OSTIS-2013
(Open Semantic Technologies for Intelligent Systems)

УДК 004.822:514

AN ANN-BASED METHOD FOR VOCAL FOLD PATHOLOGY DIAGNOSIS

Vahid Majidnezhad [*], Igor Kheidorov[*]

[*] *United Institute of Informatics Problems, National Academy of Sciences, Minsk, Belarus*
vahidmn@yahoo.com
ikheidorov@sakrament.com

Abstract: There are different 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, the first and second stages play a critical role in performance and accuracy of the classification system. In this paper 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 Principal-Component Analysis (PCA) is used for feature reduction. An Artificial Neural Network is used as a classifier for evaluating the performance of our proposed method.
Keywords: Wavelet Packet Decomposition, Mel-Frequency-Cepstral-Coefficient (MFCC), Principal-Component Analysis (PCA), Artificial Neural Network (ANN).

INTRODUCTION

Speech 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 detection of pathologies. Laryngoscopy and electromyography 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 a pathology. Moreover, these diagnostic means may cause the patients feel much discomfort and distort the actual signal so that it may be lead to incorrect diagnosis as well [Alonso et al., 2001] [Ceballos et al., 2005] [Ceballos et al., 1996] [Adnene et al., 2003]. Acoustic analysis as a diagnostic method has no drawbacks, peculiar to the above mentioned methods. It is a non-invasive diagnostic technique that allows pathologists to examine many people in short time period with minimal discomfort. The general scheme of vocal fold pathology diagnostic methods consists of three stages which are feature extraction, feature reduction and classification.

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 et al., 2004] [Rosa et al., 2000] [Mallat, 1989] [Wallen et al., 1996]. Feature extraction, using the above mentioned parameters, has shown its efficiency for a number of practical tasks. In the proposed method, we have used the Mel-Frequency-Cepstral-Coefficients (MFCCs), Energy and Shannon Entropy parameters for creating the 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 Linear Discriminant Analysis (LDA) [Ji-Yeoun et al., 2007]. In the proposed method we have used PCA for feature reduction. 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 the ANN for the classification purpose.

1. Methodology

The wavelet transform, as was shown in [Manfredi, 2000], is a flexible tool for analysis of speech signals. 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 et al., 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.

The block diagram of our proposed method is illustrated in Fig. 1. In the first stage, by the use of MFCC and Wavelet Packet Decomposition, feature vector containing 1035 features is made. In the second stage, by the use of PCA method, the dimension of feature vector is reduced. In the last stage, by the use of Artificial Neural Network (ANN), the speech signal classified into two classes: pathological or healthy.

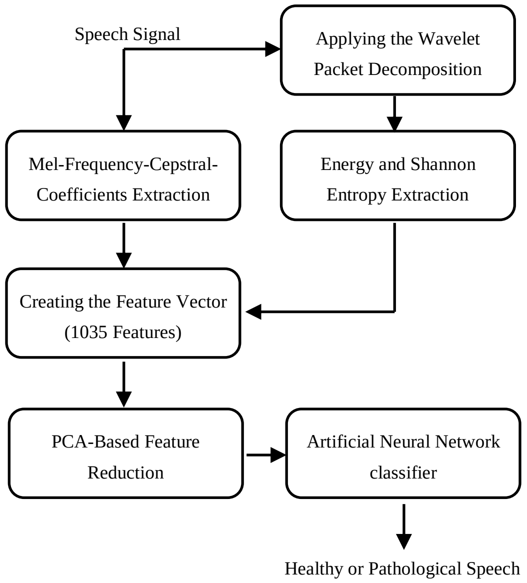

Figure 1- The scheme of the proposed method for detection of vocal fold pathology.

1.1. Feature extraction

As it is shown in Fig. 1, 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, 511 energy features along with 511 Shannon entropy features are extracted. Finally, by the combination of these features, the initial feature vector with the length of 1035 features is created.

1.1.1. *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 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 [Ji-Yeoun et al., 2007] to classify the speech into pathology and healthy class.

1.1.2. *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. In this study, mother wavelet function of the tenth order Daubechies has been chosen and the signals have been decomposed to 8 levels. The mother wavelet used in this study is reported to be effective in voice signal analysis [Guido et al., 2005] [Umapathy et al., 2005] and is being widely used in many pathological voice analyses [Fonseca et al., 2007].

