Using Vector Quantization for Universal Background Model in Automatic Speaker Verification

Djellali Hayet[1], Laskri Mohamed Tayeb[2]
[1,2] Badji Mokhtar University Annaba Algeria, Computer Science Department[1,2],
LRS Laboratory[1], LRI Laboratory[2]
Badji Mokhtar University, P-O Box 12, 23000 Annaba, Algeria

Abstract. We aim to describe different approaches for vector quantization in Automatic Speaker Verification. We designed our novel architecture based on multiples codebook representing the speakers and the impostor model called universal background model and compared it to another vector quantization approach used for reducing training data. We compared our scheme with the baseline system, Gaussian Mixtures Models and Maximum a Posteriori Adaptation. The present study demonstrates that the multiples codebook gives more verification accuracy called equal error rate but this improvement also depends on the codebook size.

Keywords: Vector Quantization, Speaker Verification, Codebook, false Acceptance, False reject, Universal Background Models, Linde Buzo Gray.

1 Introduction

The speaker verification is a field of speaker recognition which the main objective is to authenticate a person's claimed identity. The speaker voice is used to recognize him (her), we create two models, the first one is the speaker model and the second is the impostor model called universal background model UBM. The recorded speech is preprocessed, compared to speaker and UBM model in order to compute the score and finally compared to threshold.
It has been proved that the variation factors like speaker identity, utterance length, gender, session, transmission channel, speaking, affect the system performance [1][2][3]. Intra speaker variability influences the verification performance system. Thus, it is important to record each speaker at different time but also means the huge speech data.
The state of the art of text independent speaker recognition is Gaussian mixture model and Maximum a posteriori adaptation. Speaker dependent GMM are derived from the speaker independent model called universal background model (UBM) and Maximum a posteriori adaptation MAP using target speaker speech data.
Vector Quantization (VQ) model was introduced in 1980's used in data compression [4]. VQ is one of the simplest text independent speakers model, and often used for computational technique. It also provides good accuracy when combined with background model adaptation [4][5].

In VQ based speaker recognition, each speaker is characterized with the set of code vectors and is referred to as that speaker's codebook. Normally, a speaker's codebook is trained to minimize the quantization error for the training data from that speaker. The most commonly used training algorithm is the Linde-Buzo-Gray (LBG) algorithm [6].
When the speaker speech data becomes huge, it involves the time consuming problem. Gurmeet replaced the EM algorithm with LBG algorithm. Experimentally, they found that the complexity of calculation can be reduced by 50% compared to the EM algorithm. The reason is the LBG algorithm utilize apart of feature vectors for classification [7].
We applied Vector Quantization in Automatic Speaker Verification; usually, each target speaker had his own codebook, when usually the speaker independent models had two gender dependent codebook originates from impostor speakers (male, female).
Our approach aim to select the best universal background model UBM, we try another way to model VQ UBM with set of sub UBM. We divide the features vectors extracted from processing step (Mel cepstral coefficients: MFCC) in a equal size and applied for each of them the LBG algorithm to obtain its codebook (cd1,cd2,…cdK)..
The aim is to get the best sub model with LBG algorithm for impostors (UBM) and then compute the distortion error from optimal Sub UBM. We aim to reduce EER in the presence of small training data of each client and select the best sub UBM.
We organized paper as follows, modeling speakers based on vector quantization and MAP adaptation is introduced in Section 2, and the ASV architecture proposed in Section 3 followed experiments in Section 4 and conclusion in section 5.

2 Vector Quantization and MAP Adaptation

We introduce vector quantization and Maximum a posteriori adaptation in Automatic Speaker Verification:

2.1 Vector Quantization

Vector Quantization (VQ) is a pattern classification technique applied to speech data to form a representative set of speaker features. It was introduced to speaker recognition by Soong [8]. In speaker verification, Vector quantization (VQ) model were applied in Soong and Rosenberg, It is one of the simplest text-independent speaker models and usually used for computational speed-up techniques, it also provides competitive accuracy when combined with background model adaptation [5][8][9][10].

In the training phase, a speaker-specific VQ codebook is generated for each known speaker by clustering his training acoustic vectors. The distance from a vector to the closet codeword of a codebook is called a VQ distortion [4][11].

In the Test phase, an input utterance of a known voice is vector-quantized using trained codebook from proclaimed identity and the speaker independent model codebook (Universal Background Model). The total VQ distortion is computed.

In principle, when we get a large amount of training vectors representing speaker in the training vectors. We should reduce it by vector quantization. Suppose there are N vectors, to be quantized, the average quantization error is given by

$$\mathbf{E} = \frac{1}{N} \sum_{t=1}^{N} \mathbf{e}(\mathbf{x}\ \mathbf{t}) \qquad (1)$$

The task of designing a codebook is to find a set of code vectors so that E is minimized. However, the commonly used method is the LBG algorithm [6].

