ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

Identification of Laryngeal Disorders Based On MFCC and Jitter and Shimmer Features and PSO Classifier

Fatemeh Salehi*[1], Mehran Emadi[2]

PG Student, Department of computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran[1].

Assistant Professor, Department of Electrical Engineering, Mobarakeh Branch, Islamic Azad University, Mobarakeh, Iran[2].

ABSTRACT: Laryngeal disorders are very uncomfortable and unbearable due to the continuous use of the human voice. However,the better identification of laryngeal disorders always is necessary. Recently, a lot of researchers such as doctors and biomedical engineers directed to the non-invasive and easier methods to detect disorders. The advantage of this method compared with invasive methods, are its tolerable for the patient and its more speed in achieve to the final result. Including new and non-invasive methods to diagnose of laryngeal disorders are audio signal processing method, which due to the effect that anomaly puts on the human voice; its type will be detected. Most of the researches that has been done so far are separated patients and healthy persons from each other and a few of studies have done classification of several disorders from between various disorders. In the proposed work, Feature extraction is done in three ways, the first depending on MFCC features and the second depending on Jitter and Shimmer features and the third by combining MFCC and Jitter and Shimmer. Meanwhile, achieved features are used along with PSO algorithm to analyse and classify anomalies based on several classes. Also, we used four groups of anomalies and a class of normal voice as benchmark data sets and evaluated and compared the proposed method with different feature extraction strategy. Our simulations results confirm the superior performance of the proposed method, especially when the features are extracted based on combination of MFCC and Jitter Shimmer. The result from the combination is 80% and using MFCC alone is 66% and using Shimmer and Jitter is 43%.

KEYWORDS: MFCC;PSO;nodules; polyps; paralysis; edema.

I.INTRODUCTION

Speech signal is significantly depend on the Larynx. When the larynx is impaired, speech quality gets decreasing. Hence, acoustic analysis can provide meaningful information to characterize the tremulous voices.Laryngeal disorders involved the person in two forms of congenital and acquisitive and can be divided into three main groups: Organic, functional, or a combination of both. Organ disorders are including two categories: structural and neurological. Structural disorders are related to physical problems that arise in the mechanism of the vocal cords. Neurological disorders arise due to problems in the nervous system. Examples of organic-structural disorders are including cysts, nodules, polyps, ulcers and organic- neurological disorders are include paralysis, tremors, Parkinson. Functional disorders mean that the physical structure is natural, but voice mechanism is acts non-efficient and inappropriate. Vocal cord thickening disorders, ventricular compression, leaning are on the group of functional disorders. In the identification of laryngeal disorders, this can be seen as a symptom of laryngeal disorders is eclipse sound [1,2,3]. But in the more advanced stages may lead to the loss of patient's larynx. Disorders that we examine in this study are nodules, polyps, paralysis, and edema. Nodules; The main cause them is long-term or bad use of voice that is found mostly in women and children and it is the common problem of amateur singers and teachers and the form of creating it in the larynx is appearance of large and symmetrical and homochromatic mass in the larynx[4]. Polyps; the main cause of it, is the accumulation of fluid in the tissues of the having mucosa, smoking and hyperthyroidism that can be found in young man and the form of creating it in the larynx is appearance of large and symmetrical and red and inflamed mass in the larynx[5].Paralysis; the most important cause of it, is stretching of recurrent nerve of the baby during the accouchement or dysfunction in controller nerves of the laryngeal and the form of creating it in the larynx is sudden closure of the respiration way and inability of the vibration of two sides of vocal cords or one side of vocal cords.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

Edema; the most important cause of it, is irritation caused by various diseases or change the nature of the cells of tissue of the vocal cords and the form of creating it in the larynx is topical swell of surface texture of vocal cords and their becoming heavy[6].In this study, the target is detection of several types of the larynx disorders on voice signal. To extract relevant features, we use the combination of MFCC methods and Jitter and shimmer method to frequency information and time information are considered in extracted model from signals. Also it is noteworthy, to search for a match of test samples with test and in order to final implementation and detect of cases, we will use of search method of PSO that brings the best diagnostic matched by test based on optimizing of target function of the issue of detection of disorders.In this study, in diagnose of the larynx disorders, MFCC feature for two reasons has been consideration:
No need to decode a voice signalrobustness against some audio signal distortionAlso jitter and shimmer features in order to assess the quality of the voice based on CD1 changes between cycles and voice amplitude changes between cycles are defined and considered.Then first, MFCC and jitter and shimmer features are expressed and then based on them, identification of laryngeal dysfunction done by algorithm of PSO.

