INVESTIGATING THE SIGNIFICANCE OF CULTURE AND ACCENT ON THE USABILITY OF TEXT TO SPEECH TECHNOLOGY AS A TEACHING AND LEARNING TOOL

YVONNE MARY SPITTLES

DOCTOR OF PHILOSOPHY DEGREE

2009

Dedication

This work is dedicated to

my husband Derek,

my three children Claire, Cheryl and Adam,

and my four grandchildren Damion, Kieran, Daisy and George

with all my love.

Thank you for all of your support.

Also, remembering my Mum, Mary, and the promise I have fulfilled.

Acknowledgements

I am indebted to my supervisor, Professor Lynne Dunckley for all her invaluable advice, support and guidance, without whom this PhD would not be possible. It was with great sadness that Professor Dunckley passed away in December 2008 just before the final submission of my thesis. I hope that I have done her justice.

I would also like to express my gratitude to my second supervisor, Professor Andy Smith for his advice and guidance, particularly at the MPhil transfer and final stages of the PhD.

Thanks also, to Dr Roger Tucker and Pernilla Nasfors with whom I worked during the testing of the banana information system, Rachel Moore of the ACE Advisory Trust Centre and Norman Hore of Rapid English Consultants.

Abstract

A multicultural education system embracing diversity is one that is being encouraged in this, and many other, countries. E-learning systems are being increasingly implemented and text to speech elements may be critical in multicultural and special needs environments. Although independent literature on text to speech technology, culture, special needs education and usability is available, there is little evidence of research that integrates all four aspects. This thesis investigates the significance of differences in culture and accent on the usability of text to speech technology as a teaching and learning tool.

Speech has the potential to be the most intuitive human computer interface but it can be claimed that the usability of such systems has not advanced sufficiently to meet acceptable standards of use. Although many end users claim to be unable to understand text to speech, some developers hold the view that there is no point adding expressiveness or emotion to the output speech as the systems have no understanding of what they are saying. This research investigates the proposition that technology is sufficiently advanced, but that problems exist because developers and end users misunderstand each other.

In compliance with strict ethical guidelines in the design and implementation of tests with human subjects who may be vulnerable, three separate studies were undertaken in pursuance of the research objectives.

The first study was performed in the computer laboratory at Thames Valley University and focussed on the relationship between error rates and accent. Results showed that error rates were consistently lower when subjects heard speech that preserved their native culture.

The second study was conducted at the farms of banana farmers in Kenya, testing a new system that had been developed by the Local Language Speech Technology Institute, with the aim of helping farmers to become self-sufficient by being able to glean previously inaccessible information. These experiments focussed on learnability, satisfaction and error rates. The research showed that there was an unacceptably high error rate, rendering the system unusable. This was attributed to two main factors. The first factor was the large amount of speech that was incorporated into the system that proved difficult to navigate. The second factor was the absence of speech that preserved the native culture of the end users. The developers are now taking an evolutionary approach to address these factors in the interface design.

Using the results of the first two experiments, a prototype model of a talking book was created and experiments were undertaken in the final study, focussing on learnability and satisfaction components. The results showed that usability was increased for subjects when cultural differences had been considered in the interface design.

Overall this research shows that the consideration of the culture of learners and the accent of speech is a very important component in interface design, demonstrating that a person's accent is representative of their cultural upbringing which also affects their preferred learning styles. Results consistently showed that the usability of text to speech technology is improved when the speech that is output preserves the native culture of the end user.

List of Publications

Spittles, Y. & Dunckley, L. (2007). Ethical Considerations for Usability Testing in Kenya using Kiswahili and English. CHI User Centred Design and International Development *[Online]* *htttp://wwwmikebinta.gatech.edu/UCDandIDWorkshop/papers/spittles.pdf.*

Spittles, Y. & Dunckley, L. (2005). Ethical Issues in Participative Research in Usability Testing. *Edean UK: The Practical Application of Ethical Guidelines for Participative Research.* Pp 13 – 19.

Glossary

ADC	Analogue to digital converter
ANN	Artificial neural networks
DAC	Digital to analogue converter
DAM	Diagnostic acceptability measure
DPCM	Differential pulse code modulation
DRT	Diagnostic rhyme test
DSP	Digital signal device
FM	Frequency modulation
Formants	Short term correlations in the human speech signal
HCI	Human Computer Interaction
HMM	Hidden Markov Models
IMLE	Interactive Multimedia Learning Environment
ISO	International Organization for Standardization
MIDI	Musical instrument digital interface
Morphological	Words can have an internal structure
MOS	Mean opinion score
MRT	Modified rhyme test
NLP	Natural Language Processing
OCR	Optical character recognition

PC	Personal computer
PCM	Pulse code modulation
Phonological	Words consist of a sequence of distinctive sounds or phonemes
Phonemes	Basic sound units which make up human speech
Pragmatic	The speaker uttered the sentence for a reason
Semantic	Words and sentences have meaning
SSADM	Structured systems and analysis design methodology
SME	Small to medium sized enterprises
SMS	Short message services
SRS	Speech recognition software
Syntactic	Words are systematically ordered to form sentences
TBI	Telephone Based Interface
TTS	Text to speech

Table of Contents

Chapter One - Introduction

1.1 Research Background and Motivation

The use of speech could be considered to be the most intuitive human computer interface, particularly for end users who have special educational needs (SEN) and are unable to use traditional input and output devices. Indeed for some end users, the use of speech is the only interface that is accessible for them.

This research is inspired by the experiences of one such end user, who will be referred to as Jane, whom the author has taught on a one-to-one basis. Jane is a young adult female of Afro-Carribean culture and is totally blind with multiple sclerosis which renders her with total paralysis. She has no functionality in her limbs whatsoever and has to be lifted in and out of her wheelchair and bed by the means of a mechanical hoist operated by her carer. Her speech, hearing and intellect, however, are totally unaffected and despite facing great adversity, she enjoys studying for qualifications through the Hadley School for the Blind and is extremely proud of the many Braille certificates that adorn her walls. As Jane cannot read or write, the use of speech technology is her primary means of accessing and communicating information. The poor levels of usability she experiences with the software that she uses is an area that causes her much frustration. The reasons behind her frustrations are twofold – to be able to use speech as an input, significant training of the system has to be undertaken initially before the software can be used.

The training sessions are not suitable for end users with SEN as the specifically designed training speech has to be recorded in its entirety in one session, and does not allow for end users tiring. In addition, the system only accepts words in an American accent so therefore she has no choice but to endeavour to pronounce words in this way which is alien to her as Jane only rarely leaves the house and lives alone with her primary carer who is also of Afro-Carribean culture. They communicate to each other in English language with a strong Carribean accent; she has not had many opportunities of hearing other people speak in accents of other cultures. This resulted in the inability to train the system properly and this was, therefore, considered by the author to be a contributing factor to the poor resulting usability of this system.

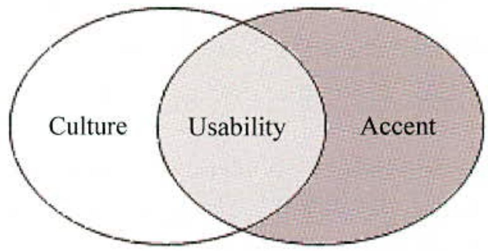
In addition to using speech as an input, Jane also uses speech as an output but unfortunately this adds further to the frustration as, again, the speech is in an American accent that she is unfamiliar with which she cannot understand properly. The author shares in her frustrations and has contemplated whether the usability of the system would be improved if the output speech was in the same accent as the end user.

Speech systems have been developed and are used with varying degrees of success in different languages, including English as an output. Such systems however are not readily available in the English language with accents. This research considers whether it is right to only provide end users of text to speech (TTS) systems with counter-cultural software which means that they will only ever be used to using culturally biased products. Griffith (1998), states that when we localise software for small indigenous cultures, we are preserving elements of that culture and indirectly preserving a wider variety of perspectives. However, there is also a chance that in the future, more culturally appropriate products, which preserve cultural values and preferences, might not be successful because the user group has already been exposed to the less appropriate (usually American) version beforehand (Yeo and Barbour, 1998).

1.2 Overview of Thesis

This research is concerned, primarily, with the usability of TTS technology as a teaching and learning tool for end users of post compulsory school age (16+) from different cultures, for whom English may not be their first language and, in addition, may have SEN, with the focus on inclusivity. It is not the intention of the author to create new TTS systems but, moreover, to consider the usability of existing and developing systems focusing, in particular, on their levels of accuracy and how end users perceive them in relation to their lack of accent, emotion, pace and tone. A series of three experiments explores the significance of the effect of differences in the culture and accents of end users, on the usability of such systems as shown in Figure 1.1.

Figure 1.1. Potential Interaction of Culture and Accent on Usability



1.3 Language

Coxhead (2005), considers that if one of the aims of the field of Human Computer Interaction is for computers to fit in with us, rather than us with them, they will have to be capable of being able to process natural language, as an ever increasing number of end users are now moving away from the traditional input devices. Natural language can be defined as one of the languages which humans naturally speak as opposed to one that has been invented, such as Esperanto. There are up to 7,000 languages spoken in the world but the majority of the population speaks one of eight languages – Mandarin (Chinese), English, Hindi, Spanish (Castillan), Russian, Arabic, Bengali and Portuguese (Coxhead, 2005).

As far as spoken language is concerned, even people who speak the same language do not use exactly the same set of words with exactly the same pronunciation, grammar and meaning; the language that an individual person uses is referred to as an idiolect. When people from the same or similar cultures communicate verbally their idiolects are very similar and this is referred to as a dialect. Examples of very similar languages which are referred to as dialects are Hindi and Urdu which are spoken in northern India and Pakistan which are associated both with political divisions and with the Hindu and Muslim religions respectively (Coxhead, 2005).

1.4 Speech Synthesis

Speech synthesis is the generation of human sounding speech without directly using a human voice. Speech synthesis systems are often called TTS because of their ability to convert text into speech. Such systems are composed of two main parts – a front end and a back end. The front end takes input in the form of text and outputs a symbolic linguistic representation. The back end takes the symbolic linguistic representation as input and outputs the synthesised speech as a waveform. The software, or hardware capable of rendering artificial speech is called a speech synthesiser and the naturalness of these refers to how much the output sounds like the speech of a real person. The two main technologies used for generating synthetic speech waveforms come under the categories of concatenative synthesis and formant synthesis.

Concatenative synthesis is based on the concatenation of segments of pre-recorded speech. This method results in the most natural sounding synthesised speech. However, natural variation in speech and automated techniques for segmenting the waveforms may result in audible glitches in the output, detracting from the naturalness. There are three main types of concatenative synthesis: unit selection synthesis, diphone synthesis and domain-specific synthesis. Formant synthesis does not use any human speech samples at runtime. Instead, the output synthesised speech is created using an acoustic model. Using parameters such as fundamental frequency, voicing and noise levels varied over time, a waveform of artificial speech is created. Most systems based on formant synthesis technology generate artificial, robotic sounding speech.

The author explores the theories that end users, who may already have SEN and already face adversity, should be offered culturally appropriate software which may increase usability. In terms of TTS technology, this would include output speech that sounds as though it is in the same accent as the end users, meaning that cultural background must be considered at the design stage of TTS systems. In addition, the quality of the speech that is output is paramount and it is important to add expressiveness to voice from a computerised system; without such expressiveness it will be very difficult for end users, particularly those with SEN, to be able to put the output into context.

Huckvale (2002), considers the lack of expression or interest or emotion in the speech of current systems is due to the fact that the systems do not actually understand what they are saying, see no purpose to the communication, nor actually have any desires of their own and that it makes no sense to add expressiveness or emotion to voice output from a mindless system.

However in contrast, Zajicek and Jonsson (2006), consider that listeners to speech based information systems attribute qualities to voices which will make them receptive, or otherwise, to the information. Humans are well tuned to detect characteristics in a voice and use that skill both when communicating with humans and with speech based computer systems (Brave et al, 2005). Humans also assess the characteristics of the voice and this affects their perception of the liking and credibility of what is said, ie the content of the information coveyed by the speech based system (Brave et al, 2005) and that both synthesised and recorded voices influence perception of content so that a happy voice makes content seem happier and a sad voice makes content seem less happy.

Lazarsfield and Merton (1964), consider that better human communication occurs between a source and a receiver who are homophilous and have a common frame of reference. Rogers and Bhowmik (1970), also consider that communication is more effective when source and receiver are similar. When two individuals share common meanings, belief and mutual understandings, communication between them is more likely to be effective. Jonsson et al (2005), considers that characteristics of voice can also influence people's attention and affect their performance.

Murray and Arnott (2008), state that all speech produced by humans includes information about the speaker, including the emotional state of the speaker. It is thus desirable to include vocal affect in any synthetic speech where improving the naturalness of the speech produced is important.

1.5 Previous Research in Usability of Software for End Users with SEN and TTS

This area of research is unique. Williams et al (2006), state that there has been a surprising lack of research into usability with regard to people with disabilities and even less concerning those with learning disabilities. Williams et al (2006), go on to state that most of the usability studies that have been performed have dealt with web-based resources and services. Harryson et al (2004), states that as far as disabled people are concerned, accessibility guidelines almost entirely support people with low vision while accessibility guidelines for people with cognitive limitations are almost non-existant. Williams et al (2006), are currently working on a research project on using ICT to facilitate self-advocacy and learning for SEN learners but this research does not consider TTS or multicultural aspects.

Harryson et al (2004), undertook a small scale usability study specifically designed to examine computer use by people with cognitive, as opposed to physical, disabilities where a small sample of seven subjects were observed as they navigated between different web pages using a standard web browser. Subjects ranged between 15 and 44 years in age and were set a series of web navigation tasks on a selection of chosen web sites. Harryson et al (2004), concluded that the processing of text could impede accessibility to the Internet for people with cognitive disabilities and suggest that screen readers and text scanning technology would support this user group.

In another research activity Brown et al (2002), also examined the usability of a system for people with learning difficulties, but used a panel of 'experts' in the study instead of the end users themselves. Brown et al (2002), developed an interactive multimedia learning environment (IMLE) designed to be used by socially excluded people, in addition to those with SEN. To examine the usability of the IMLE, a panel of representatives from a UK university, UK college, Mencap and other organisations undertook five tasks; the results suggested that accessibility should be improved by a speech alternative to the navigational buttons and to key text used within the IMLE.

The Disability Rights Commission (2004), undertook a usability study with disabled users which commissioned the Centre for Human Computer Interaction Design at City University, London to survey a large and representative sample of web sites used by the British public with the aim to investigate web site accessibility for disabled people. The investigation was confined to publicly accessibile sites. Using a commercially available software tool, the home pages of 1000 sites were tested. The impairment groups represented in the user testing were blind and partially sighted people, profoundly and slightly deaf people, people with specific learning disabilities such as dyslexia and physically impaired people.

The report concluded that many of the problems encountered by users are of a nature that designers alone cannot be expected to recognise and remedy. These problems can only be resolved by including disabled users directly in the design and evaluation.

Rowan et al (2000), also looked at usability of websites, although they did not conduct a usability study. However, they did recommend a practical methodology for use in assessing websites. Rowan et al (2000), maintain that the majority of web based information, facilities and services is unnecessarily inaccessible to people with disabilities, due to the lack of awareness of accessibility issues on the part of developers, and argue that currently available accessibility evaluation methods are unsatisfactory in the scope and presentation of their results. Consequently, there is a need for a method which utilises the strengths of current methods, but which also bridges their weakneses.

Although it has now been several years since the British government recognised the importance and benefits of the use of ICT for people with SEN, research into different aspects of using ICT for people with SEN is not considerable (Williams et al, 2006). While some groups of disabled people, particularly those with visual impairments have been the focus of many studies – such as those examining usability or accessibility, it is difficult to find any literature for people with SEN, despite evidence suggesting that this group are able to use computers and gain much benefit from doing so (Williams et al, 2006).

Banes and Walter (2002), and Florian (2004), consider that the use of ICT as a teaching and learning tool in environments with SEN can help in communication and that in addition it is also exciting to use and a positive challenge for the learners. Larcher (2000), states that success with ICT lies in the context of its use rather than in the actual software itself, maintaining that collaborative play can provide an opportunity to apply interactive communication and language skills (such as vocabulary). ICT can also help make more traditional learning activities fun. Keates (2000), considers that one of the main groups of people with special educational needs who could potentially benefit from ICT are those with dyslexia.

Blin and Munro (2008), consider that the advent of the Internet heralded predictions that the use of ICT would transform and disrupt teaching practices but the anticipated disruption has not come to fruition. Although technology is now common place in most higher education institutions, there is little evidence of significant impact on teaching practices and current implementations are accused of being focused on improving administration and replicating behaviourist, content-driven models as opposed to being used as a tool to aid teaching and learning.

There is some evidence in the literature to suggest that research is being undertaken into TTS synthesis in other languages, as opposed to accents. Yoon (2008), is developing a prosodically-sensitive concatenative TTS synthesis system in Korean, but this is for use in Korea with end users who are native to that culture.

Jones et al (2008), undertook an experiment on synthesised speech to analyse and evaluate the speech rate with native and non-native listeners. This experiment assessed the variation in speech rate on comprehension and persuasiveness of a message presented in TTS synthesis to native and non-native listeners. These experiments were conducted in Australia using banking product descriptions using subjects who were native speakers of Australian English or non-native speakers of Australian English. The findings of the experiments suggested that a faster rate lowers comprehension for both native and non-native listeners but does not influence the persuasiveness of the message. Jones et al (2008), go on to state that these findings have implications for the selection of speech rates for persuasive messages delivered to native and non-native listeners using TTS.

Moustroufas and Digalakis (2007), undertook experiments to evaluate the pronunciation of students of a foreign language without any knowledge of the uttered text. Previous attempts have shown that it is feasible to evaluate the pronounciation of a non-native speaker by having implicit or explicit knowledge of the uttered text, provided that enough utterances are available. The approach adopted by Moustroufas and Digalakis (2007), was to use characteristics of the mother tounge (source) of the speaker in the evaluation of his/her pronounciation. They then recorded 20 Greek students speaking English (target) and evaluated their pronounciation using algorithms. The findings of the experiments showed that the pronounciation scores that were based on both the target and source languages had a better correlation with the human scores than those based only on the characteristics of the target language.

Some other research that has been conducted into speech concerning non-native speakers, is in the field of speech recognition as opposed to TTS. He and Zhao (2007), undertook experiments with American English speech, British English speech and mandarin Chinese accented speech. The results of the experiments showed that the use of prior knowledge of accents enabled more reliable estimation of bias distributions.

1.6 Culture

There are many definitions of culture, for example Hall (1959), wrote the following: "Culture stands for the way of life of a people, for the sum of their learned behaviour patterns, attitudes and material things". One of the most prominent areas of research undertaken in cultural issues has been by Hofstede (1991), who identified five distinct variables: Power Distance Index (PDI); Uncertainty Avoidance Index (UAI); Individualism (IDV); Long Term Orientation (LTO) and; Masculinity (MAS).

Hall et al (1987), consider that despite popular beliefs to the contrary, one of the biggest barriers to success is the one erected by culture. The European Telecommunications Standard Institute Technical Human Factors (ETSITHF 2005), state that with the enlargement of the European Union, citizens will come from countries that have an ever-increasing range of national cultures and languages. The number of tourists and immigrant workers who use non-European languages will also increase within these countries. Therefore, the range of cultures and languages that must be supported in European communication and information services will need to grow significantly.

1.7 Aim

The overall aim of this research is to investigate whether the effects of culture and accent have any significance on the usability of TTS technology as a teaching and learning tool for end users of post compulsory school age (16+), for whom English may not be their first language and may have SEN.

1.8 Objectives

There are four objectives which are explored by the testing of hypotheses, through the conducting of three separate sets of strictly controlled experiments, in full compliance with ethical and industry standards. By using this approach, the results are reliable and enable both quantitative and qualitiative methods to be used to display the findings. The following table shows the research objectives with the associated hypotheses being tested.

Table 1.1 Research Objectives, Method and Hypotheses

Objective		**Method**	**Hypothesis**	**Chapter Reference**
(i)	Analyse and evaluate the significance of the effects of culture and accent on the usability of TTS applications.	Controlled laboratory experiments with 44 male and female subjects from English and Indian cultures, using synthesised speech in the English language with English and Indian accents representing male and female voices.	a) There is no difference in error rate when the accent of the speech is representative of the culture of the subjects.	Chapter 4
(ii)	Analyse and evaluate the significance of the effects of gender on the usability of TTS applications.	Controlled laboratory experiments with 44 male and female subjects from English and Indian cultures, using synthesised speech in the English language with English and Indian accents representing male and female voices to identify whether there is a phenomenon regarding the effects of gender , which needs to be included in the experimental designs.	b) The gender of the subjects makes no difference to the error rate.	Chapter 4
(iii)	Analyse and evaluate the significance of the effects of localised speech on the usability of TTS applications.	Controlled experiments with 10 male and female subjects from African cultures, using synthesised speech in the English language with English accent and in the Kiswahili language.	c) There is no difference in error rate when the language of the speech is representative of the culture of the subjects.	Chapter 5
(iv)	Use the results of objectives (i) to (iii) to develop a bespoke system and evaluate its learnability and satisfaction	Develop a prototype talking book and undertake controlled experiments with 40 male and female subjects from Chinese, Czech, English, German, Ghanian, Indian, Nepalese, Pakistani, Phillipino, Swedish, and Zimbabwean cultures, using synthesised speech in the English language with English accent.	d) There is no difference in learnability when the accent of the speech is representative of the culture of the subjects. e) There is no difference in satisfaction when the accent of the speech is representative of the culture of the subjects.	Chapter 6

1.9 Structure of the Thesis

In addition to Chapter One, there are six further chapters to this thesis; an overview of each follows.

Chapter Two

Chapter Two consists of an extensive literature review. Legislation, moral and ethical issues relating to SEN are discussed. Various definitions of culture are explored, particularly the cultural model by Hofstede. The author explores the theory that a person's accent is not just something that is casually adopted but, moreover, represents something far deeper; it is representative of their culture and their way of life to which they may have been subjected to since birth and that, subsequently, the use of TTS technology should consider the inclusion of accents. Previous research into culture, speech, language, learning and the links between them are detailed, exploring the philosophy of Vygotsky. Explanations of how human speech and synthetic speech are formed are given. The chapter ends with HCI, usability issues and details of previous usability tests that have been undertaken within the field of education, SEN and TTS technology, including the work of Zajicek and Brewster.

Chapter Three

In Chapter Three, there is a discussion of different research methods and explanations are given of why the post-positivist approach has been selected. Ethical and moral issues are considered fully, including the detailed procedure undertaken to submit a full testing proposal for approval which was critical to ensure that potential harm through any testing procedures were minimised, before testing could proceed with vulnerable human subjects.

Chapter Four

This chapter details the first set of experiments that were undertaken to test error rates using generic TTS software in the English language, in both English and Indian accent, using subjects from both English and Indian cultures. Subjects from English and Indian cultures were selected due to the representation of these cultures in the TTS software being tested. These experiments use word pairs from the Design Rhyme Test (DRT), which is an established method that is used for testing TTS and by doing so, an industry standard method is being applied to the testing procedure; all of the word pairs are weighted equally. However, in addition to undertaking the DRT independently where just sets of words are heard which are out of context, a modified version of this test is also used to include speech that is in context and with this test, subjects are given word pairs to complete a sentence. By using this modified method of the test, there is the opportunity to make comparisons to establish whether hearing the speech in context has any effect on error rate.

There is also an exploration of whether there is a phenomenon that needs further investigation regarding the gender of the subjects.

Chapter Five

Chapter Five details the second set of experiments that tests an information system that has been developed as an educational tool in both English and Kiswhaili languages, by the Local Language Speech Technology Institute, for use by banana farmers in Kenya, as part of an initiative by the Kenyan Ministry of Agriculture. As this research is unique, there are no existing tests available for this purpose and therefore, with strict adherence to ethical issues and ISO standards, new testing methods have been developed and undertaken with subjects from the target group of end users at their own banana farms in Kenya. The DRT cannot be used for these tests, as the subjects hear large portions of speech including words that may not be in the DRT. The error rate of the system is tested first using questions taken from the information system itself, in full awareness that there will be an element of inconclusion in the creation of the tests as some of the farmers may already know the answer. Therefore, guidance is given in the wording and subject matter of the the questions by agricultural extension workers and colleagues from the University of Nairobi. Prior to the tests, it was decided that if the error rates for the system were extremely low, testing could then be repeated on subjects outside of the target group who would not know the answers.

Chapter Six

In this chapter, the third sets of experiments are detailed which tests a talking book created as a result of the findings of this research. This talking book will be used as a teaching and learning tool for end users with SEN, from different cultures - Chinese, Czech, English, German, Ghanian, Indian, Nepalese, Pakistani, Phillipino, Swedish, and Zimbabwean – and incorporates the use of TTS technology to complement a graphical user interface (GUI). The culture of the subjects are based on convenience sampling. Tests are undertaken using the principles of the standard usability measurement inventory questionnaire method (SUMI) which is an industry wide accepted standard of measuring usability and the significance of the effects of the culture of the speech and the subjects, from multicultural backgrounds, are analysed and evaluated. Due to the structure of the talking book, it would be inappropriate to measure error rates, as the subjects are able to see the words on the screen in addition to hearing them, so the results would be inconclusive. The testing on this system, therefore, focuses on the measurement of the learnability and satisfaction components of usability.

Chapter Seven

This final chapter contains a discussion and conclusion for the research undertaken. Each objective is referred to in turn drawing together key points of research and ends with suggestions for further research and development opportunites and a conclusion.

Chapter Two – Literature Review

2.1 Chapter Introduction

With the relatively recent introduction of special needs legislation into the United Kingdom, there has been a shift of attention towards catering for the needs of learners who have learning disabilities. This legislation exists to ensure that people who have an educational need that is special in some way are not disadvantaged from those who do not have a SEN. It could be argued that all learners have SEN because everybody is unique, indeed the Every Child Matters legislation that was introduced into the UK, uses this as its underlying philosophy, to ensure that no child is disadvantaged in realising his/her highest potential socially and intellectually. The use of ICT as a teaching and learning tool has become a key area in supporting such learners, with computer related technology being adapted to assist end users, more commonly in terms of providing keyboards or mice, that have been especially adapted for them to use. However the theory is explored that an individual's culture may affect their preferred teaching and learning styles and that the way someone speaks is not just an accent that they have casually adopted, but that it represents something far deeper – it is a status of their cultural background to which they may have been subjected to since birth. Therefore when such assistive computer related technology is developed it is not enough to just consider the technology itself, but the author explores whether the cultural status of the end user is also a vital consideration, which is an area of research that up until now has been overlooked. The author explores whether developers and end users misunderstand each other in the development of educational systems for use by learners with special SEN from different cultures and whether cultural differences are taken into consideration, or overlooked, resulting in systems that more closely match the needs perceived by the developers.

This research is unique but incorporates several other areas of research that have been studied as separate objects. This chapter begins by looking at the legislation that exists regarding accessibility and inclusivity in education and sets the scene for the obligations for each educational establishment to meet the legal requirements. The background of culture is then explored, with definitions of culture given including the cultural models by Hofstede and the definition of culture adopted in this thesis.

The links between culture and language are explored, including the work of Vygotsky. An explanation is given of how human speech is formed followed by a description of how synthetic speech is created. Usability, Human Computer Interaction and examples of previous evaluation methods, in particular those used by Zajicek and Brewster are given.

2.2 Accessibility and Inclusivity in Education

The Education Act 1996 considers all learners up to eighteen years of age. Section 312 of this Act refers to SEN and states that a learner has SEN if he or she has a learning difficulty which calls for special educational provision to be made for him or her. This Act also states that a learner must not be regarded as having a learning difficulty solely because the language of the home is different from the language in which he or she is or will be taught. This implies that a student learning English as an additional language, may also have a SEN, but that this assumption must not be made if the only reason for their learning difficulties is that they do not initially speak the language of the educational establishment.

The above Act applies only to learners up to eighteen years of age. However, learning is now widely accepted and encouraged to be lifelong and subsequently, learners with SEN will be found in all stages of education and inclusive education should, therefore, be available for all. The British Psychological Society: Inclusive Education Position Paper 2002, considers that inclusive education is:

- rejecting segregation or exclusion of learners for whatever reason – ability, gender, language, care status, family income, disability, sexuality, colour, religion or ethnic origin;

- maximising the participation of all learners in the educational community of their choice;

- making learning more meaningful and relevant for all, particularly those learners most vulnerable to exclusionary pressure;

- rethinking and restructuring policies, curricula, culture and practices in schools and learning environments to accommodate diverse learning needs.

The United Nations Educational Scientific and Cultural Organisation (UNESCO) 1994, considers that inclusion and participation are essential to human dignity and to the enjoyment and exercise of human rights. Within the field of education this is reflected in the development of strategies to bring about a genuine equalisation of opportunity.

In the field of widening participation for adults with learning difficulties in continuing education, Sutcliffe and Jacobsen (2001), consider that such learners:

- are usually black or from other minority groups;

- are female;

- have profound/multiple difficulties or challenging behaviour;

- have sensory disabilities;

- have a dual label of learning difficulties and mental health difficulties;

- have a history of institutionalisation.

Sutcliffe and Jacobsen (2001), state that in 1996 the National Institute of Adult Continuing Education (NIACE) addressed the impact of the Further and Higher Education Act 1992, which applies to learners over the age of secondary education, in the area of continuing education for adults with learning disabilities. This research showed that provision for certain groups, such as older adults and those with profound and multiple learning difficulties, had diminished since the advent of the Further and Higher Education Act 1992, with reasons being given by education providers that they feared the new emphasis on accreditation and progression could exclude some groups of learners.

In 1995 the Disability Discrimination Act (DDA) was introduced in the UK with the intention of ending the discrimination encountered by many disabled people. The main focus of the Act was the employment of disabled people and their access to goods and services. However, the Act also affected education institutions in the post 16 sector who had to make what was termed to be 'reasonable adjustments' in the way they provided their services to make sure that they made them accessible to disabled people.

The 1997 Dearing Report reflected the Aims of the DDA when it was argued that access to education should be widened in order to include students who might previously have been disadvantaged because of their socio-economic status, gender, ability level, ethnic background, geographical location or special educational needs (Dearing, 1997).

The DDA was amended to include education specifically in 2001 by the introduction of The Special Educational Needs and Disability Act (SENDA); the teaching and learning components of this Act came into force in 2002 with a resulting major impact upon education within higher and further education (Banes and Brave, 2002). This Act made it an offence to discriminate against a disabled person by treating him/her less favourably than others for a reason relating to his/her disability. For the purposes of this Act, a disabled person is defined as someone who has a physical or mental impairment, which has an effect on his/her ability to carry out normal day to day activities (Willder, 2002).

The Act covers all aspects of an educational institution's student services which Willder (2002), defines as 'services of any description which are provided wholly or mainly for students' and will obviously include educational services such as teaching and learning provision.

In this legislation, SENDA describes discrimination as occurring where, for a reason relating to his/her disability, a person is treated 'less favourably' than someone without that disability and that such treatment cannot be justified.

Furthermore, discrimination can also occur when an institution fails to make a reasonable adjustment which results in the students being placed at a substantial disadvantage in comparison with someone who is not disabled ; such failure cannot be justified (Willder, 2002). However, discrimination does not occur if the institution does not know or could not have been reasonably expected to know that the student was disabled.

Frederickson and Cline (2002), consider that people are classified as having a SEN if they have learning difficulties or disabilities which significantly affect their access to traditional methods of teaching and learning; the types and levels of disability are diverse and are personal to each individual. An individual may have one disability, but more usually, a combination of several. According to Frederickson and Cline (2002), the traditional way of considering SEN was as an individual deviation from what is considered normal when the individual concerned has significant difficulties in learning compared with the majority of other people of the same age; this view holds that it is best understood by looking at individual differences between people.

An alternative more recent approach, however, argues that SEN arise when inappropriate environmental demands are placed on an individual which exceed their current capabilities (Frederickson and Cline, 2002). This final point is of particular relevance to this research, as it is important that the information systems that are used by the end users, do not exceed their current capabilities and that the environmental demands are incorporated in their cultural needs.

With the first approach of considering individual differences, the assumption is that the cause of the difficulties lies with the individual with biological, cognitive or behavioural factors within them that prevent them from functioning or developing in the same way as the majority of other individuals of the same age (Frederickson and Cline, 2002) who consider that the problems associated with this approach is that:

- it is based on assumptions which do not withstand scrutiny, for example, the assumption that all people, including those learning English as an additional language, have had the same learning opportunities as other individuals of the same age;

- considerations are not given to the differences that the individual's social context and educational environment make to the individual's level of achievement;

- focusing exclusively on the individual and not on any other factors that may be inherent, is failing to address the problem. For example, it may not be appreciated that an educational establishment is making no provision to provide extra English tuition for a learner who has recently joined the country.

With the second approach of focusing on environmental demands, the assumption is made that the learner's current attainments reflect their previous learning experiences (Frederickson and Cline, 2002). Therefore, if they are taught appropriately they will learn more successfully. It is considered that the individual's learning problems arise because of a mismatch between their current skills and what the curriculum requires of them. In this view, the responsibility for problems faced by people with physical disabilities can be placed on the environment rather than the people themselves with the theory that wheelchair users do not have impaired mobility but rather that buildings are designed in ways that are inappropriate for them. Frederickson and Cline (2002), consider however that the problem with this approach is that individual differences are also important, since different people respond to different methods of teaching and learning.

Using the analogy of the buildings being designed in ways that are inappropriate for wheelchair users, an investigation is need to explore whether the information systems that are developed for end users with SEN are designed in a way that is appropriate for them, rather than to suit the needs of the developer.

Frederickson and Cline (2002), continue to state that both of the above approaches are simplistic views and a preferable view, which is the one that is now widely accepted, relies on an interactional analysis. This views the level of need as the result of a complex interaction between a) the individual's strengths and weaknesses; b) the level of support available; and c) the appropriateness of the education being provided.

2.3 Facilitating Access Through The Use of Assistive Technology

One of the ways in which access to learning resources and teaching material can be facilitated is through the use of assistive technology which Doyle and Robson (2002), define as 'equipment and software that are used to maintain or improve the functional capabilities of a person with a disability'.

Willder (2002), considers that the use of assistive technology will become the norm due to the implementation of SENDA

Banes and Brave (2002), consider that the types of assistive technology that may be used includes technology that:

- facilitates access to a standard PC;
- facilitates access to the Internet;
- facilitates access to end manipulation of the written word;
- facilitates access to and manipulation of the spoken word;
- helps to compensate for cognitive deficits.

Assistive technology includes hardware such as scanners, adapted keyboards and speech technology (Banes and Brave, 2002). Information Technology

(IT) permeates every area of the modern educational establishment with most subjects already including some element of IT within the curriculum; any student who cannot access that technology will immediately be at a disadvantage. An increasing number of students are expected to present their coursework and assignments in printed form. However, computer use can appear to be unachievable for some disabled people (Banes and Brave, 2002), although the problem usually lies with the design or interface of the computer.

Banes and Brave (2002), state further that for students with a physical, cognitive, learning, hearing or visual disability, the standard screen, keyboard, mouse and desktop may be unproductive, uncomfortable or even impossible to access.

2.4 The Impact of European Inititatives on Education and Assistive Technology

The eEurope Action Plan 2002 (European Commission Information Society, 2002), which was agreed by the Heads of State and Government, states that 'special attention should be given to disabled people and the fight against info-exclusion'. At an e-Government ministerial conference in 2001, the topic of inclusion was discussed and concurred that 'greater account must be taken of specific needs, for example of the physically disabled' (eGovernment Ministerial Declaration, 2001). This means that e-information must be accessible to all people with disabilities (not just students).

Member states of the European Union have an obligation under the Amsterdam Treaty (Treaty of Amsterdam, 1997), to take into accounts the needs of disabled people and measures are currently being implemented to apply this commitment in the domain of IT.

The Design for All Project Team Executive Report (2000), states that 'ideally all products and services should be available to 100% of the population'.

2.5 Multicultural Education

Inasmuch as cultures vary in what they emphasise in the subjects that they teach, there is also cultural diversity in how students participate in the learning process (Samovar and Porter, 2004). How education proceeds in a culture is tied directly to the values and characteristics of the culture with cultural issues differing in teaching methods and styles – lectures versus interaction, co-operation versus competition, silence versus noise, for example.

Chen et al (1998) consider that educational systems must continually adapt to the ever-changing information age needs in the global marketplace, which is an evolutionary process. As society changes, so must the educational system by preparing learners for today's globalised, multicultural information age economy. Chen et al (1998), consider that academic exposure to the multicultural environment will provide students with the skills to excel in the real world. As the business world adjusts its views to fit a changing society, the academic environment must do the same. As students ultimately return to the world outside education, the more fully they learn to recognise and respect differences in the beliefs, values and worldviews of people of varying cultural extraction, the more effectively they will promote a multicultural society beyond the classroom.

Hirson (2002), considers that learners must be provided with intellectual awakening and growth and this intellect is exercised through encountering new people, new ideas and new social constructs. A multicultural student body is important to the experiences of both the dominant and sub-cultures alike. Hirson (2002), considers that diversity prepares students for meaningful participation in our increasingly heterogeneous and multicultural democracy. Samovar and Porter (2004), state further that such an approach to education requires that attention be given to the characteristics of the students that the system intends to educate.

Gay (2000), states that the fundamental aim of culturally responsive pedagogy is to empower ethnically diverse students through academic success, cultural affiliation and personal efficacy. Knowledge in the form of curriculum content is central to this empowerment. To be effective, this knowledge must be accessible to students and connected to their lives and experiences outside of education.

Samovar and Porter (2004), state that multicultural education emphasises that human communication is often dependent on one's knowledge of culture. In this approach, teachers and students gain cultural knowledge and this information can enhance appreciation and sensitivity of students' own as well as others' cultures. This enhanced view of culture can help misconceptions of culture that at one time may have caused miscommunication between members of diverse cultures. Samovar and Porter (2004), go on to state that a successful multicultural classroom is one in which students and teachers understand each other in their communicative interactions.

2.6 The Formation of Human Speech

Co-articulation, which is the smearing of sounds into one another is also responsible for making speech sound natural, and is also very difficult to achieve in synthetic speech. In human speech, when a particular gesture is being produced, the next one is already anticipated resulting in the ways in which the articulators move. To glean an appreciation of the complexity of synthetic speech, it is vital to understand first the intensely complex way in which speech is formed by human beings. The way human beings modulate their voices by articulating the laryngeal tones into audible oral speech, set them aside from all other species. Dutoit (1997), states that it is one of the mediums used to convey information and that it is in several non-mutually exclusive levels. The speech waveform (acoustic) coveys linguistic information such as the speaker's tone and speaker's emotion (Dutoit, 2002). Breen et al (1998), considers that speech consists of two types of variability - random variability, from the in-exactitude of the articulators and predictable variability from the linguistic structure, speakers' sex and speaking styles.

Speech is produced when air is forced from the lungs through the vocal cords and along the vocal tract, which extends from the opening in the vocal cords, called the glottis, to the mouth.

Owens (1983), states that at a linguistic level, speech can be viewed as a sequence of basic sound units called phonemes. These phonemes are abstract linguistic units and may not be directly observed in the speech signal itself. Each phoneme may be responsible for more than just one sound, depending on the other phonemes which surround it, and at the acoustic level these are referred to as allophones. When human beings speak using the same prose, they are conveying the same information using the string of phonemes but they sound very different to each other. This variation in sound is due to several factors but, in particular, changes in dialect, vocal tract length and shape and, since the shape of the vocal tract is controlled by the speech articulators, phonemes correspond directly to articulatory gestures (positions and movements). Articulatory gestures are termed to be either static or dynamic depending on whether or not the articulators move. The English language can be categorised in a set of forty phonemes or articulatory gestures (Owens, 1983).

Speech sounds can be categorised into three main groups according to their mode of excitation (Owens, 1983).

1) **Voiced sounds**, for example "aah" or "oh: are produced when the

vocal cords vibrate open and closed, thus interpreting the flow of air from the lungs to the vocal tract and producing quasi-periodic pulses of air as the excitation. The vibration of the cords produces an airflow waveform which is triangular. The rate of the actual opening and closing gives the pitch of the sound which can be adjusted by varying the shape and tension of the vocal cords and the pressure of the air behind them. In an adult, the range of pitch is 5 kHz to 25 kHz with an average value of 12 kHz for males and 15 kHz for females.

2) **Unvoiced sounds**, for example "h" result when the vocal cords do not vibrate and the excitation is a noise-like turbulence produced by forcing air at high velocities through a constriction in the vocal tract when the glottis is held open. Such sounds show little long term periodicity, although short term correlations due to the vocal tract, are still present.

3) **Plosive sounds**, for example "phh" in pin result when a complete closure is made in the vocal tract and air pressure is built up behind this closure and released suddenly. The rapid release of this pressure provides a transient excitation of the vocal tract, which may occur with or without vocal cord vibration to produce voiced or unvoiced plosive sounds.

Some sounds, however do not fall into any one of the categories specifically but are a mixture.

Vowel gestures are produced by voiced excitations of the vocal tract. The articulators remain static, there is no nasal coupling and the sound radiation comes from the mouth. Each vowel is characterised by the forward, backward, raised and lowered positions of the tongue. Vowels, therefore, are classified as front, middle or back depending on the position of the tongue during articulation (Owens, 1983). At the acoustic level, each vowel is characterised by the values of the first three or four resonances of the vocal tract. Semi-vowels are produced in a similar manner but when glides are produced, the articulators move rapidly between static vowel positions. Dipthongs, being a combination of two vowel sounds, are produced similarly although the articulators move slower between static vowel positions (Owens, 1983).

Nasals are produced by vocal cord excitation when the vocal tract is totally constricted at some point along the oral passageway; as the name suggests the sound comes from the nostrils rather than the mouth making the oral cavity act as a side branch resonator trapping acoustic energies at certain frequencies. The resonant frequencies of the oral cavity appear as anti-resonances or zeros (Owens, 1983).

Frequently referred to as stop consonants, plosives are generated by a total closure of the vocal tract allowing the air pressure to build up and then releasing it suddenly. The voiced plosives are distinguished from their unvoiced counterparts depending on whether vocal cord vibration is present or not. Plosives are characterised by intermittent bursts of energy; their properties are highly influenced by the sounds, which precede or succeed them. Affricates can be interpreted as the consonant equivalent of a diphthong and can be categorised as a voice affricate judge and an unvoiced affricate church, in the English language which are produced when a stop and fricative consonant are both shortened and combined.

2.7 Interpersonal Communication

A study has been undertaken to examine virtual community quality through socioloinguistics theory. Gefen and Ridings (2005), state that according to sociolinguistics, in oral discourse men communicate to establish superior social standing, while women communicate with the undertone of rapport, compassion, and empathy. The study undertaken by Gefen and Ridings (2005), shows that these differences carry over to the asynchronous written environment of virtual communities and affect men's and women's respective perceptions of community quality. Women go to virtual communities to give and to get social support and have a more favourable assessment of the capability of others. According to Gefen and Ridings (2005), this pattern generally holds even when comparing mostly single-gender communites and mixed-gender communities. However, on closer investigation, undertones were found in mixed-gender communites being less than in their respective mostly single-gender communities.

Gumperz (1982), refers to human communication as channelled and constrained by a multilevel system of learned, automatically produced and closely coordinated verbal and non-verbal signals. The most significant insights into how signs affect verbal changes come from studies in speaker-listener coordination (Kendon et al, 1975). When the relationship of speakership moves to listeners' responses were measured, it was found that these tend to be synchronised in such a way that moves and responses follow each other at regular rhythmic intervals. The timing of the responses is much faster than one would expect if unpredictable stimuli were responded to (Kempton, 1981), suggesting that conversational synchrony requires some degree of predictability, such as is most commonly acquired by shared culture and similarity of interactive experience.

Research has been undertaken on the communication between interethnic groups (Erickson and Schultz, 1982) where a series of student-counsellor advisory sessions were filmed and tape recorded in which ethnic backgrounds of both counsellors and students varied. The interaction that is expected in sessions such as these are either neutral or instrumental, directed towards the goal of helping the student in planning. Indeed the counsellors cannot be said to be prejudiced, as defined by the usual attitude measures. However, the findings of the study showed that there were significant differences in the amount of useful information that the student obtained, depending on the ethnicity of the participants (Erickson and Schultz, 1982). Furthermore, the results suggested that Poles communicated most comfortably and easily with other Poles, but less easily with Italians and least easily with Puerto Ricans and blacks. Black counsellors were less affected by ethnically different students than the white counsellors.

Other studies have been undertaken with new female employees, of varying cultures, at a major British airport (Gumperz, 1982). Females of Indian cultures were perceived as being surly and un-cooperative, but the findings suggested that the intonation patterns used by native English speaking customers and colleagues were strange to them and that no rudeness was intended.

Indian English speech systematically contasts with Western English speech in its prosodic treatment of simple sentences and the distribution of stress, with Indian English eiher sounding full of stress and staccato, or droning and monotonous to Western English speakers, (Gumperz, 1982). The reason for this is because Indian English speakers rarely reduce syllables and pronounce all consonants with a higher degree of articulation than native English speakers, thus in one sense employing a great deal of stress, but at the same time, no syllables are stressed significantly more than others.

The two main features of Indian English speech, in contrast to Western English speech, therefore, are a) the subdivision of utterance into small chunks and b) the rhythmic marking of stress by several words with no one syllable made tonally prominent.

2.8 Speech and Memory

Pisoni et al (1996), state that a multi-modal speech signal, where people can see the speaker's face (either in person, through the television, webcam or video telephone for example), is robust, informative and provides information that perceivers are able to exploit during perceptual analysis. It is suggested that the addition of visual information in the stimulus display about the speakers' articulation affects the efficiency of initial encoding operations at the time of perception and also results in more detailed and robust representations of the stimulus events in memory (Pisoni et al, 1996).

Mullennix et al (1989), reported that intelligibility for isolated words in noise is influenced by the number of talkers that are presented in the experimental condition. Experiments had been undertaken where listeners had been presented with lists spoken by a single talker or lists spoken by fifteen different talkers. Results were consistent and showed that identification performance was always better for words presented in single-talker lists than multiple-talker lists. Pisoni et al (1996), consider that these findings suggest that under high talker-variability conditions, listeners engage in some form of on-line recalibration each time a new voice is encountered in a set of the trials suggesting that spoken word recognition is related to the processing of the talker's voice.

Further experiments to evaluate perceptual learning demonstrated that detailed properties of a talker's voice are encoded into memory and can be used to facilitate word recognition in noise (Nygaard et al, 2004). The results of these experiments suggest that listeners incidentally encode information about the vocal source attributes when listening to different speakers. Nygaard et al (2004), argue that listeners retained a procedural memory for a talker's voice, in addition to specific details about a linguistic event. Thus, the neural representation of spoken words may encompass both an abstract phonetic description of the utterance as well as detailed information about the structural description of the talker's vocal tract.

Pisoni et al (1996), undertook experiments to evaluate the affect of multi-modal speech perception on both immediate memory span and on serial recall of isolated words. These studies compared unimodal and multimodal presentation formats to assess how these sources of information interact and influence the representation of spoken words in memory. According to Pisoni et al (1996), the results of the experiments suggest that the additional visual information about the talkers face is retained and used in subsequent recall. These results also support the findings obtained several years previously by Goldinger et al (1991). The presence of additional information about an item, such as the talker's voice, appears to provide the perceptual system with the ability to build a more detailed or robust representation. The presence of additional stimulus dimensions about a talker's voice may aid retrieval mechanisms which use discriminability and distinctiveness to recover items from memory (Pisoni et al, 1996). Multi-modal representation of speech also helps this elaboration process to work more efficiently (Pisoni et al, 1996).

2.9 <u>Speech Synthesis</u>

The ultimate aim of a speech synthesiser is to be able to synthesise speech that is intelligible, natural and be able to convey speaker characteristics such as age, accent and emotion. Therefore, there are three main characteristics that are inherent in a good quality text to speech synthesiser (How, 2003):

1) A good set of synthesis units and models that can define and represent the speech and the speaker's characteristics. The various nuances of speech could be parameterised using statistical modelling. TTS systems are increasingly dependent on statistical modelling of database speech units. These statistical models could then be applied during automatic speech units selection and as prosodic models (Bellegarda et al, 2001) during the synthesis process;

2) the ability to synthesise natural intonation and spectral content across the concatenated units in relation to the intended linguistic requirement;

3) an efficient spectral and formant smoothing process that ensures spectral and formant continuity and smoothness across the concatenated units.

Concatenative speech synthesis, also referred to as time-domain concatenative speech synthesis, is the process of synthesis by concatenating sub-speech units to form speech (Keller, 2002). In this approach, large databases of human speech are collected, and constituent speech portions comprising of segments, syllables, words and phrases are identified. During the synthesis phase, designated signal portions comprising of diphones, polyphones or even whole phrases, are retrieved from the database according to phonological selection criteria (unit selection), chained together (concatenation) and modified for timing and melody (prosodic modification). Because such speech portions are basically stored and minimally modified sections of human speech, concatenative speech consists by definition only of possible human speech sounds, which in addition preserve the personal characteristics of a specific speaker and this is a significant factor in the improved signal quality of current text to speech systems (Keller, 2002).

According to How (2003), the reasons for concatenative methods producing the most natural and intelligible sounding results are: the improvement in the quality of the output of concatenative TTS synthesis systems is usually proportional to the number of stored speech units available during synthesis. This means that the greater the number and intonational variety of stored speech units, the better the quality of the synthesised speech. Large numbers of stored speech units translates to substantial storage capacity. Spectral and fundamental frequency creates smoothness across concatenated units. This is vital to concatenative systems as the individual units which make up the whole of the synthesised speech usually comes from different sentences. The same word from a different sentence would have different spectral and prosodic characteristics. Concatenative synthesisers requires much less detailed knowledge of speech production compared to a rule based synthesier in the production of synthesised speech (How, 2003).

Speech synthesisers are used in various capacities and the type of application in which it is used will, to a certain degree, determine the storage type of the speech units used during synthesis. In some synthesisers there are only a few specific recorded utterances, for example telephone interactive systems such as BT Answer 1571. In other systems, for example text to speech synthesisers where generation of arbitrary speech is needed, speech units are stored in groups of phones which is the actual pronunciation of a phoneme as diphones (an adjacent pair of phones and the recording of the transition between the two), triphones (three adjacent phones and the recording of the transition between the three) or words (Sagisaka et al, 1992).

The choice of synthesis methods to use depends on the required capability of the system. In many cases, the preference is to use TTS synthesisers rather than systems with just a selection of a few recorded utterances, as they are more flexible, being able to generate any speech without being confined to a few utterances. This is particularly true when the sentences to be generated could not be known in advance (How, 2003) for example, speaking and reading aids for the disabled, proof reading of documents and reading of messages, including e-mails. Even when only a fixed number of utterances are needed, the preference may still be on TTS synthesis, depending on the scale of the system; exceptions to this might be systems such as voice alarm systems and telephone interactive voice response systems (IVRS), where the number of utterances remain few, fixed and unlikely to change over time.

However, for systems that are dynamic, where access to very large and rapidly changing databases are needed (How, 2003), TTS synthesis systems are more advantageous. The use of TTS synthesisers eliminates the need to constantly maintain the set of utterances, which reduces the storage space needed and also the need to recall the original speaker to re-record any new or modified phrases.

The process of being able to synthesise perceptually naturally sounding speech is a significant challenge for researchers and developers. Natural human speech not only conveys the meaning inherent in the spoken words, but also the accent, gender, emotional state and geographical origin of the speaker (How, 2003). Currently, computer synthesised speech its standards of intelligibility and naturalness are inconsistent and synthesised speech at times appears mechanical, lacking in the nuances that are so frequently present in natural human speech (How, 2003).

In addition to the improved signal quality in concatenative synthesis, another major factor in recent improvements has been the refinement of prosodic models (Keller, 2002). These models fall into two categories which are predominantly linguistic and predominantly empirical-statistic (stochastic). For many languages, early linguistically inspired models did not furnish satisfactory results, since they were incapable of providing credible predictive timing schemas or the full texture of a melodic line. As a consequence of these inadequacies, stochastic models have moved into the dominant position among high quality speech synthesis devices. These models generally implement either an array or tree structure of predictive parameters and derive statistical predictors for timing from extensive database material. Keller (2002), states that the prediction parameters do not change a great deal from language to language. Moreover, they generally concern the position in the syllable, word and phrase, the sounds making up a syllable, the preceding and following sounds and the syntactic and lexical status of the word.

Text is a written realisation of speech that conveys the same underlying linguistics information (How, 2003). The input text is first broken down and classified as various phonemic units belonging to that language in a concatenative TTS system. Subsequently, the acoustic segments in the inventory that matches these phonemic units are selected and stringed together. With a concatenative text to speech synthesiser, the procedure is to analyse the text to remove any ambiguities, such as Dr, which could be translated into either doctor or drive, for example, or misspelled words or other words, such as people's names, which do not already exist in the dictionary. The end result once the system has done this is a string of characters (How 2003). The string of characters is then parsed to determine the fundamental frequency, duration and energy structures. Representation of the string of characters in the form of phonetic transcription labels that corresponds to the speech units used by the system is then established. These units are retrieved from storage and, after signal processing, speech is synthesised.

Concatenative synthesisers can be divided into five basic modules (Black et al, 1995).

1) **Inventory.** This is the first module and has to be constructed prior to the actual process of concatenative text to speech synthesis. The inventory

is populated by speech units – and determines the effectiveness and ease of implementation of the formant and intonation and signal processing modules.

2) **Text analysis and phonetic.** This is the second module and comprises two parts – text analysis and phonetic transcription generation. The text analysis removes ambiguities in input text. In addition, it also parses the sentences to obtain the intonation information and structure of the input text. The phonetic transcription that corresponds to the speech unit labels in the inventory is generated by the second process in the module.

3) **Speech unit selection and synthesis module.** This is the third module and selects and synthesises the most appropriate speech units.

4) **Formant and intonation processing module.** This is the fourth module and generates the required targeted formants and intonation values.

5) **Signal processing module which includes the unit concatenation and discontinuity smoothing processes.** This final module generates the speech units with the required formants and intonation version of the units from the inventory. Subsequently these units are concatenated and the resultant smoothed to generate the output synthesised speech.

2.10 Language and Cultural Relationships

Language is the way that human beings primarily communicate with each other. Indeed, many aspects of our culture are explained to us by the use of language both spoken and written. In culture, the use of language is a key consideration. Culture exceeds national borders with many cultures existing within a nation. In the same way, a language can be shared by different cultures, although they may adopt a specific vocabulary, and a culture can use different languages, for example multi-lingual countries such as Switzerland where several language groups might form sub-cultures of their own preferring one specific language. Individuals learn about their culture by reading books, listening to stories, listening to songs and music, looking at paintings, body language, observing people around them, clothing, food and dance. When discussing these influences, the form of communication is language.

