

Assessment of the stability of morphological ECG features and their potential for person verification/identification

Mikhail Matveev^{1*}, Ivaylo Christov¹, Vessela Krasteva¹, Giovanni Bortolan², Dimitar Simov¹, Nikolay Mudrov¹, and Irena Jekova¹

¹Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Sofia, Bulgaria

²Institute of Neuroscience, IN-CNR, Padova, Italy

Abstract. This study investigates the potential of a set of ECG morphological features for person verification/identification. The measurements are done over 145 pairs of ECG recordings from healthy subjects, acquired 5 years apart (T1, T2 = T1+5 years). Time, amplitude, area and slope descriptors of the QRS-T pattern are analysed in 4 ECG leads, forming quasi-orthogonal lead system (II&III, V1, V5). The correspondence between feature values in T1 and T2 is verified via factor analysis by principal components extraction method; correlation analysis applied over the measurements in T1 and T2; synthesis of regression equations for prediction of features' values in T2 based on T1 measurements; and cluster analysis for assessment of the correspondence between measured and predicted feature values. Thus, 11 amplitude descriptors of the QRS complex are highlighted as stable, i.e. keeping their strong correlation (≥ 0.7) within a certain factor, weak correlation (< 0.3) with the features from the remaining factors and presenting high correlation in the two measurement periods that is a sign for their person verification/identification potential. The observed coincidence between feature values measured in T2 and predicted via the designed regression models ($r=0.93$) suggests about the confidence of person identification via the proposed morphological features.

1 Introduction

Nowadays, research on automatic person identification is a rapidly developing area, considering not only the security requirements for the cases of financial transactions, access control, travelling, but also the remote health monitoring scenarios in the clinical and emergency medicine, as well as the organization and processing of hospital databases. The efforts of many researchers are focused on the application of internal body physiological biometric characteristics, which provide robustness to hacker attacks and falsification and are available in most of the health monitoring scenarios. The analysis of the electrocardiogram (ECG) in this respect started about a decade ago, applying either methods that use measurements after detection of fiducial points or analysis of the ECG morphology.

The fiducial-based approaches are the primary considered in the field as they employ morphological features, typically measured for diagnostic purposes by ECG devices. Thus, the crucial task for precise localization of specific anchor points on the P-QRS-T segment can be managed by certified commercial ECG analysis modules with minimal intervention. The application of such a person identity approach can be easily extended to automated management of in-hospital databases. According to the data cited in literature, the identification accuracy could reach values between 97%

and 100%, based on the following analysis approaches: 12 uncorrelated diagnostic features of P-QRS-T amplitudes and durations, processed by Principle Component analysis score plots applied over a database with 20 subjects [1]; 15 P-QRS-T temporal features, fed to discriminant functions over 29 subjects under various stress conditions [2]; fiducial based temporal and amplitude measurements combined with features that capture the heartbeat patterns, tested over 31 healthy subjects: 18 with a single ECG record and 13 with more than one ECG record [3].

The fiducial independent approaches for person identification are based on assessment of the P, QRS, T waveforms similarity in the analysed ECG recordings by calculation of correlation coefficients, which are subsequently processed with different techniques. The achieved identification accuracy is: 100% via discrete cosine transform over a database with 14 subjects [4]; 96.2% provided by discriminant analysis over 61 subjects [5]; 91.4% [6] and 85.7% [7] via assessment of the maximal correlation coefficient of a single-lead ECG for 11 and 14 subjects, respectively, improved to 100% while 12-lead ECG is analysed for the database with 11 subjects [6].

