Speaker Verification System based on Type-2 Fuzzy Gaussian Mixture Models

Mrs. S. Gayathri
Assistant Professor,
Department of Electronics and Communication Engineering,
K.Ramakrishnan College of Technology,
Tiruchirappalli, Tamilnadu, India.

***Abstract*— This paper proposes the use of Type-2 Fuzzy Gaussian Mixture Models (T2 FGMMs) in a Speaker Verification system. Type-2 Fuzzy Gaussian Mixture Model is an extension of GMM based on Type-2 Fuzzy Sets (T2 FSs). It uses Footprint Of Uncertainty (FOU) and interval secondary Membership Function (MF) to handle GMMs uncertainty in estimating the parameters mean μ and covariance matrix ∑. The proposed methodology for Speaker Verification system uses speech files from the TIMIT database for training and testing phases. The test features are applied to the trained model and the verification decision is made using Generalized Linear Model (GLM). The experimental results showed that T2 FGMMs provide a low Equal Error Rate (EER) than GMM indicating that T2 FGMMs gives better performance than GMMs in a Speaker Verification system.**

***Keywords -Gaussian Mixture Model, Fuzzy Sets, Type-2Fuzzy Gaussian Mixture Model, Footprint Of Uncertainty, Generalized Linear Model, Equal Error Rate*.**

I. INTRODUCTION

Speaker Verification refers to the task of determining the claimed identity of the unknown speaker. It plays a major role in biometrics and security. It is used in Automatic Speaker Verification (ASV) systems for access control. They are also used for voice telephony, voice mail, tele-banking, tele-shopping and secure transfer of confidential information. In Speaker Verification, GMM is used to model the distribution of feature vectors of speaker utterances.

Gaussian Mixture Models (GMMs) are widely used in modeling because of their universal approximation ability. They can model any density function if they contain enough mixture components [5].GMMs are used for clustering, object tracking, background subtraction, feature selection, signal analysis, learning and modeling [6].GMM based methods have been developed to meet specific applications such as adapted GMMs [1], Mahalanobis distance based GMMs [6], wrapped GMMs [6] and Active curve axis or GMMs (AcaGMMs) [6].

Real world problems often encounter uncertainties in the system parameters due to noisy data. The various sources of uncertainties occurring in a Speaker Verification system can be grouped into [3]: (a) Insufficient or noisy training data can make parameters of the model λ uncertain, so that the mapping of the model λ *is* also uncertain. (b) The relationship between training data and unknown test data is uncertain due to limited prior information. (c) The linguistic labels can be uncertain since the same observation may mean different things to different people. All of these uncertainties can be considered as fuzziness resulting from incomplete information, i.e., fuzzy observations, fuzzy models, and fuzzy labels. The nature of uncertainty in a Speaker Verification system can be categorized into three types [8]: (i) Fuzziness (vagueness), which results from the imprecise boundaries of fuzzy sets. (ii) Non-specificity (information based imprecision) which is connected with sizes (cardinalities) of relevant sets of alternatives. (iii) Strife (discord), which expresses the conflicts among the various sets of alternatives.

The uncertainties occurring in the GMM parameters can be handled by Type-2 Fuzzy Sets (T2 FSs) [5]. The Type-2 Fuzzy Sets are used to describe the fuzziness of the GMM parameters: the mean vector μ and the covariance matrix ∑ [5]. These Type-2 Fuzzy Sets (T2 FSs) can describe and estimate the uncertainties due to their three dimensional fuzzy Membership Functions (MFs) [2]. In contrast, type-1 fuzzy sets cannot directly model the uncertainties due to their crisp MFs and two dimensional structures of the MFs [2]. Type-2 membership functions can simultaneously evaluate randomness and fuzziness by using Footprint Of Uncertainty (FOU) and interval secondary Membership Functions as shown below [4].

