Northwestern Polytechnic University

Squanto

Capstone Project Final Report

Under the Guidance of

Prof. A. Banafa

Submitted By

Bhakti Jivani(16578)
Hardik Mistry (16100)
Shyam Bhimani (16246)
Maulik Upadhyay(16577)

TABLE OF CONTENT

CHAPTER 1: INTRODUCTION

1.1 Purpose

Speech recognition technology is one from the fast growing engineering technologies. Nearly 20% people of the world are suffering from various disabilities; many of them are blind or unable to use their hands effectively. They can share information with people by operating computer through voice input. Our project is capable to recognize the speech and convert the input audio into text. For text to speech voice synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech computer or speech synthesizer, and can be implemented in software or hardware products.

Highlighting the main features of Squanto:

- Squanto is capable to convert Speech into text.
- It will also convert text into Speech.

1.2 Project Summary

"This is a proposal for a demonstration project which will seek to use an especially smart form of convert text to speech and speech to text".

1. Able to write text both voice and voice.
2. Lower operational cost.
3. Provide significant help for the people with disabilities.
4. Require less consumption of time in writing text.
5. Make pro reading effective.
6. Help with seniors who have vision problems.
7. allowing multi-tasking so that attention can be given to reading materials when time would otherwise not permit (such as during washing dishes, waiting in line, commuting, walking)

8. extending one's reading capacity when the body is too tired to visually read (even if the eyes are too tired to read, one is still capable of listening strategically)

1.3 Market Research

- According to the new market research report "Speech Recognition Market", the speech recognition market is expected to grow from USD 3.73 Billion in 2015 to USD 9.97 Billion by 2022, at a CAGR of 15.78% during the forecast period.

1.4 Methodology

- Project scope this project has the speech recognizing and speech synthesizing capabilities, though it is not a complete replacement of what we call a SIRI but still a good app for making notes using voice recognition and save it in your device.

CHAPTER-2 AIMS & OBJECTIVES OF THE WORK

2.1 Aims of the Study

The study aimed to study the feasibility of the Internet as a medium for synchronous delivery of information of subjects. The main areas of interest were audio-only, as well as video and audio transmission. In summary, the specific objectives of the project were to:

• Conduct a feasibility analysis into the development of a Speech to text and text to voice systems.

• Some people of the world are suffering from various disabilities; many of them are blind or unable to use their hands effectively. They can share information with people by operating Smart Phone through voice input.

2.2 Objective

The main objective of the Squanto is to provide speech to text and text to voice conversion.

By using the Project Squanto a mobile application can translate audio speech to text. The service supports the use of multiple grammars.

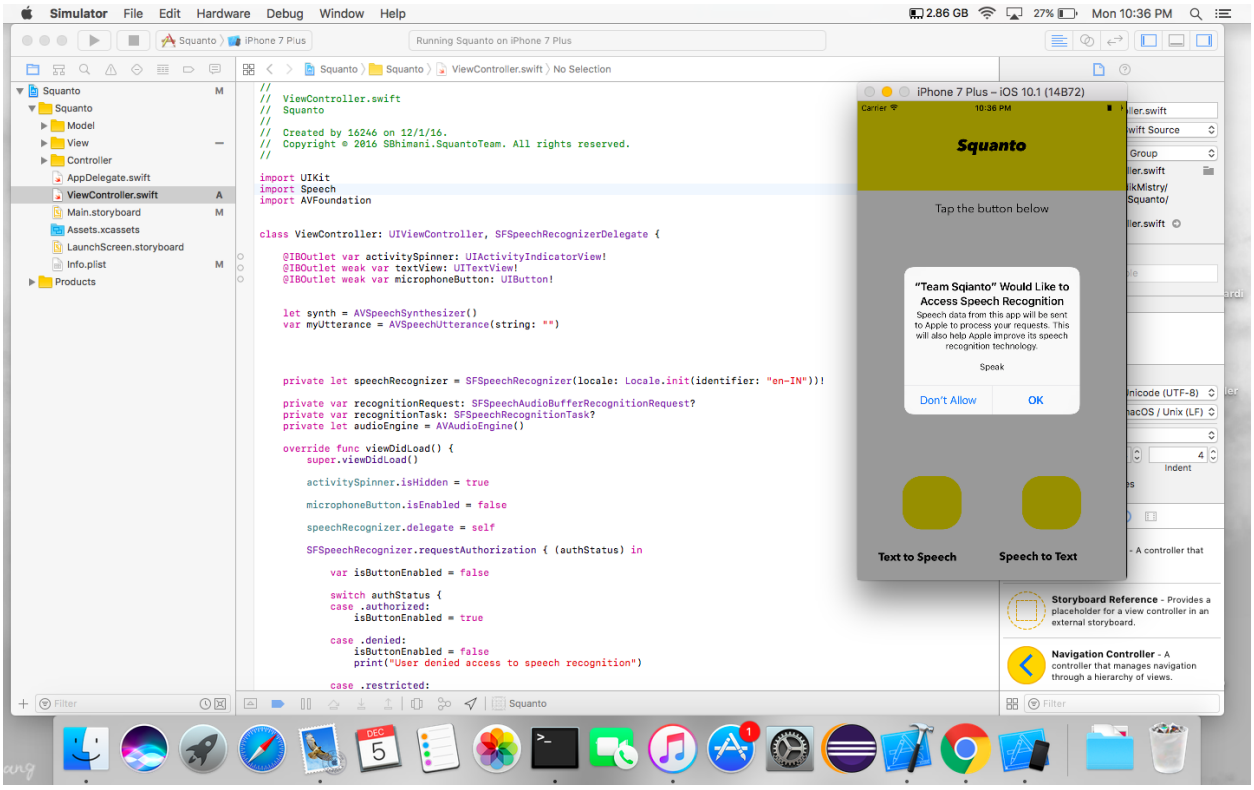
Text-to-speech synthesis -TTS - is the automatic conversion of a text into speech that resembles, as closely as possible, a native speaker of the language reading that text. Text-to speech synthesizer (TTS) is the technology which lets computer speak to you.