II.DATABASE

Data sets from KAY Elemetrics, model 4337, version 1.03 are used in this study as benchmark to assess the performance of proposed algorithm. These databases have been developed in the laboratory of voice and speech of Massachusetts clinic, and contain 57 healthy signals and the 653 patient signals. In each signal, the subject produces the sound /a/ for a few seconds. A wide variety of organic, neurological, traumatic and psychological diseases is present in this database. Each patient signal has one label or more that specifies its type. If a disease has more than one label, the labels are prioritized based on their probability. Part of the patient signals has no label and their pathological status has only been mentioned. Diseases in the database include: full and partial paralysis, nodules, polyps, vibration, extreme performance, edema, vomiting, thickening, ulceration, leaning of vocal cords and several other diseases. In the present study, 5 categories of data are selected from the database related to 4 types of abnormalities of the larynx and a normal data set [7].

III.FEATURE EXTRACTION

The proposed Feature extraction process contains two steps for diagnosis of laryngeal abnormalities:

- Training features associated with various diseases from the database.
- Processing and detection of voice typical disease based on trained data in the previous stage.

In this scheme, each node with message searches for possible path nodes to copy its message. Hence, possible path nodes of a node are considered. Using NSS, each node having message selects its path nodes to provide a sufficient level of end-to-end latency while examining its transmission effort. Here, it derives the CSS measure to permit CR-Networks nodes to decide which licensed channels should be used. The aim of CSS is to maximize spectrum utilization with minimum interference to primary system. Assume that there are M licensed channels with different bandwidth values and ydenotes the bandwidth of channel c. Each CR-Networks node is also assumed to periodically sense a set of M licensed channels. Midenotes the set including Ids of licensed channels that are periodically sensed by node i. suppose that channel c is periodically sensed by node i in each slot and channel c is idle during the time interval xcalled channel idle duration. Here, it use the product of channel bandwidth yand the channel idle duration x, tc= xy, as a metric to examine the channel idleness. Furthermore, failures in the sensing of primary users are assumed to cause the collisions among the transmissions of primary users and CR-Networks nodes.

A.MFCC2method[8,9]

In existing researches in the field of audio signal, the researchers found that in an audio signal, the more effective information of signals, mainly there are at low frequencies and as a result, to achieve more useful information of the signal should be more emphasis on this part of the signal. This idea is created a method calling MFCC that is one of the most widely used methods in the field of recognition of voice speaker. In this analysis, first the size of Fourier transform of the desired data window from voice signal is calculated. Then, size of the Fourier transform is multiplied on the bank filter calling Mel and a number of the coefficient (which depends on the number of our filters) is extracted. Then, the logarithm of coefficients obtained and by using of one of the methods for dimension reduction, a few numbers of coefficients can be extracted.

The bank filter of Mel increases effect of low-frequency of voice signal in extraction feature of voice and weakens the higher frequencies further and reduces their impact on the extraction features. It should be noted that because of using of the Mel relation is that this relation is human ear hearing relation. Block diagram of MFCC method shown in the following figure.

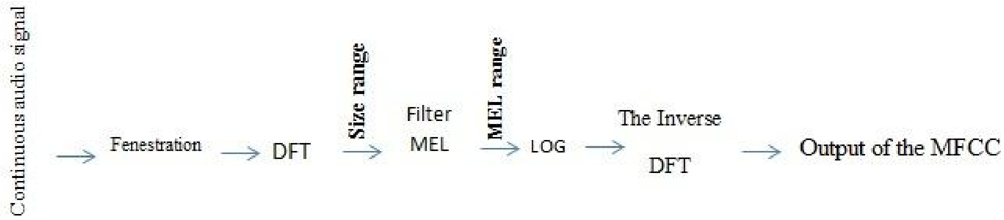

Figure 1: MFCC algorithm structure

So, the frequency spectrum of the sound signal is multiplied by Mel filter bank. If the audio signal is sampled as x (n), it Discrete Fourier transform will be as follows:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{\left(\frac{-j2\pi nk}{N}\right)}, k = 0,1, \dots, N-1 \quad (1)$$

Where, N is the number of samples in any signal window. Mel filter bank is as follow: As it can be seen, it has more focused on the low frequencies. Mel bank filter in such a way that the number of available filters in the low frequencies much more than the number of available filters in the high frequency.

