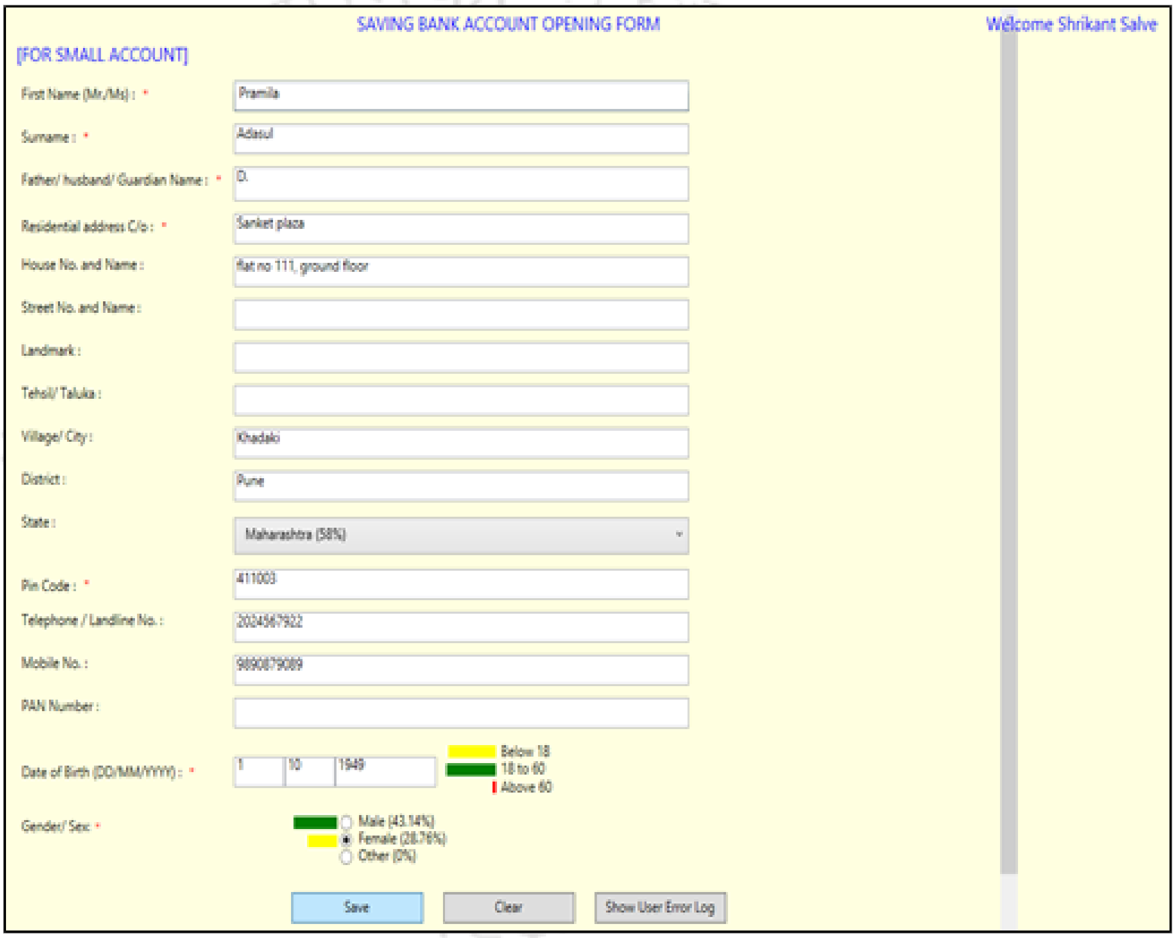


Figure 4-5: Screenshot of data entry form with dynamic widgets of *ELIIDE* - tool

In this 'data entry form' of *ELIIDE* tool, the specific fields are interlined with each other by forming probabilistic relationship between them using Bayesian network as shown in Figure 4-2. As soon as the operators enters wrong input in particular field *ELIIDE* gives error message to him/ her.

4.4.3. Error Messages

As our literature study (Grisedale, Graves, & Grünsteidl, 1997; Parikh, Ghosh, & Chavan, 2002; Chand & Day 2006; Gore *et al*. 2012) suggest that there may by influence of local language on data entry by operator who primarily educated in their local language (mother tongue). But, the data entry happens in English language. Therefore, we have designed the errors and feedback messages in their local language (i.e. *Marathi*). These errors and feedback messages are also supported with audio so as to make the operator feel that, *ELIIDE* is communicating with them and may get emotionally attached with it. The *ELIIDE* - tool displays the error messages in the *Marathi* language when the user enters an invalid or no value for particular field.

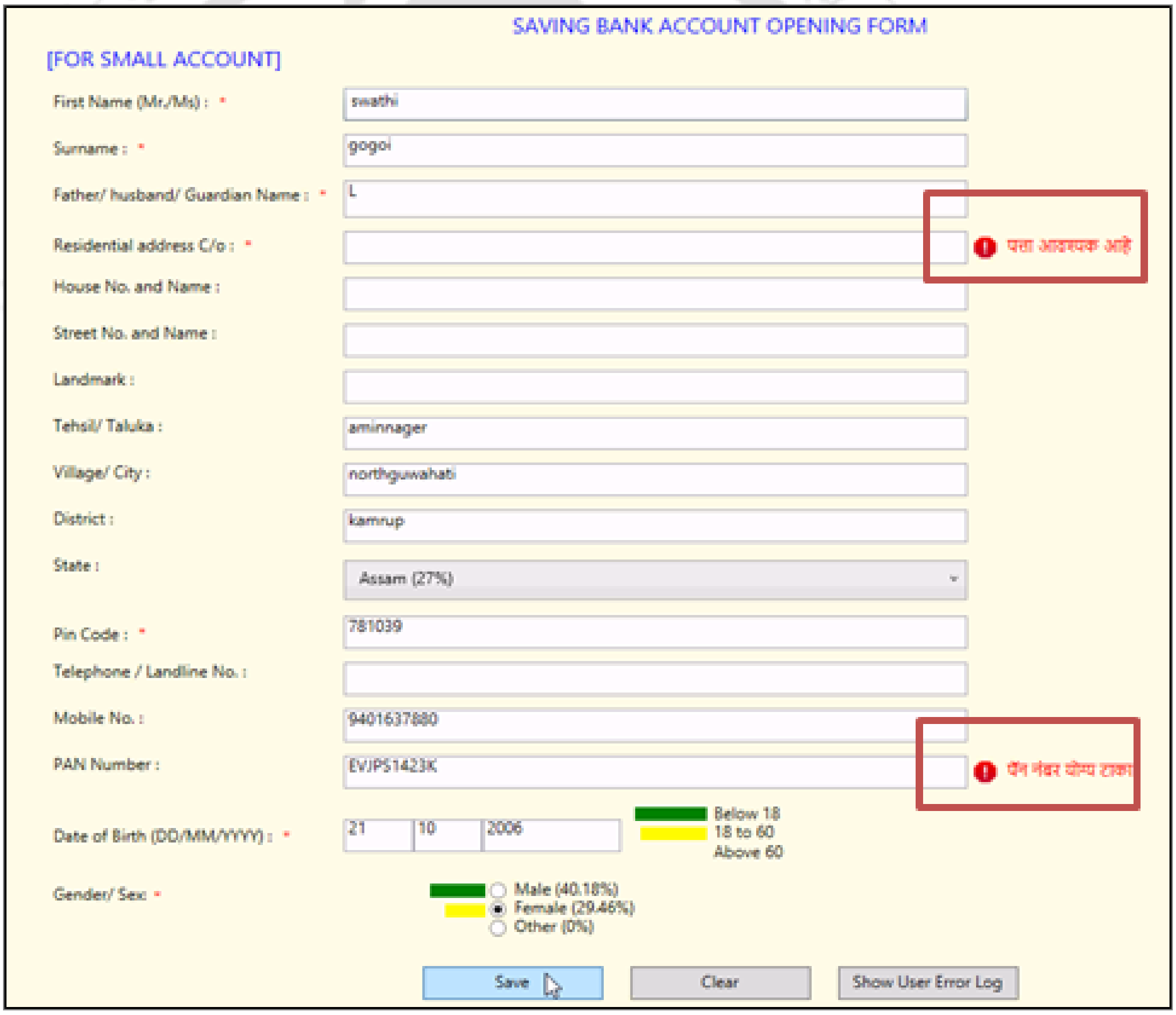
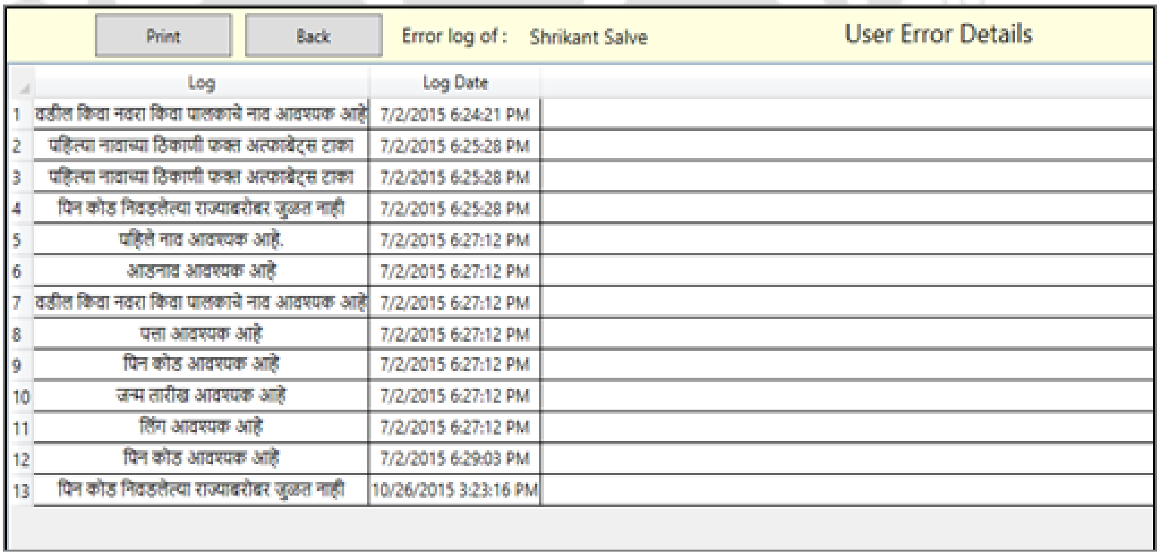


Figure 4-6: Screenshot of data entry form, red square rectangles showing error messages displayed in *Marathi* language by *ELIIDE* tool

In Figure we can see that for 'field-4' i.e. 'Residential Address' field, if the operator does not enter anything in this compulsory field (marked with '*' sign), then *ELIIDE* gives error message in *Marathi* language, English translation is “Address is compulsory”. Another field of ‘PAN Number’ where operator entered wrong PAN, so the error message shown is “Please enter correct PAN number” (English translation). Figure 4-6 depicts the screenshot of error messages highlighted with thick red rectangle.

4.4.4. Error Report Generation

The *ELIIDE* - tool, generate error report showing error log details like- type of error and when the error happened of individual operator/ user. This information is important for individual to review their faults and improve their performance. So that during data entry they will be more attentive to the most erroneous fields to mitigate errors. The error log is also important for the manager to evaluate the performance of individual operator. The most erroneous field can be identified and studied for further improvement. The error log also displays timing of error. The most ‘erroneous timing’ means the time of a day on which the operator makes most errors can be identified to boost their minds with refreshment. The operator can also take the print of individual error log record. The screenshot of this user error log is illustrated in Figure 4-7 below.



	Log	Log Date
1	वडील किंवा नवरा किंवा पालकाचे नाव आवश्यक आहे	7/2/2015 6:24:21 PM
2	पहिल्या नावाच्या ठिकाणी फक्त अल्फाबेट्स टाका	7/2/2015 6:25:28 PM
3	पहिल्या नावाच्या ठिकाणी फक्त अल्फाबेट्स टाका	7/2/2015 6:25:28 PM
4	पिन कोड निवडलेल्या राज्याबरोबर जुळत नाही	7/2/2015 6:25:28 PM
5	पहिले नाव आवश्यक आहे.	7/2/2015 6:27:12 PM
6	आडनाव आवश्यक आहे	7/2/2015 6:27:12 PM
7	वडील किंवा नवरा किंवा पालकाचे नाव आवश्यक आहे	7/2/2015 6:27:12 PM
8	पत्ता आवश्यक आहे	7/2/2015 6:27:12 PM
9	पिन कोड आवश्यक आहे	7/2/2015 6:27:12 PM
10	जन्म तारीख आवश्यक आहे	7/2/2015 6:27:12 PM
11	लिंग आवश्यक आहे	7/2/2015 6:27:12 PM
12	पिन कोड आवश्यक आहे	7/2/2015 6:29:03 PM
13	पिन कोड निवडलेल्या राज्याबरोबर जुळत नाही	10/26/2015 3:23:16 PM

Figure 4-7: Screenshot of user error detail report generated by *ELIIDE* - tool- Incorporation of local language.

4.4.5. Predictive Text Entry Widgets

The *ELIIDE* - tool supports predictive text entry mechanism (Ali & Meek, 2009). As we can see in Figure 4-8 shows the screenshot of the example of predictive text entry implemented for *ELIIDE* -tool which is highlighted using think red structure. The user is typing first name starting with 'Pra', then the widget displays corresponding predicted values.

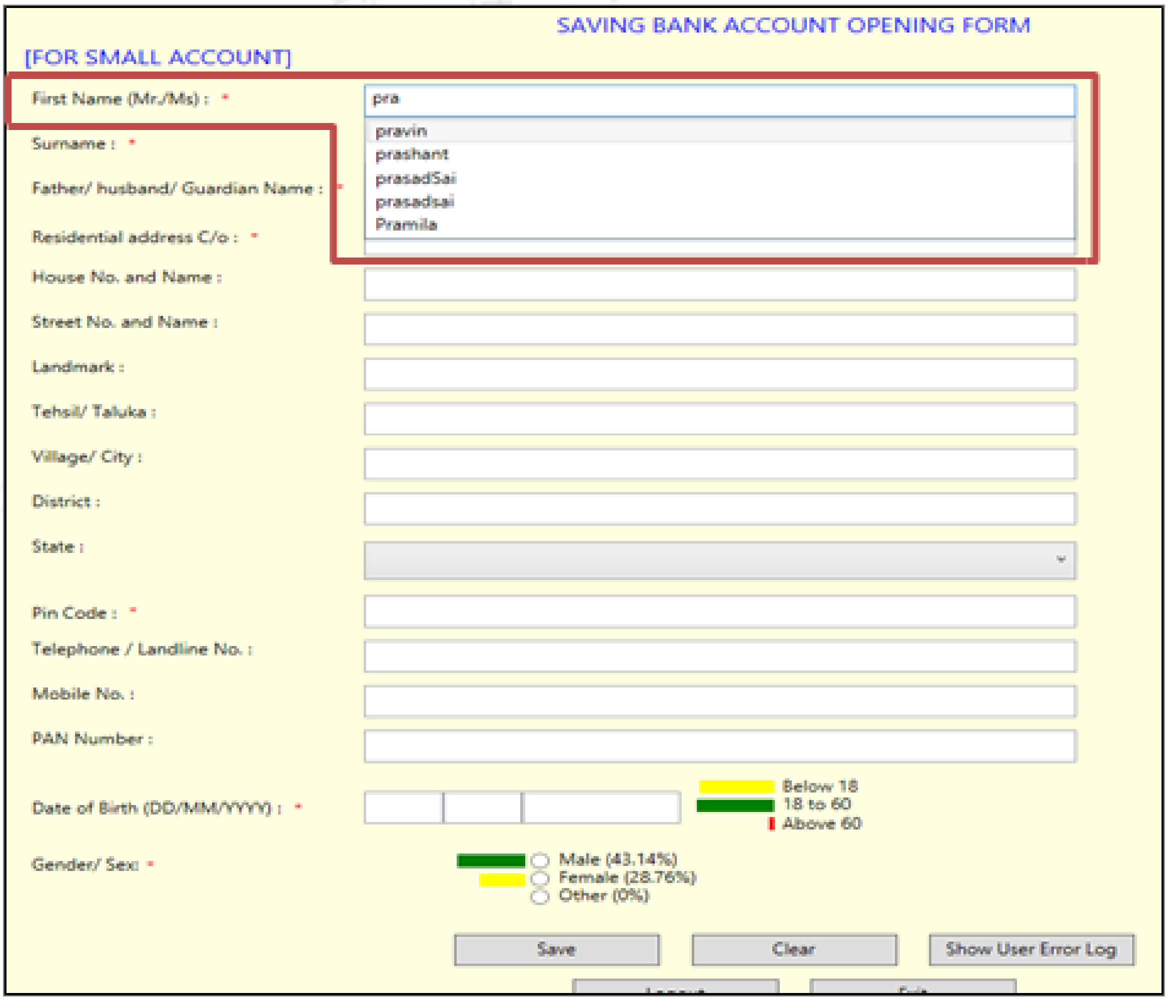


Figure 4-8: Screenshot of data entry form of *ELIIDE*, showing predictive text entry widgets highlighted by red boxes

4.4.6. Dynamic Drop-down Menu

The Figure 4-9 depicts the screenshot of dynamic drop-down menu design implemented for *ELIIDE*- tool. This type of menu design is also known as elective split menu design (Chen K., Chen, Conway, Hellerstein, & Parikh, 2011; Warren & Bolton, 1999). In this menu design, the menu items are spited into two parts. The first/ upper part is showing most frequently used five items (Miller G. A., 1956). Also this list items are supported with quantitative probability so as to judge the most relevant item entry. The second part shows the remaining items sorted alphabetically.

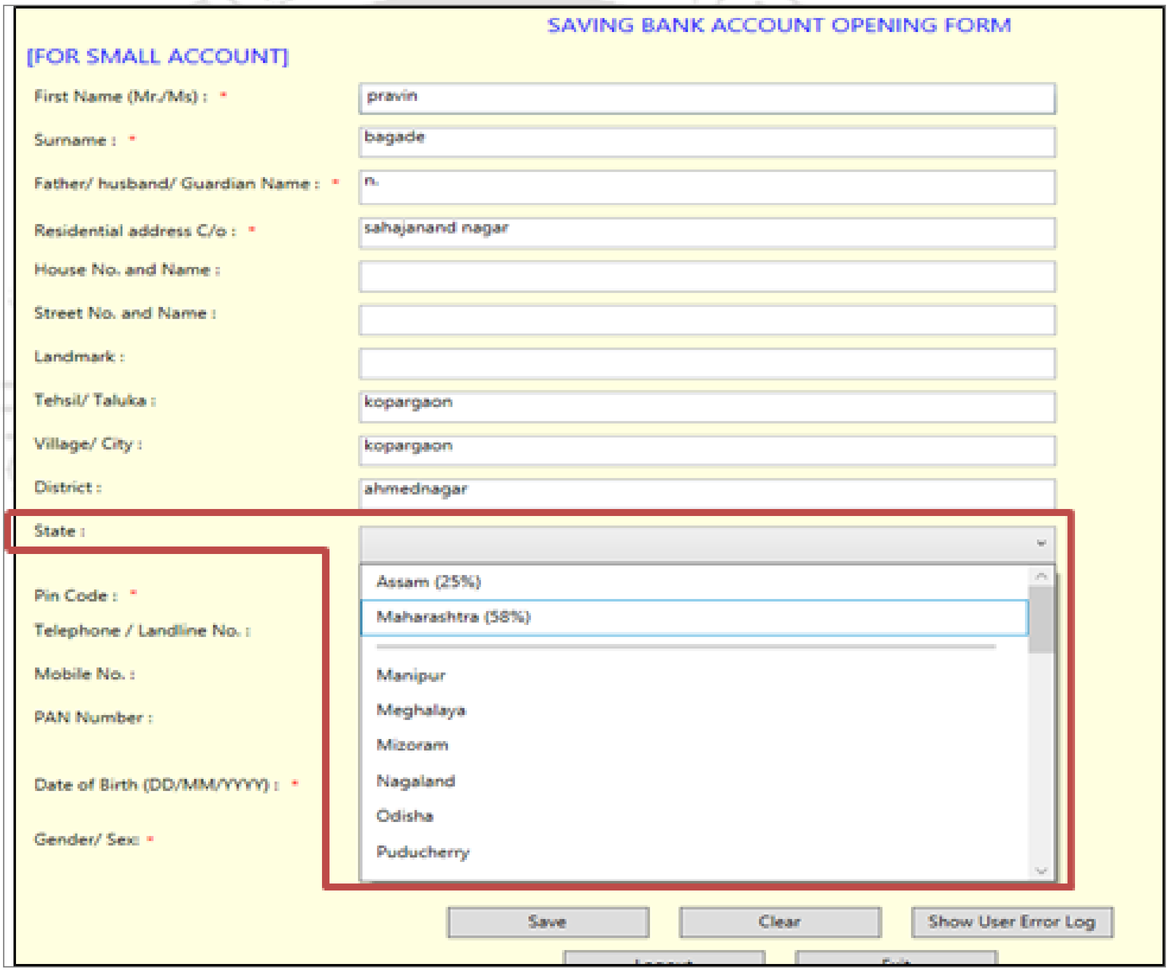


Figure 4-9: Screenshot of data entry form of *ELIIDE*, showing dynamic drop-down menu design highlighted by red boxes

4.4.7. Adaptive Feature

The *ELIIDE* provides an adaptive audio support feature, which is optional for expert users. In Figure 4-11, the additivity feature is shown by using red rectangle box. This feature was adopted by *ELIIDE* according to operator's expertise which is calculated using his performance index. The user/operator having performance index (PI) less than 5%, we called them as expert users. The performance index is calculated as,

$$\textbf{Performance Index} = \frac{\textbf{Error Made}}{\textbf{Error Opportunities}} * 100$$

Here,

Error Made = number of errors made by the operator during data entry,

Error Opportunities= total number of possible errors an operator can make in one data entry form (Figure 5-2), here it is 40.

