Efficient Respiratory Rate Extraction on a Smartwatch

David Pollreisz and Nima TaheriNejad

***Abstract*— Using the Photoplethysmogram (PPG) sensor of a smartwatch to extract Respiratory Rate (RR) is very attractive. However, existing algorithms suffer from lack of accuracy and susceptibility to noise and movement artifacts. To tackle this issue, we propose performing Frequency Domain Peak (FDP) analysis using the Frequency Modulation (FM) feature. Moreover, our analysis of existing methods show that in contrast to the common practice Smart Fusion (SFU), despite incurring extra computational costs, is very little helpful. It is hence more preferable and efficient to avoid SFU. The proposed method shows an improvement of at least 130% in the Figure of Merit (FoM) and has more than 60% smaller mean error. Therefore, it can be reliably used in a wide range of applications.**

I. Introduction

Wearable devices are used in many applications such as well-being and sport activities [1], [2], physical health [3]–[6], and mental health [1], [7], [8]. However, they face several challenges, such as constrained resources [9], changes in the environmental conditions [9]–[11], and excessive noise and artifacts introduced by users' activities [12]. Typically extra processing of the signals and more complex algorithms are adopted to tackle these challenges, however, those require additional resources which are often scarce on these devices [4], [6], [9]. Here, we propose a new algorithm for Respiratory Rate (RR) extraction, which is more reliable in face of movement artifacts, while using less steps than the State-of-the-Art (SoA).

RR is an important physiological measure used in various medical studies [13]. However, measuring RR directly, e.g., using a mouth piece, is rather impractical in daily setups since intervenes with their activities, is uncomfortable, and comes with social stigmas. On the other hand, we know that breathing influences the cardiac system [14]. These influences, visualized in Figure 1, are amplitude and frequency modulation as well as wandering of baseline. This enables us to extract RR using Photoplethysmogram (PPG) signals of wearable devices such as smartwatches. One of the main challenges of this method is the movement artifact contamination [9], [12]. Most methods for alleviating the effect of movement artifact use extra sensors, computations and power to perform extra complex calculations, an overview of which can be found in [12]. In contrast, here, we present a method that does not need any extra sensors. Moreover, by a closer look into performance factors, we show that the Smart Fusion (SFU) step -which is common in the literature- is not needed. Our analysis show that while incurring a larger amount of computation overhead, SFU provides insignificant benefits in the proposed algorithm. This makes it an even simpler and more efficient algorithm. More importantly, without any extra sensors it reduces the error to a range that is acceptable for many applications such as Early Warning Score (EWS) [10], [11] The proposed algorithm, Frequency Domain Peak (FDP) using Frequency Modulation (FM) only, is considerably more robust against movement artifacts; i.e., compared to SoA algorithms it shows an improvement of 60-72% and 130-174% in mean error and Figure of Merit (FoM), respectively. This indicates the improved reliability of the RR extraction on a smartwatch using the proposed algorithm. Further, by skipping Amplitude Modulation (AM) and Baseline Wanderer (BW) feature extraction and their fusion, it reduces necessary computations and increases efficiency.

Authors are with the Institute of Computer Technology, TU Wien, Vienna, Austria. E-mail: `nima.taherinejad@tuwien.ac.at`

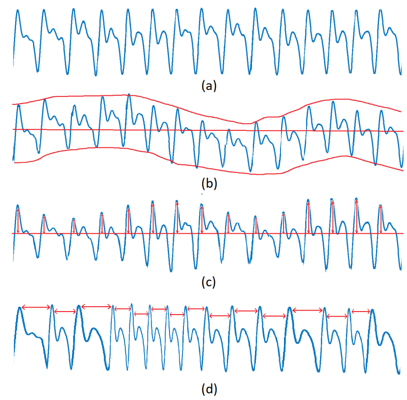

Fig. 1. Modulations caused by respiration: (a) none (b) Baseline Wanderer (BW) (c) Amplitude Modulation (AM) (d) Frequency Modulation (FM).

