International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 16 (2018) pp. 12798-12804

Speaker Recognition by Gaussian Filter Based Feature Extraction and Proposed Fuzzy Vector Quantization Modelling Technique

Satyanand Singh

Assistant Professor,
Department of Electrical and Electronics Engineering,
Fiji National University, Fiji, Island.

Orcid Id: 0000-0002-7707-031X

Abstract

Automatic speaker-recognition systems (ASR) have developed as a vital method for recognition of a person in many applications like e-commerce as well as in common interactions, law enforcement, and forensics. The performance of automatic speaker recognition system depends on the duration of the voice of test and train samples. In this paper, I address the issues in the design of codebook for outlier effect using the fuzzy vector quantization based algorithm for tasks such as speaker recognition for short voice samples. Particularly, I address the issues in partitioning the voice data for matching sequences of speaker specific feature vectors extracted from the voice signal data of utterances of a speaker. The design of codebook based on fuzzy vector quantization (FVQ) and fuzzy c-means (FCM) vector quantization has been proposed previously for matching the test and train voice signal characterized as sets of speaker-specific features vectors for tasks such as distortion measure and speaker recognition. Speaker-specific features Mel Frequency Cepstral Coefficients (MFCC) extracted by the traditional triangular filter as well as Gaussian and Tukey shapes filter. This research paper compares experimental results of three different modelling techniques namely, Fuzzy c-means, Fuzzy Vector Quantization2 (FVQ2) and proposed fuzzy vector quantization (FVQ). ASR efficiency of FVQ shows significant improvement in performance compared to FCM and FVQ2. The efficiency of proposed ASR system is 98.8 % for 2 seconds of training voice data for a set of 100 speakers taken from the Texas Instruments and Massachusetts Institute of Technology (TIMIT) database.

Keywords: Fuzzy vector quantization (FVQ), fuzzy c-means (FCM), Mel Frequency Cepstral Coefficients (MFCC), Fuzzy Vector Quantization2 (FVQ2).

INTRODUCTION

Distinguish an individual by his or her voice is a God gifted human quality most takes assure case in regular human-to-human association/ correspondence. Talking somebody via phone more often than not starts by distinguishing who is talking and, at the minimum in instances of recognizable speakers, a subjective confirmation by the audience that the person is right and the discussion can continue. ASR frameworks have developed as a vital method for authenticating individuality in several online applications and in addition all in all business collaborations, criminology, what's more, law enforcement. Aside from individual validation for access control, speaker recognition is a vital device in law implementation, national security, and legal sciences.

Extraordinarily, people routinely identify people by their voices with striking precision, particularly when the level of commonality with the subject is high. Normally, even a short nonlinguistic line, for example, a laugh quietly, is sufficient for us to recognize a well-known individual [1]. ASR research groups have produced different techniques pretty much autonomously for five decades. On the other hand, native speaker recognition is a regular capacity of people which is, on occasion, extremely exact and efficient. Late studies on mind imaging [2, 3] have uncovered numerous subtle elements on how I perform psychological based speaker recognition, which may motivate new headings for ASR approach.

To begin with, considering the general exploration area, it would be helpful to illuminate what is incorporated by the term speaker recognition, which comprises of two option undertakings: speaker identification and verification. In speaker recognition, the assignment is to distinguish an unknown speaker from the database of known speakers. There is two kind of ASR system open-set and close-set. If all speakers are known in a set then it is called as closed-set on the other hand, if the test speaker could likewise be from outside the predefined known speaker set, this turns into an open-set situation, and, along these lines, a world model or universal background model (UBM) [4] is required.

Speaker identification can be based on a voice stream that is content dependent or content free. This is more significant in speaker-verification systems in which a claimed person speaks a predefined text, like a password or personal identification number (PIN), to access system. All through this paper, the emphasis will be on text-independent automatic speaker recognition systems.

A few algorithmic and computational advances have empowered noteworthy ASR performance in the cutting edge. Approaches utilizing phonotactic data, phoneme recognizer followed by language models (PRLM) parallel PRLM, have been appeared to be very effective [5]. In this phonotactic demonstrating structure, an arrangement of tokenizes is utilized to interpret the speech information into token strings or cross sections which are later scored by n-gram dialect models [6] or mapped into a sack of trigrams highlight vector for support vector machine (SVM).

