Analysis of Embodied Conversational Agents in SecondLife for Speech Recognition

Except where reference is made to the work of others, the work described in this thesis is my own or was done in collaboration with my advisory committee. This thesis does not include proprietary or classified information.

Wanda R. Moses

Certificate of Approval:

Cheryl D. Seals Associate Professor Computer Science and Software Engineering	Juan E. Gilbert, Chair Professor Computer Science and Software Engineering
Ivan E. Watts Associate Professor Educational Foundations Leadership and Technology	George T. Flowers Dean Graduate School

Analysis of Embodied Conversational Agents in SecondLife for Speech Recognition

Wanda R. Moses

A Thesis

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Master of Science

Auburn, Alabama
December 18, 2009

Analysis of Embodied Conversational Agents in SecondLife for Speech Recognition

Wanda R. Moses

_____

Signature of Author

_____

Date of Graduation

Vita

Wanda Moses is a PhD student in the Computer Science and Software Engineering Department at Auburn University. She was born in Charleston, SC on May 2, 1962 to Annie M. and David E. Moses. Ms. Moses received a Bachelor of Science degree in Mathematics and Computer Science from South Carolina State University in May 2005. She is currently a graduate research assistant in the Human Centered Computer Lab at Auburn University. Her interests are in Human Computer Interaction, User Interface Design, Adaptive Learning Technologies, Multimodal Interfaces and Spoken Language Systems. She is a member of the Association of Computing Machinery (ACM) and Delta Sigma Theta Sorority, Inc. Ms. Moses has one daughter, Tashawnna L. Lemons and four grandchildren, T'Andreus J. N. Lemons, Kiara D. Lemons, Jadon L. Lemons and Aaliyah S. N. Lemons.

Thesis Abstract

Analysis of Embodied Conversational Agents in SecondLife for Speech Recognition

Wanda R. Moses

Master of Science, December 18, 2009
(B.S., South Carolina State University, 2005)

47 Typed Pages

Directed by Juan E. Gilbert

A virtual world is a computer simulated environment that's interactive and allows the user to become a participant in the computational space, entering a computer world that they have created. This participation gives the user the sense of presence in the environment. In a virtual world people can select or create an animated alter ego, known as an avatar. The world can be inhabited by a single avatar or simultaneously by many creating a virtual community. These virtual worlds and communities are created and used for many different reasons such as aviation and medical training, entertainment, socializing, commerce and marketing, mentoring, and educational purposes, just to name a few. When users decide to use a particular virtual world, one of their main concerns is the quality of various attributes of the software. Some of these attributes are functionality, usability, efficiency, reliability, maintainability and sometimes portability. The focus of this thesis is on a specific virtual world application, Second Life, one of the top 10 online three-dimensional virtual worlds. This study is an intensive descriptive analysis of Second Life's virtual agents and the capability of incorporating speech recognition into Second Life.

Acknowledgments

Giving thanks and praises to my God, God the Father, God the Son, Jesus the Christ, and God the Holy Spirit for guiding and keeping me throughout this journey. I would like to express my deepest appreciation to my advisor, Dr. Juan E. Gilbert, for the opportunity to continue my education and the challenge he issues to always think outside the box and to go above and beyond what's required. To Dr. Cheryl D. Seals, my Soror and encourager, I am much obliged. Dr. Homer W. Carlisle, it's been great, thank you. To my family and friends, thank you for the encouraging words, prayers and belief in me as I continued to pursue my educational goals. To my Sistahgurl, prayer partner and most confident supporter, Sherald Moses, I love you and cannot express in words my appreciation. To my benefactor, most avid supporter, and little brother, David Moses, thanks a million for everything, including the laptop. For my number 1 fan, the love of my life, my Mamanem, Annie M. Moses, "Never would have made it without you and God, of course". My deepest gratitude goes to my daily, wake up inspiration and oldest sister, Essie Claire, you know I love you and already miss our morning chats. I would like to issue a challenge to each of my wonderfully unique grandchildren, T'Andreus Joi' NaShae, Kiara D'Avianne, Jadon La'Quan and Aaliyah Shantae' Nicole Lemons, to join me in achieving at least this level of education. To the members of the HCCL at AU and other fellow graduate students, I am forever grateful for your advice, support, help, encouragement, etc., etc. Last but not least, to my biggest fan, my best inspiration, my greatest joy, my lovely daughter, my friend, my heartbeat, the awesome mother of my beautiful grandchildren, Tashawnna La'Vonshell Lemons, "Thank you for allowing me to take this time to 'do me' even though it took away from 'our time'. I love you!!!" To each and every one of you, "Be blessed, in Jesus' name!"

Style manual or journal used Journal of Approximation Theory (together with the style known as "aums"). Bibliograpy follows van Leunen's *A Handbook for Scholars.*

Computer software used The document preparation package TEX (specifically LATEX) together with the departmental style-file `aums.sty`.

Table of Contents

List of Figures

Chapter 1

INTRODUCTION AND BACKGROUND

1.1 Introduction

In the midst of the many emerging technologies you will find networked, three-dimensional virtual worlds, which is the combination of a desktop type virtual reality and a chat room type environment. The three key features of a three-dimensional virtual world are: avatars, that are physical representations of the user within the virtual world; an interactive environment where the various users communicate with each other; and an environment that gives the illusion of a three-dimensional space. There are numerous three-dimensional virtual world applications available today at varying degrees of functionality and pricing. The top 10 most popular of these according to Second Life Update.com are [24]:

1. An updated version of SecondLife that includes voice chat.
2. An updated version of World of Warcraft, a multiplayer role playing game that was created by Blizzard Entertainment in 2001.
3. IMVU, created by IMVU Inc. in 2004 is a graphical instant messaging client and has 20 million registered users.
4. Kaneva, created by Kaneva, Inc. in 2004, allows you to easily combine video, 2D, and social networks.
5. There.com, created by There Inc. in 2001 where MTV Networks run Laguna Beach Style virtual world.
6. Active Worlds, created by Active Worlds in 1997, is a virtual reality 3D world.
7. Meet Me, created by Transcosmos Inc., Japanese version, based on Tokyo Japan and G rated with lots of rules.
8. HiPiHi, created in China and has heavy government censorship.

9. A World of My Own (AWOMO), created by Sir Richard Branson in 2007, with Virgin Games as a 20% stake holder.

10. Moove, created in 1994, German based with emphasis on 3D chat and dating in rooms.

These and other virtual worlds are utilized for many different reasons, by ages ranging from pre-school to senior citizens, and provide varying degrees of functionality. In virtual worlds everything is manufactured. You can create just about whatever you can imagine and then experience being a part of what you created. It has programmable objects that have properties that can be changed. When objects are created in a virtual world they are tagged. This allows the original designer to decide whether she wants to let others make a copy of the object free of charge, for a fee, or not at all. Most virtual worlds have an area where free objects can be created and obtained by others. There are no bounds other than the functionality of the system and your creative imagination. In virtual worlds you are able to use the basic skills that you have been taught since birth. These inherent skills allow you to use gestures, communicate, move about and manipulate objects in a natural way. Instead of using the mouse to click or draw lines, the semantics should be as natural as pointing your finger, talking and/or grasping an object. When wanting to change the location of objects around you, you should not have to reverse, backspace or undo. You should be able to turn around, reach out and change or move an object to where you want it to be. A child should be able to use the system. In a virtual world, you are not bound by physical restraints, only by the system's functionality and the limits of our own imaginations. A person can encounter many unique experiences within virtual worlds that they may never have the opportunity to experience in real life. In a virtual world, they do not need to know all of the technicalities such as how the operating system works or how to program in a certain programming language. All that's required is basic thinking, minimal skills and an accurate mental model of the system. Although virtual worlds are becoming a very popular medium for educational purposes, many of them are deemed "not conducive to in class use" because the user's typing ability and experience plays a major role in their success rate

and the frustration level of the user's response. The ability to create an environment that uses voice to text or speech recognition within virtual worlds would eliminate this major drawback. With this in mind, talking would be more of a natural means of communication than the current method of texting in most virtual worlds. This thesis analyzes and details the history of virtual reality, virtual worlds and virtual communities. They will all be defined in the next section. It also provides background information on Second Life, its virtual agents and their ability to incorporate speech recognition.

1.2 Background

Virtual Reality is a simulation of a world or an environment in which a user can interact, using hardware or software, with three-dimensional objects that seem realistic. Bricken, W. states that "virtual reality is a natural interface with abstractions". He predicts that although no one at that time, 1990, knew the impact of extended exposure to virtual reality, both positive and negative, it would be commonplace by 2010 [4].

Virtual World(VW) is defined best as "a synchronous, persistent network of people, represented as avatars, facilitated by networked computers" [3]. VW is a computer simulated environment that's interactive and allows the user to become a participant in the computational space. Virtual worlds are both intuitive and experiential. A user is immersed within a multi-sensory environment in which many experience a profound sense of being ensconced in the virtual world.

A Virtual Community is a group of people that communicate, collaborate, connect, share information, network and/or interact in an online environment. According to John Hagel and Arthur Armstrong in their book, Net Gain, the virtual community provides its participants with four basic needs:

- The main purpose of a transaction oriented community is the sale and purchase of products and services, and the delivery of any information that will bring about the accomplishment of these activities (i.e. Amazon and eBay).

- An interest oriented community is characterized based on a significant amount of content generated by the members. In this type of community there are more social interactions than in the transaction oriented community (i.e. FaceBook and MySpace).

- A fantasy oriented community allows its members to create new personalities, stories and fantasy environments where they can interact with other members by sending messages or text chatting (i.e. Second Life and WorldofWarCraft).

