Analysis of Human Electrocardiogram for Biometric Recognition Using Analytic and AR Modeling Extracted Parameters

B. Vuksanovic and M. Alhamdi

***Abstract*—In this paper, a new approach for automatic analysis of single lead ECG for human recognition is proposed and evaluated. Following the pre-processing step, the ECG stream is partitioned into separate windows where each window includes single beat of ECG signal. After successful QRS detection, various temporal, amplitude and AR coefficients are extracted and used as an input to a classifier in order to identify the individuals. In this work, proposed system has been tested using records from three different publicly available ECG databases. Signal pre-processing techniques, applied parameter extraction methods and some intermediate and final classification results are presented in this paper.**

***Index Terms*—ECG, AR model, biometric, extraction.**

I. Introduction

Biometric recognition provides authentication by identifying each individual based on the biological and physiological signal characteristics. A number of identification methods have been investigated in the last decades [1], using physical features such as finger prints, face images [2] and biological signal behaviour such as ECG [3]. Analysis of ECG signals as a biological tool for individual recognition has become an active research field in the recent years[4], [5]. Validity of using ECG as a biometric tool is supported by the fact that ECG signal belonging to each individual has certain unique features [6] which can be used to distinguish it from other ECG signals. Differences between ECG signals are usually caused by the variability of heart position and orientation relative to the ribs (the ribs being the reference clinically used to place the precordial electrodes on the human body), which are highly variable among different persons. Other differences can be related to body habitus [7] sex, age, length, and weight of the subjects.

The signal classification is usually considered in the light of selection, extraction and classification of extracted features. High recognition rate has been achieved with the approach based on the autocorrelation (AC) in conjunction with discrete cosine transform (DCT) [3]. Proposed method does not require any waveform or fiducialpoint detections but AC and DCT are computationally demanding operations and require long ECG records for each patient or individual to identify them successfully. A method known as Pulse Active Width (PAW) is implemented on ECG for biometric authentication [8]. The results of this approach have indicated that PAW yields equivalent performance in terms of accuracy compared to conventional temporal and amplitude feature extraction methods. Even though, PAW is complicated process which needs powerful digital signal processors to overcome the time delay.

In this paper, a new approach for automatic analysis of single lead electrocardiogram (ECG) for human recognition and individual identification is proposed. This approach depends on on analytic (Amplitude, Time and Width) and modelling (AR) features extracted from the ECG beat. Obtained results indicate high level of accuracy and shorter processing time needed to identify the individuals. Eighteen analytic and modelling features are extracted to identify individuals and k nearest neighbour (knn) classification algorithm applied in order to classify those features and evaluate the proposed approach. ECG feature selection and extraction using AR modelling has recently been used [9] resulting in accurate classification of various arrhythmia and ventricular arrhythmia conditions.

The remainder of this paper is organized as follows. Section II gives a brief description of the techniques used in the pre-processing phase to clean ECG signals of noise and other artefacts. Section III provides a review of QRS detection methods used in this work. Feature selection and extraction methods are discussed in Section IV whilst Section V contains experimental results and discussion of those results. Conclusions are presented in Section VI.

II. Pre-Processing Phase

Variations in ECG describe the electrical activity of the heart and are related to the electrical flows inside and around the heart. ECG signal provides information about morphology, heart rate and rhythm. Typical ECG beat contains five waves (P, Q, R, S and T). ECG signal is recorded by attaching electrodes to different places on the skin, such as chest, legs, arms and neck [10]. The collected ECG data usually contain noise components of low-frequency caused by driftline wonder and a higher frequency components caused by power line interferences [11]. The presence of noise will corrupt the signal and make the feature extraction and classification process more difficult and less accurate.

A number of research papers have discussed the removal of noise and power line interference from the ECG signals. In [12] a non-linear adaptive method to eliminate power line interference from the ECG signals is presented. The wavelet coefficient threshold based hyper shrinkage function was used in [13] to detrend the raw ECG signals. In [14] a

Manuscript received March 19, 2014; revised June May 22, 2014.

