

# New scheme based on GMM-PCA-SVM modelling for automatic speaker recognition

Kawthar Yasmine Zergat · Abderrahmane Amrouche

Received: 19 October 2013 / Accepted: 21 April 2014 / Published online: 14 May 2014  
© Springer Science+Business Media New York 2014

**Abstract** Most of the existing speaker recognition systems are based on the basic GMM, the state of the art GMM-UBM, the SVM or more recently the GMM-SVM modeling. In this paper, a new scheme for Automatic Speaker Recognition (ASR), namely GMM-PCA-SVM, is presented. Dimensionality reduction using Principal Component Analysis (PCA) technique, which was previously applied in the front-end process, is now incorporated in the core of the GMM-SVM modeling part, in order to reduce the size of the adapted means vectors issued from the Universal Background Model (UBM). A Comparative study, using Mel Frequency Cepstral Coefficients (MFCC) with Cepstral Mean Subtraction (CMS) extracted from the TIMIT database is performed for speaker recognition in clean and noisy environments. It is shown that the proposed scheme is a promising way for the ASR task. In fact, the recognition performances using GMM-PCA-SVM proposed method is significantly improved compared to the conventional SVM or GMM-SVM based systems.

**Keywords** Speaker recognition · Dimensionality reduction · Support vector machine (SVM) · PCA · Gaussian supervector (GMM-SVM) · GMM-PCA-SVM · Noisy environments

## 1 Introduction

Due to the growing need for secured access or criminalistic investigations, improving Automatic Speaker Recognition (ASR) systems became an attractive challenge. ASR covers verification and identification. Automatic Speaker Verification (ASV) is the use of a machine to verify a person's claimed identity from his voice. In Automatic Speaker Identification (ASI), there is no a priori identity state, and the system decides who the person is [Campbell \(1997\)](#).

Current state of the art systems in text-independent speaker recognition use cepstral coefficients as baseline features, and speaker modeling techniques, such as Universal Background Gaussian Mixture Models (GMM-UBM) [Reynolds et al. \(2000\)](#) and Gaussian Supervector (GMM-SVM) [Campbell et al. \(2006\)](#).

This work was originally devoted to a robust ASR task using the Support Vector Machine (SVM) [Wan and Renals \(2003\)](#), [Karam and Campbell \(2008\)](#) and the hybrid GMM-SVM based recognizers. Along this study, it clearly appears that the dimensionality reduction is an attractive way to process the huge quantity of data without loss of recognition performance. Many different approaches have been studied to improve the system's accuracy with a minimum size of input data. In [Jokic et al. \(2012\)](#), the authors discuss possibilities for dimensionality reduction of the standard MFCC feature vectors by applying Principal Component Analysis (PCA). The results showed that PCA is an interesting method to reduce dimensionality without decreasing the system performance. The GMM-UBM is adopted in [Li and Dong \(2013\)](#). The MAP (Maximum A Posterior Probability) means have been improved by using the MLLR (Maximum Likelihood Linear Regression) and EigenVoice.

In [Hanihci and Ertas \(2011\)](#), in a first time, the authors made a partition of the UBM data into clusters using the

---

K. Y. Zergat (✉) · A. Amrouche  
Speech Com. & Signal Proc. Lab.-LCPTS, Faculty of Electronics and  
Computer Sciences, USTHB, Bab Ezzouar 16 111, Algeria  
e-mail: zergatyasmine@gmail.com

A. Amrouche  
e-mail: namrouche@usthb.dz

Vector Quantification (VQ) algorithm, afterward the transformation matrix is obtained by applying the PCA on the set of feature vectors in each cluster. Finally, multiple speaker models are constructed using this set of transformed feature vectors through MAP adaptation. Best results were achieved using  $K = 2$  local regions with model order  $M = 256$ . The obtained EER is less than 12.2%. In Minkyung et al. (2010), the authors propose a global eigenvector matrix based PCA for speaker recognition (SR) task, to deal with the large amount of training data when the eigenvector matrix of each speaker is calculated. The authors use training data issued from all speakers to calculate the covariance matrix and use this matrix to find the global eigenvalue and eigenvector matrix to perform PCA technique. The proposed method shows better performance while requiring less storage space.

A Fishvoice based feature fusion method incorporating with PCA, LDA is proposed in Zhang and Zheng (2013). The high dimensional input data is simply projected into a lower-dimensional subspace. Results show that this technique can effectively reduce the Equal Error Rate (EER) for utterances as short as about 2 s. In Jiang et al. (2013), the authors transform the original features extracted from speech files by PCA and KPCA (Kernel-PCA) to select effective emotional features for the Automatic Speech Emotion Recognition (ASER). Results shown that feature dimension reduction seriously improve the accuracy of the ASER system. In Lee (2004), Lee introduced local fuzzy PCA based GMM which creates the regions using fuzzy clustering algorithm followed by PCA for each region. The author concluded that this technique gives comparable performance accuracy for speaker identification task with reduced dimension of data. The best performances are reached with the proposed method. With reduced dimension, the performances are same or better to the conventional GMM, with  $k = 2$  clusters and mixture number equal to 64.