- A relationship oriented community is built based on the members' important or significant life experiences and usually leads to personal relationships between the members.

and the virtual community has five major characteristics:

- Information about a particular subject that is of common interest, whether it's generated by the members or published by the VC, brings members together.

- Integration of content and communication is accomplished via forums, chat rooms and email.

- There is great deal of experiences, comparative insights, perspectives and collective expertise that could not be matched by any one expert.

- VC allows members access to a broader range of suppliers and products, along with the ability to negotiate with them, because of the competiveness.

- By providing members with valuable resources and environments such as Amazon and eBay, VCs can make a profit and increase their power and popularity.

A lot of times the different types of virtual community characteristics overlap [1].

Chapter 2

HISTORY

(From Virtual Reality to Virtual World)

2.1 Virtual Reality and Virtual World Timeline

1962 - Sensorama by Pliny the Elder, Gaius Plinius Secundus

1968 - Virtual Reality device by Ivan Sutherland using heavy, bulky headsets, wire-frames and other sensory input simulations.

1974 - Maze War, first three-dimensional shooter game that involves multi-users was the first networked game and was represented by eyeballs within a maze.

1978 - MUD, text-based, multi-user online game that used a TELNET program.

1990 - HMD, Data Gloves

1992 - CAVE, Cave Automatic Virtual Environment, is a surround-screen, surround-sound, projection-based virtual reality system, a cube that's about the size of a room.

1995 - Active Worlds, one of the oldest and most dynamic 3D virtual world applications available online today.

1996 - OnLive! Traveler, the first 3D virtual world technology that supports multi-user, synchronous, audio communication.

1999 - Whyville, launched by Numedeon Inc.

2003 - Second Life, by Linden Lab.

2.2 The Sensorama, 1962

The idea of virtual worlds can be linked to Pliny the Elder, an important naval commander, author and philosopher. Figure 2.1 shows the Sensorama, mechanical-based, that was built in 1962. It's one of the earliest known immersive technologies. It was a machine that used many of the senses (today this concept is called multi-modal). It used vision, sound, balance, smells and touch to simulate its world. Virtual worlds evolved from generic virtual reality simulators and were fueled by the gaming industry [33].

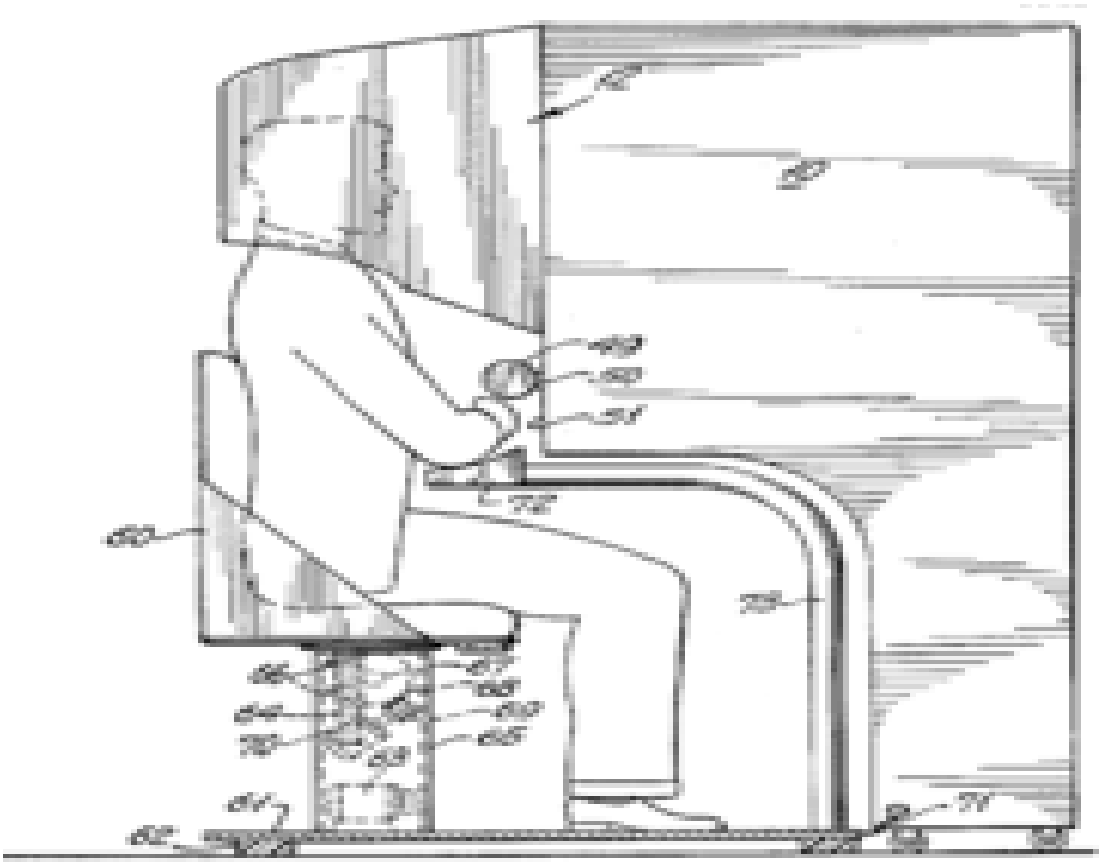

Figure 2.1: The Sensorama, from U.S. Patent #3050870

2.3 Maze War, 1974

Maze War, the first three-dimensional shooter game that involves multi-users was the first networked game and was represented by eyeball "avatars" within a maze, celebrated their 30th Year Retrospective on November 7th, 2004 at the Vintage Computer Festival in Mountain View, CA. It was first played at the NASA Ames Research Center in California using Imlac's PDS-1 in 1974. Figure 2.2 shows a picture of early documentation of the Maze War game and a picture of the Maze War game running on a Xerox workstation in 2002 [14].

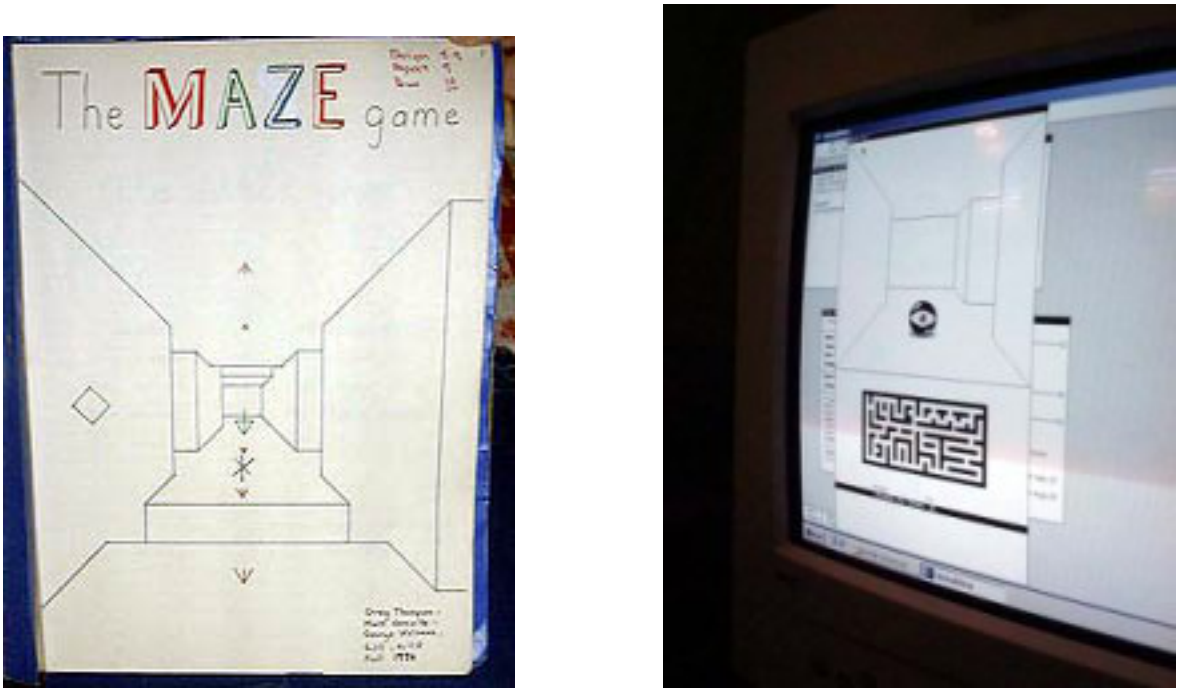

Figure 2.2: Documentation of Maze War in the early stages (left), and Maze War running on a Xerox workstation in 2002 (right).

2.4 MUD, 1978

The MUD, Multi-User Dungeon, was a text-based, multi-user, online game that used a TELNET program that allowed you to follow a link in order to play the game. It was not 3D, but could be played on any computer. It is the world's oldest virtual world. MUD was developed at Essex University in England in 1978, see Figure 2.3. Users had to read to find out about the rooms, players, objects and actions that were being performed within the virtual world. They communicated by typing commands that resembled natural language. Many of the virtual worlds today still communicate using text. Second Life and other social virtual worlds have been said to have originated with what was known as a graphical MUD [16].

Figure 2.3: MUD, Multi-User Dungeon

2.5 A CAVE, 1992

A CAVE, Cave Automatic Virtual Environment, is a surround-screen, surround-sound, projection-based virtual reality system, a cube that's about the size of a room. The user would wear special glasses when she enters the cave. This will allow her to see the special affects of the three-dimensional graphics generated within the cave by projecting the images onto mirrors which in turn project them to projection screens. The caves use electromagnetic sensors that track the movement of the user and are made of non-magnetic stainless steel to cause a minimal amount of interference with the sensors. The user's body movements actually control the projectors that are located outside of the cave and driven by computer(s). The positioning of the mirrors and projectors in a CAVE can be seen in Figure 2.4 [11].

