

The Quest for Raw Signals: A Quality Review of Publicly Available Photoplethysmography Datasets

Florian Wolling

Ubiquitous Computing, University of Siegen
Siegen, Germany
florian.wolling@uni-siegen.de

Kristof Van Laerhoven

Ubiquitous Computing, University of Siegen
Siegen, Germany
kvl@eti.uni-siegen.de

ABSTRACT

Photoplethysmography is an optical measurement principle which is present in most modern wearable devices such as fitness trackers and smartwatches. As the analysis of physiological signals requires reliable but energy-efficient algorithms, suitable datasets are essential for their development, evaluation, and benchmark. A broad variety of clinical datasets is available with recordings from medical pulse oximeters which traditionally apply transmission mode photoplethysmography at the fingertip or earlobe. However, only few publicly available datasets utilize recent reflective mode sensors which are typically worn at the wrist and whose signals show different characteristics. Moreover, the recordings are often advertised as raw, but then turn out to be preprocessed and filtered while the applied parameters are not stated. In this way, the heart rate and its variability can be extracted, but interesting secondary information from the non-stationary signal is often lost. Consequently, the test of novel signal processing approaches for wearable devices usually implies the gathering of own or the use of inappropriate data.

In this paper, we present a multi-varied method to analyze the suitability and applicability of presumably raw photoplethysmography signals. We present an analytical tool which applies 7 decision metrics to characterize 10 publicly available datasets with a focus on less or ideally unfiltered, raw signals. Besides the review, we finally provide a guideline for future datasets, to suit to and to be applicable in digital signal processing, to support the development and evaluation of algorithms for resource-limited wearable devices.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile devices**; • **Hardware** → *Digital signal processing*; • **Applied computing** → *Health informatics*.

KEYWORDS

digital signal processing, raw, reflective, photoplethysmography

ACM Reference Format:

Florian Wolling and Kristof Van Laerhoven. 2020. The Quest for Raw Signals: A Quality Review of Publicly Available Photoplethysmography Datasets. In *Proceedings of the Third Workshop on Data Acquisition To Analysis (DATA '20)*, November 16–19, 2020, Virtual Event, Japan. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3419016.3431485>

1 INTRODUCTION AND MOTIVATION

The optical measurement principle photoplethysmography (PPG) has been present as pulse oximetry in clinical studies for decades to measure the peripheral oxygen saturation (SpO_2) and the heart rate (HR), typically with a clip attached to the fingertip or earlobe. In recent years, the technique has undergone a revival in modern wearable devices, such as wrist-worn fitness trackers and smartwatches. In contrast to their medical counterparts, these devices provide tight resources and the implemented algorithms have to be optimized for limited processing power, memory, and battery capacity. The loose attachment and the deployment in real life also causes issues as the wearer's motion results in device displacement and soft tissue deformation which are perceivable as signal artifacts.

The development of efficient algorithms, or even the application of machine learning in the near future, requires large datasets of numerous individuals with different constitutions. However, obtaining recordings from long-term deployment is difficult and hence most publicly available databases, such as MIMIC-II / -III [18, 28] of the notable platform PhysioNet [14], are originated in medical studies with a clinical background and stationary devices. Although these datasets usually provide PPG measurements, the compiled findings cannot be directly transferred to wearable devices and their challenges emerging when worn in everyday life.

In clinical settings, often the standard pulse oximeters are applied while commercial wearable devices are common for in-the-wild studies. For both kind of devices the *raw* signal, directly obtained from the analog front-end, is usually not accessible. The devices apply filters to remove the predominant DC component, low-frequency baseline wandering, and high-frequency noise to obtain a detrended and smooth signal. Due to their limited memory, especially wearable devices condense the signal down to its required essence, e.g. measures of peripheral oxygen saturation (SpO_2), heart rate (HR), and heart rate variability (HRV). Only devices intended for the use in research, such as the popular Empatica E4 [15], provide pretended *raw* data which, however, are still preprocessed to ease the interpretation and to save valuable memory. Hence, the signals are already filtered, often rescaled or normalized, and flipped to be consistent with the associated arterial blood pressure (ABP).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

DATA '20, November 16–19, 2020, Virtual Event, Japan

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-8136-9/20/11...\$15.00

<https://doi.org/10.1145/3419016.3431485>

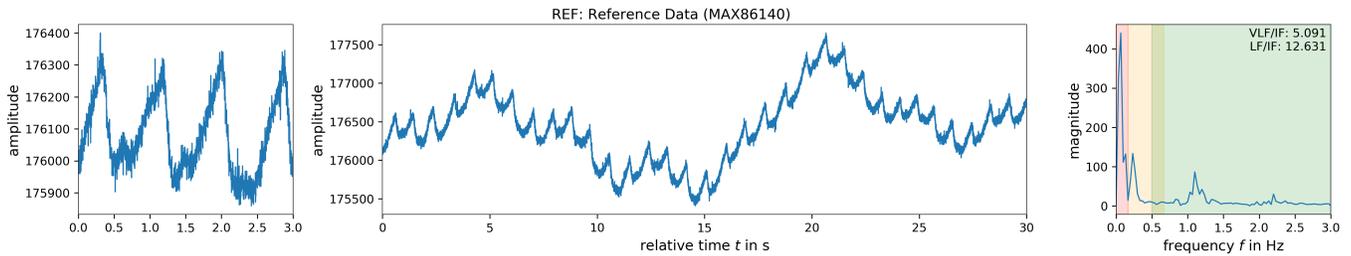


Figure 1: Excerpt from the raw reference PPG signal: green light’s channel of Maxim MAX86140EVSYS#. Short close-up (left), 30-second window (middle), respective frequency spectrum (right). Pulsatile heartbeat signal is inversely proportional to the blood volume, superimposed by low-frequency baseline wandering and high-frequency noise. Note also the large DC offset.

In contrast, researchers developing innovative sensing concepts and hardware, primarily test their proper working, but usually do not focus on the recording of long-term measurements under different conditions. Typically, only a few individuals are testing the prototypes in a lab and rarely in a real-life setting. The recordings are exclusively evaluated and presented as consolidated results, but the gathered data are then not made publicly available.

