ISSN: 2321-7782 (Online)
Impact Factor: 6.047
Volume 5, Issue 1, January 2017
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Voice Pathology Identification System using SVM Classifier

S. C. Punitha[1]
HOD,
Department of Computer Science
PSGR Krishnammal College of Arts & Science for Women,Coimbatore – India

N. A. Sheela Selvakumari[2]
Assistant Professor,
Department of Computer Science
PSGR Krishnammal College of Arts & Science for Women,Coimbatore – India

Abstract: Voice Pathology analysis plays a major role in the recent history of medical researchers. Due to the need in research, the identification and classification of voice pathology are still considered as a challenging is in the voice analysis. Commonly Patients can able to identify a minor change in voice parameters, such as hoarseness, but the voice pathologies can originate from a wide spectrum of causes, like common cold to a malicious tumor. Nowadays, human experts like otolaryngologists were detecting a number of voice pathologies from the Patients speech. Unluckily, the current classification rate of voice pathology by the human experts is only about 60-70%. Thus voice pathologies can be finding by the endoscopy procedures like laryngostroboscopy or surgical micro laryngoscopy, which distress the patient to a great extent and it is expensive also. The main objective of this research work is to aid this voice pathology finding process with computer based diagnostic tools. This voice pathology detection system works based on the support of the clinic based professional otolaryngologists, by identifying the risk of the pathology automatically without any endoscopy which increases the detection of voice pathology at the earliest. In this research work, the speech signal is analyzed by the acoustic parameters like Signal Energy, pitch, Silence removal, Windowing, Mel frequency Cepstrum, and Jitter. At the end, the classification technique i.e Support Vector Machine is used to classify the normal and pathology voice based on the features extracted in the previous phase. Based on the results & discussion mentioned below, thus the Voice pathology detection system successfully classified the normal voice and the pathological voice which helps in diagnosing and analyzing the patient.

Keywords: Voice Pathology, Classification, Acoustic Analysis, Mel Frequency Cepstrum, Support Vector Machine.

I. INTRODUCTION

In the recent medical history, a significant interest was shown in voice analysis, which is a medical study of the patient's voices, which have a problem in the vocal cords. In analyzing the patient's voice, a voice therapist needs some extensive training and experience in that particular subject. But it is not possible in all situations, because voice problems commonly originate from the vocal folds or Laryngeal muscle, which controls the voice produced by the human.

Such vocal fold analysis and to watch their movements is physically a difficult one, thus there is a need for a device or equipment to measure the voice pathology. This need creates an awareness among the researcher to find some technology to help the voice therapist in detecting the voice pathology. Such technological support can be done base on the complete understanding or studying of the most common vocal disorders, their symptoms, root cause and disease side effects.

Based on the study, researcher planned to create a voice pathology detection system which will be a modest, non- invasive, and consistent automatic system used for the detection of pathologies in the human vocal cords. Acoustic features act as a main parameter to discriminate the voice signals as normal or pathology. Acoustic features were used to describe the features or physical characteristics of speech signal at every instance of time. Based on the features extracted from the Acoustic Features, the classification system analyze and decide whether the voice signal produced from the vocal cord is affected by some disease or not.

By incorporating such techniques used in the voice analysis, this paper explores an efficient Voice Pathology Detection System. A detailed study was made in voice analysis and the dataset is constructed by recording the human voice in a noise free environment. The speech signal is then analysed in order to extract the acoustic parameters such as the Signal Energy, pitch, Mel frequency Cepstrum, and Jitter.

Next an attempt is made to analyse and to discriminate pathological voice and the normal voice using classification method. In this connection, a famous classification model i.e., Support Vector Machine Algorithm is used to classify the voice signal into normal or pathology based on the features extracted from the Acoustic parameters. Finally a successful automatic Voice Pathology Detector was designed, which will help the voice therapist to analyze and diagnose the patient's voice.

