PATHOLOGICAL ASSESMENT OF VOCAL FOLD NODULES AND POLYP USING ACCOUSTIC PERTURBATION AND PHASE SPACE FEATURES

Roozbeh Behroozmand, Farshad Almasganj, Mohammad Hassan Moradi

Faculty of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran
rbehroozmand@bme.aut.ac.ir , almas@aut.ac.ir , mhmoradi@aut.ac.ir

ABSTRACT

In this study, investigations are carried out from the speech waveform of the patients whose vocal folds vibration pattern has been affected by the presence of nodules and polyp. Acoustic, pitch and amplitude perturbation quotients, and nonlinear dynamic measures, phase space reconstruction and correlation dimension, are scrutinized to analyze a sustained vowel pronounced by these individuals. It is found that the cycle-to-cycle pitch frequency and amplitude variation in the case of polyp is higher than nodules. It is also demonstrated that patients' voice signals with nodules, in comparison with polyp, has lower-dimensional phase space dynamical characteristics. Classification procedure implemented using support vector machine shows that nonlinear dynamical features provide us with a more prominent description of such benign vocal fold pathologies as a valuable tool for clinical diagnosis applications.

1. INTRODUCTION

The effect of voice pathologies caused by physiological alterations of the vocal cords is a very important issue due to the unhealthy pattern of cords' vibration and the decrease in patients' speech signal quality. In addition, the detection of incipient damages to the cords helps us in improving the prognosis, treatment and care of such pathologies. The information derived from the speech signal helps in the detection of vocal fold pathologies in their early stages of emergence, and also enables voice therapists and vocal fold surgeons in preventing them from progressing or becoming malign. Therefore, non-invasive methods as less expensive and more convenient solution to the problem of vocal fold disorder diagnosis play an important role in the clinical applications.

In this paper two widespread types of benign vocal fold pathologies known as nodules and polyp are taken into consideration. These laryngeal disorders are defined using specific entities by laryngologists and voice pathologists, based on their anatomic location and gross appearance [1]. They are both formed by inflammation caused by stress, trauma or irritation. A nodule is defined as a small lesion occurring on both sides of the vocal folds, strictly symmetric on their border of the anterior and middle third, and usually immobile during phonation. A polyp is defined as a lesion on the anterior third of the folds. It may be sessile or pedunculated and, if pedunculated, very mobile. Ottolaryngologists and voice pathologists agree that nodules and polyp are distinguishable based on their location and size [1]. A biopsy larger than 0.3 cm could be a polyp and a biopsy less than 0.3 cm could be a nodule. It is also stated that there is no definitive histological distinction between them. In most of the cases, these lesions are present exactly at the hourglass-like gap between both folds.

The harmful effect of such vocal fold disorders causes the glottis fissure to remain continuously open so that the patients are confronted by resulted difficulties in breathing, coughing, and even speaking. During the closing phase of folds' vibration, the presence of nodules and polyp on the outer layer of vocal folds' tissue inhibits them from being completely folded on each other. This effect would be evident in the glottal waveform and the audible quality of patients' speech signal. Because the vibration pattern of vocal folds, excited by the air flow running through the glottis, is an important indicator of laryngeal function, any abnormality of the larynx will be evident by tracking the speech signal characteristic variations which often assume the aspect of noise. Incomplete closure of the vocal folds, glottal air leakage and their asymmetrical vibration, due to their biomechanical parameter alterations, are responsible for pitch frequency and air flow volume changes, amplitude and mucosal wave reduction and the noise-like turbulence of airflow in vicinity of nodules and polyp [2].

In this study, the effect of vocal fold nodules and polyp on the acoustic perturbation and nonlinear dynamical feature variations in patients' speech signal is investigated.

2. MATERIALS AND METHODS

2.1. Database

The voice samples examined in this study were selected from the Disordered Voice Database [3], model 4337, version 1.03 (Kay Elemetrics Corporation, Lincoln Park, NJ), developed by the Massachusetts Eye and Ear Infirmary Voice and Speech Lab. This database includes 19 samples of vocal fold nodules and 20 samples from patients with vocal fold polyp. 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.