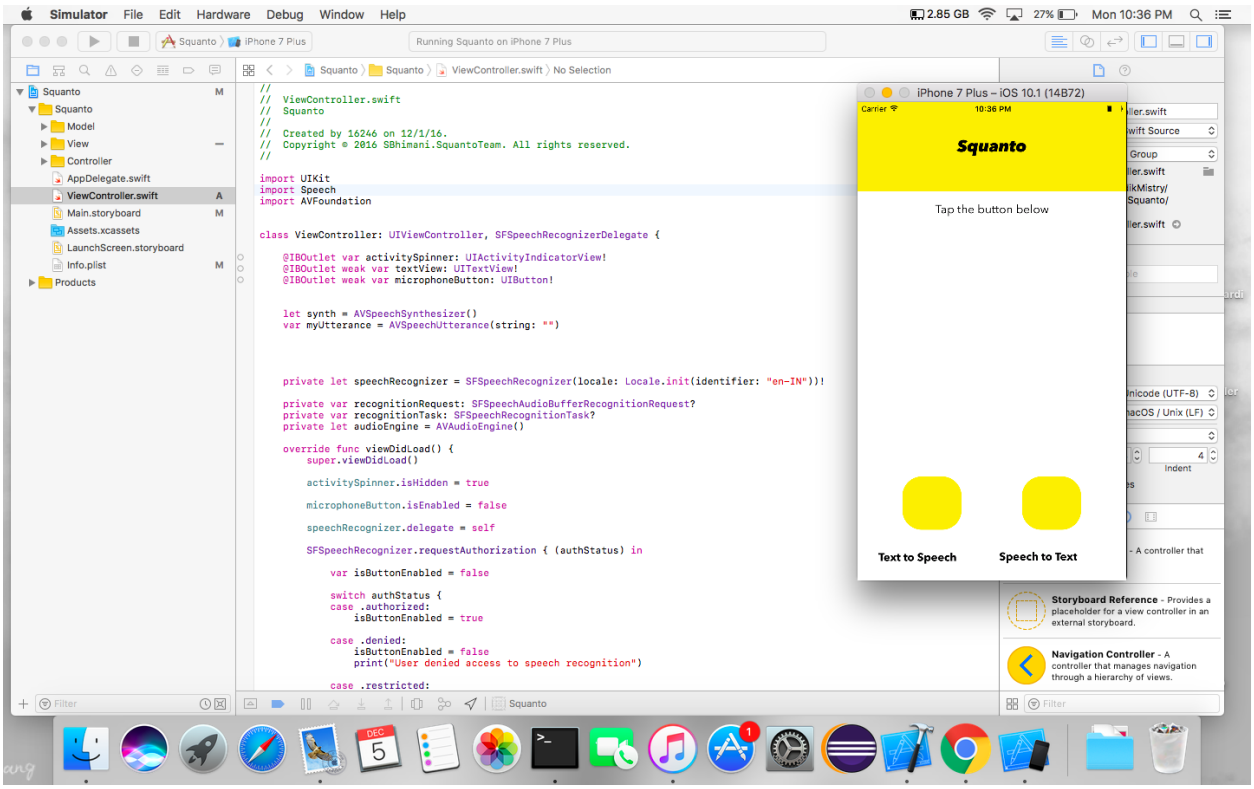
CHAPTER-3 Step By step Flow of project

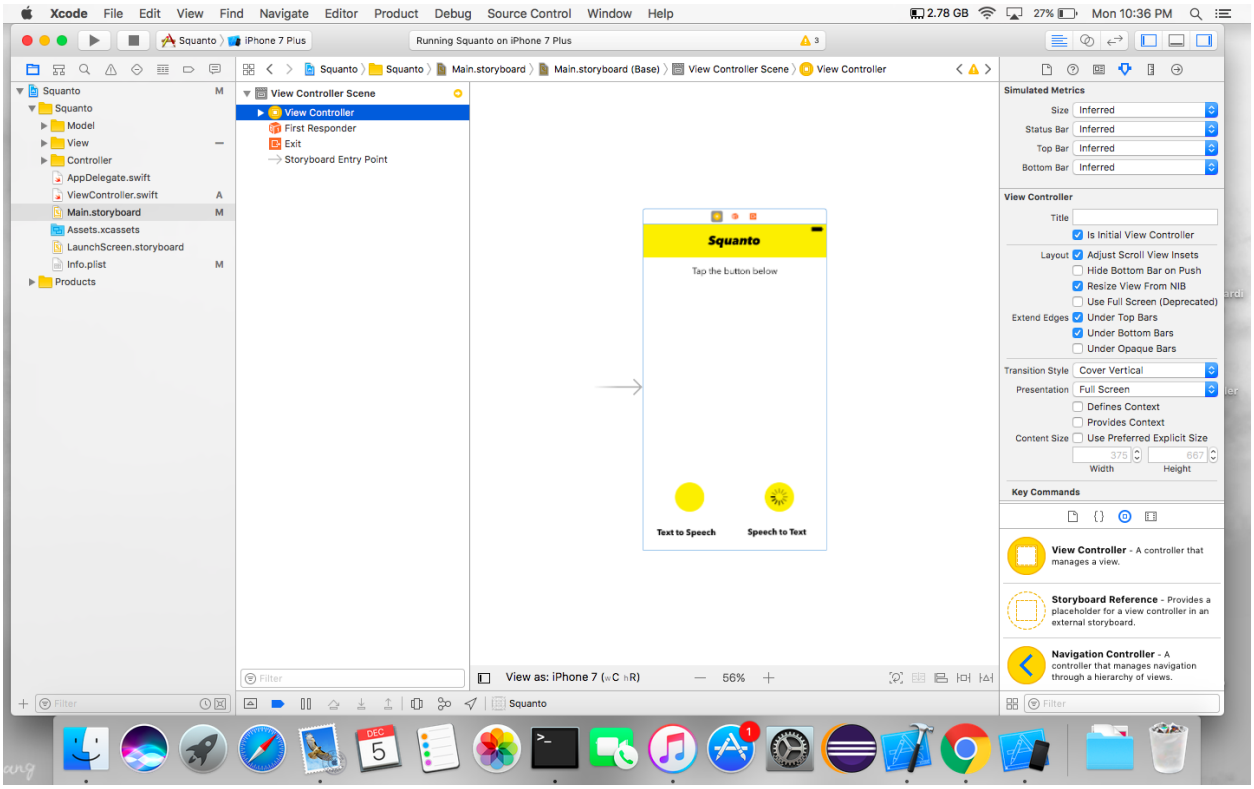
- ## Designing The App UI

For our application we create a new iOS Single View Application project with the name SpeechToTextDemo. We first go to your Main.storyboard and add a UILabel, a UITextView, and a UIButton for creating a user interface design of our application.

Our storyboard should look something like this:




Simulator File Edit Hardware Debug Window Help
iPhone 7 Plus – iOS 10.1 (14B72)
Squanto
Tap the button below
Text to Speech
Speech to Text


Xcode File Edit View Find Navigate Editor Product Debug Source Control Window Help
Squanto
Tap the button below
Text to Speech
Speech to Text

- **Using Speech Framework**
  We use the Speech framework for our application , first we have to import it and adopt the SF SpeechRecognizer Delegate protocoln and add its protocol to the ViewController class.

- **User Authorization**

  Before using the speech framework for speech recognition, we have to first ask for users' permission because the recognition doesn't happen just locally on the iOS device but Apple's servers. All the voice data is transmitted to Apple's backend for processing. Therefore, it is mandatory to get the user's authorization.