Ellis et al (1986), defines language as a set of human channels of communication which provides a give and take of information between two or more individuals in a conversation. Language differs from other channels because it is the most highly developed form of human communication and is the one that sets aside human beings and other animals. Katz (1981), argues that it is possible that language can be treated as an abstract object to be studied as a body of words separate from mental models, culture and other abstract objects. This theory proposes that a language consists of a certain number of words which are used by individuals to express themselves and that different languages exist because of isolated areas during language development. The language would restrict a person's view of the world because he or she cannot see what is expressed in words (Katz, 1981).

Language and culture hold the power to maintain national or cultural identity and that language is important in ethnic and nationalist sentiment because of its powerful and visible symbolism, (Rogers et al, 1998).

It needs to be considered whether culture influences language or whether language influences culture. Whorf (1956), considers that what we see or do is described using language suggesting that language is linked with thought, that we cannot have any thought without language. Without language describing the world around us it would be impossible to form a view of the world and consequently impossible to form a culture. In contrast Brown (1958), states that language is an aspect of culture, perhaps usable to identify sub-cultures but no different from any other cultural aspects such as economic situation and level of education and that language is not significant enough to say it defines culture.

One of the most prominent areas of research into the development of language and cultural relationships was undertaken by Vygotsky. Vygotsky's innovative research in psychology incorporates several key concepts including psychological tools, mediation, internationalization and the zone of proximal development. His work covers diverse topics such as the origin and the psychology of art, development of higher mental functions, philosophy of science and methodology of psychological research, the relation between learning and human development, concept formation, interrelation between language and thought development, play as a psychological phenomenon, the study of learning disabilities and abnormal human development (Vygotsky, 1980).

Vygotsky (1980), investigated child development and how this development was guided by the role of culture and interpersonal communication. Observations were made of how higher mental functions developed historically within particular cultural groups, as well as individually through social interactions with significant people in a child's life, particularly parents, but also other adults. An area of Vygotsky's psychology which is of particular relevance to this research is the field of cultural mediation and internalization. Vygotsky considered that through these interactions, a child came to learn the habits of mind of her/his culture, including speech patterns, written language, and other symbolic knowledge through which the child derives meaning and affected a child's construction of his/her knowledge (cultural mediation). The specific knowledge gained by children through these interactions also represented the shared knowledge of a culture (internalization).

According to Vygotsky (1980), internalization can be defined as a mastery of tools and skills which occurs through the activity of the child within society and the appropriation in which the child takes a tool and makes it his/her own, perhaps using it in a way unique to his/herself. As an example, internalizing the use of a pencil allows the child to use it very much for his/her own ends rather than draw exactly what others in society have drawn previously.

One of Vygotsky's most significant contributions to psychological research that is also of particular relevance to this research, is the inter-relationship of language, development and thought. This concept establishes the explicit and profound connection between speech (both inner speech and oral language) and the development of mental concepts and cognitive awareness. Vygotsky (1980), considers that inner speech is qualitatively different from external speech via a gradual process of internalization, with younger children only really able to think out loud. Vygotsky (1980), claims that in its mature form, it would be unintelligible to anyone except the thinker and would not resemble spoken language as we know it (in particular being greatly compressed). Hence, thought itself develops socially.

An infant learns the meaning of signs through interaction with his/her main carers, crying when an item is wanted for example or gurgling when they are happy, and that these verbal sounds can be used to learn how to conduct social interaction with the infant beginning to utilise and develop this faculty by progressing to asking for the object by name for example (Vygotsky, 1980). Language itself, starts as a tool external to the child used for social interaction. The child guides personal behaviour by using this tool in a thinking out loud manner. According to Vygotsky (1980), thinking out loud (or self-talk) is a tool of social interaction which tapers to negligible levels when the child is alone or with deaf children. Gradually this self-talk is used more as a tool for self-directed and self-regulating behaviour. Then, because speaking has been appropriated and internalized, self-talk is no longer present around the time a child starts school. Vygotsky (1980), states that self talk develops along a rising not declining, curve; it goes through an evolution, not an involution, until in the end it becomes inner speech. Inner speech, therefore, develops through its differentiation from social speech.

Vygotsky (1980), therefore, considers that speaking in a child develops along two lines – the line of social communication and the line of inner speech, by which the child mediates and regulates his/her activity through his/her thoughts which in turn are mediated by the semiotics (the meaningful signs) of inner speech. Inner speech is not comparable in form to external speech. External speech is the process of turning thought into words, whereas inner speech is the opposite, it is the conversion of speech into inward thought. As an example, inner speech contains predicates only, subjects are superfluous and words are used much more economically. One word in inner speech may be so replete with sense to the individual that it would take many words to express it in external speech.

According to Keesing (1965), culture influences you from the instant you are born and that it is a tenet of cultural anthropology that culture tends to be unconscious. Hoebel and Frost (1976), refer to the term enculturation which denotes the total activity of learning one's culture and is the conscious or unconscious conditioning occurring within that process whereby the individual, as child and adult, achieves competence in a particular culture. Rubin (1988), considers that the presence of culture is so subtle and pervasive that it simply goes unnoticed. Luckman (1999), makes the point that although culture provides strength and stability, it is never static. Cultural groups face continual challenges from such powerful forces as environmental upheavals, plagues, wars, migration, the influx of immigrants and the growth of new technologies. As a result, cultures change and evolve over time.

Samovar and Porter (2004), consider that culture is learned. From birth onwards the search for meaning becomes a lifelong endeavour. As an individual moves from word to word, event to event and person to person, they seek meaning in everything. The meanings that are given to these experiences are learned and culturally based. Samovar and Porter (2004), state further that the notion of learning is the single most important characteristic of cultures. Samovar and Porter (2004), consider that language serves two important cultural functions. Firstly, it is the means of preserving culture and secondly it is the medium of transmitting culture to new generations. Language also is important to all aspects of human interaction and that communication serves a variety of purposes that facilitate and maintain cultural, social and individual needs.

Somovar and Porter (2004), consider that intercultural communication is the circumstance in which people from diverse cultural backgrounds interact with each other; cultural diversity has the potential to make intercultural communication very difficult. The crucial element in communication is culture and the impact it has on an individual's communicative behaviour as culture strongly influences an individual's beliefs, values and world views. This is reflected in their use of language, non-verbal behaviour and how they relate to others. Intercultural communication can be described as having two distinct points of contact – international and domestic.

International contacts are those between people from different countries and cultures and domestic is between people of diverse cultural backgrounds that live within an overarching societal group. International contacts means that no culture can remain aloof or autonomous and there are three main developments that have made intercultural contact more axiomatic and pervasive a) new technology and information systems; b) changes in the world's population and c) rapid movement towards a global economy (Samovar and Porter, 2004).

2.11 Culture

In this section, the theory is explored that the world consists of people from many different cultures living in many different countries, where the majority of the population may not share the same culture and that, therefore, an individual's culture is not necessarily dictated by the country in which he or she lives but moreover culture is dictated by their social upbringing by belonging to a certain group. This is an important factor in this research, as there may be assumptions within society, and in particular within education, that just because an individual is living within a certain country, they have automatically adopted their culture and their way of learning. The consideration of factors that are culturally inherent that may affect their methods of learning may be overlooked. There are many different definitions of culture and, although there is not one particular definition that has been accepted as being definitive, all definitions appear to have some commonality.

Hall (1959), considers that culture stands for a way of life of a people, for the sum of their learned behaviour patterns, attitude and material things.

According to Kluckholn (1962), culture is a set of definitions of reality held in common by people who share a distinctive way of life. Definitions of reality include language, values and the norms that set the limits for behaviour. Kluckholn (1962), also considers that it is learned behaviour that is common to a group of people and transmitted by the older generation to its offspring. He argues that such a general statement can be made when human beings, in spite of individual differences, can be thought to think, act, feel and behave in certain ways that are approximately the same for the whole group.

In contrast to the definitions of Kluckholn (1962), Kroeber et al (1958), focus on behaviour patterns as opposed to a group of people. They conclude that culture is transmitted and creates a content of patterns of values, ideas, and other symbolic meaningful systems as factors in the shaping of human behaviour and the artefacts produced through behaviour.

Subjective culture, which can be defined as a cultural group's characteristic way of perceiving the man-made part of it environment, is expressed in objective artefacts. According to Trandis (1972), the way people view the world around them is expressed in cultural products such as paintings, books, folklore and clothing.

Trompenaars (1993), considers that culture is the way in which a group of people solve problems, taking the collective programming of the mind one step further and focusing on applying the way that they think to solve everyday problems.

One of the most prominent and comprehensive areas of research into culture was conducted between 1967 and 1973 by Hofstede, into how values in the workplace were influenced by culture. Hofstede (1980), calls culture the collective programming of the mind, which distinguishes the members of one human group from another. Hofstede (1980), considers that groups of people think in the same way because they share the same learning processes. During his research at IBM, he collected and analysed data from over 100,000 individuals from forty countries and using these results he developed a model that identified five primary dimensions to different cultures. These primary dimensions are discussed in more detail later in this chapter.

Since then Hofstede's research has received scrutiny and criticism for both his methods of defining and delineating national cultural consequences. One of the main criticisms is that this was a cultural study yet Hofstede's results, however, are categorised by country and not by cultural groups. Often, there is more than one cultural group within a country and this implies that there may be a significant deviation from Hofstede's results. Notwithstanding this, however, it has been widely acknowledged in academia and business throughout the world, that using Hofstede's dimensions analysis can play a significant role in better understanding the intercultural differences within regions and between countries.

Within each society there will be a national, dominant culture, which is not necessarily monolithic as within each culture there may be numerous co-cultures and specialised cultures. Victor (2001), suggests that culture is never a homogenous thing of one piece, in every culture there are internal contradictions or polarities. However, throughout this research, the term culture is applied to the dominant culture found in most societies, which is the one in power which allows the segment of the population to speak for the whole while setting the tone for the sub-cultures. Samovar and Porter (2004), consider that intercultural communication involves interaction between people whose cultural perceptions and symbols are distinct enough to alter the communication event and that communication is the basis of all human contact. All human beings regardless of their culture, share common universal experiences. Everyone is a member of the human species sharing universal needs, a member of a specific culture sharing common cultural patterns, and at the same time a distinct person with an individual psychology following a unique script. Regardless of an individual's culture, we all share such common emotions such as fear, love, anger, hostility, shame, guilt, envy, grief and joy (Samovar and Porter, 2004). Each culture has its own forms of ethnocentrism, ego defense, pride and forms of play and in every culture people stress manners and civility to each other. However, each individual is much more than their culture (Samovar and Porter, 2004).

Samovar and Porter (2004), consider that there are five main elements of culture: history, religion, values, social orgnaisation and language.

According to Rogers et al (1998), language and culture hold the power to maintain national or cultural identity and that language is important in ethnic and nationalist sentiment because of its powerful and visible symbolism; it becomes a core symbol. Rogers et al (1998), state that how people think and how they ultimately speak is determined to a large extent by their culture; this is a process known as linguistic relativity. They suggest that the assignment of meaning to a message concerns human perceptions about the relationship between symbols and their referents. Language is used to think as well as speak. Linguistic relativity is the degree to which languge influences human thoughts and meanings. It proposes that in human thought, language intervenes between the symbols and the ideas to which the symbols refer (Rogers et al, 1998).

A definition for a culture, therefore, should provide justification for the fact that common background makes them a group, a common background not just based on the lifetime of one individual but on lifetimes of many individuals that together form the basis of rules and knowledge from culture.

It is impossible to distinguish precisely all of the aspects in which one culture is different from another culture and to set boundaries around a culture, to say where it starts and where it stops. However, it can be concluded that culture controls behaviour in deep and persisting ways, many of which are outside of the daily awareness and beyond the conscious control of the individual. This cultural control is being extended to include language and accent.

2.12 Cultural Variables

Cultural variables measure the degree of belonging to a certain culture. These variables are important to this research as usability tests will be undertaken using subjects from many different cultures and careful consideration will have to be given to testing procedures and ethical considerations, which vary between cultures. In addition, cultural variables will also be an important factor when analysing the results of the tests to identify whether there is any diversification from the model of cultural variables.

As mentioned in the previous section Hofstede (1991), identified five distinct variables as follows:

1) **Power Distance Index (PDI).** This focuses on the degree of equality, or inequality between people in a country's society. A high power distance ranking indicates that inequalities of power and wealth have been allowed to grow within the society. These societies are more likely to follow a caste system that does not allow significant upward mobility of its citizens.

A low power distance ranking indicates the society de-emphasises the differences between citizens' power and wealth. In these societies equality and opportunity for everyone is stressed.

2) **Uncertainty Avoidance Index (UCI).** This focuses on the level of tolerance for uncertainty and ambiguity within the society ie in unstructured situations. A high uncertainty avoidance ranking indicates the country has a low tolerance for uncertainty and ambiguity. This creates rule-orientated society that institutes laws, rules, regulations and controls in order to reduce the amount of uncertainty. A low uncertainty avoidance ranking indicates the country has less concern about ambiguity and uncertainty and has more tolerance for a variety of opinions. This is reflected in a society that is less rule-oriented; more readily accepts change and takes more and greater risks.

3) **Individualism (IDV).** Individualism focuses on the degree the society reinforces individual or collective achievement and interpersonal relationships. A high individualism ranking indicates that individuality and individual rights are paramount within the society. Individuals in these societies may tend to form a larger number of looser relationships. A low individualism ranking typifies societies of a more collectivist nature with close ties between individuals. These cultures reinforce extended families and collectives where everyone takes responsibility for fellow members of their group.

4) **Long Term Orientation (LTO).** This focuses on the degree the society embraces, or does not embrace, long term devotion to traditional, forward thinking values. High long term orientation ranking indicates the country prescribes to the values of long term commitments and respect for tradition. This is thought to support a strong work ethic where long term rewards are expected as a result of today's hard work. However, business may take longer to develop in this society, particularly for an 'outsider'. A low long term orientation ranking indicates the country does not reinforce the concept of long term, traditional orientation. In this culture, change can occur more rapidly as long term traditions and commitments do not become impediments to change.

5) **Masculinity (MAS).** Masculinity focuses on the degree the society reinforces, or does not reinforce, the traditional masculine work role model of male achievement, control and power. A high masculinity ranking indicates the country experiences a high degree of gender differentiation. In these cultures, males dominate a significant proportion of the society and power structure, with females being controlled by male domination. A low masculinity ranking indicates the country has a low level of differentiation and discrimination between genders. In these cultures, females are treated equally to males in all aspects.

As stated earlier in this section, it has been widely acknowledged in academia and business throughout the world, that using Hofstede's dimensions analysis can play a significant role in better understanding the intercultural differences within regions and between countries. Subsequently, therefore this cultural model by Hofstede is the one that is being adopted. However McSweeney (2002), explicitly challenges the work of Hofstede. He argues that Hofstede's research relies on assumptions that are fundamentally flawed. He argues further that these assumptions are crucial in the sense that each is necessary for the plausibility of Hofstede's identification claims; subsequently his natural cultural descriptions are invalid and misleading. McSweeney (2002), states that Hofstede's generalisations about national level culture come from an analysis of small subnational populations which relies on the unproven and unprovable supposition that within each nation there is a uniform national culture and on a mere assertion that micro-local data from a section of IBM employees was representative of that supposed national uniformity. In addition McSweeney (2002), argues that what Hofstede identified is not national culture, but an averaging of situational specific opinions from which dimensions or aspects of national culture are justifiably inferred. The claims of McSweeney, however, contradict Hofstede who emphasised throughout his research that his findings were an average of IBM employees. However, although Hofstede categorised subjects into countries instead of cultural groups, he also undertook further analysis into cultural aspects within each country. As an example, he has provided significant analysis of religious beliefs held within each country; the author explores the theory that religious beliefs are a significant part of a cultural group. When the usability tests are conducted, subjects will include those who come from an English culture and an Indian culture. The findings will be interpreted using Hofstede's cultural model. When Hofstede conducted his research and categorised subjects as "India' he did not travel to India to conduct the tests, nor did he assume that they were all born in India but he did, moreover assume that all of the subjects from India adopted the main culture from that country. Likewise the same strategy will be adopted for analysing the results of the usability tests conducted in this thesis.

McSweeney (2002), considers that Hofstede generalises about the entire national population in each country solely on the basis of analysis of a few questionnaire responses, the respondents being categories of a employees in a single company and did not supply evidence that this was a nationally representative sample. To generalise about an entire nation on the basis of such a small number of questionnaire responses is insignificant. McSweeney (2002), continues that IBM subsidiaries had many nationally atypical characteristics. These included: the company's selective recruitment only from the middle classes; the frequent international training of employees; the technologically advanced and unusual characteristics of its products during his survey periods – which were before the development of the personal computer; the frequent personal contacts between subsidiary and international headquarters staff; its tight, internationally centralised control; US ownership during a period in which foreign direct investment was new and controversial; and the comparatively young age of its managers. Again, McSweeney (2002), does not fully consider that Hofstede was measuring differences not absolutes, with all of the environmental variables being the same, Hofstede's argument was that the differences found must be cultural.

Of the potentially huge number of cultural and non-cultural influences on the questionnaire answers, Hofstede assumed that only three – organisational, occupational and national cultures – were significant. Each respondent, therefore, was conceived of as exclusively carrying only one of these three non-interacting cultures. McSweeney (2002), considered this to be an unreasonable hypothesis and showed this as follows:

$$(OrC + OcC + NC1) - (OrC + OcC + NC2) = NC1 - NC2$$

Where

OrC = organisational culture at IBM

OcC = occupational cultures

NC = national culture

and therefore NC1 – NC2 = difference(s) between the two national cultures.

This suggests that every one occupation has one single global culture, (McSweeney 2002).

Despite the views of McSweeney, the work carried out by Hofstede remains the largest study of this type. When an individual belongs to a culture, their culture is not affected by their type of job, or level of employment within that job – culture is far deeper reaching. Hofstede has never denied that the findings were from the group of IBM employees, and although this has been criticised, it did mean that his tests could be performed consistently and that all of the subsequent analysis has been carried out in the same way across subjects from all cultures. To give an effective argument against the findings of Hofstede, McSweeney would have to conduct similar experiments to disprove Hofstede's findings, not merely attempt to interpret them in a way in which is alien to Hofstede's study.

Ford and Gelderblom (2003), have also used Hofstede's cultural models to research the effects of culture on performance achieved through the use of HCI. Their research focuses on understanding users and their individual differences which result from inter alia, differences in culture. The primary goal of their research was to determine whether Hofstede's cultural dimensions affected the performance achieved through the use of HCI. Ford and Gelderblom (2003), went on to identify the characteristics of the cultural dimensions and then identify test subjects and test interfaces displaying appropriate cultural dimensions and finally assess the impact of these cultural dimensions on the speed, accuracy and satisfaction levels achieved by test subjects using the test interfaces to perform data collection tasks. Test subjects and website interfaces were identified in terms of cultural dimension characteristics. The test subjects were selected based not only on their cultural dimensions, but also by controlling for user profiles. The data resulting from the experiments was then analysed to establish whether these dimensions had any impact on the performance achieved when using these websites. The findings of Ford and Gelderblom (2003), was that the results of the experiments did not provide sufficient evidence to conclude that any of the tested cultural dimensions affected human performance, however, the performance levels attained suggested that the usability of the interfaces was increased for all users, as a result of accommodating high uncertainty avoidance, masculinity, collectivism and high power distance characteristics into the design of the interfaces.

2.13 Human Computer Interaction and Culture

The author explores the theory that end users of computer related technology should be provided with an interface that is compatible with their own culture and that, furthermore, this may improve their satisfaction and success rates with the technology and in the case of education, raise academic achievement. This is an area that has been much overlooked so far with little evidence to suggest that the cultural needs have been taken into consideration when developing such technology. The ideal situation would be one where the system supports the context dependent cultural and language preferences of the end user.

Norman (1998), considers that the goals of HCI are to make tasks easier, more effective, more satisfying to perform and safer. Research suggests that there is now little doubt that culture affects how people interact with computers and that there is a need for designers to consider whether interfaces designed in and for one culture, will work for other cultural groups (De Souza and Dejean, 2000; Ess and Sudweeks, 2000, Dunckley and Jheita, 2004).

Earlier research on cultural determinants in the 1990s focused on the manifestations of culture that are visible including symbols and layout (Del

Galdo, 1990; Barber, 1995) and the use of colours, icons and metaphor (Duncker, 2000; Nielsen, 2000).

More recent research investigated those concepts of culture that are less visible, intangible and culture sensitive, including icons, metaphors, functionality and perceived usefulness (Abdelnour-Nocera and Dunckley, 2005; Duncker, 2000; Evers, 2000). Most of this previous research shows how users from different cultures differ in their perception and preferences and exhibit differences in their user behaviour and general attitudes towards the interface (Kralish and Bettina; 2004: Yeo and Loo, 2004).

Cultural diversity makes it unrealistic to rely on intuition or personal experience of interface designers (Dunckley and Smith, 2000). Therefore, different approaches and techniques have been explored for assessing cultural differences and international design and implementations. These design approaches can be classified as Globalisation, Internationalisation, Localisation and Culturalisation.

2.13.1 Globalisation

Bourges-Waldegg (2000), defines globalisation as the process of planning, designing and producing a product or service which can be used internationally. Globalisation is a life cycle model that includes Internationalisation, Localisation, project management, software development, testing and technical writing resulting in a globalized product that is, in essence, free from cultural bias and intended for use in a range of cultural contexts (Bourges-Waldegg, 2000; Barker et al, 2000; Hall and Webb, 2000).

Internationalisation of software separates the software into two componennts – a culture-independent and a culture-dependent component (Del Galdo, 1996).

2.13.2 Localisation

Evers (1999) states that to provide software for a particular cultural group, the localisation process is conducted. Basic parts of the localisation process include Technical Localisation, National Localisation and Cultural Localisation. Evers (1999), states further that Technical Localisation incorporates technical aspects to be considered to adapt the product into different markets. National Localisation covers translation, support to language, punctuation and formats. Cultural Localisation goes beyond translation and functionality and targets aspects like appeal, taste and beliefs of the end user, as well as localisation of icons, symbols and metaphors to ensure that they are not sensitive to the targeted culture. Cultural Localisation is also refered to as Culturalisation (Barber and Badre, 1998).

Cunliffe (2007), refers to minority languages and considers that where a minority language exists alongside a dominant one, and in the absence of any significant commercial benefit or law, arguments for the provision of minority languages are often based on social grounds. According to Cunliffe (2007), majority language provision will meet the needs of the majority of users, providing they have some skills in the majority language. Many organisations, however, have no social or ethical commitment to a minority language and even where some commitment or obligation exists, there will be a lack of support in resources, resulting in either poorly implemented language provision, or simple translations of majority language provision.

Cunliffe (2007), states further that there is a risk that a translation-based approach to minority language may fail to recognise cultural differences and that software, generally, reflects cultural conventions and social practices of the place in which it is developed and often the localisation of software follows a process that is only superficial, changing the surface elements, but without consideration for the manner in which the software has been designed to operate.

According to Abdelnour-Nocera and Hall (2004), a process of localisation which is more than the translation-based approach is necessary, which recognises the importance of cultural differences both between the producers of the software and the end users, which are factors that may make localised software unusable.

2.14 Usability Evaluation Methods

Usability refers to any part of a system with which a human being interacts. The range of interaction between human and machine is large and diverse and, therefore, usability cannot be classified as a single, one-dimensional property of a user interface but, moreover, can be categorised as having five main components: Learnability; Efficiency; Memorability: Errors: and Satisfaction (Nielsen, 1993). Abratt et al (2003), state that usability testing evalutes how easy a system is to learn and use, offering many benefits to end users. These benefits include: increased pace of learning, decreased user task time and errors and increased job satisfaction. According to ISO 9241-11 (1998), 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.

Gray and Salzman (1998), state that various usability evaluation methods (UEM) have been created, promoted and used to evaluate the interaction of the human with the computer for the purpose of identifying aspects of this interaction that can be improved to increase usability. Experimental results have shown that for user testing, only a few participants are needed to identify problems and even fewer participants are needed to identify severe problems (Virzi, 1992) who conducted three studies with 12, 20 and 20 particcpants. Nielsen and Molich (1990), conducted a study using 34 participants and Nielsen (1992) used 3 groups of 31, 19 and 14 participants. Neilsen (1993), suggests that at least 10 subjects are needed for performance measurement in full tests, but that less are acceptable for pilot tests.

According to Gray and Salzman (1998), an interest in the design of interfaces has been a persistent HCI topic but there has not been the same interest shown in the design of experiments. Small problems in how an experiment is designed and conducted have been shown to have large effects on what could be legitimately concluded from its outcomes (Gray and Salzman, 1998).

An investigation into the validity of five influential UEM studies was undertaken by Gray and Salzman (1998). Jefferies et al (1991), compared four UEMS that they called Heruistic Evaluation, Cognitive Walkthrough, guidelines and user testing to assess the usability of HP-VUE, a visual interface to the UNIX operating system. There were 4 subjects in the Heuristic Evaluation group; 1 team of 3 subjects in the guidelines group, 1 team of 3 subjects in the Cognitive Walkthrough group and 6 regular PC users who were unfamiliar with UNIX in the user testing group. The user testing was conducted by a human factors expert, practiced in user testing. According to Gray and Salzman (1998), the UEM undertaken by Jefferies et al (1991), was presented as an experimental comparison of four UEMS and several misleading conclusions were drawn resulting through low power and lack of stastics as well as uncontrolled differences among groups. According to Gray and Salzman (1998), there were also construct validity issues with how problems were counted and conclusions regarding the strengths of different types of evaluators went beyond the scope of the study. Gray and Salzman (1998), conclude that overall, the design and scope of the study did not support the inferences made regarding cause and effect or generality of the results.

A comparison of user testing was undertaken using a walkthrough technique that combined scenarios with guidelines, (Karat et al (1992). 48 subjects were drawn from a subject pool and were assigned randomly to 3 conditions: user testing (2 groups of 6 subjects), individual walkthrough (2 groups of 6 subjects) and team walkthrough (2 groups with 6 teams of 2 individuals per team). One group in each condition evaluated one integrated office system whereas the second group evaluated a second integrated office system. During a three hour session, subjects used the technique to which they had been assigned to elarn about the system, freely to explore the system, work through nine scenarious, and complete a questionnaire. Gray and Salzman (1998), state that in a this UEM study undertaken by Karat et al (1992), there is a failing in the area of statistical conclusion validity with few statistical tests being reported. Gray and Salzman (1998) consider that several of the findings of this study should be considered with caution.

Nielsen (1992), undertook a study in which he examined whether the probability of finding usability problems increased with usability expertise as well as domain (voice response systems) expertise and also a study where he classified outcomes for six Heuristic Evaluations of different user interfaces along several dimensions. In the expertise study, Nielsen had 3 groups complete a Heuristic Evaluation of a printed dialogue. Groups consisted of 31 computer science students (novices), 19 single experts and 14 double experts who had expertise in user interface issues and voice response systems. Neilsen interpreted the results of his study to provide advice regarding how expertise effects the types of problems found by Heuristic Evaluation. The assumptions made by Neilsen (1992), about the effect construct as well as the approach used to analyze the data severely weakened the validity of the conclusions of his UEM study (Gray and Salzman, 1998). Neilsen (1992), made conclusions that assumed that named problems were actual problems and that differences found by opportunistic comparisons and tested by only eyeball statistics, were real (Gray and Salzman, 1998).

Desuvire et al (1992), compared the effectiveness of three types of evaluators including experts and non-experts and compared all UEM conditions with user testing. Desurvire et al (1992), failed to recognise the limitations of their study and based many strongly worded conclusions upon scant data (Gray and Salzman, 1998). The pre-requisites for experimental study, statistical conclusion validity and internal validity, were severly lacking and due to this Gray and Salzman (1998), consider that there is nothing that can be safely concluded based upon this study.

Nielsen and Phillips (1993), failed to acknowledge the limits that their highly variable results placed on statistical conclusion validity and, subsequently, the results could be misleading (Gray and Salzman, 1998)

There are three main approaches to evaluating software: usability testing; field studies and analytical evaluation (Preece et al, 2007). Each approach has several methods associated with it which are: observing users; asking users; asking experts; user testing; inspections and modelling users' performance. Some approaches use a combination of methods, for example usability testing and field studies both involve observing users and asking users, but the conditions under which they are used, are different (Preece et al, 2007).

The defining characteristic of usability testing is that the test environment and the format of the testing is controlled by the evaluator (Koyani et al, 2003). Quantifying users' performance is a dominant theme in usability testing with tests, typically, taking place in a laboratory or laboratory-like conditions where the user is isolated from day-to-day interruptions (Preece et al, 2007).

Preece et al (2007), state that well-planned evaluations are driven by goals which aim to seek answers to clear questions, which in the case of usability testing, are stated explicitly before the tests. Preece et al (2007), propose the DECIDE framework which has six items which are dealt with iteratively:

1) Determine the goals.

2) Explore the questions.

3) Choose the evaluation approach and methods.

4) Identify the practical issues.

5) Decide how to deal with the ethical isues.

6) Evaluate, analyse, interpret and present the data.

2.15 International Usability Evaluation

Usability evaluation aims to answer the question of whether a system is good enough to satisfy all the users' needs and requirements (Nielsen, 1993). According to Neilsen (2000), the ultimate usability evaluation method is the International User Testing where, theoretically, usability tests are conducted in special usability laboratories that are equipped with cameras to record the user's comments and facial expressions. Nielsen (2000), suggests that international user testing could be administered remotely, where the user runs a self-administered test on the computer through the web, while the experimenters observe over the Internet. Neilsen (2000), also considers the use of a portable usability laboratory taken to the user's sites to be an appropriate method of conducting international usability evaluation.

Hariandja and Daams (2005), state that cross-cultural usability testing is exceptionally useful when designing products for users from different cultures. Evers (2002), suggests that some user evaluation methods are less applicable than others for a culturally diverse user base, claiming that observation methods, for example, might not be appropriate for users from some cultures. Smith and Dunckley (1998), criticise usability evaluation methods as being subject to cultural bias and practical difficulties and highlight the drawback of some of the international usability methods such as the difficulty of multicultural heuristics, as well as the costly and logistically difficult recruitment of multicultural representative users for testing in laboratories. Smith and Dunckley (1998), propose a user interface design approach based on Kreitzberg's (1996) Logical User-Centred Interaction Design (LUCID) methodology.

LUCID methodology is a management strategy that provides a structure for scheduling needed steps for interface design approaches (Kreitzberg, 1996). LUCID is designed to promote an orderly process, with iterations within a stage and predictable progress among stages; each progress is tied to specified deliverables and timely feedback (Shneiderman, 1998). Smith and Dunckley (1998), employed the LUCID approach to distinguish between the interface design factor that influences the usability across all user groups and those design factors, which are sensitive to cultural variables, so that localised and globalized versions can be developed.

Cultural comparisons are based not on individuals but rather on central tendencies from each country compared together. Instead of dealing with the whole population, they suggested testing interfaces with a few typical users who will represent the characteristics of the population (Smith and Dunckley, 1998).

Smith and Dunckley (1998), identify sub-groups of users on whom to focus the evaluation, based on objective factors for example gender or age, and subjective factors, for example Hofstede's cultural variables. By their application of the LUCID approach, Smith and Dunckley (1998), use rapid prototyping with user dichotomies and user arrays used on subjective factors.

According to them, this allows testing a fraction of the user population and still obtaining a significant amount of usability information.

According to Dray et al (2003), technology can knit the world together or it can create a schism. Abratt et al (2003), state that this is a very pertinent statement, particularly for developing countries, as IT can have a positive or negative contribution to economic development. On the negative side, the proliferation of technology in developing countries has been blamed for widening the gap between the rich and the poor, taking the form of a digital divide between those that do and those that do not have access to technology. On the positive side, however, developing countries can take advantage of IT to achieve economic gains (Abratt et al, 2003). In countries such as South Korea, Taiwan and Singapore, investments in technology have allowed these countries to estable themselves as technologically self-sufficient with these countries now exporting home-grown technology to other nations (Abratt et al, 2003).

Abratt et al (2003), state that HCI is one of the most important disciplines in making sure that IT is used in a positive manner within developing countries by designing systems to suit individuals' economic, political, social and cultural identity, technology can be a powerful catalyst towards national economic improvement.

Usability evaluations have been undertaken in Namibia, (Winschies, 2007). Winschies (2007), considers that in participatory design, the real problem, as well as the system requirements, are best determined through merging the different viewpoints of stakeholders involved. In a multicultural system design setting, the forming and interpretation of viewpoints, as a cross-cultural judgement, is problematic in that it depends on the stakeholders' perception of the environment (Winschies, 2007). However, the perception of reality and the structuring and processing of experience depends on our habits that are shaped by our culture. A non-native computer expert understands and models the environment as he perceives it through his personal culture bound perspective which rarely coincides with the view of the local users (Winschies, 2007). This obviously has a major impact on system design which in this case is based on a misconception of the initial situation, thus leading to the implementation of an undesirable system (Winschies, 2007). Furthermore, diverse disciplines advocate a diological approach to resolve or minimise those misperceptions within co-operative tasks. Yet, cross-cultural dialogue is predetermined for misunderstandings due to distinct cultural determinants (Winschies, 2007). Although the importance of communication in system design has generally been recognised and multiple methods have been put forth to facilitate communication between user and system engineer, their validity has not been studied in a multicultural context as yet (Winschies, 2007). Therfore, it is up to the system engineer to determine the stakeholders' intention and communication competency and to accordingly develop and apply culturally valid communication techniques to conduct a successful dialog.

Winschies (2007), concludes that IT, as well as the development methods and techniques cannot simply be transferred across continents and cultures but have to be determined within each development context jointly with the users. Therefore, partipatory design in a cross-cultural context goes beyond the involvement of users in the design of the product but should include an appropriation of the design process itself (Winschies, 2007).

2.16 The Use of Cultural Models in Human Computer Interaction Evaluation

According to Gillham (2004), there have been several attempts to use the Hofstede model for culturally appropriate interface development. Marcus et al (2003), have consistently advocated the link between Hofstede's cultural dimensions and characteristic factors of user interfaces and have examined a number of these dimensions and their possible impact on user-interface design. Marcus and Gould (2000), found that features of a website design such as the degree of flexibility allowed in completing a task will often reflect the Hofstede's country ranking scores for Uncertainty Avoidance and propose their guidelines to inform future designs so that they are more culturally appropriate. As an example, the design recommendations for high Power Distance cultures include: access to information should be highly structured; tall hierarchies; a strong emphasis on social and moral order and the use of business and social roles to organise and restrict information (Marcus and Gould, 2000). For cultures scoring high Individualism ranking, Marcus and Gould (2000), design guidelines suggest catering for motivation based on personal achievement; including images of success that stress materialism and consumerism; have written language that is rhetorical and tolerates extreme claims; gives prominence to youthful ideas and images; emphasises change and places the individual before the group.

Ford and Gelderblom (2003), tested the design model suggested by Marcus and Gould (2000), and having found inconclusive results Ford and Kotze (2005), sought a different way to understand culture's influence. The model that they propose identifies five general categories of variables that can influence usability and could have confounded their 2003 results (Ford and Kotze, 2005).

The five categories in the model suggested by Ford and Kotze (2005) are:

1) subjective culture;

2) the interface;

3) user acceptance;

4) speed of performance;

5) objective culture.

To account for and encompass these possible confounds, the conceptual model the authors propose for testing usability cross-culturally is broken up into three contexts: user characteristics, task characteristics and the environment. The user characteristics context is further broken up into three classes: cultural, physical characteristics and psychological characteristics. The psychological characteristics class includes several sub-classes which are investigated using the Unified Theory of Acceptance and Use of Technology (UTAT) in this work (Ford and Kotze, 2005). These characteristics include: perceived enjoyment; professional status; self-efficacy, ease of use; ease of understanding; computer anxiety and computer self-efficacy. Task characteristics are further broken down into: job category; risk; demands; linkages and task execution. The environment category consists of: organizational environment and physical enviornement. Each one of the classes under each category is further broken down into specific variables that would need to be controlled, isolated or accounted for in some way when undertaking cross-cultural usability research (Ford and Kotze, 2005).

The guidelines proposed by Marcus and Gould (2000), address front end design and the suggested model given by Ford and Kotze (2005), provides an empirical model for cross-cultural usability evaluation. Kersten et al (2002), suggest a conceptual framework for addressing back end design to make systems more culturally appropriate. Kersten et al (2002), argue that culture influences our core beliefs and behaviours and influences the way we approach business practice and the authors suggest that taking a more culturally aware approach to software development can be achieved by firstly determining which aspects of the software are culturally dependent and then designing these separately to be appropriate for each culture while maintaiing a core set of 'libraries' that are used by all instances of the software (Kerseten et al, 2002).

An alternative design approach is to take a 'cultural fingerprint' of a culture and website (Smith et al, 2006). The fingerprint is a way of using Hofstede's scores for each country to map out where a particular country sits in the four dimensional space of Individualism, Masculinity, Power Distance and Uncertainty Avoidance (Smith et al, 2006). This fingerprint can then be compared to the 'cultural fingerprint' of existing sites designed for this culture. The site's 'cultural fingerprint' would be calculated by a team of expert evaluators and by undertaking a comparison of the two fingerprints, a design team can tell where the misfit between the target culture and the site design lies (Smith et al, 2006).

Whilst there is no empirical evidence to show that cultural models can be directly used to inform design (Gillham, 2004), these models may be of some use in profiling cultures and their members. Gillham (2004), states further that cultural models can provide a subjective measurement of the gulf between the culture of origin and the localisation target culture. Hoft (1996), proposes that cultural models can be used to address some design issues, such as the extent to which localisation of a product is needed.

It appears from the above research, therefore, that there is no general consensus concerning a reliable approach for the design of interfaces for cross-cultural use.

2.17 Previous Research in Usability and Evaluation of Speech Technology

Due to the uniqueness of the field of this research, there is no literature available on such previous work incorporating TTS technology, accent, culture and SEN. However, the work of the following is relevant to identify methods that have been used previously to establish any weaknesses in testing methods to ensure that they are not repeated.

2.17.1 Evaluation of An In-Car Speech System for Older Users

As far as the testing of speech as an output is concerned, one of the main areas of research has been conducted by Zajicek, in particular the testing of speech as an output from a web browser tool (BrookesTalk) and also from in-car speech systems for older adults. Zajicek and Jonsson (2006), consider that speech systems have been shown to help older adults accomplish tasks. They can also provide useful information about the environment and things happening around them, which older people may not readily absorb for themselves. In the domain of computing, voice prompt messages have been successfully used to provide reminders concerning previous interaction for those with poor memories and enabled older people to use a computer system, which they had not done before (Zajicek and Hall, 2000). Zajicek and Jonnsson (2006), undertook laboratory based usability tests on an in-car speech messaging system designed for older users. The aim of the experiments was to identify characteristics of the voices used by the system that have the potential to influence driver's attitude to the voice and thus driving performance. The perception of the voices as well as the perception of the person's speaking was examined, in an attempt to identify characteristics of the voices that would explain the impact on driving performance for older adults which had been found by Jonsson et al (2005), where older adults showed a definite preference for information delivered by a younger voice. The younger voice used was that of a 20 year old woman and the older voice was that of a 73 year old woman. Although there will not be any distinguishment between the ages of the synthesised speech for these usability tests, there will be a distinguishment between different cultures and therefore the methods used by Zajicek and Jonsson here may be appropriate to consider.

In the study undertaken by Jonsson et al (2005), there were 12 subjects, 6 who were 55+ and 6 who were 18-25 years of age. When undertaking usability tests using vulnerable subjects, such as older people or those with learning disabilitites, it is appropriate to have smaller sized groups of subjects due to the individualness of the testing procedure and associated ethical issues. This thesis also undertakes tests on small groups of subjects for the above reasons. The subjects who took part in the experiments gave informed consent, which will also be obtained prior to the testing procedure and volunteered their time; this research will also be using volunteers only.

The subjects listened to the recordings of the two voices which they then rated using 4 questionnaires (Zajicek and Jonsson, 2006). Zajicek and Jonsson (2006), provided a standard questionnaire to identify measures of similarity. Subjects were asked to rate the statements based on a 10 point scale of contrasting statements. Although questionnaires will be used in this thesis, the design of this will not incorporate a 10 point scale due to the learning disabilities of the end users. The subjects were randomly divided into two groups, one group that listened to and rated the young voice first and one group that listened to and rated the older voice first (Zajicek and Jonsson, 2006). The randomness of using different voices in the testing procedure will also be a key consideration for this thesis in the conduct of usability tests to ensure that the subjects do not have a chance to pre-empt the voice that they are about to hear.

Once the questionnaires had been completed, the results were measured by using ANOVAs; this is a method that will also be adopted to show the comparisons of results, between participant factors and to present the findings graphically.

Although the testing referred to by Zajicek and Jonsson (2006), refers to an in-car speech system, there are several key areas of the testing procedure that has been identified as good practice which will be incorporated into the tests in this thesis.

2.17.2 Evaluation of Non-Speech Auditory Prompts

In addition to the use of speech as an output, research is also being undertaken into non-speech auditory prompts. Brewster (2008), has undertaken research into Earcons which he describes to be abstract, musical tones that can be used in structured combinations to create auditory messages, which are based on musical sounds. Several experiments have been undertaken to investigate the use of Earcons for representing hierarchical information which has included applications for Telephone Based Interfaces (TBI) and for blind computer users. Brewster (2008), considers that TBI are becoming an increasingly important method for interacting with computer systems and goes on to state that the telephone is an ubiquitous device and is many people's primary method of entry into the information infrastructure. Access to an increasing number of services is being offered over the telephone and the rapidly increasing use of mobile telephones means that people access these services at many different times and places. However Brewster (2008), continues to state that the provision of this functionality may be rendered useless if usability issues are not considered and has subsequently undertaken research into an evaluation of TBI.

Brewster (2008), considers that there are two main approaches to overcome navigation problems. The first one being to use speech recognition and TTS where the user would speak a command the system would reply in speech. This would avoid many navigation problems but users would have to remember all of the possible commands that they can say. Williams et al (1998) claim that the length of spoken menu prompts is a significant problem with speech based menu designs. With this in mind, Brewster (2008), used non-speech sounds to provide navigation cues to enhance the existing interfaces. Three main experiments were carried out and the results of each fed into better designed sounds and also into the design principles for TBI (Brewster, 2008). The first two experiments allowed Brewster to improve the design of the sounds beyond those done before the project started. The third experiment forced an extension of ideas to work with a real mobile phone and its menu structure. The experiments also led to the introduction of new interaction techniques: one to control sound volume to avoid annoyance and the other to present continuous, time-varying data in sound. Finally, a tool was produced to help interface designers quickly and efficiently use the knowledge of earcon design that had been gained from the experiments (Brewster, 2008).

The tests were based on users performing a range of typical mobile phone tasks including setting profiles and call diverts on a phone that they had not used before (Brewster, 2008). There were 24 subjects which had been selected from the University of Glasgow, which were split into two groups of 12. All of the subjects had no previous experience of using a Nokia phone that was being used in the experiments (Brewster, 2008). In addition, an endeavour was made to recruit subjects who had no experience of using mobile phones at all, there were ten from this category in each group with the remaining two in each group having a limited experience of using a mobile phone menu. All subjects were computer literate and familiar with navigation in hierarchical menus. One group used a phone sonified with Earcons and another used a standard one which just played a tone when keys were pressed (Brewster, 2008). The hypothesis being tested was that the navigation sounds should allow users to build a better mental model of the menu structure of the phone and which would translate into faster performance and fewer errors (Brewster, 2008).

The results showed that the number of key presses needed to perform tasks started at a similar level in the two groups but by the end of the experiment the sonified group took an average of 28% fewer. Similary, the number of tasks completed was initially similar but by the end, significantly more were completed correctly with the sonified phone. Subjective workload ratings showed no differences between the sonified phone and the standard one.

Contact was made with Brewster to discuss the evaluation methods that he had used with Earcons, and it was apparent that the evaluation techniques used for Earcons would be inappropriate for TTS. However, notwithstanding this, the method that was used for conducting the experiments is interesting as, again, due to the personal nature of these tests, a small sample size was used.

2.18 Influence of Literature Review on Relationships

This literature review has explored previous research into the separate fields of speech, accent and culture. However, as this research is unique there is no definition on the relationships between them. Subsequently, for the purposes of this research the stance is taken that a person's speech and accent is representative of their culture and therefore, their speech and accent represents a deep phenomenon of affinity to that culture. It is also considered that although an individual may move to a country, or to a different part of the same country that has a different culture, they may adopt a certain acculturalisation of the culture in which they find themselves living, for example dress and consumer products, but however, these are merely superficial manifestations and do not affect their existing cultural affinity.

2.19 Influence of Literature Review on Research Methodology

The key findings from this extensive literature review that will need to be considered in the methodology of the evaluations are:

- A comparative study of subjects from different cultures is appropriate.

- The evaluation method used by Zajicek and Jonsson (2006), suggests that small groups of subjects, the use of questionnaires, quantitative analysis using ANOVAs, gaining informed consent prior to the experiments and the use of volunteers are appropriate.

- The evaluation method used by Brewster (2008), uses a TBI which is also used for the banana information system. The subjects used by Brewster had no previous experience of using the 'phone that was being tested. In the experimental design, subjects will also be used who are inexperienced in using the TBI.

Several issues have been identified from the evaluation methods used by Zajicek and Jonsson (2006) and Brewster (2008) and the table belows how these issues will be incorporated into the experimental designs for this research.

Table 2.1. Key Issues Identified by Previous Evaluation Methods

Experiment	Key Points from Zajicek and Jonsson (2006)	Key Points from Brewster (2008)
Generic Software Experiments (detailed in Chapter Four)	Small groups of subjects Quantitative analysis using ANOVAs Informed consent Volunteers	No previous experience
Experiments in Kenya (detailed in Chapter Five)	Small groups of subjects Quantitative analysis Informed consent Volunteers Questionnaires	No previous experience TBI
Prototype Modelling and Testing (detailed in Chapter Six)	Small groups of subjects Quantitative analysis using ANOVAs Informed consent Volunteers Questionnaires	No previous experience

The experiments are further explained in the following chapters.

Other relevant issues include the following:

- There is a legal obligation for learning providers to ensure that education is accessible and inclusive to all, but evidence so far suggests that this is not, necessarily, always the case.

- IT is used as a tool to facilitate access to learning, but evidence suggests that the interface of speech is an area that has been overlooked. Concatenative and formant synthesis may be appropriate for different types of applications.

- Speech is not just the way that someone talks, it is representative of their culture and their way of preferred learning, which may be different across cultures.

- The use of multimodal interfaces may be more memorable than speech alone.

- The educational system must prepare learners for a multicultural information age economy. However, evidence suggests that cultural differences are an issue in system design and that usability and evaluations can be difficult and need to be well planned.

2.20 Chapter Summary

End users with SEN have a moral and ethical right not to be discriminated against within education, regardless of their culture. The introduction of legislation goes some way in addressing this, but their needs can never be fully met until cultural effects are fully considered. Language and cultural relationships have been discussed and the review of the research of other academics enforces the view that the way people speak and the accent that they use is not just something that has been casually adopted but represents something much deeper – their culture and their way of learning. It follows, therefore, that particularly for people whose needs are special in some way, that their way of learning should be included in the design of ICT systems.

Definitions of culture have been explored in this chapter. The following classification of culture has been adopted for this thesis:

People of a culture belong to a certain group, they all share a distinctive way of life and their behaviour and attitude is dictated by their culture.

Following extensive research this classification includes the principles that because behaviour and attitude is dictated by the culture to which an individual belongs, their cultural accent is also dictated in this way.

Cultural variables were discussed, with the cultural model by Hofstede being adopted for use in the analysis of some of the test results, this will be detailed fully in Chapters Four and Seven of the thesis. An explanation of how human speech and synthetic speech are formed is also given, with the chapter ending by looking at previous research in usability and evaluation of speech technology. Zajicek and Jonsson have undertaken an evaluation on an in-car speech system for older users and Brewster has undertaken an evaluation of non-speech auditory prompts for the use of navigating hierarchical telephone menus; key points were identified from these testing methods.

The next chapter discusses the research methods that will be used for conducting three separate sets of experiments which involve human subjects, and how the results will be analysed and evaluated.

Chapter Three – Research Methodology

3.1 Chapter Introduction

In the previous chapter, previous evaluation methods that have been used were discussed and key points were identified that are important to consider in developing evaluation methods within this research.

Zajicek and Jonsson (2006), evaluated an in-car navigation system for older users that used TTS as an output. It is interesting to note that they had a sample for the tests of just 12 subjects. Although at first this may seem a very small sample, older users are from a vulnerable group of subjects and to undertake tests to meet ethical standards without causing distress to them, small numbers are appropriate because of the amount of time needed to give each subject the individual attention required. Zajicek and Jonsson (2006), also used a questionnaire response which will be incorporated into this thesis, but it will have to be carefully adapted to suit end users with SEN and, furthermore, incorporating the additional considerations of a reader and a translator. In Zajicek and Jonsson's (2006), testing the voices were used in a random order. This is an important feature that will also need to be included in this thesis to ensure that the subjects cannot pre-empt what the next voice will be. Zajicek and Jonsson (2006), used subjects from the target group of older users. In this thesis, three sets of experiments to test usability are going to be undertaken. The first set of experiments will include subjects who do not have SEN. The reason for this is to identify whether or not there is a case to prove, which will then necessitate further testing with end users who do have SEN. In addition, by using this approach, any problems with the testing methods can be overcome before subjecting more vulnerable subjects.. To analyse and evaluate the results, Zajicek and Jonsson (2006), used ANOVAs and they will also be used wherever possible in this thesis to show analysis of variance.

The evaluation undertaken by Brewster (2008), is not as relevant to the tests in this thesis as the work of Zajicek and Jonsson. However, again it is very interesting to note that the sample size was small, with just 24 subjects.

The testing undertaken by Zajicek and Jonsson (2006), and Brewster (2008), have some commonality. They both have a structured approach, using small sample sizes because they were observing human subjects, and endeavours were made to eliminate bias as much as possible. For example, Zajicek and Jonsson (2006), used subjects from the target group and Brewster (2008) used subjects without any previous experience of the phone that he was using for testing.

3.2 Credibility of Testing Methods

To have credible results, there needs to be credibility of testing methods. The credibility of research is based on the validity and reliability of the research findings in addition of the findings' capability of providing a basis for scientific generalisation (El Said, 2005). Krathwohl (1997), considers that there are three concepts as follows:

Table 3.1 Credibility Concepts

Source: Krathwohl (1997)

Credibility Concept	Related Questions
Validity	Does an instrument measure what it is supposed to measure?
Reliability	Will the measure yield the same results on different occasions (assuming no real change to what is being measured)?
Generalisaility	What is the probability that patterns observed in a sample will also be present in the wider population from which the sample is drawn?

Hirschheim (1992), states that one of the major advantages of using this post-positivist approach is its capability to support research validity and reliability through the methodological pluralism concept that emphasizes the use of multiple methods of measures; this approach is also referred to as triangulation. Janesick (2000), considers that there are five types of triangulation – data, investigator, theory, methodoligical and interdisciplinary.

The following table shows how triangulation applies to this research:

Table 3.2 Triangulation

Source: Janesick, (2000)

Triangulation Type	Meaning	Application to this research
Data	The use of a variety of data sources in a study	Data is collected through the following variety of different sources: literature survey; experiments in England on generic systems; experiments in Kenya on a banana information system; comparison and contrast of two existing teaching and learning tools; and modelling and testing of a teaching and learning tool.
Investigator	The use of several different researchers or evaluators	Several researchers are involved in the experiments conducted in Kenya in collaboration with the Universites of Nairobi and Bristol. Different evaluators are used from a wide range of cultures who have differing levels of previous experience with IT and speak different languages, both within England and Kenya.
Theory	The use of multiple theoretical perspectives to interpret a single set of data	This research uses cultural model theories of human behaviour and applies to the Hofstede model the results of the usability tests, where appropriate. Hypothesis testing is also undertaken on cultural differences in using TTS technology.
Methodoligcal	The use of multiple methods to study a single problem	The following multiple methods of data gathering are used in this research: usability testing for error rates using an adopted method of the Design Rhyme Test; the design of usability tests for banana farmers in Kenya incorporating English

		and Kiswahili; interviews with the developer of a new teaching and learning tool; interviews with a centre which uses an established award winning teaching and learning tool; and satisfaction questionnaires using the SUMI technique undertaken with usability tests on a teaching and learning tool prototype model developed by the author.

Krathwohl (1997), considers that another concern about research is the sampling technique used where research can build up on a subset of a population, which is used to represent the population under study. Krathwohl (1997), states further that statistics can be subsequently used to investigate the likelihood that a pattern observed in the population be a replication of the sample pattern, thus providing a basis for research generalisation and as research can never reach an entire population, all research is only ever based on samples.

It is considered that sampling techniques are divided into probability and non-probability techniques (Krathwohl, 1997). Probability sampling involves random sampling of units from the population at some stage in the sampling process which enables the researcher to make inferences about characteristics of the population. The probability sampling technique includes simple random, stratified, systematic and cluster sampling methods.

Krathwohl (1997), considers that all methods used under the probability technique creates a sample using a random process for selection of elements from the entire population, the non-probability sampling does not. Krathwohl (1997), states further that non-probability sampling methods are procedures that do not include random sampling at some stage in the process and because of their convenience, they are common.

El Said (2005), states that non-probability sampling techniques include judgemental and purposive quota, snowball, sequential and convenience sampling methods. El Said (2005), states further that judgemental and purposive sampling involves judgements by researchers of which characteristics of the target population should be included in the sample. In quota sampling, researchers establish quotas for characteristics of individuals, often on the basis of demographic data, to ensure that they are as distributed in the sample as they are in the population.

Snowball sampling is used to discover members of the population not otherwise easily identified, by starting with known members and asking for referrals to other knowledgeable individuals. Sequential sampling involves gathering data in successive waves until some criterion of adequacy is met. Convenience sampling enables the researcher to select a number of cases whose size depends mainly of subjects' availability and ease of data collection. Krathwohl (1997), considers that the convenience sampling method, also referred to as the grab method, is the most commonly used non-probability technique.

This research uses a non-probability convenience sampling technique. By using this method it was possible to select a sample based on subjects' availability and the ease of collecting the data.