2.2. Acoustic Perturbation Analysis

The acoustic perturbation parameters for estimating the property variation of pathological voices are measured during phonation of sustained vowels. These parameters define the degree of cycle-to-cycle instability of amplitude and pitch, and indicate the level of aperiodic components, predicted by the presence of turbulent noise, frequency and amplitude modulations of voice signal due to the alterations in biomechanical properties of the vocal folds. Measures of acoustic perturbation including pitch and amplitude perturbation quotients, known as PPQ and APQ, provide a quantitative assessment of vocal quality. These parameters refer to the measurement of cycle-to-cycle variation in the fundamental frequency and amplitude of a voice signal at a specific number of periods. In general the perturbation quotient of any quantity can be calculated as follows [4]:

$$PQ=\frac{100\%}{N-K}\sum_{i=\frac{K-1}{2}}^{N-\frac{K-1}{2}}\left(U_i-\frac{1}{K}\sum_{k=-\frac{K-1}{2}}^{\frac{K-1}{2}}U_{i+k}\Bigg/\frac{1}{K}\sum_{k=-\frac{K-1}{2}}^{\frac{K-1}{2}}U_{i+k}\right) \quad (1)$$

In which U can be the frequency or amplitude, N is the total number of cycles and K refers to the length of the windowed signal. As the precise calculation of PPQ and APQ measures for acoustic analysis of a voice signal in time domain strongly depends on the detection of each glottal cycle and its duration, an accurate pitch detection algorithm referred to as modified cepstrum, proposed by Mitev and Hadjitodorov [5], is used to estimate the initial and final points of each glottal cycle in the speech signals.

2.3. Phase space dynamics analysis

The descriptions of nonlinear systems are based substantially on analyses of their behaviors in their phase space [6]. Phase space is an abstract mathematical space spanned by the dynamical variables of the system. The state of the dynamical system at a given instant in time can be represented by a point in this space. The dynamical variable changes in time traces out a path in the phase space in which the aperiodic or chaotic behavior can be observed.

In signal processing concerns, a chaotic signal is defined as a signal produces by an autonomous chaotic system in response to an initial condition that leads to aperiodic behavior [7]. The typical application involves observing a signal from a nonlinear system and attempting to classify it as chaotic or non-chaotic and to determine some quantitative measure of the degree of chaos. As these goals are especially difficult to accomplish, there are some auxiliary methods which, in combination, can increase the likelihood of a correct classification and provide an approximate measure of the degree of chaos. One of the most common of these methods is the calculation of the signal's correlation dimension in its phase space. To permit calculation of this invariant measure it is necessary to construct an attractor in a space of sufficiently high dimension using the observed signal, a process known as phase space embedding. This process proceeds as follows:

Suppose x(t) is sampled at the rate f_S such that *N*-data points x[n], $0 \le n \le N-1$, are obtained. To construct an embedding in R^m, data vectors x_j, $0 \le j \le N - km$, are created as:

$$x_j = \left(x[j], x[j+k], x[j+2k], \ldots, x[j+(m-1)k]\right)^T \quad (2)$$

where *k* is a delay that is chosen through minimizing the average mutual information between *x[j]* and *x[j+k]*, [8]. The objective in the choice of lag *k* is to ensure that all *m* various coordinates of each x_j vector convey independent information. By using each x_j which defines a point in R^m a pseudo phase plot, known as phase space plot, is constructed. The choice of the correct embedding dimension, m, is also nontrivial and there is not general agreement regarding the best method. One of the most common approaches which search for the optimum embedding dimension is referred to as the method of false nearest neighbors discussed in [7].

Knowledge of the dimension of the attractor of a phase plot provides important information about a signal and the system from which it emanates. The higher the dimension the more spatially complex is the structure of the attractor. It can be inferred that the complexity degree of a signal increases with its dimension consequently [9]. If the dimension is non-integer, then the attractor has a fractal structure and the phase plot demonstrates irregular motion. If one can obtain a long data record when dealing with real data, often a sufficiently reproducible estimate of dimension can be obtained so that one may distinguish between signals whose dimensions differ by a specific threshold.

Calculation of signal's *correlation dimension* in its phase space is a common way in estimating the signal's degree of chaos. It is based on the concept of how densely the points on an attractor aggregate around one another and its calculation can be related to the relative frequency with which the attractor visits each covering element. Given a data time series that has been embedded in a space R^m, the Euclidian distance measure can be represented using *m*-dimensional embedding vectors $x_i = [x_{i1}, x_{i2}, \ldots, x_{im}]$, in which indexes are represented by $i_m = (i+m-1)kt_s$. An alternative distance measure which is much faster to calculate, is the maximum distance between corresponding components of x_i and x_j as [7]:

$$d_{ij} = \max_k\left(\left|x_{ik} - x_{jk}\right|\right) \quad (3)$$

Using the chosen distance measure, the correlation integral function, $C^m(r)$, can be defined as [9]:

$$C^m(r) = \frac{1}{N^2}\sum_{i\neq j} H\left(r - d_{ij}\right) \quad (4)$$

In which the Heavy-side function is as follows:

$$H(\alpha) = \begin{cases} 1 & \alpha \ge 0 \\ 0 & \alpha < 0 \end{cases} \quad (5)$$