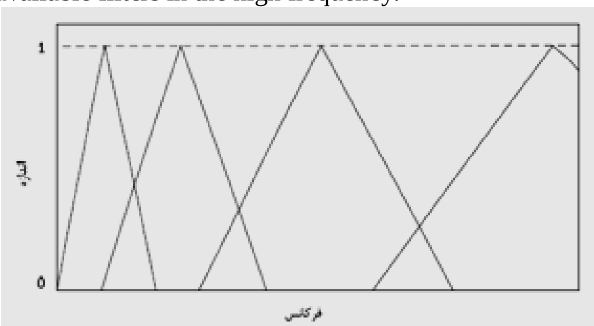

Figure 2: bank filters of MFCC method

This bank filter for every available filter in it, extract one coefficient, as the desired filter in the signal is multiplied and energy of obtained signal is calculated and can be written as a coefficient. Thus, the number of available filters in the bank filter, coefficients can be extracted.Given that the number of filters of the bank filters in low frequency, is much higher than high frequency, so the number of extracted coefficients from low-frequency is increase and this issue makes more emphasis on the low frequency. Each of the Mel bank filters can be structured as follows:

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

$$H_i(k)=\begin{cases}0 & for\ k<f_{b_{i-1}}\\ \dfrac{k-f_{b_{i-1}}}{f_{b_i}-f_{b_{i-1}}} & for\ f_{b_{i-1}}\le k\le f_{b_i}\\ \dfrac{f_{b_{i+1}}-k}{f_{b_{i+1}}-f_{b_i}} & for\ f_{b_i}\le k\le f_{b_{i+1}}\\ 0 & for\ k>f_{b_{i+1}}\end{cases} \quad (2)$$

Where, i is the number of filter and:

$$f_b(i)=\left(\frac{N}{F_s}\right).\hat{f}_{mel}^{-1}\left(f_{mel}(f_{low})+j.\frac{\hat{f}_{mel}(f_{high})-\hat{f}_{mel}(f_{low})}{M+1}\right) \quad (3)$$

$$f_{mel}=1127.ln\left(1+\frac{f_{lin}}{700}\right) \quad (4)$$

$$\hat{f}_{mel}^{-1}=700e\frac{\hat{f}_{mel}}{1127}-1 \quad (5)$$

With this filter bank, MFCC coefficients are calculated as follows:

$$C_j=\sum_{i=1}^{M}X_i.cos\left(\left(j.(i-0.5).\frac{\pi}{M}\right)\right);\ j=1,2,\ldots,J \quad (6)$$

Where, M is the number of filters in filters bank and J is the number of extracted coefficients. Xi also is calculated as a follow logarithmic:

$$X_i=log\left(\sum_{k=0}^{N-1}|X(k)|.H_i(k)\right);i=1,2,\ldots,M \quad (7)$$

Each of the Mel filters used in the filter bank is as follows:
As it can be seen that most gain of filter is the intermediate frequency[10,11].

B. Jitter and Shimmer feature

These features are defined for evaluating the quality of the audio signal and they are measured the changes of fundamental frequency and amplitude variations of the cycle to cycle of audio signals:

$$\text{Jitter}=\frac{\frac{1}{N-1}\sum_{k=1}^{N}|T_k-T_{k+1}|}{\frac{1}{N}\sum_{k=1}^{N}|T_k|} \quad (8)$$

$$\text{Shimmer}=\frac{\frac{1}{N-1}\sum_{k=1}^{N}|A_k-A_{k+1}|}{\frac{1}{N}\sum_{k=1}^{N}|A_k|} \quad (9)$$

The above equations are calculated the changes percentage. In the above equations T is time, A is amplitude and N is the number of cycles. The above parameters are specified in the example of below figure, in this example, N=4 is considered.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

IV. PROCESS OF FEATURES CLASSIFICATION

Particle swarm optimization (PSO) algorithm for pattern matching and identification[12,13]

