Efficient Modelling Technique based Speaker Recognition under Limited Speech Data

Research Proposal *in* International Journal of Image, Graphics and Signal Processing · November 2016

DOI: 10.5815/ijigsp.2016.11.06

3 authors, including:

Satya Nand
St. Peter's Engineering College
5 PUBLICATIONS **0** CITATIONS

SEE PROFILE

I.J. Image, Graphics and Signal Processing, **2016, 11, 41-48**
Published Online November 2016 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2016.11.06

Efficient Modelling Technique based Speaker Recognition under Limited Speech Data

Satyanand Singh
CMRIT, Department of ECE, Secunderabad, 500010, India
Email: yogitechno@gmail.com

Abhay Kumar and David Raju Kolluri
Research Scholar, SSSUTMS, Department of CSE, Sehore, 466001, India
Associate Professor, St.Peter's Engineering College, Department of CSE, Secunderabad, 500014, India
Email: {abhay1880, kolluridavid}@gmail.com

***Abstract*—**As on date, Speaker-specific feature extraction and modelling techniques has been designed in automatic speaker recognition (ASR) for a sufficient amount of speech data. Once the speech data is limited the ASR performance degraded drastically. ASR system for limited speech data is always a highly challenging task due to a short utterance. The main goal of ASR to form a judgment for an incoming speaker to the system as being which member of registered speakers. This paper presents a comparison of three different modelling techniques of speaker specific extracted information (i) Fuzzy c-means (FCM) (ii) Fuzzy Vector Quantization2 (FVQ2) and (iii) Novel Fuzzy Vector Quantization (NFVQ). Using these three modelling techniques, we developed a text independent automatic speaker recognition system that is computationally modest and equipped for recognizing a non-cooperative speaker. In this investigation, the speaker recognition efficiency is compared to less than 2 sec of text-independent test and train utterances of Texas Instruments and Massachusetts Institute of Technology (TIMIT) and self-collected database. The efficiency of ASR has been improved by 1% with the baseline by hiding the outliers and assigns them by their closest codebook vectors the efficiency of proposed modelling techniques is 98.8%, 98.1% respectively for TIMIT and self-collected database.

***Index Terms*—**Vector Quantization, Fuzzy c-means Vector Quantization, Fuzzy Vector Quantization2, Novel Fuzzy Vector Quantization, Objective Function.

I. INTRODUCTION

ASR is a smart machine, which is capable of recognizing a person from their speech [1].To pick out speakers from a database, ASR system needs adequate speech data with the goal that it can separate speaker well and consequently yields reliable recognition [2].ASR in a limited speech data condition, the main goal to recognise a speaker when both training and testing data are short. Since the measure of speech data of a speaker who is non-cooperative is little in the restricted information conditions, the machine is able to acquire less number of feature vectors which are not sufficient to model and discriminate the speaker [3]. Past four decades of research, the current ASR system have reached a milestone with an appreciable performance provided test/train utterance are more than enough and the signal-to-noise ratio (SNR) is large enough.

Literature survey in line with limited speech data based speaker recognition system; all authors have centered of attention on discriminative training/scoring [4], clustering [5], and neighborhood information [6] among the close-set speakers.

In this research, the issue of close-set ASR is concentrated on with the requirements of low speech data enrolments about 2-3 sec of the size of 100 speakers in training and testing. At the time of registration of speakers, if the speech signal for train /test is in few sec, then the phonemes of the spoken text will be fragmented and bring the acoustic gaps situation in train model space of speaker. At the point when a test token for an enlisted speaker contains phonetic substance that was not seen amid training of ASR system, it results in a low probability score, and potentially a wrong choice for that speaker (i.e., phonemes found during testing of ASR but not during training data for the same speaker cause the speaker model to be rejected) [7].

As on date, some efforts have been made to distinguish the person under inadequate speech signal circumstances utilizing the idea of Universal Background Model (UBM) to moderate the inadequacy, which requires extra speech data to train the GMM-UBM ASR system model [4].S. Kwon and S. Narayanan has made an attempt that by selecting just the component vectors which are segregating the speakers it is conceivable to distinguish speaker under limited speech data condition [8]. As we specified before there are just a few modelling techniques in case of limited speech data that most straightforward feasible like FCM, FVQ2 and NFVQ might be utilized for demonstrating the speakers. These modelling techniques are straightforward and less computational complex.

A. Challenges with limited speech data in ASR

Regardless of the immense success, the state of art ASR performs well only if the training and testing utterances are sufficient in data size. In several ASR, the clients are hesitant to provide sufficient voice data, especially for testing, in phone banking. In different circumstances, it is profoundly hard to gather adequate speech data, for instance in legal applications [9].

The recent research advocate if the speech data utilised during testing phase bring down 10% (from 20 sec of speech data to 2 sec of speech data) the performance of ASR degraded abruptly from 6.34% to 23.89% in terms of equal error rate (EER) [9]. In ASR application once testing speech data is less than 2 sec the performance of the system in terms of EER 35% has been reported by Mak et al. [10].