Hence, in research focusing on algorithms that are applied close to the hardware level, either available datasets from a clinical origin or self-recorded datasets with a limited evidence and reproducibility have to be taken. As stated by Charlton et al. [5], there are only few studies applying their approach to more than two datasets or even comparing multiple approaches to each other, tested on the same, larger benchmark dataset, but their comparison is important as the "performance may differ significantly between datasets". Also Reiss et al. [26] describe the problem that "existing approaches are highly parametrised and optimised for specific scenarios of small, public datasets". Pimentel et al. [25] also emphasize that "Future studies should concentrate on the use of [...] raw data sources as a benchmark for comparison". However, the immense variety of possible parameters has an essential impact on the recorded signals. The selection of light color and sampling scheme [32], but also the hardware’s individual characteristics make it even harder to generate a universal dataset which novel approaches such as spectral estimation and sparse sampling could be tested with.

2 RAW SIGNALS

In digital signal processing, datasets of original, *raw* signals are preferred over preprocessed ones which naturally limit the possible use right from the start. *Raw* data are considered to be universal and to still provide all inherent information, the noise spectrum as well as the desired signals and even hidden secondary information. Although these datasets tend to become very large quickly, it is still reasonable to record the direct measurements, if possible, as today’s computers are usually capable of handling them.

However, Gitelman et al. [13] state that "raw data is an oxymoron" as "data are always already ‘cooked’ and never entirely ‘raw’". Although rather philosophic, this statement is also valid for PPG sensing. The definition of *raw* is strongly related to the perspective and interest of the researcher. It mainly depends on the research domain and hence the intended level of abstraction, from the highest one of the devices’ consumer via the medicals’ view on HR or SpO₂ signals down to the engineers’ bits, amperes, and volts or even the physicists’ luminous flux, absorption, and reflectivity.

Consequently, in context of this paper, *raw* signals are defined to contain the maximum possible information by applying the minimum necessary preprocessing. In case of PPG this means that the recordings contain the directly captured values from the analog-to-digital converter (ADC) which receives the amplified analog signal from the photodetector, usually via a transimpedance amplifier. An excerpt from such a *raw* signal is presented in Figure 1.

3 PHOTOPLETHYSMOGRAPHY

The optical measurement principle photoplethysmography (PPG) noninvasively measures the blood volume flow pulsating in the microvascular bed of the tissue beneath the skin. It utilizes an intensive light source to illuminate the skin while a photodetector measures the light that is either passing through the tissue in transmission mode or that is reflected by the tissue in reflective mode. For decades, the transmission mode is traditionally applied to the fingertip or the earlobe at which pulse oximetry is a proven method to monitor heart rate (HR) and peripheral oxygen saturation (SpO₂) of regular ward patients. In contrast, modern wearables mostly apply the reflective mode at the dorsal wrist at which both light-emitting diodes and photodiodes are placed nearby on the skin surface to measure the light reflected by perfused tissue. [3, 31]

In both modes, short light flashes are emitted to sample the blood volume of the moment. While there is a broad consensus on the origin of the transmission mode’s signal, in which an increasing blood volume absorbs a larger amount of light, the origin of the signal modulation in reflective mode, however, is not entirely clear and still subject to research [20, 30]. Moreover, there is a lot of confusion about the direction of the original, *raw* signal’s course. For both modes, the received light intensity is inversely proportional to the blood volume by nature, but commonly flipped to be consistent with the associated arterial blood pressure (ABP) [1, 8].

Depending on the utilized light color, the pulse is captured at different tissue layers which textures result in a varying motion tolerance [9, 17]. Also the actual AC signal amplitude largely depends on the wavelength, but still comprises only about 1 to 10 % of the total signal scope [20, 21]. While the pulsating blood vessels modulate the reflected light, the smaller veins and other nearby tissue just add a DC offset. Consequently, the analog-to-digital converter has to provide a high resolution, usually ranging from coarse 12 up to 24 bit, to enable the detailed representation of the desired AC signal. While commercial devices already apply advanced sampling strategies to improve the signal’s robustness and signal-to-noise ratio, in research uniform sampling is still standard and often only one single and even fused channel is provided in the datasets [32].

Table 1: Overview of the reviewed publicly available datasets.

id	dataset	authors	year	origin	data format	#subjects	#recordings	length	link
S01	MAXREFDES100	[2]	Biagetti et al. 2020	recording	*.mat	7	105		https://www.sciencedirect.com/science/article/pii/S2352340919314003
S02	PPG-DaLiA	[26]	Reiss et al. 2019	recording	*.pkl / *.csv	15			https://ubicomp.eti.uni-siegen.de/home/datasets/sensors19/index.html.en
S03	WESAD	[29]	Schmidt et al. 2018	recording	*.pkl / *.csv	15 (12m, 3f)		~36 min	https://ubicomp.eti.uni-siegen.de/home/datasets/icmi18/index.html.en
S04	BloodLossSvm	[27]	Rejin et al. 2018	recording	*.csv / *.txt	9 (trauma) / 27 (healthy)	67 / 27	2 min	https://figshare.com/articles/NR_bloodlosssvm_zip/5594644
S05	PPG-BP	[24]	Liang et al. 2018	recording	*.txt	219	657	3 × 2.1 s	https://figshare.com/articles/PPG-BP_Database_zip/5459299
S06	BIDMC	[25]	Pimentel et al. 2017	MIMIC II [28]	*.mat / *.csv	53	53	8 min	https://physionet.org/content/bidmc/1.0.0/
S07	Wrist PPG During Exercise	[16]	Jarchi et al. 2017	recording	wfdb data	8 (5f, 3m)			https://physionet.org/content/wrist/1.0.0/
S08	Cuff-Less Blood Pressure Estimation	[19]	Kachuee et al. 2015	PhysioNet [14]	*.mat				https://archive.ics.uci.edu/ml/datasets/Cuff-Less+Blood+Pressure+Estimation
S09	IEEE SPC 2015 (TROIKA)	[35]	Zhang et al. 2015	recording	*.mat	12 (training) / 8 (test)	12 / 8		https://sites.google.com/site/researchbyzhang/ieeespcup2015
S10	IEEE SPC 2013	[22]	Karlen et al. 2013	CapnoBase [23]	*.mat	42	42	8 min	http://www.capnabase.org/index.php?id=857

#: number of

Table 2: Overview of the applied PPG sensor setups and configurations.

