Research Article
A Novel GMM-Based Feature Reduction for Vocal Fold Pathology Diagnosis

Vahid Majidnezhad and Igor Kheidorov

Department of Computer Engineering, Shabestar Branch, Islamic Azad University, Shabestar, Iran

Abstract: Acoustic analysis is a proper method in vocal fold pathology diagnosis so that it can complement and in some cases replace the other invasive, based on direct vocal fold observation, methods. There are different approaches and algorithms for vocal fold pathology diagnosis. These algorithms usually have three stages which are Feature Extraction, Feature Reduction and Classification. While the third stage implies a choice of a variety of machine learning methods (Support Vector Machines, Gaussian Mixture Model, etc.), the first and second stages play a critical role in performance and accuracy of the classification system. In this study we present initial study of feature extraction and feature reduction in the task of vocal fold pathology diagnosis. A new type of feature vector, based on wavelet packet decomposition and Mel-Frequency-Cepstral-Coefficients (MFCCs), is proposed. Also a new method for feature reduction is proposed and compared with conventional methods such as Principal Component Analysis (PCA), F-Ratio and Fisher's discriminant ratio. Gaussian Mixture Model is used as a classifier for evaluating the performance of the proposed method. The results show the priority of the proposed method in comparison with current methods.

Keywords: Gaussian Mixture Model (GMM), Mel-Frequency-Cepstral-Coefficient (MFCC), Principal Component Analysis (PCA), proposed GMM-based feature reduction, vocal fold pathology diagnosis, wavelet packet decomposition

INTRODUCTION

Vocal signal information often plays an important role for specialists to understand the process of vocal fold pathology formation. In some cases vocal signal analysis can be the only way to analyze the state of vocal folds. Nowadays diverse medical techniques exist for direct examination and diagnostics of pathologies. Laryngoscopy, glottography, stroboscopy, electromyography, videokimography are most frequently used by medical specialists. But these methods possess a number of disadvantages. Human vocal tract is hardly-accessible for visual examination during phonation process and that makes it more problematic to identify pathology. Moreover, these diagnostic means may cause patients much discomfort and distort the actual signal that may lead to incorrect diagnosis as well (Alonso *et al.*, 2001; Ceballos *et al.*, 2005, 1996; Adnene and Lamia, 2003).

Acoustic analysis as a diagnostic method has no drawbacks, peculiar to the above mentioned methods. It possesses a number of advantages. First of all, acoustic analysis is a non-invasive diagnostic technique that allows pathologists to examine many people in short time period with minimal discomfort. It also allows pathologists to reveal the pathologies on early stages of their origin. This method can be of great interest for medical institutions.
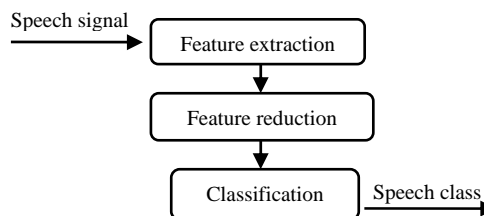

Fig. 1: The general scheme of vocal fold pathology diagnosis

In recent years a number of methods were developed for segmentation and classification of speech signals with pathology. The general scheme of vocal fold pathology diagnosis is illustrated in Fig. 1.

Different parameters for feature extraction are used. Traditionally, one deals with such parameters like pitch, jitter, shimmer, amplitude perturbation, pitch perturbation, signal to noise ratio, normalized noise energy (Manfredi, 2000) and others (Llorente and Vilda, 2004; Rosa *et al.*, 2000; Mallat, 1989; Wallen and Hansen, 1996). Feature extraction, using the above mentioned parameters, has shown its efficiency for a number of practical tasks. These parameters are frequently used in systems for automatic vocal fold pathology diagnosis, in speaker identification systems or in multimedia database indexing systems. In the proposed method, we have used the Mel-Frequency-

Corresponding Author: Vahid Majidnezhad, Department of Computer Engineering, Shabestar Branch, Islamic Azad University, Shabestar, Iran

Table 1: Summary of some pervious works

Reference	Feature set	Feature reduction approach	Classifier
Chen *et al*. (2007)	25 acoustic parameters given by MDVP	PCA	Support vector machine
Go´mez *et al*. (2005)	Spectral perturbation	PCA	K-means clustering
Michaelis *et al*. (1998)	Acoustic feature, noise	PCA	Threshold
Marinaki *et al*. (2004)	Linear prediction coefficients	PCA	K-nearest neighbors
Llorente *et al*. (2006)	Mel-frequency-cepstral-coefficients	F-ratio and Fisher's discriminant ratio	Gaussian mixture model
Ritchings *et al*. (2002)	Spectral	-	Artificial neural network

Cepstral-Coefficients (MFCCs), Energy and Shannon Entropy parameters for creating the initial features vector.

Also different approaches for feature reduction are used such as Principal Component Analysis (PCA) (Chen *et al.*, 2007; Go´ mez *et al.*, 2005; Michaelis *et al.*, 1998; Marinaki *et al.*, 2004) and Fisher's Discriminant Ratio (Llorente *et al.*, 2006). In the proposed method, we have used the proposed GMM-Based feature reduction and we have shown its priority in comparison with the PCA and Fisher's discriminant ratio.

Finally, the reduced features are used for speech classification into the healthy and pathological class. Different machine learning methods such as Support Vector Machines (Chen *et al.*, 2007), Artificial Neural Networks (Ritchings *et al.*, 2002), etc., can be used as a classifier. In the proposed method we have used Gaussian Mixture Model (GMM) for classification purpose. In Table 1, some pervious methods in the vocal fold pathology diagnosis are summarized.

