Article

An Efficient Algorithm for Automatic Peak Detection in Noisy Periodic and Quasi-Periodic Signals

Felix Scholkmann *, Jens Boss and Martin Wolf

Biomedical Optics Research Laboratory, Division of Neonatology, University Hospital Zurich, 8091 Zurich, Switzerland; E-Mails: bossj@student.ethz.ch (J.B.); martin.wolf@usz.ch (M.W.)

* Author to whom correspondence should be addressed; E-Mail: felix.scholkmann@usz.ch; Tel.: +41-(0)44-255-93-26.

Received: 3 August 2012; in revised form: 29 October 2012 / Accepted: 13 November 2012 / Published: 21 November 2012

Abstract: We present a new method for automatic detection of peaks in noisy periodic and quasi-periodic signals. The new method, called automatic multiscale-based peak detection (AMPD), is based on the calculation and analysis of the local maxima scalogram, a matrix comprising the scale-dependent occurrences of local maxima. The usefulness of the proposed method is shown by applying the AMPD algorithm to simulated and real-world signals.

Keywords: peak detection; local maxima scalogram; multiscale local maxima detection; automatic multiscale-based peak detection (AMPD) algorithm

1. Introduction

The detection of peaks in signals is an important step in many signal processing applications. Up to now, many different methods have been developed, including those based on traditional window-threshold techniques [1–4], wavelet transform [5–13], Hilbert transform [14], combining Hilbert and wavelet transform [15], artificial neural networks [16,17], techniques using templates [18,19], morphology filtering [20–22], nonlinear filtering [23–26], Kalman filtering [27], Gabor filtering [28], Gaussian second derivative filtering [29], linear prediction analysis [30], higher-order statistics [31], K-Means clustering [32], fuzzy C-Means clustering [33], Empirical Mode Decomposition [34], hidden Markov models [35], and techniques using entropy [36], momentum [37], histogram/cumulative
distribution function [38], intensity weighted variance [39], stochastic resonance [40], or a smoothed nonlinear energy operator [41].

The problem with most of the peak detection algorithms available is that the more generally applicable the algorithm, the more free parameters (i.e., window length or threshold value) have to be chosen in order to be able to apply the algorithm to the signal. In contrast, algorithms with only a few parameters are in general restricted for use in specific applications such as the detection of R-peaks in electroencephalography (EEG) signals [1,7,8,10,12,14–17,20–24,30–34] or peaks in chromatography data [3,4]. Additionally the existence of noise in the analyzed signal is a challenge for many of the peak detection algorithm available.

The unsatisfactory situation with regard to existing peak detection methods motivated the development of a simple and effective new peak detection algorithm for noisy periodic and quasi-periodic signals based on the insight that the general applicability of an algorithm could be preserved by (i) using a multiscale technique to detect all local maxima of the signal and (ii) finding the “real peaks” by automatically analyzing the results of the multiscale technique applied. In particular, the goal of our work was to develop a peak detection algorithm that has the properties of (i) not having any free parameters values that have to be chosen by the user prior to the analysis; (ii) being able to detect peaks in periodic and quasi-periodic signals; and (iii) having a peak detection efficiency that is fairly robust against high and low frequency noise.

It should be noted that the goal of the present work was to develop a peak detection algorithm specifically to detect peaks in noisy periodic and quasi-periodic signals. Thus, it was not the intention to develop a general peak detection method that is also applicable to non-periodic signals, such as signals normally obtained in the field of analytical chemistry, including mass spectra, optical spectra or gas chromatograms.

2. Algorithm

Let \( x = [x_1, x_2, \ldots, x_i, \ldots, x_N] \) be a given univariate uniformly sampled signal containing a component of periodic or quasi-periodic peaks. The first step of the automatic multiscale-based peak detection (AMPD) algorithm consists of calculating the local maxima scalogram (LMS). To this end, the signal \( x \) is first linearly detrended, i.e., the least-squares fit of a straight line to \( x \) is calculated and subtracted from \( x \). Then, the local maxima of the signal \( x \) are determined using a moving window approach whereby the window length \( w_k \) is varied \( \{w_k = 2k/k = 1, 2, \ldots, L\} \) with \( k \) the \( k \)-th scale of the signal and \( L = \lceil N/2 \rceil - 1 \) where \( \lceil z \rceil \) is the ceiling function that gives the smallest integer not less than \( z \). This is realized for every scale \( k \) and for \( i = k + 2, \ldots, N - k + 1 \) according to

\[
m_{k,i} = \begin{cases} 0, & x_{i-1} > x_{i-k-1} \land x_{i-1} > x_{i+k-1} \\ r + \alpha, & \text{otherwise} \end{cases}
\]

where \( r \) is a uniformly distributed random number in the range \([0, 1]\) and \( \alpha \) a constant factor (\( \alpha = 1 \)). For \( i = 1, \ldots, k + 1 \), and for \( i = N - k + 2, \ldots, N \) the value \( r + \alpha \) is assigned to \( m_{k,i} \). These operations of Equation 1 result in the matrix
where the $k$-th row contains the value for the window length $w_k$.

