Claudia Nagel*, Nicolas Pilia, Axel Loewe, and Olaf Dössel

Quantification of Interpatient 12-lead ECG Variabilities within a Healthy Cohort

https://doi.org/10.1515/cdbme-2020-3127

Abstract: The morphology of the electrocardiogram (ECG) varies among different healthy subjects due to anatomical and structural reasons, such as for example the shape of the heart geometry or the position and size of surrounding organs in the torso. Knowledge about these ECG morphology changes could be used to parameterize electrophysiological simulations of the human heart.

In this work, we detected the boundaries of ECG waveforms, i.e. the P-wave, the QRS-complex and the T-wave, in 12-lead ECGs from 918 healthy subjects in the Physionet Computing in Cardiology Challenge 2020 Database with the IBT openECG toolbox. Subsequently, we obtained the onset, the peak and the offset of each P-wave, QRS-complex and T-wave in the signal. In this way, the duration of the P-wave, the QRS-complex and the T-wave, the PQ-, RR- and the QT-interval as well as the amplitudes of the P-wave, the Q-, R- and S-peak and the T-wave in each lead were extracted from the 918 healthy ECGs. Their statistical distributions and correlation between each other were assessed.

The highest variabilities among the 918 healthy subject were found for the RR interval and the amplitudes of the QRS-complex. The highest correlation was observed for feature pairs that represent the same feature in different leads. Especially the R-peak amplitudes showed a strong correlation across different leads.

The calculated feature distributions can be used to optimize the parameters of populations of cardiac electrophysiological models. In this way, realistic in-silico generated surface ECGs can be simulated in large scale and could be used as input data for machine learning algorithms for a classification of cardiovascular diseases.

Keywords: ECG variabilities, ECG features, Electrophysiological Model Parameterization

1 Introduction

In-silico generated ECGs based on an electrophysiological model can be used to study and better understand the influence of physiological and pathological cardiac excitation propagation on the surface ECG. However, value ranges of different cell and tissue parameters that are to be specified for the simulation, such as for example tissue conductivities, vary among different literature reports [1]. Therefore, they need to be optimized in the simulations so that the in-silico generated surface ECGs match the statistics of clinical ECG characteristics in the best possible way to ensure that simulated ECG signals remain as realistic as possible.

Previous studies quantified the distribution of selected ECG features, such as amplitudes and durations of single ECG waveforms [2]. However, the ECG statistics reported in literature mainly contain analyses on selected and lead independent ECG features and no complete statistics of lead-specific amplitude and timing features based on the same database exists to the best of our knowledge. Furthermore, no comprehensive analysis on correlation coefficients among the extracted features is published.

To cater the need for a full analysis of lead dependent timing and amplitude features, eleven timing and amplitude features per lead are extracted for 918 healthy subjects from the Physionet CinC Challenge 2020 Database [3], resulting in a total of 132 features per patient. A statistical distribution for all extracted features was set up. Furthermore, the correlation coefficients between all features were calculated to evaluate how different features are linked to each other which also has to be considered when parameterizing electrophysiological simulations.

2 Methods

2.1 ECG Waveform Boundary Detection

The IBT openECG toolbox was used to detect the waveform boundaries in the 12-lead ECGs of all 918 healthy subjects. The algorithm detected the timestamps of the on- and offset of the P-wave (P_{on}, P_{off}), the QRS-complex (QRS_{on}, QRS_{off}) and the T-wave (T_{on}, T_{off}) as well as timestamps of the P-, R-, and T-peak (P_{peak}, R_{peak}, T_{peak}). The Q- and the S-peak (Q_{peak}, S_{peak}) were marked at the minimum signal value in the interval [QRS_{on}, R_{peak}] and [R_{peak}, QRS_{off}], respectively.

The delineation is exemplarily shown for 3 leads of the signal *A0002*.