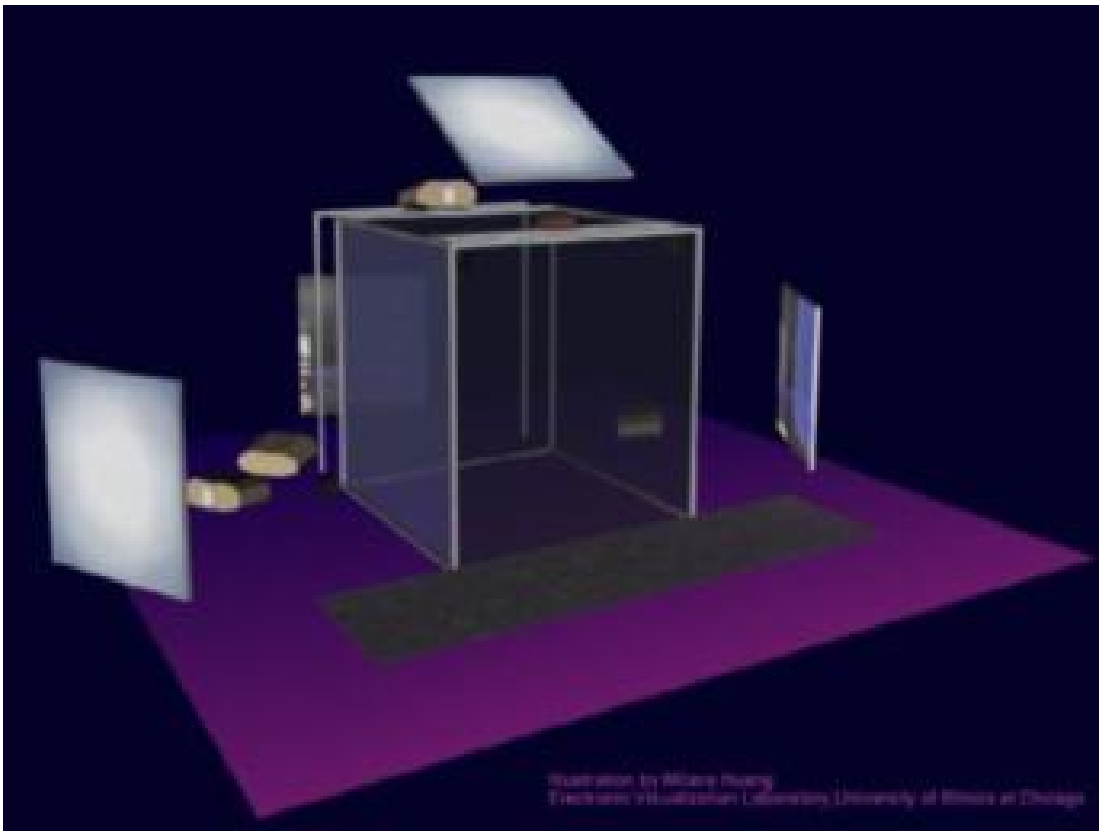

Figure 2.4: Positioning of the mirrors and projectors in a CAVE

2.6 Virtual Reality Hardware, 1990

There are a number of changes that define the paradigm shift to virtual reality technology where the user actually becomes a part of the environment. Symbol processing shifts to reality generation; viewing a monitor is replaced by wearing a computer; symbolic becomes experiential; the observer is now a participant; the interface is inclusion; physical objects are programmable; instead of only using visual sense interaction is now multimodal, using many senses; and metaphors are virtualities [4].

In virtual reality a user actually becomes a participant in the computational space by placing himself within the image. He can do this by wearing some type of hardware, Head Mounted Device(HMD), that can determine his behavior and from this, display things from his perspective [4]. The HMDs were heavy and bulky, some without audio capacity and all have cables which restrict movement. Some other interface devices used to interact with elements in the virtual world include the Polhemus Isotrack, which generates a small electro-magnetic field and tracks the movement of receptors within it; the Mandala which is a passive video tracking is used in artificial reality systems; 6-D control devices such as the Data Glove, the Vertex Glove, the Bird, and the Space Ball. Figure 2.5 shows a user wearing the Data Gloves and an HMD [30].

Figure 2.5: User wearing a Head Mounted Device and Data Gloves

2.7 Active Worlds, 1995

Active Worlds is one of the oldest and most dynamic 3D, online virtual world applications [7]. It is text based. The primary means of communication is chatting via text. Although English is the default language in Active Worlds, the browser supports other languages such as Danish, Dutch, Finnish, French, Italian, Norwegian, Portuguese and Spanish. Other languages such as Russian and Japanese are also used in Active Worlds even though they are not supported within the browser.

Active World users can be tourists or for a yearly fee of $19.95, a citizen. Citizens are allowed to select an avatar from an existing pool of avatars that include humans and others such as a bird and an alien. These avatars can not go from one world to another. Each citizen is assigned a unique identity that allows them to be contacted by other citizens, no matter what world they are in. Users can communicate with other citizens by whispering, they must be near each other, or sending short telegrams, they can be in different worlds or offline. Messages can be delayed due to distance between users, bandwith or the user's typing skills. Trying to cypher through text-based conversations can be very confusing. Figure 2.6 shows an example of text-based conversations between three users [7].

There are capabilities that allow users a 1st person view of the world or a 3rd person's view where they can also view themselves within the world. It has the capabilities to allow users to add and build in existing extensible worlds [7].

Active Worlds is very popular in the education arena. There are over a hundred individually owned worlds in the Active Worlds Educational Universe (AWEDU), that was created in 1999. An AWEDU browser is shown in Figure 2.7. Users are limited to choosing their avatars from a pre-selective group [7].

Immigration Officer: Still under construction
MD: great background
Michigan: hello MD
Lindbergh: tank yew
Michigan: Lindbergh made the Backdrop
Lindbergh: actually
Lindbergh: I think that was Fryedds work
MD: really? What did you use. they're great
Lindbergh: I just modified it, for something new
Lindbergh: umm
Michigan: he used PSP5
Lindbergh: paint shop pro 5.0
Lindbergh: I like photoshop, but
MD: really? paint shop?
Lindbergh: this was pretty cool
Michigan: Change yor avatar MD to Static Lady
Lindbergh: yeah, lots of new stuff in it
Michigan: cool
MD: did you two create most of these?
Lindbergh: yeah
Michigan: yes
MD: this is beautiful.
Michigan: everything is original

Figure 2.6: Text-based Conversations Between Three Users in Active Worlds

Figure 2.7: The Active Worlds Educational Universe Browser

2.8 Onlive! Traveler, 1996

In 1996, OnLive! Traveler came into being. It is the first 3D virtual world technology that supports multi-user, synchronous, audio communication [7]. This is accomplished by using microphones to speak with other users in real-time within a chat room environment. Users that have an avatar within hearing range will be able to hear the conversations of other users. In order to have a private conversation, the users must go to a more private area, a personal space, or use a text chat box. The avatars are heads that can be customized by the user, however the user is not allowed to build or modify anything within the world. Only 20 users are allowed to occupy a particular world at the same time. The "talking heads" have real life-like facial expression capabilities such as lip-syncing and blinking eyes that "provide a provocative if not eerie sense of embodied presence in the 3D environment" [7]. Figure 2.8 shows the Onlive! Traveler browser.

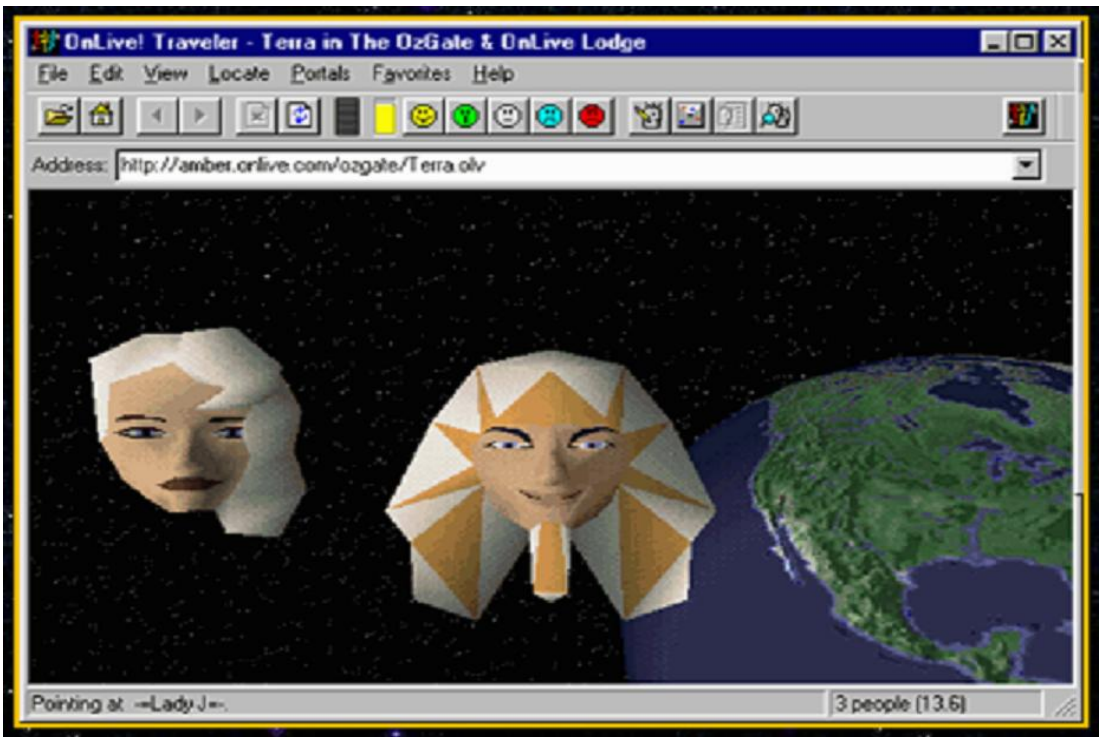

Figure 2.8: Onlive! Traveler Browser

2.9 Whyville, 1999

Whyville, launched in 1999, by Numedeon Inc., is a virtual world designed to support inquiry-based constructivist user centered self-service learning, with a strong STEM (Science, Technology, Engineering and Mathematics) base. The avatars are constructed by the user. This 3D virtual world technology is targeted toward children ages 8 to 14, with over 65% of the population being female. Their learning projects are strongly social and simulation based. It is a free browser-based technology that requires minimal network connectivity[31]. Figure 2.9 shows the Whyville browser.

