

Assessment of the Potential of Morphological ECG Features for Person Identification

I Jekova¹, I Christov¹, V Krasteva¹, G Bortolan², M Matveev¹

¹Institute of Biophysics and Biomedical Engineering, Sofia, Bulgaria

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

Abstract

This study investigates the potential of ECG morphological feature set for person identification. The measurements are done over 145 pairs of ECG recordings from healthy subjects, acquired 5 years apart. Time, amplitude, area and slope descriptors of the QRS-T pattern are analyzed in 4 ECG leads, forming quasi-orthogonal lead system (II&III, VI, V5). The inter-subject variation, the difference of means in 1st vs. 2nd recording measurements, as well as the cross-correlation between features are estimated. Thus, 2 area and 4 amplitude descriptors of the QRS complex are highlighted. The population heterogeneity in the space of the selected features is verified via Factor analysis by Principal components extraction method. It confirms the orthogonality of the 6 features (each of them has significant factor loading for a particular factor). The analysis shows that the first 3 factors have eigenvalues higher than 1, both for the measurements in the 1st and the 2nd ECG recording and they accumulate respectively 68% and 64 % of the total data variation, which is a sign for their person identification potential.

1. Introduction

Nowadays, research on automatic person identification is focused on internal body physiological biometric characteristics, robust to hacker attacks and falsification. 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 overall ECG morphology.

The methods relying on fiducial based approach report: identification based on 12 uncorrelated diagnostic features of P-QRS-T amplitudes and durations, processed by Principle Component analysis score plots, achieved 100% identification accuracy (IDA) over a database with 20 subjects [1]; identification employing 15 P-QRS-T temporal features, fed to discriminant functions, providing IDA between 97% and 100% over 29 subjects

under various stress conditions [2]; identification via fiducial based temporal and amplitude measurements combined with features that capture the heartbeat patterns [3]. The latter combined approach provides 100% IDA when tested over 31 healthy subjects: 18 with a single ECG record and 13 with more than one ECG record.

Great part of the fiducial independent approaches for identification is based on autocorrelation (AC). The subsequent processing with discrete cosine transform results in 100% IDA over a database with 14 subjects [4]; application of discriminant analysis [5] provides 96.2% IDA for 48 patients with single ECG recording and 13 healthy subjects with more than one ECG record; assessment of the maximal correlation coefficient of a single-lead assures 91.4% [6] and 85.7% [7] over databases with 11 subjects and 14 subjects, respectively, while 12-lead ECG assures 100% IDA over the database with 11 subjects [6]. Another method calculates the two-dimensional heart vector formed by the limb ECG leads and its first and second derivatives achieves 98.1% IDA by a distance based approach over 74 subjects [8]. Processing of a normalized QRS complex via Multilyer perceptron provides 96.1% IDA over a database with 30 healthy subjects [9]. ECG decomposition in a number of intrinsic mode functions combined with Welch spectral analysis for heartbeat features extraction, followed by a K-Nearest Neighbors classifier leads to 95.6% IDA over 108 subjects having one ECG record with ST-segment changes and 12 healthy subjects with more than one ECG record [10].

Majority of the cited methods are tested with small-sized ECG databases [1,2,4,7,9] or track intra-subject changes of ECG characteristics measured in very short distanced time intervals [2,3,5,9,10]. This might bias the reported high identification/verification accuracy from the real case scenario.

The aim of this study is to investigate ECG inter-individual differences observed in a set of morphological features in order to assess their potential for person identification. The study is conducted on a large ECG dataset, containing two different ECG recordings per subject acquired 5 years apart.

2. ECG database

The ECG signals used in this study are taken from a computerized ECG-ILSA database, collected for the Italian Longitudinal Study on Aging project [11]. 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) [12] are processed as follows:

- 1) Detection of QRS complexes in a combined ECG lead [13], heartbeat classification [14], 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 $0.35*RR$ before and $0.55*RR$ after the detected fiducial points is considered;
- 3) QRS-T delineation in each of the four leads [15];
- 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 [13]: 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*);
- 5) Calculation of QT/RR ratio, common for the four leads. It is reported to exhibit high intra-subject stability and essential inter-subject variability [16].

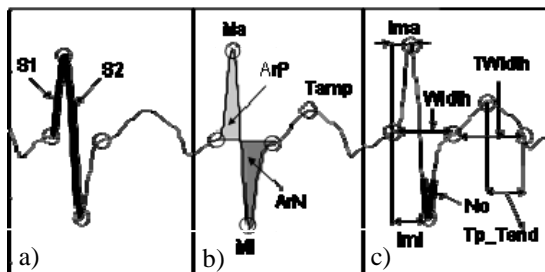


Figure 1. Graphical representation of the morphological features calculated for the averaged delineated QRS-T pattern: a) slopes; b) amplitudes, areas; c) time intervals.

