Voice Disorder Detection Based On Acoustic Analysis And Optimized Back Propagation Neural Network

Dr. N.A.Sheela Selvakumari, Dr.V.Radha,

Abstract: The diagnosis of voice disorders through aggressive medical techniques seems painful for patients. Hence, automatic speech recognition and disorder identification methods have drawn much interest in the recent years and have proved to be successful. In this Paper, voice recordings are taken from the Saarbruecken Voice Database. The signals are preprocessed to remove silence and de-noised using Hybrid Wiener Filter Discrete Wavelet Transforms (HWF DWT). Features are extracted using Cat Swarm Optimization Mel Frequency Cepstrum Coefficients (CSOMFCC). Finally, the features are classified using Classification using Modified Optimized Back Propagation Network Disorder voice Classification (MOBPNDC). The classification scheme outperforms the existing Support Vector Machine (SVM) and Back Propagation Neural Network (BPNN) methods in terms of Accuracy, Precision, Recall, F-Measure and Time period.

Keywords: Voice Pathology, Mel-Frequency Cepstrum, Cat Swarm Optimization Classification, SVM, BPNN.

1. INTRODUCTION

Pathological voice recognition has drawn much attention in the last decade. Speech processing is an outstanding tool for voice disorder detection and classification. The most interesting works are focused on the Parkinson’s Disease (PD), Multiple Sclerosis (MS) and other diseases that belong to the class of neurodegenerative diseases affecting the patients’ speech, motor and cognitive capabilities [5] & [6]. Speech analysis is complex and has been ignored for an extended period of time. Voice pathologies are a consequence of improper use of voice, stress, inhalation of tobacco smoke, gastric reflux and hormonal problems. These diseases typically affect the vocal folds and are detected by direct laryngoscopy, which visualizes the vocal folds using a camera. This method is aggressive and painful for the patient depending on the equipment used, requiring a local anesthetic. Test can also be performed by indirect laryngoscopy using a mirror which, which is less aggressive, involving less amounts of local sedatives. But this equipment is expensive and involves high maintenance costs. Assessing the features of a voice disorder aids the pathologist to identify the voice disorder, decide the form of treatment and formulate prognosis. An efficient, inoffensive, easy and low-priced method for pathologies recognition may help in initial evaluation and serve as a complementary method for the voice disorder diagnosis. In this Paper, voice signals taken from the Saarbruecken Voice Database are classified. In preprocessing, silence is removed from the signals and noise is reduced using Hybrid Wiener Filter Discrete Wavelets Transforms (HWF DWT). The predominant features are extracted using Cat Swarm Optimization Mel Frequency Cepstrum Coefficients (CSOMFCC). Finally, the classification of the features is done using Classification using Modified Optimized Back Propagation Network Disorder voice Classification (MOBPNDC).

2. LITERATURE REVIEW

In the literature, several work done in pathological voice identification focus on the categorization of healthy and unhealthy subjects, without acknowledging pathology. Some authors in pathological voice identification [2], [20], [3], [18] have used features like pitch jitter, shimmer, Harmonic-to-Noise Ratio (HNR) and also energy spectrum to identify disorders. Some authors have used MFFC [9] Wavelet analysis [11]. Features like Jitter and Shimmer will alone not be sufficient to categorize pathologies [8], [18] and [7] have compared different systems and features in pathological voice identification. Spectral modulation and SVM are used by [13] in the identification of polyps using the vowel /a/ from the MEEI corpus [4]. The features are found using spectral modulation and the discrete spectrum of the signal is modelled in sub-bands. An algorithm based on Principal Component Analysis (PCA) is used for dimensionality reduction, followed by feature selection based on mutual information.