Figure 2.9: Whyville Browser

2.10 Second Life, 2003

Second Life is an internet-based virtual world environment and was launched on June 23, 2003. Figure 2.10 shows the world map of Second Life as it was in 2002. Discussed more in next chapter.

Figure 2.10: Second Life 2002

Chapter 3

LITERATURE REVIEW

3.1 Second Life

Second Life, originally named LindenWorld, was developed by Linden Research Inc. Philip Rosedale, former Linden Lab CEO, is the founder [25]. It is an Internet-based virtual world environment and was launched on June 23, 2003. It gained international acclaim through media outlets in late 2006, early 2007 timeframe due to its use for emulating virtually anything in real life [25]. Second Life has millions of residents who are represented as avatars, from all walks of life, who can interact with each other through the Second Life viewer, which provides a social network service combined with general aspects of meta-verse. The software, that was designed using C++, consists of the client(viewer) that runs on the resident's computer and thousands of servers that are maintained by Linden Labs [17]. New versions of both being deployed on a regular basis to continually improve performance, usability and security.

Second Life has been adopted as a platform for education activities by many institutions, such as colleges, universities, libraries and government entities that have fully ascribed to it [25]. It has also been chosen as the preferred platform for many other activities such as socializing, networking, commerce and marketing, fantasizing, mentoring, aviation training, advertising, and the list goes on. There are businesses that use Second Life for project managing, using private spaces to conduct meetings with managers and representatives worldwide. Just like in real life, Second Life participants, avatars, also experience adverse conditions, such as trespassers, disputes with other residents, thieves, and other abuse.

Second Life uses Linden Dollars (L$) for currency. The exchange rate when buying Linden Dollars is $0.30 US for one (1) L$ and when selling Linden Dollars the exchange

rate is 3.5%. The Linden Dollar is used in Second Life to purchase land, purchase or rent an Estate, shop at the various businesses, buy objects, etc. The steadily increasing number of residents forces Linden Lab to create additional land in Second Life on a regular basis. Selling land in Second Life is similar to selling real estate in the real world. Land owners are charged a maintenance fee, called a tier, that's required to preserve it just as some real estate owners are charged property taxes in the real world to retain their property. The amount of the tier fees are based on "the peak amount of land held during your previous 30 day billing cycle. This includes land parcels held and land tiers donated to groups. The fee is tiered and discounted as you acquire more land. Peak usage is measured by the maximum amount of land you held – for any length of time – during your billing cycle."[12]

Just about everything in Second Life is created by the residents. To enhance your fantasy, you are allowed to design your clothes, your house, tint your skin, style and color your hair, this list is continuous. The Linden Scripting Language (LSL) is the programming language the members use to modify the objects that they create. It also allows them to add a behavior to an object. Every LSL program must have at least one state, the default state where the LSL program will begin excution, some programs have multiple states. Figure 3.1 shows the scripting code that parallels the familiar "Hello World" program segment. The "llSay" command selects a channel and allows the script to chat text to that channel the same way an avatar would.

Second Life also has what they have labeled as "Second Life URLs (SLURL)", Second Life Uniform Resource Locator. The SLURL allows you to link to Second Life from other websites or from within the Second Life viewer. When you click on the SLURL you will be automatically transported that location within Second Life. Second Life also have a SLurlBuilder. It's a graphical user interface that allows you to generate a SLURL by inputing the Sim, a virtual 65536 meters squared area of Second Life land with a unique name, and the x, y, and z coordinates for that sim [27].

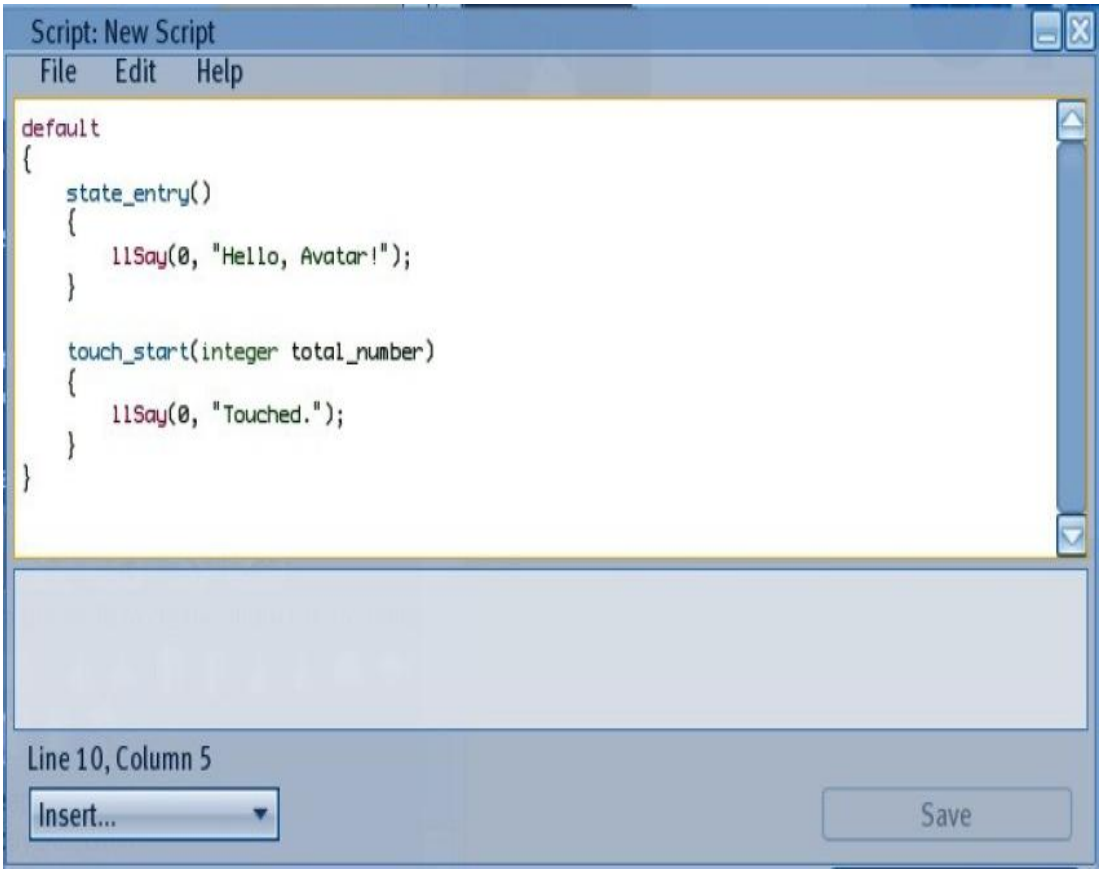

Figure 3.1: Linden Scripting Language for "Hello, Avatar!"

3.1.1 Second Life Timeline

1999 - Linden Lab was founded [9]

2001 - Work begins on Second Life (internal name LindenWorld) [9]

2002 - 1st Resident joined Second Life (Steller Sunshine) [9]

2002 - Public beta started [9]

2003 - Second Life released (No currency and no teleporting) [9]

2005 - Teen Second Life opened for 13-17 year old residents during office hours [32]

2006 - Teen Second Life opened 24/7 [32]

2006 - SL Resident, Anshe Chung, 1st Real Life millionaire of SL business, on cover of U.S. Business World magazine [9]

2008 - Second Life 1.19.1 (with windlight rendering) released [WindLight is the code

name for our “physically-accurate atmospheric rendering & lighting” project] [34]

2008 - Mark Kingdon (M Linden) announced as new CEO [20]

2008 - Open Grid Public Beta Begins [23]

2008 - Mono launched (part of 1.24 Server deploy) [22]

2009 - Announced acquisition of OnRez and XStreet SL, online marketplaces [26]

3.1.2 Pricing

In 1990 a very small network of basic virtual worlds could cost up to a quarter of a million dollars. With today's technology, there is practically an unlimited amount of low to no cost access to internationally networked virtual worlds. Second Life has three account types; basic, premium and concierge:

- The Basic account, which is free, however you cannot own land on the mainland (available via auctions at varying prices). You are allowed to rent anywhere (depends on the landowner), own Estates or Private Regions ($1000.00 per region). You cannot participate in Live Chat. There is no support offered other than what is available to everyone via frequently asked questions. You are allowed building, scripting, shopping and access to various events.

- The Premium account has a $9.95 monthly fee, or a monthly fee as low as $6.00 per month for annual accounts (Appendix A). It allows you to own land on the mainland, you can submit support tickets to get help from the support team and you can initiate Live Chat sessions.

- The Concierge account is any account that owns an Estate, has paid for an Estate or pays the tier for greater than half of a region's worth of mainland parcel. They would be able to access a special Live Chat area phone number that others are not allowed to see.

