Deep Learning-Based Automated Extraction of Cardiac Time Intervals from Electrocardiogram and Phonocardiogram for Heart Failure Assessment

Mariana Lourenço[1], Cátia Isabel Costa[2], Cristina Oliveira[3], André Lobo[2], Ricardo Fontes-Carvalho[3], Francesco Renna[1]

[1] INESC TEC, Faculty of Sciences of the University of Porto, Porto, Portugal
[2] Unidade Local de Saúde Gaia e Espinho, Porto, Portugal
[3] UnIC@RISE, Porto, Portugal

Abstract

Cardiac time intervals (CTIs) are established markers of systolic function that correlate with left ventricular ejection fraction (LVEF), but their clinical assessment often involves echocardiography. This work develops a fully automated framework for CTI extraction from synchronous electrocardiogram (ECG) and phonocardiogram (PCG) recordings and evaluates its ability to discriminate heart failure (HF) phenotypes.

Two automated pipelines were developed and compared: (i) a beat-by-beat (BbB) approach, in which U-Net models segment ECG (P, QRS, T) and PCG (S1, S2) events from multi-domain features, and (ii) a median heartbeat (MHB) approach, in which beats are aligned and averaged prior to segmentation to improve robustness. CTIs (QT, QS1, QS2, S1S2) were computed from the detected fiducial points and evaluated against manual annotations from 84 synchronous ECG-PCG recordings, stratified by LVEF-defined groups.

The framework achieved good agreement with manual annotations (MAE: S1S2 = 17.3 ms, QS2 = 46.0 ms), with the MHB approach markedly more robust than BbB, especially for PCG-derived intervals. QS2 and S1S2 were the most discriminative of HF status (Mann-Whitney U, mildly-reduced vs. preserved EF: $p = 0.043$ and $p = 0.125$), while QS1 showed moderate and QT weaker discrimination.

1. Introduction

Cardiovascular diseases remain the leading cause of death worldwide, accounting for an estimated 17.9 million deaths in 2019 (32% of all deaths) [1]. Among these, heart failure (HF) affects over 64 million people and imposes a substantial burden on healthcare systems [2]. Echocardiography (ECHO) is the clinical gold standard for assessing systolic function through cardiac time intervals (CTIs) and left ventricular ejection fraction (LVEF) [3], but it requires costly equipment and specialized personnel, which limits its accessibility and makes it operator-dependent [4].

Electrocardiography (ECG) and phonocardiography (PCG) are non-invasive, low-cost alternatives that can be combined to estimate CTIs from the timing of electrical (P, QRS, T) and acoustic (first and second heart sound, S1/S2) cardiac events [5,6]. Intervals such as the QT interval, the pre-ejection-period-related QS1 interval, the total electromechanical systole (QS2), and the left-ventricular-ejection-time-related S1S2 interval can be estimated from the timing of ECG and PCG events [7,8], and a manually-annotated analysis of the same patient cohort studied here showed that several of these intervals correlate with LVEF and differ across HF phenotypes [9]. Recent advances in deep learning have enabled automatic segmentation of ECG [10] and PCG [11] waveforms, raising the possibility of a fully automated CTI-based monitoring tool.

This paper extends our previous manually-annotated analysis [9] by presenting a fully automated framework for CTI extraction from synchronous ECG and PCG recordings. We compare a beat-by-beat (BbB) U-Net segmentation pipeline with a robust median heartbeat (MHB) pipeline, evaluating both against manual annotations and across LVEF-stratified patient groups.

2. Materials and Methods

2.1. Dataset

Synchronous ECG (500 Hz) and PCG (3000 Hz) recordings were acquired with an FDA-approved multimodal digital stethoscope (Rijuven CardioSleeve), in collaboration with the Cardiology Department of the Unidade Local de Saúde Gaia e Espinho (ULSGE). Recordings were collected at the aortic, pulmonary, tricuspid and mitral auscultation sites ($\sim$30 s each), yielding 558 ECG-PCG pairs from 167 subjects (2-93 years, $\sim$60% male). LVEF

group labels – reduced (rEF, LVEF $< 40\%$), mildly reduced (mrEF, 40-50%) and preserved (pEF, LVEF $> 50\%$) – were available for 160 recordings. Of these, 84 recordings with well-defined fiducial points (4 rEF, 26 mrEF, 54 pEF) were manually annotated with QRS onset, T-wave onset, and S1/S2 timings on the median heartbeat template [9], from which reference QT, QS2 and S1S2 intervals were derived; this subset served as ground truth for both accuracy assessment and EF-based analysis.

2.2. Signal preprocessing

ECG signals were notch-filtered at 50 Hz ($Q = 30$), band-pass filtered (6th-order Butterworth, 0.5-100 Hz) and median-corrected. PCG signals were downsampled to 1 kHz using a polyphase FIR filter and despiked using the Schmidt method [12]. Both signals were z-score standardized prior to further processing.

