International Journal of Electrical and Electronics Research ISSN 2348-6988 (online)
Vol. 6, Issue 4, pp: (9-16), Month: October - December 2018, Available at: www.researchpublish.com

Speaker Identification using Discrete Wavelet Packet Transform Technique with Irregular Decomposition and BIC

[1]Jyoti Devi, [2]Sukhvinder Kaur

[1]Student, Department of Electronics and Communication Engineering, Swami Devi Dyal Institute of Engineering and Technology, Haryana, India
[2]H.O.D., Department of Electronics and Communication Engineering, Swami Devi Dyal Institute of Engineering and Technology, Haryana, India

Abstract: **Speaker Recognition System is increasingly being deployed as a more natural means for recognition of people. Speaker Recognition (SR) is the process of automatically recognizing the person speaking on the basis of the information obtained from the speech features. In today world scenario, speaker recognition system is very popular in voice verification for identity and access control to services. Speaker identification and verification are basic part of speaker recognition system. Basically it is a process of automatically recognizing the identity of speaker on the basis of individual information extracted from speech signals. Our main purpose is to reduce error in the system and increase the efficiency. In this report, it is done with the help of different feature extraction algorithm. We have introduced a new approach for speaker recognition. This new system firstly frames the speech signals and then these signals are compressed using DWT for noise reduction and better sampling frequency. Furthermore, features of compressed signals are extracted with the help of Discrete Wavelet transform (DWT), Non-Linear Energy operator (NEO) based DWT, and Irregular Discrete Wavelet Packet transform (IRR DPWT). The distance metrics incorporated are Delta Bayesian Information criteria (delta BIC). At the end results are evaluated with Detection Error Tradeoff (DET) curve and Receiver Operator Characteristics (ROC) curve by finding the area under curve (AUC). The main purpose of speaker identification system is to provide a reliable decision, accept or reject a speaker, given a claimed identity and a recording of a speaker phrase. After comparing area under curve we can say that the best result is shown by DWT NEO with BIC distance metric.**

Keywords: **Speaker Recognition System, Discrete Wavelet transform (DWT), BIC distance metric.**

1. INTRODUCTION

In our everyday lives there are many forms of communication, for instance: body language, textual language, pictorial language & speech etc.

However amongst those forms speech is always regarded as the most powerful form because of its rich dimensions character.

Except for speech text (words), the rich dimension also refers as the gender, attitude, emotion, health situation & identify of a speaker. Such information is very important for an effective communication.

Speaker recognition is the process of recognizing automatically in speech signal who is speaking on the basis of individual information included. This technique uses the speaker voice to verify their identity and provides control access to services such as voice dialing, database access services, information services, voice mail, and security control for confidential information area, remote access to computers and several other fields where security is the area of concern.

Speech is a complicated signal produce as a result of several transformation occurring at several level: semantic, linguistic, articulatory and acoustic. Differences in these transformations are reflected in the difference in the acoustic properties of the speech signal. Beside there are speaker related differences which are a result of a combination of anatomical difference inherent in the vocal tract and the learned speaking habits of different individuals in speaker recognition. Speaker recognition system helps in the basic purpose of speaker identification. The speaker identification speaker designed has potential in several security applications. For example may include, users having to speak a PIN (Personal Identification Number).

Speaker Recognition process includes two steps:

- Speaker Identification
- Speaker Verification

Speaker Recognition can be classified into number of categories. Figure 1 below provides the various classification of speaker recognition.