***Corresponding author: Claudia Nagel, Nicolas Pilia, Axel Loewe, Olaf Dössel,** Institute of Biomedical Engineering (IBT), Karlsruhe Institute of Technology (KIT), Kaiserstraße 12, Karlsruhe, Germany, e-mail: publications@ibt.kit.edu

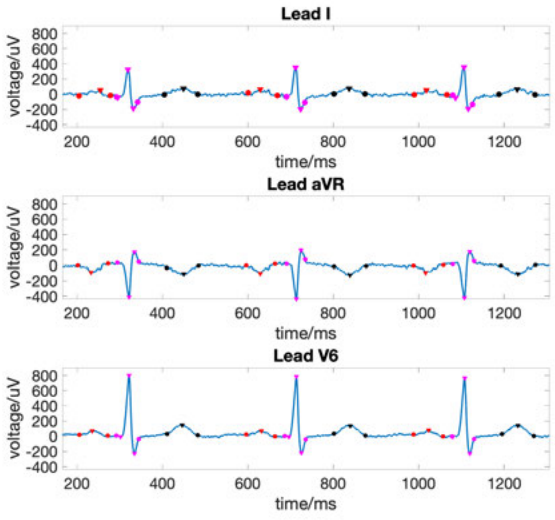

Fig. 1: Example of detected boundaries in the signal sel0303lre. P-wave, QRS-complex and T-wave markers are colored in red, magenta and black, respectively. On- and offset of the waveforms are marked with a dot, the peaks with a triangle.

2.2 Calculation of Statistical Feature Distributions

Eleven features were calculated for each beat i by means of the detected ECG waveform boundaries and peaks for each lead as follows:

- P-wave duration: $P_d[i] = P_{off}[i] - P_{on}[i]$
- QRS-complex duration: $QRS_d[i] = QRS_{off}[i] - QRS_{on}[i]$
- T-wave duration: $T_d[i] = T_{off}[i] - T_{on}[i]$
- PQ-intverval: $PQ_d[i] = QRS_{on}[i] - P_{on}[i]$
- QT-intverval: $QT_d[i] = T_{off}[i] - QRS_{on}[i]$
- RR-intverval: $RR_d[i] = R_{peak}[i+1] - R_{peak}[i]$
- P-amplitude: $P_a[i] = ECG(P_{peak}[i])$
- Q-amplitude: $Q_a[i] = ECG(Q_{peak}[i])$
- R-amplitude: $R_a[i] = ECG(R_{peak}[i])$
- S-amplitude: $S_a[i] = ECG(S_{peak}[i])$
- T-amplitude: $T_a[i] = ECG(T_{peak}[i])$

For each subject, the median value for each feature was calculated over all beats detected in one recording to obtain a representative feature set per subject as exemplarily shown for the P-wave duration below:

$$P_d = median(P_d[0], ..., P_d[N])$$

with N denoting the number of beats. Taking the median accounts for wrongly detected waveform boundaries or outliers and ensures that the length of the ECG recording does not influence the calculation of the statistical distributions of interpatient ECG variations.

The histograms of the resulting 132 ECG features (11 features $\times$ 12 leads) were calculated and the probability distribution for each feature was fitted once with a kernel and once with a normal distribution. Subsequently, the probability density function (PDF) for each feature was generated.

Furthermore, the correlation coefficients between each pair of features were computed.

3 Results

3.1 Feature Distributions

The mean and standard deviation of all feature values are listed in Tab. 1.1 and Tab. 1.2.

The timing features in Tab. 1 show that the duration of different ECG segments slightly varies between different leads. The P-wave is characterized by a relatively low dispersion of 16 ms [4], whereas the QRS-complex and the T-wave duration vary more markedly with a maximum discrepancy between their lead dependent mean values in Tab. 1 of 24ms and 40ms, respectively. The RR interval is independent of the specific lead and shows the same mean and standard deviation across all leads up to a a precision of 10 μs (cf. Tab. 1).