In this study, first, the importance of automatic methods of detection of vocal fold pathology is investigated. Also, the stages of these methods are described. Then by focusing on their second stage, which is feature reduction, a novel GMM-Based approach is proposed. In other words, the main objective of this study is to propose an efficient approach to reduce features vector in the field of vocal fold pathology diagnosis. The results of the experiments show better performance of the proposed GMM-Based method in comparison with a well-known method (PCA).

METHODOLOGY

The presence of pathology in a vocal tract inevitably leads to voice signal distortion. Depending on pathology severity the distortion may be more or less significant. Among all sounds that are produced by vocal tract, sustained vowels and some sonorant consonants are most easily distorted if pathology is present.

The wavelet transform, as was shown in Manfredi (2000), is a flexible tool for time-frequency analysis of speech signals, especially for short data frames, like separate phonemes. In Fig. 2 wavelet transform of a stressed vowel [a:], pronounced by a healthy speaker, is shown.

But the situation changes in case of pathological voices. In Fig. 3, 4 and 5 wavelet transforms of the same vowel are given, but in these cases it is pronounced by speakers with different voice pathologies. The instability of the formant frequency is obviously seen.

This led us to supposition that feature vectors based on wavelets can show good results. The idea to build feature vector on wavelets for audio classification was previously reported by Li *et al.* (2003) and Tzanetakis and Cook (2002). These authors used the Discrete Wavelet Transform (DWT) coefficients for their method of feature extraction for content-based audio classification. Kukharchik *et al.* (2007) used Continues Wavelet Transform (CWT) coefficients for their method of feature extraction. Cavalcanti *et al.* (2010) used Wavelet Packet Decomposition (WPD) nodes coefficients for their method for feature extraction. In this study we have also used the wavelet packet decomposition to create the wavelet packet tree and to extract the features.

The block diagram of our proposed method is illustrated in Fig. 6. In the first stage, by the use of MFCC and Wavelet Packet Decomposition, feature vector containing 139 features is made. In the second stage, by the use of the GMM-Based feature reduction method, the dimension of feature vector is decreased. In the last stage, by the use of GMM, the speech signal classified into two classes: pathological or healthy.

Feature extraction: As it is shown in Fig. 6, first, by the use of cepstral representation of input signal, 13 Mel-Frequency-Cepstral-Coefficients (MFCC) are extracted. Then the wavelet packet decomposition in 5 levels is applied on the input signal to make the wavelet packet tree. Then, from the nodes of resulting wavelet packet tree, 63 energy features along with 63 shannon entropy features are extracted. Finally, by the combination of these features, the initial feature vector with the length of 139 features is created.

Mel-Frequency-Cepstral-Coefficients (MFCCs): MFCCs are widely used features to characterize a voice signal and can be estimated by using a parametric approach derived from Linear Prediction Coefficients (LPC), or by the non-parametric discrete Fast Fourier Transform (FFT), which typically encodes more information than the LPC method. The signal is windowed with a hamming window in the time domain
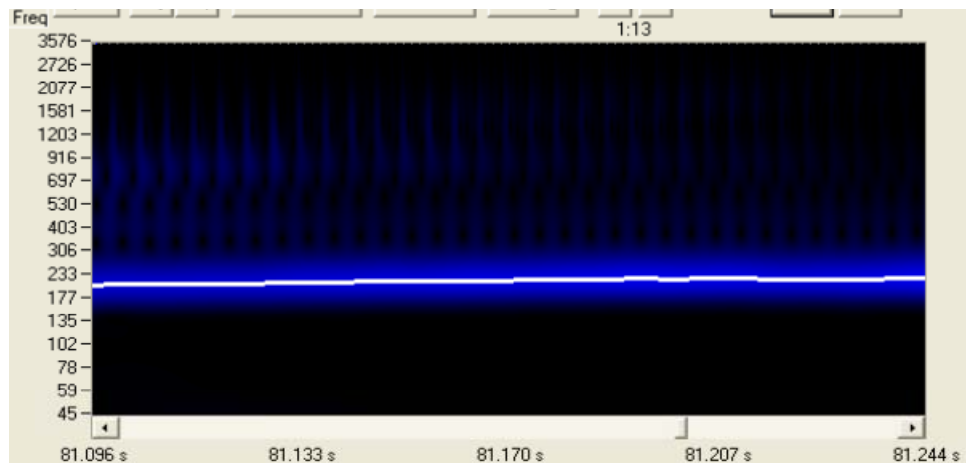
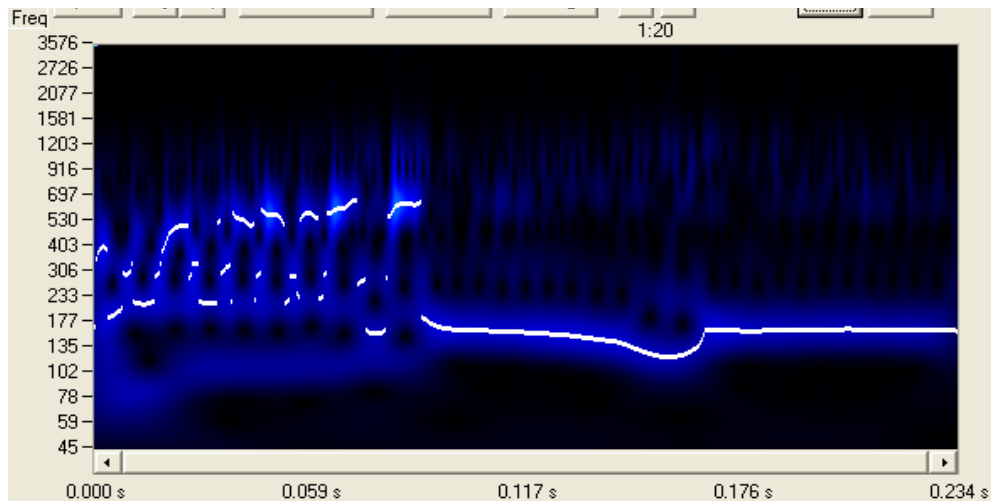Fig. 2: Wavelet transform of a stressed vowel [a:] pronounced by a healthy speaker

