Identification of Voice Disorders using Speech Samples

Jagadish Nayak
Dept. of E&C Engg, , MIT Manipal-576119
jag.nayak@mit.manipal.edu ,
P Subbanna Bhat
Dept. of E&C Engg. NITK Surathkal, Srinivasnagar –575025
p_subbannabhat@yahoo.com

Abstract--This paper attempts to identify pathological disorders of larynx using Wavelet Analysis. Speech samples carry symptoms of disorder in the place of their origin. The speech signal is subjected to wavelet analysis, and the coefficients are used to identify disorders such as Vocal Fold Paralysis. Multilayer Artificial Neural Network is used for classification of normal and affected signals.

1. INTRODUCTION

Physicians often use invasive techniques like Endoscopy to diagnose the symptoms of vocal fold disorders. However, it is possible to identify disorders using certain features of speech signals [1,2]. This paper uses wavelet analysis technique to extract a feature vector from speech samples, which is used as input to a Multilayer Neural Network classifier. Wavelet analysis provides a two-dimensional pattern of wavelet coefficients. The energy content of Wavelet coefficients at various level of scaling is used to formulate a feature vector of speech sample. Attempt is made to use this feature vector as a diagnostic tool to identify pathological disorders in the larynx. A three layer feed forward network with sigmoid activation is used for classification. Generalized Back Propagation Algorithm (BPA) is used for training of the network.

Usually, pathological disorders in larynx reflect in the quality of speech signal. The common disorders are acute infective laryngitis, chronic non-specific laryngitis, vocal fold paralysis etc. The disorders show up in speech signal in the form of disruption of Phonation (Hoarseness), Articulation (Dysarthria) and Resonance. The vocal cords are in the form two elastic bands of muscle tissue located in the larynx (voice box) directly above the trachea (windpipe). Voice is produced when air held in the lungs is released through the closed vocal cords, causing them to vibrate. When a person is not speaking, the vocal cords remain apart to accommodate breathing.

Vocal cord paralysis is a disorder that occurs when one or both of the vocal cords do not open or close properly. It is a common disorder in old age, and can occur with varying intensity. Typically vocal cord paralysis result in hoarseness of voice and inability to project the voice loudly.

0-7803-7651-X/ 03/ $17.00 © 2003 IEEE

Pitch (frequency of vibration of vocal cords) is an important feature of voiced speech signal. For normal speech, the vocal fold would vibrate at regular intervals with a fundamental frequency F_0 that is called the *Pitch*. In the event of paralysis, voiced speech samples contain irregular pitch due to improper functioning of vocal cords. Vocal fold paralysis would produce noise-like signal, which is almost similar to unvoiced speech [3]. Therefore, Pitch alone cannot be used as an indicator of disorder.

Speech being a highly non-stationary signal, Fourier Transform is not a very useful tool for analysis. Whereas, Wavelet Transform approach is a good tool for analysis of non-stationary signals, as it is useful in localizing a symptom both in time and frequency scales. In disordered speech, the non-stationary behavior of the Pitch can be analyzed using Wavelet Transforms. The energy distribution in various levels of scaling can provide information about localized irregularity in vocal fold vibration. The energy content in each scale is extracted and used as feature vector for the ANN classification.

2. WAVELET ANALYSIS

Wavelet analysis essentially involves comparing the signal with a chosen (finite duration) *Wavelet*; and recording the correlation coefficient. The Wavelet is then *translated* (shifted) along the time scale by a short distance and again the coefficient is evaluated. The process is continued to cover the entire signal duration. The General definition of the wavelet transform is given as:

$$W(a,b) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{a}} \psi^* \left(\frac{t-b}{a} \right) dt \qquad (1)$$

Where a and b are real and * denotes complex conjugate and $\psi(t)$ is the wavelet function. In Dyadic wavelet transform, frequency band is divided into the number of levels by the factor of two. Figure 1 shows the wavelet decomposition for DWT with a scale of 2 levels. Here each frequency band is having different resolution [9].

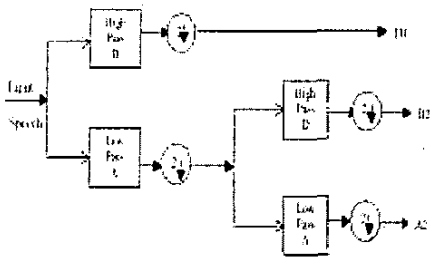

Figure 1. Wavelet Decomposition

3.FEATURE EXTRACTION

There are many parameters like pitch, jitter and shimmer, which can be used as part of feature vector for classification. Since speech is a highly random signal, sometimes pathological speech jitter and pitch may match the normal speech. Here normalized energy across the scales is being chosen as parameter of feature vector. However, in the example, the feature vector comprises of normalized energy content at various levels (scaling factor) of Wavelet coefficients. The energy content of the signal is normalized against total energy content in the signal (Figure 2). The wavelet coefficients are extracted using Filter Bank.[4]

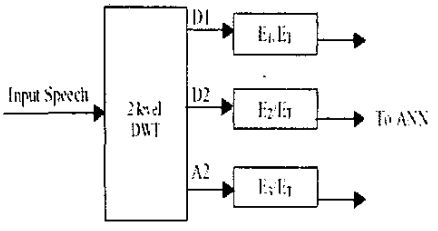

Figure 2 Feature Extraction

The observed periodicity in the lower band scale from the Wavelet transform is more dominant than in the higher band scale. The normal speech is having consistent period but the periodicity is decreasing in the pathological speech. The normalized energy across the scale *i* is

$$E_N(i) = \frac{E_i}{E_T} \qquad (2)$$

where i = 1 2 3.

