A Novel Algorithm for HRV Estimation from Short-Term Acoustic Recordings at Neck

Piyush Sharma[1] and Esther Rodriguez-Villegas[2]

Abstract— **Heart rate variability (HRV) is an important non-invasive parameter to monitor the activity of the autonomic nervous system. This paper proposes an algorithm to analyze HRV by processing the acoustic data, recorded by placing a small, wearable sensor on the suprasternal notch (at neck) of an adult subject, primarily intended to record breathing sounds. The method used an empirical data analysis approach of the Hilbert-Huang transform (HHT) to construct an instantaneous energy envelope and segment the cardiac cycle by detecting S1 and S2 sounds using the K-means algorithm. The time-domain HRV analysis for the short-term recordings of 10 subjects demonstrated a close agreement with the reference ECG signal. The instantaneous heart rate (IHR) comparisons yielded an accuracy of 95.78% and 92.35% for S1 and S2 sounds respectively. The experimental results showed that the proposed algorithm can provide an accurate HRV analysis for the cardiac signals recorded at the neck.**

I. INTRODUCTION

Heart rate variability (HRV) measures the variation in the time intervals of the successive cardiac cycles and serves as an important physiological parameter to study the effect of sympathetic and parasympathetic branches of the autonomic nervous system on cardiovascular mortalities [1]. It is conventionally studied in two forms: inter-beat interval (IBI) or instantaneous heart rate (IHR), derived from the consecutive heartbeats. The gold standard method of measuring the HRV is by extracting the RR interval tachogram from the electrocardiogram (ECG). In recent studies, the detection of systolic peaks in photoplethysmography (PPG) signal has also been used to derive pulse rate variability as a surrogate choice to HRV [2]. While these methods provide useful cardiac information, the cumbersome ECG recording setup and the reliability concerns of the PPG signal limit their application to the HRV analysis using wearable technology [3]. Cardiac auscultation using phonocardiogram (PCG) provides an alternative for measuring heart cycle variation. PCG records the heart sounds by sensing the mechanical vibrations caused by the closure of the heart valves. These mainly consists of, what are called, S1 and S2 components, also known as fundamental heart sounds [4].

Only few studies have used the PCG signal to measure HRV and the majority of these works focus on the extraction of fetal HRV by recording the acoustic signals from the abdominal maternal wall [5], [6]. To the best of author's knowledge, this study presents, for the first time, a novel algorithm to measure HRV from the acoustic signals recorded at the neck using a wearable device. Methods for S1-S2 detection reported in literature have been mainly energy-based ones, using features such as Shannon energy, Shannon entropy, absolute amplitude, squared energy, etc. [7]. However, these techniques assume the input signal to be stationary, and therefore lose the self-adaptivity towards the processing of the data. This study overcomes such a limitation by using an empirical data analysis technique which uses a *posteriori*-defined basis, based on, and derived from, the non-stationary signal, to produce meaningful physical representations.

The rest of the paper is organized as follows: Section II discusses the proposed algorithm in detail. The experimental results are documented in Section III and the conclusions about the study are extracted in Section IV.

II. METHODS

The block diagram in Fig. 1 represents different stages of the proposed algorithm to detect S1 and S2 sounds from the acoustic breathing signal, leading to extraction of HRV. The algorithm is mainly divided in two phases: segmentation and classification of the heart sounds. Segmentation is performed by constructing the instantaneous energy envelope from intrinsic mode functions (IMFs) derived using empirical mode decomposition (EMD) and Hilbert-Huang transform (HHT); whereas the heart sounds are classified by clustering the features of the energy peaks using the K-means algorithm.

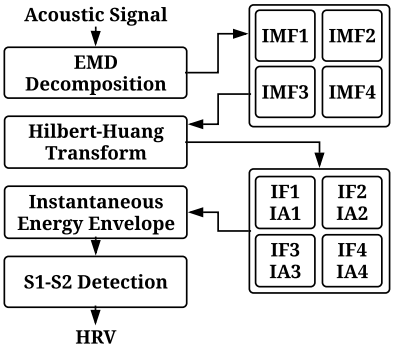

Fig. 1: Block diagram of the proposed algorithm.

