WILEY | Hindawi

Research Article

Invariant Measures Based on the U-Correlation Integral: An Application to the Study of Human Voice

Juan F. Restrepo [1,2] and Gastón Schlotthauer[1,2]

[1]*Laboratorio de Señales y Dinámicas no Lineales, Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Ruta Prov. 11, Km 10, Oro Verde, Entre Ríos, Argentina*
[2]*Instituto de Investigación y Desarrollo en Bioingeniería y Bioinformática, Oro Verde, Entre Ríos, Argentina*

Correspondence should be addressed to Juan F. Restrepo; jrestrepo@bioingenieria.edu.ar

Nonlinear measures such as the correlation dimension, the correlation entropy, and the noise level were used in this article to characterize normal and pathological voices. These invariants were estimated through an automated algorithm based on the recently proposed U-correlation integral. Our results show that the voice dynamics have a low dimension. The value of correlation dimension is greater for pathological voices than for normal ones. Furthermore, its value also increases along with the type of the voice. The low correlation entropy values obtained for normal and pathological type 1 and type 2 voices suggest that their dynamics are nearly periodic. Regarding the noise level, in the context of voice signals, it can be interpreted as the power of an additive stochastic perturbation intrinsic to the voice production system. Our estimations suggest that the noise level is greater for pathological voices than for normal ones. Moreover, it increases along with the type of voice, being the highest for type 4 voices. From these results, we can conclude that the voice production dynamical system is more complex in the presence of a pathology. In addition, the presence of the inherent stochastic perturbation strengthens along with the voice type. Finally, based on our results, we propose that the noise level can be used to quantitatively differentiate between type 3 and type 4 voices.

1. Introduction

The human voice is the most important means of communication among individuals. Thanks to vocal communication, activities like asking for help are apparently trivial in our daily routine. Thus, a voice disorder can limit our ability to cover our most basic needs, producing a negative impact on our quality of life. For this reason, it is very important not only to increase our knowledge about the mechanism of voice production but also to characterize its dynamics in normal and pathological conditions.

In the literature, several methodologies to assess human voices can be found. However, their reliability depends on the nature of the studied voice. Consequently, Titze proposed a qualitative classification for voice signals [1]. The scheme proposed by Titze divides the signals into three types: type 1 signals are nearly periodic voice signals, type 2 signals have strong modulation or subharmonics, and type 3 signals are characterized by a very irregular or even chaotic behavior. Sprecher et al. proposed a modification to this scheme. They redefined type 3 voices as being deterministic chaotic signals, adding a fourth type that is characterized by a dominant random-like behavior [2].

Figure 1 shows the time series, the state space reconstruction, and the spectrogram of each type of voice. A normal voice (first column) is characterized for a quasiperiodic time representation and a smooth attractor in the reconstructed state space. Moreover, from its spectrogram, one can easily observe the fundamental frequency and its harmonics. A pathological type 1 voice (second column) displays a more irregular time series than the normal voice. Notice that although their attractors are similar, pathological type 1 attractor is not so smooth. Furthermore, both spectrograms are also similar in the sense that one can still distinguish a fundamental frequency and its harmonics. Nevertheless, the noise content blurs the harmonics at high frequencies.

In the case of a pathological type 2 voice (third column), it is possible to observe a less regular time series, compared to the pathological type 1 voice. The volume occupied by its attractor has been reduced and it is more difficult to distinguish a smooth shape. In its spectrogram, one can find subharmonic frequencies. The pathological type 3 voice (fourth column) is characterized by an irregular time series. The state space reconstruction is similar to the one of white Gaussian noise. Its spectrogram shows the fundamental frequency but its harmonics are rapidly blurred. Finally, a pathological type 4 voice (fifth column) has an irregular time series representation and a regular state space reconstruction, like the type 3 voice. In its spectrogram, one can find a fundamental frequency but not the harmonics.

These representations are useful to differentiate between some kinds of voices, for example, between normal and pathological type 3 voices. However, it is very difficult to differentiate between a normal and a pathological type 1 voice or between a type 3 and a type 4 voice. The importance of this classification is based on the fact that traditional perturbation measures like jitter and shimmer give reliable results if they are applied to type 1 and type 2 signals. In contrast, nonlinear dynamics concepts should be used to characterize type 3 voices, whereas they are unreliable when applied to type 4 signals. Moreover, until now, the classification of type 4 voices has been done subjectively by visual inspection [2].

Over the last three decades, strong evidence about the nonlinear behavior of the voice production mechanism has been collected [3–6], leading the scientific community to develop concepts and methods based on nonlinear dynamics and chaos theories. There exist an extensive set of publications in which those concepts have been used to characterize healthy and pathological voices. The correlation dimension (D) and the correlation entropy (K_2) are two quantities that are used to characterize the complexity of a dynamical system. The former can be thought of as an estimation of the number of variables involved in the dynamics (degrees of freedom) [7]. The latter measures the rate at which information about the dynamics is lost over time. More complex systems are commonly characterized by having higher dimensions and entropy values [8].

The correlation integral is the quantity used to estimate D and K_2 [9]. It has been widely used in the biomedical field since it was proposed by Grassberger and Procaccia in the early 1980s [9].

According to what we know so far, this correlation integral has been used in all studies that have estimated D and K_2 from voice signals. However, it is well known that the Grassberger and Procaccia approach is not robust in the presence of noise, and it needs special conditions to converge [10]. More robust variants of the correlation integrals have been proposed, like the Gaussian kernel correlation integral (GCI) [11] and the U-correlation integral [12].

