Kannada Speech Recognition Enquiry System For Farmers

Mrs. Anjani
Dept. ECE MITE, Moodbidri

Pooja
Dept. ECE MITE, Moodbidri

Ranjitha V P
Dept. ECE MITE, Moodbidri

Manjunath Acharya
Dept. ECE MITE, Moodbidri

Ganesh Kamath
Dept. ECE MITE, Moodbidri

Abstract— **Under developing and developing countries agriculture provides large scale of employment in rural area. It is the backbone of economic system. For formers it is important to decide which crop to grow and to gain knowledge regarding the crops. In This paper describe the process of acquiring speech data for training an ASR system for Kannada language that will form the core of a voice interface to the webpage providing information about crops as a voice. Provision will be provided to display the information on a screen using a website.**

Keywords—google speech recognizer software ; web crawl;

I. INTRODUCTION

India is an agricultural country. Despite the advent of technology in recent years, agriculture has still remained the major source of income for majority of the population. In order to provide a stimulus to the agricultural growth, the Ministry of Agriculture, Government of India set up many websites which displays information regarding problems and solutions. But most of the farmers are far behind in technology to use website. So, we are conceptualizing a voice recognizing portal for kannada speaking farmers which can show the information about crops. The benefits of such system can be reaped if farmers are able to access this information easily. Presently, illiteracy, ignorance, lack of knowledge of English, computers and internet are acting as hurdles in accessing the information from the website. On the other hand, if this information can be made available just by speaking over a mic it would foster a large number of farmers. This would need implementation of a voice interface to the website using Automatic Speech Recognition (ASR). Speech data acquisition is the first step towards building a speech recognition system. The accuracy of recognition depends on the speech data used to train the system. The measures taken for collecting apt speech data from Kannada Speaking farmers, to develop a robust speech recognition system as a part of voice interface for agricultural information retrieval. The process of acquiring speech data for training an ASR system for Kannada language that will form the core of a voice interface to the webpage providing information about crops as a voice. Provision will be provided to display the information on a screen using a website.

The Technology Development for Indian Languages (TDIL) [1] programme of the Department of Information Technology (DIT) has initiated a nationwide project that effects such a voice interface in six Indian languages, which are Marathi, Hindi, Tamil, Telugu, Bangla and Assamese. This project is being implemented by a consortium of seven institutions, which are Indian Institute of Technology (IIT) Madras, IIT Bombay, IIT Kanpur, IIT Guwahati, International Institute of Information Technology (IIIT)Hyderabad, Tata Institute of Fundamental Research (TIFR) Mumbai and Centre for Development and Advanced Computing (C-DAC) Kolkata [3]. IIT Bombay and TIFR Mumbai are jointly involved in developing the Marathi ASR system.

II. LITERATURE REVIEW

In [1] has proposed that in daily life Speech and spoken words have always played a big role in the individual and collective lives of the people. The Speech that represents the spoken form of a language. Speech synthesis is the process of converting message written in text to equivalent message in spoken form .A Text-To-Speech (TTS) synthesizer as a computer-based system that should be able to read text. In this paper, I am explaining single text-to-speech (TTS) system for Indian languages Viz., Hindi to generate speech .This generally involves two steps, text processing and speech generation. A graphical user interface has been designed for converting Hindi text to speech in Python Swings. In India there are different languages are spoken, but each language is the mother tongue of tens of millions of people. The languages and scripts are very different from each other. The grammar and the alphabet words are similar to a large extent. It present text-to-speech (TTS) system based on the Concatenate synthesis approach. The text to speech conversion may seem effective and efficient to its users if it produces natural speech and by making several modifications to it. This system is useful for deaf and dumb people to Interact with the other peoples from society. Text to speech synthesis is a critical research and application area in the field of multimedia interfaces. The system read the input data in a natural form. The user types the input string and the system reads it from the database or data store where the words,

Special Issue - 2019

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
RTESIT - 2019 Conference Proceedings

phones, diphones, triphone are stored. In this paper, we presented the development of existing TTS system by adding spellchecker module to it for Hindi language.

