INTEGRATING IPTV AND SOCIAL NETWORKING WITH VOICE INPUT

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Master of Computer Science

In the Department of Computer Science

University of Saskatchewan

Saskatoon

By

AFSHANA SYED HUSSAIN

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science

176 Thorvaldson Building

110 Science Place

University of Saskatchewan

Saskatoon, Saskatchewan

Canada

S7N 5C9

ABSTRACT

Elderly people comprise the highest proportion of television viewers. Elderly people often struggle with new technology and reject it due to complexity. We propose a system to help people keep up with certain new technologies, such as IPTV and social networks with reduced efforts. We specifically propose a system to integrate IPTV with Twitter, a social networking website with an aid of a mobile phone. The system uses speech to text technology on mobile phone, as input to reduce the difficulty involved in the interaction with Twitter, while viewing television. As speech is a more convenient and natural way of expression than text, we anticipate that people from other age groups can also benefit from the system.

ACKNOWLEDGEMENTS

I wish to express my gratitude to my supervisor, Dr. Julita Vassileva for her constant support, patience and guidance. She has helped me extensively during the research and thesis writing phase. Her immense knowledge in the social psychology has been a major factor in my research work. I would also like to convey my heartfelt thanks to members of the supervisory committee, Dr. Ralph Deters and Dr. Gord McCalla for their encouragement, valuable inputs and comments. My Special thanks to Hashim Al-adhami for his collaborative effort in coding the IPTV platform, which was integrated into middleware of the proposed system. Thanks to all my lab mates from MADMUC Lab, for their inputs, invaluable assistance and support during presentations, research and evaluation of the system. Especially, I wish to express my love and gratitude to my beloved family for their understanding & endless love, through the duration of my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

IPTV	Internet Protocol Television
ITV	Interactive Television
TV	Television
VOD	Video-on-Demand
DSL	Digital subscriber line
ATIS	Alliance for Telecommunications Industry Solutions
CATV	Cable television
IP	Internet Protocol
LAN	Local area network
PNA	Phone-line networking alliance
DVD	Digital video disc
IIS	Internet information services
MRML	Media room markup language
IPTVCD	IPTV customer device
XML	Extensible markup language
API	Application programming interface
URL	Uniform resource locator

CHAPTER 1
INTRODUCTION

Television (TV) provides a valuable source of information about sports, news, current affairs, and other areas of life. Many TV channels nowadays are solely dedicated to the needs of viewers of various age categories. Many elderly people live in relative isolation and television fulfills their need to be connected. With the advent of social networking, the world has become open and more connected, since many social-networking sites (such as Facebook, Twitter, Myspace, LinkedIn, Flickr) provide entertainment and help people to connect with friends. They also serve as a major information source as well as a communication channel. People share information and updates regarding current news and trends they find interesting. People also use features such as status updates on social networks, commenting on blogs and emails to communicate with one another. People in all age groups use social-networking sites to stay connected to friends and family. Statistics show that the percentage of elderly population using social-networks is increasing in the developed countries, though it is still much smaller than that of the younger population. The reason for the lesser participation of elderly people on social networks is due to unfamiliarity with of computers and lack of confidence and/or need to learn new ways of interacting with technology.

TV can provide an alternative way of engaging elderly people in social-networking sites that could be easier and more natural for them, since they are well familiar with TVs. Modern TV technologies (Interactive TV and IPTV) allow the users to interact with the television content, service provider and other viewers. Although, interactive television has been around for a long time, Internet protocol television (IPTV) is a new technology. In the foreseeable future, many North-American households will have IPTV, and many television service providers are already venturing into that direction. With the entry of IPTV there is an increasing need for social-networking experience to be integrated with interactive television experience. Social networking through IPTV can help elderly people remain connected with their children, grand-children, friends and other family members. To make the navigation among the complex functionality of IPTV easier for elderly people we propose to integrate IPTV to social-networking websites with an aid of mobile phone, where speech to text technology is used as input to reduce the difficulty involved in interaction, during television viewing activity.

The thesis is organized as follows: in Chapter 2, a survey of the current technologies used by elderly people to communicate is presented. Furthermore, the specific needs of elderly people and how they use different new technologies are discussed. In Chapter 3, we present our approach that integrates speech technology and mobile phones, with web-based social-networking sites and IPTV. Chapter 4 presents the implementation details and Chapter 5 - a pilot study with the system prototype. Chapter 6 describes the evaluation of the improved system prototype. Chapter 7 discusses the potential contribution and presents directions for future work.

CHAPTER 2
LITERATURE SURVEY

This literature survey covers a wide range of related literature and topics from different areas, including an overview of internet protocol television and elderly people, in terms of their needs and participation in the social networking, television viewing, and mobile phone usage. The literature survey will also focus on the role of interactive television, recent speech to text and mobile phone technologies. In later chapters, we will focus on integrating all these different technologies to build an application which will provide convenience to the elderly and other age groups.

2.1 Internet Protocol Television

IPTV is an abbreviation for Internet Protocol Television [36]. Internet Protocol TV (IPTV) is a system through which Internet television services are delivered by the architecture and networking methods of the Internet Protocol Suite, that is using the Internet and broadband Internet access networks, instead of being delivered through traditional radio frequency broadcast, satellite signal, or cable television (CATV) formats [53],[22]. The standard services of IPTV include:

- Live - television [22, 14].
- Time-shifted programming [22, 14].
- Video on demand (VOD) [22, 14].

An official definition of IPTV is given by Alliance for Telecommunications Industry Solutions (ATIS) IPTV exploratory group in 2005:

"IPTV is defined as the secure and reliable delivery to subscribers of entertainment video and related services. These services may include, for example, Live TV, Video on Demand (VOD) and Interactive TV (iTV). These services are delivered across an access agnostic, packet switched network that employs the IP protocol to transport the audio, video and control signals. In contrast to video over the public Internet, with IPTV deployments, network security and performance are tightly managed to ensure a superior entertainment experience, resulting in a compelling business environment for content providers, advertisers and customers alike."[14].

IPTV is a part of telecommunication marketing term "quadruple play," which includes the internet, phone, television, and wireless technologies [52]. IPTV delivers content, which include high-quality television broadcast, audio and video content [36]. Unlike traditional Internet TV the IPTV architecture delivers the information content by a private network (Delivery network) or secure internet connection [52]. The IPTV supports the interaction with the content, the IPTV server and other users, in addition to traditional television functionalities [52].

IPTV interactive services include services such as gaming, time-shifted programming, personalization of television content, channel selection, multi-tasking services and integration with other services [52].

Figure 2.1 shows the main components of the IPTV architecture. The IPTV architecture has components, which include data center, delivery network (private network), IPTV customer device (IPTVCD) and home network [36]. In the following section, the various types and functions of each of the components will be discussed.

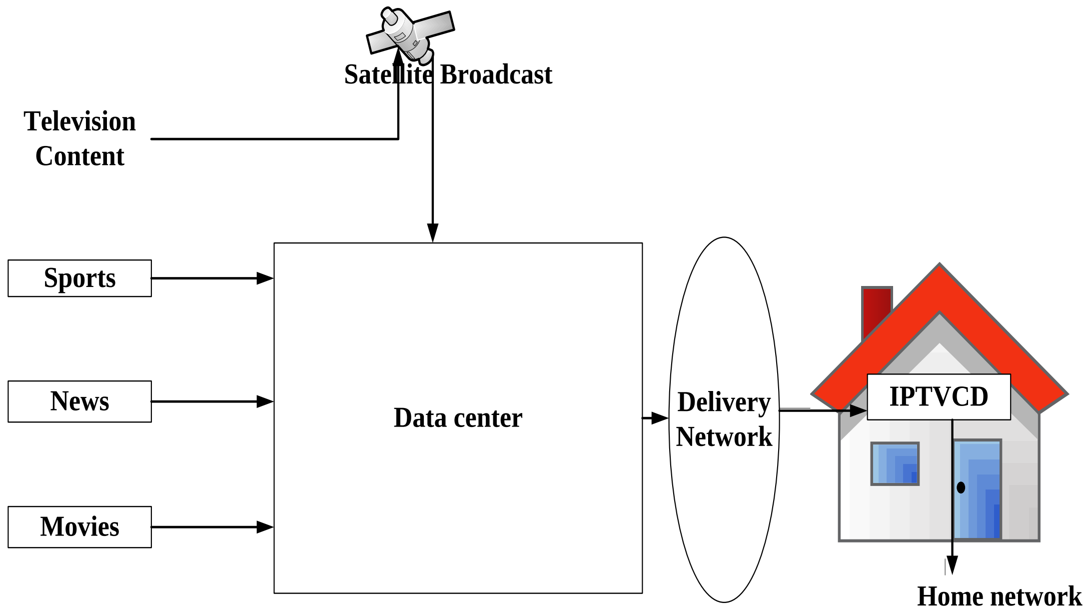

Figure 2.1 IPTV Architecture

The data center has many responsibilities [36]. It receives the content from variety of sources (for example: news, sports, movie), and prepares video content for delivery. It also manages profile and payment information of the IPTV subscribers [36].

IPTV uses a number of delivery networks to deliver the content. The types of delivery network include fiber network, DSL network, satellite network, fixed and wireless network, cable television network and the Internet. The fiber network has dedicated fiber links suitable for IPTV content delivery. The fiber networks can be implemented in various ways, which include, Fiber to the regional office (FTTRO), Fiber to the neighborhood (FTTN), Fiber to the curb (FTTC), Fiber to the home (FTTH) and Fiber to the apartment (FTTA) [36]. The main difference between each of these fiber networks is based on the type of installation. The DSL network is a digitally subscribed line technology used on telephone lines to deliver IPTV content. The satellite networks use satellites to deliver IPTV content, which have high speed and bandwidth. The fixed and wireless network uses the wireless technologies such as radio transmission and mobile phones for the delivery of IPTV content. The standard cable television network can also be used for delivering the IPTV content. In addition, the Internet can also be used to deliver the IPTV content.

The IPTV content delivered through the delivery network is received by IPTV consumer devices. The IPTV consumer devices decode the incoming IP based video stream. The standard IPTV consumer devices include: set-top boxes, gaming consoles and media servers [52, 36].

The home network enables the distribution of data, voice and video between different consumer devices [36]. The home network technologies include: Fast and Gigabit Ethernet, Wireless LAN, Home Plug AV, Universal Power-line Association-Digital Home Standard, Home PNA, Multimedia over Coax Alliance [52]. Finally, home networks deliver the television content to the IPTV for viewing.

The IPTV service providers in Canada include: Aliant TV, Bell TV, MTS TV, Sasktel, Max and Telus TV [13]. In US, the major companies offering IPTV services include AT&T and Verizon [13].

In the past, this technology has been restricted by low broadband penetration and high cost of installing and wiring capable of transporting IPTV content reliably in the customer's home. The projected growth of residential IPTV in year 2010 was about 400 million households [14]. Many telecommunications providers are exploring IPTV as a new market for revenue opportunity. There has been an increase in the IPTV installations within schools, universities, corporations and local institutions as well.

FCC Media Bureau Chief Bill Lake had said earlier that TV, and the Internet would soon be the same, but only 75 percent of homes had computers, while 99 percent had TV [14]. A

Nielsen survey conducted in 2009 verifies 99 percent of video viewing was still done on TV [14].

2.2 Elderly people needs

As people age, they have an increasing need to be connected, esp. in case of an emergency [38]. Elderly people need to feel secure and independent without violating their privacy. Many elderly people have difficulty and lack the motivation to learn new technologies. Furthermore, elderly people feel intimidated by the overwhelming complexity they perceive in the technology and have limited finances to afford these technologies [1, 37]. The affordability of the technology is a common requirement for all the age groups.

Many elderly people suffer from impaired vision or hearing, and low manual dexterity. That is why elderly-friendly technology should use simple interfaces, which require less efforts from the users [1, 37]. The common requirements for the elderly people are that any gadgets need to have an easy and intuitive design, user-friendly interface and should be affordable [1, 37]. The design should also take into account the common conditions among elderly, such as low vision and limited manual dexterity. Some common design requirements for elderly people include developing cell phones with large solid buttons and feedback or task-completion acknowledgement capabilities (for example, a button click sound) [37]. The gadgets need to serve the three basic needs for elderly people, which include security, freedom and hazard management [1]. The common telephone line is considered a typical assurance system, as it fulfills the need for safety [38]. The mobile phone and the telephone provide a sense of security to the elderly, by fulfilling the need to be connected. It is always comforting to know people are just one call away if hazard strikes, e.g. a fall. Life has become much more convenient and self-assuring with mobile phones.

So, while considering building any new technology, it is important to consider the user-friendly interface and intuitiveness of the design in accordance with the special needs of the elderly.

2.3 Television viewership and the elderly

Television viewing is considered to be a leisure activity. In today's fast-paced life people have less time to view television. The younger population turns to other online sources for information, entertainment and news. It is common knowledge that elderly adults watch more television than any other age group [15]. The elderly people (aged 65+) represent the major television viewer-ship in many countries. According to Nielsen data [15], the average TV viewing time for adults aged sixty-five years or older is between four to five hours per day, as compared to two to three hours for younger adults. The higher viewership among elderly people is due to having more free time after retirement. The television satisfies their need to be connected. Television serves as a valuable source of information, entertainment, and helps the elderly to keep up with the changing times. It gives them a sense of belonging and helps to counter the loneliness. It is also the most convenient medium, as it requires little effort to switch on/off the television or to change the channel.

Many television channels provide a variety of content to match the individual tastes and interests. It is also a cheaper source of entertainment and does not require leaving home. Nowadays, people can watch movies of choice on demand. There are facilities available to record movies and shows, and play them at a later time. The live telecast of the sports, news and entertainment is also engaging and helps elderly viewers to stay informed. According to a study [15], the elderly viewers watch more factual content compared to younger viewers. The elderly people are more interested in news and education programs, than in entertainment shows.

According to the TV viewership age division, the 65+ aged seniors account to 17.4% of total television viewers. The highest TV viewership is in the age group (55 to 64) at 17.7% [46]. So, people aged 55+ accounts for 35.1% of the TV viewership.

2.4 Email usage

Email is still the predominant way of communication on the Internet. Though the email usage by elderly people is lesser than that of other age groups, the usage has increased at a steady rate. Elderly people (aged 65+) use email very frequently according to a demographics study [10]. The study shows that older people use email within a close-knit circle consisting of either relatives or friends, with whom they share a few emails a month. However, the elderly don't use email as a means to establish new relationships [10]. The lower rate of email usage by the elderly when compared to the younger age groups is a result of the lesser inclination to use computers and other electronic devices. If the elderly have visual or mobility impairments, then typing an email can become a difficult task. Elderly people can benefit if the email clients use a more natural mode of data input instead of typing. There are many applications that offer dictation of the email to people. Though the current speech technologies are not perfect, there has been a rapid improvement in recent times [10].

2.5 Social networking usage

Senior citizens have increased their presence on social-networking websites over the recent years. A report by the Pew research center for the Internet & American Life project [54], shows that between April 2009 and May 2010, social-networking use among internet users aged 65+ grew by 100% (from 13% to 26%). In comparison, the social-networking use among users aged (18-29) grew only by 13%—from 76% to 86%. Though young adults continue to be the heavier users of social-networking compared to elderly people, research shows that elderly people are still interested in using social networking [54]. A lot of people in the elderly age group (63 to 75) use Twitter frequently. Twitter usage by the elderly is at seventeen percent (17%) in the elderly age group [21]. Most (90%) of the elderly users of social media in US, have picked up Facebook [21].

The number of elderly people, holding social-networking account has grown steadily from 10% in 2007 to 14% in 2008. In 2009, the percentage of elderly people holding a profile on the social-networking site reached 36%, which is more than one-third of the elderly population [21]. Though the study is restricted to the US and the rest of the world may show different social network penetration levels, it shows a trend that elderly people are quickly adopting social-networking websites. Elderly people are willing to learn and use the new technology to communicate with children, grand-children, other family members and friends. If the access to social-networking sites is made simpler, more elderly people will be willing to participate.

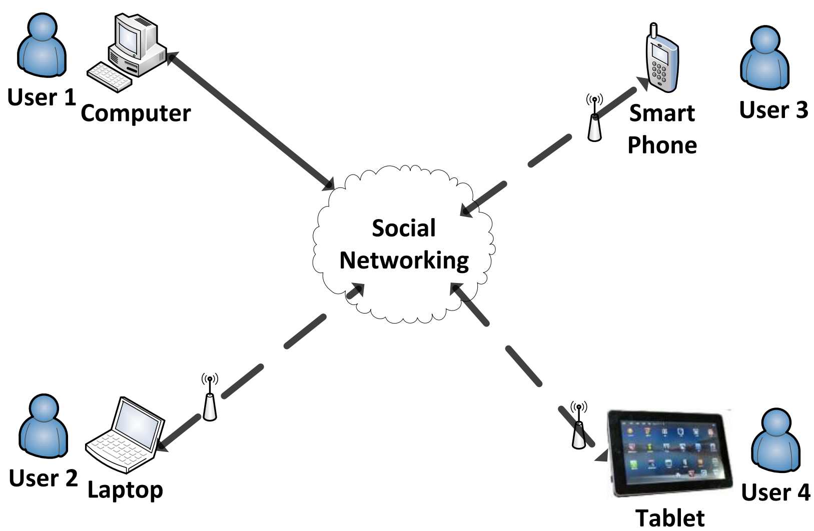

Figure 2.2 Social-networking usages

Figure 2.2 illustrates how social-networking can be done through various devices, such as desk-tops, laptops, tablets and smart phones. Facebook is the most well known social network currently. Facebook users are increasing at a very high rate as compared to the users of other social-networking websites. Facebook is used widely by people in all age groups. Currently, there are 800+ million users worldwide. As per checkfacebook.com, Facebook has 7,792,760 elderly users, which is 5% percent of the total user population of Facebook [20]. The continent with the largest number of elderly Facebook users is North America, and the country with the highest number of elderly Facebook users is the United States. It is closely followed by Asia and Europe [45].

Twitter is one of the major social-networking sites for micro-blogging, where the users share data such as instant message updates restricted to 140 characters in length, with other users who have subscribed as their followers. Twitter has fewer features than Facebook, which makes it a more streamlined platform for sharing news. That is why it is often used by celebrities and politicians to share news with their fans and supporters. It is also used within the work environment. The statistics show that between 2008 and 2010, the number of Twitter users of all age groups grew significantly [4], Even though elderly people comprise only 2% of the Twitter users, studies show a gradual many increase [41].

2.6 Interactive television

Interactive Television should be of particular interest to elderly people since it would allow them to access several services at home. The interactive television gives the elderly an opportunity to browse and carry-out e-commerce activities. It can motivate them to be more active online. It fulfills two of their basic needs, which include being independent and being socially connected [6].

By interactive television the elderly can interact with their online community and service provider, which in turn will give them a sense of control. The television viewing will become a social activity, if interactivity with other users is introduced. There are already options in interactive television for the users to send SMS, which are displayed on the right side or the bottom of the television screen. The users can choose to discuss and share information about the current television content which includes music, movies, sports and news. E-commerce activity can be extended to the television set helping the user to view, choose and buy an item of interest from the comfort of the living room. This technology can serve well to the elderly people who cannot move around freely due to physical restriction, for example allowing them to purchase groceries or clothes online and get them delivered home.

The standard remote control is a frequently used user-interface for household commercial media devices [24]. By pressing buttons, the user can switch TV channels, increase or decrease the volume, etc [24]. The arrow keys help to navigate through onscreen display menus [24]. Interactive television applications are mostly controlled using colored coded keys [24]. The conventional remote controls are not the most convenient devices for television. Firstly, there are various remote controls to operate different devices like Television, DVDs and other devices. An additional effort is required by the users to remember which remote controls are used for which devices [31]. Secondly, to allow control over increasingly complex functionality, remote controls have evolved complex layouts, in which navigating to a given feature is quite laborious. Once a specific feature is reached it is hard to back track to the main television viewing screen in an interactive mode [31]. In order to use conventional remote controls, people are expected to comprehend complex instruction manuals and remember functions of each button [28]. The layout of the remote controls is not too intuitive and places a high cognitive load on the user

[31]. The buttons on the remote control are generally small, and the users find it difficult to find the button related to the functionality. A study shows users (especially elderly) preferred face-to face interaction or human assistance instead of reading the complicated remote control manual [31]. The complicated layouts and manuals is one of the main reasons for elderly people to become disinterested in a technology [31].

One of the solutions to bypass the problems encountered in the remote control is using a gesture based TV remote control [29]. In a study, a set of gestures which can be useful in television viewing were obtained by observing the user behavior, and after that appropriate gestures related to the remote control features were designed, and later the acceptance of those gestures was evaluated [29]. Other solutions include pRemote [24], which allows creating personal interfaces by the user and helps using different templates for the different applications [24]. The pRemote has a glass mount in which different templates can be inserted [24]. The templates can be switched in the pRemote at anytime. The various templates in the pRemote are designed to control different services such as interactive games and messenger functions. The reception unit in the pRemote can identify the user using the template and provide the personalized representations on the output unit [24]. In addition, pRemote has a digital pen for writing as an alternative to text input devices or touch screen devices [24]. The advantage of using pRemote is supporting various functionalities without exchanging the underlying hardware infrastructure [24]. The pRremote has a simple interface which is easy to understand. There are also other approaches using new devices. Enns and MacKenzie study [16], presents a remote control with an additional touch pad, on which the program numbers can be written with a finger to switch TV channels. In a study by Myers et al. [34], the usability of a laser pointer to point out the hand fluctuations cause inaccuracies. MacKenzie and Jushoh study [32], examined the handling of a Gyromouse for interaction control from a distance. In Fuhrmann et al. study [19], a pen-like device enables a user to control applications via Bluetooth by using hand gestures. There is literature which focuses on using devices, which are primarily not designed as remote control, such as PDAs (Robertson [43]) or mobile phones (Barros et al. [2]) as a replacement to standard remote control. In a study, a mobile phone enabled with Bluetooth was used in place of a remote control, and the evaluation showed that it served the purpose well [2]. In a paper about sociable TV [42], it was discovered that the older adults liked a remote control interface with only six buttons, as it was easy to understand and navigate [42].

In summary, research has shown that the traditional remote control is not the best device for interactive interface, and other devices such as mobile phones can offer a more straight-forward interaction [26].

2.7 Mobile phone ownership

Cell phones have become an important part in our lives. Everyone from teenagers to seniors carries a cell phone. This has been made possible as a result of cheaper device and service costs. The other reason is the introduction of pre-paid plans, which allow low commitment and risk options to own cell phones. Many elderly people are buying cell phones as a result of affordable cell phone and services. Cell phones provide both convenience and security to their owners and aid them in being independent. Pew research center's the "Internet and American Life" project in August 2010 gives some interesting insights. It is a survey conducted for all the age categories from 18 to 75+ [3]. The primary focus was on the mobile phone usage pattern by the elderly aged 65+. In the study, the elderly were asked two main questions [3]: 1. Do they possess any mobile phone or a smart phone, such as Blackberry or iPhone? 2. Do they use their cell phone and how?

According to the survey [3], 68% of the elderly (aged 66-74), and 48% of old people (aged 75+) respondents owned cell phones. From the survey results, it can be deduced that the elderly preferred to do simpler tasks on the mobile phone, such as taking a picture and receiving/making a call. They did not prefer using the mobile phone as a major entertainment source for playing games or music, nor did they use the mobile phone much for internet access or even to send or receive email. The reason for that is the more complex interaction, and unfriendly user interfaces. The buttons on most of our new cell phones are tiny and to type a letter one needs to press the same button several times quickly. As the buttons are tiny, it becomes hard to type messages, especially for elderly users who have low vision and manual dexterity.

The new generation of mobile phones and smart phones increasingly involve speech to text technology as a form of data input, to bypass the difficulties in typing data on the small keyboard or number pad. Speech to text technology also provides a less explored way of interacting with TV and the other media equipment. The next section provides a brief overview of speech to text technologies.

2.8 Speech to text technologies and mobile phones

Speech to text technologies have been used to help the deaf and hard of hearing in recent times. Such technologies are used, for example, in lecture rooms to convert all the verbal lectures into textual format to equip deaf students [50]. Speech to text technologies has been used in education to improve the language skills of students. Project INTELL uses Carnegie Mellon's Sphinx III speech recognition, which helps in gaining proficiency in an Indian language (Telugu) [17]. Another project, called LISTEN helps in increasing the oral proficiency of children in their native language [17]. Speech-recognition software by Microsoft (Automatic Reading Tutor) helps students in learning to speak/read in the first language and second languages. Apart from these projects, there are different systems that focus on diction or utterance of individual character or long sentences for a native language speaker or second language speaker [17]. pARLING is system used in games to increase the vocabulary of second language [17]. Speech recognition software ProTutor, selects the sentences based on user's difficulty in pronouncing a few words (sounds) [17].

Speeches to text technologies have made inroads in ubiquitous computing environments, for example, car navigation systems. There are also many such applications on smart phones, such as Dragon dictation, Dragon search, Google Voice Search, Bing voice search, the Siri Personal Assistant on iPhone, Shoutout and Vlingo [35]. They assist in voice dialing, social-networking, composing emails and messages by using voice. During driving, the user can compose emails, searching for places, and use messaging with minimum hand movement. The speech to text technology aids in multi-tasking.