II. Related Work

As on date, research on limited speech data based ASR system has been reported. The joint factor analysis (JFA) can enhance the performance of ASR in the case of limited speech data by diverting the variability in different subspace [11]. The comparison of ASR performance for limited speech data based on JFA and i-vector modelling has been reported by various compensating techniques [12]. What's more, a score-based portion determination system has been proposed in [9], which estimates the superiority of every test speech portion taking into account an arrangement of companion models, and scores the test speech with the dependable fragments as it were. A relative EER attenuation of 22% was reported in ASR when the test speech data are shorter than 15 sec.

We have evaluated the efficiency of various modelling technique, computational complexity and its performance enhancing parameters. For example, a case of FCM the feature vectors are clustered into no overlapping fashion while in the case of FVQ2 the feature vectors are clustered in an overlapping fashion. Consequently, the principle of clustering is different and it is not possible to hide the existence of outliers in FCM and FVQ2, but NFVQ modelling technique replaces them by their closest codebook vectors.

The design of codebook performed by FCM and FVQ2 modelling by employing only crisp decision-making procedures [13], the intelligence is every training vector is allocated to only one cluster. Ultimately, FCM and FVQ2 techniques do not take into account the possibility that a particular training vector may also belong to another cluster. NFVQ efficiently examination and resolves complex, ill-defined and less logical system which were unable to resolve by FCM and FVQ2 [14]. Fuzzy logic is a critical thinking control framework strategy that fits execution in frameworks extending from basic, little, implanted small scale controllers to expansive, organized, multi-channel PCs or workstation-based information obtaining and control frameworks. They can be executed by equipment, programming, or a mix of both [15]. Fuzzy logic was initially considered as a superior strategy for sorting and taking care of information, however, has later turned out to be a super decision for some control framework applications where the speaker was non-cooperative. Fuzzy logic based speaker recognition system can be incorporated with anything from small hand-held items to extensive mechanized procedure control frameworks. In the course of the initial clusters, centre utilizes the clustering strategy about the comparability threshold and its minimum distance theory [16]. To start with, group the Vector samples generally with the function of the initial cluster centre. This strategy is with a specific end goal to beat the initialization vector which is very sensitive of FCM clustering and ensures access to the general cluster result. This method is in order to overcome the initialization vector sensitive shortcomings of Fuzzy C-means clustering and guarantee access to the overall cluster result [17].

III. The Application of Fuzzy Clustering for VQ

In ASR, VQ is fretful to demonstration on set of data vectors $X=\{X1,, X2 ,.....Xn\}$ which is unlabeled in nature and $\in \mathbb{R}^p$ with a set $V=\{V1,, V2 ,.....Vn\}$ with c<<n and $\in \mathbb{R}^p$ Here X_k = training vector, X=training set, every V_i =codebook vector and the set V= codebook. The key concern in VQ is to outline the codebook. It can be composed through utilizing crisp or hard options in the systems. In crisp or hard cases clustering the normal distortions indicated by D,

$$D = \frac{1}{n}\sum_{k=1}^{n} \min_{1\le i\le c}\left\{ \| x_k - v_i \|^2 \right\} \quad (1)$$

The proposed modelling technique has been compared with FCM and FVQ developed by Karayiannis and Pai in [18].

A. Modelling with FCM Algorithm

In FCM modelling technique every data point is associated with all clusters with different degree of membership. The sum of all degree of membership of a data is equal to one. The main objective of FCM clustering to minimize the J_m to improve the ASR performance and hence EER.

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

The all variables in eqn. (2) is defined as:

n = Total number of considered data points.
c = Total number of clusters in modelling technique.
m = Fuzzy controlling parameters.
X_k = k^{th} data point.
$\boldsymbol{v_i}$ = Centre of the i^{th} cluster.
$u_{i,k}$ =Degree of membership in the i^{th} cluster.

FCM performs the accompanying strides amid clustering:

i. First start the cluster membership values, $u_{i,k}$
ii. compute the cluster centers

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

iii. Update $u_{i,k}$ according to

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

iv. Compute ***Jm***.
v. Do again step ii–iv until ***Jm*** enhances by not exactly a predefined threshold or number of Iterations.

FCM clustering organizes every data point into the cluster with the highest degree of membership.

The number of iterations is controlled by clustering termination condition in such a way that the improvements need to maintain a certain value when both of the condition satisfied:

- Maximum iteration achieves a most extreme of 50.
- ***Jm*** development must be less than 0.00001 in successive iteration.