As mentioned above, the main idea of this work is to find a new scheme for speaker recognition modeling based on dimensionality reduction, with improved performance. The ability of the PCA is investigated in order to reduce the size of the adapted mean vectors issued from the GMM-UBM model. Moreover, the paper investigates the influence of the dialect Yun and Hansen (2009, 2011), Chitturi and Hansen (2007) effect on the ASR systems.

The rest of the paper is outlined as follow. Sections 2 reviews the SVM and the GMM-SVM classifiers used for the ASR task. Then, dimensionality reduction applied in the front end part of the ASR system is described in sect. 3. Section 4, detailed the proposed new scheme based on the GMM-PCA-SVM modeling. Section 5 presents the data sets used and the experimental results in both clean and noisy environments. Finally, Sect. 6 concludes this paper.

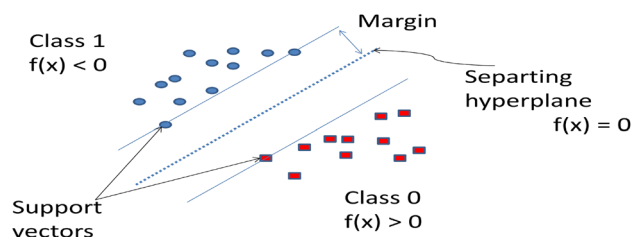


Fig. 1 Principle of support vector machine (SVM) classification

## 2 Speaker recognition using SVM and GMM-SVM

### 2.1 SVM modeling

Support Vector Machines (SVM), is a powerful discriminative classifier that is related to minimizing generalization errors. SVM aims to fit an Optimal Separating Hyperplane (OSH) between classes by focusing on the training samples that lie at the edge of the class distributions, the support vectors, and separates classes using “Maximum-Margin” hyperplane boundary (see Fig. 1).

When data are not linearly separable in the finite dimensional space, a kernel function  $k(\cdot, \cdot)$  is used, this leads to an easier separation between two classes with a Hyperplane. A linear hyperplane in the high dimensional kernel feature space, Hilbert space ( $H$ ), corresponds to a nonlinear decision boundary in the original input space. More details can be found in both Vapnik’s book Vapnik (1998) and Burges’ tutorial Burges (1998).

The SVM is constructed from the sums of a kernel function  $k(\cdot, \cdot)$  as follow:

$$f(x) = \text{sign} \left[ \sum_{t=i}^N \alpha_i t_i k(x, x_i) + b \right] \text{ with } \sum_{t=i}^N \alpha_i t_i = 0 \quad (1)$$

Where  $t_i$  are the ideal outputs,  $x_i$  represent the support vectors, which are the training data.  $\alpha_i$  are Lagrange multipliers and  $b$  represents the bias.

The Radial Basis Function (RBF) and the polynomial kernels are commonly used, and take respectively the following forms:

$$k(x, x_i) = e^{-\gamma \|x - x_i\|^2} \quad (2)$$

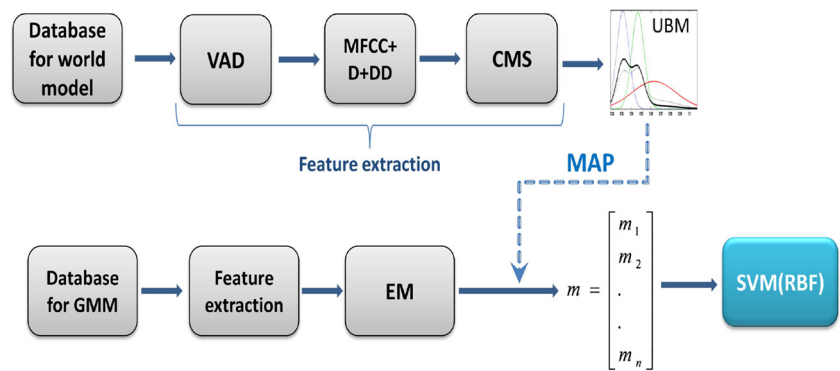
$$k(x_i, x_j) = (x_i \cdot x_j + 1)^d \quad (3)$$

where  $\gamma$  is the width of the Radial Basis Function and  $d$  is the order of the polynomial function.

