Computer Methods and Programs in Biomedicine 153 (2018) 41–51

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Correlation analysis of respiratory signals by using parallel coordinate plots

Esra Saatci

Department of Electrical and Electronics Engineering, Istanbul Kultur University, Bakirkoy, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 6 October 2016
Revised 8 May 2017
Accepted 2 October 2017

Keywords:
Statistical analysis of biomedical signals
Respiratory signal modeling
Parallel coordinate plots of the respiratory signals

ABSTRACT

Background and objectives: The understanding of the bonds and the relationships between the respiratory signals, i.e. the airflow, the mouth pressure, the relative temperature and the relative humidity during breathing may provide the improvement on the measurement methods of respiratory mechanics and sensor designs or the exploration of the several possible applications in the analysis of respiratory disorders. Therefore, the main objective of this study was to propose a new combination of methods in order to determine the relationship between respiratory signals as a multidimensional data.
Methods: In order to reveal the coupling between the processes two very different methods were used: the well-known statistical correlation analysis (i.e. Pearson's correlation and cross-correlation coefficient) and parallel coordinate plots (PCPs). Curve bundling with the number intersections for the correlation analysis, Least Mean Square Time Delay Estimator (LMS-TDE) for the point delay detection and visual metrics for the recognition of the visual structures were proposed and utilized in PCP.
Results: The number of intersections was increased when the correlation coefficient changed from high positive to high negative correlation between the respiratory signals, especially if whole breath was processed. LMS-TDE coefficients plotted in PCP indicated well-matched point delay results to the findings in the correlation analysis. Visual inspection of PCB by visual metrics showed range, dispersions, entropy comparisons and linear and sinusoidal-like relationships between the respiratory signals.
Conclusion: It is demonstrated that the basic correlation analysis together with the parallel coordinate plots perceptually motivates the visual metrics in the display and thus can be considered as an aid to the user analysis by providing meaningful views of the data.

1. Introduction

In respiratory studies, airflow to and from the lungs and mouth pressure (corresponds to the pressure inside the lungs) are concerned in order to model the lung statics and dynamics. Airflow and pressure measurements are performed by utilizing direct (via spirometry, pneumotachograph, etc.) [1] and indirect (via temperature, tracheal sound, electrocardiography, etc.) [2–4] measurement methods; in invasive (with and without ventilatory support) and noninvasive (via facemask or nose cannula) [5] conditions; and with or without physical activities (sleeping, speaking, etc.) [5]. On the other hand, the measurements of temperature and humidity within nasal cavity, and association between nasal temperature, airflow and the quality of life improvement of chronic obstructive pulmonary disease (COPD) patients are analyzed in the medical studies [6,7]. In all these studies the specific aim was either to estimate the parameters of the respiratory system underpinning the healthy or diseased lungs or to construct the fast and accurate measurement methods and/or sensor systems. However, the understanding of the bonds and the relationships between the respiratory signals, i.e. the airflow, the mouth pressure, the relative temperature and the relative humidity during spontaneous breathing may provide the improvement on the measurement methods of respiratory mechanics and sensor designs or the exploration of the several possible applications in the analysis of respiratory disorders. For instance, the knowledge of the linear relationship between the respiratory signals can shed light on the studies, in which indirect, non-invasive and basic sensors [7] are designed to be used to determine the respiratory mechanics [2,8,9] or to detect the many diseased conditions in breath testing systems [10,11].

Therefore, the main objective of this study was to propose a new combination of methods in order to determine the relationship between volumetric airflow, mouth pressure, inhaled and exhaled air temperature and humidity in noninvasive and spontaneous breathing conditions. The important motivation was to present the first development steps of representation methods to allow for the quantitative and the qualitative investigations of cor-

E-mail address: esra.saatci@iku.edu.tr

https://doi.org/10.1016/j.cmpb.2017.10.003
0169-2607/

relations between these signals. In this study, in order to reveal the coupling between the processes two very different methods were used: the statistical correlation analysis and parallel coordinate systems. The statistical correlation analysis is a well-known method in the time-domain that can shed some light on the causal relationship between two processes that generate the signals [12–14]. However, it is known that correlation can be misinterpreted if the underlying assumption of a joint normal distribution and the linearity interrelation of both signals are violated [15]. Visualization tools may be used to complement the correlation analysis [15,16]. To the best of our knowledge, neither the comprehensive correlation analysis of the respiratory signals in time-domain nor the visual correlation presentation has been studied in the literature. On the other hand, coupling between airflow and many physiological structure during breathing has been quantitatively investigated in [17] and [18] and coherence, which is the correlation in the frequency-domain has been utilized extensively in order to reveal the reliability and feasibility of the forced oscillation technique (FOT) when the spectrums of the airflow and pressure signals are fitted to the linear or nonlinear impedance model to estimate the respiratory parameters. However, none of these studies provides a qualitative and quantitative proof that yields the time lags or scales the linearity occurs between respiratory signals. Moreover, with the absence of the evidence which confirms the linear or nonlinear interactions of the signals in time scale, modeling and design issues for the respiratory system becomes obscure.

Thus, the existing correlation analysis supported by the parallel coordinate plots (PCP) are proposed in order to reveal dependencies and bonds between respiratory signals which, in turn may establish a new tool for the respiratory system studies. The advantages of the PCPs which were straightforward mapping of multi-dimensions to 2D, better human perception seeing adjacency (i.e. lines showing connectivity) and patterns (e.g. groups of lines) would help to achieve this aim. On the other hand, the reported disadvantages of PCPs were solved by using bundles for the categorical data observation, visualization metrics and axis manipulation and coloring for the correct data interaction and separating the inspiratory and expiratory plots for the clear and evident understanding of the plots.

