International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

A Pilot Research on Android Based Voice Recognition Application

Ajins Joy, R. Saranya

Abstract. In recent trend, Speech recognition has become extensively used in customer service based organization. It has acquired great deal of research in pattern matching employed machine learning (learning speech by experience) and neural networks based speech endorsement domains. Speech recognition is the technology of capturing and perceiving human voice, interpreting it, producing text from it, managing digital devices and assisting visually impaired and older adults using unequivocal digital signal processing. In this paper we have presented a comprehensive study of different methodologies in android enabled speech recognition system that focused at analysis of the operability and reliability of voice note app. Subsequently we have suggested and experimented an android based speech recognizer app viz. Annotate which predominately focus on voice dictation in five different languages (English, Hindi, Tamil, Malayalam and Telugu) and extracting text from image using Automatic Speech Recognition (ASR) and Optical Character Recognition (OCR) algorithm. Finally, we identified opportunities for future enhancements in this realm.

Keywords: SpeechRecognizer . RecognizerIntent . Amazon app . Sqlite database . Voice note.

I. INTRODUCTION

Nowadays smart gadgets have paramount impact in the present day regular habitual life in which electronic touch screen technology is an exceptional invention of it [1].When it comes to human conveyance, perception of speech takes highest priority than other senses [7,8].In this regard Speech Recognizer mechanism is one of the contemporary accomplishment in an advanced technological development field that authorize an analog signals to pertinent text, reply to the spoken instructions through decision making feature/pattern process identification and cognitive speech articulation services. There are numerous speech-empowered/incorporated online notepad / voice notes applications were deigned and available in the market, yet the challenging aspect is providing the description to the text heading is still an arduous task.

In this paper we strive to provide a recommendation kind of mobile application viz. 'Annotate' means adding note to given commands - an android application where the user can simply add the description by their own voice (uses only3.2 MB phone memory) . We have assimilated Automatic Speech Recognition (ASR) module (hardware and software-based technique to recognize and analyse human voice [9,10,11] to identify the word/ words a person has spoken with an additional feature of storing the information / content (electronically i.e., image to text) available in the paper document and later reusing it using Object Character Recognition (OCR) java classes (JDK 8) by a single click and also we have included text to voice support for English language by implementing oxford dictionary API in Android Studio IDE. The road map of this paper is organized as follows: Section 2, we presented literature review of speech recognition and voice app currently available in the market; Section 3, we reported the pilot study that delineates the information and illustrations of ingrained technologies related to the analysis of the app; finally, we conclude this interpretation work with brief note on its future applications.

Manuscript published on November 30, 2019.
* Correspondence Author
Ajins Joy*, JRF (Junior Research Fellow) in the Department of School of Computer Science and Engineering at Vellore Institute of Technology (VIT), Chennai, Tamilnadu, India. (Email: ajinsjoy@gmail.com)
Dr.R. Saranya, Assistant Professor, Department of Computer Science, Central University of Tamil Nadu, Thiruvarur, Tamilnadu, India. (Email: saranya@cutn.ac.in)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

II. LITERATURE REVIEW

In 2012, Jianliang Meng et al. [1] presents a run-through of variegated speech based pattern recognition modus operandi and technology. The methods like Hidden Markov Models (HMM) with underlying hidden parameters of a time varying signal were statistically analysed for unambiguous voice pattern outcome , but this approach capitulate the limitation of longer training sequence of given input sequence. Furthermore artificial neural network based speech (auditory) recognition of dynamic voice signal were examined by extricating different characteristics of speech signal and many knowledge sources like phonemes, vocabulary, etc. used to validate segmented voice input against neural network learning parameters. The advantage of this approach is it assists the user in a state of being convenient recognition system. Besides that significant applications like VRCP and financial based stock quotation speech recognition system were briefed with deficiency of conventional applicability of substandard dependence adaptation and manipulating speech in noisy environment.

