Nonlinear Signal Processing for Voice Disorder Detection by Using Modified GP Algorithm and Surrogate Data Analysis

Aboozar Taherkhani, Seyyed Ali Seyyedsalehi, Arash Mohammadi, Mohammad Hasan Moradi
Faculty of Biomedical Engineering
Amirkabir University of Technology
Tehran, Iran
Email: a.taherkhani@aut.ac.ir, ssalehi@aut.ac.ir, a.mohammadi@aut.ac.ir, mhmoradi@aut.ac.ir

Abstract—Acoustic voice analysis is an effective, cheap and non-invasive tool that can be used to confirm the initial diagnosis and provides an objective determination of the impairment. The nonlinearity of the voice source mechanisms may cause the existence of chaos in human voice production. Voice pathology can cause to additional colored noise to voice wave. Added noise to a chaotic signal causes reduction of the deterministic property and therefore increases correlation dimension of signal. Surrogate data analysis can measure this deviation and give a criterion for an amount of noise added to the chaotic signal. By using this criterion a threshold level is set to separate disordered voice from normal voice and 95% accuracy is achieved.

Introduction

Nowadays voice diseases are increasing dramatically, due mainly to unhealthy social habits and voice abuse. It is well known that voice diseases affect the quality of the voice register. These diseases should be diagnosed and treated at an early stage. Acoustic voice analysis is an effective and non-invasive tool that can be used to confirm the initial diagnosis and provides an objective determination of the impairment. Early detection and treatment of laryngeal tumors can reduce both morbidity and mortality.[1] Invasive methods need massive equipment and waste lots of money and time whereas in the Acoustic methods, voice can be recorded for several times without taking much time, imposing much cost and pain. Invasive methods give effective and necessary information about disease, Therefore acoustic methods can be used to primary detection of disease, the patient is sent to specialized centers to do invasive methods if it is necessary.

According to the benefit of the acoustic voice analysis, improvement of these methods is recently considered. Roland et al. presented an acoustical feature extraction paradigm that focused on jitter, shimmer, standard deviation of fundamental frequency, and the glottal-to-noise excitation ratio was used to analyses 120 voice samples. An improved artificial neural network (ANN) was used for classification. 80% of all voice samples could be classified correctly as either healthy or hoarse [2].

Chaos has been observed in turbulence, chemical reactions, nonlinear circuits, the solar system, biological populations, and seems to be an essential aspect of most physical systems. Methods based on nonlinear dynamics, including general dimension (Hausdorff dimension, information dimension, correlation dimension, etc.), entropy (Kolmogrov entropy, second-order entropy, etc.), and Lyapunov exponents, enable us to quantitatively describe chaotic behavior. Investigations of chaotic activities in physiologic systems suggest that changes in nonlinear dynamic measures may indicate states of pathophysiological dysfunction [3]. Poon and Merrill, [4] for example, found that chaotic activity decreased in electrocardiogram (ECG) signals from patients with congestive heart failure. Hornero et al [5] found that the electroencephalogram (EEG) signals generated by schizophrenic patients had a significantly lower correlation dimension than the EEG signals of normal subjects. These examples suggest that chaos theory and nonlinear dynamic methods might potentially be applied to diagnose physiological disorders and evaluate the effects of clinical treatments.

Over the last two decades, observations in computer models of the vocal folds and nonlinear dynamic analysis of human voices have established the existence of chaos in human voice production. As noted by many researchers, the nonlinearities of the voice source mechanisms (e.g., the nonlinear pressure-flow relation in the glottis, the nonlinear stress-strain curves of vocal fold tissues, and the nonlinearities associated with vocal fold collision) make this development unsurprising. Researchers have applied these new tools to studying abnormal conditions associated with laryngeal pathologies, to differentiate normal and pathologic voices and diagnose pathologies, and to assess the effects of clinical treatments. Some traditional voice analysis methods, such as jitter and shimmer, may be unreliable for analyzing aperiodic voices. Nonlinear dynamic methods provide information complementary and nonredundant to existing analysis methods [3].

