Acoustic Assessment of the Voices of Children Using Nonlinear Analysis: Proposal for Assessment and Vocal Monitoring

*Leonardo Wanderley Lopes, †Silvana Luciene do Nascimento Cunha Costa, †Washington César de Almeida Costa, Suzete Elida Nobrega Correia, and †Vinícius Jefferson Dias Vieira, *João Pessoa, Paraíba, Brazil

Summary: Objective. To analyze the accuracy of recurrence measurements, both isolated and combined, to assess the intensity of vocal disorders in children.

Method. A total of 93 children of both sexes (48 girls and 45 boys), aged between 3 and 10 years, participated. The vocal-deviation intensity was evaluated by the consensus of three speech therapists from the pronunciation of vowel /ε/ using the visual analog scale. In the acoustic analysis, eight recurrence plot characteristics were evaluated and extracted with neighborhood radius values that maintained the recurrence rate at 1%, 2%, 3%, 4%, and 5%. The classification was performed using quadratic discriminant analysis applied for individual and combined measurements. The performance was evaluated by measuring the accuracy, which related the cases correctly classified to all the analyzed cases.

Results. In the classification cases concerning individual measure performance, the trapping time and maximum length of the diagonal lines showed the best classification potential to discriminate between healthy and disturbed voices, with accuracy rates above 80%. In the healthy and mild deviation cases, the trend (TREND) measure was also relevant. For the mild versus moderate deviation classification, the best performance was obtained by the TREND measure (85.00% ± 7.64%). A gain was obtained in the classification rate when the measures of recurrence were combined, reaching an accuracy of 95.00% ± 5.00%, for discriminating between healthy voices and those with mild deviation.

Conclusions. The measures of recurrence, either alone or combined, may be useful in detecting healthy and disturbed voices and in differentiating the intensity of vocal disorders in children.

Key Words: Voice analysis–Acoustic–Children–Nonlinear analysis.

INTRODUCTION

The high prevalence of dysphonia in children requires special attention in the assessment and diagnosis of their voices by developing objective measures that provide an understanding of the intensity of vocal deviation and its manifestation in different periods aged between 3 and 9 years. 1–4

An acoustic signal is a complex product of the nonlinear interaction between the aerodynamic and biomechanical properties of the vocal production system. Therefore, a large correlation exists between laryngeal physiology and acoustic measures. 5

Acoustic assessment is considered less subjective than perceptive analysis to provide quantification of the vocal deviation. However, findings are incipient to assert that a single instrumental assessment can be consistently and strongly correlated with perceptual analysis. 6

An incessant search has been performed by clinicians and researchers to develop noninvasive measurements with high discriminatory power, enabling screening, assessment, diagnosis, and monitoring of voice disorders. In the literature, no consensus has been reached on the set of measures with greater accuracy for evaluating voice signals. Therefore, an urgent need exists for studies to investigate the discriminatory power of individual and/or combined acoustic measures that can be used in the classification of healthy and disturbed voices. 7

In recent years, techniques based on nonlinear dynamic analysis and chaos theory have been used in acoustic analysis both in the classification of healthy and disturbed voices because of the different degrees of vocal disorder and the different types of voice quality. 8,9

The methods of nonlinear dynamics can analyze irregular behavior and may be important in different studies of vocal production, including studies that evaluate the effectiveness of the treatment provided, classify voices in different degrees of modification, and differentiate between healthy and disturbed voices, and even contribute to the diagnosis of laryngeal diseases. 10–12

Among the measures used in the nonlinear dynamic analysis are the recurrence plots (RPs) and their quantification measures, which are based on the Poincaré recurrence theorem and have the advantage of working on short and nonstationary series. 13–15

RP is a square matrix that represents the evolution of a dynamic system, which can be visually analyzed qualitatively and subjectively by observing its structure of isolated points as well as the diagonal, vertical, and horizontal lines (Figure 1).

In the case of the voice production system, only one state variable is known, which is the digitized voice signal. For such representation, the temporal series under analysis needs to be
immersed in a phase space from the Takens’ immersion theorem,\textsuperscript{16} given by

\[ x_i = f(x(t_i), x(t_i + \tau), x(t_i + 2\tau), ..., x(t_i + (m-1)\tau)) \quad (1) \]

which represents the \( i \)th vector constructed from the original signal and its \( m \) lagged versions in time, where \( \tau \) is the delay applied to the voice signal and \( m \) is the immersion dimension. Using the Takens’ theorem, variables that attempt to reproduce the dynamics of speech production system are created.