### 2.2 GMM-SVM speaker recognition

Gaussian Mixture Model (GMM) is a type of density model that is used to represent the speaker and follows the probabilistic rules. The GMMs models are easy to implement,

**Fig. 2** Bloc diagram of the GMM-SVM based speaker recognition system



and are commonly used for Language Identification, Gender Identification and Automatic Speaker Recognition tasks. The GMM model obtains the likelihood of a D-dimensional Cepstral vector  $\vec{x}$  using a mixture model  $\lambda$  of M multivariate Gaussians Reynolds et al. (2000) given by:

$$p(x/\lambda) = \sum_{i=1}^M \pi_i b_i(x) \tag{4}$$

where  $\pi_i$  represents the mixture weights and  $b_i(x), i = 1, \dots, M$  are the component densities given by:

$$b_i(x) = \frac{1}{2\pi^{D/2} |\Sigma_i|^{1/2}} \exp \left[ -\frac{1}{2} (x - \mu_i)' (\Sigma_i)^{-1} (x - \mu_i) \right] \tag{5}$$

with mean vector  $\mu_i$  and covariance matrix  $\Sigma_i$ . The mixture weights satisfy the constraint that  $\sum_{i=1}^M \pi_i = 1$ . These parameters are estimated using the Expectation–Maximization (EM) algorithm Reynolds et al. (2000). For speaker recognition, each speaker is modeled by a GMM and is referred to by its model  $\lambda$ .

The UBM is generally a large GMM learned from multiple speech files to represent the speaker’s independent distribution of features, its parameters (mean, variance and weight) are found using the EM algorithm. The hypothesized speaker specific model is derived by adapting the parameters of the UBM using the speaker’s training speech and a form of Bayesian adaptation MAP Reynolds et al. (2000). The specifications of the adaptation are given below.

Given a UBM model and training vectors from the hypothesized speaker,  $X = \{x_1, x_2, \dots, x_T\}$ , we first determine the probabilistic alignment of the training vectors into the UBM mixture components. That is, for mixture  $i$  in the UBM, we compute

$$\Pr(i/x_t) = \frac{\lambda_i p_i(x_t)}{\sum_{j=1}^M \lambda_j p_j(x_t)} \tag{6}$$

$$n_i(X) = \sum_{t=1}^T \Pr(i/x_t) \tag{7}$$

$$E_i(X) = \frac{1}{n_i} \sum_{t=1}^T \Pr(i/x_t) x_t \tag{8}$$

This is the same as the expectation step in the EM algorithm. Finally, these new sufficient statistics from the training data are used to update the old UBM sufficient statistics for mixture  $i$  to create the adapted parameters for mixture  $i$  with the equations:

$$\bar{\mu}_i = \alpha_i E_i(X) + (1 + \alpha_i) \mu_i, i = 1, \dots, M \tag{9}$$

$$\alpha_i = \frac{n_i(X)}{n_i(X) + r} \tag{10}$$

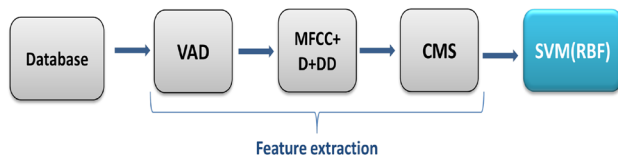
where  $r$  is a fixed relevance factor.

Another approach became more popular, which consists of using the hybrid system GMM-SVM. The main goal is to see the complementary information provided by the traditional GMM to the SVM based system. In this approach, instead of using the MFCC features directly, the hybrid classifier uses the adapted Gaussian means of the mixture components obtained from the UBM and the MAP adaptation as input to the SVM system for the discrimination and the decision task. An illustrative bloc diagram of the GMM-SVM classifier is given on Fig. 2.

### 3 Dimensionality reduction in the front-end part

In order to investigate the influence of dimensionality reduction on the ASR system, the PCA is applied to the input feature vectors (MFCCs) issued from the speech signal for the SVM based speaker recognition system.

The SVM system is based on the principle of structural risk minimization. It is considered to be more suitable for classification and therefore is used in our work. The difficulty of the SVM classifier is setting its respective optimal parameters ( $C, \gamma$ ) to achieve the lower misclassification accuracy. These parameters are calculated during the training phase,



**Fig. 3** Bloc diagram of the SVM based speaker recognition system

and the final step consists of the testing phase, which allows the evaluation of the robustness of the classifier. To calculate the classification function class ( $x$ ) in the SVM model, the RBF kernel was used. All presented results in this study, were obtained using that function.

In this paper, the SVM was trained directly on the acoustic space, which characterizes the client data and the impostor data. In this way, 15 unknown speakers were used to represent the impostors for the recognition task.

For PCA-SVM model, the PCA was applied to the feature vector in the front-end part, and was applied to each speaker independently. This leads to a better representation of the speaker's intra variability and allows reducing the effective size of the input data (MFCCs). The SVM block diagram is given on Fig. 3.

The PCA Jolliffe (2010) technique is an unsupervised feature extraction method Izquierdo-Verdiguier et al. (2014), it rotates the synchronization system in such a way that the directions of the axes are oriented with progressively decreasing variance of the data Kuncheva and Faithfull (2014). This technique allows a transformation from a number of correlated variables into a smaller number of uncorrelated ones Malarvizhi and Sivasarathadevi (2013), the Principal Components (PCs) while preserving the maximum variance during the projection process. The following paragraph details the theoretical fundamentals of the PCA routine.