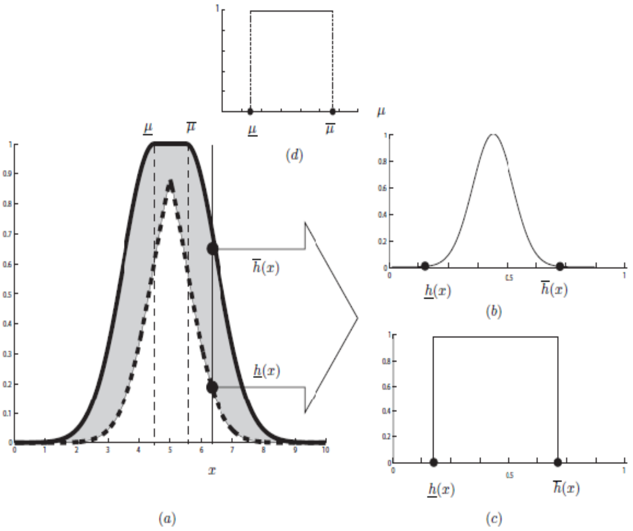

Figure (1): The three-dimensional Type-2 Fuzzy Membership Function (T2 MF) (a) shows the primary membership with the lower (thick dashed line) and upper (thick solid line) membership functions, where and are the lower and upper bounds given the input x respectively. The shaded region is the Footprint Of Uncertainty (FOU). (b) shows the Gaussian secondary membership function. (c) Shows the interval secondary membership function. (d) Shows the mean μ has a uniform membership function.

The various features of Type-2 Fuzzy Sets comprises of [3]: T2 FSs can represent more uncertainties simultaneously by using primary and secondary Membership Functions (MFs). T2 FSs can handle uncertainties covered by Foot Print of Uncertainty (FOU) efficiently by propagating the uncertainties. Different defuzzication techniques of T2 FSs may produce different results giving additional flexibility to design systems.

Based on these Type-2 Fuzzy Sets (T2 FSs), a new extension of GMM is obtained known as Type-2 Fuzzy GMM (T2 FGMM) which is the key part of the proposed Speaker Verification system. Section II describes the proposed system for Speaker Verification using T2 FGMM. Section III gives the experiments conducted with their results during the training and testing phases of the proposed Speaker Verification system. Section IV discusses the future direction and conclusion obtained from the observations of the previous section.

II. PROPOSED SYSTEM

2.1 System Description

The proposed system for Speaker Verification using T2 FGMM is shown below. The Speaker Verification system consists of training and testing phases. The speech processing modules consists of Pre-processing using VAD, Feature extraction using MFCC and Modeling the features using T2 FGMM.

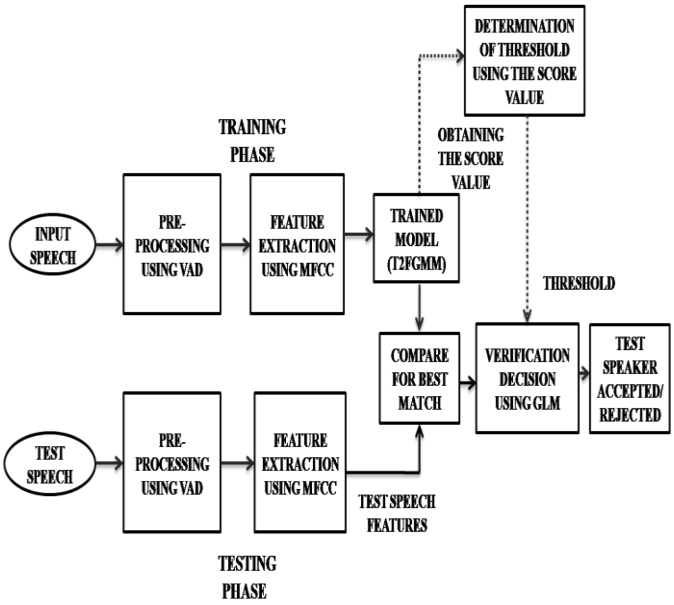
DETERMINATION OF THRESHOLD USING THE SCORE VALUE
TRAINING PHASE
OBTAINING THE SCORE VALUE
INPUT SPEECH
PRE-PROCESSING USING VAD
FEATURE EXTRACTION USING MFCC
TRAINED MODEL (T2FGMM)
THRESHOLD
COMPARE FOR BEST MATCH
VERIFICATION DECISION USING GLM
TEST SPEAKER ACCEPTED/ REJECTED
TEST SPEECH
PRE-PROCESSING USING VAD
FEATURE EXTRACTION USING MFCC
TEST SPEECH FEATURES
TESTING PHASE

Figure (2): Overview of Speaker Verification system using Type-2 Fuzzy Gaussian Mixture Model (T2 FGMM)

The initial step is pre-processing of the speech signal. Here, it is done by Voice Activity Detection (VAD). It removes the silence portion present at the beginning and end of the adjacent samples. It determines the voiced and unvoiced portion present in the speech signal. This voice activity detected signal is given as input for feature extraction. After this, Mel-Frequency Cepstral Coefficients (MFCCs) are extracted. This in turn provides us with useful feature vectors that really establish the characteristics of the speaker.