A. Pre-processing

The acoustic signals recorded at the suprasternal notch, as shown in Fig. 2(a), are not suitable for a direct cardiac analysis as they are heavily corrupted by the breathing sounds. An uncontrolled environment during the recording setup also introduces motion artifacts and background noise. A higher bandwidth of 20 to 1000 Hz for breathing sounds

[1,2]The authors are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. {piyush.sharma14,e.rodriguez}@imperial.ac.uk

allows to separate the cardiac information generally present below 150 Hz [3]. The signal is processed in blocks of 5 s duration, which contain multiple heartbeats corresponding to a heart rate of 40 to 200 bpm, for an adult subject. A third-order Butterworth bandpass filter with cut-off frequencies of 5 and 150 Hz removes the baseline wandering and higher frequencies from the signal. The filtered signal originally sampled at 2205 Hz is downsampled by a factor of 5 to reduce the number of computational cycles in later stages of the algorithm. The cardiac signal thus obtained consists of S1 and S2 sounds as shown in Fig. 2(b).

B. EMD Decomposition and HHT

The heart sounds behave as a non-stationary time-series signal as their corresponding frequencies vary as a function of time. Therefore, a suitable way of processing the cardiac signal is by computing its instantaneous characteristics using HHT which provides meaningful physical interpretations of the data from non-stationary processes [8]. HHT rely on a *posteriori*-defined basis and consists of two phases: EMD and Hilbert spectral analysis (HSA). EMD is an adaptive method which decomposes a signal into different intrinsic modes of oscillations, also called IMFs using a sifting process [9]. An IMF represents a simple oscillation and satisfies two criteria: (a) the difference between the number of local extrema and zero crossings should not differ by more than 1; (b) the mean of the envelope constructed by local maxima and local minima is zero.

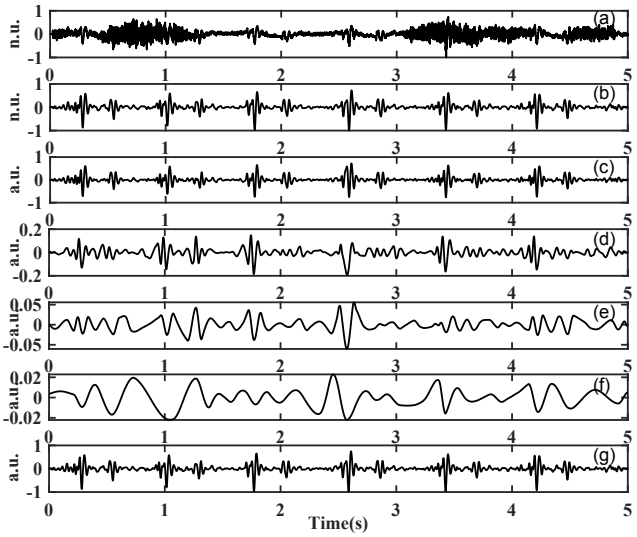

Fig. 2: EMD decomposition of the input signal. (a) Original acoustic signal, (b) Filtered cardiac signal, (c) IMF1, (d) IMF2, (e) IMF3, (f) IMF4, (g) Sum of all the four IMFs. n.u. - normalized units, a.u. - absolute units (after IMF decomposition).

