ciência e natura

A New PSO Classifier Based Method Applied to Detect Anomalies of the Larynx

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

[1] Department of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran

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

Abstract

Quality of the human voice can be affected by anomalies of the larynx due to the physical, Nerve-muscle or only nervous origins. Video Stroboscope and vocal folds movement display systems are key tools which often used to detect Laryngeal anomalies. These methods are invasive, time consuming and expensive, so researchers are trying to find non-invasive methods that lead to the final answers faster than invasive methods and contain tolerable condition for patients. Many interests are directed to the application of speech processing techniques in relevant works. In these works, researchers were used different processing methods in medical engineering to detect anomalies. Recently, variety of researches presented to detect anomalies from the audio signals of individuals based on the features that extracted from audio signals. These methods have been conducted to separate patient audio from non-patient once. These researches do not work properly when an anomaly is among several anomalies and achieve bad error rate. In this paper, we aim to propose a new method of automatic Anomalies detection which performs based on a new mechanism of feature extraction and a PSO classifier. 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 analysis 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: Vocal ford disorders, PSO Classifier, MFCC, Shimmer, Jitter.

1 Introduction

S peech 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. Voice disorders are caused by problems in structure or function of the larynx, or both, which can occur for various reasons. Anomaly is prevalent issue which Larynx encountered. This problem that is studied often following a change in voice brings especial complexity for patients who involved. In this study, we used several anomalies as polyp, nodule, edema, vocal folds paralysis to evaluate performance of the proposed method in detecting the problems. Polyp, of which the most important cause is the fluid accumulation in mucosal tissues, smoking and hyperthyroidism, is mostly found in young men and it appears as large, asymmetrical, red and inflamed masses in larynx. Polyps are soft and flexible and are usually unilateral. Signs of polyps are rough voice or affected breathing. In addition, the patient may feel that something is in his or her throat [1],[2]. Nodules are primarily due to the long-term or bad use of voice that is found mostly in women and children and is the common problem of amateur singers and teachers and appears as large and symmetrical masses of the same color in the larynx. Initial nodules are soft and flexible. The entire larynx may be somewhat inflated. In the early stages, nodules are unilateral and may be confused with polyps. Chronic nodules are often hard, white and thick. Hoarseness and affected breathing are signs of nodules. Many people complain of pain in the neck around the larynx which might radiate to upward to the ear and downward to the chest [3]. Paralysis is mainly due to the recurrent nerve stretching of infant at birth or abnormalities in the neural control of the larynx and appears as sudden closure of the larynx in breathing and bilateral or unilateral vibrating disability of vocal folds. Characteristics of vocal cord paralysis include: dysphonia of the voice quality, reduced voice phonation, reduced voice volume, uniformity of voice volume, broken voice volume level [4]. Another anomaly is vocal cord edema. The main cause of edema is the stimulations induced by various anomalies or change in the nature of the vocal fold tissue cells.

It appears as local swelling of surface tissue of vocal cords and their heaviness in the larynx [5]. There are different works that pursue the analysis of the disorder voices. For example, Marek Wisniewski and his colleagues in [6] presented a new work on detecting the disorder voice based on HMM or Saenz-Lechon considered SVM and GMM to study the effect of audio Compression in automatic detection of voice disorders in [6]. Also, Lotfi Salhi classified the voice signals by using a multi-layer neural network [7]. In [8] suggested a novel hybrid method for vocal fold pathology diagnosis. So that, first, an initial feature vector for vocal fold pathology diagnosis is proposed. Then, for optimizing the initial feature vector, a genetic algorithm is proposed. Some experiments are carried out for evaluating and comparing the classification accuracies, which are obtained by the use of the different classifiers (ensemble of decision tree, discriminant analysis and K-nearest neighbours) and the different feature vectors (the initial and the optimized ones). Finally, a hybrid of the ensemble of decision tree and the genetic algorithm is proposed for vocal fold pathology diagnosis. The experimental results show that normal and pathological signal are separated with 93 percent accuracy.

In this paper, we propose a new strategy which can be used to simultaneously diagnose multiple abnormalities. In this strategy, a new PSO based algorithm is accompanied with MFCC and Jitter Shimmer feature extracting methods for multi-class classification and analysis of anomalies. Proposed strategy and details of our work is mentioned in the next sections.