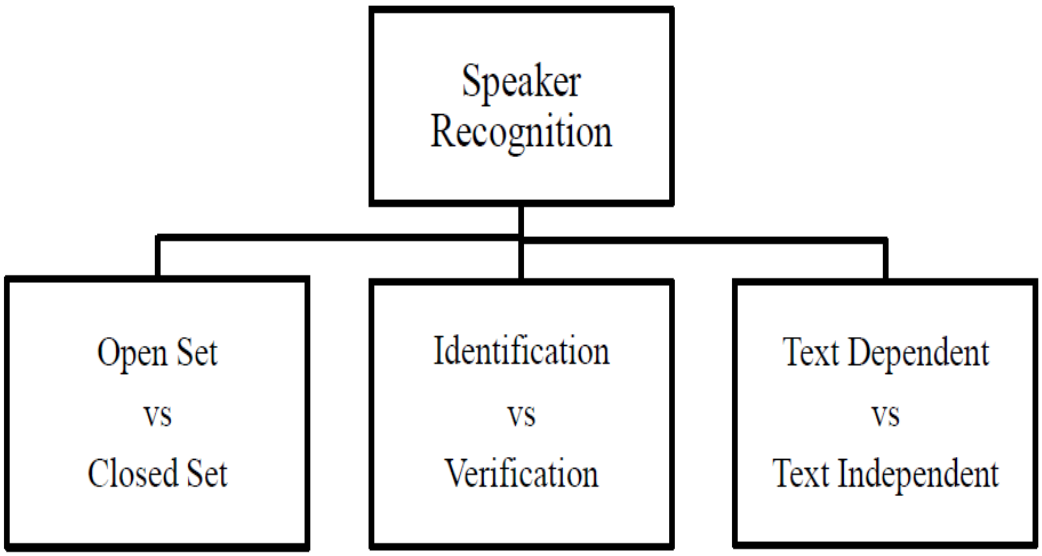

Figure 1.1 Classification of Speaker Recognition

APPLICATIONS:

Speaker recognition technologies are used in wide application areas. There is truly no limit to the applications of speaker recognition. If audio is involved, one or more of the speaker identification branches may be used. Some of the major applications of speaker recognition have been in the following few sections.

- **Speaker Recognition for Authentication**
- **Speaker Recognition for Surveillance**
- **Forensic Speaker Recognition**
- **Security**
- **Speech recognition**
- **Multi - Speaker tracking**
- **Audio and Video Indexing Applications**
- **Biometric Applications**

2. PROPOSED WORK AND RESEARCH METHODOLOGY

Designing of Proposed Speaker Identification System:

Speaker recognition process divided into two steps: Speaker identification and speaker verification.

- Speaker identification is the process of extracting the features associated with the speaker. The feature extracting tools that we will use are BIC and NEO.
- Speaker verification on the other hand is the process of accepting or rejecting the speaker identity by matching algorithm based on their feature. Distance metrics algorithm performed matching process. BIC and KL2 is one of the distance measuring algorithms. T-Test distance matric also use for to decrease the error rate and increase the efficiency.

.