Natural beings behave sometimes as a mass. One of the main streams in artificial life research is reviews how to handle of natural creatures as a mass and reimplementation of mass models on the computer. A new method of optimizing by using of replication of group behavior of natural creatures was invented in the early 1990s. Eberhart and Kennedy developed the particle swarm optimization (PSO) based on simulation of birds and fish categories mass. Every person has exchanges the previous experience in PSO. Here, we focus on the method of PSO as a collective intelligence techniques. PSO is created for solving nonlinear optimization problems with continuous variables. Moreover, unlike other evolutionary techniques such as genetic algorithms, PSO can be implemented with only a small program. This ability of PSO is one of its advantages compared with other optimization techniques. PSO is a method based on stochastic techniques that can be used to find the global minimum (non-deterministic) of non-linear planning issues. Kennedy and Eberhart developed the PSO through simulation of bird categories. Position of each operation by s and also its velocity by v is displayed. The correction of factor position can be done by using position and velocity information. The bird categories is optimized a specific objective function. Any factor knows the best value so far (pbest) and its position s. This information is the analogy of the personal experiences of each factor. Moreover, each of factors knows the best of value obtained in the group (gbest). This information is the analogous of personal experience of each factor that how to other factors acts around them. Each factor tries to changes its position by using of following information:The current position of s, the current velocity of v, pbest, gbest.This modification can be expressed by the concept of velocity. The velocity of any factor can be obtained from following equation:

$$v_i^{k+1} = wv_i^k + c_1 rand_1 \times \left(pbest_i - s_i^k\right) + c_2 rand_2 \times \left(gbest - s_i^k\right) \qquad (10)$$

Where, v_i^k is the factor velocity of i in repeat k, w is weight function, and c_j is a weight coefficient. Rand is a random number between 0 and 1, s_i^k is the current position of factor i in repeat k, pbesti is the pbest of i dot and gbest is the gbest of group. Means the right side of the above equation can be expressed as follows. The right side of this equation has three sentences. The first sentence is the previous velocity of factor. The second and third sentences are to change the factor velocity. Without second and third sentences, flying factor will continue the previous direction to hit the boundary. The factor tries to search the new areas and therefore, the first sentence is associated with variation in the search process. In other words, without the first sentence, flying velocity of factor only by using of the current position and its best position in the past is determined. The factors will try to converge to pbests or gbest.The current position of searching point in answer space can be modified by following equation:

$$s_i^{k+1} = s_i^k + v_i^{k+1} \qquad (11)$$

Each factor is modifying its current position by using of combining vectors shown in the above figure. In fact, PSO is used of several searching point, and searching points gradually by using of pbests and gbest are close to the optimum point.

A. Steps PSO[14]

Step 1. Create the initial condition for each factor. The primary researcher points of (s_i^0) and velocity of (v_i^0) each factor usually randomly is generated in the allowed range.
Step 2.Evaluation of the searching point of any factor. The amount of target function can be calculated for any factor.
Step 3. Correct any search point
Step 4. Check the situation of output. If the current repeat number is reach to the maximum of default repeat number, then stop, else go to the step 2.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

V. THE PROPOSED OVERALL ALGORITHM TO DETECT DISORDERS OF THE LARYNX

According to the described features, we offer the proposed algorithm that is composed from two following parts to identification of laryngeal disorders:
Learning the corresponsive features with various diseases from the database.
Processing and identification of type of disease of voice sample based on learning data from previous step.
In the first step, the available data in database of available diseases is processed and is obtained two categories of features for them that are MFCC and jitter and shimmer and to prevent of repeated calculations are saved into the secondary database. These cases are done by using of pre-processed of initial data of available diseases from the perspective of the amount of noise and quality of the signals. In other words, in the first step, the available data of various diseases in database are processed independently, and the data with acceptable amount of noise are selected and in the one step, noise is eliminated (For example, by using the software of noise reduction of voice signal such as noise killer, the quality of signal somewhat improved). This issue is done to better detection of features related to diseases from learning data. Then in the second step, all of processing are done based on the saved data in the secondary database that they have the ability of develop the knowledge related to laryngeal disorders, well. In this step, data or test data by using of two types of expressed features are extracted features and by PSO algorithm with available samples in the secondary database are matched. This matching in search will be possible in particular of optimize the target function in the following form:

$$\text{objective} = \text{Arg max}_{\alpha_i^*} \; F_{\text{frequency formant}} + F_{\text{quality of voice}} \qquad (12)$$

The above objective is consisting of two part of compliance of frequency features (format) compliance of voice quality. Each part refers to the set of extracted features by MFCC or jitter and shimmer. Laryngeal anomaly detection problem in research is maximizing this problem by choosing the best vector of candidate feature (α_i^*) from the data set of secondary database. The overall structure of the proposed algorithm is shown in the following block diagram.