During training, these useful feature vectors are modeled using T2 FGMM to develop a trained speaker model. In the testing phase, the features of the test speech are applied to the trained model and score values are calculated. Now, the Speaker Verification process is made by using Generalized Linear Model (GLM) based on the interval likelihoods of T2 FGMMs. Finally, the Speaker is accepted or rejected based on the threshold value.

2.2 Speech Database

The proposed system for Speaker Verification uses the TIMIT (Texas Instruments Massachusetts Institute of Technology) database. TIMIT contains a total of 6300 sentences, 10 sentences spoken by each of 630 speakers from 8 major dialect regions of the United States.

2.3 Speech Pre-Processing using VAD

Rabiner and Sambur [12], proposed an algorithm for voice activity detection that is based on measurements of energy and zero crossing rates.

The block diagram for Voice Activity Detection algorithm is shown below.

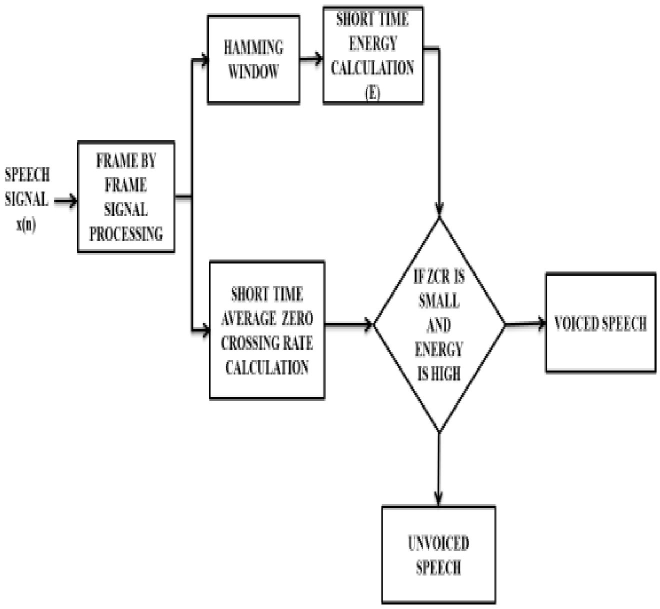
SPEECH SIGNAL x(n)
FRAME BY FRAME SIGNAL PROCESSING
HAMMING WINDOW
SHORT TIME ENERGY CALCULATION (E)
SHORT TIME AVERAGE ZERO CROSSING RATE CALCULATION
IF ZCR IS SMALL AND ENERGY IS HIGH
VOICED SPEECH
UNVOICED SPEECH

Figure (3): Voice Activity Detection (VAD)

The basic steps of the VAD algorithm are the following [11]: The input speech signal is divided into frames. Energy and Zero-crossing rate (ZCR) for each frame is calculated. Threshold value for Energy and Zero-crossing rates are fixed and compared with Energy and ZCR of each frame. If the Zero-crossing rate is small and Energy is high then, it is declared as voiced speech otherwise it is termed as unvoiced speech. This voiced speech is given as input for feature extraction.

2.4 Extraction of MFCC

Mel-Frequency Cepstral Coefficients (MFCC's) is one of the most successful feature representations in Speaker Verification. The basic blocks for MFCC extraction are shown in Fig 2.3.

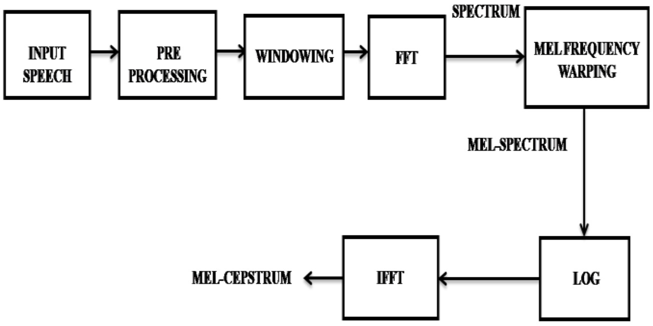

Figure (4): Feature Extraction using MFCC

The non–linear warping of the frequency axis can be modeled by the Mel-scale. The frequency
Groups are assumed to be linearly distributed along the Mel-scale.

$$f_{mel}(f) = 2595 log\left(1 + \frac{f}{700}\right) \qquad (1)$$

The Mel-frequency scale is linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. Finally, the feature vectors comprising of 13 Cepstral Coefficients are extracted.

