On the Filtering of Photoplethysmography Signals

Adewale Emmanuel Awodeyi, Stephen R. Alty and Mohammad Ghavami
Faculty of Engineering, Science and the Built Environment (FESBE)
Department of Engineering and Design
London South Bank University
London, United Kingdom
{awodeyia,steve.alty,ghavami@lsbu.ac.uk}

Abstract—Recently there has been renewed interest in the application of photoplethysmography signals for cardiovascular disease assessment. Photoplethysmography signals are acquired non-invasively using visible and infrared light passed through the finger pulp. Unfortunately, this method commonly suffers from many forms of interference and distortion such as; baseline wander, mains-line interference and random spikes or other such artifacts. This paper presents a new approach for effective filtering of the photoplethysmography signal. Specifically, a cascaded filtering method for removing the artifacts from photoplethysmography signals based on the median and polynomial filters (MdPF) is proposed. Recordings from the PhysioNet database are used to validate the proposed method. Our experimental results show that the performance of MdPF cascaded filtering method is more effective than other current methods alone in removing artifacts from photoplethysmography signals. Root mean square error measurements are used for comparison purposes. This paper follows from previous work on median based method for baseline wander removal in photoplethysmogram signals.

Keywords—*photoplethysmography, cardiovascular disease, data acquisition, pre-processing*

I. INTRODUCTION

It has been shown recently that the photoplethysmography (PPG) signal [1] is used to obtain clinical information for the early diagnosis of cardiovascular disease (CVD). However, like most physiological signals, PPG waveforms are difficult to acquire without encountering some forms of interference and distortion. Preprocessing of physiological data is an important part of this research since the noise and artifact contamination of the PPG signal can affect the efficient interpretation of clinical information and early diagnosis of disease. The main challenges in processing the PPG signals are described as follows [2]: power line interference is clearly displayed within the frequency domain as a spike at a frequency of 50Hz or (60Hz in USA) caused by mains power sources and can be eliminated by one-stage median filter (1SMdF). The power-line interference noise is generated by using a cosine wave as displayed in Fig. 1(a). The frequency of cosine wave is 50Hz whereas the corresponding amplitude is 170mV. The patient - sensor motion also induced by muscular noise is usually due to the movement of the patient or sensor and can be eliminated by ensuring that the patient lies relaxed and the sensors are properly connected. Baseline wander artifacts usually originate from respiration at frequencies varied between 0.15 and 0.5Hz [3] and can be suppressed by using two-stage median filtering (2SMdF) [4]. An artificial baseline wander is generated by using a combination of sine and cosine waves as displayed in

Fig. 1(b). The frequencies of the sine and cosine waves are 0.4Hz and 0.2Hz respectively, whereas their corresponding amplitudes are 120mV and 240mV.
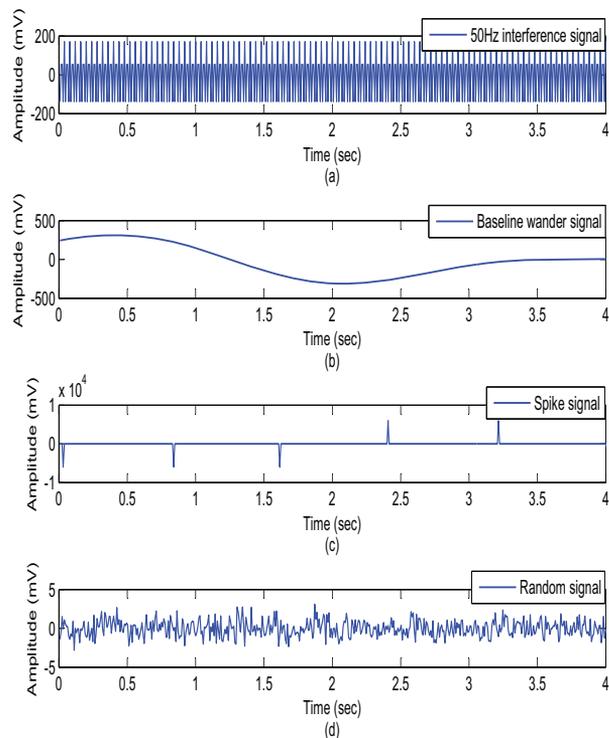

Fig. 1: Signal types (a) 50Hz interference signal (b) Baseline wander signal (c) Spike signal (d) Random signal.