In[2]. has proposed that there are about 45 million blind people and 135 million visually impaired people worldwide. Disability of visual text reading has a huge impact on the quality of life for visually disabled people. Although there have been several devices designed for helping visually disabled to see objects using an alternating sense such as sound and touch, the development of text reading device is still at an early stage. Existing systems for text recognition are typically limited either by explicitly relying on specific shapes or colour masks or by requiring user assistance or may be of high cost. Therefore we need a low cost system that will be able to automatically locate and read the text aloud to visually impaired persons. The main idea of this project is to recognize the text character and convert it into speech signal. The text contained in the page is first pre-processed. The pre-processing module prepares the text for recognition. Then the text is segmented to separate the character from each other. Segmentation is followed by extraction of letters and resizing them and stores them in the text file. These processes are done with the help of MATLAB. This text is then converted into speech. This paper is an effort to suggest an approach for image to speech conversion using optical character recognition and text to speech technology. The application developed is user friendly, cost effective and applicable in the real time. By this approach we can read text from a document, Web page or e-Book and can generate synthesized speech through a computer's speakers or phone's speaker. The developed software has set all policies of the singles corresponding to each and every alphabet, its pronunciation methodology, the way it is used in grammar and dictionary. This can save time by allowing the user to listen background materials while performing other tasks. System can also be used to make information browsing for people who do not have the ability to read or write.

In [3]. Speech is one of the oldest and most natural means of information exchange between human. Over the years, Attempts have been made to develop vocally interactive computers to realize voice/speech synthesis. Obviously, such an interface would yield great benefits. In this case a computer can synthesize text and give out a speech. Text-To-Speech Synthesis is a Technology that provides a means of converting written text from a descriptive form to a spoken language that is easily understandable by the end user (Basically in English Language). It runs on PYTHON platform, and the methodology used was Object Oriented Analysis and Development Methodology; while Expert System was incorporated for the internal operations of the program. This design will be geared towards providing a one-way communication interface whereby the computer communicates with the user by reading out textual document for the purpose of quick assimilation and reading development.

In [4]. Described that it is able to control devices by voice has always intrigued mankind. Today after intense research, Speech Recognition System, have made a niche for themselves and can be seen in many walks of life. The accuracy of Speech Recognition Systems remains one of the most important research challenges e.g. noise, speaker variability, language variability, vocabulary size and domain. The design of speech recognition system requires careful attentions to the challenges such as various types of Speech Classes and Speech Representation, Speech Pre-processing stages, Feature Extraction techniques, Database and Performance evaluation. This paper presents the advances made as well as highlights the pressing problems for a speech recognition system. The paper also classifies the system into Front End and Back End for better understanding and representation of speech recognition system in each part.

In [5] described that Text-to-speech synthesizer is an application that converts text into spoken word, by analyzing and processing the text using Natural Language Processing (NLP) and then using Digital Signal Processing (DSP) technology to convert this processed text into synthesized speech representation of the text. Here, we developed a useful text-to-speech synthesizer in the form of a simple application that converts inputted text into synthesized speech and reads out to the user which can then be saved as an mp3. file. The development of a text to speech synthesizer will be of great help to people with visual impairment and make making through large volume of text easier.

III. METHODOLOGY
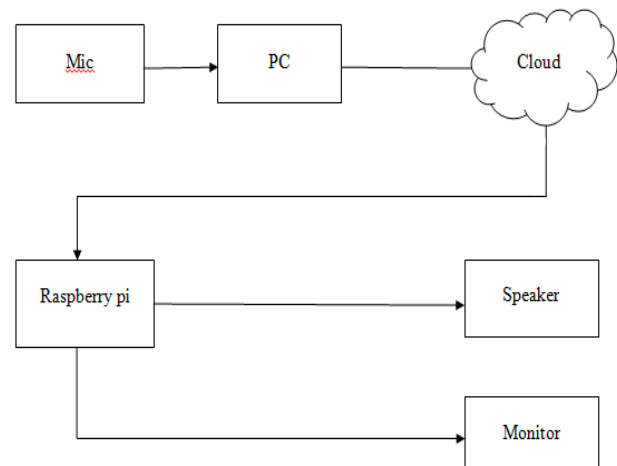
A. MIC

A mic is a transducer that converts sound into an electrical signal. Mic are used in many applications such as telephones, hearing aids, sound recording. For speech acquisition we make use of mic. Mic is connected to the system in the system which takes voice from mic then stores the voice in the form of wave file then uses google recognizer convert voice to text and that text will matching with crop name which we already given the different number for different crops, if any matches found that matching number will send to the cloud.

B. Cloud

Cloud storage is a model of computer data storsge in which the digital data is stored in logical pools.

