RESEARCH ARTICLE OPEN ACCESS

Speaker Recognition Using Machine Learning Based Method

Vaibhav Bhardwaj[1], Manish Sharma[2]
[1]*Department of Computer Engineering & Information Technology, Suresh Gyan Vihar University, Jaipur,*
[2]*Department of Computer Engineering & Information Technology, Suresh Gyan Vihar University, Jaipur,*
Corresponding Author: Vaibhav Bhardwaj

ABSTRACT: Speaker recognition is related to speaker identification and speaker verification problems. Speaker identification is a task to identify the provided speech sample (utterance) as belonging to one speaker from the set of known speakers (1:N match). Speaker verification is a task of comparison of the utterance to one given speaker identity and either accepting or rejecting the identity (1:1 match).Here we are using an artificial intelligent (AI) based technique to complete this task. Here we are trying to differentiate the sound of APJ Abdul Kalam sir and the Donald Trumph president of America. Herewe are using K-nearest neighbor (KNN) algorithm which is a powerful supervised learning tool. We have extract the four parameters of sound and these are 'power spectral density', ‘spectral centroid’, ‘spectral flatness’ and ‘roll-off frequency'.
Keywords: Power spectral density, spectral centroid, spectral flatness and roll-off frequency.

Date Of Submission: 11-09-2019 Date Of Acceptance: 29-09-2019

I. INTRODUCTION

Identifying a speaker is one of the most complicated tasks in speech recognition problems. It generally depends on the production of speech of the speaker with physiological and behavioral characteristics. These characteristics rely on the speech generation (voice source) and the envelope behavior (vocal and nasal tract). Speaker recognition can be arranged to speaker identification and speaker verification [1]. Speaker identification consists in comparing a vocal message with a set of registered utterances corresponding to different speakers and determines the one who spoke. It entails a multi-choice classification problem. Speaker verification consists in accepting or rejecting the speaker who claims to be. It entails a yes-no hypothesis testing problem. All speaker recognition systems contain two main phases: feature extraction and recognition. During the first one, a training vector is generated from the speech signal of the word (password) spoken by the user. These training vectors are stored in a database for subsequent use in the recognition phase. During the recognition phase[2-5], the system tries to identify the unknown speaker by comparing extracted features from password with the ones from a set of known speakers. Speaker recognition is related to speaker identification and speaker verification problems. Speaker identification is a task to identify the provided speech sample (utterance) as belonging to one speaker from the set of known speakers (1:N match). Speaker verification is a task of comparison of the utterance to one given speaker identity and either accepting or rejecting the identity (1:1 match).Here we are using an artificial intelligent (AI) [6-8] based technique to complete this task. Here we are trying to differentiate the sound of APJ Abdul Kalam sir and the Donald Trumph president of America. Here we are using K-nearest neighbor (KNN) [9]]algorithm which is a powerful supervised learning tool. We have extract the four parameters of sound and these are 'power spectral density'[10], ‘spectral centroid’ [11], ‘spectral flatness’ [12] and ‘roll-off frequency' [13].

II. SPEAKER INFORMATION

In this paper we have use the python programming to create the model and run the algorithms. We have chosen two speakers first is Dr. APJ Abdul Kalam and second is Mr. Donald Trumph. We have chosen these two speakers because there is a lot of variation in both of these sounds. If we take the both samples of Indian citizen then model can be create some confusion in decision making. The idea is to identify the speakers even if they speak the same password. The following figure shows the system’s architecture.

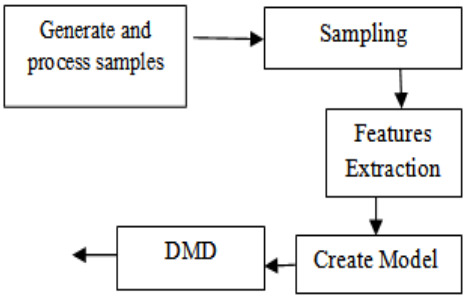

Fig.1 Architecture of SRS

2.1) Generation and Process on samples:

We have downloaded a speech of both Dr. Kalam and Trumph. After removing the noise from the speech we had a sample size of 56 min. We chunk 1099 samples from these speeches of 55 min. the sampling rate of the signal was 11025 Hz.

2.2) Sampling

Computer is a digital device. It works only on the discrete samples. Here is sampled the signal with the rate of 11025 Hz.

