

Android Speech to Text Converter for SMS Application

Ms. Anuja Jadhav* Prof. Arvind Patil**

*(Computer Science Y.C.C.E. Nagpur) India,

** (Computer Science Y.C.C.E., Nagpur) India

Abstract: The mobile phone users are increase day by day and everyone wants to develop new things which is improved version of existing one. The SMS user might want the faster SMS technology which can convert their voice into text and SMS will be transferred. This project based on evaluating voice versus keypad as a means for entry and editing of texts. In other words, we can say messages can be voice/speech typed. A speech to text converter is developed to send SMS .It is found that large-vocabulary speech recognition can offer a very competitive alternative to traditional text entry.

Keywords: Short Message Service (SMS); Speech to text conversion (STT) User Interface (UI); Neural Network (NN)

1. INTRODUCTION

People love their mobile phones because they can actually stay in touch wherever they are. That means not just for talking, but e-mailing, texting, and so on. We are constructing with the growth of mobile phone technology. As per the users are increase day by day, facilities are also increasing. Let's start with simple regular handsets which were used just for making phone calls, mobiles have changed our lives and have become part of it. Now a day's mobile phones are not use for making calls but they have innumerable uses and can be used as a Camera , Music player, Tablet PC, T.V. , Web browser etc . And with the new technologies, new software and operating systems are required. In recent years, smart phones have placed an increasing emphasis on bringing speech technologies into mainstream usage. This focus has led to products such as Speech server, which is used to implement speech-enabled telephony systems. Now let's limit our focus towards short message system it is text messaging service component of phone, using standardized communications protocols that allow the exchange of short text messages between mobile phone devices. SMS text messaging is the most widely used data application in the world, with 2.4 billion active users, or 74% of all mobile phone subscribers. The term SMS is used as a synonym for all types of short text messaging as well as the user activity itself in many parts of the world. SMS is also being used as a form of direct marketing known as SMS marketing. There is no doubt that more and more Mobile phone users are using short message service (SMS) instead of making voice calls. To satisfy the demands of users, mobile phone manufacturers are constantly adapting and innovating to ensure that they can survive in this competitive market. The important innovation in SMS technology is the speech recognition technology that can convert voice messages into text messages. In simplified words messages can be voice/speech typed.

As we all know cell phones are very important part of modern life. Many of us need to make a urgent call or message at anytime from anywhere. Many of them needs their cell phones when they can't do so e.g. At the time of driving, cooking accidents may occur because of this activity an speech to text converter for mobile design for

this purpose so to avoid accidents. Android is the best suitable operating system for this kind of system. It is the open source operating system that is use to develop application for mobile users.

2. Android Platform

Operating system has developed a lot in last 15 years. Starting from black and white phones to recent smart phones or mini computers, mobile OS has come far away. One of the most widely used mobile OS these days is ANDROID. Android does a software bunch comprise not only operating system but also middleware and key applications. After original release there have been number of updates in the original version of Android. It is the software stack of mobile devices. Android SDK provides the API's that is necessary to begin developing applications on the Android platform using the Java programming language. Android includes an embeddable browser built upon WebKit, the same open source browser engine powering the iPhone's Mobile Safari browser.

An Android application consists of one or more of the following classifications: 1) Activities: Is the application that has a visible UI is implemented with an activity. When a we selects an application from the home screen or application launcher, an activity is started.2) Services: A service should be used for any application that needs to persist for a long time, such as a network monitor or update-checking application. 3) Content providers: We A content provider's job is to manage access to persisted data, such as a SQLite database. Suppose for the bigger system like Speech to text conversion system or one that makes data available to multiple activities or applications, a content provider is the means of accessing your data.4)Broadcast receivers: This is use to launched to process a element of data or respond to an event, such as the receipt of a text message. Following figure showing the Architecture of Android platform. The key application of the speech to text conversion system is recognition technology. Speech input adds another dimension to the mobile phones. The Google's Voice Search application, which is required, is many times pre-installed on many Android devices and available in Android Market, and it provides powerful features. Now we can dictate our message instead of typing it. Just tap the new microphone

button on the keyboard, and you can speak in just about any context in which you would normally type.