Fig. 3: Wavelet transform of a stressed vowel [a:] pronounced by the speaker with hypertrophic laryngitis
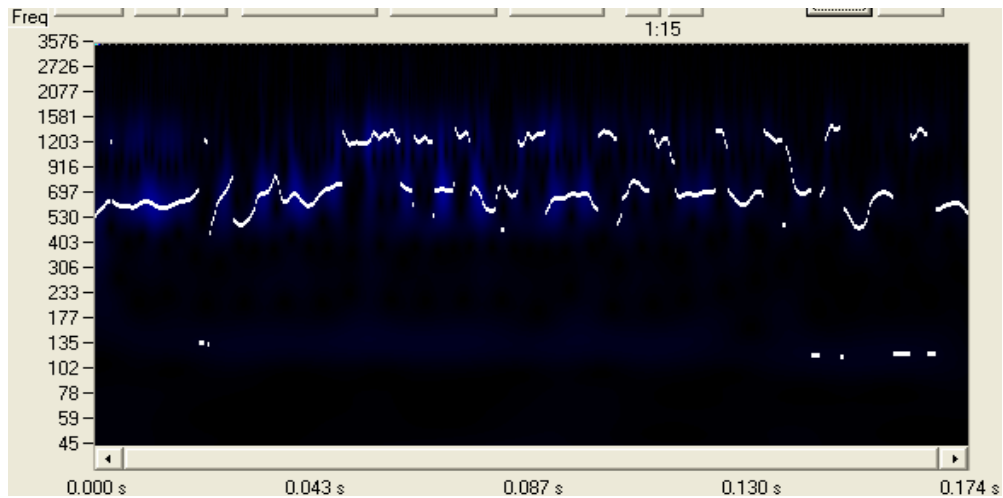

Fig. 4: Wavelet transform of a stressed vowel [a:] pronounced by the speaker with hypertonic dysphonia
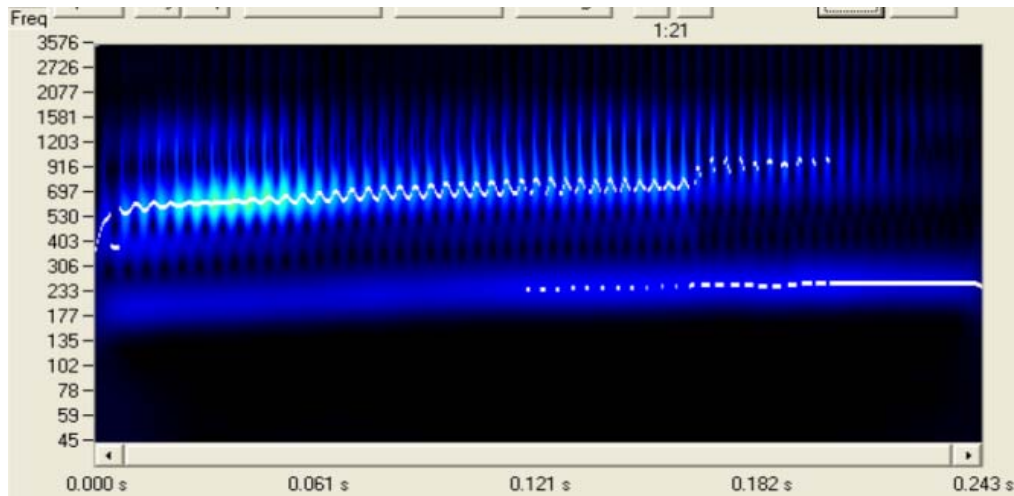
Fig. 5: Wavelet transform of a stressed vowel [a:] pronounced by the speaker with chronic catarrhal laryngitis
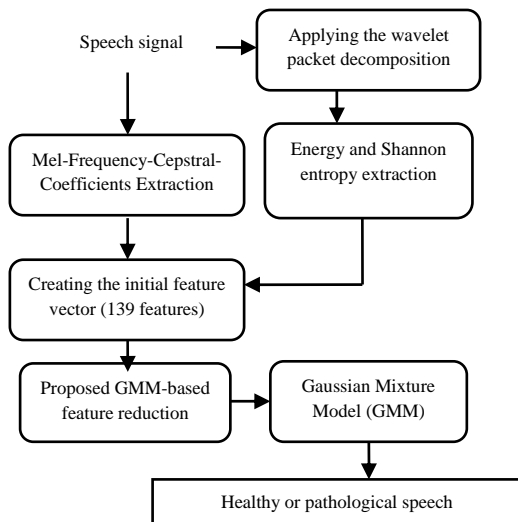

Fig. 6: The block diagram of the proposed method for detection of vocal fold pathology

and converted into the frequency domain by FFT, which gives the magnitude of the FFT. Then the FFT data is converted into filter bank outputs and the cosine transform is found to reduce dimensionality. The filter bank is constructed using 13 linearly-spaced filters (133.33Hz between center frequencies) followed by 27 log-spaced filters (separated by a factor of 1.0711703 in frequency.) Each filter is constructed by combining the amplitude of FFT bin. The MATLAB code to calculate the MFCC features was adapted from the Auditory Toolbox (Malcolm Slaney). The MFCCs are used as features in Llorente *et al.* (2006) to classify the speech into pathology and healthy class. We have used reduction of MFCC information by averaging the sample's value of each coefficient.