The distributions of the amplitudes differ among the 12 leads. Especially the R- and S-peak amplitude show differences of >1mV in their mean values between different leads. The amplitudes for lead aVR are negative (except for the R-Peak by definition). Lead II, which represents the main axis of the heart, showed the highest absolute mean values for all amplitude features among the Einthoven leads (cf. Tab. 2). The R-progression /S-degression in the Wilson leads is shown in Tab. 2. The R-Peak amplitude increases from lead V1 to lead V4, whereas the absolute value of the S-Peak amplitude decreases from lead V2 to V6. The PDFs calculated with a kernel estimation for the R- and S-peak amplitudes in the Wilson leads are shown in Fig. 2. They all follow approximately a normal distribution with increasing and decreasing mean values between V1 to V4 and V2 to V6, respectively.

3.2 Feature Correlations

All feature combination pairs with a correlation coefficient above 0.85 are listed in Tab. 3. The highest correlation coefficients are in all twelve cases obtained for pairs of the same feature in different leads. Especially for the R-peak amplitude

Tab. 1: Mean and standard deviation of all timing features in different leads.

ECG Feature	P_d in ms	PQ_d in ms	QRS_d in ms	QT_d in ms	T_d in ms	RR_d in ms
Lead I	135.53 (± 17.82)	152.22 (± 24.81)	100.41 (± 13.66)	379.62 (± 39.92)	155.36 (± 30.08)	812.14 (± 106.39)
Lead II	134.42 (± 14.31)	156.05 (± 23.52)	105.67 (± 15.83)	385.45 (± 40.83)	157.47 (± 31.88)	812.12 (± 106.39)
Lead III	142.89 (± 22.78)	178.43 (± 29.46)	99.12 (± 22.47)	409.92 (± 59.78)	191.21 (± 56.59)	812.21 (± 106.75)
Lead aVR	133.43 (± 12.67)	151.77 (± 21.56)	103.97 (± 13.27)	379.03 (± 32.67)	151.61 (± 21.81)	811.93 (± 106.47)
Lead aVL	148.71 (± 23.09)	177.03 (± 32.14)	96.08 (± 19.68)	387.18 (± 58.46)	175.81 (± 50.66)	812.15 (± 106.49)
Lead aVF	136.93 (± 17.73)	166.34 (± 27.19)	102.76 (± 19.85)	397.01 (± 53.34)	171.95 (± 47.39)	812.28 (± 106.47)
Lead V1	145.78 (± 28.23)	184.87 (± 25.60)	120.39 (± 17.99)	388.35 (± 56.07)	179.77 (± 50.05)	812.05 (± 106.12)
Lead V2	149.09 (± 26.69)	176.39 (± 28.93)	115.88 (± 13.69)	383.86 (± 55.20)	182.25 (± 46.44)	812.13 (± 106.38)
Lead V3	149.26 (± 25.72)	165.04 (± 28.60)	109.40 (± 14.47)	390.06 (± 51.93)	175.75 (± 44.94)	812.21 (± 106.50)
Lead V4	145.63 (± 23.41)	161.11 (± 29.96)	103.21 (± 13.14)	388.56 (± 44.96)	163.52 (± 37.59)	812.17 (± 106.48)
Lead V5	143.60 (± 21.15)	161.32 (± 27.55)	100.56 (± 12.19)	384.76 (± 41.78)	158.17 (± 32.89)	812.10 (± 106.41)
Lead V6	142.23 (± 19.20)	160.91 (± 26.16)	101.95 (± 12.81)	383.41 (± 41.52)	156.47 (± 31.85)	812.17 (± 106.38)

Tab. 2: Mean and standard deviation of all amplitude features in different leads.

