Supplemental Material

I. REPLICATED METHODS

A. LEE'S METHOD

In this method, the signals from all the load cells are first windowed with length of five seconds and band-pass filtered (5th order Butterworth with 1 to 20 Hz pass band). Then, the resulting signal is differentiated and passed through a non-linear transformation resulting in the Shannon entropy (SE) values. Finally, a moving average is applied on the output of the SE process for further smoothing. Peak detection is eventually performed on the automatically selected optimal result from one of the four load cells. The next window is then defined to begin 0.2 seconds after the last detected peak in the current window. Since Lee's method was proposed for data sampled at 1000 Hz, and our sampling rate is 250Hz, we also tested an up-sampled BCG for this method to assess the impact of the sampling rate mismatch on the performance of this method.

B. LYDON'S METHOD

In the Lydon method, windowing is applied on the pre-processed BCG (6th order Butterworth with 0.7 to 10 Hz pass band). The length of the window is proposed to be 0.3 seconds that will move across the BCG every 0.01 seconnds. Then the energy of each sliding window is obtained. The peaks in the resulting signal is supposed to correspond to the BCG and thus is considered for peak detection. Since this method was proposed for data sampled at 100 Hz and again different with what we have sampled our data at, a down-sampled BCG was tested for this method too. Furthermore, two moving-average filters were added to the Lydon method to reduce the large number of false positives that were found in our data; this addition can be justified by the use of a similar filter in Lee and similarity between Lee and Lydon's methods (we later learned from the original code that such a smoothing was included but not mentioned in the paper).

C. BRÜSER'S METHOD

Brüser also proposes a sliding window method, but takes a statistical approach to estimate heartbeat intervals (HBI's)

within that sliding window at each iteration constrained to two "upper and lower thresholds" based on prior knowledge of typical HR range. This method has two main parts, the "basic" and the "extended" parts of the algorithm.

In the "basic" part of the algorithm a sliding window or as they call it "analysis window" centered at a given point in time is selected. The length of the "analysis window" is restricted to less than twice the "upper threshold" for time between two heartbeats to ensure inclusion of at least two heartbeat events. A level thresholding is applied to each sliding window for motion artifact removal. Then, a local interval length is estimated in each sliding window using the estimators "Modified autocorrelation", "Modified average magnitude difference function (AMDF)" and "Maximum amplitude pairs (MAP)" at each window. The term "Modified" is prefixed for the first two estimators due to varying length of the sliding window. The mentioned estimators are then fused to obtain a "Probabilistic estimator" for the beat to beat interval. For the definitions of the mentioned estimators and the fusion algorithm in use, please refer to the original work.

The "extended" part of the algorithm makes use of the redundant peaks introduced due to the sliding window in the "basic" part. All the estimates around the centers of each sliding window are narrowed down and merged using a confidence threshold constrained under two conditions. The conditions are: 1. the peaks should be apart by approximately the length of the local interval for a given center; and 2. they must be mirrored about the center of the sliding window. Then if such peaks have the largest combined amplitudes, they were called the "boundary peaks" of the interval and the peak to the right of the center was referred to as the "anchor point". These anchor points were later used to group the interval estimates together and are "irrelevant" to the final interval estimation. In order to exclude unreliable estimates from the set, a "quality indicator" or "confidence indicator" $\overline{Q_k}$ is derived by averaging the individual confidence values associated with each estimate and then a fixed threshold th_Q to be applied to each $\overline{Q_k}$. As can be inferred from the original

paper, th_Q is a subject-specific threshold.

D. ALVARADO'S METHOD

This work applies a continuous wavelet transform (CWT) with B-Spline wavelets on the band pass filtered BCG (0.5-20Hz 2nd order). Then it is hypothesized that the zero crossings in the 5th scale of CWT of the BCG correspond to the J-Peaks of the original BCG. Once the 5th scale CWT of the BCG is obtained the rest of their work emphasizes more on how those zero crossings in the new CWT signal are automatically identified by use of adaptive level and interval thresholds to avoid false zero crossing events and consequently false J-Peaks. The process has two steps of "Learning Stage" and "Decision Stage". In the "Learning Stage", a set of detection thresholds to detect the first four beats is defined and a J-J interval average is estimated. Also, a "refractory period" of 300 ms is introduced in this stage to avoid false detection. In the "Decision Stage", the information from the "Learning Stage" is used as initial values based on which new level and J-J interval thresholds are estimated and decisions made on peak detection. This process continues in this manner till the end.