II. Proposed Method

Figure 2 depicts an overview of our RR extraction method, which we describe in the remainder of this section.

1) Pre-Processing Step: Before extracting features of the signal, the raw data needs to be prepared. This preparation includes band-pass filtering of the signal. This eliminates the offset and any noise that lays outside the range of interest and helps in extracting the location of local maxima and minima of the Blood Volume Pulse (BVP) signal. A finite impulse response filter [15], e.g., Butterworth, leads to sharper peaks and easier detection of the maxima of the BVP signal. The order of the filter, N, is determined using $N = 2f_s/25$,

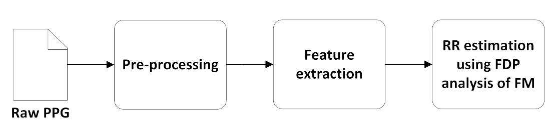

Fig. 2. Flow chart of the implemented RR extraction algorithm.

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

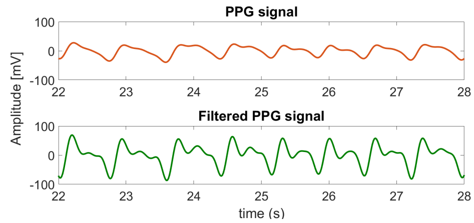

Fig. 3. (a) The original PPG signal and (b) the filtered PPG signal.

where f_s is the sampling rate. The filter coefficients [15] are

$$b_k = \begin{cases} -1 & for \quad k = 0, ... \frac{N}{2} - 1 \\ 1 & for \quad k = \frac{N}{2}, ..., N - 1 \end{cases} \tag{1}$$

Figure 3 shows the visible improvement on a sample PPG signal before and after application of the filter. Out of the filtered signal, the local extrema are searched using three criteria: (i) Extrema are recognized as such only if they are bigger than the mean value of the signal, (ii) only extrema that are $0.4 f_s$ apart are considered, (iii) a trough must be surrounded by two peaks and the other way around. Next, only the relevant peaks and troughs remain and the extraction of the features can start.

2) FM Feature Extraction: The proposed method relies only on the FM feature of the PPG signal, therefore, in contrast to the literature, we do not need to extract AM and BW. FM, is calculated by subtracting the temporal location of each peak and the the peak after it. At the end it is normalized to the mean of the signal. This gives us a continuous FM signal which we process in the next step.

3) Proposed RR Estimation: For estimation of RR, existing methods [16]–[18] process the AM, FM, and BW features and their properties in the time domain. In our algorithm, FDP-FM, we use only one feature, namely FM, which we do not process in the time domain, as others do, but rather in the frequency domain. Our algorithm searches for the Dominant Frequency (DF) in the extracted signal. First, the signal is detrended. Next, dominant peaks in the range of $0.033-2$Hz (associated to a breath rate of 2 to 120 Breath per Minute (BPM)) are found. The breath rate corresponding to the DF is the estimated RR.

III. Experiments and Results

1) Experimental Data and Setup: The data set[1] is composed of 41 samples taken from four male healthy volunteers, aged between 26 and 29 years, breathing and moving while wearing Empatica E4 smartwatch. The first task was normal breathing in the range of 10 to 15 BPM. The second, fast breathing with a breath rate over 15 and the last, slow breathing with a breath rate below 10. During the 60 seconds of measurement, they moved their arms from the table straight into the air three times and counted their breaths to use as the ground truth. This movement introduces noise, more specifically movement artifacts into the recorded signal. Figure 4 depicts an example of a Blood Volume Pulse (BVP) signal with a noisy portion in the middle introduced due to movement. Table I shows a summary of the used data.

[1]At the time of conducting the research, our institution had no formal research ethics committee or institutional review board to approach for approval of the research. We confirm that we followed the principles of the WMA Declaration of Helsinki (adopted in 1964, most recently amended in 2013); the participants were healthy volunteers who gave their written informed consent.