In this paper we used modified standard Grassberger–Procaccia (GP) algorithm that presented in [6] for estimating the correlation dimension of a time series related to vowel /a/. We apply the Surrogate data analysis and extract normalized mean sigma deviation (nmsd). Evidences show that nmsd is a better feature in comparison with correlation dimension. In Section II we first describe the methods and materials that contain brief explanations of database, modified GP algorithm.
and surrogate data analysis. In Section III the results of our experiences are brought. At last some discussions on our results are presented in Section IV.

II. MATERIALS AND METHODS

II. I. DATABASE

The voice samples examined in this study were selected from the Disordered Voice Database [7], model 4337, version 1.03 (Kay Elemetrics Corporation, Lincoln Park, NJ), developed by the Massachusetts Eye and Ear Infirmary Voice and Speech Lab. Subjects were asked to sustain the vowel /a/ and voice recordings were made in a soundproof booth on a DAT recorder at a sampling frequency of 44.1 kHz.

II. II. NONLINEAR ANALYSIS

Nonlinear systems can show chaotic behavior when some conditions are fulfilled. Chaos is the term used to describe what we consider to be simple, well-behaved systems. Chaotic behavior, when looked at casually, looks erratic and almost random. Chaos theories enable us to categorize and understand complex behavior that had confounded previous theories [8].

In order to quantifying chaos, we use the embedding space technique. Unfortunately, the theoretical underpinnings of the embedding technique are not fully developed at present and some “rules of thumb” are used for them [8]. Therefore it is necessary to develop these rules to achieve better results.

A. Modified GP Algorithm

The GP algorithm uses the delay embedding technique for the calculation of correlation dimension (D2). It creates an artificial space of dimension M with delay vectors constructed by splitting a discretely sampled scalar time series \( s(t) \) with:

\[
\tilde{x}_i = [s(t), s(t + \tau), \ldots, s(t + (M - 1)\tau)]
\]

(1)

Here the delay time \( \tau \) is chosen suitably such that the vectors are not correlated. The relative number of points within a distance \( R \) from a particular (i-th) data point is given by:

\[
p_i(R) = \lim_{N_v \to \infty} \frac{1}{N_v} \sum_{j=1, j\neq i}^{N_v} H(R - |\tilde{x}_i - \tilde{x}_j|)
\]

(2)

Where \( N_v \) is the total number of reconstructed vectors and \( H \) is the Heaviside step function. Averaging this quantity over randomly selected centers \( N_c \) gives the correlation function.

\[
C_M(R) = \frac{1}{N_c} \sum_{i}^{N_c} p_i(R)
\]

(3)

The correlation dimension \( D_2(M) \) is then defined to be the scaling index of the variation of \( C_M(R) \) with \( R \) as \( R \to 0 \). That is,

\[
D_2 = \lim_{R \to 0} \frac{\log C_M(R)}{\log(R)}
\]

(4)

For large values of \( R \), a significant fraction of the M-spheres used in the computation will typically go beyond the attractor region. This “edge effect” leads to underestimation of \( C_M(R) \) for large \( R \) and finally causes \( C_M(R) \) to saturate to unity. According to [6] a proper linear part in the \( \log C_M(R) \) versus \( \log(R) \) is identified which is called the “scaling region” and its slope is taken to be \( D_2 \).