Wide band noise (spikes and low-frequency noise) could also be caused by complex stochastic processes within a wide band signal. This artifact could be due to instrumentation amplifiers, the recording system picking up ambient electro-magnetic interference and can be removed by one-stage median and polynomial filters. Therefore, the spikes and random noise are common forms of interference to photoplethysmography signal as displayed in Fig. 1(c) and Fig. 1(d) respectively. The spikes are generated by adding transients with random signs at random points. We created random noise using Matlab "randn" function with zero mean and standard deviation of variable values. In this experiment, the sampling frequency (sampling rate) is 125Hz.
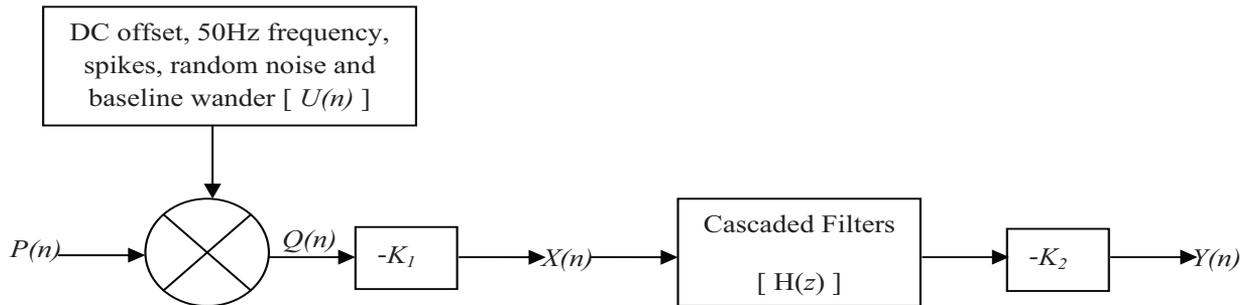
Fig. 2: PPG signal pre-processing.

II. DATA COLLECTION AND PRE-PROCESSING OF PPG SIGNALS

At the initial stage, PPG data were extracted online from PhysioNet MIMIC II databases [5] which were imported into Matlab for additional processing as expressed in (1,2,3). The length of the extracted PPG signals was 10 seconds and originally sampled; however the time was arbitrarily scaled. In order to cascade each stage in Fig. 2, the best - quality simulation output of the primary stage becomes the input of the next stage, and so on as described in Fig. 3. On the first stage, the one-stage median filter was applied (1SMdF). The second stage is a polynomial-based smoothing filter; which we will refer to as poly. In the final stage, we applied the two-stage median filters (2SMdF) [3] and then mean removal. As a comparison with the Matlab filtering method, median filter (Medfilt1) is used in the first stage. Where high pass window-based finite impulse response filter method (Fir1) or Wavelet denoising based on symmetric wavelets was in the second stage, we applied two-stages, moving average filter (2SMaF) [3] in final stage and then mean removal. As expressed in (4), the cascaded transfer function $H(z)$ is written as a product of each stage transferred functions.
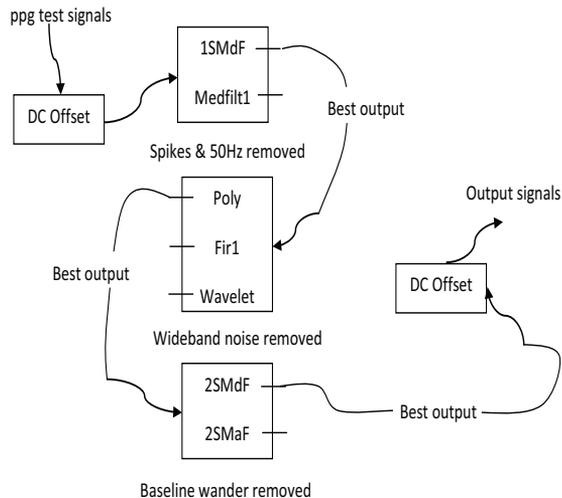

Fig. 3: Block diagram of the cascaded filters algorithm for pre-processing of PPG signals.