Majority of the cited methods are tested with small-sized ECG databases [1,2,4,7] or track intra-subject changes of ECG characteristics measured in very short distanced time intervals or even in the same session

* Corresponding author: mgm@biomed.bas.bg

[2,3,5]. This might bias the reported high identification accuracy from the real case scenario, where the influence of different factors, such as long-term ECG changes, electrode rearrangement, etc., should be considered. Our experience with large ECG databases, containing distanced in time couples of standard 12-lead ECG recordings shows lower authentication accuracies, i.e. about 86% correct verifications for a fiducial-based morphological measurements over 574 ECG couples [8]; 92.5% correct verifications for non-fiducial QRS pattern matching over 316 ECG couples [9]; 87%/78% correct verifications/identifications for non-fiducial correlation analysis of limb leads over 147 ECG couples [10].

Recently, we evaluated the inter-subject variability of morphological ECG features over large amount of time (>5 years) in a population of healthy cardiac subjects [11]. The aim of this study is to investigate the intra-individual stability of the ECG morphology using the same feature set in order to assess the potential for person identification via orthogonal ECG leads.

2 ECG database

The ECG signals used in this study are taken from a computerized ECG-ILSA database [12,13], collected for the Italian Longitudinal Study on Aging project [14]. The ECGs have been collected from individuals aged from 65 to 84 years. Each recording is with duration of 10s and includes the standard 12 leads, sampled at 500 Hz. The database contains 901 patients recorded both in the first phase (T1) and in the second phase (T2=T1+5 years). A set of 145 individuals without cardiac diseases has been selected for analysis in this study.

3 Method

Four ECG leads, forming quasi-orthogonal lead system (II&III,V1,V5) [15] are processed as follows:

- 1) Detection of QRS complexes in a combined ECG lead [16], heartbeat classification [17], calculation of mean RR interval between predominant beats;
- 2) Extraction (by best-fit correlation) of an averaged QRS-T pattern of the predominant beats for each of the 4 leads. Interval of $0.35 \cdot RR$ before and $0.55 \cdot RR$ after the detected fiducial points is considered;
- 3) QRS-T delineation in each of the four leads [18];
- 4) Calculation of 15 morphological pattern features per lead (see figure 1), as follows:

- 12 features describing the QRS morphology that have previously been approved to provide adequate heartbeat comparison and classification [14]: QRS-width (Width); amplitudes (Ma, Mi) and offsets (Ima, Imi) of maximal positive and negative peaks; slope from QRS-onset to first peak (S1); slope from first to second peak (S2); QRS positive (ArP), negative (ArN) and total area (Ar=ArP+ArN); sum of the absolute QRS velocities values (Av); number of samples crossing 70% of the maximal peak amplitude (No);

- 3 features describing the T-wave – width (TWidth), amplitude (Tamp), peak-to-end interval (Tp_Tend).

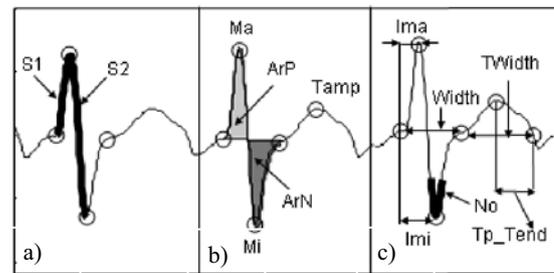


Fig. 1. Graphical representation of the morphological features calculated for the averaged delineated QRS-T pattern: a) slopes; b) amplitudes, areas; c) time intervals.

Aiming to select feature constellation, which is stable over time, thus being appropriate for the identification task, we apply the following procedures included in the multivariate analyses of the software package Statistica 12.7 (Dell Inc.):

- 1) Factor Analysis by Principal Components extraction method with 'varimax' rotation for providing orthogonal factors. The idea is to highlight the outstanding correlated (dependent) and uncorrelated (independent) features separated in the first 3 factors. The stability of the factors is independently verified over the data collected in the periods: T1, T2 and joint T1+T2. Presence of factor stability between the 3 datasets would indicate the factors' potential to represent the subjects' correspondence.
- 2) Correlation analysis for assessment of the coincidence between feature values in T1 and T2;
- 3) Multiple stepwise regression for synthesis of regression equations that could be applied for predicting the value of a certain feature in T2, based on data measured in T1.
- 4) Cluster analysis for assessment of the prediction model, e.g. the coincidence between measured and predicted feature values.