1.2. 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”. 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] [Marinaki et al., 2004].

PCA 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. 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 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.

1.3. Artificial Neural Network

An artificial neural network (ANN) as a computing system is made up of a number of simple, and highly interconnected processing elements, which processes information by its dynamic state response to external inputs. Neural networks are better suited for achieving human-like performance in the fields such as speech processing, image recognition, etc. Processing elements in an ANN are also known as neurons. These neurons are interconnected by means of information channels called interconnections. Each neuron can have multiple inputs; while there can be only one output. Inputs to a neuron could be from external stimuli or could be from output of the other neurons. Copies of the single output that comes from a neuron could be input to many other neurons in the network. When the weighted sum of the inputs to the neuron exceeds a certain threshold, the neuron is fired and an output signal is produced. The network can recognize input patterns once the weights are adjusted or tuned via some kind of learning process [Lee et al., 1992].

2. Experiments and Results

The database was created by specialists from the Belarusian Republican Center of Speech, Voice and Hearing Pathologies. We have selected 75 pathological speeches and 55 healthy speeches which are related to the vowel “a”. All the records are wave files in PCM format. The whole scheme of our proposed method is illustrated in Fig. 1. We have adopted a 10 fold cross-validation scheme to assess the generalization capabilities of the system in our experiments.

It is necessary to know how many neurons in the hidden layer are required to achieve the optimal results. So, in the first experiment the numbers of neurons in the hidden layer are investigated and the results are shown in Fig. 2. As it is obvious, using 21 neurons in the hidden layer lead to the optimal case which is 74.62% of accuracy. Finally, according to the result of the first experiment, an ANN with 21 neurons is created for using in the second experiment. The goal of the second experiment is to reduce the initial feature vector’s length so that the accuracy of the classification phase increases. For this purpose, the conventional PCA method is used and the results are shown in Fig. 3. As it is obvious, using the 51 features lead to the best result which is 89.23% of accuracy. The 51 selected features, by the means of PCA method, are shown in the table 1.

Table 1- The selected features for the construction of feature vector.

Feature Reduction Method	The selected features	Accuracy
PCA (feature vector length=51)	The 1-3, 5-6 and 10th coefficients of MFCC. Energy at the 19, 32-33, 35, 65-66, 70-71, 129, 131, 135, 140, 256-258, 260, 262-263, 265-266, 268, 271 and 274th nodes of WP Tree. Entropy at the 8, 16-17, 32-33, 64, 66, 69, 128-129, 132, 134-135, 256-259, 261, 264, 268, 270 and 280th nodes of WP Tree.	89.23%
None (feature vector length=1035)	All the 511 entropy & 511 energy & 13 MFCC coefficients.	74.62%

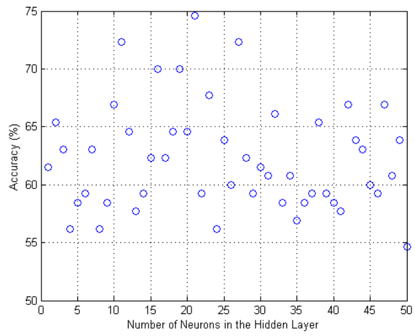

Figure 2 - The Classification Results based on the number of Neurons

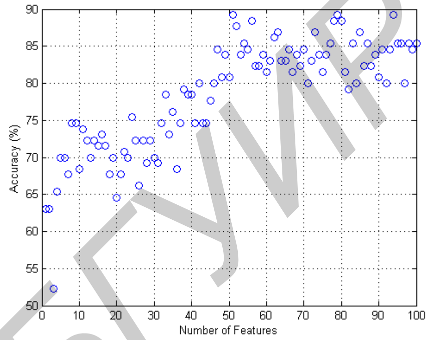

Figure 3 - The Classification Results based on the number of Features

3. Conclusion

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. In this article, an ANN-Based method for vocal fold pathology diagnosis is proposed so that in the proposed scheme, Mel-Frequency-Cepstral-Coefficients along with the wavelet packet decomposition are used for feature extraction phase. Also PCA method for the feature reduction phase is used. And finally the ANN is used for the classification phase. Two experiments are designed to investigate the optimal case for the numbers of neurons in the hidden layer of the ANN and also the optimal case for the feature vector length as the input of ANN.