For example:

If operator made 5 errors in one data entry form, then PI will be

PI= 5 / 40 = 0.12 i.e. 12 % which is greater than 5%.

But if user made 2 errors in one data entry form, then

PI=2 / 40 = 0.048 i.e. 4.87 % which is less than 5%, we called him as expert user.

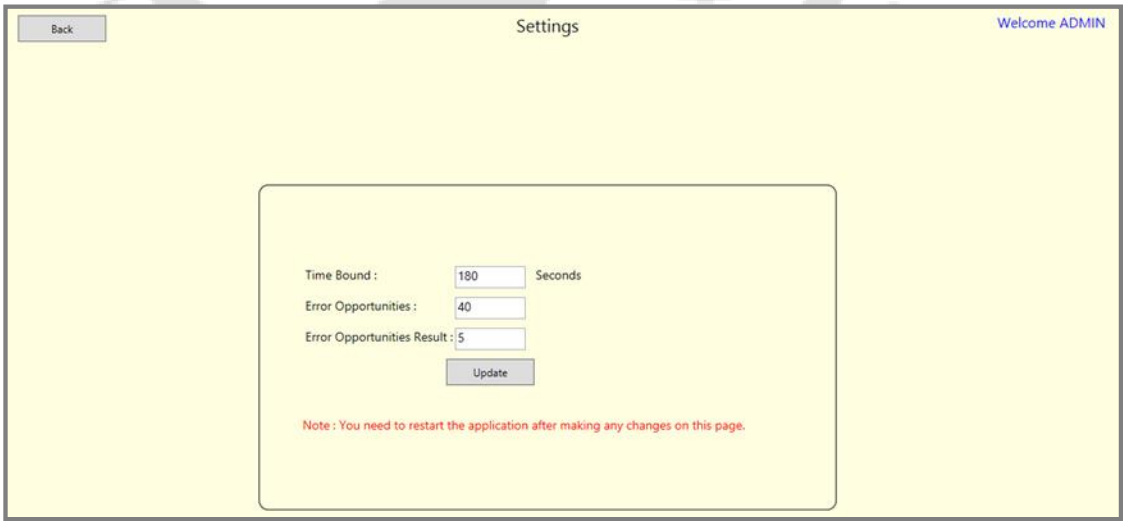


Figure 4-10: Screenshot of Settings of *ELIIDE* tool

As shown in the Figure 4-10, *ELIIDE* - tool provides the option for changing the different settings like 'Time Bound'. For checking the performance of operator on retracted time data entry this option is provided on *ELIIDE* - tool. The time bound of the entry can be adjusted. The performance index limit and error opportunities can be modified using this setting.

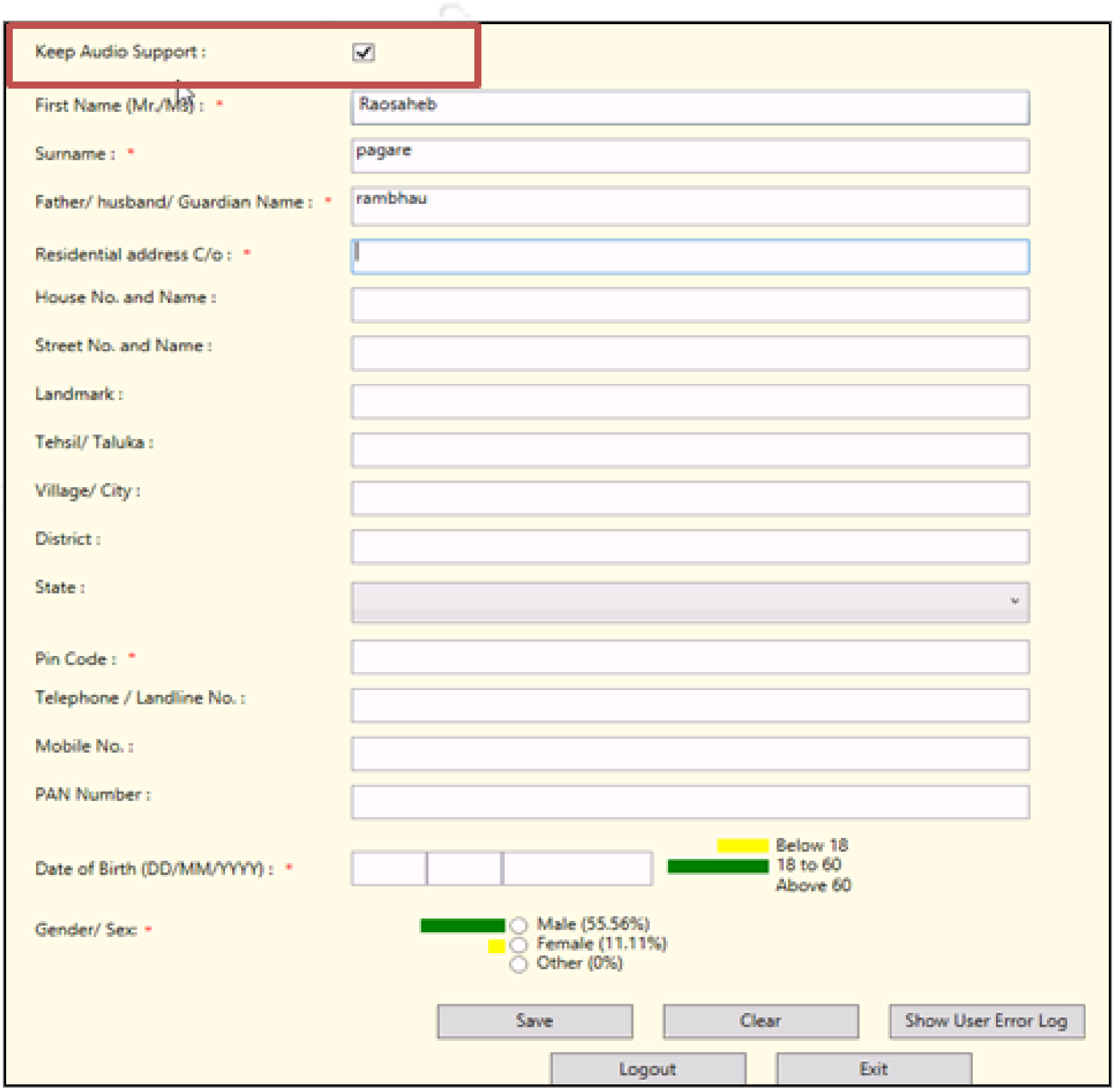


Figure 4-11: Screenshot of data entry form with adaptive feature highlighted by red box of *ELIIDE* - tool

4.4.8. Quantitative Probabilistic Approach

The Figure 4-12 depicts the screenshot of *ELIIDE* -tool showing quantitative probabilistic widgets (Chen, Hellerstein, & Parikh., 2010). The highlighted thick red rectangle box portion in Figure 4-12 shows two different design approach for widgets. First "Date of Birth" field is provided with the 'bar chart' indicating quantitative probability of existing entries for different age groups like 'below 18 years', '18 to 60 years' and 'above 60 years'. The different age groups are represented using different types colours for each bar.

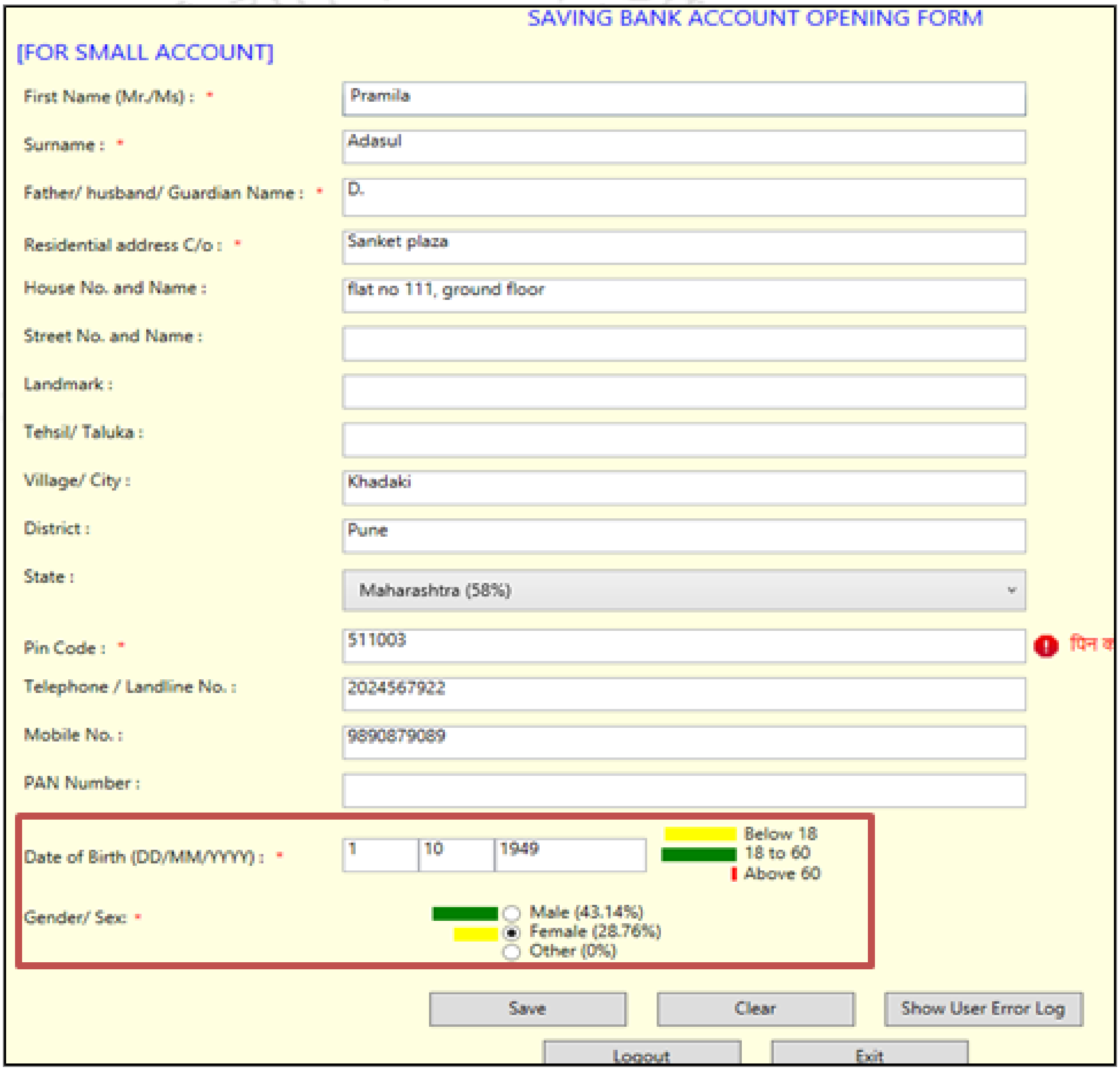


Figure 4-12: Screenshot of quantitative probabilistic widgets highlighted by thick red box, first widget of 'Date of Birth' entry shows the quantitative probability using bar graph for different age groups and second 'Gender/Sex' radio button is supported with numeric probability using percentage and bar graph for particular gender

As shown in Figure 4-12, for "Date of Birth" field the age groups 'below 18 years' is represented using yellow, '18 to 60 years' using green and 'above 60 years' is represented using red colour. The second data entry widget or radio button named as "Gender / Sex"0 was supported with the numeric probability for particular gender.

This type of practice can help operator to cross validate by referring quantitative probabilistic of the entry and make less errors. Also sometime they ignore the cross validation done by referring actual paper document and hence speed up their performance while data entry and therefore make faster entries.

This tool acts as data collection and experiment instrument and while doing this the new tool validation is also carried out. The *ELIIDE* - tool stores all records entered or transcribed (from paper form to computer) by operators during experimental process. The paper forms or data entry forms (see Figure 4-1) was prepared according to actual data entry forms used by operator at rural-BPO. The data entry forms were outsourced by banks. The designed data entry form includes only personal information block (taken from actual form referred by rural-BPO operators) containing seventeen fields.

Therefore, the *ELIIDE* generates different types of graphs to view stored information and errors. The graphs are like- overall errors observe in each data entry field, overall errors done by operators. The *ELIIDE* also provides various reports for individual operator and manager (or admin) too. The generated operators' performance report shows errors performed with date and time. The admin can see any operators' performance report. The tool gives a useful performance summary for a BPO operator indicating errors, improvements in efficiency. No tool has been reported in literature that incorporates local language related requirements.

4.4.9. Generation of Graphs

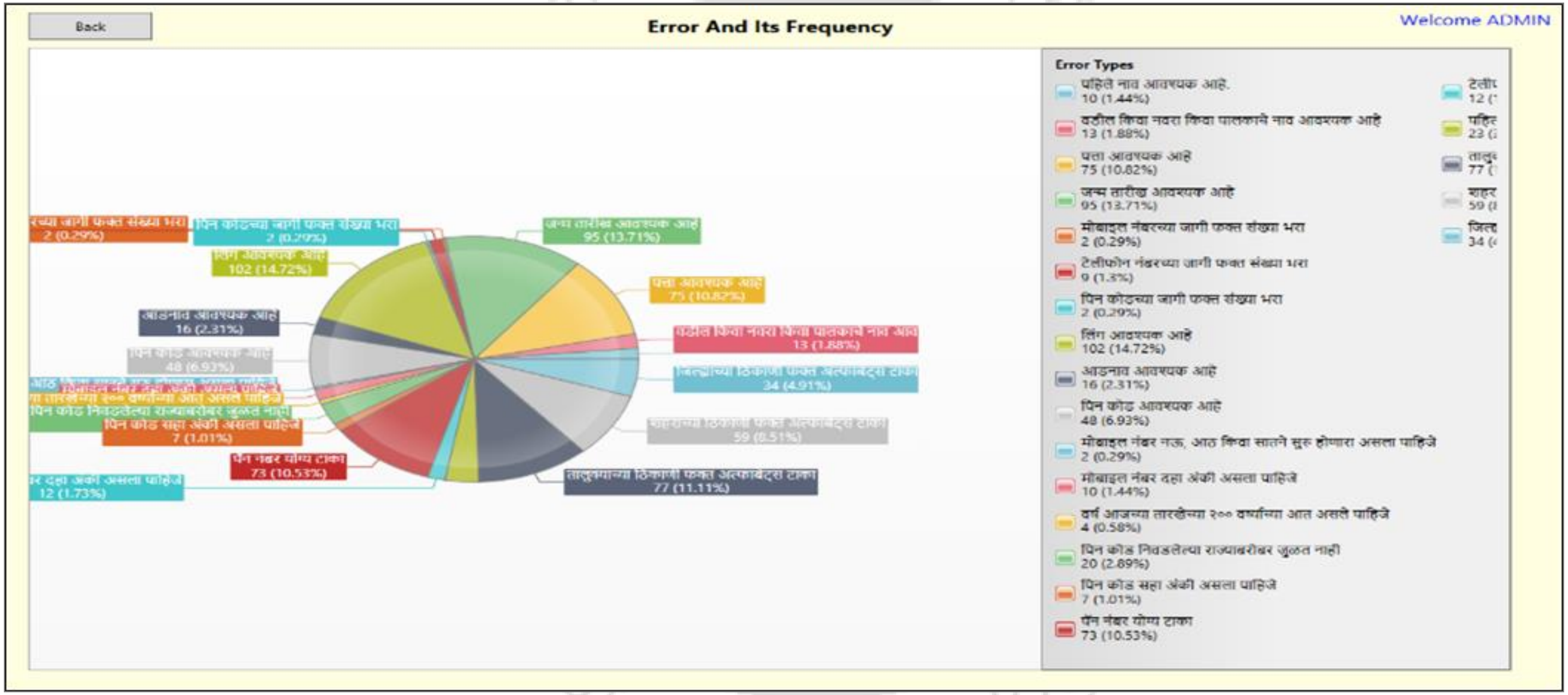


Figure 4-13: Screenshot of data entry form with dynamic widgets of *ELIIDE*.

.

The *ELIIDE* - tool not only shows the information about errors for particular operator but it also analyse and generate graphs for it. It generates different graphs for representation of information in particular way we want, for example, if we want to check errors done field-wise. Using this information, the most erroneous field can be identified and studied for further improvement. So that during data entry they will be more attentive to the most erroneous fields to mitigate errors. The *ELIIDE* - tool also provide pie-graph for showing frequency of each error. This helps to identify the type of errors found more. This information is important for designers and developers to recognize and improve erroneous fields and most committed errors.

The screenshot of one such graph is shown in Figure 4-13.

4.4.10. Additional Features

Additional features provided by *ELIIDE* - tool are,

1. Adding new user- new user can be added in *ELIIDE* - tool.
2. Showing operator's data entry report- This interface generates data entry report for single operator which is viewed by that particular operator only and all operators report is viewed by admin user.
3. Update error and feedback messages- This interface provides facility to add more error and feedback messages for a particular language. The interface has also provision of updating the language of error and feedback messages.

The flexible nature of *ELIIDE* - tool provides the facility to add/ update the error and feedback messages. In India, twenty-two regional languages being spoken, from which *Marathi* language was selected for study and accordingly errors and feedback messages were prepared. The *ELIIDE* also support modification of the language of these errors and feedback messages.

Figure 4-14 depicts the screenshot, showing the buttons to access these additional feature embedded in *ELIIDE* - tool.