4 Results and discussion

The analysis of the designed feature space ($15 \times 4 = 60$ features) via Factor analysis highlights 11 features that are relevant to the aim of the study and for which the following observations are valid:

- High (≥ 0.7) and significant ($p < 0.05$) factor loadings in one of the first 3 factors, combined with low and insignificant loadings in the remaining factors, when the factor analysis is performed over the joint data for 2×145 patients in T1+T2 records (see Table 1);
- The same feature distribution among the first 3 factors and comparable factor loadings, when the factor analysis is performed separately over the data collected during T1 (145 patients) and T2 (145 patients). The results are presented in Table 2 and Table 3.

Table 1. Factor loadings in the first 3 factors obtained via analysis of the data for T1 and T2.

Loadings with absolute values ≥ 0.7 are highlighted.

	Factor 1	Factor 2	Factor 3
ArN_II	-0.889176	0.004531	0.093518
S2_III	-0.763494	0.048325	-0.165169
Ar_III	0.928569	0.076507	0.083850
ArN_III	-0.874855	-0.078762	0.141749
Av_III	0.870994	0.011419	0.132374
Ar_VI	0.097656	0.834255	0.101620
ArN_VI	-0.042548	-0.822745	-0.167931
Av_VI	-0.009787	0.880939	0.125139
Ar_V5	0.044894	0.145240	0.727179
ArP_V5	-0.198526	0.232198	0.723374
Av_V5	-0.027200	-0.086195	0.744158

Table 2. Factor loadings in the first 3 factors obtained via analysis of the data for T1.

Loadings with absolute values ≥ 0.7 are highlighted.

	Factor 1	Factor 2	Factor 3
ArN_II	-0.900022	0.106638	0.069295
S2_III	-0.716803	-0.059771	0.006593
Ar_III	0.929317	0.043028	0.025345
ArN_III	-0.884789	0.071698	0.121484
Av_III	0.849656	0.059122	0.019912
S2_VI	0.155151	-0.893559	0.064475
Ar_VI	0.016207	0.851458	0.045255
ArN_VI	0.015608	-0.821156	-0.054854
Av_VI	-0.088712	0.900437	-0.005498
Ar_V5	0.044044	0.027269	0.877656
ArP_V5	-0.142407	0.134194	0.834679

Table 3. Factor loadings in the first 3 factors obtained via analysis of the data for T2.

Loadings with absolute values ≥ 0.7 are highlighted.

	Factor	Factor	Factor
ArN_II	-0.883309	-0.040706	0.140409
S2_III	-0.748261	0.013423	-0.169060
Ar_III	0.921478	0.108880	0.075771
ArN_III	-0.859490	-0.121004	0.125263
Av_III	0.867208	0.038183	0.118138
S2_VI	0.053774	-0.745554	-0.042857
Ar_VI	0.185672	0.758675	0.098808
ArN_VI	-0.089083	-0.790738	-0.226551
Av_VI	0.061571	0.865982	0.114185
Ar_V5	0.089063	0.206883	0.743416
ArP_V5	-0.181731	0.285029	0.764283

Table 4. Correlations coefficients between the features included in the stable groups for both measurement periods T1 and T2.

Significant correlations are highlighted ($p < 0.05$).