The objective of this article is to characterize the dynamics of normal and pathological voices through the correlation dimension, the correlation entropy, and the noise level (σ). Furthermore, we seek to statistically analyze these invariants for the four types of voices. As a novelty, we obtain these invariants using U-correlation integral through an automated

algorithm [12] which allowed us to avoid subjective judgements. Finally, we propose a new quantitative methodology to differentiate between type 3 and type 4 voices.

2. Materials and Methods

The correlation dimension and the correlation entropy are invariants that characterize the natural measure of a dynamical system [13]. These invariants can be easily computed from the correlation integral which is obtained from indirect temporal measures of one of the variables of the system [9]. For an observed stationary time series of length N, $\{x_n\}_{n=1}^N$, the reconstructed m-dimensional delay vectors $\mathbf{x_i} = (x_i, x_{i+\tau}, \ldots, x_{i-(m-1)\tau})$, $i = 1, 2, \ldots, L = N - (m-1)\tau$, must be formed. The correlation integral $G_m(h)$, in its general form, is defined as the probability that the distance $\tilde{z} = \|\mathbf{x_i} - \mathbf{x_j}\|$ between two randomly selected m-dimensional delay vectors $\mathbf{x_i}$ and $\mathbf{x_j}$ is smaller than a value h [13]:

$$G_m(h) = \int g(h, \tilde{z}) f_{\tilde{z}}^m(\tilde{z}) \, d\tilde{z}, \tag{1}$$

where $g(h, \tilde{z})$ is a kernel function and $f_{\tilde{z}}^m(\tilde{z})$ is the probability density function (pdf) of the distance between pairs of delay vectors. In this article, we used the Euclidean distance.

The Grassberger and Procaccia correlation integral $C_m(h)$ uses as a kernel function $g(h, \tilde{z}) = H(1 - \tilde{z}/h)$, where $H(\cdot)$ is the Heaviside step function [9]. On the other hand, the GCI proposed by Diks et al., $T_m(h)$, adopts $g(h, \tilde{z}) = \exp(-\tilde{z}^2/4h^2)$ [11]. The main advantage of the GCI over the Grassberger and Procaccia correlation integral is that the former allows us to model the influence of additive noise on the scaling law.

For a zero-mean time series of variance σ_c^2 with additive white Gaussian noise of variance σ_n^2, the scaling rule is [11, 14, 15]

$$T_m(h) \sim h^m \left(h^2 + \sigma^2\right)^{(D-m)/2} e^{-m\tau K_2} \tag{2}$$
$$\text{for } m \longrightarrow \infty, \ \sqrt{h^2 + \sigma^2} \longrightarrow 0,$$

where

$$\sigma = \frac{\sigma_n}{\sqrt{\sigma_c^2 + \sigma_n^2}}. \tag{3}$$

It is important to say that all time series used in this work were rescaled to have unitary standard deviation. In this sense, σ is the noise level after rescaling; that is, $\sigma \to 0$ corresponds to clean time series and $\sigma \to 1$ implies a pure stochastic process. The signal-to-noise ratio (SNR) can be calculated as

$$\text{SNR} = 10 \log_{10} \left(\frac{1 - \sigma^2}{\sigma^2}\right) \text{dB}. \tag{4}$$

In practical applications, it is very important to be able to quantify σ, the noise level in the time series. This is because it allows us to correct the estimations of D and K_2. Moreover, the conclusions driven by the interpretation of the invariants
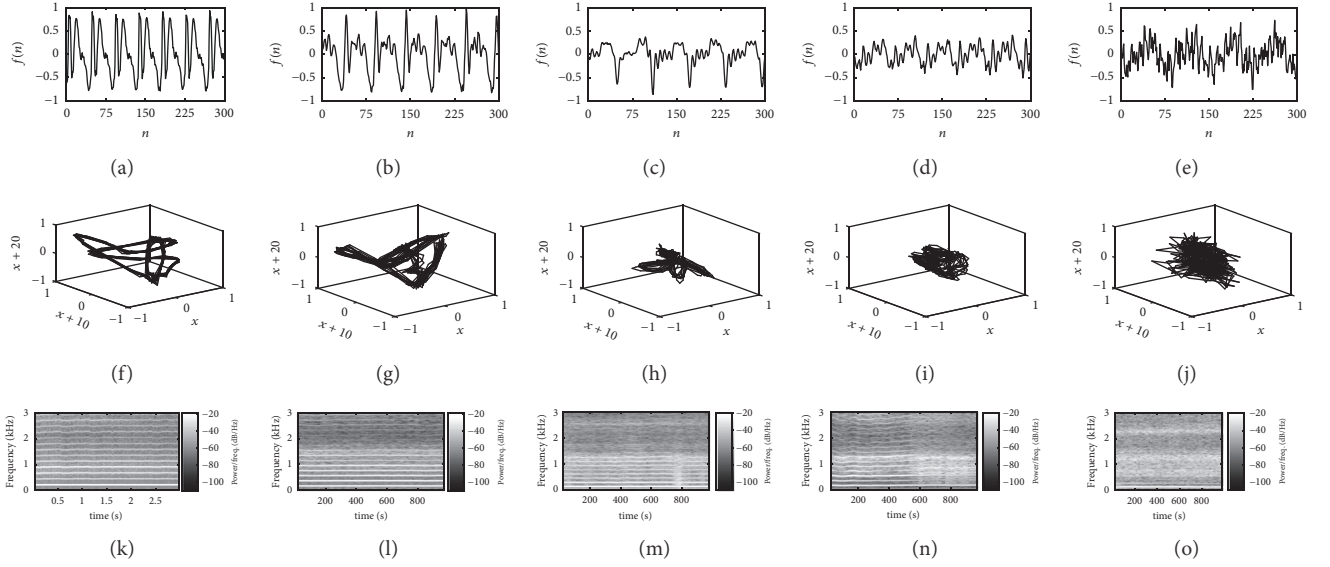
FIGURE 1: Time series, state space reconstruction, and spectrogram of normal and pathological voices. First column: normal voice; second column: pathological type 1 voice; third column: pathological type 2 voice; fourth column: pathological type 3 voice; fifth column: pathological type 4 voice.

calculated in the presence of high levels of noise should be taken carefully. For this reason, an estimate of the noise level must be also reported, allowing the readers to be aware of the limitations of the estimates. On the other hand, when the data is taken under controlled conditions, like the voices in the database analyzed herein, a high noise level can be seen as an indicator of a dominant underlying additive stochastic process.