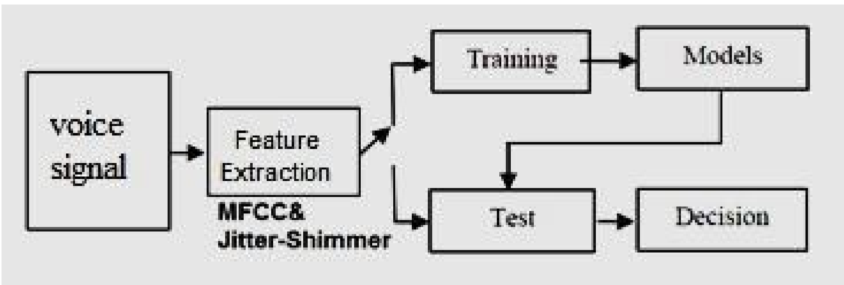

Figure 7. The overall structure of the proposed algorithm

Optimization of expression objective function and pattern matching as noted above, is done by particle swarm algorithm in duration of described the mechanism in the previous section.

VI. THE PROPOSED ALGORITHM PERFORMANCE REVIEW MECHANISM:

First, the data contained in the database are classified into two categories of training data and test data. In this study, from the all data of each class we considered approximately 60% as training data and the remaining data are used to test the proposed algorithm. In the first phase as noted previously, based on training data we extracted two categories of frequency feature (MFCC) and voice signal quality (Jitter and shimmer) and we saved into the secondary database. Then for the test data, we extracted two types of features and we matched them with available features in the secondary database by PSO algorithm. This algorithm is maximizing target function expressed in previous chapter in the search of matching the couples data of input signal and available data in secondary database. Due to the repetitive of particle swarm algorithm, convergence of it is very important. The convergence of this algorithm is demonstrated by testing the situation of the best solution found during the specified number of repetitions. For example, the convergence of the consider algorithm for a matching problem is obtained as follow, that its downside is prove of its convergence to the best possible answer.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

In the feature extraction process, 20 MFCC parameters are considered and with the jitter and shimmer features created the final vector that has a good effect in the identification process. In addition, in the MFCC feature extraction process, selected window to apply the filter bank is Hamming type. Each of the frames found during processing taken in 256 samples and Interference between the frames is applied as 50 percent (about 100).Matching results are compared by actual tag of test data and finally algorithm identification rate is achieved for different data (classes of diseases) and all of database. Considered equation to compute of detection rate is as follow:

$$\text{Recognition rate(Rr)} = \frac{I_{A\cap B}}{N} \qquad (13)$$

In the above equation, $I_{A\cap B}$ is the compliance of obtained results from proposed method with the available test data tags in the database. Closing this equation to 1, means that detection rate is high. Complete and separation results obtained from proposed algorithm on the database of kay is as follow. The number of samples correct and false detection also about data is included in this table. These data are shown in the figure () based on the single feature of MFCC or combination of two features categories of MFCC and jitter and shimmer.

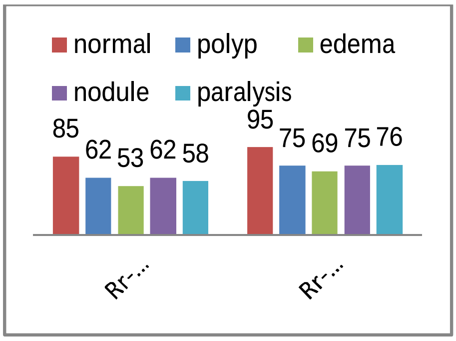

Figure9.Comparative data based on the single feature of MFCC or combination of two features categories of MFCC and jitter and shimmer for each groups

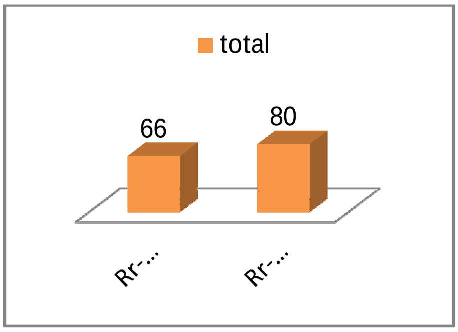

Figure10.Overall data based on the single feature of MFCC or combination of two features categories of MFCC and jitter and shimmer for each groups

VII. CONCLUSION

Due to the importance of diagnosis of laryngeal abnormalities, achieving a specific mechanism with good accuracy is very important. The best available diagnostic methods those are non-processing too, have not necessary accuracy and have aggressive nature that considered being harmful. So, providing the new processing techniques that included high accuracy, are desired. In this study, a processing method with combination extraction features by the MFCC and jitter and shimmer is provided that it has a good quality at detection of 5 types of diseases in database of kay. In the proposed method for features compliance with available samples in database, an intelligent search method particle swarm is used.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

The obtained results show that proposed method of combining of two feature categories in increase of identification rate and classification is successful while using only MFCC features has resulted a lower detection rate. Also, using of PSO method to matching patterns due to the complicated and non-linear of n-dimensional space obtained from features combining is cause of more accurately in identify of test data.