The initialization step of the system consists of the creation of the eigenspace. Let the training set be the input feature vectors (MFCCs),  $X = \{x_1, x_2, \dots, x_M\}$ . The average mean of the set is defined by:

$$\bar{X} = \frac{1}{M} \sum_{i=1}^M x_i \quad (11)$$

Each feature vector differs from the average mean  $\bar{X}$  by the vector:

$$\vartheta_p = x_p - \bar{X} \text{ with } p = 1 \dots M \quad (12)$$

The rearranged  $\vartheta_p$  vectors construct the  $\delta$  ( $N \times M$ ) matrix that will be subject to the PCA technique Kresimir delac et al. (2005). The covariance matrix of  $X$  using the  $\delta$  matrix is calculated as:

$$C = \delta \delta^T \quad (13)$$

Let  $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$  be the eigenvalues of the covariance matrix  $C$ , ordered from largest to smallest and  $\phi = \{\omega_1, \omega_2, \dots, \omega_n\}$  be the corresponding eigenvectors.  $\phi$  represents the transformation matrix which projects the original data  $X$  onto orthogonal feature space.

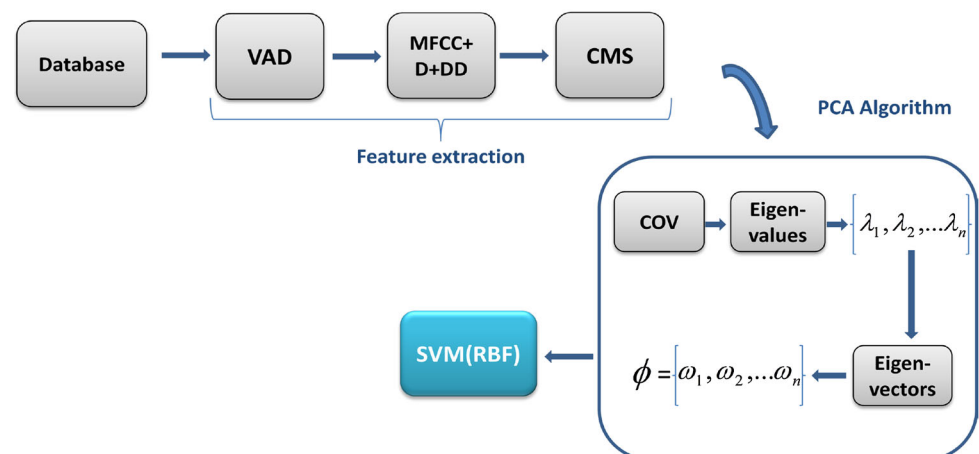
The dimensionality reduction is then made by keeping some number of the principal components that capture most of the variance in the data set, and discarding the rest. So, the transformation matrix  $\phi$  will consist of the first  $D$  eigenvectors which is associated with largest  $D$  eigenvalues, where  $D$  is the new dimension. Figure 4 illustrates the block diagram of the PCA-SVM based speaker recognition system.

#### 4 The proposed GMM-PCA-SVM based speaker recognition modeling

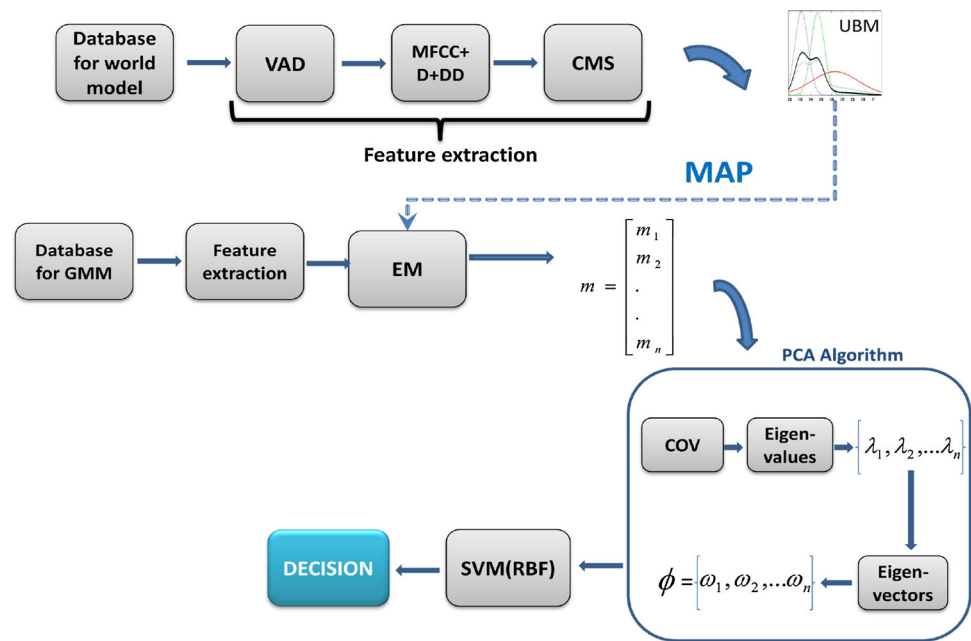
##### 4.1 System Overview

The bloc diagram of the proposed GMM-PCA-SVM based speaker recognition system is depicted in Fig. 5. First, a Voice Activity Detector (VAD) technique is used. For a

**Fig. 4** Bloc diagram of the PCA-SVM based speaker recognition system



**Fig. 5** The proposed GMM-PCA-SVM based speaker recognition system



given speech utterance, the energy of all speech frames is computed. An empirical threshold is then determined from the maximum energy of these speech frames. This classifies speech segments as either speech or silence segments. Finally, silent segments (no-speech) are removed.