Aiming to select feature constellation appropriate for the identification task, we do the following:

- 1) Calculate coefficient of variation for each feature (x):

$$CV(x) = 100 * (Standard\ Deviation(x) / Mean(x)),$$

to score its potential for providing large distance between samples in the feature space. This is important for the person identification task where enhanced inter-individual differences are beneficial.

- 2) Apply one-sided paired t-test over the scored features to verify that their mean values calculated for the measurements in T1/T2 do not differ significantly.
- 3) Set additional requirement for lack of correlation between features, thus minimizing the redundancy of information in the feature constellation.
- 4) Apply Factor analysis by Principal components extraction method on the outstanding uncorrelated features to verify the population heterogeneity.

4. Results and discussion

Analyzing the designed feature space ($15 \times 4 + 1 = 61$ features), we have highlighted two of the features that are relevant to heterogeneous population ($CV > 30\%$) in both T1 and T2: *Ar_II* ($CV(T1)=44\%$, $CV(T2)=50\%$) and *Ar_III* ($CV(T1)=76\%$, $CV(T2)=82\%$). The feature constellation is complemented by four additional features which satisfy the conditions for insignificant differences between mean values in T1 and T2, i.e. *Mi_II* (126.2 ± 14.9 vs. 128.2 ± 22.1 , $p=0.183$), *Mi_III* (124.2 ± 9.2 vs. 123.8 ± 9.5 , $p=0.358$), *Mi_V1* (124.8 ± 3.7 vs. 125.7 ± 7.4 , $p=0.096$), *Mi_V5* (130.6 ± 13.7 vs. 132.2 ± 13.3 , $p=0.157$), and show the minimal correlations to each other (Table 1a,b).

Table 1a. Correlations within the selected feature constellation observed for the measurements in T1.

Var	Mi_II	Ar_II	Mi_III	Ar_III	Mi_V1	Mi_V5
Mi_II	1.00	-0.07	0.30	0.06	0.13	0.26
Ar_II	-0.07	1.00	-0.02	0.39	-0.13	-0.21
Mi_III	0.30	-0.02	1.00	0.05	-0.04	0.08
Ar_III	0.06	0.39	0.05	1.00	-0.23	0.18
Mi_V1	0.13	-0.13	-0.04	-0.23	1.00	-0.11
Mi_V5	0.26	-0.21	0.08	0.18	-0.11	1.00

Table 1b. Correlations within the selected feature constellation observed for the measurements in T2.

Var	Mi_II	Ar_II	Mi_III	Ar_III	Mi_V1	Mi_V5
Mi_II	1.00	0.14	0.04	0.09	-0.05	0.21
Ar_II	0.14	1.00	-0.01	0.50	0.02	-0.08
Mi_III	0.04	-0.01	1.00	0.08	-0.06	-0.02
Ar_III	0.09	0.50	0.08	1.00	0.10	0.11
Mi_V1	-0.05	0.02	-0.06	0.10	1.00	-0.02
Mi_V5	0.21	-0.08	-0.02	0.11	-0.02	1.00

The features' orthogonality is also confirmed by the fact that each feature from the constellation has significant factor loading for a particular factor (see Tables 2a,b). The first 3 factors (F1, F2, F3) have eigenvalues higher than 1, both for the measurements in T1 and T2 and they accumulate 68% and 64% of the total data variation, respectively (Tables 3a,b) that is a sign for their high potential for person identification.

Table 2a. Factor loadings (Varimax normalized) for the selected 6 features, measured in T1.

Var	Factor					
	F1	F2	F3	F4	F5	F6
Mi_II	0.03	0.15	0.13	-0.07	0.98	0.03
Ar_II	-0.97	-0.01	-0.12	0.06	-0.03	0.21
Mi_III	0.01	0.99	0.03	0.02	0.15	0.02
Ar_III	-0.21	0.02	0.10	0.11	0.03	0.97
Mi_V1	0.05	-0.02	-0.06	-0.99	0.07	-0.11
Mi_V5	0.12	0.03	0.98	0.06	0.13	0.10

Table 2b. Factor loadings (Varimax normalized) for the selected 6 features, measured in T2.