ECG Feature	P_a in μV	Q_a in μV	R_a in μV	S_a in μV	T_a in μV
Lead I	57.84 (± 40.96)	-46.24 (± 41.82)	561.98 (± 278.40)	-98.81 (± 94.05)	186.61 (± 93.16)
Lead II	79.59 (± 63.89)	-64.58 (± 43.31)	808.22 (± 336.15)	-129.72 (± 124.81)	209.99 (± 103.74)
Lead III	34.67 (± 67.42)	-86.47 (± 135.08)	426.10 (± 328.25)	-141.03 (± 218.73)	29.44 (± 94.77)
Lead aVR	-64.43 (± 43.91)	-387.17 (± 341.10)	109.86 (± 78.53)	-246.27 (± 400.48)	-200.49 (± 83.24)
Lead aVL	19.13 (± 49.43)	-43.09 (± 75.48)	284.50 (± 251.76)	-136.85 (± 164.22)	81.03 (± 79.17)
Lead aVF	56.37 (± 62.30)	-53.80 (± 55.52)	575.84 (± 342.59)	-120.35 (± 137.87)	117.41 (± 87.88)
Lead V1	13.67 (± 99.82)	-72.81 (± 261.53)	229.23 (± 178.70)	-757.58 (± 439.68)	-7.92 (± 152.03)
Lead V2	112.43 (± 264.43)	-38.71 (± 305.62)	640.32 (± 372.50)	-1083.27 (± 564.17)	365.82 (± 285.77)
Lead V3	114.67 (± 282.59)	-15.65 (± 176.30)	998.17 (± 529.93)	-752.48 (± 448.13)	390.28 (± 265.82)
Lead V4	34.55 (± 169.56)	-45.05 (± 63.71)	1389.98 (± 576.85)	-476.21 (± 328.45)	386.56 (± 225.91)
Lead V5	13.34 (± 65.93)	-59.21 (± 53.68)	1324.48 (± 467.76)	-283.99 (± 234.78)	329.26 (± 176.52)
Lead V6	19.43 (± 63.52)	-61.90 (± 49.78)	1089.41 (± 365.06)	-151.85 (± 143.66)	264.89 (± 136.95)

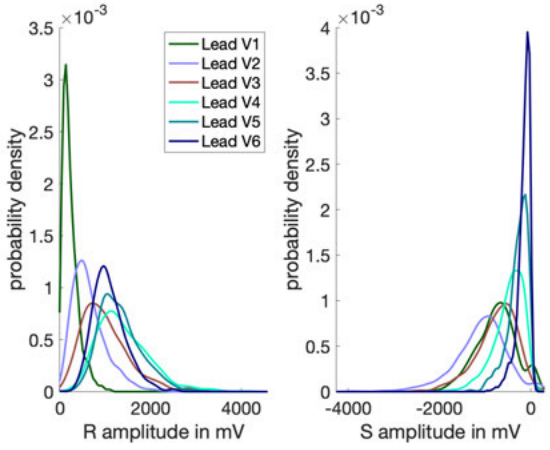

Fig. 2: Probability density functions of the R-peak and S-peak amplitudes in the Wilson leads V1-V6 calculated with a kernel estimation. The distributions for different leads are color-coded.

Tab. 3: Feature pairs with a correlation coefficient above 0.85.

ECG Feature 1	ECG Feature 2	correlation coefficient
R_a Lead III	R_a Lead aVF	is 0.9302
R_a Lead II	R_a Lead aVF	is 0.9215
R_a Lead V4	R_a Lead V6	is 0.8808
T_a Lead II	T_a Lead aVF	is 0.8752
S_a Lead V3	S_a Lead V4	is 0.8645
S_a Lead V4	S_a Lead V5	is 0.8620
R_a Lead I	R_a Lead aVL	is 0.8620

and the S-peak amplitude, a strong correlation prevails for the same feature in different leads.

The highest absolute correlation coefficient for a feature pair representing different features in the same lead is the Q-Peak and the S-Peak amplitude in lead aVR with a correlation coefficient of -0.8304. The highest correlation between different features in different leads was found for the R-Peak amplitude in lead aVR and the S-peak amplitude in lead II (-0.8161).

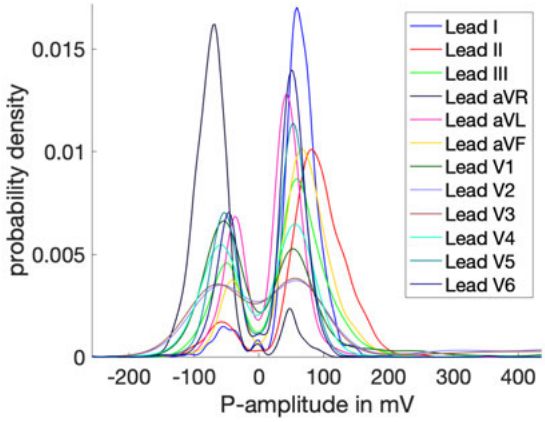

Fig. 3: Probability density functions of the P-wave amplitudes calculated with a kernel estimation. The distributions for different leads are color-coded.