The introduction of a speech interface can facilitate and motivate the elderly people to be active in using all the reviewed technologies, which include IPTV, interactive TV, social networks and mobile phones. Speech is the most natural way of human communication. Researchers have shown that most people prefer to accomplish their task by speaking rather than typing text [7]. For example, a study shows that participants from all age groups preferred speaking annotations rather than typing them [7]. Speech technologies are of interest to the elderly group, as it can help them overcome manual dexterity and low vision restrictions. Speech technologies can help elderly people to avoid the strenuous task of typing. By speech, they can accomplish the task more easily and efficiently, as correcting errors by re-recording a message is faster and easier than retyping it. Though the accuracy of speech to text conversion depends on the software and the ability of the user to speak clearly, it is still better to use speech with some editing (typing) than to type the whole sentence.

Problem Statement

Our goal is to provide a convenient way of interaction with IPTV, by incorporating speech technology on mobile phones to access social networks via IPTV.

CHAPTER 3
PROPOSED APPROACH

This chapter discusses the main objective and motivation for the proposed approach. The objective is to create a middleware, as a way to integrate different technologies such as IPTV, social networking and speech recognition technologies. The proposed approach specifically aims at providing a convenient means of interaction by using speech technologies on mobile phone, to access social-network while watching IPTV. It has two main phases. The proposed approach has 2 parts: First, the mobile phone receives speech input from the user, converts it into text messages and sends it to Twitter. Second, the middleware receives messages or updates for a social network (in our case, Twitter), and displays them on the IPTV simulator screen. The aim is to accomplish a connection between a mobile phone, IPTV and Twitter.

This chapter gives a brief introduction about the main features of the proposed design and approaches used during implementation.

3.1 Using IPTV

The thesis targets IPTV because it is a new technology, which is likely to have a strong influence on television viewing experience. The IPTV is expected to grow and become the main TV technology soon. The vision of the future of IPTV is that it will eliminate the differences between the internet and television. Everything that can be accomplished on the internet will become possible on television. However, one main difference remains, the television is designed for audience sitting at 10 feet, and the internet/computers are designed for users sitting 1.5 feet away. On computers, users are able to interact with a keyboard or a touch screen. If the interface can be changed to accommodate the internet experience on the television, it will open doors for new possibilities to integrate the TV viewing experience with internet browsing. The IPTV currently has necessary hardware and software to venture into the area. Microsoft Mediaroom provides a programming environment to simulate an IPTV system, and expertise can be used to develop prototypes exploring new interface options.

3.2 Using Twitter

As discussed in the literature survey, elderly people account for 2% of the Twitter users. So, Twitter seems to be a good option for the integration. Twitter is a light weight social-networking site. It can serve as a good starting point. Moreover, Twitter is a high response site. Integrating IPTV to Twitter can provide some interesting insights on the performance, in terms of speed. Twitter also has a well defined API that is easy to learn and use, and Twitter APIs are not frequently updated like other social networking sites, e.g. Facebook. That is why Twitter was chosen as the social network to be used in our application.

3.3 Using Blackberry Smart Phone

With the intense competition among phone producers, elderly people have better access to phones these days. One commonly used smart phone is Blackberry. There are many third party applications available for Blackberry for speech to text conversion. One advantage of Blackberry over other proprietary platforms, such as the iPhone, is that applications can be created for Blackberry using the java programming language and standard software-development platforms, such as Eclipse. For this reason, we have chosen to use the Blackberry platform to develop our application. Since other smart phones such as Android and iPhone also offer speech-to-text applications, the approach can be extended to other smart phones too.

3.4 Using Speech to Text

Speech is the most natural way of expression for people from all ages. The speech to text conversion can benefit particularly the elderly and visually or physically challenged people, as it helps them overcome manual dexterity restrictions and the strenuous task of typing. By speech, they can accomplish the task of communicating with IPTV or social networks more easily and efficiently.

For the proposed system, many speech recognition applications were explored to find a suitable option. Initially, a speech recognition application was developed with Microsoft Speech API (SAPI) and Microsoft Visual studio .NET using C# programming language. The developed speech-recognition application had a very low accuracy for speech to text conversion.

At a later time, Dragon naturally speaking (third-party) software was explored to analyze the speech to text conversion capabilities. The Dragon speech recognition software had a better speech to text conversion accuracy compared to the developed speech recognition application (with Microsoft SAPI), but it required voice training. After exploring a number of other speech recognition applications, Vlingo was identified as the best fit for speech recognition due to five reasons: First, Vlingo virtual assistant had better accuracy as compared to other speech recognition software. Second, it had an in built Twitter client interface to post updates to Twitter website using speech. Third, it can be installed and used on most smart phones. Fourth, it is available for free download on Blackberry smart phone. Fifth, development effort was not required for the integration into our application.

3.5 Overall approach

The proposed middleware will use a third party speech to text software in a smart phone which will take the speech input and convert it into text and send it to the Twitter website using

Twitter client on smart phone. There will be a custom developed middleware including the "IPTV Web Server" and "Twitter Application Server" (in Figure 3.1), which will serve as an interface between the Twitter website and the IPTV. The Twitter Application Server will be responsible for taking the incoming request of user profile data from the IPTV platform, fetching the required information related to the user profile from the Twitter website and transferring user's stream from the Twitter website to the IPTV platform.

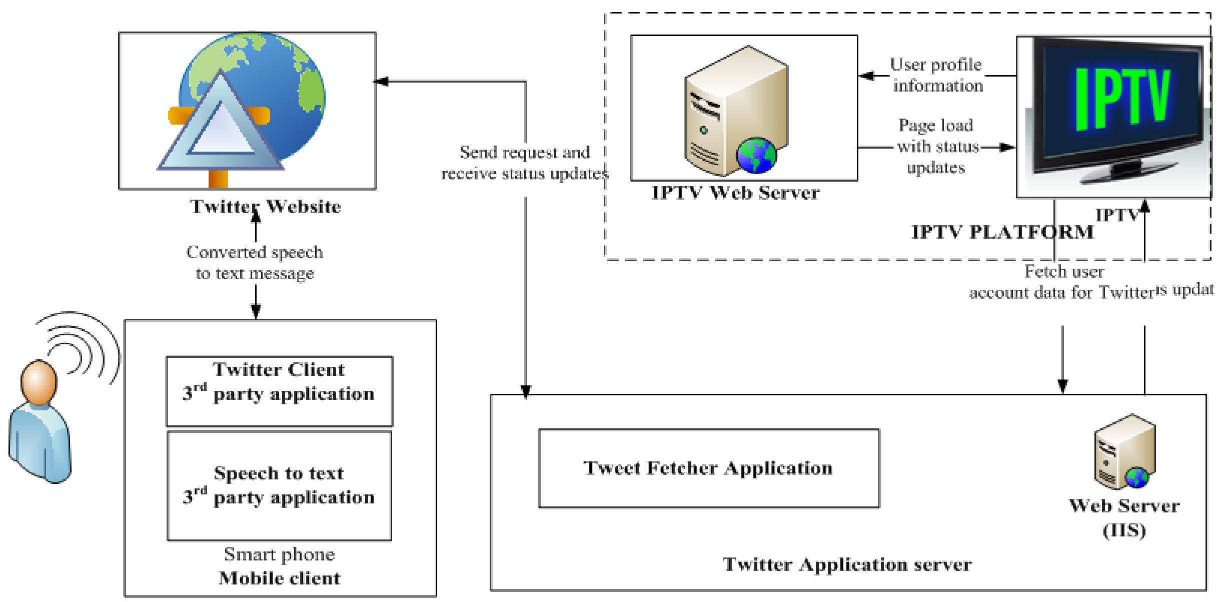

Figure 3.1 Proposed approach

The IPTV web server will be accountable for displaying the messages included in the Twitter stream on the IPTV (simulator) screen, during the page loads.

3.6 Scenario

Let's say Bob and Allen are friends who live in two different locations. Bob is 77 years old and Allen is 80 years old. Bob has a tremor in his hands and does not like typing on a tiny mobile phone, keyboard or any other mobile device. Allen on the other has limited vision and doesn't like typing either. Bob and Allen are currently in their respective living rooms, watching television. They are following a soccer game on the television. They like to discuss and comment on the game progress on a social-networking site. As neither of them is comfortable using the text-based mode to send and receive messages, they may prefer talking instead of typing. The television viewers can use their mobile phones to communicate with the social-networking web site using the speech to text software application on the smart phone.

3.7 Limitations

There are two main assumptions in our proposed system. First, in a real world setting, an IPTV viewer's interaction will be limited to the IPTV interface and mobile phone interface only. IPTV viewers are not expected to have a direct interaction with any software on the computer, as it will add an extra cognitive load on them. Second, in a real world setting, the IPTV viewers in a household will be limited to the inmates of the house.

If an IPTV viewer needs to request Twitter services on the IPTV interface, viewer has three options: First, the IPTV viewer has to enter their login credentials on the IPTV interface by means of the IPTV remote control. Second, the IPTV viewer's login credentials are already stored somewhere in the IPTV platform and the viewer can just select his Twitter profile on the IPTV interface by means of simple navigation on IPTV remote control. Third, the Twitter login credentials used on the smart phone to send the converted speech-to-text to Twitter website can be used by IPTV, to get the Twitter services on IPTV.

1. The limitation of using the first option is that the Mediaroom simulator's remote control does not have the most intuitive and user-friendly interface, which is a common problem with most remote controls today. It has a complicated layout and entering the information (username, password) into the IPTV interface by using the remote control is a laborious task.

Moreover, our primary focus of the work is to make the interaction simple and convenient to benefit targeted user group (elderly, low vision and limited manual dexterity users). Using a remote control to enter the IPTV viewer's Twitter account information on the IPTV interface contradicts with our main goal of providing convenience to our user group of interest. So, using remote control to enter text based information on IPTV interface is not a viable option.

2. The second option of storing the Twitter login credentials on the IPTV platform will be a viable solution because IPTV viewers have to make simple navigation using a combination of arrow keys and a select button on the IPTV remote control. However, this option is not without limitations. In order for the IPTV provider to store Twitter login credentials of the IPTV viewer, the IPTV service provider needs to know the Twitter account details of the IPTV viewer in advance. The IPTV service provider will typically maintain a database for information required for TV viewer's authentication on Twitter in IPTV platform. Whenever a new IPTV viewer request an addition of Twitter services with his/her Twitter account details, then the entry will be made into the database. This process will also involve processing time by the IPTV service provider, which in turn restricts creating the IPTV viewer's Twitter profile on the IPTV interface immediately. In other words, it can be said that the IPTV service provider needs to know in advance the Twitter profile information of the IPTV viewer to provide the Twitter service on the IPTV interface of the household.

3. In the third option, the twitter login credentials of the user on the mobile phone can also be used for providing the Twitter service for an IPTV viewer. However, the limitation with this method is that in a household many inhabitants own many mobile phones, while IPTV is just assumed to be one. If multiple individuals of the household are using Twitter at the same time, then the question of whose Twitter profile has to be displayed on the IPTV screen arises.

We have chosen option 2 in the above discussion for implementation of our proposed work. Due to the above limitations the proposed system currently works with two sample user accounts only. The Twitter Application Server currently uses the Oauth to fetch the user's message updates. In the Oauth method, the values of the access token and token secret are used to authenticate the user/viewer in the application and get his/her Twitter status updates. The values required for user authentication are already known for these two sample user accounts and they are hard-coded in the Twitter Application Server eliminating the need to log into the user's Twitter account through the computer. Our proposed system is just a prototype to show the possible integration of different technologies to provide convenience. It does not support the full functionality of real world IPTV services.

CHAPTER 4
ARCHITECTURE AND IMPLEMENTATION

The main objective of the proposed middleware is to integrate pre-existing technologies such as IPTV, mobile phones and social networking to provide the convenience to all the users, by introducing speech-to-text technology. This chapter focuses on the architecture and implementation details of the proposed architecture which is illustrated in Figure 4.1. The proposed architecture consists of three components: the mobile client, the Twitter application server (an application to fetch tweets and web server) and IPTV platform (in our case IPTV simulator and IPTV web server). In the following sections, we will describe each of the components in detail.

4.1 Mobile Client

The mobile client (Figure 4.2) is a Blackberry Torch Smartphone. The BlackBerry Torch 9800 is a 2010 model. It has a physical QWERTY keyboard and a sliding multi-touch screen display. The Blackberry Torch runs on BlackBerry OS 6, and has features for greater social networking integration and powerful universal search. The third party application Vlingo [35] is a voice to text conversion software, which can be installed on Blackberry Torch 9800. Vlingo software can also be installed on many other modern smart phones. The Vlingo virtual assistant converts the user’s words into actions. The Vlingo provides services like searching, texting, composing emails and accessing navigation facilities by using the speech technology. It has facilities to search restaurants in a geographic area by speech technology. In addition, it has an ability to connect and post updates or messages to social networking websites such as, Facebook and Twitter.

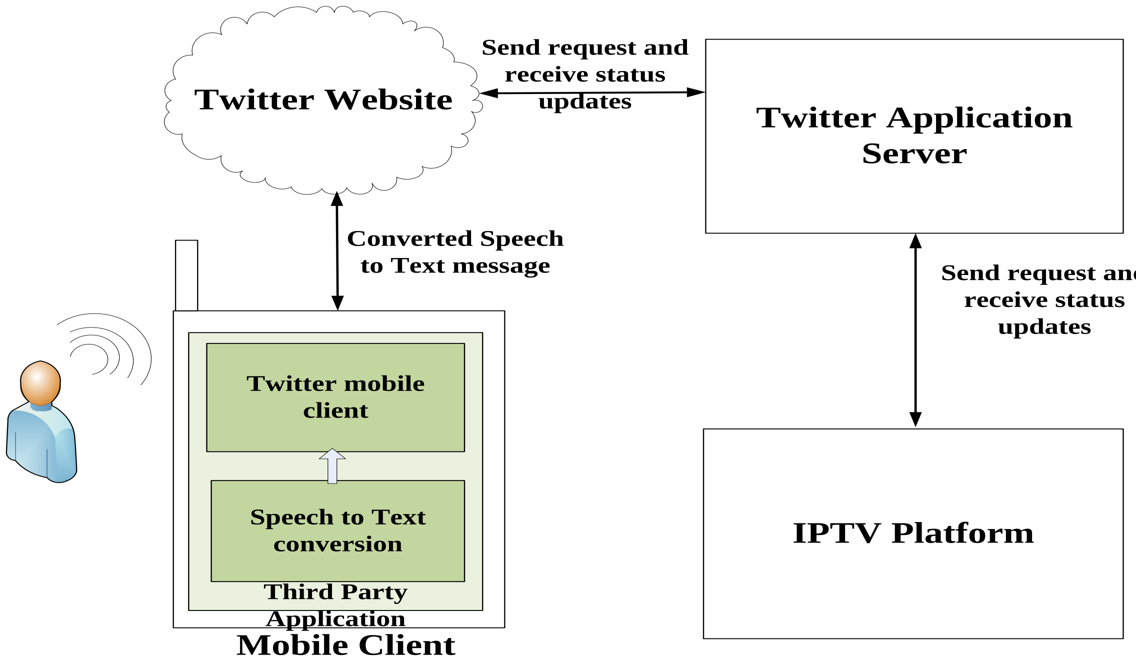

Figure 4.1 Overall architecture

Figure 4.2 The Mobile Client Interface

In the architecture, the mobile client (Blackberry phone) uses the third party application Vlingo to connect and post messages on the Twitter website, with the help of speech technology. The Vlingo contains the necessary Twitter client interface to perform the operation.

During the initial installation of the Vlingo client on the Smartphone, the Twitter login credentials of the user have to be entered, as illustrated in Figure 4.3. The login credentials can be changed at a later time, if needed. The Vlingo application on the mobile client is client software, which communicates with the Vlingo server over the Internet, where the speech to text conversion is performed.

Figure 4.3 Twitter Login Credentials on Vlingo 3rd party application

The left-side convenience key (Figure 4.4) on the Smartphone is used to start the Vlingo client by pressing it. The Vlingo client listens (Figure 4.5) to the user's spoken sentence. The first word of the sentence should be "Twitter", followed by the sentence the user intends to tweet. The first word "Twitter" is to inform Vlingo that the message is supposed to be sent to Twitter website. The left-side convenience key is pressed while the complete sentence is spoken.

The user should speak the words clearly and distinctly. Once the sentence is completed, the left-side convenience key is released by the user.

Figure 4.4 Left-side convenience key on Blackberry

Figure 4.5 Vlingo client - Listening to the user

The Vlingo client then shows a “thinking” icon (Figure 4.6) on the mobile client. During this phase, the Vlingo client communicates with the Vlingo server over the Internet, to make an appropriate text prediction for the spoken sentence. Once the text prediction is made the text message is displayed to the user on the Twitter mobile client interface (Figure 4.7).

Figure 4.6 Vlingo client - Thinking icon shows connecting to Vlingo server

If the user is unsatisfied with the converted text, then the whole process is repeated all over again. If the user is satisfied with the converted text, then the user can choose to send it to the Twitter website by clicking "Paste" button (Figure 4.7). Pressing the "Paste" button on the Twitter client interface sends the message to the Twitter website. The features on the Blackberry phone allow the users to choose different fonts and font sizes. For the users with low vision, the font size can be enlarged as per the user's requirement, so that the user can read the converted text with ease by scrolling down. Depending on the accuracy of the speech to text conversion software and user's clarity of speech, the process of dictating can be repeated. The user can also make minor corrections of the converted text by typing, which saves time and effort compared to typing the whole sentence. The editing capabilities can also accommodate minor corrections and handle the barriers related to differences in spoken and written language. For example, the speech recognition software cannot handle all the punctuations marks and symbols like @, and carriage return by using speech only. Though, Vlingo software supports some punctuations (?, ! *, -, +, =), it does not support all the punctuations and symbols. The editing can be used to insert the punctuations and symbols to increase the accuracy.

Figure 4.7 Vlingo client - Twitter screen

It is important to note that the mobile client uses the Vlingo functionality as provided by the vendor, i.e. there was no development of software or design contribution that the thesis makes to the area of the speech to text conversion. The technical contribution is in the development of the Twitter Application server and the overall idea of integrating these technologies. Refer to Appendix B – I – Figure B1.

4.2 Twitter Application Server

The Twitter Application Server illustrated in Figure 4.8, maintains the connection between the Twitter website and the IPTV platform. The Twitter Application Server, as illustrated in in Figure 4.9 has two main components. It has a Tweet fetcher application and web server application (IIS). The Tweet fetcher application fetches the user tweets from the Twitter website. The web server application is the internet information services (IIS). IIS is a Microsoft web server, which is an integral part of the Windows server family of products, offering HTTP capabilities in Windows environment. The two components will be discussed in detail in section 4.2.1 and 4.2.2.

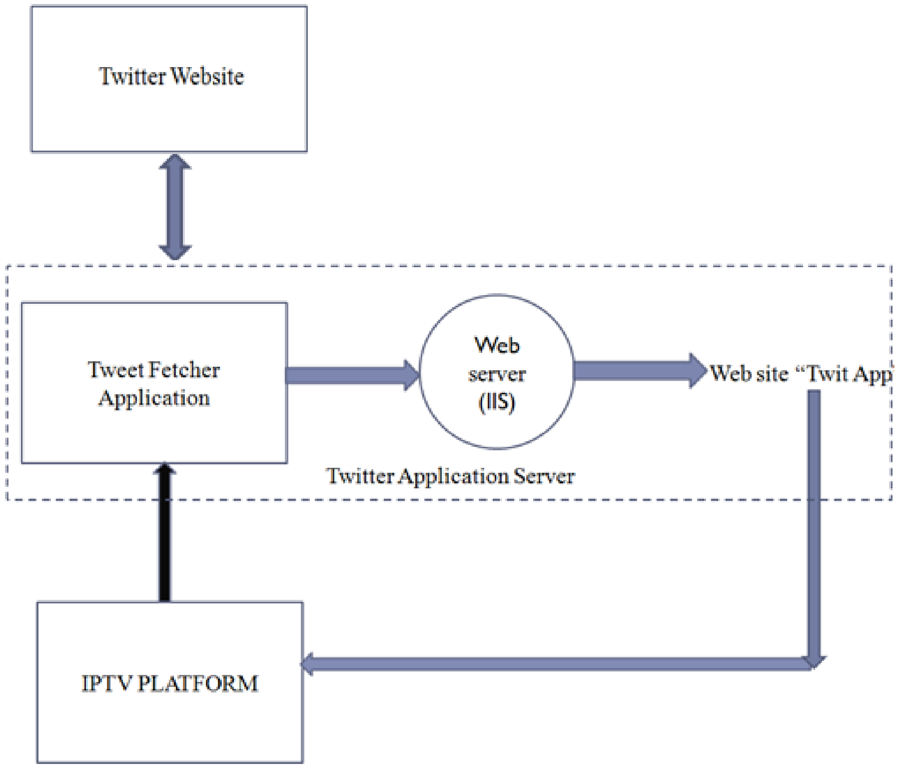

Figure 4.8 Twitter Application Server connects Twitter website and IPTV Platform

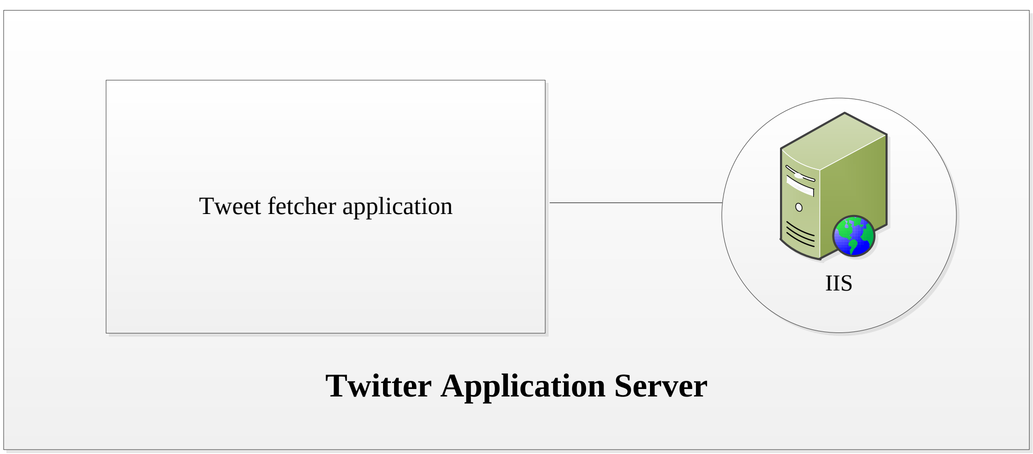

Figure 4.9 Twitter Application Server

4.2.1 Tweet fetcher application (with ASP.NET)

A web application was developed, using the C# programming language and ASP.NET. The web application was developed with approximately 80 lines of C# code. In order to develop web applications that use the Twitter API to retrieve and/or post messages, it is necessary to register an application (as a Twitter application developer), and obtain consumer key, consumer secret, access token, and access secret token. Registration is done to get consumer key, consumer secret, access token and access secret token (Appendix D – I)., which is later used in the application development (Appendix D – II), to authenticate the user to Twitter, as described in section (Appendix D – III),. The application performs the following steps.

- It performs the user authentication process, described in (Appendix D – III), with the help of a Web.config file (Appendix D – II – Figure D.1)which has the values for consumer key, consumer secret, access token value and access token secret.
- It gets the user-timeline and friends-timeline information in XML format using Tweetsharp wrapper, as described in (Appendix D – III).
- The application parses the returned XML file and re-builds a new XML file in a concise format with only desired values, as described in section (Appendix D – IV).
- The new re-built XML content is published to the website using IIS, as described in section 4.2.2.

4.2.2 Web Server (IIS)

The IIS (Microsoft internet information server) is used as a web server on the Twitter Application Server. The IIS server has two main functions. First, it receives the status updates for name of the user whose tweets are requested from the IPTV simulator index page. The username is then used as the string “s” in the Tweets fetcher application (Appendix D – III) to get the status messages of the requesting user. Secondly, after the XML content is created using the string “sample1” (Appendix D – IV), it publishes the website “TwitApp” (refer to Figure 4.10).

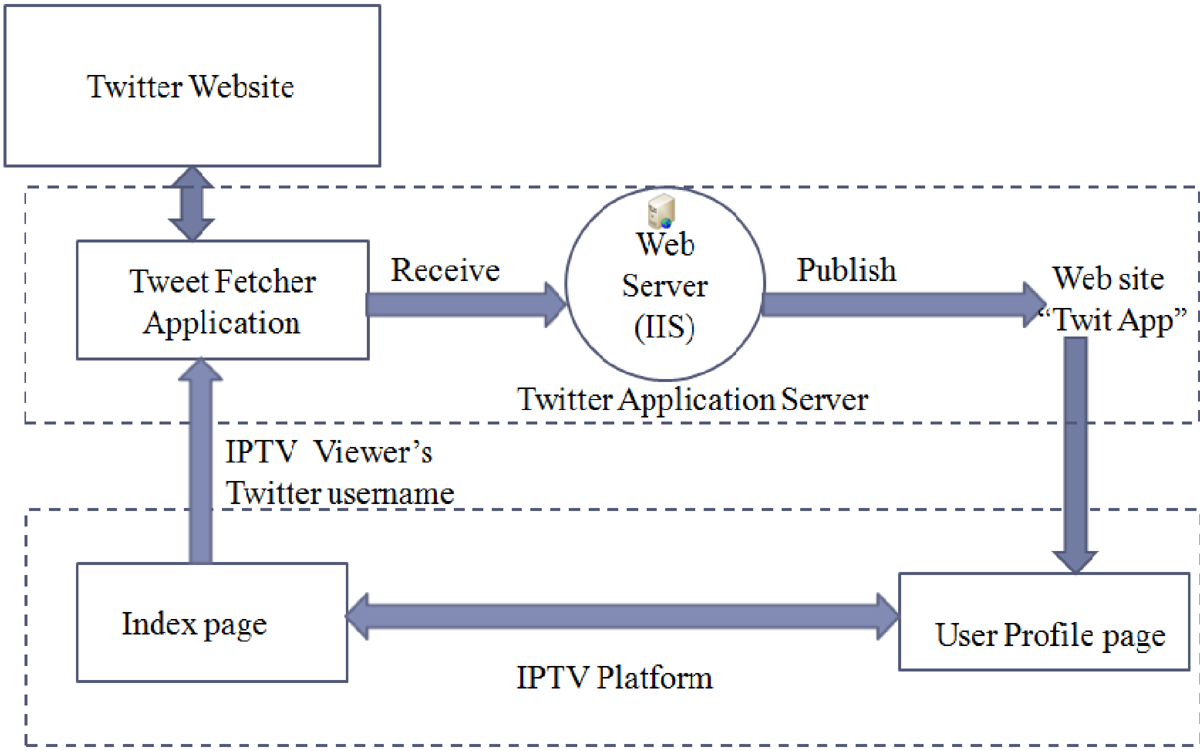

Figure 4.10 The function of Web server (IIS) in Twitter Application Server

As Microsoft Visual Studio provides many options for working with IIS for developing web sites and web applications, IIS was considered a reasonable option. In addition, Visual Studio provides the necessary upgrades and releases to support IIS. The IIS website is created using the IIS manager. After configuring IIS, the build\publish option is used to publish the content of the web application to an IIS website (TwitApp). The IIS website is now accessible to the IPTV simulator on another machine connected through the Internet. The IPTV simulator is responsible for displaying website content on simulator screen using MRML code. Refer to Appendix B – I – Figure B2.