RP is formed according to the following equation\textsuperscript{17}:

\[ \overline{R}_{ij}^{m,n} = \theta(\varepsilon - ||\overline{x}_i - \overline{x}_j||), \quad \overline{x}_i \in \mathbb{R}^n, \quad i, j = 1, ..., N. \quad (2) \]

where

- \( N \) is the number of \( \overline{x}_i \) states considered
- \( \varepsilon \) is the neighborhood radius (threshold) at point \( \overline{x}_i \)
- \( ||.|| \) is the neighborhood measure, usually the Euclidean measure
- \( \theta(.) \) is the unit step function
- \( m \) is the immersion dimension.

If \( \overline{R}_{ij} = 1 \), the state is considered recurrent; as a result, a black dot is marked on the RP. If \( \overline{R}_{ij} = 0 \), the state is “nonrecurrent,” and a white dot is marked on the RP.\textsuperscript{7,15,18,19}

Figures 1 and 2 show examples of the RPs obtained from classic signals such as a sine wave (a pure tone, a deterministic signal), without noise addition, and white noise (a random signal that contains equal power within any frequency band width). The diagonal lines of the RPs in the periodic signals are completely filled (Figure 1), whereas in the nonstationary random signals such as the white noise, the plot shows the predominance of isolated points (Figure 2).

The RPs obtained from signals of infant voices are shown in Figures 3–5, which represent the voice signals of a healthy child, with mild and moderate degree of deviation, respectively.
Recurrence rate (REC)—measures the density of the recurrence points in the RP; it is defined as

\[
REC = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j} \text{ with } i \neq j.
\]

(3)

Determinism (DET)—ratio of the number of recurrence points that form diagonal structures to all recurrence points. It is related to the predictability of the system and is defined as

\[
DET = \frac{\sum_{l=l_{\text{min}}}^{N} p^e(l)}{\sum_{l=1}^{N} R_{ij}}.
\]

(4)

where \(p^e(l) = \{i; i = 1 \ldots N_l\}\) represents the frequency distribution of lengths \(l\) of the diagonal structures in the RP and \(N_l\) is the total number of diagonal lines.

Maximum length of diagonal lines \(L_{\text{max}}\)—represents the length value of the longest diagonal line of the RP. The length of the diagonal lines is determined by the time during which two segments evolve in parallel on a trajectory. In this context, \(L_{\text{max}}\) is precisely the measure that quantifies the maximum time that two segments remain with a similar evolution over time. The longer the two segments remain with a similar pattern that has a well-defined trajectory over time, the longer is the length of the diagonal lines in the RP. It is defined as

\[
L_{\text{max}} = \max\{\{l; i = 1, \ldots, N_l\}\}.
\]

(5)

Shannon entropy (ENTR)—represents the frequency distribution of the diagonal line lengths and reflects the complexity of the deterministic structure present in the system; it is defined as follows:

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

where \(p(l) = \frac{p^e(l)}{\sum_{l=l_{\text{min}}}^{N} p^e(l)}\).

(6)

Trend (TREN)-a linear regression coefficient of the density of the recurrence points of the diagonals parallel to the main diagonal (identity line). This measure provides information regarding the nonstationarity of the process, and it is defined by

\[
\text{TREN} = \frac{\sum_{i=1}^{N} (i - \bar{N}/2)(\text{REC}_i - \langle \text{REC} \rangle)}{\sum_{i=1}^{N} (i - \bar{N}/2)^2}.
\]

(7)

Laminarity (LAM)—the ratio of recurrence points that form the vertical structures to the whole set of all recurrence points on the plot. LAM represents the occurrence of laminar states in the system without describing the length of these laminar phases. If the RP consists of more single recurrence points than vertical structures, the LAM will decrease.\(^{15}\) It is related to the occurrence of recurrent states that do not change over time, that is, the amount of laminar phases in the system (intermittency). It is defined as

\[
\text{LAM} = \frac{\sum_{\nu=\nu_{\text{min}}}^{\nu_{\text{max}}} v^e(\nu)}{\sum_{\nu=\nu_{\text{min}}}^{\nu_{\text{max}}} R_{ij}}.
\]

(8)

where \(\nu\) is the size of the vertical structure, \(p^e(\nu)\) is the probability of this vertical structure occurring within the RP, and \(\nu_{\text{min}}\) is the minimum number of vertical structures to be counted.

