

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

**Usability Engineering of Interactive Voice Response (IVR)
Systems in Oral Users of Southern Africa.**

THESIS

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

In the Department of Electrical Engineering

Of

University of Cape Town

By

Tembaletu Jama Ndwe

February 2011

Abstract

This research study focuses on the feasibility of using the telephone as a tool for information access in the oral communities of Southern Africa. The OpenPhone and BGR systems are used as case studies and their designs have been influenced by field studies with the targeted users. The OpenPhone project aims to design an Interactive Voice Response (IVR) health information system that enables people who are caregivers for HIV/AIDS infected children to access relevant care-giving information by using a telephone in their native language of Setswana in Botswana, Southern Africa. The BGR system allows soccer fans to access results of recently played matches in Premier Soccer League (PSL) of South Africa.

Both designs utilise usability engineering methodology to ascertain that the end products are usable and satisfactory to their targeted users. We hypothesise that ignoring the socio-cultural circumstances of the intended users will lead to an ineffective design. In the present study, the first challenge is to determine which input modality is preferred by the intended users: dual tone multi frequency (DTMF) input or automatic speech recognition (ASR) input. Usability evaluation experiments were conducted with a sample of the intended users and both qualitative and quantitative analyses are used to explore the results.

We have obtained additional evidence on ASR / DTMF preferences in the developing world, and we present the first evidence of the relationship between educational levels and users' interaction behaviour. We have established important findings on language preferences for numbers pronunciation which impacts how IVR systems are designed for users in the region. We also show that HCI for development research can benefit substantially from focussing on "enjoyable" applications as a testing ground for speech technology in the developing world, instead of concentrating only on serious matters such as healthcare, agriculture, public services, etc.

Title of thesis: Usability Engineering of Interactive Voice Response (IVR) Systems in Oral Users of Southern Africa

Name: Tembalethu Jama Ndwe

Address: P.O. Box 395
Pretoria
0001

Date: February 2011

Email: jamakasijadu@gmail.com

Acknowledgements

I would like to thank my supervisors: Professors Etienne Barnard, Mqhele E. Dlodlo and Daniel Mashao for their supervision, support and encouragement.

Thanks to the Human Languages Technologies (HLT) division of the Meraka Institute - CSIR for allowing me to be part of the team.

To various research associates who have contributed to the success of this project.

Thanks to Renee Koen for help with statistical analyses.

Thanks to Bryan McAlister for help with development.

To Jeffrey Nichols for your contribution to the project.

To Peter Clayton for consciously and unconsciously teaching me to write during my previous studies.

To Thato, Thuloane, Karen, Carmen and Gerhard (in Meraka Institute) for your support especially when things were going topsy-turvy.

To the late Baron Peterssen for knowing that there will come a day.

To Liz Pretorius and Ruth de Villiers for your continued support and caring.

uMzinzela nawo onke amaZizi akuthi, adlulayo nakhoyo, ndiyakhahlela.

uQAMATA neLitha lam lokukhanya ngemfesane nothando olungapheliyo.

Declaration of Free Licence

I hereby:

(a) grant the University free license to reproduce the above thesis in whole or in part, for the purpose of research;

(b) declare that:

(i) the above thesis is my own unaided work, both in conception and execution, and that apart from the normal guidance of my supervisor, I have received no assistance apart from that stated below;

(ii) except as stated below, neither the substance or any part of the thesis has been submitted in the past, or is being, or is to be submitted for a degree in the University or any other University.

(iii) I am now presenting the thesis for examination the thesis for examination for the Degree of PhD.

Signed:

Name: **Tembaletu Jama Ndwe**

To my father,
“KOBANJALO”

University of Cape Town

List of Abbreviations

ASR	Automatic Speech Recognition
AU	African Union
BBCCCoE	Botswana Baylor Children's Clinical Centre of Excellence
BGR	Beautiful Game Results
DTMF	Dual Tone Multi Frequency
HIV/AIDS	Human Immunodeficiency Virus/Acquired Immune Deficiency Syndrome
ICT	Information and Communication Technologies
ISO	International Standards Organisation
IVR	Interactive Voice Response
NEPAD	New Partnership for Africa's Development
PSTN	Public Switched Telephone Network
UCD	User-centred Design
UI	User Interface
VUI	Voice User Interface
WOZ	Wizard of Oz

University of Cape Town

Contents

Chapter 1.	<i>An Overview of the Research Study</i>	<i>1</i>
1.1	Introduction	1
1.2	Objectives and goals	2
1.3	Strategy towards objectives	4
1.4	Review of previous studies	5
1.5	Theoretical framework and research positioning	8
1.6	Development relevance to Southern Africa	10
1.7	Summary of findings of the study	11
1.8	Remainder of the thesis	12
Chapter 2.	<i>Usability and its Development Methods</i>	<i>14</i>
2.1	Introduction	14
2.2	Definition of usability	14
2.3	User-centred development	16
2.4	Motivation for selected case studies	17
	2.4.1 Motivations for the OpenPhone application	17
	2.4.2 Motivations for the BGR application	18
2.5	Conclusions	20
Chapter 3.	<i>User and Stakeholder Requirements Solicitation</i>	<i>21</i>
3.1	Introduction	21
3.2	Stakeholder meetings	22
	3.2.1 BBCCCoE staff focus group meeting	22
	3.2.2 BGR stakeholder meeting	23
3.3	Identification and recruitment of the participants	24
	3.3.1 OpenPhone participant and requirements gathering	24
	3.3.2 BGR participant and requirements gathering	28
3.4	Present coping mechanisms	29
	3.4.1 Coping means for OpenPhone's intended users	29

	3.4.2	Coping means for BGR's intended users	31
3.5		Summary of findings	31
	3.5.1	OpenPhone findings	31
	3.5.2	BGR findings	32
3.6		Conclusions	34
<i>Chapter 4. Systems Dialogue Design</i>			36
4.1		Introduction	36
4.2		Description of the fundamental design	37
4.3		Context design	38
4.4		Content design	40
4.5		OpenPhone and BGR telephony prototype systems	42
	4.5.1	WOZ prototype system in OpenPhone	43
	4.5.2	BGR ASR prototype system	44
	4.5.3	OpenPhone and BGR DTMF prototype system	44
4.6		Conclusions	44
<i>Chapter 5. Experimental Design and Procedures</i>			45
5.1		Introduction	45
5.2		Usability evaluation methods	45
5.3		Heuristic evaluation	47
5.4		User testing	47
5.5		Experiments	49
	5.5.1	Main usability tests	49
	5.5.2	Vuvuzela earcon experiment	53
	5.5.3	Experiment on language preferences when reading numbers	54
	5.5.4	A note on the number of experimental participants	55
5.6		Conclusions	56

Chapter 6. Results and Analyses 57

6.1	Introduction	57
6.2	Types of measurements	57
6.2.1	OpenPhone measurements	57
6.2.2	BGR measurements	60
6.3	Experimental results	62
6.3.1	OpenPhone objective results	62
6.3.2	BGR objective results	65
6.3.3	OpenPhone subjective results	68
6.3.4	BGR subjective results	69
6.3.5	Vuvuzela earcon experimental results	72
6.3.6	Numbers pronunciation experimental results	72
6.4	Analyses	73
6.5	Why DTMF for OpenPhone and ASR for BGR?	76
6.5.1	Nature of modality versus socio-cultural norms	76
6.5.2	Learnability issues	78
6.6	Conclusions	79

Chapter 7. Hindsight, Shortcomings, and Future Prospects 80

7.1	Introduction	80
7.2	Hindsight	80
7.3	Findings and contributions	82
7.4	Shortcomings of the study	88
7.5	Future Prospects	90
7.6	Conclusions	92

<i>References</i>	<i>96</i>
<i>Appendix A User-Centred Design Activities</i>	<i>110</i>
<i>Appendix B1 Heuristic Evaluations</i>	<i>111</i>
<i>Appendix B2 Heuristic Evaluations (continued)</i>	<i>112</i>
<i>Appendix C1 Dialog Structure of OpenPhone</i>	<i>113</i>
<i>Appendix C2 Call Flow Diagram for BGR</i>	<i>114</i>
<i>Appendix D Wizard of Oz (WOZ)</i>	<i>115</i>
<i>Appendix E Experimental Procedures Pictures</i>	<i>116</i>
<i>Appendix F1 OpenPhone Tasks</i>	<i>117</i>
<i>Appendix F2 BGR Tasks</i>	<i>118</i>
<i>Appendix G OpenPhone DTMF Post Questionnaire</i>	<i>119</i>
<i>APPENDIX H BGR WOZ Post Questionnaire</i>	<i>120</i>

Chapter 1

An Overview of the Research Study

1.1 Introduction

In the present information age it can be argued that after basic needs like food, shelter, and health care, access to information and communication is one of the most important needs in any population group. This is certainly true in developing countries where citizens need access to the information that is relevant and important to their needs in order to uplift living standards and promote development through connectivity and information sharing (Ndwe et al., 2007).

In Southern Africa, Africa and the developing world in general, it is difficult for most people to access information because the tools and the technologies used to access information such as computers and the internet, are prohibitively expensive and require training prior to operating such tools and using such technologies (Ndwe et al., 2010). This bars those who are either poor, illiterate, or without any computer skills from accessing information and these people are unfortunately the majority of the people living in Southern Africa who vary widely in terms of their educational, socio-economic and literacy backgrounds. This variance makes it an unfeasible task to design a technology user interface (UI) that will suit all. The inability to access information and participate in the present global information economy has the capacity to bring about and promote new forms of social injustices through social exclusion and marginalisation of those who are not technologically savvy.

An alternative channel to access information is through the use of technologies that are readily available and are familiar to the people in the envisaged deployment area of Southern Africa. For example, wireless radio is such a technology because the majority of households in Southern Africa have a radio. The drawback of this technology is that it is a one-way communication device and therefore it cannot be interactive and the users cannot choose the specific information that they want. Another viable solution is the telephone, which is an instrument that the majority of people are accustomed to and have used at some time. There is no prior training necessary for the majority of people in order to make a telephone call. In addition to this, according to a survey conducted in 2007, it is estimated that 72.9% of households have at least one cell phone whereas only 7.3% of households have an internet connection in South Africa (Statistics South Africa, 2007). The number of cellular phones is

growing rapidly in the region and, in fact, the cell phone penetration is growing faster in Africa than everywhere else in the world (ITU, 2006; Brown et al., 2003).

Another important outlook for pursuing telephonic voice interaction as an alternative channel for information access is the fact that speech is the fundamental means of human communications (Nass and Brave, 2005). Speech is so much a natural channel for communication that hearing people in all cultures persuade, inform, and build relationships primarily through speech (Pinker, 1994). This phenomenon is even more important to consider within the realm of the region of Southern Africa where a strong oral tradition exists amongst a large low literacy population (Barnard et al., 2003a).

This research looks at how telephonic communication can be used effectively and productively in order to access information by the intended user population of Southern Africa through the use of an interactive voice response (IVR) system. The research is motivated by the aspiration to address the marginalisation of people who could benefit immensely from the provision of relevant information access that could positively affect their livelihoods, but who are barred from such information access by their economic and social circumstances. The research looks primarily at the usability of telephonic voice access as a feasible means to extend the capabilities of ICT to include those who do not have the tools and/or the know how to access information using the contemporary tools such as computers and the internet.

This chapter introduces the research study and describes its motivation together with research objectives and also introduces the strategy that will be pursued in order to accomplish those objectives. The hypothesis that the research seeks to test is presented together with the nature of the realities that are envisaged to be uncovered as by-products of testing the hypothesis and how those realities are expected to influence the bigger overall objectives of the study. It is expected that the main contribution of the research will be positioned within those realities that will be uncovered in the process of testing the hypothesis and not the hypothesis per se. This introductory chapter also introduces each of the forthcoming chapters of the study and summarises the contents of what is discussed in the rest of the thesis.

1.2 Objectives and goals

The main goal is to establish best practices in the design of IVR systems in the perspective of the intended user population and within the anticipated user population's context of use. Such IVR systems are envisioned to provide access to information and government information-

based services that are relevant to the intended users including health information services, civic information services, social benefits, sport and recreation services, and others. In that setting, the first objective of this research is to investigate the most imperative aspects to address when designing a suitable voice user interface (VUI) within the presented environment of the intended user population in the region of Southern Africa.

Good design improves the usability of voice services – it reduces overall development costs and the time to roll-out by eliminating problems early in the design cycle
(Stentiford and Popay, 1999).

The suitability and worthiness of the technology will be decided by the intended users as they know better than anybody else what works or does not for them. The intention of the research is to follow methods that will guide the development of the envisioned products to be more fitting to the intended users' needs and the evaluation of whether these intentions have been met or not will be left to the intended users. The researchers (the primary researcher and author of this thesis, and various research facilitators) have their own preconceived notions as to what works best, but nevertheless, the awareness of such predispositions is what gives the researchers the urge in pursuing established methods that will steer the design and development of the envisioned products towards the fulfilment of the research objectives within the scope of this study. These established methods are continuous development cycles that improve the technology over a number of iterations but, because of deficiency of time, monetary and other resources, this research study focuses on the first iteration of the developments that are used as case studies. The findings of the study allow for making recommendations on improvements of further iterations that are beyond the scope of this present study.

Another objective of the study is to compare two modalities of the interaction between the user and the end products of the designed IVR systems. The two modalities in concern are:

- a) a speech-enabled system that uses speech or human voice as input commands to the system, and
- b) a dual tone multiple frequency (DTMF) system that uses telephone keys as input commands to the system.

The details of differences between the two modalities are discussed in Chapter 4 and the comparison will be done through usability evaluations of prototypes that will represent each modality of interaction.

The researchers cannot predetermine the outcomes of the investigation and cannot foretell the exact knowledge contribution that will be made by the research before the journey has begun. It is anticipated that along the voyage of the IVR design and how the intended users perceive

the technology, some realities about the design of IVR systems in the context of the given user population will be established. As a guiding light, the research hypothesises that ignoring cultural aspects and the social environment of the intended users will lead to ineffective UI design and consequently an unsuccessful system. In testing this hypothesis, the research anticipates to unveil the aforementioned realities about IVR design in the optimism that those realities may apply not only within the constraints of our specific intended products but to the general field of IVR systems design, particularly for the aforementioned deployment area.

1.3 Strategy towards objectives

The methodology used in accomplishing the objectives of this research is to make use of case studies in developing and deploying IVR systems. Since the study is concerned with establishing best practises in the design of IVR systems for a specific user population and context of use, the case study method, according to Yin (2003), is appropriate as it is used when:

- the focus of the study is to answer “how” and “why” questions – the research seeks to establish how to design appropriate IVR systems for the target users and explore the phenomena surrounding their use practices by investigating why they (the targeted users) use such systems the way they do.
- you cannot manipulate the behaviour of those involved in the study – the proposed IVR technology is built and modified around the target users’ context and not the other way around, i.e., the target users are not required to change their context and behaviour in order adapt to the proposed technology.
- you want to cover contextual conditions because you believe they are relevant to the phenomenon under study – as mentioned in the previous statement,
- the context of the target users is pivotal in the research study (Yin, 2003).

The first case study is the design of an IVR system for caregivers to HIV/AIDS infected children that enables the caregivers to access information on care-giving issues in Botswana, Southern Africa. The IVR system is called OpenPhone as it was built using open sources and standards in 2008. The major stakeholder of the OpenPhone system is the Botswana Baylor Children’s Clinical Centre of Excellence (BCCCCoE), the clinic that the caregivers get services. The BCCCCoE is located in Gaborone, the capital city of Botswana.

The second case study is the design of a soccer results retrieval system, which enables soccer fans to obtain the results of recently played games of the Premier Soccer League (PSL) of South Africa. Soccer is affectionately referred to as the beautiful game by sports presenters

and fans in South Africa and around the world (Garicano and Palacios-Huerta, 2005; Vrooman, 2007), hence the application is called Beautiful Game Results (BGR) system. The BGR was built in 2010 and it is targeted at the fans of the soccer game which is the most popular sport in both South Africa and the Southern African region.

The details and the motivation behind the selection of the particular case study applications, the target geographical areas, and the circumstances of the target user populations are discussed in the next chapter. Specifically, the research centres on the first iterations of both case study developments. It is envisaged that the lessons learnt and the experiences gathered during the design of first iteration development prototypes of OpenPhone and BGR systems will enable this research study to explore its primary and secondary objectives. It is worth noting that since the case studies were developed two years apart, there was more experience and prudence in the development of the second case study than in the first.

1.4 Review of previous studies

There have been other studies of IVR systems whereby the users were required to compare the speech-enabled and the DTMF modalities. In contrast to the current research study, the majority of those studies have been done in developed countries, with users who are generally literate, for example, professionals or college students (Delogu et al., 1998a; Foster et al., 1998; Karis, D., 1997). The intended users of this study are different from the user populations in aforementioned studies as the culture and the given users' mental models are different from that of those in other studies that were carried out in the Global North (developed countries of the world). The geographical location of the other studies is also important due to the fact that this study is conducted in Africa, where oral rather than literary orientation prevails. It is a place where news travels from person to person by spoken word (An analysis of African oral traditions, 2010). Southern Africa is not different from the rest of Africa as far as this research study is concerned because both case studies of this research are aimed completely at the indigenous African people of Southern Africa who are strongly grounded in oral traditions (Barnard et al., 2003a), whereas the aforementioned studies were done within Western cultures which are predominantly literate cultures. This is an important differentiation as orality affects various aspects that are of importance in the interaction between the intended users and the proposed technology including how people think, communicate, and learn (Sherwani et al., 2009a). The context of orality in this study is that of residual orality which is defined as verbal expression of thought within cultures that have been affected or exposed to writing and print but have not fully internalised the use of such technologies in their daily lives (McLuhan, 1962). Residual orality is different from primary

orality which is defined as verbal expression of thought within cultures that have not been affected or exposed to writing and print at all (Ong, 1982). A culture can be defined as being made up of certain values, practices, relationships and identifications (Massey, 1998). The different variables that constitute a culture make each culture to be unique in its mental models including its understanding and interaction with technology and hence unique in its user needs and context of use. As stated in section 1.2, this study hypothesises that different cultural norms and socio-cultural constitutions have different influences in the way that technology should be designed and ignoring such influences will eventually lead to an unsuccessful system. In that view, there is a need to investigate the user interactions with IVR systems according to the targeted users' social and cultural circumstances. Various cultural aspects of the intended users will be considered in the design of the IVR systems, and it is predicted that one of the main cultural aspects to be considered will be that of language since the UI of the technology under investigation is speech orientated.

The results from the previous studies conducted in the Global North have shown various outcomes with the preference of DTMF input for linear task completion and speech-enabled input for non-linear tasks (Lee and Lai, 2006). Other studies, conducted by Harris Interactive and commissioned by Nuance, have shown that speech-enabled IVR systems are preferred by users over the DTMF system signifying ease-of-use, convenience, and accessibility as the key benefits in using the speech-enabled system (Nuance Communications Inc., 2003). Delogu et al., (1998a) found preference of DTMF over speech-enabled input designed for an isolated word interface. Both case study applications in this research use isolated word input and the tasks that are to be completed by the users are all linear tasks but the present study is carried in a different context of the Global South within a setting of two different case studies that are contain contrasting subject matters.

There have been other studies within oral communities in developing countries for designing information access via speech technology but with differing cultural and user contexts, and also differing geographical locations from this study. Plauché et al. (2006) did a speech technology study for accessing agricultural information in India and established that literacy has an effect in the task completion and success rates. Sherwani et al. (2007) investigated the design of a speech-based access to health information by low-literate users in Pakistan. Sherwani et al. (2009b) established the preference of speech-enabled input compared to DTMF in a healthcare system that was intended for low literacy users. However, the abovementioned studies inevitably involved limited user populations and task domains – for example, it did not involve stigmatised and privacy issues (the healthcare system was aimed at neonatal healthcare). Also, the aforementioned studies in the developing world did not

involve a pleasurable application that is aimed at entertainment for people in developing countries. Sharma et al. (2009) utilises one of the same case studies as investigated in this research, in a stigmatised healthcare setting, and found a preference for DTMF over speech-enabled input.

Patel et al. (2009) did a comparative study between speech-enabled and DTMF inputs for an Avaaj Otalo (AO) application for agricultural information access intended for farmers in rural India and found preference of DTMF citing faster completion rate and less difficulty in providing input to the system when compared to the speech-enabled input system. However, the Patel et al. (2009) study does also not investigate privacy-related issues, and does not make use of within-subjects experiment design as in the case of the current research study. We value the within-subjects experimental design as it is important for the study participants to experience both modalities in order to ensure impartial comparison between the two modalities. In another study conducted by Patel et al. (2010), on the same AO platform, the authors lean more towards the use of speech technology as a message board in the context of social media and focussing on user input and feedback as a critical part of the interaction experience. Patel et al. (2010) focus on a longer term of experience as the AO experiment is conducted over a seven month period unlike the current study which focuses on the users' first encounter with the VUI and compares the two modalities based on a within-subjects comparison based on their first encounter. Patel et al. (2010) found user preference of DTMF over speech-enabled modality but the experimental design in that study is different from this study. Patel et al.'s (2010) study was done on one application that enabled the user to either use spoken input or dial a corresponding number to interact with the system. The choice of interaction modality was made during the introductory prompt in the root menu of the system and once the user chooses a particular modality at the beginning the system, the AO system would maintain that interaction modality throughout the interaction. For example:

AO: Welcome to Avaaj Otalo! You can get to information by saying a single word, or by dialing the number. To ask a question, say 'question', or dial 1; to listen to announcements, say 'announcements', or dial 2; to listen to the radio program, say 'radio', or dial 3.

User: (dials 1)

AO: OK, you want to ask a question. To record your own question, press 1. To listen to the questions and answers of other farmer friends, press 2.

User: (dials 1)

AO: OK, you want to record a question. Please say your question slowly and clearly after the beep.

User: How can I protect my cotton crop from mealy bugs? (Patel et al., 2010).

From the above AO dialogue it can be noted that there is a pattern of first introducing the option of spoken input and then the dialling option secondarily and once a choice is made there is no option of altering the interaction modality. It has been established that users tend to better understand the critical information of the message if it is in the final part of the sentence rather than in the initial part (Delogu et al., 1998b) and the DTMF option in the dialogue was always in the final part of the sentence during the introduction of options that the users have. This could have influenced the choice of modality during the first encounter of the interaction and unfortunately the user had no way to alter this choice for the duration of the interaction. The AO study may have been more unbiased by alternating the order in which the modalities are introduced to the user. The current research study avoids such partiality by ensuring that the interaction with the different modalities is done over two discrete interaction sessions and communication lines and the user does not have to choose one modality over the other before they have experienced both modalities. This discrete procedure is maintained in both case study applications. Patel et al.'s (2010) study also makes use of commands 'dial' and 'press' interchangeably within the same dialogue which can be confusing to the user, particularly within the context of developing countries where there is still the existence of rotary phones. A user who is accustomed to dialling numbers on rotary phone and tested on an AO touchtone phone may be confused on what to do when instructed to 'dial' a number on the keypad. The present study avoids such confusions of using different words that produce the same action and maintains consistency throughout the system instructions.

This study compares DTMF and ASR input modalities in the setting of two different applications. The first application involves accessing information about an illness that is pandemic in the region of Southern Africa and has brought about much damaging consequences in all walks of life and either infects or affects everybody in the region. The second study is in strong contrast to the first as it a passionate and fun application that allows soccer fans to access soccer results of recently played games. Soccer is the favourite sport in the region and its status as the favourite sport has recently been enhanced by the World Cup 2010 soccer games which were recently held in the region. The researchers are not aware of any published research studies that compare the two modalities in the milieu of two conflicting settings within the region, within any oral communities or anywhere else.

1.5 Theoretical framework and research positioning

This research study is conducted within the field of engineering as it is about the design of a system, but it is heavily influenced by social sciences in the sense that it is a study of how human beings interact with a technological system. The study of the human side of the system

is driven by the desire to improve the design of the technology so that it serves its users more fittingly. The study is aimed at enhancing the users' livelihoods through provision of information and services by means of a familiar technological tool, i.e. the telephone, rather than a technology that needs to be learnt before users are able to use it such as a standard multi-purpose computer. Engineering is a field that is mostly dominated by the empirical approach in research which relies heavily on measurable and observable facts and based on knowledge as exclusively a function of experience rather than hearsay, rumour, doctrine, intuition or imagination (Janse van Rensburg, E., 2001). This particular research would be limited and incomplete if it were to take the empirical approach alone, because there are still things that cannot be measured in an empirical manner (Mrazek, 1993). This is particularly so in studies concerned with human behaviour and the improvement of quality of life.

The research endeavours to design a technology according to the users' needs and expectations rather than from the novelty and functionality of the technology as perceived by the designers. This allows the users to participate in the design of the technology that will eventually be used by them, which is a way of empowering the users to produce tools that will be used by them. In turn, this enables the target users to actively participate in the information economy at first hand rather than hear about the so-called 'information age' and not knowing what it means.

Historically, society has seldom given poor people ownership over the tools of production (Castells, 1997).

As stated in the introductory paragraph, the research hypothesis provides a roadmap for the journey towards the objectives of the research and the main knowledge contribution of the research lies by the roadside as the research unveils realities about the design of IVR systems in the given setting. This implies that the main contributions will be unearthed through unbiased research questions and unforeseen occurrences, observed trends and facts that will present themselves along the research journey. The main reason for the research through discovery approach is that there have previously been few research studies with the same intentions of gathering best practises in IVR systems development within the intended population. This entails that the research has to proceed with an open mind to new discoveries without preconceived expectations. The knowledge interest in this study is of technical nature with an anticipation to tackle a social problem, i.e., information inaccessibility, in a predominantly technical manner.

1.6 Development relevance to Southern Africa

A telephone is available in many places and requires no particular skills to operate and if this familiar technology can be used to access information, it can help in bridging the information gap that exists in Southern Africa. The research is expected to benefit the intended users including communities who are technology illiterate with no access to contemporary information technology tools such as multi-purpose computers. In the context of providing government services in the aforementioned target area, as described in Barnard et al. (2003b), speech technology is envisioned to play an important role in rendering the necessary government services to the citizens. The technology will also enable governments in the region to serve their citizens more adequately and reduce the overcrowding and queues that are typical of government service offices in Southern Africa.

The research output can also be applied in the rest of the developing world, and particularly in Africa, which is in accordance with the New Partnership for Africa's Development (NEPAD) and the African Union (AU) initiatives which are aimed at political, social, and economical development that seeks to pull the entire continent out of the quagmire of poverty, underdevelopment and marginalization (Address by The President Of South Africa, 2001). It is recognised that ICT is capable of accelerating development and improving people's livelihoods (Langmia, 2006; Benner, 2003). A user centred approach to ICT development will help in making sure that the technology is truly usable to its target user population.

Speech systems are generally viewed as valuable tools for information access that will have significant impact in the developing world (Sharma et al., 2009; Barnard et al., 2008, Barnard et al., 2010b), and one of the main reasons for high regard for such technology is because of its utility. Utility refers to the extent to which the product provides the right kind of functionality so that users can do what they need or want to do (Sharp et al., 2007). Barnard et al. (2008) have reported on the utility of speech systems in the developing world and pointed out that their utility depends on several issues including cost, payoff, user needs, complexity of the application, and quality of the design and development process. Another strong incentive for the use of speech technology is that of a "motivated user", whereby inherent attributes of the user population prevent the user from interaction with information systems except through speech technology (Barnard et al., 2008). In the context of oral users in Southern Africa, the inherent problems include the aforementioned lack of access to contemporary ICT tools such as standard multi-purpose computers, the cost of such tools and the learning process involved in learning such tools as they require, at least, basic computer literacy in order to use them (Ndwe et al., 2007). Plauché et al. (2006) also support the

initiative of a “motivated user” through their finding that user acceptance is positively correlated to the difficulty that users would have in accessing the same information through other means.

1.7 Summary of findings of the study

In comparing the two modes of DTMF and ASR in IVR systems, it cannot be conclusively established that oral users prefer one over the other and this research proposes that the type of content in the interaction has repercussions on the users’ choice.

In oral users, the objective usability measures of effectiveness and efficiency, as defined by International Standards Organisation (ISO), do not have a conclusive bearing on oral users’ choice of technology.

There is a positive correlation between oral users’ choice of technology and learnability of the technology.

Oral users’ choice is affected by how the technology allows for the users’ current situational norm at the time of interaction with the technology.

In oral cultures, the variances in situational norms are carried over into the interactions with speech technologies such as IVR systems.

There is a statistically significant difference in task completion times between lower and higher literacy users.

The following three findings have been built upon recommendations that were introduced by Sherwani et al. (2009a) concerning research on speech technology in oral users. Sherwani et al. (2009a) based their recommendations upon theories that were introduced by Ong, (1982).

This research has scientifically examined these theories and the results were as follows:

- Oral users are more inclined to position a voice by the speaker
- Oral cultures promote volubility and redundancy in speech.
- Oral cultures tend to speak in an antagonistically toned manner

After having examined the theories and the results thereof, this research has established that there is a negative correlation between oral behaviour in speech technology interactions and the users’ level of education.

Culturally inclined earcons are an effective means of providing grounding about the context of the speech application that the oral users are accessing.

Culturally inclined earcons also help in evoking the situational norms and afford the users to engage easily into the situational norm that is demanded by the context of the application accessed.

In oral settings where the colonial language of English has remained the dominant *lingua franca* of the various indigenous peoples of the country even after political independence, such as in the majority of Southern African countries, English is a viable interaction language for IVR systems that primarily provide numerical information.

Careful attention to the methods used during the first encounters with the intended users during the requirements solicitation is beneficial for the development even in later development stages.

