TIME-DOMAIN FEATURES AND PROBABILISTIC NEURAL NETWORK FOR THE DETECTION OF VOCAL FOLD PATHOLOGY

M. Hariharan[1], M. P. Paulraj[2], Sazali Yaacob[3]

School of Mechatronic Engineering, Universiti Malaysia Perlis (UniMAP)
Jejawi 02600, Perlis, Malaysia
[1]wavelet.hari@gmail.com, [2]paul@unimap.edu.my, [3] s.yaccob@unimap.edu.my

ABSTRACT

Due to the nature of job, unhealthy social habits and voice abuse, people are subjected to the risk of voice problems. It is well known that most of vocal fold pathologies cause changes in the acoustic voice signal. Therefore, the voice signal can be a useful tool to diagnose them. Acoustic voice analysis can be used to characterize the pathological voices. This paper presents the detection of vocal fold pathology with the aid of the speech signal recorded from the patients. The speech samples from Massachusetts Eye and Ear Infirmary (MEEI) database are used to evaluate the scheme. Time-domain features based on energy variation are proposed and extracted from the speech to form a feature vector. In order to test the effectiveness and reliability of the proposed time-domain features, a Probabilistic Neural Network (PNN) is employed. The experimental results show that the proposed features gives very promising classification accuracy and can be effectively used to detect the vocal fold pathology clinically.

Keywords: *Acoustic Analysis, Vocal Fold Pathology, Time-Domain Features, Probabilistic Neural Network*

1.0 INTRODUCTION

Acoustic analysis and detection of vocal fold pathology have undergone substantial research and development in the last three decades. The vocal fold pathology has to be diagnosed in the early stage. To detect the vocal fold pathology, ENT clinicians and speech therapists use subjective techniques or invasive methods such as the direct inspection of the vocal folds and the observation of the vocal folds by endoscopic instruments [1]. These techniques are expensive, risky, time consuming, discomfort to the patients and require costly resources, such as special light sources, endoscopic instruments and specialized video-camera equipment. In order to circumvent the above problems, non-invasive methods have been developed to help the ENT clinicians and speech therapists for early detection of vocal fold pathology.

In the bibliography, there are many algorithms have been found for the automatic detection of vocal fold pathology by means of long-time signal analysis [1-5]. In recent years, more modern approaches have been invented which use short-time speech analysis or Electroglottograph (EGG) signals [6-8]. The short-time acoustical features extracted from the EGG signal can be examined to depict the aspects of normal or abnormal vocal fold vibration motion. The proper diagnosis of vocal fold pathology is essential. This paper presents the detection of vocal fold pathology with the aid of the speech signal recorded from the patients. Time-domain features are proposed and extracted to detect the vocal fold pathology. In order to test the effectiveness and reliability of the proposed time-domain features, a Probabilistic Neural Network (PNN) is employed.

2.0 SPEECH DATA

The speech data are taken from the commercial database distributed by Kay Elemetrics for the classification experiments [9]. The database contains 53 normal and 657 pathological voice samples developed by the Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Labs. In order to increase the size of the normal voices using 53 normal voice samples, all the normal voice samples are segmented to a length of 0.4 sec and it gives 313 normal voice samples. The acoustic samples are the sustained phonation of the vowel /ah/(1-3s) long and reading (12 seconds) of the “Rainbow Passage” from patients with normal voices and a wide variety of organic, neurological, traumatic, and psychogenic voice disorders in different stages. All the speech samples were collected in a controlled environment and sampled with 25 kHz or 50 kHz sampling rate and 16 bits of resolution. In order to test the effectiveness of the method and features, a total of 970 voices samples of sustained phonation of the vowel (657 abnormal+313 normal) are used and downsampled to 16 kHz for our analysis.