This paper indicates the literature review in section (II), Problem statement in section (III), a detailed explanation of the proposed research work, techniques and methodology implemented was discussed in section (IV). The proposed framework environment, test results, and final screens were discussed in section (V). Finally, the conclusion in noted in the section (VI) and plan for future work is discussed in section (VII).

II. Literature Review

Detection of voice pathology and its Treatment approaches followed in clinical practice have been discovered through investigations done from the speech-language pathologists [1]. The voice features were analysed in [2], the parametric approaches were used in identifying glottal signal in speech signals, whereas Time Frequency, Amplitude Modulation and Magnitude spectrum were based on the Non parametric approaches.

In [3] the most popular Acoustic feature extraction technique, MFCC i.e., Mel-Frequency Cepstrum Coefficients were used as a base measures to classify the normal and pathology voice using GMM Classifier. In [4] the input dataset used to find the pathology were compressed and stored in MP3 and the Classification Methodologies used to classify the normal and pathology voice were tested and compared among the GMM and SVM techniques.

[5] Incorporates the classification techniques like Principal Component Analysis i.e PCA and Support Vector Machine to classify the normal and pathological voice based on the 27 features extracted from the voice dataset. In [6] a different approach was initiated by selecting the Genetic Algorithm to optimize the feature set for the speech signal and for the classification the SVM kernel is used. Like [3], [7] also used the MFCC as a main feature extraction technique and to classify the voice signals it used the GMM i.e, Gaussian Mixture Model.

III. Problem Statement

The main motto of the proposed research work is to provide an automatic voice analysis system, which is used to find whether the patients voice contains pathology or not. Mostly patients went to hospital and waiting for a long to time to see the voice therapist and consult with them. If a patient has a problem in his voice, then sure he has to undergo some medicine process to get well. If the patient is waiting to know whether his voice is normal or not then, it is a problematic one as well as a time consuming process. On the other side, the voice therapist also needs additional training and experience to handle such decision making between normal voice and pathology voice. Hence to overcome this problem, our proposed work acts as a tool to analyze and detect the pathology that present in a patient's voice

IV. Proposed Research Methodology

The Voice pathology identification system classify the input voice as normal or pathology voice by processing the voice in three steps i.e i) Preprocessing the input signal ii) Extracting the meaningful features from the preprocessed signal, and iii)

Using the extracted features, classifying the voice into normal or pathology. The overview of the proposed research work to detect and classify the voice pathology is depicted in figure 1.

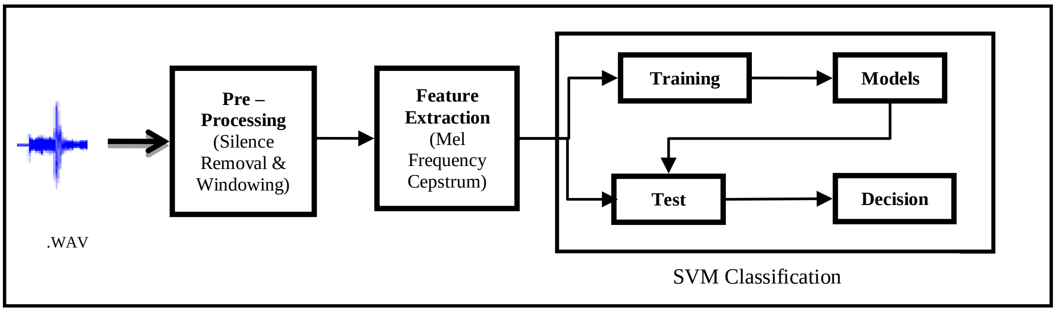

Figure 1: Research Plan for the proposed work.

A. Pre-processing:

The preprocessing step contains the silence removal from the audio signal using two simple audio features like short time energy and spectral centroid. These two features filtered the voiced samples by setting a threshold on the voice signal, finally, the voice sample was segmented and the voice samples below the threshold were treated as silence and those samples were removed from the speech segments.