id	sensing device	location	mode	illumination	f_s	ADC resolution	preprocessing
REF	MAX86140 (EVSYS#)	wrist (dorsal)	reflective	green (2x 528 nm), yellow (2x 590 nm)	512 Hz	19 bit	BS 50 / 60 Hz (hardware)
S01	MAXREFDES100#	wrist	reflective	infrared (880 nm), red (660 nm), green (537 nm)	400 Hz	16 bit	
S02	Empatica E4	wrist (non-dominant)	reflective	red (2x), green (2x)	64 Hz	0.9 nW/dig	BW removal, MA removal (combines different light waves)
S03	Empatica E4	wrist (non-dominant)	reflective	red (2x), green (2x)	64 Hz	0.9 nW/dig	BW removal, MA removal (combines different light waves)
S04		finger, forehead, ear	reflective	infrared, red	80 Hz		
S05	SMPLUS SEP9AF-2	earlobe, fingertip (?)	transmission	infrared (905 nm), red (660 nm)	1000 Hz	12 bit	BP 0.5 to 12.0 Hz
S06		wrist (dorsal) (?)	transmission (?)		125 Hz		
S07	Shimmer 3 GSR+	left hand's finger (?)	reflective	green (510 nm)	256 Hz	12 bit	Shimmer's on-board filter / cycling: LP 15 Hz (2nd order Butterworth)
S08		fingertip	transmission (?)		125 Hz		
S09		wrist (dorsal)	reflective	green (2x 515 nm, 2 cm distance)	125 Hz		BP 0.4 to 4.0 Hz (2nd order Butterworth)
S10			transmission (?)		300 Hz		

(?:) ambiguous information; BP: band-pass filter; BS: band-stop filter; BW: baseline wandering; MA: motion artifact

PPG is an active measurement principle which illuminates the skin surface with high-intensity LEDs and simultaneously measures the intensity of the light reflected. The sensor's sampling rate f_s is a very important parameter as it decides about the energy dissipation and the device's battery life on the one hand, but also on the temporal resolution and the signal's details on the other hand. Although Choi et al. [7] state that 25 Hz are sufficient to derive the HRV adequately and Wollong et al. [33] state that even 10 Hz are sufficient to derive the heart rate properly, most commercial devices use a significantly higher sampling rate to receive detailed signals with a high temporal resolution [32]. However, one has to distinguish the pulse repetition frequency which is up to several kHz and the group frequency in the tens to hundreds of Hz. For the recording of a research dataset, a higher sampling frequency is desirable. The uniform samples can later be resampled to a lower rate or even irregularly sampled for advanced sparse sampling.

Common preprocessing stages for PPG signals include traditional filters such as high-pass, band-pass, and notch filters. High-pass filters are used to detrend the *raw* signal, to remove the predominating DC component and the baseline wandering, to get a zero-centered signal. The use of a band-pass filter adds a low-pass stage to smooth the signal contour and to suppress high-frequency noise. A notch band-stop filter is often used to eliminate power line noise around 50 / 60 Hz and is usually implemented on-board, in the sensor's analog front-end hardware. The aforementioned filters are usually implemented in software as Butterworth FIR or IIR filter of 2nd order, as higher orders tend to degrade the signal. The emerging Savitzky-Golay FIR is a linear phase smoothing filter which conserves the pulsatile shape of the desired waveform on top of the non-stationary signal and non-Gaussian noise, but does not show a constant, predictable transfer function [6]. [11]

4 DATASET REVIEW

Table 1 provides an overview of the 10 reviewed datasets, their related publications, information about the length and number of subjects and contained recordings, as well as links to their publicly available repositories. Additional information about the utilized sensing devices, technical details, and applied preprocessing, as stated in their documentation, are summarized in Table 2.

4.1 Reference Data

The recording of a suitable reference dataset with multiple subjects has been prohibited by the SARS-CoV-2 pandemic in 2020. Consequently, recordings from preliminary experiments in 2019 were taken to serve as a reference for the quality review. However, the set of just a few experimental recordings is not itself suitable for a full research dataset, as the number of participants is too small. The evaluation system of the sensor Maxim MAX86140 has been utilized for the recording. The technical details are provided in Table 2, for a more detailed description please refer to Wollong et al. [33].

4.2 Decision Metrics

The suitability and applicability of a dataset is difficult to quantify in a universal way. Hence, we decided to provide the following 7 decision metrics which can be used to support the selection of a publicly available dataset, either from the list of the reviewed 10 references or from a different one by applying the developed tool.

4.2.1 Time Base. The sampling frequency f_s is often assumed to be constant. Hence, the individual samples' timestamps are usually omitted to save valuable memory and only the desired rate is stated. However, due to internal processes, devices tend to show a deviating sampling period Δt and consequently a jittering frequency. The rate's (1) mean \bar{f}_s is preferably close to the desired value and the standard deviation σ_{f_s} ought to be negligible. Despite that, especially at high f_s it is beneficial to know the exact time of a taken sample to generate a regularly sampled dataset by interpolation.

$$\bar{f}_s = \frac{1}{n} \sum_{i=1}^n \frac{1}{\Delta t_i} \quad \text{and} \quad \sigma_{f_s} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{\Delta t_i} - \bar{f}_s \right)^2} \quad (1)$$

with $\Delta t_i = t_{i-1} - t_i$

4.2.2 Signal Mean. The *raw* signal y naturally contains a very large DC component. Consequently, its mean (2) already tells a lot about the applied preprocessing. A mean of about 0 means that the signal has potentially been shifted to get a zero-centered signal or even a high-pass filter has been applied which also causes the signal to drop back to the origin. For practical reasons an error margin ϵ_z of 1.5 % is added as filtered values do not always hit exact 0.

$$\bar{y} : \begin{cases} \bar{y} \gg 0 & , \text{natural DC component} \\ \bar{y} \approx 0 & , \text{detrended, zero-centered} \end{cases} \quad (2)$$

4.2.3 Signal Scope. The signal's scope (3) represents only a minor fraction of the overall signal's extent and is often cropped at its minimum to reduce the memory demands. In many cases, the signals are even scaled and normalized in the range of $[0, 1]$. In other cases the signal is fit into the range of $[-1, 1]$, but if the signal has been zero-centered beforehand, then either the $\min(y)$ or the $\max(y)$ might not reach the lower -1 or upper 1 boundary.

$$y : \begin{cases} 0 \ll \min(y) < \max(y) & , \text{normal scope} \\ \min(y) \geq 0 \wedge \max(y) > 1 & , \text{cropped} \\ \min(y) \geq 0 \wedge \max(y) \leq 1 & , [0, 1] \text{ normalized} \\ \min(y) \geq -1 \wedge \max(y) \leq 1 & , [-1, 1] \text{ normalized} \end{cases} \quad (3)$$