E. SADEK'S METHOD

In this method the band-pass filtered BCG (Chebyshev 2.5 to 5 Hz pass band and sampled at 50 Hz (or based on the paper, Butterworth 1 to 10 Hz pass band)) is decomposed into approximation and detail coefficients using MODWT. Then the 4th level "smooth coefficient" is selected for ten-second-long windowing and peak detection. The wavelet applied in the decomposition is the 8th order Symlet (Sym8 in MATLAB). The peak detection part is not described in detail, but from our inspection of the python code we were provided with, appears very similar to the built-in function "findpeaks" in MATLAB.

F. SIMPLE CROSS CORRELATION METHOD

For the cross correlation (XCOR) method, the BCG underwent preprocessing steps identical to what is proposed in Lydon method. Then a single BCG complex is picked as a template for performing the cross-correlation step and then using it for peak detection.

II. PARAMETERS OF INTEREST IN EACH METHOD

Lee, Lee with up-sampled BCG (Lee-US), Lydon, Lydon with down-sampled BCG (Lydon-DS) and XCOR methods use two moving-average filters for smoothing purposes and the lengths for the filters had to be adjusted. The lengths of the filters were mostly subject-specific for Lee and Lydon methods and subject-independent for XCOR method with the exception of only one participant. The BCG for that particular participant had to be removed later from the study as cross-correlation method failed to obtain non-zero detect and false peaks which made it impossible to calculate the MAE_p.

For the peak detection step, the "findpeaks" built-in function of MATLAB seems very likely to have been used in the original work of Lee and we used it for Lee-US, Lydon, Lydon-DS and XCOR methods as well. We also had to tune the "Mininimum Peak Prominence" input of the mentioned MATLAB built-in function, "findpeaks". Therefore, this was a two-step optimization process. First, for a certain value of "Mininimum Peak Prominence" that produced a reasonable pair of Det. and FAR, the lengths for the moving average filters were optimized. Then using those optimal window lengths, "Mininimum Peak Prominence" was iterated over different values to produce the ROC curves and a value for "Mininimum Peak Prominence" was selected.

For Brüser Method, we had to first find an optimal value for the percent-overlap of the "analysis window". A single value for this parameter turned out to be optimal for all other BCGs from the other participants. Next, the "confidence indicator threshold", th_Q, as discussed when introducing Brüser method was tuned to yield good performance in terms of "anchor points" or peaks elaborated earlier. Per our understanding from the original work, this threshold has to be adjusted for the BCG data from each participant.

In Alvarado method, since the process is adaptive, any alterations to their proposed constants or initial values for their proposed parameters turned out to have no effect on the performance of this method. Therefore, those parameters were set to the same values as initially proposed. However, the scale of the wavelet component of the BCG used for peak detection, and the choice of wavelet for decomposition were affecting the performance of this method significantly for our data. Since scale 5 of the CWT was resulting in very poor performance for our data, we tried a range of scales until an acceptable range of scales was found. Then within that range, scale numbers specific to each BCG set was selected for optimal performance. For the choice of an optimal wavelet, MATLAB provides family of "bior" wavelets that implement a B-Spline wavelet with bi-orthogonality and different members of this family proved to be optimal for some of the BCG sets. For the BCG sets where an optimal member from the "bior" family was not found, a member from its inverse, "rbio", or a member from the "sym" family was used. It is worth mentioning that for the cases where a wavelet was not optimal, the performance was severely affected and therefore sticking to a single wavelet was avoided in order not to penalize this method harshly. A summary of wavelets and scale combination is reported in Table 1.