B. Vuksanovic and M. Alhamdiis are with the University of Portsmouth, Anglesea Building, Anglesea Road, Portsmouth PO1 3DJ, UK (e-mail: branislav.vuksanovic@port.ac.uk, mustafa.alhamdi@port.ac.uk).

simplified lattice based adaptive IIR notch filter has been suggested to remove power line interference. Other technique proposed the use of digital FIR filters [15] for the power line noise reduction. [16] suggested a method of noise removal dependent on a non-linear wavelet and wavelet packet and the extended Kalman Filter (EKF) based on a nonlinear dynamic model used for the generation of synthetic ECG signals has also been used to clean ECG signals [17]. Bayesian wavelet shrinkage denoising approach for high resolution ECG (HRECG) filtering has been used in [18]. This approach is depended on three basic steps: the dyadic Wavelet Transform (WT) computation, the shrinkage of the wavelet coefficients using adaptive Bayesian rules, and the reconstruction of the denoised signal through the inverse WT.

In this work, method proposed in [19] has been applied to decrease the effect of driftline wonder and preserve the information about the physiology of an individual's ECG. The raw ECG signals are first downsampled to the lowest sampling frequency of the ECG signals from the set (1000 Hz in this case) and then filtered through the 6th order Butterworth filter. The output of the Butterworth filter is then passed through power line interferences filer. Another Butterworth filter – 10th order low pass type is used to remove high frequency components and noise. Fig. 1 shows the ECG signalbefore and after the filtering stage.

III. QRS Detection

The QRS complex is the most important feature of the ECG signal. Without the accurate knowledge of the QRS complex location, P and T waves are hard to detect and distinguish from each other. Most of the QRS detection methods depend heavily on filtering stage followed by averaging according to a threshold value [20]. This threshold value is used to distinguish between noise signal and the QRS complex and is usually chosen according to the peak height or peak location of ECG signal [21]. A number of research papers have investigated the problem of accurate QRS complex detection depending on the applications of wavelet filter banks [22] and the modified p-spectrum to detect heart beats in ECG signals [23]. The short-time Fourier transform (STFT) was also employed after ECG filtering stage in order to detect QRS complex [24].

A number of commercial systems have been designed and implemented to perform signal processing tasks such as 12-lead off-line ECG analysis, Holter tape analysisand real-time patient monitoring [25]. All these applications require an accurate detection of QRS complex in the ECG signals.

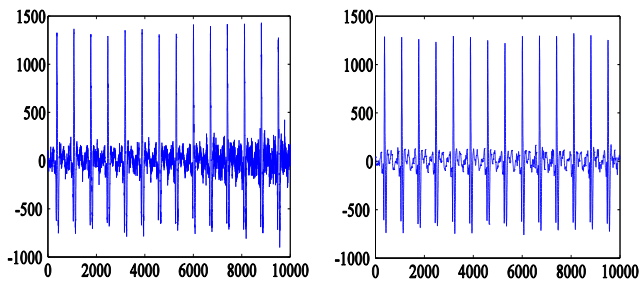


Fig. 1. Raw and filtered ECG signal.

P and T waves occur before and after the QRS complex respectively. The frequency components of QRS complex range between 10 Hz and 25 Hz. Most of the QRS detection algorithms use a filtering stage followed by the actual detection in order to attenuate unwanted signal components and artefacts. In this work, QRS detection plays an important stage as it is then used to localize the R peaks of ECG stream of beats and main indicator for the window partitioning process.

Following the location of R peaks in the analysed ECG stream, window size estimation is applied in order to extract windows such that each window contains single beat of ECG. Mathematically, the window size is estimated based on the heart rate variability (HRV) of each ECG stream. Fig. 2 shows four different samples from the ECG database. In this work, QRS detection is adapted and used to pinpoint the exact QRS complex position in the ECG stream. Filter bank method employs a bank of linear phase filters to decompose the ECG signal into subbands with uniform frequency bandwidths in order to account for the ECG signal energy distribution in the frequency domain [22]. Fig. 3 shows the R peaks detected from each ECG beat.

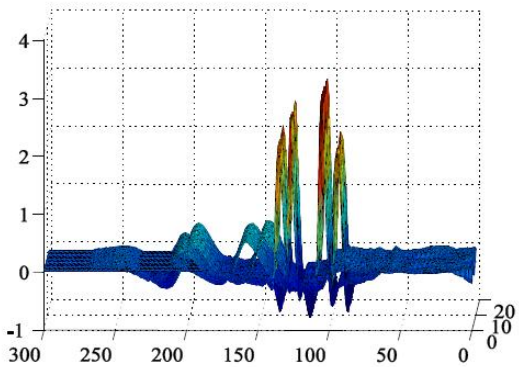


Fig. 2. Window extraction of four samples of ECG signals.