The main disadvantage of the GCI is that it requires high values of m to obtain a convergent estimate of K_2 [10, 15]. This is because the pdf of the interpoint distance changes with the embedding dimension, whereas the GCI's kernel does not (see (1)). As a result, the convergence of the GCI to K_2 is slowed down [12]. The last statement also applies to the Grassberger and Procaccia correlation integral. As a solution, we have recently proposed the U-correlation integral (UCI) [12]:

$$U_m^\beta (h) = \int \frac{\Gamma\left(\beta/2, z/h^2\right)}{\Gamma\left(\beta/2\right)} \widehat{f}_m (\sigma; z) \, \mathrm{d}z, \tag{5}$$

where $\Gamma(a, t)$ is the upper incomplete Gamma function, $\Gamma(a)$ is the Gamma function, and $\widehat{f}_m(\sigma; z)$ is the pdf of the squared interpoint distance. There are two important aspects about this correlation integral that deserve to be mentioned. First, note that the UCI's kernel function, given by

$$g\left(h, \beta, z\right) = \frac{\Gamma\left(\beta/2, z/h^2\right)}{\Gamma\left(\beta/2\right)}, \tag{6}$$

has a parameter β that is used to incorporate information about the embedding dimension. In other words, this kernel function is able to change according to m. Second, we are now working with the square of the interpoint distance, that is, $z = \widetilde{z}^2$, to reduce the computational cost.

The scaling law for the UCI is [12]

$$U_m^\beta (h) = \frac{\widehat{\phi}}{2} (2\sigma)^D \frac{\Gamma(D/2)\,\Gamma\left((\beta + m)/2\right)}{\Gamma\left(\beta/2\right)\Gamma\left(m/2 + 1\right)}$$

$$\cdot e^{-m\tau K_2} \left(\frac{h^2}{4\sigma^2}\right)^{m/2} \tag{7}$$

$$\cdot F\left(\frac{\beta + m}{2}, \frac{m - D}{2}; \frac{m + 2}{2}; -\frac{h^2}{4\sigma^2}\right),$$

where $\widehat{\phi}$ is a normalization constant and $F(a, b; c; t)$ is the Gauss hypergeometric function. The UCI is approximated with the U-correlation sum $\widehat{U}_m^\beta(h)$ which, in this article, is calculated as follows: first, we obtain the squared distances between each pair of m-dimensional delay vectors $z_\omega = \|\mathbf{x_i} - \mathbf{x_j}\|^2$, where $\omega = \{(i, j) \ / \ i = 1, 2, \ldots L, \ j = 1, 2, \ldots, L, \ i \neq j\}$, and then

$$\widehat{U}_m^\beta (h) = \frac{1}{\Omega} \sum_{\omega=1}^{\Omega} g\left(h, \beta, z_\omega\right) = \frac{1}{\Omega} \sum_{\omega=1}^{\Omega} \frac{\Gamma\left(\beta/2, z_\omega/h^2\right)}{\Gamma\left(\beta/2\right)}, \tag{8}$$

where $\Omega = L(L-1)$. The evaluation of the function $g(h, \beta, z_\omega)$ requires numerical integration which is computationally expensive. Instead, we performed the evaluation through spline interpolation of the function $\widehat{g}(t) = \Gamma(\beta/2, t)/\Gamma(\beta/2)$, where $t \in [10^{-6}, 10^2]$ for each value of β.

Once the correlation sum is calculated, we estimated D, K_2, and σ using coarse-grained estimators. These are explicit expressions for D, K_2, and σ as functions of m and h [13]. Such functions are useful because they allow us to estimate the invariants and to visually confirm a scaling regime. In [16], we have presented three coarse-grained functions: $D_m^U(h)$, $K_m^U(h)$, and $\sigma_m^U(h)$ which are based on the UCI.

TABLE 1: Analyzed subjects: number and age.

Condition	Number		Age (years) mean ± SD	
	Male	Female	Male	Female
Healthy	21	32	38.8 ± 8.5	34.1 ± 7.8
Pathologic	67	96	41.2 ± 9.2	37.6 ± 8.3

As an advantage, they can be calculated in an automated manner from two U-correlation integrals. Moreover, we have proposed an automated algorithm to estimate D, K_2, and σ from these coarse-grained estimators [16].

2.1. Coarse-Grained Estimators Based on UCI.
Here, we present the coarse-grained estimators used in this study. A more detailed description can be found in [12, 16].

We define the noise level functional as [12]

$$\Delta_m^U(h) = \frac{1}{2}\left[\dot{D}_{m+2}^{\beta=m}(h) - \dot{D}_m^{\beta=m}(h)\right] \approx \frac{4\sigma^2}{h^2 + 4\sigma^2} \tag{9}$$

$$\text{for } m \gg D,$$

where

$$\dot{D}_m^\beta(h) = \frac{d \ln U_m^\beta(h)}{d \ln h}. \tag{10}$$

This functional is the difference between the logarithmic derivatives of two U-correlation integrals: $U_m^{\beta=m}(h)$, which is the UCI calculated with squared distances z_ω coming from m-dimensional delay vectors and a kernel with the parameter $\beta = m$, and $U_{m+2}^{\beta=m}(h)$. The last one is the correlation integral obtained from distances between pairs of $(m+2)$-dimensional delay vectors and a kernel with the parameter $\beta = m$.