3.0 PROPOSED FEATURE EXTRACTION

Feature extraction from the speech signal plays very important role in the area of automatic detection of vocal fold pathology. A great amount of acoustic parameters have been proposed and its effectiveness has been proven by experimental researches. The important parameters are as pitch [10], jitter [11-13], shimmer [11,12], the harmonics-to-noise (HNR) [14,15], and the normalized noise energy (NNE) [16]. All these parameters are based on the fundamental frequency and the correct estimation of fundamental frequency of pathological voices is not easy. This paper proposes a simple feature extraction method that extracts features from the time-domain energy of speech signal in order to detect the vocal fold pathology. All the features are calculated from short-time frames extracted from the speech signals. The short-time frame length is selected as 64 ms (1024 samples per frame), with an overlap of 50% between adjacent frames. Such frame size is selected to ensure, at least, one pitch period per window.

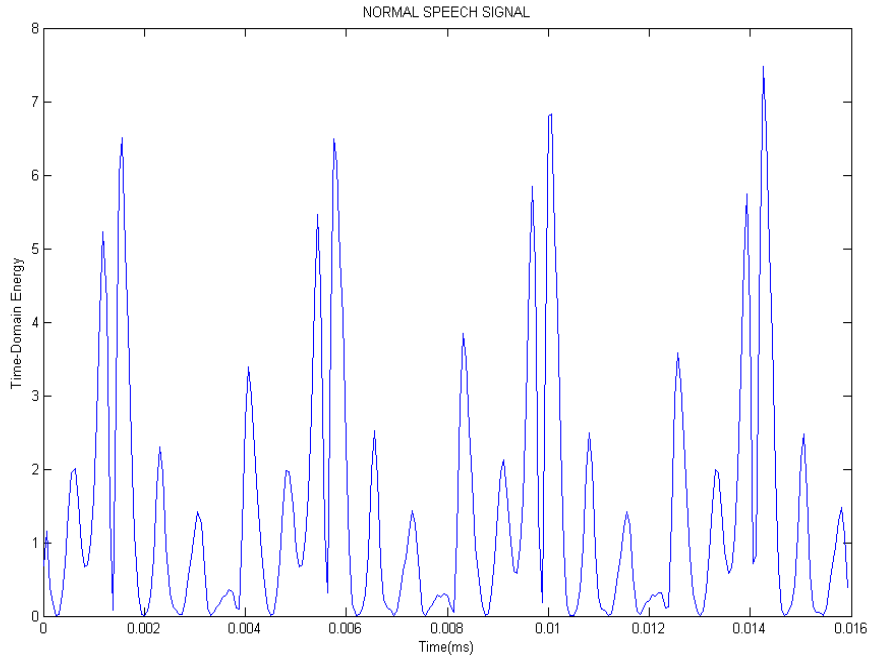

Fig. 1(a) Time-domain energy plot of normal speech signal

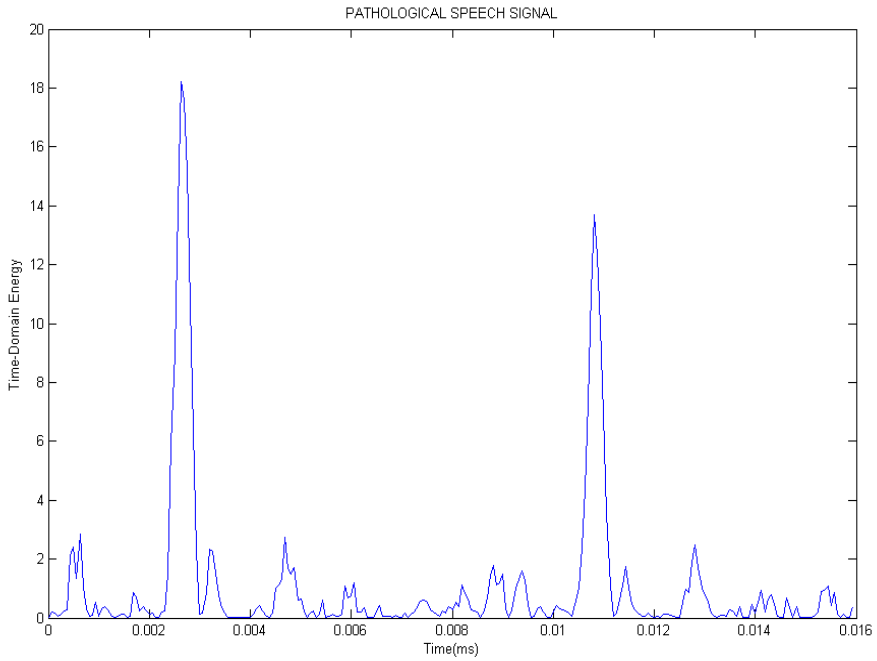

Fig. 1(b) Time-domain energy plot of pathological speech signal

Figure 1(a) and (b) shows the time-domain energy plot of normal and pathological speech signal. These energy plots clearly indicate the significant difference between the normal and pathological speech signal. Fig. 2 shows the energy peaks of a speech signal. Consider a speech signal is divided into N number of short-time frames and hamming windowed. Consider the i^{th} frame, from the i^{th} short-time frame the following parameters are extracted.

f_{i1} - First maximum energy peak
f_{i2} - Second maximum energy peak
f_{i3} - Third maximum energy peak
f_{i4} - First maximum energy peak multiplied with its location (*n1*)