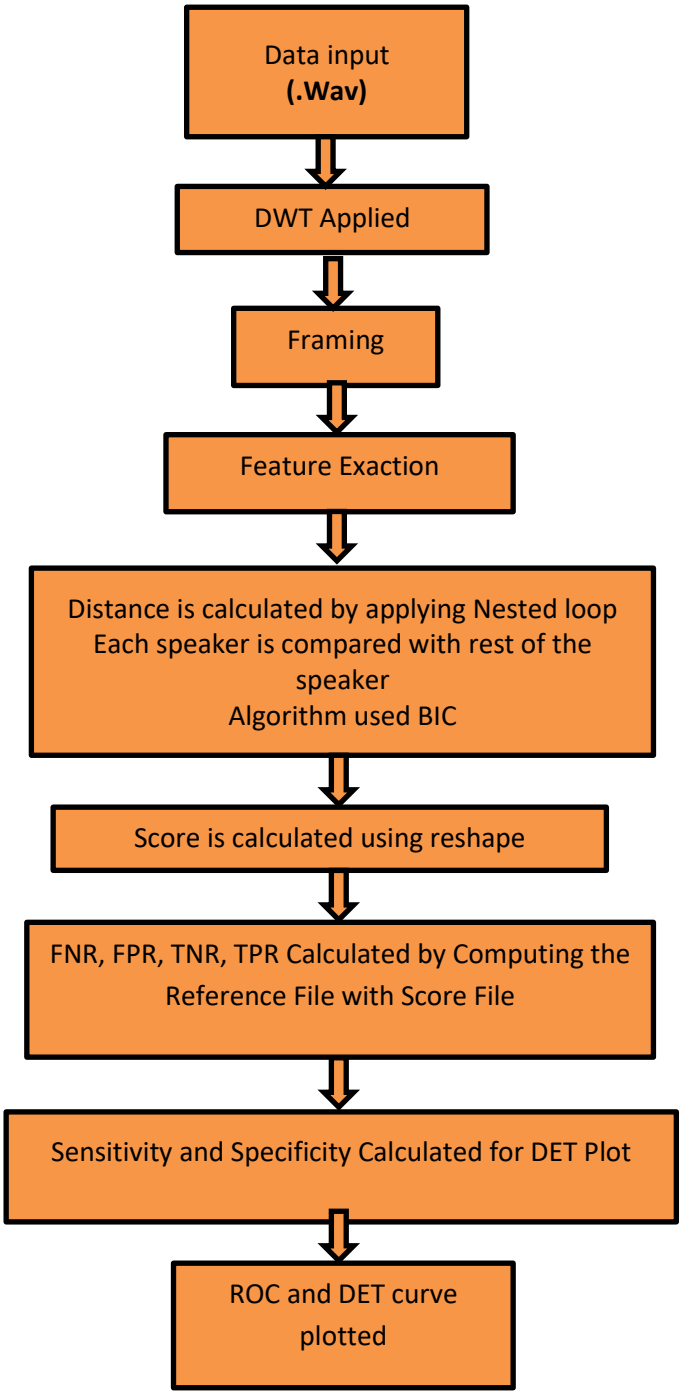

Figure 2: Algorithm Used to deign Speaker Recognition System

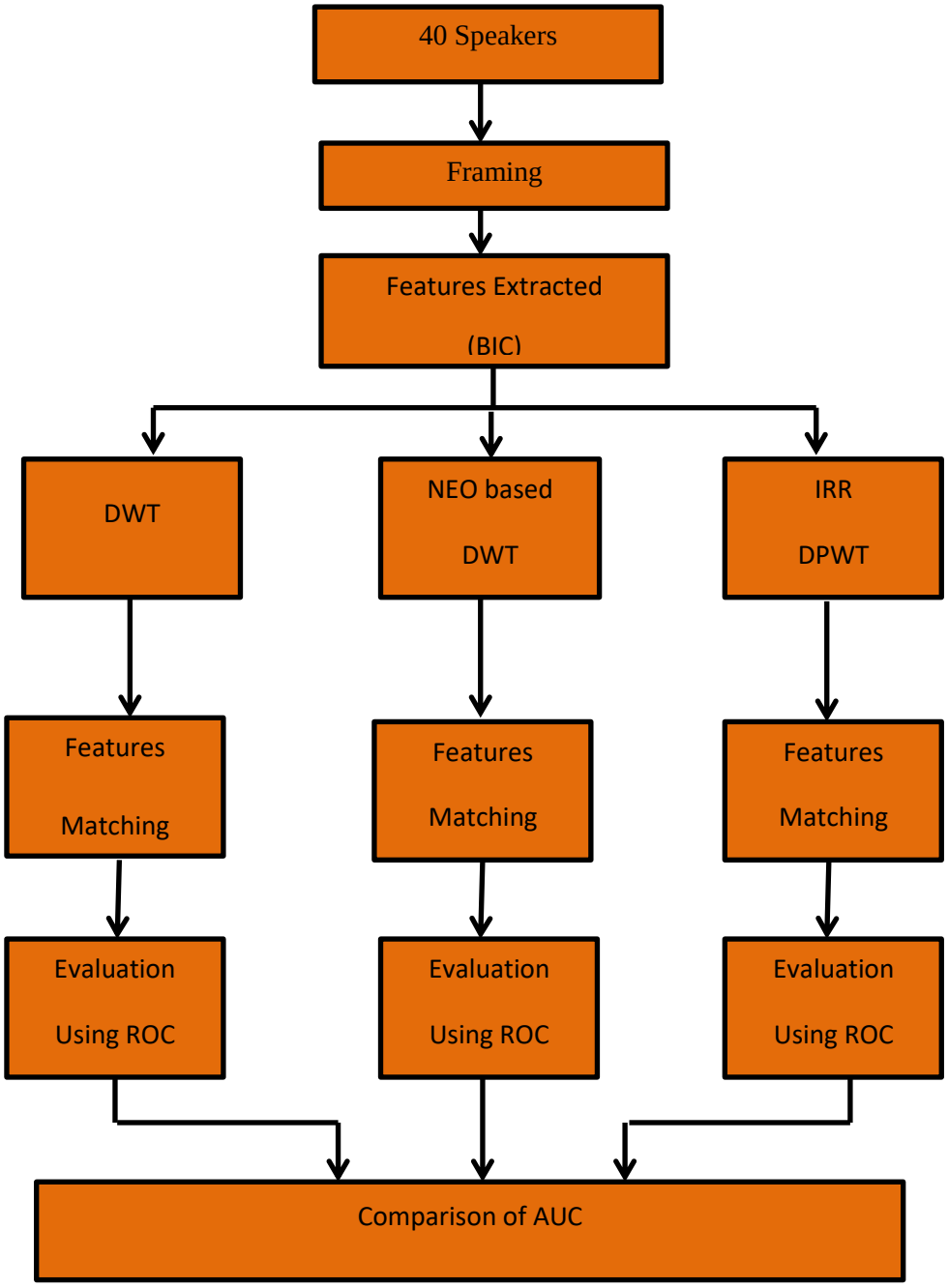

Figure 3: Steps Used in Speaker Identification System

- **Framing:** In figure 3 we can see the whole processes that are used in our thesis work. We sample the output of an analog microphone to produce the speech signal. Using amplitude quantization along with sampling in the frequency domain, we are able to represent the speech signal into the speech waveform. Numbers of different standards are available for storing and transmitting the speech waveform. Now speech is in the form of signal in MATLAB. The database of audio file is stored in .wav format with sampling frequency of 48000 samples per sec.

DWT is a powerful mathematical tool in many areas of science and engineering especially in the field of speech and image compression which uses multi resolution filters banks for the signal analysis. A wavelet is a basic idea of the wavelet transform is to represent an arbitrary signal 'S' as a super position of a set of such wavelets or basis functions. These basis functions are obtained from a signal prototype wavelet called the mother wavelet by dilation and translation. The discrete wavelet transform as shown in figure for one dimensional signal. It is defined in the following equation

$$W(i, j) = \sum i \sum j \; X(j) 2^{-\frac{i}{2}} \; \varphi(2^{-i} n-k) \qquad \ldots (1)$$

Where φ (t) is the basic analyzing function called the mother wavelet. In DWT a time – scale representation of a signal is obtained by digital filtering techniques. A low frequency component of a signal is more significant than high frequency since the low frequency component since the low frequency component have maximum information content. The DWT is computed by successive low pass filtering and high pass filtering of the discrete time domain signal. This algorithm is called the Mallat algorithm. At each level the decomposition of the signal has two kinds of outputs. The