FUZZY WAVELET PACKET BASED FEATURE EXTRACTION METHOD APPLIED TO PATHOLOGICAL VOICE SIGNALS CLASSIFICATION

B. S. Aghazadeh 1, H. Khadivi Heris 1, H. Ahmadi 2, M. Nikkhah-Bahrami 1
1Department of Mechanical Engineering, Tehran University, Tehran, Iran
2Department of Electrical Engineering, Tehran University, Tehran, Iran

Abstract: In this paper an efficient fuzzy wavelet packet (WP) based feature extraction method has been used for the classification of normal voices and pathological voices of patients suffering from unilateral vocal fold paralysis (UVFP). Mother wavelet function of tenth order Daubechies (d10) has been employed to decompose signals in 5 levels. Next, WP coefficients have been used to measure energy and Shannon entropy features at different spectral sub-bands. Consequently, to find discriminant features, signals have been clustered in 2 classes using fuzzy c-means method. The amount of fuzzy membership of pathological and normal signals in their corresponding clusters is considered as a measure to quantify the discrimination ability of features. Thus, considering this measure, an optimal feature vector of length 8 has been chosen to discriminate pathological voices from normal ones. Feature vector obtained by considering nodes’ discriminant ability with classification percentage of 100 has a better performance in comparison with the one obtained by genetic algorithm. Feature vector including equal portion of nodes for the features of energy and entropy with the approximate classification percentage of 96. The simulation results show that fuzzy WP based feature extraction is an effective tool in voice signal analysis.

Keywords: Voice disorders, feature extraction, wavelet packets, fuzzy sets

I. INTRODUCTION

Unilateral vocal fold paralysis (UVFP) occurs from a dysfunction of the recurrent or vagus nerve innervating the larynx and causes a characteristic breathy voice. UVFP most commonly occurs following a surgical iatrogenic injury to the vagus or recurrent laryngeal nerve resulting in glottal incompetence, either partial or complete, because of the poor or reduced vocal fold closure.

Physiological alterations of vocal cords cause unhealthy patterns of cords’ vibration and the decrease in patients’ speech signal quality known as voice pathologies. Therefore, the detection of incipient damages to the cords is useful in improving the physician’s treatment and care of such pathologies. Physicians often use invasive techniques like Endoscopy to diagnose symptoms of voice disorders. It is, however, possible to identify disorders using certain features of speech signal in a non-invasive way [1]. Schuck et al. [2] have used Shannon entropy and energy features of wavelet packet decomposition and the best basis algorithm for normal/pathological speech signal classification. Fonseca et al. [3] have employed mean square values of reconstructed signals in discrete wavelet transform sub-bands and least square support vector machine (LS-SVM) classifier for identification of signals from patients with vocal fold nodules and normal signals. Guido et al. [4] have tried different wavelets on the search for voice disorders. Mother wavelet of Daubechies with support length of 20 (db10) was found as the best wavelet for speech signal analysis among commonly used wavelets. Behroozmand et al. [5] have used genetic algorithm for optimal selection of wavelet packet based energy and Shannon entropy features for identification of patients’ speech signal with unilateral vocal fold paralysis (UVFP). The results showed that the decomposition level of five is the best level to analyze pathological speech signals. Local discriminant bases (LDB) and wavelet packet decomposition have been used to demonstrate the significance of identifying discriminant WP subspace in a work by Umapathy et al. [6].

Fuzzy wavelet packet based feature extraction method has been proposed by Li et al. and has been applied to biological signal classification [7]. In contrast to the standard methods of feature extraction used in WPs, this method of discriminatory feature extraction from wavelet packet coefficients is based on the fuzzy set criterion. Yang et al. [8] have applied fuzzy wavelet packet method to feature extraction from electroencephalogram (EEG) signals. The results show that this method is promising for the extraction of EEG signals in brain-computer interfaces (BCIs).

This work aims to identify patients with UVFP by extracting an effective feature vector containing less number of features and higher discrimination accuracy and lower order of computational complexity (e.g. in comparison with the one obtained by genetic algorithm based optimal feature [5]). It is based on wavelet packet transform (WPT), fuzzy sets, and artificial neural network (ANN) classifier.
II. METHODS

A. Wavelet Packet Transform

Recently, wavelet packets (WPs) have been widely used by many researchers to analyze voice and speech signals. There are many outstanding properties of wavelet packets, which encourage researchers to employ them in many widespread fields. It has been shown that sparsity of coefficients' matrix, computational efficiency, and time-frequency analysis can be useful in dealing with many engineering problems. The most important, multiresolution property of WPs is helpful in voice signal synthesis.

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 the both low and high frequency sub-bands to generate the next level of the hierarchy. WPs can be described by the following collection of basis functions:

\[
W_{2n}\left(2^{p+1}x-l\right) = \sqrt{2^{2p}} \sum_{m} h_{2p-2l} x^{-m} \quad (1)
\]

\[
W_{2n+1}\left(2^{p+1}x-l\right) = \sqrt{2^{2p}} \sum_{m} g_{2p-2l} x^{-m} \quad (2)
\]

where \(p\) is scale index, \(l\) the translation index, \(h\) the low-pass filter and \(g\) the high-pass filter.