f_{i5} - Second maximum energy peak multiplied with its location (*n2*)
f_{i6} - Third maximum energy peak multiplied with its location (*n3*)

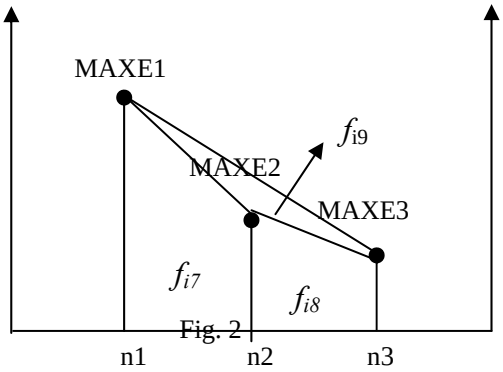

Fig. 2

Fig. 2 Illustration the energy peaks and areas between the two energy peaks of a speech signal in one short-time windows.

f_{i7} - Area enclosed by the first and second maximum energy peaks after drawing a straight line between them
f_{i8} - Area enclosed by the second and third maximum energy peaks after drawing a straight line between them
f_{i9}. - Area enclosed by the first and third maximum energy peaks after drawing a straight line between them
f_{i10} - Total energy of i^{th} short frame
f_{i11}. - Absolute difference between first and second maximum energy peaks
f_{i12} - Absolute difference between second and third maximum energy peaks
f_{i13} - Absolute difference between first and third maximum energy peaks

The following are the parameters extracted between adjacent frames.

f_{i14} - Absolute difference of first and second maximum energy peaks between adjacent frames
f_{i15} - Absolute difference of first and third maximum energy peaks between adjacent frames
f_{i16} - Absolute difference of second and third maximum energy peaks between adjacent frames

After extracting the above parameters from the i^{th} frame, the parameter set can be represented as

$$F=[f_{i1}\ f_{i2}\ f_{i3}\ f_{i4}\ f_{i5}\ f_{i6}\ f_{i7}\ f_{i8}\ f_{i9}\ f_{i10}\ f_{i11}\ f_{i12}\ f_{i13}\ f_{i14}\ f_{i15}\ f_{i16}] \quad i=1,2,...,N$$

Finally, from the parameter matrix the variation of each parameters are calculated using the following equation 1 and forms the feature vector.

$$V_j = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|F_{i,j} - F_{i+1,j}\right|}{\frac{1}{N}\sum_{i=1}^{N}F_{i,j}} \quad (1)$$

i=1,2,…,N and j=1,2,…,16

where V_j represents the variation of each parameters.