4.3 IPTV platform

To design, develop, and simulate the IPTV experience, the Microsoft Mediaroom IPTV simulator was used. Microsoft Mediaroom uses Media Room Markup Language (MRML) code to display the XML content in addition to the video stream. The IPTV related application development was done with the help of my colleague Hashim Al-adhami as a class project. Hashim Al-adhami's contribution to this project included designing the page's layouts for the IPTV, and coding efforts in the Mediaroom. This part of the project was later integrated together with the middleware already developed, to accomplish the initial prototype. About 100 lines of MRML code was developed by Hashim, on the Microsoft Mediaroom IPTV platform. The IPTV as described in Figure 4.11 has two main components namely IPTV simulator and IPTV web server.

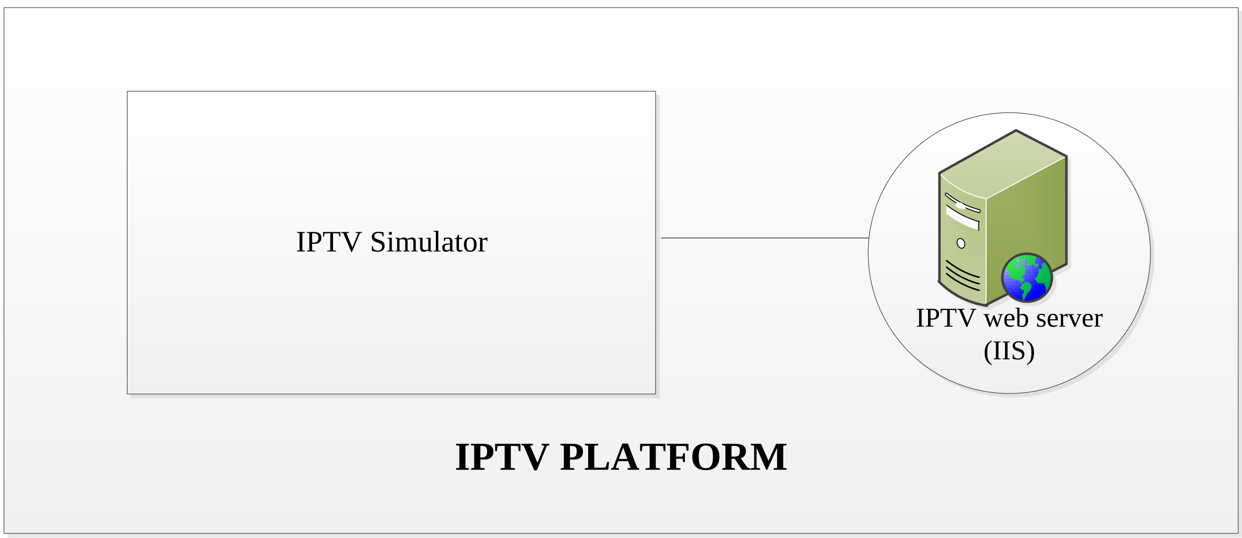

Figure 4.11 IPTV Platform

4.3.1 IPTV Web server (IIS)

The IPTV simulator content has video stream and associated XML data. The IIS server in the IPTV Platform is responsible for page loads. Initially, when the IPTV simulator starts, the IPTV web server (IIS server) loads the index page using MRML code. The IPTV Web server (IIS server) gets the URL, username and layer information and assigns the appropriate MRML code necessary for page load of the specific user with all the components in the IPTV simulator.

The IIS server also loads back the Index page with URL and layer information on user request. The Figure 4.12 illustrates the function of IIS server in IPTV Platform.

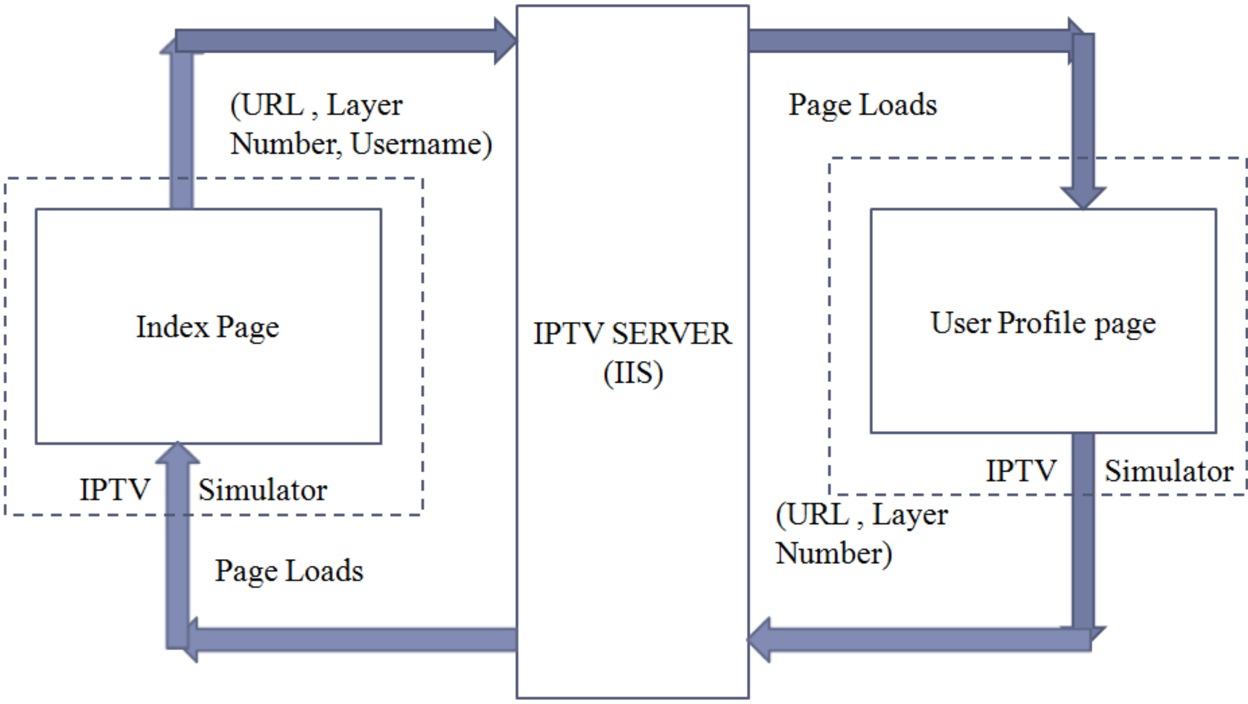

Figure 4.12The function of Web server (IIS) in IPTV Platform

Figure 4.13 shows the design of the Index page and Figure D.6 (in Appendix D – V) shows the MRML code for the Index page design based on attributes such as button position, button style and button text. The Figure D.6 (in Appendix D – V) also describes, the submit actions (SubmitAction0 and SubmitAction1) to be performed on the button click events. SubmitAction0 and SubmitAction1 are two actions which are used to select the user profile on Index page.

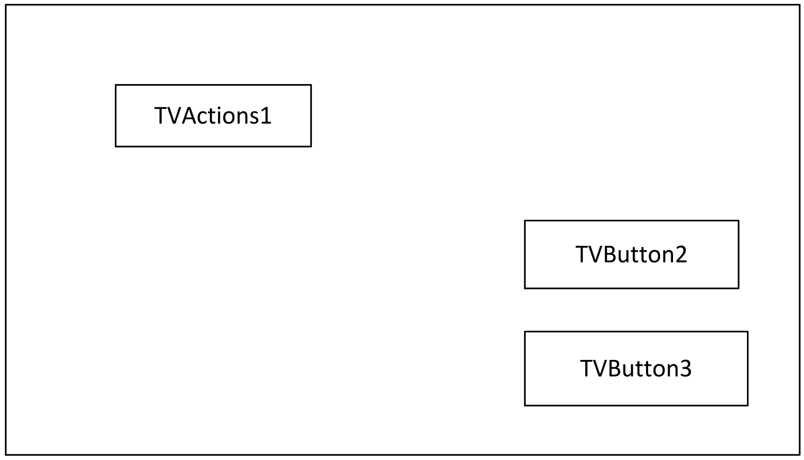

Figure 4.13 Index page design

In the index page (Figure 4.13) there are two TV buttons, namely “TVButton2” and “TVButton3”, which store the user name “User 1” (Afshana) and “User 2” (Hashim) respectively.

4.3.2 IPTV Simulator

The IPTV Simulator receives the video stream, and the xml data from the xml data source. The IPTV simulator then overlays the xml data on the video stream, and display it on IPTV simulator screen. Figure 4.14 illustrates the functions of IPTV simulator.

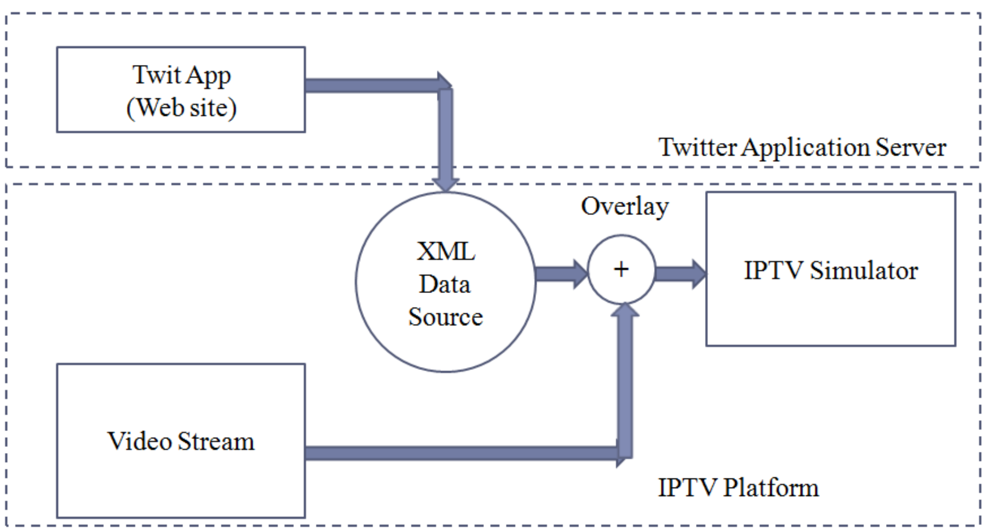

Figure 4.14 The function of IPTV simulator in IPTV Platform

The IPTV simulator displays two designed pages, namely the Index page and the User Profile page. The Index page is the first page which loads in the IPTV simulator. Refer to Appendix B – I – Figure B3.

The Index page has the profile information of the users who presumably live in the same household. The IPTV viewer can choose the user profile whose tweets are to be displayed on the IPTV simulator. Once a user profile is selected on the Index page then the following steps are carried out. First, the user's status request is sent with a user name to the Twitter Application server for further processing, to which the tweets are received in XML format and sent to the IPTV simulator's User Profile page for display. Secondly, the URL and layer number information required for User Profile page is sent to the IPTV web server.

After the button (TVButton1 or TVButton2) click in the Index page, the Mediaroompage.aspx, responsible for the display of User Profile page is initiated. The Mediaroompage.aspx has the MRML code for the display of the User Profile page. In addition, it uses the IIS website "TwitApp" (section 4.2.2), published by Twitter Application server as XML data source to display the appropriate information, such as status message, user name, picture URL, date-time and source, in the component objects on the User Profile page. Figure 4.15 illustrates the user profile page design.

In Figure 4.15, clicking the "Back" button takes the user to Index page. In Figure 4.15, the "TV Text" object holds the username of the current user profile page. Refer to Appendix B – I – Figure B4.

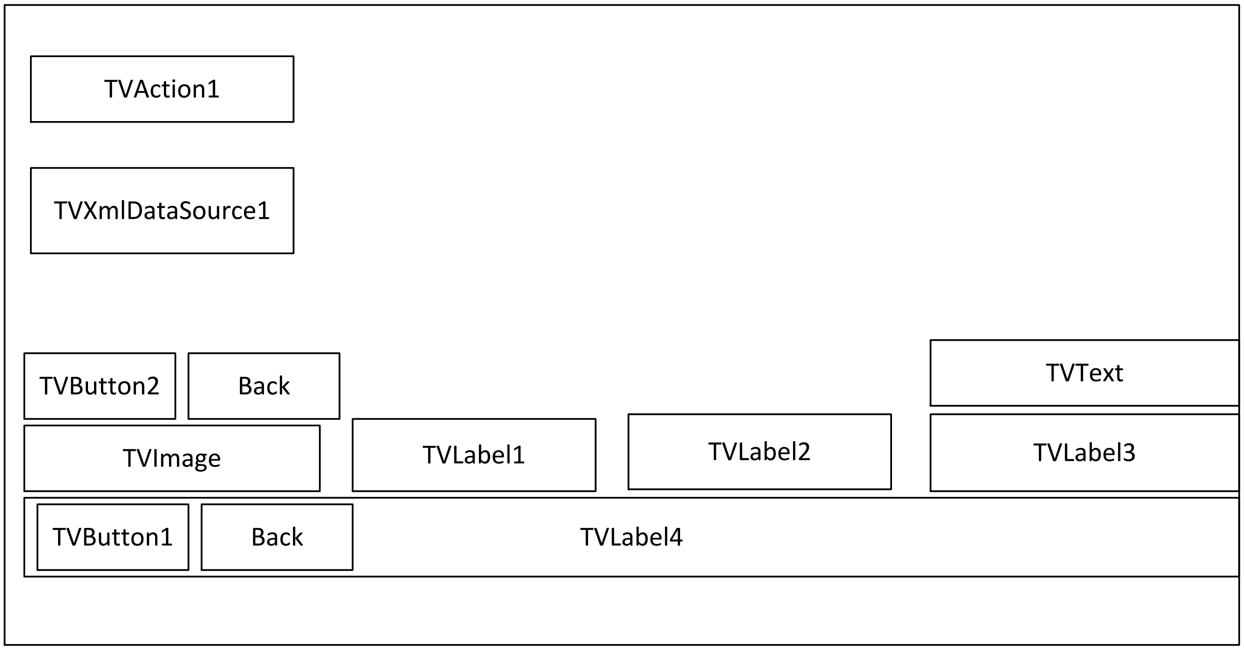

Figure 4.15 User Profile page design

The Figure D.7 (in Appendix D – VI) illustrates the MRML code for the bound list object in the User Profile page. Each element in the above bound list is contained by another TV list item template object "TvListItemTemplate". The "TvListItemTemplate" holds the information such as: status message, username; picture URL, date_time and source information. Figure D.8 (in Appendix D – VII) shows the code snippet for the TVListItemTemplate and sub-elements of TVListItemTemplate. The website "TwitApp" from the Twitter Application Server, is used as the XML data source in the Mediaroom IPTV to display the user status on User Profile page, which is described by code snippet of MRML code in Figure D.9 (in Appendix D – VIII). The content will be displayed in the User Profile page of the IPTV simulator.

The Figure D.10 (in Appendix D – IX), illustrates the code snippet about TV actions and composite actions in User Profile page.

The C# code associated with "TVText" functionality of Figure 4.15 is described in Figure 4.16. The button "TVText1" stores the username of the current User Profile page.

```
if (Request.QueryString.HasKeys())
        {
            if (Request.QueryString["TVButton3"] != null)
                //Set this url by default in case of not a postback
                TVText1.Text = Request.QueryString["TVButton3"];
            if(Request.QueryString["TVButton2"] != null)
                TVText1.Text = Request.QueryString["TVButton2"];
        }
```

Figure 4.16 Code snippet for TVText1 functionality

4.4 Design Considerations

Initially, the User Profile page was designed in four different layouts, which included transparent horizontal layout, opaque horizontal layout, transparent vertical layout, opaque vertical layout. The opaque horizontal layout and opaque vertical layout were later discarded, as they completely obstructed viewing the video content. The other two layouts, which included transparent vertical layout (Figure 4.17 and Figure 4.18) and transparent horizontal layout (Figure 4.19 and Figure 4.20), were later evaluated in a pilot study. The transparent vertical layout status updates scrolled top to bottom and vice versa. In this layout the video viewing was slightly obstructed and two status updates were shown on the screen at any given moment. The transparent horizontal layout status updates scrolled left to right and vice versa. In this layout the video viewing had no obstruction and only one status update was shown on the screen at any given moment, so the users were able to read the status updates with ease.

4.4.1 Initial transparent vertical layout design

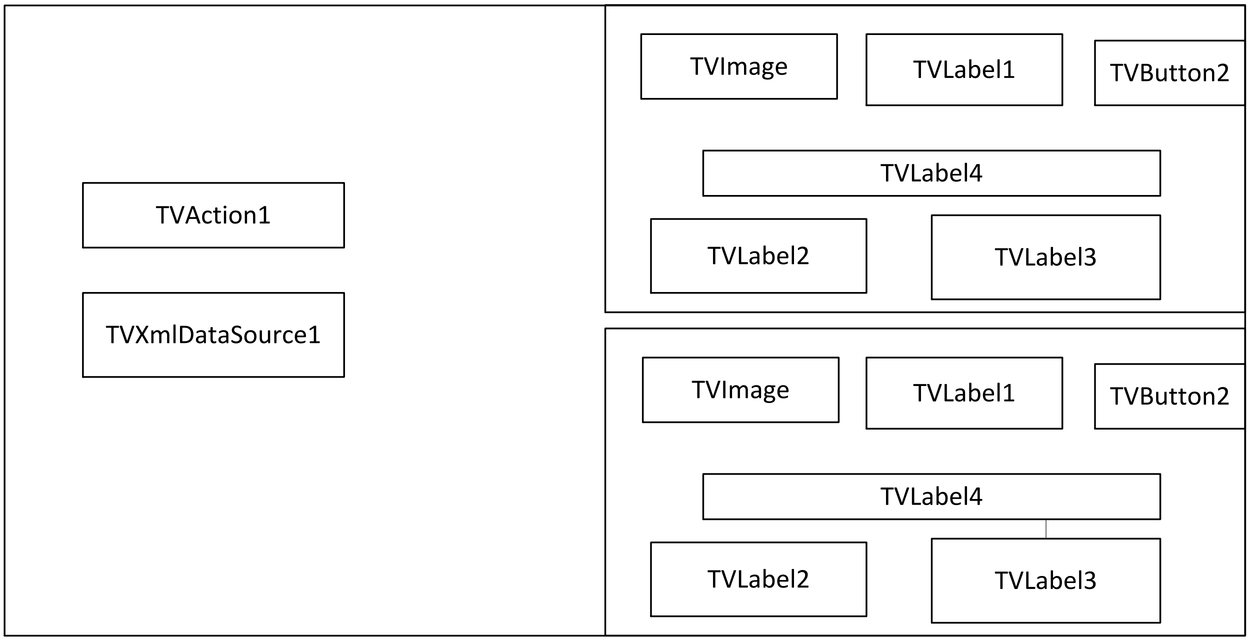

Figure 4.17 Design of transparent vertical layout of User Profile page

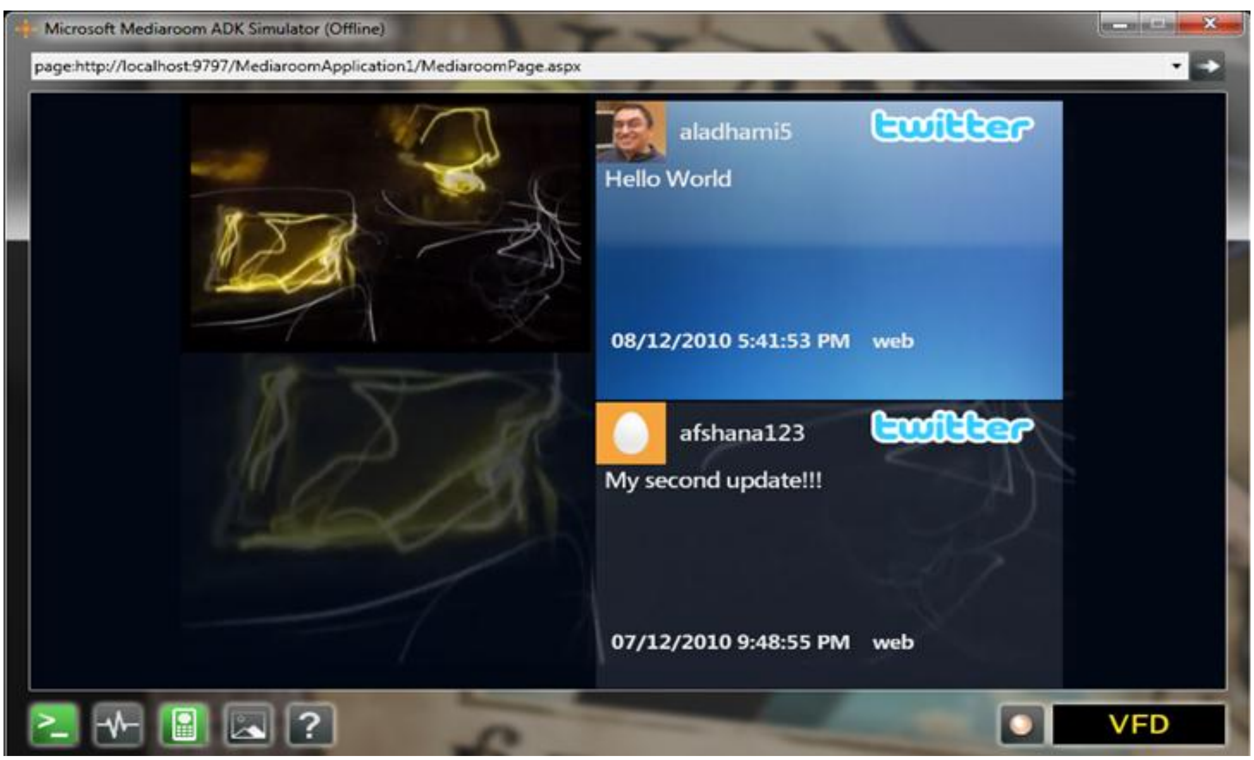

Figure 4.18 Screenshot of transparent vertical layout on IPTV simulator.