Generally, the amplitude of the voice signal differs significantly with time i.e, the breadth or range of the voice signal will differ during the speech segment and the silence segment. The amplitude of the voice is lower when the signal segment contains silence and the amplitude will be larger when the signal segment contains speech. Such amplitude variation over a period can be detected by the short-time energy function. The short time energy for an audio signal is defined as the sum of squares of voice samples in one segment. Thus the short-time energy function reflects that the energy is greater in voiced signal than the silenced or unvoiced signal.

$$\text{Short Time Energy} = \sum^{\infty} S_i^2 \qquad \text{......... Equ(1)}$$

Where S is the Short time speech segment and ‘i’ is the no of voice samples.

Like Short time Energy function, Spectral Centroid is also a function, which is used to find the brightness of a sound. It can be measured by calculating the center of gravity i.e, the weighted mean of the frequency that present in the voice signal, using the Fourier Frequency Transformation. Spectral centroid finds the midpoint of the spectral energy distribution, which is also called as a balance point of the spectrum.

$$\text{Spectral Centroid} = \frac{\sum_{n=0}^{N-1} f(n)\, x(n)}{\sum_{n=0}^{N-1} x(n)} \qquad \text{......... Equ(2)}$$

Generally, speech signals are unstable. But for a short interval of time like 10 to 30 ms, the speech signal may be stable. Hence to cut such short speech signal, windowing function will be a helpful one. For a period of time, the speech signals taken for processing is called Window. Thus the voice data present in the Window is called Frame. For every Q ms, the voice features were extracted from the signal, which is called as frame rate, hence the duration of the Window is P ms. Usually, and P is bigger than Q, which paves the way to overlap the consecutive frames in the window. Such overlapping fragments were used for speech analysis. In this, the length of the frame is called Frame Length and the overlapping length is called Frame shift Length. The Hamming windowing technique is used because of its larger working part and its smoothness in the low pass.

Feature Extraction:

Feature Extraction plays a major role in selecting the features, which is suitable for the classification process. Selecting features is an important one for a Decision-Making System. Thus MFCC i.e, Mel Frequency Cepstral Coefficients, is one among the popular feature extraction techniques. In voice analysis, Mel Frequency Cepstrum is a representation of a short-term power spectrum of a voice signal, which produced based on the linear cosine transformation of a log power spectrum on a nonlinear Mel-scale frequency.

Mel Frequency Cepstral coefficients joined as a group to make up an MFC. The Mel Frequency Cepstrum uses equally spaced frequency bands on the Mel Scale, which estimates the human auditory system's response more closely than the normal Cepstrum which uses linearly-spaced frequency bands.

Jitter is a measurement used to check the vocal stability. Actually, the frequency of a voice signal will vary from one cycle to another cycle. By finding the random period variability, vocal jitter can be calculated, which is responsible for hoarse or rough voice signal. The jitter frequency variability for a normal voice will be less than 1%. Thus Jitter has been a widely used methodology for voice pathology analysis.

Classification:

The classification algorithm is used to classify the normal as well as pathology voice, here SVM i.e, Support Vector Machine is used to classify the voice signals. SVM is a supervised Machine Learning algorithm, which is used for both classification and regression. Supervised algorithms use a training set, based on the models derived from the training set, the real-time test data set will be processed. Here the final voice signal is fed into the system as a test data, based on the training set data model, it classifies the voice signal as normal voice or pathology voice.

V. EXPERIMENTAL RESULTS

In this paper, the voice pathology identification system was designed and implemented in Matlab. For the experimental study, 10 sample (6 Pathological and 4 Normal) voice files in the .wav format were given as input to the system, which was recorded in a noise free environment using a microphone array. The view of the designed Voice Pathology Identification System was depicted in figure 2.