3.1.3 Navigation Methods

The primary method for avatars to navigate and move around in Second Life is by foot, either walking or running. To get around more quickly, they can fly about 170 meters above ground level. Scripting attachments enable them to fly higher. Just as in the real world, avatars can ride in vehicles. There are go-karts that can be obtained from the object library, vehicles created by residents, vehicles made available from auto dealers (Some car dealerships provide models of their cars to the members of Second Life.), helicopters, submarines and hot-air balloons. The maximum altitude for any object within Second Life is 4096 meters high. Another popular method of traveling is teleporting which allows an avatar to travel instantly from one location to another. This can be done by creating a link using the Linden scripting language that includes the coordinates of the location to where you want to teleport, using the world map and simply clicking on your desired location or conducting a search of where you want to go using the search window and clicking on the teleport button.

3.1.4 Current Communication Methods

The primary means of communicating in Second Life is text chatting. You can whisper if you are within 10 meters of the avatar you are chatting with. You can speak at a normal volume if you are within 20 meters of the listener or you can shout to anyone within 100 meters of your avatar. One of the downfalls with many VWs is that they do not allow the complexity and variety of non-verbal communications that is afforded in a face-to-face setting. This is discussed more in the Embodied Conversational Agents section.

You can have private chats via instant messenger (IM) with an individual or with a group. The resident does not have to be online in order for you to IM them. As long as you have their "calling card", you can IM them.

You can also use Voice in Second Life. You must have the necessary hardware on your computer. When you are a voice enabled resident and on a voice-enabled parcel of land, you can talk with other voice enabled residents. You can recognize a voice enabled resident

by a voice intensity indicator that can be seen above the resident's head. The indicator changes color and size with the level of the volume of the speaker's voice or how close they are to the microphone.

In the push-to-talk- mode (default setting), you use voice service the same way you would use a walkie talkie, by designating a button and pushing the button to talk or you can use the button on the bottom of the screen. If you decide that you do not want to use this mode, you can turn this mode off, however your microphone will remain on at all times [21].

Using voice, you can talk to an individual or talk to a group, but you cannot do both at one time.

3.2 Automated Speech Recognition

Automated speech recognition systems allows users to speak in a natural way rather than type in information on a keyboard or keypad. Speech recognition actually recognizes words and should not be confused with speech detection, which is a technique used to determine the presence of human speech or the absence thereof in areas that may include other sounds or voice recognition, which makes a distinction between voices. With speech recognition, the automated system listens to words that are spoken then converts those speech signals to machine readable words. This process is based on the Hidden Markov Model (HMM), where a sequence of states are used to build words. [15] [2] [36] [35] [8]. There are many speech recognition systems available on the market today. Quite a few of them uses the HMM. For this analysis, a particular speech recognition system is not necessary.

3.2.1 jVoiceBridge

Sun Microsystems Laboratories sponsors an experimental technology, jVoiceBridge, a software-only audio mixer written entirely in Java. jVoiceBridge handles Voice over IP (VoIP) audio communications that can be used in virtual worlds and for speech detection.

The ability for telephone support is added to the virtual world using jVoiceBridge. You have to be connected to the bridge to communicate using jVoiceBridge such as during a conference call. Project Wonderland is a 3D virtual environment that uses jVoiceBridge. It uses standard Session Initiation Protocol (SIP) communication channel to send signaling data and Real-time Transport Protocol (RTP) for audio data. Each avatar and recorded sound source within the world have adjustable audio channels. jVoiceBridge uses the softphone, a client-side application, that connects to the voice bridge, a server application. The voice bridge mixes audio for different uses and can work with any software phone that's SIP-based. Figure 3.2 shows jVoiceBrigde's architecture using SIP and RTP protocols to transmit voice data from server to client. Figure 3.3 shows an overall picture of communications within project Wonderland [10].

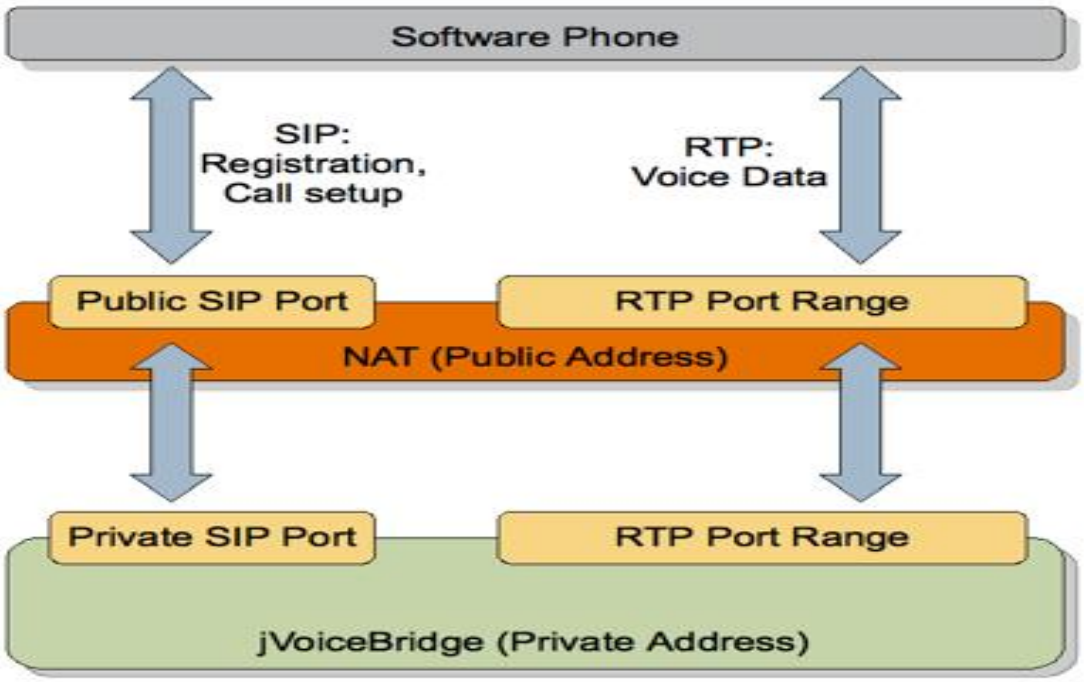

Figure 3.2: jVoiceBridge uses SIP/RTP protocols to transmit voice data

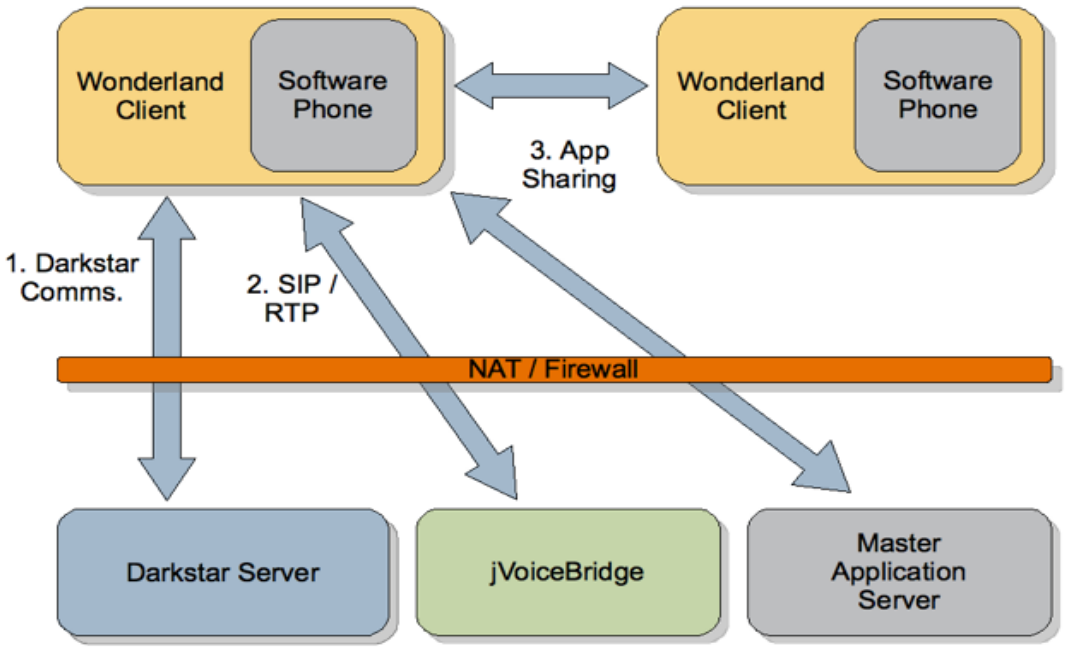

Figure 3.3: Communications in project Wonderland

3.2.2 Cairo

Cairo is a speech resource server that's also written entirely in Java. It was designed to comply with standards set forth in Media Resource Control Protocol Version 2 (MRCPv2), with which client hosts are able to control media processing resources. Figure 3.4 shows a picture the Cairo MRCPv2 server. Cairo is open source and is made of four projects:

- Cairo-server which builds on other open source speech projects. It adds functionality to allow "enterprise scale deployments of speech/telephony applications"[28].

- Cairo-client is a library that has basic speech client capabilities and an API to build clients in compliance with MRCPv2 speech servers.

- Cairo-sip is a library that has an API that supports SIP/SDP message communication between clients and servers.

- Cairo-rtp is a library that has an API that supports RTP audio streaming between a media source and sink.