Mel Frequency Cepstral Coefficients (MFCC) parameters and wavelets analysis have yielded better results [9], [10], [11], [12]. A voice recognition algorithm proposed by [16] is based on the MFCC coefficients and their first and second derivatives, feature reduction is done using F-ratio and
Fisher’s discriminant ratio and classification is performed using Gaussian Mixture Model (GMM). Voice impairments are detected using MFCC vectors and their first derivative, disregarding the second derivative. [19] have experimented on the ability of an acoustic model that is both time-based and spectral-based measures to keep track of the changes following voice disorder treatment and aid in possible treatment outcomes measure. A weighted, four-factor acoustic algorithm based on shimmer, pitch, sigma, ratio of low-to-high frequency spectral energy, and cepstral peak was used to predict severity in dysphonia [16] have dealt with classification of voice disorder using multi-layer network. The algorithm is a hybrid technique that takes energy coefficients of wavelets as input. Supervised training was conducted on a speech database of several pathological and normal voices, and normal and pathology voices were discriminated, followed by classification of neural and vocal pathologies. Non-invasive diagnostic methods have increased with the demand for simple, rapid and trouble-free tests. With the growth in technology that assists in extraction and signal processing, analytical methods are developed to comprehend the complexity of voice signals. [12] has presented a new idea that characterizes signals of healthy and pathological voice based on two mathematical tools, Wavelet Transform (WT) and Artificial Neural Networks (ANN), out of which the approach using WT seems to be suitable in discriminating healthy and disordered voices. [21] have used MFCCs, Linear Predictive Cepstral Coefficients (LPCCs), Perceptual Linear Predictive (PLP) features, Gammatone Filter outputs, Timbral Texture features, Stationary Wavelet Transform (SWT) based Timbral Texture features and Relative Wavelet Packet Energy and Entropy features were extracted from the Emotional Speech (ES) signals and its Glottal Waveforms (GW). Particle Swarm Optimization based Clustering (PSOC) and Wrapper based Particle Swarm Optimization (WPSO) were propounded to improve the recognizing and selecting ability of the features respectively. Extreme Learning Machine (ELM) was used in the classification of different types of emotions. [2] have proposed an algorithm for voice disorders identification based on two classification algorithms namely, Artificial Neural Networks (ANN) and Support Vector Machine (SVM). Mel Frequency Cepstral Coefficients (MFCC) is used for feature extraction and first and second derivatives. Linear Discriminant Analysis (LDA) is also used for feature selection so as to enhance the performance and minimize the complexity of the algorithm.

3. PROPOSED RESEARCH METHODOLOGY

The proposed classification scheme involves the following steps.

- Preprocessing
- Feature Selection and Extraction from the denoised signal
- Classification

3.1 Preprocessing

Normal as well as pathological (with disorders) voice recordings is taken as input. There may be moments of silence in the recordings. These instants of silence are removed during preprocessing.

Further, noise is reduced using Hybrid Wiener Filter Discrete Wavelet Transforms (HWFDWT) de-noising algorithm. The proposed de-noising algorithm endeavors to remove the Gaussian noise from the signal taken into consideration. The resultant denoised signal is free from noise. De-noising using wavelet techniques is very effective as it is capable of capturing the energy of a signal in limited energy transform values.

The approach of DWT based image de-noising involves the following steps:

- Transformation of the noisy image into frequency domain by DWT
- Application of Wiener filter on each sub-band using local window ‘n x n’
- Performing of inverse DWT to get the de-noised image

Wiener filter is involved in removing Gaussian noise from a corrupted signal based on the statistics assessed from a local neighborhood of the speech. This filter is highly dependent on the noise power, which is the noise variance in a corrupted signal. When the variance is large, the filter does little smoothing. On the other hand, when the variance is small, the filter performs better. This filter yields better results in contrast to other filtering techniques used for speech enhancement. In this paper, this filter with local window is applied on the DWT to remove Gaussian noise from each subband.

3.2 Cat Swarm Optimization Mel Frequency Cepstrum Coefficient (CSOMFCC) based Feature Selection and Extraction

The main aim of feature extraction is to reduce the dimensionality and also extract more useful/dominant information hidden in the signals by circumventing needless or redundant information. In signal processing, signal is pre-processed to remove noise and interferences before performing feature extraction. Cat Swarm Optimization Mel Frequency Cepstrum Coefficients (CSOMFCC) is used to extract the best features from the signal. The extracted features are stored in the database.

Two important features that are considered are:

- Degree of Voicing (DOV)
- Pitch

As described, CSO involves two sub-models, the seeking and the tracing mode. A mixture ratio (MR) is defined to combine these two modes. Cats stay at rest while they are awake. They are very careful in choosing their next point of motion or stay at the same point. Seeking mode is used to represent it as cats spend most of the time in this mode.