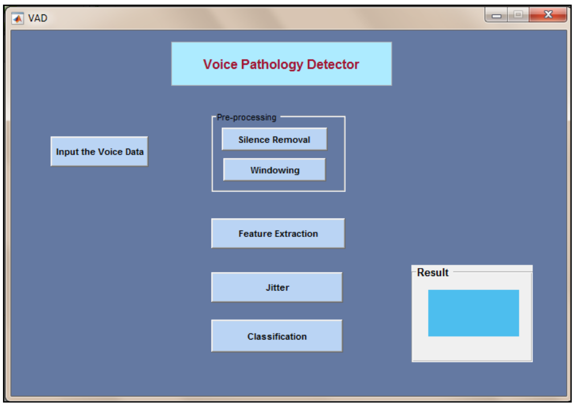

Figure 2: Voice Pathology Identification system

In the second phase, the sample speech .wav file is pre-processed using a Threshold estimation based on the short time energy and spectral centroid, silence detection and removal technique and a windowing technique. Such pre-processing techniques cleared the noise that existing in the voice file, which was depicted in figure 3 & 4.

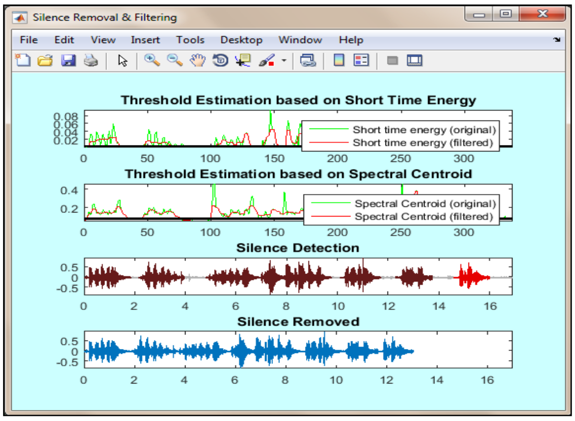

Figure 3: Silence Removal

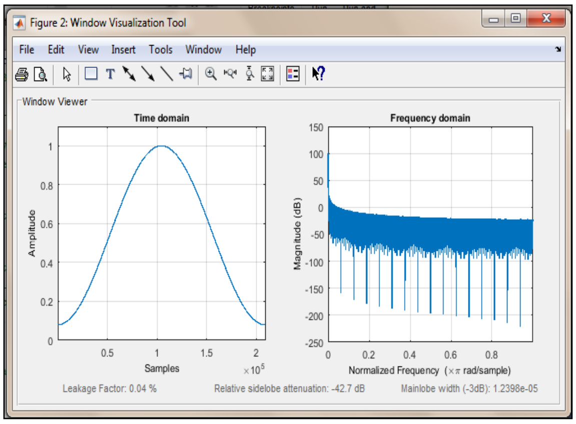

Figure 4: Windowing

In the third Phase, Mel Frequency Cepstrum Coefficients were used to find the Mel Frequency Cepstrum, which equalizes the input voice frequency bandwidth. It describes the power spectral envelope of a single frame. Here 12 Mel Frequency Cepstrum Coefficients were generated from the Log filter bank. Figure 5 & 6 represents the Mel – Frequency Cepstrum Coefficient view for the input.