3.3 Selecting an Appropriate Research Approach

Land (1992), considers that information systems are a multidisciplinary endeavour as contributions to its study come from a variety of disciplines meaning that there is no single framework which covers all of the domains of knowledge essential for its study. Furthermore, there is a wide range of philosophical assumptions made in the study of information systems regarding the underlying phenomena under investigation (Orlikowski and Baroudi, 1991).

According to Galliers (1992), it is highly unlikely that there is an information systems research approach with a universal applicability. Furthermore Galliers (1992), considers that researchers into this subject area are becoming increasingly aware of the limitations of the scientific approaches to their work within the constraints of the socio-technical nature.

In consideration of the above points, it is important that there is an awareness of the whole range of research paradigms, which are a set of beliefs that guide a researcher's action. Guba and Lincoln (1994), consider that a paradigm offers a way of categorising a body of complex worldviews that guide action which consists of three basic beliefs. Guba and Lincoln (1994), consider that the major paradigms that structure and organise social science research are positvism, post-positivism, critical theory and constructivism or interpretations as detailed as follows:

Philosophical Aspect of Ontology

This raises the basic questions about the nature of reality that can be known, (Guba and Lincoln, 1994).

Positivism	**Realist native, realism, real.** Reality exists independent of humans. It is unproblematically, apprehensible, measurable and operates according to fixed laws of a cause-effect form.
Post-Positivism	**Critical realism, real.** Reality is assumed to exist but only imperfectly. It is probabilistically apprehensible due to the imperfectability of human cognitive and intractable nature of phenomena.
Critical	**Historial realism, virtual.** Reality shaped by social, cultural and ethnic values historically crystallises over time. Humans are not confined to a particular state.

Constructivism/ Interpretivism

Realist

Reality is relative to observers, as there are many socially constructed realities that are not subject to any natural laws. Humans continuously construct and reconstruct their reality.

Philosophical Aspect of Epistemology

This considers how we know the world and what the relationship between the inquirer and the known is (Guba and Lincoln, 1994).

Positivism

Dualism/Objectivism

The observer and the observed object do not affect each other. The aim of the inquiry is to establish cause-effect relationships between the objects of its study. Empirical testing results are assumed to reflect true explanations of the object's aspects.

Post-positivism

Modified dualism/Objectivism

Preference is given to critical tradition and critical community. The aim of the inquiry is explanation and prediction of knowledge. Findings are probably true but always subject to falsifications.

Critical	**Transactional/Subjectivist** Preference of the long-term historical and ethnographic studies. The aim of the inquiry is critique and transformation of knowledge. Findings are value mediated.
Constructivism/ Interpretivism	**Transactional/Subjectivist** The observer and the object are interlocked. The aim of the inquiry is understanding and reconstruction of knowledge. Findings are therefore created by the observer's interpretation and by the investigation process.

Philosophical Aspect of Methodology

Positivism	**Experimental/Manipulative** The inquiry process seeks to verify hypotheses established as facts or laws and eliminate confounding factors so as to explain the phenomenon as it really is. The aim is to predict and control using empirical quantitative tests.
Post-positivism	**Modified Experimental/Manipulative** Hypotheses are initially assumed to be false; they are probable facts or laws. Inquiries are done in more

	natural settings and while collection of more situational data. Multiplism of several methods that may include qualitative techniques.
Critical	**Dialog/Dialectal** The inquiry involves a dialogue between the investigator and the participants to transform ignorance into more informed consciousness, initiating changes in the social relations and practices
Constructivism/ Interpretivism	**Heremeneutic/Dialectical** This inquiry involves a continuous argumentative that seeks to critique, analyse and re-analyse. The aims are to reach a joint construction of a phenomenon by those parties to it.

Denzin (2000), considers that positivist and interpretive paradigms represent opposing constellations of beliefs about how valid knowledge may be generated. Winfield (1990), considers that a major advantage of the positivist approach is that it is unprejudiced in the sense that one can attempt to replicate the findings in a different study or in a different context. Hirschheim (1992), conversely, considers that the postivist approach is a poor and misleading approach to conducting social science research as it assumes an objective external reality upon which inquiry can converge. However Winfield (1990), states that there is a strong argument that what can be discovered in the interpretivist approach are not generalisations to larger population but contextual findings.

Lincoln and Guba (2000), state that the post-positivist approach is positioned between positivism and interpretivsm and was introduced to fulfil a need for researchers to overcome challenges of postivism. Hirschheim (1992), states that many researchers give support to an appropriate post-positivst stance in research, arguing that the use of multiplism of several methods, emphasised by post-positivism is suitable for the nature of information systems studies.

3.4 Choice of Research Paradigm

The post-positivism research approach has been selected as being the most appropriate research paradigm for this study as the research is focusing, in the main part, on the effect of culture and accent on the usability of information systems that use text to speech synthesis. This method is particularly suitable for this thesis because the post-positivism inquiry seeks to verify cause and effects relationships between hypotheses, which are initially assumed to be false (Guba and Lincoln, 1994). Several null hypotheses will be tested during the course of this study to establish the effects of independent variables against the dependent variables of usability components.

Both epistemological and technical reasons exist for selecting this research approach due to the fact that an endeavour is being made to establish a cause and effect relationship between the phenomenon's constructs. In addition, there is a hybrid approach for the testing procedures used in the inquiry.

Post-positivism rejects the idea that individuals see the world perfectly as it really is and instead consider that the observer is biased and all observation can be affected (Hirscheim, 1992). Guba and Lincoln (1994), consider that methodology in post-positivism emphasised critical multiplism as a way of assessing hypotheses. Kuhn (1996), states that this approach matches with the goal of information systems research in studying reality putting in mind that there is a likelihood that one many never be able to attain that goal.

3.5 Research Philosophy

Post-positivism informs social theorising and empirical investigation and posits the existence of three levels of understanding for social phenomena (Giddens, 1984), who considers that a framework to study the phenomena under investigation is suggested to include the following methodological steps:

Subjective understanding – the author seeks to find a subjective understanding throughout an exploratory phase which explores the phenomenon and generates research constructs through elicitation of the perception of human subjects' understanding of the phenomenon.

Interpretivist understanding – the author seeks to find an interpretivist understanding based on the interpretation of the subjects' understanding which leads the author with knowledge gained from literature, to generate relationships between research constructs and to build the hypothetical model used in the research.

Positivist understanding – the author seeks to find a positivist understanding which will be targeted throughout a model testing phase where formal testing of the research hypotheses is conducted by applying empirical assessment for verifying or disconfirming the hypothetical model.

3.6 Ethical Issues

As the performance tests require the participation of human beings, it is necessary to consider certain ethical issues. Allen (1984), states that it is important to conduct tests with human subjects with deep respect for the users' emotions and well-being and that they cannot be subjected to any form of destructive testing. Although the risks of becoming physically harmed through usability testing are negligible, undertaking testing can be quite stressful for some subjects and they may feel pressure to perform and may also worry about being slow at learning the systems and suffer inferiority complexes if they struggle in the knowledge that all this is happening under observation. In particular, the most highly educated and intelligent users, are concerned about exhibiting ignorance during the testing procedure (Schrier, 1992).

In consideration of the above, therefore, it is important that the experimenter makes the users feel as comfortable as possible during and after the tests, making it clear that it is the system that is being tested as opposed to the users (Nielsen, 1993).

The Economic and Social Data Service (2005), states that it is a fundamental ethical principle of research to ensure that subjects give consent and that they enter into the research voluntarily, without coercion and in the knowledge that they can withdraw at any time, confident in the knowledge that there will be no adverse consequences of participating in the research, including the preservation of anonymity if they wish.

In addition to ethical principles, the protection of information is also a statutory requirement contained in the Data Protection Act 1988. However, although consent is crucial to data protection principles, the EU directive 95/4/EC/24/10/1995 defines the data subject's consent as: "Any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed". The Data Protection Act allows for the further use of anonymised data for research purposes only. Specific provision is also made under the Act for processing personal information which includes racial or ethnic origin, political opinions, religious or other beliefs, trade union membership, physical or mental health condition, sex life, criminal proceedings or convictions, (Economic and Social Data Service, 2005). For personal information to be considered fairly processed, at least one of several extra conditions must be met. These include: having the explicit consent of the individual or being required by law to process the information for employment purposes or needing to process the information in order to protect the vital interests of the individual or another person or dealing with the administration of justice or legal proceedings (Information Commissioner, 2005). Nielsen (1993), recommends certain main ethical considerations for testing with human subjects which are detailed below with annotations on how this will be applied to the usability testing procedures in this thesis.

Table 3.7 Ethical Considerations for Testing

Source: (Nielsen, 1993)

Ethical Consideration – Before the Test	Application to Actual Test
(i) Have everything ready before the user turns up.	(i) The test room, computers, software, tasks, response booklets and other materials will be ready in advance of the subjects' arrival at the computer lab.
(ii) Emphasise that it is the system being tested, not the user.	(ii) The subjects will be given this information in writing on their consent forms and this will also be re-iterated verbally before the test.
(iii) Acknowledge that the software is new and untested, and may have problems.	(iii) The subjects will be informed that the software, although it is not new and has been tested before, has not been tested for users from multi-cultural environments and that there may be inherent problems with the software and that, subsequently, the experimenter is testing to identify any faults so that they can be addressed in the modelling of a bespoke prototype.
(iv) Let users know that they can stop at any time.	(iv) The subjects will be given this information in writing on their consent forms and this will also be re-iterated verbally before the test.

(v) Explain any recording, keystroke logging, or other monitoring that is used.	(v) No audio or visual recording or keystroke logging will be used. Subjects will be informed in writing on their consent forms that the experimenter will be monitoring the performance of the system throughout the test.
(vi) Tell the user that the test results will be kept completely confidential.	(vi) It will be clarified to the subjects that all of the results will remain anonymous, used for the purposes of academic research only and will not be divulged to any employer or teacher. Subjects will be identified only be encoding for administration purposes by the experimenter as participant 1, participant 2, for example.
(vii) Make sure that you have answered all of the user's questions before proceeding.	(vii) Subjects will be given an opportunity to ask questions about the test and testing procedure before the start of the test.
Ethical Consideration – During the Test	**Application to Actual Test**
(i) Try to give the user an early success experience.	(i) The experimenter will provide word sets with the lowest potential user based error rates first.
(ii) Hand out the test tasks one at a time.	(ii) The experimenter will hand out the tasks for test one and when completed, then hand out the tasks for test two.
(iii) Keep a relaxed atmosphere in the test room, serve coffee and/or have breaks.	(iii) Due to Health and Safety regulations in force at Thames Valley University where the testing will take place, and the timing aspect of the testing, refreshments will be made available after the completion of the test, outside of the computer lab.
(iv) Avoid disruptions: close the door and post a sign on it. Disable telephone.	(iv) A no entry sign will be placed on the door of the computer laboratory where the test is going to be conducted to avoid distractions and disruptions. All subjects will be asked to turn off their mobile telephones. There is no main telephone.
(v) Never indicate in any	(v) The author will not interact with the

way that the user is making mistakes or is too slow.	subjects in any way during the test, except in the circumstances where the participant is distressed and wishes to stop the test or is experiencing hardware problems with the computer.
(vi) Minimise the number of observers at the test.	(vi) The experimenter will be the only observer at the test.
(vii) Do not allow the user's management to observe the test.	(vii) The subjects' employers or teachers will not be involved in the conduct of the tests in any respect. Information about student performance will not be made available to their teacher.
(viii) If necessary, have the experimenter stop the test if it becomes too unpleasant.	(viii) The experimenter will stop the test at any time if a) a participant requests this or b) the author considers that a participant is displaying signs of distress through the testing procedure.
Ethical Considerations – After the Test	**Application to Actual Test**
(i) End by stating that the user has helped you find areas of improvement.	(i) The experimenter will thank the subjects for their invaluable assistance in testing the usability of the software and that the results will be used to model a prototype that will address any weaknesses that may be identified as a result of the test.
(ii) Never report results in such a way that individual users can be identified.	(ii) The experimenter will identify subjects only by methods of encoding – participant 1, participant 2, for example – for administration purposes only.
(iii) Only show videotapes outside the usability group with the user's permission.	(iii) The experimenter will not video record any aspect of the testing procedure.

The ethical considerations detailed in the table above applies, in particular, to the tests that will be undertaken on generic software in the computer laboratory. More specific ethical considerations regarding testing in Kenya, are fully documented in Chapter Five.

Before the undertaking of full testing it was necessary to seek approval to conduct the tests from the University's Ethics Committee because of the sensitivity and nature of the tests with human subjects. This entailed the submission of a Testing Proposal and accompanying 40 page interim report to a panel of ten senior academics experienced in ethics involved in research; this was a very important procedure to ensure that there was consensus amongst researchers that the tests would not have the potential to cause distress to any of the subjects.

3.7 Research Involving Students

Students have been participating in research for many years and inviting students to participate as action researchers in a project can give them a voice in constructing their literacy experiences and accounting for their own practices, (Erickson, 1986). In America, students have been paid as co-researchers to interview classmates and peers, collect literacy artefacts and reading transcripts to identify themes. They are also involved in data reduction and analysis (Glaser and Strauss, 1967). Student researchers and teachers share with one another accounts of their worlds and work accounts that may become the basis for further inquiry as well as correctives to the hidden assumptions that each may bring about the other (Comstock, 1982). Research is not done "on" or "about" students and teachers, but rather it is done "with" these groups and their interests in mind, and "for" the possibilities that open up such work.

3.8 Statistics Used in Reseach

The data from the research will be analysed and evaluated using several different techniques. These techniques will include the use of a spreadsheet, Excel, to perform formulae and functions to find totals and percentages. The use of a statistical program, SPSS, will also be used for hypotheses testing to find significance using a variety of techniques including both an analysis of variance (ANOVA) for correlated scores and repeated measures and factorial design.

In statistics, results are called significant if they are unlikely to have occurred by chance. However, a statistically significant difference means that there is evidence of a statistical difference, but it does not necessarily mean that the difference is large. In traditional statistical hypothesis testing, the significance level of a test is the maximum probability, assuming the null hypothesis, that the statistic would be observed. The significance of a result is called its p-value; the smaller the p-value, the more significant the result.

Significance is represented by the alpha symbol α This research uses the level of significance of 5%. Therefore, if the test of significance gives a p-value equal to or lower than 0.05 which is the α Level, the null hypothesis is rejected. These results are referred to as being statistically significant. Different α. levels have different different advantages and disadvantages. A very small α. level set at 0.01, is less likely to be more extreme than the critical value and so is more significant than higher level values at 0.05.

However, smaller α levels run greater risks of failing to reject a false null hypothesis and so have less statistical power. The selection of an α level therefore, involves a compromise between signficance and power.

3.9 Pilot Testing Strategy

In addition to the full consideration of ethical issues, it was also vital that the actual testing procedures were piloted first to eliminate and reduce any errors, inconsistencies and weaknesses that may be encountered with the proposed testing strategy and testing plan. These errors may include instructions that are incomprehensible for the users; misinterpretation of the instructions; clarity and usability of questionnaires; time management to complete the tasks and appropriateness of the ease of the test, (Nielsen, 1993). Pilot testing, therefore by its very nature, is on a much lesser scale and can involve far fewer subjects. Nielsen (1993), considers that test users should still be as representative as possible of the intended users of the system for pilot testing and for comprehensive further testing, users from several different subpopulations should be included to ensure all main different categories of expected users are covered.

3.10 Hypothesis Testing

Following contemplation of the most appropriate approach to undertake the usability tests and came to the conclusion that there would not be one test that would be suitable to measure everything due to the complexity of the study, but, moreover, that several different tests would be required.

Although usability consists of five main components – errors, efficiency, memorability, learnability and satisfaction – this study gives particular emphasis on the error rates with the rationale that if the error rates are found to be unacceptably high then it may be pointless testing the other components of the system, as very high error rates are considered to be catastrophic and may render the system unusable. For a system to be classed as at least satisfactory, the error rate cannot be higher than 10% overall (Papamichalis, 1997).

A null hypothesis is a hypothesis that is set up to be nullified or refuted in order to support an alternative hypothesis. When used, the null hypothesis is presumed to be true until statistical evidence in the form of a hypothesis test indicates otherwise. The null hypothesis is used to test differences in treatment and control groups with the assumption at the outset of the experiment is that no difference exists between the two groups for the variable being compared with the null hypothesis only being rejected when it becomes evidently false when the researcher has a certain degree of confidence, in this research 95%, that the data does not support the null hypothesis.

The hypotheses that are being tested in this research are detailed in section 1.8 of this thesis and also in chapters four, five and six.

3.11 Chapter Summary

Moral and ethical issues have been discussed in this chapter with details of how ethical issues will be strictly adhered to throughout the research to minimise harm to subjects. Different research methods and testing strategies have been considered together with different research paradigms and a justification for the choice of research methods applied in this thesis.

Each set of experiments are fully documented in chapters four, five and six.

Chapter Four – Generic Software Experiments

4.1 Chapter Introduction

This chapter documents the first sets of experiments that were undertaken. These experiments use existing TTS synthesis software which is for generic use, not necessarily as a teaching and learning tool. The speech is available in the English language in American, English and Indian accents. The speech is used in an apparent random order with the tests to find out whether the usability of the software, in terms of error rates, is improved when end users hear speech that sounds as though it is in the same accent as their own.

However, before undertaking the tests, it was necessary to undertake pilot testing to identify any potential problems with the testing procedure. These pilot tests involve twelve subjects, male and female, from American, English, Chinese, Indian and Pakistani cultures. The results of the pilot tests are used to establish whether there may be a phenomenon that needs further investigation.

Following the results of the pilot tests, the chapter then goes on to detail the results of the full testing procedure where forty four subjects from English and Indian cultures were used in the experiments. These results are analysed and evaluated using tables, ANOVAs and graphical representations. All experiments took place under controlled test conditions in a computer laboratory at Thames Valley University using subjects from the student body, all first year computer science students with similar levels of computer literacy but no previous experience of TTS technology.

The purpose of the experiments are to test hypothesis a) *There is no difference in error rate when the accent of the speech is representative of the culture of the subjects* and hypothesis b) *The gender of the subjects makes no difference to the error rate.*

4.2 Subjective Measures

This software is going to be tested for intelligibility and whether or not cultural, or gender issues, have any affect on the intelligibility in terms of error rates, ie is the speech clearly understood. Subjective testing of speech is being used for these experiments and can be categorised as follows: those testing intelligibility and those testing quality (Papamichalis, 1987). The two categories, however, are not disjointed and therefore good quality implies good intelligibility while the converse is not necessarily true ie good intelligibility does not necessarily mean that the quality is also good. Papamichalis (1987), suggests the Diagnostic Rhyme Test (DRT) which is a system developed by the software developer Dynastat to test intelligibility.

Viswanathan and Viswanathan (2005), also refer to the testing methodologies suggested by Papamichalis and state that TTS systems can be effectively assessed only on the basis of reliable and valid listening tests to assess overall performance. The testing procedures used by Papamichalis (1987), and Viswanathan and Viswanathan (2005), are based on testing procedures also recommended by the telephone transmission subjective tests.

Intelligibility tests, using the DRT method, are based on the ability of listeners to distinguish phonemes with common attributes. The listener is presented with one word from a pair using synthesised speech and is asked to determine which word was spoken from two possible choices. An important note here is that although the DRT is referred to as "words" for ease of reference, they are merely phonemes with common attributes that may not necessarily be a proper word. The alternative choice is a word different from the presented word only in one phoneme, usually a consonant. DRT is an intelligibility test, created by Dynastat who develop speech synthesis systems and measure something that is well defined and their results only have a small standard deviation which means that they are accurate and are repeatable (Papamichalis, 1987). The DRT works uses a corpus of 192 words in 96 rhyming pairs as shown in the following table which previous research has shown to be frequently confused by end users of TTS (Papamichalis, 1987). In a given instance, one word of the pair is presented and the listener is asked to determine which word was spoken (Papamichalis, 1987). The two words of each pair, for example "zeal" and "seal", differ only in one attribute of the first consonant so therefore, a correct response from the listener indicates that the speech under examination preserves that attribute, (Papamichalis, 1987).

Table 4.1 Word pairs used in the DRT. Source: Papamichalis (1987).

veal-feel	bean-peen	zoo-sue	dune-tune
meat-beat	need-deed	moot-boot	news-dues
vee-bee	sheet-cheat	foo-pooh	shoes-choose
zee-thee	cheep-keep	juice-goose	chew-choo
weed-reed	peak-teak	moon-noon	pool-tool
yield-wield	key-tea	coop-poop	you-rue
gin-chin	dint-tint	vole-foal	goat-coat
mitt-bit	nip-dip	moan-bone	note-dote
vill-bill	thick-tick	those-doze	though-dough
jilt-gilt	sing-thing	Joe-go	sole-thole
bid-did	fin-thin	bowl-dole	fore-thor
hit-fit	gill-dill	ghost-boast	show-so
zed-said	dense-tense	vault-fault	daunt-taunt
mend-bend	neck-deck	moss-boss	gnaw-daw
then-den	fence-pence	thong-tong	shaw-chaw
jest-guest	chair-care	jaws-gauze	saw-thaw
met-net	pent-tent	fought-thought	bong-dong
keg-peg	yen-wren	yawl-wall	caught-taught
vast-fast	gaff-calf	jock-chock	bond-pond
mad-bad	nab-dab	mom-bomb	knock-dock
than-dan	shad-chad	von-bon	vox-box
jab-gab	sank-thank	jot-got	chop-cop
bank-dank	fad-thad	wad-rod	pot-tot
gat-bat	shag-sag	hop-fop	got-dot

In addition to using this established testing method, however, the author has also designed her own testing procedures, as despite much research, no models were found for testing with multicultural end users to evaluate whether cultural issues affected the accuracy of the output of TTS systems.

The testing strategies suggested by Papamichalis (1987), also adapted by Viswanathan and Viswanathan (2005) who were prompted by the lack of clear evidence of the conceptual content of another testing procedure known as the Mean Opinion Scale (MOS). Viswanathan and Viswanathan (2005), state that they employ state of the art psychometric techniques such as confirmatory factor analysis to the MOS to provide strong tests of psychometric properties which they consider is better suited to appraise synthesis systems since it includes items that are specific to the artefacts found in synthesised speech. The techniques used by Viswanathan and Viswanathan (2005), will not be used in these experiments as they have been designed to test quality whereas intelligibility is being tested. However, it is interesting to note that other researchers have also, successfully, adapted the testing techniques promoted by Papamichalis.

For Test 1, the testing procedure will incorporate some of the words from the corpus in the context of a sentence, the results of which will be used to analyse and evaluate whether or not this has any effect on intelligibility. Synthesised speech from male and female voices of different cultural groups will be apparently randomly selected throughout 10 pairs of rhyming words taken from the 96 pairs in the corpus – only the experimenter will be aware of the correct response and the synthesised speech coding which is being used each time. One word of each pair will be presented to the subjects using the chosen accents and the subjects will be asked to mark on an answer sheet the word they thought was spoken from a possibility of two from the corpus plus one totally unrelated word which will be used as a distractor.

For Test 2, the testing procedure will again incorporate some of the words from the corpus but instead of being in a sentence, the word will be totally out of context, the results of which will be used to analyse and evaluate whether or not this has any effect on intelligibility. Synthesised speech from male and female voices of different cultural groups will be apparently randomly selected throughout 20 pairs of rhyming words taken from the 96 pairs in the corpus – only the experimenter will be aware of the correct response and the synthesised speech coding which is being used each time. One word of each pair will be presented to the subjects using the chosen accents and the subjects will be asked to mark on an answer sheet the word they thought was spoken from a possibility of two from the corpus. There is no distractor included in these tests as it was considered that this would not be necessary as there would be a reduction in possible guessing of the answers, as there may be a risk of in Test 1, as the words are not in any sort of context.

The following methods will be used to measure the results of the tests:

1) collate the answer sheets from the subjects which are identifiable by the participant identity number

2) abstract the gender and appropriate cultural background for each of the subjects and collate by culture

3) abstract and apply the answers to each of the six synthesised speech patterns

4) calculate the error rate by applying the following equation:

$$P = \frac{W}{T} 100$$

Where: P = percentage of incorrect responses

W = number of wrong answers

T = total number of items involved

5) evaluate the error rate of a) subjects who were from the same cultural background as the synthesised speech and b) subjects who were not from the same cultural background as the synthesised speech

6) abstract the gender and appropriate cultural background for each of the subjects and collate by gender

7) abstract and apply the answers to each of the six synthesised speech patterns

8) calculate the error rate by applying the following equation:

$$P = \frac{W}{T} 100$$

Where: P = percentage of incorrect responses

W = number of wrong answers

T = total number of items involved

9) evaluate the error rate of a)subjects who were the same gender as the synthesised speech and b) subjects who were not the same gender as the synthesised speech

Due to the manner in which the word pairs have been selected, all words are weighted equally (Papamichalis, 1987) therefore it will not be necessary to repeat each set of words (Test 2) or sentences (Test 1) several times in different accents, instead they will be heard once only. In addition, hearing the tests more than once may increase the risk of subjects tiring or pre-empting sentences and words. Also, the tests aim to reflect as closely as possible what would actually be heard if the system was being used outside of the testing environment.

The typical values of DRT range between 25% and 15% error rate and, for systems to be classified as "at least satisfactory" would have a DRT of 90% accuracy, which represents an error rate of 10%, (Papamichalis, 1987).

At this stage of the testing, the intelligibility of the output is being analysed and evaluated as opposed to other aspects of the usability of input to the system. The reason for this is because the error rate is the most important aspect as inaccurate interpretation of speech can be classed as a catastrophic error because it destroys the user's work, making it difficult to recover from.

4.3 Profile of Subjects for the Pilot Tests

There were twelve subjects who took part in the pilot testing undertaken from English, Chinese, Indian, American and Pakistani cultures. The purpose of the pilot tests, however, was to test the actual procedure and to establish any significance to continue with the main testing. It is accepted that there is an unequal distribution of culture and of gender (the gender imbalance being due to computer science students taking part in the trials which is male dominated) which would prevent a reliable analysis being undertaken, but the pilot test was essential to eliminate problems before the main tests took place.

There were 25% female and 75% male subjects. Although an unbalanced representation of gender was not the ideal as the results are harder to analyse and would have an element of inconclusion, these subjects all had similar levels of IT skills and experience and it was considered, therefore, that the results of using these subjects would give suggestions on whether or not there was any significance to continue with further tests.

Subjects were invited from first year Computing and IT students at Thames Valley University; subsequently, all of the subjects had previous experience of using IT, although none had prior experience of TTS synthesis.

English was the first language of 67% of the subjects; 17% Urdu, 8% Punjabi and 8% Cantonese. All of the subjects' first language was also the first language of their parents. This sample is representative of the diverse multicultural aspect of the University. 70% of the subjects had lived in the UK all of their life, 30% were born elsewhere. They were all aged between 18-20 years.

4.4 Implementation Plan for Pilot Testing

Before undertaking the tests the following procedure was followed which involved the issuing and completion of documents (as detailed in Chapter Three), which can all be found in the Appendix.

- Information sheet given to subjects.
- Written consent obtained from subjects.
- Questionnaire asking for personal information, including culture was completed by the subjects.
- Task 1 completed.
- Task 2 completed.

Stages 1 and 2 of the Implementation Plan for Pilot Testing

The subjects were given a participant information sheet and consent form following research into models used by the Economic and Social Data Service and the Faculty of Health and Social Science at Thames Valley University which was given to subjects at a date in advance of the conduct of the tests. These documents can be found in the Appendix.

Stage 3 of the Implementation Plan for Pilot Testing

After consent had been obtained, the subjects were given a questionnaire to complete before the start of the tests. This form asked for details of their age, how long they had lived in England for, where they had been born, their nationality, their cultural origin, their gender, their first language and the first language of their parents/guardians. This information is needed to analyse and evaluate the results of the tests. This questionnaire can be found in the Appendix.

Stage 4 of the Implementation Plan for Pilot Testing

A script was read to the subjects to ensure that everyone was fully informed and then the task sheet for Task 1 was given out.

The task sheet has clear instructions on the top informing the subjects that they will hear each sentence three times using different voices of synthetic speech. They are asked to indicate with a tick in the appropriate box whether they thought the correct word to answer the sentence was produced by speech A, speech B or speech C. The maximum time allowed is 10 minutes.

As stated earlier in this chapter, all of the words that were used came from the word pairs in the DRT, with one word outside of the DRT as a distractor, with the theory that this may reduce the risk of tempting guesswork between two responses, using synthesised speech in different accents as detailed later in this chapter in the pilot testing results. The script and the task sheet can be found in the Appendix.

This test asked the students to identify the correct word to complete a sentence from a choice of three possible answers, from English speech in different accents (American and English) and to write the correct answer on an answer sheet. The correct word would only be spoken once, with two incorrect words – one from the DRT using words that have previously been confused and the other word not from the DRT acting as a distractor. Various voices of male and female sounding English American and English British synthesised speech were used from AT & T Natural Voices 16Hz in an apparent randomised order. Synthesised speech in other English language accents was not available at the time of the pilot tests.

By using the words used in the DRT in this manner, it would be possible to compare results when a subject would know what the correct word would be, such as in this case when they are given a sentence to complete, with just hearing words independently without context which will occur in Test 2.

Sentences used in Task One

Question No	Question	Speech A	Speech B	Speech C
1	Vegetarians do not eat …. **(correct answer meat)**			
2	Wild animals can be kept in a … **(correct answer zoo)**			
3	……..is pleasant to drink **(correct answer juice)**			
4	Midday is at …. **(correct answer noon)**			
5	A young horse is called a …. **(correct answer foal)**			
6	They are going shopping to buy new ….. **(correct answer shoes)**			
7	The friends enjoyed watching the …… **(correct answer show)**			
8	Ducks swim in a **(correct answer pond)**			
9	Bread is made from ….. **(correct answer dough)**			
10	The fish were caught in a ……. **(correct answer net)**			

The cultural differences were analysed first as shown below.

Table 4.2 Summary of results of Pilot Test 1 by culture and accuracy of answer

Culture of Participant	Mean error rate for all answers overall for all speech outputs	Mean error rate for all answers when speech sounded as though it was from the same culture as the participant (if available)
Indian (first language Punjabi)	30%	N/A
Chinese (first language Cantonese)	30%	N/A
American (first language English)	10%	0%
Pakistani (first language Urdu)	40%	N/A
English (first language English)	16.25%	11.67%

The error rate for all answers overall mean average is 25.25%. However when speech sounded as though it came from the same culture, the average error rate was reduced to 11.67% for English culture subjects, and a reduction of 10% to 0% for the American culture.

These results suggest that the error rate of the output of TTS synthesis may be significantly reduced when end users hear sentences of speech that sounds as though it is from the same culture to their own. However, there is an element of inconclusion in the sample due to an imbalance between cultures in addition to accents not being available for all cultures.

After abstracting the cultural data, the results for gender were then analysed as shown in the table below.

Table 4.3 Summary of results of Pilot Test 1 by gender and accuracy of answer

Gender of Participant	Error rate for all answers when speech sounded as though it was from the same gender
Male	18%
Female	20%

The error rates for male subjects when they heard sentences that were from speech that sounded as though it was from male synthesised speech of any culture is 18%; the error rates for female subjects when they heard sentences that were from speech that sounded as though it was from female synthesised speech of any culture is 20%.

This suggests that it may not just be the cultural aspects of the sound of the speech but also the gender of the synthesised speech may also have an effect on the accuracy of the output. As the overall error rate for all answers overall is 25.25%, these results indicate that the error rate of the output of TTS synthesis may be reduced when end users hear sentences of speech that sounds as though it is from the same gender as themselves. However, again, these results have an element of inconclusion because of the gender imbalance.

Subsequently, the results of the subjects when the synthesised speech sounded as though it came from both the same culture and gender were analysed.. The results were collated as shown in the following table:

Table 4.4 Summary of results of Pilot Test 1 by gender, culture and accuracy of answer

Culture of Particpant	**Error rate for all answers when speech sounded as though it was from the same gender and culture MALE**	**Error rate for all answers when speech sounded as though it was from the same gender and culture FEMALE**
Indian (first language Punjabi)	N/A	N/A
Chinese (first language Cantonese)	N/A	N/A
American (first language English)	0%	N/A
Pakistani (first language Urdu)	N/A	N/A
English (first language English)	10%	0%

These results were extremely interesting and suggest that when end users hear sentences from synthesised speech that sounds as though it is from the same gender and same culture as themselves the error rate may be significantly reduced. The male American culture shows 0% error rate; the male English culture shows 10% error rate and the female English culture shows 0% error rate.

There is an element of inconclusion in these results because the speech was not available in all cultures; it was only available in American English and English English and the gender was also imbalanced. However, the results show a significant improvement when those subjects who were able to hear their accent because it was available, heard speech that sounded as though it was from the same culture, suggests that further testing is required to investigate this finding with a more balanced sample of subjects. This is detailed later in this chapter.

4.5 Pilot Test 2

Stage 5 of the Implementation Plan for Pilot Testing

After the subjects had completed Test 1, the script was read for Test 2 and the task sheet was handed out which can be found in the Appendix.

This time the subjects heard two words for each task - one word at a time, using different voices of synthetic speech with different accents. The subjects were asked to circle the word which they thought they heard. Unlike Test 1, these words are out of context so that the subjects would not be able to apply any guess work to what they thought the word was. Again, an apparent randomised order was used in male and female voices with American English and English English accents.

ID number of participant:

QUESTIONS FOR PARTICIPATION IN SPEECH SYNTHESIS RESEARCH

TASK TWO

Project Title: Text to Speech Technology: usability in multi-cultural environments as a teaching and learning tool for end users with special educational needs

Name of Researcher: Yvonne Spittles MSc MSBT MIVA

Institution: Thames Valley University

Contact Details: Yvonne.Spittles@tvu.ac.uk

Instructions for Task Two

You will hear one word at a time, using different voices of synthetic speech. Please indicate by circling the word that you thought you heard for each question.

Question No	Word A	Word B
1	veal	feel
2	vee	Bee
3	zee	thee
4	weed	reed
5	yield	wield
6	gin	chin
7	mitt	bit
8	vill	bill
9	jilt	gilt
10	bid	did
11	hit	fit
12	zed	said
13	mend	bend
14	then	den
15	jest	guest
16	keg	peg
17	vast	fast
18	mad	bad
19	than	dan
20	jab	gab

Test 1 asked the subjects to complete a sentence with words which were in context. Papamichalis (1987), suggests that the output of TTS synthesis is more accurate when only individual words are heard one at a time; with the theory that the perceived accuracy of the synthesised speech is reduced because of its lack of prosody and intonation. Subsequently the second test required subjects to indicate the correct output from a choice of two words.

This test asked the students to identify the correct word from a choice of two possible answers and to circle the correct answer on an answer sheet. Male and female American English and English English synthesised speech were used from AT & T Natural Voices 16Hz. All of the words came from the DRT using words that have been previously proven to be confused. There were twenty word pairs. The data regarding the culture and gender of the subjects and the culture of the speech output were abstracted and the results are shown on the following page.

<u>Table 4.5 Summary of results of Pilot Test 2 by culture and accuracy of answer</u>

Culture of Participant	**Mean error rate for all answers overall from all speech outputs**	**Mean error rate for all answers when speech was from the same culture as the participant (if available)**
Indian (first language Punjabi)	30%	N/A
Chinese (first language Cantonese)	35%	N/A
American (first language English)	10%	0%
Pakistani (first language Urdu)	25%	N/A
English (first language English)	13.5%	10%

Again, there is an element of inconclusion in these results because the speech was not available in all cultures but, however, the significant improvement for the subjects whose speech was available in different cultures (American and English) suggests that further testing is required to investigate this finding with a more balanced sample of subjects.

These results support the findings of Test 1 and further suggest that the error rate of the output of text to speech synthesis may be significantly reduced when end users hear words of speech that sounds as though it is from the same culture as their own.

The error rate for all answers overall average is 22.7%. However when speech sounded as though it came from the same culture, the average error rate was reduced to 10% for the English culture and to 0% for the American culture.

The results also support the theory that the output is more accurate when single words are heard with an overall error rate of 22.7% compared to 25.25% for sentences as suggested in Test 1.

After abstracting the cultural data, the results for gender were then analysed as shown in the table below.

Table 4.6. Summary of results of Test 2 by gender and accuracy of answer

Culture of Participant	Error rate for all answers when speech sounded as though it was from the same gender
Male	17.5%
Female	26.67%

The error rates for male subjects when they heard words that were from speech that sounded as though it was from male synthesised speech of any culture is 17.5%; the error rates for female subjects when they heard sentences that were from speech that sounded as though it was from female synthesised speech of any culture is 26.67%.

In contrast to the results for Test 1, these results did not suggest that there was any improvement in the use of gender alone in the sound of the synthesised speech as the error rate for females overall increased from 25.25% overall to 26.67%.

Subsequently, the results of the subjects when the synthesised speech sounded as though it came from both the same culture and gender were analysed.

The results were collated as shown in the following table:

Table 4.7. Summary of results of Test 2 by gender, culture and accuracy of answer

Culture of Participant	**Error rate for all answers when speech sounded as though it was from the same gender and culture MALE**	**Error rate for all answers when speech sounded as though it was from the same gender and culture FEMALE**
Indian (first language Punjabi)	N/A	N/A
Chinese (first language Cantonese)	N/A	N/A
American (first language English)	0%	N/A
Pakistani (first language Urdu)	N/A	N/A
English (first language English)	7%	15%

Again, these results were extremely interesting and suggest that when end users hear words from synthesised speech that sounds as though it is from the same gender and same culture as themselves the error rate may be significantly reduced. The male American culture shows 0% error rate; the male English culture shows 7% error rate and the female British culture shows 15% error rate. This is in comparison with the mean average error rate overall of 22.7%.

4.6 Summary of Results of Pilot Tests

Despite meticulous preparation, the pilot test identified a weakness with the testing procedure where the speech was played at an erratic speed resulting in the test having to be re-started. It was vital to rectify this to ensure that it woud not be repeated in the actual tests. Due to this weakness in the testing procedure and the element of inconclusion incorporated in the unbalanced sample, the findings of the pilot tests are not used for any purpose other than procedural testing and establishing the significance of continuing with the actual tests. These findings suggest that the accuracy of the output of TTS synthesis may be significantly increased when cultural and gender aspects are considered of the end user. The pilot test has served its purpose – it has identified a weakness in the testing system which will be rectified before the full tests are undertaken with a more balanced sample of gender and culture. The results of the pilot tests also suggest that there is a phenomenon to be investigated.

4.7 Actual Tests

Following on from the pilot tests, further testing was undertaken with forty four different subjects, which was a much larger sample than in the pilot test. This time however, subjects were taken from a balanced sampling frame to eliminate inconclusion in the results.

4.8 Profile of Subjects

The decision was made to undertake usability tests using subjects from English and Indian cultures. This decision was made for three reasons. Firstly convenience sampling, as previously discussed in Chapter Three as it was possible to attract potential subjects from an equal number of Indian and English cultures from the University. Secondly, the author wanted to undertake a comparison of contrasting cultural dimensions using Hofstede's models to establish whether or not the differences in cultural dimensions affected the usability of the software. Thirdly, in addition to English and American, Indian was the only other commercially available accent in the English language at the time of testing.

There were 11 male subjects who were from an Indian culture, their first language was Punjabi, which was also the first language of their parents; they had lived in England all their life. There were 11 female subjects who were from an Indian culture, their first language was Punjabi, which was also the first language of their parents; they had lived in England all their life. It could be argued that all of the subjects were in fact English by nationality, because all of them had been born in England. However, when asked to indicate their culture, there were distinct differences. The subjects who had Indian parents, considered that their culture was Indian because, although there may be superficial manifestations of acculturation in the way that they dressed which was in a more western manner and the purchase of consumer products, they still, nonetheless, kept the deep underlying values of their Indian culture. Thee reasons for keeping these underlying values are based on strong traditional family principles, which are highly respected in Indian cultures, with multi-generational families being the accepted norm. In contrast, there is a tendency for people from English cultures to set up their own homes and live independently from their extended family, often with members of a family living considerable distances apart from each other. Within Indian cultures, arranged marriages are still commonplace, with young men travelling to India from their home in England, to meet their bride for the first time and marry her while he is still in India. In contrast, arranged marriages in England are virtually unheard of in the present day. The emphasis on marriage within Indian cultures is that it is for life, with a very low divorce rate, whereas in English cultures divorce is commonplace with many people marrying more than once.

There were 11 male subjects who were from an English culture; their first language was English, which was also the first language of their parents; they had lived in England all their life. There were 11 female subjects who were from an English culture; their first language was English, which was also the first language of their parents; they had lived in England all their life. All students were students at Thames Valley University, had similar levels of computer literacy but had no previous experience of TTS technology. They were all aged between 18-20 years.

4.9 Implementation Plan

The same testing procedure was followed as for the pilot test as detailed earlier in this chapter using the same testing implementation plan as follows (all documents can be found in the Appendix):

- Information sheet given to subjects.
- Written consent obtained from subjects .

- Questionnaire asking for personal information, including culture is completed by the subjects.

- Task 1 completed.

- Task 2 completed.

In Test 1, subjects heard each sentence three times. Each sentence was in different accents – Engish or Indian but the speech was only played in female sounding voices. This is to eliminate inconclusion in the test results as synthetic speech was only available in female Indian accent at the time of testing, and therefore the decision was made to keep the English accent to female also to eliminate bias. Two of the words were incorrect, one of which was taken from the corpus of words which are commony mistaken and the other was a distractor. They then indicated on the task sheet which one they considered was the correct answer. Each sentence had a word missing There were a total of ten questions.

The results were then analysed and entered into frequency tables and ANOVAs as follows:

Table 4.8. Test 1 - Frequency – Indian Culture, Indian Accent

(T) Total no of responses	**(W) Wrong answers**	**(R) Right answers**	**(P) % of right answers**	**(P-100) % of wrong answers**
110	5	105	95	5

These results look at the error rates when the subjects of Indian culture heard speech in an Indian accent. The results of this test shows that when subjects of Indian Culture heard speech that was also of an Indian culture, the error rate was 5%. According to the research undertaken by Papamichalis (1987), as previously discussed, error rates 10% and under are considered acceptable, so this result is encouraging.

Table 4.9. Test 1 - Frequency – Indian Culture, English Accent

(T) Total no of responses	(W) Wrong answers	(R) Right answers	(P) % of right answers	(P-100) % of wrong answers
110	20	90	82	18

These results look at the error rates when the subjects of Indian culture heard speech in an English accent. The results of this test shows that when subjects of an Indian culture heard speech that was of an English culture, the error rate was 18%. This is an increase in error rate of 13% compared to when the subjects of Indian culture heard speech that sounded as though it was in the same accent as their own. This suggests that the accuracy is increased when Indian culture subjects hear speech in an Indian accent, than when they hear speech in an English accent.

Next, the results from the subjects of English culture were analysed as follows to identify whether or not there was also an increase in accuracy for these subjects if they heard speech that was in the same accent as their own.

Table 4.10. Test 1 - Frequency – English Culture, English Accent

(T) Total no of responses	(W) Wrong answers	(R) Right answers	(P) % of right answers	(P-100) % of wrong answers
110	3	107	97	3

These results look at the error rates when the subjects of English culture heard speech in an English accent. The results of this test shows that when subjects of an English culture heard speech that was also of an English culture, the error rate was 3% and is within the acceptable range of 10% or lower. This result is lower than the subjects from the Indian culture whose error rate was 5% when they heard speech of the same culture and still in the acceptable range.

Table 4.11. Test 1 - Frequency – English Culture, Indian Speech

(T) Total no of responses	(W) Wrong answers	(R) Right answers	(P) % of right answers	(P-100) % of wrong answers
110	18	92	84	16

These results look at the error rates when the subjects of English culture heard speech in an Indian accent. The results of this test shows that when subjects of an English culture heard speech that was from an Indian culture the error rate was 16%. This is a higher error rate than when the culture of both the subjects and speech was both English and is also in line with the increase in error rates when the subjects of Indian culture heard English speech of 18%. The tests so far on subjects from both Indian and English cultures suggest that the error rate is lower in both cultures when they hear speech in the same accent as their own.

The results were then entered into ANOVAs as follows:

Table 4.12. ANOVA – English Culture Subjects for Test 1

Between-Subjects Factors

Factor	**N**
English Accent Speech	22
Indian Accent Speech	22
Male Subjects	22
Female Subjects	22

This table shows the between-subjects factors. There were 44 subjects in total, with an equal divide of male and female gender, and an equal divide of speech heard in both English and Indian accents. All synthesised speech was female gender.

Descriptive Statistics

The following table of descriptive statistics for Test 1 shows the following information.

- The first column refers to the culture of the speech that was heard ie whether it was Indian (in Indian accent) or English (in English accent).

- The second column refers to the gender of the subject (all synthetic voices were in female gender)

- The third column refers to the mean error rate

- The fourth column refers to the standard deviation

- The fifth column is the number of subjects. There were equal numbers of male and female subjects. The frequency of the accents heard (English and Indian) were also equal.

Table 4.13. Descriptive Statistics – English culture subjects for Test 1

Culture of Speech	Gender	Mean	Std. Deviation	N
English	Male	.18	.405	11
English	Female	.36	.674	11
	Total	.27	.550	22
Indian	Male	1.36	.809	11
Indian	Female	.64	.924	11
	Total	1.00	.926	22
Totals	Male	.77	.869	22
	Female	.50	.802	22
	Total	.64	.838	44

From these results, we can see that the total mean average rate for errors for when the English accent is heard is .27, which increases to 1.00 when the Indian accent is heard by English culture subjects. We can also see that the average error rate for female subjects is .50 increasing to .77 for male subjects.

These mean values suggest that there may be some signficance in the results and therefore further parts of the ANOVA were analysed.

Table 4.14. Tests of Between-Subjects Effects – English Culture Subjects

Test 1

Dependent Variable is Errors

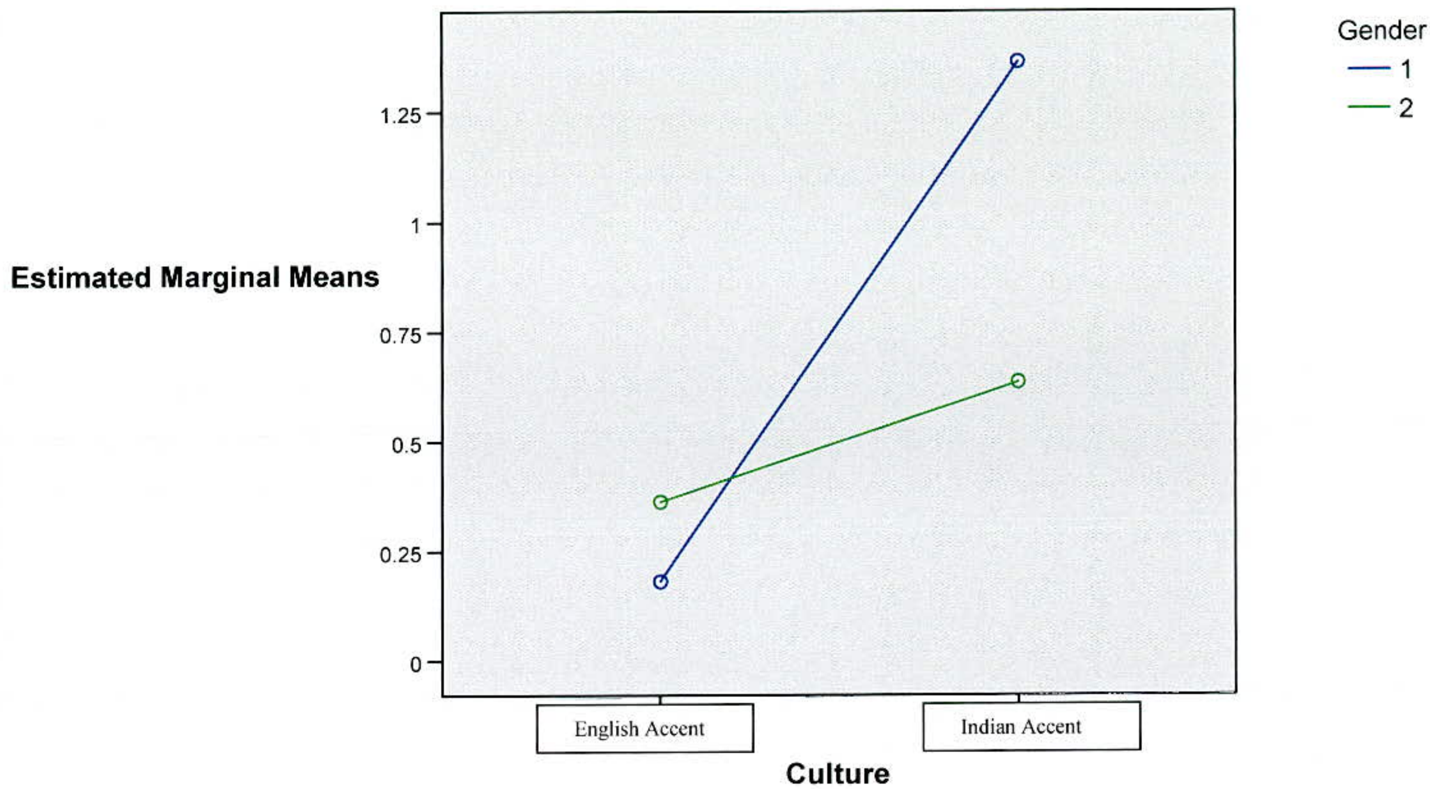
Source	Type III Sum of Squares	df	Mean Square	F	Sig
Corrected Model	8.909[a]	3	2.970	5.584	.003
Intercept	17.818	1	17.818	33.504	.000
Culture of Speech	5.818	1	5.818	10.940	.002
Gender of Subject	.818	1	.818	1.538	.222
Culture of Speech *Gender of Subject	2.273	1	2.273	4.274	.045
Error	21.273	40	.532		
Total	48.000	44			
Corrected Total	30.182	43			

[a] R squared = .295 (Adjusted R Squared = .242)

These results show that there is significant difference for the culture and culture*gender which supports the previous analysis that error rates are lower when subjects from English culture hear speech that is an English accent, than when they hear speech that is in an Indian accent.

These results are shown in the following graphical representation.

Estimated Marginal Means of Errors
English Culture Subjects Test 1



Legend: Gender 1 = male

Gender 2 = female

Summary of Test 1 Results for English Subjects

As shown in the above line graph, it can be seen that:

1) the error rate was lower for both male and female genders when speech was heard in an English accent;

2) when speech was heard in an English accent, the error rate was lower for male subjects than female subjects;

3) when speech was heard in an Indian accent, the error rate was lower for female subjects than male subjects;

4) the error rates overall for female subjects are lower than those for male subjects;

5) there is a lower error rate for both genders when speech is heard that sounds as though it is in an English accent; and

6) there is an interaction for culture and gender.

ANOVA Summary

As can be seen in Table 4.14, a two way unrelated ANOVA showed that significant effects were obtained for the Culture of Speech ($F_{1,40}$= 10.940, p = 0.002) and for the culture of speech * gender ($F_{1,40}$= 4.274, p = 0.045).

After analysing the results for the English culture subjects, the same technique was used to analyse the results for the Indian culture subjects as follows:

Table 4.15. ANOVA – Indian Culture Subjects for Test 1

Between-Subjects Factors

Factor	N
Indian Accent Speech	22
English Accent Speech	22
Male Subjects	22
Female Subjects	22

This table shows the between-subjects factors for the ANOVA. There were 44 subjects in total, with an equal divide of male and female gender, and an equal divide of speech heard in both English and Indian accents. All synthesised speech was female.

Descriptive Statistics

The following table of descriptive statistics for Test 1 shows the following information.

- The first column refers to the culture of the speech that was heard ie whether it was Indian (in Indian accent) or English (in English accent).

- The second column refers to the gender of the subject (all synthetic voices were in female gender)

- The third column refers to the mean error rate

- The fourth column refers to the standard deviation

- The fifth column is the number of subjects. There were equal numbers of male and female subjects. The frequency of the accents heard (English and Indian) were also equal.

Table 4.16. Descriptive Statistics for Indian Culture Subjects for Test 1

Dependent Variable: Errors

Culture of Speech	Gender	Mean	Std. Deviation	N
Indian	Male	.27	.467	11
Indian	Female	.64	1.027	11
	Total	.45	.800	22
English	Male	1.27	1.421	11
English	Female	.91	1.044	11
	Total	1.09	1.231	22
Totals	Male	.77	1.152	22
	Female	.77	1.020	22
	Total	.77	1.075	44

From these results, we can see that the total mean average rate for errors for when the Indian accent is heard is .45, which increases to 1.09 when the English accent is heard by Indian culture subjects. We can also see that there is no significant difference in the mean average error rate for male and female subjects at .77.

Again, these mean values suggest that there may be some signficance in the results and therefore further parts of the ANOVA were analysed.