In speaker verification, the codebook is used for classification and minimizing the quantization error. We selected LBG algorithm defined as the iterative improvement algorithm or the generalized Lloyd algorithm. Given a set of N training feature vectors, {t1, t2,.tn} characterizing the variability of a speaker, we search a partitioning of the feature vector space, {S1, S2,..., SM}, for that particular speaker where S, the whole feature space, is represented as S = S1 U S2 U...U SM.

The performance of a quantizer is designed by an average distortion between the input vectors and the final vectors, where E represents the expectation operator (equation 1).

2.2 Gaussian Mixture Models & MAP Adaptation

GMM-UBM-Maximum Likelihood Modeling: this approach is based on training UBM male model with Gaussian mixture model and the other female UBM (from female speech). The model parameters (mean, covariance and weight of the Gaussian) are trained with the EM algorithm (Expectation-Maximization).

Maximum a Posteriori approach MAP resolve the problem of maximum likelihood ML(can't generalize well to unseen speech data in low training data). MAP use prior knowledge of the distribution of the model parameters and insert it in modeling process[12][13]. The Maximum A Posteriori MAP approach is to use the world model and client training data to estimate the client model on the basis of these data and MAP Adaptation [12][13] [14][15][16].

The client model is derived from the world model by adapting the GMM parameters (mean, covariance, weights) estimated. However, experimentally, only the averages of GMM are adapted [13].

3 Speaker Verification Architecture Based on Vector Quantization

We proposed two VQ-UBM models, the first one is the baseline system, the second is VQ Sub-UBM. We describe our new modeling UBM:

3.1 Training Phase

VQ Sub UBM

The acoustics vectors obtained in features extraction were split in subset of data with the same dimension and served to create codebook {CDU1, CDU2, …, CDUk} for world model UBM. We divide UBM speech data in N subsets instead of one global UBM, in figure 1, after feature extraction, the MFCC vectors were the input of L.B.G algorithm which provide K codebook.

Codebook

There are several different approaches to finding an optimal codebook. The idea is to begin with a vector quantizer and a codebook and improve upon the initial codebook by iterating until the best codebook is found. We aim to reduce redundancy in UBM data by clustering, to do that, we implement this algorithm:

Algorithm 1: VQ Sub-UBM

Training Phase

```
Input : MFCC vectors; Output: Codebook CDU(1..M).

We divide MFCC vector in equal sub matrix and applied
LBG algorithm for each of them.

   Input[ C ]= MFCC vector(Feature Extraction).

   Split C in M equal sub matrix Ci;

Train UBM of each Ci for different size of
codebook(k=16,32,64,128,256); Result= CDU (i=1..M).
```

Test Phase

```
In recognition phase, we compute Euclidean distance and
evaluate quantization error from each codebook and test
vector,

We choose the best codebook with minimal quantization
error.
```

The quantization square error ESQ

$$MSE(X,Y) = \frac{1}{|x|}\sum_{Xi} \min \left| |Xi \quad yk| \right|^2 \qquad (2)$$

Where $yk \in Y$; Xi :vector data; yK : centroid

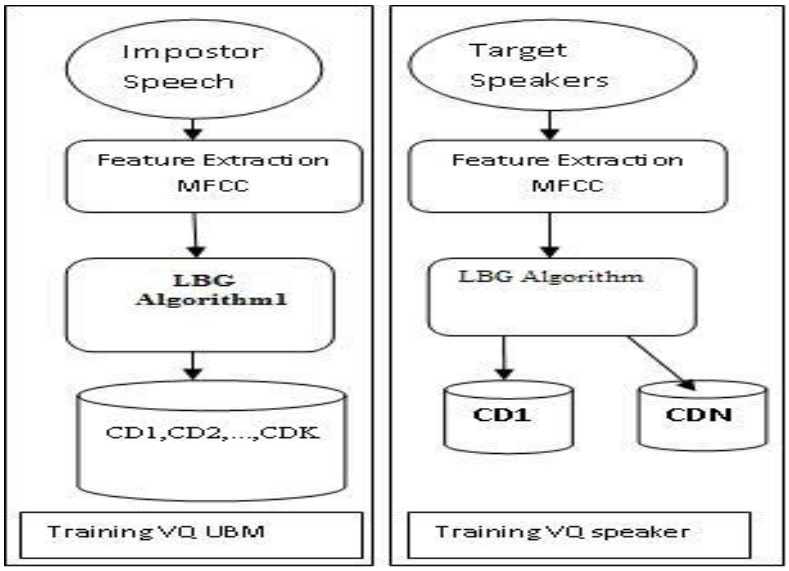

Fig 1. Vector Quantization Architecture: VQ Sub-UBM is applied for UBM only

3.2 Test Phase

We compute the threshold (CDT) from 8 male and 8 females' speakers others than UBM speakers and trained by LBG algorithm.