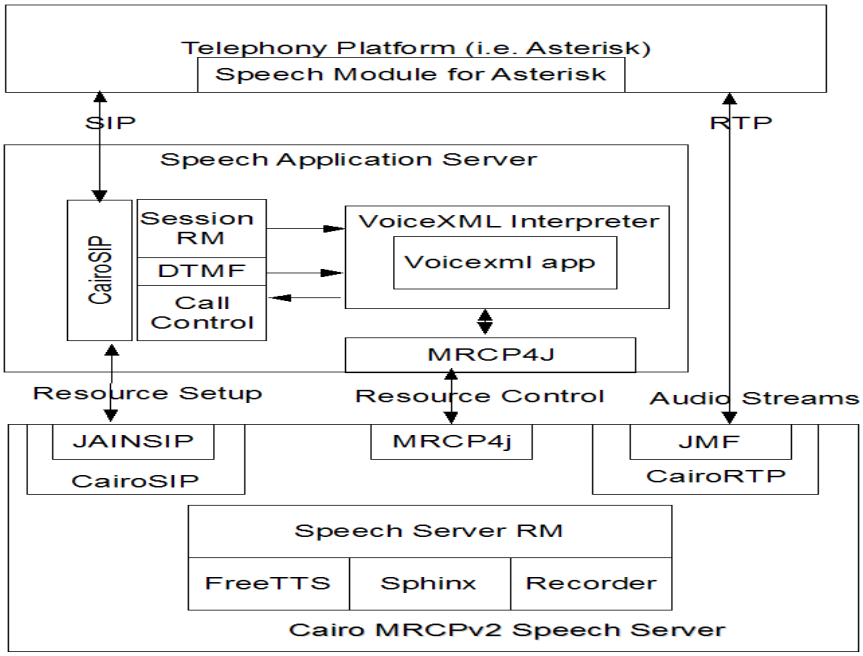

Figure 3.4: The Cairo MRCPv2 server

3.3 Embodied Conversational Agents

An embodied conversational agent(ECA) is a virtual human that can converse with humans. It produces speech, facial expressions and hand gestures. It also has a certain level of understanding. An ECA is a multimodal interface that has the modalities that are common in a human conversation such as speech, hand gestures, facial expressions and the stance of the body. Although these can be intricate parts of the communication process, some of these modalities are missing in the avatars of many virtual worlds. An ECA is a software agent that can be programmed within a computational environment (similar to an avatar) and a dialogue system that uses verbal and nonverbal devices to regulate and advance the dialogue between the user and the computer [5].

3.3.1 Embodied Conversational Agents in Second Life

Researchers from Rensselaer Polytechnic Institute(RPI) has incorporated AI into an avatar named "Eddie". Eddie is a 4 year old embodied conversational agent that interacts with people through chatting. He is able to make reasonable assumptions based on the information given. The researchers created a scenario in Second Life to test Eddie's ability to predict another avatar's response to a situation. They used an automated theorem prover along with methods that transforms Second Life's chatting into the prover's native language, formal logic. They used the RASCALS cognitive architecture, which can be seen in Figure 3.5 [18]. The first time around, Eddie made the wrong prediction, which would have been the typical response based on the average 4 year old's way of thinking. With a little adjustment to the code, the second time around Eddie would be able to make the correct prediction. The software that was used allows Eddie to be controlled by simulated keystrokes. Figure 3.6 shows a picture from the demonstation within Second life where Eddie is making the wrong prediction [19].

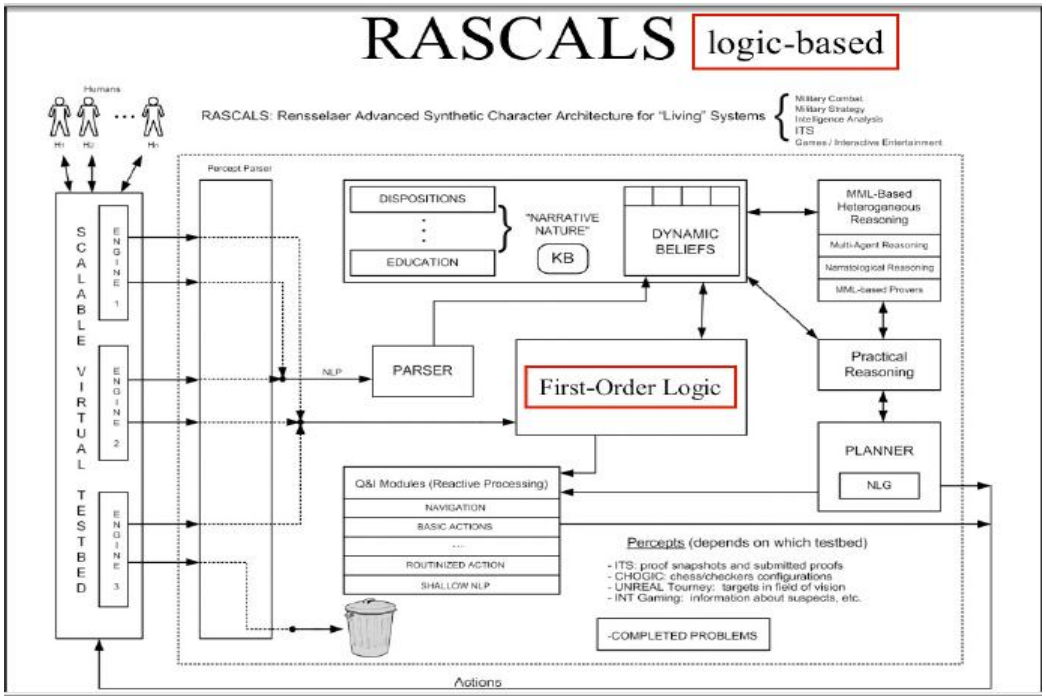

Figure 3.5: The RASCALS Cognitive Architecture

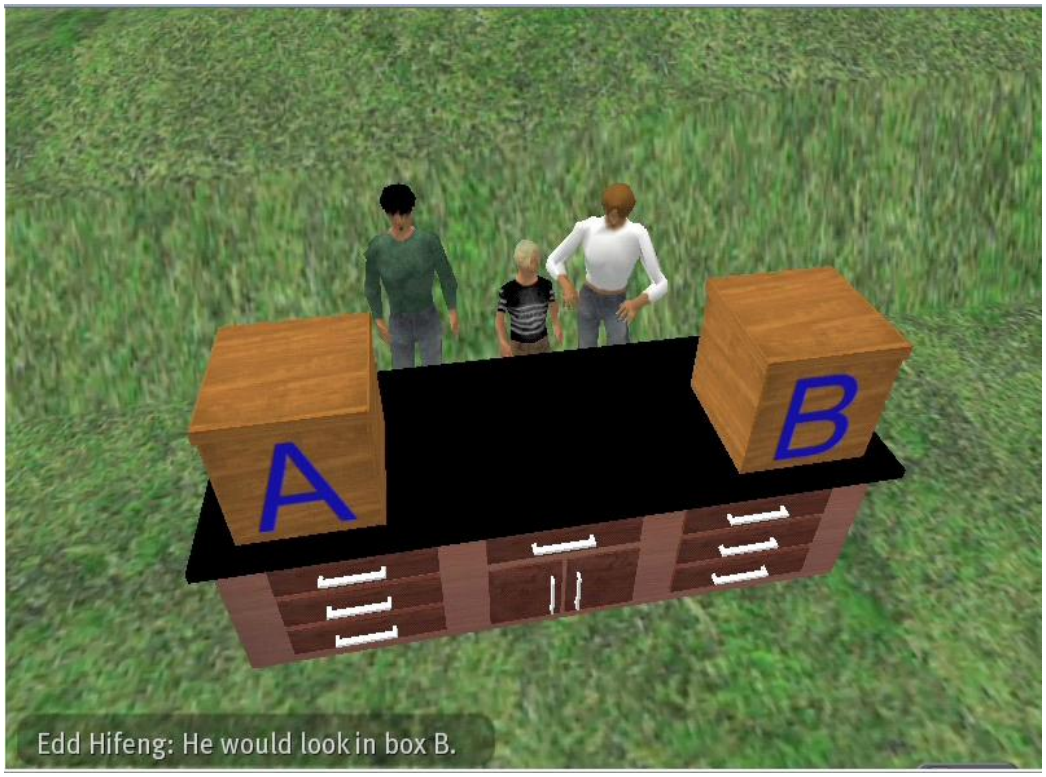

Figure 3.6: False Belief in Second Life

Chapter 4

PROBLEM STATEMENT

Currently there is no published documentation of Embodied Conversational Agents or stand-alone virtual agents in Second Life with speech recognition capabilities. If asked the question, "Why add speech recogntion to Second Life?" Some of the more prominent answers would be: Speech recognition is more natural for interacting; it could make the client more accessible to people with disabilities [29]; it's more convenient than texting (especially when your hands are busy and not available for use or you do not have the ability to use your hands); and it's great for open ended questions that may require more than a simple answer. Adding speech recognition capabilities to Embodied Conversational Agents would add a new dimension to virtual worlds, specifically to Second Life, and possibly enable other formerly excluded users with disabilities such as limited sight or typing skills to access and utilize the Second Life virtual community. After an in-depth analysis of communication within Second Life, this thesis aims to recommend a design method that would allow speech recognition by an Embodied Conversational Agent within Second Life.

Chapter 5

DESIGN

There are a couple of approaches and design methods that could be used to add speech recognition to Second Life. There is the client-side approach and the server-side approach. The various design methods within these approaches are discussed below.

5.1 Client Side Approach

There are three different design methods that will be address from the client-side approach.

5.1.1 Speech Recognition Application on Client

For this design the speech recognition application programming interface(API) could run on the client to interact with Second Life. This could be accomplished by creating an object within Second Life to capture the texting or chatting data and then transmitting the data to the speech recognition API located on the client.

The speech recognition API would process the data that's gathered from Second Life and send the results back to Second Life. In this design method, all of the processing is conducted on the client-side.

Some of the benefits are a greater range of functionality using an API. It would be easier to add or change features within the application. The initial development would not take much time and it would be easy to maintain. Figure 5.1 shows the basics of how this design would work.