We have shown in [12] that $\Delta_m^U(h)$ is a function that decreases monotonically from 1 to 0 on a scale proportional to the value of σ. From this noise functional, we can define the noise level coarse-grained estimator as [12]

$$\sigma_m^U(h) = \frac{h}{2}\sqrt{\frac{\Delta_m^U(h)}{1 - \Delta_m^U(h)}}. \tag{11}$$

The coarse-grained estimators for D and K_2 can be read as [16]

$$D_m^U(h) = \dot{D}_m^{\beta=m}(h) + \frac{\Delta_m^U(h)}{1 - \Delta_m^U(h)}\left[\dot{D}_m^{\beta=m}(h) + 2(m \right.$$

$$\left. - 1)\left(\frac{U_m^{\beta=m-2}(h)}{U_m^{\beta=m}(h)} - 1\right)\right], \tag{12}$$

$$K_m^U(h) = -\frac{1}{2\tau}\left\{\ln\left(\frac{U_{m+2}^{\beta=m+2}(h)}{U_m^{\beta=m}(h)}\right)\right.$$

$$+ \ln\left(\frac{D_m^U(h)}{m} + 1\right)\right\} + \frac{1}{2\tau}$$

$$\cdot \ln\left[\Delta_m^U(h)\left(\frac{m - \dot{D}_m^{\beta=m}(h)}{m - D_m^U(h)}\right)\right.$$

$$\left. + \left(1 - \Delta_m^U(h)\right)\left(\frac{\dot{D}_m^{\beta=m}(h)}{m} + 1\right)\right], \tag{13}$$

respectively.

The calculation of $D_m^U(h)$ requires the estimation of $\Delta_m^U(h)$ and two correlation integrals: $U_m^{\beta=m}(h)$ and $U_m^{\beta=m-2}(h)$. On the other hand, $K_m^U(h)$ depends on the correlation integrals $U_{m+2}^{\beta=m+2}(h)$ and $U_m^{\beta=m}(h)$, $\Delta_m^U(h)$, the logarithmic derivative $\dot{D}_m^{\beta=m}(h)$, and the coarse-grained estimator $D_m^U(h)$.

2.2. Database.
This study was conducted using the *Massachusetts Eye & Ear Infirmary* (MEEI) Voice Disorders Database, distributed by *Kay Elemetrics* [17]. We have chosen a subset of 219 records (53 normal and 166 pathological) of sustained phonation of vowel /a/ [18]. Each record was down-sampled to 10 kHz and the subjects' descriptive statistics are presented in Table 1. From this set of voices, we excluded 3 records because they contained less than 8000 sample points, ending with a total of 216 records to be analyzed.

For each record, a centered window of $N = 8000$ data points was selected and normalized to have zero mean and unitary standard deviation. Then, the U-correlation sums were computed for $m = \{4, 6, \ldots, 20\}$, $\tau = 10$, and $h \in [e^{-5}, e^2]$. Moreover, the nearest 10 temporal neighbors of each delay vector were discarded [7]. Finally, Algorithm 1 was applied. We must clarify that the range of values for the embedding dimension was selected taking into account the studies conducted in [6, 19–21], where a low-dimensional vocal system is suggested ($D < 5$ for pathological voices). On the other hand, the embedding lag was selected as the average lag (over all voices) where the first minimum of the mutual information function occurs [7, 10].

In a preliminary study, the pathological records were classified into Titze's scheme based on the visual observation of the time series and its spectrograms. The classification resulted in 74 type 1 records and 81 type 2 records. The remainder 11 voices where taken apart since these techniques cannot differentiate between type 3 and type 4 voices.

(1) Calculate $U_m^{\beta=m}(h)$ and $U_m^{\beta=m-2}(h)$ using Eq. (8) for $m > 2$.

(2) Calculate $\Delta_m^U(h)$ using Eq. (9) and the UCIs obtained in step (1).

(3) Compute $\sigma_m^U(h)$ with Eq. (11) and obtain the function $F_\sigma(h)$. Estimate σ within a range of h centered at the value h where $F_\sigma(h)$ is minimum.

(4) Use $U_m^{\beta=m}(h)$, $U_m^{\beta=m-2}(h)$ and $\Delta_m^U(h)$ to calculate $D_m^U(h)$ (Eq. (12)) and obtain the function $F_D(h)$. Estimate D within a range of h centered at the value h where $F_D(h)$ is minimum.

(5) Obtain $K_m^U(h)$ (Eq. (13)) using $U_m^{\beta=m}(h)$, $\Delta_m^U(h)$ and $D_m^U(h)$. Calculate the function $F_{K_2}(h)$. Estimate K_2 within a range of h centered at the value h where $F_{K_2}(h)$ is minimum.

ALGORITHM 1: Automated estimation of attractors' invariants.

Following the definition of type 4 voices given by Sprecher et al. in [2], we calculated the coarse-grained estimators $D_m^U(h)$ and $K_m^U(h)$ for these 11 voices. Then, we grouped together the voices that did show a scaling regime (finite correlation dimension and entropy) as type 3 and those that did not as type 4.

3. Results

The first result of the simulations can be observed in Figure 2. It shows the coarse-grained estimators $\sigma_m^U(h)$, $D_m^U(h)$, and $K_m^U(h)$ for a normal voice (Figures 2(a), 2(f), and 2(k)), a pathological type 1 voice (Figures 2(b), 2(g), and 2(l)), a pathological type 2 voice (Figures 2(c), 2(h), and 2(m)), a pathological type 3 voice (Figures 2(d), 2(i), and 2(n)), and a pathological type 4 voice (Figures 2(e), 2(j), and 2(o)).