Wavelet packet decomposition: Recently, Wavelet Packets (WPs) have been widely used by many researchers to analyze voice and speech signals. There are many out-standing properties of wavelet packets which encourage researchers to employ them in widespread fields. The most important, multi resolution property of WPs is helpful in voice signal synthesis (Herisa *et al.*, 2009; Fonseca *et al.*, 2007).

The hierarchical WP transform uses a family of wavelet functions and their associated scaling functions to decompose the original signal into subsequent sub-bands. The decomposition process is recursively applied to both the low and high frequency sub-bands to generate the next level of the hierarchy. WPs can be described by the following collection of basic functions:

$$W_{2n}(2^{p-1}x-1)=\sqrt{2^{1-p}}\sum_{m}h(m-2l)\sqrt{2^p}W_n(2^px-m) \qquad (1)$$

$$W_{2n+1}(2^{p-1}x-1)=\sqrt{2^{1-p}}\sum_{m}g(m-2l)\sqrt{2^p}W_n(2^px-m) \qquad (2)$$

where
p : Scale index
l : The translation index
h : The low-pass filter
g : The high-pass filter with:

$$g(k)=(-1)^k h(1-k) \qquad (3)$$

The WP coefficients at different scales and positions of a discrete signal can be computed as follows:

$$C_{n,k}^p = \sqrt{2^p} \sum_{m=-\infty}^{\infty} f(m)W_n(2^p m - k) \qquad (4)$$

$$C_{2n,l}^{p-1} = \sum_m h(m - 2l) C_{n,m}^p \tag{5}$$

$$C_{2n+1,l}^{p-1} = \sum_m g(m - 2l) C_{n,m}^p \tag{6}$$

For a group of wavelet packet coefficients, energy feature in its corresponding sub-band is computed as:

$$Energy_n = \frac{1}{N^2} \sum_{k=1}^n \left| C_{n,k}^p \right|^2 \tag{7}$$

The entropy evaluates the rate of information which is produced by the pathogens factors as a measure of abnormality in pathological speech. Also, the measure of Shannon entropy can be computed using the extracted wavelet-packet coefficients, through the following formula:

$$Entropy_n = -\sum_{k=1}^n \left| C_{n,k}^p \right|^2 \log \left| C_{n,k}^p \right|^2 \tag{8}$$

In this study, mother wavelet function of the tenth order Daubechies has been chosen and the signals have been decomposed to five levels. The mother wavelet used in this study is reported to be effective in voice signal analysis (Guido *et al.*, 2005; Umapathy and Karishnan, 2005) and is being widely used in many pathological voice analyses (Fonseca *et al.*, 2007). Due to the noise-like effect of irregularities in the vibration pattern of damaged vocal folds, the distribution manner of such variations within the whole frequency range of pathological speech signals is not clearly known. Therefore, it seems reasonable to use WP rather than DWT or CWT to have more detail sub-bands.

Feature reduction: Using every feature for classification process is not good idea and it may be causes to the increasing the rate of misclassification. Therefore, it is better to choose the proper features from the whole features. This process is called as "Feature Reduction".

The goal is to reduce the dimension of the data by finding a small set of important features which can give good classification performance. Feature reduction algorithms can be roughly grouped into two categories: filter methods and wrapper methods. Filter methods rely on general characteristics of the data to evaluate and to select the feature subsets without involving the chosen learning algorithm. Wrapper methods use the performance of the chosen learning algorithm to evaluate each candidate feature subset. Wrapper methods search for features better fit for the chosen learning algorithm, but they can be significantly slower than filter methods if the learning algorithm takes a long time to run. The concepts of "filters" and "wrappers" are described in Kohavi and Jhon (1997).

One way for feature reduction is Principal Component Analysis (PCA) which is used frequently in pervious works such as Chen *et al.* (2007), Go´mez *et al.* (2005), Michaelis *et al.* (1998) and Marinaki *et al.* (2004). PCA is a well-known filter method. Another way for feature reduction is Fisher's Discriminant Ratio which is used in pervious works such as Llorente *et al.* (2006). It's also a filter method. In this section we also propose a novel approach, the Proposed GMM-Based Feature Reduction, for the feature reduction which belongs to wrapper methods.

Fisher's discriminate ratio: The ratio (Llorente *et al.*, 2006) represents the relationship between within-class and inter-class variances under the same assumptions as the F-Ratio. The following assumptions must be enforced:

- The feature vectors within each class must have Gaussian distribution
- Features should be uncorrelated
- The variances within each class must be equal

Given a set of classes w_k, $k = 1, 2, \ldots, K$, twice scatter measurements can be defined as follows:

Within-class scatter (S_w): is a measurement of the scattering of the samples that belong to a class w_k around their respective means:

$$S_w = \frac{1}{k} \sum_{k=1}^K E\{(x - \mu_k)(x - \mu_k)^t\}, x \in w_k \tag{9}$$

Interclass scatter (S_b): measures the scattering of each class mean around the overall mean:

$$S_b = \frac{1}{k} \sum_{k=1}^K E\{(\mu_k - \mu_0)(\mu_k - \mu_0)^t\} \tag{10}$$

where,
μ_k : The mean of class w_k
μ_0 : The mean value of the whole dataset without considering the class segmentation