4.2.4 Granularity. The granularity Δ of the signal y is ideally identical to its amplitude resolution if the sampling rate f_s is sufficiently high to quantize and reconstruct the signal slopes. It is determined from the sorted list of unique values without duplicates by seeking the minimum Euclidean distance (4). The granularity helps to unveil applied preprocessing as the values directly obtained from the ADC are binary integers by nature (5).

$$\Delta = \min(y_i - y_{i-1}) \forall y_i \in \text{sort}(\text{unique}(y)) \quad (4)$$

$$\Delta : \begin{cases} \Delta \in \mathbb{N} \wedge \Delta = 1 & , \text{integer of 1 digit} \\ \Delta \in \mathbb{N} \wedge \Delta > 1 & , \text{integer, small } f_s \text{ or short } t \\ \Delta \in \mathbb{R} & , \text{floating-point} \end{cases} \quad (5)$$

4.2.5 Clipping. If the signal has been normalized either to $[0, 1]$ or to $[-1, 1]$, clipping artifacts can occur at the boundaries which cut the caps of the lowest and highest peaks. Those flat tops are detected by means of multiple successive samples that stay at the constant boundary values for a longer period t_c , then counted and averaged over 30 s windows along the entire time series.

4.2.6 Flipping. Traditional pulse oximetry sensors monitor the PPG signal proportional to the course of the arterial blood pressure (ABP) and hence have to flip it to enable this analogy. The *raw* signals of both PPG modes, however, originally show an inversely proportional course [1]. To determine the pulse direction, two measures are determined. The first one determines the pulses' center of mass which is usually originated at the systolic onset while the diastolic peak is much lighter. Based on Choi et al. [8], the second measure compares the steepness of the down and up slopes which are steeper for the systolic than for the diastolic phase.

4.2.7 Frequency Spectral Ratio. As all physiological signals, *raw* PPG signals are non-stationary and dominated by low-frequency baseline wandering. Hence, most approaches are applying a high-pass or band-pass filter to remove the low-frequency components and to limit the pulsatile signal in a constant boundary envelope. This filtering, however, prevents the option to analyze these frequency components which are associated with activity in the autonomic nervous system and particularly respiration [10, 22, 25].

The frequency spectrum is split up into four bands. The very low frequency VLF band (0.0 to 0.167 Hz) predominantly contains

random baseline wandering. The low LF band (0.167 to 0.667 Hz respective 10 to 40 bpm) mainly contains respiratory signals, but is overlapping with the intermediate IF band (0.5 to 3.0 Hz respective 30 to 180 bpm) which mainly contains the heartbeat signal [10, 12]. The high frequency HF band (>3.0 Hz) is associated with noise, but can also contain higher harmonics of the heartbeat. Disturbances through daily motion are mainly located in the 1.0 to 2.5 Hz band [31] and, consequently, might affect these metrics.

The metrics are derived from firstly the ratio of the dominant peak in the VLF versus the dominant peak in the IF band (6) and secondly the dominant peak in the LF versus the mean of the IF band (7). They are covering the most common corner frequencies applied to detrend the *raw* PPG signal.

$$\frac{\max(VLF)}{\max(IF)} : \begin{cases} \gg 1 & , \text{if very low frequencies present} \\ \leq 1 & , \text{if high-pass filtered, } f_c \geq 0.167 \text{ Hz} \end{cases} \quad (6)$$

$$\frac{\max(LF)}{\bar{IF}} : \begin{cases} \gg 1 & , \text{if low frequencies present} \\ \leq 1 & , \text{if high-pass filtered, } f_c \geq 0.667 \text{ Hz} \end{cases} \quad (7)$$

4.3 Results

Table 3 summarizes the output from the multi-varied quality analysis tool applying the presented 7 decision metrics. Most of the reviewed datasets were recorded at a sampling rate f_s of more than 100 Hz, except for the two datasets **S02** and **S03** based on the wearable Empatica E4 with fixed 64 Hz. The time bases are provided for **S01**, **S06**, and **S07**. However, as no jitter was detectable for **S01**, its samples' timestamps have probably been added subsequently, based on the desired f_s , but the real f_s according to the device's internal clock is not traceable anymore. In contrast, **S06** and **S07** provide real timestamps of the samples' moments which enables the subsequent resampling and interpolation to a regular rate. **S01** and **S05** showed a granularity of 1 which allows the conclusion that those contain actual *raw* signals directly obtained from the ADC. However, while the other metrics of **S01** indicate that it was not preprocessed at all, the ones of **S05** indicate a flipped signal course. The majority of the flipped time series are originated in a transmission mode measurement (**S05**, **S06**, **S08**, and **S10**) while only the Empatica E4 devices' signals **S02** and **S03** are also flipped, presumably to conform with the measurements of traditional pulse oximeters. For **S07** it was not possible to validate the detected direction. Although most time series were filtered, only **S02**, **S03**, **S09**, and **S10** are actually zero-centered. While **S06** is ideally fitted into $[0, 1]$, **S04** is rather $[-1, 1]$, but probably intended to be also $[0, 1]$ -normalized. The metrics VLF and LF reliably distinguish the unfiltered (**S01**, **S04**) from the less (**S07**, **S10**) and the strongly (**S08**, **S09**) high-pass filtered datasets. However, they are inapplicable for **S05** as it contains only very short signal snippets of 2.1 s which result in a very coarse and inadequate frequency spectral resolution. Only **S10** showed clipping artifacts which are typical for aged CapnoBase data. In general, the results of **S02** and **S03**, originated in the same research team, show a high similarity and accordance although the datasets contain independent recordings from different studies with different research questions. Accompanied by **REF**, only the most recent dataset **S01** proved itself to contain entirely *raw* signals, but **S01** without providing real timestamps.