The ASR systems use the short term spectrum features [Harrag et al. \(2011\)](#) to represent speaker specific features. Indeed, the short term spectrum features convey the glottal source, the vocal tract shape and length of a speaker, and thus lead to a better representation of a given speaker.

In this study, an extraction of 12 MFCCs, plus their delta and double delta Cepstral coefficients, making 36 dimensional feature vectors to represent the feature space. These features are extracted using a Hamming window with 20 ms of length and a shift of 10 ms. The window is used to taper the original signal on the sides and therefore reduces the side effects [Hanilci and Ertas \(2011\)](#). Finally, a Cepstral Mean Subtraction (CMS) [Kinnunen and Li \(2010\)](#) is applied to these features by the subtraction of the cepstral mean of the feature vectors in order to fit the data around their average.

#### 4.2 Modeling phase

In the proposed GMM-PCA-SVM scheme, the main idea consists of the introduction of the dimensionality reduction using PCA technique in the core of the recognizer. The proposed process is given on [Fig. 5](#).

The mean vectors issued from the UBM model using MAP adaptation are projected using PCA technique into an orthogonal feature space. The new reduced mean vectors are then used as input to the SVM model for scoring.

#### 4.3 Double dimensionality reduction

To better investigate on the contribution of PCA technique, dimensionality reduction is also applied in the front-end part of the proposed GMM-PCA-SVM system (see [Fig. 6](#)).

### 5 Experiment results

#### 5.1 Corpora

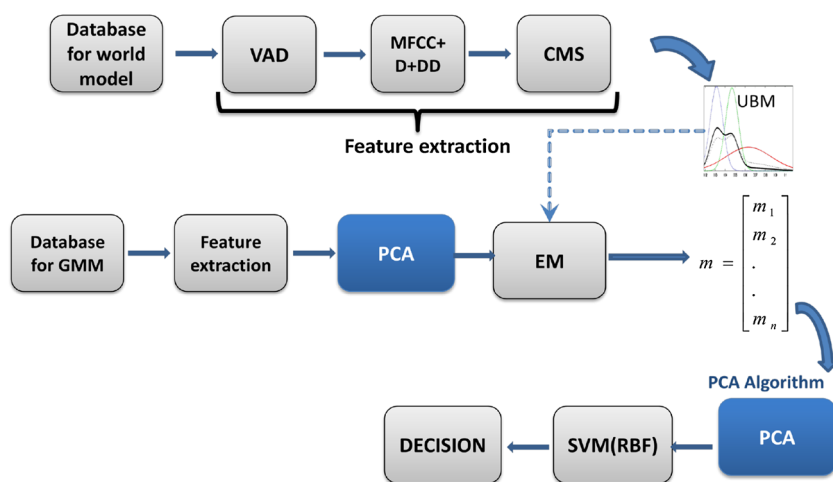
The corpus used in this work is issued from the TIMIT database [Garofolo et al. \(1993\)](#), which was one of the first corpora available that had a large number of speakers, and has been used for many speaker recognition studies. This database includes phonetic and word transcriptions as well as a 16-bit, 16 kHz speech file for each utterance and is recorded in “.ADC” format.

The database consists of a set of 8 sentences with 3s of length spoken by 491 speakers in English language and divided in 8 dialects (Dr1 to Dr8) of the United States. We have selected 5 phonetically rich sentences (SX recordings) for the training and 3 other utterances (SI sentences) different from the previous ones for the testing. In this way, the text independency of speaker recognition was preserved.

#### 5.2 Speaker recognition using SVM and GMM-SVM

To evaluate the influence of dialect and size of database on the ASR, a comparative study of the SVM and GMM-SVM systems is performed. In this study, Gaussian mixture models

**Fig. 6** Bloc diagram of the double dimensionality reduction, the PCA-GMM-PCA-SVM based speaker recognition system



**Table 1** Performance of the SVM and GMM-SVM based speaker recognition systems, in term of EER (%)

Subset	Dialect	N of Speakers	GMM-SVM (%)	SVM (%)
Dr1	New England	47	14.84	14.83
Dr6	New York City	47	14.54	16.04
Dr2	Northern	90	6.34	6.93
Dr3	North Midland	86	6.62	7.24
Dr4	South Midland	65	8.75	8.8
Dr5	Southern	65	8.45	8.71
Dr7	Western	66	8.37	8.18
Dr8	Army Brat	25	22.07	26.89

were used with  $M = 32$ . The parameter  $\alpha_i$  is calculated as in Eq. (10). For GMM-MAP training, only mean values of the Gaussian components were adapted, with a relevance factor of 16, the weight vector and the covariance matrix were not modified.