There exist several ways to quantify the discriminative power. The interclass separation measurement can be calculated comparing the relationship between within-class and inter-class scattering. The following is a standard computation:

$$J = \left| S_w^{-1} S_b \right| \tag{11}$$

A feature vector is said to be optimum if the inter-class separation is maximized. If computing these measurements is carried out for every single feature alone, such measurements are known as Fisher's discriminant ratio F_i. The higher the value of F_i, the more important the feature is: this means that feature has a low variance with respect to inter-class variance and this is the reason why it is desirable to discriminate between them. These criteria adopt a special form in the one-dimensional, two-class problem, quantifying the reparability capabilities of individual features:

$$F_i = \frac{(\mu_{ic} - \mu_{i\bar{c}})^2}{\sigma_{ic}^2 - \sigma_{i\bar{c}}^2} \tag{12}$$

where, sub-index C corresponds to normal voices and $\bar{C}$ to pathological.

Principal component analysis: This method searches a mapping to find the best representation for distribution of data. Therefore, it uses a signal-representation criterion to perform dimension reduction while preserving much of the randomness or variance in the high-dimensional space as possible (Arjmandi and Pooyan, 2012). The first principal component accounts for as much of the variability in the data as possible and each succeeding component accounts for as much of the remaining variability as possible. PCA involves the calculation of the eigenvalues decomposition of a data covariance matrix or singular value decomposition of a data matrix, usually after mean centering the data for each attribute. PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate, called the first principal component, the second greatest variance on the second coordinate and so on. The principal component W_1 of a dataset X can be defined as:

$$W_1 = \arg_{\|w\|=1} \max \, \text{var}\{W^T X\} = \\ \arg_{\|w\|=1} \max \, E\{(W^T X)^2\} \tag{13}$$

with the first $K\text{-}1$ component, the Kth component can be found by subtracting the first $K\text{-}1$ principal components from X:

$$\hat{X}_{k-1} = X - \sum_{i=1}^{k-1} W_i W_i^T X \tag{14}$$

and by substituting this as the new data set to find a principal component in:

$$W_k = \arg_{\|w\|=1} \max E\{(W^T \hat{X}_{k-1})^2\} \tag{15}$$

The karhunen Leove transform is therefore equivalent to finding the singular value decomposition of the data matrix, X:

$$X = W \sum V^T \tag{16}$$

and then obtaining the reduced-space data matrix by Y projecting X down into the reduced space defined by only the first L singular vectors, W_L:

$$Y = W_L^T X = \sum V_L^T \tag{17}$$

The matrix W of singular vectors of X is equivalently the matrix W of eigenvectors of the matrix of observed co-variances, $C = XX^T$:

$$XX^T = W \sum \sum W^T \tag{18}$$

In PCA, the optimal approximation of a random vector $X \in R^N$ in N-dimensional space by a linear combination of M $(M<N)$ independent vectors is obtained by projecting the random vector X into the eigenvectors corresponding to the largest eigenvalues of the covariance matrix of vector X (Arjmandi and Pooyan, 2012). The main limitation of PCA is that it does not consider class separately, since it does not take into account the class label of the feature vectors.

The proposed GMM-based feature reduction: As it is mentioned before, the main limitation of PCA is that it does not take into account the class labels and it just focus on the sample's value. In other words, the PCA searches for the features which their sample's value have bigger variance in comparison with others and it does not collaborate with the classifier. So, we have proposed an approach, the GMM-Based Feature Reduction, to overcome this disadvantage. For this purpose, the sample's value are divided into "train" and "test" groups and fed to the GMM classifier. Also, a distance criterion is defined which shows the distance between the results of classifier and the real classes of speeches. An empty result's vector also is defined. The aim of the proposed method is to add the specified number of features from initial feature vector into the empty result's vector so that the distance criterion is minimized. Of course this process is repeated and tested for possible situations till the finding desirable solution. The formula (19) has declared the distance criterion. The aim of the proposed method is the finding the subset of features so that they minimize the *dist*. The a_i is the result of classifier and the r_i is the real class for i^{th} speech signal. The n is the number of speech files in the "train" group:

$$dist = \sum_{i=1}^{n} |a_i - r_i| \qquad (19)$$

Gaussian mixture model: Let $x \in R^n$ be a random vector that has an arbitrary distribution. The distribution density of *x* is modeled as a Gaussian mixture density, a mixture of Q component densities, given by Llorente *et al.* (2006):

$$p(\frac{x}{\lambda}) = \sum_{i=1}^{Q} c_i . p_i(x), \sum_{i=1}^{Q} c_i = 1, c_i \geq 0 \qquad (20)$$

where, $p_i(x)$, $i = 1,...., Q$ are the component densities and c_i, $i = 1,...., Q$ are the component weights. Each component density is an n-variate Gaussian function of the form:

$$p_i(x) = \frac{1}{(2\pi)^{n/2} |C_i|^{1/2}}$$

$$\exp[-\frac{1}{2}(x - \mu_i)^T C_i^{-1}(x - \mu_i)] \qquad (21)$$

with μ_i the *n*1* mean vector and C_i the *n*n* covariance matrix.

The main motive for using the GMM as a representation of the acoustic space is that it has been demonstrated that a linear combination of Gaussian basis functions has a capacity to represent a large class of sample distributions (Llorente *et al.*, 2006).

In the proposed method, two GMMs are trained for healthy and pathological speeches. For classifying a speech, each GMM calculates the likelihood of that speech. Then the GMM with greater likelihood identifies the class of that speech which is healthy or pathological.