The rest of the paper is organized as follows: Section 2 contains a definition of MFCC and Jitter Shimmer feature extraction strategy. In section 3, data sets which are used to evaluate our proposed algorithm are introduced. The proposed method is presented in section 4 and section 5 reports its empirical results. Afterward some concluding remarks are provided in Section 6.

2 Methods Of Feature Extracting

In this section, some tools are introduced that we used to find the suitable parameters for classifying healthy and abnormal voice signals.

2.1 MFCC features

Mel frequency cepstral coefficients are defined by Eq.1.

$$\widetilde{C} = \sum_{k=1}^{K} \log(\widetilde{S}) \cos[n(k - \frac{1}{2})]\frac{\Pi}{K} \quad (1)$$

MFCC coefficients can be extracted by a 32 filter bank which is applied on 10ms (256 sample) Hamming windowed frames at 50% of overlap.

2.2 Jitter and Shimmer Features

Jitter occurs over voice production process, especially in vowel phonation and mostly defined as small fluctuations in glottal cycle lengths [10]. These features help give the vowel its naturalness in contrast to constant pitch and amplitude which result in a machinelike sound. Also, Jitter and Shimmer have significant effect on quality of speaker voice. A suitable estimation of Jitter can discriminate healthy and dysphonic speakers. Hence, Jitter shows the precent of changes in cycle duration between cycles and Shimmer indicates the precent of changes in speech amplitude between cycles. These features are defined based on the bellow equations.

$$Jitter = \frac{\frac{1}{N-1}\sum_{k=1}^{N}|T_k - T_{k+1}|}{\frac{1}{N}\sum_{k=1}^{N}|T_k|} \quad (2)$$

$$Shimmer = \frac{\frac{1}{N-1}\sum_{k=1}^{N}|A_k - A_{k+1}|}{\frac{1}{N}\sum_{k=1}^{N}|A_k|} \quad (3)$$

Where, Tk and Ak respectively are the time and amplitude, N is number of cycles.

3 Data Base

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 [10].

4 The Process of feature extraction and classification

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 the first step, each signal of each data set processed to extract MFCC and Jitter Shimmer parameters. After that, obtained parameters are stored in a secondary database to avoid repeated calculations. Also, it is must be mentioned that, before any processing we perform pre-processing on raw data of available diseases from the perspective of noise level to develop the quality of signals. In other words, in the first phase, various disease data in the database are processed independently and data with an acceptable noise level are selected and noise level decreased (For example, using audio signal noise reduction software such as noise killer signal quality is somewhat improved).This is done for the better detection of disease-related characteristics from the training data. Then in the second step, all the processing is done based on data stored in the secondary database that has the ability to express the knowledge of disorders

of the larynx. In this step, test data is processed using two types of mentioned feature extractor and extracted features are matched using PSO algorithm to the samples which stored already in the secondary database. The matching process is same as optimizing the bellow objective function:

$$\text{Objective} = \text{Arg max}_{ai^*}\, F\text{ frequency formant} + F\text{ quality of voice}$$

The above objective function consists of two parts. First part, matches the frequency features (formant) and the second part matches the voice quality. Each section refers to the set of features extracted by MFCC and Jitter and shimmer. Laryngeal anomaly detection problem in this research is to maximize the problem by selecting the best candidate feature vector (ai^*) from data collection of the secondary database. The overall structure of the proposed algorithm is shown in the block diagram of Fig.1.
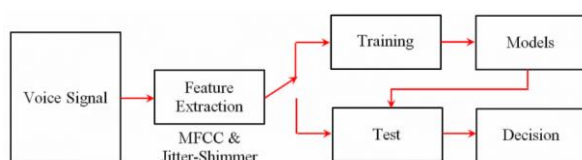

Fig. 1. The overall structure of the proposed algorithm [11]