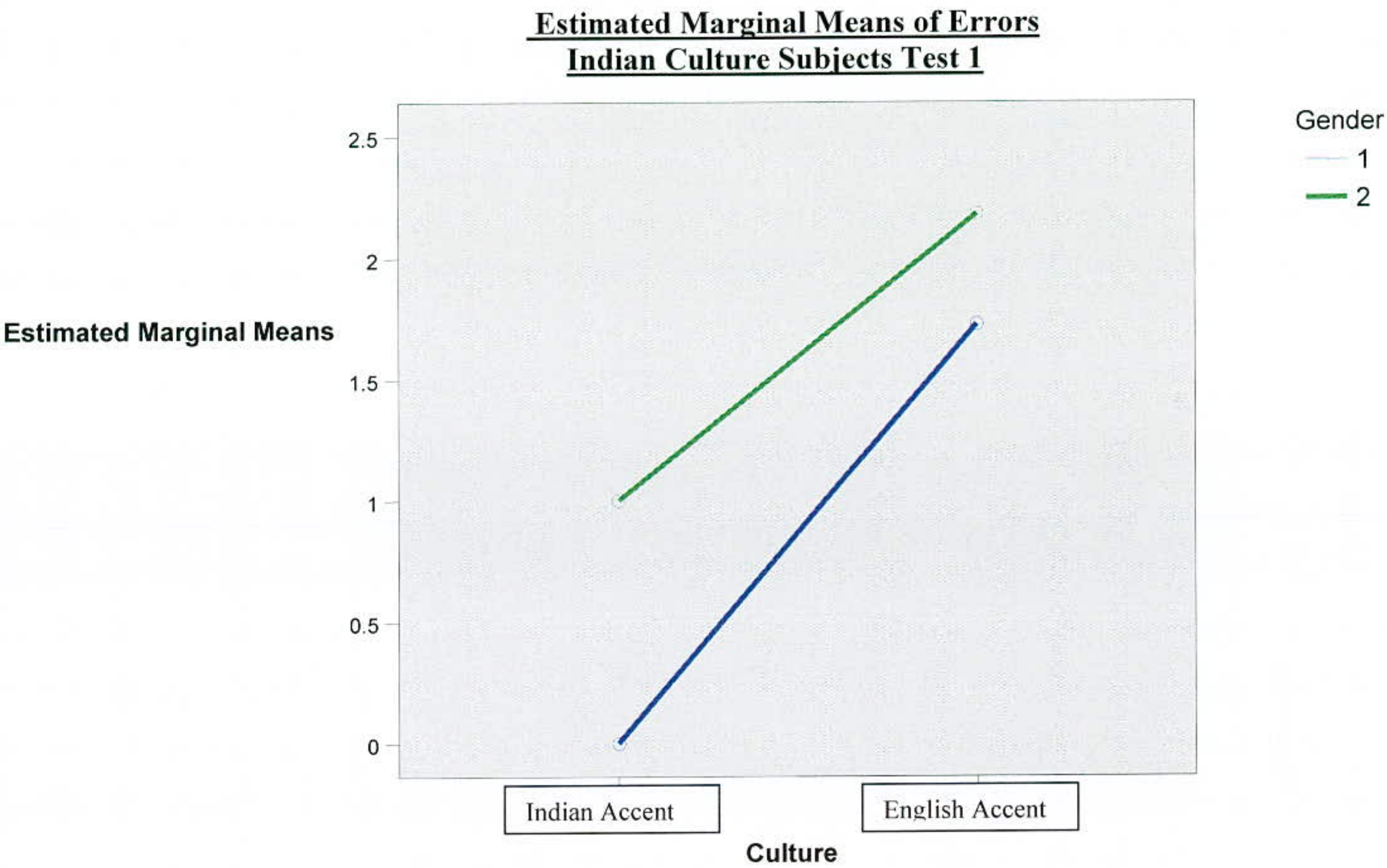
Table 4.17. Tests of Between-Subjects Effects

Dependent Variable is Errors – Indian Culture Subjects for Test 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig
Corrected Model	5.909[a]	3	1.970	1.798	.163
Intercept	26.273	1	26.273	23.983	.000
Culture of Speech	4.455	1	4.455	4.066	.050
Gender of Subject	.000	1	0.000	0.000	1.000
Culture of Speech *Gender of Subject	1.455	1	1.455	1.328	.256
Error	43.818	40	1.095		
Total	76.000	44			
Corrected Total	49.727	43			

R Squared = .119 (Adjusted R Squared = .053)

These results show that there is significant difference for the culture which supports the previous analysis that error rates are lower when subjects hear speech that is in the same culture as their own. These results are shown in the following graphical representation.



Legend: Gender 1 = male

Gender 2 = female

Summary of Test 1 Results for Indian Subjects

As shown graphically in the above line graph, it can be seen that:

1) the error rate was lower for both male and female genders when speech was heard in Indian accent;

2) when speech was heard in an Indian accent, the error rate was lower for male subjects than female subjects;

3) the error rates overall for female subjects and male subjects are the same;

4) there is a lower error rate for both genders when speech is heard that sounds as though it is in an Indian accent.

<u>ANOVA Summary</u>

As can be seen in Table 4.17, a two way unrelated ANOVA showed that significant effects were obtained for the Culture of Speech ($F_{1,40}$= 4.066, p = 0.050).

Summary of Test 1 Results for Subjects from Both Cultures

The findings from Test 1 show:

1) for subjects from English and Indian cultures, the error rate was lower for both male and female genders when speech was heard in the accent that sounded the same as their own;

2) for subjects from English and Indian cultures, the error rate was lower for male subjects than female subjects when speech was heard in the same culture as their own;

3) when speech was heard in an accent different to their culture, the error rate was lower for female subjects than male subjects for the English culture but the error rate was lower for male subjects from the Indian culture. The error rates overall for male and female subjects are not significantly different for the Indian culture, but the English culture the error rates are lower for female subjects than for male subjects.

Test 2 Results

After completing Test 1, the subjects moved onto Test 2 following the same procedure as for the pilot test as detailed earlier in this chapter, using the same implementation plan. In Test 1, subjects had to identify the correct word to complete a sentence; in contrast previous work by others (Papamichalis, 1987) has suggested that the output of TTS synthesis is more accurate when only individual words/sounds are heared one at a time with the theory that the perceived accuracy of the synthesised speech is reduced because of its lack of prosody and intonation; the findings of the pilot test support this theory. Subsequently, Test 2 required subjects to indicate the word that they heard from a choice of two words which are taken from the DRT to see if that is the case with this research. Test 2 asked the subjects to identify the word/sound that they thought they heard from a choice of two possible answers and to circle this word/sound on an answer sheet. An example of the answer sheet can be found in the Appendix. Each pair of words were in different accents - Engish or Indian but the speech was only played in female sounding voices and, as with Test 1 were randomised to prevent prior learning. This is to eliminate inconclusion in the test results, as detailed earlier in this chapter synthesised speech was only available in female Indian accent at the time of testing, and therefore the decision was made to keep the English accent to female also.

Only one of the words/sounds was correct and was taken from the corpus of words which are commony mistaken. They subjects indicated on the task sheet (page 164) which one they considered was the word that they thought they heard. Words were not in any sort of context. There were a total of twenty tasks.

The results were then analysed and entered into frequency tables and ANOVAs as follows:

Table 4.18. Test 2 - Frequency – Indian Culture, Indian Accent

(T) Total no of responses	**(W) Wrong answers**	**(R) Right answers**	**(P) % of right answers**	**(P-100) % of wrong answers**
220	8	212	96	4

These results look at the error rates when subjects of Indian culture heard speech in an Indian accent. The results of this test shows that when subjects of an Indian culture heard speech that was also of an Indian culture, the error rate was 4% and is in the acceptable range of 10% or lower.

Table 4.19. Test 2 - Frequency – Indian Culture, English Accent

(T) Total no of responses	(W) Wrong answers	(R) Right answers	(P) % of right answers	(P-100) % of wrong answers
220	39	181	82	18

These results look at the error rates when the subjects of Indian culture heard speech in an English accent. The results of this test shows that when subjects of an Indian culture heard speech that was from an English culture, the error rate was 18%. This is an increase of 14% from when they heard speech that was of the same culture.

Table 4.20. Test 2 - Frequency - English Culture, English Accent

(T) Total no of responses	(W) Wrong answers	(R) Right answers	(P) % of right answers	(P-100) % of wrong answers
220	9	211	96	4

These results look at the error rates when the subjects of English culture heard speech in an English accent. The results of this test shows that when subjects of an English culture heard speech that was also of an English culture, the error rate was 4% and is within the acceptable range of 10% or lower. The result is lower than the subjects from the Indian culture whose error rate was 5% when they heard speech of the same culture and is still in the acceptable range.

Table 4.21. Test 2 - Frequency – English Culture, Indian Accent

(T) Total no of responses	(W) Wrong answers	(R) Right answers	(P) % of right answers	(P-100) % of wrong answers
220	32	188	85	15

These results show that the error rate when subjects from the English culture heard speech that sounded as though it was from the Indian culture is 15%. This is an increase in error rate of 11% from when the English culture subjects heard speech of the same culture.

To establish whether the findings were significant, the data was entered into SPSS and a two way analyses of variance (ANOVA) for uncorrelated scores was performed. This method allows a comparison to be made of the means of the dependent variables when there are two independent variables.

Table 4.22. ANOVA – English Culture Subjects for Test 2

Between-Subjects Factors

Factor	N
English Accent Speech	22
Indian Accent Speech	22
Male Subjects	22
Female Subjects	22

This table shows the between-subjects factors for the ANOVA. There were 44 subjects in total, with an equal divide of male and female gender, and an equal divide of speech heard in both English and Indian accents.

Descriptive Statistics

The following table of descriptive statistics for Test 2 shows the following information.

- The first column refers to the culture of the speech that was heard ie whether it was Indian (in Indian accent) or English (in English accent).

- The second column refers to the gender of the subject (all synthetic voices were in female gender)

- The third column refers to the mean error rate

- The fourth column refers to the standard deviation

- The fifth column is the number of subjects. There were equal numbers of male and female subjects. The frequency of the accents heard (English and Indian) were also equal.

Table 4.22. Descriptive Statistics – English culture subjects for Test 2

Culture of Speech	**Gender**	**Mean**	**Std. Deviation**	**N**
English	Male	.73	.905	11
English	Female	.09	.302	11
	Total	.41	.734	22
Indian	Male	2.09	1.868	11
Indian	Female	1.00	.775	11
	Total	1.55	1.503	22
Totals	Male	1.41	1.593	22
	Female	.55	.739	22
	Total	.98	1.303	44

From these results, we can see that the total mean average rate for errors for when the English accent is heard is .41, which increases to 1.55 when the Indian accent is heard by English culture subjects. We can also see that the average error rate for female subjects is .55 rising to 1.41 for male subjects.

Again, these mean values suggest that there may be some signficance in the results and therefore further parts of the ANOVA were analysed.

Table 4.23. Tests of Between-Subjects Effects

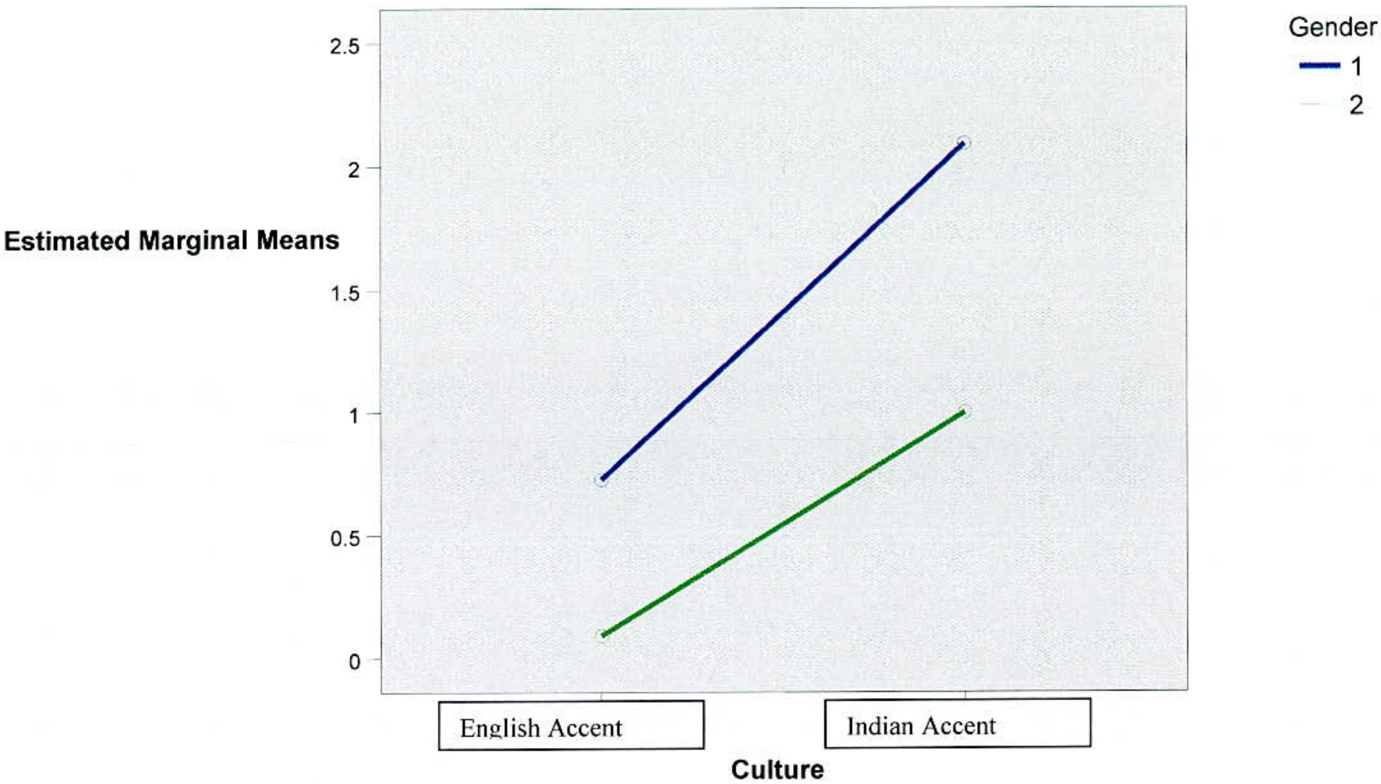
Dependent Variable is Errors – English Culture Subjects for Test 2

Source	Type III Sum of Squares	df	Mean Square	F	Sig
Corrected Model	22.977[a]	3	7.659	6.127	.002
Intercept	42.023	1	42.023	33.618	.000
Culture of Speech	14.205	1	14.205	11.364	.002
Gender of Subject	8.205	1	8.205	6.564	.014
Culture of Speech *Gender of Subject	.568	1	.568	.455	.504
Error	50.000	40	1.250		
Total	115.000	44			
Corrected Total	72.977	43			

a. R Squared = 3.15 (Adjusted R Squared = .263)

These results show that there is significant difference for the culture and gender which supports the previous analysis that error rates are lower when subjects from English culture hear speech that is in an English accent, than when they hear speech that is in an Indian accent. These results are shown in the following graphical representation.

Estimated Marginal Means of Errors
English Culture Subjects Test 2



Legend: Gender 1 = male

Gender 2 = female

Summary of Test 2 Results for English Culture Subjects

As shown graphically in the previous line graph, it can be seen that:

1) the error rate was lower for both male and female genders when speech was heard in English accent;

2) when speech was heard in an English accent, the error rate was lower for female subjects than male subjects;

3) when speech was heard in an Indian accent, the error rate was lower for female subjects than male subjects;

4) the error rates overall for female subjects are lower than male subjects.

ANOVA Summary

As can be seen in Table 4.23, a two way unrelated ANOVA showed that significant effects were obtained for the Culture of Speech ($F_{1,40}$= 11.364, p = 0.002) and the gender ($F_{1,40}$= 6.564, p = 0.014).

Table 4.24. ANOVA – Indian Culture Subjects for Test 2

Between-Subjects Factors

Factor	**N**
Indian Accent Speech	22
English Accent Speech	22
Male Subjects	22
Female Subjects	22

This table shows the between-subjects factors for the ANOVA. There were 44 subjects in total, with an equal split of male and female gender, and an equal split of speech heard in both English and Indian accents. All synthesised speech was female gender.

Descriptive Statistics

The following table of descriptive statistics for Test 2 shows the following information.

- The first column refers to the culture of the speech that was heard ie whether it was Indian (in Indian accent) or English (in English accent).

- The second column refers to the gender of the subject (all synthetic voices were in female gender)

- The third column refers to the mean error rate

- The fourth column refers to the standard deviation

- The fifth column is the number of subjects. There were equal numbers of male and female subjects. The frequency of the accents heard (English and Indian) were also equal.

Table 4.25. Descriptive Statistics – Indian culture subjects for Test 2

Culture of Speech	**Gender**	**Mean**	**Std. Deviation**	**N**
Indian	Male	.00	.000	11
Indian	Female	1.00	1.342	11
	Total	.50	1.058	22
English	Male	1.73	2.054	11
English	Female	2.18	1.328	11
	Total	1.95	1.704	22
Totals	Male	.86	1.670	22
	Female	1.59	1.436	22
	Total	1.23	1.583	44

From these results, we can see that the mean average rate for errors for when the Indian accent is heard is .50, which increases to 1.95 when the English accent is heard by Indian culture subjects. We can also see that the average error rate for male subjects is .86 rising to 1.59 for female subjects. These mean values suggest that there may be some significance in the results and therefore further parts of the ANOVA were analysed.

Table 4.26. Tests of Between-Subjects Effects – Indian Culture Subjects

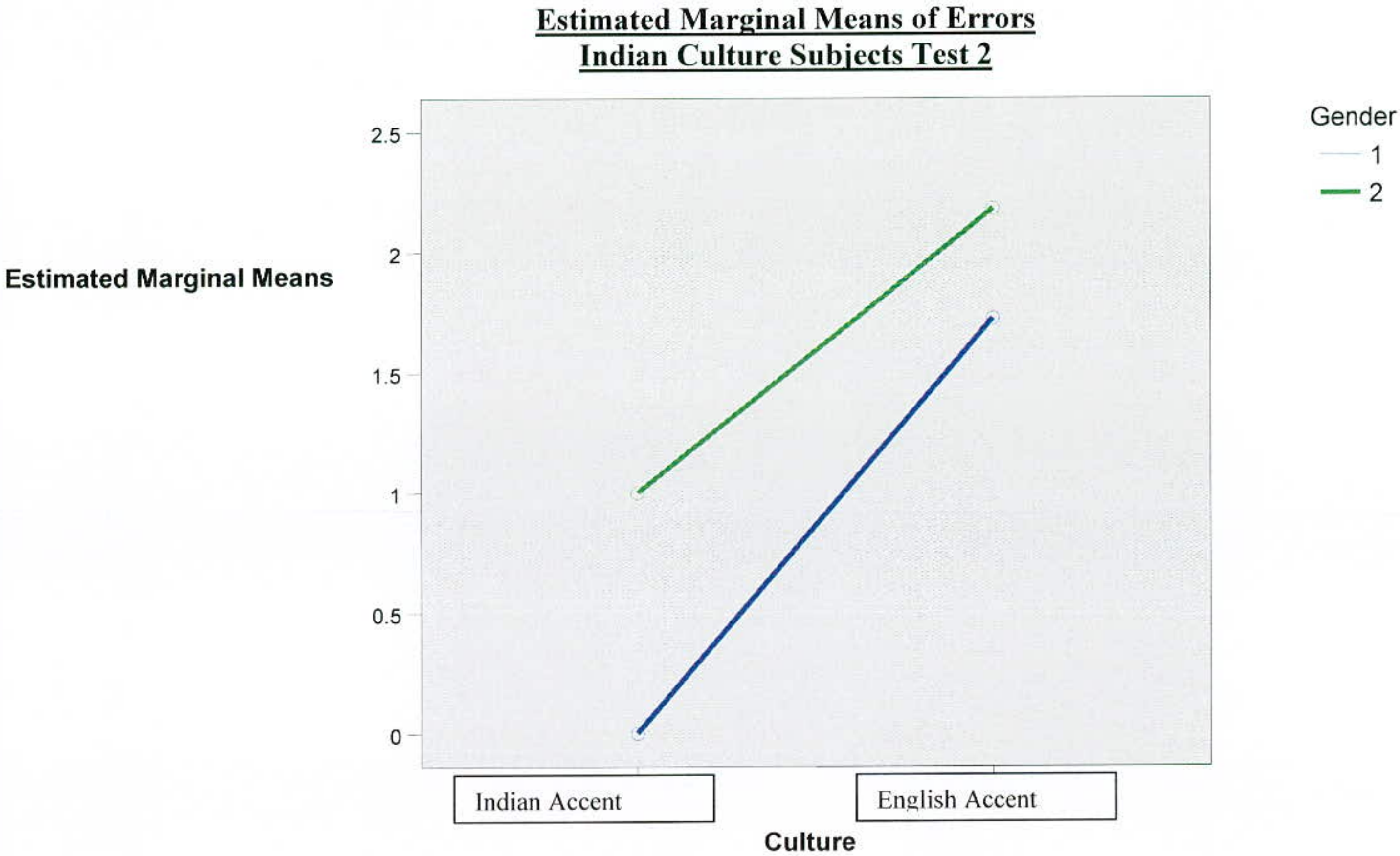
Test 2

Dependent Variable is Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig
Corrected Model	29.909[a]	3	9.970	5.125	.004
Intercept	66.273	1	66.273	34.065	.000
Culture of Speech	23.273	1	23.273	11.963	.001
Gender of Subject	5.818	1	5.818	2.991	.091
Culture of Speech *Gender of Subject	.818	1	.818	.421	.520
Error	77.818	40	1.945		
Total	174.000	44			
Corrected Total	107.727	43			

a. R Squared = .278 (Adjusted R Squared = .223)

These results show that there is a significance for the culture which supports the previous analyses that error rates are lower when subjects from Indian culture hear speech that is an Indian accent, than when they hear speech that is in an English accent. These results are shown in the following graphical representation.



Legend: Gender 1 = male

Gender 2 = female

Summary of Test 2 Results for Indian Subjects

As shown graphically in the above line graph, it can be seen that:

1) the error rate was lower for both male and female genders when speech was heard in Indian accent;

2) when speech was heard in an Indian accent, the error rate was lower for male subjects than female subjects;

3) when speech was heard in an English accent, the error rate was lower for male subjects than female subjects;

4) the error rates overall for male subjects are lower than those for female subjects.

ANOVA Summary

As can be seen in Table 4.27, a two way unrelated ANOVA showed that significant effects were obtained for the Culture of Speech ($F_{1,40}$= 11.963, p = 0.001).

Summary of Test 2 Results for Subjects From Both Cultures

The findings from Test 2 show:

1) for subjects from English and Indian cultures, the error rate was lower for both male and female genders when speech was heard in the accent that sounded the same as their own;

2) for English culture subjects, the error rate was lower for female subjects than male subjects when speech was heard in the same accent but

for Indian culture subjects, the error rate was lower for male subjects than female subjects when speech was heard in the same accent;

3) for English culture subjects when speech was heard in an accent different to their own, the error rate was lower for female subjects than male subjects but for Indian culture subjects, the error rate was lower for male subjects than female subjects when speech was heard in an accent different to their own;

4) for English culture subjects the overall error rates are lower for female subjects than male subjects for for Indian culture subjects, t he overall error rates are lower for male than female subjects.

4.10 Comparison of Results for Test 1 and Test 2

When comparing the results from both Test 1 and Test 2 it can be seen that:

1) In terms of error rates, the findings were consistent across both tests, for subjects from both English and Indian cultures. The error rate was lower for both male and female genders when speech was heard in the accent that sounded the same as their own native culture.

2) When analysing the results for errors made by subjects in their own accent, there were differences in the findings as follows. For Test 1 the error rates for both cultures was lower for male subjects than female

subjects but for Test 2, there was a lower error rate for female subjects than male subjects for English culture but the error rate was lower for male subjects than female subjects for Indian culture. The findings for the Indian culture, therefore are consistent in this respect ie the error rates for the male subjects are lower across both tests than for the female subjects.

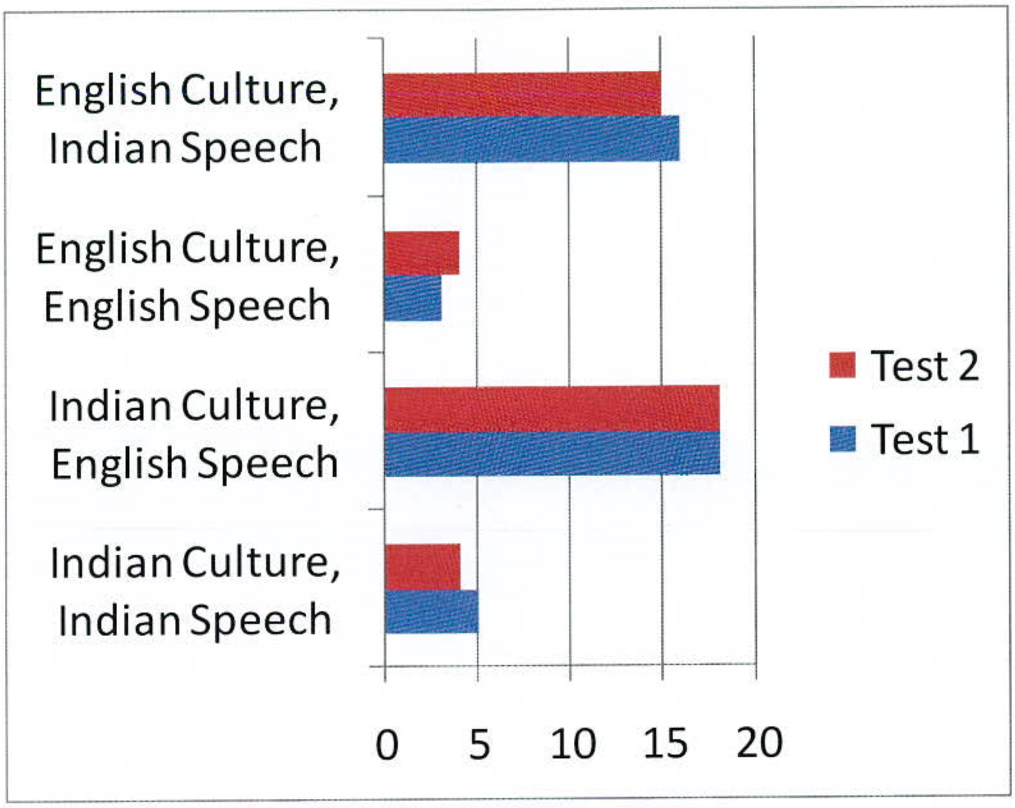
3) When analysing the results for the errors made by subjects in the accent different to their own there were differences. In Test 1, the error rate for both cultures was lower for female subjects than male subjects but in Test 2 the error rate was lower for female subjects than male subjects for the English culture but for Indian culture subjects, the error rate was lower for male subjects than female subjects. The findings, therefore in terms of gender are inconsistent in this respect.

4) In Test 1 the overall error rates for female and male subjects do not have any significant difference for the Indian culture but for the English culture, female subjects had a lower error rate than male subjects. In Test 2 the overall error rates are lower for female subjects for English culture subjects for but for Indian culture subjects, the overall error rates are lower for male than female subjects. The findings, therefore in terms of gender are inconsistent in this respect.

The findings from both of the tests show consistently, that across both cultures, the error rate is significantly lower when speech is heard that sounds as though it is in the same accent as the subject. The findings from both of the tests regarding the effect of gender is inconsistent as the results for the English subjects varies but, however, the error rates for male subjects from the Indian culture is consistently lower than for female subjects.

The following graphical representation shows the % error rates for each culture across the two tests. Here we can see that for the English culture hearing Indian accent speech, the error rate was 16% for Test 1 – this is outside of the acceptable range of 10%. For Test 2, the error rate was 15%, again outside of the acceptable range. However, when they heard speech that sounded as though it was in their accent, the error rate was 3% for Test 1 and 4% for Test 2 – this is in the acceptable range. For the Indian culture subjects hearing English accent speech, the error rate was 18% for both Tests 1 and 2, again outside of the acceptable range. However, when they heard speech that sounded as though it was in their native cultural accent, the error rate was 5% for Test 1 and 4% for Test 2 – this is in the acceptable range.

Comparison of Test Results - % Error Rates



In terms of supporting the theory held by Papamichalis (1997) that error rates are lower when subjects hear speech as individual words and that accuracy diminishes when larger chunks of text are heard due to its lack of prosody and intonation, these results are inconclusive.

However, these findings may explain why speech has not yet become the ultimate human computer interface – the error rates are too high for it to be used successfully, if the end users cannot hear speech that sounds as though it preserves their accent.

4.11 Hypothesis Testing

The purpose of the tests was to consider the following hypotheses which were considered and show whether they are rejected or supported by the findings of the tests.

Hypothesis (a) – There is no difference in error rate when the accent of the speech is representative of the culture of the subjects.

Test 1 – The ANOVAs showed that significant effects were obtained for the accent of speech:

($F_{1,40}$ = 10.940, p=0.002) for the English culture subjects and;

($F_{1,40}$ = 4.066, p=0.050) for the Indian culture subjects.

Test 2 – The ANOVAs showed that significant effects were obtained for the accent of speech:

($F_{1,40}$ = 11.364, p=0.002) for the English culture subjects and;

($F_{1,40}$ = 11.963, p=0.001) for the Indian culture subjects.

These findings reject this hypothesis.

Hypothesis (b) – The gender of the subjects makes no difference to the error rate.

Test 1 – The ANOVAs showed that significant effects were not obtained for the gender for the English culture subjects ($F_{1,40}$ = 1.538, p=0.222) or for the Indian culture subjects ($F_{1,40}$ = 0, p=1.000) ;

Test 2 – The ANOVAs showed that significant effects were obtained for the gender for the English culture subjects ($F_{1,40}$ = 6.564, p=0.014) but not for the Indian culture subjects ($F_{1,40}$ = 2.991, p=0.091)

These findings support this hypothesis.

4.12 Chapter Summary

This chapter details the procedure and results of experiments that were performed in controlled test conditions at Thames Valley University using generic text to speech synthesis software. Details of the pilot tests are given where American English and English English accents were used to identify anomalies in the testing procedure and consider whether there was a phenomenon to be investigated. The anomalies identified in the pilot testing procedure are stated together with the rationale why the results of the pilot tests are not included in the final analysis. Details of the tests that followed on forty four subjects from English and Indian cultures are given with comprehensive analysis of the results using ANOVAs for the English and Indian cultures who took part in the actual tests. Two hypotheses are tested. The null hypothesis (a) is rejected with the ANOVAs suggesting that the accent of the speech does make a difference to error rate. The null hypothesis

(b) is supported with the ANOVAs suggesting that the gender of the subject makes no difference to the error rate.

With regards to the theory held by Papamichalis (1997), that error rates are lower when words are heard individually, rather than in prose, the results of these tests are inconclusive.

In summary, these results may provide an explanation of why speech is not the ultimate HCI yet as, the error rate at greater than 10% when end users hear speech that does not preserve their accent, is just too high.

The findings of these tests are discussed in further detail in Chapter Seven, after a further two set of tests have been completed using different subjects and in different environments.

After completing experiments using generic software, it was identified that there was a significant improvement in error rate when the speech that was output was in the same accent as the end user. However, the tests had been undertaken in laboratory conditions, they were performed in the English language (although they were in different accents) and the end users did not have SEN. Therefore, further tests were undertaken. This time, however, the target group was end users who may have SEN, in the end users own environment outside of the laboratory and in a language other than English. Full details can be found in the next chapter.

Chapter Five – Experiments in Kenya

5.1 Chapter Introduction

In Chapter Four, tests are detailed that had been performed in controlled conditions in a computer laboratory at Thames Valley University. These tests had been used to solely measure error rates with the rationale that this component is the most important part of usability, as if the error rate is unacceptably high, the system can be rendered as unusable. The research of Papamichalis (1987), considers that 10% is the maximum amount of errors for a text to speech system to be considered satisfactory. Indeed, unsatisfactory error rates are catastrophic for information systems and negate any other positive aspects of the system, deeming it as unusable.

The tests in this chapter approach usability testing from a different dimension. The system that is being tested this time is an information system that has been created as part of an initiative by the Ministry of Agriculture in Kenya in collaboration with the LLSTI at the University of Bristol the University of Nairobi and the National Agriculture and Livestock Extension Programme (NALEP) who are responsible for providing extension services to farmers in providing support and information. The overall aim behind the introduction of this system is to help banana farmers become better educated in crop management and self-sufficient, subsequently reducing poverty within the country, by being able to glean previously inaccessible information, through the interface of speech.

Instead of the tests being performed in the computer laboratory, they are undertaken in Kenya in the environment in which they will be used, by subjects from the target group of end users.

The results of the evaluation of the system will be presented to the Kenyan government, NALEP and to the Kenya Society for the Blind who have all previously confirmed to the LLSTI the need for voice based services as a reference system for farmers, especially the visually impaired or illiterate, so it is therefore important that this system is evaluated using techniques that consider ethical issues and provide results that are credible.

As with the previous tests, the analysis of the error rates is the main aim of the tests. In contrast to the tests in Chapter Four, however, these tests will be available in both Kiswahili and English Language. The reason for this is that English and Kiswahili are both official languages taught in school in Kenya and one or the other (or sometimes both) may be adopted as the first language of the Kenyan people. In addition, it is expected that the subjects may have SEN as for most of the middle-aged farmers, their education stops at the end of primary school; many of them are illiterate and in addition 75% of the visually impaired in Kenya are also farmers.

If it is possible to create an effective and accurate banana information system, accessible by telephone, the LLSTI considers that it could be possible to extend the system to also provide information on other crops and livestock. This information could be made up to date, local and personal. The system could also include additional functionality such as SMS and e-mail integration.

In addition to testing error rates, these experiments have the added objectives of exploring learnability and satisfaction to establish whether being an end user from a developing country has any effect on the willingness to use ICT.

The results of the usability tests in Chapter Four suggested that usability is increased when end users hear speech which preserves their accent. However, in the previous tests, the words that they heard were spoken individually from a corpus of words used in a test which were commonly mistaken in TTS technology and not from a TTS system directly for a specific task. With this section, however, testing was undertaken with a
system, which had been developed using text in paragraphs from a standard vocabulary.

The experiments were created, analysed and evaluated by the author but were conducted in Kenya under the supervision of the LLSTI, NALEP and colleagues from the University of Nairobi.

The purpose of the experiments is to test hypothesis c) *There is no difference in error rate when the language of the speech is representative of the culture of the subjects.*

In addition, the tests will identify whether banana farmers, who may be illiterate and inexperienced in using ICT would be willing to use such a system and whether the communications network in Kenya is robust enough to support it.

5.2 Cultural Aspects of Kenya

To glean an appreciation of the environment in which the end users are expected to use this information sytem, an overview of the cultural aspects of Kenya are given below .

Owing to its diverse range of regional people, Kenya consists of people from many different cultures. It is located in Eastern Africa and borders the Indian Ocean between Somalia and Tanzania with a total area of 582,650 sq km which consists of 569, 250 sq km land and 13,400 sq km water; there is a tropical climate along the coast to an arid climate inland. The terrain consists of low plains rising to highlands and fertile plateau in the west of the country. 8.08% of the land is used for arable farming, 0.98% for permanent crops and 90.94% for other use; 670sq km of the land is irrigated, (Central Intelligence Agency, 2006). The main hazards encountered by the farming community include recurring drought followed by flooding during rainy seasons together with water pollution from urban and industrial wastes; degradation of water quality from increased use of pesticides and fertilisers; water hyacinth infestation; deforestation; soil erosion; desertification and poaching, (Central Intelligence Agency, 2006).

Kenya is a country that has excessive mortality rates due to AIDS; in 2003 there were 1.2 million people living with AIDS in Kenya which resulted in 150,000 deaths in the same year, (Central Intelligence Agency, 2006). The effect of AIDS, combined with other significant risk of infectious diseases which include bacterial and protozoal diarrhoea, hepatitis A, typhoid fever, malaria and schistosomiasis, results in lower life expectancy, higher infant mortality and death rates. This also contributes to lower population and growth rates and changes in the distribution of population by age and sex than would otherwise be expected are also a factor, (Central Intelligence Agency, 2006). The population of Kenya is around 34,000,000 and is made up of 42.5% 0-14 years, 55.2% 15-64 years and 2.3% 65 years +; the median age is 18.19 years. By the end of 2001, Kenya was host to 220,000 refugees from neighbouring countries particularly Somalia and Sudan, (Central Intelligence Agency, 2006).

The population of Kenya is sub-divided into the following ethnic groups – Kikuyu 22%, Luhya 14%, Luo 13%, Kalenjiin 12%, Kamba 11%, Kisii 6%, Meru 6% and other African 15%, non-African 1%. 45% of the population are Protestant; 33% are Roman Catholic, 10% have indigenous beliefs, 10% are Muslim and 2% other religions, (Central Intelligence Agency, 2006).

In 2003, statistics showed that 85.1% of the population aged 15+ could read and write, broken down into 90.6% male and 79.7% female yet 40% of the adult population are unemployed with 75% of those employed working in agriculture; 50% of the population are living below the poverty line, (Central Intelligence Agency, 2006). However, the majority of the population who can read and write are situated in the urban areas where they are better educated and the residents of the more rural areas, particularly those in the farming industry are more likely to have SEN due to the fees for secondary education, meaning that their education stops at the end of primary school.

The telephone system in Kenya is unreliable with little attempt made by the government to modernise except for service to business. Trunks are primarily microwave radio relay with business data being commonly transferred by a very small aperture terminal (VSAT) system. Statistics for 2003 state that there were 328.400 main line telephones in use and 1,590,800 mobile cellular telephones with 8.325 Internet hosts and 400,000 Internet users, (Central Intelligence Agency, 2006) .

English and Kiswahili, also referred to as Swahili, are the official languages of Kenya. Both are taught in school but, however, neither one of them is more popular than the other as the most common language spoken at home with rural Kenyans preferring to use their traditional tribal language. There is not an area of Kenya where people cannot speak Kiswahili but conversely, there are many places outside of the towns where English is not spoken, (World 66, 2006). Kiswahili belongs to the Sabaki subgroup of the

Northeastern coast Bantu languages and is closely related to the Miji Kenda group of languages which includes Pokomo and Jgazija. Over the period of over one thousand years, intense and varied interaction with the Middle East, Arabia, Persia, Indian and China have given Kiswahili a rich infusion of loan words from a wide assortment of languages, (Swahili Language, 2006). As in English, the proportion of loan words that are spoken changes depending on the context in which the speaker is communicating. For example, in English a discussion of bodily functions sounds much more appropriate if you use particular words commonly of Latin derivation. Similarly, an educated Kiswahili speaker will use more Arabic-derived words with English terms in polite circumstances, even though the same phrase could usually be said in Kiswahili using words of Bantu origin alone

5.3 Speech used in the Banana Information System

The Festival TTS system, a general multi-lingual speech synthesis system developed at the Centre for Speech Technology Research (CSTR) at the University of Edinburgh which is distributed under a free software license and is written in C++, was used to create the banana information system. For building the Kiswahili voice, the Festvox suite of tools was used, which allows new synthetic voices to be made. The Festival and Festvox toolset was extended by LLSTI to simplify the building of new voices (Shalanova and Tucker, 2004). Festival is a concatenative TTS system which puts together short units of speech, selected from a speech database produced from human recordings and places them into new long sequences of speech. The unit chosen for the Kiswahili voice was the diphone, which is a unit that normally starts half-way through the phone and ends half-way through the second. Diphones are often used in TTS since they are relatively easy to join and can capture a reasonable amount of the unit's context, in balance to the number of units needed to record. Jurafsky and Martin (2000), state that typically, there are between 1000-2000 diphones in a language.

To be able to cover all the necessary diphones in a language a speech database needs to be designed that contains a limited set of phonetically balanced sentences and for the Kiswahili voice, a text selection tool was used to select such sentences form a large Kiswahili text corpus, (Gakuru et al, 2004). The sentences were then recorded, using formal grammar, in a studio at the Kenya Broadcasting Corporation by the professional newscaster Ken Walibora (Gakuru et al, 2005). Tucker and Shalanova (2005), consider that selecting a normative speaker can be the single most important decision in the whole system. Indeed, the effects that this has on the usability of TTS technology is being investigated, with the findings of the usability tests that have already been performed as detailed in Chapter Four, supporting this theory.

In contrast, to having a native speaker (Male, Kiswahili), the English voice was a prototype female voice taken from the CSTR database which had an English accent and does not sound like a native Kenyan speaker. The developers are in the process of developing a voice in English language with a Kenyan accent, but this was not available at the time of the testing.

5.4 Banana Information System

Once the system had been completed, it was hosted on a website which would be accessible to the banana farmers via a TBI. However, connectivity to the Internet is a key criterion because without this, accessibility to the speech system would not be possible.

Prior to the development of the banana information system, the LLSTI spent two weeks in Kenya analysing the requirements of the end users which included observation of the farmers and life in Kenya, meetings with government representatives and NALEP. Once they had identified the end user requirements, they designed and developed the system at the LLSTI, University of Bristol. The main sources of technical information for the system came from the Fruits and Vegetables Technical Handbook from the Ministry of Agriculture and Rural Development, (Ministry of Agriculture and Rural Development, 2003) and by interviews with agricultural extension workers from NALEP. When the system had been created, Nasfors returned to Kenya to undertake further development of the system, to undertake usability testing and to implement the system.

At the present time, farmers who are growing banana crops are finding it difficult to access the information they need to be successful, for several reasons. Firstly some of them cannot read or write and where written information is available in libraries, for example, they cannot read the text. In addition, many of them are based in rural locations and cannot travel to the libraries which are based in the towns. The Ministry of Agriculture has recognised this problem and appointed extension workers to travel out to visit the farmers but they are too few in numbers. Therefore, it was decided that a system that the farmers could access remotely by telephone and could select the appropriate information and find out the answer immediately may be a solution to this problem. This information system is accessed by the end users pressing buttons on the telephone keypad to select from a menu. This method of gaining information would not require any pre-determined standards of numeracy or literacy or the need to travel. However, this method would also have other evaluation criteria that needed to be considered which included whether or not the farmers had access to telephones, whether they would be willing to embrace this technology as the overwhelming majority of them had never used ICT before and whether the telecommunications system was robust enough to support such a system.

The text representing the speech output from the system, which is available in both English (English accent) female and Kiswahili, male gender can be found in the Appendix. The user is faced, initially, with the start menu and the option to choose between the language that they wish to hear by pressing 1 for English and 2 for Kiswahili. Once they have made their selection, they then go on automatically to the main menu where they press a number on the telephone keypad to go to the appropriate section and from there they can listen to information about key points in banana farming.

The structure for the menu is shown in the Appendix.

The developers of the system consider it to be structured into different categories to provide easy navigation within the system but in contrast, however, on perusing this structure the theory was explored that the system did not provide easy navigation. The reasons for this are from the initial greeting, the end user is faced with nine options, with two further menus to navigate each with separate options; there is a lot of speech to listen to and that this may cause problems. As an example, if the end user wants to find out about the Control of Sigatoke Leaf Spots (1.7.1.6) they have to select Pests and diseases (1.7) from the first menu, there is then a further eight options and if the end user selects Control of pests and diseases (1.7.1) from the second menu, there is a further six options from the third menu before they can find the information they are looking for. Therefore, if the end user wanted to find out information about the Control of Sigatoke Leaf Spots, the end user had to listen to seven options from the main menu, one option from the second menu and six options from the third menu. Using this example, the end user would have to listen to fourteen options first. It is uncertain, therefore, how usable this system would be for farmers who may be inexperienced in using ICT, telephones, and have low levels of education.

In Chapter Two, the work of Brewster (2008), who has undertaken an evaluation of TBI, was discussed. Brewster (2008), considers that TBI are becoming an increasingly important method for interacting with computer systems and with the rapidly increasing use of mobile telephones, people can access these services at many different times and places. However, it is the view of Brewster (2008), that the provision of this functionality may be rendered useless if usability issues are not considered. Brewster (2008), considers that the use of TBI can have navigation problems with the menus being too long and difficult to navigate. Williams et al (1998), claim that the length of spoken menu prompts in TBI is a significiant problem with speech based menu designs. It is important, therefore, for these key points to be considered when undertaking the evaluation of the system. When undertaking the analysis and evaluation of the tests, therefore, emphasis needs to be placed on the accuracy of the system, a high error rate may not, necessarily, indicate that the speech that is output is mistaken but, moreover, that the end user cannot get the correct answer because there is too much speech to listen to. In addition, if the end user makes the incorrect selection and listens to some speech and then selects the correct one and listens to yet more speech, this would be total information overload and the end user may just give up. Further consideration also needed to be made as to whether speech as an output on its own is satisfactory as a tool to provide information, or whether it is more appropriately used to complement other forms of media.

5.5 The Case for User-Based Testing

Research suggests that only between 30% and 40% of information systems that are produced are ever successfully implemented and used successfully for the purpose for which they were designed. The development of such systems can be categorised into five main stages of the software lifecycle – analysis, design, testing, implementation and maintenance; the two main areas where weaknesses are more likely to occur are at the analysis and testing stages. At the analysis stage, which is the initial stage, it is critical to undertake a feasibility study that explores whether it is actually possible to create the system that has been asked for. Feasibility studies consider the following aspects:

- **Technical** – what technology in terms of telecommunications, hardware and software is available and compatible

- **Economic** – how much is the system going to cost to produce, train end users and maintain. Will the benefits of the system outweigh the financial outlay?

- **Legal** – what are the legal implications of introducing the system and what legislation is in force in the country.

- **Operational** – how would the new system fit in with current work practices and procedures.

- **Schedule** – how long will the new system take from inception to completion. Is it possible to produce the system in the given timescales.

Once a feasibility study has been undertaken, only then should the system go ahead with a thorough analysis of any exisiting systems.

The second area of the software lifecycle which is susceptible to weakness is the testing area. Gray and Salzman (1998), consider that although researchers and practitioners in HCI have been very interested in interface design, this interest has not been extended to the design of experiments and state further that reliable and valid guidance for the actual design of the interface depends on the results of the tests and that it is vital that researchers have an understanding of how small features of an experimental design can cast large shadows over the results and conclusions that can be drawn from them. Usability testing is an important strategy as part of software development but is often overlooked; there is little material on ethical issues in the HCI literature and designs of tests using human subjects require the utmost consideration of these.

5.6 Previous Evaluation Techniques Used by LLSTI

The Local Languge Speech Technology Initiative (LLSTI) was formed in 2001 and is based at the University of Bristol; the Director of the Institute is Dr Roger Tucker. The aim of this initiative is to try a new approach to the production of speech technology in local languages, particularly in the developing world. There are four languages that are being developed – Hindi, isiZulu, Ibibio and Kiswahili (Tucker and Shalonova, 2005). The resulting systems will be available through open source so that others can build on them creating an accessible community of interest in both the language and the technology.

Tucker and Shalonova (2005), state that in the developing world, where the vast majority of the population live, mobile telephone usage is growing at a phenomenal rate. Many users would value timely information on jobs, health issues, local market prices etc., but have little access to computers and know only their local or national language. Tucker and Shalonova (2005), see voice services and speech technology as an opportunity to address this need, but that they have to be developed both linguistically and culturally with different service models in order to succeed for these new users, who represent a completely different and new market for voice systems.

Prior to the development of the banana information system, an evaluation had already been undertaken of the isiZulu TTS system which did not prove to be entirely satisfactory, (Tucker and Shalonova, 2005). Tucker and Shalanova (2005) describe this testing procedure to be ad-hoc, stating that they had used one or more of the following approaches – but were not specific about which ones:

1) Developer testing. High level testing to eliminate major bugs in all text processing modules.

2) User tests. Tests that are given to native subjects asking them for vague comments.

Due to the significance of the impact that the successful implementation of the banana information system could make to the lives and economy of Kenya, the

LLSTI seeked to ensure that testing that was undertaken on this system would provide credible results.

Before creating usability tests, the author explored the procedures that had been used in this previous ad-hoc approach to usability testing, to establish whether the isiZulu system had actually been tested at all with any real end users, whether or not they came from the target group and, most importantly had ethical considerations had been given to the subjects, considering that they may have low literacy and ICT skills. It appears that the testing was carried out by Barnard and Davel (2004), in Africa who prepared a questionnaire to:

1) obtain biographic information of the evaluators (their age, home language, level of education;

2) test the subjects' understanding of the matter presented in this way;

3) query the subjects' subjective experience of the synthesised speech.

The strategy of giving subjects a written questionnaire to complete stating their level of education, without any ethical considerations to their reading and literacy levels may be inappropriate. During the testing procedure itself, the instructions were given by synthetic voice and then information clips were given. Barnard and Davel (2004), state that if it became evident during the testing procedure that they were not sufficiently literate to write down their answers, an experimenter completed the rest of the questionnaire on their behalf.

Barnard and Davel (2004), state that this system was evaluated by 23 subjects who had been canvassed in various ways ranging from personal contacts to a company-wide e-mail solicitation. This meant that the subjects included those people who were experienced in ICT as they were employed within organisations that had use of e-mail. In addition, these subjects were tested by using a web based application, not in a laboratory situation as their counterparts, who were employed outside of the companies. The author explores the theory that lack of control would lower reliability in the results of these tests.

Testing that was carried out on this software may provide results that lack credibility for the following reasons:

1) there is no evidence of any ethical considerations to the needs of the end users meaning that they could be caused unnecessary distress;

2) there was an inconsistent approach to the testing procedure. Some subjects took the tests in a laboratory (although there is no evidence to suggest that the tests were taken under controlled conditions), and some subjects took the tests via a web based application;

3) the subjects had not been selected without bias as e-mail to companies had been used which would automatically select end users who had experience of using ICT, even though the whole idea underlying this project was to be able to reach individuals who did not have use of computers.

Barnard and Davel (2004), concluded that the testing of the system had produced encouraging results, but also identified that the design of the comprehension tests for users of limited literacy requires particular attention.

5.7 Testing Procedure

In contrast to the isiZulu tests, the experiments for the banana information system was undertaken in full compliance with strict ethical guidelines and industry standards relating to undertaking tests with human subjects. Following a pilot test using two subjects, full testing took place using performance measures tests with ten subjects who had been invited from the target group of banana farmers in Kenya. On 26 May 2006, people from the farming community were informed about the tests using the information sheets that the author had produced and ten volunteers came forward to take part in the tests. In addition, to the ten volunteers two additional farmers tried out the system in a pilot test, using mobile phones for the purpose of testing the connectivity of the system and eliminate any errors in the testing procedure. At the time of the pilot tests, connectivity was satisfactory and no errors in the testing procedure were identified.

Unfortunately, however, connectivity was inconsistent. The actual tests took place on 1 June 2006, but unfortunately this time there was no connectivity for the mobile phone, so as this was the sixth attempt to undertake the tests (due to lack of connectivity), a back up plan was used where the sound files were played using a laptop computer. This meant that the subjects could still navigate the system by pressing buttons and answer the questionnaires, but it did raise the question of the robustness and the reliability of the Kenyan telecommunication system. The tests were created by the author, who sought guidance from NALEP and colleagues from the University of Nairobi on the structure of the tests. The tests were conducted in Kenya by the LLSTI under the supervision of the Kenyan Ministry of Agriculture and the University of Nairobi.

As the testing procedure was going to be carried out on very vulnerable subjects with poor levels of literacy, numeracy and ICT, utmost consideration was given to the application of ethical standards. The following table refers to the procedures developed by the author which where adhered to in Kenya to ensure that no distress was caused.

A transcriber and translator was available before, during and after the tests.

Table 5.1 Ethical Considerations and Applications to Practice

Ethical Consideration – Before the Test	**Application to Test**
(i) Have everything ready before the user turns up.	(i) The experimenter and transcriber/translator arrived early at the subjects farm to test the mobile telephones and the laptop and had all the necessary software, tasks, response booklets ready.
(ii) Emphasise that it is the system being tested, not the user.	(ii) The subjects were given this information in writing on their consent forms and this was also re-iterated verbally before the test.
(iii) Acknowledge that the software is new and untested, and may have problems.	(iii) The subjects were informed in writing and verbally before the start of the test, that the software has not been tested before and that there may be inherent problems with the software.
(iv) Let users know that they can stop at any time.	(iv) The subjects were given this information in writing on their consent forms and was also re-iterated verbally before the start of the test
(v) Explain any recording, keystroke logging, or other monitoring that is used.	(v) No audio or visual recording or keystroke logging was used. Subjects will be informed in writing on their consent forms that the experimenter will be monitoring the performance of the system throughout the test.
(vi) Tell the user that the test results will be kept completely confidential.	(vi) It was clarified to the subjects that all of the results will remain anonymous, used for the purposes of academic research only and will not be divulged to any employer or teacher. Subjects will be identified only be encoding for administration purposes by the experimenter as participant 1, participant 2, for example.

***(**vii) Make sure that you have answered all of the user's questions before proceeding.**	(vii)Subjects were given an opportunity to ask questions about the test and testing procedure before the start of the test.

Ethical Consideration – During the Test	**Application to Pilot Test**
(i) Try to give the user an early success experience.	(i) The questions on the task sheet was designed with the lowest potential user based error rates first
(ii) Hand out the test tasks one at a time.	(ii) The experimenter handed out the consent form, then the task sheet and finally the satisfaction questionnaire.
(iii) Keep a relaxed atmosphere in the test room, serve coffee and/or have breaks.	(iii) Refreshments were made available at the end of the test.
(iv) Avoid disruptions: close the door and post a sign on it. Disable telephone.	(iv) There is no main telephone.
***(v)* Never indicate in any way that the user is making mistakes or is too slow.**	(v) The experimenter did not interact with the subjects in any way during the test. They would only interact in circumstances where the participant showed signs of distress and wished to stop the test or was experiencing problems with the system.
(vi) Minimise the number of observers at the test.	(vi) The experimenter and transcriber/translator will be the only observers at the test.
(vii) Do not allow the user's management to observe the test.	(vii) The subjects' employers or teachers were not be involved in the conduct of the tests in any respect.
(viii) If necessary, have the experimenter stop the test if it becomes too unpleasant.	(viii) The experimenter would have stopped the test at any time if a) a participant had requested this or b) the experimenter considered that a participant was displaying signs of distress through the testing procedure.

<u>Ethical Considerations – After the Test</u>	<u>Application to Pilot Test</u>
(i) End by stating that the user has helped you find areas of improvement.	(i) The experimenter thanked the subjects for their invaluable assistance in testing the usability of the software.
(ii) Never report results in such a way that individual users can be identified.	(ii) The experimenter has only identified subjects only by methods of encoding – participant 1, participant 2, for example – for administration purposes only.
(iii) Only show videotapes outside the usability group with the user's permission.	(iii) The experimenter did not video record any aspect of the testing procedure.

Following consultations with NALEP, the LLSTI and colleagues from the University of Nairobi, the author carefully developed several documents that would be straightforward to complete for the subjects and decided that multiple choice options where subjects could just circle a box would be the most appropriate as this would eliminate the need for writing. As discussed previously, a translator and transcriber would be available to complete any forms for those who are illiterate.

The first document was a consent form which can be found in the Appendix. A translator was available to read the form to the subjects which asked for their consent, informed them that their participation was voluntary and that they were free to withdraw from the tests at any time and that it was the the computer program that was being tested, not them. The forms asked for the subject's age, gender, usual language of communication, other languages of communication, frequency of using the telephone and existing levels of computer literacy.

Tasks for the subjects to perform and a sheet on which to place their answers were then developed. Before deciding on the tasks themselves, considerable time was spent liasing with the colleagues stated above, who gave guidance on the most appropriate questions to ask for the tasks themselves and perusing the information system to ascertain the information that was available. The answers to all of these questions should be readily available from the system. Tasks were set to find out the minimum rainfall before irrigation becomes necessary; varieties of banana susceptible to Panama disease, planting distances; times for planting; and symptoms of pest and disease. It is accepted that using multiple choice questions is not the ideal in terms of obtaining results that do not have an element of inconclusion as some of the farmers may already know the answers, but it was decided that further testing could be undertaken outside of the target group at a later stage, but risks could not be taken on causing distress to the subjects in the target group by not giving them options to choose from. The task sheet can be found in the Appendix.

