Speech to Overhead Text Display *

Thomas Jelkin
thomas.jelkin@ndsu.edu

Scott Miller
scott.p.miller@ndsu.edu

Austin Wanner
austin.p.wanner@ndsu.edu

North Dakota State University
Electrical and Computer Engineering Department
Fargo, ND 58108-6050

May 2016

Keywords: Speech to Text,Speech Conversion, Microphone, Array, Classroom

Abstract

The goal of this project is to create a better learning environment for those with hearing disabilities. This is accomplished by creating a device that is portable and can take in an input of speech from both the professor and the students. Converting speech to text should have an accuracy of at least seventy-five percent, and have a conversion time of less than thirty seconds from when the speaker finishes a statement to when words are outputted. This device should also be able to wirelessly display converted text via projector.

*** This material is based in part upon work supported by the National Science Foundation (NSF) under Grant Number CBET-1401507. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.**

1. Introduction

The Speech to Overhead Text Display system is a device used to convert spoken word to text and display it via a wireless system in the classroom. Every student should be given the same opportunities to learn, regardless of any disabilities that they may have. Students with hearing disabilities may not have that same opportunity even when sitting in front of the class, the lecture may not be clearly audible. If they have to focus all of their attention on figuring out what the professor is saying, they may not be able to focus on the material itself, making it harder to learn. Currently these students can read handwritten/printed notes from either the professor or another student, but they are missing out on everything the professor says that is not in the notes and also any questions the other students may have.

This project is meant to solve the problems that students have with learning when they have hearing disabilities. A system that displays text of everything that the professor is saying will fix these problems by letting the student read the lecture instead of listen to it. Our system is able to listen to whoever is speaking, process the audio signal, then wirelessly send the text to a projector that is in the room.

In previous work, there have been devices that convert speech to text, but these devices do not recognize the entire English language. Some of these devices are programed to only know certain words, because the device does not need to know every word. There have also been some devices that are handheld devices to pick up speech from only one person and then display that text.

In Section 2, the previous work in speech to text devices will be examined. Section 3, will elaborate on the requirements needed for this device. The various design options pertaining to both hardware and software will be explained in Section 4. For Section 5, the final design approach, including the flowcharts and each sub-device, will be discussed. Section 6, is the conclusion. The parts that were used in the device will be listed in Section 7. Section 8 will be additional information. The various references will be listed in section 9. Finally in section 10, the approval will be stated.

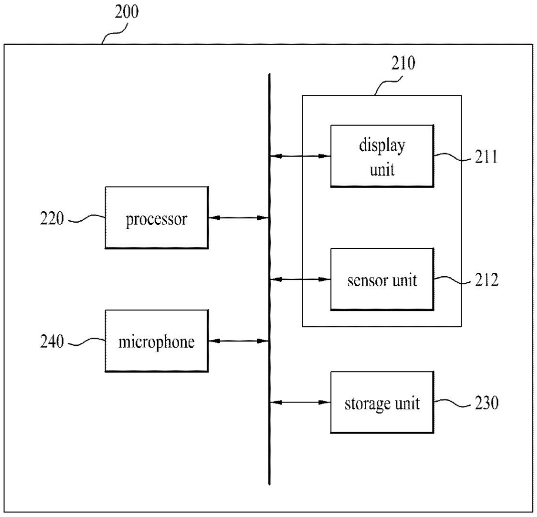
2. Previous Work

2.1 Speech-to-Text Display Device Method and Display Device for Converting Text to Speech

LG Electronics Inc. patented a device and method, which employs a storage unit to store the voice data from a microphone, a processor to translate the voice data into text, a sensor unit to detect user input to the display unit, and a display unit to display the text. The figure to the right (Fig. 1) shows the setup of their device.