- **Providing Authorization Messages**

  Apple requires all the authorizations to have a custom message from the app. In case of speech authorization, we must authorize two things:

  1. Microphone usage.

  2. Speech Recognition.

  At the end run hit the Run button, you should be able to compile and run the app without any errors.

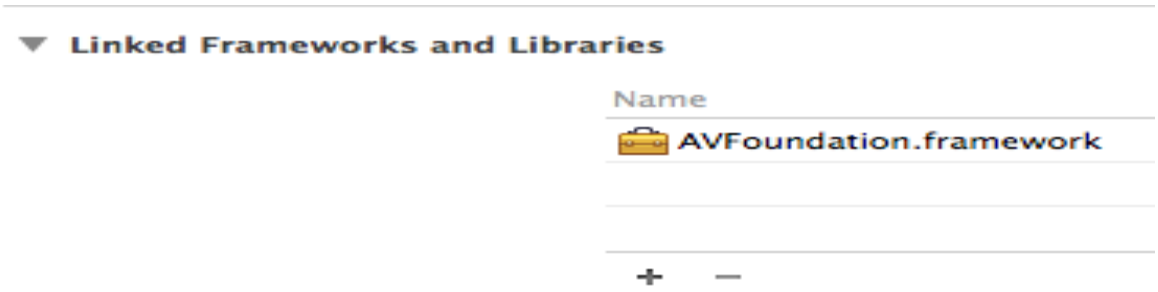
- **Handling Speech Recognition**
  Now that we have implemented the user authorization, let's move onto the implementation of speech recognition. This object handles the speech recognition requests. It provides an audio input to the speech recognizer.

  The recognition task where it gives you the result of the recognition request. Having this object is handy as you can cancel or stop the task.

  We use audio engine. It is responsible for providing your audio input and we create startRecording() function.This function is called when the Start Recording button is tapped. Its main function is to start up the speech recognition and start listening to your microphone.