2.3) Feature extraction

The following features we have constructed for speaker identification system.

i) Power spectral density

A Power Spectral Density (PSD) is the measure of signal's power content versus frequency. A PSD is typically used to characterize broadband random signals. The amplitude of the PSD is normalized by the spectral resolution employed to digitize the signal.

$$S_x(f) = \int_{-\infty}^{+\infty} R_x(\tau) e^{-jw\tau}\, d\tau$$

Where $R_x(\tau)$ is an autocorrelation function.

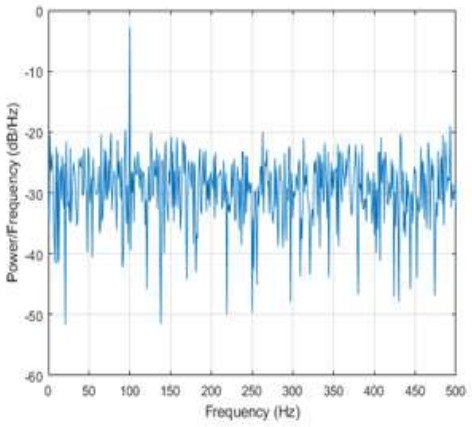

Fig. 2 Power spectrum of voice signal.

Power spectrum of voice signal is shown in Fig.1. This spectrum is always different for different kind of speech signal

ii) Spectral Centroid

The spectral centroid is commonly associated with the measure of the brightness of a sound. This measure is obtained by evaluating the "center of gravity" using the Fourier transform's frequency and magnitude information. The individual centroid of a spectral frame is defined as the average frequency weighted by amplitudes, divided by the sum of the amplitudes, or:

$$S.C. = \frac{\sum_{K=1}^{N} KF[K]}{\sum_{K=1}^{N} F[K]}$$

Here, F [k] is the amplitude corresponding to bin k in DFT spectrum. Spectral centroid is shown in Fig. 3.

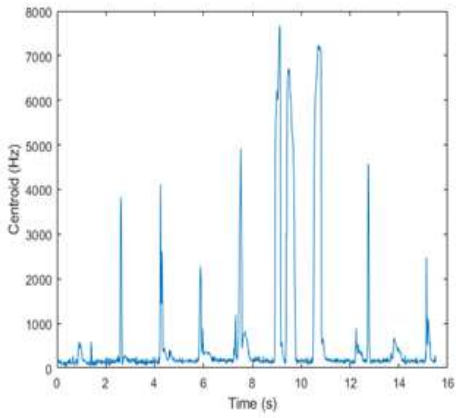

Fig. 3 Spectrum centroid of voice signal.

iii) Spectral Flatness

White noise has a flat power spectrum. So a reasonable way to measure how close a sound is to being pure noise is to measure how flat its spectrum is. Spectral flatness is defined as the ratio of the geometric mean to the arithmetic mean of a power spectrum.

iv) Roll-off Frequency

The frequency above or below which a filter begins to filter out the harmonics of the waveform. As the rolloff frequency is raised or lowered, more of the harmonics of the sound will be filtered out. Specifically, the frequency at which the response of an equalizer or other audio device is reduced by 3dB, This is also sometimes called the half-power point and can refer to both lowpass and highpass response curves. The rolloff frequencies of an amplifier are the frequencies where the output voltage drops to 0.707 of the middle range output. A decrease of the voltage by a factor of 0.707 is equivalent to -3dB, so these critical frequencies are often referred to as the 3dB down points.

III. EXPERIMENTAL SETUP

Experimental testing requires dataset with speakers andevaluation criteria.

A. Dataset

Dataset having the following features.

3. 65.5 min utterances per speaker
4. 1099 samples
5. re-sampled to 11025Hz and 16 bits.
6. hamming window and FFT size is 512
7. Training samples: 70%
8. Testing samples: 30%
9. signals are sampled at 8 kHz
10. K = 3

Throughout this study, verification experiments highlight the feasibility of using power spectrum density of the signal to improve the system's performance. The results obtained with different feature vectors show that the use of MFCC coefficients together with PSD yielding to more significant results. In our speech recognition system, for feature extraction step, the Welch algorithm is applied to speech signal to estimate the power spectrum density. For Welch's parameters, we have used hamming window and FFT size is 512. With these parameters, Welch algorithm generates a vector of 257 elements. 02 speakers are tested. Total 1099 utterances are collected in our experiments. We have used 70% utterances for training and 30% utterances for testing. In all our experiments, the speech signals are sampled at 8 kHz. We propose to identify speakers based on their vocal password. The system differentiates between speakers even if they use the same word.

