I. INTRODUCTION

Voice detection and speech recognition is become a topic of interest as world is moving towards machine learning or deep learning. In 1952 three Bell Labs researchers, [Stephen. Balashek](https://obits.nj.com/obituaries/starledger/obituary.aspx?page=lifestory&pid=158702138) , R. Biddulph, and K. H. Davis built a system called '[Audrey](https://cdn57.androidauthority.net/wp-content/uploads/2012/04/IBM-Shoebox-front.jpg)' an automatic digit recognizer for single-speaker digit recognition. Their system worked by locating the [formants](https://en.wikipedia.org/wiki/Formants) in the power spectrum of each utterance. The 1950s era technology was limited to single-speaker systems with vocabularies of around ten words. [1] That was the initial but till then problem is still the same we are using machine learning algorithm to predict whom voice is this and what is the text accordingly apart that what we are moving ourselves towards real time detection and conversion of sound to text using offline library trying to making it efficient.

This paper is divided in to three sections voice detection, Sound to text conversion and real time system.

Voice Detection contains in itself many steps

1. Noise Removal
2. Extracting features
3. Training of feature with its label accordingly
4. Prediction method after training of data

II. THE NOISE REMOVAL ALGORITHM

III. EXTRACTING FEATURES

For features extraction from a sound (.wav) file we are using librosa which is a python package for music and audio analysis.

First of all loading that wav file using load function it gives sample rate and an array while moving further we extract mfcc (Mel-scale frequency cepstral coefficient ) they are defined well in Methodology part, after extracting these mfcc features a label in a sense of identity is placed there e.g.: these features are for female or male .

Covering next step of voice detection we will repeat the above process for each data sample while appending feature in features list and label in labels list accordingly.

IV. PREDICTING THE VOICE

We are using sklearn predicting method using machine learning algorithm but it requires features of the sound which is to be predicted so we will use that load method again to get mfcc features and array and then passing that to predicting method which will return a label (identity)