low frequency components are known as the approximation a[n] and high frequency component are known as the detailed d[n]. At each decomposition level, the half band filters produce signals spanning only half the frequency band. This doubles the frequency resolution as the uncertainty in frequency is reduced by half. With this approach the time resolution becomes good at high frequencies. The filtering and decimation process is continued until the desired level is reached.

The DWT of the original signal is then obtained by concatenating all the coefficients a[n] and d[n] starting from the last level of decomposition. The successive high pass and low pass filtering of the signal can be depicted by the following equations:

$Y_{High}[k] = \sum_n x[n]g[2k - n]$ **…….(2)**

$Y_{Low}[k] = \sum_n x[n]h[2k - n]$ **……..(3)**

Where Y_{High} and Y_{Low} are the outputs of high pass and low pass filters obtained by sub sampling by 2.

- **Feature Extraction:** Any feature of speech that encompasses segments larger than the phonetic segments is called a suprasegmental feature. The feature extraction can be considered as a data reduction process that attempts to capture the essential characteristics of speaker with a small data rate. There are various techniques for extracting speech feature in the form of coefficients such as the Mel Frequency Cepstral Coefficients (MFCC), Non-Linear Energy Operator (NEO).

- **Analysis of System:** Analysis is done by distance measuring algorithm. BIC and KL2 are already exciting algorithm and we proposed a new algorithm for distance measuring. T-Test is used to calculate the similarity between two speakers. Analysis is done with Distance measuring metrics by calculating their score file and computing it with reference file.

- **Performance and Evaluation:** DET and ROC curve is plotted to measure the performance of algorithm. DET is tradeoff between error rate i.e. false positive rate and false negative rate. It gives the miss error rate and false alarm rate.

