German Medical Data Sciences 2023 — Science. Close to People.
R. Röhrig et al. (Eds.)

doi:10.3233/SHTI230718

ECG Matching: An Approach to Synchronize ECG Datasets for Data Quality Comparisons

Mohamed ALHASKIR[a,c,1*], Matteo TSCHESCHE[b*], Florian LINKE[c], Elisabeth SCHRIEWER[c], Yvonne WEBER[c], Stefan WOLKING[c], Rainer RÖHRIG[a], Henner KOCH[c], and Ekaterina KUTAFINA[a]

[a] *Institute for Medical Informatics, Medical Faculty, RWTH Aachen University, Aachen, Germany*

[b] *Institute for Mobile Autonomous Systems and Cognitive Robotics, FH Aachen University of Applied Sciences, Aachen, Germany*

[c] *Department of Epileptology, Neurology, Medical Faculty, RWTH Aachen University, Aachen, Germany*

Abstract. Clinical assessment of newly developed sensors is important for ensuring their validity. Comparing recordings of emerging electrocardiography (ECG) systems to a reference ECG system requires accurate synchronization of data from both devices. Current methods can be inefficient and prone to errors. To address this issue, three algorithms are presented to synchronize two ECG time series from different recording systems: Binned R-peak Correlation, R-R Interval Correlation, and Average R-peak Distance. These algorithms reduce ECG data to their cyclic features, mitigating inefficiencies and minimizing discrepancies between different recording systems. We evaluate the performance of these algorithms using high-quality data and then assess their robustness after manipulating the R-peaks. Our results show that R-R Interval Correlation was the most efficient, whereas the Average R-peak Distance and Binned R-peak Correlation were more robust against noisy data.

Keywords. Electrocardiography, wearable electronic device, sensors comparison, time-series synchronization

1. Introduction

Performance testing of emerging medical sensors is a crucial element in clinical validation to ensure the fulfillment of functional requirements [1]. For newly developed ECG systems, device readings are compared with those of a routinely used ECG system [1-3]. Ideally, the comparison is conducted under controlled conditions, with both the device under validation and the reference device tested on the same patient [1,2]. However, to perform quality analyses and other comparative measures, data from both devices must first be accurately synchronized. In its simplest form, synchronization can

[1] Corresponding Author: Mohamed Alhaskir, malhaskir@ukaachen.de, Institute of Medical Informatics, University Hospital RWTH Aachen, Pauwelsstr. 30, 52074 Aachen, Germany.
* Both authors contributed equally

be achieved manually, by starting the measurements simultaneously or documenting each device's starting time. Both approaches are error-prone and can be challenging and inefficient for large datasets. Applying Cross-Correlation (CC) for aligning two-time series entails substantial time complexity and memory consumption, especially when dealing with time series lasting hours or even days. Similarity search methods such as Euclidean Distance (ED) and Dynamic Time Warping (DTW) for data alignment present similar challenges. While some work has been done to optimize these methods for pattern detection in large datasets with exceptionally short times [4,5], DTW and ED methods have especially focused on the detection of specific patterns in the data. Thus, even if the data samples were generated from the same patient at the same time, such methods could be compromised by the discrepancy of the time series generated from different systems due to variations in noise and the preprocessing steps implemented in each of them. We propose reducing ECG data to their cyclic features (R-peaks) to mitigate inefficiencies, decrease computational complexity, and minimize data discrepancies arising from noise and disparate data processing methods. Furthermore, we adapted the work done on neural spike pattern analysis using correlation between binned spike times and intraspike intervals [6] for ECG data alignment. We refer to the implemented approaches as Binned R-peak Correlation (BRC) and R-R Interval Correlation (RRIC). Additionally, we implemented a method by [7] which focuses on minimizing the time differences between corresponding R-peaks to align the ECG segments. We refer to this method in this work as Average R-peak Distance (ARD). The objective of this study is to compare the performance of the three algorithms in alignment of ECG time series from different recording systems. For this purpose, the following questions for each approach must be answered:

1. How does the computational time of the algorithm vary with respect to input data length (number of R-peaks)?
2. What is the level of accuracy (i.e. the average time difference between the estimated match and the true match) that can be achieved by the algorithm?
3. How robust is the algorithm against mislabeled R-peaks?