Table 3: Results of the multi-varied quality analysis. Color highlighted indicators for the output of the 7 decision metrics with green: positive, red: negative, orange: vague declaration, blue: unverifiable. Additionally, overview of subsidiary measures.

id	dataset	sampling rate f_s (Hz)			signal characteristics				time domain			frequency domain		artifacts			
		desired	real	jitter	mean	min	max	span	#values	gran. Δ	ZC	[0, 1]	[-1, 1]		flip	VLF / IF	LF / IF
REF	Reference Data (MAX86140)	512.0	511.750	0.001	176529.1	174920.0	178594.0	3674.0	3477	1	No	No	No	No	76.9	103.0	No
S01	MAXREFDES100#	400.0	400.000		6054.6	5661.0	7387.0	1726.0	1553	1	No	No	No	No	109.5	55.3	No
S02	PPG-DaLiA	64.0			-0.002	-1647.390	1557.6	3205.0	59323	0.010	Yes	No	No	Yes	0.150	1.783	No
S03	WESAD	64.0			-0.000	-873.670	988.080	1861.8	45440	0.010	Yes	No	No	Yes	0.044	1.947	No
S04	BloodLossSVM	200.0			0.475	-0.013	0.998	1.011	1011	0.001	No	Quasi	Yes	No	246.6	15.7	No
S05	PPG-BP	1000.0			2036.9	1682.0	2587.0	905.000	511	1	No	No	No	Yes	² 0.009	² 0.047	No
S06	BIDMC	125.0	125.000	3.559	0.466	0.224	0.698	0.474	407	0.001	No	Yes	No	Yes	0.018	1.884	No
S07	Wrist PPG During Exercise	256.0	255.882	0.575	1378.3	1269.8	1498.1	228.372	23744	0.003	No	No	No	¹ No	0.066	31.0	No
S08	Cuff-Less Blood Pressure Estimation	125.0			1.840	0.000	4.002	4.002	2792	0.001	No	No	No	Yes	0.009	0.590	No
S09	IEEE SPC 2015 (TROIKA)	125.0			-0.708	-270.000	208.500	478.500	840	0.500	Yes	No	No	No	0.088	0.781	No
S10	IEEE SPC 2013	300.0			-0.215	-9.840	10.240	20.080	754	0.020	Yes	No	No	Yes	0.059	0.927	Yes

#: number of; ZC: zero-centered; VLF / IF: very low to intermediate frequency ratio; LF / IF: low to intermediate frequency ratio; 1: too noisy signal; 2: too short recordings

5 DISCUSSION

The results in Table 3 show many alerts and, besides the reference data **REF**, only dataset **S01** fulfills most of the criteria. Nevertheless also the other datasets are suitable and applicable for specific research questions when being aware of their origin and limitations.

The most important criteria are related to filtering as it significantly affects the signal and limits the expedient applications. If the signal mean is close to zero, most likely a high-pass filter has been applied which removes the low-frequency components and hence shifts the signal. Although it is an artifact of aged datasets, also clipping considerably affects the signal quality by cutting the lowest and highest peak caps and thus impedes their positioning. In contrast, the granularity as well as the $[0, 1]$ and $[-1, 1]$ normalizations do not affect the signal itself, but indicate that the *raw* signals from the ADC have been relabeled according to a physical value or even rescaled. In case of an interpolation process, e.g. due to regularization, the granularity would presumably show rather a floating point than an integer value. In general, the transition from integer to floating point values is unfavorable as the calculation with those often results in rounding errors and inaccuracies. The awareness of flipping allows to repeatedly flip the signal as most algorithms are less effective with slopes in inverse direction.

Limitations. We assume that the characteristics of a single time series are valid for the entire dataset. As the two datasets **S02** and **S03** show, this assumption applies not only within the same series, but also for the same device type. Hence, we carefully selected recordings that, in our opinion, represented a meaningful cross section of the entire dataset. Of course the tool can be applied to all particular data, serving as a basis for a statistical analysis, but this would have gone beyond the scope of this paper.

6 GUIDELINES

We provide the following guidelines to supplement the general requirements on quality datasets, e.g. [4, 29, 34], with a specific focus on PPG datasets. Unprocessed, *raw* measurements are preferable and do not limit the research in signal processing and algorithms for wearable PPG sensing right from the start. As the performance of wearables tends to be limited, their long-term deployment in-the-wild is not easy. Hence, the sensor configuration is always a trade-off between universality and reusability of the data on the one hand, but required memory and battery life on the other hand.

We would like to encourage 1) to utilize PPG sensors that are capable of recording *raw* or just slightly filtered signals; 2) to record synchronized reference signals such as ECG and RSP; 3) to use the maximum possible sampling rate as long as it does not limit and terminate your experiment early; 4) to save the unfused samples of all particular measurement channels, also the ambient light intensity, if available; 5) to enable the recording of timestamps for each taken sample. These configurations consume higher amounts of valuable memory and energy, so balance the parameters according to your research interests and mitigate others for long-term monitoring.

Further, 6) provide all technical details of the utilized sensing device as well as its configuration. This includes not only the desired sampling rate and the number of measurement channels, but also the components' names, e.g. of LEDs and photodetectors, their light wavelength and drive current or peak sensitivity, the sample duration, and the applied sampling scheme. Also 7) describe setup, attachment, and measurement location of the sensor. The location "wrist" can for example be described more precisely by adding "volar" for the palm side or "dorsal" for the back side.

A 8) detailed documentation with a brief description of your research domain and questions will help your fellows to appraise whether your experiments and the recorded data are compatible.

7 CONCLUSION

We have presented an analytical tool for the quality review of 10 publicly available photoplethysmography (PPG) datasets, based on 7 multi-varied decision metrics. Although all datasets were advertised to contain *raw* signals, the characteristics of the PPG data look quite diverse. Our developed tool enables to automatically analyze the suitability and applicability of datasets for the intended research approach and helps to identify preprocessed and filtered signals with a limited evidence. Furthermore, we argue for more quality datasets which actually contain *raw* PPG recordings that do not limit their further use right from the start. Hence, we provide guidelines for future datasets with a focus on recordings of reflective mode PPG for research in digital signal processing and the development of algorithms for resource-limited wearable devices.