In 2013, Nicholas Mulhern et al. [2] has analysed and furnished an accurate android based voice controlled command (an assistive technology) which lend a helping hand to neuromuscular disordered people.

They have traversed various voice recognition (offline-PocketSphinx) keyword commands as a template and mentioned the use of Eclipse SDK and PIC microcontroller which send signals to a smart phone via. Bluetooth exchange to run a general television remote control without a need to remunerate this device which enhances and empower individual independence in household domains.

In 2014, Amruta Pise et al. [3], used new text recognition detector (zone based feature extraction technique) for natural scene images. The methods like combined Image Centroid Zone and Zone Centroid Zone used for feature (character) extraction to localize normalized coordinates (height and width ratio) of text line. Similarly for classification and recognition feed forward propagation algorithm were used at the Pre-Processed image which is inclined at distance metric feature systems using MATLAB. The primacy of this paper faster text extraction and recognition amid different font sizes and multiplex background. In the year 2014, Kamrul Hasan Talukder et al. [4], presents text extraction from colour image using Connected Component Based Approach (conversion from grayscale to binary image). Extraction of text from image involves detection, localization, extraction, enhancement and recognition. This approach helps to extract large text from image more accurately. By using this method, we can extract text from image more accurately in short time. But the limitation of this approach is, it become less productive when the text is too small.

In 2014, Gunta Chomchalerm et al. [5], has implemented an android application viz. 'Braille Dict' a Thai language based database (using LEXiTRON based SQLite) for visually impaired individuals. It uses Braille keyboard to give the English word to this application as input and yield Thai language meaning with TTS feature (VAJA engine). This approach has given 90% performance evaluation braille output (knowing the meaning with reduced time). The drawback of this approach is few issues like Multi-Touch motions and signs (gestures) are not resolved.

Table 1 : Summary of Literature Survey

ALGORITHM/ METHODOLOGY USED	WHO, WHEN	ADVANTAGES
Hidden Markov Model (HMM)	Jianliang Meng, Junwei Zhang,Haoquan Zhao (2012)	Underlying hidden parameters of a time varying signal were statistically analysed for unambiguous voice pattern outcome.
Voice Controlled Command	Nicholas Mulhern, Neil McCaffrey, Nicholas Beretta, Eugene Chabot PhD, Ying Sun PhD (2013)	An assistant technology that human can control everything with using their own voice.
Kohonen neural network	Pratik Madhukar Manwatkar and Mr. Shashank H. Yadav (2015)	Map to read the individual contents and searching the contents from these documents line by line and word by word.
Histogram of Oriented Gradients (HOG)	Amruta Pise and S.D.Ruikar (2014)	The object detection obtained using HOG. It counts occurrences of gradient orientation in localized portions of an image.
Text-to-Speech(TTS)	Gunta Chomchalerm, Jitsupha Rattanakajornsak, Unchalee Samsrisook Damras W ongsawang and W orapan Kusakunniran (2014)	This technology uses text will be spoke by the system
Optical Character Recognition(OCR)	Sathiapriya Ramiah, Tan Yu Liong and Manoj Jayabalan (2015)	OCR is very quicker and accurate. Reduce the time taken, that's mean it work quicker and more efficiently.
And Multiclass support Vector Machine(SVM)	Nidhi Kalidas Sawant Prof. Sangam Borkar (2018)	The editable text will be extracted from the image using the multiclass support vector machine

In 2015,Pratik Madhukar Manwatkar et al. [6], deals with automatic electronic text recognition analysis from (paper) document images. The authors have used text recognition module for preliminary processing (grayscale and binary conversion of scanned paper document), training by segmentation (dividing image into objects) and matrix based pattern (feature) extraction techniques. They have suggested Kohonen neural network Self-Organizing feature map to read the individual contents and searching the contents from these documents line by line and word by word. The disadvantage of this approach is inappropriate for certain characters in the paper documents are different to the font of the character in the computer system.