TABLE I
Distribution of the used experimental data

	Normal	Fast	Slow
With Movement	12	4	7
Without Movement	10	4	4

TABLE II
Summary of the results of the proposed method.

Window Overlap	Mean|Error| [BPM]	STD	Samples [%]	Window length [s]	FoM
No	4.229	3.996	100	16	12.16
Yes	4.567	4.253	100	20	11.33

The proposed method is implemented in Matlab. For the filter we have used a 4^{th} order high-pass Infinite Impulse Response (IIR) filter, namely Butterworth, with a 0.05 Hz cut-off frequency and a low-pass one with a 5 Hz cut-off frequency. Sampling frequency of the BVP is 64 Hz.

2) Sliding Window: The performance of the system is affected by the processing window size. To find a suitable window size, a parameter sweep was performed, where the window range was swept from 4 to 30 seconds in a step size of 2 seconds. The maximum was set to 30 because the test data are 60 seconds long. If the range exceeds 30 only half of the signal could be analyzed since no second window could be fully formed. That is the reason why the range was not increased further. In addition, the same range was tested with 50% overlapping windows. The best window length was in this manner selected for each algorithm.

3) FoM: The merit of an RR extraction algorithm does not solely depends on its average accuracy, but also factors such as the amount of variation or dispersion in prediction, i.e., Standard Deviation (STD), as well as the percentage of times it is successful in calculating an RR (regardless of its accuracy). To better compare algorithms, we define a FoM, which captures all these effects under one umbrella. That is,

$$FoM = \frac{CSR}{Mean(|Error|) + STD} \tag{2}$$

where error is measured in BPM and CSR stands for the Computed Samples Ratio, that is the number of samples with successful estimation of RR divided by the overall number of samples. Thus, the proposed FoM combines the amount of error and reliability (represented by STD) of RR estimation, with the number of samples it can successfully estimate. The larger FoM, the better the algorithm.

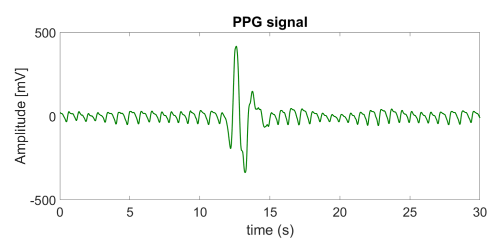

Fig. 4. The effect of movement artifact in BVP.

4) Results of the experiments: The proposed algorithm, showed to be able to successfully estimate all samples in our data set, with a mean error and standard deviation of approximately 4 BPM. This is an acceptable performance for many applications (e.g., EWS [3], [5], [10]), especially given the fact that the data was contaminated with movement artifacts. Moreover, as we can see in Section IV, it has a better FoM than other existing algorithms. A summary of the results that we obtained for the proposed algorithm is inserted in Table II. As we see in this table, the performance of the system with and without windows overlap is very similar. However, the algorithm without overlapping windows gives slightly better results, which makes it overall favorable.

IV. Comparisons

1) Algorithms in the Literature: To have a fair comparison, we implemented other principle algorithms in the literature, namely Time Domain Peak Detection (TDPD), Time Domain - Count Origin (TDCO), and Count Origin - Smart and Time Fusion (COSTF). Moreover, we implemented an algorithm based on the same analysis mode, i.e., FDP, which uses all three features (as opposed to only FM, as used in our proposed algorithm) and their SFU [19].

The first extraction algorithm [16], TDPD, uses the Peak Detection (PD) for estimation of RR. The second and third algorithms, TDCO and COSTF [17], use Count Origin (CO) method to detect peaks and troughs. In this method, a threshold as 0.2 times the 75^{th} percentile of peak values is defined, and any peaks with an amplitude smaller than this threshold is dismissed. A breath is detected as two consecutive peaks separated by only one trough (with an amplitude less than zero). COSTF uses an additional fusion method called Temporal Fusion (TFU) [18]. In Table III, a summary of all analyzed algorithms, including the proposed method (FM) and the proposed algorithm, and their features are shown. The key new feature of the proposed algorithm is using DF and only DF for its estimation.