The original data set, \( s_j \), is first transformed to a uniform deviate, \( s_j(t_j) \). Note that \( s_j(t_j) \) ranges from 0 to 1, which makes the volume of the embedding space unity. In order to take into account the edge effects correctly, it is convenient to redefine \( p(R) \) as the number of data points within an M-cube (instead of M-sphere) of length \( R \) around a data point. This is equivalent to replacing the Euclidean norm by the maximum norm. Operationally this is done by choosing randomly \( N_c \) data points as centers of M-cubes of length \( R \). Of these \( N_c \), M-cubes, only those which are within the bounding box of the embedded data are considered. Finally the correlation sum \( C_M(R) \) is obtained by averaging the number of data points within the M-cubes. The imposition of the requirement that an M-cube has to be within the embedding space ensures that there are no edge effects due to limited data points. However, this also means that for large values of \( R \), only a small fraction of the original \( N_c \) M-cubes are taken into consideration. Hence a maximum value of \( R \), \( R_{\text{max}} \), is fixed such that for all \( R < R_{\text{max}} \) the number of M-cubes which satisfy the above criterion is at least a hundredth of the total number of vectors, i.e. \( N_v/100 \). To avoid the region dominated by counting statistics only results from \( R > R_{\text{min}} \) are taken into consideration, where \( N_v C(R) > 10 \), which ensures that on average at least ten data points are considered per center. This makes sure that the region \( R_{\text{min}} < R < R_{\text{max}} \) is not affected by either “edge effects” or counting statistics.

\( C_M(R) \) is computed for several different values of \( R \) between \( R_{\text{max}} \) and \( R_{\text{min}} \), the logarithmic slope at each point is calculated and the average is taken to be \( D_2(M) \). The error on \( D_2(M) \) is estimated to be the mean standard deviation over this average. This error is an estimate of how well the region used by the scheme, \( R_{\text{min}} < R < R_{\text{max}} \), can be represented by a linear scaling region. A large error signifies that those values of \( R \) for which \( C_M(R) \) are not affected by counting statistics and edge effects do not represent a single scaling region. It should be noted that there often exists a critical embedding dimension \( M_{\text{cr}} \) for which \( R_{\text{min}} = R_{\text{max}} \) and no significant results can then be obtained for \( M > M_{\text{cr}} \). Thus our algorithm fixes an upper limit on \( M \) up to which calculations are to be repeated. For practical implementation of the above scheme, it is sufficient to choose \( N_v \) as \( 0.1 N_v \). The delay time \( \tau \) is chosen to be the value where the auto-correlation function
drops by 1/e. With these values, \( D_2(M) \) for \( M = 1 \) to \( M = M_{\sigma} \) is computed for a given data stream and a chi-squared fitting is undertaken using a simple analytical function.

\[
f(M) = \left( \frac{D_{2\text{sat}}}{M_d - 1} \right)^{-1} (M - 1) + 1 \quad \text{for } M < M_d \]
\[
= D_{2\text{sat}} \quad \text{for } M \geq M_d.
\]

The best fit value of \( D_{2\text{sat}} \) (obtained by minimizing \( \chi^2 \)) is taken to be the saturated correlation dimension with errors corresponding to \( \Delta \chi^2 = 1 \). Considering the uncertainties in the computation and statistics of the errors in \( D_2(M) \), a more sophisticated fitting procedure is perhaps not warranted. A best fit value of \( D_2 \approx M_{\sigma} \) implies that no saturation of \( D_2(M) \) was detected.

In summary, the algorithmic scheme first converts a data stream to a uniform deviate. Next, the autocorrelation function is evaluated to estimate the time delay \( \tau \). For each \( M \), \( c_M(R) \) is evaluated using \( N_c = 0.1N_v \) randomly chosen centers. The limits \( R_{\min} \) and \( R_{\max} \) are estimated and \( D_2(M) \) is computed for the region from \( R_{\min} \) to \( R_{\max} \). The process is repeated for consecutive values of \( M \) until \( R_{\max} = R_{\min} \). The resultant \( D_2(M) \) curve is fitted using function (5) which returns the saturated correlation dimension \( D_{2\text{sat}} \) with an error estimate [6].

**B. Surrogate data Analysis**

Phase has an important role in the chaotic signals. Chaotic regularity (determinant property of chaos) of a signal is related to the chaotic regularity of both phase and amplitude. This determination is highly related to the phase as a little distortion in phase disturbed chaotic regularity.