Their patent mentions the use of a sensor for touch/motion detection. We will not be utilizing any touch/motion detection, and will consequently not need a sensor unit.**[262]**

FIG. 1

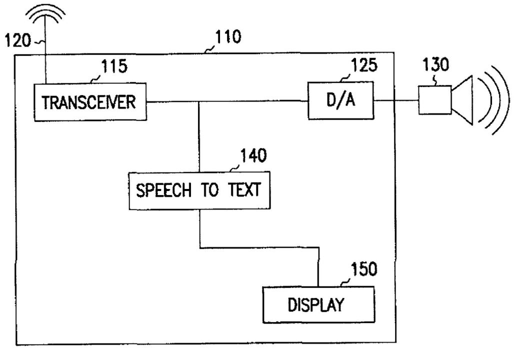


2.2 Speech and Text Conversion in Handheld Device

Gateway, Inc. has patented a device which uses a receiver to route voice to text communications to a speech-to-text-processor and then displays the text on a handheld device. Their device also has a speaker that can playback the audio, which required them to use a digital-to-analog converter in order to output the signal, which they had digitally stored. Fig. 2 on the right shows the relationship between these units.

The capability to have audio playback is not part of the requirements for our device. Therefore, we will not need a digital-to-analog converter to playback the system. Their patent indicates that their receiver pertains to integrated cellular modems. our device will also use one to send the signal to the wireless receiver connected to our projector. Our device is not handheld as this one, but it is small and compact enough to ensure portability. **[262]**

FIG. 2



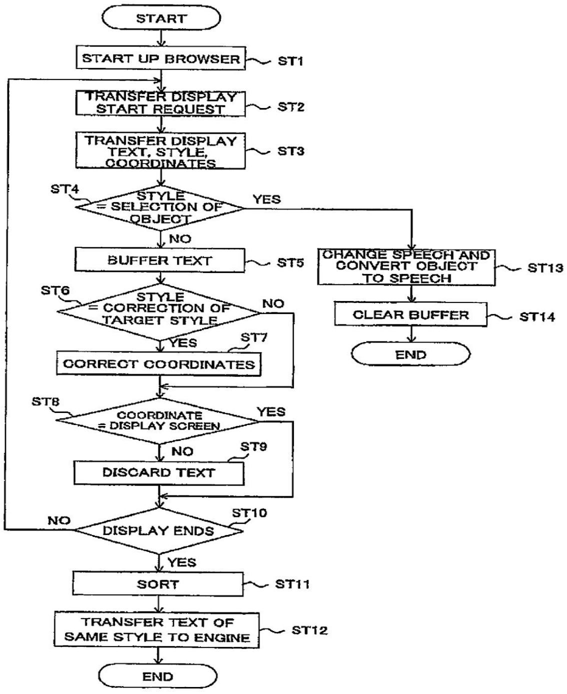
2.3 Speech-to-Text Conversion System

Display Apparatus Equipped with Speech Synthesis Method

Kyocera Corporation has patented a display apparatus that uses speech synthesis. The device uses a storage unit to store the text information, a display unit to display the text information, and a speech synthesizer for converting speech to text. Fig. 3 on the right is a flowchart for the device.**[386]**

Our device differs from this device because our device does not require any network activity for requiring speech to text information.

FIG. 3


START
START UP BROWSER
ST1
TRANSFER DISPLAY START REQUEST
ST2
TRANSFER DISPLAY TEXT, STYLE, COORDINATES
ST3
ST4
STYLE = SELECTION OF OBJECT
YES
NO
BUFFER TEXT
ST5
CHANGE SPEECH AND CONVERT OBJECT TO SPEECH
ST13
CLEAR BUFFER
ST14
END
ST6
STYLE = CORRECTION OF TARGET STYLE
NO
YES
ST7
CORRECT COORDINATES
ST8
COORDINATE = DISPLAY SCREEN
YES
NO
ST9
DISCARD TEXT
NO
DISPLAY ENDS
ST10
YES
SORT
ST11
TRANSFER TEXT OF SAME STYLE TO ENGINE
ST12
END