Var	Factor					
	F1	F2	F3	F4	F5	F6
Mi_II	0.07	0.11	-0.02	0.03	0.99	0.03
Ar_II	0.96	-0.05	0.01	-0.00	0.07	0.26
Mi_III	-0.01	-0.01	-1.00	0.03	0.02	0.04
Ar_III	0.26	0.07	-0.05	-0.06	0.04	0.96
Mi_V1	0.00	-0.01	0.03	-1.00	-0.03	0.05
Mi_V5	-0.05	0.99	0.01	0.01	0.11	0.06

Table 3a. Eigenvalues and cumulative variance of F1-F6 for the measurements in T1.

Factor	Eigen value	Total Variance	Cumulative eigenvalue	Cumulative variance
F1	1.53	25.45 %	1.53	25.45 %
F2	1.50	24.98 %	3.03	50.44 %
F3	1.06	17.60 %	4.08	68.03 %
F4	0.87	14.58 %	4.96	82.61 %
F5	0.58	9.72 %	5.54	92.33 %
F6	0.46	7.67 %	6.00	100.0 %

Table 3b. Eigenvalues and cumulative variance of F1-F6 for the measurements in T2.

Factor	Eigen value	Total Variance	Cumulative eigenvalue	Cumulative variance
F1	1.58	26.29 %	1.58	26.29 %
F2	1.20	19.94 %	2.77	46.23 %
F3	1.05	17.58 %	3.83	63.81 %
F4	0.94	15.73 %	4.77	79.53 %
F5	0.79	13.22 %	5.57	92.75 %
F6	0.43	7.25 %	6.00	100.0 %

We aim at clusterization of the population data (separately for T1 and T2) in order to find coincidence between the individual feature constellation for each subject in the feature space of T1 and T2. For that purpose, we have studied the data variation over the population in T1 and T2, using the projections of the selected six features in the planes formed by the first 3 factors (F1xF2, F1xF3, F2xF3). The examples in Figures 2a,b illustrate the relative proximity between the feature projections for T1 and T2 in the factor plane F1xF2.

Aiming to demonstrate the clusterization of the population data, we have analyzed the feature projections in the three factor planes (F1xF2, F1xF3, F2xF3) and we have selected the three features that are far apart – Ar_II, Mi_V1, Mi_V5. Figures 3a,b illustrate the data distribution in the 3D-space of Ar_II, Mi_V1, Mi_V5 for T1, T2, respectively. The similarity between the spatial distribution of the data is evident even in this case of reduced number of features applied for clusterization.

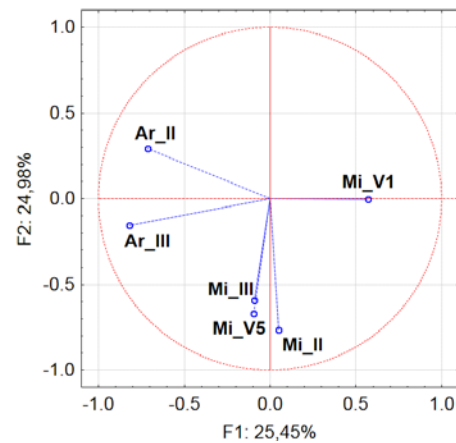


Figure 2a. Projection of the features measured in T1 on the factor-plane (F1xF2).

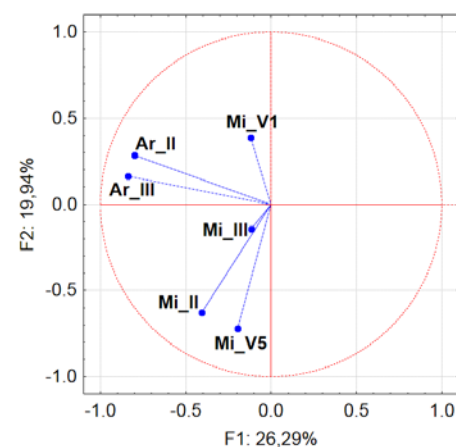


Figure 2b. Projection of the features measured in T2 on the factor-plane (F1xF2).

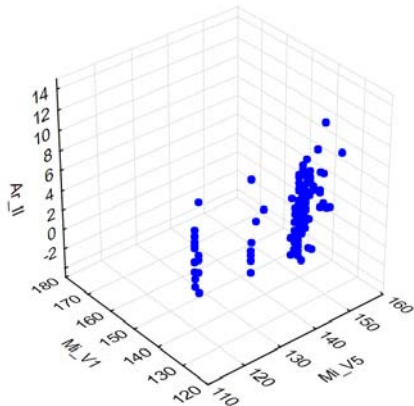


Figure 3a. 3D scatterplot of features Mi_V5 vs. Mi_V1 vs. Ar_II, measured in T1.