In 2015, Sathiapriya Ramiah et al. [7], proposed approach for recognizing text from image using coherent Tesseract OCR engine with Bing text to speech feature. In this method, the captured image is Pre-Processed (converting to grayscale) to remove noise, then it is redirected to OCR engine to acknowledge the text from the image (by extraction using open source library Leptonica). This approach is associated with Bing translator voice over text (English translation). The experimental result shows 95% accurate functionalities of this approach which assists both blind users and travellers. The only drawback is that this application does not responded for all the diversified languages.

In 2018,Nidhi Kalidas Sawant et al. [9], presents a SVM(Support Vector Machine) classifier (for editable text) based TTS conversion for Indian Devanagari Script Behaviour of Marathi language printed text. The authors have dealt with TTS based character segmentation of OCR system and HoG (Histogram of Oriented Gradients) image feature descriptor detection algorithm for transforming the printed image to text and speech output. In the year 2018,D.S.S. De Zoysa et al.[10], suggests to use mobile based audio assisted Optical Character Recognition(OCR) and Text to Speech System (TTS) configured absolute framework viz. "Bhashitha" in Sinhala language for normal and abnormal visual acuity people. Here Tesseract engine for processing different background real time Sinhala text images documents and CLUSTERGEN Statistical Parametric Synthesizer were used to perceive and train the content of the documents. Such approach minimize the workload of the user and also used to read out with the help of high linguistic accuracy and read long texts fluently. But this TTS doesn't know how to deal with the text features and some symbols like bullets or dashes and in addition OCR text works only with the printed text not for the handwritten text.

III. INGRAINED TECHNOLOGIES (SIMULATION OF ANNOTATE) & RESULTS

Various process modules are involved in implementing the idea of Annotate- A speech recognizer application using Android 8.0 (Oreo). In this paper we have focused on the following:

A.ASR deployed voice dictation via. Five different languages with Text to Speech reader and English to English Dictionary [1, 2, 3, 6]

In this module, (Automatic Speech Recognizer (ASR) Feature is been called) the user need to perform the access permissions (e.g. authorizing microphone) to do a specific action (i.e. to acquire a feature of this app). Subsequently the user will be instigated with two options viz. Voice Note and English to English Dictionary in main page of the app as given in the below figures.

Fig. 1.a. Main Page

Fig.1.b. Language selection

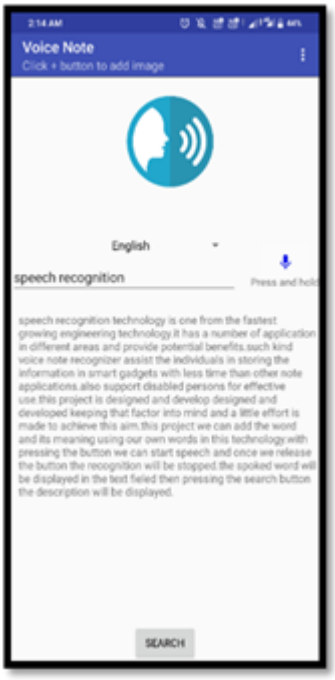

Fig 1.c. Accessing saved note

If the user chooses Voice note, the reinforcing feature is that they can add a note in three different vernacular languages of South India (i.e. Malayalam, Tamil, and Telugu) in addition to English and Hindi by pressing add choice in menu bar. After selection of language the user is required to add heading and description using their own voice (by pressing and holding add button). Furthermore the user can search the added note using their own voice by enduring List option that contain in option menu(with all five different languages) you store any note with the same name as you spoke, the description will be displayed as shown in the figures.

Fig 2.a. Append Note (English

Fig 2.b. Append Note (Hindi)

Fig 2.c. Append Note (Malayalam)

In this this application we can write our own note in 5 different languages Fig 2.a,b & c representing that adding note about speech recognition in English , Hindi and Malayalam languages.For implementing this mobile app, Android Studio 3.1.1 IDE with java platform is considered. In this all things are based on java. We can obtain the speech recognizing facility with using "Speech recognizer" inbuild class. And for storing the note SQLite data base is used. We can create database with using "SQLiteOpenHelper" inbuild class. When calling the RecognizerIntent, This application start to records the voice and also with using EXTRA_LANGUAGE specifying the language that want to record.