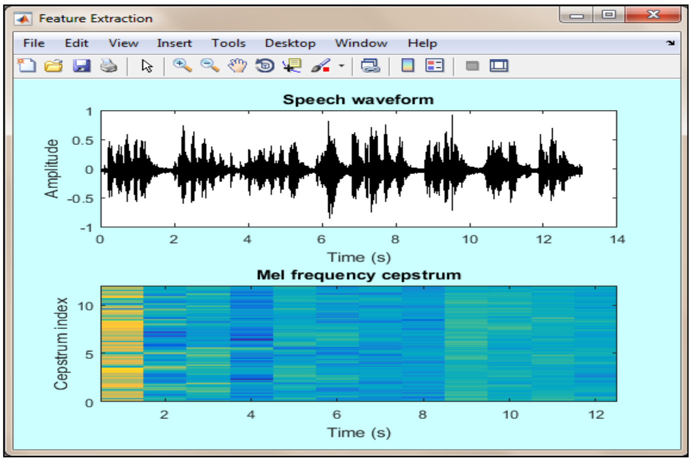

Figure 5: Normal speech signal vs Mel Frequency Cepstrum

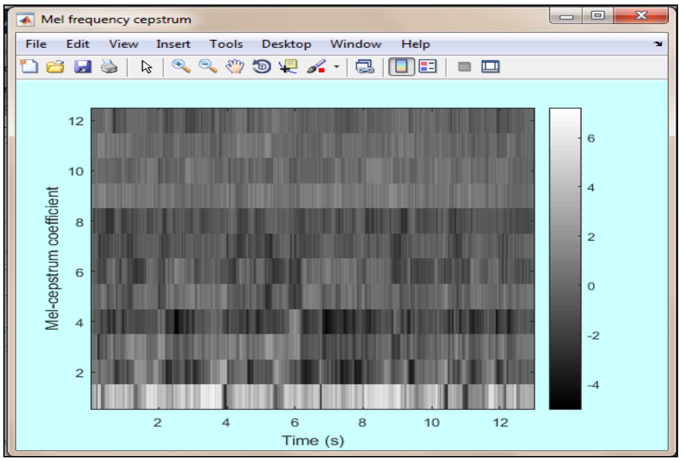

Figure 6: 12 Mel-Frequency Cepstrum Coefficient

In the fourth Phase, Support Vector Machine classification algorithm is used to classify the voice signals as normal or pathology based on the features like MFCC and Jitter extracted in the previous phase. The test data set were compared and models derived from the training data set and indicated by o's and 1's, where o indicates pathology frequency and 1 indicated normal frequency. Thus based on the majority of the frequency's the final decision will be concluded as Normal or Pathology.