2. Methods

2.1. Data

The data in this work were collected at the epileptology clinic at RWTH Aachen University hospital in the frame of a project on seizure detection. We used the names ECGa and ECGb to differentiate between the data recorded with the to-be-validated sensor and the ECG data recorded at the epilepsy clinic, respectively. ECGa was recorded with a wearable sensor produced by movisens GmbH (Karlsruhe, Germany). On the other hand, the Micromed® Group's EEG system SD LTM64 EXPRESS, which allows for parallel recording of a single ECG channel, was utilized to record ECGb data. Details of the two devices are presented in table 1.

We tested the proposed algorithms on two patients' datasets, who were recorded simultaneously with both devices. Specifically, each patient was recorded for a period of two days using the wearable sensor and during this period, three ECGb recordings were conducted. Each of the ECGb had a duration between one and four hours.

Table 1. Specification of the two devices used in the study.

	ECGa	ECGb
Device	EcgMove4	SD LTM64 EXPRESS
Software	SensorManager 1.12.7	SystemPlus EVOLUTION 1.06.0005
Manufacturer / Vendor	movisens GmbH	Micromed® Group
Sampling Frequency [s^{-1}]	1024	256
Amplitude Range [mV]	+/-5	+/-3.2
Analog-digital Conversion [bit]	12	16
Filters	Bandwidth (1.6 Hz - 33Hz)	Bandwidth (0.16 Hz - 70 Hz)

2.2. Hardware and Software

This work was evaluated on a Dell Latitude 5421 with an Intel i7-11850H Version 6.141.1 processor running at 2.50 GHz and 15726 MB RAM. The code was implemented in a Jupyter Notebook 6.5.2 using Python 3.10.6 and the libraries Scipy 1.10.1, numpy 1.24.1 and biosppy 1.0.0.

2.3. Algorithms

2.3.1. R-R Interval Correlation

In the RRIC algorithm, the relation between the R-R interval in both datasets for each time shift is evaluated using Pearson Correlation. A high correlation score would signify a possible match, whereas a low correlation value suggests the opposite.

2.3.2. Average R-peak Distance

In the context of ARD, the discrepancies between the nearest R-peaks of both datasets are calculated and subsequently averaged at each step (Figure 1A). A minimal average R-peak distance suggests that the R-peaks in both sets occur at analogous time points, signifying a potential alignment.

2.3.3. Binned R-peak Correlation

For the BRC method, time series are segmented into 100 ms bins, in which the presence of an R-peak is indicated by 1 and its absence by 0 (Figure 1B). The binned R-peaks data are subsequently assessed via Pearson's correlation coefficient. A high correlation implies that R-peaks predominantly occupy the same bins, indicating a strong alignment between the two datasets. In contrast, a low correlation shows that R-peaks are rarely placed in the same or near bin and that both datasets are not matching in the given time step.

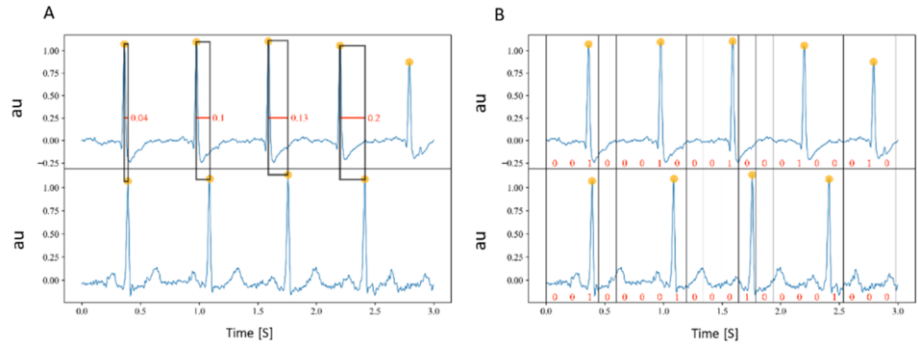

Figure 1. A) Difference between each R-peak and its nearest neighboring R-peak in the adjacent ECG time series is calculated. These R-peak differences are then utilized to compute the ARD at a specific time step. B) ECG time series data are binned using a 100 ms window and assigned labels of 1 or 0 based on the presence or absence of an R-peak, respectively.