2.5 Gaussian Mixture Models

Gaussian Mixture Models are widely used method for Speaker Modeling. The GMM parameters include mean vector (, covariance matrix (and mixture weight (w).They are trained using the Expectation Maximization (EM) algorithm [8].This algorithm uses an iterative Maximum Likelihood (ML) estimation technique.

A Gaussian Mixture Model is represented as the weighted sum of M component densities as given by the equation,

$$p(x/\lambda) = \sum_{i=1}^{M} w_i g \qquad (2)$$

where, is a D-dimensional continuous-valued data vector, - Mixture weights, $g(x/$ is a Component Gaussian density. The estimated parameters of GMM may exhibit uncertainty due to noisy data [5]. The uncertainties [3] explained in the previous section that cannot be handled by GMM are addressed by T2 FGMMs. This motivated the study towards Type-2 Fuzzy Gaussian Mixture Models (T2 FGMMs) explained in the next section.

2.6 Type-2 Fuzzy Gaussian Mixture Models

Zeng et al proposed an extension of GMM based on Type-2 Fuzzy Sets known as type-2 fuzzy GMM (T2 FGMM) [5].It can handle the uncertainties that cannot be addressed by GMM. They can model the uncertainties due to their three dimensional membership functions (MFs). The first dimension is used to describe the uncertainty of the observation whereas the second dimension is used to determine the uncertainty of the primary MFs.

Type-2 MFs are used to represent multivariate Gaussian with uncertain mean vector µ or covariance matrix ∑.These parameters are replaced by uncertain mean vector (T2 FGMM-UM) and uncertain covariance matrix (T2 FGMM- UV). Here, the mean and standard deviation and likelihood of observation are assumed to have a uniform distribution on a well- defined interval.

The process to train the T2 FGMM consists in estimating the parameters and and the factors and .

The factors and set the intervals in which the parameters vary:

$$\underline{\mu} = \mu - k_m \sigma, \qquad \overline{\mu} = \mu \qquad (4)$$

$$\underline{\sigma} = \qquad (5)$$

where, (0,3)and (0.3,1)represents the uncertainty factor for mean and covariance respectively.

Multivariate Gaussian having an uncertain mean vector can be defined as:

$$N(x;\ \tilde{\mu},\ \Sigma) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} exp\left[-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right] \dots exp\left[-\frac{1}{2}\left(\frac{x_d-\mu_d}{\sigma_d}\right)^2\right],$$
$$\mu_1 \in \left[\underline{\mu_1}, \overline{\mu_1}\right], \dots, \mu_d \in \left[\underline{\mu_d}, \overline{\mu_d}\right] \qquad (6)$$

Multivariate Gaussian with uncertain covariance matrix is given by:

$$= \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} exp\left[-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right] \dots exp\left[-\frac{1}{2}\left(\frac{x_d}{}\right.\right.$$
$$\sigma_1 \in \left[\underline{\sigma_1}, \overline{\sigma_1}\right], \dots \qquad (7)$$

where, represents the uncertain mean vector (UM) and covariance matrix (UV). The upper MF in the Gaussian function with uncertain mean is given by:

$$\overline{h}(x) = \begin{cases} f\left(x;\ \underline{\mu}, \sigma\right), x < \mu,\ 1, \underline{\mu} \leq \\ \qquad f(x;\ \overline{\mu}, \sigma), x > \overline{\mu} \end{cases} \qquad (8)$$

where,

$$f(x;\ \overline{\mu}, \sigma) \triangleq \qquad (9)$$

$$f\left(x;\ \underline{\mu}, \sigma\right) \triangleq \qquad (10)$$

The lower MF in the Gaussian function with uncertain mean is given by:

$$\underline{h}(x) = \begin{cases} f(x;\ \overline{\mu}, \sigma), \\ f\left(x;\ \underline{\mu}, \sigma\right), \end{cases} \qquad (11)$$

The upper MF in the Gaussian function with uncertain standard deviation is defined as:

(12)

The lower MF in the Gaussian function with uncertain standard deviation is defined as:

(13)

The length of the interval between two bounds of the log-likelihood interval is L= :

$$H(x) = |ln \qquad (14)$$

where and are the upper and lower membership functions of the GMM with uncertain parameters, respectively.