- **Import AVFoundation**

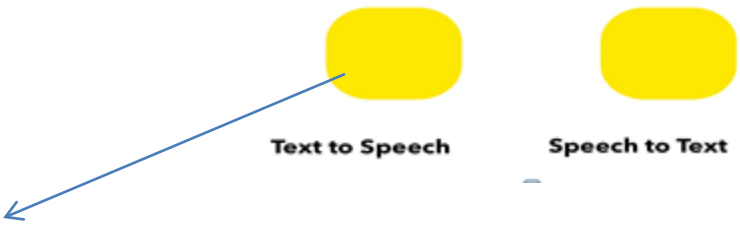
  To use the text to speech API, we'll need access to the AVSpeechSynthesizer class, which is part of the AVFoundation framework. Select your project in Xcode and go to the Linked Frameworks and Librariessection at the bottom of the General tab. Click the plus button and select the AVFoundation framework from the list.



- **Adding an Action on text to speech button**

  The button in the user interface should trigger a method named textToSpeech. To accomplish this, we need to create an action using the @IBAction type qualifier. By using the @IBAction type qualifier, the method will show up in Interface Builder as an action. Let's see how this works.

  Open the Assistant Editor and make sure the XIB file is shown on the left and the ViewController class on the right. Select the button in Interface Builder and control-drag from the button to the ViewControllerclass on the right as shown below.

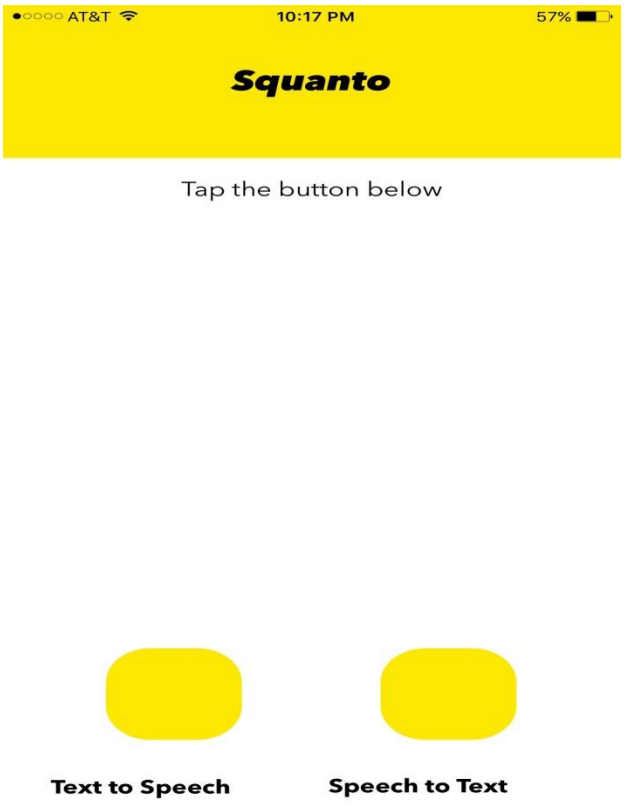


Action

- **Triggering Speech Recognition**

  We need to make sure that speech recognition is available when creating a speech recognition task, so we have to add a delegate method to ViewController. If speech recognition is unavailable or changes its status, the microphoneButton.enable property should be set. For this scenario, we implement the availabilityDidChange method of the SFSpeechRecognizerDelegate protocol.

  The implementation as seen below.


AT&T
10:17 PM
57%
Squanto
Tap the button below
Text to Speech
Speech to Text

CHAPTER-4 WORKING ENVIORMENT

4.1 Hardware Requirement

The app will be hosted on App store and User can download app on their IOS device and can enjoy.

4.1.1 client side:

Apple device having IOS 10 or later.

4.2 Software Requirements

Front End Client:

IOS 10 or later version. Need High speed internet connection or Wi-fi in order to get better performance

Web Server:

Apple Server is being used in the whole Text-to-Speech, and Speech-to-Text conversation.

Data Base Server:

No Database required as it does not store any data.

Back End:

Swift, Xcode IDE

4.3 Technology used

For developing this system we will use need Xcode 8 beta and an iOS device running the iOS 10 beta and Swift we are also use Speech framework, a useful API for speech recognition. In fact, Speech Kit is the framework which iphone Siri uses for speech recognition.

Xcode 8:

Xcode is an integrated development environment (IDE) containing a suite of software development tools developed by Apple for developing software for macOS, iOS, WatchOS and tvOS.The latest stable release is version 8 and is available via the Mac App Store free of charge for OS X El Capitan and macOS Sierra users. Registered developers can download preview releases and prior versions of the suite through the Apple Developer website. However, Apple recently made a beta version of version 8.0 of the software available to those of the public with Apple Developer accounts.

Swift:

Swift is a general-purpose, multi-paradigm, compiled programming language developed by Apple Inc. for iOS, OS X, watchOS, tvOS, and Linux. Swift is designed to work with Apple's Cocoa and Cocoa Touch frameworks and the large body of extant Objective-C (ObjC) code written for Apple products.

Speech Framework:

The Speech APIs perform speech recognition by communicating with Apple's servers or using an on-device speech recognizer, if available. To find out if a speech recognizer is available for a specific language, you adopt the SFSpeechRecognizerDelegate protocol.

CHAPTER: 5 SYSTEM ANALYSES

5.1 Study of Current Application

Before deciding scope for our project, we have studied many existing Speech recognition systems and other text to speech converter application. This study helped us lot for knowing the various features of Squanto.