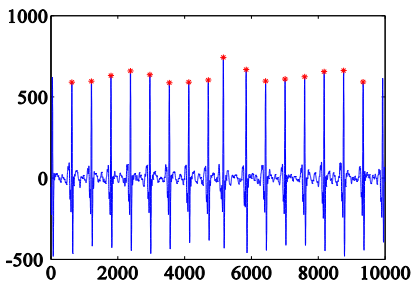


Fig. 3. Detected QRS complex.

IV. Methodology

Biometrics-based human identification is essentially a patient recognition problem which involves pre-processing, feature extraction and classification stages. The electrocardiogram is an emerging biometric modality that has seen about 13 years of development reported in peer-reviewed literature [26] and as such deserves a systematic review and discussion of the associated methods and findings. In particular, the categorization of methodologies in ECG based biometry relies on the feature

extraction and classification schemes. Finally, comparative analysis is presented to estimate performance of this approach of ECG biometric system. The comparative study includes the cases where training and testing data sets come from the same and different partition of ECG signal measured on different days.

Recently, cardiovascular signals have been studied for use in identity recognition problems using electrocardiography [27]–[29]. ECG feature extraction plays a significant role in diagnosing most of the cardiac diseases. One cardiac cycle of an ECG signal consists of the P-QRS-T waves. Feature extraction scheme proposed in this work uses amplitudes, time intervals and AR parameters of the ECG signal for subsequent analysis and identification. For the purpose of comparative study, the system proposed in this work follows the procedure of extracting eighteen features from each ECG beat listed in Table I.

In this work, identification has been attempted using extracted analytic features (amplitude, time and width) and modelling features (AR parameters). The performance of the proposed method is evaluated based on the subject rate recognition which is determined by the number of correctly classified subjects of the test database using knn classification algorithm.

TABLE I: EXTRACTED FEATURES FROM ECG PEAK

Temporal	PQ	Time interval between P and Q waves.	PS	Time interval between P and S waves.
	PR	Time interval between P and R waves.	TS	Time interval between T and S waves.
	TQ	Time interval between T and Q waves.	TR	Time interval between T and R waves.
	PT	Time interval between P and T waves.		
Amplitude	P	Amplitude height of P wave.	T	Amplitude height of T wave.
	S	Amplitude height of S wave.	R	Amplitude height of R wave.
	Q	Amplitude height of Q wave.		
Width	QRS	Width in time of QRS.	T	Width in time of T wave.
	P	Width in time of P wave.		
AR coefficients	a_1	First coefficient of AR model.	a_3	Third coefficient of AR model.
	a_2	Second coefficient of AR model.		

A. Analytic Features

Analytic features are usually being extracted by analysing and localizing certain fiducial points from the ECG signals. In order to extract analytic features of ECG window or beat, procedure similar to [30] is followed. As the R peak is already detected using QRS detection, Q, S, P and T amplitudes are localized by finding local maxima and minima separately around R position of each window. Fig. 4(a) to Fig. 4(d) show the P, Q, S and T analytic features indicate with "*" along the ECG stream.

Feature selection and extraction is the key stage in the ECG based human identification system. Previously proposed methods [27]–[29] for ECG-based identity recognition use temporal attributes as well as amplitude distances between detected fiducial points. In this work temporal, amplitude and width distances between indicated points have been used. All extracted parameters are normalized by the maximum and minimum two points of P and T waves of each ECG beat. $P_{\min}$, $P_{\max}$, $T_{\min}$ and $T_{\max}$ are used as indication points to provide less variability with respect to heart rate. These points on ECG wave are used to estimate P and T width. The time interval between $P_{\min}$ and $P_{\max}$ is the P wave width. Similar method is used to estimate T width, the interval between $T_{\min}$ and $T_{\max}$.