2.4 Speech to Text Conversion Communication System

Harris Corporation has patented a communication system and method designed to convert speech to text. It multiplexes the speech signal into a text message and wirelessly transmits this signal to a second device which demultiplexes the text message. The text message is then displayed simultaneously with the corresponding speech, which is broadcast via an audio output transducer.**[983]**

Our device does not store or output the speech signal,

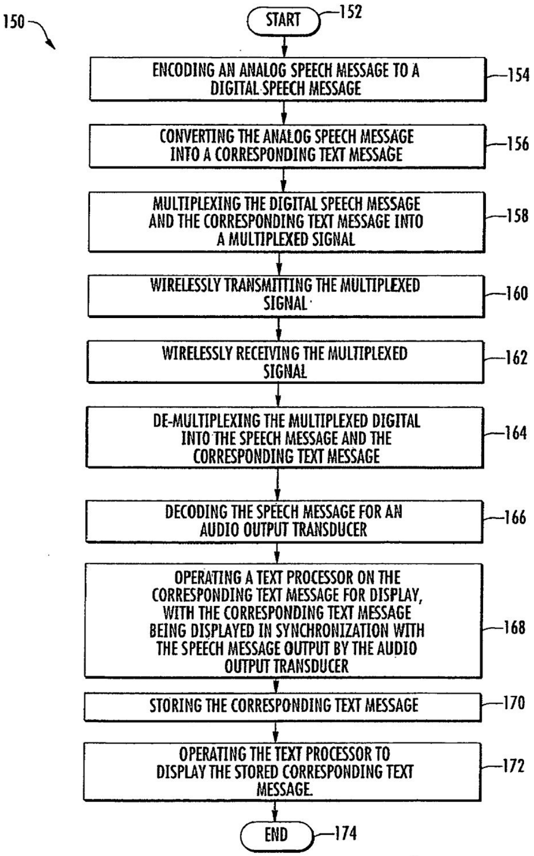

150
START
152
ENCODING AN ANALOG SPEECH MESSAGE TO A DIGITAL SPEECH MESSAGE
154
CONVERTING THE ANALOG SPEECH MESSAGE INTO A CORRESPONDING TEXT MESSAGE
156
MULTIPLEXING THE DIGITAL SPEECH MESSAGE AND THE CORRESPONDING TEXT MESSAGE INTO A MULTIPLEXED SIGNAL
158
WIRELESSLY TRANSMITTING THE MULTIPLEXED SIGNAL
160
WIRELESSLY RECEIVING THE MULTIPLEXED SIGNAL
162
DE-MULTIPLEXING THE MULTIPLEXED DIGITAL INTO THE SPEECH MESSAGE AND THE CORRESPONDING TEXT MESSAGE
164
DECODING THE SPEECH MESSAGE FOR AN AUDIO OUTPUT TRANSDUCER
166
OPERATING A TEXT PROCESSOR ON THE CORRESPONDING TEXT MESSAGE FOR DISPLAY, WITH THE CORRESPONDING TEXT MESSAGE BEING DISPLAYED IN SYNCHRONIZATION WITH THE SPEECH MESSAGE OUTPUT BY THE AUDIO OUTPUT TRANSDUCER
168
STORING THE CORRESPONDING TEXT MESSAGE
170
OPERATING THE TEXT PROCESSOR TO DISPLAY THE STORED CORRESPONDING TEXT MESSAGE.
172
END
174

FIG. 4

3. Requirements

There were defined requirements this project included in order to accomplish the objective of the client. Primary requirements were necessary components of the project. Secondary requirements were elements which could be added if time allotted. Constraints were needed in order to meet clients requests.

3.1 Primary Requirements

- · Device should have portable capabilities

Hardware:

- At least 8Gb of ram
- 120Gb of memory
- Contain a processor that runs at a minimum speed of 1.6 GHz
- Device should plug into 120V outlet

Inputs:

- Directional microphone to capture speech

Software:

- Software converts input audio signal to text
- Converts the speech to text at an average accuracy rate of at least 75%
- Should display the text no longer than 30 second after it is spoken
- Software should run as soon as the device boots up

Outputs:

- Projector should connect to the device wirelessly
- Displays converted text through projector

3.2 Secondary Requirements

Build a microphone array that is able to take in the best signal of spoken word from either the audience or primary speaker to convert it into text, rather than using a standalone microphone. This microphone array should contain a minimum of 8 microphones.