2. Methods

This section provides a brief summary of the correlation analysis and the applications of PCPs to the correlation analysis. Throughout this section, due to consistency and understandability of the methods some results from the estimations were presented in the form of mean (standard deviation). The mean and standard deviation were calculated as an average of the results from 10 separate breath cycles belonging to the one representative subject.

2.1. Pearson's correlation coefficient

The Pearson's correlation $\rho_{XY}(n_1, n_2)$ between two one-dimensional zero-mean signals $X(n)$ and $Y(n)$ is defined as:

$$\rho_{XY}(n_1, n_2) = \frac{COV(X(n_1), Y(n_2))}{\sigma_X \sigma_Y} \tag{1}$$

where n is the data points (corresponds to the time in seconds), $X(n)$ and $Y(n)$ are assumed to be jointly stationary discrete-time random processes with $COV(X(n_1), Y(n_2)) = E[X(n_1)Y(n_2)]$, $\sigma_X = \sqrt{E\left[X^2(n)\right]}$ and $E[.]$ is the expected value operator. An estimate of ρ_{XY} based on the observations $\{X(n), Y(n)\}_{n=1}^{N}$ is given by

$$\hat{\rho}_{XY} = \frac{\frac{1}{N}\sum_{n=1}^{N} X(n)Y(n)}{\left[\frac{1}{N}\sum_{n=1}^{N} X^2(n)\right]^{1/2}\left[\frac{1}{N}\sum_{n=1}^{N} Y^2(n)\right]^{1/2}} \tag{2}$$

The simplest linear relationship between $X(n)$ and $Y(n)$ can be detected from $\hat{\rho}_{XY}$ by observing the values ranges from -1 (perfect negative correlation) via 0 (no correlation) to 1 (perfect positive correlation). Although in this concept, the single measure of the correlation coefficient is easy to interpret, it is demonstrated in [15] that the estimate in (1) is biased, in terms of accuracy, toward the high correlation coefficients. Thus for the correct correlation judgment, additional assessments are required.

2.2. Cross-correlation coefficient

The cross-correlation coefficient $r_{XY}(l)$ between two one-dimensional zero-mean signals $X(n)$ and $Y(n)$ is defined as [19]:

$$r_{XY}(l) = \frac{COV(X(n), Y(n+l))}{\sigma_X \sigma_Y} \tag{3}$$

where l is the lags in data points, $X(n)$ and $Y(n)$ are assumed to be jointly stationary discrete-time random processes with $COV(X(n), Y(n+l)) = E[X(n)Y(n+l)]$. An estimate of $r_{XY}(l)$ based on the observations $\{X(n), Y(n)\}_{n=1}^{N}$ is given by

$$\hat{r}_{XY}(l) = \begin{cases} \dfrac{1}{N}\displaystyle\sum_{n=1}^{N-l} X(n)Y(n+l) & 0 \le l \le N-1 \\ \dfrac{1}{N}\displaystyle\sum_{n=1-l}^{N} X(n)Y(n+l) & -(N-1) \le l \le -1 \end{cases} \tag{4}$$

Cross-correlation coefficient (or lagged correlation coefficient) has been used in many applications to detect delayed relationship between signals [19,20]. The most basic time-domain method used to estimate the point delay (corresponds to time delay) of the signals, is simply to locate the maximum value in the cross-correlation coefficient function [20]. Thus, for the complete correlation analysis, the point delay between respiratory signals were estimated by cross-correlation coefficient. Here, the vital assumption is that respiratory signals are periodically changing between inspiratory and expiratory phases.

However, cross-correlation coefficient requires a preprocessing step by which the samples of the signals are imposed to be uncorrelated [19,21]. In this way, the cross-correlation between the signals reveals clearly the time delay as a peak at the corresponding lag. Respiratory signals have a considerable correlation between samples. Hence, decorrelation and whitening were applied to each respiratory signal before the cross-correlation analysis (see Fig. 1). In this study basic decorrelation algorithm (i.e. Mahalanobis transformation) is proposed in the preprocessing step [22].

In this method, the correlated zero mean random vector $\mathbf{x} = [X(1), X(2), \ldots, X(N)]^T$ is transformed to the decorrelated and whitened version $\mathbf{x}^w = [X^w(1), X^w(2), \ldots, X^w(N)]^T$ by the whitening matrix $\mathbf{W}$ ($\mathbf{X}^w = \mathbf{WX}$ and $\mathbf{R}_{\mathbf{X}^w} = \mathbf{E}[\mathbf{X}^w\mathbf{X}^{wT}] = \mathbf{W}\mathbf{R}_{\mathbf{X}}\mathbf{W}^T = \mathbf{I}_N$ where $\mathbf{I}_N$ is the identity matrix). In general autocorrelation matrix of $\mathbf{R}_{\mathbf{X}}$ is symmetric, usually positive definite but not diagonal. The spectral decomposition theorem claims that any symmetric matrix $\mathbf{R}_{\mathbf{X}}$ can be written as $\mathbf{R}_{\mathbf{X}} = \mathbf{U}\Lambda\mathbf{U}^{\mathbf{T}}$, where the columns of $\mathbf{U}$ are the eigenvectors with associated eigenvalues Λ which is the diagonal matrix having the eigenvalues as its diagonal elements. It can be shown easily that $\Lambda^{-1/2}\mathbf{U}^{\mathbf{T}}\mathbf{R}_{\mathbf{X}}\mathbf{U}\Lambda^{-1/2} = \mathbf{I}_N$ by which the whitened signal $\mathbf{X}^w$ is obtained by the whitening matrix

$$\mathbf{W} = \mathbf{U}\Lambda^{-1/2}$$

The main problem of decorrelation is due to the additive measurement noise components in the signals [22]. The most important step in this approach is removing the estimated additive white