Mean length of the vertical structures (TT)—also known as the time spent in a state (trapping time). This measure contains information about the amount and length of the vertical structures in the RP.\(^{15}\) It measures the mean time that the system remains in a particular state.

\[
\text{TT} = \frac{\sum_{\nu=\nu_{\text{min}}}^{\nu_{\text{max}}} v^e(\nu)}{\sum_{l=l_{\text{min}}}^{N} p^e(l)}.
\]

(9)

FIGURE 4. Recurrence plot obtained from a child’s voice with mild deviation (time delay \(\tau = 10\) and embedded immersion dimension \(m = 7\)).

FIGURE 5. Recurrence plot obtained from a child’s voice with moderate deviation (time delay \(\tau = 10\) and embedded immersion dimension \(m = 7\)).
Maximum length of vertical structures \( (V_{\text{max}}) \)—provides the maximum time that a state remains in a laminar state, that is,

\[
V_{\text{max}} = \max(\{v_i; i = 1, \ldots, N_v\}).
\] (10)

To develop an efficient tool for acoustic analysis, the choice of the most appropriate characteristics for vocal and laryngeal conditions to be searched and the choice of the classifier can decisively influence the final result of the classification obtained. Although a child’s voice signal is even more complex than an adult’s voice signal and that often, linear methods cannot account for the complexity of the child’s voice, the analysis of nonlinear dynamics could be a new opportunity for evaluating the voices of children.

In this context, the aim of this study is to analyze the accuracy of recurrence measurements, both isolated and combined, to assess the intensity of voice disorders in children.

**METHOD**

This study is a quantitative, descriptive, cross-sectional study approved by the Research Ethics Committee of the Institution of Origin under protocol number 775/10. A total of 93 children of both sexes (48 girls and 45 boys) aged between 3 and 10 years, all members of a school linked to a federal public educational institution, participated in the study.

Children with cognitive impairment, those with changes in the upper airways at the time of the data collection, and those who could not accomplish the requested speech task were excluded from the study. Only six 3-year-old children were excluded from this study because they could not complete the speech task. Four children were excluded because they refused to use the microphone and utter the requested vowel, and the other two were excluded because they did not utter the required speech in less than 3 seconds. The recording sessions were conducted in a quiet room and had a mean duration of 5 minutes.

The task requested was for the emission of the sustained vowel /e/ in the maximum phonation time. The data were collected using an HP notebook (Hewlett-Packard Development Company, Palo Alto, CA) and a microphone headset (Logitech, Fremont, CA) using PRAAT software, version 5.1.44 (P. Boersma and D. Weenink, University of Amsterdam, The Netherlands) with a sampling rate of 44.100 Hz. The study was conducted from March to October 2012.

Subsequently, the voices were edited in the Sound Forge software, version 10.0 (Sony, Tokyo, Japan) by eliminating the initial 1 second and the final 2 seconds from the vowel emission because of the large irregularity in these passages; however, a minimum time of 2 seconds for each emission was preserved. The normalization of the signals was carried out in the Sound Forge control “normalize” at a peak-level mode to standardize the audio output from \(-6\) to \(6\) dB.

For the perceptual-auditory analysis of the voice, we chose the visual analog scale (VAS) with a metric of 0–100 mm. The nearest marking to zero “0” represented a minor change, whereas the nearest to hundred “100” represented a major change. This assessment was performed by the consensus of three speech therapists (voice specialists) with expertise in the perceptual-auditory voice evaluation. The intensity (overall grade) of the vocal deviation (general grade [GG]), roughness grade, breathiness grade, strain grade, and instability grade were evaluated.

The perceptual evaluation session was performed in a quiet environment. Initially, the judges were told that the voices should be considered healthy when they were socially acceptable for a child with a grade 1 VAS (0–35.5 mm). They were also instructed that roughness would correspond to the presence of vibratory irregularity, breathiness would be related to audible air leak during the emission, strain would correspond to the perception of vocal effort, and instability would be identified by the presence of vocal quality and floating rate during the emission. Each emission of the sustained vowel was presented three times through a sound box at a comfortable intensity self-reported by the evaluators. Subsequently, the presence or absence of vocal deviation and the predominant voice in the diverted voices (rough, breathy, strained, or unstable) was identified. Finally, the judgment of the deviation intensity (GG—overall grade) was made.