4.4.2 Initial transparent horizontal layout design

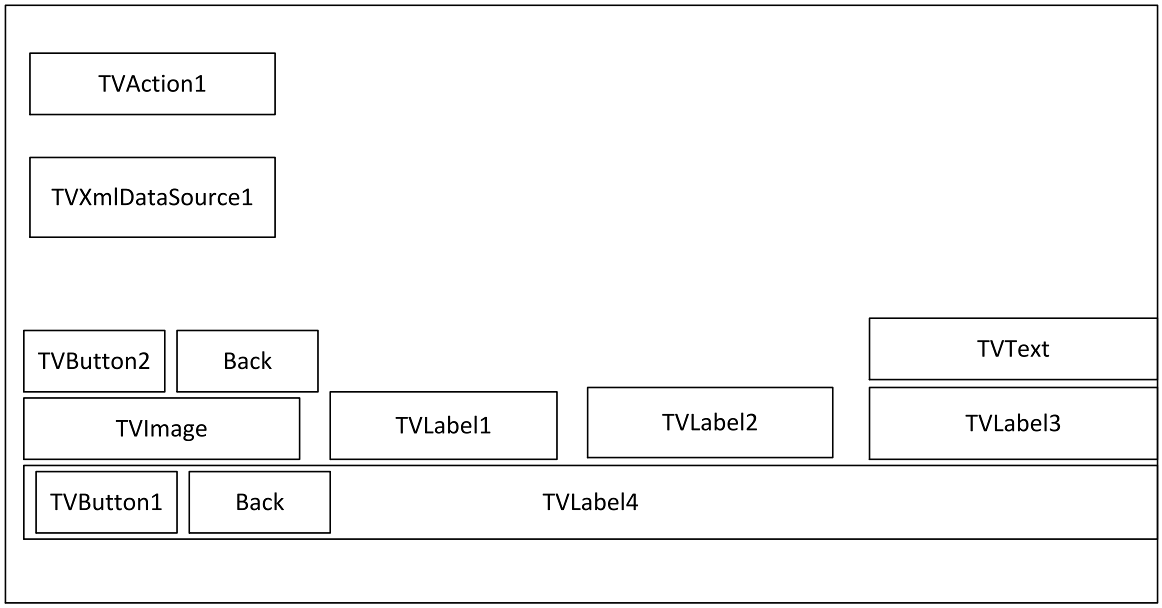

Figure 4.19 Design of transparent horizontal layout of User Profile page

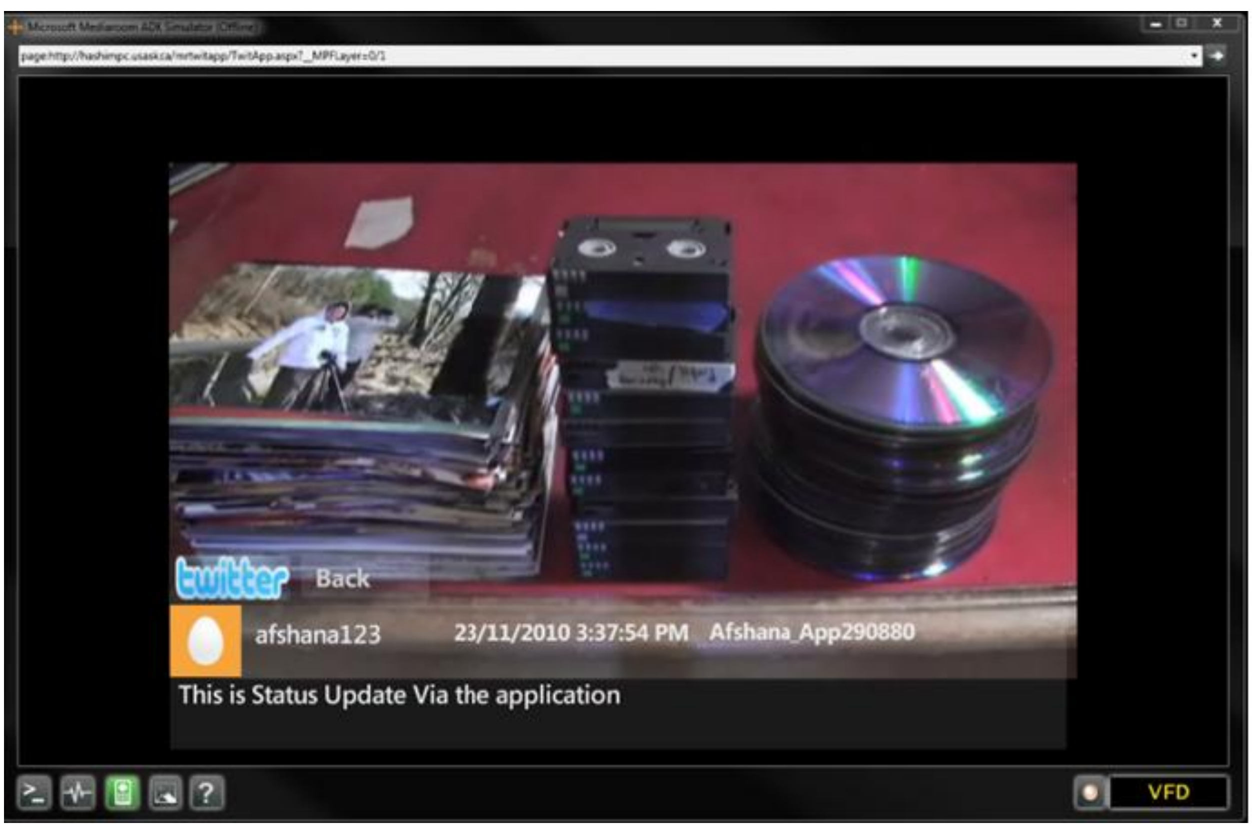

Figure 4.20 Screenshot of transparent horizontal layout on IPTV simulator.

CHAPTER 5
PILOT STUDY

This chapter focuses on a pilot study conducted with a group of regular university students in MADMUC lab, to evaluate the feasibility of the system and to get feedback for improvement of the prototype. This chapter discusses in detail the results of the pilot study. In the next chapter, we discuss the evaluation of the system prototype with students and university staff members, who mimic the low-vision and limited manual dexterity impediments of the elderly

The main question tackled in this pilot study is: the feasibility of the architecture, the appropriateness of the different layouts, the viability of speech-to-text as input method, the subject's acceptance of the speech-to-text experience.

5.1 Pilot Study

A pilot study of the system was conducted with six students from the MADMUC lab. Different parameters were measured, including system responsiveness and accuracy of the speech to text conversion. The accuracy of the speech to text conversion was measured based on different accents, sentence length, speaking rate, voice and pitch.

Initially, the pilot study was conducted using the Blackberry simulator, which was later replaced with a real Blackberry phone, to create a more natural setting for evaluation. Microsoft Mediaroom simulator was used to simulate the IPTV experience. A Vlingo client was installed on the Blackberry simulator. In addition, the Vingo client provides a Twitter client interface to send the converted text messages to the Twitter web site.

During the design and implementation phase, the system layouts were designed in four different ways to analyze the user interface preferences, which included transparent horizontal layout, opaque horizontal layout, transparent vertical layout, and opaque vertical layout. The layouts were designed with a transparent and opaque look. Layouts were also designed with a horizontal and a vertical message display styles. The horizontal layout displayed one tweet per screen, which scrolled left to right. The vertical layout displayed two tweets per screen, which scrolled top to bottom.

The pilot study was conducted in a controlled environment and was limited to two user accounts. The participants used the Twitter accounts associated with these two user accounts, and they were not able to use their own Twitter accounts. The reason for this restriction was the need to set up the Twitter accounts of the participants in advance on the IPTV web server, as we were unaware of the identity of the participants in the pilot study. So, we created two sample user accounts for testing and evaluating the system.

In the pilot study, several parameters characterizing the user experience were evaluated. First, the user's perception of the turn-around time was evaluated on the scale of quick, slow and medium. The turnaround time was also measured as the duration between the time the user speaks on the Blackberry simulator, and the time the Twitter message appears on the IPTV simulator screen. Second, the speech to text conversion of the Vlingo third-party software was evaluated with respect to different accents. The six participants had a variety of different accents including Chinese, East-Indian, South-American, and Nigerian. Evaluating the speech to text conversion accuracy of the Vlingo third party application is needed to evaluate the overall performance of the proposed system and user satisfaction. In the pilot study, the users were asked to fill a questionnaire suggesting the possible improvements, and their perceived advantages or disadvantages. In addition, the users were asked to rate the system performance using 5-point Likert scale. Likert scale is commonly used rating scale in research questionnaires.

Likert scale has values ranging from strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree.

It is assumed that the user is already logged into one of the two available user Twitter accounts of the Vlingo application residing on the Blackberry simulator. The user accounts on the Blackberry simulator are same as the user accounts used in the IPTV simulator. Initially, during the installation of the Vlingo client on the Blackberry simulator, the user is prompted to assign a left-side convenience key, which is pressed by the user to record the speech. During the evaluation, it is assumed that the left-side convenience key has been already assigned. When the user is ready to participate in the pilot study, basic instructions and guidelines were given to carry on the pilot study. The user is instructed to keep holding the left-side convenience key until the end of the spoken sentence and then release it. The instruction to the user had a procedure involving many steps.

The first step involved pressing the left-side convenience key on the Blackberry simulator. As soon as the left-side convenience key was pressed, the Vlingo client software is activated, and a “listening” icon is displayed on the Blackberry display, to inform the user that the software is ready to listen to the user speech.

In the second step, the user spoke the sentences or word, which user wanted to post on Twitter website, without releasing the left-side convenience key. The user was instructed to prefix the sentence with the word “Twitter”, to let the Vlingo client know the message was intended to be sent to the Twitter website. The users were reminded to hold onto the left-side convenience key until the sentence ends completely. The user had to speak the sentence in a clear and distinct voice.

In the third step, the user released the left-side convenience key. At this stage, the Vlingo client software residing on the Blackberry simulator communicated with the Vlingo server in the cloud, to establish a connection and make a best text prediction of the spoken sentence. The predicted text for the spoken sentence is sent back to the Vlingo client, to be displayed on the Blackberry simulator screen.

In the fourth step, the user could see the converted text message on the Twitter interface of the Vlingo client, with options to paste it (to send it to Twitter) or cancel it. If satisfied, user pressed on the paste-button. If user is not satisfied, then user could cancel and start all over again. The detailed pilot study scenario is outlined in Appendix A- I.

The IPTV simulator worked independently from the procedure followed on the mobile client. If the user wanted to see the message on the IPTV, then the user could go to his/her chosen user-profile on the IPTV simulator. The message sent to the Twitter website should be available for display on the IPTV simulator. After the pilot study, the participants were asked to fill in a questionnaire based on their suggestions for improvements, the perceived advantages or disadvantages of the system, and their satisfaction with the overall performance of the system.

The questionnaires from all the participants were collected, and the feedback from each participant was analyzed. The detailed pilot study questionnaire is outlined in Appendix A- II. Below is a summary of the feedback from each participant.

Participant 1 had a Chinese accent. The participant felt the user interface was convenient. The suggestions included improvement to the scalability to accommodate more users, and to add a better notification when the converted text appears on the Vlingo client. The participant also suggested using another speech to text conversion software instead of Vlingo, and generally comparing and evaluating the best option available to integrate into the proposed system.

Participant 1 was not satisfied with the speech to text conversion of speech with Chinese accent. However, the participant was satisfied with the turn-around time, which was about five seconds. The participant also showed a preference towards the transparent horizontal layout. In terms of overall satisfaction with the system performance, the user rated it as good on a 5-point Likert scale. The 5-point Likert scale ranged from bad, fair, neutral, good, excellent.

Participant 2 had an East Indian accent. The participant was impressed with the functionality, and felt everyone including the people with limited manual dexterity and vision can benefit from it, as speech is the natural way to communicate. The participant suggested using real Blackberry phone, as a Blackberry simulator was used for pilot study instead of the Blackberry phone. However, the participant was quite satisfied with the general turn-around time which was about five seconds. The participant also showed a preference towards the transparent horizontal layout. In terms of overall satisfaction with the system performance, the user rated it as good.

Participant 3 was from Latin America and had an American accent. The participant suggested using another speech to text software for comparison and felt the system needs improvement in overall performance. The participant was satisfied with the user-interface. The participant was satisfied with the general turn-around time of 5 seconds. The participant also showed a preference towards transparent horizontal layout. In terms of overall satisfaction with the system performance, the user rated it as excellent.

Participant 4 had an East-Indian accent. The participant was satisfied with the functionality and felt that people can benefit from it. The participant suggested using another speech to text software for comparison. The participant felt the speech-to-text software was not able to handle foreign names. However, the participant was satisfied with the general turn-around time of five seconds. The participant also showed a preference towards transparent and horizontal layout. In terms of overall satisfaction with the system performance, the user rated it as excellent.

Participant 5 had an East-Indian accent. The participant was satisfied with the functionality, but felt remote control on the IPTV simulator was not intuitive, which made it difficult to use. The participant suggested using another speech to text software for comparison. The participant felt that the speech-to-text software was good, with the exception of handling foreign names. However, the participant was satisfied with the general turn-around time of five seconds. The participant also showed a preference towards transparent horizontal layout. In terms of overall satisfaction with the system performance, the user rated it as excellent.

Participant 6 had a Nigerian accent. The participant was impressed with the functionality and the felt the system provided convenience. The participant suggested devising a constructive way to help people to adapt the technology. The participant felt that the speech-to-text software was good, with the exception of handling foreign names and long sentences. However, the participant was satisfied with the general turn-around time which was about five seconds. The participant also showed a preference towards transparent horizontal layout. In terms of overall satisfaction with the system performance, the user rated it as excellent. The detailed pilot study results are outlined in Appendix A- III.

5.1.1 Summary of the pilot study results

Most of the participants found the proposed system to be useful. The turn-around time was measured to be five seconds on average and participants were satisfied with it. Participants were not fully satisfied with the speech-to-text conversion of different accents, foreign names, and long sentences. The common suggestions included using a different speech to text application, increasing scalability to accommodate more users, adding a notification option with better messages and providing appropriate training. Participants preferred the transparent horizontal layout. The participants expressed enthusiasm and satisfaction in using the system. The participants preferred using the application on the real Blackberry phone, instead of the simulator. Participants felt the Mediaroom simulator's remote control is not intuitive enough, which made it difficult to use.

5.1.2 Conclusion

From the pilot study, we learnt that our system's accuracy of speech-to-text relied on the subject's accents to large extent. Our system handled some accents better than others. The proper diction of the participant was important for speech-to-text conversion. The system performed poorly on long sentences and foreign names. The users preferred certain layouts. Most users preferred the transparent horizontal layout. The preference of layout clearly showed that the subjects favored minimum obstruction while watching video. The subjects preferred using the real phone instead of the simulator. As discussed in the literature survey, we also observed most subjects struggled with the Mediaroom simulator's remote control and felt that the layout was complicated. The difficulty in using the remote control re-enforces the need to have simple interface. One of the participant also mentioned that this speech-to-text technology would be beneficial to all user groups including the elderly and disabled, who are targeted user groups for evaluating our system prototype in evaluation phase.

We also encountered questions relating to scalability of the system in terms of supporting maximum number of users, the system will be able to support. The issue of scalability can be addressed in future work and we will not be addressing the issue within the scope of our research work. Using different speech-to-text software instead of Vlingo was one of the main suggestions by subjects in pilot study. However, we have chosen to continue the use of Vlingo in evaluation phase due to the lower cost (it is free), and Vlingo's simple interface to work with the Twitter website. In the pilot study, we focused on the acceptance and satisfaction of user based on performance and user Interface. In the evaluation chapter, we will measure the exact time taken to carry out certain pre-scripted tasks, the error rates for each of the tasks and number of attempts to successfully perform a given task. The findings from the evaluation can give interesting insights about user's frustration level and error patterns based on users. It can also show which tasks takes take longer time in comparison.

CHAPTER 6
EVALUATION

This chapter focuses on the evaluation phase using the studies conducted with a group of participants to evaluate the feasibility of the system and to get feedback for improvement of the prototype. This chapter discusses in detail the results of the evaluation. The main question tackled in this research is: Can speech-to-text technology on mobile phones be integrated into social-networking and IPTV to provide convenience? Further there are other sub-questions to be answered which include:

(1) Will the technology benefit all user groups including elderly people, and people with limited vision and mobility?

(2) Is speech-to-text a better alternative to text input mode on a tablet or smart phone?

(3) Is there any difference in how the users interact with the system with respect to their disabilities or their skill sets?

(4) Is speech-to-text a viable method of interaction with social-networking and IPTV?

(5) Are there any particular challenges which users encounter while using the system?

We have adopted qualitative research methodology for the evaluation. The reason for adopting qualitative methodology is to get detailed and accurate conclusions from a small sample of participants in our study. Apart from detailed observations, video-taping, interview and questionnaire, we collected numerical data for the time taken to conduct a prescribed dialogue, the error rate and number of attempts involved. These were measures mainly used to observe the frustration levels and system performance for each user.

The evaluation is carried out as a comparative study, where the user experience with our system prototype (speech-to-text technology) is compared with their experience using text based mode of entry using a tablet and a mobile phone.

6.1 Experimental setup

The proposed system is focused on providing convenience to all the user groups, with particular emphasis on elderly people, people with limited vision and people with limited manual dexterity. Unfortunately, we did not have access to representatives of these user groups as subjects for the system evaluation. So we recruited university students and staff members as participants in the experiments, even though, the evaluation results may have been different had we chosen to use elderly participants based on their awareness of the speech technology, their willingness to use speech technologies and their individual frustration levels, using non-elderly participants to mimic the impairments can give some insight into the viability of the approach and the usability of user interface. Therefore, in the evaluation we simulated the difficulties encountered by these user groups by creating interaction obstacles for regular users. To simulate the user group of elderly people / people with limited manual dexterity, the participants were asked to wear rubber-bands around their fingers to limit their finger movement abilities. To simulate the user group of elderly people / with limited vision, the participants were asked to wear glasses covered with transparent packing tape to limit their vision. Each participant completed a set of predefined tasks in each of the three experimental conditions: speech-to-text, text on tablet and text on mobile.

6.2. Scenario

The following scenario was used: two users are watching the same video clip simultaneously on their own television sets (using IPTV service), in their respective living rooms, while having a conversation about the show on Twitter. Both the users are using the speech-to-text software (Vlingo) on their smart mobile phones to communicate with each other.

Each of them is using his or her user account on the Twitter website, and they mutually follow each other on Twitter.

The users exchange comments about the program they are watching. The comments contain sentences with different complexity, with respect to the words or names used. In the experimental setup, the IPTV experience is being achieved by using the MediaRoom IPTV simulator. To simulate two IPTV sets, the IPTV simulator was installed on two different machines, separated by some physical distance. Then a short TED video (8 minutes duration) was downloaded into the "Pseudo Channel" (renamed as "Channel1") folder of both the IPTV simulators. The downloaded channel was tested to see if the video works properly and simulates realistic TV experience. To support the user conversation about the show, two Blackberry Torch smart phones were used to communicate with the Twitter website. Off-the-shelf speech-to-text software Vlingo was installed on both Blackberry phones. In the real scenario, the users log into Twitter just once and their authentication is remembered by their smart phones, but as changes in the Twitter account are hard on the back end side, for the experiment two fictitious user Twitter login credentials were entered on the both the phones during the installation of the speech-to-text software on the smart-phone. These two Twitter accounts were used by all the subjects. During the pilot study, the participants were given the freedom to utter any sentence of their choice, however during the evaluation (user study) the participants were restricted to use pre-scripted messages. The purpose was to facilitate measuring and comparing the various aspects such as time taken, error rate and number of attempts with respect to the accent. The user conversation was pre-scripted and divided into three different phases, so that various parameters in the voice communication, such as accents, difficult words, foreign names and long sentences could be tested. Figure 6.1 is a snapshot of the pre-scripted user conversation (chat):

Phase 1:

User 1: Hello, How are you?

User 2: I am fine.

Phase 2:

User 2: Thanks for asking, USERNAME-1. What are you doing right now?

User 1: USERNAME-2, I am watching the TED TALK show.

User 2: What is it about?

User 1: It is about "How to build your creative confidence"

User 2: WOW!!! I am watching it too.

Phase 3

User 1: Interesting. I applaud talks like this...This is how great change is made...we are all creative creatures.

User 2: I Agree.

Figure 6.1 Phrases in the scenario

In the chat scenario, "User 1" is the name of the participant in role-1, and "User 2" is the name of the participant in role-2. Refer to Appendix C- I.

6.3 Roles

The study was conducted in three sessions with six different participants, with two participants during each session. In each session, the same scenario was run three times under different conditions. In each condition both participants used a different mode of input to communicate according to the scenario. The different conditions were: (i) text mode on tablet or computer; (ii) text mode on a mobile phone; and (iii) speech-to-text mode on a mobile phone (our proposed approach).

The participants took up two roles; role-1 and role-2 that were linked to the utterances communicated respectively by User 1 and User 2 in the scenario and involved different simulated communication impediments. In the scenario User 1 was assumed to have a manual dexterity deficiency simulated by wearing a rubber band. User 2 was assumed to have vision deficiency, simulated by wearing glasses with masking tape.

To ensure that all participants did the scenarios with different impediments, participants were required to switch their roles when repeating the scenario in each of the different conditions. For example, participant A takes the role 1 for conditions (i) and (iii) and role 2 for condition (ii), while participant B takes role 2 in conditions (i) and (iii) and role 1 for condition (ii). This is illustrated in Figure 6.2

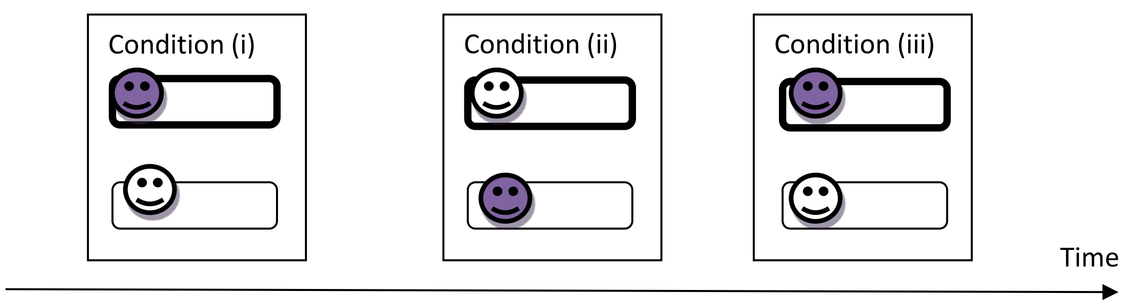

Figure 6.2 Swapping of roles by the two participants during one session

6.4 Measures

As the user experience clearly depends on the quality of the speech-to-text conversion provided by the 3rd party tool Vlingo, it was important to measure the user experience in relation to the performance of the software, which depends on the complexity of speech elements involved. For this reason, as shown in Figure 6.1, the scenario was divided into three phases, which separate the conversation according to sentence complexity and complexity of words used.

Phase-1 involves simple sentences and words like “hello” and “fine”. In Phase-2, the script includes the participant’s name, which may include easy, but also difficult foreign names, which may be difficult for speech-to-text conversion. Phase-2 also allows involves specific acronyms, such as “TED”, which are confusing for speech-to-text software, due to the existence of similar sounding words. Phase-2 also includes difficult words like “creative confidence”. Phase-3 involves long sentences.

Each of the phases was further divided into the individual statements spoken by participants. For each of the sentence, the following three parameters were measured: (i) Turn-around time, (ii) Number of attempts, and (iii) Error-rate.

Turn-around time: This is the time taken for a successful attempt. In the text mode, i.e. Conditions (i) and (ii), the turn-around time is the time difference between the time, when the user starts typing the sentence, and the time when the sentence shows up on the Twitter website. In the speech-to-text system (condition (iii)), it is the time difference between the time when the user starts uttering the sentence, to the time when the sentence shows up on the IPTV simulator screen.

Number of attempts: In conditions (i) and (ii), the number of user attempts was one, because the error rate while using text mode is negligible. Users made no mistakes while typing a sentence in the scenario because they got visual feedback of every letter they typed on the screen, immediately after pressing the key, unlike the speech-to-text, where the whole sentence shows after the complete sentence is spoken. In the speech-to-text system, maximum four attempts were allowed due to the time constraint (held with 45 minutes) of the experiment, in order to facilitate the participant to get better speech-to-text conversion. It was left to the participant's choice to attempt for four times or lesser. If the participant felt frustrated, he/she had the choice to send the sentence with errors, and to attempt the next sentence on the scenario.

Error-rate: The error-rate in a speech-to-text system is equal to the total number of words converted inaccurately divided by the total number of words in that sentence. The error-rate in a textual system is equal to the total number of words typed inaccurately divided by total number of words in that sentence.

Other measurements included the number of participants preferring to use (i) the speech to text system (ii) the speech to text with editing (iii) the textual system on mobile phone and tablets.

6.5 Participants

There were six different participants in our study. All the six participants were non-elderly users who mimicked the visual and manual dexterity limitations of elderly people by using specially designed physical impediments. The *within-subject* method was used, were the same participants evaluated all the three conditions. The advantage of the *within-subject* method is that the evaluation can be done with a small user group. As the participants received payment for particular time duration, the study was conducted with only six participants. The user study with six participants gives some interesting insights regarding the research questions however the results may vary if the number of study participants were increased. Four participants were University students, who belonged to graduate and post-graduate degree levels. Two participants were University staff members. Age division: Of the six participants, three participants were below age 30, and three participants were above age 30. There were no participants in the age group above 65.

Background (non-CS or CS): Among the six participants, three participants were related to the computer science field, and three participants worked / studied outside of computer science field.

Typing skills: Among the six participants, two participants were students who were skilled in texting on the mobile phone, and two participants were students who were not skilled in texting. There was one staff member comfortable using text on the mobile phone and another staff member who was not comfortable texting on mobile phones. This information was important to find out, if people with good texting skills would rather use text, or would prefer to use speech-to-text system. This information can also help in finding out, if the people with poor texting skills could use speech-to-text system or rather use text.

Accent: It is important to find out how the speech-to-text performs on different accents. As speech to text software can perform poorly without sufficient training for people with strong accent, the experience with a system depending on speech-to-text software could be too frustrating for users with strong accent. So, participants were chosen from different cultural backgrounds, to see whether the accuracy of speech-to-text conversion of the off the shelf software Vlingo, has any impacts based on their experience with the system. In the six participants, one participant had a Nigerian accent, one participant had a South Asian accent, two participants had Chinese accent, and two participants had Canadian accent. It is important to note that one of the participants had a Canadian accent due to long exposure to North-American culture, though she hailed originally from a middle-Eastern cultural background. The demographics questions are presented in Appendix C – III.

6.6 Consent

The participants were briefed about the purpose of the study and given an informed consent form to sign. The consent form, along with the entire study was approved by the University of Saskatchewan Behavioral Ethics Committee with protocol. The users were able to proceed to the study only after accepting the terms and conditions on the consent form.

The participants were asked to read and sign the consent form. The participants were also informed that they would be paid a sum of $15 for their participation in the study. They were informed that they had a choice to walk away from study at any time and they will be video-taped during the study. The actual consent form is presented in Appendix C – II.

6.7 User study

The user study had three facets: (a) usability facet (b) observational facet (c) comparative facet.

6.7.1 Usability Facet

The usability facet aimed to find out if the interface design was easy to use and understandable. The users had to complete a set of pre-defined tasks.

The usability facet evaluates the system features such as satisfaction with the interface, the learn-ability of the interface, and the error handling capability with respect to different accents, pitches of voice and diction.

6.7.2 Observational Facet

The observational facet aims to evaluate how the IPTV simulator and mobile smart phone are used. For the observational facet, the think-aloud method is used. The participant is asked to speak aloud about what they feel and think while interacting with the system. The users are not being disturbed or interrupted during the study. All instructions are given before the study begins. The study was video-taped.