We have studied many application which convert speech in to text and text in to speech and noticed some special features of them, that are listed as follows:

1. **Dragon Dictation [IOS app]**
   Dragon Dictation is an easy-to-use voice recognition application powered by Dragon NaturallySpeaking that allows you to easily speak and instantly see your text or email messages. In fact, it's up to five (5) times faster than typing on the keyboard.

2. **iSpeech - Text to Speech**
   Type in text and listen with the iSpeech App. iSpeech converts text to speech with the best sounding voices anywhere. Simply enter any text and iSpeech will instantly read the words.

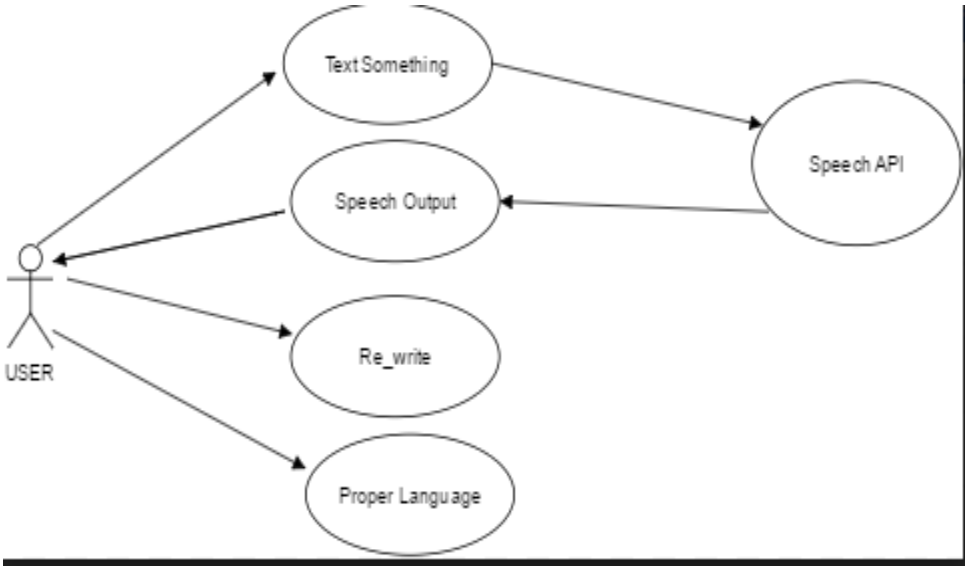
3. **Voice Brief (iPhone)**
   Voice Brief is one of the best text to speech apps for iPhone users. Fully featured text to voice converter that is integrated with lots of features such as customizable contents, 4 types of high quality voices , built-in alarm , background listing option with multi tasking , built-in web browser and many more.

4. **Voice Reader Text To Speech (iPhone & iPad)**
   Voice Reader is a very useful text reading application for iPhone and iPad which helps you to listen your text documents in more than 21 languages with 32 voices. Great app to listen your documents in background with multi tasking feature.

5.2: Diagrams

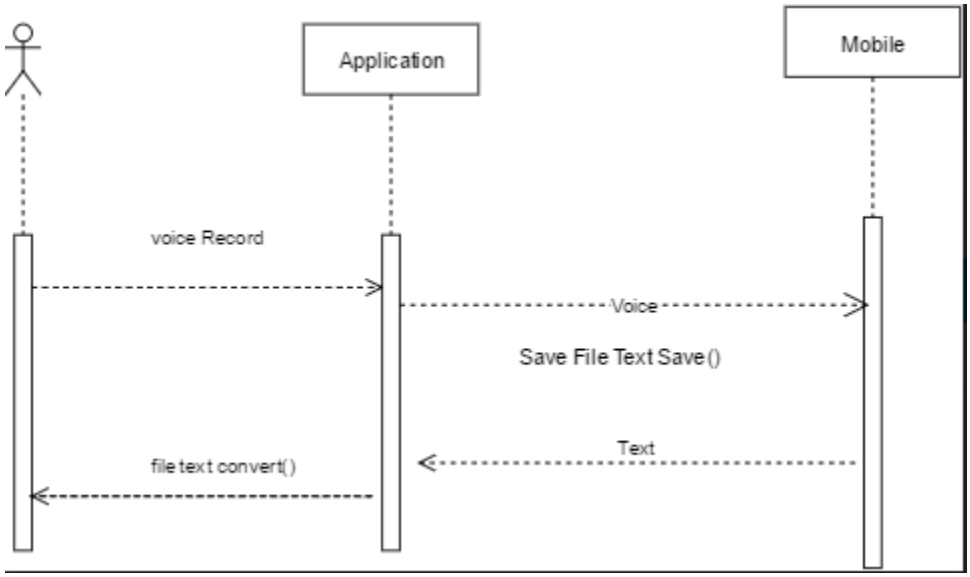
5.2.1 USE CASES

Use Case Diagram is a representation of a user's interaction with the system and depicting the specifications of a use case.