The WP coefficients at different scales and positions of a discrete signal can be computed as follows:

\[
C_{p, l}^{n} = \sqrt{2^{2p}} \sum_{m} f_{2p} x^{-m} \quad (4)
\]

\[
C_{2p, l}^{n} = \sum_{m} h_{2p-2l} C_{p, m}^{n} \quad (5)
\]

\[
C_{2p+1, l}^{n} = \sum_{m} g_{2p-2l} C_{p,m}^{n} \quad (6)
\]

For a particular sequence of wavelet packet coefficients, energy in its corresponding sub-band can be computed as:

\[
\text{Energy}_{n} = \frac{1}{N^2} \sum_{p} \left| C_{p, l}^{n} \right|^2 \quad (7)
\]

The Shannon entropy as another extracted feature for classification of signals can be computed through the following formula:

\[
\text{Entropy}_{n} = \frac{1}{N^2} \sum_{p} \left| C_{p, l}^{n} \right|^2 \log \left| C_{p, l}^{n} \right|^2 \quad (8)
\]

Due to the noise-like effect of irregularities in the vibration pattern of damaged vocal folds, the distribution manner of such variations within the whole frequency range of pathological speech signals is not clearly known. Therefore, it seems reasonable to use WP rather than discrete wavelet transform (DWT) to have more detail sub-bands.

B. Fuzzy Set-Based Feature Selection Criterion

With fuzzy sets we allow any pattern \(x_k\) to belong to several classes to varying degrees. Assuming \(u_{ik}\) a membership grade of pattern \(x_k\) to class \(i\) we have:

\[
u_{ik} = \frac{1}{\sum_{i} d(x_k, \bar{A}_i)^b} \quad (9)
\]

where \(c\) is the number of clusters, \(d(x_k, \bar{A}_i)\) is the mean of class \(i\), \(\bar{A}_i\) is the set of indexes of the training patterns belonging to class \(i\), \(N_i\) is the number of class \(i\) training patterns, \(\|x_k - \bar{A}_i\|\) is the Euclidean distance and \(b > 1\) is the fuzzification factor that modifies the shape of membership grades. For the labeled training patterns in feature space, \(X\), we define a membership function based on the criterion \(F(X) \in (0, N]\) to evaluate the classification ability of \(X\) as follows:

\[
F(X) = \sum_{i=1}^{c} \sum_{x_k \in A_i} u_{ik} \quad (10)
\]

The larger the values of \(F(X)\), the higher the classification (discrimination) abilities of the feature space \(X\).

In fuzzy set based optimal WP decomposition for each labeled original signal a full WP decomposition to maximum level of five has been performed. The mother wavelet function is chosen to be the tenth order Daubechies (db10). Consequently, features (i.e. energy and entropy) of all signals in each node have been clustered using Fuzzy Clustering Method (FCM). Most discriminant nodes have been identified according to the parameter \(F(X)\) and the signals’ energy and entropy in those nodes have been used to construct the feature vector applied to artificial neural network (ANN) classifier.

C. Database

Used in this study are sustained vowel phonation samples from subjects from the Kay Elemetrics Disordered Voice Database [9]. Subjects were asked to sustain the vowel /a/ and voice recordings were made in a sound proof booth on a DAT recorder at a sampling frequency of 44.1 kHz.
III. RESULTS AND DISCUSSION

Having signals decomposed by mother wavelet of tenth order Daubechies to 5 levels of decomposition and having on hand energy and Shannon entropy at each decomposition sub-band, fuzzy logic based feature extraction method has been applied to construct an optimal feature vector of length 8 according to the nodes’ discrimination ability, which can separate normal and pathological (UVFP) voice signals.

Table 1 shows the most discriminant nodes in terms of energy or entropy feature, with their discrimination abilities, \( F(X) / \text{number of data} \times 100 \), which are obtained from fuzzy clustering method.

A feature vector of length 8 has been extracted from the data: 1) with equal portion of discriminant energy and entropy nodes, 2) according to the best discriminant nodes in terms of energy or entropy. Consequently, approximately 65 percent of data has been used as the training data set and the remaining 35 percent are set aside as the test and validation data to train a feedforward backpropagation multilayer classifier neural network with 3 hidden layers.

Fig. 1 shows the wavelet packet tree and the participating nodes in feature vector, which are selected according to their discriminant ability. As can be seen, selected sub-bands are distributed over the whole available frequency ranges, which shows that pathological factors do not influence specific frequencies which accentuates the role of WP decomposition with equal decomposition of both high and low frequencies.

As a case in point, the coefficients’ energies of decomposed voice signals in the most discriminant node (31) have been illustrated in fig. 2. The efficiency and discrimination ability of selected node is obvious.

<table>
<thead>
<tr>
<th>Node</th>
<th>Energy</th>
<th>Entropy</th>
<th>Discrimination ability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>*</td>
<td>76.22</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>*</td>
<td>73.32</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>*</td>
<td>67.17</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>*</td>
<td>66.95</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>*</td>
<td>66.61</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>*</td>
<td>66.82</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>*</td>
<td>65.85</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>*</td>
<td>62.71</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>*</td>
<td>61.50</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>*</td>
<td>61.27</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. The most discriminant nodes in terms of signal energy or entropy
The simulation results show that fuzzy wavelet packet based feature extraction method and neural network classifiers are effective tools in voice signal analysis. Moreover, feature vector obtained considering nodes’ discriminant ability with classification percentage of 100 has a better performance in comparison with the feature vector including equal portion of nodes for the features of energy and entropy with the approximate classification percentage of 96.

IV. CONCLUSION

In this study, classification of voice signals into two groups of normal and patients with unilateral vocal fold paralysis (UVFP) has been presented. Fuzzy wavelet packet based feature extraction method has been utilized to find the optimal feature vector of length 8 from energy and Shannon entropy features in WP decomposition sub-bands. In the following, the obtained feature vector has been passed on to a neural network (NN) classifier. The simulation results show that the fuzzy wavelet packet based selected optimal feature vector of length 8 applied to a NN classifier can achieve a classification accuracy of 100 percent, which despite its relatively short length, outperforms feature vectors obtained by other methods.

REFERENCES