Pathological Voice Classification Based on Recurrence Quantification Measures

Washington C. de A. Costa, F. M. de Assis, B. G. Aguiar Neto, Silvana C. Costa and Vinicius J. D. Vieira

Abstract—This paper presents an analysis of speech signals based on quantification measures of recurrence plots. A comparison between healthy voices and voices affected by laryngeal pathologies (Reinke’s edema, nodule and vocal fold paralysis) is made. In order to classify these signals as pathological or healthy, seven recurrence quantification measures are used: Determinism, maximum length of the diagonal structures, Shannon entropy of line distribution, slope of line of best fit, laminarity, length of longest vertical line segment and mean vertical line length or trapping time. Discriminant analysis methods (linear and quadratic) are applied to each feature individually, and to the vectors formed by feature combination with cross-validation classification rates up to 95.71±4.94\% (95\% confidence interval). Results show that the employed measures present a significant discriminative potential to distinguish healthy voices from pathological ones.

Keywords—Speech signal analysis, laryngeal pathologies, recurrence quantification measures.

I. INTRODUCTION

Voice is considered the main tool of human communication. From a health science standpoint, the human voice has been shown to carry much information about the general health and well-being of an individual. Our voice reveals who we are and how we feel, providing considerable insight for the structure and function of certain parts of the body [1].

Several factors can affect the voice quality including unhealthy social habits such as smoking and alcoholism, vocal abuse, and larynx diseases. Early diagnosis of laryngeal pathologies can positively influence the treatment and cure of diseases. Some medical examinations (e.g. fiberoptic laryngoscopy) applied to observe the larynx and help the diagnosis are considered invasive, causing discomfort to patients. In the last decades, methods based on digital signal processing techniques have been developed to provide an auxiliary and noninvasive tool to detect vocal disorders caused by laryngeal pathologies. Some of these methods are based on linear model of speech production [2], [3] or they can be based on nonlinear dynamic analysis of speech [4], [5].

Acoustic analysis of speech signals to detect their disorders caused by vocal fold pathologies are based on perturbation measures obtained from fundamental frequency or their perceptual correlate pitch such as jitter, shimmer, amplitude perturbation quotient (APQ), pitch perturbation quotient (PPQ) and others [6], [7]. However, some pathological signals are so disordered that the pitch obtention is difficult or sometimes impossible in the presence of some pathologies. Other applied methods are based on inverse filtering [8]. Nevertheless, the fact that linear prediction coding (LPC) based on inverse filtering has as assumption a linear model, such methods do not behave well when pathology is present due to nonlinearities introduced by the pathology itself [2].

Since the original suggestions of Titze et al [9] to improve our understanding of voice disorders with nonlinear dynamic concepts and analysis methods, researchers have been studying new techniques to differentiate healthy and pathologic voices and diagnose laryngeal pathologies [10].

The correct classification rate obtained in previous researches to distinguish between pathological and healthy voices varies significantly: 85.8\% [11], 89.1\% [12], 91.8\% [13], 99.44\% [14], 90.1\% (Healthy x Nodule), 85.3\% (Healthy x Edema) and 88.2\% (Healthy x Pathological) [15]. However, the comparison among the researches carried out is very complex due to the wide range of measures, data sets and classifiers employed.

Although there are several methods for acoustic analysis of voices affected by laryngeal diseases, different measures respond to different disorders in different ways [13]. The research to find out which voice characteristics or a combination of them that best detect a voice disorder, caused by a specific pathology, is still an open field.

A new method based on nonlinear data analysis has become popular: Recurrence Plots (RPs). Recurrence is a fundamental property of dissipative dynamical systems. Recurrence plots visualize recurrent behavior of dynamical systems. The method allows the identification of system properties that cannot be observed using other linear and nonlinear approaches and it is especially useful for analysis of nonstationary systems with high dimensional and/or noisy dynamics [16].

Recurrence plots were firstly used as a tool to visualize the behavior of phase-space trajectories. Zbilut and Weber [17] and later Marwan et al [18] have consolidated the method as a tool in nonlinear data analysis with the application of quantification measures to analyze the RPs. Recurrence quantification analysis (RQA) can be applied to almost every kind of data such as: biological systems including electromiography (EMG) [19], intracranial electroencephalogram recordings (EEG) [20], cardiac sound evaluation [21], Earth sciences [22], Finances [23] and computer networks [24]. Recently, but still little explored, these have been applied to speech signals analysis [25], [26], [13].

