Computers in Biology and Medicine 37 (2007) 571–578

Computers in Biology and Medicine

www.intl.elsevierhealth.com/journals/cobm

Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders☆

Everthon Silva Fonseca[a,b,*], Rodrigo Capobianco Guido[a], Paulo Rogério Scalassara[a], Carlos Dias Maciel[a], José Carlos Pereira[a]

[a] SEL/EESC/USP and IFSC/USP—Department of Electrical Engineering, School of Engineering at São Carlos and Institute of Physics at Sao Carlos, University of São Paulo, SP, Brazil

[b] EE/UCLA—School of Engineering and Applied Sciences, University of California at Los Angeles, CA, USA

Abstract

This work describes a novel algorithm to identify laryngeal pathologies, by the digital analysis of the voice. It is based on Daubechies' discrete wavelet transform (DWT-db), linear prediction coefficients (LPC), and least squares support vector machines (LS-SVM). Wavelets with different support-sizes and three LS-SVM kernels are compared. Particularly, the proposed approach, implemented with modest computer requirements, leads to an adequate larynx pathology classifier to identify nodules in vocal folds. It presents over 90% of classification accuracy and has a low order of computational complexity in relation to the speech signal's length.

Keywords: Voice disorders; Wavelet transform; LPC; SVM; Pattern recognition in spoken language

1. Introduction

Discrete-time processing of recorded voice signals [1] can be used to detect different acoustical characteristics that differentiate between normal and pathologically affected human voices. Pathologies related to the glottal tract are usually identified through acoustic perceptual standards like *breathness*, hoarseness and harshness [2–4]. However, due to the complex structure of the biological system for speech synthesis, pathologies with harsh characteristics may be confused with those perceptually defined as hoarse [5]. The turbulence in glottal flow, resulting from malfunction of the vocal folds, can be quantified by the noise in spectral components of speech [6]. Pathologies caused by soft or incomplete closure of the glottis, as nodules in vocal folds, are often associated with high-frequency noise [7,8]. Thus, we intend to analyze this particular high frequency characteristic of pathologically affected voices in order to distinguish them from the normal ones.

Most of the recent computer-based algorithms for laryngeal pathology detection described in the literature are based on wavelets, fractals or neural maps and networks [9,10]. Neural maps and networks cover over 95% of the existing techniques, some of them reaching almost 100% accuracy in the results when a good procedure is used to train the classifiers, but, sometimes, with a high computational order of complexity in relation to the signal's length. Usually, in this last kind of classifier, the voices are clusterized in respect to the following parameters: formant frequencies, pitch period and its deviations, stability of pitch period during vowel phonation, degree of dissimilarity of the shape of the pitch, low-to-high energy ratio (LHER), noise-to-harmonics ratio (NHR) and harmonics-to-noise ratio (HNR). Fractal-based classifiers have about 90% classification accuracy, but they usually detect only some particular pathologies, like Friedreich's ataxia for example [11,12]. Best-basis wavelet classifiers produce about 85% of classification accuracy.

This work proposes an algorithm, with a low order of computational complexity, to identify patients with nodules [13] in vocal folds. It is based on Daubechies' discrete wavelet transform

☆ Based on *Everthon Silva Fonseca*'s Ph.D. Thesis.

* Corresponding author.

E-mail addresses: everthon@sel.eesc.usp.br (E.S. Fonseca), guido@ifsc.usp.br (R.C. Guido), scalassara@sel.eesc.usp.br (P.R. Scalassara), maciel@sel.eesc.usp.br (C.D. Maciel), pereira@sel.eesc.usp.br (J.C. Pereira).

0010-4825/$ - see front matter © 2006 Published by Elsevier Ltd.
doi:10.1016/j.compbiomed.2006.08.008

(DWT-db) [14–16], linear prediction code (LPC) [1], and least squares support vector machines (LS-SVM) [17–20], the latter being a statistical learning technique for training classifiers. LS-SVMs were originally created for two-group classification problems using a hyper-linear separating plane. The proposed algorithm produces the output, the results of the classification, much faster than other related methods. In terms of accuracy, it advances the state-of-the-art classifiers [10] for nodule pathology detection.

This article is organized as follows: Section 2 presents a brief review on DWT-db and LS-SVM, Section 3 details the proposed algorithm, and Section 4 describes the methodology and tests. Lastly, Section 5 lists the results and Section 6 presents the conclusions.