EXPERIMENTS AND RESULTS

In this section, three experiments have been designed. These experiments are simulated in MATLAB 7.11.0. The whole scheme of the proposed method is illustrated in Fig. 6. We have adopted a cross-validation scheme (Duda *et al.*, 2000) to assess the generalization capabilities of the system in our experiments.

Database description: The database was created by specialists from the Belarusian Republican Center of Speech, Voice and Hearing Pathologies. We have selected 40 pathological speeches and 40 healthy speeches randomly which are related to sustained vowel "a". All the records are in PCM format, 16 bits, mono, with 16 kHz sampling frequency.

Results: In first experiment, we apply the t-test on each feature and compare p-value for each feature as a measure of how effective it is at separating groups. The result is shown in Fig. 7. There are about 28% of features having p-values close to zero and 50% of features having p-values smaller than 0.05, meaning there is about 70 features among the original 139 features that have strong discrimination power. One can sort these features according to their p-values (or the absolute values of the t-statistic) and select some features from the sorted list. However, it is usually difficult to decide how many features are needed unless one has some domain knowledge or the maximum number of features that can be considered has been dictated in advance based on outside constraints.

One quick way to decide the number of needed features is to plot the MCE (Misclassification Error, i.e., the number of misclassified observations divided by the number of observations) on the test set as a function of the number of features.

In second experiment, we have applied PCA approach for feature reduction. Since the total number of our observations is 80, so it is better to use the lower number of features for our classification's purpose. Therefore, the MCE has computed for various numbers of features between 1 and 20. The obtained results by means of GMM with 1, 2, 3 and 4 mixture are shown in Fig. 8. In order to reasonably estimate the performance of the selected approach, it is important to use the 50 training samples to fit the PCA-Based approach and compute the MCE on the remaining 30 test observations (blue circular marks in the Fig. 8. It is illustrated why resubstitution error is not a good error estimate of the test error; we also show the resubstitution MCE using red triangular marks.

In third experiments, we have applied the proposed GMM-Based method for feature reduction in order to compare it with the PCA-Based approach. The MCE has computed for various numbers of features between 1 and 20. The obtained results by means of GMM with
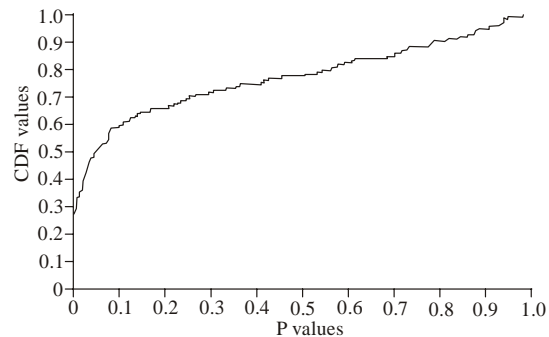