2.3.4. Search Phases

In ARD and BRC, assigning each time step as a potential synchronization point may result in significant processing time. As such, these two methods initiate the search for the optimal match using an increased step size, defined here as $1/8^{th}$ of the average R-R Interval. By doing this, timepoints with a noteworthy matching score (high correlation in RRIC or low Average R-peak distance in ARD) that suggest a possible alignment are gathered. The search for the ideal match then proceeds in the next phase, where only the areas surrounding these potential matching timepoints are examined (Figure 2).

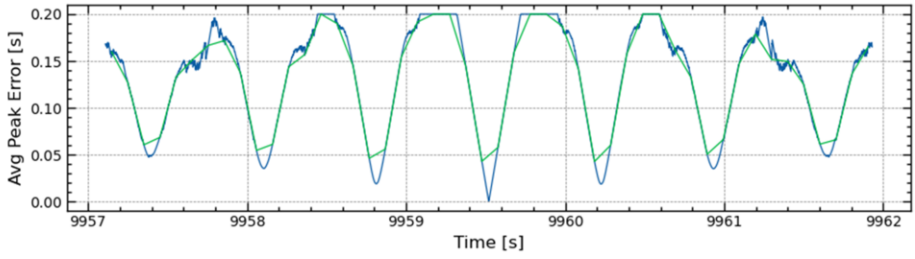

Figure 2. Average peak error of ARD method around a timepoint of interest. The blue and green lines show the average R-peak distance at different timesteps. The green line represents the first phase where a moving step of 1/8th of the average R-R Interval was used and the blue one shows the results of the second phase where the average R-peak distance for every possible time step around a candidate is computed.

In our work, we defined these possible datapoints of interest as the 50 top timesteps with the highest match score. The RRIC method, on the other hand, uses a single-phase search process that assesses the correlation in one iteration. This approach is sufficient because all potential synchronization points are boiled down to the R-peaks, disregarding the time steps in-between and leading to an expedited algorithm run time.

2.4. Evaluation Methods

Our evaluation process is divided into two separate stages. In the initial evaluation step, we established two metrics: 1) the time difference between the estimated match time and

the true match time, and 2) the algorithm's execution time. At this stage, we assessed the intrinsic performance of the algorithms utilizing manually annotated R-peaks.

In the second evaluation step, we aimed to emulate real-world data conditions by employing an automated R-peak detection algorithm and manipulating the quality of the data to examine the robustness of the proposed methodologies. Therefore, we implemented approaches to simulate the results of peak detection on noisy ECG data. These methods are random R-peak shifting, R-peak removal (random and clustered), and random R-peak addition. In the random R-peak shifting approach, each R-peak in ECGa is moved around its position by a given percentage of an average R-R interval. For the R-peak removal approach, a percentage of all R-peaks in ECGa are removed using either an equally distributed probability (random) or a distribution which favors clusters by increasing the probability of removing neighboring R-peaks by 3, 4, 7, 10 folds after removing a single R-peak (clustered). Finally, R-peak addition was performed by randomly adding a percentage of the R-peaks in the ECGa using a uniform distribution.

3. Results

3.1. First Evaluation

The data of patient 1 were used for the first evaluation. The patient dataset consists of an ECG device that was measuring continuously for two days (ECGa) while another device was capturing the ECG signal (ECGb) simultaneously at three different time points (n = 3) within these two days. Every R-peak is hand-labeled in the ECGa to allow the evaluation of the algorithm's performance in the best-case scenario. The performance was tested using different R-peaks length from ECGb and the evaluation metrics were computed for each approach at each R-peaks length (see Table 2).

Table 2. Evaluation table with the average time difference between the best match and the true match as well as the average run time of each approach while matching 10, 100, and 1000 peaks from three distinct ECGb to a 48h long ECGa. Differences greater than 3000 ms were labeled as 'N/A' and were considered to be failed matches.

Approach	Number of Used Peaks	Average Time Difference [ms]	Average Run Time [s]
ARD	10	N/A	29.994
	100	0	293.029
	1000	2	3175.830
BRC	10	N/A	1686.800
	100	0	1744.230
	1000	1	2191.860
RRIC	10	N/A	12.763
	100	2	13.065
	1000	265	14.269

Table 2 shows the results of the first evaluation scenario in which the run time of each approach is shown for different amounts of R-peaks. RRIC is the fastest approach and its run time remained relatively stable even when using a higher number of R-peaks. ARD had a lower run time in comparison to BRC when using fewer R-peaks but is slower for increasing numbers of R-peaks. The table additionally shows the difference between the true match and the algorithm’s result. When using only 10 R-peaks, none of the algorithms were able to find a match. Using 100 R-peaks showed the best results and

using more peaks resulted in larger differences from the true match, especially for the RRIC approach.