The main drawback for this approach is that the client-side libraries share processing space with the speech recognition API, as they must run on the same computer, which could

be resource intensive, using a lot of CPU and memory space. Another drawback comes into play if you decide to change speech recognition technology, as the application may very well have to be completely rewritten and for this reason, this design was not recommended.

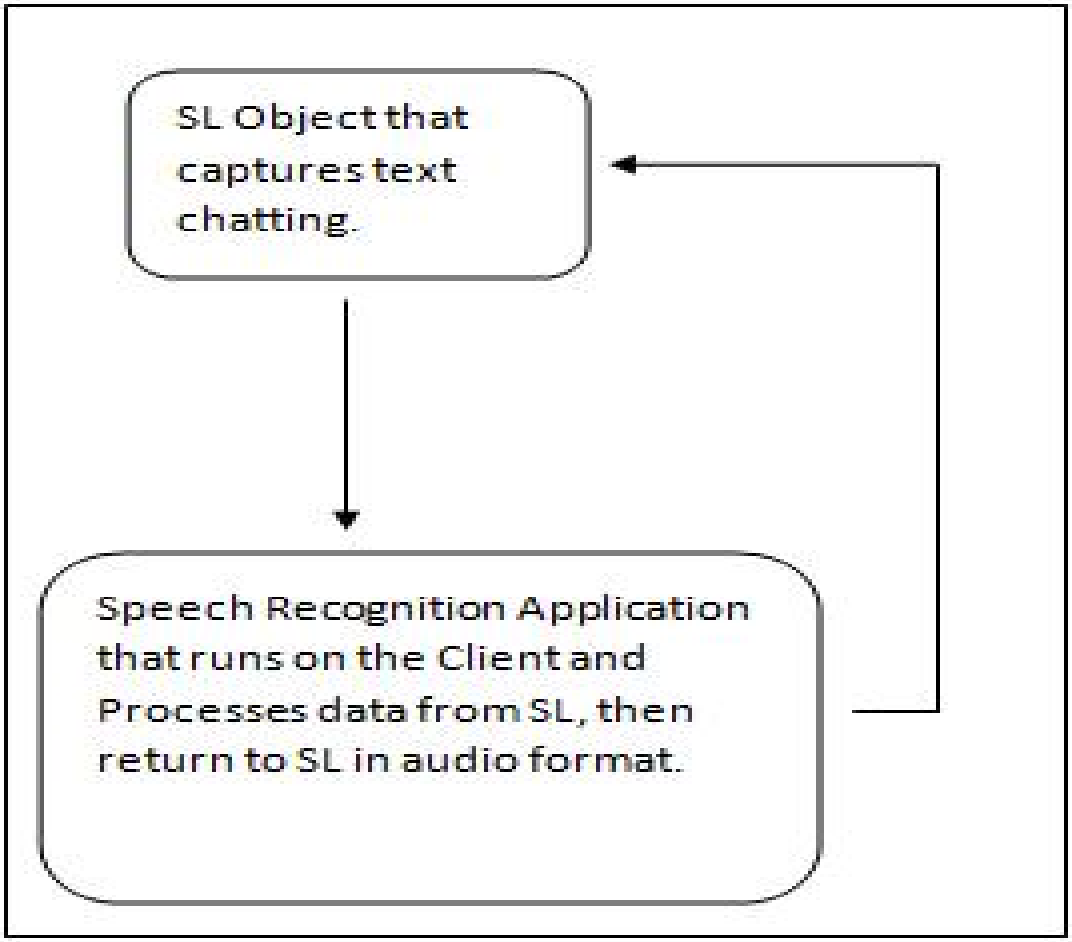

Figure 5.1: Speech Recognition Application on Client.

5.1.2 Client-Side Over the Network

This design uses a network connection in the processing of speech recognition. A speech recognition application could be developed on the client-side and used by deploying it over the network using MRPC V1, which only supports the speech synthesizer and recognizer resources and uses the Real Time Streaming Protocol(RTSP) between the client and the media resource server or MRPC V2, which uses Session Initiation Protocol(SIP) between the client and the media resource server and supports the speech synthesizer, recognizer and recorder resources along with speaker verification and identification resource types [6].

In this design method a network connection to the speech server(s) would be opened and a Java application would send information to Second Life to tell Second Life how to behave in order to incorporate speech recognition results within the world. The processing of speech recognition would occur on the client-side and the results would be transmitted to Second Life over the network.

When the speech server is updated, the speech client software would more than likely have to be updated, therefore this design was not recommended. This process is shown in Figure 5.2.

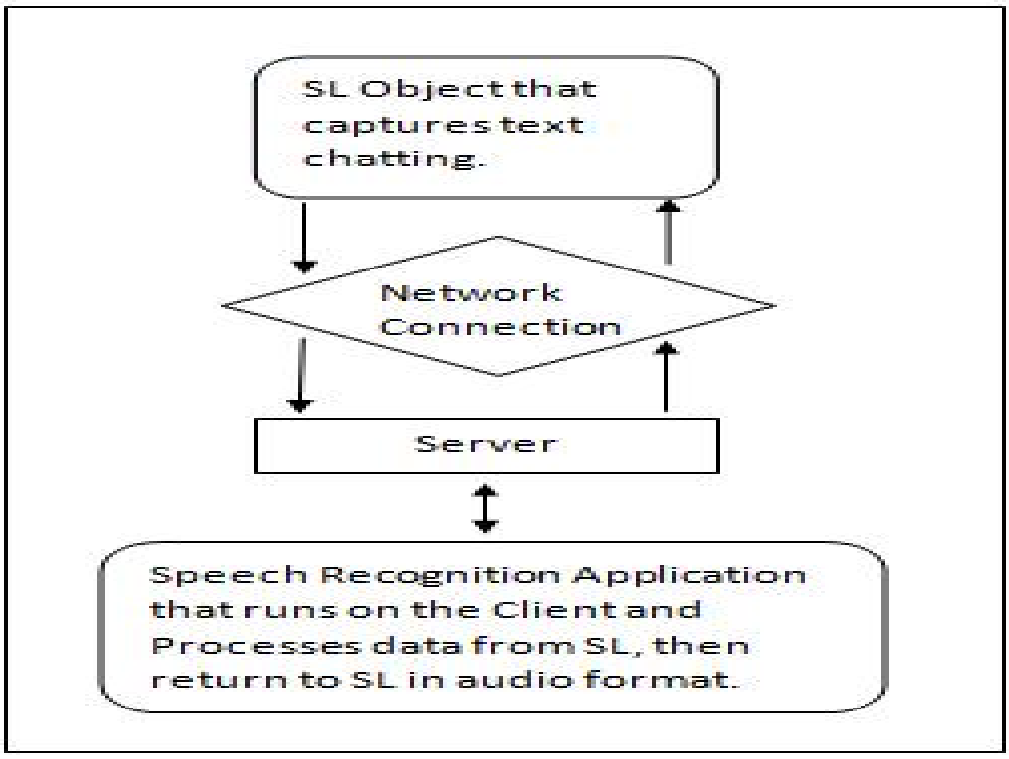

Figure 5.2: Speech Recognition Application on Client Over the Network.

5.1.3 Speech Recognition Engine

A speech recognition engine would be developed on the client, then integrate the engine into Second Life. The speech recognition engine would load the entire vocabulary, which includes all of the grammars (list of words to recognize) and the audio, which could be a microphone or telephone like application. It would then break down the audio into a mathematical wave form, analyze it for distinct features of speech, sounds and characteristics. It would separate all unnecessary noise then compare the sounds to acoustic models. The probable matches would then be returned. This design process would be very invasive and resource intensive, therefore it was not recommended.

5.2 Server Side Approach

There are two different design methods that will be address from the server side approach.

5.2.1 Server-Side Over the Network

With this design, an object could be created within Second Life to capture the texting or chatting data and then transmit the data to the speech recognition system through a network connection. A speech recognition application could be developed to be deployed over the network. In this design method a network connection to the speech server would be opened and a Java application would send information to Second Life to actually tell Second Life how to behave in order to incorporate speech recognition. This process is very similar to the client-side approach of the over the network design, however the processing of speech recognition occurs on the speech server(s), as shown in Figure 5.3.

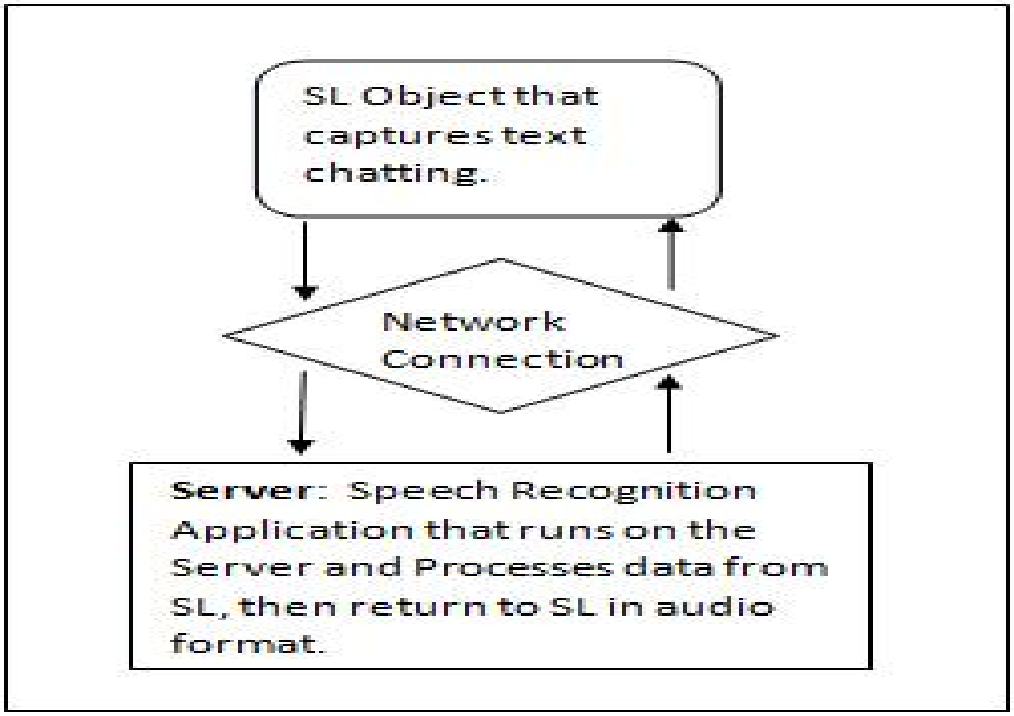

Figure 5.3: Speech Recognition Application on Server Over the Network.