At the end of the perceptual evaluation session, 10% of the samples were randomly repeated for reliability analysis by the consensus of the judges using the Cohen kappa coefficient. The kappa value was 0.80, indicating good agreement among raters.

Subsequently, matching of the numerical scale (EN) was performed for VAS with grade 1 (0–35.5 mm) indicating healthy voices, grade 2 (35.6–50.5 mm) indicating mild deviation, grade 3 (50.6–90.5 mm) indicating moderate deviation, and grade 4 (90.6–100 mm) indicating severe deviation.

For nonlinear acoustic evaluation, the immersion parameters in the speech production system (immersion dimension \( m \) and the step of reconstruction or time delay \( \tau \) ) were obtained using the TISEAN software, version 3.0.1 (Hegger and Holger Kantz, Max Planck Institute for Physics of Complex Systems, Dresden Saxony, Germany; Thomas Schreiber, Physics Department, University of Wuppertal, Wuppertal, North Rhine—Westphalia, Germany). The RQA, from which the measures of recurrence quantification (recursive least squares) were extracted, was performed using the routine package RQA, version 13.1 (Charles L. Webber Jr., Department of Cell and Molecular Physiology, Loyola University Chicago, Stritch School of Medicine, Maywood, IL).

The recurrence measures were extracted using the neighborhood radius values that maintained the rate of recurrence at five levels: 1%, 2%, 3%, 4%, and 5%. Thus, the behavior of the eight characteristics of recurrence, namely, neighborhood radius (RADIUS), DET, \( L_{\text{max}} \), ENTR, TREND, LAM, TT, and \( V_{\text{max}} \), was observed.

The signal classification was performed using the quadratic discriminant function (quadratic discriminant analysis) by the MATLAB 2009 classify function (Mathworks, Natick, MA). The classifier was tested by a 10-fold cross validation. The whole set was divided into 10 subsets, where 90% of the signals were used to train the system and 10% were used for the testing, which were chosen randomly and without repetition. This process was carried out 10 times. The performance evaluation was performed by measuring the accuracy (computed by the
average of the classification rate obtained on each repetition), which related the cases correctly classified to all the cases evaluated.

First, an individual classifier was used for each recurrence measure. Then, the eight measures were combined to increase the performance: first by two-by-two combination, then by three-by-three, four-by-four, and until all eight combinations were completed one at a time.

RESULTS

Regarding the intensity of vocal deviation (GG), 70.7% (n = 70) exhibited mild deviation and 13.1% (n = 13) had voices with moderate deviation. Only 10.1% (n = 10) of the subjects had healthy voices. No child showed severe deviation in the vocal quality (Table 1).

Because of the disparity of the group sizes, a subset of the disturbed voices was randomly selected to equate the voice sets considered in the analysis. In this way, 10 voices from each group were used in the classification system.

Table 2 lists the results obtained for the classification between the healthy (grade 1) and disturbed (grades 2 and 3) voices obtained from the individual recurrence measure analysis. The TT and $L_{\text{max}}$ measures showed significant classification potential to discriminate between grade 1 and grades 2 and 3, considering a level of 1% and 3% of recurrence rate, respectively. Therefore, the TT was better in detecting the absence of voice deviation, whereas $L_{\text{max}}$ is more accurate in detecting deviation.

The classification results for grades 1 and 2 of the individual measure performance are listed in Table 3. In this case, $L_{\text{max}}$, TREND, and TT exhibited the best accuracy rates. The TT and TREND yielded the best specificity rates, and $L_{\text{max}}$ exhibited the best sensitivity, similar to that in the earlier case (Table 2).

The results obtained for the classification between the mild (grade 2) and moderate (grade 3) deviation of voices with regard to the individual recurrence measure analysis are listed in Table 4. The best accuracies were obtained by the TREND and $L_{\text{max}}$ measures. In this case, $L_{\text{max}}$ was better in detecting the presence of mild deviation, similar to what occurred in the earlier related cases. TREND, however, was the best measure to detect moderate deviation when these two measures were compared.

Table 5 lists the performance evaluation of the classification using the combination of the quantification measures. Only the best combinations obtained in each classification case are presented. The combinations of the measures did not increase the accuracy in discriminating the healthy from the disturbed voices (grade 1 × grade 2 and grade 3). However, with regard to the classification of grade 2 versus grade 3, the rates were higher than those obtained with the best individual measures.

