Biomedical Signal Processing and Control 22 (2015) 135–145

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Detection of vocal disorders based on phase space parameters and Lyapunov spectrum

Hamzeh Ghasemzadeh [a,*], Mehdi Tajik Khass [b], Meisam Khalil Arjmandi [c], Mohammad Pooyan [d]

[a] Faculty of Electrical and Computer Engineering, Islamic Azad University, Damavand Branch, Damavand, Iran
[b] Faculty of Electrical and Computer Engineering, Tabriz University, Tabriz, Iran
[c] Dept. of Communicative Sciences and Disorders, Michigan State University, MI, USA
[d] Faculty of Biomedical Engineering, Shahed University, Tehran, Iran

ARTICLE INFO

Article history:
Received 5 February 2015
Received in revised form 20 June 2015
Accepted 7 July 2015
Available online 29 July 2015

Keywords:
Voice pathology detection
Chaos theory
Genetic algorithm
Linear discriminant analysis
Support vector machine

ABSTRACT

Previous studies have shown that the underlying process of speech generation exhibits nonlinear characteristics. Since linear features cannot represent a nonlinear system thoroughly, this paper employs new sets of non-linear measurement for assessing the quality of recorded voices. Such measurement could be exploited for implementing efficient and convenient systems for diagnosing laryngeal diseases without using invasive methods. Three sets of features based on mutual information, false neighbor fraction, and Lyapunov spectrum are investigated to this end. Furthermore, distributions of the proposed features and their discriminative property are investigated. Moreover, the described procedure benefits from the synergy between different concepts of pattern recognition. First, a genetic algorithm (GA) is invoked to find a-near optimum subset of features. Second, linear discriminant analysis (LDA) is applied to remove remaining redundancies and correlations between selected features. Finally, support vector machine (SVM) is employed for learning decision boundaries. Sensitivity and specificity of 99.3% and 94% respectively were achieved in the simulation results.

1. Introduction

Diseases related to human voice production system such as laryngeal cancer and polyps usually cause severe damages to patients' health and even their social life. Fortunately, most of these diseases can be cured if they are detected at early stages. Since laryngeal syndromes are likely to cause anomalies in a patient's voice (such as breathiness and hoarseness which are two common symptoms of human speech system's malfunction), some well-versed specialists can discern the problem just by listening to the patient's voice and decide to prescribe direct examinations like laryngoscopy. However, these examinations are very expensive and time-consuming. Moreover, they cause discomfort to patients as they are invasive. Thus, some pre-examinations are worthwhile. One major drawback of perceptual examinations is their inherent subjectivity which makes them unreliable and difficult to quantify [1,2]. To overcome these problems, researchers have been seeking a reliable way – usually based on speech signal processing techniques – to distinguish between healthy and pathologic voices. This procedure can be implemented in a mobile device or even simple applications on electronic gadgets, so everybody – especially those who are at high risks due to genetic defects or professions (like professional singers) – can do the primary examinations easily.

In order to distinguish between healthy and pathologic voices, firstly some features have to be found whose differences are distinctly vivid. Then, voices are categorized into healthy or pathological using some classifying methods such as SVM or neural network (NN). Over the past years, several measures have been introduced to quantify the abnormality of humans' speech signals. These measures include fundamental frequency (f_0) [3,4], pitch perturbation (jitter) and amplitude perturbation (shimmer) [5,6], harmonic to noise ratio (HNR) [7], low to high energy ratio (LHR) [8], glottal to noise excitation ratio (GNE) [9], normalized noise energy (NNE) [10], signal to noise ratio (SNR) [11], and mel-frequency cepstral coefficients (MFCC) [12–14]. Jitter and shimmer are measures of short-term (cycle-to-cycle) variation in the fundamental frequency and amplitude of a voice signal. All of these measures are based on the linear acoustic theory [15–18] and assume a planar sound propagation through vocal tract.

* Corresponding author at: Islamic Azad University, Damavand Branch, Daneshghah Street, Damavand, Iran. Tel.: +98 9010344171.
E-mail address: hamzeh_g62@yahoo.com (H. Ghasemzadeh).

http://dx.doi.org/10.1016/j.bspc.2015.07.002
1746-8094/

Further researches have shown that the linear model is inadequate. In another word, airflow propagation through human's vocal tract is more likely to follow the fluid dynamic rules which lead to nonlinear models [19–21]. Specifically, phenomena such as subharmonic oscillation, sudden change of vocal fold behavior, and octave jumps show that voice signal presents highly non-linear vibratory systems and therefore the chaos theory may be employed [22–25]. For example, transitions between different voice attractors can be characterized in terms of bifurcations [23,26].