It is important to observe that all coarse-grained estimators present a scaling region, except the ones for correlation dimension and K_2 entropy of type 4 voices. This highlights the suitability of these estimators to analyze normal and pathological (type 1, type 2, and type 3) voices.

As it can be observed in the first row of Figure 2, the noise level is greater for the pathological voices than for the normal voice and, in the pathological case, it increases along with the type. This means that, for both normal and pathological voices, there is an underlying stochastic component and its level increases in the presence of pathology.

The behavior of the estimator $D_m^U(h)$ is presented in the second row of Figure 2. As it can be seen, this estimator suggests a value of $D \approx 1.25$ for normal voices and slightly greater values for pathological type 1 and type 2 voices. However, it is difficult to say whether there is any difference between the analyzed normal and pathological voice. Nevertheless, these results suggest that the voice production system has a relatively low dimension.

Regarding the estimator $K_m^U(h)$, it can be observed from the third row of Figure 2 that, for the voices analyzed herein, the estimator converges to values close to zero. This suggests the presence of a strong harmonic component. Furthermore, there exists a small increase of K_2 entropy from normal voice to pathological type 1 voice and from the latter to pathological type 2 voice. This reflects an increasing degree of irregularity which is often associated with an increase in complexity.

For the type 3 voice, it can be observed in the fourth column of Figure 2 that the noise level coarse-grained estimator converges to a higher value than the one corresponding to the type 2 voice. However, it is still possible to observe a scaling range in both $D_m^U(h)$ and $K_m^U(h)$. Moreover, these estimators converge to higher values than the ones for type 2 voice. On the other hand, for the type 4 voice (fifth column of Figure 2), it is not possible to find a scaling range either in dimension or in the entropy estimator. This behavior is expected since, by definition, these types of voices have an infinite value of dimension and entropy.

As we mentioned before, this kind of nonlinear techniques should not be used with type 4 voices. Note that, from the curves of $D_m^U(h)$ and $K_m^U(h)$ (Figures 2(j) and 2(o), resp.), it is not easy to decide whether there is a scaling range. In this sense, an untrained person could erroneously determine its existence, resulting in misleading estimations.

On the other hand, a scaling range can be found for $\sigma_m^U(h)$ (Figure 2(e)), and it converges to a very high noise level value $\sigma \approx 0.5$ (SNR $\approx$ 4.77 dB). This suggests that the dynamics are mostly ruled by the underlying stochastic component. It is important to mention that this behavior is consistent in all voices that did not have a scaling region on $D_m^U(h)$ and $K_m^U(h)$ (type 4 voices). Based on this result, we suggest that a high value of σ could be a quantitative indicator of type 4 voices.

In order to find a noise level value that allows us to differentiate between type 3 and type 4 voices, we select by visual inspection all voices that did not show a scaling region on the estimators $D_m^U(h)$ and $K_m^U(h)$. Then, for each signal, we estimate the noise level value from $\sigma_m^U(h)$ and select the minimum of these values ($\sigma_{thr} = 0.4$) as a threshold.

3.1. Automated Estimation of Invariants. In order to obtain reliable estimations of D, K_2, and σ, it is essential to verify the existence of a scaling behavior over the coarse-grained estimators. Once it is found, one has to choose a range of h values from which to estimate the invariants. This choice is critical since, in practical application, the coarse-grained estimators strongly vary as a function of the scale. To avoid subjective judgements, we have proposed an algorithm (see Algorithm 1) for the automated estimation of D, K_2, and σ based on the coarse-grained estimators $D_m^U(h)$, $K_m^U(h)$, and $\sigma_m^U(h)$, respectively [16].
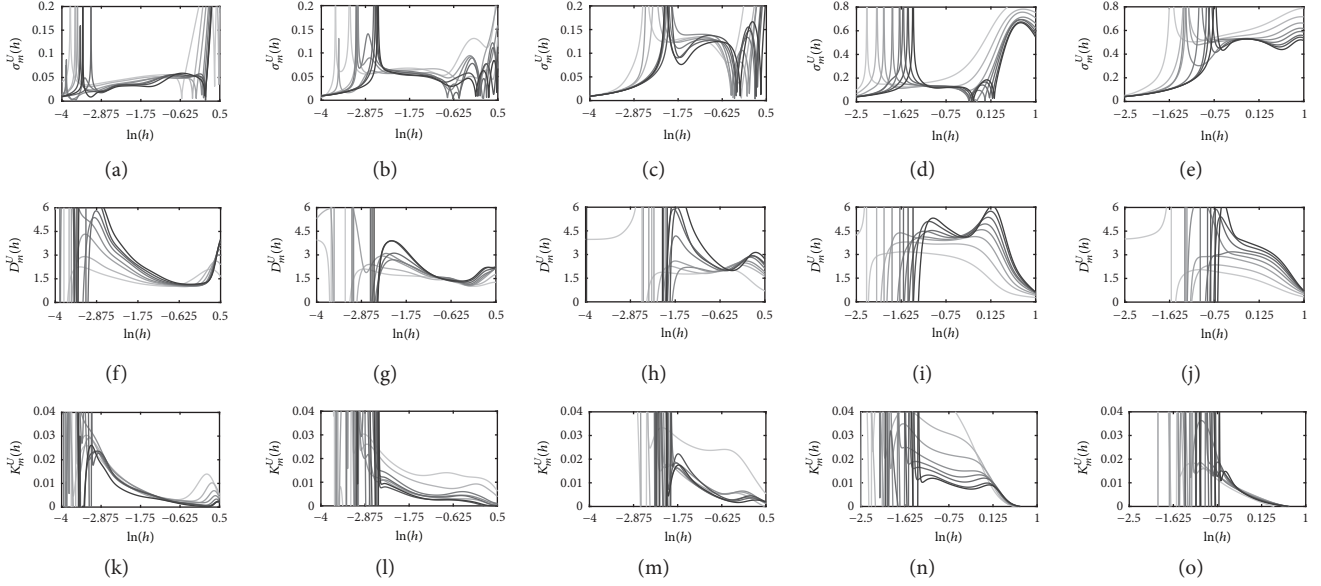
FIGURE 2: Coarse-grained estimators $\sigma_m^U(h)$, $D_m^U(h)$, and $K_m^U(h)$ for normal and pathological voices. First column: normal voice; second column: pathological type 1 voice; third column: pathological type 2 voice; fourth column: pathological type 3 voice; fifth column: pathological type 4 voice. The curves for $m = \{4, 6, \ldots, 20\}$ are color-coded in grayscale where the lightest gray corresponds to $m = 4$.