2.3. Beat-by-beat (BbB) approach

Two U-Net models (one for ECG, one for PCG), adapted from a heart-sound segmentation architecture [11], classify each sample into four classes ({baseline, P, QRS, T} for ECG; {S1, systole, S2, diastole} for PCG). Each signal was converted into four 50 Hz feature envelopes (ECG: Hilbert, Shannon energy, homomorphic and smoothed envelopes; PCG: homomorphic and Hilbert envelopes plus Morlet and Mexican-hat wavelet scalograms, 20-200 Hz), stacked into 64-sample patches (8-sample stride) and fed to an encoder-decoder U-Net (four down-/up-sampling blocks, 16-256 feature maps, softmax output). The ECG model was trained on the LUDB database [13]; the PCG model on the PhysioNet/CinC 2016 database [14].

Event segments (start, end, center) were extracted per class, and heartbeat windows were defined around reference times (QRS-segment midpoint for BbB) with width equal to the subject's median RR interval. Events were assigned to the window containing their center time (or of maximal temporal overlap at boundaries). Within each window, only combinations of events following the expected physiological order (Q before T-offset, before S2-center, before S1-center; S1 before S2) were retained, from which the QT, QS1, QS2 and S1S2 intervals were computed (S2-center $-$ S1-center; T-end $-$ Q-start; S2-center $-$ Q-start; S1-center $-$ Q-start, respectively).

2.4. Median heartbeat (MHB) approach

R-peaks were detected with NeuroKit2 [16], chosen after benchmarking against a Pan-Tompkins implementation [15] and three other open-source detectors. Validated against manual R-peak annotations on 69 recordings (50 ms tolerance), NeuroKit2 achieved precision 0.948, recall 0.969, F1-score 0.958 and mean timing error 2.0 ± 4.0 ms. For each subject, ECG windows (-0.4/$+1.0$ s around each R-peak) were extracted and averaged samplewise into a median ECG beat; PCG windows defined by the same R-peak timings were similarly averaged into a median PCG beat, exploiting the ECG-PCG synchronization. Each median beat pair was then segmented using the same U-Net models and post-processing pipeline as the BbB approach (Section 2.3) to extract one set of CTIs per recording from a single, noise-robust representative cardiac cycle. Fig. 1 summarizes both pipelines.

2.5. Statistical analysis

Estimated CTIs were compared with manual annotations using mean absolute error (MAE), root-mean-square error (RMSE) and the Pearson correlation coefficient (r); for the BbB approach, per-recording estimates were averaged across heartbeats for comparability with the single MHB estimate per recording. The coefficient of determination (R^2), Spearman correlation and Bland-Altman bias/limits of agreement (LoA) were also examined and are reported selectively, where most informative; correlation significance was assessed via a two-tailed t-test. Differences between LVEF groups (mrEF vs. pEF) were assessed with the Mann-Whitney U test; the rEF group ($n = 4$) was excluded from formal testing due to its small sample size.

3. Results

3.1. Estimation accuracy

Table 1 summarizes the agreement between automated and manually annotated CTIs. The MHB approach achieved the best accuracy for the PCG-anchored intervals: S1S2 MAE $= 17.3$ ms ($r = 0.80$, $p < 0.0001$) and QS2 MAE $= 46.0$ ms ($r = 0.40$, $p = 0.0003$; Spearman $\rho = 0.71$, $p < 0.0001$), versus MAE $= 98.5$ ms and 131.2 ms for BbB, whose estimates showed no significant correlation with ground truth for either interval. Bland-Altman analysis confirmed markedly narrower limits of agreement for MHB (S1S2: bias -11.4 ms, LoA $[-55, +32]$ ms; QS2: bias $+42.3$ ms, LoA $[-40, +120]$ ms) than for BbB (S1S2 LoA $[-400, +300]$ ms; QS2 LoA $[-600, +600]$ ms), the latter dominated by outliers. For the ECG-only QT interval, BbB slightly outperformed MHB ($R^2 = 0.31$ vs. 0.20; $r = 0.59$ vs. 0.48), both with comparable, narrower limits of agreement ($\sim \pm 75$ ms). For QS1, MHB showed a systematic overestimation bias ($+62.8$ ms) and weak, non-significant correlation with ground truth ($r = 0.16$, $p = 0.14$); BbB estimates were consistently and substantially off and are omitted as unreliable.