DISCUSSION

The data on the prevalence of voice disorders in childhood have attracted experts to provide greater attention to the child’s voice, from the development of objective measures that promote understanding of the intensity of vocal deviation and its manifestation in different periods between ages 3 and 9 years to the best approaches for voice problem solutions in this age range.1–3,22

Acoustic analysis allows integrating physiological and auditory levels, quantifying the perceptual characteristics, and inferring the larynx function because voice is a physical phenomenon that suffers from the effects of neurophysiological mechanisms.

### TABLE 1.
Occurrence, Mean, and Standard Deviation of the Intensity of Vocal Deviation by VAS

<table>
<thead>
<tr>
<th>Vocal Deviation Intensity</th>
<th>Healthy n (%)</th>
<th>Mild n (%)</th>
<th>Moderate n (%)</th>
<th>Mean VAS ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG</td>
<td>10 (10.1)</td>
<td>70 (70.7)</td>
<td>13 (13.1)</td>
<td>44.19 ± 7.67</td>
</tr>
</tbody>
</table>

**Abbreviations:** GG, general grade of vocal disorder; VAS, visual analog scale.

### TABLE 2.
Classification of the Vocal Deviation Severity by Individual Recurrence Measures—Grade 1 × (Grades 2 and 3)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Accuracy (%)</th>
<th>REC (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RADIUS</td>
<td>60.00 ± 6.67</td>
<td>1</td>
<td>40.00 ± 16.34</td>
<td>80.00 ± 13.34</td>
</tr>
<tr>
<td>DET</td>
<td>50.00 ± 7.46</td>
<td>5</td>
<td>20.00 ± 13.42</td>
<td>80.00 ± 13.34</td>
</tr>
<tr>
<td>$L_{\text{max}}$</td>
<td>80.00 ± 11.06</td>
<td>3</td>
<td>90.00 ± 10.00</td>
<td>70.00 ± 15.29</td>
</tr>
<tr>
<td>ENTR</td>
<td>60.00 ± 10.01</td>
<td>4</td>
<td>50.00 ± 16.67</td>
<td>70.00 ± 15.29</td>
</tr>
<tr>
<td>TREND</td>
<td>65.00 ± 7.64</td>
<td>2</td>
<td>40.00 ± 16.34</td>
<td>90.00 ± 10.00</td>
</tr>
<tr>
<td>LAM</td>
<td>60.00 ± 10.01</td>
<td>1</td>
<td>60.00 ± 16.34</td>
<td>60.00 ± 16.34</td>
</tr>
<tr>
<td>TT</td>
<td>85.00* ± 7.64</td>
<td>1</td>
<td>70.00 ± 15.29</td>
<td>100.00 ± 0.00</td>
</tr>
<tr>
<td>$V_{\text{max}}$</td>
<td>60.00 ± 10.01</td>
<td>1</td>
<td>30.00 ± 15.29</td>
<td>90.00 ± 10.00</td>
</tr>
</tbody>
</table>

**Abbreviations:** Grade 1, healthy; Grade 2, mild deviation of voice quality; Grade 3, moderate deviation of voice quality; RADIUS, neighborhood radius; REC, recurrence rate; DET, determinism; $L_{\text{max}}$, maximum length of diagonal lines; ENTR, Shannon entropy; TREND, trend; LAM, laminarity; TT, mean length of vertical structures; $V_{\text{max}}$, maximum length of vertical lines.

* Measures with better accuracy.
Many studies have attempted to determine the relationship between the acoustic characteristics of the speech signal and the perceived voice quality. Thus, establishing the following points has become increasingly important: correlation between these assessments such as how much the acoustic measurements are able to discriminate normal and abnormal voices as well as the discriminatory power between the different degrees of vocal disorders.2,6,23–25

Voice disorders tend to disturb the sound signal in different ways, combining different types of disturbances and noise. Therefore, some studies6,24 have indicated that a precise voice screening should use a combination of several noise acoustic measurements, which allow for each individual vocal emission to be measured by a single parameter set.25 However, the ideal would be the selection of the best acoustic characteristics, that is, those showing the highest correlation with the perceptual analysis and the major discrimination power between the normal and abnormal voices.6,24,26

A major problem in grouping the acoustic characteristics from the vocal assessment is that most of these characteristics are sensitive to different acoustic properties. The main question raised is what are the measures that yield the best evaluation regardless of the irregularity and the additive noise in the disturbed voices. It is only when these two properties are measured independently that one can show its relevance in assessing different vocal qualities identified in the perceptual analysis, such as roughness and breathiness and their physiological correlates.27 A given feature can extract specific information from the RPs, which represent the analyzed signals that are not captured by another. Different features can be complementary, and their combination becomes more efficient in the task of discriminating the vocal intensity deviation.