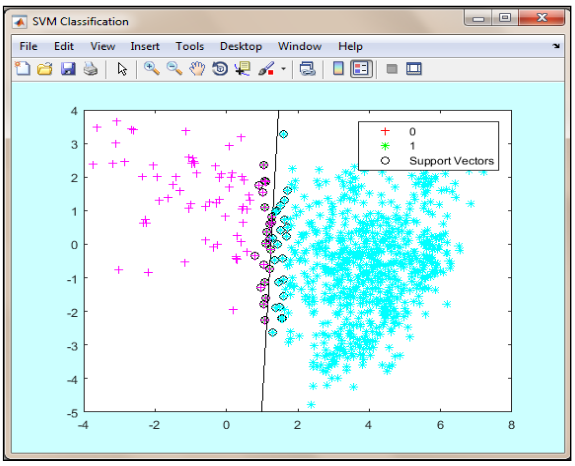

Figure 7: SVM Classification of Normal Voice

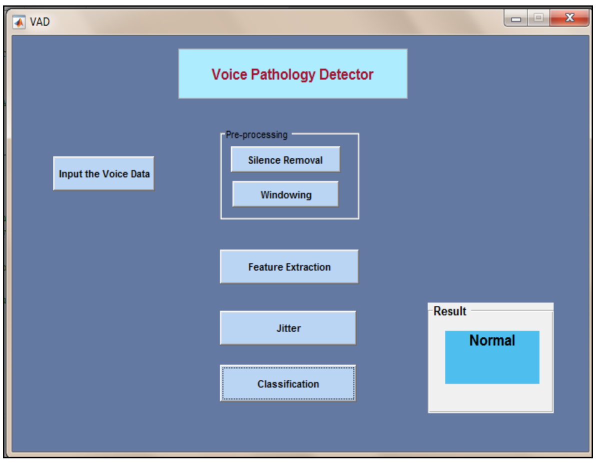

Figure 8: Sample result for a Normal voice

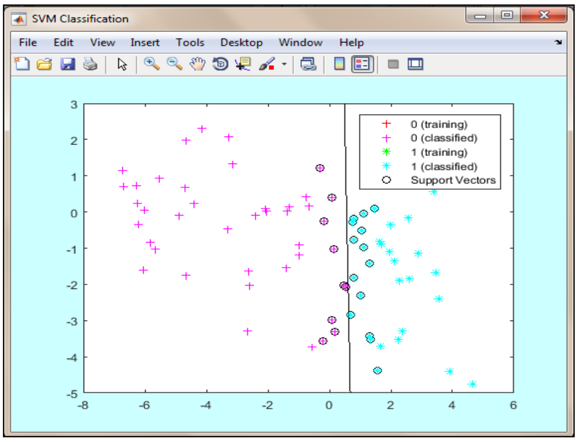

Figure 9: SVM Classification of Pathology voice

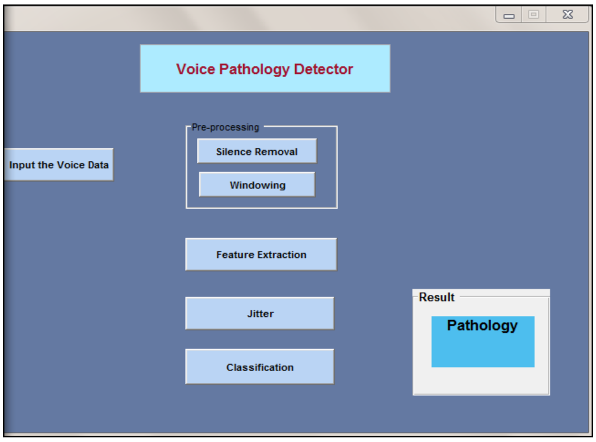

Figure 10: Sample result for a Pathology voice

VI. Conclusion

In this Experimental Study, a Voice Pathology Identification System using SVM Classifiers were proposed. Some sample .wav files were considered as own Database and taken for the experiment. The Pre-processing techniques like silence removal, filtering and windowing were done. Mel-frequency Cepstrum was taken as Feature Extraction Techniques, Classification technique adopted here was SVM classifier. Thus the Workflow of the voice analysis Research was derived and explained in this paper. By analyzing the above work, it shows that the system accurately processed and classified the voice samples into normal and pathology.

VII. Future Work

In this experimental study, we implemented only one methodology for classification and we used our own database for the experiments. In future, we wish to compare efficient classification techniques for finding the voice pathology and to adopt a standard database for our further experiments.