REFERENCES

[1] Arjmandi, MeisamKhalil and Pooyan, Mohammd and Mohammadnejad, Hojat and Vali , Mansour,'' Voice Disorders Identification Based on Different Feature Reduction Methodologies and Support Vector Machine ,"Proceedings of ICEE 2010, May 11-13, 2010 ,978-1-4244-6760-0/10/$26.00 ©2010 IEEE, 2010.

[2] Rahmani, Mohsen, "Applying the speech enhancement method as a preprocess in order to increase of the accuracy of speech diagnostic", 2001.

[3] Mehdi Zadehfar, Vida and Almas Ganj, Farshad, "Audio anomaly detection by using dynamic data features", 2011.

[4] Behroozmand , Roozbeh and Almasganj,Farshad," Comparison of Neural Networks and Support Vector Machines Applied to Optimized Features Extracted from Patients' Speech Signal for Classification of Vocal Fold Inflammation, "pp. 844-849, 2005 IEEE International Symposium on Signal Processing and Information Technology,0-7803-9314-7/05/$20.00©2005 IEEE.

[5] Vaziri, Ghazakeh and Almas Ganj, Farshad, "Application of nonlinear feature of approximate entropy in detecting disorders of laryngeal polyps and nodules", 2007.

[6] Behroozmand Rouzbeh and Almas Ganj, Farshad and Moradi, Mohammad Hasan, "Application of features of fractal dimension and non-linear dynamics of phase space of speech signals in the detection of laryngeal anomalies of vocal cords paralysis patients", 2008.

[7] Disordered Voice Database (CD-ROM), Version 1.03, MassachusettsEye and Ear Infirmary, Kay Elemetrics Corporation, Boston, MA, Voiceand Speech Lab., October 1994

[8] A. C. L., M. T. J. and R.J. P., "Speech Recognition Using Reconstructed Phase Space Features", ICASSP., 2003.

[9]Z.Ali.,M.Alsulaiman.,g.Muhammad.,I.Elamvazuthi.," Vocal Fold Disorder Detection based on Continuous Speech by using MFCC and GMM", 978-1-4799-0724-3/13/$31.00 ©2013 IEEE,novaember2013.

[10] C.Becchetti,L.P.Ricotti,Speech Recognition Theory and C++ Implementation ,JOHN WILEY & SONS,LTD,March 2000

[11] J . Hakkien, J. Sountausta, R. Hariharan,M.Vasilache, K.Launilla," Improved Feature Vector Nomalization for Noise Robyust Connected Speech Recognition .", EUROSPEECH 99 Volum 6 page 2833-2836,1999.

[12] Xia, W.J., Wu, Z.M., A Hybrid Particle Swarm Optimization Approach for the Job-Shop Scheduling Problem. International Journal of Advanced Manufacturing Technology 29, 2006, pp. 360.366.

[13] Sha, D.Y., Hsu, C.-Y., A Hybrid Particle Swarm Optimization for Job Shop Scheduling Problem Computers & Industrial Engineering 51, 2006,pp. 791.808.

[14] Poli, R., Kennedy, J., Blackwell, T., Particle Swarm Optimization: An Overview. Swarm Intelligence, 1,2007, pp. 33.57.

BIOGRAPHY

Fatemeh Salehi is post graduate student at Department of Computer Engineering,NajafabadBranch,Islamic Azad University , Iran. Her research interests include image and video processing in signal processing. He has several paper in relevant journals and conference publications in subject of her interests.

MehranEmadi received the B.S. degree in Electrical Engineering (Electronics) in 1995, the M.S. degree in Electrical Engineering (Electronics) in 1999 from Azad University, NajafAbad branch, Iran, and the Ph.D. degree in Electrical Engineering (Image processing) from University Technology Malaysia (UTM), Malaysia in 2013. His researches include semiconductors in his M.S. degree and image processing, computer vision, biometric applications in his Ph.D. degree period. There have been more than 30 technical papers, published by him and he is an assistant professor in Azad University in Iran and now he is the head of Master students in Azad University Mobarakeh branch. He has been working as the program chair, and a committee member for some international conferences, and has been training several students in Universities by webinar.