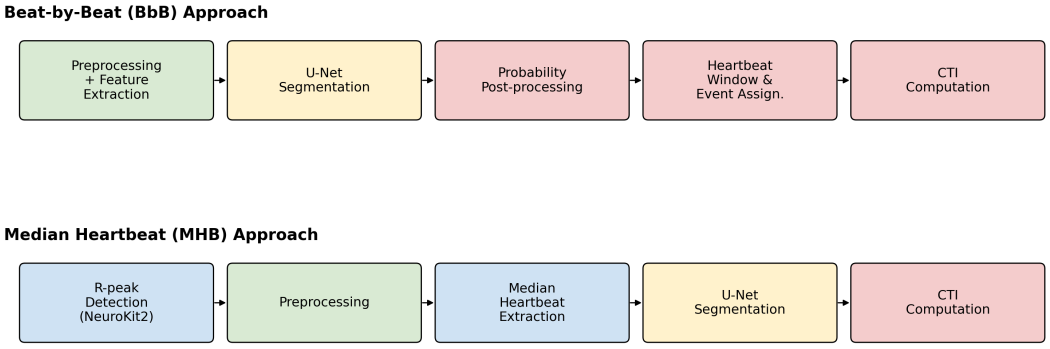

Figure 1. Overview of the beat-by-beat (top) and median heartbeat (bottom) segmentation pipelines. Common building blocks (preprocessing, U-Net segmentation, CTI computation) are shared between the two approaches.

Table 1. Estimation accuracy (vs. manual annotations) and EF-based discrimination (Mann-Whitney U, mrEF vs. pEF) for the median heartbeat (MHB) and beat-by-beat (BbB) approaches. Better-performing method per row pair shown in bold. BbB/QS1 (–): estimates were unreliable and omitted.

Interval	Method	N	MAE (ms)	RMSE (ms)	r	U	p
S1S2	MHB	82	**17.3**	**24.2**	**0.80**	**529.5**	**0.125**
	BbB	44	98.5	173.4	0.02	184.5	0.955
QT	MHB	84	30.1	42.5	0.48	724.5	0.819
	BbB	84	**30.0**	**39.6**	**0.59**	**861.0**	**0.104**
QS2	MHB	80	**46.0**	**58.9**	**0.40**	**464.5**	**0.043**
	BbB	48	131.2	332.6	0.02	265.0	0.342
QS1	MHB	83	215.4	241.5	0.16	571.5	0.274
	BbB	–	–	–	–	–	–

3.2. EF-based discrimination

Fig. 2 shows the MHB-derived S1S2 and QS2 distributions stratified by LVEF group; both intervals were descriptively shorter in the small reduced-EF group ($n = 4$), consistent with impaired systolic function, although this group could not be formally tested. Between the mrEF and pEF groups, the MHB-derived QS2 interval differed significantly (Mann-Whitney $U = 464.5$, $p = 0.043$), and S1S2 showed a similar but non-significant trend ($U = 529.5$, $p = 0.125$); neither interval discriminated the groups when estimated with BbB ($p = 0.955$ and $p = 0.342$). QS1 showed weak discrimination for MHB ($p = 0.274$; BbB omitted, cf. Section 3.1), and QT showed no discrimination for either method ($p = 0.819$, $p = 0.104$).

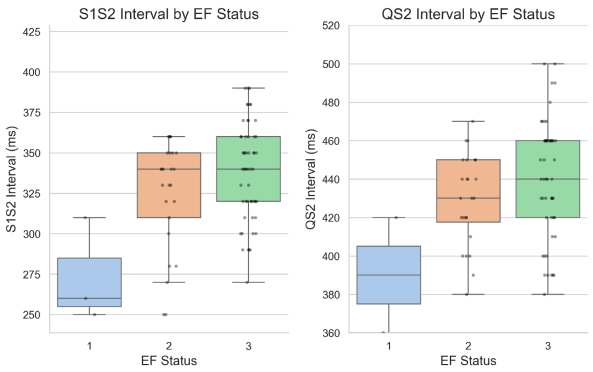

Figure 2. MHB-derived S1S2 (left) and QS2 (right) intervals stratified by LVEF group (1 = rEF, $n = 4$; 2 = mrEF, $n = 26$; 3 = pEF, $n = 54$).

4. Discussion

Consistent with our previous manual-annotation-based analysis [9], QS2 and S1S2 emerged as the CTIs most sensitive to LVEF status, while QT and QS1 were less discriminative. Importantly, these relationships were preserved when CTIs were estimated fully automatically, provided the MHB strategy was used.

The clear advantage of MHB over BbB for PCG-anchored intervals (S1S2, QS2, QS1) reflects the difficulty of segmenting individual heart sounds: S1/S2 morphology is more variable and noise-sensitive than ECG mor-

phology, so per-beat U-Net predictions are more error-prone, and averaging across the cardiac cycle before segmentation suppresses this variability. Conversely, for the ECG-only QT interval, BbB matched or slightly exceeded MHB accuracy, indicating that per-beat segmentation is adequate, and even advantageous, when only the cleaner ECG modality is involved.