The EMD decomposition was implemented in MATLAB using a sifting tolerance of 0.1 in the Cauchy type convergence criterion. It was empirically observed that the first four IMFs obtained from the filtered cardiac signal in Fig. 2(c)-(f) were able to reconstruct the signal in Fig. 2(g). It can be observed that the IMFs eliminates the riding waves and introduces symmetry in the data, providing meaningful instantaneous characteristics of the signal. For every IMF_m where $1 \leq m \leq 4$, the HSA computes the analytic signal (AS_m) by determining its complex conjugate (IMF_m^*), using the Hilbert transform to derive the corresponding instantaneous frequency (IF_m) and instantaneous amplitude (IA_m) as follows:

$$\text{AS}_m = \text{IMF}_m + \iota \text{IMF}_m^*, \quad \text{where } \iota = \sqrt{-1} \tag{1}$$

$$\text{IA}_m = \sqrt{(\text{IMF}_m)^2 + (\text{IMF}_m^*)^2} \tag{2}$$

$$\theta_m = \tan^{-1}\left(\frac{\text{IMF}_m^*}{\text{IMF}_m}\right) \tag{3}$$

$$\text{IF}_m = \frac{1}{2\pi} \times \frac{\mathrm{d}\theta_m}{\mathrm{d}t} \tag{4}$$

The instantaneous energy (IE_m) of an IMF is obtained by squaring its instantaneous amplitude.

$$\text{IE}_m = (\text{IA}_m)^2 = (\text{IMF}_m)^2 + (\text{IMF}_m^*)^2 \tag{5}$$

C. Instantaneous Energy Envelope

The pre-processing stage removes the breathing information, to an extent. Therefore, the remaining frequencies mainly correspond to the heart sounds. Since the IMFs are essentially a breakdown of the filtered cardiac signal, the proposed algorithm selects the instantaneous energy corresponding to the maximum instantaneous frequency among all the four IMFs to construct the energy envelope. It can be observed that higher energies are obtained for the S1 and S2 sounds in Fig. 3(a). The noisy transitions in the systolic and diastolic period of the cardiac cycle also introduce some energy peaks. Such erroneous peaks are removed by clipping the normalized energy envelope using an energy threshold shown by dash-dot line in Fig. 3(a). The initial threshold to process first 30 s of the data is set to 0.05. For n^{th} signal block of 5 s duration, the number of energy peaks are defined as T_n. The threshold for subsequent segments is decreased in steps of 0.01 (if required) until the desired number of energy peaks as defined in (6) are obtained.

$$\text{T}_n = \text{mean}(\text{T}_{n-1}, \text{T}_{n-2}, \text{T}_{n-3}) \tag{6}$$

Since the oscillations in the morphology of the heart sounds also introduce multiple energy peaks, all the peaks lying within a time distance of ± 50 ms [3] from the central peak are included to interpolate the envelope of the S1 and S2 sounds as shown in Fig. 3(b). This is done to preserve the actual time width of the fundamental heart sounds. A moving average filter with a span of 50 ms is further used to smoothen the energy envelope. Since the heart rate of 200 bpm results in a minimum time duration of 300 ms for a cardiac cycle, all the redundant peaks lying within this time frame and possessing lower amplitudes are removed.

D. S1 and S2 Detection

The energy envelope allows an easy segmentation of the cardiac cycle by producing peaks corresponding to the S1 and S2 sounds. The classification of these peaks is performed in two different scenarios as follows:

1) The energy threshold ensures that the number of peaks for the segment under consideration is, at least, T_n. For the signal blocks with T_n peaks, the amplitude 'Amp' and the time difference 'Δd' between the successive peaks (except the last) are determined. These features are used as an input to the K-means algorithm to divide all the data points in 3 different clusters, namely: S1, S2 and A (artifact). The data points denoted by 'o' and their centroids denoted by '$*$' are plotted in Fig. 3(c) using the fact that the systolic period is lower than the diastolic period for normal subjects. Depending on the comparison of the y-coordinate of A centroid (A_y) with S1 and S2 centroids (Sp_y) in (7), the cluster A is classified as follows:

$$\Delta d_{\mathrm{Sp}} = \frac{|\mathrm{Sp}_y - \mathrm{A}_y|}{\mathrm{A}_y} \quad p = 1, 2 \tag{7}$$

$$\mathrm{A} = \begin{cases} \mathrm{S1}, & \text{if } \Delta d_{\mathrm{S1}} \leq 0.25 \\ \mathrm{S2}, & \text{if } \Delta d_{\mathrm{S2}} \leq 0.25 \\ \mathrm{A}, & \text{otherwise} \end{cases} \tag{8}$$

As an illustration, the cluster A in Fig. 3(c) was classified as S1 and the corresponding annotations were assigned to the heart sounds in Fig. 3(d).