This means that the correlation integral function is determined by making a hyper-sphere of radius *r* around every embedded data point and counting the average number of embedded data points inside this hyper-sphere. This function is a type of spatial correlation because it expresses the extent to which embedded data points are close together.

In this study, the correlation dimension is calculated by an easier way of computing the correlation integral known as Grassberger-Procaccia algorithm [10]. It assumes that the probability of two points in the set which are in hyper-sphere of size *r* is approximately equal to the probability that two points of the same set are separated by a Euclidian distance as less than or equal to *r* :

$$C^m(r) \approx \frac{1}{(1/2)N(N-1)}\sum_{i=1}^{N}\sum_{j=i+1}^{N} H\left(r - d_{ij}\right) \quad (6)$$

Assuming that the number of points in the hyper-sphere should be proportional to the radius of the corresponding hyper-sphere raised to the power D_C ,where D_C is the dimension estimation referred to as the correlation dimension, therefore correlation integral should be proportional to r^{D_C}. By setting $C^m(r) = V.r^{D_C}$, we have:

$$D_C = \lim_{r\to 0}\left[\frac{\ln C(r)}{\ln r}\right] \quad (7)$$

The estimation of the correlation dimension from data confronts some limitations. First, there is a limited range over which the correlation dimension can be evaluated because when *r* approaches the size of the phase plot, the calculated correlation function plateaus and also for very small values of r the effects of noise and finite resolution dominate in calculated correlation integral. For such reasons it is necessary to compute the correlation integral function over a wide range of r and then search for more limited range of radius over which the estimate of dimension stays consistent [7]. This is accomplished by plotting the $\ln(C(r))$ versus $\ln(r)$ and finding the range for r over which this function is reasonably linear. This linearity can be more accurately determined by calculating the slope of the $\ln(C(r))$ versus $\ln(r)$ function and evaluating the range of r for which the slope is nearly constant. The average slope over this range would be a reliable estimation for D_C, the correlation dimension [7]. Even when a good estimate of correlation dimension is acquired, a second issue should be resolved which is the proper embedding dimension used to calculate D_C. The invariant measures of an attractor of dimension d theoretically can be recovered completely in a space of dimension $2d+1$. The usual solution is to evaluate D_C as a function of m, the embedding dimension, as it is increased until the estimated dimension ceases to increase as m increases further. This convergence of estimates may occur for $d \le m < 2d+1$.

2.4. Support vector machine

In order to classify the extracted acoustic perturbation and nonlinear dynamic features, SVM classifier is hired to determine which of these features can best differentiate nodules cases from polyp. SVM is a new technique in the field of statistical learning theory which was proposed by Vapinik [11]. It is based on the structural risk minimization principle (SRM) in which two main objectives is pursued. The first is to control the empirical risk on the training data set. The second is to control the capacity of the decision functions used to obtain this risk value. It is a method of training polynomials, radial basis function, or multilayer perceptron classifiers, in which the weights of the network are found by solving a quadratic programming problem (QP) with linear inequality and equality constraints.

Assume that the training data with k number of samples are represented by $\{x_i, y_i\}, i = 1,...., k$, where $x \in R^n$ is an n-dimensional vector and $y \in \{-1,+1\}$ is the class label. The aim is to find a hyper-plane that divides the data so that all the points with the same label are on the same side of the hyper-plane. This depends on finding w and b such that:

$$y_i(w.x_i + b) > 0 \tag{8}$$

If a hyper-plane exists that satisfies (8), the two classes are said to be linearly separable.

The SVM finds the hyper-plane with maximum Euclidian distance from the training set. According to the SRM principle, there will be just one optimal hyper-plane with a specific maximal margin, defined as the sum of distances from the hyper-plane to the closest points of each class. This linear classifier threshold is the optimal separating hyper-plane, referred to as OSH. In case of linearly separable classes, it is possible to rescale w and b so that:

$$\min_{1 \le i \le k} y_i(w.x_i + b) \ge 1 \tag{9}$$

Therefore, (8) can be revised as below:

$$y_i(w.x_i + b) \ge 1 \tag{10}$$

Regarding (10), distance to the closest point is $1/\|w\|$ and the OSH can be found by minimizing $\|w\|^2$ under constraint (9). The minimization procedure uses Lagrange multipliers and quadratic programming (QP) optimization methods [11].