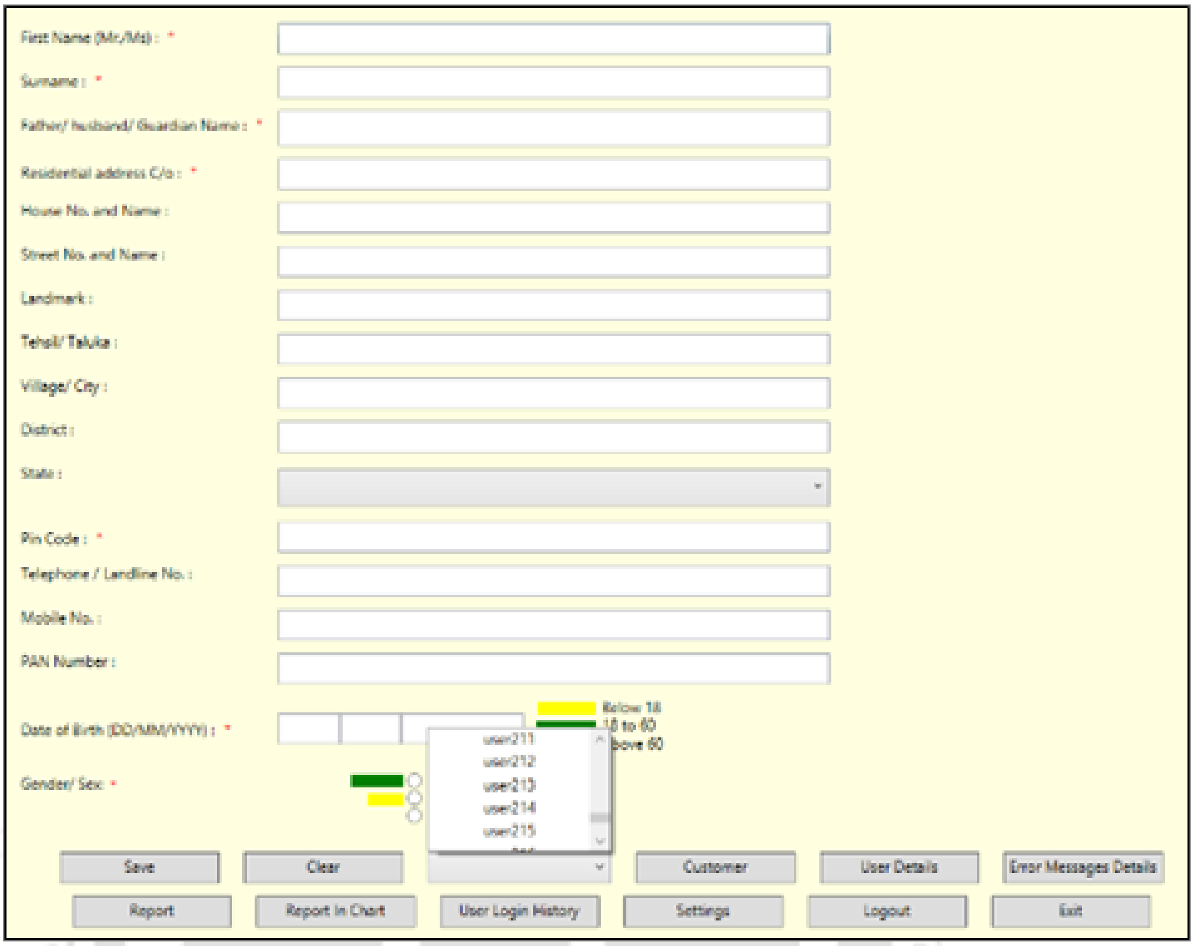


Figure 4-14: Screenshot of additional features like- adding new user, showing user data entry report of *ELIIDE* tool

4.4.11. User Performance Report

If the admin wants to know the performance of particular operator from particular period, then *ELIIDE* - tool generate this information by user performance report. The screenshot of which is presented in Figure 4-15 below.

The 'user performance report' shows error log details like- type of error and when the error happened of individual operator. This information is important for individual to review their faults and improve their performance. So that during data entry they will be more attentive to the most erroneous fields to mitigate errors. The performance report is also important for the manager to evaluate the performance of individual operator. The most erroneous field can be identified and studied for further improvement. The *ELIIDE* interface not only provides the user performance index but also generate the monthly error report. This feature can help manager to monitor the performance of individual operator and pay salary accordingly.

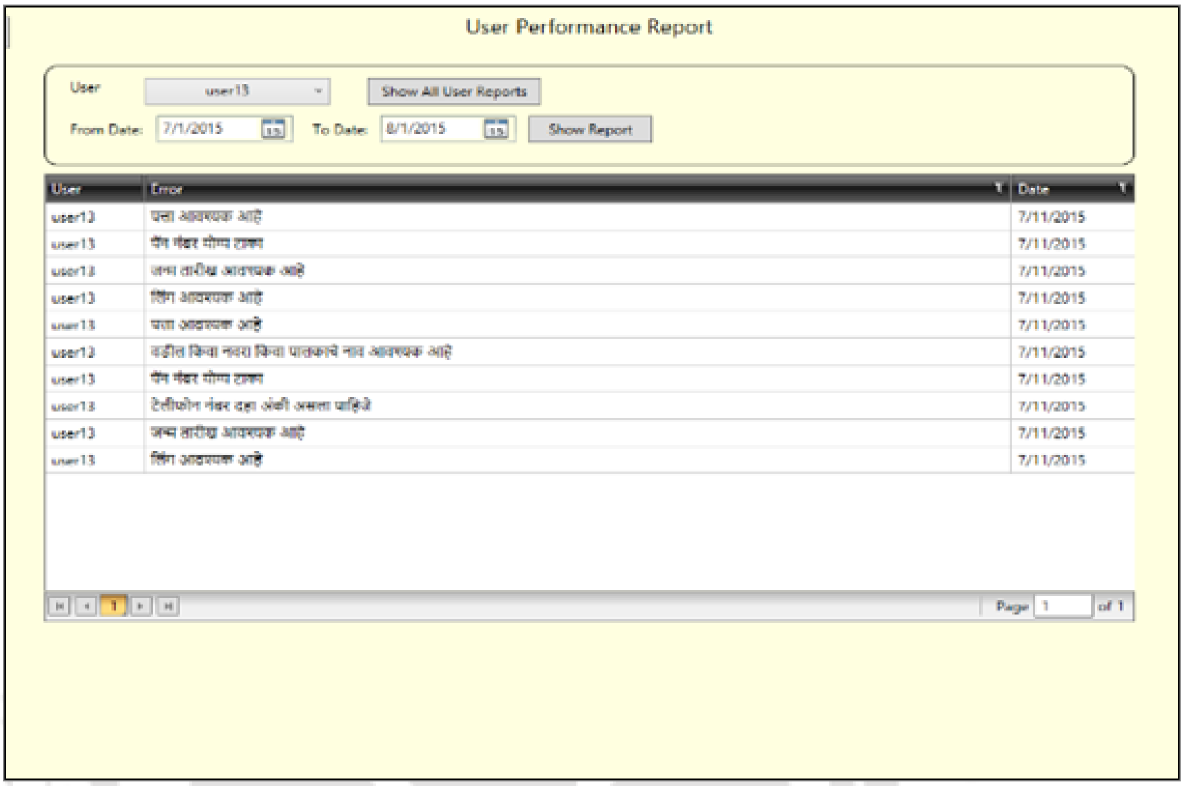


Figure 4-15: Screenshot of user performance report generated by *ELIIDE* - tool

The Table below illustrates the issues found in existing user interface and solution provided by intelligent user interface.

Issues with Existing User Interface	Solved by Intelligent User Interface
User interface is in English language	User interface gives feedback and error messages in Local language the operator speaks with audio support
Fails to correct specific field constraint	Does not allow user to enter special characters and numerical in text fields and special characters and texts in numerical fields. If user does so, *ELIIDE* consider them as errors and add it into errors list.
Does not provide clues during typing	Does provides clues to operator during typing
Fails to provide confirmation logic	Provides confirmation

Does not provide validation logic for fields	Validations provided to fields, for wrong entries it gives error message
Does not provide feedback and error messages in local language with audio support	It provides feedback and error messages in local language with audio support
No support of dynamic drop-down slip menu design	Supports dynamic drop-down slip menu design
No support of probabilistic widgets	Support quantitative probabilistic widgets
No support for Adaptive feature	Supports Adaptive feature
Does not provide user performance index	Provides user performance index
Does not provides any graph to view stored information and errors	Generates different types of graphs to view stored information and errors

This interface provides the user performance index and also generate the monthly error report. This feature can help manager to monitor the performance of individual operator and pay salary accordingly.

4.5.Conclusion

This Chapter presented the development phases and screenshots of the newly developed interface for data entry, named as *ELIIDE* - tool. The *ELIIDE* has different features like dynamic, predictive, adaptive and probabilistic. The interface uses local *Marathi* language to communicate with user / operator. The communication happens in terms of error and feedback messages. This additional feature may support rural users to get emotionally attached to interface. The generation of different reports like user log record and user performance report may also help the manager to track the performance of particular user.

Chapter 5

Experimental Methodology: User Testing, Research Methods, Experiment Design

While the last Chapter reported initial pilot studies this Chapter presents the overall main experimental methodology used to address the research questions. Initially, the working hypotheses, independent and dependent measures have been listed. The methodology involving- an instrument used for data collection, sampling framework and the procedure adapted for data collection is reported. The experiments and the methodology for them are fully described. All details of how the empirical side of the research has been conducted.

5.1. Introduction

The state of the art review in Chapter 2 has highlighted several research gaps which lead to the formation of nine research questions. The first research gap is, the interface designed with intelligent widgets like display of autocomplete suggestion for text field by ranking strategy based on likelihood, predictive text entry widget, radio button pointed with most likely options and dynamic drop-down split-menu- can have influence on data entry and the errors. The second research gap is, other factors like culture (language) and emotions can also have effect on data entry errors. The last research gap is, the influence of factors like errors, cognitive load, system usability, satisfaction, willingness to continue usage of proposed new interface design with intelligent widgets is explored.

This Chapter demonstrates the experimental methodology carried out to prove a set of hypotheses. It includes the data collection, sample selection, instruments used. A new tool was developed with which this experiment was performed. The new GUI tool is explained in detail in next Chapter only the experiment is reported in this Chapter.

5.2. User Testing (or Research Methods): Data collection, Participants, Instrument

5.2.1. Data collection methods

The data collection involves collection of specific data regarding system, activity and personnel. It includes evaluation of existing operational system through usability, error analysis and task analysis techniques (Stanton, Salmon, Walker, Baber, & Jenkins, 2006).

Interviews: We have conducted semi-structured interviews to gather information from the operators and manager at rural-BPOs. We have gather information regarding systems usability, user perceptions and reaction about system and errors during interaction.

Questionnaires: We have also provided set of questionnaires for getting reaction about system, including usability, user satisfaction, error, user opinion and attitude. More details regarding questionnaire reported in section below.

Observation: Observational method helped us to gather data regarding the physical and verbal aspects of the data entry task or scenario at rural-BPOs. These include task performed by system (types), the operators performing the tasks, task step and sequence, error made, communication between operators, the technology used by the system in conducting the tasks. We have photograph, video recorded a particular task or scenario.

5.2.2. Participants

Convenience sampling was done for selection of the subjects based on ease of access and availability. Two hundred and twenty-four (224) participants volunteered for the study from which One hundred and three (103) were professional data entry operators working in rural-BPOs like- *Source2Rural*, *RuralShores* and *Maitreya* and one hundred and twenty-one (121) students from three poly-technique institute. Both groups of participants were from different places from Maharashtra, a state in Western India. Fifty male and fifty-three female operators participated from three rural-BPOs. Fifty-eight male and sixty-four female students i.e. non-operators participated from three poly-technique institutes. Table 5.1 below gives a snapshot of the participants who took part in the study.

Table 5-1: Participants details

Gender	Operator	Non-operator
Male	50	58
Female	53	63
n	103	121
Total n= 224		

All participants had a minimum of six months of experience of using computer and had primary education (upto 10th standard) in local (i.e. *Marathi*) language. Participants were between 18-32 years old (M=22.62, SD=4.05; Operator: M=25.56, SD=4.21; Non-operator: M=20.12, SD=1.28). A within group design study was formulated for the investigation.

Table 5-2: Demographic information of participants

	Operators	Non-operators
Age group	18 to 32	18 to 30
Qualification	Diploma or Graduate in any field, E.g. BA, BCOM, BSC Etc.	Diploma final year students
Primary Education	In local language	In local language
Computer Experience	Minimum 6 months of data entry	Minimum 6 months

The Table 5-2 depicts the demographic information of participants involved in this study. Age group of participant is 18 to 32 years. The qualification of the operators was diploma and graduate in any field like BA (Bachelor of Arts), BCOM (Bachelor of Commerce), BSC (Bachelor of Science) etc. All participants were done their primary education in local language and minimum six months of experience using computer.

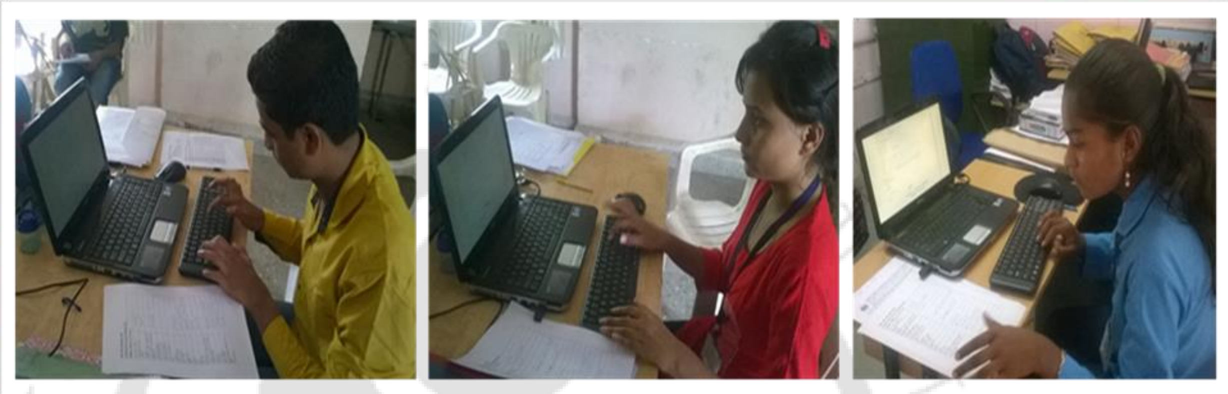


Figure 5.1: Participants performing ecperiment (Photographs used by consent)

Figure 5.1 shows the participants performing the data entry experiment.

5.2.2.1. Sample Distribution

Convenience non-probabilistic sampling technique was followed to draw a sample from the entire population. The sampling formula followed was of $\frac{Z^2*(p)*(1-p)}{c^2}$. Proposed sample drawn thus further underwent Kolmogorov-Smirnov test for normality and of Kurtosis distribution based on whether passed Kolmogorov-Smirnov test it was consider to follow or not the norms of normal distribution. Accordingly, leading them to be further subjected with parametric or non-parametric analysis. The sample drawn using above formula followed Gaussian, Poisson, Binomial distribution formula with rejection region lying at $\alpha/2=0.05$ for the distribution.

5.2.3. Instruments Used

In the current research study, questionnaires were used as an instrument to capture the participants’ subjective evaluation of the user interface. Participants had to complete the given set of tasks on provided user interfaces on computers and thereafter they were administered the questionnaires. The study conducted in this investigation employed purposive homogeneous sampling.

A. The participants were randomly selected for the experiment of data entry on provided user interfaces, which started with a demographic questionnaire.

B. This research study introduces a 'tool' or 'user interface' named as '*ELIIDE* -tool' specially designed with widgets having intelligent features, we have already discussed about this interface in the 'Chapter 4'. The *ELIIDE* -tool was divided into two parts (or screens or forms), first is 'Intelligent User Interface' and second is 'Existing User Interface'. Each participant needs to perform data entry on both parts of the interface. This tool acts as data collection and experiment instrument and while doing this the new tool validation is also carried out.

C. The data entry forms (refer Figure 5-2) were provided to each participant. The data entry forms were designed specifically for this research in three different variations like in the English language, in *Marathi* language and in mixed language (using both languages). (a sample is attached in Appendix 2). During empirical study we have observed that, the actual data entry forms presented to operators which are outsourced by different outsourcing agencies like Banks are written in two different languages that is either in English or in local language or sometime in mix language also. So, we have taken three different variations of data entry forms as shown in Figure 5-2. The Figure 5-2 below depicts the few copies of data entry forms used in this experiment.

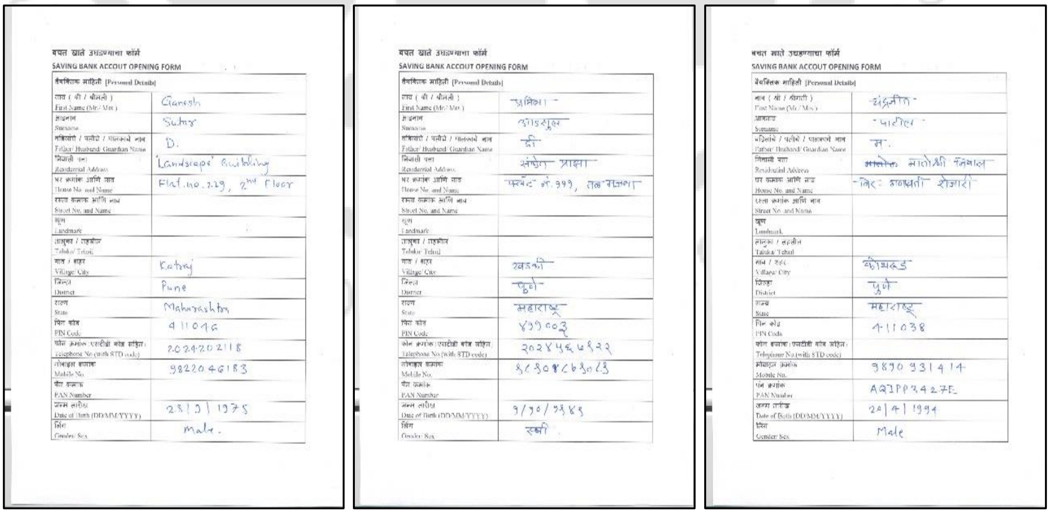


Figure 5-2: Data Entry forms in three format (a) English language (b) *Marathi* language (c) Mixed

D. After the completion of the main experiment session participants were supplied a post-test questionnaire comprising of following items:

i. National Aeronautics and Space Administration Task Load Index [NASA TLX (Hart & Staveland, 1988)]

ii. System Usability Scale [SUS (Brooke, 1996)]

iii. Questionnaire for User Interface Satisfaction (version 5.0) [QUIS (Chin, Diehl, & Norman, 1988)]

iv. Relative advantage [RA (Poelmans, Wessa, Milis, Bloemen, & Doom, 2008)]

v. Willingness to continue to use [WCU (Bickmore & Picard, 2004)].

5.3. Experiment Design

5.3.1. Experiment Variables

A within-group study has been proposed. The graphical depiction of the experimental design is presented in Figure 5-4 and experimental variables including dependent and independent variables shown in Figure 5-3 below.

A within-group design was adopted for the study. The study investigates three independent features as user interface, subject groups and task variations. The user interface feature is having two variables like Existing UI and Intelligent UI. The subject groups are operators and non-operators. The last feature task variation represents three variables as English language, *Marathi* language and Mixed language. Two factorial design was adopted. Eight dependent variables were captured during the experiment, three were quantitative and five were qualitative. Therefore, t-test and ANOVA analysis technique were adopted for the analysis of the data.

