Survey Paper on Jarvis Digital Life Assistant

***Sneha Sabale, *Madhavi Vora, **Prof. A. S. Shinde**

**B. Tech. Students, **Assistant Professor*
Department of Computer Science and Engineering
Sahakar Maharshi Shankarrao Mohite Patil Institute of Technology and Research,
Akluj, Solapur, Maharashtra, (India)

Received: 10 January, 2022 Accepted: 22 February, 2022 Online: 05 March, 2022

ABSTRACT

Jarvisis "Just Rather Virtual Assistant System" which make easiest our life. In this project Jarvis is digital life assistant which uses only human communication means such twitter, instant message and voice to create two way connection between human and his apartment, controlling light and appliances assist in booking , notify him of breaking news, facebook notification and many more. In our project we mainly use voice as a communication means so the Jarvis is basically speech recognition system. This concept of speech technology encompasses two different technologies first is Synthesizer and second is Recognizer. The speech synthesizer takes as input and produce and audio stream as output. Also speech recognizer on other hand does opposite. Also it takes an audio stream as an input and thus turns it into text transcription. The voice is a signal of infinite command. The direct analysis and synthesizing the complex voice due to too much information contained in the signal. Therefore the digital signal process such as Feature Extraction and Feature matching are introduced to represent the voice signal. In this project we directly use speech engine which use Feature Extraction technique as Mel scaled frequency cepstral. The Mel scaled frequency ceptrals coefficients (MFCCs) derived from most widely used front-ends in state-of-the-art speech recognition systems. Our aim to create more and more functionalities which can help human to assist their daily life and reduce their effort. In our test we check all these functionalities are working properly.

Keywords: *Feature Extraction, Speech recognizer, Feature Matching.*

INTRODUCTION

Speech is an effective and natural way for people to interact with applications, complementing or even replacing the mouse, keyboard, controllers and gestures. A hand-free yet accurate way to communication with application, speech lets be productive and stay informed in a variety of situations where other interfaces will not. The recognition of speech is very interesting topic that is very useful in many applications and environments in our daily life survival. Speech recognizer is a machine which understands a human and their spoken word in some way and can act thereafter. A different aspect of speech recognition is facilitating for people with functional disability or other kind of handicap. To make their daily chores easier, voice control could be helpful. Because of their voice they could operate the light switch turn ON or OFF or operate some other domestic appliances very easily.

This leads to the discussion about intelligent homes where this operation can made available for common man as well as for handicapped. According to this information presented so far one question comes naturally that is "How speech recognition is done?" To get knowledge of hoe speech recognition problem can be approached today, are view of some research high lights will be presented. The earlier attempt to devise systems for automatic speech recognition by machine were made in the1950's, When various researchers tried to exploit the fundamental ideas of acoustic-phonetics. Davis, Biddulph, Balashek built a system for isolated digit recognition for a single speaker in 1952 at Bell Laboratories. The system relied heavily on measuring spectral resonances during the vowel region of each digit.

LITERATURE REVIEW

In paper titled “Speech based Human Emotion Recognition Using MFCC”, IEEE WiSPNET year of 2017 conference proposed by author name as M.S.Likitha, Sri Raksha R. Gupta.

A database consists of voices of 60 people with different emotions. Speech signal of speakers read using the function away reading MATLAB tool. MFCC method is sued for detecting emotion from voice signals easily.

The proposed method is based on feature extraction by using MFCC algorithm and decision making using standard deviations. The speech signal made to undergo Farming, after which it is passed through Hamming window for windowing process. Fast Fourier Trans form was performed on the input signal. The standard deviation for the mean value was found, and this value was passed through as if else statement, where they obtained standard deviation of that particular motion is compared with the optimized value of standard deviation for different emotion and corresponding emotion were displayed. It can predict some basic emotion such as happy, sad and angry from MFCC waves. In[2] named “Emotion Recognition from Speech using Convolutional Neural Network with Recurrent Neural Network Architecture”, IEEE WiSPNET year of 2017 conference proposed by authors named as Saikat Basu, Jaybrata Chakraborty. For the last two decades, several intelligent systems are proposed by researches. These different systems also differ by nature of features used for classification of speech signals. There are widely used spectral feature are Mel-frequency cepstrum coefficients (MFCC) and linear predictive cepstral. Speech understanding is a major component of human machine interaction and its quality affect the user experience.