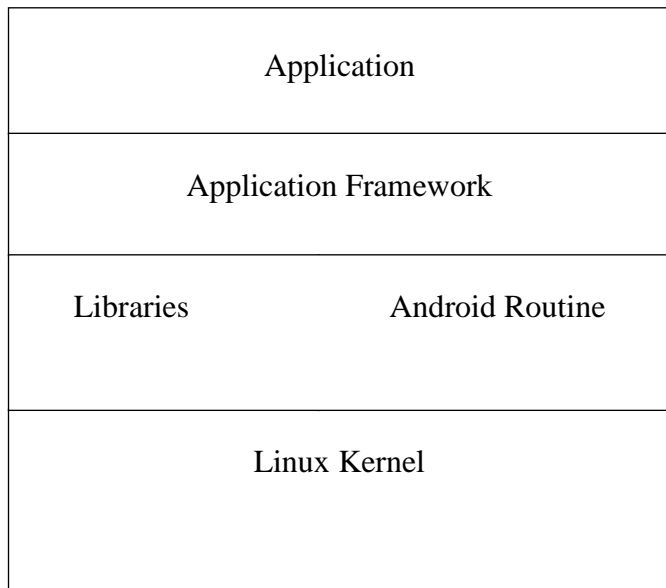


Fig1: Android Architecture

3. A Speech to Text Conversion System

Elements of Speech to Text

- 1) Speech Acquisition
- 2) Speech to text conversion
- 3) SMS transmission

This system allows us to give our voice as input and produce text as an output. Systems use "training" where an individual speaker reads sections of text. These systems analyze the person's specific voice and use it to fine tune the recognition of that person's speech, resulting in more accurate transcription. The speech to text conversion system is the ability of smartphone to identify the words which is in the spoken language and translation of it into the readable form that is in the form of text. Many speech recognition software available in the market which is available for limited vocabulary of words or if the words spoken very clearly.

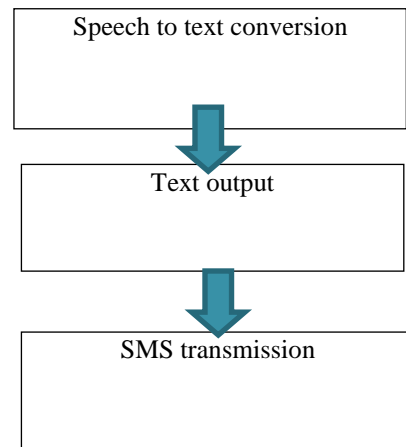
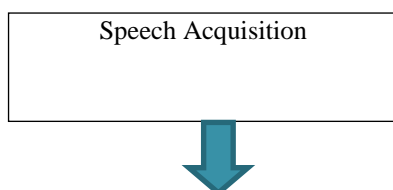


Fig 2: showing the elements of speech to text conversion system

4. Speech Acquisition

Every speech-to-text conversion system require acquisition of speech at the real time During speech acquisition, speech samples are obtained from the speaker in real time and stored in memory for pre-processing. Speech acquisition requires a voice recorder in mobile phone that has the proper amplification to receive the voice speech signal, sample it, and convert it into digital speech. To achieve speech pre-processing sphinx frame work is used this is the best tool found to acquiesce speech signals. Sphinx is design with high flexibility modularity. To achieve speech pre-processing sphinx frame work is used this is the best tool found to acquiesce speech signals. Sphinx is design with high flexibility modularity. Figure2 shows the overall architecture of sphinx. There are three modules in the sphinx frame work frontend, decoder, Linguist Front end takes the input speech signals and parameterized it into sequence of features. The Linguist translates any type of standard language model, along with pronunciation information from the Dictionary and structural information from one or more sets of AcousticModels, into a SearchGraph. The SearchManager in the Decoder uses the Features from the FrontEnd and the SearchGraph from the Linguist to perform the actual decoding, generating Results. At any time prior to or during the recognition process, the application can issue Controls to each of the modules, effectively becoming a partner in the recognition process.