FCM Objective Function ***Jm*** = 3.24288, 2.38009, 2.08069, 1.53718, 1.19842, 1.07426, 0.98673, 0.92191, 0.88394, 0.86302, 0.85027, 0.84238, 0.83758, 0.83449, 0.83218, 0.82999, 0.82740, 0.82380, 0.81861, 0.81153, 0.80300, 0.79438, 0.78711, 0.78184, 0.77837, 0.77619, 0.77482, 0.77396, 0.77341, 0.77304, 0.77280, 0.77264, 0.77252, 0.77244, 0.77238, 0.77233, 0.77230, 0.77227, 0.77225, 0.77224, 0.77223, , 0.77222.

In FCM clustering the objective function determines the clustering terminations. Here the minimum objective function achieved 0.77222 after 42 iterations.

The minimum value of objective function obtained with FCM modelling technique is ***Jm*** = 0.77222. Fig. 1 demonstrates the plot of objective function by FCM modelling technique.

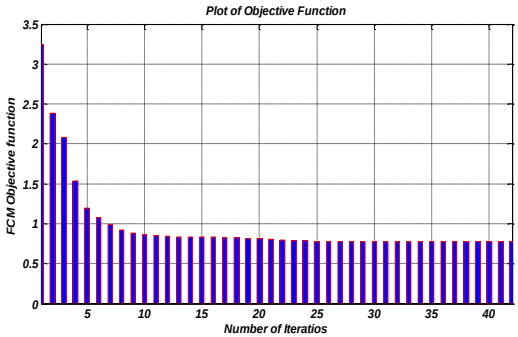

Fig.1. Plot of FCM modelling technique objective function.

At the point when the objective function ***Jm*** is minimized, resulting the number of iterations don't bring about any development of vectors between clusters and its boundaries limits get to be settled. This could be utilized as one of conceivable pointers to terminate the FCM computation. A remarkable point of FCM algorithm is its computational straightforwardness. A case of FCM clustering is shown in Fig. 2.

On account of speaker recognition the speech signal are pre-processed and a set of speaker specific feature vectors computed. There is no broad hypothetical answer to find the optimum number of clusters in FCM for any given data set. A basic methodology is to contrast the consequences of various runs and distinctive quantities of classes and pick the best one as indicated by a given paradigm. The disadvantage of over fitting can be to a great extent dispensed with by utilizing the Information theoretic construct VQ which works in light of the guideline of physical clarification of the data clusters.

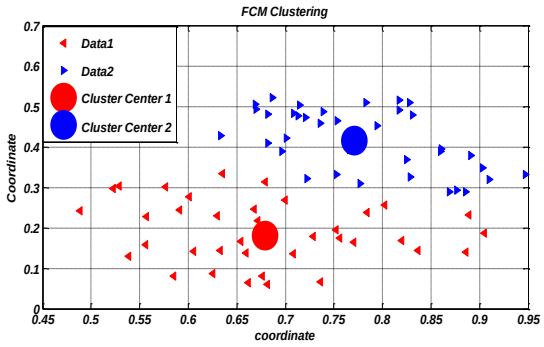

Fig.2. A plot FCM clustering for 2 clusters.

Fig. 3 shows the plot of distortion of the test speech signal with all speakers speech signal available in train database using FCM modelling technique.

Distortion of the speakers test speech signal using FCM modelling technique. D = 6.471, 6.026, 12.653, 5.605, 10.039, 5.968, 8.007, 5.688, 13.014, 6.097, 5.633, 5.843, 5.573, 5.532, 6.563, 13.892, 5.654, 4.927, 6.662, 4.666, 5.8925 14.513, 8.443, 9.994, 5.794, 10.055, 6.195, 6.675, 5.192, 7.123, 6.711, 12.066, 5.209, 5.408, 6.251.

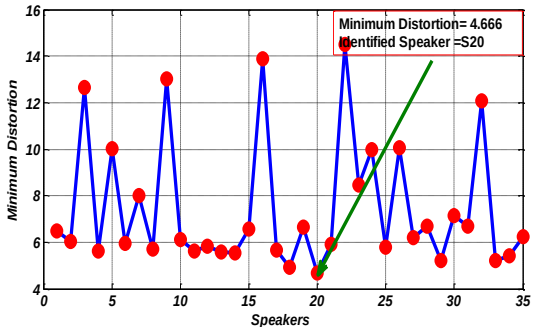

Fig.3. FCM modelling based distortion.

B. *Fuzzy Vector Quantization2*

The achievability of this decision was tried by

checking that the $u_{i,k}$ to resolve outlier circumstances. The codebook vectors can be assessed for this situation by v_i, coming about because of the minimization of Jm. The formula used in FCM for assessing the $u_{i,k}$ and v_i with the vector assignment methodology projected in section II A, same results we are going to utilize in FVQ2 algorithms. Development of Jm in this case must be less than 0.00001 in successive iteration.