The additional propounded ingrained attribute of this module is text to voice reader (voice note) used within dictionary for English language. In dictionary if the user search any word, and then if they click say button this application will speak the meaning of the explored word and the data (word) can be stored in English which can be later searched and traversed with voice using the same caption. If user needs to know meaning of any word, they can use English to English Dictionary implemented (called) from oxford Dictionary API (With "com.example.ajins.Annotate.MyDictionaryRequest").
When pressing and holding the mic button the user can spoke any English word. Then taping search button the meaning of that word will be display, if press say button the meaning will be speak by this application which is shown in below figures.

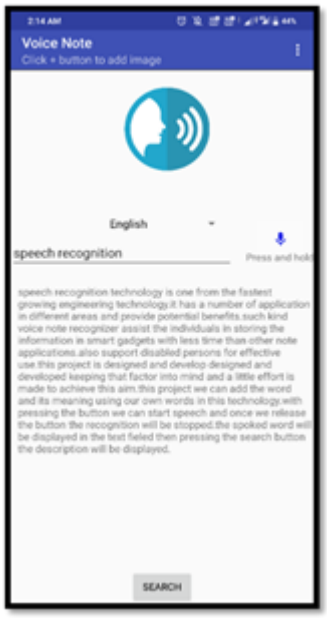

Fig 3.a. Voice note

Fig 3.b. English to English Dictionary

B. Ingression of Text from Image [6, 7, 8]

This module is manoeuvred if the user wishes to take notes directly from the real-world objects to their smart gadgets say user can add notes from a paper along with description by taking just an image using OCR tesseract tool. To do that in option menu there is a choice 'Add from Image' if they clicks new tab will pop to take picture/image via. camera or gallery. From there if user select gallery user can select image from the gadget's gallery, else use camera to take new photo of the text. Then user can fine tune the calibrations of the text they need to read, and press submit button will be added as description. Finally we have uploaded (published) this application in amazon app store so anyone can download this application for free.

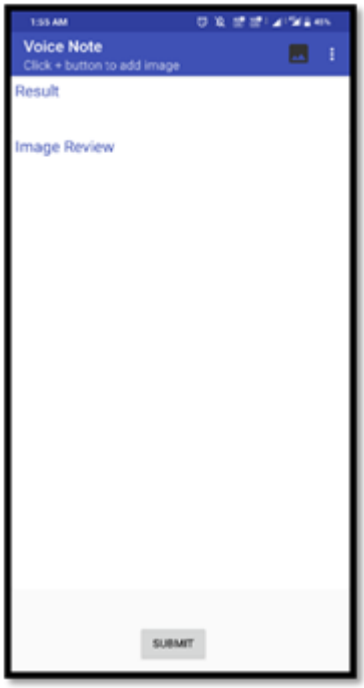

Fig 4.a. Voice Note Image review page

Fig 4.b. Retrieving text from image (paper document)

Fig 4.c. Retrieving text from image (paper document)

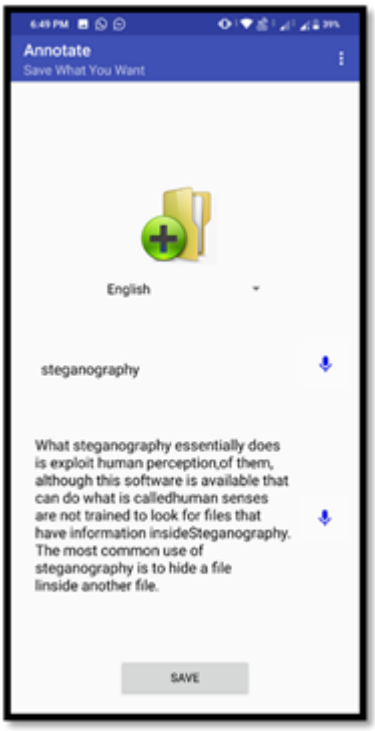

Fig4.d. Adding note from Image

Fig. 1. a figure caption. (figure caption)