For Sadek method, the first author assisted us by providing python code for it. Therefore, little ambiguity were left to address in implementing their code in MATLAB. Due to different sampling frequency (50Hz vs 250Hz which affects level selection in the wavelet transform), we did have to run some experiments to optimize for the right wavelet decomposition scale and window length. A summary of wavelet decomposition scales and window lengths used in this method for different participants is provided in Table 2. Since the python code was using "minimum peak distance"

as a parameter to help find the peaks in the peak detection code, the same parameter was used with MATLAB in this work to be consistent with the original work when comparing the performance. Clearly, different values for the mentioned parameter affect the overall performance of the method significantly and therefore this parameter was also considered for optimization. It is worth mentioning that this parameter is not specified in the paper and in the code we were provided with, a value of 1 i.e. one sample was assigned to this parameter which was introducing tremendous number of false peaks and that was another reason to optimize for that parameter. The drawback of using "minimum peak distance" is that we need to have prior knowledge about heartbeat distances in order to assign a value for this parameter which makes peak detection hard in an online scenario. In addition, the mentioned parameter was not used in Lee, Lee-US, Lydon, Lydon-US and cross-correlation methods which makes the comparison task harder. Therefore, the performance of this method is evaluated based on the "minimum peak prominence" values as well. When optimizing for this parameter, the decomposition scales determined earlier were first used to find a value for the "minimum peak prominence" parameter such that a good pair of detection and FAR was obtained. Then, using the values found for the "minimum peak prominence", the performance of the method was tested again for different decomposition scales to check if a different scale would be optimal in this scenario. During this process, different scales proved to be optimal for only BCGs from three participants. For those BCGs, a new set of "minimum peak prominence" parameters corresponding to the optimal scales were then sought.

III. WAVELETS AND WAVELET DECOMPOSITION SCALES

Table 1 summarizes the combination of wavelet and wavelet decomposition levels used to optimize the performance of Alvarado method. The "Number of Participants" column indicates the total number of participants for whom a certain combination of wavelet and decomposition level that resulted in good performance for the method.

TABLE 1. Wavelet name and wavelet component used to optimize the performance of Alvarado method for different participants

Number of Participants	Wavelet Name	Decomposation Levels
13	bior2.2	37, 30, 48, 37, 42, 32, 20, 49, 34, 50, 25, 47, 31
4	bior3.9	59, 76, 61, 41
7	bior4.4	34, 32, 48, 29, 29, 28, 19
4	rbio3.9	68, 31, 50, 50
1	sym4	25
1	sym8	39

Table 2, summarizes the wavelet decomposition scales (left) and window lengths (right) used to optimize the performance of Sadek method. Again, the "Number of Participants" column indicates the total number of participants for whom

TABLE 2. Left: Wavelet component used to optimize the performance of Sadek method for different participants. Right: Window lengths used to optimize the performance of Sadek method for different participants.

Number of Participants	Decomposition Scale
4	4
9	5
17	6

Number of Participants	Window Length
1	16
1	12
7	14
21	10

a certain wavelet component and window length maximized the performance of the method.

IV. SIGNAL QUALITY INDEX

The process of signal quality index (SQI), is based on Matched Filter approach [1]. The matching is performed by comparing the signal in question with a known signal by obtaining the correlation coefficient of the two [1]. In order to have a known BCG, BCG stream is segmented and the ensemble average of the segments is taken [1]. In order to figure out the starting point for each segment, R-peaks of the ECG were used as reference points; the ECG was simultaneously recorded with the BCG. The minimum R-R interval was used for the length of each BCG segment in [1].

In our case however, the segments' lengths were left to be the same distance between their corresponding R-R interval. This lead to unequal lengths of the segments. The length deficiencies were corrected by zero padding the short vectors; and when obtaining the ensemble averages, the zeros were excluded in the calculation.

The obtained ensemble average is considered to be the known signal [1] or the "main template" as referred to it in [1]. In addition, sub-templates were also made out of a number of adjacent BCG segments such that sum of all those numbers will add up to the total number of BCG segments in the stream.

Correlation coefficients were obtained by matching each sub-template with the main template and the average of all the resulting correlation coefficients was called SQI [1].

REFERENCES

[1] A. O. Bicen and O. T. Inan, "A signal quality index for ballistocardiogram recordings based on electrocardiogram rr intervals and matched filtering," in Biomedical & Health Informatics (BHI), 2018 IEEE EMBS International Conference on. IEEE, 2018, pp. 145–148.

● ● ●