The *raw* reference data, recorded with the MAX86140EVSYS# evaluation system, as well as the implemented Python tool, based on the presented 7 decision metrics, will be available for download to support the reproducibility and the review of new datasets:

<https://ubicomp.eti.uni-siegen.de/home/datasets/data20/>

REFERENCES

- [1] Tomas Ysehak Abay. 2016. *Reflectance Photoplethysmography for Non-invasive Monitoring of Tissue Perfusion*. Doctoral Thesis. University of London, London, UK. <http://openaccess.city.ac.uk/16923/>
- [2] Giorgio Biagetti, Paolo Crippa, Laura Falaschetti, Leonardo Saraceni, Andrea Tiranti, and Claudio Turchetti. 2020. Dataset from PPG wireless sensor for activity monitoring. *Data in brief* 29 (2020), 105044. <https://doi.org/10.1016/j.dib.2019.105044>
- [3] Dwaipayan Biswas, Neide Simoes-Capela, Chris van Hoof, and Nick van Helleputte. 2019. Heart Rate Estimation From Wrist-Worn Photoplethysmography: A Review. *IEEE Sensors Journal* 19, 16 (2019), 6560–6570. <https://doi.org/10.1109/JSEN.2019.2914166>
- [4] Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *Comput. Surveys* 46, 3 (2014), 1–33. <https://doi.org/10.1145/2499621>
- [5] Peter H. Charlton, Drew A. Birrenkott, Timothy Bonnici, Marco A. F. Pimentel, Alistair E. W. Johnson, Jordi Alastruey, Lionel Tarassenko, Peter J. Watkinson, Richard Beale, and David A. Clifton. 2018. Breathing Rate Estimation from the Electrocardiogram and Photoplethysmogram: A Review. *IEEE Reviews in Biomedical Engineering* 11 (2018), 2–20. <https://doi.org/10.1109/RBME.2017.2763681>
- [6] Ayan Chatterjee and Uttam Kumar Roy. 2018. PPG Based Heart Rate Algorithm Improvement with Butterworth IIR Filter and Savitzky-Golay FIR Filter. In *2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)*, Satyajit Chakrabarti (Ed.). IEEE, 1–6. <https://doi.org/10.1109/IEMENTECH.2018.8465225>
- [7] A. Choi and H. Shin. 2017. Photoplethysmography sampling frequency: pilot assessment of how low can we go to analyze pulse rate variability with reliability? *Physiological measurement* 38, 3 (2017), 586–600. <https://doi.org/10.1088/1361-6579/aa5efa>
- [8] Changmok Choi, Byung-Hoon Ko, Jongwook Lee, Seung Keun Yoon, Uikun Kwon, Sang Joon Kim, and Younho Kim. 2017. PPG pulse direction determination algorithm for PPG waveform inversion by wrist rotation. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2017 (2017), 4090–4093. <https://doi.org/10.1109/EMBC.2017.8037755>
- [9] W. J. Cui, L. E. Ostrander, and B. Y. Lee. 1990. In vivo reflectance of blood and tissue as a function of light wavelength. *IEEE transactions on bio-medical engineering* 37, 6 (1990), 632–639. <https://doi.org/10.1109/10.55667>
- [10] Parastoo Dehkordi, Ainara Garde, Behnam Molavi, J. Mark Ansermino, and Guy A. Dumont. 2018. Extracting Instantaneous Respiratory Rate From Multiple Photoplethysmogram Respiratory-Induced Variations. *Frontiers in physiology* 9 (2018), 948. <https://doi.org/10.3389/fphys.2018.00948>
- [11] Mohamed Elgendi. 2012. On the Analysis of Fingertip Photoplethysmogram Signals. *Current Cardiology Reviews* 8, 1 (2012), 14–25. <https://doi.org/10.2174/157340312801215782>
- [12] Susannah Fleming, Matthew Thompson, Richard Stevens, Carl Heneghan, Annette Plüddemann, Ian Maconochie, Lionel Tarassenko, and David Mant. 2011. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. *The Lancet* 377, 9770 (2011), 1011–1018. [https://doi.org/10.1016/S0140-6736\(10\)62226-X](https://doi.org/10.1016/S0140-6736(10)62226-X)
- [13] Lisa Gitelman. 2013. *Raw data is an oxymoron*. The MIT Press, Cambridge, Massachusetts. <https://cds.cern.ch/record/1530979>
- [14] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley. 2000. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 101, 23 (2000), E215–20. <https://doi.org/10.1161/01.cir.101.23.e215>
- [15] Empatica Inc. [n.d.]. Empatica E4. <https://www.empatica.com/research/e4/>. Accessed: 2020-09-15.
- [16] Delaram Jarchi and Alexander Casson. 2017. Description of a Database Containing Wrist PPG Signals Recorded during Physical Exercise with Both Accelerometer and Gyroscope Measures of Motion. *Data* 2, 1 (2017), 1. <https://doi.org/10.3390/data2010001>
- [17] Liu Jing, Zhang Yuan-Ting, Ding Xiao-Rong, Dai Wen-Xuan, and Zhao Ni. 2016. A Preliminary Study on Multi-Wavelength PPG Based Pulse Transit Time Detection for Cuffless Blood Pressure Measurement. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2016 (2016), 615–618. <https://doi.org/10.1109/EMBC.2016.7590777>