2) Simulation of Existing Algorithms: We implemented all aforementioned algorithms and ran them on our data in a similar condition. We performed extensive analysis of their performance, when the RR was estimated using one of the features only, as well as when it was obtained using a fusion algorithm. It should be noted that compared to the original implementation, in our implementation of other works, we enhanced TDPD, TDCO, and COSTF by changing the PPG peak detection from the detection of the maximum in the raw signal to finding it after applying a filter. As seen in Figure 3, application of this filter increases the quality of the input PPG signal and helps to achieve a more robust detection. In addition, the SFU algorithm was changed so that it can fuse two estimated values instead of only all three (original case) to increases the percentages of calculated samples.

In the enhanced SFU, first, the STD of the estimated values from each feature (BW, AM and FM) is calculated for each window. If the STD is below 4, the mean value of these values is calculated and is taken as the RR value of the fusion method. On the other hand, if only two estimations have a STD below 4 and this value is lower than the STD of all three estimations, the mean value of these two estimations is calculated and set as the final RR value of the fusion method. If all STDs exceed 4 then the value of the SFU is set to NaN (Not a Number). In addition to SFU, for COSTF a TFU is calculated using [18] $RR_i = 0.2RRest + 0.8RR_{i-1}$, where $RRest$ is the estimated value from the SFU algorithm. If some values are not calculated because the STD is above 4, this fusion algorithm can be used to smoothen the values. In practice TFU works similar to a low pass filter, which reduces errors due to numbers widely out of range [18].

TABLE III
A SUMMARY OF ALL ANALYZED RR EXTRACTION ALGORITHMS.

	Feature Extraction			RR Estimation			Fusion	
	AM	BW	FM	CO	PD	DF	SFU	TFU
TDPD	✓	✓	✓		✓		✓	
TDCO	✓	✓	✓	✓			✓	
COSTF	✓	✓	✓	✓			✓	✓
FDP	✓	✓	✓			✓	✓	
Prop. Alg.			✓			✓		

A summary of the results we obtained for each algorithm, including variations of the proposed method, is inserted in Table IV to Table VI. In Table IV, the STDs for TDPD in the cases of BW and SFU are 0 because only 2.44% of the samples were calculated. That is only one sample and therefore, no STD could be calculated.

3) Comparative Analysis: We have summarized the comparison for the best result of each algorithm and their respective FoM improvements in Table VII. In this table, the improvements of factor F are calculated using $Imp. = \frac{F_L - F_s}{F_L}$, where L indicated the larger number and s the smaller number between the two.

We observe that the proposed system has the best (largest) FoM, that is 12.16, with a mean error of 4.2 BPM and STD of 3.9, while successfully calculating all samples. As we see in Table VII, the mean error of other SoA algorithms are 2.5-3.6 times larger than the proposed method. Similarly, they have 1.5-1.7 times larger STD which represent their lack of reliability compared to the proposed algorithm. In particular, we note that a mean error of 11-15, associated with existing algorithms, is extremely large and being comparable to the actual number of breath per minutes, in most cases, renders it unacceptable. We observe that FoMs reflect these decisive factors too. In summary, as we can see in Table VII, the proposed method (FDP-FM), compared to other three existing methods (TDPD, TDCO, COSTF, and FDP), has a 130-174% better performance (largest FoM)

Among the various modalities of the proposed algorithm, the best combination –as seen in Table VI– is the FM, without overlapping windows. In this method, all samples were successfully calculated, with an average error of only 4.2 BPM, and a STD of 3.99. Almost as good as that method is AM without overlapping windows, where the FoM is 11.68 mainly due to the somewhat higher STD. The best window length is 16 seconds for both combinations. Lastly, we notice that the smallest mean error and STD (with a small margin) belongs to SFU of the proposed method, which ranks

TABLE IV

STATISTICAL RESULTS OF TDPD.