T1\T2	ArN_II (T2)	S2_III (T2)	Ar_III (T2)	ArN_III (T2)	Av_III (T2)	Ar_VI (T2)	ArN_VI (T2)	Av_VI (T2)	Ar_V5 (T2)	ArP_V5 (T2)	Av_V5 (T2)
ArN_II (T1)	0.91	0.50	-0.78	0.82	-0.63	-0.11	0.08	0.01	-0.13	0.17	-0.04
S2_III (T1)	0.42	0.86	-0.74	0.66	-0.82	-0.15	0.13	-0.11	-0.21	-0.10	-0.17
Ar_III (T1)	-0.72	-0.74	0.90	-0.87	0.82	0.25	-0.20	0.13	0.22	0.03	0.12
ArN_III (T1)	0.74	0.66	-0.86	0.93	-0.77	-0.24	0.14	-0.10	-0.12	0.08	-0.05
Av_III (T1)	-0.58	-0.83	0.84	-0.79	0.86	0.22	-0.17	0.13	0.23	0.07	0.16
Ar_VI (T1)	0.12	-0.02	0.04	0.01	0.04	0.70	-0.55	0.65	0.12	0.21	0.08
ArN_VI (T1)	-0.16	0.01	-0.02	-0.03	-0.02	-0.70	0.57	-0.64	-0.12	-0.23	-0.08
Av_VI (T1)	0.18	0.01	-0.04	0.10	-0.01	0.59	-0.46	0.65	0.04	0.16	0.01
Ar_V5 (T1)	0.10	-0.01	0.04	-0.01	0.03	0.21	-0.27	0.18	0.73	0.74	0.67
ArP_V5 (T1)	0.25	0.07	-0.08	0.12	-0.06	0.23	-0.24	0.20	0.65	0.76	0.61
Av_V5 (T1)	0.15	0.01	-0.03	0.04	-0.02	0.12	-0.22	0.13	0.65	0.66	0.66

The Factor analysis results support the following conclusions:

- 1) Confirm our hypothesis that the ECG leads (II&III), V1 and V5 form quasi-orthogonal lead system, considering the fact that the features, which are strongly correlated with a certain factor are from a single axis (lead) and the features measured in different axes (leads) are distributed among the 3 factors;
- 2) Confirm the stability of the formed feature subsets, considering the measurements in T1 and T2, i.e. their strong correlation within a certain factor and weak correlation with the features from the remaining factors. The formed feature subsets are as follows:

- ArN_II; S2_III; Ar_III; ArN_III; Av_III in lead (II&III);
- Ar_VI; ArN_VI; Av_VI in lead V1;
- Ar_V5; ArP_V5; Av_V5 in lead V5.

Based on the above findings, it is reasonable to consider that the best features within the formed subsets are representative for person identification.

The results from the correlation analysis applied over the feature values measured in T1 and T2 (see Table 4) give a hint about the possibility for synthesis of regression equations that could be applied for predicting the value of a certain feature in T2 based on data measured in T1. The comparison between the predicted features' values and the measured ones could be considered as an adequate preliminary sign about the confidence level for person identification via morphological ECG features measured in different time moments. This is demonstrated via one of the models designed by means of stepwise multiple regression analysis, i.e. prediction of the values for ArN_III(T2) based on the values measured for ArN_II(T1) and ArN_III(T1) – see Table 5. Based even on this single model for calculation of ArN_III values in T2 we have achieved 80.7% group indistinction, when applying K-means clustering over the measured and predicted values. Only 19.3% of the predicted values have been assigned to different cluster.

Table 5. Statistics for the regression model
 $ArN_III(T2) = f(ArN_II(T1), ArN_III(T1))$.
 $R = .929$ $R^2 = .862$ Adjusted $R^2 = .860$
 $F(2,143) = 443.95$ $p < 0.001$, Std. Error of estimate: 2.158

	Beta	Std.Err. of Beta	B	Std.Err. of B	t(143)	p-level
Intercept			-0.562	0.264	-2.129	0.035
ArN_II(T1)	0.142	0.057	0.331	0.133	2.483	0.014
ArN_III(T1)	0.806	0.057	0.922	0.066	14.063	0.000

5 Conclusions

The results achieved in this preliminary study show that in general the morphological features measured from the electrocardiogram could be applied for identification of subjects in different periods of examination. This could be used in the clinical and emergency medicine, as well as for the design and management of personalized ECG databases.