Surrogate data analysis uses this property of chaotic signal to capture a criterion for amount of deterministic property of signal. Surrogate data generated by taking the Fourier transform of original data, randomizing the phases (without changing the amplitudes), and then regenerating the surrogate data with an inverse Fourier transform. Produced data has been removed any deterministic evolution [8]. Therefore the correlation dimension of surrogate data is different from original data. By measuring this difference we can achieve a criterion for deterministic property of a chaotic signal.

Added noise to a chaotic signal causes reduction of the deterministic property and therefore increase correlation dimension of signal. Surrogate data analysis can measure this difference and give a criterion for amount of noise added to the chaotic signal. The addition of noise to the chaotic system is found to decrease the difference between the \( D_2(M) \) of the data and the surrogates. The effects of the addition of different percentage of white and red noise on surrogate analysis of data from the Rossler system are shown in Fig. 1 [6]. The upper panels (a) and (b) are for white noise contamination at 20\% and 50\% respectively. The \( D_2(M) \) values for the data are represented by filled circles and connected by solid lines, while for the corresponding ten surrogates, the curves are represented by dashed lines. The lower panels (c) and (d) are for red noise contamination at 20\% and 50\% respectively. The upper panels (a) and (b) are for white noise contamination at 20\% and 50\% respectively. The lower panels (c) and (d) are for red noise contamination at 20\% and 50\% respectively [6].
IV. RESULTS

\( C_M(R) \) of vowel \(/a/\) is calculated by GP algorithm for various amount of R and M. In Fig. 2, the diagram of \( \log(C_M) \) versus \( \log(R) \) for various M is shown. Location of \( R_{min} \) and \( R_{Max} \) are determined by “*” where each point between them is approximated by a line. The slope of these lines is correlation dimension of signal for corresponding M. In Fig. 3, diagrams of \( D_2(M) \) according to a healthy signal (black curve) with its three surrogate data (blue green reed) are shown. There is much difference between these two groups of curve; in other word these two groups are easily distinguishable. They have large nmsd (14.635). In Fig. 4, diagrams of \( D_2(M) \) according to a disordered voice (black curve) with its three surrogate data (blue green reed) are shown. These two groups are close to each others, therefore nmsd of them is small (2.041). Whereas the variation range of correlation dimension is 1.6-3.4, the variation range of nmsd is 2.041-14.635, Therefore nmsd is a better feature for classification of pathological and healthy voice in comparison with correlation dimension.

![Fig. 2. The diagram of \( \log(C_M) \) versus \( \log(R) \) for various M is shown. Location of \( R_{min} \) and \( R_{Max} \) are determined by “*” where each point between them is approximated by a line. The slope of these lines is correlation dimension of signal for corresponding M.](image)

![Fig. 3. Diagrams of \( D_2(M) \) according to a healthy signal (black curve) with its three surrogate data (blue green reed) are shown. There is much difference between these two groups of curve.](image)

![Fig. 4. Diagrams of \( D_2(M) \) according to a disordered voice (black curve) with its three surrogate data (blue green reed) are shown. These two groups are close to each others, therefore nmsd of them is small (2.041).](image)

The results of the analysis on 16 disordered voices and 5 healthy voices are shown in Table I and Table II respectively. \( D_2^{sur} \) and \( D_2^{*sur}(M) \) are correlation dimension of original signal and average of surrogate data’s correlation dimension respectively. \( M^{*sur} \) (Saturated dimension) is the dimension in which the correlation dimension \( (D_2(M)) \) of original signal saturates. \( M^{*sur} \) is the average of saturated dimension of surrogate data. nmsd is normalized mean sigma deviation (6).