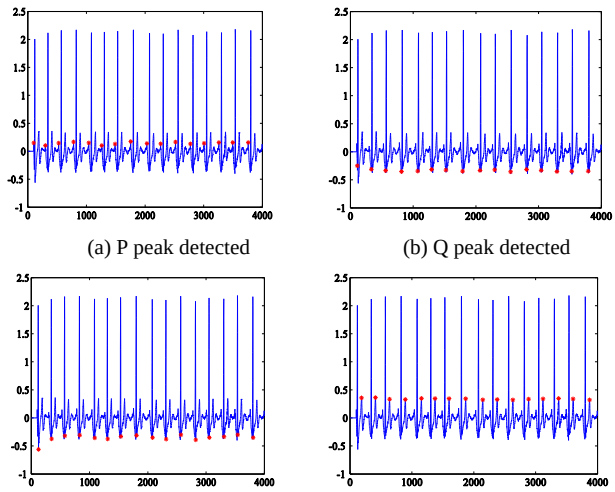
B. Modelling Features

In addition to analytic ECG features described in the previous section, proposed system uses modelling features obtained through the modelling of two or more successive ECG beats. To achieve this, a discrete form of an autoregressive (AR) signal model of order p, AR(p), is applied [31] to model the measured ECG signal. The estimated coefficients are then added to a feature set used in the final, classification stage of the proposed system.

Using the proposed AR model, a sampled signal sequence $y(n)$ can be represented by the relationship:

$$y(n) = a_1 y(n-1) + a_2 y(n-2) + \ldots + a_p y(n-p) + \varepsilon(n) \quad (1)$$

where $a_k\,(k = 1, 2, \ldots, p)$ are the model coefficients used in the classification process and the $\varepsilon(n)$ is a white noise series, innovation process with zero mean and variance σ^2.



(a) P peak detected (b) Q peak detected

(c) S peak detected (d) T peak detected

Fig. 4. Analytic points detected.

Accurate modelling of two successive beats from ECG signal can be achieved using a 3rd order AR model as described in [9]. Individual section of a pre-processed ECG signal in red and the corresponding AR(3) model in blue are shown in Fig. 5.

V. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed methods,

experiments are conducted using signals from three sets of publicly available ECG databases: PTB [32], MIT-BIH [33] and Milano [34]. The PTB database is provided by the National Metrology Institute of Germany and contains 549 records from 294 subjects. The MIT-BIH database contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects studied by the BIH Laboratory between 1975 and 1979. Each record from this database consists of conventional 12 and 3 leads ECG. The duration of the recordings vary for each subject.

A subset of 13 subjects was selected to test proposed method. The selected subjects did not exhibit significant arrhythmias. Some of ECG records present in the database contained significant amount of noise and other artefacts significantly reducing the valid heartbeat information, and were not used in our experiment.

Since the database only contains one record for each subject, records were partitioned into two halves. The first half was then used as a main "training set" and the second half as "test set" for classification trials. The training and test sets should in the final evaluation of the proposed system be an ECG recordings made on different days for the same subject. The feature matrix for classification is formed of 18 extracted parameters, described in the earlier sections and 14 windows or beats for each subject. By applying duration of ECG stream of 8 seconds length with no overlapping, different number of beats are separated into windows where each window has one beat. Several different window lengths tested in this work show approximately the same classification performance as long as full multiple beats are present in the extracted window. Figure 6 shows the scatter plot of extracted parameters including amplitude, time, width and AR coefficients for both parts (main and test) of each ECG subject in the database.

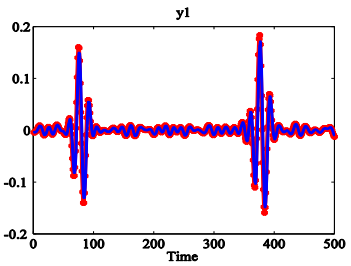


Fig. 5. Modelled and original ECG signal.

For performance and accuracy estimation of the proposed system, knn classification algorithm has been applied using "knnclassify" function from classification toolbox in Matlab. This function can be called using syntax: "C=knn(main, test, classes)", C shows the classification factors that express algorithm performance as in Table II. Main and test sets are different 8 sec time interval of each subject and classes are the number of patients included in this test. Both main and test sets represent the ECG recordings of different days of the same subject. Main, training set includes 18 extracted parameters and 14 windows or beats for 13 subjects, where test set includes another 18 parameters from other 14 windows or beats for the same 13 subjects. The k-nearest neighbour is one of the statistical classification algorithms used for classifying objects based on closest training examples in the feature space. knn is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. The knn is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbours, with the object being assigned to the class most common amongst its k nearest neighbours. The obtained results presented in Table II show that 100% correct rate is reachable by using all the extracted features. This high accuracy of identification comes from the matching parameters between the main and test sets of extracted features for the 13 subjects used in this experiment. This results in a high percentage of negative productive value when the test comes to use smaller number of parameters to identify individuals for example, just amplitude and AR parameters. Negative productive value indicates the performance of a diagnostic testing procedure. It is defined as the proportion of subjects with a negative test result who are correctly classified.