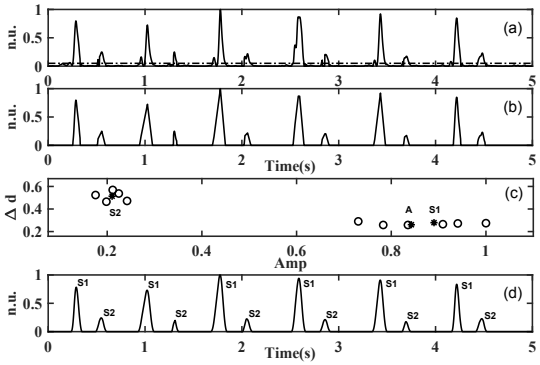

Fig. 3: Heart sounds segmentation and classification without artifacts. (a) Energy of maximum frequencies among four IMFs along with the threshold line (marked as dash-dot), (b) Interpolated energy envelope, (c) Features clustering using K-means algorithm, (d) S1-S2 classification in the averaged energy envelope.

2) The presence of artifacts in a signal block can produce a count of peaks higher than T_n. Since the artifacts occur as short bursts of energy, they usually possess lower amplitudes (Amp) and time widths (w) in the signal reconstruction after the initial filtering process, as shown in Fig. 4(a). These features are plotted in Fig. 4(b), where the data points inside the ellipse are classified as artifacts due to a lower amplitude as compared to S1 and S2 sounds. Further, the clustering algorithm as explained above is repeated to annotate the heart sounds, separating them from the artifacts in Fig. 4(c).

The time indexes t_i corresponding to the maximum amplitude of the annotated S1 and S2 sounds, denoted by '$*$' in Fig. 4(c) are recorded for every signal block. For a total of N cardiac cycles segmented and classified using the proposed algorithm, the IBI and IHR, corresponding to both heart sounds, are obtained as follows:

$$\mathrm{IBI}_i = \mathrm{t}_{i+1} - \mathrm{t}_i, \quad \text{where } i \in [1, \text{N-1}] \tag{9}$$

$$\mathrm{IHR}_i = \frac{60}{\mathrm{t}_{i+1} - \mathrm{t}_i} \tag{10}$$

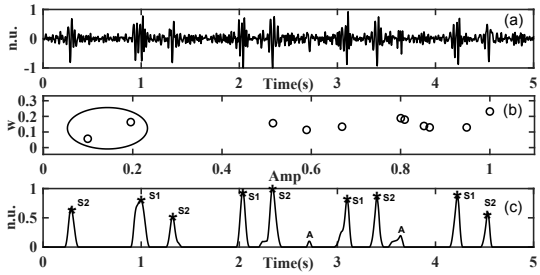

Fig. 4: Heart sounds segmentation and classification with artifacts. (a) Cardiac signal reconstructed using four IMFs, (b) Time width and amplitude of the energy peaks, (c) S1-S2 and artifact classification in the averaged energy envelope.

E. Database

The signals used to validate the algorithm, corresponding to 10 adults subjects, were from an anonymized database resulting from a study at the National Hospital for Neurology and Neurosurgery, UK. The experimental procedures involving human subjects described in this paper were approved by both, MHRA and REC and had reference number 09/H0716/2. The signals in this study were recorded at a sampling frequency of 2205 Hz whilst subjects were sleeping by placing a custom, small, wearable acoustic sensor on the suprasternal notch [10]. The ECG signal was also recorded synchronously using the FDA approved SOMNOmedics monitor [11] to provide a reference IBI and IHR time series.