FVQ2 Objective Function Jm =2.19022, 1.55145, 1.29068, 0.94758, 0.73473, 0.64221, 0.57287, 0.52184, 0.48639, 0.46198, 0.45161, 0.44769, 0.44567, 0.44437, 0.44346, 0.44279, 0.44227, 0.44185, 0.44148, 0.44114, 0.44080, 0.44045, 0.44005, 0.43960, 0.43907, 0.43848, 0.43785, 0.43725, 0.43672, 0.43631, 0.43602, 0.43583, 0.43570, 0.43562, 0.43557, 0.43554, 0.43552, 0.43551, 0.43550.

In FVQ2 clustering the objective function determines the clustering terminations. Here the minimum objective function achieved 0.43550 after 39 iterations with 0.00001 improvements.

The minimum value of objective function obtained with FCM modelling technique is Jm = 0.43550. Fig. 4 demonstrates the plot of objective function by FVQ2 modelling technique.

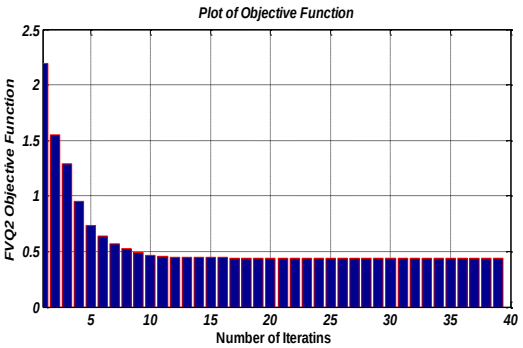

Fig. 4. Plot of FVQ2 modelling technique objective function.

A case of FVQ2 clustering is shown in Fig. 5. FVQ2 comparatively takes less computational time but it is very sensitivity to the initial guess, noisy speech data and could not able to solve the outliers. The downside of over fitting can be to a great extent dispensed with by utilizing the input speech data construct FVQ2 which works with respect to the guideline of physical elucidation of the information groups.

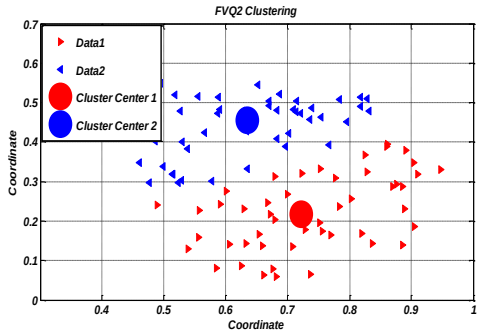

Fig.5. A plot of FVQ2 clustering for 2 clusters.

Fig. 6 demonstrates the plot of FVQ2 modelling technique distortion of the test speech signal with all speakers speech signal available in train database.

FVQ2 modelling technique and its distortion of the test speech signal. D = 6.372, 5.845, 13.192, 5.625, 10.960, 6.340, 7.939, 5.581, 13.320, 6.162, 5.297, 5.794, 5.413, 5.262, 6.013, 14.385, 5.825, 5.133, 7.015, 4.492, 5.362, 14.483, 9.017, 10.713, 5.672, 10.071, 6.0152, 6.7239, 5.204, 6.654, 7.036, 12.445, 5.197, 5.327, 6.159.

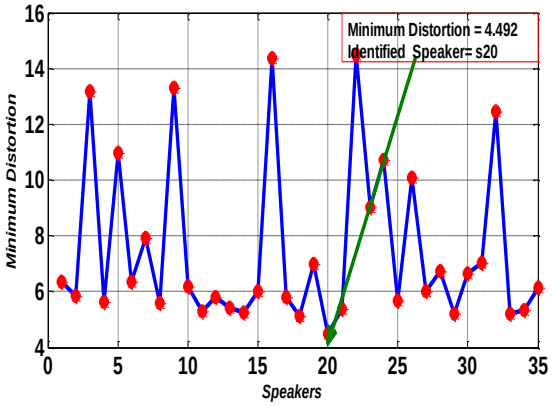

Fig.6. Plot of FVQ2 modelling based distortion.

C. *Novel Algorithm NFVQ*

The Jm of FCM algorithm can be formulated in NFVQ as follows [19]:

$$J_m = \sum_{k=1}^{n}\sum_{i=1}^{c} f(u_{ik}) \| x_k - v_i \|^2 \qquad (3)$$

$$f(u_{i?}) = \frac{1}{2}u \quad + \frac{1}{2}(u \quad)^2 \qquad (4)$$

NFVQ performs the accompanying strides amid clustering

i. First start the cluster membership values,

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

ii. compute the cluster centers,

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

iii. Update $u_{i,k}$

iv. Compute Jm

Do again step ii–iv until ***Jm*** enhances by not exactly a predefined threshold or number of Iterations.

