Pathological Voice Detection by Cepstral Analysis Using Multiple Classifiers

Raissa Tavares\(^1\), Nathália Brunet\(^2\), Suzete Correia\(^3\), Silvana Cunha Costa\(^4\)
Electrical Engineering Coordination
Federal Institute of Technological Education - IFPB
João Pessoa, Brazil
\(^1\)raissa@ieee.org, \(^2\)nathaliabrunet@gmail.com, \(^3\)suzete@ifpb.edu.br, \(^4\)silvana@ifpb.edu.br

Abstract—This paper evaluates cepstral classifiers applied to pathological voice detection problem. The goal is analyze the individual and combined performance of classifiers based on cepstral, weighted cepstral, delta cepstral, and weighted delta cepstral parameters. They are evaluated considering two different combination strategies yielding a multiple classifier that is more efficient than either individual technique. The efficiency rates obtained vary from 87% using stand alone weighted delta cepstral to 98% considering the classifiers combination.

Keywords-Acoustic signal analysis, pathological voices, cepstral analysis, multiple classifiers.

I. INTRODUCTION

The diagnosis of laryngeal pathologies is usually made by laryngoscopic exams, which are considered invasive, causing discomfort to patients. Digital signal processing techniques performing an acoustic analysis for vocal quality assessment is a simple and noninvasive measurement procedure. These techniques provide an objective diagnosis of pathological voices, and may be used as complementary tool in laryngoscopic exams [1].

Some researchers have dedicated efforts for obtaining efficient methods for discriminating normal and pathological voices using acoustic analysis [1]-[6]. However, the research for a more detailed and representative acoustic analysis of pathological voice signals is still a promising area.

In this work, a parametric analysis based on cepstral analysis is employed to discriminate pathological voices of speakers affected by vocal fold edema. Cepstral (CEP), weighted cepstral (WCEP) delta cepstral (DCEP), and weighted delta cepstral (WDCEP) parameters are used as features to detect the irregularities of the pathological voices in comparison with the normal voice. A vector quantization technique (VQ) was used associated with a distortion measurement to classify the speech signal by each parameter. The VQ was trained with voices affected by the considered pathology, and may provide a set of speech parameters that represent the pathological speech signal versus normal speech [10].

II. AN OVERVIEW ON THE CEPSTRAL ANALYSIS

Considering that speech signal is the result of convolving excitation with vocal tract sample response, by cepstral analysis, it is possible to separate these two components. All the cepstrum-related features described are obtained after Linear Predictive Coding (LPC) analysis [8],[14].

The linear prediction method estimates each speech sample based on a linear combination of the \( p \) previous samples; a larger \( p \) enables a more accurate model. It provides a set of speech parameters that represent the vocal tract [8]. It is expected that any change in the anatomical structure of the vocal tract, because of pathology, affects the LPC coefficients and also the cepstral and its derivatives. A linear predictor with prediction coefficients, \( a(k) \), is defined as a system whose output is

\[
\hat{s}(n) = \sum_{k} a(k) s(n-k),
\]

where \( p \) is the predictor order. The autocorrelation method and the covariance method are standard for computing the predictor coefficients. The prediction coefficients are computed using the Levinson-Durbin algorithm [8].

Cepstral derivatives can improve the representation of the spectral speech properties. Pathological speech presents significant spectral differences of normal voices. The cepstral coefficients can be calculated recursively from the linear predictor coefficients, \( a(k) \), by means of [9]:

\[
\begin{align*}
\alpha(1) &= -a(1) \\
\alpha(n) &= \sum_{j=1}^{n-1} \frac{1}{n} \alpha(j) c(n-j) \\
\end{align*}
\]

Cepstral coefficients obtained by (2) provide a good measure of the difference in the spectral envelope of the speech frames. These coefficients are used in order to observe the information of voice transitions in pathological speech signal versus normal speech [10].
The first derivative of the cepstral coefficients (Delta Cepstral Coefficients) is given by [10]:

\[
\frac{\Delta c(n,t)}{\Delta t} = \Delta c(n) = \phi \sum_{k} k c(n, t+k),
\]

(3)

where \( c(n,t) \) is the n-th LP coefficient at time \( t \), \( \phi \) is a normalization constant and \( 2K+1 \) is the number of frames over which the computation is performed.

The delta cepstral coefficients are obtained as a simplified version of (3), as it was proposed by [10]:

\[
\Delta c_{i}(n) = (\sum_{k} k c_{i+q}(n))G, \quad 1 \leq n \leq p,
\]

(4)

where \( G \) is a gain term (for example, 0.375), \( p \) is the number of delta cepstral coefficients, \( K=2 \), \( n \) the coefficient index and \( i \) the frame of analysis [12].