C. Raspberry Pi

In Raspberry-pi we are extracting number which is uploaded by the system and that number will match with the information of crops if match is found then the information will play in the speaker.

D. Speaker

Speaker is a device which converts an electrical audio signal into a corresponding sound. Here we can hear the crop information.

E. Monitor

In monitor the images of the crop will displayed on the monitor.
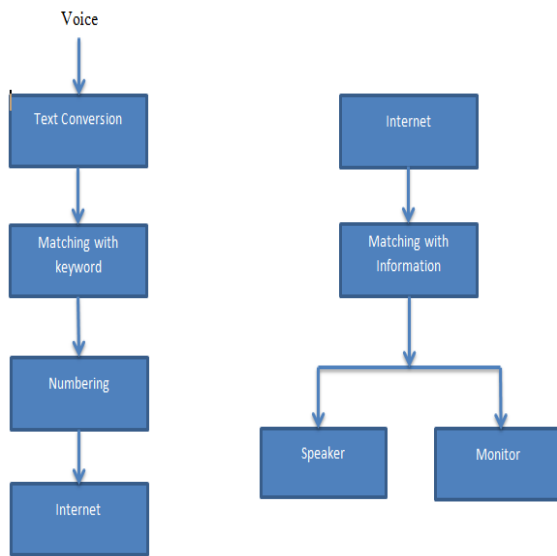

Fig. 2 Flow diagram of Transmitter and Receiver of speech recognition system

When we run the program it start to recording for 5 seconds after recording the crop name next step is convert that crop name in to text using Google speech recognizer it will convert the voice to text and then matching is used to match that converted text with the crop name if the crop name found then assign the number for that crop and send it to the cloud.

In receiver part the raspberry pi extract the information from the cloud and that number start to match and if matching is found then that matching information which is in the form of voice will be come out from the speaker and the image of the crop will be displayed on the monitor.

IV. SPEECH RECOGNITION

Speech recognition is a technology that able a computer to capture the words spoken by a human with a help of microphone. These words are later on recognized by speech recognizer, and in the end, system outputs the recognized words. The process of speech recognition consists of different steps that will be discussed in the following sections one by one.

An ideal situation in the process of speech recognition is that, a speech recognition engine recognizes all words uttered by a human but, practically the performance of a speech recognition engine depends on number of factors. Vocabularies, multiple users and noisy environment are the major factors that are counted in as the depending factors for a speech recognition engine.

A. Types of Speech Recognition

Speech recognition systems can be divided into the number of classes based on their ability to recognize that words and list of words they have. A few classes of speech recognition are classified as under:

1) Isolated Speech: Isolated words usually involve a pause between two utterances; it doesn't mean that it only accepts a single word but instead it requires one utterance at a time.

2) Connected Speech: Connected words or connected speech is similar to isolated speech but allow separate utterances with minimal pause between them.

3) Continuous speech: Continuous speech allows the user to speak almost naturally, it is also called the computer dictation.

4) Spontaneous Speech: At a basic level, it can be thought of as speech that is natural sounding and not rehearsed. An ASR system with spontaneous speech ability should be able to handle a variety of natural speech features such as words being run together, "ums" and "ahs", and even slight stutters.

PyCharm is an integrated development environment (IDE) used in computer programming, specifically for the Python language. It is developed by the Czech company JetBrains. [6] It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCSes), and supports web development with Django.

PyCharm is a Python IDE with complete set of tools for Python development. In addition, the IDE provides capabilities for professional Web development using the Django framework. Code faster and with more easily in a smart and configurable editor with code completion, snippets, code folding and split windows support.

PyCharm Features

Project Code Navigation - Instantly navigate from one file to another, from method to its declaration or usages, and through class's hierarchy. Learn keyboard shortcuts to be even more productive.

Code Analysis - Take advantage of on-the-fly code syntax, error highlighting, intelligent inspections and one-click quick-fix suggestions to make code better.

Python Refactoring - Make project-wide code changes painlessly with rename, extract method/superclass, introduce field/variable/constant, move and pull up/push down refactoring's.

Web Development with Django - Even more rapid Web development with Django framework backed up with excellent HTML, CSS and JavaScript editors. Also with CoffeeScript, Mako and Jinja2 support.

Google App Engine Support - Develop applications for Google App Engine and delegate routine deployment tasks to the IDE. Choose between Python 2.5 or 2.7 runtime.