NFVQ Objective Function ***Jm*** = 2.13237, 1.55066, 1.30806, 0.94819, 0.74236, 0.64364, 0.57082, 0.51670, 0.48705, 0.47568, 0.47165, 0.46972, 0.46831, 0.46692, 0.46533, 0.46329, 0.46057, 0.45703, 0.45272, 0.44799, 0.44348, 0.43987, 0.43745, 0.43599, 0.43515, 0.43465, 0.43433, 0.43409, 0.43387, 0.43365, 0.43339, 0.43307, 0.43266, 0.43210, 0.43134, 0.43027, 0.42882, 0.42695, 0.42492, 0.42317, 0.42202, 0.42141, 0.42112, 0.42099, 0.42092, 0.42089, 0.42088, 0.42087.

In NFVQ clustering the objective function determines the clustering terminations. Here the minimum objective function achieved 0.42087 with the improvement of 0.00001 after 49 iterations.

The minimum value of objective function obtained with FCM modelling technique is ***Jm*** = 0.42087. Fig. 7 demonstrates the plot of objective function by NFVQ modelling technique.

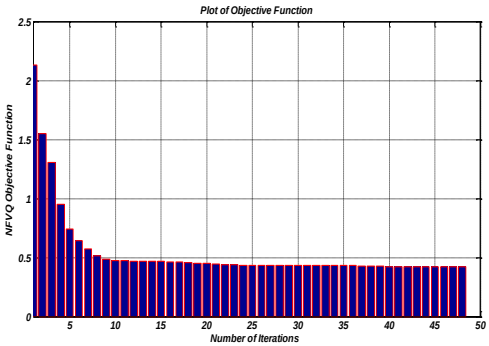

Fig.7. Plot of NFVQ modelling technique objective function.

A case of NFVQ clustering is shown in Fig. 8. NFVQ comparatively takes more computational time and capable of handling noisy speech data at the same time solve the outliers.

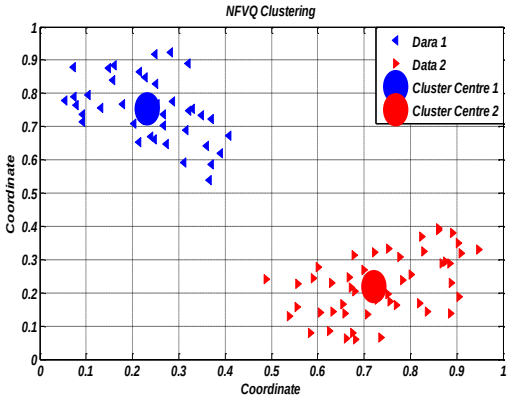

Fig.8. A plot of NFVQ clustering for 2 clusters

Fig. 9 demonstrates the plot of distortion of the test speech signal with all speakers speech signal available in train database using NFVQ modelling technique.

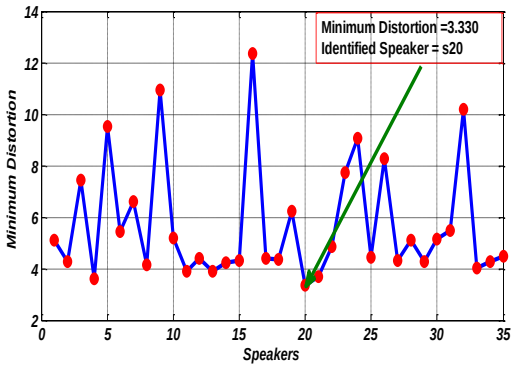

Fig.9. Plot of NFVQ modelling based distortion.

Distortion of the test speech signal of the speakers during testing phase using NFVQ modelling technique. D = 5.116, 4.288, 7.422, 3.590, 9.528, 5.442, 6.608 4.162, 10.928, 5.179, 3.897, 4.381, 3.890, 4.231, 4.307, 12.370 4.383, 4.343, 6.243, 3.334, 3.703, 4.874, 7.709, 9.045, 4.430, 8.266, 4.330, 5.113, 4.277, 5.136, 5.472, 10.205, 4.012, 4.267, 4.478.

IV. Experiments

A. *Experimental setup*

The training speech data was restricted to roughly 3 sec, while test speech data was made for 2, 3, and 4 sec. Experimental assessment of the modelling algorithms continued with the 100 speakers from TIMIT database. The TIMIT database contains voice 630 individuals that are part in subsets as indicated by the Dialect Region to which each of them has a place. Each DR has been used into train and test ASR.

We recorded speech in the DSP lab of St Peter's Engineering College, Hyderabad (India) in Electronics and Communication Engineering Department [20]. This lab was expected to give high SNR expected under the normal circumstances without inordinate "pops" because of breath noise (some pops still happen in the recordings). We have recorded voice for this examination was at one sitting for every speaker. The content of the utterance was arbitrarily chosen by the speaker. The fundamental voice recordings comprise of ninety male and ten female speakers at the rate of 16.0 kHz with 16 bits resolution per sample. Voice test was recorded at 3–4 cm far from the recording device. We considered the impact of the mask and the unconstrained discourse against content perusing. In this manner, some subjects likewise finished extra talking undertakings where they were advised to change their voice intentionally keeping in mind the end goal to be not perceived effectively, or talk suddenly on a conventional topic, for example, the climate or individual emotions [21].