This algorithm selects the scaling range where those invariants should be estimated based on the next criteria: (i) the coarse-grained estimator must be constant for a range of scale values and (ii) the value of the invariant should converge as the embedding dimension increases.

The algorithm begins by approximating the U-correlation integrals $U_m^{\beta=m}(h)$ and $U_m^{\beta=m-2}(h)$ for different m values ($m > 2$) using (8). Note that $U_m^{\beta=m-2}(h)$ must be calculated for $m > 2$ since the shape parameter of the incomplete Gamma and Gamma functions must be greater than zero. Next, the logarithmic derivatives $\dot{D}_m^{\beta=m}(h)$ and $\dot{D}_{m+2}^{\beta=m}(h)$ must be computed in order to obtain $\Delta_m^U(h)$ (see (9)). Observe that the correlation integral $U_{m+2}^{\beta=m}(h)$ is equal to the correlation integral $U_{\widehat{m}}^{\beta=\widehat{m}-2}(h)$ evaluated at $\widehat{m} = m + 2$. In this article, the logarithmic derivatives were obtained using a wavelet transform approach [22]. This produces a smooth version of the derivatives allowing us to better estimate the invariants.

The noise level must be calculated from a range of h where the coarse-grained estimator $\sigma_m^U(h)$ is nearly constant and its variation across the m values is the smallest. In this sense, for the noise level, we define the functions [16]

$$A_\sigma(h) = \frac{1}{M} \sum_{i=1}^{M} \frac{d\sigma_{m_i}^U(h)}{d \ln h},$$

$$V_\sigma(h) = \frac{1}{M-1} \sum_{i=1}^{M} \left(\sigma_{m_i}^U(h) - \widehat{\sigma}^U(h) \right)^2, \quad (14)$$

$$F_\sigma(h) = A_\sigma(h) V_\sigma(h),$$

where $m \in \{m_1, m_2, \ldots, m_i, \ldots, m_M\}$ and $\widehat{\sigma}^U(h)$ is the average of $\sigma_m^U(h)$ across m. $A_\sigma(h)$ is the average over m of the derivative of $\sigma_m^U(h)$ with respect to $\ln h$, $V_\sigma(h)$ gives the variation of $\sigma_m^U(h)$ across m, and $F_\sigma(h)$ is the product of the two aforementioned functions. We propose to estimate σ within a range of h centered at the h value at which $F_\sigma(h)$ is minimum. This way, σ is estimated in a range of h centered in a plateaued region (scaling region) of $\sigma_m^U(h)$, and its value is consistent through the parameter m.

The correlation dimension and the correlation entropy can be determined using the coarse-grained estimators $D_m^U(h)$ (see (12)) and $K_m^U(h)$ (see (13)), respectively. To find a range of h values to estimate D and K_2, we use a similar approach, but from functions $F_D(h)$ and $F_{K_2}(h)$, respectively. The functions $F_D(h)$ and $F_{K_2}(h)$ can be computed similarly to $F_\sigma(h)$ but using the coarse-grained estimators $D_m^U(h)$ and $K_m^U(h)$, respectively.

The estimation of each invariant for the whole group of signals is presented in Figure 3. Figure 3(a) shows a box plot of the noise level estimation for normal (N), pathological type 1 (PT1), pathological type 2 (PT2), pathological type 3 (PT3), and pathological type 4 (PT4) voices. As it can be observed, σ is greater for pathological voices than for normal ones. However, it is not possible to establish a statistical difference between normal and pathological type 1 and type 2 voices. On the other hand, we can differentiate normal from pathological type 3 and type 4 voices. It is very important to say that the group labeled as TP4 was selected because they have a noise level value $\sigma > 0.4$. Note that, using this threshold, we can separate type 3 from type 4 voices.

The box plot of the correlation dimension is presented in Figure 3(b). There is an increase of D from normal to
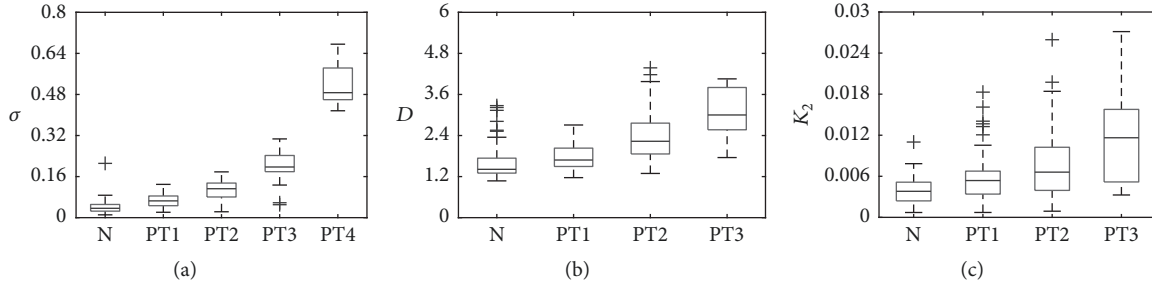
FIGURE 3: Invariant estimation box plot: (a) noise level, (b) correlation dimension, and (c) correlation entropy. Normal (N), pathological type 1 (PT1), pathological type 2 (PT2), pathological type 3 (PT3), and pathological type 4 (PT4) voices.