REFERENCES

[Alonso et al., 2001] Alonso, J.B. Automatic Detection of Pathologies in the Voice by HOS Based Parameters / J.B. Alonso, J.D. Leon, I. Alonso, M.A. Ferrer // EURASIP Journal on Applied Signal Processing, - 2001 - No. 4 – pp. 275-284.

[Ceballos et al., 2005] Ceballos, L.G. A Non-Linear Based Speech Feature Analysis Method with Application to Vocal Fold Pathology Assessment. / L.G. Ceballos, J. Hansen, J. Kaiser // IEEE Trans. Biomedical Engineering. – 2005 – No. 45(3) – pp. 300-313.

[Ceballos et al., 1996] Ceballos, L.G. Vocal Fold Pathology Assessment Using AM Autocorrelation Analysis of the Teager Energy Operator/ L.G Ceballos, J. Hansen, J. Kaiser // ICSLP-1996 Proc. – 1996 - pp: 757-760.

[Adnene et al., 2003] Adnene, C. Analysis of Pathological Voices by Speech Processing. / C. Adnene, B. Lamia // Signal Processing and Its Applications. - 2003 – No. 1(1) – pp. 365-367.

[Manfredi, 2000] Manfredi, C. Adaptive Noise Energy Estimation in Pathological Speech Signals. / C. Manfredi // IEEE Trans. Biomedical Engineering. –2000 – No. 47(11) – pp. 1538-1543.

[Llorente et al., 2004] Llorente, J.I.G. Automatic Detection of Voice Impairments by Means of Short-Term Cepstral Parameters and

Neural Network Based Detectors/ J.I.G. Llorente, P.G. Vilda // IEEE Trans. Biomedical Engineering. – 2004 – No. 51(2) – pp. 380-384.

[Rosa et al., 2000] Rosa, M.D.O. Adaptive Estimation of Residue Signal for Voice Pathology Diagnosis. / M.D.O Rosa, J.C. Pereira, M. Grellet // IEEE Trans. Biomedical Engineering – 2000 – No. 47(1) – pp. 96-104.

[Mallat, 1989] Mallat, S.G. A Theory for Multi-resolution Signal Decomposition: the Wavelet Representation. / S.G. Mallat // IEEE Trans Pattern Analysis and Machine Intelligence. – 1989 – No. 11(7) – pp. 674-693.

[Wallen et al., 1996] Wallen, E.J. A Screening Test for Speech Pathology Assessment Using Objective Quality Measures. / E.J Wallen, J.H. Hansen // ICSLP Proc. – 1996 – pp. 776-779.

[Chen et al., 2007] Chen, W. SVM-based identification of pathological voices. / W. Chen, C. Peng, X. Zhu, B. Wan, D. Wei // Proceedings of the 29th Annual International Conference of the IEEE EMBS. – 2007 – pp. 3786-3789.

[Go´mez et al., 2005] Go´mez, P. Principal component analysis of spectral perturbation parameters for voice pathology detection. / P. Go´mez, F. Dı´az, A. A´lvarez, K. Murphy, C. Lazaro, R. Martinez, V. Rodellar // Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems. – 2005 - pp: 41-46.

[Michaelis et al., 1998] Michaelis, D. Selection and combination of acoustic features for the description of pathologic voices. / D. Michaelis, M. Frohlich, H.W. Strube // Journal of the Acoustical Society of America. – 1998 – No. 103(3) – pp. 1628-1639.

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

[Ji-Yeoun et al., 2007] Ji-Yeoun, L. Classification of pathological and normal voice based on linear Discriminant analysis. / L. Ji-Yeoun, J. SangBae, H. Minsoo // Proceedings of the Eighth International Conference on Adaptive and Natural Computing Algorithms. – 2007 - pp. 382-390.

[Ritchings et al., 2002] Ritchings, R.T. Pathological voice quality assessment using artificial neural networks. / R.T. Ritchings, M.A. McGillion, C.J. Moore // Medical Engineering & Physics. – 2002 – No. 24(8) – pp. 561-564.

[Li et al., 2003] Li, T. A comparative study on content based music genre classification. / T. Li, M. Oginara, Q. Li // Proc. Of the 26th annual int.ACM SIGIR conf. on Research and development in information retrieval. – 2003 – pp. 282–289.