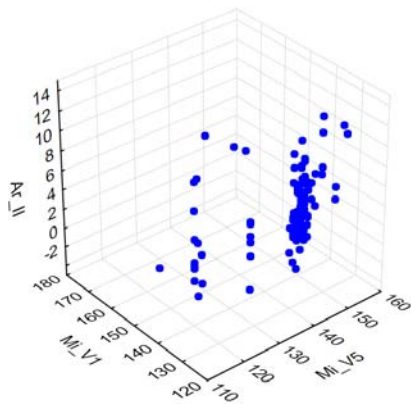


Figure 3b. 3D scatterplot of features Mi_V5 vs. Mi_V1 vs. Ar_II, measured in T2.

As a result from the performed analyses, we could speculate that the QRS morphological features are the most adequate for person identification. However, further investigations are needed over the T-wave features and QT/RR ratio in order to use their potential to improve the clusterization of the population in T1 and T2.

5. Conclusions

The presented study has the characteristics of an exploratory data analysis. The achieved results, even obtained over a reduced feature set, show that it is advisable to focus the investigations towards appropriate clusterization of the data measured in T1 and T2. This would be followed by testing the potential of the designed clusters to match a subject in T1 and T2.

Acknowledgements

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

References

- [1] Biel L, Pettersson O, Philipson L, Wide P. ECG analysis: a new approach in human identification. *IEEE Trans Instrum Meas*, 2001;50(3):808-12.
- [2] Israel S, Irvine J, Cheng A, Wiederhold M, Wiederhold B. ECG to identify individuals. *Pattern Recognition*, 2005;38:133-42.
- [3] Wang Y, Agrafioti F, Hatzinakos D, Plataniotis K. Analysis of human electrocardiogram for biometric recognition. *EURASIP Journal on Advances in Signal Processing*. 2008;article ID 148658.
- [4] Plataniotis K, Hatzinakos D, Lee J. ECG biometric recognition without fiducial detection. In *Proc. on Biometrics Symposium BSYM'06*, 2006;19-21.
- [5] Agrafioti F, Hatzinakos D. ECG biometric analysis in cardiac irregularity conditions. *Signal Image & Video Processing*, 2009;3:329-43.
- [6] Poree F, Gallix A, Carrault G. Biometric identification of individuals based on the ECG. Which conditions? *Comp in Cardiol*, 2011;38:761-4.
- [7] Jekova I. Potential of the peripheral electrocardiogram for personal verification/identification. *Proc Bulg Acad Sci*, 2015;68(4):497-504.
- [8] Wubbelier G, Stavridis M, Kreiseler D, Boussejot R, Elster C. Verification of humans using the electrocardiogram. *Pattern Recogn Lett*, 2007;28:1172-5.
- [9] Sidek K, Khalil I, Smolen M. ECG biometric recognition in different physiological conditions using robust normalized QRS complexes. *Comp in Cardiol*, 2012;39:97-100.
- [10] Zhao Z, Yang L, Chen D, Luo Y. A human ECG identification system based on ensemble empirical mode decomposition. *Sensors*, 2013;13:6832-64.
- [11] Maggi S, Zuchetto M, Grigoletto F, Baldereschi M, Candelise L, Scarpini E, et al. The Italian Longitudinal Study on Aging (ILSA): design and methods. *Aging Clin Exp Res* 1994;6:464-73.
- [12] Matveev M, Tomov L, Tomov I, Daskalov I. Lead Selection for ECG Screening. *Adv Cardiol*, 1977;19:148-9.
- [13] Christov II. Real time electrocardiogram QRS detection using combined adaptive threshold. *Biomed Eng Online* 2004;3(28)
- [14] Christov I, Jekova I, Bortolan G. Premature ventricular contraction classification by the K-th nearest neighbours rule. *Physiol Meas*, 2005;26:123-30.
- [15] Christov I, Simova I. Q-onset and T-end delineation: Assessment of the performance of an automated method with the use of a reference database. *Physiol Meas* 2007; 28:213-21.
- [16] Batchvarov V, Ghuran A, Smetana P, Hnatkova K, Harries M, Dilaveris P, et al. QR-RR relationship in healthy subjects exhibits substantial intersubject variability and high intrasubject stability. *Am J Physiol Heart Circ Physiol*, 2002; 282: 2356-63.

Address for correspondence.

Irena Jekova
 Institute of Biophysics and Biomedical Engineering
 Acad. G. Bonchev str., bl.105, 1113, Sofia, Bulgaria
irena@biomed.bas.bg