The so-called impostor model is used as an a priori for the estimation of speaker models. For this purpose, a gender balanced UBM consisting of 2048 mixture components was trained using the EM algorithm. The UBM aims to model the general acoustic space of 120 unknown speakers (impostors), 60 male and 60 female, where each speaker utters five different sequences. In a last step, an SVM classifier using the target GMM supervectors and the SVM background which represents GMM supervectors of 25 impostors labeled as (-1) for scoring is trained. Table 1 presents the results in term of EER of different dialects and different lengths of subdatabases contained in the TIMIT dataset.

Table 1 shows the EER (%) of the speaker recognition system accuracy with both SVM and GMM-SVM classifiers. As expected, in major cases, the GMM-SVM outperforms the SVM system’s performance. For example, the EER obtained with the SVM model for the Dr8 subset is equal to 26.89

% where it is less than 22.1 % for the hybrid GMM-SVM system.

Even when the three subsets of the TIMIT corpora have almost the same number of speakers, Dr4, Dr5 and Dr7 with different dialect, both GMM-SVM and SVM performance accuracies are quite the same for all these subsets. For example, for the SVM classifier, the EER in Dr4 (Dialect: *South Midland*, Number of speaker: 65) is 8.8 %, in Dr5 (Dialect is: *Southern*, Number of speaker is: 65) it is 8.71 % and in Dr7 (Dialect is: *Western*, Number of speaker is: 66) 8.18 %. But on the other hand, a difference of performance accuracy is noticed with the SVM classifier for the Dr1 and Dr6 subsets. The EER in Dr1 (Dialect: *New England*, Number of speaker: 47) is 14.83 % while it is 16.4 % for the Dr6 (Dialect is: *Southern*, Number of speaker is: 47) subset. Therefore, one cannot confirm that the dialect did have an influence on the ASR task for both systems. However, the number of speakers has a big influence on both classifiers for the speaker recognition rate. That is, the greater the number of speakers, the smaller the EER becomes. This is clearly seen with Dr8 (Dialect: *Army Brat*, Number of speaker: 25), and Dr2 (Dialect: *Northern*, Number of speaker: 90) for which the EERs are 26.89 and

**Table 2** Performance of the GMM-PCA-SVM, PCA-GMM-PCA-SVM and PCA-SVM systems, in term of EER (%)

Subset	Dialect	N of Speakers	GMM-PCA-SVM	PCA-GMM-PCA-SVM	PCA-SVM
Dr1	New England	47	5.68	12.5	7
Dr6	New York City	47	5.39	12.24	10.58
Dr2	Northern	90	4.16	6.66	4.62
Dr3	North Midland	86	3.58	6.81	3.95
Dr4	South Midland	65	4.18	8.57	4.22
Dr5	Southern	65	4.66	8.45	5.36
Dr7	Western	66	4.28	8.45	5.38
Dr8	Army Brat	25	8.19	12.90	26

6.93 % for the GMM-SVM and the SVM classifiers respectively.

### 5.3 Speaker recognition using the PCA dimensionality reduction

The main goal of the experiments described in this section is to evaluate the recognition performance of the proposed system using the PCA dimensionality reduction in the core of the classifier. Results when applying PCA in the front-end part of the ASR system are also presented in the Table 2.

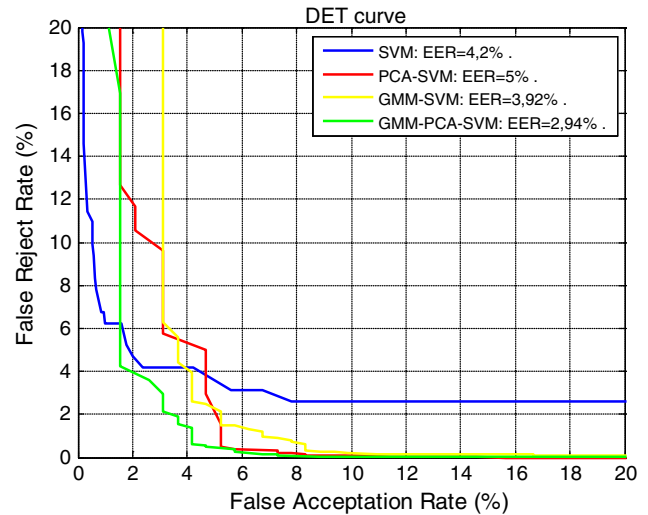
Comparing to Table 1, we can observe that using PCA dimensionality reduction leads to a notable increase in the system’s accuracy for both SVM and the hybrid GMM-SVM classifiers. It is clearly seen that, the proposed GMM-PCA-SVM system outperforms the other ones for all different subsets of the TIMIT database.