In fact traditional Hidden Markov model (HMM) based speaker verification is generally based on tokenizer class model, all tokenizations connected here to make a system [7], for example, Gaussian Mixture Model (GMM) tokenization [8], universal phone recognition (UPR) [9], articulator the property based methodology [10], deep neural network based telephone recognizer [11], just to give some examples. A late literature review LID is given in details by Li et al [7].

With the presentation of shifted delta-cepstral (SDC) acoustic components [12], promising results utilizing GMM system with the variable investigation [13, 14], super vector model [15] furthermore, maximum mutual information (MMI) based discriminative training [16] have likewise been accounted for LID. In this work, I concentrate on the acoustic level frameworks.

Acoustic-phonetic methodology, which is normally taken by specialists prepared on this, requires quantitative acoustic estimations from speech signal tests, and based on fact examination of the result. By and large, comparable phonetic units are removed from the known and addressed speech signal, and different acoustic parameters measured from these portions are evaluated. The LR can be helpfully utilized as a part of this methodology since it depends on numerical parameters [17].

In spite of the fact that the acoustic-phonetic methodology is a more target approach, it has some subjective components. For instance, an acoustic–phonetician may distinguish speech signal as being influenced by anxiety and after that perform the objective examination. In any case, whether the speaker was really under anxiety at that minute is a subjective amount controlled by the inspector through his or her experience. As on date, aggregate variability i-vector ASR modelling has achieved critical consideration in both LID and SV areas because of its remarkable efficiency, less system complexity and compact system in size. In i-vector modeling, initial, a solitary element investigation is utilized as a front end to produce a low dimensional aggregate variability space which together models dialect, speaker, and channel variability altogether. At that point, inside this i-vector space, variability costs techniques, for example, Within-Class Covariance Normalization(WCCN), Linear Discriminative examination (LDA) and Nuisance Attribute Projection (NAP) [15], are performed to diminish the variability for consequent modeling (e.g., utilizing SVM, logistic relapse (LR) and neural network for LID and probabilistic linear discriminate analysis (PLDA).

In this paper, I propose an FVQ algorithm for speaker recognition. In first step modification in fuzzy c-means objective function is done and later reformulation of the objective function. In a codebook design, analytical learning condition is analyzed to minimize the objective function. In a real-time, highly correlated data like voice signal fuzzy clustering is very natural rather a hard clustering, as items on the limits between a few classes are not compelled to completely have a place with one of the classes, but instead, are allotted enrollment degrees of membership of 0 and 1 showing their halfway participation. So the present study was undertaken with the objective of to find out the speaker recognition efficiency improving components with the help of different novel algorithms.

FUZZY CLUSTERING OF REAL TIME VOICE DATA USING VECTOR QUANTIZATION

In an automatic speaker recognition system, VQ is more concerned with the demonstration of a set of unlabeled voice data vectors $X = \{x_1,\ x_2\ \ldots.,\ x_n\}\ \in \mathbb{R}^p$ by a set $V = \{v_1,\ v_2 \ldots.,\ v_c\} \in\ \mathbb{R}^p$ by means of $c \ll n$. Every x_k $(1 \leq k\ \leq\ n)$ is represented as training data vector and the set $\mathbf{X}$ is passed on as training set data, while each v_i $(1\ \leq\ i\ \leq\ c)$ is codebook data vector and the set $\boldsymbol{V}$ is a codebook of real time voice data. Every codebook data vector related to a cluster, and called as cluster center [18]. The important subject in VQ is the designing of codebook. The codebook can be outlined by crisp or fuzzy decision-making measures.

The codebook can be evaluated by average distortion measures and represented as [19]:

$$D = \frac{1}{n}\sum_{k=1}^{n}\ \min_{1\leq i\leq c}\{\|x_k - v_i\|^2\} \tag{1}$$

Fuzzy clustering based on VQ algorithms is well described in the next section, which are known as FCM and the FVQ2 developed by Karayiannis and Pai in [20].