The ordering of the individual $H_i(z)s$ will lead to a system with completely different numerical characteristics. To determine, which give the best numerical characteristics, it may be effective to try various orderings for a given set of $H_i(z)s$ where $i = 1, 2, 3$.

$$Q(n) = P(n) + U(n) \tag{1}$$

$$K_1 = \frac{1}{N} \sum_{n=1}^{N} Q_n \tag{2}$$

$$X(n) = Q(n) - K_1 \tag{3}$$

$$H(z) = \frac{Y(n)}{X(n)} = H_1(z).H_2(z).H_3(z) \tag{4}$$

As shown in Fig. 2, where $P(n)$ is a clean PPG original signal contained in P_1 through P_N, and $(P + U)n$ is a corrupted PPG signal. Additionally, $X(n)$ is the output after mean removal (DC offset) and so input to cascaded filters whereas n is an index that runs through these values. In cascaded filters, the structure of 1SMdF can be represented as,

$$y(n_1) = \{ \quad Median(X(:, n), 2) \qquad X_1(:, j)$$

$$f(X_1(:, j)) = \{ \quad y(n_1) \qquad \forall j \in \{1, ...N\} \tag{5}$$

Where $n = max(j - floor(L/2), 1) : min(j + floor(L/2), N)$, X is a noisy PPG data. Additionally, L is averaging window length in samples, N is a dimensional size. Meanwhile, $X_1(:, j)$ is the output after spikes and 50Hz removal whereas j is an index that runs through these values. In cascaded filters, the structure of smoothing algorithm can be represented as,

$$f(y(n_2)_s) = \frac{\sum\limits_{i=n}^{n} A_i y(n_2)_{s+i}}{\sum\limits_{i=n}^{n} A_i} \tag{6}$$

This showed that a set of integers $(A[-n], A[-(n-1)]).....$

...., $A[n-1], A[n]$) could be derived and used as weighting coefficients to carry out the smoothing operation. The use of these weighting coefficients known as convolution integers, is the smoothed data point $y(n_2)_s$. In cascaded filters; the structure of 2SMdF can be represented as,

$$y(n_3) = \begin{cases} Median(X(:,n_1),2) & X_1(:,j) \\ Median(X1(:,n_2),2) & X_2(:,j) \end{cases}$$

$$f(X_2(:,j)) = \{ \ y(n_3) \quad \forall j \in \{1,...N\} \tag{7}$$

Where $n_1 = max(j - floor(L1/2),1) : min(j + floor(L1/2),N)$ and $n_2 = max(j - floor(L2/2),1) : min(j + floor(L2/2),N)$, X is a noisy PPG data. Additionally, $L1$ and $L2$ are first and second stages averaging window length in samples, respectively, N is a dimensional size. Finally, $X_2(:,j)$ is the output after baseline wander removal whereas j is an index that runs through these values. Therefore, after mean removal as expressed in (8), $Y(n)$ is a filtered PPG signal as expressed in (9).

$$K_2 = \frac{1}{N} \sum_{j=1}^{N} X_2(:,j) \tag{8}$$

$$Y(n) = X_2(:,j) - K_2 \tag{9}$$

III. METHOD AND EXPERIMENT RESULTS

The PPG data extracted from PhysioNet database [5] is first filtered before yielding a prototype clean signal; we refer to as a PPG original signal in Fig. 4(a). Typically, data can exhibit unwanted noise and distortion that last for a short duration of time. Thereby, we have corrupted the PPG original signal with measured amounts of DC offset, 50Hz frequency, baseline wander, random noise and several large spikes over the length and then mean removal as expressed in (3) to generate the PPG test signal in Fig. 4(c). The PPG test signal is verified by applying three stages of filtration to extract a suitably clean signal. Individual experiments on each stage are carried out separately as described in Fig. 5, 6 and 7. As expressed in (10), root mean square error (RMSE) is used in order to quantify the noise level. The lower RMSE value, the better quality of the signal. RMSE can be defined as the square root of the mean of the square error. In addition, we evaluate performance results using the Fast Fourier Transform (FFT) for comparison with RMSE. The statistical analysis and comparison results are shown in Table 1 and Fig. 8, respectively.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{10}$$