In this work, quantification measurements of recurrence...
A. Review of Recurrence Plots

A recurrence plot (RP) is a two-dimensional squared matrix with black and white dots and two time-axes, where each black dot at the coordinates \( i, j \) represents a recurrence of the system state \( x(i) \) at time \( j \) [16]:

\[
R_{i,j}^{m,\varepsilon} = \Theta(\varepsilon - \| \vec{x}_i - \vec{x}_j \|), \quad \vec{x}_i \in \mathbb{R}^m, \quad i, j = 1 \ldots N. \tag{1}
\]

where:
- \( N \) is the number of considered states \( \vec{x}_i \);
- \( \varepsilon \) is the neighborhood radius (threshold) at the point \( \vec{x}_i \);
- \( \| \cdot \| \) is the norm in the neighborhood, usually the Euclidean norm;
- \( \Theta(\cdot) \) is the Heaviside function;
- \( m \) is the embedding dimension of the system (degrees of freedom).

Figure 1 shows recurrence plots examples of healthy and pathological signals, respectively. From the examples in Figure 1, it is possible to detect the presence of small diagonal lines in the RP of the healthy speech signal. However, in the recurrence plot of the pathological speech signal there is a predominance of isolated points.

Recurrence plots (RP) visualization is a subjective method and has the disadvantage that the user had to detect and interpret the pattern and structures revealed by them. Zbilut and Webber [27], with the quantification of the RP structures, provided an objective and modern tool of nonlinear data analysis [28]. The quantification measures are described below.

B. Recurrence Quantification Measures

In this work, the recurrence quantification analysis applied to healthy and pathological voices is fulfilled by using the following measurements [16], [29]:

The density of recurrence points in a recurrence plot within a specified radius is represented by Recurrence Rate (REC) or per cent recurrences. This measure simply counts the black dots in the recurrence plot. This coincides with the correlation sum and is given by:

\[
REC = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j}^{m,\varepsilon}, \tag{2}
\]

Determinism (DET) - It is related with the determinism of the system, which is the fraction of recurrence points forming diagonal lines. Processes with stochastic behaviour cause none or very short diagonals, whereas deterministic cause longer diagonals and less single, isolated recurrence points. The ratio of recurrence points that form diagonal structures to all recurrence points is given by:

\[
DET = \frac{\sum_{l=l_{\min}}^{L_{\max}} P^c(l)}{\sum_{l=l_{\min}}^{L_{\max}} R_{i,j}^{m,\varepsilon}}, \tag{3}
\]

where \( P^c(l) \) represents the frequency distribution of diagonal structures lengths \( l \) in the RP and \( l_{\min} \) is the minimum number of points to form a diagonal structures in the RP.

Maximum diagonal line length (\( L_{\max} \)) - corresponding to a mean prediction time or to the inverse of the divergence of the system. Eckmann [30] has stated that this measure is related to the largest Lyapunov positive exponent:

\[
L_{\max} = \max(\{l_i, i = 1, \ldots, N_l\}) \tag{4}
\]

Entropy (ENTR) - The Shannon entropy of frequency distribution of the diagonal line lengths measures the complexity of the deterministic structure in the system:

\[
ENTR = - \sum_{l=l_{\min}}^{N} P(l) \ln p(l), \tag{5}
\]

where \( p(l) = \frac{P^c(l)}{\sum_{l=l_{\min}}^{L_{\max}} P^c(l)} \).

Slope of line of best fit (TREND) - It is a linear regression coefficient over the recurrence point density REC of the diagonals parallel to the mean diagonal (line of identity, LOI). It gives information about a nonstationarity in the process, especially a drift:

\[
TREND = \frac{\sum_{i=1}^{N} (i - \bar{N}/2)(REC_i - \langle REC \rangle)}{\sum_{i=1}^{N} (i - \bar{N}/2)^2} \tag{6}
\]

Laminarity (LAM) - Represents the fraction of recurrence points forming vertical lines (the ratio between the recurrence
A. Database