Due to lack of time, and the lack of knowledge that a microphone array will improve the accuracy, we were not able to complete the microphone array.

3.3 Constraints

The only thing that could cause problems for our device is white noise. White noise is any outside noise like air flow

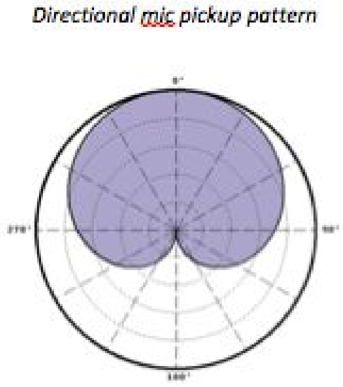
4. Design Options

The different design options we considered varied in the following; hardware selection, software, microphone type, power sources, displays. These options were either sifted through or experimented with.

4.1 Different Microphone Type(s)

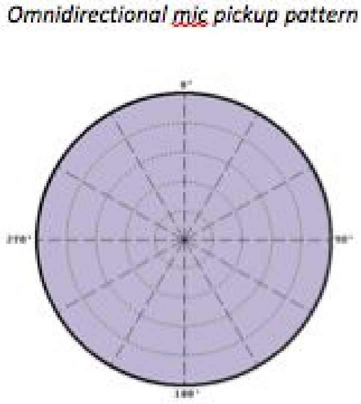
Microphone Type	Advantages	Disadvantages
Cardioid	● Channel separation capabilities ● Good at picking out a single voice and rejecting the rest	● Would need multiple microphones for the instructor and the audience ● Does not pick up noises well from the side which would limit the movement of the instructor
Omnidirectional	● Would only need one microphone to detect audio from both the instructor and the audience. ● Normally smaller and have a lower profile. ● Allows the instructor to move around while staying in the range of the microphone. ● Allows the instructor to move around with less chance of stepping into a feedback zone. ● Lower distortion levels.	● No channel separation to differentiate direct and indirect sounds. ● Will pick up more background noise

Bidirectional	● Useful for isolating one voice. ● Can cover both the instructor and the audience	● Less sensitive to sounds coming from the side, which would limit the movement of the instructor.



Eveterry21.wordpress.com

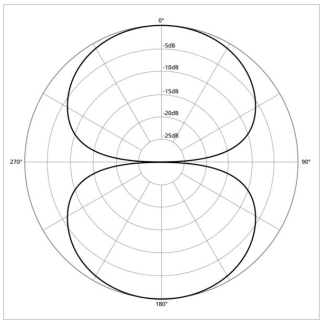
Cardioid (Directional)



www.studyblue.com

Omnidirectional

Bidirectional mic pickup pattern



www.harmonycentral.com

Bidirectional

In the end, we decided on using a microphone that has the capabilities of all three options. We decided on this microphone because it gives us more options if we change our mind on the way we will use the microphone.

4.2 Hardware Selection

4.2.1 *Speech-to-Text DSP Chip Selection: DSP Chip TDA7590*

This device or similar devices were considered because they contain the following properties; Echo cancellation, Noise cancellation, Speech recognition, Speech synthesis, Access to external RAM (16Mw) through expansion port, Large on-board memory (128k Words-24 bit) **[DSP]**

However, we did not go with this selection due to the lack of availability as well as how expensive one of the only available options were.

4.2.2 *Raspberry Pi 2*

Similar to that of a micro computer, the raspberry pi 2 has the following properties; 900 MHz Broadcom BCM2836 Arm7 Quad Core Processor, 1GB RAM4 x USB 2 ports, Full size HDMI, Micro USB power source, with 400 mAh. **[Pi]**

We originally went with the Raspberry Pi 2 as our hardware selection but found out it did not have the necessary amount of computing power we needed in order to convert speech to text in real time.