In order to account for the sensitivity of the low-order cepstral coefficients to overall spectral slope and the sensitivity of the high-order cepstral coefficients to noise, cepstral weighting (lifting) is employed.

The weighted cepstral coefficients (WCEP), \( cw(n) \), are obtained by [10]-[12]:

\[
cw(n) = c(n) \cdot w(n).
\]

(5)

The type of window used in this work was the band pass lifting (BPL), given by [10]:

\[
w(n) = \begin{cases} 
1 + \frac{L}{2} \sin \left( \frac{n\pi}{L} \right), & n=1,2,\ldots,L, \\
0, & \text{otherwise}. 
\end{cases}
\]

(6)

where \( L \) is the size of the window. The BPL weights a cepstral sequence by (6) so that the lower- and higher-order components are de-emphasized.

Weighted Delta Cepstral coefficients (WDCEP) are obtained replacing (4) in (5), resulting on

\[
\Delta cw_{i}(n) = \Delta c_{i}(n) \cdot w(n).
\]

(7)

The characteristics of weighted cepstral and delta cepstral are associated by using (7).

III. DATABASE AND METHODS

The database was recorded by the Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab [14]. The following cases were selected: 44 patients presenting vocal fold edema - 33 women (17 to 85 years old) and 11 men (23 to 63 years old), most of them (32) with bilateral edema and 53 patients with normal voices which is composed of 21 male (26 to 59 years old), and 32 female (22 to 52 years old).

In the pre-processing stage speech signals are multiplied by a 20 ms Hamming window with an overlap of 50%. A filter of pre-emphasis (0.95) is also used. Then each parameter is calculated after LP coefficients (\( p=12 \)).

To dimensionality reduction of data, a Vector Quantization (VQ) technique [15] is used that is associated with a distortion measurement. The quantization is carried out individually for each feature using just voices under vocal fold edema in the training phase. Thus, different VQ-trained distance classifiers [12] are obtained by the discrimination process. The VQ-classifiers are applied to static feature vectors, which are computed for every analysis frame of the speech samples over a dynamic input sustained vowel /a/. It is used 50% of vocal fold edema cases in the training phase. To the test phase, the other 50% of voices signals under vocal fold edema, and all the normal (53) voices are used. After the feature extraction, a codebook is generated using the Euclidean distortion measurement and the nearest neighbour rule is used to find the codevector. LBG algorithm to quantization and the least mean square distance for classification process are used [16].

IV. RESULTS AND DISCUSSION

The measurements used to evaluate the performance of the methods are the following: Correct acceptance (CA) rate; Correct rejection (CR) rate, False acceptance (FA) rate; False rejection (FR) rate; and the Efficiency representing the correct classification of a given class when that is present, given by \( E(\%) = \text{CR+CA)/(CR+CA+FA+FR)} \times 100 \) [1].

Table I shows the results obtained for each parameter individually. It can be seen that Delta Cepstral (DCEP) method gives the best Efficiency and False Acceptance rates. However, this method presents a higher False Rejection rate compared to cepstral (CEP) method.

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>CA (%)</th>
<th>FA (%)</th>
<th>FR (%)</th>
<th>E (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEP</td>
<td>89</td>
<td>91</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>WCEP</td>
<td>94</td>
<td>86</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>DCEP</td>
<td>98</td>
<td>86</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>WDCEP</td>
<td>91</td>
<td>82</td>
<td>9</td>
<td>18</td>
</tr>
</tbody>
</table>

To evaluate the combined features, makes the assumption that a speech signal must be assigned to one of the \( K \) possible classes and assume that \( L \) classifiers are available. The distortion measurement used by the \( i \)th QV-classifier is denoted as \( d_i \). Two combination rules have been employed:

- Combination by average:

\[
D = \frac{1}{K} \sum_{i=1}^{K} d_i,
\]

(8)

- Combination by Product:

\[
D = \prod_{i=1}^{L} d_i,
\]

(9)

where \( D \) denotes the distortion obtained after combination.

In order to guarantee an standardization of the that the classifier outputs, the distortion values of each VQ-classifier were normalized (values between 0 and 1). A threshold of \( D \) is chosen such as the best separation between the classes is obtained.

508
The results for average and product combination are presented in Tables II and III, respectively. The results show that, for almost all combinations, the False Rejection rates decrease significantly. In the average rule (Table II), the probability in rejecting the presence of vocal fold edema (14%) in the individual case (DCEP) that gives the best efficiency (Table I) is about 2% when combining the four parameters.