Coefficient (LPCC). Only pitch and MFCC features are used for recognition of emotion. In [3] named “Emotion Recognition from Speech using Emotional Statistical Parametric Speech Synthesis Using LSTM-RNNs ”IEEE WiSPNET of 2017conference year of 12-15 December 2017,authors name as Shumin An, Zhenhua Lingand Lirong Da. proposed two modeling approaches, emotion dependent modeling and unified modeling with emotion codes, are implemented and compared by experiments. LSTM-RNN-based acoustic model are built separately for each emotion type. By using two or four –dimensional emotion space gives better result In[4]named “Training Spoken Language Understanding System With non-parallel speech and Text”, IEEE WiSPNET of 2020 conference proposed by author named Leda Sari, Samual Thomas. End-to-end spoken language understanding (SLU) systems are typically trained on large amount of data.

PROPOSED SYSTEM

Block Diagram of Proposed System With it deep learning. Proposed model gives better accuracy for data sets. For real time imaginary large dataset is needed. The acoustic characteristics of the speech signal are Feature. The concept of feature extraction is a small amount of a data from the speech signal is extracted. In speech analysis, it is common to use two types of features: acoustic, which have physical sense and representation characteristics that correspond to values calculated over the to analyze the signal without disturbing its acoustic properties.

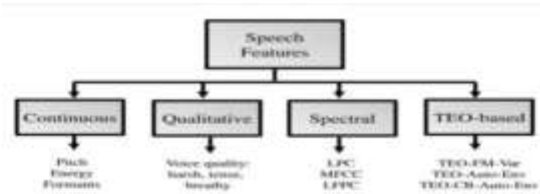
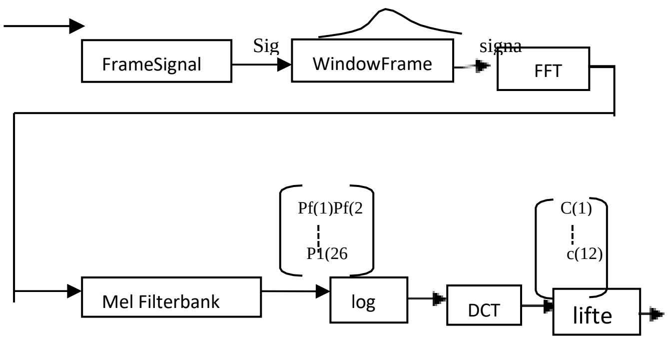


Figure 1: Types of Speech Features

METHODOLOGY

For each frame



MFCC overview

In this project we directly use speech engine which use feature extraction technique as Mel Scaled frequency cepstral coefficient (MFCCs) derived from Fouriertrans form representation of speech signal, and those that in general does not correspond to any physical sense and filter bank analysis are perhaps the most widely used front-ends in state-of-the-art speech recognition systems. After testing, we confirming that all these functionalities are working properly.

REFERENCES

[1] Kanchan Patil, AvinashKharat2, PratikChaudhary3, ShrikantBider4, Rushikesh Gavane 5 Department of Information and Technology, SRESs Sanjivani College of Engineering Kopargoan, India

[2] M.S. Likitha, Sri Raksha R. Gupta "Speech based Human Emotion Recognition Using MFCC", IEEE WiSPNET year of 2017

[3] Leda Sary, Samuel Thomas, Mark Hasegawa-johnson, Department of computer engineering, Universities of Illinoisat Urbana-chamaign.

[4] Shrutika Khobragade Department of computer Engineering.