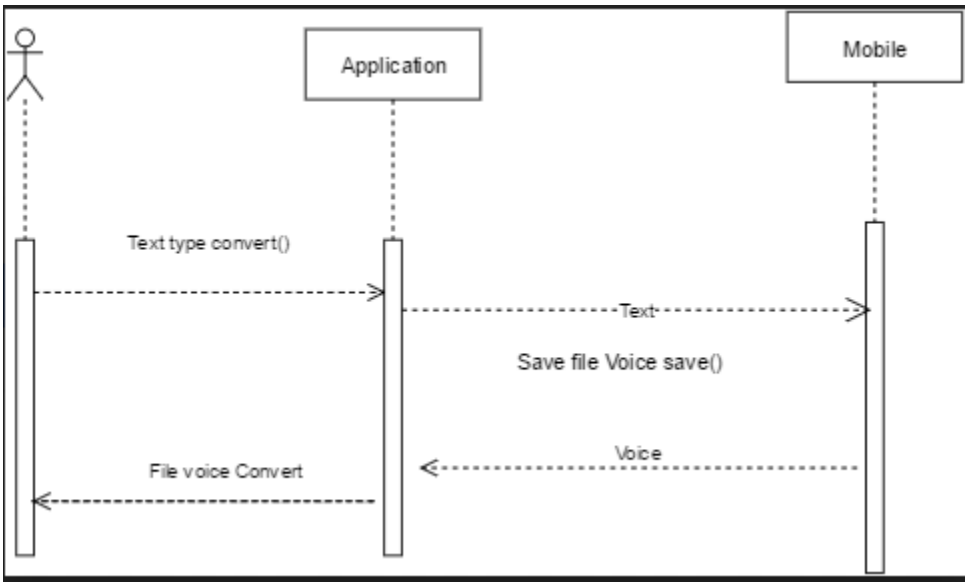


5.2.2: SEQUANCE DIAGRAM:

- Shows the static structure of the model.
- Collection of static modeling elements such as classes and their relationships connected as a graph.
- Provides visual representation of objects, relationships and their structures.
- **FIGURE: Speech to Text**



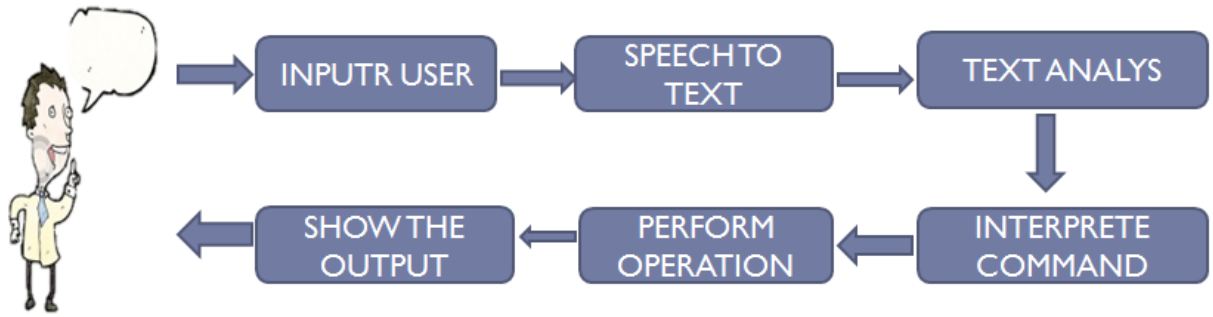
Text to Speech:



5.2.3: Flow Diagram

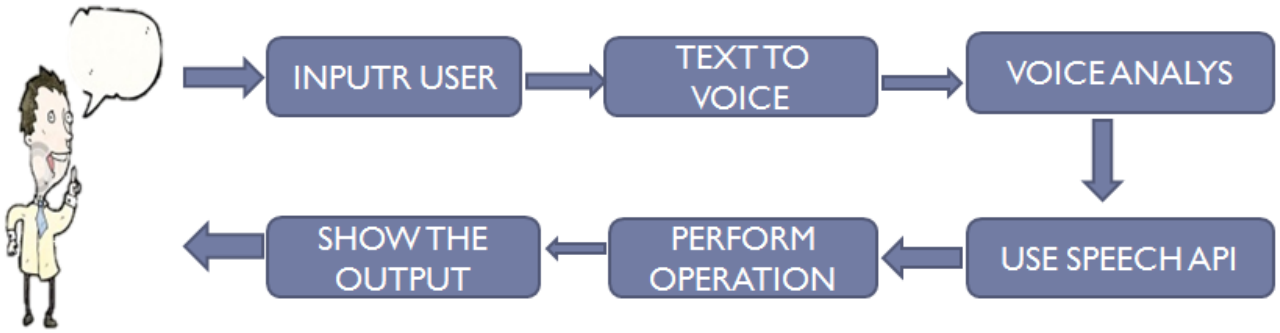
Flow Of Speech to Text:

This is a flow diagram of student. This shows which activity can done by user who is registered as student.



Flow of Text to Speech:

This is a flow diagram of student. This shows which activity can done by user who is registered as student.