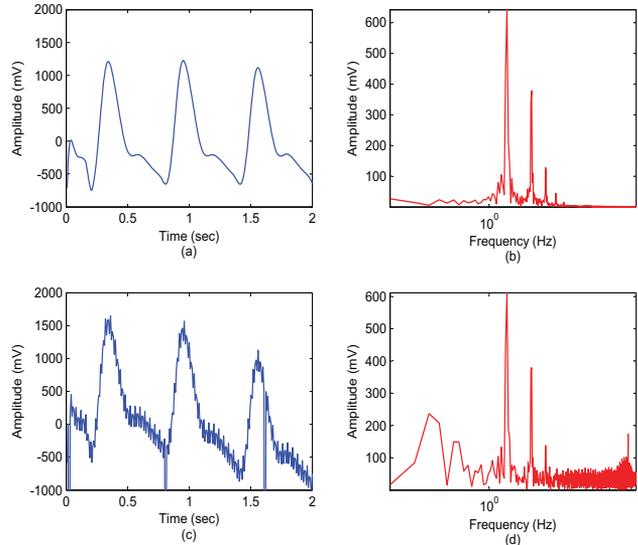Where y_i is the value actually observed, $\hat{y}_i$ is the value predicted.

Fig. 4: Signal types (a) PPG original signal (b) FFT PPG original signal (a) PPG test signal (b) FFT PPG test signal.
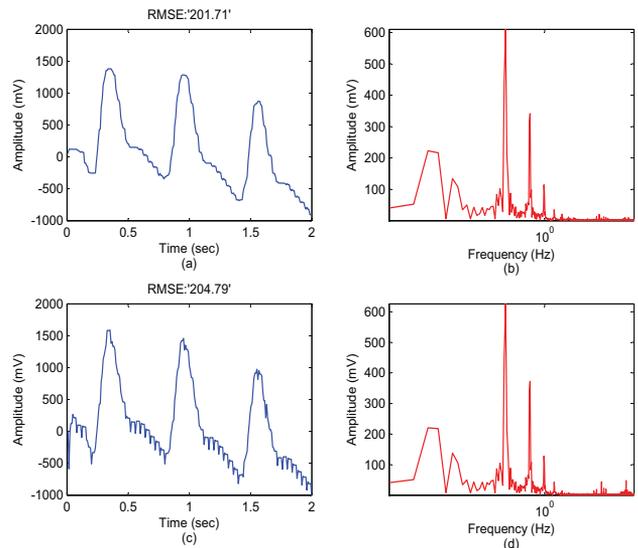
A. First Stage of Cascaded Filters

Fig. 5: Simulation results (a) 1SMdF (b) FFT 1SMdF (c) Medfilt1 (d) FFT Medfilt1.

TABLE 1: Statistical analysis

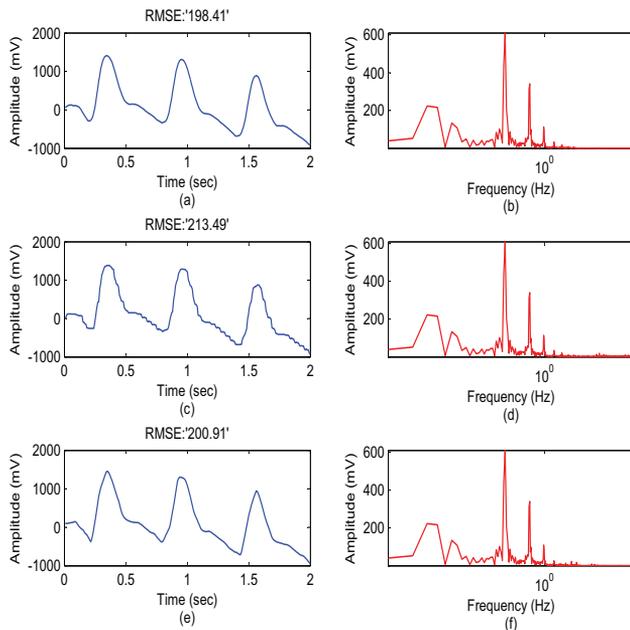| Parameter	Result obtained from					
	Spikes and 50Hz		Wideband Noise		Baseline Wander	
	1SMdF	Medfilt1	Poly	Wavelet	2SMdF	2SMaF
RMSE	201.71	204.79	198.41	200.91	65.20	67.69

B. Second Stage of Cascaded Filters

Fig. 6: Simulation results (a) Poly (b) FFT Poly (c) Fir1 (d) FFT Fir1 (e) Wavelet (f) FFT Wavelet.

A. Final Stage of Cascaded Filters

Fig. 7: Simulation results (a) 2SMdF (b) FFT 2SMdF (c) 2SMaF(d) FFT 2TSMaF.