As expressed already, optimization of the objective function Eq.4 and also the matching process are carried out by mass particles algorithm. At First, existing data in the database is divided into two categories: training and test data. In this study, approximately 60% of data of each class were considered as training data, and other data were used to test the proposed algorithm. As noted earlier, in the first phase, two groups of frequency features (MFCC) and audio signal quality (Jitter and shimmer) were extracted and stored in a secondary database based on the training data. Then two types of features were extracted from the testing data are matched using PSO algorithm with the same features which placed in the secondary database. This algorithm implements the matching of input signal with data in the secondary database to maximize the objective function introduced in the previous chapter. Due to the iterative nature of the particle mass algorithm, its convergence is very important. The convergence of the algorithm is proved by the testing the best solution found within the specified number of repetitions. For example, the convergence of the algorithm to a matching problem is obtained as follows: The descending status of it may prove its convergence to the optimal solution.
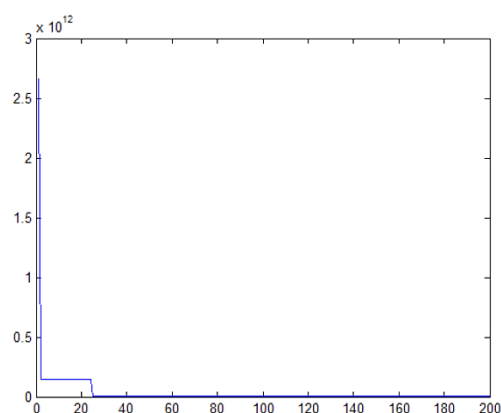

Fig. 2. Shows the convergence of the particles mass method in a matching sample

The following is a summary of the proposed spectral algorithm. The proposed algorithm maps the original space into the feature space to extract good parameters for detecting abnormality. The overall follow chart of the proposed algorithm is as follows:
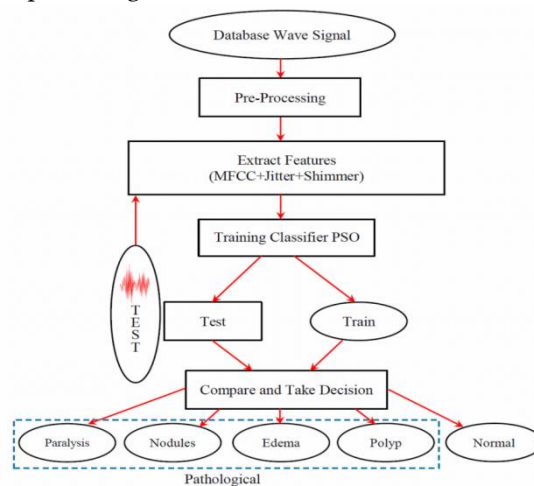

Fig. 3. Follow chart of the proposed algorithm

5 Experimental *results*

In our proposed algorithm, the feature extraction process is performed using 18 components of MFCC which coupled with Jitter and Shimmer Features. So, the final feature vector is created based on the features of Jitter and shimmer which has a positive effect on the identification process. In addition, during the

MFCC feature extraction process, a Hamming type window is selected to apply a filter bank. 256 samples considered for each frame in our processing and interaction between frames is applied as 50% (about 100 samples). Consequently, the matching result of the proposed algorithm is compared with the actual label of testing data and the recognition rate is calculated for different data (classes of diseases). In this study, the recognition rate considered as Eq.5 [12].

$$Recognition\ rate(\text{Rr}) = \frac{\text{I}_{A \cap B}}{\text{N}} \qquad (5)$$

Proposed algorithm is performed using different types of features on five data sets. Features considered as bellow:

- MFCC
- Jitter & Shimmer
- Combination of MFCC and Jitter Shimmer

As indicated in the bellow figure, combined Features has superior performance in comparison with else. The results are shown in Figures (4-5).
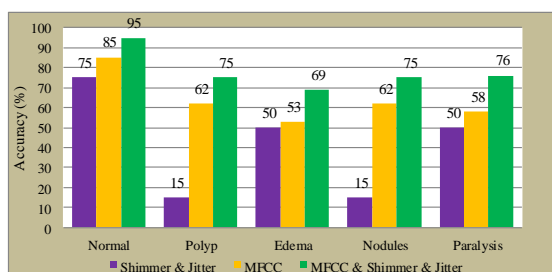