4.3 Software Considerations

4.3.1 Operating system- Ubuntu 15.10:

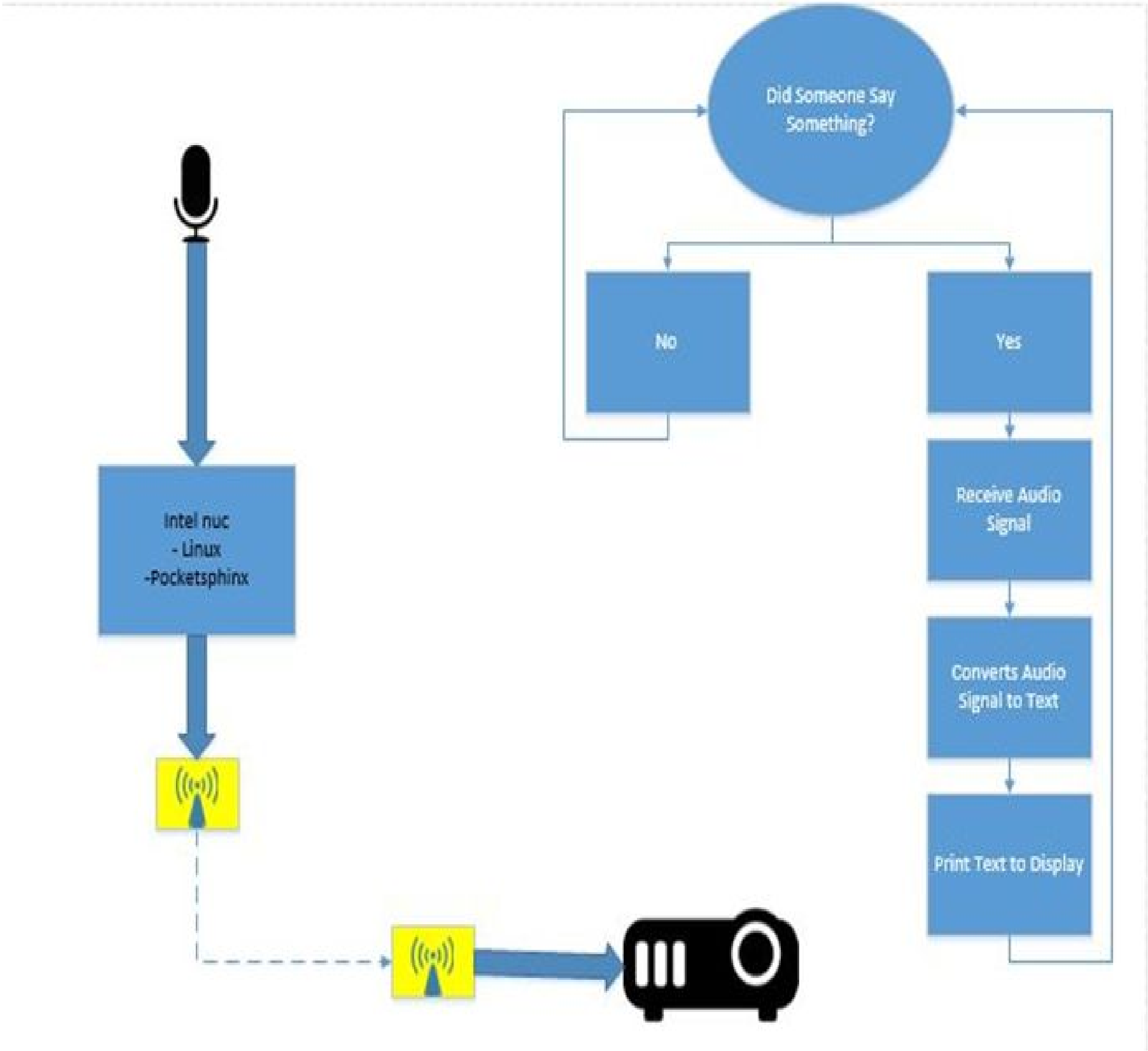
We decided to use Linux as the operating system for our hardware because it is more efficient than windows in fulfilling our applications requirements. Also the software we are using is specifically made for Linux, so this made choosing Linux as our operating system a logical choice.