Discriminant function: linear (LDA) and quadratic (QDA). MATLAB 2009 (Mathworks) using two different Recurrence Quantification Analysis software (RQA) [29]. Extracted keeping the maximum recurrence rate at 1% using MATLAB, in Eq. (4), by means of the equation:

\[
V_{\text{max}} = \max \{ V_l; l = 1 \ldots L \}. \tag{9}
\]

III. MATERIAL AND METHODS

A. Database

In this work, the quantification recurrence measurements are extracted of sustained vowels of speech signals recordings from Disordered Voice Database, Model 4337, developed by Kay Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab [31]. The database includes samples from 118 patients with a wide variety of voice disorders. All samples were collected in a controlled environment with the following features: low-noise-level, constant microphone distance, direct digital 16-bit sampling and robust signal conditioning. The duration of these vowel samples was 3 s for healthy voices and 1 s for pathological voices. A 25 or 50 (kHz) sampling rate was employed to pathological and healthy voices, respectively. All the files were down-sampled to 25 kHz and a single 400 ms frames of analysis was considered for each signal.

The selected cases comprise 118 patients with pathological voices (45 with vocal fold edema, 18 with nodule and 55 patients with vocal fold paralysis) and 53 patients with healthy voices.

B. Methodology

The methodology applied in this work is summarized in Fig. 2. First, the embedding dimension (m) and the optimum time delay (τ) for each signal are extracted using the Visual Recurrence Analysis software (VRA) [32]. After obtaining τ and m, N vectors of dimension m are formed from s(t) and with their τ, 2τ, . . . , (m − 1)τ delayed versions [33].

The quantification measures mentioned in Section II are extracted keeping the maximum recurrence rate at 1% using Recurrence Quantification Analysis software (RQA) [29]. Discriminant analysis is applied by using the classify function of MATLAB 2009 (Mathworks) using two different discriminant function: linear (LDA) and quadratic (QDA).

IV. RESULTS

The MQR classification performance is evaluated by means of 10-fold cross-validation experiments. Firstly, for each of individual features and then, to improve the classifier performance, the features are combined two-by-two, three-by-three, four-by-four, five-by-five, six-by-six and, finally, all the seven measures, forming hybrid vectors.

A. Individual Measures Classification

The overall classification performance obtained by means of cross-validation (95% confidence), for each discriminant function (LDA and QDA), to individual features are shown in Tables I-III.

TABLE I

<table>
<thead>
<tr>
<th>Measures</th>
<th>LDA</th>
<th>QDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>DET</td>
<td>78.78 ± 9.21</td>
<td>79.33 ± 7.53</td>
</tr>
<tr>
<td>L_{max}</td>
<td>82.78 ± 7.65</td>
<td>82.44 ± 7.99</td>
</tr>
<tr>
<td>ENTR</td>
<td>80.89 ± 10.85</td>
<td>80.44 ± 5.74</td>
</tr>
<tr>
<td>TREND</td>
<td>65.22 ± 10.37</td>
<td>62.11 ± 11.04</td>
</tr>
<tr>
<td>LAM</td>
<td>57.22 ± 9.48</td>
<td>57.11 ± 10.12</td>
</tr>
<tr>
<td>V_{max}</td>
<td>73.44 ± 9.97</td>
<td>71.33 ± 6.76</td>
</tr>
<tr>
<td>TT</td>
<td>65.11 ± 14.31</td>
<td>65.44 ± 8.81</td>
</tr>
</tbody>
</table>

It can be observed, in Table I, that the best classification rate (maximum mean accuracy) obtained is 82.78±7.65% for

![Classification system based on recurrence quantification analysis.](image)
using $L_{\text{max}}$ and LDA classifier. To this case, it was obtained a mean false positive rate of 23.67±13.32% and 8.00±13.82% to false negative rate.

For Healthy x Paralysis discrimination (Fig. 5), a maximum mean accuracy of 95.71±4.94% was obtained with four different combinations: $DET$, $L_{\text{max}}$, $ENTR$, $V_{\text{max}}$, and $T_T$, using the LDA classifier. In this case, a mean false positive rate of 6.00±6.91% and a mean false negative rate of 11.00±8.40% were obtained.

The results obtained in the Healthy x Paralysis discrimination presented in the Table III give a maximum mean accuracy of 91.78±5.69%, combining five measures: $DET$, $L_{\text{max}}$, $ENTR$, $V_{\text{max}}$, and $T_T$. In this case, a mean false positive rate of 6.00±6.91% and a mean false negative rate of 11.00±8.40% were obtained.