3. RESEARCH METHODS

- **Non-linear Energy Operator:** Amplitude modulation –frequency modulation (AM-FM) of speech signal plays an important role in speech perception and recognition. The Am-FM model has been successfully used in various areas of signal processing. Specifically in speech processing this model has been applied for speech analysis and modeling, speech synthesis, emotion, speech and speaker recognition. A nonlinear energy operator (NEO) is used to track the energy required to generate an AM-FM signal and separate it into amplitude and frequency components. The nonlinear energy operator can be capable of estimating the speech signal energy. The energy operator is a nonlinear differential operator derived from analysis of the energy in a second-order harmonic oscillation system. In this investigation a time domain technique is developed to extract the pitch frequencies from voiced speech signals.

$\Phi [x' (t)] = [x(t)^2] - [x (t)x''(t)]$ **……..(4)**

And the discrete version of the operator can be defined as

$\Phi (x(n)) = x^2[n] - x[n-1]x[n+1]$ **…… (5)**

The NEO approach to demodulation has many attractive features such as simplicity, efficiency, and adaptability to instantaneous signal variations.

Dyn Operator The definition of Teager Energy Operator was first given by J.F. Kaiser in the year 1991. It was a nonlinear approach to the speech signals. Similarly J. Rouat defined another nonlinear energy operator called Dyn Operator in the same year. The Dyn operator is also an energy tracking operator and it can be defined as

$\text{Dyn}[x(t)] = x(t)\, x(t)$ **…….(6)**

This operator is not so conventional as compared to the TEO and has some less importance in the field of speech signal processing. Both the energy operator defined here has some special advantages and disadvantages.

- **Irregular Discrete Wavelet Packet Transform (DPWT):** DWT decomposes the data in a dyadic form, and the recursive decomposition only act on the low frequency content that is "approximation". The high frequency content is preserved as "detail". Approximation is abbreviated as A and Detail is abbreviated as D.

- **Bayesian Information criterion (BIC):** The input stream is a Gaussian process in the cepstral space. We present a maximum likelihood approach to detect turns of a Gaussian process the decision of a turn is based on the Bayesian Information criterion (BIC), a model selection criterion in the statistics literature. The problem of the model identification is to choose one among a set of speaker models to describe a given set data set. We often have speaker of a series of models with different number of parameters. It is evident that when the number of parameters in the models is increased. The likelihood of the training data is also increased however when the number of parameters is too large, this might cause the problem of overtraining. Several criteria for model selection have been introduced in the statistics literature, ranging from non-parametric methods such as cross-validation to parametric methods such as the Bayesian Information Criterion (BIC).

BIC is a likelihood-based criterion penalized by the model complexity. This report focuses on the distance measure between test speaker feature and the database features.

Let us consider two audio segments (i,j) of parameterized acoustic vectors of $X_i=\{X_1,X_2 \dots \dots \dots X_N\}$ and $X_j=\{X_1,X_2 \dots \dots \dots X_K\}$ of length N_i and N_j respectively, and with mean and standard deviation values $\mu_i, \sigma_i, \mu_j, \sigma_j$. Each one of these segments is modeled using Gaussian processes $M_i(\mu_i, \sigma_i)$ and $M_j(\mu_i, \sigma_j)$ which can be a single Gaussian or a Gaussian Mixture Model (GMM). On the other hand, agglomerate of both the segments into X, with mean and variance μ, σ and the corresponding Gaussian process M(μ, σ).

For a given acoustic segment X_i , the BIC value of model M_i is determined by:

BIC (M_i) = log L (X_i, M_i) - $\lambda\frac{1}{2}$# (Mi) log (Ni) **…….. (7)**

log L(X_i, M_i) is log-likelihood of the data given the considered model, λ is a free design parameter dependent on the data being modeled, estimated using development data N_i is the number of frames in the considered segment and #(M_i) the number of free parameters to estimate in model (M_i). Such expression is an approximation of the Bayes Factor (BF) (Kass and Raftery (1995), Chickering and Heckerman (1997)).

In order to use BIC to evaluate the speaker change, it evaluates the GLR (Generalized Likelihood Ratio) through following equation. The closest pair of segment id updated and the process stopped when penalized minimum distance score is greater than specific threshold (typically 0).