**TABLE II. PERFORMANCE EVALUATION – THE AVERAGE RULE**

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>CR (%)</th>
<th>CA (%)</th>
<th>FA (%)</th>
<th>FR (%)</th>
<th>E (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEP and DCEP</td>
<td>92</td>
<td>95</td>
<td>8</td>
<td>5</td>
<td>94</td>
</tr>
<tr>
<td>CEP and WDCEP</td>
<td>94</td>
<td>95</td>
<td>6</td>
<td>5</td>
<td>95</td>
</tr>
<tr>
<td>CEP and WCEP</td>
<td>96</td>
<td>94</td>
<td>4</td>
<td>5</td>
<td>96</td>
</tr>
<tr>
<td>DCEP and WDCEP</td>
<td>98</td>
<td>82</td>
<td>2</td>
<td>18</td>
<td>90</td>
</tr>
<tr>
<td>DCEP and WCEP</td>
<td>94</td>
<td>93</td>
<td>6</td>
<td>7</td>
<td>94</td>
</tr>
<tr>
<td>WDCEP and WCEP</td>
<td>92</td>
<td>95</td>
<td>8</td>
<td>5</td>
<td>94</td>
</tr>
<tr>
<td>CEP, DCEP and WDCEP</td>
<td>94</td>
<td>91</td>
<td>6</td>
<td>9</td>
<td>93</td>
</tr>
<tr>
<td>CEP, DCEP and WCEP</td>
<td>96</td>
<td>95</td>
<td>4</td>
<td>5</td>
<td>96</td>
</tr>
<tr>
<td>CEP, WDCEP and WCEP</td>
<td>96</td>
<td>98</td>
<td>4</td>
<td>2</td>
<td>97</td>
</tr>
<tr>
<td>DCEP, WDCEP and WCEP</td>
<td>94</td>
<td>93</td>
<td>6</td>
<td>7</td>
<td>94</td>
</tr>
<tr>
<td>CEP, DCEP, WDCEP and WCEP</td>
<td>94</td>
<td>98</td>
<td>6</td>
<td>2</td>
<td>96</td>
</tr>
</tbody>
</table>

The best result is obtained using combination by product of CEP and WDCEP classifiers (Table III). It can be observed an improvement of at least 6% in efficiency rate, comparing with the DCEP individual classifier. For this case, the probability in detecting the presence of the edema pathology when, in real, it is not present (FA), is null.

**TABLE III PERFORMANCE EVALUATION – THE PRODUCT RULE**

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>CR (%)</th>
<th>CA (%)</th>
<th>FA (%)</th>
<th>FR (%)</th>
<th>E (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEP, DCEP</td>
<td>100</td>
<td>93</td>
<td>0</td>
<td>7</td>
<td>97</td>
</tr>
<tr>
<td>CEP, WDCEP</td>
<td>100</td>
<td>95</td>
<td>0</td>
<td>5</td>
<td>98</td>
</tr>
<tr>
<td>CEP, WCEP</td>
<td>92</td>
<td>98</td>
<td>8</td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td>DCEP, WDCEP</td>
<td>96</td>
<td>86</td>
<td>4</td>
<td>14</td>
<td>91</td>
</tr>
<tr>
<td>DCEP, WCEP</td>
<td>94</td>
<td>95</td>
<td>6</td>
<td>5</td>
<td>95</td>
</tr>
<tr>
<td>WDCEP, WCEP</td>
<td>92</td>
<td>98</td>
<td>8</td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td>CEP, DCEP and WDCEP</td>
<td>94</td>
<td>95</td>
<td>6</td>
<td>5</td>
<td>95</td>
</tr>
<tr>
<td>CEP, DCEP and WCEP</td>
<td>94</td>
<td>95</td>
<td>6</td>
<td>5</td>
<td>95</td>
</tr>
<tr>
<td>CEP, WDCEP and WCEP</td>
<td>94</td>
<td>98</td>
<td>6</td>
<td>2</td>
<td>96</td>
</tr>
<tr>
<td>DCEP, WDCEP and WCEP</td>
<td>94</td>
<td>93</td>
<td>6</td>
<td>7</td>
<td>94</td>
</tr>
<tr>
<td>CEP, DCEP, WDCEP and WCEP</td>
<td>74</td>
<td>98</td>
<td>26</td>
<td>2</td>
<td>86</td>
</tr>
</tbody>
</table>

V. CONCLUSIONS

In this paper the individual and combined performance of classifiers based on cepstral, weighted cepstral, delta cepstral, and weighted delta cepstral parameters were evaluated for the pathological voice detection problem. The results show that combination of these classifiers can yield a significant performance improvement related to individual ones. The best efficiency rate in the individual case was 92% and after the combinations, about 98%. This mean that the parameters employed are complementary and can be used to detect vocal disorders caused by the presence of vocal fold pathologies. Future work will focus in the use of others combination rules, such as an weighted average, for example, and in the use of other classifiers, such as Neural Network and/or Hidden Markov Models. Furthermore, the system performance can be tested with other laryngeal pathologies.

REFERENCES