A questionnaire, which can be found in the Appendix, was also designed to give to the subjects after they had completed the tests to be able to analyse and evaluate the usability of the system. Again, considerable consideration was given to the design of the questionnaire. Within the field of HCI, it is accepted practice to use the SUMI method of psychometric tests but it was not considered to be appropriate for this group of subjects as it does not have the correct range of questions that would allow an accurate analysis and evaluation of this particular system. It was decided, therefore, to develop an alternative satisfaction questionnaire. The SUMI technique uses standard questions that are designed for use in the western world and the questions would not, necessarily, be understood by subjects in a developing country with limited or no previous computer experience and, therefore, subjects may just select a response at random which would result in unreliable data.

To be able to undertake an analysis and evaluation of the system, it was necessary to ask specific questions which were more appropriate for the subjects because they were more straightforward to understand as they related directly to the sytem. The questionnaire, therefore, asked the following: whether or not the subjects found the system easy to use, the language they chose, how easy it was to make a selection, whether they would use this system, clarity of the speech from the system, what gender of speech they would prefer and their normal sources of information.

Once the testing had been undertaken, the results of the tests were collated and converted into tabular format; the results of the tests were then analysed.

5.8 Profile of Subjects

Testing was undertaken at the farmers own farms using ten subjects of mixed gender aged between 18 and 41+ years as detailed later in this chapter. Although the size of this sample may appear small, Nielsen (1993), states that ten is the minimum number required for this type of usability test and, in consideration of the time constraints conducting these tests with a transcriber in a vulnerable target group, it was decided that this would be satisfactory. In Chapter Two, the evaluation methods of Zajicek and Jonsson (2006), were discussed where a small sample of twelve subjects were used to evaluate a system using speech technology.

5.9 Results of Test

This test asked the subjects to identify the correct answers for five questions from a multiple choice of three answers. All of the questions relate to banana farming, although this method of testing has an element of inconclusion because it was presumed that the subjects may already know the answers, it was necessary to make the questions multiple choice of this nature to allow for ease of testing and to have a common theme of interest for all subjects to answer. It was decided that in the event that very high levels of accuracy was experienced, further usability tests would be undertaken in the computer laboratory at Thames Valley University with subjects outside of the target group of end users who would not know the answers to these questions.

However, the outcome of the tests was in direct contrast with what was expected in terms of error rate. As the tasks asked for information on banana farming, it was assumed that the error rate would be particularly low as the end users may already know the answer and was prepared to undertake further tests outside of the user group to eliminate any inconclusion in this respect.

However, the results indicate an extremely high error rate as can be seen from the following table, with the overall mean average error rate at 44%.

It is important to note here that the 10% error rate continues to be used in these experiments, but the concept is different to the experiments conducted in Chapter Four. In the previous chapter, the error rate was calculated by the number of words but in these experiments the error rate is calculated by the number of incorrect answers to questions. The ultimate goal, of course, is to have 0% error rate, but 10% remains a useful benchmark.

The subjects were also asked to comment on learnability and satisfaction components of usability via a questionnaire. Although the results are included below, these must be accepted with reservations; we simply have no way of knowing whether this is the true opinion of the subjects or whether they were just being polite because they felt privileged to be testing the system. One thing, however, is evident – there is a discrepancy between perceived satisfaction with the system and its actual usability.

Table 5.1. Error Rates for Tasks

Subject ID	Task 1	Task 2	Task 3	Task 4	Task 5	No of Correct Answers	% Error Rate
K001	Incorrect	Correct	Correct	Correct	Correct	4	20
K002	Incorrect	Correct	Incorrect	Correct	Incorrect	2	60
K003	Incorrect	Correct	Incorrect	Correct	Incorrect	2	60
K004	Incorrect	Correct	Correct	Correct	Incorrect	3	40
K005	Incorrect	Correct	Incorrect	Correct	Correct	3	40
K006	Incorrect	Correct	Incorrect	Correct	Incorrect	2	60
K007	Correct	Correct	Correct	Incorrect	Incorrect	3	40
K008	Incorrect	Correct	Incorrect	Correct	Incorrect	2	60
K009	Incorrect	Correct	Correct	Correct	Incorrect	3	40
K010	Incorrect	Correct	Correct	Correct	Correct	4	20
No Correct	1	10	5	9	3		
Mean Average Errors %	90	0	50	10	70		44

Each set of responses to each task was analysed as follows:

Task 1. This asked the subjects to find out the minimum rainfall before irrigation becomes necessary. This first question was designed to give an early success experience, however, only 10% of the subjects got this answer correct. The correct answer is 1000mm, but to find out this information, the subject has to navigate through two options – the first one is to select the language that they would like to hear and the second one is the option for ideal climatic and soil conditions.

Task 2. This asked the subjects to find out which banana variety is highly susceptible to the Panama disease (Fusarium Wilt). Interestingly, even though the subjects have to navigate through four options – the first one is to select the language that they would like to hear, the second one is the option for pests and diseases, the third one is for panama disease and the fourth one is for information and symptoms of panama disease, 100% of the subjects got this answer correct. This result suggests that the subjects may have already known the answer to this question.

Task 3. This asked the subjects to find out which varieties should be planted at a spacing of 3 times 3 meters. 50% of the subjects got this answer correct. The subjects had to navigate through two options - the first one is to select the language that they would like to hear, the second one is planting and land preparation.

Task 4. This asked the subjects to find out when planting should be done. 90% of the subjects got this answer correct. The subjects had to navigate through two options – the first one is to select the language that they would like to hear, the second one is planting and land preparation.

Task 5. This asked the subjects to find out which pest or disease shows the symptom that the older leaves turn yellow, and collapse while still green at the base. 30% of the subjects got this answer correct. The subjects had to navigate through five options – the first one is to select the language that they would like to hear, the second one is to select pests and diseases, the third one is to select overview of symptoms, the fourth one is to select panama disease and the fifth one is to select information and symptoms of panama disease. The results of the tests are in contrast to the perceived usability by the subjects where 80% of the subjects stated that it was easy to tell which option to select from the menu, with the remaining 20% stating that they were not sure whether it was easy to tell which option to select from the menu or not. In addition, 90% of the subjects stated that they found the system easy to use, with the remaining 10% stating that they were not sure whether the system was easy to use or not. This perceived usability by the subjects must be taken with some reservations for reasons detailed in this chapter. These error rates are above the 10% suggested by Papamichalis (1987), and may render this system unusable. None of the tasks could be weighted with more significance than the others, because all of the information to be gleaned from the system was of equal relevance. Question two has 100% accuracy, but, however, it is not known whether the subjects knew the answer to this question beforehand. Although this seems likely due to the constrast with the other results, the results for task two will still be included in the analysis and evaluation. Further investigations of the profile of the subjects were undertaken to identify other factors that may have influenced these results.

Table 5.2. Ages of the Subjects

Subject ID	% Error Rate	Age
K001	20	18-30
K002	60	41+
K003	60	31-40
K004	40	41+
K005	40	31-40
K006	60	31-40
K007	40	41+
K008	60	41+
K009	40	41+
K010	20	41+
Mean Average Errors % Age 18-30	20.00	
Mean Average Errors % Age 31-40	53.33	
Mean Average Errors % Age 41+	43.33	

As we can see from the above table, the mean average error rate is low for the youngest subject who is in the 18-30 category, at 20%. However, this subject is the only one with this profile. For comparison, the only other subject who achieved 20% error rate was in the 41+ category, so without further evidence of tests on 18-30 year old subjects, a conclusion cannot be drawn that the subjects age affected the results necessarily.

In the 31- 40 category, the mean error rate was 53.33% and in the 41+ category, the mean error rate was 43.33%.

Further investigations were then undertaken with the gender of the subjects.

Table 5.3. Gender of the Subjects

Subject ID	% Error Rate	Gender
K001	20	Male
K002	60	Female
K003	60	Male
K004	40	Male
K005	40	Male
K006	60	Female
K007	40	Male
K008	60	Male
K009	40	Female
K010	20	Male
Mean Average Errors % Male	40.00	
Mean Average Errors % Female	53.33	

The results of these tests show that the mean average error rates for females are higher at 53.33% than those for males at 40%. These results suggest that the gender of the subjects, may affect the error rate, but that it is still unacceptably high. Next, the frequency of previous use of using the telephone was analysed. An interesting response from the questionnaire was that 70% of subjects stated that they would prefer to hear synthesised speech in the same gender, with 30% expressing a preference not to hear speech in the same gender.

Table 5.4. Frequency of Using the Telephone

Subject ID	% Error Rate	Telephone Frequency
K001	20	Weekly
K002	60	Daily
K003	60	Weekly
K004	40	Weekly
K005	40	Daily
K006	60	More rarely than weekly
K007	40	Weekly
K008	60	More rarely than weekly
K009	40	Daily
K010	20	Weekly
Mean Average Errors % Daily use of telephone	46.67	
Mean Average Errors % Weekly use of telephone	36.00	
Mean Average Errors % More rarely than weekly use of telephone	60.00	

These results show that the subjects who used the telephone at least weekly, had a lower mean average error rate than those subjects who had very little experience of using the telephone, more rarely than once per week. This suggests that more frequent use of the telephone may affect the accuracy of the system. However, as per the analysis of the previous tests, this did not bring any of the subjects into an acceptable error range. Next the frequency of using a computer was analysed.

Table 5.5. Frequency of Using a Computer

Subject ID	% Error Rate	Computer Use Frequency
K001	20	Often
K002	60	Never
K003	60	A little
K004	40	Never
K005	40	Never
K006	60	Never
K007	40	Never
K008	60	Never
K009	40	Never
K010	20	Never
Mean Average Errors % Often use of computer	20.00	
Mean Average Errors % A little use of computer	60.00	
Mean Average Errors % Have never used a computer	45.00	

These results show that for the subject that had often used a computer before, there was a mean average error rate of 20%. However, as earlier identified this subject is in a younger age group that the other subjects, and is the only one in this category; he is more likely to have used a computer at school than his elder peers and therefore, this result is not surprising. However, the mean average of subjects who had used a computer a little before was 60% and higher than those who had never used a computer before at 45.00%. As the usability tests were undertaken pressing buttons on a computer keyboard, these results are inconclusive. The next thing to be analysed was the language used by the subjects and the language that they listened to the speech in.

Table 5.6. First Language of Subjects and Culture of Synthesised Speech

Subject ID	% Error Rate	Languages	Synthesised Speech
K001	20	English and Kiswahili	English
K002	60	English and Kiswahili	English
K003	60	English	English
K004	40	Kiswahili and English	Kiswahili
K005	40	English and Kiswahili	English
K006	60	Kiswahili and English	English
K007	40	Kiswahili	English
K008	60	English and Kiswahili	Kiswahili
K009	40	English and Kiswahili	English
K010	20	Kiswahili	Kiswahili
Mean Average Errors % Kiswahili synthesised speech	33.33		
Mean Average Errors % English synthesised speech	48.57		

As we can see from the above results, the mean average error rates when subjects heard speech that sounded as though it came from the same culture as their own is 33.33% compared with 48.57% when subjects heard speech that sounded as though it was from a different culture to their own. The subjects who heard the Kiswahili speech, heard a native speaker, whereas the subjects who heard the English speech did not hear a native speaker. These findings support the previous findings in the usability tests as detailed in Chapter Four, that the culture of the speech affects the error rates of the system for end users. However, at 33.3%, the error rate is still unacceptably high. There is a discrepancy between the high error rates and the perceived usability of the system. 70% of the subjects stated via the questionnaire that they found the speech clear and easy to understand, with the remaining 30% of the subjects stating that they were not sure whether they found the speech clear and easy to understand or not. This response, however, is not reflected in the error rates of the practical tests. The responses to the questionnaire, therefore, must be taken with reservation as we simply have no way of knowing whether the subjects were just so excited at being given the opportunity to use this system that they were just being polite or not. Certainly, the responses to the questionnaire were in contrast to verbal responses given afterwards to the developers when the subjects commented that they found the Kiswahili speech very difficult to understand because it used formal grammar which the banana farmers did not use in this particular part of Kenya. They also commented that the English speech was also very difficult to understand because it used an English accent, which again they were unfamiliar with. The developers played a test file sample of the prototype English voice that they are in the process of developing in English with a native Kenyan accent. 100% of the subjects expressed a preference for this voice, stating that it was much clearer for them to understand than the existing two voices.

Winschies and Paterson (2004), undertook a thorough analysis of usability evaluations that were undertaken in Namibia and revealed an unknown challenge in cross-cultural usability engineering; namely that the implicit western understanding of usability and its associated assumptions lead to locally inappropriate usability evaluation. For example, current usability testing methods are based on the assumption that an effective and efficient task completion correlates with user satisfaction (Winschies and Paterson, 2004). However, the findings of Winschies and Paterson (2004), imply that user satisfaction in the Namibian context is independent of task-solving success.

In consideration of the above findings, Winschies (2007), suggest a contextual redefinition of usability and determination of valid methods and has undertaken a series of investigative sessions with different Namibian user groups. Firstly, subjects were asked to brainstorm on associative and related terms/concepts for the word usability in general. Secondly, subjects elaborated on general characteristics of a good working environment. Thirdly, subjects select only the appropriate terms from the two previously produced lists which should apply to the software system that they will work with, to be considered usable. Terms that were named by more than one group were: easy, save, comfort, specific, reliable, pace, goal oriented and conducive. None of the groups mentioned terms commonly associated with usability such as speed, learnable, memorable and error rates (Winschies and Fendler, 2007).

Although the error rates are unacceptably high, the developers were encouraged by the response from the subjects to the system. 100% of the subjects stated that they would prefer to use this system to other methods of gaining information, which for 90% of the subjects is by liaison with the agricultural extension workers. However, as stated earlier we cannot be sure of the accuracy of the satisfaction questionnaire due to the politeness of the subjects and possible misinterpretation of the satisfaction questionnaire responses. In addition to these points, Dunckley et al (2007), state that subjects have high expectations but that a key issue is that changes in practice would need to be linked with demonstrated success and usefulness. Dunckley et al (2007), also state that picking out single members of the group for special treatment could be problematic as there have been instances of privileged farmers in Kambu being subjected to curses and witchcraft to which unexplained or unexpected events are often associated to.

It appears that original concerns about the quantity of speech to be listened to may also be a contributory factor to this unacceptably high error rate, in addition to the cultural aspects of the speech and therefore the author conducted an evaluation of the system herself which confirmed this. The amount of options and speech that has to be heard before the end user can find the information that they are seeking is too high and further development of this system is necessary in this respect before it can be used.

Unfortunately, the unacceptably high error rate from this system renders it unusable at the present time and that further investigation of connectivity and re-design of the system is necessary, to incorporate easier to navigate menus and speech that sounds as though it is in the same culture as the end user.

It was considered by the LLSTI, at the time, that the major reason why the majority of the subjects could not find all of the correct responses was that they could not perceive the spoken output produced by the text to speech system as they only had the opportunity to listen to the speech output once and may have absorbed more if they could have listened to the speech output twice for each menu. However, it is not the aim of this system to listen to the menus more than once for the end users to find the information they need. This would be a totally unmanageable task for end users with SEN such as low levels of literacy found in this target group. It is agreed, however, that the accuracy was increased when the subjects heard speech that sounded as though it was from the same culture as the subjects and that consideration needs to be given to modify the system including the introduction of a Kenyan native accent.

Tests are there to identify potential weaknesses in systems so that they can be put right before the system is live, so the developers can be positive that they can have the opportunity to undertake further research into overcoming connectivity problems. Through the undertaking of the tests, a technical weakness was identified in the communication technology available in Kenya; the developers are now researching further into possible approaches to overcome this as there were significant problems with the testing with at least five attempts being made to conduct the tests on the system over a four week period. Poor telecommunications meant connectivity was extremely unreliable. The telephone system in Kenya is unreliable and although the government is making attempts to modernise, one company has had the monopoly of supplying the service that has mainly been directed for business use. Trunks are primarily microwave radion relay with business data being commonly transferred by a very small apeture terminal (VSAT) system.

Statistics for 2003 state that there were 328,400 main line telephones in use and 1,590,900 mobile cellular telephones with 8.325 Internet hosts and 400,000 Internet users, (Central Intelligence Agency, 2006).

It transpired that weekends and late evenings would be the only time that end users would be able to connect to this system and this was identified as being a key weakness which would greatly restrict the accessibility and acceptance of the system by the Kenyan farming community who were already inexperienced in using a TBI.

Therefore, there were three main issues that were of major concern – the unacceptably high error rates which were caused by too much speech to listen to at one time; the speech not accurately reflecting their accents and; the unacceptably poor connectivity which was caused by poor telecommunications system within the country.

This system would be challenging to use for most end users but for end users from this target group who may have SEN and will therefore be unable to make notes as they hear the speech to remind them later, there is just too much speech to listen to. In addition, once the end user has made a selection from the initial menu, they then have to hear a minimum of two minutes synthesised speech for each option to get the information they are seeking. If the end user has made an incorrect selection from the initial menu, they have to wait until the speech for that section has finished before they can re-select from the main menu all over again. The whole process can take around ten minutes for the end user to get the information they require and this fact, combined with shorter concentration spans from this group of end users, is a contributing factor to the unacceptably high error rates.

After the analysis and evaluation of the tests, the LLSTI presented the system to NALEP who commented that they thought a system like this would fill a recognised information gap and increase the demand and interest for extension services and were convinced that this technology is important. However they also observed that the English system was difficult to comprehend and would be more comprehensible if it had a Kenyan English accent. NALEP were also presented with the sound file sample of the prototype English voice in a Kenyan accent and they were all agreed that this was much improved in terms of clarity and would be a far preferred option.

In addition, to presenting this system to NALEP, as 75% of the visually impaired in Kenya are also farmers, the LLSTI also presented the system to the Kenya Society for the Blind (KSB). One of the members of the society, who is totally blind and works as a computer teacher, tried out the system. His comments were that an improvement to the system would be to have one voice reading out the actual information and another voice reading the menus, which would make the system more interesting to listen to. Interestingly, he also considered that having musical tones, which is an area of research that Brewster (2008), is undertaking as discussed in Chapter Two, would also reduce the monotony of the system. Notwithstanding the unacceptably high error rates, the LLSTI decided to implement the system in the farming community in Kenya on 24 July 2006. On 14 May 2007, the system was withdrawn, described as being no longer operational and awaiting further development.

The LLSTI is now continuing with the development of English language speech in the native culture of the end users, in addition to simpifying the menu structure. Potential uses for integrating this system with a graphical user interface which could be used at local libraries so that the farmers can see images in addition to hearing the speech, even if they cannot read the words on the screen, are also being explored. There is a definite need to impart information to the Kenyan farming community and, therefore, any information system that the developers can produce that will help farmers is worth further exploration, providing that it is usable.

It was originally intened to undertake further usability tests at Thames Valley University, but until such time as the interface for the system is amended, further tests are not necessary.

Hypothesis Testing

One of the main purposes of the tests was to test the following hypothesis:

Hypothesis (c) – *There is no difference in error rate when the language of the speech is representative of the culture of the subjects..*

The results of the tests reject this hypothesis with a mean average error rate of 33.33% when the language of the speech is in the subjects' culture and 48.57% when speech is in a different culture.

5.10 Chapter Summary

This chapter details the usability tests that were conducted in Kenya on an information system that had been recently created for banana farmers to help them find key information on growing their crops. The mean average error rate for the system is unacceptably high at 44%, which renders this system inefficient and unsuable at the current time. The results of the tests also suggested that the error rate was lower when speech that sounded as though it was in the same culture as the subjects was heard; rejecting the hypothesis *There is no difference in error rate when the language of the speech is representative of the culture of the subjects* supporting the findings of the previous tests in Chapter Four. After undertaking an evaluation, it was concluded that the amount of speech that the end users have to listen to is a contributing factor, in addition to the culture of speech, and therefore the developer of the system is considering modifications to address this. Subjects stated that they preferred to use this technology rather than their existing methods of gaining information and, although it is disappointing for the developers that the system is no longer operational due to usability issues, they now have the opportunity to investigate further whether connectivity in Keyna is robust enough to support such a system, incorporation of accents and the possibility of incorporating a graphical user interface. The findings of the tests are discussed further in Chapter Seven.

The next chapter considers multi-modal approaches and explores whether TTS technology as a single user interface is the most appropriate approach for end users with SEN, or whether usability of such systems are improved further when multi-modal interfaces are used in addition to considerations of the culture of the speech output.

Chapter Six - Prototype Modelling and Testing

6.1 Chapter Introduction

In Chapter Four and Chapter Five, details have been given on tests that have been undertaken to test the usability of TTS software. In Chapter Four, tests were performed using generic TTS software in controlled laboratory conditions using subjects from Indian and English cultures, comparing the results to Hofstede's cultural model. In contrast, Chapter Five focussed on tests that were performed in Kenya on subjects using a TBI. The findings of the tests so far suggest that the error rate is lower when end users hear speech that sounds as though it is from the same culture as their own.

However, what is not clear from the results of the tests is whether speech as a single interface is the most appropriate for a teaching and learning tool for end users with SEN. The banana information system that has been evaluated in Chapter Five used speech solely as an interface but the findings of the tests deemed it to be unusable. The interpretation made is that the reason for this is that, in addition to the speech not accurately preserving the native culture of the subjects, another contributing factor was the vast amount of speech that has to be heard in navigating the menu hierarchy of the system before the information can be found. In consideration of this, the developers of the system are investigating whether or not it is possible to simplify the amount of speech and also whether it would be more appropriate to add a graphical interface to use with a computer rather than a telephone.

This chapter explores the concept that a multi-modal approach may increase usability for end users with SEN. To find out what is used in existing software for SEN, interviews were undertaken with a developer, Norman Hore, of an educational system which is now being used with SEN within mainstream education and within young offenders institutions, where there are high incidents of SEN. This system, Rapid English, has received government funding for its development and its aim is to improve literacy skills by using a multi modal approach incorporating TTS technology and a graphical user interface. Visits were made to Reading Young Offenders Institute and observations were undertaken of the system in use in addition to undertaking a comparative evaluation.

In addition, the ACE Centre in Oxford which focuses on the use of technology with the communication and educational needs of people with physical and communication difficulties, was visited. An interview with the assessment officer, Rachel Moore was undertaken where a demonstration was given of the technology that the centre recommends, Clicker 5. In addition a comparative evaluation of this system was undertaken.

In this chapter, both of the above systems are compared and contrasted to identify similarities and differences. Both systems use multi-modal interfaces, but neither of the software was developed to consider cultural aspects. Using the results of the evaluation of these systems, plus the findings from her research as detailed in previous chapters, a prototype model was created and tested using subjects from different cultures, who have SEN. The purpose of the experiments is to test hypothesis d) *There is no difference in learnability when the accent of the speech is representative of the native culture of the subjects* and hypothesis e) *There is no difference in satisfaction when the accent of the speech is representative of the native culture of the subject.* The results are analysed and evaluated using ANOVAs.

6.2 Case Study One

The first teaching and learning tool to be researched incorporating TTS technology as part of a multi-modal approach is Clicker Version 5. This was in liaison with the ACE Centre, which is a nationwide organisation focusing on the use of technology with the communication and educational needs of people with physical and communication difficulties; their needs may be complex and profound. The ACE centre works as follows - when people are introduced to the ACE Centre, an assessment is carried out on their individual needs by a trained assessor; the disabilities are wide ranging but are typically severe for example, not having use of their hands and needing to use switches which are activated by moving their head from side to side in a wheelchair to activate buttons to select items. This is a functionality that is incorporated into the educational systems that they use. Some of the young people are unable to speak. For end users who cannot speak or their language is not very well developed, hearing speech as an output is of the utmost importance for them to help them develop their own speech. Cerebal palsy and autism are disabilities that are widely experienced within ACE. The centre fully supports the young person, and their family, in using the technology that they provide, this includes training and regular ongoing assessments. The technology is taken home to use, but the training and assessments take place at the centre.

Supporting end users who are totally blind is extremely challenging for the ACE centre as they learn in a specific way that is different to the way sighted end users do. Using audio information alone requires very specific training and these young people are referred to other providers (Moore, 2006).

People up to the age of eighteen years are supported at ACE, with the overwhelming majority having complex educational needs. The ACE centre has researched and developed technology to support the end users and base their systems on the principle that the highest level of satisfaction is gained when a minimum amount of input produces a maximum level of output. To achieve this goal, text to speech plays a significant role; and for end users who are deaf, the uses of symbols are also extremely relevant. Clicker version 5 is a system developed by Cricksoft and is the software program that is the most highly used and recommended by the ACE centre (Moore, 2006).

Speech as an output is a main feature of Clicker and uses speech technology where digitised speech, such as a recording of words which will not change – this is used for auditory prompts giving the end user instructions, for example. Speech that has been concatenated from databases of speech is also used within the system due to the unpredictability of what the end user will want to hear. Research at the centre has suggested that the end users have difficulty in understanding the concatenated speech and that digitised speech is much preferred. Also the software developers are considering incorporating this into new versions of the software as much as possible (Moore, 2006).

Using Clicker, the end user can enter text via a keyboard or switches which is then displayed on the screen. For auditory prompts, using digitised speech, the end user cannot express any preferences in the speech but, in contrast, for the concatenative speech, the end user can set preferences of speech from the following options:

- Each letter can be spoken as it is entered into the system
- Each word can be spoken when it is added into the system
- Each sentence can be spoken when it has been punctuated
- Each word can be highlighted when they are spoken

There are also options for the type of speaker for concatenative speech which can be classified as follows:

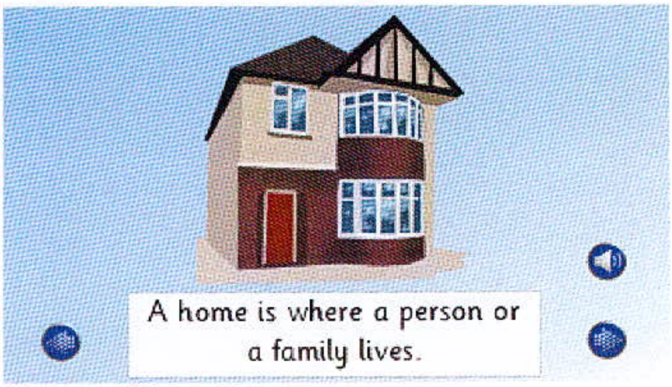
- Adult female, British English

- Adult male, British English

- Adult female, American English

- Adult male, American English

Figure 6.1 The Clicker Graphical User Interface



Source: Clicker 5

The above screen shot shows the simplistic GUI of this software. Clicker uses a design style of a graphical user interface consisting of bright colours with clear text about the image and navigational buttons, to complement the speech output. ACE considers that using speech on its own is not advisable, unless the end user is totally blind due to the shorter capacity of audio memory (Moore, 2006). In addition, the end users enjoy looking at the images which helps to put the speech in context. The research undertaken by the ACE Centre suggests that if the end users enjoy using the system, they will look forward to their sessions with it, whereas simply listening to speech alone becomes mundane and monotonous (Moore, 2006).

Moore (2006), states that research undertaken at the centre also suggests that the parents and carers of the end users have requested child like speech but this has not been supported by the young people themselves, although they have expressed preferences for speech that sounds as though it is from the same gender as the end user.

The ACE Centre has found that Clicker is an extremely effective teaching and learning tool supporting the educational needs of profoundly disabled end users with the use of speech technology being a significant component. The end users come from multicultural backgrounds with often severe speech and hearing impediments. As the end users hear speech on a daily basis in the communities they live in their accent, having speech output from the systems in their accent could enhance their teaching and learning if they have a clearer understanding of what is being said from the system; this would also help develop their own speech (Moore, 2006).

6.3 <u>Computer Assisted Learning and Reading</u>

Davidson (1994), states that computers are beneficial in the teaching of reading as computers are capable of giving large amounts of consistent practice and there is evidence to show that the more practice one has at a skill the more proficient one becomes. Anderson et al (1979), conclude that with respect to reading, the time on task is an important factor in learning; computers can provide such an opportunity.

Dowhower (1979), showed that learners who were given practice in re-reading passages doubled their rate of reading and increased their level of accuracy to over 97%. This effect was transferred to new passages, although at a reduced level. Reinking (1986), found that there were similar effects for adults and found that time spent on reading can be even more beneficial than reading instruction. Schaudt (1987), states that the use of a direct instruction approach to the teaching of reading has produced greater gains on reading tests than less structured methods because it increases the academic learning time, which produces gains in achievement.

Lewin (1998), states that the use of talking book software which replicates real books but with additional features such as computer generated words, phrases and sentence pronunciations, has the potential to support current classroom practice. They help with word recognition, enabling opportunities for engaging in independent practice (Lewin, 1998). The use of sound and graphics, harnessing motivational aspects and the ability to use the software in a variety of ways all contribute to the success of computer assisted learning as a means of providing supplementary practice in reading instruction (Reiking and Bridwell-Bowles, 1991).

Previous research has demonstrated that computer assisted reading software can make a valuable contribution to the teaching of reading (Lewin, 1998). Wise et al (1989), compared the effects of whole word, syllabic and sub-syllabic word pronunciations at the level of onset and rhyme with children experiencing difficulties in learning to read. They found that each condition improved whole word recognition, and that segmented pronunciations also improved phonological coding skills. Beginner readers who read text with computer assistance improved as much as a group who were reading to an adult, (Reitsma, 1988) even though they made little use of the computer generated speech.

6.4 Case Study Two

The second of the two teaching and learning tools that have been researched is called Rapid English which is a system that has been designed and is being used within mainstream education and within young offenders education. The use of speech as an output is also a significant component of this system.
Rapid English is based on the Pareto principle that states that, in any field of activity, 80% of return comes from just 20% of effort. When applying the same rigour for communicating in English, where 80% of communication comes from just 20% of the language, by focussing on the 20% of the language that has a high frequency of use, learners meet and practise language that has a high communicative value (Whitehead and Hore, 2007).

When first starting to use Rapid English, written samples of learners handwriting are scanned into the system, the program analyses the problems of each learner individually and presents the solutions one at a time in a clear consistent manner (Whitehead and Hore, 2007). During the course, more samples of handwriting are scanned into the system to compare progress.

Whitehead and Hore (2007), state that many learners have already developed strategies to avoid reading or speaking in public. The pronunciation system using TTS technology, allows them to make progress which they can hear for themselves and this gives them increased control and confidence (Whitehead and Hore, 2007). They rapidly find that they can apply this to any text that they read and that they can read words in context as well as pronounce longer words which were previously inaccessible to them. Whitehead and Hore (2007), state that this not only boosts their confidence, but encourages them to re-engage with learning.

According to Whitehead and Hore (2007), success that comes with such a system that incorporates methodology with technology, means breaking a cycle of failure, boosting learners' confidence and encouraging them to re-engage with learning (Whitehead and Hore, 2007). The developer of this system has been teaching English in mainstream schools for many years and, more recently, has been working with lower achievers.

Nationally, frustration has been experienced by teachers of English at the very low achievement the students make as reflected in the SATs results published by the government. Hore (2006), considers that much of the theory that was being taught within schools in English was not needed for the actual SATs tests and he subsequently undertook significant research to support this theory. Previous SATs examination papers were analysed and evaluated at Key Stage Three together with findings from thirty nine different countries; the research suggested that the mistakes made within all countries were the same and that teaching was focusing on what the teacher perceived was important, which was not always relevant to the subject matter in the tests themselves (Hore, 2006).

Following his research, Hore (2006), trialled his system in a mainstream school with students who were below the average ability and were currently underachieving in English. The system assumes very little prior knowledge and experience of being able to read and write. The system is available in different modules, depending on what level the student is working at, and uses different topic areas – for example equipment or places – to gain the students' initial interest in using the program. The end user can enter text via the keyboard and select text and images from the screen which are then output as speech. The type of speech used is non-concatenative and is taken from a corpus of digitised speech that has been input by the developer himself; the speech is always predictable. It is considered extremely important that the students can hear how the words are spoken so that they can identify the written word, and also to be able to identify how to pronounce words correctly (Hore, 2006).

The results of the trials of this teaching and learning tool are as follows:

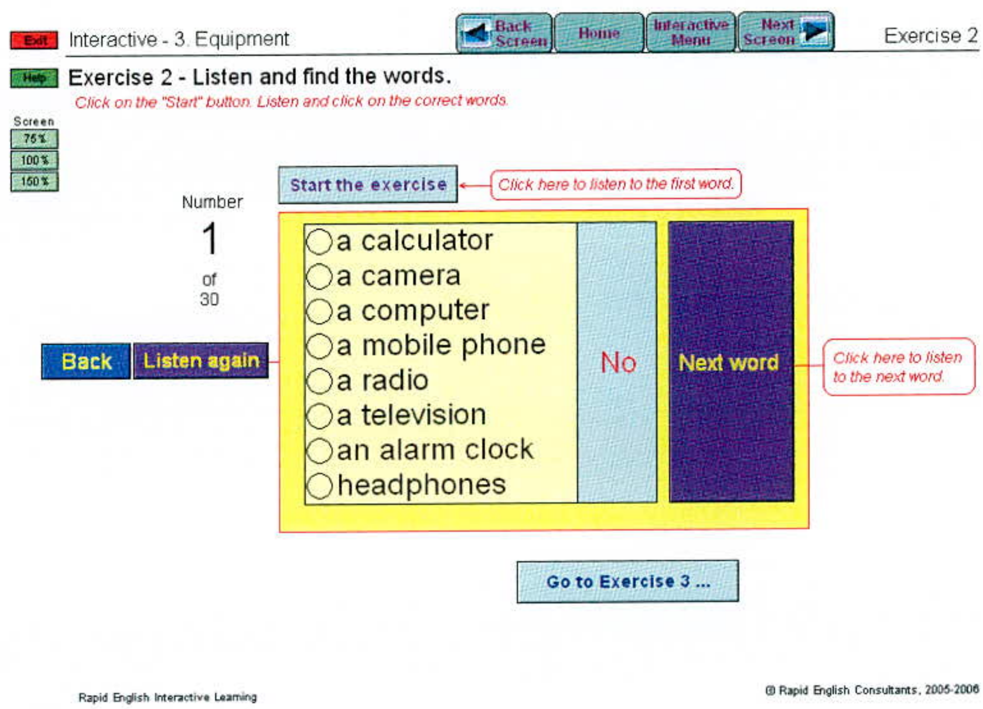
By the age of fourteen years (Year 9) the government expect students to achieve level 5 in English SATS, this demonstrates that they possess an acceptable level of English for their age. If they achieve level 6, they are excelling and if they achieve lower than level 5 they possess English language skills that are lower than fourteen years.

Hore (2006) states that in 2002 – 2003, Rapid English was used in one class of students who were in the bottom stream for English and were below average ability; a parallel group of students were taught using conventional teaching methods. 64% of students who used the Rapid English system achieved level 5 in the SATs examinations, compared with just 35% achieving level 5 who did not use the system. The same trials were undertaken in 2004-2005 with 64% of students who used the Rapid English system achieving level 5, compared with just 31% of students achieving level 5 who did not use the system. These findings suggest that using the Rapid English system where students can hear speech as an output has a significant improvement on their success.

Hore (2006), re-iterated the importance of having a multi-modal approach to make the system enjoyable and easier to use, so that the learners would want to return to use it frequently, expediting the progress that they make and as can be seen from the following screen, there are, again, clear navigation buttons, clear text and appropriate use of bright colours to complement the use of speech as an output.

Brewster (1994) defines a multi-modal interface as one that presents information in different sensory modalities, specifically visual and auditory. Brewster (1994) states that all computers have the ability to present information via different media such as graphics, text, video and sound but they are not, however, all multi-modal. Most of the media used by computer systems present information to the visual system (Brewster, 1994), with very few systems making much of their capacity to produce sound.

Figure 6.2 Rapid English Interface



Source: Rapid English

As can be seen from the above image, the Rapid English interface consists of a simplistic GUI in bright colours with clear text and navigational buttons which complement the speech output. This has some commonality with the Clicker 5 system which also incorporates the same features.

6.5 Comparison and Contrast of Clicker and Rapid English

Following the interviews, demonstrations and observations of the two systems, a comparative evalution of both systems was undertaken and is detailed in the following table:

Table 6.1 Comparison and Contrast of Clicker and RapidEnglish

Feature	Clicker	RapidEnglish
End Users	People with profound physical and learning disabilities from multi-cultural environments	People with lower than average ability in English Language, who may also have other physical and learning disabilities from multi-cultural environments
Uses of System as a Teaching and Learning Tool	Used to develop speech, communication and educational advancement in all subject areas	Used to develop English Language skills in reading, writing, listening and speaking
Type of Text to Speech Technology	Digitised speech for auditory prompts, otherwise concatenative speech	Digitised speech
User Preferences	Can choose from American English male or female and English English male or female	British English male only
Comparative Evaluation	The author did not experience any difficulty in understanding the auditory prompts, but could not always repeat the exact speech that had been output from the system	The author did not experience any difficulty in understanding the speech

The results of the comparative evaluations, suggest that the use of speech as an output is a component of these systems that complements their graphical user interfaces and makes a significant difference to the educational

advancement of students who have educational needs that are special in some way including those with profound physical and learning disabilities.
Neither of the systems, however, consider cultural issues of the end user and, subsequently, the techniques and philosophy incoporated in these systems will be included into a teaching and learning tool in the form of a talking book.

The prototype talking book will incorporate a multi-modal interface that complements speech output with a GUI and is an approach successfully used by Clicker and Rapid English. The talking book is going to be created using a generic software presentation program which is a method that can be adopted by teachers to use as a teaching and learning tool for learners with SEN without the need for buying additional software. By using this method, speech that is either digitised or concatenative can be incorporated. Therefore, such talking books can be produced to cater for various subject matters and at appropriate levels of detail for the end user.

Initially, the talking book will be developed with only English language in an English accent. The reason for this is that usability tests are going to be undertaken on the system using subjects from different cultures who have SEN, to evaluate whether a multi-modal approach is preferred to a singular interface of speech. At the present time, speech in the English language is only readily available in English, American and Indian accents; it is not available in speech that preserves the native culture of all of the subjects. Therefore, as the system is going to be tested on as wide a range of multicultural end users as possible, all speech output will be in English accent to eliminate an element of inconclusion in the test results. The tests that are performed on the talking book test the learnability and satisfaction components, they do not test error rates. The reason for this is because the extensive tests to evaluate these areas have already been undertaken as detailed in Chapter Four and Chapter Five which suggests that error rates are reduced if the end users hear speech that sounds as though it is in the same culture as their own. In addition, as the end users can see the words on the screen with this multi-modal approach, the testing of error rates would produce inconclusive results. In these tests, therefore, the focus on learnability and satisfaction. However, if there is any significance in these areas of usability due to cultural effects, then it would be expedient to proceed to introduce English speech in other accents when these become available.

6.6 The Talking Book

The first consideration for the talking book was the content. This book is going to be tested on end users aged 16-20 from multicultural backgrounds, all with SEN. It was decided that, the end users would have differing hobbies and interests and she wanted to make the book interesting to all of the end users. If, for example, a book about football was created this may appeal to some of the end users but bore others. Therefore, it was decided to create the book on the subject of the capital of England because all of the subjects who will take part in the test are studying in England and this subject will incorporate commonality for all subjects. All of the words that are included are common use words in the English vocabulary.

In consideration of the previous interfaces evaluated in this thesis and the recommendation of Hore (2006), and Moore (2006), to keep the amount of options to a very minimum, the talking book incorporates a basic GUI which consists of four pages with each page explaining an aspect of London. In addition, the amount of speech that is heard at a time is also kept to a minimum to help absorption of the information. When the page of the book is opened the speech will automatically be heard which is in English female speech. The GUI incorporates bright colours for the images, with clear navigational buttons which the end user can click on to go back or forward a page through the book.

Figure 6.3 Talking Book Page 1

This is the first page of the talking book. The only text that is heard is the title "All About London" which will be spoken as the page is opened.

The image is of the guards outside Buckingham Palace which the subjects can look at and discuss before clicking on the navigation button to go to the next page of the book. There is only one navigational button on this page to reduce any confusion on which icon to click on.

If the subject returns to this page later on in the book, the speech will automatically be repeated.

Image source: www.greenwichmeantime.com

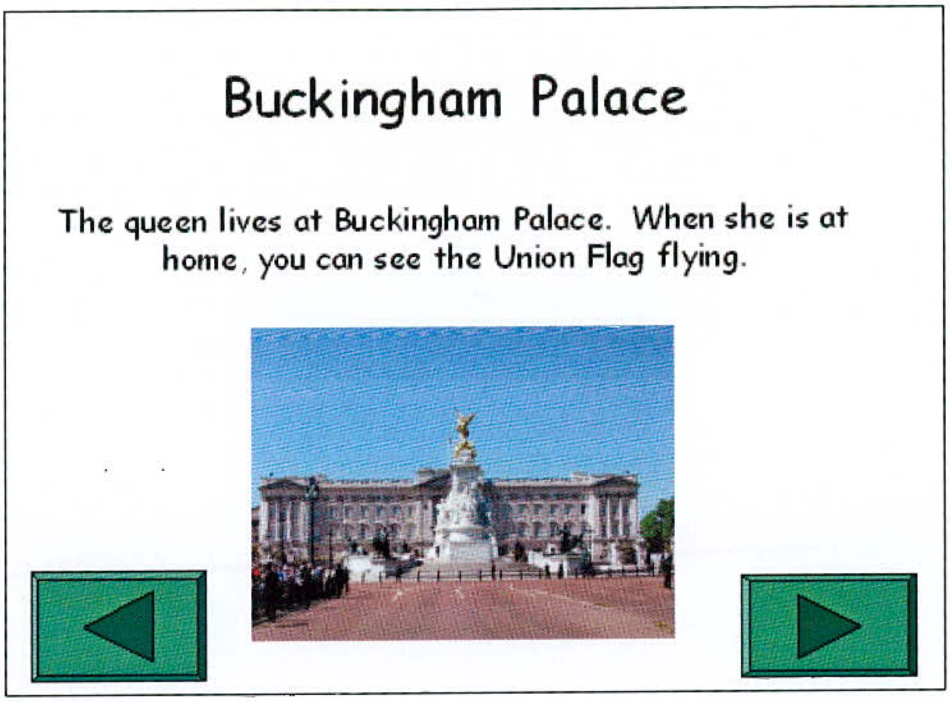


Figure 6.4 Talking Book Page 2

This is the second page of the talking book. When the page is opened, the title "Buckingham Palace" will be heard automatically followed by "The queen lives at Buckingham Palace. When she is at home, you can see the Union Flag flying". The image is of Buckingham Palace which the subjects can then look at and discuss before clicking on the navigational button to go on to the next page. Alternatively the subjects could click on the navigational button to go back to the previous page and hear the first page again.

The graphical interface retains its simplistic design, ensuring clarity of use for end users with special educational needs.

Image source: www.bcsu.net

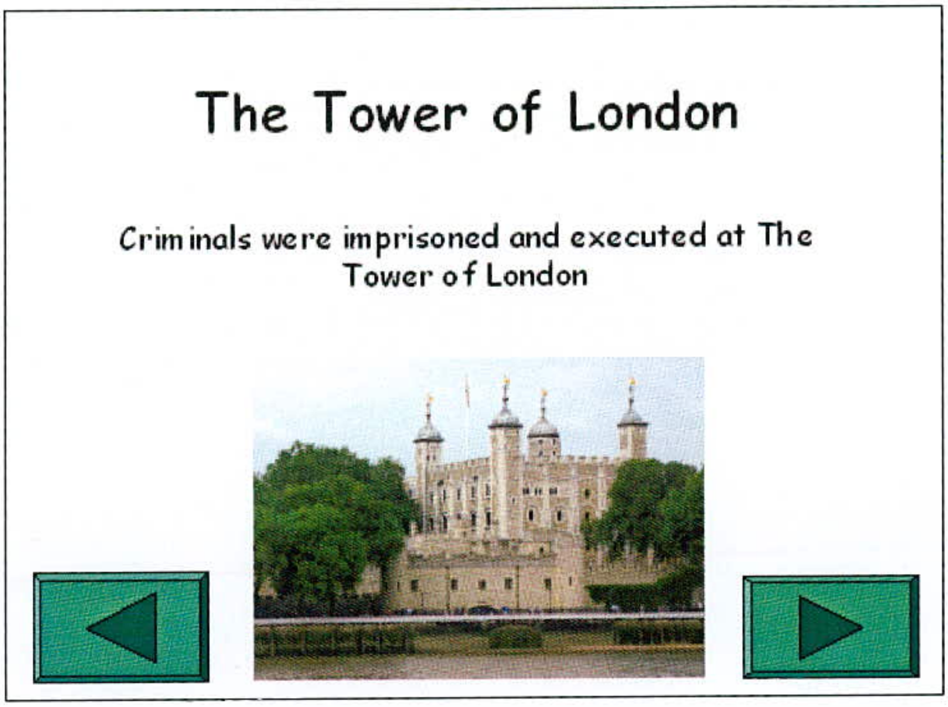


Figure 6.5 Talking Book Page 3

This is the third page of the talking book. When the page is opened the title "The Tower of London" and the text "Criminals were imprisoned and executed at The Tower of London" will be heard. The image is of The Tower of London which the subjects can then look at and discuss before clicking on the navigational button to go on to the next page. Alternatively the subjects could click on the navigational button to go back to the previous page on Buckingham Palace and hear that page again. The author considered adding an icon so that speech could be repeated, but decided to retain the simplistic design of the interface, ensuring clarity of use for end users with special educational needs which enables them to hear speech again if they re-visit the page.

Image source: www.historic.uk.com

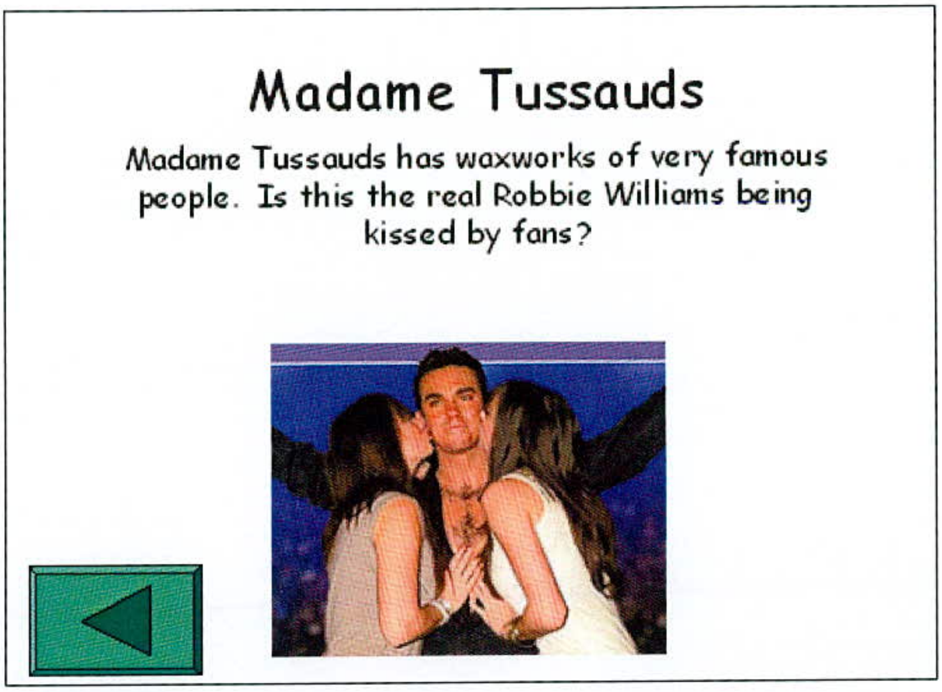


Figure 6.6. Talking Book Page 4

This is the fourth and final page of the talking book.

When the page is opened the title "Madame Tussauds" and the text "Madame Tussauds has waxworks of very famous people. Is this the real Robbie Williams being kissed by fans?" will be heard.

The image is of popstar Robbie Williams whom it was considered would be easily identified by the age range of the subjects who will be testing this system. The interface retains its simplicity with one icon which the end user can click on the green arrow to go to the previous page of the book, The Tower of London.

Image source: www.semantic.co.uk

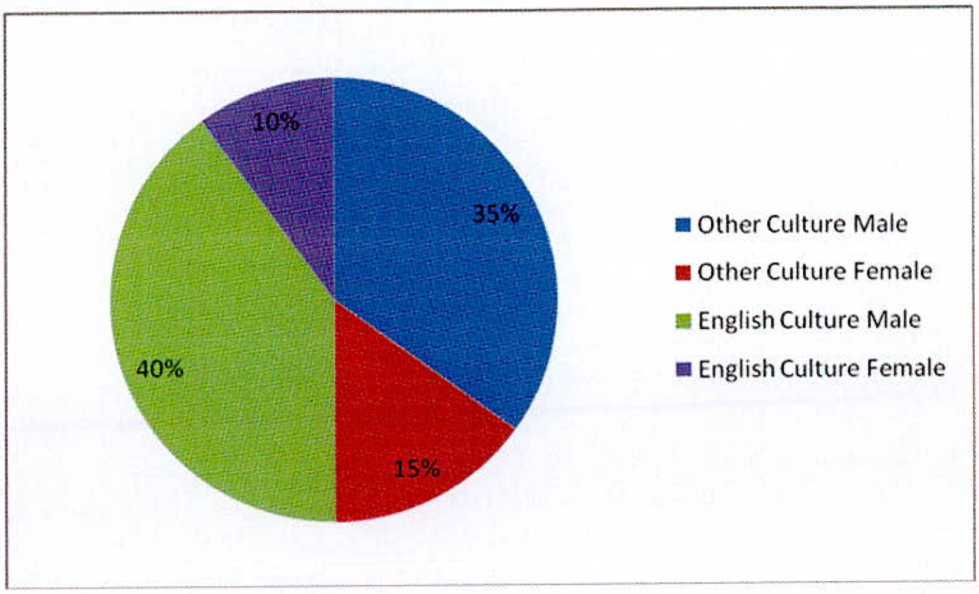
6.7 Method for Testing and Profile of Subjects

Once the prototype talking book had been created, preparations were made to to undertake experiments in the computer laboratory at Thames Valley University. As with all previous tests, the subjects were fully informed in advance of the experiments and informed consent was obtained. In addition, the whole procedure carefully adhered to ethical considerations and credibility of testing methods.

There were 40 subjects from multicultural backgrounds, aged between 16 and 20 years; all of them have SEN and currently attaining lower than a level 2 qualification, which is the benchmark for 16 year olds within this country. It was decided not to ask the subjects what their SEN were as this was not considered to be an important factor as the research investigation is designed to accommodate all SEN, not just some.

There were 40 subjects in total, 50% of the subjects were from the English culture and 50% were from Ghana, Phillipines, Nepal, Germany, Sweden, China, India, Czech Republic, Pakistan and Zimbabwe of mixed male and female gender as shown in the following graphical respresentation. The subjects have been categorised as follows for the purpose of analysis and evaluation of the results. Those who are English culture and those whose culture is other than English. As stated earlier in this chapter, speech is not available that preserves all of of their native cultures for the subjects.

Profile of Subjects



6.8 Procedure for the tests

Before undertaking the actual tests, pilot testing was undertaken with two subjects outside of the sample group. The purpose of the pilot tests was to solely identify any anomalies with the testing procedure or the technology. No anomalies were identified and the results of the two subjects for the pilot are not included in the results. At the start of the tests, the talking book was demonstrated using the LCD projector in the computer laboratory. The rationale behind demonstrating the system was that it would be reassuring for the subjects to see how to use the system in advance of being observed individually.

Following the demonstration, the subjects each had access to the talking book and each one was observed using the system independently but did not help them. It was decided to obtain responses to the learnability and satisfaction of the system by using questionnaires which were created using 18 points from the SUMI technique, with 8 points relating to satisfaction and 8 points relating to learnability. In some instances, where the subjects were unable to read or write fluently, or understand the questionnaire, the statements were read out and the questionnaire was completed on their behalf. A maximum of 10 minutes was allowed per subject for each test.

In addition to the statements on the questionnaire, the subjects were asked verbally after the tests, if they would *prefer* to hear the speech in the same culture as their own, and the response was 100% positive. Subjects were also asked if they would *prefer* to hear the speech in the same gender as their own and, again, the response was 100% positive.

6.9 Analysing the results

The results of the tests were analysed using the ISO quality in use metrics relating to software product quality to analyse human computer interfaces.

Quality in use metrics measure the extent to which a product meets the needs of specified users to achieve specified goals with effectiveness, productivity, safety and satisfaction in a specified context of use. Quality in use is assessed by observing representative users carrying out representative tasks in a realistic context of use. The measures may be obtained by simulating a realistic working environment (for example in a usability laboratory) or by observing operational use of the product. When measuring quality in use it is important that users are only given the type of help and assistance that would be available to them in the operational environment. Satisfaction metrics assess the users' attitudes towards the use of the product in a specified context of use.

The test results were collated into the two categories of English culture subjects and subjects from other cultures.

The accuracy and correctness of quality evaluation relies strongly on the metrics used on that process. Use of the metric in the same product to the same evaluation specification (including the same environment), by different test users should produce results that can be accepted as identical (ISO/IEC PDTR 9126-4, 2000); this applies to these tests which were undertaken in the same computer laboratory under the same environmental conditions.

6.10 Questionnaire

When the tests had been completed, before the results could be analysed it was necessary to apply weighting to each response so that ISO standards could be applied to show significance for the results. A new variable was computed for each response and calculated to show the overall score for each separate component for learnability and satisfaction before ANOVAs could be performed. The nearer that a response is to 3, the more significant the answer is.

Each individual statement was considered further and a decision was made on which would be the most significanct response which would show significance in the use of this software and encoded the answers as follows:

Statement 1 *I would recommend this software to my friends*

Component: Satisfaction

The most significant response would be agree

Agree (3)	Undecided (2)	*Disagree (1)*

Statement 2 *The software has at some time stopped unexpectedly*

Component: Learnability

The most significant response would be disagree

Agree (1)	Undecided (2)	*Disagree (3)*

Statement 3. *Learning to operate this software to start with is full of problems*

Component: Learnability

The most significant response would be disagree.