4.3.2 Pocketsphinx

The software we decided to use is called Pocketsphinx. Pocketsphinx is an open source speech to text software that takes an input that is a single channel audio signal with a 16 kHz sample rate. This software then splits the waveform on utterances, and uses an acoustic model, phonetic dictionary, and a language model to convert speech into text. An acoustic model contains acoustic properties for each senone. A phonetic dictionary contains a mapping from words to phones. Finally, a language model is used to restrict the word search. **[FAQ]**

5. Final Design Approach

5.1 How it all works (FlowChart)



5.2 Our project overview

In the end, after some hardware and software considerations it decided upon the following.

5.3 Intel NUC

A mini pc that contains the following; 6th generation Intel i5 processor with support up to 32Gb of primary memory. A secondary storage capacity of 250Gb with internal ssd,with built in wireless and bluetooth connection, along with 4x USB 3.0 ports (1x with charging capable) support. **[NUC]**

A greater processor and ram was needed in order to convert speech in faster amounts of time, which was more than what the raspberry pi had. Decreasing the average conversion time from 6.5 minutes down to roughly 0.5-1.5 minutes. Time varies based upon how long the user's utterances are.

5.4 Microphone

The Microphone we used is a Blue Microphones Yeti USB Microphone. This microphone has a sampling rate of 48kHz, and a bit rate of 16-bits. It requires 5V and 150mA of power. The Yeti has multiple directional capabilities including: Cardioid, Bidirectional, Omnidirectional, and Stereo. Another feature that is very beneficial about this microphone is the adjustable gain. **[Yeti]**

5.5 Wireless system

Our wireless transmitter that we used in order to send our converted text terminal to our projector is an IOGEAR Wireless Hdmi Transmitter/Receiver Kit GWHD11. This device is lightweight and compact, only 3.8x3.8x1.5 inches, and weighs 5.6 ounces. It supports video resolutions of 480p, 720p, and 1080p. It also has the capabilities of doing wireless uncompressed HD 1080p audio/video streaming with 3D support up to 30ft. The best part of about this wireless system is that no software or driver installation needed. **[IOGEAR]**

5.6 Projector

The projector that we decided to use is a ViewSonic PJD5155 LightStream SVGA Home Entertainment Projector. This particular projector features 3300 lumens. This was chosen because we wanted a projector that would be able to work and be seen well in a room with the lights on. The resolution is native SVGA 800 x 600. The projector has multiple connectivity capabilities including: HDMI, 2 x VGA, Composite Video, S-Video, 1 x VGA output, Audio in/out, Mini USB and RS232. **[Proj]**

5.7 Pocketsphinx

5.7.1 What is Pocketsphinx

Pocketsphinx is an open source speech recognition software developed by Carnegie Mellon University. Its input is a single channel audio signal with a 16 kHz sample rate. It splits the waveform based on utterances. The system uses an acoustic model, phonetic dictionary, and a language model to convert speech into text. An acoustic model contains acoustic properties for each senone. A phonetic dictionary contains a mapping from words to phones, while the language model is used to restrict word search **[Tool]**

5.7.2 Basic Concepts of Speech

The basic building blocks of sounds are phones, triphones, senones and utterances. Phones are classes of sounds. The acoustic properties of a phone can sound different based on context, speaker, and the style of speech.Some phones can sound different depending on what word they are part of. Phones in context are called Triphones.Senones are detectors for triphones. Utterances are formed by non-linguistic sounds. They are generally pauses, or filler words like “um, uh, etc..”. **[Basic]**

6. Conclusion

The speech to overhead text display device is very beneficial to the future of teaching for those with hearing disabilities. This system solves the problem of those students that can not hear very well. It will allow those students to be able to read what the professor/students say instead. By being able to read what the professor is saying, the students will be able to have any questions that were asked in the class.