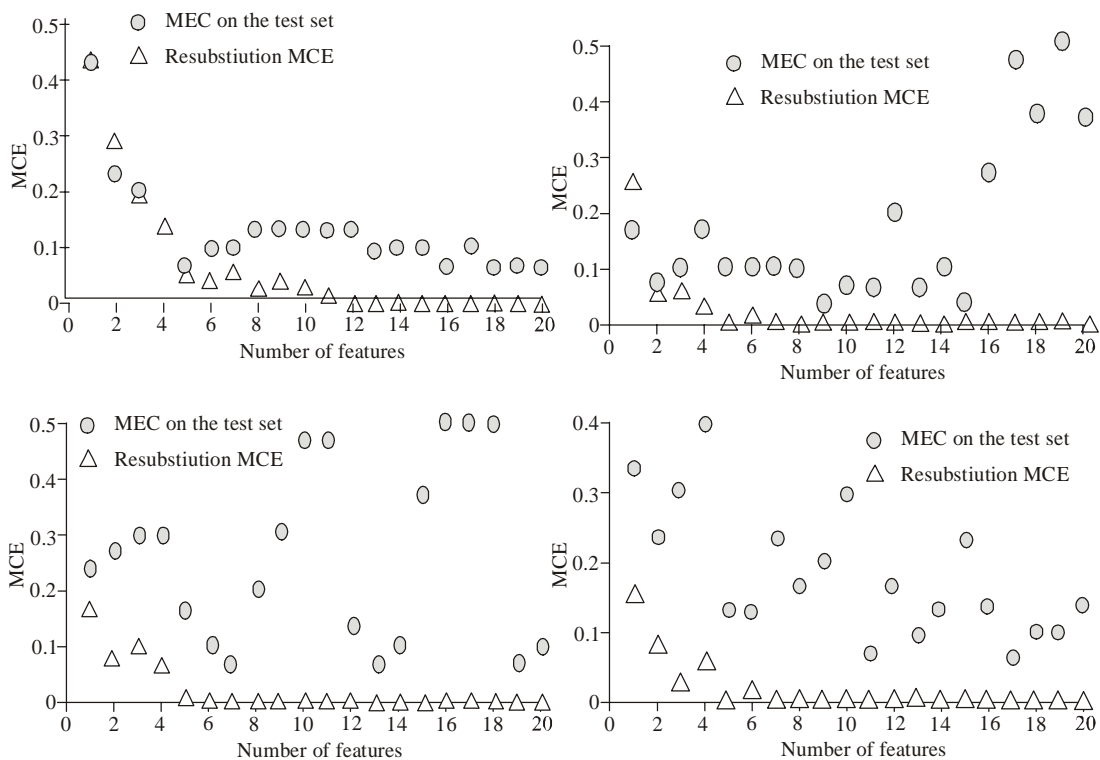Fig. 7: The p-value for features

Fig. 8: The obtained MCE by means of the PCA-based feature reduction method and different numbers of mixture for GMM
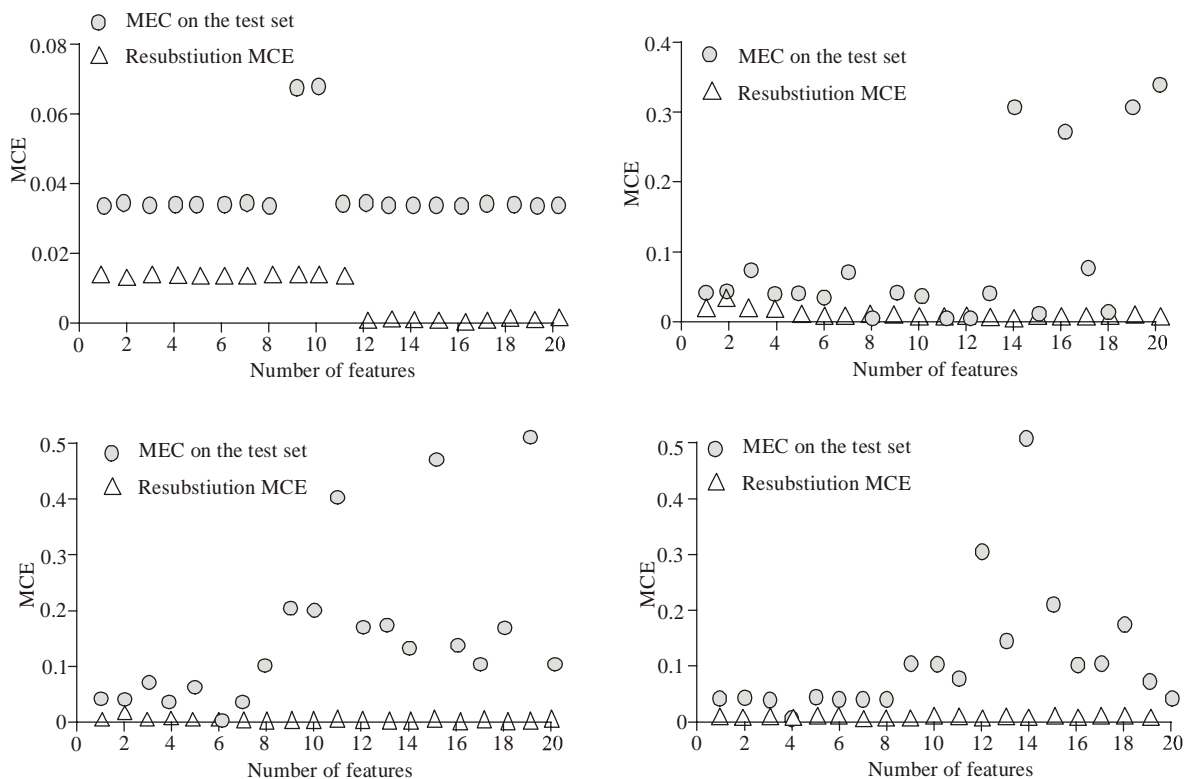

Fig. 9: The obtained MCE by means of the proposed GMM-based feature reduction method and different numbers of mixture for GMM

Table 2: The selected features and accuracy in the best case

Feature reduction method	Classifier	The selected features	Accuracy
PCA	GMM with 2 mixtures	The 1st and 4th coefficients of MFCC Energy at the 17th, 18th, 32nd and 35th nodes of WP tree Entropy at the 1st, 4th and 34th nodes of WP tree	96.67
The proposed GMM-based	GMM with 4 mixtures	The 4th coefficient of MFCC Energy at the 16th and 19th nodes of WP tree Entropy at the 2nd node of WP tree	100%
Fisher discriminant ratio (Llorente *et al.*, 2006)	GMM with 6 mixtures	The first 24th MFCC coefficients	97.35

1, 2, 3 and 4 mixture are shown in Fig. 9. In order to reasonably estimate the performance of the selected approach, it is important to use the 50 training samples to fit the GMM-Based approach and compute the MCE on the remaining 30 test observations (blue circular marks in the Fig. 9. It is illustrated why resubstitution error is not a good error estimate of the test error; we also show the resubstitution MCE using triangular marks.

As it is obvious in Fig. 8 and 9, from the MCE point of view, the performance of the proposed GMM-Based approach is better than the PCA-Based approach. Also, the best performance of the proposed GMM-Based approach by means of GMM with 4 mixtures and 4 features is better than the best performance of PCA-Based approach by means of GMM with 2 mixtures and 9 features. The selected features and accuracy, in the best case, by means of different approaches are shown in Table 2.

CONCLUSION

In this study, it is shown that features based on wavelet transformation have potential for detection of vocal fold. So, in the proposed approach, Mel-Frequency-Cepstral-Coefficients (MFCC) along with the wavelet packet decomposition is used for feature extraction phase.

Also a novel approach for the feature reduction phase in the vocal fold pathology diagnosis is proposed. Three experiments are designed to investigate the efficiency of the proposed GMM-Based method. The results of experiments show the priority of the proposed GMM-Based method in comparison with the conventional PCA-Based method and Fisher Discriminant Ratio.

In this study, the GMM is used as the classifier. One of the main advantages of GMM is its ability to classify results correctly even when classes are similar. This methodology requires a shorter time for training than other approaches such Multilayer Perceptron (MLP) or Learning Vector Quantization (LVQ). Furthermore, the GMM approach displays comparable accuracy with respect to LVQ or MLP.