Agree (1)	Undecided (2)	*Disagree (3)*

Statement 4 *I sometimes don't know what to do next with this software*

Component: Learnability

The most significant response would be disagree

Agree (1)	Undecided (2)	*Disagree (3)*

Statement 5 *I enjoy my sessions with this software*

Component: Satisfaction

The most significant response would be agree

Agree (3)	Undecided (2)	*Disagree (1)*

Statement 6 *It takes too long to learn the software commands*

Component: Learnability

The most significant response would be disagree

Agree (1)	Undecided (2)	*Disagree (3)*

Statement 7 *Working with this software is satisfying*

Component: Satisfaction

The most significant response would be agree

Agree (3)	Undecided (2)	*Disagree (1)*

Statement 8 *This software seems to disrupt the way I normally like to arrange my work*

Component: Satisfaction

The most significant response would be disagree

Agree (1)	Undecided (2)	*Disagree (3)*

Statement 9 *I feel in command of this software when I am using it*

The most significant response would be agree

Component: Satisfaction

Agree (3)	Undecided (2)	*Disagree (1)*

Statement 10 *I would not like to use this software every day*

Component: Satisfaction

The most significant response would be disagree

Agree (1)	Undecided (2)	*Disagree (3)*

Statement 11 *There is too much to read before you can use the software*

Component: Learnability

The most significant response would be disagree

Agree (1)	Undecided (2)	*Disagree (3)*

Statement 12 *Using this software is frustrating*

Component: Satisfaction

The most significant response would be disagree

Agree (1)	Undecided (2)	*Disagree (3)*

Statement 13 *The speed of this software is fast enough*

Component: Satisfaction

The most significant response would be agree

Agree (3)	Undecided (2)	*Disagree (1)*

Statement 14 *There have been times in using this software where I have felt quite tense*

Component: Satisfaction

The most significant response would be disagree

Agree (1)	Undecided (2)	*Disagree (3)*

Statement 15 *It is easy to make the software do exactly what you want*

Component: Learnability

The most significant response would be agree

Agree (3)	Undecided (2)	*Disagree (1)*

Statement 16 *This software is really very awkward*

Component: Learnability

The most significant response would be disagree

Agree (1)	Undecided (2)	*Disagree (3)*

Statement 17 I have to look for assistance most times when I use this software

Component: Learnability

The most significant response would be disagree

Agree (1)	Undecided (2)	*Disagree (3)*

Statement 18 I will never learn to use all that is offered in this software

Component: Learnability

The most significant response would be disagree

Agree (1)	Undecided (2)	*Disagree (3)*

There are an equal number of subjects from English cultures and other cultures. However, there are more male subjects than female subjects. As the main aim of this research is to identify cultural, rather than gender differences, the effects of gender will not be analysed or evaluated with these tests. As with the tests in Chapter Four, the subjects were selected through convenience sampling techniques and the diverse a range of subjects possible would be greatly restricted to a lower number if it an equal gender split was insisted upon.

6.11 Test Results

The figures from the table were then separated into two different components – learnability and satisfaction and were entered into the following table.

Table 6.2. Summary of responses to SUMI questionnaire

Culture	**Learnability**	**Satisfaction**	**Possible Total**
Other than English	341	452	540
English	465	495	540

As we can see from the above table, the maximum possible total for each component is 540. For the Learnability component, the score for the English culture was higher than the other cultures at 465 and 341 respectively. For the Satisfaction component, the score for the English culture was again higher than the other cultures at 495 and 452 respectively. Each response was then analysed individually as follows:

Responses to the Satisfaction component

Statement	**Other than English Culture**	**English Culture**
1.I would recommend this software to my friends.	**60**	**59**

This is a very encouraging response. The score for the cultures other than English is 60/60 and the score for the English culture is 59. If the subjects are so satisfied using the system that they would recommend it to their friends, then they will persevere using the system themselves and are more likely to want to use it regularly. These findings suggest that subjects from cultures other than English are prepared to use software for educational purposes, and this supports the findings of the satisfaction questionnaire as detailed in Chapter Five for the tests for the banana information system where 100% of Kenyan banana farmers said that they would prefer to use the system rather than other sources of information.

Statement	Other than English Culture	English Culture
5.I enjoy my sessions with this software	60	59

Again, this was a very positive response with the score for the cultures other than English 60/60 and the English culture 59/60. Once more this supports the response to the previous statement that subjects from other cultures are prepared to use software for educational purposes. As previously discussed, if end users enjoy using the software they are going to want to use it more (Moore, 2006).

Statement	Other than English Culture	English Culture
7.Working with this software is satisfying	53	59

Again, the scores for this statement were very close for both the cultures other than English and the English cultures with scores of 53/60 and 59/60 respectively. This supports the previous two statements that end users from other cultures are willing to use this type of technology and find it satisfying.

There does not appear to be a significant difference in scores between subjects from different cultures.

Statement	Other than English Culture	English Culture
8.This software seems to disrupt the way I normally like to arrange my work	**54**	**55**

The score for the cultures other than English is 54/60 and the score for the English culture subjects is 55/60. These scores are still high and suggest that the software does not cause any significant disruption to the way the subjects normally like to arrange their work. Even though the scores were not the maximum, these findings do suggest that the subjects could incorporate this software into the way they normally arrange their work without causing too much disruption.

Statement	Other than English Culture	English Culture
9. I feel in command of this software when I am using it	**49**	**59**

This is the first response analysed that shows a marked difference in scores between cultures. The English culture subjects scored 59/60, and the cultures other than English scored 49/60. This suggests that, although subjects from other cultures may enjoy using the software, would recommend it to their friends and that it does not disrupt the way they like to work (as detailed in the previous responses above), they are not as confident in its use as the subjects from the English culture who may have had more previous exposure to ICT.

Statement	Other than English Culture	English Culture
10.I would like to use this software every day	40	39

The scores across all cultural groups are close for this statement. The score for cultures other than English is 40/60 and the score for the English culture is 39/60. This suggests that, although the subjects have already expressed that they find using the software enjoyable, they would prefer not to use it every day. This is in line with the theory that ICT should be used as a tool to aid teaching and learning – not as a replacement for a teacher.

Statement	Other than English Culture	English Culture
12. Using this software is frustrating	58	60

Again, the scores across cultures is very close for this statement with 58/60 of the cultures other than English and 60/60 of the English culture subjects stating that they do not consider the software to be frustrating. This finding is important because it means that they may be more inclined to want to use the software and supports the findings of previous statements in this respect.

Statement	Other than English Culture	English Culture
13. The speed of this software is fast enough	47	46

The score for the cultures other than English was 47/60 and the English cultures 46/60 for this statement. This score is lower than expected and may suggest that further functionality may need to be added to the interface so that the times of the talking book can be adjusted.

Statement	Other than English Culture	English Culture
14. There have been times in using this software where I have felt quite tense	**38**	**59**

The scores for the cultures other than English is 38/60 and the scores for the English culture subjects is 59/60. There appears to be a significant difference between the scores across the cultures and suggests that subjects from other cultures are not as relaxed at using the software as the subjects from the English culture who may have been more exposed to ICT previously. This also supports the findings to Statement 9 as detailed above.

6.12 Summary of Responses for Satisfaction

The findings above suggest that across all cultures, subjects enjoy using the software and would recommend it to their friends. However, the findings also suggest that end users from English cultures are more confident at using the software and find it less stressful. To identify whether there is any significance in the difference to these results, ANOVAs were then used as detailed below:

Table 6.3. ANOVA – Satisfaction Component

Between-Subjects Factors

Factor	**N**
English	20
Other than English	20

This table shows the between-subjects factors. There were 40 subjects in total, with an equal divide between English cultures and cultures other than English. All synthesised speech was female gender.

Descriptive Statistics

The following table of descriptive statistics for the satisfaction component shows the following information.

- The first column refers to the culture of the subjects
- The second column refers to the mean score
- The third column refers to the standard deviation
- The fourth column is the number of subjects.

Table 6.4. Descriptive Statistics – Satisfaction Component

Culture of Subjects	Mean	Std. Deviation	N
English	24.75	1.118	20
Non English	22.60	1.789	20
Totals	23.68	1.831	20

From these results, we can see that the total mean average score for the satisfaction component for the English culture subjects is 24.75, which reduces to 22.60 for the Non English subjects.

These mean values suggest that there may be some signficance in the results and therefore further parts of the ANOVA were analysed.

Table 6.5. Tests of Between-Subjects Effects – English Culture Subjects

Test 1

Dependent Variable is Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig
Corrected Model	46.225[a]	1	46.225	20.775	.000
Intercept	22420.225	1	22420.225	10076.506	.000
Culture of Subject	46.225	1	46.225	20.775	.000
Error	84.550	38	2.225		
Total	22551.000	40			
Corrected Total	130.775	39			

[a] R squared = .353 (Adjusted R Squared = .336)

These results show that there is a significant result for the effect of Culture.

Summary of Satisfaction Results

1) The satisfaction score is higher for subjects from English cultures than non-English cultures.

ANOVA Summary

As can be seen in Table 6.5, a two way unrelated ANOVA showed that significant effects were obtained for the Culture of Subjects ($F_{1,40}$= 20.775, p = 0.000).

The next component to be analysed was Learnability.

6.13 Responses to the Learnability Component

Statement	Other than English Culture	English Culture
2.The software has at some time stopped unexpectedly	39	53

The score for the cultures other than English was 39/60 and the score for the English cultures was 53/60. The author is not aware of any circumstances where the software actually stopped working during the testing procedure and these results, therefore, may be interpreted to suggest that the subjects perceived that the software had stopped and did not know what to do next. This highlights the importance of careful training of subjects before they start to use the software and this may have been a weakness in the testing strategy in not giving enough information on how to navigate through the system. However, it also suggests that even with this simplistic interface, some subjects did not know what to do to move onto the next page and thought that the system had stopped working; they may have thought that the next page came automatically.

Statement	Other than English Culture	English Culture
3.Learning to operate this software to start with is full of problems	48	57

The scores for the cultures other than English subjects was 48/60, and the score for the English culture subjects was 57/60. There appears to be a significant difference in scores across cultures which supports the findings of the satisfaction component where it was considered that the findings suggest that subjects from other cultures are less confident at using ICT.

Statement	Other than English Culture	English Culture
4.I sometimes don't know what to do next with this software	55	51

The score for the cultures other than English is 55/60, and the score for the English culture subjects is 51/60. These findings re-iterate the importance of the simplistic design of the interface as, although these scores are high they are not maximum, implying that not all subjects knew what to do next to get to the next page.

Statement	Other than English Culture	English Culture
6.It takes too long to learn the software commands	60	58

The scores for the cultures other than English is 60/60 and the score for the English culture subjects is 58/60. These scores are encouraging and again supports the previous findings of a simplistic easy to use interface.

Statement	Other than English Culture	English Culture
11.There is too much to read before you can use the software	54	54

The scores across all cultures do not have any significant difference at 54/60. No instruction manuals/guides were provided and the only reading was the words on the screen which were replaced by speech. However, these findings highlight the vulnerability of this target group of subjects and how they perceive reading to be involved. In the design of the interface, therefore, it is important to keep the amount of written instructions to a minimum to accommodate use by end users who lack confidence and may be illiterate.

Statement	**Other than English Culture**	**English Culture**
15.It is easy to make the software do exactly what you want	**51**	**50**

The scores across all cultures were similar with 51/60 for cultures other than English and 50/60 for English culture subjects. Again, this supports the previous findings of the importance of a simplistic user interface as a maximum score was not achieved suggesting that not all subjects found it easy to navigate through the interface.

Statement	**Other than English Culture**	**English Culture**
16.This software is really very awkward	**53**	**51**

Again, the scores across cultures are very similar with 53/60 other than English culture and 51/60 English culture subjects. Again, these findings support the importance of a simplistic easy to use interface as some subjects still found it very awkward to use.

Statement	**Other than English Culture**	**English Culture**
17. I have to look for assistance most times when I use this software	**54**	**54**

Interestingly, there was no significant difference across all cultures with 54/60. This suggest that not all subjects were confident at using the software unaided and enforces the findings that ICT is to be used as a tool to help teaching and learning, not as a replacement as some subjects need assistance in using the software; they cannot necessarily use it alone.

Statement	**Other than English Culture**	**English Culture**
18. I will never learn to use all that is offered in this software	47	50

Again, the score across cultures is similar with 47/60 for cultures other than English and 50/60 for English culture subjects. Again, this suggests that not all subjects are fully confident in using the facilities of this softare and re-iterates previous findings that the end users need to be fully supported in its use.

6.14 Summary of Learnability Responses

The aim of this research is to explore usability issues of ICT as a teaching and learning tool for end users with SEN. The results of the learnability component re-iterates the previous findings that this vulnerable group of end users lack confidence in using the software and in their abilities and that it must be emphasised that ICT is not a replacement for a teacher but to be used as a tool to support teaching and learning. In addition, some subjects were still uncertain about the use of the interface. Although at first glance with the interface for the prototype talking book, end users who do not have SEN needs may consider it to be unncesssarily simple and are used to using far more complex interfaces, these results suggest that it is important to keep the design of the interface as simple as possible for end users who have SEN; this is true across all cultures.

To identify whether there is any significance in the difference to these results, ANOVAs were then used as detailed below:

<u>Table 6.6. ANOVA – Learnability Component</u>

Between-Subjects Factors

Factor	N
English	20
Other than English	20

This table shows the between-subjects factors. There were 40 subjects in total, with an equal divide between English cultures and cultures other than English. All synthesised speech was female gender.

Descriptive Statistics

The following table of descriptive statistics for the satisfaction component shows the following information.

- The first column refers to the culture of the subjects

- The second column refers to the mean score

- The third column refers to the standard deviation

- The fourth column is the number of subjects

Table 6.7. Descriptive Statistics – Learnability Component

Culture of Subjects	Mean	Std. Deviation	N
English	23.25	3.354	20
Non English	17.05	.224	20
Totals	20.15	3.919	40

From these results, we can see that the total mean average score for the learnability component for the English culture subjects is 23.25, which reduces to 17.05 for the Non English subjects.

These mean values suggest that there may be some signficance in the results and therefore further parts of the ANOVA were analysed.

Table 6.8. Tests of Between-Subjects Effects – English Culture Subjects

Test 1

Dependent Variable is Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig
Corrected Model	384.400 [a]	1	384.400	68.035	.000
Intercept	16240.900	1	16240.900	2874.496	.000
Culture of Subject	384.400	1	46.225	384.400	.000
Error	214.700	38	5.650		
Total	216840.000	40			
Corrected Total	599.100	39			

[a] R squared = .642 (Adjusted R Squared = .632)

These results show that there is a significant result for the effect of Culture.

Summary of Learnability Results

1) The learnability score is higher for subjects from English cultures than non-English cultures.

ANOVA Summary

As can be seen in Table 6.8, a two way unrelated ANOVA showed that significant effects were obtained for the Culture of Subjects ($F_{1,40}$= 384.400, p = 0.000).

6.15 Hypothesis Testing

After analysing and evaluating the results of the questionnaires, the the following hypotheses were tested:

Hypothesis (d) – *There is no difference in learnability, when the language of the speech is representative of the culture of the subjects.*

A two way unrelated ANOVA showed that significant effects were obtained for the Culture of Subjects ($F_{1,40}$) = 68.035 p=.000.

These findings reject the hypothesis.

Hypothesis (e) – *There is no difference in satisfaction when the language of the speech is representative of the native culture of the subjects.*

A two way unrelated ANOVA showed that significant effects were obtained for the Culture of Subjects ($F_{1,40}$) = 20.775 p=.000.

These findings reject the hypothesis

.

6.16 Chapter Summary

This chapter details an evaluation into two teaching and learning tools that incorporate TTS technology that are used successfully for end users with SEN but do not address cultural usability issues. Using the results of the evaluation plus findings from the usability tests of as detailed in Chapters Four and Chapter Five, a teaching and learning tool in the form of a talking book was developed. This prototype has a multi-modal design with a GUI and speech in English accent. On the completion of the prototype, tests were undertaken with 40 subjects from multicultural backgrounds, all of whom had SEN which were not profound and did not prohibit them from using a computer without the use of further adaptive technology, switches for example.

The tests that were undertaken evaluated the components of learnability and satisfaction as opposed to error rates. The results suggest that learnability and satisfaction are improved when the subjects hear speech that is in the same culture as their own.

The findings from these results will be discussed in more detail in the next chapter which will also contain discussions and conclusions for all of the research undertaken in this project.

Chapter Seven – Conclusion and Discussion

7.1 Chapter Introduction

Human beings have been communicating with each other for thousands of years, but traditionally this was with people of the same culture and in the same language. Nowadays, with the increased ease of travel and communication via computer related technology, many are now living in a multicultural society and find communication increasingly difficult.

Many people live in countries that have a dominant culture that may be different to their own and find that the technology available to them in that country is only available in the culture of the dominant culture. With the use of TTS technology, the experiments that have been undertaken in this research project show that this reduces the usability of such systems.

The situation may be worse for developing countries where more than 40% of the population may be illiterate, with most cases of illiteracy being in rural areas where there is also a lack of access to technology. The use of TTS technology through a TBI interface could be used as a tool in these situations but only if careful consideration is given to the exact needs of the end users with an interface that is designed to be straightforward to use in a native cultural language or accent that the end users can clearly understand.

Many cultural groups value highly their cultural characteristics and their linguistic diversity – after all this is what sets them aside from other cultures, yet technology has been slow in embracing linguistic diversity. Despite the fact that there are over 6,000 languages spoken worldwide, 75% of all web content is in the English language. Indeed, the representation of Kiswahili in computer technology is so marginal that no evidence has been found of it featuring in any surveys and there may, indeed, be a very real risk that as minority languages become far less accessible and available, they may disappear altogether.

There is inactivity amongst many software developers in promoting machine translations in native cultures and, although developing language technology may be expensive, this remains a critical area of technology that is still much overlooked.

7.2 Reflection on Research Methodology

A research approach was selected that provided measurable, creditable and repeatable results. All of the experiments were carried out with strict adherence to ethical and industry standards, ensuring that no distress whatsoever would come to vulnerable human subjects.

There were three different sets of experiments and although these were in different situations, they were all under strictly controlled conditions. The first set of experiments that were held in the computer laboratory at Thames Valley University included forty four subjects from Indian and English cultures and tested generic TTS synthesis software.

The second set of experiments were held at the farms of banana farmers in Kenya and tested the usability of a TBI in English and Kiswahili.

The final set of experiments were held in the computer laboratory at Thames Valley University again, but this time using a prototype talking book with subjects who had mild to moderate SEN.

After careful considerations of different paradigms and philosophical approaches the post-positivist approach was adopted for this research. With reference to the philosophical aspect of Epistemology, preference is given to critical tradition and critical community with the aim of the inquiry being explanation and prediction of knowledge. Findings are probably true but always subject to falsifications, (Guba and Lincoln, 1994). With reference to the philosophical aspect of Ontology, reality is assumed to exist but only imperfectly, it is probabilistically apprehensible due to the imperfectability of human cognitive and intractable nature of phenomena. With reference to the philosophical aspect of Methodology, hypotheses are initially assumed to be false; they are probable facts or laws. Inquiries are done in more natural settings and with the collection of more situational data.

The above provided an appropriate philosophical approach for the carefully controlled tests, critical analyses and evaluation of the results and hypotheses testing.

If one of the aims of HCI is for computers to fit in with us, rather than us with them, speech must have the potential to be the most intuitive interface. However, the results of the experiments undertaken in this research help to identify why it is not used yet to its fullest advantage.

On reflection, it may have been useful to have included more qualitative methods in the research methodology to complement the quantitative methods that were followed. With the first set of experiments that were held in the computer laboratory at Thames Valley University, interesting data may have been gleaned by talking to the subjects after the tests to establish what their personal feelings were towards the software and how they found it to be in terms of its emotion, pace and tone.

With the second set of experiments which were held in Kenya, due to the poor connectivity, the tests actually took place by hearing sound files on a laptop. The use of keystroke logging may have given an insight into where mistakes were being made in the use of the system, for example were all of the subjects misinterpreting a specific part of the hierarchical structure, or were all of the errors made independent of each other. The results of this logging could have been discussed with the subjects at a later date.

7.3 Reflection on Ethical Framework Followed

When undertaking experiments with human participants, it is critical to ensure that an ethical framework is adhered to that protects the participants against any unnecessary distress. Although the actual risk of any physical harm in taking part in usability tests involving software is neglible, some of the subjects were from vulnerable groups with SEN and may, in particular be anxious about participating. Therefore, before any of the experiments a thorough investigation was undertaken into ethical standards. Subsequently, the experiments in this research are based on principles from the Economic and Social Data Service (2005), who state that it is a fundamental ethical principle of research to ensure that subjects give consent and that they enter into the research voluntarily, without coercion and in the knowledge that they can withdraw at any time, confident in the knowledge that there will be no adverse consequences of participating in the research, including the preservation of anonymity if they wish.

When the experiments, consent and information sheets had been designed, it was then necessary to seek approval to conduct the tests from the University's Ethics Committee.

This entailed the submission of a Testing Proposal and accompanying 40 page interim report to a panel of ten senior academics experienced in ethics involved in research; this was a very important procedure to ensure that there was consensus amongst researchers that the tests would not have the potential to cause distress to any of the subjects.

The ethical framework adhered to, also complies with the recommendations from The British Psychological Society (2009). The society emphasises the importance of the consideration of ethical implications and ensuring that all participants are as fully informed as possible prior to testing. According to The British Psychological Society (2009), in recent years there has been an increase in legal actions by members of the general public against professionals for alleged misconduct and that researchers must recognise the possibility of such action if they infringe the rights and dignity of participants in their research.

The importance of making it clear to participants about their rights to withdraw from the research at any time is also emphasised by The British Psychological Society (2009).

7.4 Reflection on Objectives

Objectives (i) and (ii)

The first objective for this research project was to:

Analyse and evaluate the significance of the effects of culture and accent on the usability of TTS applications.

The second objective for this research project was to:

Analyse and evaluate the significance of the effects of gender on the usability of TTS applications.

Experiments were undertaken in the computer laboratory at Thames Valley University using subjects who were first year Computer Science students. They all had a similar level of previous computer experience but none of them had prior use of TTS. After obtaining informed consent, pilot testing and with strict adherence to ethical issues, the experiments took place using AT & T natural voices in English language, English accent and English language, Indian accent. There were eleven male English culture subjects, eleven female English cuture subjects, eleven male Indian culture subjects and eleven female Indian culture subjects. There were two tests – one of them asked the subjects to identify the correct word to complete a sentence and the other one asked the subjects to identify the correct word from a choice of two random words, which were not in any sort of context. An equal number of speech patterns were used in both English and Indian accent but all speech was female.

The results of the tests were then analysed and evaluated to identify if the accent of the speech had any effect on the error rates of the tests. The results showed that for each test and for each culture, the error rate was lower when subjects heard speech that sounded as though it was in their own accent.

Previous research by Papamichalis (1987), considers that speech output is more accurate when one word is heard at a time rather than in continuous prose but the results of the experiments conducted in this research were inconclusive in this respect.

A hypothesis was also tested to identify whether the gender of the subjects had any effect on the outcome of the usability but in this case, the results of the tests were also inconclusive and did not, therefore, suggest that there was necessarily any phenomenon in this respect. It is out of the scope of this project to investigate this further but when synthesised speech becomes available in Indian male accent in addition to Indian female accent, then there may be the opportunity for further research in this area.

Findings for Objectives (i) and (ii)

The error rate was significantly lower for both male and female subjects, from both English and Indian cultures, when they heard output speech that had an accent that was representative of their native culture. It was also found that when speech was heard in context ie to finish a sentence, the error rates across both cultures was lower for male subjects than female subjects. However, when speech was heard without context ie single words, the error rate was lower for female subjects from the English culture but lower for male subjects from the Indian culture.

The findings of this experiment show that culture and accent has a significant effect on the usability of TTS applications. The effects of gender, however, was not proven.

Objective (iii)

The third objective was to:

Analyse and evaluate the significance of the effects of localised speech on the usability of TTS applications.

Undoubtedly, there is a need for banana farmers in remote rural locations in Kenya to have access to up-to-date information to help them become more successful in growing their crops. The Kenyan government is fully supportive towards this initiative and the LLSTI designed and developed a TBI in both the Kiswahili and English language as an endeavour to be a solution to this problem. This is a highly commendable area of work which has the potential to help thousands of farmers throughout Kenya. However, the field of TTS technology is complex, and with the inclusion of a TBI, this complexity is increased. Unfortunately, there were several issues which rendered this system unsuable, resulting in the system being withdrawn from use just a few months after its implementation.

After obtaining informed consent, pilot testing and with strict adherence to ethical issues, the testing took place at the subjects' farms in Kenya using speech in both Kiswahili and English languages. The actual tests had been carefully designed by the author in close liaision with NALEP and the University of Nairobi to ensure that they were appropriate and to endeavour to give the subjects an early success experience; the subjects were asked to select an answer from a multiple choice. A translator and transcriber were appointed to be present at all times due to the low levels of literacy in this target group of end users.

However, conducting the experiments had many problems. Firstly, there were serious problems with the communications technology in Kenya and the experiments did not provide any evidence to suggest that the technology in Kenya could support such a system as there were five failed attempts to undertake the experiments using the TBI, through lack of connectivity. It was therefore considered that, due to the time restrictions placed on the developers, that there would be a back up plan and if, on the sixth attempt, there was still lack of connectivity, then the sound files would be played on a laptop computer and the subjects would navigate by pressing the buttons on the laptop instead of a telephone. Although not the ideal, this is what happened which still enabled the experiments to take place. After further research by the LLSTI, it became evident that there would only be certain times of the week when the farmers would be able to use the TBI, due to the monopoly of business use.

The second problem was with the speech from the system. The system was available in both Kiswahili and English languages. However, the subjects in the tests had difficulty understanding either of these. The reasons for this are that, although the Kiswahili language was spoken by a native speaker, insufficient attention had been given to the localisation process resulting in formal grammar being used which banana farmers in rural areas were unfamiliar with. The English language was in an English accent, which the subjects were also unfamiliar with and could not understand. However, at the end of the experiments, the subjects were played soundfiles from a prototype speech which is in English language with a Kenyan accent and all of the subjects commented that they found the clarity of this system much clearer. Subsequently, the LLSTI are continuing to develop this speech to incorporate into the system which may increase its clarity and the usability of the system.

The third problem was the difficulty in navigating through the system due to the large amounts of speech that had to be listened to. If the subjects made a mistake and pressed the incorrect selection, they could find themselves totally lost in such a system and spending a long time, and a lot of money on the telephone, listening to useless information.

It could be the case that, even with the improved speech in the Kenyan English accent, a speech interface on its own may not be appropriate for subjects who have SEN such as Kenyan banana farmers who may be illiterate. Therefore, the LLSTI considered the further possibility of producing a complementary graphical user interface so that the farmers could use this in addition by accessing this from the Internet, wherever access was available.

Of course, not all farmers would have access to this but it may help those who can.

Findings for Objective (iii)

The error rate from the tests was unacceptably high and the positive response from the subjects to the learnability and satisfaction components showed that there was a significant discrepancy between the subjects' perceived satisfaction with the system, and its actual usability. The error rates were, however, lower when the output speech represented the culture of the end users.

The findings of this experiment show that localised speech has a significant effect on the usability of TTS applications.

Objective (iv)

The fourth objective was to:

Use the results of Objectives (i) to (iii) to develop a bespoke system and evaluate its learnability and satisfaction.

An interview was conducted with the developer of the Rapid English system at a Young Offenders Institute where he had successfully implemented a teaching and learning tool which he had designed which incorporated TTS technology. This system was designed for learners who have SEN and who need to improve both literacy and communication. After interviewing the developer and observing end users using the system at the Institute, a comparative evaluation was undertaken and it was observed that the system had a very clear, simplistic interface. The developer emphasised the importance of having the minimal amount of text and images on the screen and that success rates had been much higher when simplistic interfaces were used, as end users with SEN found this far more straightforward to cope with as it reduced confusion and distractions.

The use of speech in this system played a pivotal role to its success. Speech output was available in English language, male only, but the end users could also record their own speech and play it back with the idea that their communication and pronunciation skills could be monitored and progression recorded. The manner in which the speech was used, by using the minimal amount of words, complemented by a GUI consisting of the minimal amount of text and icons, were very positive parts of the interface design.

Observations were made at the Ace Advisory Trust Centre on end users using Clicker 5, and an interview was conducted with the Assessment Officer and Speech and Language Therapist. The Ace Advisory Trust Centre helps learners who have profound disabilities with the use of technological aids. Clicker 5 is the system that is most frequently recommended by them to help learners with education and communication skills. In common with the Rapid English system, it was emphasised during the interview, the importance of keeping a very simplistic interface to ensure that end users are not confused or distracted - one icon at a time is the ideal.

Speech also plays a pivotal role with this system which is available as an output in both genders, in English language only in English and American accents. It was commented that parents and guardians had expressed preference for a child-like voice, but this had not been supported by the children themselves. However, it was also commented that being able to listen to speech that was in the same accent as their own has the potential to be of enormous benefit in helping them understand the speech and develop their own communication skills as many of these end users spend a lot of their time at home with carers who communicate in their accent.

A comparative evaluation of this system was undertaken. It was clear, straight forward to use with a simplistic design. However, as identified by the Ace Advisory Trust Centre, there is the potential to improve usability by incorporating speech in the native culture of the end users.

Using the findings of the previous experiments and research, a prototype talking book was created. This prototype incorporated a simplistic GUI which was complemented by speech output, which was kept to the minimum.

The talking book was on the subject of London which would be of common ground to all subjects who took part in the experiments. There were forty subjects in total – twenty were from the English culture and twenty were from a mix of cultures other than English. In previous experiments, the author has striven to ensure an equal number of subjects from each culture and gender, but however it was not possible to recruit such numbers for this experiment and therefore the author decided that she would prefer to test the system on as a diverse representation as possible and categorise the users as English culture and non-English culture; all of the non-English culture did not have English as their first language and had not been resident in England for more than one year.

After taking informed consent, the experiments were undertaken at the computer laboratory at Thames Valley University. All of the subjects had mild to moderate SEN and had not used TTS before. The talking book was demonstrated on the projector and then an observation of each subject using the system individually was undertaken. Questionnaires based on the SUMI principle, were used to gather responses on learnability and satisfaction components of the system.

The speech output from the system was in English female voice only. As this research has shown, synthesised speech is not readily available in accents other than Indian and therefore it was not possible to incorporate all of the different accents. The results of the questionnaire were analysed and evaluated to identify whether the culture of the speech had any effect on the usability of the system. These results supported all of the previous experiments undertaken that the usability was improved for end users when they heard speech that sounded as though it was in the same culture as their own. The incorporation of speech in accents, when it becomes available, therefore has the potential to improve the usability of such systems.

All of the experiments that have been undertaken have consistently shown that usability is increased when the speech that is output preserves the cultural characteristics of the end users. Due to the inability of learners with SEN to absorb large amounts of speech at a time, it is also appropriate to incorporate a GUI wherever possible to complement such systems. The prototype talking book that was developed incorporated both speech and a GUI, with the resultant testing showing that the functionality of speech output in different accents has the potential to improve the usability of the system.

Findings for Objective (iv)

The findings show that speech can play an integral and critical role as a teaching and learning tool, but that usability of TTS technology is improved when speech is used as part of a multi-modal interface.

Across all cultures, the subjects enjoyed using the software. However, the findings also show that end users from English cultures are more confident at using the software and find it less stressful. The results also showed that this vulnerable group of end users lack confidence in using the software and in their abilities meaning that a simplistic user interface is critical and, emphasising the key point, that ICT is not meant as a replacement for a teacher, but is to be used as a tool to support teaching and learning.

The findings of this experiment show culture has a significant effect on learnability and satisfaction.

7.5 Reflection on the Experiments

This research involved the undertaking of three separate sets of experiments. Each set of experiments used different subjects who were from diverse cultures, some of whom had SEN.

Before each set of experiments, pilot testing was undertaken with the main purpose to identify any weaknesses in the testing procedure. The pilot tests used fewer subjects than for the actual tests but they were representative of the subject group. Although the results of the pilot tests are incluced, they are not used for analysing and evaluating the significance of the experiments.

The First Set of Experiments

There were problems identified with the first set of experiments which were undertaken at the computer laboratory at Thames Valley University were satisfactory. The output speech ran at an erratic rate which made it difficult for subjects to absorb the information and complete the tasks. It was necessary, therefore, to stop the tests and repeat the speech. This problem was, therefore, corrected before the actual tests took place at a later date.

The actual speech was heard from two loudspeakers at the front of the laboratory. These had also been tested in the pilot and was not identified as a problem by any of the subjects, however, on reflection it may have been good practice to have provided the subjects with personal headphones as this may have helped improve their concentration and reduce any errors in glitches in sound signals.

The Second Set of Experiments

Undertaking the second set of experiments proved to be a challenge. The pilot tests went well with two volunteers from the local farming community undertaking the tests using mobile 'phones as planned. However, the scheduling for the actual tests resulted in five failed attempts due to poor connectivity in Kenya.

However, there was a back-up plan in place and as stated earlier in this thesis, the subjects proceeded to use a laptop computer to hear the sound files and use the keyboard to make their selections. On reflection, it would have been useful to have had another opportunity to allow the same subjects to test the system again using mobile 'phones and to undertake a comparison of results. However, due to time constraints and the poor connectivity, this was not an available option at the time. There may have been some interesting results as due to the poor connectivity, the sound files heard from the mobile 'phone may have been even less intelligible than those from the laptop.

The Third Set of Experiments

The third set of experiments involved subjects who were particularly vulnerable as they all had SEN, although none of them had any hearing impediments. The experiments all took place in the computer laboratory at Thames Valley University and there we no issues identified with the pilot tests, or in the actual tests. However, on analysing and evaluating the results, some subjects considered that there had been times when the software had stopped running, but this had not been the case. On reflection, it would have been useful to have taken a video recording of the experiments to try to identify what had happened to imply that a software fault had occurred.

7.5 Reflection on the Conclusions for the Experiments

Hypothesis a).

There is no difference in error rate when the accent of the speech is representative of the culture of the subjects.

This hypothesis was tested in the first set of experiments. The results were analysed using ANOVAs which showed that there were statistically significant differences between speech that was representative of the culture of the subjects and speech that was not. It was, therefore, possible to draw the conclusion, directly from the statistical evidence, that this hypothesis should be rejected.

Hypothesis b)

The gender of the subjects make no difference to the error rate.

Although it was not the overall aim of this research to explore gender inequalities, it was considered that it was important to establish, at the first set of experiments, whether there was any significance in the results which arose from gender differences. If there was any significance, then there could be an argument that statistical evidence could not be used on its own to support the fidings that the differences in error rates were solely culture, as gender may also have played a part. Therefore the above hypothesis was also tested. As with the first hypothesis, a detailed statistical analysis was undertaken using ANOVAs which showed that there was not a statistically significant difference that could be made between the gender of the subjects and the error rate. Therefore, the conclusion was drawn directly from this statistical evidence, that this hypothesis should be supported.

Based on these findings, the author was confident that it would not be necessary to incorporate gender differences into further experimental design.

Hypothesis c)

There is no difference in error rate when the language of the speech is representative of the culture of the subjects.

This hypothesis was tested in the second set of experiments in Kenya. The banana information system used speech that had been localised for the end users and speech that had not. It was important, therefore, to undertake an analysis and evaluation on how successful the localisation process had been. The testing procedure was fraught with difficulties, from poor connectivity and time restraints. Eventually, testing took place using 10 subjects from the target group. Although Nielsen states that 10 is a satisfactory number for undertaking usability tests, this number of subjects did not produce enough data to be entered into an ANOVA to identify any statistically significant differences, as the minimum number of subjects required for this method of analysis is 20.

The results of the experiments, however, were then calculated to show the percentage of tasks that had been answered correctly. Using this method of analysis and evaluation, the conclusion was drawn that the system was unusable due to the unacceptably high error rates. However, it was not possible to state, with certainty, what any statistical significant differences there were.

From the results of the experiements, the findings suggested that the hypothesis should be rejected because the amount of tasks answered correctly was higher for the speech that had been localised and was representative of the culture of the subjects. This did not, however, imply that the localisation process had been satisfactory. Furthermore, it was not possible to distinguish statistically between the intelligibility of the speech, or the poor interface.

Hypotheses d) and e)

d) There is no difference in learnability when the accent of the speech is representative of the culture of the subjects.

e) There is no difference in satisfaction when the accent of the speech is representative of the culture of the subjects.

These hypotheses were tested in the computer laboratory at Thames Valley University using 40 subjects from a mix of different cultures. The subjects were asked to score 18 questions on a SUMI based questionnaire and the subsequent scores were weighted and entered into ANOVAs. The results showed that there were statistically significant differences when the accent of the speech was representative of the culture of the subjects for both hypotheses and using this statistical information, the conclusion was drawn that both of these hypotheses could be rejected.

7.6 Reflection on the Aim

The overall aim of this research was to:

Investigate the significance of the effects of culture and accent on the usability of TTS technology as a teaching and learning tool for end users, of post compulsory school age (16+), for whom English may not be their first language and, in addition, may have SEN.

This aim has been met in full. Extensive experiments with cross-cultural human subjects, including those with SEN have been undertaken and, where appropriate, the results have been modelled against Hofstede's cultural dimensions. The results of the experiments consistently show that the usability is improved when speech is incorporated that is representative of the end user's native culture and that, therefore, this should be incorporated into text to speech technology to improve its usability.

Interviews, observations and a comparative evaluation of two existing educational software programs that incorporate TTS technology for end users with SEN have been undertaken; one of the systems was used in mainstream education for people who have SEN, the other system was used for people outside of mainstream education who have profound SEN There was much commonality between the systems and it became evident that the use of a multi-modal interface that incorporated speech and a graphical user interface was appropriate for this cohort of end users.

Using the results of all of the analyses and evalutions, a prototype talking book was developed that consisted of a multi-modal interface and it was tested on subjects who were in mainstream education but had SEN. This system was well received by the subjects, but the testing procedure did identify the importance of having the most simplistic of interfaces, which may appear far too basic for developers of more general systems.

7.7 Further Research and Development Opportunites

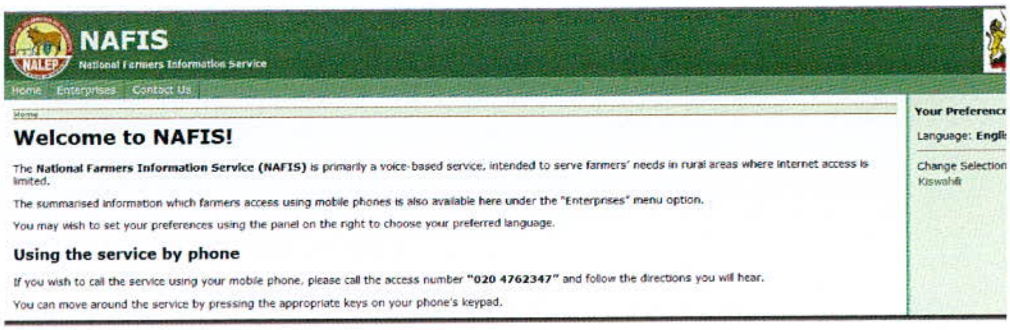
TTS technology can provide an effective teaching and learning tool if the needs of the end user are carefully considered, otherwise the effects could, conversely, be detrimental and cause much frustration. There is a need to bridge the gap between what designers want to design and what end users really need. This is a challenging enough process when developing systems for end users all from the same culture, but is compounded when the end users come from different cultures and have SEN.

This has been particularly true for the banana information system. The need for such a system cannot be denied and the dedication and commitment of the parties involved is admirable. However, until this gap is filled with a thorough analysis of what the end users need, they will continue to be disappointed with failing systems.

After the completion of the experiments and the withdrawal of the unoperational system, the developers then attempted to re-implement the banana information system in April 2008, this time by using a website to complement the speech, but by June 2008 yet again, this was withdrawn from use.

As can be seen from the following images, the web page consists mostly of text, which this research project has shown is not an appropriate design for end users who may be illiterate; this appears to be an area that has been overlooked by the developers.

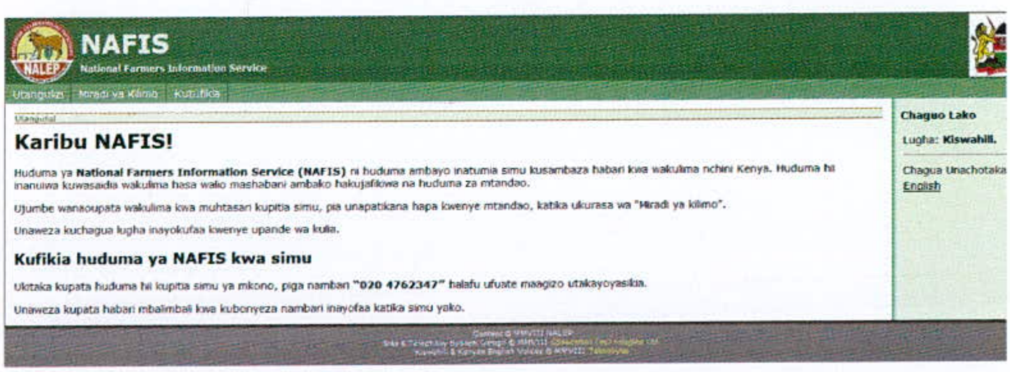
Figure 7.1. Banana Information System via The Internet in English Language



Source: www.nafis.go.ke.

There is an option to also view the system in Kiswahili but you have to be able to read the instructions to know where to click to access this first.

Figure 7.2. Banana Information System via The Internet in Kiswahili Language



Source: nafis.go.ke

The design of this particular interface contradicts all of the findings detailed in this thesis in terms of design for end users with SEN; the incorporation of of icons in the interface design would have been a preferred approach.

Although it is out of the scope of this research project, the developers are now in the process of designing and developing another interface for the information system in Kenya and, this time the developers are incorporating some positive aspects of interface design whch are more appropriate for these end users. Firstly, they are no longer going to use the TBI as the connectivity is too poor, but will instead provide the information via ipods.

The system will incorporate a GUI consisting of images to prompt the end users, text and speech. However, instead of having the choice of speech output in Kiswahili and English the speech output will only be available in English with the native Kenyan accent, which is the one that the subjects identified as being the clearest speech at the time of the experiments. However, the developers must ensure that the amount of speech that is incorporated is kept to a minimum as too much speech may, again, deem the system as unusable. There will also be other factors to take into consideration with the use of ipods in a developing country.

Statistics suggest that only 30% - 40% of information systems that are produced are ever implemented successfully, with previous research showing that the main contributing factors are weaknesses in the analysis and testing stages of the systems lifecycle, where the actual end users are overlooked and instead developers produce systems that more closely match their own needs. However, the systems lifecycle is not a strictly sequential set of stages and through continuing development, the banana information system may not just become another statistic.

Gakuru and Tucker (2009) state that the development of the system has taken five years and that it is still not complete. However, they state further that the length of time is partly due to the fact that this development has involved the imparting of skills and ownership to local developer. Gakuru and Tucker (2009), refer to a faster approach which could have included getting a large donor or government department to buy lots of expertise from the developed world, put together the system, install it and hand it over to the Kenyan government but that this latter approach is unlikely to have worked. Using their approach they now have local expertise and ownership of the system with provides possibilities for it to be maintained locally.

If local expertise and ownership was not counted on right from the beginning of the project, the development process would have terminated on the first Kiswahili TTS system, when no more funding could be secured, (Gakuru and Tucker, 2009). Further development, improvement and maintenance can now be carried out cost-effectively locally and the authors consider that this demonstrates that this approach of technology transfer, could be adopted for other development projects to ensure maximum utalisation and sustatinability of ICT projects.

Further research into building capacities in human language technology (HLT) for African languages is being undertaken by Adegbola et al (2009), who consider that the development of HLT is one of the important by-products of the information revolution but that the level of knowledge and skills for African languages remain low, with most development and research continuing within the frameworks of knowledge production for an industrial society. Adegbola et al (2009) have started on a five year initiative for the acceleration of the development of knowledge and skills in HLT for African languages. The findings of this research, may assist the development of intelligible speech in Kiswahili for use in the banana information systems and it would be very interesting to follow closely this research.

A further area where further research could be undertaken is to undertake an analysis and evaluation of the significance of the effects of gender in the usability of TTS. Although the first set of experiments investigated the possibility of the effects of gender, the results did not suggest that there was a phenomenon to investigate in this research study. However, previous research suggests that the use of computer technology can lead to gender inequalities.

7.8 Conclusion

It is vital that research continues into the development of speech that preserves the native culture of end users and that this functionality is incorporated into systems so that they can be effectively used as teaching and learning tools. The findings of this research show consistently that usability is improved when end users hear speech that preserves their native culture but the findings also show that at the present time, the needs of multicultural end users, who may have SEN are not being met. However, with continued research and the development of speech in different languages and accents, there is the potential to reach the ideal where an end user can select speech that is representative of his or her native culture.

The knowledge and skills already exist to incorporate such technology; when the time comes that it is fully implemented, we will then be in a position to consider that developers and end users in this field, have an understanding of each other.

References/ Bibliography

Abdelnour-Nocera, J. & Dunckley, L. (2005). Usefulness as a Social Construct: A Study of Cultural Contrasts between System Producers and Users. In D. Day, V. Evers and E. Del Galdo (Eds). *Designing for Global Markets 7: Bridging Cultural Differences: IWIPS' 2005 Proceedings.* The Seventh Internatioanl Workshop on Internationalization of Products and Systems. Grafisch Centrum, Amsterdam. PP 69-82.

Abdelnour-Nocera, J. & Hall, P. (2004). Global software, local voices: The social construction of usefulness of ERP system. In F. Sudweeks and C.Ess (Eds) *Proceedings of the 4th International Conference on Cultural Attitudes towards Technology and Communication.* Karlstad, Sweden. June/July. Pp 29-42. Murdoch University.

Abratt, D., Mallinson, W. & Bekker, A. (2003). Usability Testing: Recipe for Success. *StarWest 2003.* San Jose, California.

Adegbola, T. (2009). Building Capacities in Human Language Technology for African Languages. *Proceedings of the EACL 2009 Workshop on Language Technology for Africal Languages.* Nigeria. 2009. Pp 53-58.

Allen, R. B. (1984). Working paper on ethical issues for research on the use of computer services and interfaces. *ACM SIGCHI Bulletin.* **16** (1), pp 12-16.

Anderson, L.M., Evertson, C.M. & Brophy, J.E. (1979). An experimental study of effective teaching in first grade reading groups. *Elementary School Journal.* **79,** pp 193-223.

Banes, D. & Brave, J. (2002). *Accessibility and Inclusivity in further and higher education: an overview,* in: L. Phipps, A. Sutherland and J. Brave et al (Eds) Access All Areas: disability, technology and learning, JISC TechDis and ALT, **1,** pp 1-5.

Banes, D. & Walter, R. (2002). *Internet for All, 2nd Edition.* David Fulton, London.

Barber, B. (1995). *Jihad vs. McWorld:* How Globalism and Tribalism are Reshaping the World. Ballantine Books, New York, USA.

Barber, W. & Badre, N. (1998). Culturability: The Merging of Culture and Usability *The Online Proceeding of the Fourth Conference on Human Factors and the Web.* Basking Ridge, New Jersey, USA. [Online]. URL http://www.research.att.com/conf/hfweb/proceedings/barber/index.html. (Accessed 26 October 2004).

Barker, T., Barker, J. & Doolan, M. (2000). The Development of Multimedia Learning Applications for Use by Students in Different Linguistic and Cultural Contexts. In A Smith (ED). *Cultual Issues in HCI: One Day Workshop Sponsored by the British HCI Group and Optimum.Web Ltd.* University of Luton. Pp 66-74.

Barnard, E. & Davel, M. (2004). LLSTI isiZulu TTS Evaluation Report. *CSIR,* Pretonia, South Africa.

Bellegarda, J.R., Silverman, K.E.A., Lenzo, K. & Anderson, V. (2001). Statistical prosodic modelling: from corpus design to parameter estimation. *IEEE transactions on speech and audio processing,* ***9****, pp 52-66.*

Bibliography of HCI. [Online]. URL www.bathspa.ac.uk/schools/graduate-school/biblio-pri. (Accessed 26 October 2004).

Black, A.W. & Campbell, N. (1995). Optimising selection of units from speech databases for concatenative synthesis. *Eurospeech.* **1,** pp 581-584.

Blin, F. & Munro, M. (2008). Why hasn't technology disrupted academics' teaching practices? Understanding resistance through the lens of active theory. *Computer Speech and Language Journal,* **2,** pp 475-490.

Bourges-Waldegg, P. (2000). Globalization: A Threat to Cultural Diversity? In D. Day, E. Del Galdo and G. Prabhu (Eds). *Designing for Global Markets*

2: IWIPS' 2000 Proceedings. The Second International Workshop on Internationalization of Products and Systems. Backhouse Press, Maryland, USA. Pp 15-124.

Bourges-Waldegg, P. & Scrivener, S. (1998). Meaning, the Central Issue in Cross-Cultural HCI Design. *Interacting with Computers.* **9,** (3), Pp 287-309.

Brave, S., Nass, C. & Hutchinson, K. (2005). Computers that care: investigating effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies.* **62,** 2, pp 161-178.

Breen, A.P. & Jackson, P. (1998). Non-uniform unit selection and the similarity metric within BT's laureateTTS system. *Third ESCA speech synthesis workshop, pp201-206.*

Brewster, S. (1994) *Providing a structured method for integrating non-speech audio into human-computer interfaces.* PhD Thesis. University of York.

Brewster, S. (2008). *Earcons and sonically-enhanced widgets.* [Online]. URL www.dcs.gla.ac.uk. (Accessed 25 March 2008).

Brewster, S. (2008). *Principles for improving interaction in telephone-based interfaces.* [Online]. URL www.dcs.gla.ac.uk. (Accessed 25 March 2008).

British Psychological Society. (2002). Inclusive Education Position paper. *[Online]. URL www.bps.org.uk.* (Accessed 13 June 2007).

British Psychological Society. (2009). British Psychological Society Guidelines. *[Online]. URL www.bps.org.uk.* (Accessed 27 June 2009).

Brown, A., Rhodes, E. & Carter, R. (2002). Innovative attempts to support learning through continuing vocational education and training. *Paper presented at FORUM, Evora.*

Brown J. (1958). *Some Tests of the Decay Theory of Immediate Memory.* Quarterly journal of Experimental Psychology, **10,** pp 12-21.

Buckingham Palace Guards Image. *[Online]. URL www.greenwichmeantime.com.* (Accessed 7 May 2007)

Buckingham Palace Image. *[Online]. URL www.bcsu.net.* (Accessed 7 May 2007).

Central Intelligence Agency. (2006). Facts about Kenya. . *[Online]. URL www.cia.gov.uk/kenya.* (Accessed 26 March 2006).

Chen, G.M. & Starosta, W.J. (1998). *Foundations of Intercultural Communication.* Needcham Heights. MA: Allyn and Bacon pp 226.

Comstock, D.E. (1982). A method for critical research. *In E. Bredo and W Feinberg, Knowledge and values in social and educational research*, pp 370-390. Philadelphia: Temple University Press.

Correct Outloud. [Online]. URL www.sightandsound.co.uk. (Accessed 26 October 2004).

Coxhead, P. (2005). Natural Language Processing and Applications. URL www.birmingham.ac.uk. (Accessed 5 October 2005).

Cunliffe, D. (2007). Minority Languages and the Internet. In M. Cormack & N. Hourigan (Eds) *Minority Language Media: Concepts, Critiques and Case Studies.* Multilingual Matters, Clevedon. Pp 133-150.

Dankovicova, J. & Nolan, F. (1999). Some acoustic effects of speaking style on utterances for automatic speaker identification. *Journal of the International Phonetic Association.* **29**, (2), pp 115-128.

Davidson, J. (1994). The evaluation of computer-delivered natural speech in the teaching of reading. *Computers in Education Journal.* **22,** (1-2), pp 181-185.

Dearing (1997). Dearing Report. [Online] www.lifelonglearning.co.uk/dearing/. (Accessed 20 May 2008).

Denzin, N. (2000). The Practices and Policies of Interpretation. In N Denzin and Y Lincoln *Hand Book of Qualitative Research.* Sage publications, 2nd edition. Pp 897-922.

De Souza, M. & Dejean, P. (2000). Cultural Influence on Design. In D. Day, E. Del Galdo and G. Prabhu (Eds). *Designing for Global Markets 2: IWIPS' 2000 Proceedings.* The Second International Workshop on Internationlization of Products and Systems. Bakchuse Press, Maryland, USA. Pp 125-136.

Del Galdo, E. (1990). Internationalization and Translation: Some Guidelines for the Design of Human-Computer Interfaces. In J. Nielsen (Ed). *Designing User Interfaces for International Use.* Elsevier Science Publishers. Pp 1-10.

Del Galdo, E. (1996). Culture and Design. In E. Del Galdo and J. Nielsen (Eds). *International User Interfaces.* John Wiley & Sons. Pp 74-87.

Desurvire, H.W., Kondziela, J.M. & Atwood, M.E. (1992). What is gained and lost when using evaluation methods other than empirical testing. *Proceedings of the HCI'92 Conference on People and Computers VII.* Pp 89-102.

Dowhower, S.L. (1979). Effects of repeated reading on second-grade transitional readers' fluency and comprehension. *Reading Journal. 22, pp 389-407.*

Doyle, C. & Robson, K. (2002) *Accessible Curricula: Good Practice for All.* UWIC Press, **6.**

Dray, S., Siegel, A. & Kotze, P. (2003). Indra's Net: HCI in the Developing World. *Interactions.* **10,** (2), Pp 28-37.

Duncker, E. (2000). Cross-Cultural Use of Colours and Metaphors in Information Systems. In A. Smith (Ed). *Cultural Isues in HCI: One Day Workshop Sponsored by the British HCI Group and Optimum.Web Ltd.* University of Luton, UK. Pp 1-10.

Dunckley, L., Abdelnour-Nocera J., Oyugi C. & Camara S. Interim Technical Report. *VESEL – May 2007 Visit to Kenya User Sites.*

Dunckley, L. & Jheita, K. (2004). Formative Evaluation of International User Interfaces Based on Sorting Techniques. In V. Evers, E. Del Galdo, D. Cyr and C. Bonanni (Eds). *Designing for Global Markets 6: Culture, Trust and Design Innovation: IWIPS' 2004 Proceedings* The Sixth International Workshop on Internationalization of Products and Systems. BondRepro Publication. Vancouver, Cananda. Pp 91-102.

Dunckley, L. & Smith, A. (2000). A Cultural Dichotomies in User Evaluation of International Software. In D. Day, E. Del Galdo and G. Prabhu (Eds). *Designing for Global Markets 2: IWIPS' 2000 Proceedings.* The

Second International Workshop on Internatioanlization of Products and Systems. Backhouse Press, Maryland, USA. Pp 39-52.

Dutoit, T. (1997). *An introduction to text to speech synthesis.* Kluwer Academic Publishing.

Dutoit, T. (2002). Data driven techniques in speech synthesis. *Computational Liguistics.* ***4****, (29), pp 570-572.*

Economic and Social Data Service. (2005). [Online]. URL: www.esds.ac.uk (Accessed 20 June 2005).