✓ **Test Algorithm**

```
CDU: UBM codebook; CDS: Speaker codebook;

VQdist : VQ distorsion

Input : X=speaker speech ,claimed identity,

     MFCC= Feature Extraction(X)

      For i=1 to M  VQcdu=VQdistorsion(X,CDUi) End

CDUoptimal=CDU best cobdebook UBM  where
Argmin(VQdistorsion(X,CDUi))

VQdist(speaker) = VQdist(X,CDS)  – Vqdist(X,CDUoptimal)

If   VQdist(speaker)>  VQdist(CDT)  then  client acces
```

```
Else    reject
```

4 Protocol Experiment

In this section, we describe a set of experiments designed to evaluate the performance of the proposed system under a variety of condition and compare it to baseline system GMM MAP and standard VQ UBM.

4.1 Database and Baseline System

The Arabic database is recorded in Goldwave frequency 16KHz for a period of 60s for each speaker when training and 30s in the testing phase. The UBM population is 15 men's and 15 women. Four sessions are recorded for each speaker at an interval of 1 month. Ten clients are registered in the database (5 men and 5 women).

4.2 VQ Sub-UBM Model

We extract MFCC vector for all acoustics data allowed to UBM training and applied LBG algorithm for it. We obtain one centroid (NxT) by gender, where we try different value of N=k=16, 32, 64, 128, 256. In recognition phase, we compute Euclidean distance and evaluate quantization error (equation 1) from centroid and test vector, we computed codebook for each target speaker and finally evaluate the score.

TABLE I. VQ-SUB UBM PERFORMANCES

CodeBook Size	FA(%)	FR(%)
CD32	22,86	23,86
CD64	25,71	23,86
CD128	7,14	22,73
CD256	14,29	22,73

4.3 Baseline VQ UBM Model

We compute one codebook for the Baseline VQ UBM and evaluate LBG algorithm for k=16, 32, 64, 128. We built UBM models from 30 Arabic speakers; UBM male with 15 male speakers and UBM female from 15 female speakers. The global threshold is computed from other database: 8 male and 8 female speakers.

TABLE II. BASELINE VQ UBM PERFORMANCES

CodeBook Size	FA(%)	FR(%)
CD32	14,29	73,03
CD64	12,86	4,89

4.4 Baseline GMM MAP system

We train universal background model UBM gender dependent(male, female) under expectation maximization algorithm EM and create each target speaker model with GMM MAP approach, we try different sizes of GMM (8, 16, 32,64,128) and evaluate the value of false acceptance and false rejection.

TABLE III. GMM MAP SYSTEM RESULTS

#Gaussians	8	16	32	64	128
GMMMAP EER %	**19.16**	36.12	35.2	35	36.04

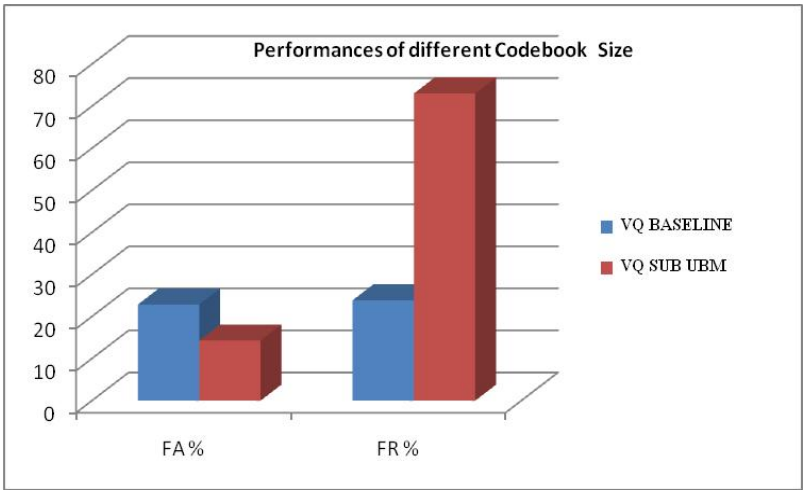

Fig 2. Comparaion of VQ Baseline and VQ SUB UBM

5 Discussions

We compare different modeling speaker techniques: VQ Sub-UBM, Baseline VQUBM and GMM MAP their performances were evaluated using the same data and front end processing.