Studies28,29 based on the nonlinear model of speech production have highlighted the importance of considering the nonlinearities inherent in the vocal production process, resulting in complex factors such as temporal variation in the vocal tract shape, resonance associated with physiology, losses because of viscous friction on the inner walls of the vocal tract, softness of these internal walls, sound radiation in the lips, nasal coupling, and flexibility associated with the vibration of the vocal folds.

In the present study, we investigated the accuracy of the recurrence measures, used individually and in combination, to assess the intensity of the deviation in the voices of children.

The measures that have better accuracy for classifying healthy (grade 1) and disturbed (grades 2 and 3) voices were

<table>
<thead>
<tr>
<th>Measure</th>
<th>Accuracy (%)</th>
<th>REC (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RADIUS</td>
<td>55.00 ± 8.98</td>
<td>1</td>
<td>30.00 ± 15.29</td>
<td>80.00 ± 13.34</td>
</tr>
<tr>
<td>DET</td>
<td>50.00 ± 12.92</td>
<td>1</td>
<td>30.00 ± 15.29</td>
<td>70.00 ± 15.29</td>
</tr>
<tr>
<td>$L_{\text{max}}$</td>
<td>80.00* ± 8.17</td>
<td>5</td>
<td>90.00 ± 10.00</td>
<td>70.00 ± 15.29</td>
</tr>
<tr>
<td>ENTR</td>
<td>40.00 ± 9.25</td>
<td>1</td>
<td>30.00 ± 15.29</td>
<td>60.00 ± 16.34</td>
</tr>
<tr>
<td>TREND</td>
<td>75.00 ± 8.34</td>
<td>1</td>
<td>60.00 ± 16.34</td>
<td>90.00 ± 10.00</td>
</tr>
<tr>
<td>LAM</td>
<td>60.00 ± 12.48</td>
<td>1</td>
<td>60.00 ± 16.34</td>
<td>60.00 ± 16.34</td>
</tr>
<tr>
<td>TT</td>
<td>75.00 ± 8.34</td>
<td>1</td>
<td>60.00 ± 16.34</td>
<td>90.00 ± 10.00</td>
</tr>
<tr>
<td>$V_{\text{max}}$</td>
<td>70.00 ± 8.17</td>
<td>3</td>
<td>40.00 ± 16.34</td>
<td>100.00 ± 0.00</td>
</tr>
</tbody>
</table>

Abbreviations: Grade 1, healthy; Grade 2, mild deviation of voice quality; Grade 3, moderate deviation of voice quality; RADIUS, neighborhood radius; REC, recurrence rate; DET, determinism; $L_{\text{max}}$, maximum length of diagonal lines; ENTR, Shannon entropy; TREND, trend; LAM, laminarity; TT, mean length of vertical structures; $V_{\text{max}}$, maximum length of vertical lines.

* Measures with better accuracy.
the TT and \( L_{\text{max}} \), as listed in Table 2. The classification of the vocal deviation intensity for the healthy voice versus that with mild deviation (grade 1 vs grade 2) also yielded the best result, together with \( L_{\text{max}} \) and followed by the TREND and TT measures (Table 3). In the classification of voices with mild deviation versus moderate deviation (grade 2 vs grade 3), the most significant measures were TREND and \( L_{\text{max}} \).

The \( L_{\text{max}} \) measure is related to the diagonal lines of the RPs that occur when part of the trajectory of a segment (in this case, the voice signal) evolves parallel to another segment of the trajectory, that is, the trajectory enters the same region in the phase space at different times. In this context, \( L_{\text{max}} \) is precisely the measure that quantifies the maximum time when the two segments with a similar evolution remain.

The TT evaluates the mean time when the system remains in the same state, which is related to the chaotic behavior of the system. TT is defined as the average duration of a laminar state and represents the average length of the vertical structures present in the RPs. All disturbed voices of the considered signals used in the test phase of the classification were detected and discriminated from the healthy ones (100\% specificity), indicating that the measure of the frequency and laminar state (TT) is a relevant measure to quantify the vocal deviation in the voices of children.