ΔBIC = R (i,j) – λP **…….(8)**

Where P is the penalty term, which is a function of the number of free parameters in the model. For a full covariance matrix it is

P = $\frac{1}{2}$(d+$\frac{1}{2}$d (d+1)) log (N) **…….. (9)**

d is the dimension of the space. The term penalty accounts for the likelihood increase of bigger models versus smaller ones.

The term R(i) can be written for the case of models composed on a single Gaussian as

R (i,j) = $\frac{N}{2}$log|$\sum X$| - $\frac{Ni}{2}$log|$\sum Xi$| - $\frac{Nj}{2}$log|$\sum Xj$| **…….(10)**

For cases where GMM with multiple Gaussian Mixture are used, eq. is written as

ΔBIC(M_i) = log L(X, M) – (log L(X, M) – (log L(X_i, M_i)) + log(X_i, M_i) – λΔ# (i,j) log(N) **…….(11)**

Where Δ# (i,j) is the difference between the number of free parameters in the combined model versus two individual models.

Although ΔBIC (i,j) is the difference between two BIC(i) criteria in order to determine which models suits better the data. The positive value of ΔBIC (i,j) shown that speaker is different. If the ΔBIC (i,j) value comes out to be negative then the speaker is same.

4. RESULTS

- **Experimental Results:**

The verification performance of speaker identification system is normally evaluated using the receiver operating characteristics (ROC) or detection error tradeoff curve (DET).DET curve indicate how miss rate and false are related with each other. This will give the efficiency and error rate as shown in table 5.1. Graph of DET curve with DWT, NEO based and irregular DWT are shown in figure 5.1. DET represent the performance of speaker identification system. In DET curve we plot error rate on both axes.

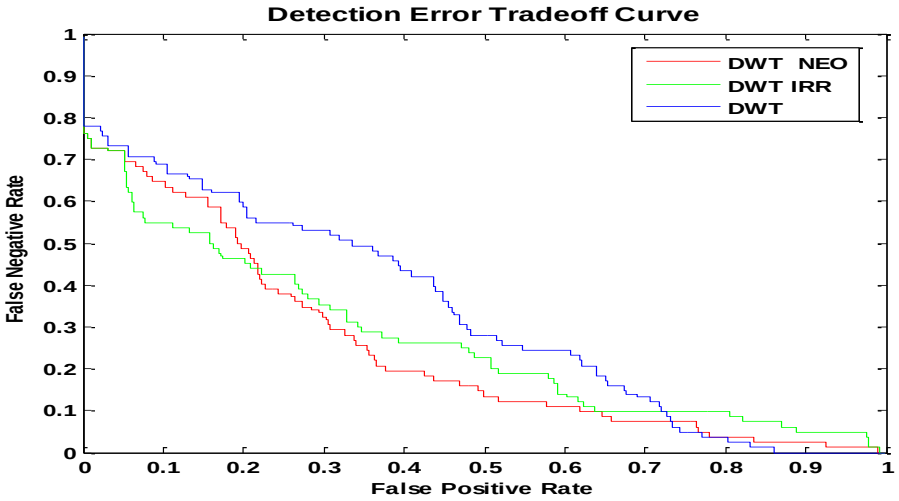

Figure 2.1 The Combine DET Curve Using DWT, NEO Based DWT and IR

Table 2.1 Efficiency and Error Result

	DWT		NEO DWT		DWT IRR	
	Efficiency	**EER**	**Efficiency**	**EER**	**Efficiency**	**EER**
BIC	66.31	33.69	75.21	24.79	72.99	27.01

We can clearly see that in table 2.1 that the minimum error occurs when NEO DWT algorithm is used with BIC. ROC curve with different feature are shown in fig 5.2 below. The ROC curve is a tool for diagnostic test evaluation. The ROC curves indicate the sensitivity and specificity pair corresponding to a particular decision threshold.

The graph in the figure 2.2 shows three ROC curve (DWT, NEO based DWT and IRR DWT) representing worthless test. Accuracy is measured by the area under ROC curve. In ROC curve the true positive rate (sensitivity) is drawn in function of the false positive rate (specificity). The ROC curve that passes through the upper left corner is 100% sensitivity and 100% specificity. Therefore the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test. In our thesis work the out of the three curve the NEO based DWT curve is more close to the upper left corner compare to the others.