Learnability curves and the gradient of the change in success rates provide a visual expression of the occurrence and magnitude of learnability.

1.8 Remainder of the thesis

Chapter 2: Usability is defined; various perspectives and the methods that need to be followed in order to achieve usability are discussed. The motivation behind the choice of the OpenPhone and BGR systems as the case studies is also discussed.

Chapter 3: The intended user population is further explained in this chapter. Requirements from both the users and the stakeholder are gathered in order to inform the design process. Personae for the two systems that represent the users are also created.

Chapter 4: The dialogue designs for the OpenPhone and BGR systems are discussed together with the manner in which the design was influenced by the circumstances of the intended users.

Chapter 5: This chapter presents the experimental designs and the procedures that were followed in conducting usability experiments with the intended users of the systems. Various usability evaluation methods are presented and the motivations for the methods that are used in this study are discussed.

Chapter 6: Types of measurements that were used in the usability evaluations are discussed. Results that have been achieved from the experiments that were conducted in the previous chapter are analysed and conclusions based on the analyses are presented.

Chapter 7: The final chapter provides the findings of the research and takes a look at the shortcomings of the study and what might have happened if things could have been done differently. The chapter presents hindsight on the road of discovery that has been travelled by the researchers in this study and also presents future prospects on similar developments within the region and the target users. Further recommendations on IVR system design within the same setting are presented.

University of Cape Town

Chapter 2

Usability and its Development Methods

2.1 Introduction

The previous introductory chapter presented the standpoint that in order to effectively provide information-based services to the intended users of Southern Africa, there has to be research into alternative means of providing relevant information to the users in the aforementioned area. Unlike contemporary ICT tools which require at least basic computer literacy, these alternative channels for information access have to be technologies that the targeted user communities are accustomed to and are also widely available. This research specifically looks into the manner that the telephony information access systems are used by the intended users and utilises development methodologies that will account for the users' circumstances and attributes. This chapter examines such development methodologies both in the general context and within the context of this research and also looks at the benefits of following such methodologies. These development methodologies are the same methods that were envisioned in the previous chapter as "methods that will guide the development of the envisioned product to be more fitting to the intended users' needs", and, "methods that will steer the design towards the fulfilment of the research objectives" as discussed in section 1.2. The OpenPhone and BGR case studies that were introduced in the previous chapter will also be discussed further together with the motivations for choosing such case studies.

2.2 Definition of usability

There are several definitions of usability (ISO 9241-11, 1998; Shackel, 1991; Nielsen, 1993) that emphasize a variety of usability characteristics or attributes as required tools for the measurement of usability. Although the different definitions of usability have common ground and shared characteristics they, nevertheless, have differing perspectives in the way that they quantify usability of a product. For the purposes of the current study, we wish to arrive at a definition that is most applicable for the context of systems design in both our case studies.

According to the ISO the definition of usability is:

Usability is the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use (ISO 9241-11, 1998).

Additionally the ISO standard describes the implied usability characteristics of effectiveness, efficiency, and satisfaction in Table 2.1.

USABILITY CHARACTERISTIC	DEFINITION
Effectiveness	Accuracy and completeness with which users achieve specified goals
Efficiency	Resources expended in relation to the accuracy and completeness with which users achieve goals
Satisfaction	Freedom from discomfort, and positive attitudes towards the use of the product

Table 2.1: Definitions of usability characteristics (ISO 9241-11, 1998)

ISO 9241-11 also provides definitions of other usability components used within the standard definition itself and within the definition of usability characteristics in Table 2.2.

USABILITY COMPONENT	DEFINITION
Context of use	Users, tasks, equipment (hardware, software and materials), and the physical and social environments in which a product is used.
User	Person who interacts with the product
Goal	Intended outcome
Task	Activities required to achieve a goal
Product	Part of the equipment (hardware, software and materials) for which usability is to be specified or evaluated
Measure (noun)	Value resulting from measurement and the process used to obtain that value

Table 2.2: definitions of usability components (ISO 9241-11, 1998)

The most salient aspect which sets this study apart from most other similar studies that have been done in investigating the usability of an IVR system is the socio-cultural circumstances of the specific user population that the product is intended for. This implies that the context of use is different from those of the other studies as indicated in section 1.4. The ISO 9241-11 standard definition of usability is becoming the main reference of usability (Jokela et al., 2003) and the standard presents a contextually oriented view of usability (Folmer and Bosch, 2004). The contextual orientation and popularity of the ISO 9241-11 standard motivated us to adopt the ISO definition as the main definition that will guide the way that usability is engaged and measured in this research. However, other viewpoints of usability will also be

integrated into the research as they are not in contradiction with the ISO definition but rather extend on it.

The usability definitions by Nielsen (1993) and Shackel (1991) expand on the ISO definition by including learnability as a characteristic of usability in addition to ISO's effectiveness, efficiency and satisfaction. Learnability is defined as change of usability over time as the user becomes more acquainted with the system. Shackel also includes the flexibility characteristic which is defined as how the system will change and adapt to the change in the context of use within the users (Shackel, 1991) which involves changes that occur over an extended period of time. As a result, flexibility is not applicable to this study due to the fact that this study is mainly concerned with the first round of development iteration of the intended products whereas flexibility arises over a longer period of time.

2.3 User-centred development

User-centred design (UCD) is an approach for employing usability (Nunes, 2006). It is a structured product development methodology that involves users throughout all stages of product development in order to create a product that meets users' needs through the consideration of the stakeholder organization's business objectives and the users' needs, characteristics, and preferences. UCD is also referred to as usability engineering or human-centred design (Jokela et al., 2003).

UCD is made up of stages that the development of the product goes through in a systematic manner and in a chronological order as follows:

1. *Understanding and specifying the context of use which involves knowing the users and their setting of use.*
2. *Specifying the user and organisational requirements entails the determination of success criteria for the product and acquiring design guidelines and constraints from both the stakeholder organisation and the intended users.*
3. *Production of design requires integration of the information from the above stages together with theoretical knowledge of UI design to produce a design solution.*
4. *Evaluation of design against requirements is done through representative tasks with representative users during usability evaluations (Jokela et al., 2003).*

The stages of UCD are presented graphically in Appendix A. Many reports recommend the development methodology of UCD as an aid to improve the usability of a product (Mayhew, 1999; Nielsen, 1993; Beyer and Holtzblatt, 1998; Vredenburg, 2002).

Nielsen (1992) provides another popular alternative methodology that is more condensed and generalised yet congruent to the above methodology of UCD and calls it the usability engineering process. Nielsen's usability engineering process is executed through chronological stages that are called the usability engineering life cycle and this methodology consists of 3 stages as follows:

1. *The pre-design stage*: this stage combines stages 1 and 2 of the above mentioned UCD methodology.
2. *The design stage*: this stage combines stages 3 and 4 of the UCD, and
3. *The post-design stage*: this stage comprises gathering information that will positively influence the subsequent iteration of the design (Nielsen, 1992).

2.4 Motivations for the selected case studies

The application selection for a case study that will manifest the larger intention of providing access to information and services via a telephone for the intended users within the targeted user area is a challenging task on its own. The development of the selected application should reflect the unique environmental and social challenges that are confronted by the people in the target deployment area. The challenge in selecting the appropriate application comes with the fact that no one application can reveal all the uniqueness of the environmental and social dimensions that exist in the target area. This section discusses the reasons for selecting the two case studies.

2.4.1 Motivations for the OpenPhone application

The OpenPhone case study has been selected for the following reasons:

- I. The Southern African region is made of different countries and some of those countries have different cultures and languages within the country. For example, The Republic of South Africa alone has 11 official languages and various cultures that are associated with those languages. The OpenPhone case study was conducted in Botswana, one of the Southern African countries that have a single dominant cultural setting whereby the indigenous people in the country belong to one culture of the Batswana people. The single cultural setting of Botswana was suitable as the initial case study for this research because the research needed to isolate and focus on one culture without the influence of various cultures to begin with. A multicultural setting would have complicated the study as it would have introduced more variables to solve whereas the research intends to begin with a single cultural variable and explore a multicultural setting thereafter.

- II. Providing health information on HIV/AIDS is a very important issue as the disease has reached pandemic proportions in the region. The top 8 countries with the highest prevalence of the disease in the world are countries in the Southern Africa (UNAIDS, 2008).
- III. In heavily-affected countries, such as Botswana and Zimbabwe, HIV is the underlying reason for more than one third of all deaths among children under the age of five (Mason, E., 2006), and the OpenPhone application is aimed at caregivers of HIV infected children.
- IV. The major stakeholder for the OpenPhone, i.e. BBCCCoE, has already adopted the strategy of dealing with the disease through education and provision of information about the disease instead of providing medication alone as most public clinics in the region do. The BBCCCoE does this by conducting lectures every weekday mornings at the clinic which educate caregivers on care-giving issues.
- V. In BBCCCoE's attempt to tackle the disease through education, they are faced with the problem that most of the caregivers are not adequately literate to use printed material as a reference for care-giving which makes it difficult for the distribution of care-giving information. The lectures at the BBCCCoE are not effective because the caregivers forget what they were told during the lectures and BBCCCoE is aware of this. Thus, there is a clear need for some other intervention into the situation of providing information on care-giving issues.
- VI. Botswana has the second largest prevalence of HIV/AIDS in the world (UNAIDS, 2008).

The combination of the abovementioned facts produced sufficient motivation for a case study in deploying an HIV/AIDS information IVR system called OpenPhone for research on telephony access system development in the region of Botswana, Southern Africa.

2.4.2 Motivations for the BGR application

The BGR case study has been selected for the following reasons:

- I. In contrast to the OpenPhone system, the BGR application has been conducted in a setting with multicultural influences rather than a single dominant culture. The research needed to find out the effects of various languages and cultures in the milieu of developing and deploying a telephony information access technology. The effect of introducing several languages and cultures is envisioned to bring about more complexity in the study particularly in view of the fact that the study is speech

oriented. Nevertheless, the language diversity reflects an important aspect of the state of affairs in much of Southern Africa, since there is a great variety of cultures and languages in the region. The BGR case study was conducted in South Africa, a country that is one of the Southern African countries with diverse language and cultural setting: there is a variety of cultures with 11 official languages. The multicultural setting of South Africa was suitable as a subsequent case study for this research following the single-cultured focus of OpenPhone. It is important to note that the BGR application is primarily designed for scientific analysis of how to design appropriate telephonic information access systems within the region and therefore the major stakeholder of the BGR application is the research team that is conducting this research.

- II. Soccer is the most popular sport in the world (The Most Popular Journal, 2006; Top 10 Most Popular Team Sports in the World, 2010); it is the most popular sport in Africa (Most Popular Sport in Africa, 2010; Exploring Africa, 2010); and also the most popular sport in South Africa. In particular, its popularity reaches a zenith amongst the indigenous African people, although people of all races play and watch the sport. Soccer boasts more supporters than any other sport in South Africa, has the highest media coverage and, notably, it is played in both rural and urban areas in the country (Makhaye, 1998). No other mainstream sport in South Africa has this feature of being equally played and popular in both rural and urban settings.
- III. Soccer, which is also called football in many countries, serves a distinctive role in political and societal transformation. It is more than just a favourite pastime, but also a political unifier of the various peoples of South Africa (SAFA, 2010). This perspective of soccer as a unifier, in a nation that was previously plagued by the social injustice of apartheid, has been reinforced by the resounding success of the Fédération Internationale de Football Association (FIFA) 2010 World Cup held recently in the country. The event has been applauded as a success both locally and abroad, with 91% of South Africans believing their country is now more unified (South Africa's FIFA World Cup, 2010).
- IV. The success of the FIFA 2010 World Cup has offered the already popular game an even more auspicious future in the country, as the money given to South Africa by FIFA, the international governing body of football, will be directed towards developing local soccer (Valcke: South Africa set a new benchmark, 2010).

The above mentioned facts motivated the development of the BGR application as a second case study for research on telephony access system development in Southern Africa.

2.5 Conclusions

Usability is a key concept in understanding how to support users in their tasks. The benefits of usability include increased productivity, enhanced quality of work, reductions in support and training costs, and improved user satisfaction (Kujala, 2002). Good usability of any product is a desirable asset and in order to ensure usability in the design of interactive products designers must deliberately engage established systematic methods that will deliver a product that is usable to its target users. Good intentions and ad hoc design based on intuition are not enough to ensure the usability of an application (Mayhew, D. J., 1999).

Two development methodologies have been proposed in this chapter and both methodologies have the same common goal of improving the usability of the resulting product. Both methods will be used together in this study as they are in accord with each other. The research will benefit from the application of the aforementioned methodologies to the design of the OpenPhone and BGR systems which will be explored in the forthcoming chapters, starting with the user requirements and analyses in the next chapter.

Chapter 3

User Requirements Solicitation and Analyses

3.1 Introduction

Usability and the significance of its application to the development of interactive information access systems have been discussed in the previous chapter. Two development methods have been discussed and selected as paths to be pursued in order to achieve usability. To develop a system that is aligned with the users' context of use this research has to depend on the information that is provided by the users or anticipated users of the system (Lynch and Gregor, 2004). It has long been acknowledged that user participation and mutual understanding between the system developers/researchers and the anticipated users are two of the most important factors in the successful development and implementation of systems (DeBrabander and Edstrom, 1977; Dickson and Simmons, 1970; Lucas, 1973; Powers and Dickson, 1973).

In this chapter, the development of collaboration between the researchers and the anticipated target users is presented together with initial information about anticipated users, their environment, and what they presently do in order to meet their information needs. The researchers also get to know the system stakeholders and their needs. It is a vital stage in the development of the system because establishing user requirements is well recognized as a critical step in the development of useful and usable systems (Curtis et al, 1988). This is due to the fact that the hard part of building systems is not building them, it is in knowing what to build—it's in acquiring the necessary knowledge (Armour, 2000).

The initial meetings that are discussed in this chapter with the intended users and the stakeholder can be viewed as first induction meetings as further collaborative meetings are planned for future iterations which are beyond the context of first development iterations of this study. This is because user involvement should be continuous throughout the project lifecycle (Gulliksen et al., 1999; Lynch and Gregor, 2004). The user and stakeholder requirements gathering for the OpenPhone system was conducted during focus group meetings that were held at the customer premises, the BBCCCoE, in Botswana over a three different gatherings on three successive days. User requirements gathering for the BGR system was conducted over a three month period and two different modes of requirements solicitation were used which were individual interviews that were conducted face-to-face and individual interviews that were conducted over the phone for participants at a distance. In this

chapter the meetings are discussed and the proceedings of the information gathering process are also described.

3.2 Stakeholder meetings

In the corporate environment, stakeholders are defined as those groups without whose support the organisation would cease to exist (Freeman and Reed, 1983). Drawing from this definition, and within the context of this research study, stakeholders are those groups or individuals without whose support the case studies would cease to exist. The stakeholders bear the largest concern for the development and deployment of each system and therefore it is imperative to address their concerns early in the development of the proposed systems. In this section the stakeholder meetings are discussed.

3.2.1 BBCCCoE staff focus group meeting

The staff focus group meeting was held at the BBCCCoE conference centre with BBCCCoE staff members composed of a doctor and 5 nurses. The objective of the focus group was to get the staff members at BBCCCoE to acquaint themselves with the proposed content for the system and review the content such that it is aligned with their needs. This content is the information that the caregivers will be accessing when using the OpenPhone system. It is essential for the researchers to have the BBCCCoE's consent on the content that is to be on the OpenPhone system as they are the major stakeholder who will be accountable for the information that is accessed by their clients, the caregivers.

The proposed content was prepared prior to the meetings by the researchers and constructed from two independent sources. The first source is a document from the BBCCCoE entitled, "Adherence Education and Psychosocial Support Guidelines: *Training Curriculum and Toolkit*". This document is the handbook that is used by the BBCCCoE staff members as a lecture guide when giving lectures to the caregivers. The second source of the proposed content is a professional HIV/AIDS consultant, Prof. Alta Van Dyk, who is a research psychologist and a professor at the University of South Africa (UNISA), in Pretoria, South Africa on the HIV/AIDS topic. She is an author of the book entitled, "HIV/AIDS Care and Counselling" (Van Dyk, A., 2005), which has been used in the proposed content for the OpenPhone system under investigation. In an interview that was held with her at UNISA, she provided advice on how to conduct focus groups with caregivers and important factors to look out for during the focus group meetings, and she kindly offered the book to the research team.

Through the staff focus group meeting the researchers got an opportunity to learn the context in which the OpenPhone system is to be used directly from the BBCCCoE. The meeting also served to strengthen the partnership between the customer and the researchers. The BBCCCoE staff members were given the proposed content in print so that they could comment and edit the content as they saw fit. Some of their heartfelt input was immediate, such as the present escalation of drug abuse problem amongst youth that was being encountered by the country of Botswana. They immediately suggested that the system should address this problem as it is very crucial for patients who are taking HIV/AIDS treatment to abstain from using any drugs as this has a negative effect on the treatment program. The research team would not have thought of this problem beforehand as it was not documented in any of the literature that the team used to formulate the proposed content. This demonstrated the significance and positive effects of conducting the focus group meeting.

3.2.2 BGR stakeholder meeting

As previously mentioned, the BGR application is primarily designed for scientific analysis of how to design appropriate IVR systems within the region which implies that the major stakeholder of the system is the research team. The researchers met before conducting user requirements solicitation to brainstorm about the development of the application. The researchers agreed that the user requirements participants would come from different parts of the country as the application is aimed at soccer fans around South Africa. The system would provide only the results of recently played games in the Premier Soccer League (PSL) and not in other leagues such as the first division and second division leagues.

It was concluded by the researchers that the system would have efficiency as an essential feature in the BGR application. In the view of the popularity of the sport in the country, the researchers envisioned that callers would call repeatedly, and this would require the system design to facilitate a speedy navigation path through the application and a minimal connection time per user which would consequently result in high system efficiency and user satisfaction. The researchers resolved to a maximum call connection time of one minute per user interaction and required the BGR design to incorporate this constraint as high priority.

The researchers also decided that the BGR system would be primarily aimed at the indigenous African people of the country because soccer is significantly played and supported by that sector of the population. This sector of the population forms nearly 80% of the whole population of the country (South Africa's population, 2010; Statistics South Africa, 2010). Since this sector is neither culturally nor linguistically homogenous, the researchers agreed

that the development would need to address the diversity in languages as a prominent issue since the proposed technology is speech oriented.

3.3 Identification and recruitment of the participants

In order to amalgamate user requirements there was a need to recruit potential users of each system and then solicit user requirements from those anticipated users. This section discusses the procedures that were followed in recruiting the prospective users and requisition of user requirements for the OpenPhone and the BGR systems.

3.3.1 OpenPhone participant recruitment and requirements gathering

Staff members at the BBCCCoE were informed about the staff and user focus group meetings that would be held at the clinic two months before the meetings actually took place. They were asked to notify the research team on the best time to conduct the focus group meetings and the dates for the visits to BBCCCoE were determined and set by the BBCCCoE. They also informed the research team that meetings with the target user participants for the planned focus groups, the caregivers, would be arranged easily as they are the people who visit the BBCCCoE for consultations and lectures on a daily basis. They informed the research group that they would ask target users to attend the focus group immediately after the usual lectures that are normally given to the caregivers every morning.

The process of requirements development and specification is complex even when designing an elementary single user system because users often cannot properly articulate their needs (Pekkola et al., 2006). Within the OpenPhone project the complexity of requirements development is fuelled by fact that the general use of IVR systems for information access is fundamentally an unknown concept to the target users.

There was no prior communication between the researchers and the target user participants. The participants were selected using a sample of convenience as they were indiscriminately recruited from the pool of caregivers that go for consultations and attend lectures at the BBCCCoE. In that way, the participants were easy to reach and they were also willing to participate, which made the aforementioned sampling technique applicable (Oates, B.J., 2006). The participants are characterised as homogenous strangers since they have something in common, which is the need for care-giving information, and yet they do not necessarily know one another.

The focus group approach was chosen as the primary methodology for acquiring user needs for the OpenPhone system. This methodology is pragmatically appropriate for this particular user group because focus groups:

- *do not discriminate against people who cannot read or write*
- *can encourage participation from those who are reluctant to be interviewed on their own (such as those intimidated by the formality and isolation of a one to one interview)*
- *can encourage contributions from people who feel they have nothing to say* (Kitzinger, 1995).

The above characteristics of focus groups fit the intended user population well in the milieu of the challenges faced by this user population. There are some guiding rules on the basic composition of a focus group which state that focus group must often:

- *Use homogenous strangers as participants*
- *Rely on relatively structured interview with high moderator involvement*
- *Have 6 to 10 participants per group* (Morgan, 1997).

These rules are only for guidance and are not rigid because, for example, focus groups can consist of as few as five members (Sampson, 1972) and as many as 20 members (Hess, 1968).

The caregiver focus group meetings had three primary objectives:

- The first was to study the user characteristics in order to compile and develop a user profile.
- The second was to allow the targeted users to voice their opinions as to what concerns they would like the proposed system to address.
- The third objective was to engage the targeted users to form a coalition in the design of the system through participatory design which intends to involve the targeted users in all phases of the project. This particular focus group engagement concentrated on introducing the system to the target user population and gathering opinions, beliefs, and attitudes about issues of interest about the proposed system.

The caregiver focus groups were held over two sessions on two consecutive days with two different groups of participants. There were 11 participants on the first day and 16 on the second and almost all were female (with only one male participant on the second day). Both sessions started with a welcome speech that also thanked the participants for attending the focus groups. All conversations and interactions with the caregivers were conducted in Setswana, the language of Batswana people of Botswana, with the aid of two moderators who were fluent in both English and the local language of Setswana. One of the moderators was a

local Botswana citizen and resident which has helped the research team in comprehending the local cultural nuances which would not have been understood by any other means, not even by the other moderator who is fluent in Setswana but not a resident nor a citizen of Botswana.

The OpenPhone system objectives were then introduced to the participants and the moderators made it clear to the participants that the research team was there to learn from them as to what was needed to be addressed by the proposed system. The moderators informed the participants that the system to be built would only succeed if the participants collaborated and partnered with the research team in building the system and that the team was respectfully asking for their cooperation. The research team was fully aware of the limitations as to how much the participants could contribute to the design of the system; but nevertheless, the research team regarded the participants as experts developing and defining tools for their own use within their own environment (Schuler and Namioka, 1993). A demonstration of a health IVR system was presented so that the participants could see the use and the interaction that is involved with using an IVR system for the purpose of health information access. Demonstration videos have been found to be beneficial in elucidating various aspects that are involved in the use of interactive information access service particularly for low-literacy users as in the case of the OpenPhone case study (Medhi and Toyama, 2007; Sherwani et al., 2007). Participants were offered snacks and fruit juice.

The participants were asked to complete consent forms about the focus group meetings and the discussions that were to proceed. Participants who were unable to complete the consent forms themselves due to inability to read or write, were helped by the research team members. The consent form also requested the participants to furnish demographic information such as age, level of education, gender, and others. This information allowed the researchers to know the targeted population better in order to inform design decisions more accurately. The participants were notified that all the discussions would be recorded and that the participants had the right to stop participation at any time that they wished. (On the second day one of the participants decided that she did not want to participate any further and then the number of participants was reduced from 16 to 15). The researchers asked no reasons as to why she had decided to quit. This demonstrated the participants' practice of the right to autonomy, including their right to abandon the meeting and thus participation at any time for any reason.

The participants then enthusiastically engaged in the discussion and brainstorming of concerns that they felt the OpenPhone system should address. The discussion was again conducted by the two moderators and two observers were taking notes of the discussion. Participants' enthusiasm was evidenced by the fact that they would chat amongst each other

and discuss amongst themselves what they thought necessary to be made available on the proposed system before giving their views to the moderators. Naturally, some participants were more talkative than others but the moderators encouraged those who were less talkative by engaging them in the discussion and asking them what they considered important and should be made available in the proposed system. The focus groups took an average of 105 minutes each and at the end of the focus groups the participants were thanked for their participation.

Prior to the focus group meetings the research team had generated a strategy on how the focus groups with the users would be conducted. The strategy for the focus group interviews was to initiate the discussion in the form of a semi-structured interview whereby the moderators generally had a framework of broad themes to be explored. The discussion would thereafter be relaxed and opened into an unstructured form whereby the participants could converse freely around the theme at hand. This open discussion would then be gradually guided to converge towards a more specific issue by the moderators.

To be effective, focus group questions need to be open-ended and move from the general to the specific (Dreachslin, 1999).

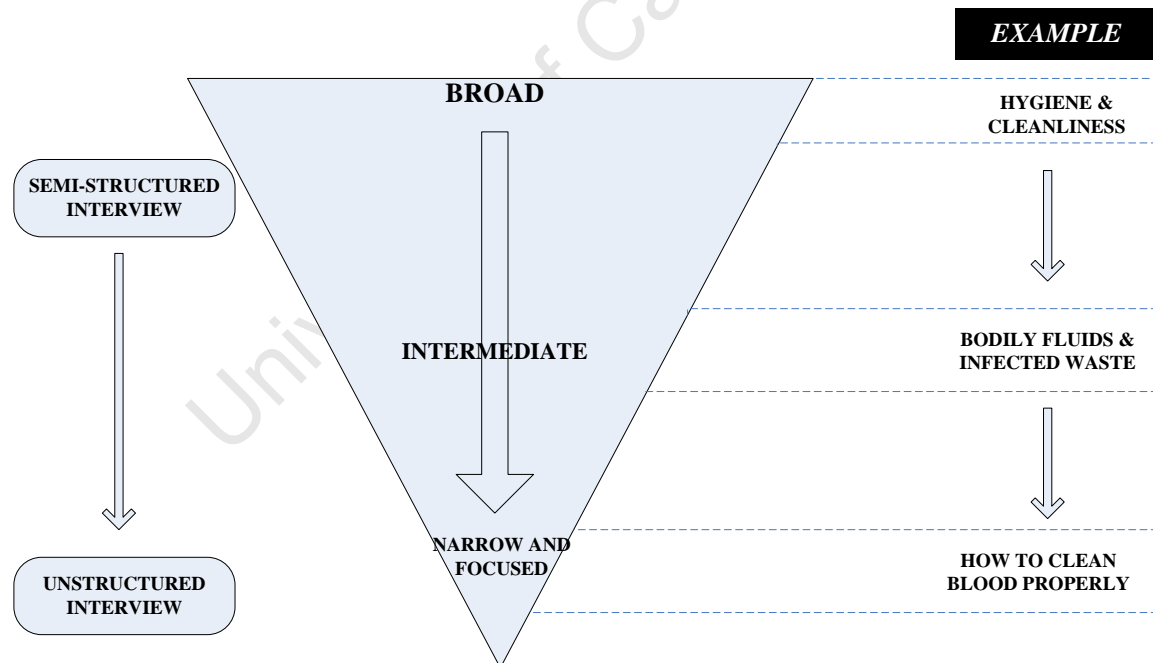


Figure 3.1: Questions generation strategy (Ndwe et al., 2008a)

Figure 3.1 depicts a graphical representation of the strategy that was used to conduct the focus groups. As an example, a broad theme of hygiene and cleanliness would be introduced by the moderators and then the participants would be allowed to discuss issues and concerns around that theme within the context of care-giving. The moderators would then direct the dialogue

towards a more focused subject of bodily fluids and infected waste through asking further questions and probing such as, “what would you like to know about hygiene and cleanliness...why?” When the participants gave answers to that the moderators would then probe more by asking questions such as, “what sort of bodily fluids are usually of concern?” This would then lead the participants to even more detailed questions such as how to clean infected spilled blood properly.

Information charts were used to collate the information that the participants were giving in the form of specific questions that they would like the system to address under each theme. The participants were then asked to prioritize the issues, through consensus, under each theme which produced a list of specific issues in descending order of importance under each theme as viewed by the participants. These specific issues are illustrated in Appendix C and discussed more in Chapter 4 as they are basis of the system dialogue design.

The research team convened after each focus group meeting to:

- Discuss the proceedings. These discussions would go over all the experiences that had been gathered during each focus group and allow the team to highlight any extraordinary elements of the meetings
- Summarize each meeting
- Analyse the summaries and write a report

3.3.2 BGR participant recruitment and requirements gathering

A preliminary user study was conducted over a period of three months to determine whether the intended target group of the BGR system would, in fact, be inclined to use such a system. Participants for the study were selected using a sample of convenience initially, and later the sampling progressed into snowball sampling technique which finds people from the target population and having gathered data from them the researchers ask about other people relevant to the research (Oates, B.J., 2006). Personal interviews were conducted in different areas of the country through a sample of convenience and respondents were asked to provide names and telephone numbers of friends from other areas who were enthusiastic about soccer.

Individual interviews were conducted face-to-face and over the phone for participants at a distance. In this way, the sample grew to include at least one respondent from each of the nine provinces of the country and the total number of participants was 43. The condition for participation was great interest in, and fondness of, soccer and those who were not interested in the game were excluded. Appropriateness for serving as an interviewee was assessed by asking potential participants questions about current soccer events and/or results of recently-

played prominent games. Only those who indicated awareness and knowledge of such events were interviewed. Participants were also requested to provide their phone numbers, so that they could be contacted for subsequent enquiries if necessary. All of them willingly did so.

The typical South African soccer fan comes from the demographic group of males in their 20s or 30s, a group that comprises a large percentage of the country's population (MarketingMix, 2009). This is in line with the selected interview participants who had an average age of 31. The researchers had no gender bias in the selection of participants, but many females declined the invitation to participate, explaining that, although they watched games from time to time, they could not be considered keen soccer followers. The appropriateness test for serving as an interviewee and the refusal to participate from some of the initial 43 participants resulted in a final sample of 25 participants, comprising five females and 20 males. All the participants were indigenous South African citizens. They came from different cultural groups and had a variety of home languages, representing seven of the nine official indigenous languages spoken in the country. Their level of education was diverse, ranging from a primary school dropout to a senior researcher with a doctoral degree.