5.2.2 Plug-In Object Over the Network

A plug-in object could be written and added to Second Life using the Linden Scripting Language that's used within the virtual world. The plug-in object could be as small and unobtrusive as you would like it to be and would act like a microphone. The texting or chatting data would be captured from within Second Life through the plug-in object then transmitted to the speech server(s) over the network; or the texting or chatting data could be captured over the network.

In either case, the Java application would process the speech recognition and return the data to Second Life through the network.

The Java application could be updated and downloaded to the network, when the client connects to the network, the updated application is executed without causing work for the client.

Since the plug-in object would be small and unobtrusive; it could be used by any Second Life resident; it would be very scalable and could have multiple speech clients interacting with multiple speech servers; it would be very portable with little to no work for the user when updated, therefore this design seems to be the best choice for the Second Life virtual world environment.

Chapter 6

CONCLUSION AND FUTURE WORKS

6.1 Conclusion

The main purpose of this study was to recommend a design that would be best to incorporate speech recognition into an Embodied Conversational Agent within Second Life. There were three different designs discussed for the client-side approach and two different designs discussed for the server-side approach. It was determined that the best approach and design to incorporate speech recognition into an Embodied Conversational Agent within Second Life would be the server-side approach using the plug-in object over the network. In this approach a plug-in object would be designed and added to Second Life using the Linden Scripting Language. The plug-in object would be small and unobtrusive; it could be used by any Second Life resident; it would not require a lot of processing resources from within Second Life; it would be very scalable and could have multiple speech clients interacting with multiple speech servers; it would be very portable with little to no work for the user when updated, therefore this would be the recommended approach and design method.

6.2 Future Works

Potential future works that can be pursued from this analysis would include the actual development and implementation of the plug-in object design. Other future works may be the development of an Embodied Conversational Agent that can pass the Turing test. With speech recognition systems, no artificial intelligence(AI) is needed, but think of the possibilities if AI were added to the equation. This would take Second Life to an entirely new level.

Bibliography

[1] Armstrong, A., Hagel, J. (1997). Net Gain. Boston Massachusetts: Harvard Business School Press.

[2] Baum, L. E. (1972). An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov process. Inequalities 3, 1-8.

[3] Bell, M. W. (2008). Toward a Definition of "Virtual Worlds", Journal of Virtual Worlds Research 1(1), July 2008. ISSN 1941-8477 [online]. Available: http://journals.tdl.org/jvwr/article/viewFile/283/237 accessed on June 13, 2009.

[4] Bricken, W. (1990). Learning in virtual reality. HITL document. Seattle, WA, Human Interface Technology Laboratory.

[5] edited by: Cassell, J., Sullivan, J., Prevost, S. and Churchill, E. F. (2000). Embodied Conversational Agents [online]. Available: http://mitpress.mit.edu/catalog/item/default.asp?tid=3494&ttype=2 accessed on June 14, 2009.

[6] Dialogic. MRCP V1 Client Library User's Guide [online]. Available: http://www.dialogic.com/products/docs/appnotes/9603_MRCP_V1_Client_Lib_User_Guide_an.pdf accessed on July 1, 2009.

[7] Dickey, M. D. (2003). 3D virtual worlds: An emerging technology for traditional and distance learning. Proceedings of The Convergence of Learning and Technology, Windows on the Future [online]. Available: http://www.oln.org/conferences/OLN2003/papers/Dickey3DVirtualWorlds.pdf accessed on April 3, 2009.

[8] Furui, S. (2002). Recent progress in spontaneous speech recognition and understanding. Proceedings of the IEEE Workshop on Multimedia Signal Processing.

[9] History of Second Life (2009). Available: http://wiki.secondlife.com/wiki/History_of_Second_Life accessed on June 8, 2009.

[10] jVoiceBridge (2004). Available: https://jvoicebridge.dev.java.net/ accessed on June 20, 2009.

[11] Kenyon, R. V. (1995). The Cave: Automatic Virtual Environment: Characteristics and Applications [online]. Available: http://www.cs.uic.edu/~kenyon/Conferences/NASA/Workshop_Noor.html accessed on June 1, 2009.

[12] Land Pricing and Use Fees [online]. Available: http://secondlife.com/whatis/landpricing.php accessed on July 16, 2009.

[13] Linhares, R. L. MediaSoft: Virtual Reality in Cyberball [online]. Available: http://www.interactive-park.com/Cyberball_Mediasoft_eng.pdf accessed on June 7, 2009.

[14] Maze War (2004). The DigiBarn's Maze War 30 Year Retrospective "The First First Person Shooter" [online]. Available: http://www.digibarn.com/history/04-VCF7-MazeWar/index.html accessed on June 1, 2009.

[15] McMillian, Y., Gilbert, J.E.(2008). Distributed Listening: A Parallel Processing Approach to Automatic Speech Recognition. Proceedings of ACL-08: HLT, Short Papers (Companion Volume), pp. 173-176, Columbus, OH.

[16] M.U.D Multi-User Dungeon (2007) [online]. Available: http://www.british-legends.com/ accessed on June 7, 2009.

[17] Open Source Portal (2009) [online]. Available: http://wiki.secondlife.com/wiki/Open_Source_Portal accessed on July 26, 2009.

[18] Rensselaer (2008). Advanced Synthetic Characters/RASCALS Cognitive Architecture [online]. Available: http://rair.cogsci.rpi.edu/asc_rca/ accessed on July 25, 2009.

[19] Rensselaer (2009). Bringing Second Life To Life: Researchers Create Character With Reasoning Abilities of a Child [online]. Available: http://news.rpi.edu/update.do?artcenterkey=2410 accessed on May 23, 2009.

[20] Second Life (2008). M Linden announced as new CEO [online]. Available: http://blog.secondlife.com/2008/04/22/announcing-our-new-ceo/ accessed on June 8, 2009.

[21] Second Life (2009). Knowledge Base in Second Life [online]. Available: https://support.secondlife.com/ics/support/default.asp?deptID=4417 accessed on June 11, 2009.

[22] Second Life (2008). Mono Launch [online]. Available: http://blog.secondlife.com/2008/08/20/mono-launch/ accessed on June 8, 2009.

[23] Second Life (2008). Open Grid Public Beta Begins [online]. Available: http://blog.secondlife.com/2008/07/31/open-grid-public-beta-begins-today/ accessed on June 8, 2009.

[24] SecondLifeUpdate (2007). Top 10 Online 3D Virtual Worlds [online]. Available: http://www.secondlifeupdate.com/2007/10/18/top-9-online-virtual-3d-worlds accessed on May 30, 2009.

[25] Second Life on Wikipedia (2009) [online]. Available: http://en.wikipedia.org/wiki/Second_Life accessed on April 25,2009.

[26] Second Life (2009). XStreet SL and OnRez to join Linden Lab [online]. Available: https://blogs.secondlife.com/community/features/blog/2009/01/20/xstreet-sl-and-onrez-to-join-linden-lab/ accessed on June 8, 2009.

[27] A Second Life on the Grid (2009). What's a SLURL? [online]. Available: http://slonthegrid.blogspot.com/2009/07/whats-slurl.html accessed on July 25, 2009.

[28] SpeechForge (2008) [online]. Available: http://www.speechforge.org/projects/cairo/ accessed on June 15, 2009.

[29] Springer R. (2009). Speech in a Virtual World [online]. Available: http://www.speechtechmag.com/Articles/Column/Voice-Value/Speech-in-a-Virtual-World-55189.aspx accessed on July 15, 2009.

[30] Strickland, J. (2007). How Virtual Reality Works [online]. Available: http://electronics.howstuffworks.com/gadgets/other-gadgets/virtual-reality.htm/printable accessed on June 10, 2009.

[31] Taxonomy of Virtual Worlds for Education Workshop, (Philadelphia, PA, March 2009)[online]. Available: http://view.scicentr.org/records/abstracts.aspx accessed on April 4, 2009.

[32] Teen Second Life (2009) [online]. Available: http://en.wikipedia.org/wiki/Teen_Second_Life accessed on June 8, 2009.

[33] Virtual World (2009) [online]. Available: http://en.wikipedia.org/wiki/Virtual_world accessed on April 1, 2009.

[34] Windlight Release (2009) [online]. Available: http://blog.secondlife.com/2008/04/02/the-dawning-of-a-new-viewer-second-life-1191-now-available/ accessed on June 8, 2009.

[35] Young, S. R. (1990). Use of dialog, pragmatics and semantics to enhance speech recognition. Speech Communication, vol. 9, pp. 551-564.

[36] Young, S. R., Hauptmann, A. G., Ward, W. H., Smith, E. T., and Werner, P. (1989). High level knowledge sources in usable speech recognition systems. Communications of the ACM, vol 31, no. 2, pp. 183-194.