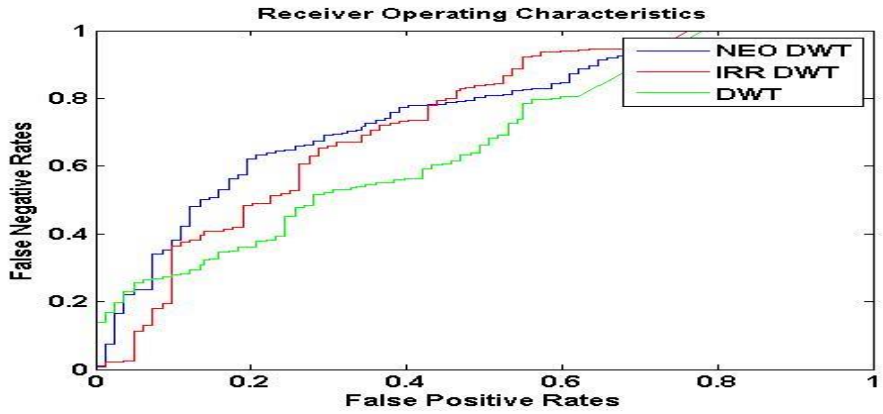

Figure 2.2 The ROC Curve With Using DWT, NEO Based DWT and IRR DWT

5. CONCLUSION

Speaker recognition system is very popular in voice verification for identity and access control to services. Speaker identification and verification are basic part of Speaker recognition system. In this report, it is done with the help of three type of feature extraction simple DWT and NEO based DWT and IRR DWT. We have introduced a new approach for speaker recognition.

This new system firstly frames the speech signals and then these signals are compressed using DWT for noise reduction and better sampling frequency. Furthermore feature of compressed signal are extracted with the help of nonlinear energy operator (NEO).

These feature are further used for identification and verification of speaker voice. The distance metrics incorporated are Delta Bayesian Information Criteria (delta BIC). At the end, results are evaluated with detection Error Tradeoff (DET) curve and Receiver Operator Characteristics (ROC) curve by finding the area under curve (AUC). The best result is shown by NEO based DWT feature.

We have compared the output of three feature i.e. simple DWT and NEO based DWT and IRR DWT and we find that NEO based DWT more efficient.. The speaker recognition system using the T-test Distance Metric count minimum False Alarm when compared with others. BIC procedure and KL2 are same in case of False Positive.

REFERENCES

[1] Ankur Maurya, Divya Kumar, R.K. Agarwal,"Speaker Recognition for hindi speech signal using MFCC – GMM approach," *Procedia computer science 125 (2018) 880-887*.

[2] Suma Paulose, Dominic Mathew (2017)," Performance evalution of different modeling method and classifier with MFCC and IHC features for speaker recognition,".

[3] Gunjan Jhawar, Prajacta Nagraj, and P. Mahalakshmi," Speech Disorder Recognition using MFCC," *International conference on communication and signal processing, April 6-8, 2016, India.*

[4] A. Maazouzi, A.Laaroussi, N. Aqili, M. Raji and A.Hammouch," Speech Recognition System using burg Method and Discrete wavelet transform," *2nd international conference on electrical and information technologies 2016.*

[5] Shanthini Pandiaraj and K.R. Shankar Kumar," Speaker Identification using Discrete wavelet Transform,".

[6] Alexandros Georgogiannis, Vassillis Digalakis, " Speech Emotion Recognition Using Non-Linear Teager Energy Based Features in Noisy Environments," *20th European Signal processing Conference,"Bucharest, Romania, August 27-31, 2012.*

[7] Nilu Singh, R.A. Khan , Raj Shree," Application of Speaker Recognition," *International conference on modeling, optimization and computing Procedia Engineering 38 (2012) 3122-3126.*

[8] Gyanendra Kumar, Verma, Uma Shankar Tiwary," Text indepent speaker identification using wavelet transform," *Computer and communication technology (ICCCT), 2010 international conference on, pp. 130-134 ,2010.*

[9] Hasan Ates F., Michael Orchard T. ," Speaker Coding Algorithm for wavelet image compression," *IEEE transaction on image processing, VOL. 18. NO.5, May 2009*