pathological voices. For normal voices, the median is $D = 1.4$; for pathological type 1 voices, $D = 1.68$; for type 2 voices, $D = 2.23$; and for type 3 voices, $D = 2.99$. From these results, we can establish that the vocal system has a low dimension, even in the presence of a pathology. The group TP4 is not shown in this plot since, by definition, its dimension is infinite. Regarding the correlation entropy, it is shown in Figure 3(c) that there is a small increase of K_2 from normal to pathological voices. However, it is not possible to find a statistical difference between normal, pathological type 1, and pathological type 2 voices. The median correlation entropy estimated for normal voices is $K_2 = 0.0038$, for pathological type 1 voices is $K_2 = 0.0053$, for pathological type 2 voices is $K_2 = 0.0068$, and for pathological type 3 voices is $K_2 = 0.012$.

4. Discussion

Bearing in mind that the used records were taken in a controlled environment, we could associate the noise level with the strength of an additive stochastic component that coexists with the dynamic producing the voice. There is a tendency of σ to increase its value from normal voices to pathological ones. This can be seen as an increase of the power of the stochastic component caused by the presence of a pathology. Moreover, in pathological voices, σ increases its value along with the voice type.

The definition of type 4 voices given by Sprecher et al. [2] states that these signals are characterized by pure stochastic oscillations; therefore, their dimension is infinite [10]. In practice, it is not possible to measure an infinite correlation dimension value since it is bounded by the embedding dimension used to calculate the correlation integral. Instead, an infinite correlation dimension is inferred if there is not a scaling regime in the coarse-grained estimator of D. In previous studies, this was done through visual inspection, which is always a subjective judgement [2, 19].

One interesting aspect of the coarse-grained estimator $\sigma_m^U(h)$ is that, in the type 4 signals analyzed here, it always presents a scaling range from which to estimate σ, although $D_m^U(h)$ and $K_m^U(h)$ have no scaling regions. These results led us to think that the noise level can be used as an objective measure to discriminate between type 3 and type 4 voices.

The threshold proposed here ($\sigma_{\text{thr}} = 0.4$) was set using the voice signals that did not present a scaling regime in $D_m^U(h)$ and $K_m^U(h)$. With this threshold, we were able to well separate type 3 and type 4 voices. Nevertheless, we are aware that this threshold was selected based on observations from this database. This finding must be validated with a more extensive study involving a larger number of records and voice care professionals.

Our estimations of correlation dimension are in concordance with other studies [6, 19–21]. In [23], Choi et al. conducted a very similar simulation over the Kay Elemetrics database. They reported a mean value of $D = 1.57$ for normal voices and an increased dimension value for pathological ones. Moreover, they obtained a decreasing SNR (it was calculated according to [24]) with the type of the voice, being the lowest for type 3 voices. However, they did not analyze type 4 voices. Another study by Zhang and Jiang conducted over the Kay Elemetrics database reported a mean correlation dimension of $D = 1.51$ for normal voices and $D = 3.17$ for a group with vocal tremor [25]. As far as we know, all researches conducted on normal and pathological voices have used the Grassberger and Procaccia correlation integral to estimate D and K_2. This methodology has the disadvantage that its estimations are sensitive to noise presence, requires large m values to converge, and does not give an estimation of the noise level [10, 13]. However, other variants like the Gaussian correlation integral have not been used.

Regarding the correlation entropy, our results suggest that pathological type 1 and type 2 voices have a slightly greater value than normal ones. Furthermore, these three types of voices have a K_2 value close to zero, suggesting nearly periodic dynamics. For type 3 voices, the values of K_2 are the greatest, meaning more irregular and unpredictable dynamics. These values are comparable with those presented by Yan et al. [26]. They reported an estimation of K_2 for normal (mean value $K_2 = 0.014$) and esophageal phonation (mean value $K_2 = 0.023$) subjects. In [27], Calawerts et al. calculated the largest Lyapunov exponent for type 1, type 2, and type 3 voices using the Kay Elemetrics database. Their results suggest an increase of the value of the largest Lyapunov exponent along with the type of the voice. Our results are in concordance with this study since we obtain an estimation of correlation entropy that not only is greater for pathological

voices than for normal ones but also increases with the type of the voice.

5. Conclusions

In this article, we have studied normal and pathological voices through the correlation dimension, the correlation entropy, and the noise level. These invariants were estimated using an automated algorithm based on coarse-grained estimators derived from the U-correlation integral. The results suggest that the voice production dynamical system has a low dimension. The value of D is greater for pathological voices than for normal ones. Moreover, its value also increases along with the type of the voice. Regarding the correlation entropy, its value is very low for normal and for type 1 and type 2 pathological voices. Although a more extensive study is still needed, this finding suggests that the system dynamics have a harmonic oscillatory behavior. On the other hand, pathological type 3 voices present higher values of K_2, implying a more complex behavior which is reflected in a more irregular dynamic. The noise level can be interpreted as the power of a stochastic perturbation intrinsic to the voice production system. Our results show that σ is greater for pathological voices than for normal ones. Furthermore, it increases along with the type of voice, being the highest for type 4 voices. This means that the presence of the stochastic component is stronger in pathological voices. Based on these results, in this work, we have proposed a quantitative criterion that can be used to differentiate between type 3 and type 4 voices. We are aware of the limitation of the reduced samples of type 3 and type 4 voices. In this sense, these preliminary results will be validated with a more extensive study involving a larger number of records and voice care professionals.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] I. R. Titze, *Workshop on Acoustic Voice Analysis: Summary Statement*, vol. 24, National Center for Voice and Speech, 1995.