**B. Results of Combined Measures**

In order to get better classification rates, the measures were combined. Figures 3-5 summarize the best results obtained for the individual and combined cases. For the Healthy x Edema (Fig. 3) rates, the maximum mean accuracy obtained was 92.64±5.27% and 5.00±11.31%, respectively.

**TABLE II**

**CORRECT CLASSIFICATION RATE (%) FOR INDIVIDUAL FEATURES - Healthy x Nodule.**

<table>
<thead>
<tr>
<th>Measures</th>
<th>LDA</th>
<th>QDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$DET$</td>
<td>84.46 ± 5.84</td>
<td>84.64 ± 7.23</td>
</tr>
<tr>
<td>$L_{\text{max}}$</td>
<td>85.89 ± 11.81</td>
<td>93.04 ± 5.27</td>
</tr>
<tr>
<td>$ENTR$</td>
<td>87.32 ± 5.79</td>
<td>87.14 ± 10.17</td>
</tr>
<tr>
<td>$TREN$</td>
<td>61.96 ± 6.82</td>
<td>56.43 ± 12.85</td>
</tr>
<tr>
<td>$LAM$</td>
<td>53.39 ± 7.19</td>
<td>53.75 ± 11.96</td>
</tr>
<tr>
<td>$V_{\text{max}}$</td>
<td>74.64 ± 10.53</td>
<td>75.89 ± 8.54</td>
</tr>
<tr>
<td>$T_T$</td>
<td>63.21 ± 12.17</td>
<td>63.21 ± 15.52</td>
</tr>
</tbody>
</table>

In the case of Healthy x Nodule discrimination (Table II), a maximum mean accuracy of 93.04±5.27% was obtained using $L_{\text{max}}$ and QDA. The mean false positive and false negative rates were 7.33±6.82% and 5.00±11.31%, respectively.

**TABLE III**

**CORRECT CLASSIFICATION RATE (%) FOR INDIVIDUAL FEATURES - Healthy x Paralysis.**

<table>
<thead>
<tr>
<th>Measures</th>
<th>LDA</th>
<th>QDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$DET$</td>
<td>81.36 ± 6.40</td>
<td>85.18 ± 3.41</td>
</tr>
<tr>
<td>$L_{\text{max}}$</td>
<td>90.82 ± 6.14</td>
<td>92.64 ± 5.15</td>
</tr>
<tr>
<td>$ENTR$</td>
<td>89.82 ± 4.79</td>
<td>88.09 ± 4.36</td>
</tr>
<tr>
<td>$TREN$</td>
<td>67.73 ± 9.83</td>
<td>65.55 ± 9.69</td>
</tr>
<tr>
<td>$LAM$</td>
<td>52.91 ± 10.58</td>
<td>60.45 ± 10.73</td>
</tr>
<tr>
<td>$V_{\text{max}}$</td>
<td>80.45 ± 10.05</td>
<td>78.91 ± 7.38</td>
</tr>
<tr>
<td>$T_T$</td>
<td>65.91 ± 8.96</td>
<td>61.36 ± 13.73</td>
</tr>
</tbody>
</table>

The results obtained in the Healthy x Paralysis discrimination presented in the Table III give a maximum mean accuracy of 92.64±5.15%, when employing $L_{\text{max}}$ and QDA. For this case, the mean false positive and false negative rates obtained were 9.67±9.96% and 5.33±15.58%, respectively.

As it can be seen in Tables I-III, the use of $L_{\text{max}}$ measure followed by $ENTR$ and $DET$ provides, individually, the best results concerning the task of discriminating healthy voices from pathological ones.
V. CONCLUSION

In this paper, the recurrence quantification measures were applied to discriminate between healthy and pathological voices. In the classification by means of the individual features, the maximum diagonal line length ($L_{\text{max}}$) provided the higher correct classification rates. This measure also appears in all the best results obtained by feature combinations. As expected, the combination of a set of appropriate measures increased the classification system performance related to the individual cases. Besides, the high standard deviation rates obtained by the individual classifiers, probably caused by the different degrees of severity of the pathologies, were reduced after feature combination. The obtained results suggest that the applied method provide a good potential for discrimination between groups of healthy voices and voices affected by edema, nodule and vocal fold paralysis.

REFERENCES