- [18] Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li-Wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data* 3 (2016), 160035. <https://doi.org/10.1038/sdata.2016.35>
- [19] Mohamad Kachuee, Mohammad Mahdi Kiani, Hoda Mohammadzade, and Mahdi Shabany. 2015. Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time. (2015), 1006–1009. <https://doi.org/10.1109/ISCAS.2015.7168806>
- [20] Alexei A. Kamshilin and Nikita B. Margaryants. 2017. Origin of Photoplethysmographic Waveform at Green Light. *Physics Procedia* 86 (2017), 72–80. <https://doi.org/10.1016/j.phpro.2017.01.024>
- [21] Yung-Hua Kao, Paul C.-P. Chao, and Chin-Long Wey. 2019. Design and Validation of a New PPG Module to Acquire High-Quality Physiological Signals for High-Accuracy Biomedical Sensing. *IEEE Journal of Selected Topics in Quantum Electronics* 25, 1 (2019), 1–10. <https://doi.org/10.1109/JSTQE.2018.2871604>
- [22] Walter Karlen, Srinivas Raman, J. Mark Ansermino, and Guy A. Dumont. 2013. Multiparameter Respiratory Rate Estimation from the Photoplethysmogram. *IEEE transactions on bio-medical engineering* 60, 7 (2013), 1946–1953. <https://doi.org/10.1109/TBME.2013.2246160>
- [23] Walter Karlen, M. Turner, Erin Cooke, Guy Dumont, and J. Mark Ansermino. 2010. CapnoBase: Signal database and tools to collect, share and annotate respiratory signals. In *2010 Annual Meeting of the Society for Technology in Anesthesia*. 25.
- [24] Yongbo Liang, Zhencheng Chen, Guiyong Liu, and Mohamed Elgendi. 2018. A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China. *Scientific Data* 5, 1 (2018), 180020. <https://doi.org/10.1038/sdata.2018.20>
- [25] Marco A. F. Pimentel, Alistair E. W. Johnson, Peter H. Charlton, Drew Birrenkott, Peter J. Watkinson, Lionel Tarassenko, and David A. Clifton. 2017. Toward a Robust Estimation of Respiratory Rate From Pulse Oximeters. *IEEE transactions on bio-medical engineering* 64, 8 (2017), 1914–1923. <https://doi.org/10.1109/TBME.2016.2613124>
- [26] Attila Reiss, Ina Indlekofer, Philip Schmidt, and Kristof Van Laerhoven. 2019. Deep PPG: Large-Scale Heart Rate Estimation with Convolutional Neural Networks. *Sensors (Basel, Switzerland)* 19, 14 (2019).
- [27] Natasa Reljin, Gary Zimmer, Yelena Malyuta, Kirk Shelley, Yitzhak Mendelson, David J. Blehar, Chad E. Darling, and Ki H. Chon. 2018. Using support vector machines on photoplethysmographic signals to discriminate between hypovolemia and euvoolemia. *PLoS one* 13, 3 (2018), e0195087. <https://doi.org/10.1371/journal.pone.0195087>
- [28] Mohammed Saeed, Mauricio Villarreal, Andrew T. Reisner, Gari Clifford, Li-Wei Lehman, George Moody, Thomas Heldt, Tin H. Kyaw, Benjamin Moody, and Roger G. Mark. 2011. Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database. *Critical care medicine* 39, 5 (2011), 952–960. <https://doi.org/10.1097/CCM.0b013e31820a92c6>
- [29] Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. 2018. Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. <https://doi.org/10.1145/3242969.3242985>
- [30] Nina Sviridova, Tiejun Zhao, Kazuyuki Aihara, Kazuyuki Nakamura, and Akimasa Nakano. 2018. Photoplethysmogram at green light: Where does chaos arise from? *Chaos, Solitons & Fractals* 116 (2018), 157–165. <https://doi.org/10.1016/j.chaos.2018.09.016>
- [31] Toshiyo Tamura, Yuka Maeda, Masaki Sekine, et al. 2014. Wearable Photoplethysmographic Sensors—Past and Present. *Electronics* 3, 2 (2014), 282–302. <https://doi.org/10.3390/electronics3020282>
- [32] Florian Wolling, Simon Heimes, and Kristof Van Laerhoven. 2019. Unity in Diversity: Sampling Strategies in Wearable Photoplethysmography. *IEEE Pervasive Computing* 18, 3 (2019), 63–69. <https://doi.org/10.1109/MPRV.2019.2926613>
- [33] Florian Wolling and Kristof Van Laerhoven. 2018. Fewer Samples for a Longer Life Span: Towards Long-Term Wearable PPG Analysis. In *Proceedings of the 5th International Workshop on Sensor-based Activity Recognition and Interaction (Berlin, Germany) (iWOAR '18)*. ACM, New York, NY, USA, Article 5, 10 pages. <https://doi.org/10.1145/3266157.3266209>
- [34] Kristina Yordanova, Jesus Favela, and Gabriela Marcu. 2019. Challenges Providing Ground Truth for Pervasive Healthcare Systems. *IEEE Pervasive Computing* 18, 2 (2019), 100–104. <https://doi.org/10.1109/MPRV.2019.2912261>
- [35] Zhilin Zhang, Zhouyue Pi, and Benyuan Liu. 2015. TROIKA: A General Framework for Heart Rate Monitoring Using Wrist-Type Photoplethysmographic Signals During Intensive Physical Exercise. *IEEE transactions on bio-medical engineering* 62, 2 (2015), 522–531. <https://doi.org/10.1109/TBME.2014.2359372>