Clustering of voice data based on Fuzzy c-Means

In ASR system FCM algorithm is most commonly used to cluster the voice data. $\mathrm{u_{i\,k}} = \{\mathrm{u_i(x_k)}, 1 \leq \mathrm{i} \leq \mathrm{c}, 1 \leq \mathrm{k} \leq \mathrm{n}\}$ is represented as membership degree of the k-th training vector to the i-th cluster. The fuzzy-c clusters constitute a constrained in X if three conditions are satisfied,

$$\begin{gathered} 0\ \leq\ u_{i\,k}\ \leq 1\,,\quad \forall\, i,k \\ 0 < \sum_{k=1}^{n} u_{i\,k} < n, \quad\ \forall\ i \\ \sum_{\mathrm{i}=1}^{\mathrm{c}} \mathrm{u_{i\,k}} = 1, \qquad\quad \forall \mathrm{k} \end{gathered} \tag{2}$$

If $\sum_{\mathrm{i}=1}^{\mathrm{c}} \mathrm{u_{i\,k}} = 1,\ \forall \mathrm{k}$ is not satisfied then it is called as unconstrained fuzzy c partition. The realization of the fuzzy c-means for real time voice data is based on the minimization of objective function represented as J_m, under the equality constraint given in eqn. (2). The objective function is mathematically defined as,

$$J_m = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{i\,k})^m\|x_k - v_i\|^2 \tag{3}$$

Real time voice data cluster centers and the membership degrees that solve the above constrained is represented by the following eqn. [21],

$$v_i = \frac{\sum_{k=1}^{n}(u_{i\,k})^m \ x_k}{\sum_{k=1}^{n}(u_{i\,k})^m} , \qquad 1 \le i \le c \qquad (4)$$

And

$$u_{i,k} = \frac{1}{\sum_{j=1}^{c}\left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|}\right)^{\frac{2}{m-1}}} , 1 \le i \le c , 1 \le k \le n \qquad (5)$$

The fuzziness in eqn. (4) and (5) is controlled by the parameter m with iterative optimization procedure. When $m \to 1$ the partition is called crisp and other hand, if m takes large values then the membership degrees of each training vector approaches 1/c. In this case, slowly all the cluster centers will be pulled towards the training vectors.

The main objective of enhancing the ASR efficiency can be met by minimizing the J_m. The objective function for 64 clusters for different iteration is represented in table.1. The objective function J_m= 0.30 with minimum distortion D = **4.6669** is achieved in 35 iterations.

Table 1: Minimum Objective Function of fuzzy Clustering

Number of Iterations	Objective Function
1	1.30
5	0.45
10	0.35
15	0.30
20	0.30
25	0.30
30	0.30
35	0.30

Distortion of the speakers using real time voice data in training the ASR system by fuzzy c-means clustering is as follows: D = 6.4711, 6.0261, 12.6531, 5.6056, 10.0395, 5.968, 8.0071, 5.6889, 13.0149, 6.0971, 5.6337, 5.8435, 5.5733, 5.5329, 6.5637, 13.8927, 5.6554, 4.9273, 6.6624, **4.6669**, 5.8925, 14.5138, 8.4431, 9.9943, 5.7945, 10.0552, 6.195, 6.6755, 5.192, 7.1233, 6.7116, 12.066, 5.209, 5.4083, 6.2515. Lowest distortion measure is highlighted in bold.

Clustering of voice data based on Fuzzy Vector Quantization2

Application of VQ in ASR system is implemented by mapping each training vector of voice data to single codebook, which is maximizing degree of membership of training vector. It has a serious effect on codebook design and hence the ASR efficiency because this approach hides the existence of outliers and replaces them by their closest codebook vectors. To overcome of outliers effect, the set T_k is updated as in the ν-th iteration the set $T_k^{(v)}$ contains $\aleph T_k^{(v)}$ codebook vectors. Defining the average distance $\check{d}_k^{(v)}$ as follows,

$$\check{d}_k^{(v)} = \frac{1}{\aleph(T_k^{(v)})} \sum_{v_i \in T_K^{(v)}} \|x_k - v_i\|^2 \qquad (6)$$

T_k is obtained in the $(v+1)th$ iteration as follows.