6.7.3 Comparative Facet

The comparative facet compares (a) Accessing social-networking using text on tablet (b) Accessing social-networking using text on mobile phone (c) Accessing social-networking using speech technology on our proposed system.

The comparative facet can be evaluated by two different methods, which include “Within-subject” method or “Between-subject” method.

In the *Within-subject method*, the same user participates in all the three comparison cases. The advantage of this method is that the same participant is used to conduct the comparison of cases, and the result of the comparison is not be influenced by other factors that depend on the subject, e.g. background, age, experience etc. The Within-subject method does not need many participants. The disadvantage in the Within-subject method is that the participant learns the system easily over the period of time of the experiment and this “learning effect” can influence the results.

In the *Between-subject method* different users participate in all the three comparison cases. The advantage of this method is that the participants cannot learn the system over-time, because each user participates in each case only one time. The disadvantages in the Between-subject method are that it needs a big pool of participants based on various criteria like age, sex, skill-set, etc, to eliminate the effects of these factors in the results.

For the comparison facet of our proposed system, we use the *Within-subject method* for comparison. Within-subject method is a more suitable method because our three cases of comparison are quite different from each other and the “learning effect” does not occur across them. The measures used in comparison facet are time taken for each task, error-rate and number of attempts. The questionnaires related to usability and comparative facet is presented in Appendix C – IV. The exit questionnaire is presented in Appendix C – V.

6.8 Procedure for experiment

The experiments took place over 3 days, with one session taking place each day with a pair of two different participants going through a pre-set scenario. The pairing of participants was done in the following way, as described in Table 6.1.

Table 6.1 Cultural Background

Days	Participant-1	Participant-2
1	Nigerian Accent	South-Asian Accent
2	Chinese Accent	Chinese Accent
3	Canadian Accent	Canadian Accent

After giving consent, the participants were asked to fill-up an entry questionnaire, which focused on collecting demographic information related to their age, education, cultural background, field of study/work and skills in typing/texting on mobile phones and computers.

After that, the participants were instructed on how the study would proceed and were given instructions on how to use the speech-to-text system. The participants were informed that they could ask questions or seek assistance at any point during the study.

The study was held in three different conditions. In each of the three sessions the same pair of participants went consecutively first through condition (i), then through (ii) and (iii) and swapped their roles according to the scheme shown in Figure 6.2.

In all the above conditions, there were assigned time-keepers / scribers for each participant. The time-keeper used stop-clock to measure the time taken in the format (minutes:

seconds: milliseconds). Apart from the time, the time-keeper/scribers also kept track of the number of attempts, and wrote down observations of particular problems, questions or behaviours of the participants.

The error rate calculation was done at the end of the experiment by comparing the pre-scripted chat messages to the messages showing up on the Twitter website.

6.9 Results

In this section, we will first discuss the results organized by the experimental conditions, and by session for each of the three sessions, i.e. participant 1 paired with participant 2, followed by participant 3 paired with participant 4, followed by participant 5 paired with participant 6. First the results of the condition representing the proposed system with speech-to-text conversion on mobile phone (condition iii) are presented, followed by condition (i) – using a tablet and text mode of entry, and finally, condition (ii) using a mobile phone and text mode of entry. For more details on results, refer to Appendix C – VI.

6.9.1 Results for Condition (iii) - Speech to Text

Table 6.2 Participants, accents, and roles

Participants		**Accent**	**User role in pre-scripted scenario**	**Obstruction tool**
Session 1	**Participant-1**	Nigerian Accent	User 1	Rubber-band
	Participant-2	South-Asian Accent	User 2	Glasses with masking tape
Session 2	**Participant-3**	Chinese Accent	User 1	Rubber-band
	Participant-4	Chinese Accent	User 2	Glasses with masking tape
Session 3	**Participant-5**	Canadian Accent	User 1	Rubber-band
	Participant-6	Canadian Accent	User 2	Glasses with masking tape

Table 6.2 describes the assignment of the impediments and user roles for all the participants during the three sessions.

Session 1

Table 6.3 shows time taken, number of attempts and error rate for participant 1. For participant 1 who had a Nigerian accent, the speech-to-text conversion worked well for simple sentences and foreign names and the error rate was zero percent. However, for difficult words like "creative confidence" and long sentences the error rate ranged from 83% to 88%. The average number of attempts was 3.

Table 6.3 Results for Participant 1

Our System (speech mode)	**Statements containing**	**Role**	**Participant-1**		
Statement number in scenario			**Time Taken (min:sec:ms)**	**No of attempts**	**Error rate**
1	Simple words like hello, how, are you?	1	0:23:9	1	0
4	Contains foreign usernames and confusing words like "TED"	1	0:12:8	3	0
6	Complicated words like "creative confidence"	1	0:12:0	4	88.88889
8	Long statement with difficult words like "applaud"	1	0:44:0	4	83.33333
Total			**1:31:17**	**12**	
Average				**3**	**43.05**

The average error rate was 43% and total time it took to play the part according to the scenario was 1 minute and 31 seconds. This participant had higher error rate with speech-to-text conversation than the other participants, but was skillful in using text on mobile phone. This participant preferred using text on mobile phone instead of using speech-to-text system, according to the feedback on the questionnaire.

Table 6.4 shows time taken, number of attempts and error rate for participant 2. For participant 2, who had a South Asian accent, the speech-to-text conversion worked well for simple sentences. For sentences containing foreign names, the error rate was around 10 percent. For the sentences containing the words “about” and “wow”, the error rate was 50% and 83% respectively. The average number of attempts was 2.4. The average error rate was 28% and the total time to play the scenario was 1 minute and 21 seconds.

Table 6.4 Results for Participant 2

Our System (speech mode)	**Statements containing**	**Role**	**Participant-2**		
Statement number in scenario			**Time Taken (min:sec:ms)**	**No of attempts**	**Error rate**
2	Simple words like "fine"	2	0:10:0	1	0
3	Contains foreign usernames	2	0:31:05	3	10
5	Simple words like "what" and "about"	2	0:13:09	4	50
7	Words like ""Wow" and "Watching"	2	0:17:04	3	83.33
9	Simple word like "Agree"	2	0:10:00	1	0
Total			1:21:18	12	
Average				2.4	28.333

This participant had higher error rate, when compared to participant-3, participant-4, participant-5, or participant-6, with speech-to-text conversation. Participant was also not skillful in using text on mobile phone. This participant preferred using text on mobile phone over the speech-to-text system, according to their feedback on the questionnaire.

Session 2

Table 6.5 shows time taken, number of attempts and error rate for participant 3. For participant 3, who had a Chinese accent, the speech-to-text conversion worked well for simple sentences, but sentences that included foreign names had an error rate around 25 percent. For the sentence containing difficult words ("creative confidence") and long sentences the error rate ranged from 11% to 38%. The average number of attempts was 2. The average error rate was 18% and total time taken was 1 minute and 53 seconds. This participant had low error rate with the speech-to-text conversation and was not skillful in using text on mobile phone. This participant preferred using speech-to-text system over text, according to their feedback on the questionnaire.

Table 6.6 shows time taken, number of attempts and error rate for participant 4. For participant 4, who also had a Chinese accent, the speech-to-text conversion was good for simple sentences. For sentences with foreign names the error rate was around 10 percent. For sentences containing words such as "wow", the error rate was 16%. The average number of attempts was 2. The average error rate was 5.3% and total time taken - 1 minute and 40 seconds. This participant had lower error rate with speech-to-text conversation and was also skillful in using text on mobile phone. This participant also preferred using speech-to-text system, and speech with editing over text, according to their feedback on the questionnaire.

Table 6.5 Results for Participant 3

Our System (speech mode)	Statements containing	Role	Participant-3		
Statement number in scenario			Time Taken (min:sec: ms)	No of attempts	Error rate
1	Simple words like hello, how ,are you?	1	0:25:0	1	0
4	Contains foreign usernames and confusing words like "TED"	1	0:15:00	3	25
6	Complicated words like "creative confidence"	1	0:17:5	1	11.11
8	Long statement with difficult words like "applauding"	1	0:56:5	3	38.89
Total			**1:53:10**	**8**	
Average				**2**	**18.75**

Table 6.6 Results for Participant 4

Our System (speech mode)	Statements containing	Role	Participant-4		
Statement number in scenario			Time Taken (min:sec:ms)	No of attempts	Error rate
2	Simple words like "fine"	2	0:15:5	1	0
3	Contains foreign usernames	2	0:36:0	3	10
5	Simple words like "what" and "about"	2	0:16:0	1	0
7	Words like ""Wow" and "Watching"	2	0:18:00	4	16.67
9	Simple word like "Agree"	2	0:15:5	1	0
Total			**1:40:10**	**10**	
Average				**2**	**5.334**

Session 3

Table 6.7 shows time taken, number of attempts and error rate for participant 5. For participant 5 who had a Canadian accent, the speech-to-text conversion was good for simple sentences, but sentences containing foreign names had error rate around 12.5 percent. For sentences containing difficult words like "creative confidence" and for long sentences the error rate was same i.e., 11.11 percent. The average number of attempts was 2. The average error rate was 8.63% and total time taken was 1 minute and 36 seconds. This participant had lower error rate with speech-to-text conversation and was not skillful in using text on mobile phone. This participant also preferred using the speech-to-text system over text.

Table 6.7 Results for Participant 5

Our System (speech mode)	**Statements containing**	**Role**	**Participant-5**		
Statement number in scenario			**Time Taken (min:sec:ms)**	**No of attempts**	**Error rate**
1	Simple words like hello, how, are you?	1	0:17:00	1	0
4	Contains foreign usernames and confusing words like "TED"	1	0:26:10	1	12.5
6	Complicated words like "creative confidence"	1	0:33:02	2	11.11
8	Long statement with difficult words like "applauding"	1	0:20:05	4	11.11
Total			1:36:17	8	
Average				2	8.68

Table 6.8 shows time taken, number of attempts and error rate for participant 6. For participant 6 who had a Canadian accent, the speech-to-text conversion worked well for simple sentences. For sentences with foreign names the error rate was around 10 percent. For sentences containing words like “wow”, the error rate was 0%. The average number of attempts was 1.4. The average error rate was 2% and total time taken was 1 minute and 17 seconds. This participant had the lowest error rate of all participants’ with speech-to-text conversation, and was skilful in using text on mobile phone. This participant also preferred using the speech-to-text system over text.

Table 6.8 Results for Participant 6

Our System (speech mode)	**Statements containing**	**Role**	**Participant-6**		
Statement number in scenario			**Time Taken (min:sec:ms)**	**No of attempts**	**Error rate**
2	Simple words like "fine"	2	0:11:89	1	0
3	Contains foreign usernames	2	0:25:86	1	10
5	Simple words like "what" and "about"	2	0:12:07	1	0
7	Words like ""Wow" and "Watching"	2	0:16:26	3	0
9	Simple word like "Agree"	2	0:13:55	1	0
Total			1:17:264	7	
Average				1.4	2

The Table 6.9 shows time taken for all participants for condition (iii). This table is useful to compare the time taken for all participants.

Table 6.9 Results for all participants in condition (iii)

Our System (speech mode))	Statements containing	Role	Participant-1	Participant-2	Participant-3	Participant-4	Participant-5	Participant-6
Statement number			Time Taken	Time Taken	Time Taken	Time Taken	Time Taken	Time Taken
1	Simple words like hello, how ,are you?	1	0:23:9		0:25:0		0:17:00	
2	Simple words like "fine"	2		0:10:0		0:15:5		0:11:89
3	Contains foreign usernames	2		0:31:05		0:36:0		0:25:86
4	Contains foreign usernames and confusing words like "TED"	1	0:12:8		0:15:00		0:26:10	
5	Simple words like "what" and "about"	2		0:13:09		0:16:0		0:12:07
6	Complicated words like "creative confidence"	1	0:12:0		0:17:5		0:33:02	
7	Words like ""Wow" and "Watching"	2		0:17:04		0:18:00		0:16:26
8	Long statement with difficult words like "applauding"	1	0:44:0		0:56:5		0:20:05	
9	Simple word like "Agree"	2		0:10:00		0:15:5		0:13:55
	Total Time		1.31.17	1.21.18	1.53.10	1.40.10	1.36.17	1.17.264

Comparison of three measures for Condition (iii)

The table 6.10 shows the average time taken, average number of attempts and average error rate for each participant while using the speech to text software in our proposed system.

Table 6.10: Average time taken, Average no. of attempts, Average error rate on Speech-to-text (Proposed system)

Participants		Accent	User role in pre-scripted scenario	Obstruction tool	Avg-Time taken (Seconds)	Avg-No. of attempts (Numeric)	Avg-Error rate (%)
Session 1	**Participant-1**	Nigerian Accent	User 1	Rubber-band	22.75	3	43.05
	Participant-2	South-Asian Accent	User 2	Glasses with masking tape	16.20	2.4	28.33
Session 2	**Participant-3**	Chinese Accent	User 1	Rubber-band	28.25	2	18.75
	Participant-4	Chinese Accent	User 2	Glasses with masking tape	20.00	2	5.334
Session 3	**Participant-5**	Canadian Accent	User 1	Rubber-band	24.00	2	8.68
	Participant-6	Canadian Accent	User 2	Glasses with masking tape	15.45	1.4	2

The participant-1, participant-3 and participant-5 are in role user-1 while running the pre-scripted chat scenario. The participant-2, participant-4 and participant-6 are in role user-2. The participants in same role can be compared and we know that role user-1 had more difficult sentences than role user-2. We can observe that in each session the average time taken by the participants in the role user-1 is always higher than the average time taken by the participants in the role user-2. In Table 6.10, we can also see that the error rate for Chinese accent is better than South Asian and Nigerian accent. We can also see that the error rate for Canadian accent is better than for Chinese accent. We also see a relationship between the average error rate and average number of attempts. For participants with lesser error rate the number of attempts is also less. We can see that the proposed system has a better accuracy for Canadian accent.

User -1 **User-2**

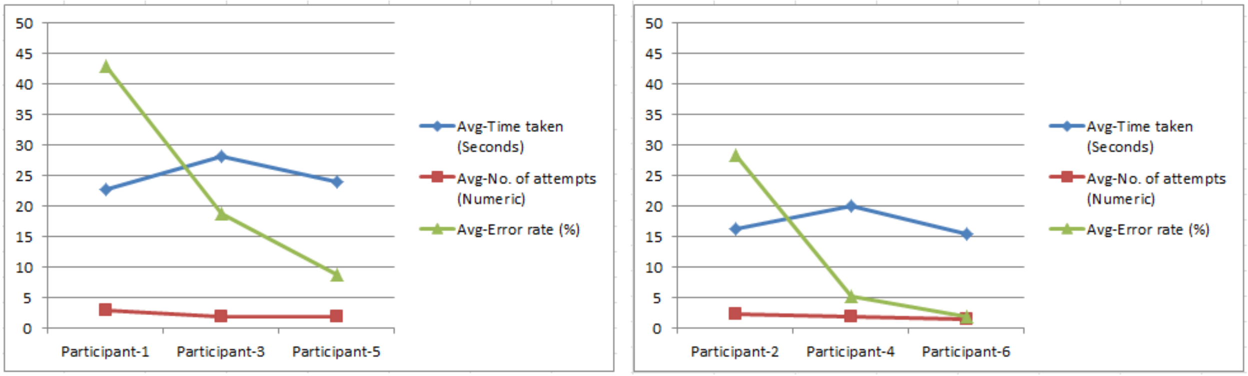

Figure 6.3 Average time taken, Average number of attempts, and Average error rate for participants in user role user-1 and user-2

Figure 6.3 is the graphical representation of average times taken, the average number of attempts, and the average error rate for the three participants in the roles of user-1 and user-2. In the graph, the blue line represents average time taken, red line represents average number of attempts and the green line represents average error rate. It shows a relationship between the average error rate and average number of attempts. For participants with lesser error rate the number of attempts is also less.

6.9.2 Results for Condition (i) - Text mode of entry on a tablet

When working in textual mode on a tablet the error rate for all participants was less than one percent. The average number of attempts was 1. So, for textual mode, we will only consider the time taken by each participant. The time to run the scenario in this mode ranges from 4 to 11 minutes, which is much longer compared to the time it took using the speech to text system.

Table 6.11 Results for all participants in condition (i)

Tablet /computer (text mode)	Statements containing	Role	Participant -1	Participant-2	Participant -3	Participant -4	Participant -5	Participant -6
Statement number			**Time Taken**	**Time Taken**	**Time Taken**	**Time Taken**	**Time Taken**	**Time Taken**
1	Simple words like hello, how ,are you?	1	2:0:0		2:10:0		1:30:00	
2	Simple words like "fine"	2		1:30:0		1:00:00		1:41:47
3	Contains foreign usernames	2		1:30:0		1::30:03		2:06:00
4	Contains foreign usernames and confusing words like "TED"	1	1:30:0		1:47:01		1:31:00	
5	Simple words like "what" and "about"	2		0:57:06		0:55:0		2:20:00
6	Complicated words like "creative confidence"	1	2:29:6		2:31:6		1:35:21	
7	Words like ""Wow" and "Watching"	2		0:58:0		0:59:00		1:54:00
8	Long statement with difficult words like "applauding"	1	2:40:03		2:45:02		2:10:00	
9	Simple word like "Agree"	2		0:36:06		0:32:0		2:23:28
	Total Time		**8:38:09**	**5:31:12**	**7:03:09**	**4:56:03**	**6:36:21**	**10:18:75**

The turn-around time for the sentences using the text based input was large when compared to speech-to-text system. Table 6.11 summaries the impediments and time taken for each participant.

It is important to note that the total time taken for each participant varies a lot, because the skillset required for using text mode varied largely between the participants. Difficulties with limited WI-FI connectivity during some of the sessions affected the time taken for the entered text to appear on the Twitter website. Table 6.11 shows time taken for all participants in condition (i).

6.9.3 Results for Condition (ii) Textual mode on a mobile phone

When working in textual mode on a mobile phone, the error rate for all participants was less than one percent. The average number of attempts was around 1. So, for textual mode, we will only consider the time taken by each participant.

The time taken to run the scenario in this mode ranges from 4 to 16 minutes, which is much longer compared to the time it took using the speech to text system. The turn-around time for the sentences using the text based input was large when compared to the speech-to-text system. The table 6.12 summarizes the impediments and time taken for each participant.

It is important to note that the total time taken for each participant varies a lot, because the skill of using text mode varied largely between the participants. Problems with limited WI-FI connectivity, affected the times. Table 6.12 shows time taken for all participants in condition (ii).

Table 6.12 Results for all participants in condition (ii)

Mobile phone (text mode)	**Statements containing**	**Role**	**Participant -2**	**Participant-1**	**Participant -4**	**Participant -3**	**Participant -6**	**Participant -5**
Statement number			**Time Taken**	**Time Taken**	**Time Taken**	**Time Taken**	**Time Taken**	**Time Taken**
1	Simple words like hello, how ,are you?	1	2:40:03		1:39:05		0:38:78	
2	Simple words like "fine"	2		4:00:00		1:22:00		0:15:00

3	Contains foreign usernames	2		2:10:00		1:30:00		1:10:00
4	Contains foreign usernames and confusing words like "TED"	1	3:00:00		2:39:08		0:59:90	
5	Simple words like "what" and "about"	2		3:20:00		1:30:05		1:50:00
6	Complicated words like "creative confidence"	1	4:30:58		1:25:06		0:43:06	
7	Words like ""Wow" and "Watching"	2		2:23:05		1:38:08		1:50:26
8	Long statement with difficult words like "applauding"	1	3:38:03		2:52:03		2:17:15	
9	Simple word like "Agree"	2		3:53:00		0:30:00		0:20:00
	Total Time		**13:48:64**	**15:46:05**	**8:35:22**	**6:30:14**	**4:37:189**	**5:25:26**

6.9.4 Compare the average time taken in all the three conditions

The table 6.13 shows the average time taken using speech on the proposed system, average time taken using text on the tablet, and average time taken using text on the mobile phone. We can observe that the average time taken using speech on the proposed system is far lesser than the average time taken using text on the tablet or the average time taken using text on the mobile phone for all the participants. We can also see that the average time taken on the tablet and the average time taken on the mobile phone varies for each participant, depending on participant's skillset in texting on the specific device.

Table 6.13 Average time taken for all the participants in all the three conditions

Participants		Avg-Time taken (speech on proposed system) in seconds	Avg-Time taken (text on tablet) in seconds	Avg-Time taken (text on mobile) in seconds
Session 1	**Participant-1**	22.75	129.5	207.01
	Participant-2	16.20	66.20	189.20
Session 2	**Participant-3**	28.25	105.75	128.75
	Participant-4	20.00	59.20	78.00
Session 3	**Participant-5**	24.0	99.00	69.26
	Participant-6	15.45	123.61	65.00

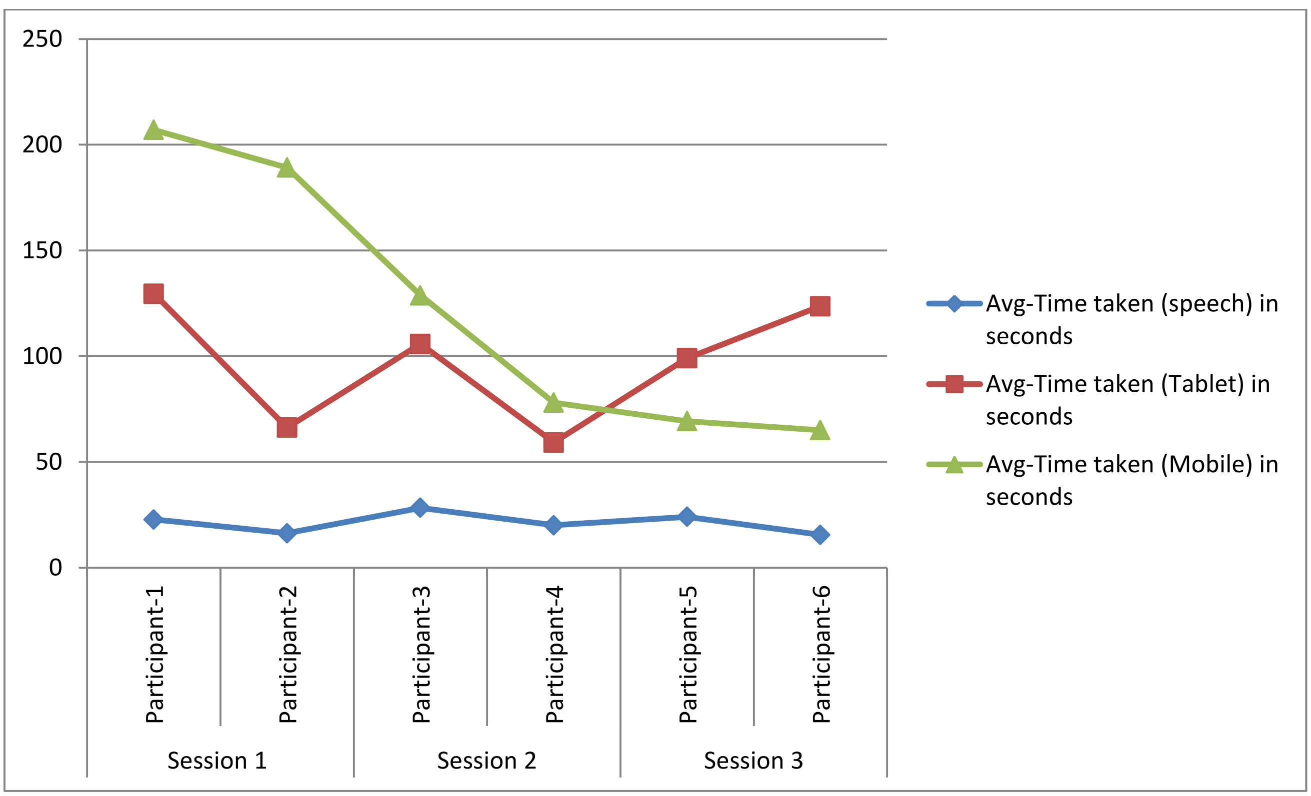

Figure 6.4 Average time taken on the proposed system (Speech), Average time taken on tablet (Text), and Average time taken on mobile phone (Text)

Figure 6.4 is the graphical representation of average time taken on the proposed system (Speech), average time taken on tablet (Text), and average time taken on mobile phone (Text). In the graph, the blue line represents average time taken using speech in our proposed system, red line represents average time taken using text on the tablet and the green line represents average time taken using text on the mobile phone. It shows that the average time taken using speech on the proposed system is far lesser than average time taken using text on the tablet or average time taken using text on the mobile phone for all the participants.

6.10 Discussion

The participants in condition (iii), i.e., using the proposed approach had much lower turn-around time for posting on Twitter, than in conditions (ii) and (i), i.e., using text entry mode on mobile phone and tablet. Thus a higher efficiency was observed in condition (iii), despite the higher number of attempts. The higher number of attempts and the errors in speech-to-text conversion in relation to complex words or foreign names led to frustration in some of the users, due to which user preferences for three different conditions were different. Two participants preferred text over speech. One participant preferred speech with editing capability over text. Three participants preferred using speech over text. Figure 6.5 shows that half of the participants were in favour of speech-to-text and 2/3 preferred using speech with editing, while only only 1/3 of participants wanted to use text only. Though, speech-to-text software has a higher error-rate, an improvised speech-to-text software (after intial voice training period) can minimize the error rate. The typing skillsets of the users was also an important factor.