The purpose and details of the manner in which the BGR system works were first explained to each participant. All of them indicated they had previously used some form of IVR system, most commonly for refilling prepaid airtime on cell phone accounts and interacting with the customer service of their cell phone service provider. The interviewees were explicitly balanced between participants living in urban areas and participants from rural areas, since the game of soccer is popular, and played, in both settings (Makhaye, G., 1998). Of the 25 participants, 10 participants lived in typically rural areas, 13 in urban areas, and two participants had lived, and continued to live, equally within both settings. Participants indicated that they wanted the interviews to be conducted in their own home languages. During the interviews, it was noted, however, that interviewees frequently switched languages and mixed their home languages with English, inserting selected terms and phrases.

3.4 Present coping mechanisms

This section discusses the present means that the intended users employ in order to get relevant information pertaining to the aims of the case study applications of this research.

3.4.1 Coping means for OpenPhone's intended users

The caregivers presently rely on lectures which are given every morning at the BBCCCoE, and each caregiver attends at least one lecture which is given during their first visit to the

BBCCCoE for consultation and registration with the centre. The BBCCCoE reported that subsequent lecture attendance is very low amongst the caregivers after the initial lecture session. Most caregivers attend repeat lectures only when they happen to be at the BBCCCoE for collection of medication and/or consultancy, but rarely do they ever attend lectures for the sake of reminding themselves and refreshing their knowledge on care-giving issues.

The lectures normally take about 90 minutes and the caregivers are allowed to ask questions at the end of the lecture. The lecturing BBCCCoE staff members have observed that the caregivers get to be passive learners during the lectures with some caregivers even falling asleep. During the interviews with caregivers it was found out that some caregivers travel about 70 km from their homesteads in order to get to the clinic. Their homesteads are primarily in rural areas of Botswana and travelling is normally done by combination of public transport and walking. The monotony of a lecturer speaking continuously for more than an hour and the exhausting travelling done by the caregivers before getting to the BBCCCoE partially explain the passiveness and the low follow-up lecture attendance amongst the caregivers. The effect of this mode of learning is that the content taught to the caregivers is easily forgotten. The staff members are very much aware that the present lecturing system is not an effective means of distributing the information that is relevant to the caregivers.

Easton et al. (2003), describes functional literacy as the ability to use reading, writing, and computational skills at a level adequate to meet the needs of everyday use. Many Southern Africans, while being able to identify letters in the alphabet, are unable to read or write well enough to apply this knowledge in a useful manner. This implies that even though a lot of people have attended primary schooling and are able to recognise letters and words, they are nevertheless unable to read and write such that they can be able to assimilate information through printed material. Thus, many people in the region are functionally illiterate.

Considering that most government services are currently paper based, being functionally illiterate poses a complex problem (Easton et al., 2003). Furthermore, Africa is a nation rich in oral tradition and news travels from person to person by spoken word (An analysis of African oral traditions, 2010). Ancient writing traditions do exist on the African continent, but most Africans today, as in the past, are primarily oral peoples (African Storytelling, 2010). This general oral tradition in Africa is evident in Southern Africa as well (Barnard et al., 2003a) and the tradition allows speech-oriented information to be reasonably more favourable to the intended users than printed material.

The OpenPhone system is envisioned as an intervention to the present method of ineffective lecturing that remotely instructs, reminds, and consolidates knowledge about care-giving issues for the caregivers as they cannot be furnished written material.

3.4.2 Coping means for BGR's intended users

During the interviews with potential users of the BGR system, the participants indicated the popularity of the television as presently the main medium of obtaining results of recently played games. The participants also indicated dependence on word of mouth from friends for obtaining results which is in agreement with the oral tradition discussed in section 3.4.1. Other means of obtaining results include radio and also print media such as newspapers and magazines specifically for urban residents. Television is also the most popular method of watching live games followed by radio and attending the live matches at the soccer stadia.

3.5 Summary of findings

This section discusses findings on the encounters with the anticipated users of the two systems and outlines how such encounters are expected to have a bearing on the design of the systems.

3.5.1 OpenPhone findings

Unexpected issues on social services such as government grants were brought up by the participants but unfortunately these issues cannot be addressed directly by the system on this first iteration. Issues that were contemplated by the designers as of high importance such as caregiver psychological support were perceived as of low priority by the participants. When the participants were probed about this issue they stated that they get psychological support through strong immediate and extended family support. This support can also be communal which is typical of the Tswana culture whereby family and close community members are supportive towards other community members especially in the rural areas where those communal values are still maintained.

Conducting the meeting and the interviews has had an impact on the researchers' beliefs in terms of testing the researchers' general predispositions whereby some of the information that the researchers had imagined as important to the caregivers was not viewed as such by the participants. The meetings have enabled the researchers to gain the targeted users' inputs on what their information requirements are and to eliminate unnecessary elements that the users don't need in the proposed system. The meetings have also helped the design team in building a persona which is a model user that the research team creates to help understand the needs

and behaviours of the target users. It was also established that the targeted users are predominantly females, ranging from semi-literate to illiterate adults but who are nevertheless numerically literate, and use the phone regularly. The findings allowed the researchers to construct a persona that would be mirrored by the voice talent as follows:

- Sound like a caring nurse willing to answer questions.
- Be a mother-tongue Setswana speaker.
- Have a full, mature female, well-articulated voice.
- Instil a sense of confidence and trust (Sharma et al., 2009)

(Note that we use the term “persona” consistently in this sense, as opposed to the other common usage which refers to the “personality” of the IVR itself).

3.5.2 BGR findings

The demographic information provided by the participants such as age, level of education, language preferences and others, enabled the researchers to formulate a persona for the BGR application in a similar fashion to the establishment of a persona for the OpenPhone system.

The persona of the BGR system demanded the voice talent to be:

- An indigenous African male capable of speaking fluent South African English
- Within the age range of the South African soccer fans as discussed in subsection 3.3.2.
- Speak in an excited voice because an excited tone is appropriate for lottery or sports results (Marics and Engelbeck, 1997).

Major findings of the user study were:

- All the participants indicated that during normal informal conversations they speak their own home languages, but that normal use of their vernacular usually involves an admixture of the home language with other South African languages, mainly English.
- When participants were asked in which language they would like the envisaged BGR system to address them, the majority preferred to interact with the system in English. Only four of the 25 participants desired to interact with the system in their home language.
- During the user study it was established that in their home language admixtures, the indigenous people exclusively addressed numerals such as telephone numbers, temperatures, distances, prices, etc., in English.
- The participants indicated that television is their main medium of obtaining results of recent soccer games. This is followed by radio and, to a lesser extent, print media.

Television is also the most popular method of tracking live games, followed by radio and by actually attending matches at the soccer stadia.

- The participants zealously showed interest in the system and stated that they would use such a system if it would be deployed and their main reason was finding scores to matches that have been recently played when they are travelling or at work and away from a television or radio. The participants explained that getting results through word of mouth is not always reliable as friends may sometimes prank them by providing unreliable information. They envisioned the system to be helpful in that regard as it would provide concise information about soccer results that they may have missed.

Since the participants had provided their phone numbers, the researchers contacted them subsequently, after the initial round of interviews, and tested the above language-related findings by asking simple numeracy questions such as: the distance between their residential homes and the nearest post office; temperatures in their current location; entrance fee paid to watch a soccer match; etc. It was notable that the participants all gave their responses in their home languages, but the parts that involved numerals were stated in English only. This response pattern also occurred with users who had chosen their home language as the desired medium of interaction with the BGR system. For example, when asked about the entrance price of a match, a person who speaks *isiXhosa* as their home language would say [English in italics]:

‘Ibikade *iyi-twenty rand* kodwa ngoku *yi-forty*’, which in English, translates to:

‘It used to be twenty rand and now it is forty’, where the ‘rand’ is the local currency.

This style of providing information in a language admixture corresponded with the occurrence of providing numerical information such as age and telephone numbers exclusively in English during the initial interviews. Even when asked about the days that PSL matches are played, the participants provided information in their language but the part that involves the actual days was pronounced in English. The language attitudes and admixture practices of South Africa and the consequent choice of interaction language have influenced the choice of language in which the BGR system interacts with users. Initially, during the stakeholder meetings, it had been assumed that the design would use all nine indigenous languages of the country as language diversity in the country was established as an important consideration in the design of the BGR application. In order to empirically establish the language attitudes within the context of BGR development, the researchers commissioned an experiment that would uncover and establish some generalities about the language situation in the country and how this situation influences the design of IVR systems. The results of that

experiment have been applied in the design of the BGR system and the experiment itself is discussed in Chapter 5.

3.6 Conclusions

User needs gathering is a way of animating and furnishing influential information into the design process that will have impact on the manner in which the systems are designed. In conducting real user observations the researchers' findings get to be based on realities, not preconceptions. A major cause for deficiency in information systems development methods is the intricacy of anticipating the system's use in its actual applied environment (Robinson, 1993). Users bring about things that the researchers would erroneously consider unimportant and, conversely, they also remark on the things that the researchers contemplate to be essential but users don't deem to be valuable as they have other ways of dealing with such issues and this fact is evidenced on user requirements for both systems. This supports Robinson's notion of difficulty in anticipating a system's use in its actual applied work environment (Robinson, 1993).

The contacts made with the actual prospective users enable strong relations between the users and the researchers to be formed. These relationships will, optimistically, create trust and understanding between the parties in order to devise an affiliation with common goals of designing a truly usable system.

We have found out that although the participants lack knowledge about technical matters on how to build an appropriate IVR system, they are rich in common sense knowledge about their needs and their typical concerns on care-giving and language usage issues. The researchers regard both scientific and common sense knowledge is important in formulating a holistic solution.

Personae are powerful tools for helping designers keep users in mind during the design process in the absence of the actual intended users. Benefits of creating personae in both applications are:

- Creating a persona has assisted the designers to approach the design more objectively, with their target user in mind, instead of their own views and beliefs.
- In using the persona as the target character, the designers are more capable of identifying how the caregivers will interact with the design. This enables the designer to gain an insight about the design and system usage that they wouldn't have gained in any other way.

- The persona puts all the researchers on the same page as to who the design is intended as the researchers have the same point of reference being the persona which also enables researchers to put themselves in the shoes of the target users during the design.

In sum the persona helps the researchers to make a smooth transition between user requirements and the design, which will benefit the overall design of the system. It is worth noting that the created personae are primary personae for the systems and are expected to evolve as the researchers gain more knowledge about the targeted users of the technology even beyond the parameters of present first iteration relating to this study. Based on the requirements solicitation meetings and the personae that have been created, the designers are able to turn the requirements information into functional specifications of the designs for the two applications. The researchers regard the requirements solicitation meetings to be a success as the gathered information enables the researchers to get started with the design of both proposed systems, which is discussed in the next chapter.

Chapter 4

Systems Dialogue Design

4.1 Introduction

The previous chapter presented group meetings between the researchers, the intended users and the major stakeholders for the two systems. The meetings were satisfactory because the researchers obtained both the stakeholder needs and input from the intended users which has enabled the researchers to start the design of the systems. The designs incorporate the user and stakeholder requirements in combination with researchers' knowledge on VUI design and relevant concepts are emulated from interaction design techniques in other types of interfaces such as graphical user interfaces (GUI) design. The careful design of the system dialogue is essential as it is the part of the system that the users interact with and directly experience. It is the part that most strongly influences the users' perception of the system, and how skilfully the system dialogue is designed and written largely determines the nature and effectiveness of telephony systems (Migneault et al., 2006; Stentiford and Popay, 1999). The dialogue design for both systems needs to be effective and satisfactory to the users both in context and in content (Ndwe et al., 2008c). The context concerns the structure of the system dialogue, the layout or arrangement of information within the system dialogue and the manner in which the user and the system take turns to interact with each other. The content concerns the precision and suitability of the information that is supplied by the system to the users as contained in the script and system utterances. The script is a document that contains the conversational speech sectors for each system utterance. The system utterances include any questions, comments, answers or instructions that will be given by the system to the user. The script also contains the range of possible inputs that will be accepted by the system from the user and the rules that are used to congregate the different system utterances and user inputs to form a coherent and interactive conversation between the user and the system.

This chapter discusses the development of the dialogue design of the systems in both context and content and it illustrates how the design was constrained by the circumstances and characteristics of the specific intended user populations. The characteristics and circumstances of the intended users are implanted in the personae that were created and, in turn, the personae are embedded in the researchers' consciousness during the design. One of the main objectives of the two case studies is to compare two input modalities, i.e. DTMF and ASR, and this objective necessitates the design of two prototypes for each case study application, one for the DTMF modality prototype and the other for the ASR modality prototype. The two modalities are nonetheless based on the same structure and the same

fundamental design in order to compare them even-handedly. This fundamental design is discussed next.

4.2 Description of the fundamental design

In order to impartially compare the two input modalities (DTMF and ASR) of each of the two systems (OpenPhone and BGR), the design of both input modalities is based on a fundamental and common design pertaining to both input modalities in each system. Appendices C1 and C2 represent the dialogue structure and the call flow diagram of the two systems, respectively.

Acomb et al. (2007) suggest three basic types of IVR dialogue applications:

- The first type is informational applications wherein the user interacts with the IVR system in order to obtain information on a particular subject, such as obtaining departure times from an airline company.
- The second type is transactional applications where the user interacts with the IVR system to complete a transaction such as making a reservation for a flight.
- The third type is problem solving applications where the user can do more complex interactions such as obtaining support and customer care services that normally require a human operator such as a request for a change of seat allocation from an airline company.

In the present pilot developments within the scope of this research study, the OpenPhone and BGR systems are informational application systems as the main objective for both systems is to deliver relevant information to respective system users.

The systems are user independent as they are aimed at the general public and therefore there is no login procedures required for access, meaning that the users access the system anonymously. This implies that the systems are not customisable for any particular single user since the same service is offered to all who access the system. Making the systems customisable would conceivably improve the overall service that they offer to users but this would have come at the cost of anonymity being lost and would have resulted in more complex systems to design. The login functionality would particularly not be appropriate for OpenPhone's user population as they are novice users and literacy challenged users who would find it challenging to keep passwords in combination with other login information (Prasad et al., 2008). Login functionality would also be inconvenient for the BGR in the milieu of the desirable efficiency feature as it would extend the call connection time and lengthen the navigation path through the dialogue of the system.

In the current OpenPhone system, the underlying system is a personal computer-based and low-fidelity prototype IVR system since the conversations and interactions between the system and the user are made with a telephone terminal connected directly with a personal computer and not via the public switched telephone network (PSTN). However, the caller will not be aware of this as the interface that they experience imitates that of a normal telephone communication as illustrated in Appendix D. In the BGR application, the system is a fully functional high-fidelity prototype that is connected to the PSTN so that users can call from anywhere in the world. For the sake of the research the two modalities were assigned different phone numbers - for the reasons that were addressed in the previous studies review in section 1.4, and also to make it simple for the researchers to monitor the usage of the two input modalities during the user experiments. Schaffer and Sorfaten (1999) have established that there is no difference in low-fidelity and high-fidelity tests if the main objective of the tests is for performing usability tests.

Both systems have a relatively small vocabulary with a single objective of providing, respectively, care-giving and soccer results information to those who access it. The following sections address the development of both the context and content of the dialogue systems and the design decisions that were taken by the researchers in order to make the systems more fitting to their respective users. The motivations behind such decisions are also discussed.

4.3 Context Design

The context is the structure and the layout of the information in the system dialogue and how this information is delivered by the system to the user and from the user to the system (Ndwe et al., 2008c). This section discusses the context of the dialogue design:

- An IVR system puts demand on short term memory resources through a technique of learning called chunking, which is the temporary storage of chunks of information like numbers and words. Studies have shown that humans have a capacity of chunking that ranges from 5 to 9 chunks of information at a time (Miller, G.A., 1956). From this premise the researchers took a design decision of not having more than 5 items per menu in both systems. This decision was taken in order to minimise the number of menu items that the users have to remember thereby reducing the users' cognitive load during the use of the system. It has been found that the number of items in a menu is directly proportional to errors during system usage (Engelbeck and Roberts, 1990). This minimalist approach was especially taken in the light that OpenPhone's intended users are functionally illiterate and novice users of an IVR

system, and also in the quest for efficiency in the BGR system as more menu items would have lengthened the interaction.

- The keywords that are used as user utterances or commands in the ASR modality of both systems were picked according to a set of characteristics that were part of the design decisions. The commands had to be:
 1. Short and precise. It was decided that the commands that were to be given in the ASR system menus were to be mostly a single noun whenever possible. This decision was made so that the users only had one word to remember as a command to utter to the system. This decision was taken because the more constraining a prompt (command), the smaller the active vocabulary at that point, and consequently the higher the recognition accuracy obtained and the fewer errors that will occur (Hone and Baber, 1995; Marics and Engelbeck, 1997).
 2. Intuitive in the sense that the commands were to be keywords that are clearly connected to the effect that they are supposed to produce. For example:
System: To learn about hygiene and cleanliness, say 'HYGIENE', whereby HYGIENE is the keyword or command to be uttered by the user. It is important that, "the wording of the menu choice should clearly represent the functionality accessed by that choice" (Marics and Engelbeck, 1997). An example from BGR:
System: To get results of a match that was played last Wednesday, say 'WEDNESDAY', whereby WEDNESDAY is the user utterance.
 3. Acoustically dissimilar in the sense that similar sounding keywords were not used within the same menu list (Ndwe et al., 2008c).
- Information is provided with incremental levels of detail in a similar approach to an interaction technique called progressive disclosure in GUI design where provision of information initially focuses on the necessary parts and then progressively reveals additional details only when necessary (Jones, 1989). In effect, when a user has not given any response either by pressing a key for the DTMF system or by uttering a command for the ASR system, a more detailed and longer explanation of the menu is given by the system.
- A barge-in facility was set-up in the systems, through interruptible messages and dial-ahead facilities for ASR and DTMF prototypes respectively, to enable the user to interact with the system at anytime they wish without having to wait for the menu list to be concluded before the user can have their input. The barge-in feature will also be convenient in the future for advanced users when the system is fully operational. Barge-in allows users who are accustomed with the system to swiftly navigate through the system dialogue menus in order to improve the time taken to access

information and thereby improving the efficiency of the interaction (Schumacher et al., 1995). Barge-in also supports recovery from errors as the user can barge-in and recover from an erroneous input.

- Both systems had a timeout functionality whereby the system times out if the user does not say anything in the case of the ASR system or does not press anything in the case of the DTMF system. Following a timeout the system repeats the current menu item and if the user continues to produce no response, then the system eventually times out and exits. This functionality prevents the system from being continuously in operation even when there has been no user response.
- Global keywords have been introduced into the design and the user is frequently reminded of about the availability of such keywords. These global keywords can be used anywhere and anytime during the interactions. For example, ‘*START*’ prompts the system to go straight to the main menu whenever the user utters the command. The equivalent of this global command in the DTMF modality is pressing the number ‘0’. The dialogue reiterates that when the user wants to start again, they can say ‘*START*’ or press ‘0’ in the case of the DTMF system.
- The items in each menu were listed and given to the user in descending order of importance. During the OpenPhone user needs focus groups the caregivers were asked to prioritise the items that they wanted the system to address by their order of importance as addressed in subsection 3.3.1. The order of menu presentation of OpenPhone reflects the way that the caregivers had prioritised their needs such that the more important issues are mentioned first in the menu. In the BGR system, the menu items were listed in a logical order. In view of the fact that menu items were three different days of the week, the order was made to be follow the culturally standard order of week days, i.e., Wednesday first, Saturday secondly and Sunday thirdly.

4.4 Content Design

The content concerns the precision and suitability of the information that is supplied by the system to the users as discussed next:

- The content in the dialogue script for OpenPhone was first verified by BBCCCoE, the major stakeholder of the system as discussed in subsection 3.2.1. The content in the BGR system was improvised for the purpose of research and did not reflect genuine scores for matches that have actually been played. (The researchers concluded that improvised results would serve the objectives of the research better than real results. The tasks that the users were required to perform during the user experiments

involved retrieving the results of particular games as provided in the system and giving answers to indicate the users' ability or inability to retrieve results from the system. To avoid users who would give the right answers for the wrong reasons, the results had to be improvised so that the users participating in the experiments would not furnish correct results just because they know the real results from games that have been actually played. The correct responses to the tasks were the results provided by the system that can only be correctly answered by navigating the system and performing the required tasks properly).

- The choice of language was such that the commands and the information contained in the OpenPhone system was in a language that the caregivers understand and avoided technical terminology that is used by medical professionals. For example, an immune system does not have a direct translation in the Setswana language and an alternative explanation of the immune system is essential since the immune system is pivotal in the HIV/AIDS illness. The nurses at the BBCCCoE who give the lectures to caregivers suggested that a term that directly translates in English to 'soldiers of the body' is well understood by the caregivers to mean the immune system as it is the system that protects the body against infections hence the term 'soldiers of the body'.

Several studies suggest that prompt messages should be simple and natural, should be expressed in a familiar language to the user, should minimize the user's memory load, and should not use different words to mean the same thing (Delogu et al, 1998a).

In the BGR system, the language used was English for reasons explained in subsection 3.5.2. The content in the BGR dialogue script is very limited as the aim of this pilot design is to probe certain user-interface issues, by providing only the results of recently played games and no further commentary about how the games were played. The information in the BGR system was indexed by the days that PSL soccer matches are played, i.e., Wednesday, Saturday, and Sunday. Soccer fans know very well that matches are only played during those days.

- A recorded voice was used for both systems instead of a synthetic voice because listening to and comprehending synthetic voices is more difficult than with natural voice. A synthetic voice requires more cognitive effort than listening to natural voice (Delogu et al., 1998b). Connie Ferguson, who is a well known actor in a popular soap opera in the local television broadcast network, was used as the voice talent for recording the voice in the OpenPhone system since the system has far more content when compared to the succinct BGR. Connie Ferguson is a female and her age is within the average age range of the caregivers. She speaks fluent Setswana in the local accent. This voice talent decision was based on the desire to employ a voice that

is familiar, in local accent, female as most caregivers are, in the same age range as the caregivers and a person who is a professional as a voice talent. The voice for an application should be selected to match the context and purpose of the interactive application (Marics and Engelbeck, 1997). The voice talent offered her services for free for the voice recording of both modalities of the OpenPhone system. Soap operas are very popular in Botswana and the use of a familiar voice is intended as a crowd puller for the system. An indigenous African male voice actor within the average age range of South African soccer fans was used for the BGR since the game of soccer is historically a masculine sport (Pelak, 2005; Mennesson and Clement, 2003; Bryson, 1987). The soccer sports commentators on both television and radio are significantly male, resembling the participants of user studies in subsection 3.3.1, where 80% of the interviewees were male. An excited tone was used as this is the appropriate voice for sports results (Marics and Engelbeck, 1997).

- Only one voice was used in each application because using different voices disrupts the flow of the application and focuses the users' attention towards the voice rather than on what is being said (Marics and Engelbeck, 1997).
- The dialogue systems follow a pattern of first introducing the effect and then the action that should be taken in order to bring about the effect instead of the other way around. For example, the system prompts are as follows:

System: For advice on washing your linen, say 'LINEN'; instead of:

System: Say 'LINEN', for advice on washing your linen.

This is because it has been found that users tend to better understand the critical information of the message if it is in the final part of the sentence rather than in the initial part (Delogu et al., 1998b). The same principle was used in the DTMF system as in the following example:

System: To get results of a match played last Wednesday, press '1'; instead of:

System: Press '1', to get results of a match played last Wednesday.

- An introduction is provided under each menu item so that the user is aware which topic they have selected and the topic that is about to be discussed. This allows the users to know exactly where they are in the dialogue so that if the system is about to discuss a topic that is not wanted by the user they can always start again.

4.5 OpenPhone and BGR telephony prototype systems

For the sake of unbiased comparison between the two modalities (DTMF and ASR) in each of the case study applications (OpenPhone and BGR), the design of both modalities has been deliberately kept the same including the system prompts, hierarchical structure, system output

voice, system output speech rate, and others. The next subsections discuss the two input modalities of ASR and DTMF. The ASR system was simulated with a Wizard of Oz (WOZ) system for the OpenPhone system, and the reason for this substitution is explained.

4.5.1 WOZ Prototype system in OpenPhone

The speech-enabled IVR system under investigation employs ASR which is a speech technology that is somewhat error-prone even in extensively researched languages such as English. The likelihood of recognition errors is more pronounced in less prominent languages such as the OpenPhone users' native language of Setswana. Another fact to consider is that OpenPhone has relatively more user input options that are expected from the user and a relatively larger vocabulary than the concise BGR system. It has been established that the probability for error is directly proportional to the active vocabulary at any given point in the dialogue (Hone and Baber, 1995; Marics and Engelbeck, 1997), which makes OpenPhone more prone to error when compared to BGR which has much less active vocabulary in its dialogue.

At the time of the development of OpenPhone, the ASR technologies for a speech-enabled IVR system in target language of Setswana were still work in progress. The imperfections and limitations of ASR technology would obscure the experimentation of comparing the two modalities since the user utterances would not be guaranteed to be correctly recognised by the ASR modality. This malfunction would compromise the accuracy and integrity of the experimentation since the experimental conditions might favour DTMF due to the inaccuracy of the speech recognition engine in the speech-enabled ASR version of the system. In the WOZ prototype the speech recognition errors are avoided since a human being recognises the speech inputs made by the users instead of an ASR engine. Also, considering the fact that, during the OpenPhone usability tests, whereby the participants tended to speak in a low and slow voice that was unconfident, a real ASR system would have produced more recognition errors which would have compromised the integrity of the experimentation as discussed.

Appendix D shows a graphical representation of the how the WOZ prototype functions. The user (experimental participant) instructs the system through voice commands and listens to the system's response through a normal telephone interface. This is done in order to make the experiment as close as possible to that of making a normal telephone call. The wizard, who is obscure to the user, listens to the user commands and responds by selecting the appropriate responses to be played by the system in the same way that a fully functional system would respond. The user is not aware that it is the wizard who is controlling the outputs that the user hears as the user thinks that they are interacting with the system directly without human

intervention. In our version of WOZ there are no speech recognition errors as the wizard is a human being who understands human utterances even if they have not been properly articulated and therefore the WOZ is always accurate in recognizing the user's inputs. (If one wishes to study the effect of recognition errors, this protocol could obviously be adapted.)

4.5.2 BGR ASR prototype system

In contrast to the OpenPhone WOZ system, the BGR system uses a fully functional ASR system in which the users use verbal commands to interact and control the system which are automatically recognised by the system. These verbal commands are chosen from a given list in the system's menu.

4.5.3 OpenPhone and BGR DTMF prototype system

In both applications, the DTMF modality allows the use of the telephone keypad as the sole input device and the modality utilises a menu dialogue strategy. The DTMF system output presents a menu that instructs the user to press a particular number on the phone keypad that corresponds to a particular effect. For example:

System: To learn about hygiene and cleanliness press 2, to learn about nutrition press 3,...

The user then reacts by pressing whichever number that corresponds to the task that they want to carry out.

4.6 Conclusions

The personae in the researchers' consciousness have helped to keep the researchers' foci on who they are designing for. Instead of asking questions such as, "how would I use this system", the researchers were asking, "how would the persona use this system". The personae have been instrumental in the design by ensuring the incorporation of the intended users' environment and circumstances in the design of the dialogue systems for both applications and the prototypes in each system. The designs have amalgamated VUI design principles together with the aforementioned knowledge from the personae and also adapted relevant principles from the design of other types of interfaces. The designs of the two prototypes for each application have prepared the researchers to embark on the usability tests for the prototypes of the applications which are discussed in the next chapter.

Chapter 5

Experimental Design and Procedures

5.1 Introduction

In the preceding chapter the design of the OpenPhone and BGR systems was presented together with the various design decisions that were taken in order to ensure that the systems are appropriate for and usable by their intended users. The logic behind the design decisions was explained and the structure of the system dialogue design was discussed. The design has produced two prototypes for each system which were tested with representative users in Botswana and South Africa. The two prototypes are a DTMF and a speech-enabled system for each case study.

This chapter presents the design of experiments and the procedures that were followed for the usability tests with prospective users of the two systems. There were three main objectives for administering the proposed experiments:

- To evaluate the usability of each system through experiments that involve anticipated users of each system in order to improve its usability.
- To subjectively and objectively compare the two IVR input modalities of DTMF and ASR for the proposed applications in the milieu of the aforementioned user groups' characteristics.
- To establish whether the systems are perceived as a useful and usable technology that may have a positive bearing on the targeted user groups.

Usability evaluations are a crucial part of development in UCD as they provide invaluable information about the interaction between the users and the technology.

Evaluating usability is now considered an essential part of the system development process and a variety of methods have been developed to support the human factors professional in this work (Dillon, A. 2001).

This chapter also presents various methods for usability evaluation and justifies the researchers' choice of methods.

5.2 Usability evaluation methods

There are generally 3 types of usability evaluation methods (Jacko et al., 2003):

- I. Inspection: this method is an expert evaluation method that involves developers and usability experts in scrutinising usability related features of the product's interface.