$$T_k^{(v+1)} = \left\{v_i \in T_k^{(v)} : \|x_k - v_i\|^2 \le \check{d}_k^{(v)}\right\} \qquad (7)$$

Each training vector initially is assigned to all of the codebook vectors to remove outliers effect. As this development proceeds, the cardinality $\aleph(T_k^{(v)})$ of the set T_k decreases, until T_k will include only one element. I choose to use the FVQ2 algorithm, because only this algorithm is directly related to the constrained minimization of the objective function given in eqn. (3). The implementation of the FVQ2 requires that the codebook vectors are updated by using eq. (4). Moreover, in the v-th iteration, if the training vector x_k is in fuzzy mode its membership degrees are calculated as,

$$u_{i\,k} = \frac{1}{\sum_{v_j \in T_k^{(v)}}\left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|}\right)^{\frac{2}{m-1}}} , with \ \ v_i \in T_k^{(v)} \qquad (8)$$

while if x_k is in crisp mode then the membership degrees are given by the next nearest neighbor condition,

$$u_{i\,k} = \begin{cases} 1 , if \ \|x_k - v_i\|^2 = \min_{1\le i\le c}\{\|x_k - v_i\|^2\} \\ 0, otherwise \end{cases} \qquad (9)$$

FVQ2 clustering minimum objective function achieved with distortion $\check{d}_k^{(v)}$=0.1887. Table 2 shows the objective function and number of iteration by FVQ2.

Table 2: Minimum Objective Function of FVQ Clustering

Number of Iterations	Objective Function
1	**1.25**
5	**0.40**
10	**0.30**
15	**0.25**
20	**0.25**
25	**0.25**
30	**0.25**
35	**0.25**

Distortion of the speakers using FVQ2is as follows:

D =6.3721, 5.8452, 13.1923, 5.6251, 10.963, 6.3408, 7.938, 5.582, 13.319, 6.163, 5.296, 5.785, 5.412, 5.264, 6.011, 14.384, 5.815, 5.131, 7.005, **4.492**, 5.362, 14.482, 9.007, 10.723, 5.671, 10.07, 6.0151, 6.7238, 5.2056, 6.6534, 7.0366, 12.4455, 5.1927, 5.3277, 6.1519. Lowest distortion measure is highlighted in bold.

Clustering of real time voice Data using proposed Fuzzy Vector Quantization

In this section Iexhibit a step by step investigation of the proposed fuzzy vector quantization algorithm and its application in ASR system. The objective function is represented by J_p and defined as follows,

$$J_p = \sum_{k=1}^{n}\sum_{i=1}^{c} f(u_{ik})\|x_k - v_i\|^2 \qquad (10)$$

With

$$f(u_{ik}) = \frac{1}{2}u_{ik} + \frac{1}{2}(u_{ik})^2 \qquad (11)$$

Where the degree of membership of the k-th training vector to the i-th codebook vector is represented as u_{ik}. The main aim is to minimize the above objective function J_p under the following equality constraint,

$$\sum_{i=1}^{c} u_{ik} = 1, \quad \forall\, k \qquad (12)$$

The degree of membership of the codebook vector that resolve the above objective function minimization difficulty are given by the following theorems,

Theorem 1

If $v_i(1 \le i \le c)$ are fixed then the $u_{ik}(1 \le i \le c : 1 \le k \le n)$ that minimize J_p in eqn. (10), under the constraint in eqn. (12), are given by the next eqn.,

$$u_{ik} = \frac{c+2}{2} . \frac{1}{\sum_{j=1}^{c}\left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|}\right)^2} - \frac{1}{2} \qquad (13)$$

Theorem 2

If $u_{ik}(1 \le i \le c; 1 \le k \le n)$ are fixed, then $v_i(1 \le i \le c)$ that minimize J_p in enq. (11) is given by the next eqn.,

$$v_i = \frac{\sum_{k=1}^{n} f(u_{ik})x_k}{\sum_{k=1}^{n} f(u_{ik})} \qquad (14)$$