4 Discussion

In this work, interpatient ECG variabilities were quantified in all twelve standard leads and statistics including the mean and the standard deviation of all feature values across different leads are provided. In contrast to previous studies, the ECG feature values were assessed specifically for each lead. By comparing the findings available in previous works with the respective mean values in this study, it gets apparent that especially the values for the QRS duration, the RR- and QT interval coincide in our and previous studies. However, the mean values we found for the P-wave duration are approximately 15-20 ms higher compared to those published in literature [2]. The reason for this can be traced back to different definitions based on which the P-wave boundaries get detected.

Also the calculation of the PDFs highly depend on correctly detected ECG boundaries. Even though using the median value among approximately 30 beats per patient for each feature minimizes the effect of incorrect detected fiducial points in the ECG, it cannot be guaranteed that all ECG boundaries are accurately identified throughout the whole signal.

Even though the assumption of normally distributed features holds for most of the features, it is violated especially for the P-wave amplitude in lead V1 where the P-wave is known to be biphasic. For this reason, the PDF of the ECG features are not only calculated with a normal but also with a kernel distribution that is shown in Fig. 3. It can be seen that the curve for lead V1 (dark green) is characterized by two peaks with one at a positive and one at a negative value. Therefore, it would be beneficial to first classify different morphologies and subsequently set up individual statistics for each of them.

The high correlation coefficients in Tab. 3 involving the features from a Goldberger and from an Einthoven lead traces back to the fact that aVR, aVL and aVF can be calculated with a linear combination of the signals from lead I, II and III and are therefore dependent on each other. The negative correlation coefficients involving amplitude features from the aVR lead is due to the predominantly negative signal in this lead [5].

5 Conclusion

With this work, we provide the mean and the standard deviation of 132 ECG features. Since most of the features are normally distributed along the cohort, the simulation parameters need to be set in such a way that the simulated population also reflects normal feature distributions. The highest correlation coefficients were found for the same ECG features in different leads meaning that when parameterizing electrophysiological simulations, the correlation of different features in different leads do not necessarily have to be considered.

With the knowledge of the feature values' variation within the healthy population, the findings of this study can not only be used for parameterizing electrophysiological simulations but also for diagnosing cardiovascular diseases by comparing the feature values of the healthy reference population with potentially diseased ones.

Author Statement

Research funding: This study is part of the project 18HLT07 Medal Care. Conflict of interest: Authors state no conflict of interest.

References

[1] M. J. Peters, G. Stinstra, M. Hendriks (2001) Estimation of the Electrical Conductivity of Human Tissue, Electromagnetics, 21:7-8, 545-557

[2] Mason, J., Ramseth, D., Chanter, D., Moon, T., Goodman, D., Mendzelevski, B. (2007). Electrocardiographic reference ranges derived from 79,743 ambulatory subjects. Journal Of Electrocardiology, 40(3), 228-234.e8.

[3] Perez Alday, E. A., Shah, A., Liu, C., Sharma, A., Seyedi, S., Bahrami Rad, A., Reyna, M., Clifford, G. (2020). Classification of 12-lead ECGs: the PhysioNet - Computing in Cardiology Challenge 2020 (version 1.0.0). PhysioNet.

[4] Aytemir, K., Ozer, N., Atalar, E., Sade, E., Aksoyek, S., Ovunc, K. et al. (2000). P Wave Dispersion on 12-Lead Electrocardiography in Patients with Paroxysmal Atrial Fibrillation. Pacing And Clinical Electrophysiology, 23(7), 1109-1112.

[5] Madias, J. (2011). aVR vs. -aVR: should aVR be left where it is?. Europace, 13(7), 918-919.