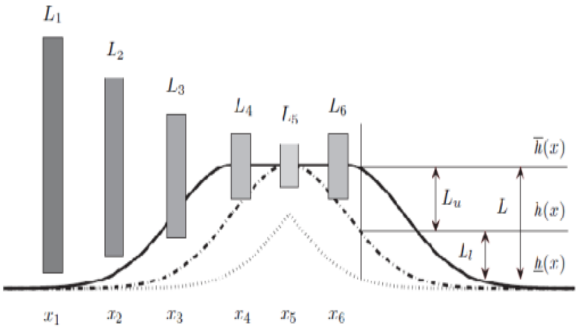

Figure (5): Log-likelihood intervals indicating the uncertainty in the model parameters

The length of the interval L describes the uncertainty of the class model to the input . The longer L the more the uncertainty. If km = kv= 0, then L = Ll = Lu = 0, which implies that there is no uncertainty so that the membership grade h ( ) is enough to make a classification decision. If k increases for a fixed deviation | − μ|, the length of the interval increases representing more uncertainty of the class model to the input x. Here, the Verification process is performed over the mean and variance intervals.

2.7 Speaker Verification using GLM

The Speaker Verification process is made using Generalized Linear Model (GLM) [4]. Here, the interval likelihoods obtained using T2 FGMMs are specified by the upper and lower bound values i.e. and respectively. Then weights are assigned to the upper and lower bound values. After that, the weighted upper and lower bounds are linearly combined to give the output for decision making. Finally for performing verification, a threshold value is fixed and score of the speaker is compared with the threshold. If the score value if greater than the threshold value, the speaker is accepted otherwise rejected.

III. EXPERIMENTS AND RESULTS

The proposed system for Speaker Verification uses the TIMIT (Texas Instruments Massachusetts Institute of Technology) database. It consists of speech files extracted from 30 female speakers and 71 male speakers. The speech data for each speaker includes 10 speech files, each of about 2-3 seconds duration. Here, the dialect region1 (dr1) speech signals from the TIMIT database was taken in which a single speaker is made to utter different sentences such as: “Don’t ask me to carry an oil rag like that” and “She had your dark suit in greasy wash water all year”. The Speaker Verification system is implemented in the MATLAB platform.

The initial step is pre-processing of the signal using Voice Activity Detection (VAD). This voice activity detection deals with removing the silence duration i.e., it removes the silence portion present at the beginning and end of adjacent samples. It determines the silence, voiced and unvoiced regions present in the speech signal. First, the input speech is divided into number of segments. Then, the energy and zero crossing rates for each segment are calculated. After estimating the energy and zero crossing rates, they are compared with a pre-defined threshold. If the segment has high energy and lesser number of zero crossings then it is labeled as a voiced portion. On the other hand, if the segment contains low energy with more number of zero crossings then it is labeled as an unvoiced portion. This VAD performed signal is given as input to feature extraction carried out by using MFCCs.

The next step after VAD is the feature extraction process. Mel- Frequency Cepstral Coefficients (MFCCs) are used to extract the feature from the speakers of the TIMIT database. Here, the first five female speakers of dialect region1 (dr1) is taken. Each speaker is made to utter eight sentences. Technique of computing MFCC is based on the short-term analysis, and thus from each frame a MFCC vector is computed. Finally, thirteen coefficients on Mel-scale were extracted.

After feature extraction process, the training data and testing data are modeled using GMM-UBM system.

(a) Training Phase: For training phase, the first five female speakers of dr1 are taken. Each speaker is made to utter eight sentences. GMM containing 10 mixtures of dimension 13 is used. Only the diagonal values are considered. Also the UBM model is developed using GMM adaptation. After that, the desired features i.e., the features that really establish the characteristics of the speaker are considered and modeled using GMM and T2 FGMM. For analysis, the results of GMM and T2 FGMM are compared with each other.