Result:Figure3 showing the result of speech acquisition where the speech input is taken at the real time and the text understandable by the systems shows as the output. Every computer system has its own voice recognition and hence it produce output accordingly for example if my input is Google but computer recognise it as "and/this/yes" that is not exact output for that we require some training to be done on the input data.

Sr. No.	input	Output
1	Google	and/this/yes
2	Yahoo	Down to/now
3	Fantastic	and to/ that

Table 1:- showing output generated by acquisition module

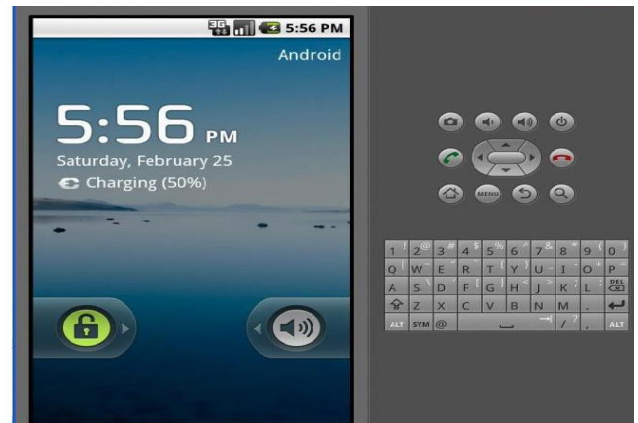


Fig4: showing android emulator

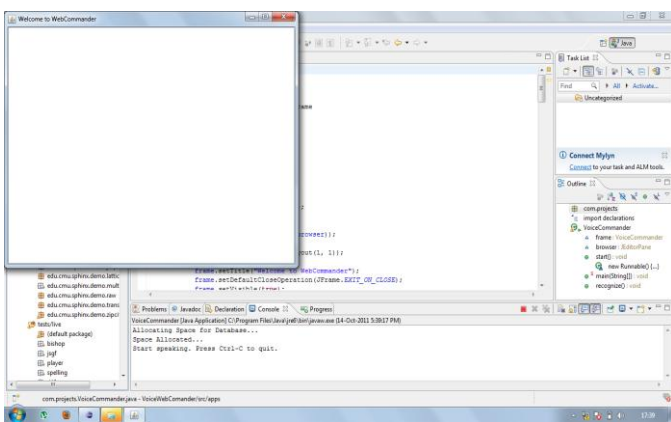


Fig3: showing the result of speech acquisition module

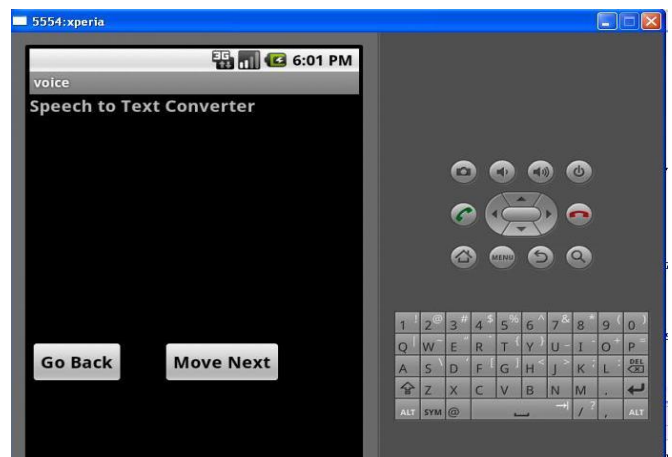


Fig5: the window where the speech input is collected

5. Speech to Text Conversion

The main requirement of every speech to text conversion system is a database which will compare peach with frequencies. If we develop the system which will convert the speech into text globally that is for any user it is very difficult job because the frequency of giving input of any user is different as that of other user. If the system is global hence we are creating it for the mobile user means our job is very much difficult there are millions and billions mobile users and sound frequency and peach comparison is again difficult. Hence we need to use database which is already provided by "Google".