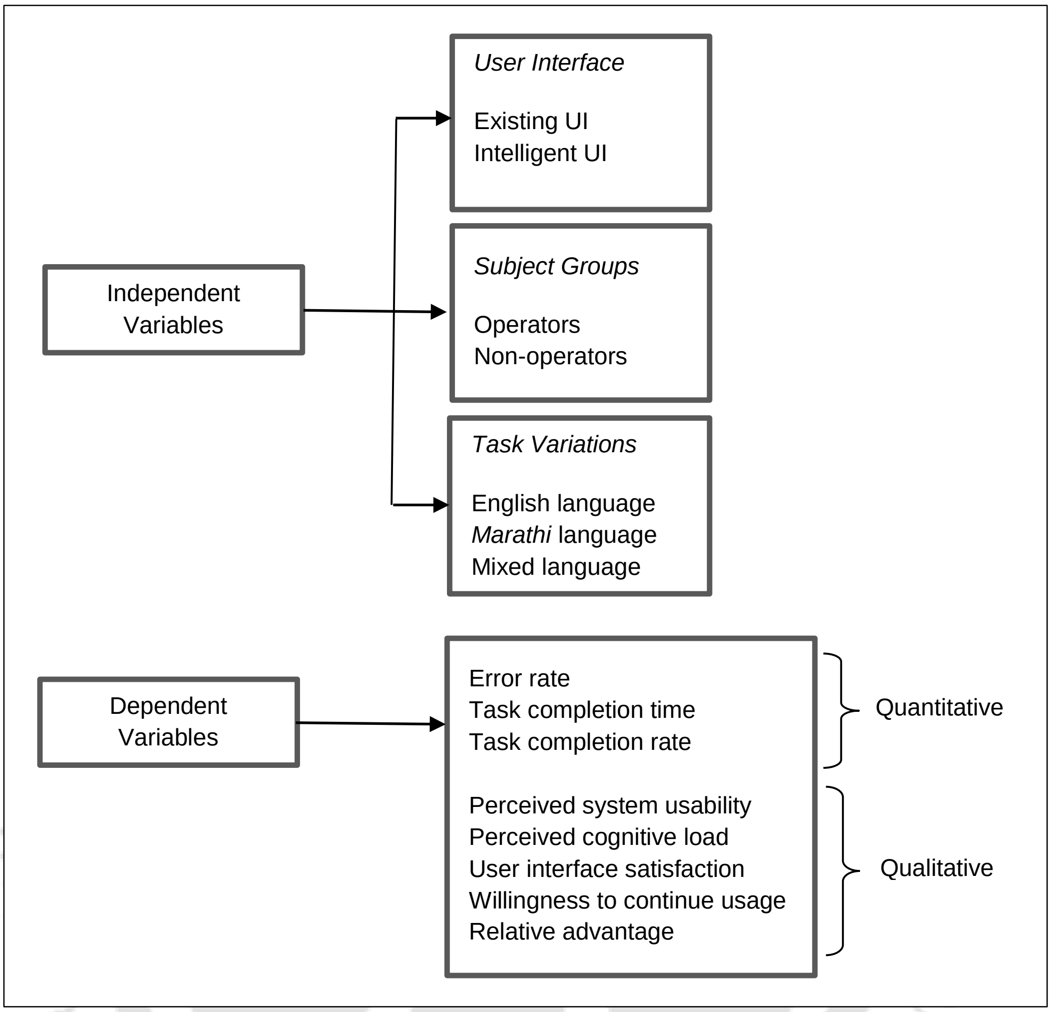


Figure 5-3: Experimental Variables; Source: Author-generated

5.3.2. Task Design

Four tasks were given to each participant but the sequence of the task may be different. Table 5-3 below depicts the names of four tasks performed by participants. Task1 consist of data entry on existing user interface, Task2 consist of data entry on existing user interface with time bound of 180 seconds and so on.

Table 5-3: Task Description

Tasks	Name of Task
Task1	Existing UI entries (EUI)
Task2	Existing UI entries with time bound (EUI+TB)
Task3	Intelligent UI entries (IUI)
Task4	Intelligent UI entries with time bound (IUI+TB)

**UI means User Interface*

The samples were equally distributed among each set of data entry forms, as shown in Table 5-4. Ninety data entry forms having three groups each of thirty were used in the study.

Table 5-4: Distribution of samples among each set of task variation forms (i.e. data entry forms)

	Forms used for data entry		
	30- English	**30- *Marathi***	**30- Mixed**
Operators (103)	34	34	35
Non-operator(121)	40	40	41

5.3.3. Procedure in details

The graphical depiction of the experimental design and detail procedure is presented in Figure 5-4. The details procedure of conduction of experiment is discussed below,

Orientation / practice session: Prior to the actual experiment, the participants were explained about the design and purpose of user interface and also provided practice session on it. They used interface half hour daily for one-week duration before actual experiment started. The participant, in the experiment, does data entry on provided interface using data entry forms / sheets.

Main experiment session: Before going for the actual experiment the participants were told to fill pre-test questionnaires (refer Appendix-1A) which include demographic information. Each participant performed four tasks (shown in Table 5-3), two tasks were data entry on existing interface (having static widgets) and other two were on the intelligent interface (having dynamic widgets). The sequence of the task was random to avoid learning effect. The tasks consist of a transcription of given data entry form (Figure 5-3) (also called as paper form) into electronic form using both interfaces. Participants were instructed to perform the tasks as quickly and accurately as possible. The computer based background recording of each participant interaction with the designed user interface have taken for calculation of the accuracy and speed.

Post-test Questionnaires: After completion of the experiment the participants were instructed to fill the post-task questionnaires (refer Appendix-1B) like -National Aeronautics and Space Administration Task Load Index (NASA TLX), System Usability Scale (SUS), Questionnaire for User Interface Satisfaction (QUIS), Relative advantage (RA) and Willingness to continue to use (WCU) to express their opinion and experience about the user interface. The subjective experience was recorded in terms of cognitive load, perceived system usability, user interface satisfaction, willingness to continue usage and relative advantage.

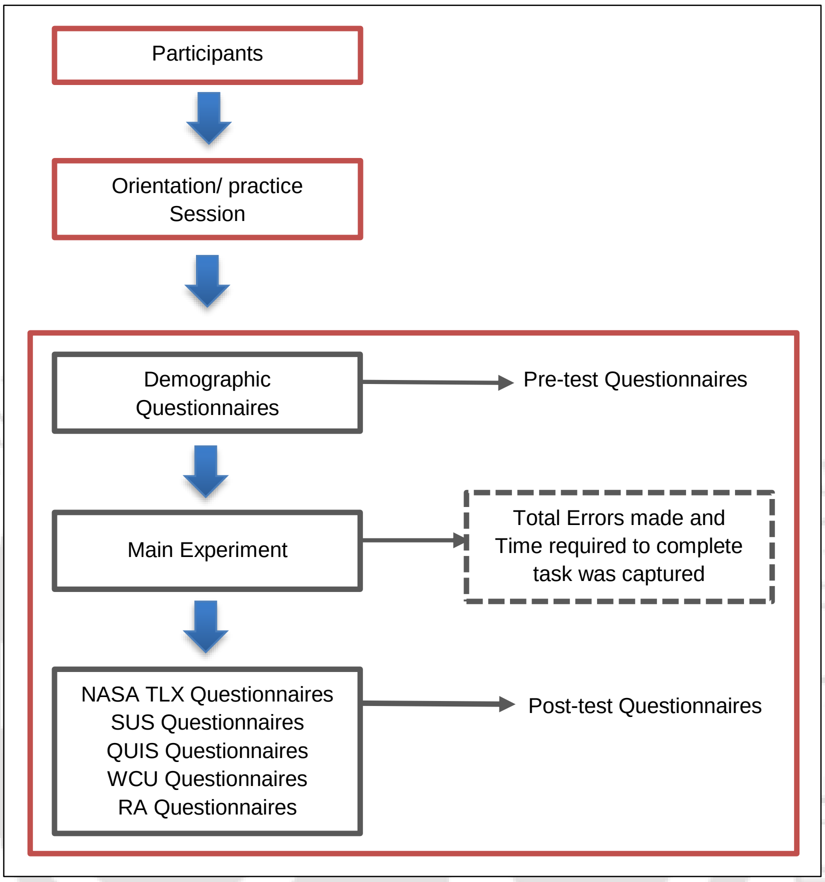


Figure 5-4: Graphical representation of the experiment design and process; Source: Author-generated

Figure 5.4 depicts the complete procedure of this experiment which is explain above.

5.4. Conclusion

This Chapter presented the overall main experimental methodology used to address the research questions and working hypotheses. The methodology involving- an instrument used for data collection, sampling framework and the procedure adapted for data collection is reported. The experiments and the methodology for them are fully described. The user testing through empirical evaluation of the research has been conducted.

Chapter 6

User Testing / Verification of Designed User Interface- *ELIIDE* tool: Results and Analysis

This Chapter reported the statistical analysis of the data collected during the experimental study. A t-test and ANOVA analysis approach has been adopted to analyse the data sets. The experimental result highlights the effect of intelligent feature on the performance in term of speed and accuracy and their subjective evaluation of the intelligent user interface (*ELIIDE* - tool). Detailed discussion of the results from the perspective of the working hypotheses of the experiments has been done.

6.1.Introduction

This Chapter reports validation of the *ELIIDE* - tool, by actual data entry operators working in rural-BPOs. The experimental result highlights the effect of intelligent feature on the performance in term of speed and accuracy and their subjective evaluation of the intelligent user interface (*ELIIDE*). As stated earlier this tool was developed both as an instrument for experimentation and also while it collects data- which after analysis can also be used to validate the tool.

6.2.Results and Analysis

To identify the significant differences among groups, separate paired sample t-tests were employed. Table 6.1 enlists the paired sample t-test tables. It is important to state here that t-tests are generally robust against violations of normality (Edgell & Noon, 1984). Therefore, a paired sample t-test carried out to compare means between the groups for within group study.

It was followed by one-way ANOVA. First, test of normality was carried out by Shapiro Wilk test for each individual measure. The ANOVA is a statistical method used to compare the means of two or more groups. There are different types of ANOVA, first is one-way ANOVA having one factor with at least two independent levels. For example, Hypothesis (H_4)- There is significant difference between errors committed by female operator compared to male. Here factor 'Gender' has two levels like female and male which is independent and error is dependent variable. Therefore, one-way ANOVA statistic was used.

The subsequent sections in this Chapter enlist the detailed results of each of the hypothesis.

6.2.1. Hypothesis (H_1)

The working hypothesis (H_1) which were tested in the experiment are given below:

$H_{1(a)}$: The user interface designed with intelligent features like- (i) display of autocomplete suggestion for text field by ranking strategy based on likelihood, (ii) predictive text entry widget, (iii) radio button pointed with most likely options and (iv) dynamic drop-down split-menu, does affect the accuracy of data entry.

The paired sample t-test conducted on the data sets revealed (Table 6-1) that there was a statistical significance ($t_{(103)}$= 3.301, p>0.01), between exiting user interface (M=1.07, SD=0.68) compared to intelligent user interface (M=0.82, SD=0.64) on errors observed during data entry by operator in regular time. A statistical difference ($t_{(103)}$= 2.205, p>0.05), between exiting user interface (M=1.21, SD=0.79) compared to intelligent user interface (M=1.03, SD=0.76) on errors observed during data entry by operator in limited time given. A statistical difference ($t_{(58)}$= 5.188, p>0.001), between exiting user interface (M=3.17, SD=1.14) compared to intelligent user interface (M=2.48, SD=1.26) on errors observed during data entry by non-operator in regular time. A statistical difference ($t_{(53)}$= 2.260, p>0.05), between exiting user interface (M=2.00, SD=1.24) compared to intelligent user interface (M=1.60, SD=1.06) on errors observed during data entry by non-operator in limited time given.

Table 6-1: Results of hypothesis 1

			User Interface				
	Dependent Value		**Existing UI**		**Intelligent UI**		***t***
		N	**Mean**	**Sd**	**Mean**	**Sd**	
Operators	Regular time	103	1.07	0.68	0.82	0.64	3.301**
	Limited time	103	1.21	0.79	1.03	0.76	2.205*
Non-operators	Regular time	58	3.17	1.14	2.48	1.29	5.188***
	Limited time	53	2.00	1.24	1.60	1.06	2.260*

* Significant at .05 level of significance (P<0.05)

** Significant at .01 level of significance (P<0.01)

*** Significant at .001 level of significance (P<0.001)

Table 6-2 depicts the classification of errors reported in the experiment.

Table 6-2: Two types of errors observed in experiment

Text Entry Errors	**Numerical Entry Errors**
Mistype/ Spelling/ Incorrect: substitutions and intrusions	Wrong
Transposition	Reverse
Doubling	Double
Case	Missing
Capture, phonetic, misinterpretation	
Omission/ Wrong field	

There are many schemes suggested by researchers (Rumelhart & Norman, 1982; Grudin, 1984; Salthouse, 1986; Lang, Graesser, & Hemphill, 1991; MacKenzie & Tanaka-Ishii, 2007; Oladimeji *et al*., 2011) for classifying errors. After studying all these classification schemes we have adopted the most suitable for this experiment depicted in Table 6-2. The total error observed in the experiment were divided into two broad categories as text entry errors and numerical entry errors as shown in Table 6-2. The classification of data entry error is described below:

(1) Mistype/ Spelling/ Incorrect: substitutions and intrusions: This error occurs when the user forgets character/s within words or whole word in transcribed text.

Presented text: ‘BWCPK1233L’
Transcribed text: 'BWCPKI233L'

In the example above while typing ‘PAN-card number’, the operator has typed ' I ' instead '1', constituting a transposition error.

(2) Transposition: This type of error occurs when the user switches/ swaps two adjacent character in a word.

Presented text: ‘IIT Guwahati’
Transcribed text: ‘iti Guwahati’

(3) Doubling: In this type user tries to create an unwanted duplicate character after a target character.

Presented text: 9899429071
Transcribed text: 9899942907

(4) Case: This refers to entering a target character in the wrong case, that is creating a uppercase letter when it is supposed to be in lowercase, or vice versa.

Presented text: *Pooja*
Transcribed text: *pooja*

(5) Capture, phonetic, misinterpretation: The capture error occurs when operator intends to type one sequence, but gets captured by another that has a similar beginning (Rumelhart & Norman, 1982). Operator reads a message and does not construct the correct meaning of the message. (Lang, Graesser, & Hemphill, 1991).

Example of phonetic error: Presented text: *Tankar*
Transcribed text: *Tanker*

Example of misinterpretation: Presented text: *Kharat*

Transcribed text: *Thorat*

(6) Omission/ Wrong field: Operator fails to perform particular operation or omitted particular character from word or performed wrong operation.

Presented text: *Sindhu*

Transcribed text: *Sidhu*

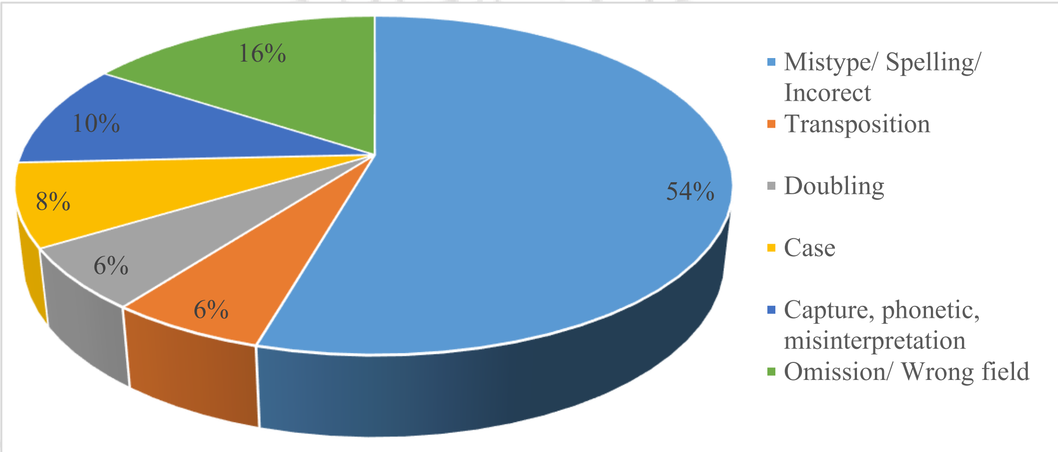
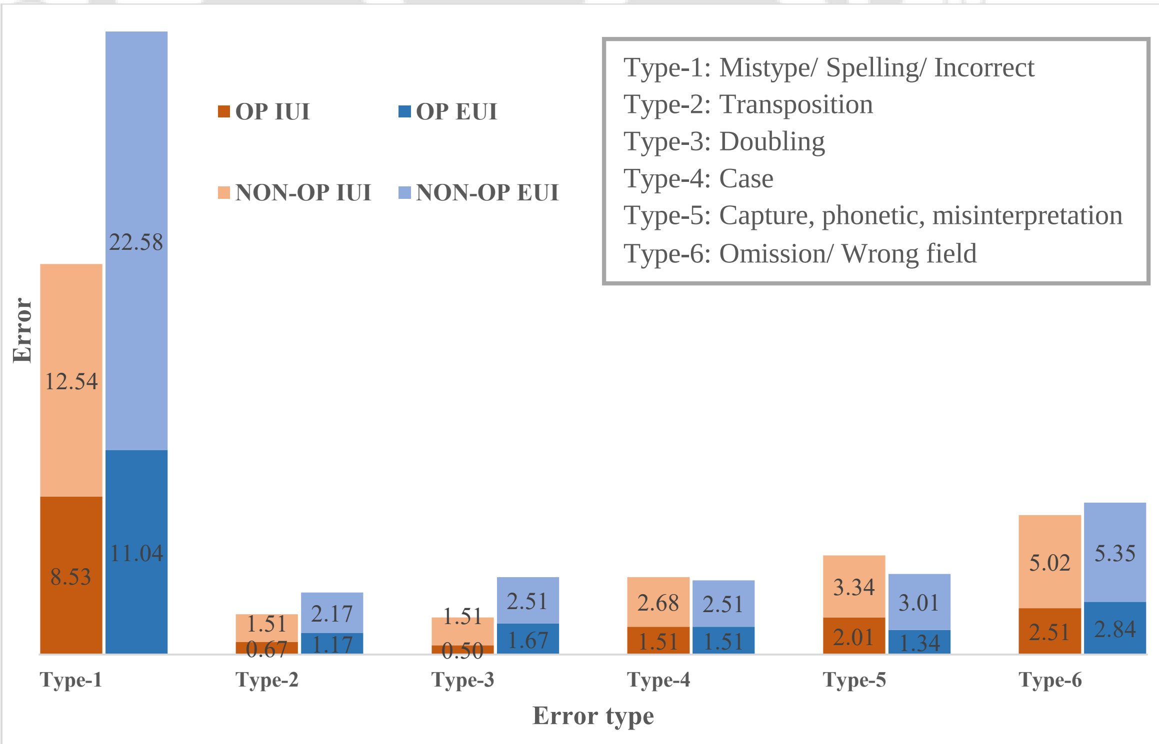


Figure 6-1: Graph showing six categories of text entry errors



Note: OP IUI: Operator Intelligent User Interface, NON-OP EUI: Non-operator Existing User Interface

Figure 6-2: Graph showing various types of text entry errors observed in intelligent and existing user interfaces

Figure 6-1 shows the pie-graph of type of text data entry errors and their frequencies in percentage. The text errors observed in the experiment were classified in six groups.

As the results show significant difference between intelligent UI and existing UI on errors, this might be because of large difference between 'Mistype / spelling / incorrect' type of errors (shown in pie-graph of Figure 6-1) for both experimental conditions. We have observed that, there are few errors due to 'language effect' e.g. the presented text is 'AQIPP3427E' and transcribed text is 'A□IPP3427E'. The error is in second character, instead of 'Q' the operator typed '□'. This may be due to 'language effect' because the operator was presented with mixed data entry form which contains *Marathi* language and English language contents. The predictive mechanism of the *ELIIDE* - tool may help to minimise their memory load and result in less errors compare to existing one.

The bar-graph shown in Figure 6-2 depicts the classification of text entry errors observed in intelligent UI and existing UI made by operator and non-operator. There is large different in 'Mistype/ Spelling/ Incorrect' and 'doubling' type of errors for two different type user interfaces. This is because the predictive mechanism provided for *ELIIDE* -tool, predicts the future entry which helps the operator to select the most relevant entry.