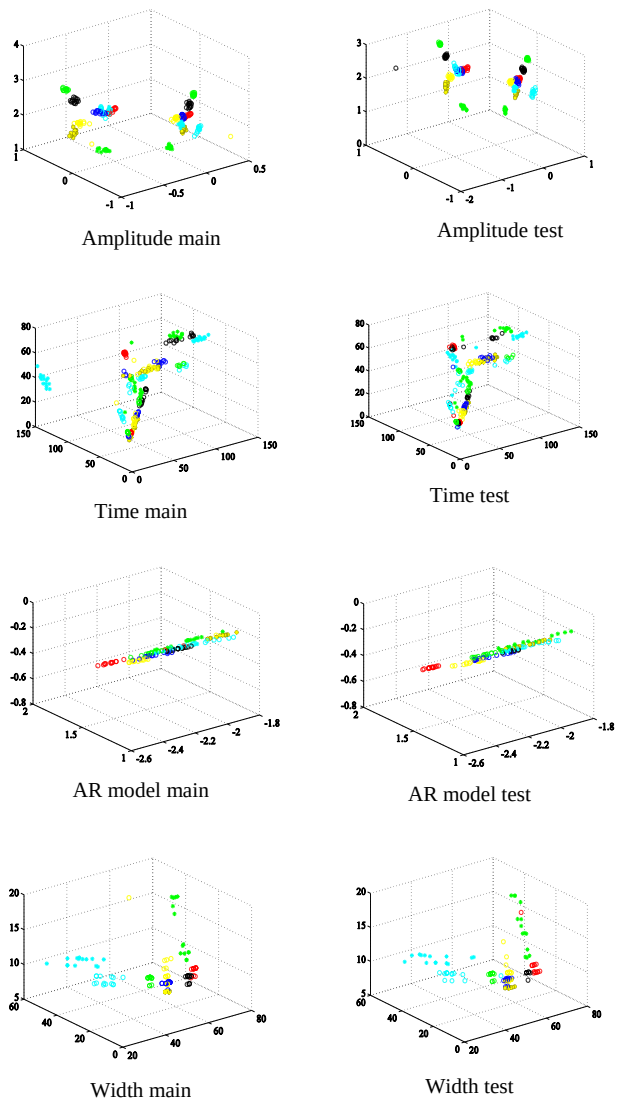


Fig. 6. Extracted parameters in 3D feature space.

Table II indicates the correct rate and comparison of experimental results for different sets of parameters used in main and test sets combination. Suggesting width parameters alone to classify 13 patients illustrate very low correct rate (55.56%).

Even though amplitude and AR parameters can reach a

high rate of classification (98.89%) percentage of negative test is high. By using all extracted features (Amplitude, Time, Width and AR model) to identify individual, it is possible to reach 100 % accuracy of classified subjects with a reasonable negative test value.

VI. Conclusion

A biometric system for automatic analysis of a single lead electrocardiogram (ECG) for human identification is presented in this work. The first stage of this system consists of a band-pass Butterworth filter used to remove noise and other artefacts present in the raw ECG signal. The R peaks of an ECG stream are then localized by filtering a signal through the bank of filters and analysing each filtered signal component individually. The individual heart beats of an ECG record are aligned by the R peak position and truncated by a window size based on the heart rate variability (HRV) of individuals' ECG. The temporal, amplitude, width and AR parameters are then extracted from each ECG signal and used to identify individual ECGs. An approach of modelling two or more successive ECG beats, using a discrete form of an autoregressive (AR) signal model of order p is applied to extract AR modelling features, AR(p).