[2] A. Sprecher, A. Olszewski, J. J. Jiang, and Y. Zhang, "Updating signal typing in voice: Addition of type 4 signals," *The Journal of the Acoustical Society of America*, vol. 127, no. 6, pp. 3710–3716, 2010.

[3] I. R. Titze, R. J. Baken, H. Herzel, and P. Arts, "Evidence of chaos in vocal fold vibration," in *Vocal fold physiology: new frontiers in basic science*, vol. 1, pp. 143–188, 1993.

[4] H. Herzel, D. Berry, I. R. Titze, and M. Saleh, "Analysis of vocal disorders with methods from nonlinear dynamics," *Journal of Speech, Language, and Hearing Research*, vol. 37, no. 5, pp. 1008–1019, 1994.

[5] A. Giovanni, M. Ouaknine, B. Guelfucci, P. Yu, M. Zanaret, and J.-M. Triglia, "Nonlinear behavior of vocal fold vibration: The role of coupling between the vocal folds," *Journal of Voice*, vol. 13, no. 4, pp. 465–476, 1999.

[6] J. J. Jiang, Y. Zhang, and C. McGilligan, "Chaos in voice, from modeling to measurement," *Journal of Voice*, vol. 20, no. 1, pp. 2–17, 2006.

[7] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, 2nd edition, 2004.

[8] J.-P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Reviews of Modern Physics*, vol. 57, no. 3, part 1, pp. 617–656, 1985.

[9] P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Physical Review Letters*, vol. 50, no. 5, pp. 346–349, 1983.

[10] M. Small, *Applied nonlinear time series analysis: applications in physics, physiology and finance, World Scientific*, vol. 52, 2005.

[11] C. Diks, "Estimating invariants of noisy attractors," *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 53, no. 5, pp. R4263–R4266, 1996.

[12] J. F. Restrepo and G. Schlotthauer, "Noise-assisted estimation of attractor invariants," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 94, no. 1, Article ID 012212, 2016.

[13] C. Diks, *Nonlinear time series analysis*, vol. 4 of *Nonlinear Time Series and Chaos*, World Scientific Publishing Co., Inc., River Edge, NJ, 1999.

[14] D. Yu, M. Small, R. G. Harrison, and C. Diks, "Efficient implementation of the Gaussian kernel algorithm in estimating invariants and noise level from noisy time series data," *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 61, no. 4, pp. 3750–3756, 2000.

[15] G. Nolte, A. Ziehe, and K.-R. Müller, "Noise robust estimates of correlation dimension and k2 entropy," *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 64, no. 1, p. 10, 2001.

[16] J. F. Restrepo and G. Schlotthauer, "Automatic estimation of attractor invariants," *Nonlinear Dynamics*, vol. 91, no. 3, pp. 1681–1696, 2018.

[17] M. Eye and E. Infirmary, *Voice disorders database, version. 1.03 (cd-rom)*, Kay Elemetrics Corp, Lincoln Park, NJ.

[18] V. Parsa and D. G. Jamieson, "Identification of pathological voices using glottal noise measures," *Journal of Speech, Language, and Hearing Research*, vol. 43, no. 2, pp. 469–485, 2000.

[19] A. Behrman and R. J. Baken, "Correlation dimension of electroglottographic data from healthy and pathologic subjects," *The Journal of the Acoustical Society of America*, vol. 102, no. 4, pp. 2371–2379, 1997.

[20] G. Vaziri, F. Almasganj, and R. Behroozmand, "Pathological assessment of patients' speech signals using nonlinear dynamical analysis," *Computers in Biology and Medicine*, vol. 40, no. 1, pp. 54–63, 2010.

[21] N. Yan, M. L. Ng, D. Wang, L. Zhang, V. Chan, and R. S. Ho, "Nonlinear dynamical analysis of laryngeal, esophageal, and tracheoesophageal speech of cantonese," *Journal of Voice*, vol. 27, no. 1, pp. 101–110, 2013.

[22] J. Luo, J. Bai, and J. Shao, "Application of the wavelet transforms on axial strain calculation in ultrasound elastography," *Progress*

in Natural Science: Materials International, vol. 16, no. 9, pp. 942–947, 2006.

[23] S. H. Choi, Y. Zhang, J. J. Jiang, D. M. Bless, and N. V. Welham, "Nonlinear dynamic-based analysis of severe dysphonia in patients with vocal fold scar and sulcus vocalis," *Journal of Voice*, vol. 26, no. 5, pp. 566–576, 2012.

[24] P. Milenkovic, "Least mean square measures of voice perturbation," *Journal of Speech, Language, and Hearing Research*, vol. 30, no. 4, pp. 529–538, 1987.

[25] Y. Zhang and J. J. Jiang, "Nonlinear dynamic mechanism of vocal tremor from voice analysis and model simulations," *Journal of Sound and Vibration*, vol. 316, no. 1-5, pp. 248–262, 2008.

[26] N. Yan, M. L. Ng, D. Wang, V. Chan, and L. Zhang, "Nonlinear dynamics of voices in esophageal phonation," in *Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2011*, pp. 2732–2735, USA, September 2011.

[27] W. M. Calawerts, L. Lin, J. C. Sprott, and J. J. Jiang, "Using Rate of Divergence as an Objective Measure to Differentiate between Voice Signal Types Based on the Amount of Disorder in the Signal," *Journal of Voice*, vol. 31, no. 1, pp. 16–23, 2017.