[Tzanetakis et al., 2002] Tzanetakis, G. Musical genre classification of audio signals. / G. Tzanetakis, P. Cook // IEEE Trans. on Speech and Audio Processing. – 2002 – No. 10(5) – pp. 293-302.

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

[Cavalcanti et al., 2010] Cavalcanti, N. Comparative Analysis between Wavelets for the Identification of Pathological Voices. / N. Cavalcanti, S. Silva, A. Bresolin, H. Bezerra, A. Guerreiro // Proceedings of the 15th Iberoamerican congress conference on Progress in pattern recognition, image analysis, computer vision, and applications. – 2010 – pp. 236-243.

[Herisa et al., 2009] Herisa, H.K. Optimal feature selection for the assessment of vocal fold disorders. / H.K. Herisa, B.S. Aghazadeh and M.N. Bahrami // Computers in Biology and Medicine. – 2009 – No. 39(10) – pp. 860-868.

[Fonseca et al., 2007] Fonseca, E. Wavelet time frequency analysis and least squares support vector machines for identification of voice disorders. / E. Fonseca, R.C. Guido, J.C. Pereira, P.R. Scalassarsa, C.D. Maciel, J.C. Pereira // Computers in Biology and Medicine. – 2007 – No. 37(4) – pp. 571-578.

[Guido et al., 2005] Guido, R.C. Trying different wavelets on the search for voice disorders sorting. / R.C. Guido, J.C. Pereira, E. Fonseca, F.L. Sanchez, L.S. Vierira // 37th IEEE International Southeastern Symposium on System Theory. – 2005 – pp. 495-499.

[Umapathy et al., 2005] Umapathy, K. Feature analysis of pathological speech signals using local discriminant bases technique. / K. Umapathy, S. Krishnan // Medical and Biological Engineering and Computing. – 2005 – No. 43(4) – pp. 457-464.

[Lee et al., 1992] Lee, K.Y. SHORT-TERM LOAD FORECASTING USING AN ARTIFICIAL NEURAL NETWORK. / K.Y. Lee, Y.T. Cha, J.H. Park // Transactions on Power Systems. – 1992 – No. 7(1) – pp. 124-132.

ДИАГНОСТИКА ПАТОЛОГИИ ГОЛОСОВОГО ТРАКТА НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ

Вахид Маджиднежад *, Игорь Хейдоров *

** Объединенный институт проблем информатики Национальной академии наук Беларуси (ОИПИ), Минск, Беларусь*

vahidmn@yahoo.com

ikheidorov@sakrament.com

В этой статье представляется метод искусственных нейронных сетей для решения задач диагностики патологии голосового тракта.

ВВЕДЕНИЕ

В настоящее время существуют разнообразные медицинские методы для непосредственного изучения и выявления патологий. Однако эти методы обладают рядом недостатков. Голосовой тракт человека труднодоступен для визуального осмотра, что не позволяет использовать его для выявления патологии. Кроме того, различные приборы для диагностики могут вызвать у пациента значительный дискомфорт, что приводит к искажению реального сигнала и в конечном счете неправильному диагнозу. Предлагаемый в данной статье акустический анализ является неинвазивным методом диагностики и не имеет недостатков, свойственных указанным выше способам.

ОСНОВНАЯ ЧАСТЬ

Для решения задачи диагностики патологии голосовых связок была разработана искусственная нейронная сеть. Для нахождения оптимального числа нейронов в скрытом слое ИНС и длины вектора признаков, подаваемого на вход ИНС, было проведено два эксперимента. В соответствии с результатами первого эксперимента была построена ИНС с 21 нейронами, которая была использована в последующих экспериментах. Цель второго эксперимента заключается в сокращении размерности вектора признаков таким образом, чтобы улучшить точность классификации. Для этого был использован метод главных компонент. Результаты эксперимента показали, что использование для классификации 51 признака приводит к наилучшим результатам(точность 89.23%).

ЗАКЛЮЧЕНИЕ

В данной работе был предложен новый тип вектора признаков на основе вейвлет-разложения пакетов и мел-частотных кепстральных коэффициентов (MFCCs). Для уменьшения размерности вектора признаков был использован метод главных компонент, а в качестве классификатора была использована искусственная нейронная сеть.