3.2. Second Evaluation

In the second evaluation, the data of the second patient were used. Since this data were unlabeled, we automatically labeled it using the Hamilton R-peak detection algorithm [8]. For the evaluation, 100 R-peaks from three different ECGb (n1, n2 and n3) were matched against a one-hour ECGa. Here, the robustness of each method is assessed using increasing levels of simulated noise using the above introduced data manipulation methods.

Table 3. Evaluation table of the approach's sensitivity to noise in form of peak-shifting, peak-removal using an equal distribution and a distribution which forces clustered removals, and peak-addition. The table shows the time difference between the correct match and the result of the approaches for three ECGb (n1, n2 and n3) under a certain percentage of noise in ECGa. Differences greater than 3000 ms were labeled as 'N/A' and were considered to be failed matches.

Approach	**% Noise**	**Peak-Shifting [ms]**			**Peak-Removal (Equally Distributed) [ms]**			**Peak-Removal (Clustered) [ms]**			**Peak-Addition [ms]**		
ECGb		n1	n2	n3	n1	n2	n3	n1	n2	n3	n1	n2	n3
ARD	0	1	1	0	1	1	0	1	1	0	1	1	0
	1	0	N/A	1	1	1	0	1	1	0	1	N/A	0
	5	6	9	3	1	1	0	1	1	0	429	687	0
	10	21	N/A	N/A	1	1	0	1	1	0	1	737	0
	20	N/A	N/A	N/A	1	1	0	1	0	0	1	1	0
	50	N/A	N/A	N/A	2	N/A	0	N/A	N/A	0	1	1	0
BRC	0	0	5	3	0	5	3	0	5	3	0	5	3
	1	0	5	1	0	5	3	0	5	3	0	5	3
	5	19	10	23	0	5	3	0	5	1	0	5	3
	10	11	1142	27	0	5	3	0	5	3	0	5	3
	20	14	1131	145	0	5	3	0	5	3	0	5	3
	50	N/A	N/A	N/A	0	N/A	N/A	0	1116	3	0	5	3
RRIC	0	N/A	0	0	N/A	0	0	N/A	0	0	N/A	0	0
	1	N/A	1	7	N/A	N/A	0	N/A	0	0	N/A	1130	0
	5	N/A	7	52	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1173
	10	N/A	100	98	N/A	N/A	N/A	N/A	N/A	2259	N/A	N/A	N/A
	20	N/A	N/A	113	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	50	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 3 shows the results of the noise sensitivity evaluation. Each approach is compared with different levels of noise. The RRIC approach is very sensitive to even small levels of noise except for peak-shifting noise. ARD and BRC show a high robustness against most noise types except for peak-shifting.

4. Discussion

In summary, three simple algorithms, ARD, BRC, and RRIC are explored to match ECG datasets from different systems automatically. The RRIC had a high accuracy and low run time if the considered R-peaks on ECGa are perfectly detected. However, in contrast to ARD and BRC, RRIC does not match properly if R-peaks are missing or incorrectly added as tested in our noise sensitivity evaluation. Also, using a low number of R-peaks (100 R-peaks) achieved a better matching result (0-2 ms deviation) than using 1000 R-peaks in all methods.

As illustrated in Table 2, when using a very low number of R-peaks (10 R-peaks), none of the algorithms were able to detect a possible match. This can be explained by the inadequacy of the information contained in these short segments, resulting in inaccurate matches. The deviation from the true match for high peak numbers (1000 R-peaks) in RRIC can be explained by the algorithm identifying an offset with the highest correlation. However, for the true match, the correlation may not be as high. This is mainly due to long-term trends such as varying sampling rates or minor time shift between both devices caused by factors including battery status and temperature changes [7].