Substituting eqn. (13) into the objective function in eqn. (10) Ican easily obtain the following reformulating function,

$$R_J = R_{J1} + R_{J2} \qquad (15)$$

$$R_{J1} = \frac{2+c}{4}\sum_{k=1}^{n}\sum_{i=1}^{c}\left\{\|x_k - v_i\|^2\left[\sum_{j=1}^{c}\frac{\|x_k - v_i\|^2}{\|x_k - v_j\|^2}\right]^{-1}\right\} - \frac{1}{4}\sum_{k=1}^{n}\sum_{i=1}^{c}\|x_k - v_i\|^2 \Rightarrow$$

$$R_{J1} = c\frac{2+c}{4}\sum_{k=1}^{n}\left[\sum_{j=1}^{c}\frac{1}{\|x_k - v_j\|^2}\right]^{-1} - \frac{1}{4}\sum_{k=1}^{n}\sum_{i=1}^{c}\|x_k - v_i\|^2 \qquad (16)$$

Relationally,

$$R_{J2} = \frac{4-c^2}{8}\sum_{k=1}^{n}\left[\sum_{i=1}^{c}\frac{1}{\|x_k - v_i\|^2}\right]^{-1} + \frac{1}{8}\sum_{k=1}^{n}\sum_{i=1}^{c}\|x_k - v_i\|^2 \qquad (17)$$

Substituting eqn. (15) and (16) into eqn. (17), the reformulating function is novel as follows,

$$R_J = K_1\sum_{k=1}^{n}\left[\sum_{i=1}^{c}\frac{1}{\|x_k - v_i\|^2}\right]^{-1} - K_2\sum_{k=1}^{n}\sum_{i=1}^{c}\|x_k - v_i\|^2 \qquad (18)$$

Where

$$K_1 = \frac{(2+c)^2}{8} \;\; and \;\; K_2 = \frac{1}{8} \qquad (19)$$

By minimizing the reformulating function in eqn. (17) with respect to the codebook vectors, the gradient-descent based learning rule for the i-th codebook vector is given as,

$$v_i(t+1) = v_i(t) - a(t)\sum_{k=1}^{n} f\big(u_{ik}(t)\big)x_k \qquad (20)$$

Where $f(u_{ik})$ is given in eqn. (11), and $a(t)$ is the learning rate parameter, which can be calculated as follows,

$$a(t) = a_0\left(1 - \frac{t}{t_{max}}\right) \qquad (21)$$

Where a_0 is the initial value for the learning parameter, and t_{max} is the maximum number of iteration. Based on the above analysis, the proposed fuzzy learning vector quantization algorithm for speaker recognition given as follows,

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 16 (2018) pp. 12798-12804

The proposed vector quantization for speaker recognition

(i) Select initial values for $v_i (1 \le i \le c)$.

Set values for the design parameters $t_{max} = 35$ and $a_{0=}0.5$.

For $t = 1$ to t_{max} .

(ii) Using eqn. (13) to calculate $u_{ik}(1 \le i \le c, 1 \le k \le n)$

(iii) Using eqn. (11) calculate $f(u_{ik})(1 \le i \le c, 1 \le k \le n)$.

For $i = 1$ to c

(iv) using eqn. (21) to update the codebook vectors.

Endfor

Endfor

End

The proposed FVQ clustering and its minimum objective function J_p=0.07 with distortion D =4.334. Table 3 shows the objective function for the proposed Fuzzy vector quantization.

Table 3: Minimum Objective Function of proposed FVQ

Number of Iterations	Objective Function
1	**0.62**
5	**0.25**
10	**0.12**
15	**0.07**
20	**0.07**
25	**0.07**
30	**0.07**
35	**0.07**

Distortion of the speakers using proposed FVQ is as follows:

D = 6.1162, 5.2875, 8.4223, 4.5901, 10.5284, 6.4421, 7.6078, 5.1621, 11.9275, 6.1786, 4.8972, 5.3811, 4.8904, 5.2311, 5.3065, 13.3695, 5.3833, 5.3433, 7.2431, **4.334**, 4.7034, 5.8735, 8.7089, 10.0453, 5.4298, 9.2664, 5.3304, 6.1129, 5.2765, 6.1359, 6.4721, 11.205, 5.0118, 5.2669, 5.4778. Lowest distortion measure is highlighted in bold. In the following section I present the experimental results.