Fig. 4. Comparison of data on Jitter and shimmer and the single feature of the MFCC and combining the two features of MFCC, Jitter and Shimmer for individual groups
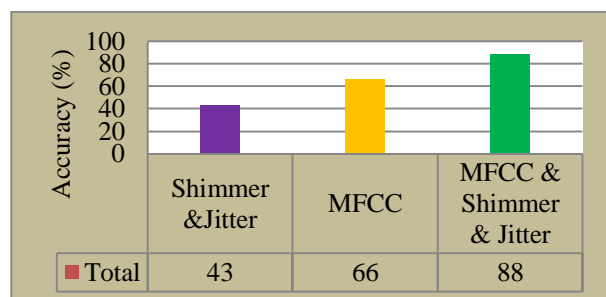

	Shimmer &Jitter	MFCC	MFCC & Shimmer & Jitter
Total	43	66	88

Fig. 5. *Overall comparison of data based on Jitter and Shimmer and the single feature of MFCC*

and combination of two features of MFCC, Jitter and Shimmer for individual groups

6 Conclusion

By following to the importance of diagnosis of laryngeal abnormalities, achieving a specific mechanism with good accuracy is very important. The best diagnostic methods that are non-processing are not enough accurate and have aggressive nature that is considered to be harmful. In this paper, we proposed a new strategy of anomaly detection based on combination MFCC and Jitter Shimmer features and PSO algorithm. Experimental results show the superior performance of the combinational features in comparison with else. Also, due to the complexity and the nonlinearity of the n-dimensional space PSO algorithm leads to the greater accuracy in detecting test data.

Refrences

[1] Byeon, H. K., Hoon Kim, J., Ho Kwon, J., Hee Jo, K., Jun Hong, H., Shik Choi, H. (2013). Clinical Characteristics of Vocal Polyps With Underlying Sulcus Vocalis, Journal of Voice, Vol. 27, No. 5, p.632-635

[2] *Lucero, J. (2011). Dynamics of the Vocal Fold Oscillation, Trends in Applied and Computational Mathematics, vol. 6, no. 1, p. 11-20.*

[3] Galata, G., Schulte, K. M. (2014). Management of the Thyroid Nodule, Endocrine Surgery, vol. 32, no. 10, p. 531-537.

[4] Tarin, T. T., Martinez, J. A., Shapiro, N. L. (2005). Familial bilateral abductor vocal cord paralysis, International Journal of Pediatric Otorhinolaryngology, vol 69, p. 1693-1696.

[5] Varma, J. R., Hazarika, B., Pradhan, S., Momin, A., (2009). Reinke's edema, Indian Jounnal of Radiology and Imaging, vol. 19, no. 4, p. 296–297.

[6] *Hansen J. H.L. et All. (1998). A non-linear Operator based speechFeature Analysis Method with Application to Vocal fold pathology Assessment , IEEE on Biomedical Engineering, vol. 45, Issue: 3, p. 300-313.*

[7] *Salhi, L., Mourad, T., Cherif, A. (2010). Voice Disorders Identification using Multilayer NeuralNetwork, International Arab journal of Information Technology, Vol. 7, No 2.*

[8]   Majidnezhad,V. (2014). A novel hybrid method for vocal fold pathology diagnosis based on Russian language, Journal of AI and Data Mining (JAIDM), vol. 2, no. 2, p. 141-147.

[9]   Tavares, R., Monteiro, N., Correia, S., Costa, S., Aguiar Neto, N., Fechine, J. (2010). Optimizing laryngeal pathology detection by using combined cepstral features, Proceedings of 20th International Congress on Acoustics, p. 1-6.

[10]   Disordered Voice Database (CD-ROM), Version 1.03, (1994), Massachusetts Eye and Ear Infirmary, Voice and Speech Lab., Boston, MA, Kay Elemetrics.

[11]   Fezari, M., Amara, F., El-Emary, IM. (2014). Acoustic Analysis for Detection of Voice Disorders Using Adaptive Features and Classifiers. Proc 2014th Int Conf on Circuits, Systems and Control, Switzerland, p. 112-117.

[12]   Vikram, C. M., Umarani, K. (2013). Pathological Voice Analysis to Detect Neurological Disorders Using MFCC and SVM. International Journal of Advanced Electrical and Electronics Engineering (IJAEEE), Vol. 2, No. 4.