Folmer et al. (2003) argue that for current software systems, most usability issues are detected during testing and deployment. Fixing usability issues at such late stage of the development proves to be very costly (Folmer et al, 2003). It is for this reason that research team of the OpenPhone and BGR systems has utilised the inspection method as it occurs prior to the development of a full prototype, after the design has been produced. The method is highly economical as it only involves the developers and the usability experts in the evaluation without any users. There are various forms of this method including cognitive walkthroughs, heuristic evaluations, feature inspections and others. The heuristic evaluation inspection method is used in this study because it is the most common informal method and it involves having usability specialists judging whether each dialogue element follows established usability principles (Nielsen and Mack, 1994). The particular principles that are used in this study are the Usability Heuristics by Nielsen (1994) which are popular and well established principles that are widely used in inspection evaluation.

- II. Inquiry: this method involves observing users interacting with the product in real usage conditions such as in a workplace situation. The observer pays attention to the user's needs and what they like or dislike about the system under evaluation through interviews, observations, surveys and other enquiry methods in real work settings. It is a less formal method when compared to the usability testing method (Jacko et al., 2003). However, this method is not appropriate for the objectives of this study. The OpenPhone and BGR are at the first development iteration and have not reached launching stage whereby they can then be tested in real usage environment through the use of an inquiry test method. In further developments, beyond the objectives of this study, the method may be used to evaluate later versions of the systems.
- III. Testing: this method involves testing the existing prototypes of the systems with the anticipated users either in field trials or in a usability laboratory. The method has been used widely in quantifying usability of products and it is "generally considered to yield the most reliable and valid estimate of an application's usability" (Dillon, A., 2001). The method's reliability is the main reason that this study utilises this method more extensively than any other method. Furthermore, the testing method includes a form of inquiry in the sense that the test participants are interviewed after the test as to what they like or dislike about the product for the purpose of establishing their satisfaction with interacting with the product. Also, the participants are gauged as to whether they feel that the product would add value to their lives and if they would use the product in real life.

The next section discusses the inspection as an evaluation method by using Nielsen's usability heuristics.

5.3 Heuristic evaluation

This study relies to a lesser extent on heuristic evaluation (HE) than on the testing method as HE is done by the researchers and therefore is biased towards the researchers' opinions which may be far different from the intended users' point of view. As stated in Chapter 1, the research values the intended users' point of view more than any other view on the usability of the systems. Nevertheless, the heuristic evaluations provide the researchers with an opportunity to do introspection about the systems that they have designed.

Four researchers were independently given the task of evaluating the OpenPhone system according to the Nielsen's 10 heuristics that are general principles for user interface design (Nielsen, 1994). Each researcher's task was to evaluate the OpenPhone system design as to how far it complies with the 10 design principles. Thereafter the researchers had a meeting and debated the heuristics and then produced a reciprocated summary of how OpenPhone complies with each of the 10 heuristics. For the BGR system, a team of eight researchers held a meeting and brainstormed around the issue of how the design complies with Nielsen's 10 heuristics. Nielsen's heuristics are general principles in UI design and not specifically for VUI design and therefore some of them had to be slightly reinterpreted and adjusted to fit within the framework of VUI design. As an example, appendices B1 and B2 summarise the researchers' opinions about the OpenPhone case study designs in the milieu of the 10 heuristics. (BGR heuristics are similar to OpenPhone). Each heuristic is presented, then a definition of that particular heuristic is given, followed by the manner in which OpenPhone comply or flout that heuristic.

The heuristic evaluation gave the researchers a fair amount of confidence that they were on the right track as most of the heuristics were adequately met by the OpenPhone and BGR designs. Those heuristics that were not adequately addressed by the designs gave the researchers the motivation to redress the designs such that the final designs were in accord with the heuristics prior to the development of full prototypes. Nonetheless, the researchers knew that the heuristic evaluation was only the beginning and the true evaluations were by means of imminent usability tests with intended users after the development of the prototypes.

5.4 User testing

It has been previously mentioned in section 2.2 that the ISO standard has been adopted as the main definition that will guide the way that usability is engaged and measured in this study, and therefore, the standard is used as the key reference in conducting the measurement of usability in the usability tests. In conjunction with the definition of usability the ISO standard

also provides a framework that demonstrates the relationship between the different usability components and characteristics as discussed in section 2.2. This framework is graphically presented in Figure 5.1.

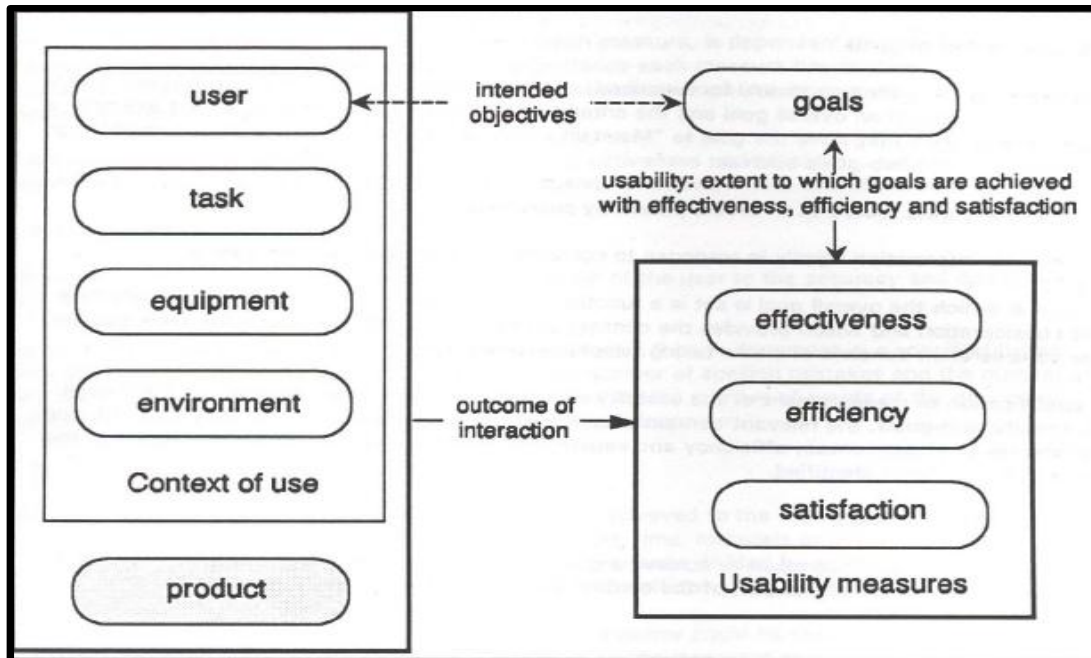


Figure 5.1: Usability framework (ISO 9241-11, 1998)

The ISO standard definition and the framework in Figure 5.1 demonstrate that the units which need to be measured are effectiveness, efficiency and user satisfaction that the users experience as they execute tasks in order to achieve their goals. The ISO also categorizes the measures of efficiency and effectiveness as performance measures or objective measures. The measure of satisfaction is categorized as attitude or subjective measure. The objective and the subjective measures are independent of each other but are complimentary and this study makes use of both measures.

The complementarity between objective and subjective measures also leads to the fact that usability can only be established through the simultaneous measurement of both aspects (Larsen, 2003).

In addition to the characteristics of usability as defined by the ISO standard, the extension of learnability by Nielsen (1993) and Shackel (1991), as described in section 2.2, is integrally used as a directive of what to measure in the usability tests. The following paragraph describes the events that occurred during the usability tests of the two input modalities of both applications. Another experiment, which was first introduced in subsection 3.5.2 with the aim of validating the way that indigenous people of South Africa pronounce numbers, is also discussed.

5.5 Experiments

This section describes the various experiments that were conducted including the main usability tests that compare the two modalities of DTMF and ASR for the two case studies. The tests comprise an experiment for the effects of a culturally inclined earcon and an experiment for establishing language preferences when reading numbers. The number of participants in the experiments is also discussed at the end of this section.

5.5.1 Main usability tests

The participants were first thanked for their participation. For OpenPhone there was a WOZ operator who controlled the WOZ system as illustrated in Appendix D. The BGR required the participants to call two distinct telephone numbers over the PSTN, one for the DTMF prototype and the other for the ASR prototype. The tests were done from 21 April through 25 April 2008 for OpenPhone and 1 December through 11 December 2010 for the BGR. The researchers emphasised that they were not testing the participants but rather how the systems perform for the intended users and if the test participants got any of requested tasks incorrect, then that would simply imply that something needs to be changed on the relevant system. This was explained in order to ensure that the users would not interpret the tasks as a test of their abilities or inabilities to use the systems but to determine how the systems could be improved, and also the fact that an incorrect task was viewed to be as good as a correct one. All communications were conducted in the indigenous language of the test participants which was Setswana for the OpenPhone and the various languages of the indigenous African people of South Africa for the BGR.

For the OpenPhone application, a demonstration video that illustrated how an IVR system works within the context of health information access was shown to each participant before they could engage in the tasks. The video had been prepared beforehand by the researchers and it featured a demonstration of people using a phone to access information about giving care to HIV/AIDS infected children. As the caregivers were not acquainted to the use of an IVR system, the purpose of the video was to demonstrate the abilities of such a system in a visual manner, and to build an expectation from the caregivers of what to anticipate out of the experience with the system. Demonstration videos have been found beneficial in clarifying the use of interactive information access services particularly for low-literacy users as in the case of the OpenPhone case study (Medhi and Toyama, 2007; Sherwani et al., 2007). The participants were asked if any of them would like to watch the video again but none of them

required so. There was no need for a demonstration or extensive explanation for the BGR application because the participants were accustomed with the use of IVR systems for information access. The BGR participants were briefed on the purpose of the study and the tasks they were being requested to perform. The participants were also informed about the operation of the system – e.g., that the information on the system was indexed by the different days that PSL games are played and that information therefore was to be primarily accessed according to the day that a specific match was played. Since the BGR was a live application that was accessible over the PSTN, it was possible for remote participants to participate in the experiment unlike OpenPhone which was only tested within a laboratory environment. All the prototypes were instrumented to log participants' interaction activities including task navigation paths, task completion times, timeouts, barge-in activities, and others. The facial expressions and gestures of participants that were tested in the vicinity of the researchers were observed and noted, i.e., all the OpenPhone participants and those BGR participants that were tested in the presence of the researchers.

For both applications there were two test stations (two communication lines in BGR) for the experiments, one for the DTMF modality session and the other for the ASR/WOZ modality session. The participants were asked to participate in both the DTMF and ASR/WOZ modalities. In total there were four sets of tasks (A,B,C, and D) that could be done by the participants and out of the four each participant was required to do only two sets. The task sets for the OpenPhone and the BGR systems are shown in appendices F1 and F2 respectively. Each set had two tasks within it (task 1 and task 2) and all the sets were different from each other in terms of the combination of tasks that were performed. This made a total of four tasks that were performed by each participant (two tasks in each set of two sets). The first set was performed during the first session and the second set was done in the second session. Half of the participants were asked to start with DTMF as their 1st session and the other half were asked to start with ASR/WOZ as their 1st session.

In order to distribute the task set execution evenly amongst the participants, the task sets were executed in a rotating manner. For example, the first participant did the 1st session in DTMF using set A tasks and the same participant was then asked to do the 2nd session in ASR/WOZ using set B tasks. The second participant was asked to do 1st session in ASR/WOZ using set B tasks and the same participant was asked to do 2nd session in DTMF using set C tasks. The third participant was asked to do 1st session in DTMF using set C tasks and the same participant was then asked to do the 2nd session in ASR/WOZ using set D tasks. The fourth participant was asked to do 1st session in ASR/WOZ using set D tasks and the same participant was then asked to do 2nd session in DTMF using set A tasks. The fifth participant

followed the same procedure as the first participant, and the sixth followed the same procedure as the second, the seventh as the third and the rotation carried on in that manner. All the tasks were independent of each other in the sense that no task depended on the accomplishment of any other task.

The above experimental procedure of half-and-half crisscrossing of task execution combined with the rotation of set tasks was done in order to make sure that:

- The same number of tasks was performed for each of the modalities.
- Half of the participants started with DTMF modality and the other half started with the ASR/WOZ modality.
- There was an even distribution of tasks performed between the 4 possible task sets.

In each set, the first of the two tasks was designed to be easier than the second task and the participants started with the easier task and then progressed to the slightly more difficult task. The 2nd tasks were more challenging and engaging in the sense that they needed the participants to pay more attention and have more patience as the task performance required the participants to access information that was mentioned later in the system dialogue and deeper in the dialogue structure.

Figure 5.2 is a graphical representation of the above task execution.

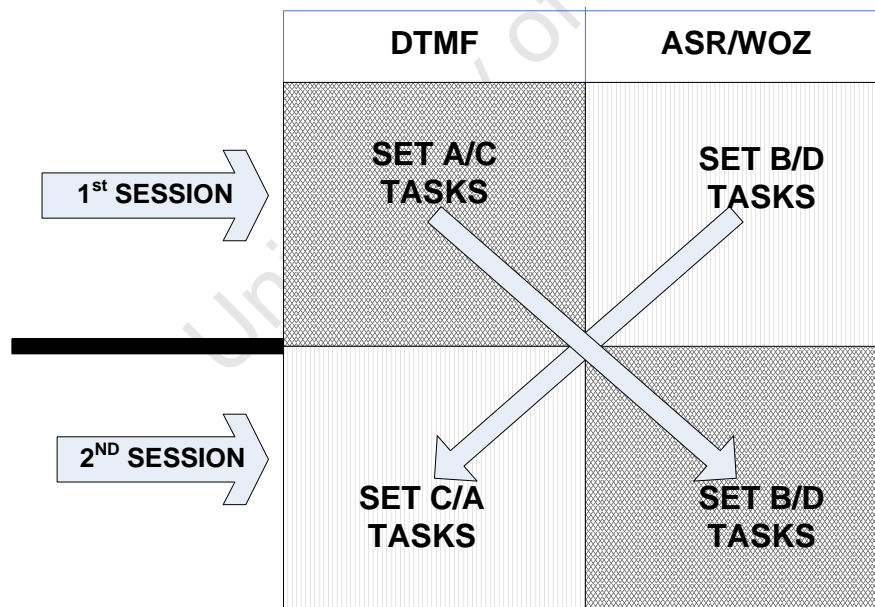


Figure 5.2: Graphical presentation of planned task execution

In-between the sessions, questions about the participants' experiences on the 1st session were asked and questions about comparing the two modalities were not asked since they had interacted with only one modality by that time.

Over the entire five days a sample of convenience of 35 participants were recruited from the caregivers who visit the BCCCoE during the weekdays for OpenPhone and any caregiver who was willing to participate irrespective of age, gender, or any other characteristics was welcome to participate. For the BGR application a combination of purposive sampling and snowball sampling whereby relevant participants, [indigenous people of South Africa within the typical soccer fan age range of 20s and 30s (MarketingMix, 2009)], were picked and, after having gathered data from them, were asked to provide suggestions about other people relevant to the research (Oates, 2006). Cluster sampling was used for making sure that participants with different indigenous language backgrounds from different provinces of the country in both rural and urban areas were represented. Participants were asked a few demographic questions which requested their age, level of education, gender, and home language and this information was very much aligned with the demographics that had been previously obtained during the user requirements solicitation focus groups in Chapter 3. Initially there were 31 participants that were recruited for the BGR tests. The tests did not always go according to the plan as demonstrated in Figure 5.2 mainly due to some technical errors in the prototypes. In the BGR, some callers ran out of cellphone battery power during the tests and others hung up because of other commitments during the remote tests. This reduced the initial number of recruited participants on both systems.

There were a total of 33 participants for OpenPhone who performed a total of 103 tasks between them. Some of the OpenPhone participants were only able to use one of the prototype systems due to various reasons including mostly technological malfunctions with the prototype systems and sometimes participants who had to leave due to other personal engagements. Out of the total 33 participants, only 22 participants experienced both prototypes and only those participants' results are employed for the direct contrast between the two prototypes as discussed in the next chapter. For the BGR system the number of participants was reduced to 27 participants and these are the participants who are used in the comparison of the two modalities in the next chapter as they performed tasks in both modalities. Nine participants performed their tasks in the presence of the researchers and 18 performed their tasks remotely. The participants covered speakers of all the nine indigenous languages of South Africa and came from all the nine provinces of the country. Six of the users were female. It was important for some of the BGR participants to execute their tasks remotely as the application needed to cover participants from both urban and rural areas in all nine provinces of South Africa. It would have been very costly for the researchers to be physically present with the participants in all nine provinces. A substantial number of the participants in the BGR were those that had been initially recruited for the user requirements gathering process as stated in subsection 3.3.2. During requirements gathering participants

were requested to provide their phone numbers so that they could be contacted for subsequent enquiries. Those participants were happy to find out that their initial contributions had culminated in the BGR system and they were delighted to participate in the usability tests for the BGR that they already knew about. The BGR participants were categorised by their level of education into three bands as follows:

- Lower band education (LBE) participants ranging from no formal education up to primary school education level.
- Middle band education (MBE) participants who are between primary and high school education levels, i.e., above primary and below high school.
- Higher band education (HBE) participants who have high school and tertiary education or professional experience.

Out of the 27 participants, nine were from the LBE, 11 from MBE, and seven from the HBE.

5.5.2 Vuvuzela earcon experiment

In the BGR experimental design an earcon was introduced in order to assess the effects of an expressive signature tone in the UI of an IVR system, within the Southern African context of oral people and this was done with the use of a vuvuzela signal. A vuvuzela is a vociferous air horn that is blown by most avid South African soccer supporters and it is synonymous with the vibrant atmosphere at South African soccer matches (Swanepoel et al., 2010). The sound of the instrument being blown at soccer games has evolved into an emblem of hope and unity for many South Africans and it is a proud and permanent symbol of its patrons (FIFA, 2009). FIFA approved the vuvuzela as part of the signature South African World Cup 2010 which was recently held in the country. The motivation behind the vuvuzela earcon experiment was from the idea expressed by Sharp et al. (2007) which states that:

One of the benefits of using expressive embellishments [in a UI] is that they provide reassuring feedback to the user and can be both informative and fun. They can, however, sometimes have the opposite effect on people, who find them intrusive, causing them at times to get annoyed and even angry (Sharp et al., 2007).

Sharp et al.'s (2007) idea is based on all kinds of user interfaces, including GUI, and the researchers wanted to find out the effects of culturally inclined earcons within the given setting of oral users.

A brief vuvuzela tone that lasted for three seconds was introduced at the beginning of the introductory prompt of the BGR system. The playing of the vuvuzela tone was programmed to play randomly half of the time that user accesses the system with an equal 50% probability of getting it played or not played. This meant that in the four tasks that the participants were

requested to perform during the BGR usability tests, each participant had a probability of hearing the vuvuzela twice. The questions that the participants were asked at the end of each experimental session included questions about their experience on the vuvuzela sound as demonstrated in Appendix H. Results of the participants' experience with the vuvuzela is discussed in the next chapter.

5.5.3 Experiment on language preferences when reading numbers

As indicated in subsection 3.5.2, BGR users exhibited an unexpected preference of English as their choice of interaction language with the proposed system. The researchers conducted a survey to authenticate these findings, which came as a surprise to the researchers as they had geared themselves to cater for the language differences between the indigenous peoples of South Africa. The survey was entitled the 'numbers pronunciation experiment' and was aimed at establishing the way that numerals are pronounced by the indigenous peoples of South Africa as the outcomes of such research have a meaningful impact in the way that IVR systems are designed for the region, specifically IVR systems that have numeric information within their content and context as in the case of the BGR system. The participants for the experiment were recruited initially through purposive sampling in public areas of various cities and towns in South Africa, in work places, in shopping malls, in traditional gatherings of rural areas and various other public places and the only qualification to participate in the experiments was to be an indigenous South African citizen. The experiment was conducted over two months prior to the BGR usability experiments and was also extended to include the participants in the usability tests. Except for the BGR usability participants, who were mostly male, the experiment was balanced between male and female both from urban and rural areas of the country and included adult participants without regard for age who were interviewed remotely and in person. Data from 72 participants around the country including those participants in the BGR usability tests was collected for the experiment.

In person participants were randomly shown numbers of one to three digit numerals on a laptop screen, on a cellphone display and on printed paper and the researchers were asking the questions, "what is this" and "what language is it"? The order of showing the numbers to the participants on the three different interfaces was randomised and did not follow any particular order. The researchers used the three different interfaces for methodological triangulation (Denzin, 1970), and thereby eliminating any bias towards any particular interface as the participants may have been inclined to pronounce numbers on a certain interface differently on other dissimilar interfaces. Remote participants were sent a message (SMS or text message) that contained a number and then interviewed telephonically and asked what was in

the message they were sent and what language was it in. All conversations with the participants both in person and remotely were conducted in the participants' spoken language of choice which were the various indigenous languages of South Africa. After the two questions each participant was also asked what language did they learn to count and what language did they do their arithmetic subjects. The results of the numbers pronunciation experiment are discussed together with the OpenPhone and BGR experimental results in subsection 6.3.6.

5.5.4 A note on the number of experimental participants

Early studies concerning the number of users that should be used in conducting usability experiments that were conducted by Virzi and later by Nielsen have suggested that five test users per iteration can reveal an average of 80% of all the usability problems that exist in a user interface (Virzi, 1992; Nielsen, 1993). The problem with the assumptions made by Virzi and supported by Nielsen is that the range of problems revealed can be anywhere from 55% up to 99% (Faulkner, L., 2003), which implies that there is a possibility that almost half of the usability problems can be missed. A usability study was conducted by Faulkner whereby 60 users were tested and random sets of five or more were sampled from the whole of 60 in order to demonstrate the risks of using only five participants and the benefits of using more. It was found that the range of usability problems that were uncovered for five test users was 55% to 99% with an average of 80%. For 10 test participants the minimum value of the range was increased to 82% with an average value of 94%, and for 15 users the minimum value was increased to 90% with an average value of 97% of problems being uncovered (Faulkner, L., 2003). The experiment showed the benefits of increasing the number of test participants in an experiment.

The common barring factor that usability professionals usually encounter in recruiting more test participants is that of budget constraints from the funding company (Faulkner, L., 2003). In the case of this study the test participants are the same people who will conceivably be using the systems that are being tested. The users were willing to participate in the experiments in the same way that their enthusiasm had been shown during the requirements gathering focus groups and did not require payment for participating in the usability tests which helped to keep the expenses for the tests low. The numbers pronunciation experiment and BGR tests needed to encompass participants of different language and cultural backgrounds from all over the country which increased the number of participants needed for the tests. It is also important to remember that the participants in all the experiments of this study were not compensated for their participation but the researchers relied on the

willingness of the participants. This is unlike user studies in the developed world whereby participants are compensated in accordance to the time that they spend on the experiments. The dependence on the participants' enthusiasm makes it difficult to know before hand the exact number of participants that will be involved in the experiments and also brings in the reality that participants who have been recruited in this manner may quit the experiment for any reason at any time and therefore it is better to aim for a higher number of participants to accommodate the unknown number of participants that may fall through. The eventual numbers of participants in experiments of this study were adequate in meeting usability test goals.

5.6 Conclusions

The experiments that have been described in this chapter have allowed the researchers to evaluate the usability of the case study applications and to formulate some realities about the language attitudes and the effects of culturally inclined earcons in the targeted user population. These findings have a bearing on how IVR systems for the targeted users should be designed. The researchers believe that getting more than required number of users is advantageous so that when some of results are inadmissible due to the matters that surround conducting usability tests in the target area, the researchers can still have enough results for analyses. The next chapter discusses the results that were obtained from the experiments.

Chapter 6

Results and Analyses

6.1 Introduction

In this chapter the results that were generated by the experimental procedures as presented in previous chapters are analysed and discussed. The findings that are presented are both the objective and subjective results which culminate in conclusions and deductions that are based on the findings. The difference between the subjective and objective forms of measurement were introduced in section 5.4 and the next section further clarifies the rationale behind the types of measurement that were used in this study.

6.2 Types of measurements

The objective measurements are quantitative factors that were obtained from empirical measurements and the subjective results are qualitative representations of the intended users' sentiments and satisfaction levels towards the prototypes. The next subsections discuss the measurements that were used in each case study application.

6.2.1 OpenPhone measurements

The objective measurements that were used are:

- Effectiveness of the system, as defined in Table 2.1, and measured as the success rate that is presented as the ratio of successfully completed tasks in relation to the total number of tasks.
- Efficiency of the system, as defined in Table 2.1, and measured as an objective measure of time taken to correctly accomplish the given tasks. In the OpenPhone system the measurement of efficiency was not computed in the same way as in the BGR system (and in other similar studies (Delogu et al., 1998a; Foster et al., 1998; Karis, D., 1997, Patel et al, 2009)), because:
 - For this particular first iteration the researchers were interested in establishing whether the intended users were able to use the IVR system technology to access information and were less concerned with the efficiency issues such as task completion time, number of steps taken to complete tasks, or whether the test participants took the shortest route in navigating towards the completion of the task. As the intended users of the OpenPhone application were novice users of IVR systems the researchers did not view efficiency as a critical factor and were

only interested in knowing if it is possible at all for the intended users to use an IVR system to access healthcare information.

- The nature of the information provided by OpenPhone contributed to the reduced emphasis on efficiency in the system. Care-giving information, within the context of this application, is information that is to be understood properly as any misunderstanding may have harmful consequences upon the patient's well-being. Even for an expert user of such a system, it is more desirable that the user should proceed with care in order to make sure that the appropriate information is accurately accessed instead of a swift and efficient navigation through the system that may lead to inaccurate task completion. For novice users, a slower navigation through the system dialogue becomes even more imperative and therefore, the success rate (effectiveness) was more valuable to the researchers than the time taken to complete the tasks (efficiency) for this application.
- The pattern of use for the OpenPhone system is envisaged to be evenly distributed throughout its use because it is not anticipated that there will be particular periods whereby the caregivers suddenly have the urge to call the system at the same time. The expected number of callers at any given time to the system is relatively low when compared to the BGR application as the OpenPhone is targeted at the patrons of the BBCCCoE and the researchers did not expect a high demand on the system resources that would justify the researchers to prioritise efficiency in the system design, as in the BGR case for efficiency that is advocated in the next subsection.
- Most significantly, using task completion time would have produced inaccurate results in OpenPhone's experimental setting. The measurement of time taken to finish the task was automatically done by the software within the two prototypes as human intervention would introduce errors. It became apparent that these results were meaningless as many of the users had a tendency to go an extra mile with their tasks by repeating the same tasks again and by accessing other information which was not part of the task requirement once they realised that they had figured out how to use the system, especially those users who got their first tasks incorrect and got their subsequent tasks correct. Even when the participants had found the answer to the given task, during the usability tests, they would continue to listen to the rest of the dialogue and not hang up to present the answer to the researchers. This resulted in OpenPhone prototypes presenting task completion times that were not genuine and the researchers could not rely on such measurements. The researchers deemed it to be improper to interrupt whilst the participants were engaged in their superfluous tasks. The

average age of the researchers was less than that of the participants and such interruptions could have been viewed as lack of respect within the local cultural norms, particularly when they come from younger people. Secondly, the participants had been requested to participate in the tests without any monetary rewards for the participants. This procedure was also unconventional when compared to other similar studies where test participants are often compensated for participating in tests. Conceivably, the researchers might have had more control in the way that the tasks were executed and would have applied stricter regulations towards task completion if the participants were to be paid for their involvement. Thirdly, the researchers did not anticipate the tendency for superfluous tasks and extended interactions with the system and it literally caught the researchers by surprise. Even though task completion times were not the main aim of the study, the researchers had prepared themselves to obtain these measurements hence the prototypes were instrumented to capture task completion times.

Efficiency in OpenPhone has been approached as a subjective measure of comparison between the attendance of lectures and visiting the clinic in order to get advice on care-giving matters in contrast to using the system to get advice. This corresponds to the definition of efficiency which defines efficiency as the resources expended in relation to the accuracy and completeness with which users achieve goals as indicated in Table 2.1. In the case of the OpenPhone system the resources expended are a phone call instead of the time and money expended in travelling a maximum of 70 km to the clinic. The information that the caregivers receive over the phone is accurate and specified according to their needs instead of an entire lecture on various issues concerning care-giving which they usually find difficult to retain. Clearly, from the subjective measurement point of view, the OpenPhone system is more efficient than the present system of attending lectures at the clinic and this perception was expressed by the participants as discussed in subsection 6.3.3.

For the subjective measures, at the end of each experimental session, users were asked to answer questions that were contained in a questionnaire that asked for their perception of the quality of interaction, of their effort in interacting with the system, and their satisfaction with the different features of the system as shown in Appendix G. In other similar studies qualitative levels of satisfaction have been expressed in a quantitative manner through the use of a Likert scale (Smeele and Waals, 2003). The Likert scale measures level of satisfaction to correspond to a particular value that is assigned to that level of satisfaction. The researchers did not use the Likert scale because:

- In the intended users' language of Setswana, differentiating between slight variations of the same condition, such as to like very much and to like less much, is not normally used in describing feelings towards artefacts such as technological tools especially for concepts that the users are not accustomed to.
- The researchers refrained from the use of the Likert scale as a yardstick for measuring the participants' level of satisfaction and Sherwani et al. (2009a) agrees with inappropriateness of the Likert scale in people from oral cultures. The questions in the questionnaires were reduced to simple enquiries that could be answered decisively with a yes or no or neutral as shown in Appendix G.

In the OpenPhone system the subjective results are used and viewed as more dependable than objective results given the exploratory nature of the experiments. Of course, human experience is generally subjective (Isomursu, 2008), and user subjective reactions to speech interfaces may well be a more important predictor of real world success (Hone and Graham, 2001). Shackel also regards user satisfaction or users' attitude towards the system as the most important aspect of usability (Shackel, 1991), and objective measures of system behaviour may not suffice in predicting system acceptability (Tractinsky, 1997). As the study is centred on the first iteration of the IVR system development, the researchers are more interested in evaluating the technology's feasibility to add value in the lives of the intended users through the measurement of the users' acceptance of the technology since the technology is new to the intended users. The success or failure of the technology depends on the human experience and the study mainly relies on the intended users' view on whether the technology is enough to serve its purpose and meet the user needs and expectations or not. Nevertheless, the study utilises both the subjective and objective results in unison and regards both as complementary. In the analyses of the results, a good conclusion is considered as one that has support of both types of results.