IV. CONCLUSION

This paper shows the advantages of cascaded filters that can be implemented in various ways. In this paper, the simulation results, FFT and RMSE measures indicate clearly that our cascaded filtering can give superior performance on the removal of artifacts in PPG signals. By cascading filters with different responses, it is possible to allow for various interference signals. We have evaluated our cascade with 10 extracted PPG data sequences drawn from the PhysioNet database and have obtained positive results in each case. To further increase the performance, it may be effective to examine different ordering of four filters. Future work will involve outlier removal and feature extraction from the PPG signals and data from various patients will be analyzed for CVD risk.

V. ACKNOWLEDGMENT

Fig. 8: Comparison between MdPF and Matlab built-in filters.

REFERENCES

[1] S. R. Alty, N. Angarita-Jaimes, S. C. Millasseau, P. J. Chowienczyk, "Predicting Arterial Stiffness from the Digital Volume Pulse Waveform," *Biomedical Engineering, IEEE Transactions on.* pp2268 - 2275 Dec, 2007.

[2] M. Elgendi, "On the Analysis of Fingertip Photoplethysmogram Signals," *Current Cardiology Reviews.* pp14 - 22 Feb, 2012.
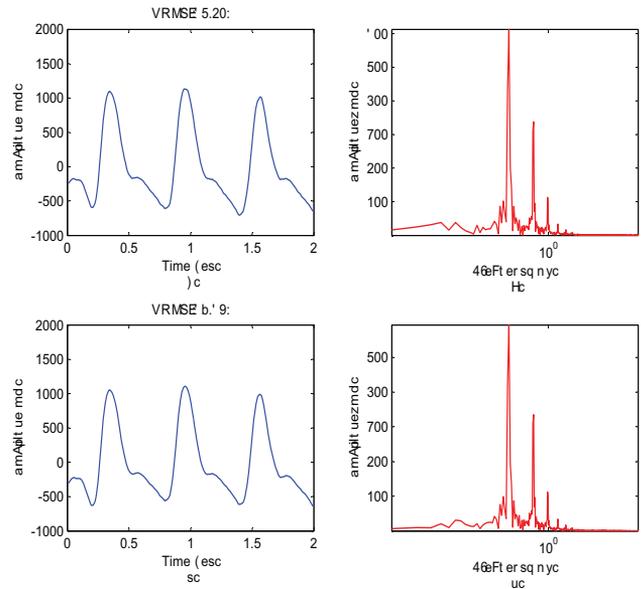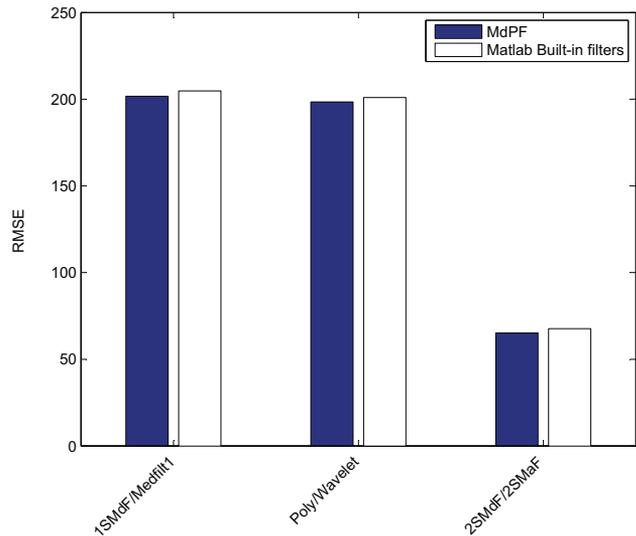
[3] Zahoor-Uddin, "Baseline Wandering Removal from Human Electrocardiogram Signal using Projection Pursuit Gradient Ascent Algorithm," *International Journal of Electrical and Computer Sciences IJECS/IJENS.* Vol: 9 No: 9, pp 11-13, 2011.

[4] A. E. Awodeyi, S. R. Alty and M. Ghavami, "Median Filter Approach for Removal of Baseline Wander in Photoplethysmography Signals," *In proceddings of European Modelling Symposium.* Manchester: pp261 - 264, November, 2013.

[5] PhysioNet Database. Available At *http://physionet.org/cgi-bin/atm/ATM* [accessed 02/06/13].