IV. RESULTS AND DISCUSSION

Table one is showing the sample data set of extracted features.

Table.1 Sample of features

582	0.380663	0.541502	0.049808	-21.649140
1615	0.281897	0.410802	0.101503	-19.665713
1606	0.323037	0.497632	0.087086	-16.026001
1381	0.344100	0.491245	0.110512	-18.204625
1400	0.271033	0.409208	0.102737	-18.368770
1395	0.489375	0.651003	0.156984	-18.850644
121	0.357549	0.614076	0.054397	-23.326698
331	0.353143	0.557076	0.059503	-21.328191
...	...	...	...	...

```
#Import scikit-learn metrics module for accuracy calculation
from sklearn import metrics
# Model Accuracy, how often is the classifier correct?
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
```

```
Accuracy: 0.90456874523
```

```
if(y1[0][0]>0.7):
    print("The speaker is APJ abdul kalam")
elif(y1[0][1]>0.7):
    print("The speaker is Donald Trump")
else:
    print("Unknown")
```

Fig.4 Result of program executed

The model is trained by KNN algorithm and taking the 3 nearest neighbor. The accuracy of the system is 90% and the value of K score is 94. The system is providing the output result in probabilistic manner. There are two classes i.e. 0,0 and 0,1. It showing that if class [0,0] probability is greater than 0.7 then the result go in the favour of Dr. kalam and other wise it will choose the Trumph.

V. CONCLUSION

In this paper, we developed a method based on power spectrum density estimation to extract speech features. We proposed a word-dependent identification system where all speakers use the same password. With the database used, the recognition rate exceeded 90%. Welch algorithm generates a vector of 257 elements. We believe that in addition of the accuracy obtained by the system, the time running is also optimized by using one vector of 257 elements. Also, more training data and good pre-processing methods can enhance the accuracy of the system.

REFERENCES

[1]. N. Dehak, Z. Karam, D. Reynolds, R. Dehak, W. Campbell, and J. Glass, "A Channel-Blind System for Speaker Verification", Proc. ICASSP, pp. 4536-4539, Prague, Czech Republic, May 2011.

[2]. N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front- End Factor Analysis for Speaker Verification", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 19, No. 4, pp. 788-798, May 2011.

[3]. R. Togneri and D. Pullella, "An Overview of Speaker Identification: Accuracy and Robustness Issues", In: IEEE Circuits And Systems Magazine, Vol. 11, No. 2 , pp. 23-61, ISSN : 1531-636X, 2011.

[4]. D. A. Reynolds, "A Gaussian Mixture Modeling Approach to Text Independent Speaker Identification", PhD Thesis. Georgia Institute of Technology, August 1992.

[5]. T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors", Speech Communication 52(1): 12-40, 2010.

[6]. D. A. Reynolds, "Robust Text-Independent Speaker Identification Using Gaussian Mixture SpeakerModel", IEEE Transactions on Speech and Audio Processing. vol. 3, n. 1, pp. 72-83, January, 1995.

[7]. D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models", Digital Signal Process., vol. 10, no. 1–3, pp. 19–41, 2000.

[8]. S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, "Hidden Markov model toolkit (htk) version 3.4 user's guide", 2002.

[9]. D. A. Reynolds, "Speaker identification and verification using Gaussian mixture speaker

models," Speech Commun., vol. 17, no. 1–2, pp.91– 108, 1995.
[10]. W. Campbell, D. Sturim, and D. Reynolds, "Support vector machines using GMM supervectors for speaker verification," IEEE Signal Process. Lett., vol. 13, no. 5, pp. 308–311, 2006.
[11]. D. Reynolds, "Experimental evaluation of features for robust speaker identifi cation," IEEE Trans. Speech Audio Process., vol. 2, no. 4, pp. 639–643, 1994.
[12]. Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J.G., Pallett, D. S., and Dahlgren, N. L., "DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus CDROM, "NIST, 1993.
[13]. A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET Curve in Assessment of Detection Task Performance", in EUROSPEECH, vol. 4, pp. 1895–1898, 1997.

Vaibhav Bhardwaj" Speaker Recognition Using Machine Learning Based Method" International Journal of Engineering Research and Applications (IJERA), vol. 9, no. 9, 2019, pp. 27-30