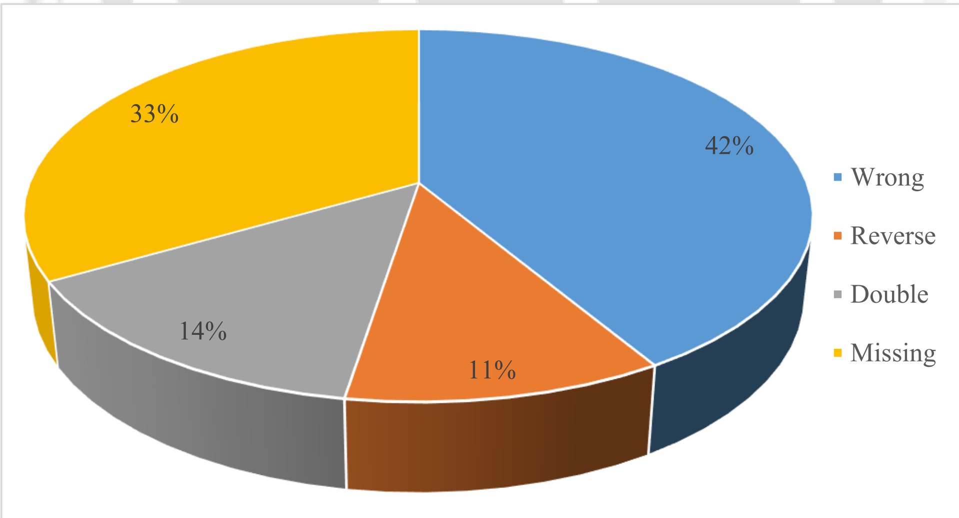
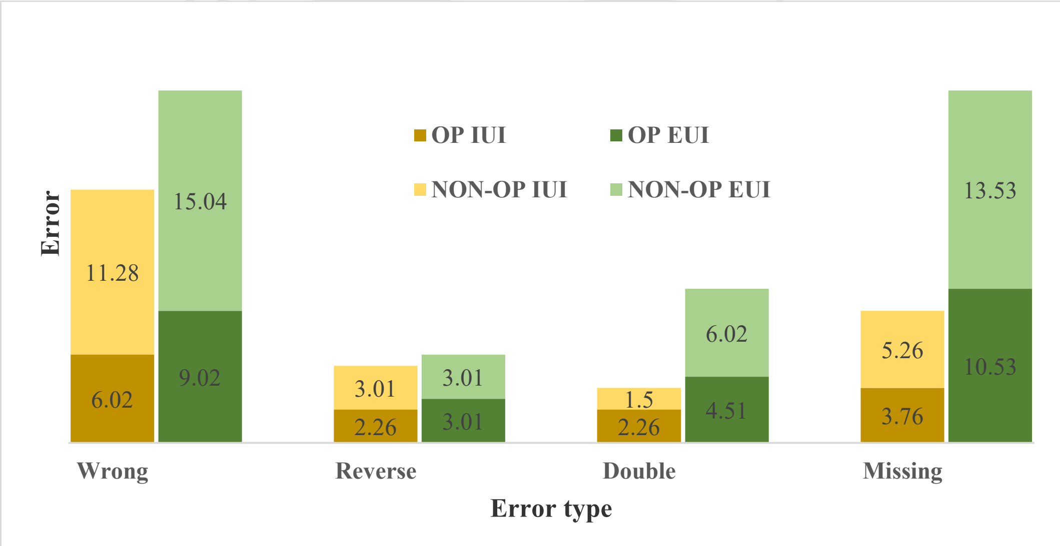


Figure 6-3: Categories of digit (numerical) entry errors

Figure 6-3 shows pie-graph of categories of numerical entry errors observed in this experiment. The percentage of 'wrong' and 'reverse' types of errors is more compare to other types. There are four numerical entry fields in the form as 'PIN code', 'Mobile number', 'Telephone number' and 'Date of Birth'. Also, few text entry fields like 'House number and Name', 'Street number and Name' and 'PAN number' where numerical entries were done by operator.

The bar-graph shown in Figure 6-4 depicts the classification of numerical entry errors observed in intelligent UI and existing UI made by operator and non-operator. There is large difference in 'missing' type of errors from intelligent UI and existing UI designs. This is because intelligent UI is supported with the field constraint, if it does not satisfy *ELIIDE* gives error. For example, the constraint for 'Mobile No.' field is that it must have ten digits and if violates error message shown.



Note: OP IUI: Operator Intelligent User Interface, NON-OP EUI: Non-operator Existing User Interface

Figure 6-4: Graph showing various types of digit (numerical) entry errors observed in this experiment

Figure 6-5 shows the bar-graph for representation of errors observed widgets-wise. It is observed that mostly the fields like 'Address', 'Taluka', 'Village/ City', 'District', which are not supported with intelligent features, the operator makes comparatively more errors.

The specially designed widgets like 'State', 'Date of Birth' and 'Gender', supported with intelligent features of intelligent user interface are observed to be less error making fields compared to fields from existing user interface.

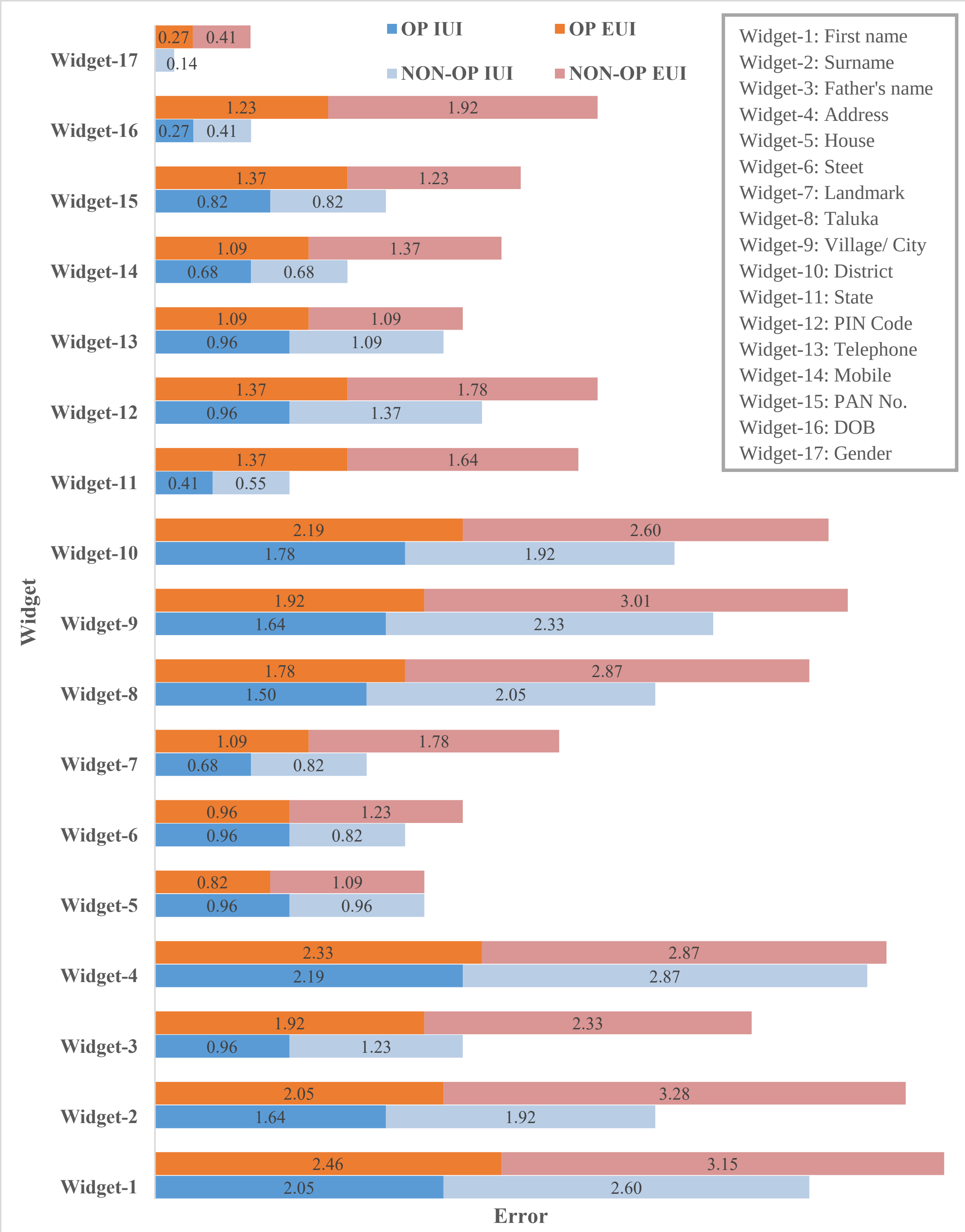


Figure 6-5: Distribution of errors according to 17-widgets

Table 6-3 depicts few more examples of errors observed during this experiment, these error types adopted after studying different classification of data entry errors from literature. The adopted classification of errors was elaborated below Table 6-2 with example.

Table 6-3: Error types with their examples

Error type	**Example**
Phonetic	P:*Sangamner* T:*Sangamnar*
	P:*Tankar* T:*Tanker*
	P:*Varsha* T:*Versha*
Mistype/ spelling/ incorrect	P:EVJPS1423K (PAN Number) T:EVJPSL423K
	P:9930805970 T:9693308059 (ERROR-written his own mobile no.)
	P: BWCPK1233L T: 'BWCPKI233L' (ERROR- trying to type 'I' instead '1')
Omission- Skip character, word	P: *Swathi* T: *Swati*
	P:*Sindhu* (F) T:*Sidhu* (M) (ERROR- changed sex)
Misinterpretation	P: *Kharat* (in *Marathi* language) T: *Thorat*
Wrong key	P:07 T:o7
	P:Q T:2 (ERROR -TWO-'□' IN *Marathi* language looks like 'Q' of English language)
	P:03612582820 T:'036=' (ERROR-Trying to enter)
Language effect	P:□□□□□□ (423601) (in *Marathi* language) T:423609 (ERROR- *Marathi* digit '1' looks like English digit '9')
	P:□□□□□□□□□□□ (02423224840) (in *Marathi* language) T:02423482248 ERROR- Transposition
	P:Q T:2 (ERROR: TWO-'□' IN Marathi language looks like 'Q' of English language)
Mistakenly written his own pin code instead presented	
Mistakenly written his own mobile number instead presented	
Wrong field	ERROR- Typed next field entry in same field
Transposition error	P: IIT Guwahati T: iti (ERROR- trying to type)
Case	P:*Pooja* T:*pooja*
Doubling	P: 9899429071 T: 98999 (ERROR- trying to type)

6.2.2. Hypothesis (H_2)

The working hypothesis (H_2) which were tested in the experiment is given below:

$H_{2(a)}$: The user interface designed with intelligent features, does affect the speed of data entry.

The paired sample t-test conducted on the data sets revealed (Table 6-4) that there was a statistical significance ($t_{(102)}$= 5.424, p>0.001), between exiting user interface (M=184.48, SD=3.18) compared to intelligent user interface (M=182.59, SD=3.54) on speed of data entry by operator. A statistical difference ($t_{(64)}$= 2.413, p>0.05), between exiting user interface (M=186.98, SD=4.99) compared to intelligent user interface (M=184.72, SD=4.92) on speed of data entry by non-operator.

Table 6-4: Results of hypothesis 2

	Dependent Value	*N*	User Interface				*t*
			Existing UI		Intelligent UI		
			Mean	Sd	Mean	Sd	
Time required	Operator	102	184.48	3.18	182.59	3.54	5.424***
	Non-operator	64	186.98	4.99	184.72	4.92	2.413*

* Significant at .05 level of significance $(P<0.05)$

*** Significant at .001 level of significance $(P<0.001)$

6.2.3. Hypothesis 3

The working hypothesis (H_3) which were tested in the experiment is given below:

$H_{3(a)}$: The user interface designed with intelligent features do effect the variables like- (i) perceived system usability, (ii) perceived cognitive load, (iii) user interface satisfaction, (iv) willingness to continue the usage and (v) relative advantage.

The paired sample t-test conducted on the data sets revealed (Table 6-5) that there was a statistical significance ($t_{(103)}$= 7.209, p>0.001), between exiting user interface (M=35.14, SD=2.93) compared to intelligent user interface (M=32.56, SD=3.64) on cognitive load by operators. A statistical difference ($t_{(108)}$= 6.69, p>0.001), between exiting user interface (M=35.07, SD=3.90) compared to intelligent user interface (M=32.06, SD=4.43) on NASA-TLX by non-operator. A statistical difference ($t_{(103)}$= 9.859, p>0.001), between exiting user interface (M=70.53, SD=6.56) compared to intelligent user interface (M=80.15, SD=8.67) on system usability by operator. A statistical difference ($t_{(108)}$= 8.912, p>0.001), between exiting user interface (M=71.48, SD=6.73) compared to intelligent user interface (M=79.40, SD=8.78) on system usability by non-operator. A statistical difference

($t_{(103)}$= 21.909, p>0.001), between exiting user interface (M=6.53, SD=0.23) compared to intelligent user interface (M=7.30, SD=0.36) on user interface satisfaction by operator. A statistical difference ($t_{(108)}$= 23.571, p>0.001), between exiting user interface (M=6.51, SD=0.20) compared to intelligent user interface (M=7.18, SD=0.33) on user interface satisfaction by non-operator. A statistical difference ($t_{(103)}$= 4.268, p>0.001), between exiting user interface (M=6.16, SD=2.08) compared to intelligent user interface (M=6.84, SD=1.95) on willingness to continue usage by operator. A statistical difference ($t_{(108)}$= 6.864, p>0.001), between exiting user interface (M=4.75, SD=1.70) compared to intelligent user interface (M=5.55, SD=1.60) on willingness to continue usage by non-operator.

Table 6-5: Results of hypothesis 3

	Dependent Value	*N*	User Interface				*t*
			Existing UI		Intelligent UI		
			Mean	Sd	Mean	Sd	
NASA-TLX	**Operator**	103	35.14	2.93	32.56	3.64	7.209***
	Non-operator	108	35.07	3.90	32.06	4.43	6.690***
SUS	**Operator**	103	70.53	6.56	80.15	8.67	9.859***
	Non-operator	108	71.48	6.73	79.40	8.78	8.912***
QUIS	**Operator**	103	6.53	0.23	7.30	0.36	21.909***
	Non-operator	108	6.51	0.20	7.18	0.33	23.571***
WCU	**Operator**	103	6.16	2.08	6.84	1.95	4.268***
	Non-operator	108	4.75	1.70	5.55	1.60	6.864***
RA	**Operator**	103	7.52	1.09	8.72	1.17	7.560***
	Non-operator	108	7.37	0.99	8.79	1.07	10.503***

* Significant at .05 level of significance (P<0.05)

** Significant at .01 level of significance (P<0.01)

*** Significant at .001 level of significance (P<0.001)

6.2.4. Hypothesis 4

The working hypothesis (H_4) which were tested in the experiment is given below:

$H_{4(a)}$: There is significant difference between errors committed by female operator compared to male.

The one-way ANOVA result found no significant difference between female and male in any of the experimental condition given in Table 6-6. Therefore, we fail to reject the null hypothesis (**$H_{4(0)}$**).

Table 6-6: Results of hypothesis 4

	Dependent Value	Gender		*F*
		Female Mean (sd)	Male Mean (sd)	
Intelligent UI	**Regular time**	1.12 (1.015)	1.33 (1.244)	1.212
	Limited time	1.29 (0.931)	1.30 (0.985)	0.001
Existing UI	**Regular time**	1.46 (1.064)	1.61 (1.329)	0.489
	Limited time	1.48 (0.926)	1.62 (1.120)	0.575

6.2.5. Hypothesis 5

The working hypothesis (H_5) which were tested in the experiment is given below:

$H_{5(a)}$: There is effect of language used on error rate in the case of, (i) English language in forms used for data entry, (ii) *Marathi* language and (iii) Mixed language (i.e. Both English and *Marathi* combined).

The one way ANOVA shows (Table 6-7) significant difference ($F_{(2,75)}$=4.49, p>0.05) between English (M=1.54, SD=0.99), *Marathi* (M=1.54, SD=0.86) and mixed (M=2.46, SD=1.79) in error by operator using intelligent UI. A significant difference ($F_{(2,72)}$=3.44, p>0.05) between English (M=2.00, SD=1.07), *Marathi* (M=2.73, SD=1.04) and mixed (M=3.00, SD=2.00) in error by non-operator using intelligent UI. A significant difference ($F_{(2,75)}$=8.72, p>0.001) between English (M=2.35, SD=0.85), *Marathi* (M=1.81, SD=0.94) and mixed (M=3.12, SD=1.51) in error by operator using existing UI.

Table 6-7: Results of hypothesis 5

`	Dependent Value	Task Variation			*F*
		English Mean (sd)	*Marathi* Mean (sd)	Mixed Mean (sd)	
Intelligent UI	**Operator**	1.54 (0.989)^a	1.54 (0.859)^a	2.46 (1.794)^b	4.489*
	Non-operator	2 (1.074)^a	2.73 (1.041)^a	3 (2)^b	3.437*
Existing UI	**Operator**	2.35 (0.846)^a	1.81 (0.939)^a	3.12 (1.505)^b	8.722***
	Non-operator	2.70 (1.068)	2.96 (1.248)	3.20 (1.824)	0.761

a, b Common superscripts show no significant difference between mean pairs (Post hoc test using Tuckey HSD)

* Significant at .05 level of significance (P<0.05)

** Significant at .01 level of significance (P<0.01)

*** Significant at .001 level of significance (P<0.001)

6.3.Conclusion

This Chapter presented the experimental results of the research investigation being argued in this thesis. The paired t-test analysis has been carried out and the results found significant difference between intelligent UI and existing UI on accuracy and time.

Table 6-8: Hypotheses test results

Hypotheses		**Results**
$H_{1(a)}$	The user interface designed with intelligent features like- (i) display of autocomplete suggestion for text field by ranking strategy based on likelihood, (ii) predictive text entry widget, (iii) radio button pointed with most likely options and (iv) dynamic drop-down split-menu, does affect the accuracy of data entry.	Proved
$H_{2(a)}$	The user interface designed with intelligent features, does affect the speed of data entry.	Proved
$H_{3(a)}$	The user interface designed with intelligent features do effect the variables like- (i) perceived system usability, (ii) perceived cognitive load, (iii) user interface satisfaction, (iv) willingness to continue the usage and (v) relative advantage.	Proved
$H_{4(a)}$	There is significant difference between errors committed by female operator compared to male.	Disproved
$H_{5(a)}$	There is effect of language used on error rate in the case of, (i) English language in forms used for data entry, (ii) *Marathi* language and (iii) Mixed language (i.e. Both English and *Marathi* combined).	Proved

The next Chapter discusses the consolidated findings of this research investigation, highlights the contribution of the research work and presents future scope of the current investigation.

Chapter 7

Discussion

This Chapter discusses the result and analysis done on previous Chapter 6. It also summarises the major findings of this research work and discusses its relation with the theory presented in Chapter 2.

7.1.Introduction

This Chapter discusses the result and analysis done on previous Chapter 6. It also summarises the major findings of this research work and discusses its relation with the theory presented in Chapter 2.

7.2.Discussions

The literature identified different factors which may affect the performance (in terms of errors) of the operator working in rural-BPOs. One of them is, design of user interface used for data entry. We found that there are many issues with existing data entry user interfaces which may cause errors like (a) poor design of user interface: fails to correct specific field constraint, does not provide clues during typing, fails to provide confirmation logic, does not provide validation logic for fields (b) User interface was in the English language: (i) the majority of operators were educated in their mother tongue (local language) and the user interface (UI) used for data entry is completely in the English language (ii) the operators speak in their local language when they are socialising at work. The language used during data entry is English, this means their thinking and conversing language is different therefore working seamlessly between two languages cause them to make errors and also takes extra time during data entry. This is a potential context for errors. (iii) error or feedback message: The operator gets confused when error or feedback message appear in English which take extra time for them to read and understand it and internalize it. Therefore, to dress these issues we have developed and implemented a new intelligent user interface especially for data entry operators in rural Indian context. The experiments were conducted to compare two user interfaces, one is newly designed interface (*ELIIDE* - tool) and second is the existing user interface; to test the number of errors made and time taken to complete given task. The t-test analysis technique was adopted for the analysis of the data.