(b) Testing Phase: In testing phase, the first three female speakers are considered to be imposters. They are made to utter only two sentences of dr1. During testing, the scores are calculated for each speaker. The mean value of the score value is kept as the threshold. Now, the T-norm score values are combined and their average value is found out. This is given to the Detection Error Tradeoff (DET) function to get the performance curves. This curve gives the value of Equal Error Rate (EER in %) at which the false alarm probability (in %) and miss probability (in %) are almost equal.

TABLE I: PERFORMANCE EVALUATION OF GMM AND T2 FGMM IN TERMS OF EER

Model	Equal Error Rate (in %)
GMM	16.0
T2 FGMM-UM	14.2
T2 FGMM-UV	13.9

The above given results shows, that the proposed system for Speaker Verification, achieves a minimum Equal Error Rate (EER) than the existing GMM based speaker verification systems.

IV.CONCLUSION AND FUTURE DIRECTIONS

The performance of proposed speaker verification system based on Type-2 Fuzzy Gaussian Mixture Models (T2 FGMMs) is analyzed. GMM model is developed in order to provide a comparative analysis between GMM and T2 FGMM. Due to noisy data in real world problems, Speaker Verification systems are more subjected to uncertainties. These uncertainties in the system are directly modeled by T2 FGMM which uses Footprint Of Uncertainty (FOU) and interval secondary membership function.The mean and standard deviation values estimated using the GMM based speaker model are used in T2 FGMM to calculate the uncertain mean vector (UM) and uncertain covariance matrix (UV) by creating a Membership Function (MF) for them. The uncertainty parameter ‘k’ used in T2 FGMM set the intervals in which the parameters μ and ∑ vary. The verification process is made over the mean and variance intervals. The speaker acceptance/rejection decision is made by using Generalized Linear Model (GLM).The proposed method for Speaker Verification performs better than the existing techniques. Our future direction is to implement the proposed method under noisy conditions and use hybrid modeling techniques for further enhancement in verification process.

REFERENCES

[1] Reynolds D. A., Quatieri T. F. and Dunn R. B., “Speaker Verification using adapted Gaussian Mixture Models”, Digital Signal Processing, vol. 10, no. 1-3, pp. 19 – 41, 2000.

[2] Mendel J.M. and John R.I.B., “Type-2 Fuzzy Sets made simple”,IEEE Transactions on Fuzzy Systems, vol. 10, no. 2, pp 117–127,2002. 3, pp. 19 – 41, 2000.

[3] Zeng J. and Liu Z-Q., “Type-2 Fuzzy Sets for handling uncertainty in pattern recognition”, IEEE International Conference on Fuzzy Systems, pp. 6597–6602, 2006.

[4] Zeng J. and Liu Z-Q., “Type-2 Fuzzy Sets for pattern recognition: the state-of-the-art”, Journal of Uncertain Systems, vol. 1, no. 3, pp.163–177, 2007.

[5] Zeng J., Xie L., and Liu Z-Q., “Type-2 Fuzzy Gaussian Mixture Models”, Pattern Recognition, vol. 41, no. 12, pp. 3636–3643,2008.

[6] Zhaojie Ju ,Honghai Liu, “ Fuzzy Gaussian Mixture Models”, Pattern Recognition, vol. 45, pp. 1146–1158,sep 2011.

[7] Jerry M.Mendel., “Advances in Type-2 Fuzzy Sets and systems”, ScienceDirect, Information Sciences 177 (2007) 84–110.

[8] Zeng J. and Liu Z-Q., “Type-2 Fuzzy Markov Random Fields and their Application to Handwritten Chinese Character Recognition, IEEE Transactions On Fuzzy Systems, Vol. 16, No. 3, June 2008.

[9] Bachu R.G., Kopparthi S., Adapa B., Barkana B.D, “Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal”.

[10] L.R.Rabiner and M.R.Sambur, “An Algorithm for determining the end points of isolated utterances”, Bell system technical journal 1975.

[11] Tsang Ing Ren, Dimas Gabriel, Hector N. B. Pinheiro and George D. C. Cavalcanti, “Speaker Verification Using Type-2 Fuzzy Gaussian Mixture Models”, 2012 IEEE International Conference on Systems, Man, and Cybernetics.

[12] Hector N. B. Pinheiro, Tsang Ing Ren, George D. C. Cavalcanti, Tsang Ing Jyh and Jan Sijbers, “Type-2 fuzzy GMM-UBM for text-independent speaker verification, 2013 IEEE International Conference on Systems, Man, and Cybernetics.