6.2.2 BGR measurements

The BGR system used the entire range of measurements that are defined in Table 2.1. It must be remembered that the BGR was developed two years after the OpenPhone and the researchers were more experienced and prudent in the development of the second case study than in the first including their approach in conducting usability tests. The researchers had learnt to have more control on the procedures of the tests and attaining what they want out of the usability tests. In addition, there were far fewer technical malfunctions on the BGR tests than what was experienced during OpenPhone tests – again, partially due to the experiences gained from the first case study.

The objective measurements that were used on BGR are:

- Effectiveness of the system was measured in the same manner as in OpenPhone, i.e., as the success rate that is presented as the ratio of successfully completed tasks in relation to the total number of tasks.
- Efficiency of the system, as defined in Table 2.1, was measured as an objective measure of time taken to successfully accomplish the given tasks. In contrast to the OpenPhone and as explained previously in subsection 3.2.2, efficiency was a high priority feature of the BGR system, because:
 - BGR users were accustomed to the use of IVR systems and knew the fact that the longer that a call lasts, the more the user pays for the call and therefore it is rational to make the call as short as possible.
 - BGR is a sports entertainment application and is envisioned to be accessed in a more cheery mood in comparison to the OpenPhone which would be conceivably accessed when a caregiver is in need of help. The nature of the information that is provided by the BGR is short numerical information about the results of a match that the user wants to find out about and making an error in the retrieval of such results is not as harmful as in the case of OpenPhone's sensitive information. In that view, the BGR users are less restricted in the speed of navigation through the dialogue, and are actually encouraged to navigate swiftly for minimising costs of the call and also for the following reason:

A very high call volume is expected in the BGR system especially when considering that the system is aimed at a very large user population. The pattern of use for the BGR system is anticipated to have peaks immediately after a popular match has been played and it is desirable that each caller spends as little call connection time as possible so that the system resources can be freed for other callers. This is the reason that the design of the BGR required a maximum call connection time of one minute per user interaction and required the BGR design to incorporate this constraint as high priority.

- Due to the experiences observed in the OpenPhone usability tests whereby participants continued to listen to the rest of the dialogue even when they had heard the answers they were looking for and the superfluous tasks that the OpenPhone users executed, the researchers explicitly requested the BGR participants to hang up as soon as they get the answers that they were looking for in the dialogue and refrain from listening to information that is irrelevant

to the required tasks. This request was re-emphasised after each task and the participants were reminded that remaining connected for longer will cost them more money, which is something that they were already aware of. The BGR procedure of subjective measurements was similar to that of the OpenPhone as shown in Appendix H.

6.3 Experimental results

The following subsections will discuss the objective and the subjective results that were obtained from the experiments.

6.3.1 OpenPhone objective results

A summary of results that compare success rates between the first and second tasks of the DTMF modality in OpenPhone are presented in Table 6.1:

	DTMF Task 1	DTMF Task 2	Total
Correct	4	11	15
Incorrect	17	10	27
Average success rate	19%	52.4%	35.7%

Table 6.1: Results for users who did both DTMF tasks

The results in Table 6.1 are for 21 users who did both DTMF tasks 1 and 2 in the proper manner, i.e., within one session, consecutively without repeating any tasks, and in the proper order of doing task 1 first and then task 2. The 21 users did not necessarily perform their WOZ tasks in the above mentioned proper manner.

By using Chi-square test for significant differences in the proportion of correct answers between task 1 and task 2 in DTMF of OpenPhone:

Chi-squared value = 5.0815, p-value = 0.0242, which means that there is a statistically significant difference in the user performance for the two DTMF tasks since $p < 0.05$.

A summary of results that compare success rates between the first and the second tasks of the WOZ modality in OpenPhone are presented in the following table:

	WOZ Task 1	WOZ Task 2	Total
Correct	8	9	17
Incorrect	12	11	23
Average success rate	40%	45%	42.5%

Table 6.2: Results for users who did both WOZ tasks

The above results, in Table 6.2 are for 20 users who did WOZ tasks 1 and 2 in the proper manner. The 20 WOZ users did not necessarily perform their DTMF tasks in the above mentioned proper manner.

By using Chi-square test for the two WOZ tasks:

Chi-squared value = 0.1023, p-value = 0.7491, which means that there is no statistically significant difference in the user performance for the two WOZ tasks since $p > 0.05$.

In order to verify the above results, McNemar's test for significant differences was used instead of the chi-square test and the same results were obtained indicating a statistically significant difference between task 1 and task 2 in DTMF modality and no difference in the WOZ modality.

It is important to recall that tasks 1 (the 1st tasks) within different sessions and different modalities were different tasks but they were equivalent in their level of difficulty and the same applies for tasks 2. Also, tasks 2 (the 2nd tasks) were more difficult than tasks 1 in both sessions because they needed the participants to navigate information that was deeper in the dialogue and mentioned later in the system prompts. The difference in the level of difficulty between the 1st and 2nd tasks could have been made to be the requirement for the participants to make certain inference about the given information. The researchers did not want to test the participants' abilities or inabilities to make inferences from the given information because the researchers were only interested in the participants' ability to navigate and use an IVR system to obtain information on this first iteration. Figure 6.1 shows the learnability curves for OpenPhone which illustrates the change in success rates as the users move from the 1st task to the 2nd task in the OpenPhone. The gradient of the change between the DTMF tasks is much greater than that of the WOZ tasks, which concurs with the statistically significant difference that was established between the two modalities concerning the change between the first and the second tasks. The implications of the different gradients are discussed under learnability issues in subsection 6.5.2.

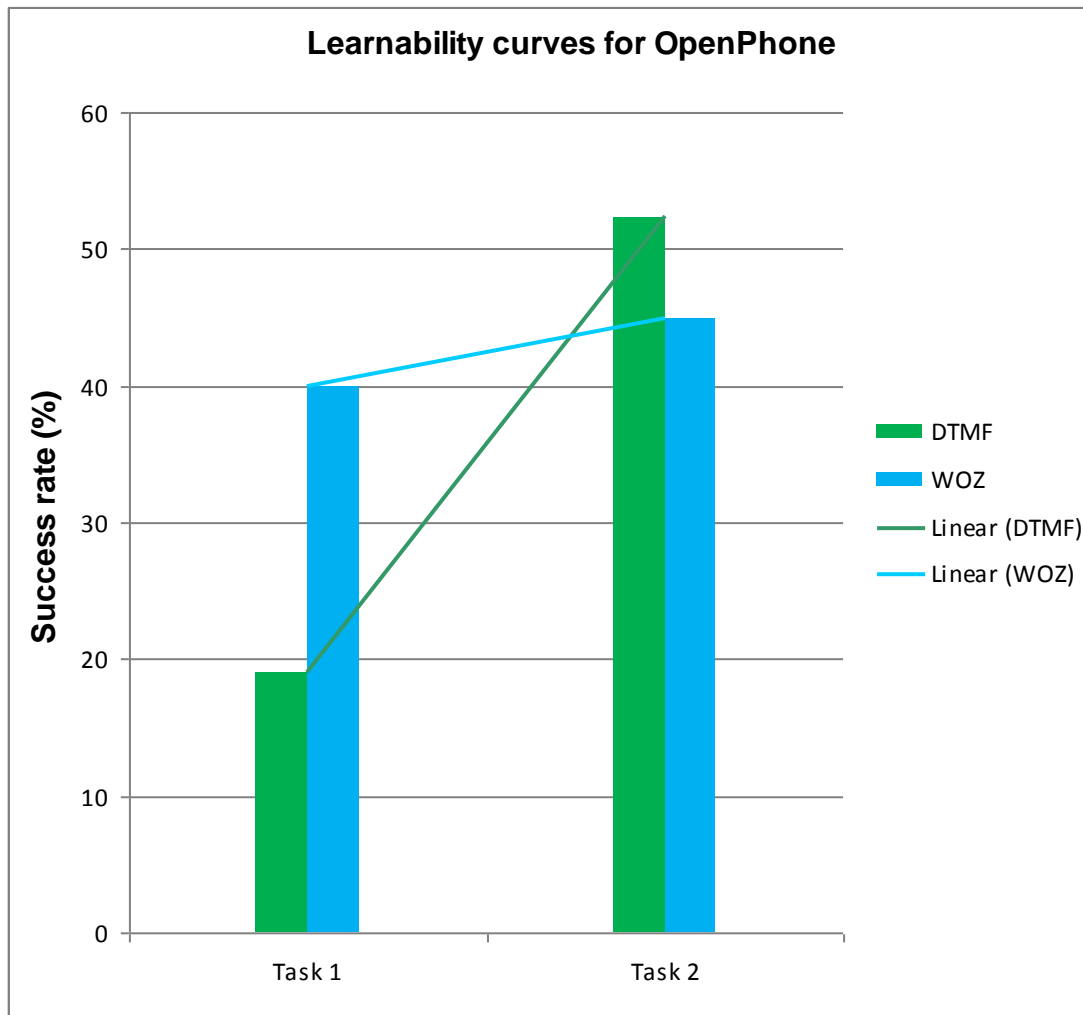


Figure 6.1 Learnability curves in OpenPhone

A summary that compares the overall success rates of the two modalities for the users who participated on both modalities is presented below in Table 6.3:

	DTMF Tasks	WOZ Tasks
Correct	15	17
Incorrect	27	24
Total tasks	42	41
Average success rate	35.7%	41.5%

Table 6.3: Results for the users who participated in both modalities

A total of 22 participants did at least one task from each modality which means that they had the experience of both modalities even though the 22 users did not necessarily perform their tasks in the proper manner. In both case studies and in all sessions there was never an instance where task 2 was done before task 1. For comparison between the two modalities it is important to use only the participants who had a functional experience of both modalities. The fact that half of the 22 users who participated in both sessions started at the DTMF station and the other half started at the WOZ station has ensured that the results are not biased towards

the modality that the users started with. It must also be remembered that in those 22 participants who participated on both modalities, some of them did not necessarily do 2 tasks per modality but since they have experienced at least 1 of each modality they have been included in the comparison. Chi-Square test for significant differences between the success rates in DTMF and WOZ are shown in Table 6.3:

Chi-squared value = 0.3965, p-value = 0.5289, which means that there is no statistically significant difference in the user performance for the two modalities since $p > 0.05$.

6.3.2 BGR objective results

BGR users exhibited high success rates in both modalities when compared to OpenPhone users. Results of the DTMF tasks that compare success rates of the DTMF modality in the three levels of education in BGR are presented below:

		DTMF Task 1	DTMF Task 2	Total
Lower Band Education	Correct	8	9	17
	Incorrect	1	0	1
	Average success rate	88.9%	100%	94.4%
Middle Band Education	Correct	11	11	22
	Incorrect	0	0	0
	Average success rate	100%	100%	100%
Higher Band Education	Correct	7	7	14
	Incorrect	0	0	0
	Average success rate	100%	100%	100%

Table 6.4: Results for users who did both DTMF tasks

The results in Table 6.4 are for 27 users who did both tasks 1 and 2 in DTMF modality in the proper manner.

Results of the ASR tasks that compare the success rates of the ASR modality in the three levels of education in BGR are presented below:

		ASR Task 1	ASR Task 2	Total
Lower Band Education	Correct	6	9	15
	Incorrect	3	0	3
	Average success rate	66.7%	100%	83.3%
Middle Band Education	Correct	10	11	21
	Incorrect	1	0	1
	Average success rate	90.9%	100%	95.5%
Higher Band Education	Correct	7	7	14
	Incorrect	0	0	0
	Average success rate	100%	100%	100%

Table 6.5: Results for users who did both ASR tasks

The results in Table 6.5 are for 27 users who did both tasks 1 and 2 in ASR modality in the proper manner. Similarly to OpenPhone, tasks 1 within different modalities were different tasks but they were equivalent in their level of difficulty and the same applies for tasks 2. Tasks 2 were more difficult than tasks 1 in both sessions because they needed the participants to navigate information that was deeper in the dialogue and mentioned later in the system prompts. All the tasks in the BGR tests were performed in the proper manner.

There was no statistically significant difference found in using chi-square test for comparing differences in the proportion of correct answers between task 1 and task 2 in both DTMF and ASR modalities of the three educational levels. The same results were obtained when using Fisher's exact test. There was no statistically significant difference between the overall success rates of the two modalities in the BGR system. The overall success rate in all bands of education levels for DTMF is 96.3% for task 1 and 100% for task 2. For ASR modality the overall success rate in all bands of education levels is 85.9% for task 1 and 92.9% for task 2. Figure 6.2 shows the learnability curves for BGR which illustrates the change in success rates as the users move from the 1st task to the 2nd task in the BGR system. The gradient of the change between the DTMF tasks is less than that of the ASR tasks, although the difference is not statistically significant. The implications of the different gradients are discussed under learnability issues in subsection 6.5.2.

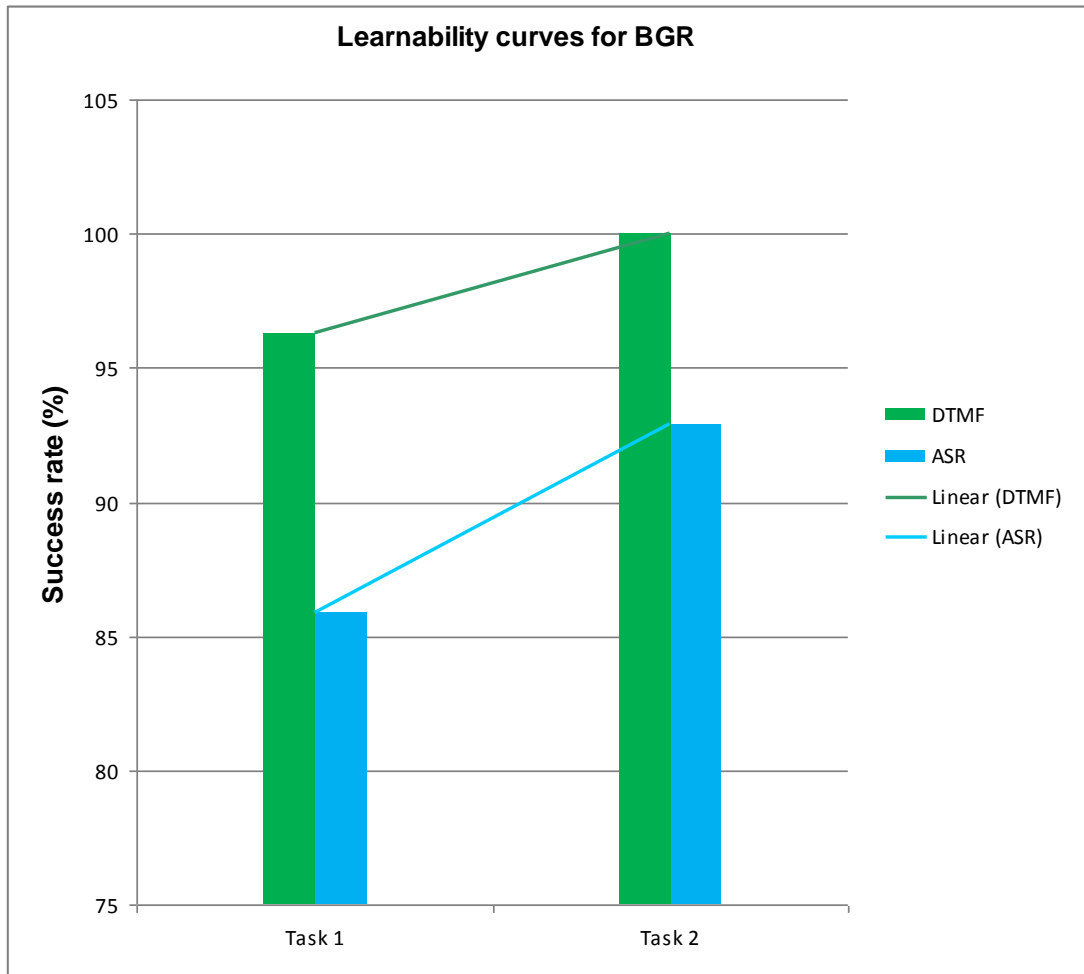


Figure 6.2: Learnability curves in BGR

As mentioned earlier, the BGR system featured efficiency as an important design attribute and the researchers needed to measure the time it took to accomplish each task. The average time that it took the users, in seconds, to correctly finish the tasks in the different categories of education levels is summarised in the Table 6.6 below:

Task	Lower Education Band	Middle Education Band	Higher Education Band
DTMF task 1	38.4	37.5	24.9
DTMF task 2	33.8	32.5	27.9
DTMF Average	36.1	35	26.4
ASR task 1	34.7	34.8	27
ASR task 2	40.7	40.2	30.9
ASR Average	37.7	37.5	29

Table 6.6: Time to complete tasks (seconds)

Table 6.7 below shows the results of t-test for differences in averages in time to complete tasks between levels of education bands and it is evident from the table that HBE users consistently and also statistically significantly took less time to complete their tasks in both

modalities when compared to the lower levels of education, i.e., LBE and MBE. The results of Wilcoxon two sample/Mann-Whitney U test concur with the t-test results presented below:

		T-Statistic	P-value	Significance at 95% level:
DTMF task 1	Between LBE and MBE:	0.2	0.8418	Not significant
	Between LBE and HBE:	-3.55	0.0041	Significant
	Between MBE and HBE:	-3.2	0.0063	Significant
DTMF task 2	Between LBE and MBE:	0.59	0.565	Not significant
	Between LBE and HBE:	-2.95	0.0106	Significant
	Between MBE and HBE:	-2.39	0.0299	Significant
DTMF all times	Between LBE and MBE:	0.41	0.6833	Not significant
	Between LBE and HBE:	-4.42	0.0001	Significant
	Between MBE and HBE:	-3.86	0.0005	Significant
ASR task 1	Between LBE and MBE:	-0.01	0.9913	Not significant
	Between LBE and HBE:	-2.42	0.0302	Significant
	Between MBE and HBE:	-2.53	0.0225	Significant
ASR task 2	Between LBE and MBE:	0.18	0.8569	Not significant
	Between LBE and HBE:	-4	0.0025	Significant
	Between MBE and HBE:	-6.01	< 0.0001	Significant
ASR all times	Between LBE and MBE:	0.09	0.9259	Not significant
	Between LBE and HBE:	-4.09	0.0003	Significant
	Between MBE and HBE:	-4.61	< 0.0001	Significant

Table 6.7: Results of t-test for differences in averages between education bands

6.3.3 OpenPhone subjective results

Immediately after completing each session the participants were asked subjective questions about their experiences with the prototype that they had just used. The subjective questions enquired about factors such as ability to understand the content, pace of the conversation, perceived system response time, and whether their expectations were met or not (as presented in Appendix G). After having completed both sessions they were also asked which system they preferred between the two prototypes and why.

Out of the 22 participants who interacted with both prototypes, 13 preferred DTMF, four preferred the WOZ system, and five were equally happy with both systems which produced a 59.1% preference of DTMF over 18.2% for WOZ, with 22.7% of users who were undecided. Two people out of the four who preferred the speech-enabled system (WOZ) said that they actually like both but they were choosing the speech-enabled system because they think that it would be a better system for elderly users who might have difficulty using the DTMF system. Another participant mentioned that she preferred the WOZ system because she envisions doing something else with her hands whilst interacting with the system. The participants who

did not have any particular preference over any of the systems claimed that the two interfaces were the same to them.

The substantial majority of 13 participants who chose the DTMF system had well-defined reasons including perceived faster speed. The most common reason was the ease of use and the fact that the DTMF was easy to follow because, “it is impossible to get lost as the system [DTMF] just tells you what to do.” Another participant remarked that in the WOZ system she had to pay more attention because if a command is missed or misunderstood or misinterpreted, then that can be problematic but the DTMF is more “straightforward.” Throughout the interactions, the participants were noted to speak in a low, slow paced and monotonous voice and their facial expressions did not show any particular emotion.

All the participants in the studies expressed their approval of the technology as something that would certainly help them and stating reasons such as cutting down the need to travel to the clinic and high availability as the telephony system would help them on weekends and after hours when the clinic is closed. This unanimous view of the usefulness of the technology was the same even for those who experienced only one of the prototypes, who were not included in the comparison exercise. The response to the question about the future usage of the system, if it was fully deployed, was zealously positive with all users.

6.3.4 BGR subjective results

The BGR subjective questions were conducted in the same manner as in the OpenPhone system by asking similar questions after each session and then asking the participants to compare the two modalities after having performed tasks in both modalities as presented in Appendix H.

Out of the nine LBE participants who interacted with both prototypes, seven preferred ASR and two preferred DTMF. Of the MBE participants 10 preferred ASR and only one preferred DTMF. In the HBE participants six participants preferred ASR and again one preferred DTMF. In total, out of the entire set of 27 participants who participated in the BGR tests, 23 preferred ASR and four preferred DTMF, which produced an 85.2% preference of ASR over 14.8% for DTMF. The BGR users were all decisive in their preferences and none of them preferred both modalities as in the case of OpenPhone.

The 23 users who preferred ASR had different reasons which were consistent within their level of education. LBE users mostly preferred ASR because it presented a more natural way

of interaction through speaking rather than the extra effort of pressing numbers. These users considered the ASR to be easier to use and they felt that they had more control when giving verbal commands than pressing numbers in the DTMF modality. The two users who preferred the DTMF mentioned that their preference was based on the fact that they are more used to DTMF than ASR as it is the only modality that they have used in IVR systems. It is worth noting that the most common application for using IVR system in the BGR user population is interacting with the participants' cellphone service provider which is always in the DTMF modality in South Africa.

The MBE users preferred the ASR because they mostly perceived it to be faster even though objective results illustrated that, on average, DTMF outperformed ASR in terms of the time taken to complete tasks although task completion times using the two systems were not significantly different. They also expressed that it was more pleasant to use only one channel of communication [listening and talking] rather than two forms as in case of DTMF where you have to press numbers and listen. MBE users preferred the ASR by a larger margin than any of the groups and were also the largest group in terms of the number of participants. An unexpected reaction from almost all the MBE users and also the majority of the LBE users was the enquiry about the person behind the voice as the users wanted to know the person speaking on the BGR system. The speaker was not a professional voice actor and but one of the researchers who offered to do the voice prompt recording and the researchers did not expect the users' inquisitiveness about the identity of the speaker.

HBE users included users with high levels of experience in using IVR systems and also generally technically literate people. They were consistently producing the fastest time for completing the task with a perfect success rate in all tasks. The first and foremost reason for ASR preference was novelty. They mentioned that ASR is trendy and it is an advanced technology that is beyond their familiarity with DTMF systems that they normally use. HBE users mentioned that they have tried ASR systems before, on a few occasions, but have been discouraged by inaccurate recognitions. The one HBE user who preferred DTMF had a very technical reason of not wanting to put trust on the application itself, as his input is more guaranteed to be correctly recognised when using DTMF. This preference was evidently conforming to the impediments that are generally encountered with the use of ASR systems as expressed by other users within this group. HBE users also made requirements for more efficiency when asked about their further comments on their experiences with the system. The HBE users mentioned that they would have loved the system to mention the scores of their favourite teams first before mentioning other scores so that they can have the option to opt out immediately after getting those scores. This comment was particularly in the light that the

second tasks required the participants to listen to the system prompts for a longer time as the answers to the task questions were mentioned later in the prompts.

Overall, it was evident how the BGR users were excited in using the BGR system by the way that they conducted the tests. An obvious observation was that users were speaking out confidently, loud and swiftly in one positively articulated utterance during the interaction with the BGR ASR prototype. This loudness tendency occurred for all levels of education groups although it was noted that the HBE users were not as openly loud as the LBE and MBE users but, nevertheless, the HBE users were still confident and fast in their utterances. The caller loudness and confident speaking contributed to the very high recognition levels that were experienced in the ASR prototype, “as confident speakers enunciate more clearly, making the recognition task much easier” (E. Barnard, January 10, 2011, e-mail message to author). This was particularly helpful for users who were calling from cellphones in environments that were not ideal for speech recognition purposes. The researchers did not request the callers to raise their voice at any time but the callers selected this style of interaction at their own discretion. Additionally, it was also noted that the LBE and MBE users, in their exuberant state of interaction, tended to use combative language when giving answers to the researchers. For example, when asked the task question for the results of the match that was played on Sunday between Orlando Pirates and Supersport United, an *isiXhosa* language speaker would answer [*isiXhosa in italics*]:

‘*iSupersport United ibhaxulwe two-one yiOrlando Pirates*’, which in English translates to:

‘Supersport United was slaughtered two-one by Orlando Pirates’.

This manner of speaking is standard among soccer fans and it is common amongst soccer sports commentators on television even when speaking in English.

Unsolicited users continued to call the two prototypes, especially the ASR system, even after the tests were done such that the researchers had to be careful in separating valid usability test calls from arbitrary calls in the log files. Even after the usability test period, calls were received by the prototype systems from unknown numbers who called at any time of the day as the prototype systems were up and running 24 hours a day. This showed enjoyment of the prototypes and, conceivably, users’ eagerness for the system to have real results of matches that have been actually played in the PSL.

6.3.5 Vuvuzela earcon experimental results

A brief vuvuzela tone was randomly played at the beginning of the introductory prompt of the BGR systems and participants were asked questions about the vuvuzela experience at the end of each experimental session, in the BGR usability tests as shown in Appendix H. Out of the 27 participants in the BGR, 25 participants positively loved it and the other two participants jokingly mentioned that they would have liked it more if it would have been played softer as they had experienced enough vuvuzela during the world cup tournament and then said that it was alright. The vuvuzela was loved irrespective of level of education or gender, except for the two HBE users who desired it to be a bit softer but still mentioned that they love the vuvuzela. The main reason for the love of the vuvuzela was the cultural grounding that it provided to the participants as they proclaimed that it is a symbol of South African soccer and it allows the user to know that whatever follows is going to be associated with South African soccer, for sure. One user, who is a former junior league player, mentioned that, “the vuvuzela is like a weapon to South African soccer in the same way that a rifle is to a soldier, it is good for both spectators and the players as it provides bravery and confidence to all”. This comment further substantiated the use of combative language in oral expressions. For the in-person participants it was easy to see the widening smile that suddenly came to the users’ face when the vuvuzela was played especially when it did not play during the first task of the first session and then played on the subsequent tasks whereby participants would certainly show more exuberant behaviour after hearing the vuvuzela. Remote participants wanted to know why the vuvuzela cannot play all the time.

6.3.6 Numbers pronunciation experimental results

The researchers conducted a survey that was aimed at establishing the way that numerals are pronounced by the indigenous peoples of South Africa because the outcomes of such research have an impact in the way that IVR systems that have numeric information in their content and context are designed for the region, including the BGR system. All the interviewed participants pronounced the numerals that they were either shown in the different interfaces or sent via text messages in English even though the interviews were in the various indigenous languages of the country. The participants would speak in their indigenous language of choice but when it came to the pronunciation of numbers, they would definitely pronounce that in English. This confirmed the results presented in the BGR user requirements findings subsection 3.5.2.

When the participants were asked the language that the numerals were in, 63 out of the 72 participants that were interviewed said that they were in English. It was surprising to find

participants with tertiary education also claiming that the numerals were ‘obviously’ in English. Seven of the remaining nine participants (who replied that the numerals were not in any language and were just symbols that can be pronounced in any language) were all participants with tertiary education in mathematically based fields such as computer science, engineering and economics. The other two also had tertiary education but not in mathematics-based subjects and it was found that they were well travelled overseas and had lived outside South Africa and could speak other world languages. One participant could speak French and the other could speak Portuguese. Numerals in both those world languages are pronounced in those languages and not in English. The researchers found that all the participants learnt to count and did all their mathematics courses from primary education up to tertiary education in English. It is worth noting that in South African formal education, the language of instruction is mainly English, which “tends to displace all other languages as the major or the first language of tuition” (PanSALB, 2000). Consequently, in no Sub-Saharan African country is an indigenous language used as the medium of education at the university level (Adegbiya, 1994). Although South Africa has eleven official languages, it is evident that ‘some are more equal than others’ and, for pragmatic and economic reasons, African languages are little used in the public and private sectors (Pienaar and Slabbert, 2002). The practicality is that, ‘English is generally understood across the country, being the language of business, politics and the media, and the country’s lingua franca, even though it only ranks joint fifth out of eleven as a home language’ (The Languages of South Africa, 2010). Furthermore, Pienaar and Slabbert (2002) suggest the existence of a perception that English is the (neutral) language of the new national identity. The combined effects of the language that the participants first learnt to count in (English), the continued usage of the language in their daily lives and the present status of English in the country have informed the researchers as the reason why the indigenous people of South Africa exclusively address numerals in English. This had an effect in the way that the BGR was designed as the language used in the interaction is English.

6.4 Analyses

In the OpenPhone system there were numerous timeouts that were experienced especially in the WOZ system. When people made mistakes and lost their bearings in the navigation with the WOZ system, they would simply keep quiet until the system timed out or alternatively they would say the utterance, “FINISH”, which was the keyword/command for the system to exit. On the contrary, DTMF participants would persevere and try different options and ended up taking more time especially on the successive DTMF tasks.