The results confirm that, there is a statistical significance between exiting user interface compared to intelligent user interface on errors observed during data entry by operator for both time limitations. This proves that the user interface designed with intelligent widgets helps the operator to improve their performance by reducing errors during data entry. This means that the specially designed widgets for data entry in the context of rural-BPO mitigate errors. The result also shows that, the time taken to complete the given data entry take on intelligent user interface was statistically significant compare to existing user interface. This shows that *ELIIDE* –tool is faster than existing UI.

This is because *ELIIDE* –tool does not allow the operator to enter special characters and numerical in text field for example, in fields like '*First Name*', '*Last Name*' etc. where only text is allowed. Also special character and text is not allowed in numerical fields like '*Telephone No*.', '*Mobile No*.' and '*Date of Birth*'. This prevent the operator from making errors if he/she mistakenly presses wrong key. The *ELIIDE* - tool supported with quantitative probability for specific widgets/ fields provides information clue during data entry. This type of practice can help operator to cross validate by referring quantitative probabilistic of the entry and make less errors. Also sometime they ignore the cross validation done by referring actual paper document and hence speed up their performance while data entry and therefore make faster entries. The predictive mechanism of *ELIIDE* could also helped operators to select appropriate predicted entries without error and it also save time of typing. The dynamic split menu design of *ELIIDE* can help the operator to select appropriate entry from most frequently used five items (Miller G. A., 1956) from menu. Also this list items are supported with quantitative probability so as to judge the most relevant item entry without error and less time efforts.

The literature studies have reported that in rural parts of India people have minimum access and familiarity with computers because of illiteracy and spoken language problems, most information systems being in English. The development cost of applications with community partners that meet their local language learning needs, is beyond the budgets of community development projects. In such a scenario the reliability and quality of rural based data entry services may also become questionable in terms of output quality. But, we have noticed that most of the data entry on computers happens in the English language. The operators speak in their local language when they are socialising at work. The language used during data entry is English, this means their thinking and conversing language is different therefore working seamlessly between two languages cause them to make errors and also takes extra time during data entry. This is a potential context for errors. The operator gets confused when error or feedback message appear in English which take extra time for them to read and understand it and internalize it.

Therefore, to overcome these limitations we have designed the errors and feedback messages in their local language (i.e. *Marathi*). These errors and feedback messages are also supported with audio so as to make the operator feel that, *ELIIDE* is communicating with them and may get emotionally attached with it. The *ELIIDE* - tool displays the error messages in the *Marathi* language when the user enters an invalid or no value for particular field.

Female operators are steadily increasing in semi-urban pockets. Issues regarding their performance and pay do exist. During our initial observation we found that in rural-BPO's the number of female operators are slightly more compared to male. Therefore, we are also investigating that, women are more accurate compared to men during data entry.

Apart from the gaps highlighted above, the research literature also notices the fact that sensitive variables like- perceived system usability, perceived cognitive load, user interface satisfaction, willingness to continue usage and relative advantage, have been ignored while designing the interfaces for data entry, on systems being used by BPO organizations. Therefore, in this study we have investigated the effect of these sensitive variable on data entry operators. We could infer that the intelligent user interface and existing interface significantly influences cognitive load, system usability, satisfaction, willingness to continue usage and relative advantages. Therefore, user interface designed with intelligent widgets can decrease cognitive load, increase system usability and satisfaction. Also users are willing to continue using this interface for data entry. It can be relatively better compared to existing interface used in data entry GUIs.

The literature suggests (depicted in Table 1-3) different classification of errors according to their context of study. Oladimeji *et. al.* (2011) have identified six categories of number entry errors (skipped, transposition, wrong digit, missing decimal, missing digit and other error). Error can be classified according to whether they occur at a skill, rule or knowledge-based level (Rasmussen, 1983; 1986); whether they are slips or lapses (automaticity errors) or mistakes (conceptual errors) (Norman, 1983); and according to whether the error occurs at a task, semantic or interactional level (Maran, 1981; Devis, 1983). The text entry by physical keyboard typing has been studied by many researchers (Rumelhart & Norman, 1982; Grudin, 1984). Gentner *et al*., 1984 have found that there is large percentage of typing errors such as substitutions, insertions and omissions. The other errors like transposition error, doubling error, alternation error, homologous error, capture error, phonetic swap; type of errors found in transcription typing.

We will now discuss the finding of this thesis outline with above paragraph which was elaborated in Section 1.3. A comprehensive list of relevant classification of data entry errors found in the context of rural- BPOs in this study is prepared, which is one of the contribution of this thesis work. The items in this list are sorted into two broad groups as- text entry errors and numerical entry errors. The text entry errors are classified into six types as- (i) Mistype/ Spelling/ Incorrect: substitutions and intrusions, (ii) transposition, (iii) doubling, (iv) case, (v) capture, phonetic, misinterpretation and (vi) omission/ wrong field. The numerical entry errors are classified into four types as- wrong, reverse, double and missing. The inference from our evaluated hypothesis such that (i) the intelligent user interface and existing interface significantly influences the error rate. (ii) the user interface designed with intelligent widgets addressing local needs can increase the accuracy during data entry and (iii) the intelligent user interface and existing interface significantly influences the time.

Therefore, intelligent widgets can help to speedy data entry. (iv) the intelligent user interface and existing interface significantly influences cognitive load, system usability, satisfaction, willingness to continue usage and relative advantages. Therefore, user interface designed with intelligent widgets can decrease cognitive load, increase system usability and satisfaction. Also users are willing to continue using this interface for data entry. It can be relatively better compared to existing interface used in data entry GUIs, suggest that in the context Indian rural-BPO worker there are more than one root causes of error like language, work culture, education background. From our inferences we realise that though the errors committed by BPO operator and others fall in the same taxonomy of error reported and research earlier, but there is a difference which we feel a significant, this difference probability due to culture, language, work culture, education background assuming that the training and the data entry infrastructure is same. When we look at errors committed by BPO we feel that they are interconnected and compounded. There could be compounded error for example, Gentner *et al.*, 1984 have found that there is large percentage of typing errors such as substitutions, insertions and omissions. The other errors like transposition error, doubling error, alternation error, homologous error, capture error, phonetic swap; type of errors found in transcription typing.

Though the classification fit with the earlier taxonomy the additional way of adding this compounded causes for this error may not simplify into the taxonomy that was earlier, so we suggest that the taxonomy of error be expanded into a compounded error due to cultural difference or mother tongue or education background.

Chapter 8

Conclusions, Contributions and Future Work

This Chapter summarises the major findings, and core contributions of this research. It also highlights on the limitations, and general implications of the results in practice. The Chapter also elaborates on the future scopes and potential paths of this research.

8.1. Introduction

This Chapter intends to summarise the main research findings and core contributions of this research along with its general implication in practice, limitations and future scope. At first, short overview and conclusion of the research is stated in Section 8.2, consolidated research findings are summarised in Section 8.3.

Section 8.4 presents the core contributions of this research. Limitations and general implications of the research findings and contributions have been elaborated in Section 8.5.

Section 8.6 discusses the future scope and potential directions for future research.

8.2. Conclusion

This research investigation was carried out on noticing the contradictory findings and various questions that remained unanswered in 'data entry error' studies in the context of rural India. The literature survey points out to the existence of conclusive research from the perspective of identifying the effect of GUIs design features on data entry in rural context. It has been argued in the literature that the influence of local language and GUI design that might affect operators' performance in a data entry in rural Indian context had not been found to be addressed earlier.

The data entry error is the focal point around which research investigation has been carried out in this thesis. It is argued that there are several factors, which may affect the performance (in terms of error/accuracy and time/speed) of operators, (a) effect of lower usability factor of software employed for data entry: There is lack of expertise in designing user interfaces for such data entry software, especially failing to address localised specific field constraints that can, if incorporated, ensure high quality of transcription (data entry) with low rate of errors (b) there may be cultural issues / challenges like differences between local spoken language and input language (English) by data entry operators - all of which needs to be investigated (c) the sensitive variables like- perceived system usability, perceived cognitive load, user interface satisfaction, willingness to continue usage and relative advantage, have been ignored while designing the interfaces for data entry, on systems being used by BPO organizations.

Therefore, to address the above challenges we have designed and implemented a data entry user interface (name as *ELIIDE* - tool) supported with intelligent features like- (i) display of autocomplete suggestion for text field by ranking strategy based on likelihood,

(ii) predictive text entry widget, (iii) radio button pointed with most likely options and (iv) dynamic drop-down split-menu. The *ELIIDE* has features like dynamic, predictive, adaptive and probabilistic. The interface uses local *Marathi* language to communicate with user / operator. The communication happens in terms of error and feedback messages. This additional feature may support rural users to get emotionally attached to interface.

The experiments were conducted to compare two user interfaces, one is newly designed interface (*ELIIDE* - tool) and second is the existing user interface the operator uses for data entry. The participants including professional data entry operators working in rural- BPOs volunteered for the study. Prior to the actual experiment, the participants were explained about the design and purpose of user interface and also provided practice session on it. Before going for the actual experiment the participants were told to fill pre-test questionnaires which include demographic information. Each participant performed four tasks, two tasks were data entry on existing interface (having static widgets) and other two were on the intelligent interface (having dynamic widgets). The sequence of the task was random to avoid learning effect. The tasks consist of a transcription of given data entry form (refer Figure 5.2) (also called as paper form) into electronic form using both interfaces. Participants were instructed to perform the tasks as quickly and accurately as possible. The computer based background recording of each participant interaction with the designed user interface have taken for calculation of the accuracy and speed. After completion of the experiment the participants were instructed to fill the post-task questionnaires to express their opinion and experience about the user interface. The subjective experience was recorded in terms of cognitive load, perceived system usability, user interface satisfaction, willingness to continue usage and relative advantage. The t-test and ANOVA analysis technique were adopted for the analysis of the data. Results highlight that intelligent user interface design features do affect the operator's performance in terms of accuracy and speed. It has also been observed that *ELIIDE* -tool can affect operators subjective experience. Consolidated finding of the experiments are discussed below.

8.3.Consolidated Findings of this Research

This study shows effect of intelligent feature used in designing user interface for data entry on speed, accuracy, system usability, cognitive load, satisfaction, willingness to continue and relative advantages. Following are the main research findings of this research.

1) We could infer that the intelligent user interface and existing interface significantly influences the error rate. Therefore, user interface designed with intelligent widgets addressing local needs can increase the accuracy during data entry.
2) Tool can also perform as error training tool and performance evaluation tool.
3) We could infer that the intelligent user interface and existing interface significantly influences the time. Therefore, intelligent widgets can help to speedy data entry.
4) We could infer that the intelligent user interface and existing interface significantly influences cognitive load, system usability, satisfaction, willingness to continue usage and relative advantages. Therefore, user interface designed with intelligent widgets can decrease cognitive load, increase system usability and satisfaction. Also users are willing to continue using this interface for data entry. It can be relatively better compared to existing interface used in data entry GUIs.
5) It is observed that there is no significant difference between female and male in error rate. Both sexes have similar error rates.
6) We could infer that English, *Marathi* language data entry task and mixed (combining both English and *Marathi*) significantly influences the accuracy in three task conditions (i) operator using intelligent UI (ii) non-operator using intelligent UI and (iii) operator using existing UI. The data entry task can influence the accuracy, therefore data entry forms can either be in English or *Marathi* language but not in mixed language.

8.4. Major Research Contributions of this Thesis

This research produces three core contributions to the field of usability and human computer interaction.

1) Design of efficient 'intelligent user interface for data entry' evaluation tool for operators working in rural-BPOs. Main features of this interface tool are:
    a. Display/ provide automatic feedback and error/ warning messages with audio support in *Marathi* language when the operator entered a wrong value: Local Language nuanced prompts that has been incorporated and is a novel approach specific to rural-BPOs in Maharashtra State.
    b. Provides monthly performance index for operator and generation different reports by graph acting as a training tool: local mental model mapped to rate of error and type of errors specific to a rural setting.

Additional features of this interface

c. Dynamic: The interface is designed with dynamic widgets which dynamically updating meaning reordering and highlighting the other likely options.

d. Adaptive: The interface adapts certain features (like- display of audio support as option for expert users) for the operator based on its previous performance.

e. Predictive: The interface provides the predictive text entry widgets.

f. Probabilistic approach

2) Evaluated this interface from actual data entry operators in terms of speed, accuracy, system usability, cognitive load, satisfaction, willingness to continue and relative advantages.
3) The interface uses local language for communication with operators making the interface tool user friendly to operators.
4) A comprehensive list of relevant classification of data entry errors found in the context of rural- BPOs in this study is prepared, which is one of the contribution of this thesis work. The items in this list are sorted into two broad groups as- text entry errors and numerical entry errors. The text entry errors are classified into six types as- (i) Mistype/ Spelling/ Incorrect: substitutions and intrusions, (ii) transposition, (iii) doubling, (iv) case, (v) capture, phonetic, misinterpretation and (vi) omission/ wrong field. The numerical entry errors are classified into four types as- wrong, reverse, double and missing.

8.5. Limitations and Generalisations of this Research

This section attempts to highlight on the limitations and generalisations that can be made from this research.

8.5.1. Limitations of this Research

Following are the major limitations of this research:

This tool was developed for rural-BPOs, only for specific type of data entry work i.e. for digitization of bank information (personal information block only). Also this interface was developed and tested only for *Marathi* language. Interface provides predictive text only for specific fields.

8.6. Scope of Future Research

In future scope, this interface can be generalize for different sectors for digitization of information in organizations like -finance, insurance, banking, government sectors, etc. This interface can also be modified for any rural language in India.

Aspects such as will audio prompts in local language whenever error is made during English language data entry, be acceptable to the operators is another research question that can be taken up in the future.

Appendix 1A

Pre-test Questionnaires

1

उपयुक्तता संशोधन प्रकल्प [Usability Research Project]

उपयुक्त अभियांत्रिकी आणि मानवी संगणक संवाद प्रयोगशाळा, डिझाइन विभाग, भारतीय तंत्रज्ञान संस्था गुवाहाटी [Usability Engineering & Human Computer Interaction Lab, Department of Design, Indian Institute of Technology Guwahati]

प्रश्नावली [Questionnaire]

सदर प्रकल्पात गोळा करण्यात येणारी माहिती शैक्षणिक संशोधन कार्याचा भाग असून हया माहितीची अंमलबजावणी उपयुक्त अभियांत्रिकी आणि मानवी संगणक संवाद प्रयोगशाळा, भारतीय तंत्रज्ञान संस्था गुवाहाटी येथे केली जाणार आहे. सर्व आपली वैयक्तिक माहिती, प्रतिमा आणि व्हिडिओ गोपनीय ठेवण्यात येणार आहे आणि हि माहिती केवळ शैक्षणिक संशोधनासाठी वापरली जाईल. कृपया प्रश्नांची उत्तरे देत असताना आपले विचार आणि भावना मोकळ्या मानाने व्यक्त करा. आपल्या सहकार्याबद्दल धन्यवाद !!!

[This data collection work is the part of an Academic Research work being executed in the Usability Engineering and HCI Lab of Indian Institute of Technology Guwahati. All your personal information, images & videos taken would be kept confidential and the statistically treated responses will be used for academic research purposes only. Please feel free to share your thoughts and feelings while you answer the questions. Thanks you for your co-operation.]

***पार्श्वभूमी**: आम्ही आपला संगणक वापरा संबंधीचा अनुभवयाची माहिती गोळा करत आहोत. तुमचे प्रतिसाद आम्हाला समजून घेण्यास मदत करतील कि कसे नवीन ग्राफिकल युजर इंटरफेस तुमच्या सारख्या युजरसाठी डीसायीन केला पाहिजे जो तुमच्या भाषेत तुमच्याशी संवाद करेन (मातृभाषाचा आवाज सांगतो कि चूक किंवा पुढील अडकलेल्या कारवाई बाबतीत).*

*[**Background**: We are collecting responses to your experience using Computers. The response will help us understand how new graphic user interfaces need to be designed so as to assist a user like you through prompts especially in the language you are comfortable with (say mother tongue voice over prompting you in case you make a mistake or are stuck for the next action)].*

***सहभागी माहिती** [Participant Information]:*

नाव [Name]:______________________ लिंग[Sex]: पुरुष[Male] / महिला[Female]

वय[Age] :______________________ शहर आणि राज्य[City & State]:____________

शिक्षण- सर्वोच्च पदवी /प्रमाणपत्र /पातळी (उदा.राजकारण विज्ञान मध्ये बी.ए.)[Education- Highest degree/ certificate/level (Ex: B.A in Politics Sc.)]:______________________

विशेष / अतिरिक्त पात्रता (निर्दिष्ट करा)[Special / additional qualifications acquired (please specify)]:______________________

आपल्या संस्थेच्या नाव [Name of your organization]:______________________

तुम्ही काय नोकरी करता आणि नोकरीतील स्थान काय आहे?(कृपया निर्दिष्ट करा) [What is your job description & position?(please specify)]:______________________

____

तुम्ही किती काळापासून काम करीत आहेत?(एकूण नोकरी अनुभव) [How long have you been working (Total Job Experience)?]:______________________

तुम्हांला समजणाऱ्या भाषा (मातृभाषा पासून सुरु करा) [Languages known starting with mother tongue]:______________________

संगणक ज्ञान आणि अनुभव [Computer Knowledge and Experience]

तुम्ही किती काळापासून कॉम्पुटर वापरत आहात? [How long have you been using computers?]

More than a year ☐ 6 to 12 months ☐

Less than 6 months		Never used	

आपण किती वेळ कॉम्प्युटरचा वापर करता? [How much time do you use computer?]

Daily(more than 4hrs)(except holidays)		Daily (less than 4hrs)	
Some time in a week		Never use	

आपण कोणत्या प्रकारचा कॉम्पुटर वापरता? खालील तक्ता भरा [What type of computers have you used? Please fill out the following table]:

कॉम्पुटरचा प्रकार [Type of computer]	कधीपासून वापरता [Length of time]	वापराचे उद्धिष्ट [Purposes]
Example: Desktop PC	*1 years*	*Personal use at home, word processing, email*
Example: Laptop	*6 months*	*Email, Web search*
Example: Tablets / Mobile Phone browser		

आपण कॉम्पुटरवर इंटरनेटचा वापर नियमित करता का?

[Do you currently have regular access to a computer having an Internet Connection?]

हो [Yes]		नाही [No]	

पहिल्या वेळेस आपणास कॉम्पुटर कसा वापरायाचा कोणी शिकवले?

Who taught you how to use a computer in the first place?

School/ College		Private Computer Centre	
Friends		Family	
Youself		Other (plase specify)	

आपण आपल्या स्वत: च्या संगणकावर साक्षरतेचे प्रमाण कसे नोंदवल?

How would you rate your own computer literacy?

पुरेसा नाही [Poor]		पुरेसा [Adequate]	
चांगले [Good]		फार चांगले [Very good]	
उत्कृष्ट [Excellent]			

आपण आपल्या स्वत: च्या इंटरनेट साक्षरतेचे प्रमाण कसे नोंदवल?

How would you rate your own Internet literacy?

Poor		Adequate	
Good		Very good	
Excellent			

आपण आपल्या वर्तमान टायपिंग कौशल्याचे प्रमाण कसे नोंदवल?

How would you rate your current typing skills?

Poor		Adequate	
Good		Very good	
Excellent			

आपण कोणत्या भाषेत टायपिंग करता? (आवश्यक असल्यास आपण एकापेक्षा अधिक पर्याय निवडू शकता)

Which language do you use for typing? (can tick multiple choices if required)

English		Mother tongue (Marathi)	
Hindi		Other (plase specify)	

आपण इंग्रजी भाषेत आपल्या वर्तमान टाइपिंग कौशल्याचे प्रमाण कसे नोंदवल?