2. A review on DWT-db and LS-SVM

2.1. DWT-db

The discrete wavelet transform (DWT), whose main idea is the process of multi-resolution analysis (MRA) proposed by Mallat [15], is one of the most appropriate techniques to make a joint time-frequency analysis of discrete-time signals. It allows one to find the time-support of frequencies. Considering $\vec{f}$ as the discrete-time signal under analysis, it is decomposed in the sum of two other vectors, $\vec{A}$ and $\vec{D}$, called, respectively, *trend* and *fluctuation* [21],

$$\vec{f} = \vec{A} + \vec{D}, \tag{1}$$

where

$$\vec{A} = \sum_{k=0}^{n/2-1} \langle \vec{f}, \vec{v_k} \rangle \vec{v_k} \quad \text{and} \quad \vec{D} = \sum_{k=0}^{n/2-1} \langle \vec{f}, \vec{w_k} \rangle \vec{w_k}.$$

Then, for a discrete-time signal $\vec{f}$ containing n samples:

- $\vec{A}$ is the projection of $\vec{f}$ onto a subspace V, with a basis containing $n/2$ vectors;
- $\vec{D}$ is the projection of $\vec{f}$ onto a subspace W, with a basis containing $n/2$ vectors;
- $V \perp W \leftrightarrow \vec{A} \perp \vec{D}$;
- $\vec{v_i} \perp \vec{w_i} \leftrightarrow \langle \vec{v_i}, \vec{w_i} \rangle = 0$.

The process above is a one-level decomposition. When n is an integer power of 2, this process can be repeated $j = \log(n)/\log(2)$ times. To do that, the resulting signal $\vec{A}$ is decomposed one or more times, creating a decomposition of j levels. This is the main idea behind the MRA analysis: decomposing a signal in several levels of resolution (Eq. (2)):

$$\vec{f} = \vec{A_j} + \sum_{i=1}^{j} \vec{D_i}. \tag{2}$$

Thus,

- $\vec{A_j}$ is the projection of $\vec{f}$ onto a subspace V_j, with a basis containing $n/2^j$ vectors;
- $\vec{D_i}$ is the projection of $\vec{f}$ onto a subspace W_i, with a basis containing $n/2^i$ vectors;
- $V_j \perp W_j \leftrightarrow \vec{A_j} \perp \vec{D_j}$;
- $\vec{v_{i,j}} \perp \vec{w_{i,j}} \leftrightarrow \langle \vec{v_{i,j}}, \vec{w_{i,j}} \rangle = 0$.

In other words,

$$f[n] = \sum_{k=0}^{n/2^j-1} H_{j,k}[n]\phi_{j,k}[n] + \sum_{t=1}^{j} \sum_{k=0}^{n/2^j-1} G_{t,k}[n]\psi_{t,k}[n], \tag{3}$$

where

- $\phi[n]$ and $\psi[n]$ form a Riezs basis [15] to write signal f;
- $\phi[n]=\sum_k h_n \phi[2n-k]$, defined recursively by dilations and translations of itself, is called *scaling function* [15];
- $\psi[n]=\sum_k g_n \phi[2n-k]$, defined recursively, is called *wavelet function* and is orthogonal to the *scaling function*;
- $H_{j,k}[n] = \langle f, \phi_{j,k}[n] \rangle$;
- $G_{t,k}[n] = \langle f, \psi_{t,k}[n] \rangle$;
- $\{0\} \leftarrow \cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots \rightarrow L^2$;
- if $f[n] \in V_j \rightarrow f[2n] \in V_{j+1}$;
- $V_{j+1} = V_j \oplus W_j$;
- the h_k coefficients form a low-pass filter;
- the g_k coefficients form a high-pass filter;
- the h_k and g_k coefficients form the *analysis filter bank*;
- a filter with k coefficients is called a filter of support k.

In practice, to compute the DWT, the coefficients of the signal under analysis, $f[\,]$, are convolved both with the low-pass filter, $h[\,]$, and the high-pass filter, $g[\,]$, as follows:

$$y_{\text{low-pass}}[\,] = f[\,] * h[\,] = \sum_{k=0}^{M-1} h[k]f[2n-k], \quad 0 \leqslant n \leqslant \frac{N}{2}, \tag{4}$$

$$y_{\text{high-pass}}[\,] = f[\,] * g[\,] = \sum_{k=0}^{M-1} g[k]f[2n-k], \quad 0 \leqslant n \leqslant \frac{N}{2}, \tag{5}$$

where $y_{\text{low-pass}}$ and $y_{\text{high-pass}}$ are the outputs, M is the length of signals h and g, N is the length of signal f, and $*$ denotes the discrete-time convolution.

In each level of decomposition, there is both a down-sampling by 2 of the transformed signals and a wrap-around process [15], because the convolutions above are, in fact, filters that allow a half-band of the original signal to pass, according to *Nyquist rate* [1]. It is not the goal of this work to detail all of the properties, but there are three special properties of a wavelet transform that are used in this work. They are:

- *Signal energy* is defined as the scalar value E in Eq. (6):

$$E = \sum_{n=0}^{N-1} f[n]^2, \tag{6}$$

N being the length of signal f;