The TREND measure presented a major potential to distinguish voices with moderate deviation from the mild ones, exhibiting an accuracy of 85.00\% ± 7.64\%. This measure provides information about nonstationarity in the process.\(^{15}\)

The more severe the vocal deviation is, the lesser is the process stationarity. Therefore, this measure could identify among the voices with deviation where the variation in their dynamics tends to be more abrupt or the dynamics of the more severe signals are more irregular.

A survey\(^{30}\) based on the nonlinear model evaluated the accuracy of single and combined acoustic measures in the classification of healthy and disturbed voices affected by laryngeal pathologies in adults. Used individually, \( L_{\text{max}} \) presented the highest accuracy rate and sensitivity in the classifier.

The findings of this study indicate that \( L_{\text{max}} \), TT, and TREND are relevant measures to detect the presence/absence of vocal disorders, and in rating the intensity of vocal deviation, when applied individually at a quadratic discriminant analysis-based classifier. Therefore, they may be useful for diagnostic and vocal screening procedures and for monitoring the intensity of vocal deviation in children throughout their development or pre- and postvocal therapy. However, the major challenge is to understand the physiological mechanism involved in this measure.

Multiple nonlinearities have been known to exist in the voice production process. The graphics and measures of recurrence emerged as among the nonlinear analysis techniques that seek to study the dynamics of the voice production process. Meanwhile, a dissipative system assumes that although small disturbances in a dissipative system cause divergence between states, the system returns to the previous state after some time and keeps evolving over time in the same manner.\(^{13}\) Thus, the recurrence measures allow quantitative analysis of the irregular activities in the vocal production mechanism.

The vibrations of the vocal folds as a dynamic system can be represented as a trajectory in the phase space that evolves over time. In the case of \( L_{\text{max}} \), the longer the two segments remain with a similar trend, that is, with a well-defined trajectory, the greater is the frequency of vibration of the vocal folds. Chaotic behavior is more evident (TT) in deviated voices. Slight irregularities in the sound signal are considered normal variations associated with the physiological mechanism of the vocal production. However, the increased values in disruption/irregularity of the acoustic signal may indicate a malfunction of the glottal mechanism. They may be used to detect the presence of a vocal disorder,\(^{7}\) which was observed in this study through the high discriminatory power of \( L_{\text{max}} \) and TT regarding the healthy and disturbed voices.

\( L_{\text{max}} \) and TREND have a direct relationship with the regularity/periodicity of a signal over time in which the RPs are related to the formation and length of the diagonal lines. Thus, the increased irregularity in voice signal, which is caused by the irregular vibration of the vocal folds, can change the values of \( L_{\text{max}} \) and TREND. Such measures can then be shown as strong predictors in identifying vocal deviation from the vibratory irregularities. The lighter the deviation is, the greater is the length of the diagonal lines. Smaller diagonal lines characterize the presence of more severe deviations (Figures 3–5).

In the perceptual level, the presence of irregular vibration may correspond to the presence of roughness and/or tension in the vocal emission. In turn, the surface roughness and strain, along with breathiness, are the most important universal parameters that detect the presence of a vocal or larynx disorder,\(^ {27}\) including those in the pediatric population.

In a study\(^ {31}\) conducted to evaluate the effectiveness of voice therapy in children with functional dysphonia, a larger change

---

**TABLE 5. Performance Evaluation of the Combined Recurrence Measures—Best Accuracy and Combination**

<table>
<thead>
<tr>
<th>Classification</th>
<th>Best Combination</th>
<th>REC (%)</th>
<th>Accuracy (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 1 × (grades 2 and 3)</td>
<td>RADIUS, ( L_{\text{max}} ), and ( V_{\text{max}} )</td>
<td>4</td>
<td>85.00 ± 7.64</td>
<td>90.00 ± 10.00</td>
<td>80.00 ± 13.34</td>
</tr>
<tr>
<td>Grade 1 × grade 2</td>
<td>RADIUS, ( L_{\text{max}} ), TREND, and TT</td>
<td>3</td>
<td>95.00 ± 5.00</td>
<td>100.00 ± 0.00</td>
<td>90.00 ± 10.00</td>
</tr>
<tr>
<td>Grade 2 × grade 3</td>
<td>RADIUS and TREND</td>
<td>2</td>
<td>95.00 ± 5.00</td>
<td>90.00 ± 10.00</td>
<td>100.00 ± 0.00</td>
</tr>
</tbody>
</table>