Thus, all elements of the $L \times N$ matrix $\mathbf{M}$ are in the range of $[0, 1 + \alpha]$. The matrix $\mathbf{M}$ is called the LMS of the signal $\mathbf{x}$.

The second step of the algorithm comprises a row-wise summation of the LMS matrix $\mathbf{M}$

$$
\gamma_k = \sum_{i=1}^{N} m_{k,i}, \text{ for } k \in \{1,2,\ldots,L\}
$$

The vector $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_i, \ldots, \gamma_L]$ contains the information about the scale-dependent distribution of zeros (and thus local maxima). The global minimum of $\boldsymbol{\gamma}$, $\lambda = \arg \min(\gamma_k)$, represents the scale with the most local maxima. This value $\lambda$ is used in the third step of the algorithm to reshape the LMS matrix $\mathbf{M}$ by removing all elements $m_{k,j}$ for which $k > \lambda$ holds, leading to the new $\lambda \times N$-matrix $\mathbf{M}_r = (m_{k,j})$, for $i \in \{1,2,\ldots,N\}$ and $k \in \{1,2,\ldots,\lambda\}$.

In the last step of the algorithm, the peaks are detected by (i) calculating the column-wise standard deviation of the matrix $\mathbf{M}_r$, according to

$$
\sigma_i = \frac{1}{\lambda - 1} \sum_{k=1}^{\lambda} \left( m_{k,i} - \frac{1}{\lambda} \sum_{k=1}^{\lambda} m_{k,j} \right)^{1/2}, \text{ for } i \in \{1,2,\ldots,N\}
$$

and (ii) find all indices $i$ for which $\sigma_i = 0$ holds. These values are stored in the vector $\mathbf{p} = [p_1, p_2, \ldots, p_q, \ldots, p_{\hat{N}}]$ with $\hat{N}$ the total number of detected peaks of the signal $\mathbf{x}$. $\mathbf{p}$ refers to the indices of the detected peaks. We found empirically that for quasi-periodic input signals the highest frequency ($f_{\text{max}}$) of the oscillation should not be larger than 4 times the lowest frequency ($f_{\text{min}}$), i.e., $f_{\text{max}} < 4f_{\text{min}}$.

Thus, using this AMPD algorithmic framework, every value of a given signal $\mathbf{x} = [x_1, x_2, \ldots, x_i, \ldots, x_N]$ is detected as a peak when the column-wise standard deviation of the matrix $\mathbf{M}_r$ meet the criterion $\sigma_i = 0$.

Figure 1 summarizes visually the steps of the AMPD algorithm and Figure 2 shows an example of applying the AMPD algorithm to a simulated noisy multicomponent signal $\mathbf{x}$, defined as $\mathbf{x} = [x_1, x_2, \ldots, x_i, \ldots, x_N]$ , with $x_i = \{a \sin(2\pi f_1/i f_s) + b \sin(2\pi f_2/i f_s) + c \sin(2\pi f_3/i f_s) + d \varepsilon_i\}$ for $i \in \{0,0.1,\ldots,N\}$ where $f_s = 800$ Hz, $f_1 = 10$ Hz, $f_2 = 70$ Hz, $f_3 = 5$ Hz, $a = 1$, $b = 1$, $c = 0.5$, $d = 0.1$ and $\varepsilon$ a normally distributed random number in the range $[0, 1]$. 

$$
\mathbf{M} = \begin{pmatrix}
  m_{1,1} & m_{1,2} & \cdots & m_{1,N} \\
  m_{2,1} & m_{2,2} & \cdots & m_{2,N} \\
  \vdots & \vdots & \ddots & \vdots \\
  m_{L,1} & m_{L,2} & \cdots & m_{L,N}
\end{pmatrix} = (m_{k,j})
$$

(2)
3. Validation

Two types of simulated data sets (type A and B) were created and used for the validation of the algorithm. The first type (A) of data set was constructed to evaluate the ability of the AMPD algorithm to detect peaks in noisy periodic signals. Therefore 5 signals (sampling frequency: 100 Hz, length: 20 s, frequency of the sinus oscillation: 1 Hz) with different amount of Gaussian white noise were used. The signal-to-noise ratios (SNR) of the final signals were 25 dB, 10 dB, 5 dB and 0 dB. Additionally one signal without noise was created. The SNR values were calculated according to the
formula SNR = 10log(\(P_{\text{signal}} / P_{\text{noise}}\)), with \(P_{\text{signal}}\) the power of the periodic signal and \(P_{\text{noise}}\) the power of the noise. The second type (B) of data set was constructed to evaluate the ability of the AMPD algorithm to detect peaks in noisy quasi-periodic signals. Therefore, five chirp signals were created, each comprising a sinus oscillation with an increasing frequency (start frequency: 1 Hz, end frequency: 3.8 Hz). As for the type A signals, different amounts of noise were added to the signals leading to the same SNR-values as for the type A signals.