EXPERIMENTAL RESULTS

The proposed fuzzy vector quantization algorithm based ASR is implemented and its efficiency is evaluated with parameters t_{max} = 35, a_o= 0.5 and codebook size = 256. The experiment uses the TIMIT set of database. The proposed algorithm implemented in MATLAB and results were compared with those of the Fuzzy c-Means, FVQ2 algorithms and MFCC Phase Information [22, 23]. A total 1000 utterances of the TIMIT database of 2 sec voice were put to test for 100 speakers. For the above cases, ASR recognition efficiency has been calculated “Efficiency” = Number of utterance correctly identified/Total Number of utterance under test. Table 4 shows that the efficiency of the ASR system for Triangular, Gaussian and Tukey [24] based filters and Fuzzy c-means, FVQ2 and proposed FVQ techniques respectively. It can be observed from this table that use of Gaussian filter and proposed FVQ show significant improvement.

Table 4: Efficiency of Speaker Recognition system

Filters	Accuracy		
	Fuzzy c-Means	**FVQ 2**	**Proposed FVQ**
Triangular Filter	96.9%	97.0%	**97.2%**
Gaussian Filter	98.1%	98.3%	**98.8%**
Tukey Filter	97.3%	97.5%	**97.9%**

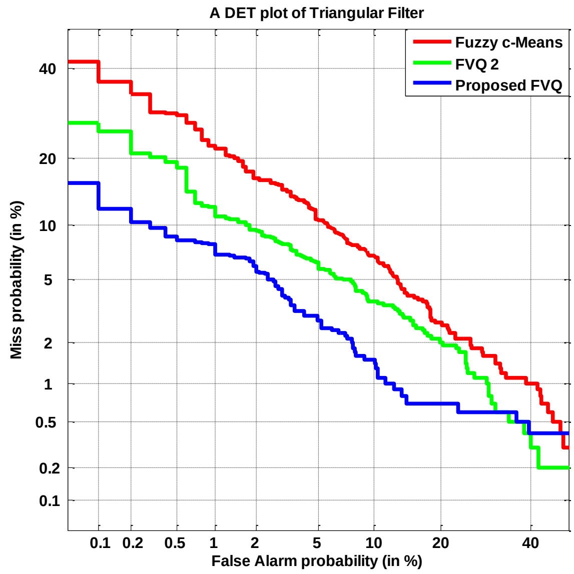

Figure 1: Equal Error Rate of Triangular Filter with Fuzzy c-means,FVQ2 and Proposed FVQ Modelling techniques.

It can be observed from the above table that the proposed FVQ algorithm associated with the use of Gaussian filter shows significant improvement up to 98.8%.

Fig. 1shows the Triangular filter based modeling technique equal error rate (EER). The triangular filter based technique ASR efficiency for fuzzy c-means, FVQ2, and proposed FVQ are 96.9%, 97.0%, and 97.2% respectively. The proposed FVQ EER with triangular filter is 3.8%.

Fig. 2shows the Gaussian filter based modeling technique equal error rate (EER). The Gaussian filter based technique ASR efficiency for fuzzy c-means, FVQ2, and proposed FVQ are 98.1%, 98.3%, and 98.8% respectively. The proposed FVQ EER with a Gaussian filter is 2.7%.

Fig. 3 shows the Tukey filter based modeling technique equal error rate (EER). The Tukey filter based technique ASR efficiency for fuzzy c-means, FVQ2 and proposed FVQ are

97.3%, 97.5%, and 97.9% respectively. The proposed FVQ EER with Tukey filter is 3.5%.

The EER for proposed FVQ modeling technique ASR system on TIMIT database improvement is 1.1% and 0.8% respectively. The efficiency of the ASR system with the combination of Gaussian and proposed FVQ shows significant improvement of 1.6% and 0.9% compared to Triangular and Tukey filter respectively.