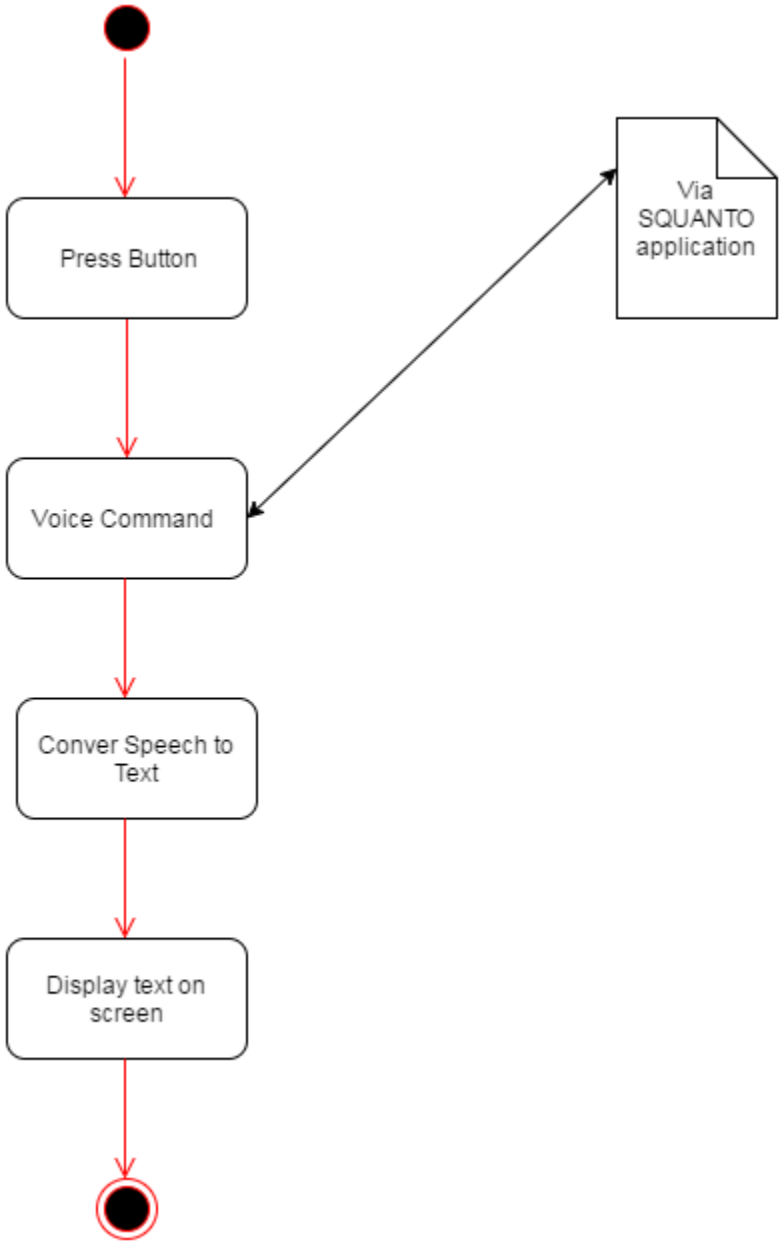


5.2.4:Activity Diagram

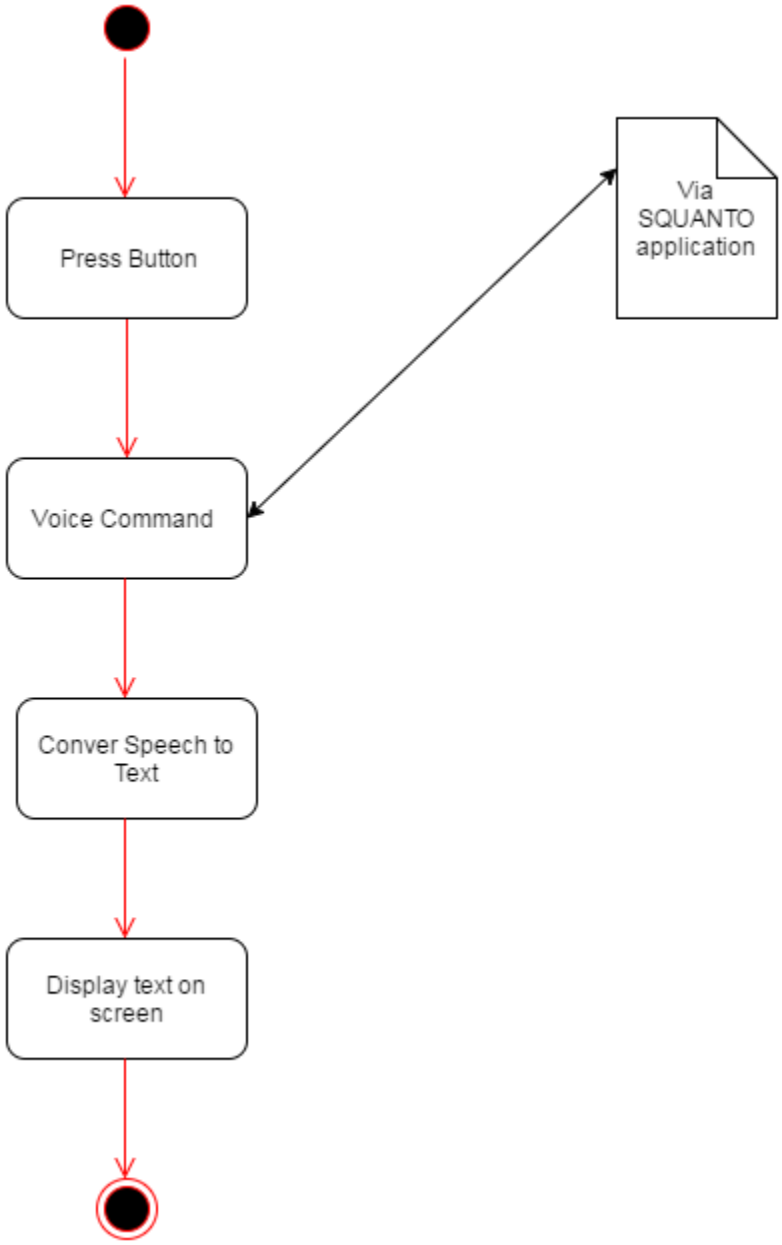
Activity diagrams are graphical representations of workflows of stepwise activities and actions[1] with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams are intended to model both computational and organizational processes (i.e. workflows) Activity diagrams show the overall flow of control.

