**RESEARCH PAPER**

**MEETING RECORDER**

College of Computing and Information Sciences, PAF Karachi Institute of Economics and Technology

**ABSTRACT:**

As world is moving towards the new era so called as the era where intelligent systems are used to solve complex problems? In this paper, we propose an intelligent framework to record meeting to avoid hassle of writing and to search each person’s data. The framework will also help in enhancing the productivity of meeting by reducing human generated errors and distractions. A problem can arise in during a meeting when a recorder (human) is there to write all of the conversation and the problem statement we have identified is that the recorder can miss any of the information during the meeting because of the noise in the audio and multiple persons can speak at the time that will be difficult to identify so we are overcoming the problem by making the computer to do that task. The framework uses voice recognition, noise reduction and real time audio processing to generates meeting minutes

**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. [3] That was the initial but till then problem is still the same. The problems we have analyzed is that the only one speaker can speak at the time so after the some span of time that audio will be converted into text and by recognizing their voice. we are using machine learning algorithm to predict whose 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 and also trying to making it efficient to convert the text using offline library. Section-I discuss current voice recognition and noise removal techniques. Section-II discuss the features extracted from the audio input. Section-III discuss the training of those features by labeling them. Section-IV discuss the model training to label out the features of voice. Section-V discuss the detection of speaker at the real time. Section-VI is about PCA reduction of data after addition of a new member and Section-VII is Oversampling of reduced data

The main modules that will be discussed in this paper will be the following:

* Noise Removal using discrete wavelet transform.
* Extracting voice features.
* Training of the voice feature with the label (Model Training).
* Prediction the person voice after training of data.
* Real time voice detection.
* Over Sampling of audio files
* Reduction using PCA

**LITERATURE REVIEW:**

In previous years research on sound and its feature was in focus about feature extracting techniques and noise removal techniques

**METHODOLOGY:**

I. SPEECH TO TEXT:

Over the years a lot of work has done in the field of speech recognition. But all of the work related to speech recognition doesn’t convert the speech into text in real time. Most of the techniques used for this purpose takes time to convert the speech into the textual form. One disclosed embodiment a speech conversion program receives audio inputs from a microphone array of a head-mounted display device [2]. A beam forming technique is applied to at least a portion of the audio inputs to identify target audio inputs. The target audio inputs are converted into text. A beam forming technique is applied to audio inputs from a microphone array in the head-mounted display device to identify target audio inputs. The thresholds of each alphabet is already defined so when person speaks it detects the voice and compare each alphabet to its threshold if the value matches then it is declare as the alphabet.

II. NOISE REMOVAL:

The principle under which the wavelet de-noising approach operates relies on the fact that for many real life signals, a limited number of wavelet coefficients in the lower bands are sufficient to reconstruct a good estimate of the original signal. Usually these coefficients are relatively large compared to other coefficients or to any other signal (especially noise) that has its energy spread over a large number of coefficients [1]. By calculating the threshold value we eliminate or reduce the noise from audio by keeping the information from the original audio.

III. REAL TIME VOICE DETECTION:

Real time voice detection is not easy and attempts are there in different ways first an attempt is made for an [3] iterative system to follow a human voice in real time, using low cost electronics and processors. Sound source localization does not require direct line of sight, and can be implemented relatively more easily than vision-based localization methods and it is a contribution in real time detection field. [4] Nowadays, we’re moving to a new world of technology and input devices designed to fit specific times and places, such like voice activation for computer games and app games. Real time speech detection and recognition this is the combine key idea.

IV. MODEL TRAINING:

After features extraction the model is trained on the extracted features dataset so that the same model can be used to do the further tasks.

* The model is trained using ‘sklearn Adaboost’classifier and ‘joblib’*.* For model training the extracted features of each audio file stored in the .npy file are used and the label of each audio feature is labeled according to the data stream. Then a file of joblib extension is generated which is actually the trained model.
* Another model is trained known as SVM for better accuracy and testing purpose we extract data from our audio files stored in .npy files as we did earlier in joblib model to find out better accuracy.
* Third one is NN (Neural Network) same work is done for it in search of better accuracy.

V. DETECTION OF SPEAKER:

Using the trained model the speaker can be recognized based on the voice features. On audio file features are extracted and then compared with the features saved in the trained model which we did earlier in model training part now it is the time to predict whose voice is it from the trained model to get the label (name) saved against the audio we are giving as input.

VI.PCA DIMENSIONALITY REDUCTION:

The key objective of PCA Dimensionality Reduction is that suppose we have a dataset consists of many attributes that are name, age, height, address, contact number and etc., so by using this algorithm we can fetch out the useful information in this dataset that could be name, age, contact number. In short PCA only kept the meaningful information form the dataset and discards the other information from the dataset. In this algorithm the lossless reduction is performed so that minimal useful data can be lost. All of the useless information or data stream is then reduced or discard form the audio stream and the final useful data stream is then received. The steps of implementing the PCA Dimensionality Reduction algorithm are:

1. Read the dataset.
2. Extract the last column from each dataset properties.
3. Normalize the dataset.
4. Calculate the covariance.
5. Find the eigen value and eigen vectors.
6. Sort the eigen value and eigen vectors.
7. Select top eigen values and its corresponding eigen vectors.
8. Formed the new dataset in the reduced dimension.
9. Plot the new dataset.

After implementing PCA dimensionality reduction the length of the data stream for 5 second audio sample reduced to half in numbers 220,159.

VII. OVER SAMPLING OF DATA:

On this reduced audio samples oversampling is done to oversample the audio sets for increasing the number of dataset to describe that how this specific technique works is, let’s have some training data which has number of samples *n* and has the features *f* in the feature space of the dataset. For simplicity mark that the dataset is continuous.