4. CLASSIFIER

In the area of automatic detection of voice pathology various classifiers have been proposed such as multi-layer perceptron [17,18], learning-vector quantization [19], Hidden Markov models [20], linear discriminant analysis [21], Gaussian mixture models [6], and k-nearest neighbourhood classifier[22,23]. In this paper, a probabilistic neural network is employed for the classification of pathological voices. Neural networks are frequently employed to classify patterns based on learning from examples. Artificial Neural Network (ANN) provides alternative form of computing that attempts to mimic the functionality of the brain [24]. Neural networks have been the subject of intensive research efforts in recent years because of their interesting learning and generalization properties and their applicability of classification, approximation and control problems. The back propagation method is a learning procedure for multilayered feedforward neural networks. But it has drawbacks of longer training time and local minima problem.

4.1. PROBABILISTIC NEURAL NETWORK (PNN)

Probabilistic neural networks can be used for classification problems. Donald F. Specht has proposed the probabilistic neural net based on Bayesian classification and classical estimators for probability density function [25]. It uses exponential activation function instead of sigmoidal activation function and also the training time is lesser compared to multi-layer feed forward network trained by back propagation algorithm. Consider the two Class problem, namely Class *A* and Class *B*. The probabilistic neural network uses the following estimator for the probability density function as given by equation 2

$$f_A(x) = \frac{1}{(2\pi)^{n/2}\,\sigma^n}\,\frac{1}{m_A}\sum_{i=1}^{m_A}\exp\left[-\frac{(x-x_{Ai})^T(x-x_{Ai})}{2\sigma^2}\right] \tag{2}$$

Where x_{Ai} is the *i*th training pattern from Class *A*, *n* is the dimension of the input vectors, m_A is the number of training patterns in Class *A*, and σ is a smoothing parameter corresponding to the standard deviation of the Gaussian distribution. The probabilistic neural net consists of four types of units, namely, input units, pattern units, summation units, and an output unit. The pattern unit computes distances from the input vector to the training input vectors, when an input is presented, and produces a vector whose elements indicate how close the input is to a training input. The summation unit sums these contributions for each class of inputs to produce as its net output a vector of probabilities. Finally, a compete transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 for that class and a 0 for the other classes.

The algorithm of the PNN is as follows [26]:

Step 0 : Initialize the weights

Step 1 : For each training input to be classified, do Steps 2 – 4.

Step 2 : Pattern units:
Compute net input:
$z_{inj} = x \,.\, w_j = x^T w_j$
Compute output using the equation 3.

$$z = exp\left[\frac{z_{inj} - 1}{\sigma^2}\right] \tag{3}$$

Step 3 : Summation unit:
Sum the inputs from the pattern units to which they are connected. The summation unit for class B multiplies its total input by the equation 4

$$v_B = -\frac{h_B\, c_B\, m_A}{h_A c_A\, m_B} \tag{4}$$

Step 4 : Output(decision) unit:
The output unit sums the signals from f_A and f_B.
The input vector is classified as Class A if the total input to the decision unit is positive.

The net can be used for classification as soon as an example of a pattern from each of the two classes has been presented to it. However, the ability of the net to generalize improves as it is trained on more examples. Varying σ gives control over the degree of nonlinearity of the decision boundaries for the net. A decision boundary approaches a hyperplane for large values of σ and approximates the highly nonlinear decision surface of the nearest neighbor classifier for values of σ that are close to zero. In this paper, PNN architecture is constructed using *newpnn( )* in MATLAB 7.0 function.

4.2. PERFORMANCE EVALUATION

The k-fold cross-validation scheme [27] is used for estimating the classifier performance. In this work, a 10-fold CVC scheme was used to increase the reliability of the results. Using this scheme, the proposed feature vectors are divided randomly into 10 sets and training is repeated for 10 times. For each run of cross validation the number of normal and pathological cases is equal. In order to test the classifier performance, several measures namely, sensitivity, specificity, positive predictivity, negative predictivity, and the overall accuracy are considered. These measures are calculated from the measures true positive (TP), true negative(TN), false positive(FP), and false negative(FN) as presented in Table 1.