How would you rate your current typing skills in English language?

Poor		Adequate	
Good		Very good	
Excellent			

जर मातृभाषेतून टाइप करत असाल, आपण आपल्या वर्तमान टाइपिंग कौशल्याचे प्रमाण कसे नोंदवल?

If typing in mother tongue, how would you rate your current typing skills?

Poor ☐ Adequate ☐
Good ☐ Very good ☐
Excellent ☐

कार्यक्षमता (चुका आणि वेळ) [Efficiency (Errors and Time)]

आपण इंग्रजी भाषेत टाइप करताना किती वेळा चुका (त्रुटी) करतात?

How often do you make errors while typing in English language (in data entry)?

खुप वेळा [Very often] ☐ अनेकदा [Often] ☐
कधीकधी [Occasionally] ☐ क्वचितच [Seldom] ☐
कधीही नाही [Never] ☐

तुमच्या सहकारी किंवा मित्राच्या तुलनेत आपण इंग्रजी भाषेत आपल्या वर्तमान टाइपिंग कौशल्याचे प्रमाण कसे नोंदवल?

How would you rate your current typing speed in English language compared to your co-workers / friends?

पुरेसा नाही [Poor] ☐ पुरेसा [Adequate] ☐
चांगले [Good] ☐ फार चांगले [Very good] ☐
उत्कृष्ट [Excellent] ☐

जर मातृभाषेतून टाइप करत असाल, आपण टाइप करताना किती वेळा त्रुटी (चुका) करतात?

If typing in mother tongue, how often do you make errors (in data entry)?

Very often ☐ Often ☐
Occasionally ☐ Seldom ☐
Never ☐

जर मातृभाषेतून टाइप करत असाल, आपण आपल्या वर्तमान टाइपिंग कौशल्याचे प्रमाण कसे नोंदवल?

If typing in mother tongue, how would you rate your current typing speed?

Poor ☐ Adequate ☐
Good ☐ Very good ☐
Excellent ☐

मातृभाषा (उदा. मराठी भाषा) [Mother tongue (Eg: Marathi- language)]

कोणत्या माध्यमातून तुम्ही प्राथमिक शिक्षण पूर्ण केले?

In which medium did you complete primary education?

मराठी [Marathi] ☐ हिंदी [Hindi] ☐
इंग्लिश [English] ☐ इतर [Other (plase specify)] ☐

कोणती भाषा आपण समाजात असतानी बोलणे पसंत करता?

Which language you prefer to speak in society?

मराठी [Marathi] ☐ हिंदी [Hindi] ☐
इंग्लिश [English] ☐ इतर [Other (plase specify)] ☐

कामाच्या ठिकाणी आपले सहकारी / व्यवस्थापकाशी बोलतांना कोणती भाषा तुम्ही बोलता?

Which language do you speak while interacting with your co-worker / manager at work place (or while performing practical)?

Marathi ☐ Hindi ☐
English ☐ Other (plase specify) ☐

कोणती भाषा कोणती भाषा आपण डाटा एन्ट्री दरम्यान वापरता?

Which language do you use during data entry work?

Marathi ☐ Hindi ☐
English ☐ Other (plase specify) ☐

तंत्रज्ञानाचा वापर (मोबाईल फोन) *[Use of technology (mobile phones)]*

आपण वापरत असलेला मोबाईल फोन कोणत्या प्रकारचा आहे?

Which type of mobile phone you are using?

4

केवळ कॉल करण्याची सुविधा असलेले साधे फोन
Simple phone having calling facility only ☐

कॅमेरा, टच स्क्रीन इत्यादी सारख्या सर्व सुविधा असलेले स्मार्ट फोन
Smart phone having all facilities like camera, touch screen, etc. ☐

कोणत्याही प्रकारचा मोबाईल फोन वापरत नाही
No using any mobile phone ☐

जर आपण सामान्य मोबाइल फोन वापरत असाल, तर तुम्हाला त्यातील काय वैशिष्ट्ये आवडत आणि काय आवडत नाही?
If you are using simple mobile phone, what features do you like and dislike about it?
कृपया वर्णन करा [Please describe]____

____

जर आपण स्मार्ट मोबाइल फोन वापरत असाल, तर तुम्हाला त्यातील काय वैशिष्ट्ये आवडत आणि काय आवडत नाही?
If you are using smart phone, what features do you like and dislike about it?
कृपया वर्णन करा [Please describe], ____

____

Appendix 1B

Post-test Questionnaires

1

तुम्ही वापरलेल्या इंटरफेसचा अनुभव उचित बॉक्सवर बाबोबरीची फुली मारून व्यक्त करा
Please tick appropriate boxes after experiencing the Interface shown to you

***प्रणाली उपयुक्तता स्केल* [System Usability Scale (SUS)]**

	Strongly agree	agree	Neither agree nor disagree	disagree	Strongly disagree
मला हा इंटरफेस वारंवार वापरायला आवडेल 1. I think that I would like to use this interface frequently					
मला हा इंटरफेस विनाकारण जटिल आढळला 2. I found the interface unnecessarily complex.					
मला हा इंटरफेस वापरण्यास सोपा वाटत होता 3. I thought the interface was easy to use.					
हा इंटरफेस वापरतांनी मला सक्षम तांत्रिक व्यक्तीच्या साहाय्याची गरज वाटते 4. I think that I would need the support of a technical person to be able to use this interface.					
मला या इंटरफेस मध्ये विविध कार्य एकात्मिक मिळाले 5. I found the various functions in this interface were well integrated.					
माझ्या मते इंटरफेस मध्ये खूप विसंगती होत्या 6. I thought there was too much inconsistency in this interface.					
मी अशी कल्पना करतो कि बहुतेक लोक या इंटरफेस वापरणे लवकर शिकू शकतील 7. I would imagine that most people would learn to use this interface very quickly.					
मला हा इंटरफेस वापरतांनी अवजड आढळला 8. I found the interface very cumbersome to use.					
मला हा इंटरफेस वापर करून खूप विश्वास वाटला 9. I felt very confident using the interface.					
मला हा इंटरफेस वापरण्यापूर्वी खूप काही जाणून घ्यावे लागेल 10. I needed to learn a lot of things before I could get going with this interface.					

युजर इंटरफेस समाधान प्रश्नावली* *[Questionnaire for User Interface Satisfaction (QUIS)]

सॉफ्टवेअर बाबतची एकूण प्रतिक्रिया *[Overall reaction to the software]*		*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*		*NA*
1	भयंकर [terrible]											विस्मयकारक [wonderful]	
2	कठीण [difficult]											सोपे [easy]	
3	डोकेदुखी [frustrating]											समाधानकारक [satisfying]	
4	अपुरी शक्ती											पुरेशी शक्ती	

2

		[inadequate power]											[adequate power]	
5		कंटाळवाणे [dull]											उत्तेजक [stimulating]	
6		ताठ [rigid]											लवचिक [flexible]	
Screen														
7	स्क्रीन वर शब्द वाचन [Reading characters on the screen]	हार्ड [hard]											सोपे [easy]	
8	ठळक सुलभ कार्य [Highlighting simplifies task]	not at all											very much	
9	माहितीची मांडणी [Organization of information]	confusing											very clear	
10	स्क्रीनची क्रमवारी [Sequence of screens]	confusing											very clear	
परिभाषा आणि प्रणाली माहिती ***[Terminology and system information]***														
11	संपूर्ण इंटरफेस मध्ये परिभाषेचा वापर [Use of terms throughout interface]	विसंगत inconsistent											सातत्य पूर्ण consistent	
12	परिभाषा कार्य संबंधित Terminology related to task	never											always	
13	स्क्रीन वर संदेश स्थिती [Position of messages on screen]	inconsistent											सातत्य पूर्ण consistent	
14	इनपुट साठी संकेत [Prompts for input]	confusing											clear	
15	संगणक त्याची प्रगती बद्दल माहिती [Computer informs about its progress]	never											always	
16	त्रुटी संदेश [Error messages]	unhelpful											helpful	
शिक्षण [Learning]														
17	इंटरफेस ऑपरेट करणे [Learning to operate the interface]	difficult											easy	
18	चाचणी आणि चूक करून नवीन वैशिष्ट्ये एक्सप्लोर करणे	difficult											easy	

	[Exploring new features by trial and error]													
19	नावांची आठवण आणि आदेशांचा वापर [Remembering names and use of commands]	difficult											easy	
20	कामे करत सोपे आहे [Performing tasks is straightforward]	never											always	
21	स्क्रीनवर मदत संदेश [Help messages on the screen]	unhelpful											helpful	
22	पुरवणी संदर्भ साहित्य [Supplemental reference materials]	confusing											clear	
***इंटरफेस क्षमता* [Interface capabilities]**														
23	इंटरफेस गती [Interface speed]	too slow											fast enough	
24	इंटरफेस विश्वसनीयता Interface reliability	unreliable											reliable	
25	इंटरफेस असल्याचे Interface tends to be	गोंगाट करणा noisy											quiet	
26	आपल्या चुका सुधारत [Correcting your mistakes]	difficult											easy	
27	सर्व प्रकारच्या युजरसाठी डिझाइन Designed for all levels of users	never											always	

***संबंधित फायदा प्रमाणात* [Relative advantage scale]**

या इंटरफेसच्या माध्यमातून डाटा एंट्री करने विद्यमान इंटरफेस पेक्षा अधिक प्रभावी आहे 1. Doing data entry through this interface is more effective than existing interface.	Strongly agree	agree	Neither agree nor disagree	disagree	Strongly disagree
हा इंटरफेस मी आतापर्यंत वापरलेल्या इतर विद्यमान इंटरफेस पेक्षा चांगला आहे 2. This interface is better than other existing interface I have had so far.	Strongly agree	agree	Neither agree nor disagree	disagree	Strongly disagree

4

***वापर सुरू ठेवण्यासाठीची इच्छा दर्शवणारी स्केल** [Willingness to Continue to Use Scale]*	
भविष्यात ह्या सॉफ्टवेअरचा (इंटरफेचा) वापर सुरू ठेवण्यासाठी आपली इच्छा रेट करा Please rate your willingness to continue working with the software in future Examinations	1 2 3 4 5 6 7 8 9 10 Lowest Highest

***नासाचे कार्य लोड निर्देशांक** [NASA Task Load Index]*

मानसिक त्रास Mental Demand	कार्य कसे मानसिक ताणाचे होते? How mentally demanding was the task?

Very Low खुप कमी — खुप जास्त Very High

शारीरिक त्रास Physical Demand	कार्यात शारीरिक कष्टाचे प्रमाण किती वाटले? How physically demanding was the task?

Very Low — Very High

ऐहिक त्रास Temporal Demand	कार्याच्या वेळेत घाई किंवा भराभर पणा किती वाटला? How hurried or rushed was the pace of the task?

Very Low — Very High

कामगिरी Performance	आपणास जे काम करण्यास सांगितले होते ते यशस्वीरित्या कसे साध्य केले? How successful were you in accomplishing what you were asked to do?

Very Low — Very High

प्रयत्न Effort	आपल्या कामगिरीचा स्तर पूर्ण करण्यासाठी तुम्हाला ह्या कामात किती कठीण परिश्रम घ्यावे लागले? How hard did you have to work to accomplish your level of performance?

Very Low — Very High

निराशा Frustration	कामात तुम्हाला असुरक्षित, निराश, संतप्त, ताण आणि चिडवल्यासारखे जाणवले का? How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low — Very High

5

इंटरफेबद्दलच्या तुम्हाला आवडलेल्या ३ ते ५ गोष्टी सांगा. [List three to five things you like the most about the interface] 1.____________________ 2.____________________ 3.____________________ 4.____________________ 5.____________________
आपण सामान्यपणे वापर असलेला इंटरफेस मध्ये कोणतीही सुधारणा करण्यासाठी सूचना सांगा. [Any suggestion for improvement in the interface you normally use]. ____________________ ____________________ ____________________ ____________________

Appendix 2A

Data Entry Forms: English Language

बचत खाते उघडण्याचा फॉर्म

SAVING BANK ACCOUT OPENING FORM

वैयक्तिक माहिती [Personal Details]	
नाव (श्री / श्रीमती) First Name (Mr./ Mrs.)	Ganesh
आडनाव Surname	Sutar
वडिलांचे / पतीचे / पालकाचे नाव Father/ Husband/ Guardian Name	D.
निवासी पत्ता Residential Address	'Landscape' Building
घर क्रमांक आणि नाव House No. and Name	Flat.no.229, 2nd Floor
रस्ता क्रमांक आणि नाव Street No. and Name	
खूण Landmark	
तालुका / तहसील Taluka/ Tehsil	
गाव / शहर Village/ City	Katraj
जिल्हा District	Pune
राज्य State	Maharashtra
पिन कोड PIN Code	411046
फोन क्रमांक (एसटीडी कोड सहित) Telephone No.(with STD code)	2024202118
मोबाइल क्रमांक Mobile No.	9822046183
पॅन क्रमांक PAN Number	
जन्म तारीख Date of Birth (DD/MM/YYYY)	23/9/1975
लिंग Gender/ Sex	Male.

Appendix 2B

Data Entry Forms: *Marathi* Language

बचत खाते उघडण्याचा फॉर्म

SAVING BANK ACCOUT OPENING FORM

वैयक्तिक माहिती [Personal Details]	
नाव (श्री / श्रीमती) First Name (Mr./ Mrs.)	प्रमिला-
आडनाव Surname	आडसुल
वडिलांचे / पतीचे / पालकाचे नाव Father/ Husband/ Guardian Name	डी
निवासी पत्ता Residential Address	संकेत प्राझा
घर क्रमांक आणि नाव House No. and Name	फ्लॅट नं. १११, तळ मजला
रस्ता क्रमांक आणि नाव Street No. and Name	
खूण Landmark	
तालुका / तहसील Taluka/ Tehsil	
गाव / शहर Village/ City	खडकी
जिल्हा District	पुणे
राज्य State	महाराष्ट्र
पिन कोड PIN Code	४११००३
फोन क्रमांक (एसटीडी कोड सहित) Telephone No.(with STD code)	२०२४५६७९२२
मोबाइल क्रमांक Mobile No.	९८९०४८६९०८९
पॅन क्रमांक PAN Number	
जन्म तारीख Date of Birth (DD/MM/YYYY)	१/१०/१९४९
लिंग Gender/ Sex	स्त्री.

Appendix 2C

Data Entry Forms: Mixed (English and *Marathi* both) Language

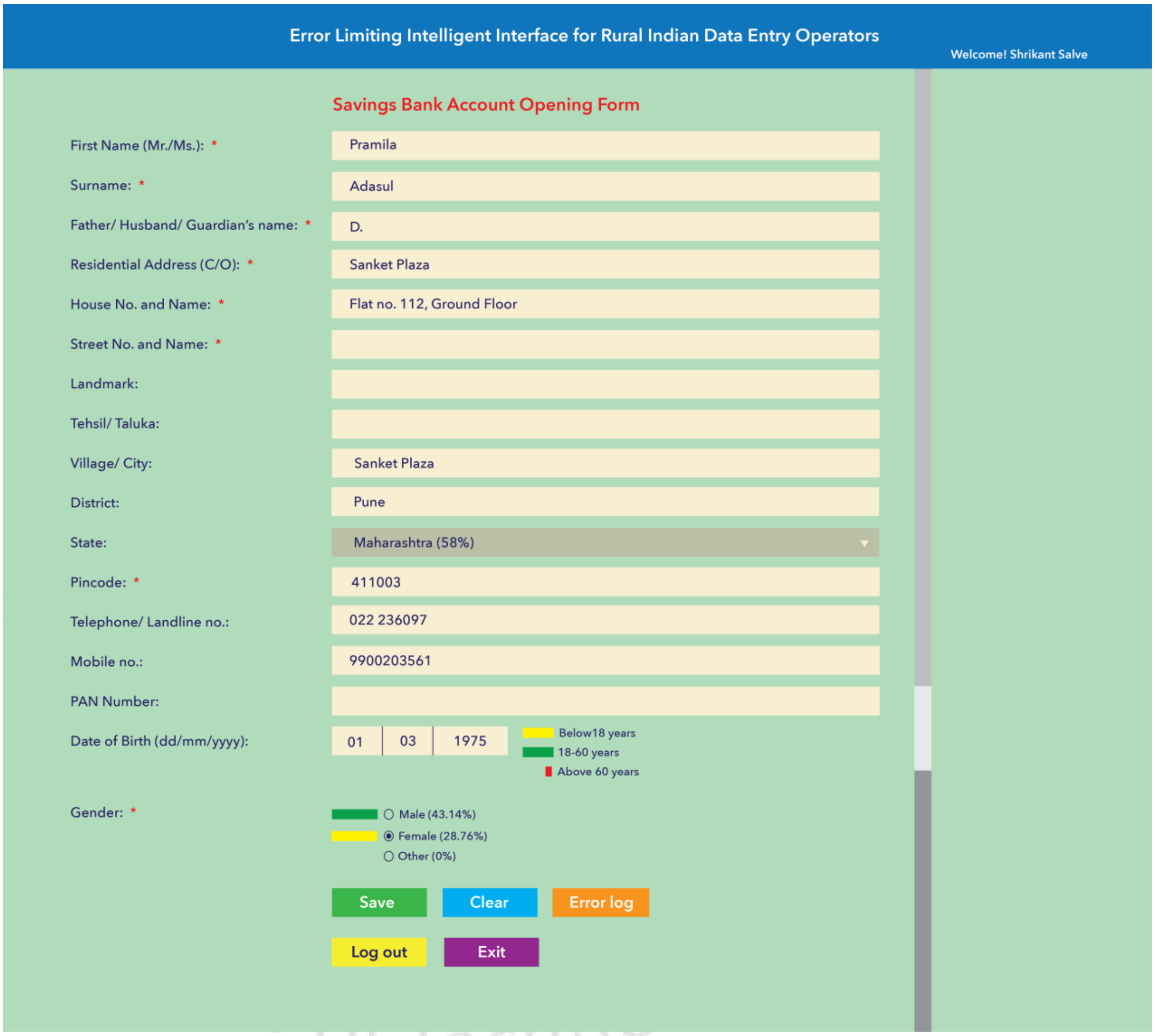
बचत खाते उघडण्याचा फॉर्म

SAVING BANK ACCOUT OPENING FORM

वैयक्तिक माहिती [Personal Details]	
नाव (श्री / श्रीमती) First Name (Mr./ Mrs.)	चंद्रजीत
आडनाव Surname	पाटील
वडिलांचे / पतीचे / पालकाचे नाव Father/ Husband/ Guardian Name	म.
निवासी पत्ता Residential Address	~~मातोश्री~~ मातोश्री निवास
घर क्रमांक आणि नाव House No. and Name	विर- गणपती शेजारी
रस्ता क्रमांक आणि नाव Street No. and Name	
खूण Landmark	
तालुका / तहसील Taluka/ Tehsil	
गाव / शहर Village/ City	कोथरुड
जिल्हा District	पुणे
राज्य State	महाराष्ट्र
पिन कोड PIN Code	411038
फोन क्रमांक (एसटीडी कोड सहित) Telephone No.(with STD code)	
मोबाइल क्रमांक Mobile No.	9890 931414
पॅन क्रमांक PAN Number	AQIPP3427E
जन्म तारीख Date of Birth (DD/MM/YYYY)	20/4/1994
लिंग Gender/ Sex	Male

Appendix 3A

Visually and graphically improved concept version of *ELIIDE*- tool shown below



Bibliography

A

Ali, A., & Meek, C. (2009). *Predictive models of form filling.* Microsoft Reserach, Technical Report. MSR-TR-2009-1.

B

Baber, C., & Stanton, N. A. (1994). Task analysis for error identification: a methodology for designing error tolerant consumer products. *Ergonomics*, 37(11), 1923–1941.

BarahaPad- http://www.baraha.com. (n.d.).

Barchard, A., & Pace, L. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behaviour*, pp 1834-1839.