In the case of non-separable training sets, the i-th data point has a slack variable ξ_i, which represents the magnitude of the classification error. A penalty function $f(n)$ represents the sum of these misclassification errors as:

$$f(\xi) = \sum_{i=1}^{l} \xi_i \tag{11}$$

In this case (10) can be written as follows:

$$y_i(w.x_i + b) + \xi_i \ge 1 \tag{12}$$

The SVM solution can be found by keeping the upper bound on the VC dimension small and by minimizing an upper bound on the empirical risk, for example the number of training errors with the following minimization, under constraint (12):

$$\min_{w,b,\xi_i}\left[\frac{1}{2}\| w \|^2 + C\sum_{i=1}^{k} \xi_i\right] \tag{13}$$

The first term in (13) is the same as in the linearly separable case to control the learning capacity, while the second term controls the number of misclassified points. The regularization constant C>0 determines the trade-off between the empirical error and the complexity term. Parameter C is chosen by the user and a large C corresponding to the assignment of a higher penalty to errors.

3. EXPERIMENTAL RESULT

The PPQ, APQ and correlation dimension features are extracted from 39 sample records, with the approximate length of 1 second (more than 40,000 samples), from the patients' speech signal with vocal fold nodules and polyp (19 with nodules and 20 with polyp). The pitch period detection is implemented using the modified cepstrum method discussed in [5]. PPQ and APQ features, as acoustic perturbation measures, are calculated at smoothing factor of 5 and 11 periods, respectively. Correlation dimension is also hired to describe the nonlinear dynamical characteristics of all voice samples. Figure1 shows the slope of the *ln(C(r))* against *ln(r)* where the curves from bottom to top correspond to the ascending embedding dimensions, m. The limited range of r over which the slope value, by increasing the embedding dimension m, saturates is also marked in this figure. Then correlation dimension measure is estimated in the embedding dimension after which the slope value remains constant (Figure2).

Table.1 shows the statistical measures for the extracted features from these two types of laryngeal disorders. It shows that the presence of polyp on the vocal fold tissue leads to the higher mean value for pitch and amplitude perturbation quotients. The estimated mean value for correlation dimension also demonstrates the low dimensional and, therefore, less complex characteristics of the phase space attractors in the speech signals of the patients with nodules in comparison with polyp.

Table1- Mean and variance of features for nodules and polyp

	Vocal fold nodules		**Vocal fold polyp**	
	mean	***variance***	***mean***	***Variance***
PPQ	*0.77*	*0.22*	*1.49*	*5.85*
APQ	*3.41*	*1.39*	*5.22*	*10.03*
D2	*1.72*	*0.16*	*2.66*	*1.46*

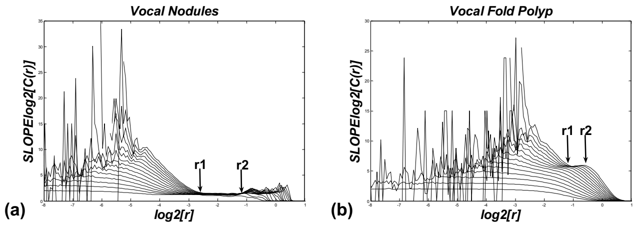

Figure1-Slope of correlation integral function vs. radius in logarithmic scale for (a) one nodules case (b) one polyp case

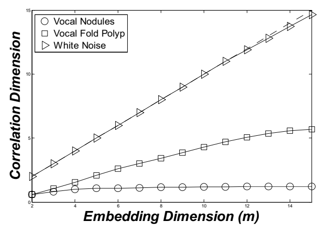

Figure2- Correlation dimension vs. embedding dimension for the specified range of radius for the nodules and polyp cases of fig.1

The result of SVM implementation for the classification of these two types of laryngeal disorders is shown in Figure3. In this procedure 60% of feature sets are selected as training data set and the remaining for test. It is shown that for all 3 tested SVM kernels, correlation dimension, as a nonlinear dynamical feature, leads to the highest correct classification percentage in differentiating vocal fold nodules from polyp.