III. Experimental Results

A. Time-domain HRV Analysis

The IBI variations for the RR, S1S1 and S2S2 intervals (generally known as NN intervals [1]) are quantified in Table I. These were obtained using the selected time-domain measures for a short-term recording of 5 minutes duration. A close correlation between the performance metrics of all the NN intervals can be observed. This was also validated using the Wilcoxon rank sum test which estimates the *p*-value to test the null hypothesis that the HRV estimations from the acoustic signal and the ECG data are in close agreement. Two *p*-values for RR-S1S1 and RR-S2S2 comparisons are provided in Table I. The *p*-values obtained at 5% significance level do not provide enough evidence to reject the null hypothesis concluding that the HRV measures obtained from the proposed algorithm yields a high degree of statistical agreement with the HRV estimations from the ECG signal.

B. IHR Comparisons

Table II lists the performance metrics of the IHR comparisons for the whole dataset. It can be observed that the mean absolute error (MAE), mean (μ) and standard

TABLE I: Time-domain HRV parameters for IBI estimated from ECG and acoustic breathing signals.

Subject	SDNN (ms)			SDSD (ms)			MeanNN (ms)			RMSSD (ms)			pNN50 (%)		
	RR	S1S1	S2S2	RR	S1S1	S2S2	RR	S1S1	S2S2	RR	S1S1	S2S2	RR	S1S1	S2S2
1	62.37	60.25	67.93	28.34	27.20	32.98	907.19	908.24	903.43	28.18	27.15	32.79	9.93	9.59	11.42
2	53.28	49.23	58.77	20.59	22.67	27.51	1093.68	1095.73	1086.96	20.51	22.65	27.46	7.68	6.19	4.30
3	94.84	82.08	90.35	61.25	54.37	59.21	910.52	911.99	898.07	61.08	54.28	59.10	32.30	30.74	39.41
4	74.84	74.24	75.88	54.97	55.89	53.63	1076.85	1076.17	1078.21	54.88	55.77	53.52	29.21	26.47	27.18
5	81.91	78.43	84.62	52.03	46.29	51.75	1448.61	1450.28	1455.23	51.95	46.23	51.58	23.18	22.51	25.32
6	42.57	49.26	52.31	25.91	27.20	35.82	753.02	751.73	754.35	25.89	27.17	35.76	5.59	9.15	12.39
7	45.25	44.89	45.57	24.78	24.66	25.43	834.09	834.00	835.27	24.76	24.63	25.32	9.15	9.81	9.43
8	22.93	33.31	29.91	28.46	26.00	31.49	1061.18	1067.85	1064.33	28.39	25.97	31.41	11.29	4.11	7.83
9	66.59	66.37	67.62	48.19	45.60	49.93	967.59	969.46	959.76	48.03	45.54	49.81	27.05	27.34	29.13
10	95.73	98.66	99.15	71.58	74.30	79.65	876.92	878.37	881.58	71.52	74.19	79.62	46.90	51.25	48.93
p	–	>0.05	>0.05	–	>0.05	>0.05	–	>0.05	>0.05	–	>0.05	>0.05	–	>0.05	>0.05

deviation (σ) in the Bland-Altman plot for RR-S1S1 and RR-S2S2 comparisons achieve low values demonstrating a high degree of agreement between the acoustic and ECG outputs. Fig. 5 illustrates such a comparison for an adult subject. The accuracy is determined by counting all the heart rate differences lying within a $\pm 5\%$ variation with respect to the ground truth. The higher energies of S1 sounds allowed its reliable detection, therefore, providing a high IHR accuracy of 95.78% as compared to 92.35% for S2 sounds.

TABLE II: Performance Metrics of the IHR determination for a total of 10 subjects.