B. *Front-End Processing*

Test and train signal investigation outline frame size were 30 msec with the time shift of 10 msec. The number of mel- frequency bins is 25 and frequency range is from 20 Hz to 7.5 KHz.

C. Experimental Result

The feature extraction of the speaker is performed based on MFCC classifier and modelling with codebook size of 64, 128, 256 and 512.

Table 1. shows the efficiency of TIMIT database for different codebook size with FCM, FVQ2 and NFVQ modelling techniques. One can observe from table 1. that codebook size 256 with NFVQ show significantly high efficiency.

Table 1. ASR efficiency of TIMIT Database

Codebook Size	*64*	*128*	*256*	*512*
FCM Efficiency (%)	96.9	97	97.4	97.2
FVQ2 Efficiency (%)	97.3	97.5	98.1	97.7
NFVQ Efficiency (%)	98.1	98.3	98.8	98.4

Table 2. ASR efficiency of self-collected Database

Code Vector Size	*64*	*128*	*256*	*512*
FCM Efficiency (%)	96.1	96.3	96.9	96.5
FVQ2 Efficiency (%)	96.3	96.7	97.6	96.8
NFVQ Efficiency (%)	97.6	97.8	98.1	97.6

It can be observed in Table I. and Table II. that both corpora demonstrate the same general trend with NFVQ perform well. The speaker recognition efficiency indicates that for FCM, FVQ2 and NFVQ modelling technique an increase of the number of cluster for the most part prompts a recognizable increment of speaker recognition efficiency when the number of codebook size increases from 128 to 256, further increment from 256 to 512 demonstrates a little debasement in execution prompting marginally bring down speaker recognition efficiency.

The principle behind the efficiency degradation is seen because of the expansion in various codeword can be attributed to the more trim data distribution. With the expanding number of codeword, the data is exceptionally dispersed and the codeword are along these lines not able to do model a specific speaker precisely, which at last falls apart the efficiency.

V. System Evaluation

The EER is the most broadly utilized in ASR for the performance measure. In biometric authentication system reference threshold, FAR, FRR goes hand in hand [22].Along these lines, the ASR performance of the Fuzzy modelling system was likewise measured utilizing the EER. Since the EER must be ascertained for a fixed number of code vectors, a codebook containing 256 codeword's was utilized to describe the performance of ASR between FCM, FVQ2, and NFVQ.

Fig. 10 and Fig. 11 represent the miss versus false alarm probability and EER values for FCM, FVQ2, and NFVQ modelling techniques utilizing codebook size of 256. Fig.10 demonstrates for TIMIT database and Fig.11 demonstrate for the self-collected database.

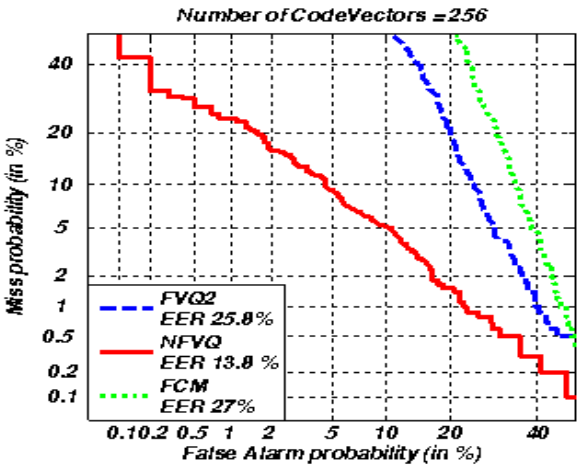

Fig.10. Plot of EER Performance of TIMIT database.

The miss likelihood measures the percent of invalid matches and the false alarm likelihood measures the percent of substantial inputs being rejected. The EER parameter speaks to the rate at which both the miss likelihood and the false alarm likelihood are level with. The lower the EER, the more precise the ASR system is considered.

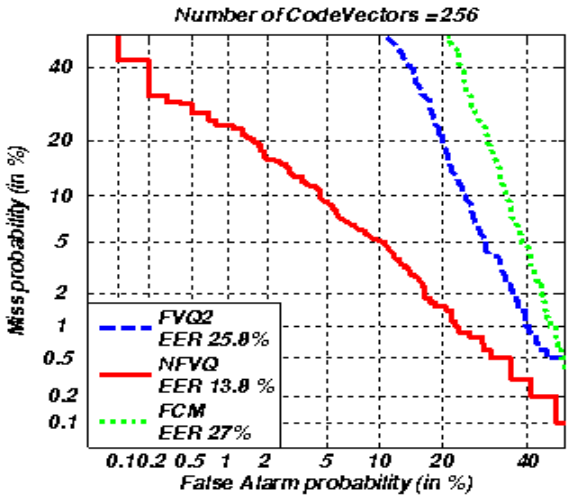

Fig.11. Plot of EER Performance of the self-collected database.