If we try to create the database manually it will create time and space problem. To create such a huge database lot of time will west and this project is for mobile users whose internal memory of mobile is generally low and the database for this is system is very much huge. Instead of creating a database for this system we need to use existing database which is available at the web server.

The speech input is collected in the container and the send it to the peach and frequency comparison the recognition take place. To get correct text which is spoken we need training on input file for that neural network (NN) is used and the output is collected in the text file and SMS is ready to send.

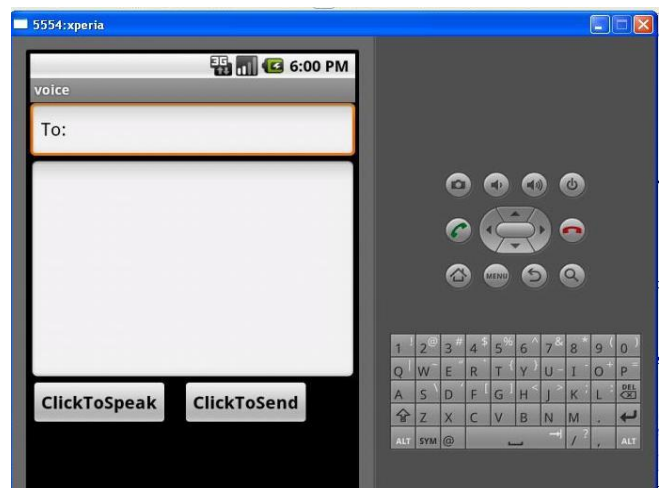


Fig6: click to speak button is created to collect the speech into container and give text as the output

6. SMS Transmission

During SMS sending phase the speech samples which is converted into text is send as an SMS to the other device. In this phase Android emulators are used two android emulators are used to send and receive the SMS. From the first emulator which we consider as mobile one text message is send and other device receive it at the inbox.

Here we use AndroidManifest.xml to add the two permissions - SEND_SMS and RECEIVE_SMS.

Next we have added here main.xml to get the phone number of sender. Next, in the SMS activity, we wire up the Button view so that when the user clicks on it, we will check to see that the phone number of the recipient and the message is entered before we send the message using the sendSMS() function, which we will define shortly.

To send an SMS message, we use the SmsManager class. Then we will call the getDefault() static method to obtain an SmsManager object. The sendTextMessage() method sends the SMS message with a PendingIntent. The PendingIntent object is used to identify a target to invoke at a later time. For example, after sending the message, you can use a PendingIntent object to display another activity.

In this case, the PendingIntent object (pi) is simply pointing to the same activity (SMS.java), so when the SMS is sent, nothing will happen. When an SMS message is sent, the first BroadcastReceiver's onReceive event will fire.

This is where we check the status of the sending process. The second PendingIntent object (deliveredPI) monitors the delivery process. The second BroadcastReceiver's onReceive event will fire when an SMS is successfully delivered.

Result:

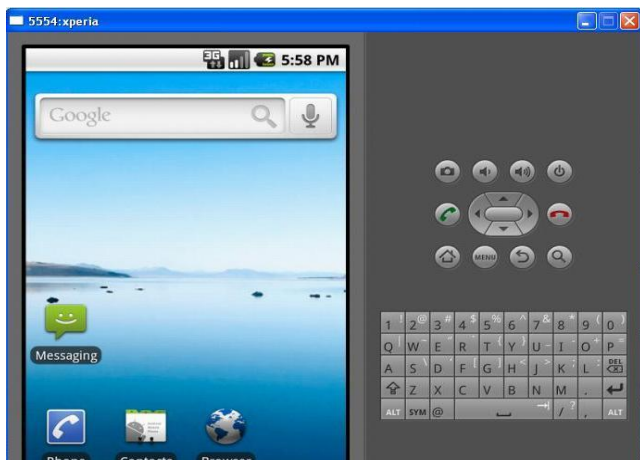


Fig7: Showing the result of sending text SMS

7. PERFORMANCE PARAMETERS OF THE SYSTEM

Depends upon following factors we can calculate the performance of this system

- 1) Accuracy of Recognition — the most important parameter in any recognition system is its accuracy. A recognition accuracy of 100% for all digits, independent of the speaker, is the goal.
- 2) Speed of Recognition — if the system takes a long time to recognize the speech, users would become restless and the system loses its significance. A recognition time of less than 1 second is required for the project.

8. CONCLUSION

This paper will demonstrate the requirement of speech to text conversion system and also it shows

how we can construct the STT for mobile users. When we are taking about mobile the main memory requirement we have consider here because we are using already existing database and that is safe and system will work faster this is the main advantage of this system.

REFERENCES

- [1] Andreas Stolcke, , Barry Chen, Horacio Franco, Venkata Ramana Rao Gadde, Martin Graciarena, , Mei-Yuh Hwang, Katrin Kirchhoff, , Arindam Mandal, Nelson Morgan, , Xin Lei, Tim Ng, Mari Ostendorf, Kemal Sönmez, Anand Venkataraman, Dimitra Vergyri, and Qifeng Zhu, "Recent Innovations in Speech-to-Text Transcription at Sricisi-uw" *IEEE Transactions On Audio, Speech, And Language Processing*, vol. 14, no. 5, september 2006, pp 1729-1744
- [2] Brandon Ballinger, Cyril Allauzen, Alexander Gruenstein, Johan Schalkwyk, "On-Demand Language Model Interpolation for Mobile Speech Input", *INTERSPEECH 2010, 26-30 September 2010, Makuhari, Chiba, Japan*, pp 1812-1815
- [3] Ryuichi Nisimura, Jumpei Miyake, Hideki Kawahara and Toshio Irino, "Speech-To-Text Input Method For Web System Using Javascript", *IEEE SLT 2008* pp 209-212
- [4] M. Tomalin, F. Diehl, M.J.F. Gales, J. Park & P.C. Woodland , "Recent Improvements To The Cambridge Arabic Speech-To-Text Systems", *ICASSP 2010* pp 4382-4385
- [5] Janet See, Umi Kalsom Yusof, Amin Kianpisheh, "User Acceptance towards a Personalised Handsfree Messaging Application (iSay-SMS)", *CSSR 2010 Initial Submission December 5-7,2010* pp 1165-1170
- [6] Panikos Heracleous, Hiroshi Ishiguro and Norihiro Hagita, "Visual-speech to text conversion applicable to telephone communication for deaf individuals" *18th International Conference on Telecommunication 2011*. pp 130-133
- [7] Y. Liu, E. Shriberg, A. Stolcke, D. Hillard, M. Ostendorf, and M. Harper, "Enriching speech recognition with automatic detection of sentence boundaries and disfluencies," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 5, pp. 1524–1538, Sep. 2006.
- [8] M.J.F Gales, F. Diehl, C.K. Raut, M. Tomalin, P.C. Woodland, and K. Yu, "Development of a phonetic system for large vocabulary arabic speech recognition," in *Proc. of ASRU, 2007*.
- [9] L. Nguyen, T. Ng, K. Nguyen, R. Zbib, and J. Makhoul, "Lexical and phonetic modeling for arabic automatic speech recognition," in *Proc. of Interspeech, 2009*.
- [10] C.C. Wong, "Enabling Ecosystem for Mobile Advertising in an Emerging Economy," Monash University Doctoral Colloquium. *Langkawi, Kedah, Malaysia, 14-16 December 2009*.