| Table I. THE RESULTS OF THE ANALYSIS ON 16 DISORDERED VOICES |
|---------------------|---------------------|---------------------|
| \( D_2^{sur} \) | \( M^{*sur} \) | \( D_2^{*sur}(M) \) | \( M^{*sur} \) | nmsd |
| 2.3321 | 5.0000 | 2.3132 | 4.0000 | 0.462 |
| 3.8410 | 8.0000 | 3.9236 | 6.0000 | 1.6699 |
| 3.6811 | 7.0000 | 3.7693 | 6.0000 | 3.3879 |
| 1.8775 | 7.0000 | 2.0795 | 6.0000 | 3.7409 |
| 1.6843 | 3.0000 | 2.4450 | 7.0000 | 5.8129 |
| 1.9120 | 9.0000 | 2.2112 | 5.0000 | 8.6273 |
| 2.2278 | 8.0000 | 2.1141 | 5.0000 | 3.1528 |
| 1.9722 | 4.0000 | 2.3785 | 4.0000 | 2.4440 |
| 1.8973 | 9.0000 | 1.7673 | 3.0000 | 3.8098 |
| 3.6300 | 8.0000 | 3.3156 | 6.0000 | 2.9489 |
| 2.4062 | 8.0000 | 2.2056 | 6.0000 | 0.9490 |
| 3.0555 | 6.0000 | 2.8827 | 7.0000 | 7.6868 |
| 0.2046 | 6.0000 | 1.2059 | 8.0000 | 2.9659 |
| 1.1502 | 5.0000 | 1.1177 | 6.0000 | 10.8247 |
| 1.6120 | 7.0000 | 1.5367 | 8.0000 | 1.6721 |
| 1.0848 | 4.0000 | 1.0568 | 4.0000 | 2.2814 |

| Table II. THE RESULTS OF THE ANALYSIS ON 5 HEALTHY VOICES |
|---------------------|---------------------|---------------------|
| \( D_2^{sur} \) | \( M^{*sur} \) | \( D_2^{*sur}(M) \) | \( M^{*sur} \) | nmsd |
| 1.1929 | 7.0000 | 2.5007 | 5.0000 | 14.6349 |
| 1.4821 | 7.0000 | 1.9621 | 4.0000 | 11.7207 |
| 1.4170 | 8.0000 | 1.9820 | 7.0000 | 19.5827 |
| 1.9292 | 4.0000 | 2.8398 | 5.0000 | 9.3704 |
| 0.7938 | 4.0000 | 3.0109 | 5.0000 | 40.6419 |
The mean values of nmsd for disordered and healthy voices are 3.902 and 19.190 respectively. If we choose nms = (3.902+19.190)/2=11.546 as differentiation threshold and use it for separation of disordered an healthy voices, all subjects are detected correctly except dashed subject. Therefore the classification accuracy is about 95%.

V. DISCUSSION

Healthy voice has lower correlation dimension in comparison with disordered voice because it has much regularity (determination property of chaos). Randomizing the phase in order to generate surrogate data causes distribution of this regularity and increase the surrogate data correlation dimension and nmsd (Fig. 3). While disordered voice has less regularity and high correlation dimension therefore has small nmsd (Fig. 4). This differentiation can be used to segregate the disordered voices from healthy voices.

The variation range of correlation dimension (0.21-3.9) is shorter than the variation range of nmsd (0.46-40.64). Therefore surrogate data analysis is powerful method for classification of disordered and healthy voices in comparison with correlation dimension.

In this paper without use of any classifier (such as neural network or support victor machine) we achieved 95% accuracy. It can be developed for classifying the kind of disorder by using a proper classifier. For example in [9] the effect of nodules and polyp on voice were investigated. Acoustic, pitch and amplitude perturbation quotients, and nonlinear dynamic measures, phase space reconstruction and correlation dimension, were used. They showed disordered voice with nodules, in comparison with polyp, has lower-dimensional phase space dynamical characteristics. Surrogate data analysis can improve their results.

ACKNOWLEDGEMENTS

We would like to thank Professor M. Hashemi Gholpayeghani for his critical discussions and valuable comments that have demonstrated in chaos course and Linguistic Laboratory of Tehran University for their useful data.

REFERENCES