E_T =Total Energy across all the levels.

E_i = Energy at each level. This shows how well the signals energy is spread across the scale.

4. CLASSIFICATION

A multilayer feed forward Artificial Neural Network with Generalized Back Propagation Algorithm is implemented for classification (Figure 3). The network has three input nodes and one output neuron, and five neurons in the hidden layer. The input is the feature vector obtained from Wavelet decomposition. The network weights updating is given by [7,8]

$$\Delta W_{kj} = \frac{1}{2}\eta(d_k - z_k)[1 - z_k^2]y_j \qquad (3)$$

for j= *1,2,3* ..... ; k=1,2,3.....

Thus the (k x j) weights of output layer are updated. Next the input layer weights *Vji* are updated using Eqn (4).

$$\Delta V_{ji} = \frac{1}{4}\eta[1 - y_j^2]x_i \sum_{k=1}^{K}(d_k - z_k)[1 - z_k^2]W_{kj} \qquad (4)$$

for i= *1, 2,3* ..... j= *1,2,3* .....

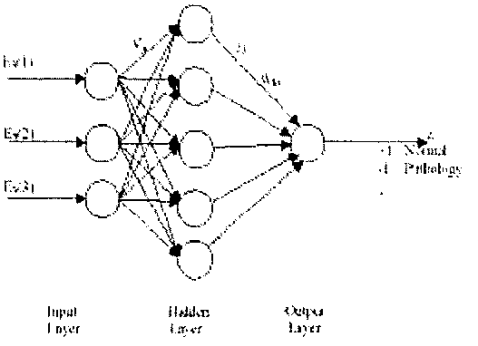

Figure 3. Multilayer ANN Structure

5. RESULTS

A total of 50 data samples were used (25 *Normal* and 25 vocal fold *Paralysis* data) for training the network [6]. The plot of normalized energy $E_N(i)$ versus scale value is shown in Figure 4. It can be seen that from energy level point of view, *Normal* and *Paralysis* samples are grouped separately. Also that in the first scale, normal speech signals shows higher Energy levels than the pathological signals. Similarly the output D2 and A2 gives better classification among normal and pathological data. In this level it is found that pathological data has higher values. This type of variation in normalized energy across the scale is very useful for classification using ANN.

The feature vector (Normalized Energy across each scale) is used as input to the ANN. Once ANN is trained for classification, it is tested with 20 samples of test data (10 *Normal* and 10 *Paralysis* data). The results are shown in Table 1. The classification results obtained show an accuracy of 80% .

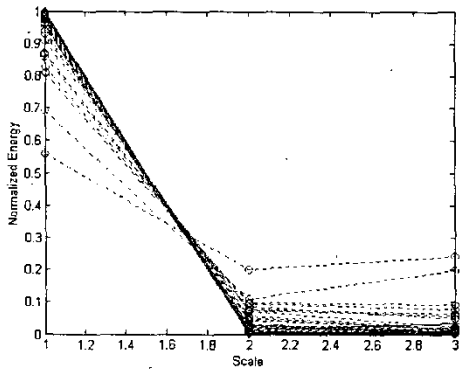

Figure 4. Variation in Normalized energy across the scale(' *' – Normal 'o'—Pathological)

6. CONCLUSION

It is suggested that a feature vector based on wavelet coefficients may be useful for classification of normal and pathological speech data. At a preliminary level, the speech data is classified into two classes under the heads Normal and Paralysis.

The multilayer-feedforward network with BPA can be used as a classifier, but its effectiveness depends upon the quality and size of training data set. In the present case, the feature vector has only three components entirely based on wavelet coefficients. However, if additional parameters of speech sample – such as Pitch, jitter, shimmer are added to the feature vector, the training of the network may be more effective, and the classification is likely to improve.

Table 1

Speech Samples	No of Test Data	Correct Classification	% Result
Normal	10	10	100
Pathology	10	8	80

REFERENCES

[1] F. Plant, H Kessler , B Cheetham, J Earis, " Speech Monitoring of Infective Laryngitis" , *Proceedings of ICSLP96, Philadelphia* , pp. 749 – 752 , 1996

[2] M.N. Viera, F.R. McInnes, M.A. Jack " Robust F0 and Jitter estimation in the Pathological voices ", *Proceedings of ICSLP96, Philadelphia,* pp.745 –748, 1996.

[3] Lowrence R. Rabiner and Ronald W. Schaffer , "Digital Processing of Speech Signals" *Prentice Hall.*

[4] Raghuveer M Rao , Ajith Bopardikar "Wavelet Transform Introduction to Theory and Applications", *Pearson Education Asia.*

[5] Cheol-Woo Jo , Dae-Hywn Kim, "Analysis of Disordered Speech signal using Wavelet Transform" *ICSLP 98,*

[6] "Disordered Voice Data base", Ver 1.03, *Kay Elemetrics Corp. 1994.*

[7] B. Yajnanarayana, "Artificial Neural Networks" *Prentice Hall of India, 1999*

[8] Simon Haykins " Neural Networks " *Prentice Hall , 1999.*

[9] S. Mallat, "A Theory for multiresolution signal decomposition: Wavelet representation" , *IEEE Trans. Pattern Analysis and Machine Intelligence.* Vol. 11. No. 7 pp674-693 July 1989.