In recent years, the chaos theory has been used as a powerful tool for analyzing nonlinear systems [27,28]. Correlation dimension (CD) [29] and the largest Lyapunov exponent (LLE) [28] are among the most popular nonlinear features. Zhang and Jiang deployed CD to discriminate between different types of speech signals [30]. Efficacy of nonlinear features for discriminating between pathological and healthy voices in the case of running voices was shown in [31]. Giovanni et al. used LLE to diagnose the unilateral laryngeal paralysis [32]. Zhang et al. applied CD and entropy measures to analyze the sustained vowels of patients before and after surgical excision of vocal polyps [33]. Vaziri et al. exploited CD and LLE for classifying voices [34]. Henríquez reported a very high accuracy based on nonlinear features and neural networks [35]. Arias et al. argued about performance of this method and noticed that normal and pathological voices have different lengths in MEEI database. They argued that high accuracy of this method could be biased by the different lengths of the files [14].

Employing nonlinear analysis of voices, this paper makes the following contributions:

- While there is a whole spectrum of Lyapunov exponents, to the best of our knowledge, previous works have only used LLE. This work uses the whole spectrum. Furthermore, two new sets of features based on false neighbor fraction (FNF) and mutual information (MI) are proposed.
- Using a mathematical approach, it is shown that the proposed features have different distributions and therefore, they have good discriminative properties. This is achieved by employing the simplified model of additive uncorrelated noise for pathological voices.
- The final decision is achieved by synergy of GA for feature selection, LDA for feature reduction, and SVM for learning decision regions.

The rest of this paper is organized as follows. Section 2 presents a brief introduction to the chaos theory and the employed mathematical model for pathological voices. Nonlinear features are defined in Section 3. Section 4 is devoted to the proposed method and analyzing the proposed features. Experimental results are presented in Section 5. Section 6 discusses the proposed method and conclusions are finally reported in Section 7.

2. Chaos in pathological voices

2.1. Chaotic model of a system

In order to apply chaotic analysis, it is necessary to construct the phase space of a system. Fortunately, Takens embedding theorem shows that if the embedding dimension is sufficiently high, the phase space can be constructed from time series samples [36]. The most important technique for constructing the phase space is the method of delays [37].

Let $x(t)$ denote time series produced by an unknown dynamic system. Then, m-dimensional vector of $\boldsymbol{s(t)}$ can be constructed by successive time delays of $x(t)$:

$$\boldsymbol{s(t)} = [x(t), x(t+\tau), \ldots, x(t+(m-1)\tau)] \tag{1}$$

where τ and m are called time-delay and embedding dimension, respectively. In practice, these parameters are not known a-prior and should be estimated [38]. Fig. 1 shows reconstructed phase space of a normal and some pathological samples.

2.2. Mathematical model of impaired voices

In order to compare distributions of features extracted from normal and pathological voices, a mathematical model is employed. Furthermore, to make this mathematical problem tractable, an admittedly simplified model of additive noise is employed. It is noteworthy that this model is only employed for providing mathematical insight into discriminating properties of the proposed features. In another words, there is no claim that all pathological impairments follow this assumption or that the only difference between normal and pathological voices is due to their noises. Also, the uncorrelated noise does not make the system chaotic, but as it is discussed in Section 4.2.1 it can alter chaotic structure of a signal.

A pathological voice, $x_P(t)$, is modeled as the sum of two independent sources:

$$x_P(t) = x_N(t) + n(t) \tag{2}$$

where $x_N(t)$ represents a normal subject voice without any pathological anomaly and $n(t)$ is an *i.i.d* random signal with zero mean and variance of σ^2.

3. Nonlinear features and their analysis

Anomalies in pathological voices stem from malfunctions of some parts of voice production system. Consequently, it is logical to assume that pathological and normal voices stem from systems with different dynamics. In this section, we will show that nonlinear features can reflect these differences.

3.1. False neighbor fraction (FNF)

In dimension m, each vector $\boldsymbol{s(t)_m}$ has a nearest neighbor $\boldsymbol{s(\tilde{t})_m}$ with respect to a distance measure. Let us denote their distance by d_m^2:

$$d_m^2 = ||\boldsymbol{s(t)_m} - \boldsymbol{s(\tilde{t})_m}||^2 \tag{3}$$

Also, let $\boldsymbol{s(t)_{m+1}}$ and $\boldsymbol{s(\tilde{t})_{m+1}}$ be maps of $\boldsymbol{s(t)_m}$ and $\boldsymbol{s(\tilde{t})_m}$ in dimension $m+1$. Then, distance between these points in the new dimension is:

$$d_{m+1}^2 = d_m^2 + (x(t+m\tau) - x(\tilde{t}+m\tau))^2 \tag{4}$$

In other words, by going from dimension m to $m+1$ two neighbor points diverge with the factor of

$$D = ||x(t+m\tau) - x(\tilde{t}+m\tau)||^2 \tag{5}$$

If the value of D is larger than a specific threshold [39,40], we call these points false neighbors. It can be argued that these points are neighbors in dimension m, because of the folding that occurs by projecting a high dimensional attractor down to dimension m (Fig. 2).

After all false neighbors are found, FNF is calculated according to:

$$\text{FNF} = \frac{\textit{No. of false neighbors}}{\textit{No. of all neighbors}} \times 100\% \tag{6}$$