In the BGR system, timeouts were experienced in both systems but for different reasons than in OpenPhone. As indicated in the success rates, a very high number of users got their tasks correct but the timeouts still occurred. BGR participants had an apparent tendency to continue listening to the rest of the system dialogue even after they had found the answer to the task questions posed to them and the extended listening time without hanging up resulted in the system eventually timing out as they were no longer interacting with the systems. This phenomenon occurred on both systems but with greater proportion in the DTMF system, particularly during the first tasks, and was observed in the LBE and MBE groups and not in the HBE group. This effectively produced, and partly explains, the higher task completion times in both modalities on the LBE and MBE groups in comparison with the HBE group as demonstrated in Table 6.6. The other contributor to the significant difference in task completion times between the different education level groups is due to differences in the use of the barge-in facility which is explained further below, in this section.

Contrary to OpenPhone, the timeouts in the BGR were not associated with getting lost in the dialogue navigation or not knowing what to do as demonstrated by the high task success rates in both modalities. Through OpenPhone experiences of continued listening and repetition of tasks, the researchers decided to disallow the prototype systems to continue interacting with the users after the end nodes that give users the final information within a menu branch in a tree (hierarchically) structured dialogue system. The BGR prototype systems were instrumented to timeout immediately after the edge nodes of the dialogue by automatically hanging up the call if the user is not interacting with the system and just listening to the system conversation. This decision was also taken as part of the resolution for efficiency in the system so that an inactive caller can be promptly removed from the system in order to allow other callers to interact with the system, as described in subsection 3.2.2.

During the tests, the BGR systems would timeout and exit (automatically hang up) but the participants would consistently give definite correct answers which showed that the participants were not lost and they actually knew what was requested of them in the task questions. The prolonged listening to the dialogue continued to happen in LBE and MBE groups even when the participants were explicitly requested to hang up as soon as they find the answer to the questions posed in the tasks. Between the tasks, participants were reminded to hang up when they have found the answer to task questions but even though most participants heeded to this request for the subsequent tasks, some participants just continued to listen until the system automatically terminated the call. When they were asked the reason for continuing to listen in, the users proclaimed, "I know it is costing me more to stay connected for longer but I wanted to hear the rest of the information".

In OpenPhone, the DTMF modality was preferred significantly over the WOZ modality notwithstanding the fact that the two modalities did not differ significantly in the overall success rates achieved. The researchers had imagined that, due to the intended users' oral tradition, the speech-enabled WOZ system would be preferred by a large margin since the participants use their local language to interact with it. The researchers had believed that the single modality of interaction of the speech recognition substitute system (WOZ) would be preferable because information is exchanged to and from the system using only speech. With the DTMF system users employ two modalities of interaction with the system, i.e., listening to speech and pressing buttons. It was interesting to observe that even people who experienced more success with the WOZ still chose the DTMF as their preferred modality of interaction. In BGR the overall success rate across all education level groups was not significantly different but the users preferred the ASR with an even greater extent than the users who preferred DTMF in the OpenPhone. In BGR, the researchers had learnt not to have any preconceived ideas about the users' preferences as they were surprised enough from the OpenPhone experiences. It must be remembered that the underlying common factor in the users and usage of the two case studies is that they are for orally grounded people, even though there are differences in their cultures, levels of education, nature of the content in the systems (OpenPhone being a health-related application, and BGR being a fun application).

In OpenPhone the barge-in facility was used by most users, particularly more in the WOZ modality, and the barge-in time roughly corresponded to the time that the correct option at the first menu level was played (Sharma et al., 2009). The barge-in activity for BGR users within the LBE and MBE groups was the same as in OpenPhone but HBE users exhibited a different behaviour. It must be remembered that the researchers briefly introduced how the BGR system works to all participants. After having learnt the system structure from the brief introduction and from practically experiencing how the information was retrieved from the first task in the first session, HBE users did not wait for the entire menu prompt that explained what needs to be done to finish, but instead barged-in and accessed the appropriate information even before it was introduced in the menu prompts. It was more obvious to the HBE users that, for example, to get results for a match that was played on Sunday, the participant had to say 'Sunday' without having to wait to be told so by the system. This contributed to the statistically significant lower task completion times in both DTMF and ASR prototypes of the HBE group in the BGR system. The fact that average task completion times for tasks in group 2 (2nd tasks) were still greater than for tasks in group 1 (1st tasks), even in the HBE group, can be explained as follows: tasks in group 2 required the participants to listen to information that was mentioned much later in the end node which consequently lengthened the interaction time and produced a higher task completion time. For this reason,

the researchers expected tasks in group 2 to take longer than tasks in group 1 in both sessions of both case study applications. Contrary to this expectation, it can be noted in Table 6.6 that tasks 1 took longer than tasks in group 2 in the DTMF modality for the LBE and MBE participants. The proposed reason for this is explained in the findings and contributions section 7.3.

6.5 Why DTMF for OpenPhone and ASR for BGR?

The participants preferred the DTMF modality in the OpenPhone and preferred ASR in the BGR even though the comparisons of the average success rates in both modalities suggest that they could have been more inclined towards making the opposite choices. In finding out why this occurred there is a need to look at how the two modalities differ in their modus operandi in view of the participants' socio-cultural ideals and traditions, as discussed next:

6.5.1 Nature of modality versus socio-cultural norms

In operation, the DTMF modality necessitates the user to interact by pressing buttons on the phone. The interaction is muted to a nearby person and therefore, except for the user, no other person would be able to comprehend the content and the context of the interaction between the user and the system. Conversely, in the WOZ/ASR interaction, and consequently the real life ASR technology, the interaction modality necessitates the user to respond and interact by using their voice as in a normal conversation with another human being. This means that a close by person would be able to overhear the content of the interaction, whether unintentionally or otherwise. Other people who have used the same system would certainly know both the content and the context of the interaction as well.

Within the culture of the caregivers, speaking out loudly on matters that are affiliated with sex is an inappropriate practice. It is well known that HIV/AIDS is an illness that is mostly spread through sex, and therefore speaking out loud about HIV/AIDS is not common unless it is through educational media such as lectures, television, radio, newspapers, etc. The stigma about HIV/AIDS is a complex situation that is interlinked with other stigmas including ethnicity, religion, gender and others (Reidpath and Chan, 2005; Chesney and Smith, 1999). Stigma is considered as one of the main impediments in controlling the disease.

It must be remembered that the caregivers are mostly parents of the HIV/AIDS infected children and consequently are HIV/AIDS infected themselves. The issue about care-giving is not just about their children but it is something that is much closer to their own lives as well.

Being an HIV infected woman, within the socio-cultural ideals of the caregivers bears even more stigmatisation, suffering in silence and shame (Duffy, L., 2005). The caregivers' socio-cultural ideals in combination with the lack of seclusion in the modus operandi of the WOZ modality support the postulation that the caregivers' choice of the DTMF modality is based on the privacy of the modality when compared to its counterpart. The researchers view the privacy issue as the most compelling reason behind the users' choice.

An inherent problem with the stigma of HIV/AIDS is that the users would largely not say that the stigma is the reason for their choice of interaction modality. This is partly because from the lectures that they attend and from the media of radio and television the effects and negative consequences of stigma have been addressed over and over, and caregivers know very well that they are not supposed to yield to stigma. Nevertheless, the HIV/AIDS stigma continues to be a problem in their communities and by nature of stigma, as it is something no one wants to talk about or be associated with, it discourages the caregivers in revealing stigma as an influence to their choice of interaction modality. In all the 22 participants who performed both modalities and were asked to compare the 2 modalities, only one participant mentioned that they prefer the DTMF because of its confidentiality, stating that in the WOZ system other people can hear what the user says. Interestingly, this particular user got both her DTMF tasks incorrect and got both her WOZ tasks correct and still chose DTMF as her modality of choice.

In contrast to the OpenPhone system's access to a stigmatised issue, the BGR provides access to an issue that is not only free of stigma but something that has a culture of being publicly debated and spoken about loudly in public places, private homes, work, and generally everywhere. It is a normal practise to find soccer fans gathered at work, in public places or over a phone call and debating issues about recent games and such conversations are typically accompanied with exuberance and loud speaking and laughter of fans teasing the supporters of teams that recently lost a game. Soccer fans love to speak out loudly about their favourite teams and this camaraderie is a way of making friends because soccer fans find it easy to start a chat with a stranger based on soccer. This loud and exuberant expression of love for soccer is part of South African soccer culture and it is compatible with the oral tradition of the indigenous people of the country as it promotes expression of opinions and chatting freely on a subject that is well-known and loved by the soccer fans. The exuberance of speaking loudly on soccer issues was evidenced on the BGR whereby participants were speaking loud and fast as they gave commands to the system, something they were not requested to do, as they were only requested to speak clearly. This was in direct contrast to the reserved, monotonous and slow style of interaction that was experienced with the OpenPhone participants. According to

studies that were conducted in human emotional sensitivity, loud and fast voice was found to correlate to happiness, joy and confidence whilst a low and slow voice was linked to boredom, grief and sadness (Scherer, 1979; Davitz, 1964). The researchers believe that the reason for the overwhelming choice of the ASR in BGR is because it allows the users to interact with the system in the same manner that they normally address soccer issues which is, simply, loud. Everything about the game in the region from the vociferous vuvuzela to the colourful apparel worn by the spectators, inside and outside the stadia, is loud. Likewise, the choice of the DTMF system in OpenPhone was due to fact that DTMF allows the users to exercise confidentiality in an interaction with a stigmatised subject matter within a society that imposes even more stigmatisation towards infected women as they are looked upon more negatively than infected males.

6.5.2 Learnability issues

In both systems there was some learning effect on both modalities whereby the 2nd tasks in both modalities were performed with equal or more success than the 1st tasks, notwithstanding the fact that the 2nd tasks were more difficult than the 1st tasks. Learnability is one of the main pillars of usability engineering which emphasizes that the user should be able to rapidly begin working with the system (Holzinger, 2005; Nielsen, 1993). Learnability is defined as change of usability over time as the user becomes more acquainted with the system. Learnability of a product can be measured by comparing the quality of use for users over time (Bevan and Macleod, 1994), and in the case of this study learnability has been achieved within a short space of time and not over continuous use of the technology over a long period and this is referred to as rapid learnability.

Figures 6.1 and 6.2 provide the change in success rates as the users move from the 1st task to the 2nd task in the OpenPhone and BGR systems respectively. It must be remembered that task 1 was always done before task 2 in all sessions on both case studies. If the gradient of the change is positive then this can be interpreted as an indication that learnability has occurred, and a zero gradient means that the 2nd task was performed with the same success as the 1st task. From this it can be inferred that, for learnability to have occurred, the gradient of the curve between the first and the subsequent task must be greater than zero and also the value of the gradient is directly proportional to the magnitude of learnability that has occurred. In the case of this study, the gradient is directly proportional to rapid learnability that has taken place. It can be noted from the comparison of the gradient of the of learnability curves in Figures 6.1 and 6.2 that the gradients of the learning curves of DTMF in OpenPhone and ASR in BGR are greater than that of WOZ in OpenPhone and DTMF in ASR respectively. This indicates

that the magnitude of learnability was more prominent in DTMF and ASR for OpenPhone and BGR respectively.

6.6 Conclusions

OpenPhone users viewed the technology as something that would improve their ability to give care to their children at any time of the day and from their homesteads instead of having to travel to the clinic in order to get help. The BGR users regarded their system as a helpful means that gives them results of recently played games whilst travelling or at work without having to wait for sports news on the television or the radio. The intended users' response in both case studies is the buy-in that was aspired by the research team in order to be able to continue with further development of the technologies. The researchers consider the present first iteration of the developments to be a success as they have gathered information that will improve the next iteration and obtained a positive affirmation from the most valuable entity in UCD, i.e., the intended users.

Chapter 7

Hindsight, Shortcomings, and Future Prospects

7.1 Introduction

In the previous chapter some experimental observations were made which influence the development of IVR systems that are intended for users from oral cultures within the region of Southern Africa. These observations were based on the experiences and lessons learnt during the development of the OpenPhone and the BGR systems in the milieu of the hypothesis which stated that overlooking cultural aspects and the social environment of the intended users will lead to ineffective UI design and consequently an unsuccessful system. In this concluding chapter, the research will revisit and summarise the journey that has been travelled and recapitulate on the lessons learnt in order to reflect on best practices and future prospects of IVR design within the targeted area. This is done through the summarisation of the research outcomes and then drawing insights and conclusions in the light of the stated hypothesis. We also discuss the shortcomings of the study. As with most journeys, there are alternative routes that can be taken and this chapter reflects on what might have happened if the research had taken other routes in comparison to the route that was chosen.

7.2 Hindsight

The beginning of this research addressed the possibility of deploying the telephone as a tool for accessing information that can be used by people in the region of Southern Africa, including people who are functionally illiterate and cannot use contemporary ICT tools as discussed in Chapter 1. The target population of the proposed technology in this research is the indigenous African people of Southern African who are characterised as oral people. Taking retrospection on the lessons learnt during the course of the two case studies we can conclude that a telephone is a feasible tool for information access when the interface for interacting with the UI is designed appropriately for its intended users. As proposed in the hypothesis in Chapter 1, the appropriateness of the UI design highlights the incorporation of cultural aspects and the social environment of the intended users. The appropriateness of the design has included the following steps in their chronological order:

- I. Identification of the need for a human-centred design has been derived from the fact that humans interact with a VUI in the same way that they interact with other human beings. Humans use the same part of the brain in both human-to-human and human-

to-machine interactions (Nass and Brave, 2005). Through human evolution, people have become so voice-attuned that when humans hear a voice they quickly associate that with people no matter if the source of the voice is man or machine (Holtgraves, 2002). Research has shown that humans respond socially to all voice interfaces whether the voice origin is of man or machine (Nass and Brave, 2005). People also assign attributes such as likeability, intelligence, kindness, and others to a VUI (Nass and Gong, 2000). These facts have provided the researchers with enough motivation for the human-centred approach in the design of the system.

- II. User requirements solicitation was done directly with the intended users in a manner that is not intrusive or intimidating to the intended users. For OpenPhone, the research found out that focus groups were successful for requirements solicitation whereas in the BGR system, the researchers conducted interviews with the intended users. The researchers also conducted a focus group meeting with the main stakeholder of the OpenPhone system, i.e. the BBCCCoE. This meeting enabled the researchers to understand the organisational needs of the stakeholder in order to align the system design with their needs. For BGR, the stakeholders were the research team and the stakeholder meeting allowed the researchers to specify their requirements and also established agreements about the critical features of system design.
- III. The preferences, concerns, and characteristics of the intended users in each case were incorporated into the design of the systems which was carried out in the setting of the intended users' personae. The stakeholder requirements in each case were also incorporated into the design together with the designers' theoretical expertise in VUI design.
- IV. The two prototypes that were produced by the design were then tested with representative users who were recruited from the targeted user groups. The test participants performed representative tasks and also answered survey questions concerning their experiences with the systems that they had just performed tasks on. Subjective and objective results were obtained from those tests and the researchers were able to draw conclusions based on the results.
- V. The positive feedback from the participating users indicated that the user requirements were met. Information about the users' preferences and observations that were noted will be incorporated into the forthcoming iterations.

The above steps that have been followed throughout the design are homologous to the UCD activities that are depicted in Appendix A and the researchers view the above steps as milestones of the exploratory journey. The journey that the research envisioned to be travelled in Chapter 1 has been completed for the first iterations and the hypothesis that was

presented at the beginning of the journey is accepted owing to the findings that are reported in the next section. The lessons and insights that the researchers contemplated on gaining along the road have been harvested. The journey travelled is also homologous to what Nielsen refers to as “Usability Engineering Lifecycle” (Nielsen, 1992), as the journey has been a voyage around the usability engineering lifecycle and the second cycle (second iteration) is not part of this thesis as initially stated. Nevertheless, the insights and experiences gained are useful not only for the second iterations but also for guiding other similar developments within the given context of use. The production of such guiding principles is what was initially envisioned as establishment of best practices in the first chapter.

There is a noticeable trend that is reflected and replicated throughout the above mentioned milestones which is user involvement and high regard for the users’ context of use and the researchers can thus claim, without reservations, that a human-centred approach in design has been followed in this research. User involvement and regard for users’ context of use have led to the acceptance of the stated hypothesis, as it is clear that without the cognisance of users’ cultural influences, the designs would have produced systems that are different and less effective from the present ones. This is especially considering that the researchers initially had their own predetermined assumptions on both case studies of what should do well for the intended user groups, but were proved wrong. Additionally, future iterations would not have been as well informed by the experiences that have been gathered during the present first iterations, if the cultural influences had been ignored.

7.3 Findings and contributions

1. In the comparison between the two modes of DTMF and ASR in IVR systems, it cannot be conclusively established that oral users prefer one over the other and this research proposes that the type of content in the interaction have repercussions on the users’ preferences. Confidential matters that are associated with stigmatised issues within certain socio-cultural norms in oral communities are better handled with the mute DTMF modality (Ndwe et al., 2010), whereas outspoken issues that are freely discussed with exuberant attitudes are preferably handled with the ASR modality. This finding contextualizes studies that have been previously conducted which suggest that DTMF is preferred by low-literacy users (Patel et al., 2009; Sharma et al., 2009) and also qualifies another study that suggests the preference of ASR (Sherwani et al., 2009b). This research proposes that the type of content in the system has a bearing on the users’ choice of technology.

2. In our users, who are all from oral cultures, the choice between DTMF and ASR is not influenced by the performance levels of the users, which implies that the effectiveness of the system does not influence users' choice. This was evidenced by the fact that in both case studies of this research the users preferred systems where they had performed equally or less successfully. Also, in the BGR study, the users chose the system that was marginally less efficient to them in terms of time taken to complete the required tasks, which implies that their choice is not influenced by efficiency of the system. From this it can be inferred that the objective usability measures of effectiveness and efficiency, as defined by ISO, do not have a conclusive bearing on oral users' choice of technology.
3. In the OpenPhone system the users preferred DTMF over WOZ modality and the success rates showed a higher learnability in DTMF than in WOZ. In BGR the users preferred ASR and there was also higher learnability in ASR than in DTMF as visually shown in the learnability curves in Figures 6.1 and 6.2 for OpenPhone and BGR respectively. Patel et al. (2010) established preference of touchtone (DTMF) over a speech input VUI, and stated that, "participants using the DTMF interface demonstrated a significantly greater performance improvement between the first and third task" (Patel et al., 2009). By deploying data triangulation, which entails gathering data through several sampling strategies, so that slices of data at different times and social situations, as well as on a variety of people, are gathered (Denzin, 1970), the current research suggests that there is a positive correlation between oral users' choice of technology and learnability of the technology. In the three aforementioned case studies, users have preferred the technology that exhibits more learnability.
4. The research has established that oral users' choice is affected by how the technology affords the users' current situational norms at the time of interaction with the technology. Different cultures around the world have different cultural norms concerning loudness in speaking. Poyatos (2002) makes several examples of these various norms, for example, in Kenya and Ghana, shouting in the street and talking loudly indoors is unacceptable and Ghanaians find Nigerians too loud whilst loudness is a cultural characteristic in Spain and Italy. In addition to these cultural norms there are situational norms that fluctuate above or below the standard cultural norm as when people enter a large office where people work quietly at their desks, in an exclusive lounge (particularly with an intimate low light level) or even during an abrupt silence whereby people immediately adjust their voice loudness accordingly, whereas other environmental situations can force people to raise their voices, such as noisy factory or in a noisy party (Poyatos, 2002). In BGR system the users chose the

ASR because it affords the users the opportunity to shift easily from their standard cultural norm to the situational norm of loud soccer culture. Conversely, DTMF was chosen by its users because it enables them to shift easily to the situational norm of being mute about a stigmatised subject.

5. The loud culture of soccer, as explained in subsection 6.5.1, evokes a situational norm that is above the cultural norm and this is attested the exuberant loud voice during participant interaction with the BGR system. On the other hand, the reality of stigma in HIV/AIDS illness within the socio-cultural norms of the OpenPhone's targeted users evokes a situational norm that is below the cultural norm as attested by the low, slow and monotonous voice during their interactions. This research proposes that in oral cultures these variances in situational norms are carried over into the interactions with speech technologies such as IVR systems, as attested by the reactions of the users when confronted with the different types of applications.
6. This research has established that there is a statistically significant difference in task completion times between lower literacy users and higher literacy users. Plauché et al. (2006) performed a speech technology study for accessing agricultural information in India and established that literacy has an effect in the task completion and success rates although there was no significance found between the literate and illiterate users. This study concurs with the view of no significant differences in the success rates of different education levels that was established by Plauché et al. (2006), but did find significant differences in task completion times between higher and lower educational levels. This finding provides an important guiding principle in the design of IVR systems for oral users as it implies that the designers should be cognisant of the level of education of their intended users before they conclude on the levels of efficiency in the design and optimise accordingly, i.e., less efficient for lower literacy and more efficient for higher literacy levels. This view was verified by the HBE users' comments during the interviews after the usability tests whereby they expressed their need for the future fully functional system to address the results of their favourite teams first before other team results so that they can have the option of hanging up immediately after getting those results, thereby completing their tasks faster. Their comments came specifically from the perspective that the second tasks were more time consuming as the users had to wait for a longer time before they could get the answer that was requested in the task questions. None of the MBE and LBE users referred to any needs about more efficiency.

The diminishing effects of oral residue in higher levels of education were observed in the following findings:

7. Oral users are more inclined to position a voice by the speaker hence LBE and MBE users requested to know who was speaking in the voice that they heard in the system. Notably, the majority of HBE users were never interested in finding out the person behind the voice but were more interested in completing the tasks as soon as possible. Although Sherwani et al. (2009a) did not investigate the issue comprehensively, they did theorise that oral people give more importance to the source of information than literate users within the context of speech technology development. This concurs with the observation in BGR users and implies that, in the design of IVR systems that are targeted for oral communities it is important to use voice talents who are trusted and associated with the content in the system. For example, in the case of an application such as BGR, it would be essential to use a well known television soccer sports caster as the voice talent so that oral users can have a human positioning of the voice in the system.
8. According to suggestions made by Sherwani et al.'s, (2009a) framework on orality-grounded HCID, which is based on theory by Ong, (1982), oral cultures promote volubility and redundancy in speech. In this study the effect of volubility and redundancy was demonstrated by the observed prolonged interactions whereby users unnecessarily kept listening to extraneous dialogue till the systems timed out and this happened in both case studies for lower literacy users, i.e. the OpenPhone, LBE and MBE users but not the HBE users. This effect resulted in prolonged task completion times in those users with lower education levels even though the MBE users' task completion times showed improvement in comparison with LBE users. In OpenPhone participants the prolonged interaction was so prominent that it became impossible to consider the time taken to complete tasks as part of the usability measures, as previously explained in subsection 6.2.1. HBE users, on the other hand, exhibited efficient task completion by barging in and abstaining from superfluous dialogues and consequently produced significantly lower task completion times in both modalities.
9. Oral cultures tend to speak in an agonistically toned manner as theorised by Ong, (1982) and Sherwani et al. (2009a) and this was evidenced in the argumentative, engaging and combative manner that soccer fans discuss matters pertaining to matches and also in the way that LBE and MBE users gave answers to task questions about results of soccer matches as described in subsections 6.3.4 and 6.5.1. The

prevalence of combative expression in lower education users occurred with only one participant in the HBE users.

The above three observations in findings 7,8,and 9 provide theoretical triangulation for the proposal that there is a negative correlation between oral behaviour in speech technology interactions and the users' level of education. The oral behaviour lessens with higher education. Ong (1982) and Sherwani et al., (2009a) have made theoretical contributions that concur with the scientific observations that have been made in this study. It is evident from the above mentioned three observations that the behaviours shown by the participants were more intense in lower levels of education users and became much more modest in HBE users. As explained in the first chapter, the type of orality referred to in this research is residual orality. McLuhan (1962) suggests that as a culture internalises the technologies of literacy [reading and writing], the oral residue diminishes, and this statement is evident in the different user behaviors between lower and higher levels of education. Although OpenPhone users' behaviour has been utilised to arrive at this triangulated finding, most of the theory positions that have been used come from the BGR observations. The main reason for the reliance on BGR observations is that it is the only system out of the two where the participants have been explicitly categorised according to their level of education in this research. Moreover, the nature of the BGR as an entertainment and cheerful application has allowed the participants to exhibit behaviours that would have been difficult to observe within cheerless applications such as the OpenPhone.

10. Culturally inclined earcons are an effective means of providing grounding about the context of the speech application that the oral users are accessing. All the participants agreed that the vuvuzela made them aware that the context of the application was about South African soccer and it indicated something concerning, "happiness", as some participants proclaimed. This 'happiness' view is assumable because the vuvuzela is played only in pleasurable settings, even beyond the context of soccer as in, for example, musical concerts, traditional marriages, and other pleasurable social gatherings.
11. In consequence to the above mentioned finding, culturally inclined earcons also help in evoking the situational norms and afford the users to engage easily into the new situational norm that is required by the context of the application accessed. This finding was evidenced by the widening smiles and amusement that immediately came upon the users' faces especially when they heard the vuvuzela for the first time. The sound of the vuvuzela encouraged the exuberant loudness as users raised their voice more after encountering the vuvuzela for the first time.

12. The unforeseen language behaviour that was exhibited by the BGR users brought about an important reality about users in Southern Africa that affects the way that speech applications in the region should be designed because language is central to speech technology. Due to the history of colonisation of the country and the mixture of cultures within the country, the present language situation has resulted in a state of admixture language practices with English being the language most likely to mix with (PanSALB, 2000). This research has established that in oral settings where the colonial language of English has remained the dominant *lingua franca* of the various indigenous peoples of the country even after political independence, such as in the majority of Southern African countries, English is a viable interaction language for IVR systems that primarily provide numerical information. Looking back at the OpenPhone system, the use of numbers in the DTMF prototype was also in English because even the illiterate people in both Botswana and South Africa pronounce numerals in English as both were British colonies. It must be remembered that even though some of the users in Southern Africa are functionally illiterate, they are nevertheless, numerically literate. This finding has also saved the research team considerable resources in the development of the pilot system and it is envisaged that there would be substantial savings in other developments within the oral communities with similar settings.
13. The focus groups, in the OpenPhone user requirements solicitation, were congruent to the users' socio-cultural norms and this enabled the users to participate and fully engage into the discussions during the focus groups. We hypothesise that the success of the focus group meetings was due to the fact that the targeted users are accustomed with communal meetings that resemble focus groups. When there are communal issues that need to be discussed within the targeted users' environment, people are usually summoned by the village chief or other relevant authorities to gather in a common place. The discussion is led by the chief in a manner that is very similar to the way that focus groups were conducted between participants and the moderators as shown in Figure 3.1. The researchers believe that this resemblance has enabled the OpenPhone focus group participants to fully partake and contribute knowledge that was critical in the development of the proposed system, even though the discussion was about a relatively alien idea since the participants are not accustomed to such a system (Ndwe et al., 2008b). In the BGR the requirements solicitation was conducted by initially using a sample of convenience and later progressed into snowball sampling technique and this combination allowed the researchers to attain a wider

user audience that covered the entire country, which was essential as the application is aimed at soccer fans nationwide. This method was successful because the participants that were contacted by the researchers had been referred by their friends and therefore it was easier to initiate conversations after informing them about the person who had supplied the research team with their name and telephone details about a subject that they were comfortable to talk about, i.e., soccer. The researchers propose that it is necessary to find an appropriate method of requirements solicitation as this is the stage whereby the researchers meet the intended users for the first time with the anticipation to form a lasting relationship. Careful attention to the methods used during the first encounters with the intended users pays off well even in later stages of the development as testified later during the usability tests stage of BGR when the researchers were recruiting participants. Some of the participants were the same participants in the initial requirements solicitation phase who had provided their phone numbers. These participants were happy to find out that their initial contributions were meaningful and had culminated in the production of the BGR system. The participants were delighted to participate in the usability tests and were happy to call the BGR prototype systems on their own account without requesting any compensation for the calls made. This was a positive contribution from the intended users especially when considering the high prices for telecommunication calls in South Africa.

14. The learnability curves and the gradient of the change in success rates provide a visual expression of the occurrence and magnitude of learnability when comparing different task performances within a user interface, as demonstrated in Figures 6.1 and 6.2.

7.4 Shortcomings of the study

This section addresses matters that the present iterations could not address due to various reasons including time, budget, and human resources constraints. These are matters that will need to be considered in forthcoming iterations.

In the OpenPhone system, when users were lost in the WOZ navigation or confused on what to do, the users had a tendency to keep quiet and then the system would time out, which resulted in more timeouts when compared to the DTMF. It could have been that the users were silent because they did not want to seem inadequate or foolish in performing their tasks (as they knew that the researchers were listening to them). The researchers tried to be

indifferent towards the participants during the tests so as to influence them as little as possible but it cannot be ignored that they were present. On the other hand, users tended to try more options and persevered more with the DTMF and this could have been caused by the confidentiality of the interaction in the DTMF modality as the users might have assumed that the researchers were not aware of which buttons they were pressing (therefore were convinced that it was more confidential to make mistakes in DTMF than being overheard in the WOZ system).

The OpenPhone study might have produced different results if the users were to be asked to conduct the WOZ experiment in privacy, so that the unpleasant effect of speaking aloud about a stigmatised subject is minimal. It is important to note that the confidentiality addressed in the above paragraph is different from the privacy addressed in this paragraph and the two are not to be confused. The confidentiality mentioned above is a perceived privacy in providing input in the DTMF modality whereby the users assume that the researchers are not aware of what numbers they have pressed and therefore feel more relaxed in attempting different options in DTMF than uttering spoken words in WOZ which are easily overheard by the researchers; this confidentiality has nothing to do with stigma. The other privacy effect is directly related to the stigmatised subject of HIV/AIDS and it is concerned with users being able to interact with the system in isolation in order to observe if this has any effects in their ability to interact with the system about a stigmatised subject. The common and underlying influence in both types of privacy is the effect of the presence of the researchers, which is meaningful in oral cultures especially in lower literate users. Sherwani et al. (2009a) concurs with the proposal of researchers' influence on oral participants by stating that:

The experimenter always has a higher socio-economic status than the participant and is usually foreign to the local community, leading to feelings of intimidation and "performance anxiety" Sherwani et al. (2009a).