Activity diagrams should be used in conjunction with other modeling techniques such as interaction diagrams and state diagrams. The main reason to use activity diagrams is to model the workflow behind the system being designed. Activity Diagrams are also useful for: analyzing a use case by describing what actions needs to take place and when they should occur; describing a complicated sequential algorithm; and modeling applications with parallel processes.

Speech to Text Activity Diagram:

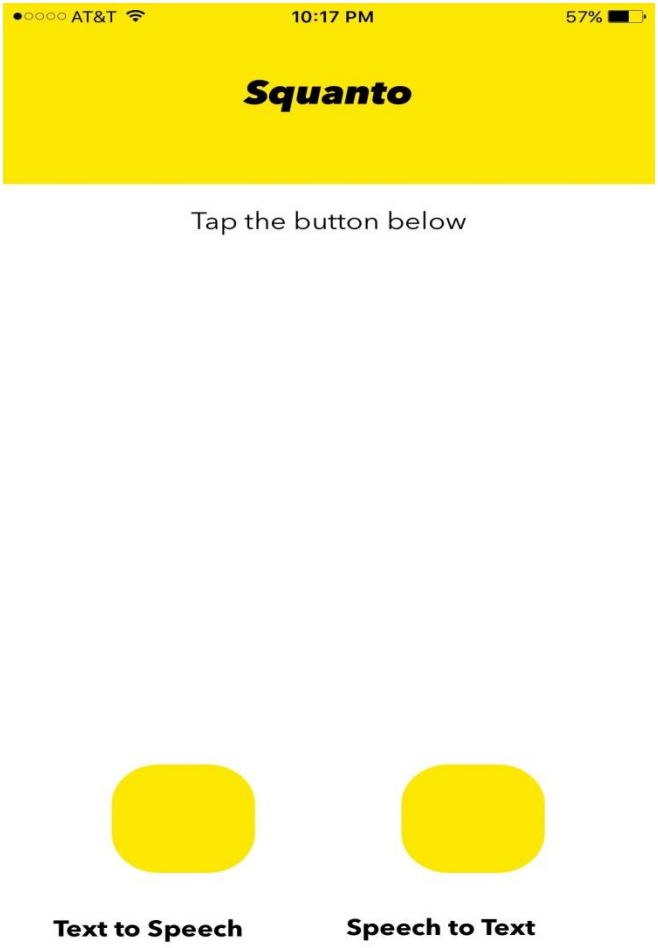


Text to Speech Activity Diagram:



CHAPTER-6: SCREEN SHORTS

6.1 Screen Shorts:



CONCLUSION

- As given in system specification, we have implemented most of the functionalities of the proposed Squanto. Our IOS app Squanto will be used for convert voice into text and text in to speech.
- Text to speech synthesis is a rapidly growing aspect of computer technology and is increasingly playing a more important role in the way we interact with the system and interfaces across a variety of platforms. We have identified the various operations and processes involved in text to speech synthesis. We have also developed a very simple and attractive graphical user interface which allows the user to type in his/her text provided in the text field in the application. Our application interfaces with a text to speech engine developed for American English.
- Speech recognition has a big potential in becoming an important factor of interaction between human and machine in the near future.

EXPEXTED OUTCOMES

- The Squanto iPhone application can be used to do a significant number of the things that you would do on iPhone, all without typing on the virtual keyboard, open a specific application, or explore any menus. With Squanto, everything you do is tap on the microphone icon and start talking. Squanto then translates your voice input and responds accordingly.

- This system can be used in Car System, Health care, Education, Telephony and other domains. Also the performance of speech recognition systems is usually evaluated in terms of accuracy and speed.

REFERENCES

1. Braina - Speech recognition software
   http://www.capterra.com/speech-recognition-software/spotlight/145652/Braina/Brainasoft

2. Apple siri
   https://developer.apple.com/ios/human-interface-guidelines/

3. Inspired- UI-mobile apps Design patterns

   http://inspired-ui.com/

4. ATMac-Using Assistive Technology to Empower Apple Users With Disabilities
   http://atmac.org/how-to-speak-to-your-idevice

5. Dragon-dictation ios mobile app
   https://itunes.apple.com/us/app/dragon-dictation/id341446764?mt=8

6. IOS app and IBM Watson
   https://www.youtube.com/watch?v=c4v2VyEgyF8

7. IOS 10 Speech recognition
   http://stackoverflow.com/questions/37959353/how-to-use-ios-10-speech-recognition

8. IOS 10 Speech recognition Tutorial

   http://www.techotopia.com/index.php/An_iOS_10_Speech_Recognition_Tutorial