Ellis, A. & Beattie, G. (1986). *The Psychology of Language and Communication.* Lawrence Erlbaum.

El Said, G.R. (2005). Cultural Effect on Electronic Consumer Behavour. PhD Thesis. Brunel University.

Erickson, F. (1986). Qualitative methods on research on teaching. *In M Whitrock (Ed), Handbook of research in teaching. 3rd edition*, pp 119-161. New York: Macmillan.

Erickson, F and SchuLtz, J.J. *The Counselor as Gatekeeper: social and cultural organization of communication in counselling interviews.* New York: Academic Press.

Ess, C. & Sudweeks, F. (2000). The Preface In F. Sudweeks and C. Ess (Eds). *Cultural Attitudes Towards Technology and Communication: CATaC' 2000 Proceedings.* The Second International Conference on Cultural Attitudes toward Communication and Technology. Murdoch University Publication. Perth, Australia.

European Telecommunications Standard Institute Technical Human Factors. (2005). *Human Factors; Multicultural and language aspects of multimedia communications,* pp 3-30.

Evers, V. (1999). *Cultural Differences in Understanding Human-Computer Interfaces.* CITE report No. 245. Centre for Information Technology in Education. The Institute of Educational Technology, The Open University, UK.

Evers, V. (2000). Cross-Cultural Understanding of Graphical Elements on the DirectED Website. In A. Smith (Ed). *Culturl Issues in HCI: One Day Workshop Sponsored by the British HCI Group and Optimum,Web Ltd.* University of Luton. Pp11-21.

Evers, V. (2002). Cross-Cultural Applicability of the User Evaluation Mehods: A Case Study amongst Japanese, North American, English and Dutch Users. In T. Loren (Ed). *Changing the World, Changing Ourselves: CHI' 2002 Proceedings.* The CHI Conference on Human Factors in Computing Systems, Menneapolis, USA.

Festival Speech Synthesis System. [Online]. URL www.cstr.ed.ac.uk. (Accessed 25 October 2004).

Florian, L. (2004). 'Uses of technology that support pupils with special educational needs' in Florian L, Hegarty J (Eds) *ICT and Special Educational Needs: a Tool for Inclusion.* Open University Press, Buckingham, pp 7-20.

Ford, G. & Gelderblom, H. (2003). The effects of culture on performance achieved through the use of human computer interaction. *ACM International Conference Proceedings.* **47,** pp 218-230.

Frederickson, N. & Cline, T. (2002). *Special Educational Needs, Inclusion and Diverstiy: A textbook.*_ Buckingham: Open University Press

Gakuru, M., Iraki, F.K., Tucker, R., Shalanova, K. & Ngugi, K. (2004). Design of Speech Data Base for Unit Selection in Kiswahili TTS. *E-Tech,* Kenya.

Gakuru, M., Iraki, F.K., Tucker, R., Shalanovam K. & Ngugi, K. (2005). Development of a Kiswahili Text to Speech System. *Interspeech,* Lisbon, Portugal.

Gakuru, M. & Tucker, R. (2009). Sustainability of Research and Development: A Case of Successful Technology Transfer in Spoken Language Technology. *CHI Conference Workshop. 2009.* Boston.

Galliers, R.D. (1992). Choosing Information Systems Research Approaches. In R.D. Gallier's (Ed). *Information Systems Research: Issues, Methods and Practical Guidelines.* Blackwell Scientific Publications.

Gay, G. (2000). *Culturally Responsive Teaching: Theory, Research and Practice.* New York: Teachers College Press, pp 111.

Gefen, D. & Ridings, C.M. (2005). If you spoke as she does, sir, instead of the way you do: a sociolinguistics perspective of gender differences in virtual communities. *ACM SIGMIS.* ***2**, (36), pp 78-92.*

Giddens, A. (1984). *The Constitution of Society: Outline of the Theory of Structure.* University of California Press.

Gillham, R. (2004). Towards a Strategic Model of Design Support for Localisation. In V. Evers, E. Del Galdo, D. Cyr and C. Bonnani (Eds). *Designing for Global Markets: Culture, Trust and Design Innovation: IWIPS' 2004 Proceedings.* The Sixth International Workshop on Internationalization of Products and Systems. BondRepro Publication, Vancourver Canada. Pp 103-113.

Glaser, B. & Strauss, A.L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Hawthorne, NY: Aldine de Gruyter.

Goldinger, S.D., Pisoni, D.B. & Logan, J.S. (1991). On the nature of talker variability effects on serial recall of spoken word lists. *JEP: Learning., Memory., and Cognition.* **17,** *Pp 152-162.*

Gray, W.D & Salzman, M.C. (1998). Damaged Merchandise? A review of experiments that compare usability evaluation methods. *Human Computer Interaction,* **13** (3), pp 203-261.

Griffith, T.L. (1998). Cross-cultural and cognitive issues in the implementation of new technology. *Interacting with Computers,* **9** (4), pp 431-447.

Guba, E.G. & Lincoln, Y.S. (1994). Competing Paradigms in Qualitative Researach. In N Denzin and Y Lincoln *Handbook of Qualitative Research.* Sage Publications. Pp 105-117

Gumperz, J.J. (1982). *Discourse Strategies*. Cambridge University Press.

Hall, E. (1959). *The Silent Language.* Doubleday, New York

Hall, E.T & Hall, M.R. (1987). Hidden Differences: Doing Business with the Japanese,. Doubleday, New York.

Hall, B. & Webb, D. (2000). Globalization Skills: A Great Niche in Great Demand. In D. Day, E. Del Galdo and G. Prabhu (Eds). *Designing for*

Global Markets 2: IWIPS' 2000 Proceedings. The Second International Workshop on Internationalization of Products and Systems. Backhouse Press, Maryland, USA. Pp 159-170.

Hariandja, J. & Daams, B. (2005). Cultural Differences in Product Use and User Needs: Cross-Cultural Usability Testing of a Food Processor in Indonesia and the Netherlands. In D. Day, V. Evers and E. Del Galdo (Eds). *Designing for Global Markets 7: Bridging Global Differences: IWIPS' 2005 Proceedings.* The Seventh International Workshop on Internationalization of Products and Systems. BondRepro Publication. Vancouver, Canada. Pp 5-6.

Harryson, B., Svenska, A. & Johansson, G.I. (2004). How people with development of disabilities navigate the Internet. *British Journal of Special Education.* **31**, (3), pp 138-142.

He, X. & Zhao, Y. (2007). Prior knowledge guided maximum expected likelihood based model selection and adaption for non-native speech recognition. **21**, (2), pp 247-265.

Heath, S. (1999). *Digital Processing of Speech.* Macmillan.

Heathcote, P. (2003). *Computing.* Payne-Gallway.

Hirschheim, R. (1992). Information Systems Epistemology: A Historical Perspective. *Information Systems Research: Issues, Methods and Practical Guidelines.* Blackwell Scientific Publications.

Hirson, R.E. (2002). Excellence and Diversity. *ABA Journal,* **88** (6).

Hoebel, E.A & Frost, E.L. (1976). *Culture and Social Anthropology.* New York: McGraw-Hill, **58.**

Hofstede, G. (1980). *Culture's Consequences.* International Differences in Work-Related Values. Sage, New York.

Hofstede, G. (1991). *Cultures and Organisations: Software of the Mind.* McGraw-Hill.

Hoft, N. (1996). Developing a Cultural Model. In E. Del Galdo and J. Nielsen (Eds). *International User Interfaces.* John Wiley and Sons. Pp 41-73.

Hore, N. (2006). *Rapid English Software Discussion.* [Personal interview].

How, P.H. (2003). *Statistical Analysis, Modelling and Synthesis of Voice for Text Speech Synthesis.* PhD thesis. Brunel University.

Huckvale, M. (2002). *Improvements in Speech Synthesis.* Wiley.

Human Computer Interaction Group. [Online]. URL www.bcs.or.guk/hci. (Accessed 25 October 2004).

Human Computer Interfaces for the Visually Impaired. URL. www.rnib.org.uk. (Accessed 26 October 2004).

Inclusion and Special Educaional Needs. [Online]. URL www.naldic.org.uk/ITTSEAL2/teaching/Inclusion.crfm. (Accessed 13 June 2007).

Information Commissioner. (1988). *Data Protection Act.* [Online]. URL: www.informationcommissioner.gov.uk (Accessed 20 June 2005).

Janesick, V. (2000). The Choreography of Qualitative Research Design. *Handbook of Qualitative Research.* Sage publications, 2nd edition.

Jeffries, R., Miller, J.R., Wharton, C., & Uyeda, K.M. (1991). User Interface evaluation in the real world: A comparison of four techniques. *Proceedings of the ACM CHI'91 Conference on Human Factors in Computing Systems.* ACM Press, New York. Pp 119-124.

Jurafsky, D. & Martin, J.H. (2000). *Speech and Language Processing.* Prentice Hall.

Jones, C., Berry, L. & Stevens, C. (2008). Speech intelligibility and persuasion: speech rate and non-native listeners. *Computer Speech and Language Journal.* **21,** (4), pp 641-651.

Jonsson, I.M., Zajicek, M, Vass, C. & Sharp, H. (2005). "Improving automotive safety by pairing driver emotion and car voice emotion". *Proceedings of CHI 2005.*

Karat, C-M., Campbell, R. & Fiegel, T. (1992). Comparison of empirical testing and walkthrough methods in user interface evaluation. *Proceedings of the ACM CHI'92 Conference on Human Factors in Computing Systems.* ACM Press, New York. Pp 397-404.

Katz, J. (1981). *Language and Other Abstract Objects.* Rowman and Littlefield, Totawa, New Jersey.

Keates, A. (2000). *Dyslexia and Information and Communications Technology.* David Futon Publishers, London.

Keesing, F.M. (1965). *Cultural Anthropology: The Science of Custom.* New York: Holt, Rinehart and Winston, **18.**

Keller, E. (2002). *Future Directions of Research in Speech Synthesis.* : The Naturalness of Synthetic Speech. Wiley.

Kempton, W. (1981). *The Rhythmic Basis of Interactional Synchrony.* Trinity University Press.

Kendon, A., Harris, R.M. and Key, M.R. (1975). *Organization of Behavior in Face-to-Face Interaction.* The Haugue: Mouton.

Kluckhohn. J. (1962). *Culture and Behaviour.* University of Arizona Press, Tucson.

Kralisch, A. & Bettina, B. (2004). Cultural Determinants of Search Behaviour on Websites. In V. Evers, E. Del Galdo, D. Cyr and C. Bonanni (Eds). *Designing for Global Markets 6: Culture, Trust and Design Innovation. IWIPS' 2004 Proceedings.* The Sixth International Workshop on Internationalization of Products and Systems. BondRepro Publication. Vancouver, Canada Pp 61-74.

Krathwohl, D. (1997) *Methods of Educational and Social Science Research: An Integrated Approach.* Addison Wesley Longman, 2nd edition.

Kreitzberg, C. (1996). Managing for Usability. In A. Alber (Ed). *Multimedia: A Management Perspective.* Wadsworth Publishing Company, 1st Edition. Pp 65-88.

Kroeber, A.L. & Parsons, T. (1958). The concepts of culture and of social systems. *American sociological review, **23**, pp 582.*

Kuhn, T. *(1996)..* *The Structure of Scientific Revolutions.* University of Chicago Press, 3rd edition.

Land, F. (1992). The Information Systems Domain. In RD Galliers *Information Systems Research: Issues, Methods and Practical Guidelines.* Blackwell Scientific Publications.

Larcher, J. (2000). 'Information technology for children with language difficulties', in Rinaldi W (Eds) *Language Difficulties in an Educational Context*. Whurr Publishers Ltd, London.

Lazarsfeld, P.F. & Merton R.K. (1964). "Friendship as Social Process: A Substantive and Methodological Analysis," in Monroe Berger and other (eds.), *Freedom and Control in Modern Society*, New York: Octagon, 1964.

Lewin, C. (1998). Talking Book Design: What do practitioners want? *Computers in Education Journal.* **30,** (1-2), pp 87-94.

Lincoln, Y. & Guba, E. (2000). Paradigmatic Controversies, Contradictions, and Emerging Confluence. In N Denzin and Y Lincoln. *Handbook of Qualitative Research.* Sage publictions, 2nd edition. Pp 163-188.

Luckin, R. (2008). The Learner Centric Ecology of Resources: a Framework for using Technology to Scaffold Learning. *Computers and Education, **50,** pp 449-462.*

Luckman, J. (1999). *Transcultural Communication in Nursing.* Albany, NY: Delmar Publishers pp 22.

Marcus, A. & Gould, E.W. (2000). Cultural Dimensions and Global Web User-Interface Design. *Interactions.* Pp 33-46.

Marcus, A., Baumgartner, V. & Chen, E. (2003). User Interface Design vs. Culture. In V. Evers, K. Roese, P. Honold, J. Coronado and D. Day (Eds). *Designing for Global Markets 6: IWIPS' 2003 Proceedings.* The Fifth International Workshop on Internationalization of Products and Systems. The Publisher of the University of Kaiserslauthern, Berlin, Germany. Pp 67-78.

McSweeney, B. (2002). Hofstede's 'Model of National Cultural Differences and Consequences: A Triumph of Faith – A Failure of Analysis. *Human Relations.* **55,** (1), pp 89-118.

Ministry of Agricultural and Rural Development. (2003). *Fruits and Vegetables Technical Handbook.* Second edition (revised). The Ministry of Agriculture and Rural Development, Nairobi, Kenya.

Moore, R. (2006). *Discussion on meeting the needs of learners with profound special educational needs through the use of assistive technology.* [Personal interview].

Moustroufas, N. & Digalakis, V. (2007). Automatic pronounciation evaluation of foreign speakers using unknown text. *Computer Speech and Language Journal.* **21,** (1), pp 219-230.

Mulleninix, J.W., Pisoni, D.B. & Martin, C.S. (1989). Some effects of talker variability on spoken word recognition. *Journal of Acoustic Society.* ***85,*** Pp 365-378.

Murray, I.R. & Arnott, J.L. (2008). Applying an analysis of acted vocal emotions to improve the simulation of synthetic speech. **22,** (2), pp 107-129.
National Farmers Information Service. (2008). [Online] *www.nfis.go.ke.* (Accesssed 20 May 2008)

Nielsen, J. (1990). Traditional Dialogue Design Applied to Modern User Interfaces. *Communications of the ACM.* ***33,*** *(10), Pp 109-118*

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. *Proceedings of the ACM CHI'92 Conference on Human Factors in Computing Systems.* ACM Press, New York. Pp 373-380.

Nielsen, J. (1993). *Usability Engineering.* Academic Press.

Nielsen, J. (2000). *Designing Web Usability: The Practice of Simplicity.* New Riders Press.

Norman, D. (1998). *The Design of Everyday Things.* The MIT Press.

Nygaard L.C., Sommers, M.S. & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Perception and Psychology.* **5,** Pp 42-46.

Orlikowski, W. & Broudi, J. (1991). Studying Information Technology in Organizations: Research Approaches and Assumptions. *Information systems research,* 2 (1), 1-28.

Owens. P. (1983). *Digital Processing of Speech.* Macmillan.

Papamichalis, P. E. (1987). *Practical Approaches to Speech Coding.* Prentice-Hall.

Pisoni, D.B., Saldana, H.M. & Sheffert, S.M. (1996). Multi-modal encoding of speech in memory: A first report. *ICLSLP Proceedings, Fourth International Conference.* **3,** Pp 1664-1667.

Reinking , D. (1986). Computers, reading and a new technology of print. In *Reading and Computers: Issues for Theory and Practice.* Teachers College Press, New York.

Reinking , D. & Bridwell-Bowles, L. (1991). Computers in reading and writing. In *Handbook of Reading Research, Vol 2 ed.* R Barr, Longman.

Reitsma, P. (1988). Reading practice for beginners: effects of guided reading, reading while listening and independent reading with computer based speech feedback. *Reading Research Quarterly*, **23**, (2), pp 219-235.

Research Access and Preservation. [Online]. URL www.esds.ac.uk (Accessed 20 June 2005).

Researching HCI. [Online]. URL www.usability.net. (Accessed 25 October 2004).

Resources for Visually Impaired. [Online]. URL www.rnib.org.uk. (Accessed 26 October 2004).

Robbie Williams Image. *[Online]. www.semantic.co.uk.* (Accessed 7 May 2007)

Rogers , E.M. & Bhowmik, D.K. (1970). Homophily-Heterophily: Relational Concepts for Communication. *Public Opinion Quarterly, **34,*** pp 523, Columbia University.

Rogers, E.M & Steinfatt, T.M. (1998). *Intercultural Communication.* Prospect Heights, IL: Waveland Press, pp 135.

Rowan, M., Gregor, P., Sloan, D. & Booth, P. (2000). Evaluating web resources for disability access. *Proceedings of the 4th International Conference on Assistive Technologies,* pp 80-84. Arlington, Virginia, US.

Rubin, B. (1988). *Communication and Human Behaviour, 3rd Edition.* New York: Macmillan, pp 384.

Sagisaka, N., Kaiki, N., Iwahaski, N. & Mimura, K. (1992). ATR – v.TALK speech synthesis system. *ICSLP.* (1), pp 483 – 486.

Samovar, L.A. & Porter R.E. (2004). *Communication Between Cultures, 5th Edition.* Thomas Wadsworth.

Schaudt, B.A. (1987). The use of computers in a direct instruction reading lesson. *Reading Psychology.* **8,** pp 169-178

Schrier, J.R. (1992). Reducing stress associated with participating in a usability test. *Proceedings of the Human Factors Society 36th Annual Meeting.* Pp 1210-1214.

Screenreaders. [Online]. URL www.screenreader.co.uk. (Accessed 26 October 2004).

Shalanova , K. & Tucker R. (2004). *Issues in Porting TTS to Minority Languages.* SALTMIL workshop on Minority Languages. Lisbon.

Shneiderman, B. (1998). *Designing for the User Interace: Strategies for Effective Human-Computer Interaction.* Addison-Wesley Longman, 3rd Edition.

Smith, A. & Dunckley, L. (1998). Using the LUCID method to Optimize the Acceptability of Shared Interfaces. *Interacting with Computers.* **9,** pp 335-345.

Speech Generating Device. [Online[. URL www.speechgeneratingdevice.com. (Accessed 25 October 2004).

Speech Synthesis. [Online]. URL www.phon.ucl.ac.uk/research/speechsynth.html. (Accessed 25 October 2004).

Sutcliffe, J. & Jacobsen, Y. (2001). *All Things Being Equal:* A Practical Guide To Widening Participation for Adults with Learning Difficulties in Continuing Education. Leciester: NIACE.

Swahili Language. (2006). *[Online].* URL www.en.wikipedia.org/wiki/Swahili (Accessed 26 March 2006)

Tatham, M. (1986). The Problem of Capturing Linguistic and Phonetic Knowledge. *Proceedings of the Institute of Acoustics.* **8,** (7), pp443-450

Text to Speech Systems. [Online]. www.keyspell.com. (Accessed 25 October 2004).

The Data Protection Act .(1988). *[Online]. www.opsi.gov.ukActs/Act1998 /UKpga_19980029-en.* (Accessed 17 June 2008).

The Design for All Project Team Executive Report 2000. [Online]. URL: www.dfaei.org.docs/D2.3.pdf. (Accessed 30 April 2008).

The Disability Discrimination Act 1995. [Online]. http://www.opsi.gov.uk/actst/acts1995/ ukpga_19950050_en_1 (Accessed 30 April 2008).

The Disability Rights Commission. (2004). *[Online]. www/p[so/gpv/ilACTS/acts1999/ukpga_1990017_en_1.* (Accessed 17 June 2008).

The Economic and Social Data Service (2005). *[Online*]. www.esds.ac.uk.. (Accessed 17 June 2008).

The Education Act 1996. [Online]. URL: http://www.opsi.gov.uk/Acts/acts1996/ukpga_19960056_en_1 (Accessed 30 April 2008).

The eEurope Action Plan 2002. [Online]. URL: http://europa.eu/rraid/presReleasesAction.do?reference=IP/03/425&format=HTML&aged=1&language=EN&guiLanguage=en. (Accessed 30 April 2008).

The eGovernment Ministerial Declaration 2001. [Online]. URL: http:..archive.cabinetoffice.gov.uk/egov2005conference/about/index.asp. (Accessed 30 April 2008).

The Further and Higher Education Act 1992. [Online]. www.opsi.gov.uk/acts/acts1992/UKpga_19920013_en_1 (Accessed 30 April 2008).

The Information Commissioner. (2005). *[Online] www.ico.gov.uk.* (Accessed 17 June 2008).

The United Nations Educational Scientific and Cultural Organisation 1994. [Online]. URL: *http://earli2007.hu/nq/home/scientific_program* (Accessed 30 April 2008).

The Special Educational Needs and Disability Act 2001. [Online]. URL: http://www.opsi.gov.uk/ACTS/acts2001/ukpga_20010010_en_1 (Accessed 30 April 2008)

Tower of London Image. *[Online]. www.historic.uk.com.* (Accessed 7 May 2007)

Trandis, H.C. (1972). *The Analysis of Subjective Culture.* New York: Wiley-Interscience.

Trompenaars, (1993). *Riding the Waves of Culture.* Nicholas Brearley Publishing, London.

United Nations Educational Scientific and Cultural Organisation. (1994). *Salamanca World Statement.* (1994). **11.**

Tucker, R. & Shalonova , K. (2005) Supporting the Creation of TTS for Local Language Voice Information Systems. *Interspeech,* Lisbon, Portugal.

Using Keystone. [Online]. URL www.abilitynet.co.uk. (Accessed 25 October 2004).

Victor, S. (2001). Election 2000 and the Culture War. *The Humanist,* ***5.***

Viswanathan, M., & Viswanathan, M. (2005). Measuring speech quality for text to speech synthesis: development and assessment of a modified mean opinion score (MOS) scale. *Computer Speech and Language Journal.* **19,** (1), pp 55-83.

Vocal Anatomy. [Online]. URL www.vocalfocus.com/vocal-anatomy.html. (Accessed 15 June 2005).

Vygotsky, L.S. (1978). *Interaction between Learning and Development.* Cambridge. MA: Harvard University Press.

Vygotsky, L.S. (1980). *Mind in society : The development of higher psychological processes. Thought and Language.* MIT Press.

Vygotsky, L.S. (1986). *Thought and Language.* MIT Press.

Vygotsky, L.S, (1987). *Problems of general psychology.* New York: Plenum Press.

Whitehead, M. & Hore, N. (2007). Rapid English project. *ALT/QIA Making e-learning work conference, Nottingham.*

Whorf, B. L. (1956). *Language, Thought and Reality: Selected writings of Benjamin Lee Whorf.* The MIT Press, Cambridge, MA.

Willder, B. (2002) *'Disability legislation: implications for learning technologists in the UK'* in Phipps, L., Sutherland, A. and Brave et al, J. Access All Areas'.

Williams, P., Jamali, H.R. & Nicholas, D. (2006). Using ICT with people with SEN: what the literature tells us. *Aslib Proceedings.* **11,** (4), pp 330-345.

Winfield, I. (1990). *Organizations and Information Technology: Systems, Power and Job Design.* Alfred Waller Ltd.

Winschies, H. (2007). Cultural Appropriation of Usability Engineering and Participatory Design Practices – A Namibian Experience. CHI User Centred Design and International Development. *[Online] htttp://wwwmikebinta.gatech.edu/UCDandIDWorkshop/papers/Winschies .pdf.* [Accessed 28 August, 2008*).*

Winschies, H. (2006). The Challenges of Participatory Design in an Intercultural Context: Designing for Usability in Namibia. *Participatory Design 2006, Trento, Italy.*

Winschies, H. & Paterson, B. (2004). Sustainable Software Development in Fulfilling the Promises of ICT. *SAICSIT 2004.* Stellenbosch, South Africa.

Winschies, H & Paterson, B. (2007). Assumptions Considered Harmful – The Need to Redefine Usability. *HCII 2007.*

Wise, B., Olson, R., Ansett, M., Andrews, L., Terjak, M., Schneider, V., Kostuch, J. & Kritho, L. (1989). Implementing a long-term computerised remedial reading program with synthetic speech feedback: hardware, software and real world issues. *Behaviour Reserch Methods, Instruments and Computers,* ***21***, pp 173-180.

Wood, D.J. (1980). *Teaching the young child: some relationships between social interaction, language and thought.* In Olsen, D. (Ed) Social Foundations of language and cognition: Essays in Honor of J.S. Bruner. New York: Norton.

Wood, D.J , Bruner, J.S. & Ross, G. (1976). *The role of tutoring in problem solving.* Journal of Child Psychology and Psychiatry, **17,** 2, pp 89-100.

World 66. (2006). . [Online]. URL www.world66.com/africa/kenya/language (Acessed 29 March 2006).

Yeo, A.W. & Barbour R.H. (1998). Strategies of internationalisation and localisation: A postmodernist's perspective. *Interacting with Computers, special issue: Shared values and shared interfaces.*

Yeo, A. & Loo, W. (2004). Identification and Evaluation of the Classification Schemes: A User Centred Approach. In V. Evers, E. Del Galdo, D. Cyr and C. Bonanni (Eds). *Designing for Global Markets 6: Culture, Trust and Design Innovation: IWIPS' 2004 Proceedings.* The Sixth International Workshop on Internationalization of Products and Systems. BondRepro Publication, Vancouver, Canada. Pp 75-87.

Yoon, K. (2008). Design and evaluation of prosodically-sensitive concatenative units for a Korean TTS system. *Computer Speech and Language Journal.* **22,** (3), pp 273-294.

Zajicek, M. & Hall, S. (2000). Solutions for elderly visually impaired people using the Internet. *HCI*, Sunderland, pp 299-307.

Zajicek, M. & Jonsson, I. (2006) In-car Speech Systems for Older Adults: Can They Help and Does the Voice Matter? *International Journal of Technology Knowledge and Society, Hyderabad.*

Faculty of Professional Studies

PARTICIPANT INFORMATION SHEET

Text to Speech Technology: usability in multi-cultural environments as a teaching and learning tool for end users with special educational needs

You are being invited to take part in usability tests where you will hear different words spoken which have been created using different types of synthetic speech which represent a sample of different gender and cultural backgrounds. The results of the tests will be analysed and evaluated to see if the level of accuracy and speed is better for people when the synthetic speech sounds as though it is from the same gender and cultural group as the person taking the test. Before you decide it is important for you to understand why the evaluation is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. **Thank you for reading this.**

1) **What is the purpose of the study?**

The purpose of the study is to find out whether the accuracy level and speed to complete a task using text to speech synthesis is higher when people hear synthetic speech which sounds as though it is from someone who is the same gender and cultural group as them.

2) **Why have I been chosen?**

You have been asked to participate because you are studying at Thames Valley University and the results of the tests may be used to advise the University and the wider world on the use of Text to Speech Synthesis in the future which may help you and others to prepare assignments and to use this technology in other contexts. In addition, this will enhance your knowledge and understanding of the cutting edge developments and research into computer science in the twenty first century.

3) **Do I have to take part?**

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason.

What will I be asked to do if I take part?

You will be asked to participate in some tests that are called usability tests. This will involve coming to one of the computer rooms with about 14 other people. You will be given two sheets of paper with words and tick boxes on them and words to circle. You will then hear two words at a time from the loudspeakers from the computer at the front of the room. You have to tick the box, or circle the word, that you thought you heard. The synthetic speech that you will hear will be both male and female and from different cultural backgrounds. This will be for a maximum of 45 minutes.

It is important to note at all times that it is the software that is being tested, not you. None of the tests are IQ tests, they are just identifying the word that you thought you heard. The results will be totally confidential and are only going to be collated with other results and used for the research project. At no time, will your results be identifiable or disclosed to any of your lecturers.

4) What will happen to the results of the research study?

The results of this phase of the evaluation will be used to provide clarity and validity to the subsequent phases of the study.

5) Who is organising and funding the research?

The research is being funded by Thames Valley University. The research is being organised by Yvonne Spittles who has a Master of Science degree in Information Technology and is experienced in conducting usability testing on software. The research is being supervised by Professor Lynne Dunckley who is a Professor in Information Technology and is extremely well qualified and experienced in this field.

6) Who has reviewed the study?

This study has been reviewed and approved by the Thames Valley University Research Degrees Committee and the Thames Valley University Faculty of Health and Human Sciences Research Review Committee and by Senior Academics in the Faculty of Professional Studies at Thames Valley University.

7) Contact for Further Information

If you would like any further information please contact the researcher as follows

Yvonne Spittles

Telephone: 0118 967 5423

E-mail: yvonne.spittles@tvu.ac.uk

Thank you for reading this.

Thames Valley University
Faculty of Professional Studies

Text to Speech Technology: usability in multi-cultural environments as a teaching and learning tool for end users with special educational needs

CONSENT FORM

Name of Lead Investigators: Yvonne Spittles

Please Initial

1. I confirm that I have read and understand the information sheet dated 20 October 2005 (version 1) for the above study and have had the opportunity to ask questions.

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

3. I agree to take part in the above study.

Name of Participant Date

Signature

Name of Person taking consent Date

(if different from Researcher)

Signature

Address..

....

Signed... (Researcher)..

Date.......................................

SCRIPT TO BE READ WHEN UNDERTAKING THE PILOT TESTS

Before the test

Good morning/afternoon. Thank you very much for agreeing to participate in these tests. Your co-operation in taking part will be of considerable help to me with my research.

Please remember that it is the software that is being tested – not you. Your results will remain strictly anonymous and confidential.

The entire procedure will take a maximum of 45 minutes. At the end, there will be complimentary beverages for you. You are free to leave at any time, and if you wish to do so, please put up your hand.

Has anyone got any questions before we proceed any further?

During the test

I have set up two tasks which I would like you to complete during this session.

You do not have to use the computers, but you are asked instead to mark your answers on the sheets I am going to hand out to you.

You will hear synthetic speech coming from the loudspeakers. This speech is from male and female voices and from some cultures – not all cultures are represented and this may be because the synthetic speech is not available yet.

Test 1

I am now going to hand out the task sheets for the first test. You will see that there are 10 questions listed with choices of Speech A, Speech B or Speech C.

You will hear each sentence three times using different voices of synthetic speech with the correct missing word at the end. Please tick in the box where you think you heard the correct word to complete the sentence.

Does anyone have any questions before we start?

The testing starts now.

The time allowed for this test is 10 minutes. Using the wall clock in the room for consistency, please write the start time on your sheet in the bottom left hand corner. When you finish, please write the finish time on your sheet in the bottom right hand corner.

Time is now up and I will take the papers in from you, thank you. Has anyone got any questions before we move on to the next task?

Test 2

I am now going to hand out the task sheets for the second test. This time, there are 20 question numbers with word A or word B next to them. There are not any sentences this time.

You will hear one word at a time, using different voices of synthetic speech. Please circle the word that you thought you heard for each question.

The time allowed for this test is 10 minutes. Using the wall clock in the room for consistency, please write the start time on your sheet in the bottom left hand corner. When you finish, please write the finish time on your sheet in the bottom right hand corner.

Does anyone have any questions before we start?

The testing starts now.

Time is now up and I will take the papers in from you, thank you.

After the test

That now completes all of the testing. Thank you very much, once again, for taking part. You will have played a key part in helping to improve the quality of text to speech synthesis development.

I will contact you again in the very near future to let you know the general outcome of the tests but individuals will not be identified or referred to in any way.

If you would now like to make your way to the Ritazza cafeteria, there is complimentary beverages and snacks for you. Good morning/afternoon.

ID number of participant:

QUESTIONS FOR PARTICIPATION IN SPEECH SYNTHESIS RESEARCH

TASK ONE

Project Title: Text to Speech Technology: usability in multi-cultural environments as a teaching and learning tool for end users with special educational needs

Name of Researcher: Yvonne Spittles MSc MSBT MIVA

Institution: Thames Valley University

Contact Details: Yvonne.Spittles@tvu.ac.uk

Instructions for Task One

You will hear each sentence three times using different voices of synthetic speech. Please indicate with a tick in the appropriate box whether you think the correct word to answer the sentence was produced by speech A, speech B or speech C.

The maximum time allowed for this is ten minutes.

Question No	Question	Speech A	Speech B	Speech C
1	Vegetarians do not eat			
2	Wild animals can be kept in a ...			
3	........is pleasant to drink			
4	Midday is at			
5	A young horse is called a			
6	They are going shopping to buy new			
7	The friends enjoyed watching the			
8	Ducks swim in a pond......			
9	Bread is made from			
10	The fish were caught in a			

Banana information service

Start

Welcome. For English, press 1. For Swahili, press 2.

1. Main menu

Welcome to the banana information service. If you want to hear the information people ask for often, you can press 9. Otherwise you can choose which stage of banana growing you want information on from the following options.

For ideal climatic and soil conditions for growing bananas, press 1. (See section 1.1)
For information about different banana varieties, press 2. (See section 1.2)
For planting material, press 3. (See section 1.3)
For planting and land preparation, press 4. (See section 1.4)
For maintaining your plants, press 5. (See section 1.5)
For harvesting, press 6. (See section 1.6)
For pests and diseases, press 7. (See section 1.7)
For market information, press 8. (See section 1.8)
For information people ask for often, press 9. (See section 1.9)
At any time in the conversation you can interrupt by pressing the star key and go back to the previous menu. You can also choose to repeat the information you have just been given by pressing 0. When you want to end the conversation you can simply hang up.

1.1 Ideal climatic and soil conditions

The crop grows well in fairly hot and humid areas that are within an altitude of 0 to 1,800 meters above sea level. Rainfall has to be at least 1,000 millimeters per year. This is more crucial at flowering time. Irrigation is therefore necessary in low rainfall areas. Bananas can be grown near riverbanks, but avoid growing bananas on exposed sites, as they are sensitive to strong wind. It is recommended to plant in wind-sheltered positions and in blocks rather than strips so the plants can protect each other against winds. Deep, well-drained fertile soils are ideal for banana growing as the crop cannot withstand water logged soil conditions. pH requirements range between 6 and 7.5. Lime can be added to soils that are acidic in order to make them less acidic and better suited for banana growing. The main areas of production in Kenya are Coast Western, Nyanza, Eastern and Central provinces. For more information, contact your local agricultural extension officer.
For information about different varieties of bananas, press 1. (See section 1.1.1)

1.2 Varieties

There are many varieties grown, some for cooking and others for dessert purposes. Some of the high yielding varieties for cooking are Uganda green (matoke), Kisii matoke, Ndizi Ngombe and Kikuyu 1. The varieties used for dessert purposes are all supposed to be yellow when they are ripened, since that is what the market prefers. The Cavendish varieties are Giant Cavendish, Williams Hybrid, Grand Nain and Dwarf Cavendish. Other ripening bananas are Sweet banana, Lacatan, Valery and Paz. Gros Michel, known locally as Kampala, is a very popular banana for ripening, but it is highly susceptible to Fusarium Wilt, also known as the Panama Disease. For more information, contact your local agricultural extension officer.

1.3 Planting material

There are two types of planting material. One material is sword suckers, which are shoots with narrow leaves. These are taken from parent plants when about 1 meter high and 15 centimeters diameter at the base. The other alternative planting material are Bull-heads from the rootstock of the parent plant. Bull-heads are used only if the sword and maiden suckers are unavailable. Propagation of bananas using tissue culture, TC, is done in a laboratory and is fast gaining popularity. Tissue culture has numerous advantages of other conventional methods of propagation. Tissue culture facilitates rapid production of quality planting materials. The TC banana seedlings should be 200 to 300 millimeters high uniform, and having at least five healthy dark leaves and wider inter-nodes at the time of transplanting. You can buy TC banana seedlings at Jomo Kenyatta University of Agriculture and Technology, at Jomo Kenyatta Village nurseries or at KARI Thika, at a cost of 75 to 100 KSh per sucker. You can get sword suckers from neighbours or buy from other farms at a cost of 20 to 40 KSh per sucker. Sword suckers may be carrying earthborn diseases such as Nematodes and can be prevented. For more information, contact your local agricultural extension officer.

To hear how to prevent and control earthborn diseases, press 1 (See section 1.7.1.3)

1.4 Planting and land preparation

Plough the land during the dry season and remove or kill all weeds, especially couch grass.

Planting should be done at the beginning of the long rains. Holes should be dug 60 centimeters deep and be 60 centimeters wide. Fill the holes with topsoil mixed with 2 debes of well-rotten farmyard manure and 226 grams of double superphosphate, DSP, per hole. Plant the suckers some 30 centimeters deep so that it is firm in the soil. The spacing depends on the variety. The short varieties should be 3 times 3 meters while the medium varieties should be 3 times 4 meters. The tall varieties are spaced at 4 times 4 meters. Bananas are sensitive to strong wind. Planting in wind-sheltered positions and in blocks rather than strips is recommended. If planted in blocks the plant protect each other against winds. Intercrop with quick growing and shallow rooted crops in the first year. Fertilizer application is recommended for banana growth. Mix 100 to 150 grams of DSP per hole with the soil at planting time. For more information, contact your local agricultural extension officer.

For information about planting material, press 1. (See section 1.3)

1.5 Maintaining your plants

Uproot any weeds by hoeing, especially when the plants are young. When the plants mature they will cover the ground and out compete any weeds. Care should be taken not to disturb the roots which are found in the top 15 centimeters of the soil. The first year, you should intercrop with quick growing and shallow rooted crops such as beans and groundnuts. After that, mulch the orchard by using stems cut during thinning. Prune all the dead leaves continuously. Fertilizer application is recommended for banana growth. Apply 100 to 125 grams of Calcium Ammonium Nitrate, C.A.N., per stool per year at the beginning of the rains. This is scattered top-dressed in a circle 2 feet away from the banana stool. Thin the orchard in order to produce large bunches and to increase yields. Generally the banana stools should be thinned to leave one bearing plant, one half grown, and one just starting to grow. Stems bearing bunches should be supported by sticks or poles to stop immature bunches from falling. Thinning tissue culture bananas is critical, but special care must be taken.
If you have any questions, you can contact your local agricultural extension officer.

For information about thinning tissue culture bananas, press 1. (See section 1.5.1)
For more information why mulching is beneficial, press 2. (See section 1.5.2)

1.5.1 Thinning of tissue culture bananas

Tissue culture bananas produce large numbers of suckers 1 to 2 months after planting. These suckers should be cut off at the ground level to allow the development of mother plant. When the mother plant has reached 1 meter tall, you should select 1 vigorous sword sucker facing eastward or up the slope, where the land is sloppy. All other suckers should be cut at the base, gorged out in the middle and killed with 2 milliliters of kerosene or diesel. When the mother plant start to flower select 1 sucker in the direction of the first selected sucker. If the sucker selection is properly done, there should be a bearing mother plant, a large daughter sucker and a small grand daughter, all aligned in one direction. For more information, contact your local agricultural extension officer.

1.5.2 Mulching

Mulching is beneficial because it retains moisture, protects the soil from erosion and compaction and increases the organic matter content of the soil as it rots down. It also suppresses weed growth. Mulch the orchard by using stems cut during thinning. For more information, contact your local agricultural extension officer.

1.6 Harvesting

Harvesting starts 13 to 18 months after planting. Cut the stem when the fruit is fully developed and it is light green and shiny in appearance. Avoid bruising the bananas at harvesting, since they will ripen faster and perhaps even rotten at the points of the bruises. Clean and wash the bananas to make them more attractive on the market. When transporting, you can wrap the bunches in grass or banana leaves to avoid bruising. Other ways of making the bananas stay longer are to package the bunches separately and to let them hang instead of lying down. Yields of 17 to 20 tons per hectare per year can be obtained and up to 40-50 tons for well-managed orchards. After about 8 to 10 years of production, a decline in productivity of banana is usually experienced. Replacement of the plantation is usually recommended. If the same plot is to be used, soil fumigation is recommended before establishment of another plantation. For more information, contact your local agricultural extension officer.

1.7 Pests and diseases

For information about control of different pests and diseases, press 1. (See section 1.7.1)
For an overview of the symptoms of different pests and diseases, press 2. (See section 1.7.2)
For information about banana weevils, press 3. (See section 1.7.3)
For information about banana silvering thrips, press 4. (See section 1.7.4)
For information about nematodes, press 5. (See section 1.7.5)
For information about the cigar-end-rot disease, press 6. (See section 1.7.6)
For information about the panama disease, press 7. (See section 1.7.7)
For information about sigatoke leaf spots, press 8. (See section 1.7.8)

1.7.1 Control of pests and diseases

Choose which pest or disease you want information on.
For banana weevil, press 1. (See section 1.7.1.1)
For banana silvering thrips, press 2. (See section 1.7.1.2)
For nematodes, press 3. (See section 1.7.1.3)
For cigar-end-rot disease, press 4. (See section 1.7.1.4)
For panama disease, press 5. (See section 1.7.1.5)
For sigatoke leaf spots, press 6. (See section 1.7.1.6)

1.7.1.1 Control of Banana Weevil

Plant clean healthy suckers. Suckers suspected of being infected should be dipped in a mixture of 100 milliliters of 18% Dieldrin in 20 liters of water. Cut all old stems at ground level and dust cut rhizomes with 5% Dieldrin and cover with soil. Sprinkle 5% Dieldrin or Aldrin dust around the bases of the stem. For more information, contact your local agricultural extension officer.
For information and symptoms of banana weevil, press 1. (See section 1.7.3)

1.7.1.2 Control of Banana Silvering Thrips

Spray with Carbaryl, Dichlorvos Fenitrothion or Diazinon. For more information, contact your local agricultural extension officer.
For information and symptoms of banana silvering thrips, press 1. (See section 1.7.4)

1.7.1.3 Control of Nematodes

Nematodes may be controlled by planting pest free suckers. Remove all roots from the suckers before planting and kill any pests on the suckers by the hot water method. Heat water until a candle just melts in it, which is at 55 degrees centigrade. Remove the water from the fire and put the suckers in the hot water for 20 minutes. Nematodes may also be controlled by strict crop rotation and by fumigating the soil with Furadan or Nemacur. For more information, contact your local agricultural extension officer.
For information and symptoms of nematodes, press 1. (See section 1.7.5)

1.7.1.4 Control of the Cigar-end-rot disease

Control by removing dry floral plants by hand 8 to 11 days after fruit bunch emergence. For more information, contact your local agricultural extension officer.
For information and symptoms of the cigar-end-rot disease, press 1. (See section 1.7.6)

1.7.1.5 Control of the Panama Disease

Control by planting resistant varieties like Cavendish, Kisigame, and Uganda green (Matoke). For more information, contact your local agricultural extension officer.
For information and symptoms of the panama disease, press 1. (See section 1.7.7)

1.7.1.6 Control of Sigatoke Leaf spots

Remove badly spotted leaves and leaf trash and destroy them. Avoid over-crowding of plants. Chemically, Mancozeb can be used to control the disease. For more information, contact your local agricultural extension officer.
For information and symptoms of sigatoke leaf spots, press 1. (See section 1.7.8)

1.7.2 Overview of symptoms

If the banana leaves turn yellow, wither and die prematurely, it can be a sign of banana weevil, so press 1 for more info. If you find silvery patches on fruits which later turn brown, it can be a sign of banana silvering thrips, so press 2 for more info. If you find yellowing of leaves, growth ceases and that the rhizome development is retarded, it can be a sign of nematodes, so press 3 for more info. If the fruit tips undergo a dry rot with an ashy appearance that look like a cigar, it is a sign of cigar-end-rot, so press 4 for more info. If the older leaves turn yellow, and collapse while still green at the base and the plants fail to produce normal fruit and die before the fruit stalk is fully developed, it can be a sign of the Panama disease, so press 5 for more info. If you find well-defined long gray spots with a black ring and an outer yellow halo, it can be a sign of sigatoke leaf spots, so press 6 for more info. If you have any questions, you can contact your local agricultural extension officer.

For banana weevil, press 1. (See section 1.7.3)
For banana silvering thrips, press 2. (See section 1.7.4)
For nematodes, press 3. (See section 1.7.5)
For cigar-end-rot disease, press 4. (See section 1.7.6)
For panama disease, press 5. (See section 1.7.7)
For sigatoke leaf spots, press 6. (See section 1.7.8)

1.7.3 Banana Weevil

This is a brown-black weevil with a curved hard shell. They are nocturnal and hide during the day in or between leaf sheaths above the ground level. They form irregular tunnels in the rhizome at ground level reducing it to a mass of rotten tissue. The leaves turn yellow, wither and die prematurely. Infested plants are easily blown over by wind. The heart leaf may also die. For more information, contact your local agricultural extension officer.
For control of the banana weevil, press 1. (See section 1.7.1.1)

1.7.4 Banana Silvering Thrips

The adult pest is dark brown, 1.5 millimeters long with 2 pairs of fringed wings. It causes silvery patches on fruits which later turn brown. The skin of heavily infested fruit may crack permitting secondary infection, which causes fruit to rot. For more information, contact your local agricultural extension officer.
For control of the banana silvering thrips, press 1. (See section 1.7.1.2)

1.7.5 Nematodes

Nematodes affect roots so that they rot away. The symptoms are yellowing of leaves, growth ceases and the rhizome development is retarded. Mature plants may topple, and brown lesions form on roots in advanced stages. For more information, contact your local agricultural extension officer.
For control of nematodes, press 1. (See section 1.7.1.3)

1.7.6 Cigar-end-rot disease

This is caused by a fungus that invades the dried flower parts and penetrates into the skin. High humidity, overcrowded plants and abundant leaf trash favor the disease incidence. The fruit tips undergo a dry rot with an ashy appearance that look like a cigar. For more information, contact your local agricultural extension officer.
For control of the cigar-end-rot disease, press 1. (See section 1.7.1.4)

1.7.7 Panama Disease

The Panama Disease, also called Fusarium Wilt, is caused by a soil-borne fungus found in banana debris in the soil. The older leaves turn yellow, and collapse while still green at the base. The emerging heart leaf may die while the stem remains standing till it decays and falls over. Diseased plants fail to produce normal fruit and die before the fruit stalk is fully developed. For more information, contact your local agricultural extension officer.
For control of the panama disease, press 1. (See section 1.7.1.5)

1.7.8 Sigatoke Leaf spots

Sigatoke leaf spots are caused by various fungi. This disease forms well-defined long gray spots with a black ring and an outer yellow halo. For more information, contact your local agricultural extension officer.
For control of sigatoke leaf spots, press 1. (See section 1.7.1.6)

1.8 Market information

Commodity prices for the bigger markets are given after the evening news on Citizen TV and radio, and on KBC radio. For market information, you can also contact your local agricultural extension officer. KACE, Kenya Agricultural Commodities Exchange, also provides information about markets and commodity prices. They have an automatic information hotline that you can call on 0900 552 055. Unfortunately, the commodities that are currently covered by this service are only maize, beans, potatoes, cabbages and tomatoes. KACE also operates 11 Market Information Points, MIPs, located at rural market centres in Kenya. An MIP serves as a source of market information and intelligence, and also as a trading floor to link buyers and sellers of commodities. Information is displayed on bulletin and writing boards at an MIP. There are currently 11 MIPs in the following areas and regions: Embu and Machakos in Eastern Province, Murang'a and Karatina in Central Province, Eldoret and Kitale in Rift Valley Province, Chwele, Kamukuywa, Mayanja and Myanga in Western Province, Kisii in Nyanza Province. Eventually, KACE plans to establish MIPs in all agricultural regions of Kenya, to link commodity surplus and deficit areas.
For Eastern Province, press 1.

For Central Province, press 2.

For Rift Valley Province, press 3.
For Western Province, press 4.
For Nyanza Province, press 5.

1.9 Frequently Asked Questions

The crop grows well in fairly hot and humid areas that are within an altitude of 0 to 1,800 meters above sea level with rainfall of at least 1,000 millimeters per year. Plough the land during the dry season and remove or kill all weeds, especially couch grass. Planting should be done at the beginning of the long rains. Holes should be dug 60 centimeters deep and be 60 centimeters wide. Fill the holes with topsoil mixed with 2 debes of well-rotten farmyard manure and 226 grams of double superphosphate, DSP, per hole. Plant the suckers some 30 centimeters deep so that it is firm in the soil. The short varieties should be planted with a spacing of 3 times 3 meters, the medium varieties with 3 times 4 meters and the tall varieties are spaced at 4 times 4 meters. Bananas are sensitive to strong wind. Planting in wind-sheltered positions and in blocks rather than strips is recommended. Fertilizer application is recommended for banana growth. Apply 100 to 125 grams of Calcium Ammonium Nitrate, C.A.N., per stool per year at the beginning of the rains. This is scattered top-dressed in a circle 2 feet away from the banana stool. Thin the orchard in order to produce large bunches and to increase yields. Generally the banana stools should be thinned to leave one bearing plant, one half grown, and one just starting to grow. Harvesting starts 13 to 18 months after planting. Cut the stem when the fruit is fully developed and it is light green and shiny in appearance. If you have any questions, you can contact your local agricultural extension officer.

Huduma ya habari kuhusu ndizi

Anza

Karibu. Kwa mazungumza kwa Kiingereza, bonyeza 1. Kwa mazungumzo kwa Kiswahili, bonyeza 2.

1. Yaliyomo

Karibu kwa huduma ya habari kuhusu ndizi.Wakati wowote kwenye mazungumzo haya, unaweza kupata habari kuhusu maswali ambayo huulizwa na watu wengi kwa kubonyeza 9 (tisa) au unaweza kuchagua mojawapo ya sehemu zifuatazo ili kupata habari iliyomo.

Kwa mazingira mwafaka na udongo bora wa kupanda ndizi, bonyeza 1. (Tazama sehemu 1.1)
Kwa habari kuhusu aina mbalimbali za mbegu za ndizi,bonyeza 2. (Tazama sehemu 1.2)
Kwa mbegu za kupanda, bonyeza 3 (Tazama sehemu 1.3)
Kwa matayarisho ya shamba na upanzi, bonyeza 4 (Tazama sehemu 1.4)
Kwa utunzaji wa mimea yako, bonyeza 5 (Tazama sehemu 1.5)
Kwa kuvuna, bonyeza 6 (Tazama sehemu 1.6)
Kwa wadudu na wanyama waharibifu, bonyeza 7 (Tazama sehemu 1.7)
Kwa habari kuhusu soko, bonyeza 8 (Tazama sehemul.8)
Kwa habari inayouliziwa sana na watu ,bonyeza 9 (Tazama sehemu 1.9)
Wakati wowote kwenye mazungumzo, unaweza kuyakatiza mazungumzo kwa kubonyeza kibonyezo cha nyota ili uweze kurudi katika yaliyomo. Unaweza pia kuchagua kuirudia taarifa ambayo unapatiwa punde tu kwa kubonyeza kibonyezo cha 0. Unapotaka kukatiza mazungumzao ya awali, unaweza kukatiza mawasiliano.

1.1 Mazingira mwafaka na udongo bora

Ndizi hukua vyema katika sehemu zenye joto na unyevu wa kadri ambazo zinapatikana katika nyanda za kati ya mita 0 hadi 1,800. Mvua inafaa kuwa takriban kadri ya milimita 1,000 kwa mwaka.Hii ni muhimu hasa wakati wa kuchanua kwa maua ya ndizi.Ndizi zinaweza kupandwa karibu na kingo za mito. Hata hivyo, unafaa kuepuka kupanda ndizi karibu na maeneo wazi yanayoathiriwa mno na upepo mkali. Inapendekezwa kwamba ndizi zinafaa kupandwa katika sehemu zisisoathiriwa mno na upepo mkali. Udongo mzuri wenye rutuba pia ni muhimu katika ukuzaji wa ndizi.Mahitaji ya pH kati ya 6 na 7.5. Madini ya *Lime* inaweza kuongezwa kwa udongo ambao una asidi ili kuipunguza.Maeneo yanayokuza ndizi kwa wingi nchini Kenya ni pamoja na Mkoa wa Pwani, Mkoa wa Magharibi, Mikoa ya Nyanza, Mashariki na Kati.
Kwa habari kuhusu aina mbalimbali za ndizi, bonyeza 1 (Tazama sehemu 1.1.1)

1.2 Aina za ndizi

Kuna aina kadha wa kadha za ndizi zinazokuwza baadhi zikiwa za kupikwa na zingine zikiwa za kukuzwa katika sehemu kame. Baadhi ya aina za mbegu zinazotoa mazao mengi na Uganda green (matoke), Kisii matoke, Ndizi Ng'ombe na Kikuyu 1. Aina ya mbegu zinazokuzwa katika sehemu kame zinatakikana kuwa na rangi ya manjano zinapoiva, kwa sababu hizi ndizo zinazopendwa sana sokoni. Aina za ndizi katika kundi la Cavendish ni Giant Cavendish, Williams Hybrid, Grand Nain Dwarf Cavendish.Aina nyingine ni Sweet Banana, Lacatan, Valery na Paz. Gros Michel ambalo hujulikana kama Kampala ni aina ya ndizi maarufu inayoliwa baada ya kuiva, lakini inaathiriwa mno kwa urahisi na maradhi ya Fusarium Wilt ambayo pia hufahamika kama Panama Disease.

1.3 Mbegu za kupanda

Kuna aina mbili za mbegu za kupanda Mojawapo ya mbegu hizi ni migomba chipukizi na na migomba yenye matawi membamba. Miche hii huchumwa kutoka kwenye mimea iliyokomaa ikiwa na urefu wa mita 1 na kipenyo cha sentimita 15 shinani.Mbegu nyingine badalia ya kupanda ni Bull-Heads kutoka kwa mgomba uliokomaa.Bull-heads hutumika tu ikiwa migomba chipukizi haipatikani. Uzalishaji wa ndizi kwa matumizi ya teknolojia ya chembechembe za vimelea au Tissue Culture (TC) hufanywa maabarani na inaendelea kupata umaarufu. Mbinu hii ya uzalishaji ina faida nyingi kuliko mbinu nyingine za uzalishaji.Ukuzaji viini huchochea uzalishaji haraka wa mbegu bora za kupanda.
Miche iliyokuzwa kwa njia hii inafaa kuwa na urefu wa milimita 200 hadi 300 na iwe na aghalabu majani matano yenye afya na vifundo vipana wakati wa kuatika miche. Unaweza kununua miche iliyokuzwa kiteknolojia katika Chuo Kikuu cha Kilimo na Teknolojia cha Jomo Kenyatta au Taasisi ya Utafiti wa Kilimo ya Kenya (KARI) huko Thika kwa bei ya shilingi 75 hadi 100 kwa kila mche chipukizi. Unaweza kupata migomba chipukizi kutoka kwa majirani au ununue kutoka kwa kwa wakulima wengine kwa bei ya kati ya

shilingi 20 hadi 40. Migomba chipukizi inaweza kuwa ina magonjwa ya shambani.