Special Issue - 2019

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
RTESIT - 2019 Conference Proceedings

Version Control Integration - Check-in, check-out, view diffs, merge — all in the unified VCS user interface for Mercurial, Subversion, Git, Perforce and other SCMs.

Graphical Debugger - Fine-tune Python or Django applications and unit tests using a full-featured debugger with breakpoints, stepping, frames view, watches and evaluate expressions.

Integrated Unit Testing - Run a test file, a single test class, a method, or all tests in a folder. Observe results in graphical test runner with execution statistics.

Customizable & Extensible - Bundled Textmate, NetBeans, and Eclipse & Emacs keyboard schemes, and Vi/Vim emulation plugin.

B. Speech Recgonition Process

*1) Voice Input :*With the help of microphone audio is input to the system, the pc sound card produces the equivalent digital representation of received audio.

*2) Digitization:*The process of converting the analog signal into a digital form is known as digitization, it involves the both sampling and quantization processes. Sampling is converting a continuous signal into discrete signal, while the process of approximating a continuous range of values is known as quantization.

3) Acoustic Model: An acoustic model is created by taking audio recordings of speech, and their text transcriptions, and using software to create statistical representations of the sounds that make up each word. It is used by a speech recognition engine to recognize speech. The software acoustic model breaks the words into the phonemes.

4) Language Model: Language modelling is used in many natural language processing applications such as speech recognition tries to capture the properties of a language and to predict the next word in the speech sequence. The software language model compares the phonemes to words in its built-in dictionary.
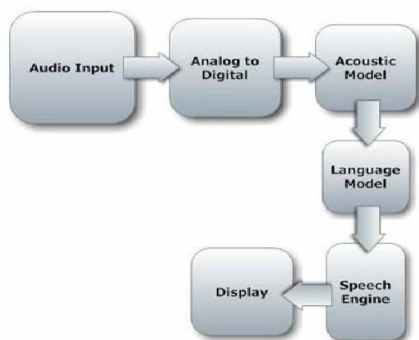

Fig. 3 Speech Recognition Process

5) Speech engine: The job of speech recognition engine is to convert the input audio into text; to accomplish this it uses all sorts of data, software algorithms and statistics. Its first operation is digitization as discussed earlier, that is to convert it into a suitable format for further processing. Once audio signal is in proper format it then searches the best match for it. It does this by considering the words it knows, once the signal is recognized it returns its corresponding text string.

C. The Future of Speech Recognition.

Accuracy will become better and better.
• Dictation speech recognition will gradually become accepted.
• Greater use will be made of "intelligent systems" which will attempt to guess what the speaker intended to say, rather than what was actually said, as people often misspeak and make unintentional mistakes.
• Microphone and sound systems will be designed to adapt more quickly to changing background noise levels, different environments, with better recognition of extraneous material to be discarded.

D. Fundamentals to speech recognition

Speech recognition is basically the science of talking with the computer and having it correctly recognized. To elaborate it we have to understand the following terms.

Utterances: When user says some things, then this is an utterance in other words speaking a word or a combination of words that means something to the computer is called an utterance. Utterances are then sent to speech engine to be processed.

Pronunciation: A speech recognition engine uses a process word is its pronunciation that represents what the speech engine thinks a word should sounds like. Words can have the multiple pronunciations associated with them.

Grammar: Grammar uses particular set of rules in order to define the words and phrases that are going to be recognized by speech engine, more concisely grammar define the domain with which the speech engine works. Grammar can be simple as list of words or flexible enough to support the various degrees of variations.

Accuracy: The performance of the speech recognition system is measurable; the ability of recognizer can be measured by calculating its accuracy. It is useful to identify an utterance.

Vocabularies: Vocabularies are the list of words that can be recognized by the speech recognition engine. Generally, the smaller vocabularies are easier to identify by a speech recognition engine, while a large listing of words is difficult task to be identified by engine.

Training: Training can be used by the users who have difficulty of speaking or pronouncing certain words, speech recognition systems with training should be able to adapt.