Table1 Confusion Matrix

		Actual Classification	
Predicted Classification		Pathological	Normal
	Pathological	TP	FN
	Normal	FP	TN

TP= True Positive, the classifier classified as pathology when pathological samples are present
TN= True Negative, the classifier classified as normal when normal samples are present.
FN= False Negative, the classifier classified as normal when pathological samples are present.
FP= False Positive, the classifier classified as pathological when normal samples are present.
Sensitivity = TP/(TP+FN)
Specificity = TN/(TN+FP)
Positive predictivity = TP/(TP+FP)
Negative predictivity = TN/(TN+FN)
Overall accuracy = (TP+TN)/(TP+TN+FP+FN)

5.0 RESULTS AND DISCUSSION

Many research works have already been done in the area of automatic detection of voice pathology. In most of the studies, a two-class classification is carried out to categorize voice signal as normal or pathological class. The correct classification rate obtained in different works, when solving the two-class classification problem varies between 85% and 98.7% [17-23]. The main of this work is to develop simple and efficient feature extraction method without computing fundamental frequency, since the correct estimation of fundamental frequency of pathological voices is not easy. The time-domain features are extracted using the method as discussed in section 2. The data samples of 970 include 657 pathological and 313 normal speeches from the MEEI database are taken for our analysis. The mean and variance of each features of normal and pathological are tabulated in Table 2.

Table 2 Mean and Variance of the proposed features of normal and pathological voices

Features	Mean		Variance	
	normal	pathological	normal	Pathological
1	4.16	8.61	5.69	46.20
2	7.18	16.38	10.57	100.47
3	6.90	15.67	9.96	94.65
4	6.87	14.96	93658.00	85.68
5	10.22	20.46	15.96	104.12
6	13.03	20.51	19.55	93.97
7	14.95	20.87	22.64	82.42
8	64.17	139.29	1929.90	1544.36
9	73.65	111.70	1011.06	1404.95
10	79.13	119.93	1347.17	1219.87
11	84.15	1.02	453.67	39.43
12	87.65	1.32	458.26	55.31
13	54.76	1.07	279.51	17.42
14	81.22	69.07	439.85	226.74
15	80.13	72.90	421.59	207.23
16	79.16	72.08	479.90	213.98

From the table 2, it can be observed that the proposed features can be used to discriminate the voice as normal or pathological clinically. Using the proposed features, PNN network is trained for various spread factors of 0.01, 0.02, 0.03, 0.04, and 0.05. The results of the PNN classifier are tabulated in table 3 in terms of sensitivity, specificity, positive predictivity, negative predictivity, and the overall accuracy.

Table 3 Classification performance of the PNN network

Spread Factor	Positive Predictivity(%)	Negative Predictivity(%)	Sensitivity (%)	Specificity (%)	Overall Accuracy(%)
0.01	99.56	88.21	89.41	99.50	95.89
0.02	99.39	98.40	98.42	99.38	99.07
0.03	99.39	98.34	98.36	99.38	99.05
0.04	99.39	98.47	98.48	99.38	99.09
0.05	99.39	98.66	98.67	99.39	99.15
Average	**99.42**	**96.42**	**96.67**	**99.41**	**98.45**

From the table 3, it is observed that, the overall accuracy of the PNN classifier is 98.45%, the over all number of correctly classified pathological samples is 99.42%, the over all number of correctly classified normal samples is 96.42%, the overall specificity is 99.41% and the overall sensitivity is 96.67%. The performance of the proposed method cannot be directly compared with the previous works, since their computation, database handling and

selection of classifiers, presentation of results are not comparable. In order to prove the validity and reliability of the proposed schemes, the 10- fold cross validation scheme is used. The results show the reliability and effectiveness of the proposed features and it can be applied to diagnose the vocal fold pathology clinically.

