

Mentors

Dr. Larry F. Hodges LFH@clemsun.edu
Dr. Amy Ulinski aulinski@clemsun.edu
Toni Bloodworth tbloodw@g.clemsun.edu



Author:
Freddie Lee Dunn III

School:
Georgia Institute of Technology

Contact:
gth745c@mail.gatech.edu

Introduction

This summer research was about virtual reality and virtual environments. Our goal was to create a virtual character that interacts with the user and tells them some of the benefits of being a computer science major and all that the School of Computing here at Clemson has to offer.

We created two different virtual humans named CARA, for Clemson Automated Research Agent, and CLEM, for Clemson Learning Experience Mentor. The challenge was to have the virtual humans interact with the user in a natural way.

Background

Voice recognition technology allows a user to speak naturally while the computer interprets the speech into text that can be used. With each new product the recognition accuracy and capabilities continue to expand. Some of the more common voice recognition programs existing at the moment are:

- Microsoft SAPI
- Dragon Naturally Speaking
- Sphinx
- SALT

Microsoft SAPI is something that has been released and included with Microsoft Vista so that everyone has access and can use it if they so choose. Dragon Naturally Speaking, or dragonspeak for short, is supposedly one of the best on the market, but while this program has many great features, the cost may outweigh the benefits if compared to some of the free versions available.

Text-to-Speech (TTS) is a feature that has also been included with most of these programs that allows for users to type in text and have the computer read back what was said. For dragonspeak, the voices were recorded because all the logic relating to when they would play was outside of Second Life; however, for SAPI, a TTS voice was used because all the logic relating to when it would be played was within C#.

Research

For the two demos that were constructed, Dragon Naturally Speaking was used in the creation of CARA while Microsoft SAPI was used in the creation of CLEM.



Figure 1: CARA and CLEM the virtual humans

Dragon Naturally Speaking

Initially the plan was to have users interact with Dragonspeak by transferring our speech into the chat box for Second Life, which is the only way it can recognize inputs externally; however, the user still had to manually press enter in order for the text to be recognized by the program.

This problem led to first using a few command words that, when spoken, typed into the chat box the intended phrase and press enter afterward.

Another problem that came about from this was that every word/phrase/name had to be turned into a keyword so that it can be recognized by the voice recognition. While this did work, adding new commands to the grammar file was difficult.

The last problem that was encountered was that dragonspeak became unreliable due to the program randomly not listening at times or just freezing altogether. When the time came for us to put together CLEM, we decided to design a different solution.

Microsoft SAPI

SAPI demonstrated that it could automatically detect when a user finished speaking and solved the problem of keywords that was necessary with dragonspeak. Also, by creating custom grammar files, which were easier compared to dragonspeak, we were able to increase vocal recognition and accuracy which increased the reliability and functionality of this program.

Implementation



Figure 2: CARA demo in action

Figure 2 demonstrates how the CARA demo is run. In this demo, I was able to test and show how voice recognition integrated into Second Life, via Dragonspeak, was able to work. This demo was run with multiple groups including everyone from middle school teachers to a group of 7th graders attending a computer camp and everyone else in the middle.

Impact

Voice recognition had a big impact on the overall project. This technology allowed the user to do a variety of things:

- Recognize user voices and interpret what was said
- TTS capabilities that allowed interaction with the user
- Create custom grammar files
- Work for a variety of users which is useful for demonstrations or larger future projects and endeavors

I learned a lot from getting a chance to work with the REU program. This opportunity gave me plenty of experience such as:

- The ability to work in multiple programming languages such as C#, SQL, Linden Script, Dragon Script, and XML neither of which I had prior experience in
- Learning a skill that is on the front end of technology and has multiple capabilities in future applications.
- Critical thinking skills that can be used for later projects

This project was unique in that it hadn't been attempted before and what was done was never the original intention of the designers of Second Life. The accomplishment of a completed project that is different than intended shows the ability to design creatively on multiple levels.

Conclusion

Through the course of this project, I learned a variety of information about the voice recognition technology that was available to us.

Dragon Naturally Speaking

- Advantages
 - The many features gave users a variety of ways to do voice recognition
 - Easy to adjust the grammar files or make your own (all inclusive)
- Disadvantages
 - Could not be programmed to automatically know when users have stopped speaking
 - Did not contain a complete version of Visual Basic Script (needed SDK libraries)
 - Would freeze quite often when running which made it unreliable

Microsoft SAPI

- Advantages
 - Very customizable use due to everything being coded in C#
 - Automatically detects when user is finished speaking
 - Ability to import custom grammar files written in XML and compiled into a CFG
- Disadvantages
 - Accurate voice recognition was dependent upon the grammar files (the inclusive one was too broad to be of use)

Future Work

I plan to use what I have learned here to help in future projects in order to make them more able to interact with users. One of the most immediate plans I have for this knowledge is to use it for my senior project design this fall at Georgia Tech.



Figure 3: The avatars in Second Life for the REU group