3.3. Classification

Once the features are extracted, classification is done based on the selected features. The performance of classifier is dependent on the quality of signals pre-processing and the feature extraction. The features of the voice recordings are classified using Modified Optimized Back Propagation Network Disorder voice Classification (MOBPND). The features are classified as Male, Female or Pathological (Disorder). MOBPND is a modified back propagation learning algorithm which involves functional constraints. It deals with the increase in the convergence...
speed to yield better performance. The cost function is designed for output layer and hidden layer separately. The cost function includes the functional constraints namely, Penalty at the hidden layer, weight decay at the output layer with linear and nonlinear error terms. The cost function grows with the convergence speed of training the hidden layer neural network. The BP algorithm is modified by adding functional constraints [4] to the cost function [3] to train single hidden layer neural network. The algorithm includes the first order derivatives of the hidden neurons’ activation functions and the second order derivatives of the output neurons’ activation functions into a total of linear and nonlinear quadratic errors of output neurons. As the modified training algorithm is independent on the nonlinearity approximation [4], it converges very quickly when compared to the original hybrid algorithms. The quadratic linear error involves a weighted coefficient that plays vital role in convergence.

4. IMPLEMENTATION

In this research, we used two databases, namely Saarbrucken Voice Database (2000 persons) and Private Real time Dataset (80 samples) collected from the Department of Pathology, Faculty of Medical Sciences and Research, Karpagam University, Coimbatore. Voice samples of people, whose age is in the range of 30-35, are considered. This chosen dataset are recorded as voice samples in (.wav) audio format. This sample contains the recordings of vowels /a/, /i/, /u/ and phrases like “good morning”, “how are you”, “hello”, “where are you”, “welcome” produced at different pitch levels like low, normal and high. Also, the length of the audio clips with sustained vowels is 2 seconds, and all recordings are sampled at 50 kHz with 16-bit resolution. These samples mainly concentrate on the five different diseases like Laryngitis, Laryngoceles, Dysphonia, Diplophonia and Chorditis, since these are the major voice disorder types which affect the voice organs like vocal cord, vocal fold, vocal box, bulge appearance in neck and the voice parameters. Finally, the Testing has been carried out with our research methodology.

5. RESULTS AND DISCUSSION

The quality of the produced output is measured using the Accuracy, Specificity and Sensitivity. Based on these parameters the efficiency of the developed system is shown by following Roc curve. A Receiver Operating Characteristic (ROC) Curve is a way to compare diagnostic tests. It is a plot of the true positive rate against the false positive rate. A ROC plot shows the relationship between sensitivity and specificity. A ROC curve passes through the upper left corner. Therefore, the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test.

ROC Curve = Plot (FPR, TPR) 

Figure 2. Shows the performance analysis of Pathological voices classification with respect to the Sensitivity and Specificity done for the developed system. The parameters performance of the three different classifiers (SVM, BPNN and MOBPND) receiver operating characteristic (ROC) curves is produced by the MOBPND, support vector machine (SVM), and BPNN. It clearly shows the proposed algorithm outperforms well in terms of classification Specificity and Sensitivity.

Figure 3. Shows the performance analysis of Pathological voices classification Comparison of two different data set (namely Saarbrucken Voice Database (2000 persons) and Private Real Time Dataset (80 samples)) done for the developed system considering certain performance parameters like Sensitivity, Specificity and Accuracy. The proposed Modified Optimized Back Propagation Network Disorder Voice Classification MOBPND, plotted with the other two classifiers SVM and BPNN. It clearly shows the proposed algorithm outperforms high classification Sensitivity, Specificity and Accuracy.

6. CONCLUSION

The voice signals were preprocessed to remove silence and de-noised using Hybrid Wiener Filter Discrete Wavelet Transforms (HWFDT). Cat Swarm Optimization Mel Frequency Cepstrum Coefficients (CSOMFCC) based scheme was used in feature extraction. Finally, the features were classified using Classification using Modified Optimized Back Propagation Network Disorder voice Classification (MOBPND). From the results, it is evident that the proposed MOBPND classification scheme
Outdoes the existing Support Vector Machine (SVM) and Back Propagation Neural Network (BPNN) methods in terms of Accuracy, Specificity and Sensitivity.

REFERENCES