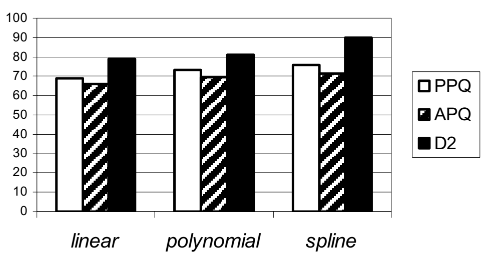

Figure3- nodules & polyp classification using three SVM kernels

4. DISCUSSION

It has been stated that the presence of nodules and polyp on the outer layer of vocal folds' tissue leads to the alteration in normal vibration pattern of the cords, and consequently, the waveform of the patients' speech signal. These effects are caused by the consistent opening of the glottal gap and the variation in biomechanical parameters of the cords, resulting in the glottal air leakage, vortex flows, asymmetrical vibration and the pitch or amplitude perturbations. Acoustic measures are mostly useful for describing the differences between the negative effects of nodules and polyp on the cycle-to-cycle variation of pitch frequency and the amplitude of patients' speech signal. However, it has been expressed that the nonlinear dynamic features are responsible for understanding the degree of the chaotic behavior of the patients' speech signals. It is also demonstrated that the correlation dimension constitutes a more reliable set of features to express the differences in the degree of complexity which is imposed on the behavior of the signals' phase space attractor. Due to their smaller size and anatomical location, nodules lead to less chaotic behavior of the signals in their phase space. The differences in the number of nonlinear variables, used to describe the attractors' behavior, for nodules and polyp is more substantial than their effect on the signals' PPQ and APQ features. The supreme ability of nonlinear dynamic features, regarding the acoustic perturbation ones, leads to the higher correct classification rate in making distinguishes between vocal fold nodules and polyp cases.

5. CONCLUSION

In this paper the ability of PPQ, APQ and correlation dimension features in describing the pathological effect of nodules and polyp on the vibration pattern of vocal folds have been discussed. The classification procedure using SVM showed that the correlation dimension feature, by estimating the degree of chaotic behavior in signals' phase space, provides us with a better explanation of the differences in the negative effect of such vocal folds' lesions. In comparison with the acoustic perturbation measures, it is concluded that the nonlinear dynamic features, play a more important role in the clinical diagnosis of vocal fold disorders such as nodules and polyp.

6. REFERENCES

[1] Lesly Wallis, Cristina Jackson-Menaldi, Wayne Holland, and Alvaro Giraldo, "Vocal Fold Nodule vs. Vocal Fold Polyp: Answer from Surgical Pathologist and Voice Pathologist Point of View" *Journal of Voice*, Vol. 18, No. 1, pp. 125–129, 2003.

[2] Caitriona Mc Hugh-Munier, Klaus R. Scherer, Willy Lehmann, and Ursula Scherer, "Coping Strategies, Personality, and Voice Quality in Patients with Vocal Fold Nodules and Polyps" *Journal of Voice* Vol. II, No. 4, pp. 452-461 (1997).

[3] Disordered Voice Database, Version 1.03, October 1994, *Massachusetts Eye and Ear Infirmary, Voice and Speech Lab, Boston, MA, Kay Elemetrics Corp*.

[4] Stefan Hadjitodorov, Petar Mitev , "A computer system for acoustic analysis of pathological voices and laryngeal diseases screening", *Medical Engineering & Physics* 24 (2002) 419–429.

[5] Petar Mitev, Stefan Hadjitodorov,"Fundamental frequency estimation of voice of patients with laryngeal disorders", *Information Sciences* 156 (2003) 3–19

[6] Andrew C. Lindgren, Michael T. Johnson, Richard J. Povinelli "Speech recognition using reconstructed phase space features", *ICASSP 2003*

[7] Eugene N. Bruce, "Biomedical Signal processing and Signal Modeling", *Wiley Series in Telecommunications and Signal Processing, 2001*

[8] Vassilis Pitsikalis, Petros Maragos, "Speech Analysis and Feature Extraction using Chaotic Models", *ICASSP 2002*

[9] "Nonlinear Biomedical Signal Processing", Volume II, Dynamic Analysis and Modeling, Edited by Metin Akay , *IEEE Press Series on Biomedical Engineering , 2001*

[10] Grassberger P, Procaccia I. Measuring the strangeness of strange attractors*.* *Physica* D. 1983; 9:189–208

[11] Vapinik, V., 1998, Statistical Learning Theory, *Wiley*, NY