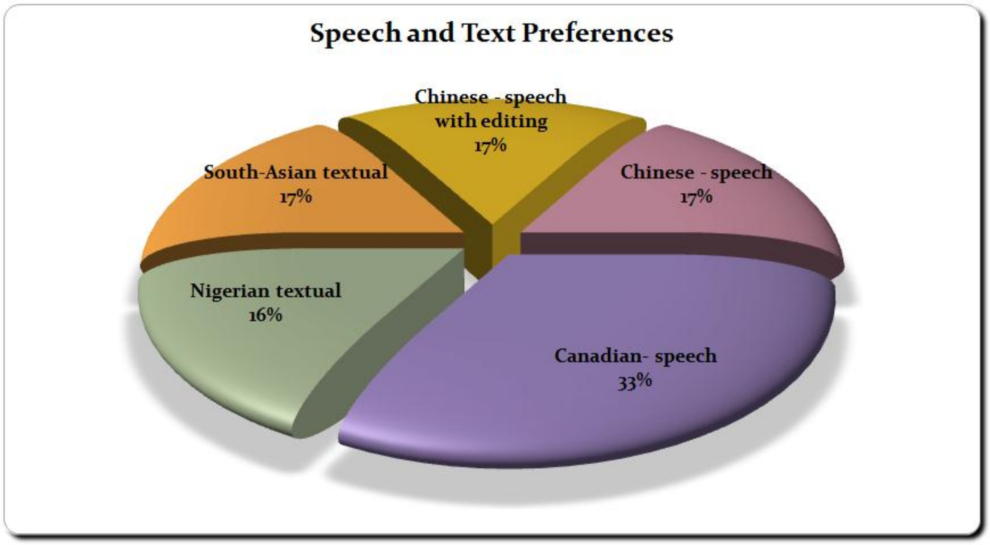

Figure 6.5 Speech and text preferences by participants with different accents

For users with better texting skills the time taken in entering the text on the mobile phone and tablet in the first two experimental conditions was much shorter when compared to time taken by the users, who weren't skilled in texting. In addition, the users with better texting skills preferred text when the system did not handle their accents well. The accent of users also makes a difference. We saw that the speech-to-text system performed much better (with lower error rates) with Canadian accents. In addition, we also found out that the diction while speaking mattered. In case of participant-4 who had a Chinese accent the error rate was around 5%, while participant-5 who had a Canadian accent had a error rate of 8%. The higher error rate for participant-5 was due to less clear diction, whereas participant-4 with a Chinese accent spoke clearly. Yet, most speech-to-text converters, including Vlingo, improve with usage, as they learn to recognize the speaker's voice, accent and specific pronounciation/diction. In this experiment, due to the short duration, it was impossible to train the software for particular users or accents, so the system was tested in harder conditions than in typical use. Still, the user acceptance of the system will depend on their initial experience with it, and it needs to be acceptable from the start, in order to encourage them to continue using it and to allow for the speech-to-text conversion to adapt to the individual.

The participants felt that it was harder to use the textual mode on the mobile phone and tablet with impediments. The Rubber-band restricted their finger movement, and the glasses with masking, made it difficult to see the keys on the tablet and the mobile phones. The participants often struggled a lot to complete the task and the impediments affected the time taken to finish the task in condition (i) and (ii). The impediments did not affect much of their performance in condition (iii), because there was only minimal finger movement and eye-sight required (just to verify the correctness of the speech-to-text conversion).

So the evaluation shows that

1) The proposed approach is feasible, as we have built a system prototype and have tested it successfully.
2) The speech-to-text does provide convenience for users especially when they have to struggle with low vision or manual dexterity deficiency.
3) Speech input (or) speech with editing can serve as a better alternative to text input.
4) People's skill set in using the mobile phones and computers has an effect on the time-spent on the system and their comfort level with system.

The evaluation showed also that numerous technical problems of the involved technologies can make the proposed system perform sub-standard, e.g. loss of Wi-Fi connection, or temporary unavailability of the IIS Twitter server. We used experimental technologies and an

IPTV simulation to demonstrate the feasibility and user acceptance of the approach in principle. Yet all IPTV and speech-to-text technologies are quickly maturing and we can expect that real applications with more reliable performance will appear very shortly (if not already there).

6.11 Limitations

While the study was intended as a proof of feasibility for the combination of technologies proposed to address the problem of providing means for elderly users to connect via social media during TV watching, much more work is required to find out if it would be a viable on the market. There are several important limitations in our user study. The user study was conducted with non-elderly users mimicing the impairments of elderly users with physical impediments. If elderly user group participants were used for the user study then results may have been different. We used a small group of 6 participants. If the number of participants were increased then the study results may change. In both the pilot study and evaluation, we used the same speech to text software - Vlingo. With the increasing number of speech recognition software available in the market, the accuracy of the speech to text conversion may vary based on the software. We did not get a chance to explore all the third party speech recognition software in the market due to financial restrictions and time. Testing the proposed system with an alternative speech software may be one possible direction for future work. We did not get a chance to explore the system with various video clips with different durations and genres, such as reality shows and sports. The participants tested the proposed system in a controlled environment with headphones to listen to the sound of the IPTV. In the real world, television users may experience some difficulties due to the noise introduced by the television viewing during the speech to text conversion, inspite of holding the mobile phone at a close proximity to their faces.

CHAPTER 7
CONCLUSION

In this thesis, we show that integrating different technologies holds a promise to make it easier for elderly people to access social networks during IPTV viewing. In chapter 2, we saw that IPTV is likely to transform television viewing in the next few years, merging it with social experience and richer interaction. Based on the current technological trends, it can be expected that in near future Internet and television will lose their individual identities. People will be able to communicate, network, shop, play games, and watch news, sports and movies together even at a distance.

The main aim of our work is to provide convenience to the elderly but also to regular users by introducing speech-to-text in the social networking and integrating it to IPTV. In the literature survey, we had found out that the elderly people comprise the largest proportion of television viewers. We also saw that elderly people commonly lack the technical skills to interact with mobile phones and other technology. In addition, health problems related to old age impose constrains on their vision and manual dexterity. Most elderly people are intimidated by the technology due to unfriendly interface.

In chapter 2, we also saw that social-networking usage is on a rise in all age groups. The social-networking usage growth rate is considerably higher in the elderly age group. If convenient user-interface is provided to the elderly, then they can be encouraged to be more active on social networking. Incorporating speech-to-text technology can benefit elderly and other age groups and help them use easier social networking in familiar context of TV watching. The elderly people can adapt to social networking, in addition to their television viewing activity, and thus regain the needed support and security in feeling connected with others.

The recent speech-to-text technologies are showing a great improvement. Recently, there have been many commercial applications for mobile phones. We know mobile phones are becoming more accessible. Using speech-to-text technology on mobile phones can help the elderly to use social-networking, while viewing television. Moreover, speech is the most natural way to communicate, and it can serve people of all age groups. Hence, integrating IPTV and social-networking will enhance the user's experience, and fulfill their need to be connected.

Many smart phones can run various third party applications, including speech-to-text conversion software, which can make them serve as a good interaction tool. The speech to text feature can be extended in the future to other activities, such as composing email and web searches.

In our approach, we integrate several different technologies -- speech-to-text technology, mobile phone, social networking, and IPTV -- to provide convenience to the elderly. There are lot of free third party applications, which does speech-to-text conversion on the recent smart phones. For our proposed work, we used a free third party application called Vlingo. We used blackberry torch as the mobile smart phone.

A system prototype was developed and improved after performing a pilot study. The evaluation was through a qualitative user study. A comparative facet was evaluated by user study conducted while using social networking in textual mode by means of a mobile phone, using social networking in textual mode by means of a tablet, using social networking using speech-to-text. Students and staff members participated in the user study.

From evaluation we found out 2/3 of participants preferred speech or speech with editing, while only 1/3 preferred only text among the six participants. We also saw that our proposed system had shorter turn-around time and was very quick. We also saw that people's diction and accent mattered while using the speech-to-text software. Some people who were very skilled at using text performed the texting tasks faster and preferred texting.

The main contribution of our approach is the integration of various technologies to provide convenience to elderly and regular users and encourage them to use social networking.

As a part of future work, the proposed system could also be extended to accommodate other social-networking web sites such as Facebook, Google plus etc. As Vlingo software is freely available software for most mobile phone platforms with reasonably accuracy, we choose to use it for the pilot and user study. For future work, we can explore other speech to text conversion software available on the market, to compare the accuracy of speech to text conversion, and performance of the proposed system, when used with different software in terms of turn-around time and number of attempts. It will be interesting to compare different speech recognition software with respect to the punctuation marks and symbols commonly used in our written language.

As elderly people are a vulnerable user group, we were not able to test our system with the target user group. For future work, it would be important to test the prototype with the actual target audience i.e. elderly people to evaluate in real time the system performance with respect to the target user group. As we evaluated the system prototype with limited number of participants, it will be a good idea to evaluate the system prototype with a large number of participants. It is important to test the system with real targeted user group to know if elderly people would be willing to use our system and measure the effect (for example, levels of frustration) during the usage. If the elderly people are satisfied with our system, we can apply speech technologies to the other areas of interest to them.

It would be also important to evaluate the system performance in more realistic home conditions, for example of higher noise level, for example a television set at high volume and no headphones while using the speech to text software.

For future, we can also explore offering a menu selection of a pre-stored set of fairly standard comments in combination to speech to text and editing. Further, it can also be extended to allow sending emails while watching IPTV. Elderly people can benefit, if email clients use a more natural mode of communication than typing. Speech to text technology has been deployed in a number of applications that offer the possibility to dictate email instead of typing. The improvement in user interaction on IPTV would allow users to interact naturally with web-content and applications.

LIST OF REFERENCES

[1]. Abascal, J., and Civit, A. (2001). Universal access to mobile telephony as a way to enhance the autonomy of elderly people. In Proceedings of the 2001 EC/NSF workshop on Universal accessibility of ubiquitous computing: providing for the elderly (WUAUC'01). ACM, New York, NY, USA, 93-99.

[2]. Barros, G., Zuffo, M. K., and Benini, J. (2006). Home Entertainment UI Continuum: Cell Phone, HTPC and iTV. In CHI2006 Workshop: Investigating new user experience challenges in iTV: mobility & sociability. Montreal, Canada.

[3]. Berthiaume, D. (2011). Marketing Chart: Do you ever use your phone to... (% of cell phone users). February 09 2011. Last accessed: February 16, 2012:http://www.marketingcharts.com/direct/85-of-us-adults-have-cell-phone-15958/pew-cell-phone-actions-feb-2011jpg/

[4]. Brian, M. (2010). Twitter's User Demographics Visualised [Infographics]. December 30, 2010. Last accessed: February 16 2012: http://thenextweb.com/twitter/2010/12/30/twitters-user-demographics-visualised-infographic/

[5]. Cathey, G. (2010). Booleanblackbelt: LinkedIn Demographics and Visitor Statistics. September 20, 2010. Last accessed: February 16 2012: http://www.booleanblackbelt.com/2010/09/linkedin-demographics-and-visitor-statistics/

[6]. Cesar, P., Chorianopoulos, K., and Jensen, J. (Eds.). (2007). Interactive TV: a shared experience: 5th European conference, EuroITV 2007, Amsterdam, the Netherlands, May 24-25, 2007: proceedings. Springer-Verlag, Berlin, Heidelberg. Pg: 74.

[7]. Chalfonte, B., Fish, R., and Kraut, R. (1991). Expressive richness: a comparison of speech and text as media for revision. In Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology (CHI '91), Scott P. Robertson, Gary M. Olson, and Judith S. Olson (Eds.). ACM, New York, NY, USA, 21-26.

[8]. Daniel. (2010). Speech recognition. December 12, 2010. Last accessed: February 16, 2012 : http://en.wikipedia.org/wiki/Speech_recognition

[9]. Darnell, M. (2008). Making digital tv easier for less-technically-inclined people. In Proceeding of the 1st international conference on Designing interactive user experiences for TV and video (UXTV '08). ACM, New York, NY, USA, 27-30.

[10]. Datagems. (2010). Age Demographics of U.S. Webmail Visitors. October 2010. Last accessed: February 16, 2012: http://www.comscoredatamine.com/2010/10/age-demographics-of-u-s-webmail-visitors/

[11]. Decosta, D. (2010). Importance of mobile phones. 20 July, 2010. Last accessed: December 12 2010: http://www.articlealley.com/article_ 43486_47.html

[12]. Dhasmana, S. (2010). How mobile phone make our life different. December 2010. Last accessed: December 12 2010: http://www.rifcomobile.com/articles/how-mobile-phone-make-our-life-different

[13]. Digital home. (2009). Canada's Largest Digital Service Providers. November 23, 2009. Last accessed: February 16, 2012 : http://www.digitalhome.ca/2009/11/canadas-largest-digital-service-providers/

[14]. Erwinkarim. (2010). IPTV. October 23, 2010. Last accessed on: February 16, 2012: http://en.wikipedia.org/wiki/IPTV

[15]. Elderly and the Media - Aging and Levels of Viewing, What Do Older Adults Watch?, Other Media Use. Last accessed: February 16, 2012: http://encyclopedia.jrank.org/articles/pages/6515/Elderly-and-the-Media.html

[16]. Enns, N. R. N., and MacKenzie, I. S. (1998). Touchpad-based remote control devices. In CHI 98 conference summary on Human factors in computing systems.1998, ACM Press: Los Angeles, California, United States, 18-23.

[17]. Epp, C, D. (2010). Protutor: A pronunciation tutor that uses historic open learner models. M.Sc. Thesis. University of Saskatchewan: CANADA.

[18]. Fisher, J. (2010). Internet and the Elderly. 2010. Last accessed: February 09, 2011: http://www.followitlive.com/internet-and-the-elderly/45157

[19]. Fuhrmann, T., Klein, M., and Odendahl, M. (2003). Bluewand-A versatile remote control and pointing device. In KiVS Kurzbeiträge 2003. 81-88.

[20]. Gonzalez, N. (2012). Global Audience, CheckFaceBook.com. (2012). Last accessed: February 16, 2012: http://www.checkfacebook.com/

[21]. Grove J, V. (2010). Baby Boomers and Seniors Are Flocking to Facebook [STATS]. January 28, 2010. Last accessed: February 16, 2012: http://mashable.com/2010/01/28/baby-boomers-social-media/

[22]. Hei, X., Liang, C., Liang, J., Liu Y., Ross, K.W. A Measurement Study of a Large-Scale P2P IPTV System. IEEE Transactions on Multimedia, , vol.9, no.8, pp.1672-1687, 2007

[23]. Hepburn, A. (2010). Facebook vs Twitter Demographics. December 21, 2010. Last accessed on: February 16, 2012: http://www.digitalbuzzblog.com/infographic-facebook-vs-twitter-demographics-2010-2011/

[24]. Hess, J., Küstermann, G., and Pipek, V. (2008). Premote: a user customizable remote control. In *CHI '08 extended abstracts on Human factors in computing systems* (CHI EA '08). ACM, New York, NY, USA, 3279-3284.

[25]. Jaanus. (2006). Elderly people like Skype in a Finnish test. March 2006. Last accessed: February 16 2012: http://blogs.skype.com/en/2006/03/elderly_people_like_skype_in_a.html

[26]. Jacko, J. (Ed.). (2007). Proceedings of the 12th International Conference on Human-Computer Interaction: Intelligent Multimodal Interaction Environments. Springer-Verlag, Berlin, Heidelberg. (Pg 979)

[27]. Java, A., Song. X., Finin, T., and Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities. In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 work shop on Web mining and social network analysis (WebKDD/ SNA -KDD '07) ACM, New York, NY, USA, 56-65.

[28]. Kim, L., Park, W., Cho, H., Park, S. (2010). A universal remote control with haptic interface for customer electronic devices. Consumer Electronics, IEEE Transactions on, 56(2), 913 -918.

[29]. Kim, S. H., Ok, J., Kang, H. J., Kim, M. C., and Kim, M. (2004). An interaction and product design of gesture based TV remote control. In *CHI '04 extended abstracts on Human factors in computing systems* (CHI EA '04). ACM, New York, NY, USA, 1548-1548.

[30]. Kurniawan, S., Mahmud, M., and Nugroho, Y. (2006). A study of the use of mobile phones by older persons. In CHI '06 extended abstracts on Human factors in computing systems (CHI '06). ACM, New York, NY, USA, 989-994.

[31]. Kurniawan., S. (2007). Older women and digital TV: a case study. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility* (Assets '07). Elsevier Ltd, ACM, New York, NY, USA, 251-252.

[32]. MacKenzie, I. S., and Jusoh, S. (2001). An Evaluation of Two Input Devices for Remote Pointing. In Proc. of the 8th IFIP International Conference on Engineering for Human-Computer Interaction (EHCI '01), Murray Reed Little and Laurence Nigay (Eds). Springer-Verlag. London, Uk, 235-250

[33]. Montpetit, M., Klym, N., Mirlacher, T. (2009). The future of IPTV: Adding social networking and mobility. Telecommunications, 2009. ConTEL 2009. 10th International Conference on, vol., no., pp.405-409.

[34]. Myers, B. A., Bhatnagar, R., Nichols, J., Peck, C.H., Kong, D., Miller, R. , and Long, A. C. (2002). Interacting at a distance: measuring the performance of laser pointers and other devices. In Proc. of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves. 2002, ACM Press: Minneapolis, Minnesota, USA, 33-40.

[35]. Nczempin. (2011). List of Speech recognition software. March 29, 2011. Last accessed: February 16, 2012 : http://en.wikipedia.org/wiki/List_of_speech_recognition_software

[36]. O'Driscoll, G. (2008). *Next Generation IPTV Services and Technologies*. Wiley-Interscience, New York, NY, USA.

[37]. Plos, O., Buisine, S. (2006). Universal design of mobile phones: A case study. CHI'06 International Conference on Human Factors in Computing Systems (work-in-progress). ACM, New York, NY, USA, 1229-1234.

[38]. Pollack, M. (2007). Intelligent Assistive Technology: The Present and the Future. In Proceedings of the 11th international conference on User Modeling (UM '07), Springer-Verlag, Berlin, Heidelberg, 5-6.

[39]. Quantcast. (2011). Facebook.com. December 2011. Last accessed: February 09 2011: http://www.quantcast.com/facebook.com

[40]. Quantcast. (2012). Quantcast: Youtube.com. January 2012. Last accessed : February 16, 2012: http://www.quantcast.com/youtube.com

[41]. Reuters. (2009). Elder like Twitter, Too. June 30, 2009. Last accessed: February 16, 2012: http://www.pcmag.com/article2/0,2817,2349535,00.asp

[42]. Rice, M., and Alm, N. (2007). Sociable TV: exploring user-led interaction design for older adults. In *Proceedings of the 5th European conference on Interactive TV: a shared experience* (EuroITV'07), Pablo Cesar, Konstantinos Chorianopoulos, and Jens F. Jensen (Eds.). Springer-Verlag, Berlin, Heidelberg, 126-135.

[43]. Robertson, S., Wharton, C., Ashworth, C., and Franzke, M. (1996). Dual device user interface design: PDAs and interactive television. In Proc. of the SIGCHI conference on Human factors in computing systems: common ground. (CHI '96), Michael J. Tauber (Ed.). ACM, New York, NY, USA, 79-86.

[44]. Rubin, A. M., and Rubin, R. B. (1982). Older persons' tv viewing patterns and motivations. Sage Publications, Inc. Communication Research, Vol 9 No (2): April 1982, 287-313.

[45]. Rucker, J, D. (2011). Facebook Demographics: A Comprehensive Guide (Amazing Squeezed Into One Page. January 22, 2011. Last accessed: February 16, 2012: http://soshable.com/facebook-demographics-a-comprehensive-guide-amazingly-squeezed-into-one-page/

[46]. Sditch. (2009). Prime time tv viewers by age demographics. May 10, 2009. Last accessed: February 16, 2012: http://www-958.ibm.com/software/data/cognos/manyeyes/visualizations/prime-time-tv-viewers-by-age-demog

[47]. Solis, B. (2010). The age of social networks. March 17, 2010. Last accessed: February 16, 2012: http://www.briansolis.com/2010/03/the-age-of-social-networks/

[48]. Spinks, J. (2012). Mobile_phone. February 9, 2012. Last accessed: February 16, 2012: http://en.wikipedia.org/wiki/Mobile_phone

[49]. Srinivasan, A. (2009). Twitter User age demographics. February 13, 2009. Last accessed : February 16, 2012: http://gorumors.com/crunchies/twitter-user-age-demographics/

[50]. Stinson, M .S., Elliot, L.B., Kelly, R.R., Liu, Y. (2009). Deaf and Hard-of-hearing Students' Memory of Lectures with speech-to-Text and Interpreting/Note Taking Services. Journal of Special Education (2009), 43(1), pp.52-64.

[51]. Sysomos Inc. (2010). Sysomos: Inside YouTube Videos. February 2010. Last accessed February 16, 2012: http://www.sysomos.com/reports/youtube/

[52]. Teknocrat. (2011). what is IPTV? IPTV Providers companies services. August 16, 2011. Last accessed : February 11, 2012 **:** http://www.teknocrat.com/what-is-iptv.html

[53]. Xiao, Y., and Du, X., Zhang, J., Hu, F., Guizani, S. (2007). Internet Protocol Television (IPTV): The Killer Application for the Next-Generation Internet. Communications Magazine, IEEE , vol.45, no.11, pp.126-134.

[54]. Yedinak, G. (2010). Social Networking Use for 50+ Demographic Doubles in 2010. August 30, 2010. Last accessed: February 16, 2012 http://seniorhousingnews.com/2010/08/30/social-networking-use-for-50-demographic-doubles-in-2010/

[55]. ZDNet Research. (2006). US Internet Users by age group. January 27, 2006. Last accessed : February 16 2012: http://www.zdnet.com/blog/itfacts/us-internet-users-by-age-group/10015

APPENDIX A
I. SCENARIO FOR PILOT STUDY

Assumption 1: The user is pre-logged into Vlingo's Twitter interface using one of the two Twitter user accounts.

1. Press left side convenience key on the Blackberry phone.
2. Please speak the sentence you intend to tweet followed by the word "Twitter". Please speak the sentences in distinctly voice while keeping the left-side convenience key pressed until the end of the sentence.
3. Once the sentence is over, release the left-side convenience key and wait to see if the message has been converted from speech to text on the Blackberry phone Twitter screen.
4. The Blackberry phone Twitter screen will show the converted speech as text. If you are satisfied with it, then approve it by clicking on the "Post" button. If you are unsatisfied with conversion, please cancel and start over at step 1.
5. If you approve the converted message, then wait on the Mediaroom simulator screen (IPTV screen) for the tweeted message to show up in the Twitter stream of the account currently visible in the Mediaroom simulator.

Instructions for user while using IPTV Mediaroom simulator:

1. If the current Mediaroom simulator screen doesn't show the user-profile page of the user used for the sending message on the Blackberry phone, then click on the "back" button on the user profile page which will take you to the "Index page".
2. On the Index page, user can select one of the two preset user accounts.
3. After the appropriate selection of user profile on the Mediaroom simulator, the tweets sent by the Blackberry phone will be displayed on the simulator's user profile page.

II. QUESTIONNAIRE USED IN THE PILOT STUDY

1. What do you think about the system in terms of functionality and general usage?

2. What are your suggestions for improvement?

3. What are the things (features) you like about the system?

4. What are the things (features) you dislike about the system?

5. From a scale of 1 to 5 rate the system?
   - Bad
   - Not good
   - Neutral
   - Good
   - Excellent
6. Do you prefer transparent or solid (opaque) layout in IPTV Mediaroom simulator?

7. Do you prefer horizontal or vertical message display?

8. What do you feel about the overall turn-around time of the system?

   - Quick
   - Slow
   - Moderate/Medium

III. RESULTS FROM PILOT STUDY

Participant 1 had a Chinese accent. The participant felt the user interface was convenient. The suggestions included improvement to the scalability to accommodate more users, and to add a better notification when the converted text appears on the Vlingo client. The participant also suggested using another speech to text conversion software instead of Vlingo, and generally comparing and evaluating the best option available to integrate into the proposed system. Participant 1 was not satisfied with the speech to text conversion of speech with Chinese accent. However, the participant was satisfied with the turn-around time, which was about five seconds. The participant also showed a preference towards the transparent horizontal layout. In terms of overall satisfaction with the system performance, the user rated it as good on a 5-point Likert scale. The 5-point Likert scale ranged from bad, fair, neutral, good, excellent.

Participant 2 had an East Indian accent. The participant was impressed with the functionality, and felt everyone including the people with limited manual dexterity and vision can benefit from it, as speech is the natural way to communicate. The participant suggested using real Blackberry phone, as a Blackberry simulator was used for pilot study instead of the Blackberry phone. However, the participant was quite satisfied with the general turn-around time which was about five seconds. The participant also showed a preference towards the transparent horizontal layout. In terms of overall satisfaction with the system performance, the user rated it as good.

Participant 3 was from Latin America and had an American accent. The participant suggested using another speech to text software for comparison and felt the system needs improvement in overall performance. The participant was satisfied with the user-interface. The participant was satisfied with the general turn-around time of 5 seconds. The participant also showed a preference towards transparent horizontal layout. In terms of overall satisfaction with the system performance, the user rated it as excellent.

Participant 4 had an East-Indian accent. The participant was satisfied with the functionality and felt that people can benefit from it. The participant suggested using another speech to text software for comparison. The participant felt the speech-to-text software was not able to handle foreign names. However, the participant was satisfied with the general turn-around time of five seconds. The participant also showed a preference towards transparent and horizontal layout. In terms of overall satisfaction with the system performance, the user rated it as excellent.