TABLE II: Extracted Features from ECG Peak

Classification Type	Correct Rate	Negative Productive Value (NP)
Width	55.56%	33%
Amplitude	98.89%	86%
Amplitude and AR	95%	87%
Amplitude, Time, Width and AR model	100%	17%

A subset of thirteen subjects was selected in this work in order to test proposed system. Extracted parameters for each subject are fitted into knn classification algorithm to classify and eventually identify "test" parts of each signal. 100% identification accuracy for thirteen different ECG recordings has been achieved in this project when each ECG signal is split into "train" and "test" section. This leads to a conclusion that the proposed ECG based biometric system might lead to an accurate identification in some practical applications.

References

[1] Y. L. Singh and P. Gupta, "ECG to individual identification," in *Proc. 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems, BTAS 2008*, pp. 1-8, 2008.

[2] S.-H. Lin, "An introduction to face recognition technology," *Informing Science special Issue on Multimedia Informing Technologies*, part. 2, vol. 3, no. 1, pp. 1-7, 2000.

[3] Y. Wang, F. Agrafioti, and D. Hatzinak, "Analysis of human electrocardiogram," *Hindawi Publishing Corporation, EURASIP Journal on Advances in Signal Processing*, p. 11, 2008.

[4] L. Biel, P. Wide, L. Philiphson, and O. Patterson, "ECG analysis: A new approach in human identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 3, pp. 808-812, 2001.

[5] B. K. Wiederhold, L. W. Gavshon, and J. M. Irvine, "Heart rate variability: A new biometric for human identification," in *Proc. the International Conference on Artificial Intelligence (IC-AI '01)*, 2001, pp. 1106–1111.

[6] R. Hoekema, "Geometrical aspects of the interindividual variability of multilead ECG recordings," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 5, pp. 551–559 , 2001.

[7] R. H. Selvester, M. Sanmarco, R. Sarma, and I. Eisenstein, "The electrocardiogram in obesity," *J. Electrocardiol.*, vol. 15, no. 2, pp. 115–118, 1982.

[8] S. Safie, J. Soraghan, and L. Petropoulakis, "ECG biometric authentication using Pulse Active Width (PAW)," *2011 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS)*, pp. 1-6, 2011.

[9] B. Vuksanovic and M. Alhamdi, "ECG based system for arrythmiadetection and patient identification," *ITI2013*, pp. 1-7, 2013.

[10] D. Gari and E. Macshary, "Biomedical engineering: Advanced methods and tools for ECG data analysis," *Artech House, Inc.*, 2006.

[11] I. Duskalov, "Developments in ECG acquisition, preprocessing, parameter measurement, and recording," *IEEE Engineering in Medicine and Biology Magazine*, vol. 17, no. 2, pp. 50-58, 1998.

[12] A. Ziarani and A. Konrad, "Non linearadaptive method of elimination of power line interference in ECG signals," *IEEE Transactions on Biomedical Eng*., vol. 49, no. 6, pp. 540-544, 2010.

[13] S. Pooranchandra and N. Kumaravel, "A novel method for elimination of power line frequency in ECG signal using hyper shrinkage functions," *Digital Signal Processing*, vol. 18, no. 2, pp. 116-126, 2010.

[14] S. S. Dhillon and S. Chakrabarti, "Power line interference removal from electrocardiogram using asimplified lattice based adaptive IIR notch filter," in *Proc. the 23rd Annual EMBS International conference*, Istanbul, Turkey, 2001.

[15] M. Chavan, R. Aggarwala, and M. Uplane, "Interference reduction in ECG using digital FIR filters based on Rectangular window," *WSEAS Transactions on Signal Processing*, vol. 4, no.5, pp. 340-49, 2008.

[16] P. Tikkane, "Non-linear wavelet and wavelet packet denoising of electrocardiogram signal," *Biological Cybernetics*, vol. 80, no. 4, pp. 259-267, 1999.

[17] R. Sameni, M. Shamsollahi, and C. Jutten, "Filtering electrocardiogram signals using the extended Kalmanfilter," in *Proc. IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai*, pp. 1-4, 2005.

[18] M. Popescu, P. Cristea, and A. Bezerianos, "High resolution ECG filtering using adaptive Bayesian wavelet shrinkage," in *Proc. IEEE conference in Computers in Cardiology*, pp. 401-404, 1998.