1.4 Utayarishaji shamba na Upanzi

Tayarisha shamba wakati wa kiangazi kwa kulima na kuondoa magugu yote hususan ukoka.Upanzi unafaa mvua nyingi ianzapo kunyesha. Mashimo yanafaa kuchimbwa kina cha sentimita 60 yenye upana wa sentimita 60. Jaza mashimo hayo kwa udongo wa juu uliochanganywa na debe 2 za samadi ya ng'ombe iliyooza vizuri na garamu 226 za mbolea aina ya Double Superphospahte (DSP) katika kila shimo. Panda miche kwa kina cha sentimita 30 ili iwe imara. Nafasi kati ya mche mmoja na mwingine hutegemea aina ya ndizi. Ndizi fupi zinafaa kupandwa katika umbali wa mita 3 kwa 3 ilihali zile zenye urefu wa kadri zikifaa kutenganishwa kwa umbali wa mita 3 kwa 4. Aina ya ndizi ndefu kwa umbali wa mita 4 kwa 4. Unaweza kupanda ndizi pamoja na mimea mingine ya vyakula inayokua haraka kando na kuwa na mizizi isiyopenya mno kwenye ardhi katika mwaka wa kwanza. Utiaji mbolea unapendekezwa kutekelezwa kuimarisha ukuaji wa ndizi. Changanya mchanga katika kila shimo la kupanda na gramu 100 hadi 150 za mbegu ya DSP wakati wa kupanda.
Kwa habari kuhusu mbegu za kupanda, bonyeza 1 (Tazama sehemu 1.21)

1.5 Utunzaji wa mimea yako

Ng'oa amagugu yoyote kwa kupalilia hasa wakati mimea ni michanga. Mimea ikikwisha kukomaa, itafunika ardhi na kushinda magugu. Wakati wa kupalilia, uangalifu unafaa kudumishwa ili mizizi katika udongo wa juu isikatwe. Katika mwaka wa kwanza, ndizi zinaweza kupandwa katika kipande kimoja cha ardhi pamoja na mimea kama maharagwe na njugu ambayo ina mizizi yenye kina kufupi ardhini. Baada ya hapo, unaweza kutandaza majani ya migomba inayopogolewa shambani. Kadhalika, matumizi ya mbolea yamependekezwa katika kukuza ndizi. Tumia gramu 100 hadi 125 za Calcium Ammonium Nitrate (CAN) katika kila shimo kila mwaka mvua inapoanza kunyesha. Weka mbolea hii juu ya udongo katika mduara futi 2 kutoka shimo la mmea. Punguza miche shambani ili ndizi ziweze kuzaa mikungu mikubwa na kuongeza mazao. Kwa jumla mimea yote kwenye shimo moja inafaa ipunguzwe ili ubaki mmea mmoja ambao unaanza kukua. Mashina yenye mikungu yanafaa kuhimiliwa kwa nguzo ili yasianguke kabla ya kukomaa. Upunguzaji wa ndizi zilizokuzwa kwa njia ya kiteknolojia ni shughuli ngumu na nyeti hivi kwamba, tahadhari kubwa inafaa kuchukuliwa.
Kwa habari zaidi kuhusu upunguzaji wa mimea ya ndizi iliyokuzwa kiteknolojia, bonyeza 1. (Tazama sehemu 1.5.1)
Kwa habari zaidi kuhusu manufaa ya mtandazo shambani, bonyeza 2.(Tazama sehemu 1.5.2)

1.5.1 Upunguzaji wa ndizi zilizokuzwa kiteknolojia

Ndizi zilizokuzwa kiteknolojia huotesha vimelea vingi mwezi mmoja hadi miwili baada ya kupandwa. Vimelea au chipukizi hizi zinafaa kukatiwa mashinani ili kuruhusu ukuaji wa mmea mkuu. Mmea mkuu unapotimu urefu wa mita moja, unafaa kuchagua chipukizi moja yenye nguvu unaoelekea upande wa mashariki au juu ya mteremko ambako ardhi ina urowevu. Chipukizi nyingine zote zinafaa kukatiwa mashinani na kuuliwa kwa mashina hayo kwa matumizi ya milimita 2 za mafuta ya taa au dizeli. Mmea mkuu unapoanza kutoa maua, chagua mche mmoja unaoelekea upande wa chipukizi ya mwanzo.Iwapo uchaguzi wa mche unafanywa vizuri, kunafaa kuwepo mmea mkuu unaozaa matunda, mche mkubwa na mdogo, yote ikielekea upande mmoja.

1.5.2 Kuweka Matando

Kuweka matandazo kuna manufaa kwa sababu huhifadhi unyevu, hukinga mmomonyoko wa udongo na huongeza rutuba kwenye udongo pindi yanapooza.Kadhalika, matando huzuia ukuaji wa magugu.Weka matando shambani kwa kutumia mabaki ya migomba inayokatwa wakati wa kupunguza mimea.

1.6 Kuvuna

Uvunaji huanza kati ya miezi 13 hadi 18 baada ya kupanda. Kata shina la mgomba wakati matunda yenyewe yamekuwa kamili na yenye rangi nyepesi ya kijani kibichi na sura ya kunga'aa. Wakati wa kusafirisha, funika mikungu ya ndizi kwa nyasi au majani ya ndizi ili kuepuka kuvalia.Mazao ya tani 17 hadi 20 kwa kila hekta moja kwa mwaka yanaweza kupatikana.Hata hivyo, unaweza kupata mazao ya tani 40 hadi 50 kutoka kwenye mashamba yaliyotunzwa vizuri. Baada ya takariban miaka 8 hadi 10 ya uzalishaji mazao ya ndizi yanaweza kupungua. Wataalamu hupendekeza kuwa mkulima abadili mimea ya ndizi.Ikiwa shamba bado litatumiwa tena, ufukizaji wa wa udongo hupendekezwa kabla ya upandaji tena wa ndizi.

1.7 Vitilifu na Magonjwa

Kwa habari kuhusu uzuiaji vitilifu na magonjwa tofautitofauti,bonyeza 1 (Tazama sehemu 1.7.1)
Kwa maelezo kuhusu dalili za vfukusi vya ndizi, bonyeza 2. (Tazama sehemu 1.7.2)
Kwhabari kuhusu Vifukusi vya Ndizi, bonyeza 3. *(Tazama sehemu 1.7.3)*
Kwa habari kuhusu banana Silvery Thrips, bonyeza 4 (Tazama sehemu 1.7.4)
Kwa habari kuhusu Nematodes, bonyeza 5. (Tazama 1.7.5)
Kwa habari kuhusu ugonjwa wa Ciger-end rot, bonyeza 6. (Tazama sehemu 1.7.6)
Kwa habari kuhusu Panama Disease, bonyeza 7.(Tazama sehemu 1.7.7)
Kwa habari kuhusu Sigatoke leaf Spots, bonyeza 8. (Tazama sehemu 1.7.8)

1.7.1 Uzuiaji wa vitilifu na magonjwa

Chagua kitilifu au ugonjwa unaotaka habari juu yake.
Kwa vifukusi vya ndizi bonyeza 1 (Tazama sehemu 1.7.1.1)

Kwa 'banana silvering thrips' bonyeza 2 (Tazama sehemu 1.7.1.2)
Kwa nematoda, bonyeza 3 (Tazama 1.7.1.3)
Kwa ugonjwa wa 'Cigar-end-rot', bonyeza 4 (Tazama sehemu 1.7.1.4)
Kwa ugonjwa wa panama, bonyeza 5 (Tazama sehemu 1.7.1.5)
Kwa 'sigatoke leaf spots,' bonyeza 6, (Tazama sehemu 1.7.1.6)

1.7.1.1 Uzuiaji was Vifukusi vya ndizi

Panda miche chipukizi ilio safi na yeny afya. Miche chipukizi inayotumiwa kuwa imeambukizwa inafaa kutiwa ndani ya mchanganyiko wa milimita 100 wa asilimia 18 ya Dieldrin ndani ya lita 20 za maje. Katia chini ya ardhi mashina yote yaliyochakaa na upake rizomu zilizokatwa na asilimia 5 ya Dieldrin na kufunika kwa udongo. Nyunyiza asilimia 5 ya Dieldrin au dawa ya unga ya Aldrin maeneo yote ya chini ya shina.

Kwa habari na dalili za vifukusi vya ndizi, bonyeza 1 (Tazama sehemu 1.7.3)

1.7.1.2 Uzuiaji wa Banana Silvering Thrips

Nyunyizia dawa ya Carbaryl, Dichlorous Fenitrothion au Diazinon.

Kwa habari na dalili za banana 'silvering thrips', bonyeza 1 (Tazama Sehemu 1.7.4)

1.7.1.3 Uzuiaji wa Nematoda

Wadudu wa nematoda wanaweza kuzuiwa kwa kupanda miche chipukizi isiyo na vitilifu kwa kupuna mizizi yote kabla ya kupanda , mbadilisho wa mimea na kwa kufukiza udongo kwa dawa ya Furadan au Nemacur.

Kwa habari na dalili za nematoda, bonyeza 1 (Tazama sehemu 1.7.5)

1.7.1.4 Uzuiaji wa ugonjwa wa Cigar-end-rot

Zuia kwa kuondoa sehemu zote za maua zilizokauka kwa mkono siku 8 hadi 11 baada ya mkungu wa ndizi kujitokeza.

Kwa habari na dalili za ugonjwa wa 'cigar-end-rot', bonyeza 1 (Tazama sehemu 1.7.6)

1.7.1.5 Uzuiaji wa ugonjwa wa Panama

Zuia kwa kupanda aina ya mimea inayostahilimi kam Cavendish, Kisigane, na Uganda green (Matoke)

Kwa habari na dalili za ugonjwa wa Panama, bonyeza 1. (Tazama 1.7.7)

1.7.1.6 Uzuiaji wa Sigatoke leaf spots

Ondoa majani yenye madoadoa na takataka za majano na kuyaharibu. Epusha msongamano wa mimea. Kikemikali, dawa ya 'Mancozeb' inaweza kutumika kuuzuia ugonjwa huo.

Kwa habari na dalili za ugonjwa wa 'sigatoke leaf spots' bonyeza 1. (Tazama 1.7.8)

1.7.2 Maelezo ya dalili

Ikiwa majani ya ndizi yatageuka kuwa ya rangi ya manjano, kunyauka na kukauka kabla ya wakati wake basi inaweza kuwa dalili ya vifukusi vya ndizi kwa hivyo bonyeza 1 kwa habari zaidi. Ikiwa utapata madoa ya kifedha kwenye ndizi ambayo baadaye hugeuka kuwa ya rangi ya kikahawia, inaweza kuwa dalili ya 'banana silvering thrips', kwa hivyo bonyeza 2 kwa habari zaidi. Ikiwa majani yatageuka kuwa ya rangi ya manjano, ukuaji kukoma na ukuaji wa rizomu kudumaza, inaweza kuwa dalili ya nematoda, kwa hivyo, bonyeza 3 kwa habari zaidi. Ikiwa ncha za ndizi zitakuwa na uozo mkavu na sura ya kijivu inayoonekana kama sigara, ni dalili ya 'cigar-end-rot' kwa hivyo, bonyeza 4 kwa habari zaidi. Ikiwa majani yaliyokomaa yatageuka kuwa manjano, na kuanguka yakiwa bado na rangi ya njano katika sehemu ya chini na migomba kutozaa matunda ya kawaida na kukauka kabla ya kikonyo cha ndizi kukua kikamilifu, inaweza kuwa dalili ya ugonjwa wa panama, kwa hivyo, bonyeza 5 kwa habari zaidi. Ikiwa utapata madoa marefu ya kijivu pamoja na duara nyeusi na duara ya nje ya manjano, inaweza kuwa dalili ya 'sigatoke leaf spots' kwa hivyo, bonyeza 6 kwa habari zaidi.

Kwa vifukusi vya ndizi, bonyeza 1 (Tazama sehemu 1.7.3)
Kwa 'banana silvering thrips', bonyeza 2. (Tazama sehemu 1.7.4)
Kwa nematoda, bonyeza 3 (Tazama sehemu 1.7.5)
Kwa ugonjwa wa 'Cigar-end-rot', bonyeza 4 (Tazama sehemu 1.7.6)
Kwa ugonjwa wa Panama, bonyeza 5. (Taxama sehemu 1.7.7)
Kwa sigatoke leaf spots, bonyeza 6. (Tazama sehemu 1.7.8)

1.7.3 Vifukusi vya Ndizi

Mdudu huyu ana rangi ya hudhurungi nyeusi na kaka gumu lililopinda. Huonekana usiku na hujificha wakati wa mchana ndani au kati ya majani juu ya udongo. Hutoboa mashimo kwenye rizomu na kuzifanya zioze. Majani hugeuka kuwa ya rangi ya manjano, hunyauka na kukauka kabla ya kupevuka. Mimea iliyoathirika huangushwa kwa urahisi na upepo. Pia jani la kati huweza kukauka.

Kwa uzuaji wa vifukusi vya ndizi, bonyeza 1 (Tazama sehemu 1.7.1.1.)

1.7.4 Banana Silvering Thrips

Mdudu mkubwa huwa na rangi ya kahawia iliyoiva, urefu wa milimita 1.5 urefu na jozi mbiliza mbawa. Husababisha madoa ya kifedha kwenye matunda ambayo baadaye huguka kuwa ya hudhurungi. Ngozi ya ndizi lililoathiriwa mno huweza kupasuka na kuruhusu maambukizo zaidi, ambayo husababisha ndizi kuoza.

Kwa uzuiaji wa 'banana silvering thrips' bonyeza 1. (Tazama sehemu 1.7.1.2)

1.7.5 Nematoda

Nematoda huathiri mizizi ambayo huoza. Dalili zake ni majani kugeuka rangi ya manjano ukuaji kunakokwama na kuviza rizomu. Mimea iliyokomaa inaweza kuanguka na mizizi kuwa na vidonda vidonda vyenye maji ya hudhurungi baadaye.

Kwa uzuiaji wa nematoda, bonyeza 1 (Tazama sehemu 1.7.1.3)

1.7.6 Ugonjwa wa Cigar-end-rot

Ugonjwa huu husababishwa na ukungu ambao huvamia sehemu za maua zilizokauka na kupenya ndani ya ngozi. Unyevu mwingi, mimea iliyosongamana na wingi wa taka za majani huchangia katika kuenea kwa ugonjwa huu. Nche za matunda hufanya madoa makavu yaliyooza yenye sura za kijivu yanaoonekana kama sigara.

Kwa uzuiajiwa ugonjwa wa 'cigar-end-rot', bonyeza 1. (Tazama sehemu 1.7.1.4)

1.7.7 Ugonjwa wa Panama

Ugonjwa wa Panama ambao pia huitwa Fusarium Wilt, husababishwa na ukungu unaopatikana udongoni katika mabaki ya ndizi yaliyo udongoni. Majani yaliyokomaa hugeuka na kuwa ya manjano na kuanguka yangali bado na rangi ya kijano sehemu za chini. Jani la kati linalochipuka linaweza kukauka ilhali mgomba unabaki wima mpaka pale unapooza na kuanguka. Mimea iliyoathirika hukosa kuzaa matunda na kukauka kabla ya kikonyo cha ndizi kukua kikamilifu.

Kwa uzuiaji wa ugonjwa wa Panama, bonyeza 1 (Tazama sehemu 1.7.1.5)

1.7.9 Ugonjwa wa 'sigatoke leaf spots'

Sigatoke leaf spots husabishwa na kungu za aina mbalimbali. Ugonjwa huu husababisha madoa marefu ya kijivu yenye duara nyeusi na duara ya nje ya rangi ya manjano.

Kwa uzuiaji wa 'sigatoke leaf spots' bonyeza 1. (Tazama sehemu 1.7.1.6)

1.8 Habari kuhusu soko

Shirika la Kubadilishana bidhaa za Kilimo la Kenya, KACE hutoa habari kuhusu masoko na bei za bidhaa. Shirika hili lina simu ya moja kwa moja unayoweza kupiga. Simu hiyo ni nambari: 0900 552 055. Kwa bahati mbaya, bidhaa zinazoshughulikiwa kwa sasa kwa huduma hii ni mahindi, maharagwe, viazi, kabeji na nyanya. Shirika la KACE pia huendesha vituo 11 vya habari kuhusu soko, vilivyowekwa katika vituo vya masoko ya vijijini nchini Kenya. Kituo cha habari ya soko huhudumu kama chanzo cha habari kuhusu soko, na pia kama msingi wa kibiashara wa kuunganisha wanunuzi na wauzaji wa bidhaa. Habari huonyeshwa (taarifa rasmi ya habari) na katika vibao vya kuandikia katika kituo cha habari kuhusu soko. Kwa sasa, kuna vituo 11 vya habari ya soko katika mahali na maeneo yafuatayo: Embu na Machakos katika Mkoa wa Mashariki, Murang'a na Karatina katika Mkoa wa Kati, Eldoret na Kitale katika Mkoa wa Bonde la Ufa, Chwele, Kamukuywa, Mayanja na Myanga katika Mkoa wa Magharibi na Kisii katika Mkoa wa Nyanza. Mwishowe shirika la KACE linapanga kunzisha vituo vya habari kuhusu soko katika maeneo yote ya kilimo nchini Kenya ili kuunganisha maeneo yenye uziada na uhaba wa bidhaa.

Kwa Mkoa wa Mashariki, bonyeza 1

Kwa Mkoa wa Kati, bonyeza 2

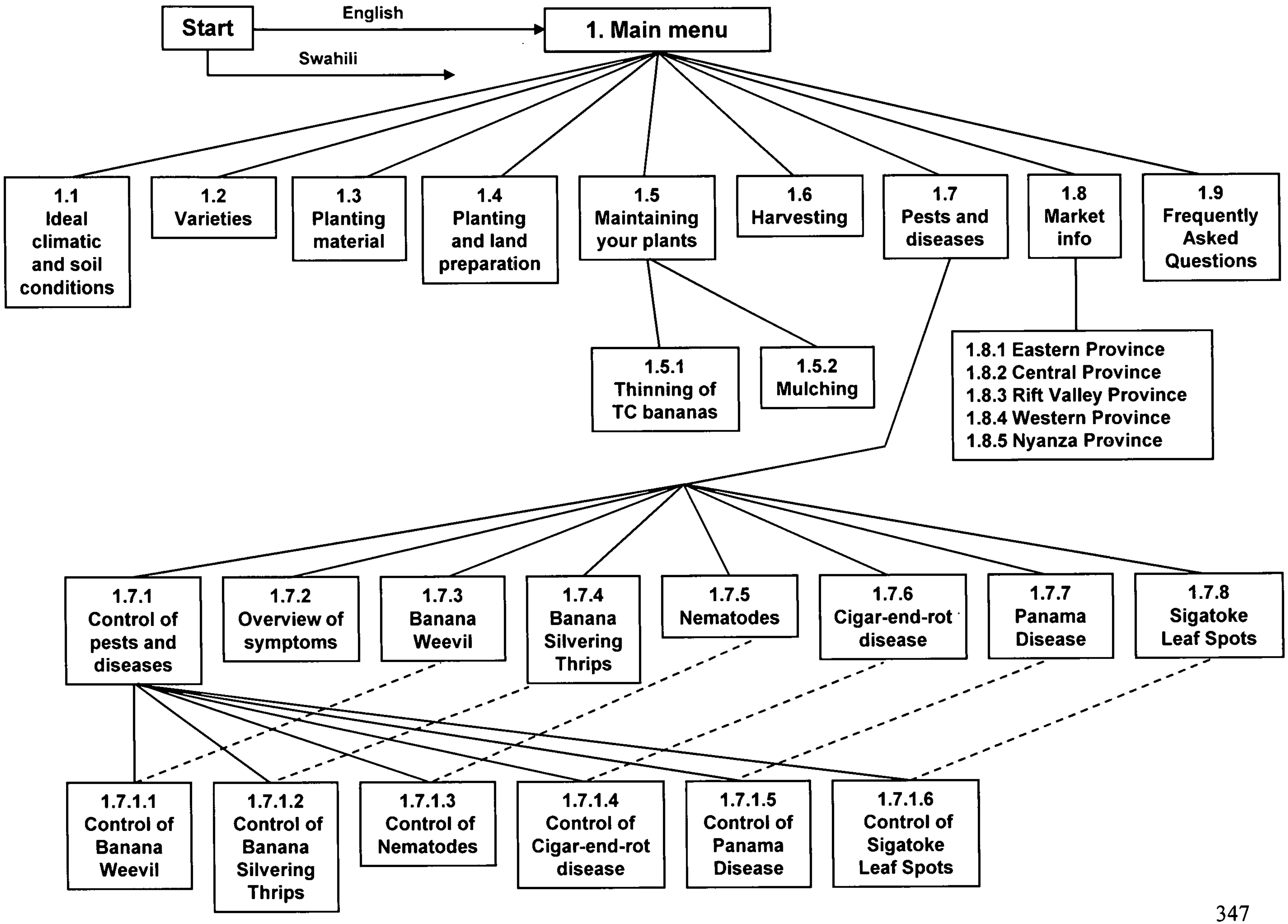
Kwa Mkoa wa Bonde la Ufa, bonyeza 3

Kwa Mkoa wa Magharibi, bonyeza 4

Kwa Mkoa wa Nyanza, bonyeza 5

1.9 Maswali yaulizwayo mara kwa mara

Ndizi hunawiri vizuri katika maeneo yenye joto na unyevu unyevu kadiri yaliyo katika mwinuko wa mita 0 hadi 1,800 kutoka usawa wa bahari pamoja na angalau mvua ya milimita 1,000 kwa mwaka. Tayarisha shamba wakati wa kiangazi na kuondoa au kuua magugu yote, hususan ukoka. Upanzi unafaa kufanywa wakati wa mvua ya masika kuanza. Uvunaji huanza baada ya kipindi cha miezi 13 hadi 18 baada ya kupanda. Mashimo yanafaa kuchimbwa kina cha sentimita 60 na sentimita 60 upana. Jaza mashimo na udongo wa juu juu uliochanganywa na madebe 2 ya samadi iliyooza vizuri na gramu 226 za mbolea ya Double Superphosphate, DSP kwa kila shimo. Panda mimea chipukizi kwa kina cha sentimita 30 ili iwe imara udongoni. Aina fupi ya ndizi inafaa kupandwa kwa kutenganisha kwa mita 3 kwa 3, aina ya ndizi yenye urefu wa kadiri zitenganishwe kwa mita 3 kwa 4 na aina ndefu ya ndizi zipandwe kwa kuacha nafasi ya mita 4 kwa 4. Ndizi huathiriwa na upepo mkali. Upanzi wa ndizi katika maeneo yaliyokingwa dhidi ya upepo na katika mapande makubwa kuliko mistari unapendekezwa.


Start
English
Swahili
1. Main menu
1.1 Ideal climatic and soil conditions
1.2 Varieties
1.3 Planting material
1.4 Planting and land preparation
1.5 Maintaining your plants
1.6 Harvesting
1.7 Pests and diseases
1.8 Market info
1.9 Frequently Asked Questions
1.5.1 Thinning of TC bananas
1.5.2 Mulching
1.8.1 Eastern Province
1.8.2 Central Province
1.8.3 Rift Valley Province
1.8.4 Western Province
1.8.5 Nyanza Province
1.7.1 Control of pests and diseases
1.7.2 Overview of symptoms
1.7.3 Banana Weevil
1.7.4 Banana Silvering Thrips
1.7.5 Nematodes
1.7.6 Cigar-end-rot disease
1.7.7 Panama Disease
1.7.8 Sigatoke Leaf Spots
1.7.1.1 Control of Banana Weevil
1.7.1.2 Control of Banana Silvering Thrips
1.7.1.3 Control of Nematodes
1.7.1.4 Control of Cigar-end-rot disease
1.7.1.5 Control of Panama Disease
1.7.1.6 Control of Sigatoke Leaf Spots

CONSENT FORM TO TAKE PART IN TESTING A BANANA INFORMATION SYSTEM

1. I confirm that I have been fully informed about the testing that is going to take place and that I have had the opportunity to ask questions.
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.
3. I understand that it is the new computer program that is being tested not my skills and that all of the results will remain completely anonymous at all times.

YOUR NAME...

YOUR SIGNATURE..**DATE**........................

NAME OF PERSON TAKING CONSENT...

SIGNATURE OF PERSON TAKING CONSENT...

PLEASE ANSWER THE FOLLOWING QUESTIONS BY CIRCLING THE CORRECT BOX

1. How old are you?

18-30	31-40	41+

2. Are you male or female?

Male	Female

3 What language do you usually communicate in?

English	Swahili	Other

4. What other languages can you communicate in fluently?

English	Swahili	Other

5. How often do you make a phone call?

Daily	Every week	More rarely

6. Have you used a computer before?

Often	A little	Never

YOUR NAME..

TASK SHEET

Following the instructions that have been given to you on how to use the system, connect to the banana information service and use the options available to try to answer the following questions. One box contains the correct answer, the other two boxes contain incorrect answers – please put a circle around the one that you think is correct.

1. What is the minimum rainfall before irrigation becomes necessary?

a) 1,000 mm	**b)** 900mm	**c)** 1,100 mm

2. Which banana variety is highly susceptible to the Panama disease (Fusarium Wilt)?

a) Uganda green (Matoke)	**b)** Kisigame	**c)** Gros Michel (Kampala)

3. Which varieties should be planted at a spacing of 3 times 3 meters?

a) Tall	**b)** Medium	**c)** Short

4. When should planting be done?

a) End of the long rains	**b)** Middle of the long rains	**c)** Beginning of the long rains

5. Which pest or disease shows the symptom that the older leaves turn yellow, and collapse while still green at the base?

a) Cigar-end-rot disease	**b)** Nematodes	**c)** Panama disease

Thank you, now please turn over to complete the questionnaire which asks for your impressions of the system.

QUESTIONNAIRE

YOUR NAME..

Consider the following statements and put a circle around the box depending on whether you agree, are not sure, or disagree.

1. I found the system easy to use.

a) I agree	b) Not sure	c) I disagree

2. What language did you listen to the speech in?

a) English	b) Swahili

3 I could tell easily which number I needed to press from the menu to get the information I wanted.

a) I agree	b) Not sure	c)I disagree

4 I would use this system instead of using other resources such as the Internet or books.

a) I agree	b) Not sure	c) I disagree

5. The speech from the system was very clear to understand.

a) I agree	b) Not sure	c) I disagree

6. Would you prefer to hear a male or female voice?

a) Male	b) Female

7. If you have an issue about growing bananas where do you normally go to for information?

a) Other farmers	b) Agricultural Extension Officers	c) Other

TALKING BOOK – ALL ABOUT LONDON

YOUR NAME:

DATE:

The attached questionnaire has 18 statements. Please answer every one of them. Against each statement there are three boxes.

You should mark the first box if you generally AGREE with the statement.

Mark the central box if you are UNDECIDED or can't make up your mind.

Mark the right box if you generally DISAGREE with the statement.

In marking the left or right box you are not necessarily indicating STRONG agreement or disagreement, but just your general feeling most of the time.

TALKING BOOK – ALL ABOUT LONDON

QUESTIONNAIRE

No	Statement	Agree	Un-decided	Disagree
1	I would recommend this software to my friends			
2	The software has at some time stopped unexpectedly			
3	Learning to operate this software to start with is full of problems			
4	I sometimes don't know what to do next with this software			
5	I enjoy my sessions with this software			
6	It takes too long to learn the software commands			
7	Working with this software is satisfying			
8	This software seems to disrupt the way I normally like to arrange my work			
9	I feel in command of this software when I am using it			
10	I would not like to use this software every day			
11	There is too much to read before you can use the software			
12	Using this software is frustrating			
13	The speed of this software is fast enough			
14	There have been times in using this software where I have felt quite tense			
15	It is easy to make the software do exactly what you want			
16	This software is really very awkward			
17	I have to look for assistance most times when I use this software			
18	I will never learn to use all that is offered in this software			

ID number of participant:

PARTICIPANT QUESTIONNAIRE FOR SPEECH SYNTHESIS RESEARCH
ALL RESPONSES WILL BE KEPT STRICTLY CONFIDENTIAL AND
ANONYMOUS AND IN FULL COMPLIANCE WITH THE DATA PROTECTION ACT

Project Title: Text to Speech Technology: usability in multi-cultural environments as a teaching and learning tool for end users with special educational needs
Name of Researcher: Yvonne Spittles MSc MSBT MIVA
Institution: Thames Valley University
Contact Details: Yvonne.Spittles@tvu.ac.uk

- The information I supply will be preserved at Thames Valley University and will be kept confidential unless I give permission for my name to be used.
- The material will be preserved as a permanent research resource and may be used for research and publications now and in the future.
- My contribution will be kept safely and securely with access only to those with permission from the above researcher.
- I understand that I can withdraw my consent at any time by contacting the researcher.
- I have been given information about the research project and the way in which my contribution will be used.
- I give my permission for the information I am about to give/have given to be used for research purposes only (including research publications and reports) with strict preservation of anonymity.

Please answer the following questions:

1) **How old are you?**

2) **How long have you lived in England for?**

3) **Were you born in England?**

4) **If you were not born in England, what country were you born in?**

5) **What is your nationality?**

6) **What do you consider to be your cultural origin?**

7) **Are you male or female?**

8) **What is your first language?**

9) **What is the first language of your parents/guardians?**

Ethical Considerations for Usability Testing in Kenya using Kiswahili and English

Yvonne Spittles,
Institute for IT,
Thames Valley University, UK

Lynne Dunckley,
Institute for IT,
Thames Valley University, UK

ABSTRACT

This paper describes a user-based evaluation of an Information and Communication Technology (ICT) system that is being developed as part of an initiative by the Kenyan Ministry of Agriculture in collaboration with the University of Nairobi and the Local Language Speech Technology Institute based at the University of Bristol. The system would provide key information for farmers about growing bananas to enable farmers to become aware of better farming procedures, including planting requirements and identification and prevention of disease. The system uses telephone text as an input and speech as an output, both in Kiswahili and English. The participants used either English or Kiswahili as their language and had varying levels of literacy; the majority had never used ICT before.

Author Keywords

User testing, rural users, literacy, cross-cultural evaluation.

ACM Classification Keywords

H5.m. Information interfaces and presentation (HCI): H1.2 Human Factors

INTRODUCTION

Kenya consists of residents from many different cultures. 8.08% of the land is used for arable farming, 0.98% for permanent crops and 90.94% for other use; 670sq km of the land is irrigated (Central Intelligence Agency, 2006) – it is amongst these conditions that farmers need support and direction to plant, tend crops, treat disease, harvest and export bananas successfully. Access to agricultural information is difficult as the farmers are often in remote rural locations with no access to books or the Internet and rely on support from visits by national extension agricultural workers. The population of Kenya is around 34,000,000 and is made up of 42.5% 0-14 years, 55.2% 15-64 years and 2.3% 65 years+; the median age is 18.19 years. By the end of 2001, Kenya was host to 220,000 refugees from neighbouring countries particularly Somalia and Sudan (Central Intelligence Agency, 2006). The population of Kenya is sub-divided into the following ethnic groups – Kikuyu 22%, Luhya 14%, Luo 13%, Kalenjiin 12%, Kamba 11%, Kisii 6%, Meru 6% and other African 15%, non-African 1%. In this case, the users in the usability tests belonged to the Kikuyu ethnic group.

English and Kiswahili are the official languages of Kenya. Both are taught in school and are equally popular, however rural Kenyans prefer to use their traditional tribal language. There is not an area of Kenya where people cannot speak Kiswahili but there are many places outside of the towns where English is not spoken.

User Based Testing

Research suggests that only between 30% and 40% of ICT systems produced are ever successfully implemented and used for the purpose for which they were designed. While it is widely recognized that user-based testing can improve this, Gray et al (1998) consider that although practitioners in Human Computer Interaction (HCI) have been very interested in interface design, this interest has not been extended to the design of experiments and that

reliable and valid guidance for actual interface design depends on the results of the tests. It is vital that researchers have an understanding of how small features of an experimental design can cast large shadows over the results and conclusions that can be drawn from them.
In this case the interface of the ICT system being tested relies on end-users accessing information by pressing telephone buttons to select from a menu which gives speech output.

Ethical Background

Ethics can be regarded as the rules or standards governing the conduct of individuals or groups. One of the problems is the confused approach to computer ethics by society, governments and academics. The principal question for ethics is not, as the formalist would have it, "What rules should I obey?" but rather, "What intentions should I adopt?" The Economic and Social Data Service (2005) states that it is a fundamental ethical principle of research to ensure that participants give consent and that they enter into the research voluntarily, without coercion and in the knowledge that they can withdraw at any time, confident in the knowledge that there will be no adverse consequences of participating in the research, including the preservation of anonymity if they wish. Ethical issues regarding research methods have been primarily concerned with protecting subjects involved in research. Cassell (1980) states that over-attendance to hierarchically imposed guidelines may be irrelevant in the field research and may lead investigators to ignore ethical guidelines that seem inapplicable to their work, or to deny that there are even ethical problems associated with their research. He states that inappropriate regulations may do more harm than good and such a narrow conception of ethics not only fails to recognise the breadth of ethical decisions and judgments that are made, but it also serves to broaden the gap between researcher and researched. .Allen (1984) states that it is important to conduct tests, with human subjects, who cannot be subjected to any form of destructive testing, with deep respect for the users' emotions and well-being.

Although risks of becoming physically harmed through usability testing are negligible, undertaking testing can be stressful for participants who may feel pressure to perform and may worry about being slow at learning the systems, suffering inferiority complexes if they struggle in the knowledge that all this is under observation. Therefore, it is important that the evaluator makes the users feel as comfortable as possible during and after the tests, making it clear that it is the system that is being tested as opposed to the users (Nielsen, 1993). Case studies relate how western researchers undertook usability testing in Asia with users with specific learning difficulties who abandoned the testing in a state of distress because the experimenters had not considered ethical issues. A key ethical consideration is literacy as residents in the more rural areas are more likely to be illiterate and this must be one of the most sensitive areas to address in usability testing. The question "can you read or write" is pointless if the participant cannot read, likewise it is against ethical guidelines to subject the participant to any humiliation by directly asking them this question. Therefore, it was decided to approach this ethical dilemma by having a transcriber present at all tests.

Usability Evaluation in Practice

There were significant problems with the user testing with five attempts to conduct the usability tests on the system over a four-week period. Poor telecommunications meant connectivity was extremely unreliable. Although the Kenyan government is making attempts to modernise, one company has had the monopoly of supplying the service which is directed to business use. Trunks are primarily microwave radio relay with business data being commonly transferred by a very small aperture terminal (VSAT) system. Statistics for 2003 state that there are 328,400 main line telephones in use, with 1,590,900 mobile cellular telephones and 8,325 Internet hosts with 400,000 Internet users (Central Intelligence Agency, 2006). · Thus weekends or late evenings would be the only time that end users were able to connect reliably to the system. This was identified as a key weakness that would restrict the accessibility and acceptance of the system by the farming community who were already inexperienced in using ICT. We designed the tests which were undertaken in Kenya under the direction of the University of Nairobi. Following the pilot test using 3 participants in which connectivity issues were identified, full testing later took place with a further 10 participants at their own farms. At the start of each test, the users were asked verbally if they would like the tester to complete the forms for them. This eliminated any embarrassment that they could not read or write. The consent forms and task sheets were formatted using tick boxes and multiple choice questions so that the participants had to write as little as possible – because even if someone can read, it does not mean that they are confident to write also.

Another ethical consideration was for participants who could only communicate in Kiswahili as the tests were all written in English. It was therefore decided that a translator should be present at all tests but ideally this should be the same person as the transcriber to minimize the amount of persons present to reduce anxiety for the participant.

Findings

Following strict guidelines for ethical issues and overcoming technical difficulties, testing was successfully completed in Kenya on the target user group of banana farmers, all from the Kikuyo tribe without causing any distress or humiliation whatsoever resulting in valuable data being obtained which is being analysed and evaluated. 60% of the participants could be classed as "older users" aged 41+ who are traditionally less inclined to want to embrace ICT. 80% of the participants had never used a computer before and only 30% used a telephone on a daily basis; 10% of the participants were illiterate. From testing this system technical weaknesses were identified in the ICT available in Kenya; the technical developer is now researching further into possible approaches to overcome this.

REFERENCES

1. Allen, R,B. (1984). Working paper on ethical issues for research on the use of computer services and interfaces. *ACM SIGCHI Bulletin.* **16,** (1), pp 12-16.
2. Cassell et al. (1980).Ethical problems of fieldwork. *Social problems (special issue). 27, pp 259-378*
3. Central Intelligence Agency. (2006). Facts about Kenya. *[Online}. URL* www.cia.gov.uk/kenya. (Accessed 26 March 2006).
4. Economic and Social Data Service. (2005). *[Online]. URL www.esds.ac.uk.* (Accessed June 2005)
5. Gray et al. (1998). Damaged Merchandise? A review of experiments that compare usability evaluation methods. *Human Computer Interaction,* **13,** (3), pp 203-261.
6. Neilsen, J. (1993). Usability Engineering. Academic Press.

ETHICAL ISSUES IN PARTICIPATIVE RESEARCH IN USABILITY TESTING

Yvonne Spittles and Lynne Dunckley

Thames Valley University

20 December 2005

ABSTRACT

User-based testing is an important strategy as part of software development particularly for user interface design and is a key part of user participation in agile methodologies. However, there is little material on ethical issues in the HCI literature. In this paper the authors address the ethical issues involved in the user-based evaluation of interactive user interfaces as part of software development. The paper considers the ethical guidelines available in other domains and goes on to describe a case study which provides pointers for the development of ethical guidelines and checklists for usability evaluation.

INTRODUCTION

Ethics can be regarded as the rules or standards governing the conduct of a person or the conduct of the members of a profession. There are two philosophical approaches to ethics. The first takes it as self-evident that there exists a criterion of goodness or rightness must consist in some law or principle of behaviour. This approach could be simplified as "legalism" or "formalism"- the view that without strict rules of conduct, people will be governed by private preference or personal advantage. The main objection to an ethic of laws is that it disregards the consequences of the actions it prescribes. Provided only that an act conforms to the rules, it is good, regardless of the harm that may result.

The second approach, currently much in fashion, takes an humanist approach in that it tries to establish an ethic not based on the straightjacket of inflexible moral codes, but one that takes account of human needs and aspirations. Its watchwords are humanistic ones like fulfilment, tolerance, creativity, and freedom, and all of which are stifled when the highest good is conceived as obedience to law. The trouble with this approach is that all these goods are difficult to pin down and turn out to be subjective; that is, relative to the individual or to a particular profession or society. (Macmurray 1938) claims these two schools of ethics -"legalism" and "relativism"- turn out to be half-truths. He anchors his thought in the reality of human freedom, which he establishes by the strictest reasoning. Where he differs from both the legalist and relativist is in his relentless pursuit of the logical implications of freedom. Macmurray also distinguishes an intention from an ideal:
An ideal of life is inherently reflective and contemplative ,...an idea of how life might be lived or ought to be lived .... It is concerned with judgment, not with action... If we are to realize our ideals, we shall first have to form an intention

to act in a way that we believe will realize them. The ideal itself is not an intention.

If in reality human life is constituted by and goals, which may or may not reflect an ideal, then the principal question for ethics is not, as the formalist would have it, "What rules should I obey?" but rather, "What intentions should I adopt?"

The Economic and Social Data Service states that it is a fundamental ethical principle of research to ensure that participants give consent and that they enter into the research voluntarily, without coercion and in the knowledge that they can withdraw at any time, confident in the knowledge that there will be no adverse consequences of participating in the research, including the preservation of anonymity if they wish. In addition to ethical principles, the protection of information is also a statutory requirement contained in the Data Protection Act 1988.

Ethical issues regarding research methods have been primarily concerned with protecting human subjects involved in research, (Meriam, 1991), with the first set of principles guiding researchers in conducting experiments with human subjects dating back to the Nuremberg Code in 1945. Further professional standards have been developed to protect human subjects of behavioural and biomedical research. (Cassell et al, 1980) states that over-attendance to hierarchically imposed guidelines may be irrelevant in the field research and may lead investigators to ignore ethical guidelines that seem inapplicable to their work, or to deny that there are even ethical problems associated with their research.

The case for user based evaluation

A number of researchers have observed that many software developers are still not actively engaging users in the design process. There was some evidence that this was the result either of a lack of knowledge of the available techniques or a perception that usability evaluation techniques were too expensive and too difficult to apply. In 1989 (Milstead et al. 1989) had reported that only 21% of Danish software developers knew about thinking aloud methods and of those only 6% actually used them. Perusal of more recent survey reports and much of the current literature on object oriented design of user interfaces suggests there has been relatively little progress in terms of the incorporation of user evaluation methods. Although (or perhaps because) many evaluation methods have previously been proposed this remains a controversial area. While (Gray & Saltzmann 1998) have challenged the effectiveness of many reported usability methods, (Landauer 1995) provides a review of the value of user-centred design (UCD) based on a wide variety of reported studies where 'some kind of UCD was intentionally applied'. While the average gains were impressive (50%) individual studies reported gains in performance from 0 (NASA) to 720% (IBM). Thus within UCD there is a wide range of both methods and resulting performance. Landauer claims based on his studies that without UCD a user interface has typically forty flaws. He suggests that small performance gains from UCD are the result of incomplete design cycles and the use of 'quick and dirty' UCD activities. He also (Landauer, 1995, p218) outlined how designers tend to take a system-centred point of view and programmers who are involved day and night with a program cannot put themselves in the place of a new user. Developers who are working on their own do not have sufficient domain knowledge, either to be sure of recognising all usability problems (and consequently the design issues relating to them) or to be confident of accurately prioritising these with reference to actual user concerns.

For these reasons the interface design process should be user-centred and needs to be founded on the principles of participation and co-operation between users and developers. However participation and co-operation needs to be carefully planned, an issue clearly supported by Macaulay (1996) when stating:

> *'whilst it can be argued that the requirements process would be enriched by co-operation between stakeholders .... it is by no means clear that interaction between people with ... a diversity of expertise and motivations would result in anything but chaos.'*

Effective participative tools and techniques are therefore required to maximise the contribution that each stakeholder group can make to this part of the interface design process.

However most of the focus on user centred design has focused on the positive benefits of user involvement, and particularly user based evaluation. This has been from the developers' point of view and the improvement of relations between users and developers. User Involvement is also important for building belief in the software product in individualistic users (Tudhope et al. 2000). Users and developers can be seen as on opposing sides, as (Tudhope et al. 2000) describe but these attitudes can change after participating in design teams.

The case for an ethical approach

As the performance tests require the participation of human beings, it is necessary to consider certain ethical issues. (Allen, 1984) states that it is important to conduct tests, with human subjects, with deep respect for the users' emotions and well-being and that they cannot be subjected to any form of destructive testing. Although the risks of becoming physically harmed through usability testing are negligible, undertaking testing can be quite stressful for some participants and they may feel pressure to perform and may also worry about being slow at learning the systems and suffer inferiority complexes if they struggle in the knowledge that all this is happening under observation. In particular, the most highly educated and intelligent users, are concerned about exhibiting ignorance during the testing procedure (Schrier, 1992). In consideration of this points, it is important that the experimenter makes the users feel as comfortable as possible during and after the tests, making it clear that it is the system that is being tested as opposed to the users (Nielsen, 1993).

The positive user-developer interactions described above are not always evident, unfortunately. This is partly the result of the way many evaluations are carried out so that there is an imbalance in power between the developers and the users. This situation has been known for some time. For example (Falzon 1990) describes dialogues between experts and non-experts (e.g. patient-doctor) which are analogous to developer-user situations where the expert speaker soon assumes control of the conversation and the remaining exchange of information follows a sequence of 'yes/no' questions and answers. This is in contrast with the situation where the expert perceives the other speaker as expert in the domain.

The imbalance of power and control particularly occurs when

a) user and developer are from different cultures
b) user and developers are perceived as having different status
c) user and developers are perceived as having different abilities such as occurs in accessibility testing

However the issues of dealing with vulnerable users have seen scant attention in the HCI literature. This neglect can result in distressful experiences for users. For example a recent unpublished report described an evaluation where the users and developers were from a different culture and some of the users were described as illiterate. The task scenarios were then adapted for the illiterate users in a way which was clearly humiliating for them. The western evaluators were then surprised that the illiterate users only completed one of the three tasks and then stopped participating in the evaluation. They also would not agree to complete the extensive pre and post study questionnaires, but how were they expected to do so?

A similar imbalance of power and control can occur in a situation where the evaluators are 'teachers' and the users are students. (Gray & Saltzmann 1998) have criticised evaluations based on these arrangement as potentially unreliable but they did not dwell on the ethical issues raised by these studies.

Experience from other domains

Collaborative, change-oriented inquiry that is based in dialogue, invites entanglement (Lather, 1986). When University and student or teacher researchers work together, they engage in a change-enhancing, interactive, contextualised approach to knowledge-building. Ethics becomes entangled with questions of epistemology because each decision about what counts as knowledge becomes an ethical choice about engaging with and valuing another perspective; these perspectives may differ so radically as to be incommensurable (Bernstein, 1991).

The case study

Usability refers to any part of a system with which a human being interacts. The range of interaction between human beings and machine is large and diverse and, therefore, usability cannot be classified as a single, one-dimensional property of a user interface but, moreover, can be categorised as having five main components: learnability; efficiency of use; memorability; errors and; satisfaction (Nielsen, 1993). This study is testing generic speech synthesis programs, rather than those allocated to a specific system or field of work and therefore participants will be selected with different computer skills levels, different ages, cultural backgrounds and different potential end user requirements for speech synthesis and by so doing this, test results will be obtained which will accurately reflect its full usability for many potential types of systems and uses.

This study uses the standard performance measures as stated in the ISO standards which is particularly suitable at the analysis and final testing stage of the systems lifecycle. A minimum of 10 end users were required to test the system with statistics resulting which are straightforward to compare. Although this evaluation method selected does not find individual usability problems, it is a particularly appropriate method to test human participants.

Before undertaking full testing it is vital that the actual testing procedure is piloted first to eliminate and reduce any errors, inconsistencies and weaknesses that may be encountered with the proposed testing strategy and testing plan. These errors may include instructions that are incomprehensible for the users; misinterpretation of the instructions; clarity and usability of questionnaires; time management to complete the tasks and appropriateness of the ease of the text (Nielsen, 1993). Pilot testing, therefore by its very nature, is on a much lesser scale and can involve far fewer participants. (Nielsen, 1993) considers that test users should still be as representative as possible of the intended users of the system for pilot testing and for comprehensive further testing, users from several different subpopulations should be included to ensure all main different categories of expected users are covered. (Nielsen, 1993) suggests that at least ten participants are needed for performance measurement in full tests.

As part of this research study, it was necessary to undertake an analysis and evaluation of usability of text to speech synthesis, with particular emphasis on diversity with the collation of data of end users of differing gender, age, differing levels of computer skills and cultural orientation. This paper explains the procedures that have been undertaken to ensure that all ethical issues associated with usability testing including human subjects, have been addressed prior to the undertaking of such tests. (Nielsen 1993) recommends certain main ethical considerations for testing with human participants as shown in the following tables with annotations on how this is to be applied to the usability testing procedures

Ethical considerations and applications to practice

Ethical Consideration – Before the Test	**Application to Pilot Test**
(i) Have everything ready before the user turns up.	(i) The test room, computers, software, tasks, response booklets and other materials will be ready in advance of the participants' arrival at the computer lab.
(ii) Emphasise that it is the system being tested, not the user.	(ii) The participants will be given this information in writing on their consent forms and this will also be re-iterated verbally before the test.
(iii) Acknowledge that the software is new and untested, and may have problems.	(iii) The participants will be informed that the software, although it is not new and has been tested before, has not been tested for users from multi-cultural environments and that there may be inherent problems with the software and that, subsequently, the experimenter is testing to identify any faults so that they can be addressed in the modelling of a bespoke prototype.
(iv) Let users know that they can stop at any time.	(iv) The participants will be given this information in writing on their consent forms and this will also be re-iterated verbally before the test.
(v) Explain any recording, keystroke logging, or other monitoring that is used.	(v) No audio or visual recording or keystroke logging will be used. Participants will be informed in writing on their consent forms that the experimenter will be monitoring the performance of the system throughout the test.
(vi) Tell the user that the test results will be kept completely confidential.	(vi) It will be clarified to the author's participants that all of the results will remain anonymous, used for the purposes of academic research only and will not be divulged to any employer or teacher. Participants will be identified only be encoding for administration purposes by the experimenter as participant 1, participant 2, for example.
(vii) Make sure that you have answered	(vii) Participants will be given an

all of the user's questions before proceeding.	opportunity to ask questions about the test and testing procedure before the start of the test.
Table 2	
Ethical Consideration – During the Test	**Application to Pilot Test**
(i) Try to give the user an early success experience.	(i) The experimenter will provide word sets with the lowest potential user based error rates first.
(ii) Hand out the test tasks one at a time.	(ii) The experimenter will hand out the tasks for trial one and when completed, will then hand out the tasks for trial two.
(iii) Keep a relaxed atmosphere in the test room, serve coffee and/or have breaks.	(iii) Due to Health and Safety regulations in force at the University where testing will take place, and the timing aspect of the testing, refreshments will be made available after the completion of the test, outside of the computer lab.
(iv) Avoid disruptions: close the door and post a sign on it. Disable telephone.	(iv) A no entry sign will be placed on the door of the computer laboratory where the test is being conducted to avoid distractions and disruptions. All participants will be asked to turn off their mobile telephones. There is no main telephone.
(v) Never indicate in any way that the user is making mistakes or is too slow.	(v) The author will not interact with the participants in any way during the test, except in the circumstances where the participant is distressed and wishes to stop the test or is experiencing hardware problems with the computer.
(vi) Minimise the number of observers at the test.	(vi) The experimenter will be the only observer at the test.
(vii) Do not allow the user's management to observe the test.	(vii) The particpants' employers or teachers will not be involved in the conduct of the tests in any respect. Information about student performance will not be made available to their teacher.
(viii) If necessary, have the experimenter	(viii) The experimenter will stop the test

stop the test if it becomes too unpleasant.	at any time if a) a partipant requests this or b) the author considers that a participant is displaying signs of distress through the testing procedure.

Table 3

Ethical Considerations – After the Test	**Application to Pilot Test**
(i) End by stating that the user has helped you find areas of improvement.	(i) The experimenter will thank the participants for their invaluable assistance in testing the usability of the software and that the results will be used to model a prototype that will address any weaknesses that may be identified as a result of the test.
(ii) Never report results in such a way that individual users can be identified.	(ii) The experimenter will identify participants only by methods of encoding – participant 1, participant 2, for example – for administration purposes only.
(iii) Only show videotapes outside the usability group with the user's permission.	(iii) The experimenter will not be video recording any aspect of the testing procedure.

Reflection in practice

Because of the sensitivity of the nature of the tests and the inclusion of students, it was necessary to submit a proposal for pilot testing to the Thames Valley University Ethics Committee, which consists of ten senior academics with particular expertise in ethical issues; written authorisation from the committee is essential before any testing can take place. This submission included a forty page report outlining the types of test that would be carried out, how ethical issues would be addressed together with questionnaires collating personal information on students which includes: age; how many years they have lived in England; their place of birth; nationality; cultural origin; sex; first language and first language of their parents/guardians. Consent forms to participate in doctorate research, information sheets and the test questions were also submitted to the committee.

For this study, the ethical guidelines set out by Nielsen were followed. However there is a strong case for revising and extending these guidelines to deal with evaluations dealing with accessibility and cross-cultural issues which may involve vulnerable adult users or children. Taking Macmurray's approach to ethics we need to consider not just the rules but the intensions and consequences of the guidelines. We know about the problems with the illiterate users because the evaluators were prepared to be candid about their experiences. We would all benefit from a 'no blame' culture were HCI professionals could report their failures as well as their successes so that the community of practice could learn and document these errors in the same way as is now common in the medical profession.

CONCLUSION

It was not expected, at the time of commencing the doctorate study, that it would be necessary to seek approval from the University's Ethics Committee. The research into ethical issues, preparation of a forty page report and all necessary consent forms was extremely time consuming but however, from this research, an appreciation has been gleaned of how vital ethical considerations are when including human participants, and how it is important not just to protect the participants, but also the experimenter as case studies have also been looked at where experimenters in America undertook usability testing in Asia with participants with specific learning difficulties who left the testing in a state of distress because the experimenters had not considered ethical issues.

REFERENCES

Allen, R. B. (1984). Working paper on ethical issues for research on the use of computer services and interfaces. *ACM SIGCHI Bulletin.* **16.** (1), pp 12-16.

Bernstein, R.J. (1991). *The new constellation: The ethical-political horizons of modernity/post modernity.* Cambridge: Polity Press.

Cassell, J, Murray L. (1980) Ethical problems of fieldwork. *Social problems (special issue).* **27,** pp 259-378.

Economic and Social Data Service. (1993). [online] www.esds.ac.uk. Accessed 10 October 2005.

Falzon, P (1990). Human-computer interaction: Lessons from human-human communication. In P.Falzon (ed.) Cognitive Ergonomics (pp 51-68). London: Academic Press.

Gray, W.D. and Saltzman, M.C. (1998) Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction*, 13, 3, 203-261

Landauer, T. K. (1995) *The Trouble with Computers*. MIT Press.

Lather, P. (1986). Research as praxis. *Harvard Educational Review,* **56** (3), pp 257-277.

Macaulay, L. (1996), Requirements for requirements engineering techniques, Proceedings of the second international conference on requirements engineering, (ICRE'96), IEEE Computer Society Press, pp 157-164.

Macmurray, J. (1938) *The Clue to History*; London: SCM Press.

Maiden, N. A. M. and Rugg, G. (1996), ACRE; selecting methods for requirements acquisition, *Software Engineering Journal*, 11(3), 183 - 192.

Meriam, J.L. (1991). The decline of academic standards. *Engineering Education.* **81,** pp 405-407.

Milstead, U., Varnild, A. and Jorgensen, A.H. (1989) Assuring the quality of user interfaces in system development (in Danish). Proc. Of NordDATA'89 Joint Scandinavian Computer Conference, pp 479-484.

Nielsen, J. (1993). *Usability Engineering.* Academic Press.

Schrier, J.R. (1992). Reducing stress associated with participating in a usability test. *Proceedings of the Human Factors Society 36th Annual Meeting. Pp 1210-1214.*

Tudhope, D., Beynon-Davies, P. and Mackay, H. (2000), Prototyping Praxis: Constructing Computer Systems and Building Belief. *Human-Computer Interaction*, 15, 353-383.