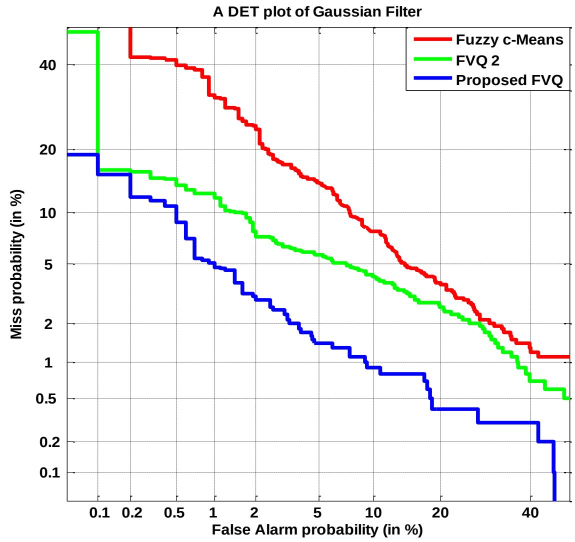

Figure 2: Equal Error Rate of Gaussian Filter with Fuzzy c-means,FVQ2 and Proposed FVQ Modeling techniques.

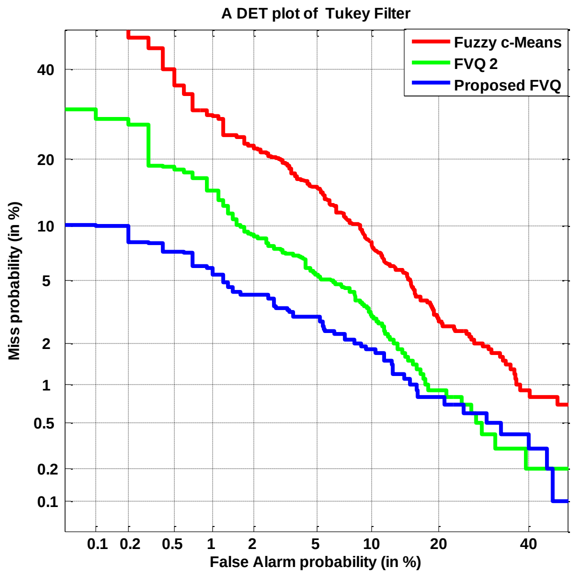

Figure3: Equal Error Rate of Tukey Filter with Fuzzy c-means FVQ2 and Proposed FVQ Modeling techniques.

CONCLUSION

In this paper, I proposed the FVQ algorithm to overcome the outlier effect in clustering for ASR system. The speaker recognition experiments were conducted on the TIMIT database which consists of 630 speakers. The proposed algorithm was designed to capture the advantages provided by fuzzy decision-making processes while maintaining the computational capabilities achieved by crisp decision-making processes. This was accomplished by developing and reformulating objective function for the well known fuzzy c-means clustering technique. Several simulations were performed by MATLAB, in which the proposed FVQ algorithm was compared to other techniques found in the literature [21]. The objective function is minimized and distortion of the proposed FVQ approach is reduced when compared with the objective function and distortion of Fuzzy c-means, and FVQ2. The result of this comparison shows that the proposed FVQ algorithm can be used as a reliable tool in speaker recognition applications. In our future research work, I intend to investigate the performance of the ASR system on real-time voice data for mobile application.

REFERENCES

[1] S. Singh, “Forensic and Automatic Speaker Recognition System,” International Journal of Electrical and Computer Engineering, vol. 8, no 5, 2018.

[2] P. Belin, R. J. Zatorre, P. Lafaille, P. Ahad, and B. Pike, ‘Voice-selective areas in human auditory cortex,” Nature, Vol. 403, 2000, pp. 309–312.

[3] E. Formisano, F. De Martino, M. Bonte, and R. Goebel, ‘Who’ is saying ‘what’? Brainbased decoding of human voice and speech,” Science, vol. 322, 2008, pp. 970–973.

[4] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, “Speaker verification usingvadapted Gaussian mixture models,” Digital Signal Process., Vol. 10, no. 1–3, 2000, pp.19–41.

[5] S.Singh, “Support Vector Machine Based Approaches For Real Time Automatic Speaker Recognition System,” International Journal of Applied Engineering Research,vol. 13, no. 10, pp. 8561-8567, 2018.