Although it may be possible to try to build a complete multi-class classification system with a hierarchy of support vector machines so that detection of different type of pathological speech will be possible. For this propose, we suppose that further research for more sophisticated feature extraction phase.

REFERENCES

Adnene, C. and B. Lamia, 2003. Analysis of pathological voices by speech processing. Signal Proc. Appl., 1(1): 365-367.

Alonso, J.B., J.D. Leon, I. Alonso and M.A. Ferrer, 2001. Automatic detection of pathologies in the voice by HOS based parameters. EURASIP. J. Appl. Signal Proc., 2001(4): 275-284.

Arjmandi, M.K. and M. Pooyan, 2012. An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine. Biomed. Signal Process., 7(1): 3-19.

Cavalcanti, N., S. Silva, A. Bresolin, H. Bezerra and A. Guerreiro, 2010. Comparative analysis between wavelets for the identification of pathological voices. Proceedings of the 15th Iberoamerican Congress Conference on Progress in Pattern Recognition, Image Analysis, Computer Vision and Applications, Springer-Verlag Berlin, Heidelberg, pp: 236-243.

Ceballos, L.G., J. Hansen and J. Kaiser, 1996. Vocal fold pathology assessment using AM autocorrelation analysis of the teager energy operator: ICSLP-1996. Proceeding of 4th International Conference on ICSLP, Philadelphia, PA, pp: 757-760.

Ceballos, L.G., J. Hansen and J. Kaiser, 2005. A Non-Linear based speech feature analysis method with application to vocal fold pathology assessment. IEEE T. Biomed. Eng., 45(3): 300-313.

Chen, W., C. Peng, X. Zhu, B. Wan and D. Wei, 2007. SVM-based identification of pathological voices: Engineering in Medicine and Biology Society, 2007, EMBS 2007. Proceedings of the 29th Annual International Conference of the IEEE, Lyon, pp: 3786-3789.

Duda, R.O., P.E. Hart and D.G. Stork, 2000. Pattern Classification. 2nd Edn., Wiley Interscience. ISBN-10: 0471056693, ISBN-13: 978-0471056690.

Fonseca, E., R.C. Guido, J.C. Pereira, P.R. Scalassarsa, C.D. Maciel and J.C. Pereira, 2007. Wavelet time frequency analysis and least squares support vector machines for identification of voice disorders. Comput. Biol. Med., 37(4): 571-578.

Go´mez, P., F. Dı´az, A. A´lvarez, K. Murphy, C. Lazaro, R. Martinez and V. Rodellar, 2005. Principal component analysis of spectral perturbation parameters for voice pathology detection. Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems, ISBN: 0-7695-2355-2, pp: 41-46.

Guido, R.C., J.C. Pereira, E. Fonseca, F.L. Sanchez and L.S. Vierira, 2005. Trying different wavelets on the search for voice disorders sorting. Proceedings of the 37th IEEE International Southeastern Symposium on System Theory, pp: 495-499.

Herisa, H.K., B.S. Aghazadeh and M.N. Bahrami, 2009. Optimal feature selection for the assessment of vocal fold disorders. Comput. Biol. Med., 39(10): 860-868.

Kohavi, R. and G. John, 1997. Wrappers for feature subset selection. Art. Int., 97(1-2): 272-324.

Kukharchik, P., D. Martynov, I. Kheidorov and O. Kotov, 2007. Vocal fold pathology detection using modified wavelet-like features and support vector machines. 15th European Signal Processing Conference (EUSIPCO), Pozan, Poland, pp: 2214-2218.

Li, T., M. Oginara and Q. Li, 2003. A comparative study on content based music genre classification. Proceeding of the 26th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, ON, Canada, pp: 282-289.

Llorente, J.I.G. and P.G. Vilda, 2004. Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. IEEE T. Biomed. Eng., 51(2): 380-384.

Llorente, J.I.G., P.G. Vilda and M.B. Velasco, 2006. Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters. IEEE T. Biomed. Eng., 53(10): 1943-1953.

Mallat, S.G, 1989. A Theory for multi-resolution signal decomposition: The wavelet representation. IEEE T. Pattern Anal., 11(7): 674-693.

Manfredi, C., 2000. Adaptive noise energy estimation in pathological speech signals. IEEE T. Biomed. Eng., 47(11): 1538-1543.

Marinaki, M., C. Kotropoulos, I. Pitas and N. Maglaveras, 2004. Automatic detection of vocal fold paralysis and edema. Proceedings of Eighth International Conference on Spoken Language Processing-ICSLP.

Michaelis, D., M. Frohlich and H.W. Strube, 1998. Selection and combination of acoustic features for the description of pathologic voices. J. Acoust. Soc. Am., 103(3): 1628-1639.

Ritchings, R.T., M.A. McGillion and C.J. Moore, 2002. Pathological voice quality assessment using artificial neural networks. Med. Eng. Physic., 4(8): 561-564.

Rosa, M.D.O., J.C. Pereira and M. Grellet, 2000. Adaptive estimation of residue signal for voice pathology diagnosis. IEEE T. Biomed. Eng., 47(1): 96-104.

Tzanetakis, G. and P. Cook, 2002. Musical genre classification of audio signals. IEEE T. Speech Audio Proc., 10(5): 293-302.

Umapathy, K. and S. Krishnan, 2005. Feature analysis of pathological speech signals using local discriminant bases technique. Med. Biol. Eng. Comput., 43(4): 457-464.

Wallen, E.J. and J.H. Hansen, 1996. A screening test for speech pathology assessment using objective quality measures. ICSLP Proc., Duke University, Durham, pp: 776-779.