Overall, the speech to overhead text display system makes classes easier for those with hearing impairments, and even those who are taking notes and they did not catch what the professor said.

7. Parts List

- Mini PC—Intel® NUC Kit NUC6i5SYH
- Samsung 850 EVO 250 GB M.2 SSD (MZ-N5E250BW)
- Crucial CT2KIT102464BF160B 16GB Kit DDR3-1600 MT/s 204-Pin SODIMM Notebook Memory
- Blue Microphones Yeti USB Microphone

- ViewSonic PJD5155 LightStream SVGA Home Entertainment Projector
- IOGEAR Wireless Hdmi Transmitter/Receiver Kit GWHD11

8. More Information

For more information on this and other adaptive technology projects at the North Dakota State University Electrical and Computer Engineering Department, contact:

Samee Khan, Ph.D.
Associate Professor
Samee.Khan@ndsu.edu
Phone: 7012317615

9. References:

[262] "Patent US20030097262 - Handheld Device Having Speech-to Text Conversion Functionality." Google Books,
http://www.google.com/patents/US20030097262
22 May 2003. Web. 05 Oct. 2015.

[386] "Patent US20060224386 - Text Information Display Apparatus Equipped with Speech Synthesis Function, Speech Synthesis Method of Same, and Speech Synthesis Program." Google Books,
http://www.google.com/patents/US20060224386
05 Oct. 2006. Web. 12 Oct. 2015.

[983] "Patent US20140163983 - Display Device for Converting Voice to Text and Method Thereof." Google Books,
http://www.google.com/patents/US20140163983
12 June 2014. Web. 12 Oct. 2015.

[Basic] "CMUSphinx." Basic Concepts of Speech,
http://cmusphinx.sourceforge.net/wiki/tutorialconcepts,
March 4, 2015, accessed November 17, 2015

[DSP] TDA7590. STMicroelectronics,
http://www.st.com/st-web-ui/static/active/jp/resource/technical/document/datasheet/CD00103223.pdf
23 Sept. 2013. PDF. Web 12 Oct. 2015

[FAQ] "CMUSphinx." Frequenty Asked Questions (FAQ) [Wiki].
http://cmusphinx.sourceforge.net/wiki/faq ,
27 Mar. 2016. Web. 16 Nov. 2015.

[IOGEAR] "Wireless HDMI Transmitter and Receiver Kit." *IOGEAR*.
https://www.iogear.com/product/GWHD11/
Web. 08 Mar. 2016.

[NUC] "Mini PC Intel® NUC Kit NUC6I5SYK." Intel.
http://www.intel.com/content/www/us/en/nuc/nuc-kit-nuc6i5syk.html
Web. 05 May 2016.

[Pi] "Raspberry Pi 2 Model B." Raspberry Pi Raspberry Pi 2 Model B Comments.
https://www.raspberrypi.org/products/raspberry-pi-2-model-b/
Web. 12 Oct. 2015.

[Proj] "PJD5155 Standard Resolution 4:3 (SVGA) 3,300 Lumens Value Business Projector - Projector." - Products.
http://www.viewsoniceurope.com/uk/products/projectors/PJD5155.php
Web. 08 Mar. 2016.

[Tool] "CMUSphinx." Overview of Toolkit,
http://cmusphinx.sourceforge.net/wiki/tutorialbeforestart ,
December 09, 2014, accessed November 16. 2015.

[Yeti] "YETI." Blue Microphones.
http://www.bluemic.com/products/yeti/
Web. 12 Oct. 2015.

10. Approved By

Advisor Name ___________Samee Khan, Ph.D.________

Advisor Signature ____________________________________ Date ___________________