The AMPD algorithm was implemented in Matlab (The Mathworks, Natick, MA, USA). The results of applying the algorithm to the simulated data sets are visualised in Figures 3 and 4. All peaks were correctly detected (based on visual inspection) for all types of signal characteristics used.

**Figure 3.** Results of applying the AMPD algorithm to the simulated type A data sets. (a) no noise; (b) SNR = 25 dB; (c) SNR = 10 dB; (d) SNR = 5 dB; (e) SNR = 0 dB.
Figure 4. Results of applying the AMPD algorithm to the simulated type B data sets. 
(a) no noise; (b) SNR = 25 dB; (c) SNR = 10 dB; (d) SNR = 5 dB; (e) SNR = 0 dB.

4. Application Examples

In this section, we present a demonstration of the proposed algorithm on five different time series originating from different research areas (i.e., astrophysics, biomedical optics, physiology, geophysics and chaos theory) in the field of natural sciences.

4.1. Sunspot Numbers

The time series analyzed comprises the total monthly number of sunspots, i.e., the number of dark spots on the solar surface caused by interactions between the solar magnetic field and the Sun’s plasma [42]. The time series comprises 3151 values and spans a period from January 1749 to July 2011. The data was downloaded from the website of the Solar Influences Data Analysis Center (SIDC) of the Royal Observatory of Belgium (SIDC-team) [43]. The results of the analysis are given in Figure 5. All 23 peaks were correctly detected (no false negative errors) and no false detection occurred (no false positive errors).
Figure 5. Results of applying the AMPD algorithm to a time series comprising the monthly total number of sunspots.

4.2. Blood Volume Pulse in fNIRS Signals

Signals measured using functional near-infrared spectroscopy (fNIRS) contain a strong oscillation that is caused by the blood volume pulse (BVP) related to the arterial blood pressure changes (systole, diastole). fNIRS is a technique allowing non-invasive measurement of changes in the blood concentrations of oxyhemoglobin ([O$_2$Hb]) and deoxyhemoglobin ([HHb]) by irradiating light in the near-infrared spectrum (600–950 nm) into the brain and analyzing the reflected light intensity changes [44–46]. The information of the BVP in fNIRS signals is relevant for various reasons, i.e., for a proper removal of the BVP from fNIRS signals to improve the detection of fast neuronal optical signals [47,48], or for the usage of a surrogate to the heart rate variability measured with electrocardiography [49].

In order to demonstrate the applicability of the AMPD algorithm to detect the BVP-peaks in a fNIRS time series, we used a self-recorded [O$_2$Hb] signal measured on the prefrontal cortex (sampling frequency: 10 Hz, source-detector distance: 2.5 mm, device: MCP-II, multichannel continuous-wave NIR-spectroscope, for technical details see [50]). The time series analyzed had a duration of 200 s (~3.3 min). The results of the analysis are given in Figure 6. All 225 peaks were correctly detected (no false negative errors) and no false detection occurred (no false positive errors).
4.3. Maximum Concentration of Expired CO₂

The CO₂ concentration of the exhaled air, measured with capnography, is an important parameter in medicine and basic physiological research. The waveform of CO₂ concentration changes during the respiratory cycle comprises four phases (I: inspiratory baseline, II: expiratory upstroke, III: expiratory plateau, and IV: inspiratory downstroke) where the phase IV marks the peak of the whole waveform. This peak value is called end-tidal CO₂ (P_{ET}CO₂) and is a best approximation of the alveolar CO₂ level, i.e., the partial pressure of CO₂ in the arterial blood (PaCO₂) [51]. A precise and robust measurement of the phase IV peak is indispensable to correctly determine the P_{ET}CO₂ values.

The time series analyzed comprised a 500 s long CO₂ time series measured on a resting subject (sampling frequency: 5 Hz, device: Nellcor N1000 gas analyzer, Nellcor. Inc, Hayward, USA). The results of the analysis are given in Figure 7. All 133 peaks were correctly detected (no false negative errors) and no false detection occurred (no false positive errors).
4.4. QRS Peaks in ECG Signals

Electrocardiography (ECG) time series contain information about the direction and magnitude of the electrical activity caused by electrophysiological effects (depolarisation and repolarisation) of the atria and ventricles. The time series comprises a recurrent sequence of characteristic signal waveforms (P, QRS, T) where the QRS complex is the most striking element. The maximum of the QRS complex is the R-peak. The sequence of RR intervals is the heart rate. The variability of the RR interval lengths is called heart rate variability (HRV), associated with the state of the autonomic nervous system [52].