IV. CONCLUSION

In this paper , an android voice note application named Annotate to capture voice , translate it with text (in five different languages) along with extracting text from image has been developed. Finally this app has been tested, results were demonstrated and published in amazon app store. Speech Recognition has become one of the research topic for the past years. Today, most the system (smart) are automated with inbuilt speech recognizer application through process identification and understanding. To achieve highly accurate (flawless) speech recognition system we require further optimization algorithm

REFERENCES

1. Jianliang Meng, Junwei Zhang,Haoquan Zhao: " Overview of the Speech Recognition Technology", at Fourth International Conference on Computational and Information Sciences, August 2012.
2. Nicholas Mulhern, Neil McCaffrey, Nicholas Beretta, Eugene Chabot PhD, Ying Sun PhD: "Designing Android Applications Using Voice Controlled Commands", at 39th Annual Northeast Bioengineering Conference, April 2013.
3. Amruta Pise, S.D.Ruikar: "Text Detection And Recognition in Natural Scene Images", at International Conference on Communication and Signal Processing, April 2014.
4. Kamrul Hasan Talukder, Tania Mallick:" Connected Component Based Approach for Text Extraction from Color Image", 17th International Conference on Computer and Information Technology (ICCIT),2014
5. Gunta Chomchalerm, Jitsupha Rattanakajornsak, Unchalee Samsrisook: "Braille Dict, Dictionary Application For The Blind On Android Smartphone", at Third ICT International Student Project Conference (ICT-ISPC2014), March 2014.
6. Mr. Pratik Madhukar Manwatkar, Mr. Shashank H. Yadav: "Text Recognition from Images", at IEEE Sponsored 2nd International Conference on Innovations in Information, March 2015.
7. Sathiapriya Ramiah, Tan Yu Liong, Manoj Jayabalan: " Detecting Text Based Image with Optical Character Recognition For English Translation And Speech Using Android" at IEEE Student Conference on Research and Development, Dec. 2015.
8. Dr. Jayashri Vajpai, Avnish Bora: "Industrial Applications of Automatic Speech Recognition Systems", Journal of Engineering Research and Applications, March 2016.
9. Nidhi Kalidas Sawant, Prof. Sangam Borkar: "Devanagari Printed Text To Speech Conversion Using OCR", at Proceedings of the Second International conference, August 2018.
10. D.S.S. De Zoysa , J.M. Sampath , E.M.P De Seram , D.M.I.D. Dissanayake , L.Wijerathna and S. Thelijjagoda: " Project Bhashitha - Mobile Based Optical Character Recognition And Text-To-Speech System", at 13th International Conference on Computer Science & Education (ICCSE 2018), August 2018.
11. https://www.techopedia.com/definition/6044/automatic-speech-recognition.

AUTHORS PROFILE

Ajins Joy, is a JRF (Junior Research Fellow) in the Department of School of Computer Science and Engineering at Vellore Institute of Technology (VIT), Chennai, since June 2019. He has a Masters in Computer Science in Central University of Tamil Nadu, 2019

Dr. R Saranya is an Assistant Professor currently working in the Department of Computer Science at Central University of Tamil Nadu, India since 2016. She has more than 7 years of teaching experience in Postgraduate and undergraduate courses. Her research areas are in Computer Networks, Cybersecurity, Software Engineering and Internet of Things. Her current research is in DDOS with C-Worm detection. She is also a member of academic bodies like CSI, IEEE, and IE. She has presented more than 15 research papers in International (IEEE) and National Conferences. She has also published more than 10 research articles in peer-reviewed journals.