As demonstrated in Fig. 10 and Fig. 11 both corpora demonstrate the same pattern with NFVQ beating both FCM and FVQ2 modelling techniques.

The FCM modelling technique gave the most elevated EER 27% and 19.8 %, FVQ2 gave medium EER 25.8% and 18.6% for self-collected and TIMIT database respectively. At last, the NFVQ gave the least EER 13.8% for the self-collected database and 10.2% for TIMIT.

VI. Discussion

It is clear that with limited speech data another modelling technique in ASR is not going to handle the situation. The proposed NFVQ speaker modelling technique gives quantifiable changes overall corpora and test conditions. One can observe a significant improvement in speaker recognition efficiency for TIMIT and self-collected corpus (EER improvement 19.8-12.2 % and 27-13.8%). Since the self-collected database is made in real working condition, what's more, speaker information is gathered in a single recording session, this speaks to an ideal working environment for the state of art ASR. In Fig. 10 and Fig. 11 one can see the same trends of EER improvement. The proposed algorithm gives the highest efficiency for limited speech data but the system complexity is high. The number of iteration performed by FCM, FVQ2 and NVVQ 42, 39 and 49 respectively with minimum distortion of 4.669, 4.4920 and 4.3300.

As NFVQ gives less distortion with the training speech data it enhances the recognition efficiency. But NFVQ increases the system complexity and make ASR little sluggish. In Fig. 8 clearly indicates the overlapping and outlier has been removed and hence ASR performance is enhanced in NFVQ algorithm.

VII. Conclusion and Future Work

In this paper, we have studied the experimental assessment of different modelling techniques for ASR. The NFVQ modelling technique was designed to capture the advantages provided by fuzzy decision-making processes while maintaining the computational capabilities achieved by crisp decision making processes. Thus, we propose that the NFVQ modelling technique can be efficiently used in ASR to model the speaker data in limited speech data to enhance the efficiency. Positive aspects of the algorithm are co-operative and non-cooperative speaker can be identified as the developed system is capable of working for even one sec of speech data and it is the text-independent system. It recognizes the speaker at a rapid rate as compared to the previously developed system. System complexity is the less as it processes short duration of test and train speech data.

Observed deficiencies are once the number of speaker's increases in system database the identification accuracy decreases and the unknown voice must come from a fixed set of known speaker which has been already used in training the system. It is not suitable for open set speakers.

At present, we are utilizing Euclidean distance based computation in ASR to measure the separation of limited data that are highly correlated a superior distance measuring technique should be investigated to enhance the efficiency of speaker recognition. The feasibility of NFVQ modelling technique should be confirmed for large size database under limited speech data.

Acknowledgment

The author would like to thank to Dr. B. Yegnanarayana, Director IIIT, Hyderabad for his support in providing the TIMIT database that has resulted in bringing out this original research work.

References

[1] Prashanthi, Satyanand Singh, Dr. E.G. Rajan and Pat Krishanan, "Sparsification of Voice Data Using Discrete Rajan Transform and its Applications in Speaker Recognition," IEEE International Conference on Systems, Man, and CyberneticsOctober 5-8, San Diego, CA, USA, pp. 437-442, 2014.

[2] H.S. Jayana, S. R. M. Prasanna, "Fuzzy vector quantization for speaker recognition under limited data conditions," TENCON 2008 , IEEE Region 10 Conference, pp. 1-4, Nov. 2008.

[3] Satyanand Singh and ajeet Singh, "Accuracy Comparison using Different Modeling Techniques under Limited Speech Data of Speaker Recognition Systems," The Global Journal of Science Frontier Research, Vol 16, No 2-F , 2016.

[4] P. Angkititrakul and J. H. L. Hansen, "Discriminative In-Set/Out-of-Set Speaker Recognition," IEEE Trans. Audio Speech Language Processing, vol. 15(2), pp. 498-508, Feb. 2007.

[5] P. Angkititrakul, J. H. L. Hansen, and S. Bagahaii, "Cluster-dependent modeling and confidence measure processing for in-set/out-of-set speaker identification," in Proc. Odyssey 2004 Speaker Lang. Recognition Workshop, pp. 2385–2388, 2004.

[6] P. Angkititrakul and J. H. L. Hansen, "Identifying in-set and out-of-set speakers use neighbourhood information," in Proc. ICASSP'04, pp. 393–396, 2004.

[7] Soumendu Das and Sreeparna Banerjee, "An Algorithm for Japanese Character Recognition," I.J. Image, Graphics and Signal Processing Vol. 7, No. 1, PP.9-15, December 2014.

[8] S. Kwon and S. Narayanan, "Robust speaker identification based on selective use of feature vectors," Pattern Recognit. Lett, vol. 28, pp. 85–89, 2007.