Acknowledgements

This study is supported by the Bulgarian Scientific Research Fund, grant T02/11.

References

1. L. Biel, O. Pettersson, L. Philipson, P. Wide, "ECG analysis: a new approach in human identification", *IEEE Trans. Instr. Meas.*, **50(3)**: 808-12, (2001)
2. S. Israel, J. Irvine, A. Cheng, M. Wiederhold, B. Wiederhold, "ECG to identify individuals" *Pattern Recognition*, **38**, 133-42 (2005)
3. Y. Wang, F. Agrafioti, D. Hatzinakos, K. Plataniotis, "Analysis of human electrocardiogram for biometric recognition", *EURASIP Journal on Advances in Signal Processing*, Article ID 148658 (2008)
4. K. Plataniotis, D. Hatzinakos, J. Lee, "ECG biometric recognition without fiducial detection", *Proc. on Biometrics Symposium BSYM'06*, 19-21, (2006)
5. F. Agrafioti, D. Hatzinakos, "ECG biometric analysis in cardiac irregularity conditions", *Signal Image & Video Processing*, **3**, 329-43 (2009)
6. F. Poree, A. Gallix, G. Carrault, "Biometric identification of individuals based on the ECG. Which conditions?", *Comp. in Cardiol.*, **38**, 761-4, (2011)
7. I. Jekova. "Potential of the peripheral electrocardiogram for personal verification/identification" *Compt. Rend. de l'Acad. Bulg. des Sci.*, **68(4)**: 497-504, (2015)
8. I. Jekova, V. Krasteva, R. Leber, R. Schmid, R. Twerenbold, C. Müller, T. Reichlin, R. Abächerli, "Intersubject variability and intrasubject reproducibility of 12-lead ECG metrics: Implications for human verification", *J. Electrocardiol.*, **49**: 784-9, (2016)
9. V. Krasteva, I. Jekova, R. Abächerli, "Analysis of QRS Patterns in 15-Lead ECG for Person Verification", *Annual J. Electronics*, **9**: 10-13, (2015)
10. I. Jekova, G. Bortolan, "Personal verification/identification via analysis of the peripheral ECG leads. Influence of the personal health status on the accuracy", *Biomed. Research Int.*, **2015**, Article ID 135676, (2015)
11. I. Jekova, I. Christov, V. Krasteva, G. Bortolan, M. Matveev, "Assessment of the Potential of Morphological ECG Features for Person Identification", *Comput. in Cardiol.*, **42**: 921-4, (2015).
12. S. Maggi, M. Zuchetto, F. Grigoletto, M. Baldereschi, L. Candelise, E. Scarpini, et al., "The Italian Longitudinal Study on Aging (ILSA): design and methods", *Aging Clin. Exp. Res.*, **6**: 464-73 (1994)
13. G. Bortolan, M. Bressan, F. Golferini, ILSA Study Group, "QT dispersion in the elderly. The ILSA study", *Aging Clin. Exp. Res.*, **16(5)**: 342-8, (2004)
14. G. Bortolan, M. Bressan, F. Golferini, "Serial analysis in the ECG-ILSA database: study of longitudinal modifications", *Comput. In Cardiol.*, **26**: 667-70, (1999)
15. M. Matveev, L. Tomov, I. Tomov, I. Daskalov, "Lead Selection for ECG Screening", *Adv. Cardiol.*, **19**: 148-9, (1977)
16. I. Christov, "Real time electrocardiogram QRS detection using combined adaptive threshold", *Biomed. Eng. Online*, **3(28)**, (2004)
17. I. Christov, I. Jekova, G. Bortolan, "Premature ventricular contraction classification by the K-th nearest neighbours rule", *Physiol Meas.*, **26**: 123-30, (2005)
18. I. Christov, I. Simova, "Q-onset and T-end delineation: Assessment of the performance of an automated method with the use of a reference database", *Physiol Meas.*, **28**: 213-21, (2007)