Special Issue - 2019

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
RTESIT - 2019 Conference Proceedings

E. Software Description

The PyCharm is an integrated development environment (IDE) used in computer programming, specifically for the Python language.
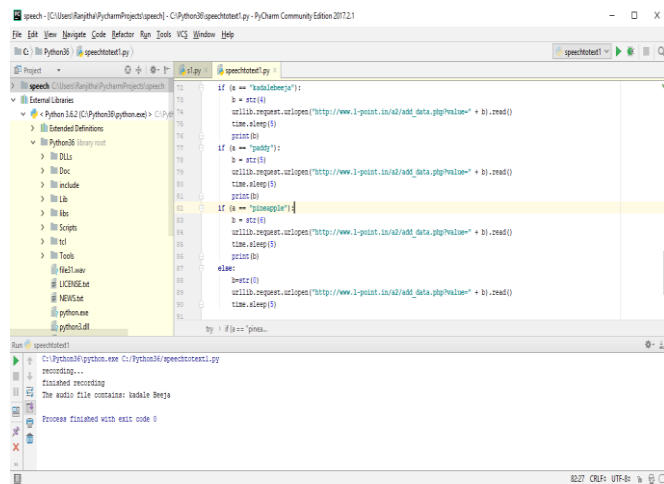

Fig. 4 Displays the content of crop Kadale beeja

Here we using the Pycharm, when the farmer speakes the crop name it starts to recording and after completion of recording it takes some time (milisecs) and start searching for the information of the crop which is farmer spoke. If the crop information is available in the program it displays the audio file content.
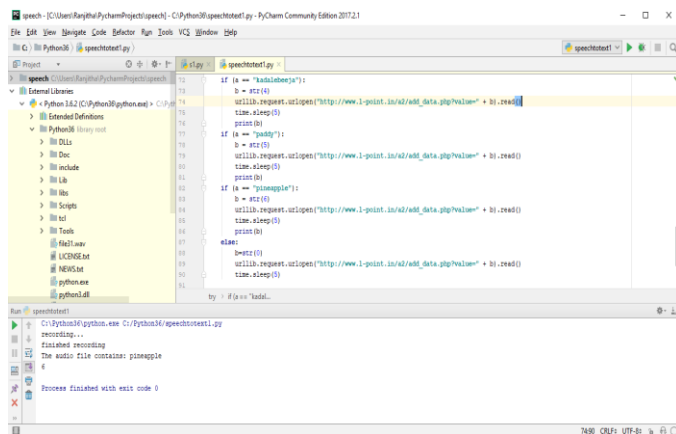

Fig. 5 Displays the content of pineapple

V. CONCLUSION

Kannada speech recognition enquire system for farmers will be very helpful for the uneducated farmers who wants to have the information of their crop. These system can be deployed in all the places where it will be helpful for the framers. The process of acquiring speech data for training an ASR system for Kannada language that will form the core of a voice interface to the cloud providing information about crops as a voice. Provision will be provided to display the image of the crop. This system can be developed to different local language also. In later days we can develop this project in Android phones so not only farmers everyone can get crop information.

REFERENCES

[1] Kaveri Kamble, Ramesh kagalkar "*Translation of Text to Speech Conversion for Hindi Language*" India. International Journal of Science and Research (IJSR). 2016

[2] JishaGopinath, PoojaChandran, Saranya S, Aravind S *"Text to Speech Conversion System using OCR"* International Journal of Emerging Technology and Advanced Engineering. 2015

[3] Swathi Ahlawat, rajiv Dahiya, "*A Novel Approach of Text to speech Conversion Under Android Environment", IJCSMS International Computer, 2015.*

[4] Nitin Washani, Sandeep Sharma "*Speech Recognition System computer applications"* proceedings of IEEE ICASSP. 2015.R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[5] Itunuoluwalsewon, JeliliOyelade and OlufunkeOladipupo Department of Computer and Information Sciences*, "Android Application to get Word Meaning through Voice",* International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) 2014.

[6] D.Sasirekha, E.Chandra, "*Text to Speech: A Simple Tutorial",* International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012.

[7] C. Kim and R. M. Stern, "Feature extraction for robust speech recognition based on maximizing the sharpness of the power distribution and on power flooring", in Proc. ICASSP, pp. 4574– 4577, 2010.

[8] V. Tyagi, "Fepstrum features: Design and application to conversational speech recognition", IBM Research Report, 11009, 2011.

[9] V. Mitra, M. McLaren, H. Franco, M. Graciarena, N. Scheffer, "Modulation Features for Noise Robust Speaker Identification," Proc. of Interspeech, pp. 3703-3707, Lyon, 2013.