Abbreviations: Grade 1, healthy; Grade 2, mild deviation of voice quality; Grade 3, moderate deviation of voice quality; RADIUS, neighborhood radius; REC, recurrence rate; DET, determinism; \( L_{\text{max}} \), maximum length of diagonal lines; ENTR, Shannon entropy; TREND, trend; LAM, laminarity; TT, mean length of vertical structures; \( V_{\text{max}} \), maximum length of vertical lines.
occurred in the overall grade, roughness, and breathiness with a minor variation in the strain and instability parameters in terms of the perceptual-auditory evaluation. This fact may indicate the strain and instability parameters as most common in the speech pattern of children and are not always markers of the presence of vocal disorders or tissue changes in the larynx, unlike the breathiness and roughness parameters.

The vibration irregularities of the vocal folds can happen when the aerodynamic and biomechanical parameters, including subglottic pressure, stiffness, mass, and strain, exceed certain normal values because of the presence of asymmetry between vocal folds, mass lesion, or even dysfunctional settings that cause hyperkinetic frameworks with excessive phonatory tussae.

In the case of the voices of children, we have to remember that the vocal folds have not yet completed the process of differentiation of the intermediate and deep layers of the lamina propria. In addition, the infant is still under the development phase of neuromuscular control; thus, we expect that measures of disruption/irregularity find more alteration in the voices of children, particularly up to 5 years.

The best accuracy values for the combined characteristics were obtained with recurrence rates of 2–4%, indicating that the higher amount of recurrence points on the graph influenced the increase in the correct classification of signals. The presence of irregularities and noise during emission already compromised the representation of the sound signal in the graph; thus, the increased rate of recurrence allowed greater evidence of points, diagonal, and vertical structures, enabling improvement in the classification of signals.

We can observe that the combination of measures did not increase the accuracy of the classification of voices in the healthy or disturbed voices (grade 1 vs grades 2 and 3). However, in the classification of deviation intensity in the disturbed voices, the accuracy increased from 80.00% ± 8.17% with \( L_{\text{max}} \) to 95.00% ± 5.00% (grade 1 vs grade 2) when the RADIUS, \( L_{\text{max}} \), TREND, and TT measures were combined. The RADIUS measure increased the discriminative power of the other measures, which presented a relevant discriminative potential when applied individually. The neighborhood radius variation also played an important role in the detection of the deviations in grades 2 and 3. When associated with TREND, a variation in the RADIUS measure implies more or lesser recurrent points, increasing the perception of the system dynamic details. After the combination, both the achieved sensitivity and specificity rate values were higher than 90%. This result reinforces the possibility of using the recurrence measures in both screening procedures when monitoring the voice during voice therapy.

In a research on the combination of recurrence measures to evaluate the voice signals from a healthy larynx and those with benign lesions in the vocal folds, the highest accuracy value was 92.48% ± 6.49% for all measures of \( L_{\text{max}} \), entropy of the frequency distribution of the lengths of diagonal lines, LAM, and \( V_{\text{max}} \).

The methods of nonlinear dynamics can analyze the behavior with different degrees of irregularity and may provide complementary, rather than redundant, information over existing traditional methods mainly because they can analyze voices with greater aperiodicity.

Vocal production is a system that evolves over time, tending to equilibrium and presenting repeating cycles, but it may show simultaneous changes and irregular behavior. Thus, analysis by nonlinear models appears appropriate for the study and evaluation of voices, even those in the pediatric population.

However, although the nonlinear methods of analysis are potentially valuable tools for the study of speech, problems arise in applying these methods, particularly with regard to the effects of noise and the complexity of numerical algorithms. The development of simple and robust methods to analyze the chaotic activity in vocal production is important, particularly in the clinical screening routine, assessment, and diagnosis and monitoring of laryngeal and vocal disorders.

More studies are needed to examine the clinical relevance of these new methods that provide theoretical contributions and analysis of the voice signal. Practical analyses of these methods should be developed and tested by more extensive evaluations, higher number of subjects, and greater differences in age groups with different laryngeal lesions, as well as by verifying the applicability of the analysis of nonlinear dynamics in clinical practice.

**CONCLUSIONS**

The recurrence measures, either individually or in combination, may be useful in detecting healthy and disturbed voices and in the classification of voice deviation intensity in children. Thus, they can be used both in screening procedures and in monitoring during voice therapy.

**REFERENCES**