6.0 CONCLUSION

A simple feature extraction method to detect the vocal fold pathology is proposed based on the study of time-domain energy level with the aid of the speech signal recorded from the patients. In order to test the effectiveness and reliability of the proposed time-domain features, a Probabilistic Neural Network is employed. The experimental results show that the proposed features give very promising classification accuracy of 98.45% with less computational complexity in feature extraction. The proposed features can be used as additional acoustic indicators and can also be used as a valuable tool for researchers and speech pathologies to detect the vocal fold pathology. In the future, it is proposed to extend this method to detect the specific type of disorders and also to develop an online diagnosing system.

REFERENCES

[1] Carlos Hernandez-Espinosa, Mercedes Fernandez-Redondo, Santiago Aguilera-Navarro, Pedro-gomez-Vilda,Godino-Llorente J.I, Gomez-Vilda P, "Diagnosis of Vocal and Voice Disorders by the Speech Signal", *Proceedings of IEEE-INNS-ENNS International Joint Conference on Neural Networks*,2000, vol.4, pp.253-258.

[2] Boyanov B. and Stefan Hadjitodorov, "Acoustic Analysis of Pathological Voices. A voice analysis system for the screening of laryngeal diseases", *Proceedings of IEEE International Conference on Engineering in Medicine and Biology Society*, 1997,vol.16, pp.74-82.

[3] Kausya H, Ogawa,S., Mashima,K., "Normalized noise energy as an acoustic measure to evaluate pathologic voice", *Journal of the Acoustical Society of America*, 1986, vol.80,pp.1329-1334.

[4] Martinez Cesar E, Rufiner hugo L., "Acoustic Analysis of Speech for Detection of Laryngeal Pathologies", *Proc. of 22nd Annual IEEE International Conference on Engineering in Medicine and Biology Society*, 2000, vol.3, pp.2369-2372.

[5] Dimitar D. Deliyski, "Acoustic Model and Evaluation of Pathological Voice Production", *Third International Conference on Speech Communication and Technology*, EUROSPEECH'93, 1993, pp1969-72.

[6] Godino-Llorente J.I, Gomez-Vilda P, and Blanco-Velasco M., "Dimensionality Reduction of a Pathological Voice Quality Assessment System Based on Gaussian Mixture Models and Short-term Cepstral Parameters", *IEEE Transactions on Biomedical Engineering*, 2006, vol.53, pp.1943-1953.

[7] Ritchings, T., McGillion, M., and Moore, C., "Pathological voice quality assessment using artificial neural networks", *Medical Engineering and Physics*,2002, vol.24, no.8, pp.561-564.

[8] Childers,D.G, Keun Sung Bae, "Detection of Laryngeal Function Using Speech and Electroglottograph Data", *IEEE Transactions on Biomedical Engineering*, 1992, vol,39,pp.19-25.

[9] Massachusetts Eye and Ear Infirmary, "Voice Disorders Database", Version 1.03(CDROM), *Kay Elemetrics Corporation*, 1994, Lincoln Park, NJ, USA.

[10] Boyanov B, Ivanov Tm Cholet G, "Robust hybrid pitch detector", *Electronics letters*, 1980, vol.29, pp. 1924-1926.

[11] Feijoo S, Hernandez C, "Short-tem stability measures for the evaluation of voice quality", *Journal of Speech and Hearing Research,* 1990, vol. 33, pp.324-334.

[12] Kasuya H, Endo and Sliu S, "Novel acoustic measurements of jitter and shimmer characteristics from pathological voice", *Proceedings of EUROSPEECH'93*,1993, pp.1973-1976.

[13] Ludlow C, Bassich C, Connor N, Coulter D, Lee Y, "The validity of using phonatory jitter and shimmer to detect laryngeal pathology", *Laryngeal function in phonation and respiration*, Brown & Co., Boston, 1987, pp.492-508.

[14] Yunik M and Boyanov B, "Method for evaluation of the noise-to harmonic-component ratios in pathological and normal voices", *Acoustica*, 1990, vol. 70, pp. 89-91.