Participant 5 had an East-Indian accent. The participant was satisfied with the functionality, but felt remote control on the IPTV simulator was not intuitive, which made it difficult to use. The participant suggested using another speech to text software for comparison. The participant felt that the speech-to-text software was good, with the exception of handling foreign names. However, the participant was satisfied with the general turn-around time of five seconds. The participant also showed a preference towards transparent horizontal layout. In terms of overall satisfaction with the system performance, the user rated it as excellent.

Participant 6 had a Nigerian accent. The participant was impressed with the functionality and the felt the system provided convenience. The participant suggested devising a constructive way to help people to adapt the technology. The participant felt that the speech-to-text software was good, with the exception of handling foreign names and long sentences. However, the participant was satisfied with the general turn-around time which was about five seconds. The participant also showed a preference towards transparent horizontal layout. In terms of overall satisfaction with the system performance, the user rated it as excellent.

APPENDIX B

I. SCREEN SHOTS

Figure B.1 Blackberry screen shots (i) Blackberry Home page (ii) Blackberry Application folder with "Vlingo"

(iii) Blackberry screen Twitter client Interface

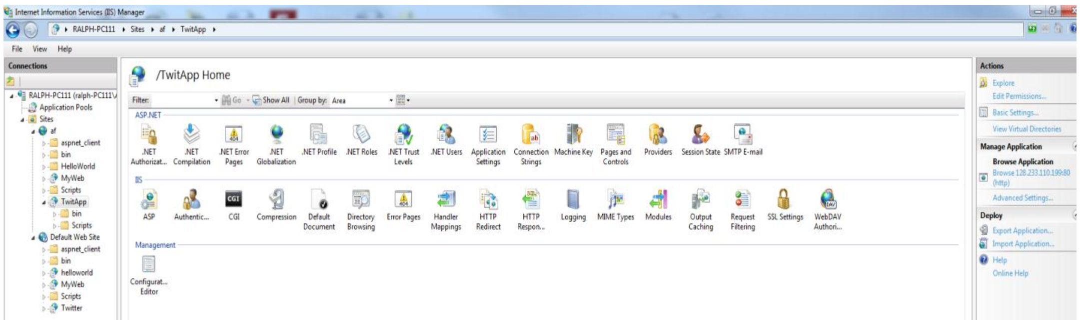

Figure B.2 IIS of Twitter web service

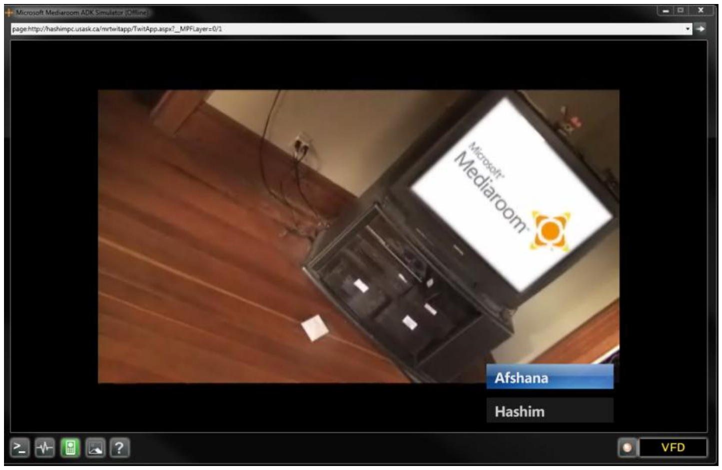

Figure B.3 Index Page with user accounts

Figure B.4 User Profile Page when the tweets are hidden

Figure B.5 IPTV remote control

APPENDIX C

I. SCENARIO

Phase 1:

User 1: Hello, How are you?

User 2: I am fine.

Phase 2:

User 2: Thanks for asking, USERNAME-1. What are you doing right now?

User 1: USERNAME-2, I am watching the TED TALK show.

User 2: What is it about?

User 1: It is about "How to build your creative confidence"

User 2: WOW!!! I am watching it too.

Phase 3

User 1: Interesting. I applaud talks like this...This is how great change is made...we are all creative creatures.

User 2: I Agree.

II. CONSENT FORM

UNIVERSITY OF SASKATCHEWAN

DEPARTMENT OF COMPUTER SCIENCE

UNIVERSITY OF SASKATCHEWAN

You are invited to participate in a study entitled *("Integrating IPTV to Social networking")*. Please read this form carefully, and feel free to ask the researchers any questions you might have.

Researchers:

Julita Vassileva, Department of Computer Science (966-2073), jiv@cs.usask.ca

Afshana Hussain Syed Hussain, Department of Computer science, afs371@mail.usask.ca

The system will integrate the social networking to IPTV and provide convenience to the elderly, people with limited dexterity, people with limited vision and other regular users to use Social networks by means of speech technology on smart phone. The goal of this study is to understand if users with limited vision and dexterity can navigate the interface easily, understand, and to find out what users think of the System. The evaluation takes around 30 minutes, and can be done at a computer in the MADMUC/ IMAGE Lab, if they wish so.

There are no known risks in this study.

(the following paragraph describes in brief the purpose and method of the particular study, in this case Usability Evaluation of our system)

Findings from the study will be used to evaluate user-friendliness of the interface for the targeted user group, and it can be helpful in enhancing the interface of the system. You will be given the wearable glasses to simulate the users with limited vision, and rubber-band will be wrapped around the fingers to simulate the limited dexterity users. You will be observed and videotaped during your use of the system, and then you will need to fill a short questionnaire and participate in a brief interview about your experience with the system.

<u>You will receive a $15 honorarium for your participation in the study</u>.

The research data and video tapes will be stored on a password-protected computer system and will be available only to the researchers. Personally identifying information will be destroyed upon completion of data collection, and pseudonyms (alias) will be used to refer to the participants. The data will be kept by the researchers for a minimum of five years upon the completion of this study in a secure storage.

Aggregate results will be used in a M.Sc. thesis and articles published in peer reviewed conferences and scientific journals. However, any information that can be linked to a specific participant will be removed or altered.

Your participation is voluntary, and you may withdraw from the study for any reason, at any time, without penalty of any sort. You may refuse to answer individual questions. If you withdraw from the study at any time, any data that you have contributed will be destroyed at your request.

If you have any questions concerning the study, please feel free to ask at any point; you are also free to contact the researchers if you have questions at a later time. This study has been approved on ethical grounds by the University of Saskatchewan Behavioral Research Ethics Board on (insert date). Any questions regarding your rights as a participant may be addressed to that committee through the Ethics Office (966-2084). Out of town participants may call collect. You may find out about the results of the study through the MADMUC website (http://madmuc.usask.ca) or by contacting the researchers.

I have read and understood the description provided above; I have been provided with an opportunity to ask questions and my questions have been answered satisfactorily. I consent to participate in the study described above and I understand that I may withdraw this consent at any time. A copy of this consent form has been given to me for my records.

____________________________________ ________________________________

(Name of Participant) (Date)

___________________________________ ________________________________

(Signature of Participant) (Signature of Researcher)

III. DEMOGRAPHIC QUESTIONNAIRE

1. Gender	○ Male ○ Female
2. Age	○Above 65 ○ Below 65
3. Are you currently (and mainly)	○ A student ○ working ○ Retired
4. How many hours a day do you spend on the Television?	○ 0 – 2 hours ○ 3 – 5 hours ○ 6 – 8 hours ○ over 8 hours
5. How comfortable are you with computers or tablets?	○ Very basic skills ○ Take time to learn ○ Comfortable ○ Skilled
6. How comfortable are you with Mobile phone use?	○ Very basic skills ○ Take time to learn ○ Comfortable ○ Skilled

7. Where do you most often access the Internet from? Please check all that applies	☐ Home ☐ Work ☐ University ☐ Recreation Centre ☐ Other
8. What is your native language?	☐ English speaker ☐ Non-English speaker
9. Cultural Background?	☐ Asian ☐ African ☐ American ☐ European ☐ Other
10. Whom do you mainly contact when online? Please check all that applies	☐ Family ☐ Students ☐ Co-workers ☐ Teachers / Professors ☐ Friends ☐ Strangers ☐ Others

IV. USABILITY & COMPARATIVE STUDY QUESTIONNAIRES

Measures	Accessing Social networking using text on tablet	Accessing Social networking using text on mobile phone	Accessing Social networking using speech-technology on our proposed system
Time Spent			
PHASE-1			
PHASE-2			
PHASE-3			
Number of attempts			
PHASE-1			
PHASE-2			
PHASE-3			
Error (Number of words)			
PHASE-1			
PHASE-2			
PHASE-3			

Usability & Comparative study questionnaires using 5-Likert scale

Measures	Accessing Social networking using text on tablet	Accessing Social networking using text on mobile phone	Accessing Social networking using speech-technology on our proposed system
Satisfaction			
Learn-ability			
Error-Handling			
Turn-around time			

V. EXIT QUESTIONNAIRE

1) How easy was it to understand the goal of our system? (5 Likert Scale)

2) How easy was it to understand and use the IPTV simulator screen interface? (5-Likert Scale)

3) Any particular difficulty encountered while using the IPTV simulator screen interface?

4) How easy was it to navigate through the IPTV simulator remote? (5-Likert Scale)

5) Any particular difficulty encountered while navigation of IPTV simulator remote?

6) How easy was to understand and use the speech to text s/w on Blackberry phone? (5-Likert Scale)

7) Any particular difficulty encountered while using speech software?

8) Were you satisfied with the system's accuracy in speech-to-text conversion? (5-Likert Scale)

9) Any concerns/comments about the accuracy?

10) Were you satisfied with the system's accent handling in speech-to-text conversion? (5-Likert Scale)

11) Any concerns/comments about the accent handling?

12) Were you frustrated while using the system? Please explain…

13) Would you rather use Text on the Mobile phone (or) Text on the Tablet to access the social networking instead of using our system? Which one?

14) Show the design of horizontal and vertical screen and ask for preference?

15) Were you satisfied with the overall performance of the system? (5-Likert Scale)

VI. RESULTS FROM EVALUATION

The study was held in 3 different rounds on each participant. It included

1. Accessing social networking on tablet/computer (textual mode).
2. Accessing social networking on mobile phone (textual mode).
3. Accessing social networking on our system (speech mode).

Scenario: In the scenario the two users are communicating with each other using social networking (twitter) via IPTV was to be simulated. So, a Ted talk show of about 7 minutes was downloaded and added to the IPTV simulator channels on different machines. The video was played at the same time and users were asked to communicate with pre-scripted user chat script. The chat script was scripted in such a way that it includes hard and difficult words, long sentences and the foreign name. The scenario itself is further divided into three phases to evaluate the time taken, error rate and number of user attempts in each phase and individual sentences.

SESSION 1:

Participant 1:

Background: The female participant was aged below 30 and is employed. The participant typically spent 0 to 2 hours on the television set every day and had comfortable skills in using handling computers and mobile phones. The participant preferred using internet at home and work. The participant is a Nigerian and non-English speaker. The participant has a Nigerian accent. The participant used the internet to stay in touch with family and friends.

Experiments and technical difficulties: It is important to note that the participant were involved in the study in a passive manner. As the user had to wait for the other participants to respond before carrying out speaking the assigned dialogue, user was also involved in texting and talking to people in the room while watching the participants. The participant asked questions in between study. The user had to repeat the sentences many times to the complicated sentences and long sentences. The IPTV server went down and we had to re-run the IPTV server, because the messages stopped appearing on the IPTV simulator screen. The Twitter website went down once. The participant and the co- participant went out of sync while performing the task and swapped each-others dialogues. The user did not pay much attention to video in the IPTV simulator screen.

First round: Initially, the participant started using the social networking on the computer/tablet using the text mode, which is the first part of the experiment in our comparative study. While using textual mode on the tablet, the participant wore glasses to simulate low-vision. While using the textual mode on the computer, participant felt fairly satisfied (rated “3” on 5-point Likert scale). The participant was satisfied with the error-handling capabilities (rated “4” on 5-point Likert scale). The participant was satisfied with the turn-around time (rated “4” on 5-point Likert scale). The participant was satisfied with the learnability of the system (rated “5” on 5-point Likert scale).

Second round: For the second experiment, the participant started using the social networking on the mobile phones using the text mode, in our comparative study. While using textual mode on the mobile phone, the participant wore rubber-bands around the finger to simulate limited manual dexterity. While using the textual mode on the mobile phone, participant felt not satisfied (rated “2” on 5-point Likert scale). The participant was fairly satisfied with the error-handling capabilities (rated “3” on 5-point Likert scale). The participant was satisfied with the turn-around time (rated “4” on 5-point Likert scale). The participant was satisfied with the learnability of the system (rated “4” on 5-point Likert scale).

Final round: For the final set of experiment, the participant was asked to use our system prototype which has speech to text capabilities. The participant was instructed at the beginning on the experiment and was video-taped while conducting the experiments. While using the speech mode on the mobile phone, participant was not satisfied (rated “2” on 5-point Likert scale). The participant was not satisfied with the error-handling capabilities (rated “2” on 5-point Likert scale). The participant was satisfied with the turn-around time (rated “4” on 5-point Likert scale). The participant was satisfied with the learnability of the system (rated “4” on 5-point Likert scale).

Feedback/suggestions: The participant felt that understanding the overall goal of the system was very easy, by rating the ease of understanding at “5” on a 5-point Likert scale. The participant also rated the IPTV simulator screen interface and remote interface (Refer to

Appendix B – B5) as not easy to understand or use ("2" on the Likert scale). The participant felt that the speech-to-text was easy to use, by rating it as "5" on a 5-point Likert scale. The participant felt that speech-to-text software had difficulty while converting long sentences and difficult words. The participant was not satisfied with the accuracy of the speech-to-text conversion (rated "1" on a 5-point Likert scale). The participant felt the accuracy has to be improved considering the error rate. The participant felt that accent handling capability was fair (rated "3" on a 5-point Likert scale). The user was frustrated with using long, complicated sentences. Participant repeated the sentences to get the right conversion. The participant chooses to use the text mode. The participant felt currently using only speech is not a good option. The participant rated the overall system as satisfying, by rating it at "3" on 5-point Likert scale.

Participant 2:

Background: The male participant was aged above 30 and is a student. The participant typically spent 0 to 2 hours on the television set every day and had comfortable skills in using handling computers and takes time to learn about mobile phones. The participant preferred using internet at home and university. The participant is primarily a South-Asian and non-English speaker. The participant has a South Asian accent. The participant used the internet to stay in touch with teachers, co-workers and friends.

Experiments and technical difficulties: It is important to note that the participant were involved in the study in a passive manner. As the user had to wait for the other participants to respond before carrying out speaking the assigned dialogue, user was also involved in texting and talking to people in the room while watching the participants. The participant asked questions in between study. The user had to repeat the sentences many times to get the correct conversion for complicated sentences. The IPTV server went down and we had to re-run the IPTV server, because the messages stopped appearing on the IPTV simulator screen. The Twitter website went down once. The participant and the co- participant went out of sync while performing the task and swapped each-others dialogues. The Wi-Fi connection went down thrice, while using the speech to text software on the mobile phone and the participants needed assistance. So, the participants had to wait till the WI-FI was working properly. This participant was engaged and paid more attention to the video and the IPTV simulator. The user also sometimes pressed the "send" button twice, and the Vlingo showed a warning for "duplicate message" and user was instructed to press the send button only once. Sometimes, the user send the message all the way to the twitter and wanted to attempt the same message again in which case the Message had to be erased from Twitter website of the user account. The user tried to focus on the video in the IPTV simulator screen and always wore the head phones.

First round: Initially, the participant started using the social networking on the tablet using the text mode, which is the first part of the experiment in our comparative study. While using textual mode on the tablet, the participant wore rubber-bands around the finger to simulate limited manual dexterity. While using the textual mode on the tablet, participant felt not satisfied (rated "2" on 5-point Likert scale). The participant was satisfied with the error-handling capabilities (rated "4" on 5-point Likert scale). The participant was satisfied with the turn-around time (rated "4" on 5-point Likert scale). The participant was satisfied with the learnability of the system (rated "4" on 5-point Likert scale).

Second round: For the second experiment, the participant started using the social networking on the mobile phones using the text mode, in our comparative study. While using textual mode on the mobile phone, the participant wore glasses to simulate low-vision. While using the textual mode on the mobile phone, participant felt satisfied (rated "4" on 5-point Likert scale). The participant was satisfied with the error-handling capabilities (rated "4" on 5-point Likert scale). The participant was fairly satisfied with the turn-around time (rated "3" on 5-point Likert scale). The participant was satisfied with the learnability of the system (rated "4" on 5-point Likert scale).

Final round: For the final set of experiment, the participant was asked to use our system prototype which has speech to text capabilities. The participant was instructed at the beginning on the experiment and was video-taped while conducting the experiments. While using the speech mode on the mobile phone, participant was fairly satisfied (rated "3" on 5-point Likert scale). The participant was not satisfied with the error-handling capabilities (rated "2" on 5-point Likert scale). The participant was completely satisfied with the turn-around time (rated "5" on 5-point Likert scale). The participant was satisfied with the learnability of the system (rated "4" on 5-point Likert scale).

Feedback/suggestions: The participant felt that understanding the overall goal of the system was very easy, by rating the ease of understanding at "5" on a 5-point Likert scale. The participant also rated the IPTV simulator screen interface as easy ("4" on the Likert scale). The participant felt that he did not really use the IPTV remote or IPTV simulator a lot. The participant rated the IPTV simulator remote as “3” on a 5-point Likert scale. The participant felt that the speech-to-text was easy to use, by rating it as “4” on a 5-point Likert scale. The participant felt that speech-to-text software had difficulty while converting the words which sounded similar. For example, the word “talking” was converted to “walking”. The participant was fairly satisfied with the accuracy of the speech-to-text conversion (rated “3” on a 5-point Likert scale). The participant felt the accuracy has to be improved considering the error rate. The participant felt that accent handling capability was good (rated “4” on a 5-point Likert scale). The participant felt the speech software will perform better for native English language speaker. The user was frustrated with the technical problems and associated delays while conducting experiment. The participant preferred to use the text mode. The participant felt currently using only speech is not a good option. The participant also felt horizontal layout is preferable on the IPTV simulator screen. The participant rated the overall system as satisfying, by rating it at “3” on 5-point Likert scale.

SESSION 2:

Participant 3:

Background: The female participant was aged below 30 and is a student. The participant typically spent 0 to 2 hours on the television set every day and had comfortable skills in using handling computers and mobile phones. The participant preferred using internet at home, work and university. The participant is Chinese and non-English speaker. The participant has a Chinese accent. The participant used the internet to stay in touch family, teachers, co-workers and friends.

Experiments and technical difficulties: It is important to note that the participant were involved in the study in a passive manner. As the user had to wait for the other participants to respond before carrying out speaking the assigned dialogue, user was also involved talking to people in the room. The participant asked questions in between study. The user had to repeat the sentences many times to the complicated sentences and long sentences. The user tried to focus on the video in the IPTV simulator screen.

First round: Initially, the participant started using the social networking on the tablet using the text mode, which is the first part of the experiment in our comparative study. While using textual mode on the tablet, the participant wore glasses to simulate low-vision. While using the textual mode on the tablet, participant felt completely satisfied (rated “5” on 5-point Likert scale). The participant was satisfied with the error-handling capabilities (rated “4” on 5-point Likert scale). The participant was fairly satisfied with the turn-around time (rated “3” on 5-point Likert scale). The participant was completely satisfied with the learnability of the system (rated “5” on 5-point Likert scale).

Second round: For the second experiment, the participant started using the social networking on the mobile phones using the text mode, in our comparative study. While using textual mode on the mobile phone, the participant wore rubber-bands around the finger to simulate limited manual dexterity. While using the textual mode on the mobile phone, participant felt satisfied (rated “4” on 5-point Likert scale). The participant was fairly satisfied with the error-handling capabilities (rated “3” on 5-point Likert scale). The participant was fairly satisfied with the turn-around time (rated “3” on 5-point Likert scale). The participant was completely satisfied with the learnability of the system (rated “5” on 5-point Likert scale).

Final round: For the final set of experiment, the participant was asked to use our system prototype which has speech to text capabilities. The participant was instructed at the beginning on the experiment and was video-taped while conducting the experiments. While using the speech mode on the mobile phone, participant was satisfied (rated “4” on 5-point Likert scale). The participant was fairly satisfied with the error-handling capabilities (rated “3” on 5-point Likert scale). The participant was completely satisfied with the turn-around time (rated “5” on 5-point Likert scale). The participant was completely satisfied with the learnability of the system (rated “5” on 5-point Likert scale).

Feedback/suggestions: The participant felt that understanding the overall goal of the system was very easy, by rating the ease of understanding at “5” on a 5-point Likert scale. The participant also rated the IPTV simulator screen interface as very difficult to understand (“2” on the Likert scale). The navigation from/to different user profile was difficult to understand on the interface of the IPTV simulator screen. In addition, the participant also felt need for more icons on the IPTV simulator screen and meaningful buttons on the IPTV remote control. The participant rated the IPTV simulator remote as “3” on a 5-point Likert scale. The participant felt that the speech-to-text was very easy to use, by rating it as “5” on a 5-point Likert scale. The participant felt that speech-to-text software had difficulty while converting the difficult words and requires slower pace of speaking. The participant was fairly satisfied with the accuracy of the speech-to-text conversion (rated “3” on a 5-point Likert scale). The participant felt the accuracy has to be improved and the speech software should make meaningful predictions. The participant felt that accent handling capability was not good (rated “2” on a 5-point Likert scale). The participant chooses to use speech. The participant also felt horizontal layout is preferable on the IPTV simulator screen. The participant rated the overall system as satisfying, by rating it at “3” on 5-point Likert scale.

Participant 4:

Background: The female participant was aged below 30 and is a student. The participant typically spent 0 to 2 hours on the television set every day and had comfortable skills in using handling computers and mobile phones. The participant preferred using internet at home. The participant is Chinese and non-English speaker. The participant has a Chinese accent. The participant used the internet to stay in touch family and friends.

Experiments and technical difficulties: It is important to note that the participant were involved in the study in a passive manner. As the user had to wait for the other participants to respond before carrying out speaking the assigned dialogue, user was also involved talking to people in the room. The participant asked questions in between study. The user had to repeat the sentences many times to the complicated sentences and long sentences. The user tried to focus on the video in the IPTV simulator screen. The Wi-Fi connection went down once. The Twitter gave warning for duplicate entries and the participant was assisted.

First round: Initially, the participant started using the social networking on the tablet using the text mode, which is the first part of the experiment in our comparative study. While using textual mode on the tablet, the participant wore rubber-bands around the fingers to simulate minimum manual dexterity. While using the textual mode on the tablet, participant felt satisfied (rated “4” on 5-point Likert scale). The participant was completely satisfied with the error-handling capabilities (rated “5” on 5-point Likert scale). The participant was fairly satisfied with the turn-around time (rated “3” on 5-point Likert scale). The participant was completely satisfied with the learnability of the system (rated “5” on 5-point Likert scale).

Second round: For the second experiment, the participant started using the social networking on the mobile phones using the text mode, in our comparative study. While using textual mode on the mobile phone, the participant wore glasses to simulate low vision. While using the textual mode on the mobile phone, participant felt not satisfied (rated “2” on 5-point

Likert scale). The participant was satisfied with the error-handling capabilities (rated “4” on 5-point Likert scale). The participant was fairly satisfied with the turn-around time (rated “3” on 5-point Likert scale). The participant was completely satisfied with the learnability of the system (rated “5” on 5-point Likert scale).

Final round: For the final set of experiment, the participant was asked to use our system prototype which has speech to text capabilities. The participant was instructed at the beginning on the experiment and was video-taped while conducting the experiments. While using the speech mode on the mobile phone, participant was hardly satisfied (rated “1” on 5-point Likert scale). The participant was hardly satisfied with the error-handling capabilities (rated “1” on 5-point Likert scale). The participant was completely satisfied with the turn-around time (rated “5” on 5-point Likert scale). The participant was fairly satisfied with the learnability of the system (rated “3” on 5-point Likert scale).

Feedback/suggestions: The participant felt that understanding the overall goal of the system was very easy, by rating the ease of understanding at “5” on a 5-point Likert scale. The participant also rated the IPTV simulator screen interface as very difficult to understand (“2” on the Likert scale). The participant felt strongly that a user manual was required to understand the interface and functioning of the IPTV simulator screen. In addition, the participant also required a better feedback on the IPTV simulator, while different arrow keys are used on the IPTV remote. The participant rated the IPTV simulator remote as “2” on a 5-point Likert scale. The participant felt that IPTV simulator remote interface was confusing and difficult to understand. The participant felt that the speech-to-text was very easy to use, by rating it as “5” on a 5-point Likert scale. The participant felt that speech-to-text software had difficulty while converting the spoken punctuation marks, names of people and difficult words. The participant was not satisfied with the accuracy of the speech-to-text conversion (rated “1” on a 5-point Likert scale). The participant felt that accent handling capability was very good (rated “4” on a 5-point Likert scale), except for handling the names of the people. The participant felt frustrated while using the speech mode to communicate with the social network, as she had to do it multiple times before getting it right. The handling of long sentences by the speech software was poor. The participant chooses to use speech with textual corrections/editing followed by the text mode. The participant felt currently using only speech is not a good option. The participant also felt horizontal layout is preferable on the IPTV simulator screen. The participant rated the overall system as satisfying, by rating it at “3” on 5-point Likert scale.

SESSION 3:

Participant 5:

Background: The female participant was aged above 30 and held a job. The participant typically spent 0 to 2 hours on the television set every day and had comfortable skills in using handling computers and took time to learn using mobile phones. The participant preferred using internet at home and work. The participant is a Canadian and English speaker. The participant has a Canadian accent. The participant used the internet to stay in touch family.