Table I shows the value of false acceptance and false rejection for different codebook size (32, .., 256) in VQ SUB UBM approach and observe that the best value is designed for 128 codebook size (7.14% and 22.73%). The result in table 1 provide

more accuracy recognition than table II for codebook size=32(FA=22.86%; FR=23.86%) and worst for codebook size 64. We observe that the size of codebook influences the performance and the multiple UBM provide better result.

Figure 2 demonstrate the performance of VQ SUB UBM is worst than VQ UBM in false rejection, however we tested only VQ UBM with 32 and 64 codebook size.

In Baseline GMM MAP system, Equal error rate is 19.16% for 8 mixtures and between [35%-36.12%] for model order M=16...128. The performances decreases because the reduced speech data and didn't apply normalization technique like Tnorm.

6 Conclusions

VQ SUB UBM achieved (FA=7.14% and FR=22.73%) for 128 codebook size and improved the performance of vector quantization applied in speaker verification compared to baseline vector quantization. The codebook size influences the verification accuracy. The size of speech data should be increased in order to validate our experiments in large database.

References

1. Campbell, J.: Speaker Recognition, A Tutorial. Proc. IEEE 85 (9), pp. 1437--1462 (1997)
2. Reynolds D. A, Rose R. C. : Robust Text Independent Speaker Identification Using Gaussian Mixture Speaker Models, IEEE Trans. Speech Audio Processing, vol. 3, pp. 72-- 83 (1995)
3. Doddington, G., Liggett, W., Martin, A., Przybocki, M., Reynolds, D.A.. : Sheeps, goats, lambs and wolves., : A Statistical Analysis of Speaker Performance in the NIST 1998 Speaker Recognition Evaluation. In Proc. of ICSLP (1998)
4. Wan-Chen C., Ching-Tang H., Chih-Hsu H,. : Robust Speaker Identification System Based on Two-Stage Vector Quantization, Tamkang Journal of science and engineering, Tamkang Journal of Science and Engineering, Vol. 11, No. 4,. pp. 357-- 366 (2008)
5. Jialong H, Li L, Gunther P, : A Discriminative Training Algorithm for VQ-Based Speaker Identification.IEEE Transactions on Audio and Signal Processing,vol 7, (1999)
6. Linde Y., Buzo A., Gray R.M., : An Algorithm for Vector Quantizer Design," IEEE Trans. Commun., vol. 20, pp. 84 --95 (1980)
7. Gurmeet S, Panda S, Bhattachryya S. Srikanthan S., : Vector Quantization Technique for GMM Based Speaker Verification. IEEE International conference on acoustics speech and signal processing, USA. pp. 65 -- 68 (2003)
8. Soong F.K., Rosenberg A.E, Rabiner L.R,Juang B.H 1985, : A Vector Quantization Approach to Speaker Recognition. IEEE International Conference on Acoustics speech and signal Processing, pp. 387 -- 390 (1985)
9. Rosenberg A. E., Soong F. K.. : Evaluation of a Vector Quantization Talker Recognition System in Text Independent and Text Dependent Modes. Comput. Speech Lang, vol .22, pp. 143 -- 157 (1987)
10. Jenq-Shyang P, Thesis: Improved Algorithms For VQ Codeword Search, Codebook Design and Codebook Index Assignment. University of Edeinbourgh (1996)

11. Kinnunen, T., Saastamoinen, J., V., Hautomaki, M., Vinni, P., Franti, : Comparing Maximum a Posteriori Vector Quantization and Gaussian Mixture Models in Speaker Verification, Pattern recognition letters, (2008)
12. Preti, A. : Thesis, Surveillance de Réseaux Professionnels de Communication par la Reconnaissance du Locuteur. Académie d'Aix Marseille, Laboratoire d'informatique d'Avignon (2008)
13. Bimbot, F., Bonastre, J.F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., Merlin, T., Ortega-Garcia, J., Petrovska-Delacretaz, D., Reynolds, D.A. : A tutorial on Text- Independent Speaker Verification. J. Appl. Signal Process. 4. pp. 430 -- 451 (2005)
14. B Vesnicer, F Mihelic., : The Likelihood Ratio Decision Criterion for Nuisance Attribute Projection in GMM Speaker Verification, Hindawi Publishing Corporation Eurasip Journal On advances in Signal Processing volume (2008)
15. Reynolds, D.A. Quatieri, T.F. Dunn, R.B. Speaker Verification Using Adapted Gaussian Mixture Models. Digital Signal Processing. 10. pp. 19 -- 41 (2000)
16. Furui S, : Speaker Dependent Feature Extraction, Recognition and Processing Techniques ,NTT Human interface Laboratories, Japan Speech Communication , Elsevier Science Publishers North-Holland. pp. 505--520 (1991)