### 5.4 Speaker recognition in noisy environment

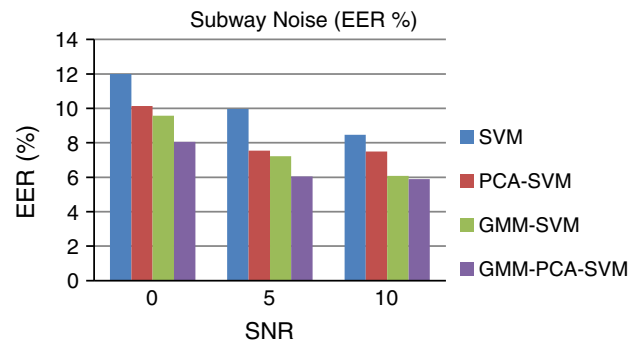
The SVM and the GMM-PCA-SVM classifiers have been performed in both clean and noisy environments. A set of 176 speakers issued from the eight subsets of the TIMIT database is used.

For real world setting, two different noisy environments, Train station and Subway noises issued from the NOISEUS database have been used within Signal-to-Noise Ratio, SNR = 0, 5 and 10 dB. The experimental protocol is the same as that one detailed previously in this paper. In clean environment, the obtained results are express by the Detection Error Tradeoff (DET) curve (See Fig. 7).

In the clean case, a low degradation is noticed when applying the PCA technique in the front-end part of the SVM classifier. For example, the EER increased from 4,2 % (SVM alone) to 5 % (PCA-SVM). For the proposed GMM-PCA-SVM model, the PCA gives an important contribution for the recognition accuracy. In fact, the EER decreases from 3,92 % for the conventional GMM-SVM based classifier to 2,94 % obtained with the proposed GMM-PCA-SVM one.



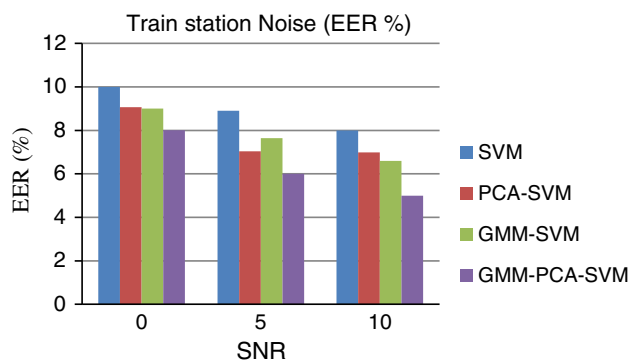
**Fig. 7** Speaker recognition in clean environment



**Fig. 8** Comparative performance of the speaker recognition systems using speech corrupted with Subway noise

Figures 8, 9 present the performance accuracy of the proposed system in different noisy environments. It is clearly seen that the proposed GMM-PCA-SVM speaker recognition system is more robust compared to the conventional SVM or GMM-SVM based speaker recognition systems.

Concerning the noisy environment, the contribution of the PCA is clearly noticed for both SVM and GMM-SVM systems. Best performances accuracies are reached with the proposed GMM-PCA-SVM system. Applying PCA in the front-



**Fig. 9** Comparative performance of the speaker recognition systems using speech corrupted with Train station noise

end part of the SVM system brings also interesting results. For instance, for Subway noise and at SNR = 0 dB, the EER is 12 % for SVM based system alone, while it is less than 10, 2 % for the PCA-SVM based system.

In the speech signal, it is expected that subsets of variables are highly correlated with each other. These variables are quite redundant and consequently share the same powerful rule in defining the outcome of interest. Consequently, the system is trained on unnecessary samples which lead to a loss of time and performance. Furthermore, when speech data is corrupted with different noises, the information within this particular data (the redundant samples/less significant samples) is totally lost and hence causes a serious degradation in system accuracy.

The basic solution is to combine, using the PCA technique, these variables into a smaller number that will account for most of the variance in the observed data. One of the principal assumptions of PCA technique is assuming that components with big variance correspond to interesting dynamics and lower ones correspond to noise. Though, the purpose of this paper consists of the use of PCA in the modeling phase of the classifier, which transforms the reduced adapted mean vectors into an orthogonal feature space and allows throwing out the low weight transformed features. This considerably enhances performances by removing correlations between variables.

## 6 Conclusion

In this paper, a new GMM-PCA-SVM scheme has been proposed for ASR. The concept, based on the dimensionality reduction, consists of applying the PCA technique to the adapted mean vectors in the modeling phase of the GMM-SVM based speaker recognition system. Comparative study proven that, this new scheme brings interesting results in both clean and noisy environment.

In addition, dimensionality reduction was also applied to both front-end stage and speaker modeling core, but in this last case, the overall reduction method was not more effective due to the huge loss of data caused by the repeated reduction.