A PLOTS

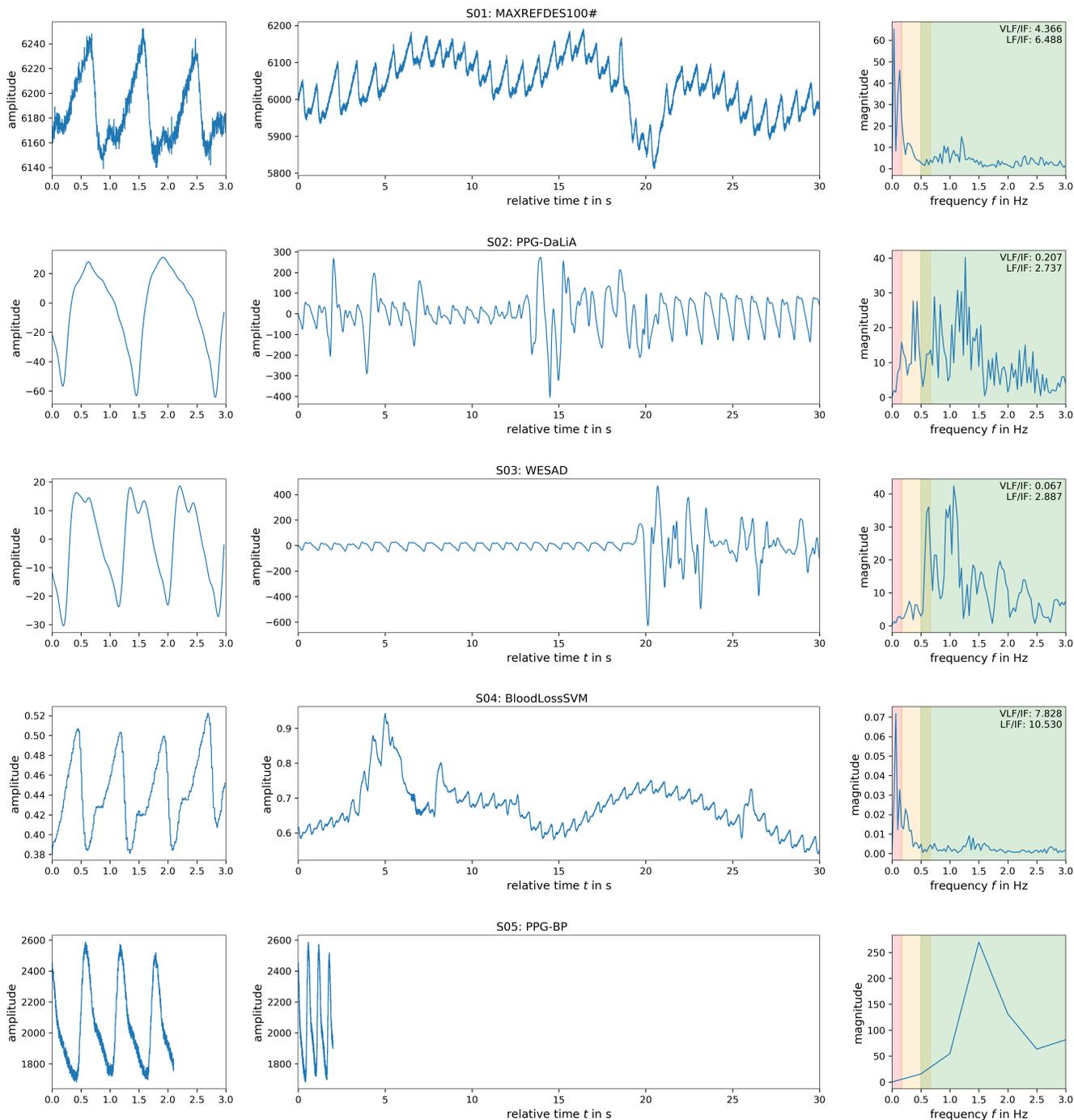


Figure A.1: Excerpts from the first five reviewed datasets: Short close-up of few pulses on the left, a 30-second window in the middle, and its respective frequency spectrum (FFT) on the right. Note that the PPG-BP dataset (bottom) contains only snippets of 2.1 s length. Frequency bands: very low frequency (VLF, < 0.167 Hz, red), low frequency (LF, 0.167 to 0.667 Hz, orange), and intermediate frequency (IF, 0.5 to 3.0 Hz, green) while the high frequency (HF, > 3.0 Hz) noise and harmonics are clipped.

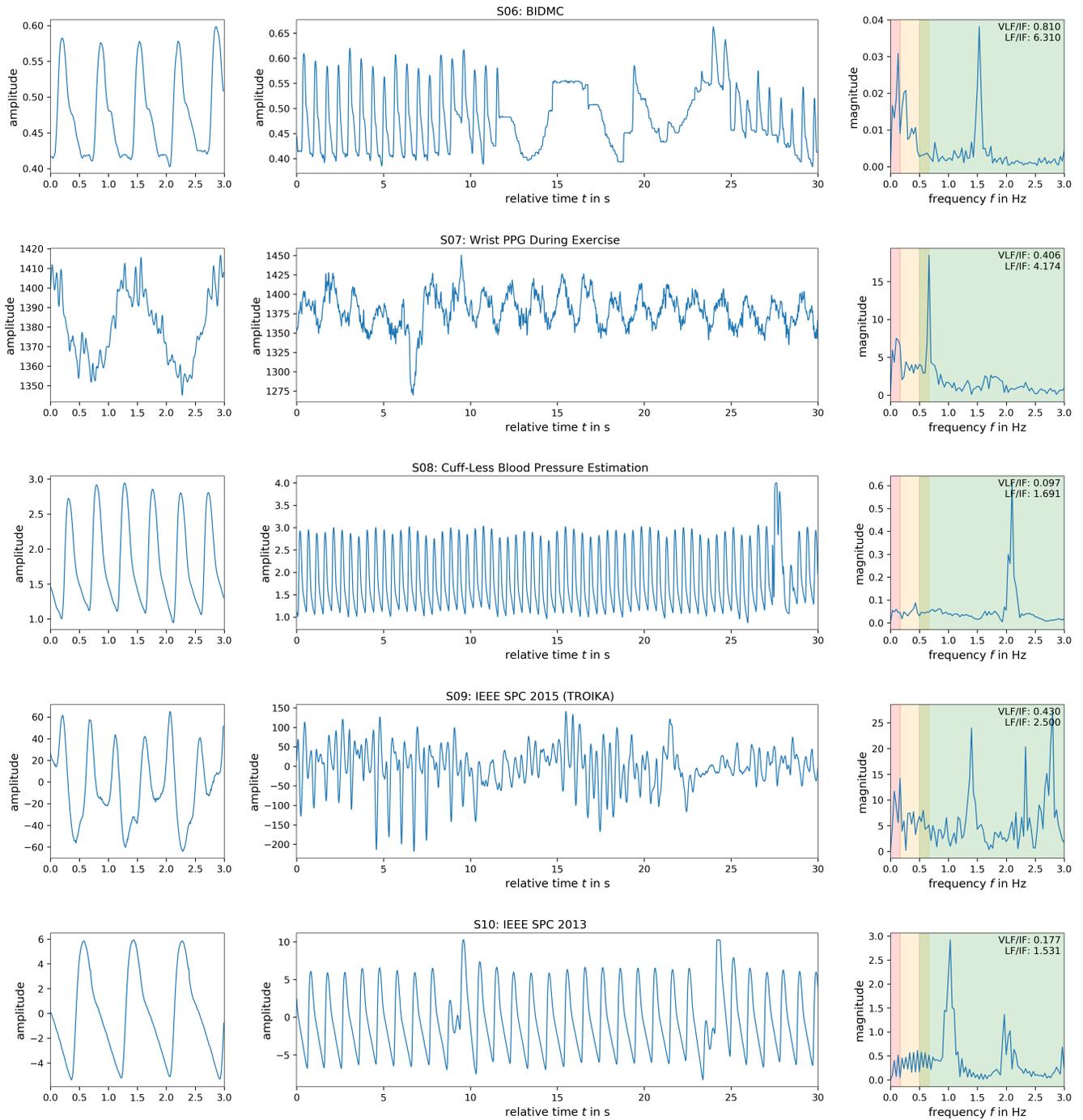


Figure A.2: Excerpts from the last five reviewed datasets: Short close-up of few pulses on the left, a 30-second window in the middle, and its respective frequency spectrum (FFT) on the right and cannot be analyzed in the frequency domain. Frequency bands: very low frequency (VLF, < 0.167 Hz, red), low frequency (LF, 0.167 to 0.667 Hz, orange), and intermediate frequency (IF, 0.5 to 3.0 Hz, green) while the high frequency (HF, > 3.0 Hz) noise and harmonics are clipped.