Several limitations should be noted. The 20 ms effective temporal resolution of the U-Net feature envelopes limits precision for short intervals such as QS1. ECG and PCG are processed independently, so errors in either modality propagate to cross-modal intervals (QS2, QS1); a fused multimodal architecture could mitigate this. The reduced-EF group was too small ($n = 4$) for formal statistical testing, and, unlike MHB, the BbB approach cannot yet reliably compute instantaneous, RR-corrected CTIs, which are known to correlate more strongly with EF than uncorrected values. Larger, more balanced datasets and multimodal fusion are natural next steps.

5. Conclusion

We presented a fully automated deep-learning framework for extracting cardiac time intervals from synchronous ECG and PCG recordings, comparing beat-by-beat and median-heartbeat segmentation strategies. The median heartbeat approach provided robust, clinically meaningful CTI estimates – particularly for PCG-derived intervals – with QS2 and S1S2 retaining their ability to discriminate HF phenotypes even without manual annotation. These results support the feasibility of low-cost, automated CTI monitoring as a scalable complement to echocardiography for HF assessment.

Acknowledgments

This work is co-funded by the European Regional Development Fund (ERDF) through the Innovation and Digital Transition Programme (COMPETE 2030) under Portugal 2030 and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) within project PULSE, with reference 17172 (COMPETE2030-FEDER-00864900) (doi.org/10.54499/17172-COMPETE2030-FEDER-00864900).

References

[1] World Health Organization. Cardiovascular diseases (CVDs). WHO Fact Sheets, 2019.

[2] Savarese G, Lund LH. Global public health burden of heart failure: an updated review. Cardiac Failure Review 2017;3(1):7-11.

[3] Paiva RP, Carvalho P, Couceiro R, Henriques J, Antunes M, Quintal I, Muehlsteff J. Beat-to-beat systolic time-interval measurement from heart sounds and ECG. Physiol Meas 2012;33(2):177-94.

[4] Lewis RP, Rittogers SE, Froester WF, Boudoulas H. A critical review of the systolic time intervals. Circulation 1977;56(2):146-58.

[5] Rafie N, Kashou AH, Noseworthy PA. ECG interpretation: clinical relevance, challenges, and advances. Hearts 2021;2(4):505-13.

[6] Zhu B, Zhou Z, Yu S, Liang X, Xie Y, Sun Q. Review of phonocardiogram signal analysis: insights from the PhysioNet/CinC challenge 2016 database. Electronics 2024;13(16):3222.

[7] Klum M, Urban M, Tigges T, Pielmus AG, Feldheiser A, Schmitt T, Orglmeister R. Wearable cardiorespiratory monitoring employing a multimodal digital patch stethoscope: estimation of ECG, PEP, LVET and respiration using a 55 mm single-lead ECG and phonocardiogram. Sensors 2020;20(7):2033.

[8] Roos M, Toggweiler S, Zuber M, Jamshidi P, Erne P. Acoustic cardiographic parameters and their relationship to invasive hemodynamic measurements in patients with left ventricular systolic dysfunction. Congest Heart Fail 2006;12(s1):19-24.

[9] Petersen FT, Lobo A, Oliveira C, Costa CI, Fontes-Carvalho R, Schmidt SE, Renna F. Predicting the left ventricular ejection fraction using bimodal cardiac auscultation. Computing in Cardiology 2025.

[10] Jimenez-Perez G, Alcaine A, Camara O. U-Net architecture for the automatic detection and delineation of the electrocardiogram. Computing in Cardiology 2019.

[11] Renna F, Oliveira J, Coimbra MT. Deep convolutional neural networks for heart sound segmentation. IEEE J Biomed Health Inform 2019;23(6):2435-45.

[12] Schmidt SE, Holst-Hansen C, Graff C, Toft E, Struijk JJ. Segmentation of heart sound recordings by a duration-dependent hidden Markov model. Physiol Meas 2010;31(4):513-29.

[13] Kalyakulina A, Yusipov I, Moskalenko V, Nikolskiy A, Kosonogov K, Zolotykh N, Ivanchenko M. Lobachevsky University Electrocardiography Database (LUDB), v1.0.1. PhysioNet, 2021.

[14] Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ, et al. An open access database for the evaluation of heart sound algorithms. Physiol Meas 2016;37(12):2181-213.

[15] Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Trans Biomed Eng 1985;32(3):230-6.

[16] Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, Schölzel C, Chen SHA. NeuroKit2: a Python toolbox for neurophysiological signal processing. Behav Res Methods 2021;53(4):1689-96.

Address for correspondence:

Francesco Renna
Faculty of Sciences of the University of Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal
francesco.renna@fc.up.pt