Further, RRIC's susceptibility to noise compromises its robustness (see Table 3). The RRIC's sensitivity is likely due to the distorted correlation by the added noise. This limitation poses challenges when applied to real-world data, which are inherently presumed to be noisy. The ARD and BRC techniques display comparable accuracy and speed for both 100 and 1000 R-peak datasets while exhibiting greater robustness in comparison to RRIC. In the case of the ARD, the higher robustness is likely due to the averaging effect, which focuses on the overall alignment of the ECG time series and remains relatively unaffected by low levels of noise in the data. Additionally, with BRC, the correlation between both the presence and absence of R-peaks are considered, while the RRIC solely emphasizes the time intervals between R-peaks. These distinctions make RRIC less robust to noise compared to the BRC and ARD.

The explored and examined ECG data alignment methods presented in this work may not be as fast as DTW and ED [4,5]. Nevertheless, they circumvent the variability exhibited between different ECG systems to align the data by reducing it to their cyclic features. Conversely, DTW and ED focus on identifying distinct patterns within the data to determine the alignment time point, which can be difficult considering the dissimilarity of the segments due to their origin from various sources. The noise sensitive behavior of the algorithm can be further improved by detecting and preprocessing erroneous RR-intervals before carrying out time series alignment. Also, to improve the run times, the algorithms could be parallelized by using multiple kernels. With these modifications, the proposed algorithms seem to have high potential for efficient and robust data aligning between different ECG systems. We are working on testing the performance of the developed methodologies on a larger dataset to test the validity of our preliminary findings.

Declarations

Ethical vote: This project was approved by the ethical committee of the RWTH Aachen University (Vo.-No 381/19, chair: Günther Schmalzing).

Conflict of Interest: The authors declare that there is no conflict of interest.

Author contributions: MA, EK: conception of the work; MA: study design; YW, SW, RR, HK, EK: project supervision; ES, FL: data acquisition; MA, MT: data analysis and interpretation; MA, MT: writing the manuscript; ES, YW, SW, RR, HK, EK: substantial revising of the manuscript. All authors approved the manuscript in the submitted version and take responsibility for the scientific integrity of the work.

Acknowledgement: Funded by German Federal Ministry of Education and Research (No 13GW0576C).

References

[1] Fensli R, Gundersen T, Snaprud T, Hejlesen O. Clinical evaluation of a wireless ECG sensor system for arrhythmia diagnostic purposes. Med Eng Phys. 2013 Jun, doi: 10.1016/j.medengphy.2013.03.002.

[2] Van Heuverswyn F, De Schepper C, De Buyzere M, Coeman M, De Pooter J, Drieghe B, et al. Clinical validation of a 13-lead electrocardiogram derived from a self-applicable 3-lead recording for diagnosis of myocardial supply ischaemia and common non-ischaemic electrocardiogram abnormalities at rest. European Heart Journal - Digital Health. 2022 Oct 21;3(4):548–58, doi: 10.1093/ehjdh/ztac062.

[3] Kleiman R, Darpo B, Brown R, Rudo T, Chamoun S, Albert DE, et al. Comparison of electrocardiograms (ECG) waveforms and centralized ECG measurements between a simple 6‐lead mobile ECG device and a standard 12‐lead ECG. Annals of Noninvasive Electrocardiology. 2021 Jul 19, doi: 10.1111/anec.12872.

[4] Rakthanmanon T, Campana B, Mueen A, Batista G, Westover B, Zhu Q, et al. Searching and mining trillions of time series subsequences under dynamic time warping. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. 2012 Aug 12, doi: 10.1145/2339530.2339576.

[5] Hundt C, Schmidt B, Schömer E. CUDA-Accelerated Alignment of Subsequences in Streamed Time Series Data [Internet]. IEEE Xplore. 2014 [cited 2023 Apr 5]. p. 10–9, doi: 10.1109/ICPP.2014.10.

[6] Stella A, Quaglio P, Torre E, Grün S. 3d-SPADE: Significance evaluation of spatio-temporal patterns of various temporal extents. Bio Systems [Internet]. 2019 Nov 1, doi: 10.1016/j.biosystems.2019.104022.

[7] Vollmer, Marcus & Bläsing, Dominic & Kaderali, Lars. (2019). Alignment of Multi-Sensored Data: Adjustment of Sampling Frequencies and Time Shifts. 10.22489/CinC.2019.031.

[8] Hamilton P, Tompkins WJ. Quantitative Investigation of QRS Detection Rules Using the MIT/BIH Arrhythmia Database. IEEE Transactions on Biomedical Engineering. 2017, doi: 10.1109/TBME.1986.325695.