Estimation Method	Mean|Error| [BPM]	STD	Samples [%]	Window Length [s]	Window Overlap	FoM
AM	15.923	9.032	9.76	8		0.39
BW	12.248	0	2.44	8		0.20
FM	10.738	6.921	92.7	24		5.25
SFU	9.226	0	2.44	12		0.26
AM	16.222	10.001	9.76	8	✓	0.37
BW	17.590	0	2.44	8	✓	0.13
FM	10.656	6.911	92.7	24	✓	5.27
SFU	6.877	0	2.44	12	✓	0.35

TABLE V

STATISTICAL RESULTS OF TDCO AND COSTF.

Estimation Method	Mean|Error| [BPM]	STD	Samples [%]	Window length [s]	Window Overlap	FoM
AM	14.618	6.914	100	28		4.64
BW	16.789	6.057	100	30		4.37
FM	13.937	6.450	100	24		4.90
SFU	15.418	6.353	97.6	22		4.48
TFU	16.257	6.753	100	28		4.35
AM	14.589	5.883	100	28	✓	4.88
BW	17.154	5.883	100	22	✓	4.34
FM	13.924	6.523	100	22	✓	4.89
SFU	15.323	6.156	95.1	22	✓	4.42
TFU	15.676	6.877	100	30	✓	4.43

TABLE VI

STATISTICAL RESULTS OF THE FDP AND FDP-FM.

Estimation Method	Mean|Error| [BPM]	STD	Samples [%]	Window length [s]	Window Overlap	FoM
AM	4.226	4.335	100	16		11.68
BW	8.157	5.009	100	12		7.59
FM	**4.229**	**3.996**	**100**	**16**		**12.16**
SFU	3.148	3.214	75.61	28		11.88
AM	4.568	5.524	100	16	✓	9.91
BW	7.838	5.261	100	16	✓	7.63
FM	4.567	4.253	100	20	✓	11.34
SFU	3.351	3.387	75.61	28	✓	11.22

TABLE VII

COMPARISON OF THE BEST PERFORMANCE OF ALL ALGORITHMS.

Algorithm	Mean|Error| [BPM]	Impr.	STD	Samples [%]	FoM	Impr.
TDPD - FM	10.656	60%	6.911	92.7	18.977	130%
TDCO - FM	13.937	70%	6.450	100	20.387	148%
COSTF - SFU	15.323	72%	6.156	95.1	22.431	174%
Proposed - FM	4.229	-	3.996	100	8.225	-

second in terms of FoM. The reason for that is its relatively lower success rate in estimating samples, that is, only 75%. In other words, in every 4 samples, the RR of one cannot be properly estimated, which -as properly reflected in FoM- affects its merit and reliability. Moreover, we need to bear in mind that for SFU, all three features (AM, BW, and FM) need to be extracted and their associated RR estimated and fused, whereas in the proposed algorithm (FDP-FM) only one feature (FM) needs to be extracted and its associated RR estimated. All this extra calculations require hardware and energy resources and to lead to only 1 BPM improvement in the mean error. Hence, the proposed algorithm (FDP-FM), not only has a better FoM, but also is simpler and requires less calculations than FDP-SFU, making it more efficient.