Comparison	MAE (bpm)	μ (bpm)	σ (bpm)	Accuracy (%)
RR-S1S1	1.26	0.01	2.64	95.78
RR-S2S2	1.97	0.03	3.19	92.35

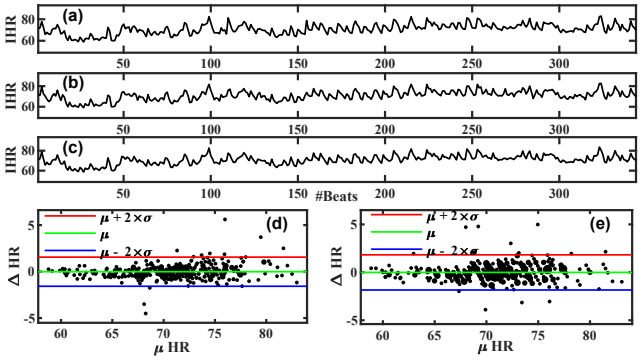

Fig. 5: IHR time series of an adult subject generated from (a) S1S1, (b) S2S2, (c) and RR intervals. (d) Bland-Altman comparison for S1, (e) and S2 sounds with respect to ECG outputs.

IV. Conclusion

This work proposed an algorithm to utilize the breathing sounds recorded at neck, and localize the fundamental heart sounds to determine the HRV of an adult subject. The cardiac cycle was processed using an empirical data analysis approach which provided meaningful representations of the non-stationary cardiac signal. The validation of the study included both the IHR comparisons and the time-domain HRV analysis using standard measurements defined by the Task Force [1]. The algorithm achieved a high IHR accuracy and obtained time-domain HRV measures in close agreement with the ECG signal, despite the absence of a specific characteristic point in the morphology of the heart sounds. This is however just a pilot study and hence, full future validation would require a greater number of subjects.

Acknowledgment

The research leading to these results has received funding from European Research Council (ERC), grant agreement no. 724334.

References

[1] M. Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, and P. J. Schwartz, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *European heart journal*, vol. 17, no. 3, pp. 354–381, 1996.

[2] E. Gil, M. Orini, R. Bailón, J. Vergara, L. Mainardi, and P. Laguna, "Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions," *Physiological measurement*, vol. 31, no. 9, p. 1271, 2010.

[3] P. Sharma, S. A. Imtiaz, and E. Rodriguez-Villegas, "An algorithm for heart rate extraction from acoustic recordings at the neck," *IEEE Transactions on Biomedical Engineering*, 2018.

[4] A. K. Dwivedi, S. A. Imtiaz, and E. Rodriguez-Villegas, "Algorithms for automatic analysis and classification of heart sounds a systematic review," *Access, IEEE*, vol. PP, no. 99, pp. 1–1.

[5] A. Jiménez, M. Ortiz, M. Peña, S. Charleston, R. González, A. Aljama, and S. Carrasco, "Performance of a method to generate fetal cardiotachograms using fetal phonocardiography," *Computers in Cardiology*, pp. 453–456, 2001.

[6] R. Ortiz, R. González, M. Peña, S. Carrasco, M. Gaitán, and C. Vargas, "Differences in foetal heart rate variability from phonocardiography and abdominal electrocardiography," *Journal of medical engineering & technology*, vol. 26, no. 1, pp. 39–45, 2002.

[7] H. Liang, S. Lukkarinen, and I. Hartimo, "Heart sound segmentation algorithm based on heart sound envelogram," in *Computers in Cardiology 1997*. IEEE, 1997, pp. 105–108.

[8] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings: Mathematical, Physical and Engineering Sciences*, pp. 903–995, 1998.

[9] P. Sharma and K. C. Ray, "Efficient methodology for electrocardiogram beat classification," *IET Signal Processing*, vol. 10, no. 7, pp. 825–832, 2016.

[10] E. Rodriguez-Villegas, G. Chen, J. Radcliffe, and J. Duncan, "A pilot study of a wearable apnoea detection device," *BMJ open*, vol. 4, no. 10, p. e005299, 2014.

[11] SOMNOmedics GmbH : (2014) SOMNOscreen. [Online]. Available: http://www.somnomedics.eu/