The researchers believe that this "performance anxiety" becomes less intense in applications that are for fun and cheerful as the anxiety was not evident in BGR participants, although this could have been due to the BGR users' confidence in using IVR systems as opposed to the OpenPhone users. In the BGR system, this researcher-participant relationship was positively stimulated by the researchers' general knowledge of the beautiful game's current events in both local and international soccer scenes. This greatly improved the interactions with the BGR participants as it was easy for the researchers to break the ice and start a comfortable conversation with the participants. The researchers believe that it would have been, otherwise, difficult to overcome the reservations that people tend to have when conversing with strangers especially in the context of participating in a usability test. This resulted in a researcher-participant relationship with minimal "performance anxiety". There is a need to explore the

effects of researcher-participant relationship in oral user studies and the correlation of such a relationship to the type of content provided by the system. The researchers' ability to speak in the language of the participants also improves the researcher-participant relationship.

User expectations of command words to be used in the speech-enabled WOZ/ASR menu should be studied further. In this study the terms were established from the lectures that the caregivers get from the BBCCCoE and the command words used were not necessarily the users' preference. Further studies would need the assistance of a professional linguist in collaboration with the users themselves in order to investigate the appropriateness of the command words.

Even when a logical term was assigned to a menu item, users could be misled when their expectations of the item did not match the term. Due to these constraints, appropriate and careful naming is critical to reduce ambiguity in a menu selection system (Ziefle, M., 2002).

The cataloguing of system commands must consider command words that are close to the users' mental models which match the task scenarios (Huang et al., 2006). The findings by Ziefle (2002) and Huang et al. (2006) were formulated within the context of visual menus on a cell phone but these findings are very applicable to audio menus as well.

The researchers are aware that if the BGR application was descriptive and content-intensive, providing, for example, explanations of how the goals were scored, and by whom, the choice of interaction language might have been different from the present choice, namely English. This requires further investigation.

7.5 Future prospects

Investigations into the effects in provision of grounding for oral users are necessary as the vuvuzela has unmistakably provided grounding for soccer fans such that they all knew that the application was primarily about South African soccer. Future research could explore whether other culturally oriented earcons in different contexts provide the same grounding, such as investigating if an ambulance siren provides grounding for emergency applications. A siren is well known in Southern Africa and associated with medical emergencies. The context of an emergency would be very different from that of a fun application and the research would need to establish the acceptability of an earcon within that context in oral users.

In the context of the above mentioned emergency situation it would also be interesting to establish the users' reaction to emergencies in the context of speech technology and the situational norm that would be exhibited by the participants when interacting with the

application. Their choice of interaction mode would also shed more light into their preferences and their language choices would also be informative.

Future research, specifically in the area of language attitudes, will investigate if this effect is the same in other post-colonial countries of Africa where the colonial language has remained the dominant *lingua franca* of the various peoples of the country even after political independence. This research should include countries where this *lingua franca* is not English such as Portuguese in Mozambique and Angola, and French in Morocco and Cameroon. As stated in the shortcomings of this study that the users might have chosen other interaction languages if the BGR application was descriptive and content-intensive, providing extra information such as explanations of how the goals were scored, and by whom. This requires further investigation. It is also possible that even if the application is descriptive the users may still prefer English because of the influence by media and other factors. In South Africa soccer is generally a sport that is internationally influenced as fans tend to have both a local favourite team as well as an international favourite. The loud and exuberant discussions by soccer fans are not always about local soccer events but also include international matters about the beautiful game of soccer. Some of the BGR participants claimed to like their international teams even more than their local teams. In this view, it is possible that soccer fans may still like the BGR to be in English even if it was content intensive like the OpenPhone application and that research is required to clarify the correlation between type of content and language preferences in speech technology.

The choices made by the users cannot be taken as conclusive; hence, the research suggests that the choice of technology depends on certain factors that have been discussed in the findings and research contributions. It would be interesting to find out about the use of numbers as commands given by users in the WOZ/ASR modality in applications with sensitive content such as OpenPhone instead of using verbs/nouns that can reveal the context of the interaction to an observer. Even though the content in the application would be in the language of the users, the numbers would conceivably be in the aforementioned *lingua franca* of the particular region as established in findings of this research. The spin-off of using the *lingua franca* would be improved recognition as speech recognition technology in those languages is more advanced than developing world languages with limited speech resources. Of course, the use of numbers in the *lingua franca* would depend on the language attitudes of the targeted region.

It was resolved that the design of both systems would not incorporate login functionality because of literacy levels in OpenPhone and the desire for high efficiency in BGR. HBE users

of the BGR system expressed their desire to have the results of the matches involving their personal favourite teams to be mentioned first before the other results as they would be more likely call to find out about the results of their own teams, primarily. In the view of this demand for future developments, users would be required to provide information about their favourite teams during their initial call to the system and thereafter the system would identify the caller through the caller ID service which transmits the caller's number to the system. This would improve the service quality and efficiency of the system as the users would be directed to the information that they most likely want, i.e., the results of their favourite teams before they hear any other results which they may not be necessarily interested in accessing.

There is a general lack of speech technology research in the developing world but even more scarcity of research in applications that are for entertainment. Except in this research, the researchers are not aware of any documented research in speech technology applications for entertainment that have been aimed at oral users in the developing world. Most applications that are meant for the developing world focus on evident needs such as healthcare and agricultural information access and this is understandable given the challenges in the developing world. Nevertheless, it is important to conduct research from diverse views in order to gather data that is more comprehensive which will, optimistically, lead to a more thorough understanding of user-interface design for oral users.

7.6 Conclusions

The effects of socio-cultural circumstances in the design of technological artefacts, within the paradigm of UCD, are not a new subject in the field of ICT for development. This research study further substantiates the very critical importance of sensitising research according to those socio-cultural contexts when developing products that are aimed at the developing regions of the world.

According to Hofstede's Five Dimensions of Culture (Hofstede, 2001), people living in rural Southern African areas exist in collectivism because in most cases they reside according to their kinship (Easton et al., 2003). This means that people in rural areas live in communal existence whereby the community as a whole is more prominent than individuals, and everybody is more inclined to help and be helped by everybody else. This is in contrast to the way that urban inhabitants live in the region, as they are more inclined to exist in individualism where individuals are more concerned about their own personal and close family's livelihoods and nobody else's (Easton et al., 2003). People living in collectivism are led by collective standards that are determined by their culture and socially accepted norms

instead of individualistic approach of self-determined norms. Collectivism of the caregivers was evidenced during the requirements solicitation focus groups. The researchers had hypothesised that one of the most critical issues to be addressed by the OpenPhone was the issue of psychological support for care-giving as the nature of the disease is emotionally challenging. The caregivers denied that they need such support within the system as they have other means of support with the psychological aspects of the disease (which is obtained from extended family and communal support as discussed in Chapter 3). For this reason, the psychological support was not included in the OpenPhone system dialogue design. Research and technological design for such users should be cognisant of this social existence of collectivism (in contrast to the more likely influence of individualism on researchers who are most likely to come from urban areas with individualistic lifestyles). This further substantiates the need for researcher-participant relationship in oral user studies as addressed in section 7.4.

In BGR, the researchers initially overlooked the present language attitudes of the multicultural South Africa and the growing perception of English as the neutral language of the new national identity (Pienaar and Slabbert, 2002). If it was not for the consistent pronunciation of numerical data in English by the intended users, the researchers might have missed this important finding which affects the design of applications that are numerically oriented in the region. An example of such applications can be found in telephone banking which is a growing industry in the region. Terms that are frequently used in the banking industry such as current accounts, savings accounts, and others, do not have a direct translation in the indigenous languages and English is used for such terms by the indigenous people of all cultures and language backgrounds. The combination of such terms and pronunciation of numbers in English would conceivably result in the interaction language for banking IVR systems being in English. This research has also manifested the challenging issues of deploying speech-enabled technology in a multilingual scenario such as that of South Africa, where an admixture of different languages is the conventional way of speaking. In this view, Barnard et al. (2010b) suggest that:

Users are likely to switch between their local language and the shared language based on a complex set of factors. A useful system may therefore have to cater for two or more languages in parallel, and seamless operation in these languages without explicit choice by the user will be required (Barnard et al., 2010b).

In general, the current research proposes that cultural issues have to be looked through thoroughly and with unassuming attitudes from the researchers as some of the most critical issues cannot be revealed from the objective or subjective evaluations. These critical issues may be implicit issues that are taboo to disclose in such experimental evaluations. The most

critical issues may only be unearthed through reflections and insight into the mechanisms of the local socio-cultural circumstances as evidenced in the privacy issues of OpenPhone and loud exuberance of BGR case studies.

The need for cognisance towards cultural issues is intensified even more with speech applications as humans use the same parts of the brain when communicating with voice systems as they do with other human beings (Nass and Brave, 2005). This phenomenon makes the user interaction with a VUI to be more of a social issue when compared to other types of user interfaces. This effect is even more pronounced within the context of the region of Southern Africa where a strong oral tradition exists amongst a substantial pre-literate or semi-literate population (Barnard et al., 2003a).

Learnability refers to how easy a system is to learn to use. It is well known that people don't like spending a long time learning how to use a system. They want to get started straight away and become competent at carrying out tasks without too much effort (Sharp et al., 2007).

Learnability provides a user buy-in mechanism for the technology that is not only constrained to oral societies but is envisaged to work effectively with the general population as well. This research has established positive correlation between rapid learnability and the preference of technology in oral users (albeit with a limited data set). Even more interestingly, this rapid learnability has been achieved despite the fact that the 2nd tasks were designed to be more difficult than the 1st tasks. It is reasonable to presume that most people would welcome a technology more if they realise that it is something that is within their operational reach.

People are frustrated by long learning times, and new users of software tools demand rapid absorption of tool capabilities (Haramundanis, 2001).

From the experiences derived from the case studies presented above, there are some prerequisites in order to support rapid learnability. The prerequisites entail UI aesthetics that are inspired by the socio-cultural circumstances and norms of the intended user population as postulated in the hypothesis. These socio-cultural norms may prove to have more influence than other factors such as system efficiency and effectiveness.

Utility of speech technology is evident in both case studies of the present study. There is a clear need for care-giving information beyond the lectures that the caregivers presently get as the lectures are not effective as discussed in subsection 2.4.1. OpenPhone users stated that the technology will improve their lives by offering assistance in giving care to their children without the inconvenience and expenses of travelling to the BBCCCoE clinic for lectures that can easily be forgotten. OpenPhone access would be made available in a technology that the

users are familiar with (the telephone) and at all times even beyond the normal operating hours of the clinic. Similarly, the BGR users expressed their appreciation of the technology as it would enable them to get results of recently played games during times when they have no access to the television or any other media platforms that would provide such information as quickly as the BGR system. In both case studies the concept of “motivated user” is fitting as accessing the same information in another form of technology is frequently not as feasible as when accessing it through the speech technology. OpenPhone users cannot be furnished printed material as their levels of literacy are not enough to assimilate care-giving information in textual form. For the BGR users, the technology brings the ability to access information about soccer results through the use of a familiar technology instantly without having to wait for news on television or newspapers, in an environment where Internet penetration is very low.

Despite the lack of entertainment speech technology research in oral users, such applications present important issues about the behaviours of users in the developing world that are unlikely to be unearthed by any other means except by looking at ICT for development through the colourful lens of entertainment rather than the conventional stance that is based on needs. This is in contrast to Barnard et al.’s (2010a) expectation that significant benefits [in IVR systems deployment] can be envisioned if information is provided in domains such as agriculture, health care and government services. Clearly, such domains are of necessity in the developing world but, as stated before, there are oral behaviours that are more perceptible in fun applications as evidenced in the current study.

References

Acomb, K., Bloom, J., Dayanidhi, K., Hunter, P., Krogh, P., Levin, E., Pieraccini, R., (2007). Technical Support Dialog Systems: Issues, Problems, and Solutions. In *Proceedings of HLT 2007 Workshop: "Bridging the Gap: Academic and Industrial Research in Dialog Technology"* (Rochester, NY, April 26, 2007). Association for Computational Linguistics, New Brunswick, NJ, pp. 25-31.

"Address by The President Of South Africa, Mr Thabo Mbeki, At Tsinghua University, Beijing, China, 11 December 2001". [Online] Available:
<http://www.info.gov.za/speeches/2001/0112111046a1003.htm> [Accessed 28 December 2010].

Adegbija, E., (1994). *Language Attitudes in Sub-Saharan Africa: A Sociolinguistic Overview*. Clevedon, UK: Multilingual Matters.

Adherence Education and Psychosocial Support Guidelines: Training Curriculum and Toolkit, (2006). Botswana-Baylor Children's Clinical Centre of Excellence, Gaborone, Botswana.

African Storytelling, (2010). [Online] Available:
<http://web.cocc.edu/cagatucci/classes/hum211/afirstory.htm> [Accessed 22 November 2010].

An analysis of African oral traditions, (2010). [Online] Available:
<http://www.helium.com/items/1008065-an-analysis-of-african-oral-traditions> [Accessed 22 November 2010].

ANSI, (2001). *Common Industry Format for Usability Test Reports (ANSI-NCITS 354-2001)*. American National Standards Institute, Washington, DC, USA.

Armour, P. G., (2000). The five orders of ignorance. In *Commun. ACM* 43, 10 (Oct. 2000), pp. 17-20.

Barnard, E., Cloete, J.L., Coetzee, L., and Patel, H., (2003a). [Towards effective telephone-based delivery of government services](#). *9th International Winelands Conference*, Stellenbosch, 2003.

Barnard, E., Cloete, L., and Patel, H., (2003b). Language and Technology Literacy Barriers to Accessing Government Services. In *Lecture Notes in Computer Science*, R. Traunmüller, Ed., Springer-Verlag, Berlin, Germany, Issue 2739, pp. 37-42.

Barnard, E., Plauché, M. and Davel, M., (2008). The Utility of Spoken Dialogue Systems. In *Proceedings of the 2008 IEEE Workshop on Spoken Language Technology (SLT)*, December 2008, pp. 13-16.

Barnard, E., Davel, M.H., Van Huyssteen, G.B., (2010a). Speech Technology for Information Access: A South African Case Study. *AAAI Symposium on Artificial Intelligence*, Stanford University, USA, March 2010, pp. 8-13.

Barnard, E., Schalkwyk, J., Van Heerden, C., and Moreno, P.J., (2010b). Voice Search for Development. International Speech Communication Association (Interspeech), Makuhari, Japan, 26-30 September 2010, pp 282-285.

Benner, C., (2003). Digital development and disruption in South Africa: Balancing growth and equity in national ICT policies. *Perspective on Global Development and Technology*, vol. 2, No. 4, pp. 1-24.

Bevan, N., and Macleod, M., Usability measurement in context. *Behaviour and Information Technology*, 13, pp.132-145 (1994).

Beyer, H. and Holtzblatt, K., (1998). *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann Publishers, San Francisco.

Brown, I., Cajee, Z., Davies, D., and Stroebel, S., (2003). Cell phone banking: Predictors of adoption in South Africa—An exploratory study, *International Journal of Information Management* 23 (5) (2003), pp. 381–394.

Bryson, L., (1987). Sport and the Maintenance of Masculine Hegemony. *Women's Studies International Forum*, Vol. 10, pp. 349-360.

Castells, M., (1997). *The Power of Identity*. Blackwell Publishers, MA, USA.

Chesney, M.A., Smith, A.W., Critical Delays in HIV Testing and Care: The Potential Role of Stigma, *American Behavioural Scientist*, Vol. 42, No. 7, 1162-1174 (1999).

Curtis, B., Krasner, H., and Iscoe, N., (1988). A field study of the software design process for large systems. In *Communications of the ACM*, 31, 11 (1988), pp. 1268-1287.

Davitz, J. R., (1964). *The communication of emotional meaning*. New York: McGraw-Hill.

DeBrabander, B. and Edstrom, A., (1997). Successful Information Systems Development Projects, In *Management Sci.*, 24, 2 (1977), pp. 191-199.

Delogu, C., Di Carlo, A., Rotundi, P., and Sartori, D. (1998a). A comparison between DTMF and ASR IVR services through objective and subjective evaluation. In *Proceedings of IVTTA '98* (Sydney, Australia, November 30–December 4, 1998), pp. 145–150.

Delogu, C., Conte, S., and Sementina, C. (1998b). Cognitive factors in the evaluation of synthetic speech. In *Speech Commun.* 24(2). May, 1998. pp. 153-168.

Denzin, N. K. (1970). *The Research Act in Sociology*. Aldine, Chicago, USA.

Dickson, G. W. and Simmons, J.K., (1970). The Behavioral Side of MIS. In *Business Horizons*, 13, 4 (August 1970), pp. 1-13.

Dillon, A. (2001). Usability evaluation. In *Encyclopaedia of Human Factors and Ergonomics*, W. Karwowski, Ed. London: Taylor and Francis.

Dreachslin, J.L., (1999). Focus groups as a quality improvement technique: a case example from health administration education. In *Quality Assurance in Education*, 7, 4 (1999), pp. 224 – 233.

Duffy, L., Suffering, shame, and silence: The stigma of HIV/AIDS. *Journal of the Association of Nurses in AIDS care*, Volume 16, Issue 1, January 2005, pp. 13 – 20.

Easton, H., Nakene, M., Naidoo, S., and Kafidi, P.L., (2003). Validating the user-centred design process within a developing context. *CHI-SA 2003 Conference*, Johannesburg, 2003.

Engelbeck, G. and Roberts, T., (1990). The Effects of Several Voice-Menu Characteristics on Menu Selection Performance. In *Technical Report ST0401*, US West Advanced Technologies.

Faulkner, L., (2003), Beyond the Five-user Assumption: Benefits of Increased Sample Sizes in Usability Testing. *Behaviour Research Methods, Instruments and Computers*, 35(3), pp. 379-383.

FIFA: Fédération Internationale de Football Association - Vuvuzela: a symbol of South Africa, (2009). [Online] Available: <http://www.fifa.com/confederationscup/news/newsid=1073689.html> [accessed 28 December 2010].

Folmer, E., and Bosch, J., (2004). Architecting for usability: a survey. *The Journal of Systems and Software*, 70 (1-2) 2004, pp 61–78.

Folmer, E., Van Gorp, J., and Bosch, J., (2003). A Framework for Capturing the Relationship between Usability and Software Architecture. In *Software Process: Improvement and Practice*, 8(2) 2003, pp 67–87.

Foster, J.C., McInnes, F.R., Jack, M.A., Love, S., Dutton, R.T., Nairn, I.A., and White, L.S., (1998). An experimental evaluation of preferences for data entry method in automated telephone services. In *Behaviour & Information Technology*, Volume 17, Issue 2, pp. 82-92.

Freeman, R. E., Reed, D.L., (1983). Stockholders and Stakeholders: A new perspective on Corporate Governance. In *California Management Review*, Vol. 25, Issue 3, p88-106.

Garicano, L., and Palacios-Huerta, I., (2005). Sabotage in Tournaments: Making the Beautiful Game a Bit Less Beautiful. *Center for Economic Policy Research Discussion Paper No. 5231*, September 2005.

Grice, P., (1975). Logic and conversation. In *Syntax and Semantics*, Vol. 3, Speech Acts, P. Cole and J.L. Morgan, Eds., Academic Press, New York, NY, pp. 41-58.

Gulliksen, J., Lantz, A., and Boivie, I., (1999). User Centered Design in Practice – Problems and Possibilities: A summary of 1998 PDC and CSCW workshop. In *SIGCHI Bulletin* 31(2), 1999, pp. 25-35.

Haramundanis, K., (2001). Learnability in information design. In *Proceedings of the 19th Annual international Conference on Computer Documentation*. Sante Fe, New Mexico, USA, October 21 - 24, 2001. SIGDOC '01. ACM, New York, NY, pp. 7-11.

Hess, J. M., (1968). Group Interviewing. In *1968 ACR Fall Conference Proceedings*, Robert L. King, ed. Chicago: American Marketing Association, pp. 193-196.

Hofstede, G. (2001). *Culture's Consequences*. Sage Publications, California, USA.

Holtgraves, T., (2002), *Language as a Social Action: Social Psychology and Language*. Lawrence Erlbaum Associates, Mahwah, NJ, USA.

Holzinger, A. 2005. Usability engineering methods for software developers. *Commun. ACM* 48, 1 (Jan. 2005), pp. 71-74.

Hone, K. S. and Baber, C. (1995). Using a simulation method to predict the transaction time effects of applying alternative levels of constraint to user utterances within speech interactive dialogues. In *Proceedings of the ESCA Workshop on Spoken Dialogue Systems*, P. Dalsgaard, L. Larsen, L. Boves, and I. Thomsen, Eds. ESCA, Vigso, Denmark, pp. 209–212.

Hone, K. S. and Graham, R., (2001). Subjective assessment of speech-system interface usability. In *EUROSPEECH-2001*, pp. 2083-2086.

Huang, S., Chou, I., Bias, R., (2006). Empirical evaluation of a popular cellular phone's menu system: theory meets practice.

International Telecommunications Union, (2006). *World Telecommunication Development Report 2006. Mobile Cellular. Executive Summary*. Geneva, Switzerland.

Isomursu, M., (2008). In *IFIP International Federation for Information Processing, Volume 272, Human-Computer Interaction Symposium*, Peter Forbig, Fabio Paterno, Annelise Mark Pejtersen; (Boston: Springer), pp. 47-52.

Jacko, J.A., Stephanidis, C., and Harris, D., (2003). *Human-Computer Interaction: Theory and Practice*, Lawrence Earlbaum Associates.

Janse van Rensburg, E., (2001). *An orientation to research*, Rhodes University, Environmental Education Unit, research methods short course 2001.

Jokela, T., Iivari, N., Matero, J., and Karukka, M., (2003). The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11. In *Proceedings of the Latin American Conference on Human-Computer Interaction*. Rio de Janeiro, Brazil, August 17 - 20, 2003. CLIHC '03, vol. 46. ACM, New York, NY, pp. 53-60.

Jones, M. K., (1989). *Human-computer interaction: A design guide*. Englewood Cliffs, NJ: Educational Technology Publications.

Karis, D., (1997). Speech recognition systems: Performance, preference, and design. In *Proceedings of 16th International Symposium on Human Factors in Telecommunications*, May 12–16, 1997, Oslo, Norway, pp. 65–72.

Kitzinger J., (1995). Qualitative Research: Introducing focus groups. In *BMJ* 311, pp. 299–302.

Kujala, S., (2002). *User Studies: A Practical Approach to User Involvement for Gathering User Needs and Requirements*, Thesis (PhD), Helsinki University of Technology.

Langmia, K., (2006). The role of ICTs in the economic development of Africa: The case of South Africa. *International Journal of Education and Development using ICT*, vol. 2, No.4. Available: <http://ijedict.dec.uwi.edu/viewarticle.php?id=200>.

Larsen, L.B., (2003) *On the usability of spoken dialogue systems*. PhD thesis. Aalborg University.

Lee, K. M. and Lai, J., (2006). Speech Versus Touch: A Comparative Study of the Use of Speech and DTMF Keypad for Navigation. *International Journal of Human-Computer Interaction*, 19 (3), pp. 343-360.

Lucas, H. C., Jr., (1973). A Descriptive Model of Information Systems in the Context of the Organization, In *Database*, 5, 2 (Winter 1973), pp. 27-36.

Lynch, T. and Gregor, S., (2004). User participation in decision support systems development: influencing system outcomes. In *Eur. J. Inf. Syst.* 13, 4 (Dec. 2004), pp. 286-301.

Makhaye, G., (1998). Shosholozza's Goal: Educate Men in Soccer. *Agenda* vol. 39, pp. 93-96.

MarketingMix: A Marketer's Guide to SA Soccer, (2009). [Online] Available:
<http://www.marketingmix.co.za/pebble.asp?relid=7010> [Accessed 1 October 2010].

Marics, M.A., & Engelbeck, G. (1997). Designing Voice Menu Applications for Telephone Applications. In *Handbook of Human-Computer Interaction* (2nd Edition), Helander, M.G., Landauer, T.K., & Prabhu, P.V. (Eds.) North Holland: Amsterdam, Netherlands.

Mason, E., (2006). Positioning paediatric HIV in the child survival agenda. *Presentation to UNICEF-WHO consultation*. January 11-13, 2006. New York, NY, UNICEF.

Mayhew, D.J., (1999). *The Usability Engineering Lifecycle*. Morgan Kaufman, San Francisco, USA.

McLuhan, M., (1962). *The Gutenberg Galaxy: The Making of Typographic Man*. University of Toronto Press, Toronto, Canada.

Medhi, I. and Toyama, K., (2007). Full-Context Videos for First-Time, Non-Literate PC Users. In *IEEE/ACM International Conference on Information and Communication Technologies and Development*, Bangalore, India.

Menesson, C., and Clement, J.P. (2003). Homosociability and homosexuality: The case of soccer played by women. *International Review for the Sociology of Sport Journal*, vol. 38, No. 3, pp. 311-330.

Migneault, J. P., Farzanfar, R., Wright, J. A., and Friedman, R. H. 2006. How to write health dialog for a talking computer. *J. of Biomedical Informatics* 39, 5 (Oct. 2006), pp. 468-481.

Miller, G. A., (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review*, 63, pp. 81-97.

Morgan, D. L., (1997). *Focus Groups as Qualitative Research*, Sage Publications, Thousand Oaks, CA.

Most Popular Sport in Africa, (2010). [Online] Available:
<http://www.driftingonfoot.com/most-popular-sport-in-africa/> [Accessed: 26 August 2010].

- Mrazek, R., (1993). *Alternative Paradigms in Environmental Education Research: Monographs in Environmental Education and Environmental Studies*, 3. North American Association of Environmental Education, Troy, USA.
- Naidoo, S., and Barnard, E., (2005). Applying the Nass Model to developing world telephone interfaces, *The 4th International Cyberspace Conference on Ergonomics (Cyberg)*, Johannesburg: International Ergonomics Association Press.
- Nass, C. and Brave, S., (2005). *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*, MIT Press, USA.
- Nass, C., and Gong, L., (2000), Speech interfaces from an evolutionary perspective: Social psychological research and design implications, *Communications of the ACM*, 43, pp. 36–43.
- Ndwe, T.J., Dlodlo, M.E., Mashao, D.J., (2007). Usability engineering of an interactive voice response system in a diverse-cultured and multilingual setting. *International Conference on Systems, Computing Sciences and Software Engineering (SCSS 2007)*. 3-12 December 2007, pp. 554-559.
- Ndwe, T.J., Barnard, E., Dlodlo, M., Mashao, D., Kuun, C., Sharma, A., (2008a). OpenPhone user engagement and requirements solicitation in low literacy users. *Human-Computer Interaction Symposium*. Milan, Italy, September 2008, pp. 189-193.
- Ndwe, T.J., Dlodlo, M.E., Kuun, C., Sharma, A., (2008b). Focus groups in rural user studies. *International Conference on Engineering Education, Instructional Technology, Assessment, and E-learning (EIAE 07)*, pp. 180-184.
- Ndwe, T.J., Sharma, A., Kgampe, M., Kuun, C., (2008c). Dialogue design for the OpenPhone system. *Science real and relevant: The 2nd CSIR Biennial Conference*. Pretoria, South Africa, 17-18 November 2008, pp. 1.
- Ndwe, T.J., Dlodlo, M.E., Nichols, J.W., (2010). Comparison of Touch and Speech-enabled IVR System in Low Literacy Users. *IEEE International Conference on User Science and Engineering (i-USER 2010)*. Shah Alam, Malaysia, 13-15 December 2010, pp. 264-269.
- Nielsen, J., (1993). *Usability Engineering*. AP Professional, Boston, MA, USA.

Nielsen, J., (1992). The usability engineering life cycle. *Computer*, vol.25, no.3, pp.12-22, Mar 1992.

Nielsen, J. Chapter 2: Heuristic Evaluation. In *Usability Inspection Methods*, Nielsen, J. and Mack, R.L., Eds. John Wiley and Sons, New York, 1994, pp. 25-62.

Nielsen, J. (1994). *Usability Engineering*. Morgan Kaufmann, San Francisco, CA.

Nuance Communications Inc., (June, 2003). Speech Satisfies a Broad Consumer Market.

[Online]. Available:

http://investor.nuance.com/ireye/ir_site.zhtml?ticker=nuan&script=410&layout=6&item_id=425519 [Accessed: 28 July 2003].

Nunes, I.L., (2006). Ergonomics and Usability – key factors in Knowledge Society. *Foresight Studies on Work in the Knowledge Society*, New University Lisbon/Faculty Sciences Technology, Portugal, Oct 19-20, 2006.

Oates, B.J., (2006). *Researching Information Systems and Computing*. SAGE Publications Limited, London.

Ong, W.J., (1982). *Orality and Literacy: The Technologizing of the Word*. Routledge, London.

PanSALB: Pan South African Language Board, (2000). *Language Use and Language Interaction in South Africa: A National Sociolinguistic Survey*. PANSALB, Pretoria, South Africa, December 2000.

Patel, N., Agarwal, S., Rajput, N., Nanavati, A., Dave, P. and Parikh, T.S., (2009). A Comparative Study of Speech and Dialed Input Voice Interfaces in Rural India. In *Proceedings of CHI 2009*, pp. 51-54.