[6] Gauvain, J., Messaoudi, A., Schwenk, H., “ Language recognition using phone lattices,” in: Proc. ICSLP, 2004.

[7] Li, H., Ma, B., Lee, K.A., “Spoken language recognition: From fundamentals to practice,” Proceedings of the IEEE 101, 2013, pp. 1136-1159.

[8] Torres-Carrasquillo, P., Singer, E., Kohler, M., Greene, R., Reynolds, D., Deller Jr, J., “Approaches to language identification using gaussian mixture models and shifted delta cepstral features,” in: Proc. ICSLP, 2002, pp. 89-92.

[9] Siniscalchi, S.M., Reed, J., Svendsen, T., Lee, C.H., “Exploiting context-dependency and acoustic resolution of universal speech attribute models in spoken language recognition,” in: Proc. INTERSPEECH, 2010, pp. 2718-2721.

[10] Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., et al., "Deep neural networks for acoustic modeling in speech recognition," The shared views of four research groups. IEEE Signal Processing Magazine 29, 2012, pp. 82-97.

[11] Deng, L., Li, X., "Machine learning paradigms for speech recognition: An overview," IEEE Transactions on Audio, Speech, and Language Processing 21, 2013, pp. 1060-1089.

[12] Torres-Carrasquillo, P., Singer, E., Kohler, M., Greene, R., Reynolds, D., Deller Jr, J., "Approaches to language identification using gaussian mixture models and shifted delta cepstral features," In: Proc. ICSLP, 2002, pp. 89-92.

[13] Kenny, P., Ouellet, P., Dehak, N., Gupta, V., Dumouchel, P., "A study of interspeaker variability in speaker verification," IEEE Transactions on Audio, Speech, and Language Processing 16, 2008, pp.980-988.

[14] Campbell,W., Sturim, D., Reynolds, D., "Support vector machines using gmm supervectors for speaker verification," IEEE Signal Processing Letters 13, 2006, pp. 308-311.

[15] Burget, L., Matejka, P., Cernocky, J., "Discriminative training techniques for acoustic language identification," in: Proc. ICASSP, 2006.

[16] S.Singh, Mansour H. Assaf, Sunil R.Das, Emil M. Petriu, and Voicu Groza, "Short Duration Voice Data Speaker Recognition System Using Novel Fuzzy Vector Quantization Algorithms", IEEE International Instrumentation and Measurement Technology Conference, May23-26, 2016.

[17] Martinez, D., Plchot, O., Burget, L., Glembek, O., Matejka, P., "Language recognition in ivectors space," in: Proc. INTERSPEECH, 2011, pp. 861-864.

[18] S. Singh, Abhay Kumar, David Raju Kolluri, "Efficient Modelling Technique based Speaker Recognition under Limited Speech Data", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol, 8, No.11, 2016, pp.41-48.

[19] S.Singh and Ajeet Singh "Accuracy comparison using different modeling techniques under limited speech data of speaker recognition systems" Global Journal of Science Frontier Research: F Mathematics and Decision Sciences, Vol. 16, 2016, pp.1-17.

[20] N. B. Karayiannis, P.-I. Pai, "Fuzzy Vector Quantization Algorithms and Their Application in Image Compression," IEEE trans. Vol. 36(8), 2008, pp. 957-971

[21] Jayanna, H.S., Prasanna, S.R. Mahadeva., "Fuzzy Vector Quantization for Speaker Recognition under Limited Data Conditions," IEEE Region 10 Conference, TENCON, Hyderabad India, Nov 2008, pp.1-4.

[22] Seiichi Nakagawa, "Speaker Identification and Verification by Combining MFCC and Phase Information," IEEE transaction s on audio , speech and language processing, Vol. 20, 2012, pp. 1085-1095.

[23] S.Singh and E.G. Rajan "Vector quantization approach for speaker recognition using MFCC and Inverted MFCC" International Journal of Computer Application, Vol. 17, No 1,2011, pp. 1-7.

[24] S.Singh and Dr. E.G. Rajan "Application of different filters in mel frequency cepstral coefficients feature extraction and fuzzy vector quantization qpproach in speaker recognition" International Journal of Engineering Research & Technology, vol. 2, pp.213-220, 2013.