Moreover, the results show that the dialect did not have a visible effect on the system's performances. However, the size of the database (number of speakers) affected strongly the performance accuracy of both classifiers.

For future work, additional features, such as prosodic and voice quality features can be merged with the proposed method to ameliorate the speaker recognition performance accuracy.

## References

- Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2), 1–47.
- Campbell, W., Sturim, D., Reynolds, D. A., & Solomonoff, A. (2006). SVM based speaker verification using a GMM supervector kernel and Nap variability compensation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toulouse, France (pp. 97–100).
- Campbell, J. P., Jr. (1997). Speaker recognition: a tutorial. In *Pro. IEEE*, 85(9), 1437–1462.
- Chitturi, R., & Hansen, J. H. L. (2007). Multi-stream dialect classification using SVM-GMM hybrid classifiers. In *IEEE Workshop on Automatic Speech Recognition & Understanding*, Kyoto, Japan (pp. 431–436).
- Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., Dahlgren, N., et al. (1993). *TIMIT acoustic-phonetic continuous speech corpus*. Philadelphia: Linguistic Data Consortium.
- Hanilci, C., & Ertas, F. (2011). VQ-UBM based speaker verification through dimension reduction using local PCA. In *19th European Signal Processing conference*, Spain (pp. 1303–1306).
- Harrag, A., Mohamadi, T., & Harrag N. (2011). LDA fusing of acoustic and prosodic features: application to speaker recognition. In *Colloquium on Humanities, Science and Engineering Research*, Penang (pp. 245–248).
- Izquierdo-Verdiguier, E., Gomez-Chova, L., Bruzzone, L., & Camps-Valls, G. (2014). Semisupervised kernel feature extraction for remote sensing image analysis. *IEEE transactions on geoscience and remote sensing*, PP(99), 1–12.
- Jiang, J., Wu, Z., Xu, M., Jia, J., & Cai, L. (2013). Comparing feature dimension reduction algorithms for GMM-SVM based speech emotion recognition. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Kaohsiung, China (pp. 1–4).
- Jokic, I., Jokic, S., Gnjatovic, M., Delic, V., & Peric, Z. (2012). Influence of the number of principal components used to the automatic speaker recognition accuracy. *Electronics & Electrical Engineering*, 123, 83–86.
- Jolliffe, I. T. (2010). *Principal component analysis* (2nd ed.). New York, NY: Springer-Verlag.
- Karam, Z. N., & Campbell, W. M. (2008). A multi-class MLLR kernel for SVM speaker recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, USA (pp. 4117–4120).
- Kinnunen, T., & Li, H. (2010). An overview of text-independent speaker recognition: from features to supervectors. *Speech Communication*, 52, 12–40.



- Kresimir delac, K., Grgic, M., & liatsis, P. (2005). Appearance based statistical methods for face recognition. In 47th international symposium EL-MAR, Zadar, Croatia (pp. 151–158).
- Kuncheva, L. I., & Faithfull, W. J. (2014). PCA feature extraction for change detection in multidimensional unlabeled data. *IEEE transactions on neural networks and learning systems*, 25(1), 69–80.
- Lee, K. Y. (2004). Local fuzzy PCA based GMM with dimension reduction on speaker identification. *Pattern Recognition Letters*, 25, 1811–1817.
- Li, H., & Dong, Y. (2013). EigenVoice used in speaker recognition with a few training samples. *Advanced Materials Research*, 823, 618–621.
- Malarvizhi, A., & Sivasarathadevi, K. (2013). Performance analysis of HDM and PCA, ICA In teeth image recognition, In Proceedings of International Conference on Optical Imaging Sensor and Security, Coimbatore, India (pp. 1–5).
- Minkyung, K., Eunyong, K., Changwoo, S., & Sungchae, J. (2010). Speaker verification and identification using principal component analysis based on global eigenvector matrix. *Hybrid Artificial Intelligence Systems*, 6076, 278–285.
- Reynolds, D. A., Quatieri, T. F., & Dunn, R. B. (2000). Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, 10(1–3), 19–41.
- Vapnik, V. (1998). *Statistical learning theory*. New York: John Wiley.
- Wan, V., & Renals, S. (2003). SVMSVM: support vector machine speaker verification methodology. In IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings (ICASSP), Hong Kong, China (pp. 221–224).
- Yun, L., & Hansen, J.H.L. (2009). Factor analysis-based information integration for Arabic dialect identification. In IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan (pp. 4337–4340).
- Yun, L., & Hansen, J. H. L. (2011). Dialect classification via text-independent training and testing for Arabic, Spanish, and Chinese. *Audio, Speech, and Language Processing, IEEE Transactions on Biometrics Compendium*, 19(1), 85–96.
- Zhang, C., & Zheng, T.F. (2013). A fisher voice based feature fusion method for short utterance speaker recognition. In 2013 IEEE China Summit and International Conference on Signal and Information Processing, Beijing, China (pp. 165–169).