Patel, N., Chittamuru, D., Jain, A., Dave, P., and Parikh, T.S., (2010). Avaaj Otalo — A Field Study of an Interactive Voice Forum for Small Farmers in Rural India. In *Proceedings of CHI 2010*.

Pekkola, S., Kaarilahti, N., and Pohjola, P. 2006. Towards formalised end-user participation in information systems development process: bridging the gap between participatory design and ISD methodologies. In *Proceedings of the Ninth Conference on Participatory Design: Expanding Boundaries in Design - Volume 1* (Trento, Italy, August 01 - 05, 2006). PDC '06. ACM, New York, NY, pp. 21-30.

Pelak, C.F. (2005), 'Negotiating Gender/Race/Class Constraints in the New South Africa: A Case Study of Women's Soccer', *International Review for the Sociology of Sport*, 40(1), pp. 53-70.

Pienaar, M., and Slabbert, S., (2002). Determining the Appropriate Code in a South African Business Environment. *South African Linguistics and Applied Language Studies*, Volume 20, Issue 4, pp. 263 - 270.

Pinker, S., (1994). *The Language Instinct: How the Mind Creates Language*, Morrow and Company, New York, N.Y., USA.

Plauché, M., Nallasamy, U., Pal, J., Wooters, C., and Ramachandran, D., (2006). Speech Recognition for Illiterate Access to Information and Technology. In *Proceedings of 1st International Conference on ICT and Development*, UC Berkeley. [Online]. Available: <http://www.icsi.berkeley.edu/pubs/speech/speech-illiterate.pdf> [Accessed: February 2007].

Powers, R. F., and Dickson, G.W., (1973). MIS Project Management: Myths, Opinions, and Reality. In *California Management Rev.*, 15, 3 (1973), pp. 147-156.

Prasad, A., Medhi, I., Toyama, K., and Balakrishnan, R., (2008). Exploring the feasibility of video mail for illiterate users. In *Proceedings of the Working Conference on Advanced Visual interfaces* (Napoli, Italy, May 28 - 30, 2008). AVI '08. ACM, New York, NY, pp. 103-110.

Reidpath, D. D. and Chan, K. Y. *AIDS Care*, Volume 17, Number 4, May 2005 , pp. 425-432(8).

Robinson, M., (1993). Design for unanticipated use. In *Proceedings of the Third Conference on European Conference on Computer-Supported Cooperative Work* (Milan, Italy, September 13 - 17, 1993). G. de Michelis, C. Simone, and K. Schmidt, Eds. ECSCW. Kluwer Academic Publishers, Norwell, MA, pp. 187-202.

SAFA: South African Football Association, (2010). About SAFA: Introduction to SAFA, [Online] Available: <http://www.safagoal.net/index.php?page=safaintro> [Accessed: 26 August 2010].

Sampson, P., (1972). Qualitative Research and Motivation Research. In *Consumer Market Research Handbook*, Robert M. Worchester, ed. Maidenhead, England: McGraw-Hill Book Co. (U.K.), Ltd., pp. 7-27.

Schaffer, E. and Sorfaten, J., (1999). Interactive Voice Response (IVR) Case Study: Testing Your Telephone-Based e-Commerce. *The Journal of Electronic Commerce*, 12(2), pp. 78–84.

Scherer, K.R., (1979). Nonlinguistic vocal indicators of emotion and psychopathology. In *Emotions in Personality and Psychopathology*, C.E. Izard, (ed.), Plenum, New York, pp. 495–529.

Schuler, D. and Namioka, A., (1993). Participatory design: Principles and practices. Hillsdale, NJ: Erlbaum.

Schumacher, R. M., Hardzinski, M. L., and Schwartz, A. L., (1995). Increasing the Usability of Interactive Voice Response Systems: Research and Guidelines for Phone-Based Interfaces. In *Human Factors*, Volume 37, Issue 2, pp. 251-264.

Shackel, B., (1991). Usability—context, framework, definition, design and evaluation. In *Human Factors For informatics Usability*, B. Shackel and S. J. Richardson, Eds. Cambridge University Press, New York, NY, pp. 21-37.

Sharma, A., Plauché, M., and Barnard, E., (2009). HIV Health Information Access Using Spoken Dialogue Systems: Touchtone vs. Speech. In *IEEE International Conference on Information and Communications Technologies and Development '09 (ICTD 09)*, Doha, Qatar, pp 95-107, April 2009.

Sharp, H., Rogers, Y., and Preece, J., (2007). Interaction Design: beyond human-computer interaction, 2nd ed. John Wiley and Sons Ltd, West Sussex, England, UK.

Sherwani, J., Ali, N., Mirza, S., Fatma, A., Memom, Y., Karim, M., Tongia, R., and Rosenfeld, R., (2007). HealthLine: Speech-based access to health information by low-literate users. In *IEEE/ACM International Conference on Information and Communication Technologies and Development*, Bangalore, India.

Sherwani, J., Ali, N., Rosé, C., and Rosenfeld, R., (2009a). Orality-Grounded HCID: Understanding the Oral User. *Information Technologies & International Development*, 5(4), pp. 37-49.

Sherwani, J., Palijo, S., Mirza, S., Ahmed, T., Ali, N., and Rosenfeld, R., (2009b). Speech vs. touch-tone: Telephony interfaces for information access by low literate users. In *Proceedings of the IEEE International Conference on ICTD*, pp. 447–457.

Shneiderman, B., (1997). *Designing the User Interface*, 3rd ed. Addison-Wesley, Reading, MA, USA.

Smeele, P., Waals, J., (2003). Evaluation of a Speech-driven Telephone Information Service Using the PARADISE framework: A closer Look at Subjective Measure. In *Proceedings of Eurospeech 2003*, pp. 1949-1952.

South Africa's FIFA World Cup – a success at home and abroad, (2010). [Online] Available: <http://www.fifa.com/worldcup/organisation/media/newsid=1305767/index.html> [Accessed: 24 September 2010].

Statistics South Africa, (2007). Community Survey 2007: Statistical Release P0301, [Online] Available: <http://www.statssa.gov.za/publications/P0301/P0301.pdf> [Accessed: 26 June 2010].

Stentiford, F.W.M. and Popay, P.A., (1999). The Design and Evaluation of Dialogues for Interactive Voice Response Services. In *British Telecommunications Technology Journal*, 17(1).

Swanepoel, D., Hall III, J.W., Koekemoer, D., (2010). Vuvuzela – good for your team, bad for your ears. *South African Medical Journal*, vol. 100, no. 2, pp. 99-100.

The Languages of South Africa, (2010). [Online] Available: <http://www.southafrica.info/about/people/language.htm> [Accessed: 25 September 2010].

The Most Popular Journal: Most Popular Sport in the World, (2006). [Online] Available: <http://most-popular.net/sport-played-world> [Accessed: 26 August 2010].

Top 10 Most Popular Team Sports in the World, (2010). [Online] Available: <http://www.worldsbiggests.com/2010/04/top-10-most-popular-team-sports-in.html> [Accessed: 26 August 2010].

Tractinsky, N., (1997). Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, United States, March 22 - 27, 1997). S. Pemberton, Ed. CHI '97. ACM, New York, NY, pp. 115-122.

UNAIDS, 2008. Report on the Global AIDS Epidemic, July 2008, available at: http://www.unaids.org/en/KnowledgeCentre/HIVData/GlobalReport/2008/2008_Global_report.asp. [Accessed: 10 November 2008].

Valcke: South Africa set a new benchmark, (2010). [Online] Available: <http://www.fifa.com/worldcup/news/newsid=1306612/index.html> [Accessed: 24 September 2010].

Van Dyk, A., (2005). HIV/AIDS Care and Counselling: A Multidisciplinary Approach, Pearson Education, South Africa.

Virzi, R.A., (1992). Refining the Test Phase of Usability Evaluation: How many subjects is enough? In *Human Factors*, 34, pp. 457-468.

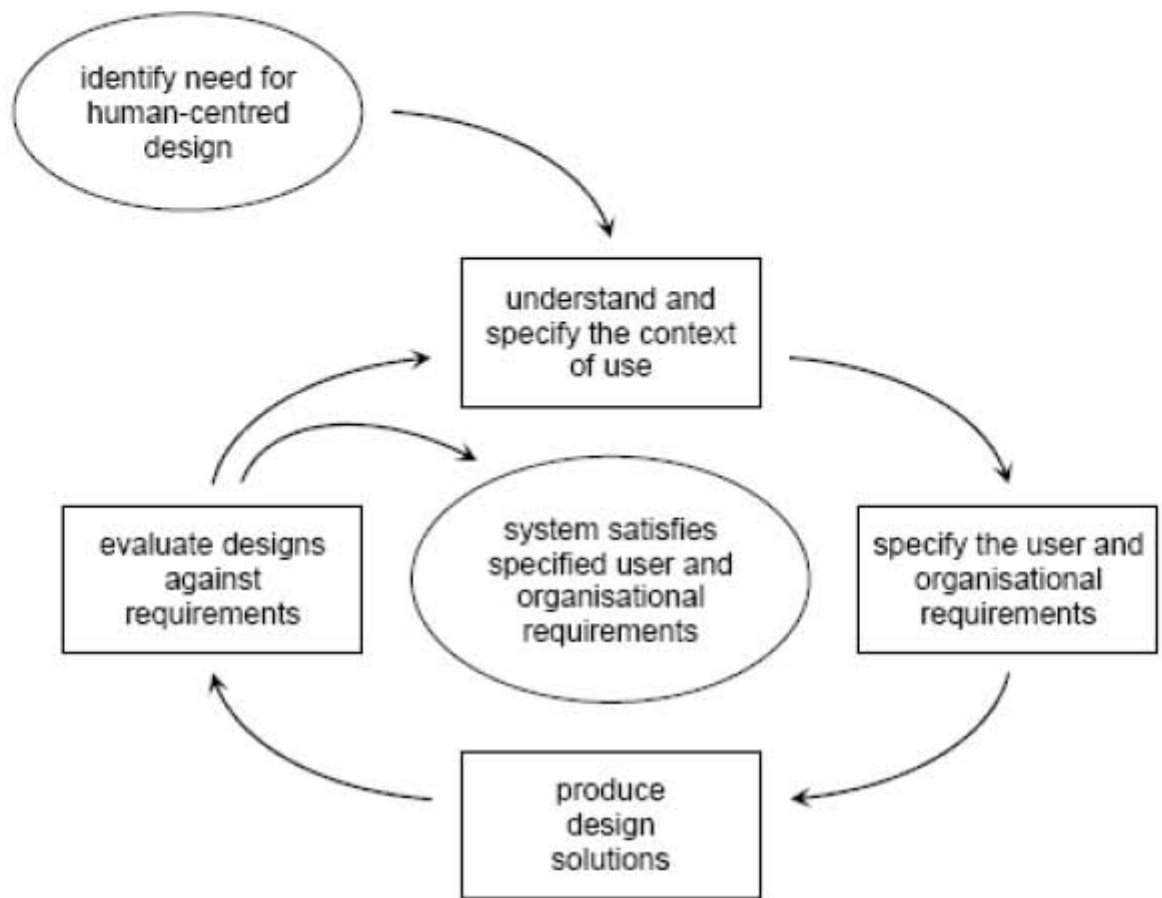
Vredenburg, K., Isensee, S. and Righi, C., (2002). User-Centred Design: An Integrated Approach. Prentice Hall, Upper Saddle River, USA.

Vrooman, J. (2007). Theory of the Beautiful Game: The Unification of European Football. *Scottish Journal of Political Economy*, Volume 54, Issue 3, pp. 314–354.

Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.

Ziefle, M., (2002). The influence of user expertise and phone complexity on performance, ease of use and learnability of different mobile phones. *Behaviour and Information Technology*, 21(5), pp. 303-331.

Appendix A: User-Centred Design Activities (Jokela et al., 2003)



APPENDIX B1: Heuristic Evaluations

HEURISTIC	DEFINITION	OPENPHONE COMPLIANCE
Visibility of system status	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.	The system always alerts the user of their current state by providing a preamble statement to an action that they may or may not choose, for example, the system will prompt the user, “to learn about HIV spread, say SPREAD”. This indicates to the user that after saying the command SPREAD they will be interrelate with information that deals with the spread of HIV. This procedure is maintained throughout the system dialogue.
Match between system and the real	The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.	The system dialogue refrains from using scientific and medical nomenclature but instead words that are familiar to the user population are used to describe phenomena. For example, the immune system is referred to as “soldiers of the body” in the system dialogue.
User control and freedom	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.	Throughout the system dialogue the user is reminded that at anytime that they would like to start over again (as an “emergency exit” when they make mistakes) they can go back to the main menu by saying the command START in the ASR version or pressing 0 in DTMF.
Consistency and standards	Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.	The system dialogue has been steadfastly designed to have no reuse of the same utterances or commands unless they mean exactly the same thing. The system is a finite state machine whereby, for example, pressing the digit 3 in one state produces different results when pressing the same digit in another state. Confusion in pressing the same digit is curbed by the fact that the user is always made aware as to which state that they are currently in as discussed in the visibility of system status.
Error prevention	Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.	Errors that are inherent of the speech recognition technology cannot be avoided but the design of the system is such that it accommodates those technology limitations in the form of, for example, short and precise commands that are intuitive and easy to remember, and are also acoustically dissimilar. Furthermore, good error messages are provided for when errors happen such as when a system timeout is encountered due to user silence, the system says, “I am sorry, I did not understand you. Please clearly say one of the following words...” This procedure is maintained throughout the system dialogue.

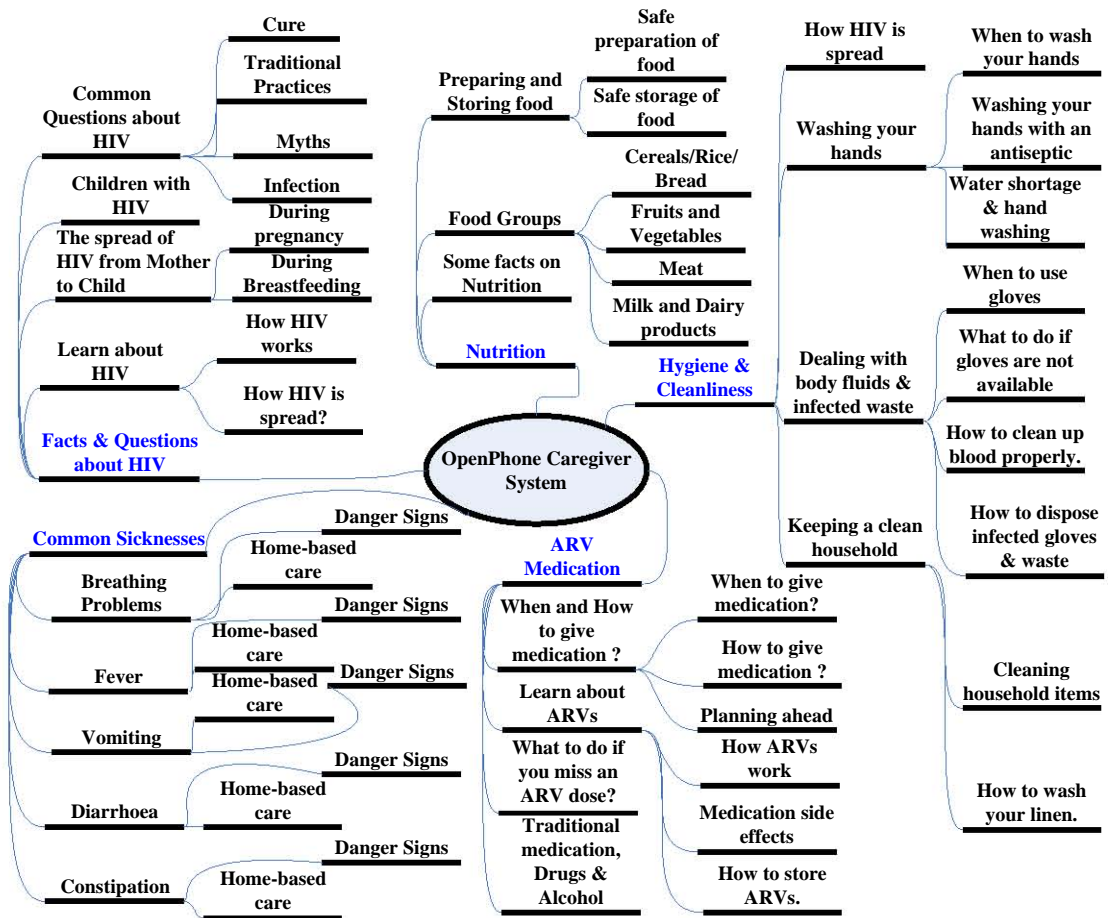
APPENDIX B2: Heuristic Evaluations (continued)

HEURISTIC	DEFINITION	OPENPHONE COMPLIANCE
Recognition rather than recall	Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.	The number of menu items is decisively limited to be no more than 5 at the root menu and the depth of each menu is limited to 3 levels as demonstrated the system dialogue structure in figure 3 in order to minimize the users' cognitive load. The menu-driven dialogue structure ensures that the user is guided towards their goals.
Flexibility and efficiency of use	Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.	The barge-in feature is supported so that advanced users can interrupt the system and execute commands as they feel necessary without having to wait for the system to finish system output if they know the commands they want to execute. At the same time novice users are guided by the system towards their goals.
Aesthetic and minimalist design	Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.	Grice's maxim of quantity ¹ for cooperative conversations is adhered to (Grice, 1975).
Help users recognize, diagnose, and recover from errors	Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.	Error messages are provided in a humble, clear and concise language. A universal command that is applicable at all levels of the system dialogue provides an emergency exit to go back to the main menu and this command is reiterated in several instances in the system dialogue.
Help and documentation	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.	There is no documentation provided since the system is for people who may not be able to read. A feature for users to leave questions is provided so that further help can be provided later.

¹ Grice's maxim of quantity states that:

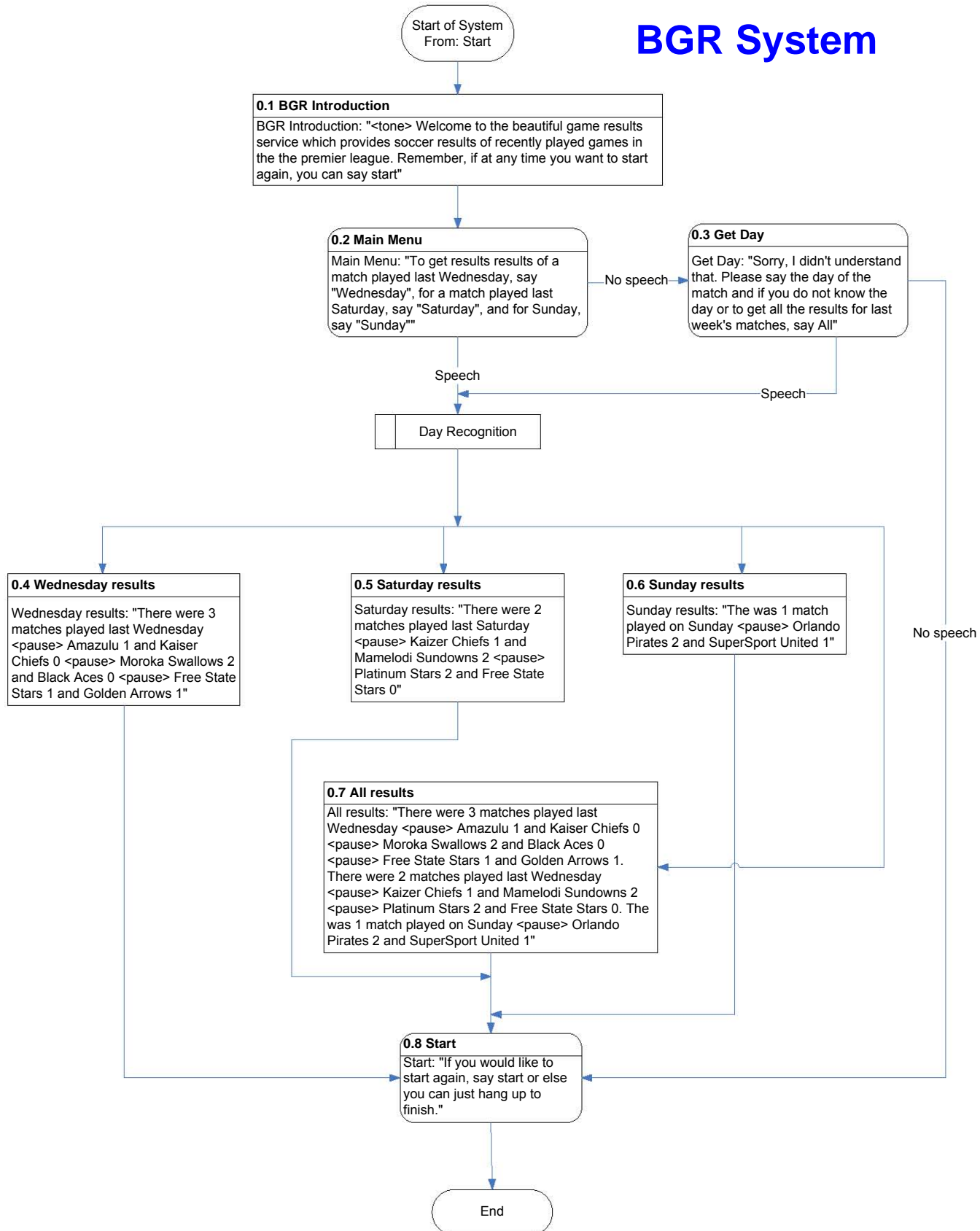
- I. *Make your contribution as informative as is required.*
- II. *Do not make your contribution more informative than is required (Grice, 1975).*

Appendix C1: Dialogue Structure of OpenPhone

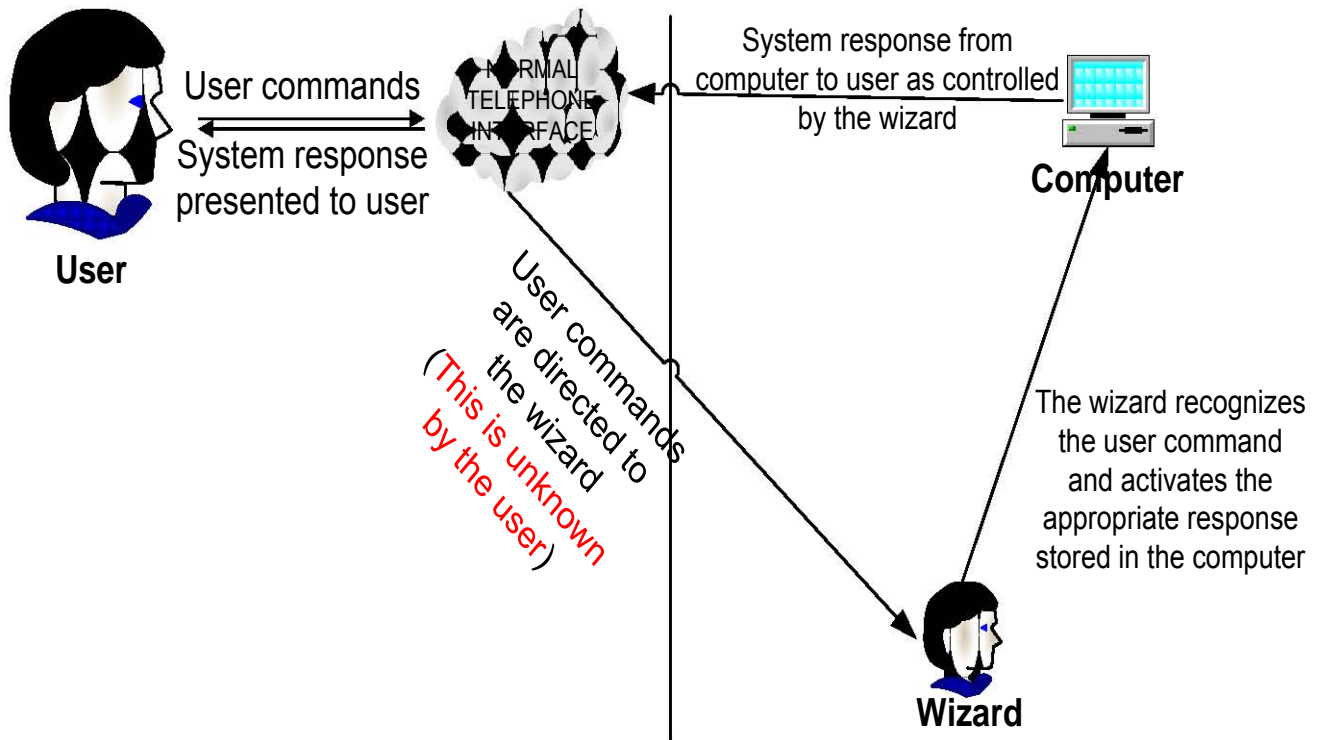


Appendix C2: Call Flow Diagram for BGR

BGR System



Appendix D: Wizard of Oz (WOZ)



Appendix E: Experimental Procedures Pictures



User participant interacting with the system (a)



User participant interacting with the system (b)

Pictures courtesy of Christiaan Kuun

Appendix F1: OpenPhone Tasks

SET A

1. Your friend knows that she must wash her hands frequently to keep her family safe from disease but she has very little water at home so she doesn't know what to do. You've heard that Baylor has a phone number (help line) that can answer many health questions. What information does the help line provide on what to do when there is shortage of water?
2. Your friend comes to your house saying that her child is very sick; the child feels very warm to touch and has a fever. Your friend wants to know what she can do to care for the child at home. You know that Baylor has a phone number (help line) that can answer health questions. What information does the help line provide on what to do at home when a child has fever?

SET B

1. Your friend knows that she must use gloves when touching any blood or body fluids to keep her family safe from disease but she does not have any gloves so she doesn't know what to do. You've heard that Baylor has a phone number (help line) that can answer many health questions. What information does the help line provide on what to do when there are no gloves?
2. Your friend comes to your house saying that her child is very sick; he/she is weak and has diarrhoea. Your friend wants to know what she can do to care for the child at home. You've heard that Baylor has a phone number (help line) that can answer many health questions. What information does the help line provide on what to do when the child has diarrhoea?

SET C

1. Your friend knows that she must wash her hands frequently to keep her family safe from disease but she has very little water at home so she doesn't know what to do. You've heard that Baylor has a phone number (help line) that can answer many health questions. What information does the help line provide on what to do when there is shortage of water?
2. Your friend comes to your house saying that her child is very sick; the child feels very warm to touch and has a fever. Your friend wants to know what she can do to care for the child at home. You know that Baylor has a phone number (help line) that can answer health questions. What information does the help line provide on what to do at home when a child has fever?

SET D

1. Your friend knows that she must use gloves when touching any blood or body fluids to keep her family safe from disease but she does not have any gloves so she doesn't know what to do. You've heard that Baylor has a phone number (help line) that can answer many health questions. What information does the help line provide on what to do when there are no gloves?
2. Your friend comes to your house saying that her child is very sick; he/she is weak and has diarrhoea. Your friend wants to know what she can do to care for the child at home. You've heard that Baylor has a phone number (help line) that can answer many health questions. What information does the help line provide on what to do when the child has diarrhoea?

Appendix F2: BGR Tasks

SET A

1. How many matches were played last Saturday?
2. What were the results of the match that was played last Sunday between Orlando Pirates and Supersport United?

SET B

1. How many matches were played last Wednesday?
2. What were the results of the match that was played last Wednesday between Free State Stars and Golden Arrows?

SET C

1. How many matches were played last Sunday?
2. What were the results of the match that was played last Saturday between Kaiser Chiefs and Mamelodi Sundowns?

SET D

1. How many matches were played last Wednesday?
2. What were the results of the match that was played last Wednesday between Moroka Swallows and Black Aces?

Appendix G: OpenPhone DTMF Post Questionnaire

No	Question	Yes	Neutral	No
1	Was it easy to understand the advice you heard? (Content)			
2	In this conversation, did the system understand what you wanted? (User Input recognition)			
3	In this conversation, was it easy to get the advice you wanted? (Task Ease)			
4	Was the pace of conversation ok? Too fast---Too slow (Interaction Pace)			
5	In this conversation, did you know what you could say or do at any time? (User Expertise)			
6	Did you ever feel that it was taking a long time to get the advice you wanted? (System Response and Length of content)			
7	Did this conversation go as you expected? (Expected Behaviour)			
8	ONLY IF user has done BOTH sessions. Which did you prefer to use between Buttons and Speech for getting advice? (Comparable Interface)			
9	After calling this number, do you think you would use this number to get health advice when you are at home? (Future Use)			
10	Is there anything else you would like to tell me?			

APPENDIX H: BGR ASR Post Questionnaire

No	Question	Yes	Neutral	No
1	Was it easy to understand the task you were required to perform? (Content)			
2	During the interaction, did the system understand what you wanted? (User Input recognition)			
3	Was it easy to get the answer you wanted? (Task Ease)			
4	Was the pace of interaction ok? Too fast---Too slow (Interaction Pace)			
5	Did you get lost or confused on what to do at any time? (User Expertise)			
6	Did you ever feel that it was taking a long time to get the answer you wanted? (System Response and Length of content)			
7	Did the interaction go as you expected? (Expected Behaviour)			
8	Which did you prefer to use between Buttons and Speech for getting advice? (Comparison of interface, only if user has done both)			
9	After calling the system, do you think you would use the system for real results access? In which modality? (Future Use)			
10	Did you hear a vuvuzela at anytime during the interaction, and if yes, how many times, do you like its presence?			
11	Why do you like/not like the vuvuzela?			
12	Is there anything else you would like to tell me about the interaction?			