[19] S. K. Jagtap, "The impact of digital filtering to ECG analysis: Butterworth filter application," in *Proc. 2012 International Conference on Communication, Information & Computing Technology (ICCICT)*, 2012, pp. 1-6.

[20] H.-Y. Zhou and K.-M. Hou, "Embedded real-time QRS detection algorithm for pervasive cardiac care system," in *Proc. 9th International Conference on Signal Processing*, 2008, pp. 2150-2153.

[21] J. F. Moraes, "QRS complex detection algorithm using electrocardiogram leads," *Computers in Cardiology*, pp. 205-208, 2002.

[22] G. Kokturk, "A real-time simulated QRS detection systemconstructed using wavelet filteringtechnique," *IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pp. 281- 284, 1998.

[23] P. Qiuand and K. J. R. Liu, "A robust method for QRS detection based on modified p-spectrum," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 501-504, 2008.

[24] N. Uchaipichat and S. Inban, "Development of QRS detection using short-time fourier transform based technique," *IJCA Journal*, pp. 7-10, 2010.

[25] B.-U. Kohler, C. Hennig, and R. Orglmeister, "The principles of software QRS detection," *Engineering in Medicine and Biology Magazine*, vol. 21, no. 1, pp. 42 - 57, 2002.

[26] I. Odinaka, A. Kaplan, J. O'Sullivan, and L. Po-Hsiang, "ECG biometric recognition: A comparative analysis," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1812-1824, 2012.

[27] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "ECG analysis: A new approach in human identifications," *IEEE Trans. Instrum. Meas.*, vol. 50, no. 3, pp. 808–812, 2001.

[28] M. Kyosoand and A. Uchiyama, "Developmentofan ECG identification system," in *Proc. 23rd Ann. EMBS Int. Conf.*, 2001.

[29] B. K. Wiederhold, L. W. Gavshon, and J. M. Irvine, "Heart rate variability: A new biometric for human identification," in *Proc. Int. Conf. Artificial Intelligence (ICAI 2001)*, 2001.

[30] D. Wiederhold, K. Wiederhold, and A. Israel, "ECG to identify individuals," *Pattern Recognition*, pp. 133-142, 2005.

[31] C. Anderson, E. Stolz, and S. Shamsunder, "Discriminating mental tasks using EEG represented by AR models," in *Proc. IEEE 17th Annual Conference on Engineering in Medicine and Biology Society*, 1995, vol. 2, pp. 875-876.

[32] A. Goldberger, L. Amaral, L. Glass, and Hausdorff, "Components of a new research resource for complex physiologic signals," *Physio Bank, Physio Toolkit, and Physio Net, Circulation*, vol. 101. no. 23, p. e21.

[33] R. Mark, "The MIT-BIH arrhythmia database on CD-ROM and software for use with it," in *Proc. Computers in Cardiology*, 1990, pp. 185-188.

[34] D. Milano, *Politecnicobiosignalsarchives on CD-ROM*, (Copyright (C) Politecnico. 1992), 1992.

Branislav Vuksanovic graduated from the University of Belgrade, Serbia with degree in electrical and power engineering. He holds MSc degree in measurement and instrumentation from South Bank University, London and a PhD in active Noise control from the University of Huddersfield, UK. Previously, he worked as a project engineer for Croatian Electricity Board in Osijek, Croatia. During his academic career he worked as a research fellow at Sheffield and Birmingham Universities on optical brain imaging and medical video compression projects. Currently he works as a senior lecturer at the University of Portsmouth, School of Engineering. He has published papers in the field of active noise control, biomedical signal processing and pattern recognition for intrusion detection and knowledge based authentication. He published one book in Digital Electronics and Microcontrollers field. His current research interests are in the application of pattern recognition techniques for power systems and analysis of ground penetrating radar and ECG data.

M. Alhamdi was born in Basra, Iraq in 1985. He gained a BSc degree in electronic and communication engineering in 2007 and a MSc degree in communication systems analysis in 2010 both from the University of Portsmouth, UK.

He is currently pursuing a PhD research at the same institution in the area of ECG signal processing and analysis. So far, he has published two conference and two journal papers resulting from this research work. Previously, heworked in Mobile Communication System (CDMA2000), and in a heart clinical center for pacemaker and ICD transplants surgery. Mr. Alhamdi is a member of IEEE institute.