To demonstrate the usefulness of the proposed AMPD algorithm, we used a self-recorded ECG signal (sampling frequency: 128 Hz; device: 1-channel ECG, MK3-ETA, TOM-Medical, Graz, Austria). The analyzed time series had a duration of 200 s (~3.3 min). The results of the analysis are given in Figure 8. All 221 R-peaks were correctly detected (no false negative errors) and no false detection occurred (no false positive errors).
4.5. Peaks in the Variation of the Earth’s Length of Day due to Lunar Tidal Forcing

Since the Earth’s rotation rate is not a constant, each day has a different duration (length of day, LOD) when comparing the astronomically determined duration of a day on Earth with the duration of a day according to the SI units definition (1 day = 86,400 s) measured using atomic clocks [53]. The variations in LOD are highly non-stationary and comprise different oscillations. One of the strongest oscillations present in the LOD time series has the average period of 13.6 days. This oscillation is the result of a lunar tidal force that the Moon exerts on the Earth’s rotation due to changes in the lunar declination during the lunar cycle [54].

The proposed AMPD algorithm was tested to establish its ability to detect the peaks of this lunar-related oscillation in the unfiltered (raw) LOD data. To this end, (i) daily averaged LOD data were obtained from the International Earth Rotation and Reference Service (IERS) [55]; and (ii) the peaks of the LOD time series from 2008 to 2011 (1461 days) were detected. The results are given in Figure 9. In total 105 peaks were detected. One peak was not detected (one false negative error) but no false detection occurred (no false positive errors). As shown in Figure 9, the missed peak has low amplitude and is superimposed with a rapidly decreasing part of a slower oscillation (low frequency noise) which makes the detection difficult. The missed peak detection could be prevented by high-pass filtering the signal prior to analysis with the AMPD algorithm.
The average duration of the intervals between the peaks was $13.7885 \pm 2.1974$ days, which corresponds well with the value reported in literature [54,56].

**Figure 9.** Results of applying the AMPD algorithm to a LOD time series.

4.6. **Chaos Theory: Lorenz System**

The three-dimensional Lorenz system, developed by Edward N. Lorenz in 1963 [57], is a simplification of the seven-dimensional system developed by Barry Salzman in 1962 [58]. The Lorenz system is a classic example of a nonlinear system exhibiting chaotic behaviour. It is given by three ordinary differential equations (Lorenz equations):

$$
\frac{dx}{dt} = \sigma(y - x), \quad \frac{dy}{dt} = x(\rho - z) - y, \quad \frac{dz}{dt} = xy - \beta z
$$

with the system parameters $\sigma, \rho$ and $\beta$ that determine the dynamic behaviour of the system.

To demonstrate the capability of the proposed AMPD algorithm, (i) the system of ordinary differential equations given by (5) was solved for the parameter values $\sigma = 10$, $\rho = 28$, $\beta = 8/3$, and $t \in [1,3000]$; and (ii) the peaks of the time series $x(t)$ were detected. The results are given in Figure 10. All 53 peaks were correctly detected (no false negative errors) and no false detection occurred (no false positive errors).
5. Discussion and Conclusions

In this paper, we presented a new method, called automatic multiscale-based peak detection (AMPD), for automatic detection of peaks in noisy periodic and quasi-periodic signals. Compared to traditional methods for peak detection, the proposed AMPD algorithm has the following advantages: (i) no free parameter values have to be chosen by the user prior to analysis; (ii) peaks in periodic and quasi-periodic signals can be detected; and (iii) the peak detection efficiency is fairly robust against high and low frequency noise. Hence, the presented method results in a high efficiency in detection of peaks.

Using simulated and real-world signals, the usefulness of the AMPD algorithm was evaluated and demonstrated. For signals with high amplitude low frequency noise, the performance of the algorithm can be optimized by high-pass filtering the signal prior to analysis. However, the high-pass filtering is not necessary for most of the signals as demonstrated.

For future work it would be interesting (i) to further explore the information given in the LMS in order to find, for example, additional rules for the automatic identification of peaks; (ii) to evaluate the performance of the AMPD algorithm compared to other peak detection algorithms available; and (iii) to use and test the algorithm in real world applications.

Acknowledgments

The authors thank Rachel Folkes for assisting with proofreading and Patrick Wespi for his helpful comments on the draft.
References


© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).