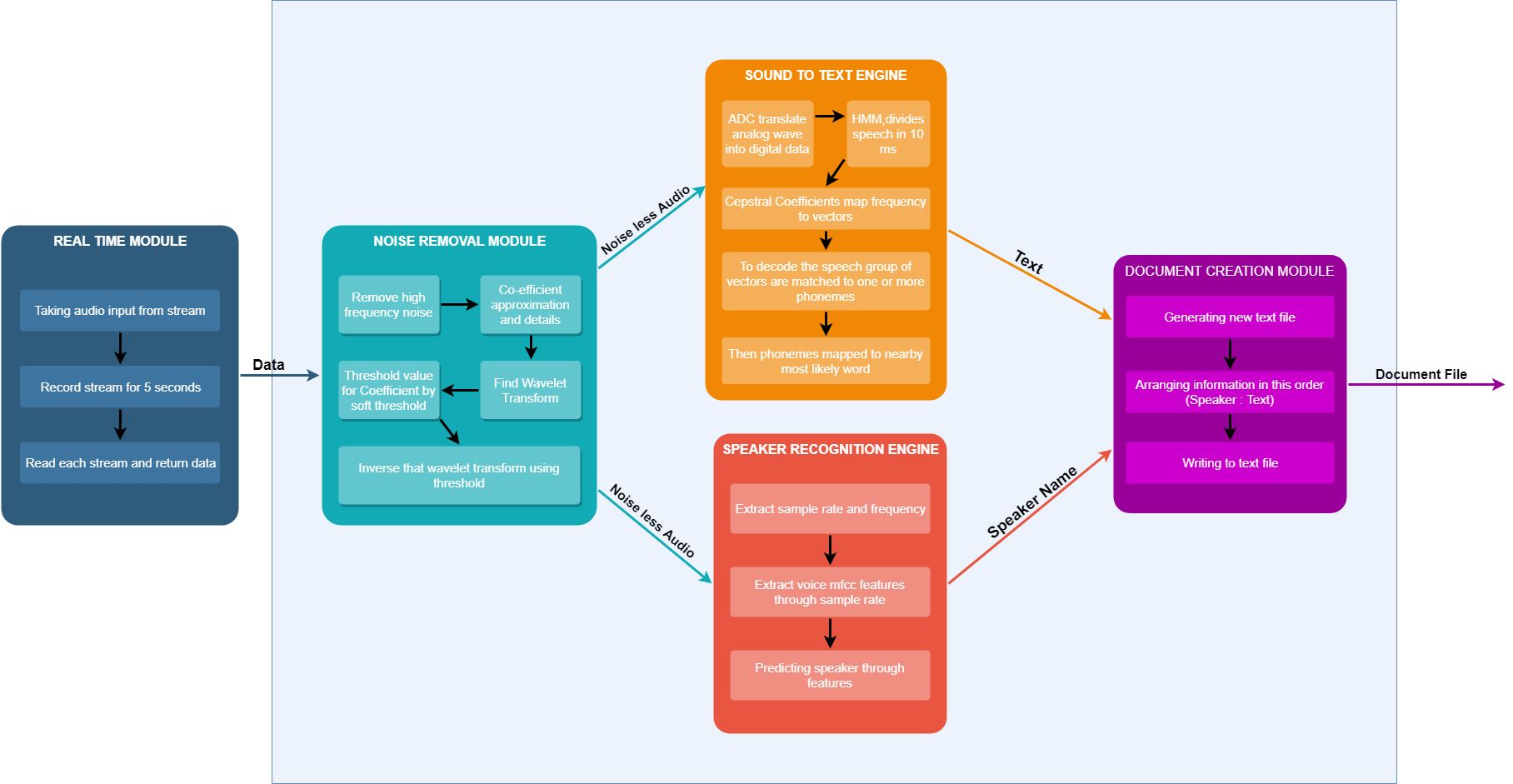
This algorithm takes the data that needs to be over sampled and the number of sample need to be created and the value of nearest neighbor. The value of nearest neighbors use to maintain the accuracy of the oversampled datasets.

In our case the Smote () function takes the parameters mentioned above but the values of the parameters are:

1. Dataset on which oversampling will be done.
2. Number of samples need to be created that is 50000.
3. Nearest Neighbors value that is k=1. At k=1 only the closest neighbor of the same class is advised to use.

After oversampling on the datasets the number of data will be increased that will create enough number of samples to train the model.

**FRAMEWORK DIAGRAM:**

****

**RESULTS:**

**REFERENCES:**

[1] Audio Signals Noise Removal Real Time System Sorin Zoican, Ph.D. POLITEHNICA University of Bucharest

[2] <https://patentimages.storage.googleapis.com/5f/d4/6f/3c9183dc9778ae/US9280972.pdf>

[3] [https://www.eurasip.org/Proceedings/Eusipco/Eusipco2017/wpapers/ML2.pdf](https://www.eurasip.org/Proceedings/Eusipco/Eusipco2017/wpapers/ML2.pdf?fbclid=IwAR1azUFuFbZbm3ba6KyFnYCbRPrXtHxjfzElq_jG0h25NvBZlM8WwX3CtYk)  
[4] [https://link.springer.com/chapter/10.1007/978-3-319-60639-2\_24](https://link.springer.com/chapter/10.1007/978-3-319-60639-2_24?fbclid=IwAR20COHyqiIVuNYBs9YxoG1jXkhOkY7kHze4FNYM2kd4CL2Nqz9jpH5bpW0)