[15] Eiji Yumoto, Wilbur J, Gould, "Harmonics to noise ratio as an index of the degree of hoarseness", *The Journal of the Acoustical Society of America*, 1990, vol.87, no. 3, pp.1278-1289.

[16] DeKrom G, "A cepstrum based technique for determining a harmonics to noise ratio in speech signals", *Journal of Speech and Hearing Research*, 1993, vol. 36, pp. 254-266.

[17] Tim Ritchings, Mark A. McGillion, Christopher J. Moore, "Objective assessment of pathological voice quality using multi-layer perceptrons", *Proc. of the First Joint EMBS/BMES Conference*, 1999, vol. 2, pp.925.

[18] Tim Ritchings, Mark A. McGillion, Christopher J. Moore,GV Conroy, "Objective assessment of pathological voice quality", *IEEE International Conference on Systems, Man, and Cybernetics*, 1999, vol.6,pp. 340-345.

[19] J. I. Godino-Llorente and P. Gomez-Vilda, "Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors," *IEEE Transactions on Biomedical Engineering*, 2004,vol. 51, no. 2, pp. 380–384.
[20] Mirjam Wester, "Automatic classification of voice quality: comparing regression models and hidden Markov models", *Proceedings of VOICEDATA98, symposium on databases in voice quality research and education*, 1998, pp. 92-97.
[21] K. Umapathi, S. Krishnan, V. Parsa, and D. G. Jamieson, "Discrimination of pathological voices using a time-frequency approach," *IEEE Transactions on Biomedical Engineering*, 2005, vol. 52, no. 3, pp. 421–430.
[22] T. Ananthakrishna T, Kumara Shama, and U.C. Niranjan, "k-means nearest neighbor classifier for voice pathology", *IEEE INDIA Annual Conference,* 2004, pp.352-354.
[23] Kumara Shama, Ananthakrishna, and Niranjan U. Cholayya, "Study of harmonics-to-noise ration and critical band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology", *EURASIP Journal on Advances in Signal Processing*, Hindawi Publishing Corporation, 2007, vol. 2007, pp.1-9.
[24] S.N Sivanandam and Paulraj M P, "An Introduction to Artificial Neural Networks",2003, *Vikhas Publication*, India.
[25] Specht, D, "Probabilistic neural networks", *Neural Networks,*1990, vol.3. no.1, pp. 109-118.
[26] Laurene Fausett, "Fundamentals of Neural Network", 1994,Prentice Hall, New Jersey, USA.
[27] Kohavi, R. "A study of cross validation and bootstrap for accuracy estimation and model selection" In *Proceedings of the 14th International Conference on Artificial Intelligence*, 1995, pp. 1137-1143.

BIOGRAPHY

M. Hariharan has obtained BE in Electrical and Electronics Engineering from Bharathiar University, India and Master degree from Anna University, India. He is currently doing Ph.D in Mechatronic Engineering at Universiti Malaysia Perlis, Malaysia. His research interests include Speech Signal Processing, Wavelet Transform, Image Processing and Artificial Neural Network. He is a graduate student member of IEEE and Acoustical Society of America, USA.

Paulraj M P is currently working as an Associate Professor at Universiti Malaysia Perlis, Malaysia. His research interests are in the areas of artificial intelligence, fuzzy systems, speech processing, acoustic engineering and biosignal processing applications. He has published more than 100 papers in referred journals and conferences. He has authored a book titled as "Introduction to Artificial Neural Networks". He is a member of the Institution of Engineers, India and MISTE, India.

Sazali Yaacob is currently working as a Professor at Universiti Malaysia Perlis, Malaysia. He has published more than 150 papers in Journals and Conference Proceedings. His research interests are in Artificial Intelligence applications in the fields of Acoustics, Vision and Robotics. He received Charted Engineer status by the Engineering Council, United Kingdom in 2005 and also a member of the IET (UK).