Experiments and technical difficulties: It is important to note that the participant were involved in the study in a passive manner. As the user had to wait for the other participants to respond before carrying out speaking the assigned dialogue, user was also involved talking to people in the room. The participant asked questions in between study. The user had to repeat the sentences many times to the complicated sentences and long sentences. The user tried to focus on the video in the IPTV simulator screen.

First round: Initially, the participant started using the social networking on the tablet using the text mode, which is the first part of the experiment in our comparative study. While using textual mode on the tablet, the participant wore rubber-bands around the fingers to simulate minimum manual dexterity. While using the textual mode on the tablet, participant felt fairly satisfied (rated “3” on 5-point Likert scale). The participant was fairly satisfied with the error-handling capabilities (rated “3” on 5-point Likert scale). The participant was fairly satisfied with the turn-around time (rated “3” on 5-point Likert scale). The participant was completely satisfied with the learnability of the system (rated “5” on 5-point Likert scale).

Second round: For the second experiment, the participant started using the social networking on the mobile phones using the text mode, in our comparative study. While using textual mode on the mobile phone, the participant skipped wearing glasses to simulate low vision, as the participant already suffered from low-vision. So, the participant was just advised to take her glasses off to conduct the experiment. While using the textual mode on the mobile phone, participant felt very satisfied (rated “4” on 5-point Likert scale). The participant was very satisfied with the error-handling capabilities (rated “4” on 5-point Likert scale). The participant was very satisfied with the turn-around time (rated “4” on 5-point Likert scale). The participant was completely satisfied with the learnability of the system (rated “5” on 5-point Likert scale).

Final round: For the final set of experiment, the participant was asked to use our system prototype which has speech to text capabilities. The participant was instructed at the beginning on the experiment and was video-taped while conducting the experiments. While using the speech mode on the mobile phone, participant felt extremely satisfied (rated “5” on 5-point Likert scale). The participant was satisfied with the error-handling capabilities (rated “4” on 5-point Likert scale). The participant was completely satisfied with the turn-around time (rated “5” on 5-point Likert scale). The participant was completely satisfied with the learnability of the system (rated “5” on 5-point Likert scale).

Feedback/suggestions: The participant felt that understanding the overall goal of the system was very easy, by rating the ease of understanding at “5” on a 5-point Likert scale. The participant also rated the IPTV simulator screen interface as easily understandable and straight-forward (“5” on the Likert scale). However, the participant rated the IPTV simulator remote as “4” on a 5-point Likert scale. The participant felt that IPTV simulator remote interface was difficult to use, but not difficult to understand. The participant felt that the speech-to-text was very easy to use, by rating it as “5” on a 5-point Likert scale. The participant felt that speech-to-text software had difficulty while converting the longer words, names of people and difficult words. The participant was quiet satisfied with the accuracy of the speech-to-text conversion (rated “4” on a 5-point Likert scale). The participant felt accuracy of the speech-to-text was not good enough with names. The participant felt that accent handling capability was excellent (rated “5” on a 5-point Likert scale), except for handling the names of the people. The participant felt frustrated while using the text mode to communicate with the social network, due to the rubber band on the fingers and the discomfort in using tablets, and preferred speech-to-text. The participant also felt horizontal layout is preferable on the IPTV simulator screen. The participant rated the overall system as completely satisfying, by rating it at “5” on 5-point Likert scale.

Participant 6:

Background: The female participant was aged above 30 and held a job. The participant typically spent 3 to 5 hours on the television set every day and had excellent skills in using handling computers and mobile phones. The participant preferred using internet at home and work. The participant is South-Asian and non-English speaker. However, having spent many years in North-America, the participant has a Canadian accent. The participant used the internet to stay in touch family and friends.

Experiments and technical difficulties: It is important to note that the participant were involved in the study in a passive manner. As the user had to wait for the other participants to respond before carrying out speaking the assigned dialogue, user was also involved talking to people in the room. The participant asked questions in between study. The user had to repeat the sentences many times to the complicated sentences and long sentences. The user tried to focus on the video in the IPTV simulator screen.

First round: Initially, the participant started using the social networking on the tablet using the text mode, which is the first part of the experiment in our comparative study. While using textual mode on the tablet, the participant wore rubber-bands around the fingers to simulate minimum manual dexterity. While using the textual mode on the tablet, participant felt not satisfied (rated “2” on 5-point Likert scale). The participant was fairly satisfied with the error-handling capabilities (rated “3” on 5-point Likert scale). The participant was fairly satisfied with the turn-around time (rated “3” on 5-point Likert scale). The participant was completely satisfied with the learnability of the system (rated “5” on 5-point Likert scale).

Second round: For the second experiment, the participant started using the social networking on the mobile phones using the text mode, in our comparative study. While using textual mode on the mobile phone, the participant wore glasses to simulate low vision. While using the textual mode on the mobile phone, participant felt not satisfied (rated “2” on 5-point Likert scale). The participant was fairly satisfied with the error-handling capabilities (rated “3” on 5-point Likert scale). The participant was fairly satisfied with the turn-around time (rated “3” on 5-point Likert scale). The participant was completely satisfied with the learnability of the system (rated “5” on 5-point Likert scale).

Final round: For the final set of experiment, the participant was asked to use our system prototype which has speech to text capabilities. The participant was instructed at the beginning on the experiment and was video-taped while conducting the experiments. While using the speech mode on the mobile phone, participant felt extremely satisfied (rated “5” on 5-point Likert scale). The participant was fairly satisfied with the error-handling capabilities (rated “3” on 5-point Likert scale). The participant was satisfied with the turn-around time (rated “4” on 5-

point Likert scale). The participant was fairly satisfied with the learnability of the system (rated “3” on 5-point Likert scale).

Feedback/suggestions: The participant felt that understanding the overall goal of the system was very easy, by rating the ease of understanding at “5” on a 5-point Likert scale. The participant also rated the IPTV simulator screen interface as easily understandable and usable (“5” on the Likert scale). However, the participant rated the IPTV simulator remote as “2” on a 5-point Likert scale. The participant felt that IPTV simulator remote interface was confusing and difficult to understand. The participant felt that the speech-to-text was very easy to use, by rating it as “5” on a 5-point Likert scale. The participant felt that speech-to-text software had difficulty while converting the spoken punctuation marks, names of people and depended on the pace of talking. The participant was quiet satisfied with the accuracy of the speech-to-text conversion (rated “4” on a 5-point Likert scale). The participant felt accuracy of the speech-to-text was not good enough with names. The participant also felt that user had to speak at a much slower pace for a better accuracy rate. In addition participant felt, the distance at which the television viewer sat from the television may affect the accuracy of the speech-to-text conversion, if the television had a high volume. The participant felt that accent handling capability was very good (rated “4” on a 5-point Likert scale), except for handling the names of the people. The participant felt frustrated while using the text mode to communicate with the social network and preferred speech-to-text mode to communicate. The participant also felt horizontal layout is preferable on the IPTV simulator screen. The participant rated the overall system as very satisfying, by rating it at “4” on 5-point Likert scale.

APPENDIX D

I. Registering Twitter application in the Twitter developer website

The first step involved in fetching the tweets from the Twitter website involves registering the user application at the Twitter developer's website (https://dev.twitter.com/apps). The developer needs to create a new application on the developer website, by clicking the "create a new application" button. Once the button is clicked the developer is led to another page, where the developer needs to fill in the information such as application name, application description, callback URL, websites, etc. Once the application is registered the developer gets consumer key, consumer secret, access token and access secret token, which can be used for user authentication and retrieving information from Twitter website.

II. Making changes to Web.config file in application

After the consumer key, consumer secret, access token and access secret token are obtained for the application, the Web.config file in the MS Visual Studio web application project is modified to assign values to appropriate keys. Figure D.1 shows the code snippet illustrating the assignment of values in the Web.config file.

```
<appSettings>
      <add key="consumerKey" value="n31XsEwdI8uhTno1XADvw"/>
      <add key="consumerSecret" value="wgX3ARkLzSFWGMbsMOxb7N532Ng0RNi15vaQRs"/>
      <add key="accesstoken" value="158853915-GQ3Fka2WOqbVSm24QT7kGKTcRD0uYLb3u7Ne3qiS"/>
      <add key="accessSecrettoken" value="u6Q2ZwOYBop27Wudn2bfEJvtvxhp4q9doDErMOIrOA"/>
</appSettings>
```

Figure D.1 Code snippet of Web.Config file.

III. User Authentication process and obtaining user status updates

Once the Web.config file is changed appropriately to accommodate the consumer key, consumer secret, access token, and access secret token values, then the C# application (default.aspx.cs) can use the values to fetch status updates from Twitter after the authentication.

As the access token and access token secret are already known, the application can use Twitter API to perform the get and post operation. Figure D.3 shows the code snippet illustrating the user authentication process.

The application uses OAuth for authentication. The OAuth allows the developer to write applications that access the Twitter API, without the need to enter username and password. It has three main benefits. First, the application will continue to work, even if the Twitter login values for user accounts are changed. Secondly, the developer has more control over stopping/resuming the application using Twitter. Thirdly, it is a very secure method as it does not have individual username and password information of the user. The OAuth authentication flow as described in Figure D.2 has the following sequence.

1. The consumer requests the "Request Token", which includes Consumer_key, Consumer_Secret, and a Call back URL. The consumer is the Tweet Fetcher Application. The code snippet in Figure D.3 gives more details on this step.
2. The service provider grants "Response Token" to the consumer. If the request token is null then appropriate error message is shown in the Tweet Fetcher Application. The service provider is Twitter. The code snippet in Figure D.3 gives more details on this step.
3. The consumer requests service provider for status updates of the user in XML format, with Consumer_key, Consumer_Secret, Token and Token_Secret. The

Token and Token_secret are used to authenticate the user and obtain the status updates. The Token and Token secret are used in the application to by-pass the user login into the Twitter account. The code snippet in Figure D.4 gives more details on this step.

4. The service provider grants the request and sends the status updates of the user to the consumer. The code snippet in Figure D.4 gives more details on this step.

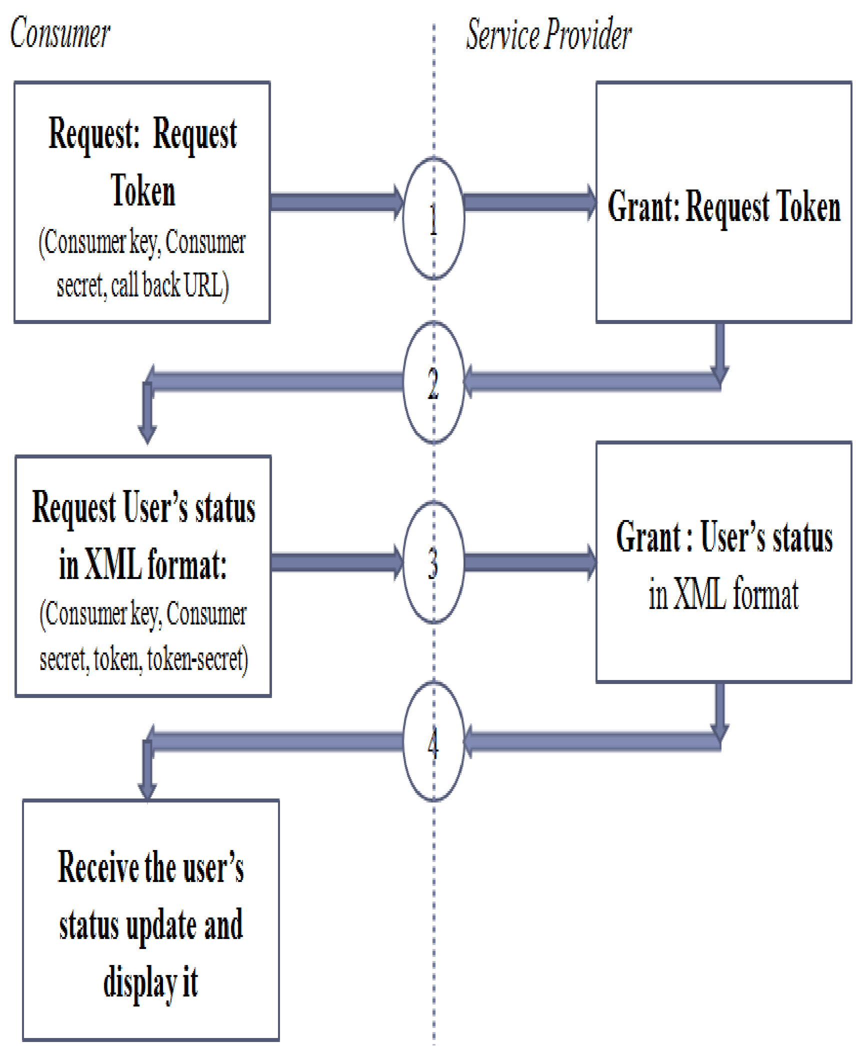

Figure D.2 OAuth Authentication flow

```
// add these to web.config
            _consumerKey = ConfigurationManager.AppSettings["consumerKey"];
            _consumerSecret = ConfigurationManager.AppSettings["consumerSecret"];
            _token = ConfigurationManager.AppSettings["accesstoken"];
            _token_secret = ConfigurationManager.AppSettings["accessSecrettoken"];

            // This callback URL overrides the one set up via Twitter application settings
            var request = GetRequestToken("http://localhost:4235");

      .

      .

      .
private OAuthToken GetRequestToken(string callbackUrl)
{
var requestToken =
FluentTwitter.CreateRequest().Authentication.GetRequestToken(_consumerKey,
_consumerSecret, callbackUrl);

            var response = requestToken.Request();
            var result = response.AsToken();

            if (result == null)
            {
                var error = response.AsError();
                if (error != null)
                {
                    throw new Exception(error.ErrorMessage);
                }
            }

            return result;
 }
```

Figure D.3 Code snippet for authentication and call back URL

We use Tweetsharp in combination with C#. Tweetsharp is a fast and clean wrapper around the Twitter API. It has a well developed set of facilities to get status updates. Using Tweetsharp the user timeline information (status messages) and friend timeline information (status messages) are obtained in an XML format. Figure D.4 illustrates obtaining the user timeline information in XML format. The code snippet in Figure D.4 shows the FluentTwitter request is created using the 4 values of consumer key, consumer secret, token and token secret for authentication, to fetch the 100 status updates using friends-timeline Twitter API, in a XML format. Based on the value of string "s", which contains the username the user1's status message (or) user2's status messages are fetched. The user1 and user2 are the two user accounts used in the proposed system prototype to evaluate the system.

```
var twitterFL = FluentTwitter.CreateRequest().AuthenticateWith(_consumerKey,
_consumerSecret, _token, _token_secret).Statuses().OnFriendsTimeline().Take(100).AsXml();

if (s.Contains("user1"))
{
 twitterFL = FluentTwitter.CreateRequest().AuthenticateWith(_consumerKey, _consumerSecret,
_token, _token_secret).Statuses().OnFriendsTimeline().Take(100).AsXml();

 }
else
 {
  twitterFL =
FluentTwitter.CreateRequest().Statuses().OnUserTimeline().Take(10).For("user2").AsXml();
  }
```

Figure D.4 Code snippet to get user timeline (XML format)

IV. XML Parser and re-builder

The user timeline information is obtained from the Twitter website in an XML format. The file is parsed and rebuilt as a new XML file. The new file has only few filtered values, which are of interest to the developer. Figure D.5 illustrates parsing and rebuilding new XML file. In the code snippet D.5, the response is obtained (for the request made in code snippet D.4) and casted (converted) to type status, using the responseFL.AsStatuses(). A string builder "sb3" is used to construct a new XML format, with the values of status message, user name, picture URL, date-time and source.

```
var responseFL = twitterFL.Request();
                var statuses1 = responseFL.AsStatuses();

                //Try for XML format
                StringBuilder sb3 = new StringBuilder();
                StringBuilder date = new StringBuilder();
                sb3.Append("<?xml version=\"1.0\"?>");
                sb3.Append("<Catalog>");

                foreach (var status in statuses1)
                {
                    sb3.Append("<Status>");

                    sb3.Append("<Status_message>");
                    String temp = status.Text;
                    sb3.Append(temp);
                    sb3.Append("</Status_message>");


                    sb3.Append("<username>");
                    String temp1 = status.User.ScreenName;
                    sb3.Append(temp1);
                    sb3.Append("</username>");


                    sb3.Append("<pictureUrl>");
                    String temp2 = status.User.ProfileImageUrl;
                    sb3.Append(temp2);
                    sb3.Append("</pictureUrl>");


                    sb3.Append("<Date_Time>");
                    String temp3 = status.CreatedDate.ToString();
                    sb3.Append(temp3);
                    sb3.Append("</Date_Time>");


                    sb3.Append("<Source>");
                    String temp4 = status.Source;
                    sb3.Append(temp4);
                    sb3.Append("</Source>");


                    sb3.Append("</Status>");


                }
                sb3.Append("</Catalog>");


                String Sample = sb3.ToString();
                //Response.Write(String.Format("{0} ", Sample));

                if (Sample.Contains("&"))
                    Sample1 = Sample.Replace("&", "&amp;");
                else
                    Sample1 = Sample;
                Response.ContentType = "text/xml";

            Response.Write(String.Format("{0} ", Sample1));

            //End xml format
```

Figure D.5 Code snippet for parsing and rebuilding new XML file

After the XML format is created the data stream is written to string “sample”. The string “sample” is then re-written to new string “sample1”, by replacing all “&” to “&amp”. The symbol “&” was replaced with “&amp” to avoid the XML parser error.

V. INDEX PAGE - MRML CODE

In code snippet (Figure D.6), action “SubmitAction1” or “SubmitAction0” is performed, based on which button is clicked. If “TVButton2” is clicked, then action “SubmitAction1” takes place. If “TVButton3” is clicked, then action “SubmitAction0” takes place. The action “SubmitAction0” and “SubmitAction1” send the appropriate URL and layer-number to IPTV web server based on the user identity. The “TVAction1” in Index page is an invisible action responsible to handle the button click events of “TVButton2” and “TVButton3”.

```
<%@ Page Language="C#" AutoEventWireup="true"  CodeFile="TwitApp.aspx.cs" Inherits="_Default"
%>
<%@ Register Assembly="TVControls" Namespace="Microsoft.TV.TVControls" TagPrefix="mrml" %>
<%@ Register assembly="TVControls" namespace="Microsoft.TV.TVControls.Actions"
TagPrefix="mrml" %>
<%@ Register assembly="TVControls" namespace="Microsoft.TV.TVControls.Animations"
TagPrefix="mrml" %>
<form id="form1" runat="server">
<mrml:TVPage ID="TVPage1" runat="server" style="position:absolute;">
    <mrml:TVButton ID="TVButton2" runat="server"
    style="position:absolute; top: 397px; left: 477px;" Text="Afshana"
        onclick="SubmitAction1">
    </mrml:TVButton>
    <mrml:TVButton ID="TVButton3" runat="server" onclick="SubmitAction0"
        style="position: absolute; top: 438px; left: 477px;" Text="Hashim">
    </mrml:TVButton>
    <mrml:TVActions ID="TVActions1" runat="server" Actions-Capacity="4"
        style="position: absolute; top: 33; left: 36;" Version="8">
        <mrml:SubmitAction Gadgets="TVButton3" Name="SubmitAction0"
            Url="MediaroomPage.aspx?__MPFLayer=0/1" X="0" Y="0" />
        <mrml:SubmitAction Gadgets="TVButton2" Name="SubmitAction1"
            Url="MediaroomPage.aspx?__MPFLayer=0/1" X="0" Y="0" />
    </mrml:TVActions>
</mrml:TVPage>
</form>
```

Figure D.6 Code snippet of MRML code for the index page design.

VI. USER PROFILE PAGE - MRML CODE FOR BOUND LIST

The Figure D.7 illustrates the MRML code for the bound list object in the User Profile page. The TVBoundList object is a list, which acts as a place holder to display the user tweets. The list has a capacity to display 33 tweets in horizontal display layout. The horizontal display layout defines tweets scrolling from left to right. Each element in the list holds exactly one tweet. At any given moment, exactly one user tweet is displayed on the simulator screen with a facility to scroll to the next tweet on the right (or) left.

```
<mrml:TVBoundList ID="TVBoundList1" runat="server"
        style="position:absolute; top: 369px; left: 1px;
height: 104px; width: 634px;"
        VisibleItemCount="1"
        DataBinder="XmlDataBinder,TVXmlDataSource1,//
Catalog" TotalCountPath="33"
        FocusCanLeave="True" AutoFlow="Horizontal"
IsVisible="False">
</mrml:TVBoundList>
```

Figure D.7 Code snippet for the bound list in the User Profile page

VII. USER PROFILE PAGE - MRML CODE FOR TV LIST ITEM TEMPLATE

Each element in the above bound list is contained by another TV list item template object “TvListItemTemplate”. The “TvListItemTemplate” holds the information such as: status message, username; picture URL, date_time and source information. Figure D.8 shows the code snippet for the TVListItemTemplate and sub-elements of TVListItemTemplate. The MRML code has values defining screen position and text for the TVLabels such as status message, username, date_time and source. The MRML code also has values defining the screen position for the TVImage such as pictureUrl.

```
<mrml:TVListItemTemplate ID="TVListItemTemplate1" runat="server"

            style="position: absolute; top: 0px; left: 0px; height: 101px;
width: 632px;"
            DataBinder="TemplateName:Status;ValuePath:;">
            <mrml:TVLabel ID="TVLabel2" runat="server"
DataPath="Status_message"

                style="position:absolute; top: 50px; left: 6px; height:
44px; width: 622px; bottom: 68px;"
                Text="TVLabel" WhiteSpace="Preserve">
            </mrml:TVLabel>
            <mrml:TVLabel ID="TVLabel3" runat="server"
                style="position: absolute; top: 8px; left: 60px; width:
134px;" Text="TVLabel"
                DataPath="username">
            </mrml:TVLabel>
            <mrml:TVImage ID="TVImage1" runat="server"
DataPath="pictureUrl"
                style="position: absolute; right: 582px;" >
            </mrml:TVImage>
            <mrml:TVLabel ID="TVLabel4" runat="server" DataPath="Date_Time"
                FontStyle="bold16"
                style="position: absolute; top: 8px; left: 199px; width:
177px; font-size: xx-small; height: 25px; right: 267px;"
                Text="TVLabel">
            </mrml:TVLabel>
            <mrml:TVLabel ID="TVLabel5" runat="server" DataPath="Source"
FontStyle="bold16"
                style="position: absolute; top: 7px; left: 380px; width:
245px;"

                Text="TVLabel">
            </mrml:TVLabel>
        </mrml:TVListItemTemplate>
```

Figure D.8 Code snippet for the TVListItemTemplate.

VIII. USER PROFILE PAGE - MRML CODE FOR XML DATA SOURCE

The website “TwitApp” from the Twitter Application Server is used as the XML data source in the Mediaroom IPTV to display the user status on User Profile page, which is described by code snippet of MRML code in Figure D.9. The content will be displayed in the User Profile page of the IPTV simulator.

```
<mrml:TVXmlDataSource ID="TVXmlDataSource1" runat="server"
        style="position: absolute; top: 102px; left: 15px;
height: 26px; width: 216px;"
        Url="http://sazedul.usask.ca/TwitApp/"
onready="TimeoutAction0"
        RequestData=" TVText1">
    </mrml:TVXmlDataSource>
```

Figure D.9 Code snippet for the XML data source from Twitter Application server

IX. USER PROFILE PAGE - MRML CODE FOR TV ACTIONS AND COMPOSITE ACTIONS.

The Figure D.10 illustrates the code snippet about TV actions and composite actions in User Profile page. The TV action “TVAction1” is invisible composite action to perform the following actions.

1. Handle the focus event by moving the focus from one TVButton to another TVButton.
2. Handle showing/ hiding tweet messages.
3. Handle showing/ hiding Twitter icon.
4. Composite action to carry out multiple actions on single events.

```
<mrml:TVActions ID="TVActions1" runat="server" Actions-Capacity="8"
        style="position: absolute; top: 42; left: 37;" Version="53">
        <mrml:NavigateAction Data="back" Name="NavigateAction0" X="0" Y="0" />
        <mrml:RefreshAction Name="RefreshAction0" Target="TVXmlDataSource1" X="0"
            Y="0" />
        <mrml:SubmitAction Gadgets="" Name="SubmitAction0" Url="TwitApp.aspx" X="0"
            Y="0" />
        <mrml:TimeoutAction Action="RefreshAction0" Delay="5000"
            Name="TimeoutAction0" X="0" Y="0" />
        <mrml:CompositeAction Name="CompositeAction0" X="0" Y="0">
            <mrml:FocusAction Name="FocusAction0" Target="TVButton2" X="0" Y="0" />
            <mrml:HideAction Name="HideAction0" Targets="TVButton1 TVButton5" X="0" Y="0" />
            <mrml:ShowAction Name="ShowAction0" Targets="TVButton2 TVButton4" X="0" Y="0" />
            <mrml:ShowAction Name="ShowAction2" Targets="TVBoundList1" X="0" Y="0" />
        </mrml:CompositeAction>
        <mrml:CompositeAction Name="CompositeAction1" X="0" Y="0">
            <mrml:FocusAction Name="FocusAction1" Target="TVButton1" X="0" Y="0" />
            <mrml:HideAction Name="HideAction1" Targets="TVButton2 TVBoundList1 TVButton4"
                X="0" Y="0" />
            <mrml:ShowAction Name="ShowAction1" Targets="TVButton1 TVButton5" X="0" Y="0" />
        </mrml:CompositeAction>
    </mrml:TVActions>
```

Figure D.10 Code snippets for the TV actions and composite actions.

The TVAction1 also handles the page refreshes. In composite action 0, the Twitter icon and back button show up above the status bar. In composite action 1, the Twitter icon and back button show up below the status bar.