[9] Lantian Li, Dong Wang, Chenhao Zhang, and Thomas Fang Zheng, "Improving Short Utterance Speaker Recognition by Modeling Speech Unit Classes," IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 24, Issue. 6, pp. 1129-1139, 21 March 2016.

[10] M.W. Mak, R. Hsiao, and B. Mak, "A comparison of various adaptation methods for speaker verification with limited enrollment data," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2008, vol. 1. IEEE, 2006, pp. I–I.

[11] R. J. Vogt, C. J. Lustri, and S. Sridharan, "Factor analysis modelling for speaker verification with short utterances," in The Speaker and Language Recognition Workshop. IEEE, 2008.

[12] A. Kanagasundaram, R. Vogt, D. B. Dean, S. Sridharan, and M. W.Mason, "i-vector based speaker recognition on short utterances," in Proceedings of the 12th Annual Conference of the International Speech Communication Association. International Speech Communication Association (ISCA), 2011, pp. 2341-2344.

[13] Melek, W.W., Emami, M.R., Goldenberg, A.A., "An improved robust fuzzy clustering algorithm," Fuzzy

Systems Conference Proceedings, FUZZ-IEEE '99. IEEE International vol.3, pp.1261-1265, 1999.
[14] S.Singh and Dr. E.G. Rajan, "Application Of Different Filters In Mel Frequency Cepstral Coefficients Feature Extraction And Fuzzy Vector Quantization Approach In Speaker Recognition," International Journal of Engineering Research & Technology, Vol. 2 Issue 6, pp.419-425, June 2013
[15] Moyen Mohammad Mustaquim., "Fuzzy-Logic Controller for Speaker-Independent Speech Recognition System in Computer Games," Universal Access in Human-Computer Interaction. Applications and Services Vol. 6768 of the series Lecture Notes in Computer Science, pp. 91-100, July. 2011.
[16] Jasdeep Kaur and Manish Mahajan, "Hybrid of Fuzzy Logic and Random Walker Method for Medical Image Segmentation," I.J. Image, Graphics and Signal Processing, Vol. 7, No. 2, January 2015.
[17] Jacek M. Leski, Marian Kotas, "Generalized fuzzy c-means clustering strategies using L p norm distances," Journal Fuzzy Sets and Systems., Vol 279 Issue C , pp. 112-129, Nov. 2015.
[18] N. B. Karayiannis, P.I. Pai, "Fuzzy Vector Quantization Algorithms and Their Application in Image Compression," IEEE Trans Image Processing, vol. 4, no.9, pp. 1193-1201, 1995.
[19] 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, pp. 1-6, 23-26 May 2016.
[20] S.Singh and Dr. E.G. Rajan, "MFCC VQ Based Speaker Recognition and Its Accuracy Affecting Factors.," International Journal of Computer Application. Vol. 21, No 6, pp 1-6, May-2011.
[21] Shashidhar G. Koolagudi , Kritika Sharma , K. Sreenivasa Rao, "Speaker Recognition in Emotional Environment," International Conference, ICECCS 2012, Kochi, India, August 9-11, 2012.
[22] Jyoti Malik,Dhiraj Girdhar,Ratna Dahiya, and G. Sainarayanan, "Reference Threshold Calculation for Biometric Authentication," I.J. Image, Graphics and Signal Processing, Vol. 6, No. 2, pp. 46-53, 2014.

Authors' Profiles

Dr. Satyanand Singh received the M.E. and Ph.D. degrees in Electronics and Communication Engineering from NIT Rourkela and the JNTU Hyderabad, India, in 2002, and 2016, respectively. Presently he is working with CMR Institute of Technology, Hyderabad as Associate professor in ECE ddepartment. His primary research interests include speaker recognition, robust speech modeling and feature extraction, pattern recognition and biometrics.

Abhay Kumar is a Ph.D research scholar at the Sri Satya Sai University of Technology & Medical Sciences (SSSUTMS), Sehore, Madhya Pradesh, India. He has Masters' degrees in Computer Science and Engineering from College of Engineering, Guindy, Chennai, India. Professionally he has over fifteen years of experience in teaching Computer Science subjects at Under Graduate and Post Graduate level.

David Rau Kolluri received M.Tech Degree in Computer Science & Enginnering From JNTUH in 2010, B.Tech(CSE) from JNTUH in 2002 and Diploma in CSE from Kakatiya University in 1996. He is currently pursuing PhD (CSE) from Rayalaseema University in the area Data Mining. His research interests include Big data and Machine Learning. He is having total 15 years of teaching experience. He is currently working with St.Peter's Engineering College as a associate professor in the department of Computer Science and Engineering and strong in programming language subjects.

How to cite this paper: Satyanand 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, pp.41-48, 2016.DOI: 10.5815/ijigsp.2016.11.06