References

1. Rose, M., Ferguson, A., Power, E., Togher, L. and Worrall, L.," Aphasia rehabilitation in Australia: Current practices, challenges, and future directions", International Journal of Speech-Language Pathology, [Early On-line], 1–12, 2013.
2. Sonu, R. K. Sharma "Disease Detection Using Analysis of Voice Parameters'' International Journal of Computing Science and Communication Technologies, VOL.4 NO. 2, January 2012.
3. Alireza A. Dibazar, Theodore W. Berger, and Shrikanth S. Narayanan "Pathological Voice Assessment" IEEE EMBS, NEW YORK, 2006.
4. N. Saenz_Lechon, et al., "Methodological issues in the development of automatic systems for voice pathology detection," Biomed. Signal Processing Control 1, PP. 120– 128, 2006.
5. C. Pend, Q. J. Xu, B. K. Wan and W. X. Chen, "Pathological Voice Classification Based on Features Dimension Optimization," Transactions of Tianjin University, Vol. 13, No. 6, 2007.
6. M. Farra, J. Hernando and P. Ejarque, "Jitter and Shimmer Measurements for Speaker Recognition," TALP Research Center, Department of Signal Theory and Communications, Universitat Politecnica de Catalunya, Barcelona.
7. D. Pravena, S. Dhivya and A. Durga Devi, "Pathological Voice Recognition for Vocal Fold Didease," International Journal of Computer Applications (0975-888), Vol. 47, No. 13, 2012.
8. Raissa Tavares , Nathalia Monteiro , Suzette Correia , Silvana C. Costa , Benedito G. Aguiar Neto and Joseana Macedo Fechine "Optimizing laryngeal pathology detection by using combined cepstral features'' Proceedings of 20th International Congress on acoustics, ICA 2010 23- 27 August 2010, Sydney, Australia ICA 2010.
9. Julian D. Arias-Londono, Juan I. Godino-Llorente, German Castellanos- Dominguez, Nicolas Saenz-Lechon, Victor Osma-Ruiz "Complexity Analysis of Pathological Voices by means Markov Entropy measurements'', 31st Annual International Conference of the IEEE EMBS Minneapolis, Minnesota, USA, September 2-6, 2009.
10. Darcio G. Silva, Luıs C. Oliveira and Mario Andrea "Jitter Estimation Algorithms for Detection of Pathological Voices" Hindawi Publishing Corporation, EURASIP Journal on Advances in Signal Processing Volume, Article ID 567875, 2009.
11. Miltiadis Vasilakis, Yannis Stylianou "Voice Pathology Detection Based on Short-Term Jitter Estimations in Running Speech'' Folia Phoniatr Logop, PP. 153–170, 2009.

12. Jacques Koremana, Manfred Putzer, Manfred Just "Correlates of Varying Vocal Fold Adduction Deficiencies in Perception and Production: Methodological and Practical Considerations" Folia Phoniatr Logop, PP: 305–320, 2004.
13. J. H. L. Hansen et al., "A Non_Linear operator based speech feature analysis method with application to vocal fold pathology assessment," IEEE Trans. Biomed. Eng., 1995.
14. L. Salhi, T. Mourad, and A. Cherif, "Voice Disorders Identification using Multilayer Neural Network," Int. Arab. J. Inf. Technol., Vol. 7, PP. 177–185, 2010.
15. J. I. Godino_Llorente, P. Gomez_Vilda, N. Saenz_ Lechon, M. Blanco_Velasco, F. Cruz_Roldan, and M. A. Ferrer_Ballester, "Support Vector Machines Applied to the Detection of Voice Disorders", Springer_ Verlag, Berlin, PP. 219–230, 2005.
16. Zulfiqar Ali, Irraivan Elamvazuthi, Mansour Alsulaiman, Ghulam Muhammad, "Detection of Voice Pathology using Fractal Dimension in a Multiresolution Analysis of Normal and Disordered Speech Signals", Journal of Medical Systems, Springer, 2015.
17. Zulfiqar Ali, Irraivan Elamvazuthi, Mansour Alsulaiman, Ghulam Muhammad, "Automatic Voice Pathology Detection with Running Speech by Using Estimation of Auditory Spectrum and Cepstral Coefficients Based on the All-Pole Model", Journal of Voice, Elsevier, 2015.
18. Zulfiqar Ali, Mansour Alsulaiman, Irraivan Elamvazuthi, Ghulam Muhammad, Tamer A. Mesallam, Mohamed Farahat, Khalid H. Malki, "Voice Pathology Detection based on the Modified Voice Contour and SVM", Biologically Inspired Cognitive Architectures (BICA), Elsevier, 2015.
19. Zhijian Wang, Ping Yu, Nan Yan, Lan Wang, Manwa L. Ng, "Automatic Assessment of Pathological Voice Quality Using Multidimensional Acoustic Analysis Based on the GRBAS Scale. Signal Processing Systems 82(2), PP: 241-251, 2016.
20. Ping Yu, Zhijian Wang, Shanshan Liu, Nan Yan, Lan Wang, Manwa L. Ng, "Multidimensional acoustic analysis for voice quality assessment based on the GRBAS scale", ISCSLP, PP: 321-325, 2014.