Bickmore, T. W., & Picard, R. W. (2004). Towards caring machines. *In CHI'04 extended abstracts on Human factors in computing systems* (pp. pp. 1489-1492). ACM.

Black, N. T., & Ertel, W. (2011). *Introduction to artificial intelligence.* Springer Science & Business Media.

Booth, P. A. (December 1991). Errors and theory in human-computer interaction. *Acta Psychologica*, Volume 78, Issues 1–3, Pages 69-96.

Boring, R. L., Griffith, C. D., & Joe, J. C. (2007). The Measure of human error: Direct and indirect performance shaping factors. *IEEE 8th Human Factors and Power Plants and HPRCT 13th Annual Meeting* (pp. pp. 170-176). Monterey, CA, USA: IEEE.

Bradley, M., & Lang, P. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behaviour therapy and experimental psychiatry*, 25, 1, pp 49-59.

Broeck, J. V., Mackay, M., Mpontshane, N., Luabeya, A. K., Chhagan, M. L., & Bennish, M. (2007). Maintaining data integrity in a rural clinical trial. *Clinical Trials*, 4: 572–582.

Brooke, J. (1996). SUS: A “quick and dirty” usability scale. In *Usability evaluation in industry* (pp. pp. 189-194). Taylor & Francis.

Busse, D. (1999). On Human Error and Accident Causation. *Interact.*

Byrne, M. D., & Bovair, S. (1997). A Working Memory Model of a Common Procedural Error. *Cognitive Science*, 21: 31–61.

C

Cairns, P., Pandab, P., & Power, C. (2014). The Influence of Emotion on Number Entry Errors. *In Proceedings of the 32nd annual ACM conference on Human factors in computing systems (CHI '14)* (pp. 2293-2296). New York, NY, USA: ACM.

Calculate Levenshtein Distance. (2013, October 10). Retrieved from http://www.unit-conversion.info/texttools/levenshtein-distance

Card, S. K., Newell, A., & Moran, T. P. (1983). *The Psychology of Human-Computer Interaction.* NJ, USA: L. Erlbaum Assoc. Inc., Hillsdale.

Causse, M., Dehais, F., Péran, P., Sabatini, U., & Pastor, J. (2013). The effects of emotion on pilot decision making: A neuroergonomic approach to aviation safety. *Transportation Research Part C: Emerging Technologies*, pp 272-281.

Chand, A. (2002). Designing for the Indian rural population: Interaction design challenges. *Proc. Development by Design Conference.*

Chand, A., & Dey, A. (2006). Jadoo: a Paper User Interface for Users Unfamiliar with Computers. *Computer-human Interaction (CHI) 2006* (pp. pp. 1625-1630). Proc. CHI 2006.

Chapanis, A. (1999). *The Chapanis Chronicles: 50 years of human factors research, education, and design.* Santa Barbara, California: Aegean.

Chen, K., Chen, H., Conway, N., Hellerstein, J. M., & Parikh, T. S. (2011). Usher: Improving data quality with dynamic forms. *IEEE Transactions on Knowledge and Data Engineering*, 1138-1153.

Chen, K., Hellerstein, J. M., & Parikh., T. S. (2010). Designing adaptive feedback for improving data entry accuracy. *In Proceedings of the 23nd annual ACM symposium on User interface software and technology (UIST '10)* (pp. 239-248). New York, NY, USA: ACM.

Chepenik, L. G., Cornew, L. A., & Farah, M. J. (2007). The influence of sad mood on cognition. *Emotion*, Vol 7(4), 802-811.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '88)* (pp. 213-218). New York, NY, USA: ACM.

Clore, G., & Storbeck, J. (2006). Affect as s information about liking, efficacy, and importance . *In J. Forgas (Ed). Hearts and Minds: Affective influences on social cognition and behaviour. New York: Psychology Press.*

Coombes, S., Janelle, C., & Duley, A. (2005). Emotion and Motor Control: Movement Attributes Following Affective Picture Processing. *Journal of Motor Behavior*, 37, 425-436.

Coombes, S., Tandonnet, C., Fujiyama, H., Janelle, C., Cauraugh, J., & Summers, J. (2009). Emotion and motor preparation: A transcranial magnetic stimulation study of corticospinal motor tract excitability. *Cognitive, Affective, & Behavioral Neuroscience*, 9, 380-388.

D

Dix, A., Finlay, J. E., Abowd, G. D., & Beale, R. (2003). *Human-Computer Interaction.* Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Dusch, B., Crilly, N., & Moultrie, J. (2011). *Developing a Framework for Mapping Sustainable Design Activities.*

E

Edgell, S. E., & Noon, S. M. (1984). Effect of violation of normality on the t test of the correlation coefficient. *Psychological Bulletin*, 95(3), 576.

Ernst, & Young. (January 2011). *Rural Shoring – Upcoming Avenues for ITeS.* Techbytes-Issue 7, EY technology newsletter.

F

Forgas, J. (2013). Don’t worry, Be Sad! On the Cognitive, Motivational, and Interpersonal Benefits of Negative Mood. *Current Direction in Psychological Science*, 22, 225-232.

G

Gentner, D. R., Grudin, J. T., Larochelle, S., Norman, D. A., & Rumelhart, D. E. (1984). A glossary of terms including a classification of typing errors. *In W. E. Cooper (Ed.), Cognitive aspects of skilled typewriting. New York: Springer-Verlag*, 39–43.

Ghosh, S., Samanta, D., & Sarma, M. (2012). Cost of Error Correction Quantification with Bengali Text Transcription. *4th International Conference on Intelligent Human Computer Interaction (IHCI)* (pp. pp.1-6). IEEE.

Gore, K., Lobo, S., & Doke, P. (2012). GappaGoshtiTM:Digital inclusion for rural mass. *Fourth International Conference on Communication Systems and Networks (COMSNETS)*, (pp. pp.1,6).

Gray, J. R. (2001). Emotional modulation of cognitive control: Approach–withdrawal states double-dissociate spatial from verbal two-back task performance. *Journal of Experimental Psychology: General*, Vol 130(3), 436-452.

Grisedale, S., Graves, M., & Grünsteidl, A. (1997). Designing a graphical user interface for healthcare workers in rural India. *In Proceedings of the ACM SIGCHI Conference on Human factors in computing systems (CHI '97)* (pp. 471-478). New York, NY, USA,: ACM.

Grudin, J. T. (1984). Error patterns in skilled and novice transcription typing. *In W. E. Cooper (Ed.), Cognitive aspects of skilled typewriting. New York: Springer-Verlag.*, 121–143.

H

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload*, 1(3), 139-183.

Hermens, L. A., & Schlimmer, J. C. (1994). A machine-learning apprentice for the completion of repetitive forms. *IEEE Expert: Intelligent Systems and Their Applications*.

Hewett, T. T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., . . . Verplank, W. (1992). *ACM SIGCHI curricula for human-computer interaction.* ACM.

http://muzaffarnagar.nic.in/admin.htm. (2016, October). Retrieved from http://muzaffarnagar.nic.in/admin.htm: http://muzaffarnagar.nic.in/admin.htm

http://www.censusindia.gov.in. (2016, October). Retrieved from http://www.censusindia.gov.in: http://www.censusindia.gov.in/Tables_Published/Admin_Units/Admin_links/subdistrict_nomeclature.html

https://www.tin-nsdl.com/. (2016, October). Retrieved from https://www.tin-nsdl.com/: https://www.tin-nsdl.com/

J

Jeon, M., Yim, J.-B., & Walker, B. (2011). An angry driver is not the same as a fearful driver: effects of specific negative emotions on risk perception, driving performance, and workload. *In Proc. of AutomotiveUI '11* (pp. pp 137-142). USA: ACM.

K

Kam, M., Kumar, A., Jain, S., Mathur, A., & Canny, J. (2009). Improving literacy in rural India: cellphone games in an after-school program. *In Proceedings of the 3rd international conference on Information and communication technologies and development (ICTD'09)* (pp. 139-149). Piscataway, NJ, USA: IEEE Press.

Kleinman, K. (2001). Adaptive double data entry: a probabilistic tool for choosing which forms to reenter. *Controlled Clinical Trials*.

L

Lang, K. L., Graesser, A. C., & Hemphill, D. D. (1991). Understanding Errors in Human-Computer Interaction. *SIGCHI Bull. 23*, 43-45.

Lazonder, A. W., & Meij, H. V. (1994). The effect of error-information in tutorial documentation. *Interacting with computers*, 6 (1). pp. 23-40. ISSN 0953-5438.

Lee, D., & Tsatsoulis, C. (2005). Intelligent data entry assistant for XML using ensemble learning. *Proc. ACM 10th Int'l Conf. Intelligent User Interfaces (IUI).* ACM.

M

MacKenzie, S., & Tanaka-Ishii, K. (2007). *Text Entry Systems: Mobility, Accessibility, Universality.* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Madduri, V. R., Gupta, M., De, P., & Anand, V. (2010). Towards mitigating human errors in it change management process. *In Service-Oriented Computing. Springer Berlin Heidelberg.*, pp. 657-662.

Meganathan, R. (2009, October). *English language education in rural schools of India: the situation, the policy and the curriculum*. Retrieved from BBC World Service: http://www.teachingenglish.org.uk/blogs/rama-meganathan/english-language-education-rural-schools-india-situation-policy-curriculum

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review,* Vol 63(2), 81-97 .

Mitchell, J., & Shneiderman, B. (1989). Dynamic versus static menus: an exploratory comparison. *SIGCHI Bulletin*, 33-37.

Moore, G. C., & Benbasat, I. (1991). Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. *Information Systems Research*, 192-222.

N

Nath, P., & Pradhan, R. (2012). Does Feeling Happy Contributes to Flexible Thinking: Exploring the Association Between Positive Emotions and Cognitive Flexibility. *Psychological Studies*, 59, 180-190.

Neerincx, M., & Streefkerk, J. (2003). Interaction in desktop and mobile context: emotions, trust and task performance. *Ambient intelligence: EUSAI 2003* (pp. pp 119-132). Lecture notes in computer science, Berlin: Springer.

Nielsen, J. (1993). Chapter 2 - What Is Usability?, In Usability Engineering. In *Usability Engineering* (pp. 23-48). San Diego: Morgan Kaufmann.

Nielsen, J. (2012, January 4). *usability-101-introduction-to-usability*. Retrieved from www.nngroup.com: https://www.nngroup.com/articles/usability-101-introduction-to-usability/

Norman, D. A. (1983). *Position paper on human error.* Italy: In NATO Advanced Research Workshop on Human Error.

Norman, D. A. (1988). *The psychology of everyday things.* New Tork: Basic Books.

O

Oladimeji, P., Thimbleby, H., & Cox, A. (2011). Number entry interface and their effects of error detection. *Human-Computer Interaction (INTERACT 2011)* (pp. pp 178-185). Springer Berlin Heidelberg.

P

Parikh, T., Ghosh, K., & Chavan, A. (2002). Design studies for a financial management system for micro-credit groups in rural india. *In Proceedings of the 2003 conference on Universal usability (CUU '03)* (pp. 15-22). New York, NY, USA: ACM.

Patel, N. A. (2008). Experiences designing a voice interaction for rural India. *Spoken Language Technology Workshop (SLT 2008)* (pp. pp 21-24). 15 Patel, N., Agarwal, S., Rajput, N., Nanavati, A., Dave, P. and Parikh T.S., "Experiences designing a voice IEEE.

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. (pp. 5. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 51-54). ACM.

Patnaik, S., Brunskill, E., & Thies, W. (April 2009). Evaluating the accuracy of data collection on mobile phones: A study of forms, SMS, and voice. *In International Conference on Information and Communication Technologies and Development (ICTD)* (pp. pp. 74-84). IEEE.

Poelmans, S., Wessa, P., Milis, K., Bloemen, E., & Doom, C. (2008). Usability and acceptance of e-learning in statistics education, based on the compendium platform. *In International Conference of Education, Research and Innovation (ICERI 2008)*, (pp. pp. 1-10).

R

Rasmussen, J. (1983). Skills, rules, and knowledge: signals, signs, and symbols and other distinctions in human performance model. *IEEE Transactions: Systems. Men & Cybernetics*, SMC-13, 257-267.

Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering.* New York, NY, USA: Elsevier Science Inc.

Ravi, V. V., & Venkatrama Raju, D. (2013). Rural Business Process Outsourcing in India-Opportunities and Challenges. *International Journal of Business and Management Invention ISSN (Online): 2319-8028, ISSN (Print): 2319-801X*, Volume 2, Issue 8, PP 40-49.

Reason, J. (1990). *Human Error.* Cambridge Press.

Rooney, J., Heuvel, L., & Lorenzo, D. (2002). Reduce Human Error. In *Quality Progress Handbook* (pp. pp 27-36). www.asq.org.

Rumelhart, D. E., & Norman, D. A. (1982). Simulating a Skilled Typist: A Study of Skilled Cognitive-Motor Performance. *Cognitive Science*, 1-36.

Russell, S., & Norvig, P. (2003). *Artificial Intelligence a modern approach* . Prentice-Hall, Egnlewood Cliffs, 25, 27.

S

Salthouse, T. A. (1986). Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological bulletin,* 99(3), 303.

Sarma, M., & Sarma, K. (2012). Segmentation of Assamese Phonemes using SOM. *In Emerging Trends and Applications in Computer Science (NCETACS), 3rd National Conference on* (pp. pp. 121-125). IEEE.

Sauro, J. (2012, May 15). *Measuring Errors In The User Experience*. Retrieved from Measuring U: http://www.measuringu.com/blog/errors-ux.php

Schlimmer, J. C., & Wells, P. C. (1996). Quantitative results comparing three intelligent interfaces for information capture: a case study adding name information into an electronic personal organizer. *Journal of Artificial Intelligence Research,* 329-349.

Sears, A., & Jacko, J. A. (2009). *Human-Computer Interaction Fundamentals.* CRC Press.

Sears, A., & Shneiderman, B. (1994). Split menus: effectively using selection frequency to organize menus. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 1(1), 27-51.

Senders, J. W., & Moray, N. P. (1991). *Human Error.* NJ: LEA: Hillsdale.

Shneiderman, B., & Plaisant, C. (1987). *Designing the user interface: Strategies for effective human-computer interaction.* Addison-Wesley. 672p.

Shwartz, L., Rosu, D., Loewenstern, D., Buco, M., Guo, S., Lavrado, R., . . . and Singh, J. (2010). Quality of IT service delivery - Analysis and framework for human error prevention. *Service-Oriented Computing and Applications (SOCA)* (pp. 1-8). IEEE.

Singh, G., Findlater, L., Toyama, K., Helmer, S., Gandhi, R., & Balakrishnan, R. (2009). Numeric paper forms for NGOs . *International Conference on Information and Communication Technologies and Development (ICTD 2009).* IEEE.

Smelcer, J. B. (1995). User errors in database query composition, , , http://dx.doi.org/10.1006/ijhc.1995.1017. *International Journal of Human-Computer Studies*, Volume 42, Issue 4, Pages 353-381, ISSN 1071-5819.

Smith, A., Joshi, A., Liu, Z., Bannon, L., Gulliksen, J., & Li, C. (2007). Institutionalizing HCI in Asia. *In Proceedings of the 11th IFIP TC 13 international conference on Human-computer interaction* (p. Volume Part II). (INTERACT'07).

Soukoreff, R., & MacKenzie, I. (2001). Measuring Errors in Text Entry Tasks: An Application of the Levenshtein String Distance Statistic, . . ACM, . *Extended Abstracts of CHI 2001* (pp. pp. 319-320). New York: ACM.

Spering, M., Wagener, G., & Funke, J. (2005). The role of Emotions in complex problem solving. *Cognition and Emotion*, 19(8), pp 1252-1261.

Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C., & Jenkins, D. P. (2006). *Human Factors Methods: A Practical Guide for Engineering and Design.* Brookfield, VT, USA: Ashgate Publishing Company.

T

Thimbleby, H., & Cairns, P. (2010). Reducing number entry errors: Solving a widespread, serious problem. *Journal of the Royal Society Interface*, pp 1429-1439.

W

Walia, W., & Carver, J. (2009). A systematic literature review to identify and classify software requirement errors. *Journal of Information and Software Technology*, 1087–1109.

Warren, J. R. (1996). Intelligent data entry for physicians by machine learning of an anticipative task model. *Australian and New Zealand Con-ference on Intelligent Information Systems* (pp. pp.64-67). doi: 10.1109/ANZIIS.1996.573890.

Warren, J. R., & Bolton, P. (1999). Intelligent split menus for data entry: a simulation study in general practice medicine. *Proceedings of the AMIA Symposium*, (pp. 450-454).

Weyers, B., Burkolter, D., Kluge, A., & Luther, W. (2010). User-Centered Interface Reconfiguration for Error Reduction in Human-Computer Interaction. *Advances* *in Human-Oriented and Personalized Mechanisms, Technologies and Services (CENTRIC), Third International Conference on Digital Object*, (pp. 52-55).

Wickens, C. D. (1992). *Engineering psychology and human performance.* New York: Harper Collins.

Wiseman, S., Cairns, P., & Cox, A. (2011). A taxonomy of number entry error. *HCI 2011: The 25th BCS Conference on Human-Computer Interaction.*

World Bank. (2013, May 26). *Rural India Population*. Retrieved from The World Bank: http://go.worldbank.org/8EFXZBL3Y0

Z

Zhu, Y., Tian, X., Wu, G., Gasso, G., Wangl, & Canu, S. (2013). Emotional Influence on SSVEP based BCI. *In: Proceedings of ACII'13*, (pp. 859-864).

List of Publication resulting out of the research work reported in this thesis

1. Shrikant Salve and Pradeep Yammiyavar. (2013) Influence of local 'language' in data entry errors: A pilot study in the rural Indian setting. Human Computer Interactions (ICHCI), 2013 International Conference on, vol., no., pp.1,4. [Online] Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6887812&tag=1

2. Shrikant Salve and Pradeep Yammiyavar. (2014) "Towards proposing an intelligent error limiting User Interface for rural Indian data entry operators", Australian Journal of Intelligent Information Processing Systems, 13(4). Retrieved February 4, 2014, from http://cs.anu.edu.au/ojs/index.php/ajiips/article/view/1255

3. Shrikant Salve and Pradeep Yammiyavar. (2014). "A study on efficiency of input devices on native language during numerical data entry", HWWE'14, McGraw Hill Education, ISBN (13): 978-93-392-1970-3, ISBN (10): 93-392-1970-8.

4. Shrikant Salve, Shanu Shukla and Pradeep Yammiyavar. (2015). Affect Component and Errors During Numerical Data Entry-A Study. In ICoRD'15–Research into Design Across Boundaries Volume 1 (pp. 573-583). Springer India. [Online] Available: http://dx.doi.org/10.1007/978-81-322-2232-3_50

5. Shrikant Salve and Pradeep Yammiyavar. (2015). "Trade-off between time and error during numerical data entry by rural / semi-urban Indian users", National Conference on Modeling, Optimization and Control, "NCMOC - 2015", 4th-6th March 2015, Organized by Vishwakarma Institute of Technology, Pune.

6. Shrikant Salve and Pradeep Yammiyavar, "Quantitative Probabilistic Widgets as a Method to Improve Usability Performance of Data Entry Tasks", presented as International Conference on Humanizing Work and Work Environment HWWE 2015 at IIT Bombay on 6-9 Dec.2015

7. Shanu Shukla, Shrikant Salve, Sanjram Premjit K. & Pradeep Yammiyavar. "Does Emotion Modulation Influences Speed-Accuracy Trade-off in Numerical Data Entry Task?" Submitted for Journal of Computational Cognitive Science, under review.

8. Shrikant Salve and Pradeep Yammiyavar, "Can Dynamic Widgets Improve Data Entry Efficiency?", Submitted to ICoRD-2017, under review.

9. Shrikant Salve and Pradeep Yammiyavar, "An intelligent GUI tool for the use of rural Indian data entry operators for training in error reduction", Submitted to Journal of Human-Computer Interaction, under review process.