V. CONCLUSION

In this work, we proposed a new RR extraction algorithm, FDP-FM, which uses DF and FM features to estimate the RR. The proposed algorithm showed to be significantly more reliable than existing ones. Our algorithm has an average error of only 4.2 BPM and a STD of 3.9, while successfully calculating all samples. For a better comparison, we introduced a FoM, which combines the mean error, STD, and the percentage of successfully estimated samples. Compared to others, the proposed algorithm shows more than 130% improvement in the FoM. A noteworthy observation of our study is that in the presence of movement artifacts, in almost all existing methods, the estimations based on the FM feature showed to be more reliable than other combinations. In other words, the common practice of using fusion methods, which incur considerable extra calculations costs, seem to bring no significant additional value to the table compared to the ones using the FM feature (especially, considering their additional calculations and complexity costs). This observation invites further investigations.

REFERENCES

[1] P. Parego *et al.* *Wireless Mobile Communication and Healthcare*. Springer International Publishing, 2018.

[2] H. Yin *et al.* Smart healthcare. *Foundations and Trends® in Electronic Design Automation*, 2(14):401–466, 2018.

[3] M. Götzinger *et al.* Enhancing the early warning score system using data confidence. In *International Conference on Wireless Mobile Communication and Healthcare*, pp. 91–99. Springer, 2016.

[4] F. Forooghifar *et al.* Self-aware wearable systems in epileptic seizure detection. In *Euromicro Conference on DSD*, pp. 7, 2018.

[5] A. Anzanpour *et al.* Self-awareness in remote health monitoring systems using wearable electronics. In *Proceedings of Design and Test Europe Conference (DATE)*, Lausanne, Switzerland, March 2017.

[6] F. Forooghifar *et al.* Resource-aware distributed epilepsy monitoring using self-awareness from edge to cloud. *IEEE Transactions on Biomedical Circuits and Systems*, 13(6):1338–1350, Dec 2019.

[7] N. TaheriNejad and D. Pollreisz. Assessment of physiological signals during happiness, sadness, pain or anger. In *6th MobiHealth*, 2016.

[8] D. Pollreisz and N. TaheriNejad. A simple algorithm for emotion recognition, using physiological signals of a smart watch. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2353–2356, July 2017.

[9] N. TaheriNejad. Wearable medical devices: Challenges and self-aware solutions. In *IEEE Life Sciences*, pp. 5–6, 2019.

[10] M. Götzinger *et al.* Enhancing the self-aware early warning score system through fuzzified data reliability assessment. In *International Conference on Wireless Mobile Communication and Healthcare*. Springer, 2017.

[11] M. Götzinger *et al.* IoT-based early warning score system augmented with fuzzified confidence assessment. *ACM/Springer Mobile Networks and Applications*, pp. 1–18, 2019.

[12] D. Pollreisz and N. Taherinejad. Detection and removal of motion artifacts in PPG signals. *Mobile Networks and Applications*, Aug 2019.

[13] P. Charlton *et al.* Extraction of respiratory signals from the electrocardiogram and photoplethysmogram: Technical and physiological determinants. *Physiological Measurement*, 38(5):669–690, 3 2017. Creative Commons Attribution license.

[14] R. A. Cernat *et al.* Real-time extraction of the respiratory rate from photoplethysmographic signals using wearable devices. *European Conference on Ambient Intelligence*, 2014.

[15] C. C. Mayer. Pulse contour analysis model for non-invasive determination of hemodynamical parameters. *Phd Thesis, TU Wien*, 2007.

[16] A. Schaefer and K. W. Kratky. Estimation of breathing rate from respiratory sinus arrhythmia: Comparison of various methods. *Annals of Biomedical Engineering*, 36, 03 2008.

[17] S. A. Shah. Vital sign monitoring and data fusion for paediatric triage. *Phd Thesis, Oxford University, UK.*, 2012.

[18] J. Lazaro *et al.* Deriving respiration from photoplethysmographic pulse width. *Medical and Biological Engineering and Computing*, vol. 51:pp.233–242, 02 2013.

[19] D. Pollreisz and N. TaheriNejad. Reliable respiratory rate extraction using PPG. In *11th IEEE Latin American Symposium on Circuits and Systems – LASCAS*, pp. 1–4, San José, Costa Rica, February 2020.